SECOND ANNUAL

Arthropod Genomics

S Y M P O S I U M

April 10-13, 2008 • Kansas City

POSTER ABSTRACTS

1

SIBO: The Social Insect Behavior Ontology

*Abbas, Ilma; Smith, Christopher D.
San Francisco State University, Dept. of Biology, San Francisco, CA 94132

Invasive social insects, including the red imported fire ant, Africanized honeybee, and Argentine ant, exhibit unusually cooperative and aggressive behavior in their non-native ranges and result in billions of dollars of agricultural and ecological damage. Since pesticides are ineffective at controlling these pests, a better understanding of the genes involved in social behavior may provide insight into new ways to control these species. One challenge to this approach is that behaviors are often incompletely or qualitatively described and often presented in textbooks or scientific articles that are poorly accessible. Behavior ontologies are useful tools to organize the data in the scientific literature into a form suitable for advanced data mining. Ontologies can be used both to standardize behavior nomenclature and to map the relationships between behaviors and associated genes or chemicals so that they are more amenable to computer-aided analyses. While, the Gene Ontology (GO) has eased interspecies comparison of the 'molecular functions' in many annotated genomes, there are currently only 14 terms for 'social behavior' (GO:0035176). Therefore, we started development of SIBO, the social insect behavior ontology. SIBO contains terms for complex social behaviors exhibited by different ant and bee species, the semiochemicals involved in group communication, and the anatomical parts where these chemicals are synthesized, stored, or sensed. We have already developed a map of ant anatomical parts and their interrelationships as well as a catalog of semiochemicals used for communication. Currently, we are curating scientific literature to create ontologies for social behaviors in honeybees and the Red Harvester Ant, *Pogonomyrmex barbatus,* including 'colony defense', 'foraging', and 'patrolling.' We propose to integrate SIBO with existing gene function and chemical ontologies to better understand the gene networks that underlie social behavior and to identify potential new targets for biological control.

2

An FPC-Based Integrated Genetic and Physical Map of the Hessian Fly Genome

*Aggarwal, R.; Benatti, T.; Fellers, J. P.; Chen, M-S.; Zhao, C.; Schemerhorn, B. J.; Stuart, J.
(RA, TB, CZ, JS) Department of Entomology, Purdue University, West Lafayette, IN 47907; (JPF) USDA-ARS and Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; (MSC) USDA-ARS and Department of Entomology, Kansas State University, Manhattan, KS 66506; (BJS) USDA-ARS and Dept. of Entomology, Purdue University, West Lafayette, IN 47907

The Hessian fly (*Mayetiola destructor*) is an important pest of wheat (*Triticum* spp.). We have developed a physical map of the Hessian fly genome by employing high-throughput fingerprinting of bacterial artificial chromosomes (BACs). The BAC DNA was digested with five different enzymes and labeled with SNaPshot Primer Extension Kit. The restricted fragments were sized using ABI 3730 capillary DNA analyzer. DNA fingerprints of 13,614 BAC clones were generated that provided ~12-X coverage to the Hessian fly genome. The BAC fingerprints were assembled into contigs using FPC (v. 8) software at a cutoff value of $1e^{-29}$ and a tolerance value of 5. These analyses assembled 264 contigs with 4542 BACs (~4.3-X coverage) ranging in length from 8 to 73 BACs per contig.

The Hessian fly polytene chromosomes were divided into 26 regions numbered from A to Z. Each of the 264 contigs was assigned to a particular region of the polytene chromosome using fluorescent *in situ* hybridization (FISH). Microsatellite (SSR) markers and salivary gland derived Expressed Sequence Tags (ESTs) were integrated onto their respective contigs. In addition, each of the BACs has been end-sequenced. The results have been made publicly available as an WebFPC: http://genome.purdue.edu/WebAGCoL/Hfly/WebFPC/. The results provide a solid foundation for a future Hessian fly genome sequencing effort.

3

ESTs uncover unique salivary enzymes of tarnished plant bugs

Allen, Margaret L.
USDA Agricultural Research Service, Biological Control of Pests Research Unit, Building 8, NBCL
59 Lee Rd., Stoneville, MS 38776

The tarnished plant bug, *Lygus lineolaris*, feeds with piercing-sucking (haustellate) mouthparts. When the mouthparts are inserted into plant host tissue, enzymes are injected. These salivary enzymes partially digest plant tissues so that the insect can suck out nutrients. *L. lineolaris* is common in North America and has a wide host range, and is not susceptible to the *Bacillus thuringiensis* (Bt) crystal proteins used to engineer pest-resistant crops. Thus, pest control for Lygus infestations is limited to chemical applications, and resistance in some populations is increasing dramatically. Alternative and novel methods to control Lygus are needed, so gene sequencing efforts have been initiated. Two small EST libraries were sequenced and deposited in NCBI GenBank. The first of these was from female adults, and many of the identifiable sequences were associated with reproduction: vitellogenins, for example. To seek genes associated with feeding and digestion, male nymphs at the final (5th instar) nymphal stage of development were used for a second EST library, and three sequences were targeted for further analysis. One of the primary plant structural constituents that Lygus saliva degrades is pectin, a carbohydrate component of the plant cell wall. Three ESTs from the male nymph library appeared to encode polygalacturonase (PG), a pectin-degrading enzyme most often associated with fungal pathogenicity. The three *L. lineolaris* PG genes were amplified in their entirety by RACE, and found to encode three unique enzymes. Semiquantitative PCR of these genes indicated that they were differentially expressed in all feeding stages. To examine the individual roles of the specific PGs, quantitative PCR was undertaken with PG2 and PG3 in response to different feeding treatments. Results of these studies will be presented.

The results presented here demonstrate the progress of genomic methods towards expansion of applicability from model insects to specific insect problems. Hemimetabolous insects are underrepresented in GenBank, but substantial progress has been made to increase EST submissions in the NCBI dbEST database, particularly within the orders Hemiptera and Orthoptera. While *L. lineolaris* is a member of the order Hemiptera, it is highly taxonomically divergent from the pea aphid, *Acyrothosiphon pisum*, which belongs to the suborder Sternorrhyncha; and from the planthoppers, within the Archaeorrhyncha. Lygus bugs are true bugs (Heteroptera) within the Prosorrhyncha. Within the true bugs, the vast majority of EST submissions are from the Chagas disease vectors in the family Reduviidae, which are blood-feeding insects and therefore expected to have salivary genes distinct from the phytophagous *L. lineolaris,* family Miridae. The current breadth of gene sequence available, sophisticated search tools, plus technology facilitating rapid EST sequencing, make targeted studies like this one possible.

4

Application of insect genomics in the identification of resistance mechanisms and novel target sites

*Alves, Analiza P.; Lorenzen, Marcé D.; Beeman, Richard W.; and Siegfried, Blair D.
(APA, BDS) Department of Entomology, University of Nebraska – Lincoln, Lincoln, NE, USA, 68583-0816; (MDL, RWB) Grain Marketing and Production Research Center, ARS-USDA, Manhattan, KS, USA, 66502.

Bacillus thuringiensis (Bt) is a valuable source of insecticidal proteins for use in insect pest control either in conventional spray formulations or in transgenic crops. However, the evolution of insect resistance in field populations is an important threat to this technology especially with transgenic plants that express the Bt toxins. The western corn rootworm (WCR), *Diabrotica virgifera virgifera* LeConte, one of the insect pests targeted with Bt transgenic plants, has displayed an amazing capacity to develop resistance to most management strategies, including soil insecticides, behavioral resistance to crop rotation, and foliar adulticides. Therefore, a critical need exists for new and effective management options. The objective of this project is to develop a system to identify genes and pathways important as possible target sites and in conferring insecticide resistance to WCR. To conduct this study, disruption of selected genes were obtained through RNA interference (RNAi), based on the synthesis and injection of gene specific double stranded RNA. To validate the RNAi technique in WCR, silencing

of WCR *laccase*, a gene involved in cuticle tanning, was successfully obtained in our lab. Gene silencing can be visualized by sustained lack of pigmentation of injected larvae after molting. Further investigations in larval WCR RNAi were conducted in a midgut specific gene chitin synthase. This research will provide the basis for conducting large-scale identification of genes related to insect resistance and to vital pathways representing potential insecticide target sites.

5

Expression and Analysis of Two Major Cuticle Proteins in *Tribolium castaneum*

*Arakane, Yasuyuki; Muthukrishnan, Subbaratnam; Beeman, Richard W; Kramer, Karl J; Kanost, Michael R (YA, SM, KJK, MRK) Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA; (RWB, KJK) Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA

The insect exoskeleton or cuticle is a complex extracellular matrix consisting of several functional layers. Cuticular proteins (CPs) and the polysaccharide chitin are the major components of the exo- and endocuticular layers or procuticle. During cuticle tanning (sclerotization and pigmentation), CPs are cross-linked by quinones or quinone methides produced by oxidation of catechols catalyzed by the phenoloxidase, laccase. CPs contain a conserved sequence known as the Rebers & Riddiford (R&R) motif, which may function as a chitin-binding domain that helps to coordinate the interaction between chitin fibers and the protein network. More than 100 CP-like genes have been identified in *Drosophila melanogaster*, with a similar number present in the red flour beetle, *Tribolium castaneum*. Different CPs have been found in the different developmental stages and in different regions of the exoskeleton, indicating that specific CPs are required to produce a cuticle with the proper morphology and mechanical properties.

To study the function(s) of insect CPs, we focused on adult elytra of the *T. castaneum*. In protein samples extracted from elytra of newly emerged adults, yielding two highly abundant proteins with apparent sizes of 10 and 20 kDa. To characterize these major proteins further, each was digested with trypsin, and the resulting peptides were analyzed by MALDI-TOF mass spectrometry. Comparison of this result with conceptual trypsinization of the computed proteome of *Tribolium* identified two candidate genes, which we call *TcCP10* and *TcCP20*. We cloned the full-length cDNAs corresponding to these genes. Both proteins belong to the RR-2 family (cuticle DB, http://biophysics.biol.uoa.gr/cuticleDB/) and possess a putative signal peptide sequence. TcCP20 is an ortholog of the *Tenebrio* adult cuticle protein, acp22. Both *TcCP10* and *TcCP20* were mapped to linkage group 3. The expression patterns revealed by real-time PCR showed that the transcript levels of both genes dramatically increase at the pharate adult stage (1 -2 days before molting), and then decline after adult eclosion.

RNA interference (RNAi) was used to investigate the functions of *TcCP10* and *TcCP20*. Specific dsRNAs for each gene were injected into last-instar larvae. Following dsRNA injections, expression of *TcCP10* and *TcCP20* genes was analyzed by real-time PCR and SDS PAGE to evaluate transcript and protein levels. Both genes were substantially and specifically down-regulated at the mRNA and protein levels after RNAi treatment. Injection of ds*TcCP10* or ds*TcCP20* had no effect on larval-larval, larval-pupal or pupal-adult molts. The elytra of the resulting adults, however, were malformed and abnormal. Elytra of ds*TcCP20* adults did not fully elongate and did not extend far enough to cover the abdomen. These elytra were slightly separated and not smooth as compared to those of control insects. Like the ds*TcCP20* phenotype, the resulting adults obtained after injection ds*TcCP10* also exhibited elytral defects, but they were more severe. These elytra were short, wrinkled, bumpy, warped and fenestrated. CP10-deficient insects could not fold their hind wings properly and eventually died approximately 7 days after eclosion, probably because of dehydration. TcCP10 and TcCP20 were also extracted from the pronotum cuticle of *Tribolium*. These results demonstrate that two cuticle proteins, TcCP10 and TcCP20, play critical roles in adult cuticle formation of *T. castaneum*.

6

Expression Profiles and Functional Analysis of Genes Encoding Chitin Deacetylases, Extracellular Matrix-Modifying Proteins in *Tribolium castaneum*

*Arakane, Yasuyuki; Begum, Khurshida; Dixit, Radhika; Park, Yoonseong; Specht, Charles A; Kramer, Karl J; Beeman, Richard W; Muthukrishnan, Subbaratnam
(YA, RD, KJK, SM) Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, Kansas 66506, USA; (KB, YP) Department of Entomology, Kansas State University, 123 West Waters Hall, Manhattan, Kansas 66506; (CAS) Department of Medicine, University of Massachusetts, Worcester, MA 01605, USA; (RWB, KJK) Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, Kansas 66502, USA

In insects chitin, the *β*-(1,4)-linked *N*-acetylglucosamine homopolymer, is one of the major components of the exoskeleton or cuticle. The extracellular matrix (ECM) of this exoskeleton is modified in different ways to make it either very rigid and thick or thin and flexible, or to generate specialized structures such as mandibles with sharp cutting edges. Chitin deacetylases (CDAs) are secreted proteins belonging to a family of extracellular chitin-modifying enzymes that deacetylate chitin to form chitosan, a polymer of *β*-(1,4)-linked D-glucosamine residues. This modification possibly contributes to the affinity of chitin for a variety of proteins distinct from those that bind uniquely to chitin and also to mechanical properties of the cuticle.

Two genes encoding putative chitin deacetylases in *Drosophila*, *serpentine* (CG32209) and *vermiform* (CG8756), play critical roles in shaping of tracheal tubes as well as in regulating the structural properties of epidermal cuticle. By searching Beetlebase (http://bioinformatics.ksu.edu/BeetleBase/), we identified nine genes encoding CDA-like proteins in *Tribolium*. A comparative analysis of CDA families in other insect species including *Drosophila, Anopheles* and *Apis* indicated that the number of *CDA* genes varies from species to species. Of these, *T. castaneum* has the greatest number, including some recently duplicated genes found clustered in its genome. The expression profiles of all *CDA* genes during development were studied by RT-PCR. The *Tribolium* genes *TcCDA1* and *TcCDA2* are orthologs of *DmSerp* and *DmVerm*, respectively, and are expressed throughout all stages of insect development. In contrast *TcCDA6 – 9* are expressed predominantly during the larval feeding stages. *in situ* hybridization revealed that *TcCDA1* and *TcCDA2* are expressed in epidermal cells, whereas *TcCDA6* through *TcCDA9* are expressed in cells lining the midgut lining.

To study the function(s) of *TcCDA* genes, double stranded RNA (dsRNA) specific for each gene was injected into insects at different developmental stages and phenotypes were observed. No visible phenotypic changes were observed after injection of dsRNAs for *TcCDA3* to *9,* while injection of dsRNAs for *TcCDA1* or *TcCDA2* affected all molts, including the larval-larval, larval-pupal and pupal-adult. Insects treated with these dsRNAs did not shed their old cuticles and were trapped in the exuviae. Interestingly, different adult phenotypes were observed after injection of dsRNAs specific for alternatively spliced transcripts of *TcCDA2*, namely*TcCDA2A* and *TcCDA2B*. Neither ds*TcCDA2A* nor ds*TcCDA2B* had any effect on molting. However, the adults that developed after injection of ds*TcCDA2A* into larvae were unable to bend or unbend their legs, at the femoral-tibial joints. In contrast, adults that emerged after larvae were injected with ds*CDA2B* exhibited defects in epidermal cuticle morphology, including roughened elytra. When adult females were injected with ds*TcCDA1* or ds*TcCDA2A*, 70-90% of their progeny were unable to hatch from the egg, and those that did hatch could not molt to second larval instar. These results support the hypotheses that; 1) TcCDA1 and TcCDA2 are critical for egg hatch and also larval-larval, larval-pupal and pupal-adult molting; 2) alternatively spliced variants of *TcCDA2*, *TcCDA2A* and *TcCDA2B*, have different roles in determining adult cuticle morphology; and 3) Proper ratios of chitin/chitosan levels at different stages and/or in different tissues are critical for insect development.

7

Sequencing and annotation of the West Nile encephalitis mosquito genome: *Culex pipiens quinquefasciatus*

Peter Arensburger[1], Peter W. Atkinson[1], Bruce Birren[2], Frank H. Collins[3], Chinnapa Kodira[2], Karyn Megy[4], Marc Muskavitch[5], Sinéad O'Leary[2], Linda Hannick[6], Brian Haas[6]
[1] University of California, Riverside, Riverside, CA; [2] The Broad Institute, Cambridge, MA; [3] University of Notre Dame, Notre Dame, IN; [4] European Bioinformatics Institute, Cambridge, UK; [5] Boston College, Chestnut Hill, MA; [6] J. Craig Venter Institute, Rockville, MD

The genome of the Southern house mosquito, *Culex pipiens quinquefasciatus*, has been sequenced and annotated through a collaboration of the Broad Institute, the J. Craig Venter Institute, VectorBase, and the *Culex* genome consortium. Members of the *Culex pipiens* species complex are the primary vectors of several human pathogens, including those responsible for West Nile encephalitis and lymphatic filariasis. The assembled genome is 540 Mb in length, divided into 3171 scaffold sequences, with 6.14X sequencing coverage. A merge of three predicted gene annotations was made public in January 2008. Both the assembly sequence and gene

information is hosted at VectorBase. The genome appears to be somewhat polymorphic at the nucleotide level. It also has over 30% more predicted gene transcripts than the genomes of the other two sequenced mosquitoes: *Anopheles gambiae* and *Aedes aegypti*. Possible explanations for these unexpected observations are discussed.

8

Chromosome Evolution in Response to Sex Determination of the Hessian fly

*Benatti, T; Aggarwal, R; Fellers, J. P.; Chen, M-S; Zhao, C; Schemerhorn, B.J.; Stuart, J.
(TB, RA, CZ, JS) Purdue University, Department of Entomology, West Lafayette, IN 47907; (JPF) Kansas State University, USDA-ARS and Department of Plant Pathology, Manhattan, KS 66506; (MSC) Kansas State University, USDA-ARS and Department of Entomology, Manhattan, KS 66506; (BJS) Purdue University, USDA-ARS and Department of Entomology, West Lafayette, IN 47907

Hessian fly (*Mayetiola destructor*) females typically produce offspring of only one sex. In order to map the genetic factor controlling this trait, individual gynogenic and androgenic females were fingerprinted using AFLP-PCR. A BAC clone (Mde37d21) containing a marker (2709) linked to the sex factor was then physically positioned near the end of the long arm of the polytene chromosome A1. This analysis revealed two chromosome A1 inversions associated with sex determination. BAC clones Mde37d21 and Hf5d5 were diagnostic for these inversions and used to investigate their distribution, inheritance, and segregation in Hessian fly populations.

Three variants of the A1 chromosome were observed: $A1^{AX}$ – an androgenic chromosome lacking inversions, $A1^{GY1}$ – a gynogenic chromosome containing only the most distal inversion, and $A1^{GY2}$ – another gynogenic chromosome containing both the distal and proximal inversions. $A1^{AX}$ was present in every individual examined and was homozygous in every male. Females were either homozygous $A1^{AX}$ or heterozygous for $A1^{AX}$ and one of other two A1 variants. Heterozygous $A1^{AX}/A1^{GY1}$ and $A1^{AX}/A1^{GY2}$ females were both observed in four populations, but never within the same all-female family. To investigate the effect of these inversions on genetic recombination, we mapped simple sequence repeat (SSR) markers on $A1^{AX}$, $A1^{GY1}$, and $A1^{GY2}$. Taken together, results suggests that the A1 inversions have evolved in response to a sex determining mutation on the A1 long arm in a manner analogous to the evolution of Y (or W) chromosomes in vertebrates.

9

Wnt genes in the *Tribolium* genome and their role in development

*Bolognesi, Renata; Fisher, Tamara & Brown, Susan J.
Kansas State University, Division of Biology, Manhattan, KS 66506

wingless (wg)/Wnt family genes encode secreted glycoproteins that function as signaling molecules in the development of vertebrates as well as invertebrates. In a survey of *Wnt* family genes in the newly sequenced *Tribolium* genome, we found a total of nine *Wnt* genes. As in *Drosophila, Wnt1, 6 and 10* are clustered in the genome. Comparative genomics indicates that *Wnt9* is also a conserved member of this cluster in several insects for which genome sequence is available. Another *Tribolium Wnt* gene (*Tc-WntD/8*) is related to the diverged *Drosophila WntD* gene, both of which phylogenetically cluster with *Wnt8* genes. We further attempt to knockdown the expression of each *Tribolium* Wnt gene by RNA interference (RNAi). Results showed that depletion of *Tc-WntD/8* produces dramatic segmentation defects; irregular germband elongation results in embryos lacking abdominal segments. In contrast, *Tc-Wnt1/wg* depletion does not reduce the number of segments; *Tc-Wnt1* RNAi embryos elongate normally and defects appear during germband retraction. RNAi for the other *Tribolium Wnt* genes did not produce obvious embryonic phenotype. In an attempt to deplete all Wnt signaling in the posterior growth zone, we performed RNAi for *wntless (wls),* assuming it is required for the secretion of all Wnt ligands. Our results show that depletion of *Tc-wls* causes severe segmentation defects that resemble those produced by depletion of *Tc-Wnt1*, but not *Tc-WntD/8*. Both *Tc-Wnt1* and *Tc-wls* RNAi produce highly compact, limbless embryos that still contain the correct number of segments. More severe phenotypes are produced by double RNAi combining *Tc-WntD/8* with *Tc-wls* or *Tc-Wnt1*. Thus, the *Tc-wls* RNAi phenotype is likely due to effects on secretion of *Tc-Wnt1*, while the effects of *Tc-WntD/8* depletion are *Tc-wls*-independent. Our studies describe the *Tribolium* Wnt gene family and provide the first evidence of a Wnt ligand involved in posterior patterning in short germ insects, and that Wls is not required for the secretion of all the Wnt proteins, as previously thought.

10
The Black Fly Genome Project.
C.L. Brockhouse[1], R.J. Post[2], L. Crainey[2], D. Benissan-Messan[1], B. Brei[1] and J.K. Colbourne[3]
[1]Creighton University Department of Biology, Omaha, [2]Natural History Museum, London, UK, and [3]Center for Genomics and Bioinformatics, Indiana University (Bloomington)

We propose a full genome sequencing project and an accompanying cDNA sequencing project for the Simuliidae (Black Flies). The full genome sequence will be an invaluable resource for the insect genomics community, that will allow order-wide functional genomic comparative analysis of genomic contents and their organization, as well as functional analyses of critical parameters such as insect attributes linked to their capacity to transmit disease agents. These attributes include blood feeding (haematophagy), parasite/pathogen transmission, symbiosis, and insecticide resistance. The EST project is critical to assembling the full genome in the face of the absence of genetic maps and the presence of inversion polymorphisms, and will enormously enhance efforts to genetically map the black fly genome and explore gene regulation, differences among species, and symbiosis. The cDNA project is a critical component of this proposal, to facilitate the annotation of the genome sequence and to produce reliable microarrays that will be used to explore the conservation of transcriptional regulators under conditions that are shared by species that vector the agents of disease. This project aims to make a significant impact in furthering genomic knowledge of vector biology, by promoting comparative research on a disease vector that has close phylogenetic relationships to both mosquitoes (*Anopheles*, *Aedes*) and non-haematophagous insects. Biological material will be supplied by participating laboratories (Noblet, Brockhouse, Adler, Post, McCreadie, Wilson, Boakye), while the CGB will carry out the genomic projects in support of the genome sequencing and effort, including cDNA library construction/screening, sequence assembly validations, EST characterization, and related bioinformatics. The principal investigators will solicit the involvement of the growing insect genomics research community for the overall analysis and annotation. The resulting database will be incorporated within the proposed InsectBase.

11
Annotation and Expression Profiling of Apoptosis Related Genes in the Yellow Fever Mosquito, *Aedes aegypti*
Bryant, Bart*; Blair, Carol D.; Olson, Ken E.; and Clem, Rollie, J.
(BB and RJC) Molecular, Cellular, and Developmental Biology Program, Arthropod Genomics Center, Division of Biology, Kansas State University, Manhattan, KS; (CDB and KEO) Arthropod-Borne and Infectious Diseases Laboratory, Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO

Apoptosis has been extensively studied in *Drosophila* by both biochemical and genetic approaches, but there is a lack of knowledge about the mechanisms of apoptosis regulation in other insects. In mosquitoes, apoptosis occurs during plasmodium and arbovirus infection in the midgut, suggesting that apoptosis plays a role in mosquito innate immunity. We searched the *Aedes aegypti* genome for apoptosis related genes using *Drosophila* and *Anopheles gambiae* proteins as queries. In this study we have characterized eleven caspases, three inhibitor of apoptosis (IAP) proteins, a previously unreported IAP antagonist, and orthologs of Ark, Dnr1, and BG4 (also called dFadd). (Waterhouse R et. al., 2007) initially characterized most of these genes, however the annotation of these genes were improved using gene prediction programs, existing ESTs, and full length cDNA sequences. We examined the developmental expression profile for these genes in *Ae. aegypti* larvae, pupae and adults. We also studied the function of the novel IAP antagonist, IMP. It's expression in mosquito cells caused apoptosis, indicating that IMP is a functional pro-death protein. Further characterization of these genes will help elucidate the molecular mechanisms of apoptosis regulation in *Ae. aegypti*.

12
A Complete System For Community Genome Annotation
*Cain, Scott; Hu, James; Just, Eric; and the GMOD Consortium
(SC) Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724 USA; (JH) Department of Biochemistry and Biophysics, 2128 TAMU, Texas A&M University, College Station, TX 77843-2128 USA;(EJ) dictyBase, Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

The Generic Model Organism Database (GMOD) Project is an open source project to develop a complete set of software for creating and administering a model organism database. Here we present a complete system for enabling community annotation of a genome. It is implemented in a VMware appliance, a virtual machine running its own operating system, in this case, the Ubuntu Linux distribution. The use of VMWare allows users to get the

complete system and run it as a server without the need to install and configure all of the components themselves. The components consist of Chado, an organism-agnostic database schema; Apollo, a Java GUI application for editing gene models; and MediaWiki, the well known web application that is used for Wikipedia. Here it is used in conjunction with the TableEdit extension to allow users to annotate genes in an easy to use application, with the results being saved back to the Chado database. All that the users need to provide is the initial genome data in the form of sequence and computational results like gene predictions and similarity results. This work was supported by grants from the National Institutes of Health and the USDA Agricultural Research Service.

13

Identifying Specialized Salivary Gland Transcripts in Pea Aphid Using Bioinformatics Tools

*Caragea, D; Kallumadi, S; Dittmer, N; Chellapilla, S; Mutti, N; Feng, C; Pierson, M; Herman, M; Culbertson, C; Reese, J; Edwards, O; *Reeck, G.
(DC, SK) Kansas State University, Department of Computer Science and Informatics, Manhattan, KS 66506; (SC) Kansas State University, Bioinformatics Center; (ND, CF, MP, MH, GR) Kansas State University, Department of Biochemistry; (NM) Arizona State University, School of Life Sciences, Arizona State University; (CC) Kansas State University, Department of Chemistry; (JR) Kansas State University, Department of Entomology; (OW) CSIRO, Perth (Floreat), Australia

We are in the process of analyzing over 9,000 pea aphid salivary-gland ESTs, along with nearly 100,000 NCBI-deposited ESTs from other pea aphid libraries. The central goal of the analysis is to identify transcripts that are strongly enriched in the salivary-gland libraries compared to other pea aphid libraries (antennae, digestive tract, head, parthenogenetic embryo, whole body) and from those to predict putative secreted proteins. We are using the "R-statistic" of Stekel et al. (2000) to identify transcripts strongly enriched in the salivary-gland libraries.

The starting point for our analysis is contig formation using the KSU Bioinformatics Center's recently developed AthropodEST pipeline. The principal advantage of this local pipeline is facile control of parameters for contig formation and the ability to handle very large data sets efficiently. For the 9453 salivary-gland ESTs analyzed, the AthropodEST pipeline formed 826 contigs and found 2662 singletons (for a total of 3488 unigenes). The contig formation and singleton identification step was followed by unigene annotation using the Blast2GO (Conesa et al., 2005) annotation tool. Blast2GO found the greatest number of hits in *Drosophila*, *Aedes*, mouse, human and *Tribolium*, with most of the hits coming from the UniProt database. An intriguing aspect of the annotation was that 2368 of the 3488 unigenes had no blast hits. Preliminary analysis shows that at least some of these unigenes clearly encode proteins. Furthermore, many of the salivary-gland-EST abundant no-hit-transcripts appear to have high R-statistic values (estimated by comparing salivary-gland transcripts to their corresponding whole-body transcripts obtained from a non-normalized pea aphid whole body library). This suggests that the salivary-gland-EST abundant no-hit-transcripts may encode specialized salivary proteins.

To improve the pea aphid contigs obtained from salivary-gland libraries, we are forming contigs from the combined set of all EST libraries available for the pea aphid (containing approximately 109,000 ESTs), including several non-normalized libraries and one normalized whole-body library. R values will be calculated for the resulting contigs (based on the non-normalized libraries) and those that appear to be specialized salivary-gland transcripts will be thus identified. A thorough investigation of the specialized salivary-gland transcripts for which there are no blast-hits available is planned.

The programs resulted from our study (for analyzing the pipeline output and calculating the R-statistic) will be incorporated into the package of programs available through the Bioinformatics Center.

Acknowledgments: AGC research grant (CC, GR), CRC for Plant Biosecurity (OW, GR) KSU Bioinformatics Center (DC, SC), K-INBRE (DC, SC).

14

Characterization of an endoplasmic reticulum protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*

*Feng Cui[1,5], Huaien Dai[1], Yasuaki Hiromasa[1], Doina Caragea[2], Changzhong Sheng[1], John Reese[3], Owain Edwards[4,5], Gerald Reeck[1]
[1]Department of Biochemistry, Kansas State University, USA, [2]Department of Computing and Information Sciences, Kansas State University, [3]Department of Entomology, Kansas State University, [4]CSIRO Entomology, Australia, [5]CRC for Plant Biosecurity

A protein called ARP was cloned from a salivary gland cDNA library of the pea aphid (*Acyrthosiphon pisum*), expressed in *E.coli* and purified using Ni-NTA and ion exchange chromatography. The ARP ORF is 525 bp and encodes 174 amino acid residues, including a signal peptide in the N-terminus. The molecular mass of

recombinant ARP is 19.2 kDa, confirmed by mass spectrometry. Its calcium binding potential is disclosed qualitatively by reaction with ruthenium red. ITC assay (isothermal calorimetric titration) demonstrates that there are two types of binding sites within this protein, one tightly binding 1 calcium, and the other loosely binding 4 calciums. ARP exists as a monomer in the presence of calcium. This protein has a high content of alpha helix and calcium doesn't influence this content. The gene of ARP includes four encoding exons and three introns inferred from the alignment between the ORF and the genomic sequence. Eight conserved cysteines exist in all ARP homologues from insects, mammals and nematodes, apparently forming four disulfide bonds. In the C-terminus, there is an ER-retention signal, KEEL. It seems all ARPs have variants of this 4-residue signal. ARP may therefore be in the class of proteins called reticuloplasmins, proteins that reside in the endoplasmic reticulum (ER) lumen and have varied functions, such as molecular chaperones during protein assembly and degradation and as calcium buffers [1].

15

Role of the *Knickkopf* (*Knk*) Gene Family in Cuticle Turnover in *Tribolium castaneum*

*Chaudhari, Sujata; Arakane, Yasuyuki; 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; (KJK, RWB) Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA

Proteins associated with insect chitin metabolism and cuticle organization are potential targets for novel biopesticides. Recent studies using *Drosophila* embryos have identified Knickkopf (Knk) as a protein whose expression is necessary for tracheal tube expansion. Although the precise function of Knk is still uncertain, it is thought to be involved in chitin metabolism and/or assembly of chitin microfibrils in the insect epidermal cuticle and tracheal lining. In the present study, we have identified three putative orthologs of *Drosophila Knk's* in *Tribolium castaneum* and designated them as *TcKnk1*, *TcKnk2* and *TcKnk3*. We have also investigated their function using dsRNA-mediated RNA interference (RNAi). Protein domain analysis of *Tribolium* Knk's indicated the presence of putative DM13 domain, DOMON domain and GPI-signal of which at least one (DOMON) is predicted to bind sugars. Stage-specific expression analysis of these three genes suggested that they may be differentially regulated. While *TcKnk1* and *TcKnk3* are expressed at all stages of *Tribolium* development, expression of *TcKnk2* is absent at the embryonic stage and is significantly reduced at the late larval stage. RNAi of *TcKnk1* resulted in lethality due to molting defects evident at larval-larval, larval-pupal and pupal-adult molts, a phenotype very similar to that found after depletion of chitin synthase A by RNAi in *Tribolium*. RNAi of *TcKnk2* showed no obvious defects in molting at the larval-larval and larval-pupal molts. However, mortality was observed in >50% of the *Knk2* dsRNA-injected insects at pharate adult stage. Collectively our results indicate that the different *Tribolium* Knk's may have multiple and different roles in the synthesis, degradation or organization of chitin.

16

ArthropodEst: A Pipeline for automated EST data analysis

Chellapilla, Sanjay; Kallumadi, Surya; Park,Yoonseong; Caragea, Doina; Brown Susan J
(SC, SJB) Athropod Genomics Center and KSU Bioinformatics Center, Division of Biology, Ackert Hall KSU, Manhattan KS 66506; (SK, DC)Department of Computer and Information Science, Nichols Hall, KSU, Manhattan, KS 66506; (YP) Entomology Department, Waters Hall, KSU, Manhattan, KS 66506

Expressed Sequence Tags **(**ESTs), produced by single-pass end-sequencing of cDNA clones, generate large datasets that are instrumental in gene discovery and gene sequence determination. Several EST data analysis pipelines are available as web-servers, (ESTpass, EGassembler, and ESTexplorer). However, they all have limitations, e.g., with respect to the amount of data that can be uploaded at one time or the type and format of the

annotations and statistics they provide. The KSU Bioinformatics Center has developed an in-house EST analysis computational pipeline, ArthropodEST, using existing open source software tools. The pipeline cleans and assembles cDNA/EST sequences into contigs and singletons, which are then assigned functional annotations via BLAST2GO

The WWW-accessible pipeline performs the following steps in order:
[1] Optional screening for vectors and contaminants using standard vector databases (NCBI UniVec and EMBL EmVec) and user-provided vector and contaminant sequences.
[2] Screen for interspersed repeats and low complexity.
[3] Clustering and assembly into contigs and singletons according to user-specified parameters.

A summary report including results from each step is automatically created and provided to the user in addition to the final contigs and singletons after assembly. To view the contigs, the WWW URLs to two widely-used graphical viewers, Consed and Clview, are provided in the final report. We will describe ArthropodEST steps and paramenters in detail, comparing its speed and accuracy with other pipelines that are available through web-servers.

17
GMOD: Database Resources for Emerging Model Organisms
Clements, Dave*; Lapp, Hilmar; Vision, Todd
National Evolutionary Synthesis Center (NESCent), 2024 W. Main Street, Suite A200, Durham, NC 27705

With the advent of more affordable sequencing technologies many smaller research communities now have an unprecedented volume of data available to them. These emerging model organism communities now face the major challenge of managing, sharing and exploiting this newfound wealth of data with only minimal informatics budgets. The Generic Model Organism Database project (GMOD) is a collection of interoperable open source software tools for managing, annotating and visualizing genomic data. GMOD tools are used in diverse contexts, ranging from genome annotation projects within individual labs to major model organism databases, and are widely used inside the arthropod research community (AphidBase, BeeBase, BeetleBase, FlyBase, HeliconiusBase, VectorBase, wFleaBase). Recently, the National Evolutionary Synthesis Center (NESCent) has joined the GMOD collaborative with the goal of lowering the barrier to entry for research communities focused on emerging model organisms. A new GMOD help desk is responsible for maintaining up-to-date documentation, providing users with ondemand assistance and offering training workshops and online tutorials. Over the past year, the GMOD wiki (http://gmod.org) has been completely revamped, and a variety of tutorials are being developed for popular GMOD components. The GMOD help desk also aims to grow the developer community and disseminate best practices in the use and extension of GMOD tools, particularly as applied to comparative and evolutionary data. In addition to sponsoring the help desk, NESCent, as part of its work on HeliconiusBase, is also extending the GMOD data model (Chado) to better support evolutionary data types.

18
Genomic tools for the European corn borer, *Ostrinia nubilalis*
*Coates, B.S.; Sumerford, D.V.; Lewis, L.C.
United States Department of Agriculture, Agricultural Research Service, Corn Insects & Crop Genetics Research Unit, Ames, IA, 50011

Expressed sequence tags (ESTs) from larval midgut and fatbody tissue, and end sequencing data from bacterial artificial chromosome (BAC) clones has been generated for *Ostrinia nubilalis*. 1745 midgut ESTs have an average read length of 457.2 bp, and annotation predicted the presence of 535 unique gene sequences (GenBank dbEST accessions EL928389 – EL930130). Multiple families of proteinase and lipase genes may result from gene duplication or convergent evolution. Preliminary annotation of > 500 fatbody ESTs suggested > 120 unique genes may be present. A BAC library, named OnB1, contains 36,864 clones each carrying ≥ 120 kb of genomic DNA for an estimated 8.8-fold coverage. BAC end sequence (BES) reads (GenBank dbGSS accessions ET217010 to ET217273) showed homology to *O. nubilalis* and *Bombyx mori* EST sequences. Screening an approximate 2.76-fold genome equivalent of OnB1 clones by PCR has determined physical position of nearly 100 sequence tag site (STS)-, BES-, and EST-based markers. Included are Z chromosome-linked markers *tpi*, *ket*, *ldh*, and *OnZ*1 that will be used to construct contigs for a physical map of that chromosome.

19

Developing Novel Analytical Methods to Sample and Identify Aphid Salivary Proteins

*Culbertson, Christopher T.; Meyer, Amanda R.; Lamabadasuriya, Manuja; Reeck; Gerald R.; Reese, John C.; Campbell, Peter R.; and Edwards, Owain
(CTC, ARM, ML) Kansas State University, Department of Chemistry, Manhattan, KS 66506; (GRR) Kansas State University, Department of Biochemistry, Manhattan, KS 66506; (JCR) Kansas State University, Department of Entomology, Manhattan, KS 66506 (PRC, OE) CSIRO Entomology GPO Box 1700 Canberra ACT 2601

Aphids are important plant and crop pests. They are known vectors for the transfer of many plant viruses, and they lower plant yields. From a physiological standpoint, the proteins and enzymes secreted by aphids while feeding on plants are crucial to the aphid's ability to successfully obtain nourishment from a plant, and may act as elicitors of plant responses. But at a biochemical level, very little direct information in known about the aphid secretome. The lack of direct information is mostly due to the extremely small volume of saliva secreted by aphids and the difficultly in collecting these secretions without significant dilution. In order to try to solve this problem we have been pursuing a 3 pronged approach that includes genomics, transcriptomics and proteomics. The combination of these 3 lines of research is critical to the success of this project especially in the area of identifying novel proteins in the saliva. The salivary transcripts provide a base from which to search to mass spectral data sets obtained from the aphid saliva. This presentation, however, will focus on the proteomics part of the problem and the isolation of the aphid salivary secretions. In order to try to overcome the volume and dilution limitations that have historically hindered aphid saliva analysis we have been exploring to potential of using miniaturized chemical analysis instrumentation (i.e. microfluidic devices) and capillary HPLC to collect and purify sufficient quantities of "concentrated" aphid saliva that can then submitted to mass spectrometric analysis. Microfluidic devices consist of a series of small channels etched in glass. We can make one surface of these devices out of a material (Parafilm®) that aphid stylets can penetrate and then after several hours of feeding collect the feeding media left in the channels (~0.3 to 3 uL). This media should contain a relatively concentrated sample of aphid salivary secretions. We will present results and discuss the capabilities and limitations of these devices to collect aphid salivary secretions and the various sample cleanup and HPLC steps that have been developed in order to make the samples amenable to mass spectrometric analysis.

20

Comparative genomics and database construction for Lepidoptera

d'Alençon, E.; Sezutsu, H.; Legeai, F.; Quesneville, H.; Gordon, K.; Samain, S.; Gagneur, C.; Gimenez, S.; Cousserans, F.; Flutre, T.; Shimomura, M.; Brun, A.; Jacquin-Joly, E.; Weissenbach, J.; Mita, K.; *Fournier, P.; Feyereisen, R.
(EA, CG, SG, FC, PF) INRA & Montpellier University, France; (HS, MS, KM) NIAS, Tsukuba, Japan; (HS, AB, RF) INRA-Sophia & Nice University, France; (FL) INRA, AgroCampus Rennes, France; (HQ, TF) INRA-Versailles & Paris VI University, France ; (EJJ) INRA-Versailles, France ; (KG) CSIRO, Canberra, Australia ; (SS, JW) Genoscope, Every, France.

We are developing a study on Lepidoptera microsynteny, *i.e.* an analysis on gene similarity, conservation of gene order, orientation and distance along limited chromosomal regions. As the the annotated and integrated *Bombyx mori* genome becomes available, it is now possible to investigate the phylogenetic links between *B. mori* and two Noctuidae species, *Spodoptera frugiperda* and *Helicoverpa armigera*. More than 60 gene probes were designed for both noctuids on genes related to various physiological pathways. Probes were hybridized against BAC genomic filters. Positive clones were fingerprinted and one BAC clone of each selection was sequenced. Gene annotation was provided by the Kaikogaas tool (developed for the *Bombyx* genome), and synteny was visualized using Apollo software.

More than 35 BACs of each species have already been selected and sequenced, representing 3.8 Mb (*Ha*) and 4.5 Mb (*Sf*). From these, 16 triads have already been compared between these species and *B. mori* corresponding regions. The probe genes were found surrounded by same known genes, but gene duplications or inversions were frequent. Translocation events occurred in few cases, and synteny blocks were generally interrupted by several hypothetical protein coding genes that differed from one species to another.

Our groups have developed a concerted project of ESTs sequencing, using various tissues or cell lines for *Spodoptera frugiperda*. These data will be helpful for gene annotation, but also for functional genomics studies, including microarray based analysis of the transcriptome.

We are now developing and maintaining a complete bioinformatics environment dedicated to the analysis of these Lepidoptera genomics data. Based on the GMOD infrastructure, we set up a Chado database and filled it with available *Helicoverpa armigera*, *Spodoptera frugiperda* and *Bombyx mori* corresponding genomic sequences. We added functional and structural annotations delivered by Kaikogaas, identification of putative repeated segments and transposons provided by the Repet package, and similarities with some insect peptides (honeybee or *Drosophila*). It is possible to localize colinear region between couples of sequences. Everything can be visualized though the common Gbrowse interface and annotations can be corrected by any scientist of our consortium by use of Apollo. By using and participating to French and international development around the GMOD environment, we are planning to improve our platform by giving access to data mining or visualization tools through a well-documented web portal.

21

Using a Genotyping Array to Reveal QTL for Mitochondrial OXPHOS Pathway Enzyme Efficiencies in Hybrids of *Nasonia* Parasitoid Wasps

*Gibson, Joshua D.; Niehuis, Oliver; Gadau, Jürgen
Arizona State University, School of Life Sciences, Tempe, AZ 85287

The genus *Nasonia* (Hymenoptera: Pteromalidae) is a complex of three closely related species of parasitoid wasps. These species are reproductively isolated due to their infection with different strains of *Wolbachia* endosymbionts. Once the three species are cured of their *Wolbachia*, they can be crossed to produce interspecific F_1 hybrids. While the F_1 hybrids are fully viable and fertile, haploid F_2 hybrids in crosses between *N. vitripennis* and *N. giraulti* suffer from increased mortality (*i.e.*, hybrid breakdown). Determining what causes this increased mortality is the focus of the present research project.

Studying a mapping population of haploid F_2 hybrids, one expects to find the two parental alleles of any given locus in equal frequency if meiosis is fair and no genotype specific larval or adult mortality occurs. In *N. vitripennis* x *N. giraulti* F_2 hybrids, however, this ratio is significantly distorted for certain loci. A previous study by us revealed that the observed distortion is the result of differential mortality of F_2 hybrids during larval development. A significant proportion of this mortality is caused by genic incompatibilities between nuclear and mitochondrial factors. We also found that the efficiency of mitochondrial enzyme complexes involved in oxidative phosphorylation (OXPHOS), the generation of ATP, is significantly reduced. Since four of five OXPHOS enzyme complexes are composed of both nuclear and mitochondrial subunits, a disruption of the OXPHOS pathway could be the cause for the observed F_2 hybrid mortality.

To study whether quantitative trait loci (QTL) for OXPHOS enzyme complex efficiencies in F_2 hybrids of *N. vitripennis* x *N. giraulti* map to the same position as the hybrid incompatibility loci that were previously found, we annotated the nuclear encoded genes of the OXPHOS pathway in the *Nasonia* genome. We identified 60 putative OXPHOS genes in *Nasonia* out of the 65 genes reported from *Drosophila* and human. Single nucleotide polymorphisms (SNPs) differentiating *N. vitripennis* and *N. giraulti* were identified within these genes and were designed onto a genotyping array in order to allow us to determine which allele is present at a given locus in the F_2 hybrids . We added 1431 additional SNPs to the array that are spread across the entire genome, concentrating on the 300 largest scaffolds of the *Nasonia* genome assembly 1.0. *N. vitripennis* x *N. giraulti* crosses were carried out and the genotyping arrays are currently being run with DNA from the F_2 progeny of these crosses.

We expect to find QTL that explain the increased variation that we see in mitochondrial OXPHOS efficiencies of F_2 hybrids. Should incompatibilities between nuclear and mitochondrial encoded subunits of OXPHOS enzyme complexes be responsible for the increased F_2 hybrid inviability, at least some QTL should map at the same position as the previously studied hybrid incompatibility loci. We intend to introgress these loci into the reciprocal species to obtain individuals with nuclear genomes that are identical to the pure species except for the locus under investigation. These individuals will allow us to observe the effect of each locus on the efficiency of the mitochondrial complexes and will give us further insight into the role of nuclear-mitochondrial interactions in hybrid inviability.

22
Unlocated Arthropod genes, and ways to find them
Gilbert, Don, Indiana University, Biology Department, Bloomington, IN 47405, gilbertd@indiana.edu

The *Daphnia pulex* genome is rich in tandem duplicate genes, some 20% of its 30,000+ genes. However some gene predictors have missed or incorrectly located almost half of these. Estimates from genome-wide tile expression suggest an additional 5,000 genes have been missed.

Gene prediction for new genomes such as this first crustacean is still an uncertain task. Even in species clades with a well-characterized model such as *Drosophila*, gene finding remains an uncertain task.

Prediction tools are increasingly sophisticated and accurate. Today's methods draw on the range of available gene evidence and improved modelling of gene structures. Yet they are sensitive to available gene data and expected structures. They find well-known genes, but fail at accurate detection of novel and diverged genes.

Measures from gene duplication and genome-wide tile expression can more accurately locate those genes missed by other methods. Computational methods are being developed to turn these signals to accurate gene models.

Application of these methods to athropod genomes, including *Daphnia* and *Drosophila*, uncovers 10% to 25% additional species specific and diverged genes. This work includes development of new automated genome analysis pipelines on NSF TeraGrid shared cyberinfrastructure, as part of the Generic Model Organism Database project.

23
The Genetics of Honey Bee Colony Defense
Hunt Lab: Ammons, A.; *Andino, G.; Arechavaleta, M.; Emore, E.; Guzmán, E.; Hunt, G.; Schlipalius. D.
(AA, GA, CE, GH) Department of Entomology, Purdue University, West Lafayette, IN 47907; (MA) Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal INIFAP, Querretaro, Mexico; (EG) Department of Environmental Biology, University of Guelph, Ontario, Canada

The high recombination rate of bees resulted in fine-scale quantitative trait loci (QTL) mapping and ID of candidate genes from genome sequence for colony stinging behavior. The results obtained from quantitative real time-PCR showed that some candidate genes for 3 QTL for stinging and guarding behaviors had trends in differential expression between gentle and defensive bees. The candidate genes *huntingtin*, *14-3-3*, *Arrestin* (*AmArr4*), *GABA BR1* receptor, Homer protein and several transcriptions factors are thought to be involved in different neural transport functions and some of them modulate olfactory and visual signaling. In addition, results from many sets of crosses strongly suggest that paternal effects of African origin increase the defensive behavior of hybrid colonies.

24
Characterization, Gene Organization, Expression and Functional Analysis of Peritrophins in *Tribolium castaneum*
*Jasrapuria, S.; Arakane, Y.; Begum, K.; Kramer, K. J.; Beeman, R. W.; Park, Y.; Muthukrishnan, S.
(SJ, YA, SM, KJK) Kansas State University, Department of Biochemistry, Manhattan, KS, 66506; (KB, YP) Kansas State University, Department of Entomology, Manhattan, KS, 66506; (RWB) Grain Marketing and Production Research Center, ARS-USDA, Manhattan, KS, 66502.

This study is focused on the characterization and expression of peritrophic membrane protein genes (*PMP-1* and *PMP-2*) in the gut of the red flour beetle, *Tribolium castaneum*. These genes were identified by a bioinformatics search for *Tribolium* proteins containing peritrophin-A domains, which have been identified in proteins extracted from peritrophic membranes of other insects. The two genes are arranged in tandem (head-to-tail) in the *Tribolium* genome and are separated by only 331 base pairs. Sequences of cDNAs encoding *PMP-1* and *PMP-2* were deduced from genomic sequences, available ESTs and by 5′RACE and 3′RACE. Subsequently full length cDNAs were cloned using appropriate primers. Both the developmental and tissue-specific expression patterns of these two genes were analyzed during insect development by RT-PCR and q-PCR. *PMP-1* (predicted molecular weight 75 kDa) encodes a protein of 706 amino acids containing 9 tandem repeats of the CBM14 chitin-binding Peritrophin-A domain (pfam01607), also known as chitin-binding domain type 2 (ChtBD2=smart00494). *PMP-2* (predicted molecular weight 147 kDa) encodes a protein of 1306 amino acids composed of 14 tandem repeats of

this same domain. Each repeat has 6 cysteine residues and is 77 amino acids in length. The repeats are connected by linker regions rich in proline and threonine which show some variability in length in *PMP-2*. Both the *PMP-1* and *PMP-2* genes are expressed only in the posterior part of the midgut during the feeding stages, as revealed by both the RT-PCR and *in situ* hybridization. dsRNA-mediated RNAi did not result in any observable phenotypes, even though transcript depletion could be demonstrated by q-PCR. Two other closely-related peritrophins, namely GLEAN_08506 and GLEAN_04216 also have the Peritrophin-A=ChtBD2 domain (7 and 4 tandem copies, respectively) and are very similar to those in PMP1-2. These results will form the basis for future expression of recombinant proteins and analysis of chitin binding activity.

25

Genome Size of *Metaseiulus occidentalis* (Acari: Phytoseiidae) Among Smallest known in Arthropods

*Jeyaprakash, Ayyamperumal; Hoy, Marjorie A.
Department of Entomology and Nematology, University of Florida, Gainesville, FL 32611

The genome size of the predatory mite *Metaseiulus occidentalis* needs to be estimated before whole genome sequencing is done, but little sequence information is available for this species. *Drosophila melanogaster Actin* and *EF1* sequences were used to design degenerate primers to amplify two single copy nuclear genes from *M. occidentalis*. Four different *Actin* genes and one *EF1* sequence were amplified from *M. occidentalis*, cloned and sequenced. One unique *Actin* sequence allowed designing a *M. occidentalis*-specific primer pair. Only one copy of *EF1* is known to exist in the *M. occidentalis* genome and another specific primer pair was designed from this sequence. A quantitative high-fidelity real-time PCR procedure was used to estimate the copy number of these two unique sequences from a known quantity of *M. occidentalis* genomic DNA and compared with nine serially-diluted plasmid templates with known copy numbers. The genome size in base pairs was estimated using the formula [(0.978 X 10^{-9}) X DNA content (pg)]. The genome size of *M. occidentalis* was estimated to be 87.6 Mb. When compared to other arthropods this is slightly larger than a spider mite (75 Mb) genome, but far smaller than those of fruitflies (125-180 Mb), honeybee (200 Mb), mosquitoes (278-800 Mb), housefly (295 Mb), silkworm (530 Mb), ticks (2.3-7.1 Gb), spiders (723 Mb-5.5 Gb) or horseshoe crab (2.5 Gb).

26

Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses

*Jiang, Haobo; Zou, Zhen; Najar, Fares; Wang, Yang; Roe, Bruce
(HJ, ZZ, YW) Oklahoma State University; Department of Entomology and Plant Pathology, Stillwater, OK 74074; (FN, BR) University of Oklahoma, Department of Chemistry and Biochemistry, Norman, OK73019

The tobacco hornworm *Manduca sexta* is widely used as a model organism to investigate the biochemical basis of insect physiological processes but little transcriptome information is available. To acquire a broad view of the larval hemolymph proteins, particularly those related to immunity, we synthesized and sequenced cDNA fragments from a mixture of eight total RNA samples: induced fat body and hemocytes, naïve fat body, hemocytes, integument and trachea, as well as fat body and hemocytes from wandering larvae. Using massively parallel pyrosequencing, we obtained 95,458 expressed sequence tags (ESTs) at an average size of 185 bp per read. A majority of the sequences (69,429 reads) were assembled into 7,231 contigs, 1178 of which had significant similarity with *Drosophila* genes from various functional groups. Only ~8% (606) of the contigs matched known *M. sexta* cDNA sequences, representing 186 of the 375 unique NCBI entries. The remaining 6,625 contigs, including 142 with putative functions, represented DNA segments newly discovered from this well studied biochemical model insect. A search of the dataset using *Tribolium castaneum* immune protein sequences revealed 197 cDNA contigs with significant similarity (E-value $< 1\times10^{-5}$). These included 81 previously unknown *M. sexta* sequences coding for putative defense molecules such as pattern recognition receptors, serine proteinases, serpins, Toll-like receptors, intracellular signaling molecules, and antimicrobial proteins.

27

Comparative genomic analysis of the *Tribolium* immune system

*Jiang, Haobo; Zou, Zhen; Evans, Jay D.; Lu, Zhiqiang; Zhao, Picheng; Williams, Michael; Sumathipala, Niranji; Hetru, Charles; Hultmark, Dan
(HJ, ZZ, ZL, PZ, NS) Oklahoma State University, Department of Entomology and Plant Pathology, Stillwater, OK 74078; (JDE) USDA-ARS Bee Research Laboratory, Beltsville, MD 20705; (MW, DH) Umeå University, Umeå Centre for Molecular Pathogenesis, Umeå S-901 87, Sweden; (CH) Institut Biol Moléc Cell, CNRS, Strasbourg 67084, France

Tribolium castaneum is a species of Coleoptera, the largest and most diverse order of all eukaryotes. Components of the innate immune system are hardly known in this insect, which is in a key phylogenetic position to inform us about the genetic innovations accompanying the evolution of holometabolous insects. We have annotated immunity-related genes and compared them with homologous molecules from other species. Around 300 candidate defense proteins are identified based on sequence similarity to homologs known to participate in immune responses. In most cases, paralog counts are lower than those of *Drosophila melanogaster* or *Anopheles gambiae* but are substantially higher than *Apis mellifera*. The genome contains probable orthologs for nearly all members of the Toll, IMD, and JAK/STAT pathways. While total numbers of the clip-domain serine proteinases are approximately equal in the fly (29), mosquito (32) and beetle (30), lineage-specific expansion of the family is discovered in all three species. Sixteen of the thirty-one serpin genes form a large cluster in a 50 kb region resulted from extensive gene duplications. Among the nine Toll-like proteins, four are orthologous to *Drosophila* Toll. The presence of scavenger receptors and other related proteins indicates a role of cellular responses in the entire system. The structures of some antimicrobial peptides drastically differ from those in other orders of insects. A framework of information on *Tribolium* immunity is established, which serves as a stepping stone for future genetic analyses of defense responses in a nondrosophiline genetic model insect.

28

Comparison of the Cytochrome P450 Genes in the Genomes of *Apis mellifera* and *Nasonia vitripennis*: Effects of Haplodiploidy and Eusociality on the P450 Repertoire

Johnson, Reed M.*; Berenbaum, May R.
University of Illinois—Urbana-Champaign, Department of Entomology, Urbana, IL 61802

The recent sequencing of the honey bee (*Apis mellifera*) genome has revealed that bees have among the smallest repertoire of cytochrome P450 monooxygenase genes of any insect--only 46 sequences, compared to 89 in *Drosophila melanogaster* and 111 in *Anopheles gambiae.* Other insects are known to withstand exposure to natural toxins and gain pesticide resistance through elevated metabolic detoxification made possible by the P450 family of enzymes. There are two explanations for the low P450 number in the honey bee. First, haplodiploidy could be the cause, because selection for P450s associated with resistance occurs directly on hemizygous males, leading to more rapid fixation of these genes. Second, eusociality in bees, and the high level of nest homeostasis eusociality provides, could insulate the queen from exposure to toxins, making P450-mediated detoxificationless critical. The sequencing of the genome of the haplodiploid, but solitary, jewel wasp (*Nasonia vitripennis*), and its 90 encoded P450s, has made clear that the social behavior and the well regulated nest environment of the eusocial honey bee may be the principal factor in the low number of P450 genes encoded by the honey bee genome.

29

Functional Annotation of the Expressed Sequence Tags from the Gut of the European corn borer (*Ostrinia nubilalis* Hübner)

*Khajuria, C.; Zhu, Y.; Chen, M.; Buschman, Lawrent L.; Subbaratnam, M.; Zhu, K.
(CK, MC, LLB, KZ) Department of Entomology, Kansas State University, Manhattan, KS 66506; (MS) Department of Biochemistry, Kansas State University, Manhattan, KS 66506; (YZ) USDA-ARS-JWDSRC, 141 Exp Stn Rd, Stoneville, MS 38776

Two cDNA libraries were constructed from the gut of European corn borer (ECB, *Ostrinia nubilalis*) larvae. A total of 15,000 random clones from both the libraries were sequenced. These sequences were assembled into 4,451 unique sequences, including 2,787 singletons and 1,664 contigs. Among these sequences, 78% encoded for putative proteins which shared significant sequence similarity with known sequences in GenBank with E-values ≤ 10^{-3}. These sequences includes many genes encoding for various digestive enzymes such as chymotrypsin-like proteases, trypsin like-proteases, elastases, cysteine proteases, carboxypeptidases, aminopeptidases, and cathepsin L-like proteases. These putative proteins were further analyzed and sorted into different categories of

molecular functions, biological processes and cellular components. This work provides foundation for the future research to use microarray technology for analyzing the global changes of gene expression in response to Bt protoxins/toxins as well as the genetic difference between Bt resistant and susceptible strains of ECB.

30

Cotesia plutellae bracovirus genome and its function in altering insect physiology

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

Polydnavirus is a group of animal DNA virus mutually associated with some ichneumonoid wasp. Its relatively large size of genome has been considered as a major source of the parasitoid function to manipulate developmental and immunological processes of target parasitized insects. *Cotesia plutellae* bracovirus (CpBV) is a polydnavirus derived from *C. plutellae*, which parasitizes the diamondback moth, *Plutella xylostella*. Parasitized *P. xylostella* exhibits altered physiological symptoms in development and immune reactions. Though several other parasitic factors such as ovarian proteins, venom, and teratocytes are identified, CpBV has been more focused on elucidating various host physiological alterations occurring due to the parasitism, which has driven the CpBV genome project. CpBV attains a typical bracovirus structure by its single unit membrane envelope, in which multiple nucleocapsids are enclosed. Its genome DNAs are segmented and located on the genome of *C. plutellae*. Its replication begins at adult tissue development during pupal stage. An apparent genome size is 471 kb estimated from 27 segments separated on 5% agarose gel. A current work on the genome has been completely sequenced 24 genomic segments and analyzed their genomic structure. The aggregated genome size is 351,299 bp long and exhibits an average GC content of approximately 34.6%. Average coding density is about 32.3% and 125 putative open reading frames are predicted. Though more than half (52.5%) of predicted genes are annotated as hypothetical, the annotated CpBV genes share amino acid sequence homologies with those of other bracoviral genomes. The annotated genes are classified into the known bracoviral families, in which a family of protein tyrosine phosphatase is the largest including 36 ORFs, suggesting a significant role during parasitization. In addition, 8 and 7 ORFs encode $I\kappa\beta$-like and EP1-like, respectively. Some predicted genes are known only in *Cotesia*-associated bracoviral genomes. Finally, two homologous genes, CpBV15α/β, are unique in CpBV genome, which are not matched to any other known polydnaviral genes. Their homology with malarian circumsporozoite toxin and eukaryotic translation inhibition factors suggests their function in host translation inhibitory factor.

31

Evolution of Acyl-CoA Desaturase Genes in Moths and Flies

*Knipple, Douglas C.; Jeong, Seong Eun
(DCK) Department of Entomology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456 USA; (SEJ) Department of Life Sciences, Hannam University, 133 Ojung-Dong, Daedeok-Gu,Daejeon 300-791, Korea

Lepidopteran sex pheromones are volatile chemical blends that are synthesized from fatty acid precursors in terminally differentiated cells or glands of adult females and released at appropriate times to attract conspecific males for mating. Three decades of investigations of the chemical constituents and biosynthetic pathways of lepidopteran sex pheromones has demonstrated that most of the hundreds of unique constituent molecules of lepidopteran sex pheromones are derived from unsaturated fatty acid precursors with variable hydrocarbon chain length, and variable number, location and geometry of double bonds. These findings suggest that pheromone biosynthetic pathways of the Lepidoptera evolved through the recruitment of genes involved in normal fatty acid metabolism and that the diversity of chemical structures used as sex pheromones in this large taxonomic group has arisen, in part, through the evolution of novel biochemical properties of key biosynthetic enzymes. It is apparent that acyl-CoA desaturases play a particularly significant role in the generation of structural diversity of sex pheromone components as a consequence of their acquisition of novel substrate specificities, regiospecificities and the generation of variable amounts of *Z* and *E* geometric isomers. In this poster we examine recent progress made through molecular genetic and bioinformatic approaches towards understanding the structure, function, and evolution of insect acyl-CoA desaturases, and their role in sex pheromone biosynthesis in moths.

32

The Use of Microarrays to Find a Marker for Rotation Resistance in the Western Corn Rootworm

*Knolhoff, Lisa M.; Walden, Kimberly K. O.; Ratcliffe, Susan T.; Onstad, David W.; and Robertson, Hugh M.
(LMK, KKOW, HMR) University of Illinois at Urbana-Champaign, Department of Entomology, Urbana, IL 61801; (STR) University of Illinois at Urbana-Champaign, Department of Crop Sciences, Urbana, IL 61801; (DWO) University of Illinois at Urbana-Champaign, Department of Natural Resources and Environmental Science, Urbana, IL 61801

The western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae), is a major pest of corn in the United States. A principal method of management is crop rotation, as this insect typically completes its entire life cycle in corn. However, the western corn rootworm has adapted to crop rotation in much of Illinois and Indiana. Rotation-resistant females lay eggs in fields that are in annual rotation with corn, namely soybean. The mechanism for this behavioral adaptation is likely to be related to heightened levels of flight and locomotory activity, increasing the odds that rotation-resistant females enter non-corn crops. A molecular marker is needed to clearly distinguish between rotation-resistant and susceptible individuals. Previously, a cDNA library of heads of gravid females was used to construct a microarray of expressed sequence tags (ESTs) at the University of Illinois. We used microarray analysis to screen for candidate genes that are differentially expressed between the two behavioral types. A total of 20 ESTs show at least a twofold expression difference at a stringent level of significance. Candidates with large and significant differences in expression were validated using real-time PCR; at least three are potential markers for the rotation resistance trait.

33

Characterization of Differential Transposon Activity in Vector Mosquitoes

Kuo, Hsien-Ying; Pledger, David
Biological & Health Sciences, Texas A&M-Kingsville, Kingsville, TX 78363

It is widely accepted that transposons have played an important role in the molecular evolution of genomes. DNA transposons also represent power genetic tools for functional genomics studies in animals. Enhancer trapping and insertion mutagenesis have been performed in a range of taxa by remobilizing previously integrated transposons. However, while transposon mediated germ line transformation of several medically important mosquito species has been achieved, transposons have not been observed to remobilize in mosquito germ lines. Analyses of transposon integrations in mosquito chromosomes and interplasmid events in mosquito embryos have indicated that transposon mobility may be altered in mosquitoes, as evidenced by observed increases in rates of non-canonical integration compared to those reported in other insect hosts. However, no direct evidence of interactions between mosquito proteins and transposons or transposases has been reported. We hypothesized that, if mosquito host factors do exist which influence transposon activity, there is likely to be interspecies variation in these processes. Thus, we characterized the activity of the *Mos1 mariner* transposon in *Aedes aegypti* and *Anopheles stephensi* mosquitoes by interplasmid transposition assay in an effort to observe whether any significant interspecies variation in the rates of aberrant or non-canonical integrations could be observed. While the transposition frequency of *Mos1* was observed to be similar in these two species, a significantly lower rate of non-canonical integrations occurred in *An. stephensi*, compared to that observed in *Ae. aegypti*. These observations of differential *Mos1* activity between these species serve as strong evidence of the existence of mosquito specific factors that influence transposon mobility. Based on these data, *Mos1* transposition in *Ae. aegypti* may represent a useful model for identifying host mechanisms that influence transposon mobility in other animals, as homologies in these processes may be shared across taxa.

34
VectorBase: A genome resource for arthropod vectors of human pathogens
*Lawson, Daniel & VectorBase.
EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom CB10 1SD

VectorBase (www.vectorbase.org) is an NIH-NIAID bioinformatic resource centre focused on invertebrate vectors of human pathogens. Three mosquito genomes are currently represented within VectorBase: *Anopheles gambiae* PEST strain, Aedes aegypti Liverpool isolate and the *Culex pipiens* Johannesburg isolate. Other species which VectorBase has responsibility include the body louse *Pediculus humanus* and the tick *Ixodes scapularis*.

Anopheles gambiae PEST strain was sequenced and published in 2002 by an international consortium and since then has undergone several rounds of re-annotation. Here we present the latest gene set AgamP3.4 with an emphasis on the integration of manually appraised gene predictions and the ongoing approaches employed to remove erroneous predictions and generally improve the quality of the predicted proteome.

Aedes aegypti Liverpool isolate was sequenced by The Institute of Genome Research (TIGR) and the Broad Institute and analysed in collaboration with VectorBase and the Aedes community. We shall present the published gene set AaegL1.1 and discuss the annotation process developed during the project which is the model employed in VectorBase for all genome annotations. A first round of re-annotation is planned for early 2008 which will integrate a significant number of new predictions as well as correcting erroneous ones in the current set.

Finally, the *Culex pipiens* Johannesburg isolate has been sequenced by the J. Craig Venter Institute (JCVI) and the Broad Institute. We shall present the initial gene set CpipJ1.1 and discuss some preliminary comparative analysis of the three mosquito genomes.

We shall discuss methods of data access and interrogation available through VectorBase and our relationship with the wider nucleotide and protein databanks. VectorBase is committed to maintain and re-annotate these genomes in the light of new experimental data or improvements in gene prediction algorithms. The re-annotation of the mosquito genomes is a community effort and we strongly encourage researchers to communicate errors and discrepancies in the data sets by sending e-mails to info@vectorbase.org

35
Desiccation-induced changes in gene expression in the red flour beetle *Tribolium castaneum*
*Li, Bin; Lord, Jeff; Hartzer, Kris; Lorenzen, Marcè D.; Oppert, Brenda; Beeman, Richard W.; Park, Yoonseong (BL, YP) Kansas State University, Department of Entomology, Manhattan, KS 66506; (JL, KH, ML, BO, RWB) United States Department of Agriculture, Agricultural Research Service, Grain Marketing and Production Research Center, Manhattan, KS 66502.

The red flour beetle, *Tribolium castaneum*, is a serious pest of stored products and represents a group of desiccation-resistant coleopteran insects. The genome sequence assembly includes >97% of the euchromatin, of which > 90% has been mapped onto the linkage groups. Based on computerized predictions for ~16,000 putative genes, we have designed an oligonucleotide microarray (Combimatrix Corp.) containing ~12,000 unigene 35-mer probes for this insect. We used this microarray to investigate genome-wide changes in gene expression in response to desiccation stress. Larvae were reared at 75% relative humidity (RH) for 15 days. The experimental group was then exposed to 22% RH for 24 hours, in four biological replications, while corresponding controls were maintained at 75% RH.

Based on the microarray data, we ranked candidate genes by the following criteria: low P values, high fold-changes, and high normalized fluorescence values. A total of 206 candidate genes were found to be differentially expressed (>1.7x). Forty one candidates were further examined by quantitative real-time reverse transcriptase PCR (qRT-PCR) to validate the microarray data. qRT-PCR revealed five up- and eight down-regulated genes in the desiccation stress treatment, among those genes that had shown >1.7-fold changes with $P<0.05$ after microarray analysis. Therefore, ~32% of the candidates in the microarray experiment were validated by qRT-PCR. Up-regulated genes included purine biosynthesis protein PURH (2x) and chitinase 9 (1.8x). Down-regulated genes included J domain-containing heatshock protein binding protein (11x), arylsulfatase B (3.5x), and Cathepsin B (4.3x). The list of genes indicates that the response to desiccation-stress involves general mechanisms of stress response, with possible involvement of changes in regulation of cuticle metabolism.

36
The VectorBase Manual and Community Annotation Submission pipeline
Neil F. Lobo[1], Robert V. Bruggner[1], Ryan Butler[1], Nate Konopinski[1], Daniel E. Lawson[2], Martin Hammond[2], Karine Megy[2], Kathryn S. Campbell[3], Susan Russo[3], Robert M. MacCallum[4], Seth Redmond[4], Pantelis Topalis[5], Emmanuel G. Dialynas[5], William M. Gelbart[3], Ewan Birney[2], Greg Meady[1], Frank H. Collins[1]
[1] University of Notre Dame, Notre Dame, IN USA.
[2] EMBL- European Bioinformatics Institute, Hinxton UK.
[3] Harvard University, Cambridge, MA, USA.
[4] Imperial College, London, UK.
[5] Institute of Molecular Biology and Biotechnology, Heraklion, Greece.

VectorBase, an NIH-NIAID Bioinformatics Resource Center for invertebrate vectors of human pathogens, currently hosts the genomes of Anopheles gambiae, Aedes aegypti, Culex pipiens, Ixodes scapularis and Pediculus humanus. One of the main challenges has been the incorporation of annotations from a wide variety of sources: sequencing centers, annotators, the community and the public sequence databases. VectorBase has developed a pipeline for data submission that has both rapid display on the genome browser yet retains quality control checks on the submitted data. Gene models are submitted using a web interface through formatted Excel spreadsheets and stored in a Chado schema relational database. Curators appraise the submissions using a web-based system and approved gene models are immediately displayed with appropriate accreditation on the genome browser. These gene models are then incorporated into the relevant organism gene build during the next round of genome annotation. The system is flexible and can be applied to other online genomic repositories to enable them to accept annotation updates from the community.

37
Transcriptional perturbations associated with the action of maternal-effect, selfish genes in *Tribolium*
*Lorenzen, Marcé D.; Brown, Susan J.; Beeman, Richard W.
(MDL, RWB) Grain Marketing and Production Research Center, ARS-USDA, Manhattan, KS, 66502; (SJB) Kansas State University, Division of Biology, Manhattan, KS, 66506.

Maternal-Effect Dominant Embryonic Arrest (*Medea*) genes are a novel class of maternally-acting, selfish genes that are widespread in natural populations of *Tribolium*, but are completely unknown in the remainder of the invertebrate world. *Medea* (*M*) genes are unique in combining a “maternal poison” and a “zygotic antidote” to gain a postzygotic survival advantage. *M* factors are also lethal in combination with the hybrid-incompatibility factor *H*. M^1, one of two known 3rd linkage group *M* loci, was positionally cloned, and proved to be associated with a 21-kb insertion adjacent to a 230-nt hairpin. While the M^1 region has been characterized molecularly, the detailed mechanisms of maternal lethality, zygotic protection, and *H*-incompatibility are still a mystery. The specific goals of this project are to: 1) develop *Tribolium*-specific whole-transcriptome microarrays and whole-genome tiling arrays; 2) compare transcriptomes of M^1M^4/H, $M^1/+$, M^1 and H prior to the time of *M*-induced developmental arrest; 3) conduct a genome-wide survey of microRNA expression in above genotypes prior to the time of *M*-induced developmental arrest; and 4) use quantitative RT-PCR to verify *M*-associated misregulation of identified transcripts.

38

Genomic Characterization and Evolution of Arthropod G-protein Coupled Receptors

Meyer, Jason M.*; Van Zee, Janice P.; Giraldo-Calderón, Gloria I.; Hill, Catherine A.
Purdue University, Department of Entomology, West Lafayette, IN 47907

G-protein coupled receptors (GPCRs) comprise a large gene family of transmembrane receptors that bind extra-cellular ligands and then initiate intracellular signal transduction cascades via their associated G-proteins. The GPCRs are integral to important biological functions such as neurotransmission, vision, and response to hormones. Many arthropod GPCRs may be potential targets for the development of novel pesticides, so work is ongoing to identify and characterize promising candidates in medically-important arthropods. We previously identified the repertoire of visual and non-sensory GPCRs in the malaria mosquito, *Anopheles gambiae*, and the yellow fever mosquito, *Aedes aegypti* (Hill et al., 2002; Nene et al., 2007), using bioinformatics approaches and manual annotation. Here we report the identification of the equivalent GPCRs from the genomes of the body louse, *Pediculus humanus*, and the Lyme disease tick, *Ixodes scapularis,* and provide a comparative analysis to investigate the evolution of GPCRs among these species.

Anopheles, Aedes, Pediculus and *Ixodes* each have more than 100 non-sensory GPCRs which were grouped into four major classes based on amino acid similarity to eukaryotic GPCRs. The body louse has the smallest number of non-sensory GPCRs (104) of any arthropod investigated to date. This is consistent with its small genome size and may reflect the simplified biology of this obligate ecto-parasite, relative to the other arthropods examined. Genome-genome comparisons have enabled the identification of GPCRs that are highly conserved in all four species. This suggests an ancient origin for some GPCRs that pre-dates the split between Chelicerata and Mandibulata (most recent common ancestor ~ 500 MYA) and points to the maintenance of common GPCR-mediated pathways in these lineages. However, GPCRs with potential novel gene functions have been identified within these common pathways. For example, the opsins, putative visual receptors, appear to have undergone expansions in both *Aedes* and *Anopheles* but not in the louse or tick, possibly reflecting the evolution of visual processes specific to mosquitoes. Despite conservation of amino acid sequence, many GPCRs display considerable variability in gene structure between these species. In general, *Pediculus* GPCRs are more fragmented, having additional exons and introns compared to their orthologs in these mosquitoes and the tick. Research is underway to confirm gene expression of each GPCR using reverse transcriptase-polymerase chain reaction (RT-PCR) analyses to validate our gene predictions and improve manual annotations. Initial experiments to determine tissue-specific expression of GPCRs using *in situ* hybridization have involved the *Aedes* opsins, which have advanced our understanding of mosquito vision at the molecular level.

39

The transcriptional profiles of mosquito salivary glands upon malaria parasite invasion

Michel, Kristin*; Zobnov, Evgeny M.
(KM) Division of Biology, Kansas State University, Manhattan, KS 66502, USA; (EMZ) Dpt. de Medecine Genetique et Developpement, Universite de Geneve, CH 1211 Geneva 4, Switzerland

We aim to gain insight into the contribution of salivary gland epithelial immune responses to malaria parasite transmission in the African malaria mosquito *Anopheles gambiae*. We recently identified an epithelial immune response that salivary glands mount against *Plasmodium berghei* sporozoites. So far, we know that the serine protease inhibitor SRPN6 partakes in this response, which not only limits to the number of parasites available for transmission but also influences their infectivity. However, neither the mode of action of this immune response nor it control is currently understood.

The specific hypothesis of this proposal is that mosquito salivary glands express a number of transcriptionally regulated immune factors that interact with the malaria parasite and influence parasite transmission. The aims are designed to explore the molecular make-up of the SRPN6-dependent salivary gland immune response and to gain insight into its regulation.

The aims of the proposed project are:
(1.) Identification of genes that are transcriptionally co-regulated within salivary glands during rodent malaria parasite infection by Affymetrix microarrays. Research towards this aim will be performed with support from the Gene Expression Facility and Bioinformatics Center at KSU.
(2.) Identification of potentially common transcriptional control mechanisms in salivary glands by bioinformatics analysis of upstream transcriptional regulatory elements.

This project is designed to provide two important results. The first constitutes one or several gene lists of candidates that potentially are involved in salivary gland-specific innate immune reactions against parasite invasion. The second will be a candidate list of transcription factors that potentially regulate innate immune reactions in mosquito salivary glands. Both results provide valuable preliminary data for future research and its funding by agencies such as NIH; they will aide subsequent, more detailed analyses of potential contributions of these genes to Plasmodium parasite killing within its mosquito vector.

40

Systemic RNAi in the red flour beetle, *Tribolium castaneum*

*Miller, Sherry C., Brown; Susan J.; & Tomoyasu, Yoshinori
Kansas State University, Division of Biology, Manhattan, KS 66506

RNA interference (RNAi) is a widely conserved mechanism of post-transcriptional gene regulation and has become a common method to knock down gene function in many model systems. However, in some model systems the method of introducing dsRNA to initiate the RNAi response poses inherent problems. In *Drosophila melanogaster*, the leading insect model system, we show that most larval tissues are incapable of mounting an RNAi response after injection of dsRNA into the hemolymph. Conversely, RNAi can be applied easily in post-embryonic stages of organisms that exhibit a systemic RNAi response, in which cells uptake dsRNA and transmit the RNAi effect throughout the entire body. The molecular basis responsible for the difference between organisms with and without a systemic RNAi response is unknown. Understanding the molecular basis of systemic RNAi in organisms with this response may provide insight regarding the current limitations of RNAi as a genetic tool in some organisms.

In the last few years, several insects have been shown to have a systemic RNAi response. Fortuitously, one of these insects is the red flour beetle, *Tribolium castaneum*, an emerging model system whose genome has recently been sequenced. In this project, we have characterized the systemic RNAi response in *Tribolium* by analyzing important features of dsRNA (such as size and dose) as well as tissue vulnerability and duration of the RNAi effect. We are also characterizing *Tribolium* homologs of the core components of the RNAi pathway as well as the genes essential for dsRNA uptake in other organisms. These approaches will reveal the conserved and divergent aspects of the systemic RNAi machinery, perhaps revealing ways to improve the application of RNAi as well as giving insight into the evolutionary history of this response.

41

A comparative annotation of Drosophilid dicistronic genes

*Motlhabi, Lala; Smith, Christopher D.
San Francisco State University, Dept. of Biology, San Francisco, CA 94132

Unlike viruses and bacteria, which can encode many proteins in a single messenger RNA (mRNA), most eukaryote genes express only one protein from each transcript. Due to the requirement for a modified 5'cap, the majority of eukaryote mRNAs have a single open reading frame (ORF) per transcript. In rare cases, eukaryotes have dicistronic genes with two non-overlapping open reading frames encoded in a single mRNA transcript. While the upstream gene is expressed through normal cap-dependent translation, it is unknown how the ribosome initiates translation of the downstream ORF. One mechanism used by viruses to internally recruit ribosome is a RNA structure called an internal ribosomal entry site (IRES). It is unclear to what extent eukaryotes dicistronic genes use IRES sequences, but whole genome sequence data and cDNA-verified gene annotations in the fruitfly, *Drosophila melanogaster*, reveal that dicistronic genes are much more prevalent than previously thought. The availability of 12 Drosophilid genomes allows for a comparative genomic annotation the ~50 well-supported *D. melanogaster* dicistronic genes and their orthologs. We hypothesize that conserved dicistronic genes will also share regions important for regulation of the downstream ORF, including putative IRES structures or cryptic promoters. We also predict that differences in the intron-exon structure of dicistronic genes between species will elucidate whether their regulation is conserved or the result of more recent gene structure changes. We have annotated several dicistronic genes in at least six Drosophilids (*D. simulans, D. sechellia*, *D. yakuba*, *D. erecta*, *D.*

pseudoobscura, *D.virliis)*. We used the FlyBase database along with Genscan, BLAST, Repeatmasker, and the MAKER annotation tool to identify the ORFs, matching cDNA sequences, align proteins, and find repeats present in syntenic dicistronic gene contigs. We also used CLUSTALX to align dicistronic annotations in six Drosophilid species and the Apollo annotation tool to refine these gene models. We found interesting trends in the gene structures of dicistronic genes across the species. For example, in some cases the Inter-Cistronic Region (ICR) region between the two ORFs reveals conservation across the species studied. We also observed decreases in ICR length with increasing evolutionary distance. To verify the presence and stability of monocistronic versus dicistronic transcripts *in vivo*, we are presently designing primers and will use quanitative PCR in several *Drosophilid* species. Future experiments will focus on employing RNA folding software (eg INFERNAL) to scan for potential IRES structures that can be tested *in vivo*. We conclude that comparative genomic annotation is a useful tool to dissect and elucidate the conserved regulatory and gene structure features of dicistronic genes.

42

Arthropod Genome Resources at NCBI

Murphy, Terence*; Church, Deanna; Kitts, Paul; Maglott, Donna; Tatusova, Tatiana; Pruitt, Kim; and the NCBI Genome Annotation Team.
NCBI/NLM/NIH/DHHS, 45 Center Drive, MSC 6510 Building 45, Bethesda, MD 20892-6510.

The National Center for Biotechnology Information (NCBI) provides genomic resources to support arthropod research at multiple levels including finding information about genomes, genes, sequences, expression, variation, and publications. These resources are extensively cross-linked and facilitate navigation across a broad range of biological information. A wide spectrum of resources are available for a variety of insects, including *Drosophila melanogaster* (fruit fly), *Anopheles gambiae* (African malaria mosquito), *Apis mellifera* (honey bee), *Nasonia vitripennis* (jewel wasp), and *Tribolium castaneum* (red flour beetle), and support for an additional 14 arthropod genomes is planned in the near future.

Entrez Gene is a gene-oriented database that includes protein-coding genes, genes encoding functional RNAs, and pseudogenes. Gene models for some organisms, such as *Drosophila melanogaster* and *Anopheles gambiae*, are based on annotation data provided by model organism databases such as FlyBase and VectorBase. For other organisms, gene models are based on sequences of known genes submitted to GenBank and predictions calculated with NCBI's Gnomon gene prediction tool. Gene provides RefSeq and GenBank accession numbers, map information, nomenclature, publications, and external links including links to model organism databases for each locus. It also supports GeneRIF (Gene References into Function), which allows users to contribute functional annotation in Gene.

The NCBI Map Viewer displays organism-specific maps produced by NCBI's genome annotation pipeline, including annotated genes, gene models, RefSeqs, DNA and RNA sequences aligned to the genome, UniGene clusters, and curated or GLEAN gene models produced through community annotation efforts. Map Viewer also has the capability to include genetic, radiation hybrid, and physical maps generated by research groups. It can cross-link the map data using common markers, and correlate maps to sequences and genes when appropriate sequence-based data is available. Map Viewer includes similar data for many organisms, encouraging discovery by comparative genomic analysis.

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

43

A Microarray-Based Analysis Of Transcriptional Compartmentalization In The Alimentary Canal Of *Anopheles gambiae* (Diptera: Culicidae) Larvae

*Neira Oviedo, M.; vanEkeris, L.; Corena-McLeod, M.D.P.; Linser, P.J.
(MNO, LVE, PJL) The Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL; (MDPC) Neuropsychopharmacology laboratory, The Mayo Clinic, Jacksonville, FL.

The alimentary canal of the larval mosquito displays a considerable degree of physiological compartmentalization among its different anatomical sub-divisions. We performed a comparative microarray analysis in order to identify transcripts which are particularly enriched in each gut section of the 4th instar larva of *Anopheles gambiae*. Based on the available annotation of the selected transcripts, we suggest that the metabolism and absorption of proteins and carbohydrates takes place mainly in the gastric caeca and posterior midgut, whereas the anterior midgut specializes in the metabolism and absorption of lipids. Transcripts encoding anti-microbial peptides were found to be enriched in the gastric caeca, and a high enrichment of transcripts associated with enzymes involved in xenobiotic detoxification was found in the anterior midgut. Furthermore, our data supports the notion that the region encompassing the hindgut and Malpighian tubes plays important roles in avoiding the excretion of nutrients, as well as in xenobiotic detoxification.

44

The salivary transcriptome of *Anopheles gambiae* larvae: a microarray-based study

*Neira Oviedo, Marco V.; Ribeiro, J.M.; vanEkeris, L.; Linser, P.J.
(MNO, LVE, PJL) The Whitney Laboratory, University of Florida, St. Augustine, FL; (JMR) Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD

Although the salivary transcriptome of adult mosquitoes has been thoroughly described in several recent papers, very little information exists regarding the biological role of larval salivary glands in the Culicidae. We used whole-transcriptome Affymetrix® chips to compare the transcriptional profiles of Anopheles gambiae larval (L4) salivary glands and whole larvae. We identified a total of 277 transcripts as being significantly enriched in the salivary glands. Based on available annotation for the known or predicted protein sequences encoded by these transcripts, 41 were identified as corresponding to secreted proteins, 233 as corresponding to non-secreted (housekeeping) proteins, and 3 as encoding proteins of unknown function. Based on functional annotation of the putatively secreted gene products, we propose that larval salivary secretions have roles in nutrient digestion, detoxification, immunity, and mouthpart lubrication. Interestingly, several components of the larval saliva (e.g. apyrase and serine proteases) have also been reported to exist in adult female saliva, where they are though to help regulate a vertebrate host's immune response to bloodfeeding. In conclusion, our results suggest that the salivary glands are important components of both the digestive and immune systems of larval mosquitoes, and that their study might provide clues about the evolution of adaptations to bloodfeeding observed in adults.

45

The Regulation of Innnate Immune Responses to the Endoparasitoid *Campoletis sonorensis* (Hymenoptera: Ichneumonidae) in the Lepidopteran Host *Heliothis virescens*

*Nusawardani, Tyasning[1)]; Shelby, Kent S.[2)]; Popham, Holly J.[2)]; Webb, Bruce A[1)].
[1)] University of Kentucky, Department of Entomology, S-225 Agricultural Science Center North, Lexington, KY 40546; [2)] USDA Agricultural Research Service, Biological Control of Insects Research Laboratory, 1503 S. Providence Rd., Columbia, MO 65203

Lepidopteran insects are one of the most important insect orders causing detrimental and global impacts on human nutrition and economics. Thus, many efforts have been directed to control the pest populations. Collecting genomics data from lepidopteran insects can provide a new tool to study the molecular mechanisms of many signaling pathways, including: those that play a role in insect immunity. In this study, we propose that the inhibition of immune response may determine the successful parasitization of the endoparasitoid *Campoletis sonorensis* in the lepidopteran host *Heliothis virescens.* We will investigate the regulation of innate immune responses by parasitization using custom microarrays of *H. virescens* ESTs. We present some preliminary findings from the identification of immune genes from *H. virescens* cDNA libraries derived from immune-challenged fat body tissues and hemocytes. From the cDNA libraries we obtained 44,844 ESTs that are assembled into contiguous and singleton sequences as putatively different transcripts. All putative transcripts will be used to construct a custom microarray to observe the profile of immune gene expressions in the *H. virescens*. To date we identified some putative immune genes that belong to some signaling pathways that are involved in

the insect immune responses, such as: prophenoloxidase cascade, NF-κB signaling, JAK-STAT signaling, and Ras/raf/mitogen-activated protein kinase signaling. The findings of this study will provide insights into lepidopteran innate immune responses and support efforts that target the immunity response pathway for lepidopteran insect pest control.

46

The *Helicoverpa armigera* (Lepidoptera:Noctuidae) genome project

Oakeshott, John G.; Fitt, Gary P.; Gordon, Karl H.; Edwards, Owain R.; Tay, Wee T.; *Cameron, Stephen L.; Batterham, Phillip.
(JGO, GPF, KHG, ORE, WTT, SLC) CSIRO Entomology, PO Box 1700, Canberra, ACT, 2601, Australia; (PB) University of Melbourne, Dept of Genetics & Bio21 Institute, Melbourne, 3010, VIC, Australia.

We announce a project to sequence the genome of the World's worst insect pest of agriculture, *Helicoverpa armigera*, which causes damage and control costs of at least $5 b/yr globally. It is a highly successful polyphagous coloniser and three of its close relatives are also major pests. *H. armigera* has evolved resistance to every class of chemical pesticides applied to it and is showing early signs of resistance to the two biopesticides also now used against it. It is probably the best understood pest insect in the world in terms of its ecology, behaviour and pathology and has proven remarkably phenomenon-rich (e.g. in respect of stress response, facultative diapause, feeding behaviour, rapid resistance evolution, mobility and migration). However, it is difficult to sustain or inbreed colonies in the lab and the genetic bases for all its ecological, evolutionary and biochemical phenomena have remained intractable. Existing genomics resources for *H. armigera* include sequences from over 60 BACs, 10000 ESTs and relatively inbreed lines established by CSIRO Entomology. Sequencing this genome in 2008 will allow us to develop genomics resources over the coming years including QTLs for resistance, host plant adaptations etc ; microarray covering the complete gene set; re-sequencing and functional/population genomic approaches to profile gene sequence diversity and the profiling of gene expression in response to factors such as host plant feeding, stresses and biopesticides.

47

Isolation and characterization of *ScOr83b*, a highly conserved odorant receptor from the stable fly (*Stomoxys calcitrans*)

*Olafson, Pia U.; Lohmeyer, K. H.
USDA-ARS, U.S. Livestock Insects Research Laboratory, Kerrville, TX 78028

The stable fly is a significant economic pest of confined and pastured cattle, with infestations impacting herd physiological and nutritional responses. As in other blood-feeding dipterans, olfaction guides many aspects of stable fly behavior including host/food source location, mate seeking, and ovipositional site selection. In an effort to identify genes that play a role in stable fly olfaction, we utilized degenerate primers to isolate a stable fly transcript encoding a putative odorant receptor, *ScOr83b*. *ScOr83b* displays amino acid sequence similarity to *CcOr83b* of the Mediterranean fruit fly and *AgOr7* of *Anopheles gambiae*, orthologues of an atypical odorant receptor that is highly conserved across diverse insect species. *ScOr83b* is expressed in both heads and bodies of unfed adult females and males, a spatial expression pattern similar to that reported for the corresponding odorant receptor in the red flour beetle (*TcOr1*). We are currently assembling relevant cDNA libraries that may enable the identification of additional sequences that play a role in the chemoreception pathway of the stable fly, i.e. odorant and chemosensory binding proteins.

Current stable fly control methods rely heavily on the management of fly breeding sites and the use of insecticides. Such cultural control practices are becoming difficult to implement from a labor cost standpoint, and chemical treatments are costly and less effective due to the fly's reproductive behavior. An understanding of the molecular mechanism of olfaction in the stable fly may provide an avenue for the development of alternative control methods.

48

ButterflyBase: A framework for comparative genomics in butterflies and moths

* Papanicolaou, Alexie; Gebauer-Jung, Steffi; McMillan, W. Owen; Jiggins, Chris D.; Vogel, Heiko; Heckel, David G.
(AP, SGJ, VH, DGH) Entomology, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany; (WOM) Department of Genetics, North Caroline State University, NC 27695-7614, USA; (CDJ) Department of Zoology, University of Cambridge, CB2 3EJ, UK.
* alexie@butterflybase.org

With over 100,000 species, a dozen model systems, a long tradition of research, a large community of evolutionary biologists, population ecologists, pest biologists and genome researchers, the Lepidoptera are an important insect group for studying both applied and basic science questions. A diversity of model species allows us to ask phylogeny-wide questions such as: How do the genomes of a butterfly and a moth differ? A moth and a fruitfly? How does diet and habitat shape the genome of a species? How does the gene-content of the Z/W sex chromosomes differ across the phylogeny and how does it compare with the ca. 28 autosomes? As technology allows us to generate a diverse collection of inexpensive partial genomes, such questions are a sample of what multi-disciplinary comparative approaches in Lepidoptera can focus on. Even though Lepidopterists have not fully entered the genomic era, the data available can be embedded in a phylogenetic framework to create a unique dataset. We describe the use of Sanger and 454 Expressed Sequence Tag (EST) datasets for Lepidoptera to perform transcriptome-wide surveys for the presence of selection using Single Nucleotide Polymorphisms (SNPs). We also present our current attempts to provide a quality assessment in multi-species orthologue prediction in the presence of missing data. These tools are being integrated into the user-friendly platform for comparative genomics of Lepidoptera: ButterflyBase (http://www.butterflybase.org). This nascent platform already supports many needs of the lepidopteran research community, including molecular marker development for synteny research, orthologue prediction, gene annotation & guidance of functional experiments, and detection of Lepidoptera-specific or rapidly evolving proteins likely to be involved in important ecological processes. Currently, ButterflyBase is expanding to include non-cDNA derived sequence, ecological and mapping data for key species. The ultimate aim of ButterflyBase is to evolve into a multi-disciplinary platform for Lepidoptera with the vision that any researcher, irrelevant of their computing skills and discipline, can acquire and compare experimental data of their favourite protein, gene family, developmental process or organism.

49

Molecular analysis of of entomopathogenic fungi *Metarhizium anisopliae*

*Krutmuang Patcharin
Department of Entomology, Faculty of Agriculture, Chiang Mai University, 50200 Thailand
Contact Address: E-mail: p-charin@chiangmai.ac.th

Metarhizium anisopliae is a mitosporic entomopathogenic fungus that has been exploited extensively as biological control agent (BCA) against several pests. *Metarhizium anisopliae* isolates from several insect hosts and from various sugar cane growing areas of Thailand, were examined for genetic diversity using polymerase chain reaction (PCR)-based technology, involving amplified fragment length polymorphism (AFLP) was used to assess the genomic variability between 4 isolates of *Metarhizium spp* strains. Amplified fragment length polymorphism (AFLP) analysis of entomopathogenic fungus evidence provides a means of obtaining a reproducible DNA profile in a relatively short period of time in species for which no sequence information is available. Genomic DNA from mycelium of each strain was optimised and the use of cetyltrimethyl ammonium bromide (CTAB) and sodium chloride (NaCl) was incorporated. All strains could be typed in these conditions. DNA were double-digested by two restriction endonucleases (EcoRI and MseI) and ligated to oligonucleotide adapters. Two consecutive PCR reactions (pre-amplification and selective amplification) were performed using a modification of the AFLP protocol described by Gibco (Invitrogen, Rockville, MD). The DNA fragments were separated by electrophoresis using silver staining for band visualisation. Based on 23 AFLP primer combinations, a total of 1504 bands were detected. An average of approx. 65 bands were scored for each primer pair. Among of which 3 polymorphic fragments (obtained from E-AGG/M-CAA, E-AGG/M-CAA, E-AGG/M-CAA) were identified as potentially a strain specific. DNA fragments of between 0.26 and 0.38 kp were obtained. These markers have practical utility for (1) establishing conspiracy in the cultivation and distribution of *Metarhizium sp* (2) identifying geographic sources. The results also suggest that AFLP markers may be useful for the tracking of specific biocontrol strains in the field.

50
Observing Argentine ant foraging behaviors at the molecular level
Placek, Jennifer*; Smith, Christopher. D.
San Francisco State University. Department of Biology, San Francisco, CA 94132.

Linepithema humile (Argentine ant) is an invasive species that negatively affects other plants, animals, and whole ecosystems. It's success is due, in part, to its extreme aggression, cooperation, and its successful foraging. The foraging gene (*for*) has been identified in several insect species including the honeybee, harvester ant, fruit fly and wasps. While it is known that the honeybee foraging gene is up-regulated in individuals outside of the nest, it is down-regulated in foraging harvester ants, and the expression level of the foraging gene in other hymenopteran species remains unstudied. We have used bioinformatics to identify the foraging gene from Argentine ants and will use quantitative PCR to compare the expression levels of the gene between foraging and non-foraging ants. We will also expose foraging ants to various semiochemicals and use computer-assisted video monitoring to find chemicals that increase or decrease foraging and/or aggressive behaviors in Argentine ants. These studies will be helpful to understand the molecular basis of invasive behavior, identify more benign chemicals to assist control and eradication programs, and potentially allow reintroduction of native species to the environment.

51
Whole Genome Re-Sequencing Strategies In Drosophila: Short Reads Vs. Long Reads Vs. Assembled Sequence
Stephen Richards*[1], Trudy Mackay[2], Jeff Reid[1], James Knight[3], Mathew Bainbridge[1], Maithreyan Srinivasan[3], Jason Affourtit[3], Brian Desany[3], Michael Egholm[3] and Richard Gibbs[1]
[1]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX. [2]Department of Genetics, North Carolina State University, Raleigh, NC. [3]454 Life Sciences, Branford, CT.

Drosophila is an excellent model system for the accurate measurement of quantitative traits. We have proposed a Drosophila Genetic Reference Panel of 300 strains to be sequenced allowing the association of phenotype and genotype by the fly community.

To date, we have sequenced 3 genomes of inbred phenotyped strains and the reference strain of *D. melanogaster*. Line 360 was sequenced to 12X coverage on the 36bp short read illumina platform, 12X sequence coverage on the 250bp read-length 454FLX platform. As a line 360 gold standard we produced 1Mb of high quality (>Q40) unique Sanger sequence alignments. We assembled the 454 data using the Atlas assembly suite of tools. The other lines were sequenced to 12X 454 coverage – with line 375 sequenced using the extra long read FLX kit (500bp reads) and assembled using the Newbler platform.

Comparisons of all these data at different coverage levels aligned to the reference sequence using Mosaik suggests the following conclusions: 454 high quality completeness of genome analysis increases from low levels at 2X coverage until plateau at 99% of the gold standard sequences between 10 and 12X coverage. Short reads are harder to align, and thus give lower genome coverage and lower quality consensus sequence. Snp detection was fine in both platforms in regions of high quality multiple read coverage, but short read platforms require additional coverage due to the higher substitution error rate. Alignment issues with 36bp reads prevent the identification of indels greater than 3bp. 454 reads are better at indel detection, with the exception of 1 and 2bp length changes in longer homopolymers. Re-sequencing of the reference strain at 12X 454 FLX coverage identified ~ 2,000 polymorphic bases, and indicated an extremely low false positive rate. We will additionally present snp and indel detection statistics for assembled contigs alignment in comparison to read alignment data.

52
Whole genome annotation of the spider mite *Tetranychus urticae*
*Robbens, Steven; Rouzé, Pierre; Grbic, Miodrag; Van de Peer, Yves
(RS, PR, YVdP) Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium. Bioinformatics and Evolutionary Genomics, Department of Molecular Genetics, Technologiepark 927, Ghent University, B-9052 Ghent, Belgium; (GM) University of Western Ontario, Department of Biology, 1151 Richmond St, London, ON N6A5B7, Canada

Application of chemical pesticides in agriculture represents one of the major costs of agricultural production and a key source of environmental pollution, destruction of wildlife, and introduction of known carcinogens in the human food chain. The current need for novel methods of pest control coincides with unprecedented advances in genomic analyses of crop plants which open novel avenues for biotechnology. In contrast, genomic resources of

pest species, necessary for development of new pest control strategies through plant breeding and/or biotechnology, are lagging behind. The current gap in knowledge on pest genetics, genomics and plant-pest interactions is a major obstacle for the development of alternative pest control strategies. To circumvent these problems, whole genome sequencing of a major agricultural pest, the two-spotted spider mite *Tetranychus urticae*, has been recently approved. The spider mite whole genome sequencing project, funded by US Department of Energy (http://www.jgi.doe.gov/sequencing/why/CSP2007/spidermite.html) will create one of the first genome sequences of the pest herbivore providing a unique resource for the development of novel pest control strategies.

Tetranychus urticae belongs to the Chelicerates, which is the second largest group of arthropods and occupies a basal phylogenetic position within the Arthropoda. They are considered as one of the major pests in agriculture. Due to its polyphagous feeding strategy (feeding on over a 1000 plant species) and its rapid development (generation time 7 days in hot season) this species represents a key pest in greenhouse crops including tomatoes, cucumbers, peppers and ornamental plants. It is also a major pest in annual field crops such as maize, soybeans and barley. *T. urticae* attacks major horticultural crops including apples, pears and grapes. In addition, spider mite is a pest in hot and dry climates and recent computer modelling predicts that its significance will increase in EU as a consequence of global warming.

For the genome annotation of *T. urticae*, we will use EuGène, a highly integrative protein coding gene prediction platform for eukaryotes. Instead of a priori matching a given mathematical model to the gene finding problem, EuGène collects all the sources of evidence exploited for gene prediction and the typical list of evidences for expert annotation can be divided easily into two categories, namely 1) statistical evidence, or more generally, in silico evidence provided by dedicated mathematical models with the purpose of catching the specific properties of gene components (e.g. local information that a region seems to be coding for a protein), 2) similarity with sequences of previously documented molecules. Once a structural annotation has been obtained, we will perform a functional annotation, where a (potential) function will be assigned to the proteins that are encoded by the gene models defined by the structural annotation. Finally, the obtained genes will be used, together with genes from other members of the Arthropoda, for creating gene families. *T. urticae* specific gene families can potentially be used as new targets for biotechnological pest control (RNAi).

53

Predicting Heterochromatic Regions and Genes in Insect Genomes

Rossi, Fabrizio; Celniker, Susan E.; Dimitri, Patrizio; Karpen, Gary H.; *Smith, Christopher D.
(CDS) San Francisco State University, Dept. Biology, San Francisco, CA 94132; (SEC, GHK, CDS) Drosophila Heterochromatin Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; (FR, PD) University of Rome La Sapienza, Dept. Genetics & Molecular Biology, Rome Italy.

Metazoan genomes are composed of two general types of sequence, euchromatin and heterochromatin. Euchromatin is typically repeat poor and contains most of the protein-coding genes. In spite of its reputation as being repetitive ‘junk’ DNA, heterochromatin contains actively transcribed genes, sequences essential for chromosome stability (i.e. centromeres & telomeres), and has been shown to contain non-coding RNAs essential for suppressing transposable elements throughout the genome. However, in most genome sequencing projects the heterochromatic regions are under-represented in the final genome assembly because they are repeat-rich, difficult to clone, and hard to assemble with current software. Thus, while the wealth of sequence data from over 20 assembled insect genomes has greatly aided gene annotation and comparative studies of euchromatin features, little light has been shed on large-scale genome organization or genes found in the heterochromatin. We have undertaken extensive sequencing and annotation of the heterochromatic regions of the *Drosophila melanogaster* genome and discovered several hundred genes embedded within this high-repeat content region. To test whether these genes were also heterochromatic in other species, we used a comparative genomics approach to identify putative orthologs in other insect genomes and then measured the repeat content of the surrounding regions. We found that *D. melanogaster* heterochromatin gene orthologs are often embedded in repeat-rich regions in other insects and have used FISH to confirm that these genes are close to the centric heterochromatin in several other Drosophilids. We observed that many orthologs were localized to less repetitive scaffolds with increasing evolutionary distance, suggesting that some *D. melanogaster* heterochromatin genes may be derived from a euchromatic ancestor. However, we also noted several cases of genes that appeared to be embedded in repeat-rich regions, which may indicate that some genes require being in a heterochromatic context for proper function. Our results indicate that local repeat content may be a sufficient proxy to estimate whether individual genes or whole scaffolds are heterochromatic in other species and that tracking the repeat density surrounding gene orthologs through evolution may be a useful tool to study heterochromatin in less assembled genomes.

54

Developing genomics tools for *Frankliniella occidentalis*, the western flower thrips

*Rotenberg, Dorith; Fellers, John P.; Chanbusarakum, Lisa J.; Ullman, Diane E.; and Whitfield, Anna E.
(DR, AEW) Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506;
(JPF) USDA-ARS-PSERU, Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506;
(LJC, DEU) Department of Entomology, University of California, Davis, CA 95616

The western flower thrips, *Frankliniella occidentalis,* is the most economically-important agronomic pest within the insect order *Thysanoptera*. Despite the importance of thrips as pests, there is little knowledge of the *F. occidentalis* genome or gene functions at this time. The aim of our research is to develop sequence and genomics tools for *Frankliniella occidentalis.* In future studies, we will use these tools to investigate the molecular interactions between *Tomato spotted wilt virus* (TSWV) and *F. occidentalis*. We hypothesize that TSWV infection of *F. occidentalis* alters the insect transcriptome, resulting in the up-regulation of insect genes involved in recognizing and responding to the invading virus. To test our hypothesis, we are pursuing the following objectives: 1) create expressed sequence tag (EST) resources for larval thrips and 2) identify differentially-expressed genes in TSWV-infected insects. We constructed a normalized cDNA library from larval thrips. The 5' and 3' sequences of 1,248 clones were determined and assembled into 281 contigs and 1,281 singletons. These ESTs represent the first significant exploration of the thrips transcriptome.

55

Developing genomics tools for *Peregrinus maidis*

*Rotenberg, Dorith; Hogenhout, Saskia A.; Nelson, James C.; and Whitfield, Anna E.
(DR, JCN, AEW) Kansas State University, Department of Plant Pathology, Manhattan, KS, 66506; (SAH) The John Innes Centre, Department of Disease and Stress Biology,
Norwich Research Park, Norwich, United Kingdom

Over 70% of plant-infecting viruses are transmitted from one host to another by arthropod vectors. The corn planthopper, *Peregrinus maidis,* is an economically important agronomic pest because it is both a direct pest of crops and an efficient vector of plant viruses. However, there is little knowledge of the *P. maidis* genome or gene functions. The aim of our research is to develop arthropod sequence and genomics tools for *P. maidis.* We will use these tools to investigate the molecular interactions between *Maize mosaic virus* (MMV) and *P. maidis*. We hypothesize that MMV infection of *P. maidis* alters the insect midgut transcriptome, resulting in the up-regulation of insect genes involved in recognizing and responding to the invading virus. To test our hypothesis, we are pursuing the following objectives: 1) create expressed sequence tag (EST) resources for the *P. maidis* midgut; 2) identify differentially-expressed genes in MMV-infected *P. maidis* midguts; and 3) determine the biological role of candidate genes in MMV-infected insects. To date, we have constructed normalized cDNA libraries from alimentary canals harvested from larval planthoppers reared on healthy (MMV– library) and virus-infected corn plants (MMV+ library). Preliminary analyses of a subset of sequences obtained from each library reveal diverse, high quality sequences. We have sequenced 20,000 clones picked from the two gut libraries and analysis of these sequences is under way.

56

Whitefly Genomics and Proteomics Resources for Elucidating Virus-Vector Interactions

Saripalli, Chandra.; Cicero, Joseph M.; Hunter, Sarah C.; Brown, Judith K.*
Department of Plant Sciences, The University of Arizona, Tucson, AZ 85721 USA

The whitefly *Bemisia tabaci* (Gennadius) complex is the vector of the genus *Begomovirus* (family, Geminiviridae). Begomoviruses cause debilitating crop diseases worldwide. New World begomoviruses are transmitted in a persistent manner but are non-replicative and non-transovarial in the whitefly vector. The *B. tabaci* genome sequence is not yet available and thus, a number of important biological questions that could be addressed using genomic and proteomic approaches remain unexplored.

The begomovirus transmission pathway is hypothesized to involve receptor-mediated exo/endocytosis and possibly other mechanisms by which virions cross the midgut and salivary gland membrane barriers. Once in the hemolymph, virions are thought to interact with a primary endosymbiont synthesized 60S heat shock protein that may protect, stabilize, and/or mask virions while in transit to the salivary glands. It is none-the-less hypothesized

that during and/or following the ingestion-acquisition phases the whitefly innate immune system may be stimulated in response to virion circulation, producing 'stress' and other 'defense' proteins.

To investigate these and other relevant hypotheses, a large-scale 5' EST (expressed sequence tag) project was undertaken in which a non-subtracted cDNA library was constructed, cloned, and sequenced for the B biotype of *B. tabaci*. The sources of mRNA for library construction consisted of non-viruliferous adults, instar (crawler to pupae), eggs, and adults that had been allowed an acquisition access feeding period (AAP) on *Tomato yellow leaf curl virus*- and *Tomato mottle virus*-infected tomato plants (Leshkowitz et al., 2006, BMC Genomics 7:79-98.).

We undertook a re-annotation of whitefly ESTs using current NCBI database entries (to February 2008). The EST analysis and assembly were carried out using stackPack v2.2 (http://gene3.ciat.cgiar.org/stackpack/) resulting in 10,651 high quality ESTs in 659 clusters, 3,896 singletons and 4,429 contaminants from a total of 18,976 clones sequenced. The output consensi were then subjected to BLAST-N and BLAST-X analysis to search available ESTs, unigenes and proteins generated by 9 ongoing or completed insect genome projects. Translated proteins were assigned to one or more functional Gene Ontology (GO) (http://www.geneontology.org) categories that describe organismal gene and gene product attributes. A suite of quantitative or 'real-time' polymerase chain reaction (qPCR) primers were designed to amplify and quantify gene expression for candidate genes representing GO categories with an emphasis on those with predicted involvement(s) in biological stress responsiveness, some of which may be up- or down-regulated as whiteflies feed and/or as virions circulate in the transmission pathway.

Additionally, our group has initiated an investigation of the proteome of the B biotype of *B. tabaci* allowed a 48-hr AAP on *Squash leaf curl virus* -infected pumpkin or on healthy pumpkin control plants. To date we have made protein extracts from whole adult whiteflies and from mass extirpations of individual organs of the digestive system (alimentary canal and salivary glands) and analyzed the proteins in the bands resolved by 1-dimensional electrophoresis. From these preliminary experiments, we have annotated 750 unique proteins using the NCBI non-redundant database. The prevalent protein families identified via gel electrophoresis followed by MudPIT analysis include cell surface proteins, chaperone-associated proteins, and kinases.

57

Unriddling the genome: How to identify (new) genes involved in head development
or
The GEKU insertional mutagenesis screen

Schinko, J.B.*; Küster, E.; Trauner, J.; Lorenzen, M.; Beeman, R.W.; Brown, S.; Klingler, M.; Wimmer, E.A.; Bucher, G.
(JBS, EK, GB, EAW) Georg-August-University Göttingen, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Developmental Biology, Göttingen, Germany; (JT, MK) Friedrich-Alexander-University, Institute for Biology, Div. Developmental Biology, Erlangen, Germany; (ML, RWB) USDA-ARS-GMPRC, Manhattan, KS, USA; (SB) Kansas State University, Division of Biology, Manhattan, KS, USA

In order to learn about the genes involved in development, various homologous genes have been isolated from the beetle and their expression and function have been compared to their fly orthologs. However, this approach will not identify genes that are crucial in the beetle but not in the fly. In order to identify such genes, the GEKU consortium (**G**öttingen – Gregor Bucher and Ernst Wimmer; **E**rlangen - Martin Klingler, **K**ansas State University - Sue Brown; **U**SDA grain marketing and production research center - Richard Beeman) perform the **"GEKU"** insertional mutagenesis screen.

New insertions are efficiently generated by crossing helper and donor strains. The helper strain carries a *Minos* transposable element encoding *piggyBac* transposase. The donor strain carries a *piggyBac* element with the eye-specific 3xP3-EGFP reporter expression and an easily detectable enhancer trap in the abdomen. In offspring that has lost the enhancer trap but retains the eye specific expression the mutator must have jumped from its original position to a novel locus.

We present the screening procedure and the results of the first 3.000 new insertions that we have generated in Göttingen and that we have screened for enhancer traps, lethality, sterility and embryonic cuticular phenotypes. This has identified a number of genes required for head development, segmentation and the development of appendages and some larval and pupal organs.

58

Preliminary analyses of the *Heliothis virescens* transcriptome

*Shelby, Kent. S.; Terenius, O.; Popham, Holly J.R.
(KSS, HJRP) USDA-ARS-BCIRL, 1503 S. Providence, Columbia, MO 65201; (OT) University of California--Irvine, Molecular Biology and Biochemistry, Irvine, CA 92697.

Generation and analysis of genomic resources for the tobacco budworm, *Heliothis virescens*, are necessary for detailed functional genomic studies of the physiology and biochemistry of this highly destructive pest. In this study we present preliminary analyses of the ~45,000 publicly available *H. virescens* ESTs. ESTs were retrieved, assembled and annotated into 8074 contigs, and 7231 singletons; totaling 15,301 putative transcripts. 6497 of these have a significant BLAST to known homologs from other organisms, primarily insect species. Expression of selected immune response homologs in the fat body, midgut and hemocytes were monitored using real time PCR following bacterial and baculoviral infection. Results will be discussed.

59

Genomics and proteomics of neuropeptide in the black-legged tick *Ixodes scapularis*

*Šimo, Ladislav; Park, Yoonseong
Kansas State University, Entomology Depatment, Manhattan, KS 66506

Peptidergic signal system is an ancient cell-cell communication mechanism and involved in numerous behavioral and physiology events of multicellular organisms. We have been interested in exploring neuropeptidergic system in an Ixodes tick that transmits Lyme disease. Because of the importance of *I. scapularis* to human health, this species was chosen for a whole genome sequencing project (Hill and Wikel, 2008, Trends Parasitol. *21*, 151-153). Immunohistochemistry using a total of 14 different antibodies, originally raised against various insect neuropeptides, were performed to identify immunoreactive cells and neural projections. Specific staining patterns in the immunohistochemitry indicate that the tick nervous system is a rich source of diverse neuropeptides related to those identified in insects and other arthropods. We also have performed the search of putative genes encoding neuropeptides from the Ixodes genomic database and expressed sequence tag sequences. Several criteria applied in the search were 1) homology to known neuropeptide sequences, 2) presence of multiple active peptide sequences that are homologous one another, and 3) presence of canonical processing sites at the carboxy terminal of each mature peptide. We have identified more than 25 candidate neuropeptide genes by using these criteria from the currently available data. We also used peptidomics tools (proteomics) to identify the neuropeptides in synganglia extract using desorption ionization-time of flight/time of flight (MALDI-TOF/TOF). We identified a 1322 Da peptide with the sequence ASD<u>W</u>NRLSGM<u>W</u>amide, having the signature of typical myoinhibitory peptide (MIP) W(6X)Wamide motif. The MIP immunoreactivity indicates that the MIP is produced from a number of cells in the synganglion and their projections reach to the salivary glands, suggesting its role in controlling the salivary glands in tick. This sequence was also identified in a genome shotgun sequence having two consecutive MIP-like peptides interrupted by a canonical processing site. This is the first report showing a neuropeptide innervating the salivary gland of tick in our knowledge.

60

How Do Insects Fly: Bioinformatics Analysis of the Myofibrillar Protein Projectin

Southgate, Richard; Philipp, Drew; Ayme-Southgate*, Agnes.
College of Charleston, Department of Biology, Charleston SC 29401

Insect flight is powered by two types of muscles known as either synchronous or asynchronous. More basal insects have synchronous-type flight muscles, and the acquisition of asynchronous physiology is believed to have occurred on multiple occasions during insect evolution. It is likely that the transition from synchronous to asynchronous was a multi-step process, requiring adaptations through changes in both protein composition and structure. The asynchronous mode relies on stretch-activation, a process leading to multiple contractions for each nerve impulse. Stretch activation is made possible, at least in part, by the presence of connecting filaments that link Z-bands and myosin filaments and provide high resting stiffness. In *Drosophila melanogaster,* these structures are composed of two proteins, projectin and kettin. *Drosophila* projectin is a very large protein, 1,000kDa in size, with a modular organization using two repeated motifs, the fibronectin III (Fn) and Immunoglobulin (Ig) domains, as well as several unique sequences. One of these unique sequences, the PEVK domain, is known to work as an elastic region together with the NH_2-terminal Ig domains (N-Ig) in other muscle proteins such as vertebrate titin. The PEVK domain is so called because of its high content (50% and above) in the amino acids P, E, V and K, allowing unusual folding and unfolding properties. The physiological mechanism of asynchrony may, therefore, have evolved partly through sequence modifications in projectin's PEVK and N-Ig

domains. To address this question we initiated a molecular analysis of projectin in various insects with either synchronous or asynchronous flight muscles. We obtained genomic sequence information in thirteen different insects using data available from insect genome projects, as well as, some partial sequences from direct sequencing for other non-project insects. Data will be presented describing the gene annotation process, the analysis of the gene structure (exon-intron pattern), and its evolution from basal to derived insects. Evolution of different projectin domains will be presented in relation to the proposed functions of the protein in stretch activation and myofibrillar assembly. Our current data for the PEVK and N-Ig regions indicate that indeed part of the PEVK domain presents a highly variable sequence in different insect orders. Complex alternative splicing of the PEVK domain is also present in many insects leading to PEVK regions with variable length and PEVK content. Phylogenetic analysis of the projectin gene and protein will be presented, in particular the interesting discrepancy with the current insect phylogenetic tree, in relation to the position of the Hymenoptera (bees and wasps).

61
Bioinformatic analysis of ESTs from termite castes
Matthew Steller and Srini Kambhampati
Department of Entomology, Kansas State University, Manhattan, KS 66506

Termites (Isoptera) are separated into morphologically and behaviourally distinct castes of sterile workers and soldiers, and reproductive alates. Previous research has implicated that caste differentiation is controlled genetically. We compared the gene expression patterns of each caste and two larval life stages of *R. flavipes* in an attempt to better understand caste-specific gene expression. Five caste- and life stage-specific cDNA libraries were constructed , cloned, and sequenced to generate an EST library of almost thirteen thousand sequences. All sequences were trimmed and arranged into contiguous sequences using Sequencher 4.7. Putative gene functions were assigned based on a tblastx swissprot search. Expressed gene patterns for each library were compared using Blast2GO. Seventy to ninety percent of sequences had no significant (e-value<1E-10) homology to sequences in existing databases. ESTs of workers and soldiers were from the widest array of functional classes. All libraries showed sequences with putative functions assigned to reproduction, which is unexpected in the non-reproductive soldier and worker castes. The biological process ontology was most variable for the worker caste with three unique terms (reproductive process, biological adhesion, and growth), however soldiers did have 10% of the ontology terms represented as sequences that made up less than 1% of the total, while workers consisted of 8% indicating increased functional variability in the soldier ESTs. As expected, each caste appears to have unique expression patterns; of special interest are genes associated with transcriptional control and reproduction. We plan on performing fine-scale analysis of caste-specific expression patterns using a microarray of over 2,000 ESTs.

62
Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*
*Tomoyasu, Yoshinori; Miller, Sherry C.; Tomita, Shuichiro; Schoppmeier, Michael; Grossmann, Daniela; Bucher, Gregor.
(YT, SCM) Division of Biology and K-State Arthropod Genomics Center, Kansas State University, Manhattan, Kansas 66506, USA; (ST) Insect Genome Research Unit, National Institute of Agrobiological Sciences, 1-2, Owashi, Tsukuba, Ibaraki 305-8634, Japan; (MS) Universitat Erlangen, Institut fur Biologie, Abteilung fur Entwicklungsbiologie, Staudtstr., D-91058 Erlangen, Germany; (DG, GB) Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Georg-August-Universität Göttingen, Abteilung Entwicklungsbiologie, Justus-von-Liebig-Weg, 37077 Göttingen, Germany

Background: RNA interference (RNAi) is a highly conserved cellular mechanism. In some organisms, such as *Caenorhabditis elegans*, the RNAi response can be transmitted systemically. Some insects also exhibit a systemic RNAi response. However, *Drosophila*, the leading insect model organism, does not show a robust systemic RNAi response, necessitating another model system to study the molecular mechanism of systemic RNAi in insects.

Results: We used *Tribolium*, which exhibits robust systemic RNAi, as an alternative model system. We have identified the core RNAi genes, as well as genes potentially involved in systemic RNAi, from the *Tribolium* genome. Both phylogenetic and functional analyses suggest that *Tribolium* has a somewhat larger inventory of core component genes than *Drosophila*, perhaps allowing a more sensitive response to double-stranded RNA (dsRNA). We also identified three *Tribolium* homologs of *C. elegans sid-1*, which encodes a possible dsRNA channel. However, detailed sequence analysis has revealed that these *Tribolium* homologs share more identity

with another *C. elegans* gene, *tag-130*. We analyzed *tag-130* mutants, and found that this gene does not have a function in systemic RNAi in *C. elegans*. Likewise, the *Tribolium sid*-like genes do not seem to be required for systemic RNAi. These results suggest that insect *sid-1*-like genes have a different function than dsRNA uptake. Moreover, *Tribolium* lacks homologs of several genes important for RNAi in *C. elegans*.

Conclusion: Although both *Tribolium* and *C. elegans* show a robust systemic RNAi response, our genome-wide survey reveals significant differences between the RNAi mechanisms of these organisms. Thus, insects may use an alternative mechanism for the systemic RNAi response. Understanding this process would assist with rendering other insects amenable to systemic RNAi, and may influence pest control approaches.

63

Gene Duplication and Genome Organization of the Ixodidae

*Van Zee, Janice P.; Meyer, Jason; Schlueter, Shannon; Schlueter, Jessica; Hill, Catherine.
(JPVZ, JM, CH) Purdue University, Department of Entomology, West Lafayette, IN, 47907; (SS) Purdue University, Computer Information and Technology Department, West Lafayette, IN, 47907; (JS) Purdue University, Agronomy Department, West Lafayette, IN, 47907, USA
Contact information: Janice Van Zee, Dept of Entomology, Purdue University, 901 W State St, West Lafayette, IN 47907 USA; Phone: 765-496-1513; Fax: 765-496-1219; Email: jpage1@purdue.edu.

Ixodid ticks (subphylum Chelicerata, class Arachnida, suborder Ixodida) are obligatory hematophagous ectoparasites of worldwide medical and veterinary importance. Ticks transmit a variety of viruses, bacteria and protozoa and cause direct damage to their host via attachment and feeding. Our research team is analyzing the genomes of several ixodid ticks to improve understanding of tick biology and transmission of tick-borne diseases. Previous work has shown that species of hard ticks (family Ixodidae) and soft ticks (family Argasidae) have large genomes with haploid genome sizes ranging from 1 to more than 7Gbp. Preliminary evidence suggests that repetitive DNA and possibly segmental duplications may contribute significantly to the genome size of ixodid ticks. However, genome composition and organization have been largely unstudied in the Ixodida and very little is known about the evolution of ixodid tick genomes.

Gene and segmental duplication events and the accumulation/proliferation of repetitive elements are recognized as major driving forces in eukaryotic evolution providing biological variants and generating biodiversity. In this study we used bioinformatics approaches to investigate the frequency of gene duplication events in four species of hard ticks, namely the prostriate tick *Ixodes scapularis* (Lyme disease tick), and the metastriate ticks *Amblyomma variegatum* (Lone start tick), *Rhipicephalus appendiculatus* (Brown ear tick) and *Rhipicephalus* (Boophilus) *microplus* (Southern cattle tick). We analyzed tentative consensus sequences (TCs) derived from tick ESTs using Vmatch software to identify putative paralogs and investigate the possibility of whole genome or segmental duplications in the pro and metastriate lineage of ticks. Synonymous substitution rates were used to estimate the timing of putative gene duplication events in each species. Our data suggest that a significant fraction (20-25%) of the expressed genome may have recently duplicated in all four species. The ratios of nonsynonymous to synonymous nucleotide substitution rates indicate that approximately 25% of these putative duplicated genes are under positive selection and suggest that they are evolving novel functions. Bioinformatics analyses are also ongoing to identify and characterize LTR retrotransposons in the *I. scapularis* genome and the contribution of these elements to genome size. These studies are contributing to our understanding of tick genome evolution and complement ongoing efforts to sequence and analyze pro- and metastriate ticks.

64

Characterizing the Transcriptome of a Nonmodel Organism, the Butterfly *Melitae cinxia*, using 454 Pyrosequencing

*Vera, J. Cristobal; Wheat, Christopher W.; *Fescemyer, Howard W.; Frilander, Mikko J.; Crawford, Douglas L.; Hanski, Ilkka; Marden, James H.
(JCV, CWW, HWF, JHM) The Pennsylvania State University, Department of Biology, 208 Mueller Lab, University Park, PA 16802; (CWW, IH) University of Helsinki, Department of Biological and Environmental Sciences, PL 65, Viikinkaari 1, 00014 Finland; (MJF) University of Helsinki, Institute of Biotechnology, PL 56, Viikinkaari 9, 00014 Finland; (DLC) University of Miami, Rosenstiel School of Marine and Atmospheric Sciences, 4600 Rickenbacker Causeway, Miami, FL 33149

A *de novo* assembly of *454* pyrosequencing data was used to deduce much of the transcriptome from the Glanville fritillary butterfly (*Melitaea cinxia*, Lepidoptera: Nymphalidae). This butterfly is a prominent species in population biology but had no previous genomic data. Sequencing runs on a Roche Genome Sequencer 20

using two normalized cDNA collections from two genetically diverse pools consisting of larvae plus pupae and adults, respectively, yielded 608,053 ESTs with a mean length of 110 nucleotides. Files containing these EST sequences can be obtained from the NCBI Short Read Archive, accession SRA000207.

These ESTs were assembled with Lasergene SEQMAN Pro 7.1 into 48,354 contigs (sets of overlapping DNA segments) and 59,943 singletons. Alignment (NCBI BLASTX) comparisons were performed against Sanger sequences we generated from an *M. cinxia* plasmid cDNA library, three expanded sub-volumes of the Uniprot 9.2 annotated protein database, and insect specific sequences including the *in silico* predicted *Bombyx mori* protein set, *Drosophila melanogaster* unigenes, *Heliconiu erato* clustered ESTs, and Butterflybase v2.91. All sequences, BLAST results, and annotations were stored in a normalized MySQL database. These comparisons confirmed the accuracy of the sequencing and assembly, and indicated the presence of about 9,000 unique genes, along with more than 6,000 additional genes from unannotated contigs confirmed with microarrays. Average depth of coverage was 6.5 fold for the longest 4,800 contigs, which were 348 to 2,849 bp in length. This coverage is sufficient for detecting large numbers of SNPs. BLAST analyses revealed whole-insect derived RNA to contain transcripts from non-metazoan taxa, with most of these sequences annotating to genes from microsporidia, a common intracellular parasite.

Oligonucleotide probes (60-mer, N = 207,149) were designed from the assembled sequences using Agilent's eArray web tool. These probes included six different oligos per annotated contig or singleton and an oligo for each complimentary strand of unannotated sequences, which allowed inference of strand orientation for unannotated sequences. All probes plus Agilent controls were printed on microarrays featuring 244 K spots and hybridized with Cy3-labeled cRNA derived from an even mixture of the two original RNA pools used in making the cDNA that used in *454* sequencing. Analysis of the microrray spot intensities found highly repeatable hybridization intensity, provided further validation of sequence quality and assembly, and enabled us to determine the best-performing probes for use in designing smaller, 44 K feature arrays. These smaller arrays were hybridized with labeled aRNA derived from the transcript of individual butterflies revealing excellent repeatability (r^2 = 0.98) and biological differences among individuals including microsporidial infection.

When performed to provide sufficient coverage depth, *454* pyrosequencing, allows *de novo* transcriptome assembly and leads to fast, cost effective, and reliable development of functional genomic tools for non-model species. With the advent of pyrosequencing technology, obtaining EST sequences from the transcriptomes of any arthropod is now within reach of efforts by an individual laboratory and funding levels available from NSF and NRI. This technology advance narrows the gap between approaches based on model organisms with rich genetic resources versus species that are most tractable for ecological and evolutionary studies.

65

Comparative genomic analysis of the Odorant Binding Protein multigene family across *Arthropoda*

*Vieira, Filipe G.; Sanchez-Gracia, Alejandro; Rozas, Julio
(FGV, JR) Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona 08028, Spain; (ASG) Department of Physiology and Molecular Biodiversity, Institute of Molecular Biology of Barcelona, CSIC, Barcelona 08034, Spain

Chemoreception is a widespread mechanism involved in critical biological processes, being essential for the detection of sources of food, egg-laying substrates, mates and predators, and allowing for communication and social coordination.
The insect peripheral olfactory-system comprises three major multigene families, the olfactory receptor (OR), the gustatory receptor (GR) and the odorant-binding protein (OBP) families. The analysis of the OBP and CSP (chemosensory protein, an OBP related family) families in twelve *Drosophila* species allowed the identification of genes encoding putative members of these families. Particularly, we have identified 54 ortholog groups (43 of them present in all twelve species) in the OBP gene family. The analysis shows that duplicate members mainly arise from tandem gene duplication and progressively diverge in DNA and amino acid sequence and that the pseudogenization events are predominant in the external branches of the phylogenetic tree. We observed that the OBP arrangement in clusters is maintained across the *Drosophila* species, pointing to some functional significance of this organization. The phylogenetic relationship among orthologous and paralogous OBP members noticeably supports the birth-and-death model for the evolution of this multigene family, with an estimated birth-and-death rate of 0.002-0.008 per gene and per million years.

Recently, we have extended our analysis to nine additional *Arthropoda* species (seven *Hexapoda*, one *Crustacea* and one *Chelicerata*), allowing the identification of a large number of putative members of the OBP and CSP multigene families.

66
Full-Arthropods: A Full-length cDNA Database of Arthropods
*Watanabe, J.; Hiranuka, K.; Kibukawa, E.: Katayama, T.; Kawashima, S.; Wakaguri, H.; Suzuki, Y.; Sugano, S.; Sugimoto, C.; Toyoda, A.; Vineet, S.; Taylor, T.; Usui, M.; Nogami, S.; Maeda, R.
(WJ) The University of Tokyo, Institute of Medical Science, Department of Parasitology, Minatoku, Tokyo 108-8639, JAPAN; (KT, KS) The University of Tokyo, Institute of Medical Science, Human Genome Center, Minatoku, Tokyo 108-8639, JAPAN; (HK); (KE); (WH, SY, SS) The University of Tokyo, Graduate School of Frontier Sciences, Department of Medical Genomics, Kashiwa, Chiba 277-8561, JAPAN; (SC) Hokkaido University, Center for Zoonosis Control, Sapporo, Hokkaido JAPAN; (TA, VS, TT) RIKEN Genomic Sciences Center, Yokohama 230-0045, JAPAN; (UM, MR)\ Obihiro University of Veterinary Medicine and Agriculture, Obihiro, 080-8555, JAPAN; (NS) Nihon University, College of Bioresource Sciences, Department of Veterinary Medicine, Fujisawa 252-8510, JAPAN

Full-Arthropods is a database of full-length cDNAs from various arthropods of medical importance. We have produced full-length cDNA libraries from Anopheles mosquito, Tsetse flies, Tsutsugamushi mite and dust mite using the oligo-capping method that we have originated. It essentially replaces the cap structure at the 5'-end of the intact eukaryote mRNAs with a synthetic RNA primer and preferentially clones the full-length cDNAs that represent the entire structure of the expressed genes. They also specify the exact transcription start sites and thus enable to identify the upstream promoter regions. At this moment Full-Arthropods contains 5'-end and 3'-end-one-pass sequences of the oligo-capped cDNA clones from *Anopheles gambiae* mapped onto the genome sequences that had been downloaded from the public database. The genome browser allows annotation of genes on the basis of full-length cDNA sequences. We are now in the process of incorporating larva sequences. Construction of databases of other arthropods including Tsetse fly and dust mite is underway.

67
Genetic Mapping of Avirulence Genes in the Hessian Fly
*Zhao, C; Aggarwal, R; Benatti, T; Chen, M-S; Schemerhorn, BJ; Harris, MO; Stuart, J
(ZC, AR, BT, SJ) Purdue University, Dept. of Entomology, West Lafayette, IN 47907; (CMS) Kansas State University, USDA-ARS and Dept. of Entomology, Manhattan, KS 66506; (SBJ) Purdue University, USDA-ARS and Dept. of Entomology, West Lafayette, IN 47907; (HMO) North Dakota State University, Dept. of Entomology, Fargo, ND 58105

Since the discovery of the gene-for-gene relationship by H. H. Flor in the middle of previous century, knowledge regarding the interaction between plant resistance (*R*) genes and their corresponding avirulence (*Avr*) genes in plant pathogenic bacteria and fungi has grown tremendously. Similar genetic interactions are known to exist between certain plants and highly adapted phytophagous insect species. Nevertheless, *Avr* genes in insect gene-for-gene interactions have yet to be identified. Advances in genomic technologies promise to eliminate this problem.

The Hessian fly (*Mayetiola destructor*) is perhaps the best example of an insect with a gene-for-gene interaction with its host, wheat (*Triticum* ssp.). Toward improving our understanding of the wheat-Hessian fly interaction, we have been pursuing a map-based strategy for cloning Hessian fly *Avr* genes that correspond to three *R* genes in wheat (*H6*, *H9*, and *H13*). These *Avr* genes are named *vH6*, *vH9*, and *vH13* in accordance with their corresponding *R* gene in wheat.

Using an FPC-based physical map and molecular genetic markers (microsatellites, BAC-end sequences, and AFLP markers) progress in mapping all three *Avr* genes has been made. *vH6* has been positioned between two adjacent contigs near the centromere on the long arm of chromosome X2. Using a mapping population consisting of 152 individuals, we found that only 3 recombination events separate these *vH6*-flanking contigs. *vH9* has been located within a single BAC contig near the telomere of the short arm of chromosome X1. Using a similar mapping population, we found that only 3 recombination events exist between the nearest genetic marker and *vH9*. *vH13* has been positioned within the sequence of a single BAC clone present near the telomere of the short arm of chromosome X2. Genetic mapping using a recombinant inbred line is refining the position of *vH13* within this sequence.

68
Evolution of the *morgue* Cell Death Gene in the Phylum *Arthropoda*
Ying Zhou*; John R. Nambu
(YZ)University of Massachusetts Amherst, Neuroscience and Behavior Program, Amherst, MA 01003; (JRN) University of Massachusetts Amherst, Biology Department, Amherst, MA 01003

Morgue is a novel gene cloned in *Drosophila melanogaster*, which can enhance the levels of programmed cell death (PCD) induced by RHG proteins (*reaper*, *hid*, or *grim*). *Morgue* encodes a protein including a zinc finger motif, an Fox domain and a ubiquitin E2 conjugase domain, in which the active Cysteine is replaced by a Glycine. The presence of an F box and a conjugase domain implies that Morgue acts in protein ubiquitination and the ability of Morgue to associate with the DIAPI suggests that Morgue may regulate PCD via ubiquitination and turnover of cell death regulators. To gain insight into the evolution of this unique protein and its roles in ubiquitination and PCD, we have initiated studies to identify and characterize related sequences from other species. Using bioinformatics and molecular approaches, we have found that Morgue distribution appears to be restricted to Arthropods. Thus far we have identified Morgue homologues in all the *Drosophila* species examined, as well as several other representative insects including the honeybee (*Apis mellifera*), silkworm (*Bombyx morii*), and red flour beetle (*Tribolium castaneum*). In addition, we have identified Morgue homologs in two crustacean species, the American lobster (*Homarus americanus*) and the water flea (*Daphnia pulex*). Each of these Morgue homologues contains a related zinc finger, F box, and variant conjugase domain in a similar organization. Furthermore, the Glycine substitution in the conjugase domain is conserved, suggesting specific functions for this residue. In several species the three domains are encoded by separate exons. Significantly, we have also identified a putative Morgue homologue from the wandering spider (*Cupiennius salei*), which only consists of a conjugase domain with the conserved Glycine substitution. Spider Morgue does not appear to contain a zinc finger or an F box. We are currently working to confirm this finding and identify Morgue-related zinc finger and F box sequences in the spider. Together, our data suggests that the *morgue* gene arose via exon shuffling events during divergence of the phylum *Athropoda*.

69
Analysis and Functional Annotation of Expressed Sequence Tags from the Aquatic Midge (*Chironomus tentans*)
* Zhu, Kun Yan; Li, Xiuwei; Zhang, Xin; Starkey, Sharon R.
Kansas State University, Department of Entomology and Ecological Genomics Institute, Manhattan, KS 66506 (Contact: *kzhu@ksu.edu*)

Environmental stressors, including pesticides, heavy metals, nutrients, and oxygen depletion, are detrimental to organisms and have now become a growing global concern. This research is to develop an expressed sequence tag (EST) database from the aquatic midge (*Chironomus tentans*). From 10,368 random cDNA clones sequenced, we retained 10,367 high-quality ESTs with an average length of 896 bp after preprocessing. These ESTs were assembled into 2,912 unique sequences including 2,095 singletons and 817 contigs. Blast search resulted in 2,152 putative transcripts that were functionally annotated and sorted into 10 molecular function categories: binding (37.6%), catalytic (37.5), transporter (12.1), structural molecule (6.7), molecular transducer (1.3), antioxidant (1.1), enzyme regulator (1.0), translation regulator (1.0), transcription regulator (1.0), and motor (0.5) activities. Eighty nine transcripts likely encode different types of hemoglobin involved in oxygen storage and/or transport, whereas at least 28 encode various cytochrome P450 monooxygenases involved in xenobiotic metabolism. These ESTs are expected to help us develop a DNA microarray for studying the mechanism of stressor’s effects, and cellular and molecular responses of the midge to environmental stressors.

70
BeeSpace: An Interactive Environment for Analyzing Nature and Nurture in Societal Roles