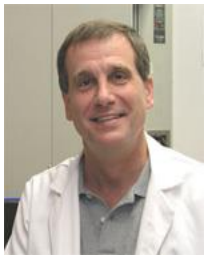
# Arthropod Genomics Symposium

# Platform/Oral Presentations

## (In order of speaker)

## Keynote:



***Insect genetic technologies and their applications to physiological genetics and functional genomics***

**David A. O'Brochta**
**University of Maryland College Park**

Technological advances in DNA sequencing technologies and bioinformatics are leading to a deep understanding of the biology, genetics, ecology and evolution of insects. Although there have been great advances in insect genomics over the past 20 years that continue today at increasingly rapid rates, we have just scratched the surface of most insect systems and to go deeper researchers will need genetic technologies that will allow them to manipulate the genomes of insects with greater ease, enabling them to more fully explore the functions of genes, genetic networks and physiological systems. Powerful genetic technologies that enable the functional genomic analysis of animal and plant systems are evolving rapidly and many have been or could be used successfully in insects. Today there are at least a dozen different types of genetic technologies used as functional genomic tools in insects and while some hold the promise of being transformative, can that promise be fulfilled and what will be required to make it so? New modes of genetic technology delivery are needed if current genetic technologies are going to more fully impact the study of insects. Closer synergistic relationships with human gene therapy researchers and bioengineers who face genetic technology delivery challenges that are similar in many ways to the challenges faced by many insect scientists are likely to be important for enabling insect scientists working on diverse insect groups to capitalize on what is no less than a genetic technology revolution.

## i5K/Emerging Genomes

***Origins of spider biodiversity: Genomic approaches at the nexus of ecology, behavior, and evolution***

**Rosemary Gillespie**
**University of California, Berkeley**

The processes through which biodiversity appears, whether through colonization from elsewhere or *in situ* diversification, and how it adapts and changes in the context of the environment, is fundamental to understanding not only the scale of biodiversity and how it will be modified as the planet changes, but also how it can be managed for environmental, medical, or other applied uses. Within this context, spiders are particularly important, being one of the most diverse orders of animals, and in which diversification has been tightly

linked to two important products, silk and venom: The importance of silk comes from its material properties as a light, stretchy, and strong material, designed for stopping and trapping prey; venom from its immense pharmaceutical repertoire, designed to chemically paralyze and preserve prey. However, while genomic research has advanced, with tremendous progress made in many other taxa, the development of genomic tools in spiders has lagged behind. Three major issues that have hindered progress in spider genomics are the large size of the genome, the very low GC content, and its repetitive nature. Here, we focus on major ongoing research in our group, in which we use genomic tools to understand processes involved in the establishment, adaptation, and species proliferation, in different lineages of spiders. Focusing on newly established populations and those undergoing adaptive radiation, we examine the role of both natural and sexual selection, in particular: (1) Natural selection associated with feeding and niche differentiation, examining the role of venom and silk in providing a basis for proliferation. (2) Natural selection associated with predation, in particular how color has evolved in allopatry and sympatry in response to different selective pressures. (3) Sexual recognition in the course of diversification, and the use of genomic tools in determining the nature of selection and the role of visual, behavioral, and chemical cues. At the same time, we evaluate how genomic tools can be used to understand (1) the potential role of admixture and fusion/ fission cycles in the successful establishment of populations, and subsequent diversification. And (2) potential co-option of similar genetic pathways in the allowing repeated evolution of similar form and allowing prediction of the role of adaptive shifts versus differentiation in allopatry in dictating the trajectory of diversification.

***Genome sequence of the bed bug, Cimex lectularius***

**Joshua Benoit, Biological Sciences, University of Cincinnati**

Bed bugs have re-established as a pest throughout much of the world during the last two decades. Various factors including increased pesticide resistance leading to poor control and their ability of long-term survival without food and water have been linked to this resurgence. In this study, we describe the genome of the common bed bug with a specific focus on obligate blood feeding in both sexes, mating through traumatic insemination, starvation resistance and factors responsible for pesticide resistance. Analyses of the C. lectularius genome and predicted protein coding genes provides the first complete representation of genes that are linked to traumatic insemination, reduced olfaction/chemosensory processes, host-symbiont interactions and unique sets of genes associated with multiple mechanisms of insecticide resistance. Data provided by this genome sequencing and analysis will establish a foundation for research devoted to insecticide resistance and uncovering distinctive mechanisms that are integral to bed bug biology.

***The i5K house spider genome reveals unexpected diversity and shifts in the expression of genes associated with the extreme toxicity of black widow venom.***

**Jessica Garb, Department of Biological Sciences, University of Massachusetts Lowell**

Venoms are protein-rich secretions with biomedical significance that have independently evolved in numerous animal lineages, but have only been studied at the whole-genome level in a few species. The genomes of the common house spider (*Parasteatoda tepidariorum)* and the Western black widow spider (*Latrodectus hesperus*) have been sequenced as part of the i5k pilot, and though these species are closely related, black widow venom exhibits far greater toxicity than house spider venom. Black widow venom is dominated by latrotoxins and latrodectins, two protein families that are virtually unknown outside of a few closely related species. Thus, we investigated the diversity, evolution and expression of latrotoxin and latrodectin genes with the recently annotated house spider genome using tissue-, sex- and stage-specific Illumina expression data. We discovered ≥47 latrotoxin genes in the house spider genome, many of which are tandem arrayed. House spider latrotoxins vary extensively in predicted structural domains and stage and tissue specific expression, suggesting novel functional roles. We uncovered far fewer latrodectin genes, but expression and phylogenetic evidence shows their recruitment for venom function from a

neuropeptide hormone family following gene duplication, inversion and coincident domain truncation. While latrodectins and other peptides are highly expressed in house spider venom glands, latrotoxins represent a far smaller proportion of house spider venom-gland expressed transcripts. Phylogenetic analyses show house spider latrotoxins are highly divergent from black widow latrotoxins, and this, along with the lower expression of latrotoxins in house spider venom, may explain the far greater potency of black widow venom. We will present plans to extend this work to the i5k black widow genome, including strategies to improve the current genome assembly.

***Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in* Manduca sexta**

**Haobo Jiang, Entomology and Plant Pathology, Oklahoma State University**

Serine protease (SP) and serine protease homolog (SPH) genes in insects encode a large family of proteins involved in digestion, development, immunity, and other processes. While 68 digestive SPs and their close homologs are reported in a companion paper (Kuwar et al., 2015), we have identified 125 other SPs/SPHs in Manduca sexta and studied their structure, evolution, and expression. Fifty-two of them contain cystine-stabilized structures for molecular recognition, including clip, LDLa, Sushi, Wonton, TSP, CUB, Frizzle, and SR domains. There are nineteen groups of genes evolved from relatively recent gene duplication and sequence divergence. Thirty-five SPs and seven SPHs contain 1, 2 or 5 clip domains. Multiple sequence alignment and molecular modeling of the 54 clip domains have revealed structural diversity of these regulatory modules. Sequence comparison with their homologs in Drosophila melanogaster, Anopheles gambiae and Tribolium castaneum allows us to classify them into five subfamilies: A are SPHs with 1 or 5 group-3 clip domains, B are SPs with 1 or 2 group-2 clip domains, C, D1 and D2 are SPs with a single clip domain in group-1a, 1b and 1c, respectively. We have classified into six categories the 125 expression profiles of SP-related proteins in fat body, brain, midgut, Malpighian tubule, testis, and ovary at different stages, suggesting that they participate in various physiological processes. Through RNA-Seq-based gene annotation and expression profiling, as well as intragenomic sequence comparisons, we have established a framework of information for future biochemical research of nondigestive SPs and SPHs in this model species.

***i5K@OrthoDB***

**Robert Waterhouse, Department of Genetic Medicine and Development, University of Geneva Medical School**

Orthology delineation is a cornerstone of comparative genomics, offering qualified hypotheses on gene function by identifying “equivalent” genes in different species, as well as highlighting shared and unique genes that offer clues to understanding species diversity. The almost 90 arthropod species included in the latest release of the OrthoDB hierarchical catalog of orthologs (OrthoDB v8, www.orthodb.org, Kriventseva *et al*, NAR, 2015) offers the most comprehensive orthology resource for arthropod comparative genomics, leading to a maturing understanding of the composition of the insect gene repertoire (Waterhouse, COIS, 2015). Mapping of the annotated gene sets from newly-sequenced and annotated genomes of *i*5K pilot project species to OrthoDB orthogroups allows researchers to make the most of their newly-sequenced arthropod genomes. With now more than 100 sequenced and annotated arthropod genomes, thanks to the progress of the *i*5K pilot project, comparative genomics approaches are becoming ever more powerful tools to improve and extend genome annotation and interpretation for newly-sequenced species.

***The Asian longhorned beetle, Anoplophora glabripennis (Cerambycidae) genome yields new insights into the genomic basis of phytophagy and the evolution of beetle megadiversity***

**Duane D. McKenna**
**University of Memphis**

The order Coleoptera (beetles) contains at least 389,000 described living species, more than any other order of animals. The causes of this apparent "inordinate fondness" (Hutchinson 1959) are controversial, but interactions with plants are widely considered to have played an important role. I will begin this talk with a brief discussion of the phylogeny and evolution of beetles. I will then discuss the Asian longhorned beetle (ALB) genome, including new insights into the genomic basis of phytophagy (feeding on plants) and polyphagy (feeding on multiple hosts), gained through studies recently completed by the International Asian Longhorned Beetle Genome Consortium. The ALB genome encodes an arsenal of glycoside hydrolases, plant cell wall degrading enzymes, digestive proteinases, and detoxification genes – some obtained from microbes via lateral gene transfer – that facilitate phytophagy and broad polyphagy. Its metabolic capacity is further expanded through affiliations with gut microbes. ALB can thus degrade the main polysaccharide networks in plant cell walls, and has the metabolic plasticity needed to feed on tree species with different chemistries, including novel hosts in areas where it has been introduced. Large expansions of chemosensory genes may play a role in the reception of pheromones and plant semiochemicals used for host plant and mate finding.

***Progress and insights from the milkweed bug genome project: towards a reference species for the Hemiptera***

**Kristen Panfilio, Institute for Developmental Biology, University of Cologne**

The Hemiptera are the most species-rich order of hemimetabolous insect, characterized by piercing and sucking mouthparts that have allowed them to exploit diverse food sources and ecological niches across the globe. A recent i5k pilot project has generated draft genomes for a number of hemipterans, providing the first opportunity to investigate the underlying genomic features that correlate with lineage specific adaptations and that are shared across the order, making a bug a bug. Here I will present the initial findings from the genome project of the milkweed bug, *Oncopeltus fasciatus*. Having served as a research model for physiology and development since the 1960s, *O. fasciatus* provides arguably the best reference species for the Hemiptera. So, how representative is it? To date, the draft genome has led to new gene discovery and functional investigations as well as assessments of its gene repertoire and trends in gene structure. These results will be compared for *O. fasciatus* as a specialist seed feeder in relation to other hemipterans, including the sap-feeding pea aphid with its sexual-asexual life cycle, and the bed bug, an obligate blood feeder.

***From sequence to locus: Using GBS and whole genome sequencing to identify the genetic basis of white pupae in SIT colony Medflies***

**Sheina Sim, Department of Plant and Environmental Protection Sciences, University of Hawaii, Manoa**

The Mediterranean fruit fly *Ceratitis capitata* (Wiedemann), commonly known as the medfly, is a destructive agricultural pest and the object of expensive population eradication and suppression efforts within state and federal departments of agriculture. Area-wide integrated pest management programs control medfly populations through the release of sterile males which must be massively produced. Mass-rearing and release of sterile males is facilitated by two sex-linked traits *white pupae* (*wp*) and *temperature sensitive lethal* (*tsl*). Though these two sex-linked traits in what is known as a genetic sexing strain was

developed over 20 years ago, the genetic basis of *wp* is unknown. The purpose of this project was to identify SNP loci tightly linked to the causative mutation for *wp* in mass-reared sterile insect colonies. A high-quality reference genome from whole-genome sequences from individual flies and assembled using DISCOVAR was used along with GBS sequences from an F4 mapping population to identify SNPs linked to *wp*. The genomic region containing the SNPs tightly linked to *wp* was then investigated further to generate a list of candidate SNPs resulting in the *wp* mutation. This data has been used to develop a genetic assay for differentiating between recaptured sterile insect released males and wild males, will be useful for the improvement of existing sterile insect colonies, and can be used to identify an orthologous gene in other species to create novel sterile insect strains.

***The i5K pilot project: progress on sequencing a phylum***

**Monica Poelchau, USDA-ARS, National Agricultural Library**

In this talk, we report on the progress of sequencing, annotating and curating 28 Arthropod genomes by the i5k pilot project. Sequencing and automated annotation of the 28 genomes was completed in 2014, and manual curation of the resulting datasets is still underway. Three genomes – the Asian long-horned beetle (*Anoplophora glabripennis*), the bed bug (*Cimex lectularius*), and the milkweed bug (*Oncopeltus fasciatus*) have completed the manual curation phase to date, and write-up and publication of these results are ongoing. We review 1) biological insights gained from individual i5k pilot projects; and 2) process and progress of manual curation efforts. Finally, we provide suggestions for the i5k community on genome sequencing and analyses based on experiences gleaned from the pilot project.

# Vector Genomics

***Leveraging 'omics technologies to develop new initiatives for controlling vector-borne plant pathogens***

**Michelle Cilia**
**USDA Agricultural Research Service, Cornell University, and Boyce Thompson Institute for Plant Research**

A majority of plant pathogens and a large number of important animal pathogens are transmitted by insect vectors. Plant pathogens are divided between those that are carried on the cuticle linings of mouthparts and foreguts and those that circulate in their vectors. Insects in the order Hemiptera are among the most prolific vectors of plant pathogens. Our lab focuses on plant pathogens that are exclusively transmitted by hemipteran pests in a circulative manner. Phloem-retention facilitates circulative transmission by these insects, which spend prolonged periods of time feeding in the phloem tissue. This presentation will highlight our efforts to develop and apply advanced proteomics technologies to enable us to explore the dynamic pathogen-vector interface. Several examples of proteomics data will be discussed to illustrate the power of these technologies to further our basic understanding of the molecular pathways involved in circulative transmission in plants and aphids and the excellent agreement of our data with previously published studies on the biology of circulative transmission. Finally, examples from our data will also be presented to show how proteomics technologies can enable us to develop novel strategies that disrupt pathogen movement within and between hosts.

***Genome improvement using third-gen sequencing***

**Scott Emrich, Computer Science and Engineering, University of Notre Dame**

Many arthropods have difficult to assemble genomes, especially biomedically relevant vectors. As an objective of multiple related consortia, An. gambiae is unique in that it has been sequenced twice with Sanger methods, once with Illumina and PacBio WGS, and was a key member of a large phylogenetic cluster. We hypothesized that low coverage but longer "third-gen" read data would successfully improve prior assemblies at much lower cost. Here, we first assess advantages of each individual sequencing method using all available data, and report hybrid techniques that we are developing at Notre Dame for combining independent assemblies (i.e., metaassembly) and for scaffolding assemblies using limited third-gen sequencing data. We will show that our new scaffolding tool, pbSandwich, outperforms alternatives with low coverage PacBio data, and validate these overall results with a complete annotation, an optical map, and whole-genome alignments to six very closely related genomes. Finally, we will conclude with preliminary results of new hybrid genome assembly framework to another bio medically important mosquito vector, An. funestus.

***Towards cheap, decent reference genome assemblies: biological and methodological insights from Anopheles***

**Daniel E. Neafsey**
**Broad Institute of MIT and Harvard**

The recent release of *de novo* genome assemblies for many arthropod species previously lacking reference genomes has demonstrated the value of such resources for understanding diverse aspects of arthropod biology, evolution, and control. The methods used to generate many of these assemblies are resource intensive and often rely on specialized expertise, however, making them hard to scale to the innumerable species without reference genome assemblies. We review lessons learned in the generation of *de novo* reference genome assemblies for 16 *Anopheles* mosquito species, and we present an assessment of alternate methodologies aimed at generating *de novo* genome assemblies without specialized sequencing library protocols or informatics expertise, including DISCOVAR *de novo.* We find that a DISCOVAR assembly of *Anopheles arabiensis*, constructed inexpensively using reads from a single library type, is superior in quality to a transcriptome assembly of the same species in terms of gene recovery, and nearly as complete as a reference assembly built using a more complicated methodology and incorporating reads from three libraries. We discuss applications of such assemblies and opportunities to further democratize *de novo* genome assembly.

***RNA-seq analysis reveals mechanisms associated with tsetse fly-symbiont dynamics during larval development***

**Joshua Benoit, Biological Sciences, University of Cincinnati**

Tsetse flies harbor three distinct symbionts that are transmitted from the mother to her offspring. These symbionts are critical to tsetse fly physiology, in specific have been documented to play a role in immune system development and adult reproduction. In this study, we utilized RNA-seq analyses followed by functional examination to determine the effects of symbiont removal on larval development and progeny health. These analyses revealed that there are over 2500 genes with differential expression after symbiont removal. In specific, there is an enrichment for genes associated with B vitamin metabolism, larval development and chitin interactions. The reduction of gene associated with B vitamin metabolism is likely due to reduced B vitamins concentraitons that are normally produced by the obligate symbiont, Wigglesworthia, and a reduction in chitin-associated genes likely accounts for impaired peritrophic matrix

development previ/ously documented. In addition to these factors, we noted 10-fold increase in expression for two odorant-binding proteins (obps), obp6 and obp11. Knockdown of obp6 can be accomplished by injection of siRNAs into the mother during lactation. Offspring with reduced levels of obp6 yield adults with a phenotype of reduced or dysfunctional crystal cells, which act in the initiation of the melanization cascade via the release of prophenoloxidase. These results provide mechanisms underlying immune development and symbiont dynamics in tsetse flies and implicate odorant-binding proteins in the process.

### *The* Anopheles gambiae *1000 genomes (Ag1000G) project*

**Martin James Donnelly on behalf of the *Anopheles gambiae* 1000 genomes (Ag1000G) consortium.**
**Liverpool School of Tropical Medicine and the Wellcome Trust Sanger Institute, UK**

*Anopheles gambiae s.l.* is the primary vector of *Plasmodium falciparum* malaria in sub-Saharan African and is largely responsible for the high rates of malaria morbidity and mortality that the continent experiences. The Ag1000G project is a consortial venture led by the Wellcome Trust Sanger Institute that uses deep re-sequencing of population samples to characterise genome variation in this vector complex (http://www.malariagen.net/ag1000g). Our aim is to use genome sequencing to both illuminate the evolutionary biology of these fascinating insects and to direct vector control targeting, monitoring and evaluation. To date we have sequenced in excess of 2000 individual *Anopheles gambiae* and *Anopheles coluzzii* malaria vectors drawn from across sub-Saharan Africa. All data from this project are released to the community in advance of publication and access is facilitated by bespoke genome browsing tools (http://www.malariagen.net/apps/ag1000g/).
Mosquitoes collected in 13 countries in Africa, that span the breadth of transmission settings, have been contributed to the project, and data from 9 collections from eight countries (phase 1) are now available. The phase 1 data set comprises variant calls from 765 individuals sequenced to a nominal median sequence depth of 30. In this presentation I will describe levels of diversity and genome accessibility and illustrate how we can use the data to track the demographic changes associated with successful control and the continent wide emergence and spread of insecticide resistance.

### *High-throughput cis-regulatory element discovery in the vector* mosquito Aedes aegypti

**Molly Duman-Scheel, Medical and Molecular Genetics and Eck Institute for Global Health, Indiana University School of Medicine and the University of Notre Dame**

Despite substantial progress in mosquito genomic and genetic research, few cis-regulatory elements (CREs), DNA sequences that control gene expression, have been identified in mosquitoes or other non-model insects. This deficiency—a significant gap in the basic knowledge of mosquito genetics—has resulted in a lack of drivers to manipulate gene expression in selected tissues at specific times. Such tools, which have revolutionized research in genetic model organisms, would facilitate genetic studies and benefit all avenues of mosquito research. FAIRE-seq, formaldehyde-assisted isolation of regulatory elements paired with DNA sequencing, is emerging as a powerful new high-throughput tool for global CRE discovery. During the FAIRE process, chromatin is cross-linked with formaldehyde, sheared, and then phenol-chloroform extracted, allowing for preferential recovery of open chromatin DNA fragments that are not bound by nucleosomes, an evolutionarily conserved indicator of regulatory activity. FAIRE has many advantages over alternative methods, one being that the recent pairing of this technique with next-generation sequencing, FAIRE-seq, permits straightforward and genome-wide high-throughput detection of CREs. Despite the power of this approach, FAIRE-seq has not been applied to the study of non-model organisms, and most FAIRE-seq studies to date have been performed in human cell lines. We recently performed a FAIRE-seq study in the dengue and yellow fever vector mosquito *Aedes aegypti*. Following

optimization of the FAIRE DNA isolation procedure in *A. aegypti* embryos, next-generation sequencing (HiSeq 50 bp paired-end) was used to generate ~150 million reads for each of three biological replicate embryonic FAIRE DNA preparations. Sequences were aligned to the reference genome *(Aedes-aegypti* Liverpool_SCAFFOLDS_AaegL1.3) using BWA (version 0.5.9-r16). SAM files of replicates were pooled for further analysis, and FAIRE peaks were called with MACS2 (version 2.1.0.20140616). FAIRE sequences mapped to 5' flanking regions/promoters and first introns of genes throughout the genome. Known transcriptional enhancer sites were found to be enriched in the FAIRE DNA sequences, and CREs that had been verified in our previous functional studies or which were computationally predicted in an ongoing parallel study were detected in the FAIRE-seq data set. We are working to make these data accessible to all researchers in the Vectorbase Genome browser. We have also initiated a high-throughput screen in transgenic insects that will examine the ability of hundreds of these elements to promote gene expression *in vivo*. In addition to validating the FAIRE data set, the screen has been designed to select for elements that have a high potential to drive gene expression in tissues of vector importance in multiple vector insect species. This research will generate a toolkit of gene drivers for mosquito research, promote use of FAIRE-seq in additional insect species, and encourage the study of mosquito gene regulatory networks.

**Epigenomics (partial session)**

***Is methylation one of the drivers of virulence in Diuraphis noxia (Kurd.) Hemiptera: Aphididae?***

**Anna-Maria Botha, Professor, Genetics Department, Stellenbosch University**

Diuraphis noxia (Kurdjumov, Hemiptera: Aphididae), a specialist phloem feeder, is an economically important aphid pest afflicting wheat and barley yield in dry-land production regions of the USA, Argentina and South Africa. Populations sharing similar ecoagricultural regions and expressing different levels of virulence towards their hosts are called biotypes, and the number of D. noxia biotypes reported continues to increase, posing multiple threats to global food security. With the availability of the draft genome of D. noxia and confounding evidence of genomic plasticity, we set out to determine the extent of methylation in the genome of Diuraphis noxia in order to determine if methylation contributes to changes in virulence. To this end, the global levels of methylation as well as the methylation profiles of the different biotypes were investigated, the former done by measuring fluorescent adaptor levels when aphid DNA was restricted with isoschizomers HpaII and MspI. The latter involved the use of Methylation-Sensitive Amplification Polymorphism, and also provided insight into local regions of methylation in the genome. The global methylation results suggest an inversely proportional relationship between virulence levels and methylation, hypomethylation being associated with increased virulence. Methylation profiles of the biotypes, whilst similar, did show some clear differences indicating that differential methylation of certain genes could indeed contribute to differences in virulence. This study, being the first of its kind for Diuraphis noxia, has provided the groundwork for future research into methylation of this insect, and adds to a growing body of knowledge on the Russian Wheat Aphid.

# Population Genomics

***Climatic selection, population history, and developmental constraints shape genome-wide divergence in speciating* Rhagoletis *flies***

**Greg Ragland**
**Kansas State University**

Theory predicts that ecological speciation, mediated by strong selection, will leave a clear signature of genetic divergence during the earliest stages of the speciation process. However, the nature of such a genetic signature may be complicated by past patterns of selection on traits that are genetically correlated, either by pleiotropy or linkage disequilibrium (LD). For example, geographic populations of *Rhagoletis pomonella* flies experience a latitudinal gradient of selection on several aspects of phenology, mediated through dormancy. These same aspects of phenology are under strong divergent selection among host races (nascent species) that infest different host plants with different seasonal fruiting times. Genome-wide divergence among these host races, as revealed through RADseq, correlates well with two primary, diapause-related traits known to experience strong divergent selection. However, the directionality of divergence at the genetic level is the opposite of that predicted at the phenotypic level. I will discuss the historical patterns of natural selection, evidence for LD-driven genetic correlations, and emergent developmental constraints that we believe explain these counterintuitive patterns.

***Evolution of GOUNDRY, a cryptic subgroup of Anopheles gambiae s.l., and its impact on susceptibility to Plasmodium infection***

**Jacob Crawford, Integrative Biology, University of California, Berkeley**

The recent discovery of a previously unknown genetic subgroup of Anopheles gambiae sensu lato, named GOUNDRY, that is highly susceptible to Plasmodium infection in the laboratory and does not rest indoors as adults, underscores our incomplete understanding of complex speciation dynamics in Anopheles and its potential to impede malaria control efforts. Initial description of GOUNDRY suggested it differed from other known closely related subgroups in surprising and sometimes contradictory ways, raising a number of questions about its age, population size, and relationship to known subgroups. To address these questions and understand this new subgroup further, we sequenced and analyzed a panel of complete Anopheles genomes including 12 GOUNDRY individuals. We show that GOUNDRY is most closely related to Anopheles coluzzii, and the timing of cladogenesis, approximately 110 Kya, substantially predates the advent of agriculture. We find a large region of the X chromosome that has swept to fixation GOUNDRY within the last 100 years, which may be an inversion that serves as a partial barrier to contemporary gene flow. Lastly, we show that GOUNDRY has a history of inbreeding that is significantly associated with susceptibility to Plasmodium infection in the laboratory. Our results illuminate the evolution of one of an unknown number of cryptic, ecologically specialized subgroups of Anopheles and provide a potent example of how speciation dynamics could complicate malaria eradiation and control efforts.

***The genetic basis of female mate preference and species isolation***

**Amanda Moehring**
**Western University, Ontario**

The formation of new species gives rise to the wide array of biodiversity we see today, yet relatively little is known about the underlying genetic basis of species isolation. In particular, no genes have been identified for behavioral isolation, which is thought to be the first species barrier to arise. A novel genetic mapping approach was used to identify three genes that affect behavioral isolation. All three genes affect species-specific female mate preference, and appear to act via neural patterning. The contributing male trait and subsequent female neural processing has been evaluated for one of these genes, and results from these studies will be presented.

***Molecular Population Genomics of Chromosomal Inversions in*** **Drosophila pseudoobscura**

**Stephen Schaeffer, Biology, The Pennsylvania State University**

The third chromosome of *Drosophila pseudoobscura* is polymorphic for over 30 different gene arrangements that were generated through a series of overlapping paracentric inversions. Geographic and altitudinal gradients in gene arrangement frequency provide indirect evidence that selection has acted to increase and maintain the inversion mutations in populations, but the exact nature of what might be selected within the arrangements is unclear. We generated complete genome sequences for 54 *D. pseudoobscura* third chromosomes to test hypotheses about the genetic mechanisms that led to the origin and maintenance of this gene arrangement polymorphism. This sample included 8 to 15 copies of each of six chromosomal arrangements. The hypotheses tested fall into four broad classes: (1) direct effects of the inversion mutation where the breakpoints created selectable variation; (2) indirect effects of inversions as recombination suppressors where the presence of rearrangements holds particular allelic combinations together; (3) selective sweeps associated with a positively selected allele; or (4) genetic drift. We used population genetic analyses on the site frequency spectrum as well as estimates of extended homozygosity to determine genes that depart from selectively neutral models. In addition, we collected transcriptomic data from each of the six different arrangements to test for differential gene expression among inversion types. The data show that multiple genes within and immediately adjacent to the inverted regions depart from neutral expectations or show evidence for differential gene expression among arrangements. These data support the hypothesis that inversions are selected because of the role that they play in suppressing recombination and thus, maintaining linkage of alleles at multiple loci in different arrangements. The inferred functions of the selected genes include detoxification genes such as P450s, sensory genes such as odorant receptors, and digestive enzymes such as trypsins. These functions suggest that the different gene arrangements help *D. pseudoobscura* deal with environmental differences in toxins and food sources.

# Epigenomics

***The condition-responsive ontogeny of* Onthophagus *beetles: transcriptome dynamics, developmental mechanisms, and the integration of sex, nutrition, and tissue type in evo devo***

**Armin P. Moczek**
**Indiana University**

Individuals of most sexually reproducing organisms can be conceptualized as mosaics of body regions, organs, and tissue types whose development is more or less influenced by internal or external conditions, such as sex or nutrient availability. Yet we know little about how tissue- specific responsiveness to conditions is achieved in development, integrated systemically during ontogeny, and modified in evolution.

Here, I investigate the condition-responsive development of *Onthophagus* beetles, an emerging model system in evo devo. In these organisms, males and females differ substantially in tissue-specific growth responses to nutritional variation, generating complex sexual- as well as within-male dimorphisms. These dimorphisms have, in turn, diversified across species with respect to degree, direction, and body region or tissue type under consideration.

In this presentation I highlight our most recent findings with respect to the transcriptional dynamics associated with conditional development, present examples of developmental mechanisms that facilitate nutrition- and sex-specific differentiation and their integration, and close by presenting case studies on how the developmental machinery underlying condition-responsive development is diversifying across populations and species.

***DNA methylation patterns of arthropod genomes***

**Frank Lyko**
**German Cancer Research Center, Heidelberg, Germany**

Eukaryotic species use (cytosine-5) DNA methylation to facilitate phenotypic adaptation to their environments, which can include both the modulation of developmental and adaptive gene expression programs. Variations in the complement of cytosine methyltransferase enzymes have been interpreted to reflect multiple versions of a toolkit for phenotypic adaptation. During evolution, specific parts of this toolkit could have been contracted or expanded to facilitate specific requirements for genome regulation. We are using whole-genome bisulfite sequencing to investigate this hypothesis and to establish genome methylation maps of various model systems at single-base resolution. Our results define three groups of arthropod methylomes with fundamental differences that will be discussed in detail: The first group is defined by Drosophila and is characterized by the complete absence of recognizable DNA methylation patterns. The second group is defined by the honeybee and is characterized by the highly selective methylation of specific CpG residues. The third group is characterized by pervasive genome-wide methylation and we will present the marbled crayfish (*Procambarus virginalis*) as a novel model system to understand the relevance of DNA methylation for phenotypic variation.

***Sex chromosome dosage compensation in Lepidoptera: insights from nymphalid butterflies, coddling moth, and demasculinized silkworms***

**James Walters, Ecology & Evolutionary Biology, University of Kasas**

Dosage compensation is the equalization of gene expression levels in response to differences in gene dose or copy number. It is classically considered to play a critical role in the evolution of heteromorphic sex chromosomes. As the X and Y diverge through degradation and gene loss on the Y (or the W in female-heterogametic ZW taxa), it is expected that dosage compensation will evolve to correct for sex-specific differences in gene dose. Although this is typically observed in male-heterogametic (XY) species, recent genome-wide expression studies in other taxa have revealed striking exceptions, especially in ZW taxa such as birds and snakes. In these taxa, the single Z of females is under expressed relative to males. These results fuel speculation that incomplete dosage compensation may be a defining characteristic of female-heterogamety. However, Lepidoptera (moths and butterflies) are also female-heterogametic, and evidence is accumulating that at least some species show balanced expression between sexes on the Z chromosome, contradicting the emerging consensus that ZW taxa lack complete dosage compensation.

Here we report on patterns of sex chromosome dosage compensation inferred from RNA-seq in two major lepidopteran lineages not yet surveyed: butterflies and Tortricid moths. In butterflies we focus on *Heliconius melpomene*, using a fully sequenced genome. Analyses in the Tortricid moth, *Cydia pomonella* (coddling moth), are based on *de novo* transcriptome assemblies. Results from both species show that average Z chromosome expression is significantly lower than autosomes in both sexes, similar to previous reports in bombycoid moths, and suggesting a novel mechanism of dosage compensation exists in the Lepidoptera. Notably, *C. pomonella* carries a neo-Z chromosome arising from the fusion of the ancestral Z with an autosome; this neo-Z also appears to be down-regulated in males. In *Heliconius*, but not *C. pomonella*, we detect a significantly greater global Z expression in males over females, indicating dosage compensation is imperfect in this species. However, the magnitude of this dosage effect is much less than the magnitude of reduced Z expression in both sexes. Patterns of sex-biased expression show an excess of male-biased genes on the Z chromosome, consistent with predictions from sexual antagonism theory.

Finally, we have extended the analysis of RNA-seq data from Kiuchi et al (Nature, 2014) that assayed genome-wide expression in *Bombyx mori* after RNAi knockdown of the sex-determining pathway. Our efforts confirm that knockdown of the masculinizing protein increases expression specifically on the Z chromosome in males, as previously reported. Additionally, we have uncovered a similar, though weaker, pattern in females that was not previously reported. These results further support the epigenetic down-regulation of the Z chromosome in males, but also suggests a similar mechanism may be operating in females in a dose-dependent manner.

***Using genomic tools to unravel parasitic wasp genome evolution via acquisition of viral genomes***

**Anne Nathalie Volkoff**
**INRA (UMR 1333), Université de Montpellier, France**

Parasitic wasps are among the most abundant and diverse on earth, and many develop inside another insect possessing an efficient immune system. As a result, they display innovative strategies to ensure their parasitic success. Amazingly, thousands of species rely on virus-like particles to deliver “virulence” molecules (proteins or DNA molecules encoding virulence genes) in the insect host. The particles are formed in a specialized tissue of the female genital tract (named the "calyx"), stored in the oviducts and then transferred in the insect host during oviposition. The transferred virulence molecules disturb major physiological functions of the insect host and make it suitable for the development of the parasitoid offspring.

We recently conducted extensive genomics analyses coupled with proteomics to unravel the origin of the virus-like particles produced in 2 related ichneumonid species from the *Campopleginae* subfamily, *Venturia canescens* and *Hyposoter didymator*. *V. canescens* carries virus-like particles (VcVLPs) devoid of DNA whereas *H. didymator* carries Ichnovirus particles (HdIV; PolyDNAvirus family) enclosing about 50 circular dsDNA molecules.

Our data clearly show that the production of particles used for the transfer of virulence molecules from the wasp to the insect host derived from the integration of a viral genome into the genome of the wasp ancestors. Surprisingly, our results showed that the viral ancestor differs between HdIV particles and VcVLPs indicating that at least two independent events of acquisitions of a virus occurred during the evolution of the *Campopleginae* wasps.

***Insect virus discovery using a viral metagenomics approach***

**Sijun Liu**
**Iowa State University**

Metagenomics, in conjunction with next generation sequencing (NGS) technology, has become a primary mechanism for the discovery of insect virus sequences. In the past several years, we have analyzed deep sequencing data derived from sequencing of the DNA, RNA and small RNA of major pests of agricultural importance and other invertebrate species for the presence of viral sequences. We discovered sequences derived from more than 50 viruses from NGS data including that of corn rootworms, stink bugs, aphids, leafhoppers, and plant hoppers. The vast majority of viruses identified are novel viruses, primarily with RNA genomes. The fact that insects harbor many previously unknown and diverse viral species has important implications for the use of insect cell lines and laboratory insect colonies for research. In addition, we discovered viral sequences integrated into insect genomes confirming recent reports to this effect. Taken together, we conclude that most insects harbor multiple viruses, many of which result in covert infection and are unclassified species, and that in some cases, viral sequences are integrated into the host genome.

***Psyllid-symbiont transcriptome: a two-organism negotiation***

**Cecilia Tamborindeguy, Department of Entomology, Texas A&M University**

Next-generation sequencing technology is allowing us to eavesdrop on the dialog between insects and their symbionts. Phloem feeding insects can associate with several microorganisms including obligate endosymbionts housed in special cells, bacteriocytes, or plant pathogenic bacteria. By using Illumina technology to sequence insect transcriptome it is possible to elucidate how insect and symbiont gene expression are integrated and respond to the physiological needs of the insect host. The potato psyllid, *Bactericera cockerelli*, is a major pest of solanaceous crops and vector of '*Candidatus* Liberibacter solanacearum', a plant pathogen. Taking advantage of the reduced genome size and high AT content of '*Candidatus* Carsonella ruddii', we obtained the genome sequence of this primary endosymbiont by transcriptome analyses of potato psyllids. Similarly, we investigated the interaction of psyllid-'*Ca.* Carsonella ruddii' and psyllid-'*Ca.* Liberibacter solanacearum' at the transcriptome level. This approach allows to understand how partner genomes interact and has the advantage of improving bacterial genome annotation.

***Nested genomes: a hologenomic apporach to honey bee health***

**Jay Evans, Bee Research Laboratory, USDA-ARS**

Genomic resources are available for honey bees and each of the primary biological associates found in honey bee colonies, from viruses to mites and beetles. A hologenomic approach attempts to answer questions by leveraging genomic insights from each member of a biological community. This approach is being used to identify parasites and pathogens involved in bee declines (1,2), to characterize the gut microbiota (3) of bees, and to determine how strain variation within bees (4) and their associated biome (5) can play a role in bee health. This presentation will present current insights into the genetics of bees and their associates as well as future directions enabled by better tools and resources.

# POSTER ABSTRACTS

# i5K/Emerging Genomes

**1 - i5K/EG**
**Characterization and regulation of expression of an antifungal peptide from hemolymph of an insect, *Manduca sexta***

Alsouhail, Qasim; Biochemistry and Molecular Biophysics, Kansas State University
Additional authors:
Hiromasa, Yasuaki; Kansas State University
Rahnamaeian, Mohammad; Justus-Liebig-University of Giessen,Germany
Vilcinskas, Andreas; Justus-Liebig-University of Giessen,Germany
Kanost, Michael; Kansas State University

Insects secrete antimicrobial peptides as part of the innate immune response. These peptides are often cationic, low molecular peptides with diverse structures. Most antimicrobial peptides from insects have antibacterial, but not antifungal activity. We have characterized a newly identified antifungal peptide from the hemolymph of the lepidopteran insect, *Manduca sexta* (tobacco hornworm). The antifungal peptide, named diapausin-1, was isolated by size exclusion chromatography from hemolymph plasma of larvae. Fractions containing activity against *Saccharomyces cerevisiae* were analyzed by SDS-PAGE and MALDI-TOF MS/MS and were found to contain a 45-residue peptide encoded by sequences identified in *M. sexta* transcriptome and genome data bases. Diapausin-1 is a member of the diapausin family of peptides, found in several insect species. The *M. sexta* genome contains 14 genes with high similarity to diapausin-1, each with 6 conserved Cys residues. A cDNA for diapausin-1 was cloned, and used for expression of a recombinant protein. Purified natural and recombinant diapausin-1 were active against *S. cerevisiae*, with $IC_{50}$ of 12μM, but they had no detectable activity against bacteria. Diapausin-1 mRNA level in the fat body strongly increased after larvae were injected with yeast or with *Micrococcus luteus*. Yeast treated with diapausin-1 had altered morphology, whereas fungal hyphae treated with diapausin-1 showed reduced and branched hyphal growth. Diapausin-1 mRNA levels increased in the fat body and midgut of naïve larvae in the wandering stage compared with the feeding stage fifth instar. Our results indicate that synthesis of diapausin-1 is part of the innate immune response to infection in *M. sexta*.

**2 - i5K/EG**
**Transriptomes and emerging genomes identify the gene repertoires underlying functional differentiation of spider silk glands**

Ayoub, Nadia; Biology, Washington and Lee
Additional authors:
Clarke, Thomas; 1. Department of Biology, Washington and Lee University
Garb, Jessica; 2. Department of Biological Sciences, University of Massachusetts, Lowell
Hayashi, Cheryl; 3. Department of Biology, University of California, Riverside

Spiders (Araneae) are exceptional among silk producing arthropods for the diversity of silk types and functions found within and among species. Functionally and mechanically distinct silk fibers are composed primarily of unique proteins synthesized in specialized abdominal glands. Araneoid spiders (a mega-diverse clade that include orb-web, sheet-web, and cobweb weavers) possess up to seven gland types, each producing silk fibers or glues with distinctive mechanical properties that correspond to a particular function. Almost all molecular studies of spider silks have focused on members of the gene family that encode the fibers' primary structural proteins – spidroins. Recently, high throughput sequencing of genes expressed in the silk glands of the Western black widow identified hundreds of transcripts that were significantly more abundant in silk glands than other tissues, suggesting a far more complex silk protein system than previously recognized. Here we describe gene expression patterns of all seven of the functionally differentiated silk gland types in the Western black widow and two close relatives. GO term analysis of differentially expressed transcripts identified similar functions enriched in each of the differentiated gland types in all three species, but the identity of these transcripts in each of the

individual gland types is unique. We then identify homologs of these differentially expressed genes in the recently sequenced house spider genome and determine if silk gland specific expression is conserved in this more distantly related cobweb weaving species.

**3 - i5K/EG**
**Variations in thermal history lead to dissynchronous diapause development**

Bennett, Anna, Insect Genetics and Biochemistry Research Unit, USDA-ARS
Additional authors:
Yocum, George; Insect Genetics and Biochemistry Research Unit, USDA-ARS
Rinehart, Joseph; Insect Genetics and Biochemistry Research Unit, USDA-ARS

The alfalfa leafcutting bee, *Megachile rotundata*, is the world's most intensively managed solitary bee for commercial pollination. It is the primary pollinator for alfalfa seed production. Managed bees are subjected to thermal regimes for overwintering and subsequent adult emergence in time for alfalfa bloom. Mating, foraging and nesting occurs in the field and larvae develop in brood cells provisioned by the mother. In nature, larval development of the first generation is typically completed by mid-July when a portion of bees undergo diapause, a period of suppressed metabolism and development, and overwinter as prepupa. However, a proportion of the population will avert diapause to produce a second generation, the progeny of which will predominately enter prepupal diapause if they can complete larval development before the onset of winter. Management practices during the prepupal stage affect the physiology of the adult bee and have implications for overwintering survival. Therefore, an understanding of diapause regulation will help improve management practices and enable this bee to be used in additional agricultural markets.

Diapausing prepupae produced early and late in the season were removed from nests on September 1, 2010 and divided between two management groups, those kept at a constant 4-5°C and those kept outdoors, exposed to natural temperature fluctuations. Each of these four treatment groups was sampled monthly from October to June. Samples from four time points (November, January, March and May) were chosen to span the diapause maintenance, termination and post-diapause quiescence stages of development. Two lanes of paired-end Illumina sequencing was performed on RNA from three bees from each of the four months (48 samples). Within month treatment comparisons of differentially expressed genes indicates all four treatment groups represent distinct populations, demonstrating the plasticity of the bee's phenotype in response to their environment. Additionally, while the diapause process synchronizes spring emergence, it does not synchronize the physiology of bees oviposited at different points in the season. Between month treatment comparisons of differentially expressed genes is providing a foundation for understanding the timing of processes regulating diapause development. These results confirm results from studies in other species conducted by ARS scientists in Fargo, ND showing that laboratory studies are necessary, but not sufficient to explain diapause physiology under field conditions. This multi-factorial characterization of *M. rotundata* diapause will not only aid in commercial management optimization, but will provide insights into mechanisms of quiescence plasticity.

**4 - i5K/EG**
**Genome sequence of the bed bug, Cimex lectularius**

Benoit, Joshua; Biological Sciences, University of Cincinnati
Additional authors:
Bed bug genome consortium

Bed bugs have re-established as a pest throughout much of the world during the last two decades. Various factors including increased pesticide resistance leading to poor control and their ability of long-term survival without food and water have been linked to this resurgence. In this study, we describe the genome of the common bed bug with a specific focus on obligate blood feeding in both sexes, mating through traumatic insemination, starvation resistance and factors responsible for pesticide resistance. Analyses of

the C. lectularius genome and predicted protein coding genes provides the first complete representation of genes that are linked to traumatic insemination, reduced olfaction/chemosensory processes, host-symbiont interactions and unique sets of genes associated with multiple mechanisms of insecticide resistance. Data provided by this genome sequencing and analysis will establish a foundation for research devoted to insecticide resistance and uncovering distinctive mechanisms that are integral to bed bug biology.

**5 - i5K/EG**
**The search for regulatory DNA controlling the expression of cuticle protein orthologs**

Brown, Katie; Biology Department, Northern Michigan University

Cuticle proteins comprise a large family based on motifs shared by the family members. Although these proteins have been heavily studied, there is little known about the *cis*-regulatory elements involved in the regulation of the genes encoding these proteins. I will be looking at putative orthologs of MSCP14.6 (*Manduca sexta* cuticle protein) and studying the 5' flanking DNA of these orthologs. The inspiration for this study came from research conducted in 1997 that compared the 5' flanking DNA of MSCP14.6 with HCCP12 (*Hyalophora cecropia* cuticle protein). These results showed a remarkable alignment (60% identity) of the flanking DNA for the cuticle gene under study using these two species. The high degree of conservation observed in the 5' flanking DNA suggests that there are functional elements (*cis*-regulatory elements) that are under selection pressure. Cuticle genes for other species, including but not limited to Lepidopterans, will be identified and analyzed for sequences that are vital for regulating cuticle protein synthesis.

**6 - i5K/EG**
**LepBase – a multi-genome database for the Lepidoptera**

Challis, Richard, Institute of Evolutionary Biology, University of Edinburgh
Additional authors:
Kumar, Sujai; University of Edinburgh
Dasmahapatra, Kanchon; University of York
Jiggins, Chris; University of Cambridge
Blaxter, Mark; University of Edinburgh

Technical and economic barriers to generating genome-scale datasets are falling away rapidly and researchers face daunting bioinformatic challenges associated with organizing, integrating, and disseminating the welcome deluge of data.

Simple deposition of data (sequence and annotations) in major public databases is not sufficient, as it is difficult to access and analyse such 'raw' genomes in a comparative context. However, single-species, single-lab databases are typically too brittle, fragmented, and inefficient to meet the needs of the wider community.

Recognizing the need for clade-level integration of species-level genomic resources that can flow 'upstream' into pan-genomic databases, a multi-tiered approach to aggregation, integration, and dissemination of the wealth of genomic information arising from community-driven genome projects is now a leading model. Where single-species (Tier 1) databases are aggregated into clade-level (Tier 2) databases from where they can be pushed into pan-genomic (Tier 3) databases, such as Ensembl Metazoa.

LepBase is a Tier 2 database for the Lepidoptera, providing not only an aggregation of existing genomic data into a single, consistent portal but also a platform for comparative analysis and integrated, systems-oriented inferences that are not possible in single-species programmes. Our position as a clade level resource allows us to work closely with the Lepidoptera research community to incorporate the data and analyses that are of greatest value to current and future research programmes on this diverse clade.

Our initial release in February this year comprised an Ensembl genome browser with five Lepidopteran species (http://lepbase.org/lepbase-ensembl-release-0-5/). We expect to host over 50 genomes within the next year, along with comparative genomics analyses such as whole-genome alignments, orthology predictions, and protein family analyses. Future releases will also include additional tools to allow gene-centric data exploration and novel visualisations for variant data from resequencing in population genetics studies.

**7 - i5K/EG**

**Two strategies for differential gene expression analysis of egg case silk glands in the spider *Argiope argentite***

Chaw, Crystal; Department of Biology, University of California, Riverside
Additional authors:
Arensburger, Peter; California State Polytechnic University
Ayoub, Nadia; Washington and Lee University
Hayashi, Cheryl; University of California, Riverside

Spider silk synthesis is a topic of intense research in a variety of fields including ecology, molecular evolution, biomimetics, and materials science. Many spiders in the genus *Argiope* have been the subject of silk studies, ranging from the behavioral ecology of web construction to the molecular organization and mechanical properties of *Argiope* silk. An individual *Argiope* spider can spin up to seven functionally distinct silks, which are primarily composed of various combinations of fibrous spider silk proteins (spidroins). For example, female *Argiope* produce egg case swathing silk from specialized tubuliform silk glands. These glands synthesize and store tubuliform spidroin (TuSp1), the main protein in egg case silk. Although it is known that *TuSp1* is highly expressed in tubuliform silk glands, much remains unknown about the genetics of tubuliform silk production. A powerful approach to functional genomics questions is to identify differentially expressed genes with deep sequencing of cDNA. However, *de novo* transcriptome assembly and differential gene expression analysis using short sequencing reads are particularly challenging with spider silk glands. *De novo* transcriptomes can be highly fragmented, and spidroins, which feature a lengthy repetitive region flanked by non-repetitive amino and carboxyl- terminal regions, are especially difficult to assemble. Fragmented repetitive regions can cause spidroin transcripts to be overrepresented. Thus, we hypothesize that targeting the non-repetitive terminal region of spidroins for sequencing will provide an accurate count of spidroin transcript abundance. Most differential gene expression analyses use one of two types of RNA-sequencing data: full-length RNA-sequencing (RNASeq) or 3' tag Digital Gene Expression (3' DGE). RNASeq data are widely utilized because they provide coverage over the length of a given transcript. In contrast, 3' DGE reads provide a single read per transcript from the 3' region. We assembled a *de novo* transcriptome using RNASeq data from the silver garden spider *A. argentata*. Then, we compared the results of differential gene expression analyses on tubuliform silk glands and cephalothorax control tissues using RNASeq or 3' DGE reads. Overall, we find that the 3' DGE differential gene expression analysis successfully identifies spidroins in tubuliform silk glands, and that the 3' DGE analysis outperforms the RNASeq analysis with respect to identifying lowly expressed spidroins with carboxyl-terminal variations. From both analyses, we identify a subset of genes that are highly differentially expressed in *A. argentata* tubuliform silk glands. These likely include new candidate genes that may be crucial to the synthesis of egg case and other types of spider silk.

**8 - i5K/EG**
**Ancient Whole Genome Duplication Event in Ancestor of Spiders Influential in Evolution of Silk Gland Transcriptome**

Clarke, Thomas; Department of Biology, Washington and Lee University
Additional authors:
Garb, Jessica; University of Massachusetts Lowell
Hayashi, Cheryl; University of California Riverside
Sharma, Prashant; American Museum of Natural History
Wierschin, Torsten; Ernst Moritz Arndt University Greifswald
Ayoub, Nadia; Washington and Lee University
Matthias Pechmann, Georg-August-Universität Göttingen
Richard Gibbs, Baylor College of Medicine
Mario Stanke, Ernst Moritz Arndt University Greifswald
Nikola-Michael Prpic, Georg-August-Universität Göttingen;
Nico Posnien, Georg-August-Universität Göttingen
Stephen Richards, Baylor College of Medicine
Alistair P. McGregor'

The emergence of specialized tissues with novel functions, such as the silk synthesizing glands in spiders, has been hypothesized to rely on large-scale gene duplication events and subsequent paralog divergence. Such an event has been proposed for spiders, but not tested. Based on phylogenetic analyses of thousands of gene families with representatives from *Parasteatoda* (i5k genome), three additional theriidiid transcriptomes, 24 other arachnids, a chelicerate and a manidubulate, we found numerous duplicates indicative of a whole genome or segmental duplication. We found that a duplication event likely occurred near the divergence of scorpions (Order Scorpionida) and spiders (Order Araneae), in addition to younger and older duplication events. Gene families containing at least one transcript that is over-expressed in the silk glands of a theridiid are more likely to contain an ancient paralog pair from this event than gene families lacking silk specific transcripts. Likewise, the ancient duplication nodes with a descendant that is an over-expressed silk transcript retain significantly more paralog pairs than duplication nodes that lack a silk transcript. Thus the ancient large-scale gene duplication event likely provided genetic material for subsequent silk gland transcriptome evolution in the spider phylogeny.

**9 - i5K/EG**
**The foundation of complex social interactions: Lessons from the subsocial beetle *Nicrophorus vespilloides***

Cunningham, Christopher; Genetics, University of Georgia, Athens
Additional authors:
Ji, Lexiang; Department of Genetics, University of Georgia, Athens GA 30602
Sheldon, Jennifer; KSU Bioinformatics Center, Division of Biology, Kansas State University, Manhattan KS 66506
Schmitz, Robert; Department of Genetics, University of Georgia, Athens GA 30602
Brown, Susan; KSU Bioinformatics Center, Division of Biology, Kansas State University, Manhattan KS 66506
Moore, Allen; Department of Genetics, University of Georgia, Athens GA 30602

The genetic basis of complex social behavior of insects has been addressed in several eusocial species where genomic information is available. However, the generality of the patterns that have been found is unknown as genomes from taxa that represent evolutionary intermediates between eusocial and non-social insects, such as subsocial species, are not yet readily available. To begin to fill this gap, we sequenced the genome of the subsocial beetle *Nicrophorus vespilloides* to investigate the genomic underpinning of complex social interactions, as well as to help bridge the gap between social and non-social taxa. *Nicrophorus vespilloides* is usually a solitary species but they will form a family unit when a reproductive resource (a vertebrate carcass) is located. Both male and female *N. vespilloides* care for

offspring, including regurgitation of predigested food into the mouths of begging larvae. Afterwards the parents and offspring will disperse from the carcass and again be solitary. We have assembled >90% of the genome into ~4600 scaffolds with an N50 of ~123kb using Illumina short reads, PacBio long reads, and a BioNano optical map. Because *N. vespilloides* has transient but drastic changes in behavior, we also used bisulfite-treated genome sequencing to identify and characterize any DNA methylation, which might play a role in the behavioral changes seen in this species. *Nicrophorus vespilloides* provides important independent tests about the evolution and mechanisms of both sociality and parental care, and provides insight into how ubiquitously old genes are co-opted to influence new behaviors.

**10 - i5K/EG**
**Improving the *Heliconius melpomene* genome with linkage mapping, Pacific Biosciences sequencing, and haplotype merging**

Davey, John; Department of Zoology, University of Cambridge
Additional authors:
Simpson, Fraser; University College London
Dasmahapatra, Kanchon; University of York
Mallet, James; Harvard University
Jiggins, Chris; University of Cambridge

Genome resequencing of *Heliconius* butterflies show evidence of gene flow across 40% of the genome in one species pair. To study this further, high quality, contiguous genome sequences of these butterflies are required. Our initial draft assembly of the *H. melpomene* genome, published in 2012, contained over 4,000 scaffolds, but a further 8,000 scaffolds were removed as they were identified as candidate haplotype material. 85% of the genome was placed on chromosomes using a linkage map built by RAD sequencing a mapping cross, with over 200 misassemblies fixed based on chromosome assignments. The primary assembly was 273 Mb long with a scaffold N50 of 194 kb, but did not include the haplotype scaffolds. Also, because RAD markers are on average 10 kb apart, misassemblies were imprecisely identified.

We have now considerably improved this genome assembly using three approaches: 1. Incorporating haplotype scaffolds using HaploMerger, which halved the number of scaffolds and increased the scaffold N50 to 223 kb; 2. Generating 20x coverage of the *H. melpomene* genome with Pacific Biosciences sequencing, assembling this sequence and incorporating it into our draft genome, and again merging with HaploMerger, which increased the scaffold N50 to 634 kb; 3. Whole genome sequencing our mapping cross to incorporate 96% of scaffolds on the chromosome map and precisely identify our known misassemblies and a further ~200 misassemblies down to a few hundred base pairs. This assembly incorporates an extra 10 Mb of material and has a scaffold N50 of over 1 Mb. We believe these approaches can assist many heterozygous insect genome assembly projects.

**11 - i5K/EG**
**The i5K house spider genome reveals unexpected diversity and shifts in the expression of genes associated with the extreme toxicity of black widow venom.**

Garb, Jessica; Department of Biological Sciences, University of Massachusetts Lowell
Additional authors:
Gendreau, Kerry; University of Massachusetts Lowell
Haney, Robert; University of Massachusetts Lowell
Schwager, Evelyn; Oxford Brookes University
Ayoub, Nadia; Washington and Lee University

Venoms are protein-rich secretions with biomedical significance that have independently evolved in numerous animal lineages, but have only been studied at the whole-genome level in a few species. The genomes of the common house spider (*Parasteatoda tepidariorum)* and the Western black widow spider (*Latrodectus hesperus*) have been sequenced as part of the i5k pilot, and though these species are closely

related, black widow venom exhibits far greater toxicity than house spider venom. Black widow venom is dominated by latrotoxins and latrodectins, two protein families that are virtually unknown outside of a few closely related species. Thus, we investigated the diversity, evolution and expression of latrotoxin and latrodectin genes with the recently annotated house spider genome using tissue-, sex- and stage-specific Illumina expression data. We discovered ≥47 latrotoxin genes in the house spider genome, many of which are tandem arrayed. House spider latrotoxins vary extensively in predicted structural domains and stage and tissue specific expression, suggesting novel functional roles. We uncovered far fewer latrodectin genes, but expression and phylogenetic evidence shows their recruitment for venom function from a neuropeptide hormone family following gene duplication, inversion and coincident domain truncation. While latrodectins and other peptides are highly expressed in house spider venom glands, latrotoxins represent a far smaller proportion of house spider venom-gland expressed transcripts. Phylogenetic analyses show house spider latrotoxins are highly divergent from black widow latrotoxins, and this, along with the lower expression of latrotoxins in house spider venom, may explain the far greater potency of black widow venom. We will present plans to extend this work to the i5k black widow genome, including strategies to improve the current genome assembly.

**12 - i5K/EG**
**Strategies and considerations to improve assemblies of non-model arthropod genomes towards chromosomal scale datasets**

Geib, Scott; Pacific Basin Agricultural Research Center, USDA-ARS
Additional authors:
Calla, Bernarda; USDA-ARS-PBARC
Sim, Sheina; USDA-ARS-PBARC/U. Hawaii

To successfully generate a high quality assembly of a non-model arthropod, considerations must be made in terms of what resources are available to generate the most comprehensive dataset. Depending on the goal of the sequencing project, one might be interested in a single individual genome, composite assembly of multiple related organisms or generating comparable assemblies between related species. We compare utilization of several sequencing strategies towards generation of chromosome-scale assemblies of Tephritid fruit fly genomes. This includes considerations during sample selection, library prep, assembly algorithms, as well as accessory data to work towards high quality assemblies. Super-scaffolding techniques, such as linkage map construction and contact map generation are compared in terms of cost effectiveness, accuracy, and completeness to other techniques and emerging long-read technologies. Overall improved genomes can be generated while also reducing the cost in non-model arthropods.

**13 - i5K/EG**
**Phylogenetic distribution of transferrin receptors and secreted ferritins**

Gorman, Maureen; Biochemistry and Molecular Biophysics, Kansas State University
Additional authors:
Brummett, Lisa; Kansas State University
Ragan, Emily; Metropolitant State University of Denver

The transport of iron from one cell to another is essential for many aspects of an animal's life, including energy metabolism, cell division and detoxification. The mechanisms of iron export and iron uptake by mammalian cells have been known for some time, and the process of iron export by insect cells is well understood; however, surprisingly little is known about iron transport in other types of animals. In mammals, iron is transported out of cells by an iron permease, loaded onto transferrin, and transported to cells that take up transferrin-bound iron via receptor-mediated endocytosis. Mammalian ferritin is a cytoplasmic iron storage protein. In insects, iron is exported by the secretion of iron-loaded ferritin. Insect transferrin is also secreted into the hemolymph, but insect transferrin receptors have not been identified. One hypothesis about the function of insect transferrin is that it may act as an immune protein rather than an iron transport protein. Our working model of iron transport in insects is that after holoferritin is secreted

into the hemolymph, it binds to an unidentified ferritin receptor on iron-deficient cells, and is taken up by receptor-mediated endocytosis. Given the significant differences in iron transport mechanisms in mammals versus insects, we were curious about iron transport in other types of animals. If only vertebrates have identifiable transferrin receptor genes, then perhaps only vertebrates use transferrin as an iron transport protein. If only insects have secreted ferritins, then perhaps only insects use ferritin as an iron transport protein. If some types of animals have no identifiable transferrin receptor genes and also no secreted ferritins, then they must have an undiscovered iron transport mechanism. To address these hypotheses, we searched a diverse set of metazoan genomes for the presence of transferrin receptor and secreted ferritin genes. Our results indicate that the mechanisms of iron transport in most types of animals are still unknown.

**14 - i5K/EG**
**Assembly and annotation of the marbled crayfish genome**

Gutekunst, Julian; Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center
Additional authors:
Falckenhayn, Cassandra; Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center
Raddatz, Guenter; Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center
Lyko, Frank; Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center

Marbled crayfish are the only freshwater crayfish known to reproduce by cloning (apomictic parthenogenesis). Notably, among genetically identical offspring raised in the same environment phenotypic differences can be observed. Such non-genomic characteristics render the marbled crayfish an interesting laboratory model, especially for the field of epigenetics. We experimentally determined the genome size at approximately 3.8 Gbp by k-mer analysis and flow cytometry. Two individual females (Koelle, Steuerwald) were sequenced using shotgun sequencing with various insert sizes generating 350 Gbp and 196 Gbp of data respectively. High coverage sequencing data of Koelle was used to produce a first de novo draft assembly with a length weighted median contig size (N50) of 809 bp and scaffold N50 of 41 kb. To unambiguously demonstrate clonal reproduction in the marbled crayfish we are currently evaluating sequencing data from the second individual (Steuerwald). Transcriptome data provides additional information for quality control and assembly refinement. Genome wide comparisons to other arthropods will allow us to define characteristic features of decapods as ecologically and economically keystone species.

**15 - i5K/EG**
**Discovery of potential regulatory elements controlling tissue-specific expression in the common house spider (Parasteatoda tepidariorum)**

Haney, Robert: Biological Sciences, University of Massasschusetts Lowell
Additional authors:
Schwager, Evelyn; Oxford Brookes University
Garb, Jessica; University of Massachusetts Lowell

Complex organisms are composed of many distinct cell and tissue types with diverse functions, the development and maintenance of which depends on the differing composition of the underlying tissue transcriptomes. High-throughput sequencing technologies permit robust characterization of tissue-biased transcriptomes and facilitate identification of the genomic regions which control their expression. In the common house spider, Parasteatoda tepidariorum, we sequenced the transcriptomes of two tissues, silk glands and venom glands, critical to spider functional ecology. Together with data from ovarian tissue, we defined sets of genes with highly biased expression in each tissue, together with a broadly expressed control set. We identified tissue-biased transcription factors that may contribute to tissue-biased expression of downstream genes. Furthermore, utilizing the recently sequenced genome of this species, we were able to identify putative promoter regions based on the position of annotated transcription start

sites for both tissue-biased and broadly-expressed control sets of genes. Promoter regions were explored for known and novel DNA motifs overrepresented in tissue-biased promoters, which represent candidate cis-regulatory elements controlling tissue-biased expression.

**16 - i5K/EG**
**Genome size diversity in Arthropods**

Hanrahan, Shawn; Department of Entomology, Texas A&M University
Additional authors:
Johnston, J. Spencer; Texas A&M University

Genome size diversity among Arthropods remains poorly sampled, with far less than one percent of species surveyed. To date, the genome size for these species has ranged from 98 Mbp for a midge (Chironomidae) all the way up to 17,000 Mbp for a shorthorn grasshopper (Acrididae). This variation is not equally distributed among orders. The largest genome sizes are almost exclusively found within the Orthoptera, with many acridid species surpassing 10,000 Mbp. The size of these genomes has limited complete sequencing efforts. We surveyed genome size over 300 species across 21 orders. These estimates are combined with the 1400 records of arthropods from the Animal Genome Size Database to show the known range and average genome size of each sampled order of the arthropods.

The full range of genome size has yet to be discovered for any order. It is evident that this range will change, as additional sampling is performed. Because genome size will directly correlate with the cost of any sequencing project, a genome size estimate is an important part of any complete sequencing project. Moreover, an accurate estimate is essential at the end of a complete sequencing effort as a measure of the completeness of that effort. Closely related species may differ very significantly in genome size; and knowledge of the genome size of a relative can be a poor predictor of the genome size of a given species.

**17 - i5K/EG**
**Hox gene modification is driving mimetic color pattern variation in bumble bees**

Hines, Heather; Departments of Biology and Entomology, Pennsylvania State University
Additional authors:
Franzini, Luca; Pennsylvania State University
Ezray, Briana; Pennsylvania State University

Bumble bees are involved in an exceptional phenotypic diversification, whereby the 250 global species have diversified into over 600 largely segmental color patterns of their thick setal pile. These rapid color changes are considered primarily driven by convergence of bees onto local mimicry patterns. This diversification and convergence, and the segmental nature of their color shifts, makes these bees ideal for gaining insights into the evolutionary genetic and developmental mechanisms underlying phenotypic diversification. Given evidence from *Drosophila* coloration, we predicted that segmentation genes are driving pigmentation shifts through cis-regulatory modification of key pigment genes. To begin to address the genetics underlying this diversification, we studied the genetic basis of red-black convergence onto two North American mimicry complexes in the bumble bee species, *Bombus melanopygus*. Using whole genome sequencing of individuals across the hybrid zone between Pacific Coastal and Rocky Mountain zones, alignment to the genome of close relative *B. impatiens*, and genotype-phenotype association analysis, we have targeted the genomic interval involved in regulating this color variation. This region falls between Abd-A and Abd-B, the major developmental Hox genes that regulate abdominal segment identity. These results are counter to our predictions and general predictions of Evo-Devo, demonstrating that the evolutionarily conserved Hox genes themselves are altered and not genes downstream.

**18 - i5K/EG**
**Genome sequencing of the damselfly *Calopteryx splendens* (Arthropoda:Odonata)**

Ioannidis, Panagiotis; Medical School, University of Geneva
Additional authors:
Ioannidis, Panagiotis; University of Geneva Medical School, rue Michel-Servet 1, 1211 Geneva, Switzerland
Simao Neto, Felipe; University of Geneva Medical School, rue Michel-Servet 1, 1211 Geneva, Switzerland
Niehuis, Oliver; Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany
Misof, Bernhard; Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany
Zdobnov, Evgeny; University of Geneva Medical School, rue Michel-Servet 1, 1211 Geneva, Switzerland

Insects are one of the most diverse animal groups, representing more than half of all described species on the planet. The number of known species is more than one million, while it is estimated that the total number of species is close to eight million species. Their diversity is owed to the many adaptations they have, which enable them to thrive in virtually every habitat. In particular, adaptations like wings or metamorphosis had a large positive effect on this widespread occurrence of insects. Of course, the study of the evolution of these phenotypes and the associated genes must include insect species in which these traits are still in the plesiomorphic state. Such species include members of the early-diverging insect lineages, such as Odonata, Ephemeroptera, Thysanura and Archaeognatha. To this end, and within the framework of the i5k Initiative (5,000 Insect Genome Project), we proposed the genome sequencing of a number of insects from early-diverging lineages. Among them is that of the damselfly *Calopteryx spendens* (Odonata), for which preliminary results will be presented.

**19 - i5K/EG**
**The evolution of the *oskar* gene in insects**

Jones, Tamsin; Organismic and Evolutionary Biology, Harvard University
Additional authors:
Extavour, Cassandra; Harvard University

*oskar,* an insect-specific gene, is the only animal gene known to be both necessary and sufficient to specify germ cells. In the fruit fly *Drosophila*, germ cells are specified by maternal inheritance of a specialized cytoplasm in the oocyte, named germ plasm, and Oskar protein is responsible for the formation of germ plasm. *oskar* is a novel gene within the insect lineage, and was long thought to be restricted to the holometabolous insects (those that undergo complete metamorphosis). It was hypothesized that the appearance of *oskar* coincided with the evolution of germ plasm in holometabolous insects, and thus that its ancestral function was in germ cell specification. However, recent work in our laboratory showed that *oskar* evolved before the emergence of higher insects, as it is present in the basally branching insect *Gryllus bimaculatus* (the two-spotted cricket), which does not have germ plasm. *Gryllus* specifies its germ cells via a different mechanism, whereby germ cells are induced later in development by BMP signaling from neighbouring somatic cells. Further, *oskar* does not have a germ line role in *Gryllus*. Instead, it is expressed in the developing nervous system and is required for correct neural development. In this study we have searched for the *oskar* gene in genome and transcriptome datasets from insects and non-insect hexapods in order to further examine the evolutionary history of this gene. Here we will present the results of this study, including identification of previously unknown *oskar* orthologs in several insect lineages, and discussion of *oskar*'s sequence evolution throughout the insect phylogeny.

**20 - i5K/EG**
**A Draft Genome Assembly of the Lesser Wax Moth, *Achroia grisella***

Koseva, Boryana; Ecology and Evolutionary Biology, University of Kansas
Additional authors:
Kelly, John; Ecology and Evolutionary Biology, University of Kansas
Macdonald, Stuart; Molecular Sciences, University of Kansas

The lesser wax moth *Achroia grisella* is a symbiont of the honeybee *Apis mellifera* with an interesting reproductive behavior - while in most moth species females use signals to attract males, in the lesser wax moth the roles are reversed. In this study, we sequenced and assembled genomic DNA into 74,159 scaffolds of total size 418Mb with N50 of 87.3Kb. We performed CEGMA analysis to assess the completeness of our assembly. Of the 248 core eukaryotic genes (CEGs), our assembly contains 196 (79.03%). To explore synteny between *A. grisella* and *Bombyx mori*, we aligned their genomic sequences using PROmer. We found that 80% of the scaffolds contain segments that primarily align to a single *B. mori* chromosome with a rate of 90%. Finally, we used Maker to predict and annotate genic regions, and the pipeline generated 6,527 predictions, with an average intron length of 2Kb. The goal of this study is to provide a quality draft genome to serve as a molecular tool to explore the genetics of traits of evolutionary interest.

**21 - i5K/EG**
**Long non-coding genes in honey bees and their prospective alliance with virus disease**

Kwon, Hyung Wook; Major in Biomodulation, Department of Agricultural Biotechnology, Seoul National University

Additional authors:
Murukarthick Jayakodi
Je-won Jung
Tae-Jin Yang

Long non-coding RNAs (lncRNAs) are a class of RNAs that do not encode proteins. Recently, lncRNAs have gained special attention for their roles in various biological process and diseases.
In an attempt to identify long intergenic non-coding RNAs (lincRNAs) and their possible involvement in honey bee development and diseases, we analyzed RNA-seq datasets generated from Asian honey bee (*Apis cerana*) and western honey bee (*Apis mellifera*). We identified 2,470 lincRNAs with an average length of 1,011 bp from *A. cerana* and 1,514 lincRNAs with an average length of 790 bp in *A. mellifera*. Comparative analysis revealed that 5% of the total lincRNAs derived from both species are unique in each species. Our comparative digital gene expression analysis revealed a high degree of tissue-specific expression among the seven major tissues of honey bee, different from mRNA expression patterns. A total of 863 (57%) and 464 (18%) lincRNAs showed tissue-dependent expression in *A. mellifera* and *A. cerana*, respectively, most preferentially in ovary and fat body tissues. Importantly, we identified 11 lincRNAs that are specifically regulated upon viral infection in honey bees, and 10 of them appear to play roles during infection with various viruses. This study provides the first comprehensive set of lincRNAs for honey bees and opens the door to discover lincRNAs associated with biological and hormone signaling pathways as well as various diseases of honey bee.

**22 - i5K/EG**

**Improvement of the assembly of heterozygous genomes of non-model organisms, a case study of the genomes of two *Spodoptera frugiperda* host strains.**

Legeai, Fabrice, IGEPP/IRISA, INRA INRIA
Additional authors:
Anaïs Gouin[1], Anthony Bretaudeau[2], Karine Labadie[3], Jean-Marc Aury[3], Emmanuelle d'Alençon[4], Claire Lemaitre[1] and Fabrice Legeai[1,2]
[1] INRIA/IRISA équipe GenScale, Campus Beaulieu 35042 Rennes CEDEX
[2] INRA IGEPP, domaine de la Motte 35653 LE RHEU CEDEX
[3] CEA Genoscope rue Gaston Crémieux 91000 Evry
[4]INRA DGIMI, université de Montpellier 1, 34000 Montpellier

The extraction of biological information from the draft genomes of non-model organisms may result in unattainable, incomplete, or even wrong conclusions. In particular, the combination of a high level of heterozygosity and short reads sequencing may have major impact in the annotation of genes [1,2]. This wrong gene content assessment is usually the consequence of the high fragmentation of the genome sequence but it may also come from an overestimation of the genome size. The latter because the assembly of an heterozygous region for which there is a significant divergence between the two haplotypes leads sometimes to the construction of two different contigs, instead of one consensus sequence. To date, new assemblers such as *Platanus* [3], have been developed in regard to heterozygous data. But, the complete re-assembly of a genome leading to new automatic and manual annotations process is very cost-effective, and may still produce erroneous scaffolds and annotations. Thus, we set up a « soft » method to detect and correct false duplications due to heterozygosity in draft assemblies. In addition, to the identification and removal of the allelic regions (i.e. unmerged haplotypes), our protocol is able to relocate and merge supernumerary gene annotations.

We applied this method as a pre-requisite for the comparison of the genomes of 2 *Spodoptera frugiperda* (Lepidoptera: Noctuidae) strains, in the frame of the WGS project supported by the Fall armyworm International Public Consortium (FAW-IPC). This moth is a well-known pest of crops throughout the Western hemisphere. This species consists of two strains adapted to different larval host-plants: the first feeds preferentially on corn, cotton and sorghum whereas the second is more associated with rice and several pasture grasses. While, the paired-end reads of the rice-variant have been directly assembled using *Platanus* [3], we cleaned up and corrected the first release of the corn-variant, leading to a drastic reduction of the genome assembly, with the removal of 88Mbp (17%) and the increase of the N50 from 39,593 to 52,781bp. The suppressed fragments included 3,746 gene predictions; about 80% of them have been either relocated or merged with their complementary allele.

Subsequently, in order to identify new candidate genes or genomic regions involved in the host-plant adaptation, we compared the genomes and proteomes of the 2 different strains in order to identify orthologous genes, collinear regions and genome rearrangements, taking into consideration the inflated occurrence of splitted genes due to the high fragmentation of the genome.

**23 - i5K/EG**
**Genome sequencing and annotation of the cricket *Gryllus bimaculatus*, a hemimetabolous insect model**

Mito, Taro; Department of Life Systems, Institute of Technology and Science, Tokushima Unversity
Additional authors:
Itoh, Takehiko; Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Japan
Morimoto, Hiroya; Graduate School of Bioscience and Biotechnology, Tokyo Institute of Tech., Japan
Kajitani, Ray; Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Japan
Toyoda, Atsushi; Advanced Genomics Center, National Institute of Genetics, Japan
Tomonari, Sayuri; Institute of Technology and Science, Tokushima University, Japan
Masao Fuketa, Institute of Technology and Science, Tokushima University, Japan
Takahito Watanabe, Center for Collaboration among Agriculture, Industry and Commerce, Tokushima University, Japan
Yuji Matsuoka, Institute of Technology and Science, Tokushima University, Japan

The cricket *Gryllus bimaculatus* is a model for the study of hemimetabolous insect development. RNAi, transgenesis and, more recently, gene knockout/knockin via genome editing have been achieved for analyzing gene function in this species. We have been conducting whole genome sequencing of *G. bimaculatus* to further promote functional and comparative genomics. The draft genome has been generated by nextgen sequencing using Illumina Hiseq and by assembling short reads using a de novo assembler, Platanus. The assembled genome consists of 49,588 scaffolds (5.77 Mb N50) with total length 1.54Gb and 22,195 predicted protein-coding genes. Our annotation revealed huge clusters of Hox genes and Wnt genes. Our ongoing progress on *G. bimaculatus* genome analysis will be presented and discussed from an evolutionary viewpoint.

**24 - i5K/EG**
**Identification and annotation of *Frankliniella occidentalis* V-ATPase gene family members**

Oliver, Jonathan; Department of Plant Pathology, Kansas State University
Additional authors:
Whitfield, Anna; Kansas State University
Rotenberg, Dorith; Kansas State University

The genome of the western flower thrips, *Frankliniella occidentalis* (Pergande), was sequenced by the Baylor College of Medicine – Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/content/i5k-western-flower-thrips) as part of a pilot study for an international initiative to sequence 5,000 arthropod species (i5k). In the current assembly, there are 415 Mb of sequence with 17,553 predicted genes distributed across 6,263 scaffolds. Annotation of this genome is currently underway, and as the first genomic resource available for the insect order *Thysanoptera* it is expected to provide valuable tools for thrips research. To examine the usefulness of the genome assembly, a search was carried out for members of the highly conserved vacuolar ATP synthase (V-ATPase) multigene family within the *F. occidentalis* genome assembly. The V-ATPase family of genes code for the individual subunits of the multi-subunit vacuolar ATPase complex, which functions as an ATP-driven proton pump in a wide range of organisms to acidify intracellular compartments and the extracellular environment. Thirteen members of the V-ATPase multigene family were identified and manually-annotated for the i5K *F. occidentalis* genome housed at the National Agricultural Library i5K Workspace using the Web Apollo interface to view predicted gene models within the original assembly and Roche-454 Newbler-assembled contigs previously generated from a cDNA library prepared from isolated *F. occidentalis* mRNA. Eleven of thirteen identified members of this family were found to be present in single copies within the assembly, while three copies each of V-ATPase V0 subunits a and c were identified, with each V-ATPase subunit gene located on a different scaffold. Previous work in our laboratory with *F. occidentalis* documented that silencing V-ATPase V1 subunit B resulted in decreased fertility and increased mortality, and the annotation of other members of this gene family will enable additional functional studies of thrips

genes. Overall, this work illustrates that these new genomic tools can be utilized for gene annotation and will undoubtedly open new avenues for *F. occidentalis* research.

**25 - i5K/EG**
**Update on sequencing the genome of the lesser grain borer, *Rhyzopertha dominica***

Oppert, Brenda; SPIERU, USDA ARS CGAHR
Additional authors:
Morgan, Thomas; USDA ARS CGAHR
Friesen, Kenlee; USDA ARS CGAHR
Hartzer, Kris; USDA ARS CGAHR
Perkin, Lindsey; USDA ARS CGAHR
Campbell, James; USDA ARS CGAHR
Timothy Smith, U.S. Meat Animal Research Center, Clay Center, NE USA David Schlipalius, Department of Agriculture, Fisheries, and Forestry, Queensland, Australia

This report is a summary of our efforts to attain a draft assembly of sequences from the genome of the lesser grain borer, *Rhyzopertha dominica*. This insect is an economically-important storage pest worldwide, with the immature stages developing within a wheat kernel. The most effective control method for *R. dominica* infestations is phosphine fumigation. However, like other control methods for problematic insect pests, the efficacy of phosphine fumigation has been reduced because of resistant insect populations. We have phosphine-resistant and -susceptible colonies of *R. dominica*, but lack genetic resources for our genetic and transcriptomic studies. The haploid genome size of *R. dominica* is estimated to be 476 Mb, with 94 markers mapping to nine linkage groups. We have four strains of *R. dominica* that were inbred for 20 generations, and genomic data has been acquired from one inbred strain through various sequencing platforms: PGM (2x coverage with adults, 3.2x male pupae, 3.6x female pupae); Illumina paired-end (2x300) MiSeq (17x adults); and PacBio (8x adults, P5 chemistry; 31x mixed pupae, P6). We are obtaining an additional 20-30x PacBio sequencing (mixed sex pupae, P6). Assembly has been in-house with SeqManNGen (DNAStar, 64 bit MacPro, 64-128 GB RAM), or off-site computing resources, such as iPlant (Soap deNovo), ECTools and MHap (Nimbix), and Falcon (PacBio). An assembly of only PGM and MiSeq data yielded 33,378 contigs with an N50 of 17 kb (mean contig = 11,184; max contig = 167,123); assembly with PacBio reads as well as additional hybrid assemblies are in progress. I will provide updated details of the assembly and progress in annotation, as well as an example of how these resources have been used to compare the transcriptome in *R. dominica* adults resistant or susceptible to phosphine. (Funding for this research has been provided by ARS CRIS Project 5430-4300-032-00D)

**26 - i5K/EG**
**Making a genome sequence work for you and the community**

Papanicolaou, Alexie; Hawkesbury Institute for the Environment, University of Western Sydney

The molecular biology community has overcome a major bottleneck: access to near-complete genome sequence from multiple species. Such diversity of sequence information underpins a new type of genomic capability and greatly benefits the comparative genomics field. How do we, however, can make best use of this information? In this talk I will be outlining lessons learned from sequencing, annotating and sharing information from multiple genome projects. First, I will outline some best practices for designing a sequencing strategy for insects. Second, I will present the Just_Annotate_My_Genome, a platform that allows nascent bioinformaticians to structurally annotate genomes at the same standard provided by NCBI. Third, I will showcase DEW and the Just_Annotate_My_Proteins software, a visualisation platform for functional annotation via gene expression and sequence comparisons that biologist without computational training can actually use. Fourth, I will outline the progress of curating major genome sequencing efforts, namely the Mediterranean fruit fly, Helicoverpa armigera/zea and Heliconius erato genome projects, and

finally I will close with a status update on our NESCent-funded Community Curation working group (http://nescent.org/science/awards_summary.php?id=377).

**27 - i5K/EG**
***De novo* characterization of transcriptome and genome assemblies from the Asia I mtCOI genetic clade of *Bemisia tabaci***

Patel, Mitulkumar; Natural Resources Institute, University of Greenwich
Additional authors:
Colvin, John; Natural Resources Institute
Bailey, David; Natural Resources Institute
Seal, Susan; Natural Resources Institute

*Introduction:* Next generation sequencing of transcriptomes and the genome of the Asia I mtCOI genetic clade of *Bemisia tabaci* has been undertaken. A genetic comparison of this putative species with other important *B. tabaci* and related insect species may provide targets for the development of more effective whitefly control strategies, as well as improved diagnostic markers.

*Materials and Methods:* Three transcriptomes of adult *B. tabaci* (unnormalized and normalized female, and unnormalized male cDNA libraries) were constructed and sequenced. Contigs were assembled and highly expressed genes identified. The top 1,000 of these genes were mapped onto a draft Asia I genome assembly produced from corresponding whitefly DNA sequences, using GMAP. The BLASTX program of the standalone BLAST2GO package was used for gene annotation by searching the non-redundant protein sequence database at NCBI. Gene Ontology (GO) terms were assigned to annotated contigs, and phylogenetic comparisons were made of Asia I sequences with closely related species using a range of different software programs.

*Results and Conclusion:* Roche 454 pyrosequencing of total mRNA from adult male and female 'Asia I *B. tabaci*' cDNA libraries generated between 300,000 and 560,000 sequence reads for each library. Contig assemblies, constructed from these sequences using the software program CLC Genomics Workbench, generated ~30,000 core contigs. GMAP analysis against a draft genome sequence assembled from corresponding genomic DNA, confirmed the accuracy of the transcriptome assemblies, with 7,967 transcripts mapping with 100% correspondence, followed by 3,392 transcripts with 99-99.9% identity, and 4,191 with 90-99% identity. Over 1,000 full-length *B. tabaci* gene sequences have been determined. Phylogenetic comparisons of these with *B. tabaci* published sequences, as well as unpublished transcriptomes from other *B. tabaci* populations, will be presented.

Acknowledgements: Sequencing was performed at the Gene Pool Sequencing Facility (Edinburgh, UK) and The Genome Analysis Centre (TGAC, Norwich, UK). The authors acknowledge helpful discussions with Stephen Bridgett, Mark Blaxter at Gene Pool, and Shabhonam Caim and Mario Caccamo at TGAC.

**28 - i5K/EG**
**Toxicogenome of Hyalella azteca: Exploring adaptation and gene expression in response to environmental pollution**

Poynton, Helen; School for the Environment, University of Massachusetts, Boston
Additional authors:
Major, Kaley; University of Massachusetts, Boston
Blalock, Bonnie; University of Massachusetts, Boston
Weston, Donald; University of California, Berkeley

Hyalella azteca is a cryptic species complex of freshwater epibenthic crustaceans of interest to both ecotoxicology and evolutionary biology. It is the primary invertebrate crustacean used in the U.S. for freshwater sediment toxicity testing and has been the subject of recent gene expression studies in

ecotoxicogenomics. In addition, our own work has documented the emergence of resistance to two distinct classes of insecticides in several populations of H. azteca. Using the recently sequenced genome as well as several gene expression data sets of H. azteca exposed to pesticides, metals, and emerging contaminants, we intend to annotate and characterize the major gene families involved in sequestration, detoxification, oxidative stress, and toxicant response. In addition, we are particularly interested in genes potentially involved in insecticide resistance including target site resistance (e.g. voltage-gated sodium channel, acetylcholinesterase) and metabolic resistance (e.g. cytochrome P450s, glutathione-s-transferases, and esterases). This effort will greatly facilitate the development of genomic tools for environmental assessments with H. azteca.

**29 - i5K/EG**
**Eleven insect genomes sequenced at University of Illinois**

Robertson, Hugh; Department of Entomology, University of Illinois at Urbana-Champaign

I will update our progress on sequencing ten insect genomes, those of the giant honey bee Apis dorsata, eastern bumblebee Bombus impatiens, the alfalfa leafcutter bee Megachile rotundata, the wheat-stem sawfly Cephus cinctus, a trap-jaw ant Odontomachus brunneus, a braconid parasitoid wasp Microplitis demolitor, a Rhagoletis parasitoid wasp Diachasma alloeum, the parsnip webworm Depressaria pastinacella, the navel orange worm Amyelois transitella, the snowberry fly Rhagoletis zephyria, and the western corn rootworm Diabrotica virgifera. Some are now published, most are in acceptable draft form undergoing annotation and analysis, while the last one remains insufficiently assembled.

**40 - i5K/EG**
**Dressing up the new Frankliniella occidentalis genome: Using a Tuxedo RNA-seq pipeline to quantify the response of western flower thrips during Tomato spotted wilt virus infection**

Schneweis, Derek; Plant Pathology, Kansas State University
Additional authors: Yes
Whitfield, Anna; Kansas State University
Rotenberg, Dorith; Kansas State University

The interplay between a persistent-propagative virus and its insect vector is dependent on interactions between both virus and vector molecules. Much is known about Tomato spotted wilt virus proteins and the roles they play in virus replication, infection, and movement within the plant host. While the western flower thrips vector, *Frankliniella occidentalis* (Pergande), is also undoubtedly an important player with respect to molecular interactions, there is very little known about thrips molecules and their role in the virus life-cycle. Our research provides an investigation of thrips gene expression changes during virus infection. With the advent of the i5K genome for *Frankliniella occidentalis*, we are now able to explore molecular interactions at a new level. The i5K genome provides, for the first time, a direct reference comparison with transcriptomics data from *Frankliniella occidentalis*. The genome currently consists of 6,263 scaffolds (N50 = 948.9kb). We constructed a genome-reference-based transcriptome assembly from thrips coding RNA derived from sequence reads of Illumina RNA-Seq libraries from four independent biological replications of TSWV-exposed and non-exposed thrips. RNAseq reads were aligned to the genome using Bowtie2 as a short read aligner and subsequently Tophat2 to identify splice junctions between exons. Cufflinks2 was exploited to construct a reference-based transcriptome assembly. Differential gene expression due to TSWV infection was quantified, and differentially expressed transcripts were identified. This necessarily large-scale transcriptomics is one of few methods available to explore host molecules involved in virus-host interactions on a global scale, and it is important for establishing an understanding of the dynamic and complex molecular interaction between TSWV and *F. occidentalis*. We are excited to report the use of the i5K genome for differential gene expression analysis, and we hope that many more discoveries arise from its creation.

**41 - i5K/EG**
**A Bioinformatic Analysis of the Extracellular Interactome from Multiple Insects**

Seeger, Mark; Molecular Genetics, The Ohio State University
Additional authors:
Schaffer, Patrick; Molecular Genetics, The Ohio State University
Glasbrenner, David; Molecular Genetics, The Ohio State University

Cell surface and secreted proteins play central roles during metozoan development and particularly in development of complex nervous systems. These extracellular proteins are often built from different combinations of conserved protein motifs, such as immunoglobulin domains, fibronectin type 3 repeats, semaphorin domains, and leucine-rich repeats to name a few. The repertoire of this extracellular interactome has expanded significantly during the evolution of multicellular organisms. While many of these proteins are conserved in diverse organisms such as planaria and humans, others are less conserved and may have unique functions in restricted groups of organisms, such as insects versus crustaceans or perhaps holometabolous versus hemimetabolous insects. The diversity of insects and increasing availability of sequenced genomes from the i5k project provides an opportunity to examine the diversity and evolution of this extracellular interactome. Results from our ongoing bioinformatic analysis of these gene families from a variety of insects across multiple insect orders will be presented.

**42 - i5K/EG**
**From sequence to locus: Using GBS and whole genome sequencing to identify the genetic basis of white pupae in SIT colony Medflies**

Sim, Sheina; Department of Plant and Environmental Protection Sciences, University of Hawaii, Manoa
Additional authors:
Calla, Bernarda; USDA-ARS Daniel K. Inouye PBARC
Hall, Brian; University of Hawaii, Manoa
DeRego, Theodore; University of Hawaii, Hilo
Yoneishi, Nicole; University of Hawaii, Hilo
Kauwe, Angela; University of Hawaii, Hilo
Raul Ruiz USDA-APHIS CPHST Norman Barr USDA-APHIS CPHST Scott Geib USDA-ARS Daniel K. Inouye PBARC

The Mediterranean fruit fly *Ceratitis capitata* (Wiedemann), commonly known as the medfly, is a destructive agricultural pest and the object of expensive population eradication and suppression efforts within state and federal departments of agriculture. Area-wide integrated pest management programs control medfly populations through the release of sterile males which must be massively produced. Mass-rearing and release of sterile males is facilitated by two sex-linked traits *white pupae* (*wp*) and *temperature sensitive lethal* (*tsl*). Though these two sex-linked traits in what is known as a genetic sexing strain was developed over 20 years ago, the genetic basis of *wp* is unknown. The purpose of this project was to identify SNP loci tightly linked to the causative mutation for *wp* in mass-reared sterile insect colonies. A high-quality reference genome from whole-genome sequences from individual flies and assembled using DISCOVAR was used along with GBS sequences from an F4 mapping population to identify SNPs linked to *wp*. The genomic region containing the SNPs tightly linked to *wp* was then investigated further to generate a list of candidate SNPs resulting in the *wp* mutation. This data has been used to develop a genetic assay for differentiating between recaptured sterile insect released males and wild males, will be useful for the improvement of existing sterile insect colonies, and can be used to identify an orthologous gene in other species to create novel sterile insect strains.

**43 - i5K/EG**
**BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**

Simao, Felipe A.; Department of Genetic Medicine and Development (GEDEV), University of Geneva Medical School / Swiss Institute of Bioinformatics
Additional authors:
Waterhouse, Robert; University of Geneva Medical School / Swiss Institute of Bioinformatics
Ioannidis, Panagiotis; University of Geneva Medical School / Swiss Institute of Bioinformatics
Kriventseva, Evgenia; University of Geneva Medical School / Swiss Institute of Bioinformatics
Zdobnov, Evgeny; University of Geneva Medical School / Swiss Institute of Bioinformatics

BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs Felipe A. Simão†, Robert M. Waterhouse†, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov Department of Genetic Medicine and Development, University of Geneva Medical School and Swiss Institute of Bioinformatics, rue Michel-Servet 1, 1211 Geneva, Switzerland. † Equal contribution. High-throughput genomics has revolutionised biological research, however while the number of sequenced genomes grows by the day, quality assessment of the resulting assembled sequences remains complicated and mostly limited to technical measures like N50. We propose a measure for quantitative assessment of genome assembly and annotation completeness based on evolutionarily informed expectations of gene content. We implemented the assessment procedure in open-source software, with large lineage specific sets of Benchmarking Universal Single-Copy Orthologs named BUSCOs. Our arthropod dataset contains over two thousand highly-conserved single-copy genes, allowing high-resolution quality quantification of newly sequenced genomes, transcriptomes and annotated gene sets. The initial set of gene annotations built by BUSCOs provides an excellent source of information for training gene-finding programs.

**44 - i5K/EG**
**Interactions between Manduca sexta β-glucan recognition protein and hemolymph protease 14 initiate prophenoloxidase activation pathway**

Takahashi, Daisuke; Biochemistry and Molecular Biophysics, Kansas State University
Additional authors:
Kanost, Michael; Kansas State University

Insect □-glucan recognition protein (□GRP) detects □-1,3-glucan on the surface of fungi, and this recognition triggers the auto-activation of an initiating serine protease, which subsequently activates a series of downstream serine proteases comprising the immune cascades for prophenoloxidase (proPO) activation. Here we studied the molecular basis underlying the auto-activation of HP14, an initiating serine protease in hemolymph of *Manduca sexta*, upon the recognition of □-1,3-glucan by □GRP2. Biochemical analysis using HP14 zymogen (proHP14) and □GRP2 and the recombinant proteins as truncated forms showed that LA domains, the amino-terminal regulatory domains within HP14, are required for the auto-activation of proHP14 through its interaction with carboxyl-terminal glucanase-like domain of □GRP2. Consistent with this conclusion, recombinant LA domains inhibit the activation of proHP14 and proPO, likely by interfering with the interaction between □GRP2 and LA domains within HP14. Surface plasmon resonance analysis further demonstrated that immobilized HP14 LA domains weakly bound to □GRP2 with a $K_d$ value of approximately 26 □M in the presence of calcium. Our results suggest that proHP14 is recruited to the □GRP2-□-1,3-glucan recognition complex through its LA domains and that the interaction between HP14-LA domains and □GRP2-glucanase-like domain leads to the autoactivation of HP14, thereby initiating proPO activation pathway.

**45 - i5K/EG**
**The *Oncopeltus* genome: getting the most out of a draft genome**

Vargas Jentzsch, Iris; Institute for Developmental Biology, University of Cologne
Additional authors:
Panfilio, Kristen; University of Cologne

The milkweed bug *Oncopeltus fasciatus* is one of the emerging insect model organisms complementing the well established *Drosophila* and *Tribolium* models. Due to its phylogenetic position prior to holometabolous radiation, *Oncopeltus* has frequently been used for comparative studies of insect development (*e.g*., embryonic axis determination, neural development, tissue specification). The *Oncopeltus* genome has been sequenced, assembled and automatically annotated as part of an i5k pilot project. Here, we give insight into our experience in creating and curating the manual annotation dataset to produce the official gene set (OGSv1.1). Manual annotation was performed in a community effort among 18 different research groups using the interactive genome browser Web Apollo. As a result, 1430 gene models were edited and/or added *de novo* (corresponding to about 7% of all predicted genes), allowing us not only to explore gene composition and structure in *Oncopeltus*, but also to identify misassemblies in the genomic scaffolds. Multiple rounds of gene model curation led to the resulting OGSv1.1, which was fed into downstream orthology analyses and more specific pipeline analyses, such as for metabolic network reconstruction and methylation signal detection. As the first i5k species to release an OGS, our experience will be of great value for genome projects that are just starting or are midway through the process of manual annotation. We discuss some of the main pitfalls that can lead to a significant delay in the construction of an OGS. Furthermore, we present a case study of how the information generated during manual annotation can be used to further optimize the quality of the genomic assembly, including when data from a hybrid assembly with PacBio re-sequencing are incorporated.

**46 - i5K/EG**

**Evolution of minor ampullate silk genes in orb-web and cobweb-weaving spiders and their transcriptional response to diet**

Vienneau-Hathaway, Jannelle; Biology, Washington and Lee
Additional authors:
Miller, Jeremy; Washington and Lee University
Brassfield, Elizabeth; Washington and Lee University
Garb, Jessica; University of Massachusetts, Lowell
Hayashi, Cheryl; University of California, Riverside
Ayoub, Nadia; Washington and Lee University

Spiders spin up to seven task-specific fibers or glues synthesized in specialized abdominal glands. Cobweb weaving spiders use silk synthesized in minor ampullate glands to wrap prey, while orb-web weavers use it for temporary capture spirals. These functions could select for differences in the minor ampullate silk proteins' (MiSp) sequences and physical properties. Furthermore, spiders potentially modulate the protein components of their silk in response to environmental factors, like prey. Because minor ampullate fibers contribute to prey capture in cobweb and orb-web weavers we predicted that spiders would change expression levels of MiSp in response to prey availability. Using previously published sequences, our own cDNA libraries, and the common house spider genome, *Parasteatoda tepidariorum* (recently sequenced by the i5K pilot project) we compared amino acid sequences of minor ampullate spidroins (MiSp) among four cobweb and four orb-web weaving species. We also tested if the common house spider changed expression levels of MiSp when fed different prey items by quantifying MiSp transcript abundance. We found that cobweb and orb-web weavers share MiSp amino acid motifs, but that within cobweb weavers there may have been an increase in elasticity-related motifs (e.g. increase in Proline) within one species. We also found that house spiders fed crickets were significantly heavier than those fed flies, but do not yet know if the two groups differed in MiSp transcript abundance.

**47 - i5K/EG**
**i5K@OrthoDB**

Waterhouse, Robert; Department of Genetic Medicine and Development, University of Geneva Medical School

Orthology delineation is a cornerstone of comparative genomics, offering qualified hypotheses on gene function by identifying “equivalent” genes in different species, as well as highlighting shared and unique genes that offer clues to understanding species diversity. The almost 90 arthropod species included in the latest release of the OrthoDB hierarchical catalog of orthologs (OrthoDB v8, www.orthodb.org, Kriventseva *et al*, NAR, 2015) offers the most comprehensive orthology resource for arthropod comparative genomics, leading to a maturing understanding of the composition of the insect gene repertoire (Waterhouse, COIS, 2015). Mapping of the annotated gene sets from newly-sequenced and annotated genomes of *i*5K pilot project species to OrthoDB orthogroups allows researchers to make the most of their newly-sequenced arthropod genomes. With now more than 100 sequenced and annotated arthropod genomes, thanks to the progress of the *i*5K pilot project, comparative genomics approaches are becoming ever more powerful tools to improve and extend genome annotation and interpretation for newly-sequenced species.

# Agricultural Vector Genomics

**48 - AVG**
**The immune signaling pathways of Manduca sexta**

Bhattarai, Krishna; Department of Biochemistry and Molecular Biology, Oklahoma State University
Additional authors:
Cao, Xiaolong; Department of Biochemistry and Molecular Biology, Oklahoma State University
He, Yan; Department of Entomology and Plant Pathology, Oklahoma State University
Hu, Yingxia; Department of Biochemistry and Molecular Biology, Oklahoma State University
Wang, Yang; Department of Entomology and Plant Pathology, Oklahoma State University
Jiang, Haobo; Department of Entomology and Plant Pathology, Oklahoma State University
Yun-Ru Chen Boyce Thompson Institute, Cornell University, Ithaca, NY Gary Blissard Boyce Thompson Institute, Cornell University, Ithaca, NY

Signal transduction pathways and their coordination are critically important for proper functioning of animal immune systems. Our knowledge on constituents of the intracellular signaling network in insects mainly comes from the genetic analyses of *Drosophila melanogaster*. To facilitate future studies of similar systems in the tobacco hornworm and other lepidopteran pests, we have identified and examined the homologous genes in the genome of *Manduca sexta*. Based on the 1:1 orthologous relationships in most cases, we consider the Toll, Imd, MAPK-JNK-p38 and JAK-STAT pathways are intact and operative in this species and so are most of the regulatory mechanisms. Similarly, cellular processes such as autophagy, apoptosis and RNA interference probably function in similar ways, because their mediators and modulators are mostly conserved in this lepidopteran species. We have annotated a total of 185 genes encoding 196 proteins, studied their domain structures and evolution, and examined their mRNA levels in tissues taken at different life stages. Such information provides a genomic perspective of the intricate signaling system in a non-drosophiline insect.

**49 - AVG**
**iBeetle: The power of phenotypic RNAi screens in arthropod genomics**

**Bucher, Gregor**; Department of Evolutionary Developmental Genetics, Georg-August-University Göttingen
Additional authors:
Johann-Friedrich-Blumenbach Institute of Zoology and Anthropology, Georg-August-University Göttingen

Ongoing sequencing is revealing the gene sets of ever more arthropods. However, it remains a major challenge to relate genes to their functions in certain biological processes. The classic approach has been forward genetic screens where relevant phenotypes are selected after random mutagenesis and the affected genes are eventually cloned. However, this phenotype based approach is limited to highly developed model species like *Drosophila melanogaster.* An alternative approach is based on the expression of genes. Here, the specific regulation of genes in certain tissues or treatments is used to hypothesize a function, which is subsequently tested by reverse genetics. However, many specifically expressed genes do not show phenotypes and relevant genes may escape detection because they are not differentially expressed.

Large scale unbiased RNAi screens combine reverse genetics with the strength of phenotypic screening. They allow identification of relevant genes irrespective of previous knowledge and independent of their expression. We present the results of the *iBeetle* screen, where the phenotypes of appr. 9,000 randomly selected genes have been determined in the red flour beetle *Tribolium castaneum*. Indeed, we identified essential functions of genes that were not differentially expressed. Further, we found novel genes involved in basic biological processes although they had extensively been screened for in *Drosophila* before. This shows that such screens are essential to reveal a comprehensive picture of gene functions. Finally, we use our data to estimate the portion of genes required for developmental processes, which had not been revealed by previous work in *Drosophila*.

**50 - AVG**
**Integrated modeling of protein-coding genes in the *Manduca sexta* genome using RNA-Seq data from the biochemical model insect**

Cao, Xiaolong; Biochemistry and Molecular Biology, Oklahoma State University
Additional authors:
Jiang, Haobo; Advisor

The genome sequence of *Manduca sexta* was recently determined using 454 technology. Cufflinks and MAKER2 were used to establish gene models in the genome assembly based on the RNA-Seq data and other species' sequences. Aided by the extensive RNA-Seq data from 50 tissue samples at various life stages, annotators over the world (including the present authors) have manually confirmed and improved a small percentage of the models after spending months of effort. While such collaborative efforts are highly commendable, many of the predicted genes still have problems which may hamper future research on this insect species. As a biochemical model representing lepidopteran pests, *M. sexta* has been used extensively to study insect physiological processes for over five decades. In this work, we assembled *Manduca* datasets Cufflinks 3.0, Trinity 4.0, and Oases 4.0 to assist the manual annotation efforts and development of Official Gene Set (OGS) 2.0. To further improve annotation quality, we developed methods to evaluate gene models in the **M**AKER2, **C**ufflinks, **O**ases and **T**rinity assemblies and selected the best ones to constitute MCOT 1.0 after thorough crosschecking. MCOT 1.0 has 18,089 genes encoding 31,666 proteins: 32.8% match OGS 2.0 models perfectly or near perfectly, 11,747 differ considerably, and 29.5% are absent in OGS 2.0. Future automation of this process is anticipated to greatly reduce human efforts in generating comprehensive, reliable models of structural genes in other genome projects where extensive RNA-Seq data are available.

**51 - AVG**
**The best of two worlds: the use of CRISPR/Cas9 and transposable elements to generate a genome-wide mutagenesis system for the Western corn rootworm, Diabrotica virgifera virgifera**

Chu, Fu-Chyun; Entomology, North Carolina State University
Additional authors:
Gorski, Stephanie; NC state university
Cardoza, Yasmin; NC state university
Lorenzen, Marcé; NC state university

The Western corn rootworm (WCR) is a major pest of maize and is notorious for rapidly adapting biochemically, behaviorally, and developmentally to a variety of control methods. Transformation-based applications such as transposon tagging, enhancer trapping, and genome-wide mutagenesis have facilitated the genetic dissection of model species, such as Drosophila melanogaster. Following this paradigm, we are developing a germline transformation system for WCR. In an effort to recapitulate an efficient piggyBac-based system from Tribolium castaneum, we are using a sequence-specific genome editing tool, CRISPR/Cas9, in conjunction with transposable elements, which can insert randomly throughout the genome. Here we report our progress towards this aim on several fronts, including results from: 1) driving marker gene (EGFP and DsRed) expression from heterologous promoters (Tc-alpha-Tubulin and Dm-heat-shock-70); 2) germline transformation (piggyBac and Minos elements); and 3) establishment of WCR “helper” strains (i.e. beetles that expresses piggyBac transposase). We also report plans and progress towards our remaining objectives: 1) site-specific insertion of a marked piggyBac element (3’UTR of a muscle actin gene), and 2) creation of a white-eyed mutant strain to enhance the usefulness of eye-specific fluorescent marker genes. Taken together this, combination of transposon- and CRISPR-based technologies is expected to bring a wide-range of transformation–based tools to bear on understanding WCR biology.

**52 - AVG**
**Elucidating the transmission of the emerging and widespread *Grapevine red blotch-associated virus***

Cieniewicz, Elizabeth; Plant Pathology and Plant-Microbe Biology, Cornell University
Additional authors:
Fuchs, Marc; Graduate Advisor

Grapevine red blotch disease is a recently recognized viral disease of *Vitis* species to which *Grapevine red blotch-associated virus* (GRBaV) is associated. GRBaV is comprised of a single stranded, circular DNA genome (3,206 nt), and is the only member of a putative new genus in the family *Geminiviridae*. GRBaV has been detected in all major grape-growing regions of the United States and in Canada, likely as a result of the dissemination of infected propagation and planting material. Of major concern to the grape industry is whether GRBaV is capable of being transmitted in vineyards. While spread has not been observed in most areas where GRBaV has been detected, within-vineyard spread via insect vector is suspected in California and in other western grape-growing regions. The primary objective of this research is to identify insect vector candidates of GRBaV. To accomplish this, a detailed census of hemipteran insects feeding on grapevines in a selected CA vineyard is being performed. Hemipteran species that are present in high numbers within the vineyard and are shown to be carriers of GRBaV will be assessed in greenhouse transmission assays for their ability to transmit GRBaV to healthy grapevines. Upon identification of an insect vector, the transmission mechanism will be characterized. We hypothesize that GRBaV will be transmitted in a circulative manner, since this mode of transmission is conserved among members of the family *Geminiviridae*. Transmission assays, in which the vector is subject to a range of feeding periods, will be useful in determining the optimal feeding times for virus acquisition and inoculation. In addition, fluorescence localization of GRBaV in the insect vector will help us to understand the mechanism of transmission, which will inform proper management strategies for control of the insect vector.

**53 - AVG**
**Host microRNAs regulate symbiont flagellar proteins in the aphid/Buchnera symbiosis**

Feng, Honglin; Department of Biology, University of Miami
Additional authors:
Wilson, Alex; Department of Biology, University of Miami

Small non-coding RNAs (snRNAs) from the reproductive manipulator Wolbachia pipientis have previously been shown to regulate host genes. Conversely, host microRNAs may target symbiont genes thereby shaping host/symbiont interactions. Here we report the identification of host microRNAs regulating symbiont genes as cross kingdom regulators that in some cases target genes that are also targeted by symbiont snRNAs. Using Illumina HiSeq, we sequenced small RNAs expressed in whole insect, gut and bacteriocyte tissues of the green peach aphid, Myzus persicae. We mapped small read RNA data to the Buchnera and Myzus genomes. Using miRDeep2, we interrogated the genomic DNA flanking mapped small reads, identifying 180 Myzus microRNAs (Myz-miRs) and 19 eukaryotic microRNA-like Buchnera snRNAs (BsnRNAs). When examining the annotated list of genes targeted by these miRNAs/snRNAs we were intrigued to find that 14 of the 23 Buchnera flagellar proteins are potential targets. While Buchnera cells are non-motile and lack a flagellum they retain a large set of flagellar genes. These flagellar genes are expressed, translated and integrated to form the flagellar hook-basal body. Buchnera cell surfaces are dotted with hundreds of hook-basal bodies that have been hypothesized to function as protein transporters centrally important to symbiont function. Of the 23 Buchnera genome encoded flagellar proteins, 14 are miRNAs/snRNAs targets. Of these 14, both Myz-miRs and BsnRNAs target six and the remaining eight are targeted by only Myz-miRs. Our results advance understanding of host/symbiont genomic integration and regulation in a model insect nutritional holobiont.

**54 - AVG**
**Breaking Down *Medea*: Using New Genome Editing Tools to Solve an Old Mystery**

Grubbs, Nathaniel; Entomology, North Carolina State University
Additional authors:
Chu, Fu-Chyun;
Lorenzen, Marcé;

*Medea* is a maternal-effect selfish genetic element which occurs naturally in certain populations of the red flour beetle, *Tribolium castaneum*. Offspring of heterozygous *Medea* mothers must inherit at least one copy of the *Medea* element in order to survive, but, aside from being associated with a 21.5-kb insertion, the cause of this phenotype remains elusive. New tools for specific excision of genomic regions have finally made possible a systematic dissection of the *Medea*-associated insertion, allowing us to pinpoint the DNA sequence responsible for the *Medea* phenotype. It is our intention to employ the RNA-guided nuclease, Cas9, to delete and replace distinct portions of the *Medea*-associated insertion with a fluorescent marker gene, then test marked females for the continued function of *Medea*. This method is expected to crack the enigma that is *Medea*, and finally unlock its potential as a gene-drive system for the control of pests and vector-borne diseases.

**55 - AVG**
**Coleopteran Cadherins: Critical to Cry Toxicity? Using CRISPR/Cas9 to determine the functional domains of a coleopteran cadherin**

Gutzmann, Nicole: Entomology, NCSU
Additional authors:
Oppert, Brenda; USDA
Lorenzen, Marce; NCSU

The bacterium *Bacillus thuringiensis* (Bt) produces a wide variety of crystal (Cry) toxins, many of which have proven to be effective biopesticides. In fact, these toxins protect over 80% of US corn and cotton. While the mode of action of Cry toxins has been established for lepidopterans, it is still contested for coleopterans. The recent literature has demonstrated that in Coleoptera, cadherins bind Cry toxins, RNAi knockdown of cadherin transcripts results in resistance, and feeding cadherin fragments to beetles increases toxicity, therefore my goal is to determine the functional domains of a coleopteran cadherin. To achieve this goal I will use the CRISPR-Cas9 gene editing system to replace a portion of the cadherin gene in a resistant beetle species (*Tribolium castaneum*) with the equivalent portion from a susceptible species (*Tenebrio molitor*). I hope to effectively knock-in Bt susceptibility and therefore identify the sequence of the toxin binding region and confirm that midgut cadherins are an integral part of the mode of action of Bt Cry toxins in Coleoptera.

**56 - AVG**
**100 Ways to Build a Beetle: How mutants can complement and inform molecular based genetics**

Haas, M. Susan; Stored Product Insect and Engineering Research Unit (SPIERU), Center for Grain and Animal Health Research (CGAHR)
Additional authors:
Oppert, Brenda; Center for Grain and Animal Health Research, Adjunct Professor - Entomology

The *Tribolium* stock collection maintained at the Center for Grain and Animal Health Research (CGAHR) in Manhattan, Kansas, has approximately 200 mutants, 43 wild-type lines and 40 transgenic lines. In addition, we have accumulated four insecticide-resistant beetle strains. The stock collection, began by Dick Beeman in the early 1980s, has been a valuable genetic resource to researchers worldwide. We provide a sample of the wide variety of unique mutant phenotypes found in the *Tribolium* stock collection. Mutations altering the external anatomy of many different areas and structures of the body are shown and represent highlights of the extensive collection. Physical mutants may have phenotypes that are difficult or impossible to obtain via RNA interference (RNAi). Combining both molecular-based and mutant-based approaches can be a very useful and powerful way to more fully elucidate gene function. This collection is a rich beetle genetic resource that is currently available upon request.

**57 - AVG**
**A genome-wide analysis of antimicrobial effector genes and their transcription patterns in Manduca sexta**

He, Yan; Entomology and Plant Pathology, Oklahoma State University
Additional authors:
Cao, Xiaolong; Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA
Li, Kai; Institute of Biological Sciences, Donghua University, Songjiang, Shanghai 310029, China
Hu, Yingxia; Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA
Chen, Yun-ru; Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
Blissard, Gary; Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
Michael R. Kanost, Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506, USA Haobo Jiang, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

Antimicrobial proteins/peptides (AMPs) are effectors of innate immune systems against pathogen infection in multicellular organisms. Over half of the AMPs reported so far come from insects, and these effectors act in concert to suppress or kill bacteria, fungi, viruses, and parasites. In this work, we have identified 86 AMP genes in the Manduca sexta genome, most of which seem likely to be functional. They encode 15 cecropins, 6 moricins, 6 defensins, 3 gallerimycins, 4 X-tox splicing variants, 14 antifungal peptides, 15 whey acidic protein homologs, 11 attacins, 1 gloverin, 4 lebocins, 6 lysozyme-related proteins, and 4 transferrins. Some of these genes (e.g. attacins, cecropins) constitute large clusters, likely arising after rounds of gene duplication. We compared the amino acid sequences of M. sexta AMPs with their homologs in other insects to reveal conserved structural features and phylogenetic relationships. Expression data showed that many of them are synthesized in fat body and midgut during the larval-pupal molt. Certain genes contain one or more predicted κB binding sites and other regulatory elements in their promoter regions, which may account for the dramatic mRNA level increases in fat body and hemocytes after an immune challenge. Consistent with these strong mRNA increases, many AMPs become highly abundant in the larval plasma at 24 h after the challenge, as demonstrated in our previous peptidomic study. Taken together, these data suggest the existence of a large repertoire of AMPs in M. sexta, whose expression is up-regulated via immune signaling pathways to fight off invading pathogens in a coordinated manner.

**58 - AVG**
**Phylogenetic analysis and expression profiling of the pattern recognition receptors: insights into molecular recognition of invading pathogens in *Manduca sexta***

He, Xuesong; Entomology and Plant Pathology, Oklahoma State University

Pattern recognition receptors (PRRs) detect microbial pathogens and trigger innate immune responses. Previous biochemical studies have elucidated the physiological functions of eleven PRRs in *Manduca sexta* but our understanding of the recognition process is still limited, lacking genomic perspectives. While 34 C-type lectin-domain proteins and 13 Toll-like receptors are reported in the companion papers, we present here 122 other putative PRRs identified through the genome annotation. These include 78 leucine-rich repeat (LRR) proteins, 14 peptidoglycan recognition proteins, 6 EGF/Nim-domain proteins, 5 β-1,3-glucanase-related proteins, 4 galectins, 4 fribrinogen-related proteins, 3 thioester-containing proteins, 5 immunoglobulin-domain proteins, 2 hemocytins, and 1 Reeler. Sequence alignment and phylogenetic analysis reveal the evolution history of a diverse repertoire of proteins for pathogen recognition. While functions of insect LRR proteins are mostly unknown, their structure diversification is phenomenal: 22 extracellular LRR proteins have a signal peptide; 18 intracellular ones lack that; 38 LRR-TMPs with a transmembrane region may be non-Toll receptors on the surface of cells. Expression profiles of the 122 genes in the 52 tissue samples reflect complex regulation in various developmental stages and physiological states, some by Rel family transcription factors via κB motifs in the promoter regions. This

collection of information is expected to facilitate future biochemical studies detailing their respective roles in this model insect.

**59 - AVG**
**Structural features, evolutionary relationships, and transcriptional regulation of C-type lectin-domain proteins in Manduca sexta**

Hu, Yingxia: Biochemistry and Molecular Biology, Oklahoma State University
Additional authors:
Rao, XiangJun; Anhui Agricultural University
Cao, Xiaolong; Oklahoma State University
He, Yan; Oklahoma State University
Zhang, Xiufeng; Oklahoma State University
Chen, Yunru; Cornell University
Gary Blissard: Cornell University Michael R. Kanost: Kansas State University XiaoQiang Yu: University of Missouri-Kansas City *Haobo Jiang: Oklahoma State University *Corresponding author

C-type lectins (CTLs) are a large family of $Ca^{2+}$-dependent carbohydrate-binding proteins recognizing various glycoconjugates and functioning primarily in immunity and cell adhesion. We have identified 34 CTLDP (for CTL-domain protein) genes in the Manduca sexta genome, which encode proteins with one to three CTL domains. CTL-S1 through S9 (S for simple) have one or three CTL domains; immulectin-1 through 19 have two CTL domains; CTL-X1 through X6 (X for complex) have one or two CTL domains along with other structural modules. Nine simple CTLs and seventeen immulectins have a signal peptide and are likely extracellular. Five complex CTLs have both an N-terminal signal peptide and a C-terminal transmembrane region, indicating that they are membrane anchored. Immulectins exist broadly in Lepidoptera and lineage-specific gene duplications have generated three clusters of fourteen genes in the M. sexta genome, thirteen of which have similar expression patterns. In contrast to the family expansion, CTL-S1~S6, S8, and X1~X6 have 1:1 orthologs in at least four lepidopteran/dipteran/coleopteran species, suggestive of conserved functions in a wide range of holometabolous insects. Structural modeling suggests the key residues for $Ca^{2+}$-dependent or independent binding of certain carbohydrates by CTL domains. Promoter analysis identified putative κB motifs in eighteen of the CTL genes, which did not have a strong correlation with immune inducibility in the mRNA or protein levels. Together, the gene identification, sequence comparisons, structure modeling, phylogenetic analysis, and expression profiling establish a solid foundation for future studies of M. sexta CTL-domain proteins.

**60 - AVG**
**Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in Manduca sexta**

Jiang, Haobo; Entomology and Plant Pathology, Oklahoma State University
Additional authors:
Cao, Xiaolong; Oklahoma State University
He, Yan; Oklahoma State University
Hu, Yingxia; Oklahoma State University
Zhang, Xiufeng; Oklahoma State University
Wang, Yang; Oklahoma State University
Zhen Zou, Institute of Zoology, Chinese Academy of Sciences Yunru Chen, Cornell University Gary W. Blissard, Cornell University Michael R. Kanost, Kansas State University

Serine protease (SP) and serine protease homolog (SPH) genes in insects encode a large family of proteins involved in digestion, development, immunity, and other processes. While 68 digestive SPs and their close homologs are reported in a companion paper (Kuwar et al., 2015), we have identified 125 other SPs/SPHs in Manduca sexta and studied their structure, evolution, and expression. Fifty-two of them contain cystine-stabilized structures for molecular recognition, including clip, LDLa, Sushi, Wonton, TSP,

CUB, Frizzle, and SR domains. There are nineteen groups of genes evolved from relatively recent gene duplication and sequence divergence. Thirty-five SPs and seven SPHs contain 1, 2 or 5 clip domains. Multiple sequence alignment and molecular modeling of the 54 clip domains have revealed structural diversity of these regulatory modules. Sequence comparison with their homologs in Drosophila melanogaster, Anopheles gambiae and Tribolium castaneum allows us to classify them into five subfamilies: A are SPHs with 1 or 5 group-3 clip domains, B are SPs with 1 or 2 group-2 clip domains, C, D1 and D2 are SPs with a single clip domain in group-1a, 1b and 1c, respectively. We have classified into six categories the 125 expression profiles of SP-related proteins in fat body, brain, midgut, Malpighian tubule, testis, and ovary at different stages, suggesting that they participate in various physiological processes. Through RNA-Seq-based gene annotation and expression profiling, as well as intragenomic sequence comparisons, we have established a framework of information for future biochemical research of nondigestive SPs and SPHs in this model species.

**61 - AVG**
**Virulent polydnaviral genes originated from an endoparasitoid wasp genome of *Cotesia plutellae***

Kim, Yonggyun; Bio-Sciences, Andong National University

An endoparasitoid wasp, *Cotesia plutellae*, possesses a symbiotic virus, *C. plutellae* bracovirus (CpBV). This symbiotic relationship between eukaryote and virus is required for parasitism. CpBV genome consists of encapsidated genome (ECG) and unencapsidated genome (UECG) evident during viral replication. ECG consists of virulence factors, which alter various physiological processes of parasitized lepidopteran host. It helps the survival and growth of host wasp in the parasitized host. In contrast, UECG has been presumed to be involved in the viral replication by encoding coat proteins and replication-assisting factors. To determine the origin of ECG, *C. plutellae* genome needed to be determined. An isofemale line was inbred with fifteen generations and the resulting haploid males were used to sequence a whole genome using HiSeq2000 with 116x depth. The whole genome size estimation from the assembled contigs was 241,606,599 bp. Subsequent scaffolding and gap filling procedure resulted in 2,897 scaffolds with 389,450 bp in N50. From this whole genome, 21,427 genes were predicted. A viral lectin gene encoded in CpBV shared a high homology with one of four lectin genes encoded in *C. plutellae*. A viral histone H4 also exhibited a high homology with a histone gene encoded in *C. plutellae*. Five viral ankyrin repeat genes shared homology with an IkB gene of *C. plutellae*. A host PTP gene (PTP15) may be a common origin of 36 viral PTP genes encoded in CpBV. These results suggest that virulent genes encoded in CpBV are originated from host wasp genome via horizontal gene transfer.

**62 - AVG**
**Development of RNA Aptamers that block CLas transmission by the Asian citrus psyllid**

Kruse, Angela; Plant Pathology and Plant-Microbe Biology, Cornell University/Boyce Thompson Institute for Plant Research
Additional authors:
Ozer, Abdullah; Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY
Sturgeon, Kasie; Subtropical Insects and Horticulture Research Unit, U.S. Horticultural Research Laboratory, Fort Pierce, FL
Warwick, Erica; Subtropical Insects and Horticulture Research Unit, U.S. Horticultural Research Laboratory, Fort Pierce, FL
Lis, John; Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY
Shatters, Robert; Subtropical Insects and Horticulture Research Unit, U.S. Horticultural Research Laboratory, Fort Pierce, FL
Michelle Cilia, Section of Plant Pathology and Plant-Microbe Biology, School of Integrative Plant Sciences, Cornell University, Ithaca, NY; Boyce Thompson Institute for Plant Research, Ithaca, NY; Emerging Pests and Pathogens Research Unit, Robert W. Holl

Citrus greening disease (also known as Huanglongbing, HLB) is associated with plant infection by the gram-negative bacterium *Candidatus* Liberbacter asiaticus (CLas). The Asian citrus psyllid (ACP) transmits CLas from plant to plant in a circulative manner. Circulative transmission involves the movement of CLas across multiple cellular barriers in the insect, including the gut, the hemolymph (blood) and salivary glands. Circulative interaction between vector and bacterium typically relies on highly co-evolved protein-protein interactions that initially allow the bacterium to invade the insect gut barrier, and ultimately allow the CLas to enter the salivary gland and become transmissible. Access to diseased plants as nymphs is required for efficient transmission. Control strategies are being sought based on the idea that interdiction molecules that block these protein-protein interactions may interfere with the bacterium's ability to cross the nymph gut membrane and thus block acquisition. We hypothesize that RNA aptamers can be identified that bind specifically to the psyllid nymph gut and interfere with protein-protein interactions necessary for efficient acquisition of CLas. RNA aptamers are single-stranded molecules of RNA whose secondary structure imparts physicochemical properties that allow target specific interactions based on variations in RNA sequence. These targets can include proteins, glycans, nucleic acids, and even small molecules. Aptamers provide the specificity and affinity for targets similar to that of antibodies, but they are non-immunogenic and can be synthesized *in vitro* and *in vivo*. Aptamers that bind psyllid nymph guts will be selected in a process called SELEX (Systemic Evolution of Ligands by Exponential Enrichment) in which a library containing 5 x$10^{14}$ unique aptamers will be hybridized to homogenized gut tissue for several cycles. With each cycle, the library of aptamers becomes less diverse but has higher affinity for the gut tissue. Aptamers binding to the gut target will be sequenced using next generation sequencing technology, and their binding activity to psyllid tissues will be characterized *in vitro* and *in vivo*. In parallel, the same selection process will be used to identify aptamers binding to the psyllid stylet sheaths. Psyllids deposit a polymeric sheath during the feeding process, and the sheath is thought to protect the stylet against plant defenses so the insect can have prolonged feeding times on the phloem. Since sheath polymerization is necessary for successful establishment of insect feeding, we hypothesize that aptamers that bind the sheaths may inhibit polymerization and therefore may interfere with transmission of CLas.

**63 - AVG**
**Identification of genes involved in the chemical warfare of the red flour beetle Tribolium castaneum**

Lehmann, Sabrina; Developmental Biology, Johann-Friedrich-Blumenbach-Institute for Zoology and Anthropology, Georg-August-University Göttingen, Germany
Additional authors:
Li, Jianwei; Center for Medical, Agricultural, and Veterinary Entomology, Agricultural Research Service, U.S. Department of Agriculture, Gainesville, FL, USA
Wimmer, Prof. Ernst A.; Johann-Friedrich-Blumenbach-Institute for Zoology and Anthropology, Developmental Biology, Georg-August-University Göttingen, Germany

Chemical warfare is the most common defence strategy in the insect world. A broad range of coleopteran beetles react to predators, invaders and parasitic microbes with the release of toxic and repellent substances. Some leaf beetles are able to sequester their defensive chemicals from food plants, whereas the red flour beetle Tribolium castaneum produces bacteriostatic p-benzoquinones and 1-alkenes de novo in specialized secretory organs – referred to as odoriferous defensive stink glands –, which are constructed pairwise in the thorax as well as in the most posterior part of the abdomen. A stink gland consists of maximal two different types of secretion producing cells and a reservoir to store the secretions ready to use. Although the morphology of the secretion producing cells has been studied in detail in the past, to date only little is known about the biochemical processes of secretion production and their arrangement within the cell to avoid self-intoxication of the beetle. Here we present a subset of genes that are involved in the beetle's biosynthesis of p-benzoquinones with a special focus on the phenoloxidase Laccase2 (Lac2). Potential candidates have been selected from (I) a genome-wide RNAi knockdown screen (iBeetle) and (II) stink gland transcriptome data. Lac2 has been selected from a list of potential phenoloxidases in Tribolium castaneum based on its specifically high expression in the stink glands. It is expressed in a subgroup of secretory stink gland cells and the gland secretion composition is strikingly altered in RNAi-mediated knockdown beetles. Enzymatic activity assays on freshly dissected stink gland tissue of wild type and RNAi-mediated knockdown situations emphasize the involvement of Lac2 in p-benzoquinone biosynthesis. Based on our data we suppose that Lac2 is the enzyme responsible for the oxidation of p-benzoquinone precursors in the stink glands of the red flour beetle.

**65 - AVG**
**Eaten alive: lessons from the venom gland transcriptome of the wasp, *Anisopteromalus calandrae***

Perkin, Lindsey, Agricultural Research Service, USDA
Additional authors:
Friesen, Ken; USDA ARS Center for Grain and Animal Health Research
Flinn, Paul; USDA ARS Center for Grain and Animal Health Research
Oppert, Brenda; USDA ARS Center for Grain and Animal Health Research

The wasp *Anisopteromalus calandrae* is a small (2.25 mm) ectoparasitoid that targets pest beetle larvae within stored products. A female wasp drills into a wheat kernel and injects a venom cocktail with her ovipositor into the beetle larva, paralyzing and suppressing larval development. She then deposits an egg, and the parasitoid larva emerges and feeds on the immobilized host larva. *A. calandrae* has been used as a biological control agent of some stored product insects, but components of the venom are unknown, such as how they immobilize the prey but keep the larva alive for the development of their offspring. To address these questions, we dissected female wasps and collected venom and associated glands. We sequenced the transcriptome to identify venom and venom-related transcripts. Sequences aligned into 45,433 contigs, 25,726 of which had BLAST hits. The majority of hits were to *Nasonia vitripennis* as well as other bees, wasps, and ants. Gene ontology grouped sequences into eleven molecular functions, among which binding and catalytic activity were overrepresented. We highlight several interesting sequences with potential for pest control, such as those encoding metalloproteases, superoxide dismutases (SOD), and Kazal-type (KPI) and small serine protease inhibitors (SPI). Metalloprotease transcripts were abundant in the *A. calandrae* venom gland transcriptome and are also found in venomous snakes, where the enzymes interfere with blood coagulation and break down the extracellular matrix of

their victims. In other parasitoid wasps, metalloproteases affect host development, preventing molting between larval instars. SODs provide defense against reactive oxygen species in many organisms. In the wasp venom, SODs may be one of the processes that reduce damage to host tissue during developmental arrest. Lastly, we point out KPIs and small SPI transcripts. These protease inhibitors are found in the hemolymph of arthropods and have been described in many types of insect venom. They likely protect host larvae from bacterial or fungal infection while in developmental arrest. This work is the first to characterize *A. calandrae* venom and is being studied to provide alternatives in pest control and other uses.

**66 - AVG**
**Cathepsin B activity modulates polerovirus transmission efficiency by aphids**

Pinheiro, Patricia; Entomology, Cornell University
Additional authors:
Cilia, Michelle; Cornell University

Myzus persicae is an efficient vector of Potato leafroll virus (PLRV); however, PLRV transmission efficiency is significantly reduced when aphids are reared on turnip as compared to Physalis floridana. Turnip-reared aphids are also larger and have a fitness advantage. To glean insight into the molecular differences between aphids reared on turnip or P. floridana, we identified differentially expressed proteins by 2-D Difference Gel Electrophoresis (DIGE) coupled with mass spectrometry as well as a 1-D, reverse phase, nanoscale liquid chromatograph separation coupled to high resolution mass spectrometry and spectral counting. The M. persicae genome is sequenced but not annotated; therefore we developed a new database search strategy to rapidly annotate the expressed M. persicae proteins. Both approaches revealed that the cysteine protease cathepsin B was up-regulated in aphids reared on turnip. DIGE analysis also revealed multiple size and charge cathepsin B isoforms were differentially expressed. Three distinct cathepsin B proteins derived from three cathepsin B-encoding transcripts were identified. The expression and up-regulation of each protein isoform in M. persciae reared on turnips was validated using selected reaction monitoring mass spectrometry. Using the cathepsin inhibitor E64, we tested the effect of cathepsin B inhibition on PLRV transmission in aphids reared on either turnip or P. floridana. After feeding aphids on a diet with E64, PLRV transmission from aphids reared on P. floridana was decreased in an E64 dose-dependent manner. In contrast, after feeding aphids on a diet with E64, aphids reared on turnip showed an increase in transmission efficiency. Cathepsin activity assays in both cohorts of aphids shows that although the protein level is higher in aphids reared on turnip, cathepsin is less than half in these insects as compared to those reared on P. floridana. By varying enzyme and substrate concentration, we show that the enzyme activity is regulated differently in the insects on the different hosts, suggesting the presence of a co-factor, possibly plant derived. Contributions from the plant (or animal) host on the vectoring capacity of an insect vector is largely unexplored and undocumented in the vector biology literature. Here we report the first example where an insect vector modulates its transmission efficiency according to interactions with its host plants and show the mechanism of this modulation is due to the activity of cathepsin B.

# Medical Vector Genomics

**67 - MVG**
**RNA-seq analysis reveals mechanisms associated with tsetse fly-symbiont dynamics during larval development**

Benoit, Joshua; Biological Sciences, University of Cincinnati
Additional authors:
Vigneron, Aurelien; Yale University
Wu, Yineng; Yale University
Aksoy, Serap; Yale University
Weiss, Brian; Yale University

Tsetse flies harbor three distinct symbionts that are transmitted from the mother to her offspring. These symbionts are critical to tsetse fly physiology, in specific have been documented to play a role in immune system development and adult reproduction. In this study, we utilized RNA-seq analyses followed by functional examination to determine the effects of symbiont removal on larval development and progeny health. These analyses revealed that there are over 2500 genes with differential expression after symbiont removal. In specific, there is an enrichment for genes associated with B vitamin metabolism, larval development and chitin interactions. The reduction of gene associated with B vitamin metabolism is likely due to reduced B vitamins concentraitons that are normally produced by the obligate symbiont, Wigglesworthia, and a reduction in chitin-associated genes likely accounts for impaired peritrophic matrix development previously documented. In addition to these factors, we noted 10-fold increase in expression for two odorant-binding proteins (obps), obp6 and obp11. Knockdown of obp6 can be accomplished by injection of siRNAs into the mother during lactation. Offspring with reduced levels of obp6 yield adults with a phenotype of reduced or dysfunctional crystal cells, which act in the initiation of the melanization cascade via the release of prophenoloxidase. These results provide mechanisms underlying immune development and symbiont dynamics in tsetse flies and implicate odorant-binding proteins in the process.

**68 - MVG**
**Toll Signaling in *Anopheline* Mosquitoes**

Davidson, Victoria; Biology, Kansas State University
Additional authors:
Michel, Kristin; Kansas State University

The Toll pathway is a central regulator of immunity in insects and acts to induce transcription of genes that limit infection by viruses, bacteria, fungi, and protozoans. Although this pathway is conserved in insects, the full molecular makeup of this signaling cascade in mosquitoes is still unknown. We aim to identify putative immune-functioning Toll-like receptors (TLRs), responsible for initiating this intracellular signaling cascade, in the African malaria vector, *Anopheles gambiae*. Phylogenetic analysis highlighted putative orthologs of immune-functioning *Drosophila melanogaster* genes, *Toll* and *Tehao,* in *A. gambiae*. These genes have undergone duplication events in *A. gambiae* leading to the presence of four distinct genes: *Toll1A, Toll1B, Toll5A,* and *Toll5B.* However, in the closely-related species *Anopheles stephensi*, Toll1B and Toll5B are absent from the draft genome, suggesting that the phylogenetic history of putative immune-acting TLRs in *Anopheles* spp. is more complex than initially described. We present here our efforts to utilize the entomopathogenic fungus *Beauveria bassiana* strain I93-825 to probe the Toll pathway to identify immune-functioning TLRs in *A. gambiae.* The cuticle of 2-4 day old female adults was exposed by direct contact to an oil suspension of fungal spores ($1.24 \times 10^9$ spores/mL) applied to coated filter paper substrate. To test for the ability of this infection to activate the Toll pathway, the expression of key regulators REL1 (transcription factor) and Cactus (inhibitor) were knocked down by RNAi (injection of 69nL dsRNA [3µg/µL]). Overall, we found that infection with *B. bassiana* decreased significantly the median survival of *A. gambiae* by 10.17 ± 1.364 days compared to controls (T test $P = 0.0033$). Knockdown of the

Toll pathway inhibitor, cactus, increased survival after infection, while knockdown of the transcription factor, REL1, decreased survival. RT-qPCR analysis revealed a 2-fold transcriptional upregulation of Cactus 4 days post *B. bassiana* exposure (2-Way ANOVA, $P = 0.0258$). The data show that *B. bassiana* can successfully probe this signaling cascade in *A. gambiae*, as knockdown of critical pathway components by RNAi alters survivorship to fungal challenge. Future studies aim to utilize this assay to search for putative immune-acting Toll-like receptors, as the receptor(s) involved in the activation of the Toll pathway in *A. gambiae* is currently unknown. We are grateful to Dr. Matt Thomas for providing *B. bassiana* strain I93-825 and for assistance with exposure design. Partial funding provided by the National Institutes of Health through R01-AI095842.

**69 - MVG**
**Heritable CRISPR/Cas9-mediated genome editing in Aedes aegypti**

Dong, Shengzhang; Department of Veterinary Pathobiology, University of Missouri
Additional authors:
Lin, Jingyi; Department of Veterinary Pathobiology, College of Veterinary Medicine, 209C Connaway Hall, University of Missouri, Columbia, MO 65211
Held, Nicole; Department of Veterinary Pathobiology, College of Veterinary Medicine, 209D Connaway Hall, University of Missouri, Columbia, MO 65211
Clem, Rollie; Molecular, Cellular, and Developmental Biology Program, Division of Biology, 116 Ackert Hall, Kansas State University, Manhattan, KS 66506.
Passarelli, Lorena; Molecular, Cellular, and Developmental Biology Program, Division of Biology, 116 Ackert Hall, Kansas State University, Manhattan, KS 66506
Franz, Alexander; Department of Veterinary Pathobiology, College of Veterinary Medicine, 303 Connaway Hall, University of Missouri, Columbia, MO 65211

Targeted genome editing is a powerful method to study gene function in any given organism. Thus far, two tools have been used for targeted gene disruption in the yellow fever mosquito, *Aedes aegypti*: zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN). These tools involve the use of engineered DNA binding proteins to introduce target site-specific double-strand breaks in the genomic DNA. However, both tools are complicated to assemble using standard molecular techniques. The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/ CRISPR-associated sequence 9) system has previously been utilized for genome editing in a number of organisms including insects and promises to be a true 'do-it-yourself' genome editing tool.
Recently, we employed the CRISPR/Cas9 system to disrupt a marker gene in transgenic *Ae. aegypti*. We injected *in vitro* transcribed Cas9 mRNA and two sgRNAs targeting different regions of the enhanced cyan fluorescent protein (ECFP) gene into embryos of a transgenic mosquito line expressing both Dsred and ECFP from the eye tissue-specific 3xP3 promoter in separated but tightly linked expression cassettes. In outcrossed G1 larvae, successful ECFP knockout was determined by identifying individuals showing only DsRed expression but no longer ECFP expression in their eyes. We recovered four different G1 pools (5.5% knockout efficiency) with this phenotype. PCR amplification, cloning, and sequencing of PCR amplicons revealed indels in the ECFP target gene ranging from 2-27 nucleotides. These results showed for the first time that CRISPR/Cas9 mediated genome editing is achievable in *Ae. aegypti*, paving the way for further functional genomics related studies in this mosquito species.

**70 - MVG**

**Genome improvement using third-gen sequencing**

Emrich, Scott; Computer Science and Engineering, University of Notre Dame
Additional authors:
Steele, Aaron; University of Notre Dame
Lanc, Irena; University of Notre Dame
Koren, Serguei; National Biodefense Center
Phillippy, Adam; National Biodefense CenterVirginia Tech
Sharakov, Igor; Virginia Tech
Nora Besansky University of Notre Dame

Many arthropods have difficult to assemble genomes, especially biomedically relevant vectors. As an objective of multiple related consortia, An. gambiae is unique in that it has been sequenced twice with Sanger methods, once with Illumina and PacBio WGS, and was a key member of a large phylogenetic cluster. We hypothesized that low coverage but longer "third-gen" read data would successfully improve prior assemblies at much lower cost. Here, we first assess advantages of each individual sequencing method using all available data, and report hybrid techniques that we are developing at Notre Dame for combining independent assemblies (i.e., metaassembly) and for scaffolding assemblies using limited third-gen sequencing data. We will show that our new scaffolding tool, pbSandwich, outperforms alternatives with low coverage PacBio data, and validate these overall results with a complete annotation, an optical map, and whole-genome alignments to six very closely related genomes. Finally, we will conclude with preliminary results of new hybrid genome assembly framework to another bio medically important mosquito vector, An. funestus.

**71 - MVG**
**Pathogen Acquisition in *Dermacentor andersoni* Ticks is Dependent on Microbiome Composition**

Gall, Cory; Veterinary Microbiology and Pathology, Washington State University
Additional authors:
Scoles, Glen; USDA
Reif, Katie; USDA
Noh, Susan; USDA
Brayton, Kelly; Veterinary Microbiology and Pathology

Ticks are ecoparasitic arthropods that feed on vertebrates. Ticks are the first and second most important vector of disease for livestock and humans, respectively. The Rocky Mountain wood tick, *Dermacentor andersoni*, vectors several human and animal pathogens, including *Anaplasma marginale*, which causes bovine anaplasmosis, the most widespread tick-borne bacterial pathogen of cattle worldwide. Like other arthropods, multiple microorganisms colonize ticks to make up the microbiome, with pathogenic species accounting for only a small proportion. Although *D. andersoni* is a medically important vector, little is known regarding the bacterial microbiome and it's role in vector competence. We hypothesize **that the bacterial microbiome of *D. andersoni* has a significant influence on pathogen susceptibility.** To address our hypothesis, we identified two populations of *D. andersoni* ticks with differing pathogen susceptibility. We characterized the bacterial microbiome before and after disruption using PacBio sequencing. Cohorts of ticks were then exposed to either *A. marginale* or *F. novicida* to determine if microbiome alteration influenced pathogen susceptibility. Our results demonstrated that pathogen level, but not infection rates, of both *A. marginale* and *F. novicida* differed between treatment groups. The microbiome alteration resulted in shifts of the proportions of several symbionts in both populations, as well as demonstrated specific endosymbiont-pathogen relationships. The endosymbiont *Rickettsia bellii*, which increased in proportion and quantity in the microbiome post-alteration, demonstrated a negative correlation with *A. marginale* pathogen load; whereas decreased proportion of *Francisella* endosymbionts resulted in lower *F. novicida* infection levels, demonstrating a positive relationship between related endosymbiont and pathogen.

**72 - MVG**
**The University of Maryland Insect Transformation Facility: A resource for Insect Genetic Modification**

Harrell, Robert; IBBR, University of Maryland
Additional authors:
Aluvihare, Channa; University of Maryland
O'Brochta, David; University of Maryland

The University of Maryland Insect Transformation Facility (UM-ITF) has been providing the insect functional genomics community access to transgenic and non-transgenic genome altering technologies for the past 9 years. The mission of the UM-ITF is to aid researchers in the creation of genetically modified insects through; fee for service microinjection of insects with developed genome altering protocols, collaboration to develop genome altering protocols for insects without such protocols, training for researchers who are interested in employing these technologies and access to “State-of-the-art” equipment for these purposes.

In the past year the UM-ITF has been working hard to improve the services it provides. Now in addition to the reliable transgenesis service the facility provides for *Aedes aegypti* and *Anopheles stephensi*, the facility now has standardized services for production of transgenic *Anopheles gambiae*. Currently, the UM-ITF project success rate for producing transgenic *Anopheles gambiae* is 88% and is rapidly approaching the 97% success rate of *Aedes aegypti* and *Anopheles stephensi* projects. This makes the UM-ITF the only facility in the world that offers fee for service genetic modification of *Anopheles gambiae*.

The UM-ITF has also built a successful track record with clients utilizing the CRISPR/Cas-9 system in the creation of *Aedes aegypti* mutants, see Kistler et. al. 2014. The facility is also working to get the CRISPR/Cas-9 system into Anopheles and has already successfully produced *Anopheles gambiae* mutants through this system.

In an effort to further expand access to the technology of insect genomic modification the UM-ITF is actively involved with training and workshops devoted to teaching researchers from the Insect functional genomics community. The UM-ITF will take part in the International Short Course on Insect Transgenesis: Shanghai, China March 16-20, 2015 and the IGTRCN Technical Workshop: Rockville, Maryland August 17-21, 2015.

It is through efforts such as these that the UM-ITF is continually working to improve the technology of insect genetic modification and helping to make this technology more accessible to the insect functional genomics community.

**73 - MVG**
**The effect of bacterial challenge on the midgut physiology and development of the sand fly Lutzomyia longipalpis**

Heerman, Matthew; Entomology, Kansas State University
Additional authors:
Weng, Ju Lin; Kansas State University
Hurwitz, Ivy; University of New Mexico School of Medicine
Ramalho-Ortigao, Marcelo; Kansas State University

The use of genetically modified (GM) commensal bacteria has gained significant momentum and interest has grown over their applications to control disease vectors and agricultural pests. Sand flies vector a number of pathogens, with *Leishmania* being the most important. Our group is interested in the use of GM bacteria in the control of sand flies and in reducing *Leishmania* transmission in endemic areas. It has been

shown that bacterial infection in the insect gut lead to activation of innate responses including epithelial homeostasis pathways that culminate in a restructuring of the midgut epithelium. Interestingly, for holometabolous insects such as sand flies, energy conservation during late larva (L4) and pupa stages is critical for successful metamorphosis and survival. Here, we infected sand fly L3 larvae with GFP-expressing Gram+ *Bacillus subtilis* and Gram- *Pantoea agglomerans*. Bacteria distribution within the larval gut is driven in part by pH, accompanied by differential cellular apoptosis and restructuring of the midgut. Further, contrary to what had been previously suggested, these two bacteria do not survive pupation and/or metamorphosis, and bacteria loss is likely due to the midgut rearrangement that take place during these processes. Gene expression analyses suggest that upregulation of the negative regulator Pirk, beginning at 12h post infection (pi), with concomitant down regulation of the effector molecules attacin and IMPer may favor autophagy over an immune response even though insects were fed with bacteria. This is supported by the downregulation of USP36 at 24h pi. We are currently assessing the effect of Pirk knockdown on survival, ability to complete metamorphosis, and changes in immune response after bacterial challenge. A switch to autophagy would allow larvae to conserve and recycle nutrients critical to complete metamorphosis. This study provides novel insights on the effects of bacteria in the gut of insects, and a framework for paratransgenesis appliocations to control sand fly vectors.

**74 - MVG**

**On the necessity of species delimitation studies of Southern American Triatoma (Hemiptera: Reduviidae), Chagas disease vectors**

Justi, Silvia; Department of Biology, University of Vermont
Additional authors:
Galvão, Cleber; Instituto Oswaldo Cruz, FIOCRUZ
Schrago, Carlos; Universidade Federal do Rio de Janeiro

The family Reduviidae (Hemiptera: Heteroptera), or assassin bugs, is among the most diverse families of the true bugs, with more than 6,000 species. The subfamily Triatominae (kissing bugs) is noteworthy not simply because it is the only subfamily of the Reduviidae whose members feed on vertebrate blood but particularly because all 147 described members of the subfamily are potential Chagas disease vectors. The genus *Triatoma* is the most diverse of the subfamily, with more than half of the known diversity distributed throughout South America. Triatominae phylogenies show that most Southern American *Triatoma* form a single clade, however the relationship between species assigned to the *Triatoma* subcomplexes are not well resolved. The lack of phylogenetic resolution could be an indicator of recent divergence time and, based on the available fossils and molecular genetic markers for Reduviidae (16S, 18S, 28S and Wingless) we calculated that this *Triatoma* clade diversified no earlier than 20mya. Taking into account the number of lineages comprised by this clade and the necessary number of generations needed to reach reciprocal monophyly in the whole genome, the lack of phylogenetic resolution is expected. For that reason, and more importantly, because some of those lineages are important Chagas disease vectors, the identification of independent evolving lineages is a matter of Public Health. Once the genomes from these lineages are available, we will be able to apply appropriate approaches concerning species delimitation and independently evolving lineages, and identify characteristics specific for the most relevant vector lineages to better manage vector control.

**75 - MVG**

**Transcriptome of the lone star tick, *Amblyomma americanum*, revealing the molecular interaction between the vector and the pathogen *Ehrlichia chaffeensis***

Kim, Donghun; Entomology, Kansas State University
Additional authors:
Herdon, Nic; Division of Biology, Kansas State University
Jaworski, Deborah; Center of Excellence for Vector-Borne Diseases, Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University
Ganta, Roman; Center of Excellence for Vector-Borne Diseases, Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University
Park, Yoonseong; Department of Entomology, Kansas State University

Ticks are obligatory ectoparasites of many vertebrates and transmit pathogens causing diseases such as Heartland virus and Ehrlichiosis. The lone star tick, *Ambloyomma americanum* L., is the primary vector of *Erhlichia chaffeensis*, which causes human monocytic Ehrlichiosis. We aimed to investigate the genomic levels of gene regulation underlying the processes of pathogen acquisition and development of immunity towards the pathogen. We tested six experimental groups, including male or female *A. americanum, which were* either *E. chaffeensis* positive, negative, or pathogen free. One hundred cycles of single direction sequencing were performed in the Illumina HiSeq 2500 for six libraries with tagging. Raw sequence reads (more than 209 million) were trimmed and filtered based on minimum quality score (Q-value >30) and size (>40nt) for *de novo* assembly. Assembly using Trinity pipeline produced 140,574 contigs from trimmed and filtered sequence reads (about 117 million reads, 56% of raw data). For quality control of the *de novo* assembly of transcripts, we filtered out the sequences for mitochondrial, *E. chaffeensis*, and transposable elements, and tested for contig redundancy and gap separations of the assembled sequences. Analyses of differential expression by using RSEM and edgeR for 61,802 contigs were followed by Blast2GO analyses for annotations of contigs and enriched-gene ontology (GO) term analyses in pairwise comparisons of the libraries. Pathogen exposed female ticks expressed ion transporting genes at higher levels than pathogen free female ticks. Further investigation of pathogen induced gene expression would provide better understanding of pathogen-vector interactions, which will allow us to prevent of pathogen transmission by disrupting the interactions between pathogens and ticks.

**76 - MVG**

**Using targeted next-generation sequencing to characterize genetic variation in insecticide resistance genes in Culex pipiens complex mosquitoes**

Kothera, Linda; Division of Vector-Borne Diseases, Centers for Disease Control and Prevention
Additional authors:
Savage, Harry; Centers for Disease Control and Prevention

Insecticide resistance is an ongoing concern in areas where chemical pesticides are used to control vector mosquitoes. Moreover, while generally thought to be the result of multiple genes, resistance is often studied as a single gene phenomenon. Next generation sequencing technologies allow the simultaneous examination of large amounts of sequence data, and further refinements permit researchers to select genes of interest and sequence them in several individuals simultaneously in one multiplexed assay. Our study involved developing a multiplexed panel to sequence insecticide resistance related genes in *Culex pipiens* complex mosquitoes, a vector of several arboviral diseases including West Nile disease. We used Life Technologies' Ampliseq process, which uses pools of primer pairs to amplify targets in a genome. The first phase involved a pilot project with 28 genes in a small number of phenotypically resistant and susceptible individuals. We are in the process of adding genes to the panel, with the goal of having approximately 100 genes sequenced per individual, per run. Data analysis will involve determining whether associations exist between known resistance mutations and other detected SNP variation, and whether certain genotypes are more likely to be associated with resistance, even when well-known mutations such as *kdr* and ACE1 are not present. In order to make our results useful to local vector control programs, we aim to develop diagnostic assays to test for relevant mutations revealed by data analyses.

**77 - MVG**
**Leveraging genomic polymorphism for more informative *de novo* assemblies**

Love, R. Rebecca; Department of Biological Sciences, University of Notre Dame
Additional authors:
Weisenfeld, Neil I.; Broad Institute of MIT and Harvard
Jaffe, David B.; Broad Institute of MIT and Harvard
Besansky, Nora J.; University of Notre Dame
Neafsey, Daniel E.; Broad Institute of MIT and Harvard

Many arthropod species, including medically important vectors like the *Anopheles gambiae* complex, are recalcitrant to *de novo* genome assembly because of extensive polymorphism and barriers to inbreeding. Creating a quality reference genome that is robust for downstream applications can require money and technical capabilities beyond the reach of many labs. Innovations in sequencing technology and assembly programs offer the potential for improvement, but have been tested most extensively in humans and other similarly tractable mammals.

We tested one such program, the assembler DISCOVAR *de novo*, on the mosquito *An. arabiensis*, a major malaria vector in sub-Saharan Africa. With 250 bp paired-end reads from a PCR-free library of a single insert size, we generated an assembly with contig N50 approaching 21K. (For comparison, contig N50s for eleven recently published anopheline genomes generated using multiple long-range libraries from inbred colonies ranged from 20K to 200K.) To determine to what extent the quality of the DISCOVAR *de novo* assembly was attributable to the increased read length, we made another assembly from the same 250 bp paired-end reads using a program designed for heterozygous organisms, and found a contig N50 of 7K.

We evaluated the utility of the DISCOVAR *de novo*-created assembly for gene-based approaches by examining a set of universal benchmarking genes. These genes were found at rates and qualities approaching those of AaraD1, the existing *An. arabiensis* assembly created with fragment, jump, and “fosill” libraries. Additionally, the DISCOVAR *de novo* assembly spanned several gaps present in AaraD1, including some in commonly hard-to-assemble regions of heterochromatin.

DISCOVAR *de novo* also retains heterozygosity usually discarded by other assemblers. By parsing this information, we found multiple signatures of polymorphism that corresponded with independently validated nucleotide diversity. Polymorphism recoverable from the assembly was not limited to single nucleotide polymorphisms (SNPs) and simple indels, but included more complex, rarely seen variants, including several hundred over ten kilobases long. We used the existing AaraD1.2 gene set to identify potential impacts of the recoverable variation and assess its relevance.

We show that DISCOVAR *de novo* not only produces a quality assembly in a heterozygous arthropod of medical importance, from a single library, but also retains additional data about polymorphism within the resulting genome. This sequencing approach therefore offers arthropod researchers the opportunity to obtain, in the same effort, a reference assembly and basic population genomic information to guide further investigation.

**78 - MVG**
**Functional Characterization of the 5-HT Receptor in Aedes aegypti and Anopheles gambiae**

Ngai, Michelle; Biological Sciences, University of Notre Dame
Additional authors:
Shoue, Douglas; University of Notre Dame
McDowell, Mary Ann; University of Notre Dame

Insect repellents and/or insecticides are often heavily relied upon for the control of arthropod-borne diseases due to the absence of a viable vaccine. Identifying novel insecticides with alternative modes of action is of utmost importance given the worldwide development of insecticide resistance. G-protein coupled receptors (GPCR) are considered lucrative drug targets due to their involvement in a diverse number of physiological processes. In mosquitoes, GPCRs have been reported to mediate pathways that facilitate its ability to transmit disease. For example, the serotonin (5-HT) GPCR group is implicated in the olfaction system, which plays a role in mediating their preference for human hosts. 5-HT is also known to modulate circadian rhythm and memory. To determine if the 5-HT receptors are suitable targets as insecticides and/or insect repellants, the functional and pharmacological characterization of this receptor family is essential. The 5-HT receptor family is large, with 7 classes (5-$HT_1$ through 5-$HT_7$) and 14 known subtypes. With the exception of 5-$HT_3$, all serotonin receptors are coupled either to an $\alpha G_i/G_o$, $\alpha G_s$, or $\alpha G_q/G_{11}$ protein, activating independent intracellular signaling processes to ultimately decrease intracellular cyclic adenosine monophosphate (cAMP), increase cAMP or increase calcium levels, respectively. The identification of 9 and 7 putative 5-HT target genes in *Aedes aegypti* and *Anopheles gambiae*, respectively, by probing for highly conserved regions of amino acids via sequence similarity and homology-based searches was possible due to whole genome deep-sequencing projects on various species, including *Drosophila melanogaster*. The stage-specific expression profile of several 5-HT receptors was determined by quantitative RT-PCR, identifying significant levels of expression in the female head and body. Additionally, since the complete genomic sequences for two 5-HT receptor genes (AAEL0011844 and AAEL009573) are available, the target receptors were cloned into the pF9A CMV *hRluc*-neo Flexi® Vector and expressed transiently in the human embryonic kidney (HEK)-293 cells. The receptors were validated using GloResponse reporter systems and specificity was assessed against other biogenic amines including octopamine, tyramine, dopamine, and histamine. Due to the presence of a specific response to 5-HT, stable cell lines were established in the HEK-293 cells. Potential agonist activity was analyzed using luminescence based cell assays and dose response curves. This study seeks to address the strong need for novel insecticides/repellants by focusing on the discovery of novel chemistries disrupting the 5-HT gene target, effectively altering the normal gene function and ultimately, the behavior of the mosquito.

**79 - MVG**
**Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti**

Overcash, Justin; Entomology, Virginia Tech
Additional authors:
Basu, Sanjay; Virginia Tech
Aryan, Azadeh; Virginia Tech
Samuel, Glady; Virginia Tech
Anderson, Michelle; Virginia Tech
Dahlem, Timothy; University of Utah
Kevin Myles, Virginia Tech; Zach Adelman, Virginia Tech

Conventional control strategies for mosquito-borne pathogens such as malaria and dengue are now being complemented by the development of transgenic mosquito strains reprogrammed to generate beneficial phenotypes such as conditional sterility or pathogen resistance. The widespread success of site-specific nucleases such as transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 in model organisms also suggests that re-

programmable gene drive systems based on these nucleases may be capable of spreading such beneficial phenotypes in wild mosquito populations. Using the mosquito Aedes aegypti, we determined that mutations in the FokI domain used in TALENs to generate obligate heterodimeric complexes substantially and significantly reduce gene editing rates. We found that CRISPR/Cas9- based editing in the mosquito Ae. aegypti is also highly variable, with the majority of guide RNAs unable to generate detectable editing. By first evaluating candidate guide RNAs using a transient embryo assay, we were able to rapidly identify highly effective guide RNAs; focusing germ line-based experiments only on this cohort resulted in consistently high editing rates of 24–90%. Microinjection of double-stranded RNAs targeting ku70 or lig4, both essential components of the end-joining response, increased recombination-based repair in early embryos as determined by plasmid-based reporters. RNAi-based suppression of Ku70 concurrent with embryonic microinjection of site-specific nucleases yielded consistent gene insertion frequencies of 2–3%, similar to traditional transposon- or ΦC31-based integration methods but without the requirement for an initial docking step. These studies should greatly accelerate investigations into mosquito biology, streamline development of transgenic strains for field releases, and simplify the evaluation of novel Cas9-based gene drive systems.

**80 - MVG**
**Correcting genome misassemblies by physical mapping in the Neotropical malaria vector *Anopheles albimanus***

Peery, Ashley; Entomology, Virginia Tech
Additional authors:
Artemov, Gleb; Institute of Biology and Biophysics, Tomsk State University, Tomsk, Russia
Steginy, Vladimir; Institute of Biology and Biophysics, Tomsk State University, Tomsk, Russia
Sharakhova, Maria; Department of Entomology and Fralin Life Science Institute, Virginia Tech, Blacksburg, VA, USA
Sharakhov, Igor; Department of Entomology and Fralin Life Science Institute, Virginia Tech, Blacksburg, VA, USA

*Anopheles albimanus* is a major malaria vector that transmits disease in an area stretching from Mexico and Central America into northern South America. Until recently, genetic tools for this species, have been limited to a low-resolution photo map and a linkage map for chromosome 2. The photomap, published in 2000, established arm homology with *An. gambiae* based on 17 physically mapped markers. Although the 2000 map is a significant improvement over its predecessors, this map was not designed for physical mapping purposes. Recent publication of the *An. albimanus* genome sequence as part of the 16 genomes project revealed a relatively small genome assembly of 170 Mb and long scaffolds with an N50 of ~18 Mb. These promising statistics present an opportunity to create a high-quality physical genome map for this species with considerably less effort than mapping endeavors for other species. However, *in silico* comparison of gene order with *An. gambiae* predicted several misassemblies within the *An. albimanus* genome sequence. We physically mapped ~97% of the *An. albimanus* genome and confirmed 13/15 predicted misassemblies within 6 scaffolds. We then compared the position of our probes with those of the linkage map published for chromosome 2 in 2009, and established correspondence between the two maps. As has been demonstrated by the *An. gambiae* and *An. stephensi* genomes, physical mapping and genome sequencing combine to yield chromosomal assemblies which are extremely valuable resources. Here we present our physical genome map— with 97% of the genome sequence assembled into chromosomes, our chromosomal assembly surpasses all mosquito genomes published to date. This map can assist in studies of genome landscape, chromosome evolution, population genetics and genome editing to name just a few.

**81 - MVG**
**Identifying novel genetic contributors to the siRNA immune pathway in *Aedes aegypti***

Saadat, Angie; Entomology, Virginia Tech
Additional authors:
Adelman, Zach; Virginia Tech, Department of Entomology
Anderson, Michelle; Virginia Tech, Department of Entomology
Myles, Kevin; Virginia Tech, Department of Entomology
Hoeschele, Ina; Virginia Bioinformatics Institute at Virginia Tech

*Aedes aegypti* is an important vector of human pathogens including yellow fever, dengue and chikungunya viruses. The small interfering RNA pathway is a critical immune response for controlling viral replication in *A. aegypti*. The goal of this research is to identify components of the *A. aegypti* genome that influence this innate immune pathway. A transgenic mosquito strain that reports the status of the siRNA pathway via GFP intensity was employed to differentiate silencing abilities among individuals. Extreme GFP expression phenotypes, representing efficient and poor silencing abilities, were enriched over five generations. Transcriptome sequencing and analysis will be performed from pools of individuals from each final phenotype enrichment. QTL crosses were performed from single mate pair crosses of one $F_0$ parent from each enriched extreme phenotype line. $F_1$ progeny were self-crossed and the $F_2$ progeny demonstrating each of the extreme phenotypes were selected and archived with their $F_0$ grandparents for DNA sequencing and SNP analysis.

**82 - MVG**
**AT LAST, RESOLUTION OF HOW RR-1 AND RR-2 SUBFAMILIES OF CPR CUTICULAR PROTEINS ARE DEPLOYED**

Vannini, Laura; Cellular Biology, University of Georgia
Additional authors:
Willis, Judith H; University of Georgia

*Anopheles gambiae* assigns almost 2% of its protein coding genes to structural cuticular proteins (CPs). These have been classified into 13 (14) families based on sequence characteristics (See Willis, 2010 for review). Sequence domains, homology models and experimental work revealed that members of some CP families contribute to the cuticle by binding chitin (Rebers and Willis, 2001; Tang et al., 2010; Jasrapuria et al., 2010; Willis et al., 2012). The largest family, CPR, is defined by the presence of the Rebers and Riddiford (R&R) Consensus (pfam00379) known to confer chitin-binding properties (Rebers and Willis, 2001). Two main distinct forms of the Consensus, RR-1 and RR-2, have been recognized and named by Andersen (1998, 2000). RR-1-bearing proteins have been isolated from flexible cuticles, while RR-2 proteins have been associated with hard cuticle. This generalization was based on relatively few cases, and it has also been suggested that RR-2 proteins will contribute to exocuticle while RR-1 will be found predominantly in endocuticle (Andersen, 2000). This issue has not been resolved, even with the extensive expression data that are available in Togawa et al., 2007.

It is known from several works that cuticle from different stages and different anatomical regions had different cuticular proteins and moreover different cuticle proteins can be expressed at the same time in the same tissue but in different part of the cuticle itself: i.e. CPLCG3/4 and CPF3 have been found on 1-day-old adult legs in the endo- and in the exocuticle, respectively (Vannini et al., 2014).

By electron microscopy immunolocalization with colloidal gold we found evidence that CPRs bearing the RR-1 are expressed in the soft cuticle (intersegmental membrane) without any preference to endo- or exocuticle. RR-2 CPRs were only localized in the hard cuticle (abdominal sclerites).

The clear association of different CPRs to different cuticular portions underlined the complex organization of the cuticle as well as why *An. gambiae* devotes so many genes to structural CPs.

**83 - MVG**
**Islands with genetically isolated populations of *Anopheles gambiae sensu stricto* mosquitoes in Lake Victoria, Uganda could function as field sites for testing transgenic drive**

Wiltshire, Rachel; Biological Sciences, University of Notre Dame
Additional authors:
Teberg, Colin; Department of Computer Science and Engineering, University of Notre Dame
Kayondo, Jonathan; Uganda Virus Research Institute, Entebbe, Uganda
Emrich, Scott; Department of Computer Science and Engineering, University of Notre Dame
Collins, Frank; Department of Biological Sciences, University of Notre Dame

Malaria (clinical disease through infection with species of *Plasmodium* parasites) is the third largest cause of pediatric death in the developing world. Almost 75% of the 583,000 deaths attributed to it in 2013 were African children under the age of five years. Mosquitoes in the genus *Anopheles* transmit *Plasmodium* parasites to humans by injecting sporozoites (infective stage) through the proboscis whilst taking a blood meal. Current vector control programs target indoor feeding (endophagic) mosquitoes through the distribution of insecticide residual spraying (IRS) and long-lasting insecticide-treated bed nets (LLINs). It has been shown that certain mosquito populations have responded to these control efforts by demonstrating resistance to the insecticides or shifting feeding habits to coincide with early evening outdoor human activities (meals/socializing) to maximize their chances of successfully obtaining a blood meal. These changes have; therefore, limited the use of available vector control tools requiring entomologists to consider alternative technologies such as genetically modified mosquitoes. Many studies have investigated the use of mosquito transgenes *in vitro*; however, to date, this research has not been taken outside of the laboratory and tested seriously in genetically isolated populations in the field. Six populations of *Anopheles gambiae sensu stricto* mosquitoes were sampled from different locations in the Lake Victoria area of southern Uganda and sequenced with restriction site- associated DNA (RAD) chemistry. Following bioinformatics processing, the resulting single nucleotide polymorphism (SNP) sets were analyzed for population structure using principal components analysis and the software packages fastSTRUCTURE v. 1.0 and ADMIXTURE v. 1.23. We saw an emerging pattern in the population genetic structure between island and mainland sites that was suggestive of small effective population size ($N_e$) allele frequency drift combined with relatively low rates of gene flow, marking these particular sites as suitable locations for releasing transgenic mosquitoes and monitoring genetic drive.

# Population Genomics

**85 - PG**
**Genetic basis of post-diapause development of the Glanville fritillary butterfly**

Ahola, Virpi, Metapopulation Research Center, Department of Biosciences, University of Helsinki
Additional authors:
Rastas, Pasi; Department of Zoology, University of Cambridge, UK
Ikonen, Suvi; Department of Biosciences, University of Helsinki, Finland
Fountain, Toby; Department of Biosciences, University of Helsinki, Finland
Wong, Swee Chong; Department of Biosciences, University of Helsinki, Finland
Hanski, Ilkka; Department of Biosciences, University of Helsinki, Finland

In the Glanville fritillary butterfly post-diapause larval and pupal development is related to reproductive success of females and mating success of males. Larval and pupal development traits are known to be highly heritable, but genetic basis of these traits is largely unknown. We used F2 experimental crosses

between individuals from two Glanville fritillary populations to find genetic loci associated with post-diapause larval and pupal development traits. We carried out quantitative trait mapping (QTL) for more than 700 individuals and 240 SNPs. The results show several QTL associated with studied traits with low heritablity. To fine map the found loci, we resequenced the whole genome from 190 individuals. We identified several potential genes associated with post-diapause larval development time, growth rate and pupal weight.

**86 - PG**
**Population Genomics of the Mountain Pine Beetle (*Dendroctonus ponderosae*)**

Dowle, Eddy; Entomology, Kansas State University
Additional authors:
Bracewell, Ryan; University of Montana
Pfrender, Michael; University of Notre Dame
Mock, Karen; Utah State University
Bentx, Barbara; Utah State University; USDA Rocky Mountain Research Station
Ragland, Greg; Kansas State University

The Mountain Pine Beetle (MPB) is native to the pine forests of western North America, reaching from northern Mexico to Canada. Current MPB ranges are expanding as a result of climate change and infestations cause extensive damage to susceptible forests. Applying genome-wide RAD sequencing to over 700 individuals from 37 populations, we are examining the population genetics of MPB across its range. Strong population structuring occurs between geographic regions with extensive admixing occurring in the sky island populations residing in the Great Basin. However results also suggest that there are distinct genetic boundaries between some geographically proximate regions. Experimental crossing experiments have previously found postzygotic isolation between geographically distant populations. Moreover crosses between the geographically close populations of Oregon and Idaho also showed a reduction in hybrid male fitness despite their proximity. Population structuring of the autosomes and sex chromosomes show that these populations are permeable to gene flow at autosomal sites but Y chromosome sites in particular show clear disjunction between regions. Ongoing analyses are also exploring signatures of local adaptation using genome-scans and environmental correlations.

**87 - PG**
**VectorBase Population Biology (PopBio) browser and map**

Emrich, Scott, Department of Computer Science and Engineering, University of Notre Dame
Additional authors:
Giraldo-Calderón, Gloria I.; University of Notre Dame
MacCallum, Bob; Imperial College London
Kirmitzoglou, Ioannis; Imperial College London
Lawson, Daniel; European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL - EBI)
Collins, Frank H.; University of Notre Dame
Members of the VectorBase Consortium

VectorBase is a Bioinformatics Resource Center for invertebrate vectors of human pathogens, supported by the National Institute of Allergy and Infectious Diseases and National Institutes of Health, NIAID/NIH. The VectorBase database is updated and expanded every two months. In 2014 and 2015 we have updated the gene builds for our 35 genomes and also added a new genome, the house fly (*Musca domestica*), to facilitate comparative analysis with dipteran vectors. One of our tools, the Population Biology Browser (PopBio), is part of our ongoing efforts to integrate genomic, phenotypic and population data. PopBio projects are from a wider range of vector species (*i.e*., include more than our 35 genomes) and are organized in samples and assays, such as WHO and CDC bottle insecticide resistance assays. Until recently, PopBio only entry points for researchers were text search and basic browsing. Here we present 1) a map-based interface for the interrogation and visualization of the geotagged data that

constitutes the majority of VectorBase's PopBio data; 2) a new query interface to allow users to isolate subsets of interest, with summary statistics; 3) interchangeable background map layers, including retrospective climate and vegetation cover data, further equip the tool for hypothesis generation and; 4) a new data submission interface. If you have any questions or were not able to visit our poster for a demo, please send us your comments to info@vectorbase.org.

**88 - PG**
**The genetic basis of age-related change in stress tolerance and fitness in *Drosophila melanogaster***

Everman, Elizabeth; Department of Biology, Kansas State University
Additional authors:
Hunter, Kate; Iowa State University
Morgan, Theodore; Kansas State University

Organisms occur in environments that vary spatially and temporally throughout their lifespans, and genetic architecture plays an important role in controlling the expression of traits. In addition, mutation accumulation and antagonistic pleiotropy play important related roles in the change in trait expression through the aging process. Resistance to cold stress is an important fitness trait that is expected to decline with age; however this general expectation is based on the response of a small number of *Drosophila* isogenic lines. To further characterize this age-related change in cold stress resistance, we measured cold stress resistance of flies reared under standard rearing conditions and flies that experienced a short mild acclimation prior to cold stress exposure. This acclimation is referred to as Rapid-Cold Hardening (RCH) and is meant to model the environment experienced when the thermal environment cools as a cold front arrives or during diurnal thermal cycles. We measured RCH on 100 isogenic lines of the *Drosophila melanogaster* in the *Drosophila* Genetic Reference Panel (DGRP) at early (5-7 days post eclosion) and late (20-22 days post eclosion) age. Consistent with previous investigations and predictions of RCH and the DGRP lines, we observed a wide range of variation between lines at both early and late age points. In addition, cold stress resistance differs significantly ($p << 0.05$) between early and late aged flies. However, we observed that the direction of the change in expression of cold stress resistance varies among lines as well, with several lines increasing in cold stress resistance as they age. Association mapping shows that the genetic architecture underlying the response to cold stress changes with age as well, with several age-specific SNPs associated with cold stress phenotypes in each age group. Analyses of genetic variance indicate that aging in these lines may be due to the mutation accumulation (the accumulation of deleterious alleles in a population that are expressed late in life). To investigate the consequences of RCH exposure on fitness we measured mating activity on 20 lines with diverse RCH responses. These results suggest that the RCH response is positively correlated with courtship efficiency ($p<0.05$), while basal levels of cold tolerance are negatively correlated with courtship efficiency ($p<0.1$). Taken together, these data provide insight to the fitness consequences associated with stress tolerance in *Drosophila melanogaster.*

**89 - PG**
**Temporal Heterogeneity in Genetic Variation In Wild-Caught *Drosophila Simulans***

Freda, Philip; Department of Entomology, Kansas State University
Additional authors:
DiMeglio, Matthew; Saint Joseph's University
Braverman, John; Saint Joseph's University

The phenotypic composition of natural populations is not constant but rather changes through time. Underlying this phenotypic change are shifts in allele frequency due to forces like natural selection and genetic drift. However, little research has been performed to estimate variation at the DNA sequence level over short time intervals. The objective of this project was to test for heterogeneity in genetic variation across or between seasons in natural populations of *Drosophila simulans*. Tests of temporal variation in fifteen microsatellite loci were performed using wild specimens of *D. simulans* collected on the campus of

Saint Joseph's University in Philadelphia and Lower Merion, Pennsylvania, U.S.A. Specimens were collected over twelve dates in 2011 and 2012. Pairwise tests of genic differentiation revealed significantly different allelic distributions at a number of loci with the most significant changes observed between the end of the 2011 collection season (11/11/11) and the beginning of the 2012 collection season (7/31/12). Likewise, significant deviations from Hardy-Weinberg Equilibrium were strongest and observed heterozygosity lowest from late fall through early spring. These results suggest a steady decline in population size during the fall followed by a rebound of population size the following spring. Changes in allele frequency accompanying these demographic changes suggest that selection and/or drift cause annual, evolutionary shifts in the population. Local populations of *D. simulans* appear to be dynamic and cyclical, with allelic distributions changing year after year.

**90 - PG**
**Improving Bt Resistance Risk Assessment and Management by Genomic Monitoring**

Fritz, Megan; Entomology, North Carolina State University
Additional authors: Yes
Gould, Fred; North Carolina State University

The US government views the insecticidal properties of Bacillus thuringiensis (Bt) as a "public good", and mandates that Bt be used in a manner that reduces the risk of insect pests evolving resistance. Both the US-EPA and USDA have concluded that monitoring for resistance could improve agricultural management practices aimed at decreasing resistance risk. Current monitoring methods are inadequate, however. We are using new genomic tools in concert with field data to better assess both the current extent of Bt resistance in Heliothis virescens and Helicoverpa zea moths and the rate at which resistance is increasing. We developed a Double-Digest Restriction-site Associated DNA Sequencing (ddRAD-seq) approach capable of detecting allele frequency changes throughout the insect genome. We scanned the genomes of archived H. virescens and H. zea specimens from 1997 through 2012 and discovered a number of genomic regions that have changed over time, likely in response to Bt selection. In some cases, these genomic regions were located near to previously described Bt resistance genes. In other cases, new candidate genes appeared to be evolving in response to selection by Bt crops. Ultimately, our work will confirm whether insect resistance to Bt-expressing crops can be predicted by quantifying patterns of genomic change over time.

**91 - PG**
**Transcriptomics must take into account unexpected levels of endoreduplication and underreplication**

Hjelmen, Carl; Department of Entomology, Texas A&M University
Additional authors: Yes
Mynes, Melissa; Texas A&M University
Johnston, J. Spencer; Texas A&M University

When analyzing the significance of transcriptomic data, it is generally assumed that each transcript is coming from one copy of the gene; however, this may not always be the case. Organisms have been able to manipulate their genome in order to gain evolutionary advantages. One way this is accomplished is through the process of endoreduplication, in which the genome undergoes replication yet the cell does not mitotically divide. These endopolyploid cells have been found to be common in biosynthetically specialized cells, such as *Drosophila* salivary glands, silkworm moth silk glands, and the muscle cells of Hymenopteran males. Not only has this been documented in these specialty cells, but a variation of endopolyploidy (underreplication) has been found in a majority of thoracic cells in *Drosophila* species. Underreplication is a process in which replication of the genome begins, but does not complete, largely replicating the expressed euchromatic regions and not all of the heterochromatin. Having multiple copies of the genome can dramatically affect how people should interpret transcriptomic data, as one gene may not be contributing all of the expressed transcripts. Therefore, we have estimated levels of underreplication

in the thoraces of *Drosophila* species in order to gain a better understanding of this unique process in relation to genome size and its importance in a phylogenetic aspect.

**93 - PG**
**Population differences in seasonality via diapause regulation among the apple maggot fly Rhagoletis pomonella**

Schieferecke, Adam; Department of Entomology, Kansas State University
Additional authors:
Ragland, Gregory; Department of Entomology, Kansas State University

Physiological mechanisms controlling diapause have been well studied, but little is currently known about how variation in these mechanisms give rise to diapause adaptation. Diapause adaptation is important as insects expand into new geographic regions or experience changes in seasonality, the latter of which has been occurring as global climates change at an increased rate. *R. pomonella* has rapidly evolved into two genetically distinct populations that infest host plants with different fruiting times. Differences in the timing of diapause allow each population to synchronize with their respective host plant. Previous studies suggest that regulatory mechanisms underlying this shift in seasonality may act during the winter independent of any overt cues for diapause termination, such as the typical change from colder to warmer temperatures during the winter-spring transition. To test for potential physiological differences, we sampled diapausing pupae in a narrow time window encompassing three time points: 1) directly out of an overwintering cold treatment, 2) 24 and 3) 48 hours after transfer into a warm treatment. The results from performing RNAseq on the samples show that, although temperature has strong effects on gene expression (consistent with the idea that temperature is a diapause termination cue), there were also pronounced differences in expression between the two populations that persisted across all time points, including prior to warm temperature exposure. It can be concluded from our results that a timing mechanism independent of temperature cue likely causes population differences in seasonality via diapaus timing. Further, we discuss candidate genes and pathways underlying population differences in diapause regulation.

**94 - PG**
**Landscape genomics approaches in agricultural pest management**

Schoville, Sean; Entomology, University of Wisconsin Madison
Additional authors:
Crossley, Michael; University of Wisconsin-Madison

There is a growing need to identify the genetic basis of local adaptation in both model and non-model species. In agricultural landscapes, the identification of adaptive variation in arthropod pests and the spatial distribution of this variation could be used to enhance management strategies that seek to mitigate pest evolution. By leveraging large genomic and environmental data sets, landscape genomics approaches offer a novel framework to test patterns of adaptive genetic variation in a spatially explicit context. Using genome scans, I examine the distribution of adaptive genetic variation among Midwestern populations of the Colorado potato beetle (*Leptinotarsa decemlineata*) and link this variation to environmental variables, including pesticide usage intensity. I discuss the implications of this approach for informing resistance management practices at large spatial scales.

**95 - PG**
**Molecular evolution of spermatophore proteins in *Heliconius* butterflies**

Shives, Channing; Ecology and Evolutionary Biology, University of Kansas
Additional authors:
Walters, James; University of Kansas

Spermatophores are male-produced capsules that aid in the transfer of sperm to females during insect reproduction, but very little is known about the repetitive structural proteins that compose them. This research aims to characterize the molecular evolution of spermatophore proteins (spermatophorins) in *Heliconius* butterflies and targets putative spermatophorins previously identified during an expressed sequence tag (EST) transcriptomic study of male accessory gland proteins. Initial results from EST analysis were ambiguous concerning gene copy number of three distinct groupings of spermatophorin sequences. BLAST searches for these spermatophorin protein sequences in the complete *H. melpomene* genome and determined each grouping reflects a single-copy gene. Sequence alignments between the genomic and EST sequences suggest substantial non-synonymous polymorphism for these loci. Efforts are ongoing to further characterize polymorphism within *H. melpomene* and to identify homologous loci in genome assemblies of related species of *Heliconius* butterflies in order to estimate parameters of molecular evolution and population genetic diversity. We ultimately aim to test whether the unusually high divergence observed for these proteins reflects relaxed purifying selection or positive Darwinian selection.

**96 - PG**
**Phenotypic and genetic divergence among natural populations of *Daphnia pulex***

Won, Jihyun; Department of Biological Sciences, University of Notre Dame
Additional authors:
Lopez, Jacqueline; University of Notre Dame
Clifford, Benjamin; University of Notre Dame
Pfrender, Michael; University of Notre Dame

One of the key objectives in evolutionary biology is to understand the underlying sources of phenotypic variation in natural populations. Here we investigate phenotypic variation and genome-wide patterns of genetic variation in four populations of the freshwater zooplankton *Daphnia pulex*. Our goal is to address three questions: How much phenotypic and genetic variation exist among populations? What are the underlying genomic regions associated with population differentiation? What are the patterns of phenotypic and genetic associations among populations? Our life-history assay shows that four populations have little variation in morphological and life-history traits. We used 19215 SNPs from RAD-seq data to examine genetic differentiation across the genome. We examined the patterns of population structure using a principal component analysis (PCA) and a model-based clustering approach implemented in STRUCTURE. Overall, more variation at the genomic level was detected than at the phenotypic level among four populations. Finally we performed a GWAS analysis to identify SNPs associated body size. Although GWAS analyses had limited power to detect associations due to the small size of our populations, our results provide candidate genomic regions to investigate for future studies.

# Epigenomics

**97 - E**
**Is methylation one of the drivers of virulence in Diuraphis noxia (Kurd.) Hemiptera: Aphididae?**

Botha, Anna-Maria; Professor, Genetics Department, Stellenbosch University

Diuraphis noxia (Kurdjumov, Hemiptera: Aphididae), a specialist phloem feeder, is an economically important aphid pest afflicting wheat and barley yield in dry-land production regions of the USA, Argentina and South Africa. Populations sharing similar ecoagricultural regions and expressing different levels of virulence towards their hosts are called biotypes, and the number of D. noxia biotypes reported continues to increase, posing multiple threats to global food security. With the availability of the draft genome of D. noxia and confounding evidence of genomic plasticity, we set out to determine the extent of methylation in the genome of Diuraphis noxia in order to determine if methylation contributes to changes in virulence. To this end, the global levels of methylation as well as the methylation profiles of the different biotypes were investigated, the former done by measuring fluorescent adaptor levels when aphid DNA was restricted with isoschizomers HpaII and MspI. The latter involved the use of Methylation-Sensitive Amplification Polymorphism, and also provided insight into local regions of methylation in the genome. The global methylation results suggest an inversely proportional relationship between virulence levels and methylation, hypomethylation being associated with increased virulence. Methylation profiles of the biotypes, whilst similar, did show some clear differences indicating that differential methylation of certain genes could indeed contribute to differences in virulence. This study, being the first of its kind for Diuraphis noxia, has provided the groundwork for future research into methylation of this insect, and adds to a growing body of knowledge on the Russian Wheat Aphid.

**98 - E**
**Alternatively Spliced mRNA Isoforms Are Differentially Expressed Between Polyphenic Morphs in the Pea Aphid (*Acyrthosiphon pisum*)**

Chaffee, Mary; Department of Ecology and Evolutionary Biology, University of Rochester
Additional authors:
Brisson, Jennifer; University of Rochester, Department of Ecology and Evolutionary Biology. Rochester, NY 14627

Polyphenism is an extreme form of phenotypic plasticity where a single genotype can mature into multiple, alternative phenotypes that are discrete and environmentally adaptive. Alternative splicing creates a diversity of mRNA isoforms that can be used to respond to environmental conditions by influencing phenotypic traits. However, very little research has focused on alternative splicing in polyphenisms. In pea aphids, winged or wingless females can be produced due to environmental stress, and asexual or sexual morphs develop based on the photoperiod experienced by the mother. All female morphs are genetically identical. Preliminary quantitative PCR results have shown that alternative splicing exists between winged and wingless asexual females and alternative mRNA isoforms of the same gene are expressed at different levels between the morphs. We tested the hypothesis that alternative splicing is utilized equally or more between polyphenic morphs than between the sexes with the following comparisons: winged versus wingless asexual female morphs, sexual versus asexual females morphs, and males versus sexual females. Using RNA-seq data from three different strains of pea aphids, significantly different alternatively spliced genes between morphs were identified with DEXseq, MISO, and cuffdiff. These programs were chosen to compare alternative exon usage (DEXseq), and mRNA isoform expression levels (cuffdiff & MISO) between the morphs. We will present the results of these analyses.

Funded by the Biology Department at the University of Rochester

**99 - E**
**Understanding the tension between genome defense and genomic "autoimmunity" by piRNAs**

Erwin, Alexandra; Ecology and Evolutionary Biology, University of Kansas
Additional authors:
Galdos, Mauricio; University of Kansas
Wickersheim, Michelle; University of Kansas
Blumenstiel, Justin; University of Kansas

Natural selection on transposable elements (TEs) favors their proliferation within genomes. In response to this threat, piRNA mediated genome defense has evolved to limit their transposition. Recent studies have shown that in addition to TEs, genes may also be the target of piRNA silencing. Genic targeting by piRNAs can result from proximal TE insertions that drag flanking sequences into piRNA-mediated gene silencing in cis, resulting in “off-target” silencing. Using a dysgenic syndrome in Drosophila virilis, we show that patterns of genic off-targeting can differ significantly between strains. In contrast to piRNA off-targetting caused by proximal TE insertions, we find that many genes seem to be targeted in a different way, without evidence of nearby TEs. In addition, we find that when TEs are chronically expressed in the dysgenic germline, genic piRNA targetting increases. The tension between silencing TEs and consequent effects on gene expression may be an important evolutionary determinant in the evolution of the piRNA machinery.

**100 - E**
**Ecological adaptation and speciation of *Spodoptera frugiperda* host-plant strains: a comparative transcriptomic and epigenetic approach**

Mone, Yves; Diversity, Genomes and Insects-Microorganisms Interactions, INRA (UMR 1333) Université de Montpellier
Additional authors:
Orsucci, Marion; UMR 1333 INRA Université de Montpellier
Gimenez, Sylvie; UMR 1333 INRA Université de Montpellier
Nègre, Nicolas; UMR 1333 INRA Université de Montpellier
d'Alençon, Emmanuelle; UMR 1333 INRA Université de Montpellier

Among phytophagous insects, adaptation to different host plant species drives their evolution and may result in the emergence of new host races or new species through ecological speciation. Variation in gene expression patterns could play a key role in ecological speciation by promoting adaptation to new environment (here a new host plant) and by affecting adaptive genetic divergence in traits causing reproductive isolation. Our aim is to characterize such variation in the fall armyworm *Spodoptera frugiperda* (Lepidoptera Noctuidae). This Lepidopteran pest is found in the American continent as two sympatric strains that can be differentiated by their host-plant preference: one mostly associated to corn (corn variant) and the other mostly associated to rice (rice variant). We performed a functional comparative study of these two variants reared on two host plants (rice and corn) to detect differentially expressed genes and their regulators related to the adaptation to host-plant. First, RNA pools of larval tissues for the different modalities (variant/host) were collected to perform a global study of the genetic processes involved in the larval performance. This was done by a RNA-seq analysis aiming at identifying the gene families and the biological functions putatively involved in host specialization of moth larvae. Then, to identify transcriptional and post-transcriptional regulators, we started to compare the epigenetic landscapes of the two variants. We focused on the genomic distribution of various epigenetic marks such as histone modifications by performing chromatin immunoprecipitation experiments (ChIP-seq), open chromatin regions which are associated with regulatory activity using FAIRE-seq technique (Formaldehyde-Assisted Isolation of Regulatory Elements) and small non coding RNAs (sRNAs) involved in gene silencing and that could trigger chromatin-silencing modification (histones modification and DNA methylation).

**101 - E**
**RNA-seq Analysis of Daphnia pulex Under Different Environmental Stressors**

Regan, Kerry; Biology, University of Notre Dame
Additional authors:
Lopez, Jacqueline; University of Notre Dame
Pfrender, Michael; University of Notre Dame

We produced an RNA-seq dataset using the ecological model organism, *Daphnia pulex*. *D. pulex* is a cyclical parthenogen that is found in vernal pools throughout North America. We used a single genotype, which was collected from the wild, and exposed it to six different environmental stressors, along with a laboratory control. The RNA-seq data was mapped to the *D. pulex* v1.1 genome using Tophat and transcript counts were obtained using HT-seq Count. We used DEseq 2 to determine differential gene expression between each of the environmental stressors and the control. Approximately 10,000 genes were represented across the treatments and over 2500 genes were differentially expressed in any one of the environmental stressors. We also found sets of genes that are differentially expressed across all treatments to determine a generalized stress response across these environmental stressors.

**102 - E**
**An Integrated Annotation Data Warehouse at the Hymenoptera Genome Database**

Tayal, Aditi; Animal Sciences, University of Missouri
Additional authors:
R. Unni, Deepak; University of Missouri, Animal Science

The Hymenoptera Genome Database (HGD; http://HymenopteraGenome.org) is a genome informatics resource for hymenoptera insect species. HGD includes genomic data for three honey bee species (Apis mellifera, A. florea, A. dorsata), two bumble bee species (Bombus terrestris and B. impatiens), nine ant species (Acromyrmex echinatior, Atta cephalotes, Camponotus floridanus, Cardicondyla obscurior, Harpegnathos saltator, Linepithema humile, Pogonomyrmex barbatus, Solenopsis invicta, Wasmannia auropunctata), a parasitoid wasp (Nasonia vitripennis) and a sweat bee (Lasioglossum albipes). We have used InterMine to deploy a new data warehouse called HymenopteraMine to allow fast and flexible queries of annotation and ortholog data across species. HymenopteraMine integrates complex biological information from various data sources including RefSeq, UniProt, InterPro, OrthoDB, Pubmed and Gene Ontology. Users may perform a “Quick Search” or more detailed searches using predefined or customizable search “Templates” or “Query Builder”. The “Genome Region Search” and “Overlapping Feature Search” allow users to download annotations with a specified genomic context. Links between genes and publications facilitate access to relevant scientific literature. Users may download query results in various formats, such as tab-delimited files, GFF, Fasta, BED, JSON and XML. HymenopteraMine also provides widgets for Enrichment analysis or for displaying information in a tabular format and graphical format. In addition to HymenopteraMine, users may leverage the HGD genome browsers (GBrowse and JBrowse), BLAST searching and data download pages to access genome assemblies, computed and manually-annotated genes, protein homologs, cDNA sequences, non-coding RNA sequences, RNAseq-based expression data and genetic markers. For users who would like to programmatically access HymenopteraMine, they can make use of the “Intermine API” which is available in Perl, Python, Ruby and Java. Users can make a list of their own queries, store and analyze the data in their personal accounts.

**103 - E**

**Sex chromosome dosage compensation in Lepidoptera: insights from nymphalid butterflies, coddling moth, and demasculinized silkworms**

Walters, James; Ecology & Evolutionary Biology, University of Kasas
Additional authors:
Hardcastle, Thomas; Plant Biology, Cambridge University
Gu, Aloy; Cornell University
Jiggins, Chris; Zoology, Cambridge University

Dosage compensation is the equalization of gene expression levels in response to differences in gene dose or copy number. It is classically considered to play a critical role in the evolution of heteromorphic sex chromosomes. As the X and Y diverge through degradation and gene loss on the Y (or the W in female-heterogametic ZW taxa), it is expected that dosage compensation will evolve to correct for sex-specific differences in gene dose. Although this is typically observed in male-heterogametic (XY) species, recent genome-wide expression studies in other taxa have revealed striking exceptions, especially in ZW taxa such as birds and snakes. In these taxa, the single Z of females is under expressed relative to males. These results fuel speculation that incomplete dosage compensation may be a defining characteristic of female-heterogamety. However, Lepidoptera (moths and butterflies) are also female-heterogametic, and evidence is accumulating that at least some species show balanced expression between sexes on the Z chromosome, contradicting the emerging consensus that ZW taxa lack complete dosage compensation.

Here we report on patterns of sex chromosome dosage compensation inferred from RNA-seq in two major lepidopteran lineages not yet surveyed: butterflies and Tortricid moths. In butterflies we focus on *Heliconius melpomene*, using a fully sequenced genome. Analyses in the Tortricid moth, *Cydia pomonella* (coddling moth), are based on *de novo* transcriptome assemblies. Results from both species show that average Z chromosome expression is significantly lower than autosomes in both sexes, similar to previous reports in bombycoid moths, and suggesting a novel mechanism of dosage compensation exists in the Lepidoptera. Notably, *C. pomonella* carries a neo-Z chromosome arising from the fusion of the ancestral Z with an autosome; this neo-Z also appears to be down-regulated in males. In *Heliconius*, but not *C. pomonella*, we detect a significantly greater global Z expression in males over females, indicating dosage compensation is imperfect in this species. However, the magnitude of this dosage effect is much less than the magnitude of reduced Z expression in both sexes. Patterns of sex-biased expression show an excess of male-biased genes on the Z chromosome, consistent with predictions from sexual antagonism theory.

Finally, we have extended the analysis of RNA-seq data from Kiuchi et al (Nature, 2014) that assayed genome-wide expression in *Bombyx mori* after RNAi knockdown of the sex-determining pathway. Our efforts confirm that knockdown of the masculinizing protein increases expression specifically on the Z chromosome in males, as previously reported. Additionally, we have uncovered a similar, though weaker, pattern in females that was not previously reported. These results further support the epigenetic down-regulation of the Z chromosome in males, but also suggests a similar mechanism may be operating in females in a dose-dependent manner.

# Metagenomics and Horizontal Transfer

**104 - MHT**
**Nested genomes: a hologenomic apporach to honey bee health**

Evans, Jay, Research Scientists
Bee Research Laboratory, USDA-ARS
Additional authors:
Schwarz, Ryan; USDA Bee Reseaarch Lab

Genomic resources are available for honey bees and each of the primary biological associates found in honey bee colonies, from viruses to mites and beetles. A hologenomic approach attempts to answer questions by leveraging genomic insights from each member of a biological community. This approach is being used to identify parasites and pathogens involved in bee declines (1,2), to characterize the gut microbiota (3) of bees, and to determine how strain variation within bees (4) and their associated biome (5) can play a role in bee health. This presentation will present current insights into the genetics of bees and their associates as well as future directions enabled by better tools and resources.

**105 - MHT**
**Host Cell Compensation for Genomic Deterioration of an Intracellular Bacterial Symbiont**

Luan, Junbo; Department of Entomology, Cornell University
Additional authors:
Chen, Wenbo; Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA
Shan, Hong-Wei; Ministry of Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Sciences, Zhejiang University, 866 Yuhangtang Street, Hangzhou 310058, China
Liu, Shu-Sheng; Ministry of Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Sciences, Zhejiang University, 866 Yuhangtang Street, Hangzhou 310058, China
Douglas, Angela E.; Department of Entomology, Cornell University, Ithaca, NY 14853, USA
Daniel K. Hasegawa, Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA Alvin M Simmons USDA-Agricultural Research Service, U.S. Vegetable Laboratory, Charleston, SC 29414, USA William M Wintermantel USDA-Agricultural Re

The genetic and evolutionary mechanisms underlying metabolic coadaptation between intracellular bacteria and host cell are poorly understood. One widely-spread whitefly Bemisia tabaci harbors the obligate symbiont Portiera with drastically degenerate genome and facultative symbiont Hamiltonella in bacteriocytes. Comparative genomic and transcriptomic approaches revealed that bacteriocytes are metabolically specialized to supply metabolites required by Portiera. Host cell exerts metabolic regulation of intracellular bacterial symbionts by provisioning precursors or terminal metabolites, compensating missing symbiont genes and metabolic duplication. Such metabolic interactions reflect the ongoing coevolution between symbionts and host. Hamiltonella was not integrated into host-Portiera interaction network by and large.

**106 - MHT**
**Does genetic meltdown of Carsonella symbionts leave *Eucalyptus* psyllids in need of new microbial partners?**

Morrow, Jennifer; Hawkesbury Institute for the Environment, University of Western Sydney
Additional authors:
Hall, Aidan; University of Western Sydney
Riegler, Markus; University of Western Sydney

Psyllids (Hemiptera) are plant-sap feeding insects that are obligately associated with their co-evolved primary symbiont, *Carsonella*. Such obligate symbioses are often exemplars of the phenomenon of extreme genome reduction, where gene loss has resulted in incomplete metabolic pathways.

Metagenomics shows that bacterial diversity in individual psyllid species is very low, but secondary symbionts, including *Wolbachia*, *Sodalis*-like and *Arsenophonus*-like bacteria have been detected and may play important complementary roles, e.g. in symbiont-based nutrient acquisition.

Psyllids of the genus *Cardiaspina* are lerp-forming *Eucalyptus* specialists, and recent devastating outbreaks of taxonomically not fully resolved *Cardiaspina* species in Australia highlighted the need for understanding their taxonomy, host plant associations and symbioses. Using genomics and transcriptomics we investigated the ecological and evolutionary relationships of a group of *Cardiaspina* spp. with their obligate symbiont *Carsonella*, and their dominant secondary symbionts. Targeting the bacterial symbionts housed in the bacteriome with NGS, the *Carsonella* genomes from six *Cardiaspina* host species associated with different *Eucalyptus* species, as well as other more distantly related gall-forming and free-living psyllids from Australia, were sequenced and assembled. Comparative genomics indicated, as reported in other psyllids, that genetic meltdown of the conserved *Carsonella* genome has occurred. Further analysis of the genomes of secondary symbionts, which most closely resemble *Arsenophonus* spp. in all but one of the six *Cardiaspina* psyllids, will demonstrate the extent metabolic complementarity enables co-occurring symbionts and host to thrive.

**107 - MHT**
**Host Plant Defenses Disrupt Insect-Microbial Interactions in Larval Asian Longhorned Beetles (Anoplophora glabripennis)**

Scully, Erin; Grain, Forage, and Bioenergy Research Unit, USDA-ARS
Additional authors:
Geib, Scott; USDA-ARS Tropical Crop and Commodity Protection Research Unit
Carlson, John; Pennsylvania State University
Tien, Ming; Pennsylvania State University
CJ, Tsai; University of Georgia
Scott, Harding; University of Georgia
Chen, Han-Yi University of Georgia Hoover, Kelli Pennsylvania State University

The Asian longhorned beetle (ALB; *Anoplophora glabripennis*) is an invasive, wood-boring pest capable of thriving in the heartwood of over 47 tree species worldwide where it faces a number of nutritional challenges, including digestion of lignocellulose and hemicellulose and acquisition of nitrogen, essential amino acids and nutrients that are present in low abundances in woody tissue. Through transcriptome sequencing, we have previously demonstrated that this insect possesses a rich repertoire of metabolic machinery that enables it to degrade major hardwood polysaccharides, such as cellulose, xylan, and pectin; detoxify host plant defensive compounds; recycle essential nutrients; and efficiently acquire protein and nitrogen from woody tissue or microbes that inhabit the gut. Furthermore, through metagenome and metatranscriptome sequencing efforts, we have also demonstrated that the taxonomically diverse gut microbiota encode diverse suites of genes that complement and augment ALB's endogenous physiological capacities, including the abilities to convert xylose sugars into compounds that can be directly utilized by ALB for the synthesis of fatty acids and amino acids, fix atmospheric nitrogen and recycle nitrogenous waste products, and synthesize several essential amino acids and nutrients that are present in low abundances in woody tissue. Further, it also encodes genes with the capacity to degrade large aromatic compounds and may collaborate with ALB to facilitate digestion of the lignin biopolymer. Thus, the metabolic potential of the gut community encodes an extensive suite of enzymes, which has been hypothesized to contribute to ALB's broad host range. Despite its broad host range, there are several tree species which display considerable resistance to ALB and other wood-boring pests, but the mechanisms underlying this resistance have not yet been characterized. In this study, we investigate the impacts of feeding in a resistant poplar tree on ALB and its gut microbiota, revealing that feeding in this resistant host causes substantial disruptions to the gut bacterial and fungal communities, interferes with the expression of beetle genes with predicted roles in detoxification and interactions with gut microbes, and reduces the abundances of proteins with key roles in digestion.

# Author Index (cross referenced with poster number)

| Poster # | Name | Last |
|---|---|---|
| 85 - PG | Virpi | Ahola |
| 1 - i5K/EG | Qasim | Alsouhail |
| 2 - i5K/EG | Nadia | Ayoub |
| 3 - i5K/EG | Anna | Bennett |
| 67 - MVG | Joshua | Benoit |
| 4 - i5K/EG | Joshua | Benoit |
| 48 - AVG | Krishna | Bhattarai |
| 97 - E | Anna-Maria | Botha-Oberholster |
| 5 - i5K/EG | Katie | Brown |
| 49 – AVG | Gregor | Bucher |
| 50 - AVG | Xiaolong | Cao |
| 98 - E | Mary | Chaffee |
| 6 - i5K/EG | Richard | Challis |
| 7 - i5K/EG | Crystal | Chaw |
| 4 (IGT-RCN) | Bonirath | Chhay |
| 51 - AVG | Fu-Chyun | Chu |
| 52 - AVG | Elizabeth | Cieniewicz |
| 8 - i5K/EG | Thomas | Clarke |
| 9 - i5K/EG | Christopher | Cunningham |
| 10 - i5K/EG | John | Davey |
| 68 - MVG | Victoria | Davidson |
| 69 - MVG | Shengzhang | Dong |
| 86 - PG | Eddy | Dowle |
| 70 - MVG | Scott | Emrich |
| 87 - PG | Scott | Emrich |
| 99 - E | Alexandra | Erwin |
| 104 - MHT | Jay | Evans |
| 88 - PG | Elizabeth | Everman |
| 53 - AVG | Honglin | Feng |
| 89 - PG | Philip | Freda |
| 90 - PG | Megan | Fritz |
| 71 - MVG | Cory | Gall |
| 11 - i5K/EG | Jessica | Garb |
| 12 - i5K/EG | Scott | Geib |
| 13 - i5K/EG | Maureen | Gorman |
| 54 - AVG | Nathaniel | Grubbs |
| 14 - i5K/EG | Julian | Gutekunst |
| 55 - AVG | Nicole | Gutzmann |
| 56 - AVG | M. Susan | Haas |
| 15 - i5K/EG | Robert | Haney |
| 16 - i5K/EG | Shawn | Hanrahan |
| 72 - MVG | Robert | Harrell |
| 57 - AVG | Yan | He |
| 58 - AVG | Xuesong | He |
| 73 - MVG | Matthew | Heerman |
| 17 - i5K/EG | Heather | Hines |
| 91 – PG | Carl | Hjelmen |
| 59 - AVG | Yingxia | Hu |

| Poster # | Name | Last |
|---|---|---|
| 18 - i5K/EG | Panagiotis | Ioannidis |
| 60 - AVG | Haobo | Jiang |
| 19 - i5K/EG | Tamsin | Jones |
| 74 - MVG | Silvia | Justi |
| 61 - AVG | Yonggyun | Kim |
| 75 - MVG | Donghun | Kim |
| 20 - i5K/EG | Boryana | Koseva |
| 76 - MVG | Lin | Kothera |
| 62 - AVG | Angela | Kruse |
| 21 – i5K/EG | Kwon | Hyung Wook |
| 22 - i5K/EG | Fabrice | Legeai |
| 63 - AVG | Sabrina | Lehmann |
| 77 - MVG | R. Rebecca | Love |
| 105 - MHT | Junbo | Luan |
| 12 (IGT-RCN) | Mary | Mills |
| 23 - i5K/EG | Taro | Mito |
| 100 - E | Yves | Mone |
| 106 - MHT | Jennifer | Morrow |
| 78 - MVG | Michelle | Ngai |
| 24 - i5K/EG | Jonathan | Oliver |
| 25 - i5K/EG | Brenda | Oppert |
| 79 - MVG | Justin | Overcash |
| 26 - i5K/EG | Alexie | Papanicolaou |
| 27 - i5K/EG | Mitulkumar | Patel |
| 80 - MVG | Ashley | Peery |
| 65 - AVG | Lindsey | Perkin |
| 66 - AVG | Patricia | Pinheiro |
| 28 - i5K/EG | Helen | Poynton |
| 101 - E | Kerry | Regan |
| 29 - i5K/EG | Hugh | Robertson |
| 81 - MVG | Angela | Saadat |
| 93 - PG | Adam | Schieferecke |
| 40 - i5K/EG | Derek | Schneweis |
| 94 - PG | Sean | Schoville |
| 107 - MHT | Erin | Scully |
| 41 - i5K/EG | Mark | Seeger |
| 95 - PG | Channing | Shives |
| 42 - i5K/EG | Sheina | Sim |
| 43 - i5K/EG | Felipe A. | Simao |
| 44 - i5K/EG | Daisuke | Takahashi |
| 102 - E | Aditi | Tayal |
| 82 - MVG | Laura | Vannini |
| 45 - i5K/EG | Iris | Vargas Jentzsch |
| 46 - i5K/EG | Jannelle | Vienneau-Hathaway |
| 103 - E | James | Walters |
| 47 – i5K/EG | Robert | Waterhouse |
| 83 - MVG | Rachel | Wiltshire |
| 96 - PG | Jihyun | Won |