Sequencing the Navel Orangeworm (NOW) Genome to Identify Genes Associated with Detoxification and Insecticide Resistance

Project No. **13-ENTO1A-Berenbaum**

Project Leader: May Berenbaum
Department of Entomology
320 Morrill Hall
University of Illinois, Urbana-Champaign
505 S. Goodwin
Urbana, IL 61801
217.333.7784
maybe@illinois.edu

Project Cooperators and Personnel:
Hugh Robertson, University of Illinois Urbana-Champaign
Joel Siegel, ARS/USDA - Parlier

Objectives:

We propose to sequence the genome of the navel orangeworm *Amyelois transitella* (NOW) using Illumina sequencing and manual annotation. Accomplishing these objectives will provide tools for opening up entirely new avenues for NOW management, including identifying new attractants for use in ovitraps, designing specific inhibitors for larval feeding and pesticide detoxification, and for monitoring resistance if it develops. In addition we will continue our research on characterizing cytochrome P450 genes that may potentially influence insecticide resistance, as well as our research on manipulating the rate of detoxification using phytochemicals present in almonds.

Interpretive Summary:

An organism's genome is the complete set of the DNA-based genetic instructions for every aspect of its existence. Sequencing an organism's genome provides not only powerful insights into its biology but also versatile tools for manipulating that organism; in the case of pest species, genomic tools can be used to develop new approaches for sustainable management. The availability of the NOW genome can enable scientists to determine which genes are likely responsible for pesticide metabolism, including those contributing to pesticide resistance, identify genes that encode enzymes in the larvae that allow them to feed on a wide range of hosts, and identify receptors that allow adult females to identify oviposition sites and that allow caterpillars to recognize and consume their host plants, thereby causing crop damage.

Materials and Methods:

We have sequenced the genome following a strategy pioneered by the Bejing Genomics Institiute (BGI) using ILLUMINA sequencing and their SOAPdenovo assembly program. Briefly, we generated an 180bp shotgun library from a single individual moth to reduce the level of polymorphism in our basic assembly of contigs. This helps with both avoiding multiple alleles assembling separately, which is undesirable, but also helps prevent microsatellites and

other major length polymorphisms from breaking the assembly. In the end it appears that the strain of NOW employed (laboratory strain SPIRL provided by Joel Siegel) is highly inbred and there is essentially no polymorphism in the population, which greatly assisted with the assembly. This library was sequenced in two lanes on a HiSeq2000 machine generating roughly 600m reads of 100 bases each and with paired-ends, yielding a total of approximately 36X coverage of the genome. This information was assembled into contigs that represent essentially all of the unique sequence in the genome. To connect these contigs into longer scaffolds of oriented and appropriately spaced contigs, we generated single lanes of sequence from an additional 1kb fragment length shotgun library, as well as 5, 10, and 20 kb mate pair libraries, all generated from DNA extracted from pooled individuals as these libraries require much more DNA. Assembly with these reads, plus use of a gap-closure algorithm that attempts to use this information to close gaps between contigs within scaffolds, led to our final assembly. We have evaluated this assembly using both manual methods by searching a small set of proteins against it using TBLASTN, and automated methods using the program BUSCO to assess the presence of over 1896 highly conserved 1:1 orthologs in insects, using the proteins of the monarch butterfly *Danaus plexippus* as queries in automated TBLASTN searches.

Results and Discussion:

The genome assembly consists of 424,988 contigs totaling 306Mbp arranged in 80,027 scaffolds totaling 423Mbp, with 116 Mbp of NNNNNs in the gaps between contigs within scaffolds. While it might appear that this is a large number of scaffolds, the vast majority of these ~~is~~are short and consists of short contigs with large gaps between them, and likely originates from the highly repetitive centromeric and peri-centromeric regions of the genome. A standard method to assess the quality of a genome assembly is the scaffold N50 statistic, which indicates the size of the scaffold above which half of the genome assembly is contained. This number for our assembly is 1,529 kb, which is excellent compared with most ILLUMINA-only insect genome assemblies. Thus, most of the genome is assembled in scaffolds larger than this size, with the largest scaffold being 9,749kb. TBLASTN searches of the assembly using 50 proteins representing some of the conserved chemoreceptors, vision proteins, clock proteins, and a set of very long neural proteins revealed that all of them were represented in single scaffolds. Similarly, the BUSCO analysis revealed that all 1896 highly conserved orthologs are present in the assembly, with 95% of them more than 95% long. Together these analyses indicate that the assembly is both virtually complete and of high quality.

Research Effort Recent Publications:

The next step is to submit this assembly for genome-wide automated annotation, which will be undertaken by our collaborator Mark Yandell at the University of Utah using his well-regarded and widely-used MAKER pipeline, as well as submit the assembly to GenBank where it will automatically be annotated using their GNOMON pipeline. We will choose the final set of gene models to employ for a genome paper, and also recruit colleagues to assist with manual corrections and improvements inevitably needed for families of genes of particular interest to us, such as the chemoreceptors and detoxification enzymes. Finally, a genome manuscript will be prepared and published.

References Cited:

Noble K, G Niu, J Siegel, M R Berenbaum, Pyrethroid tolerance of navel orangeworm (*Amyelois transitella*) after dietary exposure to almond phytochemicals. J Pest Science: in revision