

---

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

---

**Project No.:** 11-ENT01-Berenbaum

**Project Leader:** May Berenbaum  
Department of Entomology  
University of Illinois at Urbana-Champaign  
505 S. Goodwin Ave  
Urbana, IL 61801-3795  
217.333.2910  
maybe@illinois.edu

**Project Cooperators and Personnel:**  
Joel P. Siegel, USDA-ARS, Parlier CA

## Objectives:

- Sequence the navel orangeworm (NOW) genome using Illumina sequencing and manual annotation.
- In addition, continue research on characterizing cytochrome P450 genes that may potentially influence insecticide resistance, as well as research on manipulating the rate of detoxification using phytochemicals present in almonds.

## Progress:

The NOW genome sequence assembly has been finalized at 423 Mb in ~80,000 scaffolds with a scaffold N50 of 1.53 Mb. This appears to be an excellent assembly, but to assess the quality of the assembly further, in particular with regard to the likelihood of comprehensive genome-wide automated gene modeling; we performed a recently published test known as BUSCO. It attempts to model genes encoding 2,675 conserved 1:1 orthologous proteins across insects; it succeeded in building gene models for 82% of them, with fragmented models for another 12%, and 5% missing models (these 5% are not actually missing from the genome, they are perhaps too fragmented to be built by the BUSCO methods). These values compare well with several published insect genomes and are better than most of the 28 arthropod genomes recently sequenced for the i5k pilot project by the Baylor College of Medicine Human Genome Sequencing Center.

Accordingly, we have submitted the assembly to the NCBI (National Center for Biotechnology Information, accession number LACK00000000.1), along with all the supporting raw reads to the SRA (Short Read Archive), as well as the xanthotoxin-exposed midgut RNAseq reads. These have all recently been made publicly available, and now this genome is in the queue for genome-wide automated gene modeling using their GNOMON pipeline, which is one of the best in the world.

Once the gene models are available, we will submit the assembly, all the gene models, and all the supporting data from RNAseq to the i5k Workspace portal at the National Agriculture

Library, where it will be displayed using their WebApollo browser. This will allow interested members of the community to perform manual gene modeling and refinement, followed by analyses of the genome sequence and its gene content, all of which will eventually lead to a genome paper on NOW.