Report to DOE on AMNH Grant: Microbial Genomes/From Genomes to Life

Project 1. Microbial Genomes: A genomic approach to understanding the evolution of virulence. *Actinobacillus actinomycetemcomitans* and *Haemophilus aphrophilus* genome projects.

Overview: This project seeks to use the genomes of two close relatives, *A. actinomycetemcomitans* and *H. aphrophilus,* to understand the evolutionary changes that take place in a genome to make it more or less virulent. Our primary specific aim of this project was to sequence, annotate, and analyze the genomes of *Actinobacillus actinomycetemcomitans* (CU1000, serotype f) *and Haemophilus aphrophilus*. With these genome sequences we have then compared the whole genome sequences to each other and to the current Aa (HK1651 www.genome.ou.edu) genome project sequence along with other fully sequenced Pasteurellaceae to determine inter and intra species differences that may account for the differences and similarities in disease. We also propose to create and curate a comprehensive database where sequence information and analysis for the Pasteurellaceae (family that includes the genera Actinobacillus and Haemophilus) are readily accessible.

And finally we have proposed to develop phylogenetic techniques that can be used to efficiently and accurately examine the evolution of genomes. Below we report on progress we have made on these major specific aims. Progress on the specific aims is reported below under two major headings - experimental approaches and bioinformatics and systematic biology approaches.

Experimental Approaches: Our laboratories are now fully functional with respect to whole genome sequencing. Prior to DOE funding the AMNH Molecular Systematics Labs focused on sequencing multiple a priori chosen genes and genome regions. With the awarding of DOE funds, we have been able to upgrade the technology in the laboratory and allow the AMNH labs to become proficient in whole genome shotgun sequencing approaches. The DOE funds have allowed us to purchase a colony picker, a hydro shearer and provided funds to hire a postdoctoral fellow. The postdoctoral fellow (Dr. MariaPia DiBonaventura) has worked out protocols and optimizations for all shotgun sequencing methods and we have now generated nearly 3X coverage of the *Haemophilus aphrophilus* genome. We have initiated library construction for the *Actinobacillus actinomycetemcomitans* (CU1000, serotype f) genome. Clearly our objectives to initiate whole genome sequencing and to gear the AMNH labs up to this technology have been met.

Bioinformatics and Systematic Biology Approaches: The introduction of whole genome sequencing to the AMNH laboratories has required us to rethink bioinformatics tools and techniques. The installation of PHRED, PHRAP and a multitude of other programs to adequately assemble and annotate the genome level information. We have spent most of the past year installing these programs on our server here at the AMNH and in training personnel to use the software. In addition, we have developed several annotation tools that are based on phylogenetic approaches. The first tool is called ORFcurator (http://www.dbmi.columbia.edu/~ins7001/research/genomeCurator/ORFcurator/), a web based tool that searches completed and incomplete genomes for linkage or clusters of

genes. The second tool is also web based and it is called orthologID (http://orthologid.amnh.org; Username: plant; Password: polytomies). This program uses phylogenetic principles to establish orthology statements in simple genomes like bacterial genomes and can also be extended to more complex eukaryotic genomes. Finally we have built a comprehensive database of all known finished and incomplete genomes of microbial species from a diverse array of genome centers (Table 1).

Table 1. Source genome centers for the AMNH bacterial whole genome data base.

Joint Genome Institute/Department of Energy (JGI/DOE)
www.jgi.doe.gov
National Center for Biotechnology Information (NCBI)
www.ncbi.nlm.nih.gov
University of Ohio
www.microbial-pathogenesis.org
University of Illinois
www.salmonella.org
Sanger Institute
www.sanger.ac.uk
The Institute for Genomic Research (TIGR)
www.tigr.org
University of Oklahoma
www.genome.ou.edu
Stanford University
sequence-www.stanford.edu
Genscope, Inc.
www.genoscope.cns.fr
University of Minnesota
www.cbc.umn.edu
and are currently exploring this data base

Project 2. From Genomes to Life: *Drosophila* development in space and time.
Overview: In the original proposal, we propose to utilize the *Drosophila* embryo as a platform for determining the interaction of several genes in a gene regulatory network. In particular we propose to: 1) Refine techniques for embryo staining using polyclonal antibodies across a broad phylogenetic perspective within the Drosophilidae. 2) Obtain staining patterns for 14 development genes (see below) in at least ten *Drosophila* species (see below). 3) Determine the gene regulatory circuitry involved in the regulation of early development for these the ten *Drosophila* species. 4) Using a phylogenetic approach we intend to apply the gene circuit method to organismal diversity to address how gene interactions change through time and the role these changes may or may not have in the evolution of morphological diversity. We separate report on progress on this project by describing the experimental approaches that have been imported to the museum and implemented by scientists funded by the DOE grant.

Experimental approaches:
To date we have accomplished the following experimental goals. We have explored culturing and manipulating a broad array of family Drosophilidae species (*Chymomyza amoena, Scaptodrosophila pattersoni, Zaprionus sp., D. albimicans, D. virilis, D. pseudoobscura, D. simulans, and D. grimshawi*). We have also developed some interesting and effective methods of culturing some of the more difficult to culture species as well as new methods for obtaining embryos for antibody staining. Along the way we have collected at least five thousand staged embryos for most of the species listed above so that we can move on to the antibody staining phase of the project. We have also discovered that most of the antibodies that we have in hand work quite well with four phylogenetically diverse species (*Chymomyza, Zaprionus, D. pattersoni and D. melanogaster*). These antibodies are: Hunchback, Bicoid, Caudal, Kruppel, Knirps, Giant, Tailless, Even-Skipped, Fushi-tarazu, Hairy, Odd-skipped Paired, Sloppy-paired, and Runt. We have also performed "staging" analyses of several of the species listed above. One of the problems with comparing the patterns we are generating using antibodies is that different stages give different patterns so to make a valid comparison across taxa, we need to determine that they are in the same embryonic stage. We have accomplished this staging using Nomarski optics and confocal optics. We are confident that the developmental staging is accurate and that the interpretation of the antibody staining approaches will be valid. A summary of the results from this exploratory research can be accessed on the following website.

http://research.amnh.org/molecular/Arnold's%20web%20site.htm

In addition to the fly work we have also expanded to annelid development with the addition of a postdoctoral fellow (Dr. Al Phillips) whose specialty is leech neural development. Dr. Phillips' work dovetails nicely with the fly work and many of the techniques. In particular, Dr. Phillips has set up an injection facility (using AMNH setup funds) for making transgenic embryos and doing RNAi experiments with these harder to culture Drosophilidae species.