

U.S. Database Coordination Activities

*Supported by Allotments of Multi-State Research Funds, Hatch Act
For the Period 1/1/05-12/31/05*

Overview: Coordination of Database/Bioinformatics under the National Animal Genome Research Program (NAGRP) is an effort of Iowa State University (ISU). CSREES support is allocated via NRSP-8. The NAGRP is made up of the membership of the Animal Genome Technical Committee.

FACILITIES AND PERSONNEL: James Reecy, Department of Animal Science, ISU, serves as Coordinator along with Sue Lamont, Max Rothschild and Chris Tuggle, Department of Animal Science, ISU, as Co-Coordiators. Iowa State University provides facilities and support.

OBJECTIVES: 1. Develop high-resolution comparative genome maps aligned across species that link agricultural animal maps to those of the human and mouse genomes, 2. Increase the marker density of existing linkage maps used in QTL mapping and integrate them with physical maps of animal chromosomes, and 3. Expand and enhance internationally shared species genome databases and provide other common resources that facilitate genome mapping.

PROGRESS TOWARD OBJECTIVE 3: Database and other map resources.

Researchers at universities and other research institutions are conducting multifaceted research to develop bioinformatics programs and database resources for livestock species and this research is supported in part by the NAGRP. Continued efforts to inform scientists and lay persons about genome databases have been made and many new entries are now available at www.animalgenome.org. The NAGRP genome databases were accessed over 2.2 million times by over 140,000 users world-wide.

QTL database. The NAGRP database has a newly developed Porcine QTL database that graphically displays QTL from over 70 experiments and can be used at <http://www.animalgenome.org/QTLdb/>. All information has been cross-listed at NCBI and can be viewed at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&term=pig+QTL>. Over the past year, links have been added to the viewer allowing researchers to visualize QTL on the human genome. In addition, we will expand this visualization option to additional sequenced organisms, such as the mouse, rat and dog. The database can now be used for any species. We are currently working with the Aquaculture community to enter their data into the QTL database.

PCR Primer Design. A bioinformatics program (Expedito) was developed to design primers for livestock species. This program takes advantage of the information from the human genome and applies it to livestock cDNA. This program can be used at <http://www.animalgenome.org/~hu/expeditor>.

Blast Analysis of Livestock and Aquaculture DNA sequence. Programs were developed to automatically download livestock sequence data from and synchronize them with NCBI. The available sequence databases include: (1) EST sequence from all non-human, non-mouse animals; (2) Whole genome shotgun (WGS) sequences from all livestock animals; (3) TIGR gene indices for 6 species, and can be used at <http://www.animalgenome.org/blast/>.

Phenotype Ontology Editor. With the dramatic increase in sequence information in livestock and aquaculture species, it will be imperative to be able to link phenotype, genotype, proteomic, and gene expression data in a queryable format. Development of a formal phenotype ontology will complement on-going gene and anatomy ontology efforts. However, due to the small size of the livestock and aquaculture species communities, it will be important to do so in a cooperative manner. Toward this end, we have developed an ontology editor that can be used simultaneously by several annotators. We anticipate that this resource will be shortly available to the livestock community.

Porcine EST Cluster and Annotation Analysis. Over the course of the past year, funds were used to support work by Dr. Chris Elsik's lab (Texas A&M) to cluster and annotate porcine EST clusters. This work was to support the development of a new porcine long-oligo array in collaboration with the Swine Genome Coordinator.

Genetic Program Database. With the rapid progress in genetics and genomes, there are an increasing number of genetic analysis software programs. Each program has its own pre-defined scope, assumptions, and applicability. With the large number of available programs, it can be a challenge to identify suitable options. Thus, we have created a database and related tools to effectively archive, annotate, and manage the wealth of the software information so that researchers can easily identify, locate, and retrieve appropriate software. We have also introduced ontology concepts and tools to manage the proper classification and feature annotation of this resource. To date, there are 331 software programs listed in the category of “genetic analysis”. We plan to add other genomics and computational biology software in the near future. <http://www.animalgenome.org/soft>

Database Activities: As in past years, the Livestock Genome Databases have received considerable updating.

ArkDB statistics (December, 2005)

	cattle	pig	chicken	sheep	horse	salmon	tilapia
loci	2725	4081	2530	2030	1470	230	243
(genes)	746	1588	765	543	465	10	1
(linkage)	2209	5141	3400	14909	0	0	163
(cytogenetic)	659	1927	307	843	688	0	0
clones	1	602	316	206	153	27	0
library	0	96	32	0	5	3	0
experiments	6020	5680	2529	2934	1851	293	135
references	509	1254	573	518	217	86	243

These activities help to promote cooperation and facilitate progress in livestock genomics.

Newsletter: The Bioinformatics Newsletter is published and distributed through our Homepage, and electronically on the AnGenMap email discussion group.

Meetings: In the past year, we helped support the Third International Symposium on Genetics of Animal Health, July 13-15, 2005, at Iowa State University. There were over 120 participants from 14 countries. Many livestock and aquaculture scientists attended the joint Plant and Animal Genome XIII meeting held last January, held jointly with the annual NAGRP meeting. Coordination funds helped support attendance at PAG-XIII and will do so again for the upcoming PAG-XIV in January, 2006.

Future Activities. Suggestions from researchers to help this coordination and facilitation program grow and succeed are always appreciated. (Please send them to jreecy@iastate.edu). We will continue our efforts on QTL database curation and expand these efforts to additional species. To aid in this effort we are developing a submission editor so that labs can directly deposit QTL data into the database. Furthermore, we will link new maps and sequence information to QTL as new information becomes available. In addition, we will continue phenotype ontology development, BLAST tools development, and software curation. In the next year, we will develop a genome sequence assembly program that will allow researchers to develop genomic contigs in species with incomplete genome sequence information.

There is an incredible need for species-specific genome databases, which can serve as central repositories for integrated genomic data. For instance, as livestock genomes are sequenced there is a preliminary level of annotation that will be completed by Ensembl and NCBI, but there is not consensus on what annotation is correct. Furthermore, continued annotation efforts are necessary to realize the full power of the genomic sequence. Although these efforts are well beyond the scope of the currently funded NRSP-8 project, the database coordination program plans to support community efforts toward this end. The recent USDA-NRRCGP RFA presents a wonderful opportunity in its call for bioinformatics proposals. However, the funds available to support this effort are not adequate to meet the immediate needs of the avian, bovine, and porcine communities.