U.S. Database Coordination Activities
Supported by Allotments of Regional Research Funds, Hatch Act
For the Period 1/1/06-12/31/06

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, including the Database Subcommittee.

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

OBJECTIVES: 1. Enhance and integrate genetic and physical maps of agriculturally important animals for cross species comparisons and sequence annotation, 2. Facilitate integration of genomic, transcriptional, proteomic and metabolomic approaches toward better understanding of biological mechanisms underlying economically important traits, and 3. Facilitate and implement bioinformatic tools to extract, analyze, store and disseminate information.

PROGRESS TOWARD OBJECTIVE 1. Enhance and integrate genetic and physical maps of agriculturally important animals for cross species comparisons and sequence annotation.

The genetic linkage maps of the cattle, chicken and swine have provided a framework for numerous QTL and other mapping experiments and a platform on which genome sequences have been assembled and linked to chromosomes. Over the past year, we have worked to link QTL data to the human and livestock genome sequences, to allow researchers to seamlessly transfer information between maps (http://www.animalgenome.org/QTLdb/). In addition, we have included SNP minor allele frequencies and microarray features (where available).

PROGRESS TOWARD OBJECTIVE 2. Facilitate integration of genomic, transcriptional, proteomic and metabolomic approaches toward better understanding of biological mechanisms underlying economically important traits.

Over the past decade, several hundred manuscripts have been published in which quantitative trait loci or association tests have been reported for all livestock species. We have focused on the curation of cattle, chicken, and swine QTL information (http://www.animalgenome.org/QTLdb/). In the past year with the help of Dr. Jill Maddox (Australia) we were able to expand the database to include Sheep QTLs.

PROGRESS TOWARD OBJECTIVE 3: Facilitate and implement bioinformatic tools to extract, analyze, store and disseminate information.

Efforts under this objective included communications with curators of other relevant databases, compilation of information about those databases, assessment of the content and function of those databases, and prioritization of the efforts of US coordination efforts in the areas of highest priority and utility, given the landscape of public databases already developed and maintained by others. The following described the publicly available resources, and the Coordinator's activities.

The chicken genome sequence, along with a variety of options and tools, can be accessed at three different browsers: the UCSC Chicken Genome BrowserGateway, (http://genome.ucsc.edu/cgi-

bin/hgGateway?org=Chicken&db=0&hgsid=30948908); the NCBI Chicken Genome Resources, (http://www.ncbi.nlm.nih.gov/genome/guide/chicken /); and the EBI's Ensembl Chicken Genome Browser, (http://www.ensembl.org/Gallus_gallus /). The ChickFPC browser at http://www.bioinformatics.nl/gbrowse/cgi-bin/gbrowse/ChickFPC allows for various searches of the BAC contig map. Similarly, BAC locations denoted by BAC end sequences can be found on other sequence browsers noted above. The SNP data generated by the Beijing Genomics Institute (described above) can be accessed on the UCSC or Ensembl browsers, but more extensive descriptions are available at the BGI site at http://chicken.genomics.org.cn/index.jsp . Information on 657 Chicken QTL can be found at http://www.animalgenome.org/QTLdb/chicken.html.

The latest version of ChickGBASE developed by the Roslin Institute is at http://www.thearkdb.org/arkdb/do/getChromosomeDetails;jsessionid=B8A6A5EA698B84AF80EE99BE7530B04E?accession=ARKSPC00000004 . US Poultry Genome coordinator maintains a homepage for the NRSP-8 U.S. Poultry Genome project (http://poultry.mph.msu.edu) that provides a variety of genome mapping resources, including the latest EL maps and mapping data, descriptions of available resources, the latest cytogenetic map, and access to a host of other information relating to both genetic and physical maps, including our newsletter archive.

Recently, Carl Schmidt (Univeristy of Delaware) started Birdbase.net (http://birdbase.net/), which contains links to many useful tools for the bird community.

Roslin Institute maintains the Arkdb, that hosts linkage maps for a number of livestock species (http://www.thearkdb.org/), include pigs, cattle, chicken, sheep, and others.

AgBase developed by Mississippi State University contains information from their active annotation of Gene Ontologies for cattle (http://www.agbase.msstate.edu/information/cowgo.html), pigs (http://www.agbase.msstate.edu/information/PigGO.html), chicken (http://www.agbase.msstate.edu/information/chickgo.html), sheep (http://www.agbase.msstate.edu/information/sheepgo.html) and several aqua culture species, such as catfish, trout, and salmon (http://www.agbase.msstate.edu/).

Cattle sequence, along with a variety of options and tools, can be accessed at three different browsers: the UCSC Cow Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway?org=cow), the NCBI Cow Genome Resources (http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow/), and the EBI's Ensembl cattle Genome Browser (http://pre.ensembl.org/Bos_taurus/index.html). In addition, cattle genome browser is set up at Texas A&M to aid the annotations at http://bovinegenome.org/ and in Australia (http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/bova3/). Cattle BAC contigs can be visualized at http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/baccontig/. Ruminant-human genome comparison can be visualized at http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/rumhum/. Bovine SNPs can be visualized at http://www.livestockgenomics.csiro.au/ibiss/. Cattle QTL information can be found at 3 databases: Texas (http://bovineqtl.tamu.edu/), Iowa (1125 curated QTLs; http://www.animalgenome.org/QTLdb/cattle.html) and Australia (http://www.vetsci.usyd.edu.au/reprogen/QTL_Map/).

1.8 million cattle SNPs from the bovine genome sequencing project have been aligned to cattle QTL maps. URL: http://www.animalgenome.org/QTLdb/cattle.html

NCBI announced the release of its annotation of the bovine genome assembly, build 3.1. This build includes annotation of the Btau_3.1 genome assembly from Baylor and a complete mitochondrial genome. The data can

be viewed in Map Viewer: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9913 downloaded by ftp:ftp://ftp.ncbi.nlm.nih.gov/genomes/Bos_taurus/ and searched by BLAST: http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9913

A guide to bovine genome resources is available at: http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow/ Please note two new items in this release: 1 -The Map Viewer includes a new BAC-end sequence map for build 3.1 (BES clone), e.g.:http://tinyurl.com/y3mehr 2 - The previous build, 2.1 based on Btau_2.0, is still available for display in Map Viewer: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9913&build=prev

The pig genome sequencing is actively carried out at Sanger Institute (http://www.sanger.ac.uk/Projects/S_scrofa/) and preliminary sequence results can be found at Ensembl site (http://pre.ensembl.org/Sus_scrofa/index.html) and regularly updated into the NCBI database (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9823). Pig QTL information has been actively updated at the AnimalQTLdb (1675 curated QTLs; http://www.animalgenome.org/QTLdb/pig.html).

Additional 838 porcine microsatellites with repeat motives of three to six bases have been added to the Pig QTLdb for alignment. This is an in addition to the previously aligned 4,528 porcine microsatellite markers. Thanks to Peter Karlskov-Mortensen and Merete Fredholm from the Royal Veterinary and Agricultural Univ. of Denmark for providing the data. URL: http://www.animalgenome.org/QTLdb/tmp/03219.html

A preliminary experiment to align and display microarray expression data appears to have been successful. Thanks to Shu-hong Zhao from Huazhong University for providing testing data.URL: http://www.animalgenome.org/QTLdb/pig.html

Sheep genome information can be found at several databases: NCBI sheep genome resources (http://www.ncbi.nlm.nih.gov/genome/guide/sheep/), International Sheep Genomics Consortium (http://www.sheephapmap.org/). Sheep BAC clone and FPC information can be found at http://bacpac.chori.org/library.php?id=162 , and sheep virtual genome at http://www.sheephapmap.org/isgc_virtualgenome.htm. Information on 57 Sheep QTL can be found at http://www.animalgenome.org/QTLdb/sheep.html.

A number of on-line tools have been developed http://www.animalgenome.org/bioinfo/tools/. Tools have been developed to generate PCR primers across livestock exon or between two conserved sequences. A Gene Ontology term counter has been developed to aid in the determination of the number GO parent and sibling numbers. Graph drawing tools can also be accessed as well.

This past year we focused on the development of an Animal Trait Ontology (http://www.animalgenome.org/bioinfo/projects/ATO/). As a result, for the first time individuals have an on-line resource where they can find standardized trait terms, which will help to improve communication among different groups within the livestock community. Anyone interested in helping to improve the ATO is encouraged to contact either James Reecy (jreecy@iastate.edu) or (zhu@iastate.edu).

We hosted a bovine genome annotation meeting in conjunction with Dr. Chris Elsik at PAG on January 15th. In addition, we will hosted a QTL database meeting to discuss minimal information necessary for publication of QTL data, development of a phenotype ontology, and sharing of curated data. Efforts are underway to publish the results of this meeting and subsequent discussions with interested individuals. A copy of the manuscript is available to any interested parties. We hosted a bovine genome annotation training session at Iowa State in August, 2007. Furthermore, we participated in a Bovine genome Jamboree in June, 2007 to manually annotate the bovine genome.

We have developed and will shortly release a tool that allows researchers to compare the overlap between QTL within a given species of interest. Researchers can identify regions of overlap between different QTL for a given trait of interest or they can identify regions of overlap between QTL for two traits of interest.

A new linux cluster computer with 8 nodes, and a total of 16 processors was purchased and brought on-line. We have installed the mpiblast, and developed web tools to accept, queue blast jobs and send results to users. This is a great boost to the blast capacity we have. URL: http://www.animalgenome.org/blast/mpiblast.html. To make better use of its computing power for the community, we seek input from the NRSP-8 community.

Meetings: Over 2000 scientists attended the joint Plant and Animal Genome XIV meeting held last January, held jointly with the annual NAGRP meeting. Coordination funds helped support attendance at PAG-XIV and will do so again for the upcoming PAG-XV in January, 2008.

PLANS FOR THE FUTURE.

OBJECTIVE 1. Enhance and integrate genetic and physical maps of agriculturally important animals for cross species comparisons and sequence annotation.

We successfully obtained USDA-NRI funding to develop a web based tool that will allow interested individuals to compare QTL across cattle, chickens, human, mouse, rat, pigs and sheep. This effort is a collaborative effort with researchers at the Rat Genome Database (Medical College of Wisconsin) and the University of Iowa. In addition, this grant will allow us to expand our efforts in curation of QTL, gene association and trait information.

OBJECTIVE 2. Facilitate integration of genomic, transcriptional, proteomic and metabolomic approaches toward better understanding of biological mechanisms underlying economically important traits.

We will expand the QTL database to house individual animal genotype and phenotype data. This will help research labs that lack expertise with relational databases and will facilitate comparison of QTL information across experiments. In addition, we will work to expand the QTL database to house microarray data, which will facilitate the identification of candidate genes when researchers are searching for causal mutations.

OBJECTIVE 3: Facilitate and implement bioinformatic tools to extract, analyze, store and disseminate information.

We will continue to work with bovine, mouse, rat, and human QTL databases to develop minimal information for publication standards. In addition, we will work with these same databases to improve a phenotype ontology, which will facilitate transfer of QTL information across species. As a result of successfully obtaining USDA-NRI funding NRSP-8 funding will be expanded to developing and hosting of relational databases for NRSP-8 members. We will provide knowledge, labor and hardware toward this effort.

(Prepared 1/05/08)