

**U.S. Database Coordination Activities**  
Supported by Allotments of Regional Research Funds, Hatch Act  
For the Period 1/1/08-12/31/08

**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:** The NRSP-8 project was renewed as of 10/01/08. As a majority of the reporting period was within the previous project, the objectives for that project are listed. 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/>). With the help of Dr. Jill Maddox (Australia), we have been able to expand the Sheep QTL database.

**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 describes some 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&hgid=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/](http://www.ensembl.org/Gallus_gallus/)). 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>. Chicken QTL can be found at <http://www.animalgenome.org/QTLdb/chicken.html>.

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 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 (University of Delaware) started Birdbase.net (<http://birdbase.net/>), which contains links to many useful tools for the bird community.

Cattle sequence (Build 4), 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://www.ensembl.org/Bos\\_taurus/Info/Index](http://www.ensembl.org/Bos_taurus/Info/Index)). In addition, cattle genome browser is set up at Georgetown University to aid gene annotation at <http://genomes.arc.georgetown.edu/bovine/> and in Australia <http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/bova4gff3/>. Bovine and sheep SNPs can be visualized at <http://www.livestockgenomics.csiro.au/ibiss/>. Alternatively, an independent assembly of the bovine genome can be obtained at [ftp://ftp.cbcb.umd.edu/pub/data/Bos\\_taurus/Bos\\_taurus\\_UMD\\_2.0/](ftp://ftp.cbcb.umd.edu/pub/data/Bos_taurus/Bos_taurus_UMD_2.0/). Cattle QTL information can be found at 3 databases: Texas (<http://genomes.sapac.edu.au/bovineqtl/>), Iowa (<http://www.animalgenome.org/QTLdb/cattle.html>) and Australia ([http://www.vetsci.usyd.edu.au/reprogen/QTL\\_Map/](http://www.vetsci.usyd.edu.au/reprogen/QTL_Map/)). Bovine Hapmap information can be obtained at [http://bfgl.anri.barc.usda.gov/cgi-bin/hapmap/affy2/m\\_session.pl](http://bfgl.anri.barc.usda.gov/cgi-bin/hapmap/affy2/m_session.pl).

The pig genome sequencing is actively carried out at Sanger Institute ([http://www.sanger.ac.uk/Projects/S\\_scrofa/](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](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](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9823)). Updated pig genome sequencing information can be found at <http://www.animalgenome.org/pigs/genomesequence/>. Pig QTL information has been actively updated at the Animal QTLdb (<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.livestockgenomics.csiro.au/perl/gbrowse.cgi/vsheep2/>. Information on Sheep QTL can be found at <http://www.animalgenome.org/QTLdb/sheep.html>.

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

*Ontology development* - 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. See Hughes et al. (2008) for additional information. Anyone interested in helping to improve the ATO is encouraged to contact James Reecy (jreecy@iastate.edu), Cari Park (caripark@iastate.edu) or Zhiliang Hu (zhu@iastate.edu). We are collaborating with researchers at INRA (France) and within EADGENE and SABRE, EU funded projects, to expand the utility of the ATO. Furthermore, we are working with MGI and RGD to develop a Vertebrate Trait Ontology, which will facilitate information transfer across species. Finally, we are working to develop a livestock breed ontology based on the Oklahoma State University Livestock Breeds web resource.

*Software development* – A number of on-line tools have been developed <http://www.animalgenome.org/bioinfo/tools/>. We develop a bioinformatic pipeline to assemble genome sequence using a seed sequence. The program utilizes BLAST and CAP3 to reiteratively retrieve genome sequence and assemble it. As a proof of principle, we were able to assemble bovine genes with as little as 3X genome coverage (Koltjes et al., 2008). We also develop a web-based program to help investigators categorize Gene Ontology terms. See Hu et al. (2008) for additional information. In addition, we develop new computational methods to identify GO terms that are differentially expressed. See Nettleton et al. (2008) for additional information. Graph drawing tools can also be accessed as well.

*Minimal standards development* – We have worked with the MIBBI project [http://www.mibbi.org/index.php/Main\\_Page](http://www.mibbi.org/index.php/Main_Page) to help define minimal standards for publication of QTL and gene association data (<http://miqas.sourceforge.net/>). See Taylor et al. (2008) for additional information.

*Expanded Animal QTLdb functionality* – All bovine and chicken QTL data can be viewed against respective genomes at <http://www.animalgenome.org/cgi-bin/gbrowse/cattle/> and <http://www.animalgenome.org/cgi-bin/gbrowse/chicken/>. Tabulated QTL data for pigs, cattle, and chicken can now be downloaded from each respective chromosomal view page. Once the swine genome is completed, we will map all swine QTL to the genome sequence and make it publicly available as well.

*Facilitating research* – Throughout the year, we helped many research groups with their research projects. Our involvement has ranged from data transfer to data assembly to data analysis. Please continue to contact us as you need help with bioinformatic issues.

**Meetings:** Over 2000 scientists attended the joint Plant and Animal Genome 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 in January, 2009.

## **PLANS FOR THE FUTURE.**

**OBJECTIVE 1.** 1) Create shared genomic tools and reagents and sequence information to enhance the understanding and discovery of genetic mechanisms affecting traits of interest.

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 is a collaborative effort with researchers at the Rat Genome Database (Medical College of Wisconsin) and the University of Iowa. A comparative QTL viewer is under development (<http://bioneos.com/VCMMap/>). In addition, this grant will allow us to expand our efforts in curation of QTL, gene association and trait information. Next year, we will participate in an effort to standardize gene nomenclature across species.

**OBJECTIVE 2.** Facilitate the development and sharing of animal populations and the collection and analysis of new, unique and interesting phenotypes.

We are working with the PRRS CAP Host Genome consortium to develop a relational database to house individual animal genotype and phenotype data (<http://www.animalgenome.org/lunney/index.php>). This will help the consortium whose individual research labs lack expertise with relational databases share information across among consortium members

**OBJECTIVE 3:** Develop, integrate and implement bioinformatics resources to support the discovery of genetic mechanisms that underlie traits of interest.

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. In addition, we will work to expand the QTL database to house microarray data (<http://www.anexdb.org/>), which will facilitate the identification of candidate genes when researchers are searching for causal mutations. We will work to develop a Bioinformatics Blueprint, similar to the Animal Genomics Blueprint recently published by USDA-CSREES, to help direct future livestock oriented bioinformatic/database efforts.

### **Publications**

- Koltjes, J.E., Z-L. Hu, E.Fritz, J.M. Reecy. 2008. BEAP: The BLAST Extension and Alignment Program- a tool for contig construction and analysis of preliminary genome sequence. BMC Research Notes (accepted).
- Hu, Z-L., J. Bao, J.M. Reecy. 2008. "CateGORizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories". Online Journal of Bioinformatics 9(2):108-112.
- Taylor, C.F., D. Field, S.-A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C. A. Ball, P.-A. Binz, M. Bogue, T. Booth, A. Brazma, R. R. Brinkman, A. M. Clark, E. W. Deutsch, O. Fiehn, J. Fostel, P. Ghazal, F. Gibson, T. Gray, G. Grimes, N. W. Hardy, H. Hermjakob, R. K. Julian, Jr., M. Kane, C. Kettner, C. Kinsinger, E. Kolker, M. Kuiper, N. Le Novère, J. Leebens-Mack, S. E. Lewis, P. Lord, A.-M. Mallon, N. Marthandan, H. Masuya, R. McNally, A. Mehrle, N. Morrison, S. Orchard, J. Quackenbush, J.M. Reecy, D. G. Robertson, P. Rocca-Serra, H. Rodriguez, H. Rosenfelder, J. Santoyo-Lopez, R. H. Scheuermann, D. Schober, B. Smith, J. Snape, K. Tipton, P. Sterk, A. Untergasser, J. Vandesompele, S. Wiemann. 2008. Promoting Coherent Minimum Reporting Requirements for Biological and Biomedical Investigations: The MIBBI Project. Nature Biotechnology 26(8):889-96.
- Li, P., Z-L. Hu, S. J. Moon, K. T. Do, Y. K. Ha, H. Kim, M. J. Byun, B. H. Choi, M. F. Rothschild, J.M. Reecy, K. S. Kim. 2008. Development of an in silico Coding Gene SNP map in pigs. Animal Genetics 39(4):446-50.
- Hughes, L.M., J. Bao, Z-L. Hu, V. Honavar, J. M. Reecy. 2008. Animal Trait Ontology (ATO): The importance and usefulness of a unified trait vocabulary for animal species. Journal of Animal Science. 86(6):1485-91.
- Nettleton, D., J. Recknor, J.M. Reecy. 2008. Identification of Differentially Expressed Gene Categories in Microarray Studies Using Nonparametric Multivariate Analysis. Bioinformatics 24(2):192-201. Epub 2007 Nov 27
- Karlskov-Mortensen P., Z-L. Hu, J.M. Reecy, M. Fredholm. 2008. A data resource of 838 porcine microsatellite sequences with repeat motifs of three to six bases. Animal Genetics 39(1):85-6. Epub 2007 Nov 1.

(Prepared 1/02/09)