

**U.S. Database Coordination Activities**  
Supported by Allotments of Regional Research Funds, Hatch Act  
For the Period 1/1/09-12/31/09

**OVERVIEW:** Coordination of the CSREES National Animal Genome Research Program's (NAGRP) Bioinformatics is primarily based at, and led from, Iowa State University (ISU) and is supported by 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 Susan J. Lamont, Max Rothschild, Chris Tuggle, and Shane Burgess as Co-Coordinators. Iowa State University provides facilities and support.

**OBJECTIVES:** The NRSP-8 project was renewed as of 10/01/08, with the following objectives: 1. Create shared genomic tools and reagents and sequence information to enhance the understanding and discovery of genetic mechanisms affecting traits of interest; 2. Facilitate the development and sharing of animal populations and the collection and analysis of new, unique and interesting phenotypes; and 3. Develop, integrate and implement bioinformatic resources to support the discovery of genetic mechanisms that underlie traits of interest.

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

In a separate project, the Tuggle group, in collaboration with others including James Reecy, have developed a public open-source database and website ([www.ANEXdb.org](http://www.ANEXdb.org)) for storage and analysis of functional genomics data in livestock (Couture et al., in press). During 2009, the Database Coordination team has begun efforts to migrate this database into [www.animalgenome.org](http://www.animalgenome.org) for long-term maintenance and to expand its capabilities, which currently consist of: a) tools to facilitate storage of Affymetrix-based gene expression data from any species with an available GeneChip<sup>®</sup>; b) submission of user data to NCBI Gene Expression Omnibus; and c) a comprehensive annotation of all porcine expressed sequences. The current efforts will expand the data storage capabilities to all widely used livestock gene expression profiling tools, as well as create comparative annotation of the sequence elements on these profiling tools.

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

Over the past year, we have partnered with researchers at Kansas State University, Michigan State University and U.S. Department of Agriculture to develop relational databases to store and disseminate phenotypic and genotypic information from large genomic studies in farm animals. For example, 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 among consortium members and thereby facilitate data analysis.

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

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 U.S. 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.

### Poultry

The chicken genome sequence, along with a variety of options and tools, can be accessed at three different browsers: the UCSC Chicken Genome Browser Gateway (<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/](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>.

The U.S. 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.

In response to a request from the NRSP-8 avian community at PAG 2009, a team lead by Parker Antin (U. Arizona), Shane Burgess and Carl Schmidt (U. Delaware) have developed a draft Avian Model Organism Database (MOD) called "Birdbase" (<http://birdbase.net/>). Birdbase will be released for trial by all NRSP-8 members at the PAG meeting in January 2010. This MOD could also serve as an example for the other species within NRSP-8 that also lack MODs, especially the aquaculture community, because of the similarities in systems and species diversity.

### Cattle

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, a cattle genome browser is set up at Georgetown University to aid gene annotation (<http://genomes.arc.georgetown.edu/bovine/>) and in Australia (<http://www.livestockgenomics.csiro.au/cow/>). 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/](ftp://ftp.cbcb.umd.edu/pub/data/Bos_taurus/). Cattle QTL information can be found at 3 databases: University of Adelaide (<http://genomes.sapac.edu.au/bovineqtl/>), Iowa State University (<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).

### Porcine

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 latest sequence assembly and genome annotation results can be found at the Ensembl site ([http://www.ensembl.org/Sus\\_scrofa/Info/Index](http://www.ensembl.org/Sus_scrofa/Info/Index)) 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 AnimalQTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>).

In the past year, we participated in the development of the Swine SNPchip (See Ramos et al., 2009 for full details).

### Sheep

Sheep genome information can be found at two databases: NCBI sheep genome resources (<http://www.ncbi.nlm.nih.gov/genome/guide/sheep/>) and 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/sheep/>. Information on sheep QTL can be found at <http://sphinx.vet.unimelb.edu.au/QTLdb/>.

### Aquaculture

Many useful links for aquaculture can be found at <http://www.animalgenome.org/aquaculture/>. Salmon genome resources can be found at (<http://grasp.mbb.sfu.ca/>). Recently, we developed a Zebrafish GBrowse in which researchers could visualize catfish (zebrafish or catfish?) BAC and EST sequences (<http://www.animalgenome.org/cgi-bin/gbrowse/zebrafish/>). The second-generation rainbow trout genetic map can be found at <http://www.animalgenome.org/cgi-bin/host/rainbow/viewmap>. With data from Craig Sullivan of North Carolina State University and Robert Chapman of South Carolina Department of Natural Resources, a database has been set up for Striped Bass Ovarian Transcriptome Sequence Contigs. The contig data is annotated by BLAST to GO sequence database (<http://www.animalgenome.org/aquaculture/s.bass/contigs.php>).

### Multi-species

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

### Ontology development

This past year, we focused on the integration of the Animal Trait Ontology (<http://www.animalgenome.org/bioinfo/projects/ATO/>) into the Vertebrate Trait Ontology (<http://www.animalgenome.org/cgi-bin/amion/browse.cgi>). 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, rat, mouse and human research communities. Anyone interested in helping to improve the ATO/VT is encouraged to contact James Reecy ([jreecy@iastate.edu](mailto:jreecy@iastate.edu)), Cari Park ([caripark@iastate.edu](mailto:caripark@iastate.edu)) or Zhiliang Hu ([zhu@iastate.edu](mailto: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. Finally, we are working to develop a livestock breed ontology based on the Oklahoma State University Livestock Breeds web resource.

### Software development

Several on-line tools have been developed (<http://www.animalgenome.org/bioinfo/tools/>). We developed 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 proof of principle, we were able to assemble bovine genes with as little as 3X genome coverage (Koltjes et al., 2009).

We have developed a whole-genome association visualization tool (SNPlotZ; <http://www.animalgenome.org/bioinfo/tools/snplotz/>) for cattle and swine. Individual SNP can be visualized in the context of the genome in GBrowse. This tool can be expanded for use with any species and any

SNPchip that is available.

As a result of collaborations between Iowa State University, the Medical College of Wisconsin, and University of Iowa, we are happy to release a preliminary trial version of the Virtual Comparative Map (VCMMap) tool (<http://bioneos.com/VCMMap/>). Please feel free to try things out, and send any feedback to: [vcmap@bioneos.com](mailto:vcmap@bioneos.com).

The CateGORizer tool is improved to allow users to add their own annotations to a dataset. This is useful if users have multiple datasets and wish to compare their differences or similarities.  
<http://www.animalgenome.org/bioinfo/tools/catego/>

#### Minimal standards development

We have worked with the MIBBI project (<http://www.mibbi.org/>) 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, chicken, and porcine QTL data can be downloaded in gff format and visualized in GBrowse. A couple of thousand (about 2000?) new livestock QTL have been added to the database. Currently, there are 5621 curated porcine QTL, 2359 curated bovine QTL and 1863 curated poultry QTL (<http://www.animalgenome.org/cgi-bin/QTLdb/index>). All livestock QTL data have been ported to NCBI.

#### Facilitating research

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

**Meetings:** Over 2000 scientists attended the joint Plant and Animal Genome meeting held last January, jointly with the annual NAGRP meeting. Coordination funds helped support attendance at PAG-XVI and will do so again for the upcoming PAG meeting in January 2010.

## **PLANS FOR THE FUTURE**

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

Enhance ANEXdb.org capabilities for storage and analysis of gene expression data for all livestock species.

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

We will seek to partner with any NRSP-8 members wishing to warehouse phenotypic and genotypic data in customized relational databases. This will help consortia/researchers whose individual research labs lack expertise with relational databases to warehouse and share information.

**OBJECTIVE 3:** Develop, integrate and implement bioinformatic 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 database curators to develop minimal

information for publication standards. We will also work with these same database groups to improve a phenotype (trait?) ontology, which will facilitate transfer of QTL information across species. In addition, we will expand the QTL database to house microarray data (<http://www.anexdb.org/>) (should this link be to QTLdb?), which will facilitate the identification of candidate genes for researchers seeking causal mutations. We will work with colleagues at USDA-ARS, as well as throughout Europe, 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.

BirdBase will be trialed for 2 months within the NRSP-8 community in general and especially the NRSP-8 avian community. An associated user-community survey will be analyzed by Shane Burgess and the results presented to the NRSP-8 Bioinformatics Committee co-coordinators. A decision about future support/investment in BirdBase via NRSP-8 will be made at that time.

### Publications

1. Couture, O., K. Callenberg, N. Koul, S. Pandit, R. Younes, Z.-L. Hu, J. Dekkers, J.M. Reecy, V. Honavar, C. Tuggle (2009). **ANEXdb: an integrated animal ANnotation and microarray EXpression database**. *Mamm Genome*, DOI 10.1007/s00335-009-9234-1
2. Ramos, A.M., R.P. Crooijmans, N.A. Affara, A.J. Amaral, A.L. Archibald, J.E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M.S. Hansen, J. Hedegaard, Z-L. Hu, H.H. Kerstens, A.S. Law, H-J. Megens, D. Milan, D.J. Nonneman, G.A. Rohrer, M.F. Rothschild, T.L. Smith, R.D. Schnabel, C.P. Van Tassell, J.F. Taylor, R.T. Wiedmann, L.B. Schook, M.A. M. Groenen. (2009) **Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology**. *PLoS ONE* 4(8): e6524. doi:10.1371/journal.pone.0006524
3. The Bovine Genome Sequencing and Analysis Consortium, C.G. Elsik, R.L. Tellam, K.C. Worley, et al (2009). **The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution**. *Science* 24 April 2009: Vol.324. no.5926, pp.522-528
4. Steibel J.P., Wysocki M., Lunney J.K., Ramos A.M., Hu Z.L., Rothschild M.F., Ernst C.W. (2009). **Assessment of the swine protein-annotated oligonucleotide microarray**. *Anim Genet*. 2009 Jun 8. [Epub ahead of print]
5. Koltjes, J., Z.-L. Hu, E. Fritz and J.M. Reecy (2009). **BEAP: The BLAST Extension and Alignment Program - a tool for contig construction and analysis of preliminary genome sequence**. *BMC Research Notes* 2009, 2:11

(Prepared 12/31/09)