U.S. Database Coordination Activities
Supported by Allotments of Regional Research Funds, Hatch Act
For the Period 1/1/10-12/31/10

Overview: Coordination of the CSREES National Animal Genome Research Program's (NAGRP) Bioinformatics is primarily based at, and led from, Iowa State University (ISU), with additional activities at Mississippi State University (MSU) 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 (ISU), Max Rothschild (ISU), Chris Tuggle (ISU), and Shane Burgess (MSU) 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.

See activities listed below.

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, Iowa State University and U.S. Department of Agriculture to further 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.

The following describes the project's activities over this past year.

Poultry
A total of 588 new QTL have been curated into the Chicken QTLdb (http://www.animalgenome.org/QTLdb/chicken.html). Chicken QTL can be visualized against the genome at http://www.animalgenome.org/cgi-bin/gbrowse/chicken/. In addition, we continue to mirror Carl Schmidt's Gallus genome browser (http://www.animalgenome.org/cgi-bin/gbrowse/gallus/).

In response to a request from the NRSP8 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/).

Cattle
In the past year, 2423 new cattle QTL have been added . In addition, cattle QTL can now be viewed relative to the UMD assembly; http://www.animalgenome.org/cgi-bin/gbrowse/bovine/) and Btau4.2 assembly; http://www.animalgenome.org/cgi-bin/gbrowse/cattle/.

A web site for the Ruminant Genome Biology Consortium activities is now online. It is hosted at the NAGRP Bioinformatics site and can be seen at either http://www.animalgenome.org/ruminants/ or http://www.ruminants.org/

Porcine
The pig genome sequencing is actively carried out at Sanger Institute (http://www.sanger.ac.uk/Projects/S_scrofa/) and the latest sequence assembly and genome annotation results can be found at the http://www.animalgenome.org/cgi-bin/gbrowse/ssc/. More 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/QTLdb/pig.html; 723 new QTL have been added this past year).

By working with the pig genome annotation consortium groups, we have developed a pig gene WishList to facilitate the annotation activities (http://www.animalgenome.org/cgi-bin/host/ssc/gene2bacs). This website lists the BAC to which specific gene sequences have been mapped using BLAST. Several groups, primarily the Immune Response Annotation Group (IRAG) headed by Chris Tuggle and Claire Rogel-Gaillard (INRA-Jouy en-Josas), are using this website information to manually annotate gene-encoding regions on the porcine genome sequence. The IRAG group has annotated more than 1,000 regions out of a planned 1,400 total regions thought to contain immune response genes. The NAGRP blast server was updated in time for users to blast their genes of interest against the newly assembled pig genome. It has been heavily used and quite useful to the community.

Sheep
The Sheep QTLdb has been migrated from its Australia site to the site at Iowa State University (http://www.animalgenome.org/QTLdb/sheep.html; 264 new sheep QTL have been added to the Sheep QTLdb).

Aquaculture
Many useful links for aquaculture can be found at http://www.animalgenome.org/aquaculture/. In collaboration with John Liu, Auburn Univeristy, we have set up a Catfish SNP Project web site (http://www.animalgenome.org/catfish/cbarbel/), a Teleost Alternative Splicing Database (http://www.animalgenome.org/tasd), and a Catfish COI DNA Barcode Database (http://www.animalgenome.org/fishid/). We have helped Moh Salem of West Virginia University to set up web blast and data download of the rainbow trout transcriptome data characterized using Sanger and Next GENeration sequencing data (http://www.animalgenome.org/aquaculture/salmonids/rainbowtrout/EST_WV.html).

Multi-species
Gene Ontology annotation for cattle, pigs, chicken, sheep, cat, and several aquaculture species, such as catfish, trout, and salmon as well as experimental genome annotations and tools for analyzing functional genomics data. For a detailed update please see: McCarthy FM, Gresham CR, Buza TJ,Chouvarine P, Pillai LR, Kumar R, Ozkan S, Wang H, Manda P, Arick T, Bridges SM, Burgess SC. AgBase: supporting

functional modeling in agricultural organisms. Nucleic Acids Res. 2011 Jan;39(Database issue):D497-506. Epub 2010 Nov 12. PMID: 21075795

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) for storage and analysis of functional genomics data in livestock (Couture et al. 2009). During 2010, ANEXdb was expanded beyond Affymetrix array data to include two-color array data as well. In addition, the database was migrated to NRSP-8 computer resources.

Ontology development
This past year we continued to focus 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).   Anyone interested in helping to improve the ATO/VT is encouraged to contact James Reecy (jreecy@iastate.edu), Cari Park (caripark@iastate.edu) or Zhiliang Hu (zhu@iastate.edu). In addition, we have begun working with the Rat Genome Database to integrate ATO terms that are not applicable to the Vertebrate Trait Ontology into the Clinical Measures Ontology (under development). We are collaborating with researchers at INRA (France) and within EADGENE and SABRE, EU funded projects, to expand the utility of the ATO, including the development of an ontology devoted to traits of interest in livestock species. 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/).   Recent additions to the tool-box include GFF3 File Formattor, a web tool to convert mapping data to GFF3 format to facilitate analysis of user data by alignment against respective genomes. For example, part of the tool works with the whole genome association visualization tool (SNPlotZ; http://www.animalgenome.org/bioinfo/tools/snplotz/) for cattle and swine, allowing users to convert their subset of SNPs into GBrowse-ready GFF3 files for use 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 an updated version of the Virtual Comparative Map (VCMap) tool (http://bioneos.com/VCMap/). Please feel free to try things out, and send any feedback to: vcmap@bioneos.com.

We have also worked with Jill Madox from Australia to set up a web site for user interactions in improving the CRIMAP software (http://www.animalgenome.org/tools/share/crimap/).

Minimal standards development
We have worked with the MIBBI project 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, chicken, and porcine QTL data can be downloaded in BED and GFF formats and visualized in GBrowse.   Cytogenetic G-banded chromosome ideograms are added to QTL maps of all species within QTLdb. Two QTL meta-plot tools are now available to help users with their own QTL meta-analysis (see http://www.animalgenome.org/QTLdb/doc/meta/). A total of 3998 new livestock QTL have been added to the database. Currently, there are 6344 curated porcine QTL, 4682 curated bovine QTL, 2451 curated

poultry QTL and 348 curated sheep QTL in the database (http://www.animalgenome.org/QTLdb/).   All included livestock QTL data have been ported to NCBI. With the exception of the chicken QTL, all QTL data can be visualized relative to the Illumina SNPchip data. Finally, we have started to curate all SNP association studies for all livestock species into the database.

Facilitating research
We have set up a Data Repository for community members to share their genome analysis data. Currently actively used data sharing is mostly for cattle and pigs (http://www.animalgenome.org/repository). Our helpdesk is here to assist community members. We have also set up a web site for the Ruminant Genome Biology Consortium (http://www.animalgenome.org/ruminants/) to facilitate their communications. Throughout the year, we have helped >98 research groups/individuals with their research projects and questions. 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.

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 phenotype and measurement ontologies, which will facilitate transfer of QTL information across species. In addition, we will expand the QTL Database to house whole-genome association data, which will facilitate the identification of candidate genes for researchers seeking causal mutations. We will continue 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 was released for trial by all NRSP8 members in Jan 2010. Since this time, minor changes and additions have been made to Birdbase. Birdbase serves the avian clade rather than as a single model organism database. It can serve as an example for other NRSP-8species, especially the Aquaculture community because of the similarities in systems and species diversity. An update on the BirdBase site and results of the user survey will be presented at the 2011 PAG poultry meeting (NC1170) by Parker Antin, Shane Burgess, Fiona McCarthy and Carl Schmidt. An associated user-community survey will be analyzed by Burgess and the results presented to the NRSP8 Bioinformatics committee co-coordinators. A decision will be made at that time about whether future support/investment in BirdBase via NRSP-8 will be

requested.

We are working to integrate microarray element sequences with genome data for improved orthology/comparative systems biology analyses as well as the integration of expression and genome location. This work is being initiated in the pig to facilitate genome annotation but could be extended to other species using available expression profiling tools and genome sequences.

Publications

1. Jianguo Lu, Eric Peatman, Qing Yang, Shaolin Wang, Zhi-Liang Hu, James Reecy, Huseyin Kucuktas and Zhanjiang Liu (2011). The catfish genome database cBARBEL: an informatic platform for genome biology of ictalurid catfish. Nucleic Acids Res. 2010 Oct 8. [Epub ahead of print]
2. Zhi-Liang Hu, Carissa A. Park, Eric R. Fritz and James M. Reecy (2010). QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes. Invited Lecture with full paper published on The 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany August 1-6, 2010.
3. Yu Gao, Laurence Flori, Jerome Lecardonnel, Diane Esquerre, Zhi-Liang Hu, Angelique Teillaud, Gaetan Lemonnier, Francois Lefevre, Isabelle P Oswald and Claire Rogel-Gaillard (2010). Transcriptome analysis of porcine PBMCs after in vitro stimulation by LPS or PMA/ionomycin using an expression array targeting the pig immune response. BMC Genomics 2010, 11:292.

(Prepared 1/11/11)