

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

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 Bioinformatic 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 Fiona McCarthy (MSU) as Co-Coordinators. Iowa State University and Mississippi State University provide 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, partnered with researchers at Kansas State University, Michigan State University, Iowa State University and U.S. Department of Agriculture, we have further developed and improved the web-interfaced relational databases to store and disseminate phenotypic and genotypic information from large genomic studies in farm animals and better serve the needs of researchers. 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 991 new QTL have been curated into the Animal 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/> and aligned with chicken 60K SNPs along with NCBI-annotated gene information (<http://www.animalgenome.org/cgi-bin/gbrowse/chicken/>) on genome build GG_4.0. In addition, we continue to mirror Dr. Carl Schmidt's Gallus genome browser while the

original site is undergoing restructuring (<http://www.animalgenome.org/cgi-bin/gbrowse/gallus/>).

NRSP-8 funds were used to support the development of BirdBase resources. Specifically, Chicken Gene Nomenclature Committee (CGNC) database was developed, updating records from the Roslin GeneNames effort and providing access to chicken gene nomenclature. It is now possible for biocurators and community experts to add nomenclature, download current nomenclature. In addition this resource is now linked the NCBI Entrez Gene chicken gene pages (with thanks to Janet Weber). During 2012, we expect to link in with Ensembl Genes and with the HGNC.

Via BirdBase, a bird comparative genome browser is being developed. This will initially include the chicken, turkey and zebra finch genomes and will then be expanded to include all bird genomes (including bird genomes sequences as part of the BGI-10K Genomes project). Currently we have accessed funds for this comparative browser (and its set up) and are in the process of buying the server. We expect the bird comparative genome browser to be online during 2012.

Cattle

In the past year, 1118 new cattle QTL have been added to the Animal QTLdb (<http://www.animalgenome.org/QTLdb/cattle>). In addition, cattle QTL can now be viewed relative to both the UMD3.1 assembly (<http://www.animalgenome.org/cgi-bin/gbrowse/bovine/>) and Btau4.2 assembly (<http://www.animalgenome.org/cgi-bin/gbrowse/cattle>). Cattle 770K high-density SNPs and 4.1M dbSNP data are now available in GBrowse to align with QTL and in SNPlotz for genome analysis (<http://www.animalgenome.org/tools/snplotz/>).

Porcine

The pig genome sequencing information has been updated at <http://www.animalgenome.org/pigs/genome/> and a new pig genome database has been under active development (<http://www.animalgenome.org/pig/genome/db/>). In the past year a total of 1104 new QTL have been added to the AnimalQTLdb (<http://www.animalgenome.org/QTLdb/pig>). The pig gene Wishlist (<http://www.animalgenome.org/cgi-bin/host/ssc/gene2bacs>) has continued to serve the pig genome annotation activities throughout 2012.

Sheep

In 2012, 114 new sheep QTL have been added to the Animal QTLdb (<http://www.animalgenome.org/QTLdb/sheep>). Active updates have been continued for the NRSP-8 web site for activities in the sheep genome community (<http://www.animalgenome.org/sheep/>).

Upon request from Jill Maddox (jillian.maddox@alumni.unimelb.edu.au), a new mailing list “Sheep Models” (www.animalgenome.org/sheep/community/SheepModels) has been set up and is being actively used.

GBrowse alignments for sheep 54K SNP and BAC clones were set up on OAR Build 3.1.

Aquaculture

Many useful links for aquaculture can be found at <http://www.animalgenome.org/aquaculture/>. Thanks to collaborative efforts by researchers from the USDA National Center for Cool and Cold Water Aquaculture, new QTL continue to be entered into the QTLdb. In 2012, 115 new QTL data for rainbow trout have been curated into the Animal QTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/OM/index>).

Multi-species

A local copy of Biomart software has been kept up-to-date on the AnimalGenome.ORG server to serve the cattle, chicken, pig, and horse community (<http://www.animalgenome.org:8181/>). New data sources and species continue to be updated.

Ontology development

This past year we continued to focus on the integration of the Animal Trait Ontology into the Vertebrate Trait Ontology (<http://bioportal.bioontology.org/ontologies/1659>). We have continued working with the Rat Genome Database to integrate ATO terms that are not applicable to the Vertebrate Trait Ontology into the Clinical Measurement Ontology (<http://bioportal.bioontology.org/ontologies/1583>). Traits specific to livestock products continue to be incorporated into a Livestock Product Trait Ontology (PT; <http://animalgenome.org/cgi-bin/amido/browse.cgi>). We have also continued mapping the cattle, pig, chicken, and sheep QTL traits to Vertebrate Trait Ontology (VT), Product Trait Ontology (PT) and Clinical Measurement Ontology (CMO) to help standardize the trait nomenclature used in the QTLdb. Anyone interested in helping to improve the ATO/VT is encouraged to contact James Reecy (jreecy@iastate.edu), Cari Park (caripark@iastate.edu) or **Error! Reference source not found.** (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, including the development of an ontology devoted to traits of interest in livestock species (<http://www.atol-ontology.com/index.php/en/1-ontology/visualisation>). The new VT/PT/CMO cross-mapping has been well employed by the Animal QTLdb and VCMMap tools. Finally, we have made plans to expand the livestock breed ontology with updated data from Oklahoma State University, FAO, and from China.

Via AgBase, we were recruited by the Phenotype Ontology Research Coordination Network (NSF DBI 0956049) to develop an avian anatomy ontology with the goal of integrating this with other, existing ontologies to describe phenotypes. Using MSU and AgBase funds we have about 650 terms that cover avian musculoskeletal and integument systems. The information for these terms includes relationships, synonyms, definitions, and comments (homologies to mammalian structures; species differences). The development of this ontology is done in collaboration with Duncan Davidson (UK Medical Research Council) and in conjunction with the larger Phenotype RCN Vertebrate Working group. This work will continue with BirdBase collaborators during 2012 (as funding permits) and will enable poultry researchers to describe phenotype in a standard way that promotes data sharing.

Software development

The NRSP-8 Bioinformatics Online Tool Box has been actively updated (<http://www.animalgenome.org/bioinfo/tools/>). Software upgrades were made continually to SNPlotz, Gene Ontology CateGORizer, BEAP, and the Expeditor. .

As a result of collaborations between Iowa State University, the Medical College of Wisconsin, and University of Iowa, the Virtual Comparative Map (VCMMap) tool has passed its initial development stage and is now transferred to AnimalGenome.ORG (<http://www.animalgenome.org/VCmap/>). More application development, improvement, and testing has continued. Online help materials have been added, including a written user manual and a video tutorial. To improve links between AgBase and the NRSP-8 website, AgBase now also provides a link to the Virtual Comparative Map (VCMMap). Please feel free to try things out and send any feedback to vcmap@animalgenome.org.

The web site and user forum listserv for CRI-MAP user interactions for improvement of the CRI-MAP software (<http://www.animalgenome.org/tools/share/crimap/>) has been actively used.

Minimal standards development

We have continued to work on 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

In 2012, a total of 2552 new QTL have been added to the database. Currently, there are 7451 curated porcine QTL, 5920 curated bovine QTL, 3442 curated poultry QTL, 753 curated sheep QTL, and 88 curated rainbow trout QTL in the database (<http://www.animalgenome.org/QTLdb/>). All included livestock QTL data have been ported to NCBI. Since we started to curate SNP-association data for all livestock species, there have been 5037 association data added to the database.

As a result of our continued efforts developing the Animal QTLdb into something more useful for the community, a publication to summarize all new functionalities is to appear in the 2013 Nucleic Acids Research Database Issue.

.

Facilitating research

The Data Repository for the aquaculture, cattle, chicken, and pig communities to share their genome analysis data has been proven to be very useful (<http://www.animalgenome.org/repository>). More species data is currently being added. Our helpdesk is here to assist community members. The online data file-sharing tool has been actively used. Newly added functions include authenticated access for small consortium groups and/or projects. Throughout the year, we have helped to reformat large datasets to meet the needs of wet lab researchers. We have helped more than 70 research groups/individuals with their research projects and questions. Our involvement has ranged from data transfer, data assembly, and data analysis, to software applications, code development, etc. Please continue to contact us as you need help with bioinformatic issues.

The ANGENMAP listserv has been heavily used in the past year. Now the annual posts sent through the list have grown from about 300 to over 400 per year. We have approximately 2300 subscribers, which is 170 more than last year (on average +130 per year for the past 10 years).

PLANS FOR THE FUTURE.

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. We will continue working with U.S. and European colleagues 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.

Zhi-Liang Hu, Carissa A. Park, Xiao-Lin Wu and James M. Reecy (2012). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research*, 41 (D1): D871-D879.

Publications

1. Zhi-Liang Hu, Carissa A. Park, Xiao-Lin Wu and James M. Reecy (2012). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research*, 41 (D1): D871-D879.
2. Martien Groenen, Alan Archibald, Hirohide Uenishi, Christopher Tuggle, Yasuhiro Takeuchi, Max Rothschild, Claire Rogel-Gaillard, Chankyu Park, Denis Milan, Hendrik-Jan Megens, Shengting Li, Denis Larkin, et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491, 393-398 (15 November 2012).
3. Endale Ahanda ML, Fritz ER, Estellé J, Hu Zhi-Liang, Madsen O, Groenen MA, Beraldi D, Kapetanovic R, Hume DA, Rowland RR, Lunney JK, Rogel-Gaillard C, Reecy JM, Giuffra E. (2012). Prediction of Altered 3'-UTR miRNA-Binding Sites from RNA-Seq Data: The Swine Leukocyte Antigen Complex (SLA) as a Model Region. *PLoS One*. 2012;7(11).
4. Marie-Laure Endale Ahanda, Eric R Fritz, Jordi Estellé, Zhi-Liang Hu, Ole Madsen, Martien AM Groenen, Dario Beraldi, Ronan Kapetanovic, David A Hume, Robert RR Rowland, Joan K Lunney, Claire Rogel-Gaillard, James M Reecy, Elisabetta Giuffra (2012). Prediction of Altered 3'-UTR miRNA-Binding Sites from RNA-Seq Data: The Swine Leukocyte Antigen Complex (SLA) as a Model Region. *PLoS One*. 2012; 7(11): doi: 10.1371/journal.pone.0048607. Epub 2012 Nov 6
5. Xiao-Lin Wu and Zhi-Liang Hu (2012). *Methods in molecular biology* (Clifton, NJ), 871: 145.
6. Zhi-Lian Hu, JM Reecy, and XL Wu (2012). Design database for quantitative trait loci (QTL) data warehouse, data mining, and meta-analysis. *Methods in molecular biology* (Clifton, NJ) 871, 121

(Prepared 1/11/13)