

U.S. Bioinformatic Coordination Activities
Supported by Allotments of Regional Research Funds, Hatch Act
For the Period 1/1/11-12/31/11

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 by the 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 285 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/> and aligned with chicken 60K SNPs along with NCBI annotated gene information (<http://www.animalgenome.org/cgi-bin/gbrowse/chicken/>). In addition, we continue to mirror Dr. Carl Schmidt's Gallus genome browser (<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, 525 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>. Thanks to efforts by Steven Salzberg's group at University of Maryland for mapping the STS Cattle 770K high density SNPs and 4.1M dbSNP data are made available both in GBrowse to align with QTL and in SNPlotz for genome analysis (<http://www.animalgenome.org/tools/snplotz/>).

Porcine

The updated pig genome sequencing information are 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/>). 88 new QTL have been added to the AnimalQTLdb this past year (<http://www.animalgenome.org/QTLdb/pig>). The pig gene Wishlist (<http://www.animalgenome.org/cgi-bin/host/ssc/gene2bacs>) continue to serve the pig genome annotation activities throughout the 2011.

Sheep

In 2011, 291 new sheep QTL have been added to the Sheep QTLdb. A web site for NRSP-8 activities continue to be updated as a sheep genome community information hub (<http://www.animalgenome.org/sheep/>).

Aquaculture

Many useful links for aquaculture can be found at <http://www.animalgenome.org/aquaculture/> . Thanks to collaborative efforts by Yniv Palti et all from the USDA National Center for Cool and Cold Water Aquaculture, rainbow trout is now a new species in the Animal QTLdb species families. To date 27 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 installed on the AnimalGenome.ORG server to serve the cattle, chicken, pigs and horse community (<http://www.animalgenome.org:8181/>). New data sources and species are being added.

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 have been incorporated into a new Livestock Product Trait Ontology (PT; <http://animalgenome.org/cgi-bin/amido/browse.cgi>). As the first stage outcome, we have mapped 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. Finally, we have developed a livestock breed ontology based on the Oklahoma State University Livestock Breeds web resource. This breed ontology is also found its application in the Animal QTLdb.

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 kept updated (<http://www.animalgenome.org/bioinfo/tools/>). Several major software upgrades were made to SNPlotz, Gene Ontology CateGORizer, BEAP, and the Expeditor. The new addition to the tool box is an online File Sharing Platform, with which any NRSP-8 members can freely use the tool to share large data files individually or publicly (e.g. within a consortium).

As a result of collaborations between Iowa State University, the Medical College of Wisconsin, and University of Iowa, the Virtual Comparative Map (VCMaP) tool has passed its initial development stage and now transferred to the AnimalGenome.ORG (<http://www.animalgenome.org/VCMaP/>). More application development will be expected. To improve links between AgBase and the NRSP-8 website, AgBase now also provides a link to the Virtual Comparative Map (VCMaP). Please feel free to try things out, and send any feedback to: vcmap@animalgenome.org.

Genome2Seq, an online tool that rapidly retrieves a fasta file of sequences based on genome co-ordinates generated from RNA-Seq data, was developed. Users specify either bovine, horse, chicken or pig genomes. The input file is a tab-delimited text file containing a unique ID, chromosome number, start location, and stop location in that order. The output is (a) a list of matching genes and their associated GO annotation; and (b) a fasta file of sequences for any co-ordinates that do not match any annotated genes. Genome2Seq is available via AgBase and the NRSP-8 website.

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

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

In 2011, a total of 1216 new QTL have been added to the database. Currently, there are 6432 curated porcine QTL, 5207 curated bovine QTL, 2736 curated poultry QTL, 639 curated sheep QTL, and 27 curated rainbow trout 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.

In addition, The QTLdb has been revamped to accommodate genome map (map unit: bp) based QTL and association data (e.g. GWAS data). The map view and QTL details view have updated to integrate these new data types. A new user option allows one to view "QTL", "eQTL", "Associations", or combinations of these data. Some minor but useful improvements continue to be made. For example users can see a list of references for the QTL within any chromosomal view window.

Facilitating research

The Data Repository for aquaculture, cattle, chicken and pig community to share their genome analysis data has been proven to be very useful. (<http://www.animalgenome.org/repository>). More species data is being added. Our helpdesk is here to assist community members. . Throughout the year, we have helped >88 research groups/individuals with their research projects and questions. Our involvement has ranged from data transfer, data assembly, data analysis, software applications, code development, etc. Please continue to contact us as you need help with bioinformatic issues.

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. colleagues, 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.

Publications

1. Zhi-Liang Hu, James E Koltes, Carissa A. Park, Eric R. Fritz and James M. Reecy (2011). Review: Bioinformatics approaches to livestock animal genomics research. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 2011 6, No. 044.
2. Zhi-Liang Hu, Antonio M. Ramos, Sean J. Humphray, Jane Rogers, James M. Reecy and Max F. Rothschild. (2011). Use of Genome Sequence Information for Meat Quality Trait QTL Mining for Causal Genes and Mutations on Pig Chromosome 17. *Frontiers in Livestock Genomics*, 2:43. doi: 10.3389/fgene.2011.00043.
3. Xiao-Lin Wu, Daniel Gianola, Zhi-Liang Hu and James M. Reecy (2011). Meta-Analysis of Quantitative Trait Association and Mapping Studies using Parametric and Non-Parametric Models. *Biometrics & Biostatistics, J Biomet Biostat* 2011, S1. doi: 10.4172/2155-6180.S1-001.
4. Zhi-Liang Hu, Xiao-Lin Wu and James M. Reecy (2011). Extension of Animal QTLdb: QTL Meta-analysis on the Fly. Poster paper published on ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB), August 1-3, 2011, Chicago, IL, USA.
5. 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]
6. McCarthy, F.M., Gresham, C.R., Buza, T.J., Chouvarine, P., Pillai, L.R., Kumar, R., Ozkan, S., Wang, H., Manda, P., Arick, T., Bridges, S.M., Burgess, S.C. (2011) AgBase: supporting functional modeling in agricultural organisms. *Nucleic Acids Res*, 39, D497-506.
7. Nail, C., Chouvarine, P., Weber, J., Spencer., Kumari, S., Burgess, S.C., Schmidt, C, Antin, P., McCarthy, F.M. Manual biocuration to support standardized chicken gene nomenclature at CGNC. The 6th International Chick Meeting Roslin Institute, Sept. 17-20, 2011.

(Prepared 1/11/11)