

U.S. Bioinformatic Coordination Activities

Supported by Allotments of Regional Research Funds, Hatch Act
For the Period 1/1/13-12/31/13

Overview: Coordination of the NIFA National Animal Genome Research Program's (NAGRP) Bioinformatics is primarily based at, and led from, Iowa State University (ISU), with additional activities at the University of Arizona (UA) 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 (UA) as Co-Coordinators. Iowa State University and University of Arizona 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, thereby facilitating 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 the past year.

Poultry

A total of 477 new QTL were 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/>).

The Chicken Gene Nomenclature Committee (CGNC) database was initiated with NRSP-8 funds to provide standardized gene nomenclature for chicken genes. As of 30 December 2013, we have assigned nomenclature for 14,800 genes using orthology to HGNC assigned gene names and a further 1,684 genes have been manually assigned nomenclature by biocurators supported by Arizona state funds. We also responded to 11 community requests to provide chicken gene nomenclature. The standardized chicken gene names are publically displayed at the NCBI Entrez Gene database and we are working with Ensembl to ensure they are able to also access this data.

Cattle

In the past year, 2000 new cattle QTL were 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/>).

We have also updated the initial cattle gene nomenclature provided by the Bovine Genome Database, providing standardized gene nomenclature for 9,910 *Bos taurus* genes based upon homology to assigned human gene nomenclature. This data is available at http://www.animalgenome.org/genetics_glossaries/bovgene.

Swine

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 1,547 new QTL were 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 support the pig genome annotation activities.

Sheep

In 2013, 36 new sheep QTL were 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/>).

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 2013, 39 new QTL data for rainbow trout were 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 communities (<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. A new web page is set up to reflect this development (<http://www.animalgenome.org/bioinfo/projects/ato/alt>), with new sites at <http://www.animalgenome.org/bioinfo/projects/vt/>, <http://www.animalgenome.org/bioinfo/projects/pt/>, and <http://www.animalgenome.org/bioinfo/projects/cmo/> respectively. 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). 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, Food and Agriculture Organization, and from China.

The chicken adult anatomy is complete, and consists of 2,284 ontology terms cross referenced with the Vertebrate and Uberon Ontologies. The information for these terms includes relationships, synonyms, definitions, and comments (homologies to mammalian structures, species differences). In January 2013, Drs Frances Wong (Roslin Institute) and Fiona McCarthy (University of Arizona) collaborated to begin integration of adult and embryological anatomy terms for the chicken ontology. Dr Wong's visit to the UA was partially supported by a Collaborative Exchange award from the Phenotype Ontology Research Coordination Network (NSF-DEB-0956049). Continuation of this work awaits further funding opportunity.

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; <http://www.animalgenome.org/VCMap/>) tool has passed its initial development stage and is at a stable working status serving the community. 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.

Mailing lists and user forums

We have been hosting a couple dozen mailing lists / web sites for various research groups in the NAGRP community. The most active groups include the AnGenMap (www.animalgenome.org/community/angenmap/), The "CRI-MAP users" (<http://www.animalgenome.org/tools/share/crimap/> for user interactions to improve CRI-MAP software), "Sheep Models" (www.animalgenome.org/sheep/community/SheepModels), etc.

Upon request from Hasan Khatib (hkhatib@wisc.edu), a new mailing list "EPIgroup" (www.animalgenome.org/community/epigroup/) was set up to promote epigenetics research in livestock species. It currently has 198 members.

Upon request from Frank Nicholas (frank.nicholas@sydney.edu.au), a new mailing list “OMIA-Support Group” (www.animalgenome.org/community/omia-support/) was set up to facilitate OMIA development activities. It currently has 80 members.

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 2013, a total of 4099 new QTL have been added to the database. Currently, there are 9862 curated porcine QTL, 6305 curated bovine QTL, 3919 curated chicken QTL, 789 curated sheep QTL, and 127 curated rainbow trout QTL in the database (<http://www.animalgenome.org/QTLdb/>). We are adding Horse QTLdb to the Animal QTLdb family to collect horse QTL/association data. All included livestock QTL data have been ported to NCBI. In 2013, we have worked with UCSC and Ensembl to port the livestock animal QTL data to UCSC Golden tracks and to Ensembl databases. Now users can fully utilize the tools at NCBI, Ensembl, and UCSC to mine animal QTL data..

The January 2013 Nucleic Acids Research Database Issue contains a paper describing the latest developments we made on the Animal QTLdb.

Facilitating research

The Data Repository for the aquaculture, cattle, chicken, and pig communities to share their genome analysis data has proven to be very useful (<http://www.animalgenome.org/repository>). New data is continually being added. Frequent data downloads include over 140 data files in 6 different animal species. The newly added data includes rainbow trout genome assembly draft, chicken 60K SNP information, etc. In parallel to the public data repository, the online data file-sharing tool has also been actively used to facilitate data sharing among collaborators and/or groups.

Our helpdesk is here to assist community members. Throughout the year, we have helped more than 50 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.

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-NIFA, to help direct future livestock-oriented bioinformatic/database efforts.

Publications

1. Zhi-Liang Hu, Carissa A. Park, Xiao-Lin Wu and James M. Reecy (2013). **Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era.** *Nucleic Acids Research*, 41(D1):D871-9.
2. Carissa A Park, Susan M Bello, Cynthia L Smith, Zhi-Liang Hu, Diane H Munzenmaier, Rajni Nigam, Jennifer R Smith, Mary Shimoyama, Janan T Eppig and James M Reecy. **The Vertebrate Trait Ontology: a controlled vocabulary for the annotation of trait data across species.** *Journal of Biomedical Semantics* 2013, 4:13.
3. Bouabid Badaoui, Christopher K Tuggle, Zhi-Liang Hu, James M Reecy, Tahar Ait-Ali, Anna Anselmo and Sara Botti (2013). **Pig immune response to general stimulus and to porcine reproductive and respiratory syndrome virus infection: a meta-analysis approach.** *BMC Genomics* 2013, 14:220; doi:10.1186/1471-2164-14-220.



James Reecy
Professor, Department of Animal Science
NRSP-8 Bioinformatics Coordination Project
Iowa State University
2255 Kildee Hall
Ames, Iowa 50011 USA
+1-515-294-9269 (ph)

(Prepared 1/11/14)