

## U.S. Bioinformatic Coordination Activities

Supported by Allotments of Regional Research Funds, Hatch Act

For the Period 1/1/14-12/31/14

January 10, 2015

**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 continued to further develop and improve 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>). 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 this past year.

The NAGRP data repository has been actively used by the horse community to share the Variant Call Format (VCF) files in their collaborative research.

### **Multi-species support**

The Animal QTLdb and the NAGRP data repository have been actively serving multiple species research activities. A state-of-the-art online alignment tool (Jbrowse) has been set up on the AnimalGenome.ORG server to serve the cattle, chicken, pig, sheep, and horse communities for QTL/association data alignment with annotated genes and other genome features (<http://i.animalgenome.org/jbrowse>). The advantage of Jbrowse is that it easily allows user quantitative data— XYPlot/Density, in BAM or VCF format—to be loaded directly to a user's browser for comparisons in the local environment. New data sources and species continue to be updated. This complements GBrowse, which features multiple HD SNP chip, OMIA genes, and STS marker alignments against QTL/association data for cattle, chicken, pig, sheep, and horse. Recently a dedicated virtual machine platform was set up to develop cyber resources for the Striped Bass Genome Database activity, a project led by Benjamin Reading and Charles Opperman at North Carolina State University (<http://stripedbass.animalgenome.org/>).

### **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/VT>). 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/CMO>). 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, sheep, and horse 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 links to the three new sites for VT, PT, and CMO respectively. At the request of community members, at least 45 new terms were added to the VT in 2014. 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)). The new VT/PT/CMO cross-mapping has been well employed by the Animal QTLdb and VCMMap tools. Annotation to the VT is now also available for rat QTL data in the Rat Genome Database and for mouse strain measurements in the Mouse Phenome Database. 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.

Continuing work on the chicken anatomy ontology is based upon UA biocurator funds, with work focusing on (1) linking adult chicken anatomy terms with the Uberon ontology (of generic anatomical terms) and (2) adding developmental terms provided by Prof Burt's group at the Roslin Institute. Currently the chicken anatomy ontology contains 14,627 terms, cross-referenced with the Uberon ontology (and other related anatomy ontologies). Since this ontology will be required for the Functional Annotation of Animal Genomes (FAANG) Project, during 2015 we will seek competitive funding for a full-time biocurator to complete this ontology.

### **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, and the Expeditor. The CateGORizer is now bundled with

a new external tool, ReviGO, for the convenience of users to take CateGORizer outputs directly to ReviGO for a semantic representative subset analysis.

In collaboration with Dr. Shengsong Xie from Shanghai, Yuhua Fu and Dr. Shu-hong Zhao from Wuhan, China, a sRNAPrimer designing tool has been made available through AnimalGemome.ORG (<http://www.animalgenome.org/cgi-bin/host/sRNAPrimer/d>).

As a result of collaborations between Iowa State University, the Medical College of Wisconsin, and University of Iowa, the Virtual Comparative Map (<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 have continued. Online help materials have been added, including a written user manual and a video tutorial. AgBase and the NRSP-8 websites provide multiple reciprocal reference links to facilitate resource sharing. Please feel free to try things out and send any feedback to [vcmap@animalgenome.org](mailto:vcmap@animalgenome.org).

### **Gene nomenclature standard**

During 2014 the Chicken Gene Nomenclature Committee (CGNC) updated nomenclature to support new annotations from both NCBI and Ensembl. We currently provide standardized nomenclature for 16,422 genes and this data is now routinely distributed to both NCBI Entrez and Ensembl. During 2014 funding to support chicken gene nomenclature was provided by NIH NIGMS Project number 5R24GM079326-02 and during 2015 we will be seeking continued competitive funds for this project.

The initial cattle gene nomenclature is provided by the Bovine Genome Database. Currently we have standardized gene nomenclature for 9,910 *Bos taurus* genes based upon homology to assigned human gene nomenclature ([http://www.animalgenome.org/genetics\\_glossaries/bovgene](http://www.animalgenome.org/genetics_glossaries/bovgene)). We are also working with HGNC to support the development and use of standardized gene nomenclature for livestock species.

### **Minimal standards development**

We have continued to work on the MIQAS project to help define minimal standards for publication of QTL and gene association data (<http://miqas.sourceforge.net/>). The most recent works were to develop documentations how this was done in Animal QTLdb.

### **Expanded Animal QTLdb functionality**

In 2014, a total of 9,063 new QTL have been added to the database. Currently, there are 12,618 curated porcine QTL, 13,415 curated bovine QTL, 4,379 curated chicken QTL, 1,005 curated horse QTL, 791 curated sheep QTL, and 127 curated rainbow trout QTL in the database (<http://www.animalgenome.org/QTLdb/>). All included livestock QTL data have been ported to NCBI, Ensembl, and UCSC genome browser. Now users can fully utilize the browser and data mining tools at NCBI, Ensembl, and UCSC to explore animal QTL/association data. In addition we have continued to improve existing and add new QTLdb curation tools and user portal tools. The new additions include a batch data loading tool to speed up the curation process and a new API tool set to facilitate programming access to the database (see our poster #1157 for details).

### **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. A total of 1,126 data files on different animal genomes,

supplementary data files to publications, and other sharing purpose have been made available to community users. More than 50 data files were shared/transmitted through the online data file-sharing tool by collaborators and/or groups in the community. Our helpdesk is here to assist community members. Throughout 2014, we have helped more than 60 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.

### **Community support and user services at AnimalGenome.ORG**

We have been maintaining and actively updating the NRSP-8 species web pages for each of the six species. We have been hosting a couple dozen mailing lists/web sites for various research groups in the NAGRP community (<http://www.animalgenome.org/community/>). This includes groups like AnGenMap, "CRI-MAP users", "Sheep Models", etc. The most recent addition is a new web site for the Functional Annotation of ANimal Genomes (FAANG) project, with list mailing, user forum, wiki pages, and online publishing capabilities to support coordinated international action to accelerate Genome to Phenome. An increasing number of web hits and data downloads continued in 2014. For example, AnimalGenome.ORG received over 3.7 million web hits from 237,000 individual sites (visitors), which made 970,000 data downloads that generated almost 2 TB internet traffic.

### **Reaching out**

We have been sending periodic updates to over 2,500 users worldwide to inform them of the news and updated information we develop or host at AnimalGenome.ORG. More than 38 new items were updated to the community in 2014.

### **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**

Hu, Zhi-Liang, James M. Reecy, Fiona McCarthy, and Carissa A. Park. Standard Genetic Nomenclature. In: *The Genetics of the Cattle*. Edited by Dorian Garrick and Anatoly Ruvinsky. 2014, CABI Publishing, Wallingford UK.