

U.S. Bioinformatic Coordination Activities

Supported by Allotments of Regional Research Funds, Hatch Act

For the Period 1/1/15-12/31/15

January 8, 2016

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 10/01/13, 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.

The partnership with researchers at Kansas State University, Michigan State University, Iowa State University, and the U.S. Department of Agriculture continues as the database and website interface developed for this collaboration (<http://www.animalgenome.org/lunney>) have been improved, and continued data generation by the group has increased the amount of data that is housed in the database. This resource continues to help the consortium by offering a localized source of information and continued facilitation of 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 is still actively used in 2015 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 supporting the research activities for multiple species. A collaborative site at iPlant has been set up to share some of the web traffic, including the JBrowse 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 (users need to learn how to use JBrowse). New data sources and species continue to be updated. We have recently set up a virtual machine site to host the Online Mendelian Inheritance in Animals (OMIA) database created and maintained by Dr. Frank Nicolas at the University of Sydney (<http://omia.animalgenome.org/>). This is part of our collaborative effort to migrate OMIA to NAGRP platforms.

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 (LPT), which has now been added to NCBO’s BioPortal (<http://bioportal.bioontology.org/ontologies/LPT>). 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. At the request of community members, at least 15 new terms were added to the VT in 2015. 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). The VT/PT/CMO cross-mapping has been well employed by the Animal QTLdb and VCMMap tools. Annotation to the VT is also available for rat QTL data in the Rat Genome Database and for mouse strain measurements in the Mouse Phenome Database. We have also been integrating information from multiple resources, e.g. [FAO - International Domestic Livestock Resources Information](#), [Oklahoma State University - Breeds of Livestock](#) web site, and Wikipedia, to continue development of a Livestock Breed Ontology (LBO; <http://www.animalgenome.org/bioinfo/projects/lbo/>) with an AmiGO display of the hierarchy. The LBO data has also been deposited into BioPortal (<http://bioportal.bioontology.org/ontologies/LBO>).

As of October 15, 2015, AgBase provides 1,539,447 GO annotations for 310,971 gene products from 504 species, including more than 40 agriculturally important species and their pathogens. This information includes 392,101 GO annotations for 57,589 avian gene products, 96% of which are associated with chicken and turkey.

Software development

The NRSP-8 Bioinformatics Online Tool Box has been actively maintained for use by the community (<http://www.animalgenome.org/bioinfo/tools/>). Software upgrades and bug fixes were made continually to SNPlotz, Gene Ontology CateGORizer, and the Expeditor. Bundled with ReviGO, the CateGORizer is serving users for both GO term categorization and semantic representation.

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 AnimalGenome.org 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.

Gene nomenclature standard

During 2015, the Chicken Gene Nomenclature Committee (CGNC) biocurators worked closely with NCBI Entrez curators to ensure that updated gene annotations had revised nomenclature. This year, we reviewed and updated 2,735 genes (including >100 genes we annotated in conjunction with NCBI Entrez curators). We currently provide standardized nomenclature for 22,172 chicken genes and have a pending grant application to support continued annotation of bird genes.

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 (Fiona McCarthy; 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 work has been to develop documentation indicating how this was done in Animal QTLdb.

Expanded Animal QTLdb functionality

In 2015, a total of 31,976 new QTL have been added to the database. Currently, there are 14,479 curated porcine QTL, 42,019 curated bovine QTL, 5,196 curated chicken QTL, 1,125 curated horse QTL, 1,090 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. 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 genome-wide plot of QTL/association data queried in several ways (see our poster #21315 for details).

Further development of Animal Trait Correlation Database (CorrDB)

We have started a developmental process to add public curator/editor tools to the CorrDB to allow continued curation of trait correlation data into the database. Currently the efforts are geared towards making use of resources and tools in the QTLdb for trait ontology development and management, literature management, and bug reporting tools for data quality control. The tools are expected to be released in early 2016.

Facilitating research

The Data Repository for the aquaculture, cattle, chicken, horse, pig, and sheep 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 1033 data files on different animal genomes, supplementary data files to publications, and data for other sharing purposes 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 2015, 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.

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.

We have actively maintained and developed the web site for the Functional Annotation of ANimal Genomes (FAANG) project, with new mailing list, user forum, wiki pages, interactive meeting sites, platform for collaborative funding applications, and online publishing capabilities to support coordinated international action to accelerate Genome to Phenome (<http://faang.org/>). We helped to support the GO-FAANG meeting that took place at the National Academy of Sciences building in Washington DC on October 7-8, 2015.

An increasing number of web hits and data downloads continued in 2015. For example, AnimalGenome.org received over 14.4 million web hits from 624,000 individual sites (visitors), which made 3.8 million data downloads that generated over 3 TB of internet traffic.

Reaching out

We have been sending periodic updates to over 2,800 users worldwide to inform them of the news and updated information we develop or host at AnimalGenome.org. New items were updated to the community on an ongoing basis in 2015.

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

Zhi-Liang Hu, Carissa A. Park and James M. Reecy (2016). [Developmental progress and current status of the Animal QTLdb](#). *Nucleic Acids Research (Database issue; Advance Access)*: doi: 10.1093/nar/gkv1233)..