

USDA NRSP-8 BIOINFORMATICS COORDINATION PROGRAM ACTIVITIES

Supported by Regional Research Funds, Hatch Act for the Period 10/1/15-9/30/16
James Reecy, Sue Lamont, Chris Tuggle, Max Rothschild and Fiona McCarthy,
Joint Coordinators

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 continually been improved, and continued data generation by the group has increased the amount of housed data. 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.

Multi-species support

The Animal QTLdb and the NAGRP data repository have been actively supporting the research activities for multiple species. The database has been set up to accommodate curation of catfish

QTL/association data. The collaborative site at iPlant continues to play an integral role in sharing the web traffic load by hosting the JBrowse server to serve the cattle, chicken, pig, sheep, goat, 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. The virtual machine site to host the Online Mendelian Inheritance in Animals (OMIA) database (Dr. Frank Nicholas at the University of Sydney; <http://omia.animalgenome.org/>) and [Striped Bass Genome Database](http://stripedbass.animalgenome.org/) (Benjamin Reading of North Carolina State University; <http://stripedbass.animalgenome.org/>) continues to provide collaborative researchers convenient tools to create, maintain, and manage their sites with complete control. The migration of OMIA to NAGRP platforms is in the planning stages.

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 is available on NCBO's BioPortal (<http://bioportal.bioontology.org/ontologies/LPT>). We have also continued mapping the cattle, pig, chicken, sheep, and horse QTL traits to the Vertebrate Trait Ontology (VT), LPT, and Clinical Measurement Ontology (CMO) to help standardize the trait nomenclature used in the QTLdb. At the request of community members, at least 14 new terms were added to the VT in 2016. Now the VT data download has been made possible through the Github portal (<https://github.com/AnimalGenome/vertebrate-trait-ontology>) where users can automate their data updates. 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/LPT/CMO cross-mapping has been well employed by the Animal QTLdb and VCMAP 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, as well as requests from community members, 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 is also available on BioPortal (<http://bioportal.bioontology.org/ontologies/LBO>).

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 continually made.

The Virtual Comparative Map (VCmap) tool, developed from a collaboration between Iowa State University, the Medical College of Wisconsin, and University of Iowa (<http://www.animalgenome.org/VCmap/>), has been set up to run on new hardware. Continued development and maintenance work have been performed. Application development, improvement, and testing have continued. Online help materials have been added, including a written user manual and a video tutorial. Please feel free to try things out and send any feedback to vcmap@animalgenome.org. AgBase and the AnimalGenome.org websites provide multiple reciprocal reference links to facilitate resource sharing.

Minimal standards development

The Animal QTLdb has been continually developed to use MIQAS for data curation and data integration (<http://www.animalgenome.org/QTLdb/doc/minfo/>). We have continued to work on refining MIQAS to help define minimal standards for publication of QTL and gene association data (<http://miqas.sourceforge.net/>).

Expanded Animal QTLdb functionality

In 2016, a total of 57,229 new QTL/association data have been curated into the database, representing a 47% data increase from a year ago. Currently, there are 16,516 curated porcine QTL, 95,332 curated bovine QTL, 6,633 curated chicken QTL, 1,245 curated horse QTL, 1,412 curated sheep QTL, and 127 curated rainbow trout QTL in the database (<http://www.animalgenome.org/QTLdb/>). All data have been ported to NCBI, Ensembl, and UCSC genome browser in a timely fashion. 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 improvements include accommodation of multiple genomes for QTL/association mapping, allowing the inclusion of “supplementary data” to QTL/association publications as part of the supporting evidence to curated data, allowing the ‘rs’ SNPs to be curated prior availability of their official ‘rs’ numbers, automated inclusion of validated SNPs upon successful checks/curation of related QTL/association data, etc.

Further development of Animal Trait Correlation Database (CorrDB)

We have continued development of curation tools for the CorrDB to allow ongoing curation of trait correlation data into the database. Efforts have been made to make use of resources and tools already in the QTLdb for trait ontology development and management, literature management, and bug reporting tools for data quality control. The tools are nearly ready to release for public data entry.

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 1,201 data files on different animal genomes, supplementary data files to publications, and data

for other sharing purposes have been made available to community users. The data downloads from the repository generated over 5TB of data traffic in 2016. Our helpdesk is here to assist community members. Throughout 2016, the helpdesk handled over 80 inquiries/bug reports/requests for services from groups/individuals with their research projects. Our involvement has ranged from data transfer, data deposition, formulation for best data representation, 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, FAANG international consortium, CRI-MAP users, etc.

The Functional Annotation of ANimal Genomes (FAANG) web site (<http://www.faang.org/>) is hosted by AnimalGenome.ORG. The web site has been developed and maintained to serve not only as a FAANG-related information hub, but also as a platform for this international consortium's communication, collaboration, organization, and interaction. It serves over 430 members and 10 working committees and sub-committees, with 12 listserv mailing lists, a bulletin board, and a database for membership and working group management. The actively hosted materials include meeting minutes, presentation slides, and video records of scientific meetings and related events, all interactively available to members through the web portal. A recent addition to the site is an interactive "Funding Opportunities" page where scientists can open discussions and build collaborations.

An increasing number of web hits and data downloads continued in 2016. AnimalGenome.org received over 23.1 million web hits from 56,200 individual sites (visitors), which made 5.6 million data downloads that generated over 7 TB of internet traffic.

Site maintenance

A significant portion of our 2016 efforts involved migration of the AnimalGenome.ORG site to new hardware (a 2x8 core DELL server with 64GB RAM and 25TB storage running a new version of RHEL7). It took a significant amount of time for the planning, preparation, migration, and deployment of multiple servers/services from the old hardware. These servers/services included the AnGenMap listserv and web site, Animal QTLdb, Animal CorrDB, VCmap, NAGRP Data Repository, Cri-Map Users listserv and web site, EPIgroup listserv and web site, OMIA-Support group listserv and web site, etc. The migration also involved setting up a couple dozen functional software tools within the NAGRP Toolbox, as well as virtual machine (VM) sites for Online Mendelian Inheritance in Animals (OMIA) database and Striped Bass Genome Database (SBGD) developments. Much of the effort was invested in solving a number of hardware/software compatibility problems. We also devised an internal structure utilizing multiple computer servers to best host various services, such as GBrowse, JBrowse, Biomart, NCBI Blast, etc.

Reaching out

We have been sending periodic updates to about 3,000 users worldwide to inform them of the news and updates regarding AnimalGenome.org. “What’s New on AnimalGenome.ORG web site” emails were sent out 3 times in 2016.

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.