

NRSP-8 BIOINFORMATICS COORDINATION PROGRAM 2017 ACTIVITIES

Supported by Regional Research Funds, Hatch Act
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 (<https://www.animalgenome.org/lunny>) have continually been improved and updated with newly generated 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 QTLdb has been accommodating active curation of

QTL/association data for seven species (cattle, catfish, chicken, horse, pig, rainbow trout, and sheep). The collaborative site at iPlant continues to play an integral role in sharing the web traffic load by hosting JBrowse for interactive QTL/association data map 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 user's 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 the Hybrid Striped Bass website (Benjamin Reading of North Carolina State University; <http://stripedbass.animalgenome.org/annotator/index>) continues to provide collaborative researchers convenient tools to create, maintain, and manage their sites with complete control.

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 24 new terms were added to the VT in 2017. 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, CorrDB, 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; <https://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 (<https://www.animalgenome.org/bioinfo/tools/>). Software upgrades and bug fixes were continually made. The CateGORizer, ExpeditoR, VCmap, and other tools are continually used by community members.

AgBase and the AnimalGenome.org websites provide multiple reciprocal reference links to facilitate resource sharing.

Minimal standards development

The Animal QTLdb and CorrDB have been continually developed to use MIQAS for data curation and data integration (<https://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 2017, a total of 41,093 new QTL/association data were curated into the database, bringing the total number of data to 145,842 QTL/associations. Currently, there are 26,076 curated porcine QTL, 108,040 curated bovine QTL, 8,363 curated chicken QTL, 1,304 curated horse QTL, 1,932 curated sheep QTL, and 127 curated rainbow trout QTL in the database (<https://www.animalgenome.org/QTLdb/>). All data have been ported to NCBI, Ensembl, UCSC genome browser, and Reuters Data Citation Index 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; providing a "permanent record locator" for reviewers and authors to locate QTL/association data; and allowing mapping of trait data across the QTLdb and CorrDB databases using VT, LPT, and CMO ontology terms. In addition, trait-centric views and gene-centric views of QTL/association data have been developed to facilitate user analysis of data in terms of interrogating and displaying genotype-to-phenotype information in a synopsis. We have also enabled a function to allow web users to identify traits with both QTL/association data and genetic/phenotypic/environmental correlation data, and to traverse between the QTLdb and CorrDB for information.

Further developments of Animal Trait Correlation Database (CorrDB)

Our development of the CorrDB focused on co-development of curator tools and curation environments with that of the QTLdb. This helped with resources and tool sharing on trait ontology development and management, literature management, breed ontology management, and bug reporting tools for data quality control. The newly developed CorrDB curator tools are available to the public for any user to register for an account to curate correlation data.

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 and has been actively used (<https://www.animalgenome.org/repository>). New data is continually being added. A total of 345 new data files on different animal genomes and supplementary data files to publications have been added to the repository over the last reporting period, representing a 33% data increase. Nearly 50 researchers and/or labs used the NAGRP data share platform to transfer or share their

data files. This was double the number in the previous year. Over 20 groups chose to use our Supplementary Data platform to host files for their new publications, which was a 3-fold increase from the previous year.

The data downloads from the repository generated over 12TB of data traffic in 2017. Throughout 2017, our helpdesk at AnimalGenome.ORG handled over 90 inquiries/requests for services affecting community research projects. Our involvement ranged from data transfer and hosting, data deposition, web presentation, and data analysis, to software applications, code development, etc.

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/websites for various research groups in the NAGRP community (<https://www.animalgenome.org/community/>). This includes groups like AnGenMap, FAANG international consortium, CRI-MAP users, and recent meetings like “Livestock High-Throughput Phenotyping and Big Data,” “Genome to Phenome: A USDA Blueprint for Improving Animal Production,” etc. A web service to facilitate gathering of signatures Calling for Restoration of NIFA/AFRI Foundational Program to Support the Animal Breeding, Genetics and Genomics Research played a positive role for the efforts.

The Functional Annotation of ANimal Genomes (FAANG) website (<https://www.faaang.org/>) is hosted by AnimalGenome.ORG. The website 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 467 members and 11 working committees and sub-committees, with 14 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. The “Funding Opportunities” information service has been improved to accommodate varying situations and to allow scientists to engage in open or private discussions to facilitate collaborations. Increases in the number of web hits and data downloads continued in 2017. AnimalGenome.org received over 8.6 million web hits from 625,282 individual sites (visitors), resulting in about 1 million data downloads that generated over 1 TB of internet traffic.

Site maintenance

Along with newly acquired computer servers, the NAGRP program has also retained and made good use of old hardware to form an internal networked development environment, where loads for data backup, virtual machine management, customer portal hosting, databases, and web services can be well distributed.

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 2017.

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.