NRSP-8 BIOINFORMATICS COORDINATION PROGRAM 2018 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/lunney) have continually been improved and updated with newly generated data. This resource continues to help the consortium by offering updated information and continued facilitation of data sharing and analysis for the consortium members.

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). In 2018, a total of **18,420** new QTL/association data were curated into the database, bringing the total number of data to **164,262** QTL/associations. Currently, there are 28,720 curated porcine QTL, 120,122 curated bovine QTL, 10,944 curated chicken QTL, 2,023 curated horse QTL, 2,325 curated sheep QTL, and 128 curated rainbow trout QTL in the database (over the past 5 years, a total of 141,260 QTL/association data were curated into the database (cattle: 111,817; chicken: 7,025; horse: 2,023; pig: 18,858; sheep: 1,536), representing an overall >87% of all data in the database. https://www.animalgenome.org/QTLdb/). An overhaul has been made to the Animal CorrDB with all new curation/query tools and completely re-curated data (a total of **12,515** correlations data on **448** traits, and **2,227** heritability data on **684** traits in **4** livestock animal species). The NAGRP data repository is hosting over 1,450 data files from catfish (61 files), cattle (231 files), chicken (125 files), horse (161 files), pig (228 files), rainbow trout (5 files), sheep (63 files).

The collaborative site at CyVerse 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). Eight (8) updated versions of the data set were released to the public throughout 2018. 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). Two (2) updated versions of the LPT were released this year. 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 55 new terms were added to the VT in 2018. The VT data download is available 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 continued to integrate 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 was updated and released 8 times during 2018, and 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 over the past 5 years.

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 using MIQAS for data curation and data integration (https://www.animalgenome.org/QTLdb/doc/minfo/) during and beyond past 5 years. 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/). New development in the Animal Trait Correlation Database (CorrDB) also benefited from the MIQAS framework.

Expanded Animal QTLdb functionality

All curated QTL/association data have been automatically 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 efforts included accommodating multiple genomes for QTL/association mapping/curation; introducing the use of DOI (Digital Object Identifiers) on data that have a DOI record, introducing the use of permanent data locators for each publication, development of an API tool set, improving QTLdb curator/editor tools to include management of a trait list with modifiers (for traits that are the "same" by nature but slightly different in terms of attached information); and making dynamic hyperlinks available within JBrowse for each alignment to link back to Animal QTLdb to traverse for linked information. New capability was added to allow the inclusion of "supplementary data" to QTL/association publications. In addition, a new genome-wide plot of QTL/association data and data enrichment analysis tool has been implemented for users to quickly identify over-represented data sets. QTL/association data and genetic/phenotypic/environmental correlation data links between the QTLdb and CorrDB have been further enhanced.

Further developments of Animal Trait Correlation Database (CorrDB)

Over a period of 3 years from 2016 to 2018, the Animal CorrDB has been undergone a re-development for higher data quality control standard and managements, improved curator tools, and better user interface. 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 improved 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. In 2018, 11,063 correlation data on 448 livestock traits, and 1,689 heritability data on 684 livestock traits were curated into the CorrDB. The public data web portals continue to undergo improvement.

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 curated. A total of 215 new data files on different animal genomes and supplementary data files have been added to the repository, representing an 18% data increase over the previous year. About 40 researchers and/or labs used the NAGRP data share platform to transfer or share more than 110 data files in 2018. Almost 40 groups chose to use our Supplementary Data platform to host files for their publications, which almost doubled the scale of the increase from the previous year.

The data downloads from the repository generated over 11TB of data traffic in 2018. Throughout the year, our helpdesk at AnimalGenome.ORG handled over 100 inquiries/requests for services affecting community research activities. Our involvement ranged from data transfer and hosting, data deposition, web presentation, and data analysis, to software applications, code development, advice for tool developments, 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 NRSP-8 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, and CRI-MAP users, among other user forums (https://www.animalgenome.org/community).

The Functional Annotation of ANimal Genomes (FAANG) website (https://www.faang.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 500+ members and 11 working groups and sub-groups, 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 2018. AnimalGenome.org received over 6.5 million web hits from 439,281 individual sites (visitors), resulting in about 1 million data downloads that generated over 1.2 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 were well distributed.

Reaching out

We have been sending periodic updates to over 3,000 users worldwide (https://www.animalgenome.org/community/angenmap/) to inform the animal genomics research community of the news and updates regarding AnimalGenome.org. "What's New on AnimalGenome.ORG web site" emails were sent out 3 times in 2018, in a consistent pace/pattern over the past 5 years.

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, and James M. Reecy (2019). **Building a livestock genetic and genomic information knowledgebase through integrative developments**

of Animal QTLdb and CorrDB. *Nucleic Acids Research, gky1084 (Published early online, Nov., 8 2018).*

2. Lisa Harper, Jacqueline Campbell, Ethalinda Cannon, Sook Jung, Monica Poelchau, Ramona Walls, Carson Andorf, Elizabeth Arnaud, Tanya Z Berardini, Clayton Birkett, Steve Cannon, James Carson, Bradford Condon, Laurel Cooper, Nathan Dunn, Christine G Elsik, Andrew Farmer, Stephen P Ficklin, David Grant, Emily Grau, Nic Herndon, Zhi-Liang Hu, Jodi Humann, Pankaj Jaiswal, Clement Jonquet, Marie-Angélique Laporte, Pierre Larmande, Gerard Lazo, Fiona McCarthy, Naama Menda, Christopher J Mungall, Monica C Munoz-Torres, Sushma Naithani, Rex Nelson, Daureen Nesdill, Carissa Park, James Reecy, Leonore Reiser, Lacey-Anne Sanderson, Taner Z Sen, Margaret Staton, Sabarinath Subramaniam, Marcela Karey Tello-Ruiz, Victor Unda, Deepak Unni, Liya Wang, Doreen Ware, Jill Wegrzyn Jason Williams, Margaret Woodhouse, Jing Yu, Doreen Main (2018). **AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture**. *Database,1:1–32.*

3. Zhi-Liang Hu, Carissa A. Park, and James M. Reecy (2018). **Development of Animal QTLdb and CorrDB: Resynthesizing Big Data to Improve Meta-analysis of Genetic and Genomic Information**. *The 11th World Congress on Genetics Applied to Livestock Production (WCGALP). New Zealand, February 7-16, 2018.*