

ANNUAL REPORT OF COOPERATIVE REGIONAL PROJECTS
Supported by Allotments of the Regional Research Funds, Hatch Act
January 1 to December 31, 2017

PROJECT: NRSP-8 Cattle Genome Coordinators

COOPERATING AGENCY AND PRINCIPAL LEADERS:

University of California, Davis: Juan F. Medrano

University of California, Davis: Alison Van Eenennaam, Co-coordinator

University of Missouri-Columbia: Jerry Taylor, Co-coordinator

Overview: The Cattle Genome Coordination Program is under the National Animal Genome Research Program (NAGRP). NIFA (National Institute of Food and Agriculture, formally) support is allocated from NRSP-8 and provided to the Agriculture Experiment Stations by off-the-top funding. The NAGRP is made up of the membership of the Animal Genome Technical Committee, including the Cattle Species Subcommittee.

NRSP8 Objectives: Objective 1: Create shared genomic tools, reagents and sequence information to enhance the understanding and discovery of genetic mechanisms affecting traits of interest. **Objective 2:** Facilitate the development and sharing of animal populations and the collection and analysis of new, unique and interesting phenotypes. **Objective 3:** Develop, integrate and implement bioinformatics resources to support the discovery of genetic mechanisms that underlie traits of interest.

Progress toward Objective 1. Shared genomic tools and reagents and sequence information.

An important focus of the community has been towards improving the bovine genome assembly.

Development of a new Dominette bovine genome reference assembly

A group of collaborating scientists (Tim Smith, Juan Medrano, Ben Rosen, Sergey Koren, Robert Schnabel, Derek M. Bickhart, Aleksey Zimin, and Chris Elsik) worked toward developing an improved bovine reference genome assembly and its annotation. Multiple data types were generated, such as PacBio long-read sequences, Illumina paired-end sequences, an optical map and improved gene predictions based on RNA-Seq and Iso-Seq data. All of the data were derived from tissue samples from L1 Dominette 01449 (Dominette), the reference animal. The goal of the group was to generate an improved “gold level” cattle reference genome assembly. Approximately 80X PacBio and 80X Illumina paired-end coverage of Dominette were produced. A Falcon *de-novo* assembly was generated followed by scaffolding of contigs using the Dominette Optical Map, Dovetail Genomics Chicago library/HiRise and HiC and a recombination map of 59K autosomal SNPs. In parallel, the Falcon assembly was refined using specialized software, CANU and MaSuRCA, resulting in chromosome length scaffolds. Additionally, full-length transcripts were sequenced using Iso-Seq from 30 tissues from Dominette to support improved annotation. Chris Elsik (University of Missouri, Columbia) has led the annotation effort for the new sequence assembly. A public version of the new ARS-UCD assembly will be submitted to NCBI in January 2018, shortly after the PAG meeting. Funding for this project came from Cattle Genome Coordinator Funds, USDA/MARC, UC Davis, Neogen/Geneseek and Zoetis.

As described by Bob Schnabel et al. (PAG abstract W151) comparing imputation accuracy between markers ordered according to the the new ARS-UCD assembly and UMD3.1, approximately 2,000 markers changed positions by greater than 10 Mb or are on different chromosomes in the new assembly. For these markers, average accuracy increased between 1–13%. The overall mean accuracy was increased modestly, but the largest increases in accuracy occurred for low minor allele frequency (<0.03) variants, which has significant implications for imputation to sequence level genotypes.

In addition to the Dominette ARS-UCD reference assembly, 2017 was a productive year to generate new cattle assemblies that will provide useful tools for research and commercial applications:

- Zoetis (G. Rincon, PAG abstract W157) generated a complete high quality *de-novo* Holstein bull annotated reference assembly.
- A collaboration between USMARC, University of Adelaide and NHGRI (A. Rhie PAG abstract W149) reported the development of a unique assembly from an F1 hybrid (Angus x Brahman) and a novel method for resolving completely phased haplotype assemblies from both breeds.
- LIC, New Zealand, Ben Rosen at USDA/Beltsville and others are completing long read *de-novo* assemblies of a Holstein bull and a Jersey bull.
- A draft genome assembly and annotation for river buffalo was published (Giga Sci 6(10), 2017).

Functional Annotation of Animal Genomes (FAANG) initiative

Huaijun Zhou and collaborators at U.C. Davis with the support of both USDA NIFA and NRSP8 Bovine Genome Coordinator funds are following the blueprint of the human and mouse ENCODE projects for identifying the functional roles of regulatory elements in the cattle genome. The goals are to annotate promoter, enhancer, and silencer region specific chromatin marks, and to determine the functional roles of regulatory regions in relevant tissues in each species. Cattle tissues were collected from four (2 males and 2 females) 14 month old Line 1 Hereford animals. The samples from eight tissues (skeletal muscle, liver, adipose, spleen, hypothalamus, brain cortex/whole, cerebellum, and lung) from the animals were processed for RNA-Seq, DNase-Seq, ATAC-Seq, ChIP-Seq, DNA-methylation, Hi-C and other assays. Data are being integrated to functionally annotate regulatory elements within the bovine genome.

We expect that the FAANG cattle initiative will be significantly expanded with added collaborations and assays supported by the recent FAANG Program Area NIFA funding.

Progress towards Objective 3: Bioinformatics and database resources

Cattle GRIN Genomics Database

Harvey Blackburn at USDA-ARS National Animal Germplasm Program (NAGP), Colorado State University Experimental Station and EMBRAPA have joined efforts to develop a genomic database that will serve as a repository for DNA data from the large animal genomics projects which have valuable data that need permanent archiving for future research. This effort, coupled with the existing capacities to store phenotypic and production system data in the Animal-GRIN database as well as germplasm/tissue samples, will facilitate the communities' efforts to maintain genomic data for future use. Efforts have also been ongoing to interface the Animal-GRIN system with the Internet 2 effort, which the USDA ARS, as a whole, has been engaged in developing. The

genomics database is ready to accept data. Web-interface on front-ends/back-ends are being built to facilitate user requests.

Database and bioinformatics activities are also coordinated by Jim Reecy (NRSP8 Bioinformatics Coordinator) at the NAGRP site (<http://www.genome.iastate.edu/cattle/>).

Meetings: Coordinator funds supported student travel awards for PAG-XXVI in January 2018, and will do the same for PAG XXVII in January 2019.

Plans for the future: Priorities are to support the continued efforts towards the annotation and release of the bovine genome reference assembly and the cattle FAANG initiatives, and data sharing and the creation of sample and data repositories that will benefit other cattle research investigators. We will expand our efforts to include the cattle industry and international collaborators in the application and the utilization of cattle genomic resources. For any informational items that you would like distributed via future newsletters please contact Alison Van Eenennaam (alvaneennaam@ucdavis.edu) or either of the two other co-coordinators. Constructive suggestions from researchers on areas to support in bovine genomics are also welcomed.