ANNUAL SUMMARY REPORT

PROJECT: NRSP-8
PROJECT TITLE: National Animal Genome Research Project
PERIOD COVERED: January 1 to December 31, 2016
ANNUAL MEETING DATE: January 14-15, 2017

OBJECTIVES

Objective 1: Advance the status of reference genomes for all species, including basic annotation of worldwide genetic variation, by broad sequencing among different lines and breeds of animals.

Objective 2: Develop strategies to identify and exploit genes and allelic variation that contribute to economically relevant phenotypes and traits, in part through improving functional annotation of the genomes of our species.

Objective 3: Facilitate analysis, curation, storage, distribution and application of the enormous datasets now being generated by next-generation sequencing and related "omics" technologies with regard to animal species of agricultural interest.

MINUTES OF THE ANNUAL BUSINESS MEETING:

The NRSP-8 business meeting was preceded by two days of species workshops, area subcommittees, and the combined Animal Genome Workshop presented on Sunday afternoon. A total of 105 participants from 22 countries attend the workshop.

The combined workshop included four plenary presentations as follows:

Dr. Zhanjiang Liu from Auburn University delivered the NRSP8 Distinguished Lecture entitled “Underwater Genomics: Exploiting Fishes' Unique Biology to Meet the Needs of the Aquaculture Industry”. Dr. Noelle Cocket from Utah State University presented “Recent Advancements in Sheep Genomics Research”. Dr. Claire Rogel-Gaillard from GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France presented “Microbiomes As Actors of Host Phenotype Variability: An Increasing Research Field in Animal Biology”. Dr. Gerald Quon from University of California, Davis presented “Computational Integration of Epigenomic and Functional Genomic Data for Fine-Mapping Complex Trait Loci”. Dr. Trey Ideker from University of California, San Diego presented “Building Hierarchical Models of the Cell from Genomics Data”.

The business meeting was called to order by the Chair, Huaijun Zhou (roughly 5:30 pm), with Mohamed Salem serving as the secretary. Huaijun Zhou called on each species coordinators to give their report. Below is a list of the species reports. The specifics of each report were submitted in their annual report. Notable highlights of these 7 reports follows the list.

a. Equine – Ernie Bailey
b. Swine – Chris Tuggle and Cathy Ernst
c. Aquaculture – Steven Roberts (in place of John Liu)
d. Cattle – Juan Medrano, Alison Van Eenennaam, Jerry Taylor
e. Sheep/Goat – Stephen White
f. Poultry – Hans Cheng
g. Bioinformatics – James Reecy

Notable highlights: Equine – Ernest Bailey: New FAANG project funded by NIFA. Swine – Chris Tuggle: New genome assembly with improve annotation by adding more tissues, industry representative suggested for this workshop. Aquaculture – Steven Roberts: New catfish and oyster genomes, 50K SNP-Chip for rainbow trout, workshop for aquaculture genetic improvement held at Auburn University, FAASG new consortium for all salmonid genome annotation. Cattle – Juan Medrano: A greatly improved genome reference (include 80X PacBio); 110 tissues from 4 animals were collected and are available for FAANG community; a new pheno/geno database. Sheep/Goat – Stephen White: Sheep genome improved assembly, tissues from 31 individuals collected for FAANG genome annotation. Poultry –Hans Cheng; Greatly improved built-6 genome assembly (PacBio & HiSeq), 6 micro-chromosome still missing in the assembly. Bioinformatics – James Reecy, "cyberinfrastructure" meeting this summer, funds available and suggestions are welcome; efforts have been made to move to a new server, AnGenlist and Listserv have been heavily used during last year.

1. Administrator reports

a. Eric Young

New NRSP-8 proposal is due by Jan 15, 2018

b. Lakshmi Matukumalli

NIFA greatly support the work of this group. AFRI Food security and foundational program call will be available with a deadline in May.

2. Call for old business: no items requested or presented.
3. Nominations for next business meeting: no items requested or presented. Location confirmed as San Diego for next year.
4. Nominations for Secretary/Chair-elect:

Molly McCue from University of Minnesota was nominated for the Secretary in 2018 and Chair in 2019. Second was obtained from and the motion passed unanimously. Molly had been notified of the intention to nominate and had agreed to accept the nomination prior to the business meeting.

5. "pass the gavel" to Mohamed Salem.

After passing the gavel, the NRSP8 community thanked Huaijun Zhou for his leadership in the last year during a time of transition for the NRSP8 projects.

ACCOMPLISHMENTS AND IMPACTS:

AQUACULTURE TECHNICAL REPORT

Coordinator:	Dr. John Liu
Co-Coordinator:	Dr. Steven Roberts
Administrative Supervisor:	Dr. Susan Brown
National Program Leader:	Dr. Lakshmi Kumar Matukumalli
Industry Representatives:	Mitt Walker, Dr. Scott LaPatra
Species Leaders	
Salmonids:	Dr. Yniv Palti
Catfish:	Dr. Sylvie Quiniou

Oysters:	Dr. Dina Proestou
Striped Bass:	Dr. Craig Sullivan

Objective 1

Draft white bass (Morone chrysops) genome sequence assembly was produced and is being annotated and compared to the corresponding striped bass (M. saxatilis) sequence in research on superior growth traits of Morone hybrids.

The channel catfish reference genome sequence was published in Nature Communications in June 2016. The assembly was validated by concordance of markers on the reference genome sequence and on the genetic linkage map constructed by using the 690K SNP arrays.

The genome for the immunome reference channel catfish CCBL1 and a doubled haploid blue catfish genome was sequenced using Pacific Biosciences technology.

A genome assembly project for Chinook salmon was initiated and several genome assembly approaches were evaluated.

A new tilapia genome assembly has been completed. It has been released under NCBI accession MKQE00000000. https://www.ncbi.nlm.nih.gov/assembly/GCA_001858045.2/

The genome of the Eastern oyster has been sequenced and a draft genome assembly of the Olympia oyster is available.

Objective 2

A family of host-defense peptides (piscidins) and their patterns of tissue expression during development were characterized in striped bass and its hybrids. Machine learning models were used to accurately predict striped bass egg quality from proteomics and RNA-Seq data and confirm involvement of dysfunctional cell cycle. Epigenetic markers of sperm quality (methylation status) and gene were examined.

A high-throughput micro satellite genotyping protocol was developed in catfish. This technique provides flexibility in locus selection and the ability to genotype microsatellites, SNPs, and simple indels in the same multiplexed PCR. MAG is a useful adjunct to SNP arrays for rapid parentage and kinship analyses in large populations. In addition to the QTL identified for columnaris disease resistance, QTLs for disease resistance against ESC disease, for head size, heat tolerance, and low oxygen tolerance have been developed. .

A 50K cSNPs array was developed from allelic-imbalance analysis of pooled RNA-Seq samples that may be associated with muscle yield and fillet quality traits and also with bacterial cold water disease survival in rainbow trout.

Identified patterns of selection in natural populations of various salmonid species (steelhead, sockeye salmon, and cutthroat trout) and candidate genes under selection (Hand et al. 2016; Hess et al. 2016a; Kovach et al. 2016; Nichols et al. 2016).

The CRISPR targeted genome editing technique was successfully applied in rainbow trout to reduce and eliminate expression of a functional tyrosinase protein, which resulted in fish with variations in eye and skin pigmentation. This proof-of-concept study determined that the new technology is a viable approach to reduce or knockout gene expression in rainbow trout.

A high-density SNP array containing ~40K Crassostrea gigas SNPs and ~15K Ostrea edulis SNPs has been developed.

Using C. gigas RNAseq resources, ~12,000 long intergenic noncoding RNAs (lincRNA) were identified and expression characterized to better understand their role in the development of early life stages (through metamorphosis and settlement).

High-throughput RNA-sequencing data used to conduct genome-wide survey of alternative splicing (AS) in C. gigas. 16% of oyster genes undergo AS, and many of them are involved in stress adaptation.

Objective 3
A JBrowse integrated web portal of the draft striped bass genome resource including gene predictions, relational BLAST matches, and InterProScan results, and data downloads is hosted online for use as an unrestricted public resource. Database URL: http://appliedecology.cals.ncsu.edu/striped-bass-genome-project/.

Salmonid researcher contributed to development of FishGen.net database for storage of large-scale genotypes for genetic tagging and monitoring studies.

The new rainbow trout reference genome including chromosome sequences and gene annotation was submitted to the NCBI.

Proposal to hold comparative genomics workshop funded through NRSP8 Aquaculture program. This workshop will preview the draft eastern oyster genome assembly and highlight how this resource can be exploited to answer key questions in molluscan biology.

CATTLE TECHNICAL REPORT

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.

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, Aleksey Zimin, Derek M. Bickhart) are working toward improving the bovine reference genome assembly and its annotation. Multiple data types have been generated, such as an optical map, Illumina paired-end, PacBio sequence, 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, the reference animal. The goal of the group has been to generate an improved "gold level" reference genome assembly. Approximately 80X PacBio and 80X Illumina paired-end coverage of Dominette (L1 Dominette 01449) has been produced. A Falcon de-novo assembly was generated followed by scaffolding of contigs using the recently generated Optical Map using technology from Dovetail Genomics Chicago library/HiRise 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 with the exception of the X chromosome. Additionally, full-length transcripts were sequenced using Iso-Seq from 30 tissues from Dominette to support improved annotation. Chris Elsik (University of Missouri, Columbia) will lead the annotation effort of the new sequence assembly. A public version of the new ARS-UCD assembly is expected to be released early in 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. Similar efforts are being implemented in pig and chicken, as part of the AGENCODE project. 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 steers. The samples from eight tissues (skeletal muscle, liver, adipose, spleen, hypothalamus, brain cortex/whole, cerebellum, lung) from the bulls were processed for RNA-seq, ATAC-seq, ChIP-seq, and other assays. Data will be integrated to functionally annotate regulatory elements of the bovine genome. Additionally, a large number of tissues were collected from the four animals which are available to collaborators to perform these assays available on the FAANG site (http://www.faang.org).

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 needing 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. Further web-interface work on front-ends/back-ends is being completed to facilitate user requests. It is anticipated that this system will be fully functional in 2017.

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

Meetings: Coordination funds supported student travel awards for PAG-XXV in January 2017, and will do the same for PAG XXVI in January 2018.

Plans for the future: A priority is to support the continued efforts towards the completion and release of the bovine genome reference sequence assembly in 2017, and to support data sharing and the creation of sample and data repositories that will benefit other cattle research investigators. We will expand our efforts to include international collaborators and the cattle industry, and expect to keep the Cattle Genome Community informed of developments and activities of the Cattle Genome Coordinators through an annual newsletter. If you have any informational items you would like distributed via this newsletter please contact Alison Van Eenennaam (alvaneenennaam@ucdavis.edu) or either of the two other co-coordinators. Constructive suggestions from researchers on areas to support in bovine genomics are also welcomed.

HORSE TECHNICAL REPORT

NRSP8 Horse Coordinators: Ernest Bailey (University of Kentucky); Molly McCue (University of Minnesota), Samantha Brooks (University of Florida)

Progress towards objectives:

Objective 1:

New Reference Genome Assembly

Ted Kalbfleisch Ted, Jamie MacLeod and Ludovic Orlando were funded by the Morris Animal Foundation to create a new assembly of the reference sequence, the putative Ecab 3.0. Partial support for a postdoctoral student is provided by USDA-NRSP8 coordinators' funds. The grant proposal and work is underpinned by data provided by workshop participants including whole genome sequence information from TWILIGHT

(reference horse) and from horses of other breeds. The work began in 2015 and a report provided during the July 2016 conference of the International Society for Animal Genetics (ISAG) in Salt City, Utah. The new assembly improved gene annotation, increased contig N50 from 112 KB to 1.4 MB, and eliminated most of the regions with ambiguous sequence ("Ns"). Some revisions are being made and the work will be published in 2017.

Objective 2:
New SNP assay tool
The 670K SNP chip was made available in late 2015 for research on horses. This initiative was driven by Dr. Molly McCue (co-coordinator) of the University of Minnesota with support of students, co-workers and funding from several agencies including the USD-NRSP8 coordinators' fund. Geneseek (NE) is a commercial laboratory offering testing. The Illumina SNP70 continues to be available. These resources have been used for gene discovery during 2016 as reported in the accompanying bibliography, as well as reports made at various conferences during 2016.

Objective 3:
In 2016, three scientists (Carrie Finno of UC Davis, Rebecca Bellone of UC Davis and Jessica Peterson of the University of Nebraska) were awarded funds by the Grayson Jockey Club Research Foundation to collect tissues and begin conducting assays associated with the FAANG program. The work is also supported by the Horse Genome workshop community and funds have been provided from the coordinators USDA-NRSP8 funds to support tissue collection. The workshop is using an "adopt a tissue" approach for scientists to fund characterization of tissues of interest. And organizational meeting, supported by the Dorothy Russell Havemeyer Foundation, was held at the July 2016 ISAG conference to coordinate sharing of the tissues and work to characterize their gene expression.

In connection with the FAANG initiative, coordinator funds were used to support participation by Dr. Ted Kalbfleisch at the European Hackthon, an initiative designed to make the annotation information derived from the FAANG activities available for research.

Database Activities:
Several genome browsers have been developed at the University of California, Santa Cruz, ENSEMBL and NCBI: http://www.genome.ucsc.edu/cgi-bin/hgGateway?hgsid=95987985&clade=vertebrate&org=Horse&db=0; http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9796; http://www.equinegenome.org/Equinegenome.org.htmlhttp://pre.ensembl.org/Equus_caballus/index.html.
A SNP database is available: http://www.broad.mit.edu/mammals/horse/.
A major entry point for databases and other relevant information about the horse genome workshop and participants is the workshop website: http://www.uky.ledu/AG/Horsemap.

International Efforts: The horse genome technical committee is an international activity with approximately one third of the participants coming from Europe, Africa and Australasia while the other half come from North America. Approximately 60 people participated in the workshop meeting in San Diego during January 2016. Jessica Petersen (Jessica.petersen@unl.edu) will chair the 2017 Horse Workshop and Stephen Coleman (Stephen.coleman@colostate.edu) will serve as vice-chair.

Communication: Communication within the horse genome workshop is facilitated by an email list for sharing information by the Horse Genome Coordinator and through the website: http://www.uky.edu/AG/Horsemap. One of the major aspects of the website is to increase its value for informing members of the horse industry about the scientists using horse genomics to solve important problems and to explain the value of horse genomics

Travel and Meeting Support: For the 2016 Hackathon meeting in Denmark, coordinator funds were used to support participation by Ted Kalbfleisch. For the January 2016 PAG conference, travel awards were provided to 19 students, including one Jorgenson award, and travel support for two invited speakers to the Horse Genome Workshop.

Future Activities: During 2017 a workshop on Horse Genomics will be conducted at the Conference of the International Society for Animal Genetics (ISAG) in Dublin, Ireland.

POULTRY TECHNICAL REPORT

FACILITIES AND PERSONNEL: Mary Delany, Dept. of Animal Science, UC Davis, served as Coordinator with Hans Cheng, USDA, ARS, ADOL, as Co-Coordinator.

Objective 1
Given the continued interest in further improvements to the chicken genome reference, a new build (Gallus_gallu-6.0) was generated that included PacBio long single molecule reads. For this build, most of the assembled contigs were ordered and oriented using a proximity-based map into 39 scaffolds at 94% of the assembly size; 865 remain as unplaced contigs. Compared to Gallus_gallus-5.0, there are 87 scaffolds and 28 contigs that required manual breaks due to de novo assembly or scaffolding errors. Gallus_gallus-6.0 assembled bases encompass 1.05 Gb, with an N50 contig and scaffold lengths of 18 and 35 Mb, respectively. Of the 1.05 Gb genome, ~94% of the assembled bases have been anchored to chromosomes. The next steps will be to assign large unplaced contigs to the remaining unknown autosomes after validating no genetic linkage exists to any known autosomes. In addition, we plan to add novel reference sequence to clusters of immune system genes (MHC) that are challenging to de novo assemble and genotype. Gallus_gallus-6.0 reference is a substantial improvement in base contiguity and autosomal assignment, however, the use of more targeted approaches will be required to elevate the chicken genome to levels comparable to human. Several members requested and received seed funding for special genomics projects including further sequencing the MHC-RFP-Y complex on GGA 16, sequencing of GGA 12 region responsible for the

wingless-2 mutation, microchromosome sequencing, and a study of highly pathogenic NDV susceptible/resistant lines.

Objective 2
DNA from the East Lansing international reference mapping population continues to be provided to many laboratories throughout the world. Similarly, DNA from the junglefowl used to generate the reference sequence assembly has been widely distributed, especially for copy number variant studies. Also member labs share specific poultry resources among community members (e.g., UC Davis provides hatching eggs for UCD 001 Red Jungle Fowl, UCD 003 Single Comb White Leghorn) including NRSP-8 members and also members of other multistates. A proposal was submitted for NRSP-8 poultry coordinators from Purdue University for seed funding for creation of a special synthetic population for the research community and its development will ensue sourcing from lines held at multiple member institutions and potentially commercial stocks.

Objective 3
Database activities are led by the NRSP-8 Bioinformatics Coordinator, Jim Reecy, and Susan Lamont, along with Fiona McCathy, representing poultry interests on the advisory committee for this group.

Impact:

1. Lay the foundation to generate a comprehensive data resource to be used by animal genome community and livestock industry.
2. Enable the application of molecular phenotype to the prediction of complex phenotypes and further our understanding of genetic mechanisms.

POULTRY FAANG UPDATE:
More than 39 tissues were collected from 4 F1 birds (line 6 X 7, 2 males and 2 females) in ADOL. 16 RNA-seq libraries from two biological replicates (male) across eight tissues: adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle and spleen were analyzed. Tissue specific long non-coding RNAs were identified and compared with pig and cattle with the same tissues. An analysis of 15 DNase-seq libraries show that identified tissue-specific DNase hypersensitivity (DHS) sites are associated with genes that relate to unique biological functions of the organs or tissues. Twelve dataset from H3K4me3 and H3K27me3 histone modification in chickens were analyzed. Integrative analysis of DHS sites, ChIP-seq (H3K4me3 and H3K27me3 histone modification marks), and RNA-seq allows the identification of genome-wide active and inactive promoter regions in chickens, which enables an in-depth comparison of the regulatory landscapes of multiple tissues within these species. ATAC-seq in chicken tissues was optimized and some promising results have been achieved.

MEETINGS:
Over 3,700 scientists attended the joint Plant and Animal Genome XXIV meeting January 2017, held jointly with the annual NAGRP meeting and the well-attended poultry workshop with invited speakers and also presentations from NRSP-8 and NC-1170 members. Coordination funds helped support attendance.

Travel support for students - Ma Li (University of Maryland), Recipient of Neal Jorgenson Genome Travel Award**.** We also provided travel support to students, post-docs, member PIs and speakers (~15).
Travel support for NRSP8 speakers - Dr. Gerald Quon (UC Berkeley)

IMPACT: This project is generating continues to generate improved resolving tools through which the genome sequence can be used to locate inherited production trait alleles and apply the DNA sequence to ascertain the physiological basis for those traits. It has resulted, among other things, in the generation of the complete sequence of the chicken and now the turkey genome. Commercial breeders are using the sequence and SNP we generated to characterize and improve production lines using genomic selection. In simpler terms, we are now moving closer to understanding the cause of phenotypic variation that is relevant to the agricultural use and improvement of poultry.

SHEEP AND GOATS TECHNICAL REPORT

A. NRSP-8 Sheep Genome Coordination

COOPERATING AGENCY AND PRINCIPAL LEADERS:

Utah State University: Noelle E. Cockett
Washington State University: Stephen White

PARTICIPANTS:
Louisiana State University: James E. Miller*1
Oklahoma State University: Udaya DeSilva*
Pennsylvania State University: Wansheng Liu*1
Texas A&M University: Clare Gill*, Penny Riggs*
University of Florida: Raluca Mateescu*1
University of Massachusetts-Amherst: Janice Telfer*, Cynthia Baldwin*
USDA/ARS: Michelle R. Mousel*1, Stephen N. White*1
Utah State University: Noelle E. Cockett*1
Virginia State University: Brian Sayre*1, Glenn Harris*
Virginia Tech: Rebecca Cockrum*1

*Voting member. In attendance at the 2017 NRSP-8 meeting.

BRIEF SUMMARY OF MINUTES OF ANNUAL MEETING:

The 2017 annual meeting of the NRSP-8 Cattle, Sheep, and Goat committee was held on Jan 14-15, 2017 in conjunction with the Plant and Animal Genome XXV meeting. The morning session of the scientific meeting on Jan 14th was held as a joint session in with the Swine Committee with 4 presentations on genome assembly improvements and application of genomic information for phenotypic improvement. The Saturday afternoon and Sunday morning sessions of the combined Cattle/Sheep/Goat workshop included 12 presentations covering a wide variety of topics from genome assemblies,

genome annotation, genomic control of the microbiome, pleitotropic variants, genetic interaction, a causal variant in sheep fleece variation, and genomics/genetics of water intake and bovine respiratory disease. Attendance at the sessions was good with more than 100 people attending the scientific sessions, including delegates from at least 11 countries, including Australia, Canada, United Kingdom, France, South Africa, Austria, Denmark, China, Switzerland, Russia, and Kuwait. Jared Decker was thanked for serving as President of the NRSP-8 Cattle, Sheep and Goat Committee in 2016-17. Brenda Murdoch will serve as President in 2017-18. Rebecca Cockrum was elected as the 2017-2018 Secretary, and she will serve as President in 2018-2019.

ACCOMPLISHMENTS AND IMPACTS:

Objective 1

WC1 co-receptors belong to the scavenger receptor cysteine-rich (SRCR) superfamily and are encoded by a multi-gene family. Each type I transmembrane bovine WC1 protein contains an extracellular SRCR domain arrangement that can be characterized as [a1-(b2-c3-d4-e5-d6)-(b7-c8-d9-e10-d'11)] or [a1-(b2-c3-d4-e5-d'11). The repetitive nature of the exon duplication and gene duplication makes annotation of the WC1 locus challenging; thus techniques worked out in this project will have an impact on annotation of other gene and exon duplicated loci.
Expression of particular WC1 genes defines functional subpopulations of lymphocytes known as WC1 gamma delta T cells. WC1 genes can be grouped as WC1.1-type or WC1.2-type based on the sequence of their N-terminal a1 SRCR domain, which is 4 amino acids longer for WC1.2 type a1 SRCR domains. Reciprocal expression of either WC1.1-type or WC1.2-type proteins is correlated with gamma delta T cell responsiveness to pathogens, which explains how so many large homologous open reading frames have been maintained for millions of years. We have previously identified complete genomic sequences for 13 different bovine WC1 genes through annotation of the bovine genome Btau_3.1 build. The genomes of sheep, goat and swine did not yield as many WC1 genes as we would have predicted. Extrapolating from our work in cattle, and of those genes we annotated, many appeared truncated. In 2015, to obtain an estimate of the extent of the WC1 gene array in livestock species other than bovine, we PCR-amplified N-terminal a1 SRCR domain sequence. We focused on the a1 domain because: (i) it is the signature domain for WC1 genes (vs. CD163 genes, which do not contain a1 SRCR domains), (ii) there is only one a1 domain in each bovine WC1 molecule, (iii) this domain varies most among bovine WC1 genes distinguishing them from one another, and (iv) it is coded for by a single exon. Based on a sequence of 100 ovine, 80 caprine, and 23 swine clones, we estimated ~26-30 WC1 genes in the ovine and caprine genomes and 3 in the swine genome.

In 2016, we annotated the PacBio-sequenced San Clemente goat genome in collaboration with Derek Bickhart & Tim Smith. We also obtained WC1 cDNA evidence from Boer goat and Dorset sheep. Annotation of the San Clemente goat genome led to 16 predicted WC1 gene models, including one gene with 11 duplicated a1 SRCR domains at the N-terminus instead of a single a1 SRCR domain. This a1 SRCR domain duplication

complicates using the number of WC1 a1 SRCR domains in the genome as equivalent to the number of WC1 genes in the genome. Most of the a1 SRCR domains from the annotated genome corresponded to those we previously sequenced, with the notable exception of an a1 SRCR domain that is the homolog of the bovine WC1-12 a1 SRCR domain. This is notable because of the bacterial binding properties of the bovine WC1-12 a1 SRCR domain, which predicts the same for the caprine homolog. Full-length cDNA evidence was obtained for 4 of the 16 WC1 gene models, 1 cDNA was found that does not have a counterpart in the San Clemente goat genome, and one apparent chimeric cDNA was found, casting doubt on two of the gene models (5 and 12). We also obtained two ovine full-length WC1 clones. In light of the apparently large number of caprine and ovine WC1 genes and the difficulty of cloning large and extensively alternatively spliced inserts, we are considering using next-generation sequencing on PCR-amplified WC1 templates.

Using the 5' RACE technique on swine Duroc x Yorkshire F1 cross cDNA template, we obtained cDNA evidence for the 5' UTR, signal sequence, and a1 SRCR domain for 6 porcine WC1 genes. There are 3 WC1 genes containing an N-terminal WC1.2 type a1 SRCR domain, 1 WC1 gene containing an N-terminal WC1.1 type a1 SRCR domain, and 2 WC1 genes containing an N-terminal d1 SRCR domain. Confirmation of a signal peptide followed by N-terminal d1 SRCR domain represents a novel WC1 gene structure. There are two predicted WC1 proteins annotated in the current porcine genome assembly in GenBank: ssCD163L1 and ssCD163L2. ssCD163L1 (gene id 100144477) has been corrected to link a WC1.2-type a1 SRCR domain with a WC1.2-type b2 SRCR domain. The ssCD163L2 (gene id 100627089) assembly has been corrected to remove a CUB domain and contains a signal peptide followed by an N-terminal d1 SRCR domain in place of an a1 SRCR domain. This initially appeared to be an assembly error, as all bovine WC1 genes have an N-terminal a1 SRCR domain. We predict based on the cDNA evidence that there are 4 WC1 genes missing from the current porcine genome assembly in Genbank: one with a WC1.1-type a1 SRCR domain at the N-terminus, two with a WC1.2-type a1 SRCR domain at the N-terminus and one with a d1 SRCR domain at the N-terminus. We are currently annotating a PacBio sequenced swine genome in collaboration with Derek Bickhart and Tim Smith, employing 3' RACE to determine variability at the 3' end, and attempting to obtain full-length or contiged cDNA evidence.

Objective 2

On April 26, 2016, a female U.S. Rambouillet sheep was humanely euthanized and 102 unique tissues were collected from this animal as part of the ovine FAANG project. Genomic DNA from the ewe had been previously used for a new de novo assembly of the sheep genome using 70-fold, PacBio long-read sequencing. Thus, data generated from the FAANG core assays can be annotated directly onto the new ovine reference genome. Samples from all tissues were either snap frozen in liquid nitrogen and can be used for RNA-seq, ChIP-seq and other FAANG assays, or slowly frozen or processed into DNA for chromatin accessibility assays. Thirty people were involved in the tissue collection including ovine FAANG members Noelle Cockett, Alisha Massa, Brian Sayre, Michelle Mousel and Brenda Murdoch, as well as Utah State University pathologists and

veterinarians including Tom Baldwin, Rusty Stott, Arnaud Van Wettere, Gordon Hullinger, Jaqueline LaRose, Holly Mason, and Kerry Rood. Special thanks to Tracy Hadfield for organizing the collection event, Alisha Massa for aligning the sample collections, Codie Durfee (with assistance from Brenda Murdoch and Brian Sayre) for processing DNA for ATAC-seq assays and Michelle Mousel for coordinating and conducting blood and immune cell collections. Other FAANG members who participated in the planning process included Kim Worley, Stephen White, Brian Dalrymple, James Kijas, Tim Smith, and Mike Heaton.

The Ovine FAANG Project has been awarded a 2016 NIFA/AFRI grant (B. Murdoch PD). This project will contribute to the core activities of FAANG by providing transcriptome data and detailed annotation of genes and regulatory features in the sheep. The overarching goals are to deliver enhanced functional annotation of the ovine genome in order to enable research in sheep and facilitate the understanding of gene regulation in all livestock species by generating and distributing curated transcriptome data. This will be accomplished by characterizing the complexity of the ovine transcriptome and the regulatory signals that control the expression of the transcriptome using assays for coding and non-coding transcript isoforms and alternative splicing, promoters and cis-acting regulatory elements, histone modifications, DNA methylation and open chromatin across a wide range of sheep tissues. We will also annotate the most current ovine reference genome assembly so that all transcriptome and regulatory features are included and available for analysis. These tools and resources will be made available for studying the ovine genome through online databases, including those supported by the international FAANG Consortium and provide training in their use.

Lambs are sometimes born with a condition called entropion in which the lower eyelid is inverted, causing the bottom eyelashes to rub on the cornea which can lead to blindness if not treated. Treatment is commonly done by unrolling the eyelid and surgically stapling it in correct alignment for a few weeks. Previous reports on entropion have indicated that it is genetically controlled. In this study, samples from five paternal half-sibling families segregating for entropion were collected in 2014 and 2015. Two of the five sires were born at the Utah State University sheep facility; one was from a flock with high incidence of entropion and born with the condition while the other sire, normal at birth, was from a flock with no recorded entropion births in the last 7 years. The other three sires were purchased and their eye condition at birth is unknown. Forty eight of the 159 lambs produced by these five rams were born with entropion. In an attempt to identify genetic regions involved with the entropion eye condition, genomic DNA was extracted from all lambs, sires and dams in the five families and the DNA samples genotyped with the Illumina HD SNP chip. Analysis of the SNP genotypes and entropion was done using SNP & Variation Suite v8. (Golden Helix, Inc.). Preliminary results suggested associations between the entropion condition and SNP markers on ovine chromosomes 1 and 3. Additional analyses are underway to localize the significant regions and identify underlying genes or genetic regulatory factors.

While sheep scrapie has well-defined genetic resistance, the genetics of goat scrapie are not as well understood. There are multiple amino acid substitutions in the PRNP gene

which are known to be underrepresented in scrapie positive goats, but for which the incubation times from inoculation to development of disease have not been characterized. An oral challenge experiment has demonstrated highly significant extended scrapie incubation in animals singly heterozygous for either PRNP S146 or K222 that now extend beyond an average age of 6 years, which is also longer than the commercial lifetimes of many goats.

Entropion is thought to be a recessive genetic disorder and is defined as an inversion of the eyelid margin causing lashes or external hairs to rub against the ocular surface. A GWAS with over 500,000 SNP and 473 sheep found 3 genomic regions significantly associated entropion. Two of these chromosomes have previously been associated with entropion. Further evaluation of these regions is ongoing to identify the underlying causal mutation(s). Identification of the causal mutation(s) would allow producers to eliminate carriers of this defect and more quickly progress to an entropion free flock.

Ovine progressive pneumonia virus (OPPV), a lentivirus of sheep, infects a quarter of U.S. sheep. A collaborative research project identified a deletion near ZNF389 as consistently associated with severity of lentiviral infection in multiple sheep flocks under varying conditions. However, the underlying gene(s) and mutation(s) responsible were not identified, so efforts are ongoing to fine map the QTL in this region. Both OPPV and HIV are macrophage-tropic lentiviruses with similar genomic structure, and understanding of this variant may contribute to human medicine as well as animal agriculture.

Coxiella burnetii is a zoonotic gram-negative organism broadly endemic in most of the world, and domestic ruminants (especially small ruminants) are blamed for disease outbreaks in human populations. We are generating data in ruminants to begin host genetic dissection of C. burnetii susceptibility and shedding traits.

Genetic disorders affecting erythrocyte morphology can have harmful effects on health, such as increased red blood cell fragility, high cellular turnover, and inefficient trafficking of oxygen. A GWAS in sheep found one SNP was associated with increased mean corpuscular hemoglobin concentration. A divergent artiodactyl MYADM-like repeat was identified as potentially causative and was associated with increased ewe lifetime kilograms of lamb weaned. This mutation has been fully sequenced from a BAC and additional sequence variations have been identified. Once verified, these mutations might be used by producers in marker-assisted and/or genomic selection.

Selection for GI nematode resistance based on correlation of ewe and lamb fecal egg count. Gastrointestinal nematodes (GIN) have a major impact on small ruminant production in most of the US. Haemonchus contortus is the primary pathogenic GIN due to its blood feeding behavior, that can lead to anemia and death if left untreated. GIN have developed widespread resistance to most available anthelmintics and alternative methods for control in both conventional and organic production are needed. Previous work on identifying genetic markers to select GIN resistant sheep failed to provide such markers due to the complexity of the animal's genetic response to infection. Therefore, the more traditional approach of looking at desired phenotypes for selection may be more

realistic. Selecting lambs for low fecal egg counts (FEC) has been used to genetically enhance resistance to GIN in growing lambs. Recording lamb FEC has been incorporated into some national sheep, i.e., Katahdin Hair Sheep International (KHSI), genetic improvement programs. The evaluations are conducted by the National Sheep Improvement Program (NSIP). Ewes are vulnerable to parasite infection during the peri-parturient period and commonly experience a rise in FEC. This study assessed factors associated with the peri-parturient rise in Katahdin ewes and associated FEC in their lambs. Data was evaluated for 1,487 lambings by 931 Katahdin ewes from 11 farms. Ewe FEC was done at approximately 0, 30, and 60 d postpartum and lamb FEC was done at approximately 60, 90, and 120 d of age (approximately 1,400 lambs at each age). Responses across measurement times were evaluated by repeated-measures analyses. Ewe FEC peaked at approximately 30 d postpartum, and lamb FEC increased through approximately 120 d of age. Previous work has estimated heritability of FEC in Katahdin lambs from 0.26-0.46, depending on the statistical model used, at 90 d of age. The observed correlation of ewe (30 d postpartum) and lamb (90 d of age) FEC (0.09) was highly significant (<0.001). Thus, if estimated heritabilities of ewe and lamb FEC were assumed to be 0.4, the observed correlation of 0.09 would correspond to a realized correlation of 0.45. Therefore, this might have potential to select lambs as susceptible or resistant to GIN infection based on their dams FEC at 30 d postpartum. However, maternal input is only half of the genetic input. KHSI has established FEC EBV values for sires that participate in the program. Combining high –EBV value sires with low ewe FEC at 30 d postpartum should increase the value of the selection process. Thus, results support the presence of a genetic relationship between ewe and lamb FEC for parasite resistance.

Two undergraduate students were awarded a summer research fellowship through UMass CAFE and cloned and confirmed by sequencing 14 SRCR domains from goat and swine. They successfully expressed one of them as a recombinant protein and found that its binding affinity for Mycobacteria spp. was comparable to that of a bovine WC1 a1 SRCR domain recombinant protein. We will use amino acid polymorphism and conservation among WC1 genes from individuals and between WC1 SRCR domains to attempt to predict the gamma delta T cell response to Mycobacteria and other pathogens.

CNVs of PRAMEY and bull fertility in Holstein. In collaboration with Select Sires, we have recovered a lost Y-chromosome lineage (HO09101) by IVF with the frozen semen and embryo transferring technology. Four bull calves have been produced and their Y-chromosomes are under analyzed for the copy number of the PRAMEY gene.

Male fertility evaluation with a custom-made SNP chip in cattle. We reported the development and evaluation of a custom-made 384-SNP chip for bull fertility analysis in last few years. During 2015-2016, we further analyzed these SNP markers to identify the most effective SNPs for male fertility prediction. We divided the genotypic data of ~1000 bulls into two data sets: training and testing. The training datasets was randomly sampled for 40%, 50%, 60%, 70%, 80%, or 90% of the total bulls, while the testing datasets was the remaining (60%, 50%, 40%, 30%, 20%, or 10%) bulls for validation purpose. The predictive accuracy for the traits of interests, such as sperm total motility, by using

different number of the top ranking SNPs. This is an ongoing analysis to select and validate the top marker before we finalize the panel for the male fertility-specific genetic diagnostic assay.

Subcellular localization of the bovine PRAMEY protein in bovine spermiogenesis. We continued on the characterization of the PRAMEY protein in testes and spermatozoa. To further study the subcellular localization of PRAMEY, we performed immunogold electron microscopy (IEM) with mature bull testicular samples collected from a local slaughter house. The results showed that the bovine PRAMEY protein was expressed in all steps of spermatids and acrosome and flagellum of spermatozoa. The enrichment of gold particles was observed in several cell organelles, including Golgi vesicles, chromatoid body (CB), acrosomal granule, centrioles and flagellums at certain steps. In the Golgi phase (steps 1-3), clusters of labeling were first seen in small vesicles that contain a CB-like structure. The protein was consistently expressed during the formation of the CB. Clusters of gold labeling were also seen in the nucleus. In the cap/acrosome phase (steps 4-12), significant enrichment of gold particles was observed in the CB before the formation of the acrosomal granule, and during migration of the CB toward the caudal pole of nucleus. Along the formation of acrosome, significant gold labeling was observed constantly within the matrix of the acrosomal granule, where the gold particles formed a unique but unknown pattern that differed from the clusters of labeling observed from other organelles. The gold labeling was then seen clearly after flattening of the acrosomal granule, particularly in the matrix of acrosome. The PRAMEY labeling was observed in both proximal and distal centrioles, mainly in those centrioles that have migrated into the caudal pole of nucleus. Clusters of labeling were also observed in the space between the proximal centriole and nucleus, as well as microtubules in the distal centriole during the formation of flagellum. In the maturation phase (step 13-14), the gold particles were primarily located in the acrosomal matrix and the out layer of flagellum. Our results strongly suggest that the bovine PRAMEY plays a functional role during spermiogenesis.

Determine the functional role of PRAME in spermiogenesis using a Prame-knock-out (KO) mouse model. We have purchased the Prame first allele KO mice for generating a conditional KO (cKO) model for the spermiogenesis study. We have been also working to generate a mouse Pramel1 cKO. By collaboration with Dr. Jon Oatley, we are generating a direct KO mouse model for Pramel1 by CRISPR/Cas9.

Meiotic recombination - Homologous recombination or cross-overs (CO) contribute to genetic variation and importantly ensure proper chromosome segregation. Importantly, failure or improper placements of recombination represent a significant contribution to fetal loss and infertility. Despite the importance of these issues, we know very little about the factors that control meiotic recombination rates in livestock species. We characterized and quantified the number of recombination events from Targhee, Icelandic and Suffolk rams. The total number of CO per meiocyte was quantified and CO locations on the chromosomes were measured. In addition, we have characterized individual variation within each breed.

Sustainable Small Ruminant Production through Selection for Resistance to Internal Parasites. The project funded by USDA-NIFA-CBGP is the process of evaluating DNA markers to develop early-life indicators of resistance in small ruminant populations to internal parasitism. The project assessed the physiological conditions affected by selection for resistance to internal parasites. Current work includes evaluation of economic and management considerations of whole herd/flock selection for resistance to internal parasites. We are continuing to disseminate potential benefits of selection for resistance to internal parasites for adoption by small ruminant producers.
In 2016, DNA extractions were finished and analysis of potential DNA markers were performed. Fourteen candidate SNPs were identified in the exonic regions of two different locus: MHC Class II DR Alpha (DRA) and the MHC Class II DR Beta and only three SNPs were evaluated. The need to respond to pathogens such as gastrointestinal nematodes that evade immune responses led to high allelic diversity observed at ovine MHC loci. The ovine MHC Class II locus plays a central role in the immune system by presenting peptides derived from extracellular proteins to lymphocytes T activating TH2 immune response and is considered highly polymorphic. In order to identify substitutions associated with the control of nematode populations within the host, OLA-DRA and OLA-DRB genes were analyzed to genotype two hundred and ninety three animals by using High Resolution Melting assays. Animals were dewormed with levamisole (12.5 mg/kg of live weight) and albendazole (10 mg/kg of live weight) and infected with 10,000 L3 of H. contortus per kg of body weight per oral route and fecal samples were obtained to determine fecal egg count on day 21, 28, 35 and 42 post-infection. General Linear Models were fitted with MPCV, DMI, ADG, specie, year, breed and genotype as predictors and a mean of FEC as the response variable. According to the results, the best significant predictors to fit the models were Genotype, Breed (Species), Specie and Year ($p < 0.05$). In conclusion, the polymorphisms in the MHC loci could have an important role in the immune mechanisms against H. contortus infections in sheep and goats. Also, Spanish goats and St. Croix sheep had the lowest mean fecal egg counts.

Genomics of Resilience in Sheep to Climatic Stressors
The study aims to identify differences among sheep breeds, ecotypes and individuals within ecotypes in resilience to climate mediated stress, and characterize the genetic background of this difference using landscape genetics, population genetics, and genomics approaches. The information obtained will be useful to identify genetic markers that can be used in breeding programs that take into account environmental resilience in addition to production capacity.
The following progress have been made so far:
- Sheep samples were collected from different eco-climate domains of the US. These included Dorper, Katahdin and St. Croix breeds of sheep from Midwest, Northeast Southwest and Southern parts of the US. All three breeds were sampled from all four parts of the US. A total of 115 ewes were sampled randomly from all locations and breeds, and 71 old ewes were sampled from all locations and breeds.
- Phenotyping is currently being carried out at the American Institute of Goat Research, Langston University for three resilience traits, namely: minimization of energy

use with a limited nutritional plane; water conservation with restricted availability and heat tolerance with high heat load index.

- DNA was extracted from ear punch samples after which genotyping was carried out using Illumina 50k BeadChip.

Objective 3

A sheep genomes database, created by the ISGC using funds from a second 2013 NIFA/AFRI grant (N. Cockett, PD), currently contains whole genome sequences from 455 sheep collected from around the world and aligned to Oar v3.1. Analysis of 250 of the sequences revealed over 80 million SNPs and indels with high confidence using two variant-calling platforms. Data in the database is publicly available via European Variation Archive (EVA), which delivers key advantages concerning public access of genome information, data storage and variant accessioning. Variants are available as i) raw (unfiltered) or ii) filtered following application of a comprehensive QC protocol. This resource, referred to as the SheepGenomes DB, will speed discovery and innovation for scientists working in the area of livestock genomics. Another 950 genomes will be analysed in Run 2 and incorporated into the database in 2017.

IMPACT / USEFULNESS OF FINDINGS:

The Ovine FAANG project will provide the enhanced functional annotation of the ovine genome in order to enable research in this sheep. Furthermore, this project will facilitate the understanding of gene regulation in sheep and other livestock species by generating and distributing curated transcriptome data.

Although only a small number of genetic markers for economically important traits in sheep are currently available, the identification of genetic regions associated with traits is a first step in identifying useable markers. Whole genome sequences from animals with interesting traits and phenotypes compared back to the ovine reference sequences in the Sheep Genomes Database will significantly accelerate searches for genetic regions and genes influencing phenotypes in sheep. Annotation of the ovine reference sequence with transcriptome data and regulatory features through the Ovine FAANG Project will inform our understanding of biological processes underlying a phenotype and an estimate of the probability of a particular variation in the genome sequence affecting the phenotype of interest. This increased understanding of biological processes can be used to improve the management of the animals so they can to reach their genetic potential. In addition, understanding the biological processes underlying the phenotype may enable us to define phenotypes better, thereby reducing complex phenotypes to a series of simple phenotypes based on different biological processes. The better definition of phenotypes will in turn lead to better prediction of animal performance.

Global scrapie eradication programs use trace-back data on positive animals to identify additional possible exposed animals. The addition of long incubation time data for S146 and K222 heterozygotes suggests longer trace-back histories for goats with this relatively common genotype to avoid missing sources of scrapie exposure. Identification of a

divergent MYADM-like repeat associated with mean corpuscular hemoglobin in sheep may be of interest in other species with erythrocyte morphology disorders. Identification of phenotype associations for differentially selected regions among common U.S. wool breeds identifies potential targets for additional research and marker assisted selection.

Selection of lambs based on maternal (and paternal) resistance to GIN infection would reduce costs of control and improve productivity. In addition, selecting periparturient ewes for resistance has the potential to reduce pasture contamination that their lambs would be exposed to.

The characterization of WC1 genes in ruminants and swine has utility in several areas: improved vaccine design against multiple important pathogens targeted to recruit gamma delta T cells, potential selective breeding based on SNPs in WC1 genes linked to resistance to pathogens, and the possible involvement of WC1 as a receptor in PRRSV infection of swine. Because WC1 genes are present in most mammalian and avian species, elucidation of WC1 genes in agriculturally important ruminants and swine has global application potential in improving food security and in containing zoonotic diseases in non-human animal reservoirs.

Animal breeders in the dairy industry were supervised by our findings that only two ancestral Y chromosome survived in today's Holstein population. Our work suggests minimal genetic diversity on the Y chromosome in Holsteins, and provides a start point for investigating the impact of the extremely limited number of Y-lineages on male reproduction and other traits important for the future of the Holstein breed.
The custom-made 384-SNP chip has been confirmed to be valuable for bull semen quality and fertility evaluation in dairy cattle. We are developing a smaller chip by reducing the SNP markers from the 384-SNP chip in hope to commercialize this technology for bull fertility selection in the future.

Our work on the PRAME/PRAMEY, a cancer-testis antigen (CTA), provides insights into the molecular mechanism underlying the functions of PRAME/PRAMEY during spermiogenesis, which will advance the field by adding understanding of the complexity (i.e. functions of CTAs) of the development of male germ cells. Disclosure of the role of PRAME/PRAMEY during male gametogenesis may provide insights into the molecular mechanisms shared between gametogenesis and tumorigenesis.

The meiotic recombination research provides important information regarding recombination differences in sheep spermatocyte and a greater understanding of the genes that control genetic variation to enhance reproduction, improve genetic predictions, and advance selection strategies towards the sustainability of the sheep industry.

Total Leveraged Funding Summary for 2014-2016

Federal Funding:	$ 3,402,000
State/Local/Institutional:	$ 22,000
Industry:	$ 0

Total $ 3,424,000

SWINE TECHNICAL REPORT

OVERVIEW: Coordination of Pig Genome Coordination Program is under the National Animal Genome Research Program (NAGRP) and is the effort of personnel at Iowa State University (ISU) and Michigan State University (MSU).
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 Swine Species Subcommittee.
Facilities and personnel: Chris Tuggle, Department of Animal Science, ISU, and Cathy Ernst, Department of Animal Science, MSU, have served as Joint Coordinators since 201 3 and have a five - year appointment. Iowa State University staff help support the national pig genome coordination effort as part of Iowa State University's contribution.

Policy Updates: We have developed an Advisory Committee, who will provide guidance on policy as well as help evaluate requests for funding. The members of this Advisory Committee represent the swine industry, swine genomics and biotechnology researchers, NRSP-8 Stations and participating USDA labs. The members are: Jack Dekkers (ISU), Chris Hostetler (National Pork Board), Joan Lunney (USDA-BARC), Randy Prather (U. Missouri), and Juan P. Steibel (MSU). Thanks to this group for volunteering for this important role!

Database Activities: The Pig Genome Database continues to receive considerable updating through the work of the Bioinformatics team. The PigQTLdb (http://www.animalgenome.org/QTLdb/pig) is an excellent repository for QTL and candidate gene association results. As of January 4, 2016, in the Animal QTLdb there are 16,516 pig QTLs from 566 publications curated into the database, a 14 % increase over the end of 2014. Those QTLs represent 626 different traits. Throughout 2016, the NAGRP bioinformatics team has continued their efforts to make improvements to the Animal QTLdb. Users are encouraged to register an account to enter new QTL data. Find out more from http://www.animalgenome.org/QTLdb. In addition, the pig genome build 10.2 annotations are continuing to be updated in the BioMart (http://www.animalgenome.org:8181) and for the Animal QTLdb.
Shared Materials and Funding: NRSP8 funds are available to support community activities to find associations with many different traits. In 2014, a policy was developed and approved by the Advisory Committee that for swine genomics projects to be eligible for NRSP8 Coordination support, the project must materially involve two or more NRSP8 member groups (university or ARS research locations) and that substantial funding will only be provided for projects that have matching funding from another agency. Any questions on this policy, please contact the Coordinators.

Porcine SNP chips update: In addition to the 60K Illumina and the GeneSeek GGP-Porcine LD and HD chips, a new high density SNP chip is being developed by Affymetrix, and was announced in 2015. As described above, an NRSP-8 supported project will provide validation of this chip for integration with 60K and GeneSeek

chip data.

National and International Efforts: Communication with several national and international groups and individuals is excellent. Several international meetings were organized and/or held in 2016 that had a national or international component.
1. At the 2016 Joint Annual Meeting (JAM) of the ASAS, ADSA, CSAS and other organizations a day of programmatic overlap with the 2016 International Society of Animal Genetics (ISAG) meeting was scheduled for July 23, 2016 in Salt Lake City, Utah. Several FAANG members including Chris Tuggle and Stephen White (NRSP-8 Sheep Coordinator) helped develop a Symposium on FAANG for this day. Talks were presented by speakers who are experienced leaders and scientists in the ENCODE project as well as newly emerging researchers in systems biology/functional genomics in livestock species. 450-500 persons attended, and the program is provided at the ISAG website (direct link: http://www.isag.us/Docs/DomAnimSequencing2016.pdf).
2. The ASAS meeting was attended by Dr. Ernst and both she and Dr. Tuggle attended the following ISAG meeting. Short meeting reports were provided in the October 2016 Pig Genome Update.

Communication: The Pig Genome Update has now published 124 issues and has been distributed electronically to over 2,800 people worldwide. PGU will be electronically published three times a year, and in addition to general updates, the issues will be published to coincide with major events of interest to the genome community:
Feb (a wrap-up report of the PAG meetings);
June (summer meetings reminders); and
October (summer meetings report, PAG abstract submission deadlines, preparations).

Travel and Meeting Support: Travel of several scientists was partially funded to attend important pig genomics meetings in the reporting period. These included:
Jeremy Howard, North Carolina State University, 2016 Neal Jorgenson Travel Award winner
We also partially supported the travel of speakers to the 2016 Cattle/Swine joint and Swine Workshops: Randy Prather, University of Missouri, and Bruce Whitelaw, Roslin Institute, Cattle/Swine Workshop Min-Kyeung Choi, Konkuk University, and Francisco Peñagaricano, University of Florida, Swine Workshop.

2017 commitments:
Haibo Liu, Iowa State University, 2017 Neal Jorgenson Travel Award winner
Vanmathy Kasimanickam, Washington State University, Pullman, WA
Alan Archibald, Roslin Institute, Edinburgh, UK

Research Support Activities: The goals are to help support all of the objectives of this project. Major activities included helping facilitate collection of phenotypes and sharing use of SNP chips in the future. New bioinformatic tools relevant to the swine genomics community will also be developed with help of the bioinformatics team. Constructive suggestions from researchers to help this coordination and facilitation program grow and

succeed are appreciated. **Reminder: funding is available for new projects- preliminary ideas can be a starting point and are welcome- please contact the Coordinators!

Prior approved Projects:
1. FAANG project led by Huaijun Zhou, University of California-Davis. This project also had funding promised by the NRSP8 Bovine and Poultry Coordinators, as well as funding by the National Pork Board.
2. PEDV genetics resistance project led by Max Rothschild with collaborators Daniel Ciobanu and Canadian swine genetics companies.
3. A proposal submitted by Jack Dekkers along with Cathy Ernst and Juan P. Steibel (MSU) to validate the new Affymetrix 650K chip and provide initial data on integration with 60K genotype data.
4. A proposal by Tim Smith and Dan Nonneman of USDA-MARC along with Chris Tuggle to add additional tissues to a PacBio IsoSeq project for functional annotation of the genome of the animal whose genome is being sequenced at MARC. It is important to note that for projects #3 and #4, the Swine Genome Coordinators had a co-PI role, so the proposals were vetted through the Advisory Committee for approval.

Newly approved projects during reporting period:
1. To best make use of available NRSP-8 funding in 2016-2017, a part-time postdoctoral position was proposed to be supported, to be housed at Iowa State University. This postdoc will be tasked with bioinformatic analysis of the PacBio Isoseq/Illumina RNAseq data being created and analyzed in the Smith/Nonneman/Tuggle project above, to maximize the use of these data across nine tissues to annotate the imminent new TJ Tabasco assembly. Additional projects on analysis of data from multi-station projects is projected. The Advisory Committee approved this use of funds, and a search for a qualified person to fill this position is being made.

Annual meeting workshop:
2016 Chair (for 2017 Workshop): Kiho Lee (Virginia Tech University); kiholee@vt.edu
2016 Chair elect (for 2018 Workshop): Chris Tuggle (Iowa State University); cktuggle@iastate.edu

The 2017 NRSP-8 Swine Workshop was held January 14, 2017 in San Diego, CA in conjunction with the Plant and Animal Genome XXV Conference. A joint session was held with the Cattle, Sheep and Goat Workshop in the morning focused on correlating miRNA profile of boar sperm and their potential traits, and whole genome shot-gun sequencing approach. The afternoon Swine Workshop program started with an invited presentation by a young scientists from University in South Africa, introducing their work on identifying the source of Colibacillosis in Neonatal and Weaning Piglets in South Africa using genome sequencing. Then Alex Clop from Spain introduced their recent study on identifying genetic regions influencing semen quality traits. The Jorgensen Pig Travel Award winner, Haibo Liu from Iowa State University, was introduced and he gave a lightening talk on his area of research. There were also nine presentations from seven different NRSP-8 participating locations. The presentations covered a range of topics from gene-editing approach, functional genomics,

transcriptome and microarray analysis, infectious diseases, as well as a broad range of important phenotypes, and sparked discussion among attendees throughout the workshop. Drs. Chitnis and Matukumalli gave administrator's reports, and Drs. Tuggle and Ernst gave their coordinators' report, as well as conducted a discussion on community needs and resources. At the morning joint session, 103 attendees signed in, including attendees from 11 different countries outside the US. In the afternoon, the Swine Workshop had 58 people sign in, although it is estimated that more people were present for session. Among those signing in, there were 40 institutions and 8 private companies. During the business meeting, Dr. Christian Maltecca from North Carolina State University was elected as the new chair-elect, and Dr. Chris Tuggle from Iowa State University will chair the 2018 Swine Workshop.

Impacts 2016

1. Presentation from Dr. Alan Archibald indicates that improved swine genome assembly will be available in the near future.

BIOINFORMATICS TECHNICAL REPORT

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.

ROGRESS 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.

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 (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 'ss' 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.