**ANNUAL SUMMARY REPORT**

**PROJECT:** NRSP-8

**PROJECT TITLE:** National Animal Genome Research Project

**PERIOD COVERED:** January 1 to December 31, 2015

**ANNUAL MEETING DATE:** January 9-10, 2016

**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 175 participants from 14 countries attend the workshop.

The combined workshop included four plenary presentations as follows:

Timothy P.L. Smith from US-MARC presented “Approaches Taken, Progress Made, and Enhanced Utility of Long Read-based Goat, Swine, Cattle and Sheep Reference Genomes”. Dr. Carole Charlier from University of Liege presented “The Role of Mobile Genetic Elements in the Bovine Genome”. Dr. Susan J. Lamont, Iowa State University delivered the NRSP8 Distinguished Lecture entitled “Genomics of Response to Environmental Challenges of Poultry”. The last plenary lecture was presented by Terje Raudsepp, Texas A&M University with the title “Comparative Studies of Mammalian Sex Chromosomes: From Cytogenetics to NGS”. Following the plenary presentation Dr.

Christopher K. Tuggle, Iowa State University gave an update of the “The Functional Annotation of Animal Genomes (FAANG) Initiative

The business meeting was called to order by the Chair, Dr. Daniel Ciobanu (University of Nebraska), and was recorded by the Secretary, Huaijun Zhou (University of California-Davis) with approximately 40 members in attendance. Coordinator reports were presented for the species/topic groups of Cattle, Poultry, Swine, Sheep and Small Ruminants, Equine, Aquaculture, and Bioinformatics. Dr. Eric Young (North Carolina State University) provided the administrative report. Dr. Lakshmi Matukumalli provided a brief update; AFRI call will be available on February 2016 on “Food security and foundational programs”. It was stated and confirmed that the 2016 NRSP8 meeting will again be held in conjunction with the Plant and Animal Genome conference in San Diego. Dr. Huaijun Zhou (University of California-Davis) assumed the NRSP-8 Chair for 2016-2017, and Dr. Mohamed Salem from Middle Tennessee State University was nominated for the Secretary in 2017 and Chair in 2018.

**ACCOMPLISHMENTS AND IMPACTS:**

**AQUACULTURE TECHNICAL REPORT**

Coordinator: Dr. John Liu

Co-Coordinator: Dr. Caird Rexroad

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

**2015 Progress towards NRSP8 Objectives**

Many species lack sufficient data to inform the development of strategies that will provide reliable sequence assemblies. In addition to supporting continued refinement of whole genome reference sequences and transcriptomes for the traditional US aquaculture species of catfish, trout, and oysters, NRSP8 Coordinators Funds were used to generate preliminary data for sablefish (*Anoplopoma fimbria*), blue tilapia , and Pacific whiteleg shrimp (*Litopenaeus vannamei*). These data provide insights into the complexity of these genomes and transcrptomes, and enable better optimization when developing strategies that target assembly of more comprehensive data sets.

*Objective 1*

***Catfish***

The channel catfish reference genome sequence was assembled with a N50 contig size of 77.2 Kb and a scaffold N50 of 7.7 Mb. A total of 761 Mb of the genomic sequence has been anchored to the 29 chromosomes through linkage analysis using high density SNP arrays. The reference genome sequence was validated by genetic mapping of over 54,000 SNPs, and annotated with 26,661 predicted protein-coding genes.

***Oyster***

NRSP8 Coordinator funds were used to support resource coordination workshops focused on oysters and other shellfish. Sequencing of the eastern oyster (*Crassostrea virginica*) genome (using a single gynogen oyster) has been initiated and a draft genome of Pearl oyster (*Pinctada fucata*) is in progress (NCBI BioProject ID PRJDB2628). A second generation linkage map, based on >1100 type I SNP markers and 66 microsatellite markers was constructed for *Crassostrea gigas*. This higher density linkage map reveals errors in the original genome assembly.

***Salmonids***

A. NRSP8 Coordinator funds were used to support the development of a genome reference sequence and the identification of genetic variation associated with economically important traits. A new and improved reference genome assembly was generated for rainbow trout coupled with a 50K SNP linkage map. The total size of the new assembly is nearly 2.2Gb with N50=1.7Mb. Approximately 82% of the new assembly has been anchored and ordered onto the new chromosome linkage maps.

B. A new de novo transcriptome with gene annotation and tissue gene expression atlas was assembled for rainbow trout (Salem et al., 2015).

***Striped bass***

NRSP8 Coordinator’s funds were used to support genome reference sequencing for striped bass. A 585.1 Mbp striped bass genome sequence assembly containing ~35 K scaffolds was produced from Illumina short-read sequence (66-fold genome coverage) and Pacific Biosciences single molecule, long-read sequence (2.8-fold genome coverage). *Ab-initio* and evidence-based gene predictions performed using the MAKER Annotation Pipeline identified 27,485 protein-coding genes. An jBrowse website providing access to the annotated genome was made available online at N.C. State University at https://appliedecology.cals.ncsu.edu/striped-bass-genome-project/.

*Objective 2*

***Catfish***

Columnaris causes severe mortalities among many different wild and cultured freshwater fish species, but understanding of host resistance is lacking. We have used the interspecific hybrid backcross populations and mapped the resistance genes. To identify genes associated with columnaris resistance, we performed a genome-wide association study (GWAS) using the catfish 250k SNP array with 340 backcross progenies. A genomic region on chromosome 7 was found to be significantly associated with columnaris resistance. Within this region, five have known functions in immunity, including pik3r3b, cyld-like, adcyap1r1, adcyap1r1-like, and mast2. In addition, 3 additional suggestively associated QTL regions were identified on chromosomes 7, 12, and 14. The resistant genotypes on the QTLs of chromosomes 7 and 12 were found to be homozygous with both alleles being derived from channel catfish. The paralogs of the candidate genes in the suggestively associated QTL of chromosome 12 were found on the QTLs of chromosome 7. Many candidate genes on the four associated regions are involved in PI3K pathway that is known to be required by many bacteria for efficient entry into the host. Strikingly, the candidate genes may be arranged as functional hubs; the candidate genes within the associated QTLs on chromosomes 7 and 12 are not only co-localized, but also functionally related, with many of them being involved in the PI3K signal transduction pathway, suggesting its importance for columnaris resistance.

***Oyster***

A. The fully sequenced *C. gigas* genome has led to the intense study of several expanded gene families associated with innate immunity and stress response. These families include: C1q domain-containing proteins (which act as pathogen recognition molecules), Fibrinogen-related proteins (FREPs; which exhibit versatile immune functions), and the Tumor Necrosis Factor (TNF) superfamily.

B. RNAseq analysis was performed using an oyster sample with a high viral load to elucidate interactions between host (*C. gigas*) and pathogen (OsHV-1; which causes significant mortality in juvenile oysters worldwide). This research resulted in a high quality OsHV-1 transcriptome and identified several molecular pathways in *C. gigas* that are activated in the presence of OsHV-1.

C. 295 SNPs were identified and validated in 90 genes involved with glycogen content in *C. gigas*. Statistically significant associations between genotype and glycogen content were detected for three SNP markers.

D. The heritability of DNA methylation variation was investigated in diploid and triploid *C. gigas*. Genome-wide methylation patterns did not differ between diploid and triploid oysters. Transmission of methylation status between parents and offspring was largely stable; however at some loci, methylation was observed more frequently in offspring.

***Salmonids***

A. NRSP8 Coordinators funds were used to demonstrate usefulness of high-density genotyping arrays for identifying disease resistance genes in rainbow trout.

B. An evaluation of genome-enabled selection strategies for bacterial disease resistance in a commercial rainbow trout population has demonstrated the advantage of genome selection (GS) over traditional BLUP-based breeding values. The predictive ability of offspring performance using different GS models has more than doubled that of the traditional pedigree-based method, translating to much more rapid gains in disease resistance with the potential of increased profitability for the trout aquaculture industry and large reduction of antibiotic use for trout farming.

C. The 57K SNP chip was used in genome wide association studies (GWAS) for bacterial cold water disease resistance in an experimental and commercial rainbow trout populations, which identified 14 loci with moderate to strong effects that are shared by the two populations. Similarly, two loci with moderate effect were identified in a GWAS for fillet yield in rainbow trout.

D. Allelic-imbalance analysis of pooled RNA-Seq samples was used to identify genetic markers that may be associated with muscle development in rainbow trout.

E. A new study identified differentially expressed long non-coding RNAs in response to F. psychrophilum infection in rainbow trout.

F. RNA-Seq analysis of miRNAs in myosatellite cells exposed to estrogen at different levels (biological or high) revealed dose dependent miRNA expression profiles.

G. Bisulfite sequencing of the regulatory region of MyoD gene revealed epigenetic regulation of the gene (methylation of a CpG island) under the influence of estrogen. This epigenetic regulation possibly resulted in a decreased expression of the gene in estrogen treated-muscles.

H. RNA-Seq analysis of miRNAs from eggs of different qualities (assessed by fertilization rate) identified egg quality-associated miRNAs including 4 known miRNAs (omy-miR-193b-3p, omy-miR-203c-3p, omy-miR-499-5p and omy-miR-7550-3p) and two novel miRNAs (omy-miR-nov-95-5p and omy-miR-nov-112-5p).

***Striped bass***

Artificial neural networks and supervised machine learning were used to further evaluate relationships between ovary gene expression (transcriptome) profiles and egg quality (fertility) in striped bass. Findings compared to those obtained for other farmed fishes using conventional analytical methods were published in an invited review. Advanced proteomics and targeted transcript studies revealed heretofore unknown molecular mechanisms of egg yolk protein and lipid formation in striped bass, white perch, and other teleosts.

*Objective 3*

***Oyster***

A. Resource coordination workshop focused on oysters and other shellfish held in conjunction with the National Shellfish Association annual meeting in Monterey, CA, March 22-26, 2015. Organizer: Steven Roberts.

*B. C. gigas* transcriptome information derived from 2.2 billion sequences from 114 RNAseq datasets has been organized and deposited into a publicly available database: GigaTON. The user interface provides powerful and user-friendly tools to search and retrieve annotation,

**MEETINGS:**

The Aquaculture Genome Workshop organized by Chair Mohamed Salem was held Saturday January 9th, 2016 in Dan Diego, CA. There were 84 attendees representing 18 Countries (US, Canada, Mexico, Norway, Australia, China, Netherlands, Thailand, Malaysia, Germany, Chile, UK, France, New Zeeland, Turkey, Columbia, Taiwan, Japan) and 53 institutes. The workshop included four invited presentations, sixteen contributed presentations, and 42 poster presentations.

Invited presentations included:

1. Editing Fish Genome with CRISPR

**Wenbiao Chen**, *Vanderbilt University School of Medicine*

1. The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates

**Yann Guiguen**, *INRA-SCRIBE*

1. Genomics in Fish Breeding Programs for Developing Countries

**John A.H. Benzie**, *WorldFish*

1. Regulatory Approval of Genetically Engineered AquAdvantage Salmon

**John Buchanan**, *Center for Aquaculture Technologies*

Contributed presentations included four travel award recipients **highlighted in blue**:

1. Functional Studies in Atlantic Salmon (*Salmo salar L.*) Reveals Candidates for Sterility Vaccines

**Anna Troedsson-Wargelius**, *Institute of Marine Research*

1. Genomic Selection For Bacterial Cold Water Disease Resistance Reveals Large Within-Family Variation That Cannot Be Exploited In Traditional Family-based Selective Breeding In Rainbow Trout

**Roger L. Vallejo**, *USDA-ARS-NCCCWA*

1. Weighted ssGBLUP Improves Genomic Selection Accuracy for Survival in a Rainbow Trout Population

**Breno O. Fragomeni**, *University of Georgia*

1. Genome Scan for Selection Signatures in Atlantic Salmon Populations Using a High Density SNP Array

**María Eugenia López Dinamarca**, *University of Chile*

1. The vgll3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (*Salmo salar L.*) Males

**Fernando Ayllon**, *Institute of Marine Research*

1. Genome-Wide Association Study for Identifying Genome Loci That Affect Fillet Yield in Rainbow Trout (*Oncorhynchus mykiss*)

**Dianelys Gonzalez-Pena**, *USDA-ARS-NCCCWA*

1. A Genome-Wide Association Study for Low Oxygen Tolerance in Catfish using the 250K SNP Array

**Xiaozhu Wang**, *Auburn University*

1. Candidate genes for ESC Disease Resistance of Catfish as Revealed by a Genome Wide Association Study

**Tao Zhou**, *Auburn University*

1. A New and Improved Rainbow Trout (*Oncorhynchus mykiss*) Reference Genome Assembly

**Guangtu Gao**, *USDA-ARS-NCCCWA*

1. Progress of the Shrimp Genomic Sequencing Project

**Jianhai Xiang**, *Institute of Oceanology, Chinese Academy of Sciences*

1. Comparative Transcriptome Analysis of the Swimbladder Reveals Expression Signatures in Response to Hypoxia in Channel Catfish, Ictalurus punctatus

**Qiang Fu**, *Auburn University*

1. The Catfish MicroRNAome: Identification, Annotation and Expression Profiling in Response to Bacterial Infections and Hypoxia Stress

**Shikai Liu**, *Auburn University*

1. Allelic-Imbalance Analysis in Pooled RNA-Seq Samples Identifies Muscle-Associated Genetic Markers in Rainbow Trout: Improved Bioinformatics Practices

**Rafet Al-Tobasei**, *Middle Tennessee State University*

1. Role of Long Non-Coding RNAs in Bacterial Cold Water Disease Pathogenesis in Rainbow Trout

**Bam D Paneru**, *Middle Tennessee State University*

1. Genotyping in Thousands By Sequencing (GT-seq): A Low Cost, High-Throughput, Targeted SNP Genotyping Method

**Nathan Campbell**, *Columbia River Inter-Tribal Fish Commission*

1. Development of the Catfish 690K SNP Arrays for Analysis of Quantitative Traits

**Qifan Zeng, Qiang Fu, Shikai Liu and Yun Li**, *Auburn University*

NRSP8 Coordinator Funds also supported travel for Hsinyuan Tsai from the Roslin Institute who presented “Genomic prediction of host resistance to sea lice (L. salmonis) in Atlantic salmon (S. salar).”

The business meeting was held immediately following the conclusion of the workshop. New officers were nominated and approved through a voice vote as follows:

2017 Workshop Chair – Mr. Nathan Campbell

2017 Workshop Chair Elect – Dr. Geoff Waldbieser

2017 NRSP8 Chair Elect – Dr. Mohamed Salem

Dr. Steven Roberts from the University of Washington was selected to replace outgoing Co-coordinator Dr. Caird Rexroad.

**CATTLE TECHNICAL REPORT**

**A. BRIEF SUMMARY OF MINUTES OF ANNUAL MEETING:**

The 2015 annual meeting of the NRSP-8 Cattle, Sheep, and Goat committee was held in conjunction with the Plant and Animal Genome XXIV meeting on January 9-13, 2016. The morning session of the scientific meeting on Jan 9th was held as a joint session in with the Swine Committee with 4 presentations centering around a theme of gene editing. This session was attended by approximately 170 people of which approximately 46% were international attendees, representing 21 different countries. The afternoon session of the scientific program, Cattle/Sheep/Goat Workshop-1, consisted of 7 presentations followed by 15 station reports, 2 coordinator reports (Dr. N. Cockett –Sheep and Dr. J.F. Medrano –Cattle) and an update from USDA-NIFA by Dr. Lakshmi Kumar. The morning session on Jan 10th, Cattle/Sheep/Goat Workshop-2, consisted of 12 presentations. The presentations were excellent and covered a wide variety of topics from comparative genomics, low budget reference grade genome assembly, feed efficiency and methane emissions, genotype by sequencing, genome annotation, omics data resource, comparative proteomics, recombination mapping, copy number variation, genomic selection, male fertility, and an overview of genome Canada funded research. Attendance at the sessions was exceptional with approximately 200 people attending each one of the scientific sessions. Out of these attendees, ~ 44% were international attendees from at least 29 countries, including Australia, Belgium, Brazil, Canada, Chile, China, Denmark, Egypt, Ethiopia, France, Germany, Ireland, Italy, Japan, Kenya, Mexico, Netherlands, New Zealand, Norway, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Uganda, UK, Uruguay. Approximately 25 people attended the station reports session. Stephanie McKay was thanked for serving as President of the NRSP-8 Cattle, Sheep and Goat Committee in 2015-16. Jared Decker will serve as President in 2016-17. Brenda Murdoch was elected as the 2017-2018 Secretary, and she will serve as President in 2017-2018.

**PARTICIPANTS:**

Colorado State University: Milton Thomas\*

University of Wisconsin: Brian Kirpatrick\*

University of Missouri: Jared Decker\*

Louisiana State University: James E. Miller\*

University of Florida: Raluca Mateescu\*

Oklahoma State University: Udaya DeSilva\*

 Pennsylvania State University: Wansheng Liu\*

 Purdue University: Christopher A. Bidwell\*

 Texas A&M University: Clare Gill\*, Penny Riggs\*

University of California Davis: Juan F. Medrano\*, Alison Van Eenennaam\*

University of Wyoming: Kristi Cammack\*

 USDA/ARS: Michelle R. Mousel\*, Stephen N. White\*

 Utah State University: Noelle E. Cockett\*, Thomas D. Bunch\*

Virginia State University: Brian Sayre\*

Washington State University: Holly Neibergs\*

 University of Vermont: Stephanie McKay**\***

 University of Massachusetts: Cynthia L. Baldwin & Janice C. Telfer

\*In attendance at the 2015 NRSP-8 meeting.

**B. ACCOMPLISHMENTS AND IMPACTS:**

**Objective 1:**

An important focus of the community has been towards improving the bovine genome assembly. Following important landmark accomplishments in 2015 a new bovine reference assembly will be developed in 2016.

David Schwartz, Shigou Zhou(University of Wisconsin, Madison) and collaborators generated a whole genome optical map (BtOM1.0) of Dominette 01449. Optical mappingis a high throughput approach for producing ordered, genome-wide, high-resolution [restriction maps](https://en.wikipedia.org/wiki/Restriction_map) from single molecules of DNA. The map is produced by stretching single DNA molecules in a microfluidic system, cutting the single DNA molecules with a restriction enzyme and by recording an image of the size of the restriction fragments using an epifluorescence microscope. The spectrum of resulting DNA fragments serve as a unique "fingerprint" for that sequence and it is a very useful system to independently validate the accuracy of genome assemblies. The method was developed by David Schwartz in the 90s. In order to validate the two current bovine genome assemblies (UMD3.1 and Btau4.6) a high resolution BamHI optical map (BtOM1.0) of the reference animal, Dominette, was developed. The OM spans 2,575.30 Mb and comprises 78 optical contigs, featuring an average restriction fragment size of 8.91 Kb. Comparisons between BtOM1.0 and UMD3.1, or Btau4.6, revealed that Btau4.6 presented far more discordances (7,463) than UMD3.1 (4,754). Overall, Btau4.6 presented almost double the number of discordances than UMD3.1 representing misassembles, extraneous sequences, missing sequences and inverted/translocated sequences. The OM map is valuable structural resource of Dominette and is currently being used for ordering contigs and scaffolds for the new PacBio de-novo bovine assembly. Coordinator funds were used to develop the first iteration of the Optical Map.

Tim Smith (Meat Animal Research Center, Nebraska) and Juan Medrano (University of California, Davis) have generated approximately 83X PacBio coverage of Dominette with funding from NRSP-8 Coordinator Funds, USDA/MARC, UC Davis and funds leveraged from donations from Zoetis and Neogen/GeneSeek. A PacBio *de-novo* assembly is currently being generated, followed by scaffolding of contigs using the recently generated Optical Map and Dovetail Genomics' Chicago library/HiRise scaffolding approach. The assembly is being produced by Computomics and will be completed in early 2016. The PacBio data were shared with Aleksey Zimin at the University of Maryland who is supported by USDA NIFA funds to generate a hybrid Pac-Bio/Sanger and independent Pac-Bio *de-novo* assemblies. Ben Rosen will work on the initial comparison of the Hybrid and De-Novo Assemblies. Aleksey Zimin has also agreed to undertake the task of merging the De-Novo and Hybrid assemblies to end up with a single high quality Dominette reference genome assembly. Annotation of the assembly will follow.

In addition, Kim Worley at Baylor, who is also supported by USDA/NIFA, has produced a PacBio+PBJelly Dominette assembly that has been submitted to GenBank and data are available in SRA. These will be incorporated into the *de-novo* assembly.

Chris Elsik(University of Missouri) has developed unique bioinformatics approaches for bovine genome annotation using RNA-seq and PacBio Iso-Seq single-molecule transcript sequencing data of 7 tissues generated at MARC. As part of the Dominette sequencing initiative the Iso-Seq data set will be expanded to at least 20 tissues and these data will be provided to Chris Elsik for genome annotation. Using Iso-seq data allows assembling full length transcripts, identification of isoforms and updating new and existing gene sets.

The FAANG initiative (Functional Annotation of Animal Genomes) is following the blueprint of the human and mouse ENCODE projects for identifying the functional roles of regulatory elements in the genome. This group has implemented a similar effort in cattle, pig and chicken, initiating the AGENCODE project. The goals are to identify promoter, enhancer, and silencer region specific chromatin marks, and to determine functional roles of regulatory regions in relevant tissues in each species. In cattle, the effort supported by both USDA NIFA and Bovine Genome Coordination Fund has collected a variety of tissues from 4 Line 1 Herefords (2 males and 2 females) for assays of RNA-seq, ChIP-seq, and DNase-seq. The data will be integrated to functionally annotate the bovine genome. Huaijun Zhou (UC Davis) is the PI on this project.

Jerry Taylor and collaborators (University of Missouri), in collaboration with GeneSeek, have developed an Illumina 250K Bead Infinium functional variant assay (GGP F250). This assay was developed with USDA NIFA funding which supported whole genome sequencing for variant discovery and the purchasing of a sufficient numbers of chips to support the design. The assay was designed using sequence data on over 400 individuals from multiple taurine breeds and sequence data from the 1000 Bull Genomes Project. The chip is focused on the detection of genic variants likely to be functional in taurine cattle. The assay will first be used to genotype samples from the USDA NIFA supported “Bovine Respiratory Disease CAP,” “Feed Efficiency” and “Heifer Fertility” Projects. The assay will be made available to the public through GeneSeek.

Clare Gill’s group (Texas A&M University), has identified copy number variants from genome sequences of the founders of a Nellore-Angus cross population and further characterized a locus affecting coat color reddening in cattle. Womack's group advanced the reference genomes for both buffalo and goats through continued use of radiation hybrid maps to facilitate comparative mapping in buffalo and the goat genome assembly. Womack's group (Texas A&M University) also recently demonstrated that there are 4 genes for NK-Lysin in cattle as opposed to 2 genes as indicated in currently annotated assemblies (Chen et al, PNAS, in press).

**Objective 2:**

Zhihua Jiang and Holly Neibergs (Washington State University) have performed Genome-wide analyses of 1000 Holsteins (for fertility traits) and 5100 beef cattle (2000 for susceptibility to BRD, 100 for fertility traits, 3000 for feed efficiency). Almost all of these analyses were conducted with the bovine HD BeadChip. Fine mapping to include NGS of 30 Holsteins.

Functional genomic study of high altitude disease (HAD) and its impact on beef production systems (J. Rouse Endowment and CSU Beef Improvement Center) is being performed by Milt Thomas (Colorado State University), A. Canovas and J.F. Medrano (U.C. Davis) and others. OBJECTIVE: Use functional genomic approaches to identify SNP/variants, which predict measures of disease and (or) performance traits in Angus cattle. EXPERIMENTAL APPROACH AND PROCEDURES**:** Replicate and expand procedures developed in AES-Hatch, USDA-AFRI, and McMasters Fellowship (multi-omics approach). Currently genotyped (BovineSNP50/HD) 3,000 cattle with pulmonary arterial pressure (PAP) measures and other measures of hyptertension. Collected 45 tissues for RNA-Seq (High (healthy vs sick) vs Low PAP steers). Tissues are primarily pulmonary and cardiac tissues.

Specifically, High-altitude (>1800m) disease is a challenging problem for beef cattle. The disease is a consequence of hypoxia-induced right ventricular (RV) heart failure as per vascular inflammation of the pulmonary artery (PA) and hypertension. An indicator trait of this condition, pulmonary arterial pressure (PAP) has moderate to high heritability (0.2-0.4) when measured in yearling cattle; however, knowledge of candidate genes that can be used in genetic improvement for this trait and physiological-disease condition are minimal. The transcriptomes of RV, LV, PA, aorta, muscle, and lung were examined from fattening-yearling Angus steers phenotyped to be of low or high PAP (LPAP and HPAP; n=10/group). RNA-Seq data and analyses revealed the highest number of differentially expressed genes (DEG) between groups were in RV (n=1394) and aorta (n=1173; p<0.01;Fold-change>2). Pathway analyses of DEG in RV suggested importance of IL-8/IL-10 signaling, factors promoting cardiogenesis, coagulation, thrombin and cardiac hypertrophy. Systems biology analyses of RV data suggested that 101 genes were acting as key transcriptional regulators of 705 DEG between LPAP and HPAP steers. Most of the key regulators had roles in cardiomyopathy (NFATC1), movement of leukocytes/neutrophils (OLR1,PLAUR), failure of heart (CTGF), ventricle hypertrophy (TREM1,GATA2,P38-MAPK) and vascularization (SYVN1). Several SNP variants segregated specifically in either the LPAP or HPAP animals in key regulator genes. The integration of structural and functional genomic data associated with high-altitude disease will help to develop more robust approaches for genetic selection in beef cattle. (PAG Abstract P108)

Jared Decker’s group (University of Missouri) is creating genomic predictions for meat tenderness and other carcass traits based on SNP haplotypes, which provides increased accuracy in multiple breed predictions.

Weigel, Rosa and Kirkpatrick from University of Wisconsin are co-PIs in NIFA supported CAP projects examining genomics of feed efficiency of dairy cattle, dairy cattle reproduction and susceptibility and resistance to Johnes disease in cattle, respectively. Accomplishments to date are the development of resource populations and preliminary genomic analyses. Additionally, Kirkpatrick is engaged in studies characterizing a major gene for bovine ovulation rate in cattle. Current work involves fine-mapping of the locus and characterization of mRNA and protein expression differences associated with the high ovulation rate allele. Khatib is engaged in study of the genomics of fertilization rate and embryo survival.

Wansheng Liu at Pennsylvania State University has been working with the Y chromosome in order to discover the gene content of the ruminant Y chromosome male-specific region (MSY), to define the genomic structure of the MSY genes, to study the gene expression profile, and to identify Y-linked variations/polymorphisms that impact spermatogenesis and fertility early on in the selection process of sires. Our research focus has been moved from structure genomics to functional genomics on Y-linked genes in spermatogenesis and male fertility. Details for each of Wansheng Liu’s work is shown below (A-E).

 (A) A limited number of Y chromosome lineages present in North American Holsteins.Holsteins are the most numerous dairy cattle breed in North America and have undergone intensive selection for improving milk production and conformation. Theoretically, this intensive selection could lead to the reduction of the effective population size and a reduced genetic diversity. The objective of this study was to investigate the effective population size of the Holstein Y chromosome and the effects of limited Y chromosome lineages on male reproduction and the future of the breed. Paternal pedigree information of 62,897 Holstein bulls born between 1950 and 2013 in North America and 220,872 bulls evaluated by MACE genetic evaluations of Interbull were collected and analyzed. The results indicated that the number of Y chromosome lineages in Holsteins has undergone a dramatic decrease during past 50 years as a consequence of artificial selection and the application of artificial insemination (AI) technology. All current Holstein AI bulls in North America are the descendants of only two ancestors (Hulleman and Neptune H) born in 1880. These two ancestral Y-lineages are continued through three dominant pedigrees from the 1960s, namely Pawnee Farm Arlinda Chief, Round Oak Rag Apple Elevation, and Penstate Ivanhoe Star, with a contribution of 48.78%, 51.06% and 0.16% to the Holstein bull population in the 2010s, respectively. The Y-lineage of Penstate Ivanhoe Star is nearly eliminated from the breed. The genetic variations in the two ancestral Y-lineages were evaluated among 257 bulls by determining the copy number variations (CNVs) of three Y-linked gene families, PRAMEY, HSFY and ZNF280BY, which are spread along the vast majority (95%) of the bovine Y chromosome male-specific region (MSY). No significant difference was found between the two ancestral Y-lineages although large CNVs were observed within each lineage. This study 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.

(B) Male fertility evaluation with a custom-made 384-SNP chip in cattle.We reported the development and evaluation of a custom-made 384-SNP chip for bull fertility analysis two years ago (see 2013 report). During 2014-2015, we have genotyped a total of 935 AI bulls from the Genex population, including 702 Holsteins, 114 Jersey and 119 Angus. Associations of genotypes with sperm quality traits were calculated using linear model procedures in R: Yi = μ + Gi + ei with two approaches, combined-breed and individual breed. We have finished the combined-breed analysis and identified 37 significant SNPs (*p* < 0.05) across all traits with corresponding effects and false discovery rate (FDR)-corrected p-values. No significant SNPs were identified for sperm morphology 1 (Morph1). These 37 significant SNPs were from 35 genes that map to 19 different autosomes in the bovine genome. The Venn diagram analysis indicated that Morph2 appears to have a larger impact on both Total Morph and Total Mot (motility) than Morph1 and Morph3. As to the overlapped SNPs between Total Morph and Total Mot, all but one of the Total Morph SNPs overlapped with the Total Mot SNPs, whereas an additional 17 significant SNPs are unique to Total Mot, indicating that bull sperm motility variations are likely to be affected by more genes (than morphology), or as an alternative explanation, that motility has not been selected as extensively as morphology.

(C) Subcellular localization of the PRAMEY protein in bovine spermiogenesis. The PRAME (*pr*eferentially expressed *a*ntigen in *me*lanoma) proteins are cancer-testis antigens (CTA) with leucine-rich repeat (LRR) domains that fold into a horseshoe shape for protein-protein interactions. The PRAME gene family has been amplified in the genome of eutherian mammals. In bovid lineage, PRAME has been transposed to the Y chromosome through an ‘autosome-to-Y’ transposition, an evolutionary mechanism that usually produces novel sex-linked gene to enhance male reproduction. The Y-linked PRAME gene, named PRAMEY, has expanded on the bovine Y with a median copy number of 13 - 26 among different cattle breeds. The copy number variations (CNVs) of PRAMEY were found to be associated with testicular size, semen quality and male fertility. Our previous western blot analysis with a custom-made anti-PRAMEY antibody on the bovine testes at different developmental stages (20d, 3m, 8m and 2y) indicated that PRAMEY was first expressed at the age of 8 months. Immunofluorescent staining further revealed that PRAMEY was expressed in the acrosome of spermatids and mature spermatozoa, as well as in the sperm flagellum. The objective of the present study was to investigate the subcellular localization of the PRAMEY protein by immunogold electron microscopy (IEM). Adult testis tissues and mature epididymal spermatozoa were collected from a local abattoir and used in the IEM experiments. The results indicated that the immunogold particles of PRAMEY are restricted to the ground substance/matrix of the pre-acrosomal granule (or the acrosome) without evidence of any membrane association in the Stage I round spermatids and the Stage II round spermatids when the spermatids just begin to elongate. It appears that the gold particles were nonrandomly distributed in the pre-acrosomal granule and were associated with the matrix structure. Compared to the very low (almost no) background labels across the cell organelle, gold particles were also found in regions around Golgi complex. Along the formation and expansion of the acrosomal vesicle (spreading over the nucleus) during the differentiation of the spermatids, enriched gold particles were distributed in the acrosomal matrix at the top of the head (near the apical ridge) and a tendency of association with the inner acrosomal membrane (IAM) on the bottom of the head was observed in the elongated spermatids and mature spermatozoa. The gold particle labels were also seen in the flagellum of sperm. Our preliminary IEM data suggests that the PRAMEY protein is localized in Golgi complex, the pre-acrosomal granule that is enveloped in the Golgi vesicle, and acrosomal matrix, signifying a functional role of PRAMEY in acrosome formation during spermiogenesis.

(D) Genome-wide methylation patterns in Holstein leukocytes.This is a collaborative research with Dr. Chad Dechow. The objective of this project was to describe genome-wide DNA methylation patterns in leukocytes from 8 Holstein cows with variable phenotypes. Following DNA extraction from peripheral blood leukocytes, methylated DNA fragments were captured via methylated DNA immunoprecipitation (MeDIP) and 49 bp paired-end reads were sequenced. The number of reads covering each nucleotide was extracted and the geometric mean number of reads (GMR) across cows was determined. GMR was merged with details for n Ensembl genes. The majority (78.4%) of CpG sites were sequenced at least once and repetitive element reads were primarily mapped to satellite (36.4%), SINE (29.1%), and LINE (23.7%) regions. GMR were lower at the centromeric end than for the remainder of the chromosome. There were significant variations at a chromosome level with the lowest DNA methylation for X and the highest for BTA 19. Genomic regions with more annotated genes generally had higher GMR; however, partially methylated domains in gene-rich areas were evident. GMR were lower in upstream than in intra-genic and downstream regions, with the nadir GMR occurring ~100 bp upstream of the transcription start site. GMR was relatively low over the first exon when compared with later exons, whereas GMR was high near intron-exon junctions and lower in the center of introns. DNA methylation levels varied across the genome and tended to be highest in gene-rich regions. DNA methylation was lower upstream and across initial gene exons, corresponding to regions controlling gene transcription levels.

(E) Differential expression analysis of placentae from overfed/obese ewes and lean ewes fed only to requirements.This is a collaborative research with University of Wyoming and the National Center for Genome Resources (NCGR) (at Santa Fe, NM). Sheep are a relevant biomedical model for studying human pregnancy. Maternal diet plays an important role in providing appropriate nutrients to a fetus and promoting healthy growth and development. In several mammalian species including sheep and humans, obesity during pregnancy predisposes offspring to obesity, insulin resistance, and cardiovascular disease. Obesity among pregnant women in the U.S. is now reported at between 18-38% and is increasing. Gaining an increased understanding of how maternal obesity affects an offspring’s predisposition to obesity and its associated diseases is a priority. This study investigated how maternal diet alters gene expression in the placenta, the organ responsible for providing maternal nutrients to the developing fetus. The maternal nutrient environment to which the placenta is exposed determines maternal:fetal nutrient exchange and subsequent programming of gene expression patterns in the conceptus. Deviation from a healthy maternal diet results in altered offspring phenotype and leads to predisposition to obesity and other diseases. Differential expression analysis was performed on placental tissue from pregnant ewes fed either a control (100% of requirements) or an obesogenic (150% of requirements) diet. We performed pathway and enrichment analysis on the differentially expressed genes and identified individual genes and broader pathways that are candidates for further investigation.

Jim Womack’s group (Texas A&M University) has identified variants in chicken antimicrobial peptide genes that are associated with enhanced antimicrobial activity and thus potentially contribute to economically relevant phenotypes (Lee et al, 2014). His group participated in a major GWAS study in cattle that has identified SNPs associated with resistance to bovine respiratory disease (Neibergs et al, 2014).

Stephanie McKay’s group (University of Vermont) has derived global measures of DNA methylation for regions of the limbic system of the bovine brain. DNA from five regions for each of eight cattle will undergo Whole Genome Bisulfite Sequencing in 2016. Additionally, conservation of methylation has been identified at a gene level and this work will be explored on a genome wide level in 2016.

**Objective 3:**

Harvey Blackburn (USDA-ARS National Animal Germplasm Program, NAGP) has developed the Animal – GRIN Genomics Database. The database will serve as a repository for DNA data from the animal genomics research projects. 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 and provide accessibility to valuable data for future research use.

Chris Elsik(University of Missouri) has developed a comprehensive bovine data mining and warehousing system (BovineMine). This is a unique tool that collects all structural and functional information on the cattle genome, and allows users to explore the data constructing queries that integrate the BovineMine dataset. The tool is unique, allowing the examination of tissue specific expression together with genomic variation data.

Database and bioinformatics activities are also coordinated by Jim Reecy (NRSP-8 Bioinformatics Coordinator) at the NAGRP site (http://www.genome.iastate.edu/cattle/).

Zhihua Jiang and Holly Neibergs (Washington State University) have storage and development of large data and DNA sets for dairy and beef cattle.

Gianola, Weigel and Rosa (University of Wisconsin) are developing improved methodologies for incorporation of genomic data in animal improvement programs, specifically genomic selection. Accomplishments include evaluation of alternative methodologies such as Bayesian neural networks and Reproducing Kernel Hilbert Space models.

Alison Van Eenennaam (UC Davis), as part of the Bovine Respiratory Disease Coordinated Agricultural Project (BRD CAP; USDA NIFA grant 2011-68004-30367) the eBEEF.org website (the beef genetics/genomics community of practice within eXtension) was officially launched at BIF 2015 in Biloxi, MS. eBEEF.org is part of the national eXtension program with the goal of being a one-stop site for beef cattle genetics and genomics information. Beef cattle specialists from six land grant institutions have joined forces to provide educational materials that are pertinent to today’s beef cattle producers, without searching multiple sites or filtering through countless hits on a search. The site contains factsheets, short frequently asked question (FAQ) video clips, relevant conference recordings and webinars, a blog and links to other useful beef sites.

Another goal of the eBEEF.org website is to archive the information generated from current and future beef genetics integrated grants funded by USDA-NIFA. All eBEEF.org team members are a part of one or more of the three current grants (Integrated Program for Reducing Bovine Respiratory Disease Complex in Beef and Dairy Cattle; National Program for Genetic Improvement of Feed Efficiency in Beef Cattle; and Identification and Management of Alleles Impairing Heifer Fertility While Optimizing Genetic Gain in Beef Cattle. For more information or to make suggestions please contact any of the eBEEF.org team members. The other team members are Dr. Darrh Bullock (University of Kentucky); Dr. Jared Decker, University of Missouri; Dr. Megan Rolf, Oklahoma State University; Dr. Matt Spangler, University of Nebraska; and Dr. Bob Weaber, Kansas State University.

As part of a collaborative grant entitled “Identification and management of alleles impairing heifer fertility while optimizing genetic gain in Angus cattle” (USDA NIFA grant 2013-68004-20364), collaborator Dr. Jerry Taylor at University of Missouri is sequencing more than 250 Angus, Hereford, Simmental, Charolais, Gelbvieh, Red Angus, Maine Anjou, Beefmaster and Limousin bulls to identify loss of function alleles. These will then be tested on a populations of 10,000 heifers with reproductive and breeding data that are part of the Show-Me-Select heifer replacement program. The ultimate objective of this research project is to identify lethal recessive alleles and develop tools for the implementation of strategic mating. I have been working with Dr. Brian Kinghorn, at the University of New England in Armidale, Australia to develop a computer program to optimize mating between animals carrying multiple loss of function alleles. Results from a simulation exercise were presented at the Applied Reproductive Strategies in Beef Cattle Conference. Computerized mating programs can help to simplify the use of DNA information to ensure that carriers are mated strategically to minimize the incidence of affected offspring, while still utilizing their genetics when the value of their merit overrides the discount associated with their carrier status.

A NIFA conference grant (USDA NIFA grant 2015-67015-23693) was obtained for the Applied Reproductive Strategies in Beef Cattle Conference (ARSBC) that was held at UC Davis from August 17th-18th, 2015. A total of 183 attendees registered for the meeting. Summaries, PowerPoint slides and audio of speaker presentations are published online on the Newsroom tab of the conference website ([www.appliedreprostrategies.com](http://www.appliedreprostrategies.com)). Other conferences that Alison helped organize this year included a half-day BRD Symposium is being planned in conjunction with the annual American Association of Bovine Practitioners (AABP) meeting which was held in September 2015 in New Orleans; and the 10th Transgenic Animal Conference will was held at Tahoe City, CA August 9-13, 2015.

**MEETINGS:**

Coordination funds supported student travel awards for PAG-XXIII in January 2015, and PAG XXIV in January 2016.

**C. IMPACT STATEMENTS:**

* We expect the inclusion of the Optical Map, PacBio de-novo assembly and improved bioinformatics approaches to enhance the quality of the bovine reference assembly and subsequent analysis involving the reference assembly.
* We expect that the diverse strategies undertaken to identify allelic variation that contributes towards economically relevant phenotypes will be the ability to incorporate this new information in animal breeding improvement programs.
* Distribution and storage of DNA and tissue repositories has impacted the ability to proceed with research endeavors, especially those involving functional genomics.
* We expect that an impact of this research is the ability to incorporate this new information in animal breeding improvement programs.

**D. GRANTS AWARDED SINCE OCTOBER 1, 2013: Total Funds $5,226,248**

**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:*

A. New Reference Genome Assembly. Ted Kalbfleisch, 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 is planned for completion in August 2016.

B. Whole Genome Sequences. In connection with research projects, many of which are cited in the reference section, over 200 horses have had their whole genomes sequenced. Many of those sequences are being used for the new assembly described in the previous paragraph and were used to identify SNPs for construction of the 670K SNP assay tool described below.

C. Access to reference DNA. The primary CHORI 241 BAC library was moved from the Children’s' Hospital of Oakland to the laboratory of Samantha Brooks (co-coordinator) at the University of Florida. This action was necessitated by the closing of the commercial operation of the BAC library. This move ensures continued research access to the library.

*Objective 2:*

We anticipate that discoveries in the future will be based on use of this newer higher density SNP chip. However during 2015 the molecular basis was discovered for many horse diseases affecting muscle, nervous system, respiratory system and immune system. Genetic signals were identified for yet other disease genes. Furthermore, progress was made in identifying the effect of genes previously shown to influence racing performance (MSTN) and gait (DMRT3). Other studies reported approaches to identifying estimated breeding values for performance of Thoroughbred, harness racing horses and warmblood horses.

A. 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.

*Objective 3:*

 During 2015, the horse workshop group was invited to join the FAANG initiative. Jamie MacLeod of the University of Kentucky serves on the guiding committee for FAANG and also oversees cooperation among scientists studying the horse. During 2015, three scientists (Carrie Finno of UC Davis, Rebecca Bellone of UC Davis and Jessica Peterson of the University of Nebraska) applied for funds to collect tissues and begin conducting assays associated with the FAANG program. We are waiting to hear the results of that application. Regardless, we have committed NRSP8 coordinator funds to collection of tissues to benefit this or future applications. A meeting of the horse genome workshop committee is being planned for the ISAG meeting to be held in Utah in July 2016 to make plans.

**Funding Sources Summary for 2013-2015**

(summaries provided by scientists from 16 participating laboratories and therefore represent a partial account of research funds obtained for horse genomics research in 2014)

Federal Funding: $8,883,000

Industry: $1,818,000

Local/Institutional: $3,854,000

Total $14,555,000

**Impacts 2013-2015**

1. Mutations responsible for horse diseases involving muscle, nervous system, respiratory system and immune system were identified and diagnostic tests created for use by veterinarians and horse owners.

2. The diverse use of horses, the diverse methods for measuring performance and the small population sizes present problems for applying quantitative genetic approaches to evaluation. With the advent of SNP assays, the developments of genomic selection assays for performance are being investigated for use by horse breeders.

**Meetings**

A. Workshop Chair for 2016 PAG meeting: Ted Kalbfleisch (University of Louisville) ted.kalbfleisch@louisville.edu

Workshop Chair for 2017 PAG meeting: Carrie Finno (University of California, Davis) cjfinno@ucdavis.edu

Workshop Chair for 2018 PAG meeting: Jessica Petersen (University of Nebraska)

jessica.petersen@unl.edu

The workshop participants met on Saturday and Sunday, January 9-10, 2016 at the Plant and Animal Genome Conference in San Diego. The NRSP8 meeting was held in conjunction with the Dorothy Russell Havemeyer workshop activities. Approximately 80 people attended the sessions with participants from at least 10 countries (USA, Brazil, China, Japan, Korea, Denmark, United Kingdom, Italy, Argentina, Ireland). Ted Kalbfleisch served as chair of the 2016 workshop. He will step down after this year and the next chair will be Carrie Finno. At the meeting, Jessica Petersen was elected as vice-chair and will assist Carrie Finno in 2017 and assume full leadership of the workshop in 2018.

Jessica Petersen

jessica.petersen@unl.edu

(402) 472-6328

Department of Animal Science

A218g Animal Science Building

University of Nebraska-Lincoln

Lincoln, NE 68588

B. The workshop meeting at PAG followed a July 2015 Havemeyer Conference. During the Havemeyer Conference there were 42 podium presentations and 26 poster presentations over 3 days ranging in topics from studies of ancient DNA, comparison of extant populations, investigation of structural variations among horses, discovery of mutations affecting disease and performance and the use of genomics to better understand the quantitative genetics underlying performance and diseases of horses. In addition there were reports on development of infrastructure tools.

C. During the PAG conference (January 2016) two workshop sessions were held on horse genomics with 14 podium presentations. At the PAG meeting the focus of the meeting was on development of infrastructure and specifically, use and development of databases, creation of a new genome assembly and development of tools for investigating gene expression more effectively. In addition, scientists present podium talks on discoveries related to horse genomics and 36 posters were presented related to horse genomics.

**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*

**A. Reference linkage map.** To improve the genetic map, unassigned sequence contigs from the latest genome assembly build (galGal 5) were surveyed. This screened identified 9,585 contigs 2 Kb or larger in size of UCD001 (Jungle Fowl and one parent of the East Lansing mapping population); theoretically, this suggest that approximately 1.4 Mb of the 76.6 Mb remains unassigned in the chicken genome sequence. Aligning existing UCD003 (White Leghorn and other parent of the East Lansing mapping population) reads, 4,160 scaffolds were identified that cover ~39.6 Mb with at least 1 SNP of reasonable confidence; thus, 5,425 scaffolds that account for ~37 Mb had no SNPs for potential genetic mapping. Further filtering based on Affymetrix design scores gave 5,907 SNPs on 547 contigs that could be potentially genotyped. However, in the end, 3,440 SNPs and 510 contigs were assayed and successfully scored. Following genetic mapping, 3,437 SNPs were assigned to 29 new linkage groups (called E101 to E129), none of which were linked to existing markers. Three SNPs were unlinked and assigned to E00. This information is being used to build new sequence contigs for the genome assembly (see below).

**B. Chicken genome sequence.** Wes Warren, The Genome Institute at Washington U., and co-workers released the latest build, known as Galgal5.0, which incorporated information from 30x PacBio coverage from a 10 Kb library (funded by Cobb-Vantress). This improved build saw a gain of 180 Mb in assembled bases, an increase of N50 contig size from 252 Kb to 2.8 Mb, and reduction in the number of gaps from 2,687 to 381. Of the 1.21 Gb genome, approximately 90% of the sequence has been anchored to chromosomes, which include autosomes 1-28 and 30-33, one additional linkage group (LGE64), and sex chromosomes W and Z. Upon mapping of reference reads to chromosomes of each version, there is a 1% average gain in the reads aligned to Galgal5.0. Galgal5.0 when annotated for gene content shows a gain of 4,679 genes, 2,768 non-coding and 1,911 protein-coding above Gallus\_gallus4.0.

*Objective 2*

DNA from the East Lansing international reference mapping population has been sent 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.

*Objective 3*

Database activities are led by the NRSP-8 Bioinformatics Coordinator, Jim Reecy, and Susan Lamont, along with Shane Burgess, represent poultry interests on the advisory committee for this group. Poultry bioinformatics has also benefitted from support at several other locations. We maintain a homepage for the NRSP-8 U.S. Poultry Genome project (<http://poultry.mph.msu.edu>) that provides a variety of genome mapping resources, including our newsletter archive. The Poultry Genome Newsletter is published quarterly and is distributed through our Homepage and on the angenmap email discussion group.

**FAANG ACTIVITIES (general NRSP-8 and poultry-specific)**

To address genome to phenome challenge, the FAANG Consortium was launched and the white paper with the goals of a first phase of FAANG including initial guidance on parallel deep sample and metadata collection from species with high-quality genome assemblies, as well as defining specific data types and infrastructure needed for this initial research was published in Genome Biology. A Steering Committee was created to develop and implement policy to advance the FAANG project, with representatives from major species and many countries involved. Four specific activity committees were formed, including Animals, Samples and Assays (ASA), Metadata and Data Sharing (M&DS), Bioinformatics and Data Analysis (B&DA), and Communications (COM). These committees frequently meet by conference call, to develop policies and procedures to advance FAANG.

Subsequently, two FAANG specific workshops have been held. GO-FAANG in Washington DC with 100 more participants and FAANG Workshop at the PAG meeting with 175 participants including scientists, administrators, representatives of funding agencies and commodity groups to discuss the latest advancements of the Consortium and new perspectives.

Several FAANG associated projects were started or initiated. These projects include: NIFA UC Davis (chicken, pig and cattle focusing on 8 representative tissues), horse FAANG initiative, and sheep FAANG initiative (Washington State University).

**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. Transcripts detected in these tissues show good coverage of the Ensembl gene sets, and an initial analysis has identified putative long non-coding RNAs, both tissue-specific and expressed across all 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.

**MEETINGS:**

Over 3,700 scientists attended the joint Plant and Animal Genome XXIII meeting January 2015, held jointly with the annual NAGRP meeting and the well-attended poultry workshop with invited speakers and presentations from NRSP-8 and NC-1170 members. Coordination funds helped support attendance.

**Travel support for students** - Tae Hyun Kim, Recipient of Neal Jorgenson Genome Travel Award from UC Davis (Zhou, PI); John Hsieh, Iowa State U. graduate student (Lamont, PI); Perot Saelao, UC Davis (Zhou, PI); Claire Stice, Cornell University (Johnson, PI**).**

**Travel support for poultry workshop speakers** - Dr. Steffen Weigend (Friedrich-Loeffler-Institut, Institute for Farm Animal Genetics, Germany); Dr. Doris Bachtrog (UC Berkeley)

GO-FAANG October 2015 (Washington, DC) conference travel support for attendance of Drs. Hans Cheng (USDA-ARS-ADO) and Huaijun Zhou (UC Davis).

**Additional support:** GO-FAANG conference matching fund support.

**IMPACT:** This project is generating 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 of poultry.

Funding Sources Summary for 2013-2015

Federal Funding: $19,614,189

Industry: $ 307,919

Local/Institutional: $ 231,948

Total $20,154,056

**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

PROGRESS OF THE WORK AND PRINCIPAL ACCOMPLISHMENTS:

The NRSP-8 sheep co-coordinators are participants in the International Sheep Genome Consortium (ISGC; http://www.sheephapmap.org/) and the Internaional Goat Genome Consortium (IGGC; www.goatgenome.org). These multi-institutional organizations have assumed a key role in the coordination and prioritization of ovine and caprine genomic resources. The following objectives have been greatly aided through the ISGC and IGGC efforts.

*Objective 1* (Advance reference genomes):

The domestic goat genome was the first livestock genome published that leveraged whole genome optical mapping. This work was published in Nature Biotechnology in 2013.

An ongoing project of the ISGC is development of a whole genome reference assembly. In 2010, sequence data were generated at two sequencing facilities (Beijing Genomics Institute and the Roslin Institute) from DNA of a Texel ewe and a Texel ram, respectively. A paper was published in Science in June, 2014 describing the ovine whole genome assembly (Oar v3.1), the RH map and the linkage map.

Kim Worley (BCM-HGSC) received funding from a 2013 USDA/AFRI grant to fill in gaps and improve the Texel sheep assembly using PacBio sequence data with PBJelly (scaffolding and gap filling). The sequence reads were long (up to 10 kb average) and allowed 66% of the existing gaps in the draft sequence to be closed. The assembly is more contiguous, with the contig N50 increased by 3.7 fold (from 41.7 kb to 165.2 kb). The PacBio work will be in Oar v4.

As part of the Sheep Genomes Database, another whole genome reference sequence has been constructed at BCM-HGSC using de novo techniques. The sequenced animal is a Rambouillet female whose tissues will contribute to the ovine FAANG project that is under development. The use of a single animal for the reference sequence and functional annotation will allow better alignment and characterization of the genetic elements controlling gene expression.

*Objective 2* (Identify and exploit genes and allelic variants):

The ISGC developed a high density array with 600,000 SNP to complement an existing 50,000 SNP array technology. The new HD chip fills in gaps in genome spacing and haplotype coverage for gene/variant discovery with a wide range of traits. This HD chip also provides state-of-the-art genotypes to enable genomic selection at unprecedented resolution in domestic sheep.

The IGGC has contributed to development of a goat 50,000 SNP array. This is the first large-scale SNP array for goats. It enables both genome-wide association and genomic selection approaches that have never before been possible in goats.

*Objective 3* (Facilitate analysis, distribution, and application):

Utah State University, CSIRO Australia and the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) are currently developing a resequencing database of sheep whole genome sequences with extensive annotation of variants. To date, sequences have been obtained for more than 450 animals and aligned to Oar v3.1 as well as a new ovine reference sequence generated from a Texel animal donated by USDA, ARS MARC. 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 will be publicly available via EVA in April, 2016 or directly from the Sheep Genome database in February, 2016. Variants will be 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.

An ovine FAANG project is being organized in which up to 45 tissues will be collected from the Rambouillet animal being used for the PacBio long read de novo sequence. Investigators from USDA/ARS MARC, USDA/ARS Beltsville, Washington State University, Utah State University, BCM-HGSC, University of Missouri and Virginia State have expressed interest in being participants on the project. Sufficient quantities of tissues will be collected and either processed at time of collection or snap frozen in liquid nitrogen so that several techniques and assays can be performed by participating labs. This project will be linked with the Roslin Atlas project conducted on sheep in order to maximize connection and minimize redundancy.

Funding Sources Summary for 2013-2015

Federal Funding: $2,644,000

Industry: $ 109,000

Local/Institutional: $ 445,000

Total $3,198,000

B. BRIEF SUMMARY OF MINUTES OF ANNUAL MEETING:

The 2016 annual meeting of the NRSP-8 Cattle, Sheep, and Goat committee was held on Jan 9-10, 2016 in conjunction with the Plant and Animal Genome XXIV meeting. The morning session of the scientific meeting on Jan 9th was held as a joint session in with the Swine Committee with 4 presentations centering around a theme of gene editing. The Saturday afternoon and Sunday morning sessions of the combined Cattle/Sheep/Goat workshop included 19 presentations covering a wide variety of topics from comparative genomics, low budget reference grade genome assembly, feed efficiency and methane emissions, genotype by sequencing, genome annotation, omics data resource, comparative proteomics, recombination mapping, copy number variation, genomic selection, male fertility, and an overview of genome Canada funded research. Attendance at the sessions was exceptional with approximately 200 people attending each one of the scientific sessions, including delegates from at least 29 countries, including Australia, Belgium, Brazil, Canada, Chile, China, Denmark, Egypt, Ethiopia, France, Germany, Ireland, Italy, Japan, Kenya, Mexico, Netherlands, New Zealand, Norway, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Uganda, UK, Uruguay. Stephanie McKay was thanked for serving as President of the NRSP-8 Cattle, Sheep and Goat Committee in 2015-16. Jared Decker will serve as President in 2016-17. Brenda Murdoch was elected as the 2017-2018 Secretary, and she will serve as President in 2017-2018.

ACCOMPLISHMENTS AND IMPACTS:

*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.

Faculty from VSU has been involved in several projects with the IGGC consortium members. The sequencing of the goat genome and development of a more refined genome assembly has continued over the past year. The San Clemente Island goat, “Papadum”, sequencing and assembly by the USDA and VSU has been evolving as the sequencing and assembly technologies evolve. Currently, we have about 67X coverage of Illumina HiSeq sequence data and 68X coverage with PacBio sequences. The initial assemblies were developed through collaboration with Sergey Koren and Adam Phillipy for the initial corrected sequence set and assembled contigs. Collaboration with Ivan Liachko, Shawn Sullivan and Jay Shendure, University of Washington, on development and use of Hi-C techniques and BioNano (Jeff Robinson & Alex Hastie) for optical mapping technologies for assembly of contigs into scaffolds and chromosomes. The RH map that we created a few years ago was an integral tool used to determine the quality of assemblies and final chromosome identification. The project is expected to have a data freeze in the next month and a publication to follow quickly.

*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.

WC1 co-receptors belong to the scavenger receptor cysteine-rich (SRCR) superfamily and are encoded by a multi-gene family. 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. Expression of either WC1.1-type or WC1.2-type proteins is correlated with gamma delta T cell responsiveness to pathogens. 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. To obtain an estimate of the extent of the WC1 gene array in livestock species other than bovine, we PCR-amplified N-terminal a1 domain sequence. We focused on the a1 domain because: (i) it is the signature domain for WC1 genes, (ii) there is only one a1 domain in each 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. Using genomic DNA from a single animal to minimize allelic polymorphism, we cloned the PCR products into a sequencing vector (pCR2.1-TOPO-TA, Invitrogen) and sequenced 100 ovine, 80 caprine, and 23 swine clones. We aligned the sequences using ClustalX and found 2 porcine sequences with 10-fold redundancy and 3 unique sequences that differed by one amino acid from the redundant groups. In contrast, the 59 unique ovine and 55 caprine sequences showed low redundancy: the caprine sequences had only 15 groups of 2 to 5-fold redundancy, and 41 groups with unique sequences, suggesting a potentially large gene array. Based on the number of unique sequences and allowing for allelic polymorphisms and single nucleotide mutations introduced by DNA polymerase, we estimate that there are ~25 WC1 genes in the ovine and caprine genomes, but the actual number remains to be determined. While most ovine and caprine sequences cluster together by phylogenetic analysis, several of each cluster with either bovine WC1.1-type or WC1.2-type a1 domains. There is at least one ovine and one caprine a1 sequence that cluster close to bovine a1 domains of WC1-3, -4, -8,-10, and-12, which we have shown binding to Leptospira or Mycobacteria. Our two porcine WC1 a1 sequences can be also be divided into WC1.1-type and WC1.2-type because WC1-2-type WC1 a1 domains have a 4 amino acid insertion in the C-terminal end. Two WC1 proteins have been predicted from the assembled porcine genome sequences:XP\_005655753.1 (gene id 100144477) and XP\_005655754.1 (gene id 100627089). There are some apparent errors in the genome assembly: (i) neither genes have the sequence encoding the WC1.2-type WC1 a1 domain. The WC1.1-type WC1 a1 sequence is present in gene id 100144477; (ii) The 5’ end of the gene id 100627089 sequence includes a CUB domain and a d SRCR domain in the place of an a1 SRCR domain, followed by [b-c-d-e-d’]-TM-cytoplasmic tail. We could not amplify sequence between the CUB domain and the d domain, suggesting that this is a genome assembly error; and (iii) the b2 domain that is continuous with the WC1.2-type a1 from cDNA obtained from a Duroc x Yorkshire F1 cross is 99% identical to that in gene id 100144477, which instead contains the WC1.1-type a1 domain sequence. We have obtained 5’-RACE sequence from the WC1.2-type genes to address this problem.

Differential expression analysis of placentae from overfed/obese ewes and lean ewes fed only to requirements. This is a collaborative research with University of Wyoming and the National Center for Genome Resources (NCGR) (at Santa Fe, NM). Sheep are a relevant biomedical model for studying human pregnancy. Maternal diet plays an important role in providing appropriate nutrients to a fetus and promoting healthy growth and development. In several mammalian species including sheep and humans, obesity during pregnancy predisposes offspring to obesity, insulin resistance, and cardiovascular disease. Obesity among pregnant women in the U.S. is now reported at between 18-38% and is increasing. Gaining an increased understanding of how maternal obesity affects an offspring’s predisposition to obesity and its associated diseases is a priority. This study investigated how maternal diet alters gene expression in the placenta, the organ responsible for providing maternal nutrients to the developing fetus. The maternal nutrient environment to which the placenta is exposed determines maternal:fetal nutrient exchange and subsequent programming of gene expression patterns in the conceptus. Deviation from a healthy maternal diet results in altered offspring phenotype and leads to predisposition to obesity and other diseases. Differential expression analysis was performed on placental tissue from pregnant ewes fed either a control (100% of requirements) or an obesogenic (150% of requirements) diet. We performed pathway and enrichment analysis on the differentially expressed genes and identified individual genes and broader pathways that are candidates for further investigation.

The objective of a study conducted at Utah State University was to map the genetic determinant controlling the production of four horns in two sheep breeds. Genome wide association mapping was performed using 125 animals from the Jacob and Navajo-Churro breeds that contain two and four horned individuals. A case – control design analysis of four-horned versus two-horned animals using the ovine HD SNP Beadchip revealed a strong association signal of sheep chromosome 2. This signal on OAR2 is clearly different than the locus controlling the presence or absence of horns in sheep which is found on ovine chromosome 10 (OAR10). The ten most strongly associated SNP on OAR2 are all located in a region spanning Mb positions 131.9 to 132.6, indicating the genetic architecture underpinning the production of four horns is likely to involve a single gene. The closest genes to the most strongly associated marker (OAR2\_132568092) are MTX2 and the HoxD cluster, located approximately 93 Kb and 251 Kb upstream respectively. The occurrence of an eyelid malformation across both breeds was restricted to polled animals and those carrying more than two horns. This suggests the eyelid abnormality may be associated with departures from the normal developmental production of two horned animals, and that the two conditions are developmentally linked. The study was conducted in collaboration with James Kijas, CSIRO Australia.

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 a study conducted at Utah State University, 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 include more animals in the analysis and localize the significant regions so that underlying genes or genetic regulatory factors can be identified.

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. The S127 allele was one of those, but collaborative work has shown the GS127 heterozygotes can have incubation times of 647-1333 days, compared to a mean of 289 days for “wild-type” GG127 homozygotes.

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. There is evidence to suggest a genetic basis for control of C. burnetii transmission from domestic ruminants, and a new research focus is the identification of responsible loci.

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.

Ongoing Study 1. Microbial profiling to predict feed efficiency in sheep. Objective. Identify rumen microbial characteristics that can be used to predict feed efficiency status in sheep. Experimental Approach and Procedures. Conducted sheep trial in which ewe lambs were ranked for feed efficiency; DNA from rumen fluid of ewe lambs divergent for feed efficiency was sequenced for taxa identification. Profiles of taxa associated with feed efficiency status are being tested on a naïve set of lambs that were feed efficiency tested. One Ph.D. student is finishing this project, with a projected completion date of February, 2016. RNA sequencing was also performed to determine functional changes/differences in the rumen associated with divergence for feed efficiency; this analysis is nearing completion. Collaborative Partners. University of Missouri (Gavin Conant; William Lamberson); University of Alberta (Leluo Guan)

Ongoing Study 2. Rumen microbial changes associated with high sulfur drinking water. Objective. Determine changes in rumen microbial species associated with exposure to high sulfur drinking water in sheep. Experimental Approach and Procedures. Conducted sheep trial in which lambs were administered high sulfur drinking water for a 28 d period. Rumen fluid collections were conducted on d 0, 7, 14, 28 and 35. DNA sequencing has been performed on the rumen fluid samples, and analyses are currently underway to identify microbial species important to the response, and perhaps adaptation, to high dietary sulfur (drinking water, in this case). Collaborative Partners. University of Missouri (Gavin Conant; William Lamberson)

New Study 1. Genetic and maternal influences on progeny rumen microbiome and feed efficiency. Objective. Determine the relative contributions of 1) genotype, 2) perinatal maternal environment, and 3) postnatal environment on offspring rumen microbiome and feed efficiency in beef cattle. Experimental Approach and Procedures. We will use two biologically diverse cattle breeds, and also cross-fostering and c-sections to elucidate these relative effects. This is a new project scheduled to start in January, 2016. Collaborative Partner. University of Missouri (Gavin Conant; William Lamberson); New Mexico State University (Shanna Ivey); Virginia Tech (Becca Cockrum)

We are continuing collaboration with the USDA-ARS Beltsville faculty and research groups from ILRI, ICARDI, ASARECA working in Africa and international collaborators from Brazil, Austria, UK, China, New Zealand and Australia, on a project to improve management of goat genetic resources and the goat production value chain in Africa. The project has collected >2500 samples from ~55 sites in 17 African countries. The extractions of DNA are complete for most of those samples and are currently working on genotyping these samples with the Illumina 50K SNP panel.

We have utilized the Illumina 50K SNP panel to begin characterization of U.S. meat goat breeds. We have genotyped samples from the Boer, Myotonic, Kiko and Spanish breed populations. We are using the panel to analyze the breed relationships and origins as well as inbreeding levels. This data set has been used in a comparative study with goats from Egypt in a collaboration with Max Rothschild at Iowa State University.

We have been working with an international set of collaborators to develop the goat ADAPTmap project with the goal of organizing goat genotyping and re-sequence DNA data in a combined database to identify factors associated with adaption in the goat. The project has progressed to the creation of the database and collection of genotypes on over 4,000 individuals. Currently, subgroups of researchers are developing analysis projects using the data. I am coordinating the comparative genomics subgroup, but actively involved in several of the subgroups. My focus is on analysis with runs of homozygosity. We expect to have an initial set of papers on the data completed by the end of this year.

*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.

Utah State University, CSIRO Australia and the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) are currently developing a resequencing database of sheep whole genome sequences with extensive annotation of variants. To date, sequences have been obtained for more than 450 animals and aligned to Oar v3.1 as well as a new ovine reference sequence generated from a Texel animal donated by USDA, ARS MARC. Analysis of 250 of the sequences have revealed over 80 million SNPs and indels with high confidence using two variant-calling platforms. Data in the database will be publicly available via EVA in April, 2016 or directly from the Sheep Genome database in February, 2016. Variants will be available as i) raw (unfiltered) or ii) filtered following application of a comprehensive QC protocol.

We are continuing to develop our bioinformatics tools and activities. Over the past several years, we have been developing of a novel mathematical model for candidate gene finding utilizing bioinformatics and model systems analysis approach. We have developed methods for identifying protein-protein interactions, transcription factor binding sites and epigenetic factors from multiple species and are working on methods to combine these data types for an improved prediction of candidate genes in a cross-species analysis. This model has progressed well and we currently are confident in the outcomes using simulated data sets. The current research goal is to develop the algorithm into a tool for the Animal QTL Database in a collaboration with James Reecy and continue to refine the algorithm.

PARTICIPANTS:

Louisiana State University: James E. Miller\*1

Oklahoma State University: Udaya DeSilva\*1

Pennsylvania State University: Wansheng Liu\*1

Texas A&M University: Clare Gill\*1, Penny Riggs\*

University of Florida: Raluca Mateescu\*1

University of Massachusetts-Amherst: Janice Telfer\*1, Cynthia Baldwin\*

University of Wyoming: Kristi Cammack\*1

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\*

\*Voting member.

1In attendance at the 2016 NRSP-8 meeting.

**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.

*Objective 1*

A. 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. 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).

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.

Newly approved projects during reporting period:

1. 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.

2. 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 both of these projects, the Swine Genome Coordinators had a co-PI role, so the proposals were vetted through the Advisory Committee for approval.

*Objective 2*

A. 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.

*Objective 3*

*A.* 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 14,479 pig QTLs from 507 publications curated into the database, a 4 % increase over the end of 2014. Those QTLs represent 592 different traits. Throughout 2015, the NAGRP bioinformatics team has continued their efforts to make improvements to the Animal QTLdb, which includes a new mirror site in China, facilitate the addition of gene network analysis data, improved search tools and data analysis tools. 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.

Communication:

The Pig Genome Update has now published 122 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:

MEETINGS:

A. The Swine Genome coordinators have been working with a large number of individuals in many countries to develop a new initiative, called Functional Annotation of ANimal Genomes (FAANG). This group proposes a project to identify all functional elements in animal genomes, and has presented their plans on a website organized by the Swine Coordination effort (see www.faang.org). A first international Workshop (GO-FAANG), chaired by Chris Tuggle and held in Washington DC. Approximately 100 attendees were from 24 US states and 13 non-US countries; many additional scientists attended via the web. An important component of this Workshop was the participation of 6 funding agencies from US, Canada, and Europe, who were very supportive of the goals of FAANG and provided guidance on obtaining funding for FAANG-related research. Talk videos are available: http://www.animalgenome.org/community/FAANG/bbs?s=go-faang..txt.

B. A unique opportunity to advance discussions on genomics presented itself when the organizers of the 2016 Joint Annual Meeting (JAM) of the ASAS, ADSA, CSAS and other organizations scheduled a day of programmatic overlap with the 2016 International Society of Animal Genetics (ISAG) meeting. Both groups meet on July 23, 2016 in Salt Lake City, Utah, and several FAANG members including Chris Tuggle and Stephen White (NRSP-8 Sheep Coordinator) proposed that this day be devoted to a Symposium on FAANG. The proposal was approved and this Symposium is being planned. See https://www.asas.org/meetings/isag2016/program for current details.

C. Travel of several scientists was partially funded to attend important pig

genomics meetings in the reporting period.

2015:

Melanie Trenhaile, University of Nebraska-Lincoln, 2015 Neal Jorgenson Travel Award winner; Elisabetta Giuffra, INRA, 2015 NRSP-8 special speaker on FAANG; Huaijun Zhou, University of California-Davis, Midwest ASAS Functional Genomics Worksho.p 2016:

Jeremy Howard, North Carolina State University, 2016 Neal Jorgenson Travel Award winner. We are also partially supporting the travel of speakers to the 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 Peagaricano, University of Florida, Swine Workshop.

D. NRSP-8 Swine Genome Committee Report (January 1, 2015 – December 31, 2015).

2015 Chair (for 2016 Workshop): Cathy Ernst (Michigan State University); ernstc@msu.edu

2015 Chair-elect: Kiho Lee (Virginia Tech University); kiholee@vt.edu

The 2016 NRSP-8 Swine Workshop was held January 9, 2016 in San Diego, CA in conjunction with the Plant and Animal Genome XXIV Conference. A joint session was held with the Cattle, Sheep and Goat Workshop in the morning focused on the theme of Genome Editing. The afternoon Swine Workshop program included invited presentations by four young scientists from Konkuk University in South Korea, the University of Barcelona, the Roslin Institute, and the University of Florida, who spoke about their work in emerging areas of pig genomics. The Jorgensen Pig Travel Award winner, Jeremy Howard, from NSCU was introduced and he gave a lightening talk on his area of research. There were also ten presentations from nine different NRSP-8 participating locations. The presentations covered a range of topics from functional genomics to SNP analysis, as well as a broad range of important phenotypes, and sparked discussion among attendees throughout the workshop. Drs. Clutter 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, 176 attendees signed in, including 80 (45%) from 20 different countries outside the US. Attendees from the US represented 21 universities, 14 industry companies and 2 government agencies. In the afternoon, the Swine Workshop had 33 people sign in, although it is estimated that at least 75 were present for the invited talks. Among those signing in, 12 attendees represented 7 countries outside the US, and the 21 US attendees were from 10 universities, 3 industry companies and 2 government agencies. During the business meeting, Dr. Chris Tuggle from Iowa State University was elected as the new chair-elect, and Dr. Kiho Lee from Virginia Tech University will chair the 2017 Swine Workshop.

E. Partial summary of funding awarded to Swine committee scientists (2013-2015)

Federal $5,329,559

Industry $274,107

Internal/Institutional $182,090

Total $5,785,756

**Impacts 2015**

1. Gene editing technology was applied at the University of Missouri to develop pigs that are resistant to Porcine Reproductive and Respiratory Syndrome virus (PRRSv), the most devastating disease in the swine industry.

2. NRSP8 leadership has contributed to organization of the Functional Annotation of ANimal Genomes (FAANG) initiative, which proposed to identify all functional elements in animal genomes, and the Swine Coordination effort led an international workshop (GO-FAANG) in 2015.

**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.

**PROGRESS TOWARD OBJECTIVE 1**: Create shared genomic tools and reagents and sequence information to enhance the understanding and discovery of genetic mechanisms affecting traits of interest. (See activities listed below.)

**PROGRESS TOWARD OBJECTIVE 2**: Facilitate the development and sharing of animal populations and the collection and analysis of new, unique, and interesting phenotypes

The partnership with researchers at Kansas State University, Michigan State University, Iowa State University, and the U.S. Department of Agriculture continues as the database and website interface developed for this collaboration (http://www.animalgenome.org/lunney) have been improved, and continued data generation by the group has increased the amount of data that is housed in the database. This resource continues to help the consortium by offering a localized source of information and continued facilitation of data analysis.

**PROGRESS TOWARD OBJECTIVE 3**: Develop, integrate, and implement bioinformatic resources to support the discovery of genetic mechanisms that underlie traits of interest.

The following describes the project's activities over this past year.

**Multi-species support**

The Animal QTLdb and the NAGRP data repository have been actively supporting the research activities for multiple species. A collaborative site at iPlant has been set up to share some of the web traffic, including the JBrowse server to serve the cattle, chicken, pig, sheep, 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 dataâ€” XYPlot/Density, in BAM or VCF formatâ€”to be loaded directly to a userâ€™s browser for comparisons in the local environment (users need to learn how to use JBrowse). New data sources and species continue to be updated. We have recently set up a virtual machine site to host the Online Mendelian Inheritance in Animals (OMIA) database created and maintained by Dr. Frank Nicolas at the University of Sydney (http://omia.animalgenome.org/). This is part of our collaborative effort to migrate OMIA to NAGRP platforms.

**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 has now been added to NCBOâ€™s BioPortal (http://bioportal.bioontology.org/ontologies/LPT). We have also continued mapping the cattle, pig, chicken, sheep, and horse QTL traits to Vertebrate Trait Ontology (VT), Product Trait Ontology (PT) and Clinical Measurement Ontology (CMO) to help standardize the trait nomenclature used in the QTLdb. At the request of community members, at least 15 new terms were added to the VT in 2015. 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/PT/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, 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 has also been deposited into BioPortal (http://bioportal.bioontology.org/ontologies/LBO).

As of October 15, 2015, AgBase provides 1,539,447 GO annotations for 310,971 gene products from 504 species, including more than 40 agriculturally important species and their pathogens. This information includes 392,101 GO annotations for 57,589 avian gene products, 96% of which are associated with chicken and turkey.

**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 made continually to SNPlotz, Gene Ontology CateGOrizer, and the Expeditor. Bundled with ReviGO, the CateGOrizer is serving users for both GO term categorization and semantic representation.

As a result of collaborations between Iowa State University, the Medical College of Wisconsin, and University of Iowa, the Virtual Comparative Map (http://www.animalgenome.org/VCmap/) tool has passed its initial development stage and is at a stable working status serving the community. Application development, improvement, and testing have continued. Online help materials have been added, including a written user manual and a video tutorial. AgBase and the AnimalGenome.org websites provide multiple reciprocal reference links to facilitate resource sharing. Please feel free to try things out and send any feedback to vcmap@animalgenome.org.

**Gene nomenclature standard**

During 2015, the Chicken Gene Nomenclature Committee (CGNC) biocurators worked closely with NCBI Entrez curators to ensure that updated gene annotations had revised nomenclature. This year, we reviewed and updated 2,735 genes (including >100 genes we annotated in conjunction with NCBI Entrez curators). We currently provide standardized nomenclature for 22,172 chicken genes and have a pending grant application to support continued annotation of bird genes.

The initial cattle gene nomenclature is provided by the Bovine Genome Database. Currently we have standardized gene nomenclature for 9,910 Bos taurus genes based upon homology to assigned human gene nomenclature (Fiona McCarthy; http://www.animalgenome.org/genetics\_glossaries/bovgene). We are also working with HGNC to support the development and use of standardized gene nomenclature for livestock species.

**Minimal standards development**

We have continued to work on the MIQAS project to help define minimal standards for publication of QTL and gene association data (http://miqas.sourceforge.net/). The most recent work has been to develop documentation indicating how this was done in Animal QTLdb.

**Expanded Animal QTLdb functionality**

In 2015, a total of 31,976 new QTL have been added to the database. Currently, there are 14,479 curated porcine QTL, 42,019 curated bovine QTL, 5,196 curated chicken QTL, 1,125 curated horse QTL, 1,090 curated sheep QTL, and 127 curated rainbow trout QTL in the database (http://www.animalgenome.org/QTLdb/). All included livestock QTL data have been ported to NCBI, Ensembl, and UCSC genome browser. 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 additions include a genome-wide plot of QTL/association data queried in several ways (see our poster #21315 for details).

**Further development of Animal Trait Correlation Database (CorrDB)**

We have started a developmental process to add public curator/editor tools to the CorrDB to allow continued curation of trait correlation data into the database. Currently the efforts are geared towards making use of resources and tools in the QTLdb for trait ontology development and management, literature management, and bug reporting tools for data quality control. The tools are expected to be released in early 2016.

**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 1033 data files on different animal genomes, supplementary data files to publications, and data for other sharing purposes have been made available to community users. More than 50 data files were shared/transmitted through the online data file-sharing tool by collaborators and/or groups in the community. Our helpdesk is here to assist community members. Throughout 2015, we have helped more than 50 research groups/individuals with their research projects and questions. Our involvement has ranged from data transfer, data assembly, 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, "CRI-MAP users", "Sheep Models", etc.

We have actively maintained and developed the web site for the Functional Annotation of ANimal Genomes (FAANG) project, with new mailing list, user forum, wiki pages, interactive meeting sites, platform for collaborative funding applications, and online publishing capabilities to support coordinated international action to accelerate Genome to Phenome (http://faang.org/). We helped to support the GO-FAANG meeting that took place at the National Academy of Sciences building in Washington DC on October 7-8, 2015.

An increasing number of web hits and data downloads continued in 2015. For example, AnimalGenome.org received over 14.4 million web hits from 624,000 individual sites (visitors), which made 3.8 million data downloads that generated over 3 TB of internet traffic.

**Reaching out**

We have been sending periodic updates to over 2,800 users worldwide to inform them of the news and updated information we develop or host at AnimalGenome.org. New items were updated to the community on an ongoing basis in 2015.

**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.