

SUMMARY OF NRSP-8 ACCOMPLISHMENTS (2013-2017)

Overview of accomplishments for all NRSP-8 technical committees

The most important accomplishment of the NRSP-8 has been the formation of a large community of scientists working worldwide to advance animal genomics through the sharing of resources, development of open-access multi-species bioinformatic tools, sequencing and assembly of genomes, organization of workshops and conferences, communication of results, support for travel for students and invited speakers, preparation of multi-institutional grant proposals, and formation of large collaborative research groups. The communication and sharing of information among the different species technical committees fostered by NRSP-8 has led to significant achievements under each of the three objectives outlined for 2013-2017. Across committees, the experience of one group has often informed and influenced the directions and approaches taken by other groups and this shared knowledge has accelerated tool development and discovery for all supported species. A summary of the important accomplishments and impacts for each of the technical committees (aquaculture, cattle, horse, poultry, sheep/goat, swine and bioinformatics) are included below. Here we highlight a few of the accomplishments shared across multiple species for each of the three 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. Between 2013 and 2017, reference genomes were assembled for the pig, turkey, sheep, goat, catfish pacific oyster, rainbow trout and striped bass. In addition, genome reference assemblies were improved for the chicken, cow, horse, pig and rainbow trout with researchers capitalizing on short-read sequencing technologies, optical mapping, Pacific Biosciences sequencing and other technologies. Across all species, these improved assemblies reached high-quality chromosome levels, eliminated most of the regions with ambiguous sequences, and in some cases provided sequence for previously unsequenced autosomes. Genome annotation and gene predictions were enhanced in several species using a variety of methods including RNA sequencing (RNA-Seq) of protein coding RNA, micro RNA (miRNA), and long non-coding RNA (lncRNA), full-length transcript sequencing using Iso-seq, and coordinated efforts to manually annotate genes.

Efforts were initiated in the cow, pig, chicken, horse, sheep, rainbow trout and pacific oyster to annotate additional functional elements of the genome as part of a new initiative, the Functional Annotation of ANimal Genomes (FAANG) consortium. The FAANG consortium was formed in 2014 with the goal of accelerating genome-to-phenome discovery in NRSP-8 species. In the first phases of this effort, a number of investigations have been proposed or initiated across 80-105 tissues, depending on the species. These include whole genome sequencing; whole genome bisulfite sequencing; RNA sequencing (mRNA, miRNA, ncRNA) and transcriptome assembly; ATAC-seq; ChIP-seq with DNase I, histone modification marks, insulator-binding protein CCCTC-binding factor, and important transcription factors; and the study of the genome-wide chromatin interactome using Hi-C. Is it worth noting that this is the first time some of these technologies have been applied to some of these species. Work is ongoing among members of the FAANG project to standardize collection techniques, experimental protocols, and data analysis pipelines to maximize the utility of the data produced by this effort.

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. From 2013 to 2017, single nucleotide polymorphism (SNP) high-throughput genotyping arrays were developed for several species including Equine (54K, 65K, 670K and 2M arrays), chicken (670K), cattle (250K functional allele array), sheep (600K), goat (52K), swine (670K) and rainbow trout (57K, 50K functional allele arrays). For all species, the impacts of these SNP genotyping arrays include: permitting genome-wide analyses such as genome-wide association studies (GWAS) and genomic signatures of selection for identification of genomic regions harboring alleles for traits of interest; allowing for, and improving the accuracy of predicted breeding values; enabling genomic selection; and permitting estimation of genetic diversity in breeds and populations of interest. Across species tools developed under this objective have allowed for identification of alleles responsible for important economic and disease traits, including alleles important in infectious diseases such as GBP5 associated with resistance/susceptibility to primary PRRS virus infection in pigs. In addition, the dairy industry has used SNP-chips to genotype nearly over one million dairy cattle allowing application of genomic selection which has reduced

animal selection generation interval (from 5 years to less than one year) and has increased genetic merit prediction accuracy by more than 30 percent with an estimated annual benefits of \$100 million per 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. Successful efforts have been made to develop platforms to facilitate collaborative research for collection and analysis of new, unique, and interesting phenotypes, and to develop, integrate, and implement bioinformatic resources to support the discovery of genetic mechanisms underlying agriculturally important traits. For example, the Animal Quantitative Trait Loci database (Animal QTLdb) was updated with 104,272 new quantitative trait loci (QTL). To date, the database contains 95,332 cattle, 6,633 chicken, 1,245 horse, 16,516 pig, 1,412 sheep and 127 rainbow trout loci that have been associated with many traits of interest. Further, 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 for the community with 1,140 data files, totaling 140 GBGb, shared through this platform. Finally, a collaborative VCF information-mining platform was developed to allow for sharing discovered genetic variants between researchers.

In addition to direct contributions to each of the three objectives, NRSP-8 participants have leveraged the NRSP-8 investment in tools and infrastructure into at least \$94.5 million dollars in funding to study diverse animal models to investigate fundamental mechanisms of genome biology and physiology and pathophysiology affecting production efficiency, product quality, animal health, disease resistance and food safety and to develop additional bioinformatics resources (see **Table 1**). Finally, the annual NRSP-8 workshops have become an essential component for the development of collaborations, training and dissemination of new information to government, academic, and industry stakeholders in animal agriculture. NRSP-8 species coordinators' funds have been used to support travel for 146 postdoctoral and graduate students to the NRSP-8 meetings that are held in conjunction with the annual Plant and Animal Genome (PAG) meeting.

	Federal	Private/Industry	Intramural	Total
Horse	\$14,605,017	\$4,799,843	\$3,440,344	\$22,845,204
Poultry	\$18,675,963	\$150,000		\$18,825,963
Cattle	\$27,831,461	\$73,000		\$27,904,461
Swine	\$8,229,905	\$621,591	\$533,500	\$9,384,996
Aquaculture	\$5,553,103			\$5,553,103
Sheep/goat	\$3,644,000	\$109,000	\$467,000	\$4,220,000
Bioinformatics /multispecies	\$5,753,033			\$5,753,033
Total	\$84,292,482	\$5,753,434	\$4,440,844	\$94,486,760

AQUACULTURE

<http://www.animalgenome.org/aquaculture/>

Direct contributions to Objective 1:

- Reference genome for catfish (2016). **Impact(s):** The genome reference will allow understanding the genes controlling performance traits. Technologies can be developed based on this information allowing superior catfish breeds that will help farmers increase profits.
- Rainbow trout high-density 57K SNP chip was developed and characterized (2013). Approximately 50K of the SNPs were validated in a panel of 18 rainbow trout populations at the standard 97% call rate of the Affymetrix SNP polisher software. **Impact(s):** The SNP chip allowed improved accuracy of predicting breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture.
- Reference genome for the Pacific oyster (2012). **Impact(s):** This genome provides a basis for numerous phenotype studies and provides insight into performance under changing environmental conditions.
- Striped Bass Genetic Map (2012). The first genetic map of the genome of the striped bass was developed and published. **Impact(s):** This medium-density linkage map was based on 298 microsatellite markers and is enabling detection of QTL affecting production traits.
- Rainbow Trout Reference Genome sequence (2012): A pooling and tagging scheme was used for sequencing of ~15,000 clones from the BAC fingerprinted physical map minimal tiling path (MTP). **Impact(s):** The map helped in assembling the trout genome.
- Improved Rainbow Trout Reference Genome sequence (2017): The longest available read length of the Illumina technology was used to improve the genome sequence producing longer and better anchored scaffolds to chromosomes. **Impact(s):** The genome assembly led to SNP genotyping tools that are being used to accelerate genetic improvement.
- Striped bass genome sequence assembly containing ~35 K scaffolds was produced (2015). **Impact(s):** The assembly should accelerate analysis of the striped bass genome, to identify and characterize genes affecting important production traits.

Direct contributions to Objective 2:

- A 675K SNPs array was developed for catfish (2017). **Impact(s):** This array allowed for genetic mapping and validation of the reference genome sequence assembly as well as for identification of a genetic markers associated with aquaculture production traits in catfish.
- A 57K SNPs array was developed for rainbow trout (2014). **Impact(s):** This array allowed for genetic mapping and improving assembly of the reference genome and evaluation genomic selection in rainbow trout.
- A 50K cSNPs array was developed for rainbow trout (2016). **Impact(s):** This array allowed for allelic-imbalance analysis of genes that are associated with muscle yield and fillet quality traits and also with bacterial cold-water disease survivability.
- Bulk segregant RNA-seq (BSR-Seq) was used to analyze differentially expressed genes and associated SNPs with disease resistance against enteric septicemia of catfish (ESC) (2013). A total of 1,255 differentially expressed genes were found between resistant and susceptible fish. **Impact(s):** These genes are candidates for further functional genomics work to validate their role in providing catfish with susceptibility to ESC.
- QTL mapping families for stress response and bacterial cold-water resistance (BCWD) in rainbow trout (2013). **Impact(s):** The families are being used to study genes responsible for stress response and BCWD.

- Illumina GoldenGate genotyping arrays were designed for *Crassostrea gigas* and *Ostrea edulis* (2014). **Impact(s):** These assays were used to genotype 1,000 individuals from wild and selected populations as well as families bred for commercially important traits.
- Large intergenic noncoding RNAs (lncRNAs) were identified by RNA-Seq analysis of rainbow trout transcriptome (2016). **Impact(s):** Many of the lncRNAs are tissue-specific and functionally associated with important biological processes including resistance to the rainbow trout BCWD and muscle growth.
- RNA-Seq analysis of miRNAs associated with different production quality traits in trout (2015 and 2017). **Impact(s):** Several miRNAs with epigenetic role associated with egg quality and muscle quality traits were identified.

Direct contributions to Objective 3:

- Rainbow trout QTL database (2012) available through the Animal Genome website of the NRSP-8 bioinformatics group (<http://www.animalgenome.org/cgi-bin/QTLdb/index>) and is being continually updated. **Impact(s):** QTLs are available for industry to improve aquaculture production traits in rainbow trout.
- Bioinformatics pipeline was developed for genotyping SNPs from raw sequence data for the GT-seq method (2014). **Impact(s):** The pipeline provides significant cost reduction for genotyping.
- *C. gigas* transcriptome information derived from 2.2 billion sequences from 114 RNA-seq datasets has been organized and deposited into a publicly available database: GigaTON (2015). **Impact(s):** The user interface provides powerful and user-friendly tools to search and retrieve annotation, expression, and polymorphism information of important genes related to aquaculture traits.

Communication:

- A strategic planning workshop for aquaculture genomics, genetics and breeding was held at Auburn University (2016). **Impact(s):** The workshop led to a white paper published in BMC Genomics that placed goals and priorities for future research in the aquaculture genomics, genetics and breeding in the US.
- NRSP-8 Aquaculture leaders participated in establishing the FAASG (Functional Annotation of All Salmonid Genomes) consortium. **Impact(s):** The consortium will allow coordinating data sharing and establish an infrastructure for providing high quality functional annotation of salmonid genomes.

Research support mini-grants (coordinator grants):

- Approximately 25 mini-grants (~\$10,000/each) supported projects that fall under all three primary objectives and include a variety of species.

Travel support and opportunities for trainings:

- Travel of 25 students/postdocs was funded to attend the Aquaculture workshop at PAG meetings (2012-2016). The purpose of the travel award program is to help graduate students and postdocs to travel to the annual PAG meeting to present their research.

Leveraged funds and stakeholders' use of project outputs

Leveraged funds from diverse projects totaling more than \$5,553,103 from federal sources. Selected grants are highlighted below.

- Whole genome mapping of disease resistance/susceptibility-associated SNPs in catfish. USDA National Institute of Food and Agriculture Competitive Grant no. 2015-67015-22975. **\$500,000**. John Liu (PD).

This project is designed to address the following two objectives: 1). Genome wide scan of QTLs conferring resistance to ESC and columnaris using F2 and F4 fish using the 250K catfish SNP array; and 2). Fine QTL analysis by genotyping a large number of F2 and F4 individuals using evenly-spaced markers from mapped QTL regions. **Impact(s):** The impact of this grant will be to determine genes that will be used to select for fish resistant catfish

- Closing the tilapia genome assembly. USDA National Institute of Food and Agriculture Competitive Grant no. 67015-23088. **\$270,000**. Tom Kocher (PD).
This project aims to improve the platform for genetic improvement of tilapia by developing a definitive sequence of the tilapia genome. **Impact(s):** The ultimate goal of the project is to improve the health and production of tilapia and related aquaculture species.
- Sequencing the Genome of the Eastern Oyster. USDA National Institute of Food and Agriculture Competitive Grant no. 2015-67016-22942. **\$242,051** Marta Gonez-Chiarri (PD).
This project propose to develop these key resources and tools by performing the sequencing, assembly, and annotation of a reference genome and transcriptome for the Eastern oyster *C. virginica*. Genome researchers and bioinformatics experts, in collaboration with the Eastern Oyster Genome Consortium, will use state-of-the-art sequencing and assembly strategies to achieve these aims. **Impact(s):** The reference genome and transcriptomes for the Eastern oysters will aid the research community in the discovery of candidate genes and markers associated with traits of commercial, biological, and ecological importance in oysters.
- Development of 675K SNP arrays for whole genome mapping and genetic studies in catfish. USDA National Institute of Food and Agriculture Competitive Grant no. 2015-67015-22907. **\$485,000** John Liu (PD).
This project addresses major limitations to adopting genome technologies in aquaculture that currently are the lack of a high- throughput technology for the analysis of genomic variations in relation to phenotypic variations. We need a high-density SNP array technology that allows high- efficiency, cost-effective, whole-genome coverage, analysis of genetics of important performance traits such as disease resistance. This project is poised to resolve these challenges, with three specific objectives: 1) Developing the catfish 675K SNP arrays; 2). Genetic mapping of whole genomic sequence scaffolds; 3). Enhancing and validating the catfish whole genome assembly. **Impact(s):** This project will address the most significant problem currently existing in catfish genomics. This project will develop a technology for the most efficient analysis of performance traits, and will literally transform the isolated whole genome sequence tags into a well-assembled reference genome assembly, thereby enabling its application in breeding and selection programs.
- Homozygous clonal rainbow trout lines as genomic resources. USDA National Institute of Food and Agriculture Competitive Grant no. 2016-67015-24472. **\$485,000**. Gary Thorgaard (PD).
Doubled haploid lines have unique value as genomic tools because they have minimal heterozygosity and allow full chromosomal haplotypes to be identified. These lines have been used for the rainbow trout genome sequencing project and for QTL studies. The experimental objectives will include: (1) Establish at least 12 lines within the USDA. (2) Transfer cryopreserved semen from each line as an ongoing repository. (3) Generate a repository of frozen tissues and genomic DNA. (4) Conduct baseline karyotype analysis and SNP typing by re-sequencing of two of the lines. (5) Attempt to induce sex reversal to females in the YY lines and test their fertility. **Impact(s):** The project will assure continued availability of the lines to the trout research community, develop sperm, tissue and DNA repositories and improve propagation methods.
- SNP markers for muscle, growth and fillet quality traits in rainbow trout. USDA National Institute of Food and Agriculture Competitive Grant no. 2014-67015-21602. **\$500,000**. Mohamed Salem (PD).
The project aims to find genes and to develop genetic markers that would be used in USDA marker-assisted selection programs to develop food fish strains with superior muscle growth and fillet quality in rainbow trout. This project is expected to produce a large number of true genetic markers that provide a valuable resource for determination of genetic merit of growth and carcass traits in rainbow trout. Project outcomes including, knowledge, expertise, methods, tools, and technologies, will be disseminated to the US aquaculture industry/stakeholders including the US largest producers of rainbow trout food fish and eggs. **Impact(s):** These

genetic markers may be quickly adapted to other species and give the US aquaculture industry a competitive advantage.

Major impact products (could be potential impact):

Recently, the first genome-wide SNP arrays have been developed and applications of the SNP-chip technology in genomic selection have just begun to be adapted by commercial breeders of some aquaculture species including catfish (600K), salmon (286K), rainbow trout (57K) and oyster (190K).

CATTLE

<http://www.animalgenome.org/cattle/>

Direct contributions to Objective 1:

- An improved bovine genome reference sequence assembly of Dominette (the reference animal) based on multiple data types developed by the bovine community (optical map, Illumina paired-end, PacBio sequence, and improved gene predictions based on RNA-Seq and Iso-Seq data) will be released in Fall of 2017. **Impact(s):** This improved assembly will help with the identification of genetic markers associated with economically important traits in cattle.

Direct contributions to Objective 2:

- A 250K functional variant assay was made available to cattle researchers. The assay was designed using various sources of sequence data derived from AFRI-funding and is focused on the detection of genic variants likely to be functional in taurine cattle. **Impact(s):** This tool will assist researchers to identify genetic identifying causative SNPs that are associated with economically important traits and which are likely to be useful in marker-assisted selection across multiple breeds.

Direct contributions to Objective 3:

- A database (Animal-GRIN) has been developed to serve as a permanent archive for DNA data, germplasm/tissue samples, and phenotypic and production system data from large animal genomics projects. **Impact(s):** This database will allow for future data mining and value capture from the data and samples collected by publicly funded research.

Communication:

- A bovine genome newsletter was prepared by the bovine coordinators and distributed to the AnGenMap listserv. **Impact(s):** This helped inform the bovine research community of ongoing developments with the bovine genome.
- Two industry conferences were held, the “2015 Applied Reproductive Strategies in Beef Cattle (ARSBC) Conference Grant”, and “New Approaches to Bovine Respiratory Disease Prevention, Management, and Diagnosis” with support from USDA, National Institute of Food and Agriculture Conference Grants (2014-67015-21562; 2015-67015-23693). Proceedings were produced from both conferences, and the latter was published in Animal Health Reviews.

Research support mini-grants (coordinator grants):

- Livestock FAANG project (lead by H. Zhou, P. Ross and I Korf) Coordinator funds (\$30,000). [The project allowed for sample collection from 4 individuals (2 males and 2 females). These funds were used as leverage that resulted in ~\$500,000 grant from the USDA NIFA and another ~\$100,000 from National Pork Board, Aviagen etc. for FAANG data collection on these samples.

Leveraged funds and stakeholders’ use of project outputs:

From 2013-2017, the investigators and stakeholders leveraged the tools and resources generated through NRSP-8 to obtain at least \$27,904,461 additional funding from federal sources, in funding from private foundations and industry sources. Selected grants are highlighted below.

- Integrated program for reducing bovine respiratory disease complex in beef and dairy cattle. USDA National Institute of Food and Agriculture Competitive Grant no. 2011-68004-30367. **\$9,750,000**. Jim Womack (PD). The objective of this Coordinated Agricultural Project was to use genomic tools to identify genetic markers associated decreased susceptibility to bovine respiratory disease. This is the most important disease in both the beef and dairy cattle industry with estimated losses of more than one billion dollars annually. This project used genome sequences to fine map genetic variants associated with respiratory disease, with the aim of delivering a tool that the industry can use to select for cattle that are less susceptible to respiratory disease. **Impact(s):** The impact of this grant will be decreased morbidity and antibiotic use in cattle production, and improved animal health and welfare.
- 2015-2017: USDA NIFA 2015-67015-23183. “Application of a functional variant assay and sequence imputation to identify large-effect QTL underlying feed efficiency and component traits in beef cattle.” Taylor JF, RD Schnabel, JE Decker, CS Seabury and HL Neibergs. 4/1/15-03/31/17. **\$500,000**. This grant supported the development of the GGP-F250 functional assay. The accomplishment is that we successfully designed an assay for which 173,609 variants can be assayed with a marker call rate of at least 90%. These variants are highly enriched for rare functional variation within the bovine genome and include 82,979 variants that alter amino acids within gene products, 665 Indels that either alter frame or add/delete amino acids, 2017 splice site variants and 44,358 variants within untranslated regions. The assay is currently publicly available through GeneSeek. **Impact(s):** Impacts of this grant include 23,541 variants within QTL regions detected in the BRD and Feed Efficiency grant that were identified and included on the assay and 1978 BRD case-control and 4609 Feed Efficiency project animals have been genotyped with the assay to fine-map QTL. The assay also contains 2,224 variants for which no homozygotes were detected. These are currently being mapped to genes known to be essential for life to identify candidates for lethal alleles segregating in cattle. Finally, the assay is expected to aid in the process of imputing genotypes to whole genome sequence, because, contrary to the currently used assays which are strongly enriched for common variants, the GGP-F250 is enriched for rare variants and the linkage disequilibrium that exists among rare variants will aid in the imputation of genotypes for this class of variant.
- 2013-2017: USDA-NIFA-AFRI. 2013-68004-20364. “Identification and management of alleles impairing heifer fertility while optimizing genetic gain in Angus cattle.” Patterson DJ, JF Taylor, A Van Eenennaam, S Brown and M Smith. **\$2,997,040**. This grant supported the whole genome sequencing of the 100 registered Angus bulls that have had the greatest impact on the breed as determined by the number of registered descendants. These animals, along with sequences obtained on 162 additional animals from 12 other taurine breeds, were used to identify variants genome wide. With support from the three other USDA grants (Bovine Respiratory Disease, Feed Efficiency and Functional Variant), we designed the GGP-F250 assay for which 173,609 variants can be assayed with a marker call rate of at least 90%. These variants are highly enriched for rare functional variation within the bovine genome and include 82,979 variants that alter amino acids within gene products, 665 Indels that either alter frame or add/delete amino acids, 2017 splice site variants and 44,358 variants within untranslated regions. **Impact(s):** The assay is currently publicly available through GeneSeek allowing genetic gain assessment for important production phenotypes in Angus cattle.
- 2016-2019: NIH 1R01HD084353. “Linking Fertility-Associated Gene Polymorphisms to Aberrant Sperm Phenotypes.” Sutovsky P, RD Schnabel, JF Taylor. 7/1/2016-6/30/21. **\$2,149,000**. This grant has just started but plans to sequence 100 bulls with either sperm abnormalities or with extreme differences for conception rate to identify mutations in genes known to be expressed in sperm that are responsible for the defects and variants that are candidates for genetic variation in male fertility. We have begun the collection of sperm samples from US and Canadian AI companies. **Impact(s):** The project is expected to identify and validate sperm phenotype biomarkers encoded by fertility associated polymorphic genes, and to improve sire management by genetic selection and automated semen evaluation. This project will also yield new methods and potentially new treatments for human male and idiopathic infertility.
- Gene Seek and Zoetis provided industry funds and support to leverage the cost of developing the new bovine genome assembly, **\$73,000**.

Travel support and opportunities for trainings:

- Funding was used to bring students to the annual PAG meeting based on a competitive travel award. Coordinator funds were also used on several occasions to help support the NRSP8 speaker at PAG.

Major impact products (could be potential impact):

- Genomic selection has dramatically improved the rate of genetic progress within the US dairy industry. The dairy industry has used SNP-chips to genotype over 1 million dairy cattle. Application of GS reduced animal selection generation interval (from 5 years to less than one year) and has increased prediction accuracy by more than 30 percent for an estimated annual benefits of \$100 million per year.
- Genomic selection is starting to be implemented in the US beef industry.
- Development of a 173,609 SNP functional variant assay containing variants highly enriched for rare functional variation within the bovine genome and including 82,979 variants that alter amino acids within gene products, 665 Indels that either alter frame or add/delete amino acids, and 2017 splice site variants. The assay is currently publicly available through GeneSeek.

HORSE

<http://www.uky.edu/Ag/Horsemap/>

Direct contributions to Objective 1:

- A new reference genome build (EqCab 3.0) was created for the horse and shared among workshop participants. Public release and publication is expected in late 2017. Morris Animal Foundation, NRSP-8 coordinator and other federal funds supported this work. **Impact(s):** The new assembly improved gene annotation, increased contig N50 from 112 Kb to 1.4 Mb, and eliminated most of the regions with ambiguous sequence (“Ns”). The improved reference will increase the power and efficacy of genomics research to discover the genes and alleles underlying disease and economically important performance traits in the horse.
- The annotation of the horse genome was improved through investigations of gene expression and splice variation that occurs among diverse tissues. Data supporting wide- scale annotations of the horse genome were published in several reports (2013-2016). **Impact(s):** Improved annotation provides context for the discoveries by making it possible to identify the functional aspects of genetic variation.
- SNP and insertion-deletion polymorphism discovery was performed using whole genome sequence from 153 horses as part of an effort to design 2M and 670K SNP Affymetrix SNP arrays. **Impact(s):** This work documents the extent of variation that exists among 24 horse breeds and made genotypes from 485 horses across 2M SNPs publically available providing raw material for use in developing research tools. (Schaefer RJ, *et al.* Developing a 670k genotyping array to tag~ 2M SNPs across 24 horse breeds. *BMC Genomics* 18.1 (2017): 565).

Direct contributions to Objective 2:

- An assay tool to assay ~65K SNPs (SNP70) was developed to replace the ~54K SNP (SNP50) tool in 2013. The development of this tool was a collaborative activity of the NRSP-8 community and made publicly available. An imputation pipeline between these two moderate-density arrays was developed. (McCoy AM, McCue ME. Validation of imputation between equine genotyping arrays. *Animal Genetics* 45:153, 2014.

PMCID: PMC4000747.) **Impact(s):** Developing this tool and imputation pipeline made it possible to continue to perform genome-wide analyses that impact the health and welfare of horses.

- SNP discovery based on whole genome sequence from 153 horses was used to construct the next generation 2M and 670K SNP Affymetrix SNP arrays for equine whole genome analyses. The 670K array is designed for imputation and enables data from prior lower density SNP arrays to be imputed up to ~1.8M SNPs. The equine 670K SNP chip was made available in 2015. (Schaefer RJ, *et al.* Developing a 670K genotyping array to tag~2M SNPs across 24 horse breeds. *BMC Genomics* 18.1 (2017): 565.). **Impact(s):** This 670K array and imputation resource improves genome coverage more than 30-fold over the medium density (54K and 65K) SNP arrays. More than 20,000 670K genotyping arrays have been used to date. This increase in SNP density will allow for GWAS in genetic diverse breeds of economic importance such as the American Quarter Horse (~4 million registered individuals).
- Because of the closing of the commercial operation of the BAC library, 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. **Impact(s):** This will ensure continued access to the library for equine researchers. This resource is key for investigating the broader aspects of structure and organization of the horse genome.

Direct contributions to Objective 3:

- Horse technical committee members joined the FAANG initiative to generate gene expression data for 38 of tissues from two horses. In connection, competitive, extramural industry funding was obtained to further develop this dataset. **Impact(s):** This resource will empower research in the area of functional genomics.
- The horse genomics community actively utilized the collaborative resources provided in the AnimalGenome.org Data Repository. The site hosts large shared files, prepublication works and polymorphism data.
- Horse specific transcriptome assemblies not yet curated by NCBI were made available at AnimalGenome.org and through GitHub (https://github.com/drtamermansour/horse_trans). **Impact(s):** This resource increases the publically availability of equine transcriptional data and will improve genome annotation.
- With the assistance of horse genome researchers, the AnimalQTL database added horse to the species list. **Impact(s):** This resource provides rapid access to 1,245 equine QTL and associated metadata.
- Horse genome workshop members deposited 1,524 genomic SRA archives for the horse. These accessions contain many fully re-sequenced genomes, as well as targeted datasets generated by diverse NGS platforms. **Impact(s):** This resource increases the publically availability of equine whole genome sequence and transcriptomic data.

Communication:

- Additional workshops were conducted for NRSP-8 participants in connection with the International Society of Animal Genetics ([ISAG] 2016 [Salt Lake City], ISAG 2014 [Xian, China], ISAG 2012 [Cairns, Australia], and ISAG 2017 [Dublin, Ireland]). **Impact(s):** These meetings facilitated communication and collaborations among international scientists working on all species and extended discussions conducted at the annual NRSP-8 workshops.
- Additional workshops were conducted with support of the Dorothy Russell Havemeyer Foundation that focused on issues related to horse genomics (2013 [Azores, Portugal], 2015 [Hannover, Germany], 2018 [planned, Pavia, Italy]). **Impact(s):** These workshops include the entire international horse genomics research community and facilitate exchange of information and collaboration between scientists.
- Following the identification of critical needs in coordinating collaborations across institutions for new and evolving projects, an initiative to provide a database of ongoing work is now hosted through collaboration

with the Interbull.org service. **Impact(s):** This database currently provides a listing of projects recruiting samples, but may eventually expand to include file sharing for exchange of SNP and NGS datasets.

Research support mini-grants (coordinator grants):

- Matching funds provided to support development of the SNP70 SNP genotyping array (~65K SNPs) for discovery research on the genomics of horses.
- Matching Funds provided to support development of the 670K SNP genotyping array. Primary funding from USDA-NIFA (Dr. Molly McCue PI) along with coordinator funds were used to develop a 2M test array. 670K SNPs were selected to tag ~1.8M SNPs across 24 horse breeds.
- Matching funds provided for EqCab 3.0. Primary funding from the Morris Animal Foundation. Improved predictions from assembly.
- Matching funds provided to develop FAANG resources for horse; primary funding came from Grayson-Jockey Club Research foundation project Developing resource for functional genomics research.

Leveraged funds and stakeholders' use of project outputs:

From 2012-2017, the equine investigators leveraged the tools and resources generated through NRSP-8 to obtain **\$22,845,204** in additional funding. This included **\$14,605,017** in funding from federal sources, **\$4,799,843** in funding from private foundations and industry sources and **\$3,440,344** in intramural funding. Selected grants are highlighted below.

- **“Genetic diversity and selection in the domestic horse.”** Dr. Molly McCue PI, Dr. James Mickelson Co-I, and others **\$499,481** USDA-AFRI. **Impact(s):** This proposal quantified genetic diversity and to identify functional alleles that cause variation in size, locomotion and athletic phenotypes among 36 domestic horse breeds.
- **“Tools to Link Genotype to Phenotype in the Horse.”** Dr. Molly McCue PI, Dr. James Mickelson Co-I, and others. **\$499,727** USDA-NIFA. In this proposal builds upon the recent development of high-density SNP arrays to develop tools that further facilitate GWAS in the horse by: **1)** enabling complementary GWAS approaches including gene, haplotype, and pathway-based analyses through SNP-to-gene mapping and the construction of a haplotype map; **2)** increasing marker density by developing an imputation resource; and **3)** constructing context-specific co-expression networks for integrated network-based association analysis. Prioritization of candidate genes is assisted by: **4)** refining the physical annotation of mRNAs, lncRNAs, and miRNAs; and **5)** improving functional annotation of these loci through tissue-specific gene expression and gene co-expression networks. Finally, the identification of functional alleles will be accelerated by **6)** developing a comprehensive catalog of genetic variants from WGS of >450 horses.
- **“Functional Prioritization of Candidate Genes and Alleles for Equine Metabolic Syndrome.”** Dr. Molly McCue PI, Dr. James Mickelson Co-I. **\$499,815** USDA-NIFA. Genome wide association in Welsh Ponies (WP) and Morgan horses has identified >180 chromosomal regions of interest (ROI) harboring >3,000 positional candidate genes associated with Equine Metabolic Syndrome (EMS) phenotypes. The objectives of this proposal are to **1)** prioritize candidate genes using skeletal muscle and/or adipose tissue gene expression or alterations in serum metabolite abundance to support their role in EMS pathophysiology; and **2)** identify the functional alleles underlying EMS phenotypes.
- **“Discovering Causal Variants for Complex Disease Using Functional Networks in the Horse.”** Dr. Rob Schaefer PI, Dr. Molly McCue mentor. **\$150,000** USDA-NIFA. The goals of this grant are to develop software tools to integrate available sources of genomic data and functional data (WGS, SNP, RNA-sequencing, proteomics and metabolomics) in agricultural species to better understand complex phenotypic traits using metabolic syndrome in the domestic horse as a test case.

- “**Protein Networks Mediating Airway Hyper-Responsiveness In Equine Airways.**” Dr. Chipper Swiderski **\$438,153** USDA-AFRI. This grant seeks to better understand the etiopathogenesis of Recurrent Airway Obstruction in the horse through proteomic studies and improved annotation of genes expressed during disease exacerbation.
- “**Comparative Genomics in Qatar.**” Dr. Doug Antczak and Dr. Samantha Brooks, **\$1,030,000** Qatar National Research Foundation- National Priorities Research Program. This project will document variation and signatures of selection in desert breeds of horse, as part of a larger effort to improve genomic resources in desert adapted hoof stock.
- “**Identification of Genetic Factors Responsible for Establishment of Equine Arteritis Virus Carrier State in Stallions.**” Dr. Uri Balasuriya PI, Dr. Ernie Bailey Co-PI and others. **\$2,930,000** USDA-AFRI.

Travel support and opportunities for training:

- Travel of 64 students/postdocs was funded to attend the Equine workshop at PAG meetings (2012-2016). The purpose of the travel award program is to help graduate students and postdocs to travel to the meeting to present their research.
- Support for five NRSP-8 members to attend GO-FAANG workshop in Washington DC to provide leadership horse group in connection with this initiative.
- Member sent to participate in Hack-a-thon in Europe 2016 in support of FAANG activities. Integration with international efforts to develop functional genomics databases for animal genomics.

Major impact products (could be potential impact):

- Development of 4 SNP genotyping arrays (54K, 65K, 670K and 2M). **Impact(s):** These arrays allow for efficient and economic performance of dozens of genome-wide analyses in the horse.
- Genomic diagnostics in the horse have now expanded to include commercially available tests over 100 markers contributing to more than 40 diverse traits. **Impact(s):** Costs per test are falling, and as adoption of genomic selection and mandatory genetic testing increases across the industry, translating in to reduced economic losses due to genetic disease.
- Diagnostic tests created for markers related to performance, disease and color, including *DMRT3* and gait, *TBX3* and dun color, *SHOX* and dwarfism, *BAGALT7* and dwarfism, *ACAN* and dwarfism, *RFWD3* and Appaloosa color pattern, *SERPINB11* and hoof quality, *KIT* and spotting in donkeys, *HOXD3* and occipitoalantoaxial malformation, *CXCL16* and susceptibility to equine arteritis virus.
- Additional targets for investigation were identified through genome-wide analyses including signatures of selection in 38 horse breeds, genomic loci contributing to osteochondrosis, recurrent laryngeal neuropathy and others.
- Molecular tests to identify chromosome abnormalities were reported and additional test are being developed. **Impact(s):** Chromosome abnormalities are the most common genetic cause of infertility and disease amongst horses and molecular tests are much less expensive than conventional karyotyping.
- The major histocompatibility complex plays a major role in the occurrence and consequences of allergic and infectious diseases. Determinants playing a role in specific diseases were identified and methods were developed to improve our ability to identify yet other MHC determinants. **Impact(s):** MHC is a genetically complex region but plays a major role in immune responses. Knowledge of the MHC remains incomplete for all species and but research is turning up applications, especially with respect to vaccine design and immune therapy.

POULTRY

<https://www.animalgenome.org/poultry/>

Direct contributions to Objective 1:

- The chicken genome build (Gallus_gallus-5.0) was released to the public in 2015. **Impact(s):** This improved build, which was aided by long single molecular sequencing and finished BACs, yielded a gain of 180 Mb in assembled bases and provided coverage to 3 previously missing autosomes. As the reference genome, this invaluable resource greatly enhances the ability to identify genes and genetic variations associated with traits of agronomic interest.
- A turkey draft genome was generated from next generation sequencing and a turkey BAC contig (physical) map.
- Guidelines for standardized gene nomenclature for chicken genes were developed to assign nomenclature to (1) MHC genes; (2) genes highly expressed in egg white, yolk and eggshell; (3) histone; and (4) myosin genes. **Impact(s):** This nomenclature was shared with NCBI and Ensembl.

Direct contributions to Objective 2:

- Very high-density SNP mapping (ca. 600K SNP) panels have been developed and along with 60K SNP chips. **Impact(s):** These genotyping arrays are being employed in genome-wide association studies (GWAS) and genomic selection (GS).
- Efforts have been initiated to annotate the chicken genome, especially with respect to regulatory elements. In brief, datasets for transcripts, histone marks, methylation and more have been integrated to identify promoters, enhancers, and silencers. **Impact(s):** This information is vital to help connect genotypic variation to phenotypic variation.
- Transcript and comparative genome hybridization arrays were developed and distributed.

Direct contributions to Objective 3:

- Over 40 unique chicken research lines and their derived materials have been shared with amongst investigators to expand studies on the chicken genome.
- 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.

Communication:

- Provided support for members to attend GO-FAANG meeting and/or other multi-state research project meetings to enhance communications of activities, communicate about resources.

Research support mini-grants (coordinator grants):

- Provided \$30,000 in funds towards the USDA AFRI Animal ENCODE proposal; Huaijun Zhou, UC Davis – PI.
- Financial support provided to W. Warren, Washington U., St Louis, for sequencing of microchromosomes, which has aided to fill in gaps in the genome assembly.
- Financial support provided to M. Delany, UC Davis, to create a capture arrays and sequence the wg-2 mutation in the Wingless-2.331 congenic inbred line.

- Financial support provided to H. Zhou, UC Davis, for challenge experiments involving highly pathogenic Newcastle Disease Virus (NDV) and the Fayoumi and Leghorn strains in order to characterize genetic resistance.
- Financial support provided to B. Muir, Purdue U. to create a synthetic resource population using 8 diverse genetic lines to fine map genetic resistance to Marek's disease.
- Financial support provided to M. Miller, City of Hope, for further sequencing of microchromosome 16 BAC clones to improve understanding of the MHC/Rfp-Y complex.

Leveraged funds and stakeholders' use of project outputs:

From 2013-2017, the investigators and stakeholders leveraged the tools and resources generated through NRSP-8 to obtain at least \$18,675,963 additional funding from federal sources and \$150,000 in funding from industry sources (Cobb Vantress supported efforts towards the improvement of the chicken genome assembly). Selected grants are highlighted below:

- 2013-2017: USDA NIFA 2013-67015-21357. "Improving the chicken genome assembly and annotation." Warren W, CT Brown, H Cheng H, and J Dodgson. **\$485,690**. This grant supported the improvement of the chicken genome assembly and annotation by filling in known gaps within and between existing scaffolds, and refining microchromosome linkage maps for localization of unplaced sequences. **Impact(s):** With the biology becoming reliant on a genome assembly, the higher quality chicken assembly aided all efforts especially with respect to identifying genes and pathways of agronomic importance. Furthermore, other avian genomes were improved as they also rely on the chicken genome assembly as their reference as well.
- 2012-2017. USDA NIFA 2012-67015-19419. "Enhancing genetic resistance to Marek's disease in chicken via allele-specific expression screens and genome-wide selection." This grant supported the identification of genes and genetic markers associated with resistance to Marek's disease (MD), a herpesvirus-induced lymphoma of chickens. Hypothesizing that differences in gene expression (when, where, and how much) are the major contributors of phenotypic variation for complex traits such as disease resistance, SNPs that exhibit allele-specific expression (ASE) in response to Marek's disease virus infection were identified. These ASE SNPs were found to account for over 83% of the genetic variance and were 125% more accurate in genomic selection compared to pedigree evaluation (i.e., BLUP). **Impact(s):** These results support the hypothesis that phenotypic variation in traits is primarily due to changes in regulation of gene expression rather than other sources such as differences in protein composition. Furthermore, we have identified most of the genes that confer MD genetic resistance, which should help reduce the ~\$1-2 billion in annual losses associated with MD.
- 2013-2018. USAID AID-OAA-A-13-00080 "Improving food security in Africa by enhancing resistance to disease and heat in chickens; Feed the future innovation lab for genomics to improve poultry" Zhou H. Bunn D, Gallardo G, Lamont S, Dekkers J. et al. **\$6,000,000**. This grant uses contemporary high-throughput genetic technologies of SNP chips and functional genomics, along with targeted genome resequencing and extensive statistical and bioinformatic analyses to dissect and identify the genetic factors of the chicken that enhance its resistance against NDV and heat stress by assessing diverse populations of chickens, including well- characterized research lines and highly relevant local African ecotypes. **Impact:** The project has significantly improved institute capacity (infrastructure has developed in Africa) and human capacity building, including by training of students and scientists both on-site in Africa and in the US in essential skills that enable the African partners to sustain and disseminate the results of this project. Project outcomes are expected to reduce poverty, hunger, and malnutrition, and empower women through increased agricultural productivity achieved by decreasing the major losses that currently occur as a result of Newcastle disease and heat stress in African chickens.
- 2015-2017. USDA NIFA 2015-67015-22940 "Genome wide identification and annotation of functional regulatory regions in livestock species" H. Zhou, P. Ross, I. Korf. **\$500,000**. This grant supported research

effort in functionally annotating regulatory elements in the three major farm animal species by integrative bioinformatic analysis of RNA-seq, DNase-seq and ChIP-seq data from the eight most important tissues. **Impact:** This will generate first line of re-annotation of gene structure and landscape of functional regulatory elements in chicken, bovine, and swine genomes, and will develop a framework to continue a more in-depth functional annotation of these genomes and other agricultural animals.

- 2011-2016. USDA NIFA “System Biology Analysis & Modeling Of Complex “Omic” Data: A Service Center Approach”. Zhou H. Drake K. **\$750,000**. This grant has supported an effort in collaboration with Seralogix, to provide sophisticated systems biology and modeling analysis with visualization for a total of 12 projects generating more than 100 data analysis module reports. These projects include microarray and RNA-seq data from cattle, sheep, chicken and mice in the areas of nutrition, reproduction, growth and disease. **Impacts:** Results have contributed greatly to our understanding and formulation of new hypotheses that are advancing the fields of animal infection, nutrition, reproduction, and physiology.
- 2015-2018. USDA NIFA 2015-67015-23093 and BBSRC BB/M028208/1. “US-UK Collaborative Research: Host Resistance to Avian Pathogenic E. coli” Lamont, S.J. (PD), Wolc, A; Kaiser, P. (dec.), Stevens, M., Vervelde, L. **\$499,999** (USDA). This grant supported the genomic, molecular and cellular characterization of the host-pathogen interactions between chickens and avian pathogenic E. coli (APEC), through the use of unique inbred chicken lines in both countries that differ in resistance to avian pathogenic E. coli (APEC), analysis of transgenic chickens in which all cells of the myeloid lineage express a fluorescent protein to aid the phenotyping of APEC-infected cells, definition of the transcriptome of infected cells, association of resistance with bird genetic variation (in structure and expression) through GWAS and RNA-seq analysis, and validation of selected research findings for translation into industry application. **Impact(s):** The impacts of this grant will be a reduction of the negative impact of respiratory APEC on the poultry industry, improved poultry health and vaccine strategies, and decreased use of antibiotics in food animals.

Travel support and opportunities for trainings:

- Provided financial support for over 40 students, postdocs, members, and speakers to attend the PAG Poultry Workshop (2013-2017).

Major impact products (could be potential impact):

- Genomic selection is now routinely implemented in both the meat (broiler) and egg (layer) breeding companies. This has greatly accelerated the genetic progress required by the industry to meet the growing consumer demand. Furthermore, poultry health and welfare have been enhanced.
- The chicken genome assembly reached the stage that scientists can confidently identify genes and genetic variations associated with biological traits, many of which are highly relevant to the poultry industry.
- The draft assembly of the turkey genome has been released, which affords the opportunity for efforts similar to those in the poultry industry, e.g., biological characterization, genomic selection.

SHEEP/GOAT

<http://www.animalgenome.org/sheep/>

Direct contributions to Objective 1:

- Reference genomes were published for goat in Nature Biotechnology in 2013 and for sheep in Science in 2014. An improved reference genome for the goat was published in Nature Genetics in 2017 that leveraged single molecule sequencing plus chromatin conformation capture to create a genome assembly with chromosome length scaffolds. **Impact(s):** The reference genomes advanced the status of mammal genome assembly and annotation technology at the time of publication. They also enabled comparative genomic insight into rumen biology, and expanded understanding of genes underlying numerous economically important traits. The improved goat genome reference elevated the standard for quality of mammal reference genome assemblies. Together, these results will underpin all future efforts to improve genetics of productive efficiency in sheep and goats.

Direct contributions to Objective 2:

- A sheep 600K SNP chip and a goat 52K SNP chip were both released for research in 2014. **Impact(s):** This dramatic increase in SNP density for sheep and the first genome-wide medium density panel for goat have enabled powerful new tools including genome-wide association and genomic selection to dissect and improve numerous traits in sheep and goats.

Direct contributions to Objective 3:

- A sheep genomes database has been assembled to provide a public, large-scale warehouse for global sheep genetic diversity. The database now includes re-sequencing data from almost 1,000 sheep genomes with an overall total of nearly 100 million identified sequence variants. **Impact(s):** This resource will accelerate the identification of causal variants for numerous traits and enable previously inconceivable analyses.

Communication:

- Strategic planning conference calls with international attendance were held in 2015 and 2016, in addition to annual meetings at PAG and biennial meetings at International Society for Animal Genetics (ISAG). **Impact(s):** These contributed to the development and implementation of the Ovine FAANG Project tissue collection in the U.S. and recent successful leveraged grant funding.

Research support mini-grants (coordinator grants):

- Ovine FAANG (Functional Annotation of ANimal Genomes) Project tissue collection (\$15,000 Coordinator funds). Sample collection of 100 tissues from a new reference genome sheep was conducted in 2016. **Impact(s):** Use of the reference genome animal to generate functional annotation data will enable state-of-the-art annotation, and this was recognized by a \$500,000 grant from USDA-NIFA (led by Brenda Murdoch and Noelle Cockett) to generate annotation data for FAANG using these samples.

Leveraged funds and stakeholders' use of project outputs:

From 2013-2017, the investigators and stakeholders utilized the tools and resources generated through NRSP-8 to acquire leveraged funds from diverse projects totaling **\$4,220,000**. The overall total leveraged funds include **\$3,644,000** from combined federal sources, **\$467,000** in local/institutional funds, and **\$109,000** from industry. **Impact(s):** Diverse leveraged funds supported projects to improve the functional genome annotation, provided

large-scale genetic variant datasets and analyses, and informed causal variant searches for numerous important traits. Selected grants are highlighted below.

- **Ovine FAANG Project.** USDA National Institute of Food and Agriculture Competitive Grant 2017-67016-26301. **\$500,000.00.** Brenda Murdoch (PD). The majority of a mammalian genome does not consist of coding genes but rather copious portions that influence the timing, tissue, and spatial regulation of the expression of genes. This project will generate data from numerous tissues of the reference genome sheep to elucidate transcription details at unprecedented resolution, identify numerous regulatory elements including promoters, enhancers, silencers and insulators, and perform numerous other assays including measures of DNA methylation and open chromatin. **Impact(s):** The outcomes of this project will provide the sheep research community with cutting edge functional genome annotation resources and enable comparison of genome regulation across animal species by the global FAANG (Functional Annotation of ANimal Genomes) consortium.
- **Genome-Wide Association Analysis for the Identification of Alpine Goats With High Milk Production.** USDA National Institute of Food and Agriculture Competitive Grant 2015-67031-23776. **\$100,000.00.** Yonathan Tilahun (PD). This project used genome-wide association analysis to identify genomic regions and genes associated with dairy production in Alpine goats. **Impact(s):** These results provide a foundation for improved understanding of lactation, and they enable enhanced selective breeding in the economically important Alpine goat breed.
- **Genomics of Resilience in Sheep to Climatic Stressors.** USDA National Institute of Food and Agriculture Capacity Building Competitive Grant 2013-38821-21389. **\$600,000.00.** Arthur Goetsch (PD). Environmental extremes are a key constraint on agricultural production, and sheep have genetic diversity relating to many traits. This project will identify genetic variation important for resilience to stressors present in extreme environments by surveying sheep from divergent breeds and geographic locations. **Impact(s):** This project has defined useful phenotypic measures for assessing resilience to conditions including heat load as well as feed and water availability. Assessment of genomic bases for these traits is progressing. This will enable selective breeding to improve sheep resilience to environmental extremes.
- **Building the Sheep Genomes Database.** USDA National Institute of Food and Agriculture Competitive Grant 2013-67015-21372. **\$500,000.00.** Noelle Cockett (PD). This project addresses the need for generation and collection of sheep sequences including genetic and genomic variants to benefit the whole research community. Objectives include sequence data generation on a genetically diverse set of sheep, identification of genetic variants, and public accessibility of these data plus sequence data from outside sources. **Impact(s):** Sequence data from nearly 1000 animals including over 80 million genetic variants have been made public. This resource has identified many genetic variants and provided valuable assessments of their frequency and geographic distribution among global domestic sheep.

Travel support and opportunities for trainings:

- Travel of 12 students was funded to attend the Sheep and Goat workshop at Plant & Animal Genome (PAG) meetings (2012-2017). **Impact(s):** The purpose of the travel award program is to help graduate students to travel to the annual PAG meeting to present their research.

Major impact products (could be potential impact):

- Development of 45 genetic locus tests, overarching genomic selection, and potential for gene editing systems. Genetic locus tests for 43 economically important traits plus 2 biomedical model mutations have been developed since 2014, including tests for loci with large influence effects on reproductive prolificacy, growth/milk production, muscular development, and pathogen susceptibility. In addition, the first genomic selection results for each species hold promise for incorporating unprecedented numbers of genes and traits into

genomics-enabled breeding gains moving forward. Finally, reference genome sequences underpin the broad utility of new knockout and gene editing systems now available for both sheep and goats.

SWINE

<http://www.animalgenome.org/pigs/>

Direct contributions to Objective 1:

- A global group with significant contributions by several NRSP-8 scientists created the first draft assembly of the swine genome (2012). **Impact(s):** The genome assembly led to SNP genotyping tools that have been used by the swine genetics industry to accelerate genetic improvement. The assembly has accelerated the analysis of the swine genome, to understand and manipulate genes of major importance (such as a CD163 KO pig that is resistant to porcine reproductive and respiratory syndrome virus (PRRSV) infection).
- A global group led by NRSP-8 scientists formed the Immune Response Annotation Group (IRAG) and worked closely with the Wellcome Trust Sanger Institute on the community annotation of immunity associated genes in the pig genome (2012), resulting in the detailed manual annotation of over 1,400 genes in the first draft assembly of the swine genome. **Impact(s):** This annotation work expanded swine bioinformatic capabilities for immune response associated genes, as well as identifying hundreds of missing genes in the assembly. These data are of high importance as geneticists use the genome assembly for genetic improvement of pigs.
- A dramatically improved swine genome assembly (2016) was created by researchers in the UK with contributions to gene annotation by NRSP-8 scientists. **Impact(s):** The one hundred-fold increase in quality of the sequence data and gene annotations will dramatically improve the use of the porcine genome, through increasing the accuracy of SNP analysis, gene mapping of traits and regulatory network analysis.

Direct contributions to Objective 2:

- A public-private collaboration, supported with NRSP-8 Swine Coordination funds, assessed the recently released GeneSeek Genomic Profiler for Porcine LD (GPP-LD) SNP chip list for accuracy of imputing genotypes for the Illumina Porcine SNP60. **Impact(s):** This project demonstrated the accuracy and cost-efficient use of the GPP-LD for implementation of genotype imputation for genomic selection in pigs. Data and tools are publicly available at https://www.msu.edu/~steibelj/JP_files/imputation.html.
- Methods were published for conducting meta-analyses to combine genome-wide association (GWAS) studies, and these methods were used to combine SNP and phenotypic data for a meta-analysis of meat quality data for three populations from MSU and US-MARC (collaboration funded by the National Pork Board). SNP marker data for National Swine Registry (NSR) populations has been used to establish baselines for breed identification. **Impact(s):** A strategy has been proposed for implementing use of SNP markers for breed composition estimation in the Yorkshire breed. (Funded by NSR). This project provided improved methods for genomic selection, GWAS and eQTL analyses in pigs.
- A candidate genetic marker (GBP5) associated with resistance/susceptibility to primary PRRSV infection has been identified based on transcriptomic data performed with PRRS Host Genetics Consortium (PHGC) trial samples. **Impact(s):** These results provide a promising candidate gene for further functional work to validate its role in the host response to PRRSV infection. This genetic variant (now included on the GPP-LD SNP chip) will help breeders select for pigs, regardless of breed, with the favorable genotype for PRRS resistance
- Important progress was made in the dissection of the QTL associated with PCV2 viremia located at the proximal end of SSC12. Potential genes and functional polymorphisms were evaluated using long-reads genomic sequences, gene annotation and RNA-seq. **Impact(s):** Initial results indicate novel genes and mechanisms potentially involved in PCV2 innate immune response. Marker-assisted selection for the favorable allele will decrease susceptibility to PCVAD and will provide a complementary solution to vaccination.
- Identified SNP markers and combinations of markers associated with age at puberty (AP), in particular early onset puberty, and sow reproductive longevity by conducting a genome-wide characterization of the cross between commercial Landrace and Nebraska Index Line (NIL), which has been subjected to long-term selection

for litter size. **Impact(s):** Candidate genes and SNPs located in the QTL region for litter size traits identified by GWAS, genome and RNA sequencing could explain some of the observed phenotypic variation. Following validation, some of these polymorphisms could be used in selection to improve sow fertility.

- Genomic and phenotypic predictors of PRRS resistance in pregnant gilts and their fetuses have been identified in collaboration between NRSP-8 and Canadian researchers. Gene expression data suggested that the immune response to PRRSV infection is faster in resistant versus susceptible fetuses, thus possibly contributing to different levels of fetal pathology. **Impact(s):** Overall, the findings provide insight into anti-viral immune responses that may help researchers to determine new ways to reduce the impact of PRRS in pregnant gilts and to select for genetically resistant breeders.

Direct contributions to Objective 3:

- The PRRS Host Genomic Consortium (PHGC) Database <http://www.animalgenome.org/lunney/index.php> was developed by scientists at Iowa State, Kansas State and BARC to house the large amounts of phenotypic and genotypic data that are collected across several research labs in the PRRS research community. The database has been expanded to include SNP genotype data and tissue gene expression data from arrays, next generation sequencing, RNA-seq and NanoString codeset data. **Impact(s):** The PHGC database is continuing to be used to identify genes and pathways linked to phenotypes recorded in the database. The website has been used by many researchers interested in the genetic control of the phenotypic response to PRRSV infection and vaccination. It has enabled real-time data sharing among users from geographically disparate locations; the website has had up to 93,000 visitors in the last few years, from >1,000 unique Internet IP addresses.
- A public website providing information on genomic data for the USDA-NIFA funded project 2011-68004-30336 was created to increase stakeholder awareness of genome-based analyses in swine phenotypic traits of economic importance (www.Swinefeedefficiency.com). **Impact(s):** The website is used by visitors from over 1,000 unique IP addresses each month.

Communication:

- NRSP-8 members and others wrote and published a refereed journal article describing the goals of the Functional Annotation of Animal Genomes (FAANG) consortium in 2015. **Impact(s):** Cited 27 times by PubMed articles through February 2017, this paper greatly increased the visibility of this new project, as well as established the standards and methods required to participate, a critical organizational step for the FAANG consortium.
- The Swine and Bioinformatics Coordination Program develop and continues to host an open website (www.faang.org) which provides information and coordination of the community developing this effort. **Impact(s):** The website is a clearinghouse for information on FAANG for contributors as well as the general public. The site has been visited over 65,000 times by unique IP addresses since 2015.
- The Swine and Bioinformatics Coordinators and others wrote and published a refereed journal article describing a public Workshop on FAANG “Gathering On-FAANG” ([GO-FAANG], October 7-8, 2015) in Washington DC. **Impact(s):** Increased visibility of the FAANG project to non-attendees through this publication will advance both the understanding of FAANG goals and opportunities for funding of animal genomics research worldwide.

Research support mini-grants (coordinator grants):

- A public-private collaboration between 2 NRSP-8 supported Experiment Stations (MI, IA) and Affymetrix, Inc. successfully tested a high-density 670K (HD) SNP genotyping platform for swine. **Impact(s):** Science-based evaluation of the HD chip across US swine breeds will expand proven genetic tools available for researchers and industry for genetic improvement.

- A second public-private collaboration featured NRSP-8 organization of support from the National Pork Board (\$30,000) as well as coordinator support to leverage Federal research dollars. With this support, a USDA-NIFA funded project was funded, to functionally annotate the genome of swine and other species (2015-67015-22940 “Genome wide identification and annotation of Functional Regulatory regions in Livestock Species” Zhou, H., Korf, I., Ross, PA. **\$499,842**. The overall goal of this project is to generate a comprehensive resource of functional regulatory elements for the chicken, cattle, and pig genomes. All work relies on the genomic and bioinformatics tools developed by NRSP-8 researchers and researchers across the globe. **Impact(s):** This pilot project will lay the foundation for improved use of the swine genome for genetic improvement by industry. Financial support by a major commodity group also demonstrated the clear support of industry for this pre-competitive research.
- Contributions from Bioinformatics, Poultry, Sheep and Swine Genome Coordination Programs (\$10,000) were leveraged to successfully obtain conference funding (~\$30,000) from Government (USDA-NIFA, NSF) and industry (Illumina, Inc.) to hold a Workshop on the new Functional Genomics of Animal Genomes (FAANG) in Washington DC, attended by six participating funding agencies across three countries, as well as private companies. **Impact(s):** The livestock genetics and federal agencies communities became familiar with the importance of data that identifies the functional components of the genome for any species. Through in-depth discussions in two small groups, the FAANG community created excellent working relationships and developed a plan for sharing of resources and information.

Leveraged funds and stakeholders’ use of project outputs:

Leveraged funds from diverse projects totaling \$9,384,996. The overall total leveraged funds include \$8,229,905 from combined federal sources, \$533,500 in local/institutional funds, and \$621,591 from industry. Selected grants are highlighted below.

- 2013-2017: USDA-NIFA-AFRI. 2013-68004-20362. “Genetically Improving Resistance of Pigs to PRRS Virus infection.” Dekkers, J. C.; Tuggle, C. K.; Stalder, K.; Reecy, J., **\$2,998,931**. The long-term objective is to enhance the ability of pigs to resist and minimize the impact of the most costly infectious disease in the US pork industry: Porcine Reproductive and Respiratory Syndrome (PRRS). The project is building on prior genetic analyses as well as novel such analyses, both made possible by genomic tools derived from the work of NRSP-8 members. **Impact(s):** To date, this group has validated the effect of a genetic marker previously identified to be associated with improved host response to PRRS virus infection under experimental PRRS-only challenge, under conditions that are closer to conditions that pigs experience in the field. They also demonstrated that genetic selection for improved host response of pigs to PRRS infection is possible. The identified genetic marker is available to the industry, which will help to address the enormous economic and animal welfare impacts of PRRS in the US and global industry.
- 2013-2017: USDA-NIFA-AFRI “Translational Genomics For Improving Sow Reproductive Longevity” Ciobanu, D. C.; Kachman, S. D.; Lents, C. A.; Safranski, T. J. **\$1,166,650**. The long-term goal of this project is to develop a panel of genetic markers that will predict at weaning, gilts with superior propensity for reproductive longevity, which in turn will reduce culling rates and the cost associated with developing replacement sows. This project is capitalizing on prior genetic results using SNP panels created using NRSP-8-supported genomic and bioinformatics resources. **Impact(s):** The integration of RNA and genome sequence data and evaluation of a novel approach for genome-wide associations has the potential to lead to the discovery of functional mutations that could reduce age at puberty, improve fertility and reproductive longevity, leading to an increase in sow net values in the commercial herds.
- 2015-2017 USDA-NIFA-AFRI “The TET Family regulates embryogenesis by controlling both DNA and histone methylation.” Lee, K. **\$334,263** This project is studying the action of TET family genes, which are epigenetic regulators in various cell types, during early porcine embryogenesis to expand our knowledge of how epigenetic marks are regulated during embryogenesis. The researchers are testing the specific involvement of TET family genes on epigenetic modifications, occurring after fertilization, through recent molecular analysis technologies. These studies are using the genome assembly to design tools for mutating

specific TET genes, as well as use the annotation of the porcine genome assembly to perform the epigenetic analyses. **Impact(s):** Understanding the action of key epigenetic regulators, such as TET family, during embryogenesis will allow improvements in availability of in vitro-derived embryos and increase embryo viability in domestic species.

- 2014-2017 USDA-NIFA-AFRI “Reducing production losses using behavioral and genomic tools to identify pigs suited for group living.” Siegford, J.M. **\$749,059**. In pigs, social behaviors can affect performance of all pigs in a group or result in injury or lameness. This overall goal of this project is to ensure global food security by improving use of behavior in breeding programs to reduce loss from non-infectious diseases in group-housed pigs. The scientists will use genotyping tools and resources supported by NRSP-8 project outputs to estimate genetic parameters and perform genome-wide prediction and association of social behavioral traits expressed by group-housed pigs to determine underlying genomic control of these traits. **Impact(s):** By examining social behaviors of group-housed pigs and understanding genetic control of these traits, we will develop tools allowing producers to better select and manage group-housed pigs, including gestating swine. Importantly, such information will reduce animal losses occurring from non-infectious diseases such as injury and lameness.

Travel support and opportunities for trainings:

- NRSP-8 Swine Coordination program provided competitively award funds for five graduate students to travel and present their genomics research at the Plant and Animal Genome meetings from 2013 to 2017. **Impact(s):** Increased visibility of young scientists and additional experience in presenting their research work will improve the education and professional development of these promising animal genome scientists.

Major impact products (could be potential impact):

- Porcine reproductive and respiratory syndrome (PRRS) is the most economically important disease of pigs worldwide and isolates vary greatly. The PRRS Host Genetics Consortium (PHGC) researchers compared host responses to genetically different PRRS virus isolates. NRSP-8 researchers statistically evaluated and affirmed that the resistance associated swine chromosome 4 (SSC4) allele had an effect on viremia with faster virus clearance and positive effects on weight gain. Estimates of heritability of genetic resistance to PRRSV infection were moderate for different PRRSV isolates. **Impact(s):** Genetic selection in the swine industry for increased resistance to genetically distinct PRRSV isolates is now possible.

BIOINFORMATICS

<http://www.animalgenome.org/bioinfo/>

Direct contributions to Objective 3:

- **Continued curation and development of the Animal Quantitative Trait Loci database (QTLdb).**
 - Horse and catfish were added to the Animal QTLdb.
 - Animal QTLdb database (regularly updated), a total of 104,272 new QTL/association data has been curated into the Animal QTLdb. The database contains Cattle: 95,332; Chicken: 6,633; Horse: 1,245; Pig: 16,516; Sheep: 1,412; Rainbow trout: 127 QTL reports. **Impact(s):** The database provides a set of useful molecular breeding tools allowing selection for improved productions traits in various species.
 - Development of curation tools for Animal QTLdb—allow batch entry of pre-tabulated data. **Impact(s):** This has greatly accelerated the speed of data entry for large volumes of data using a semi-automated method.
 - Additional support tools for Animal QTLdb have been released—enabled multiple genomes for cattle, chicken, and sheep. **Impact(s):** Allows QTL/association data alignment to each respective genome.
 - New data alliances with UCSC Genome Browser and the Monarch Initiative have been established. **Impact(s):** This allows users to obtain Animal QTLdb data via more avenues.
 - Integration of the Animal Trait Ontology into the Vertebrate Trait Ontology (<http://bioportal.bioontology.org/ontologies/VT>). We have continued working with the Rat Genome Database to integrate ATO terms that are not applicable to the Vertebrate Trait Ontology into the Clinical Measurement Ontology (<http://bioportal.bioontology.org/ontologies/CMO>). Traits specific to livestock products continue to be incorporated into a Livestock Product Trait Ontology (LPT), which is available on NCBO's BioPortal (<http://bioportal.bioontology.org/ontologies/LPT>). We have also continued mapping the cattle, pig, chicken, sheep, and horse QTL traits to the Vertebrate Trait Ontology (VT), LPT, and Clinical Measurement Ontology (CMO) to help standardize the trait nomenclature used in the QTLdb. Now VT data download has also been made possible through the Github portal. **Impact(s):** Allows users to query Animal QTLdb data more effectively.
 - The VT/LPT/CMO cross mapping has been employed by the Animal QTLdb and VCMMap tools. Annotation to the VT is also available for rat QTL data in the Rat Genome Database and for mouse strain measurements in the Mouse Phenome Database. We have also been integrating information from multiple resources, e.g. FAO - International Domestic Livestock Resources Information, Oklahoma State University - Breeds of Livestock web site, and Wikipedia, as well as requests from community members, to continue development of a Livestock Breed Ontology (LBO; <http://www.animalgenome.org/bioinfo/projects/lbo/>).
 - The Animal QTLdb has undergone continued upgrades to use MIQAS for data curation and data integration (<http://www.animalgenome.org/QTLdb/doc/minfo/>). Minimal standards for curation into the Animal QTLdb have been established (<http://www.animalgenome.org/QTLdb/doc/minfo/>). **Impact(s):** This will increase the quality of the data present in Animal QTLdb.
 - Expansion of the Data Repository for the aquaculture, cattle, chicken, horse, pig, and sheep communities to share their genome analysis data. **Impact(s):** This resource has proven useful for the community (<http://www.animalgenome.org/repository>). To date there are 1,140 data files, totaling 140 Gb, shared through this platform. This includes supplementary data files to publications deposited by the authors.
- **Development of a Collaborative VCF Info Mining Platform.** A recent effort was to establish a partner relationship with the European Variation Archive (EVA) in order to make shared VCF files more accessible and adherent to a better file format standard. **Impact(s):** This platform is continually used by the community and allows for more efficient on-line file sharing.
- **Collaboration with and support of community research groups.**

- PRRS project database and web site: Partnered with researchers at Kansas State University, Michigan State University, Iowa State University, and U.S. Department of Agriculture, to further developed and improve the web-interfaced relational databases to store and disseminate phenotypic and genotypic information from large genomic studies in farm animals and better serve the needs of researchers. New animal genotype and phenotype data were added. The database/web site operations have been smooth with little or no down-time (<http://www.animalgenome.org/lunney>).
- Set up a virtual machine site to host the Online Mendelian Inheritance in Animals (OMIA) database created and maintained by Dr. Frank Nicholas at the University of Sydney (<http://omia.animalgenome.org/>).
- Provided a dedicated virtual machine platform to a USDA-funded project to develop cyber resources for the Striped Bass Genome Database activities, led by Benjamin Reading and Charles Opperman at North Carolina State University (<http://stripedbass.animalgenome.org/>). The database and web site have been maintained for the past 3 years.
- Developed and maintained the web site and tools for the Functional Annotation of ANimal Genomes (FAANG) consortium. The efforts included setting mailing lists for each committee and subcommittee, a user forum, interactive meeting sites, platform for collaborative funding applications, tools for the coordination of sample collections, and online publishing capabilities, etc., to support this large international collaboration (<http://www.faang.org/>). We also helped to support the GO-FAANG meeting in Washington DC on October 7-8, 2015.
- A collaborative site at iPlant has been set up to share some of the web service loads, 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>).
- **NAGRP Tool Box has been actively maintained for online use by the community** (<http://www.animalgenome.org/bioinfo/tools/>).
- **Overall impact(s):** The Bioinformatics helpdesk annually helps more than 60 research groups. Involvement has ranged from data transfer, data assembly, and data analysis, to software applications, code development, information resources, etc.