

# NRSP 8, National Animal Genome Research Program

**Duration:** 10/01/18 – 9/30/23

## Prerequisite Criteria

*How is the NRSP consistent with the mission?*

## INTRODUCTION

Livestock and poultry are top US agricultural commodities accounting for more than half of the agricultural cash receipts, at least \$100 billion each year. On the other hand, the US imports ~90% of its seafood, at a deficit of more than \$14 billion per year. Therefore, advances in livestock, poultry and aquaculture production capacity and efficiency are essential to advance national and global food security, especially with increasing consumer demand for animal welfare, growing restrictions on antibiotic use, and changing environments. To enable agri-animal industries to increase production and meet this growing demand, we need robust animals with superior health and production traits. A major requirement for increasing production efficiency is genetically improved strains of animals with sufficient genetic diversity to allow continued improvements in traits including yield, disease and pest resistance, efficient reproduction, and stress tolerance, while assuring nutritional quality and animal welfare. The continued development of tools that enable genomic selection for production, disease resistance and genetic diversity will greatly benefit these agri-animal industries.

Infrastructure, including genomic tools (e.g., maps, annotated reference genome assemblies, genotyping platforms), integrative bioinformatics tools, databases and genetic resource populations play a critical role in enabling genomic discoveries in livestock, poultry and aquaculture species. Since 2003, NRSP-8 National Animal Genome Research Program (NAGRP) has played a major role in enabling genomic discoveries in livestock, poultry and aquaculture species. As outlined in the “2014 USDA Research, Education, and Economics Action Plan,” ([https://www.ree.usda.gov/ree/news/USDA\\_REE\\_Action\\_Plan\\_03-2014.pdf](https://www.ree.usda.gov/ree/news/USDA_REE_Action_Plan_03-2014.pdf)) infrastructure is the critical foundation upon which genomics-oriented discovery science is built; and these tools, coupled with the education and training of students, scientists, and the public, are critical to translating genomic discoveries into substantial advances in agricultural practices. *The mission of NRSP-8 is to develop and sustain these crucial infrastructure components.* This infrastructure is, in turn, used to leverage additional funding for agri-animal genomics research from both public and private sources. As detailed in the **Past Successes** section and the **Accomplishments Report**, NRSP-8 has been highly successful in the previous period in substantially advancing animal genomic infrastructure, catalyzing a broad array of research activities resulting in > **\$94 million dollars in leveraged funds**, and facilitating communication across all agri-animal species groups. Prior success and the rapidly evolving and complex genomic technologies and landscape underscore the need for continuation of NRSP-8 and its mission of supporting the agri-animal genomics research community and industry partners.

The constant advancement of next-generation sequencing (NGS) technologies and assay applications and the exponential decreases in sequencing costs and increases in throughput, have produced seismic shifts in research approaches and have substantially broadened the scope of animal genomics. For example, while production of high-quality annotated reference genomes for agricultural animals remains a considerable undertaking—the infrastructure needed to assemble billions of multi-platform reads, annotate genes, genome features, and nucleotide variants, and curate this information in a manner accessible to

stakeholders—outcomes far outweigh the sequencing costs. As whole genome sequences become available for most target species, broad NGS re-sequencing efforts and high-throughput genotyping arrays (‘chips’) have become valuable approaches for characterizing variation in breeds and strains, and for linking this variation to economically relevant phenotypes. In the past, reference genome assemblies were mostly generated at large sequencing centers, today NGS has “democratized” sequencing such that most universities and many companies can now use it to study their own populations of interest. These changing strategies are rapidly generating enormous datasets that have quickly overcome the informatics resources within individual labs and small species groups. These new “big data” challenges are reflected in our updated objectives for this renewal, as briefly outlined below and expanded on in the **Objectives section**. Harnessing the power of big data in agri-animal genomics research is only possible through coordinated teamwork efforts such as those successfully exhibited by the NRSP-8 species consortia.

NRSP-8 is an umbrella organization to promote the needs of the animal genomics community. With limited funds available to species coordinators, NRSP-8 helps to provide critical infrastructure for agri-animal genomic discoveries including genomics and bioinformatics tools and databases, genetic resource populations with economically-important phenotypes, and education and training of students, scientists, and the public.

Our first objective is to advance the quality of reference genomes for all agri-animal species through providing high contiguity assemblies, deep functional annotations of these assemblies, and comparison across species to understand structure and function of animal genomes. The NRSP-8 community, organized around species groups, the AnGenMap list-server, and NRSP-8 meetings in conjunction with the annual Plant and Animal Genome Conference (PAG), provides the core infrastructure and leadership necessary for organizing these efforts, including sharing of DNA resources, collaborative use of computational power, and community-based genome annotation.

Our second objective is to advance genome-to-phenome prediction by implementing strategies to identify and validate genes and allelic variants predictive of biologically and economically important phenotypes and traits. The tremendous power of genome-wide variant analysis is leading to the identification of markers linked to important production traits in agri-animal species. These analyses also suggest the involvement of candidate genes, whose functions are often unknown in the studied species. NRSP-8 support is critical to developing powerful functional “-omic” strategies, from improved annotation to proteomics, metabolomics, metagenomics, and metatranscriptomics to knock-out/down and genome-editing approaches.

Our third objective is to advance analysis, curation, storage, application, and reuse of heterogeneous big data to facilitate genome-to-phenome research in agricultural animal species of agricultural interest. This bioinformatic capacity is a critical component of infrastructure support enabled by NRSP-8. NRSP-8-supported approaches in this area include continued expansion of the tools and capabilities of the bioinformatics coordination program (<http://www.animalgenome.org/>), development of community-wide pipelines and portals for variant analysis and curation, training of students, scientists, and stakeholders in high-throughput data analysis, and development of collaborative platforms that facilitate species-group level access to timely genomic data.

Across all objectives of this renewal, the NRSP-8 proposes to continue providing enabling technologies and support activities, and to disseminate new data/genetic material across the animal research community. The organizational structure of NRSP-8 serves to bring together domestic scientists within a given livestock/poultry/aquaculture group, and facilitate rapid transfer of successful approaches to commercial

national and international stakeholders.

“Blueprint for USDA Efforts in Agricultural Animal Genomics 2008-2017,” was used as a guide in formulating the objectives for the previous NRSP-8 objectives. While this proposal was due for submission, a new Blueprint document “Genome to Phenome: An USDA Blueprint for Animal Production” was being prepared. The updated blueprint, when available will be used as a guide to adjust the objectives of this proposal.

*How does this NRSP pertain as a national issue?*

#### **A. National Scope**

The membership of the NRSP-8 encompasses scientists from the dairy and beef cattle, poultry, equine, sheep, goat, swine, and aquaculture sectors, and is impacting agriculture in every state and region of the U.S. The use of genomics to improve the genetics of US animal-based commodities has been adopted by nearly all the food and fiber animal-breeding industries for which genomic tools have been developed. NRSP-8 scientists across the country were instrumental in creating many of these genomic tools, showing the broad scope and applicability of the NRSP-8 project.

Another indication of the broad scope of NRSP-8 can be seen in the participation in and utilization of NRSP-supported forums, tools, and conferences. The AnGenMap discussion server has about 4,000 members across the nation and around the world, of them about 3,000 are on AnGenMap mailing list. The QTLdb program, developed by the NRSP-8 Bioinformatics Coordination Project team, is used by hundreds of scientists globally- for example, QTLdb for cattle has been cited in over 770 peer-reviewed publications. Similarly, the Plant and Animal Genome (PAG) conferences are held jointly with the annual NRSP-8 meeting, with 776 animal genome scientists participating in this international meeting in 2017; cumulatively over 9,200 animal genome scientists have attended PAG from 1999 to 2017. The chance to communicate with colleagues throughout the genomics community and to access the most recent developments from genomics industry representatives has been critical for much of the past progress outlined under the Past Successes section and the Accomplishments report.

In addition to these broad indications of the importance, impact and national scope of NRSP-8 activities, animal genomics research can be directly related to goals and challenges put forth in two recent national policy documents on agricultural research, education and engagement:

1. Animal genomics is directly relevant to two Strategic Goals (SG) in the document “2014 USDA Research, Education, and Economics Action Plan” introduced above—SG#3 “Help America promote agricultural production and biotechnology exports as America works to increase food security” and SG#4 “Ensure That All of America’s Children Have Access to Safe, Nutritious, and Balanced Meals.” In SG#3, animal genomics has a national role in Strategy #2 – “USDA will generate fundamental knowledge through research in genomic sciences and biotechnology to enhance agricultural sustainability and productivity.” Genomics is accelerating the genetic improvement in sustainable production through increasing feed efficiency in cattle, pigs and poultry as well as other major commodities. In addition, animal genomics has a clear role to play in improving resilience of animals to infectious disease, and thus addresses ‘sustainability and productivity’ as well as the safety of animal food products (SG#4). Public health is also impacted as many animal pathogens also infect people.

2. In the just-published document “The Challenge of Change” (<http://www.aplu.org/projects-and-initiatives/international-programs/challenge-of-change/index.html>), the overarching Grand Challenge is described as “Sustainably feed an expanding population and improve prospects for food and nutrition security for all.” The specific Grand Challenges in this document are described in detail in the Rationale section. Here, we highlight that animal genomics is directly addressing several of the listed challenges, including those related to *Food Availability* through genetic improvement. These include Challenge 1 “Increase yields, profitability, and environmental sustainability simultaneously,” and Challenge 2 “Develop the varieties and breeds needed for sustainable food systems.” NRSP-8 scientists are also addressing Challenges related to *Food Access* including “Create and share resources that serve all populations,” as many of the efforts in NRSP-8 focus on creation of genetic/genomic information resources that anyone nationally or globally can use for genetic research and improved food access. Finally, as discussed above, animal genomics research is directly addressing the *Food Utilization* Challenge 7 “Ensure a safe and secure food supply that protects and improves public health” through research on identifying the genetic architecture of animal resilience to pathogens that can cause human disease, and creating tools to improve such resilience.

The applicability of animal genomic research (and its support via NRSP-8) to a wide array of national issues and animal production needs is reflected both in recent USDA-NIFA proposal requests, which often incorporate a genomics component, and in the diverse applications of the genomic techniques employed by USDA-affiliated researchers. Many modern biological questions, especially those related to practical questions related to food and agricultural research, now use genome-wide tools for screening, detection, and analysis of genes, RNAs and proteins. These projects range from nutritional genomics to comparative immunology to marker-assisted breeding, but often share the same core genomic technologies and needs for bioinformatics support. For example, an algorithm developed originally for swine genome variant analysis in North Carolina may be readily adapted by researchers working with similar data in chicken populations in Arkansas.

## **B. Continued National Need for Genomic Tools and Resources**

As detailed above, the NRSP-8 mission closely aligns with the challenges described in “2014 USDA Research, Education, and Economics Action Plan” and “The Challenge of Change.” These challenges for food production are long-term in nature and are heightened by rising world human populations and environmental obstacles such as climate change. As animal genomic tools and technologies are increasingly integrated into the breeding and production practices of a broad range of species, the need for national genomics infrastructure support (as provided by NRSP-8) is only expected to increase. While genomics has been applied primarily to breeding and selection, a key deliverable is enhanced biological information on traits of interest. Such information can be applied in other research areas important for animal science, including nutrition, reproduction, meat science, welfare, and disease resistance.

Whole genome reference sequences for several species groups were delivered and/or refined in the last five years, while sequencing was initiated and nears completion in several others. **The focus of the proposed renewal turns, therefore, from the structural assembly of primary “reference” sequences to the further refinement of these resources and the capture and utilization of trait-affecting genome variation through a variety of approaches.** These new approaches are based on having the reference assembly in hand and therefore depend on the quality of this initial reference sequence. It is difficult to overstate the magnitude of this change, which can be compared to parallel developments in computing and digital communications. When NRSP-8 began in the 1990s, few among us expected that we would see

complete genome sequences available for our species, and many of our stakeholders found minimal relevance of genomics to their work. *Today, many, if not most, breeding companies across the nation and the world are routinely obtaining whole genome sequences for representative individuals from production lines* and can do so in a matter of weeks or days! This progression marks an exciting era for animal genomics, as diverse phenotypes are connected to genotypes and science translates to practice. However, this transition is also marked by significant new challenges. These include, among others, development of databases suitable for holding and examining, and querying vast amounts of re-sequence data, and development of data and tools to facilitate multilevel functional annotation and curation of genome assemblies to maximize the value of genome-wide association studies (GWAS). While many GWAS studies identify single nucleotide polymorphisms (SNPs) associated with phenotypes in animal populations, such SNPs commonly occur outside of the coding portions of genes. These regions may regulate the expression of genes and are often in more highly variable regions of the genome; thus, these results may not be transferable to another population within the same species or even to crossbred animals including the breed for which the GWAS was performed. One way to increase the value of GWAS results is to identify those variants that impart functional changes. Such variants, if present in other populations, are much more likely to also control phenotypic variation in these other populations. Identifying functional variants/SNPs is one of the objectives of the newly organized Functional Annotation of Animal Genomes (FAANG) consortium, and for which NRSP-8 researchers are integral members. As commercial stakeholders/collaborators increasingly implement genome-enabled animal selection programs, portals for data access and entry of phenotypic information will need to become more sophisticated, and yet more user-friendly. NRSP-8 funding will continue to be needed to address these diverse and evolving national challenges.

The structure, past success, and established interpersonal relationships of NRSP-8 and its members are tremendous assets in facing the new realities of animal genome sciences. Additionally, the broad participation of members with expertise across species and sub-disciplines (e.g., transcriptomics, SNP mapping, functional genomics databases, and bioinformatic tools) allows for rapid dissemination of genome advances and provides a larger combined base of expertise available to the animal genomics stakeholder community. These aspects of NRSP-8 strongly position it for continued success in its mission to enhance the global infrastructure for agricultural animal genomics.

## **RATIONALE**

### *Priority Established by ESCOP/ESS*

NRSP-8 leverages funding to enhance the sharing of genomic resources: tools, reagents, data, animal populations, and bioinformatics. The resulting products support the research community, commodity groups, industry and government regulatory agencies by providing efficient and accurate tools and information on genome sequences, genetic variation, and the relationship of genotype to phenotype that can be broadly applied in many ways. Below is a summary of the relationship of animal genomics research, for which NRSP-8 provides the necessary infrastructure, to help combating the seven food security challenge areas determined by the “Challenge of Change commission.” APLU (Association of Public and Land-grant Universities) established this commission to identify food security challenges and make recommendations on the actions needed to meet global food demands by 2050.

**Grand Challenge 1:** *We must enhance the yields, profitability and environmental sustainability of food and agricultural systems.*

Improvement of feed conversion ratios via genetic selection decreases impact of livestock production upon environment. A key means by which advances in agricultural production can be made is through selection of animals with superior trait performance (e.g., carcass yield, disease resistance, feed- conversion ratio, robustness, athleticism). While applying genetic tools to improve production and resilience of animals, we must account for animal health and welfare, as well as social, cultural, and ethical considerations. Genetic methods of improvement must become more efficient while minimizing environmental impacts. Genomics selection of superior genotypes will help addressing this challenge i.e. enhancing the animal production yields and profitability while reducing the adverse environmental impacts.

**Grand Challenge 2:** *We must develop the breeds needed for sustainable food systems to adapt to rapidly growing human population, along with climate change and weather variability.*

Genome-enabled selection practices need to increasingly account for the effects of genotype on the utilization and assimilation of different animal diets and the potential downstream environmental impacts. Cross-disciplinary research in nutrition and genomics promises tailored diets to maximize genetic growth potential while minimizing environmental impact. Additionally, genomics can enhance selection of animals better adapted for changing climatic conditions. Sustainable management includes efficient livestock utilization of nutrients, water, and space with high survival and minimal environmental impact. Selection of animals best suited for production in a given environment via genomic technologies (precision management systems) optimizes agricultural sustainability.

**Grand Challenge 3:** *We must take measures to decrease food loss and waste through more efficient distribution systems.*

Pre- and post- harvesting environments affect shelf-life, nutritional content and quality of food-animal products. Various genomics approaches allow understanding the biochemical processes through which the animal interacts with the environment pre-, post- and during the harvesting process. Genomics research on local breeds and species will allow us to optimize regionally available food systems. Selection of animals best suited for production in a given environment via genomic technologies (precision management systems) can reduce food loss and waste by creating efficient local distribution systems and by selection for animals that produce products with improved shelf-life and quality.

Also, we must expand genetics/genomics education so that specialists in small urban communities are trained to develop opportunities for urban agriculture to increase availability of nutritious food options in urban environments.

**Grand Challenge 4:** *We must create and share resources that serve all populations.*

We must provide genomics-based advice and education domestically and globally to increase access to the best and regionally most suitable animal management and breeding practices. Animal genomics research must contribute to ‘big data’ for the assessment of needs based on geocoded information on nutritional outcomes and food distribution.

**Grand Challenge 5:** *Ensure inclusive and equitable food systems.*

Healthy, reasonably-priced animal protein choices are the result of vigorous and profitable animal agriculture industries. A profitable and growing agricultural sector leads to strong farm families and farm

communities. Agri-animal genomics is focused on enhancing animal improvement programs to lower production costs, increase throughput and yields and decrease animal disease, thereby increasing profitability. Concurrently, these genomic technologies can add aspects of sustainability and biosecurity to production practices, leading to long-term market stability.

**Grand Challenge 6:** *We must address the dual burdens of undernutrition and obesity to ensure full human potential.*

Animal products are excellent source of proteins, fats, vitamins and minerals. Animal proteins contain all the essential amino acids needed for healthy growth especially in children. Genome-based enhancement of beneficial nutritional profiles of livestock can lead to healthier dietary options and improved human health affected by both undernutrition and overnutrition. In addition, models of overnutrition and obesity in livestock species (e.g. swine, equine) provide a unique opportunity to study the molecular pathophysiology of these diseases and underlying genetic susceptibilities.

**Grand Challenge 7:** *Ensure a safe and secure food supply that protects and improves public health.*

Genomics can enhance the safety of the food supply through animal identity and traceability systems based on molecular markers. Genomics allows rapid selection of lines of livestock resistant to infection, increasing animal welfare while optimizing productivity. Animal genomics research must collaborate with veterinary and human medicine to improve our understanding of the animal-human-environmental health continuum through the ONE Health initiative. Genomics-based selection must gradually reduce and find alternatives for the use of antibiotics in animal production. This is an important measure of treating antibiotic resistance in consumers of animal products.

## **RELEVANCE TO STAKEHOLDERS**

### **A. Stakeholders**

The membership of NRSP-8 includes researchers with an interest in animal genomes, and as such, they constitute the primary group benefiting directly from its activities. However, the beneficiaries are much broader than the participants. Obviously, genotype (the genome) and its relationship to biological function (the phenome) is of critical consideration for many disciplines of animal science whose researchers benefit from NRSP-8 supported resources and knowledge. International scientists are another important stakeholder group (**Figure 1**). The global animal research community, draws on the resources developed by NRSP-8, and the beneficiaries are the global agri-animal industries including commercial breeders, producers, and other entities that rely on agricultural animals. Ultimately, the general public constitutes the ultimate stakeholder group for the efforts of NRSP-8. Thus, it is hard to overstate the impact that accelerated genetic improvement enabled by genomic information has had on decreasing the global environmental footprint of animal protein production.

**Figure 1.** Worldwide distribution of ~3,000 members on AnGenMap server and/or mail list.



## B. Stakeholder Involvement

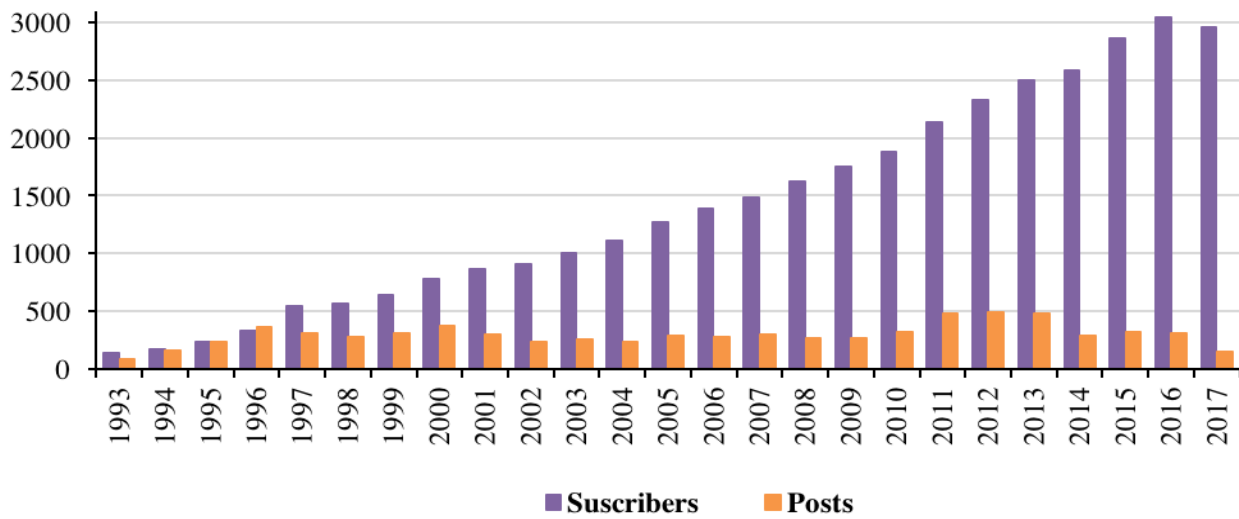
Members participate in governance and all interested parties can have input. One mechanism is via the NRSP-8 supported AnGenMap (<https://www.animalgenome.org/community/angenmap/>), which allows for daily conversation among the ~ 3,000 members, which has grown considerably over time (**Figure 2**). AnGenMap provides each species group with its own website, with links to maps, databases, newsletters, resources and community links, and these are combined at the Bioinformatics Coordination Site ([www.animalgenome.org](http://www.animalgenome.org)). NRSP-8- materials that provide objective science-based information in regard to animal genomics and agriculture are generated by species coordinators and member researchers who engage in education/training, interviews, and meetings/workshops. Each NRSP-8 species committee incorporates industry representatives that participate in governance, and many additional industry scientists and leaders attend PAG, utilize AnGenMap, receive newsletters and employ genomic tools/data deriving from NRSP-8-supported infrastructure.

The annual PAG meeting in San Diego is the largest agri-genomics meeting in the world, and hence, provides a unique opportunity for researchers from around the globe to come together and share research findings, discuss problems and plan for the future. Approximately one-third of the PAG animal scientist-attendees and one-third to one-half of AnGenMap members are international. Out of such discussions have come bold initiatives such as the 1000 Bull Genomes Project and Genomes and Functional Annotation of Animal Genomes (FAANG) Project. Additionally, this meeting attracts industry stakeholders who are keen to understand how genomic developments might be integrated into their genetic improvement programs.



Industries are beneficiaries of the genomic advancements, and a number are now using genomic information to accelerate the rate of genetic improvement. Companies are directly contributing funds, in kind support and genetic resources to help advance genomic understanding. For example, the University of Missouri, in collaboration with GeneSeek, developed an Illumina SNP-chip for genotyping 250,000 locations in the beef cattle genome (GGP F250). The assay was designed using sequence data from the 1000 Bull Genomes Project and over 400 individuals sequenced from multiple taurine breeds, for which some of the sequencing was supported by breed associations who will ultimately benefit from the research. Similarly, companies from other animal industries have been active collaborators and often co-funders of re-sequencing, GWAS, and applied genomic selection experiments of relevance to their industry.

**Figure 2.** Number of subscribers and post per year on the AnGenMap list serve. All data is cumulative per year except for 2017, which is cumulative through August 1, 2017.



### C. Renewal Justification.

As can be seen from the **2013-2017 Accomplishment Report**, previous NRSP-8 funding assisted in genome reference assemblies and improvements for multiple species, with these improved assemblies eliminating most of the regions with ambiguous sequences and in some cases provided sequence for previously unsequenced autosomes (**2013-2017 objective 1**). During this time, one or more SNP genotyping arrays were created for all the major livestock, poultry and aquaculture species and alleles contributing to major economic traits were identified in all species (**2013-2017 objective 2**). In addition, bioinformatics platforms to facilitate collaborative research were developed including a collaborative VCF information-mining platform (**2013-2017 objective 3**). From 2013-2017, NRSP-8 participants leveraged the NRSP-8 investments in tools and infrastructure with more than \$94.5 million dollars of research funding towards investigating fundamental mechanisms of genome biology and physiology and pathophysiology affecting production efficiency, product quality, animal health, disease resistance and food safety.

This renewal seeks to advance the progress enabled by the initial sequencing efforts. It is hard to imagine that just 10 years ago, there were no high-density SNP chips being used in animal breeding programs. The

implementation of genomic selection is now routine in many industries, and over a million US dairy cows have been genotyped. Despite these successes, there are still vexing problems that remain in translating genomic data into useful information. Although many species now have reference genome assemblies, there remain significant gaps and mis-assemblies that complicate analysis and the discovery of genome to phenome relationships. Additionally, the importance of regulatory elements and noncoding regions of the genome on quantitative traits is becoming increasingly evident. Data from the human research community suggest that more than 80% of the SNPs associated with disease traits are in regulatory elements and non-coding regions. The ultimate goal of NRSP-8 is to understand genomic variation and translate that basic mechanistic research into application. Ideally the three objectives outlined in this renewal will result in new knowledge that will improve the utility and predictive ability of genomic information. As detailed below, research described under Objective 1 will be a particularly important contribution to the goals of the FAANG consortium by better identifying and annotating the functional elements that are important in regulating gene expression and ultimately phenotype. Objective 2 outlines specific aims that will be important in understanding trait variation and its genetic causes, and simultaneously developing agricultural animals as biological model systems while maximizing the economic gains from genetic and genomic selection. The advent of genome editing tools opens up an exciting opportunity to precisely investigate and dissect the genetic variation associated with quantitative trait variation. It is clear that the connection between the genome and phenotype is complex. Consequently, a wide group of scientists with diverse skills and expertise with access to high-throughput genomic tools and bioinformatics must work together to achieve an understanding of this connection in order to generate practical results of value for animal agriculture and human health. Objective 3, the analysis, curation, storage, application, and reuse of heterogeneous big data to facilitate genome-to-phenome research, is perhaps the most challenging. How will the tremendous amount of sequence data that is being generated, be curated and harvested into useful information? As commercial stakeholders/collaborators increasingly implement genome-enabled animal selection programs, data access and entry of phenotypic information will need to become more plug and play. This challenge of taking big genomic data and basic science and producing actionable, predictive information is perhaps the grand challenge of our time. While previous iterations of NRSP-8 were essential to produce the genomic information now available, these next five years will be the real challenge – and that is how to integrate data and systems thinking to produce knowledge that can be translated into the desired outcomes.

## OBJECTIVES

**Objective 1:** Advance the quality of reference genomes for all agri-animal species by providing high contiguity assemblies, deep functional annotations of these assemblies, and comparison across species to understand structure and function of animal genomes.

Recent advances in second and third generation sequencing technologies have enabled the generation of draft reference genome sequence assemblies for almost all economically important agricultural animals, as described in the Accomplishments Report from the current NRSP-8 cycle. However, these genome assemblies still contain numerous gaps and local mis-assemblies, especially for highly repetitive sequence regions, such as those found in centromeres. In some cases, whole chromosomes are absent/unassigned, such as the Y chromosome in some mammals and several micro-chromosomes for all the sequenced birds. Functional annotation, at the transcriptomic, proteomic and epigenetic levels, of accurate reference assemblies are essential for most genome-wide analyses, including the dissection of the genetic architecture of complex traits and enhanced breeding using genomic selection strategies. Furthermore, identifying conserved genomic elements across animal species will provide insights into gene function and underlying regulatory mechanisms. This objective advances the quality of reference genomes for all agri-animal species

through providing deeply annotated, high contiguity assemblies, obtained through comprehensive and integrated analyses of transcriptomic data with chromatin architecture and modification data across a wide range of tissues/cells and biological states. This Objective will provide a resource to enhance the research specifics in Objectives 2 and 3.

The specific aims for this objective include:

1. Initiate creation of draft genome assemblies for economically important species and breeds for which assemblies are not yet available.
2. Improve existing genome assemblies to close gaps and improve assembly order, especially for highly repetitive sequence regions, such as those in centromeres and on the sex chromosomes.
3. Coordinate analysis of re-sequencing data to identify SNPs, CNVs, and InDels, and better annotate gene models within the reference assemblies.
4. Develop and apply approaches for identifying and functionally annotating regulatory elements through comprehensive and integrated analyses of transcriptomic data with chromatin architecture and histone modification data across a wide range of tissues and cells at different biological states.
5. Associating functional information (e.g., tissue expression, physiological processes and interaction information) with regulatory and expression elements in the genomes.
6. Compare structural and functional components of genomes across animal species to understand biological function through identifying conserved genomic elements.
7. Expand the training of students and postdocs in application of next-generation technologies to structural and functional animal genomics.

**Objective 2:** Advance genome-to-phenome prediction by implementing strategies and tools to identify and validate genes and allelic variants predictive of biologically and economically important phenotypes and traits.

The goals of this Objective are to advance genome-to-phenome prediction by facilitating translation of genomic findings into biologically relevant information for genetic improvement of all species. To achieve this goal, genome-phenome association studies, functional validation methods, deep phenotype collection, comparative genomics, metagenomics and phenomics are important approaches that will produce valuable insights. In addition, analytic tool development to effectively leverage all information as well as implementation will be critical for agri-animal research communities to realize the potential of these findings to simultaneously develop agri-animals as biological model systems while maximizing the economic gains from genetic and genomic selection.

The specific aims for this objective include:

1. Exploit the power of high-resolution SNP-chips, genome re-sequencing, and genotyping-by-sequencing in genome-phenome association studies for detection and validation of genomic variants that are predictive of economically important phenotypes.
2. Advance strategies, models, algorithms, pipelines and analytic tools to facilitate the identification, validation and incorporation of novel genetic elements and causal and/or highly predictive variants to allow the most accurate prediction of phenotypic performance based on genotypes. Newly identified variants will be used in Objective 1 to further annotate genomes.
3. Develop and adopt various approaches, including (e.g., CRISPR/Cas9 genome editing) for functional analysis, and evaluation and verification of functional allelic variants of causal genes important for production traits.
4. Support deep phenotyping of important traits at the molecular, cellular, tissue and organismal levels, including the use of high-throughput technologies such as transcriptome sequencing, proteomics and

metabolomics studies, whole-animal parameter monitoring, and in vitro gene mutagenesis screening analyses and other tools to support precision monitoring.

5. Use comparative approaches to identify genetic variation within and across species that is associated with phenotypic variation that results from common treatments or environments, or between natural and domesticated populations.
6. Support comparative phenomics, with an emphasis on the use of farm-animal models that mimic human processes to benefit human health.
7. Develop well-characterized animal populations/genetic lines/models that allow for study of biology of various economically important phenotypes and traits.
8. Advance metagenomic studies to help in discovery of novel pathogens, understanding host-pathogen interaction and determining the role of microbiota in agri-animal nutrition, health and reproductive performance.
9. Train the next generation of animal breeders in applying and developing new methods based on high-throughput genomic data to make genetic progress.

**Objective 3:** Advance analysis, curation, storage, application, and reuse of heterogeneous big data to facilitate genome-to-phenome research in animal species of agricultural interest.

The genomic “Big Data” era is here. In the last five years, advances in next-generation DNA sequencing technologies have allowed to produce enormous amounts of data in all livestock, poultry and aquaculture species. To better understand the genetic mechanisms that underlie important traits, in the next five years agri-animal scientists will direct tremendous attention to data refining in addition to data collection. To achieve this goal, bioinformatics/computational tools, resources and expertise will be necessary on an unprecedented scale. The bioinformatics team of the NRSP-8 project will provide in-house bioinformatics tools, databases and resources in addition to assisting researchers to utilize other open-source tools and resources housed elsewhere. The team will provide resources for “real-time communication” to facilitate research coordination needed for Objectives 1 and 2. Furthermore, Objective 3 will address the need to train students/postdocs to be future leaders in agriculture-oriented computational science.

The specific aims for this objective include:

1. Facilitate livestock/poultry/aquaculture genomic research by development/promotion of data sharing standards, workflows and tools necessary to integrate these resources.
2. Facilitate communication and training within the livestock, poultry and aquaculture groups to promote community discussion and awareness of community needs, current events, available resources, and other items of interest.
3. Facilitate genome-wide research approaches to understand the genotype-to-phenotype basis of important traits in livestock, poultry and aquaculture through providing opportunities to curate and reuse data created by the community.
4. Help in training students/postdocs to be future leaders in agriculture-oriented computational science.

## **MANAGEMENT, BUDGET, AND BUSINESS PLAN**

### **A. Management and Activities Plan**

The purpose of NRSP-8 is to provide and/or enhance infrastructure for the animal genome community as a whole. NRSP-8 accomplishes its mission by working with a structure that is organized by animal species and managed by species-specific coordinators and co-coordinators. In addition, due to the need to address

the critical and rapidly growing need for data management and analysis, the infrastructure includes a bioinformatics coordinator and co-coordinators. These individuals facilitate the availability of shared genomic resources (e.g., tools, reagents, protocols, samples, and data) to the members of each species committee, often by leveraging the limited resources of the coordinator with those available from federal (e.g., NIH, USDA AFRI), international, industry, and state sources. The coordinators and co-coordinators, working with the members of their respective committees and amongst themselves, help to develop, stimulate and coordinate new initiatives by providing leadership that is recognized nationally and internationally. Finally, NRSP-8 fosters communication among its members and stakeholders through, among other things, e-mail newsletters and participation in meetings and conferences. Species coordinators and co-coordinators stimulate this communication and provide focal points for input from the committee members and other stakeholders.

The coordinator and co-coordinator model has served NRSP-8 well; it is flexible, and the structure allows for continuous refinements and shared responsibility. Coordinators and co-coordinator will meet with their respective species committee members at least once each year at annual meetings (e.g., typically the species workshop at PAG) and solicit ideas and plans for project development and resources to share. Coordinators will communicate by e-mail, newsletters, and other means to make resources known to community members and solicit input throughout the year. Coordinators and co-coordinators will meet with the NAGRP Director and Administrative Advisors at least once each year and as needed through conference calls and e-mail communications. Coordinators may employ subcommittees within each of the species committees to take on special tasks or to provide recommendations to the coordinator. The frequency of communication within and across each species group affords many opportunities for periodic review of the current management plan and potential revisions. This latter point is critical given how quickly the field of genomics advances and how new technologies can alter plans and objectives.

Application for species coordinators and co-coordinators will take place prior to funding, or when needed during the funded period. Either individual coordinators or co-coordinator groups may apply. When coordinators change, the out-going coordinator will work with the new coordinator to achieve a smooth transition, or in the case of group efforts, some co-coordinators may stay on while others change. In addition, each species committee has rotating chairs that interact with coordinators and may be suitable replacements or may assist during transitions. Applications will be solicited by the NIFA administrator 60 days in advance of the start date for an approved renewal proposal and be widely publicized. Applications shall include a 2-page letter of application and intent, complete *curriculum vitae*, and up to five letters of recommendation. The NIFA Administrator and the Regional Administrative Advisers shall review applications and select the coordinators based on merit, experience, and involvement within the species community.

As a federally-funded organization, we strive to make all reagents and information available to the entire public as soon as possible. Of particular note is the Bioinformatics Coordination webpage at [www.animalgenome.org](http://www.animalgenome.org). Due to growing inputs from companies and other organizations regarding intellectual property, efforts will be made to encourage sharing of information while still protecting their proprietary information, when needed.

## **B. Budget**

We request support of \$500,000 per year. Funds for the Bioinformatics Coordinator (\$110,000 per year) will be used for: (1) \$80,000 in salary costs for the individual that maintains and develops resources at [www.animalgenome.org](http://www.animalgenome.org), (2) \$20,000 in salary cost for a part-time post-doctoral fellow to support

bioinformatics analysis of FAANG related data, (3) \$3,000 to support host servers, and (4) \$7,000 to support Coordinator and PAG student travel. For details, see appendix titled Budget.

Each of the six Species Coordinators will receive \$65,000 per year. The requested funding will be used by the coordinators to provide for communication and sharing of resources to enhance the progress of the membership of the overall Technical Committee. Funds will not be used to directly fund their personal research programs. Based on past history, these funds likely will support the development and annotation of their respective genomes, subsidizing the generation of shared reagents (e.g., SNP chips), support to produce preliminary data for grant submissions, travel to PAG or other relevant meetings (**see Accomplishments Report**). These resources and services are essential for genomics-driven research, help form cohesion within each species group, and often provide preliminary data for larger efforts in the future. As noted in the Integration and Documentation section, the coordination funds are heavily leveraged, in the sense that they assist or provide seed resources that eventually generate Federal grant and/or industry funding to members. While many genomics tools have become more efficient in terms of cost-per-data unit, they still require a substantial expense to initiate and, thus, require coordination within each species community, as each has different needs. It is also important to note that, as a federally-funded organization, we strive to make all information generated available to the entire public as soon as possible.

### **C. Business Plan**

Species and bioinformatics coordinators work with the NIFA Director of the NAGRP each year to justify current and past expenditures and plan for future budget periods. Although the requested budget outlines the proposed distribution, this can be altered through discussions between the NAGRP Director and the coordinators, if special needs arise. The NAGRP Director in consultation with the Administrative Advisers makes the final decision annually as to the distribution of funds among the coordinators. As noted above, each coordination group will receive funding through a single Experiment Station budget affiliated with the coordinator or one of the co-coordinators. In addition to flexibility in annual budgets among NRSP-8 coordinators, we request that all coordinators (or coordinator groups), with permission of the relevant Experiment Station Director, be afforded the option of carrying funds over from one fiscal year to a subsequent fiscal year within the proposed 5-year project term. Opportunities to provide genomic resources change from year to year, and coordinators cannot predict what tool or resource will be most effective and useful to the genome community 3-5 years from now. Not only do the tools and resources change as new technology advances, but frequently a critical opportunity will require more support than is feasible within a single FY budget to a single coordinator. A carryover option would allow for more responsible allocation of resources within the overall 5-year plan.

## **INTEGRATION AND DOCUMENTATION OF RESEARCH SUPPORT**

### **A. Integration with extension or academic programs and use by other potential stakeholders**

Modern agriculture will require revolutionary steps to meet projected demands in the 21<sup>st</sup> century. Contemporary high-scale basic and applied agricultural genetics and genomics research necessitate high-level multidisciplinary collaboration between academic, government, private industry and international partners. Most projects are performed by multidisciplinary and collaborative teams consisting of experts in molecular and advanced genomic technologies, data management and statistical analysis of large “omic” datasets, specific phenotyping of large populations, and devising feasible downstream applications. External and internal research proposals to private, institutional, or federal funding agencies are typically evaluated to some degree on the quality, integration and specific components of the research team. Federal

funding opportunities sometimes require involvement of multi-disciplinary research teams. In addition, involvement of stakeholder, especially industrial, perspectives is very important. This integrated research team approach greatly strengthens the scientific discovery process and will continue for the foreseeable future of agriculture research.

The main mission of the NRSP-8 is to provide a structure in which collaborative animal genomics research teams can be established and succeed, and in which both fundamental and applied knowledge can be generated in the most efficient fashion. To accomplish this aim, six main research consortia for aquaculture, cattle, equine, poultry, sheep, and swine, joined by a bioinformatics team, meet regularly and continuously to share data, generate knowledge and provide expertise. Results of this animal genome research are presented at multiple regional, national, and international meetings, and published in a wide variety of peer-reviewed, publically -available journals. Furthermore, industry groups, in communication with applied and extension experts, will continue to be informed, acquiring major results and the translational implications of our research via trade journal articles written by scientists familiar with their reader's backgrounds and interests. We predict that the NAGRP bioinformatics website, individual species' websites, and genomics-related websites such as FAANG ([www.faaang.org](http://www.faaang.org)), will be increasingly utilized by research teams to communicate results to the public and, on a case by case basis, provide education materials to stakeholders. This means of communication will be actively encouraged by the species coordinators.

The NRSP-8 members serve the USDA educational mission. It is important to note that the vast majority of NRSP-8 members are in academicians and engage in teaching, extension services, and outreach as part of their daily responsibilities. Information and, in some cases, the support that these members receive from NRSP-8 assist them in these activities. Most NRSP-8 research is accomplished through the efforts of undergraduate and graduate students, as well as postdoctoral trainees. Thus, the collaborations generated by NRSP-8 directly influence the training and careers of these junior scientists who clearly represent our future scientific leaders. As noted below, the coordinators actively encourage and financially support the specific attendance of such lab trainees at our annual meetings and other relevant venues, thus providing opportunities for training the next generation animal breeders.

## **B. Partnerships and engagement**

The framework and partnerships generated in the existing NRSP-8 project will continue to be utilized and in many cases enhanced. Species-specific collaborations, websites, and databases, particularly within the NAGRP bioinformatics program, will be maintained. The NRSP-8 bioinformatics team has been receptive to users' suggestions and needs, and the current portfolio of sites, programs, and databases is a direct result of these interactions. Our programs will increasingly evaluate and use other platforms, such as CyVerse, to assist with the complex and computationally intensive tasks using supercomputing cyber-infrastructure. This program takes advantage of the more universal aspects of high-throughput sequence and genotype analysis to provide a common means to perform our work and enable data sharing. This area is rapidly changing, and we are open to evaluating new systems for data management and analysis that meet our needs and promote efficient sharing of data and technologies. This is especially true for the International FAANG Consortium. It is expected that significant amount of epigenomic data will be generated and shared with the animal genome community nationally and internationally.

The ability of the NRSP-8 program to leverage acquisition of additional animal genomics research funds is supported by data indicating that between 2013 and 2017 NRSP-8 members leveraged funds to acquire an additional **\$84,292,482** in federal, **\$5,753,434** in industry and **\$4,440,844** in internal funding to support

animal genomics research. These successes are encouraging and likely indicate our success in providing information that industry sees the generated information as being essential to achieving their goals. Further, NRSP-8 researchers were instrumental in additional new funding being directed to animal genomics research through the USDA NIFA AFRI program; in 2018, it is expected that \$5 million will be provided for development of resources for agri-animal functional genomics research associated with the FAANG consortium.

Species coordinators and, indeed, many NRSP-8 members have close relationships with relevant animal agriculture industries and engage in joint projects with, or receive financial support from these companies. As noted below, industry is well represented and actively participates in our annual meetings. In the past five years, many breeding companies have embraced genomics methods (dense SNP genotyping, genomic selection, etc.) and have directly contributed to NRSP-8 efforts. As a result, NRSP-8-industry cooperation has grown in parallel, and we anticipate that this trend will continue.

### **C. Support of research activities nationwide**

We will continue to utilize annual species group meetings, currently associated with the annual PAG conference. With a total annual attendance of ~3,400, PAG is the largest international scientific conference reporting on non-human genomics research. Over the course of five days, numerous opportunities are available for presenting and discussing current and future research and promoting essential interactions within and across species groups. This has often been the means by which multi-disciplinary research teams are brought together and the momentum of established teams is maintained. The NRSP-8 and PAG meeting has also been a major opportunity for the animal genomics community to acquire knowledge on advances and the newest product offerings from the private sector, including the latest high throughput sequencing and genotyping technologies and the means of storing, analyzing, and sharing data. The PAG meeting is also an efficient mechanism by which the needs of the research communities and the stakeholders can be brought to the attention of the species coordinators for group discussion and potential support from their coordinator funds.

The PAG meeting attendance has increased by 30% since 2010, and total scientific attendance in 2017 was 3,399, with 57% U.S. attendees and 43% international attendees. Of those attendees, 776 indicated an affiliation with the animal genome programs. Overall, >1,700 scientific abstracts were submitted, and 121 scientific and 27 industry workshops were organized. Animal-related workshops at the 2017 PAG conference included Aquaculture, Buffalo Project, Cat & Dog, Cattle/Sheep/Goat, Camelids, Equine, Poultry, Cattle/Swine workshop, Animal Epigenetics, Animal Genomics and Adaptation to Climate Change, Avian Genomics-Going Wild, Domestication Genomics, and the multi-species NRSP-8 Animal Genome workshop. General workshops of interest to animal genome researchers included, but were not limited to, Functional Genomics, Gene Expression Analysis, Gene Introgression, Host-Microbe Interactions, NCBI, Sequencing Complex Genomes, and FAANG. In addition to NRSP-8 lab scientists, the various relevant commodity groups and Federal agencies are well represented at the PAG meeting and are encouraged to participate in our meetings. Representatives of national media are also in attendance.

The species coordinators have traditionally invested their funds in the promotion of communication and efforts that enhance the entire group's effort (see Accomplishments report). Examples include partial PAG meeting travel support for graduate and postdoctoral students, support of the AnGenMap forum, support of scientific meetings of a more specific nature that occur outside of the main NRSP-8 venue at PAG, and specific activities that generate essential group tools and resources, such as SNP panels and arrays, HapMap data, microarrays, and now certain essential whole genome and/or RNA sequencing, and



epigenomic data. We envision that this flexible and diverse use of funds will continue.

## **Outreach and Communications Plan**

**Target Audience.** The primary beneficiaries of NRSP-8 include:

1. Research scientists directly engaged in animal genome research.;
2. Scientists utilizing genomic data in complementary areas of animal science and veterinary medicine.;
3. Graduate and postdoctoral students who benefit from training opportunities in animal genomics.;
4. Commercial animal breeders and producers who apply genomic data and related technologies to enhance animal health, welfare, productivity, and management.: and
5. Government agencies and breed organizations utilizing genomic methods for animal identification, disease surveillance, and regulatory monitoring.

Support of research, education, and communication activities in animal genomics also benefits all consumers of animal agriculture through resulting improvements in the quality, safety, wholesomeness, economic efficiencies, and environmental soundness of animal products and production systems.

**Engagement of Target Audiences.** Research scientists will continue to be engaged through NRSP-8 visibility at national and international scientific meetings/workshops. The leading example of this engagement is the annual International Plant and Animal Genome (PAG) Conference. NRSP-8 is showcased in species-specific and comparative genomics sessions. In each species-specific workshop, station reports are required from scientists at participating Experiment Stations. NRSP-8 has held its annual meeting jointly with PAG for 16 years and NRSP-8 representatives play a major role in the organization and programming for this high-profile conference. PAG is also a primary avenue for engagement of the next generation of animal genomic researchers. Each species supports a competitive award for graduate student travel to the annual PAG conference; and most species coordinators also provide additional travel support for graduate students and postdoctoral fellows presenting their work. Participation by students at this conference ensures early exposure to animal genomics research and its applications. Additionally, the Poultry Genome Coordinator, Swine Genome Coordinator, and Aquaculture Genome Coordinator each publish separate newsletters that go well beyond the actual members of NRSP-8 and are circulated via the AnGenMap listserv to about 3,000 scientists, students, and extension specialists.

**Engagement of Stakeholders.** The genome committee for each animal species will have one or more representatives from an industry group or other appropriate stakeholder selected in consultation with national industry organizations. Industry participants will be invited to attend all technical committee meetings and serve as liaisons between industry and the committee and the species coordinators. Species coordinators will continue to present and interact at national and regional producer meetings, breed organizations and other appropriate venues. Coordinators will also facilitate communication of research outcomes to the general public. Currently the poultry, swine and aquaculture coordinators circulate quarterly genome newsletters. Moving forward all species coordinators will provide quarterly newsletters applicable to target audiences and stakeholders. The AnGenMap discussion group will also be used to engage stakeholders worldwide.

**Measurement of Accomplishments and Impacts.** Scientific productivity is the most important measurements of NRSP-8 impact. The species committees in conjunction with the species coordinators will document scientific productivity in an annual progress report. These reports will detail the scientific accomplishments including:

1. Published peer-reviewed scientific papers;
2. Presentations at scientific meetings by coordinators and NRSP-8 members;
3. Scientists trained in animal genomics and bioinformatics by NRSP-8 members while contributing NRSP-8 projects;
4. Attendance at the PAG species workshops, the PAG NRSP-8 workshop and the technical committee meeting; and
5. Additional funding obtained by NRSP-8 members from leveraging NRSP-8 investments in animal genomics.

**Communication of Activities, Accomplishments, and Impact.** Annual technical committee meetings facilitate interactions between scientists working with different animal species. NRSP-8 will continue to coordinate its meetings with other multi-state technical committees and with scientific conferences and workshops including the PAG meeting. NRSP-8 members communicate advances in genomics through invited presentations at scientific conferences, offering information and technology transfer among international scientists across species groups and with researchers in public and private sectors. AnGenMap, an online discussion group for the animal genomics community, includes about 3,000 members from more than 50 countries/regions, continues to be used for communication among scientists (**Figures 1 and 2**). Each species has a genome page established by NCBI that includes access and direct links to existing genomic databases and related resources.

Other mechanisms of communication include the species-specific quarterly genome newsletters that communicate scientific advances to both scientists and interested non-scientists. Scientific advances are disseminated to the public through popular press articles, webinars and industry presentations. Species-specific genome webpages reach agricultural scientists, industry members, consumers, students and others interested in the application of genomics to animal agriculture. These homepages, constructed and maintained by individual animal species coordinators, frequently provide additional information and helpful links to genomic resources within the agricultural research community. Coordinators and other members of NRSP-8 promote advances in genomics for their respective species through the development and promotion of meetings involving stakeholder groups. NRSP-8 leaders will continue to participate in the writing of white papers and other activities to advance animal genome research. Such activities have proven critical in attracting the cooperation of the NSF, NIH, international consortia, industrial partners, and breed organizations for animal agricultural genomics endeavors.

**Distribution of Research Results.** The primary in-house mechanism for distribution of results and accomplishments will be via annual reports (available on NIMSS). The technical committee chair prepares summary reports with input from the species chairs and species coordinators. In addition to reporting results and accomplishments, annual reports outline future work plans for each species. Other reporting mechanisms include peer-reviewed research publications and reviews, as well as conference presentations. A majority of the requested funding is for species coordinators to distribute among community driven projects. To better promote and share these project results, they will be featured on the [www.animalgenome.org](http://www.animalgenome.org) site with direct links to any genomic data that is generated from these projects. This will extend the impact of these projects and promote interaction among species groups.

### **Past Successes**

NRSP-8 impacts are best measured by the major milestones in animal genomics as listed in the attached Accomplishments Report. Most targeted species have advanced past the whole genome sequencing phase

with the position of genes determined to a large extent, allowing dissection of the functional polymorphisms responsible for the variation of economically important traits. This reflects extensive coordination of effort among scientists, facilitated in no small part by effective outreach and communication strategies of NRSP-8. The project has made good use of its annual technical committee meetings, species and comparative workshops, newsletters, and internet home pages to foster effective communication between member scientists and other stakeholders worldwide. These efforts have achieved the advancement of animal genomics at a pace that has substantially exceeded most predictions.