Charting the Road Map for Long-Term USDA Efforts in Agricultural Animal Genomics: Summary of the USDA Animal Genomics Workshop – September 2004

R. D. Green[1], M. A. Qureshi[2], and J. A. Long[1]

[1]USDA-ARS, Beltsville, Maryland, USA
[2]USDA-CSREES, Washington, District of Columbia, USA

INTRODUCTION

Agricultural animal research has been immensely successful over the past century in developing technology and methodologies that have dramatically enhanced production efficiency of the beef, dairy, swine, poultry, sheep, and aquaculture industries. Most of this research effort has been conducted within the broad disciplines of genetics, physiology, and nutrition, but has become increasingly narrowly defined into multiple sub-disciplines over time. While a vast body of knowledge has been generated from this effort, it has now become clear that the majority of the potential for future improvements in animal production efficiency, quality of animal products, and animal health lies in the elucidation and understanding of interactions of the various components of the biology of the animal in concert with all of the parameters of the production environment. To begin to fully understand these interactions, a redirection of the traditional "reductionist science" approach to a "systems biology" approach is required.

In the past two decades, molecular biology has changed markedly the face of agricultural animal research, primarily in the arena of genomics and the relatively new offshoot areas of functional genomics, proteomics, transcriptomics, metabolomics and metagenomics. Publication of genetic and physical genome maps in the past 15 years has given rise to the possibility of being able finally to understand the molecular nature of the genetic component of phenotypic variation. While quantitative geneticists have been remarkably successful in improving production traits, genomic technology holds potential for being able to lead to more accurate and rapid animal improvement, especially for phenotypic traits that are difficult to measure.

Recently, the agricultural research community has been able to capitalize on the infrastructure built by the human genome project (Collins et al., 2003; Gibbs et al., 2005) by sequencing two of the major livestock genomes (*Gallus domesticus* (Wilson et al., 2004; Andersson et al., 2004) and *Bos taurus* (Gibbs et al., 2002)). The 2005 calendar year is truly unprecedented in the history of agricultural animal research since annotated draft genome sequences were completed for chickens and cattle and draft genome sequences will be in progress for swine in early 2006. We now have in place the foundation of a powerful toolbox for understanding the genetic variation underlying economically important and complex phenotypes.

Over the past few years, new challenges have emerged for animal agriculture. Enhancements in production efficiency have not come without some negative side effects on animal well-being and longevity in production environments, including losses in reproductive efficiency, increased stress susceptibility, increased animal waste issues, and increased susceptibility to animal metabolic and infectious diseases. When considered in concert with increased societal concerns in the areas of natural resource conservation and protection, animal welfare, and food safety, it is clear that publicly supported agricultural research must be focused on enhancing the functionality and well-being of livestock and poultry in environmentally neutral production systems in the future.

Realizing the great potential for animal genomics to address these and other issues, a workshop was convened by the U. S. Department of Agriculture (USDA) in Washington, DC in September of 2004. The workshop was entitled *"Charting the Road Map for Long Term USDA Efforts in Agricultural Animal Genomics"*. The objective of this paper is to summarize the proceedings of the workshop and the resulting recommendations.

BACKGROUND

The 20th century was an immensely prolific time in developing methods to enable genetic improvement in livestock and poultry. In the first half of the century, geneticists who were busy trying to understand the nature and behavior of chromosomes converged with biometricians who had developed considerable statistical theory that could be used to describe variation observed in genetically defined populations of animals and plants. These two fields coalesced into what became known as "quantitative genetics" around the time of World War II. Over the following forty years, sophisticated genetic prediction methodologies were developed for the dairy, beef, swine, sheep, and poultry industries for a number of the economically important production traits measured on seedstock populations of animals. This body of work was based upon the assumption that differences observed in performance phenotypes could be attributed to underlying heritable genotypic differences. It was assumed as well in these approaches that for most of the economically important traits (e.g., milk yield, growth rate, meat yield, etc.) a large number of unknown genes contributed to this heritable variation, which led to breeding value predictions based on an infinitesimal gene model "black box" approach. An entire "genetics industry" developed around this framework and created many tools, such as predicted differences in dairy cattle and expected progeny differences in beef cattle. Even though it was probabilistic in nature, this approach has been highly successful in allowing directed genetic change to occur in all of these species. An excellent example of this approach is the coupling of genetic selection and efficacious use of artificial insemination over the past 50 years that led to a more than 100% increase in annual milk yield per cow.

In the mid-1980s, a new window of opportunity opened in livestock production science. In 1986, the new term *"genomics"* was coined to refer to the new technologies that were developed and applied to the study of mammalian DNA, such as the application of bacterial restriction endonucleases for rudimentary visualization of differences in the sequence of DNA in particular chromosomal locations through "restriction mapping". This was followed

quickly by the development of the polymerase chain reaction in 1987 that opened up an entirely new world for the study of differences in the DNA sequence of animals. Coupled with the discovery of short tandem repeat DNA markers, PCR became a powerful tool that quickly allowed the development of genetic maps of the livestock genomes, primarily based on linkage of microsatellite DNA markers.

In 1990, an Allerton Conference entitled "*Mapping Domestic Animal Genomes: Needs and Opportunities*" was hosted by the University of Illinois. This conference provided the first opportunity for scientists, producers, industry, and government representatives to come together to discuss how emerging molecular technologies could be employed to bring about major innovations for animal agriculture. Participants at this workshop recommended to USDA that genetic maps be developed to a 20 cM saturation level for each of the agriculturally important species (cattle, swine, poultry, fish, and horses). This recommendation was implemented and resulted in the publication of a number of important genetic linkage maps in the mid-1990s.

With the initial genetic maps in place, a second Allerton Conference, entitled "*Genetic Analysis of Economically Important Traits in Livestock*", was convened in 1996 to address capitalizing on animal genomics research. The workshop focused on statistical approaches to mapping complex quantitative traits and discerning how any DNA markers identified through such mapping could be used in selection programs. By this time, a number of research groups around the world had developed resource family populations that were being employed, using the previously developed linkage maps, to identify regions of the genome appearing to harbor genes giving rise to phenotypic variation in complex traits (so-called Quantitative Trait Loci (**QTL**)). Once DNA markers anchoring these QTL regions were identified, it was postulated that "marker-assisted selection" could be used to make directed genetic change in the desired traits using this technology. The primary recommendation of the Allerton II workshop was a call for building the research infrastructure necessary to enable researchers to identify important genes that control economically important traits and, eventually, gain an understanding of the function of individual genes and their interactions across the genome.

By the end of the century, everyone recognized that more genomic tools and resources were necessary for the fulfillment of the promise of livestock and poultry genomics. Even though a large number of putative QTL had been identified for a wide spectrum of traits, only a handful of simply inherited traits had been elucidated through this approach. In all of these successful cases, the fine mapping of the identified genes had relied on comparative mapping approaches to make use of the denser information available in the human, mouse, and rat maps. Despite having some improved tools, such as bacterial and yeast artificial chromosome libraries, it became clear that without the availability of the whole genome sequence as a scaffold from which to work, the time and expense of fine QTL mapping was much greater than initially envisioned. Fortunately, new high-throughput technologies were being developed that made the sequencing of whole genomes more practical, efficient, and cost effective. The human genomics research community quickly recognized this opportunity and the government and privately funded human genome sequencing projects were launched.

As the 21st century began, and the human genome moved toward an initial draft sequence, other new technologies also became available that allowed livestock and poultry researchers for the first time to move into large-scale gene expression studies. By coupling expressed sequence tags with new microarray technologies, researchers were able to visualize changes in levels of expression of hundred or thousands of genes in specific tissues under a wide variety of conditions. This began to broaden genomics research into the functional realm and initiated open discussions on how genomics might be used to bridge various disciplines into a "systems biology" framework.

In 2001, the Alliance for Animal Genome Research was formed by a group of universities, private industry parties, producer groups, and scientific societies to advocate for public funding for domestic animal genomics. This group was successful in working with the National Academy of Sciences to organize a public workshop held in 2002 entitled "*Exploring Horizons for Domestic Animal Genomics*" with the goal of identifying research goals and public and private funding needs (National Academy of Sciences, 2002). There was overwhelming consensus at the workshop that funding should be identified to produce high-coverage, draft genome sequences of the major domestic animal species (cattle, chicken, swine, dog, and cat) for deposit into the public domain databases. NHGRI had previously established a process for prioritizing species for sequencing based upon the ability of a species to better inform annotation of the human genome sequence through evolutionary comparisons. The workshop participants felt that these species would be excellent candidates to meet that objective in addition to the fact that they had been used heavily as biomedical models and were important agricultural or companion animal species. Furthermore, it was recommended that there would need to be appropriate scaling-up of bioinformatics resources to make effective use of the volumes of information that would result from the genome sequencing projects. Based upon the experiences of the National Plant Genome Initiative, it was also recommended that funding for such large-scale projects would need to come from a variety of sources, including the U.S. Federal government, private industry, and international partners.

In July of 2002, a third Allerton Conference entitled "*Beyond Livestock Genomics*" brought together leading investigators from a broad spectrum of disciplines (genetics, physiology, reproduction, animal health, and nutrition) to develop an initial plan for the full utilization of genomic information to promote animal health and productivity, while more broadly contributing to the greater life sciences. The overarching recommendation from this workshop was that additional basic research was needed to identify genomic mechanisms and novel genes / proteins in a variety of tissues under a variety of environmental conditions (Hamernik et al., 2003). Functional genomics was recognized as the vehicle for capitalizing on the investment of obtaining whole genome sequence information. The need to increase bioinformatics infrastructure and teaching and outreach efforts in animal genomics was recognized also.

In response to a request by USDA Undersecretary Joe Jen, a new Interagency Working Group (IWG) on Domestic Animal Genomics was chartered in September of 2002 by the U.S. National Science and Technology Council with the mission of enhancing

communication and awareness of the importance of livestock and companion animal species of importance to the food and agriculture system; increasing leverage of Federal investments in large-scale genome sequencing and genome analysis across government agencies; positioning the food and agriculture system as a critical element of the national genomics program; enhancing dialogue and cooperation among Federal agencies, universities, and industry in the nation; and promoting international cooperation on domestic animal genomics research. The membership of the IWG has included representatives from the Department of Agriculture (USDA), Department of Energy (DOE), Food and Drug Administration (FDA), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), Office of Management and Budget (OMB), Department of Homeland Security (DHS) and the U.S. Agency for International Development (USAID). The IWG subsequently identified the following broad strategic goals:

- Bring into place the programmatic elements needed to advance the study and understanding of domesticated animal genomes, including large-scale DNA sequencing; functional characterization of expressed genes (functional genomics); tools for data storage, analysis and visualization (bioinformatics); and study of similarities among genomes of different species (comparative genomics).
- Leverage the national infrastructure for large-scale DNA sequencing that has been established for the Human Genome Project and other vertebrate and model organism genomes.
- Advance and utilize the enabling tools and infrastructure of functional genomics and bioinformatics to enhance the understanding not only of basic science and disease mechanisms, but also to address critical agricultural missions, including animal health and well-being, food safety, and human nutrition.
- Ensure that genomics data are freely available in the public domain and genomics reagents and resources are available to the public.
- Increase the training opportunities for genomics and bioinformatics at all levels of education.
- Coordinate and encourage international cooperation to achieve these goals.

The IWG determined that large-scale sequencing, data management and bioinformatics, and functional genomics were the *specific goals to be achieved in fiscal years 2003 to 2007*. The IWG called for, among other things:

- Large-scale sequencing to produce draft genome sequences (8-fold sequence coverage) of honeybee, chicken, dog, cattle, swine, and cat species.
- Data management and bioinformatics to specifically support agriculturally important species, including significant improvements in data management and analysis software, allow for greater data accessibility and secure long term maintenance, increase capabilities to deal with rapidly accumulating data complexity as databases include functional information, and provide more powerful tools to mine large genomes.
- Recognition that an increase in data for livestock genomes requires a concomitant investment in functional genomics to support genome annotation, the study of gene regulation and expression, and species evolutionary relationships.

Since 2002, considerable progress has been achieved towards the goal of placing whole genome sequence and associated tools into the public domain for high priority domestic animal species. Annotated draft sequences have been published for the honeybee, chicken, and dog genomes and the bovine genome sequencing project has reached >6-fold coverage and is entering the final gene prediction and annotation phase. Lighter coverage sequencing of the cat genome has been completed and BAC-skim sequencing of the swine genome was launched in January 2006. Developed concomitantly with these genome projects has been a suite of associated tools including EST libraries, BAC maps, integrated physical and linkage maps, full-length cDNA libraries, microarrays or gene chips, and identification and validation of a large number of single-nucleotide polymorphism markers. All of these efforts have required leveraging of efforts between agriculture and the biomedical sciences, as well as unprecedented partnerships between U.S. Federal research agencies, international groups, universities, and private industry.

In early 2004 as the sequencing goals of the IWG appeared to be within reach, further study of how to best address the remaining two areas of greatest importance – bioinformatics and functional genomics – was warranted. Specifically, the charge was given by the IWG to the USDA to evaluate how programs in these areas should be developed further to allow full utilization of annotated genome sequences and associated tools.

USDA ANIMAL GENOMICS WORKSHOP OVERVIEW

Life sciences research activities in the USDA are administered by two separate agencies. The Cooperative State Research, Education, and Extension Service (CSREES) funds extramural research efforts conducted primarily at land grant and 1890s universities. The Agricultural Research Service (ARS) is the intramural research arm of the USDA and funds long-term, high-risk research on an ongoing basis in its 108 labs throughout the U.S. In fiscal year 2004, USDA funding for animal genomics research totaled $46.4M (ARS - $22.7M and CSREES - $23.7M).

The USDA Animal Genomics Workshop, as called for by the IWG, was designed to facilitate open input and discussion from leading scientists working in the field of animal genomics for USDA administrators. Participating scientists were selected to reflect a balance of funding sources (16 from CSREES and 18 from ARS), species of primary interest (balance between poultry, swine, cattle, sheep, and aquaculture), and area of research emphasis or expertise (gene mapping, bioinformatics and statistical genetics, functional genomics). In addition to a number of program administrators from ARS and CSREES, colleagues from NIH, NSF, and DHS also participated in the workshop.

The workshop was organized in three modules. Each module consisted of presentations by an invited panel of speakers followed by three simultaneous breakout groups to discuss the long-term needs in that area. Group reports were then assimilated into a consensus set of recommendations emanating from the event.

Structural Genomics Priorities in Domestic Animal Genomics. The opening session focused on the structural genomics needs facing animal genomics researchers today. Noelle Cockett, Utah State University, provided an overview to set the stage for further discussion. Generally, scientists have approached genomics primarily by building structural genomics resources, with ventures into functional genomics observed only in more recent years. The animal genomics research community has been successful in prioritizing needs in annual or semi-annual meeting venues, such as the International Plant and Animal Genome Workshops and the International Society of Animal Genetics. Through such international collaborations and efforts, linkage and comparative maps for all livestock species were made available. The recent and ongoing development of whole genome sequence maps of the chicken, honeybee, dog, and cattle species is a major step forward. Other important agricultural species including swine, several aquaculture species, and sheep are attempting to develop funding resources to enter the sequencing pipeline. Single-nucleotide polymorphism (SNP) based-maps now being developed from the chicken and cattle whole genome sequencing projects will be of enormous value in evaluating genetic diversity, fine mapping of QTL, and development of DNA-based animal identification systems. While the current trend toward internalization of genomics research in private companies indicates the potential value of genomic tools, it also was pointed out as a major concern. There was a consensus that we must complete the basic genome infrastructure for all major species and deposit such information in public databases, as this was viewed as absolutely essential to facilitate rapid discovery and the development of commercially usable technologies for agricultural and biomedical sciences and industries

A major advantage of using agricultural animal species in genomics research is the widespread availability of large, pedigreed research animal populations. Many of these populations have been in existence for fifty years or more and have been phenotyped widely for a variety of economically and biologically important characteristics. In the past two decades, a number of sub-populations were set up as resource families for use in QTL detection and subsequently for validation of putative QTL. Participants agreed that it is imperative in the post-genome sequencing era that the value of these populations, and tissue repositories derived from them, be recognized and supported.

Participants agreed that animal genomics is poised to impact several avenues of animal production, life sciences, and biomedical research, but physical and financial resources are crucial to capitalizing on past investments. The utilization of resources and human capital must, however, be carefully directed toward achieving outcomes and deliverables that are measurable in application, promote rapid commercialization, and enhance education of the public and the next generation of scientists. The need for a cohesive, comprehensive long-term plan for all of USDA's research efforts in genomics was evident at the workshop. Further integration of the efforts of CSREES and ARS appears warranted to achieve the greatest return on investment.

Specific Recommendations from the Structural Genomics Module:

1) *Sequence the swine genome to a minimum of 6-fold coverage for deposit into the public domain.*

2) *Obtain BAC maps and 2-fold sequence coverage for the catfish, goat, horse, salmon, trout, and turkey genomes.*
3) *Develop comprehensive full-length cDNA libraries to allow functional annotation to be achieved to acceptable levels for each of the genome assemblies listed under #2.*
4) *Complete integration of genetic linkage, radiation hybrid, and physical maps should be achieved for each genome listed under #2.*
5) *Discover and validate SNP markers and develop haplotype maps for all species to increase the density of maps for fine mapping of QTL and eventual "whole genome selection".*
6) *Develop standardized population and phenotype resources for each of the species.*
    a. *Preserve long-term, unique, experimental animal populations to capitalize on their value in functional genomics research and further develop and maintain diverse animal resource families.*
    b. *Couple these animal populations with genotypic and phenotypic information and obtain funding support for appropriate long-term tissue repositories for tissue cultures, DNA and RNA.*
    c. *Explore options to bring the agricultural animal genomics community in line with the laboratory mouse and rat communities (i.e. the Jackson Labs model). [The National Animal Germ Plasm Program, currently administered by USDA, may provide a foundation upon which to build for this function. The IWG should study this carefully to avoid any unnecessary duplication of effort and resources across Federal agencies.]*

Long-Term Challenges in Making Use of Genome Sequence Information through Functional Genomics. The second module of the workshop was an open discussion of the challenges facing agricultural animal genomics researchers in capitalizing on the structural genomics infrastructure through downstream applications in functional genomics, proteomics, metabolomics, and metagenomics. Three working groups were assigned in advance of the workshop to develop presentations representing the genomics research communities working with poultry, swine, and ruminants.

Jerry Dodgson, Michigan State University, presented an overview of chicken genome research and associated challenges. With the availability of the chicken physical maps, ESTs, microarrays, and, most importantly, the release of the draft genome sequence in 2004, the chicken became the first avian genome sequence available to scientists worldwide. It was stated that the domestic chicken has retained enormous genetic diversity, based on comparative SNP-based studies of three chicken breeds (broiler, layer and Chinese silkie) relative to the Red Jungle Fowl used for the genome sequence. Chickens possess an abundance of quantitative variation in production and disease resistance traits and are a unique biomedical model in addition to being a leading source of high quality animal protein worldwide. For these reasons, a case was made for finalizing the draft genome sequence of chicken. The chicken genome community will face grand challenges similar to those faced by the human genome community in the post-genome sequencing era, including: 1) identifying the structural and functional components encoded in the genome; 2) elucidating the genetic networks and protein pathways and their relation to phenotypes; and 3) understanding and applying the heritable variation in the genome.

Harris Lewin, University of Illinois at Urbana-Champaign, presented the long-term challenges associated with making use of genome sequence information through functional genomics in ruminants. An improved understanding of the genomic basis for traits of economic importance to the dairy, beef and sheep industries was identified as an important goal. Research problem areas where functional genomics would contribute include, but are not limited to, embryonic development (pre- and post- implantation), lactation (efficiency, composition and product quality), wool growth, muscle growth and meat quality, feed efficiency, immunobiology of infectious diseases, and animal well-being. Ultimately, the selection of candidate genes and the identification of allelic variation associated with the phenotypes are important products of functional genomics. For cattle, genomic resources available to accomplish the functional genomics goals include linkage and radiation hybrid maps with a large number of markers, BAC libraries, and physical maps developed through the International Bovine BAC Map Consortium, and whole genome sequence information. Germplasm repositories, animal resources, QTL maps, ESTs, microarrays and associated databases are in place and available as additional resources to accomplish the functional genomics goals. The "grand challenges" for the ruminant genomics community are: 1) functional annotation of cattle (and other ruminant) genes; 2) complete description and understanding of cellular pathways (e.g., metabolism, proliferation, differentiation, cell-cell interaction); 3) genomic-environment interaction (e.g., developmental pathways, abiotic stresses such as heat, cold, and drought, nutritional genomics, and infectious diseases); and 4) the development of an encyclopedia of economic trait loci. A need for additional biological resources (e.g., tissue banks, animal germplasm, cell lines), genomic technologies (e.g., RNAi, genotyping services, cloning and transgenics) and integrative databases and informatics was identified.

Larry Schook, University of Illinois at Urbana-Champaign, presented on behalf of the swine genomics research community. It was emphasized that to answer key biological questions, it is essential to have a whole genome sequence to harness comparative functional genomics across species. A minimum 6-fold coverage of the swine genome was recommended for carrying out a high quality, functional genomics program and to remain compatible with the NIH genomics programs. He outlined the International Swine Genome Sequencing Consortium's efforts in identifying the needs and resources for the swine genome sequencing initiative. A timetable was presented and it was emphasized that the international researchers and industry leaders are in agreement that a swine genome sequence is needed and the effort is timely.

General Discussion. Downstream or post-genomic applications, such as functional genomics, proteomics and metabolomics, clearly are the areas where agricultural species will benefit from the genome sequencing research investments. These benefits have begun to be realized with the completion of the human genome sequence; for example, over 40 genes have been identified subsequently for a variety of conditions, including macular degeneration (Yamagishi et al., 2005), cleft palate (Frebourg et al., 2005), lymphoproliferative disease (Nichols et al., 2005), mental retardation (Jensen et al., 2005), and testicular cancer (Diederichs et al., 2005). For both human and agricultural species, the post-sequencing challenge will be to understand the operation and function of genomic information. In

particular, the primary issue for agricultural species will be translating the respective genome sequences into enhanced productivity of the phenotypes they control or influence (e.g., disease resistance, behavior, growth, product quality, reproduction).

The post-sequencing era will move rapidly from crudely defined genomic relationships with phenotypes, such as QTL, to a rapid dissection of those relationships in the context of true functional genomics. Some examples of QTL that should progress rapidly from chromosomal localization to industrial application include meat quality and product yield in beef cattle, milk production and mastitis resistance in dairy cattle, litter size and uterine capacity in swine, product yield and parasite resistance in sheep, and coccidia resistance in poultry. The availability of genome sequences for agricultural species will enhance significantly fine mapping of individual genes in two key ways. First, an exponential increase in the numbers of SNPs distributed throughout the linkage maps will enable fine mapping of QTL at a level previously not possible. For example, poultry genomics is poised to realize this benefit with the placement of some 3 million SNPs on a 1.2 Gb genome. Second, comparative genomics will increase the likelihood of QTL identification by virtue of the highly conserved regions of genes throughout mammalian species (e.g., myostatin gene responsible for double-muscling condition in cattle [Grobet et al., 1997; Casas et al., 1998; Yang et al., 2001]).

The majority of economically important traits exhibit complex or multifactorial inheritance patterns that are influenced by environmental factors; therefore, the principal challenge is not simply detecting the QTL, but rather unraveling the genes and the regulatory elements that control gene expression (Andersson and Georges, 2004). This will require the integration of numerous resources, including genetic and physical maps, QTL markers, EST libraries, microarrays and the whole genome sequence to delineate the molecular mechanisms that control complex biological systems. Agricultural species have an advantage in that phenotypes are well characterized and diverse because they have been closely monitored and specifically modified through selected breeding.

Expression profiling of large numbers of genes across diverse tissues, populations, and environmental states also will use increasingly sophisticated quantification platforms. For example, the expression of literally thousands of genes can be studied simultaneously already using DNA chips or microarrays. The molecular biologist will be able to bypass traditional laborious processes, such as screening BAC libraries, and instead clone genes “in silico” (Wong, 2004). Proteomic technologies, including new developments in mass spectrometry and database searching, are leading to rapid advances in monitoring genome activity at the protein level. We can expect further advances in understanding the structural biology of proteins when comparative and evolutionary approaches to sequencing are utilized. Proteome analysis will elucidate groupings of genes that regulate metabolic pathways. Additionally, by following gene expression fluctuations over time and in response to specific signals, the position occupied by the protein end product of a particular gene, relative to others in metabolic and signaling pathways, can be inferred (Roberts, 2001). It follows, then, that fields, such as metabolomics, will allow genomic characterization of “systems” of proteins and their applications to animal health and nutrition, as well as human nutrition and obesity. Whereas genes and proteins set the stage for what happens in the cell, much of the

actual activity is at the metabolite level: cell signaling, energy transfer, and cell-to-cell communication are all regulated by metabolites (Schmidt, 2004).

New technologies will continue to be developed at a rapid pace to improve both the precision and efficiency of the various 'omics' approaches. For instance, the phenomenon of RNA interference (RNAi) has evolved rapidly into a powerful technique to silence gene expression in eukaryotic cells. Agricultural researchers have begun to use this technology to study gene function in porcine granulosa cells (Hirano et al, 2004) and bovine ooctyes (Paradis et al., 2005). Because RNAi technology can be used to knock out genes across a genome, having the complete genome sequence will greatly improve identification of 'targets' (proteins) for existing drugs. For example, parasitologists at CSIRO Livestock Industries are using this approach in an effort to control internal and external parasites of cattle and sheep. Another emerging technology, metagenomics, is poised to develop rapidly and have profound impacts on functional genomics research in agricultural species. Metagenomics is a new field combining molecular biology and genetics in an attempt to identify and characterize the genetic material from environmental samples and apply that knowledge. Genetic diversity is assessed by isolation of DNA followed by direct cloning of functional genes from the environmental sample. The metagenomics field was pioneered when researchers used whole genome shotgun sequencing to sequence microbial populations en masse from the Sargasso Sea (Venter et al., 2004). It is not hard to envision application of this technology to ascertain the microbial populations of the bovine rumen or porcine intestine, for example, and how the dynamics/interactions among bacterial and protozoan species create a unique microenvironment that promotes growth.

Although the field of transgenic animal production is not new in comparison to RNAi and metagenomic tools, this is an example of existing technology where needed improvements will accelerate and culminate in the development of model animal systems using livestock species. Larger domestic animals are valid biomedical research models by virtue of their anatomical and physiological similarity to humans. For example, the retina of the rhodopsin transgenic pig (Petters et al., 1997) shares many cytological features with human retinas exhibiting retinitis pigmentosa, a degenerative loss of cone photoreceptors that gradually leads to blindness. Most recently, this transgenic animal model has been used to develop surgical transplantation of normal neuroretinal grafts (Ghosh et al., 2004). The widespread use of existing transgenic domestic animal models has been limited by the relatively low success rates of nuclear transfer and cloning. To illustrate the low efficiency of producing transgenic animals, consider that a minimum of 1,200 microinjected eggs were required to produce one transgenic sheep, goat, or cow and that only about 50% of offspring express the transgene (Wall et al., 1997). A prime example of the enormous potential of transgenic technology is the recent production of transgenic dairy cattle with resistance to mastitis (Wall et al., 2005). The production of 8 transgenic calves, however, required embryonic transfer of 927 good quality blastocysts that were created from over 4,000 nuclear donor cells. Using similar technology, dairy cattle also present a great potential for producing amounts of therapeutic proteins secreted into milk. Likewise, eggs from transgenic hens are a potential high-throughput mechanism for production of therapeutic proteins. Additionally, the avian transgenic system may confer post-translational glycosylation processing more similar to humans than other species currently used for transgenic production of proteins. It is clear

that gene transfer technologies will have renewed focus in the post-sequencing era of genomics.

Perhaps the most intriguing example of new technology development on the heels of genome sequencing was the call for proposals from the National Human Genome Research Institute (NHGRI) in 2004 seeking the next generation of technologies that would reduce the cost and increase the throughput of DNA sequencing. In short, the goal of NHGRI is to lower the cost of sequencing one individual's genome (human or animal) to $1000 (USD). Once in place, these technologies will further revolutionize the post-sequencing era for agricultural species.

With all of the expected and rapid increases in knowledge in the near future, it is imperative that the methodology for defining phenotypes be clear and standardized. The systematic classification and characterization of phenotypes is essential for ultimately mapping the genes responsible for normal and abnormal development and physiology. More importantly, any search for mutations or altered functional expression depends on phenotypic screening and the ability to detect variation from normal. The challenge, then, is to develop efficient, systematic, and comprehensive phenotypic screening procedures and tools that will permit comparison among laboratories. For example, the current phenotypes of highly pathogenic avian influenza (HPAI) were formulated over 10 years ago when the only virus known to have mutated to virulence was the HPAI responsible for the 1983–84 Pennsylvania epizootic (Alexander, 2002). Cumulative evidence, however, suggests that HPAI viruses actually arose from low-pathogenicity avian influenza (LPAI) H5 or H7 viruses infecting poultry after spreading from free-living birds. At present, it can only be assumed that all H5 and H7 viruses have this potential and mutation to virulence is a random event. Therefore, the longer the presence and greater the spread in poultry, the more likely it is that HPAI virus will emerge (Alexander, 2002). This example illustrates how major research efforts in phenotypic screening are needed to characterize traits that have been difficult to measure until now.

Concomitant with the advent of functional genomics, the types and amounts of data that need to be stored in databases have changed dramatically. Many types of information that were previously collected on an *ad hoc* basis now need to be stored in a more structured manner. Additional data sets for gene expression, proteomics, and protein-protein interactions are growing increasingly complex. To analyze the data computationally in an efficient manner, there is a need for consistency between expressions in different phenotypic domains as well as in different species. The term "phenotype" can be used in different ways in different fields in biology and by different researchers in those fields. It may mean anything from the complete set of phenotypic attributes (traits) that describe an individual to a single phenotypic attribute that distinguishes an individual from other, "normal" individuals (Gkoutos et al., 2004). The development of phenotypic ontologies for livestock is critical to our ability to connect heterogeneous data types back to animal. It would be best to define the ontology in a proactive manner so that future applications will not be confounded by unraveling duplicative and/or mismatched phenotypic designations.

Equally important is to approach functional genomics, proteomics and metabolomics from an integrative systems biology perspective. Within a systems biology approach, each type of

biological information (DNA, RNA, protein, protein interactions, biomodules, cells, tissues, etc.) also has individual elements (e.g., specific promoters, genes or proteins), and the interrelationships of all these elements and types of biological information must be determined and integrated to obtain a view of the system as a whole. What is ultimately desired is the ability to unravel the complexities of epistatic (genotype by genotype; GXG) and genotype by environment (GXE) interactions and how they affect phenotypic expression. A typical GxE interaction of concern for agricultural production would be the change in performance of a set of genotypes among differing environments. Deciphering these complexities requires a holistic approach that describes and understands the biology underlying phenotypes.

The post-genome sequencing era will bring enormous quantitative and phenotypic data to the table. The USDA is the logical organization to lead this systems biology approach for agricultural species. It is suggested that compartmentalization of genomics programs, as has been done in the past for both CSREES and ARS program management, should be shifted toward integration of functional genomics approaches into all program areas and disciplines (e.g., animal growth and production, animal health, animal well-being, aquaculture, food safety, animal waste management, animal and human nutrition, etc.). A cross-disciplinary research effort will be required to integrate the global genomics data into information that is usable and applicable across the diverse landscape of agricultural production.

Specific Conclusions and Recommendations from the Functional Genomics Module:

1) *Downstream work in functional genomics and proteomics will be where the big payoff from animal genomics research is reaped.*
2) *Develop a clear and standardized methodology for defining phenotypes for success, particularly in the emerging areas of animal health and well-being.*
3) *Utilize a "big science", "holistic" approach to unravel the complexity of epistatic and genotype by environment interactions. [Agricultural animal genomics research is ideally suited to an integrative systems biology approach]*
4) *Significantly enhance the bioinformatics capacity within the public agricultural animal research enterprise to handle the increasing complexity and volume of genomic and proteomic data.*
5) *Substantially increase the comprehensive funding for downstream functional genomics, proteomics, metabolomics, and metagenomics research in agricultural animal species to capitalize on the previous investments in genomic resources, tools, and reagents.*
6) *End the previous separation of genomics efforts within USDA research portfolios as the integration of functional genomics approaches as a foundation in all program areas and disciplines is warranted.*
7) *To integrate genomic approaches across disciplines, improve the coordination and effectiveness between ARS and CSREES by developing and implementing a long-term strategic plan for USDA animal genomics research.*

Focusing on Bioinformatics Resources. The third and final module of the workshop focused on bioinformatics needs. Elliott Margulies of the NIH's Intramural Sequencing

Center discussed the vision of the post-sequencing era after sequencing the human genome (Collins et al, 2003). Because the human genome is extraordinarily complex and its function is poorly understood, the grand challenge for NIH is to catalogue, characterize, and comprehend the entire set of functional elements encoded in the human genome. Embedded within the complexity of the genome is the fact that only 1 to 2% of the DNA sequence actually encodes proteins, and the full complement of protein-coding sequences still remains to be established. Consequently, a major role for comparative sequence analysis will be the identification of functionally important non-coding sequences. These sequences will be hard to identify, as virtually no complementary datasets are available across various species to assist with computational predictions. Nevertheless, methodologies for multi-species, comparative sequence analysis relative to the human genome exist and can be used to gain insight regarding species divergence, as well as substitution rates within coding or non-coding regions under natural selection pressures over time.

John Keele, USDA/ARS U.S. Meat Animal Research Center, provided an overview of the current bioinformatics infrastructure for agricultural animal genomics. The most useful databases and bioinformatics resources included those at The Institute for Genomics Research (TIGR), National Center for Biotechnology Information (NCBI), MS-Access or MySQL software for local sequence and genotyping databases, and DNAStar (soon to be replaced by SeqWeb). In addition, the Generic Model Organism Database (GMOD), funded by the NIH and the ARS, was mentioned as a unique tool for genome database visualization, curation, and ontology. Adequate databases and tools are available to manage and analyze ESTs, SNPs, microarray, SAGE and proteomics information; however, there remain unique personnel, skills, and software needs for each of these tools. It was noted that the general lack of bioinformatics personnel and minimal integration of relational databases with all aspects of research are the two critical factors that are limiting progress in the field of bioinformatics.

Specific Recommendations from the Bioinformatics Module:

1) *Focus USDA resources on its unique capabilities, such as phenotypic characterization, population and quantitative genetics, physiology, etc., and be careful to not "re-invent" the bioinformatics capabilities already in place in other genomics research communities.*
2) *Immediately provide training programs and associated support for faculty sabbaticals, postdoctoral associates, and graduate students focused on integrating biology and computing, since one rate-limiting step for USDA in bioinformatics is awareness and literacy in use of existing tools and lack of basic training programs to bring new bioinformatics personnel online.*
3) *Develop standard descriptions of phenotypes as this is a second rate-limiting step for USDA in bioinformatics and it is a problem that will be exacerbated when functional genomics research moves into the more challenging areas of animal health and well-being in the near future.*
4) *Create a USDA bioinformatics working group at the Research, Education, and Economics mission area level to: a) coordinate and define ARS and CSREES efforts among data producers, tool developers, analysts, and consumers; and b) to better coordinate with other Federal and international agencies.*

5) *To best serve the bioinformatics needs of agricultural animal genomics, leverage USDA resources with others to develop expertise and new tools.*
    a. *Support species-specific annotation;*
    b. *Organize curation groups for management of livestock genome sequence resources, in concert with existing groups (i.e,. NCBI, UCSC, Ensembl), to help build browsers with characteristics important to current and future animal genomics research.*
    c. *Link genomic data to published literature in the animal sciences.*
    d. *As large numbers of SNP are discovered and validated, develop databases linking haplotypes with phenotypes and further tools (e.g., NCBI, dbSNP) to facilitate QTL mapping and association studies for multiple species.*
    e. *Develop a centralized and standardized system for microarray analysis and gene expression databases by requiring agreement on database platform (s) for microarray target annotation and gene expression data mining with intentions to link to genome assemblies and associated gene and protein databases.*

CONCLUSIONS

There is little doubt that the investments made to date in animal genomics will yield enormous dividends in the future for the producers and consumers of animal products and for the biomedical sciences. However, this workshop clearly identified a number of areas that need significant programmatic and funding attention within the USDA research infrastructure for this potential to be realized in a timely manner. Opportunities appear to exist and should be explored further for leveraging of future efforts with other Federal programs, given the wealth of genotypic and phenotypic information catalogued on pedigreed agricultural animal populations. Furthermore, there was strong consensus that in the post-sequencing era, research employing genomics techniques and tools should be integrated across all disciplines engaged in the animal sciences as opposed to being separated into "genomics" program areas. An overwhelmingly clear message from the workshop was that it is critical for USDA research leaders to develop and implement a visionary, long-term plan for animal genomics research as soon as possible. Such a plan will ensure that the full potential of past, current, and future efforts and investments in animal genomics will have a positive impact on animal producers and the public in the post-sequencing era.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the scientific contributions of D. Adelson, L. Alexander, S. Burgess, A. Capuco, H. Cheng, N. Cockett, L. Cogburn, E. Connor, P. Coussens, J. Dodgson, C. Elsik, C. Ernst, K. Eversole, L. Gasbarre, J. Keele, H. Lewin, H. Lillehoj, E. Marguiles, J. Ostell, J. Reecy, C. Rexroad, III, R.M. Roberts, G. Rohrer, M. Rothschild, D. Schneider, L. Schook, T. Smith, T. Sonstegard, J. Vallet, C. Van Tassell, G. Waldbeiser, W. Warren, J, Womack, K. Worley, and K. Zuelke. Additional appreciation is extended to P. Brayton, P. Burfening, L. Ellis, D. Hamernik, J. Jen, P. Johnsen, S. Kappes, M. Mirando, J.

Peterson, and W. Zamer for their participation in the planning and implementation of the workshop. Finally, the efforts of the national program staff of CSREES, especially V. Martin, and ARS, especially L. Mangra, for handling the logistics of the workshop are sincerely appreciated.

REFERENCES

Alexander D.J. (2002) Should we change the definition of avian influenza for eradication purposes? *Avian Diseases* 47:976–81.

Andersson, L. et al. International Chicken Polymorphism Consortium (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432:717-22.

Andersson L. and Georges M. (2004) Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet.* 5(3):202-12.

Casas, E., Keele, J.W., Shackelford, S.D., Koohmaraie, M., Sonstegard,,T.S., Smith, T.P., Kappes, S.M., and Stone R.T. (1998) Association of the muscle hypertrophy locus with carcass traits in beef cattle. *Journal of Animal Science* 76(2):468-73.

Collins, F.S., Green, E.D., Guttmacher, A.E., and Guyer, M.S.. (2003) *Nature* 422:835-47.

Gibbs, R.A. et al. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299-1320.

Gibbs, R.A., Weinstock, G., Kappes, S.M., Schook, L.B., Skow, L., and Womack, J. (2002) Bovine genomic sequencing initiative: De-humanizing the cattle genome. http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/BovSeq.pdf.

Diederichs, S., Baumer, N., Schultz, N., Hamra, F.K., Schrader, M.G., Sandstede, M.L., Berdel, W.E., Serve, H., and Muller-Tidow, C. (2005) Expression patterns of mitotic and meiotic cell cycle regulators in testicular cancer and development. *International Journal of Cancer* Mar 30; [Epub ahead of print].

Frebourg, T., Oliveira, C., Hochain, P., Karam, R., Manouvrier, S., Graziadio, C., Vekemans, M., Hartmann, A., Baert-Desurmont, S., Alexandre, C., Lejeune Dumoulin, S., Marroni, C., Martin, C., Castedo, S., Lovett, M., Winston, J., Machado, J.C., Attie, T., Jabs, E.W., Cai, J., Pellerin, P., Triboulet, J.P., Scotte, M., Le Pessot, F., Hedouin, A., Carneiro, F., Blayau, M., and Seruca, R. (2005) Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. *Journal of Medical Genetics* 14; [Epub ahead of print].

Ghosh, F., Wong, F., Johansson, K., Bruun, A., and Petters, R.M. (2004) Transplantation of full-thickness retina in the rhodopsin transgenic pig. *Retina* 24(1):98-109.

Gkoutos, G.V., Green, E.C., Mallon, A.M., Hancock, J.M., and Davidson, D. (2004) Building mouse phenotype ontologies. *Pac Symp Biocomput.* 9:178-89

Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., Fries, R., Hanset, R., and Georges, M.A. (1997) Deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics* 17:71-4.

Hamernik, D.L., Lewin, H.A., and Schook, L.B. (2003) Allerton III. Beyond livestock genomics. *Animal Biotechnology* 14:77-82.

Hirano, T., Yamauchi, N., Sato, F., Soh, T., and Hattori, M.A. (2004) Evaluation of RNA interference in developing porcine granulosa cells using fluorescence reporter genes. *J Reprod Dev.* 50(5):599-603.

Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P., Van Esch, H., Kleefstra, T., Hamel, B., Moraine, C., Gecz, J., Turner, G., Reinhardt, R., Kalscheuer, V.M., Ropers, H.H., and Lenzner, S. (2005) Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *American Journal of Human Genetics* 76(2):227-36.

National Academy of Sciences. (2002) *Exploring Horizons for Domestic Animal Genomics: Workshop Summary*. (Ed. Pool R and K Waddell). National Academy Press, Washington DC (42 pp).

Nichols, K.E., Hom, J., Gong, S.Y., Ganguly, A., Ma, C.S., Cannons, J.L., Tangye, S.G., Schwartzberg, P.L., Koretzky, G.A., and Stein, P.L. (2005) Regulation of NKT cell development by SAP, the protein defective in XLP. *Nat Med.* 11(3):340-345.

Paradis, F., Vigneault, C., Robert, C., and Sirard, M,A. (2005) RNA interference as a tool to study gene function in bovine oocytes. *Mol Reprod Dev.* 70(2):111-21.

Petters, R.M., Alexander, C.A., Wells, K.D., Collins, E.B., Sommer, J.R., Blanton, M.R., Rojas, G., Hao, Y., Flowers, W.L., Banin, E., Cideciyan, A.V., Jacobson, S.G., and Wong, F. (1997) Genetically engineered large animal model for studying cone photoreceptor survival and degeneration in retinitis pigmentosa. *Nature Biotechnology* 15(10):965-70.

Roberts, R.M. (2001) The place of farm animal species in the new genomics world of reproductive biology. *Biology of Reproduction* 64:409-17.

Schmidt, C.W. (2004) Metabolomics: what's happening downstream of DNA. *Environmental Health Perspectives* 112(7):A410-5.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., and Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74.

Wilson, R. et al. Inernational Chicken Genome Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.

Wall, R.J., Kerr, D.E., and Bondioli, K.R. (1997) Transgenic dairy cattle: genetic engineering on a large scale. *Journal of Dairy Science* 80(9):2213-24.

Wall, R.J., Powell, A.M., Paape, M.J., Kerr, D.E., Bannerman, D.D., Pursel, V.G., Wells, K.D., Talbot, N., and Hawk, H.W. (2005) Genetically enhanced cows resist intramammary Staphylococcus aureus infection. *Nature Biotechnology* 23(4):445-51.

Wong, E. (2004) Poultry Research in the Post-Genome Era. ISB News Report.

Yamagishi, S., Nakamura, K., Inoue, H., and Takeuchi, M. (2005) Met72Thr polymorphism of pigment epithelium-derived factor gene and susceptibility to age-related macular degeneration. *Med Hypotheses* 64(6):1202-4.

Yang, J., Ratovitski, T., Brady, J.P., Solomon, M.B., Wells, K.D., and Wall, R.J. (2001) Expression of myostatin pro domain results in muscular transgenic mice. *Mol Reprod Dev.* 60(3):351-61.