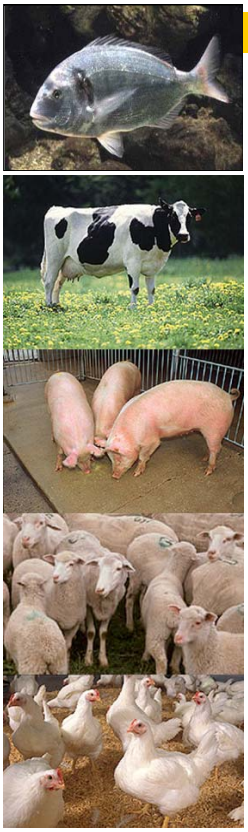
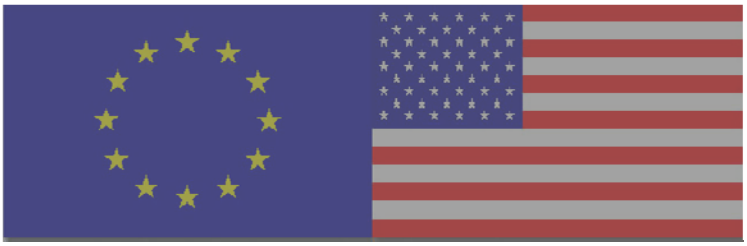
EU-US Animal Biotechnology Working Group



Strategic Priorities to Realize the Promise of Genomics to Animal Health, Well-being & Production

Report on a workshop held in Washington DC
8-10 November 2011



Under the auspices of the EU-US Task Force on Biotechnology Research

Writing Teams:
Reference sequences - Chris Elsik, Martien Groenen, Tad Sonstegard, Dave Adelson

Cyberinfrastructure – Jim Reecy, Alessandra Stella, Fiona McCarthy, Dave Adelson, Shane Burgess, Kent Reed, Chris Elsik, Elena Sarropoulou, Hossein Jorjani

Genotypic, phenotypic, epigenetic variation – Gary Rohrer, Elisabetta Giuffra, Michel Georges, Joan Lunney, Beth Cleveland, Elena Sarropoulou, Vivek Kapur, Shane Burgess, Kent Reed

Genotype to phenotype – Sue Lamont, Ross Houston, Greg Wiens, Joan Lunney, Shane Burgess, Miguel Perez-Enciso, Gary Rohrer, Elisabetta Giuffra

Communication – Johan Vanhemelrijck, Joan Lunney, Chris Warkup, Elisabetta Giuffra

Training – Chris Elsik, Martien Groenen, Fiona McCarthy, Johan Vanhemelrijck, Lakshmi Matukumalli, Elena Sarrapoulou, Greg Wiens, Miguel Perez-Enciso, Hossein Jorjani

Editors- Jeffrey Silverstein, Anne-Sophie Lequarre, Lakshmi Matukumalli, Gary Rohrer

Preface

Since 1990, the ***EU-US Task Force on Biotechnology Research*** has been coordinating transatlantic efforts to guide and exploit the ongoing revolution in biotechnology and the life sciences. The Task Force was established in June 1990 by the European Commission and the White House Office of Science and Technology Policy. This mandate has been renewed five times. The Task Force has acted as an effective forum for discussion, for coordination and for developing new ideas for the last 22 years. Through sponsoring ***workshops***, and other activities, the Task Force brings together scientific leaders and early career researchers from both sides of the Atlantic to forecast research challenges and opportunities and to promote better links between researchers. Over the years the task force has become a successful think tank on Biotechnology Research with a strong forward looking approach and with concrete examples of successful coordination in the areas of environmental biotechnology, plant and animal biotechnology, bio-based products, marine genomics, etc. This successful trans-Atlantic collaboration is a unique forum that is expected to grow in the future.

The taskforce is composed by 6 Working Groups. Among them the Animal Biotechnology Working Group (ABWG) stresses the urgent need to provide food and resources for the predicted global population of 9 billion by 2050 against a backdrop of environmental challenges, as climate change and water scarcity, in a cost effective and sustainable manner. Towards that aim we need to define critical livestock phenotypes and deploy the latest genomics and biotechnological approaches to control these phenotypes. Recent innovations such as high-throughput sequencing provide new opportunities to understand and exploit genetic mechanisms to increase productivity and control diseases. However, the large datasets generated using these technologies provide novel challenges in data storage, processing, and analysis.

Since 2008 the ABWG members have met regularly and conducted workshops to jointly discuss these issues and devise solutions. In 2009 they held a workshop in Cambridge, United Kingdom to review the existing resources and emphasize the need for strengthening the various aspects of animal biotechnology research such as computing (infrastructures, databases, and algorithms), data generation and analysis (phenomics, quantitative genetics, and annotation), and standards (data quality/availability, reference populations/data sets). The workshop outcome supported international cooperation and enhanced research funds. All participants echoed the need for training the next generation of animal scientists in bioinformatics and computational skills.

In 2010, a round table discussion was held at the 2nd International Symposium on Animal Genomics for Animal Health (AGAH) in Paris, France to foster a dialogue between scientists working at the cutting edge of animal genomics and animal health research. Stakeholders were included to provide input on priorities. It was agreed that the next step would be to organize a workshop to conduct a priorities assessment of the recommendations.

A recent review on USDA-funded programs in Animal Genetics, Genomics and Bioinformatics drew the needs for research that would build on the Blueprint for USDA Efforts in Agricultural Animal Genomics. www.csrees.usda.gov/nea/animals/pdfs/animal_genomics_blueprint.pdf.

This outline was used to build the agenda for the 2011 workshop held in November 8-10, in Beltsville, Maryland. The conclusions of this important meeting for establishing priorities are summarized in this report

The coordinators of the activity were Jeff Silverstein (US USDA), Anne-Sophie Lequarre (EU EC Directorate). The scientific co-chairs were Shane Burgess (University of Arizona, USA), Gary Rohrer (US Meat Animal Research Center, USA) and Martien Groenen (Wageningen University, Netherlands), Elisabetta Giuffra (INRA, France) and Michel Georges (Liège University, Belgium)

US Task Force Co-Chair

Jane Silverthorne

Division Director

Integrative Organismal Systems
National Science Foundation

EU Task Force Co-Chair

Maive Rute

Director Biotechnologies, Agriculture and Food

DG Research and Innovation
European Commission

Executive Summary

The EU-US Animal Biotechnology Working Group (ABWG) held its workshop, Strategic Priorities to Realize the Promise of Genomics to Animal Health, Well-being and Production, in Beltsville, Maryland, in early November 2011 to:

- Review the scientific progress made since the publication of the 'Blueprint for USDA Efforts in Agricultural Animal Genomics', a document developed in concert with discussions among international partners;
- Look forward and develop a vision document describing the current challenges; and
- Make recommendations for future research through EU-U.S. collaborations (and also allowing for partnering with other countries) and by leveraging funding to develop common resources and help solve common problems.

Funding for U.S. participants was primarily through a USDA/National Institute of Food and Agriculture (NIFA) conference award, with some additional funds provided by ARS. The EC Directorate funded the participation of the EU members and two Australians (one speaker and one workshop participant). Representatives from animal agricultural industries were also invited and six participants joined the meeting, three from the United States and three from the European Union side. In total, there were 38 attendees, including several early career scientists.

This was an umbrella meeting intended to generate ideas and direction for the ABWG over the next several years. Speakers were invited from the EU and the United States, as well as from Australia, to speak on Genomics Applications, Cyber-infrastructure and Bioinformatics, and Phenotypes and Phenotyping, primarily from outside the typical animal agriculture community. The first day of the meeting was Webcast, and recorded (recording posted at ARS: (www.ars.usda.gov/sp2UserFiles/Place/00000000/NPS/videos/EU_US_ABWG_Workshop_Webinar.asx), highlighting talks from internationally recognized experts.

On Day 2 of the workshop, discussion centered on the key needs, investments, and milestones for delivering on the promise of genomics to improve animal health, well-being, and productivity, and on priorities for reference sequence improvement. Breakout groups discussed development of genomic and cyber-infrastructure resources; better characterization of phenotypic variation; sources of variation and development of ontology structures; and developing a grand challenge around the prediction of phenotype from genotype. These topics were identified for future focused workshops. Furthermore, training and public communication were identified as key elements for ABWG to highlight. Several writing teams were created to produce a workshop report, the topics include:

- Improve agricultural and aquaculture animal reference genome sequence assemblies (develop and finish);
- Characterize genotypic, phenotypic and epigenetic variation;
- Application of resources to predict phenotype from genotype (G2P);
- Develop vision for common trans-Atlantic cyber-infrastructure;
- Education and Training; and
- Communication/Public Acceptance.

It was discussed that training and public communication elements should be included in every workshop and activity of the ABWG. One specific positive development topic to follow up on from the workshop is

related to cyber-infrastructure. Steve Goff and Dan Stanzione, PI and co-PI of the iPlant Collaborative (http://www.iplantcollaborative.org/), not only gave great informative talks, they also were open to help and discuss the ABWG efforts. Shane Burgess, the U.S. scientific co-chair, followed up quickly with them to develop an iAnimal site for a cyber-infrastructure discovery environment for animals. They are actively developing a plan to enable animal researchers to utilize the power available to plant researchers in iPlant's cyber-infrastructure environment. The iAnimal prototype interface was rolled out by the ABWG Scientific co-chair Shane Burgess at the Plant and Animal Genome Conference (PAG) 2012 during the NRSP-8 (A Bioinformatics Coordination Program, funded by NIFA) session, which was attended by a large fraction of the international animal genetics/genomics community. Over time additional features will naturally be added to this environment but a comprehensive plan to build specific cyber-infrastucture utility for the animal community will require dedicated funding to iPlant. Regardless, co-utilization of this cyber-infrastructure is likely to benefit both groups as paradigms are shared between the two especially in the area of genotype to phenotype characterization.

In September 2012, there will be a follow-up workshop in the United Kingdom in conjunction with a livestock genome annotation meeting to be held at EBI (www.ebi.ac.uk/~laura/livestock_meeting_2012.html). This workshop will be focused on the development of cyber-infrastructure together with potential specific applications and tools for bridging genotypes to phenotypes in livestock. As part of the workshop, a training opportunity to introduce investigators to iAnimal is also being arranged. This group is an ideal pilot study group and their feedback will be important in future iPlant/iAnimal development as well as workshop improvement.

High Quality Reference Genome Assemblies

General context

The list of agricultural and aquacultural animal genomes sequenced is increasing every year. A number of agriculturally and aquaculturally central genomes are currently represented in the Ensembl genome browser, including cow, pig, horse, chicken, turkey, cod, and tilapia with more in the pre-Ensembl genome browser in the process of being annotated. With rapidly declining costs of whole genome sequencing, the number of animal genomes is expected to further increase in the years to come. These genome resources already have had a tremendous impact on the use of genomic data in farm animal research as well is in animal breeding. Nevertheless, all of these genomes are still draft versions, requiring further improvement to reap the benefit from the investments that have generated these resources.

High quality reference genome sequences with comprehensive annotations of functional elements and variants are critical to the translation of genomic discovery to improved animal production. A reference genome sequence not only allows the identification of important functional elements encoded within the sequence itself, but also serves as a framework to connect diverse sets of data related to genes, regulatory elements, polymorphisms, genotypes, and phenotypes. While the initial model for a reference genome assembly was based on a single consensus representation of the sequence, a critical need to represent complex allelic diversity in animal genomes has been recognized[1]. Thus, reference genome assemblies must also serve as frameworks to catalog large-scale differences, such as single nucleotide polymorphisms (SNP) and copy number variation (CNV), between individuals within species.

<u>High quality reference assemblies and annotation are needed to:</u>

- Generate complete and correct annotations of all functional elements.
- Align next generation sequence (NGS) information from individuals to discover single nucleotide polymorphisms (SNPs), which can be used to characterize linkage disequilibrium, map genetic defects or QTL, and identify signatures of selection.
- Correctly order SNPs to enable haplotype-based imputation of genotypes.
- Map NGS transcriptome information to reliably annotated genes within the reference genome to correctly quantify gene expression.
- Map epigenetic information to understand the relationship between epigenetic modification and genes.
- Discover complex genomic variation within species.
- Accurately compare genomes of agricultural animal species with genomes of model organisms and human.
- Integrate complex genomic data types, including diverse genome assembly paths, using the reference sequence coordinate as a filing system.

For many model organisms, and in particular for humans, dedicated facilities and resources such as Genome Reference Consortium (GRC) are in place for continued improvement of the assemblies and

[1] Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H. C., Agarwala, R., McLaren, W. M., Ritchie, G. R., Albracht, D., Kremitzki, M., Rock, S., Kotkiewicz, H., Kremitzki, C., Wollam, A., Trani, L., Fulton, L., Fulton, R., Matthews, L., Whitehead, S., Chow, W., Torrance, J., Dunn, M., Harden, G., Threadgold, G., Wood, J., Collins, J., Heath, P., Griffiths, G., Pelan, S., Grafham, D., Eichler, E. E., Weinstock, G., Mardis, E. R., Wilson, R. K., Howe, K., Flicek, P. and Hubbard, T. Modernizing reference genome assemblies. *PLoS Biol 9(7): e1001091*.

their annotation. It is essential that facilities, such as a core group of annotators/curators, be established for the (major) farm animals as well. This will pave the way for the initiation of a rolling funding mechanism to secure the further annotation and curation of farm animal genomics data for the foreseeable future.

Even the most polished genome assemblies that currently exist for agricultural animal species, such as those for *Bos taurus* and *Gallus gallus*, contain large- and small-scale errors that can lead to incorrect scientific inferences.

Problems caused by incorrect reference genome assemblies include:

- Missing genes or regions, which make it impossible to catalog variation in gene family members or alleles at these regions.
- Misassembly of genes, which can cause incorrect prediction of coding sequences (and protein sequences), promoters, untranslated regions and regulatory elements.
- Misassembly of larger regions, which lead to 1) inaccurate ordering of SNP markers, and thus incorrect haplotype-based imputation of genotypes, and 2) the inability to reliably identify large-scale allelic variation.
- "Humanization" (mainly in the case of mammals) in regions where sequence is of low enough quality such that a better quality genome, such as human, is required to guide the assembly. The same problem with reliance on model organism genomes may occur in non-mammal species.

Currently, there are many on-going large scale re-sequencing projects in farm animals (cattle, pigs, chicken, and others) that are targeted to sequence hundreds or even thousands of individuals from a variety of breeds. These studies are aimed at the analysis of the genetic variation present in these different individuals and the analysis is based on the alignment of short read sequences obtained by next generation sequencing (NGS) methods. Many of these projects are funded by public funding agencies such as the USDA and the European Commission (EC) involving millions of dollars and Euros. Maximizing the success from these projects may be critically dependent on the availability of correct highly improved reference genome sequences and good annotations.

While the specific needs for individual sequencing projects may vary, they can include:

- Greater sequencing depth to reduce the number of gaps.
- Improved ordering of scaffolds, possibly by improved genetic or physical maps.
- Species-specific experimental information to improve the annotation of gene models, non-coding RNA, promoters, CpG islands, and other regulatory elements that regulate gene expression and functional interaction.
- Cataloging genetic variation, such as SNPs, CNVs, and other structural variants.

Information about potential regulatory sequences is currently still limited in farmed animals. Dedicated, coordinated projects (e.g. ChiP-seq techniques) to obtain this information to further improve the annotation are of high importance. The identification and understanding of the many aspects of gene regulation, such as transcription factor binding sites, miRNAs, and miRNA binding sites, are key prerequisites to understanding the systems underpinning the functioning of cells and organisms. Understanding the multiple interactions that influence gene regulation and being able to predict the effect of changes in the complex regulatory networks that control gene expression is a challenging area of research. The availability of well annotated farm animal reference genomes are essential research areas requiring further support.

Cataloguing intra-species variation is of tremendous importance in farm animals and a main research objective in many industry funded research projects. The large scale use of high density SNP chips as well as the analysis of CNV is growing exponentially in farm animals. Farm and companion animals provide a rich resource of different well-defined breeds, with a variety of highly interesting phenotypes. These genetic resources provide a potential goldmine for geneticists further strengthened by the specific population structure and possibilities to generate experimental crosses with greater power to detect functional genetic variation.

To achieve the goal of pursuing high-quality reference genome assemblies, the following points must be addressed:

- Given that sequencing costs are ever decreasing and that many species that were not previously considered as sequencing candidates will be sequenced, should all sequenced species have high-quality reference genomes? If not, how does one decide which ones should be sequenced to high quality?
- What metrics will be used to define a “high-quality reference genome”?
- What threshold for each metric will be required?
- Do all reference genomes need to meet the same standards?
- How will we report and maintain reference genome metadata, such as quality metric values?
- How do we plan for future sequencing technologies and the effects they might have on metrics and standards?

Resources : Genotypic, Phenotypic, Epigenetic variation

General context

The extensive biodiversity of domesticated livestock from populations naturally adapted to different environments (climate, pathogens, seasonal patterns) and/or subjected to various levels of stress in intensive housing conditions creates a huge collection of potentially inherited phenotypes to be measured, unraveled and interpreted on a genomic and genetic basis.

Worldwide views converge in identifying the 'ideal animal of the future' for being 'robust, adapted, and healthy', and 'producing a safe and nutritious food.' By efficiently coping with environmental issues (impact caused by animal husbandry changes, fluctuating climate conditions, escapees, etc.) and societal/industrial needs (e.g. changes in rural lifestyle reducing the workforce, economical issues), animal agriculture will lead efforts to feed the world's population in the future.

The future of livestock and aquaculture genomics relies on the ability to establish animal resources and a common infrastructure, both for the coordination of existing research and the implementation/interconnection of new globalized research and applications.

This implies desires to:

- Capitalize on knowledge and methodological/technological tools developed in other organisms, with a significant input available from the research frameworks and infrastructures developed for animal genomics and for plant species of agricultural relevance.
- Emphasize the scientific value of livestock in comparative genomics within the wider scientific community, as livestock species represent all vertebrate taxa (fish, reptiles, birds, and mammals); they often serve as important 'intermediates' to classical models for human genomics and in providing different types of biological resources for the wider scientific community.
- Foster mission-oriented studies targeting agricultural traits of major relevance for food animal species.

The priority resources are identified in the following paragraphs.

1. Populations and Biological Resources

In animal genomics, commercial populations are an important resource often exploited by joint investigations of academics and industry. However, only a limited number of traits can be effectively explored in commercial conditions due to economic issues, market requirements, and limited possibilities for inter-crossing breeds. It is conceivable that for many innovative phenotypes, experimental populations specifically bred, or traditional breeds raised and preserved in specific areas, will be a key resource to generate basic knowledge to be validated later in the commercial context. The existing connections between industry and academia need to be reinforced by proactive involvement of industry at each step of the decision-making process, to allow exploitation of the best resources on a case by case basis.

The creation and maintenance of animal cell lines from target species, populations, and animals would provide the scientific community easy access to common reference materials and standardized culturing protocols, similarly to what already is realised in the framework of the Encyclopedia of DNA Elements

(ENCODE) consortium for humans. Ideally, these cell lines should be created from the same living prioritized/preserved populations of reference.

There are a large number of animal and fish livestock species each with a different degree of genome sequence annotation, and with experimental animal populations already existing at the national level in Europe and the United States. These resources should be further safeguarded and/or developed in a coordinated way. Existing experimental animal resources in individual countries should first be identified, relevant phenotypic data recorded into central structured databases, and prioritized by ABWG. International governance measures for regulation and access should be designed and proposed by academics in the EU and United States. Such measures would optimize maintenance efforts and breeding costs, strongly sustain funding requests at each national level, and promote scientific exchange internationally. An important component of this effort would be to reinforce EU and U.S. policies for the valorisation of local breeds from emerging countries, as these populations may harbour unique genotypes relative to environmental adaptation and pathogen resistance.

Farm animal genetic research has been revolutionized by the mapping and sequencing of whole genomes – chicken, cattle, swine (on-going), and multiple other species (especially fish and shellfish). This achievement provides a powerful platform for comparative genomic studies in the evolution of species, gene identification across species, as well as the mechanisms of gene function and disease. For many sequenced genomes, it is possible to query a public Web site to view sequence-based comparisons between organisms ranging from human to mouse, cow, chicken, or tetraodon. However, while most genome browsers and comparative genomics tools include conserved synteny and gene content information, there is a paucity of resources that also integrate mapped disease-related or other traits across species.

2. Bioinformatic and Database Resources

Progress in livestock species would be greatly facilitated if genomics data could be easily transferred across species. Toward this goal, multiple species already have ample genomic tool-sets that include not only genomic sequence, but also genetic markers, genetic linkage maps, radiation hybrid (RH) maps, expressed sequence tags (ESTs), partial and full length cDNA, BAC contig maps, and sequence tagged sites (STSs). These resources remain critical to creating a genomic framework for organisms not in queue for whole-genome sequencing or those with low sequence coverage. In addition, much effort has been placed in using the genomic resources across species for studying the evolution of species, for gene identification across species as well as the study of mechanisms of gene function and disease. The results of these great efforts are multiple large data sets that contain a myriad of publicly available information to benefit livestock, model organisms and humans. The challenge is in managing and efficiently mining these datasets to find links to biological function.

Ontology development for biological domains has become prevalent in recent years (see the Open Biomedical Ontologies (OBO) repository at http://obo.sf.net). Ontologies provide a framework for structured data annotation that promotes cross-sharing between bioinformatic resources, structured querying for novel information discovery, and computational analysis of high-throughput data. For example, the Gene Ontology (GO), which describes gene function, is the most widely used biological ontology. Efforts are currently underway to develop ontologies for annotating trait and phenotype data. One approach is to describe phenotype using a combination of ontologies to describe physiological, anatomical, and behavioral traits combined with information about the quality of the trait derived from the Phenotypic Qualities Ontology (PATO). Another approach is to use a combination of trait, clinical

measurement, and measurement method ontologies to capture phenotype information. Ontologies such as these will facilitate the transfer of biological information across species. Despite these efforts, additional ontologies are needed to fully categorize information generated in livestock genetic studies. Capturing phenotypic data requires not only the development of ontologies to capture anatomy and behavior for agricultural animal species, but also development of processes for uniformly annotating data to these ontologies. Moreover, while phenotypic data relies on the interaction between ontologies representing different biological concepts (such as physiology and anatomy), even more complex concepts may relate to behavior and disease which may be described as phenotypes plus values (such as duration, intensity, onset). For example, disease/behavior may be considered as compound phenotypes represented by networks of clinical signs or behaviors. Developing ontologies to represent this data will support discovery based on common phenotypes and their values.
Additional database needs are articulated in the following sections.

3. Farm Animal Genome Variation

A global assessment of genomic variation must be available to researchers for the livestock and aquaculture species of importance to maximize scientific progress. To accomplish this goal, a broad sample of animals representing all major and many minor breeds or strains should be sampled. In addition to commercial and experimental farm animal populations (see 1.), this sampling effort can benefit from *in situ* and *ex situ* existing breed collections. For example, the EU Biodiversity Action Plan for Agriculture has promoted the European Community program on the conservation, characterization, collection, and utilization of genetic resources in agriculture and aquaculture to preserve genetic diversity (http://ec.europa.eu/agriculture/index_en.htm).

Ideally, the genomes of these sampled animals will be sequenced at a moderate coverage. Then, sequences would be mapped back to reference sequences to document the genetic variation represented by each strain and the level of individual variation segregating within the populations assessed. The number of strains and animals per strain included will be dependent upon the species studied. For species that have been domesticated longer, more local populations will exist, but those populations may be more homogeneous (depending on selection imposed), so fewer animals per strain will need to be sampled. For recently domesticated species, fewer strains likely exist, but the genetic diversity within each strain may be much greater, so more animals per strain should be sampled to characterize each strain.

The genotypic and phenotypic information gathered in these experiments needs to be deposited into a publicly available database so that all researchers may access it. This will foster and accelerate the rate at which discoveries can be made and implemented to enable increased production of food for the growing world population.

4. Epigenome and Interactome Projects

Epigenetics is an adaptive process that results from interactions between the genome and the environment. In vertebrate genomes, methylation at position 5 of the cytosine in CpG di-nucleotides is a heritable ‘‘epigenetic’’ mark that has been connected with both transcriptional silencing and imprinting. Analysis of these markers can provide insight into the basis of different phenotypes, unravel aspects of the phenotype (e.g. disease), which cannot be explained by genetic polymorphisms, and explicate modifications in gene expression profiles that arise.

The mechanisms underlying genotype-phenotype relationships remain only partially explained. Combinations of identical genotypes and nearly identical environments do not always give rise to identical phenotypes. The way beyond these challenges is to decipher the properties of basic biological systems. Systems biology approaches must be used to reveal the molecular networks within cells (Interactome) and their response to perturbations. Interactions can occur within and between molecules belonging to different biochemical families, e.g. protein-protein (proteomics) and protein–DNA interactomes. An important challenge is to identify and decipher the complexity of all functional RNA classes, e.g., non-coding RNAs (ncRNAs), and epigenomic markers that are encoded in the genome.

Support for development of diverse bioinformatic approaches will be crucial in this context. Scientists worldwide must exploit international frameworks which have developed following the advanced genome annotation of human and model organisms. The ENCODE Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). Using experimental, computational, and statistical analyses, the aim of ENCODE is to identify all functional elements in the human genome sequence, i.e. genes, ncRNAs, and transcriptional regulatory regions, as well as DNA binding proteins that interact with regulatory regions. ENCODE is revealing the complexities of transcription factors, different versions of histones and other markers, and DNA methylation patterns that define states of the genome in various cell types.

To accomplish its goals, ENCODE has developed standards for each experiment type to ensure high-quality, reproducible data and used novel algorithms to facilitate data analysis and provided a freely accessible database (www.genome.gov/10005107). To aid in the integration and comparison of data produced using different technologies and platforms, the ENCODE Consortium has designated common cell types, including cell lines and primary cell types, that can be used by all investigators. Moreover, plans are being made to explore the use of primary tissues and embryonic stem (ES) cells. The satellite modENCODE project aims to provide the biological research community with a comprehensive encyclopedia of genomic functional elements in the model organisms *C. elegans* and *D. melanogaster*.

An ENCODE project for relevant food animal species must be developed for livestock genomics and downstream applications. It will help to reveal the influence of the epigenetic status on phenotype. Once implemented, epigenetic markers may be used to trace back the environment to which an animal has been exposed (e.g. under or over-nutrition, endocrine disruptors) and to predict future agronomical potential of the animal in a given environment. High-throughput technologies to investigate methylation markers are already extensively used by the ENCODE community using different cell types in model species. High-throughput profiling of samples from livestock and aquaculture species should be conducted. Samples should be recorded from different tissues of each species (from embryo to adult life) and monitored for a broad range of relevant phenotypes with and without challenge by various environmental (biotic and abiotic) stressors. A coordinated EU-U.S. network could obtain and collect information on such samples from a wide genetic background, with a goal of capturing and analyzing the mutual effects of genetic variation and epigenetic genome modifications in different breeds. These data would help build a comprehensive reference database of epigenetic modifications (and phenotypes) per species and per environmental stressor. Parallel collection of cells and tissues would enable future experiments evaluating new or more refined phenotypes in the long term (see below G2P).

An integrated approach adopted for the epigenome will provide materials and data for several parallel interactome projects. ENCODE projects for humans and model species provide input to the study of complex biological systems and cellular networks that may underlie most genotype to phenotype relationships. Similar opportunities will arise following a more advanced knowledge of livestock

genomes coupled to the availability of reference populations. This would set the basis, for example, to develop system biology approaches to study host responses to infection, both *in vitro* and *in vivo*.

5. Microbiome / Virome

Advances in DNA sequencing technologies have created the foundation for analyses of the metagenome, the complex of microbial genomes recovered from animal and environmental samples. The metagenomic approach allows analysis of genetic material derived from complete microbial communities harvested from natural environments. In humans, the analysis of the genomes of the microbiome in selected body sites is assessing whether core sets of microbial communities are shared by all humans and determining how they differ following pathogen and environmental challenges. By using the reference databases, it may be possible to identify microbial markers, predict the functional capabilities of unknown microbial species on the basis of similarities with known genes, and assess the role of the microflora in health and disease.

The vast amounts of metagenomic data produced worldwide by existing consortia has required a coordinated international approach in which common techniques are used to collect samples, extract DNA, record sequences, and annotate data – hence, the recent creation of the International Human Microbiome Consortium (IHMC, www.human-microbiome.org/), launched in Heidelberg, Germany, in October 2008. The goals of the IHMC are to generate a shared resource of human microbiome data and protocols, to coordinate international efforts to reduce redundancy, and provide a venue for exchange of results and strategies. Currently, US$250 million has been committed to the IHMC from around the globe. Full members include the Consortium MetaHIT (www.metahit.eu) for the European Union and the National Institutes of Health (NIH), through the Human Microbiome Project (HMP: http://commonfund.nih.gov/hmp/index.aspx) for the United States.

Several research efforts are underway to build up reference catalogues of microbial metagenomes for major livestock species, with special emphasis to date on the gut and rumen microbial communities. The goal is to achieve understanding of the network of interactions between the microbiota, the environment, and the genotype of the host to understand and be able to modulate the efficiency of digestion and absorption of feed and nutrients, the health/welfare status and resistance to disease, the production of greenhouse gases, the excretion of phosphorus and nitrogen, and the resulting product quality.

Application: Genotype to Phenotype (G2P)

General context

There are major challenges facing modern animal production, including the need to provide high-quality food for a rapidly expanding global population in a sustainable and cost-effective manner. This has to be achieved within the context of intense competition for land resources, often limited water supply, altered weather patterns due to climate change, and, ultimately, carbon-neutrality. The opportunities offered by the application of the enabling genomic technology and computing advances described in the earlier sections of this document are substantial. Researchers can now rapidly create and analyze large datasets to facilitate problem-driven animal science research in a wide range of strategic fields with the ultimate goal of understanding and selecting/modifying individual and population genotypes to enhance their phenotype for a given environment (G2P research) (Fig. 1).

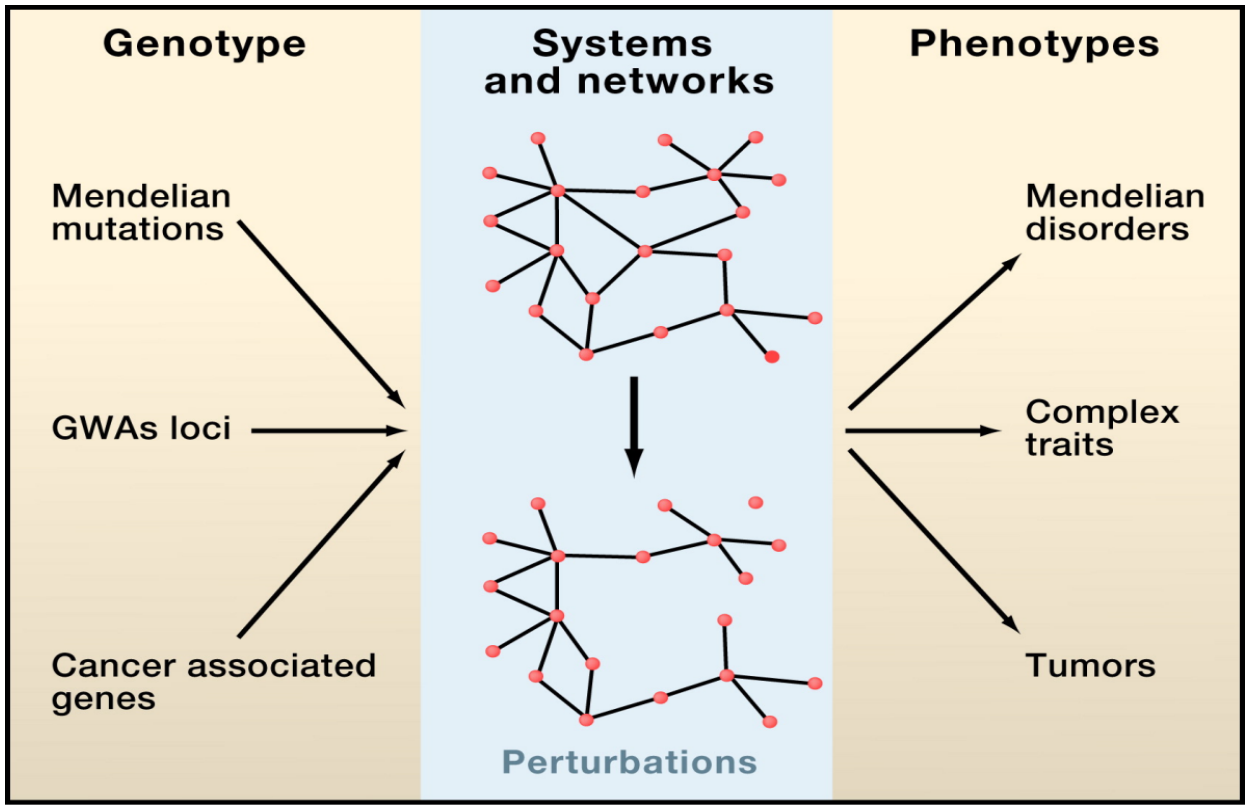


Figure 1. Complexity of systems and network interactions influencing phenotypic traits based on organism genotype. The link between perturbations in network and systems properties and phenotypes, such as Mendelian disorders, complex traits, and cancer, might be as important as that between genotypes and phenotypes. *http://www.sciencedirect.com/science/article/pii/S0092867411001309*

G2P research uses a broad range of fundamental biology (from molecules, proteins, cells, and tissues to whole animals, populations, and environmental interactions) and numerics (statistics, computational tools, and bioinformatics) to improve animal health, welfare, and production efficiency with resulting improvements in environmentally and economically sustainable agricultural production and human nutrition and health.

To maximize the impact of G2P research, improvements are needed in the quantity and rate of academic research translated into the commercial environment, which can be facilitated by:

- Collaboration/partnership between industry and academia;
- Sharing of industrial resources/data where appropriate;
- Knowledge transfer centers and staff embedded within academic institutions;
- Incentivize knowledge transfer for academics in addition to publications and grants;
- Team science; and
- Establishment of well-defined animal resources (ENCODE).

To define key priority target areas wherein G2P research can deliver maximum impact in major global challenges, this section builds on the broader animal genomics research priorities discussed and highlights opportunities in specific G2P priority areas.

Goal-driven topics for "Application: G2P" research to address high-priority agricultural and environmental issues:

1. Host:pathogen interactions:

Infectious diseases are a major limitation to animal agricultural production. The presence of pathogens reduces animal health and welfare, as well as lowering production efficiency. Trade barriers prevent animal movement and decrease product export markets. Zoonotic diseases threaten human health and decrease the confidence of consumers in animal products as a wholesome food. Further, evolving and emerging infectious diseases can threaten the viability of entire sectors of the animal production industry. In particular, climate change (see section 4 below) is currently contributing to the emergence of new infectious disease and pest pressure in previously unaffected areas.

Presently, the full scope of host genetic control of resistance/tolerance to pathogens is poorly understood for most pathogens and most food animal species. Outbred farm animals harbor extensive genetic variation that has been largely overlooked in model species research. Given the large genetic variation in host resistance/tolerance, genomic approaches that survey and define the genetic variation between animals are an important first step for defining host-pathogen interaction. Accurately defining and obtaining the needed phenotypic measurements is difficult for farm animals. Breeding companies, which can effectively measure very large numbers of animals for phenotypes such as growth, lactation, and reproduction, have limited ability and/or incur additional expense to conduct major studies on pathogen resistance in their animal populations. The practice of protecting elite breeding animals from exposure to pathogens exacerbates the issue of producing commercial animals that can resist the myriad pathogens to which they will be exposed in the field. Thus, developing a greater understanding of the genetics and genomics of host:pathogen interactions, toward the goal of identifying biomarkers to improve disease resistance/tolerance in commercial animals, is a challenge best addressed in the public research sector. However, effective long-term academic-industry collaboration is required to ensure that early-stage research results can be rapidly applied to commercial populations through validation studies, and that commercial populations can be utilized for statistically well-powered genomic studies (e.g. GWAS or genomic selection) of disease trait records that are routinely obtained in the commercial environment, especially if coupled with veterinary record databases.

Use of high-throughput genomic and phenomic approaches will expedite the identification of genetic and other bio-markers to predict disease resistance/tolerance and genetically improve populations, which can then be rapidly translated to tests. However, there is currently inadequate understanding of the basic mechanisms, networks, and emergent behaviors that underpin complex host:pathogen

interactions, and what host responses lead to successful resistance/tolerance to infection, pathogen spread, pathology, and disease. Greater understanding of the fundamental mechanisms is needed to inform a greater level of precision in phenomic screening. Some of the basic scientific questions to be answered include:

- What are the important ligands on pathogens and the corresponding receptors on host cells?
- Which specific disease phenotypes are most predictive of resistance/tolerance?
- Can pathogen-derived molecules be used to accurately assess animal response to the pathogen without exposure to the infectious agent?
- Can host immune molecules be produced in recombinant form and then administered to enhance immunity?
- Do non-commercial populations have resistance/tolerance alleles that no longer exist, or are in very low frequency, in commercially selected populations?
- What is the interaction between the gene variants regulating host response to different pathogens?
- How does the natural microbiome (Resources: Genotypic, Phenotypic, Epigenetic variation, section 5) protect against disease incursion?

Selective breeding utilizing natural genetic variation for disease resistance/tolerance is being accomplished for some species, notably aquaculture animals and poultry, and to some extent dairy cattle. These species have life history and reproductive characteristics that generate large numbers of offspring that in turn allows the possibility of extensive phenotyping of full-sib animals. Research is needed to develop high-throughput disease phenomics that integrates genome-wide SNP analyses, gene expression, disease pathology, and measurement of associated production parameters. Establishing common pedigreed animal populations for phenotyping will allow the search for causal mutations that contribute to disease resistance. Cross-species platforms are needed to integrate data and pathways discovered in model organisms with agriculturally important species.

Vaccines have been highly effective for controlling animal diseases. Genomic tools can facilitate a better understanding of how effective vaccines work in outbred animals. Immunity represents the dynamic interactions of hundreds to thousands of proteins/genes, and new tools and technologies are needed to simultaneously quantify the interactions of both innate and acquired immune system components. Opportunities exist to improve design of vaccines, as well as to develop vaccine-ready animal populations that match host genotype with a specific vaccine.

Other priority areas and opportunities within the field of host-pathogen interactions include:

- Incorporate large-scale phenotyping with genomic selection for increased disease resistance.
- Develop evolutionary framework that facilitates selection of host genetic variation that minimizes disease and/or pathogen effects, and at the same time does not increase selective pressure on the pathogen to increase virulence.
- Utilize genomic technology and phenotyping to identify pattern-recognition receptors that confer broad-spectrum resistance.
- Utilize genomic tools to assess the immunological repertoire of animal populations.
- Understand the emergent properties of interacting host/microbial genetic networks.

2. *Feed efficiency and adaptation to sustainable diets*

Feed represents the major cost of animal production in all sectors. The efficiency with which a farmed animal can convert feed to a healthy, edible product in an ethically acceptable production system is

therefore a critical trait to target in G2P research. Encompassed within this goal area is the need to minimize waste produced and its environmental impact.

Heritability estimates for feed efficiency traits are typically moderate, meaning there is much opportunity to apply selective breeding for more efficient animals. Detailed recording of individual feed intake traits will be essential, and improvements in phenotyping will lead to a more accurate definition of the genetic regulation of these traits. An understanding of the mechanisms underlying this heritability will be useful for developing new gene markers for selection, as well as novel feed components. Characterizing the host interaction with the gut microbiome will be critical to a full understanding of these traits, and can be facilitated by the development of the new genomic and bioinformatic tools discussed earlier. Certain farmed animal species will be required to transfer from animal-based diets to increasing vegetarian diets (for example, farmed aquaculture species). This change is critical for sustainability of production, and research is needed to address the response of the animal to these new diets and the quality of product obtained. There is also opportunity for interaction with the plant genetics and genomics community to address the possibility of breeding plants for improved digestibility and nutrient content to use in animal feed.

3. *Biodiversity utilization*

In intensive modern livestock farming, genetic variation probably has eroded at a faster rate than it is created anew. Therefore, local or 'unimproved' breeds can be regarded as repositories to search for alleles that can be of economic importance in the future, to develop robust genotypes that can thrive in diverse environments. Nevertheless, livestock species exhibit levels of genetic variation that are at least as high as that in the human genome. With the large number of genomes being sequenced to completion in the coming years, an important task will be to develop new methods to analyze this information, in particular to infer selective events that may highlight genes of socioeconomic interest. This information and the analyses are complementary to classical association or linkage analyses. By scanning the genomes of breeds with unique phenotypes (e.g., health, thermal-stress resistance, group behavior) alleles or genes can be identified for inclusion in selection programs in commercial production populations.

4. Climate change adaptation

All livestock species are sensitive to increased magnitude and duration of heat episodes, but it should be noted that the effects of climate change extend beyond simply temperature. For instance, new parasites which are normally encountered only in tropical climates will extend their range outside their current geographical limits. Therefore, understanding the physiological effects of temperature is necessary, but not sufficient to fully adapt animal production to climate change and for expansion into new areas of the globe. Ideally, breeding values should be provided for a range of environments, accounting for genotype by environment interaction – and this poses formidable statistical and computational challenges.

5. Microbiome

The microbial community associated with gut and mucosal surfaces is recognized as a large, complex and changing assemblage; potentially comprising thousands of species and millions of genes and very many more gene products interacting in complex networks. For farm animals, the microbiome, and its impact on host physiology are poorly understood. Furthermore, most microbial species remain uncultured and this limits investigation of metabolic capabilities and their role in the environment. Work in humans and mice shows that the microbiome likely impacts important phenotypic traits (such as feed efficiency and the development of the immune system) in farm animals. Genomic tools can be used to

characterize the phylogenetic diversity, transcript/protein/metabolite response, and stability of the microbiome associated with farm animals under different production environments.

Potential priority areas and opportunities within the field of host-pathogen interactions include:

- Sequence and phenotype the microbiomes of major agricultural species
- Determine which microbes are commensal/symbiotic/pathogenic under different circumstances.
- Identify how immune homeostasis in the gut and other colonized areas is maintained.
- Determine the impact of different disease control strategies on the microbiome.
- Determine how perturbations of the host animal affect its microbiome, and conversely, whether modulation of the microbiome can be utilized to enhance phenotypic performance.

Overall it is clear that by understanding animal genotypic variation there will be major applications for advancement of animal and human health and well being (Fig. 2).

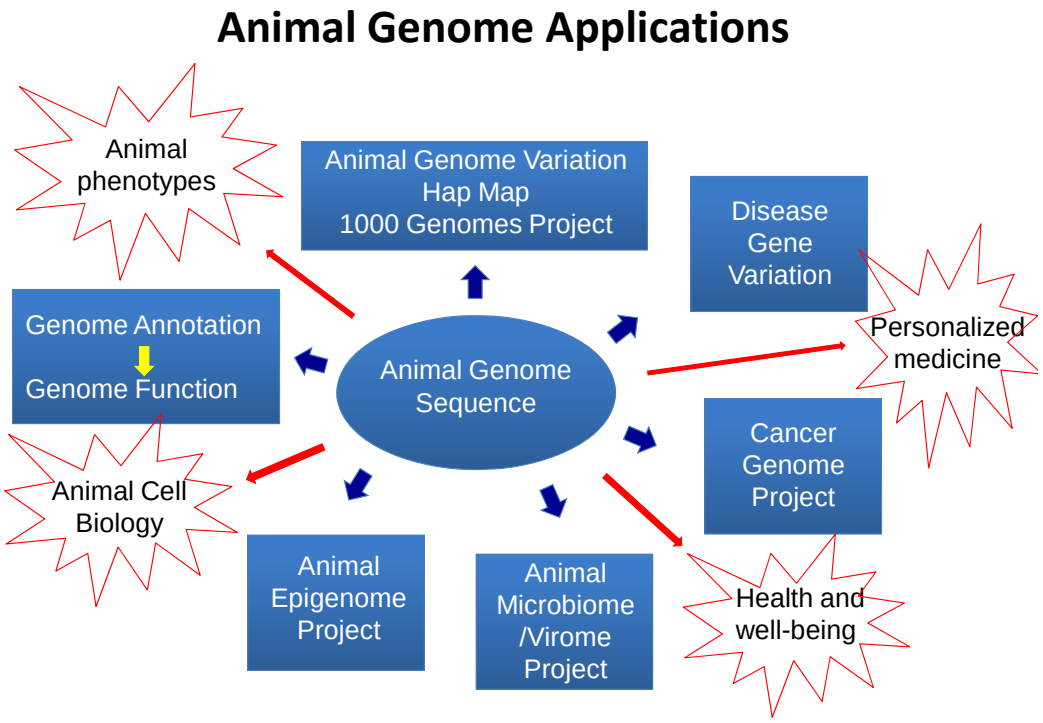


Figure 2. Agricultural Animal Genome Applications. Basic understanding of animal genomic variation (blue boxes) will numerous processes that impact agricultural animal and human health and well-being (stars). Image kindly provided by Jane Rogers, The Genome Analysis Centre (TGAC), UK.

Grand Challenges - Cyberinfrastructure

Each of the key topics described in this report will require cyberinfrastructure to support data storage, management, integration, mining, visualization and processing services. Key cyberinfrastructure requirements include support for large scale genome sequencing, assembly, annotation, and analysis; identifying genetic variation; modeling regulatory sequences; comparative mapping of genetic elements, diseases, and traits; storage of large data sets; collection, annotation, and analysis of epigenetic and complex gene interaction data; and supporting metagenomic and metabolomics analyses. An immediate need identified is the development of workflows to support high-throughput sequencing analysis and submission of -omics data to public data repositories (e.g. NCBI GEO, EBI ArrayExpress, EBI Proteomics Identifications). Furthermore, there is a need to develop analytic methodologies to facilitate integration of genotype association, expression data (protein, mRNA, miRNA), and epigenome, metabolome, and metagenomic data to fully characterize phenotypes of interest and thereby develop novel strategies to improve animal production.

For animal genomes, Ensemble (www.ensembl.org/) in Europe and NCBI (www.ncbi.nih.gov/) in the United States have the key existing cyberinfrastructure for reference genomes. Both sites currently support the reference genomes of many food animal species (cow, chicken, pig, turkey). Ensemble has developed a large number of tools to mine these reference genomes and to compare the information across different species. Strong support is needed to be able to incorporate the genomes of additional animal species that will be sequenced in the future. A strong interaction between these resources and resources developed for efficient data mining will be crucial.

Today, a platform has already been developed that helps plant researchers use tools and data more easily and efficiently (iPlant). It provides sustainable access to high performance computing, interoperable software analysis, and large data sets. The cyberinfrastructure is accessible for all levels of expertise, ranging from students to traditional biology researchers and computational biology experts. It is the vision of agricultural animal scientists to take advantage of this infrastructure and expand it to meet the needs of the agricultural animal community through the development of iAnimal and its interaction with existing animal genome resources, such as Ensembl.

We have already begun working with the iPlant Collaborative (www.iplantcollaborative.org/) to utilize and eventually further develop their existing cyberinfrastructure for animal biotechnology with tools and resources required to support animal genomics. Cross-disciplinary, community driven research groups will work collaboratively with iPlant to develop a cyberinfrastructure foundation to support the computational needs of the research community and facilitate progress toward solving major problems in animal science. This cyberinfrastructure will support a diverse group of animal researchers and bring together experts from various fields of biology, as well as computer science.

Genotype to Phenotype (G2P)

Elucidating the relationship between animal genotypes and phenotypes in complex and diverse environments is one of the foremost challenges in animal science. Animal phenotypes are determined by intricate interactions between host genetics and environmental conditions.

Very similar to the biomedical and plant communities, addressing the G2P challenge for animals requires the integrated efforts of specialists from disciplines such as biochemical; functional, quantitative, and

computational genetics/genomics; systems biology; bioinformatics; modeling; physiology; and computer science.

To better derive insights into how an animal's genotype influences phenotype, it is imperative for researchers to have ready access to an integrative cyberinfrastructure. The resources, described in the "Resources : Genotypic, Phenotypic, Epigenetic variation" section, need to be made publicly and globally available. Furthermore, it will be important for the agricultural animal community to adhere to standards and best practices so that these resources are open and interoperable. Moreover, these resources need to be species-independent to facilitate comparative genomics that can derive insights into how common phenotypes are expressed in different species.

Other key resources include the development of metagenomics and metabolomics to support G2P studies and complement genetics studies. While genetics and epigenetics can identify individual variation in the genome, metabolomics – measuring the metabolic signature – and metagenomics – measuring the microbial populations – can tell us how individuals or groups are affected, not only by small genetic changes, but also environmental changes. By combining these extra layers of information with SNP and epigenetic studies, researchers can target how individual variation affects the RNA and protein molecules that combine to generate phenotype. The development of expertise and resources that support metagenomics and metabolomics studies are likely to bridge the gap between genome and phenome. The new innovations in the iPlant project appear to provide the needed cyberinfrastructure to support these diverse collaborations using a cloud-based distributed architecture. Expansion of this cyberinfrastructure to include iAnimal, and further strengthening of existing cyberinfrastructures (e.g. Ensembl) should enable analysis of large, complex datasets and empower researchers with a variety of expertise to conduct analyses without being computational experts. The emerging cyberinfrastructure could also mirror the ELIXIR initiative (www.elixir-europe.org/about), a system with a central hub and integrated regional or national nodes.

Data Integration

Understanding genotype to phenotype (G2P) relationships can require the integration of numerous data types such as genes/proteins/metabolites/etc. with known or suspected roles. Impediments to data access and analysis include incompatible indexing, non-standardized storage or display formats, and the difficulty of maintaining awareness of new data sets as they are formed or, in time, superseded by continually evolving technologies.

Methods to describe and unify data sets into virtual systems will support animal project activities. The cyberinfrastructure needs to allow and build upon existing middle-ware systems that use metadata to achieve situational awareness of available data, the logical relationships between different data sets, and tools that enable users to find relevant information even when they are not sure what data may exist. The system needs to support data intended to be publicly distributed, as well as secure, private and/or user-local repositories, and enable information to be imported into statistical inference, visualization, and/or modeling tools and applications.

International coordination between initiatives like iPlant, iAnimal, and the EU ELIXIR program, as well as the incorporation of many smaller national initiatives to deal with the increased amounts of data being generated by next generation sequencing, will be a high priority.

Next Generation Sequence Pipeline

Next Generation Sequencing (NGS) technologies are rapidly changing the way we approach fundamental questions in biology. Technological advances in nanotechnology and sequencing chemistry have together revolutionized our ability to obtain low-cost, high-throughput DNA sequence. As these technologies advance at a rapid pace, they pose new challenges for standardizing sequence information and in automating computational tasks.

Establishment of an informatics pipeline that allows members of the animal research community to process NGS data using simple, user friendly interfaces is a critical need. Internationally, many groups (biomedical, life-science, plant, and animal) have established procedures to cope with huge data flows. It is essential that these bioinformatic resources become accessible for all researchers. Developers of such resources need to make tools that would allow users to import NGS sequencing files and output data using the algorithms commonly used by the community for processing DNA and RNA data sets.

The sequence analysis efforts would need to enable users to upload DNA or RNA sequencing data from their desktop, a remote server, or from the NCBI and/or EBI Sequence Read Archives, then view, manage, and perform basic analyses on the data in a user-centric workspace. Data management capabilities include annotation with metadata and pre-processing sequence data to remove non-biological sequence production artifacts (e.g. linkers, primers, etc). Scientists would need to be able to perform basic analytical workflows using their post-processed sequence data in a relatively short period of time with user-friendly utilities. Necessary utilities should permit sequence quality control of DNA/RNA sequence data; variant detection, including SNPs, insertions and deletions (indels), and CNVs in a test sequence compared to a reference sequence; and transcript quantitation and annotation.

Visual Analytics

A cyber infrastructure challenge in biology is in data integration and iterative, flexible analysis of experimental results. There are numerous types of data in the livestock sciences that are collected to represent the genetic, phenotypic, and environmental variation that drive complex biological systems. The list of data types affecting an individual species can be enormous, and many of these data types share little similarity. For this reason, data integration and visualization (or analysis results integration) is critical for the life sciences.

A goal would be the generation, adaptation, and integration of visualization tools capable of displaying diverse types of data from laboratory, production systems, in silico analyses and simulations, and other sources specific to G2P research. These displays would need to be designed to reveal underlying patterns, lead to novel hypotheses, provide concise syntheses, and support publication, collaborations, education, and other forms of dissemination (Web, podcast, etc.).

Discovery Environment

A Discovery Environment would be one of the ways users could interact with iAnimal type cyberinfrastructure. Rather than managing computing resource details, or learning new software for every type of analysis, the Discovery Environment would allow a user to handle all aspects of their bioinformatics workflow (e.g. data management, analysis, sharing large datasets, etc.) in one space.

A Discovery Environment could provide a Web interface and a platform to access computing, data storage, and analysis application resources as envisioned for iAnimal. The Discovery Environment would need to be designed to facilitate data exploration and scientific discovery by integrating analytical tools as modular components that may be used individually or in workflows, accessing data, and seamlessly

running tools on local or high performance computing nodes depending on the throughput and resource needs of the analysis. In addition, the Discovery Environment would need to employ provenance tracking of both primary and derived files to track and reproduce experiments, and collaboration tools enabling users to share data, workflows, analysis results, and data visualizations.

The Discovery Environment would need to be integrated with a data management system and computer resources, creating a unified environment in which researchers can access tools and data with an unprecedented degree of scalability. It features easy integration of new tools and data by any user, resulting in an ever-expanding set of analytical tools.

The Discovery Environment would need to have standard analytical tools and allow for additional tools to be added all the time. Standard tools would need to support data analyses in the following research areas:

- Clustering and network analysis
- QTL mapping and genome-wide association studies
- Sequence alignments and phylogenetic tree building
- Next Generation Sequencing
- Phylogenetic comparative methods and trait evolution

A scalable, distributed system for storing data files should be built upon highly redundant, high performance storage arrays that are geographically replicated. It needs to provide a framework for describing and locating data, a set of protocols for maximizing the transfer rate of very large sets of data, and a diversity of tools to access those data. Users could deposit their data in their iAnimal Data Store through a variety of means, and those data could be immediately accessible for downstream processing and dissemination from within the Discovery Environment.

Education/Training

If animal science research is to be translated into applications for the animal industry in a timely manner, bioinformatics/computational biology and statistical training will be absolutely required when training the next generation of animal scientists. These issues have strong implications for education. It is essential for animal science researchers to anticipate the importance and role of computational and statistical strategies, as well as practices in data management and data mining. There is also a critical need for researchers that are experts in the application and development of these strategies. Bioinformatics should be considered as a very dynamic discipline of its own and as an essential component that is critical for the success in many modern research projects. Revision of the biology curricula will be critical to overcome the current limitations in bioinformatics education.

The need to focus on training in bioinformatics and biostatistics and to recruit the next generation of scientists with a passion for agriculture was expressed clearly in this workshop. Although technical challenges of data portability and tool access remain, the creative challenge of developing tools and algorithms to better represent and visualize the huge volumes of data and add to the understanding of functional genomics are stimulating new collaborative opportunities. During this workshop, attention was drawn to the importance of reference animal populations, the paucity of animal bioinformatics applications and tools, and the need for data repositories with common/compatible standards. Each (large) research project should have appropriate resources for bioinformatics with plans for making data available and accessible, including genome annotation. The need for funding opportunities for international collaborative teams continues to be highlighted.

High priority need for collaboration:

- Training/education of the next generation and existing workforce with a passion for agriculture in appropriate bioinformatic concepts, tools, and methods.
- Development of tools to integrate, manage, represent, visualize, and analyze dense animal biology datasets.
- Design of reference resources with compatible standards and records from reference populations, including genomes and phenomes.

Modern research technologies enable biologists to embrace complexity and undertake holistic approaches in the study of biological systems at various levels and scales. They have also led to the generation and accumulation of an unforeseen and unprecedented volume and diversity of data. Although various data sets are inter-dependent and thereby reinforce or complement each other, thus far only a small minority of researchers have the necessary skills and expertise to make these connections in practice.

Communication/Public Acceptance

Effective communication about hazards and risks of a complex subject such as genomics in animals can generate trust and confidence in the competence of research. This understanding would serve as an important foundation for independent and fair decisions by regulators and public officials, thus strengthening the legitimacy of government action and improving compliance and enforcement.

The principal objective of the communication/public acceptance plan is to make recommendations on benefit/risk assessment and perceived benefit/risk communication during the process of expanding scientific knowledge base and innovative uses of animal genomics. It will involve basic researchers as well as commercial and governmental partners. It should emphasize that EU and U.S. agricultural research is committed to producing a safe and secure food supply that is enhanced by the use of the latest technologies.

This Benefit/Risk communication should be trustworthy, therefore open and transparent, based on facts and scientific knowledge, but also simple and understandable, focused on the target audience and in line with the societal questions and values of each region in Europe, the United States, and beyond. Communication will only be effective if the communicators are reliable and trustworthy spokespersons. The ABWG should therefore identify the lead communicators, and the relations they need to have with other stakeholders, the authorities and the public. A Genomics communication project should be started and supported by the United States and Europe; this will be called hereafter the Genomics Communications Group.

The Genomics Communication Group, and the whole EU-US Animal Biotechnology Working Group (ABWG), should systematically use good communication practices. This should not be done in isolation, but in collaboration with involved political authorities to form an integral part of the overall policy framework for Benefit/Risk management. The systematic use of good communication practices should result in commitment towards consistency of government decision making based on facts and known or perceived risks. The group should identify, or if necessary develop, objective Web sites that will help educate the public and regulators regarding the importance of genomics and biotechnology in improving food production.

The Genomic Communications Group should develop a communication package for individual scientists to adapt their own scientific descriptive presentations towards more goal-oriented presentations. By involving Public Relations professionals, the group could expand their communication training. This would help to avoid defensive posturing and involve all stakeholders, both positive and, potentially, negative groups.

The Genomics Communications Group should monitor scientific progress. It should monitor other sources of information and have a clear mandate to publicly react, if there is under- or overstatement of benefits and risks. Avoidance of amplification of benefit or risk is essential to maintain political and public trust.
In different parts of the world, the level of influence of pressure groups on the risk management decisions after an objective risk evaluation of a new method or product varies. It is recommended that the impact assessment of this influence on the approval process of certain “improvements under discussion” would include the question: What if the risk assessment would be disregarded in one geographic region and not in other regions?

The European Group on Ethics (EGE) advising the European Commission President has issued a set of ethical opinions, reports including criteria and recommendations based on their knowledge, perception, and interviews. The Genomics Communications Group and the ABWG should devote time to study these for potential use in adapting them to discussion of the evolution of science and societal changes. They should be informed about the values and principles that would be acceptable and those that are generating questions or even opposition. This will help delineate a clear set of limits that would need to be identified.

- **Opinions 2008-2009**

Opinion of the European group on Ethics in Science and New Technologies to the European Commission Commission, Ethics of synthetic biology. No 25. 17/11/2009

Opinion of the European group on Ethics in Science and New Technologies to the European Commission, Ethics of modern developments in agricultural technologies. No 24. 17/12/08

Opinion of the European group on Ethics in Science and New Technologies to the European Commission, Ethical aspects of animal cloning for food supply. No 23. 16/01/2008

APPENDICES:

I. Workshop Agenda
II. Speakers
III. Participants

Strategic Priorities to Realize the Promise of Genomics to Animal Health, Well-being & Production: workshop of the EU-US Animal Biotechnology Working Group of the EU-US Task Force on Biotechnology

Agenda

8th of November 2011

7:45-8:00 Meeting introduction.

8:00-10:15 **Genome sequencing and the development and use of genomics-enabled tools**
Moderators: Michel Georges (EU), Shane Burgess (US)
- Brett Tyler, Systems biology of infectious disease in plants, Virginia Bioinformatics Institute
- Stephen Goff, cyberinfrastructure collaborative for the plant sciences, University of Arizona
- Jane Rogers, The Genome Analysis Centre (TGAC), UK

10:15- 10:30 Break

10:30-12:45 **Informatics and cyberinfrastructure**
Moderators: Martien Groenen (EU), Chris Elsik (US)
- Dan C. Stanzione Jr, Engineering Discovery Environment, Texas Advanced Computing Center
- Paul Kersey, European Life Sciences Infrastructure For Biological Information, EBI, UK

12:45-14:00 **Lunch**

14:00-16:15 **Phenotypes and Phenotyping**
Moderators: Elisabetta Giuffra (EU), Gary Rohrer (US)
- Tom Weaver, Mouse Phenotyping Consortium, Mary Lyon Centre, Oxford, UK
- Monte Westerfield, Zebrafish database, University of Oregon, USA
- Mark Crowe, Australian Plant Phenomics Facility, University of Adelaide, Australia

16:15 **Break**

18:30-21:00 Dinner in Sheraton meeting/dining room and moderated open discussion with invited speakers.

9th of November 2011

7:30-8:30 Meet, breakfast, get a seat. Load PowerPoints.

8:30-10:30 Lightning talks from the participants. Four minutes each, one (optional) slide:

1. Who you are, where you are, what you do.
2. Priorities and needs for the future in the context of the meeting's remit.
3. Why you are at this meeting.
4. What you see as the most pressing priorities in each of the topic areas.
5. What you took from the previous day's speakers and discussion.

10:30-11:00 **Break**

11:00-13:00 Moderated open discussion and brainstorming about the key needs, investments and milestones for delivering on the promise of genomics for animal health, wellbeing and productivity.

13:00-14:00 **Lunch**

14:00-17:00 Break into three working groups to identify a plan for the specific path to meet the key needs, investments and milestones in a EU/US and other country perspective. Each working group chaired by an ABWG scientific member and another to take notes.

Group One
Moderators: Michel Georges (EU), Shane Burgess (US)
Group Two
Moderators: Martien Groenen (EU), Chris Elsik (US)
Group Three
Moderators: Elisabetta Giuffra (EU), Gary Rohrer (US)

17:00-18:00 Three groups reconvene to summarize their discussions.

ABWG scientific members will meet to compile a document for circulation and discussion on Day Three.

10th of November 2011

7:30-8:00 Meet, breakfast, get a seat. Circulate discussion document.
8:00-10:30 Open discussion on document and assigning writing responsibilities

10:45-12:00 Identify specific topics for working group meetings in 2012 and 2013 that will review, critique and refine the document to be produced in 2012 and to further specifically advise the agencies.
12:00 **Meeting ends**

Speakers EU-US ABWG Workshop		
Name	**Affiliation**	**Country/Region**
Genome Sequencing and Tangible Value		
Tyler, Brett	Virginia Bioinformatics Institute	USA
Goff, Stephen	University of Arizona iPlant Collaborative	USA
Rogers, Jane	The Genome Analysis Centre	EU
Informatics and Cyberinfrastructure		
Stanzione, Dan C., Jr.	Texas Advanced Computing Center, iPlant; Extreme Science and Engineering Discovery Environment	USA
Kersey, Paul	European Bioinformatics Institute	EU
Phenotypes and Phenotyping		
Weaver, Tom	Director of Mary Lyon Centre, MRC Harwell Harwell Science and Innovation Campus	EU
Westerfield, Monte	University of Oregon, Institute of Neuroscience, Zebrafish Database (ZFIN)	USA
Crowe, Mark	The Plant Accelerator®, University of Adelaide	AUS

Participant List EU-US ABWG Workshop		
Name	**Affiliation**	**Country/Region**
Adelson, David	University of Adelaide	AUST
Albers, Gerard	Hendrix Genetics	EU
Bellis, Diane	United Soybean Board	USA
Boggess, Mark	United States Department of Agriculture, Agricultural Research Service (ARS)	USA
Burgess, Shane	University of Arizona	USA
Cleveland, Beth	United States Department of Agriculture, Agricultural Research Service (ARS)	USA
Elsik, Chris	Georgetown University	USA
Gaines, Mallory	National Cattlemen's Beef Association	USA
Georges, Michel	Liège University	EU
Giuffra, Elisabetta	INRA (French National Institute for Agricultural Research)	EU
Groenen, Martien	Wageningen University	EU
Houston, Ross	Roslin Institute	EU
Jorjani, Hossein	Interbull	EU
Kappes, Steven	United States Department of Agriculture, Agricultural Research Service (ARS)	USA
Kapur, Vivek	Penn State	USA
Lamont, Susan	Iowa State University	USA
Lequarre, Anne-Sophie	European Commission	EU
Lunney, Joan K.	United States Department of Agriculture, Agricultural Research Service, Animal and Natural Resources Institute (ANRI)	USA
Matukumalli, Lakshmi	United States Department of Agriculture, National Institute of Food and Agriculture (NIFA)	USA
McCarthy, Fiona	The AgBase Databases, Mississippi State University	USA
Mileham, Alan	Genus/PIC	EU/USA
Perez-Enciso, Miguel	Universitat Autònoma de Barcelona	EU
Qureshi, Muquarrab	United States Department of Agriculture, National Institute of Food and Agriculture (NIFA)	USA
Reecy, James M.	Iowa State University	USA
Reed, Kent	University of Minnesota	USA
Rohrer, Gary	United States Department of Agriculture, Agricultural Research Service (ARS)	USA
Sarropoulou, Elena	Hellenic Centre of Marine Research	EU
Sheely, Deborah	United States Department of Agriculture, National Institute of Food and Agriculture (NIFA)	USA
Silverstein, Jeff	United States Department of Agriculture, Agricultural Research Service (ARS)	USA
Sonstegard, Tad	ARS, Bovine Functional Genomics Lab	USA
Stella, Alessandra	Parco Tecnologico Padano	EU
Vanhemelrijck, Johan	Secretary General of EuropaBio	EU
Warkup, Chris	KTN Biosciences	EU
Wiens, Greg	United States Department of Agriculture, Agricultural Research Service (ARS)	USA