***Aquaculture: Combined Coordinator and Technical Report*** (Caird Rexroad III and John Liu)

**Genome Sequencing:** The Oyster Genome Consortium was successful in establishing a whole genome sequencing project for the Pacific oyster (*Crassostrea gigas*). Sequencing and assembly of a catfish reference genome is also underway with participants from ARS Catfish Genetics Research Unit, Auburn U., USDA-ARS Bovine Functional Genomics Laboratory, and U. of British Columbia. Gene transcripts from various tissues of multiple individual catfish with diverse genetic background were also sequenced. Projects to identify EST and define the transcriptomes of various tissues were conducted in catfish, rainbow trout, brook trout and striped bass.

**Genome Mapping:** The USDA-ARS National Center for Cool & Cold Water Aquaculture (NCCCWA) rainbow trout map was used for producing a first generation integrated physical and genetic map. A high density RAD (restricted site associated DNA) genetic map of Swanson x Whale Rock recombinant double haploids is being constructed using approximately 7,600 SNPs to aid in future assembly of a reference genome sequence for trout. The 2nd generation NCCCWA rainbow trout genetic map is now available through G-browser at the Animal Genome website of the NRSP-8 bioinformatics group. A first SNP genetic map for Pacific white shrimp was built with 418 SNP markers mapped onto 45 sex-averaged linkage groups. This SNP genetic map lays the foundation for future shrimp genomics studies. Scientists from the USDA-ARS NCCCWA, VIMS and N. Carolina State U. (NCSU) developed a linkage map for striped bass by genotyping two half-sib families at 289 microsatellite DNA markers and assembled a map with 26 linkage groups.

The USDA-ARS SNARC generated 192 crosses of *Morone* using National Breeding Program foundation stocks and completed studies evaluating heritability of phenotypic variation growth of hybrid striped bass as tank-reared fingerlings. Scientists from SNARC and the U. Arkansas at Pine Bluff evaluated the genetic and phenotypic influence of parental traits on hybrid striped basslarval size and quality, and the influence of genetic factors on metabolic and stress-related traits, discovering that female phenotype does not significantly affect larval traits (e.g. growth) but that genotype does have a significant affect. This finding is significant because any increase in larval size at hatch resulting from selection would reduce the need for live feeds, which could make year-round tank production of fry and fingerlings economically viable for industry. SNARC and NCSU researchers also distributed advanced fingerlings and mature broodfish from National Breeding Program stocks to HSB producers engaged in propagation of commercial domesticated broodstocks.

**Database Activities**: Many useful links for aquaculture can be found at http://www.animalgenome.org/aquaculture/. In collaboration with John Liu, Auburn U., a Catfish SNP Project web site (http://www.animalgenome.org/catfish/cbarbel/), a Teleost Alternative Splicing Database (http://www.animalgenome.org/tasd), and a Catfish COI DNA Barcode Database (<http://www.animalgenome.org/fishid/>) have been established. The bioinformatics coordinators have helped Moh Salem of West Virginia U. to set up web blast and data download of the rainbow trout transcriptome data characterized using Sanger and Next GENeration sequencing data (http://www.animalgenome.org/aquaculture/salmonids/rainbowtrout/EST\_WV.html).

***Cattle: Coordinator Report*** *(Juan Medrano)*

**Bovine Genome sequence:** Currently, two genome assemblies have been produced from the sequence data generated by Baylor College of Medicine from Line 1 Hereford cattle, Btau\_4.2 and UMD3.1. About 26,000 genes are identified on both assemblies. Genome annotation between the two assemblies is slightly different and needs improvement, and many problems exist related to gene structure. It is clear to the community that one new improved universal reference assembly is critically needed in order to facilitate the definition of gene models, gene annotation, haplotypes definition and identification and mapping of copy number variants, as well as to allow for a comprehensive transcriptome analysis and gene discovery. A white paper describing the status and future direction of the bovine assembly was prepared in collaboration with USDA, US university scientists and international collaborators, and submitted to USDA-NIFA in October 2010. The white paper went directly to NIFA for rfp development and for use with their review panels, and to NIH for consideration for additional funding support. The development of the bovine genome assemblies illustrates the value of the support and collaboration through NRSP-8. Multiple efforts currently exist around the world to sequence elite sires for the purpose of developing the next generation of animal evaluation tools. Coordination efforts will be placed to catalog this resource and to promote sharing of information.

**SNP genotyping chips**: Two new high density genotyping chips have been developed by Illumina and Affymetrix including ~800k SNP each. The development of these chips was a joint effort from investigator members of NRSP-8 and private companies (e.g., Pfizer). These chips represent a unique resource for refining QTL position for fine mapping, identification of copy number variation and expanding the application of genome selection to a larger group of cattle breeds. The Illumina HD chip has been evaluated by several investigators who have reported on the very high quality of the call rates and reliability of genotype calls of this chip.

**Cattle microarrays:** Coordination was provided for the development of a new bovine Affymetrix “gene sampling array”, Custom WT Btau 4.0 Array. This array includes ~550k probes representing most exons for increased sensitivity to measure gene expression. The creation of this array was coordinated with input from US and German investigators and is being distributed by Affymetrix.

**Database Activities:** New additions to the NAGRP site (http://www.genome.iastate.edu/cattle/) are the bovine UMD 3.1 assembly with new NCBI annotation on GBrowse and a data sharing repository. In the past year, 2,423 new cattle QTL have been added . In addition, cattle QTL can now be viewed relative to the UMD assembly; http://www.animalgenome.org/cgi-bin/gbrowse/bovine/) and Btau4.2 assembly; http://www.animalgenome.org/cgi-bin/gbrowse/cattle/.

***Equine Coordinator Report*** *(Ernie Bailey)*

**Map Development:** During 2010, many research applications were made with the existing sequence. Most of the workshop effort has gone into research applications. However, because of variation in genome organization and gene duplication and deletion within a species, it is clear that the workshop needs to take a renewed interest in genome mapping for the horse, most likely with development of whole genome sequences for additional horses using next generation sequencing.

**Comparative Mapping:** The incentive of NHGRI for sequencing the genome of the horse was to compare the organization of genes of humans to that of the horse. To a large extent, homology was assumed in context of the annotation of the horse using genes known from other species. However, use of RNAseq data is providing gene expression information for horse tissues as well as more accurately identifying the structure and splice variation of horse genes.

**Marker Density and QTL Mapping:**  During 2010 numerous studies were published or presented at scientific meeting by scientists from NRSP-8 member stations on the topics of developmental bone diseases (osteochondrosis and related diseases), muscle diseases (polysaccharide storage myopathy), neurological disease, growth and stature, dwarfism, bone fracture and aspects of performance and infectious disease. A special issue of Animal Genetics was published devoted to this topic and is described below..

**Shared Resources**: DNA and relevant analyses for radiation hybrid mapping are available through NRSP-8 member scientists at Texas A&M. BAC library clones are available through a commercial enterprise at the Children’s Hospital of Oakland Institute as well as through the INRA at Jouy-en-Josas, France and Texas A&M University. Samples from horses phenotyped for MHC and other hereditary traits were shared among participants. A commercial SNP assay system for 55,000 SNPs (Illumina Equine SNP50) was used extensively during 2010 for genetic mapping and gene discovery. NRSP-8 funds were used to help support administration of that resource. At the end of 2010 Illumina ceased production of that product and workshop participants collaborated to develop a partnership with Geneseek (Lincoln, Nebraska) to develop a new 74K Illumina chip which will be genotyped at the Geneseek facility (see below). The use of that assay has been very effective and subject of multiple manuscripts published in a special issue of Animal Genetic devoted to horse genomics research (Animal Genetics 42, supplement 2). The issue was sponsored by the Dorothy Russell Havemeyer Foundation, but coordinator’s funds were used for incidental expenses associated with the issue. In parallel, there are several efforts to develop tools for investigation of gene expression including hybridization and sequencing methods. Information about obtaining access to these resources is available at the website for the Horse Genome Workshop: http://www.uky.edu/AG/Horsemap.

**Database Activities**: A major entry point for databases and other relevant information about the horse genome workshop and participants is the workshop website: <http://www.uky.ledu/AG/Horsemap>. Two databases compile published genetic data for horses: http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/intro2.pl?BASE=horse; http://www.thearkdb.org/. . Several genome browsers have been developed at the University of California, Santa Cruz, ENSEMBL and NCBI. A SNP database is available: http://www.broad.mit.edu/mammals/horse/.

The MacLeod Lab (University of Kentucky) launched its Equine Genome Browser. A consensus protein-coding equine gene set was generated by combining in silico gene structure predictions from Ensembl and NCBI with experimental structural annotation determined by RNA-sequencing (RNA-seq) experiments (Coleman et at 2010). This browser was developed to support the analysis of equine gene structural annotation. The browser displays consensus gene models along with their supporting Ensembl and NCBI predictions and RNA-seq derived structures (<http://macleod.uky.edu/equinebrowser/>)

***Poultry Coordinator Report*** *(Jerry Dodgson and Hans Cheng)*

**Reference linkage map:** Linkage mapping has transitioned almost solely into high throughput SNP (single nucleotide polymorphism) assays. Coordination funds have been committed to SNP chip development and distribution. Very high density SNP mapping (ca. 500,000 SNP) panels are being developed and will likely be employed in GWAS and genome-assisted selection efforts.

**Physical and comparative maps:** Physical mapping of the turkey genome is complete, along with the construction of a detailed comparative chicken-turkey BAC contig-based comparative map that was used for the assembly of the first draft turkey genome sequence (see below).

**Chicken genome sequence:** Next generation (next gen) sequencing has been applied to the chicken genome in hopes of obtaining the roughly 5% of missing sequence (predominantly on the microchromosomes) in the current chicken assembly, but so far this has made limited progress. A new build of the chicken genome sequence that combines the original reads, next gen reads (Roche and Illumina) and the near-finished quality of the Z sequence done by Bellott et al. (Nature 466:612-616, 2010) is being completed by the U. of Maryland Center for Bioinformatics and Computational Biology. Further efforts to capture missing, microchromosomal sequence have been proposed in a whitepaper submitted for review by USDA NIFA and NIH. A number of additional chicken genomes have been or are being sequenced (e.g., Rubin et al., Nature 464:587-591, 2010). The cost of next gen sequencing is now low enough that coordination funds have been committed to add new genomes of wide interest to participants.

**Turkey genome sequence:** The Turkey Genome Sequencing Consortium has generated a first draft sequence of the turkey genome (Dalloul et al., PLoS Biology 8(9):e1000475) using a combination of next gen reads, along with the turkey BAC contig-based comparative map alignments noted above. Coordination funds were committed to aid in this effort which also enjoyed support from VaTech, BARC and U. of Minnesota, among others (the effort also garnered support to both Virginia Tech and BARC from USDA-NIFA-AFRI). Sequence assembly was led by Aleksey Zimin, Steven Salzberg and colleagues at the U. of Maryland Center for Bioinformatics and Computational Biology. Efforts are on-going to improve the annotation of genes and fill gaps in the turkey sequence.

**Chicken microarrays:** In the past, coordination funds have been used to provide samples of the 44K element long oligonucleotide chicken array made by Agilent Corp. to several NRSP-8 participants, along with a new 244K whole genome long oligo array that can be used for comparative genome hybridization and whole genome transcriptional profiles. Alternatively, other participants chose to be provided GeneChip® Chicken Genome arrays from Affymetrix, Inc. Some coordination support has also been committed to Illumina RNA-sequencing and Agilent chip-based transcriptional profiling, partly in hopes of filling in missing sequences.

**Database activities:** The NRSP-8 Bioinformatics Coordinator, Jim Reecy, and Susan Lamont, along with Shane Burgess, represent poultry interests on the advisory committee for this group. Poultry bioinformatics has also benefitted from support at several other locations. A survey of chicken QTL (Abasht et al., Poultry Science 85:2079-2096, 2006) is made available from the NRSP-8 Bioinformatics team at http://www.animalgenome.org/QTLdb/chicken.html. Gene Ontology information for chicken genes is available at AgBase (http://www.agbase.msstate.edu/), mainly through the efforts of Shane Burgess and colleagues at Mississippi State. GEISHA (http://www.geisha.arizona.edu/geisha/microarray.jsp) also provides functional genomics data with an emphasis on graphical presentation of in situ hybridization during embryonic development. GEISHA is led by Parker Antin and colleagues at the U. of Arizona. Dr. Antin also led the effort that obtained NIH recognition for chicken as a model biomedical species (http://www.nih.gov/science/models/gallus/) and has also led the development of "BirdBase", an Aves-specific Model Organism Database (MOD) that can be used as a fundamental resource for all avian research communities: http://birdbase.arizona.edu/birdbase/. Carl Schmidt (U. of Delaware) has led the effort to develop Gallus GBrowse and, more recently, Turkey GBrowse which is delivered through the BirdBase website. We maintain a homepage for the NRSP-8 U.S. Poultry Genome project (http://poultry.mph.msu.edu) that provides a variety of genome mapping resources, including our newsletter archive.

**Impact:** This project is generating tools through which the genome sequence can be used to locate inherited production trait alleles and apply the DNA sequence to ascertain the physiological basis for those traits. It has resulted, among other things, in the generation of the complete sequence of the chicken and now the turkey genome. Industries have begun to apply the sequence and SNP we generated to characterizing and improving production lines using genome-wide marker-assisted selection. Since publication of the first draft of the chicken genome sequence, a shift has been made from providing and supporting physical genomics resources to those focused on gene expression and function.

***Sheep Coordinator Report*** *(Noelle Cockett)*

**Ovine Linkage Map:** The latest release of the linkage map (SM5) contains 2,528 loci across 3,800 cM, with 1,420 unique locations and average marker spacing of 2.5 cM. About 1,100 loci on SM5 are SNPs from a 1.5K pilot chip and the rest are primarily microsatellites, all genotyped across the International Mapping Flock (IMF). The linkage map can be viewed at (http://rubens.its.unimelb.edu.au/~jillm/jill.htm) on the Australian Gene Mapping Web Site, which is maintained by Jill Maddox, University of Melbourne, Australia. Genotypes from the 50K SNP BeadChip generated across the IMF are being analyzed by Dr. Maddox, with over 44,000 SNPs currently assigned to a chromosomal location.

**Ovine Radiation Hybrid Panel:** Sheep Coordinator funds have contributed to the development of an ovine radiation hybrid (RH) 5,000 rad panel (USUoRH-5,000). Around 300 markers have been added to the existing ovine whole-genome RH map within the last year using the USUoRH-5,000 panel. The addition of these markers increased marker density from 1.51 Mb/marker to 1.13 Mb/marker and the total map size increased ~ 37% in comparison to the previous version of the RH map. In addition, cross-species comparative maps based on marker-dense maps and high-coverage genome sequences were used to identify homologous synteny blocks (HSBs) and chromosome evolutionary breakpoint (EBRs) between sheep and other mammalian species. The number of homologous synteny blocks and chromosomal breakpoints between sheep and the human, cattle, horse and dog genomes were 216/54, 95/39, 122/61 and 135/75, respectively. Of the 229 conserved chromosomal segments, seventeen on human chromosomes (HSA1, 2, 3, 4, 6 and 21) and three on bovine chromosomes (BTA19, 27 and 28) had not been previously identified.

The 50K SNP BeadChip has also been typed on the USUoRH-5,000 panel and the INRAoRH-12,000 panel. Because the genomic constitution of RH clones differs significantly from the simple diploid organization of genomic DNA, a dedicated algorithm was needed to call the RH panel SNP genotypes from the raw intensities provided by the Illumina typing platform. Using this algorithm, an RH map was constructed for each ovine chromosome and then combined into a whole genome RH map comprised of 39,856 SNPs. The RH chromosome maps were developed using a comparative mapping approach that established the virtual sheep genome (VSG) as a reference for comparing alternative orders of markers.

**Sheep Genome Reference Sequence:** The ISGC is now working on the completion of a whole genome reference sequence. Sequence data for this project were generated at two sequencing facilities (Beijing Genomics Institute and the Roslin Institute) from DNA of a Texel ewe and a Texel ram, respectively. The first step of the reference sequence assembly involved the de novo assembly of 75X reads from the Texel ewe into contigs and scaffolds. Once that was completed, sequences from both animals were used for gap filling. The assembled scaffolds (2.71 Gb) cover approximately 92% of the ovine genome. In order to define the expressed portion of the genome, mRNA-seq was performed on seven tissue samples (heart, liver, ovary, kidney, brain, lung, and white fat) of the Texel ewe. This information will be used for annotation of the genome sequence. In addition to the reference assembly, about 5 million SNPs were identified in separate analyses of the male and female Texel sequences. The genetic and RH maps are contributing independent and complementary information to the ongoing assembly of the ovine whole genome reference sequence. Comparison of contig positions on the sequence scaffolds with locations in the genetic and RH maps have allowed improvement of the assemblies of scaffolds and super-scaffolds.

**Database Activities:** The Sheep QTLdb has been migrated from its Australia site to the site at Iowa State University (http://www.animalgenome.org/QTLdb/sheep.html; 264 new sheep QTL have been added to the Sheep QTLdb).

***Swine Coordinator Report*** *(Max Rothschilds)*

**Sequencing Efforts:** The Swine Genome Sequencing Consortium (SGSC) continued its efforts this past year and considerable advances have been made. Build 10 for the Sus scrofa reference genome sequence was released Monday, September 20 thanks to the efforts of many people and great collaboration across the world. The sequence and accompanying information was in a final version and released from TGAC's ftp (FTP site: ftp.tgac.bbsrc.ac.uk ; User: pig10 ; Password: Sscrofa10 ). This final version was based on the latest freeze of the physical map. The assembly is the result of the integration of all the sequenced clones and contigs produced by SOAPdenovo and Cortex whole genome shotgun (WGS) assemblies. These WGS assemblies were generated using Illumina reads sequenced at BGI and the Sanger Institute (~40X coverage). As part of the release AGP files with information about the source of every contig were provided. The WGS contigs were submitted to EMBL/Genbank, and after that the WGS contigs were to be renamed in the AGP with the corresponding accession numbers. This assembly provides an almost complete coverage of the pig genome. Additional details will be presented as they become available. The “marker” paper has recently been published in which the Consortium sets outs its plans for the analysis and publication of a draft pig genome sequence. These plans were presented to participants in the Pig Genome III conference held at the Wellcome Trust Sanger Institute, 2-4 November 2009 when a series of analysis working groups were established. Please see BMC Genomics 2010, 11:438 (http://www.biomedcentral.com/1471-2164/11/438

**Map Development Update:** New gene markers were identified with the development of the 60K SNP chip. These new markers are being integrated with the development of Build 9 and the new build 10 as maps now are based on the pig sequencing efforts.

**QTL, Candidate Genes and Trait Associations:** QTL, SNP and trait associations have continued to be reported on all chromosomes for many traits. Candidate gene analyses have proven successful with several gene tests being used in the industry for many traits including, fat, feed intake, growth, meat quality, litter size and coat color. The PigQTLdb (http://www.animalgenome.org/QTLdb/pig.html) is an excellent repository for all of these results.

**Porcine SNP chip:** Illumina and the International Porcine SNP Chip Consortium developed a porcine 60K+ SNP and has shipped it to many researchers worldwide. Researchers that did not place an order can contact Illumina for further information or questions at http://www.illumina.com/contactMe.ilmn?CS=1. The original publication was Ramos et al. 2009.

**Database Activities:** New QTL continue to be curated into the Pig QTL Database. There are now 5,986 QTLs in the database representing 581 pig traits and can be seen at (http://www.animalgenome.org/QTLdb/pig.html). Efforts are being made to update the newest pig genome information in several areas including (1) alignment with pig QTL among other genome features (http://www.animalgenome.org/gbrowse/) and (2) blast service to allow the community pig gene analysis and annotation activities. The NAGRP Bioinformatics Team has set up a pig gene Wish List which is seen at (http://www.animalgenome.org/cgi-bin/host/ssc/gene2bacs) which is playing an active role to help the pig genome annotation activities. The pig genome sequencing is actively carried out at Sanger Institute

(http://www.sanger.ac.uk/Projects/S\_scrofa/) and the latest sequence assembly and genome annotation results can be found at the http://www.animalgenome.org/cgi-bin/gbrowse/ssc/. More updated pig genome sequencing information can be found at http://www.animalgenome.org/pigs/genomesequence/.