

U.S. Poultry Species Coordination Activities
Supported by Allotments of Regional Research Funds, Hatch Act
For the Period 1/1/08-12/31/08

Overview: Coordination of Poultry Genome Mapping under the National Animal Genome Research Program (NAGRP) is a joint effort of Michigan State University (MSU) and the USDA, ARS, Avian Disease and Oncology Laboratory (ADOL). CSREES support is allocated via NRSP-8. The NAGRP is made up of the membership of the Animal Genome Technical Committee, including the Poultry Species Subcommittee.

FACILITIES AND PERSONNEL: Jerry Dodgson, Department of Microbiology & Molecular Genetics, MSU, serves as Coordinator with Hans Cheng of ADOL as Co-Coordinator. Both MSU and ADOL provide facilities and support.

OBJECTIVES. Note: NRSP-8 was renewed as of 10/01/08. Since most of the current reporting period was within the previous term, the objectives for that period are listed here. Note that we have achieved or exceeded all of these goals over the five year period of that project (10/1/2003-9/30/08). 1. Develop high resolution comparative genome maps aligned across species that link agricultural animal maps to those of the human and mouse genomes, 2. Increase the marker density of existing linkage maps used in QTL mapping and integrate them with physical maps of animal chromosomes, and 3. Expand and enhance internationally shared species genome databases and provide other common resources that facilitate genome mapping.

PROGRESS TOWARD OBJECTIVE 1. High resolution poultry genome maps.

The Reference Linkage Map(s). The genetic linkage map of the chicken has provided a framework for numerous QTL and other mapping experiments and a platform on which genome sequences have been assembled and linked to chromosomes. In connection with the genome sequence, the Beijing Genomics Institute randomly sequenced 0.25X, each, of a broiler, layer and Silkie genome, generating 2.8 million potential SNPs for high resolution linkage mapping experiments (International Chicken Polymorphism Map Consortium, *Nature* 432:717-722, 2004). In work supported by a consortium of industry, NRI Tools & Reagent grant, ARS and NRSP-8 funding, Illumina Corp was contracted to obtain ~3000 SNP genotypes, each, from ~5300 birds. About 88% (2733) of the SNP assays worked and almost all of the submitted DNAs were successfully typed. Since members of the East Lansing and Wageningen reference linkage families were included among the panel, these data greatly enhanced the chicken linkage map, more than doubling the number of markers, and were critical in the second build of the genome sequence. Not only have these data assisted in framework and QTL linkage mapping, they provide the most thorough picture to date of the genetic diversity of commercial breeding stock, worldwide (Muir et al., *P.N.A.S.* 105:17312-17317, 2008). A parallel effort by EU scientists genotyped approximately 13,000 SNPs in a variety of birds, including a Red Junglefowl (RJJ) x White Leghorn F2 cross, similar to the East Lansing reference backcross, which has resulted in a consensus map containing 9,000+ markers (Groenen et al., *Genome Research*, accepted). Currently, a 60K SNP Illumina iSelect genotyping array is being developed, and coordination funding in the coming FY will be committed to aid in its use to enhance the SNP linkage map. Another related project, currently underway, is to develop an even more dense SNP map specifically useful for the East Lansing reference linkage map population by employing reduced representation, high throughput sequencing of the UCD003 genome to generate additional SNP between the UCD001 and UCD003 parents of that population.

PROGRESS TOWARD OBJECTIVE 2. Physical maps and map integration.

Chicken BAC libraries, prepared in part with NRSP-8 and NRI Tools & Reagent funding, were fingerprinted extensively and integrated with linkage and gene maps. These data were employed to generate a second generation BAC contig map comprised of 260 contigs, most of which have been anchored to the genetic linkage/chromosome map (Wallis et al., *Nature* 432:761-764, 2004). The BAC contig physical map was updated in parallel with the second build of the chicken genome sequence. The first assembly of the draft 6.6X chicken sequence was done at the Washington U. Genome Sequencing Center (WUGSC) and released on March 1, 2004. Additional sequence data, physical, RH and SNP data were used to assemble a second, improved build of the chicken genome, released in May, 2006. The second build moved a large portion of the previously unplaced sequence contigs into specific chromosomal locations and enhanced the general contiguity and accuracy of the sequence assembly. High throughput next generation sequencing now is being used at WUGSC in hopes of obtaining the ~5% of missing sequence (predominantly on the microchromosomes) in the current chicken assembly and generally improving coverage in the next assembly. These sequences will be aligned, in part, using the 60K SNP data mentioned above, as well as pre-existing map data.

Physical mapping of the turkey genome is also on-going. A second turkey BAC library has been constructed to supplement the existing turkey CHORI-260 library. Both libraries have been used to generate over 43,000 BAC end sequences, along with over 23,000 BAC overgo hybridization assignments on the CHORI-260 library to distinct markers or genes. This includes most of the existing STS genetic linkage map markers available in turkey, as well as numerous chicken markers. Over 85,000 turkey BAC fingerprints have also been generated, allowing the construction of a first generation BAC contig physical map and a comparative chicken-turkey BAC map. Colleagues at Virginia Tech have begun a pilot project to begin high throughput sequencing of the turkey genome. These sequence data will be aligned with the chicken sequence and the comparative turkey-chicken BAC map described above.

NCBI's dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>) presently lists about 600,000 chicken ESTs and nearly 17,500 turkey ESTs. These have been critical in a variety of gene discovery efforts, especially in annotating the genome sequence and in array development.

PROGRESS TOWARD OBJECTIVE 3: Database and other map resources.

Sequence, Map and QTL: The sequence, along with a variety of options and tools, can be accessed at several different browsers: the UCSC Chicken Genome BrowserGateway, (<http://genome.ucsc.edu/cgi-bin/hgGateway?org=Chicken&db=0&hgsid=30948908>); the NCBI Chicken Genome Resources, (<http://www.ncbi.nlm.nih.gov/genome/guide/chicken/>); the EBI's Ensembl Chicken Genome Browser, (http://www.ensembl.org/Gallus_gallus/), and the Gallus Genome GBrowse website, <http://birdbase.net/cgi-bin/gbrowse/gallus/>. The latter site integrates many different types of genome data for several avian species. The ChickFPC browser at <http://www.bioinformatics.nl/gbrowse/cgi-bin/gbrowse/ChickFPC> allows for various searches of the BAC contig map. The SNP data generated by the Beijing Genomics Institute (described above) can be accessed on the sequence browsers, but more extensive descriptions are available at the BGI site at <http://chicken.genomics.org.cn/index.jsp>. 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 colleagues at Mississippi State. GO classifications are essential for functional genomic studies such as transcription profiling and proteomics. 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 is seeking funding to develop "BirdBase", an Aves-specific Model Organism Database (MOD) that can be used as a fundamental resource for all avian research communities. **ChickGBASE:** The latest version of ChickGBASE developed by the Roslin Institute is at <http://www.thearkdb.org/arkdb/do/getChromosomeDetails;jsessionid=B8A6A5EA698B84AF80EE99BE7530B04E?accession=ARKSPC00000004>. The chicken RH webserver, <http://chickrh.toulouse.inra.fr/>, is maintained by A. Vignal and colleagues at INRA (Toulouse, France). Most of the above resources are linked within the overall AvianNET, Avian Information Network at <http://www.chicken-genome.org/>, an effort led by Dave Burt (Roslin Institute). **US Poultry Genome Homepage:** 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 the latest EL maps and mapping data, descriptions of available resources, and access to a host of other information relating to both genetic and physical maps, including our newsletter archive. **Newsletter:** The Poultry Genome Newsletter is published quarterly and is distributed through our Homepage, electronically on the angenmap email discussion group and via direct email.

Reference Panel DNA: DNA from the East Lansing international reference population has been sent to many laboratories throughout the world.

Primer Kits: One or more kits have now been provided to over 130 different labs, worldwide. Our focus has shifted now to supporting SNP genotyping and most microsatellite panels have been discontinued.

Physical Mapping Resources: At least three public BAC libraries for chicken and two for turkey are now available. Filter arrays of BAC clones are being distributed for both the Texas A&M and CHORI-261 chicken BAC libraries. Filter arrays of the turkey CHORI-260 library are also available.

Chicken Microarrays: In this past FY, coordination funds again were 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. See <http://www.affymetrix.com/products/arrays/specific/chicken.affx>. Operon Biotechnologies Inc. distributes a *Gallus gallus* (chicken) Roslin/ARK CoRe Array V1.0 ready-to-spot oligonucleotide set that contains 20,673 long oligo probes for creating microarrays. See http://www.operon.com/arrays/oligosets_chicken.php. A chicken whole genome long oligo array is also available at the U. of Arizona, www.grl.steelecenter.arizona.edu and can also be purchased on a custom basis from Nimblegen, www.nimblegen.com.

Meetings: Over 2000 scientists attended the joint Plant and Animal Genome XVI meeting held last January, held jointly with the annual NAGRP meeting. Coordination funds helped support attendance at PAG-XVI and will do so again for the upcoming PAG-XVII in January, 2009. The Delivering Value from Avian Genomics meeting, followed by the Avian Gene Ontology (GO) and Microarray Data Modeling Workshop, were held May 19-22, at Mississippi State University.

PLANS FOR THE FUTURE. **Note:** objectives from the recently renewed NRSP-8 term (10/01/08-09/30/13) are incorporated below.

OBJECTIVE 1. Create shared genomic tools and reagents and sequence information to enhance the understanding and discovery of genetic mechanisms affecting traits of interest.

High throughput SNP genotyping arrays have become a primary sequence-based mapping tool that will receive continued coordinator support for development and application. Industries have already begun to apply high throughput mapping data in modifying their breeding strategies. The data may also be of use with regard to biosecurity and food safety issues. A major consortium effort designed to assess the feasibility of genome-wide marker-assisted selection (GMAS) in chickens has been assembled and funded by USDA and industry sources. The 60K SNP platform described above is a critical element in this effort. Initial results from the 3K SNP analysis (see Muir et al., referenced above) suggest that linkage disequilibrium in layer chickens may be adequately assessed at the 60K SNP level, but that in broilers may require even higher resolution maps. We have begun efforts that we hope will eventually enable 300K or even larger SNP arrays at reasonable cost.

While physical maps and associated resources will continue to be of use, especially for poultry other than chicken, there will be a shift towards high throughput sequence-based resources. We will support efforts to sequence the turkey genome and align this with that of the chicken. A separate international team will do likewise for the duck genome. Several whole genome re-sequencing projects employing various chicken lines are underway and more are anticipated.

Transcriptional profiling of additional chicken cells, lines and tissues at various developmental stages and under various environmental challenges will continue. These will employ both array and deep sequencing technologies. Profiling of chicken micro-RNA expression and the relationships of those profiles to phenotypes will be required. Proteomic profiling will need to be expanded. Both transcriptional and proteomic profiling will require improved gene annotation, a process that still requires much further effort for poultry.

OBJECTIVE 2. Facilitate the development and sharing of animal populations and the collection and analysis of new, unique and interesting phenotypes.

Mechanisms for sharing DNA panels, phenotypic and genotypic data need further development. An example is the 60K SNP, GMAS project described above that's already underway. Additional methods for supporting germplasm conservation and utilization are needed. Techniques for modifying the germline (transgenesis and RNA interference) in live birds are under-developed and under-employed in chicken research.

OBJECTIVE 3: Develop, integrate and implement bioinformatics resources to support the discovery of genetic mechanisms that underlie traits of interest.

The availability of the genome sequence has generated invaluable support at NCBI, Ensembl, and UCSC, and among the general bioinformatics community, for chicken genomics. With the rapid pace of sequencing technology development, genome sequencing is being "democratized". It will be a major challenge to keep publicly available sequence databases (and, especially, browsers) up to date with the wealth of genome re-sequencing data that is being generated at local core facilities. More rapid systems that can incorporate these data, re-assemble and re-annotate genomes are needed.

Efforts also need to be made to enhance bioinformatic support for transcriptional and proteomic profiling, in particular, improved gene annotation for the chicken (and, by extension, other poultry). Poultry-specific phenotypes of agricultural interest need to be added to ontologies.

As sequencing, profiling and other “omics” technologies advance, an increasing problem will be simply housing the enormous databases that are generated, in addition to integrating them, so that they will be of widespread use. This will be an even larger problem for human and mouse geneticists, so we hope to learn/benefit from their experience, as we have in the past. A variety of individual bioinformatic efforts for the chicken are underway as outlined above. The Poultry Genome Newsletter and homepage information will be continue to be distributed and enhanced.

(Prepared 11/24/08)