Functional Avian Endocrinology
A Dawson & P J Sharp (Eds)

Genomic tools for endocrine research

J Burnside, LA Cogburn, R Talbot[1] and JB Dodgson[2]

Department of Animal and Food Sciences, University of Delaware, Newark DE,19717
[1]Dept. of Genomics & Bioinformatics,#Roslin Institute, Roslin, Midlothian EH25 9PS, UK
[2]Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824

Introduction

The recent release of a draft chicken genome sequence (1), as well as the development of a large collection of expressed sequence tags (ESTs), genome-wide-BAC-based physical maps and supporting bioinformatics, has dramatically changed the landscape of avian biology. These new resources re-establish the chicken as a major experimental model organism. Furthermore, it seems certain that sequence and mechanistic similarities to other birds will extend the use of chicken genome information throughout avian biology. In particular, the chicken sequence, along with comparative avian genomics research, offers the long-range potential to connect complex avian phenotypes, such as growth and behaviour, to their fundamental molecular mechanistic causes. This chapter reviews the current status of chicken genomics and the resources newly available to avian researchers.

Chicken genome sequence and resources

Although individual chicken gene sequences have been emerging for over 25 years, sequencing the full genome began following the submission of a white paper to NIH by J.D. McPherson, J.B. Dodgson, R. Krumlauf and O. Pourquié (2). This highlighted the role of the chicken not only as a food animal but also as a major model organism in developmental biology, evolutionary biology, and disease research. The National Human Genome Research Institute (NHGRI) designated this a "high priority" project and sequencing began at the Washington University Genome Sequencing Center (WUGSC) (3). The strategy was to produce 6-fold whole-genome shotgun (WGS) coverage of the genome of a single female (the heterogametic sex) of the inbred UCD001 Red Jungle Fowl line. A single inbred bird was chosen to minimize DNA sequence polymorphisms and UCD001 was chosen based on its previous use in developing linkage maps and BAC libraries.

The Red Jungle Fowl is the nearest wild relative of the domestic chicken and thus represents a "wild type" genome to which subsequent domestic sequences can be compared. Parallel efforts had already begun to generate a full genome BAC contig physical map of the genome (1,4), and these were eventually employed in assembling the 6.6X draft sequence released on March 1, 2004 (5). The term "draft" sequence is used to indicate that a substantial number of gaps and occasional ordering errors remain in the sequence, as is to be expected in a 6.6X WGS assembly, despite the integration of BAC map, EST and cDNA data to improve accuracy. Retrieved sequences may contain short gaps (whose length will often be known based on "end pair" reads from a cloned fragment of known size), and approximately 5% of genes known from cDNA sequences are substantially or completely missing. Furthermore, a portion (12%) of the sequence has yet to be assigned to a specific chromosome location.

The genome sequence, along with a variety of viewing options and analytical tools, can be accessed at three different browsers: the UCSC Chicken Genome Browser Gateway (6); the NCBI Chicken Genome Resources (7); and the EBI's Ensembl Chicken Genome Browser (8). Each browser offers a variety of options and but uses the same genome assembly. The Blat program of the UCSC browser quickly maps a sequence to the genome (similar to BLAST at NCBI and Ensembl, along with SSAHA at Ensembl). The Ensembl and UCSC browsers are highly configurable for displaying (or not) information of interest. In particular, they offer powerful tools to download specific data sets for further analysis (e.g. Ensmart in Ensembl). A more recent addition to the NCBI browser is a presentation the chicken genome in MapViewer including the annotation of partial, complete and predicted genes with extensive documentation and tutorials. Several other options are available at each site, and all browsers have excellent help features to guide the user.

Gene hunters may also access the genome-wide, BAC contig-based physical map that facilitated the assembly of the genome sequence (4). Those BACs that have been aligned with the assembly via their end sequences are available as an optional display section in the UCSC and Ensembl browsers); however, a more complete version of the BAC physical map is available (9) as part of the Wageningen U. ChickAce, which was developed for storage of extensive mapping information, with particular emphasis on linkage, cytogenetic, radiation hybrid and BAC contig maps as well as phenotypic traits (QTL) (10). BAC clones, DNA pools and filters, as well as the BAC libraries themselves are also available (see Resources and/or The BAC Page section of recent Newsletters at http://poultry.mph.msu.edu/ for further details).

In parallel to the WUGSC sequence, the Beijing Genomics Institute in China generated a map of genetic variation by 1/4X shotgun sequence coverage of the genomes of three different domestic chickens (a broiler from the United Kingdom (Roslin Institute), a White Leghorn layer from Sweden (Uppsala University) and a Silkie from China). Nearly three million genetic variation sites, mostly single nucleotide polymorphisms (SNPs) have been identified (11). These data have been integrated into the various genome browsers, and also can be searched at the Chicken Variation Database (12). In addition to being of fundamental biological interest with regard to the domestication and selection of the modern chicken, most of these SNPs are prevalent within domestic lines; therefore, they can be used in high resolution linkage mapping of any phenotypic trait of interest.

A portal to information on the chicken genome and chicken biology was created at the 1st international chicken genome workshop held in Hinxton, UK in 2003 and designated AvianNet (13). This serves as a gateway to a network of WWW sites covering the chicken genome, chicken developmental biology, chicken genetics, biodiversity, chicken immunology, and other species and tools with a shared biological interest. In addition, a mailing list (14) has been established to facilitate rapid sharing of information.

Chicken cDNAs and ESTs

Sequenced cDNAs and ESTs constitute a critical resource for chicken biologists, and the number of chicken ESTs has grown dramatically in the last few years. The BBSRC large scale EST sequencing project produced approximately 340, 000 ESTs from normalized cDNA libraries prepared from 12 adult and 9 embryonic tissues (15). The ESTs can be searched using keywords or the Blast engine at the BBSRC ChickEST website (16). Since these sequences have been isolated from many animals and various strains, mining of the data has identified a large number of coding SNPs which are available on the website. The site also includes an RNAi gene prediction tool and a search engine for non-coding RNAs. The bursal EST project has generated a large (46,595) EST collection from 2-week-old chicken bursa (17). One of the libraries in this project was designed to capture the 5'end and therefore the collection contains a large subset of full-length clones, to aid in obtaining full insert sequence. The EST project at the University of Delaware (18) has focused on generating ESTs (49,000) from tissues of economic importance: liver, muscle, neuroendocrine, reproductive and lymphoid tissues. Clones from the BBSRC, UD and bursal EST projects are available through the web sites or via ARK Genomics (24). The Uppsala project produced 21,869 ESTs from brain and testes. A multi-

tissue (brain, ovary, embryos, hypothalamus, skin, and pituitary gland) library approach was taken by the group at INRA, where 12,000 cDNA clones were sequenced from both the 5' and 3' ends (23,273 ESTs total). These and several other projects have made important contributions to the now (November 2004) 527,686 *G. gallus* ESTs in GenBank (Table 1). In other species, full-length cDNA sequences have proven to be essential in accurately annotating genes and their alternative mRNA processing mechanisms through which a single gene may give rise to multiple protein products. Further analysis of all the EST sequences identified a set of unique cDNA clones coding for full length or nearly full length cDNAs. These have been fully sequenced to provide a set of 19,626 fully annotated clones which are searchable through the BBSRC ChickEST website.

The Institute for Genomic Research (TIGR) maintains a gene index of all chicken cDNAs and ESTs in GenBank. The TIGR *G. gallus* Gene Index (GgGI) integrates public data from all *G. gallus* gene sequencing projects with the goal of representing a non-redundant compilation of all chicken genes, including data on their expression patterns, cellular roles, functions, and evolutionary relationships (19). This index now contains 42,988 tentative consensus sequences assembled from all available data. In addition, 72,941 singleton ESTs are indexed. Theses numbers are in excess of the number of genes predicted in the chicken genome (approximately 20-25,000 protein coding genes) (20) and include alternatively-spliced products and non-coding RNAs, as well as lower quality sequences that could not be assembled with the stringent criteria of the TIGR CAP3 assembly program. Along similar lines, NCBI has created a chicken Unigene database that contains 21,000 entries (21). Thus, we are well on our way to having the annotated sequence for every chicken gene.

Gene expression tools

ESTs are invaluable for gene discovery, genome sequence annotation, and identification of orthologous relationships among evolutionarily distinct organisms. In addition, collections of ESTs can provide the basis for the development of microarrays. A number of centers have tackled the production of chicken microarrays (Table 2). A 13,000-element chicken cDNA array is available from the Fred Hutchinson Cancer Research Center (FHCRC) and ARK Genomics, (Roslin Institute, UK) at a cost of $150 per slide. Clones on this array were selected from a total of 363,838 chicken ESTs representing 24 different adult or embryonic tissues. A subset of 11,447 non-redundant ESTs were selected and added to an existing collection of clones (4,162) from immune tissues and a chicken bursal cell line (DT40). The array provides broad coverage of

Table 1 Chicken ESTs in GenBank

Institution	Tissues	# GenBank Entries
University of Delaware	Fat, Liver, Muscle, Pituitary, Oviduct, Ovary, Testis, Lymphoid tissues	49,393
USDA/ARS/BARC	Small intestine from birds infected with coccidia	14,376
Fred Hutchinson Cancer Research Center (FHCRC)	DT40 cells	1,116
BBSRC, UK	Embryo (9 stage or tissue-specific libraries) Adult (12 tissue-specific libraries)	330,096
Roslin Institute, UK	Brain, Embryo, Liver	5,194
Heinrich-Pette-Institute, Germany (Japan)	Bursa of Fabricius	46,595
INRA/Agenae, France	Multi-tissues: brain, ovary, embryos, hypothalamus, skin, pituitary gland	23,273
Uppsala University, Sweden	Brain, Testis (Red Jungle Fowl, Leghorn)	21, 869
University of Sao Paulo ESALQ, Brazil	Embryo (Limb buds, breast muscle)	14,395
All Others		21,379
	Total EST Sequences 10/2004	527,686

mRNAs expressed in many tissues; in addition, clones with expression unique to various tissues can be detected. L. Cogburn and collaborators have produced an 8K metabolic/somatic system array, a 7K neuroendocrine/ reproductive system array, and a 14K integrated systems array (22, 23) (and chapter this volume). These arrays are being used for transcriptional profiling across multiple tissues of divergently selected broiler chicken lines. At ARK genomics (24) four chicken

microarrays have been prepared: three targeting specific tissues, neuroendocrine, immune and bone/shell gland, as well as the general chicken 13K array. A range of studies including one on photostimulation in both chicken and quail highlight the ability to use these chicken microarrays in other avian species. Recently, Affymetrix has designed a chicken Gene Chip that contains probe sets representing about 33,000 transcripts. Since the Affymetrix Gene Chips employ short oligonucleotide probes, mostly from 3' untranslated regions of mRNAs, they are less likely to be useful in other avian species. We are rapidly approaching the availability of genome-wide arrays for the chicken, at least some of which should be applicable to other gallinaceous species and probably even more distantly related birds.

Another valuable resource for transcriptional profiling in the chick embryo is the GEISHA gene expression database developed by P. Antin (25). This site collects and displays *in situ* hybridization data at different developmental stages for a large collection of genes. This project is complemented by the Chick-Atlas project where high quality images will serve as references for in-situ hybridisation results (26).

Concluding remarks

Sequencing the chicken genome is not only a significant advance for avian biology but also a milestone in the Human Genome Project. Birds are among the more diverse orders, but their genomes demonstrate a relatively high degree of evolutionary conservation at the karyotype level. From the perspective of vertebrate evolution, the chicken sequence provides a valuable tool in comparative genomics. For example, the chicken genome presently is the nearest "outgroup" with which to analyze mammalian genome evolution. This aids in re-construction of the likely genome arrangement of the common mammalian ancestor and even the common vertebrate ancestor (1). Furthermore, in comparisons made between mammalian genomes (e.g., human vs. mouse), it is often difficult to differentiate evolutionary (i.e., selected for function) conservation from background similarity (especially in non-coding regions). The greater evolutionary distance between birds and mammals reduces the un-selected similarity to near-zero, so any significant conservation observed is almost certainly of functional significance. Third, the chicken genome reveals genes that are unique to, or much more widely used in, mammals, as well as genes that may have been lost in mammals but retained in the chicken and gene families that specifically expanded in avian evolution (1). Continued analysis of the chicken genome will improve our general understanding of how genes evolve, how they function, and the role of conserved non-coding DNA, as well as identifying those genes that give birds their unique character.

Table 2 Chicken Microarray Resources

Institution	Tissues Represented	# Elements
University of Delaware	Lymphoid tissue	5K
University of Delaware	Integrated Systems (Metabolic/Somatic/ Neuroendocrine)	14K
USDA/ARS/BARC	Small Intestine	10K
Fred Hutchinson Cancer Research Center (FHCRC)	Multi-tissue	13K
ARK Genomics, Roslin Institute	Neuroendocrine, Immune, Bone/Shell Gland, Multi-tissue	5K and13K
Uppsala University, Sweden	Brain, Testis	14K
University of Sao Paulo, ESALQ, Brazil	Embryo, muscle, pituitary	5K
China Agricultural University	Liver, muscle, brain, fat, ovary, etc	10K
Affymetrix	Genome-wide Gene Chip®	33K

Finally, the chicken sequence serves as the "model genome" for the ~9600 avian species. Since gene order has been highly conserved during avian evolution, one can expect that genes identified either by molecular analysis or by phenotype in other birds will be located in the analogous position along the sequence to that of their chicken homologs. The location, sequence and function of a chicken homolog provide an entry point for detailed analysis of a trait and/or gene(s) in a bird of interest. Birds, as a group, provide a wealth of species of interest for behavioural, reproductive, nutritional, and physiological traits, and now the chicken sequence now provides a scaffold on which much of this avian-specific biology can be functionally annotated.

References

1 Wallis JW, Aerts J, Groenen M, Crooijmans R, Layman D, Graves T, Scheer D, Kremitzki C, Higgenbotham J, Gaige T, Mead K, Walker J, Albracht D, Divito J, Yang S-P, Leong S, Chinwalla A, Hillier L, Sekhon M, Dodgson JB, Romanov MN, Cheng H, de Jong PJ, Zhang H, McPherson

JD, Krzywinski M, Schein J, Mardis E, Wilson R & Warren WC (2004). A physical map of the chicken genome. *Nature* **432** 761-764
2 www.genome.gov/Page/Research/Sequencing/SeqProposals/Chicken_Genom.pdf
3 http://www.genome.wustl.edu/projects/chicken/index.php
4 Ren C, Lee M-K, Yan B, Ding K, Cox B, Romanov MN, Price JA, Dodgson JB & Zhang H-B (2003). A bac-based physical map of the chicken genome. *Genome Research* **13** 2754-2758
5 http://www.genome.gov/11510730
6 http://genome.ucsc.edu/cgi-bin/hgGateway?org=Chicken&db=0&hgsid=30948908
7 http://www.ncbi.nlm.nih.gov/genome/guide/chicken.
8 http://www.ensembl.org/gallus_gallus.
9 http://hbz.Tamu.Edu/.
10 https://acedb.asg.wur.nl/.
11 International Chicken Polymorphism Map Consortium (2004). A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms. *Nature,* **432** 717-722.
12 http://chicken.genomics.org.cn/chicken/jsp/about.jsp
13 http://www.chicken-genome.org/.
14 chicken-genome@lists.Bbsrc.Ac.Uk.
15 Boardman E, Sanz-Ezquerro J, Overton IM, Burt DW, Bosch E, Fong WT, Tickle C, Brown WRA, Wilson SA & Hubbard SJ (2002). A comprehensive collection of chicken cDNAs. *Current Biology* **12** 1965-1969
16 http://www.chick.umist.ac.uk/
17 http://swallow.gsf.de/dt40.html
18 http://www.chickest.udel.edu
19 http://www.tigr.org/tdb/tgi/gggi
20 International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432** 695-716.
21 ftp://ftp.ncbi.nlm.Nnh.gov/repository/unigene/.
22 Cogburn L, Wang X, Carre W, Rejto L, Porter T, Aggrey S & Simon J (2003). Systems-wide chicken DNA microarrays, gene expression profiling, and discovery of functional genes. *Poultry Science* **82** 939-951
23 Cogburn L, Wang X, Carre W, Rejto L, Aggrey SE, Duclos MJ, Simon J and Porter TE (2004). Functional genomics in chickens: Development of integrated-systems microarrays for transcriptional profiling and discovery of regulatory pathways. *Comparative and Functional Genomics* **2** 253-261
24 www.arkgenomics.org
25 http://geisha.biosci.arizona.edu/
26 http://genex.hgu.mrc.ac.uk/Collaborations/ChickAtlas/intro.html, Duncan Davidsson)