

PROPOSAL TO SEQUENCE THE GENOME OF THE CHICKEN**John D. McPherson, Washington University,
Genome Sequencing Center****Jerry Dodgson, Michigan State University****Robb Krumlauf, Stowers Institute****Olivier Pourquié, Stowers Institute****I. Overview and Rationale of the Sequencing Strategy.**

The chicken genome has a haploid content of 1.2×10^9 base pairs (bp) of DNA; approximately 40% that of either mouse or human. The strategy that will be employed to sequence this genome is to assemble 6-fold whole-genome shotgun coverage of the genome and to order and orient the resulting sequence scaffolds by alignment to end-sequences of BACs in a comprehensive contig map. This is in contrast to the BAC-by-BAC approach used by the Human Genome Project laboratories to sequence the human genome but is similar to the first phase of the strategy currently being used to sequence the mouse genome. Initial analysis of an assembly of available 5-fold mouse whole-genome shotgun sequence indicates that approximately 98% coverage of the genome has been achieved in reads and about 90% is covered in assembled bases. These scaffolds are aligned to a BAC contig map using corresponding BAC end-sequences. The sequence continuity and quality of the resulting mouse sequence supports the proposed strategy and level of whole-genome shotgun coverage for the chicken genome. The resulting sequence will provide an invaluable tool for the chicken, and other research communities. The generation of finished sequence for selected regions will be readily achievable utilizing the combined BAC map and sequence. In addition, they will provide a suitable basis for generating a complete finished sequence of this genome in the future.

II. Biological Rationales for the Utility of the Chicken Genome Sequence.

Chicken as a model for studies of health/disease and biology. The chicken is the premier non-mammalian vertebrate model organism. It is one of the primary models for embryology and development as its embryonic development occurs *in ovo* rather than *in utero*. The chicken is a major model organism for the study of viruses and cancer. The first tumor virus (Rous sarcoma virus) and oncogene (*src*) were identified in the chicken, and the avian leukosis viruses (ALV) remain among the most intensively studied retroviruses. The chicken immune system provided the first indications of the distinctions between T- and B-cells, with the B-cell nomenclature based on the avian bursa of Fabricius (e.g., Cooper et al., J. Exp. Med. 123:75-102, 1966). The chicken has also provided an important model system for studies of gene regulation. For example, many of the pioneering studies of steroid hormone regulation involved the chicken oviduct (e.g., ovalbumin gene regulation). Because chicken red blood cells retain their nucleus, they have been the classical model system in which chromatin structure has been studied. As one of many possible examples, the CCCTC-binding factor (CTCF) that now appears to play critical roles in imprinting and X inactivation was first discovered

via its activity at the chicken *c-myc* gene. (Ironically, chickens lack both X chromosomes and imprinting.) The numbers of hits from PubMed searches (1995 to present) with the keyword combinations below confirm the continued relevance of the chicken to major aspects of human disease and biology, and the fact that the user community for the chicken exceeds that for other non-mammalian, vertebrate species, including zebrafish and fugu, whose genome sequences are already nearing completion.

Table 1. PubMed entries, 1995 to present (01/24/02)

Keyword	total hits	and development	and virus	and cancer	and neuron	and immunology	and pathology	and disease
chicken	17,335	3,488	2,087	854	1,053	2,536	1,067	1,673
Xenopus	12,087	2,320	298	290	1,067	414	147	264
zebrafish	2,190	1,046	46	41	337	57	57	61
fugu	150	ND	ND	ND	ND	ND	ND	ND

ND: not determined

Some novel aspects of avian biology relevant to human biology. Among the novel aspects of avian genetics is the use of a Z/W sex determination system. In contrast to mammals, the female is the heterogametic sex (ZW), whereas the male is homogametic (ZZ). Although much remains to be learned about the genetic determination of sex in birds, the evidence suggests that sex chromosomes evolved separately in mammals and avians, yet at least some of the sex determining genes are conserved between chicken and human (Nanda et al., *Cytogenet. Cell Genet.* 89:67-78, 2000). Thus, studies in the chicken can provide an illuminating contrast to sex determination in mammals. Similarly, immunoglobulin diversity (Reynaud et al., *Adv Immunol.* 57:353-378, 1994) in the chicken is mainly the result of somatic mutation (probably via gene conversion using sequences from flanking pseudogenes) rather than the VDJ recombination mechanisms that predominate in man and mouse. A third example is the much greater variation of chromosomal size in birds vs. mammals (birds being more similar to reptiles in this regard). Chicken chromosome 1 is nearly 600 cM (150-200 Mb), whereas 30 of the 38 chicken autosomes, termed "microchromosomes", range between 5 and 20 Mb. There's evidence that gene density and genetic/physical map ratios vary considerably between the larger chromosomes and the microchromosomes. The mechanism by which microchromosomes are stably maintained is of both general biological relevance and of applied interest in the design of artificial chromosomes and the study of non-disjunction.

Informing the human sequence. The rapid progress of the mouse genome sequence has provided both an informative and intriguing comparison to the human sequence. About 3% of the two sequences show significant similarity, only about half of which can be accounted for by exonic regions (R. Waterston, personal communication). The remaining half likely includes a range of sequences from those whose functions are of critical evolutionary importance to others whose similarity reflects the limited divergence time rather than a conserved function. Comparison to the fugu sequence is useful in determining functional exons, but shows little additional homology. The chicken is well positioned from an evolutionary standpoint to provide an intermediate perspective

between those provided by mouse and fugu. This viewpoint has been clearly documented in the comparative sequence analysis of the stem cell leukemia, *SCL*, gene region by Göttgens et al. (Nature Biotech. 18:181-186, 2000), the major histocompatibility complex (Kaufman et al., Nature 401:923-925, 1999), and the α - and β -globin gene regions (e.g., Flint et al., Hum. Mol. Genet. 10:371-382, 2001). Further evidence is beginning to emerge from comparative sequencing of five additional multi-Mb target regions (E. Green, personal communication). In all cases the corresponding chicken sequence is significantly more compact with smaller introns and intergenic regions, consistent with its nearly 3-fold smaller genome. In addition to identification of new regions of functional homology, the chicken sequence also provides an evolutionary intermediate that will help connect and date the sequence divergence of exons between the human sequence and those of zebrafish, fugu, and *Ciona*.

Evolution of gene arrangement. In addition to pure sequence comparisons, the chicken genome provides an excellent model with which to understand the evolution of gene order and arrangement. Studies from several labs (Burt et al., Nature 402:411-413, 1999; Groenen et al., Genome Research 10:137-147, 2000; Waddington et al., Genetics 154: 323-332, 2000; Suchyta et al., Animal Genet. 32, 121-18, 2001) have demonstrated a remarkable level of conservation in gene order between mammalian and chicken genomes. Indeed, although the evidence remains far from complete, the number of genome rearrangements separating the chicken and human genomes (on a gross level) appear to be similar to, if not fewer than, those separating the mouse and human genomes. Overall gene order is often conserved over segments of many tens of cM between chicken and human (average conserved segment length is about 30-40 cM) with the evidence clearly suggesting a higher rate of inversions vs. translocations having occurred in the 600 M years separating the two genomes (300 My from each present day genome back to the last common ancestor). This level of conservation is especially remarkable, given the fact that the chicken genome is only about 1/3 the size of mammalian genomes and that sequence conservation outside exons and regulatory regions is minimal. It has been proposed that the chromosome rearrangement rate (at least within the germline) is substantially slower in birds (and reptiles) than in mammals (Burt et al., op. cit.), but the reasons for this remain unknown. A full chicken genome sequence would provide a much more detailed and accurate picture of the comparative avian/mammalian map and a fascinating evolutionary counterpoint to the human and mouse sequences.

Chicken, the model bird. The chicken is the only avian for which significant genomics resources now exist (see below). Birds are one of the more diverse and widespread vertebrate orders, and they have long been a source of fascination to scientists and lay persons, alike. Various avian species provide models of evolution, behavior and ecology (e.g., sexual selection and courtship, vocal communication, evolution of color patterns and other morphological features, rearing of young, migration). As noted above, avian genomes appear to show relatively high levels of conservation, at least of karyotype, and

the genome sequence and related information and infrastructure that is available for the chicken provide added benefits for the genetic analysis of all wild and domestic birds.

Chicken, a food animal. Although the obvious importance of chicken meat and eggs as food is not of direct relevance to the mission of NIH, clearly, both nutrition and food safety are critical elements of human health, both in the U.S. and, even more so, in the developing world. Chickens provide one of the most important and rapidly growing sources of meat protein in the world and in the U.S. (~41% of meat produced). U.S. broiler production has grown over 65% in the last decade to about 30.4 billion pounds in 2000. Per capita broiler consumption in the U.S. has grown nearly 30% over the same period. Perhaps more important, the increasing demand for high quality protein in the developing world is expected to be one of the most important trends in the future of agriculture (Rosegrant et al., 2020 Global Food Outlook, a report of the International Food Policy Research Institute, 2001), and chicken meat and eggs are likely to be the major contributor to meeting this demand. (Establishing a functioning poultry production operation is usually less cost-intensive than for other meat animals, and cattle and/or swine consumption is not compatible with religious practices in some developing nations.)

Complementary studies in agriculture and human health. Beyond the importance of a safe and nutritious food supply to human health, the enormous world-wide interest in raising poultry for food provides a collateral source of scientific data that inform our understanding of human health and biology in general. As a simple example, more doses of Marek's Disease vaccine are made and administered than any other vaccine, human or otherwise. (Marek's Disease Virus is an oncogenic herpesvirus of fowl, the only tumor virus in any species for which an effective vaccine presently exists.) The enormous commercial populations of chickens (far exceeding those of any other captive model species) means that large scale breeding studies can be done in the chicken. In many cases the traits that are of interest (nutrition, growth, disease resistance, reproductive success) to the poultry industry who possess the largest flocks are traits that are of similar importance to human health, so studies of chicken genetics and human medicine are often complementary.

The chicken as a model for QTL analysis. Interest in human genetic disease is rapidly shifting towards so-called "complex" or multi-gene traits. Quantitative genetics, the scientific treatment of multi-genic traits, was pioneered, and continues to be heavily influenced, by agricultural geneticists, who generally employ the term quantitative trait loci, QTL, in describing such inherited characters. Although QTL analysis is also done in model species such as mouse and *Drosophila*, agricultural animals continue to be important experimental systems for quantitative genetics. Among domestic animals, the chicken is ideal for genetic mapping and QTL analysis. It reproduces rapidly, enabling several generations of large families to be generated in a reasonable timeframe. The chicken is unique among major agricultural species in that a number of highly inbred lines are available. Inbred lines, of course, have been critical to the utility of the mouse

as a model species. Like the mouse, in addition to experimental lines, there are a variety of unusual chickens that have been developed by chicken "fanciers", and there are a limited number of mutant lines that have developmental defects of various types. The commercial and experimental lines generally fall into two very different classes: layers, selected for high egg production, and broilers, selected for rapid growth and meat production. (For a full discussion of poultry diversity, see Piseni et al., *Avian and Poultry Biology Reviews* 12:1-102, 2001.) Other birds, most notably quail, a close relative of chicken, have also provided useful model species for QTL genetic analysis. Since most such studies to date have been funded by agricultural concerns, the traits of interest have been mainly related to growth, disease resistance, and egg production, although there is an increasing interest in studying behavioral traits in chickens and other avians, as these traits relate to both productivity and animal welfare.

III. Strategic Issues in Acquiring New Sequence Data.

The community. The overall size of the community of chicken biologists is best addressed by the PubMed citation data cited above. In fact, scientists interested in the chicken have really been two nearly separate communities - those interested in the chicken as a model biological system and those interested in agricultural productivity. However, the advent of genomics has begun to bridge these two communities. QTL need no longer be hypothetical/anonymous genes or alleles, and the prospect now exists to actually identify their molecular foundations. Thus poultry scientists (and companies) are increasingly interested in molecular biology. As noted above, geneticists are becoming increasingly interested in complex traits. This enhances their interest in the techniques and theories that previously have been the focus of the agriculture community. A complete genome sequence of the chicken would provide a critical bridge to bring the two communities of scientists (and their respective sponsors) closer together to focus on common goals. Even more important, a chicken sequence would be the avenue by which chicken biologists of every sort can access, utilize and apply the results of human and mouse genomics to their own research, thus stimulating further growth in the avian biology community.

To give an indication of the enthusiasm and support of the chicken research community for obtaining the sequence of this genome, an email request for letters of support was sent out shortly prior to the submission of this white paper. The response was immediate with 52 emails from 8 countries received all strongly supporting this effort and indicating how the sequence would enhance their research. A few excerpts from these emails follow.

“Chickens have a remarkable capacity for hair cell regeneration that results in spontaneous recovery from forms of deafness and balance disorders that are permanent when they occur in humans and other mammals. ... If the chick genome were sequenced that could aid our work [in this area] profoundly and might reveal answers to questions that could lead to significant improvements in quality of life for many individuals who are hard of hearing.”

“The chick has been the mainstay of developmental biologists and remains unique in that mid embryonic development cannot be studied easily in either *Xenopus* or zebrafish. The recent development of chick ES cells and the ability to make transgenics makes this a truly versatile model which would be bolstered by having the entire genome sequenced.”

“I do not think that there is another land vertebrate taxon that would be more informative to sequence in terms of the evolution of land vertebrates.”

“The chicken is a premier experimental research model with enormous potential for the future, but woefully information-poor in terms of genomic information. Knowledge of the genome sequence will be of enormous benefit to “chicken” researchers and to those who use the chicken as a model organism for basic biology (developmental biology) and biomedicine (vaccine construction and production, human therapeutics).”

“We have developed a new model for studying the effects of exposure to nicotine during embryogenesis that takes advantage of the facility with which manipulations such as drug delivery can be done in the developing chick. This model provides evidence of substantial functional alterations in central synapses in embryos treated chronically with nicotine. This is a model animal system for trying to understand at the single cell level the consequences of maternal cigarette smoking on human fetal development.”

“[The genome sequence] should enable us to identify potential candidate genes for the QTL in these regions. Towards this end it is essential to know all the genes within the identified QTL regions, which therefore is another reason why it would be essential for our research to have the availability of the sequence of the complete chicken genome. The traits for which we have identified QTL might also have relevance for the identification of genes involved in complex traits in human. The traits we are working on are as follows:

- (1) Pulmonary Hypertension Syndrome (5 QTL regions identified)
- (2) Fat deposition (6 QTL regions identified)
- (3) Immune response (Analysis in progress)
- (4) Behavioral traits (Analysis in progress)”

Status of chicken genomics. Genetic linkage map. Numerous labs worldwide have cooperated in mapping DNA-based polymorphic markers by genotyping samples from the same two international reference crosses, the Compton population (Bumstead and Palyga, *Genomics* 13, 690-697, 1992), and the East Lansing (EL) population (Crittenden et al., *Poultry Science* 72, 334-348, 1993). Subsequently, the map has been enhanced by genotyping of a third cross, the Wageningen population, by Martien Groenen and colleagues (Groenen et al., *Genomics* 49, 265-274, 1998). A consensus map based on all three populations has been published (Groenen et al., *Genome Res.* 10:137-147, 2000). Recent updates (Schmid et al., *Cytogenet. & Cell Genet.* 90:169-218, 2000) bring the number of markers on the consensus map to nearly 2000, placed into 50 linkage groups, covering around 4000 cM. The EL map alone has expanded to nearly 1200 markers on 40 linkage groups. (The excess of linkage groups in the consensus map over the number of chromosomes is due to several very small linkage groups, some of which have only been observed in one of the three base populations. Most of these are expected to be incorporated into larger linkage groups eventually.) Development and use of the map has been enhanced by the distribution of several panels of free microsatellite primer pairs (see <http://poultry.mph.msu.edu>) by the USDA National Animal Genome Research Program (NAGRP) Poultry Genome Coordination effort. As noted above, evidence continues to grow that gene order is conserved between the human and chicken genomes to a remarkable extent. Several studies have appeared (cited above) that estimate

between 100 and 200 breakpoints in gene order separating the human and chicken genomes, with blocks of conserved synteny (not necessarily gene order) exceeding 100 cM.

Physical map. The chicken genome has a haploid content of 1.2×10^9 base pairs (bp) of DNA organized on 38 autosomes plus the Z and W sex chromosomes. Most autosomes are small "microchromosomes" which cannot be distinguished by size alone. However, most microchromosomes are now marked by FISH hybridization tags and have been associated with the appropriate linkage group (Schmid et al., op. cit.). Several chicken large insert bacterial artificial chromosome (BAC) libraries have been generated. Zhang (Texas A&M), Dodgson and colleagues now have generated a library of over 115,000 BACs (Table 2), with three different partial digest restriction fragment inserts. The source DNA for these libraries is a single female from an inbred ("wild type") line based on the Red Jungle Fowl progenitor of domestic chickens. This line was one of the two polymorphic inbred lines used to generate the EL backcross mapping population described above. It contains several dominant markers that are part of the consensus linkage map and serves as a good source of SNP variation in comparison to commercial or experimental lines of both broilers and layers. Comparative sequence analysis of a variety of markers shows that it differs in sequence by about 1% from, for example, the UCD 003 inbred White Leghorn that is the other parent of the EL backcross map population. The Zhang laboratory has fingerprinted ~65,000 (>20,000 from each sub-library) of the BAC clones, using the dual digest, end-label, sequencing gel approach. Fingerprint data are being edited and entered for eventual contig analysis by FPC. The Washington University Genome Center has been provided a copy of the full BAC library and is in the process of fingerprinting it using the complementary technique of agarose gel-based fingerprinting. The goal of these projects is to generate a contig map with >95% coverage. The Mapping Group at Washington University has great experience in assembling BAC contig maps of complex genomes. This includes a recent BAC map of the mouse that currently has just 330 contigs encompassing this 3.0 Gb genome.

Table 2. The chicken (Jungle Fowl UCD 001) BAC libraries.

Libraries	Mean insert size (kb)	No. of clones	Genome coverage	Vector
<i>Bam</i> HI inserts	115	38,400	5.2 x	pBeloBAC11
<i>Eco</i> RI inserts	130	38,400	5.2 x	pECBAC1
<i>Hind</i> III inserts	138	38,400	5.9 x	pECBAC1
Total:	128	115,200	16.4 x	

Another chicken BAC library, also made at Texas A&M by Crooijmans, Groenen, and their colleagues, is also publicly available (Crooijmans et al., Mamm. Genome 11:360-363, 2000). The source DNA was an outbred White Leghorn female. This library consists of about 50,000 clones with an average insert size of 134 kb. The Groenen lab is in process of fingerprinting their full library using agarose gel similar to the Washington University effort. Different run parameters and marker usage present a challenge to directly combining fingerprint data; however, Washington University has previous experience with such endeavors and it is expected that an integrated Jungle

Fowl/Leghorn BAC map will be generated through a collaborative effort. In addition, the Groenen group has associated specific known markers with their BACs (400 at present) which will assist in anchoring these maps to the genetic map.

The extraordinarily small number of contigs comprising the mouse genome BAC map is likely due at least in part to the larger average insert size of this library (200 kb). The chicken genome mapping project will benefit from one additional comparable library that can be readily integrated into the mapping effort. These combined mapping efforts will undoubtedly generate an overall BAC contig map for the chicken genome suitable for anchoring assembled whole-genome shotgun sequence.

Integration of physical and linkage maps. Two main mechanisms are being used to link the physical and BAC contig maps at MSU and at the USDA Avian Disease and Oncology Lab (H. Cheng). One of these uses overgo probe hybridizations (3-dimensional arrays of 216 probes, hybridized in pools of 36). This has been completed for markers on chromosomes 1 and 2 and now is being extended across the genome. The second will involve semi-automated PCR screening. This is necessary for some of our relatively small microsatellites with insufficient single-copy sequence available for high quality overgo design. Plate, row and column BAC pool DNA preps have been obtained for a portion of the BAC library from Research Genetics. PCR pool-based screening is now beginning. In the future, we will expand these approaches to include additional ESTs that will also be located on the linkage map to enhance both map integration and the comparative human-chicken map. In addition, the USDA-NAGRP Poultry Genome Coordination effort provides free arrayed BAC filter sets to various users on the condition that they include their BAC identification results in a public database now being assembled (<http://poultry.mph.msu.edu>).

ESTs. Although ESTs cannot be used to answer questions of gene organization, chromosome organization and QTL analysis, they provide a useful intermediate resource and will be critical in interpretation of the genomic sequence. J. Burnside, L. Cogburn, and R. Morgan at the U. of Delaware and P. Neiman (Fred Hutchison Cancer Ctr.) and several European groups are generating EST collections (Tirunagaru et al., *Genomics* 66:144-151, 2000) and beginning to apply them to microarray analysis. A consortium of European Union scientists (especially, W. Brown of the U. of Manchester Inst. of Science and Technology, along with colleagues at the U. of Nottingham, U. of Dundee, and the Roslin Institute) and Incyte Genomics announced the sequencing of over 300,000 chicken ESTs from a wide variety of tissues and developmental stages in December. Data are posted at www.chick.umist.ac.uk. Intensive cluster analysis has not yet been done for all the available chicken ESTs, but it's estimated that the total collection includes about 40,000 unique cDNAs. This should be a substantial fraction of all chicken genes, but surely misses many genes (a limited set of cDNA libraries have been included to date) and, especially, splice variants. These EST collections (and future expansions of them) will be of enormous utility to chicken biologists, and will be useful in the analysis of the genome sequence.

Radiation hybrids. Useful radiation hybrid (RH) panels have been more difficult to construct for the chicken, probably due to the evolutionary distance involved (earlier attempts were compromised by low retention frequency). However, A. Vignal and colleagues at INRA (Castanet-Tolosan, France) have recently generated usable chicken RH panels (personal communication) and are in the process of making these generally available. These panels will provide a useful complement to the BAC contig mapping efforts noted above.

Chicken cell lines. While chicken cells have gained notoriety for being difficult to immortalize other than with viruses, in fact, a variety of chicken cell lines are now available. There is space only to mention a few of the most important. DF1 (Himley et al., *Virology* 248:295-304, 1998) is an immortalized fibroblast line that is widely used as both a viral host and a model for cell cycle/immortalization studies. DT40 (Baba et al., *Virology* 144:139-151, 1985) is a virus-immortalized lymphoid cell line notable for its very high frequency of homologous recombination after introduction of exogenous DNA (Buerstedde and Takeda, *Cell* 67:179-188, 1991.). DT40 has been widely used for gene targeting and somatic cell gene knock-out studies and has the potential to be used to make detailed *in vivo* manipulations of isolated genes for mutagenesis and/or artificial chromosome purposes. Chicken ES cell lines have also been developed (Pain et al., *Development* 122:2339-2348, 1996), as have related cell lines that show promise in studies of chimeras and transgenic birds (see below).

Transgenic chickens. Although it has been possible to make transgenic chickens using viral vectors for quite some time (Salter et al., *Poult Sci.* 65:1445-1458, 1986; Bosselman et al., *Science* 243:533-535, 1989), widespread application of transgenesis in chickens has yet to occur. This is primarily due to the rather low efficiency of germline integration of the transgene, and the resultant high cost involved in generating and maintaining a genetically homogeneous transgenic chicken line. Generation of chimeric chickens via transplant of embryonic blastodermal cells (EBC) is feasible (Petitte et al., *Development* 108:185-189, 1990), and methods of transfecting the EBC and selecting transgenic cells prior to their use in making chimeras are under active study (J. Petitte, personal communication). Similarly, generation of transgenics using transfected, immortalized primordial germ cells and using new approaches to sperm-mediated transgenesis (Kroll and Amaya, *Development* 122:3173-3183, 1996) are under active study by several university and commercial labs. Several companies in Europe and in the US are actively developing systems for the production of recombinant proteins for human medicine in chicken egg albumin. (Since there are relatively few contaminating proteins in egg albumin and it obviously can be generated in large quantities at very low cost, the chicken egg offers a potentially cost-effective route to produce pharmaceutical proteins with appropriate post-translational modifications.) While promising results have been reported, it can't yet be said that the era of routine generation of transgenic chickens is yet upon us. (The use of the chicken *tva* receptor for ALV viral vectors as a mechanism

for making cell-targeted transgenics in mouse has also been described, Fisher et al., *Oncogene* 18:5253-5260, 1999.)

Bioinformatics. The latest version of the chicken genome database, ChickGBASE, is part of the comparative mapping database for farm and other animals, Arkdb. Arkdb was primarily developed by Andy Law, Dave Burt, and Alan Archibald at the Roslin Institute. Arkdb provides a public repository for genome mapping data including details of loci and markers and is available at <http://www.thearkdb.org>. A mirror site for the poultry database is at Iowa State, <http://www.genome.iastate.edu/>. An additional homepage for the Poultry Genome is maintained at MSU. This homepage provides the latest maps and mapping data, an updated list of published microsatellites, descriptions of microsatellite kits, the latest cytogenetic map, and access to a host of other information. It can be accessed at <http://poultry.mph.msu.edu>. A chicken genomics homepage has also been developed (in the Acedb format) by Martien Groenen at <http://www.zod.wau.nl/vf>. Once the BAC contig information is available from Texas A&M, it will be publicly accessible, including a web-based Genomics Information System for efficiently manipulating and accessing the database (<http://hbz.tamu.edu/GIS>). This allows users direct access to the database, to capture, view, and manipulate data in a variety of ways. EST information is at www.chick.umist.ac.uk and at www.chickest.udel.edu. Washington University will provide access to the BAC contig map via its web site as it does for the human, mouse and *A. thaliana* genomes (<http://genome.wustl.edu>). All sequence traces will be deposited in the GenBank Trace Repository. The whole-genome shotgun sequence will be assembled using the recently developed public domain software packages (ARACHNE, MIT; PHUSION, Sanger Centre and JAZZ, JGI). Interactive displays of the assembled sequence and integrated BAC map will be available from the Washington University web site. Views of these data will also be integrated into the Ensembl human and mouse web displays using the DAS protocol where appropriate.

Cost and other funding sources. The cost of sequencing a 6-fold representation of the chicken genome by whole genome shotgun techniques is estimated at \$30M. The BAC map is well underway and the generation of approximately 150,000 BAC end sequences will provide a means of anchoring the whole-genome assemblies. These latter activities will add approximately \$1M to the project cost. Recent experience utilizing this strategy to sequence the mouse genome indicates that the resultant sequence map would provide coverage of approximately >90% of the chicken genome. Other groups have expressed interest in sequencing the chicken genome as well. These include the Beijing Genomics Institute (BGI), the BBSRC in the United Kingdom and possibly The Wellcome Trust/Sanger Centre. Should any of these groups maintain an interest in this genome and procure funds for sequencing it, the whole-genome shotgun strategy provides an easy means for the integration of such efforts with our own. Any additional sequence coverage beyond the 6-fold proposed here will enhance the sequence continuity and accuracy of the final product.