Supplementary Background

Origin of the chicken genome consortium

Given the many uses of the chicken in research and its importance as an agricultural commodity and based on the existence of a significantly large community of interested scientists, a proposal submitted to the National Human Genome Research Institute (NHGRI) in 2002 outlining the rationale for sequencing the chicken genome (McPherson et al.; http://genome.wustl.edu) was awarded a high priority\(^1\). Key aspects of that rationale are described below. The physical map (Wallis et al., this issue) and draft sequence of the 1.2 Gb chicken genome were completed at the Washington University Genome Sequencing Center in just 9 months. The draft genome sequence was released on March 1, 2004 (http://www.genome.gov/11510730). Additional whole genome sequencing was carried out by the Beijing Genome Institute on three domestic chicken varieties to identify and evaluate chicken genome variation (International Chicken Polymorphism Map Consortium, this issue). In parallel, more than 400,000 ESTs and 24,000 full length cDNA sequences were generated from a variety of different tissues to provide experimental evidence for transcription, to increase the precision of gene prediction and to validate the genome sequence assembly\(^2\).

Background

Evolution

The chicken genome serves as a model for those of ~9600 extant avian species, along with their evolutionary ancestors. Birds are part of the diapsid branch of vertebrates that split from the synapsid line (including mammals) over 310 My ago\(^3,4\) (Fig. 1, main text). Birds and crocodiles (and possibly turtles), along with their extinct progenitors, including
dinosaurs, constitute the archosauromorphs and, together with lepidosauromorphs (e.g., lizards, snakes), are the primary extant diapsids. The earliest fossils specifically assigned to the avian lineage are those of *Archaeopteryx*, dating to the late Jurassic period, about 150 My ago. Most studies suggest that many extant orders of birds were present prior to the Cretaceous-Tertiary boundary[^5-7] and date the split between *Galliformes* (land fowl, including all four Jungle Fowl species) and *Anseriformes* (water fowl) at 90 My ago and the origin of the Jungle Fowl genus itself, *Gallus*, at 8-9 My ago.

**Domestication and natural history**

Archeological evidence suggests chickens were domesticated in Asia at least by 5400 BC or perhaps even earlier in the Neolithic period (~ 8000 BC[^8]), as chicken bones were found associated with human artifacts of this age in Northeast China. The chickens kept by the Harappan Culture (2500-2100 BC) of the Indus Valley (today this region encompasses Pakistan and Western India) are considered the main source for subsequent global dispersal of domestic animals[^9,10]. Darwin[^11] suggested that the Red Jungle Fowl (RJF) was the nearest ancestor to the domestic chicken because it can interbreed with domestic birds producing fertile offspring, unlike the other Jungle Fowl (Grey, Green, Ceylon). Support for this view was provided by mitochondrial DNA analysis[^12] which indicated the Red Jungle Fowl native to Thailand are the monophyletic ancestor of the domestic chicken. It is hypothesized that chickens were originally utilized primarily for religious ceremonies and sporting purposes and, in fact, eating chicken was likely taboo in many cultures (and still is for some)[^13,14].

In addition to their use in agriculture and research, unusual varieties of chickens continue to be bred by poultry fanciers for exhibition purposes. Phenotypic variations found among domestic chickens based on physique alone (size, shape, plumage and comb) are remarkable; there are more than 300 combinations of features in over 120
breed standards with numerous varieties and an equal number of miniature versions (bantams). The extensive use of the chicken in literature, mythology, popular symbolism and for comic relief (what came first the chicken or the egg?) among cultures on a global scale speaks to interesting and positive ties to human cultural development (http://www.yale.edu/agrarianstudies/chicken/program.html).

**Agricultural relevance**

The chicken is the first agricultural animal to have its genome sequenced. Chickens continue to grow in importance as a source of high quality protein, with over 50 million tons of eggs and nearly 57 million tons of meat produced by the allied poultry industries in 2000\(^\text{15}\). Until the middle of the last century, most chicken breeds were raised for both meat and eggs, generally in fairly small flocks. However, the demand for more efficient production led to increasing genetic specialization and intensive selection, such that modern meat-type (broiler) and egg-type (layer) industrial breeds differ remarkably from each other and from their source breeds. Poultry breeders continue to make annual advances in productivity using quantitative genetic methods of selection. The molecular basis for the continued genetic adaptability of the chicken is almost completely unknown, but this genome sequence provides the opportunity to explore it (International Chicken Polymorphism Map Consortium, this issue). It is also worth noting that backyard-barnyard chicken flocks surviving on forage and food scraps remain a dependable source of animal protein for people in developing nations throughout the world.
One hundred years of chicken research: implications for human and chicken biology

Genetics: map and karyotype resources

Chickens have been a primary animal model for genetics for over 100 years since Bateson and Saunders’ classic experiments\(^ {16,17}\) established that Mendel's laws applied to animals, and Spillman\(^ {18}\) showed that feather-barring was sex-linked. Chickens are relatively straightforward to maintain, reproduce rapidly, and large crosses are generated easily. As with the laboratory mouse, but unlike most agricultural animals, inbred lines of chickens have been developed to standardize genetic backgrounds. Over the last 70 years well-characterized research resources have been developed, including mutant stocks with physiologic, metabolic, developmental and cytogenetic variants\(^ {19}\). Ironically, just as new opportunities are being afforded by the genome sequence, there is continuing loss of specialized chicken genetic lines, and there remains a lack of long-term conservation planning for the extant resources\(^ {20,21}\).

The first chicken genetic linkage maps based on morphological and physiological phenotypes were published by Serebrovsky and Petrov\(^ {22}\) and Hutt\(^ {23}\) (updated and reviewed by Bitgood and Somes\(^ {24}\)). The development of molecular DNA markers and the creation of internationally shared mapping populations dramatically improved these early maps. A consensus linkage map was published in 2000\(^ {25}\) based on three resource mapping populations\(^ {26-28}\) comprising 1889 loci and spanning 3800 cM. An additional ~300 loci subsequently have been added, bringing the current count to 2172 loci with a length of ~4000 cM. The consensus map consists of 51 linkage groups, several likely representing the same microchromosome. Thirty-one of these groups now have been assigned to a specific chromosome\(^ {29}\). Radiation hybrid (RH) mapping has met with
limited success in the chicken, and only recently has a useful chicken RH panel been established\textsuperscript{30}. Framework RH maps for individual chicken chromosomes are just beginning to appear\textsuperscript{31}. First generation BAC-based physical maps have also been produced for both individual chromosomes\textsuperscript{32,33} and for the genome as a whole\textsuperscript{34}. A more complete BAC contig physical map, developed in parallel with the genome sequence, is reported by Wallis et al. (this issue).

The karyotype of most birds consists of 40 pairs of chromosomes (2n=80) of dramatically different length; their number and appearance being quite distinct as compared to the standard mammalian karyotype. The chicken karyotype includes 38 autosomes and two sex chromosomes (Z and W). Although the chromosomes fall along a gradual size continuum, distinct size classes are obvious and here, for the purposes of analysis, we designated three groups: large macrochromosomes (GGA1-5), intermediate chromosomes (GGA6-10) and 28 microchromosomes (GGA11-38). In terms of relative scale, the macro- and intermediate chromosomes are similar in size to human chromosomes\textsuperscript{35}, whereas microchromosomes range down to sizes barely visible at the light microscope level of resolution. The inability to distinguish between the majority of microchromosomes has been a significant obstacle for cytogenetic mapping in chicken, until recently allowing for a standardized G-banded karyotype only for autosomes 1 through 8 plus the Z and W sex chromosomes\textsuperscript{36}. Recent developments in chromosome painting and fluorescence in situ hybridization (FISH) using bacterial artificial chromosome (BAC) probes has resulted in the cytogenetic identification of a large number of the microchromosomes\textsuperscript{29,37,38}.

**Sex chromosomes, genes and mechanisms**
In birds, the female is the heterogametic sex, having Z and W chromosomes; the male is homogametic having ZZ sex chromosomes. Although the avian Z and W are not orthologous to the mammalian X and Y, some general features are common. The Z chromosome, like the mammalian X, is conserved among avian lineages and is a large chromosome, whereas the W, like Y, is smaller, rich in heterochromatin and gene-poor. It is not yet known what gene triggers the avian sex determination pathway, analogous to the role of SRY for male determination in mammals, although a number of promising candidates are under study, including DMRT1 on the Z, and ASW(HINTW) and FET1 on the W. In fact, it remains uncertain whether the W plays a dominant role (analogous to Y) in avian sex determination or if dosage of the Z is critical. Chickens that are aneuploid solely for sex chromosomes are not available, but triploid (infertile) ZZW birds initiate development as females and then appear to sex reverse post-hatch, suggesting W initiation of femaleness but Z-dosage maintenance of maleness.

Additional models have been developed to suggest that Z and W gene interactions control avian sex determination rather than depending on a single dominant gene. Recent results indicate equal expression in males and females of several Z-linked genes, suggesting an active dosage compensation mechanism exists in chickens. However, genes are transcribed from both Z chromosomes in males, so any dosage compensation must be regulated by post-transcriptional mechanisms.

**Developmental biology**

Chickens have been immensely valuable as a model for developmental biologists. Chicken embryo development is morphologically similar to that of mammals with even specialized avian features such as scales and feathers providing
insight for development of homologous vertebrate structures. However, in chicken this
process occurs in ovo, allowing greater accessibility for experimental analysis of the fate
of embryonic tissues. Furthermore, the early chicken embryo develops along a flat plane
and is transparent, so the morphogenetic movements of cells and cell layers during
blastulation, gastrulation, neurulation, and somitogenesis are visible and accessible.
Embryos can also be cultured ex ovo. Thus, a wealth of classical and experimental
embryological literature has accumulated, dating from the writing of Hippocrates,
Aristotle’s description of the development of the chicken embryo (4th century BC) and
Hieronymus Fabricius’ accurate drawings chronicling daily development (16th century49)
to the classical work of the last century on the stages of chicken embryogenesis50-53.

The major concepts of developmental biology, such as embryonic induction and
embryonic fate maps, were extended to avian embryos in the 1930’s54,55. Due to its
experimental advantages, the chicken embryo rapidly became a major model organism
for the study of organogenesis. Saunders and Wolpert established avian limb bud
development as a model system for understanding embryonic patterning and
morphogenesis56. This work led to several concepts of modern developmental biology
such as positional information and morphogens. Other major breakthroughs in our
understanding of the molecular mechanisms underlying establishment of embryonic
segmentation were performed in the chick embryo57,58. Hamburger and Montalcini
pioneered the study of programmed cell death through transplantation studies, leading to
the identification of nerve growth factor59. The identification of morphological
differences between the nuclei of quail and chick cells was used by Le Douarin to
develop an extremely robust fate mapping technique60. The introduction of fluorescent
dye and green fluorescent protein labeling coupled to time-lapse imaging opened new
avenues for studying cellular dynamics in the chick embryo61, and the ability to introduce
exogenous DNA into developing tissues has facilitated the characterization of several
morphogenetic signaling pathways. Retroviral vectors have been employed in gain of function of experiments and have proven particularly useful in understanding the role of signaling systems in limb bud patterning\(^6\). More recently, \textit{in ovo} electroporation of the embryo was used in gain of function experiments in the developing nervous system and other tissues\(^6\). An alternative approach involves grafting local sources of secreted proteins or chemical inhibitors\(^6\). The lack of knock-out technology to generate loss of function mutations was overcome, in part, using electroporation of morpholinos, oligonucleotides, RNAi or ribozymes to induce local gene inactivation within the embryo\(^6\). The application of RNAi technology in the developing chicken embryo\(^6\) should boost the use of the chicken as a model for the analysis of gene function.

Developmental mutations (many inherited as single gene recessives) uncovered in large experimental and industry flocks during the 20\(^{th}\) century provide valuable assets for vertebrate developmental analysis (e.g., cleft palate, dwarfing, digit malformations, limbless and wingless, integument disorders, etc.). Over 30 lethal developmental mutations were initially described by Romanoff\(^5\); see Pisenti et al.\(^1\) and Delany\(^2\) for an updated listing of mutant lines held in N. America. Many of the mutations were carefully characterized for phenotype and inheritance pattern, and in several cases mechanistic explanations are now available. However, most of the underlying genes remain to be discovered, and the developmental mechanisms are still to be explored.

\textbf{Viral Oncogenesis}

The initial discovery by Peyton Rous in 1911 that injections of tumor filtrate produced tumors in healthy chickens initiated the field of viral oncology, leading to the characterization of Rous sarcoma virus, the subsequent discovery of oncogenes and proto-oncogenes, Temin's provirus hypothesis, reverse transcriptase and retroviral receptors (see Vogt\(^6\) for a chronology. The avian sarcoma-leukosis virus group remains
a key model for retrovirology today. The oncogenic Marek’s Disease herpesvirus provides a unique model for DNA tumor virology, and it is the only DNA tumor virus, to date, for which an effective vaccine is available. Host resistance to retroviruses, later shown to be a property of the viral receptor protein, was discovered in poultry \(^70\), and the chicken genome continues to be explored for alleles that confer resistance to a variety of pathogens, e.g. \(^71\).

**Immunology**

Studies of the chicken immune system led to the first distinction between B-(Bursa of Fabricius, the organ where Ig-producing lymphocytes are generated in avians) and T-cells (Thymus) \(^72\). The chicken immunoglobulin gene repertoire is diversified by a novel system of somatic mutation based on gene conversion \(^73\). The chicken major histocompatibility gene complex (MHC) is also of great interest. The chicken MHC is composed primarily of two large multigene clusters, the B-complex and Rfp-Y region that flank the nucleolar organizer region (18S-5.8S-28S rDNA complex) \(^74,75\).

An important application of the unique properties of the avian immune system has grown out of the observation of Buerstedde and Takeda \(^76\) that viral-transformed chicken lymphoid cell lines exhibit remarkably high rates of homologous recombination. This led to the wide use of the DT40 chicken B-cell line for genetic engineering of both avian and mammalian genes \(^48\) (see also [http://swallow.gsf.de/dt40.html](http://swallow.gsf.de/dt40.html)). Gene targeting in DT40 is largely undertaken to discover general aspects of gene function. The sequence described herein expedites the identification of worthwhile candidate genes for disruption and the subsequent analysis and interpretation of mutant phenotypes.
**Zoonoses**

Chickens and humans are infected by a number of common or related pathogens and share several disease resistance/susceptibility mechanisms. Example zoonotic diseases include salmonellosis, campylobacter, Newcastle disease and avian tuberculosis. A recent reminder of the importance of such processes was the transfer of avian influenza from chicken to human and the spread of avian influenza to commercial poultry flocks in the U.S. and elsewhere.\(^{77}\) Similarly, recent evidence suggests a partial avian origin for the SARS coronavirus.\(^ {78,79}\) Of note and concern regarding zoonotic disease is the fact that many human (and other) vaccines are produced in chicken embryonic cells. In addition to its significant contribution to human vaccine production, the chicken also provides a valuable model for vaccination strategies, for example, Marek’s Disease vaccination programs have pioneered the use of embryonic rather than post-natal administration.

**Cellular aging mechanisms: genome stability**

Several features of chicken telomere biology suggest that the chicken model is more similar to human than the rodent. Telomerase activity is developmentally down-regulated and telomeres shorten in the majority of terminally-differentiated somatic tissues, whereas telomerase activity re-emerges in transformed cells.\(^ {80-82}\) Interestingly, both chicken and human cells are more difficult to immortalize than rodent cells,\(^ {83-85}\) suggesting existence of common mechanisms/pathways governing genome stability and cellular immortalization. Notably, many avians exhibit maximum life expectancies similar to humans, and thus a stringent telomere clock may keep aging cells resistant to immortalization. Interestingly, birds also exhibit a higher basal metabolic rate (and body temperature) than mammals and may possess special protective mechanisms to respond to the generation of DNA-damaging oxidative radicals.
The ability to now “see”, in part, the edges of telomeric repeats (see below) within the draft sequence is important because researchers can identify and tag telomere-adjacent regions to study sequence and gene content. This has implications for future research on the impact of shortening of telomeric DNA on adjacent regions. Further, interstitial telomeric DNA segments have been implicated in model organisms as “hotspots” for recombination, and, more recently, terminal telomeric recombination hotspots have been observed in chicken (Rodrigue et al., submitted for publication). A key area for further research will be the exploration of the genetic mechanisms underlying enhanced recombination within the chicken genome and the role of repetitive elements in fostering recombination.

**Multifactoral factorial inheritance: Advantages of chicken QTL analysis**

A number of chicken genetic lines are models for similar physiologic conditions found in humans, *e.g.*, thyroiditis, dwarfing, muscular dystrophy, cleft palate, scleroderma, vitilligo, scoliosis, limb and digit malformations, neurological malformations, integument malformations, and retinal degeneration 19,86. The genome sequence of the chicken enhances the value of these models for analysis of the underlying genetic and physiologic mechanisms and possible therapeutic responses. In addition to specific chicken mutants, QTL studies of agricultural interest are of potential relevance to human multifactorial diseases as well. Indeed, there is considerable overlap between study of the chicken as a model organism and as a food animal87. The chicken is especially valuable for QTL analysis because it reproduces rapidly, and a number of highly inbred lines and resource populations are available.

Although several QTL searches have now been accomplished (for an overview, see [https://acedb.asg.wur.nl/](https://acedb.asg.wur.nl/)) or are in progress, only a small fraction of the interesting allelic diversity available in chickens has been examined to date. The specific
characteristics of chicken as a model in QTL studies will provide additional advantages in understanding more complex genetic phenomena, such as the importance of epistatic interactions in complex multigenetic traits. Recently, several causative mutations underlying QTL in livestock species were shown to be located within conserved regulatory modules (CRMs)\(^{89,90}\); the identification of a large number of such CRMs in the chicken genome sequence, therefore, is of relevance to the identification of the molecular basis of QTL. QTL mapping is expected to have significant impact in improving agricultural phenotypes such as disease resistance, meat quality and behavioral traits that are difficult to select in commercial poultry breeding. In addition, public concerns have already resulted in greater emphasis on animal well-being traits, whose understanding and application should benefit from new tools such as the genome sequence.
References