Minireview

Insights into vertebrate evolution from the chicken genome sequence
Rebecca F Furlong

Address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. E-mail: rebecca.furlong@zoology.ox.ac.uk

Abstract

The chicken has recently joined the ever-growing list of fully sequenced animal genomes. Its unique features include expanded gene families involved in egg and feather production as well as more surprising large families, such as those for olfactory receptors. Comparisons with other vertebrate genomes move us closer to defining a set of essential vertebrate genes.

The earliest bird fossils, from the genus *Archaeopteryx*, date back to the upper Jurassic period [1], around 150 million years ago. They show a mixture of dinosaur-like and bird-like features and lend support to the now widely accepted theory that birds evolved from dinosaurs. Birds are thus, along with most extant reptiles, members of the diapsid lineage, which split from the mammalian (synapsid) lineage around 310 million years ago. Chickens were domesticated over 7,000 years ago (reviewed in [2]) and are still of tremendous agricultural importance, and they have long been a model for biological research in fields ranging from embryology and development to virology and cancer. In addition, the phylogenetic position of the chicken, between fish and mammals, makes it ideal for comparative genomic analyses. It therefore came as no surprise when, in March 2003, the first complete avian genome sequence was initiated using the model for the undomesticated chicken, the red jungle fowl (*Gallus gallus*). Remarkably, barely one year later an initial draft assembly based on a 6.6X coverage of the genome was released into the public databases. The International Chicken Genome Sequencing Consortium [3] now reports an analysis of these data; here, I discuss some of the preliminary results from the chicken protein-coding gene dataset and the implications for our understanding of vertebrate evolution.

Gene data from each new fully sequenced genome contribute several different levels of information. Comparative analysis of complete genomes can be used to find conserved sequence elements, which may include previously unknown genes. The divergence of the compared genomes will determine the type of conservation found. Comparison of two closely related species, like human and mouse, will find many conserved regions within coding and non-coding DNA, but it may be impossible to determine which of these are functionally important. In contrast, a comparison between distantly related groups, such as fish and humans, may only detect well-conserved exonic sequences [4]. The chicken represents an intermediate-level comparison for the human, making it very useful for determining the essential features of the vertebrate genome. Comparison with genomes from other species can answer basic questions about how each lineage has diverged in gene content. In addition, similar comparisons can be used to assess the genomic changes that have taken place during evolution, such as chromosomal rearrangements and changes in the rate of evolution. Genomic analysis can provide us with a great deal of information about the organism itself; in the case of the chicken, this information will have applications in agriculture as well as in many different fields of basic research.

Conservation of gene order between chicken and human
The chicken has a haploid genome size of around 1.2×10^9 base-pairs, around 40% the size of mammals; it is estimated

that the genome sequence contains around 20,000-23,000 genes [3], a slightly smaller number than in mammals [5-7]. Many of the genes have been mapped to chromosomes, and these maps can be compared to other genomes to discover syntenic regions, where the same genes occur in a similar order along the chromosomes of different organisms. This does not just allow analysis of gene order in the chicken itself; the chicken genome can be used as an outgroup to the human and mouse genomes, allowing rates of gene rearrangement in the human genome and the architecture of the ancestral mammalian genome to be investigated. This approach uncovers a number of interesting features [3]. The rate of rearrangement in the human lineage is very slow compared to that of mouse, and that inferred for the mammalian common ancestor is slower still. When a fish outgroup is added, the analysis reveals that the rate of rearrangement on the chicken lineage is comparable to that of the mammalian common ancestor [3]. This supports a previous observation that synteny is more conserved between human and chicken than it is between human and mouse [8], and suggests that the stability of the chicken genome makes it a good candidate for future studies of vertebrate genome architecture.

Lineage-specific evolution of gene families

The chicken gene set can also be compared with those of mammalian genomes to discover lineage-specific changes to protein-coding genes or gene families, such as duplication or loss. In many cases, these changes mirror phenotypic change. For example, mammals appear to have lost several genes associated with egg production, in particular the avidin gene family [3]. These genes encode egg-white proteins and have homologs in invertebrates, indicating that they have been lost in mammals, probably in association with the reduction in egg size and internalization of the embryo on this lineage. The chicken genome appears to have fewer innovations and an enhanced rate of loss compared with other animal genomes [3]. Because the genome sequence is not finished and no other diapsid genomes are available for comparison, specific losses on this lineage cannot be discussed with confidence, but gain (or duplication) of genes can be determined with more certainty.

Gene-family expansion plays a substantial role in lineage-specific evolution. For example, both mammals and chickens have expanded their keratin gene repertoire by gene duplication, but in quite different directions [3]. Birds use a large, avian-specific family of keratin genes to form proteins for scales, claws and feathers. Mammals have undergone an independent expansion of a different keratin family, which is used to form hair fibers. A more surprising finding is that chickens have at least 218 non-identical genes that are orthologous to the human *OR5U1* and *OR5BF1* olfactory receptor genes [3]. Not only is this an exceptionally large expansion, but it is traditionally thought that birds have a

poor sense of smell [9]! The chicken genome sequence reveals that, thanks to this expansion, chickens have a similar number of olfactory receptor genes to humans [5,10], suggesting that their sense of smell may play more of a part in their behavior than previously thought.

A 'core' of essential vertebrate genes?

Comparisons with human and pufferfish (*Takifugu rubripes*) reveal around 7,000 chicken genes that have 1:1 orthologs in both species, suggesting a 'core' of genes that may have an essential role in all vertebrates [3]. The sequences in this core tend to be more conserved than other human/chicken orthologs, indicating that strong purifying selection is acting upon them, furthering the case for their functional importance. The results also suggest that these are genes that are expressed in many different tissues; this is not unexpected, as previous mammalian studies have suggested that rapidly evolving genes are expressed in fewer tissues [11,12]. The chicken genome [3] supports this theory: genes that can be found as expressed sequence tags (ESTs) from many tissues tend to be well conserved between human and chicken, whereas those expressed in few tissues are more divergent. The authors [3] also imply that a high proportion of the core genes are involved in cytoplasmic and nuclear functions, such as protein and intracellular transport. It would be interesting to discover how many of these core genes have previously been defined as mammalian housekeeping genes [13]. It should also be possible to examine whether any of these genes are also conserved across the invertebrates, to determine whether there is an animal-specific core of genes and how this differs from the vertebrate-specific core. It is generally accepted that an enhanced repertoire of developmental genes has played a role in the many innovations on the vertebrate lineage [14], but comparisons of this housekeeping dataset with invertebrate genomes - such as those of the fruit fly or sea squirt - could provide evidence for other sources of vertebrate novelty.

The chicken genome sequence assembly is currently estimated to cover 97% of the genome [3]. It is still very much in the draft phase, and a great deal of future work is likely to be necessary to refine the data. Despite the incompleteness of the protein-coding dataset, many new observations can be made about the structure and content of the avian gene set and how it compares with mammalian genomes. Analysis of the chicken genome also highlights the importance of sequencing genomes that lie in key positions on the tree of life: complete sequences of genomes from across the vertebrates, for example, would allow us to reconstruct the genome architectures of species at each node along this lineage. Closely related genomes can also reveal much, as in the case of rat and mouse genome analyses [7,15]. But no matter what organism it comes from, each new genome sequence has a fascinating story to tell, and adds more detail to our knowledge of genome evolution and organization.

References

1. Carroll RL: *Vertebrate Palaeontology and Evolution.* New York: WH Freeman; 1987.
2. Fumihito A, Miyake T, Sumi S, Takada M, Ohno S, Kondo N: **One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds.** *Proc Natl Acad Sci USA* 1994, **91:**12505-12509.
3. International Chicken Genome Sequencing Consortium: **Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432:**695-716.
4. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, *et al.*: **Comparative analyses of multi-species sequences from targeted genomic regions.** *Nature* 2003, **424:**788-793.
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
6. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420:**520-562.
7. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, *et al.*: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428:**493-521.
8. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D: **The dynamics of chromosome evolution in birds and mammals.** *Nature* 1999, **402:**411-413.
9. Jones RB, Roper TJ: **Olfaction in the domestic fowl: a critical review.** *Physiol Behav* 1997, **62:**1009-1018.
10. Malnic B, Godfrey PA, Buck LB: **The human olfactory receptor gene family.** *Proc Natl Acad Sci USA* 2004, **101:**2584-2589.
11. Duret L, Mouchiroud D: **Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate.** *Mol Biol Evol* 2000, **17:**68-74.
12. Zhang L, Li WH: **Mammalian housekeeping genes evolve more slowly than tissue-specific genes.** *Mol Biol Evol* 2004, **21:**236-239.
13. Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J, Dillon W, Lee KF, Clark KE, Haverty P, *et al.*: **A compendium of gene expression in normal human tissues.** *Physiol Genomics* 2001, **7:**97-104.
14. Furlong RF, Holland PWH: **Were vertebrates octoploid?** *Phil Trans R Soc Lond B Biol Sci* 2002, **357:**531-544.
15. Bourque G, Pevzner PA, Tesler G: **Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes.** *Genome Res* 2004, **14:**507-516.