Gallus GBrowse: a unified genomic database for the chicken

Carl J. Schmidt[1,*], Michael Romanov[2], Oliver Ryder[2], Vincent Magrini[3], Matthew Hickenbotham[3], Jarret Glasscock[3], Sean McGrath[3], Elaine Mardis[3] and Lincoln D. Stein[4]

[1]Department of Animal and Food Sciences, University of Delaware, Newark, DE 19706, [2]Conservation and Research for Endangered Species, Zoological Society of San Diego, San Diego, CA 92112-0551, [3]Genome Sequencing Center, Washington University School of Medicine, St. Louis, MO 63108 and [4]Cold Spring Harbor Laboratory, Bungtown Road, Cold Spring Harbor, NY 11724, USA

ABSTRACT

Gallus GBrowse (http://birdbase.net/cgi-bin/gbrowse/gallus/) provides online access to genomic and other information about the chicken, *Gallus gallus*. The information provided by this resource includes predicted genes and Gene Ontology (GO) terms, links to *Gallus In Situ* Hybridization Analysis (GEISHA), Unigene and Reactome, the genomic positions of chicken genetic markers, SNPs and microarray probes, and mappings from turkey, condor and zebra finch DNA and EST sequences to the chicken genome. We also provide a BLAT server (http://birdbase.net/cgi-bin/webBlat) for matching user-provided sequences to the chicken genome. These tools make the *Gallus* GBrowse server a valuable resource for researchers seeking genomic information regarding the chicken and other avian species.

INTRODUCTION

The chicken (*Gallus gallus*) has played important roles in both scientific research and the general health and welfare of humans. For example, in the field of developmental biology, the chicken embryo model has provided insight into many developmental processes including cell migration (1–3), limb development (4,5) and eye formation (6–8). The discovery of avian oncogenic viruses helped highlight the importance of specific genes in tumorigenesis and the chicken continues to be a popular model system for cancer and other diseases (9–11). As a food source, the chicken was domesticated in Asia ~7000–10 000 years ago and has undergone intensive selection for both egg and meat production over the past 60–70 years. In 2005, the United States (source: USDA National Agricultural Statistics Service) alone produced and consumed 30 billion pounds and exported another 5 billion pounds of chicken meat. In that same year, 90 billion eggs were produced in the United States. Clearly, the chicken plays an important role as both a model organism and as a food resource.

An enormous amount of genomic information and resources are available for the chicken. The genomic sequence of ~1 billion nucleotides was completed (12) and released in 2004 and then updated in 2006. A total of 3 335 290 SNPS (13) have been deposited in GenBank and over 1000 microsatellite (MS) and other genetic markers have been identified (14,15). At least five microarray platforms are available, and the *Gallus In Situ* Hybridization Analysis (GEISHA) (16) project is providing detailed descriptions of the embryonic expression pattern of many chicken genes. A centralized, web accessible, chicken database would provide a valuable resource for common access to this data. To begin providing such a resource, we have developed a Generic Model Organism Database (GMOD) (17) *Gallus* GBrowse site along with a BLAT server for searching the chicken genome. This site provides access to many chicken resources, along with mappings of turkey, condor and zebra finch nucleotide sequences to the chicken genome.

GALLUS GBROWSE DATA

The draft chicken genomic sequence (V2.1), produced by the Genome Sequencing Center at Washington University of St. Louis, was downloaded from the UCSC Genome Browser Gateway. The GMOD GBrowse viewer (17) in combination with a MySQL database management system is used to store, search and display annotation of the chicken genome. The GBrowse web page provides user access and is organized along themes including genes, gene expression platforms, gene expression data,

*To whom correspondence should be addressed. Tel: +1 302 831 1334; Fax: +1 302 831 2822; Email: schmidtc@udel.edu

Gene Ontology (GO) and pathways, markers and SNPs and other avian species.

Genes

The gene positions were defined based upon NCBI RefSeq and Ensembl cDNA predictions. These are provided as separate tracks in the GBrowse. In addition, predicted non-coding RNA genes and exon/intron positions are provided based on Ensembl predictions.

Gene expression platforms

These allow visualizing the positions of probes from five array platforms in the context of the chicken genome. Probe sequences for the Delmar (18), Avian Macrophage (19,20), Chicken 13K (21) and the Chicken Oligo microarray (http://www.grl.steelecenter.arizona. edu/products.asp) were aligned with the chicken genomic sequence using BLAT (22). The probe positions for the Affymetrix Chicken Genome Array were obtained from the NetAffx alignment file provided by Affymetrix.

Gene expression

Currently, two sets of gene expression data are accessed from *Gallus* GBrowse: GEISHA (16) and Unigene (23). The GEISHA project aims to describe the expression pattern of genes in the chicken embryo between Hamburger and Hamilton stages 1–25. The Unigene information is derived from the Unigene expression profiler, which describes the expression pattern for a gene based on EST analysis.

Gene ontologies and pathways

One set of tracks displays GO (24,25) terms for a given gene. GO terms were obtained from the Gene Ontology Annotation (GOA) Database via the NCBI database gene2go file. Hovering the mouse over the glyph will display the assigned GO term, while clicking on the link will connect to the Amigo term definition.

Reactome (26,27) is a human-centric curated knowledge base of biological pathways and pathways for other species are predicted by gene ortholog relationships. The *Gallus* GBrowse Reactome glyph links to the gene summary page in the Reactome knowledge base for the corresponding chicken gene. From the Reactome summary page, one can then access all pertinent information regarding the gene, including the reactions, pathways and molecular complexes the gene product participates in, as well as the gene's orthologs in human and other model species.

Markers and SNPS

Markers were obtained from the NCBI UniSTS ftp site, or from a sequence file provided by Dr Martien Groenen (Wageningen University). The genomic locations of these sequences were then determined by BLAT analysis. SNPs were also mapped to the genome by BLAT using the flanking sequence obtained from the NCBI dbSNP database. Because of the high density of SNPs mapped (>3 000 000) to the chicken genome, the SNP track is only visualized at a zoom scale of 250 000 nucleotides or lower. Clicking on an individual SNPs glyph will link to the NCBI cluster report for that SNP.

Other avian species

To help integrate analysis of the chicken with other avian species, genomic and cDNA data from the turkey (28–31), condor and zebra finch (32,33) have been mapped to the chicken genome by BLAT. Turkey DNA and zebra finch DNA sequences were obtained from NCBI along with the condor MS sequences. The condor 454 sequences were derived from fibroblast ESTs determined using the 454 sequencing technology (34).

DNA

This track visualizes the DNA sequence of the current region. The nucleotide sequence is only presented at a zoom of 100 base pairs. At higher zoom levels, the %GC content is displayed.

QUERY TOOLS

The *Gallus* GBrowse web page provides an integrated query interface. Specific chromosomal regions of 10 megabases or less can be accessed with known nucleotide coordinates using the Landmark or Region search box (Figure 1). This same search box can be used to locate specific information stored in the GBrowse database. For example, one can search for all genes annotated with the GO term 'apoptosis' by inserting 'GO:apoptosis' in the Landmark or Region box (Figure 2). This yields a total of 14 genes that have been annotated with 'apoptosis' in the chicken genome. A complete listing of all query prefix terms (such as GO) is provided in the *Gallus* GBrowse help pages.

One of the more challenging aspects of using many genomic databases is searching based on a gene name. As a convention, *Gallus* GBrowse uses chicken gene names assigned by NCBI and entering a search in the syntax 'NCBI:gene name' will typically recover the desired information. Another approach can be to use a homologous nucleotide or protein sequence and the BLAT server (below) to identify the chromosomal location of the gene of interest.

A BLAT server is provided (http://birdbase.net/cgi-bin/webBlat) to allow searching the chicken genome with either nucleotide or protein sequences. Two databases are provided, one containing the nucleotide sequence (Chicken Genome untranslated) and the other containing the chicken genome translated in all reading frames (Chicken Genome translated). To successfully execute a BLAT search, the appropriate database must be selected for nucleotide (untranslated) or protein (translated) input sequence. Results from the BLAT analysis are returned as two web links, one showing the alignment of the query sequence with the matched chicken genomic sequence, and the second displaying the *Gallus* GBrowse viewer focused on the region of the aligned query sequence.

Figure 1. *Gallus* GBrowse. A portion of chicken chromosome 19 (nucleotides 5129862–5170001) shown with glyphs depicting predicted genes (chicken), links to Unigene, Reactome and Gene Ontology annotation, SNPs and the location of turkey, zebra finch and condor ESTs that have been mapped to the chicken genome.
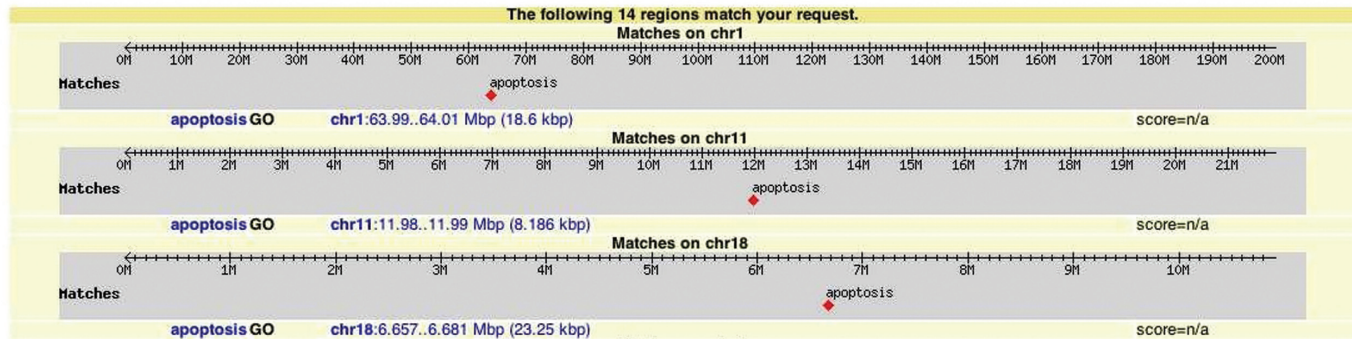
Figure 2. Searching *Gallus* GBrowse for specific entries. The search term GO:apoptosis was entered in the Landmark or Region box followed by pushing the Search button. A total of 14 entries were found, only three are presented here. Using the mouse to click on the chromosome link will open the browser to that location.

FUTURE DIRECTIONS

The *Gallus* GBrowse will be updated as new relevant information becomes available. One near term objective is to incorporate the position of repetitive sequence elements into the GBrowse database. An additional goal is incorporating both microarray and high-throughput EST sequencing data to describe gene expression patterns. Initially this will likely to reflect a simple interpretation of whether or not a gene was detected above background and allow users to determine if a given gene is expressed under the experimental conditions of the microarray or sequencing assay. *Gallus* GBrowse will also be improved by linking genes with the curated ontology efforts of AgBase (35,36). The current GO entries are derived from uncurated, electronic annotation and the AgBase effort should provide a far more reliable and accurate assignment of GO terms. Finally, a long-term goal is to continue incorporating genomic information from other avian species with the adoption of additional GMOD tools and the Chado database schema. We hope to ultimately provide an integrated resource for comparative avian genomics.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Dupin,E., Ziller,C. and Le Douarin,N.M. (1998) The avian embryo as a model in developmental studies: chimeras and in vitro clonal analysis. *Curr. Top. Dev. Biol.*, **36**, 1–35.
2. Funk,P.E. and Thompson,C.B. (1996) Current concepts in chicken B cell development. *Curr. Top. Microbiol. Immunol.*, **212**, 17–28.
3. Qureshi,M.A., Heggen,C.L. and Hussain,I. (2000) Avian macrophage: effector functions in health and disease. *Dev. Comp. Immunol.*, **24**, 103–119.
4. Tickle,C. (1999) Morphogen gradients in vertebrate limb development. *Semin. Cell Dev. Biol.*, **10**, 345–351.
5. Viallet,J., Garcia,A. and Weydert,A. (2003) Protein phosphatase 2A as a new target for morphogenetic studies in the chick limb. *Biochimie*, **85**, 753–762.
6. Bailey,T.J., El-Hodiri,H., Zhang,L., Shah,R., Mathers,P.H. and Jamrich,M. (2004) Regulation of vertebrate eye development by Rx genes. *Int. J. Dev. Biol.*, **48**, 761–770.
7. Glover,J.C. (2003) The development of vestibulo-ocular circuitry in the chicken embryo. *J. Physiol. Paris*, **97**, 17–25.
8. Naito,J. and Chen,Y. (2004) Morphological features of chick retinal ganglion cells. *Anat. Sci. Int.*, **79**, 213–225.
9. Davison,T.F. (2003) The immunologists' debt to the chicken. *Br. Poult. Sci.*, **44**, 6–21.
10. Witter,R.L. (1997) Avian tumor viruses: persistent and evolving pathogens. *Acta Vet. Hung.*, **45**, 251–266.
11. Fadly,A.M. (1997) Avian retroviruses. *Vet. Clin. North Am. Food Anim. Pract.*, **13**, 71–85.
12. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
13. Wong,G.K., Liu,B., Wang,J., Zhang,Y., Yang,X., Zhang,Z., Meng,Q., Zhou,J., Li,D. *et al.* (2004) A genetic variation map for

chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, **432**, 717–722.

14. Fillon,V., Vignoles,M., Garrigues,A., Pitel,F., Morisson,M., Crooijmans,R.P., Groenen,M.A., Gellin,J. and Vignal,A. (2003) The chicken cytogenetic map: an aid to microchromosome identification and avian comparative cytogenetics. *Br. Poult. Sci.*, **44**, 795–797.

15. Groenen,M.A., Cheng,H.H., Bumstead,N., Benkel,B.F., Briles,W.E., Burke,T., Burt,D.W., Crittenden,L.B., Dodgson,J. *et al.* (2000) A consensus linkage map of the chicken genome. *Genome Res.*, **10**, 137–147.

16. Bell,G.W., Yatskievych,T.A. and Antin,P.B. (2004) GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos. *Dev. Dyn.*, **229**, 677–687.

17. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

18. Cogburn,L.A., Wang,X., Carre,W., Rejto,L., Aggrey,S.E., Duclos,M.J., Simon,J. and Porter,T.E. (2004) Functional genomics in chickens: development of integrated-systems microarrays for transcriptional profiling and discovery of regulatory pathways. *Comp. Funct. Genom*, **5**, 253–261.

19. Bliss,T.W., Dohms,J.E., Emara,M.G. and Keeler,C.L.Jr (2005) Gene expression profiling of avian macrophage activation. *Vet. Immunol. Immunopathol.*, **105**, 289–299.

20. Keeler,C.L., Bliss,T.W., Lavric,M. and Maughan,M.N. (2007) A functional genomics approach to the study of avian innate immunity. *Cytogenet. Genome Res.*, **117**, 139–145.

21. Burnside,J., Neiman,P., Tang,J., Basom,R., Talbot,R., Aronszajn,M., Burt,D. and Delrow,J. (2005) Development of a cDNA array for chicken gene expression analysis. *BMC Genomics*, **6**, 13.

22. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.

23. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R. *et al.* (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **35**, 5–12.

24. The Gene Ontology Consortium (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, **34**, 322–326.

25. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

26. Joshi-Tope,G., Gillespie,M., Vastrik,I., D'Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.

27. Vastrik,I., D'Eustachio,P., Schmidt,E., Joshi-Tope,G., Gopinath,G., Croft,D., de Bono,B., Gillespie,M., Jassal,B. *et al.* (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.*, **8**, R39.

28. Chaves,L.D., Rowe,J.A. and Reed,K.M. (2005) Survey of a cDNA library from the turkey (*Meleagris gallopavo*). *Genome*, **48**, 12–17.

29. Reed,K.M., Knutson,T.P., Krueth,S.B., Sullivan,L.R. and Chaves,L.D. (2005) In silco mapping of ESTs from the turkey (Meleagris gallopavo). *Anim. Biotechnol.*, **16**, 81–8102.

30. Dranchak,P.K., Chaves,L.D., Rowe,J.A. and Reed,K.M. (2003) Turkey microsatellite loci from an embryonic cDNA library. *Poult. Sci.*, **82**, 526–531.

31. Smith,E., Shi,L., Drummond,P., Rodriguez,L., Hamilton,R., Powell,E., Nahashon,S., Ramlal,S., Smith,G. *et al.* (2000) Development and characterization of expressed sequence tags for the turkey (*Meleagris gallopavo*) genome and comparative sequence analysis with other birds. *Anim. Genet.*, **31**, 62–67.

32. Wade,J., Peabody,C., Coussens,P., Tempelman,R.J., Clayton,D.F., Liu,L., Arnold,A.P. and Agate,R. (2004) A cDNA microarray from the telencephalon of juvenile male and female zebra finches. *J. Neurosci. Methods*, **138**, 199–206.

33. Li,X., Wang,X.-J., Tannenhauser,J., Podell,S., Mukherjee,P., Hertel,M., Biane,J., Masuda,S., Nottebohm,F. *et al.* (2007) Genomic resources for songbird research and their use in characterizing gene expression during brain development. *Proc. Natl Acad. Sci. USA*, **104**, 6834–6839.

34. Margulies,M., Egholm,M., Altman,W.E., Attiya,S., Bader,J.S. and Al,E. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

35. McCarthy,F.M., Bridges,S.M., Wang,N., Magee,G.B., Williams,W.P., Luthe,D.S. and Burgess,S.C. (2007) AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Res.*, **35**, D599–D603.

36. McCarthy,F.M., Wang,N., Magee,G.B., Nanduri,B., Lawrence,M.L., Camon,E.B., Barrell,D.G., Hill,D.P., Dolan,M.E. *et al.* (2006) AgBase: a functional genomics resource for agriculture. *BMC Genomics*, **7**, 229.