International Chicken Genome Workshop - Final Report

Hinxton Genome Campus, Cambridge – 10-11 March 2003

Sponsored by Cobb, Aviagen, Intervet, Genesis Faraday Partnership and ARK-genomics

Many will be pleased to hear that sequencing of the chicken genome started to run at full production in March 2003 at the Washington University Genome Centre (St Louis, USA) and the assembly of the draft sequence is expected in August 2003. In order to coordinate the relevant materials with the release of this information a group of over 40 scientists recently held a workshop at the Wellcome Trust Genome Campus in Cambridge (UK).

The idea behind this workshop started in autumn-2002 when it was clear that the Washington University Genome Centre would sequence the chicken genome by the end of 2003 with funding from NHGRI (USA). This has been a wish of many investigators for the past ten years – we did not imagine it would happen so soon and so fast!

The aim of the workshop was to present the current status of the chicken genome project, prospects and future needs. The group also wanted to discuss important issues relating to the annotation of the chicken genome and bring everyone up to date on the international resources being developed in chicken biology (EST, full-length cDNA clones, microarrays, expression pattern databases, QTL mapping, SNP detection, etc…). In this way the workshop hoped to identify important data sources currently available and identify those required. Finally, an action plan was proposed to coordinate the development of these resources through national and international funding sources.

The aim of this report is to briefly summarise the workshop and its recommendations. We hope you will agree, that the information from the chicken genome will have a significant role to play not only in animal breeding and animal health but also in basic biology, development, medicine and a wide range of other applications.

Workshop Presentations

John McPherson started off the workshop with an excellent summary of the sequencing programme at the Wash U Genome Center. This programme is based on the experience gained from the human and mouse genome projects and will include a mixture of targeted and whole genome sequencing. Currently a physical map based on a 20-fold genome coverage of overlapping BAC clones is being assembled as a collaboration between BAC clones from Wash U (*John McPherson*), Texas A & M (*Hongbin Zhang*), MSU (*Jerry Dodgson*) and Wageningen University (*Martien Groenen*). This map will be integrated with other chicken mapping resources; including FISH mapped clones (cytogenetic map), genetic markers (genetic map) and radiation hybrid panels (RH map, currently 1000 markers mapped, 1-cM ~ 5-cR, ~250-kb, *Alain Vignal*). Low sequence coverage of selected BAC clones, and BAC and Fosmid end-sequences will be used to merge the BAC map with a 8-fold whole genome shotgun (WGS). Preliminary assemblies based on 2.5-Mb of finished sequence have shown that this assembly approach should be very efficient for the chicken. The high repeat content (~50%) in mammalian genomes has been a major problem with the WGS strategy, however in the chicken the repeat content is only 15%. This was shown dramatically in the repeat plots of finished genome sequences from human, mouse and chicken. Also, more than 75% of WGS sequences from chicken BAC clones were contained in a single contig compared to only 25% in the mouse. The sequence will be based on an inbred Jungle Fowl used by *Jerry Dodgson* (MSU) and *Hans Cheng* (ADOL) to create the East Lansing genetic

linkage map. The BAC map and the sequence assembly will be available from the Wash U www site (http://genome.wustl.edu/) starting April 2003.

Bin Liu from the Beijing Genome Institute (BGI) talked about their plans to sequence 100,000 chicken ESTs and chicken genome sequences (with support from the Wellcome Trust, UK). There was much discussion on what strains BGI should sequence and it was agreed that a one-fold genome coverage of both a White Leghorn (WL, from *Leif Andersson*, Upssala) line and a Broiler (B, *Dave Burt*, Roslin Institute) line would be sequenced. When compared to the Jungle Fowl (JF) sequence this would generate a rich resource of single-nucleotide polymorphism's (SNPs) that could be related back to the quantitative trait loci (QTL) being mapped in the original JF x WL (*Leif Andersson*, Upssala) or B x WL (*Dave Burt,* Roslin Institute) populations.

Ewan Birney from the EBI team then gave a summary of the Ensembl infrastructure – a key window on many metazoan genomes (human, mouse, zebrafish, rat etc). The complex Ensembl pipeline (gene prediction, genome annotation, ESTs, SNPs, etc) was explained which could be adapted for any other genome, including the chicken. The role of automatic and community annotation was discussed. Finally, the power of comparative genomics was discussed not only for improved gene prediction but also in new exciting areas such as genome evolution and gene regulation. *Rolf Apweiler* (EBI) continued the theme with thoughts on the prediction of protein function. Rolf provided an overview of databases resources, including SwissProt (1064 chicken/122,564 proteins), TrEMBL (2047 chicken/830,525 proteins) and functional databases InterPro and GOA. *Midori Harris* (EBI) continued with gene ontology's currently covering molecular function, biological processes and cellular components. Future developments include ontology's to model biology and experimentation, areas we expect the chicken consortium to contribute. It is expected that the chicken will enrich both these areas both as a model bird and a vertebrate. *Andy Law* (Roslin Institute) talked about problems and potential solutions for integration of diverse data types common to the chicken and other genome projects. This was followed by a series of talks, which illustrated the richness of these data sources already available in the chicken or ones needed in the near future.

During the last 12 months there has been dramatic progress in the development of chicken EST resources (most available from www.ark-genomics.org), with the chicken as one of the top four in the current release of dbEST. *Simon Hubbard* (UMIST) gave an excellent overview of the characteristics of the UK-EST resource (~350,000 sequences in GenBank). Other EST programmes by *Larry Cogburn* (Delaware) and *Jean-Marie Buerstedde* (Munich), *Leif Andersson* (Upssala) and *Dave Burt* (Roslin) were described. In total there are now over 600,000 chicken ESTs (UK, USA, France, Sweden, China) and various analyses suggest this collection represents ~ 32,000 gene clusters, however alignment with the chicken genome sequence will improve this estimate – which is about the same as other higher vertebrates, including human. Future plans are to sequence 13,000 full-length cDNA sequences within the next 12 months from these various collections (*Sanger Centre and Munich*). EST collections are also a rich source of SNP information (1 SNP per 1000-bp transcribed DNA *Simon Hubbard*, which is 10-fold lower than in non-coding DNA, *Dave Burt*) and when combined with the genome sequencing efforts described above will provide a powerful predictive tool for gene association studies in the chicken.

The availability of large collections of sequenced chicken cDNA clones from diverse tissue sources has provided new opportunities for gene expression studies in physiology and developmental biology. *Larry Cogburn* (Delaware) described the use of a number of specialised cDNA microarrays being used studies on metabolism and immune responses.

International collaboration between the UK and USA form the basis of the current generic cDNA microarrays being fabricated at ARK-genomics (*Dave Burt*) on two 14K arrays. Microarrays provide a high throughput technology for gene expression studies but lack resolution at the cellular level. *Parker Antin* (Arizona) discussed how chicken cDNA clones were being used in high throughput whole-mount *in situ* hybridisation studies to examine patterns of gene expression during chick development. Parker described *Geisha*, the first attempt to build a chicken gene expression database to store whole mount *in situ* hybridisation images. Future developments require GO annotation, Developmental ontology, more stages and scale up. *Duncan Davidson* (Edinburgh) described progress in this field in the mouse, and how we could adapt the tools and ideas for the chicken.

The availability of these gene expression tools opens up new opportunities for the analysis of pathways and gene networks active during development and physiological processes. This was illustrated very clearly by *Claudio Stern* (London) who described gene networks initiated from the Hensen Node; a key structure in early development. The work itself used a number of tools, including whole mount *in situ* hybridisation, misexpression of genes within the Hensen Node pathway, selection of DNA binding motifs using purified transcription factors (Churchill, a Zinc finger gene) and prediction of target genes, then further functional studies knocking out gene function using morpholinos. Finally, by looking at conserved regions in mouse and human genes, Claudio predicted regulatory regions and was able to test their function in chicken embryos using GFP-gene fusions! In a similar vein, *Paul Neiman* (Fred Hutchison Cancer Research Centre) described experiments in the chicken that through light on the role of myb-pathways in tumour development. *Jean-Marie Buerstedde* (Munich) described how gene knockouts in the chicken DT40 cell line are being used very effectively to understand immune function and metabolism. *Stuart Wilson* (UMIST) described progress in the use of *RNAi* to knockdown gene expression and how he has extended these ideas to chicken embryos. Clearly the chick has come of age and is primed for post-genomic developmental studies.

The chicken genome project had its origins in a number of genome mapping projects started ten years ago, in which the nature of quantitative traits was of primary interest. *Leif Andersson* (Upssala), *Dave Burt* (Roslin Institute) and *Martien Groenen* (Wageningen) each provided overviews on their work to map QTL for a wide range of traits, including growth, body weight, carcass composition, egg production, fatness, ascites, feather pecking, stress, etc. So far, these and other groups have defined over 250 QTL. Plans were presented on how the genes at QTL were to be defined – the proposed SNP map of the chicken would be a crucial tool in this search – the story that is emerging is that the chicken is a powerful tool for the study of the molecular basis of quantitative genetic variation.

Finally, *Elliott Margulies* (NHGRI) gave a truly excellent talk on the power of multi-species genome comparisons for the prediction of coding and regulatory regions. Eric Green's lab at the (NHGRI, USA) is currently sequencing about 50 selected genome regions from over 20 species. They have developed a number of bioinformatic tools to visualise comparisons between these species, able to correct for differences in base substitution rates and phylogenetic distances. Through these comparisons Elliot was able to define "Multi-species Conserved Segments" or MCS's apart from coding regions these are likely to represent regulatory or cis-acting functional elements. The exciting finding was that comparisons with the chicken rather than mammals or fish, were the most successful in detecting MCS's both in coding and non-coding regions.

Action Plan and the Future

Dave Burt (Roslin Institute) and *Olivier Pourquie* (Stowers) ended the meeting with all the workshop participants by discussing plans for the future of chicken genome research. These are summarised: -

- During the workshop an **International Chicken Genome Consortium** was established (**Appendix 1**), co-chaired by *Dave Burt* and *Olivier Pourquie*. The role of this committee is to provide coordination and leadership for research that will benefit from the chicken genome sequence (annotation of genome, sequencing lines in search of SNPs, cDNA microarrays, gene expression patterns, EST resources, etc). To facilitate this research the consortium has been organized around a **Steering Group** and a number of **Technical Groups**, each responsible for delivering the research (sequencing, gene expression, chicken biology, proteomics).
- Finally, the workshop agreed on a list of research priorities listed in **Appendix 2**, in order of priority, including coordinating labs, funding sources and current status of funding.
- The chicken community has long suffered from its division between the agricultural world, interested in improving existing breeds by genetics, in their immune function and associated pathologies and the academic world, for the which the chick embryo has constituted an important model for decades, and has also provided models such as the DT40 cells which exhibit a recombination rate similar to that seen in yeast. Therefore, the release of this genome sequence is without doubt going to be an important event for scientific communities as diverse as developmental biologists, geneticists, genome biologists, immunologists and others. In addition to the international consortium it is clear we need a chicken web site to facilitate the exchange of information. During the meeting this idea was discussed and it was proposed that we should establish **ChickNET** (http://www.chicken-genome.org), a network of www sites with an interest in the chicken genome, developmental biology, genetics, biodiversity, immunology, links to other species through shared biological interest (Flybase, ZFIN, etc), etc.
- *Andy Law* (Roslin Institute) has set up a mail list chicken-genome@lists.bbsrc.ac.uk, and general information about the mailing list is at https://www.lists.bbsrc.ac.uk/mailman/listinfo/chicken-genome
- Finally, the co-chairs *Dave Burt* (Roslin Institute) and *Olivier Pourquie* (Stowers) will organise another genome workshop (~100 participants?) likely to be held at the Stowers Institute (USA) mid-2004, at a time when the chicken genome data is ready to interpret.

APPENDIX 1: International Chicken Genome Consortium

Co-Chairs: Dave Burt (UK) & Olivier Pourquie (USA)

Steering Group

Leif Andersson (Sweden), Parker Antin (USA), Nat Bumstead (UK), Joan Burnside (USA), Jerry Dodgson (USA), Martien Groenen (The Netherlands), Ning Li (China), Cheryll Tickle (UK), Wes Warren (USA) and Stuart Wilson (UK)

Technical Groups

Genome Sequencing

Raw sequence: Ewan Birney, Bin Liu, John McPherson, Wes Warren
Ensembl: Ewan Birney, Dave Burt
Physical mapping: John McPherson, Hongbin Zhang, Martien Groenen, Dave Burt, Alain Vignal
SNPs: Leif Andersson, Ewan Birney, Dave Burt, Martien Groenen, Bin Liu

Gene Expression

ESTs: Leif Andersson, Joan Burnside, Jean-Marie Buerstedde, Dave Burt, Larry Cogburn, Simon Hubbard, Jacques Samarut, Bertrand Pain, Ning Li, Cheryll Tickle, Madeleine Douaire
Microarrays: Nat Bumstead, Joan Burnside, Dave Burt, Larry Cogburn, Paul Neiman, Ning Li
Atlas of Anatomy: Duncan Davidson, Claudio Stern, Cheryll Tickle
Adult tissues and pathology: Nat Bumstead, Hans Cheng, Duncan Davidson
Gene expression database: Parker Antin, Dave Burt, Duncan Davidson, Olivier Pourquie, Claudio Stern, Cheryll Tickle

Chicken Biology

QTL: Leif Andersson, Ewan Birney, Dave Burt, Hans Cheng, Andy Law, Martien Groenen
Chicken mutants: Dave Burt, Mary Delany
Loss and Gain of Function: Helen Sang, Stuart Wilson, Claudio Stern, Cheryll Tickle, Jean-Marie Buerstedde, Bertrand Pain, Jacques Samarut

Proteomics

Proteome database: Rolf Apweiler, Rob Beynon, Shane Burgess, Simon Hubbard

APPENDIX 2: Research Priority Areas

Priority	Research Task	Coordinating Lab(s)	Funding Source	Status
1	Chicken genome sequence assembly	WASHU-BGI-EBI	NHGRI-USA	Funded
2	Chicken Ensembl database	Roslin-EBI-Stowers-WASHU	BBSRC-USDA?	Funded
3	SNP discovery from broiler/layer lines	BGI-Sanger-EBI-Roslin-Upssala	Wellcome Trust-?	Part-funded
4	Full-length cDNA sequencing	UMIST-Dundee-Nottingham-Sanger	BBSRC	Funded
5	QTL database	Roslin-Wageningen	BBSRC	Funded
6	Microarray gene expression database	Roslin-FHCRC	BBSRC-?	Part-funded
7	Atlas of chick embryo anatomy	GEISHA (UArizona) HGU-Dundee-	NIH?	New
8	Large Scale in situ screen	GEISHA, (UArizona) HGU-	NIH?	New
9	Repository of *in situ* images	GEISHA (UArizona) HGU-	NIH?	New
10	Ontology of adult tissues and pathology	UADOL-?	?	New
11	Large-scale loss of function using *RNAi*	UMIST-Roslin-?	?	New
12	Chicken proteome database	EBI-UMIST	EMBL-?	Part-funded
13	Genome conservation, gene networks, co-regulated genes, promoter screens	EBI-UCL-Roslin-?	?	New

Note: Shaded Boxes indicate GEISHA project priorities.

APPENDIX 3: Web sites

Andy Law: http://www.thearkdb.org

Claudio Stern: http://sternlab.anat.ucl.ac.uk

Dave Burt: http://www.ark-genomics.org

dbEST: http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html

Duncan Davidson: http://www.hgu.mrc.ac.uk/Research/Devgen/MouseAtlas/richdunc.htm

Elliott Margulies: http://www.genome.gov/Staff/Green

Ewan Birney: http://www.ensembl.org

Hongbin Zhang: http://hbz.tamu.edu

Jean-Marie Buerstedde: http://swallow.gsf.de/dt40Est.html

Jerry Dodgson: http://poultry.mph.msu.edu/index.html

John McPherson: http://genome.wustl.edu

Larry Cogburn: http://udgenome.ags.udel.edu/~cogburn

Martien Groenen: http://www.zod.wau.nl/vf

Midori Harris: http://www.geneontology.org

Parker Antin: http://geisha.biosci.arizona.edu

Rolf Apweiler: http://www.ebi.ac.uk

Simon Hubbard: http://www.chick.umist.ac.uk

APPENDIX 4: Contact List

1.	Andersson, Leif	Leif.Andersson@bmc.uu.se
2.	Antin, Parker	pba@U.Arizona.EDU
3.	Apweiler, Rolf	Rolf.Apweiler@EBI.ac.uk
4.	Bentley, James	james.bentley@merial.com
5.	Beynon, Rob	r.beynon@liverpool.ac.uk
6.	Birney, Ewan	birney@ebi.ac.uk
7.	Buerstedde, Jean-Marie	buersted@gsf.de
8.	Brown, William	William.brown@nottingham.ac.uk
9.	Bumstead, Nat	Nat.Bumstead@bbsrc.ac.uk
10.	Burgess, Shane	burgess@cvm.msstate.edu
11.	Burnside, Joan	joan@UDel.Edu
12.	Burt, Dave	Dave.Burt@bbsrc.ac.uk
13.	Cheng, Hans	hcheng@pilot.msu.edu
14.	Clamp, Michele	michele@sanger.ac.uk
15.	Cogburn, Larry	cogburn@UDel.Edu
16.	Davidson, Duncan	duncan.davidson@hgu.mrc.ac.uk
17.	Degen, Winfried	Winfried.Degen@intervet.com
18.	Delany, Mary	medelany@ucdavis.edu
19.	Douaire, Madeleine	mdouaire@roazhon.inra.fr
20.	Ellis, Leland	Leland.Ellis@NPS.ARS.USDA.GOV
21.	Game, Alf	Alf.Game@bbsrc.ac.uk
22.	Griffin, Darren	darren.griffin@brunel.ac.uk
23.	Groenen, Martien	martien.groenen@alg.vf.wau.nl
24.	Hall, Tony	tony.hall@cherryvalley.co.uk
25.	Harris, Midori	midori@ebi.ac.uk
26.	Harry, Dave	deharry@ix.netcom.com
27.	Hubbard, Simon	sjh@bms.umist.ac.uk
28.	Hubbard, Tim	th@sanger.ac.uk
29.	Hicks, Anne	Anne.Hicks@bbsrc.ac.uk
30.	Law, Andy	Andy.law@bbsrc.ac.uk
31.	Li, Ning	ninglbau@public3.bta.net.cn
32.	Liu, Bin	liub@genomics.org.cn
33.	McKay, Jim	JMcKay@aviagen.com
34.	McPherson, John	jmcphers@watson.wustl.edu
35.	Margulies, Elliott	elliott@nhgri.nih.gov
36.	Neiman, Paul	pneiman@fred.fhcrc.org
37.	Pain, Bertrand	Bertrand.Pain@ens-lyon.fr
38.	Peterson, Jane	petersoj@exchange.nih.gov
39.	Pevzner, Igal	PEVZNERI@cobb-vantress.com
40.	Pourquie, Olivier	OLP@Stowers-Institute.org
41.	Rogers, Jane	jrh@sanger.ac.uk
42.	Samarut, Jacques	jsamarut@ens-lyon.fr
43.	Sang, Helen	Helen.Sang@bbsrc.ac.uk
44.	Stern, Claudio	c.stern@ucl.ac.uk
45.	Tickle, Cheryll	c.a.tickle@dundee.ac.uk
46.	Vignal, Alain	vignal@toulouse.inra.fr
47.	Warren, Wes	wwarren@watson.wustl.edu
48.	Warkup, Chris	chris.warkup@bbsrc.ac.uk

49. Wilson, Stuart s.a.wilson@umist.ac.uk
50. Zhang, Hongbin hbz7049@tamu.edu