Farm animal genomics and informatics: an update
Ahmed Fadiel*, Ifeanyi Anidi and Kenneth D. Eichenbaum1

Yale University School of Medicine, Yale Center for Research On Reproductive Biology, New Haven, CT 06511, USA and 1Mount Sinai School of Medicine, New York, NY 10029, USA

Received July 7, 2005; Revised August 15, 2005; Accepted October 10, 2005

ABSTRACT
Farm animal genomics is of interest to a wide audience of researchers because of the utility derived from understanding how genomics and proteomics function in various organisms. Applications such as xenotransplantation, increased livestock productivity, bioengineering new materials, products and even fabrics are several reasons for thriving farm animal genome activity. Currently mined in rapidly growing data warehouses, completed genomes of chicken, fish and cows are available but are largely stored in decentralized data repositories. In this paper, we provide an informatics primer on farm animal bioinformatics and genome project resources which drive attention to the most recent advances in the field. We hope to provide individuals in biotechnology and in the farming industry with information on resources and updates concerning farm animal genome projects.

PRELUDE
The omics and bioinformatics
Genomics is the scientific study of structure, function and interrelationships of both individual genes and the genome in its entirety (1). The field has evolved from identifying short nucleotide strings of DNA to the sequencing of an organism’s complete genome. Current progress in genomics research has facilitated comprehensive mapping of the building blocks of biology (2). Ultimately, researchers hope to gain mastery over the fundamental description of cellular function at the DNA level (3). This would encompass gene regulation, in which proteins often regulate their own production or that of other proteins in a complex web of interactions. Databases can be developed to provide solutions to problems that people encounter when dealing with massive amounts of data. However, the abundance and growing complexity of the data tools and resources required for analysis are overwhelming (http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html). The same is true of researchers focusing on the study of function and structure of proteins and protein–protein interactions.

Proteomics, or the science of protein structure and function, is now a hot spot in biomedicine. A key component to the next revolution in the ‘post-genomic’ era will be the increasingly widespread use of protein structure in rational experimental design. New computational methodologies now yield structure models that are, in many cases, quantitatively comparable to crystal structures, at a fraction of the cost (4). The technical challenge is the complete coverage of physicochemical properties for thousands of proteins (5). Thus, by analytically investigating genes and proteins, researchers have developed the umbrella study of bioinformatics, the science of analyzing biological data using cutting-edge computing techniques.

Bioinformatics deals with methods for storing, retrieving and analyzing biological data, such as nucleic acid (DNA/RNA) and protein sequences, structures, functions, pathways and genetic interactions: the computational management of all kinds of biological information (http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html). Rather than merely a mixture of computer science, data management and genome sciences, Bioinformatics now encompasses both conceptual and practical tools for the propagation, generation, processing and understanding of scientific ideas and biological information (6).

DNA and protein sequences data banks
The availability of many complete genome sequences from different species can bring insight into the function of conserved non-coding regions of DNA sequence (7). Organization of the data into coding (genes) and non-coding sequences, in addition to organizing these data into databases for ‘DNA databases’ and proteins (Table 1), is central for bioinformatics (6,8,9). Parsing the sequence of DNA or protein is a first step, followed by curation then analysis and sometimes re-curation based on the analysis. The presence of these databases and derivative search engines gave rise to programs, such as FASTA (10) and PSI-BLAST (11), which are DNA and protein sequence analysis tools, respectively (12). These tools facilitate searches for sequences that resemble one another and homologous relationship inferences. Utilization
of similarity search programs, which operate on sequences, has helped in augmenting annotated databases, which house information about sequence domains. Information recording, retrieval and cataloging continue to advance further database capabilities (13). Collaborating research centers, particularly in the US, Japan and the UK, have been actively collecting sequence data and making it accessible to public (14) (Table 1).

### Farm animals: an unexploited gold mine for biotech

Farm animals are quite valuable as resources, often notable as models for pathology and physiology studies. Magnussen (15) discusses how a variety of these farm animal models are used. The reproductive physiology of farm animals is more similar to humans than that of rodents because farm animals have longer gestation and pre-pubertal periods than mice (16). Specific farm animal physiology, such as the digestive system of the pig is similar to that of humans (17). These attributes of farm animals reveal that they are an unparalleled resource for research replicating human physiological function (15).

For decades, breeders have altered the genomes of farm animals by first searching for desired phenotypic traits and then selecting for superior animals to continue their lineage into the next generation (18). This genomic work has already facilitated a reduction in genetic disorders in farm animals, as many disease carriers are removed from breeding populations by purifying selection (19). By studying diverse phenotypes...
over time, researchers can now monitor mutations that occur as wild species become domesticated (20). Farm animal food safety will remain a concern for some time; however, advancements such as the discovery of *Escherichia coli* resistant genes in the pig (21) can mediate most of the problems. Moreover, resources devoted to investigating the genomes of farm animals can bring eventual economic benefits. For example, isolation of DNA from animal tissue can be used as an inexpensive method for tracking the origin of a meat sample, providing the recipient with the quality assurance of that food (22).

### 2005: the year of the chicken

There has been significant interest in the first complete analysis of the draft genome sequence of the chicken. This sequence has given rise to chicken genome array chips and a number of web based mapping tools. The great importance of the chicken as a scientific resource can be seen from the research on avian leucosis virus, which has led to developments in the areas of proviral insertion-mediated oncogenesis and vertebrate viral-host interactions (23,24). Still, many chicken lines are being lost due to facility downsizing and closings (http://www.grcp.ucdavis.edu/publications/doc20/full.pdf). The release of the first draft of the chicken genome in March 2004 spawned the boom in chicken genomic research (25) (Table 2). From an evolutionary standpoint, investigation of the chicken genome will provide significant information needed to understand the vertebrate genome evolution, since the chicken is between the mammal and fish on the evolutionary tree (26). Furthermore, the chicken remains significant as a food animal which comprises 41% of the meat produced in the world and serves as a reliable model for the study of diseases and developmental biology (26). With this sequenced genome, chicken breeders will have a framework for investigating polymorphisms of informative quantitative traits to continue their directed evolution of these species. The chicken genome is also effective as a comparative genomic tool that sheds additional light on various aspects of our own genome (27). In addition, complementary DNA microarrays for the chicken have already been produced for the study of metabolic and other systemic processes (27,28).

### The swine sequence

The sequencing of the pig genome generated an invaluable resource for advancements in enzymology, reproduction, endocrinology, nutrition and biochemistry research (29,30). Since pigs are evolutionarily distinct to both humans and rodents, but have co-evolved with these species, the diversity of selected phenotypes make the pig a useful model for the study of genetic and environmental interactions with polygenic traits (31). The sequencing of the pig genome is also instrumental in the improvement of human health. Clinical studies in areas such as infectious disease, organ transplantation, physiology, metabolic disease, pharmacology, obesity and cardiovascular disease have used pig models (32) (Table 3). In the near future, the sequencing of the porcine genome will allow gene markers for specific diseases to be identified, assisting breeders in generating pig stocks resistant to infectious diseases (33). Furthermore, as researchers investigate the swine genome and isolate genes that may impact the economics of breeding, members of the commercial pig industry are able to use this information to garner benefits (30).

### The *Bos taurus* genome

The mammalian order Cetartiodactyla (possessing *B.taurus* or cattle) is of great interest since it represents a group of eutherian mammals phylogenetically distant from primates (34,35). Working with the cow species, *B.taurus*, is significant because the cow is such an economically important animal. This form of livestock makes up the beef and milk production industry, which is one of the largest industries in the United States. The identification of numerous single-nucleotide polymorphisms (SNPs) makes it possible for geneticists to find associations between certain genes and cow traits that will eventually lead to the production of superior-quality beef (36). After the completion of the September 2004 *B.taurus* draft assembly (Table 4), this genome has functioned as a vehicle for studies on non-primate and non-rodent genomes as well as in comparative genomics (37). Similar to the pig, the cow also serves as a good animal model for obesity, infectious diseases and female health (34,38).

### Sheep (*Ovis aries*) genome project

Bioinformatics researchers from New Zealand, US, UK and Australia have come together to work on the sheep genome map. The focal point of interest in sheep is based on the quest to maximize sheep meat and cotton wool production. This sector of the corporate farming industry is so intent on this biotechnology project that AWI, Meat and Livestock Australia and nine other partners are investing $50 million into the Sheep Genome Project to ensure its completion (http://www.wool.com.au/LivePage.aspx?pageId=116). As a resource in biological science, researchers have mapped a subset of genes that have also been mapped in humans and mice (39). These studies have revealed the existence of mutations that yield phenotypes unique to the sheep, demonstrating that genetic analysis of the sheep can enhance our knowledge of biological pathways in other mammalian species (39,40). Table 5 provides information on current genome informatics resources available on the Internet related to sheep.

### The fish genome project

Fish are one of the most studied organisms. Researchers have investigated the genome of zebrafish, medaka, pufferfish (*Fugu rubripes* and *Tetraodon nigroviridis*) and sticklebacks. The zebrafish genome has attracted the attention of various pharmaceutical and biotech companies owing to the ease with which scientists can use this fish to study the gene function (http://www.wellcome.ac.uk/en/genome/geneandbody/hg05f006.html). The number of databases and informatics resources related to fish genomic study on the Internet is consistently growing (Table 6). Japanese researchers are interested in medakas from an evolutionary standpoint (40). Other fish genome sequences, such as the pufferfish genome are similar to the Fugu genome in size and are ~7.5 times smaller than the human genome. This has led many to believe that the genes conserved between these two species would reveal the minimal number of genes required for a vertebrate organism (41). Further investigation of the Fugu genome supported this
hypothesis and showed the ability of the Fugu genome to aid in the study of vertebrate functional non-coding sequences (41). Other studies of fish have focused on the stickleback, for its variable body shapes, ecology or its behavior (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/SticklebackSEQ.pdf). In general, fish genomic work is of interest to the commercial fish farming community. Based on transgenic studies, antifreeze protein and fish

| Database                        | Site content                                                                 | URL                                      |
|---------------------------------|------------------------------------------------------------------------------|------------------------------------------|
| Poultry and avian research      | A detailed stock and curator listing of available avian research stocks are provided | http://animalscience.ucdavis.edu/AvianResources/index.htm |
| Chicken variation database      | An integrated information system for storage, retrieval, visualization and analysis of chicken variation data | http://chicken.genomics.org.cn/index.jsp |
| Chicken genome browser          | The genome browser zooms and scrolls over chromosomes, showing the work of annotators worldwide | http://genome.ucsc.edu/cgi-bin/hgGateway?org=Chicken&db=0&hsid=30948908 |
| NCBI chicken genome resources   | Provides information on chicken-related resources from NCBI and the chicken research community | http://www.ncbi.nlm.nih.gov/genome/guide/chicken/ |
| Wellcome trust chicken genome browser | This site presents an annotation of the first draft chicken genome assembly | http://www.ensembl.org/Gallus_gallus/ |
| WUGSC chicken genome site       | Possesses helpful information on the chicken genome                          | http://genome.wustl.edu/genome.cgi?GENOME=Gallus%20gallus |
| Chicken genome array            | The GeneChip® chicken genome array is a key research tool for the study of chicken genomics and chicken viral pathogens | http://www.affymetrix.com/products/arrays/specific/chicken.affx |
| ChickCmap                       | ChickCmap allows the alignment of the different available maps in chicken     | http://www.animalsciences.nl/cmap/       |
| ChickFPC                        | Search using a sequence name, gene name, locus or other landmark             | http://www.animalsciences.nl/ChickFPC/   |
| Chicken database                | Online public database browser                                               | http://www.thearkdb.org/browser?species=chicken |
| AvianNet                        | A portal to information on the chicken genome and chicken biology            | http://www.chicken-genome.org/information/about.html |
| ChickEst                        | Provides access to 339,314 *Gallus gallus* ESTs                              | http://chick.umist.ac.uk/                |
| US poultry genome project       | Supported national animal genome research programme to serve the poultry genome mapping community | http://poultry.mph.msue.edu/index.html |
| ChickAce                        | Mainly intended to store mapping information in chicken                      | https://acedb.asg.wur.nl/                |
| University of Delaware ChickEST database | Contains over 40,000 EST sequences from the chicken cDNA libraries in the UD collection | http://www.chickest.udel.edu/ |
| TIGR *G. gallus* gene index     | Integrates research data from international *G. gallus* EST sequencing and gene research projects | http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=g_gallus |
| Bacterial artificial chromosome (BAC)-based physical and genetic map | A genome-wide, BAC-based, integrated genetic and physical map of the chicken genome | http://hbz7.tamu.edu/homelinks/phymap/chicken/chick_home.htm |
| Consensus linkage map of chicken genome | Wageningen University chicken consensus linkage map | http://www.zod.wau.nl/abg/his/research/molecular/intro.html |
| Chick RH server                 | RH mapping on INRA chicken radiation hybrid panel                            | http://chickrh.toulouse.inra.fr/         |
| Chicken-IMAGE                   | Improvement of chicken immunity resistance to disease based upon analysis of genome | http://www.vjf.cnrs.fr/image/chicken/     |
| Chickmap                        | Aim to construct an integrated genetic and physical map of the chicken genome | http://www.projects.roslin.ac.uk/chickmap/about_chickmap.html |
| Chicken genome array            | Key research tool for the study of chicken genomics and chicken viral pathogens | http://www.servicecs.com/products/       |
| Chicken genome browser          | Ensembl home ensembl chicken exportview                                      | http://atlas.cnio.es/Gallus_gallus/exportview |
| Chicken chromosome linkage map  | Provides comparative mapping information                                      | http://poultry.mph.msue.edu/resources/Resources.htm#Chicken%20Chromosome%20Linkage%20Map |
| Chicken BAC library             | A BAC library of the chicken genome has been constructed                     | http://www.zod.wau.nl/vf/research/chicken/body_bac_library.html |
| GEISHA                          | *G. gallus* (chicken) EST and in situ hybridization analysis database         | http://geisha.biosci.arizona.edu/        |
| NCBI chicken genome map viewer  | *G. gallus* (chicken) genome                                                 | http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9031 |
raised, it is clear that this type of genetic engineering is at an
debate. Despite the controversy that the ‘shining rabbit’ and the ‘cotton rabbit’
also for future medical experimentation models. Despite the

The French National Institute of Agronomic Research
for research have quickly arisen in the research community.
and about the practice of manipulating genes in animals
Alba (Amanda onion
inserted into a fertilized rabbit egg cell that eventually grew
make its glowing properties twice as powerful. This gene,
for commercial reasons. French scientists created Alba using
developed as glowing mutants that shine under special light
models, such as the Alba or the ‘mighty lighting rabbit’, were
with other animal models, such as the mouse, rat and fruit
Little genomics attention is paid to rabbits in comparison
rates (42).

Table 3. Genomic and proteomic internet sites devoted to the study of the swine genome

| Database                          | Site content                                                                 | URL                                                                 |
|-----------------------------------|------------------------------------------------------------------------------|----------------------------------------------------------------------|
| NCBI pig genome resources         | Brings together information on porcine-related resources from NCBI          | http://www.ncbi.nlm.nih.gov/projects/genome/guide/pig/                |
| Pig EST database                  | Pig EST database accommodates 98,988 pig ESTs, which were obtained from various sources | http://pigest.genome.iastate.edu/                                    |
| Genomic targets for comparative sequencing | Genomic targets of the pig                                                  | http://www.nsc.nig.gov/projects/zooseq/comp_seq_org_targets.cgi?org=pig |
| Pig QTL database                  | The database makes it possible to compare on pig chromosomes the most feasible location for a gene responsible for quantitative trait important to pig production | http://www.animalgenome.org/QTLDb/                                   |
| Swine genome maps                  | Swine gene mapping information                                              | http://www.marc.usda.gov/genome/                                     |
| Porcine genome physical mapping project | A physical map of the porcine genome                                      | http://www.sanger.ac.uk/Projects/S_scrofa/                           |
| Cytogenetic map of the pig        | List of genes mapped on pig cytogenetic map                                 | http://www.toulouse.inra.fr/lgc/pig/cyto/cyto.htm                    |
| PigBase                           | PigBase is a computer database that includes information on papers published about gene mapping in the pig | http://www.thearkdb.org/browser?species=pig                         |
| The US pig genome project         | US pig gene mapping                                                         | http://www.animalgenome.org/pigs/                                   |
| Pig genome mapping                | Listing of pig genome databases                                             | http://projects.roslin.ac.uk/pigmap/                                |
| Blast pig genome                  | Blast pig sequences                                                         | http://www.ncbi.nlm.nih.gov/genome/seq/SscBlast.html                  |
| Pig genome sequencing project     | Porcine genome project                                                      | http://www.piggenome.dk/                                            |
| Monsanto swine genome project     | Produce genetic information from cDNA libraries made from swine tissue     | http://asmcswine.missouri.edu/Description.html                        |
| Animal genome research program    | Contains pig EST data explorer                                              | http://animal.dna.affrc.go.jp/agp/index.html                         |
| ArKB-pig mapping database         | A genome database of the pig                                                | http://bioresearch.ac.uk/browse/mesh/D013552.html                    |
| Animal genome database            | Contains available pig cDNA clones                                          | http://w54.iai.affrc.go.jp/mbase.html                                |
| ARK-genomics                      | Collaborative center for functional genomics in farm animals                | http://www.ark-genomics.org/index.php                                |

growth hormone have been introduced into fish genomes,
creating fish with greater cold tolerance and faster growth rates (42).

The forgotten rabbit

Little genomics attention is paid to rabbits in comparison with other animal models, such as the mouse, rat and fruit fly in the pre- and post-genomic era. However, experimental models, such as the Alba or the ‘mighty lighting rabbit’, were developed as glowing mutants that shine under special light for commercial reasons. French scientists created Alba using a process called zygote microinjection. In this process, the scientists plucked a fluorescent protein from the fluorescent jellyfish *Aequorea victoria*. Then, they modified the gene to make its glowing properties twice as powerful. This gene, called EGFP (enhanced green fluorescent gene, was then inserted into a fertilized rabbit egg cell that eventually grew into Alba (Amanda onion). Debates about the project itself and about the practice of manipulating genes in animals for research have quickly arisen in the research community. The French National Institute of Agronomic Research hesitated to release these rabbits owing to protests over its development.

Chinese scientists have placed rabbit genes in cotton plants, producing cotton fibers as bright and soft as rabbit hair but stronger and warmer. This indicates not only the rise of the rabbit in experimental models for genetics engineering, but also for future medical experimentation models. Despite the controversy that the ‘shining rabbit’ and the ‘cotton rabbit’ raised, it is clear that this type of genetic engineering is at an early stage. Currently available sources of information about the rabbit on the web can be seen in Table 7.

Farm animal genomics: current statistics

As researchers delve into the composition of farm animal genome sequences, new functional and biological data emerge. Table 8 is compiled from the Ensembl database (http://www.ensembl.org/) and illustrates a summary of current analysis on coding regions within genomes for selected farm animals. For example, in the cow genome, ~239,000 exon regions and just over 29,000 transcripts have been tallied from among the 565 million bp found in the genome.

One of the major sources of information on farm animal genomics is the ArKB (43), which is available through the Roslin Institute (UK) and Texas A&M University (USA). The ArKB provides detailed genomic mapping data on sheep, chicken, cow and pig genomes (44), including data on PCR primers, genetic linkage map assignments of specific loci and markers, and cytogenetic map assignments. Information about farm animals in this database is displayed in Table 9 and provides a snapshot of current resources available for each genome. Quantitative functional information, such as the number of clones, microsatellites and associated mapping assignments, can lend insight into the complexity of the models available. For example, the number of primers tabulated in the swine genome currently far exceeds that known for the cow, indicating more varieties of genes are available for study in the swine. Much of this information remains at an early stage and with increasing experimentation, compilation and analysis, will be refined.
### Table 4. Cow genomics and proteomics resources on the Internet

| Database                        | Site content                                                                 | URL                                                                 |
|---------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------|
| NCBI cow genome resources       | Brings together information on cow-related resources from NCBI              | http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=10708 |
| Bovine genome sequencing        | Aim to produce a draft assembly of the genomic sequence of cow *B. taurus*    | http://genome.gov/12512284                                           |
| Ensembl genome browser          | Produces and maintains automatic annotation on selected eukaryotic genomes  | http://www.ensembl.org/Bos_taurus/index.html                         |
| Blast cow sequence              | Perform sequence comparisons for cow-specific sequences                     | http://www.ncbi.nlm.nih.gov/genome/seq/BtaBlast.html                 |
| CowBase                         | The cow genome database                                                     | http://www.thearkdb.org/browser?species=cow                         |
| Baylor bovine genome project    | Working to sequence *B. taurus* using a whole genome shotgun approach       | http://www.hgsc.bcm.tmc.edu/projects/bovine/                         |
| Cattle genome database          | The CGD is part of an international collaboration to map the bovine genome  | http://www.cgd.csiro.au/                                             |
| The interactive bovine *in silico* SNP (IBISS) database | IBISS creates an *in silico* SNP database by harnessing the vast number of bovine EST sequences | http://www.livestockgenomics.csiro.au/ibiss/                        |
| The US cattle genome project    | National cattle genome coordination program                                 | http://www.genome.iastate.edu/cattle/community/                     |
| Bovmap database                 | Cattle and cattle on human chromosomes physical mapping                     | http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl                   |
| Bovine QTL viewer               | Contains all available public domain bovine QTL data for both dairy and beef traits | http://bovineqtl.tamu.edu/                                           |
| Dairy cattle QTL map database   | Combined QTL map of dairy cattle traits                                      | http://www.vetsci.usyd.edu.au/reprogen/QTLMap/                      |
| TIGR cattle gene index          | Integrates research data from international *B. taurus* EST sequencing and gene research projects | http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=cattle      |
| NCBI SNP database               | Cattle single nucleotide polymorphism database                              | http://www.ncbi.nlm.nih.gov/SNP/snp_batchSearch.cgi?org=9913&type=SNP |
| Beef genomics initiative        | Uses structural and functional genomics approaches to study the bovine genome | http://www.afns.ualberta.ca/Hosted/Bovine%20Genomics/Index.asp?Research |
| Bovine genome browser           | Aligns all features to the bovine genome scaffolds from college bovine genome project | http://www.livestockgenomics.csiro.au/cattle.shtml                   |
| Cattle EST gene family database | EST family database through Gbrowse                                         | http://www.genome.iastate.edu/cattle/maps/db.html                   |
| Unigene: *B. taurus*             | UniGene is an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters | http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?taxid=9913             |
| Cow (*B. taurus*) genome browser| UCSC genome browser created by the bovine bioinformatics group of UC Santa Cruz | http://genome.ucsc.edu/cgi-bin/hgGateway?clade=vertebrate&org=Cow    |

### Table 5. Internet informatics resources pertaining to the sheep genome

| Database                                      | Site content                                                                 | URL                                                                 |
|-----------------------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------|
| UK sheep genome mapping project               | The genome database of the sheep                                           | http://www.projects.roslin.ac.uk/sheepmap/front.html                 |
| US sheep genome mapping project               | Public database browser section of the sheep database                       | http://www.thearkdb.org/browser?species=sheep                       |
| Sheep genome resources                        | Genome effort will generate an important resource for gene discovery, affecting health and biology and the growing sheep genome database | http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?taxid=9913             |
| Australian sheep gene mapping site            |                                                                            | http://rubens.its.unimelb.edu.au/%7Ejillm/jill.htm                  |
| Genomic targets for comparative sequencing     |                                                                            | http://www.nisc.nih.gov/projects/zooseq/comp_seq_org_targets.cgi?org=sheep |
| Mendelian inheritance in sheep Blast sheep sequences | A compiled annotated trait/locus list Blast your sequence against sheep sequences | http://www.angis.org.au/Databases/BIRX/mis/                        |
| Only cytogenetics of animals (OCOA)            |                                                                            | http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?taxid=9913             |
|                                              |                                                                            | http://www.angis.org.au/Databases/BIRX/ocoa/                        |
Expression sequence tags

Expression sequence tags (ESTs) are small pieces of cDNA sequence (usually 200–500 nt long), which are useful as markers for a desired portion of RNA and DNA that can be used for gene identification and positional mapping within a genome (http://www.ncbi.nlm.nih.gov/About/primer/est.html). The NCBI provides the most comprehensive EST database for many farm animals. Tabulation of chicken EST information, including quantities of ESTs and their tissue sources can be found in a table provided in the text.

Table 6. Fish genomic resources on the internet

| Database | Site content | URL |
|----------|--------------|-----|
| Queen mary University of London Genosope (the french national sequencing center) | The fugu genomics project Tetraodon genome browser | http://fugu.biology.qmul.ac.uk/ http://www.genoscope.cns.fr/externe/tetranew/ |
| Institute of molecular and cell biology Joint genome institute Broad institute | Fugu genome project blast Fugu genome browser Tetraodon nigroviridis database | http://www.fugu-sg.org/BLAST/Blast2.htm http://genome.jgi-psf.org/fugu6/fugu6.home.html http://www.broad.mit.edu/annotation/tetraodon/ |
| NCBI entrez genome NCBI entrez genome Ensembl tetraodon Ensembl fugu UCSC genome browser UCSC genome browser Roslin institute Zebradorn gene FishBase | Blast the fugu genome Blast zebradorn sequences Tetraodon genome browser Fugu genome browser Tetraodon genome browser gateway Fugu genome browser gateway Salmon public database browser Characterization of mapped zebradorn genes. Genomic mapping Description of fish species. Zoology | http://www.ncbi.nlm.nih.gov/BLAST/Genome/fugu.html http://www.ncbi.nlm.nih.gov/BLAST/Genome/fugu.html http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ http://genome.ucsc.edu/cgi-bin/hgGateway?org=tetraodon http://genome.ucsc.edu/cgi-bin/hgGateway?org= termedeb&clade=vertebrate&org= Fugu&db=0&hgsid=42757856 http://www.thearkdb.org/browser?species=salmon http://zfin.org/cgi-bin/webdriver?Mval=aa-ZDB_home.apg http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ http://www.ensembl.org/Tetraodon_nigroviridis/ http://www.ensembl.org/FuguRubripes/ |

Table 7. Rabbit informatics databases

| Database | Site content | URL |
|----------|--------------|-----|
| HRS articles | Information/articles pertaining to many rabbit diseases written by veterinarians | http://www.hrschicago.org/articleslay.html |
| INRA RabbitMap database | Mapping the rabbit genome. The site contains data about loci, genes, microsatellites, polymorphisms in rabbit | http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/intro2.pl?BASE=rabbit |
| Online mendelian inheritance in animals (OMIA) | Genes and phenes that have been documented in a wide range of animal species (other than human, rat and mouse). Complementary to human’s OMIM | http://omia.angis.org.au/ |
| PIR non-redundant reference protein database | Provide an integrated public resource of functional annotation of protein data | http://www-nbrf.georgetown.edu/cgi-bin/nfspecies.pl |

Table 8. Ensemble (gene build July 2005) and NCBI statistics of available farm animal genomes

| No. | Information source/type | Chicken | Cow | Pig | Sheep | Fish (zebrafish) |
|-----|-------------------------|---------|-----|-----|-------|-----------------|
| 1   | eGene predictions       | 17 784  | 22 013 | NA  | NA   | 22 877          |
| 2   | Genscan predictions    | 77 600  | 103 597 | NA  | NA   | 49 697          |
| 3   | egene exons            | 185 326 | 239 889 | NA  | NA   | 231 799         |
| 4   | egene transcripts      | 28 491  | 29 363 | NA  | NA   | 32 143          |
| 5   | Base pairs             | 1 054 180 845 | 565 382 643 | NA  | 1 688 467 974 | |
| NCBI (entrez 2005 records) | | | | | |
| 1   | Nucleotide            | 923 899 | 1 961 869 | 628 750 | 28 716 | 912 447 |
| 2   | Protein               | 29 634  | 51 726 | 883 5 | 3160 | 47 458 |
| 3   | UniSTS                | 2936    | 13 320 | 7851 | 2408 | 27 616 |
| 4   | Gene                  | 18 449  | 39 815 | 2388 | 625 | 18 429 |
| 5   | HomoloGene            | 14 433  | 12 909 | 10 795 | 1016 | 11 026 |
| 6   | EST                    | 554 084 | 624 652 | 461 891 | 10 960 | 651 991 |

*The ESTs in this table are based on the dbEST release 082605-August 26, 2005 (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) BLAST alignment data, associated Human UniGene and RH maps (EBI GenomeMap’99 and NCBI GeneMap’99), pig–human comparative map information and mapped pig genes in the pig genome database (PigBase).
Markers

| No. | Element       | Chicken | Pig | Cow | Sheep |
|-----|---------------|---------|-----|-----|-------|
| 1   | Loci          | 2530    | 4081| 2725| 2030  |
| 2   | Designated genes | 765    | 1588| 746 | 543   |
| 3   | 5.3 Sine-PCR   | 47     | 18  | 146 | 3     |
| 4   | Microsatellite | 1277   | 1673| 1219| 2257  |
| 5   | RFLP          | 255    | 258 | 250 |       |
| 6   | Clones        | 316    | 602 | 1   | 206   |

Map assignments

|   | Linkage | 3400 | 5141 | 2209 | 14909 |
|---|---------|------|------|------|-------|
|   | Cytogenetic | 307  | 1927 | 659  | 843   |
|   | Total    | 3707 | 7068 | 2868 | 15752 |

Table 9. Current status summary provided by the ArkDB

be found in the BBSRC ChickEST Database (http://chick.umist.ac.uk/) (45,46). ChickEST provides access to ~339,000 Gallus ESTs, generated from 21 different embryonic and adult tissues. Cow EST information is available on ArkDB, Ensembl (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9031) and the NCBI Cow genome EST. A collection of pig ESTs is also available through the pig EST database (http://pigest.genome.iastate.edu/index.html) in two versions. The first version contains ~14,000 EST sequences obtained from porcine whole embryo, term placenta, anterior pituitary, hypothalamus and ovary. The second, newer version of the database accommodates 99,000 pig ESTs. Pig sequence homology comparisons can be readily studied on the Iowa State University’s website. However, database updates were carried out only until January 2002. Among farm animals, there is sparse information about rabbit ESTs. NCBI does not provide comprehensive EST information, with virtually no physical maps on the rabbit. However, NCBI is still the central source for information about rabbit ESTs, mining 2275 GenBank rabbit ESTs (dbEST release 082605).

Physical mapping

Genome maps in livestock species have been under development for the last decade. Mapping, which involves understanding the relative distances between genes on a chromosome, has been performed for the many farm animals. The NCBI provide comprehensive mapping information for many farm animal species. Mapping information is available at http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9031, with specific animal links available by substituting the ‘?’ sign in the previous URL with the species’ taxonomic number, such as 9031 for the chicken, 9913 for the cow, 9823 for the pig, 7955 for the zebrafish and 9940m for the sheep. For example, the chicken mapping site can be found at NCBI MapView website (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9031). The US Livestock Species Genome Projects (http://www.animalgenome.org/) suggest using ArkDB (mentioned above), for mapping information on farm animals. The Laboratory of Biochemistry and Cytogenetics (Jouy-in-Josas, France) at the National Institute for Agricultural Research (INRA) provides the rabbit ‘Lepine’ genome mapping database (http://dga.jouy.inra.fr/cgi-bin/lgbc/main.pl?BASE=rabbit). The database provides access to information about rabbit genomic information which includes so far 410 loci with 1327 homologue loci on 129 species, 21 polymorphisms, 314 genes and their physical location on chromosomes. Other information about primers, probes and enzymes are also included. Genes in the RabbitMap database are hyperlinked to a separate page that provides locus information. Physical mapping of the zebrafish is available through the Ensemble database at http://www.ensembl.org/Danio rerio/index.html. The Ensemble database also provides a comprehensive map viewer for zebrafish and the chicken (http://www.ensembl.org/Gallus_gallus/index.html). High-resolution maps for sheep chromosomes are available at http://www.ncbi.nlm.nih.gov/genome/guide/sheep/index.html.

Quantitative trait loci and genetic linkage

One of the primary challenges in modern biology is the understanding of the genetic basis of phenomic diversity within and among species (20). The foundation for this diversity lies in genetic governance of both how traits are expressed and the associated linkage maps. Genes mapped in some farm animals serve as ‘anchors’ across the comparative maps of other species (39). Quantitative trait loci (QTLs) play a major role in farm animals and the related biotechnology industry, as they can further the identification of traits related to meat and milk production. A number of studies have been conducted to detect QTLs that can be used for determining gene variances (48,49). The third generation of sheep linkage map (50) contains 1062 loci (941 anonymous loci and 121 genes) and is a compilation of genotype data generated by 15 laboratories using the IMF population (http://rubens.its.unimelb.edu.au/~jillm/jill.htm). The ArkDB provides current summaries of linkage and cytogenetic map assignments, polymorphic marker details, PCR primers and two point linkage data (see Table 9). The ArkDB also is a major source for mapped sheep loci, with SheepBase containing almost 1500 loci. The Roslin Institute has developed the ‘resSpecies’ (http://www.reSpecies.org/) database to study genetic linkage maps, QTLs, alleles and other markers (44,51). Bovine QTL databases are available from several sources (52,53). Various traits, such as milk yield and composition, are available along with QTL dot maps, which provide varied positional and statistical information. The pig QTL (http://www.animalgenome.org/QTLD/db/) includes two release versions from NCBI in May 2004 and from NAGRP in December 2004. The pig QTL database, or pig QTDb, contains all published pig QTL data from the past decade. The user can locate genes responsible for quantitative traits central to pig production. To date, 791 QTLs from 73 publications, representing 219 different traits are incorporated into the database in the first release and 1129 QTLs, from 86 publications representing 235 different traits in the second release. The project also provides maps at http://www.animalgenome.org/cgi-bin/QTDb/viewmap and links to the NCBI pig EST Map database (http://pigest.genome.iastate.edu/index.html).

FARM ANIMAL GENOMICS: THE LEGAL SIDE

Genome protections: lessons from the human genome

Similar to other areas of scientific research, the field of farm animal genomics is ingrained with various issues requiring
ethical consideration. Although a number of gene-based technological advances, such as the prediction of illnesses or adverse drug response are beneficial for commercial interests (54), many look to protect how genetic information is used. The corollary in human genomics is the recent Genetic Information Nondiscrimination Act of 2005 (http://www.genome.gov/PolicyEthics/). This bill, which is pending in the US Congress, prohibits among other things ‘a health insurance issuer from: adjusting premiums on the basis of genetic information’ (http://thomas.loc.gov/cgi-bin/bdquery/z?d109:SN00306r.@@@L&summ2=m&). Concerns over such actions reveal the understanding by the US government that the use of human genomic data must be closely regulated in order to hinder any commercial abuses. Although genetic discrimination and health insurance might not be issues relevant to farm animal genomics, a great need exists to establish a similar legal framework for farm animals in order to ensure the proper use of farm animal genomic information.

### Genomes for transgenic animals

The federal government’s role as a legal authority for protecting the environment from scientifically introduced transgenic organisms has not been clearly outlined (55). This underscores the need for greater governmental attention to address legal problems that will arise as animal biotechnology continues to grow. Many argue that it is not only a moral imperative to respect the intrinsic sentence of animals but also a legal one (http://www.foodethicscouncil.org/library/consultations/BBSRCgenomics.pdf). The European Union’s Treaty of Amsterdam states that animals must be treated humanely, as sentient beings, which some argue implies respect for their intrinsic nature and protection against such infringements of that respect in studies such as transgenic animal experiments (http://www.eurotreaties.com/amsterdambrief.html). Furthermore, US food regulatory agencies, such as the United States Department of Agriculture and the Food Safety and Inspection Service, charged with the duty of making sure that transgenic animals anticipated for human consumption are correctly labeled and wholesome, have already moved forward to institute policies regulating the slaughtering of non-transgenic animals created from transgenic animal experiments (56).

### Ethical reservations of farm animal genomic study

As farm animal species continue to be sequenced (one of the latest being the cow, B.taurus), farming companies are using this information to perform genetic profiles of traditionally bred animals and to genetically engineer or clone other animals (57). Although many farmers may enjoy access to greater genetic resources, animal rights organizations oppose the use of animals such as transgenic goats that produce silk proteins used to make Biosteel fibers because of the allegedly inhumane treatment of animals during these studies (58). Proponents of such experiments point to the ability of transgenic mouse mammary glands to assemble and secrete recombinant human fibrinogen (59). These studies raise many concerns about research ethics and species integrity (60,61). The potential of farm animal genomics

Farm animal genomic studies continue to attract audiences excited by the multitude of applications (62). The meat industry can now use cow and chicken genomic data to confirm the quality of meat products. For example, meat producers can now confirm the parentage of an Angus cattle breed by performing a genetic blood test or attempt to identify the SNPs associated with high-quality beef (57). Other companies are using genomic information to determine disease-resistant genes in shrimp and then are selectively mating the shrimp that carry them in order to create disease resistant strains (57). In the healthcare arena, farm animal genomic work will aid in enterprises such as xenotransplantation (the transfer of animal tissues or organs into humans). Though animal organs may be used someday to satiate organ donor shortage, genomic work in this area is still in its early stage (63). Many of the immediate practical applications of farm animal genomics show potential for growth in this field.

### ACKNOWLEDGEMENTS

We thank Professor William Dynan, Medical College of Georgia for helpful discussions and careful editing of the manuscript. We thank Ms J. Wade and T. Scalise for assistance during the processing of this manuscript. Special thanks goes to the reviewers for their valuable comments and suggestions. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

### REFERENCES

1. Bazer,F.W. and Spencer,T.E. (2005) Reproductive biology in the era of genomics biology. *Theriogenology*, 64, 442–456.
2. Bilello,J.A. (2005) The agony and ecstasy of ‘OMIC’ technologies in drug development. *Curr. Mol. Med.*, 5, 39–52.
3. Chan,E.Y. (2005) Advances in sequencing technology. *Mutat. Res.*, 573, 13–40.
4. Maggio,E.T. and Ramnarayan,K. (2001) Recent developments in computational proteomics. *Trends Biotechnol.*, 19, 266–272.
5. Drake,R.R., Deng,Y., Schweger,E.E. and Gravenstein,S. (2005) Proteomics for biodefense applications: progress and opportunities. *Expert Rev. Proteomics*, 2, 203–213.
6. Rhee,S.Y. (2005) Bioinformatics. Current limitations and insights for the future. *Plant Physiol.*, 138, 569–570.
7. Koyanagi,K.O., Hagiwara,M., Itoh,T., Gojobori,T. and Imanishi,T. (2005) Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene*, 4, 169–176.
8. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimamochi,T. and Tasumi,M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112, 535–542.
9. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, 28, 235–242.
10. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, 85, 2444–2448.
11. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.
59. Butler, S.P., O’Sickey, T.K., Lord, S.T., Lubon, H., Gwazdauskas, F.C. and Velander, W.H. (2004) Secretion of recombinant human fibrinogen by the murine mammary gland. *Transgenic Res.*, **13**, 437–450.

60. Christiansen, S.B. and Sandoe, P. (2000) Bioethics: limits to the interference with life. *Anim. Reprod. Sci.*, **60–61**, 15–29.

61. Smith, T.J. (1994) Commercial exploitation of transgenics. *Biotechnol. Adv.*, **12**, 679–686.

62. Appels, R., Francki, M., Cakir, M. and Bellgard, M. (2004) Looking through genomics: concepts and technologies for plant and animal genomics. *Funct. Integr. Genomics*, **4**, 7–13.

63. Cascalho, M. and Platt, J.L. (2001) Xenotransplantation and other means of organ replacement. *Nature Rev. Immunol.*, **1**, 154–160.