Status of the Cattle Genome Map

D.M. Larkin
Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK

Abstract
During the last 30 years, the cattle genome map has been expanded from 4 genes linked on chromosome X to over 22,000 genes identified in the cattle genome sequence assembly. This progress has been achieved due to numerous projects on linkage and physical mapping of the cattle genome driven by its agricultural and scientific significance. Indeed, the high-resolution mapping and functional analysis of the genome led to the discovery of major quantitative trait loci (QTL) regions and several quantitative trait nucleotides (QTNs), as well as some disease genes in the cow population. In addition, a comparison of the cattle genome to the genomes of other mammals has revealed its unique features gained during the speciation and adaptation. With the development of non-expensive sequencing techniques, the analysis of the cattle genome will shift towards the identification of differences between breeds or individuals within breeds that account for the unique features of each breed. This approach holds promise for the development of effective tools for the marker assistant selection and disease diagnostics in cattle.

In 2009, the cattle genome sequencing and assembly has been completed. The success of this effort was the result of an international collaboration between 6 countries. An assembly of the genome became possible due to numerous projects started in the 1970s to understand the organization of cattle chromosomes [Heuertz and Hors-Cayla, 1978; Womack and Moll, 1986], to perform the microsatellite [Barendse et al., 1997] and gene mapping [Itoh et al., 2003], and to construct high-resolution physical [Everts-van der Wind et al., 2005; Snelling et al., 2007] and linkage maps [Ihara et al., 2004]. With the availability of the genome sequence and accurate assembly it became feasible to perform the analysis of the genome at a level that no one could imagine until very recently [The Bovine Genome Sequencing and Analysis Consortium, 2009].

One of the major drivers of the cattle genome studies is an attempt to understand the genetic nature of the quantitative traits and diseases affecting economically important traits, such as milk or meat quality. In the recent years, a significant progress has been achieved in this area leading to the identification of several quantitative trait nucleotides (QTNs) that contribute to such traits in cattle [Grisart et al., 2002; Cohen-Zinder et al., 2005]. The availability of the whole-genome annotation in conjunction with non-expensive sequencing and genotyping
techniques opens a new exiting opportunity for the identification and genotyping of all single-nucleotide mutations in any breed. Together with the genome-wide association studies (GWAS), this leads to the detection of all common and some individual QTNs.

The cattle genome is a great resource for studying mammalian genome evolution. Unique genome features formed by cattle in the course of speciation and adaptation are reflected in its genome by gene mutations, sequence losses, duplications, and repositions due to multiple chromosomal rearrangements that distinguish the cattle genome from other mammalian genomes and a putative mammalian ancestor [Murphy et al., 2005; Larkin et al., 2009; The Bovine Genome Sequencing and Analysis Consortium, 2009; Rijnkels et al., 2010]. However, when compared to other sequenced mammalian genomes, the cattle genome in some chromosomal regions represents an ancestral organization, allowing for the detection of evolutionary events that happened in the course of genome evolution in other species [Murphy et al., 2005].

In this work, we will briefly summarize results of the cattle genome mapping efforts, annotation, and the evolutionary history analysis. We will start with earlier efforts on cattle genome analysis, including early works on somatic cell hybrid mapping, linkage mapping, and later present the advances achieved with the use of radiation hybrid mapping and fingerprint map construction. Together, these efforts have gradually built a basis for understanding cattle QTL (quantitative trait loci) and some mendelian traits, facilitated genome assembly, and made possible functional and evolutionary study of the cattle genome.

**Somatic Cell Hybrid Maps**

The first work using interspecies hybrids of somatic cells for establishing a linkage between cattle genes reported linkage of genes G6PD, PGK, GALA, and HPRT located on cattle chromosome X [Heuertz and Hors-Cayla, 1978]. Later, after the construction of a hamster-cattle somatic cell hybrid panel [Womack and Moll, 1986] containing 31 independent clones, a large number of cattle markers were mapped using this approach. Now, the cattle somatic cell hybrid map contains over 2,700 genes, 1,400 of which were genotyped on the cattle-hamster somatic cell hybrid panel by a Japanese research group [Murphy et al., 2005; Larkin et al., 2009; The Bovine Genome Sequencing and Analysis Consortium, 2009; Rijnkels et al., 2010]. However, when compared to other sequenced mammalian genomes, the cattle genome in some chromosomal regions represents an ancestral organization, allowing for the detection of evolutionary events that happened in the course of genome evolution in other species [Murphy et al., 2005].

In this work, we will briefly summarize results of the cattle genome mapping efforts, annotation, and the evolutionary history analysis. We will start with earlier efforts on cattle genome analysis, including early works on somatic cell hybrid mapping, linkage mapping, and later present the advances achieved with the use of radiation hybrid mapping and fingerprint map construction. Together, these efforts have gradually built a basis for understanding cattle QTL (quantitative trait loci) and some mendelian traits, facilitated genome assembly, and made possible functional and evolutionary study of the cattle genome.

**Linkage Maps**

The first cattle whole-genome linkage maps were published in 1994 [Barendse et al., 1994; Bishop et al., 1994]. These maps contained over 200 and 300 polymorphic markers, respectively, with an average interval between markers >10 cM. Individual linkage groups were assigned to cattle chromosomes using an overlapping set of markers placed on the somatic cell hybrid or cytogenetic physical maps.

Significant progress in cattle linkage mapping has been achieved with the construction of the USDA-MARC linkage map containing ~1,200 polymorphic markers with an average spacing of 2.5 cM and a total genome length of 2,990 cM [Kappes et al., 1997]. This map has provided a basis for the integration of 4 linkage maps [Barendse et al., 1994; Bishop et al., 1994; Georges et al., 1995; Ma et al., 1996] and increased the power of QTL detection. The next significant improvement of the MARC map was an addition of 2,277 microsatellite markers resulting in the generation of a 3,802 microsatellite map with an average interval between markers of 1.4 cM [Ihara et al., 2004]. Later, BES (bacterial artificial chromosome end sequence) and EST-based SNPs were added to the linkage map resulting in a 4,585 marker map [Snelling et al., 2005]. After the cattle genome sequence became available, the cattle linkage maps were enriched for dual allele SNP markers. For example, Arias et al. [2009] published a cattle linkage map containing 6,924 SNP markers from Affymetrix 10K bovine SNP array. The total length of the genome on this high-resolution linkage map is 3,249 cM and average marker spacing is 1 cM. This map was integrated with the
latest build of the cattle genome assembly (Btau 4.0) and provides a direct connection between mapped QTL intervals and actual genome regions. In addition, discrepancies between the genome assembly and linkage map point to the regions that need to be carefully checked for accuracy in the map and assembly.

An important application of moderate-resolution cattle linkage maps was a whole-genome scan for QTL affecting milk production traits in Holstein cattle [Georges et al., 1995; Heyen et al., 1999; Keele et al., 1999]. In addition, 31 chromosomal regions affecting milk production QTLs were detected using Finnish Ayrshire dairy cattle [Viitala et al., 2003]. Several monogenic disorders were identified using the genetic linkage map information and genome-wide association analysis. For example, a missense mutation in the bovine ATP2A1 gene was found to be associated with congenital pseudomyotonia of Chianina cattle and potentially can be used as an animal model of human Brody disease [Drögemüller et al., 2008]. A deletion of the myostatin gene causes the double-muscled phenotype in cattle [Grobet et al., 1997]. The USDA-MARC map was also used to link radiation hybrid linkage groups to cattle chromosomes [Band et al., 2000; Everts-van der Wind et al., 2004].

### Radiation Hybrid Maps

Introduced by Cox et al. [1990] for the human genome high-resolution mapping, radiation hybrid (RH) mapping has been widely used in other species to build high-resolution ordered maps with marker spacing between millions and thousands of base pairs. In cattle, this approach has been first applied by Yang et al. [1998] for the creation of a comparative map of cattle chromosome 19 and human chromosome 17. A 5,000 Rad radiation hybrid panel constructed by Womack et al. [1997] was used to build 3 generations of Illinois-Texas (IL-TX) whole-genome cattle RH maps containing 1,087, 1,913, and 3,484 markers, respectively. The first-generation moderate-resolution IL-TX RH map contained 768 gene markers and 319 microsatellites which were used to link RH linkage groups to the USDA-MARC linkage map [Band et al., 2000]. On this RH map, 638 markers had known orthologs in the human genome, and an estimated comparative coverage of the human genome was ~50%. Regardless of a relatively small number of markers, this map had provided a great resource for predicting positions of cattle BAC end sequences using the ‘comparative mapping by annotation and sequence similarity’ (COMPASS) approach that utilizes comparative maps of cattle and human genomes for the prediction of positions of cattle genomic sequences on cattle chromosomes [Ma et al., 1998; Rebeiz and Lewin, 2000; Larkin et al., 2003].

To generate a higher resolution cattle IL-TX RH map, 870 new markers with predicted positions in gaps of cattle-human comparative coverage were selected for a new mapping project. As the result, 1,913 markers were placed on a new version of the cattle RH map. This provided ~66% comparative coverage between the human and cattle genomes and almost maximum resolution and coverage of the cattle genome that could be achieved using EST markers because of uneven distribution of genes in mammalian genomes. Most of the large gaps in the comparative coverage between the human and cattle genomes were located in gene poor regions. To build next, the 3rd generation whole-genome IL-TX RH map of the cattle genome, a set of genomic markers rather than ESTs, was used [Everts-van der Wind et al., 2005]. This set of markers was generated by the International Cattle BAC Mapping Consortium and represented ~500 bp terminal end sequences of cattle BAC clones [Larkin et al., 2003; Snellling et al., 2007]. These sequences were compared to the human genome sequence and over 3,000 cattle sequences with evenly spaced (~1 Mb apart) unique BlastN hits in human chromosome sequences were placed on the cattle RH map. This map, containing 2,516 ordered BESs, 736 ESTs and 232 microsatellites, was integrated with a physical fingerprint map. The 3rd generation IL-TX cattle RH map had ~91% comparative coverage of the human genome and demonstrated ~93% agreement in the order of markers with the cattle physical fingerprint map containing the same BAC clones (fig. 1). Whereas the focus of IL-TX RH maps was on mapping markers with known orthologs in the human genome, other groups have built RH maps that contained a significant number of microsatellite and SNP markers. For example, a 3,966 marker map built using Roslin 3,000 Rad panel contained 1,072 microsatellite markers, 1,999 genes, BESs, and AFLP markers [Jann et al., 2006]. Another map (SUNbRH, 7000 Rad) contained 5,593 markers of which 3,216 markers were microsatellites and 2,377 were ESTs [Itoh et al., 2005]. An attempt of bioinformatics-based integration of several RH and linkage maps into a single integrated map resource has been made [Snellling et al., 2007]. The resulting composite map contained 17,254 markers and was integrated with the cattle fingerprint map. The latest version of the map was used to assign scaffolds to chromosomes and to establish their order on the Maryland cattle genome assembly (UMD 2.0) [Zimin et al., 2009].
A British Columbia Cancer Research Center (BCCRC) fingerprint physical map of the cattle genome has been built by the International Bovine BAC Mapping Consortium [Snelling et al., 2007] and radiation hybrid [Everts-van der Wind et al., 2005] maps. This map contains 290,797 BAC clones from three cattle BAC libraries generated from different breeds, 200,064 clones from CHORI-240 (Hereford male), 94,848 from RPCI-42 (Holstein male), and 44,948 from TAMBT (Angus male). The initial set of ~13,000 contigs has been merged by FPC software into a set of 655 large contigs, containing 257,914 clones. Comparative data obtained from the alignment of cattle BES with the human genome allowed for selection of probable merge points between contigs which were examined by FPC program using relaxed threshold criteria. The use of

**Fig. 1.** Integration of the cattle chromosome 2 sequence assembly (Btau 4.0, in the middle) with physical fingerprint contig [Snelling et al., 2007] and radiation hybrid [Everts-van der Wind et al., 2005] maps. Lines connect positions of BAC clones from CHORI-240 library placed on all 3 maps. The graphics were generated by AUTOGRAPH web server [Derrien et al., 2007].

**Fingerprint Maps**

A British Columbia Cancer Research Center (BCCRC) fingerprint physical map of the cattle genome has been built by the International Bovine BAC Mapping Consortium [Snelling et al., 2007]. This map contains 290,797 BAC clones from three cattle BAC libraries generated from different breeds, 200,064 clones from CHORI-240 (Hereford male), 94,848 from RPCI-42 (Holstein male), and 44,948 from TAMBT (Angus male). The initial set of ~13,000 contigs has been merged by FPC software into a set of 655 large contigs, containing 257,914 clones. Comparative data obtained from the alignment of cattle BES with the human genome allowed for selection of probable merge points between contigs which were examined by FPC program using relaxed threshold criteria. The use of
a comparative information and a high number of fingerprinted clones allowed for significant decrease in the number and increase in the length of contigs compared to another cattle fingerprint map constructed at INRA. INRA fingerprint map contained 6,615 contigs designed from ~105,000 clones from INRA BAC library and ~27,000 clones from CHORI-240 library [Schibler et al., 2004].

The BCCRC physical map has been integrated with the 3rd generation IL-TX RH map by ~3,000 BEs. These independent maps showed about 93% agreement in the order or clones placed on both maps, indicating a high quality of these resources. Several hundred RH and linkage map markers were assigned to BAC clones from the BCCRC fingerprint map using PCR analysis or in silico comparative mapping against the human genome. This whole-genome contig map provided the highest level of resolution of the cattle genome until the sequence assembly became available. The human-cattle comparative map based on the fingerprint map and BES hits in the human genome has been used for discovery of long regions of amniote chromosomes that are non-randomly maintained during the chromosomal evolution [Larkin et al., 2009]. A skim of ~19,600 overlapping BAC clones from CHORI-240 library from this map has been selected to complement the whole-genome shotgun (WGS) sequence for the genome sequencing [The Bovine Genome Sequencing and Analysis Consortium, 2009].

**Genome Assembly**

The cattle genome assembly (~7.1× Sanger reads) has been generated at Baylor College of Medicine combining BAC sequencing from a Hereford male CHORI-240 library and the whole-genome shotgun sequences of DNA taken from a Hereford dam, L1 Dominette 01449 [The Bovine Genome Sequencing and Analysis Consortium, 2009]. The overlapping set of BAC clones for sequencing has been selected from the BCCRC fingerprint map. Combining BAC and shotgun sequences, a set of scaffolds with N50 of 1.9 Mb was generated. The latest published build of the cattle genome (Btau 4.0) has ~90% of the cattle genome sequence placed on 29 autosomes and chromosome X. An estimated cattle genome size is 2.87 Gb.

Another assembly of the cattle genome was built at the University of Maryland (UMD 2.0) [Zimin et al., 2009]. For this assembly, the same set of raw sequences was used as for the Baylor assembly. However, a different assembly approach and software were used. This allowed for 5% more sequence to be placed on chromosomes compared to Btau 4.0 and resulted in larger N50 size of the sequence contigs. There are significant discrepancies between Btau 4.0 and UMD 2.0 assemblies. An additional effort will be required to resolve them case by case.

An annotation of the cattle genome was performed by the Bovine Genome Sequencing and Analysis Consortium on the Baylor version of the cattle genome assembly. The number of genes in the cattle genome was estimated as >22,000 protein-coding and 496 miRNA genes. The cattle genome contains a large number of ruminant-specific transposable elements that compose 27% of the genome. Some transposable elements from the BOV-B group have intact open reading frames and could still be active in the cattle genome. An analysis of orthologous gene pairs between the human, cattle, dog, mouse, rat, opossum, and platypus genomes has revealed 1,217 genes that could be placental-specific because they are not present in opossum and platypus genomes. About 3.1% of the cattle genome is in segmental duplications. Seventy-six percent of segmental duplications contain complete or partial gene duplications. This set is enriched for genes involved in the interactions of the organism with external environment, e.g. immune proteins and olfactory receptors.

A comparison of the cattle chromosome architecture to the chromosomes of other mammals has revealed 124 evolutionary breakpoint regions in the cattle lineage of which 24 are shared by cattle and pig chromosomes (artiodactyl-specific) and 100 were found only in cattle chromosomes (see fig. 2 for an example chromosome). Nine additional breakpoints were shared by all ferungulate species (cattle, pig, dog) and represent events that originated in the ferungulate ancestor. Interestingly, there is a strong negative correlation between the positions of cattle and artiodactyl-specific breakpoints and some LINE and SINE transposable elements, whereas more recent LINE-L1 and LINE-RTE elements are significantly enriched in these breakpoint regions. Another group of repeats, tRNA\textsuperscript{Glu}-derived SINEs originating in the common ancestor of all artiodactyls, has a higher than expected density in artiodactyl-specific breakpoint regions, but not in the cattle-specific breakpoints. This suggests that evolutionary breakpoints tend to happen in the genome regions with a high density of repetitive elements that are active, and therefore have high sequence similarity between different copies required for a non-allelic homologous recombination. In confirmation of this observation, an analysis of the cattle genome re-
revealed a high density of large (>10 kb) segmental duplications in cattle and artiodactyl-specific breakpoint regions (fig. 2), phenomena previously reported for primate [Murphy et al., 2005] and murine rodent [Armengol et al., 2005] genomes.

**Comparative Studies**

Among the livestock species, cattle has one of the best and most detailed set of comparative maps available, mostly due to cattle economical importance. Whereas somatic cell hybrid maps and cross-species chromosome painting with the human and other species’ DNA probes have provided an important but patched correspondence between the cattle, human, mouse, and pig genomes [Womack and Moll, 1986; Hayes, 1995; Chowdhary et al., 1996; Schmitz et al., 1998], the real breakthrough in cattle comparative studies started with the introduction of high-resolution ordered RH maps [Band et al., 2000; Everts-van der Wind et al., 2004, 2005] and the COMPASS-based approach of marker selection for mapping [Rebeiz and Lewin, 2000; Larkin et al., 2003]. The whole-genome high-resolution ordered comparative maps iden-

---

**Fig. 2.** Comparison of 12 amniote species chromosomes using cattle chromosome 13 as the reference. Grey areas correspond to the blocks of homologous synteny as defined from the Ensembl BioMart orthologous gene set. White areas correspond to the evolutionary breakpoint regions. To the right, arrows indicate positions of cattle, artiodactyl, hominid, primate, and eutherian breakpoint regions. A heatmap shows the density of cattle segmental duplications as defined by The Bovine Genome Sequencing and Analysis Consortium [2009].
tify approximately 201–211 large blocks of homologous syntenic blocks (HSBs) being reorganized in the artiodactyl-specific EBR. The analysis of cattle genome duplications has confirmed the previous observation that evolutionary breakpoint regions in different mammalian species are enriched for this type of sequences. Therefore, one important lesson learned from the cattle-whole genome analysis and comparison to other mammalian genomes is that evolutionary breakpoint regions could play an important role in the adaptation of the genome to the environment because they are reorganizing the genes that contribute to the species’ response to external stimuli, immune response, and other functions related to the lineage-specific features [Larkin et al., 2009; Lemay et al., 2009].

With the introduction of next-generation sequencing platforms, genome sequencing becomes a trivial task. However, de novo assembly of sequences generated from a large number of short or even pair-end reads is still a difficult and resource-consuming endeavor. Next-generation resequencing projects in cattle will be focused on the assembly using a reference genome as the basis for contig construction. Like in humans, [Levy et al., 2007; Wheeler et al., 2008] resequencing projects in cattle will become the major source of polymorphic markers (mostly SNPs and indels) for QTL association studies and QTN discoveries [Eck et al., 2009].

References

H. takes advantage of the extensive mapping data generated in the bovine genome project. Bovine genomic sequences are available from the Btau-4.0 assembly at the UCSC Genome Browser (http://genome.ucsc.edu). The bovine genome assembly has been used in a number of studies to identify QTL and to identify candidate genes for these QTL. For example, the genome-wide association study (GWAS) performed by Schenker et al. (2010) identified a number of QTL that are associated with milk production traits. These QTL were mapped to specific regions of the bovine genome, and candidate genes were identified in these regions.

In conclusion, the bovine genome has provided a wealth of information that has been used to understand the genetics of important traits such as milk production. The availability of this information will continue to be important in the development of new technologies and in the improvement of dairy cattle.