General introduction: The aims of genomics in the 21's century era

Genomics is the scientific study of structure, function and interrelationships of both individual genes and the genome in its entirely.
Recognition of DNA as the hereditary material, determination of its structure, elucidation of the genetic code, development of recombinant DNA technologies and establishment of increasingly automatable methods for DNA sequencing set in the 1990 the stage for Human Genome Project (HGP) and parallely the stage for others genome projects regarding microorganisms, invertebrates, fish and mammals, in particular the mouse, the rat and the farm animals.
Current progress in genetics, comparative genomics, biochemistry and bioinformatics can bring insight into the functioning of organism in health and disease at the cellular and DNA level. The genomics becomes the central and cohesive discipline addressed to biomedical research and the genome sequences, the complex of information that guides biological development and function of organisms, lie at the beginning of any molecular discovery.
The main aim of the genomics after the complete sequencing of some model organism genomes, like, for example, *Caenorhabitis elegans, Drosophila melanogaster, Mus musculus* and, ultimately in 2003, *Homo sapiens*, is to enlarge bases knowledge in order to improve human health and well-being. In particular the genomics needs to extend the knowledge of all the components encoded in the human genome, determine how they function in an integrated manner to perform cellular and organism functions, understand how genome changes and takes on new functional roles.
Actually the human's genome structure is extraordinarily complex and its function poorly understood. Only 1-2% of its bases encode proteins and an equivalent amount of the non-coding genome is under active selection, suggesting an important function in the controlling the expression of 30000 protein-coding genes and myriad other functional elements, like non-coding genes and sequences determinants of chromosome dynamics. Even less is known about the function of half of the genome, that consists of highly repetitive sequences or the remaining non-coding, non-repetitive DNA.
A first objective of genomics is to catalogue, characterize and comprehend the entire set of functional elements encoded in human and other genomes. Comparisons of genome sequences from evolutionary distant species have emerged as a powerful tool for identifying functionally important genomic elements; from the vertebrate genome sequences analyses many previously undiscovered protein-sequencing gene were revealed; mammal-to-mammal sequence comparisons have revealed large numbers of homologies in non-coding regions, defining them in functional terms. Not only the study of genome sequences inter- species is crucial to the functional characterization of the human genome, but also the study of sequence variation intra- species will be important in defining the functional nature of some sequences.
As a larger knowledge of genome function is acquired new computational tools for the prediction of the identity and behaviour of functional elements has emerged. Moreover

genomics has to understand the interactions between genes and genes products, the complex networks that give rise to working cells, tissues, organs and organisms.
The finding of the study of simple model organisms, like bacteria and yeast, have been extended to more complex organisms, such as the mouse and the human. Also few well-characterized systems in mammals have been useful to discover biological molecular pathways. A complete understanding of the working cells required information from several levels : it was necessary to simultaneously monitor the expressions of all genes in a cell and to measure in real-time the localization, the modifications and activity of the gene products. For this reason new molecular techniques arose : the microarray, to analyze the transcriptome, the entire set of transcripts of a cell; the in-situ hybridization, to follow the presence of a protein in a tissue in vivo; the bidimensional electrophoresis to study the abundance and the composition of a set of proteins present in a cell or in a tissue, giving birth to the proteomics. Many other techniques that modulate temporally and/or spatially gene expressions *in vitro* or *in vivo*, like gene-knockout methods, knock-down approaches and the recent use of small-molecule inhibitors of specific transcript, developed after the discovery of a new regulatory class of small non-coding RNA and their mechanism of action, generally called the RNA-interference.
The final objectives will be to identify the genes responsible for human phenotypic differences, or traits, and in particular the variations in DNA sequence that are correlated to common diseases and responses to pharmacological agents, even if the expression of a pathology is a condition that has a complex origin, and involves the interplay between multiple genetic factors and non-genetic factors, like environmental influences. For these reasons several projects aimed to identify all the single nucleotides polymorphism (SNP) in the DNA sequence (i.e. single base deletions and insertions) of the human and model organisms genome, have been established along the creation of large-scale genetic association studies.
Moreover it should be considered that the genetic variation responsible of normal and disease state, is also a result of the modifications of the genome subjected to the forces of evolution. Thus, a complete elucidation of genome function requires the parallel understanding of the sequence differences across species, in order to : identify functional elements; provide insight into the distinct anatomical, physiological and developmental features of different organisms; define the genetic basis of speciation; characterize the mutational process, which drives not only long-term evolution, but that is also the cause of inherited genetic disease.
The sequencing of human genome provides an unparalleled opportunity to advance our understanding about the role of genetic factors in human health and disease, and to apply this insight to the prevention, diagnosis and treatment of diabetes, cancer, obesity, heart disease, Alzheimer's disease, etc. . The actual genomics knowledge and the new molecular tools are able to understand and reclassify all the human illnesses. In fact, the systematic analyses of somatic mutations, epigenetic modifications, genes and proteins expression and protein modifications should allow the definition of a new molecular taxonomy of illness, that could

be the basis for developing better methods for the disease detection and more effective treatments. Such ‘sentinel methods’ might include analysis of gene expression in circulating leukocytes, proteomics analysis of body fluids, advanced molecular analyses of tissue biopsies. The genetics discoveries will favour also the therapeutic design and the drug development, if we consider that at the present the pharmaceuticals on the market target approximately 500 human products, comparing to the 30000 protein-coding genes present in the human. A particular promising example of the gene-based approach to therapeutics is the application of chemical small molecules that act as positive or negative regulators of individual gene products, pathways or cellular phenotypes, after the screening and the understanding of biological functions of small RNA molecules, like microRNA (Collins et al., 2003).

Genomics now provides more and more powerful tools for unravelling the molecular basis of phenotypic diversity also in domestic animals, but genome research in livestock differs in several respects from that in humans or in experimental organisms, because it is not oriented to the identification of monogenic loci responsible of inherited disease. For decades breeders have altered the genomes of farm animals in search of a desired phenotypic trait and then selecting for it. 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.
Nowadays genomic research in farm animals is oriented to the study of traits of economical interest, like growth, milk production and meat quality, that have a multifactor background and that are controlled by an unknown number of quantitative trait loci (QTL).
Quantitative traits, such as weight and length, show a continuous distribution of phenotype values rather than the discrete values observed for a qualitative trait. They are usually controlled by multiple genes and influenced by environmental factors. A quantitative trait locus is defined as a genomic region that contains one or more genes affecting the same quantitative trait. The number of QTL that controls a given trait is not absolute and, in a statistical model, could be infinite, each genes carrying an infinitesimal effect on the phenotype. The main goal of genome research in livestock is to map and to characterize trait loci controlling various phenotypic traits. This requires powerful genome resources (Andersson, 2001).
Livestock genomics has followed in the footsteps the human genome research, adopting both its successful strategies and technologies. In turn, livestock genomics contributes to inform human genomes and to understand evolutionary history and its underlying mechanisms. Moreover farm animals were shown to be quite valuable resources as models for pathology and physiological studies. For example the reproductive physiology of domestic animals is more similar to humans than that of rodents, because farm animals have longer gestations and pre-pubertal periods than mice; specific physiological traits, such as the digestive system of the pigs, are similar to those of humans.

In addition agricultural science has a unique responsibility to human health and social stability, that is feeding an expanding world population while minimizing environmental and ecological risks. The identification of DNA variation in livestock genomes that predisposes health and productivity with less reliance on hormones, antibiotics and pesticides, will remain a concern for some time. Ultimately DNA analysis from animal tissue can be used as an inexpensive method for tracking the origin of meat sample, providing the quality assurance for the consumers.

Early attempts to construct whole-genome maps of livestock species were based on the two technologies underlying the first human genome maps : somatic cells genetics and *in situ hybridizations* (Womack and Moll, 1986, Yerle et al., 1995). These early maps defined synteny (genes on the same chromosome but not necessarily linked) and cytogenetic locations of sequences hybridizing specific DNA probes. These finding were extremely important for the first comparative mapping because the markers were genes or gene products highly conserved across mammalian genomes.

Modern genomics in livestock had its formal origins in a series of conferences in the early 1990 in which international teams of animal geneticists launched both formal and informal genome projects for some of the most widely used livestock species. From that moment dense microsatellite maps, large-insert yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) libraries, radiation hybrid panel (RH) were used for some livestock species, like cattle, pigs, sheep, horses, river buffaloes, goats, rabbit, chicken and some fish like zebrafish, medaka, pufferfish and the sticklebacks in order to localize trait loci. Linkage genetic maps, using microsatellite on the first rough genetic maps, the clonage and the characterization of interesting loci in the BAC and YAC libraries, high-resolution comparative map using the RH strategy, and the first physical maps were developed.

The development of species-specific array and the production of specific transcript profiles started after the development of large collection of sequenced cDNA clones and the corresponding production of the expressed sequence tags (ESTs) for many farm animals. 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 gene localization within a genome. The National Center of Biotechnology Information (NCBI) provides the most comprehensive EST database for many farm animals, while in the Ensembl database (http://www.ensembl.org/) is possible to find a summary of current analyses on coding regions within genomes for selected farm animals. Mapping information are available on the NCBI site http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=? substituting the last '?' sign with the species taxonomic number (i.e. 9031 for the chicken, 9913 for the cow, 9823 for the pig, 7955 for the zebrafish, 9940 for the sheep, etc. ..).

Selection for desirable traits, or conversely, selection against undesirable traits, has been practiced since the domestication of animals begun more than 10000 years ago. There has

been a long tradition of collecting and analysing data on phenotypic traits for breeding purpose in farm animals, and the most common strategy for finding trait loci was to use existing pedigree. This approach was easy in farm animals because of the large family size; for example, the artificial insemination in cattle allows to have a 1000 progeny from a single male. The promise of more accurate, efficient and economical selection that will produce offspring with desirable phenotypes, underpins a substantial portion of the founding for livestock genome projects over the past two decades.

The early linkage maps for most livestock species were constructed as tools for mapping traits and for developing molecular markers useful in marker-assisted selection (MAS). However the ultimate goal when mapping trait loci, the ultimate marker for MAS, is the identification of the causative mutations underlying the selected phenotype. Positional candidate cloning is the main strategy for this purpose. High-resolution mapping is necessary to restrict the region of interest that could contain the QTL and the number of potential candidate genes. Information on map location and gene function is then combined to identify more precisely positional candidate genes, which are subsequently evaluated by mutation screening and functional analysis. The difficulty in identifying a QTL could increase if the QTL mutations is situated in regulatory rather than in coding regions and the phenotypic effect is shifty, compared with simple loss-of-function mutations that cause inherited disorders.

Although mapped QTLs in livestock number in the hundreds, very few mutations underlying quantitative trait variation have been identified. The trait loci for which the causative gene and mutation have been identified or for which this is expected in the near future are monogenic traits of economic and biological interest : the coat colour of the pig, in which the dominant white colour is determined by a mutation in the KIT gene, encoding the mast/stem-cell growth factor receptor; the body composition, in particular the relative proportion of muscle to fat tissue, in pigs, cattle and sheep, in which different genes have been proposed as candidate genes provoking particular phenotypes, like the double-muscling phenotypes in cattle or the muscular hypertrophy in sheep; fertility traits are also studied in different species like sheep and pigs; monogenic disorder like the bovine leukocyte adhesion deficiency, caused by missense mutations in *ITGB2*.

Others monogenic disorders have been analysed and the corresponding causative mutations have been catalogued in the 'Online Mendelian Inheritance In Animals (OMIA)' database (http://www.angis.org.au/oma/). In this site is possible to find the list of all the single-locus traits mapped in cattle, pig, sheep, horse, and goat, which counts hundreds of genes, and the relative proportion of genes for which the causative mutations have been identified, approximately one-third of them. Till October 2006 (Womack et al., 2006) there are only two example of the causative mutation underpinning the QTLs, both in dairy cattle, and both controlling the fat composition of milk : the first discovery of quantitative trait nucleotide (QTN) was found in the DGAT1 locus on chromosome 14 (Grisart et al., 2004), and the second one was found in the *ABCG2* gene on chromosome 6 (Cohen-Zinder et al., 2005).

Ultimately the general disease resistance to pathogens is attracting attention both to improve animal welfare and to reduce losses in production due to disease. Several studies on the relationship between genetic variation and disease resistance have focused on major histocompatibility complex genes. Target diseases are the trypanosomiasis in cattle; the oedema disease in pigs, that is caused by the susceptibility to E*scherichia coli* infections; the Marek's disease (MD) in the chicken, that provoke a lymphoproliferative disease. The identification of QTLs for disease resistance in livestock may be the next frontier for the domestic animal genomics, in order to the understand the host-pathogen interaction and the subsequent improvement of both animal and human health. Linkage disequilibrium mapping will be a very powerful approach for mapping and finding trait loci in domestic animals once dense SNP maps become available and the cost for genotyping is reduced. Current initiative to develop complete BAC contigs of farm animal genomes will produce large-insert contigs covering the region of interest as soon as a trait locus is mapped. Such large-insert contigs can then be used to build a preliminary transcript map of the region by high-resolution comparison with the corresponding region in humans or mice. The completion of the farm animal genome sequencing will provide the researchers with the possibility to analyze the phylogenetic conservation of a causative mutation and its functional role, that will be evaluated later by experimentation. In this way it could be possible to unravel the molecular basis for a variety of phenotypic traits of agricultural, biological and medical significance.

In this thesis two different studies are proposed.
The first part of my work describes a research included in the E.U. funded project 'BovGen', aimed to develop advanced genomic tools useful to study the molecular and genetic control of important traits in cattle. In particular, only an aspect of the project is described : the construction of a high density RH map of bovine genome, which was developed under the initiative and the responsibility of the Institute of Zootechnics of the Faculty of Agriculture of the Catholic University of Piacenza, (Italy), having the professor P. Ajmone Marsan as supervisor.
The second part discusses the involvement of microRNA, an important class of expression regulatory elements in the genome, during the normal development of the mammary gland in a model organism, the mouse. The study of these regulatory elements intends to enlarge bases knowledge about the genetic mechanisms that control the proliferation, differentiation and apoptosis of cells in the tissues composing mammary gland during the reproductive cycle. This work was supported and conducted by the Laboratory of Biochemical Genetic and Cytogenetic (LGBC) at the INRA (Institut National de la Recherche Agronomique) of Jouy-en-Josas (France) under the responsability of F. LeProvost.
The study of some functional elements of the mouse genome, required mouse sequence information available on the Ensembl database, thanks to a previous work of construction of physical maps and genome sequencing in the mouse, analogous to what has been done for the cattle in the BovGen project. The complete genome sequencing of the bovine was considered

an important task in genomic research, a necessary step not only to increase genetic data on this economically important species, but also because of its general utility in the construction of comparative maps and in the identification of new genes or new regulatory conserved elements. Moreover the study of the microRNA function in mammary gland opens the way to the discovery of biological mechanism of cellular proliferation, that could be correlated to the development of breast cancer, but also to the discovery of molecular mechanism that guides epithelial tissue differentiation till the production of milk. In the future it could be possible that the new finding in the mouse could be applied to the bovine, to increase the milk production or to control the timing of lactation.

Recently a new study (Clop et al., 2006) about a QTL controlling meatiness in Texel sheep, demonstrated that the causal mutation in this species is located in the myostatin gene (GDF8) and that a G to A transition in the 3' UTR of the gene creates a target site for two known microRNA, miR-1 and miR-206, which causes translational inhibition of myostatin gene and the muscular hypertrophy, showing how the knowledge of the mechanism of action of microRNA and the use of instruments like genetic map can fuse and focus on particular biological aspects, like the study of economically important QTLs.

First part : A High-density radiation hybrid map construction

I-Introduction

I-I The objectives of livestock genomics

The detection of loci affecting economically important traits represents a major objectives in livestock genomics. It should ultimately lead to more efficient breeding schemes (marker-assisted selection or MAS) and improve the accuracy and intensity of selection programs (Georges and Andersson, 1996; Haley, 1995). In this perspective genetic maps have been constructed in various livestock species, like bovine, sheep and goat, to detect regions containing genes and QTL. The identification of genes and cloning of the corresponding genes may be achieved by standard positional cloning, taking advantage of the existence of large insert libraries and searching for transcribed sequences in these regions.
Cattle are a major economic resource worldwide, therefore there has been considerable interest in the identification of genes that are involved in improved cattle production. Numerous reports have identified genomic regions corresponding to economically important traits in cattle (Georges and Andersson, 1996; Georges, 1999), based on low to medium density genetic linkage maps of the bovine genome.

I-II Genetic maps : brief history

A genetic map shows the relative position and order of markers along the chromosomes of the genome. Genetic mapping is based on the examination of a segregating population, that could be experimental, created for example by cross-breeding experiments, or natural, such as a family, following the principle of inheritance as first described by Mendel in 1865 in his two lows of Genetics, about the segregation of independent genes.
The first genetic maps were constructed in the early decades of the 20th century for organisms such as fruit fly and used simple features inherited on genetic base like markers, even before the discovery that genes are segments of DNA. Genes were looked on as abstract entities responsible for the transmission of heritable characteristics from parents to offspring. To be useful in genetic mapping a heritable characteristic must exist in two alternative forms or phenotypes, each specified by a different allele of the corresponding gene. In the beginning the only genes that could be studied were those specifying phenotypes that were distinguished by visual examinations, like genes for the body color, eye color, wing shape, but soon it was realized that only a limited number of genes has a clear phenotype and in many cases the analyses is complicated because more than one gene affects a single physical feature. It was necessary to find characteristics that were more numerous, more distinctive and less complex than visual ones. The next markers used were biochemical phenotypes, easy to detect in microbes and humans, like antibiotic resistance or amminoacid request for the bacteria and

yeast growth, or the blood groups and immunological proteins such as human leukocyte antigens (the HLA systems) in humans.
Soon it was accepted that a map based entirely on simple phenotypes is not detailed because the genes are widely spaced out in the genome with large gaps between them and moreover only a fraction of the total number of genes exist in allelic forms that can be distinguished conveniently.

I-III Molecular markers

Mapped polymorphisms that are not genes are called DNA or molecular markers. To be useful they must exist in at least two allelic forms.
Many types of molecular markers with different characteristics were developed using different molecular techniques that analyze the variation in the sequence of DNA.
The first ones were the restriction fragment length polymorphisms (RFLP), produced after treating the DNA with a restriction endonuclease. The set of fragments produced can vary if there are single base variations in the DNA sequence of the restriction sites, leading to a length polymorphism of the fragments.
Others molecular markers that are generated from singular base variations of the sequence of DNA were developed later and they can be produced after sequencing of DNA, such as the Single Nucleotide Polymorphism, the SNP markers, or using the PCR (Polymerase Chain Reaction), like the Random Amplification Polymorphic DNA or RAPD markers, or by a combined use of restriction endonuclease and PCR, such as the Amplifyed Fragment Length Polymorphism or AFLP markers.
Another class of molecular markers, widely used in the construction of high-density genetic map, are the Simple Sequence Length Polymorphism or SSLPs markers, that comprise the minisatellites and the microsatellites. The SSLPs are tandemly repeated sequences that show length variation, in the minisatellite the repeats units comprises from tens to a few hundred nucleotides, while in the microsatellite the repeats are shorter, usually di-, tri- or tetranucleotide units. These variations of the number of repeat sequences in the DNA take origin from “errors” during the duplication of DNA during meiosis. It is possible to identify the SSLPs markers by PCR because the sequence flanking them are usually single copy sequence in the genome. Microsatellites are more popular and used compared to the minisatellites, because microsatellites are more conveniently spaced and distributes throughout the genome and because they are shorter and therefore easily to type by PCR.

I-IV Genetic linkage maps

A genetic linkage map is based on the principle of genetic linkage, first discovered by Bateson, Saunders and Punnet in 1905, but not fully understood until Thomas Hunt Morgan began his work with fruit flies in 1910-11. This principle sets that chromosome are inherited as intact units and then pair of genes located on the same chromosome are physically linked together and should be inherited together if any crossing-over event recombins homologous portion of two paired chromosomes during the meiosis. The probability that two different genes localized on the same chromosome are inherited together is proportional to the physical vicinity of the two genes considered and inversely correlated to the number of crossing over events that could occur between two genes localized in distant part of a chromosome. The localizations and orders of markers along a chromosome in genetic linkage map reflect a measure of probability. The distance between markers is not physical, but it is measured in centiMorgans (cM), 1 cM corresponding to 1% of frequency recombination between genes. The real distance in base pair, kilobase or megabase between markers and genes is measured only in physical maps, that are not produced using information from breeding experiments or pedigrees, but examining directly the DNA with molecular biology techniques in order to localize markers on different portions of a chromosome.

SNP and microsatellites, due to their high abundance in the genome, are getting more and more importance in linkage genetic maps and identification of QTLs. Microsatellites are excellent genetic markers because of their high polymorphism, different alleles containing different numbers of repeat units, comparing to the SNP, which has only two alleles.

Genetic linkage maps, based primarily on highly polymorphic, anonymous microsatellite markers, have been important in identifying chromosomal regions influencing economically important traits in cattle (Casas et al., 2001; MacNeil and Grosz, 2002; Li et al., 2002).

Cattle genetic linkage maps were constructed in 1997 with 746 markers (Barendse et al., 1997) and 1250 markers (Kappes et al., 1997), the latter one, spanning 2990 cM, was characterized by an average interval of nearly 3.0 cM.

This cattle genetic map was probably sufficient to assign hereditary phenotypes to specific chromosomes, but not to fine-map them. An intensive efforts to develop more markers to narrow the critical region was required. However, the time, labor and cost per marker of isolating DNA markers from a specific chromosomal region was substantially greater than randomly isolating markers.

Thus a random isolation of microsatellite, from microsatellite-enriched libraries (Stone et al., 1995), was chosen to enrich markers across the genome. The microsatellites were genotyped and assigned to chromosomes by multipoint linkage analysis using the CRIMAP software and a new high density bovine genetic map consisting of 3960 markers, including 3802 polymorphic microsatellite and 79 SNPs, with an average marker interval of 1.4 cM, covering 3160 cM for each of the 30 bovine chromosomes, was produced. This map represented a

powerful resource for fine-mapping of QTLs and a genetic backbone for the development of well-annotated gene maps in cattle and other related species.
Recently Ihara et al. (2004) improved this cattle genetic map and developed a microsatellite-based high-density genetic map on the basis of more than 880000 genotypes across the USDA MARC cattle reference families with a potential genetic resolution of 0.8 cM at the 95% confidence level (approximately 800 kb in the bovine genome).

I-V Somatic hybrids and FISH

There are different kind of physical maps, produced with many molecular techniques, that have different degree of resolution in the assignment of genes to chromosomes.
The first crude mapping of genes on chromosomes was obtained in human by Ruddle in 1972 fusing irradiated human cells with rodent cells and observing the generation of mononucleate hybrid cell lines capable of indefinite multiplication that, after the application of selective media, express human biochemical markers in association with the retention of human chromosomes. In the hybrid cells most of the human chromosomes were rapidly and preferentially eliminated and with appropriated stained preparations it was possible to identify the human chromosomes detecting their specific banding patterns (Goss and Harris, 1975).
The correlations between the retention of human biochemical markers in hybrids cells with the retention of identifiable chromosomes permitted to assign 50 human genes to specific chromosomes. The identifications of the position of genes within the chromosome has been achieved in the beginning by exploiting translocations that segregate linked markers (Boone et al., 1972; Gerald et al., 1974), even if this method couldn’t be applied to every genes, but only to the genes that are localized into a segment of chromosome large enough to be identified in a translocations.
Recently a bovine/hamster hybrid cell panel consisting of 30 independent hybrids was developed to locate genes (Itoh et al., 2003). The characterization of the panel by typing 279 microsatellites markers revealed the presence of all bovine chromosomes in either entire or fragmented form. The panel was also characterized with EST and 1400 EST were assigned to specific chromosomes, thus making this panel a useful tool to the regional mapping of new genes to cattle chromosomes.

The most direct way to localize a genomic segment on a chromosome is to use locus specific-probes in the *in situ hybridizations*, that is able to visualize the target within a particular banding patterns along chromosomes. The recent development of the in situ hybridization is the fluorescent in situ hybridization, or FISH, able to analyze the position of more than one probe on chromosomes at the same time, by labeling different probes whit different fluorescent molecules and the FIBER-FISH, which gives the possibility to hybridize specific probes directly on a single starnd of DNA attached to a solid support.

However the resulting cytogenetic map has lower degree of resolution compared to other kind of physical map constructed with different techniques, for example analyzing by restriction-based fingerprinting large fragments of DNA, even of megabase, contained in BAC clone library.

I-VI BAC-based physical maps

A BAC (Bacterial artificial chromosome) clone is a bacterial clone that contain one artificial chromosome made fusing casually large fragments of the genome of interest with two arms of the bacterial chromosome, that have to contain the centromer and the telomer, or only the telomer, and which carries a marker of selection on each arms.
The wide use of BAC libraries is due to the clone fidelity, to a low level of cloning artifacts, to the easy of separate the BAC DNA from the host's DNA, to the fact that often individual clones contain complete genes embedded in their genomic environment and then the clones can be used for functional studies in cell lines or transgenic applications.
A bovine artificial chromosome BAC library of 105984 clones was constructed in the vector pBeloBAC11 and organized in 3-dimensional pools in 2001 at the INRA of Jouy-en-Josas (France), (Eggen A. et al., 2001). The average insert size was estimated 120 kb after isolation by field inversion gel electrophoresis (FIGE) of digested fragments of 388 clones. Assuming that the bovine genome contains $3x10^9$ bp the total library corresponded to a four genome coverage. The library was also screened by PCR with 164 microsatellite markers to verify the homogeneous distribution of fragments from all the genome in the clones. FISH was performed for over 50 BAC clones and no one was found chimeric. This bovine BAC library contributed to increase the genome coverage of the cattle of the already existing bovine BAC libraries of 2.7 (Buitkamp et al., 2001), 6 (Cai et al., 1995), 10 (Warren et al., 2000), and 5 (Zhu et al., 1999) genome equivalents, bringing the total coverage of the bovine genome represented in BAC libraries to 28.
An analogous bovine BAC library was constructed and called the 'CHORI 240 cattle BAC library' (http://www.chori.org/bacpac). This library contains approximately 200000 clones and was created by cloning partially digested *Mbo*I genomic DNA isolated from a Hereford bull into the *Bam*HI cloning site of the pTARBAC1.3 vector.

Currently BAC libraries have been extensively used to build numerous chromosome specific or whole genome sequence physical maps by BAC fingerprintings and BAC-end sequencing. Whole genome maps have been constructed for a number of organisms including rat, cow, zebrafish, sorghum, maize and tomato (see www.genome.clemson.edu/fpc and www.bcgsc.edu for links to the corresponding web sites).
A first generation bovine BAC-based physical maps was constructed in 2004 at the INRA of Jouy-en-Josas (Schibler L. et all., 2004). This map was assembled analyzing the totality of the

clones of the bovine BAC library of the INRA and part of the CHORI-240 BAC library (26500 clones) by fluorescent double digestion fingerprinting and sequence tagged site (STS) screening.

DNA preparation was performed using a modified alkaline lyses procedure for each clone. 300-400 ng of BAC DNA was submitted to a double digestion (*Hind*III and *Hae*III), which on average generates about 40 bands of 55 to 750 bp, and simultaneously to a dye labeling. The restriction profiles of the samples was analyzed by capillary electrophoresis using a 1000 automated 96 capillary DNA sequencer. The runs were analyzed with the Genetic Profiler software developed to perform the genotyping analyses on the MEGABACE. The map was constructed starting from an initial stringent build and using an incremental process, which consisted in joining together assembled and ordered part of DNA sequence, contigs, based on end-end comparison. The map was validate and the contigs were anchored using the PCR screening information for a total of 1303 markers (451 microsatellites, 471 genes, 127 EST, 254 BAC ends). The final map, which consisted of 6615 contigs assembled from 100923 clones selected from the two libraries, was considered a valuable tool for genomics research in ruminants, including targeted marker production, positional cloning or targeted sequencing of region of specific interest. This map provided also a good framework to initiate a strategy similar to that of Gregory et al. (Gregory et al., 2002) to establish high-resolution sintenies among ruminant, human and mouse genomes.

I-VII Comparative maps

An important step for efficiently sequencing a new mammalian genome is to have a high-quality, comparatively anchored physical map.

Fujiyama et al. (2002) produced a comparative clone-based map of the human and chimpanzee genomes using paired chimpanzee BAC-end sequences (BESs) aligned by BLAST with the human genome sequences and founding that approximately 98% of chimpanzee BESs has BLAST hits in the human genome that identify putative orthologs. Gregory et al. (2002) produced a detailed comparative physical map of the mouse and human genomes by combining BAC-end sequencing with a whole-genome BAC contig created by BAC fingerprinting, revealing remarkable colinearity of the mouse and human genome.

Larkin et al. (2003) used a large-scale BAC-end sequencing strategy to built the first sequence-based physical and multi-species comparative maps of cattle. They sequenced at both ends a total of 40224 bovine BAC inserts of the CHORI-240 cattle BAC library and generated approximately 60500 high-quality cattle BESs whit an average read length of 515 bp. These BESs comprise more than 14 Mbp of non repetitive cattle DNA, thus providing a resource for anchoring cattle genomic sequences to the human and mouse genomes. The non repetitive cattle BESs were then tested for similarity to human and mouse genome sequence (NCBI Build 30) using BLASTN, revealing 29,4% and 10,1% significant hits, respectively

and howing that random cattle BESs had 3.3-fold higher similarity hits to the human genome than the mouse genome. More than 60% of all cattle BES hits in both the human and mouse genome were shown to be located in within known genes, including coding and non coding regions.

I-VIII Radiation hybrid maps

In order to construct a high-resolution physical map for each specific chromosome, basic tool to assist the final high-quality sequence assembly of the genome, and comparative mapping information from maps of the annotated human and mouse genome can be utilized efficiently. The location of bovine loci that are homologous of human genes may be predicted from the current knowledge about the conservation of synteny between genomes, but comparative mapping can sometimes produce errors, because it is based on the colinearity between two different genomes even if some genomic regions are not colinear, thus the position of a locus has to be actually proven by direct mapping on genome.
Radiation hybrid (RH) mapping has been shown to be a powerful tool to integrate comparative genome data with information from existing genetic and physical maps to generate high-resolution maps (Itoh et al., 2005).
The technology for generating physical maps using irradiation and fusion gene transfer was first developed more than 20 years ago by Goss and Harris (1975). This technology was employed in an isolating mapping experiment of human X chromosome genes ten years later by Williard et al.(1985), but it was not systematically used as a human gene mapping instrument until the work of Cox et al. (1990) of construction of a high-resolution map of the human chromosome 21. This map was constructed using hybrids generated by irradiation fusion gene transfer between a donor somatic cell hybrid containing a single human chromosome and the recipient rodent cell line. Mapping the entire human genome with this approach was impractical because it required a panel of 100-200 hybrids for each chromosome and a screening of over 4000 hybrids to generate a genomic map. For this reason Walter et al. (1994) reverted to the original method of whole genome radiation hybrid (WG-RH) of Goss and Harris, that is the use of diploid cell line like a donor genome at the place of a single chromosome of interest from a somatic cell hybrid, to demonstrate that a panel of hybrids of a diploid human cell line with a rodent recipient line could be used to map any human chromosome. Later Gyapay et al. (1996) and Hudson et al. (1995) demonstrate the emergence of WG-RHs as stand-alone mapping tools publishing two WG-RH maps of the human genome opening the way to the RH maps development.

I-VIII-a Advantages of RH maps

In contrast to linkage maps, which exploit the frequency of natural recombination between markers to calculate distances and orders of markers, RH maps are constructed using the probability of breaks between markers induced by radiation. The retention frequency, that is the measure of the proportion of donor genome retained in hybrids, of two markers is proportional to their vicinity in the genome, and inversely correlated to the number of breaks that could occur between the two markers. The retention pattern of markers for each hybrid is compared to determine linkage and map distances between markers. These distances are measured by centiRay, 1 centiRay (N rad) corresponding to a 1% frequency of breakage between these two markers after exposure to a radiation dose of N rad of X-rays. (McCarthy, 1996).

Radiation hybrids allow a clear determination of a linear order of markers along a chromosome and radiation hybrid mapping has two major advantages over physical mapping and genetic mapping: it has much higher resolution and the markers don't need to be polymorphic to be included in the map. It is an especially powerful tool for comparative gene mapping, since chromosomal order can be established for expressed genes that are usually conserved between species, but often recalcitrant to linkage mapping for lack of allelic variation. Moreover the radiation hybrids maps bridge the gap between genetic and physical maps because they offers the possibility to anchor the large DNA insert of the bacterial artificial chromosome and to identify their orientation.

I-VIII-b Principle of construction of RH panels

To generate RH panels, the donor cell line is irradiated with a lethal dose of X-rays or γ rays, and fused with the recipient cell line, using either Sendai virus or polyethylene glycol (PEG). Non-recombinant donor cells die whitin a week of irradiation. The recipient cell line will contain a selectable marker; the most frequently used are thymidine kinase deficiency (TK-) or hypoxanthine phosphoribosyl transferase deficiency (HGPRT-). Cells containing either of this marker will not grow in media containing HAT (hypoxantine, aminopterin, thymidine). The only post-fusion cells that will grow in HAT medium are recipient cells containing all their complete genome added with casual portion of donor DNA containing both the wild-type TK or HPRT gene. The hybrid colonies are expanded for DNA extraction and 96-well microplates are filled whit the hybrid DNA and the control DNA in order to be screened by PCR for the retention of genetic markers.

I-VIII-c RH panel characteristics and uses

In radiation hybrids the irradiation is utilized both to kill the donor line and to induce chromosomal breaks producing hybrids with the desired fragments size.

Increasing the irradiation dose from 5 to 25 Krads Siden et al. (1992) observed a 5- to 10-fold reduction in the size of the fragments, as well as a dramatic reduction in the retention frequency from 27 to 3%. The optimal radiation doses chosen to construct a panel of radiation hybrids is dependent upon the intended use of the lines. Low dosages results in decreased resolution of a chromosome map, while at very high dosages (greater than 10000 rads) no significant linkage between loci is observed due to extensive fragmentation and loss.

Higher-dosage hybrids which carry small fragments of DNA from a region of biological interest have been used for constructing recombinant DNA libraries and DNA probes (Florian et al., 1991).

It is generally believed that breakage along the chromosome, as well as the rejoining of the broken ends, is a random process (Heddle, 1965). However stabilization of a fragment in the hybrid requires the rejoining of the fragment with elements needed for replication and stable mitotic segregations. The preferential retention of the centromere in radiation hybrids has been observed in a number of radiation hybrids panels (Benham et al., 1989; Goodfellow et al., 1990; Ceccherini et al. 1992; Abel et al., 1993; etc.).

FISH has been used to determine the number and relative size of human fragments carried in hybrids. The number of fragments appeared to be independent of the irradiation dose used to generate the hybrids. FISH was used also as a screening procedure to identify hybrids containing human DNA, which are subsequently used for marker analyses.

The first issue in the design of a radiation hybrid mapping experiment is the number of hybrids required to achieve optimal resolution. This problem has been reviewed by Lunetta and Boehnke (1994). They calculated the resolving power of radiation hybrid panels of varying sizes as a function of retention frequency, assuming that retention frequency is the total number of radiation hybrids retaining a given marker divided by the total number of radiation hybrids tested with the marker. They suggested that a radiation hybrid panels of 90-100 lines is adequate for most mapping experiments.

The protocol for scoring markers on a radiation hybrids panel is a critical step in building the map. Markers scored as present (+) or absent (-) are completely informative; thus, false positives and false negatives bias the map. Ambiguous data can be entered as unknown (?). Testing of the markers is commonly carried out by visual inspections of ethidium bromide-stained PCR products from sequence-tagged site (STS) markers. The problem of scoring many markers across the panel is variation in the relative sensitivity of the marker tested. The problematic markers are those that show abnormally high or low retention frequency and it is normal to avoid them as anchor points in initial radiation hybrid map construction.

The first phase of analyses is a test of each marker against all the other tested markers, or two-point analyses. The two-point analyses can be used to estimate distances between markers, and to identify linkage groups to subject to multipoint analyses, that represent the second phase of the analyses. Multipoint analyses can define the trial orders of markers inside a linkage group and between clusters of markers. Normally this analyses is carried out using as small as possible linkage groups because it is computationally intensive, with N!/2 possible orders to consider for N markers present in each group. It is efficient to subdivide the problem into clusters of markers to be ordered within cluster, then order and orient the ordered clusters (Leach and O'Connell, 1995).

I-VIII-d Software used to construct RH maps

When a marker is tested on the RH panel the pattern of the presence (+) or absence (-) across the panel defines a cytogenetic placement; those markers with the same pattern of + and – are localized in the same cytogenetic 'bin'. Ordering of the bins is carried out either by the ordering of the known cytogenetic breakpoints, or by minimization of the obligate breakpoints under the assumption that the majority of the rearranged chromosomes arise from a single breakage event. These analyses have been carried out in the beginning manually, nowadays analyses packages are available.
One of the software used to produce RH maps for each chromosome is the Microsoft Windows versions of 'Chartagene' (Schiex et al., 2002), available publicly from www.inra.fr/bia/T/CarthaGene'.
The other programs available for building radiation hybrid maps are RH map (Vanderstop et al., 1991), RHMAPPER (Soderlund et al., 1998) and multi-map.
RH, cytogenetic and linkage maps can compared by using Anubis software (www.roslin.ac.uk/cgi-bin/anubis).

I-VIII-e RH bovine panels and maps

Whole genome-radiation hybrid (WGRH) panels have now been used to create medium to high resolution chromosomal maps in several species, including human (Gyapay et al., 1996), mouse (Schmitt et al., 1996; McCarthy et al., 1997), rat (Watanabe et al., 1999), pig (Yerle et al., 2002), horse (Chowdharhary et al., 2002), chicken (Morrison et al., 2004), zebrafish (Geisler et al., 1999), dog (Priat et al., 1998) and cattle (Womack et al., 1997; Rexroad et al., 2000; Williams et al., 2002; Itoh et al., 2005; Band et al., 2001).
Four whole genome radiation hybrid panels available for cattle have been used to construct RH maps: the Womack-5000 rad panel of 90 RH clones (Womack et al., 1997), the Womack-

12000 rad panel of 180 RH clones (Rexroad et al.,1999); the TM112-3000 rad panel of 94 RH clones (William et al., 2002) the SUNbRH 7000 rad panel of 90 RH clones (Itoh et al., 2005).

The first RH bovine panel was developed in 1997 using like a bovine donor cells a normal diploid fibroblast culture established from an Angus bull, JEW38. The cells were irradiated with a cobalt 60 source delivering 185 rad/min for a total dose of 5000 rad. The recipient cell line was the Chinese hamster TK- fibroblast line A23. Six markers were genotyped in all 101 RH lines.

RH panels are generally characterized and anchored to existing genetic maps using microsatellite markers. The Womack-5000 rad panel was screened with six markers spanning each of the linkage maps of bovine chromosome 1, 13 and 19 to create the first whole-genome-RH radiation bovine hybrid map. Later the same RH panel was used to create a cattle-human whole-genome comparative map (Band et al., 2000).

Williams et al. (2002) constructed and characterized a 3000-rad RH panel in order to create an outline bovine RH map. This map was developed testing on the RH panel and incorporating in the map the majority of markers available on published bovine linkage maps.

This RH panel was constructed using like donor cell line a primary bovine fibroblast cell line established from a male Holstein calf by explants culture. Cells were exposed to a 3000 rads of X-rays and fused with the HGPRT-deficient Chinese hamster cell line, Wg3H (Goss and Harris, 1975). 224 cell lines were established and screened with 33 microsatellite markers. A subset of 100 hybrids whit higher average retention frequency was selected and a final panel of 94 hybrids was produced, whose DNA is publicly available for purchase from the Res Gen Invitrogen Corp (cat no. RH10, Huntsville, Ala., USA).

In order to link the 3000-rad RH panel to the genetic (Barendse et al., 1997; Kappes et al., 1997, http://www.marc.usda.gov/genome/genome.html, www.cgd.csiro.au) and physical maps that were published for the cattle till that moment, a total of 1238 markers were typed by PCR on the RH panel (http://www.roslin.ac.uk/radhyb/), of which 1148 are microsatellite loci and 90 are genes or markers within genes. Between them 64 could not be placed, so that 1174 markers were included on the RH-maps of 29 autosomes and the two sex chromosomes.

In most cases the order of markers was consistent between the RH maps, the published linkage maps, the current RH chromosomes maps (chr1: Rexroad et al., 1999; chr 15: Amarante et al., 2000; chr 19: Yang et al., 1998; chr 23: Band et al., 1998) built by using the Womack panel, and the low-density whole genome maps of Band et al. (2000).

Itoh et al. (2005) used the whole genome 7000-rad radiation hybrid (RH) panel, SUNbRH (7000-rad), to build a high-resolution RH map. The Shirakawa-USDA linkage map served as a scaffold to construct a map of 3216 microsatellites on which 2377 ESTs were ordered. The resulting RH map provided essentially complete coverage across the genome, with 1 cR7000 corresponding to 114 kb.

I-VIII-f Integration of bovine RH map data in the construction of comparative maps

RH maps are considered a useful resource for creating comparative maps between bovine and human chromosomes through the alignment of the loci derived from coding sequences (Amaral et al., 2002; Goldammer et al., 2002; Gautier M et al., 2002; Gautier M et al., 2003; Larkin et al., 2003; Everts-van der Wind et al., 2004; Everts-van der Wind et al., 2005).

Larkin DM et al. used the cattle-hamster 5000-rad RH panel of Womack et al.(1997) to confirm *in silico* predictions of cattle chromosome positions of bovine BAC end sequences (BESs). 60547 BESs were previously anchored to the human and mouse genome by BLASTN search, like we have already described, thus the cattle chromosome locations had been predicted for the cattle BESs with significant BLAST hits in the human genome using the COMPASS Perl scripts software (COMPASS III), producing a virtual map of BESs on the cattle chromosomes.

The COMPASS strategy (comparative mapping by annotating and sequence similarity) permits the predictions of chromosome map location based upon sequence similarity of orthologous genes, if comparative map information is available for two species (Band et al., 2000; Rebeiz and Lewin, 2000).

In that case the chromosome location of BESs was predicted using data from the first-generation cattle-human comparative RH map (Band et al., 2000). Furthermore they confirmed *in silico* predictions of cattle chromosome location for a total of 109 BESs having a single high-confidence human hit on HSA11. Oligonucleotides able to discriminate cattle from rodent sequences were designed for these BESs and 89% of them gave distinct PCR product after screening of the RH panel. 84 BESs were mapped on BTA15 or BTA29 after two-point linkage and multipoint map analyses, carried out with RHMAPPER 1.22 (Slonim et al., 1997) software. Thus the high degree of accuracy (approximately 86%) of BLAST-COMPASS approach was demonstrated and a cattle-human comparative map with greater than 1-Mbp resolution was created, 84 BAC ends were added to the existing cattle RH map.

Recently Everts-van der Wind et al. (2005) used the same approach, to construct a high-resolution whole-genome cattle-human comparative map and to add new markers (cattle BESs) to the current high resolution cattle 5000-rad RH map (Band et al., 2000; Everts-van der Wind et al., 2004) collectively known as the Illinois-Texas 5000-rad radiation hybrid panel (IL-TX RH 5000).

They screened by PCR the RH panel of Womach et al. with BES from the CHORI-240 BAC library selected by BLAST for having a single significant match in the human genome, distant one from one other 1 Mbp in the human genome, and having preferentially an orthologous hit in the mouse genome. Approximately 3000 cattle bacterial artificial chromosome end sequences were added to the previous RH map, increasing the number of markers 4 time. The number of comparative points in the human genome was increased 5-fold.

An important advance made possible by mapping cattle BESs is that the RH map could be anchored directly to the whole-genome BAC fingerprinting contig. Comparisons of BES order on the RH maps and whitin the fingerprinting contigs is used to identify inconsistency in the maps and markers or clones that are presumably 'out of place' on the basis of their cattle –human comparative map location. This comparison will ultimately be important in selecting the correct minimum tiling path for the BAC-skim sequencing and correctly assembling the cattle genome sequence. Moreover also the additional mapping information coming from the integration of RH and linkage map would greatly improve the bovine genome sequence assembly (Snelling et al:, 2004; Weikard et al., 2006).

I-VIII-g Integration of bovine RH map data with genetic linkage maps

Linkage maps have been important in identifying chromosomal regions influencing economically important traits in cattle (Casas et al., 2001; MacNeil et al., 2001; Li et al., 2002), but because the lack of recombination between closely linked markers limits resolution, linkage maps are of limited value for ordering closely linked markers and identifying genes underlying quantitative trait loci. The radiation hybrid mapping provides higher resolution for ordering close markers, but high breakage frequency RH data are less reliable than linkage data for ordering widely separated groups of markers (Schiex et al., 2001).

Integrating linkage and RH data into a single map not only will refine marker order to facilitate genomic sequencing, but will also increase the efficiency of identifying genes associated with QTL.

Integration of linkage and RH maps has been reported for a number of species (NIH News Release, http://www.genome.gov/page.cfm?pageID=10506668), like the dog (Breen et al., 2001), the rat (Steen et al., 1999), the feline (Sun et al., 2001) and individual bovine chromosomes (Amarante et al., 2000; Rexroad et al., 1999; Drogemuller et al., 2002). The general approach to integrated mapping has been to score several markers from linkage maps on the RH panel, then align the independent maps via common markers.

While Nadkarni (1998) and White et al. (1999) described procedures to synthesize information from multiple independent analyses into a single merged map, Snelling et al., (2004), differently, used directly data from independent analyses to contribute to the construction of two maps and then merged independent data sets with common markers to built a single integrated map.

Agarwala et al. (2000) developed procedures for integrating RH maps, where markers common to independent RH panels contributed to the solution of a comprehensive RH map, while Schiex et al. (2001) developed and released CarthaGene software (CarthaGene home page, http://www.inra.fr/bia/T/CarthaGene) to merge and solve integrated maps representing multiple linkage and RH data sets.

The bovine chromosome (BTA) 15 was considered interesting and was chosen from Snelling WM et al. (2004) to study the integration of linkage and RH data and to compare the bovine and human genome because a QTL for meat tenderness was reported to be present on this chromosome (Keele et al., 1999; Rexroad et al., 2001) and because comparative mapping indicated that alternative segments of human chromosome (HAS) 11 are conserved on BTA 15 and 29 (Amarante et al., 2000; Rexroad et al., 2001; Gautier et al., 2002). They used the second-generation linkage map of bovine genome (Kapper et al., 1997), and the radiation hybrid data for 109 markers from the ComRad project radiation hybrid panel (94 cells lines, Williams et al., 2002; Gautier et al., 2002) to construct an integrated BTA15 map representing 145 markers, whose 42 shared by both data sets, 36 unique to the linkage data and 67 unique to RH data.

Another study that aimed to the construction of a high-resolution map of a specific chromosome was carried out from Weikard et al. (2006) on the bovine chromosome 6 (BTA 6) because a number of different QTL for various phenotypic traits, including milk production, functional, and conformation traits in dairy cattle as well as growth and body composition traits in meat cattle, have been mapped consistently in the middle region of this chromosome (Bovine QTL Viewer at texas A&M University 2005, http://bovineqtl.tamu.edu/ ; Reprogen QTL Mapof Dairy Cattle Traits 2005 http://www.vetsci.usyd.edu.au/reprogen/QTL_Map/).

The objective of the study was to construct a high-resolution 'gene rich' RH map for the target chromosomal region of BTA6 containing candidate genes underlying the QTL for milk production traits (Cohen-Zinder et al., 2005; Olsen et al., 2005; Schnabel et al., 2005; Weikerd et al., 2005) in order to dissect the different QTL at the gene-based level.

A total number of 237 loci including 115 genes and expressed sequence tags (ESTs) and markers from the recently published bovine genetic map (Ihara et al., 2004) were typed on the cattle-hamster 12000-rad WG-RH panel (Rexroad et al., 2000) and the new RH map, with a total of 234 loci, displayed a substantial increase in loci density compared to existing physical BTA6 maps. The average retention frequency of the markers was 15.2% and the average inter-loci interval on the targeted BTA6 region covered on the RH map was 17.8 cR12000, corresponding to approximately 300 kb. The order of loci determined in the new map for the targeted BTA 6 region was generally consistent with that reported on previous published RH (Itoh et al., 2005, Everts-van der Wind et al., 2004) and linkage map (Ihara et al., 2004; Snelling et al., 2005).

High-resolution RH maps integrate anonymous markers, ESTs, and genes from currently available bovine linkage and RH maps as well as high number of comparative anchor loci derived from the orthologous human chromosomes. Although a number of links to the currently existing genetic, cytogenetic, and RH maps are possible, a multitude of contigs and scaffolds of the available bovine genome sequences resources still have to be anchored and/or oriented on the chromosomes. Connecting animal phenotypes associated with the QTL anchored on genomic level with putative underlying genes would accelerate the identification

of sequence polymorphisms and gene variants and the development of SNP markers for validation of association substantially.

I-IX International bovine projects

I-IX-a International physical map and Bovine sequencing projects

An international bovine physical map project (www.bcgsc.bc.ca/projects/bovine_mapping, http://www.livestockgenomics.csiro.au/cattle.shtml) was proposed to analyze single digest fingerprintings obtained from 280000 BAC clones to identify new fusions between contigs from the two BAC-based physical maps derived from the BAC library of the INRA (Schibler et al., 2004) and the CHORI-240 BAC library. Additional mapping information can facilitate the ordering of fingerprinting contigs for the construction of physical BAC maps covering whole chromosomes and ultimately provides a valuable starting point for whole genome sequencing projects, like it happened for the human (Cao et al., 1999), for the mouse (Gregory et al., 2002) and in Drosophila (Hoskins et al., 2000).
The ultimate map for a species is the correctly assembled genome sequence.
The U.S. National Institute of Health (NIH) has given high priority to the complete genome sequencing of two Cetartiodactyl species, *Bos taurus* (cattle) and *Sus scrofa domestica* (pig; http//www.genome.gov/page.cfm?pageID=10002154) to make progress the mammalian comparative genomics because the mammalian order Cetartioactyla comprises a philogenetically distant clade of eutherian mammals relative to primates, having diverged from a common ancestor approximately 85 million years ago (Kumar and Hedges, 1998), and, on the basis of a limited amount of sequence information for orthologous regions in a number of mammals (Thomas et al., 2002), it is clear that a Cetartiodactyl genome will play an essential role in informing the human genome for conserved non coding structural and regulatory elements, for properly annotating exon/intron boundaries, and for the identification of novel genes.
The bovine genome sequencing project started in 2003 and used a combination of whole genome shotgun sequencing (WGS) and sample sequencing of a minimum tiling path of BAC clones spanning the genome. An international Bovine Genome Sequencing Consortium was established.
In october 2004 the initial draft of the bovine genome sequence using was released (NCBI Bos taurus Genome Resources 2005-http://www.ncbi.nlm.nih.gov/genome/guide/cow/; Human Genome Sequencing Center at Baylor College of Medecine_Bovine Genome project 2005-http://www.hgsc.bcm.tmc.edu/projects/bovine).
Preliminary assemblies of the current bovine genome sequence update representing a 6x coverage were established and announced in October 2005 (Pre! Ensembl (Btau 2.0) in NCBI Bos taurus genome mapview (build 2.1), (Pre! Ensembl Bos taurus Genome Assembly Site 2005, http://www.ensembl.org/Bos_taurus/index.html; NCBI Bos taurus Map Viewer Site

2005, http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9913), while recently a new assemblies (build 3.1) is available (www.livestockgenomics.csiro.au/perl/gbrowse.cgi/bova3/).

I- IX-b The BovGen project

The BovGen project started the first January 2003 and involved the work of European and extra-european scientific groups belonging to different institutes (Rosline Institute-UK, University of Alberta-Canada, INRA-France, Catholic University of Piacenza-Italy, Tuscia University of Viterbo-Italy, Max Planck Institute for Molecular Biology of Berlin, Germany).
The project had the objectives to develop advanced genomic tools to provide the necessary infrastructure for researchers to study the molecular and genetic control of important traits in cattle. Information on these traits could then be applied to the selection of cattle that are best suited to producing healthier food products of the desired quality in appropriate production systems.
As the project progressed the international project to sequence the bovine genome made very rapid progress and an addition priority objective was included in the BovGen project: to work closely with the international Bovine Genome Sequencing Consortium to aid the assembly of a high quality bovine sequence.

In details the intended molecular tools to improve or create were:
1) the best characterised bovine expression array available with around 20,000 unique expressed sequences (ESTs) to give the possibility to examine gene expression profiles in target cells under various physiological conditions such as, fed or starved, healthy and diseased as an important route to gene discovery and understanding gene function.
It was planned that the expression arrays should contain a non-redundant , or “unigene” set of 20000 unique ESTs identified in cDNA clones from a bovine brain cDNA library. The non-redundant set of ESTs was created by the Max Planck Institute starting from the brain as it was supposed that this organ expresses the greatest diversity of genes in the body and the 20000 ESTs were estimated, from the human sequence, to represent about 30-40% of all genes in the genome.
2) a high resolution RH bovine map which let the construction of cattle-human comparative map that could contain not only more than 300 and 400 links between bovine and human genomes, like the actual RH comparative maps has, but at least more than 3000 links, as the actual mouse-human comparative map, in order to place cattle genomics information on a par with the mouse-human comparative information.
3) the construction of long genome spanning BAC contigs. In this project the INRA bovine BAC library, with 105000 clones, including 20000 clones from the CHORI 240 BAC library, was available and was characterised with the ESTs sequenced to increase the immediate

utility of the BAC library and provide access points to BAC clones for local sequencing objectives.

4) the ultimate bovine genome sequencing. An international consortium competing with the international Bovine Genome Sequencing Consortium was established to sequence the bovine genome. The corner stone to the sequencing work was a whole genome BAC contig. The characterisation of the BAC library in this project was an important input to the assembly of the genome wide BAC contig. An additional contribution of the Bovgen Project was the ordering of Sequence scaffolds on chromosomes, which was achieved using markers identified within the sequences to align them with the chromosomal maps.

Almost the totality of objectives of the project were achieved and 30 publications and numerous international conference presentations were produced from this work, that made significant contribution to the international bovine sequencing project.

II-Objective

Several approaches can be used to determine the order of loci on chromosomes and hence develop maps of the genome. However, all mapping approaches are prone to errors either arising from technical deficiencies or lack of statistical support to distinguish between alternative orders of loci. Errors in maps can greatly affect the ability to map and isolate genes for complex and Mendelian traits (Risch and Giuffra 1992; Feakes et al., 1999; Goring and Terwilliger 2000), for the identification of QTL.

Inaccuracies in genetic maps can result from genotyping errors, as well as from the use of a limited number of informative meiosis to generate maps. A higher confidence in genetic-map order can be obtained by creating maps using a likelihood-ratio criterion of >= 3, as opposed to using a minimum-recombination map (Morton 1955).

Errors in the order of markers on physical maps can be due to problems with assembly or to incorrect identification of marker positions. Even when the order of markers is known to be without error, accurate estimates of recombination fractions will play an important role in linkage and associations studies (Clerget-Darpoux et al., 1986; Risch and Giuffra, 1992; Goddard et al., 2000; Collins et al., 2001; Reich et al., 2001).

The accuracy of the genome maps could in principle be improved if information from different maps (genetic, comparative with other species, RH submitted to different radiation intensity, physical, sequence assembly) was combined to produce integrated maps.

The publicly available bovine genomic sequence assembly is a draft that contains errors. Correcting the sequence assembly requires extensive additional mapping information to improved reliability of ordering of sequence scaffolds on chromosomes.

RH panels represent a powerful tool to construct high-resolution maps.

RH panels are generally characterised using microsatellite markers; however the number of these markers is often insufficient to join all the linkage groups and assemble complete maps, particularly for high-resolution panels. The development of additional anonymous markers can be a time-consuming task, and generally other types of markers, particularly ESTs, are used to saturate RH maps. These ESTs also serve to link the RH map with maps in other species (Schlapfer et al., 2002; Weikard et al., 2002).

The objective of the work described is the construction of a bovine high-density RH map, one of the main aim of the BovGen project, which could be used for the construction of an integrated map and could contribute to the International Sequencing Project to aid the final assembly of the bovine genome sequence.

It is discussed the presence of possible errors in the RH map comparing with other recently published RH and genetic maps (the Illinois-Texas (ILTX) RH map and the MARC 2004 linkage map) aligning the sequence of the corresponding mapped markers. All the bovine maps were aligned with the 6x bovine assembly (Btau_2.0 sequence) to identify its potential inconsistencies.

III-Material and Methods

III-I Sequencing of ESTs

A non-redundant "unigene" set of ESTs was selected by oligo-nucleotide fingerprinting and clustering of cDNAs from a brain library (Herwig *et al.*, manuscript in preparation). This non-redundant cDNA clone set contains 23040 bovine clones grouped by sequence assembly of ESTs into 14989 unique cDNA clusters and singletons. The cDNA clones of the "unigene" set were amplified in a 384-well microplate format by PCR consisting of an initial denaturing for 2 min at 95°C, denaturing for 45 sec at 94°C, annealing and elongation for 4 min at 65°C in 30 cycles. PCR primers were complementary to the insert-flanking vector sequences. The PCR mix contained 5 pmol of forward and reverse primers (table 1), 0,1 mM dNTP's, 1,5 M Betain, 1x PCR buffer, 0,1 mM Cresol Red and 1 U per reaction Taq DNA polymerase. PCR buffer consisted of 0,5 M KCl, 1% Tween20, 15 mM $MgCl_2$, 350 mM TrisBase, 150 mM Tris/HCl pH 8,3. PCR fragments were subjected to sequence analyses using BigDye-terminator chemistry (Applied Biosystems) and a 3700 DNA sequencer (Applied Biosystems). Average sequence read length was 750 bp. The individual EST sequence data were submitted to GenBank and are publicly available under accession numbers CO871676-CO897060.

Table1. Sequence of primers used to amplify the cDNA inserts

forward primer	**GGATCTATCAACAGGAGTCCAAGCTCAGCT**
reverse primer	**TCACCATCACGGATCCTATTTAGGTGACAC**

III-II Primer design

Maximum sequence information for annotation was achieved by aligning the ESTs data with available public cattle transcript sequences contained in the TIGR bovine gene index. TIGR clusters and corresponding ESTs cattle sequences produced were aligned and the resulting 14989 cluster sequences (consensus) were used for the subsequent construction of primers. Cluster sequences were aligned with bovine genomic sequences and only those showing clear splicing were used to define the precise exon-intron boundaries for the final primer selection. The primer design was carried out using dedicated software now in the public domain (Polyprimers, http://www.unitus.it/SAG/primers.zip). The software uses the nearest-neighbour method (SantaLucia et al., 1996) to predict the complementarity of primers and secondary structures (dimers, hairpin etc.) and is able to process large number of sequences in batches, picking primers in designated regions. To minimize the amplification of hamster DNA contained within the RH panel cell lines, primer pairs were designed with one primer within exon, the other within the adjacent intron or non coding sequence. The primer design was standardized to achieve a maximum of uniformity in their amplification conditions.

Primer details are available to the public in the ArkDB database (ArkDB Public database browser, http://www.thearkdb.org).

III-III Screening of the Roslin RH panel

2473 marker loci were successfully typed on the 94 cell lines of a 3000-rad bovine/hamster RH panel as described by Williams *et al.* (Williams et al., 2002). Vectors of 262 AFLP markers (Gorni C et al., 2004) were added to the dataset.

III-IV RH data analyses

RH vectors were assigned to chromosomes by analysing 2-pt linkage with mapped loci (Gorni et al., 2004) using RH mapper (Slonim et al., 1997). Multipoint maps were constructed using the default algorithm of the Carthagene software (Schiex and Gaspin, 1997). The initial multi-point map was improved by an iterative process of inspection of marker loci and removal and alternative addition of badly linked or disrupting loci. This process resulted in the removal of 122 loci that could not be reliably fitted into the chromosome maps with highest probability. The best maps generated by this process were compared to the ComRad RH-map (Gorni et al., 2004) and the MARC 2004 linkage map (Ihara et al., 2004) and regions showing discrepancies were examined in detail to identify the presence of problem markers. Marker positions on the maps of each chromosomes are available from the ArkDB database at http://www.thearkdb.org.

III-V Mapping of marker associated sequences against the bovine sequence assembly

ESTs sequences used to design the primers for mapped loci were aligned with the assembled 6x bovine sequence assembly (Btau_2.0) using BLAST (Altschul et al.,1990) and Spidey (Wheelan et al., 2001). To filter out incorrect alignments the BLAST e-value was set to a maximum of 1e-20 and minimum percent identity to 90%. In addition, the relative length of the BLAST hit (i.e. coverage, or length of the hit divided by the length of the query sequence) had to be at least 80%. Where ambiguous alignments were observed higher stringency filters were applied (sequence similarity higher than 97.5% and coverage higher than 90%).

III-VI Diagrammatic representation of chromosomal maps

Visual representation of map alignments for figures 2-5 was achieved using cMap (GMOD Generic Software Components for Model Organism Database, http://www.gmod.org/cmap/). For figure 1, a custom ruby script was used in combination with the bioruby toolkit (BioRuby http://www.bioruby.org).

IV-Results

IV-I Radiation hybrid map

A total of 2735 markers were added to those, 1231 markers, on the first-generation whole-genome RH maps (Williams et al., 2002), of which 2473 are newly mapped loci and 262 are previously reported AFLP markers (Gorni et al., 2004), giving a total of 3966 markers, of which 1999 are within genes, 1072 are microsatellite loci, 262 are AFLP markers, 376 are BAC end sequences and 257 are from ESTs sequences that do not show convincing similarity to the annotated bovine sequence (table 1). The RH maps for the 30 bovine chromosomes constructed from this data can be viewed and information can be downloaded from the ArkDB database (http:// www.thearkdb.org).

The total length of the RH map, including all bovine autosomes and the X chromosome is 760 Rays (R). The map of BTA 28 is the shortest one, 1141 cR, and the longest one is that of BTA7, 4408 cR. The average marker interval over the whole genome is 19 cR ranging between 12 cR (BTA29) to 29 cR (BTA20). Distance comparisons between common markers on the RH map, MARC linkage map and the bovine sequence suggests, on average, that 1 cR on the BovGen RH map is equivalent to 0,04 cM and 23 Kbp respectively, although this varies considerably across the genome.

Table1. Statistics of the RH maps by chromosome.

[a] BAC end sequences; [b]ESTs which could not be assigned to an annotated sequence; [c]average over whole genome

BTA	marker numbers		marker types					map length			marker density	
	no of markers	new mapped	AFLP [26]	BES[a]	genes	micro-satellites	unknown ESTs[b]	cR	cM	Mbp	cR/ marker	Mbp/ marker
1	158	83	8	2	69	70	9	3695.80	154.67	102.83	23.39	0.65
2	169	112	11	3	98	46	11	3487.70	128.88	86.54	20.64	0.51
3	206	160	10	3	134	35	24	4405.40	128.90	85.36	21.39	0.41
4	113	56	10		64	36	3	2605.20	119.93	69.56	23.05	0.62
5	275	220	9	94	105	40	27	4210.30	135.60	76.43	15.31	0.28
6	59	28	6		33	20		1699.40	134.42	69.62	28.80	1.18
7	217	153	13		141	45	18	4408.80	135.56	69.14	20.32	0.32
8	93	51	7	10	35	38	3	2458.60	128.62	62.12	26.44	0.67
9	133	62	8		58	64	3	3004.10	116.17	64.65	22.59	0.49
10	162	108	17	2	91	40	12	2760.50	118.83	70.00	17.04	0.43
11	178	123	14	3	102	48	11	3658.90	130.97	87.17	20.56	0.49
12	89	37	16	1	34	36	2	1553.30	109.95	48.61	17.45	0.55
13	128	76	14		64	36	14	2275.40	105.38	62.72	17.78	0.49
14	223	189	1	169	16	33	4	2667.10	103.95	50.71	11.96	0.23
15	147	43	12	2	77	38	18	2434.10	109.75	53.82	16.56	0.37
16	120	79	9	1	72	33	5	2957.20	94.46	56.99	24.64	0.47
17	115	65	15		68	27	5	2654.10	95.86	45.92	23.08	0.40
18	200	155	9	5	141	25	20	3182.10	84.38	56.51	15.91	0.28
19	156	104	7	53	58	35	3	2136.70	109.61	56.39	13.70	0.36
20	58	26	9		26	23		1712.20	82.94	42.95	29.52	0.74
21	62	18	9		18	33	2	1449.00	83.79	49.72	23.37	0.80
22	111	74	5	5	69	31	1	2031.90	88.10	48.33	18.31	0.44
23	130	78	8	1	80	20	21	2345.70	80.05	41.60	18.04	0.32
24	68	15	6	5	19	35	3	1588.70	78.13	45.25	23.36	0.67
25	133	93	6	1	89	31	6	2256.20	68.42	41.41	16.96	0.31
26	75	34	6		36	30	3	1715.10	79.39	35.72	22.87	0.48
27	65	31	3	11	18	29	4	1191.20	67.37	31.57	18.33	0.49
28	65	31	5	4	25	22	9	1141.30	61.66	34.89	17.56	0.54
29	150	110	4		103	34	9	1895.70	69.73	45.82	12.64	0.31
X	108	59	5	1	56	39	7	2390.70	146.50	47.90	22.14	0.44
Total	3966	2473	262	376	1999	1072	257	75972.4	3151.97	1740.25	19.16[c]	0.44[c]

IV-II Comparison with the ILTX RH map

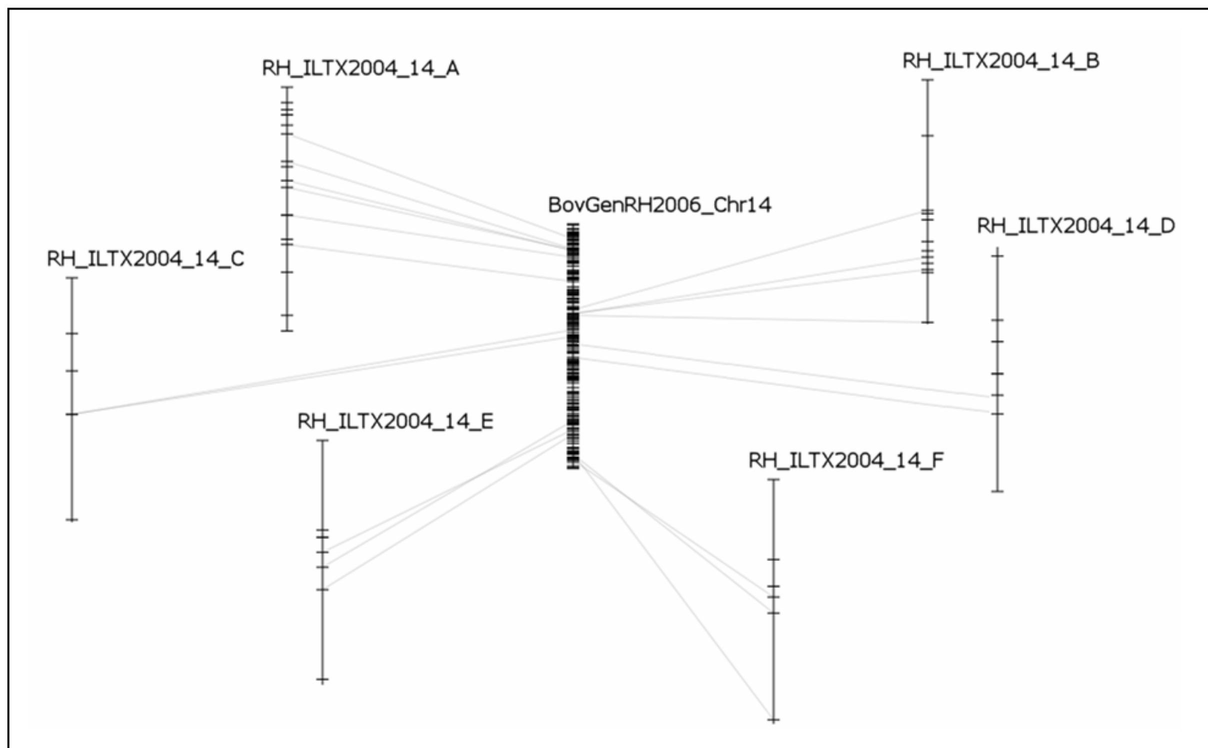
There are 241 marker loci in common between the BovGen RH map described here and the Illinois-Texas (ILTX) RH map, comprising 71 linkage groups (Everts-van der Wind et al., 2004). All of these common loci were assigned to the same chromosomes on both maps.

Correspondences in 32 linkage groups cannot be assessed for consistency of their order because the groups contain only one or two markers common between these maps. For the remaining 39 linkage groups 21 are in perfect agreement with the BovGen RH map and 14 have only one inconsistently positioned marker.

For example, the BovGen RH map of chromosome 14 has 20 markers in common with the ILTX RH map. These are divided into six linkage groups (14_A to 14_F), which are located consecutively along the chromosome. The groups contain 2 to 6 markers which are in common and the order generally agrees between both maps (figure 1). In four linkage groups (5_A, 7_A, 27_B and X_C) discrepancies between the maps are observed with more than one displaced marker. One of those, 5_A is relatively consistent despite four discrepancies in

order, as it contains 26 correspondences and covers a complete chromosome, and the discrepancies are minor. In contrast 7_A, 27_B, 30_C contain fewer correspondences (6 each) but all have several inconsistencies. Each of the three groups cover approximately half a chromosome and differ from the BovGen RH map in their marker order at 4, 4 and 5 correspondences, respectively.

Figure 1. BovGen RH map of the chromosome 14 compared with the corresponding six linkage groups of the ILTX RH map. Lines between maps connect markers common in both maps. Marker names were omitted to improve perceptibility.



IV-III Comparison with MARC 2004 linkage map

There are 885 marker loci in common between the BovGen RH and the MARC 2004 linkage maps (Ihara et al., 2004) which allows a detailed comparison of map orders and chromosome assignment.

Inconsistencies in chromosomal assignment are found for 5 of these 885 loci. In all these cases only individual markers are involved. The marker order on 13 chromosomes (BTA 4, 10, 11, 13, 14, 16, 18, 21, 23, 24, 25, 27 and 28) is in very close agreement between the BovGen RH maps and MARC 2004 maps. For example the order of the 27 markers on chromosome 4 which are in common shows only minor inversions of two pairs of linked loci (*BMS1840* and *MAF70* and also *BMS2571* which appear on the different sides of the co-mapping markers *BMS779* and *BMS3002*) (figure 2). Despite of the similarity in both cases the marker order as suggested by the MARC map is inconsistent with the multipoint map BovGen RH data, as the MARC order gives a much lower p-value.

On a further 13 chromosomes minor discrepancies between these maps are observed. On BTA 3, 5, 8, 9, 12, 17, 19, 22 and X the order of markers is essentially the same, but with a number of individual markers at different positions. For BTA 1, 2, 6 and 26 differences are observed involving the orientation of linkage groups, but with the order of markers within the linkage group is conserved. For example on BTA 26 the marker order is in general consistent between the BovGen RH and the MARC 2004 linkage map, however two small linkage groups 26_A (*BMS882*, *TGLA429*, *BMS2567* and *BM6041*) and 26_B (*MAF36*, *ILSTS091*, *MAF92* and *BM804*) have the same marker order in both maps, but are inverted with only one marker (*BM7237*) at divergent position (figure 3).

On four chromosomes major inconsistencies are observed, where complete linkage groups map to different chromosomal positions (BTA 7, 29) or where the order of markers differs within several linkage groups (e.g. BTA 7, 15 and 20). On BTA 7 for example, the position of two linkage groups 7_A (limited by the markers *CSKB071* and *TGLA303*) and 7_B (limited by the markers *BM6105* and *BM2607*) is exchanged. In addition 7_A is in a different orientation in both maps, while the marker order in 7_B is inconsistent (figure 4). Nevertheless, these discrepancies only involve about a quarter of the chromosome, and 12 out of the 38 common markers. The map positions of the other 26 markers are in close agreement between the two maps.

Figure 2. BovGen RH maps of the chromosome 4 compared to the MARC 2004 linkage map. The number of markers in each map is indicated in brackets. Lines between the maps connect markers common in both maps. Only marker names common in both maps are displayed.

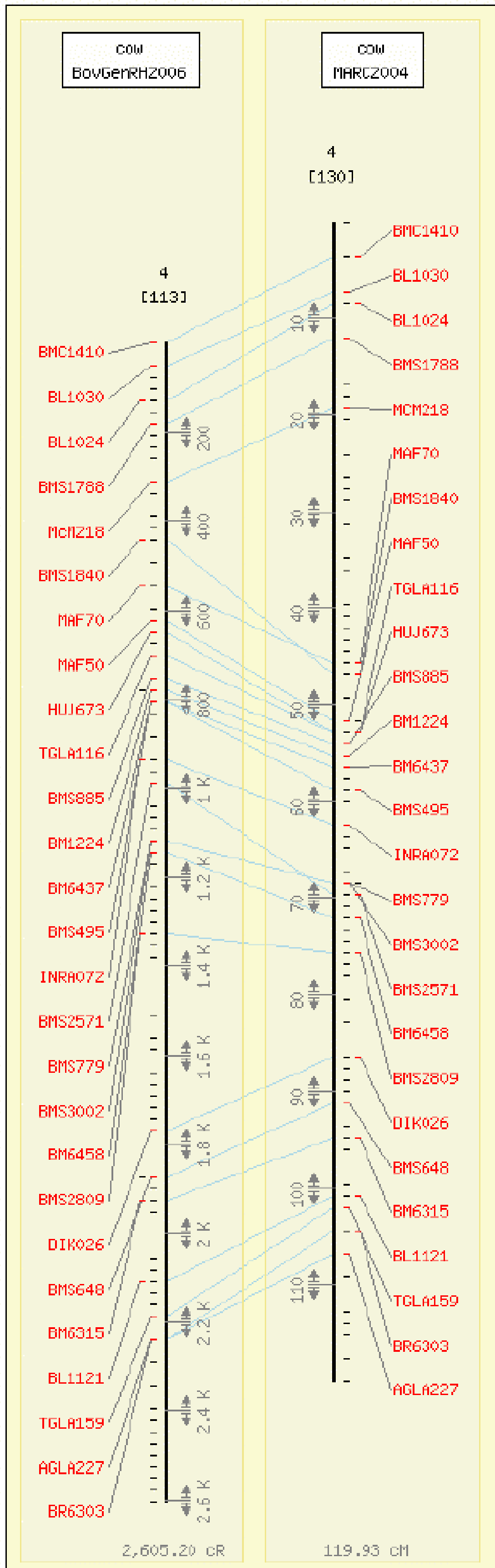


Figure 3. BovGen RH map of chromosome 26 compared with the corresponding MARC 2004 linkage map. The number of markers in each map is indicated in brackets. Lines between the maps connect markers common in both maps. Only markers names common in both maps are displayed. Locations of discussed linkage groups and their orientation are indicated by arrows.

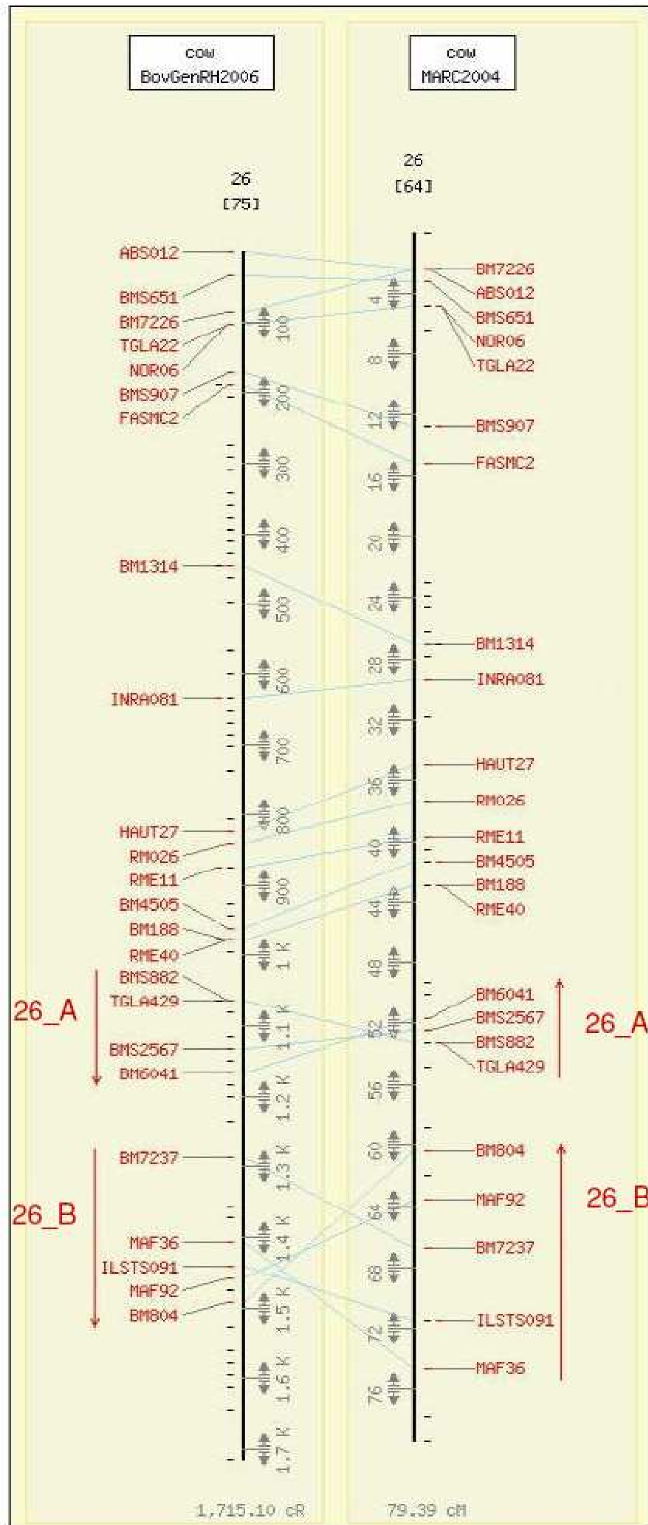
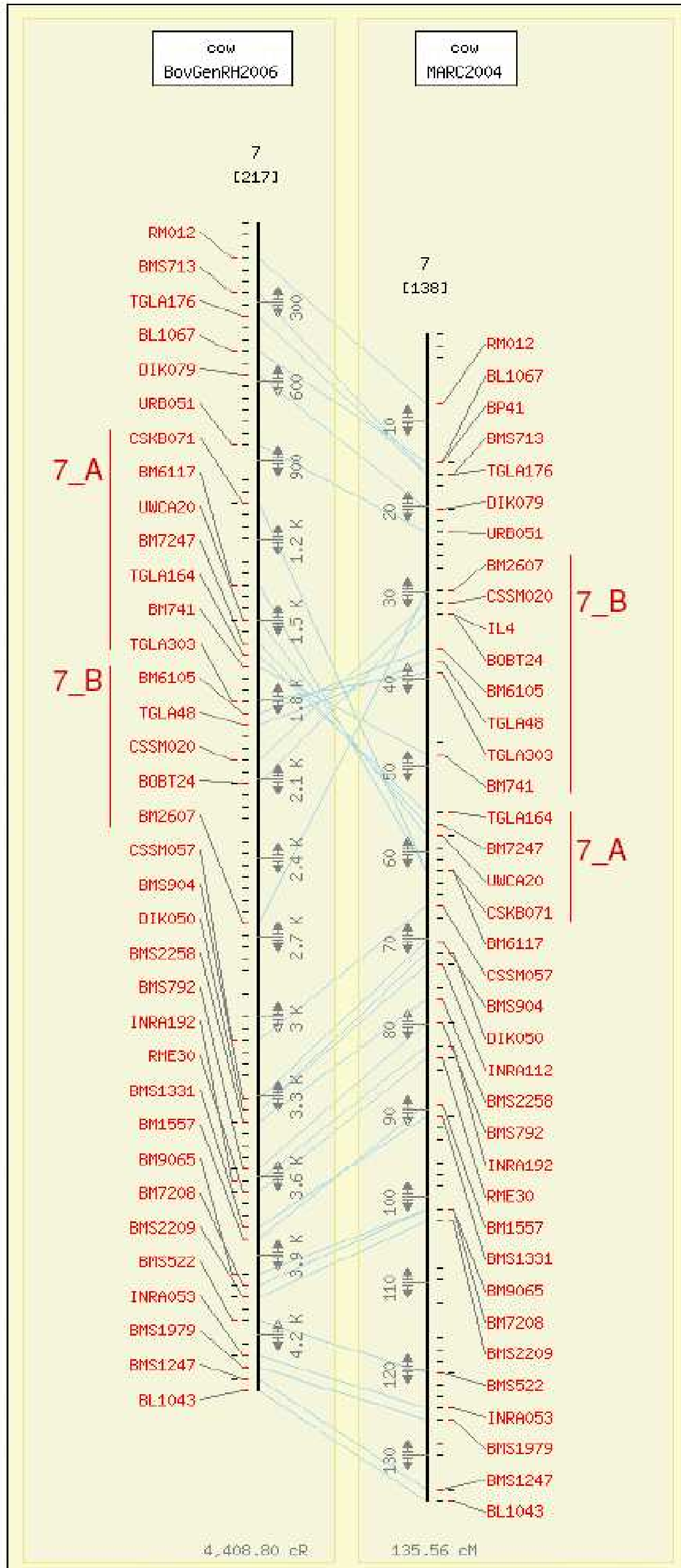


Figure 4. BovGen RH map of chromosome 7 compared with the corresponding MARC 2004 linkage map. The number of markers in each map is indicated in brackets. Lines between the maps connect markers common in both maps. Only markers names common in both maps are displayed. Locations of discussed linkage groups and their orientation are indicated by arrows.



IV-IV Comparison with the 6x bovine assembly

Of the 3966 markers successfully included in the RH map, 2898 could be unequivocally assigned to a position in Btau_2.0 bovine sequence, 2767 were assigned to the same chromosome, but 131 mapped on different chromosomes between the BovGen RH map and the sequence. On seven chromosomes inconsistent assignments involving groups with three or more markers were observed (table 2).

On most chromosomes there were many differences between the map order and the sequence: only on two chromosomes, BTA 9 and 14, the discrepancies were minor, involving a few individual markers in a different order. On most chromosomes large discrepancies involving complete linkage groups and/or large numbers of individual loci were seen particularly on chromosomes 5, 7, 16, 22, 25 and 29. On chromosome 16, six linkage groups are located at different position on the sequence compared with the BovGen RH maps (figure 5).

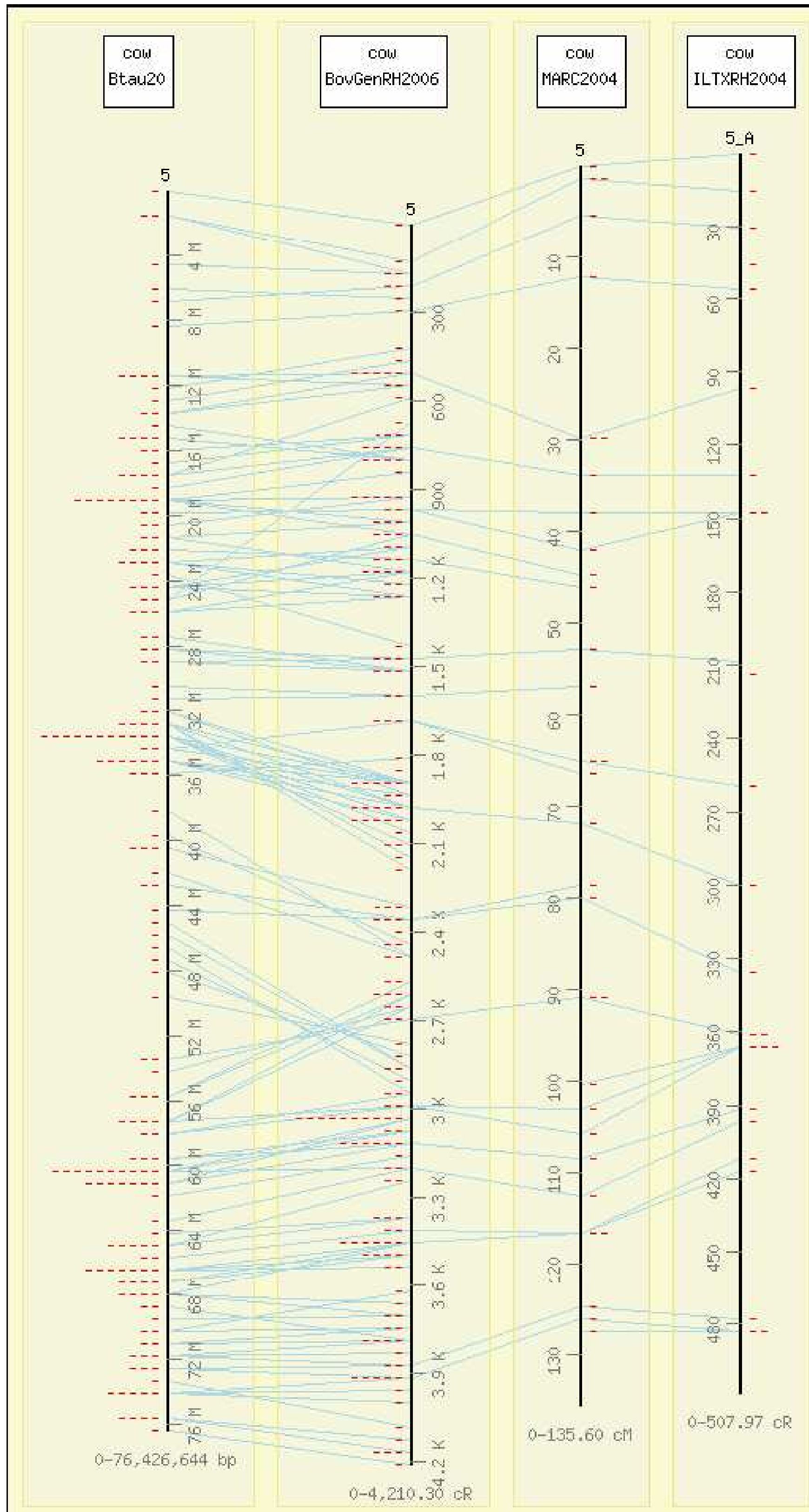
When markers that were at inconsistent positions between the BovGen RH and either the ILTX or MARC linkage maps were removed, 217 common markers with the ILTX RH map and 771 common markers with the MARC2004 linkage map remained where the available mapping data were in agreement. The mapping order of these markers was then compared with the order in the bovine sequence. Using only the markers that are consistent between the BovGen and other RH or linkage maps, the comparison with the Btau_2.0 sequence reveals considerable discrepancies across the whole genome. On chromosome 5 six markers which could be assigned to positions in the sequence assembly appeared with inconsistent positions (*BP1*, *AGLA293*, *ILSTS022*, *CSSM022*, *ILSTS066*). The remaining markers are in close agreement between the three maps and reveal significant inconsistencies with the sequence assembly (figure 6).

Table 2. Inconsistent chromosome assignments between the BovGen RH map and Btau_2.0 sequence. Only the seven most significant cases are listed, involving at least three linked markers. HSA4 is a homologue to BTA6, MM15 and HSA8 to BTA8, HSA14 to BTA21 and HSA17 to BTA19. Most 8 likely assignments are indicated by bold font.

Case	Markers involved	Assignment BovGen RH BTA	Assignment Btau_2.0 BTA	Other assignments species, chromosome [reference]
1	*BMS4030* *BOVGEN_158* *BOVGEN_91*	**1**	4	**BTA1** [40]
2	*MAF23* *BZ855103* *BZ864360*	**5**	4	**BTA5** [41]
3	*ADH7* *ADH1A* *ADN1C*	**6**	21	**HSA4**[a] [42, 43, 44]
4	*PTK2B* *BZ948637* *B4GALT1*	**8**	5	**BTA8** [1], **MM15**[a] [29], **HSA8**[a] [30]
5	*KIAA0284* *Q9Y4F5* *KNS2* *BTBD6*	11	**21**	**HSA14**[a] [31]
6	*BZ850749* *CC517527* *CC471629*	14	25	-
7	*ACLY* *KLHL11* *SC65* *JUP* *E0362G17*	**19**	23	**HSA17**[a] [45, 46]

Figure 5. BovGen RH map of the chromosome 16 compared with the 6x bovine assembly and the MARC 2004 map. The corresponding sequence similarity hits are connected by lines. The number of markers in each map is indicated in brackets. Only marker names common in both maps are displayed. Locations of discussed linkage groups are indicated.

Figure 6. BovGen RH map of chromosome 5 compared with the 6x bovine assembly and with the MARC 2004 and the ILTX RH map. Markers which were inconsistently mapped between the two RH and the MARC linkage mp and also assigned to a position of the sequence assembly were removed. Lines between the maps connect common markers.



V-Discussion

The resolution of genome maps differs between approaches, and all approaches, including the assembly of a whole genome sequence, are prone to errors: in some cases insufficient information is available to assign the correct order or positioning of loci, while data errors can introduce distortions in the maps. The ultimate genome map of a species is the correctly ordered DNA sequence. Achieving the correct sequence assembly uses several levels of information. Sequence information from other species, including the human genome could be used as a template, but should be treated with extreme caution as local species specific variations are known (Ranz et al., 2001).

Direct sequence information is used for local assembly of shot-gun sequence reads into contigs, and these contigs are then assembled into scaffolds using additional information, such as overlapping clones, and sequences from paired clone ends. The ordering of these scaffolds on chromosomes and assembly of the final sequence relies on additional mapping information, including BAC fingerprint contig maps, linkage maps and RH maps.

In this work it was described a RH map with approximately 4000 mapped loci which will contribute to the assembly of the bovine genome sequence.

V-I Comparison with other linkage and RH maps

The reliability of different maps can be assessed by examining consistency in alignment of common loci, however it is important that the information used when assembling the maps is independent, as circular arguments can give a false measure of agreement. In contrast to the approach of Itoh et al. (2005) it was not used a linkage map as template for the construction of the RH maps presented here because the aim was to assemble the most likely maps using only the RH information. This independent data can then be used to assess potential errors across different maps. It was carried out an alignment of the BovGen RH maps with the other available maps of the bovine genome and with the Btau_2.0 sequence assembly, but only after the maps were constructed. This approach could result in maps that are less consistent with other published information, but it is important to realise that is the only way to contribute new information. This independent mapping information can be used to develop a combined map which carries a measure of map confidence, based on similarity and differences between maps using independent data.

The BovGen and ILTX RH maps (Band et al., 2000; Everts-van der Wind et al., 2004; Everts-van der Wind et al., 2005) appear to be more consistent with each other than with the MARC 2004 linkage map. Some inconsistencies between linkage and RH maps may be due to the different mapping approaches, however; the observation of the apparent higher consistency between the RH maps must be treated with care. The BovGen RH map has fewer loci in common with the ILTX map than with the MARC 2004 linkage map and so fewer

discrepancies could be detected. Moreover, the ILTX map consists of 71 unordered linkage groups which are a major source of the inconsistencies.

V-II Comparison with the sequence assembly

Sequence similarity search algorithms used to align maps with Btau_2.0 have a considerable risk of errors as the algorithms might also detect gene duplications or similar motifs in different genes. To minimize this problem it was used very stringent parameters for minimum homology and maximized the required length of overlap between sequences. In addition sequence matches were assessed manually before accepting hits as correct. Thus the loci aligned between the different maps and the sequence carry a very high probability of correctly assigned homology. Differences in the position of individual markers in different maps could be simple technical variations explained by using different parameters and algorithms to construct the multipoint maps. Inconsistencies in the chromosomal assignment of individual markers may also have simple explanations, such as poor primer design resulting in amplification of related loci, and not the target locus. Of greater importance for the interpretation of the map information are inconsistencies affecting whole linkage groups. To minimise the propagation of errors in individual maps we eliminated markers that were inconsistently mapped from further analyses against the sequence assembly.
While the BovGen RH map is in general agreement with the ILTX map and the MARC 2004 map, chromosomal regions of high agreement with the Btau_2.0 sequence are quite rare. Many differences in the marker order between the Btau_2.0 sequence and the BovGen RH map cannot be detected when comparing the two RH and the MARC linkage map. Therefore, after eliminating regions and markers that were inconsistent between these maps, we found that there was poor overall consistency between the RH and linkage maps with the Btau_2.0 bovine sequence assembly. For example on chromosome 4 the marker order on the BovGen RH map is in agreement with the MARC 2004 and ILTX map, but is inconsistent with the sequence assembly. The extent of the inconsistencies detected with the sequence assembly reveals the need for improvement by inclusion of further combined mapping information (figure 6).
If we consider regions where there are inconsistencies between the different mapping methods, e.g. on chromosomes 7, 25 and 29, the assembled sequence is most consistent with the linkage map. Recalculating the maps for these three chromosomes using only markers that can be located in the bovine sequence gives a map length for chromosomes 7, 25 and 29 of 3780,7 cR, 1788,5 cR and 1683,1 cR respectively, when the markers are ordered according to the original BovGen RH maps. If the common markers are forced into the order they appear in the sequence assembly: the map lengths increases to 567,.6 cR for chromosome 7, 2680,5 cR for chromosome 25 and 2683,3 cR for chromosome 29, and the $\log_{10}$ likelihood decreases from -1306,58 to -1615,01 (BTA 7), from -763,13 to –982,82 (BTA 25) and from -741,18 to -976,64 (BTA 29). The marker order suggested by the bovine assembly and the MARC

linkage map is therefore incompatible with the data underlying the BovGen RH maps for these chromosomes.

V-III Assignment of markers to different chromosomes

The most significant problem in the genome assembly is that of erroneous chromosome assignments. By comparing assignments among different RH and linkage maps and also using comparative human or mouse information, it seems likely that the assignment in the bovine assembly is most often at fault (table 2). For example the markers *PTK2B*, *BZ948637* and *B4GALT1* (table 2, case 4) are closely linked on the BovGen RH map of BTA 8 and the linkage map of Barendse et al. (1997) which also locates the genes on BTA 8. This is also consistent with data from Fiedorek and Kay (1995) who mapped *PYK2B* (alias *PTK2B* or *Fadk*) on murine chromosome 15 and Inazawa et al. (1996) who mapped the gene on human chromosome 8 at positions which share conservation of synteny with BTA 8 (Everts-van der Wind et al., 2005).

In contrast these marker loci are placed on chromosome 5 in the Btau_2.0 sequence assembly. All three markers are located on a single sequence scaffold (chr 5.80), suggesting that the chromosomal assignment of this scaffold is wrong.

The linkage group formed by the markers *KIAA0284*, *Q9Y4F5*, *KNS2* and *BTBD6* was assigned to chromosome 11 on the BovGen RH maps; however the assignment is not consistent with other mapping data (table 2, case 5). The human homologues of these loci map to human chromosome 14 (Goedert et al., 1996) suggests that this group is correctly assigned in the Btau_2.0 sequence to chromosome 21 and that the BovGen RH assignment is incorrect. Nevertheless the linkage of this group to other markers on BTA 11 is convincing with LOD linkage values up to 13,8 between the extreme marker *KIAA0284* and the neighbouring markers on the BovGen RH map. If this linkage group is tested with markers located on BTA 21 using the BovGen RH datasets it shows no linkage. In the Btau_2.0 assembly this linkage group is at an extreme telomeric position and suggests that the statistical support for this assignment is weak and may have been made on the expected position derived from the supposed conservation of synteny between human and cattle chromosomes.

The markers *BZ850749*, *CC517527* and *CC471629* are assigned to chromosome 14 on the BovGen RH map and to chromosome 25 in the Btau_2.0 sequence assembly (table 2, case 6). These markers are derived from BAC end sequences of clones from the CHORI-240 library and are not present on other maps which could be used for comparison. All these markers are assigned to the scaffold Chr25.84 and are in a chromosomal region of the assembly with a low density of corresponding markers. In contrast on the BovGen RH map, the markers in the same region are at a higher density. This suggests that these markers are more tightly linked on the BovGen RH map and correctly positioned. No further information is available to resolve this inconsistency.

VI-Conclusions

There is striking consistency between the RH maps presented here, the MARC linkage map and the ILTX RH map. Using this data it is possible to identify possible errors in the assembly of the current bovine genome sequence and hence aid the improvement of the next sequence build. The inconsistencies between the BovGen RH, the Illinois-Texas RH and the MARC linkage maps fall into three categories, markers that are assigned to different chromosomes, which are few, minor rearrangements, which account for the majority of discrepancies, and major rearrangements of marker, or linkage group order, which also are few. When the major discrepancies between these maps are removed a large number of inconsistencies still remain with the bovine sequence assembly. Using the combined mapping information available from the high-resolution RH maps presented here together with the additional map data available from publicly available RH and linkage maps should allow the next assemble of the bovine genome sequence to be improved considerably.

VII-Reference

Abel KJ, Boehnke M, Prahalad M, Ho P, Flejter WL, Watkins M, VanderStoep J, Chandrasekharappa SC, Collins FS, Glover TW, et al. A radiation hybrid map of the BRCA1 region of chromosome 17q12-q21. Genomics. 1993;17: 632-41.

Agarwala R, Applegate DL, Maglott D, Schuler GD, Schaffer AA. A fast and scalable radiation hybrid map construction and integration strategy. Genome Res. 2000; 10: 350-64.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215: 403-410.

Amaral ME, Kata SR, Womack JE. A radiation hybrid map of bovine X chromosome (BTAX). Mamm. Genome 2002; 13: 268-271.

Amarante MR, Yang YP, Kata SR, Lopes CR, Womack JE. RH maps of bovine chromosomes 15 and 29: conservationofhumanchromosomes11and5. Mamm Genome. 2000; 11: 364-8.

Andersson L. Genetic dissection of phenotypic diversity in farm animals. Nat. Genetics 2001; 2: 130-137.

ArkDB Public database browser (http://www.thearkdb.org).

Band M, Larson JH, Womack JE, Lewin HA. A radiation hybrid map of BTA23: identification of a chromosomal rearrangement leading to separation of the cattle MHC class II subregions.Genomics. 1998; 53: 269-75.

Band MR, Larson JH, Rebeiz M, Green CA, Heyen DW, Donovan J, Windish R, Steining C, Mahyuddin P, Womack JE, Lewin HA. An ordered comparative map of the cattle and human genomes. Genome Res. 2000;10: 1359-68.

Barendse W, Vaiman D, Kemp SJ, Sugimoto Y, Armitage SM, Williams JL, Sun HS, Eggen A, Agaba M, Aleyasin SA, Band M, Bishop MD, Buitkamp J, Byrne K, Collins F, Cooper L, Coppettiers W, Denys B, Drinkwater RD, Easterday K, Elduque C, Ennis S, Erhardt G, Ferretti L, Flavin N, Gao Q, Georges M, Gurung R, Harlizius B, Hawkins G, Hetzel J, Hirano T, Hulme D , Jorgensen C, Kessler M, Kirkpatrick BW, Konfortov B, Kostia S, Kuhn C, Lenstra JA, Leveziel H, Lewin H, Leyhe B, Lil L, Martin Burriel I, McGraw RA, Miller JR, Moody DE, Moore SS, Nakane S, Nijman IJ, Olsaker I, Pomp D, Rando A, Ron M, Shalom A, Teale AJ, Thieven U, Urquhart BGD, Vage DI, Van de Weghe A, Varvio S, Velmala R, Vilkki J, Weikard R, Woodside C, Womack JE. A medium-density genetic linkage map of the bovine genome. Mamm. Genome 1997; 8: 21-28.

Benham F, Hart K, Crolla J, Bobrow M, Francavilla M, Goodfellow PN. A method for generating hybrids containing non selected fragments of human chromosomes.Genomics. 1989; 4: 509-17.

BioRuby (http://www.bioruby.org).

Boone C, Chen TR, Ruddle FH. Assignment of three human genes to chromosomes (LDH-A to 11, TK to 17, and IDH to 20) and evidence for translocation between human and mouse chromosomes in somatic cell hybrids (thymidine kinase-lactate dehydrogenase A-isocitrate

dehydrogenase-C-11, E-17, and F-20 chromosomes). Proc. Natl. Acad. Sci. U S A. 1972; 69: 510-4.

Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, Holmes NG, Cheron A, Suter N, Vignaux F, Bristow AE, Priat C, McCann E, Andre C, Boundy S, Gitsham P, Thomas R, Bridge WL, Spriggs HF, Ryder EJ, Curson A, Sampson J, Ostrander EA, Binns MM, Galibert F. Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. Genome Res. 2001;11: 1784-95.

Buitkamp J, Kollers S, Durstewitz G, Welzel K, Schafer K, Kellermann A, Lehrach H, Fries R. Construction and characterization of a gridded cattle BAC library. Anim. Genet. 2000; 31: 347-51.

Cai L, Taylor JF, Wing RA, Gallagher DS, Woo SS, Davis SK. Construction and characterization of a bovine bacterial artificial library. Genomics 1995; 29: 413-425.

Cao Y, Kang HL, Xu X, Wang M, Dho SH, Huh JR, Lee BJ, Kalush F, Bocskai D, Ding Y, Tesmer JG, Lee J, Moon E, Jurecic V, Baldini A, Weier HU, Doggett NA, Simon MI, Adams MD, Kim UJ. A 12-Mb complete coverage BAC contig map in human chromosome16p13.1-p11.2. Genome Res. 1999; 9: 763-74

Casas E, Stone RT, Keele JW, Shackelford SD, Kappes SM, Koohmaraie M. A comprehensive search for quantitative trait loci affecting growth and carcass composition of cattle segregating alternative forms of the myostatin gene. J. Anim. Sci. 2001; 79: 854-60.

Ceccherini I, Matera I, Sbrana M, Di Donato A, Yin L, Romeo G. Radiation hybrids for mapping and cloning DNA sequences of dista l16p. Somat. Cell. Mol. Genet. 1992; 18: 319-24.

Chowdhary BP, Raudsepp T, Kata SR, Goh G, Millon LV, Allan V, Piumi F, Guerin G, Swinburne J, Binns M, Lear TL, Mickelson J, Murray J, Antczak DF, Womack JE, Skow LC. The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. Genome Res. 2003 Apr;13(4):742-51. Erratum in: Genome Res. 2003;13:1258.

CHORI-240 Bovine BAC Library (http://bacpac.chori.org/bovine240.htm).

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. Biometrics. 1986; 42: 393-9.

Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI, Ron M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 2005; 15: 936-44.

Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature 2003; 422: 835-847.

Cox DR, Burmeister M, Price ER, Kim S, Myers RM. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 1990; 250: 245-50.

Clop A, Marcq F, Takeda A, Pirottini D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat. Genetics 2006; 38: 813-818.

DeWan AT, Parrado AR, Matise TC, Leal SM. The Map Problem: A comparison of Genetic and Sequence-Based Physical Maps. Am. J. Hum. Genet. 2002; 70: 101-107.

Drogemuller C, Bader A, Wohlke A, Kuiper H, Leeb T, Distl O. A high-resolution comparative RH map of the proximal part of bovine chromosome 1. Anim Genet. 2002 ; 33: 271-9.

Eggen A, Gautier M, Billaut A, Petit E, Hayes H, Laurent P, Urban C, Pfister-Genskow M, Eilertsen K, Bishop MD. Construction and characterization of a bovine BAC library with four genome equivalente coverage. Genet. Sel. Evol. 2001; 33: 543-548.

Everts-van der Wind A, Larkin DM, Green CA, Elliot JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. PNAS 2005; 102: 18526-18531.

Everts-van der Wind A, Kata SR, Band MR, Rebeiz M, Larkin DM, Everts RE, Green CA, Liu L, Natarajan S, Goldammer T, Lee JH, McKay S, Womack JE, Lewin HA. A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome Res. 2004; 14: 1424-1437.

Fadiel A, Anidi I, Eichenbaum KD. Farm animals genomics and informatics: an update. Nucleic Acids Res. 2005; 33: 6308-6318.

Feakes R, Sawcer S, Chataway J, Coraddu F, Broadley S, Gray J, Jones HB, Clayton D, Goodfellow PN, Compston A.
Exploring the dense mapping of a region of potential linkage in complex disease: an example in multiple sclerosis. Genet Epidemiol. 1999; 17: 51-63.

Fiedorek FT Jr, Kay ES. Mapping of the focal adhesion kinase (Fadk) gene to mouse chromosome 15 and human chromosome 8. Mamm Genome. 1995; 6: 123-6.

Florian F, Hornigold N, Griffin DK, Delhanty JD, Sefton L, Abbott C, Jones C, Goodfellow PN, Wolfe J. The use of irradiation and fusion gene transfer (IFGT) hybrids to isolate DNA clones from human chromosome region9q33-q34. Somat Cell Mol Genet. 1991; 17: 445-53.

Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, Fu G, Saitou N, Osoegawa K, de Jong PJ, Suto Y, Hattori M, Sakaki Y. Construction and analysis of a human-chimpanzee comparative clone map. Science. 2002; 295: 131-4.

Gautier M, Hayes H, Eggen A. An extensive and comprehensive radiation hybrid map of bovine Chromosome 15: comparison with human Chromosome 11. Mamm. Genome 2002; 13: 316-319.

Gautier M, Hayes H, Bonsdorff T, Eggen A. Development of a comprehensive comparative radiation hybrid map of bovine chromosome 7 (BTA 7) versus human chromosomes 1 (HSA 1), 5 (HSA 5) and 19 (HSA 19). Cytogenet. Genome Res. 2003; 102: 25-31.

Geisler R, Rauch GJ, Baier H, van Bebber F, Bross L, Dekens MP, Finger K, Fricke C, Gates MA, Geiger H, Geiger-Rudolph S, Gilmour D, Glaser S, Gnugge L, Habeck H, Hingst K, Holley S, Keenan J, Kirn A, Knaut H, Lashkari D, Maderspacher F, Martyn U, Neuhauss S, Neumann C, Nicolson T, Pelegri F, Ray R, Rick JM, Roehl H, Roeser T, Schauerte HE, Schier AF, Schonberger U, Schonthaler HB, Schulte-Merker S, Seydler C, Talbot WS, Weiler C, Nusslein-Volhard C, Hafft r P. A radiation nhybrid map of the zebrafish genome. Nat. Genet. 1999; 23:86-9.

Georges M, Andersson L. Livestock genomics comes of age. Genome Res. 1996; 6: 907-21.

Georges M. Towards marker assisted selection in livestock. Reprod. Nutr. Dev. 1999; 39: 555-61.

Gerald PS, Brown JA. Proceedings: Report of the Committee on the Genetic Constitution of the X Chromosome. Cytogenet. Cell. Genet. 1974; 13: 29-34.

Goedert M, Marsh S, Carter N. Localization of the human kinesin light chain gene (KNS2) to chromosome14q32.3byfluorescenceinsituhybridization. Genomics. 1996; 32: 173-5.

Goldammer T, Kata SR, Brunner RM, Dorroch U, Sanftleben H, Schwerin M, Womack JE. A comparative radiation hybrid map of bovine chromosome 18 and homologous chromosomes in human and mice. Proc. Natl. Acad. Sci. U S A. 2002; 99:.2106-2111.

Goring HH, Terwilliger JD. Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. Am J Hum Genet. 2000; 66:1298-309.

Gorni C, Williams JL, Heuven HCM, Negrini R, Valentini A, van Eijk MJT, Waddington D, Zevenbergen M, Ajmone Marsan P, Peleman JD. Application of AFLP®1 technology to radiation hybrid mapping. Chromosome Res. 2004; 12: 285-297.

Goss SJ, Harris H. New method for mapping genes in human chromosomes. Nature 1975; 225: 680-684.

Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, Burridge PW, Cox TV, Fox CA, Hutton RD, Mullenger IR, Phillips KJ, Smith J, Stalker J, Threadgold GJ, Birney E, Wylie K, Chinwalla A, Wallis J, Hillier L, Carter J, Gaige T, Jaeger S, Kremitzki C, Layman D, Maas J, McGrane R, Mead K, Walker R, Jones S, Smith M, Asano J, Bosdet I, Chan S, Chittaranjan S, Chiu R, Fjell C, Fuhrmann D, Girn N, GR C, Guin R, Hsiao L, Krzywinski M, Kutsche R, Lee SS, Mathewson C, McLeavy C, Messervier S, Ness S, Pandoh P, Prabhu AL, Saeedi P, Smailus D, Spence L, Stott J, Taylor S, Terpstra W, Tsai M, Vardy J, Wye N, Yang G, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Shvartsbeyn A, Gebregeorgis E, Krol M, Russell D, Overton L, Malek JA, Holmes M, Heaney M, Shetty J, Feldblyum T, Nierman WC, Catanese JJ, Hubbard T, Waterston RH, Rogers J, de Jong PJ, Fraser CM,

Marra M, McPherson JD, Bentley DR. A physical map of the mouse genome. Nature 2002; 418: 743-750.

Grisart B, Farnir F, Karim L, Cambisano N, Kim JJ, Kvasz A, Mni M, Simon P, Frere JM, Coppieters W, Georges M. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. U S A. 2004; 101: 2398-403.

Haley CS. Livestock QTLs--bringing home the bacon? Trends Genet. 1995; 11: 488-92.

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN. A radiation hybrid map of the human genome. Hum. Mol. Genet. 1996; 5: 339-46.

Heddle JA. Randomness in the formation of radiation-indued chromosome aberrations. Genetics. 1965; 52:1329-34.

Hoskins RA, Nelson CR, Berman BP, Laverty TR, George RA, Ciesiolka L, Naeemuddin M, Arenson AD, Durbin J, David RG, Tabor PE, Bailey MR, DeShazo DR, Catanese J, Mammoser A, Osoegawa K, de Jong PJ, Celniker SE, Gibbs RA, Rubin GM, Scherer SE. A BAC-based physical map of the major autosomes of Drosophila melanogaster.Science. 2000 ; 287: 2271-4.

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, Hu X, Colbert AM, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA, Nahf RW, Horton LT Jr, Anderson MO, Collymore AJ, Ye W, Kouyoumjian V, Zemsteva IS, Tam J, Devine R, Courtney DF, Renaud MT, Nguyen H, O'Connor TJ, Fizames C, Faure S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES. An STS-based map of the human genome. Science. 1995; 270: 1945-54.

Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennet GL, Sugimoto Y. A Comprehensive genetic Map of the Cattle Genome Based on 3802 Microsatellites. Genome Res. 2004; 14: 1978-1998.

Inazawa J, Sasaki H, Nagura K, Kakazu N, Abe T, Sasaki T. Precise localization of the human gene encoding cell adhesion kinase beta (CAK beta/PYK2) to chromosome 8 at p21.1 by fluorescence in situ hybridization. Hum Genet. 1996; 98:508-10.

Itoh T, Takasuga A, Watanabe T, Sugimoto Y. Mapping of 1400 expressed sequence tags in the bovine genome using a somatic cell hybrid panel. Anim. Genet. 2003; 34: 362-370.

Itoh T, Watanabe T, Ihara N, Mariani P, Beattie CW, Sugimoto Y, Takasuga A. A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. Genomics 2005; 85:413-424.

Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP, Lopez-Corrales NL, Beattie CW. A second-generation linkage map of the bovine genome. Genome Res. 1997; 7: 235-249.

Kumar S, Hedges SB. A molecular time scale for vertebrate evolution. Nature. 1998; 392: 917-20.

Larkin DM, Everts-van der Wind A, Reibeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, Liu L, Osoegawa K, Womack JE, de Jong PJ, Lewin HA. Cattle-Human Comparative Map Built with Cattle BAC-Ends and Human Genome Sequence. Genome Res. 2003; 13: 1966-1972.

Leach RJ, O'Connel P. Mapping of Mammalian Genomes with Radiation (Goss and Harris) Hybrids. Advances in Genetics 1995; 33: 63-103.

Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Moore SS. The identification of common haplotypes on bovine chromosome 5 within commercial lines of Bos taurus and their associations with growth traits. J. Anim. Sci. 2002; 80: 1187-94.

Lunetta KL, Boehnke M. Multipoint radiation hybrid mapping: comparison of methods, sample size requirements, and optimal study characteristics. Genomics. 1994; 21: 92-103.

McCarthy L. Whole genome radiation hybrid mapping. Comment 1996; 12: 491-493.

McCarthy LC, Terrett J, Davis ME, Knights CJ, Smith AL, Critcher R, Schmitt K, Hudson J, Spurr NK, Goodfellow PN. A first-generation whole genome-radiation hybrid map spanning the mouse genome. Genome Res. 1997; 7: 1153-61.

MacNeil MD, Grosz MD. Genome-wide scans for QTL affecting carcass traits in Hereford x composite double backcross populations. J. Anim. Sci. 2002; 80: 2316-24.
McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, HumphR SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, T 1 rask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H; International Human Genome Mapping Consortium. A physical map of the human genome. Nature 2001; 409: 934-941.

Morisson M, Jiguet-Jiglaire C, Leroux S, Faraut T, Bardes S, Feve K, Genet C, Pitel F, Milan D, Vignal A. Development of a gene-based radiation hybrid map of chicken Chromosome 7 and comparison to humanandmouse. Mamm Genome. 2004; 15: 732-9.

Morton Ne. The inheritance of human birth weight. Ann Hum Genet. 1955; 20: 125-34.

Nadkarn P. Mapmerge: merge genomic maps. Bioinformatics 1998;14:310-6.

Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen KK, Svendsen M, Meuwissen TH. Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. Genetics 2005; 169:2 75-83.

Priat C, Hitte C, Vignaux F, Renier C, Jiang Z, Jouquand S, Cheron A, Andre C, Galibert F. A whole-genome radiation nhybrid map of the dog genome. Genomics. 1998; 54: 361-78.

Ranz JM, Casals F, Ruiz A. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus Drosophila. Genome Res. 2001; 11: 230-9.

Rebeiz M, Lewin HA. Compass of 47,787 cattle ESTs. Anim Biotechnol. 2000; 1: 75-241.

Rexroad CE, Schlapfer JS, Yang Y, Harlizius B, Womack JE. A radiation hybrid map of bovine chromosome one. Anim Genet. 1999; 30:325-332.

Rexroad CE 3rd, Owens EK, Johnson JS, Womack JE. A 12,000 rad whole genome radiation hybrid panel for high resolution mapping in cattle: characterization of the centromeric end of chromosome 1. Anim. Genet. 2000; 31: 262-5.

Risch N, Giuffra L. Model misspecification and multipoint linkage analysis. Hum Hered. 1992; 42: 77-92.

Ruddle FH. Linkage analysis using somatic cell hybrids.Adv. Hum. Genet. 1972; 30: 173-235.

SantaLucia JJ, Allawi HT, Seneviratne PA. Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 1996; 35: 3555-3562.

Schiebler L, Roig A, Mahé MF, Save JC, Gautier M, Taourit S, Boichard D, Eggen A, Cribiu EP. A first generation bovine BAC-based physical map. Genet. Sel. Evol. 2004; 36: 105-122.

Schiex T, Gaspin C. Carthagene: constructing and joining maximum likelihood genetic maps. In Proceedings of ISMB'97, Halkidiki, Greece Porto Carras 1997: 258–267.

Schlapfer J, Stahlberger-Saitbekova N, Comincini S, Gaillard C, Hills D, Meyer RK, Williams JL, Womack JE, Zurbriggen A, Dolf G. A higher resolution radiation hybrid map of bovine chromosome 13. Genet. Sel. Evol. 2002; 34: 255-67.

Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet. 2005; 36: 408-16.

Siden TS, Kumlien J, Schwartz CE, Rohme D. Radiation fusion hybrids for human chromosomes 3 and X generated at various irradiation doses. Somat. Cell. Mol. Genet. 1992; 18: 33-44

Slonim D, Kruglyak L, Stein L, Lander E. Building human genome maps with radiation hybrids. J. of Computational Biology 1997; 4: 487-504.

Snelling WM, Gautier M, Keele JW, Smith TP, Stone RT, Harhay GR, Bennet GL, Ihara N, Takasuga A, Takeda H, Sugimoto Y, Eggen A. Integrating linkage and radiation hybrid mapping data for bovine chromosome 15. BMC Genomics 2004; 5:77-90.

Soderlund C, Lau T, Deloukas P. Z extensions to the RHMAPPER package. Bioinformatics 1998; 14: 538-9.

Steen RG, Kwitek-Black AE, Glenn C, Gullings-Handley J, Van Etten W, Atkinson OS, Appel D, Twigger S, Muir M, Mull T, Granados M, Kissebah M, Russo K, Crane R, Popp M, Peden M, Matise T, Brown DM, Lu J, Kingsmore S, Tonellato PJ, Rozen S, Slonim D, Young P, Jacob HJ, et al. A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. Genome Res. 1999; 9:793.

Stone RT, Pulido JC, Duyk GM, Kappes SM, Keele JW, Beattie CW. A small-insert bovine genomic library highly enriched for microsatellite repeat sequences. Mamm. Genome 1995; 6: 714-24.

Sun S, Murphy WJ, Menotti-Raymond M, O'Brien SJ. Integration of the feline radiation hybrid and linkage maps. Mamm Genome 2001; 12:436-41.
Thomas JW, Prasad AB, Summers TJ, Lee-Lin SQ, Maduro VV, Idol JR, Ryan JF, Thomas PJ, McDowell JC, Green ED. Parallel construction of orthologous sequence-ready clone contig maps in multiple species. Genome Res. 2002; 12:1277-85

Walter MA, Spillett DJ, Thomas P, Weissenbach J, Goodfellow PN. A method for constructing radiation hybrid maps of whole genomes. Nat. Genet. 1994; 7: 22-8.

Warren W, Smith TPL, Rexroad III CE, Fahrenkrub SC, Allison T, Shu CL, Catanese J, De Long PJ. Construction of and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. Mamm. Genome 2000; 11: 662-663.

Watanabe TK, Bihoreau MT, McCarthy LC, Kiguwa SL, Hishigaki H, Tsuji A, Browne J, Yamasaki Y, Mizoguchi-Miyakita A, Oga K, Ono T, Okuno S, Kanemoto N, Takahashi E, Tomita K, Hayashi H, Adachi M, Webber C, Davis M, Kiel S, Knights C, Smith A, Critcher R, Miller J, Thangarajah T, Day PJ, Hudson JR Jr, Irie Y, Takagi T, Nakamura Y, Goodfellow PN, Lathrop GM, Tanigami A, James MR. A radiation hybrid map of the rat genome containing 5,255 markers.Nat Genet. 1999; 22: 27-36.

Weikard R, Goldmammer T, Laurent P, Womack JE, Kuehn C. A gene-based high-resolution comparative radiation hybrid map as a framework for genome sequence assembly of a bovine chromosome 6 region associated with QTL for growth, body composition, and milk performance traits. BMC Genomics 2006; 7: 53-67.

Wheelan SJ, Church DM, Ostell JM. Spidey: a tool for mRNA-to-genomic alignments. Genome Res. 2001; 11: 1952-1957.

Williams JL, Eggen A, Ferretti L, Farr J, Gautier M, Amati G, Ball G, Caramorr T, Critcher R, Costa S, Hextall P, Hills D, Jeulin A, Kiguwa SL, Ross O, Smith AL, Saunier K, Urquhart B, Waddington D. A bovine whole-genome radiation hybrid panel and outline map. Mamm. Genome 2002; 13:469-474.

Womack JE. Advances in livestock genomics: Opening the barn door. Genome Res. 2005; 15: 1699-1705.

Womack JE, Johnson JS, Owens EK, Rexroad CE, Sclapfer J, Yang YP. Mamm. Genome 1997; 8: 854-856.

Womack JE, Moll YD. Gene map of the cow: conservation of linkage with mouse and man. J Hered. 1986; 77: 2-7.

Yang YP, Rexroad CE 3rd, Schlapfer J, Womack JE. An integrated radiation hybrid map of bovine chromosome 19 and ordered comparative mapping with human chromosome 17. Genomics. 1998; 48: 93-9.

Yerle M, Lahbib-Mansais Y, Mellink C, Goureau A, Pinton P, Echard G, Gellin J, Zijlstra C, De Haan N, Bosma AA, et al. The PiGMaP consortium cytogenetic map of the domestic pig (*Sus scrofa domestica*). Mamm. Genome. 1995; 6: 176-86.

Yerle M, Pinton P, Delcros C, Arnal N, Milan D, Robic A. Generation and characterization of a 12,000-rad radiation hybrid pane for fine mapping in pig. Cytogenet. Genome Res. 2002; 97: 219-28.

Zhu B, Smith JA, Tracey SM, Konfortov BA, Welzel K, Schalkwyk LC, Lehrach H, Kollers S, Masabanda J, Buitkamp J, Fries R, Williams JL, Miller JR. A 5x genome coverage bovine BAC library: production, characterization,anddistribution. Mamm. Genome. 1999;10: 706-9.