Bovine Genomics: challenges and applications

R. Fitzpatrick1, O. Casey1,2 and D. E. MacHugh3
1Teagasc Research Centre, Athenry, Co. Galway
2Department of Microbiology, National University of Ireland, Galway
3Department of Animal Science and Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4

Introduction
Recent advances in high through-put DNA sequencing, gene expression technologies and bioinformatics combined with improved gene mapping procedures are having a major impact on genomic research. The complete genome sequence of more than 160 organisms ranging from humans to mice and representative plant and prokaryotic species have now been determined (GOLD, 2004). These sequences, together with advances in functional and structural genomics, offer scientists an unprecedented opportunity to accelerate gene discovery and to determine how gene function directs biological activity. In 2003 the entire three billion base pairs encoding the human genome was completed. It is now known that the human genome encodes up to 40,000 unique genes (Xuan et al., 2003) and this number is increasingly being revised upwards as more sophisticated gene identification procedures are developed and applied. To date less than 50% of the 40,000 genes discovered have been named and have had their function determined. Some of these genes and their encoded sequences have been identified and associated with numerous diseases including cancer, muscle disease, deafness and blindness (Maubaret et al., 2002, Balmain et al., 2003). Approximately three million single nucleotide polymorphisms (SNPs) have also been identified, facilitating high through-put SNP screening of DNA sequences associated with such common diseases as cardiovascular disease, diabetes and asthma (Herbon et al., 2003). Beyond these more applied areas of study, genomic research also presents opportunities for increasing basic knowledge of the underlying biochemical, metabolic and molecular mechanisms controlling tissue physiology and biological function.

Bovine genomics resources
The rapid progress in genomic science and a glimpse into its potential applications has stimulated similar interest in developing bovine and other farm animal genomic research projects. The bovine genome is the best characterised of all the farm animals and contains 28 autosomes, two sex chromosomes and consists of ~3.1 x 10^9 base pairs of DNA encoding an estimated 30,000 to 40,000 genes (Archibald, 2000). The resources for bovine genomic research are well advanced and include representative BAC (bacterial artificial chromosome) and YAC (yeast artificial chromosome) genomic libraries, radiation and somatic cell hybrids, medium resolution linkage and physical maps, numerous cDNA libraries, high density cDNA arrays and over 250,000 expressed sequenced tags (ESTs). Many of these ESTs have been physically mapped to the bovine genome (Itoh et al., 2003). In addition, more than 2,000 polymorphic microsatellites and some 1,500 genes have been genetically and/or physically mapped (INRA, 2004). Perhaps the most exciting development in bovine genomics was the announcement in June 2002 by a leading biotechnology based company, Metamorphix, that they had sequenced the entire bovine genome to ~1X coverage and generated >600,000 SNPs (Adam, 2002). These sequences will be of immense value in identifying genetic markers associated with economically important traits. In addition, separate public efforts to independently sequence the bovine genome were officially launched in 2003 by an international consortium of scientists from the United States, France and Canada, costing an estimated $50 million (Veneman, 2003).

Analysis of bovine traits
The relationship between genotype and phenotype is complex. In cattle, some traits such as coat colour, horn development, double muscling and certain diseases are determined by genotype and are controlled by one or at most a few genes and are termed qualitative traits (Georges et al., 1996). In contrast, complex or quantitative traits such as carcass quality, growth rate, milk yield and composition, disease resistance, and fertility are multifactorial, and are influenced by both environmental parameters and multiple genes acting in concert (Willis 1998). Some of these traits such as disease resistance and fertility have low heritabilities and are difficult to improve by classical breeding. Furthermore, some traits are difficult to measure. For example, evaluation of carcass quality typically involves destructive tests whereas quantifying disease resistance requires complex challenge models that can be difficult to interpret. The accurate measurement of phenotypic traits and how they vary within and between animal populations is an essential component of genomic research (Andersson, 2001).

Considerable progress has been made in the identification and mapping of chromosomal segments termed quantitative trait loci (QTLs) that harbour genes influencing economically important production traits that could be exploited by beef and dairy breeders (Georges et al., 1995, Casas et al., 1998, Smith et al., 2000). However, the full value of these QTLs will only be realised when the actual genes and polymorphisms responsible for the specific trait(s) are identified and understood. Characterisation of gene expression and the factors that alter gene expression is fundamental to understanding the cellular processes that dictate phenotype. The expression and function of individual genes may be altered by mutations in coding and, or, regulatory regions and environmental and physiological stimuli. One of the principle objectives of bovine genomics is therefore to determine the location, structure, function, and factors effecting the expression of genes involved in health, reproduction, production and product quality.

Genomics can be divided into two main areas:- i) the sequencing and genetic and physical mapping of entire genomes (structural genomics) and ii) the analysis of global patterns of gene expression (functional genomics). A combination of these approaches is now being used to identify and characterise gene(s) controlling traits important to the dairy and beef industries.

109
**Structural Genomics**

The study of genome organisation and variation—from chromosomal rearrangements all the way down to the primary DNA sequence is termed structural genomics. The field encompasses an array of technologies and activities including the construction of genetic linkage maps, physical maps and comparative maps, the final endpoint being the production of the ultimate genetic map—the fully annotated genome sequence of the organism of interest. During the last fifteen years, the technologies underpinning structural genomics have been revolutionised by major advances in nucleic acid biochemistry, laboratory automation and computational biology. For example, during the mid to late 1980’s many teams of scientists worked flat-out for many months or years to pinpoint the location of single genes responsible for simple human monogenic disorders such as cystic fibrosis or Duchenne muscular dystrophy (Koenig et al., 1987, Rommens et al., 1989). Using current technologies, these once ground-breaking linkage mapping studies could be repeated by a single graduate student in a matter of days (Cullis, 2002). More significantly, complete or draft genome sequences have now been published for a number of mammalian species including human, mouse, rat and dog and the domestic cow will join this exalted group within the next few years (Couzin, 2003).

During the last ten years major efforts have been underway to develop a high density genetic linkage map for cattle. At the present time (early 2004), the cattle genome mapping database contains mapping information for over 4,000 genetic loci, of which more than 1,500 represent coding genes (INRA, 2004). Dense genetic linkage maps such as the current bovine genome map can be used in conjunction with appropriate cattle mapping families to identify chromosomal segments that contain molecular polymorphisms responsible for relatively straightforward single-gene traits such as presence or absence of horns (polled gene) (Georges et al., 1993) or quantitative trait loci (QTLs) that contribute to complex, partially genetic multifactorial traits such as milk yield, growth rate or meat quality (Andersson, 2001). Two case studies are presented to illustrate how structural and comparative genomics can be applied to elucidate the genetic basis of economically important cattle traits.

A single-gene trait: the myostatin gene and the double-muscling phenotype in beef cattle

The Belgian Blue beef breed is well-known for displaying a form of hypertrophy (overgrowth) called double-muscling. This Mendelian trait is also observed in some other European beef breeds including Charolais, Piedmontese, Gasconne and Parthenais. Double-muscled cattle have leaner carcasses and exhibit greater muscle mass with less fat than non-doubled muscled cattle (Hanset et al., 1987). Despite these obvious advantages, double-muscling has one major drawback—a greatly increased incidence of calving difficulty, such that Caesarian sections are the rule for term deliveries in these breeds. In 1995, a team led by Michel Georges at the University of Liege (Charlier et al., 1995) performed one of the first large-scale genome mapping studies in cattle and showed that the mh (muscular hypertrophy) locus is located on BTA2 (Bos taurus chromosome 2). Using a panel of 213 hypervariable microsatellite markers and a three-generation pedigree obtained by back-crossing mh/+ heterozygotes (produced by crossing double-muscle mh/mh Belgian Blue sires with normal +/- Friesian dams) to double-muscled mh/mh sires, Charlier and colleagues (1995) were able to pinpoint the location of the causative gene close to the centromeric end of BTA2 and also showed that a monogenic model for the double-muscling trait was correct. Parallel studies in mice revealed that knockout mutants of the myostatin gene developed extreme muscularity, comparable to the double-muscling phenotype in cattle (McPherron et al., 1997). These studies also revealed that the myostatin protein acts as a negative regulator of skeletal muscle growth. Subsequently, a number of research groups showed that the bovine homolog to the mouse myostatin gene maps to the same segment of BTA2 as the mh locus and that double-muscled cattle are homozygous for a series of loss-of-function mutations in the myostatin gene (Grobet et al., 1997, Kambadur et al., 1997, McPherron & Lee, 1997). These research efforts have subsequently led to relatively low cost polymerase chain reaction (PCR) based diagnostic tests that can be used for marker-assisted selection for or against the double-muscling trait (for example, the GenMARK Piedmontese myostatin test: www.genmarkag.com).

A multifactorial trait: the search for milk yield quantitative trait loci (QTLs)

During the last decade whole genome scans for QTLs contributing to production traits have become commonplace and a wide range of QTLs have been identified for performance traits in domestic cattle (for reviews see Andersson, 2001, Schwerin, 2001). A general overview of the resources and approaches used in QTL analysis are shown in Fig. 1. One of the first milk performance QTL mapping studies, carried out almost a decade ago, elegantly overcame the limitations of the primitive cattle genome maps available at that time by taking advantage of detailed progeny testing in elite Holstein-Friesian populations (Georges et al., 1995).

![Fig. 1. A general overview of the resources and approaches used in QTL analysis](image-url)
In dairy cattle, young sires produced from planned matings of sires and dams with the highest breeding values (BV) are genetically evaluated, based on the milking performance of 50-100 of their daughters. The milking records of the daughters are now normally collected as part of nationwide record-keeping systems. Because this system is widely implemented, it is straightforward to identify pedigrees and collect DNA from large sets of progeny-tested paternal half-brothers. The experimental design, referred to as the granddaughters design (GDD) uses a molecular genome scan to map QTLs underlying milk production traits. Marker genotyping and linkage analysis are performed in the sons, using averages of their respective daughter phenotypes as quantitative measurements. Georges et al., (1995) identified 14 bulls with 33-208 sons each (1,518 in total) with performance data from 50-100 daughters per son. They subsequently genotyped 159 polymorphic microsatellites in all of the sons—forming a map covering two-thirds of the bovine genome. Figure 2 shows a schematic illustrating this experimental design. Statistical analyses of these data in conjunction with their daughters’ lactation records identified five QTLs (on BTA1, 6, 9, 10 and 20) that had a significant effect on one or more of the traits studied including milk, fat and protein yield.

Significant advances have been achieved since the pioneering work of Georges and his co-workers and a plethora of QTLs have now been described for dairy, beef, disease and welfare traits. Most notably, recent genomic studies have identified -quantitative trait nucleotides (QTNs) (Mackay, 2001)—polymorphisms directly responsible for the statistically detected QTLs in dairy populations. For example, Blott et al., (2003) have demonstrated that the underlying mutation responsible for the BTA20 QTL originally detected using the GDD in 1995 (see above) is actually an A to T transversion that causes a phenylalanine to tyrosine substitution in the transmembrane domain of the growth hormone receptor (GHR) gene. In addition, Grisart and co-workers have pinpointed the causative QTN for a BTA14 QTL with a major effect on milk fat content as an Apa to GspC dinucleotide substitution that causes a lysine to alanine substitution in the diacylglycerol acyltransferase (DGAT1) gene (Grisart et al., 2002). These recent findings mark the beginning of the next phase of livestock genomics and point towards rational marker-assisted selection (MAS) for complex production traits using polymorphisms within genes directly affecting the trait of interest.

**Functional Genomics**

Many of the production and quality traits important to the dairy and beef industries are controlled by multiple gene expression events in a variety of different tissues making these traits difficult to study using traditional molecular techniques. Large scale gene identification surveys and high through-put gene expression technologies are now being used to improve our understanding of the molecular mechanisms controlling complex traits. A general overview of the resources, approaches and potential impacts of functional genomics is shown in Fig. 3.

**Fig. 2.** Schematic outlining the granddaughter Design (GDD) used to identify chromosomal regions harbouring QTLs influencing milk performance in elite Holstein-Friesian cattle.

**Fig. 3.** General resources, approaches and potential impacts of functional genomics

*Generation and analysis of bovine expressed sequence tags (ESTs)*

The EST strategy is widely used as a rapid and cost effective way of discovering new genes as well as providing useful resources for gene mapping and cDNA array construction. Expressed sequence tags are generated by single pass 5’ or 3’ DNA sequencing of clones randomly picked from cDNA libraries. The identity or function of the EST is then obtained by searching for homologous sequences in the public DNA and protein databases using BLAST (basic local alignment search tool) located at the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST/). Putative identification is typically given to EST sequences which have an E-value <10^-5 which is the statistical probability that the match occurred by chance. This EST approach has been used to accelerate gene discovery in a variety of bovine tissues including the ovary (Ma et al., 1998, Casey et al., 2003), mammary gland (Sonstegard et al., 2002), embryo (Ponsuksili et al., 2002), foetus (Taniguchi et al., 2001) and a generic cDNA library constructed from a variety of economically relevant tissues (Smith et al., 2001). To manage the exponential increase in bovine ESTs, specialised databases [TIGR (www.tigr.org/tgdh/) and UniGene (www.ncbi.nlm.nih.gov/UniGene)] have been developed to generate non-redundant sets of bovine
ESTs including data on their expression patterns, cellular functions and evolutionary relationships. The current number of bovine ESTs deposited in Genbank exceeds 250,000, representing over 53,000 non-redundant clones (TIGR). Several groups have established the chromosomal location of many of these ESTs (Karall-Albrecht et al., 2000, Goldammer et al., 2002). In one study a total of 1,400 ESTs were mapped to the bovine genome using a somatic cell hybrid panel (Itoh et al., 2003). Increasing the marker density of the bovine genome has important downstream applications in QTL analysis and comparative genomic studies. In addition a number of ESTs have been associated with economically important traits including milk production and energy metabolism (Durroch et al., 2001, Looft et al., 2001).

**Developments in bovine DNA array technology**

All bovine cells with the exception of red blood cells and certain reproductive cells contain the capacity to express the 30,000 to 40,000 genes encoded by the bovine genome. However, only a subset of these genes (~12,000) is expressed in any one individual cell resulting in a variety of specialised and functionally diverse tissues. A major challenge for functional genomics is to relate global patterns of gene expression with phenotype and to identify genes that are trait-associated. The application of DNA array technology is proving to be an effective approach for analysing differential gene expression in tissues isolated from animals exhibiting different physiological characteristics. One of the main advantages of array technology over traditional methods for studying gene expression is its ability to quantify the expression of thousands of genes simultaneously in a single experiment. The existing bovine arrays, are cDNA based, printed on either glass or nylon support. The basic steps for analysing gene expression using cDNA arrays involve i) labelling the experimental mRNA with fluorescent dyes (glass arrays) or radioactive nucleotides (nylon arrays), ii) hybridisation of the labelled mRNA to the cDNA array, iii) image analysis using a fluorescent scanner (glass arrays) or a phosphor imager (nylon arrays), iv) identification and quantification of signal intensities using specialised bioinformatic software programs, v) verification of selected differentially expressed genes using Northern blot analysis or Real Time RT-PCR. cDNA arrays have been used to compare gene expression profiles in a variety of bovine tissues under different physiological states including, uterine/liver tissue isolated from pregnant and non-pregnant animals (Ishiwata et al., 2003 and Herath et al., 2004), regressed versus functional corpus luteum tissue (Casey et al., 2004), adult versus fetal spleen tissue (Band et al., 2002), lactating versus non-lactating mammary tissue (Suchyta et al., 2003a) and in neutrophils isolated from parturient versus non-parturient dairy cows (Madsen et al., 2004). In this way differentially expressed genes or groups of genes were identified providing new insights into the molecular and biochemical pathways controlling tissue function and how this is modified by physiological processes. As the resources for bovine functional genomics continue to be developed such as the recently constructed 18,263 generic cDNA array (Suchyta et al., 2003b) it is likely that this technology will be increasingly used to gain a better understanding of:- i) the moderating influences of physiological processes on gene expression and tissue function, ii) the effects of animal husbandry practices on gene expression and its impact on production and quality traits and iii) the identification of molecular markers controlling important trait characteristics.

**Current and future applications of bovine genomics**

Identification of genetic markers associated with economically important traits is a growing area of research that has considerable practical potential. A number of commercial DNA diagnostic tests for beef and milk quality traits have been developed and are summarised in Table 1. These tests allow the identification of animals with increased potential to produce quality beef and dairy products and are therefore important tools for improving genetic merit through marker assisted selection (MAS). PCR-based diagnostic tests for several congenital diseases have also being developed (Table 1.) providing the means to identify animal carriers and reduce incidences of inherited disorders. Within the next few years the entire sequence of the bovine genome should be determined expanding the potential to identify additional candidate genes regulating economically important traits. Traits most likely to benefit from DNA diagnostic tests and MAS include disease resistance, carcass quality, reproductive efficiency and growth rate.

Benefits arising from DNA array technology are likely to impact on many aspects of dairy and beef production. For example a better understanding of gene expression in metabolic, immune and reproductive tissues and how this is modified by the high energy demands of milk production should aid the development of new methodologies to combat recurring problems of metabolic and infections disease and reduced fertility in high yielding dairy cows. The effects of diet on gene expression and how this affects quality traits such as carcass composition, marbling and meat tenderness will help optimise production systems to improve product quality. Similarly, knowing which genes come into play when cattle are stressed by environmental factors will help develop new management systems to promote animal health and welfare.

The recent occurrence of BSE and foot and mouth demonstrated how deeply an economy can be affected by problems in the livestock industries and also highlighted some of the limitations of conventional animal identification systems. Developments in DNA technology now provide improved methods for identifying individual animals and their products. Using PCR and a set of polymorphic microsatellite markers a unique DNA profile can be obtained that is animal specific. Products derived from tested animals can then be unequivocally linked to an individual animal by comparing DNA profiles. A leading Irish food retailer is now using this technology to provide full traceability for over 100,000 carcasses thereby improving consumer confidence on the safety and origin of meat products (Cunningham and Meghen, 2001). Increased application of these methods will prove invaluable in future disease monitoring and eradication programs.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Trait</th>
<th>Company</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroglobulin</td>
<td>Marbling</td>
<td>Genetic Solutions</td>
<td><a href="http://www.geneticsolutions.com.au">www.geneticsolutions.com.au</a></td>
</tr>
<tr>
<td>Leptin</td>
<td>Marbling</td>
<td>Merial</td>
<td><a href="http://www.merial.com">www.merial.com</a></td>
</tr>
<tr>
<td>Calpastatin</td>
<td>Tenderness</td>
<td>Genetic Solutions</td>
<td><a href="http://www.geneticsolutions.com.au">www.geneticsolutions.com.au</a></td>
</tr>
<tr>
<td>μ-Calpain</td>
<td>Tenderness</td>
<td>Frontier Beef Systems</td>
<td><a href="http://www.frontierbeefsystems.com">www.frontierbeefsystems.com</a></td>
</tr>
<tr>
<td>Myostatin</td>
<td>Double muscling</td>
<td>GenMARK</td>
<td><a href="http://www.genmarkag.com">www.genmarkag.com</a></td>
</tr>
<tr>
<td>Argininosuccinate synthetase</td>
<td>Citrullinemia</td>
<td>GenMARK and Genetic Visions</td>
<td><a href="http://www.genmarkag.com">www.genmarkag.com</a></td>
</tr>
<tr>
<td>Uridine Monophosphate Synthase</td>
<td>Deficiency of Uridine Monophosphate Synthase</td>
<td>Genetic Visions and ImmGen</td>
<td><a href="http://www.geneticvisions.net">www.geneticvisions.net</a>, <a href="http://www.immgen.com">www.immgen.com</a></td>
</tr>
</tbody>
</table>

**Glossary**

**Allele** Alternative versions of a gene or other segment of a chromosome

**Autosome** Any chromosome other than the sex chromosomes

**BAC** Bacterial artificial chromosome: high capacity bacterial vector for cloning large segments of genomic DNA (100 to 300 kb)

**Bioinformatics** The study of genetic and other biological information using computer and statistical techniques

**BLAST** A computer program that identifies homologous genes in different organisms

**cDNA** A doubled stranded DNA copy of an mRNA transcript

**cDNA array** Collection of cDNAs printed on nylon or glass support used to detect and quantify expression of large number of genes simultaneously in a single experiment

**cDNA library** Collection of cDNA clones representing the entire complement of genes expressed in a particular cell or tissue

**Chromosome** thread like structures of DNA found in the cell nucleus on which genes are carried

**Comparative genomics** A research strategy that uses information obtained from one genome to make inferences about map position and functions of genes in a second genome

**DNA marker** A DNA sequence that exists as two or more versions used to mark a map position

**EST (Expressed sequence tag)** A unique DNA sequence (~500-700 bp) within a coding region of a gene: used for gene identification

**Functional genomics** The study of genomes to determine biological function of all genes and their products

**Gene expression** Conversion of biological information encoded by a gene, first into mRNA and then protein

**Genetic linkage map** A map of the relative position of genes and other regions on a chromosome

**Genetic profile** The banding pattern revealed after electrophoresis of PCR products directed at a range of microsatellite loci

**Genomics** The comprehensive study of whole sets of genes and their interactions

**Genotype** The genetic make-up of an animal

**Locus** Chromosomal location of a gene or other piece of DNA

**Microsatellite** Non-coding DNA sequences containing short motifs, repeated in tandem, microsatellite DNA are often polymorphic and are therefore useful in mapping studies.

**Mutation** An alteration in the nucleotide sequence of a DNA molecule

**PCR** Polymerase chain reaction: a technique for amplifying regions of DNA quickly and cheaply

**Physical map** Map location of identifiable markers spaced along the chromosomes

**Phenotype** Observable characteristics resulting from interaction between genotype and the environment

**Polymorphism** A variation in DNA sequence within a population

**Qualitative traits** Traits that are distinct and can be counted rather than measured typically controlled by one or a few genes and are often simply inherited

**Quantitative traits** Traits that are measured and which show continuous variation, controlled by many genes and environmental parameters

**Real Time RT-PCR** Procedure for detecting and quantifying individual gene expression at the mRNA level

**Sex chromosomes** Pair of chromosomes determining gender (XX female and XY male)
SNP Single nucleotide polymorphism: common, single base pair variations in DNA
Structural genomics The sequencing and mapping of entire genomes
Trait Characteristic or feature of an animal
YAC Yeast artificial chromosome: high capacity yeast vector for cloning large segments of genomic DNA (up to 1000 kb)

References
Chromosome Res. 10: 411-8


Veneman (2004). *USDA / Press Release No. 0420.03*
