BOVINE GENOMICS AND COMPARATIVE MAPPING TO IDENTIFY CANDIDATE GENES

PM Corva[1] and JF Medrano[2]
[1] Departamento de Producción Animal, Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata. C.C. 276 7620 Balcarce, Argentina,[2] Department of Animal Science, One Shields Ave., University of California, Davis, CA 95616-8521, U.S.A.

Key words: genetics, cattle genomics, gene mapping

In the last decade we have witnessed a dramatic change in the way biology is conceived. A new paradigm in science states that all genes in living organisms will be known, and at that stage investigation in biology will be merely a theoretical exercise (15). The predicted new era in biology is becoming a reality. Entire genomes have been sequenced and made available to the scientific community, including several microorganisms, yeast (*Saccharomyces cerevisiae*), worm (*Caenorhabditis elegans*), and fruit fly (*Drosophila melanogaster*). Sequencing of the human genome will be completed soon, while sequencing of the mouse genome is already advancing.

Despite this wealth of genetic information, the practical application to the field of animal science, in particular to the genetic improvement of beef and dairy cattle, will not be straightforward. In the case of economically important traits (E.I.T.) in cattle, at least two important limitations exist. Most E.I.T. in cattle are quantitative traits, regulated by many genes and influenced by environmental factors. As such, it is difficult to establish a direct correspondence between genotype and phenotype, which makes gene mapping and identification more difficult. Those chromosomal regions harboring genes that control quantitative traits have been called QTL (Quantitative Trait Loci). The other limiting step in the process of gene characterization will be the unequivocal assignment of function to a gene considered a candidate for a QTL (28).

The intrinsic characteristics of quantitative traits required the development of innovative strategies to map, clone and characterize genes responsible for E.I.T. in cattle. Those strategies will be reviewed here with special emphasis on the role of comparative genomics in the discovery of genes underlying E.I.T.

Linkage analysis

The first requisite for the search of genes underlying QTL is the development of linkage maps of molecular markers. Linkage maps comprise at least two main classes of markers (27). Type I markers are coding genes, that are usually conserved among species. Type II markers are anonymous, hypervariable DNA sequences evenly distributed on the genome. Contrary to Type I markers, Type II markers are highly polymorphic and suitable for a mapping project. However, they are not conserved among species. The construction of the map is based on recombination events between markers on the same chromosome (Figure 1). The occurrence of recombination between markers allows the estimation of their order and their distance (Figure 2). Because linkage analysis is based on meiotic recombination, genetic distances do not necessary reflect physical distances. Table 1 shows the current status of linkage maps in model organisms and domestic species.

Table 1: Available mapping information for selected species (27).

Species	Chromosome number (n)	Number of markers		Genome Length (cM)
		Type I	Type II	
Human	23	>30,000	~8,000	3,300
Mouse	20	6,992	7,377	1,450
Bovine[1]	30	697	2,170	2,990-3,552
Pig	19	369	~1,000	2,300-2,500

[1] BovMap (http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl)

Based on the availability of linkage maps, the traditional approach to QTL mapping has been to establish statistical associations between segregating markers and phenotypic differences of E.I.T. (Figure 3). The significance of those statistical associations indicates the most likely location of a QTL. In the case of dairy cattle, the structure of commercial populations with sires that are extensively used for artificial insemination, and the availability of reliable phenotypic information, has made it possible to map QTL for milk production (33). The most common experimental designs are the daughter design and the granddaugher design, which will be described in a companion paper. Attempts to map QTL in commercial populations of beef cattle are still limited. However, extensive mapping work has been done in experimental crossbred populations (4, 32).

The capacity of conventional mapping methods to resolve the position of a QTL is limited. Usually, the position of a locus is assigned to 20 to 30 cM intervals (7). This level of resolution precludes further progress toward gene identification for several reasons. First, it is not feasible to develop a physical map of such a large chromosomal fragment. Second, even if the sequence were available, it could potentially contain hundreds of genes. Novel strategies are being developed to overcome these limitations.

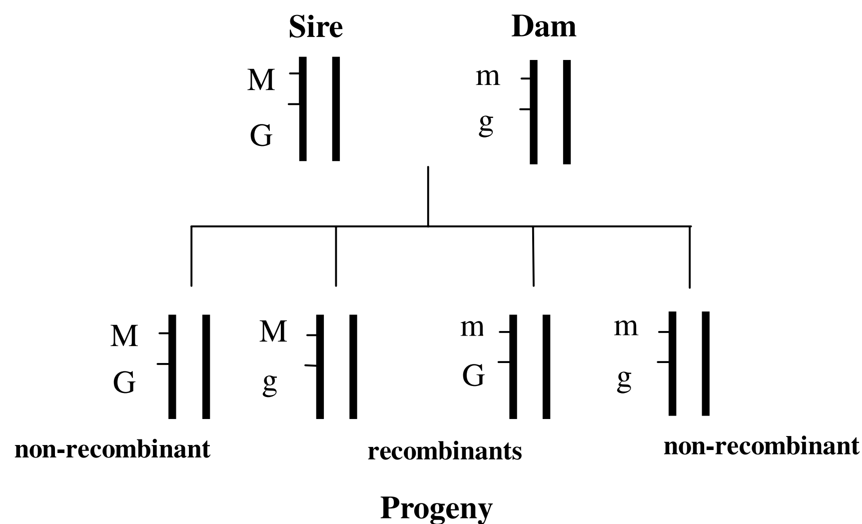

Figure 1: Example of linkage mapping. The rate of recombination (θ) between two markers (M and G in the example) is a function of the distance between them. By calculating the proportion of recombinants in a cross, one can estimate the genetic distance between two markers. For unlinked markers, $\theta = 0.5$. For linked markers, $\theta < 0.5$. The proportion of parental combinations in the progeny is equal to $(1 - \theta)$, while that of recombinants is equal to θ.

In theory, the development and mapping of new markers would help in increasing the resolution of QTL locations. However, such an approach would be limited at some point by the number of sampled meioses (10). In turn, the number of meioses depends directly on sample size (number of animals). Darvasi (9) estimated the sample size needed for high-resolution mapping of a QTL in mouse crosses to be in the order of thousands of animals. This makes the application of this strategy very unlikely in livestock. Coppieters et al. (7) have proposed to exploit the existence of "historical" recombination events in the mapping population instead of creating them by adding more animals. This approach is based on the principle of "identity by descent" (IBD) and it has been successfully applied to map a disease locus (mulefoot) (5) and QTL for milk production (29).

QTL mapping is considered the first stage in a mapping project that leads to positional cloning of genes underlying a phenotype of interest. Unlike functional cloning strategies, positional cloning methods were developed with the objective of identifying anonymous genes underlying complex phenotypes without previous knowledge about their physiological role (6, 20). However, the increasing availability of transcript maps for humans and mouse gave origin to a variant of that method, designated cloning by a positional candidate approach (6). The new method involves searching for genes that because of their location and their function could be assigned to a previously mapped QTL. As the number of genes placed on the bovine map increase, the comparative genomics approach will become more feasible.

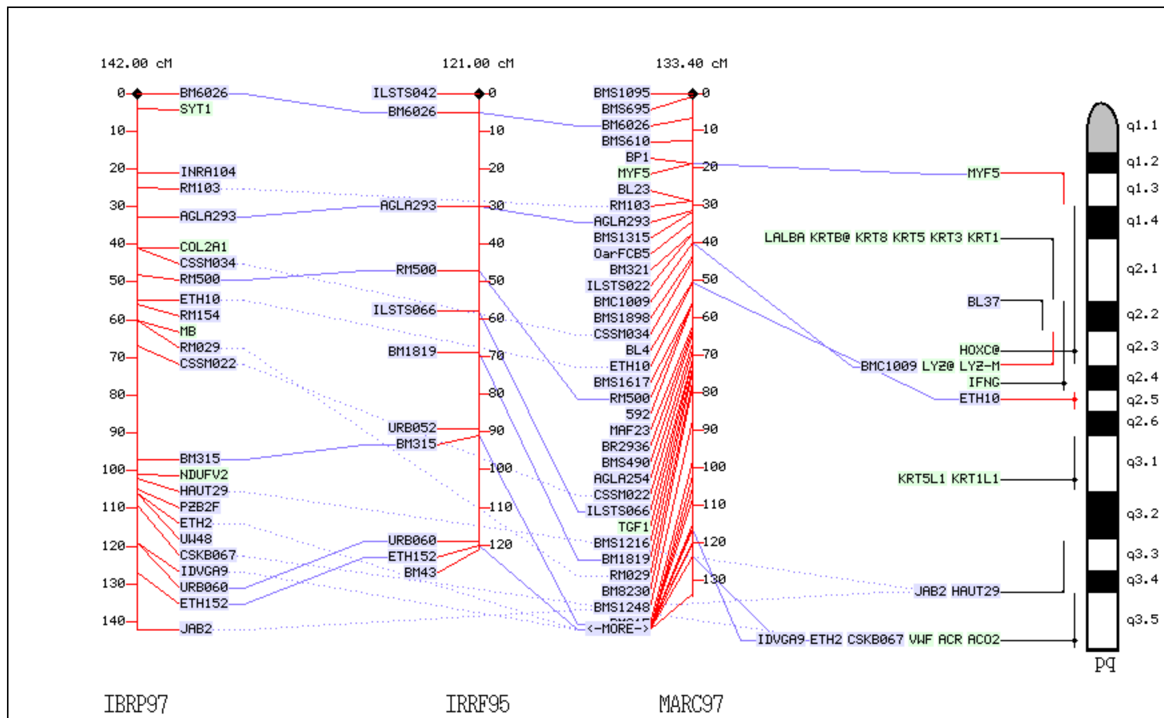

Figure 2: Linkage and cytogenetic maps of bovine chromosome 5. Three linkage maps made by different groups and using different mapping populations are presented: IBRP97 (International Bovine Reference Panel, http://spinal.tag.csiro.au/cgd.html), IRRF95 (Illinois Reference/Resource Families, http://cagst.animal.uiuc.edu/) and MARC97 (Roman L. Hruska Meat Animal Research Center, ARS-USDA, http://sol.marc.usda.gov/genome/cattle/cattle.html). Markers in common allow anchoring the different linkage maps and arriving to a consensus map of the chromosome. Source: Anubis database (http://www.ri.bbsrc.ac.uk/cgi-bin/mapviewer?species=cattle)

Comparative genomics

The approach of comparative genomics rests on the fact that all known species of mammals evolved from a common ancestor (27). The striking variability that those species show in terms of number and size of chromosomes is the result of cytogenetic events such as inversions, deletions, fusions and translocations that were part of the process of speciation (26). The consequence of this evolutionary process is that at present, the haploid number of chromosomes in mammals goes from three in the Indian muntjac (*Muntiacus muntjak*) to sixty-seven in the black rhino (*Diceros bicornis*) (27).

Another feature of this process of genome evolution is that variability among vertebrates seems to be based more on gene regulation than on gene creation or disappearance (1, 12, 26). This suggests that the same genes have similar functions in different species. Therefore, and despite the extent of variation in genome organization, it is possible to take advantage of the fact that mammals had a common origin to follow orthologous genes in different species. In turn, such information may increase the efficiency of mapping experiments since the information can be extrapolated from one genome to another.

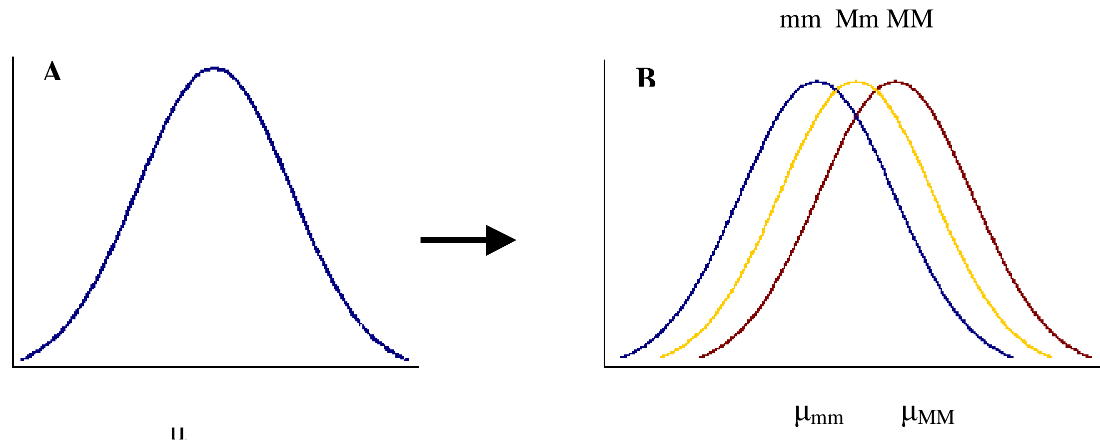

Figure 3: Classification of individuals from a mapping cross based on their genotype at a marker M. Graph A shows the distribution of phenotypes in the population. If a marker M is linked to a QTL (Graph B), a significant difference will be detected between the means of subpopulations with genotype MM and mm. The position of the curve corresponding to individuals with genotype Mm depends on the dominace effect of the QTL (in this example, d=0).

The more distant two species are, the smaller the chromosomal fragments conserved between them (26). In order to follow the chromosomal rearrangements between species and to make inferences about genome evolution, some kind of mapping information is needed. Extensive linkage information is available for humans and mouse but not for domestic species. Initially somatic-cell hybrids were used to build bovine-human synteny maps (35). Cytogenetic techniques such as chromosome painting have confirmed the correspondence among syntenic chromosomal segments (16). However, these maps do not allow comparing gene order.

A powerful tool to assign gene locations is radiation hybrid (RH) mapping (8). In this method, chromosome breakage is induced by x-ray treatment of cells of a donor species. These cells are fused to recipient rodent cells. A statistical procedure allows the estimation of the frequency of breakage between any two markers, which is a function of their distance. Because polymorphisms are not needed, Type I markers can be mapped and gene order compared (36). RH mapping has been successfully applied to the mapping of human ESTs (Expressed Sequence Tags). ESTs are fragments of anonymous cDNA sequences that were isolated from tissue-specific libraries and partially sequenced. Because of the great amount of information that the human RH map has accumulated over the years, it has become an obligate reference for mapping projects in any species. The mouse RH database is available at the Jackson Laboratory (http://www.jax.org/resources/documents/cmdata/rhmap/RHIntro.html). As for bovine, a radiation hybrid panel has been recently constructed (34). One of the most important advantages of sharing these resources is that individual laboratories can map their genes of interest in the RH panels and the resulting information integrated to the consensus map and made rapidly available to the scientific community. Detailed comparison of gene order permitted the development of comparative maps that show the homology among species (Figure 4). For the first time, comparative maps allowed the sharing of information among "map-rich" species, like humans, mouse and rat, and "map-poor" species, like most domestic species.

One of the first attempts to facilitate the task of comparative mapping among species was the development of a universal set of primers that would amplify a fairly large number of Type I markers (genes). These markers have been denominated CATS (Comparative Anchor Tagged Sequences) (22). Based on the same concept, Jiang et al. (17) developed TOASTs (Traced Orthologous Amplified Sequence Tags) to be used in pig mapping experiments.

In order to compare gene order in bovines and other species the method called "Parallel Radiation Hybrid Mapping" was developed (36). The method is based on simultaneous mapping of a set of orthologous genes in human and bovine RH panels. Application of this method to human chromosome 17 and bovine chromosome 19 demonstrated that although both chromosomes are syntenic, extensive rearrangement has taken between them (Figure 5).

Another strategy to map bovine genes and ESTs is called COMPASS (Comparative Mapping by Annotation and Sequence Similarity) (23). The position of bovine ESTs is predicted using comparative mapping information from genomic databases. The predicted position is then confirmed by RH mapping.

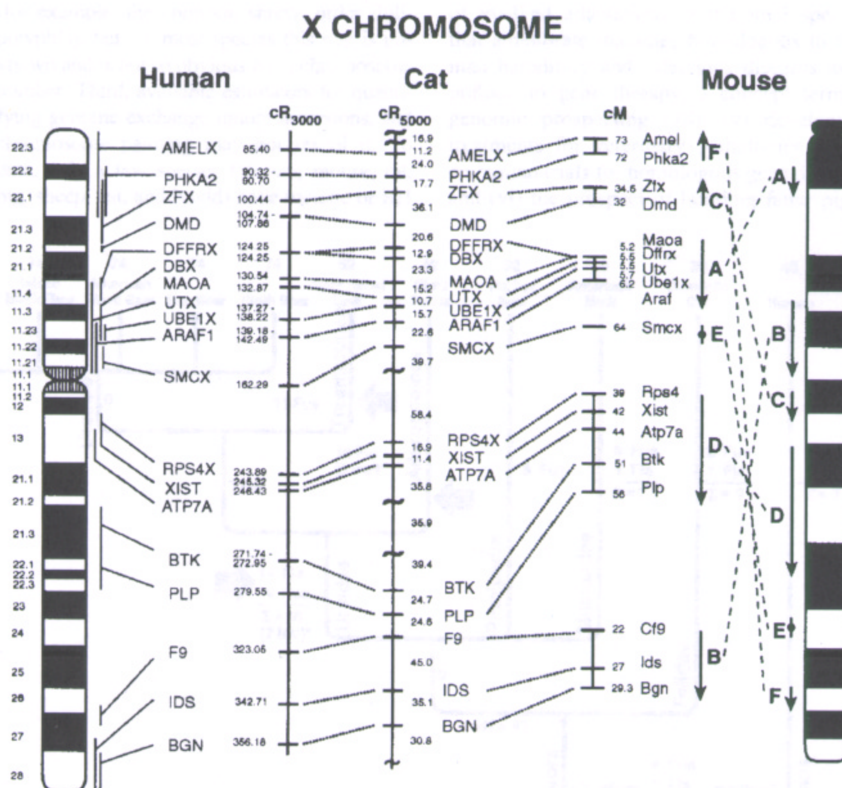

Figure 4: Comparative map of Chromosome X between human, cat and mouse (27). Gene order is identical between human and cat. Arrows indicate the polarity of mouse gene order.

The application of comparative genomics is not restricted to humans and mice. Useful information can be recovered from other model organisms. One interesting example is the *Tetraodontidae* family of fish. One member of the family, the Fugu (*Fugu rugripes*) has less than 8% of repetitive DNA, compared to 60% in humans (1, 12). Moreover, non-coding regions such as introns are significantly reduced. However, the gene number is similar to human. These characteristics make the Fugu an ideal model for genomic analysis (http://Fugu.hgmp.mrc.ac.uk). Such an example demonstrates another consequence of the new genomic era, that the barriers between species have been broken. The unifying factor is the DNA itself.

Bioinformatics

Another feature of this new era is the integration of computer science into the field. As the sequencing projects progress, large amounts of data are produced daily. These data need to be analyzed, annotated and stored in databases that facilitate access by the scientific community. Some of the available web sites pertaining to bovine genomics are presented in Table 2.

Functional genomics

The most challenging step in the process of identification of genes underlying economically important traits is the confirmation of the role of a candidate gene. Although the existence of linkage disequilibrium allows proceeding to the application of Marker Assisted Selection (M.A.S) schemes without further knowledge of the

corresponding genes, complete characterization of the genetic basis of the trait requires the identification of the underlying genes. In this situation comparative genomics can offer assistance.

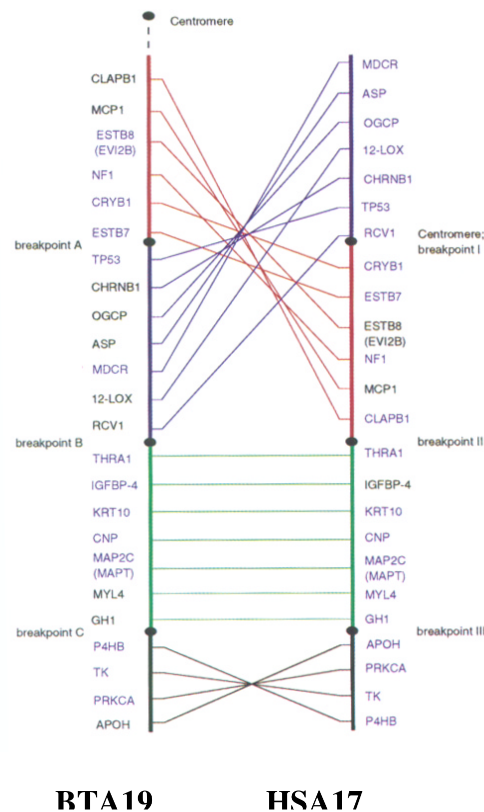

Figure 5: Radiation hybrid maps of bovine chromosome 19 and human chromosome 17, showing chromosomal rearrangements between the two species (36).

Model organisms have not only well-developed genetic maps, but they are also very well-characterized in terms of their phenotype. The role of a gene in metabolism identified in domestic species can be tested first in model organisms to confirm its role as a QTL. Alternatively, once a gene has been cloned and characterized in the model organism, the orthologue gene can be evaluated in cattle (14). New techniques are being developed to establish gene function on a large scale. One of the most promising is the microarray technology (11, 13, 19). This new technology makes it possible to study the expression of thousands of genes at a time. It is expected that entire genomes would be monitored in this way in the future. Microchips containing human and mouse genes are commercially available. Also, research groups are working on custom made chips.

The mouse is undoubtedly the best experimental model to establish the function of genes in mammals. Different experimental strategies have been developed to characterize the function of new genes in the mouse. Systematic mutagenesis in the mouse has been used as a way to characterize gene function (2, 30). Also, a comprehensive database of transgenic and knockout mouse models is available on-line, at the Jackson Laboratory web site (http://tbase.jax.org/).

Several examples demonstrate the power of parallel genetic analysis of traits of economic importance using model organisms. In one study, the mouse model allowed the identification of the loci responsible for resistance to Trypanosomiasis (18). In this experiment, three resistance loci were successfully mapped in an F_2 cross between resistant and susceptible strains.

The myostatin gene is a member of the TGF-ß superfamily. Its function in the mouse was described in a gene knockout experiment (21, 24). Mice lacking functional myostatin genes have a generalized increase in muscle mass, suggesting that the gene could be a negative regulator of muscle growth. The myostatin gene is highly conserved among species. The search for mutations in other species demonstrated that the double-muscled breeds of cattle have mutations on the myostatin coding sequence (25). Furthermore, the bovine

homologue has been mapped to chromosome 2 (31), where it was expected to be located based on comparative maps. The phenotype of myostatin knockout mice is very similar but not identical to the one seen in cattle, implying that even when the same gene is mutated, species-specific genetic factors modulate the influence of the mutation on the phenotype. Therefore, the last step in the process of gene identification will always be the characterization of the gene in the species of interest. However, substantial knowledge about the gene can be gained in the parallel analysis of model organisms.

Table 2: Web sites with information on bovine and animal genomics.

AGIS: Livestock Animal Genome Databases (USDA\ARS)
http://probe.nalusda.gov:8300/animal/index.html
Animal Improvement Programs Laboratory (Belstville, MA)
http://aipl.arsusda.gov/
Animal version of OMIM: disease genes
http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=omia
Bovine Genome Database of the National Animal Genome Research Program(NAGRP)
http://bos.cvm.tamu.edu/bovgbase.html
BovMap (Jouy, France)
http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/bovmap/Bovmap/main.pl
MARC: Meat Animal Research Center (Nebraska)
http://sol.marc.usda.gov/genome/cattle/cattle.html
National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov/
National DHIA (Dairy Herd Improvement Association)
http://www.dhia.org/home.htm
National Association of Animal Breeders
http://www.naab-css.org/index.html
Roslin Institute, Anubis program
http://locus.jouy.inra.fr/cgi-bin/lgbc/anubis27

In the last years, our laboratory has been involved in the characterization of the *high growth* (*hg*) locus, a mutation that increases body size in the mouse (3). In order to understand the genetic regulation of growth in mammals, we mapped recently growth QTL in a cross with an *hg* background (Corva and Medrano, unpublished). A very significant QTL associated with growth rate and lean mass was identified on the distal region of mouse chromosome 2, and the Growth Hormone Releasing Hormone (*Ghrh*) gene was proposed as a candidate for that locus. Interestingly, a QTL associated to carcass composition maps to the syntenic region on bovine chromosome 13 (32). Further analysis and characterization of the mouse QTL could shed light on the identity of the QTL mapped in cattle.

Acknowledgements:
Funding to the laboratory of J.F. Medrano by "Dairy Management Inc." and the "California Dairy Research Foundation", U.S.A. is gratefully acknowledged.

References

1 Angrist, M (1998) Less is more: compact genomes pay dividends. *Genome Res.* 8: 683-685.
2 Battey, J, Jordan E, Cox D and Dove W (1999) An action plan for mouse genomics. *Nat. Genet.* 21: 73-75.
3 Bradford, GE and Famula TR (1984) Evidence for a major gene for rapid postweaning growth in mice. *Genet. Res.* 44: 293-308.
4 Casas, E, Shackelford SD, Keele JW et al. (2000) Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *J. Anim. Sci.* 78: 560-569.
5 Charlier, C, Farnir F, Berzi P et al. (1996) Identity-by-descent mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15. *Genome Res.* 6: 580-589.
6 Collins, FS (1995) Positional cloning moves from perditional to traditional. *Nat. Genet.* 9: 347-350.

7 Coppieters, W, Blott S, Farnir F et al. (1999) From phenotype to genotype: towards positional cloning of quantitative trait loci in livestock? In: *From Jay Lush to Genomics: Visions for animal breeding and genetics*, 16-18 May 1999, Ames, USA. Edited by J. C. M. Dekkers, S. J. Lamont and M. F. Rothschild. AgBiotechNet Proceedings 001, Paper 5. (available at http://www.agbiotechnet.com).

8 Cox, DR, Burmeister M, Price ER et al. (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250: 245-250.

9 Darvasi, A (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet.* 18: 19-24.

10 Darvasi, A, Weinreb A, Minke V et al. (1993) Detecting Marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943-951.

11 Duggan, DJ, Bittner M, Chen Y et al. (1999) Expression profiling using cDNA microarrays. *Nat. Genet.* 21(Suppl. 1): 10-14.

12 Elgar, G, Sandford R, Aparicio S et al. (1996) Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). *Trends Genet.* 12: 145-150.

13 Epstein, CB and Butow RA (2000) Microarray technology - enhanced versatility, persistent challenge. *Curr. Opin. Biotechnol.* 11: 36-41.

14 Georges, M (1997) QTL mapping to QTL cloning: mice to the rescue. *Genome Res.* 7: 663-665.

15 Gilbert, W (1991) Towards a paradigm shift in biology. *Nature* 349: 99.

16 Hayes, H (1995) Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes. *Cytogenet. Cell Genet.* 71: 168-174.

17 Jiang, Z, Priat C and Galibert F (1998) Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mamm. Genome* 9: 577-587.

18 Kemp, SJ, Iraqi F, Darvasi A et al. (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat. Genet.* 16: 194-196.

19 Lander, ES (1999) Array of hope. *Nat. Genet.* 21 (Suppl. 1): 3-4.

20 Lander, ES and Schork NJ (1994) Genetic dissection of complex traits. *Science* 265: 2037-2047.

21 Lee, SJ and McPherron AC (1999) Myostatin and the control of skeletal muscle mass. *Curr. Opin. Genet. Dev.* 9: 604-607.

22 Lyons, LA, Laughlin TF, Copeland NG et al. (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* 15: 47-56.

23 Ma, RZ, van Eijk MJ, Beever JE et al. (1998) Comparative analysis of 82 expressed sequence tags from a cattle ovary cDNA library. *Mamm. Genome* 9: 545-549.

24 McPherron, AC, Lawler AM and Lee SJ (1997) Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* 387: 83-90.

25 McPherron, AC and Lee SJ (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. USA* 94: 12457-12461.

26 Nadeau, JH and Sankoff D (1998) Counting on comparative maps. *Trends Genet.* 14: 495-501.

27 O'Brien, SJ, Menotti-Raymond M, Murphy WJ et al. (1999) The promise of comparative genomics in mammals. *Science* 286: 458-462; 479-481.

28 Rikke, BA and Johnson TE (1998) Towards the cloning of genes underlying murine QTLs. *Mamm. Genome* 9: 963-968.

29 Riquet, J, Coppieters W, Cambisano N et al. (1999) Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proc. Natl. Acad. Sci. USA* 96: 9252-9257.

30 Schimenti, J and Bucan M (1998) Functional genomics in the mouse: phenotype-based mutagenesis screens. *Genome Res.* 8: 698-710.

31 Smith, TP, Lopez-Corrales NL, Kappes SM and Sonstegard TS (1997) Myostatin maps to the interval containing the bovine mh locus. *Mamm. Genome* 8: 742-744.

32 Stone, RT, Keele JW, Shackelford SD et al. (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *J. Anim. Sci.* 77: 1379-1384.

33 Weller, JI, Kashi Y and Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73: 2525-2537.

34 Womack, JE, Johnson JS, Owens EK et al. (1997) A whole-genome radiation hybrid panel for bovine gene mapping. *Mamm. Genome* 8: 854-856.

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

36 Yang, YP and Womack JE (1998) Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison. *Genome Res.* 8: 731-736.