Reprod. Nutr. Dev. 39 (1999) 555–561

Review

Towards marker assisted selection in livestock

Michel Georges

Department of Genetics, Faculty of Veterinary Medicine, University of Liège, B43, 20 Boulevard de Colonster, B4000 Liège, Belgium

(Received 21 June 1999; accepted 12 July 1999)

Abstract — In recent years, genomic tools have become available for most livestock species and are now being used routinely to map quantitative trait loci underlying the genetic variance for numerous economically important traits. Fine-mapping methods are being devised to refine the initially coarse map positions of the quantitative trait loci to the point required for marker assisted selection and, eventually, the positional cloning of the underlying genes. Mapping information on QTL is beginning to be used to increase genetic response by enhancing genetic variance, selection accuracy, selection intensity and by reducing the generation interval. Optimal use of MAS will require the development of more robust methods for the routine genotyping of preimplantation embryos for multiple markers. © Inra/Elsevier, Paris

marker assisted selection / QTL mapping / livestock production

Résumé — Vers une sélection assistée par marqueurs en production animale. Ces dernières années, les cartes génétiques ont été développées chez la plupart des espèces domestiques. Ces outils sont maintenant couramment utilisés pour localiser les loci impliqués dans la variabilité de nombreux caractères quantitatifs (QTL) économiquement intéressants. Des méthodes de cartographie plus précises ont ensuite été développées afin d'affiner les localisations initiales et ainsi de pouvoir, soit utiliser ces informations pour des stratégies de sélection assistée par les marqueurs génétiques (MAS), soit éventuellement de cloner les gènes responsables de ces caractères quantitatifs. Les informations cartographiques des régions renfermant ces QTL sont déjà utilisées pour accroître les potentiels génétiques en augmentant la variabilité génétique, la précision et l'intensité de sélection, et en réduisant les temps de génération. L'utilisation optimale de la sélection assistée par marqueurs, nécessite de développer des méthodes fiables permettant de déterminer en routine le génotype d'embryons pré-implantatoire pour de nombreux marqueurs génétiques. © Inra/Elsevier, Paris

sélection assistée par marqueurs / cartographie de QTL / production animale

E-mail: michel.georges@ulg.ac.be

1. INTRODUCTION

Domestication rhymes with gene manipulation. Indeed, by selecting animals exhibiting the desired properties, early agriculturalists were already sorting alleles, albeit unwittingly. More recently, by modelling an individual's phenotype as the sum of an environmental and genotypic component, quantitative genetic theory has allowed a considerable increase in genetic response. Although it is recognized that for most production traits the genotypic component likely reflects the joint contribution of multiple 'polygenes', their actual identity and precise mode of action remains unknown.

With the advent of recombinant DNA it has become feasible – at least in principle – to dissect this 'black box' into its individual Mendelian components. This prospect has received a lot of attention, not only because an understanding of the molecular architecture of complex heritable traits is of fundamental interest, but also because it might contribute to the improvement of breeding designs.

Today, the preferred strategy for identifying the genes that account for the genetic variation of a trait of interest, as observed either within or between populations, is referred to as 'positional candidate cloning' [5]. In a first stage, this approach requires the chromosomal localisation of the pursued genes using linkage analysis or related strategies, followed by the actual identification of the causal gene and sequence variants amongst the 'candidate' genes known from transcript maps to be located in the corresponding chromosomal area. While comprehensive transcript maps will not become available for domestic species in the near future, the remarkable conservation of synteny observed between mammals will allow animal geneticists to benefit from the human transcript map, which will soon be complete.

The success of the positional candidate cloning approach amongst animal geneticists is not only due to recent methodological breakthroughs – catalysed by the Human Genome Initiative – that have rendered this strategy so effective, but also to the perspective that mapping data alone could lead to more effective marker assisted selection (MAS) in the near future.

2. AVAILABLE MICROSATELLITE MAPS ALLOW FOR EFFECTIVE QTL MAPPING IN LIVESTOCK, BUT CONVENTIONAL APPROACHES LACK BOTH ACCURACY AND PRECISION

In recent years we have seen the development of comprehensive microsatellite maps for the major livestock species. The number of microsatellites is now in excess of 1 000 for cattle, pig, sheep and poultry [8]. This allows for the selection of panels of 200–300 well-distributed informative markers in most populations, which is a suitable density to map QTL using conventional linkage analysis. Using a panel of 300 autosomal microsatellites, we obtained an average information content from excess of 70 % across the bovine map from a previously described Holstein-Friesian grand-daughter design [6].

QTL influencing a broad range of economically important traits have been mapped in livestock using such panels and either experimental crosses (i.e. F2 and BC) [1] or outbred pedigrees [9]. Experimental populations produced from crosses between parental lines diverging for the traits of interest have been used in cattle, sheep, pigs and poultry. In dairy cattle, the preferred design has been the grand-daughter design that takes advantage of i) existing large paternal half-sib pedigrees obtained by A.I., and ii) the progeny-testing scheme reducing environmental noise [21].

A number of robust QTL mapping methods are available for such designs, including multipoint regression, maximum likelihood and rank-based methods. The most com-

monly applied methods are two-generation methods, i.e. one looks for the segregation of putative parental QTL alleles within full- or half-sib pedigrees. In experimental crosses, marker information from grandparents is included to establish marker–marker and marker–QTL linkage phase. More powerful methods, attempting to capture information from more complex pedigree relationships and allowing us to deal with missing genotypes are being developed [2].

Generally speaking, however, the mapping resolution that can be obtained from such approaches remains poor. Support intervals are often in the 20–30 cM range. As an example, we will show the location scores obtained along the chromosome 20 marker map for five milk yield and composition traits in the same Holstein-Friesian grand-daughter design. The same graph also reports the frequency distribution of the most significant map position as obtained by bootstrapping for the trait yielding the highest statistic: protein percentage. Although the evidence for the presence of a QTL on this chromosome is very significant, the support interval essentially encompasses its entire distal half, i.e. as much as 30 million base pairs or of the order of 1 000 genes.

Similar figures would be expected when dealing with experimental crosses. Depending on the type of cross (F2 or BC) and degree of dominance, Darvasi [7] reported confidence intervals ranging from approximately 15–70 cM when mapping a QTL with an additive effect of 0.25 phenotypic standard deviations using a pedigree comprising of the order of 500 offspring.

In addition to their lack of precision, most commonly applied QTL mapping methods potentially suffer from a lack of accuracy. Indeed, applying mapping methods that postulate the presence of a single QTL on a chromosome when in fact it carries two or more, is susceptible to reveal 'ghost' QTL, with the position most likely coinciding with that of either of the actual QTL [13].

Obviously the mapping resolution that will be achieved in most QTL mapping experiments will be insufficient to seriously envisage positional cloning and will often limit the efficiency of MAS. Methods to improve the location of QTL after their initial detection are therefore needed very much.

3. OPTIONS FOR GENETIC FINE-MAPPING OF QTL

Factors that limit the achievable mapping resolution are: i) marker density, ii) cross-over density, and iii) trait complexity.

3.1. Marker density

Obviously, the size of the marker interval to which the QTL can ultimately be assigned depends on the available marker density in the region of interest. With the maps presently available in livestock species, marker densities are of the order of 0.5–3 cM on average [9]. In cattle, efforts are underway to more than double this density across the entire genome (C. Beattie, pers. comm.). Moreover, it is relatively trivial to develop additional markers in a specific region of interest. Marker density, therefore, is unlikely to be a major limiting factor when attempting to fine-map QTL.

3.2. Cross-over density

Dense marker maps, however, are useless without offspring that have inherited recombinant chromosomes in the region of interest. As shown by Boehnke [3], the maximal mapping resolution achievable in a given pedigree corresponds to the interval defined by the nearest flanking cross-overs. One therefore attempts to collect as many recombinants as possible in the region of interest, as only these individuals contribute to refining the map position of the studied

gene. When dealing with experimental crosses [7] this can be achieved by producing more offspring (F2 or BC). Note that between 7 000 and 20 000 such offspring are needed to refine the map position of a QTL with an additive effect of 0.25 standard deviation to 1 cM! Phenotyping costs can be reduced substantially by considering only those offspring that recombine in the interval of interest. All individuals would, however, still have to be produced and marker genotyped. Alternatively, one could use advanced intercross lines (AIL; i.e. F3, F4, ... Fn lines) to increase the crossover density in the analysed generation. The main disadvantage of this approach, however, is the time required to produce such material when dealing with species with long generation intervals. Therefore, methods which consist in the generation of new recombinants by directed breeding will most often be impractical in livestock species.

Rather than produce recombinants de novo, however, one can attempt to exploit 'historical' recombinants, i.e. exploit linkage disequilibrium that might exist around the QTL of interest. Such an approach was recently applied with success to map a QTL influencing a psychological trait to a 0.8 cM interval using a heterogeneous stock (HS) resulting from 58 generations of intercross involving eight inbred mouse strains [20]. To be efficient, the marker density required for such an approach needs to be of the order of the genetic distance over which linkage disequilibrium can be expected in the population of interest. In the human population, for instance, linkage disequilibrium typically extends over subcentimorgan regions, pointing towards the need for fairly dense marker maps. We attempted to quantify the extent of linkage disequilibrium that might exist in the Holstein-Friesian dairy cattle population using genotypes corresponding to approximately 300 microsatellites and a sample of approximately 1 000 phase-known chromosomes which were considered representative of the Holstein-Friesian population. To our surprise we found very strong evidence for long range linkage disequilibrium for all autosomes, linkage disequilibrium extending over regions of 20 cM. Moreover, we found strong evidence as well for disequilibrium or gametic associations between non-synthenic loci (Farnir, in preparation). A number of factors could contribute to this disequilibrium, particularly drift due to the very limited effective population size characterizing the Holstein-Friesian dairy cattle population, selection and maybe admixture. This observation, therefore, suggested that linkage disequilibrium might be readily exploitable to fine-map genes in livestock populations without the need for very high density marker maps. We are in the process of evaluating the mapping power and precision that can be obtained with existing marker maps in livestock when using multipoint association tests or a transmission disequilibrium test (Blott et al., unpublished data).

3.3. Trait complexity

When dealing with complex traits, however, the major factor limiting mapping resolution is the complex relationship between QTL genotype and phenotype. Because the actual QTL genotype of a given individual can never be determined unambiguously from its phenotype, the position of the QTL with respect to chromosomal breakpoints cannot be determined with certainty. Several of the strategies proposed to fine-map QTL consist in methods aiming to clarify the QTL genotype of offspring that have inherited recombinant chromosomes in the critical region. Examples of such approaches applicable when dealing with experimental crosses were reviewed by Darvasi [7] and include recombinant progeny testing, the production of interval-specific congenic strains and the recombinant inbred segregation test. All these methods share the fact that offspring have to be produced from recombinant individuals allowing for the estimation of their actual QTL genotype

either from the phenotype or from a combination of phenotypes and marker genotypes of their descendants. As these methods require breeding of specific individuals de novo, they are again difficult and costly to implement with most livestock species.

As part of our efforts to fine-map QTL influencing milk production in cattle, we therefore developed an alternative approach that combined both linkage disequilibrium and QTL genotype determination by marker assisted segregation analysis in existing pedigrees. The proposed strategy is twofold. The first phase consists in identifying individuals with predictable genotypes for a specific QTL. Specifically in dairy cattle, we would identify sires that would be of the QTL genotype '*Qq*' by performing a marker assisted segregation analysis either on their sons ('grand-daughter design') or daughters ('daughter design'). By doing so, essentially we transform a complex polygenic problem into a series of monogenic entities. The second phase consists in analysing the limited number of selected '*Qq*' individuals with a high density marker map in the region of interest in order to identify shared, identical-by-descent (IBD) chromosome segments flanking a hypothetical QTL allele with large substitution effects. This approach therefore postulates that QTL alleles with large substitution effects will often be 'genetically homogeneous' within dairy cattle breeds in a manner reminiscent of what has been found for recessive diseases such as syndactyly [4] or double-muscling [10].

The proposed strategy was applied to a previously described QTL having a major effect on milk yield and composition and mapping to the centromeric end of bovine chromosome 14 [16]. Seven sires were found to be heterozygous '*Qq*' for this QTL based on the observed effect of the corresponding chromosome region on the sons' breeding values for milk yield and composition. Novel microsatellites and single nucleotide polymorphisms were developed in the corresponding chromosome region and the seven sires genotyped for the corresponding markers. Analysis of the phase-known sire chromosomes indicated that the seven chromosomes increasing fat percentage in the sons shared an identical-by-state haplotype spanning of the order of 5 cM. A multipoint maximum likelihood approach was used to demonstrate that this haplotype was indeed highly significant and could not be attributed to 'background' inbreeding for either the Holstein-Friesian population in general or for the seven selected sires in particular. The same haplotype was shown to be associated with increased fat percentage in the general population as well, providing additional support in favour of the location of the QTL within the corresponding interval. Additional marker development is in progress to more accurately define the boundaries of the shared chromosome segment.

An additional impediment to the accurate fine-mapping of QTL occurs when the identified QTL reflects the joint effect of multiple linked genes. We have preliminary evidence that this situation indeed applies for some of the QTL identified in dairy cattle. In experimental crosses, the application of two-QTL models [11], composite interval mapping (CIM [22]) or multiple QTL models (MQM [12]) may alleviate these problems. The utility of such methods is only just beginning to be explored in outbred populations [19].

4. TOWARDS MARKER ASSISTED SELECTION: NEED FOR SYNERGIES WITH REPRODUCTIVE BIOTECHNOLOGY

It is well known from classical quantitative genetic theory that the achievable genetic response is affected by i) the degree of genetic variation pre-existing within the population of interest, ii) the accuracy of selection or the correlation between the predicted and real breeding value, iii) the selection intensity or the proportion of individu-

als selected as parents for the next generation, and iv) the generation interval. Knowledge about the genes underlying the genetic variation for traits of interest is susceptible to enhance genetic response by affecting each of these four factors.

Several of the ongoing QTL mapping projects aim at identifying the genes explaining the differences observed between strains or breeds. As an example, QTL mapping experiments are conducted to understand the prolificity of Chinese pig breeds, or the trypanotolerance shown by N'Dama cattle. Practical benefits of such experiments may come from the possibility to introgress the identified QTL into commercial populations not showing these traits, using so-called marker assisted introgression (MAI). During MAI, marker information is used not only to monitor the introgression of the chromosomal segments of interest in the donor strain, but to monitor the prompt recovery of the recipient genome as well. Essentially, MAI contributes to improving genetic response by increasing genetic variation.

Adding the information on QTL genotype to the conventional phenotypic information measured on the individual itself and/or its relatives is susceptible to enhance selection accuracy and therefore genetic response [17]. This is particularly relevant for traits for which the accuracy obtained using conventional approaches is limited. This is the reason why low heritability traits such as disease susceptibility/resistance or fertility are often considered to be prime targets for MAS. Note that identifying QTL for low heritability traits is also notoriously difficult. Other prime targets for MAS are traits which are difficult to measure on the live animal. These include sex-limited traits such as milk production or carcass quality traits. While the accuracy of selection for such traits may become very high when implementing progeny testing, the latter breeding designs are costly and time consuming, therefore hampering genetic progress by increasing the generation interval. In dairy cattle, nucleus schemes have been proposed as an alternative to progeny testing: bulls are selected based on the milking performances of their full-sisters rather than their daughters. The loss in selection accuracy is expected to be more than compensated for by the potential shortening of the generation interval. Nucleus schemes, therefore, offer an interesting niche for MAS, allowing for the recovery of at least part of the loss in selection accuracy while maintaining the shortened generation interval [15].

One of the major benefits of MAS, however, may come from its potential to increase selection intensity. All approaches described so far assume that the number of selection is canditates is limited by practical factors. QTL information, however, may allow one to envisage an animal combining favourable alleles at most if not at all identified loci. Such an animal may not be present amongst the available candidates given its a priori rarity. This issue would be particularly relevant if the number of traits targeted by selection were to be increased by the addition of, for instance, disease resistance and fertility. Finding such individuals would require sorting amongst larger numbers of candidates without compromising the achievable progress at unmapped QTL. This could be achieved if MAS were to be combined with reproductive biotechnologies allowing one to produce a large number of gametes from selected matings. The genotype of the gametes could for instance be determined on an embryo biopsy. The 'bottom-up' scheme for the preselection of young dairy sires prior to progeny testing is an effort towards that goal [14, 18]. Routine genotyping of embryos for multiple genetic markers remains, however, a major challenge. Research programmes are underway to increase the amount of genomic DNA available for diagnosis using either in vitro (e.g. PEP or DOP-PCR) or in vivo amplification (e.g. blastomere culture), and will be briefly described.

REFERENCES

[1] Andersson L., Haley C.S., Ellegren H., Knott S.A., Johansson M., Andersson K., Andersson-Eklund L., Edfors-Lilja I., Fredholm M., Hansson I., Håkansson J., Lundström K., Genetic mapping of quantitative trait loci for growth and fatness in pigs, Science 263 (1994) 1771–1774.

[2] Bink M.C.A.M., van Arendonk J.A.M., Detection of quantitative trait loci in outbred populations with incomplete marker data, Genetics 151 (1999) 409–420.

[3] Boehnke M., Limits of resolution of linkage studies: implications for the positional cloning of disease genes, Am. J. Hum. Genet. 55 (1994) 379–390.

[4] Charlier C., Farnir F., Berzi P., Vanmanshoven P., Brouwers B., Georges M., IBD mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15, Genome Res. 6 (1996) 580–589.

[5] Collins F.S., Positional cloning moves from perditional to traditional, Nat. Genet. 9 (1995) 347–350.

[6] Coppieters W., Kvasz A., Arranz J.J., Grisart B., Farnir F., Mackinnon M., Georges M., A rank-based non parametric method to map QTL in outbred half-sib pedigrees: application to milk production in a grand-daughter design, Genetics 149 (1998) 1547–1555.

[7] Darvasi A., Experimental strategies for the genetic dissection of complex traits in animal models, Nat. Genet. 18 (1998) 19–24.

[8] Georges M., Andersson L., Livestock genomics comes of age, Genome Res. 6 (1996) 907–921.

[9] Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E., Hoeschele I., Mapping quantitative trait loci controlling milk production by exploiting progeny testing, Genetics 139 (1995) 907–920.

[10] Grobet L., Poncelet D., Royo Martin L.J., Brouwers B., Pirottin D., Michaux Ch., Menissier F., Zanotti M., Dunner S., Georges M., Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle, Mamm. Genome 9 (1998) 210–213.

[11] Haley C.S., Knott S., A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, Heredity 69 (1992) 315–324.

[12] Jansen R., Controlling the type I and II errors in mapping quantitative trait loci, Genetics 138 (1984) 871–881.

[13] Knott S.A., Haley C.S., Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses, Genet. Res. Camb. 60 (1992) 139–151.

[14] Mackinnon M., Georges M., A bottom-up approach towards marker assisted selection, Livest. Prod. Sci. 54 (1998) 229–250.

[15] Meuwissen T.H.E., van Arendonk J.A.M., Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes, J. Dairy Sci. 75 (1992) 1651–1659.

[16] Riquet J., Coppieters W., Cambisano N., Arranz J.-J., Berzi P., Davis S., Grisart B., Farnir F., Karim L., Mni M., Simon P., Taylor J., Vanmanshoven P., Wagenaar D., Womack J.E., Georges M., Identity-by-descent fine-mapping of QTL in outbred populations: application to milk production in dairy cattle, Proc. Natl Acad. Sci., USA (1999) in press.

[17] Smith C., Simpson S.P., The use of genetic polymorphisms in livestock improvement, J. Anim. Breed. Genet. 103 (1986) 205–217.

[18] Spelman R.J., Garrick D.J., Genetic and economic responses for within-family marker-assisted selection in dairy cattle breeding schemes, J. Dairy Sci. 81 (1998) 2942–2950.

[19] Spelman R.J., Coppieters W., van Arendonk J.A.M., Bovenhuis H., Quantitatve Trait Loci analysis for five milk production traits on chromosome six in the dutch Holstein-Friesian population, Genetics 144 (1996) 1799–1808.

[20] Talbot C.J., Nicod A., Cherny S.S., Fulker D.W., Collins A.C., Flint J., High-resolution mapping of quantitative trait loci in outbred mice, Nat. Genet. 21 (1999) 305–308.

[21] Weller J.I., Kashi Y., Soller M., Power of daughter and granddaughter designs for determining linkage between marker loci and quantitatve trait loci in dairy cattle, J. Dairy Sci. 73 (1990) 2525–2537.

[22] Zeng Z.B., Precision mapping of quantitatve trait loci, Genetics 136 (1994) 1457–1468.