Strategies to optimize QTL detection designs in dairy sheep populations: The example of the Sarda breed

Casu S., Carta A., Elsen J.M.

in

Gabiña D. (ed.), Sanna S. (ed.).
Breeding programmes for improving the quality and safety of products. New traits, tools, rules and organization?

Zaragoza : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 55

2003
pages 19-23

Strategies to optimize QTL detection designs in dairy sheep populations: The example of the Sarda breed

Sara Casu*, A. Carta* and J.M. Elsen**
*Istituto Zootecnico e Caseario per la Sardegna, Str. Sassari-Fertilia km 18.6, 07040 Olmedo, Italy
**Station d'Amélioration Génétique des Animaux, INRA, BP 27, 31326 Castanet Tolosan, France

SUMMARY – This paper faces the problem of reducing costs of genotyping and recording traits not routinely measured in QTL detection programs for dairy sheep outbred populations. With this aim, starting from the actual population structure of the Sarda breed selection nucleus, QTL detection power of daughter design, granddaughter design, and alternative strategies as selective genotyping or selective phenotyping was calculated. Results were discussed focusing on global costs and practical implications of applying QTL detection designs in dairy sheep outbred populations.

Key words: Sarda breed, QTL detection power, selective genotyping, selective phenotyping.

RESUME – *"Stratégies pour optimiser les dispositifs de détection des QTL chez les populations ovines laitières : L'exemple de la race Sarde". Ce papier concerne le problème, dans les programmes de détection des QTL chez les races pures des ovins laitiers, de la réduction des coûts de génotypage et d'enregistrement des performances de caractères qui ne sont pas mesurés en routine. Dans ce but, en partant de la vraie structure de population du noyau de sélection de la race Sarde, la puissance des protocoles fille, petites-filles et de stratégies alternatives telles que le génotypage ou phénotypage sélectif, a été calculée. Les résultats ont été discutés en considérant les coûts globaux et les implications de l'application des protocoles de détection de QTL dans les races pures des ovins laitiers.*

Mots-clés : *Race Sarde, puissance de détection des QTL, génotypage sélectif, phénotypage sélectif.*

Introduction

Size of mapping QTL experiments is often limited by the need of containing costs related to procreation and rearing of animals, performances recording and molecular DNA assays. These constraints are particularly strong for dairy sheep breed populations considering the lower individual income if compared with dairy cattle. Therefore, QTL mapping programs in dairy sheep should explore all possible strategies to limit the number of the on-farm measurements and genotypings. A first step may be the detection of traits and markers of interest by limited mapping QTL designs applied to experimental populations. Crosses between breeds to produce F2 or backcross populations allow to exploit linkage disequilibrium for genes differing between breeds and to detect genome regions controlling the traits of interest. In the light of this, an experimental Sarda × Lacaune backcross resource population of around 1000 ewes was procreated in 1999 by INRA (France) and IZCS (Italy) in order to detect QTL both on milk production traits (milk yield, protein and fat content) and traits related to the reduction of production costs (milkability, functional traits, longevity), health (resistance to mastitis or parasitic diseases), safety and nutritional value of food (milk content in fatty acids related to human health) by means of whole genome intermediate density genetic map (Barillet *et al.*, this volume, p. 13). The recording of most of these traits is time consuming and labour intensive and their study can be currently carried out only in limited experimental populations. Results from this experimental protocol, should be useful to plan QTL detection designs in pure breeds with a limited number of investigated genome regions and traits, and to develop tools for the on-farm recording of traits costly and difficult to measure. Anyway, the specific design to be applied to the pure populations must be optimised either for routinely measured traits or traits difficult to record. Daughter designs are often very expensive in terms of genotyping costs. Several experimental protocols have been proposed for reducing DNA analysis costs without loosing in power. Among those the Granddaughter Design (GDD) allows halving the molecular assays needed to obtain the same power with Daughter Design (DD) in the case of outbred populations, such as dairy cattle, characterized by large use and efficiency of artificial insemination (Weller *et al.*, 1990). The Selective Genotyping (SG) of the high

and low tails of a population completely measured for a given trait is another possible strategy to obtain a considerable reduction of DNA analysis with no proportional power loss (Lander and Botsein, 1989; Darvasi and Soller, 1992). The general principle exploited in SG is that most linkage information can be inferred by individuals showing extreme phenotypic value. Selection induces a positive correlation between residuals effect and marker mean that determines an increasing of the marker contrast and a consequent increase of power (Lander and Botsein, 1989; Bovenhuis and Spelman, 2000). Thus, both SG and GDD can substantially increase power at the expense of growing and phenotyping additional progeny, which is not a constraint for routinely measured traits in outbred populations. In some instances, however, QTL experiment size is limited because direct measurements of the trait of interest can be costly. In these cases, a correlated trait may be available for which the measurement is less time-consuming, labour intensive and more convenient. Selecting individuals to phenotype for the trait of interest from the high and low tails of the correlated trait (SP) can be an useful strategy (Medugorac and Soller, 2001). The principle is that part of the inflation of the marker contrast derived from selecting the tails of the correlated trait also affects the marker contrast for the investigated trait (Bovenhuis and Spelman, 2000). Simulation results obtained by Casu and Carta (2001) showed that the SP strategy applied to a DD actually allows an increase in power that, provided the presence of QTL influencing both traits, strongly depends on the QTL effect on the correlated trait and, to a lesser extent, on the correlation between the two traits.

The purpose of this paper is to compare different strategies to maximise overall power of QTL detection designs starting from the actual population structure of the Sarda breed flock-book.

Materials and methods

Population structure

The flock book of the Sarda breed currently included around 200,000 milk recorded ewes. Although milk yield is the only trait presently included in the selection objective, other traits are routinely recorded on the primiparous ewes, such as udder morphology since 1999, milk fat and protein content and somatic cell count, since 1998. Since early ’90, the Artificial Insemination (AI) has reached a quite large diffusion. Currently, around 20,000 fresh semen doses per year are distributed to progeny test young males and to realize planned matings between elite rams and ewes. The optimal combining between AI and natural controlled mating currently produces a genetic gain of around 3 litres per year-ewe. The widespread use of AI has also increased the size of families and allows to verify the possibility of applying QTL detection designs. The dataset used for the official genetic evaluation of 2001 included the performances of 96,113 ewes with a regular first lactation, born in the last four years from 3632 sires. The number of families to include in the simulation study of different QTL mapping designs was chosen on the basis of a threshold of 100 daughters per sire for the DD and 20 sons and at least 400 granddaughters per grandsire for the GDD (on average 41 sons and 48 daughters per son). This resulted in a subset of 90 half sibs families available for a DD and 20 grandsires families for a GDD.

Starting from this population structure two scenarios were envisaged.

Scenario 1 – QTL detection for routinely measured traits

In this scenario we focused our attention on the detection of QTL affecting only a routinely recorded trait, i.e. all phenotypes are available without additional costs.

Power of QTL detection was calculated for four designs: (1a) a complete DD; (1b) a DD with the selective genotyping of the 50% of the population chosen from the high and low tails of the within family phenotypic distribution [SGDD(50%)]; (1c) as 1b, but with the selective genotyping of the 25% within family extreme tails [SGDD(25%)]; and (1d) a complete GDD.

Scenario 2 – QTL detection for not routinely measured traits

In this scenario we focused our attention on the detection of QTL affecting a trait difficult to measure, but with an available correlated routinely measured trait. Four designs were analysed: (2a) DD with the selective phenotyping and genotyping of the 50% of the population, chosen from the

within family high and low tails of the correlated trait's phenotypic distribution; (2b) as 2a, but with the selective phenotyping and genotyping of the 25% within family extreme tails of the correlated trait; (2c) GDD with the selective phenotyping of the 50% of the population, chosen, within grandsire, among the half sibs families showing extreme phenotypic average for the correlated trait; and (2d) as 2c, but with the selective phenotyping of the 25% of the population.

Power and relative efficiency calculation

Power of QTL detection was calculated by stochastic simulations. A single QTL was supposed to be completely linked to a single marker locus. The percentage of daughters (DD) or sons (GDD) for which it was possible to identify the marker allele received from the sire was fixed to 60% and the rate of marker heterozygous sires or grandsires was set to 70%. QTL was supposed biallelic, with equal allelic frequencies. When two correlated traits were considered, the QTL effects were supposed to be of the same size (0.1, 0.2 and 0.3 within QTL genotype residual standard deviation) and the residual (additive genetic plus environmental) correlation between traits was fixed to 0.70. Polygenic h^2 was assumed to be 0.30 for both traits. For each analysed design 10,000 F tests for the marker contrast were calculated either under the null or the alternative hypothesis. A first type error α of 5 percent was chosen to determine the threshold of rejection of the null hypothesis. Power of QTL detection was estimated as the percentage of replicates in which the F calculated value under the alternative hypothesis exceeded this threshold.

In order to compare different designs, Relative Efficiency (RE) was calculated as the ratio of the absolute power with the number of genotypings, in the first scenario, or the number of the additional trait phenotypings, in the second one.

Results and discussion

Power of the different designs foreseen in the framework of the first scenario (QTL detection for routinely measured traits) is reported in Table 1.

Table 1. Power and relative efficiency of QTL detection designs for traits routinely measured with different allelic substitution effects (scenario 1)

Design	Gene effect (sd)	N grandsires	N sires	N daughters	N genotypings	Power	RE
DD	0.1					11	17
	0.3		90	100	6390	87	138
	0.5					99	157
SGDD(50%)	0.1					10	31
	0.3		90	50	3240	83	256
	0.5					99	306
SGDD(25%)	0.1					9	54
	0.3		90	25	1665	68	408
	0.5					99	594
GDD	0.1					8	135
	0.3	20	41	48	594	51	858
	0.5					99	1667

The complete DD requires a large amount of individual to be genotyped, equal to the Number of Sires (NS = 90) plus the progeny of the marker heterozygous sires (100 × (90 × 0.70) = 6300). The complete DD reaches a sufficient absolute power for a QTL effect of 0.3 within QTL residual Standard Deviation (sd) unit, but at the expense of a large number of genotypings and low RE. The selection of 50% of the population results in quite similar absolute power but with higher RE mainly for the intermediate value of QTL effect (0.3 sd unit). The further reduction of animals to be genotyped up to 25%, although producing an important increase in RE, does not allow obtaining sufficient absolute

power for intermediate QTL effects. For the GDD, the number of individuals to be genotyped equals to the Number of Grandsires (NGS = 20) plus the progeny of the marker locus heterozygous grandsires (41 × (20 × 0.70) = 574). As far as the population structure of the Sarda breed flock book is concerned, the GDD resulted in a low absolute power for intermediate effect QTL although a strong increase in RE. Thus, the GDD seems to be adequate to detect only large effect QTL. If interest is also in intermediate QTL effects a SG-DD with 50% of selected population allows to reach sufficient absolute power at the expense of lower RE and a number of individual to be genotyped more than 5 times higher. This difference becomes more and more costly with the number of investigated genome regions and analysed markers.

Table 2 shows the results of proposed design for QTL detection on traits difficult to measure when a correlated routinely measured trait is available. In a DD, reducing the measurements on the trait of interest and the genotypings to the 50% of daughters that show extreme phenotypic values for a correlated trait [SPDD(50%)], gives very promising results. The absolute power on the trait of interest is sufficient for QTL effects on both traits of 0.3 sd. By contrast SPDD(25%), even if more interesting in terms of RE, does not reach adequate absolute power for intermediate QTL effects. As far as the GDD is concerned, given the large amount of individual performances to record for the trait of interest, the application of SP resulted of null interest because of the lower absolute power, compared to DD, and very low RE. On the whole, results should be interpreted bearing in mind that also genotyping costs have to be included. Thus, the choice of the most convenient strategy has to be made considering also the relative ratio between genotyping and phenotyping costs. As an example, the SPDD(25%) becomes more and more convenient as the ratio between genotyping and phenotyping costs increases. At the extreme, the SP applied to GDD might be convenient when the unit cost of individual phenotypic measurement is much lower than genotyping cost.

Table 2. Power of QTL detection design for traits not routinely measured for different allelic substitution effects (scenario 2)

Design	Gene effects (sd)	N grandsires	N sires	N daughters	N phenotypings†	Power	RE
SPDD(50%)	0.1					10	22
	0.3		90	100	4,500	81	180
	0.5					99	220
SPDD(25%)	0.1					10	44
	0.3		90	50	2,250	65	288
	0.5					99	440
SPGDD(50%)	0.1					7	4
	0.3	20	41	48	19,680	44	22
	0.5					87	44
SPGDD(25%)	0.1					7	7
	0.3	20	41	48	9,840	32	33
	0.5					74	75

†Trait difficult to measure.

Conclusion

It appears that the present population structure of the Sarda breed is suitable for QTL detection programs focusing on those traits already routinely measured and based on a GDD, which, on the basis of our hypothesis, shows the highest relative efficiency and requires the lowest economic effort, with no additional costs other than those due to the genotyping of rams. This becomes particularly true when the interest is in large effect QTL and several markers are involved in the analysis. A DD combined with a SG limiting the number of genotypings to the paternal half-sibs that show extreme phenotypic values, might be combined with GDD in order to increase the power of detection for intermediate effect QTL and, despite the trait specificity of SG, also on traits correlated to that considered for the genotyping selection (Bovenhuis and Spelman, 2000), as it will be the case for different traits of udder morphology, milk yield and composition. As far as the selective phenotyping is

concerned, its application to a DD appears the most adequate design for detecting QTL on traits difficult or expensive to record, such as mastitis and parasite resistance, which have been assuming an increasing interest among sheep breeders, or such as some specific fat or protein components. Direct measurements of these traits (microbiological analysis of milk samples, worms count, HPLC or gas chromatography) are too expensive to be performed on the whole population, but correlated traits easier and more convenient to measure, such as somatic cell count or faecal egg count, fat and protein rate are available. If a genomic region affecting their expression was identified, the selective phenotyping of the 50% of the population showing extreme values of these correlated traits, seems to be a feasible strategy to detect QTL on the trait of interest.

References

Bovenhuis, H. and Spelman, R.J. (2000). Selective Genotyping to detect QTL for multiple traits in outbred populations. *J. Dairy Sci.*, 83: 173-180.

Casu, Sara and Carta, A. (2001). Power of QTL mapping experiments based on selection of individuals on phenotypic values of correlated traits. In: *Proc. ASPA XIV Congress*, Firenze (Italy), 2001, pp. 43-45.

Darvasi, A. and Soller, M. (1992). Selective Genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.*, 85: 353-359.

Lander, E.S. and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.

Medugorac, I. and Soller, M. (2001). Selective Genotyping with a main and a correlated trait. *J. Anim. Breed. Genet.*, 118: 285-295.

Weller, J.I., 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: 2525-2573.