Influence of Relatedness in a Pedigree Design on Estimates of IBD Probabilities

N. Vukasinovic[1] and *G. Freyer*[2]

Introduction

The most popular design for QTL mapping in dairy cattle has been the grand-daughter design (**GDD**) (Weller *et al*. 1990). This method has been criticized for not taking into account all relationships within a pedigree and thus not achieving maximum power for QTL search. The methodology to detect QTL in general pedigrees exploiting marker information was proposed by Fernando and Grossman (1989), based on a model where both the allelic QTL effects and the polygenic components are assumed random. The covariance between individuals for a putative QTL is modeled by the probabilities of sharing alleles identical by descent (**IBD**), based on linked marker genotypes. The major advantage of this approach is the ability to account for relationships among individuals in different families that can be traced back to a remote ancestor, as well as inbreeding. Estimates of IBD probabilities are the key to successful QTL mapping based on (co)variance components method. However, various factors such as complex pedigree structure or missing and uninformative markers can impact IBD probability estimates. Positive impact of involving inbred individuals in a pedigree (i.e. sires of final offspring) of a QTL mapping design was reported by Freyer *et al*. (2009). The goal of this paper is to report how the level of relatedness within a pedigree, including inbreeding, with equally distributed simulation parameters, influences estimates of IBD probabilities with respect to practical QTL mapping.

Material and Methods

A basic pedigree design (FS0) with zero inbreeding, common for pedigrees in commercial dairy herds, was chosen as a starting point for simulations. Pedigrees with 850 individuals and four generations were simulated, originating from two unrelated great-grand sires (GGS1 and GGS2). Different mating designs were applied to achieve a successive increase in inbreeding level of sires (S) of final offspring, starting from mild (FS1), over stronger (FS3, FS3a, FS3b, FS4, FS5, FS88), up to an extremely high inbreeding level (FS99). The sub-pedigrees of GGS1 and GGS2 were intertwined in all family structures except for FS5. Sixty repetitions were obtained for each family structure by varying marker heterozygosity, marker distances and information content (i.e. changes regarding missing values). Combinations of simulation parameters were the same for all family structures. Details on data simulation based on PEDSIM (Schelling *et al*. 1998) were described in Freyer *et al*. (2009). Animals were assumed genotyped for 11 unevenly spaced markers within a 55cM long putative QTL region containing one QTL.

[1] Newsham Choice Genetics, Chesterfield, MO-63005, USA
[2] Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany

The IBD probabilities between each pair of gametes were calculated using a recursive algorithm that follows transmission of marker alleles from parents to offspring (Fernando and Grossman, 1989). A rapid deterministic algorithm by Pong-Wong *et al*. (2001) was adapted to obtain IBD probabilities. The algorithm only uses the closest informative (phase known) marker bracket and omits markers that are uninformative, missing or where the linkage phases could not be determined. To avoid loss of information, a method by Knott and Haley (1998) was used to determine IBD probabilities among sibs' gametes, whereas a method by Vukasinovic and Martinez (2002) was used to determine IBD probabilities among half-sibs. These additional methods were applied to offspring of base individuals where linkage phases could not be determined, and to all full and half sib families for which average number of informative markers (across both maternal and paternal gametes) in the family using the original recursive method was smaller than the average number of informative markers used by the (half) sib method.

Results and Discussion

The overall inbreeding stayed relatively low (max. 1.6%) in all family structures except for FS99, due to limiting inbreeding to individual sires. Inbreeding coefficients F_x of sires of final offspring increased from 0.063 in FS1 (one inbred sire) up to 0.375 and 0.426 in FS88 and FS99, respectively, where all sires were inbred (Table 1). Final offspring (n=524) were generally not inbred, except in FS5, FS88, and FS99.

Means and variances of overall IBD probability at the true QTL position showed a FS-specific pattern (not shown in detail). Except for FS3a, FS3b and FS5, both the average and the variance of IBD probability increased with increasing level of inbreeding from FS0 ($\bar{x}$ = 0.013), over FS1, FS3, FS4 and FS88 to FS99 ($\bar{x}$ = 0.314). FS5 ($\bar{x}$ = 0.019) was an exception from the pattern, as mean IBD probability falls between FS1 ($\bar{x}$ = 0.017) and FS3 ($\bar{x}$ = 0.020). There was a gap in the size of IBD- probabilities between FS99 and all other inbred family structures, fully consistent with F_x. The average IBD probability did not significantly depend on marker distances and number of marker alleles. The variance of IBD probabilities, although homogeneous for varying marker distances, slightly increased with increasing number of marker alleles in all family structures. In relation to the mean, the variance of IBD probabilities decreased with increasing inbreeding level (i.e. 0.8 in FS0 and 0.3 in FS99).

Table 1 Inbreeding coefficients (F_x) in total and F_x of sires of final offspring in family structures FS1, FS3, FS4, FS5, FS88 and FS99

	F_x Total						
		S1	S2	S3	S4 + S5	S6 + S7	S8
FS1	0.0004	0.063	–	–	–	–	–
FS3	0.0005	0.125	0.125*	0.125**	–	–	–
FS4	0.0013	0.125	–	0.375	–	–	–
FS5	0.0164	0.125	0.375	0.375	–	–	–
FS88	0.0101	0.125	0.250	0.375	0.250	0.063	0.313
FS99	0.2784	0.352	0.250	0.375	0.344	0.289	0.426

Comparison of marker heterozygosity (*H*) in final offspring showed a dramatic change in the extremely inbred family structure (FS99). In FS0, as well as in most of the inbred family structures, the marker allele frequency did not deviate significantly from starting values (*H* ~ 0.85). FS99 was the only family structure where some marker alleles fully disappeared and heterozygosity in final offspring visibly diminished (*H* ranging from 0.38 to 0.75), comparable to genetic drift in a small, isolated population. Remarkable differences between the highest and the lowest marker- heterozygosity of remaining family structures were found in FS5 (*H* ranging from 0.79 to 0.83) and in FS88 (*H* ranging from 0.77 to 0.84).

Behavior of IBD probabilities: Changing shapes of profiles of correlations of IBD probabilities with the IBD- probability at the true QTL position along the chromosomal map illustrate deviating behavior of IBD probabilities among different family structures in combination with the simulation parameters (Figure 1). In general, correlations between IBD probabilities at the true QTL position and all other positions on the map were smaller with increasing distance from the true position. The reduction was most pronounced with a larger number of marker alleles (shown in Figure 1). More marker alleles should lead to more precise estimates of the QTL position, because a sharply peaked correlation profile around the true QTL position indicates better mapping precision (Grapes *et al*. 2006, based on a simulation with equally distributed markers). We observed this tendency, although not in all combinations. The reason was likely that our maps were based on unevenly spaced markers. FS88 changed its shape dramatically with varying number of marker alleles. FS99 yielded always the most peaked and the steepest profile. FS0, being the only family structure with zero inbreeding, showed less changes in the shapes, tending to flat profiles in combinations with two marker alleles. FS99 was an exception, as the sires and almost all offspring were inbred. Although some trend could be observed, the correlation profiles for the other inbred family structures did not show a clear trend following inbreeding levels. Most likely, the simulated inbreeding level was still too mild to get a clear distinction between, e.g., FS3 and FS88. FS99 was advantageous in all combinations of simulation parameters. A very high level of inbreeding apparently led to a better exploitation of information from the chromosomal region containing high marker density and increased very expressively with high number of marker alleles.

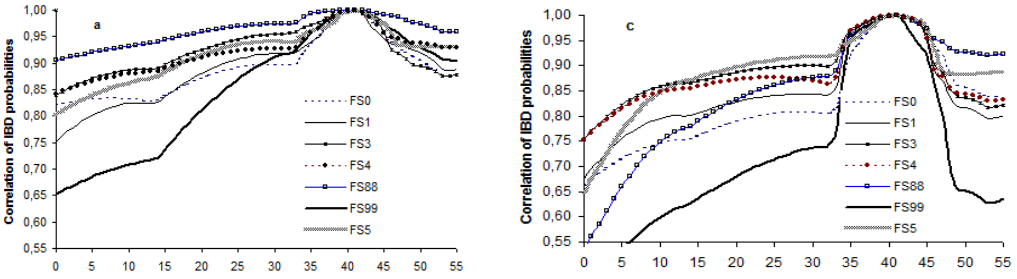

Figure 1 Correlation profiles of IBD at the current map position (*x*- axis) and the correct QTL position at 41.5 cM in combinations with two marker alleles (a) and six marker alleles (c)

Sharing of ancestral IBD in final offspring: Ancestral IBD GGS- allele sharing in final offspring reflects the strength of linkage disequilibrium. The IBD sharing probability of GGS1 and GGS2 increased with increasing number of their offspring. This, however, did not hold for FS5, because the two sub-pedigrees were not intertwined, and GGS1 and GGS2 yielded almost the same IBD sharing probability. The highest IBD sharing probability was in FS99 and GGS1. GGS2 reached its highest IBD-sharing in FS5, where the difference between both great grand sires was the smallest (Table 2).

Table 2 Average of IBD- sharing of great grand sires GGS1 and GGS2 and their final offspring within family structures at true QTL- position averaged over all simulation parameters

	Family structure								
	FS0	FS1	FS3	FS3a	FS3b	FS4	FS5	FS88	FS99
GGS1	0.055	0.112	0.114	0.119	0.112	0.130	0.143	0.149	0.317
GGS2	0.034	0.062	0.038	0.021	0.051	0.071	0.160	0.081	0.135

Conclusions

The reported results suggest that inbred animals and their offspring may be advantageous for estimating IBD probabilities used for QTL mapping in dairy cattle if those relationships are considered in the analysis. This could be shown by several characteristics of IBD, e.g., lower relative IBD variance and higher IBD- sharing in inbred family designs. The main reason for positive effects of involving inbred sires is that it helps to infer the parental phases the parental origin of the offspring haplotypes. Our suggestion would be to include existing inbred animals (sires and dams with their ancestors and offspring) in the QTL mapping panel in dairy cattle. This might reduce the number of individuals needed for genotyping and consequently, lower costs in comparison to approaches that ignore inbreeding or exclude such animals.

References

Fernando, R. and Grossman, M. (1989): *Genet Sel Evol* 21:467-477.

Freyer, G., Vukasinovic, N. and Cassell, B. (2009): *J Dairy Sci* 92:765-772.

Grapes, L., Firat, M., Dekkers, J., Rothschild, M. and Fernando, R. (2006): *Genetics* 172:1955-1965.

Knott, S. and Haley, C. (1998): *Heredity* 81:48-54.

Pong-Wong, R., George, A., Woolliams, J. and Haley, C. (2001): *Genet Sel Evol* 33:453-471.

Schelling, M., Stricker, C., Fernando, R. and Künzi, N. (1998): In *Proc. 6th WCGALP*, volume 27, pages 475-476.

Vukasinovic, N. and Martinez, M. (2002): In *Proc. of the Joint Statistical Meetings, New York, NY, American Statistical Association*, CD-ROM.

Weller, J., Kashi, Y. and Soller M. (1990): *J Dairy Sci* 73:2525-2537.