The Dairy Bull DNA Repository: A Resource for Mapping Quantitative Trait Loci

Y. Da[1], M. Ron[2], A. Yanai[2], M. Band[2], R. E. Everts[1], D. W. Heyen[1]
J. I. Weller[2], G. R. Wiggans[3], and H. A. Lewin[1]

[1]Department of Animal Sciences, University of Illinois, Urbana, IL 61801, USA; [2]Institute of Animal Sciences, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel; [3]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

SUMMARY

The Dairy Bull DNA Repository (DBDR) was established in 1992 as a DNA resource for mapping quantitative trait loci (QTL) associated with milk production, composition and health traits. Currently, the core of the DBDR consists of 16 grandsire families. The 16 core grandsire families have 40 or more sons per grandsire with a total of 1513 sons and at least 527,846 grand-daughters. For a marker with five alleles and equal allele frequency, on average we expect 970 informative sons with 338,407 daughter records for the detection of QTL. Frequency of heterozygous grandsires for 31 markers ranged from 11% to 100%. Predicted frequency of informative sons averaging over heterozygous grandsires was in the range of .5 to .84.

INTRODUCTION

Recent developments in bovine gene mapping (Bishop et al., 1994; Barendse et al., 1994) make possible whole genome scanning for quantitative trait loci (QTL). A logical extension of QTL mapping is the incorporation of marker information into animal breeding (Fernando and Grossman, 1989). This procedure, termed marker-assisted selection (MAS), is expected to increase the accuracy of selection (Smith and Simpson, 1986; Lande and Thompson, 1990), and can be used to select directly for QTL at an early age, thus shortening the selection cycle and reducing the time and costs for genetic evaluations. Use of MAS may also provide a new approach to select traits with low heritability.

Germplasm from many generations of bulls is generally available in the form of frozen semen, providing an abundant DNA resource for gene mapping. The grand-daughter design (Weller et al. 1990) provides an analytical method for mapping QTL in dairy cattle using DNA from the sire-path. A cooperative USDA Regional Research Project was initiated in 1992 for the purpose of mapping QTL in dairy cattle using the grand-daughter design. Within this project, a Dairy Bull DNA Repository (DBDR) was established to ensure that a DNA supply is available for QTL mapping efforts. Nine North American AI organizations have contributed semen to the DBDR. The DBDR currently is housed in the laboratory of H. A. Lewin at the University of Illinois. Legally binding agreements governing participation in the collaboration have been developed and executed. Herein, we report the current status of the DBDR and present an analysis of grandsire pedigrees and marker informativeness in the sons of the DBDR grandsires.

RELATIONSHIPS AMONG GRANDSIRES

Thirty-five US grandsires originally were sought for the DBDR based on the criterion of having at least 50 progeny tested sons. Of the 35 US grandsires, 29 are related by the sire-path, originating from three grandsires (Figure 1). Only two grandsires share the same dam (#25 and #26 are maternal halfsibs). These relationships may require special attention in the QTL analysis.

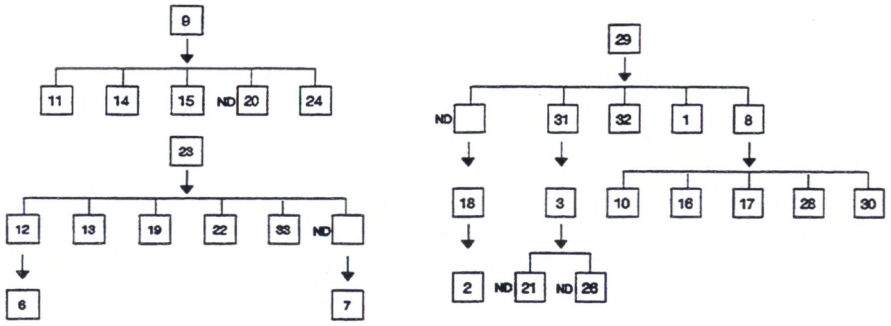

Figure 1. Grandsires related by the sire-path. ND indicates no DNA, and a square without a number indicates that the bull is not one of the 35 grandsires.

INFORMATIVENESS OF SONS OF THE DBDR GRANDSIRES

Of the 35 US grandsires, we have been able to collect semen from 16 with 40 or more sons. These 16 grandsires have a total of 1513 sons (Table 1) and 527,846 grand-daughter records. Most sons (1167) have at least 10 daughter records available for QTL detection. Two additional grandsires from Israel with a total of 51 sons are also available.

To optimize the genotyping scheme, the expected number of informative sons can be predicted for each sire and genetic marker prior to genotyping. Under the grand-daughter design, sires and their sons will be genotyped. The expected frequency of informative sons for a heterozygous sire (A_iA_j) for a locus is $i_1 = 1 - \frac{1}{2}(p_i + p_j)$, where p_i and p_j are allele frequencies of A_i and A_j (Ron et al., 1993; Da and Lewin, 1994), and the expected number of informative sons can be calculated as NIS = $i_1 \times$ (number of sons). Based on the above formula, if the 16 grandsires are all heterozygous, they are expected to have 762 and 1212 informative sons for a marker with two and five alleles respectively (Table 1). For equal frequencies, expected heterozygosity is H = .5 or H = .8 for a marker with two alleles or five alleles. When the expected frequency of heterozygous grandsires is considered, the 16 grandsires are expected to have 8 or 14 heterozygous grandsires, respectively. On average, for a marker with two alleles, 8 heterozygous grandsires are expected to have $.5 \times 762 =$ 381 informative sons and $(381/1513) \times 527{,}846 = 132{,}920$ grand-daughters. For a marker with five alleles, 14 heterozygous grandsires are expected to have $.8 \times 1212 = 970$ informative sons and $(970/1513) \times 527{,}846 = 338{,}407$ grand-daughters for QTL detection. Thus, for moderately heritable traits, the DBDR families have the statistical power of > .8

to detect a QTL effect of .3 SD or greater (Weller et al., 1990). Collection of additional sons of these grandsires will improve the sensitivity and accuracy of QTL detection.

Location of 31 genetic markers to be used for QTL mapping, frequency of heterozygous grandsires per marker, number of sons per heterozygous grandsire are presented for 16 US grandsires and 2 Israeli grandsires (Table 2). The frequency of grandsires heterozygous for genetic markers ranged from 11% to 100%. Mean number of sons per heterozygous grandsire for each marker varied from 60 to 181. Using 26 grandsires, i_1 was assessed for 21 of the 31 genetic markers with known allele frequencies. Allele frequencies in the US and Israeli populations were deduced from genotypes of 7 unrelated US sires and from maternal alleles for genetic markers analyzed on 101 sons of 7 Israeli sires (Ron et al., 1993). Averaging over grandsires, i_1 was in the range of .5 to .84 for the different genetic markers (Table 2). A broader range for i_1 was observed between sires for specific genetic markers. This is due to the occurrence of predominant alleles for several genetic markers, resulting in i_1 fluctuations depending on the alleles of the heterozygous sire. Thus, preselection of families with high i_1 for each genetic marker should be cost effective for the QTL search.

Table 1. Current 16 core grandsire families in the DBDR and expected number of informative sons

Grandsire family #	# of sons	NIS	
		n = 2	n = 5
1	181	91	145
2	169	85	135
3	160	80	128
4	132	66	106
5	137	69	110
6	101	51	81
7	85	43	68
8	81	41	65
9	80	40	64
10	73	37	58
11	71	36	57
12	66	33	53
13	49	25	39
14	46	23	37
15	40	20	32
16	42	21	34
Total	1513	762	1212

NIS = expected number of informative sons.
n = number of alleles at each locus, assuming equal allele frequencies.

Most genetic markers on the bovine map (Barendse et al., 1994) are microsatellites with (TG)n tandem repeats. In this study we used two microsatellites (Band and Ron, 1994) containing (AGC)n tandem repeats (ARO22 and ARO25). The heterozygosity values of these markers were comparable to other TG markers. These markers are characterized by high resolution of alleles and accurate genotyping. Selection of additional markers to be typed on the DBDR families is being coordinated by the collaborators to avoid extensive overlap.

THE DBDR AS REFERENCE FAMILIES FOR GENETIC MAPPING

Based on the linkage information content for two loci (Da and Lewin, 1994), the 16 grandsire families could detect a recombination frequency of 35% with a minimum LOD score of 3.0 and a power of .95. With this kind of power, an accurate male-specific genetic map will arise from the DBDR collaboration. About 1200 sons of the 16 grandsires have dams that are also daughters of the 35 US grandsires. This structure provides the possibility

that a dam's genotype can be determined through the genotypes of the son and the sire. When the genotype of the dam is known, the expected frequency of informative sons is increased to $i_2 = 1 - p_i p_j$ from $i_1 = 1 - \frac{1}{2}(p_i + p_j)$ (Da and Lewin, 1994), thus increasing the power for QTL detection. The large halfsib families may also provide an opportunity to compare recombination frequencies between bulls.

ACKNOWLEDGMENT

This study was supported in part by grants from the United States Department of Agriculture, grant No. IS-1939-91R from the Binational Agricultural Research and Development Fund, and by Project No. 35-308 (NC-209) of the Agricultural Experiment Station, College of Agriculture, University of Illinois at Urbana-Champaign.

REFERENCES

Band, M. and Ron, M. (1994) Anim. Genet. (submitted).

Barendse et al. (1994) Nature Genetics 6:227.

Bishop et al. (1994) Genetics. 136:619.

Da, Y. and Lewin, H. A. (1994) Theor. Appl. Genet. (submitted).

Fernando and Grossman. (1989) Genet. Sel. Evol. 21:467.

Lande, R. and Thompson, R. (1990) Genetics. 124:743.

Ron, M., Band, M., Wyler, A., and Weller J. I. (1993) Anim. Genet. 24:171.

Smith, C. and Simpson, S. P. (1986) J. Anim. Breed. Genet. 103:205.

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

Table 2. Marker polymorphism and informativeness: markers (Locus), chromosomal location of genetic markers (Chr), frequency of heterozygous grandsires per marker (FHG), average number of sons per heterozygous grandsire (Son), and predicted informativeness (i_1) of sons

Locus	Chr	FHG	Son	i_1 (%)	
				mean	range
CSSM19	1	8[a]/15[b]	60	79	72-93
TGLA57	1	3/16	138	ne[c]	ne
ETH121	2	13/15	80	83	75-90
ARO28	2	13/15	79	63	54-86
UWCA7	3	9/15	88	61	57-79
MGTG4B	4	13/15	71	84	79-86
AGLA293	5	5/16	111	ne	ne
UWCA20	7	12/15	73	64	50-95
TGLA48	7	14/16	86	ne	ne
CSSM47	8	15/15	81	50	50
UWCA9	9	13/15	85	79	75-90
ETH225	9	15/15	81	78	71-86
TGLA73	9	1/9	181	ne	ne
BRN	10	10/15	70	71	63-90
TGLA131	13	5/15	70	50	50
CSSM66	14	10/15	81	78	71-90
HBB	15	10/15	80	70	63-90
GBFSH	15	5/7	102	ne	ne
TGLA53	16	4/16	75	ne	ne
ETH185	17	12/15	69	80	72-93
ILSTS002	18	11/15	80	69	64-86
TGLA227	18	9/16	81	ne	ne
TGLA126	20	6/10	86	ne	ne
ETH131	21	13/15	84	84	72-90
TGLA122	21	10/16	97	ne	ne
CYP21	23	11/15	83	72	60-93
MGTG7	23	9/16	81	ne	ne
CSSM23	24	10/15	72	72	65-93
ARO25	26	8/15	66	56	50-81
ARO26	?	9/15	90	74	68-90
ARO22	?	9/15	79	50	50

[a] number of heterozygous grandsires.
[b] number of grandsires tested.
[c] ne - not estimated.