J. Dairy Sci. 86:4137–4147

Genetic Diversity and Background Linkage Disequilibrium in the North American Holstein Cattle Population

R. L. Vallejo,* Y. L. Li,* G. W. Rogers,† and M. S. Ashwell‡
*Department of Dairy and Animal Science,
The Pennsylvania State University, University Park 16802
†Department of Animal Science,
University of Tennessee, Knoxville 37996
‡Bovine Functional Genomics Laboratory, Agricultural Research Service,
US Department of Agriculture, Beltsville, MD 20705

ABSTRACT

The objectives of this study were to 1) identify highly heterozygous Holstein bulls that are as unrelated as possible and widely used in the US dairy industry; 2) quantify the level of genetic diversity in US Holsteins; and 3) determine the extent of background linkage disequilibrium (BLD) and disease trait associated linkage disequilibrium (DLD) in the US Holstein population. Twenty-three Holstein bulls that are not closely related but were widely used in the US dairy industry were genotyped for 54 microsatellite loci. The genotyping was performed on automated DNA sequencers (PE Applied Biosystems, CA), following polymerase chain reaction amplification with fluorescent dye-labeled primers. The heterozygosity for the sampled population ranged from 0.43 to 0.80. This wide range of heterozygosity allows selection of the most heterozygous bulls to develop informative families for gene mapping studies. The degree of genetic diversity in this population is significant and allows selection for traits of economic importance. As expected, there is extensive linkage disequilibrium (LD) in the US Holstein population. About half of the syntenic marker pairs presented a typical pattern of LD produced by DLD. Most of the nonsyntenic marker pairs had a typical pattern of LD arising from BLD. These results suggest that the observed LD is not purely due to genetic drift and migration and that a portion might be due to DLD. This raises our hopes of successful fine-localization of genes for complex traits using LD mapping.
(**Key words:** background linkage disequilibrium, genetic diversity, Holstein, linkage disequilibrium)

Abbreviation key: BLD = background linkage disequilibrium, **DLD** = disease trait associated linkage disequilibrium, **D′** = Lewontin's normalized pair-wise disequilibria, **GE** = genetic equilibrium, **GHR** = growth hormone receptor, **HFD** = haplotype frequency distribution, **HWE** = Hardy-Weinberg equilibrium, **LD** = linkage disequilibrium.

INTRODUCTION

Artificial breeding schemes used in dairy cattle result in reduced allele numbers and loci heterozygosity. The average marker heterozygosity in bovine genome scans is approximately 0.56 (Georges et al., 1995). When using half-sib family-based designs (daughter and granddaughter designs) to identify markers linked to QTL, marker loci that are heterozygous for the parental sires will be the only loci providing information about linkage. Ideally, a panel of marker loci for which each locus is heterozygous for all (or most) of the parental sires would be used. The information extracted from pedigrees used for linkage analysis can be maximized by first selecting potential parental sires with high levels of average loci heterozygosity (sire selection) and then defining a panel of marker loci with the highest heterozygosity possible for the selected heterozygous parental sires (marker selection).

During the past two decades, linkage analysis has been successful in localizing genes for Mendelian diseases and traits in human and livestock populations. Linkage disequilibrium (**LD**) analysis has often complemented the final phases of gene localization. These successes have fueled hopes that similar approaches will be effective in mapping genes for complex traits. Encouraged by the success of LD mapping of Mendelian disorders in isolated populations (de la Chapelle and Wright, 1998), many investigators are currently using these genetic isolates in the search for loci underlying complex diseases (Sheffield et al., 1998; Wright et al., 1999; Peltonen, 2000). Similarly, in dairy cattle, it is encouraging to see successful efforts towards the positional cloning of QTL affecting milk yield and composition using linkage and LD approaches (Grisart et al., 2002; Blott et al., 2003).

Received December 19, 2002.
Accepted May 22, 2003.
Corresponding author: R. L. Vallejo; e-mail: rvallejo@psu.edu.

The identification of a large number of densely spaced microsatellite markers has led to empirical investigations into the distribution of LD in the human (Laan and Paabo, 1997; Service et al., 2001; Devlin et al., 2001) and bovine genomes (Farnir et al., 2000). Quantifying the degree of such "background" LD (**BLD**; i.e., marker-marker loci LD) is a crucial undertaking in paving the way for whole genome association studies. To demonstrate that LD between a disease trait and marker loci is meaningful, the likelihood of simply detecting BLD should be evaluated (Freimer et al., 1997).

For genome-wide association screens to be successful, the LD signal due to the association with a shared disease allele must stand out from the BLD signal. Theoretical studies have suggested that such BLD is highly dependent on the history of a population (Slatkin, 1994), with rapidly growing populations showing less BLD than populations of constant size. In samples of affected individuals sharing a phenotype, and possibly sharing a susceptibility allele at the same disease locus, the amount of LD around the shared disease locus should be greater in a younger population. It has been suggested that if such a young population has also undergone rapid growth, it would be ideal for LD mapping of disease loci (Freimer et al., 1997). In this latter case, BLD should be less extensive in the population.

The North American Holstein population can be considered a relatively young population of constant effective size, which should be suitable for mapping chromosomal regions that underlie complex diseases and traits (i.e., QTL mapping). It was reported that LD extends over 10-cM of genetic map distance in the Dutch Friesian Holstein cattle and that most of this LD is due to random genetic drift (Farnir et al., 2000). This suggests that high levels of BLD must be the rule in dairy cattle populations and that the fine-localization of genetic factors for complex diseases and traits will not be trivial. Given this likely scenario, it is important to determine whether any proportion of observed LD resembles a pattern of haplotype frequency distribution (**HFD**) likely produced by disease trait associated LD (**DLD**; i.e., marker-disease loci LD).

To date, no study on the genetic diversity and BLD distribution in contemporary North American Holstein cattle using a population-based sample has been reported. Genetic diversity in French (Maudet et al., 2002) and North-East Asian (Kim et al., 2002) cattle breeds has been reported using a limited number of microsatellite markers. The report on genome LD by Farnir et al. (2000) may not reflect true population estimates because they used pedigree-based samples (i.e., granddaughter designs) of Dutch Holstein Friesian cattle. Furthermore, although the Dutch Holstein population has some North American Holstein influence, it does not represent the North American Holstein population. Migration generates BLD so the US and the Dutch Holstein populations may have different levels of BLD because of the rapid use and migration of the US Holstein families into the Dutch population in the 1980s and 1990s.

The specific objectives of this study were to 1) identify highly heterozygous elite Holstein bulls that are as unrelated as possible to use as parental sires to develop informative families for mapping genes for complex diseases; 2) quantify the level of genetic diversity in the US Holstein cattle; and 3) determine the extent of BLD and DLD in the contemporary North American Holstein cattle population.

MATERIALS AND METHODS

Bull Sampling and DNA Extraction

Twenty-three elite Holstein bulls that are not closely related to one another and that have large numbers of daughters in the US dairy industry were identified (Table 1) using records from the Animal Improvement Programs Laboratory, USDA (P. VanRaden and D. Norman, personal communication) and dairy breeding companies. Semen samples were acquired from the Cooperative Dairy DNA Repository (CDDR; Ashwell and Van Tassell, 1999). Genomic DNA was extracted from semen samples using a phenol/chloroform protocol previously described (Ashwell et al., 1996).

Microsatellite Genotyping

A sample of 54 microsatellite loci (DNA markers) spanning most of the bovine autosomal chromosomes (Table 2) were selected from the genetic maps available in the Meat Animal Research Center (MARC) database (Kappes et al., 1997). The DNA samples from the selected 23 Holstein bulls, three quality control samples, and one blank control sample (for a total of 27 samples) were genotyped for 54 microsatellite loci. To assess the quality of marker genotype data, three duplicated-blind DNA samples (i.e., quality control set) were genotyped for each marker locus. These genotypes were used to estimate the "genotyping error rate" for each marker locus. Genotyping of microsatellite loci was performed on ABI PRISM 377 and ABI PRISM 3700 DNA sequencers (PE Applied Biosystems, CA), following PCR amplification with fluorescent dye-labeled primers. Automated marker genotype scoring and data analysis were performed using Genescan and Genotyper software (PE Applied Biosystems).

Table 1. Comparison[1] of observed heterozygosity in a sample of elite North American Holstein bulls using microsatellite markers.

Bull[2]	Sire[3]	Dam[3]	Bull services (no.)	Bull avg. covariance	Loci typed (no.)	Heteroz. loci (no.)	Homoz. loci (no.)	Observed heteroz.
11HO3243	1879149	12052401	11,930	0.092	54	43	11	0.80
1HO4316	1890669	12843578	7850	0.104	54	39	15	0.72
1HO4480	1986164	13463545	28,234	0.123	54	38	16	0.70
7HO5255	1912270	14770455	8122	0.118	53	36	17	0.68
1HO3365	2035598	13520885	10,943	0.102	54	36	18	0.67
7HO3707	1879085	11525254	19,543	0.115	49	33	16	0.67
1HO4438	2012343	13372706	12,830	0.109	54	35	19	0.65
7HO3257	1773417	10935707	12,863	0.119	47	30	17	0.64
29HO8375	2071864	14334791	5773	0.069	53	32	21	0.60
7HO4637	383622	13586096	17,307	0.119	53	32	21	0.60
9HO1833	352790	11316154	11,546	0.096	53	32	21	0.60
1HO4333	1983348	12798676	7491	0.135	54	32	22	0.59
1HO5021	2070579	14199846	6959	0.121	54	32	22	0.59
29HO9500	392457	12876116	7856	0.096	54	32	22	0.59
7HO4638	1841366	14096052	12,275	0.139	53	31	22	0.59
9HO1817	1957210	13138862	5080	0.110	54	32	22	0.59
7HO3948	1929410	12600262	32,866	0.148	54	31	23	0.57
7HO4985	2027062	14493791	11,046	0.114	53	30	23	0.57
7HO5112	2055652	13759225	8117	0.064	54	31	23	0.57
1HO3390	2078290	13930245	11,366	0.136	54	30	24	0.56
29HO7673	2020049	14115091	5512	0.147	54	29	25	0.54
23HO453	1858047	12593031	3745	0.107	54	28	26	0.52
11HO4025	1883228	14361567	4969	0.115	53	23	30	0.43

[1]Comparison of observed heterozygosity was performed using the comparison of m proportions (Fleiss, 1981); $\chi^2 = 27.982$, 22 df, $P > \chi^2 = 0.176$.

[2]National Association of Animal Breeders (NAAB) bull number.

[3]Holstein Association USA registration number.

Measuring Genetic Diversity and Testing Genetic Equilibrium

The numbers of homozygous and heterozygous loci was determined for each bull and used to estimate the observed heterozygosity for each elite Holstein bull. The proportions of observed heterozygosity per bull were compared using the comparison of m proportions (Fleiss, 1981). The numbers of alleles, allele frequencies, observed heterozygosity, and expected heterozygosity under Hardy-Weinberg equilibrium (**HWE**) for each microsatellite locus were determined using the computer program Mega2 version 2.2 (Mukhopadhyay, 1999). Fisher's exact tests for deviations from HWE were performed using the computer program Mendel version 4.0 (Lange et al., 1988); exact P-values were estimated using 10,000 permutations. Genetic equilibrium (**GE**) tests permit a combined testing for HWE and linkage equilibrium. Fisher's exact tests for GE between adjacent marker pairs were performed using the computer program Mendel version 4.0 (Lange et al., 1988).

Measuring Linkage Disequilibrium

The extent of LD between syntenic marker pairs and gametic phase disequilibrium between nonsyntenic marker pairs were determined using the computer program Arlequin version 2.0 (Schneider et al., 2000). Exact LD P-values for the observed allelic association under the null hypothesis of random allelic assortment were estimated by Monte Carlo approximation (10,000 simulations) using the computer program Arlequin version 2.0 (Schneider et al., 2000).

Background Linkage Disequilibrium and Disease Trait-Associated LD

The factors that affect LD (e.g., genetic drift, mutation, linkage) lead to different expectations of haplotype distributions; thus, BLD and DLD are expected to produce different patterns of HFD (Freimer et al., 1997). A pattern typical of BLD is presented in Figure 1a. In this example, the marker UWCA20 has four alleles, and the marker HUJII77 has six alleles. The two haplotypes deviating most from linkage equilibrium involve different alleles, namely haplotypes 3-3 and 2-1. In contrast, Figure 1b represents a pattern typical of DLD. In this example, the marker INRA048 has six alleles, and the marker BM719 has four alleles. In this case, the common ancestral haplotype (or founder chromosome) was 4-3 because of an over representation of haplotypes with allele 4 (frequency 0.591) from INRA048 or allele 3

Table 2. Fisher's exact test for Hardy-Weinberg equilibrium[1] for microsatellite loci typed in elite North American Holstein bulls.[2]

Locus	BTA	Posit. cM	Heteroz. MARC[3]	Alleles MARC (no.)	Alleles observ. (no.)	Heteroz. observ. (no.)	Heteroz. expect. (no.)	*P*-value[4]	SE
BM8139	1	8.2	0.59	7	3	0.38	0.50	0.096	0.006
BMS2519	2	101.5	0.62	12	5	0.78	0.70	0.694	0.009
HUJII77	3	87.4	0.66	11	6	0.70	0.68	0.748	0.009
BR4502	3	103.2	0.38	11	5	0.77	0.57	0.621	0.010
BMC4214	3	123.0	0.86	13	6	0.78	0.80	0.049	0.004
BL1024	4	3.9	0.79	7	4	0.70	0.64	0.376	0.010
TCRB	4	97.2	0.21	2	2	0.00	0.15	0.002	0.001
BMS1095	5	0.0	0.76	13	7	0.74	0.81	0.219	0.008
CSSM22	5	71.1	0.66	9	6	0.59	0.72	0.067	0.005
ILSTS093	6	0.0	0.83	17	7	0.70	0.77	0.272	0.009
INRA133	6	8.2	0.52	6	4	0.48	0.39	1.000	0.000
BMS360	6	66.5	0.59	11	5	0.78	0.75	0.883	0.006
ILSTS035	6	81.0	0.79	18	5	0.41	0.48	0.434	0.010
BM4311	6	89.7	0.61	7	3	0.57	0.60	0.421	0.010
BM7160	7	0.0	0.66	8	5	0.57	0.63	0.150	0.007
BM6105	7	35.7	0.69	10	6	0.22	0.77	0.000	0.000
UWCA20	7	59.9	0.62	9	4	0.52	0.56	0.546	0.010
BM711	8	83.6	0.72	9	4	0.70	0.60	0.623	0.010
BMS2377	9	63.8	0.41	6	3	0.85	0.54	0.006	0.002
BM875	10	46.5	0.62	7	3	0.05	0.13	0.025	0.003
CSSM46	10	92.9	0.86	13	6	0.95	0.81	0.749	0.009
BM827	11	0.0	0.17	4	2	0.30	0.23	1.000	0.000
ILSTS049	11	55.0	0.28	6	2	0.14	0.11	1.000	0.000
RM150	11	65.5	0.59	14	7	0.96	0.78	0.900	0.006
BL1103	11	90.9	0.59	12	5	0.74	0.74	0.184	0.008
BMS585	12	79.7	0.52	4	3	0.65	0.55	0.168	0.007
BMS1316	12	98.7	0.55	14	7	0.65	0.81	0.034	0.004
BMS2319	13	85.3	0.69	7	3	0.68	0.60	1.000	0.000
BMS947	14	52.8	0.59	12	5	0.57	0.51	0.743	0.009
BL1036	14	78.7	0.62	10	7	0.83	0.81	0.325	0.009
BMS2055	14	84.1	0.62	10	6	0.74	0.69	0.556	0.010
BR3510	15	1.0	0.72	10	6	0.78	0.77	0.790	0.008
BMS2533	15	5.2	0.86	16	8	0.74	0.82	0.241	0.009
BMS1004	15	7.2	0.55	11	4	0.41	0.33	1.000	0.000
BM848	15	81.6	0.48	5	4	0.35	0.68	0.000	0.000
BM121	16	24.4	0.76	19	7	0.68	0.70	0.272	0.009
TGLA53	16	40.6	0.75	13	9	0.39	0.83	0.000	0.000
INRA048	16	73.0	0.69	15	5	0.57	0.54	0.135	0.007
BM719	16	78.0	0.68	9	6	0.64	0.60	1.000	0.000
BM1862	17	86.3	0.76	16	6	1.00	0.80	0.538	0.010
BM8151	18	42.3	0	0	4	0.57	0.47	0.800	0.008
HEL10	19	15.9	0.55	7	5	0.48	0.55	0.047	0.004
BMS2142	19	44.7	0.69	13	5	0.87	0.75	0.404	0.010
BP20	19	46.5	0.71	8	4	0.70	0.71	0.796	0.008
BM846	21	65.9	0.66	8	6	0.27	0.71	0.000	0.000
RM185	23	45.1	0.82	8	6	0.70	0.72	0.572	0.010
BMS2269	23	60.8	0.62	10	6	0.65	0.67	0.161	0.007
BM226	24	6.0	0.52	10	4	0.57	0.54	0.391	0.010
BMS1926	24	57.4	0.69	9	3	0.65	0.46	0.140	0.007
BM737	25	27.8	0.66	11	6	0.74	0.73	0.314	0.009
BMS1353	25	45.3	0.76	15	8	0.74	0.76	0.306	0.009
CSSM43	27	34.1	0.69	9	11	0.48	0.84	0.000	0.000
BMC6020	28	2.5	0.9	18	6	0.70	0.73	0.660	0.009
BM4602	29	0.0	0.79	12	8	0.83	0.78	0.776	0.008
Average			0.63	10.2	5.2	0.61	0.63		

[1]The computer program Mendel version 4.0 (Lange et al., 1988) was used to estimate the Fisher's exact test for Hardy-Weinberg equilibrium.

[2]Twenty-three elite North American Holstein bulls were typed for 54 microsatellite loci.

[3]USDA Meat Animal Research Center (MARC) reference families.

[4]Exact *P*-value estimated using 10,000 permutations.

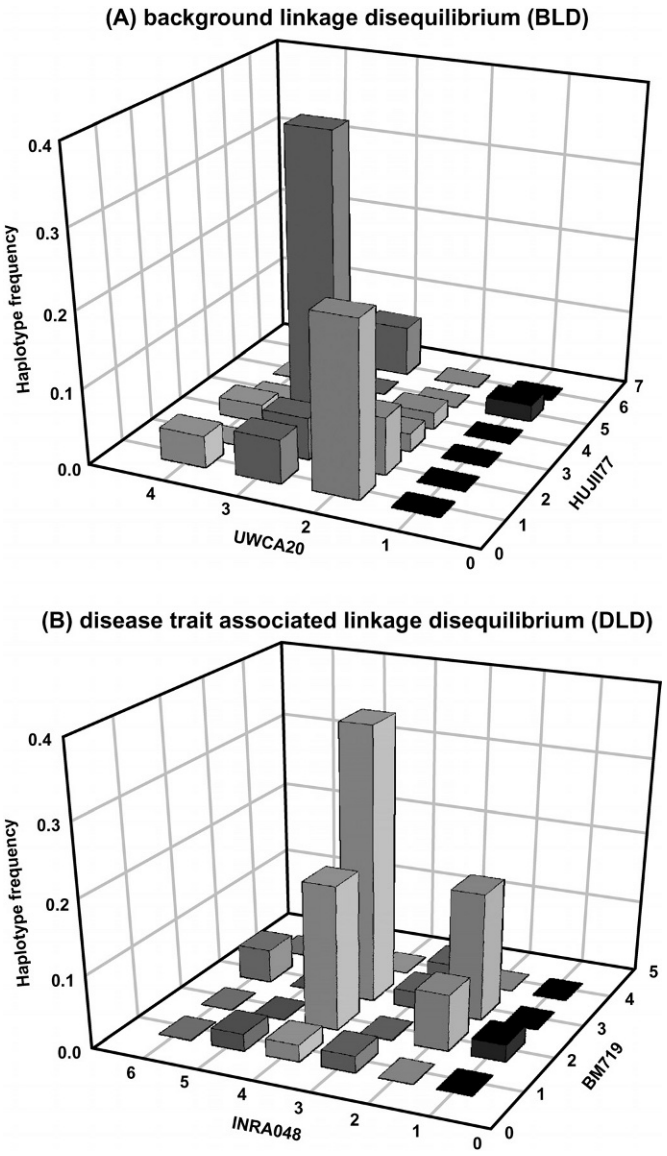

Figure 1. Expected haplotype frequency distribution under (a) background linkage disequilibrium (BLD), and (b) disease trait associated linkage disequilibrium (DLD).

(frequency 0.614) from BM719. The exceptionally high frequency of haplotypes (and alleles) in family- and population-based samples allows identification of a hypothetical common ancestral haplotype (Bitti et al., 2001; Gaspar et al., 2001; Shinar et al., 2002).

In this study, seven syntenic marker pairs that had LD *P*-value < 0.05 and a sample of 20 nonsyntenic marker pairs that had the lowest LD *P*-value (from 132 nonsyntenic marker pairs that had LD *P*-value < 0.05), for a total of 27 marker pairs, were evaluated for HFD that may resemble the expected pattern of LD produced by either BLD or DLD. Briefly, normalized pair-wise disequilibria (**D′**; Lewontin, 1964) and HFD for each of the 27 marker pairs were estimated using the computer program 3Locus version 5.0 (Long, 1999). The HFD for each marker pair was evaluated to determine whether the observed LD resembled a typical pattern of HFD produced by either BLD or DLD. The evaluation criteria were as follows. First, if the two haplotypes deviating most from linkage equilibrium involved completely different alleles (i.e., overrepresentation of any haplotype), then the HFD resembles a pattern typical of BLD (Figure 1a). Second, if the two (or more) haplotypes deviating most from linkage equilibrium involved haplotypes that derived from a common ancestral haplotype (i.e., which implies an overrepresentation of haplotypes involving either allele of this common ancestral haplotype), then the HFD resembles a pattern typical of DLD (Figure 1b). In this example, the hypothetical common ancestral haplotype was 4-3 because in descendant chromosomes (i.e., the sample of bulls tested in this study) there was an overrepresentation of haplotypes that included either allele 4 at marker INRA048 or allele 3 at marker BM719. Finally, if the pattern of HFD did not clearly resemble either BLD or DLD, then the HFD was labeled as a nondetermined (**ND**) pattern.

RESULTS AND DISCUSSION

Genetic Diversity and Heterozygosity

The identification of half-sib families that are highly informative for the mapping of genetic factors underlying complex diseases using linkage and LD methods is important in the US Holstein cattle population. Twenty-three elite Holstein bulls that are least related and widely used in North America were selected to screen for high levels of heterozygosity (Table 1). These selected bulls were genotyped for a sample of 54 microsatellite loci scattered across the bovine genome (Table 2). The observed heterozygosity in these elite Holstein bulls ranged from 0.43 to 0.80 (Table 1). Although the heterozygosity estimates were not statistically different, the range of heterozygosity among elite bulls was wide and allows selection of the most heterozygous bulls to develop informative families for gene mapping studies.

For the microsatellite loci typed in this study, the average observed heterozygosity in the elite Holstein sample was 0.61, which was close to the average heterozygosity reported for the MARC reference families (Table 2). For the evaluated microsatellite loci, there was a positive relationship (r = 0.43) between levels of heterozygosity reported in the MARC reference families and the heterozygosity observed in the sample of elite Holstein bulls. The average heterozygosity observed in the elite Holstein sample was also close to the average het-

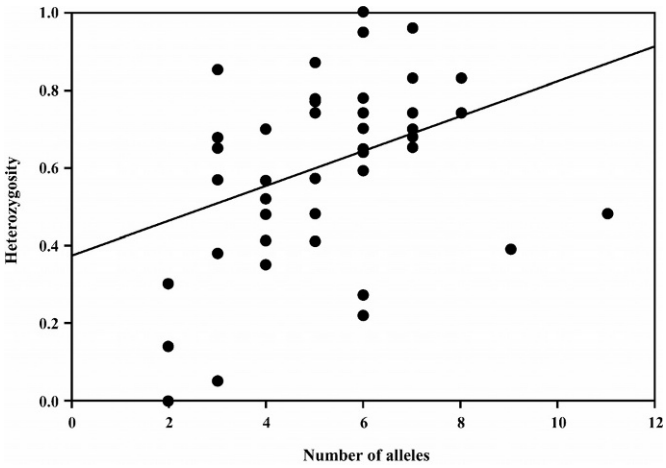

Figure 2. Relationship between heterozygosity and allele number for microsatellite loci in the North American Holstein cattle. Regression of heterozygosity on number of alleles ($b_{y.x}$) was 0.05 and correlation between heterozygosity and number of alleles (r_{xy}) was 0.38.

erozygosity of 0.56 reported for microsatellite loci typed in bovine genome-wide scans (Georges et al., 1995). The average number of alleles observed in the MARC reference families was larger than those observed in the sample of elite Holstein bulls (Table 2). The smaller number of alleles per microsatellite locus observed in the elite Holstein sample might be due to the small sample size used in this study and also because the MARC reference families were developed using a four-way cross of different cattle breeds (Bishop et al., 1994). In the sample of Holstein bulls, in agreement with population genetic theory, a positive relationship between degree of heterozygosity and number of alleles per microsatellite locus was observed (Figure 2). The heterozygosity increased 0.05 units for each unit increase in number of alleles per microsatellite locus ($b_{y.x} = 0.05$).

The degree of genetic variation observed in this purebred Holstein population is still high, which allows selection for traits of economic importance. The relatively high degree of heterozygosity observed for microsatellite loci indicates that half-sib or full-sib families of Holstein cattle that are informative for linkage and LD mapping of genetic factors underlying complex traits can be identified.

Hardy-Weinberg and Genetic Equilibrium

The notions of HWE and linkage equilibrium are central in population genetics theory. In contrast to HWE, linkage equilibrium may be reached very slowly even under ideal conditions. Hardy-Weinberg equilibrium and linkage equilibrium generally simplify the statistical analysis and are assumed when performing genetic linkage analysis. However, it is important to know how valid these simplifying assumptions are by testing the marker loci for HWE and GE.

Fisher's exact test for HWE for each microsatellite locus is presented in Table 2. Eleven microsatellite loci (20%) showed significant deviation from HWE (P-value < 0.05). Fisher's exact tests for GE between adjacent microsatellite loci were also estimated (Table 3). Genetic equilibrium tests permit a combined testing for HWE and linkage equilibrium. Most of the marker pairs were in GE, and only five marker pairs (9%) showed significant deviation (P-value < 0.05) from GE. From the marker pairs that were not in GE, two of them had significant LD (BL1036-BMS2055 and INRA048-BM719; Table 4). Few of the loci departing from HWE proportions had a high number of alleles (Table 2) with low allelic frequency (data not shown). The departures from HWE and GE are expected in some degree because several key assumptions such as random mating, nonoverlapping generations, and infinite population size cannot be met in dairy cattle populations.

In this study, a small fraction of loci did not fulfill the simplifying assumptions of genetic linkage analysis, which are HWE and GE. With this information at hand, these loci can be either discarded or used in the analysis. If deciding to use them, these loci should be scrutinized with caution before claiming that compelling evidence of linkage or LD exists.

Linkage Disequilibrium

Exact LD P-values were estimated for all syntenic and nonsyntenic marker pairs, and a summary of these results is presented in Table 5. The proportion of marker pairs with significant LD (LD P-value < 0.05) for syntenic and nonsyntenic marker pairs was 0.15 and 0.10, respectively. The hypothesis that these proportions are statistically similar could not be rejected. For syntenic marker pairs, significant LD P-values were observed for genetic distances greater than 10 cM (Figure 3). As expected, the extent of LD tends to dissipate with genetic map distance illustrated by a positive relationship between LD P-values and map distances ($r_{xy} = 0.21$).

The extent of LD observed for syntenic and nonsyntenic marker pairs in the North American Holstein sample is similar to that reported for a Dutch Holstein population (Farnir et al., 2000) in which LD between syntenic loci extended over several tens of centimorgan. They also suggested that most of the observed LD was due to random genetic drift. The effect of genetic drift is accentuated in Holstein populations worldwide by the breeding structure of the dairy industry and the

Table 3. Fisher's exact test for genetic equilibrium[1] (GE) between adjacent microsatellite loci typed in elite North American Holstein bulls.[2]

1st Locus BTA	1st Locus Posit. (cM)	1st Locus	2nd Locus	1st–2nd loci GE (*P*-value)[3]	SE
1	8	BM8139	BMS2519	0.865	0.007
2	102	BMS2519	HUJII77	0.487	0.010
3	87	HUJII77	BR4502	0.593	0.010
3	103	BR4502	BMC4214	0.981	0.003
3	123	BMC4214	BL1024	0.433	0.010
4	4	BL1024	TCRB	0.593	0.010
4	97	TCRB	BMS1095	0.532	0.010
5	0	BMS1095	CSSM22	1.000	0.000
5	71	CSSM22	ILSTS093	0.179	0.008
6	0	ILSTS093	INRA133	0.456	0.010
6	8	INRA133	BMS360	0.822	0.008
6	67	BMS360	ILSTS035	0.010	0.002
6	81	ILSTS035	BM4311	0.111	0.006
6	90	BM4311	BM7160	0.247	0.009
7	0	BM7160	BM6105	0.771	0.008
7	36	BM6105	UWCA20	0.319	0.009
7	60	UWCA20	BM711	0.299	0.009
8	84	BM711	BMS2377	0.301	0.009
9	64	BMS2377	BM875	0.209	0.008
10	47	BM875	CSSM46	0.316	0.009
10	93	CSSM46	BM827	0.617	0.010
11	0	BM827	ILSTS049	0.168	0.007
11	55	ILSTS049	RM150	0.223	0.008
11	66	RM150	BL1103	0.115	0.006
11	91	BL1103	BMS585	0.096	0.006
12	80	BMS585	BMS1316	0.317	0.009
12	99	BMS1316	BMS2319	0.339	0.009
13	85	BMS2319	BMS947	0.194	0.008
14	53	BMS947	BL1036	0.406	0.010
14	79	BL1036	BMS2055	0.021	0.003
14	84	BMS2055	BR3510	0.923	0.005
15	1	BR3510	BMS2533	0.590	0.010
15	5	BMS2533	BMS1004	0.764	0.008
15	7	BMS1004	BM848	0.048	0.004
15	82	BM848	BM121	0.813	0.008
16	24	BM121	TGLA53	0.263	0.009
16	41	TGLA53	INRA048	0.224	0.008
16	73	INRA048	BM719	0.005	0.001
16	78	BM719	BM1862	0.463	0.010
17	86	BM1862	BM8151	0.012	0.002
18	42	BM8151	HEL10	0.142	0.007
19	16	HEL10	BMS2142	0.668	0.009
19	45	BMS2142	BP20	0.593	0.010
19	47	BP20	BM846	0.536	0.010
21	66	BM846	RM185	0.705	0.009
23	46	RM185	BMS2269	0.131	0.007
23	61	BMS2269	BM226	0.835	0.007
24	6	BM226	BMS1926	0.580	0.010
24	57	BMS1926	BM737	0.982	0.003
25	28	BM737	BMS1353	0.682	0.009
25	45	BMS1353	CSSM43	0.656	0.010
27	34	CSSM43	BMC6020	0.289	0.009
28	3	BMC6020	BM4602	1.000	0.000
29	0	BM4602			

[1]The computer program Mendel version 4.0 (Lange et al., 1988) was used to estimate Fisher's exact test for GE.

[2]Twenty-three elite North American Holstein bulls were typed for 54 microsatellite loci.

[3]Exact *P*-value estimated using 10,000 permutations.

relatively small effective size of dairy cattle populations.

The extent and distribution of LD in the bovine genome will affect the goals of testing for association and gene localization in different ways. It is simpler to test for association if LD extends over long distances around the disease mutation, because not as many markers are needed to scan for associations. However, at a later

Table 4. Pattern of haplotype frequency distribution (HFD) that resembles linkage disequilibrium (LD) produced by background LD (BLD) and disease trait associated LD (DLD) in syntenic and nonsyntenic marker pairs with significant LD *P*-values.[1]

BTA[2]	Marker 1	Marker 2	Total normalized disequilibrium (D′)[3]	Pattern of HFD[4]
6	ISTS093	BMS360	0.565	ND
6	ILSTS035	BM4311	0.426	DLD
14	BMS947	BL1036	0.649	DLD
14	BL1036	BMS2055	0.639	BLD
15	BR3510	BMS1004	0.567	ND
16	INRA048	BM719	0.266	DLD
23	RM185	BMS2269	0.642	DLD
5, 13	BM121	BM226	0.547	BLD
6, 27	BM6105	CSSM43	0.590	BLD
11, 15	BM8139	BMS1095	0.403	BLD
15, 19	BM875	BM121	0.684	ND
9, 21	BM875	BMS2319	0.686	ND
10, 16	BMC4214	BM7160	0.593	ND
10, 13	BMS1316	TGLA53	0.519	BLD
11, 13	BMS2319	BR3510	0.613	BLD
16, 24	BMS2377	BM846	0.445	DLD
1, 3	BMS2519	CSSM46	0.618	BLD
3, 7	BMS2533	BMS2142	0.646	BLD
5, 10	BMS585	BMS1004	0.691	ND
13, 15	CSSM22	BM875	0.794	ND
3, 7	CSSM22	BMS2319	0.487	ND
7, 27	CSSM46	BMS1926	0.764	ND
11, 19	HUJII77	UWCA20	0.686	BLD
12, 16	ILSTS035	CSSM43	0.785	ND
2, 10	ILSTS049	BMS2319	0.679	ND
10, 24	RM150	BMS2142	0.663	BLD
12, 15	RM150	BMS2533	0.707	BLD

[1]Exact LD *P*-value < 0.05. Exact LD *P*-value estimated using 10,000 permutations and the computer program Arlequin version 2.0 (Schneider et al., 2000).

[2]*Bos taurus* autosome.

[3]Pair-wise HFD and total normalized disequilibrium (D′; Lewontin, 1964) estimated using the computer program 3locus version 5.0 (Long, 1999).

[4]Pattern of typical HFD for DLD and BLD; ND, Non-determined pattern of HFD which is likely due to BLD.

stage, when the goal is to infer gene location, long-ranging LD is potentially problematic. This means that strong associations may be observed far from the causative site(s), and these associations could lead to effort spent in the wrong genomic regions.

In the human genome, the distribution and extent of LD is quite variable and much smaller, respectively. In the bovine genome, significant LD extends over large distances (Farnir et al., 2000), and little empirical information on the distribution of LD in the bovine genome is known. In human populations, reports on LD are quite variable and extend from 5 kb to 4 Mb (Huttley et al., 1999; Pritchard and Przeworski, 2001; Service et al., 2001). As a consequence, the number of markers that will be needed to scan the human genome for association is very large. In contrast, fewer markers may be needed to perform genome association studies in the bovine genome. However, the fine-localization of these

Table 5. Comparison[1] of proportions of syntenic and nonsyntenic marker pairs with significant[2] linkage disequilibrium (LD) *P*-value.

Type of marker pair	Significant LD	Nonsignificant LD	Total	Proportion with significant LD	χ^2	$P > \chi^2$, 1 df
Syntenic	7	40	47	0.149	1.49	0.223
Nonsyntenic	132	1252	1384	0.095		
Total	139	1292	1431	0.097		

[1]Comparison of proportions performed using the comparison of *m* proportions (Fleiss, 1981).

[2]Exact LD *P*-value < 0.05. Exact LD *P*-value estimated using 10,000 permutations and the computer program Arlequin version 2.0 (Schneider et al., 2000).

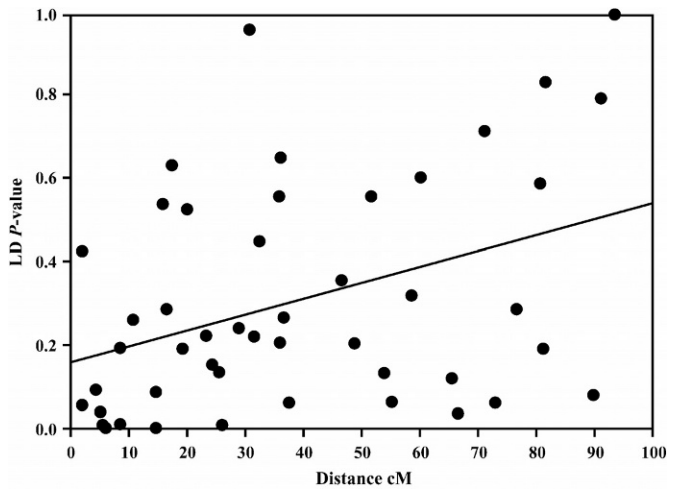

Figure 3. Distribution of linkage disequilibrium (LD) *P*-values for syntenic marker pairs as a function of genetic distance (cM) in the North American Holstein cattle. Regression of LD *P*-value on genetic distance ($b_{y.x}$) was 0.001 and correlation between LD *P*-value and genetic distance (r_{xy}) was 0.21.

Table 6. Number of marker pairs with haplotype frequency distribution (HFD)[1] that resembles linkage disequilibrium (LD) produced by background LD (BLD) and disease trait associated LD (DLD) in syntenic and nonsyntenic marker pairs that had significant LD *P*-values.[2]

Pattern of HFD[3]	Syntenic marker pairs	Nonsyntenic marker pairs[4]
BLD	1	10
DLD	4	1
ND	2	9
Total	7	20
Proportion of DLD[5]	0.57	0.05

[1]Pair-wise disequilibria (Lewontin, 1964) and HFD were estimated for each marker pair using the computer program 3Locus version 5.0 (Long, 1999).

[2]Exact LD *P*-value < 0.05; exact LD *P*-values estimated using 10,000 permutations and the computer program Arlequin version 2.0 (Schneider et al., 2000).

[3]Pattern of BLD and DLD determined by analyzing the pattern of HFD of each marker pair; ND, Nondetermined pattern of HFD that is likely due to BLD.

[4]A sample of 20 marker pairs with lowest LD *P*-value (from a total of 132 nonsyntenic marker pairs with LD *P*-value < 0.05).

[5]Comparison of *m* proportions (Fleiss, 1981), $\chi^2 = 57.4$; $P > \chi^2$ (1 df) = 0.00001.

genes may be a difficult (if not impossible) task in dairy cattle populations. Recently, a successful positional cloning of a QTL was reported in dairy cattle (Grisart et al., 2002). This was possible because several ideal conditions were met: large gene effects, one single mutation in a gene (absence of allelic genetic heterogeneity), and an easily interpretable missense mutation (rather than a regulatory promoter mutation). Forthcoming QTL cloning experiments are likely to be more complicated, because these ideal conditions may not apply.

Background and Disease Trait Associated LD

To estimate the proportion of marker pairs that resembled a typical pattern of BLD and DLD, pair-wise HFD and D′ were estimated for 27 marker pairs: seven syntenic marker pairs with significant LD (LD *P*-value < 0.05, Table 5), and a sample of 20 nonsyntenic marker pairs that had the lowest LD *P*-value (sampled from 132 marker pairs with LD *P*-value < 0.05; Table 5). To illustrate this analysis, the HFD of two marker pairs are shown in Figure 1: a syntenic marker pair (INRA048-BM719) displaying a typical pattern of HFD produced by DLD (Figure 1b); and a nonsyntenic marker pair (HUJII77-UWCA20) presenting a typical pattern of HFD due to BLD (Figure 1a). The pattern of observed HFD for the 27 marker pairs is presented in Table 4. Using D′, most of these marker pairs exhibited strong LD, except one marker pair (INRA048-BM719). Approximately half of the syntenic markers pairs (57%) presented a typical pattern of DLD (Table 6). As expected, few of the nonsyntenic marker pairs (5%) had a HFD that resembles those produced by DLD (Table 6).

The fact that a portion (57%) of observed LD in syntenic marker pairs resembles those likely produced by DLD is encouraging. These results suggest that the observed LD in the US Holstein population is not purely due to genetic drift and that a portion might be due to marker-disease trait loci LD or tight linkage. This raises our hopes of successful fine-localization of genes affecting complex disease traits using LD mapping in the US Holstein cattle population.

Previous reports and findings reported here indicate that LD extends over large distances in dairy cattle populations and that most is due to random genetic drift. Based on the current breeding structure of the dairy industry, it is reasonable to predict that dairy cattle populations will not expand quickly and that they will display high levels of BLD mostly due to genetic drift and migration in the foreseeable future. Given this likely scenario, the extent of LD observed in dairy cattle populations will be useful in mapping chromosomal regions containing genes affecting complex disease traits. The success in pinpointing the causal genes for a QTL effect will greatly depend on the study design, accuracy of phenotype measurement, size of gene effects, level of genetic heterogeneity, extent and distribution of BLD and DLD, and the use of refined statistical methods that account for BLD to minimize the rate of false positive findings.

In a few QTL cloning experiments, some ideal conditions will be met, and the LD mapping methods used will be successful in pinpointing the gene(s) and polymorphism(s) responsible for the effect (Grisart et al., 2002). However, in most QTL cloning experiments, these ideal conditions will not be met and cloning will be a complicated task. For example, Blott et al. (2003) report a nonsynonymous mutation (F279Y) in the growth hormone receptor (**GHR**) gene that contributes to the QTL effect on milk yield and composition. This mutation accounts for 3 to 5% of total trait variation that indicates that additional genes might contribute to the QTL effects observed on *Bos taurus* autosome 20. The fact that the maximum log of the odds score is distal to the GHR gene (42-cM far and outside of the 95% QTL CI) indicates either the mutation F279Y is not the causative mutation or other closely linked genes may be responsible for the observed effects on *Bos taurus* autosome 20. Thus, when LD extends over long regions and is mostly due to genetic drift and migration (i.e., BLD), refined statistical methods that account for BLD must be used or incorrect candidate genes (or genomic regions) may be identified and studied.

If one is applying haplotype analysis methods or searching for shared chromosomal segments, the high levels of BLD will increase the rate of false positives. Therefore, shared segment approaches are liberal due to the BLD, whether or not a disease trait allele exists in their vicinity. Furthermore, LD generated by genetic drift is not expected to present itself in the form of predominantly shared segments or haplotypes. As a result, such approaches to gene mapping are not very powerful when BLD is present since it will not take this form. In contrast, single marker analysis should benefit from the marker-marker correlations, and multiple two-point analysis is expected to be close to optimal for detecting this type of LD (Terwilliger et al., 1998).

The difference between a rapidly growing population and one that remains of constant size is that substantial LD between closely linked loci can be created by genetic drift alone in a population of constant size but not in one that has grown sufficiently rapidly (Slatkin, 1994). In relatively young populations of constant size, such as the bovine genome, genome-wide LD mapping will be feasible even without dense marker maps (for mapping chromosomal regions).

However, BLD will confuse the interpretation of LD analysis for mapping complex disease trait loci, as most methods of LD analysis assume linkage equilibrium between markers in control chromosomes (i.e., individuals not affected with the disease or trait of interest and sampled independently from one another). For LD mapping to succeed, it will be necessary to develop statistical methods that distinguish DLD from BLD, by either accounting for observed BLD or modeling the population history through coalescent methods.

As geneticists move from the mapping of relatively tractable Mendelian disorders to the identification of loci underlying complex disease traits, the utility of LD mapping approaches either in large farm animals or biomedical research remains a challenging task. Effective experimental design and sampling scheme based on adequately justified criteria defined by sound population genetic principles and empirical information on the distribution of LD in the bovine genome will be crucial in the mapping of complex disease trait loci through genome-wide association studies.

CONCLUSIONS

The heterozygosity for the elite Holstein bulls used in this study ranged from 0.43 to 0.80. This wide range of heterozygosity allows us to select the most heterozygous bulls to develop informative families for linkage and LD mapping of complex disease trait loci. The degree of genetic diversity observed in this purebred Holstein population is significant and actually allows breeders to select for traits of economic importance.

As expected, there is extensive LD in the US Holstein cattle population that confirms previous reports on the distribution of LD in Dutch Holstein cattle (Farnir et al., 2000). Approximately half of the syntenic marker pairs presented a typical pattern of disease trait associated LD and, as expected, few of the nonsyntenic marker pairs had a HFD produced by DLD. These results suggest that the observed LD in the US Holstein population is not purely due to genetic drift and that a portion may be due to DLD. This raises our hopes of successful fine-localization of genes affecting complex disease traits using LD mapping in the US Holstein cattle population.

Background LD should be studied in cattle populations using a population-based sample and a reference set of closely linked and evenly spaced highly polymorphic microsatellite markers and single nucleotide polymorphisms. These studies will clarify factors that influence the distribution and magnitudes of BLD in the bovine genome, aid in the dissection of BLD from LD associated with disease trait loci, and facilitate the design of optimal genome-wide association studies.

ACKNOWLEDGMENTS

We thank P. VanRaden and D. Norman (Animal Improvement Programs Laboratory, USDA) for facilitating the use of data records on Holstein dairy cows. We also thank J. B. Cooper, R. DeLuca, and R. Schnabel

for their expert technical support. We are grateful to the CDDR contributors for donating the semen samples. This research was partly funded by USDA Grant No. 2001-34163-10544 to RLV and USDA-ARS-CRIS Project No. 1265-31000-073 to MSA. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by The Pennsylvania State University, University of Tennessee, or the U. S. Department of Agriculture.

REFERENCES

Ashwell, M. S., and C. P. Van Tassell. 1999. Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. J. Dairy Sci. 82:2497–2502.

Ashwell, M. S., C. E. Rexroad, R. H. Miller, and P. M. VanRaden. 1996. Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. Anim. Genet. 27:235–242.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. F. Sunden, G. A. Hawkins, S. Solinas Toldo, R. Fries, M. D. Grosz, J. Yoo, and C. W. Beattie. 1994. A genetic linkage map for cattle. Genetics 136:619–639.

Bitti, P. P., B. S. Murgia, A. Ticca, R. Ferrai, L. Musu, M. L. Piras, E. Puledda, S. Campo, S. Durando, C. Montomoli, D. G. Clayton, A. P. Mander, and L. Bernardinelli. 2001. Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. Genet. Epidemiol. 20:271–283.

Blott, S., J. J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a QTL: A phenylalanine to tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253–266.

de la Chapelle, A., and F. A. Wright. 1998. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc. Natl. Acad. Sci. USA 95:12416–12423.

Devlin, B., K. Roeder, C. Otto, S. Tiobech, and W. Byerley. 2001. Genome-wide distribution of linkage disequilibrium in the population of Palau and its implications for gene flow in remote Oceania. Hum. Genet. 108:521–528.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D, Wagenaar, and M. Georges. 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. 10:220–227.

Fleiss, J. L. 1981. Statistical Methods for Rates and Proportions. 2nd ed. John Wiley & Sons, New York, NY.

Freimer, N. B., S. K. Service, and M. Slatkin. 1997. Expanding on population studies. Nat. Genet. 17:371–373.

Gaspar, C., I. Lopes-Cendes, S. Hayes, J. Goto, K. Arvidsson, A. Dias, I. Silveira, P. Maciel, P. Coutinho, M. Lima, Y. X. Zhou, B. W. Soong, M. Watanabe, P. Giunti, G. Stevanin, O. Riess, H. Sasaki, M. Hsieh, G. A. Nicholson, E. Brunt, J. J. Higgins, M. Lauritzen, L. Tranebjaerg, V. Volpini, N. Wood, L. Ranum, S. Tsuji, A. Brice, J. Sequeiros, and G. A. Rouleau. 2001. Ancestral origins of the Machado-Joseph disease mutation: A worldwide haplotype study. Am. J. Hum. Genet. 68:523–528.

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

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. Genome Res. 12:222–231.

Huttley, G. A., M. W. Smith, M. Carrington, and S. J. O'Brien. 1999. A scan for linkage disequilibrium across the human genome. Genetics 152:1711–1722.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard, T. P. L. Smith, N. L. Lopez Corrales, and C. W. Beattie. 1997. A second-generation linkage map of the bovine genome. Genome Res. 7:235–249.

Kim, K. S., J. S. Yeo, and C. B. Choi. 2002. Genetic diversity of North-East Asian cattle based on microsatellite data. Anim. Genet. 33:201–204.

Laan, M., and S. Paabo. 1997. Demographic history and linkage disequilibrium in human populations. Nat. Genet. 17:435–438.

Lange, K., M. Boehnke, and D. E. Weeks. 1988. Programs for pedigree analysis: Mendel, Fisher, and dGene. Genet. Epidemiol. 5:471–472.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49–67.

Long, J. C. 1999. Documentation for program 3locus.pas version 5.0. Laboratory of Neurogenetics, National Institute on Alcoholism and Alcohol Abuse, National Institutes of Health, Bethesda, MD.

Maudet, C., G. Luikart, and P. Taberlet. 2002. Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. J. Anim. Sci. 80:942–950.

Mukhopadhyay, N., L. Almasy, M. Schroeder, W. P. Mulvihill, and D. E. Weeks. 1999. Mega2, a data handling program for facilitating genetic linkage and association analysis. Am. J. Hum. Genet. 65(Suppl.):A436. (Abstr.)

Peltonen, L. 2000. Positional cloning of disease genes: Advantages of genetic isolates. Hum. Hered. 50:66–75.

Pritchard, J. K., and M. Przeworski. 2001. Linkage disequilibrium in humans: Models and data. Am. J. Hum. Genet. 69:1–14.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. version 2.0. Genetics and Biometry Lab., Department of Anthropology, University of Geneva.

Service, S. K., R. A. Ophoff, and N. B. Freimer. 2001. The genome-wide distribution of background linkage disequilibrium in a population isolate. Hum. Mol. Genet. 10:545–551.

Sheffield, V. C., E. M. Stone, and R. Carmi. 1998. Use of isolated inbred human populations for identification of disease genes. Trends Genet. 14:391–396.

Shinar, Y., B. Ben-Zeev, N. Brand, H. Lahat, V. Gross-Zur, D. MacGregor, T. Bahan, D. L. Kastner, and E. Pras. 2002. A common ancestral haplotype in carrier chromosomes from different ethnic backgrounds in vacuolating megalencephalic leucoencephalopathy with subcortical cysts. J. Med. Genet. 39:54–57.

Slatkin, M. 1994. Linkage disequilibrium in growing and stable populations. Genetics 137:331–336.

Terwilliger, J. D., S. Zollner, M. Laan, and S. Paabo. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: "drift mapping" in small populations with no demographic expansion. Hum. Hered. 48:138–154.

Wright, A. F., A. D. Carothers, and M. Pirastu. 1999. Population choice in mapping genes for complex diseases. Nat. Genet. 61:151–161.