Mammalian Genome 12, 219–226 (2001).
DOI: 10.1007/s003350010269

Mammalian Genome
Incorporating Mouse Genome

Interleukin-8 haplotype structure from nucleotide sequence variation in commercial populations of U.S. beef cattle

Michael P. Heaton, Carol G. Chitko-McKown, W. Michael Grosse,* John W. Keele, James E. Keen, William W. Laegreid

USDA, ARS, U.S. Meat Animal Research Center (MARC), State Spur 18D, P.O. Box 166, Clay Center, Nebraska 68933-0166, USA

Received: 29 August 2000 / Accepted: 17 November 2000

Abstract. The aim of the present study was twofold: first, to design a panel of 96 sires that reflects the breadth of genetic diversity in U.S. beef cattle, and second, to use this panel to discover nucleotide sequence diversity and haplotype structures of interleukin (*IL*)-*8* in commercial populations. The latter is a requisite for epidemiological studies designed to test whether *IL8* alleles are risk factors for acquiring or maintaining bacterial infections in production environments. *IL-8* encodes a proinflammatory cytokine that plays a central role in cell-mediated immunity by attracting and activating neutrophils in the early stages of host defense against bacterial invasion. Seven single-nucleotide polymorphism (SNP) markers were identified by sequencing two *IL8* DNA segments amplified from the panel of 17 popular cattle breeds (MARC beef cattle diversity panel, version 2.1). Assays for automated genotype scoring by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) were developed to independently verify the seven SNP alleles in the 96 bulls and 313 cattle from the MARC reference population. Five haplotype structures, spanning the two *IL8* DNA segments, were unambiguously defined for the set of seven *IL8* SNPs. Based on the breadth of germplasm in bovine diversity panel, the five haplotype structures for *IL8* are estimated to represent >98% of those present in these DNA segments in commercial populations of U.S. beef cattle. The frequencies of the five respective haplotypes in the eight Angus sires of the diversity panel (0.75, 0.25, 0.00, 0.00, 0.00) were similar to those scored in 150 purebred Angus cattle from six herds in four Midwestern states (0.82, 0.18, 0.01, 0.00, 0.00), suggesting that the diversity panel may also be useful for estimating allele frequencies in commercial populations.

Introduction

Infectious diseases in cattle are a significant source of economic loss and represent a potential risk to human health (Buzby et al. 1996). Improvements in herd health and food safety may result if individuals with the highest risk for infectious disease are eliminated from the production cycle. Our aim is to utilize an animal's DNA sequence information to estimate its risk of acquiring or maintaining infections.

Predicting an individual's genetic predisposition for infection from its DNA sequence will be a complex undertaking and require specific knowledge of genotype/phenotype associations in populations exposed to a variety of environmental conditions (Terwilliger and Goring 2000). One route for identifying these associations begins with case-control studies of gene regions that have a fundamental biological relationship to the phenotype (Lander and Schork 1994; Collins and Morton 1998). Two general strategies have been proposed for utilizing single-nucleotide polymorphism (SNP) markers to elucidate genotype/phenotype associations in case-control studies. The first is a direct approach that aims to identify each SNP over the entire candidate gene region to directly identify the disease-predisposing mutation. The second is an indirect approach that aims to identify a fraction of SNPs within the candidate gene region and detect the disease-predisposing mutation through linkage disequilibrium. The latter approach is popular because it requires fewer SNPs and has the potential of detecting allelic associations with disease-predisposing mutations across a wider genomic region. When this latter approach is combined with well-designed case-control studies, an association may be detected if the marker allele occurs at a significantly higher frequency in affected individuals compared with control individuals (Lander 1996; Collins et al. 1997, 1999; Chakravarti 1999; Kruglyak 1999; Terwilliger and Goring 2000). Because the information content of an individual SNP is inherently low (biallelic), the set of SNP alleles residing on a specific segment of a chromosome may be used as a group (i.e., DNA segment haplotype) to discern additional allelic variants in the candidate gene DNA segment. Haplotype analysis of candidate genes in target populations is important because it lends needed power to SNP-association studies designed to detect the disease-predisposing mutations through linkage disequilibrium (Wright et al. 1999; Zollner and von Haeseler 2000).

A previous report demonstrated an increased abundance of *IL8* RNA in bovine epithelial cells in response to *Escherichia coli* O157:H7 lipopolysaccharide (Heaton et al. 1999). *IL8* encodes a proinflammatory cytokine that attracts and activates neutrophils in the early stages of host defense against bacterial invasion (Zlotnik et al. 1999). Two haplotypes were described in the MARC reference cattle population for an *IL8* DNA segment spanning intron 2 (Heaton et al. 1999). Whereas these two haplotype alleles were sufficient for determining the *IL8* location on the bovine linkage map, efficient use of gene-based SNP markers in commercial populations of cattle will require: 1) identification of the majority of haplotype alleles for each gene; 2) development of robust, accurate assays for scoring the genotypes; and 3) estimation of allele frequencies in relevant breeds (Heaton et al. 2001). The present report identifies five haplotype structures for *IL8*, four matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) genotype assays for automatically scoring the seven SNPs, and provides an estimation of the allele and genotype frequencies in 17 breeds from commercial populations.

Materials and methods

Animal groups and genomic DNA samples. Animal pedigrees from leading suppliers of U.S. beef cattle semen were analyzed by breed and

**Present address:* CuraGen Corporation, 322 East Main Street, Branford, CT 06405, USA.

Correspondence to: M.P. Heaton; E-mail: heaton@email.marc.usda.gov.

Table 1. Oligonucleotides for *IL-8* amplification, DNA sequencing, and MALDI-TOF MS genotyping.

DNA segment[a]	SNP marker	Oligo	Oligo sequence[b]	Orientation	Function	Termination mixture	Mass (*m/z*)[c]	Multiplex marker(s)	Annealing temp. (°C)	Amplicon length (bp)
IL8DS5	—[d]	AF061521:2U21	cacattccacacctttctacc	Sense	amplification	—	—	—	55	318
IL8DS5	—	AF061521:302L18	cttcaaacaccaccttgc	Antisense	amplification	—	—	—	55	318
IL8DS5	—	AF061521:302L18B2	B2-cttcaaacaccaccttgc	Antisense	amplification	—	5932.6	—	55	318
IL8DS5	—	AF061521:42U21	tgagagttattgagagtgggc	Sense	sequencing	—	—	—	62	—
IL8DS5	—	AF061521:274L22	agtcatggttccttgagtacag	Antisense	sequencing	—	—	—	62	—
IL8DS5	AH2-1	AH2-1.19U19	ggtactaggagtacatagt	Sense	PROBE primer	ddA/ddT	5891.9	AH2-2	40	—
IL8DS5	AH2-1	AH2-1.19U20AT.A	ggtactaggagtacatagta	Sense	A allele analyte	ddA/ddT	6189.1	AH2-2	40	—
IL8DS5	AH2-1	AH2-1.19U21AT.G	ggtactaggagtacatagtga	Sense	G allele analyte	ddA/ddT	6518.3	AH2.2	40	—
IL8DS5	AH2-2	AH2-2.16U16	agtcagcctagaattg	Sense	PROBE primer	ddA/ddT	4905.3	AH2-1	40	—
IL8DS5	AH2-2	AH2-2.16U17AT.A	agtcagcctagaattga	Sense	A allele analyte	ddA/ddT	5202.5	AH2-1	40	—
IL8DS5	AH2-2	AH2-2.16U18AT.C	agtcagcctagaattcga	Sense	C allele analyte	ddA/ddT	5820.9	AH2-1	40	—
IL8DS5	AH2-3	AH2-3.22U22	ttagtcagcctagaattgagac	Sense	PROBE primer	ddA/ddT	6696.5	—	40	—
IL8DS5	AH2-3	AH2-3.22U23AT.A	ttagtcagcctagaattgagaca	Sense	A allele analyte	ddA/ddT	7055.7	—	40	—
IL8DS5	AH2-3	AH2-3.22U24AT.G	ttagtcagcctagaattgagacgt	Sense	G allele analyte	ddA/ddT	7375.9	—	40	—
IL8DS4	—	AF061521:487U21	atccttttccattgcttcta	Sense	amplification	—	—	—	55	344
IL8DS4	—	AF061521:814L17	aaggcgcagttcaacag	Antisense	amplification	—	—	—	55	344
IL8DS4	—	AF061521:814L17B1	B1-aaggcgcagttcaacag	Antisense	amplification	—	5633.5	—	55	344
IL8DS4	—	AF061521:509U21	gaattcctcagtaaagatgcc	Sense	sequencing	—	—	—	62	—
IL8DS4	—	AF061521:775L18	caggtgagggttgcaaga	Antisense	sequencing	—	—	—	62	—
IL8DS4	AH2-4	AH2-4.16U16	tgtgggtctggtgtag	Sense	PROBE primer	ddA/ddT	4999.3	AH2-7	40	—
IL8DS4	AH2-4	AH2-4.16U17AT.A	tgtgggtctggtgtaga	Sense	A allele analyte	ddA/ddT	5296.5	AH2-7	40	—
IL8DS4	AH2-4	AH2-4.16U19AT.G	tgtgggtctggtgtagggt	Sense	G allele analyte	ddA/ddT	5945.9	AH2-7	40	—
IL8DS4	AH2-5	AH2-5.17U17	gaatatttttcagtgt	Sense	PROBE primer	ddC/ddG	5205.5	AH2-6	40	—
IL8DS4	AH2-5	AH2-5.17U18.CG.C	gaatatttttcagtgtc	Sense	C allele analyte	ddC/ddG	5478.6	AH2-6	40	—
IL8DS4	AH2-5	AH2-5.17U19.CG.T	gaatatttttcagtgttg	Sense	T allele analyte	ddC/ddG	5822.9	AH2-6	40	—
IL8DS4	AH2-6	AH2-6.20U20	tcctgctttttttgttttcgg	Sense	PROBE primer	ddC/ddG	6000.0	AH2-5	40	—
IL8DS4	AH2-6	AH2-6.20U21.CG.C	tcctgctttttttgttttcggc	Sense	C allele analyte	ddC/ddG	6335.2	AH2-5	40	—
IL8DS4	AH2-6	AH2-6.20U26.CG.T	tcctgctttttttgttttcggttttttg	Sense	T allele analyte	ddC/ddG	7896.2	AH2-5	40	—
IL8DS4	AH2-7	AH2-7.21U21	tttgttaatcttgcaacctc	Sense	PROBE primer	ddA/ddT	6322.2	AH2-4	40	—
IL8DS4	AH2-7	AH2-7.21U22AT.A	tttgttaatcttgcaaccctca	Sense	A allele analyte	ddA/ddT	6619.4	AH2-4	40	—
IL8DS4	AH2-7	AH2-7.21U25AT.G	tttgttaatcttgcaaccctcgcct	Sense	G allele analyte	ddA/ddT	7518.0	AH2-4	40	—

[a]Amplicon sequences for IL8DS5 and IL8DS4 are contained within GenBank reference sequence AF061521.
[b]Sequences listed 5′ to 3′. B1 is biotin with mass 406.4 Da; B2 is biotin with mass 569.6 Da.
[c]Mass-to-change ratio.
[d]Not applicable.

Table 2. Composition of MARC beef cattle diversity panel, version 2.1.

Animal group	Registrations[a]	Fraction of total registrations[b]	Animal genomes in panel	Unshared haploid genomes[c]	Minimum allele frequency required for its detection[d]
Angus	239,476	0.32	8	15.7	0.17
Hereford	106,608	0.14	8	15.9	0.17
Limousin	61,462	0.08	8	15.6	0.18
Simmental	51,390	0.07	7	13.8	0.20
Charolais	49,223	0.07	6	11.7	0.23
Beefmaster	47,349	0.06	5	9.5*	0.27
Red Angus	33,875	0.05	6	11.8	0.22
Gelbvieh	30,178	0.04	6	11.5	0.23
Brangus	27,727	0.04	5	9.3	0.28
Salers	20,144	0.03	5	9.0*	0.28
Brahman	16,000	0.02	6	12.0	0.22
Shorthorn	15,474	0.02	5	9.9	0.26
Maine-Anjou	12,300	0.02	5	10.0*	0.26
Texas Longhorn	12,000	0.02	4	7.8*	0.32
Santa Gertrudis	11,000	0.01	4	8.0	0.31
Chianina	7,864	0.01	4	7.9	0.32
Other beef breeds	7,205	0.01	0	na[e]	na
Beef total:	749,275	1.00	92	179.2	0.017
Holstein	312,116	0.80	4	7.9*	0.32
Other dairy breeds	78,867	0.20	0	na	na
Dairy total:	390,983	1.00	4	7.9	0.32
Total	1,140,258	na	96	187.1	0.016

[a]National Pedigreed Livestock Council Annual Report, February 1999.
[b]Fraction of either total beef or dairy registrations, respectively.
[c]An average based on shared ancestors in pedigrees with at least four generations present. Asterisk (*) denotes the presence of some pedigrees with only three generations. Sires within breed were selected first on genetic relatedness and second on availability of semen.
[d]This "threshold" frequency was defined as the minimum allele frequency at which the probability of observing the allele at least once in an animal group was 0.95. The probability of observing an allele at least once is $1 - (1 - p)^n$ where "p" is the frequency of the allele and "n" is the number of independent samplings, or, in this case, the number of unshared haploid genomes for diploid organisms. This assumes that samplings (haploid genomes) are independent and identically distributed (the same p applies to all animals). Setting power or the probability of observing the allele at least once to 0.95 results in the equation: $0.95 = 1 - (1 - p)^n$. Solving this equation for p yields $p = 1 - (0.05)^{1/n}$ for all p between 0 and 1.
[e]Not applicable.

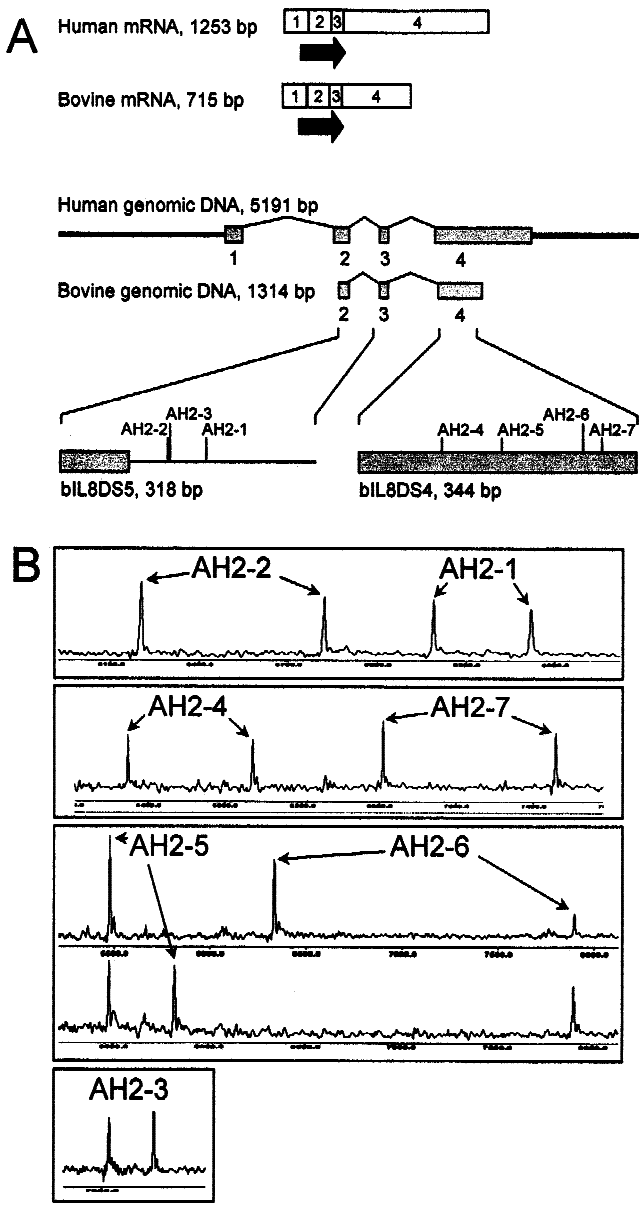

Fig. 1. Physical maps of bovine *IL8* and MALDI-TOF MS spectra of *IL8* SNP markers. Panel **A:** Comparison of the complete human *IL8* sequence (GenBank accession number M28130, (Mukaida et al. 1989) and the partial bovine *IL8* sequences. The partial bovine mRNA sequence was reconstructed from GenBank accession numbers S82598 (Morsey et al. 1996) and AF061521 (Heaton et al. 1999); the latter sequence was edited to include new amplicon sequence (intron 3, 1268 bp) generated with amplification primers AF061521:2U21 and AF061521:814L17. The symbol legend is as follows: shaded rectangles, genomic exon regions; lines connecting shaded rectangles, introns; open rectangles, cDNA sequence; dashed vertical lines, ends of partial cDNA sequences; and solid arrows, coding sequence. The SNP marker names in the MARC cattle database (http://sol.marc.usda.gov) begin with an "AH" prefix, and their relative position is indicated by vertical lines within the amplicon. Panel **B:** MALDI-TOF MS spectra of seven *IL8* SNP markers for heterozygous individuals. Assays were developed and executed as described in the Methods section. All spectral peaks represent singly charged ions whose mass-to-charge ratio (*m/z*) was compared with calibrates for mass determination. The respective masses for each indicated SNP spectral peak (allele analyte) are presented in Table 1.

sorted to identify sires with the minimum amount of shared relationships. DNA was extracted from semen as previously described (Heaton et al. 2001), and 5-μl aliquots of 10 ng/μl solutions were arrayed in 96-well microtiter plates with a Multimek 96-channel pipetting instrument (Beckman Coulter Inc., Fullerton, Calif.). The resulting plates of DNA samples (MARC beef cattle diversity panel, version 2.1; MBCDP2.1) were dried in laminar flow hoods, sealed with adhesive tape, and stored at room temperature until use. The MARC reference cattle population (Bishop et al. 1994) was used as a source of DNA for validating SNP segregation and detecting allele amplification problems. An additional 79 progeny were included (262 total) to increase the power of segregation analysis in 313 members of the reference population and thereby screen against markers displaying non-Mendelian inheritance patterns. DNA from this population was arrayed in 96-well microtiter plates as described above and was stored at room temperature until use. To compare allele frequencies between Angus sires in MBCDP2.1 with those in larger commercial populations, DNA from 150 purebred Angus cattle sampled from six herds in four Midwestern states was arrayed 96-well microtiter plates as described above. The herd selection was based on the owners' willingness to cooperate. The four American bison genomic samples used for comparison were described elsewhere (Heaton et al. 2001).

PCR amplification, DNA sequencing, and MALDI-TOF MS assays. Primers (Table 1) were designed from DNA sequence information present in the National Center for Biotechnology Information databases (http:/www.ncbi.nlm.nih.gov, GenBank accession no. AF061521) with primer analysis software Oligo version 6.0 (Molecular Biology Insights, Inc., Cascade, Colo.). Two regions of *IL8* were targeted for amplification. The first region (bIL8DS5) encompassed the two previously known SNPs (AH2-1, AH2-2). The second region (bIL8DS4) encompassed a portion of exon 4 that is predicted to encode 3′ untranslated mRNA. In each case, the selection of amplification primers was influenced by the availability of known genomic sequence and the prediction of stringent amplification. For early detection of major allele amplification problems (e.g., amplification of more than one genomic region), primers were first used to amplify four families (24 individuals) in the MARC reference population and screened against the presence of apparent heterozygous sites in all individuals. Amplicons were then tested with individuals in the MBCPD2.1, and PCR products were sequenced from both strands with nested primers (Table 1) as previously described (Grosse et al. 1999). Animal sequences from both strands were analyzed with polyphred software (Nickerson et al. 1997) in conjunction with the phred/phrap/consed software (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). Candidate SNPs were manually identified with consed software and verified with a MALDI-TOF MS system that used primer oligonucleotide base extension (PROBE), nano-liter dispensing of extension products onto silicon chips (Sequenom, Inc., San Diego, Calif.) (Little et al. 1997c), and fully automated mass spectrometric analysis. A description of the reagents and analytes for the *IL8* MALDI-TOF MS assays is provided in Table 1. The principles of the PROBE assay and chip-based mass spectrometry (Braun et al. 1997; Little et al. 1997a, 1997b, 1997c) and the technical procedures (Heaton et al. 2001) were previously described. MALDI-TOF MS assays of candidate SNPs were used to score the MBCDP2.1, and the genotypes were compared with those obtained by electrophoretic DNA sequence. SNPs that were apparent by both methods were further characterized by segregation analysis with 313 members of the MARC reference population. SNPs with apparent Mendelian inheritance patterns were advanced to haplotype analysis.

IL8 *haplotype assignment.* Segregation analysis was used to assign haplotypes that were present in 28 founders of the MARC reference population. In addition, some animals from commercial populations contained homozygous genotypes, and thus their haplotypes were unambiguously assigned. Remaining haplotypes were assigned by sequencing a set of individual cloned PCR products from heterozygous sires as previously described (Heaton et al. 2001).

Results

Panel design for SNP discovery in U.S. beef cattle. A panel of animal DNA was assembled to broadly sample the genetic diversity of commercial populations, yet limit the group to 96 samples for automated DNA sequencing and genotype analysis (Table 2). This animal group contained 92 sires from 16 beef breeds and four sires from one dairy breed. The beef breeds comprise greater than

Table 3. SNP allele and genotype frequencies for *IL8* in beef cattle[a].

	AH2-1					AH2-2					AH2-3				
	Allele[c]		Genotype			Allele		Genotype			Allele		Genotype		
Animal Group[b]	G	A	G,G	G,A	A,A	A	C	A,A	A,C	C,C	G	A	G,G	G,A	A,A
Angus MBCDP2.1	0.75	0.25	0.63	0.25	0.13	0.75	0.25	0.63	0.25	0.13	1.00	—	1.00	—	—
Angus Herds 1 to 6 combined	0.82	0.18	0.68	0.28	0.04	0.82	0.18	0.68	0.28	0.04	1.00	—	1.00	—	—
Angus Herd 1	0.58	0.42	0.33	0.50	0.17	0.58	0.42	0.33	0.50	0.17	1.00	—	1.00	—	—
Angus Herd 2	0.77	0.23	0.65	0.26	0.10	0.77	0.23	0.65	0.26	0.10	1.00	—	1.00	—	—
Angus Herd 3	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
Angus Herd 4	0.95	0.05	0.89	0.11	—	0.95	0.05	0.89	0.11	—	1.00	—	1.00	—	—
Angus Herd 5	0.80	0.20	0.61	0.39	—	0.80	0.20	0.61	0.39	—	1.00	—	1.00	—	—
Angus Herd 6	0.86	0.14	0.76	0.19	0.05	0.86	0.14	0.76	0.19	0.05	1.00	—	1.00	—	—
Hereford MBCDP2.1	0.94	0.06	0.88	0.13	—	0.94	0.06	0.88	0.13	—	1.00	—	1.00	—	—
Limousin MBCDP2.1	0.69	0.31	0.38	0.63	—	0.69	0.31	0.38	0.63	—	1.00	—	1.00	—	—
Simmental MBCDP2.1	0.86	0.14	0.71	0.29	—	0.86	0.14	0.71	0.29	—	1.00	—	1.00	—	—
Charolais MBCDP2.1	0.75	0.25	0.50	0.50	—	0.75	0.25	0.50	0.50	—	1.00	—	1.00	—	—
Beefmaster MBCDP2.1	0.70	0.30	0.40	0.60	—	0.70	0.30	0.40	0.60	—	1.00	—	1.00	—	—
Red Angus MBCDP2.1	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
Gelbvieh MBCDP2.1	0.67	0.33	0.67	—	0.33	0.67	0.33	0.67	—	0.33	1.00	—	1.00	—	—
Brangus MBCDP2.1	0.70	0.30	0.60	0.20	0.20	0.70	0.30	0.60	0.20	0.20	1.00	—	1.00	—	—
Salers MBCDP2.1	0.90	0.10	0.80	0.20	—	0.90	0.10	0.80	0.20	—	1.00	—	1.00	—	—
Brahman MBCDP2.1	1.00	—	1.00	—	—	1.00	—	1.00	—	—	0.92	0.08	0.83	0.17	—
Shorthorn MBCDP2.1	0.50	0.50	—	1.00	—	0.50	0.50	—	1.00	—	1.00	—	1.00	—	—
Maine-Anjou MBCDP2.1	0.40	0.60	0.20	0.40	0.40	0.40	0.60	0.20	0.40	0.40	1.00	—	1.00	—	—
Texas Longhorn MBCDP2.1	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
Santa Gertrudis MBCDP2.1	0.50	0.50	0.25	0.50	0.25	0.50	0.50	0.25	0.50	0.25	1.00	—	1.00	—	—
Chianina MBCDP2.1	0.75	0.25	0.75	—	0.25	0.75	0.25	0.75	—	0.25	1.00	—	1.00	—	—
Holstein MBCDP2.1	0.13	0.88	—	0.25	0.75	0.13	0.88	—	0.25	0.75	1.00	—	1.00	—	—
MARC reference parents	0.77	0.23	0.54	0.46	—	0.77	0.23	0.54	0.46	—	1.00	—	1.00	—	—
Bison	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
MBCDP2.1 Total	0.74	0.26	0.59	0.29	0.11	0.74	0.26	0.59	0.29	0.11	0.99	0.01	0.99	0.01	0.00

[a]Within each animal group, the allele and genotype frequencies are presented as the fraction of haploid and diploid genomes, respectively. A frequency of zero is denoted with a dash.
[b]Genotypes for 28 founders of the MARC reference population and four bison are presented for comparison.
[c]Allele genotype frequencies that sum to 0.99 or 1.01 are the result of rounding errors.

99% of the germplasm used in the U.S. beef cattle industry, based on the number of registered progeny for each breed (NPLC Report 1999). The number of animals used for each breed group was influenced by the number of registered progeny and the probability of allele detection (Table 2, footnote "d"). The lower limit (four genomes) allows identification of the "major" alleles in the breed group (frequencies greater than 0.32). The upper limit (eight genomes) increases the ability to detect "minor" alleles in the breed group (frequencies between 0.17 and 0.32). For selection of individual sires within each breed, the aim was to maximize the total number of unshared haploid genomes by selecting sires with few relationships within and between pedigrees. The resulting panel of 96 sires (MBCDP2.1) was estimated to contain 187 unshared haploid genomes and was expected to allow a 95% probability of detecting any allele with a frequency greater than 0.016 in the group.

Identification of IL8 SNPs in cattle. Two regions of the *IL8* gene were targeted for PCR amplification for use in SNP discovery (Fig. 1A; exon2/intron 2, exon 4). Comparison of the DNA sequences amplified from 187 haploid genomes in MBCDP2.1 showed that the consensus sequences for both amplicons were 100% identical to their respective regions of bovine *IL8* (GenBank accession no. AF061521). The sequence alignment also indicated seven sites for candidate SNPs, two of which were known to be present in the MARC reference population (Heaton et al. 2000).

Four MALDI-TOF MS assays were used to confirm the animal genotypes from electrophoretic sequencing and score the allele and genotype frequencies in 96 sires from MBCDP2.1 and 313 members of the MARC reference population. Both alleles from all seven SNPs were observed in MALDI-TOF MS assays with the discovery panel (Fig. 1B). Five of the seven SNP markers were present in the reference population (AH2-1, 2, 4, 5, and 7), and the observed segregation was consistent with the expected Mendelian patterns. Based on the correspondence of genotypes from two methods and the consistent segregation of individual SNPs, these seven *IL8* SNPs appear to be genuine, and the four MALDI-TOF MS assays appear to provide accurate automated genotype scoring.

Although the SNP discovery panel, MBCDP2.1, was primarily designed to capture the breadth of allelic diversity in commercial populations, it was suggested that this panel may also provide a useful estimation of SNP allele and genotype frequencies for commercial herds of purebred cattle. This hypothesis was tested by scoring the seven *IL8* SNP genotypes in 150 purebred Angus cattle sampled from six herds in four Midwestern states and comparing the SNP frequencies with the eight Angus sires present in the MBCDP2.1. Analysis of all seven *IL8* SNPs showed little difference between the allele and genotype frequencies averaged for six Angus herds and those from the MBCDP2.1 Angus sires. This result suggests that the allele frequencies summarized from the MBCDP2.1 (Table 3) may be useful in predicting which SNPs have optimum informativity in various breeds and provide a reasonable starting point for designing genetic tests that are broadly applicable in commercial populations.

Two features are apparent from the haplotype frequencies presented in Table 3. First, some SNPs were present in 16 of the 17 breeds (e.g., AH2-5), whereas one SNP was apparent in only a single heterozygous Brahman sire (AH2-3). Second, some pairs of SNPs were tightly linked, i.e., had identical allele and genotype frequencies in all groups. For example, the "A" allele frequency of AH2-4 is identical to the "T" allele frequency of AH2-5, regardless of which animal group is analyzed. Thus, if one site had contained the causal mutation for a hypothetical phenotype, the other site would be a useful marker for the adjacent

Table 3. (Continued)

AH2-4					AH2-5					AH2-6					AH2-7				
Allele		Genotype			Allele		Genotype			Allele		Genotype			Allele		Genotype		
A	G	A,A	A,G	G,G	T	C	T,T	T,C	C,C	T	C	T,T	T,C	C,C	G	A	G,G	G,A	A,A
0.75	0.25	0.63	0.25	0.13	0.75	0.25	0.63	0.25	0.13	1.00	—	1.00	—	—	0.75	0.25	0.63	0.25	0.13
0.81	0.19	0.68	0.27	0.05	0.81	0.19	0.68	0.27	0.05	1.00	—	1.00	—	—	0.81	0.19	0.68	0.27	0.05
0.58	0.42	0.33	0.50	0.17	0.58	0.42	0.33	0.50	0.17	1.00	—	1.00	—	—	0.58	0.42	0.33	0.50	0.17
0.74	0.26	0.65	0.19	0.16	0.74	0.26	0.65	0.19	0.16	1.00	—	1.00	—	—	0.74	0.26	0.65	0.19	0.16
1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
0.95	0.05	0.89	0.11	—	0.95	0.05	0.89	0.11	—	1.00	—	1.00	—	—	0.95	0.05	0.89	0.11	—
0.80	0.20	0.61	0.39	—	0.80	0.20	0.61	0.39	—	1.00	—	1.00	—	—	0.80	0.20	0.61	0.39	—
0.86	0.14	0.76	0.19	0.05	0.86	0.14	0.76	0.19	0.05	1.00	—	1.00	—	—	0.86	0.14	0.76	0.19	0.05
0.94	0.06	0.88	0.13	—	0.94	0.06	0.88	0.13	—	1.00	—	1.00	—	—	0.94	0.06	0.88	0.13	—
0.63	0.38	0.25	0.75	—	0.63	0.38	0.25	0.75	—	1.00	—	1.00	—	—	0.63	0.38	0.25	0.75	—
0.86	0.14	0.71	0.29	—	0.86	0.14	0.71	0.29	—	1.00	—	1.00	—	—	0.86	0.14	0.71	0.29	—
0.58	0.42	0.17	0.83	—	0.58	0.42	0.17	0.83	—	1.00	—	1.00	—	—	0.58	0.42	0.17	0.83	—
0.60	0.40	0.40	0.40	0.20	0.60	0.40	0.40	0.40	0.20	0.90	0.10	0.80	0.20	—	0.70	0.30	0.40	0.60	—
1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—	1.00	—	1.00	—	—
0.67	0.33	0.67	—	0.33	0.67	0.33	0.67	—	0.33	1.00	—	1.00	—	—	0.67	0.33	0.67	—	0.33
0.60	0.40	0.40	0.40	0.20	0.60	0.40	0.40	0.40	0.20	1.00	—	1.00	—	—	0.60	0.40	0.40	0.40	0.20
0.90	0.10	0.80	0.20	—	0.90	0.10	0.80	0.20	—	1.00	—	1.00	—	—	0.90	0.10	0.80	0.20	—
0.50	0.50	0.33	0.33	0.33	0.50	0.50	0.33	0.33	0.33	0.67	0.33	0.50	0.33	0.17	0.83	0.17	0.67	0.33	—
0.50	0.50	—	1.00	—	0.50	0.50	—	1.00	—	1.00	—	1.00	—	—	0.50	0.50	—	1.00	—
0.30	0.70	0.20	0.20	0.60	0.30	0.70	0.20	0.20	0.60	1.00	—	1.00	—	—	0.30	0.70	0.20	0.20	0.60
0.88	0.13	0.75	0.25	—	0.88	0.13	0.75	0.25	—	1.00	—	1.00	—	—	0.88	0.13	0.75	0.25	—
0.50	0.50	0.25	0.50	0.25	0.50	0.50	0.25	0.50	0.25	1.00	—	1.00	—	—	0.50	0.50	0.25	0.50	0.25
0.38	0.63	0.25	0.25	0.50	0.38	0.63	0.25	0.25	0.50	1.00	—	1.00	—	—	0.38	0.63	0.25	0.25	0.50
0.13	0.88	—	0.25	0.75	0.13	0.88	—	0.25	0.75	1.00	—	1.00	—	—	0.13	0.88	—	0.25	0.75
0.71	0.29	0.46	0.50	0.04	0.71	0.29	0.46	0.50	0.04	1.00	—	1.00	—	—	0.71	0.29	0.46	0.50	0.04
—	1.00	—	—	1.00	—	1.00	—	—	1.00	1.00	—	1.00	—	—	1.00	—	1.00	—	—
0.66	0.34	0.48	0.35	0.17	0.66	0.34	0.48	0.35	0.17	0.97	0.03	0.96	0.03	0.01	0.68	0.32	0.50	0.36	0.14

Table 4. Haplotype allele and animal genotype frequencies for *IL8* in cattle[a]

	Haplotype Allele[c]						Animal Genotype										
Animal Group[b]	1	2	3	4	5	6	1,1	2,2	3,3	4,4	5,5	1,2	1,4	1,6	2,3	2,4	3,4
Angus MBCDP2.1	0.75	0.25	—	—	—	—	0.63	0.13	—	—	—	0.25	—	—	—	—	—
Angus Herds 1 to 6 combined	0.81	0.18	—	0.01	—	—	0.68	0.04	—	—	—	0.27	—	—	—	0.01	—
Angus Herd 1	0.58	0.42	—	—	—	—	0.33	0.17	—	—	—	0.50	—	—	—	—	—
Angus Herd 2	0.74	0.23	—	0.03	—	—	0.65	0.10	—	—	—	0.19	—	—	—	0.06	—
Angus Herd 3	1.00	—	—	—	—	—	1.00	—	—	—	—	—	—	—	—	—	—
Angus Herd 4	0.95	0.05	—	—	—	—	0.89	—	—	—	—	0.11	—	—	—	—	—
Angus Herd 5	0.80	0.20	—	—	—	—	0.61	—	—	—	—	0.39	—	—	—	—	—
Angus Herd 6	0.86	0.14	—	—	—	—	0.76	0.05	—	—	—	0.19	—	—	—	—	—
Hereford MBCDP2.1	0.94	0.06	—	—	—	—	0.88	—	—	—	—	0.13	—	—	—	—	—
Limousin MBCDP2.1	0.63	0.31	—	0.06	—	—	0.25	—	—	—	—	0.63	0.13	—	—	—	—
Simmental MBCDP2.1	0.86	0.14	—	—	—	—	0.71	—	—	—	—	0.29	—	—	—	—	—
Charolais MBCDP2.1	0.58	0.25	—	0.17	—	—	0.17	—	—	—	—	0.50	0.33	—	—	—	—
Beefmaster MBCDP2.1	0.60	0.30	0.10	—	—	—	0.40	—	—	—	—	0.40	—	—	0.20	—	—
Red Angus MBCDP2.1	1.00	—	—	—	—	—	1.00	—	—	—	—	—	—	—	—	—	—
Gelbvieh MBCDP2.1	0.67	0.33	—	—	—	—	0.67	0.33	—	—	—	—	—	—	—	—	—
Brangus MBCDP2.1	0.60	0.30	—	0.10	—	—	0.40	0.20	—	—	—	0.20	0.20	—	—	—	—
Salers MBCDP2.1	0.90	0.10	—	—	—	—	0.80	—	—	—	—	0.20	—	—	—	—	—
Brahman MBCDP2.1	0.50	—	0.25	0.17	—	0.08	0.33	—	0.17	—	—	—	0.17	0.17	—	—	0.17
Shorthorn MBCDP2.1	0.50	0.50	—	—	—	—	—	—	—	—	—	1.00	—	—	—	—	—
Maine-Anjou MBCDP2.1	0.30	0.60	—	0.10	—	—	0.20	0.40	—	—	—	0.20	—	—	—	0.20	—
Texas Longhorn MBCDP2.1	0.88	—	—	0.13	—	—	0.75	—	—	—	—	—	0.25	—	—	—	—
Santa Gertrudis MBCDP2.1	0.50	0.50	—	—	—	—	0.25	0.25	—	—	—	0.50	—	—	—	—	—
Chianina MBCDP2.1	0.38	0.25	—	0.38	—	—	0.25	0.25	—	0.25	—	—	0.25	—	—	—	—
Holstein MBCDP2.1	0.13	0.88	—	—	—	—	—	0.75	—	—	—	0.25	—	—	—	—	—
MARC reference parents	0.71	0.23	—	0.05	—	—	0.46	—	—	—	—	0.43	0.07	—	—	0.04	—
Bison	—	—	—	—	1.00	—	—	—	—	—	1.00	—	—	—	—	—	—
MBCDP2.1 Total	0.66	0.26	0.02	0.06	—	0.01	0.48	0.11	0.01	0.01	—	0.27	0.07	0.01	0.01	0.01	0.01

[a] Within each animal group, the allele and genotype frequencies are presented as the fraction of haploid and diploid genomes, respectively. A frequency of zero is denoted with a dash.

[b] Genotype results for 26 founders of the MARC reference population and four bison are presented for comparison.

[c] An *IL8* haploytype allele is defined as the concatenated set of *IL8* SNP alleles present on a chromosome. Haplotype definitions: 1, GAGATTG; 2, ACGGCTA; 3, GAGGCCG; 4, GAGGCTA; 5, GAGGCTG; 6, GAAGCCG. Allele or genotype frequencies that sum to 0.99 or 1.01 are the result of rounding errors.

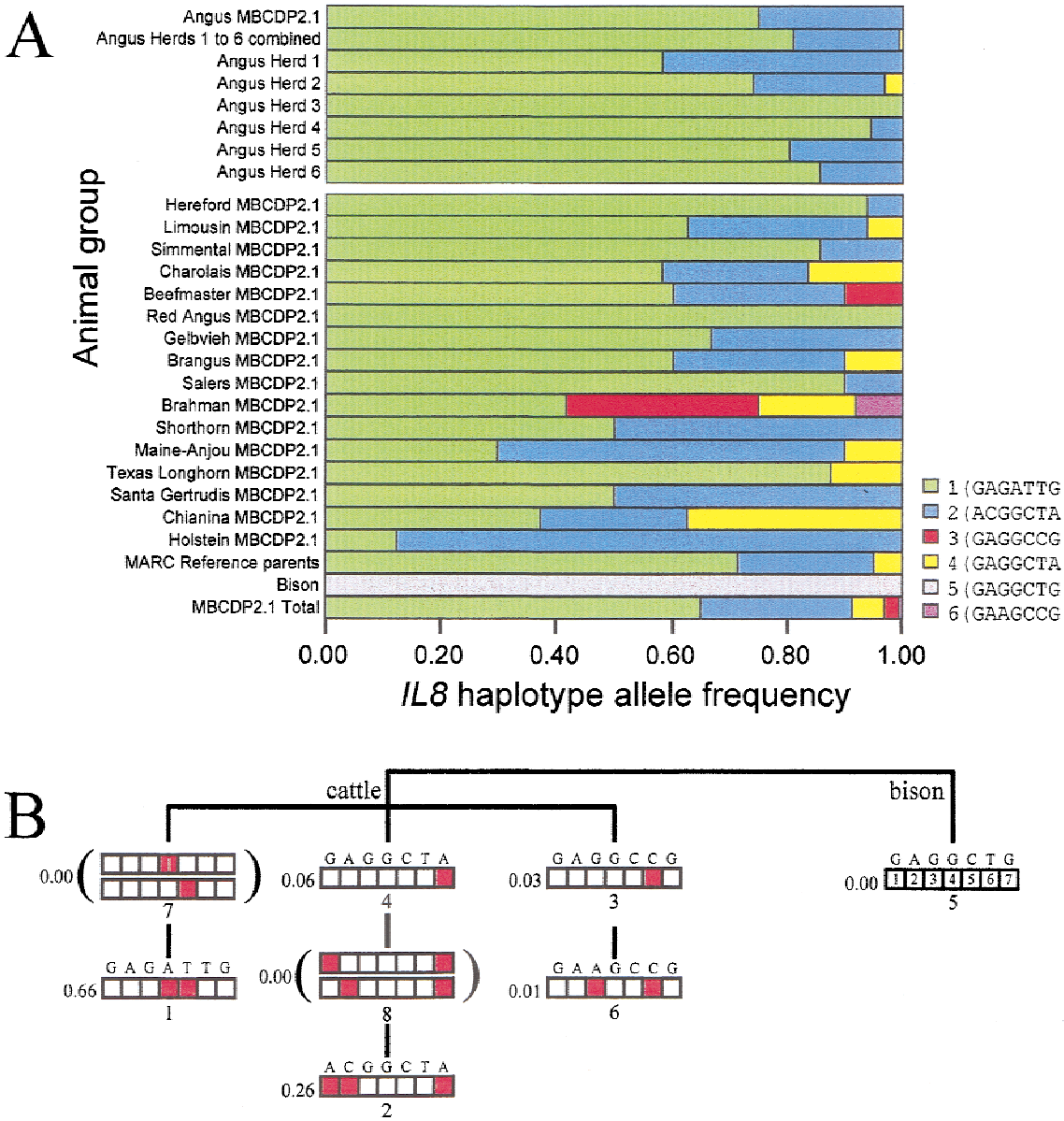

Fig. 2. *IL8* haplotype allele frequencies and haplotype lineages in cattle from commercial populations of U.S. cattle. Panel **A:** The colored stacked bars represent different haplotypes for the *IL8* locus. The bar length corresponds to their respective frequencies in the animal groups. The haplotype allele number corresponds to that presented in Table 4. Panel **B:** Each haplotype is presented as a concatenated set of biallelic SNPs (AH2-1 through AH2-7). The respective SNP alleles are represented as red or white squares, depending on their occurrence in the reference species (bison, white). The lines connecting the haplotypes indicate the proposed (unrooted) relationship between haplotype alleles. Haplotypes that are predicted to have occurred but have not yet been observed are shown with parentheses around them. They are presented as an ambiguous pair of which one, but not both, is predicted. The number beneath the squares corresponds to the haplotype allele numbers presented in Table 4, whereas the number to the left is the DNA segment haplotype allele frequency in MBCDP2.1.

mutation, since recombination has not yet eroded this apparent SNP haplotype.

IL8 *haplotype structures.* The set of SNP alleles present on a DNA segment of a single chromosome comprises a haplotype allele. For small (amplicon-sized) regions of the chromosome, the number of haplotypes (H) predicted to arise from a particular set of sequential mutations (S) follows the simple relationship: $H = S + 1$ (Heaton et al. 2001). This assumes that reversible mutations and recombination events within the DNA segment are rare. Thus, seven SNPs for *IL8* provide the opportunity to identify eight *IL8* haplotype structures in commercial populations.

Segregation analysis of the MARC reference population revealed three SNP haplotype alleles (Table 4; alleles 1, 2, and 4) with consistent Mendelian inheritance patterns. A fourth haplotype (Table 4; allele 3) was identified from a Brahman sire in MBCDP2.1 because it was homozygous for all seven SNPs. A fifth haplotype (allele 6) was predicted to occur in a heterozygous Brahman sire with allele 3. The prediction was confirmed by cloning and sequencing a 1268-bp PCR product that spanned all seven SNPs (data not shown). Combinations of these five *IL8* DNA segment haplotype alleles accounted for 100% of the genotypic variation observed in the MBCDP2.1 sires and the six purebred Angus herds (Fig 2A). The breed-based comparison of *IL-8* haplotypes illustrates two features. First, two SNP haplotypes pre-

dominate in U.S. beef cattle. The most common haplotype (1, green) was present in each cattle breed group, whereas the next most common haplotype (2, blue) was present in all cattle breed groups except Brahman and Red Angus. The latter haplotype may occur at some frequency in purebred Brahman and Red Angus cattle but was not detected in those individuals from MBCDP2.1. The second point illustrated in Fig. 2A is that two haplotypes were detected only in breeds influenced by germplasm from *Bos indicus*. For example, the number 6 haplotype (purple) was detected only in Brahman, and the number 3 haplotype (red) was detected only in Brahman and Beefmaster (a composite of Brahman, Hereford, and Shorthorn breeds). The latter results indicate that DNA segment haplotypes may be useful for identifying the ancestry of different genomic regions.

Genomic DNA from American bison, a related bovid, was amplified and sequenced to help determine simple phylogenetic relationships between the *IL8* haplotype structures. Four bison DNAs were tested, and all produced high quality sequence information for both *IL8* amplicons. All four bison contained the same homozygous genotypes for all seven SNP sites, and the combination of these SNP alleles revealed a novel bovid haplotype structure (Fig. 2A, allele 5). In addition to the unique haplotype, all four bison individuals contained an apparent inter-species difference at nt 282 of *IL8*DS5 ("A" in cattle; "G" in bison). A model showing the relationships between haplotypes is shown in Fig. 2B. This simplified model does not predict the relative age of these haplotypes, but simply identifies potential haplotype lineages and predicts structures for haplotypes that have not yet been observed. Knowledge of these predicted structures broadens the scope of automated haplotype assignment in populations with undefined pedigrees.

Discussion

Knowledge of DNA segment haplotype variation in candidate gene regions facilitates genotype/phenotype correlation studies. The present results show the *IL8* haplotype structure for a panel of cattle representing U.S. commercial populations. The seven SNPs identified from this panel are estimated to represent more than 98% of the SNP alleles present in this region for this group. The five haplotype alleles present in sires from this panel represent the majority of the eight possible nonrecombinant alleles. Three remaining haplotypes are predicted to be either rare (frequency less than 0.02) or extinct in the U.S. beef cattle herd. Recombinant *IL-8* haplotypes were not apparent in the group and are predicted to be rare in U.S. cattle as a group.

The allele and genotype frequencies presented in Tables 3 and 4 provide breed-based estimates of informativity for individual SNPs and haplotypes. For example, SNP AH2-4 is predicted to be rather informative in the Santa Gertrudis breed as a whole, based on the 1:2:1 genotype frequency ratios observed for that group in MBCDP2.1. Conversely, the same SNP marker is predicted to be infrequently polymorphic in the Red Angus breed. These estimates of informativity may be particularly useful in designing tests to confirm parentage, select sires with enhanced heterosis, conserve germplasm, and identify quantitative trait loci.

An accurate and automated platform for scoring genotypes is a requirement for making widespread use of SNP markers in cattle. The MALDI-TOF MS assays designed for the seven *IL8* SNPs were robust in >4000 genotypes scored and accurate in 31 families and 262 progeny tested. The results presented here indicate that the SNP assays will give accurate results for herds that are similar to the MARC reference population. However, this does not exclude the possibility that commercial populations may contain some alleles that will not amplify because cryptic polymorphisms occur in the genomic region where the amplification primer binds. This underscores the importance of performing preliminary segregation analysis in untested herds to detect novel allele amplification problems.

Past reports (Grosse et al. 1999; Heaton et al. 1999, 2000) have demonstrated that DNA sequencing is an efficient method to identify SNPs in cattle. An advantage of a selective gene-based approach for SNP and haplotype discovery is the flexibility to test specific gene/phenotype associations in cattle as they become known in other species. For example, detailed analysis of candidate genes with known gene/phenotype associations in humans and mice was critical in identifying β2 integrin mutations as the cause of leukocyte adhesion deficiency in Holstein cattle (Shuster et al. 1992) and in identifying myostatin mutations as the cause of double muscling beef cattle (Grobet et al. 1997; Kambadur et al. 1997; McPherron and Lee 1997). These and other examples (Collins et al. 1999) support the assertion that high-resolution tests within a candidate gene region are indispensable for genetic epidemiology.

Acknowledgments. We thank Drs. J.M. Fox, S.M. Kappes, D.P. Little, and T.P.L. Smith for critical comments regarding this manuscript; J.K. Carnahan, L.S. Flathman, R.T. Lee, and S.J. Simcox for expert technical assistance; S.L. Kluver and J.K. Rosch for secretarial assistance. We thank Dr. L.V. Cundiff for advice in designing the MARC beef cattle diversity panel; Dr. D. Coover (SEK Genetics, Inc., Galesburg, Kans.), D. Frank (ABS Global, Inc., DeForest, Wis.), B. House (Select Sires, Inc., Plain City, Ohio), and C. Rugg (Bovine Elite, Inc., College Station, Tex.) for assistance in sire selection; and Dr. R. Bohlender for providing blood samples from commercial Angus herds. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

References

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL et al. (1994) A genetic linkage map for cattle. Genetics 136, 619–639

Braun A, Little DP, Koster H (1997) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 43, 1151–1158

Buzby JC, Roberts T, Jordan Lin C-T, MacDonald JM (1996) Bacterial foodborne disease: medical costs and productivity losses (Washington, D.C.: Food and Consumer Economics Division, Economic Research Service, U.S. Department of Agriculture)

Chakravarti A (1999) Population genetics—making sense out of sequence. Nat Genet 21, 56–60

Collins A, Morton NE (1998) Mapping a disease locus by allelic association. Proc Natl Acad Sci USA 95, 1741–1745

Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278, 1580–1581

Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96, 15173–15177

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8, 186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175–185

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8, 195–202

Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, et al. (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat Genet 17, 71–74

Grosse WM, Kappes SM, Laegreid WW, Keele JW, Chitko-McKown CG, et al. (1999) Single nucleotide polymorphism (SNP) discovery and linkage mapping of bovine cytokine genes. Mamm Genome 10, 1062–1069

Heaton MP, Laegreid WW, Beattie CW, Smith TPL, Kappes SM (1999) Identification and genetic mapping of bovine chemokine genes expressed in epithelial cells. Mamm Genome 10, 128–133

Heaton MP, Grosse WM, Kappes SM, Keele JW, Chitko-McKown CG et

al. (2001) Estimation of DNA sequence diversity in bovine cytokine genes. Mamm Genome, in press
Kambadur R, Sharma M, Smith TP, Bass JJ (1997) Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. Genome Res 7, 910–916
Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22, 139–144
Lander ES (1996) The new genomics: global views of biology. Science 274, 536–539
Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265, 2037–2048
Little DP, Braun A, Darnhofer-Demar B, Frilling A, Li Y et al. (1997a) Detection of RET proto-oncogene codon 634 mutations using mass spectrometry. J Mol Med 75, 745–750
Little DP, Braun A, Darnhofer-Demar B, Koster H (1997b) Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry. Eur J Clin Chem Clin Biochem 35, 545–548
Little DP, Cornish TJ, O'Donnell MJ, Braun A, Cotter RJ, et al. (1997c) MALDI on a chip: analysis of arrays of low-femtomole to subfemtomole quantities of synthetic oligonucleotides and DNA diagnostic products dispensed by a piezoelectric pipette. Anal Chem 69, 4540–4546
McPherron AC, Lee SJ (1997) Double muscling in cattle due to mutations in the myostatin gene. Proc Natl Acad Sci USA 94, 12457–12461
Morsey MA, Popowych Y, Kowalski J, Gerlach G, Godson D, et al. (1996) Molecular cloning and expression of bovine interleukin-8. Microb Pathog 20, 203–212
Mukaida N, Shiroo M, Matsushima K (1989) Genomic structure of the human monocyte-derived neutrophil chemotactic factor IL-8. J Immunol 143, 1366–1371
Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25, 2745–2751
NPLC Report (1999) National Pedigreed Livestock Council Annual Report, Brattleboro, Vt.
Shuster DE, Kehrli ME Jr., Ackermann MR, Gilbert RO (1992) Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. Proc Natl Acad Sci USA 89, 9225–9229
Terwilliger JD, Goring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 72, 63–132
Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23, 397–404
Zlotnik A, Morales J, Hedrick JA (1999) Recent advances in chemokines and chemokine receptors. Crit Rev Immunol 19, 1–47
Zollner S, von Haeseler A (2000) A coalescent approach to study linkage disequilibrium between single-nucleotide polymorphisms. Am J Hum Genet 66, 615–628