2006

Linkage disequilibrium across six prion gene regions spanning 20 kbp in U.S. sheep

Benedict T. Green
USDA-ARS, Ben.Green@ars.usda.gov

Michael P. Heaton
USDA-ARS

Michael L. Clawson
USDA-ARS

William W. Laegreid
USDA-ARS, laegreid@email.marc.usda.gov

Incorporating Mouse Genome
Mammalian Genome
Genes and Phenotypes

Linkage disequilibrium across six prion gene regions spanning 20 kbp in U.S. sheep

Benedict T. Green, Michael P. Heaton, Michael L. Clawson, William W. Laegreid

U.S. Department of Agriculture, Agricultural Research Service, U.S. Meat Animal Research Center (USMARC), State Spur 18D, P.O. Box 166, Clay Center, Nebraska 68933, USA

Received: 5 April 2006 / Accepted: 6 July 2006

Abstract

Single nucleotide polymorphisms (SNPs) and haplotype alleles within the prion gene (*PRNP*) coding sequence of domestic sheep (*Ovis aries*) are associated with genetic predisposition to scrapie, a transmissible spongiform encephalopathy disease of sheep. This report describes regions of linkage disequilibrium (LD) throughout the *PRNP* gene region in U.S. sheep and provides a genetic framework for identifying additional *PRNP* determinants associated with scrapie resistance. Four sequence tagged sites (i.e., STS or amplicons) totaling 3869 bp and spanning 20 kbp of genomic *PRNP* sequence were sequenced in a diverse panel of 90 sires representing ten popular U.S. breeds of sheep. Analysis of these sequences identified 36 previously unreported polymorphisms. In combination with two previously characterized STS, 62 polymorphisms were analyzed in a 20-kbp *PRNP* region in this panel of U.S. sheep. Two regions of strong LD and ten common haplotypes were identified. The haplotype encoding amino acid residues A, R, and Q at codons 136, 154, and 171, respectively, was observed on nine larger haplotypes spanning *PRNP* from the promoter region to the 3′ untranslated region. The haplotype encoding VRQ was observed on two larger haplotypes, whereas ARR, ARH, and AHQ were each present on a single haplotype. The existence of multiple haplotypes encoding ARQ raises the question of whether sheep bearing these different haplotypes are equally susceptible to scrapie. The haplotype structure within the 20-kbp region of *PRNP* identified in this study is important for higher-resolution analysis of genetics contributions to scrapie susceptibility.

Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession number DQ077504.

Correspondence to: W.W. Laegreid; E-mail: laegreid@email.marc.usda.gov

Introduction

Transmissible spongiform encephalopathies (TSEs) are a group of fatal neurologic disorders characterized by the accumulation of host-encoded proteinaceous particles also known as prions (PrP). TSE disorders are associated with a conformational change in PrP to a highly stable, protease-resistant isoform (Prusiner 1982, 1998). Although TSE disorders occur in a variety of mammalian species (for review, see Prusiner 1998, 2004), scrapie is the naturally occurring TSE disorder of sheep and is an important problem where it occurs.

In sheep, specific single nucleotide polymorphisms (SNPs) within the protein-coding sequence of the prion gene (*PRNP*), and the resulting amino acid changes, have been associated with scrapie susceptibility and disease progression (for review, see Baylis and Goldmann 2004). The three most studied codons associated with scrapie susceptibility in sheep are those encoding amino acid residues at positions 136, 154, and 171 in PrP. Although there are a number of codon variants at these positions, the most common residues include valine (V) or alanine (A) at position 136, arginine (R) or histidine (H) at position 154, and glutamine (Q), arginine (R), or histidine (H) at position 171. The most common variants of codons 136, 154, and 171 define five *PRNP* codon haplotypes: ARQ, ARR, ARH, AHQ, and VRQ. These five codon haplotypes result in 15 genotype combinations, each with an associated degree of risk. For example, animals homozygous for the VRQ codon haplotype are generally considered highly susceptible and, when infected, may rapidly progress to clinical signs of scrapie. Conversely, animals homozygous for the

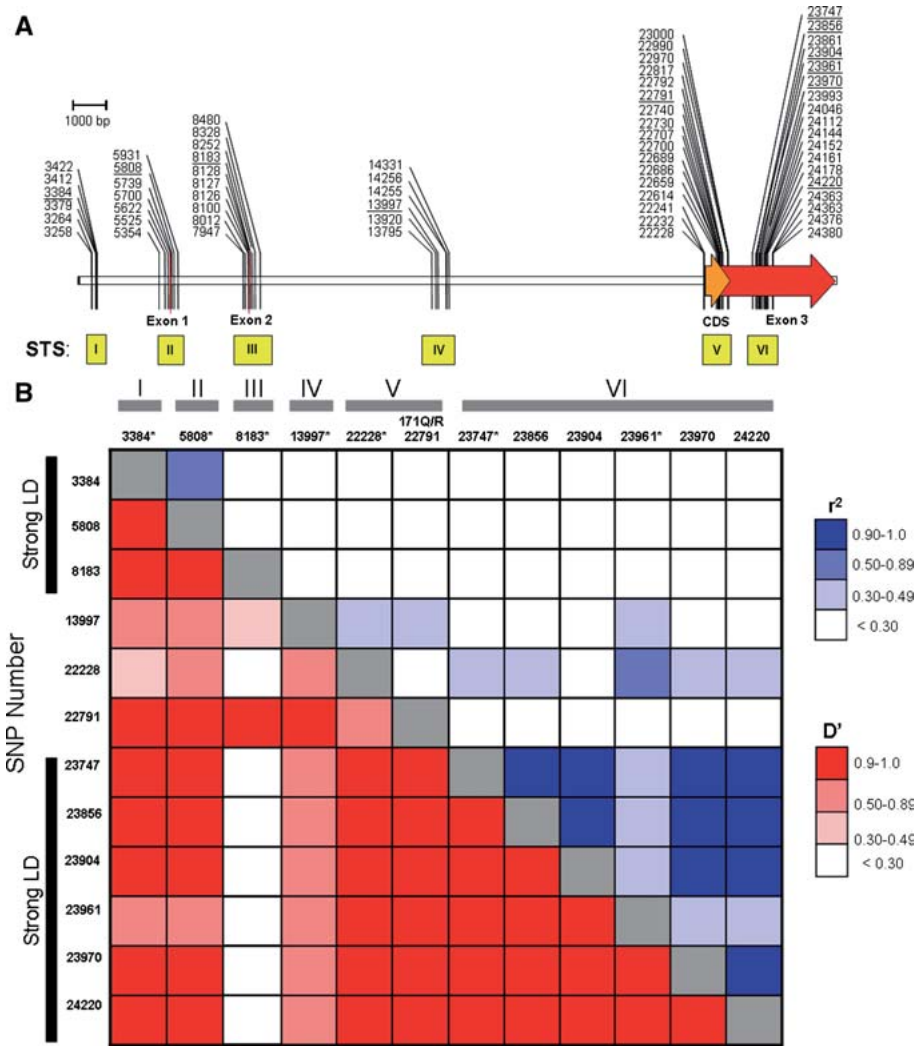

Fig. 1. Physical and LD map of *PRNP*. **(A)** *PRNP* physical map with polymorphism numbers corresponding to their location in GenBank accession number DQ077504 (tick marks). STS are represented as yellow blocks, the protein-coding region in orange, and exons in red. STS I represents GenBank accession number AY326428; STS II-IV GenBank accession number DQ077504; STS V GenBank accession number AY326330; and STS VI GenBank accession number DQ077504. Underlined polymorphism numbers were used in the inference of *D'* and r^2 in panel B. Polymorphisms 22686, 22740, and 22791 are those previously associated with the *PRNP* codon changes A136V, R154H, and Q171R, respectively. **(B)** A pairwise r^2 and *D'* plot for 12 polymorphism alleles with a minor allele frequency of 0.10 or greater in the MSDP1.1 panel. Each square in the bottom left half of the plot corresponds to the *D'* (red) and the upper right half of the plot corresponds to the r^2 (blue) between a pair of polymorphisms. *D'* values from 0.9 to 1.0 were defined as strong LD, values from 0.50 to 0.89 were defined as intermediate LD, and values less than 0.50 defined as weak LD. The numbers on the vertical and horizontal axes of the plot represent the corresponding polymorphisms from the physical map above that were used in the determination of *D'* and are listed in Table 2. Regions of strong LD are indicated by the solid line on the vertical axis. Gray bars on the horizontal axis represent the STS sequence from panel A. SNP 22791 is the Q171R polymorphism associated with scrapie progression. Asterisks denote htSNPs.

ARR codon haplotype are considered scrapie resistant (Belt et al. 1995; Clouscard et al. 1995; Hunter et al. 1997a; O'Rourke 1997; Baylis et al. 2002). The association of specific *PRNP* codon haplotypes with variation in scrapie susceptibility is the basis for genotype-based scrapie eradication programs around the world (Houston et al. 2003; Baylis and Goldmann 2004).

Although specific *PRNP* genotypes like ARQ/ARQ are associated with variation in scrapie susceptibility, many susceptible animals with this genotype do not develop clinical signs of the disease following exposure. For example, in closed flocks of Suffolk or Romanov sheep naturally exposed to scrapie, more than 50% of genetically susceptible animals were without clinical signs of disease

Table 1. Oligonucleotides used for the amplification and sequencing of *PRNP*

MARC primer number	*STS number*	*Position in DQ077504*	*Oligonucleotide sequence*[a]	*Orientation*	*Function*[b]	*PCR annealing temp. (°C)*	*Amplicon length (bp)*
18015	1	2976–2996	ctc acc att tca gaa tac ctc	Sense	A/S	58	593
8702	1	3552–3568	gca aat ggc cct aca tc	Antisense	A/S	58	593
40469	2	5326–5349	aaa taa tga gca gga act ga	Sense	A/S	55	786
40475	2	6098–6117	ttg ttc ttc tga tct ccc ca	Antisense	A/S	55	786
40472	2	5766–5786	cct gcg gcc gag gcc cta gg	Sense	S	–[c]	–
40471	2	5480–5499	aac taa cac cgc ctc cct gg	Antisense	S	–	–
40476	3	7702–7721	ccc aca ggg ctc aga atg at	Sense	A/S	58	1041
40481	3	8832–8851	ccc ttg acg tga tca cac aa	Antisense	A/S	58	1041
40477	3	7972–7991	agt gtc atg ttt gct gct tt	Sense	S	–	–
40479	3	8311–8330	ttg tga tcc cag act ttc ag	Antisense	S	–	–
40480	3	8562–8586	act aca gct tag aaa aac tgt tac c	Antisense	S	–	–
40482	4	13523–13542	gta att tga ttt ggg aag tc	Sense	A/S	50	831
40487	4	14498–14517	taa aca aag ggg aaa caa ag	Antisense	A/S	50	831
40483	4	13756–13784	ttg ttt caa act tag tcc tg	Sense	S	–	–
40486	4	14378–14402	aaa gta att tag aag ttg ggt tgg c	Antisense	S	–	–
44928	5	22207–22227	tgc tgc aga ctt taa gtg att	Sense	A/S	58	893
18014	5	23084–23099	ccc caa cct ggc aaa g	Antisense	A/S	58	893
40488	6	23615–23639	ttc caa cat atg gaa gag gtg ccc t	Sense	A/S	50	931
41452	6	24532–24550	gct agg tga caa tat tga aca	Antisense	A/S	50	931
40490	6	23896–23914	ctt gga ctg aaa gga gat ca	Sense	S	–	–
41451	6	24256–24277	agt cca aca cta gca ctg gtt c	Antisense	S	–	–

[a]Oligonucleotide sequence presented in the 5′ to 3′ direction.
[b]A = Amplification; S = Sequencing.
[c]Not applicable.

(Hunter et al. 1997b; Elsen et al. 1999). One plausible hypothesis is that the most common susceptible codon haplotype (i.e., ARQ) is present on the background of multiple haplotypes in the *PRNP* gene region which vary in their association with disease.

The present report identifies haplotypes spanning 20 kbp of ovine *PRNP* and defines linkage disequilibrium (LD) between alleles in U.S. sheep. Characterizing these haplotypes and patterns of LD may facilitate a more comprehensive evaluation of host genetic influences on scrapie susceptibility.

Materials and methods

Animals. DNA from the U.S. Meat Animal Research Center (USMARC) Sheep Diversity Panel version 1.1 (MSDP1.1), consisting of samples from 90 rams representing nine popular breeds of U.S. sheep, was sequenced and genotyped (Freking et al. 2002). The breeds included USMARC Composite III, Dorper, Dorset, Finsheep, Katahdin, Suffolk, Texel, Rambouillet, and Romanov (Leymaster 1991). The sires within breeds were selected for minimal relationships within and between pedigrees. Fourteen families in which one or both parents carried the VRQ codon haplotype were used to verify haplotype segregation.

Primers, PCR conditions, DNA sequencing, and analysis. In this study, six sequence tagged sites (STS, i.e., amplicons) spanning the ovine *PRNP* gene were analyzed: four reported here and two reported previously (Heaton et al. 2003). The amplicons included all or parts of the promoter region, exon 1, intron 1, exon 2, intron 2, the CDS, and 3′ untranslated region (UTR) (Fig. 1A). Specific oligonucleotide primer sequences for amplification and sequencing of the *PRNP* regions were designed from GenBank accession number U67922 (Lee et al. 1998) using Vector NTI 9.0.0 (Invitrogen Corporation, Carlsbad, CA) (Table 1). Genomic DNA was used as the template for amplification, and PCR products were directly sequenced as previously described (Clawson 2004). Sequences were aligned, analyzed, and visually inspected using Phred, Phrap, Polyphred, and Consed software (Ewing and Green 1998; Ewing et al. 1998; Nickerson et al. 1997). To identify potential polymorphisms in the amplification primer binding sites, a second set of amplification primers (flanking each original amplicon) was used for sequencing the original primer binding sites, thus, minimizing allelic dropout.

Haplotype prediction and LD estimation. Unphased genotype data for the selected diallelic loci that were consistent with Hardy-Weinberg equilibrium (HWE) expectations (chi squared, $p >$

Table 2. Allele and genotype frequencies of *PRNP* polmorphisms in MSDP1.1

GenBank accession number	*Position*[b]	*Region*	*Alleles (1,2)*[c]	*Frequency*[a]				
				Allele		*Genotype*		
				1	*2*	*1,1*	*1,2*	*2,2*
AY326428[d]	3258	Promoter	atgaa-(C,T)-gaat	0.99	0.01	0.99	0.01	0.00
AY326428	3264	Promoter	gaaat-(A,T)-tttct	0.64	0.36	0.49	0.30	0.21
AY326428	3379	Promoter	tgtgg-(G,A)-agatR	0.96	0.04	0.91	0.09	0.00
AY326428	3384	Promoter	Ragat-(G,A)-agaag	0.88	0.12	0.79	0.19	0.02
AY326428	3412	Promoter	gtctt-(C,T)-atggt	0.96	0.04	0.91	0.09	0.00
AY326428	3422	Promoter	tgcca-(T,C)-aacct	0.96	0.04	0.91	0.09	0.00
DQ077504[e]	5354	Exon 1	gttcc-(C,A)-gaaat	0.84	0.16	0.79	0.11	0.10
DQ077504	5525	Exon 1	tctca-(A,C)-ctcgt	0.99	0.01	0.98	0.02	0.00
DQ077504	5622	Exon 1	tcccc-(G,C)-ccccc	0.96	0.04	0.94	0.03	0.02
DQ077504	5700	Exon 1	agcgt-(C,T)-ttctc	0.98	0.02	0.97	0.02	0.01
DQ077504	5739	Exon 1	tcctt-(T,G)-aaacc	0.91	0.09	0.84	0.13	0.02
DQ077504	5808	Exon 1	actgg-(C,G)-tggga	0.89	0.11	0.81	0.17	0.02
DQ077504	5931	Exon 1	ggtta-(G,C)-gagag	0.99	0.01	0.98	0.02	0.00
DQ077504	7947	Exon 2	cacag-(T,C)-tttct	0.91	0.09	0.83	0.16	0.01
DQ077504	8012	Exon 2	atctc-(6T,4T)-aaaaa	0.96	0.94	0.96	0.00	0.04
DQ077504	8100	Exon 2	aatcc-(A,G)-ttctt	0.93	0.07	0.89	0.08	0.03
DQ077504	8126	Exon 2	agttt-(T,G)-Qaagg	0.93	0.07	0.90	0.07	0.03
DQ077504	8127	Exon 2	gtttKK(7T,9T,TC9T)-aagga	N.A.[f]				
DQ077504	8183	Exon 2	caagc-(T,C)-gaagc	0.76	0.24	0.62	0.28	0.10
DQ077504	8252	Exon 2	cctgc-(A,G)-gaatc	0.57	0.43	0.40	0.33	0.27
DQ077504	8328	Exon 2	gatca-(C,G)-aaatc	0.83	0.17	0.73	0.19	0.08
DQ077504	8480	Exon 2	aaaaa-(T,C)-agtat	0.93	0.07	0.90	0.07	0.03
DQ077504	13920	Intron 2	agttc-(A,G)-ctcaa	0.99	0.01	0.99	0.01	0.00
DQ077504	13975	Intron 2	gtaca-(A,G)-aacta	0.91	0.09	0.84	0.13	0.02
DQ077504	13997	Intron 2	gacaa-(G,A)ttgta	0.52	0.48	0.32	0.39	0.29
DQ077504	14255	Intron 2	aaaaa-(C,T)-Yatca	0.99	0.01	0.99	0.01	0.00
DQ077504	14256	Intron 2	aaaaY-(C,T)-atcag	0.91	0.09	0.83	0.14	0.02
DQ077504	14331	Intron 2	ctaga-(T,C)-agcta	0.97	0.03	0.96	0.03	0.01
AY326330[d]	22228	Intron 2	tgatt-(C,T)-ttaYg	0.61	0.39	0.41	0.40	0.19
AY326330	22232	Intron 2	tYtta-(C,T)-gtggg	0.98	0.02	0.97	0.03	0.00
AY326330	22241	Intron 2	ggcat-(A,T)-tgatg	0.94	0.06	0.87	0.13	0.00
AY326330	22614	Exon 3	caaca-(T,C)-gaagc	0.95	0.05	0.91	0.09	0.00
AY326330	22659	Exon 3	agggg-(G,C)-ccttg	0.99	0.01	0.99	0.01	0.00
AY326330	22684	Exon 3	aagtg-(C,T)-caYga	0.97	0.03	0.94	0.06	0.00
AY326330	22686	Exon 3	tgYca-(T,C)-gaRca	0.97	0.03	0.95	0.04	0.01
AY326330	22700	Exon 3	ggcct-(C,T)-ttata	0.99	0.01	0.99	0.01	0.00
AY326330	22707	Exon 3	tatac-(A,G)-ttttg	0.93	0.07	0.89	0.09	0.02
AY326330	22730	Exon 3	aggac-(C,T)-gttac	0.98	0.02	0.96	0.02	0.01
AY326330	22740	Exon 3	ctatc-(G,A)-tgaaa	0.97	0.03	0.94	0.04	0.01
AY326330	22791	Exon 3	ggatM-(A,G)-Ktata	0.66	0.34	0.47	0.39	0.14
AY326330	22792	Exon 3	gatMR-(G,T)tatag	0.97	0.03	0.94	0.04	0.01
AY326330	22817	Exon 3	ttgtg-(C,T)-atgac	0.99	0.01	0.99	0.01	0.00
AY326330	22970	Exon 3	accaa-(A,C)-ggggg	0.90	0.10	0.82	0.17	0.01
AY326330	22990	Exon 3	atcct-(C,G)-ttttc	0.90	0.10	0.82	0.17	0.01
AY326330	23000	Exon 3	cttcc-(C,T)-ctcct	0.99	0.01	0.99	0.01	0.00
DQ077504	23747	Exon 3	cttct-(A,G)-gacac	0.82	0.18	0.70	0.24	0.06
DQ077504	23856	Exon 3	tgatg-(C,T)-ttttR	0.83	0.17	0.71	0.23	0.06
DQ077504	23861	Exon 3	Ytttt-(G,A)-aacta	0.91	0.09	0.84	0.13	0.02
DQ077504	23904	Exon 3	gactg-(A,C)-aagga	0.83	0.17	0.72	0.22	0.06
DQ077504	23961	Exon 3	ctcca-(G,A)-tactt	0.63	0.37	0.44	0.38	0.18
DQ077504	23970	Exon 3	tttgg-(T,C)-cacct	0.83	0.17	0.71	0.23	0.06
DQ077504	23993	Exon 3	tgaag-(G,A)-cagga	0.93	0.07	0.87	0.13	0.00
DQ077504	24112	Exon 3	ttggc-(G,A)-atgga	0.96	0.04	0.91	0.09	0.00
DQ077504	24144	Exon 3	agtcc-(A,G)-tggtg	0.91	0.09	0.84	0.13	0.02
DQ077504	24152	Exon 3	gtgtc-(G,T)-cagag	0.93	0.07	0.86	0.14	0.00
DQ077504	24161	Exon 3	agtcg-(G,A)-acacg	0.93	0.07	0.86	0.14	0.00
DQ077504	24178	Exon 3	tgact-(A,G)-aattg	0.93	0.07	0.86	0.14	0.00
DQ077504	24220	Exon 3	agata-(C,T)-aaaaa	0.83	0.17	0.71	0.23	0.06

(Continued)

Table 2. Continued

GenBank accession number	Position[b]	Region	Alleles (1,2)[c]	Frequency[a]				
				Allele		Genotype		
				1	2	1,1	1,2	2,2
DQ077504	24363	Exon 3	tctta-(AA,A)ttgtc	0.98	0.02	0.98	0.00	0.02
DQ077504	24376	Exon 3	aaaaa-(C,G)-aaagt	0.88	0.12	0.81	0.13	0.06
DQ077504	24380	Exon 3	aSaaa-(A,G)-ttagg	0.95	0.05	0.92	0.06	0.02

[a]Allele or genotype frequencies that sum to 0.99 are the result of rounding.
[b]Nucleotide position in GenBank accession number DQ077504.
[c]Nucleotide codes: K, G/T; M, A/C; Q, Insertion/Deletion; R, A/G; Y, C/T.
[d]Heaton et al. (2003).
[e]This report.
[f]Not applicable, at least three alleles were present, the genotypes of some individuals were not resolvable.

0.01) were used to infer the *PRNP* haplotypes. Haploview 3.0, which uses an expectation-maximization (EM) algorithm, was used to infer haplotypes and to identify a minimal set of polymorphisms that define all observed haplotypes (haplotype-tagging SNPs, htSNPs) (Barrett et al. 2005). The htSNPs were validated in the MSDP1.1 panel using matrix-assisted, laser desorption-ionization, time-of-flight mass spectroscopy (MALDI-TOF MS)-based assays to confirm the haplotypes inferred from DNA sequencing. Phase version 2.0, which uses a Bayesian algorithm, was also used to infer haplotypes using default settings (Stephens et al. 2001; Stephens and Donnelly 2003).

Because of the difficulty inferring the phase of haplotypes, which include SNPs with low-frequency minor alleles, a minor allele frequency cutoff of 0.10 was used in this study (Fallin and Schork 2000; Jeffreys et al. 2001; Stumpf 2004). Deviations from HWE and a low minor allele frequency may bias the prediction and frequency estimation of haplotypes in a population (Fallin and Schork 2000). Exclusion of such polymorphisms helps mitigate the effects of genotyping error and unrecognized admixture in sample populations.

PRNP haplotype relationships were analyzed with a median-joining network algorithm (Network 4.1.1.0; Bandelt et al. 1999). Only haplotypes with a frequency of greater than 0.02 in MSDP1.1 were included in this analysis. STS data from additional animals homozygous for AHQ, ARH, and VRQ codon haplotypes were used to establish the relationship of these codon haplotypes to the median-joining network because they were present at a low frequency in MSDP1.1 (GenBank accession numbers AY909542, AY907683, and AY907685).

Results

Thirty-three previously unreported SNPs and three insertion/deletions were identified in 3869 bp of sequence from the four STS. In combination with two previously reported STS, 62 polymorphisms were analyzed in the *PRNP* region in this panel of U.S. sheep. On average, there was one polymorphic site identified for every 82 bp sequenced across 5075 bp in this diverse set of 90 U.S. sheep. There was no evidence of allelic dropout in the data set used for SNP scoring and haplotype inference. Of the 62 polymorphisms, 12 met the criteria for inclusion in haplotype predictions, i.e., a minor allele frequency of 0.10 or greater, diallelic, and genotype distribution similar to HWE expectations (Table 2). The only polymorphism affecting the amino acid sequence of PrP and which met the criteria for inclusion in haplotype predictions was 22791 (Q171R).

Pairwise D' and r^2 values for the 12 selected polymorphisms were used to assess LD in the multibreed sheep panel. The two regions of strong LD (D' values from 0.9 to 1.0) were identified in *PRNP* (Fig. 1B). The first region included part of the promoter region and all of exons 1 and 2. The second region contained the entire *PRNP* coding region and a portion of the 3′ UTR. This second region had a

Table 3. *PRNP* haplotypes in the multibreed sheep panel

	Haplotype[a]
1[b]	GCTGC**G**ACAGTC
2	GCTATAACAATC
3	GCCACAACAGTC
4	GCTATAGTCACT
5	GCCATAGTCACT
6	AGTGCAACAGTC
7	GCCGCAACAGTC
8	GCTACAACAGTC
9	GCCATAACAATC
10	ACTACAACAGTC

[a]Other haplotypes identified at frequency of less than 0.02: 11, GCTGCAACAGTC; 12, GCCGTAGTCACT.
[b]Haplotype 1 is defined by the presence of the minor allele of SNP 22791 (bold, underlined) corresponding to the 171R allele associated with scrapie resistance.

Table 4. *PRNP* haplotype frequencies in the multibreed sheep panel

	Haplotype[a]										
Breed (n)[b]	*1*[c]	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Other*[d]
Composite (20)	0.47	0.16	0.00	0.11	0.05	0.00	0.21	0.00	0.00	0.00	0.00
Dorper (15)	0.25	0.06	0.38	0.06	0.00	0.00	0.00	0.00	0.25	0.00	0.00
Dorset (19)	0.68	0.05	0.05	0.00	0.00	0.11	0.00	0.05	0.00	0.00	0.05
Finnsheep (20)	0.10	0.00	0.20	0.25	0.30	0.05	0.05	0.00	0.00	0.00	0.05
Katahdin (18)	0.50	0.00	0.11	0.00	0.17	0.22	0.00	0.00	0.00	0.00	0.00
Rambouillet (19)	0.37	0.05	0.00	0.00	0.00	0.21	0.21	0.16	0.00	0.00	0.00
Romanov (20)	0.05	0.30	0.25	0.05	0.15	0.05	0.00	0.00	0.00	0.15	0.00
Suffolk (19)	0.32	0.37	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Texel (20)	0.50	0.35	0.00	0.05	0.00	0.00	0.00	0.10	0.00	0.00	0.00
Total (170)	0.36	0.15	0.11	0.09	0.08	0.07	0.05	0.04	0.02	0.02	0.02

[a]Haplotype frequencies that add up to 0.99 or 1.01 are the result of rounding.
[b]Number of *PRNP* haplotypes resolved per breed.
[c]Haplotype 1 is defined by the presence of the minor allele of SNP 22791 corresponding to the 171R allele associated with scrapie resistance.
[d]Other haplotypes (frequency < 0.02): 11, GCTGCAACAGTC; 12, GCCGTAGTCACT.

higher degree of correlation between alleles, as indicated by r^2 values ranging from 0.30 to 1.0.

Ten common haplotypes, present at a frequency of 0.02 or greater in the MSDP1.1, were predicted by both Bayesian and EM analysis of genotypes at 12 high-frequency (MAF ≥ 0.10) polymorphic loci (Table 3). The ten haplotypes accounted for 93% of the haploid genomes in the sheep panel (Table 4). All common haplotypes, except 5 and 10, were unambiguously identified in homozygous individuals. Two other haplotypes (11, 12), with a frequency of less than 0.02, were also predicted by both Bayesian and EM algorithms but were each observed in a single heterozygous individual and, thus, were excluded from further analysis. Haplotypes 1–7 and the ARR, ARQ, and VRQ codon haplotypes were observed in 14 families and segregated in a manner consistent with Mendelian inheritance, further validating the Bayesian and EM predictions of haplotype phase.

Of the 12 polymorphisms for which D' and r^2 were estimated, seven were identified by Haploview 3.2 as htSNPs. This set of htSNPs was able to define 100% of the common haplotypes in the MSDP1.1.

Based on median-joining network analysis, there were multiple loops and 12 intermediate haplotypes not observed in the MSDP1.1 panel (Fig. 2). Haplotype 1, the most frequent haplotype, was a terminal node and exclusively observed with the ARR codon haplotype. The ARQ codon haplotype was observed with haplotypes 2–10. Though present at low frequencies in our sample, it was nonetheless possible to infer the ARQ haplotype most closely related to the ARH, AHQ, and VRQ haplotypes associated with susceptibility to scrapie. The ARH and AHQ codon haplotypes were observed to occur exclusively on the background of haplotypes 2 and 8, respectively, including the observation of homozygous individuals. The VRQ codon haplotype was observed on the background of haplotypes 4 and 6, which were widely separated across the median-joining network.

To confirm that the VRQ codon haplotype was on the background of haplotypes 4 and 6, 14 families with at least one parent bearing a VRQ codon haplotype and two offspring were sequenced to determine gametic phase. The VRQ codon haplotype was observed on the background of haplotypes 4 and 6 and segregated with those haplotypes. Moreover, a lamb from a family of VRQ heterozygous parents was homozygous for the VRQ codon haplotype and heterozygous for haplotypes 4 and 6. Thus, in this sample of sheep, the VRQ codon haplotype corresponding to amino acid residues at positions 136, 154, and 171 of the prion protein was associated with two divergent haplotypes.

Discussion

In this study, we identified 36 novel polymorphisms, 2 regions of strong LD, 10 common haplotypes, and complex evolutionary relationships underlying ovine *PRNP* codon haplotypes. There are well-documented associations between the ovine *PRNP* codon haplotypes (i.e., ARR, ARQ, ARH, AHQ, and VRQ) and susceptibility to scrapie in sheep. However, these *PRNP* codon haplotypes do not fully explain the clinical observations and epidemiologic complexities of scrapie. Genetically susceptible (i.e., VRQ/VRQ) sheep do not all develop scrapie subsequent to experimental inoculation (O'Rourke et al. 1997). Furthermore, there have been multiple cases of scrapie in ARR/ARR sheep following natural exposure (Ikeda et al. 1995; Orge et al. 2004; Buschmann

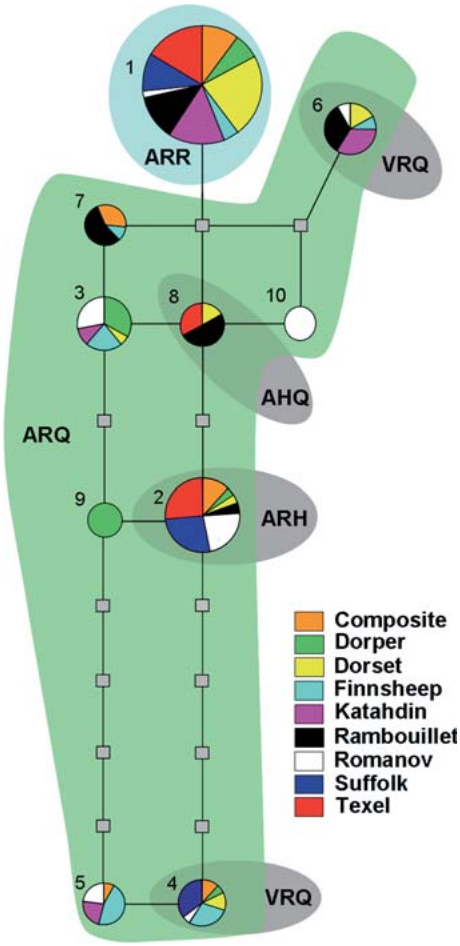

Fig. 2. *PRNP* haplotype relationships as determined by median-joining network analysis. The numbers to the left of each circle correspond to the haplotypes as defined in Table 3. Each circle is proportional in size relative to the haplotype frequency in the MSDP1.1 panel. Areas within each circle represent the relative contribution of each breed to the frequency of that haplotype in the panel. Blue and green backgrounds depict the relationships of the ARR and ARQ codon haplotypes to the network, respectively. The predicted relationships of codon haplotypes AHQ, ARH, and VRQ to the ARQ-bearing haplotypes in the network are indicated by the gray ovals. The SNPs defining these codon haplotypes were present at frequencies below 0.10 and therefore were not used to construct the network. Visual examination of genotypes indicates that the codon haplotypes are most closely related to the haplotype enclosed by the gray oval. Predicted haplotypes not observed in the MSDP1.1 panel are indicated by gray boxes.

et al. 2004). These observations and the emergence of the Nor98 scrapie strain to which the AHQ codon haplotype is most susceptible (as opposed to the VRQ haplotype) indicate that genetic resistance or susceptibility to scrapie is complex (Baylis and McIntyre 2004; Moum et al. 2005). In other species, nucleotide sequence variation in noncoding regions of *PRNP* may significantly affect susceptibility to TSEs and disease progression, a possibility just beginning to be explored in sheep (Mead et al. 2001; O'Neill et al. 2003). By including polymorphisms outside the coding region, the haplotypes defined by this study provide a framework which may allow for a more complete understanding of genetic susceptibility to scrapie and disease progression in sheep.

Haplotype 1, associated with codon haplotype ARR, is a terminal node in the median-joining network analysis and is present at varying frequencies in all breeds examined by this study. The relatively high frequency of haplotype 1 may reflect the accumulation of selective pressure on the *PRNP* gene region based on resistance to scrapie. This haplotype will likely increase in frequency due to positive selection driven by genetics-based scrapie eradication programs in the European Union and United States of America (Baylis and Goldmann 2004).

The ARQ codon haplotype is associated with nine haplotypes identified by this study. In epidemiologic studies of scrapie susceptibility, the ARQ codon haplotype has been treated as a homogeneous genotypic classification. This approach does not consider the possibility of potential functional polymorphisms outside the coding region of *PRNP*. The increased resolution of the haplotypes identified in this study, by encompassing polymorphisms in noncoding regions, provides a means to examine the possibility of other loci within ovine *PRNP* influencing scrapie susceptibility or disease progression. Diversity within the ARQ codon haplotype has been suggested by Goldmann et al. (2005), who identified nine ARQ codon haplotypes. However, these observations included only the coding region and all but one of the haplotypes occurred at a low frequency. In contrast, of the nine haplotypes associated with the ARQ codon haplotype identified in this study, five were present at a frequency of 7% or greater in a diverse panel of sheep.

The ARR codon haplotype was exclusively observed as the most common haplotype in our sample, haplotype 1. It may be descended from ARQ-bearing haplotype 7 or 8, but this relationship is not resolved in the median-joining network (Fig. 2). The codon haplotypes AHQ and ARH were each most closely related to ARQ-bearing haplotypes 2 and 8, respectively. The AHQ and ARH codon haplotypes are likely descendents of haplotypes 2 and 8, however, because of the low frequency of these codon haplotypes in our sample, the exact nature of the relationship could not be determined. Interestingly, the VRQ codon haplotype was most closely related to haplotypes 4 and 6. This result is remarkable because haplotypes 4 and 6 are distantly related, sharing only 2 of 12 SNP alleles (Table 3, Fig. 2). There are a number of ways

the V136 allele could have become associated with two haplotypes having such distinct evolutionary histories, including multiple recombination events, gene conversion, or the possibility of independent dual mutations on the backgrounds of haplotypes 4 and 6. However, there is insufficient data in the present study to implicate any particular mechanism. The functional consequences of the relationship of VRQ to divergent ARQ-bearing haplotypes remain to be determined.

The genetic structure of *PRNP* in sheep populations is complex. The 36 polymorphisms identified, combined with previously reported polymorphisms, predict ten common haplotypes in the MSDP1.1 panel. The structure of the ten common haplotypes identified clarifies the relationships of coding polymorphisms previously shown to associate with scrapie progression in sheep. Further evaluation of these ten haplotypes will allow for a more complete understanding of *PRNP* effects on scrapie susceptibility and disease progression and will complement codon haplotype-based testing and culling programs for scrapie control.

Acknowledgments

The authors thank Drs. G.P. Harhay, J.W. Keele, and S.N. White for help with data analysis, Drs. L.M. Durso, J.E. Keen, and M.R. Mousel for critical comments regarding the manuscript, R.A. Godtel, J.K. Carnahan, and T.A. Sorensen for technical assistance, and S.L. Kluver and J.K. Rosch for secretarial assistance. Names are necessary to accurately report the methods and results, and the use of names by USDA implies no approval of the product to the exclusion of others that may also be suitable. B.T. Green was partially supported by DARPA Project 5438-32000-023-01R.

References

1. Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16, 37–48
2. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265
3. Baylis M, Goldmann W, Houston F, Cairns D, Chong A, et al. (2002) Scrapie epidemic in a fully PrP-genotyped sheep flock. J Gen Virol 83, 2907–2914
4. Baylis M, Goldmann W (2004) The genetics of scrapie in sheep and goats. Curr Mol Med 4, 385–396
5. Baylis M, McIntyre KM (2004) Transmissible spongiform encephalopathies: scrapie control under new strain. Nature 432, 810–811
6. Belt PB, Muileman IH, Schreuder BE, Bos-de Ruijter J, Gielkens AL, et al. (1995) Identification of five allelic variants of the sheep PrP gene and their association with natural scrapie. J Gen Virol 76, 509–517
7. Buschmann A, Luhken G, Schultz J, Erhardt G, Groschup MH (2004) Neuronal accumulation of abnormal prion protein in sheep carrying a scrapie-resistant genotype (PrPARR/ARR). J Gen Virol 85, 2727–2733
8. Clawson ML, Heaton MP, Chitko-McKown CG, Fox JM, Smith TP, et al. (2004) Beta-2-microglobulin haplotypes in U.S. beef cattle and association with failure of passive transfer in newborn calves. Mamm Genome 15, 227–236
9. Clouscard C, Beaudry P, Elsen JM, Milan D, Dussaucy M, et al. (1995) Different allelic effects of the codons 136 and 171 of the prion protein gene in sheep with natural scrapie. J Gen Virol 76, 2097–2101
10. Elsen JM, Amigues Y, Schelcher F, Ducrocq V, Andreoletti O, et al. (1999) Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. Arch Virol 144, 431–445
11. Ewing B, Green P (1998) Base-calling of automated sequencer traces using *phred*. II. Error probabilities. Genome Res 8, 186–194
12. 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
13. Fallin D, Schork NJ (2000) Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am J Hum Genet 67, 947–959
14. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, et al. (2002) Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res 12, 1496–1506
15. Goldmann W, Baylis M, Chihota C, Stevenson E, Hunter N (2005) Frequencies of PrP gene haplotypes in British sheep flocks and the implications for breeding programmes. J Appl Microbiol 98, 1294–1302
16. Heaton MP, Leymaster KA, Freking BA, Hawk DA, Smith TPL, et al. (2003) Prion gene sequence variation within diverse groups of U.S. sheep, beef cattle, and deer. Mamm Genome 14, 765–777
17. Houston F, Goldmann W, Chong A, Jeffrey M, González L, et al. (2003) Prion diseases: BSE in sheep bred for resistance to infection. Nature 423, 498
18. Hunter N, Goldmann W, Foster JD, Cairns D, Smith G (1997a) Natural scrapie and PrP genotype: case-control studies in British sheep. Vet Rec 14, 1137–1140
19. Hunter N, Moore L, Hosie BD, Dingwall WS, Greig A (1997b) Association between natural scrapie and PrP genotype in a flock of Suffolk sheep in Scotland. Vet Rec 140, 59–63
20. Ikeda T, Horiuchi M, Ishiguro N, Muramatsu Y, Kai-Uwe GD, et al. (1995) Amino acid polymorphisms of PrP with reference to onset of scrapie in Suffolk and Corriedale sheep in Japan. J Gen Virol 76, 2577–2581

21. Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29, 217–222
22. Lee IY, Westaway D, Smit AFA, Wang K, Seto J, et al. (1998) Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res 8, 1022–1037
23. Leymaster KA (1991) Straightbred comparison of a composite population and the Suffolk breed for performance traits of sheep. J Anim Sci 69, 993–999
24. Mead S, Mahal SP, Beck J, Campbell T, Farrall M, et al. (2001) Sporadic—but not variant—Creutzfeldt-Jakob disease is associated with polymorphisms upstream of PRNP exon 1. Am J Hum Genet 69, 1225–1235
25. Moum T, Olsaker I, Hopp P, Moldal T, Valheim M, et al. (2005) Polymorphisms at codons 141 and 154 in the ovine prion protein gene are associated with scrapie Nor98 cases. J Gen Virol 86, 231–235
26. 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
27. O'Neill GT, Donnelly K, Marshall E, Cairns D, Goldmann W, et al. (2003) Characterization of ovine PRP gene promoter activity in N2a neuroblastoma and ovine foetal brain cell lines. J Anim Breed Genet 120, 114–123
28. Orge L, Galo A, Machado C, Lima C, Ochoa C, et al. (2004) Identification of putative atypical scrapie in sheep in Portugal. J Gen Virol 85, 3487–3491
29. O'Rourke KI, Holyoak GR, Clark WW, Mickelson JR, Wang S, et al. (1997) PrP genotypes and experimental scrapie in orally inoculated Suffolk sheep in the United States. J Gen Virol 78, 975–978
30. Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. Science 216, 136–144
31. Prusiner SB (1998) Prions. Proc Natl Acad Sci USA 95, 13363–13383
32. Prusiner SB (2004) *Prion Biology and Diseases*, 2nd ed. Cold Spring Harbor Monogr Ser (Cold Spring, NY: Cold Spring Harbor Laboratory Press)
33. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293, 489–493
34. Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73, 1162–1169
35. Stumpf MP (2004) Haplotype diversity and SNP frequency dependence in the description of genetic variation. Eur J Hum Genet 12, 469–477