Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle

Christopher M. Seabury[a,1], Paul M. Seabury[b], Jared E. Decker[c], Robert D. Schnabel[c], Jeremy F. Taylor[c], and James E. Womack[a,1]

[a]Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843-4467; [b]ElanTech, Inc., Greenbelt, MD 20770; and [c]Division of Animal Sciences, University of Missouri, Columbia MO 65211-5300

Contributed by James E. Womack, November 11, 2009 (sent for review September 29, 2009)

The Toll-like receptor (*TLR*) and peptidoglycan recognition protein 1 (*PGLYRP1*) genes play key roles in the innate immune systems of mammals. While the *TLRs* recognize a variety of invading pathogens and induce innate immune responses, *PGLYRP1* is directly microbicidal. We used custom allele-specific assays to genotype and validate 220 diallelic variants, including 54 nonsynonymous SNPs in 11 bovine innate immune genes (*TLR1-TLR10*, *PGLYRP1*) for 37 cattle breeds. Bayesian haplotype reconstructions and median joining networks revealed haplotype sharing between *Bos taurus taurus* and *Bos taurus indicus* breeds at every locus, and we were unable to differentiate between the specialized *B. t. taurus* beef and dairy breeds, despite an average polymorphism density of one locus per 219 bp. Ninety-nine tagSNPs and one tag insertion-deletion polymorphism were sufficient to predict 100% of the variation at all 11 innate immune loci in both subspecies and their hybrids, whereas 58 tagSNPs captured 100% of the variation at 172 loci in *B. t. taurus*. PolyPhen and SIFT analyses of nonsynonymous SNPs encoding amino acid replacements indicated that the majority of these substitutions were benign, but up to 31% were expected to potentially impact protein function. Several diversity-based tests provided support for strong purifying selection acting on *TLR10* in *B. t. taurus* cattle. These results will broadly impact efforts related to bovine translational genomics.

peptidoglycan recognition protein | bovine translational genomics | bovine Toll-like receptors | single nucleotide polymorphism

The bovine genome sequence and first-generation HapMap projects (1, 2) will soon enable genome-assisted selective breeding (3), marker-assisted vaccination (as diagnostics for enhanced vaccine design or animal response), and the development of innate immunologicals used as anti-infectives (4) as methods to mitigate economically important diseases. The mammalian innate immune system provides host defense against a variety of pathogens without requiring prior exposure (5, 6), and genes modulating innate immunity have often been considered candidate loci for improving host resistance to disease in agricultural species. The recently evolved field of innate immunity was originally catalyzed by the discovery that TOLL, a *Drosophila* protein governing developmental polarity, was also required for an effective antifungal immune response in adult *Drosophila* (7–10). To date, nine members of the TOLL gene family have been identified in *Drosophila*, whereas mammals possess 10 or 12 functional Toll-like receptor (TLR) genes (*TLR1-TLR10* in humans and cattle (*TLR1-TLR9* and *TLR11-TLR13* in mouse) (6, 9–14).

Members of the mammalian TLR gene family facilitate the recognition of pathogen-associated molecular patterns (PAMPs) and elicit host innate immune responses (5, 6) to invading bacteria, viruses, protozoa, and fungi. The mammalian TLRs encode type-I transmembrane proteins of the IL-1 receptor (IL-1R) family that possess N-terminal leucine-rich repeats involved in ligand recognition, a transmembrane domain, and a C-terminal intracellular Toll/IL-1 receptor homologous (TIR/IL-1R) domain for signal transduction (5, 6, 15). Mammalian TLR gene family members are primarily expressed by antigen-presenting cells, such as macrophages or dendritic cells, and previous investigations have elucidated the ligand specificities for most mammalian TLRs, with six gene family members (*TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, *TLR9*) known to recognize microbial (bacteria, fungi, protozoa) and synthetic ligands, and five gene family members (*TLR3*, *TLR4*, *TLR7–TLR9*) known to recognize viral components (1, 11, 15). *TLR10* is the only functional member of the human TLR gene family for which specific ligands are yet to be identified (16). However, given evidence that the TLR10 protein forms functional heterodimers with both TLR1 and TLR2 (16), it is possible that TLR10 may participate in the recognition of a variety of microbial PAMPs, including those recognized by TLR2 (16–18).

Unlike the TLR proteins, mammalian peptidoglycan recognition proteins (PGRPs; encoded by *PGLYRP1-PGLYRP4*) modulate innate immunity via PAMP recognition and microbicidal activity (19, 20). Mammalian PGLYRP2 hydrolyzes bacterial peptidoglycan, with the remaining three functioning as bactericidal or bacteriostatic proteins (20–22). Bovine PGLYRP1 binds to a variety of microbial components and kills diverse microorganisms (19). Consequently, bovine PGRPs may be generalists in both antimicrobial affinity and activity (19), rendering them potentially important for suppressing infectious diseases in food animal species.

To date, several studies have demonstrated that some naturally occurring TLR variants enhance the risk of severe infections in humans, mice, and most recently, domestic cattle (23–26). In addition to the emerging molecular and population-based case-control data, bovine health-related quantitative trait loci (QTL) have been localized to genomic regions either proximal to or directly overlapping one or more TLR loci, and at least two health-related QTL are proximal to *PGLYRP1* (25–31). Unfortunately, the resolution of the highest density SNP assay available for cattle, the Illumina BovineSNP50 (32), is inadequate to resolve the identity of any of the TLRs as underlying these QTL. Therefore, a considerable need exists to characterize the naturally occurring variation and haplotype structure within the bovine *TLRs* to evaluate the involvement of these loci in bovine disease susceptibility. This information will facilitate translational genomics related to marker-, and ultimately, whole-genome-assisted methods of animal selection to develop cattle populations with increased resistance to infectious diseases.

Author contributions: C.M.S. and J.E.W. designed research; C.M.S. performed research; C.M.S., P.M.S., and J.E.W. contributed new reagents/analytic tools; C.M.S. analyzed data, with some network contributions from J.E.D., R.D.S., J.F.T., and J.E.W.; C.M.S. wrote the paper; C.M.S. and KBioscience performed genotyping; J.E.W. provided DNA; P.M.S. engineered software to compile and manage data; and J.F.T. performed regression.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: cseabury@cvm.tamu.edu or jwomack@cvm.tamu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0913006107/DCSupplemental.

GENETICS

Herein, we characterized 220 bovine polymorphisms within the 10 *TLRs* and *PGLYRP1* (27–29, 33) in 37 cattle breeds representing *Bos taurus taurus, Bos taurus indicus*, and subspecific hybrids. We also comprehensively report on bovine *TLR* and *PGLYRP1* haplotype structure, haplotype sharing among breeds and subspecific lineages, and provide median joining networks as putative representations of haplotype evolution (34). These results provide insights into the evolution of 11 innate immune genes within two bovine lineages and enable focused efforts in bovine translational genomics.

Results

Bovine Polymorphism Validation, Minor Allele Frequency Spectrum, and Haplotype Inference. We genotyped 325 SNPs and seven indels distributed within 11 bovine innate immune genes (*TLR1-TLR10* & *PGLYRP1*) (27–29, 33) via custom fluorescent assays applied to a cohort of elite cattle sires ($n = 101$) representing 37 breeds. Ten SNPs failed to cluster within their National Center for Biotechnology Information RefSeq coordinates based on bovine genome assembly Btau4.0, and were excluded from further analysis. Of the remaining 322 loci, 220 were validated as diallelic and were reliably scored (216 SNPs + 4 indels; Table S1), 37 were monomorphic, and 65 failed quality control because of either poor genotype clustering ($n = 54$) or discrepancies between the original sequences and the fluorescent genotyping assays ($n = 11$). Of the 216 SNPs, 54 (25%) were predicted to encode nonsynonymous substitutions (nsSNPs). Overall, 214 SNPs and four indels were successfully incorporated into 144 discrete haplotypes via Bayesian reconstructions (35) (Table 1). Two validated SNPs (*TLR2 rs68268259*; *TLR4 rs8193072*) could not be incorporated into haplotypes with best-pair phase probabilities ≥ 0.90. The total number of predicted haplotypes, percentage of sires with haplotype phase probabilities ≥ 0.90, number of variable loci with minor allele frequency (MAF) ≤ 0.10, genic distributions of validated variable sites, size of the investigated regions, and average estimates of linkage disequilibrium (LD; r^2) between adjacent sites are presented in Table 1. The MAF spectrum ranged from 0.005 to 0.50 and four SNPs were determined by Haploview (36) to deviate from Hardy-Weinberg equilibrium (HWE) at autosomal loci ($n = 101$ sires, 37 breeds; *TLR10 rs55617227*; *TLR3 rs42851896*, *rs42851897*, *rs55617164*). However, because theoretical panmixia best applies to individual breeds, differences in allele frequencies among breeds because of drift or selection are likely responsible for departures from HWE.

Intragenic LD Architecture and tagSNPs/Indels. Examination of the intragenic patterns of LD via 95% confidence intervals constructed for D′ (36, 37), application of the four-gamete rule (36), and estimates of recombination between adjacent sites (38, 39) revealed one or more blocks of strong LD within each of the 11 innate immune genes. Evidence for historical recombination was detected within *TLR3*, *TLR4*, and *TLR10*, resulting in at least two detectable LD blocks within each gene. All other genes exhibited a single block of strong LD spanning all, or the majority of all validated SNPs and indels ($n = 218$), as indicated by majority rule of all three analyses (36–39). Pairwise comparisons of r^2 revealed moderate to high levels of average LD across most genes, with some discrete evidence for historical recombination events within *TLR3*, *TLR4*, and *TLR10*. Regions of LD and historical recombination were also detected and enumerated using the general model for varying recombination rate between adjacent sites ($n = 11$ genes) (38, 39). Five comparisons between adjacent SNP sites [*TLR3* (3), *TLR4* (1), and *TLR10* (1)] produced median recombination rate estimates that exceeded estimates for the background rate ($\bar{\rho}$) (38, 39) by a factor ≥ 2.0. The highest median recombination rate estimate was observed in *TLR4* (between *rs8193059* and *rs8193060*), and exceeded the background rate by a factor ≥ 6.1. Additional fine-scale analyses (38, 39) failed to identify recombination hotspots within *TLR3*, *TLR4*, or *TLR10*. Analyses to identify tagSNPs/Indels which captured 100% of the variation at 218 variable sites within all 11 innate immune genes for 37 breeds yielded 99 tagSNPs and 1

Table 1. Summary data for 11 bovine innate immune genes investigated in 37 cattle breeds

Bovine gene	BTA assign[a]	Total haps[b]	Sires phased (%)[c]	MAFs ≤ 0.10[d]	Avg r^2 all[e]	Avg r^2 *B.t.t.*[e]	Valid. SNPs[f]	Hap SNPs[g]	Valid. indels[h]	Valid. nsSNPs[i]	Region size[j] (kb)	QTL or assoc.[k]
PGLYRP1	BTA18	8	95	2	0.08	0.14	7	7	NA	1	1.6	Q
TLR1	BTA6	7	97	2	0.23	0.47	4	4	NA	1	1.5	Q, A
TLR2	BTA17	17	93	11	0.35	0.60	31	30	1	13	3.0	A
TLR3	BTA27	33	85	15	0.45	0.72	50	50	ND	3	10.6	Q
TLR4	BTA8	13	91	19	0.14	0.37	22	21	NA	6	9.8	Q, A
TLR5	BTA16	13	100	13	0.40	0.59	29	29	3	3	5.1	No
TLR6	BTA6	15	94	4	0.27	0.29	15	15	NA	7	1.9	Q
TLR7	BTAX	7	100	8	0.27	0.23	10	10	NA	1	3.9	Q
TLR8	BTAX	3	100	0	0.93	0.92	9	9	NA	6	2.3	Q
TLR9	BTA22	9	98	2	0.35	0.40	11	11	NA	1	4.3	Q
TLR10	BTA6	19	91	22	0.31	0.45	28	28	NA	12	3.8	Q
Total/Avg[l]		144	95[l]	98	0.34[l]	0.47[l]	216	214	4	54	4.3[l]	

[a]BTA assignments based on National Center for Biotechnology Information Refseq (Btau4.0) and radiation hybrid mapping (12).
[b]Total haplotypes predicted from all validated markers and best-pair reconstructions (35) with probabilities ≥0.90.
[c]Percent of sires ($n = 101$) exhibiting best pair phase probabilities ≥0.90. BTAX haplotypes were direct observations.
[d]Total polymorphisms with minor allele frequencies ≤0.10.
[e]Average intragenic r^2 values estimated for adjacent SNP and indel sites for all cattle or for *B. t. taurus* (*B.t.t.*).
[f]Numbers of putative SNPs validated as polymorphic.
[g]Numbers of validated SNPs placed on discrete haplotypes.
[h]Numbers of putative indels validated as polymorphic. NA = not applicable; ND = not determined.
[i]Numbers of putative nonsynonymous SNPs validated as polymorphic.
[j]Size of the genic region rounded to the nearest 100 bp. Kb = kilobase.
[k]Bovine health-related QTL overlapping or proximal to investigated gene (Q), or intragenic variation associated (A) with disease susceptibility in case-control studies (25–31).
[l]Average across all investigated genes.

tagIndel (Table S2). All four SNPs within *TLR1* were required to explain 100% of the genetic diversity. Within *B. t. taurus* breeds, 58 tagSNPs were required to capture 100% of the variation detected at 172 variable sites (Table S2). The need for fewer tagSNPs within *B. t. taurus* cattle reflects higher levels of intragenic LD and fewer variable sites.

Bovine Haplotype Networks and Haplotype Sharing. Evaluation of all median joining haplotype networks (Fig. 1, Fig. S1, and Table S3) revealed that: (*i*) The specialized *B. t. taurus* beef and dairy breeds could not be differentiated based on haplotypes predicted for all 11 innate immune genes, despite an average polymorphism density of one SNP/Indel per 219 bp; (*ii*) The 250 Kyr divergence between *B. t. taurus* and *B. t. indicus* (40) was revealed in many, but not all, haplotype networks (i.e., *TLR2*, *TLR4*, *TLR5*, *TLR6*, *TLR10*); and (*iii*) Low levels of haplotype sharing between *B. t. taurus* and *B. t. indicus* breeds were predicted for every gene. Because of the recent formation of *B. t. taurus* × *B. t. indicus* hybrids in North America, predicted haplotypes for these breeds predominantly fell within network nodes dominated by both taurine and indicine breeds, with some haplotypes also localizing to shared (*B. t. taurus* and *B t. indicus*) or unique network nodes. Predicted haplotypes for breeds generally considered to be of ancient origin (Braunvieh, Scottish Highland, White Park) were frequently found within network nodes representing the highest frequency haplotypes. Network nodes representing high-frequency haplotypes predicted to be shared at low levels among *B. t. taurus* and *B. t. indicus* also contained haplotypes derived from one or more of the ancient breeds for all loci except *TLR9*. The *B. t. indicus* breed composition within high-frequency nodes containing low levels of haplotypes shared among subspecific lineages was not restricted to a single indicine breed or sire, with one or more Brahman and/or Nelore sires contributing to these nodes.

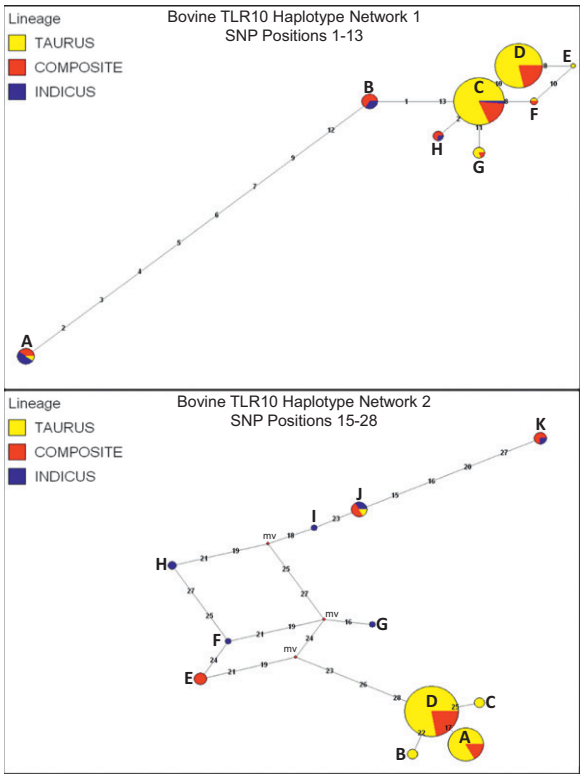

Fig. 1. Median joining (MJ) haplotype networks for bovine *TLR10* using haplotypes predicted for all cattle (n = 101; 37 breeds). Because MJ networks require the absence of recombination (57), each network represents intragenic regions of elevated LD. Haplotypes predicted for *B. t. taurus*, *B. t. indicus*, and hybrids (termed "composite") are color coded. Numbers indicate SNP positions (28) in numerical order (1–21 are coding) (Table S1). Node sizes are proportional to haplotype frequency and all branch lengths are drawn to scale. Alphabetized letters at nodes represent the breed distribution of each haplotype (Table S3). Median vectors are indicated as "mv."

Predicted Impact of Bovine Amino Acid Substitutions and Evolutionary Inferences. Using PolyPhen (41) and SIFT (42), we predicted that 37 of 54 (69%) amino acid substitutions encoded by nsSNPs are benign and tolerated, with 17 of 54 (31%) possibly eliciting phenotypic effects on the mature proteins (43). Of these, 12 of 17 (71%) were observed at frequencies ≤ 0.076, but 5 of 17 (29%) located in *PGLYRP1*, *TLR2*, *TLR3*, and *TLR6* were detected at frequencies ≥ 0.126. Results from PolyPhen and SIFT were congruent for 3 of the 17 (18%) nsSNPs (Table 2).

To examine the potential magnitude of functional and/or selective constraint related to perturbation of bovine TLR protein function, we compared the proportion of "benign" or "tolerated" amino acid substitutions (37/53) to the proportion predicted to impact protein function (16/53) using a likelihood-ratio test and found there to be fewer nsSNP likely to impact protein function than benign or tolerated amino acid substitutions ($P \leq 0.054$). This is consistent with the hypothesis that purifying selection operates to preserve the products of most protein-coding genes (43–45). However, conservative amino acid substitutions were less frequent than substitutions predicted to impact protein function (1:2 ratio) for *TLR3* and *TLR5*, considering the combined PolyPhen and SIFT results. Notably, *TLR10* was the only gene for which we were able to technically ascertain (KASPar system) and validate every previously reported coding region SNP (28) while also achieving acceptable phase probabilities (≥ 0.90) for most (91%) of the sampled cattle (Table 1). Evaluation of *TLR10* coding variation within *B. t. taurus* breeds (n = 130 chromosomes; 26 breeds) using Tajima's D (46) and Fu and Li's frequency distribution tests (D^*, F^*) (47) revealed significantly negative values for all statistics (D = −1.69898, $P < 0.05$; D^* = −4.38979, $P < 0.01$; F^* = −4.04479, $P < 0.01$). Fu's Fs statistic (48), a haplotype-based frequency distribution test, also produced a negative value (−1.420), but was not significant ($P > 0.05$). Finally, 11 of the 16 segregating sites observed in *TLR10* for *B. t. taurus* cattle were singletons. Collectively, these analyses suggest strong purifying selection acting on *TLR10* in *B. t. taurus* cattle. A limited sample size precluded similar analyses for *B. t. indicus* cattle.

A regression-based approach that considered all of the variable sites and the effective number of alleles at each site also demonstrated that *TLR10* possesses significantly less diversity than the other eight autosomal loci ($P \leq 0.05$) (Fig. S2). However, a similar haplotype-based approach did not support a reduction in *TLR10* haplotype diversity relative to the other autosomal genes. Moreover, the haplotype distribution for *TLR10* contained multiple moderate and many low frequency haplotypes.

Discussion

Our assays provide a 55-fold increase in marker density relative to the Illumina BovineSNP50 assay, which queries only four SNPs within (*TLR6*, *TLR10*) or proximal to (*TLR7*, *TLR8*) 4 of the 11 innate immune genes investigated. Validated polymorphisms, haplotype distributions, and tagSNPs/Indels elucidated in this study will directly facilitate fine mapping of existing bovine health-related QTL while also enabling independent evaluation of SNPs tentatively associated with susceptibility to Johne disease (25–31). Because of limitations related to our genotyping assays, we were able to technically ascertain all previously described coding polymorphisms (27–29, 33) only for bovine *TLR10*, and *TLR1* variants posed the greatest technical challenge because of sequence similarity with *TLR6*. For this reason, at least partial DNA sequencing

Table 2. Summary data for 17 nonsynonymous SNPs predicted to impact protein function

Bovine gene	SNP[a]	dbSNP ID	GenBank protein ID	AA Subst.[b]	Protein domain[c]	PolyPhen result[d]	SIFT result[d]	SNP Freq[e]
PGLYRP1	T > C	*rs68268284*	NP_776998	Y76H	PGRP-Ami_2	PD	T	0.265
TLR2	T > A	*rs68268251*	NP_776622.1	F227L	NCP	PD	T	0.070
	G > A	*rs68268260*	NP_776622.1	R563H	LRRCT	B	AF	0.220
TLR3	G > T	*rs42852439*	NP_001008664.1	S664I	LRRCT	PD	T	0.425
	G > A	*rs55617272*	NP_001008664.1	G426S	LRR-12	PD	AF	0.045
TLR4	A > C	*rs8193049*	NP_776623.5	N151T	LRR-3	PD	T	0.030
	A > G	*rs8193055*	NP_776623.5	K381R	LRR-6	B	AF	0.021
TLR5	G > A	*rs55617251*	XP_594146.3	A659T	Transmembrane	B	AF	0.025
	G > A	*rs55617166*	XP_594146.3	E842K	NCP	B	AF	0.015
TLR6	T > G	*rs68268270*	NP_001001159.1	L43R	NCP	PD	AF	0.020
	A > G	*rs68268272*	NP_001001159.1	R87G	LRR-1	B	AF	0.070
	A > G	*rs68268275*	NP_001001159.1	T395A	LRR-3	B	AF	0.126
TLR8	G > A	*rs55617351*	NP_001029109.1	S468N	NCP	B	AF	0.475
TLR10	G > A	*rs55617437*	NP_001070386.1	R18H	Signal Peptide	PD	T	0.076
	C > G	*rs55617286*	NP_001070386.1	I134M	LRR-4	B	AF	0.059
	A > C	*rs55617297*	NP_001070386.1	K753T	TIR	PD	AF	0.020
	G > A	*rs55617343*	NP_001070386.1	R763H	TIR	B	AF	0.046

[a]SNPs were previously described (27–29, 33) and validated in the present study.
[b]Amino acid (AA) substitutions predicted from reference proteins (GenBank Protein ID) and/or previous studies (27–29, 33).
[c]Protein domain locations predicted by SMART (http://smart.embl-heidelberg.de/). Only confidently predicted domains are depicted [NCP, no confident prediction; leucine-rich repeats (LRR) are named in order of prediction].
[d]Results from PolyPhen and SIFT (41, 42). Results other than "Benign (B)" or "Tolerated (T)" are predicted to be Possibly Damaging (PD) or Affect protein Function (AF).
[e]Observed frequency of nonsynonymous SNP allele in all 37 cattle breeds.

from long-range PCR products designed to specifically amplify each locus will be needed to ascertain all polymorphisms present within these two genes.

Evaluation of r^2 values between adjacent variable sites within the autosomal genes provided evidence for enhanced intragenic LD within *B. t. taurus* breeds (69 cattle; 26 breeds), relative to the combined sample (101 cattle, 37 breeds; *B. t. taurus*, *B. t. indicus*, *B. t. taurus* × *B. t. indicus* hybrids) (Table 1). This result is consistent with recent studies of bovine haplotype structure and LD over relatively short physical distances (2, 49), and reflects the fact that 250 Kyr of divergence between *B. t. taurus* and *B. t. indicus* cattle (40) has allowed drift and/or selection to drive different haplotypes to high frequency between the subspecies. In contrast to the autosomal loci, average r^2 values between adjacent variable sites within the X-linked genes (*TLR7*, *TLR8*) increased when all subspecific breeds were combined (Table 1). We expected uniformly higher LD within the X-linked genes because of the smaller effective population size (chromosomal) and female-limited recombination, as compared to autosomal loci. However, we did not expect phase relationships to be preserved across subspecies, which is required for an increase in r^2 on pooling of the subspecific samples, with strong selection also required to converge on a small number of effective *TLR7* and *TLR8* haplotypes in both subspecies. Notably, for variable sites evaluated herein, only three haplotypes were predicted for *TLR8*, and only two *TLR7* haplotypes were found in 85% of the surveyed cattle (Fig. S1 and Table S3). LD is extremely high within bovine *TLR8*, moderate in *TLR7*, and high in *TLR3* (Table 1), as compared to other loci, which suggests differences in intragenic selection, as recently suggested for the human TLRs (50). Future studies in which all polymorphisms are technically ascertained and analyzed in larger samples representing individual breeds will be necessary to properly assess deviations from a strictly neutral model of evolution for these genes.

Remarkably, median joining haplotype networks for all loci (Fig. 1 and Fig. S1) were unable to discriminate between the specialized *B. t. taurus* beef and dairy breeds, but did reveal low levels of haplotype sharing between the *B. t. taurus* and *B. t. indicus* breeds (Table S3). An inability to distinguish between specialized beef and dairy breeds based upon haplotypic data has previously been demonstrated (49), but not at the high-resolution SNP densities used here. Because some level of haplotype sharing among *B. t. taurus* and *B. t. indicus* breeds was predicted at all investigated loci, it seems unlikely that hybrid introgression has occurred at every locus, including all regions of genes that have been dissected by historical recombination (Fig. 1, Fig. S1, and Table S3). Therefore, two alternative explanations seem plausible: (*i*) Haplotype sharing across the *B. t. taurus* and *B. t. indicus* lineages represents retention of conserved ancestral variation that predates subspecific divergence, or (*ii*) Both lineages have evolutionarily converged on a relatively small number of innate immune haplotypes at these loci. Comprehensive sequencing and phylogenetic analyses that include *B. t. taurus*, *B. t. indicus*, and other members of the subfamily Bovinae will likely be necessary to elucidate the origin of *Bos* spp. haplotype sharing.

Our analysis of amino acid substitution phenotypes using PolyPhen and SIFT (Table 2) indicates that the bovine *TLR* family has evolved under functional and selective constraints, with < 31% of amino acid replacements expected to potentially affect protein function; the majority of these (71%) were observed at low frequencies (≤ 0.076). These findings are consistent with SIFT/PolyPhen estimates for human protein coding genes, where 25 to 32% of amino acid substitutions are predicted to affect protein function (43), thereby supporting the hypothesis that some level of purifying selection operates to preserve the products and function of most protein coding genes (43–45). However, several nsSNPs encoding amino acid substitutions predicted to alter protein function were observed at moderate to high frequencies (Table 2).

Frequency distribution tests applied using all coding region polymorphisms previously described within bovine *TLR10* (28) indicated an excess of rare and/or singleton variants in *B. t. taurus* cattle (n = 130 chromosomes; 26 breeds). Significantly negative values for Tajima's (D) and Fu and Li's tests (D^*, F^*) are often interpreted as evidence for either purifying or directional selection, but may also indicate a recent population expansion, or violations of either the mutation-drift equilibrium assumption (46) or random sample requirement (47) for these tests. Although Fu's Fs statistic was negative (−1.420), coalescent simulations to estimate the critical values of Fs did not support significant directional selection of

TLR10 haplotype variation within *B. t. taurus* cattle. Likewise, a regression-based comparison of all 11 innate immunity genes revealed less than expected allelic diversity in *TLR10* relative to the other autosomal loci, but no evidence of reduced haplotypic diversity (Fig. S2). Recent studies provide evidence of population bottlenecks at the time of domestication and breed formation in modern cattle (2, 49), which are expected to drive frequency distribution tests toward more positive values because of the loss of rare variants. Similarly, ascertainment bias because of sequencing small numbers of individuals results in the discovery of primarily common variation which biases frequency distribution tests toward more positive values (51), yet an excess of rare and/or singleton *TLR10* variants were present in our *B. t. taurus* panel. One plausible explanation for this is that strong purifying selection has constrained *TLR10* within the *B. t. taurus* lineage. This explanation is supported by evidence that the vertebrate *TLRs* evolved under purifying selection (52, 53), thereby preserving host ability to recognize specific PAMPs. Interestingly, *TLR10* is the only functional member of the human *TLR* gene family for which a specific ligand has not been identified (16). Therefore, the elucidation of one or more relevant ligands may provide further insight into bovine *TLR10* evolution.

Conclusions

Our analysis of haplotype structure, LD architecture, and tagSNP/Indel assignment for 11 bovine innate immune genes will enable studies which assess the relationships between variation within these genes and differential susceptibility to disease (25–31). Moreover, our analyses will contribute to the design of next-generation bovine genotyping assays for mapping variation underlying bovine health-related traits.

Perhaps one of the most interesting outcomes of this study was our inability to completely discriminate between specialized *B. t. taurus* beef and dairy breeds, or even *B. t. taurus* and *B. t. indicus* breeds, based on high-resolution haplotypes for 11 innate immune genes. Because the high- versus low-intensity management of dairy versus beef herds results in different disease profiles among cattle, we expect these results to tangibly impact future initiatives to translate bovine innate immune variation into health-related trait information. In addition to the interbreed and subspecific haplotype sharing, evidence for strong purifying selection within *B. t. taurus* *TLR10* was unexpected. Clearly, evolution under strong purifying selection would help ensure bovine TLR10 ligand recognition. However, what specific ligand or ligands recognized by TLR10 are important enough to elicit strong purifying selection in *B. t. taurus* cattle? The answer to this question is currently unknown.

Methods

DNA Samples and Genotyping. Bovine DNA samples (n = 101) representing *B. t. taurus*, *B. t. indicus*, and their hybrids were isolated from spermatozoa (54). Bovine subspecies designation, breed names, and sample sizes (in parentheses) were: *B. t. taurus*: Angus (three), Belgian Blue (four), Blonde d'Aquitaine (five), Braunvieh (four), Brown Swiss (two), Charolais (five), Chianina and Chiangus (five), Corriente (one), Gelbvieh (four), Hereford (three), Holstein (three), Jersey (one), Limousin (three), Maine-Anjou (four), Murray Gray (two), Normande (one), Pinzgauer (one), Red Angus (three), Red Poll (one), Salers (three), Senepol (two), Shorthorn (five), Simmental (one), Tarentaise (one), Scottish Highland (one), White Park (one); *B. t. indicus*: Brahman (three), Nelore (two); Hybrids, termed Composites: Beefmaster (five), Braford (three), Brahmousin (two), Brangus (five), Piedmontese (one), Red Brangus (two), Romagnola (two), Santa Gertrudis (four), Simbrah (three). Subspecies were assigned based on phenotype and breed origin (http://www.ansi.okstate.edu/breeds/cattle/).

SNPs and indels were genotyped using the KASPar allele-specific fluorescent genotyping system (Kbiosciences). Thermal cycling parameters and reaction concentrations followed manufacturer's recommendations, with some modifications to $MgCl_2$ concentrations. Primer sequences and $MgCl_2$ concentrations are available on request. Genotype clustering and calling was performed using KlusterCaller software (Kbiosciences) and the endpoint genotyping module incorporated within the Roche LC480 (Roche Applied Science). Both methods produced congruent genotype calls. Genotype quality was assessed by comparing KASPar-derived genotypes to those derived from sequencing data previously reported (27–29, 33), with inconsistent genotypes categorized as quality control failures. Variable sites failing quality control were excluded.

Haplotype Inference, LD Estimates, and Variant Tagging. Unphased diploid genotypes were compiled and cross-checked for parsing errors using two custom software packages (GenoConvert and GenoConvert2; ElanTech Inc.). Haplotype reconstruction and missing data imputation (≤1.71%) was performed with PHASE 2.1 (35, 55, 56) using all validated intragenic polymorphisms, all cattle (n = 101), and the –X10 option. Haplotype estimation using PHASE 2.1 is not sensitive to departures from HWE (35, 55, 56). Predicted haplotype phases with best pair probabilities ≥0.90 were retained for further analysis. Bovine X-linked haplotypes (*TLR7*, *TLR8*) were directly established by genotype homozygosity. Estimates of recombination across each gene were also assessed in PHASE 2.1 using the general model for varying recombination rate (38, 39). Deviation from the average background recombination rate ($\bar{\rho}$) (38, 39) by a factor ≥ 2.0 between adjacent sites was considered evidence for historical recombination. The potential for recombination hotspots within *TLR3*, *TLR4*, and *TLR10* was modeled by specifying a location based on deviations ≥2.0 from $\bar{\rho}$, and by allowing PHASE 2.1 to independently estimate hotspot locations (38, 39). Statistical support for a putative hotspot was defined by ≥95% of the full posterior distribution of recombination parameters deviating from the background rate by a factor ≥5.0.

Intragenic LD was visualized within Haploview (36) using unphased diploid autosomal genotypes and phase-known X-linked data (*TLR7*, *TLR8*) for *B. t. taurus* (n = 69) and all cattle combined (n = 101). LD patterns and blocks were estimated via majority rule from: 95% confidence intervals constructed for D′ (36, 37), application of the four-gamete rule (36) (fourth gamete > 0.02), and estimates of recombination between adjacent sites (38, 39). To further evaluate the patterns of LD decay, pairwise r^2 values were estimated with Haploview for all validated markers within each gene for *B. t. taurus* and all cattle combined. A minimal set of tagSNPs/Indels capturing 100% of the variation ($r^2 > 0.80$) segregating in *B. t. taurus* and all cattle combined was deduced using the Tagger algorithm in Haploview.

Median Joining Haplotype Networks. Because median joining (MJ) networks require the absence of recombination (57), genes displaying evidence of historical recombination (*TLR3*, *TLR4*, *TLR10*) were partitioned into two regions of elevated LD. Haplotypes were reconstructed (35) for each intragenic region (*TLR3*, *TLR4*, *TLR10*) and best pairs were used for MJ network analyses (34). This approach improved the proportion of cattle with best pairs phase probabilities ≥0.90 and eliminated regions displaying overt evidence of recombination. All MJ networks were constructed using Network 4.5.1.0 (Fluxus Technology Ltd) and the suggested character weights of 10 for SNPs and 20 for indels. Results were visualized, annotated, and adjusted within Network Publisher (Fluxus Technology Ltd). Network branch angles were adjusted to ensure proper magnification and clarity without changing branch lengths.

Amino Acid Substitution Phenotypes and *TLR10* Evolutionary Analyses. Bovine amino acid substitution phenotypes were predicted using PolyPhen (41) and SIFT (42) (http://genetics.bwh.harvard.edu/pph/; http://sift.jcvi.org/) with the default settings. A likelihood ratio test was performed assuming a binomial distribution for nsSNPs versus silent SNP classes to test the hypothesis that they occurred at equal frequency. Frequency distribution tests, including Tajima's D (46) and Fu and Li's Tests (D^*, F^*, Fs) 47–48), were performed in DnaSP v4.90.1 (58) using all *TLR10* coding region polymorphisms (28). Significance levels for frequency distribution tests were defined by confidence intervals estimated for each test statistic via coalescent simulation (10,000 replicates) (58). Simulations were performed given empirical estimates of θ and the observed number of segregating sites, both with and without recombination (58, 59).

At each polymorphism we estimated the effective number of alleles as $E_i = 1/[1 - 2p_i(1 - p_i)] = 1/[p_i^2 + (1 - p_i)^2]$ = 1/(expected HWE frequency of homozygotes) where p_i is allele frequency at the i^{th} locus. Thus, a measure of polymorphism diversity is $\log_2(E_i)$, which also represents the information content of each SNP. For monomorphic SNPs $\log_2(E_i) = 0$ and for SNPs with $p_i = 0.5$, $\log_2(E_i) = 1$. Thus, by summing across the N_j polymorphisms within the j^{th} gene we obtain the diversity index $I_j = \sum_{i=1}^{N_j} E_i$. We used regression analysis to examine the relationship between I_j and N_j for these genes and to test for outliers using 95% confidence estimates for the fitted regression. We also computed the expected number of haplotypes at each gene as E_j = 1/(expected HWE frequency of haplotype homozygotes) and performed a similar regression analysis to examine the relationship between observed and expected numbers of haplotypes.

ACKNOWLEDGMENTS. We thank Dr. Alejandro Rooney for critical review of the manuscript. This project was supported by the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service National Research Initiative Grant 2009-35205-05058 (to C.M.S.), Department of Homeland Security funding to the National Center for Foreign Animal and Zoonotic Disease Defense, the Robert J. Kleberg Jr and Helen C. Kleberg Foundation funding (to J.E.W.), and the Texas AgriLife Research and the College of Veterinary Medicine, Texas A&M University.

1. Elsik CG, et al.; Bovine Genome Sequencing and Analysis Consortium (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324:522–528.
2. Bovine HAPMAP Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528–532.
3. VanRaden PM, et al. (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci* 92:16–24.
4. Rosenthal KL (2006) Tweaking innate immunity: The promise of innate immunologicals as anti-infectives. *Can J Infect Dis Med Microbiol* 17:307–314.
5. Vasselon T, Detmers PA (2002) Toll receptors: a central element in innate immune responses. *Infect Immun* 70:1033–1041.
6. Kaisho T, Akira S (2006) Toll-like receptor function and signaling. *J Allergy Clin Immunol* 117:979–987. quiz 988.
7. Stein D, Roth S, Vogelsang E, Nüsslein-Volhard C (1991) The polarity of the dorsoventral axis in the *Drosophila* embryo is defined by an extracellular signal. *Cell* 65:725–735.
8. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA (1996) The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* 86:973–983.
9. Tauszig S, Jouanguy E, Hoffmann JA, Imler J-L (2000) Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. *Proc Natl Acad Sci USA* 97: 10520–10525.
10. Beutler B (2004) Inferences, questions and possibilities in Toll-like receptor signalling. *Nature* 430:257–263.
11. Akira S, Takeda K (2004) Toll-like receptor signalling. *Nat Rev Immunol* 4:499–511.
12. McGuire K, et al. (2005) Radiation hybrid mapping of all 10 characterized bovine Toll-like receptors. *Anim Genet* 37:47–50.
13. Goldammer T, et al. (2004) Mastitis increases mammary mRNA abundance of beta-defensin 5, toll-like-receptor 2 (TLR2), and TLR4 but not TLR9 in cattle. *Clin Diagn Lab Immunol* 11:174–185.
14. White SN, Kata SR, Womack JE (2003) Comparative fine maps of bovine Toll-like receptor 4 and Toll-like receptor 2 regions. *Mamm Genome* 14:149–155.
15. West AP, Koblansky AA, Ghosh S (2006) Recognition and signaling by Toll-like receptors. *Annu Rev Cell Dev Biol* 22:409–437.
16. Hasan U, et al. (2005) Human TLR10 is a functional receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. *J Immunol* 174:2942–2950.
17. Ozinsky A, et al. (2000) The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. *Proc Natl Acad Sci USA* 97:13766–13771.
18. Mukhopadhyay S, Herre J, Brown GD, Gordon S (2004) The potential for Toll-like receptors to collaborate with other innate immune receptors. *Immunology* 112: 521–530.
19. Tydell CC, Yuan J, Tran P, Selsted ME (2006) Bovine peptidoglycan recognition protein-S: antimicrobial activity, localization, secretion, and binding properties. *J Immunol* 176:1154–1162.
20. Wang M, et al. (2007) Human peptidoglycan recognition proteins require zinc to kill both gram-positive and gram-negative bacteria and are synergistic with antibacterial peptides. *J Immunol* 178:3116–3125.
21. Gelius E, Persson C, Karlsson J, Steiner H (2003) A mammalian peptidoglycan recognition protein with N-acetylmuramoyl-L-alanine amidase activity. *Biochem Biophys Res Commun* 306:988–994.
22. Lu X, et al. (2006) Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J Biol Chem* 281:5895–5907.
23. Merx S, Zimmer W, Neumaier M, Ahmad-Nejad P (2006) Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. *Hum Mutat* 27:293.
24. Texereau J, et al. (2005) The importance of Toll-like receptor 2 polymorphisms in severe infections. *Clin Infect Dis* 41 (Suppl 7):S408–S415.
25. Mucha R, Bhide MR, Chakurkar EB, Novak M, Mikula I, Sr (2009) Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle. *Vet Immunol Immunopathol* 128:381–388.
26. Bhide MR, et al. (2009) Novel mutations in TLR genes cause hyporesponsiveness to *Mycobacterium avium* subsp. *paratuberculosis* infection. *BMC Genet* 10:21.
27. Cargill EJ, Womack JE (2007) Detection of polymorphisms in bovine Toll-like receptors 3, 7, 8, and 9. *Genomics* 89:745–755.
28. Seabury CM, Cargill EJ, Womack JE (2007) Sequence variability and protein domain architectures for bovine Toll-like receptors 1, 5, and 10. *Genomics* 90:502–515.
29. Seabury CM, Womack JE (2008) Analysis of sequence variability and protein domain architectures for bovine peptidoglycan recognition protein 1 and Toll-like receptors 2 and 6. *Genomics* 92:235–245.
30. Kühn Ch, et al. (2003) Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J Dairy Sci* 86:360–368.
31. Heyen DW, et al. (1999) A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol Genomics* 1:165–175.
32. Van Tassell CP, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5:247–252.
33. White SN, Taylor KH, Abbey CA, Gill CA, Womack JE (2003) Haplotype variation in bovine Toll-like receptor 4 and computational prediction of a positively selected ligand-binding domain. *Proc Natl Acad Sci USA* 100:10364–10369.
34. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
35. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.
36. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
37. Gabriel SB, et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229.
38. Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213–2233.
39. Crawford DC, et al. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 36:700–706.
40. Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci USA* 93:5131–5135.
41. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894–3900.
42. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073–1081.
43. Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7:61–80.
44. Hughes AL, et al. (2003) Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc Natl Acad Sci USA* 100:15754–15757.
45. Subramanian S, Kumar S (2006) Higher intensity of purifying selection on >90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol Biol Evol* 23:2283–2287.
46. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
47. Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
48. Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.
49. Villa-Angulo R, et al. (2009) High-resolution haplotype block structure in the cattle genome. *BMC Genet* 10:19.
50. Barreiro LB, et al. (2009) Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet* 5:e1000562.
51. Ramírez-Soriano A, Nielsen R (2009) Correcting estimators of θ and Tajima's *D* for ascertainment biases caused by the single-nucleotide polymorphism discovery process. *Genetics* 181:701–710.
52. Roach JC, et al. (2005) The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* 102:9577–9582.
53. Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP (2009) Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci USA* 106:7073–7078.
54. Seabury CM, Honeycutt RL, Rooney AP, Halbert ND, Derr JN (2004) Prion protein gene (*PRNP*) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. *Proc Natl Acad Sci USA* 101:15142–15147.
55. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169.
56. Marchini J, et al.; International HapMap Consortium (2006) A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet* 78:437–450.
57. Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 16:37–45.
58. Rozas J (2009) DNA sequence polymorphism analysis using DSP. *Methods Mol Biol* 537:337–350.
59. Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genet Res* 50:245–250.