DIVERSITY AND EVOLUTION OF THE BOVINE AND EQUINE TOLL-LIKE

RECEPTOR GENE FAMILY: APPLICATIONS TO ANIMAL DISEASE

A Thesis

by

COLLEEN ANN FISHER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Christopher M. Seabury
Committee Members,	Michael Criscitiello
	William Murphy
Head of Department,	Linda Logan

December 2012

Major Subject: Biomedical Science

ABSTRACT

Genes modulating innate immunity in mammals are generally considered the first line of defense with respect to invading pathogens and therefore it has become important to characterize naturally occurring genetic variation, and subsequently determine whether this variation is likely to be benign, beneficial, or detrimental to the host. Relevant to this study, the mammalian Toll-like receptor proteins (TLR), encoded by members of the *TLR* gene family, have the capacity to recognize a wide variety of pathogen ligands, and mutations within these genes have been shown to influence disease susceptibility or resistance within mammalian species.

Two studies which sought to determine the frequency and distribution of naturally occurring genetic variation within the bovine and equine *TLR* genes revealed a large number of discrete point mutations, which were subsequently used to reconstruct haplotypes for each investigated gene across a large number of samples. Detailed analyses of haplotypes provided evidence for extensive haplotype sharing among specialized breeds, subspecies, and even divergent species. Classical and new tests of selection provided evidence for significant deviations from a strictly neutral model of molecular evolution for both cattle as well as equids, with some of the same *TLR* genes deviating from a strictly neutral model among divergent species. As a first step toward determining whether naturally occurring bovine *TLR* variation is likely to be benign, beneficial, or detrimental, we tested validated variation from bovine *TLR* genes capable

of recognizing components of Mycobacteria for associations with *Mycobacterium avium* subspecies *paratuberculosis* (MAP) infection in dairy cattle, and found several SNPs that were nominally associated with disease status, thereby providing evidence for small-effect loci potentially influencing risk for differential susceptibility to Johne's disease.

DEDICATION

To my parents, for all they have done for me.

NOMENCLATURE

TLR	Toll-like Receptor
SNP	Single Nucleotide Polymorphism
LRR	Leucine-rich Repeat
TIR	Toll/Interleukin-1 Receptor
IBK	Infectious Bovine Keratoconjunctivitis
PAMP	Pathogen Associated Molecular Pattern
IL-1R	Interleukin-1 Receptor
TIR/IL-1R	Tol/IL-1 Receptor
MAP	*Mycobacterium avium* subspecies *paratuberculosis*
QTL	Quantitative Trait Loci
Indel	Insertion-deletion Mutation
AF	Allele Frequency
MAF	Minor Allele Frequency
LD	Linkage Disequilibrium
AA	Amino Acid
FDR	False Discovery Rate
GWAS	Genome Wide Association Study
SD	Standard Deviation
HWE	Hardy-Weinberg Equilibrium
MJ	Median Joining

PAMP   Pathogen Associated Molecular Pattern

TMNRR   Theoretical Minimum Number of Reads Required

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Members of the *Toll-like receptor* (*TLR*) gene family occupy key roles in the innate immune system by functioning as sentries for the detection and elimination of invading pathogens without requiring prior exposure [1, 2]. Historically, while studying *Drosophila* development, it was discovered that TOLL, a fly protein governing developmental polarity, was also instrumental in the elicitation of an effective antifungal immune response in adult flies [3-6]. Further study revealed that the TOLL gene was actively involved in differential susceptibility to fungal infections, and was but one member of a gene family that served to detect and catalyze the elimination of foreign pathogenic invaders [3-6]. Studies in vertebrate animals showed similar receptors, appropriately named the Toll-like receptors, which maintained functionality in the capacity of host innate immunity [7]. Mammals are generally considered to posses 10 or 12 functional *TLR* genes (*TLRs* 1-10 in human; *Tlr*s 1-9 and 11-13 in mouse) [5-8].

It is important to note that naturally occurring genetic variation within the mammalian *TLR* genes has been associated with differential disease susceptibility to a wide range of infectious diseases, protozoan parasites, and severe inflammatory responses (for review see [9-12]). One recent study has shown that 3 human *TLR5* single nucleotide polymorphisms (SNPs, 2 missense, 1 nonsense), located within the leucine-rich repeat (LRR) and/or toll/interleukin-1 receptor (TIR) domains, effectively abolish ligand

induced signaling [13]; one SNP of the three has been significantly associated with Legionnaire's disease in humans [1]. A study using mice treated with CpG DNA (or synthetic CpG oligos) to stimulate one or more *TLR* loci demonstrated that treated mice were resistant to infections from high doses of bacteria and viruses (for review see [1]). With current evidence displaying *TLR* variants as the cause of differential susceptibility to infectious bovine keratoconjunctivitis (IBK) [14] the need for comprehensive *TLR* studies in economically important animals is highlighted. Further understanding of the natural variation in the *TLRs* and the impact of these variants on disease susceptibility will allow for expanded development and research into the use of these receptors as both whole genome selection mechanisms and innate immunologicals [1].

.

CHAPTER II

EVOLUTION OF THE BOVINE *TLR* GENE FAMILY AND MEMBER ASSOCIATION WITH *MYCOBACTERIUM AVIUM* SUBSPECIES *PARATUBERCULOSIS* INFECTION

Introduction

The ultimate goal of bovine genomics is the identification of genetic variation that modulates corresponding variation in economically important production traits, differential susceptibility to disease, and favorable host response to vaccines, which is expected to enable the improvement of these phenotypes via informed genomic selection (for review see [15]). The bovine genome sequence and first-generation HapMap projects [16, 17] have directly enabled genome-assisted selective breeding [15], nascent investigations of non-traditional traits such as marker-assisted vaccination (as diagnostics for enhanced vaccine design or animal response), the development of a new class of anti-infectives known as innate immunologicals [1], and the elucidation of loci that have evolved under strong selection, thus providing important computational evidence for genomic regions which may underlie economically important traits.

Relevant to the suppression of infectious diseases, the mammalian innate immune system provides host defense against a variety of pathogens without requiring prior exposure [2, 7]. Consequently, genes that modulate innate immunity have often been

considered as candidate loci for improving host resistance to disease in agricultural species [14, 18-20]. Among mammals, the Toll-like receptor genes (*TLR*s) facilitate host recognition of pathogen associated molecular patterns (PAMPs), thereafter eliciting host innate immune responses [2, 8] aimed at suppressing invading bacteria, viruses, protozoa, and fungi. Essential to their role in host defense, the mammalian *TLR*s encode type I transmembrane proteins of the Interleukin-1 receptor (IL-1R) family with N-terminal leucine-rich repeats (LRR) involved in ligand recognition, a transmembrane domain, and a C-terminal intracellular Toll/IL-1 receptor homologous (TIR/IL-1R) domain for signal transduction [2, 7, 21]. The mammalian *TLR* genes are primarily expressed by antigen-presenting cells (i.e., macrophages or dendritic cells), and most of the TLR ligand specificities have been experimentally elucidated, with six gene family members (*TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, *TLR9*) known to recognize microbial (bacteria, fungi, protozoa) and/or synthetic ligands, and five (*TLR3*, *TLR4*, *TLR7-TLR9*) known to recognize viral components [8, 21]. Presently, *TLR10* remains the only functional human *TLR* gene family member for which natural and/or synthetic ligands have not been fully elucidated [22]. However, given evidence for functional mammalian *TLR* protein heterodimers (*TLR10/TLR1*; *TLR2/TLR10*) [22], the host protein encoded by *TLR10* may collaboratively enable recognition of a diverse array of microbial PAMPs, including those recognized by *TLR2* [22-25].

Several studies have demonstrated that some naturally occurring *TLR* variants enhance the risk of severe infections in humans, mice, and domestic cattle, including the potential

4

for increased susceptibility to Johne's disease, a debilitating and economically important disease of ruminants caused by infection with *Mycobacterium avium* spp. *paratuberculosis* (MAP) (for review see [10, 13, 26-29]). Furthermore, several important bovine health-related quantitative trait loci (QTL) have also been localized to genomic regions either proximal to or directly overlapping one or more *TLR* loci (for review see [19, 30-34]). Therefore, we utilized massively parallel pyrosequencing of a pooled *TLR* amplicon library (*TLRs* 1-10) to comprehensively evaluate nucleotide variation and haplotype structure for 31 cattle breeds representing *Bos taurus taurus*, *Bos taurus indicus*, and their subspecific hybrids (composites). Overall, 276 single nucleotide polymorphisms (SNPs) and 4 insertion-deletion (indel) mutations were detected and validated. Bovine *TLR* SNPs and indels leveraged from the pyrosequencing study were used in a case-control analysis to identify risk factors underlying differential susceptibility to MAP in U.S. dairy cattle. In addition, we also comprehensively report on bovine *TLR* haplotype structure, the extent of haplotype sharing among specialized breeds and subspecific lineages, and provide median joining networks as putative representations of bovine *TLR* haplotype evolution [35]. Finally, we provide computational evidence for several bovine *TLR* genes evolving under disparate modes of non-neutral evolution, thereby underscoring their potential importance to bovine innate immunity and health-related traits. The results of this study will enable bovine translational genomics, QTL refinement, and ultimately, genome-assisted methods for animal selection to develop cattle populations with enhanced disease resistance and favorable vaccine response.

Results

Bovine TLR pyrosequencing, SNP detection, variant validation, and haplotype inference

For 96 elite bovine sires representing 31 domestic cattle breeds (*B. t. taurus*; *B. t. indicus*; and composites), we generated and purified 81 amplicons targeting all 10 bovine *TLR* genes (n= 7,776 total amplicon targets; see methods). The majority of the amplicons were pooled (n= 6,816) to form a normalized fragment library (Table A1, Figure A1) which was subjected to a workflow involving Roche 454 Titanium pyrosequencing with downstream variant detection using the Neighborhood Quality Standard algorithm as recently described [36], and the remaining purified amplicons (n= 960) were analyzed by standard dye-terminator cycle sequencing (Sanger) with alignment-based variant detection [30-32]. Sanger sequencing was necessary for amplicons that were intolerant to the addition of 59 oligonucleotide barcodes for PCR amplification. In total, 474 variable sites were predicted from intragenic analyses of all sequence data, which included 212 previously validated SNPs [37], 4 known indels [37], and 258 new putative SNPs. Evaluation of the genic distributions of all newly predicted *TLR* variable sites detected within the pyrosequencing data revealed that $\geq 62\%$ of the 258 new putative SNPs were located either within or immediately flanking homopolymer repeats. Nevertheless, to allow for inclusion of all possible SNPs in downstream analyses, we investigated all 474 variable sites via fluorescent allele-specific genotyping assays [37]. Collectively, we validated 280 biallelic *TLR* variants (276 SNPs + 4 indels; Table A2) using custom genotyping assays applied to the

6

sequencing discovery panel (n= 96 elite sires; 31 breeds), a panel of Holstein dairy cattle (n= 405; 3 herds), and a panel of purebred Angus beef cattle from a single herd (n= 48).

Of the 276 validated SNPs, 71 were predicted to encode nonsynonymous substitutions (nsSNPs), and one was predicted to encode a nonsense mutation in bovine *TLR5* (AA substitution R125*; SNP C2332T). For the validated SNPs detected via pyrosequencing (n= 244), we investigated the relationship between minor allele frequencies (MAFs) estimated from the analysis of pyrosequencing data, as compared to corresponding allele frequencies derived from individual fluorescent allele-specific genotyping assays, and found significant correlations across all 10 *TLR* genes (discovery panel; Table 1). Moreover, an analysis performed across all genes (n= 244 SNPs) revealed that there was little or no bias in the estimates of allele frequencies produced via targeted pyrosequencing (P = 0.999846; Ho: slope =1; Figure 1).

Collectively, 266 SNPs and 4 indels were successfully incorporated into 243 unique haplotypes via Bayesian reconstructions [37, 38] (Table 2), which included one discrete haplotype carrying the putative *TLR5* nonsense SNP. Ten SNPs (*TLR2*: 9431, 10047, 12121; *TLR3*: 3624, 3804, 5201, 6382; *TLR4*: 8166; *TLR5*: 1562, 1685; see Table A2) could not be incorporated into discrete haplotypes with best-pair phase probabilities ≥ 0.90. Summary data representing the total number of predicted haplotypes, number of cattle with phase probabilities ≥ 0.90, total number of variable sites with MAF ≤ 0.10, genic distributions of validated variable sites, size of the investigated

Table 1. Relationship between minor allele frequencies estimated from pyrosequencing and allele-specific genotyping of 96 individuals from 31 breeds.

Bovine Gene	Total 454 SNPs[a]	Overall Correlation (r)[b]	Overall RSQ (r^2)[c]
TLR1	4	0.998	0.996
TLR2	44	0.935	0.874
TLR3	39	0.958	0.918
TLR4	28	0.948	0.898
TLR5	39	0.942	0.887
TLR6	15	0.879	0.773
TLR7	15	0.959	0.920
TLR8	13	0.877	0.769
TLR9	22	0.975	0.950
TLR10	25	0.749	0.561
Totals/Avg	**244**	**0.922**	**0.855**

[a] Total SNPs detected via pyrosequencing
[b] $P < 0.05$ for all *TLR* genes
[c] RSQ is the squared correlation coefficient (r^2)
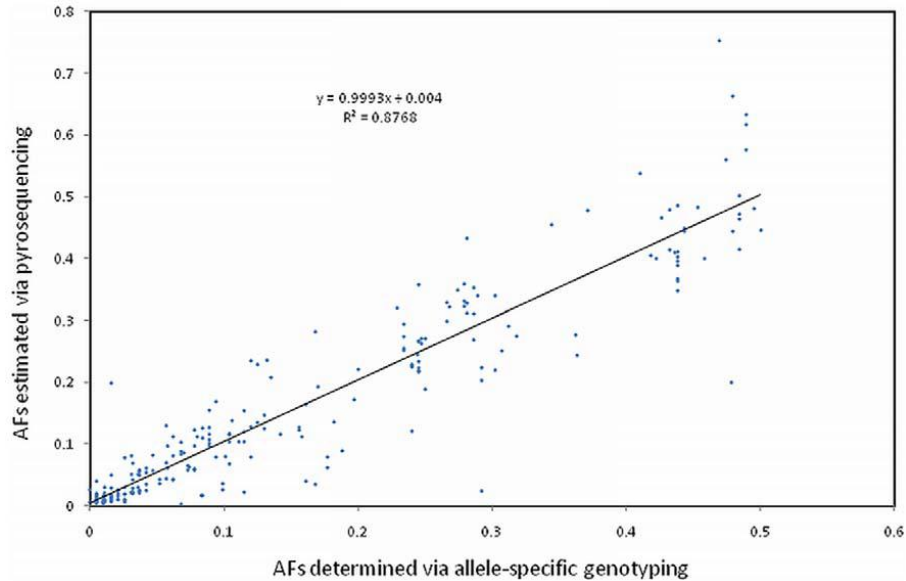
Figure 1. **For validated bovine *TLR* SNPs detected via pyrosequencing (n = 244), a regression analysis was performed for pyrosequencing allele frequency (AF) estimates corresponding to the true minor alleles (< 0.5), as defined by allele-specific genotyping assays, and minor AFs (MAFs) directly ascertained by genotyping (n = 96 elite sires; 31 breeds).** The true minor alleles (< 0.5) were correctly identified for 236/244 (97%) SNPs via pyrosequencing. This analysis provided strong statistical evidence (P = 0.999846; Ho: slope = 1) for little or no bias in the pyrosequencing-based estimates of allele frequency.

Table 2. Summary data for validated polymorphisms detected in 10 bovine innate immune genes

Bovine Gene	BTA Assign[a]	Total Haps[b]	Sires Phased[c]	MAF ≤ 0.10[d]	Avg r² all[e]	Avg r² *B.t.t.*[e]	Valid. SNPs[f]	Hap SNPs[g]	Valid. Indels[h]	Valid. nsSNPs[i]	Region Size[j] (Kb)	QTL or Assoc.[k]
TLR1	BTA6	8	547	3	0.24	0.49	5	5	0	2	2.184	Q
TLR2	BTA17	38	532	38	0.19	0.24	44	41	1	20	3.224	Q, A[l]
TLR3	BTA27	40	78	20	0.29	0.57	56	52	0	3	9.469	A
TLR4	BTA8	29	532	23	0.10	0.08	28	27	0	7	3.470	Q, A
TLR5	BTA16	30	526	29	0.20	0.18	43	41	3	9	5.334	No
TLR6	BTA6	20	526	13	0.09	0.12	15	15	0	6	2.327	Q, A[l]
TLR7	BTAX	9	96	7	0.28	0.28	15	15	0	1	4.285	Q
TLR8	BTAX	6	96	1	0.70	0.69	13	13	0	8	3.702	Q
TLR9	BTA22	20	545	9	0.27	0.29	22	22	0	3	5.033	Q
TLR10	BTA6	43	524	34	0.27	0.15	35	35	0	13	3.859	Q[l]
Total/Avg		**243**	**96%**	**177**	**0.26**	**0.31**	**276**	**266**	**4**	**72**	**42.887**	

[a]BTA assignments based on NCBI Refseq (Btau5.2).

[b]Total number of haplotypes predicted from all validated markers and best pair reconstructions [28] with probabilities ≥ 0.90.

[c]Number of cattle exhibiting best pair phase probabilities ≥ 0.90. BTAX haplotypes were directly observed. 96 animals were genotyped for *TLR3*, *TLR7* and *TLR8*, for all other loci 549 animals were genotyped.

[d]Number of polymorphisms with minor allele frequencies ≤ 0.10.

[e]Average intragenic linkage disequilibrium (r²) values estimated for adjacent SNP and indel sites for all cattle or for *B. t. taurus* (*B.t.t.*).

[f]Number of putative SNPs validated as polymorphic.

[g]Number of validated SNPs incorporated in discrete haplotypes.

[h]Number of putative indels validated as polymorphic.

[i]Number of putative nonsynonymous SNPs validated as polymorphic, including the *TLR5* nonsense SNP.

[j]Size of the genic region. 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 [26-34, 53].

[l]Tentative association in this study

regions, and average estimates of linkage disequilibrium (LD; r^2) between adjacent variable sites are depicted in Table 2. Across all investigated loci (n =549 cattle; 31 breeds), the MAF spectrum derived from allele specific genotyping assays ranged from 0.001 to 0.498, with 64% of the validated SNPs possessing MAFs ≤ 0.10 (Table 2).

Characterization of LD architecture, recombination, and intragenic tagSNPs/Indels
Evaluation of the intragenic patterns of LD across all 31 breeds of cattle via 95% confidence intervals constructed for D' [39, 40], application of the four gamete rule [39], and estimates of recombination between adjacent variable sites [41, 42] revealed one or more blocks of strong LD within each of the 10 bovine *TLR* genes. Statistical evidence for historical recombination was detected within *TLR2*, *TLR3*, and *TLR6*, resulting in at least two detectable LD blocks within each gene. All other genes exhibited a single block of strong LD spanning either all, or the majority of all validated intragenic SNPs and indels, as supported by the majority rule of all three analyses [39-42]. A comparison of average intragenic r^2 values calculated between adjacent variable sites across all 10 genes revealed a dynamic range of LD (0.09-0.70; all cattle, 31 breeds; Table 2). Discrete regions of high and low LD, the latter due to historical recombination, were also detected using the general model for varying recombination rate [38, 41, 42]. Cumulatively, four adjacent SNP sites [*TLR2* (1), *TLR3* (2), and *TLR6* (1)] produced estimates of median recombination rates that exceeded the background rate ($\bar{\rho}$; [38, 41, 42]) by a factor of at least 2.5. The highest median estimate of recombination

rate was observed in *TLR3* (between SNP positions rs42851925, rs55617222; rs55617241, rs55617451, Table A2), and exceeded the background rate by a factor of at least 5.2. Analyses to identify tagSNPs/Indels which predictively captured 100% of the variation at 280 validated variable sites within all 10 genes for all cattle yielded 160 tagSNPs and 2 tagIndels (Table A3). Similar analyses restricted to the *B. t. taurus* breeds demonstrated that only 118 tagSNPs and 1 tagIndel were predicted to capture 100% of the variation at 235 variable sites (Table A3). Interestingly, the cumulative tagging efficiency (total tags predicted/total number of validated variable sites) was similar for both analyses (all cattle vs. *B. t. taurus*), with this result largely due to the preponderance of taurine cattle in the total sample (94.4%), and the significant sharing of SNPs, indels, and haplotypes among the subspecific lineages.

High resolution bovine TLR haplotype networks and breed distributions

Median joining haplotype networks (Figures 2, 3, 4, Figures A2-A10; Table A4) constructed for all 10 genes revealed that: 1) The specialized *B. t. taurus* beef and dairy breeds cannot be genetically discriminated despite an average polymorphism density (266 SNPs+ 4 indels; see Table 2) of one variable marker per 158 bp; 2) Haplotype sharing occurs among *B. t. taurus* and *B. t. indicus* breeds at all 10 loci; 3) Shared haplotypes were often the highest frequency haplotype(s) within a network; 4) Despite haplotype sharing between the subspecific lineages, the 250 Kyr divergence [43] between *B. t. taurus* and *B. t. indicus* was evident in most, but not all, haplotype networks (i.e., *TLR1-7, TLR10*). With very few exceptions (i.e., *TLR3* Network 1, *TLR4*,

11

TLR10), the high frequency network nodes demonstrating subspecific haplotype sharing often included at least two indicine sires. Using summary data derived from the median joining networks (Table A4), we estimated the relationship between the total number of discrete *TLR* haplotypes predicted (*TLR1-10*) in seven major U.S. taurine beef breeds [44] (Angus, Charolais, Gelbvieh, Hereford, Limousin, Red Angus, Simmental), and four U.S. taurine dairy breeds (Braunvieh, Brown Swiss, Holstein, Shorthorn), and found a significant correlation (r = 0.71, P ≤ 0.0224). This correlation was driven by the large number of haplotypes predicted to be shared among the beef and dairy breeds. For the investigated beef breeds, we predicted 84 discrete haplotypes across all 10 *TLR* loci, and at least 60 (71.4%) were predicted to be shared with the four dairy breeds. However, we also detected disparities between the numbers of haplotypes predicted for *TLR4* and *TLR5*, with the dairy breeds possessing 3.8X and 2.3X more discrete haplotypes for these loci, respectively, than did our beef cattle. Exclusion of these two outlying loci resulted in a nearly perfect correlation (r = 0.98, P < 0.0001) between the numbers of discrete haplotypes predicted in beef and dairy breeds across the remaining *TLR* loci. Interestingly, the single haplotype possessing the *TLR5* putative nonsense mutation was almost exclusively predicted in Holstein cattle (Figure A1, *TLR5* Node Q; n = 53 Holstein, n= 1 Braford).
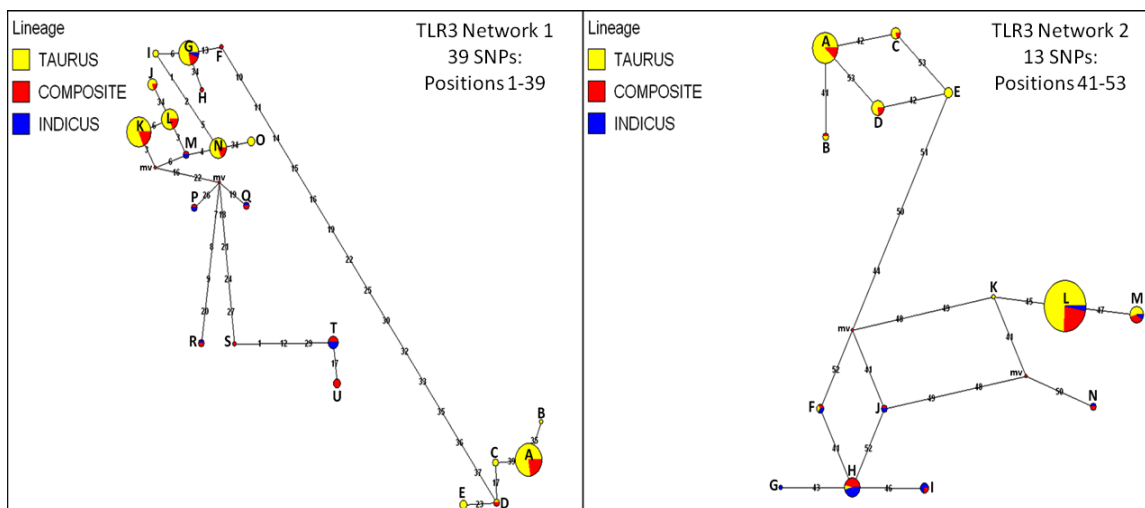
Figure 2. Median joining (MJ) haplotype networks for bovine *TLR3* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds). Because MJ networks require the absence of recombination [73], each network represents intragenic regions of elevated LD. Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
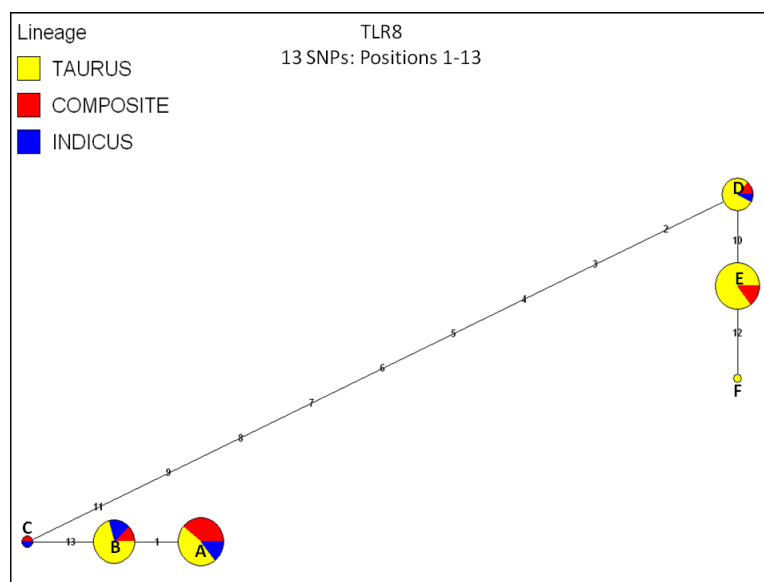
Figure 3. Median joining (MJ) haplotype network for bovine *TLR8* using haplotypes directly ascertained for all cattle (n = 96 AI sires, 31 breeds). Haplotypes observed for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4).

13
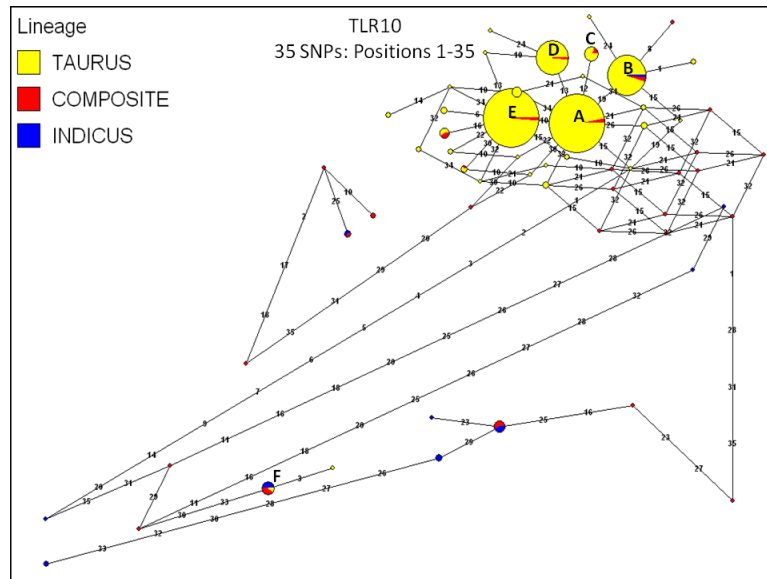
Figure 4. Median joining (MJ) haplotype network for bovine *TLR10* using haplotypes directly ascertained for all cattle (n = 96 AI sires, 31 breeds). Haplotypes observed for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4).

Functional modeling of bovine amino acid (AA) substitutions and tests of selection

Using both PolyPhen [45] and SIFT [46] to evaluate the putative functional effects of AA substitutions encoded by *TLR* SNPs, we determined that 54/72 (75%) of AA substitutions were predicted to be benign and tolerated, whereas 23/72 (32%) were predicted to impact protein function [47] by at least one of the analytical methods employed (Table 3). For those mutations predicted to impact protein function, 18/23 (78%) were detected at frequencies < 0.05, and 5/23 (22%) located in *TLR2* (1), *TLR3* (2), *TLR5* (1; putative nonsense SNP), and *TLR8* (1) were observed at frequencies ≥ 0.05, with moderate frequency substitutions detected in *TLR8* (0.562) and *TLR3* (0.432; see Table 3). The MAF for the *TLR5* putative nonsense SNP, as estimated from 405

14

Holsteins in three herds was 0.068 (Table 3). Across all polymorphisms involving AA substitutions, PolyPhen and SIFT produced analogous predictions for 61/72 (85%) observed replacements. To collectively estimate the extent of functional and/or selective constraint(s) related to bovine *TLR* protein function, we used a goodness of fit test to examine disparities between the observed distributions of AA phenotypes (PolyPhen + SIFT results; benign/tolerated vs. damaging/affect). Assuming equal probabilities for the occurrence of both classes of AA phenotypes across all bovine *TLRs*, we found there to be significantly fewer substitutions predicted to impact protein function than those classified as benign or tolerated (P = 0.00022). This is consistent with some degree of functional and/or selective constraints that generally operate to maintain the functional products of most protein coding genes [47-49]. However, this result describes a general trend across the bovine *TLR* gene family, and does not provide locus-specific insights regarding the evolutionary origin and magnitude of these constraints.

To elucidate gene-specific departures from a strictly neutral model of molecular evolution, we used Tajima's frequency distribution test (D statistic) [50], as applied to the discovery panel samples (all cattle from 31 breeds vs. *B. t. taurus*), and evaluated the significance of the observed values (D) via coalescent simulation (Table 4). Departures from neutrality were detected for *TLR3*, *TLR8*, and *TLR10*. However, the direction of the deviation was not uniform across all three loci (Table 4), suggesting that disparate modes of evolution (i.e., selection) may have influenced genetic diversity within these genes, and that there may be differences among cattle lineages (Table 4, *TLR10*).

Table 3. Summary data for 23 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]
TLR2	G > T	*ss470256478*	NP_776622.1	W119L	LRR_TYP1	PrD	AF	0.008
	T > A	*rs68268251*	NP_776622.1	F227L	NCP	PsD	T	0.015
	C > T	*ss470256481*	NP_776622.1	T311M	NCP	PrD	AF	0.006
	C > T	*ss470256483*	NP_776622.1	S485F	LRR_TYP2	PrD	AF	0.015
	G > A	*rs68268260*	NP_776622.1	R563H	LRRCT	B	AF	0.066
	G > C	*ss470256484*	NP_776622.1	E738Q	TIR	PsD	AF	0.001
TLR3	G > A	*rs55617272*	NP_001008664.1	G426S	LRR8	PsD	AF	0.058
	G > T	*rs42852439*	NP_001008664.1	S664I	LRRCT	PsD	T	0.432
TLR4	A > C	*rs8193049*	NP_776623.5	N151T	LRR3	PsD	T	0.009
	A > G	*rs8193055*	NP_776623.5	K381R	LRR6	B	AF	0.005
	A > G	*ss469376075*	NP_776623.5	H587R	LRRCT	PrD	AF	0.003
TLR5	C > T	*ss469376099*	NP_001035591.1	R125*	NCP	PsD		0.053[f]
	G > A	*ss469376101*	NP_001035591.1	R262H	NCP	PrD	T	0.004
	C > G	*ss469376107*	NP_001035591.1	F643L	NCP	B	AF	0.003
TLR6	T > G	*rs68268270*	NP_001001159.1	L43R	NCP	PrD	AF	0.003
	A > G	*rs68268272*	NP_001001159.1	R87G	LRR1	B	AF	0.017
	T > A	*ss469376113*	NP_001001159.1	F494I	LRR5	PrD	AF	0.024
TLR7	A > G	*ss469376123*	NP_001028933.1	N439S	NCP	PrD	AF	0.021
TLR8	G > A	*rs55617351*	ABQ52584.1	S477N	NCP	B	AF	0.562
	A > C	*ss469376137*	ABQ52584.1	K903T	TIR	PsD	AF	0.010
TLR10	G > A	*rs55617437*	NP_001070386.1	R18H	SigPep	PsD	T	0.018
	C > G	*rs55617286*	NP_001070386.1	I134M	LRR3	B	AF	0.013
	A > C	*rs55617297*	NP_001070386.1	K753T	TIR	PsD	AF	0.010

[a]SNPs with "rs" numbers were previously described [30-32, 37, 66] and validated in this study.
[b]Amino acid (AA) substitutions predicted from corresponding SNPs, GenBank Proteins, and previous studies [30-32, 37, 66].
[c]Protein domain locations predicted by SMART (http://smart.embl-heidelberg.de/). Only confidently predicted domains are depicted (NCP = no confident prediction; LRRs are named in order of prediction).
[d]Results from PolyPhen and SIFT [45, 46]. Results other than "Benign (B)" or "Tolerated (T)" are predicted to be Possibly Damaging (PsD), Probably Damaging (PrD), or Affect Protein Function (AF).
[e]Observed frequency of nonsynonymous SNP allele in all 31 cattle breeds.
[f]The frequency of this SNP in U.S. dairy cattle (n = 405, 3 Herds) was 0.069.

Table 4. Summary data for tests of selection across all members of the bovine *TLR* gene family

Gene	Sires Phased[a]	Tajima's *D* all[b]	Coalescent *P*-value[c]	Sires Phased[a]	Tajima's *D taurus*[b]	Coalescent *P*-value[c]
TLR1	95 (99%)	0.55535	*P* > 0.05	64 (98%)	1.49328	*P* > 0.05
TLR2	92 (96%)	0.51385	*P* > 0.05	64 (98%)	-0.06547	*P* > 0.05
TLR3	78 (81%)	**2.35965**	***P* < 0.03**	54 (83%)	**3.63792**	***P* < 0.001**[e, f]
TLR3-1[d]	83 (86%)	**2.12744**	***P* < 0.04**	59 (91%)	**3.59176**	***P* < 0.001**[e, f]
TLR3-2[d]	94 (98%)	**2.07897**	***P* < 0.05**	63 (97%)	**2.65634**	***P* < 0.02**
TLR4	89 (93%)	-0.83191	*P* > 0.05	64 (98%)	0.93683	*P* > 0.05
TLR5	86 (90%)	0.69344	*P* > 0.05	59 (91%)	0.44166	*P* > 0.05
TLR6	91 (95%)	0.16727	*P* > 0.05	65 (100%)	-0.71248	*P* > 0.05
TLR7	96 (100%)	-0.19828	*P* > 0.05	65 (100%)	-1.70370	*P* > 0.05
TLR8	96 (100%)	**3.53957**	***P* < 0.001**[e]	65 (100%)	**3.28763**	***P* < 0.001**[e]
TLR9	95 (99%)	1.15800	*P* > 0.05	64 (98%)	1.26794	*P* > 0.05
TLR10	92 (96%)	-0.29809	*P* > 0.05	61 (94%)	**-1.78285**	***P* < 0.03**

[a] Number and proportion of cattle from the sequencing discovery panel with best-pair phase probabilities ≥ 0.90 for all cattle (n = 96), and for *B. t. taurus* cattle (n = 65).
[b] Tajima's *D* statistic [50] for all cattle and for *B. t. taurus* breeds.
[c] Significance levels were estimated by coalescent simulation using 10,000 replicates [73]. All bolded loci were also significant (*P* < 0.05) via application of the beta distribution [73].
[d] Phased variation within *TLR3* Network 1 and *TLR3* Network 2.
[e] Significant after correction for multiple tests (α / n locus-specific tests; α = 0.05).
[f] Significant after adding in the best-pairs of haplotypes for taurine sires with probabilities ≤ 0.90 and correction for multiple testing (α = 0.05).

For both *TLR3* and *TLR8*, a significantly positive Tajima's D reflected an excess of moderate frequency alleles, whereas a large negative value for *TLR10* (*B. t. taurus*) reflected an overabundance of rare, low frequency variants consistent with purifying selection [37]. Therefore, it is important to note that although a significant nonrandom trend toward benign or tolerated AA substitutions was detected across all investigated loci, the underlying reason for this functional and/or selective constraint appears to be fundamentally different between some gene family members (i.e., *TLR3*, *TLR8* vs. *TLR10*). Notably, we observed at least one moderate frequency AA substitution that was predicted to impact protein function in both *TLR3* and *TLR8* (Table 3), whereas all AA

substitutions predicted to impact protein function in *TLR10* were detected at very low frequencies (Table 3). To further investigate the overall magnitude and origin(s) of the most significant deviations from a strictly neutral model (Tajima's D; pyrosequencing discovery panel; Table 4), we used Fu's F_S statistic [51] to estimate the probability of observing a number of haplotypes less than or equal to that predicted in our samples for *TLR3* (*B. t. taurus*), *TLR3-1* (*B.t. taurus*), and *TLR8* (all cattle; *B. t. taurus*). For *TLR3*, we recognized that the inability to phase all individuals in the pyrosequencing discovery panel could lead to the absence of some low frequency alleles, thus potentially driving both Tajima's *D* and Fu's F_S toward larger positive values. Consequently, we calculated Fu's F_S and Tajima's *D* for *TLR3* (*B. t. taurus*) and *TLR3-1* (*B. t.taurus*) using the following approach: 1) Both test statistics were first calculated only for sires that could be phased with best-pairs probabilities $\geq$ 0.90, as depicted in Table 4; and 2) If a significant result was achieved in this analysis, we then added the taurine haplotypes with phase probabilities < 0.90 into our analyses (*D*; F_S) by choosing the best haplotype pairs reconstructed for each sire. For Fu's F_S, only *TLR8* displayed unequivocal evidence for a departure from neutrality (All cattle F_S = 10.2712, P < 0.01; *B. t. taurus* F_S =10.296, P < 0.01), with levels of significance that withstood conservative correction for multiple testing (correction = α/n locus-specific tests, 0.05/2= Minimal P $\leq$ 0.025). For Tajima's *D*, inclusion of the best *TLR3* haplotype pairs for sires with phase probabilities < 0.90 resulted in very similar test statistics (*TLR3 B. t. taurus D* = 3.6034, P < 0.001; *TLR3-1 B. t. taurus D* =3.4895, P < 0.002; Table 4), with levels of significance that endured correction for multiple testing (0.05/8= Minimal P $\leq$ 0.00625).

A regression-based approach considering all validated variable sites and the effective number of SNPs at each site [37] also demonstrated that *TLR3* and *TLR8* possess significantly more gene diversity than do the eight other *TLR* loci (P ≤ 0.05; Figure 5) in taurine and all cattle combined. In contrast, both regression analyses (all cattle; *B. t. taurus* only) indicated that *TLR10* and *TLR2* possess significantly less gene diversity than other members of the bovine *TLR* gene family (Figure 5). With the exception of *TLR2*, these results are precisely congruent with the results of Tajima's test (*D*; Table 4).

Single marker and haplotype association tests with MAP infection

Unphased diploid genotypes for a subset of the validated SNPs and indels (n =35; nonsynonymous, putative nonsense, 5'upstream regions, and introns) within bovine *TLR* genes either known or postulated to primarily recognize bacterial ligands (*TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, *TLR9*, *TLR10*) were tested for associations with bacterial culture status for MAP (fecal and/or tissue) in three Holstein dairy herds (n= 68 cases, 270 controls). All nonsynonymous *TLR* SNPs previously associated with MAP infection [24] (*TLR1*, *TLR2*, *TLR4*) were monomorphic in our samples (n= 549; 31 breeds). Conditional logistic regression models were constructed for each of 35 variable sites meeting our selection criteria (see methods) to estimate the relative odds of MAP infection given the defined diagnostic criteria adjusted for the effects of herd and age. Collectively, six SNPs produced suggestive associations, as evidenced by uncorrected P-values (Table 5). Interestingly, three SNPs in *TLR2* and one in *TLR6* were associated with increased odds of MAP infection in animals with 1 or more copies of the minor

allele (Table 5). Two SNP loci, 1 in *TLR4* and 1 in *TLR10*, were associated with decreased odds of infection given increasing copies of the minor allele (Table 5). Following locus-specific correction of the P-values using the false discovery rate (FDR) method (http://sdmproject.com/utilities/?show=FDR) [52], two SNPs (*TLR6*-rs43702941; *TLR10*-rs55617325) remained significant (P ≤ 0.05), and three SNPs (*TLR2*-rs68268245, ss470256479,rs43706433) displayed P-values (P ≤ 0.053) that were suggestive of a potential recessive genetic association with MAP infection (Table 5). Two of these SNPs (*TLR2*-ss470256479, rs43706433) were recently hypothesized to occur on a haplotype associated with an increased risk for Johne's disease [53]. Consequently, we used PHASE 2.1 [38] to test the hypothesis that haplotype frequencies for bacterial-sensing *TLRs* differ between cases and controls. However, none of the investigated loci possessed significantly different haplotype distributions between cases and controls (P > 0.05; 1,000 permutations).

Table 5. Summary statistics for single marker association tests with risk of Johne's disease.

Marker	Model	Odds Ratio	P Value[c]	95% Confidence Interval[a]	
				Lower Bound	Upper Bound
TLR2-SNP 9564	Recessive	3.20	0.032[d]	1.11	9.24
TLR2-SNP 10511	Recessive	3.21	0.031[d]	1.11	9.25
TLR2-SNP 10540	Recessive	2.51	0.020[d]	1.15	5.48
TLR4-SNP 9788	Additive	0.27[b]	0.026	0.09	0.86
TLR6-SNP 14578	Additive	2.58[b]	0.012[e]	1.23	5.43
TLR10-SNP 774	Additive	0.53[b]	0.041[e]	0.29	0.97

[a] 95% Confidence interval for odds ratio.
[b] Odds ratio adjusted for the effect of birth year.
[c] *P*-value not corrected for multiple comparisons.
[d] *P*-value marginal (0.053) after locus-specific correction.*TLR1*, *TLR6*, and *TLR10* were considered a single locus when correcting for multiple tests.
[e] *P*-value < 0.05 after locus-specific correction [52; http://sdmproject.com/utilities/?show=FDR].

Discussion

Our methodological workflows resulted in the robust identification of SNPs with precise estimates of MAF for the bovine *TLR* genes (see methods), as evidenced by the regression of MAFs derived from the analysis of pyrosequencing data and allele-specific genotyping assays (Figure 1). For these genes, our genotyping assays provide a 70 fold increase in marker density relative to the Illumina BovineSNP50 assay, which queries four SNPs either within (*TLR6*, *TLR10*) or proximal to (*TLR7*, *TLR8*) the targeted loci, and a greater than 3 fold increase in marker density relative to the new Illumina BovineHD assay (777K), which possesses an average marker interval density of approximately 1 SNP/3.5 kb. Notably, the new BovineHD assay includes 84 SNPs that are either within or proximal to ($\leq$ 2 Kb) the 10 *TLR* genes (i.e. *TLR1* [3]; *TLR2* [6]; *TLR3* [8]; *TLR4* [6]; *TLR5* [22]; *TLR6* [23]; *TLR7* [3]; *TLR8* [4]; *TLR9* [5]; *TLR10* [4]), including one SNP implicated by our case-control study (*TLR2*-rs43706433; Table 5). Validated polymorphisms, reconstructed haplotypes, and the tagSNPs/Indels identified in this study will directly facilitate the fine mapping of bovine health-related QTL [30-34], while also enabling further evaluation of SNPs tentatively associated with differential susceptibility to Johne's disease (MAP infection) [26-29, 53] (Table 5). While large numbers of tightly clustered SNPs are sometimes difficult to genotype, we endeavored to validate all detected variants by redesigning primers and manipulating PCR conditions for problematic markers. Accordingly, we successfully validated several SNPs for which assays had previously failed [37], and we also validated the majority of the newly identified putative SNPs (pyrosequencing data) that were not associated with

homopolymer repeats. Furthermore, some regions of *TLR1* posed the greatest technical challenge due to sequence similarity with *TLR6*. For this reason, at least some DNA sequencing from medium-range PCR products designed to specifically amplify each locus is needed to exhaustively ascertain all possible variants spanning the *TLR1-TLR6* gene cluster.

Across all adjacent variable sites within the bovine *TLR* gene family, we observed higher levels of LD (r^2) in *B. t. taurus* cattle (0.32) than in the combined sample (0.26) of *Bos t. taurus*, *Bos t. indicus*, and composite breeds (Table 2). This is generally consistent with previous studies of bovine subspecific divergence, haplotype structure, and LD across short to moderate physical distances [17, 54], including our previous study on bovine *TLR* haplotype structure [37]. However, in this study intragenic estimates of r^2 increased for several loci upon pooling (all cattle), including *TLR4*, *TLR8*, and *TLR10*, which was not predicted given previously reported trends in LD [17, 37, 54]. We previously found that r^2 values were enhanced after pooling only for *TLR7* and *TLR8* [37]. This result indicates that phase-relationships have been preserved across bovine subspecies and specialized breeds for these loci, perhaps due to selection (Table 4), and is only apparent at high genotyping densities. Moreover, this observation may represent a signature of selection on some individual variable sites, with detectable levels of intragenic selection only becoming apparent (Table 4) with increasing numbers of variable sites subject to selection, and/or uniformly higher selection coefficients. For all genes except *TLR2* (Network 1 only), *TLR3* (Network 1 only), *TLR5*, *TLR8*, and *TLR9*, one or two

predominant haplotypes were predicted for the majority of the cattle investigated (Figures 2-4, Figure A1; Table A4). Moreover, significantly positive values for Tajima's D were detected for genomic regions encoding *TLR3* and *TLR8* (Table 4) despite correction for multiple testing, and for *TLR3*, the addition of best haplotype pairs for sires with phase probabilities < 0.90 produced very similar test statistics (D) for *B. t. taurus* cattle, indicating that D is not falsely inflated by the absence of rare alleles within the sires that could not be stringently phased. Additionally, a regression based test also demonstrated that *TLR3* and *TLR8* possess significantly more diversity than do all other *TLR* loci (P $\leq$ 0.05; Figure 5). Significantly positive values for Tajima's D are often interpreted as evidence for a recent population bottleneck, or for some form of balancing selection [55-57], with D being the most powerful test in its class [58], but may also indicate violations of the mutation-drift equilibrium assumption or random sample requirement. Worthy of discussion is the fact that variation within *TLR3* displayed the second highest average r^2 values between adjacent variable sites (Table 2), which in conjunction with a large, significantly positive D statistic for taurine cattle (Table 4) suggests that this gene is under selection. However, unlike *TLR8*, high r^2 ($\geq$ 0.50 for 10/13 SNPs in *TLR8*) did not persist across the majority of all adjacent variable sites in *TLR3*, and therefore, it is relatively unsurprising that our analysis of *TLR3* revealed no evidence for a deficiency of total discrete haplotypes in *B. t. taurus* cattle (i.e., F_S was not significant).
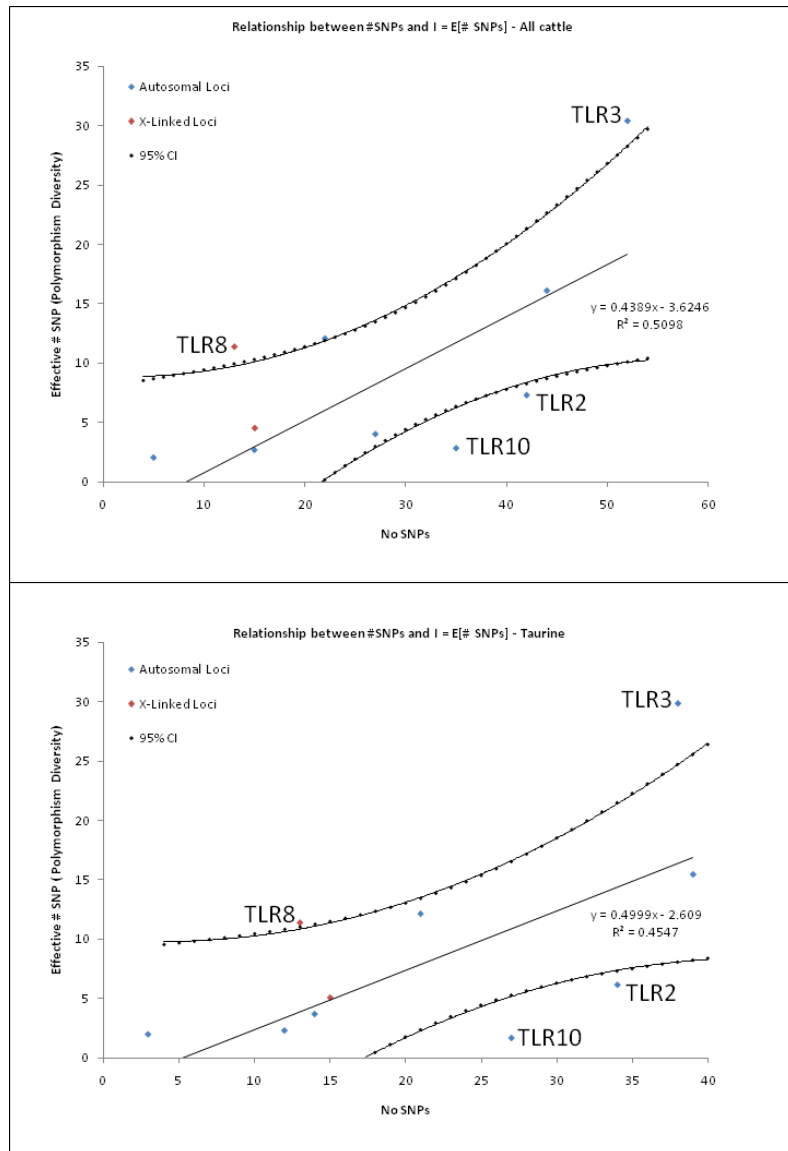
Figure 5. Relationship between the number of validated SNPs and SNP diversity here denoted as the effective number of SNPs across all 10 *TLR* loci in A) all cattle, and B) taurine cattle. The linear regressions and estimated 95% confidence intervals are shown in each panel.

Surprisingly, the region of *TLR3* demonstrating the strongest deviation from neutrality does not include the two nonsynonymous SNPs predicted to impact protein function (Table 3, Table 4), but includes a 59 bp putative promoter region (PROSCAN 1.7: http://www-bimas.cit.nih.gov/molbio/proscan/index.html) [30] harboring several

transcription factor binding sites (NF-kB, PEA1, AP-1,TFIID; Positions 2852041-2852291 of NW_001494406.2) as well as the first two exons and introns of *TLR3*. No variation was detected within the predicted promoter itself. However, 40 validated SNPs were found to flank the putative promoter (see Table A2 for coordinates), with nearly half of this variation occurring immediately upstream (n=19 SNPs). Further evaluation of LD between adjacent variable sites for taurine cattle revealed two regions of *TLR3* with persistent, unbroken $r^2 > 0.50$ between all adjacent sites as follows: 1) Variable sites 1-5 upstream of the predicted promoter (Table A2); and 2) Variable sites 10-19, which span the predicted promoter. This unbroken pattern of persistent r^2 was also detected in our pooled analysis of all cattle, but did not extend across as many adjacent variable sites (Table A2, sites 13-17; region also spans the predicted promoter), and was only found in one upstream region. Therefore, it is possible that selection is primarily operating on noncoding variation within the genomic regions flanking the predicted promoter. Future functional studies will be needed to determine whether the SNPs flanking the predicted *TLR3* promoter actually modulate differences in gene expression.

Notably, only *TLR8* displayed a significant, positive value for Fu's F_S, indicating a lower than expected number of haplotypes, as would be predicted given a recent population bottleneck or strong balancing selection. However, the high r^2 that persists across nearly all adjacent variable sites strongly implies selection (Table 2). While previous studies have suggested that population bottlenecks may have occurred at the time of domestication and breed formation for modern cattle [5, 54], these are expected to drive

frequency distribution tests (D, F_S) toward more positive values because of the loss of rare genetic variation at all loci. In particular, the effects of bottlenecks are expected to be uniform and potentially dramatic for proximal, evolutionarily related X-linked loci (*TLR7*, *TLR8*) performing similar functions [7, 8, 21], especially given smaller effective population size (chromosomal) and female limited recombination. However, *TLR7* possesses a fundamentally different frequency distribution trend (D = -0.19828 all cattle; D = -0.17037 *B. t. taurus*) as compared to *TLR8* (*TLR7* ≤ 103 Kb from *TLR8*; Btau5.2), with no evidence for a significant deviation from a strictly neutral model (Table 4). A regression based test also provided no evidence for the effects of a population bottleneck or selection operating on variation within *TLR7* (P ≥ 0.05; see Figure 5). Therefore, it seems unlikely that historic bottlenecks are responsible for deviations from neutrality for bovine *TLR8*, and more likely that balancing selection is operating to preserve a limited number of functionally divergent haplotypes. Interestingly, the haplotypes observed for *TLR8* were partitioned into two main functional groups, as classified by our AA modeling (Table 3) and median joining haplotype networks (Figure 3). Specifically, haplotypes that fell into network nodes A, B, and C differed from haplotypes falling into nodes D, E, and F by eight nonsynonymous SNPs encoding AA substitutions (Table A2), with at least two (S477N; K903T) that were predicted to impact protein function (Table 3; Figure 3). Additionally, the four most common haplotypes (nodes A, B, D, and E) differed only by one synonymous SNP (nodes A vs. B; encoding S10S) and one putatively benign or tolerated nonsynonymous SNP (nodes D vs. E; encoding S492N; see Table A2; Table 3). For these reasons, functional studies are now needed to

comprehensively assess the dynamic range of ligand-induced *TLR8* signaling in domestic cattle.

In addition to *in silico* determined signatures of selection, we also provide evidence for associations between several bovine *TLR* SNPs and differential susceptibility to the causative agent of Johne's disease (Table 5). Unlike most previous studies [26-29, 53], we detected associations for which *TLR* variation both enhanced and decreased the risk of MAP infection. Furthermore, the SNPs demonstrating associations in this study (Table 5) were within bovine *TLR* genes that are either known or postulated to recognize ligands that would facilitate MAP detection and signaling [18, 21, 26-29, 53, 59]. While two recent genome wide association studies (GWAS) employing the Illumina BovineSNP50 assay provided no evidence for *TLR* involvement in differential susceptibility to Johne's disease in cattle [60, 61], the stringency of multiple testing employed during GWAS may have failed to identify *TLR* loci modulating relatively small effects. Moreover, the marker density of the BovineSNP50 assay is insufficient to detect all possible associations with bovine *TLR* variation [37] (Table A2). The SNP density for the new Illumina BovineHD assay also may not be sufficient to detect all disease associations with *TLR* loci, and therefore, additional association and functional studies are needed to clarify the involvement of *TLR2*, *TLR6*, and *TLR10* with respect to differential susceptibility to MAP infection in Holstein cattle.

Methods

DNA samples for SNP discovery

Bovine DNA samples (n= 96) representing *B. t. taurus*, *B. t. indicus*, and their hybrids were isolated from spermatozoa as previously described [30, 32, 37]. Bovine subspecies designation, breed names, and sample sizes (in parentheses) were: *B. t. taurus* -Angus (5), Belgian Blue (2), Blonde d'Aquitaine (1), Braunvieh (4),Brown Swiss (2), Charolais (6), Chianina-Chiangus (4), Corriente (1), Gelbvieh (4), Hereford (3), Holstein (6), Limousin (4), Maine-Anjou (3), Red Angus (4), Red Poll (1), Salers (2), Senepol (2), Shorthorn (4), Simmental (5), Texas Longhorn (2); *B. t. indicus* -Brahman (8), Nelore (2); Hybrids, termed Composites – Beefmaster (4), Braford (2), Brahmousin (2), Brangus (3), Piedmontese (1), Red Brangus (2), Romagnola (2), Santa Gertrudis (2), Simbrah (3). Bovine subspecies were assigned based on phenotype and breed origin (http://www.ansi.okstate.edu/breeds/cattle/).

Bovine TLR sequencing and SNP detection

Procedures involving primer design, PCR amplification with gene-specific primers, and standard dye-terminator cycle sequencing (Sanger) of all 10 bovine *TLRs* have previously been described [30-32, 67]. For this study, we synthesized gene-specific amplification primers with a unique 10 bp 5' barcode (Roche MIDs) for each of the 10 bovine *TLR* genes (Table A5). Thereafter, we standardized all 96 discovery panel DNAs to 50 ng/ml and created three DNA pools, with each pool consisting of 32 elite sire DNAs mixed at equal concentrations. Notably, larger-scale DNA pooling in a human

28

amplicon study supports the accuracy and reliability of this approach when coupled with Roche 454 pyrosequencing [68]. Three bovine DNA pools were used to amplify all *TLR* targets via barcoded primers (Table A5), with PCR conditions and thermal parameters as previously described [30-32, 67]. Targets that were intolerant to the addition of 5' oligonucleotide barcodes for PCR amplification were amplified using standard primers in conjunction with downstream dye-terminator cycle sequencing methods previously described [30-32, 67], with one exception: A second set of DNA pools (n= 12) was created, with each pool containing equal concentrations of DNA from eight elite sires derived from the sequencing discovery panel. Importantly, both sets of DNA pools (Sanger and Roche 454) were seeded with one or more reference DNAs that had previously been sequenced and/or SNP genotyped across all 10 bovine *TLR* genes [30-32, 67], which collectively included $\geq$ 12 reference DNAs possessing 216 validated diallelic variants (212 SNPs + 4 indels) [37]. All amplicons were purified using the Qiaquick PCR purification kit (Qiagen,Valencia, CA) as previously described [31, 32], and the concentrations were estimated by Nanodrop. For preparation of a Roche 454 Titanium fragment library, we standardized all barcoded amplicons to 10 ng/ml and devised a normalization procedure that accounted for differences in amplicon size (Table A1). Because the *TLR* amplicons differed in size, an adjustment was necessary to ensure balanced 454 pyrosequencing results. Specifically, using amplicon size, we computed the mean (bp) and standard deviation (SD; bp) across all PCR targets. Thereafter, any amplicon deviating from the mean by $\geq$ 0.5 SDs in either direction was subject to proportional adjustment within the fragment library (Table A1). The direction of

adjustment (plus or minus) was determined by the direction of the deviation (i.e., smaller = proportionally less template; larger = proportionally more template; Table A1). Because the emulsion PCR process involved in the preparation of Roche 454 Titanium fragment libraries favors smaller fragments, amplicons smaller than the mean by ≥ 0.5 SDs must be proportionally reduced in the final library, whereas the opposite is true for larger amplicons. Following normalization, the bovine *TLR* sequencing library was constructed via random ligation of sequencing adaptors provided with the GS FLX Titanium library kit (Roche Applied Science, Indianapolis, IN). All library preparation, emulsion PCR, quantitation, and sequencing steps followed the manufacturer's protocol (Roche Applied Science).

SNP detection analyses for the resulting pyrosequencing data employed the Neighborhood Quality Standard algorithm [69, 70] implemented within CLC Genomics Workbench (v3.7.1), as previously described [36]. Putative SNPs were filtered using a method devised from *a priori* knowledge of biallelic controls (212 SNPs + 4 indels) [37] that were purposely seeded into the amplicon library. Briefly, we considered the possibility that some SNPs may only be found as one allele in a single elite sire (1/192 total alleles; see reference 30 for examples). Therefore, we filtered all putative SNPs predicted from our analysis of the pyrosequencing data using the following formula: 1/192 x (Total SNP Coverage) = Theoretical minimum number of reads, which represents the smallest number of reads required to shuttle putative SNPs into a validation workflow involving custom, allele-specific genotyping assays. This method

proved valuable for the discovery and validation of many low frequency SNPs, including those that occurred as one allele for a single discovery panel sire (i.e., *TLR5* putative nonsense SNP = 1/192 alleles in the discovery panel). For SNP discovery using standard dye-terminator sequencing reads, we used an alignment-based method of variant detection within the program Sequencher 4.6 [30, 32]. Briefly, high quality electropherograms were manually inspected for any evidence of a double peak. Individual nucleotide sites displaying any evidence of heterozygosity within $\geq$ 1 sequencing read were shuttled to our SNP validation workflow.

SNP validation and genotyping

All 96 DNAs from the pyrosequencing discovery panel were also used for allele-specific genotyping. Additionally for bovine *TLRs* recognizing bacterial ligands, we also utilized the following industry-relevant DNA panels: Beef (48 Purebred Angus, 1 Herd); Dairy (405 Holstein dairy cows, 3 Herds). SNPs and indels were genotyped using the KASPar allele-specific fluorescent genotyping system (Kbiosciences, Hertfordshire UK), as previously described [36, 37]. 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). Genotype quality was assessed by manually inspecting the clustering data for every individual marker, and by comparing KASPar-derived genotypes to those derived from previously reported sequence data [30, 32, 42]. Poor clustering or

31

inconsistent genotypes precipitated the following workflow: 1) Further optimization and/or redesigning the SNP assay followed by; 2) Genotyping the inconsistent samples again. Notably, to minimize the frequency of missing genotypes from a very low proportion of failed assays, most SNPs were genotyped multiple times for every DNA sample.

Haplotype inference, LD estimates and variant tagging

Unphased diploid genotypes were compiled and cross-checked for parsing errors using two custom software packages [37]. Haplotype reconstruction and missing data imputation (< 0.58%) was performed with PHASE 2.1 [38, 71, 72] using all validated intragenic polymorphisms, all cattle for a given locus, and the –X10 option. Haplotype estimation using PHASE 2.1 is not sensitive to departures from Hardy-Weinberg equilibrium (HWE) [38, 71, 72]. Predicted haplotype phases with best pair probabilities $\geq$ 0.90 were retained for further analysis. Bovine X-linked haplotypes (*TLR7*, *TLR8*) were directly ascertained by genotype homozygosity in our sire panel used for SNP discovery. Estimates of recombination across each gene were also assessed in PHASE 2.1 using the general model for varying recombination rate [40-42]. Deviation from the average background recombination rate ($\bar{\rho}$) [41, 42] by a factor $\geq$ 2.5 between adjacent sites was considered evidence for historical recombination.

Intragenic LD was visualized within Haploview [39] using unphased diploid autosomal genotypes and phase-known X-linked data (*TLR7*, *TLR8*) for *B. t. taurus* samples, and all

32

cattle combined. LD patterns and blocks were estimated via majority rule from: 95% confidence intervals constructed for D' [39, 40]; application of the four gamete rule [39] (4th gamete > 0.02); and estimates of recombination between adjacent sites [41, 42]. To further evaluate 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 predicted to capture 100% of the variation ($r^2 > 0.80$) segregating in *B. t. taurus* and all cattle combined was deduced using the Tagger algorithm implemented in Haploview.

Median joining haplotype networks

Because median joining (MJ) networks require the absence of recombination [73], genes displaying evidence of historical recombination (*TLR2*, *TLR3*, *TLR6*) were each partitioned into two regions of elevated LD. Haplotypes were reconstructed [38] for each intragenic region and best pairs were used for MJ network analyses [35]. This approach improved the proportion of cattle with best pairs phase probabilities $\geq$ 0.90 and eliminated regions displaying overt evidence of recombination. MJ networks were constructed using Network 4.5.1.0 (Fluxus Technology Ltd, Suffolk, England), and the default character weights of 10 for SNPs and 20 for indels. Results were visualized, annotated, and adjusted within Network Publisher (Fluxus Technology Ltd, Suffolk, England). Branch angles were adjusted to ensure proper network magnification and clarity without changing branch lengths.

33

AA substitution phenotypes and TLR10 evolutionary analyses

Bovine AA substitution phenotypes were predicted using PolyPhen [45] and SIFT [46] (http://genetics.bwh.harvard.edu/pph/;

http://genetics.bwh.harvard.edu/pph/pph_help.html;   http://sift.jcvi.org/;

http://sift.jcvi.org/www/SIFT_help.html) with the default settings. Results other than ''benign'' or ''tolerated'' were categorized as substitutions predicted to impact protein function [37, 45, 46]. To assess the potential for functional and/or selective constraint across the entire bovine *TLR* gene family, a goodness of fit test (χ^2) was performed assuming equal probabilities for benign or tolerated AA phenotypes versus those predicted to impact protein function. Frequency distribution tests, including Tajima's *D* [50] and Fu's F_S [51], were performed in DnaSP v4.90.1 [74] using all validated SNPs. Significance levels for frequency distribution tests were defined by confidence intervals estimated for each test statistic via coalescent simulation (10,000 replicates) [74]. Simulations were performed given the observed number of segregating sites, both with and without recombination [74, 75].

At each polymorphism we estimated the effective number of alleles as $E_i = 1/(1 - 2p_i(1-p_i)) = 1/(p_i^2 + (1 - p_i)^2) = 1/(\text{expected HWE frequency of homozygotes})$ where p_i is allele frequency at the ith locus. Thus a measure of polymorphism diversity is $\log_2 (E_i)$ which also represents the information content of each SNP [37]. 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 jth 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.

Association tests with MAP infection status

A case-control study was performed to estimate the association between specific *TLR*

genotypes and MAP infection in Holstein cattle. The study population was derived from

an established repository [76] that included whole blood samples preserved from adult

Holstein cattle in three herds that were characterized on the basis of: 1) MAP bacterial

culture of feces; 2) MAP bacterial culture of tissues for harvested cattle; 3) ELISA

values for MAP-specific antibody. Cattle from which MAP was cultured in the feces

and/or the tissues collected at harvest were selected as cases (n= 68). Herd-matched

controls (n= 270) were selected from those cattle in the repository with negative ELISA

and bacterial culture data. Cattle with multiple negative tests were preferentially selected

to reduce the probability of misclassification relative to infection status due to the low

sensitivity of available diagnostic methods for MAP. DNA was extracted from available

blood specimens using a commercial kit (MoBio DNA non-spin, Carlsbad, CA) and

assessed for quality as well as concentration by standard spectrophotometric methods.

Genotypes for validated SNPs and indels in the 5' upstream regions, introns, and those

associated with nonsynonymous or putative nonsense mutations in bovine *TLR* genes

recognizing bacterial ligands (*TLR1, TLR2, TLR4, TLR5, TLR6, TLR9, TLR10*) (see refs

[21, 23]) were evaluated for further analysis. Loci fixed for the major allele in our dairy population were excluded, leaving 35 nonsynonymous and 1 putative nonsense substitution, and 37 other SNP loci within the 5' upstream regions or intragenic introns. For these 73 variable sites, we excluded SNPs and indels with MAFs < 0.01 in our infected cases, leaving 32 SNPs and 3 indels for association tests (see Table A1).

Conditional logistic regression models were constructed for each of the 35 variable loci to estimate the relative odds of being infected with MAP based on the defined diagnostic criteria adjusted for the effects of herd using the PHREG procedure of SAS (SAS v. 9.2, SAS, Cary, NC). Effects of genotype were estimated using 3 different covariate specifications. First, an additive mode of inheritance was examined whereby the odds of infection associated with each additional copy of the minor allele was modeled as a single continuous covariate. Second, a recessive mode of inheritance was modeled, where the odds of infection in cattle homozygous for the minor allele were estimated relative to cattle heterozygous and homozygous for the major allele. Finally, each genotype was modeled as an indicator variable and effect estimates were generated for cattle homozygous for the minor allele, and for heterozygous cattle, both relative to cattle homozygous for the major allele. This allowed evaluation of assumptions in the additive model with respect to the effect of the additional copies of the minor allele being linear in the log odds, and potential intermediate effects of the minor allele not captured in the other models. Potential confounding by age was examined by including birth year as a fixed covariate (where available), and was defined as a change in the

relative odds of greater than 20% after addition of the birth year term. For models where evidence of confounding by age was detected, birth year was retained in the model to adjust genotype estimates for this effect. With the exception of *TLR1*, *TLR6*, and *TLR10*, all single marker P-values were corrected for multiple testing by applying the FDR correction (http://sdmproject.com/utilities/?show=FDR) [52] to the raw P-values derived from each investigated gene (locus-specific correction). Given the close physical proximity of *TLR1*, *TLR6*, and *TLR10* on BTA6, these genes were considered a single locus for correction of multiple tests. However, it should be noted that none of the variable markers within *TLR1* met our inclusion criteria (MAFs > 0.01), and therefore, locus-specific correction was only applied to raw P-values from *TLR6* and *TLR10*.

Haplotype association tests were performed in PHASE 2.1 [38]. Briefly, for dairy cattle with disease classifications based on bacterial culture status of MAP, we tested the hypothesis that haplotypes differ among cases and controls for all genes evaluated in the single marker association analysis (68 cases, 270 controls, n =338 total). For maximum LD-based resolution of haplotypes, we used all variable markers within seven bovine *TLR* genes that recognize bacterial ligands. Significance was estimated via 1,000 permutations.

CHAPTER III

DIVERSITY AND EVOLUTION OF THE EQUINE *TLR* GENE FAMILY

Introduction

Following the establishment of a reliable equine genome map, for orientation toward candidate genes of equine traits of interest [77], studies of the equine genome have primarily focused on either athletic ability or aspects of equine animal health [78]. With the advent of massively parallel sequencing technologies, and medium density single nucleotide polymorphism (SNP) arrays [79, 80], it has since become possible to use modern bioinformatic techniques to elucidate genomic regions and variation that differ between horses with a variety of disparate phenotypes [81-83]. Similar to the natural progression of science and modern animal husbandry practices currently underway in domestic cattle [84], it is possible that when the equine genome is further explored, and the genetic components modulating equine traits of interest are more fully resolved, a surge is likely to ensue in equine genome-assisted selective breeding [1]. At present, one popular avenue of equine research relates to the search for genetic variation that either influences or is causal for common equine problems including differential susceptibility to infectious diseases and other important health concerns (for review see http://www.uky.edu/Ag/Horsemap/hgpprojects.html). Although important, the potential for enhanced disease resistance and athletic performance through genome or marker-assisted selection is not the only focus of equine genomics initiatives, as many other

needs currently exist, including the potential for marker-assisted vaccination, where genotypes are used as indicator variables for enhancing vaccine design, or as predictors of host response [37, 85].

Toll-like receptors (*TLR*), expressed from the *TLR* family of genes, act as molecular sentries for the innate immune system by responding to pathogen associated molecular patterns (PAMPs) and triggering a host immune response without needing prior exposure [2, 7]. Of interest to the fields of equine health and innate immune biology, the mammalian *TLR* loci encode proteins that recognize a variety of different pathogen ligands, with six gene family members (*TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, *TLR9*) known to recognize microbial (bacteria, fungi, protozoa) and/or synthetic ligands, and five (*TLR3*, *TLR4*, *TLR7-TLR9*) known to recognize viral components [8, 21]. Although *TLR10* was considered the only orphan member of the *TLR* gene family for which one or more specific ligands had not been identified [9], studies indicate that human TLR10 forms functional heterodimers with both TLR1 and TLR2, which is hypothesized to enable the resulting protein complexes to recognize a diverse array of microbial ligands [10]. Moreover, a recent study has provided further clarity by demonstrating that amino acid (AA) substitutions in human TLR1 and TLR10 negatively impacted receptor function [10, 13], with TLR10 ligand specificity determined to be similar to those established for TLR1 [10].

At present, relatively few equine *TLR* studies exist [78, 86-89], with the objectives of these studies primarily limited to the discovery and characterization of equine *TLR* transcripts, levels of endogenous expression in selected equine tissues, and factors that may potentially alter equine *TLR* expression. Moreover, because mammalian innate immune studies have clearly demonstrated that some naturally occurring *TLR* variants enhance the risk of severe infections in humans, mice, and domestic cattle [10, 66, 85, 90], a need currently exists to comprehensively evaluate the frequency, distribution, and putative functional implications of naturally occurring equine *TLR* variation. Herein, we provide a detailed study of equine *TLR* variation with haplotype inference, variant tagging, and functional modeling of AA replacements encoded by validated equine *TLR* SNPs. The results of this study will directly facilitate equine case-control studies aimed at determining the relationship between naturally occurring *TLR* genetic variation and equine health traits.

Results

Equine TLR pyrosequencing, SNP detection, variant validation, and haplotype inference

Using 96 sample equines representing 42 horse and pony breeds as well as the donkey, we generated and purified 10,560 amplicons targeting 9 equine innate immune genes (*TLR1-TLR4*, *TLR6-TLR10*). All amplicons were pooled to form a normalized fragment library (Table B1, Figure B1) which was subjected to an established pyrosequencing and variant detection workflow [86]. Collectively, 337 variable sites were predicted from our intragenic analyses of the equine *TLR* pyrosequencing data, which included 10 recently

validated SNPs [89]. Further examination of the raw sequencing data and corresponding read-pileups for each *TLR* gene revealed evidence for ≥ 69 read errors. Like our previous cattle *TLR* study [85], many equine *TLR* SNPs were also predicted either within or immediately flanking homopolymer repeats, which is a known problem associated with Roche 454 pyrosequencing chemistry [85]. Using custom genotyping assays, we validated 179 biallelic variants (67%; 179/268) across the 9 investigated *TLR* genes. In order to assess the global accuracy of our variant discovery and validation workflow, we compared corresponding minor allele frequencies (MAFs) across the 9 *TLR* genes (Table 6) using a regression based approach previously described [85]. An analysis performed across all genes revealed that there was little to no bias in the estimates of allele frequencies produced via targeted pyrosequencing ($P=0.99018$; Ho: slope=0.9968; Figure 6).

Table 6. Relationship between minor allele frequencies estimated from pyrosequencing and allele-specific genotyping of 96 individuals from 42 breeds of horse and one breed of donkey.

Bovine Gene	Total 454 SNPs[a]	Overall Correlation (r)[b]	Overall RSQ (r^2)[c]
TLR1	6	0.999	0.998
TLR2	11	0.753	0.568
TLR3	31	0.931	0.866
TLR4	63	0.892	0.795
TLR6	2	1.000	1.000
TLR7	31	0.823	0.677
TLR8	17	0.836	0.699
TLR9	10	0.900	0.810
TLR10	8	0.855	0.731
Totals/Avg	**179**	**0.831**	**0.691**

[a] Total SNPs detected via pyrosequencing
[b] $P < 0.05$ for all *TLR* genes
[c] RSQ is the squared correlation coefficient (r^2)
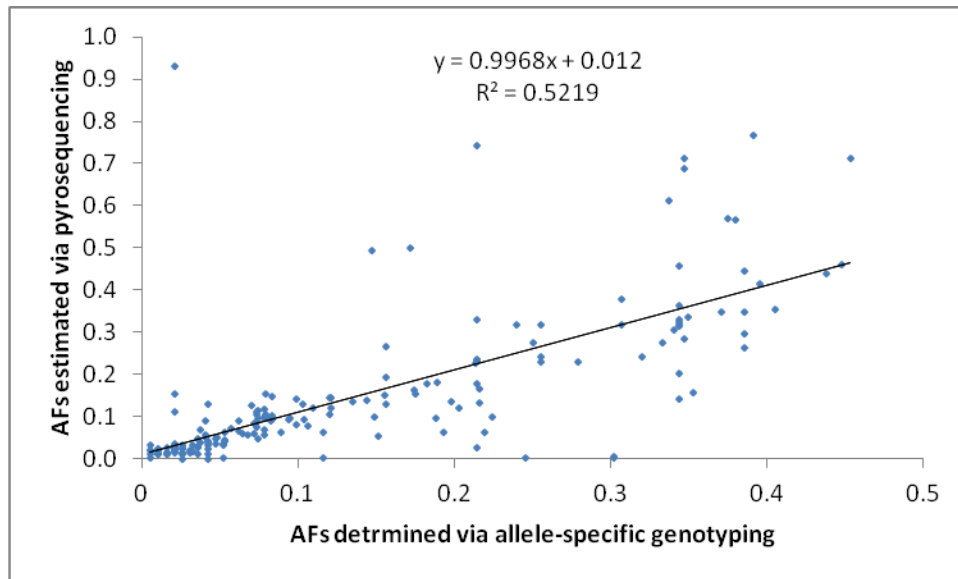
Figure 6. For validated equine *TLR* SNPs detected via pyrosequencing (n = 179), a regression analysis was performed for pyrosequencing allele frequency (AF) estimates corresponding to the true minor alleles (< 0.5), as defined by allele-specific genotyping assays, and minor AFs (MAFs) directly ascertained by genotyping (n = 96 samples, 43 breeds). The true minor alleles (< 0.5) were correctly identified for 170/179 (95%) SNPs via pyrosequencing. This analysis provided strong statistical evidence (P=0.999018; Ho: slope = 1) for little or no bias in the pyrosequencing-based estimates of allele frequency.

Altogether, 175 SNPs were successfully incorporated into 144 unique haplotypes (Table 7). Four SNPs (*TLR*3: 13787, 14310; *TLR*4: 1030; *TLR*9: 3749; Table B2) could not be incorporated into discrete haplotypes with best-pair phase probabilities ≥ 0.90. Across all investigated loci, the MAF spectrum derived from allele-specific genotyping assays ranged from 0.101 to 0.499, with 53% of the validated SNPs possessing MAFs ≤ 0.10 (Table 2).

Characterization of LD architecture, recombination, and intragenic tagSNPs

When evaluating LD across all equine samples, each *TLR* gene revealed one or more blocks of strong LD. Evidence for historical recombination was detected within *TLR*3,

Table 7. Summary data for 9 equine innate immune genes investigated

Equine Gene	ECA Assign[a]	Total Haps[b]	Sires Phased[c]	MAFs ≤ 0.10[d]	Avg r2 alle[e]	Valid. SNPs[f]	Hap SNPs[g]	Valid. nsSNP[h]	Valid tagSNPs[i]	Region Size (Kb)[j]
TLR1	3	7	88	5	0.008	6	6	6	5	3.0
TLR2	2	10	95	6	0.186	11	11	3	8	3.2
TLR3	27	24	88	18	0.037	31	29	3	25	14.5
TLR4	25	29	87	27	0.226	63	62	6	62	10.6
TLR6	3	3	94	1	0.011	2	2	2	2	2.7
TLR7	X	24	96	19	0.107	31	31	2	25	23.6
TLR8	X	35	92	10	0.008	17	17	5	16	9.3
TLR9	16	7	92	5	0.228	10	9	2	8	4.4
TLR10	3	7	95	3	0.027	8	8	6	7	2.8
Total/Avg		**144**	**(96%)**	**94**	**0.093**	**179**	**175**	**35**	**158**	**74.1**

[a]EquCab assignments based on NCBI Refseq (EquCab2.0).
[b]Total haplotypes predicted from all validated markers and best pair reconstructions [42] with probabilities ≥ 0.90.
[c]Proportion of horses exhibiting best pair phase probabilities ≥ 0.90.
[d]Total polymorphisms with minor allele frequencies ≤ 0.10.
[e]Average intragenic r^2 values estimated for adjacent SNP and indel sites for all horses.
[f]Numbers of putative SNPs validated as polymorphic.
[g]Numbers of validated SNPs placed on discrete haplotypes.
[h]Numbers of putative nonsynonymous SNPs validated as polymorphic.
[i]Numbers of tagSNPs as detected by Haploview [39].
[j]Size of the genic region rounded to the nearest 100 bp. Kb = Kilobase.

TLR4, and *TLR8*, resulting in at least two detectable LD blocks within each gene. All other genes exhibited a single block of strong LD spanning either all, or the majority of all validated intragenic SNPs, as supported by the majority rule of three analyses: confidence intervals constructed for D' [39, 40], application of the four gamete rule [39], and estimates of recombination between adjacent variable sites [31, 32]. A comparison of average intragenic r^2 values calculated between adjacent variable sites across the *TLR* genes revealed a dynamic range of LD (0.008-0.228; Table 7), with a total of 5 SNPs that produced estimates of median recombination rates that exceeded the background rate by a factor of at least 2.5. The highest estimate of median recombination rate was observed in *TLR8*, which exceeded the background rate by a factor of at least 7.2. Analyses to identify tagSNPs predicted to capture 100% of the variation at all 179 validated variable sites yielded 158 total tagSNPs (Table 7; Table B3).

High resolution equine TLR haplotype networks and breed distribution

Median joining haplotype networks (exemplified by Figures 7-9) constructed for 9 equine *TLR* genes revealed that: 1) We cannot fully discriminate between specialized breeds (Pony, Light-horse, Draft-horse) using these markers, despite an average density of one variable marker per 414 bp; 2) The estimated 10 Myr divergence between *E. caballus* and *E. asinus* [91] was only revealed in one haplotype network (*TLR*4 block 2; Figure 8); and 3) Haplotypes shared between specialized equine breeds and the donkey were often some of the highest frequency nodes within a gene-specific haplotype network.
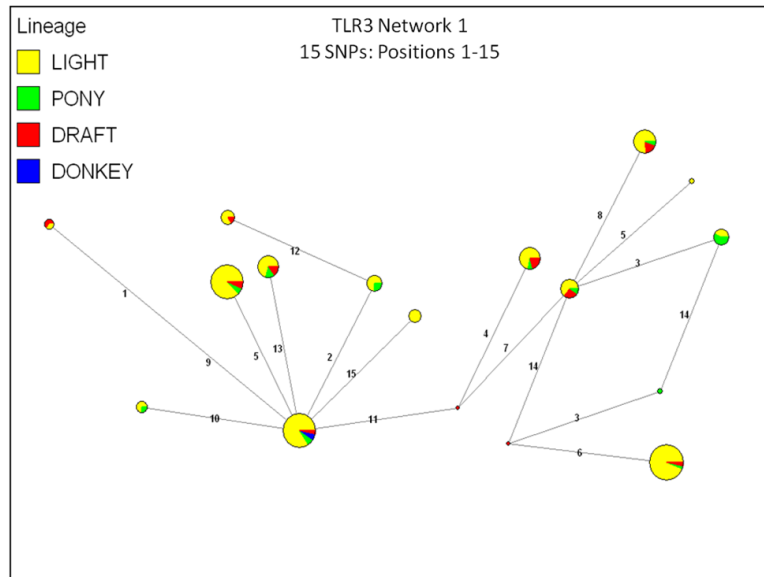
Figure 7. Median joining (MJ) haplotype network for equine *TLR3*. Because MJ networks require the absence of recombination [73], each network represents intragenic regions of elevated LD; this network represents the first block of elevated LD in *TLR3*. Haplotypes predicted for light horses, ponies, draft horses, and donkeys are color coded. Numbers indicate SNP positions in numerical order (see Table B2 for SNP information). Node sizes are proportional to haplotype frequency, and all branch lengths are drawn to scale.
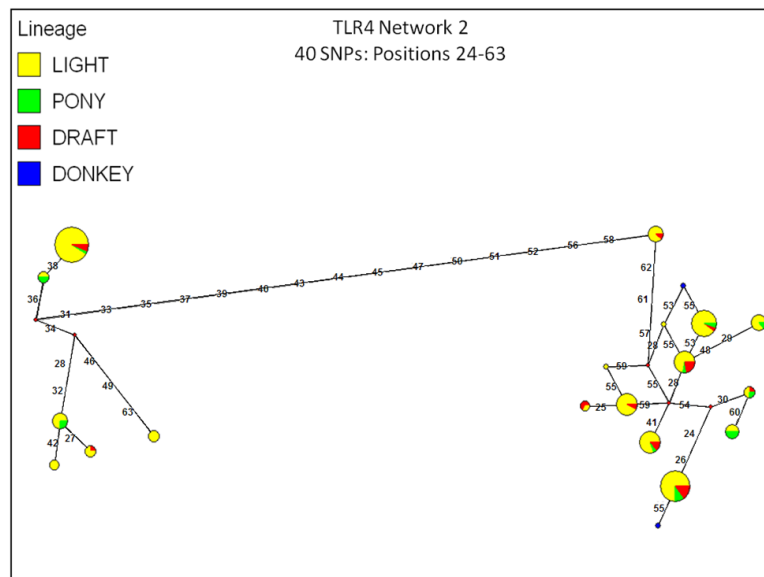
Figure 8. Median joining (MJ) haplotype network for equine *TLR4*. Because MJ networks require the absence of recombination [73], each network represents intragenic regions of elevated LD; this network represents the second block of elevated LD in *TLR4*. Haplotypes predicted for light horses, ponies, draft horses, and donkeys are color coded. Numbers indicate SNP positions in numerical order (see Table B2 for SNP information). Node sizes are proportional to haplotype frequency, and all branch lengths are drawn to scale.

45
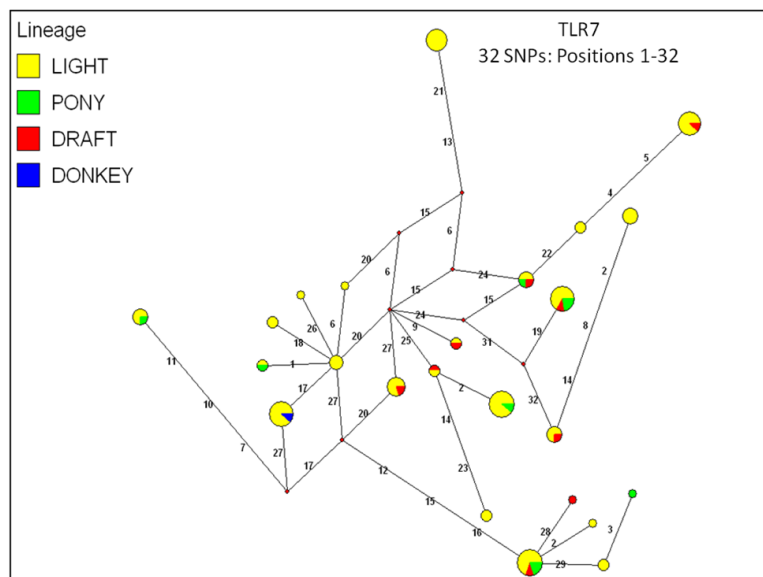
Figure 9. Median joining (MJ) haplotype network for equine *TLR7*. Haplotypes predicted for light horses, ponies, draft horses, and donkeys are color coded. Numbers indicate SNP positions in numerical order (see Table B2 for SNP information). Node sizes are proportional to haplotype frequency, and all branch lengths are drawn to scale.

Functional modeling of predicted equine amino acid (AA) substitutions and tests of selection

Both PolyPhen [45] and SIFT [46] were used to evaluate the putative functional effects of AA substitutions encoded by *TLR* SNPs, and we subsequently determined that 22 of 33 (67%) AA substitutions were likely to be benign and/or tolerated, whereas 11 of 33 (33%) were predicted to impact protein function by at least one analytical method (Table 8). For those mutations predicted to impact protein function, 4/11 (36%) located in *TLR4* (1), *TLR7* (1), *TLR8* (1), and *TLR10* (1) were detected at frequencies < 0.05, and 7/11 (64%) were observed at frequencies ≥ 0.05, with the highest frequency substitution detected in *TLR2* (0.498). Across all polymorphisms encoding AA substitutions, PolyPhen and SIFT produced analogous predictions for 27/33 (82%) AA replacements.

Table 8. Summary data for 11 nonsynonymous SNPs predicted to impact protein function

Equine Gene	SNP	GenBank Protein ID	AA Subst.[a]	Protein Domain[b]	PolyPhen Result[c]	SIFT Result[c]	SNP Freq[d]
TLR1	A > G	XP_001498694.2	Y236C	NCP	PsD	T	0.053
	G > A	XP_001498694.2	D573N	LRRCT	B	AF	0.078
TLR2	G > A	NP_001075265.1	R579H	LRRCT	PsD	T	0.828
TLR4	C > T	NP_001093239.1	P3L	NCP	PrD	AF	0.026
	A > G	NP_001093239.1	N310D	NCP	PrD	AF	0.344
TLR7	G > T	NP_001075240.1	L223F	LRR_1	PrD	AF	0.010
TLR8	A > G	NP_001104771.1	N39S	NCP	PsD	T	0.021
TLR10	A > G	XP_001498728.1	T117A	LRR_3	B	AF	0.698
	T > G	XP_001498728.1	F355L	NCP	PsD	AF	0.005
	T > C	XP_001498728.1	F637L	TIR	PrD	AF	0.052

[a]Amino acid (AA) substitutions predicted from corresponding SNPs, GenBank Proteins, and previous studies [30-32, 37, 66].
[b]Protein domain locations predicted by SMART (http://smart.embl-heidelberg.de/). Only confidently predicted domains are depicted (NCP = no confident prediction; LRRs are named in order of prediction).
[c]Results from PolyPhen and SIFT [45, 46]. Results other than "Benign (B)" or "Tolerated (T)" are predicted to be Possibly Damaging (PsD), Probably Damaging (PrD), or Affect Protein Function (AF).
[d]Observed frequency of nonsynonymous SNP allele in all equine samples.

Similar to our cattle *TLR* investigation [85], we endeavored to collectively estimate the extent of functional and/or selective constraint(s) related to equine *TLR* protein function, and therefore, we used a goodness of fit test to examine disparities between the observed distributions of AA phenotypes (PolyPhen + Sift results; benign/tolerated vs. damaging/affecting). Assuming equal probabilities for the occurrence of both classes of AA phenotypes across all equine *TLR*s, we found there to be significantly fewer substitutions predicted to impact protein function than those classified as benign or tolerated ($P \leq 0.01$); a result that is fully compatible with results from a similar *TLR* analysis performed for domestic cattle.

Discussion

The equine genome has experienced an exponential expansion with respect to the discovery and utilization of variable genetic markers since the initial genomes maps and original genome sequence was released (for review see [92]). From the first equine gene map [93], the marker density for modern maps has increased by more than 14 fold [77]. Additionally, more recent studies have also focused on the detection and validation of a genome-wide set of variable genetic markers, many of which were incorporated within the Illumina EquineSNP50 assay which features over 54,000 common SNPs spaced throughout the equine genome, with an average marker interval density of one marker every 43.2kb (15 horse breeds) [80]. Nevertheless, like most domesticated species for which low-to-medium density SNP arrays have been developed [94, 95], the underlying genome-wide marker interval density equates to very poor coverage for most genes of interest, with genome-wide association studies using these assays actually conditioned upon the technical limitations of the SNP chip itself [37]. Therefore, given ample precedence for the importance of the mammalian *TLR* gene family with respect to mammalian innate immunity [10, 66, 85, 90], an obvious need existed to develop high density polymorphism data for the equine *TLR* loci, with subsequent utilization of that information to construct custom equine *TLR* genotyping assays. To this end, we validated SNPs from 42 different horse breeds and one donkey, resulting in an average *TLR* marker interval density of approximately one SNP every 414 bp, which represents an 18 fold increase in the localized average marker density, as compared to the average density of the Illumina Equine SNP50. Importantly, equine *TLR* variation detected and

48

validated in this study, in conjunction with tagSNPs and fundamental knowledge of equine *TLR* haplotype structure, will directly facilitate future case-control studies aimed at evaluating the potential for single-marker and haplotype-based associations with susceptibility to infectious diseases as well as vaccine phenotypes in horses.

Examination of the observed patterns of LD for the equine *TLRs* targeted in this study revealed evidence for historical recombination in at least three of the nine genes investigated (i.e. *TLR3, TLR4, TLR8*). Moreover, estimates of linkage disequilibrium (r^2) across all adjacent variable sites within each equine *TLR* gene (Table 7) are strikingly lower than those observed for domestic cattle [37, 85]. The underlying reason for this is the apparent preservation of phase-relationships among bovine *TLR* variants across breeds and subspecies [37, 85]; a phenomenon that is less common among diverse horse and pony breeds (Table 7).

Interestingly, several lines of evidence indicate that functional and/or selective constraint(s) are dominant forces that have contributed to the observed patterns of variation within the equine *TLRs*. For example, a goodness of fit test designed to examine the observed pattern of nonsynonymous variation in terms of the expected null model (i.e. Ho: no functional or selective constraint) provides significant statistical support for the presence of functional or selective constraints within the equine *TLRs* (Table 8; see Results section). Moreover, the directionality of this observation (significantly fewer AA replacements predicted to impact protein function than

expected) is highly compatible with our results from the Seabury-Taylor test of selection (Figure 10; Table 8), where *TLR3*, *TLR7*, and *TLR8* all appear to possess significantly less diversity than other members of the equine *TLR* gene family.
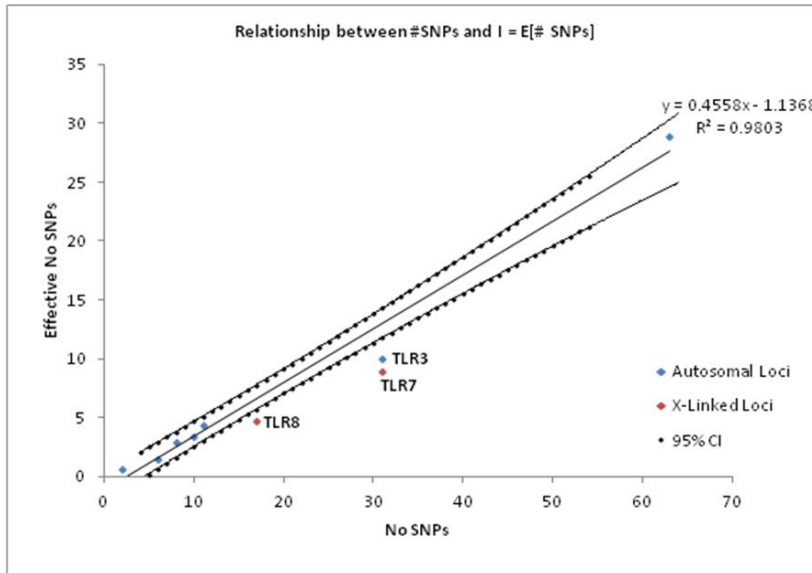
Figure 10. Relationship between the number of validated SNPs and SNP diversity here denoted as the effective number of SNPs across 9 *TLR* loci in equines. The linear regression and estimated 95% confidence interval is shown.

Likewise, median-joining haplotype networks generated for *TLR3*, *TLR7*, and *TLR8* (Figure 7, Figure 9, Figure B2) possess one unifying feature; an abundance of low frequency haplotypes, which is often considered a signature of purifying or directional selection [37]. Collectively, our analyses of these data indicate that purifying and/or directional selection is/are the most likely forces to have shaped natural variation within equine *TLR3*, *TLR7*, and *TLR8*. However, it is also possible that similar forces may have shaped variation within other equine *TLR* genes, and that our regression-based test of selection lacks power to detect these signatures. Therefore, future studies involving

branch-specific tests of selection for representative species of equidae is expected to help further resolve and characterize of locus-specific selective forces detected in this study.

Methods

DNA samples for SNP discovery

Equine DNA samples (n = 96) were isolated from whole blood. All samples were male except for Twilight, which is the female thoroughbred horse used for genome sequencing, assembly, and annotation [79]. Breeds and sample sizes were: Akhal Teke (2), American Bashkir Curly (2), American Cream Draft Horse (1), Andalusian (4), Appaloosa (4), Arabian (4), Belgium Draft (1), Canadian (2), Canadian Draft Horse (1), Caspian (3), Clydesdale (2), Connemara (3), Dutch Warmblood (1), Exmoor (2), Florida Cracker (2), Friesian (6), Haflinger (2), Hanoverian (3), Irish Draught (1), Irish Sporthorse (1), Lusitano (2), Marsh Tacky (1), Miniature Horse (4), Missouri Fox Trotter (1), Morgan (4), Mustang (3), Norwegian Fjord (2), Paint (2), Paso Fino (1), Percheron (2), Peruvian Paso (4), Poitou (1), Polish Primitive (2), Pura Raza Espanola (1), Quarter Horse (2), Selle Francais (1), Shire (1), Spanish Colonial (1), Spanish Mustang (4), Standardbred (2), Tenn Walker (4), Thoroughbred (1), Warmblood (3). Breeds were classified into the category of Draft, Light, Pony, and Donkey based on phenotype and breed origin (http://www.ansi.okstate.edu/breeds/horses/).

TLR sequencing and SNP detection

Procedures involving primer design**,** PCR amplification with gene-specific primers (Table B4), and standard dye-terminator cycle sequencing (Sanger) for target verification have previously been described [30-32, 66], with all equine *TLR* genes initially sequenced, assembled, and verified using DNA from Twilight; the horse used for genome sequencing [79]. Thereafter, we synthesized gene-specific amplification primers with a unique 10 bp 5' barcode (Table B4) (Roche MIDs) for 9 equine *TLR* genes that were annotated in the equine genome assembly (EquCab2.0). Prior to amplification, we standardized all 96 equine discovery panel DNAs to 50 ng/μl and created three pools consisting of 32 samples mixed at equal concentrations. Notably, larger-scale DNA pooling in a human amplicon study supports the accuracy and reliability of this approach when coupled with Roche 454 pyrosequencing [72], as does a recent cattle *TLR* study whereby little or no bias was observed in the SNP allele frequencies estimated by pyrosequencing, as compared to those directly ascertained by individual genotyping assays [85]. The three equine DNA pools were used to amplify all *TLR* targets via barcoded primers, with PCR conditions and thermal parameters following methods previously described [30-32, 66] (see Table B1 for details). All amplicons were subsequently purified using the Qiaquick PCR purification kit (Qiagen, Valencia, CA) as previously described [31, 32], and the concentrations were estimated by Nanodrop. For preparation of a Roche 454 Titanium fragment library, we standardized all barcoded amplicons to 10 ng/μl. Because the *TLR* amplicons differed in size, an adjustment was necessary to ensure balanced 454 pyrosequencing results, as previously described for

domestic cattle [85]. Following normalization, the *TLR* sequencing library was constructed via random ligation of sequencing adaptors provided with the GS FLX Titanium library kit (Roche Applied Science, Indianapolis, IN). All library preparation, emulsion PCR, quantitation, and sequencing procedures followed the manufacturer's recommended protocol (Roche Applied Science).

SNP detection analyses for the resulting pyrosequencing data employed the Neighborhood Quality Standard algorithm [68, 69] implemented within the CLC Genomics Workbench (v3.7.1), as previously described [36]. Putative SNPs were filtered using a method that was previously devised from *a priori* knowledge of bovine biallelic controls (212 SNPs + 4 indels) that were purposely seeded in a bovine *TLR* amplicon library, which allowed for discovery of low frequency variation (i.e., one allele in the total sample) [37]. Therefore, considering the possibility that some equine SNPs may also only be found as one allele in the total experimental sample (1/192 total alleles for equines) we filtered all putative SNPs predicted from our analysis of the pyrosequencing data using the following formula: (1 / Total Number of Alleles)*(Total SNP Coverage) = Theoretical Minimum Number of Reads Required (TMNRR), which represents the smallest number of reads required to elicit a validation workflow for putative SNPs using custom, allele-specific genotyping assays. This method proved valuable for the discovery and validation of low frequency SNPs in domestic cattle [85] as well as this study.

SNP validation and genotyping

All equine DNAs from the pyrosequencing discovery panel were also used for allele-specific genotyping. Specifically, SNPs were genotyped using the KASPar allele-specific fluorescent genotyping system (Kbiosciences, Hertfordshire UK) [37]. Thermal cycling parameters and reaction concentrations followed the manufacturer's recommendations, with some modifications to $MgCl_2$ concentrations. Genotype clustering and calling was performed using KlusterCaller and SNPviewer2 software (Kbiosciences). Genotype quality was assessed by manually inspecting the clustering data for every marker [30-32, 37]. Poor clustering or genotypes that were inconsistent precipitated the following workflow: 1) Further optimization and/or redesigning the SNP assay followed by; 2) Genotyping the inconsistent samples again. To minimize the frequency of missing genotypes from a very low proportion of failed assays, most SNPs were genotyped multiple times (> 2) for every DNA sample.

Haplotype inference, LD estimates and variant tagging

Unphased diploid genotypes were compiled and cross-checked for parsing errors using two custom software packages [37]. Haplotype reconstruction and missing data imputation (< 0.59%) was performed with PHASE 2.1 [38, 70, 71] using all validated intragenic polymorphisms, all samples for a given locus, and the –X10 option. Notably, haplotype estimation using PHASE 2.1 is not sensitive to departures from Hardy-Weinberg equilibrium (HWE) [38, 70, 71]. Predicted haplotype phases with best pair probabilities ≥ 0.90 were retained for further analysis. Unlike the autosomal *TLRs*,

Equine X-linked haplotypes (*TLR7* and *TLR8*) were directly ascertained by genotype homozygosity for males included in our DNA panel. Estimates of recombination across each gene were also assessed in PHASE 2.1 using the general model for varying recombination rate [38, 41, 42], where deviation from the average background recombination rate ($\bar{\rho}$) [41, 42] by a factor ≥ 2.5 between adjacent variable sites was considered evidence for historical recombination.

In addition to the general model for varying recombination rate, intragenic LD was also estimated and visualized within Haploview [39] using unphased diploid autosomal genotypes and phase-known X-linked data (*TLR7*, *TLR8*) for all equine samples. Consensus LD patterns and blocks were estimated via majority rule from: 95% confidence intervals constructed for D′ [39, 40], application of the four gamete rule [39]) (4th gamete > 0.02), and estimates of recombination between adjacent variable sites [41, 42]. To further evaluate patterns of LD decay, pairwise r^2 values were estimated within Haploview for all validated SNPs within each equine *TLR* gene (all samples included). A minimal set of tagSNPs predicted to capture 100% of the variation ($r^2 > 0.80$) segregating in our equine discovery panel was deduced using the Tagger algorithm implemented in Haploview.

Median joining haplotype networks

Because median joining (MJ) networks require the absence of recombination [72], genes displaying evidence of historical recombination (*TLR3*, *TLR4*, *TLR8*) were partitioned

into regions of elevated LD. Haplotypes were reconstructed [38] for each intragenic region and the best haplotype pairs were used for MJ network analyses [35]. This approach improved the proportion of samples with best pairs phase probabilities ≥ 0.90 and eliminated regions displaying overt evidence of recombination. MJ networks were constructed using Network 4.5.1.0 (Fluxus Technology Ltd, Suffolk, England), with the default character weight of 10 for SNPs. Results were visualized, annotated, and manually adjusted within Network Publisher (Fluxus Technology Ltd, Suffolk, England). Specifically, branch angles were adjusted to ensure proper network magnification and clarity without changing branch lengths.

AA substitution phenotypes and evolutionary analyses

AA substitution phenotypes were predicted using PolyPhen [45] and SIFT [46] (http://genetics.bwh.harvard. edu/pph/; http://sift.jcvi.org/) with the default settings. Similar to our cattle study [85], to assess the potential for functional and/or selective constraint across the nine equine *TLR* genes investigated, a goodness of fit test (χ^2) was performed assuming equal probabilities for benign or tolerated AA phenotypes, versus those predicted to impact protein function. We also used the Seabury-Taylor test of selection [37, 85] to examine diversity across all equine *TLR* genes. Specifically, at each polymorphism we estimated the effective number of alleles as $E_i = 1/(1 - 2p_i(1-p_i)) = 1/(p_i^2 + (1 - p_i)^2) = 1/(\text{expected HWE frequency of homozygotes})$ where p_i is allele frequency at the ith locus. Thus a measure of polymorphism diversity is $\log_2(E_i)$ which also represents the information content of each SNP [37, 85]. 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 jth 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.

CHAPTER IV

CONCLUSIONS AND FURTHER INVESTIGATIONS

Bovine Conclusions

Our detailed analysis of the haplotype structure, LD architecture, and tagSNP/Indel prediction for all 10 bovine *TLR* genes will enable studies aimed at assessing the statistical and functional relationships between validated variation, and differential susceptibility to infectious disease [26-34, 53] (Table 5). Moreover, because extensive haplotype sharing was confidently predicted for specialized beef and dairy cattle breeds, the deliverables of this study will broadly impact many facets of bovine health research, including the potential for marker-assisted vaccination; using genotypes as indicator variables for enhanced vaccine design or as predictors of animal response.

In view of the emerging global interest in genomic selection in beef and dairy cattle, we provide evidence for balancing selection on at least two of the *TLR* genes (*TLR3* and *TLR8*), with detection of a weaker selective signal consistent with purifying selection in *TLR10* [37] (Table 4). Interestingly, *TLR3* and *TLR8* encode molecular sentries that recognize invading double-stranded (ds) and single-stranded (ss) RNA viruses, respectively, thereafter eliciting host innate immune responses [8, 21]. Importantly, selection on *TLR3* and *TLR8* may have direct implications on aspects of differential susceptibility to major viral production diseases such as bluetongue (dsRNA;

Reoviridae), foot and mouth disease (ssRNA; *Picornaviridae*), bovine viral diarrhea (ssRNA; *Flaviviridae*), calf coronavirus (ssRNA; neonatal diarrhea; *Coronaviridae*),and bovine parainfluenza 3 (ssRNA; *Paramyxoviridae*) (see [62, 63]). Moreover, evolution under repeated exposure to many of these diseases may provide some explanation for the observed patterns of variation detected within *TLR3* and *TLR8*. However, it is also possible that more ancient host-pathogen interactions (i.e., eradicated Rinderpest, ssRNA, *Paramyxoviridae*; etc) may have contributed to the signatures of selection detected in this study. It should also be noted that because frequency distribution tests generally lack power to detect selection [58], departures from neutrality noted in this study are likely to underscore the strength of the selective signals observed (for review see [64]). For these reasons, future studies involving all species of the subfamily Bovinae are needed to help elucidate whether selective signals in *TLR3* and *TLR8* extend beyond modern domestic cattle lineages. Moreover, variation within these genes should be comprehensively evaluated with respect to differences in ligand-induced signaling, disease susceptibility, and the potential for marker-assisted vaccination in domestic cattle.

In addition to selective signals observed for *TLR3* and *TLR8*, several tentative associations were detected between bovine *TLR* SNPs (Table 5) and differential susceptibility to MAP infection which have not previously been reported, with one implicated locus (*TLR10*) also exhibiting evidence of purifying selection (Table 4) [37]. However, because the natural ligand(s) for *TLR10* have yet to be comprehensively

59

elucidated, the precise origin of this selective signal remains unclear. Previous studies [22, 65] indicate that human *TLR10* forms functional heterodimers with both *TLR2* and *TLR1*, thereby enabling the resulting protein complexes to recognize a wide variety of microbial ligands [65], including those derived from *Mycobacteria* [8, 21, 23, 66]. Similarly, *TLR2* is also known to form functional heterodimers with *TLR6* [23]. Recently, AA substitutions in human *TLR1* and *TLR10* were demonstrated to negatively impact receptor function [65, 66], with *TLR10* ligand recognition similar to the known range of ligands established for *TLR1* [65]. The results of our single marker association tests indirectly support the biological concept of functional unity with respect to bovine *TLR2*, *TLR6*, and *TLR10*, with variation at all three loci categorically linked to a common microbial phenotype (bacterial culture status for MAP) in Holstein cattle.

Equine Conclusions

Detailed characterization and validation of naturally occurring genetic variation within nine members of the equine *TLR* gene family provided a natural segue for elucidating equine *TLR* haplotype structure, LD architecture, and tagSNPs that may help reduce genotyping costs in future studies. Moreover, given the robust signatures of selection detected for some *TLR*s in cattle, we also aimed to determine whether similar signatures existed within the equine *TLR* genes. Our analysis of haplotype structure demonstrated evidence for haplotype sharing across all equine samples for a majority of the investigated genes, thereby indicating that our research is likely to be very applicable to a diverse variety of horse and pony breeds as well as the donkey. Moreover, future

studies focusing on equine health traits and vaccine studies are likely to make use of the validated SNPs, inferred haplotype structure, and tagSNPs elucidated herein. Notably, the mammalian *TLR*s have already shown tangible potential as innate immunologicals used as anti-infectives [1], and variation within these genes may become important for marker-assisted vaccination and/or marker assisted breeding. Importantly, the Seabury-Taylor test of selection demonstrated that equine *TLR3*, *TLR7*, and *TLR8* all displayed significantly less diversity than the other investigated loci (Figure 10), which is most likely due to purifying or directional selection. Our amino acid modeling analyses, which included a goodness of fit test designed to approximate the expectations of a strictly neutral model, provided ample statistical support for functional and/or selective constraint(s) across all 9 equine *TLR* genes, with an underlying trend that included significantly fewer amino acid replacements predicted to impact protein function than expected. Notably, the protein products of *TLR3*, *TLR7*, and *TLR8* are capable of detecting either ds- (*TLR3*) or ssRNA viruses (*TLR7*, *TLR8*) [8, 21], with signatures of selection detected in this study that may potentially have manifested by way of historic and/or contemporary exposures to specific equine viral diseases.

Future Investigations

Herein, we provide evidence for selection (i.e. natural and/or manual) for members of the bovine and equine *TLR* gene family. The precise biological and temporal origins of these signals are currently unknown. Therefore, future studies involving phylogenetic approaches involving the inference of ancestral *TLR* sequences [96], with a variety of

61

terminal taxa representing extant members of Equidae and Bovidae are needed to help clarify the evolutionary history of selection within the *TLR* gene family.

In order to explore the putative biological impact of both selection as well as discrete amino acid replacements encoded by naturally occurring variation within the bovine and equine *TLR* genes, it is possible that a species-specific cell culture system with reporter-style assays [13, 97-99] expressing a variety of naturally occurring *TLR*s haplotypes possessing validated variation would provide key insight regarding heritable differences in ligand-induced signaling. Notably, purified ligands for all of the mammalian *TLR* genes are commercially available. This approach is likely to elucidate key protein domains and amino acid positions that are functionally intolerant to some naturally occurring genetic variants, including domains and amino acid residues that have not historically been linked to receptor function (i.e. low complexity, intrinsic disorder, etc). Moreover, complex bovine and equine *TLR* haplotypes possessing multiple missense SNPs may have an additive, compensatory, or antagonistic effect on PAMP recognition, and therefore, should be thoroughly evaluated. Information gained from functional studies using reporter-style assays will further inform modern genomic selection and marker-assisted vaccination strategies leading to enhanced livestock health and production.

REFERENCES

1. Rosenthal KL (2006) Tweaking innate immunity: The promise of innate immunologicals as anti-infectives. Can J Infect Dis Med Microbiol 17: 307–314.

2. Vasselon T, Detmers PA (2002) Toll Receptors: A central element in innate immune responses. Infect Immun 70: 1033–1041.

3. 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.

4. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell. 1996, 86 : 973-983

5. Tauszig S, Jouanguy E, Hoffman 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.

6. Beutler B (2004) Inferences, questions, and possibilities in Toll-like receptor signaling. Nature 430: 257-263.

7. Kaisho T, Akira S (2006) Toll-like receptor function and signaling. J Allergy Clin Immunol 117: 979–987.

8. Akira S, Takeda K (2004) Toll-like receptor signaling. Nat Rev Immunol 4: 499–511.

9. Schroder NWJ, Schumann RR (2005) Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Lancet Infect Dis 5:156-164.

10. Texereau J, Chiche JD, Taylor W, Choukroun G, Comba B, et al. (2005) The importance of Toll-like receptor 2 polymorphisms in severe infections. Clin Infect Dis 41 Suppl 7: S408–S415.

11. Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG, et al. (2007) Full-exon resequencing reveals Toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS ONE 12: e1318 (Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0001318. Accessed 2012 June 7).

12. Smit LAM, Bongers SIM, Ruven HJT, Rijkers GT, Wouters IM, et al. (2007) Atopy and new-onset asthma in young Danish farmers and CD14, TLR2, and TLR4 genetic polymorphisms: A nested case-control study. Clin Exp Allergy 37: 1602-1608.

13. Merx S, Zimmer W, Neumaier M, Ahmad-Nejad PA (2006) Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. Hum Mutat 27: 293.

14. Kataria RS, Tait Jr. RG, Kumar D, Ortega MA, Rodiguez J, et al. (2011) Association of toll-like receptor four single nucleotide polymorphisms with incidence of infectious bovine keratoconjunctivitis (IBK) in cattle. Immunogenetics 63: 115–119.

15. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, et al. (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92: 16–24.

16. Bovine Genome Sequencing, Analysis Consortium, Elsik CG, Tellam RL, Worley KC (2009) The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324: 522–528.

17. Bovine HAPMAP Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324: 529–523.

18. Plain KM, Purdie AC, Begg DJ, de Silva K, Whittington RJ (2010) Toll-like receptor (TLR) 6 and TLR1 differentiation in gene expression studies of Johne's disease. Vet Immunol Immunopathol 137: 142–148.

19. Jann OC, King A, Corrales NL, Anderson SI, Jensen K, et al. (2009) Comparative genomics of Toll-like receptor signaling in five species. BMC Genomics 10: 216.

20. Glass EJ, Baxter R, Leach RJ, Jann OC (2011) Genes controlling vaccine responses and disease resistance to respiratory viral pathogens in cattle. Vet Immunol Immunopathol. In press.

21. West AP, Koblansky AA, Ghosh S (2006) Recognition and signaling by Toll-Like receptors. Annu Rev Cell Dev Biol 22: 409–437.

22. HasanU,Chaffois C,Gaillard C, SaulnierV,Merck E, 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.

23. Ozinsky A, Underhill DM, Fontenot JD, Hajjar AM, Smith KD, 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 U S A 97: 13766–13771.

24. 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.

25. Govindarai RG, Manavalan B, Lee G, Choi S (2010) Molecular modeling-based evaluation of hTLR10 and identification of potential ligands in Toll-like receptor signaling. PLoS ONE 5(9):e12713. (Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943521/?tool=pubmed. Accessed 2011 Jul 3).

26. 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.

27. Bhide MR, Mucha R, Mukula I Jr., Kisova L, Skrabana R, et al. (2009) Novel mutations in TLR genes cause hyporesponsiveness to *Mycobacterium avium* subsp. *paratuberculosis* infection. BMC Genet 10: 21.

28. Pinedo PJ, Buergelt CD, Donovan GA, Melendez P, Morel L, et al. (2009) Candidate gene polymorphisms (BoIFNG, TLR4, SLC11A1) as risk factors for paratuberculosis infection in cattle. Prev Vet Med 91: 189–196.

29. Pinedo PJ, Wang C, Li Y, Rae DO, Wu R (2009) Risk haplotype analysis for bovine paratuberculosis. Mamm Genome 20: 124–129.

30. Cargill EJ, Womack JE (2007) Detection of polymorphisms in bovine toll-like receptors 3, 7, 8, and 9. Genomics 89: 745–755.

31. 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.

32. Seabury CM, Womack JE (2008) Analysis of sequence variability and protein domain architectures for bovine peptidoglycan receptor protein 1 (PGLYRP1) and Toll-Like Receptors 2 and 6. Genomics 92: 235–245.

33. Kuhn CH, Bennetwitz J, Reinsch N, Xu N, Thomsen H, et al. (2003) Quantitative trait loci mapping of functional traits in the German Holstein cattle population. J Dairy Sci 86: 360–368.

34. Heyen DW, Weller JI, Ron M, Ban M, Beever JE, et al. (1999) A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol Genomics 1: 165–175.

35. Bandelt HJ, Forster P, Rohl A (1999) Median joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37–48.

36. Seabury CM, Bhattarai EK, Taylor JF, Viswanathan GG, Cooper SM, et al. (2011) Genome-wide polymorphism and comparative analyses in the whitetailed deer (*Odocoileus virginianus*): A model for conservation genomics. PLoS ONE 6: e15811.

37. Seabury CM, Seabury PM, Decker JE, Schnabel RD, Taylor JF, et al. (2010) Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. Proc Natl Acad Sci U S A 107: 151–156.

38. StephensM, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.

39. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

40. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. Science 296: 2225–2229.

41. Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165: 2213–2233.

42. Crawford D, Bhangale T, Li N, Hellenthal G, Rieder MJ, et al. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet 36: 700–706.

43. 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.

44. Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5: 247–252.

45. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: Server and survey. Nucleic Acids Res 30: 3894–3900.

46. Kumar P, Henikoff S, Ng P (2009) Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

47. Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genom Human Genet 7: 61–80.

48. Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, et al. (2003) Widespread purifying selection at polymorphic sites in human protein-coding loci. Proc Natl Acad Sci U S A 100: 15754–15757.

49. 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.

50. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

51. Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. Genetics 147: 915–925.

52. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B 57: 289–300.

53. Ruiz-Larranaga O, Manzano C, Iriondo M, Garrido JM, Molina E, et al. (2011) Genetic variation of toll-like receptor genes and infection by *Mycobacterium avium* ssp. *paratuberculosis* in Holstein-Freisian cattle. J Dairy Sci 94: 3635–3641.

54. Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, et al. (2009) High-resolution haplotype block structure in the cattle genome. BMC Genet. 10: 19.

55. Hiwatashi T, Okabe Y, Tsutsui T, Hirmatsu C, Melin AD, et al. (2010) An explicit signature of balancing selection for color-vision variation in new world monkeys. Mol Biol Evol 27: 453–464.

56. Tennessen JA, Blouin MS (2008) Balancing selection at a frog antimicrobial peptide locus: Fluctuating immune effector alleles? Mol Biol Evol 25: 2669–2680.

57. Osier FH, Weedall GD, Verra F, Murungi L, Tetteh KK, et al. (2010) Allelic diversity and naturally acquired allele-specific antibody responses to *Plasmodium falciparum* apical membrane antigen 1 in Kenya. Infect Immun 78: 4625–4633.

58. Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141: 413–429.

59. Shey MS, Randhawa AK, Bowmaker M, Smith E, Scriba TJ, et al. (2010) Single nucleotide polymorphisms in Toll-like receptor 6 are associated with altered lipopeptide-and mycobacteria-induced interleukin-6 secretion. Genes Immun 11: 561–572.

60. Zanella R, Settles ML, McKay SD, Schnabel R, Taylor J, et al. (2010) Identification of loci associated with tolerance to Johne's disease in Holstein cattle. Anim Genet 42: 28–38.

61. Neibergs HL, Settles ML, Whitlock RH, Taylor JF (2010) GSEA-SNP identifies genes associated with Johne's disease in cattle. Mamm Genome 21: 419–425.

62. Fauquet C, Mayo MA, Maniloff J, Desselberger U, Ball LA (2005) Virus Taxonomy. London: Elsevier Academic Press. Pages 441-1128.

63. Cahn CM, Line S, eds (2005) The Merck Veterinary Manual. Whitehouse Station: Merck Sharp & Dohme Corp. Pages 441-1128.

64. Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, et al. (2002) A strong signature of balancing selection in the 59 cis-regulatory region of CCR5. Proc Natl Acad Sci U S A 99: 10539–10544.

65. Guan T, Ranoa DR, Jiang S, Mutha SK, Li X, et al. (2010) Human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. J Immunol 184: 5094–5103.

66. Uciechowski P, Imhoff H, Lange C, Meyer CG, Browne EN, et al. (2011) Susceptibility to tuberculosis is associated with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. J Leukoc Biol 90: 377–388.

67. 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.

68. Ingman M, Gyllensten U (2009) SNP frequency estimation using massively parallel sequencing of pooled DNA. Eur J Hum Genet 17: 383–386.

69. Altshuler D, Pollara VJ, Cowles CR, Etten Van WJ, Baldwin J, et al. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407: 513–516.

70. Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, et al. (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res 18: 763–770.

71. Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73: 1162–1169.

72. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, et al. (2006) A comparison of phasing algorithms for trios and unrelated individuals. Am J Human Genet 78: 437–450.

73. Posada D, Crandall KA (2001) Intraspecific gene genealogies: Trees grafting into networks. Trends Ecol Evol 16: 37–45.

74. Rozas J (2009) DNA sequence polymorphism analysis using DnaSP. Methods Mol Biol 537: 337–350.

75. Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. Genet Res 50: 245–250.

76. Pradhan AK, Mitchell RM, Kramer AJ, Zurakowski MJ, Fyock TL, et al. (2011) Molecular epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* in a longitudinal study of three dairy herds. J Clin Microbiol. 49: 893–901.

77. Raudsepp T, Gustafson-Seabury A, Durkin K, Wagner ML, Goh G, et al. (2008) A 4,103 marker integrated physical and comparative map of the horse genome. Cytogenet Genome Res 122: 28-36.

78. Berndt A, Derksen FJ, Venta PJ, Ewart S, Yuzbasiyan-Gurkan V, et al. (2007) Elevated amount of Toll-like receptor 4 mRNA in bronchial epithelial cells is associated with airway inflammation in horses with recurrent airway obstruction. Am J Physiol Lung Cell Mol Physiol 292:L936-943.

79. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326: 865-867.

80. McCue ME, Bannash DL, Petersen JL, Gurr J, Bailey E (2012) A high density SNP array for the domestic horse and extant Perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet 8(1): e1002451. (Available: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002451. Accessed 2012 June 7).

81. Pulos WL, Hutt FB (1969) Lethal dominant white in horses. J Hered 60: 59-63.

82. Hill EW, McGivney BA, Gu J, Whiston R, Machugh DE (2010) A genome-wide SNP-association study confirms a sequence variant (g.66493737C > T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. BMC Genomics 11:552.

83. Corbin LJ, Blott SC, Swinburne JE, Sibbons C, Fox-Clipsham LY (2012) A genome-wide association study of osteochondritis dissecans in the Thoroughbred. Mamm Genome 23: 294-303.

84. Haves BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92: 433-443.

85. Fisher CA, Bhattarai EK, Osterstock JB, Dowd SE, Seabury PM (2011) Evolution of the bovine TLR gene family and member associations with *Mycobacterium avium* subspecies *paratuberculosis* infection. PLoS ONE 6(11): e27744. (Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0027744. Accessed 2012 June 7).

86. Singh Suri S, Janardhan KS, Parbhakar O, Caldwell S, Appleyard G, et al. (2006) Expression of toll-like receptor 4 and 2 in horse lungs. Vet Res 37: 541-551.

87. Gornik K, Moore P, Figueiredo M, Vandenplas M (2011) Expression of Toll-like receptors 2, 3, 4, 6, 9 and MD-2 in the normal equine cornea, limbus, and conjunctiva. Vet Opthalmol 14: 80-85.

88. Kwon S, Vandenplas ML, Figueiredo MD, Salter CE, Andrietti AL, et al. (2010) Differential induction of Toll-like receptor gene expression in equine monocytes activated by Toll-like receptor ligands or TNF-α. Vet Immunol Immunopathol 138: 213-217.

89. Astakhova NM, Perelygin AA, Zharkikh AA, Lear TL, Coleman SJ, et al. (2009) Characterization of equine and other vertebrate TLR3, TLR7, and TLR8 genes. Immunogenetics 61: 529-539.

90. Richez C, Blanco P, Rifkin I, Moreau JF, Schaeverbeke T (2011) Role for toll-like receptors in autoimmune disease: The example of systemic lupus erythematosus. Joint Bone Spine 78: 124-130.

91. Kumar S, Hedges SB (2011) TimeTree2: Species divergence times on the iPhone. Bioinformatics 27: 2023-2024.

92. Chowdhary BP, Raudsepp T (2008) The horse genome derby: Racing from map to whole genome sequence. Chromosome Res 16: 109-127.

93. Bailey E, Blinns MM (1998) The horse gene map. ILAR J 39: 171-176.

94. Boichard D, Chung H, Dassonneville R, David X, Eggen A, et al. (2010) Design of a bovine low-density SNP array optimized for imputation. PLoS One 7(3): e34130. (Available:

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0034130. Accessed 2012 June 7).

95. Magee DA, Park SD < Scraggs E, Murphy AM, Doherty ML, et al. (2010) Technical note: High fidelity of whole-genome amplified sheep (*Ovis aries*) deoxyribonucleic acid using a high-density single nucleotide polymorphism array-based genotyping platform. J Anim Sci 88: 3183-3186.

96. 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.

97. Merx S, Neumaier M, Wagner H, Kirschning CJ, Ahmad-Nejad P (2007) Characterization and investigation of single nucleotide polymorphisms and a novel TLR2 mutation in the human TLR2 gene. Hum Mol Genet 16: 1225-1232.

98. Gewirtz AT, Vijay-Kumar M, Brant SR, Duerr RH, Nicolae DL, et al. (2006) Dominant-negative TLR5 polymorphism reduces the adaptive immune response to flagellin and negatively associates with Crohn's disease. Am J Physiol Gastrointest Liver Physiol 290: G1157-1163.

99. Schröder NW, Diterich I, Zinke A, Eckert J, Draing C, et al. (2005) Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immuneactivation by *Borrelia burgdorferi* and protects from late stage Lyme disease. J Immunol 175: 2534-2540.

Table A1. Normalization Protocol

Amplicon	Size(BP)	Mean	Difference	Adjust	% Diff	Adjust	Add ul
TLR1-1	470	624.943662	154.943662	Yes	0.247932208	1.504135584	1.5 ul
TLR1-2	890	624.943662	-265.056338	Yes	0.424128372	2.848256744	2.85 ul
TLR2-1	816	624.943662	-191.056338	Yes	0.305717698	2.611435397	2.61 ul
TLR2-2	668	624.943662	-43.05633803				2.0 ul
TLR2-3	681	624.943662	-56.05633803	Yes	0.089698226	2.179396453	2.18 ul
TLR2-4	774	624.943662	-149.056338	Yes	0.23851164	2.477023281	2.48 ul
TLR2-5	730	624.943662	-105.056338	Yes	0.168105294	2.336210588	2.34ul
TLR2-6	436	624.943662	188.943662	Yes	0.302337112	1.395325776	1.4 ul
TLR3-6	598	624.943662	26.94366197				2.0 ul
TLR3-7	669	624.943662	-44.05633803				2.0 ul
TLR3-8	629	624.943662	-4.056338028				2.0 ul
TLR3-9	528	624.943662	96.94366197	Yes	0.155123842	1.689752316	1.69 ul
TLR3-10	527	624.943662	97.94366197	Yes	0.156723986	1.686552027	1.69 ul
TLR3-11	597	624.943662	27.94366197				2.0 ul
TLR3-12	612	624.943662	12.94366197				2.0 ul
TLR3-14	636	624.943662	-11.05633803				2.0 ul
TLR3-15	641	624.943662	-16.05633803				2.0 ul
TLR3-16	701	624.943662	-76.05633803	Yes	0.121701111	2.243402222	2.24 ul
TLR3-17	635	624.943662	-10.05633803				2.0 ul
TLR3-18	577	624.943662	47.94366197				2.0 ul
TLR3-19	590	624.943662	34.94366197				2.0 ul
TLR3-20	518	624.943662	106.943662	Yes	0.171125285	1.657749431	1.65 ul
TLR3-22	509	624.943662	115.943662	Yes	0.185526583	1.628946835	1.63 ul
TLR3-23	437	624.943662	187.943662	Yes	0.300736968	1.398526064	1.40 ul
TLR4-1	288	624.943662	336.943662	Yes	0.539158459	0.921683081	0.92 ul
TLR4-2	384	624.943662	240.943662	Yes	0.385544612	1.228910775	1.22 ul
TLR4-3	486	624.943662	138.943662	Yes	0.2223299	1.5553402	1.55 ul
TLR4-4	508	624.943662	116.943662	Yes	0.187126727	1.625746546	1.63 ul
TLR4-5	541	624.943662	83.94366197	Yes	0.134321967	1.731356066	1.73 ul
TLR4-6	539	624.943662	85.94366197	Yes	0.137522256	1.724955489	1.72 ul
TLR4-7	554	624.943662	70.94366197	Yes	0.113520092	1.772959816	1.77 ul
TLR4-8	533	624.943662	91.94366197	Yes	0.147123121	1.705753758	1.70 ul
TLR4-9	535	624.943662	89.94366197	Yes	0.143922832	1.712154335	1.71 ul
TLR5-1	642	624.943662	-17.05633803				2.0 ul
TLR5-2	661	624.943662	-36.05633803				2.0 ul
TLR5-3	563	624.943662	61.94366197	Yes	0.099118794	1.801762412	1.80 ul
TLR5-4	541	624.943662	83.94366197	Yes	0.134321967	1.731356066	1.73 ul
TLR5-5	687	624.943662	-62.05633803	Yes	0.099299092	2.198598184	2.20 ul
TLR5-6	700	624.943662	-75.05633803	Yes	0.120100967	2.240201934	2.24 ul
TLR5-8	592	624.943662	32.94366197				2.0 ul
TLR5-9	541	624.943662	83.94366197	Yes	0.134321967	1.731356066	1.73 ul
TLR5-10FixR	764	624.943662	-139.056338	Yes	0.222510198	2.445020396	2.45 ul
TLR6-1	876	624.943662	-251.056338	Yes	0.401726353	2.803452706	2.80 ul
TLR6-2	805	624.943662	-180.056338	Yes	0.288116112	2.576232224	2.58 ul
TLR6-3	845	624.943662	-220.056338	Yes	0.352121881	2.704243763	2.70 ul
TLR6-4	604	624.943662	20.94366197				2.0 ul
TLR7-1	715	624.943662	-90.05633803	Yes	0.14410313	2.288206261	2.29 ul
TLR7-2	851	624.943662	-226.056338	Yes	0.361722747	2.723445494	2.72 ul
TLR7-3	892	624.943662	-267.056338	Yes	0.427328661	2.854657321	2.85 ul
TLR7-4	822	624.943662	-197.056338	Yes	0.315318564	2.630637128	2.63 ul
TLR7-5	871	624.943662	-246.056338	Yes	0.393725632	2.787451263	2.79 ul
TLR7-6	669	624.943662	-44.05633803				2.0 ul
TLR8-1	583	624.943662	41.94366197				2.0 ul
TLR8-2	659	624.943662	-34.05633803				2.0 ul
TLR8-3	601	624.943662	23.94366197				2.0 ul

TLR8-4	592	624.943662	32.94366197				2.0 ul
TLR8-6	526	624.943662	98.94366197	Yes	0.158324131	1.683351739	1.68 ul
TLR8-7	665	624.943662	-40.05633803				2.0 ul
TLR9-1	413	624.943662	211.943662	Yes	0.33914043	1.321719141	1.32 ul
TLR9-2	414	624.943662	210.943662	Yes	0.337540285	1.324919429	1.32 ul
TLR9-3	565	624.943662	59.94366197	Yes	0.095918505	1.808162989	1.80 ul
TLR9-4	537	624.943662	87.94366197	Yes	0.140722544	1.718554912	1.72 ul
TLR9-5	499	624.943662	125.943662	Yes	0.201528025	1.59694395	1.60 ul
TLR9-6	718	624.943662	-93.05633803	Yes	0.148903563	2.297807126	2.30 ul
TLR9-7	651	624.943662	-26.05633803				2.0 ul
TLR9-8	546	624.943662	78.94366197	Yes	0.126321246	1.747357508	1.75 ul
TLR9-9	428	624.943662	196.943662	Yes	0.315138266	1.369723468	1.37 ul
TLR9-10	837	624.943662	-212.056338	Yes	0.339320727	2.678641455	2.68 ul
TLR10-2	768	624.943662	-143.056338	Yes	0.228910775	2.45782155	2.46 ul
TLR10-3	623	624.943662	1.943661972				2.0 ul
TLR10-6	868	624.943662	-243.056338	Yes	0.388925199	2.777850398	2.78 ul
Average	624.943662	624.943662					
Median	604						
Stand Dev	134.9570183						
Half Stand Dev	67.47850914						
	Edges						
Left	Right						
489.9866437	759.9006802						
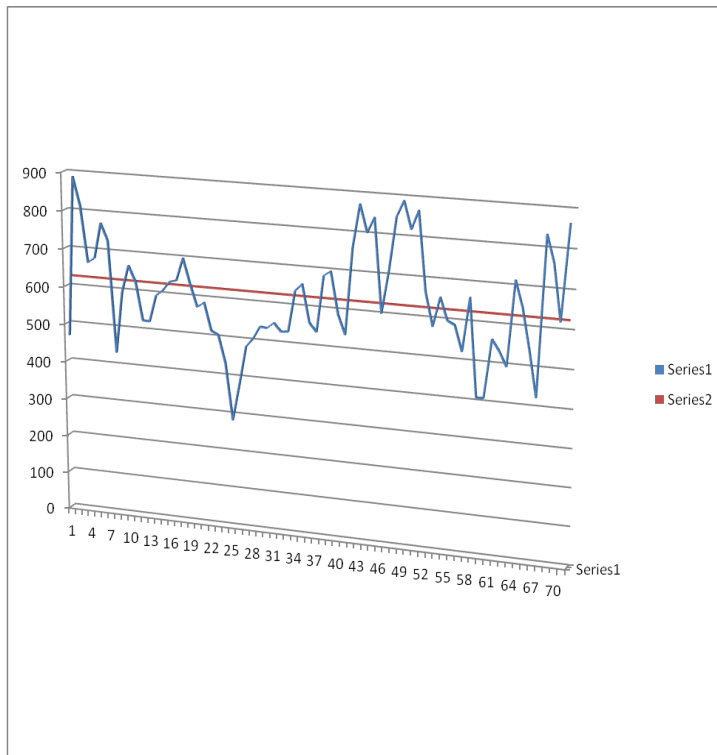
Figure A1. Graph of amplicon sizes.

Table A2. Validated SNPs and Indels.

Gene_SNP or Indel	Variation	BTA	dbSNP	Non-synonymous	MAF all	MAF taurus
TLR1_A42G	A/G	6	rs55617254	X	0.010	0.004
TLR1_G77A	G/A	6	ss469376045	X	0.002	0.000
TLR1_G363A	G/A	6	rs55617441		0.002	0.000
TLR1_C603T	C/T	6	rs43702940		0.455	0.434
TLR1_G1521A	G/A	6	rs55617193		0.430	0.430
TLR2_A9163G	A/G	18	rs68343161		0.014	0.000
TLR2_-9214TG	-/TG	18	rs68268241		0.065	0.039
TLR2_A9399G	A/G	18	ss470256470		0.003	0.002
TLR2_G9416A	G/A	18	ss470256472		0.002	0.002
TLR2_G9431A	G/A	18	rs68343163[a]		0.001	0.001
TLR2_C9564T	C/T	18	rs68268245		0.177	0.185
TLR2_C9570T	C/T	18	ss470256473		0.003	0.001
TLR2_G9579A	A/G	18	rs68343166		0.046	0.040
TLR2_A9589C	A/C	18	rs68268246		0.066	0.041
TLR2_G9644A	G/A	18	ss470256474		0.001	0.001
TLR2_C9708T	C/T	18	rs68268248		0.069	0.043
TLR2_G10018A	G/A	18	ss470256475	X	0.004	0.004
TLR2_C10047T	C/T	18	ss470256476[a]		0.002	0.001
TLR2_C10077T	C/T	18	ss470256477		0.004	0.004
TLR2_T10095C	C/T	18	rs68268249		0.068	0.042
TLR2_G10098T	G/T	18	rs55617172	X	0.337	0.318
TLR2_G10111A	A/G	18	rs68268250	X	0.056	0.039
TLR2_G10265T	G/T	18	ss470256478	X	0.008	0.003
TLR2_G10364A	A/G	18	rs43706434	X	0.171	0.178
TLR2_G10511A	G/A	18	ss470256479	X	0.158	0.167
TLR2_A10540G	A/G	18	rs43706433	X	0.338	0.316
TLR2_T10590A	A/T	18	rs68268251	X	0.015	0.000
TLR2_C10841A	G/T	18	ss470256481	X	0.006	0.007
TLR2_G10854T	G/T	18	rs68268253		0.048	0.041
TLR2_T10887A	A/T	18	rs68343167	X	0.067	0.041
TLR2_G10919A	A/G	18	rs68343168	X	0.067	0.041
TLR2_A10938C	A/C	18	rs68268254		0.003	0.000
TLR2_A11117G	A/G	18	ss470256482	X	0.004	0.004
TLR2_C11123T	C/T	18	rs68268255	X	0.005	0.000
TLR2_A11159G	A/G	18	rs68268256	X	0.071	0.044
TLR2_A11217C	A/C	18	rs68268257		0.068	0.041
TLR2_C11363T	C/T	18	ss470256483	X	0.015	0.016
TLR2_T11413G	G/T	18	rs68268258	X	0.005	0.000
TLR2_T11541C	C/T	18	rs68268259		0.150	0.159
TLR2_G11597A	A/G	18	rs68268260	X	0.066	0.039
TLR2_T11616C	C/T	18	rs41830058		0.235	0.216
TLR2_C11723T	C/T	18	rs68343170	X	0.010	0.000
TLR2_C11748T	C/T	18	rs68268262		0.003	0.000
TLR2_C11904G	C/G	18	rs68268263	X	0.068	0.042
TLR2_T11934C	C/T	18	rs68343171		0.068	0.041
TLR2_T11964C	C/T	18	rs68268264		0.068	0.041
TLR2_G12033C	C/G	18	rs68268265		0.005	0.003
TLR2_G12121C	G/C	18	ss470256484[a]	X	0.001	0.000
TLR2_G12123A	A/G	18	rs68268266		0.068	0.041
TLR2_C12204T	C/T	18	rs68268267		0.066	0.040
TLR3_A580G	A/G	27	rs42851894		0.484	0.485
TLR3_A697G	A/G	27	rs42851895		0.438	0.500
TLR3_A739T	A/T	27	rs42851896		0.281	0.338
TLR3_T746C	T/C	27	rs42851897		0.409	0.310
TLR3_G753A	G/A	27	rs42851898		0.453	0.485
TLR3_C764T	C/T	27	rs55617276		0.255	0.300
TLR3_G835A	G/A	27	rs55617196		0.011	0.000
TLR3_C961T	C/T	27	rs55617242		0.010	0.000
TLR3_A1152G	A/G	27	rs55617217		0.010	0.000
TLR3_A1285C	A/C	27	rs42851900		0.271	0.323
TLR3_T1315C	T/C	27	rs42851901		0.284	0.336
TLR3_G1565A	G/A	27	rs55617186		0.068	0.000

TLR3_G2042T	G/T	27	rs42851909		0.284	0.336
TLR3_C2060T	C/T	27	rs42851910		0.279	0.331
TLR3_G2107T	G/T	27	rs42851911		0.279	0.336
TLR3_G2185A	G/A	27	rs42851912		0.411	0.328
TLR3_G2509A	G/A	27	rs42851914		0.268	0.300
TLR3_T2512A	T/A	27	rs55617216		0.089	0.000
TLR3_G2608C	G/C	27	rs42851915		0.286	0.331
TLR3_T2750C	T/C	27	ss469376047		0.016	0.000
TLR3_A3081G	A/G	27	rs55617184		0.094	0.008
TLR3_T3344C	T/C	27	rs55617229		0.410	0.325
TLR3_G3345A	G/A	27	ss469376049		0.016	0.023
TLR3_C3381G	C/G	27	rs55617207		0.089	0.000
TLR3_C3435A	C/A	27	rs55617345		0.279	0.331
TLR3_T3458C	T/C	27	ss469376051		0.021	0.000
TLR3_A3610G	A/G	27	rs55617271		0.089	0.000
TLR3_T3624C	T/C	27	ss469376053[a]		0.005	0.000
TLR3_G3741A	A/G	27	rs55617234		0.068	0.000
TLR3_C3762G	C/G	27	rs42851919		0.279	0.336
TLR3_G3804A	G/A	27	ss469376055[a]		0.005	0.000
TLR3_T3954C	T/C	27	rs42851920		0.279	0.336
TLR3_G4086A	T/C	27	rs42851921		0.281	0.338
TLR3_T4328C	T/C	27	rs55617462		0.057	0.062
TLR3_C4332T	C/T	27	rs55617278		0.266	0.320
TLR3_C4633T	C/T	27	rs42851922		0.281	0.331
TLR3_G4783A	G/A	27	ss469376057		0.274	0.331
TLR3_G5201A	G/A	27	ss469376059[a]		0.005	0.008
TLR3_G5304A	G/A	27	rs42851924		0.234	0.277
TLR3_C5350T	C/T	27	rs42851925		0.197	0.230
TLR3_C5765G	C/G	27	rs55617222		0.135	0.008
TLR3_A6281G	A/G	27	rs42851929		0.250	0.312
TLR3_C6382A	C/A	27	ss469376061[a]		0.005	0.000
TLR3_C6707T	C/T	27	rs42852432		0.302	0.385
TLR3_T7039G	T/G	27	rs42852435		0.453	0.392
TLR3_A8009G	A/G	27	rs55617204	X	0.021	0.000
TLR3_G8270A	G/A	27	rs55617272	X	0.058	0.055
TLR3_A8902G	A/G	27	rs42852438		0.426	0.385
TLR3_G8985T	G/T	27	rs42852439	X	0.432	0.385
TLR3_G9079A	G/A	27	rs42852440		0.312	0.385
TLR3_A9586G	A/G	27	rs42852441		0.302	0.385
TLR3_C9704T	C/T	27	rs55617164		0.120	0.000
TLR3_G9739A	G/A	27	rs55617241		0.229	0.292
TLR3_C10467T	C/T	27	rs55617451		0.286	0.346
TLR3_C10848A	C/A	27	rs55617344		0.078	0.000
TLR3_A10859C	A/C	27	rs55617353		0.422	0.454
TLR4_G374A	G/A	8	ss469376063		0.004	0.003
TLR4_A534C	A/C	8	rs8193042		0.033	0.023
TLR4_C539A	C/A	8	rs8193043		0.005	0.000
TLR4_T545C	T/C	8	rs8193044		0.012	0.001
TLR4_T610C	T/C	8	ss469376065		0.109	0.113
TLR4_T5054G	T/G	8	rs8193045		0.017	0.001
TLR4_C5086T	T/C	8	ss470682348		0.030	0.031
TLR4_A5088G	G/A	8	rs8193046		0.497	0.488
TLR4_G5135A	G/A	8	rs8193047		0.015	0.001
TLR4_G8000A	G/A	8	ss469376069		0.498	0.486
TLR4_C8166T	C/T	8	ss469376071[a]		0.001	0.000
TLR4_A8219C	A/C	8	rs8193049	X	0.009	0.004
TLR4_C8807A	C/A	8	rs8193053	X	0.005	0.001
TLR4_A8886G	A/G	8	rs8193054		0.021	0.002
TLR4_A8909G	A/G	8	rs8193055	X	0.005	0.002
TLR4_T8934G	T/G	8	rs8193057		0.016	0.000
TLR4_A9288G	A/G	8	rs8193059		0.022	0.001
TLR4_C9423T	T/C	8	rs8193060		0.356	0.346
TLR4_A9463C	A/C	8	ss469376073	X	0.003	0.002
TLR4_A9527G	A/G	8	ss469376075	X	0.003	0.002
TLR4_T9534C	T/C	8	rs8193061		0.007	0.000
TLR4_T9594C	T/C	8	rs8193062		0.005	0.001

75

TLR4_G9715A	G/A	8	rs8193066	X	0.008	0.002
TLR4_C9759A	C/A	8	rs8193067		0.004	0.001
TLR4_C9788T	C/T	8	rs8193069	X	0.107	0.107
TLR4_T9795C	T/C	8	rs8193070		0.010	0.000
TLR4_C9990T	C/T	8	ss469376077		0.002	0.001
TLR4_T10308C	T/C	8	rs8193071		0.003	0.003
TLR5_C159A	C/A	16	rs55617268		0.385	0.364
TLR5_A314G	A/G	16	rs55617159		0.388	0.366
TLR5_C322T	C/T	16	rs55617368		0.361	0.363
TLR5_365Y	T/C	16	rs55617149		0.015	0.002
TLR5_-374C	-/C	16	rs55617312		0.378	0.362
TLR5_C580T	C/T	16	rs55617365		0.386	0.364
TLR5_C584T	C/T	16	rs55617173		0.386	0.364
TLR5_-628CTCCTTCTGATCAGCTGTAAATTGT	-/25bp	16	rs55617435		0.386	0.371
TLR5_T753G	T/G	16	ss469376079		0.006	0.001
TLR5_T788G	T/G	16	ss469376081		0.005	0.000
TLR5_G1137C	G/C	16	rs55617208		0.380	0.363
TLR5_T1163C	T/C	16	ss469376083		0.006	0.001
TLR5_G1189T	G/T	16	rs55617262		0.381	0.364
TLR5_C1254T	C/T	16	rs55617167		0.381	0.364
TLR5_-1369C	-/C	16	rs55617256		0.280	0.279
TLR5_C1498T	C/T	16	rs55617141		0.006	0.001
TLR5_C1562T	C/T	16	ss469376085[a]		0.006	0.006
TLR5_G1598T	G/T	16	rs55617358		0.006	0.001
TLR5_G1650A	G/A	16	rs55617432		0.386	0.365
TLR5_G1685A	G/A	16	ss469376087[a]		0.001	0.001
TLR5_G1687A	G/A	16	ss469376089		0.005	0.000
TLR5_A1778G	A/G	16	ss469376091		0.004	0.003
TLR5_A1865C	A/C	16	ss469376093		0.005	0.005
TLR5_G2135A	G/A	16	ss469376095	X	0.008	0.002
TLR5_A2326G	A/G	16	ss469376097	X	0.004	0.004
TLR5_C2332T	C/T	16	ss469376099	Putative Nonsense	0.052	0.054
TLR5_A2463G	A/G	16	rs55617233		0.016	0.003
TLR5_A2500G	A/G	16	rs55617168	X	0.016	0.003
TLR5_G2744A	G/A	16	ss469376101	X	0.004	0.003
TLR5_C2964T	C/T	16	ss469376103		0.005	0.000
TLR5_C3090T	C/T	16	rs55617142		0.006	0.001
TLR5_T3720C	T/C	16	rs55617187		0.385	0.364
TLR5_T3726C	T/C	16	ss469376105		0.004	0.002
TLR5_C3888G	C/G	16	ss469376107	X	0.003	0.003
TLR5_C3897T	C/T	16	rs55617178		0.004	0.003
TLR5_G3934A	G/A	16	rs55617251	X	0.006	0.001
TLR5_T3994C	T/C	16	ss469376109	X	0.016	0.002
TLR5_G4167C	G/C	16	rs55617161		0.023	0.003
TLR5_G4419A	G/A	16	rs55617337		0.413	0.424
TLR5_G4483A	G/A	16	rs55617166	X	0.007	0.001
TLR5_A4580G	A/G	16	rs55617176		0.359	0.362
TLR5_C4846T	C/T	16	rs55617200		0.023	0.003
TLR5_G4979C	G/C	16	rs55617158		0.359	0.362
TLR5_G4988A	G/A	16	rs55617177		0.010	0.005
TLR5_G5150A	G/A	16	rs55617322		0.358	0.361
TLR5_A5199G	A/G	16	rs55617240		0.006	0.001
TLR6_T14066G	T/G	6	rs68268270	X	0.003	0
TLR6_A14121G	A/G	6	rs68268271		0.029	0.003
TLR6_A14197G	A/G	6	rs68268272	X	0.017	0.001
TLR6_G14578A	G/A	6	rs43702941	X	0.365	0.355
TLR6_G14589A	G/A	6	rs68268273		0.028	0.003
TLR6_C15060T	C/T	6	rs68268274		0.071	0.073
TLR6_A15121G	A/G	6	rs68268275	X	0.025	0.003
TLR6_G15138A	G/A	6	ss469376111		0.006	0.000
TLR6_T15213C	T/C	6	rs68268276		0.030	0.003
TLR6_C15312T	C/T	6	rs68268277		0.009	0.000
TLR6_T15418A	T/A	6	ss469376113	X	0.024	0.026
TLR6_C15492T	C/T	6	ss469376115		0.058	0.061
TLR6_C15555G	C/G	6	rs68343176	X	0.019	0.003

TLR6_C15753T	C/T	6	rs68268280		0.357	0.364
TLR6_A15966G	A/G	6	rs68343178		0.022	0.003
TLR7_T301C	T/C	X	rs55617449		0.104	0.108
TLR7_G390A	G/A	X	ss469376117		0.062	0.092
TLR7_C475T	C/T	X	rs55617377		0.438	0.492
TLR7_C527T	C/T	X	rs55617163		0.115	0.123
TLR7_A607G	A/G	X	ss469376119		0.104	0.108
TLR7_C868T	C/T	X	rs55617433		0.115	0.123
TLR7_A1360G	A/G	X	ss469376121		0.010	0.015
TLR7_A1878G	A/G	X	ss469376123	X	0.021	0.031
TLR7_C2260G	C/G	X	ss469376125		0.062	0.077
TLR7_G3820A	G/A	X	rs55617323		0.042	0.031
TLR7_G3863C	G/C	X	rs55617439		0.106	0.109
TLR7_G3938C	G/C	X	rs29012404		0.292	0.369
TLR7_G3971A	G/A	X	ss469376127		0.062	0.077
TLR7_G4072A	G/A	X	ss469376129		0.062	0.077
TLR7_G4176C	G/C	X	ss469376131		0.104	0.108
TLR8_C400G	C/G	X	rs55617249		0.292	0.2
TLR8_C1027G	G/C	X	rs55617319		0.438	0.462
TLR8_A1247C	A/C	X	ss469376133		0.432	0.469
TLR8_C1408T	C/T	X	rs55617165		0.438	0.462
TLR8_T1415C	T/C	X	rs55617354	X	0.436	0.453
TLR8_A1500T	A/T	X	rs55617174	X	0.438	0.462
TLR8_A1523C	A/C	X	rs55617259	X	0.438	0.462
TLR8_C1594A	C/A	X	rs55617145	X	0.438	0.462
TLR8_G1800A	G/A	X	rs55617351	X	0.438	0.462
TLR8_G1845A	G/A	X	ss469376135	X	0.292	0.369
TLR8_C2686A	C/A	X	rs55617390	X	0.438	0.462
TLR8_A3078C	A/C	X	ss469376137	X	0.010	0.015
TLR8_G3606A	G/A	X	ss469376139		0.458	0.462
TLR9_G149A	A/G	22	rs55617357		0.421	0.422
TLR9_G201C	G/C	22	ss469376141		0.003	0.003
TLR9_T258C	T/C	22	ss469376143		0.002	0.000
TLR9_G367A	G/A	22	ss469376145		0.002	0.002
TLR9_G398A	G/A	22	ss469376147		0.002	0.002
TLR9_G713C	G/C	22	rs55617314		0.003	0.002
TLR9_A945G	A/G	22	rs55617138		0.425	0.441
TLR9_G1174A	G/A	22	ss469376149		0.005	0.004
TLR9_T1349C	T/C	22	rs42015526[b]		0.421	0.422
TLR9_G1401A	G/A	22	ss469376151		0.404	0.422
TLR9_C1561T	C/T	22	rs42015525[b]		0.146	0.134
TLR9_C2418A	C/A	22	ss469376153		0.003	0.002
TLR9_G2700A	G/A	22	rs55617140		0.425	0.441
TLR9_C2788T	C/T	22	ss469376155	X	0.430	0.445
TLR9_G2822A	G/A	22	rs55617258	X	0.421	0.436
TLR9_G2945A	G/A	22	ss469376157	X	0.005	0.002
TLR9_A3156G	A/G	22	ss469376159		0.423	0.439
TLR9_A3264G	A/G	22	rs55617255		0.424	0.439
TLR9_G3474C	G/C	22	rs42015524		0.410	0.414
TLR9_T4050C	T/C	22	ss469376161		0.429	0.444
TLR9_G4095A	G/A	22	rs55617221		0.406	0.424
TLR9_G4377A	G/A	22	rs55617220		0.003	0.002
TLR10_T71C	T/C	6	rs55617310	X	0.030	0.005
TLR10_G117A	G/A	6	rs55617437	X	0.018	0.003
TLR10_C361A	C/A	6	rs55617269	X	0.013	0.002
TLR10_C414A	C/A	6	rs55617137	X	0.013	0.002
TLR10_C466G	C/G	6	rs55617286	X	0.013	0.002
TLR10_A475G	A/G	6	rs55617206		0.018	0.008
TLR10_T617C	T/C	6	rs55617348		0.013	0.002
TLR10_C697A	C/A	6	rs55617155		0.001	0.000
TLR10_T723C	T/C	6	rs55617455	X	0.013	0.002
TLR10_T774A	T/A	6	rs55617325	X	0.427	0.442
TLR10_C776G	C/G	6	ss469376163	X	0.002	0.000
TLR10_T865C	T/C	6	ss469376165		0.010	0.009
TLR10_G904A	G/A	6	rs55617387		0.065	0.068
TLR10_A956T	A/T	6	rs55617197	X	0.016	0.005

TLR10_A1022G		A/G	6	rs55617324	X	0.032	0.008
TLR10_A1100G		A/G	6	rs55617311	X	0.018	0.003
TLR10_C1132T		C/T	6	rs55617152		0.006	0.001
TLR10_T1186C		T/C	6	rs55617131		0.006	0.000
TLR10_G1237A		G/A	6	rs55617298		0.081	0.081
TLR10_C1262T		C/T	6	rs55617266		0.018	0.004
TLR10_A2035G		A/G	6	rs55617153		0.019	0.006
TLR10_A2322C		A/C	6	rs55617297	X	0.010	0.005
TLR10_G2352A		G/A	6	rs55617343	X	0.012	0.000
TLR10_G3266A		G/A	6	rs55617308		0.003	0.003
TLR10_C3395T		C/T	6	rs55617336		0.016	0.000
TLR10_A3691G		A/G	6	ss469376167		0.016	0.003
TLR10_C3698A		C/A	6	ss469376169		0.013	0.000
TLR10_T3702C		T/C	6	ss469376171		0.014	0.000
TLR10_G3704A		G/A	6	rs55617227		0.021	0.002
TLR10_C3756A		C/T	6	rs55617328		0.022	0.006
TLR10_G3788A		G/A	6	rs55617457		0.031	0.002
TLR10_C3819T		C/T	6	ss469376173		0.020	0.007
TLR10_C3885T		C/T	6	rs55617156		0.012	0.002
TLR10_C3893G		C/G	6	ss469376175		0.017	0.018
TLR10_T3908A		T/A	6	rs55617212		0.031	0.002
Totals			280		72		

a indicates that this SNP could not be placed on a discrete haplotype with best-pair phase probability ≥ 0.90

b indicates that genotypes for these SNPs are represented by the reverse complement in the raw data file, which is simply a function of assay design and SNP calling

Table A3. Tag SNPs and Indels

Gene	tagSNP/Indel all cattle	Total Alleles Captured all cattle	tagSNP/Indel taurus	Total Alleles Captured taurus
TLR1	rs55617254 ss469376045 rs55617441 rs43702940	5 of 5 (100%) with these 4 tags	rs55617254 rs43702940	3 of 3 (100%) with these 2 tags
TLR2	rs68268241 ss470256470 ss470256472 rs68268245 ss470256473 rs68343166 ss470256474 ss470256475 ss470256476 ss470256477 ss470256478 rs43706434 rs43706433 rs68268251 ss470256481 ss470256482 rs68268255 ss470256483 rs68268258 rs41830058 rs68343170 rs68268262 rs68268265 ss470256484	45 of 45 (100%) with these 24 tags	ss470256470 ss470256472 rs68343163 rs68268245 ss470256473 ss470256474 rs68268248 ss470256475 ss470256477 ss470256478 rs43706434 rs43706433 ss470256481 ss470256482 ss470256483 rs41830058 rs68268265	37 of 37 (100%) with these 17 tags
TLR3	rs42851894 rs42851895 rs42851896 rs42851897 rs55617276 rs55617242 rs42851912 ss469376047	56 of 56 (100%) with these 29 tags	rs42851896 rs42851898 rs55617276 rs55617184 rs55617229 ss469376049 rs55617462 ss469376059	40 of 40 (100%) with these 15 tags

78

	rs55617184 ss469376049 ss469376051 ss469376053 rs55617234 ss469376055 rs55617462 ss469376057 ss469376059 rs42851924 rs42851925 rs55617222 ss469376061 rs42852435 rs55617204 rs55617272 rs55617164 rs55617241 rs55617451 rs55617344 rs55617353		rs42851924 rs42851925 rs55617222 rs42852435 rs55617272 rs55617241 rs55617353	
TLR4	ss469376063 rs8193042 rs8193043 rs8193044 ss469376065 rs8193045 rs8193046 rs8193047 rs8193049 rs8193053 rs8193054 rs8193055 rs8193057 rs8193059 rs8193060 ss469376073 ss469376075 rs8193061 rs8193062 rs8193066 rs8193067 rs8193069 rs8193070 ss469376077 rs8193071	28 of 28 (100%) with these 25 tags	ss469376063 rs8193042 ss469376065 rs8193045 rs8193047 ss469376069 rs8193049 rs8193054 rs8193055 rs8193060 ss469376073 ss469376075 rs8193066 rs8193069 ss469376077 rs8193071	23 of 23 (100%) with these 16 tags
TLR5	ss469376081 rs55617208 rs55617256 ss469376085 ss469376087 ss469376091 ss469376093 ss469376095 ss469376097 ss469376099 rs55617233 ss469376101 rs55617142 ss469376105 ss469376107 rs55617178 rs55617337 rs55617200 rs55617177	46 of 46 (100%) with these 19 tags	ss469376079 rs55617167 rs55617256 ss469376085 ss469376087 ss469376091 ss469376093 ss469376095 ss469376097 ss469376099 rs55617233 rs55617168 ss469376101 ss469376107 rs55617178 ss469376109 rs55617337 rs55617166 rs55617200 rs55617177	43 of 43 (100%) with these 20 tags

TLR6	rs68268270 rs68268272 rs43702941 rs68268274 rs68268275 ss469376111 rs68268277 ss469376113 ss469376115 rs68343176 rs68343178	15 of 15 (100%) with these 11 tags	rs68268271 rs68268272 rs43702941 rs68268273 rs68268274 ss469376113 ss469376115	12 of 12 (100%) with these 7 tags
TLR7	ss469376117 rs55617377 rs55617163 ss469376121 ss469376123 ss469376125 rs55617323 rs29012404	15 of 15 (100%) with these 8 tags	ss469376117 rs55617377 rs55617163 ss469376121 ss469376123 ss469376125 rs55617323 rs29012404	15 of 15 (100%) with these 8 tags
TLR8	rs55617249 rs55617354 ss469376135 ss469376137	13 of 13 (100%) with these 4 tags	rs55617249 ss469376133 ss469376135 ss469376137	13 of 13 (100%) with these 4 tags
TLR9	ss469376141 ss469376143 ss469376145 ss469376147 ss469376149 rs42015526 rs42015525 rs55617140 ss469376157 rs55617220	22 of 22 (100%) with these 10 tags	rs55617357 ss469376141 ss469376145 ss469376147 rs55617314 ss469376149 rs42015525 ss469376157 rs55617255	21 of 21 (100%) with these 9 tags
TLR10	rs55617310 rs55617437 rs55617206 rs55617348 rs55617155 rs55617455 rs55617325 ss469376163 ss469376165 rs55617387 rs55617324 rs55617311 rs55617152 rs55617131 rs55617298 rs55617266 rs55617153 rs55617297 rs55617343 rs55617308 rs55617336 ss469376171 rs55617227 rs55617328 rs55617457 ss469376173 rs55617156 ss469376175	35 of 35 (100%) with these 28 tags	rs55617310 rs55617437 rs55617269 rs55617206 rs55617348 rs55617325 ss469376165 rs55617387 rs55617197 rs55617324 rs55617311 rs55617152 rs55617298 rs55617266 rs55617153 rs55617297 rs55617308 ss469376167 rs55617328 ss469376173 ss469376175	28 of 28 (100%) with these 21 tags
Totals	162 tags	280 Total variable sites	119 Tags	235 Total variable sites
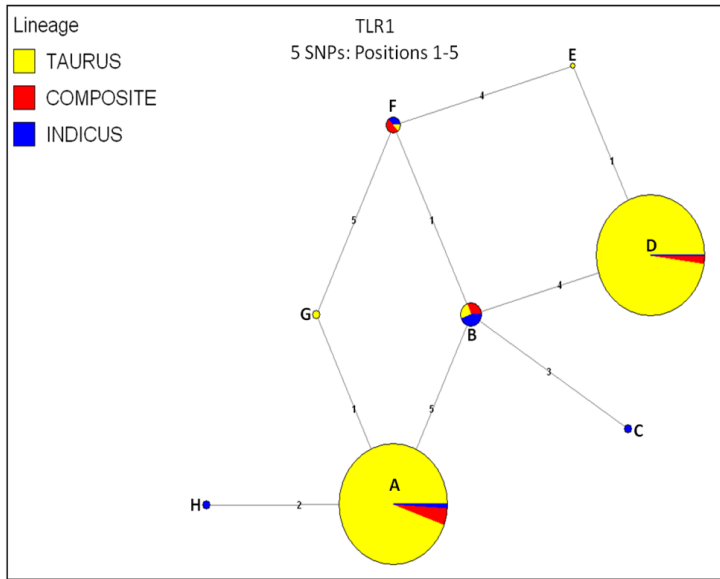
Figure A2. MJ haplotype network for bovine *TLR*1 using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein). Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
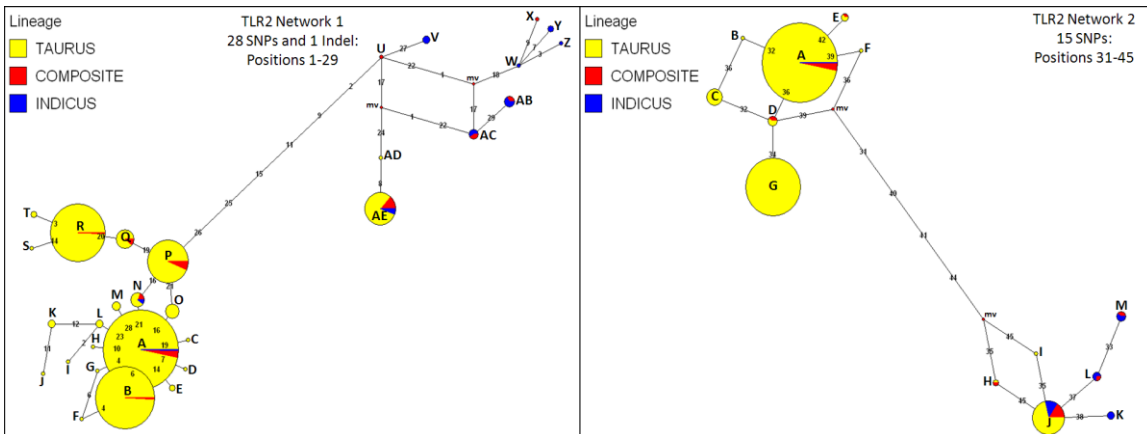
Figure A3. MJ haplotype networks for bovine *TLR2* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein). Because MJ networks require the absence of recombination [73], each network represents intragenic regions of elevated LD. Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
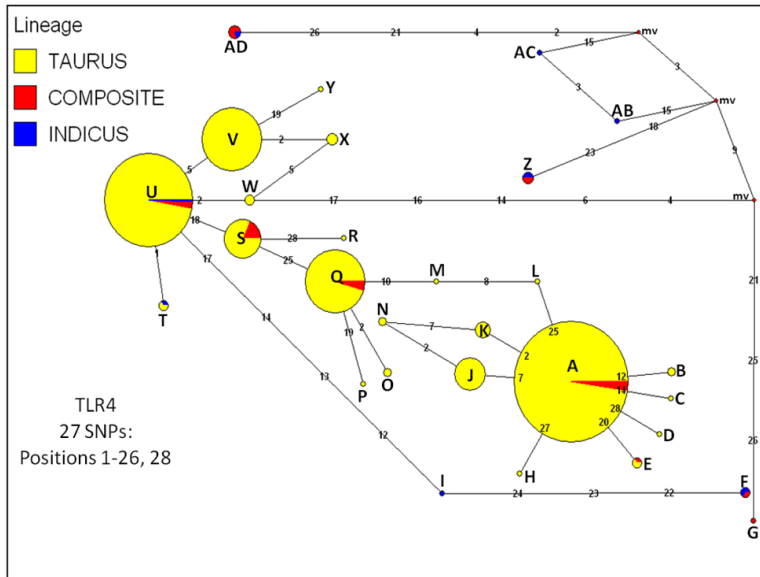
Figure A4. MJ haplotype network for bovine *TLR4* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein). Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
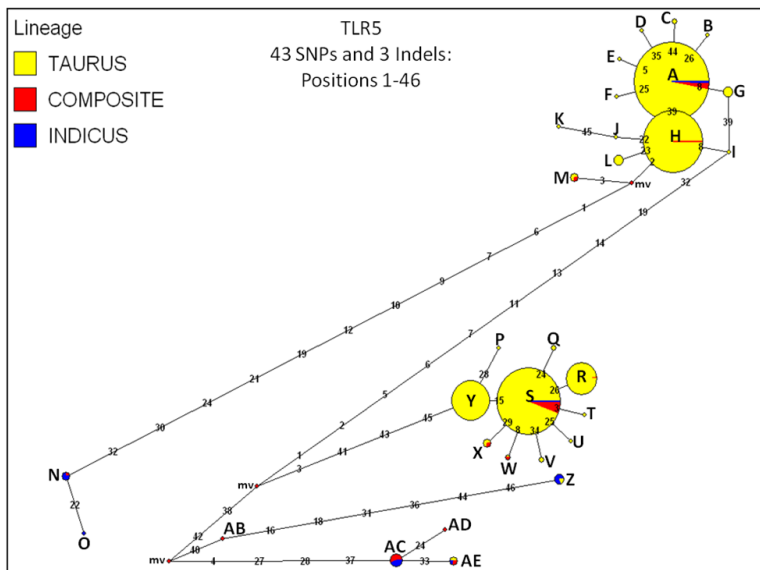
Figure A5. MJ haplotype network for bovine *TLR5* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein). Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
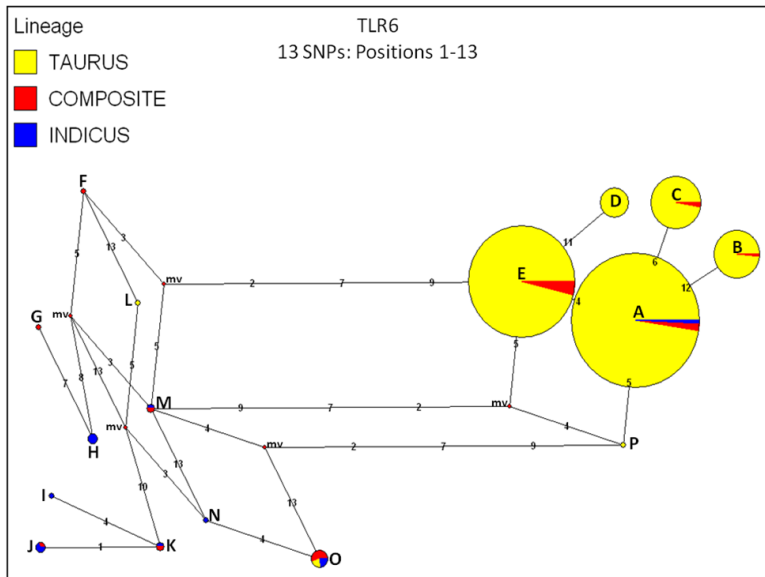
Figure A6. MJ haplotype networks for bovine *TLR6* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein). Haplotypes predicted for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
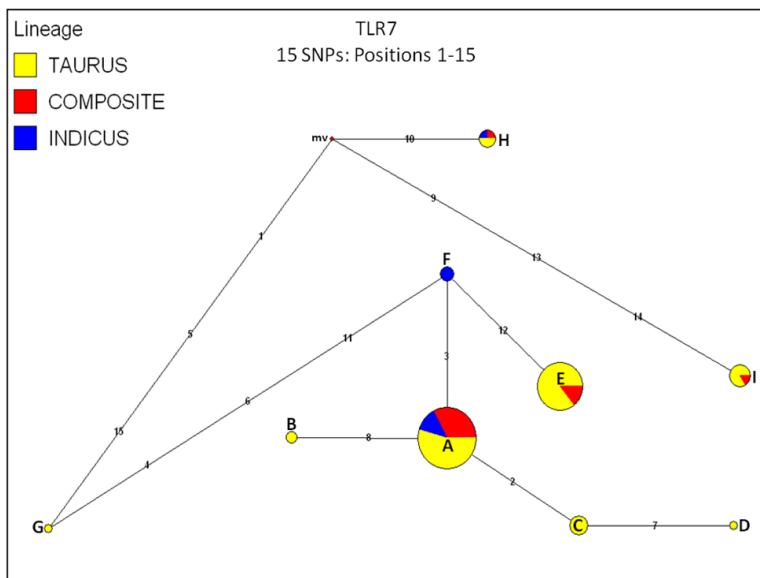
Figure A7. Median joining (MJ) haplotype network for bovine *TLR7* using haplotypes directly ascertained for all cattle (n = 96 AI sires, 31 breeds). Haplotypes observed for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.
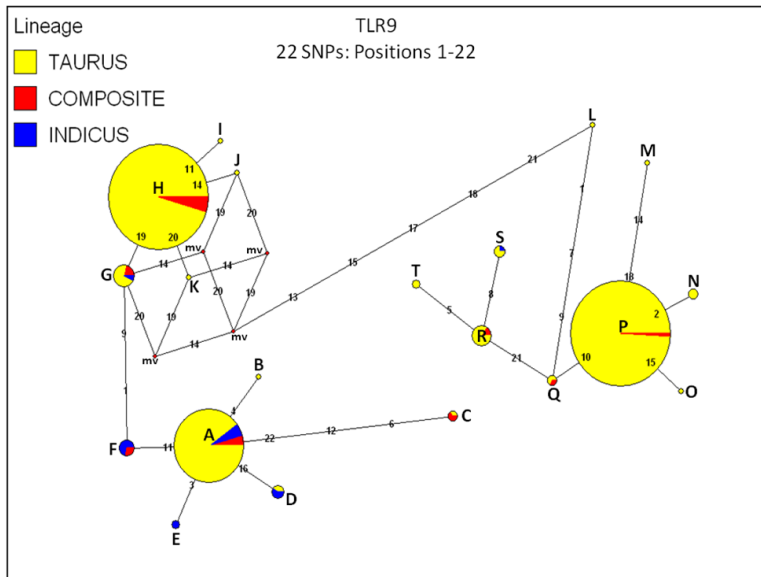
Figure A8. Median joining (MJ) haplotype network for bovine *TLR9* using haplotypes predicted for all cattle (n = 96 AI sires, 31 breeds; 48 Purebred Angus; 405 Holstein)). Haplotypes observed for *B. t. taurus*, *B. t. indicus* and hybrids (termed ''composites'') are color coded. Numbers indicate SNP positions in numerical order (see Table A2 for SNP information). 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 A4). Median vectors are indicated as ''mv''.

Table A4. Network Node Breed Key and Frequency Data.

Gene	Node	Total Frequency	Breeds	Beef	Dairy	Dairy + Beef	Beef + Dairy	Beef Sum	Dairy Sum	Shared1	Shared2	Notes
TLR1	A	469	Angus	1	1	1	1					
			Beefmaster									
			Belgian Blue									
			Blonde d' Aquitaine									
			Braford									
			Brahman									
			Brahmousin									
			Brangus									
			Braunvieh									
			Brown Swiss									
			Charolais									
			Chianina/Chiangus									
			Gelbvieh									
			Hereford									
			Holstein									
			Limousin									
			Maine-Anjou									
			Nelore									
			Red Angus									
			Red Brangus									
			Romagnola									
			Salers									
			Santa Gertrudis									
			Senepol									
			Shorthorn									
			Simbrah									
			Simmental Black									
			Simmental Red									
			Texas Longhorn									
TLR1	B	16	Beefmaster		1							
			Braford									
			Brahman									
			Holstein									
			Maine-Anjou									
			Nelore									
			Red Brangus									
			Texas Longhorn									
TLR1	C	2	Brahman									
TLR1	D	594	Angus	1	1	1	1					

			Beefmaster								
			Belgian Blue								
			Blonde d' Aquitaine								
			Brahman								
			Brangus								
			Braunvieh								
			Brown Swiss								
			Charolais								
			Chianina/Chiangus								
			Corriente								
			Gelbvieh								
			Hereford								
			Holstein								
			Maine-Anjou								
			Piedmontese								
			Red Angus								
			Red Brangus								
			Red Poll								
			Romagnola								
			Santa Gertrudis								
			Senepol								
			Shorthorn								
			Simbrah								
			Simmental Black								
			Simmental Red								
			Texas Longhorn								
TLR1	E	1	Gelbvieh	1							
TLR1	F	8	Brahman		1						
			Nelore								
			Red Brangus								
			Senepol								
			Simbrah								
	G	2	Holstein								
	H	2	Brahman								
Total	8	1094	99.63570128					3	4	2	2
TLR2 Network1	A	492	Angus	1	1	1	1				
			Beefmaster								
			Belgian Blue								
			Blonde d' Aquitaine								
			Braford								
			Brahman								
			Brahmousin								
			Brangus								
			Braunvieh								
			Brown Swiss								
			Charolais								

Chianina/Chiangus
Gelbvieh
Hereford
Holstein
Limousin
Maine-Anjou
Piedmontese
Red Angus
Red Brangus
Red Poll
Salers
Santa Gertrudis
Senepol
Shorthorn
Simmental Black
Simmental Red
Texas Longhorn

TLR2 Network1	B	189	Angus	1	1	1	1
			Beefmaster				
			Charolais				
			Chianina/Chiangus				
			Corriente				
			Holstein				
			Red Brangus				
			Senepol				
TLR2 Network1	C	1	Holstein		1		
TLR2 Network1	D	1	Angus	1			
TLR2 Network1	E	2	Braunvieh		1		
			Holstein				
TLR2 Network1	F	1	Angus	1			
TLR2 Network1	G	1	Holstein		1		
TLR2 Network1	H	1	Belgian Blue				
TLR2 Network1	I	1	Simmental Red	1			
TLR2 Network1	J	1	Angus	1			
TLR2 Network1	K	3	Chianina/Chiangus	1	1	1	1
			Holstein				
			Red Angus				
TLR2 Network1	L	3	Simmental Black	1			
			Simmental Red				
TLR2 Network1	M	4	Holstein		1		
TLR2 Network1	N	10	Brahman	1	1	1	1
			Brahmousin				
			Brangus				
			Chianina/Chiangus				
			Corriente				

			Holstein				
			Simmental Black				
TLR2 Network1	O	11	Angus	1	1	1	1
			Charolais				
			Holstein				
			Red Poll				
			Shorthorn				
TLR2 Network1	P	90	Angus	1	1	1	1
			Beefmaster				
			Blonde d' Aquitaine				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Holstein				
			Limousin				
			Maine-Anjou				
			Piedmontese				
			Red Brangus				
			Romagnola				
			Salers				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR2 Network1	Q	19	Charolais	1	1	1	1
			Gelbvieh				
			Holstein				
			Limousin				
			Romagnola				
			Salers				
			Shorthorn				
			Simbrah				
TLR2 Network1	R	163	Braunvieh	1	1	1	1
			Hereford				
			Holstein				
			Santa Gertrudis				
TLR2 Network1	S	1	Holstein		1		
TLR2 Network1	T	2	Holstein		1		
TLR2 Network1	U	1	Brangus				
TLR2 Network1	V	3	Brahman				
TLR2 Network1	W	1	Brahman				
TLR2 Network1	X	1	Santa Gertrudis				
TLR2 Network1	Y	2	Brahman				

TLR2 Network1	Z	1	Brahman								
TLR2 Network1	AB	6	Brahman Brahmousin Nelore Simbrah								
TLR2 Network1	AC	5	Brahman Nelore Romagnola Simbrah								
TLR2 Network1	AD	1	Angus	1							
TLR2 Network1	AE	51	Angus Belgian Blue Braford Brahman Brangus Braunvieh Brown Swiss Gelbvieh Hereford Holstein Nelore Red Angus Red Brangus Simbrah Texas Longhorn	1	1	1	1				
Total	30	1068	97.26775956					15	15	9	9
TLR2 Network2	A	832	Angus Beefmaster Belgian Blue Blonde d' Aquitaine Brahman Brahmousin Brangus Braunvieh Brown Swiss Charolais Chianina/Chiangus Corriente Gelbvieh Hereford Holstein Limousin Maine-Anjou Piedmontese Red Angus Red Brangus	1	1	1	1				

			Red Poll				
			Romagnola				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR2 Network2	B	1	Holstein		1		
	C	15	Charolais	1	1	1	1
			Gelbvieh				
			Holstein				
			Salers				
			Shorthorn				
TLR2 Network2	D	5	Limousin	1			
			Romagnola				
			Simbrah				
TLR2 Network2	E	3	Beefmaster		1		
			Holstein				
TLR2 Network2	F	1	Holstein		1		
TLR2 Network2	G	165	Hereford	1	1	1	1
			Holstein				
TLR2 Network2	H	2	Hereford	1			
			Simbrah				
TLR2 Network2	I	1	Angus	1			
TLR2 Network2	J	55	Angus	1	1	1	1
			Belgian Blue				
			Braford				
			Brahman				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Gelbvieh				
			Holstein				
			Nelore				
			Red Angus				
			Red Brangus				
			Santa Gertrudis				
			Simbrah				
			Texas Longhorn				
TLR2 Network2	K	3	Brahman				
TLR2 Network2	L	4	Brahman				
			Nelore				

			Romagnola					
TLR2 Network2	M	5	Brahman					
			Brahmousin					
			Nelore					
			Simbrah					
Total	13	1092	99.45355191			7	7	4	4
TLR3 Network1	A	40	Angus	1	1	1	1	
			Beefmaster					
			Belgian Blue					
			Blonde d' Aquitaine					
			Brahmousin					
			Brangus					
			Braunvieh					
			Brown Swiss					
			Charolais					
			Corriente					
			Gelbvieh					
			Holstein					
			Limousin					
			Salers					
			Senepol					
			Simmental Black					
			Simmental Red					
TLR3 Network1	B	1	Maine-Anjou					
TLR3 Network1	C	2	Shorthorn		1			
TLR3 Network1	D	2	Blonde d' Aquitaine					
			Santa Gertrudis					
TLR3 Network1	E	3	Angus	1				
			Salers					
TLR3 Network1	F	1	Romagnola					
TLR3 Network1	G	23	Beefmaster	1	1	1	1	
			Brahman					
			Braunvieh					
			Chianina/Chiangus					
			Gelbvieh					
			Holstein					
			Limousin					
			Piedmontese					
			Romagnola					
			Senepol					
			Shorthorn					
			Simmental Black					
			Texas Longhorn					
TLR3 Network1	H	1	Brangus					
TLR3 Network1	I	2	Shorthorn		1			

TLR3 Network1	J	5	Brown Swiss	1	1	1	1
			Gelbvieh				
			Maine-Anjou				
			Red Brangus				
TLR3 Network1	K	32	Angus	1	1	1	1
			Braford				
			Brangus				
			Braunvieh				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Romagnola				
			Simbrah				
			Simmental Black				
			Simmental Red				
TLR3 Network1	L	16	Angus	1	1	1	1
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Holstein				
			Maine-Anjou				
			Red Brangus				
			Santa Gertrudis				
			Simmental Black				
			Simmental Red				
TLR3 Network1	M	2	Brahman				
			Simbrah				
TLR3 Network1	N	17	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Charolais				
			Chianina/Chiangus				
			Holstein				
			Limousin				
			Piedmontese				
			Santa Gertrudis				
			Shorthorn				
			Texas Longhorn				
TLR3 Network1	O	3	Brown Swiss	1	1	1	1
			Gelbvieh				

Network	ID	Count	Breed(s)								
			Maine-Anjou								
TLR3 Network1	P	2	Brahman Simbrah								
TLR3 Network1	Q	2	Brahman Brangus								
TLR3 Network1	R	2	Nelore Simbrah								
TLR3 Network1	S	1	Braford								
TLR3 Network1	T	6	Braford Brahman Nelore Red Brangus								
TLR3 Network1	U	3	Beefmaster Brahmousin								
Total	21	166	86.45833333					8	9	7	7
TLR3 Network2	A	36	Angus Beefmaster Belgian Blue Blonde d' Aquitaine Brahmousin Brangus Braunvieh Charolais Corriente Gelbvieh Hereford Holstein Maine-Anjou Salers Senepol Simmental Red	1	1	1	1				
TLR3 Network2	B	2	Brangus Simmental Red	1							
TLR3 Network2	C	5	Blonde d' Aquitaine Braunvieh Salers Santa Gertrudis		1						
TLR3 Network2	D	9	Beefmaster Brown Swiss Holstein Limousin Red Angus Simmental Black	1	1	1	1				
TLR3 Network2	E	4	Angus Braunvieh	1	1	1	1				

Network	Group	Count	Breed				
			Gelbvieh				
			Limousin				
TLR3 Network2	F	3	Braford				
			Brahman				
			Nelore				
	G	1	Brahman				
TLR3 Network2	H	15	Beefmaster				
			Braford				
			Brahman				
			Brahmousin				
			Nelore				
			Red Brangus				
			Simbrah				
TLR3 Network2	I	4	Brahman				
			Nelore				
			Simbrah				
TLR3 Network2	J	2	Brahman				
			Simbrah				
TLR3 Network2	K	1	Brown Swiss	1			
TLR3 Network2	L	93	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Braford				
			Brahman				
			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Piedmontese				
			Red Angus				
			Red Brangus				
			Red Poll				
			Romagnola				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				

			Texas Longhorn				
TLR3 Network2	M	11	Angus	1	1	1	1
			Beefmaster				
			Brahman				
			Chianina/Chiangus				
			Holstein				
			Limousin				
			Piedmontese				
			Red Poll				
			Simbrah				
			Texas Longhorn				
TLR3 Network2	N	2	Brahman				
			Brangus				
Total	14	188	97.91666667	6	7	5	5
TLR4	A	485	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Blonde d' Aquitaine				
			Braford				
			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Corriente				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Romagnola				
			Salers				
			Santa Gertrudis				
			Shorthorn				
			Simmental Black				
			Simmental Red				
TLR4	B	2	Holstein	1			
TLR4	C	1	Holstein	1			
TLR4	D	1	Salers				
TLR4	E	3	Beefmaster				
			Holstein	1			
TLR4	F	3	Beefmaster				
			Brahman				

			Nelore				
TLR4	G	1	Simbrah				
TLR4	H	1	Holstein	1			
TLR4	I	1	Brahman				
TLR4	J	29	Holstein	1			
TLR4	K	7	Holstein	1			
TLR4	L	1	Holstein	1			
TLR4	M	1	Holstein	1			
TLR4	N	2	Holstein	1			
TLR4	O	2	Holstein	1			
TLR4	P	1	Holstein	1			
TLR4	Q	110	Angus	1	1	1	1
			Belgian Blue				
			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Gelbvieh				
			Holstein				
			Limousin				
			Maine-Anjou				
			Piedmontese				
			Red Angus				
			Salers				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR4	R	1	Chianina/Chiangus				
TLR4	S	43	Angus	1			
			Beefmaster				
			Brangus				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Romagnola				
			Senepol				
			Simmental Black				
TLR4	T	3	Brahman				
			Holstein	1			
TLR4	U	239	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				

			Blonde d' Aquitaine								
			Brahmousin								
			Brangus								
			Braunvieh								
			Brown Swiss								
			Charolais								
			Chianina/Chiangus								
			Corriente								
			Gelbvieh								
			Hereford								
			Holstein								
			Limousin								
			Maine-Anjou								
			Nelore								
			Piedmontese								
			Red Angus								
			Red Poll								
			Romagnola								
			Salers								
			Santa Gertrudis								
			Senepol								
			Shorthorn								
			Simmental Black								
TLR4	V	108	Angus	1	1	1	1				
			Braunvieh								
			Gelbvieh								
			Holstein								
			Red Angus								
			Senepol								
TLR4	W	3	Holstein		1						
TLR4	X	4	Holstein		1						
TLR4	Y	1	Holstein		1						
TLR4	Z	4	Brahman								
			Nelore								
			Red Brangus								
			Simbrah								
TLR4	AB	1	Nelore								
TLR4	AC	1	Brahman								
TLR4	AD	5	Beefmaster								
			Braford								
			Brahman								
Total	29	1064	96.90346084						5	19	4	4
TLR5	A	431	Angus	1	1	1	1				
			Belgian Blue								
			Braford								
			Brahman								

			Brangus				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Holstein				
			Limousin				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Red				
			Texas Longhorn				
TLR5	B	2	Holstein		1		
TLR5	C	1	Charolais	1			
TLR5	D	1	Holstein		1		
TLR5	E	1	Holstein		1		
TLR5	F	6	Holstein		1		
TLR5	G	196	Angus	1	1	1	1
			Belgian Blue				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Holstein				
			Limousin				
			Maine-Anjou				
			Piedmontese				
			Red Angus				
			Senepol				
			Shorthorn				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR5	H	1	Holstein		1		
TLR5	I	1	Holstein		1		
TLR5	J	1	Holstein		1		
TLR5	K	5	Holstein		1		
			Red Poll				
TLR5	L	4	Braunvieh		1		
			Holstein				
			Romagnola				
TLR5	M	4	Beefmaster				

98

TLR5			Brahman					
			Nelore					
TLR5	N	1	Brahman					
TLR5	O	1	Holstein	1				
TLR5	P	2	Holstein	1				
TLR5	Q	54	Braford					Nonsense Hap
			Holstein	1				
TLR5	R	228	Angus	1	1	1	1	
			Beefmaster					
			Braford					
			Brahman					
			Brahmousin					
			Brangus					
			Braunvieh					
			Charolais					
			Chianina/Chiangus					
			Corriente					
			Gelbvieh					
			Hereford					
			Holstein					
			Limousin					
			Maine-Anjou					
			Piedmontese					
			Red Angus					
			Red Brangus					
			Romagnola					
			Shorthorn					
			Simbrah					
			Simmental Black					
			Simmental Red					
			Texas Longhorn					
TLR5	S	1	Charolais	1				
TLR5	T	1	Charolais	1				
TLR5	U	2	Blonde d' Aquitaine	1				
			Holstein					
TLR5	V	2	Simmental Black	1				
			Simbrah					
TLR5	W	4	Blonde d' Aquitaine	1				
			Charolais					
			Red Poll					
			Simbrah					
TLR5	X	81	Chianina/Chiangus					
			Holstein	1				
TLR5	Y	6	Brahman					
			Nelore					

			Texas Longhorn								
TLR5	Z	1	Brangus								
TLR5	AB	9	Brahmousin Brangus Nelore Brahman Simbrah								
TLR5	AC	1	Santa Gertrudis								
TLR5	AD	4	Chianina/Chiangus Nelore Red Brangus Senepol								
Total	29	1052	95.81056466					8	17	3	3
TLR6	A	523	Angus Beefmaster Belgian Blue Blonde d' Aquitaine Brahman Brangus Braunvieh Brown Swiss Charolais Chianina/Chiangus Corriente Gelbvieh Hereford Holstein Maine-Anjou Piedmontese Red Angus Red Poll Romagnola Santa Gertrudis Senepol Shorthorn Simbrah Simmental Black Simmental Red Texas Longhorn	1	1	1	1				
TLR6	B	63	Charolais Holstein Red Angus Red Brangus	1	1	1	1				
TLR6	C	76	Angus Blonde d' Aquitaine Brahmousin	1	1	1	1				

100

			Braunvieh				
			Brown Swiss				
			Charolais				
			Holstein				
			Limousin				
			Romagnola				
			Salers				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR6	D	26	Angus	1	1	1	1
			Maine-Anjou				
			Shorthorn				
TLR6	E	339	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Braford				
			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Romagnola				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simmental Black				
			Texas Longhorn				
TLR6	F	1	Beefmaster				
	G	1	Brangus				
TLR6	H	3	Brahman				
TLR6	I	1	Brahman				
TLR6	J	3	Braford				
TLR6			Brahman				
	K	2	Nelore				
TLR6			Red Brangus				
	L	1	Chianina/Chiangus				

TLR6	M	2	Nelore								
			Simbrah								
TLR6	N	1	Brahman								
TLR6	O	9	Beefmaster								
			Brahman								
			Brangus								
			Red Brangus								
			Senepol								
			Simbrah								
			Texas Longhorn								
TLR6	P	1	Chianina/Chiangus								
Total	16	1052	95.81056466					5	5	5	5
TLR7	A	46	Angus	1	1	1	1				
			Beefmaster								
			Braford								
			Brahman								
			Brahmousin								
			Brangus								
			Braunvieh								
			Brown Swiss								
			Charolais								
			Chianina/Chiangus								
			Gelbvieh								
			Hereford								
			Holstein								
			Limousin								
			Maine-Anjou								
			Nelore								
			Piedmontese								
			Red Angus								
			Red Brangus								
			Romagnola								
			Salers								
			Santa Gertrudis								
			Simbrah								
			Simmental Black								
			Simmental Red								
TLR7	B	2	Charolais	1	1	1	1				
			Holstein								
TLR7	C	5	Belgian Blue	1	1	1	1				
			Brown Swiss								
			Chianina/Chiangus								
			Limousin								
			Simmental Black								
TLR7	D	1	Simmental Red	1							
TLR7	E	28	Angus	1	1	1	1				

TLR	Group	Count	Breed								
			Beefmaster								
			Belgian Blue								
			Blonde d' Aquitaine								
			Brangus								
			Braunvieh								
			Charolais								
			Chianina/Chiangus								
			Gelbvieh								
			Hereford								
			Holstein								
			Maine-Anjou								
			Red Angus								
			Red Poll								
			Salers								
			Santa Gertrudis								
			Senepol								
			Shorthorn								
			Simmental Black								
TLR7	F	3	Brahman								
TLR7	G	1	Shorthorn		1						
TLR7	H	4	Beefmaster								
			Limousin	1							
			Nelore								
			Senepol								
TLR7	I	6	Corriente	1	1	1	1				
			Red Angus								
			Shorthorn								
			Simbrah								
			Texas Longhorn								
Total	9	96	100					7	6	5	5
TLR8	A	28	Angus	1	1	1	1				
			Beefmaster								
			Braford								
			Brahman								
			Brangus								
			Brown Swiss								
			Charolais								
			Chianina/Chiangus								
			Gelbvieh								
			Holstein								
			Maine-Anjou								
			Nelore								
			Piedmontese								
			Red Angus								
			Red Brangus								
			Romagnola								

			Salers Santa Gertrudis <mark>Shorthorn</mark> Simbrah <mark>Simmental Black</mark> <mark>Simmental Red</mark>				
TLR8	B	24	Braford Brahman <mark>Braunvieh</mark> <mark>Charolais</mark> Corriente <mark>Gelbvieh</mark> <mark>Hereford</mark> <mark>Holstein</mark> Maine-Anjou <mark>Red Angus</mark> Red Brangus Senepol <mark>Shorthorn</mark> Simbrah Texas Longhorn	1	1	1	1
TLR8	C	2	Beefmaster Brahman				
TLR8	D	14	Belgian Blue Brahmousin <mark>Braunvieh</mark> <mark>Brown Swiss</mark> Chianina/Chiangus <mark>Gelbvieh</mark> <mark>Limousin</mark> Nelore <mark>Simmental Black</mark> <mark>Simmental Red</mark>	1	1	1	1
TLR8	E	27	<mark>Angus</mark> Beefmaster Belgian Blue Blonde d' Aquitaine Brangus <mark>Braunvieh</mark> <mark>Charolais</mark> Chianina/Chiangus <mark>Gelbvieh</mark> <mark>Hereford</mark> <mark>Holstein</mark> Maine-Anjou <mark>Red Angus</mark>	1	1	1	1

			Red Poll									
			Salers									
			Santa Gertrudis									
			Senepol									
			Shorthorn									
			Simmental Black									
TLR8	F	1	Angus	1								
Total	6	96	100					5	4	4	4	
TLR9	A	148	Angus	1	1	1	1					
			Beefmaster									
			Braford									
			Brahman									
			Chianina/Chiangus									
			Holstein									
			Nelore									
			Red Angus									
			Red Brangus									
			Santa Gertrudis									
			Senepol									
			Simbrah									
TLR9	B	1	Holstein		1							
TLR9	C	3	Chianina/Chiangus									
			Romagnola									
TLR9	D	5	Brahman									
			Chianina/Chiangus									
			Nelore									
			Texas Longhorn									
TLR9	E	2	Brahman									
TLR9	F	7	Brahman									
			Red Brangus									
			Santa Gertrudis									
TLR9	G	14	Beefmaster	1	1	1	1					
			Brahman									
			Brahmousin									
			Gelbvieh									
			Holstein									
			Limousin									
			Maine-Anjou									
			Red Angus									
			Red Brangus									
			Texas Longhorn									
TLR9	H	444	Angus	1	1	1	1					
			Beefmaster									
			Belgian Blue									
			Blonde d' Aquitaine									
			Braford									

TLR9			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Corriente				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Piedmontese				
			Red Angus				
			Red Brangus				
			Red Poll				
			Romagnola				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR9	I	1	Holstein		1		
TLR9	J	1	Holstein		1		
TLR9	K	1	Charolais	1			
TLR9	L	1	Limousin	1			
TLR9	M	1	Holstein		1		
TLR9	N	3	Angus	1	1	1	1
			Holstein				
TLR9	O	1	Hereford	1			
TLR9	P	436	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Brangus				
			Braunvieh				
			Charolais				
			Hereford				
			Holstein				
			Red Angus				
			Salers				
			Simmental Black				
TLR9	Q	3	Beefmaster	1	1	1	1
			Braunvieh				

			Charolais				
TLR9	R	12	Belgian Blue	1	1	1	1
			Brahmousin				
			Holstein				
			Senepol				
			Simbrah				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR9	S	4	Brahman		1		
			Holstein				
			Shorthorn				
TLR9	T	2	Corriente		1		
			Holstein				
Total		20	1090	99.27140255		10	13	7	7
TLR10	A	390	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Braford				
			Brahmousin				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Red Angus				
			Romagnola				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				
TLR10	B	81	Angus	1	1	1	1
			Brahman				
			Brahmousin				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Gelbvieh				

			Holstein				
			Maine-Anjou				
			Piedmontese				
			Romagnola				
			Salers				
			Shorthorn				
			Simmental Black				
			Texas Longhorn				
TLR10	C	11	Angus	1	1	1	1
			Brangus				
			Chianina/Chiangus				
			Maine-Anjou				
			Red Angus				
			Red Brangus				
			Shorthorn				
			Texas Longhorn				
TLR10	D	59	Holstein	1	1	1	1
			Red Angus				
			Red Brangus				
TLR10	E	426	Angus	1	1	1	1
			Beefmaster				
			Belgian Blue				
			Blonde d' Aquitaine				
			Braford				
			Brangus				
			Braunvieh				
			Brown Swiss				
			Charolais				
			Chianina/Chiangus				
			Corriente				
			Gelbvieh				
			Hereford				
			Holstein				
			Limousin				
			Maine-Anjou				
			Red Angus				
			Red Poll				
			Salers				
			Santa Gertrudis				
			Senepol				
			Shorthorn				
			Simbrah				
			Simmental Black				
			Simmental Red				
			Texas Longhorn				
TLR10	F	10	Brahman				

		Nelore Red Brangus Senepol Simbrah					
Total	6	977	88.97996357	5	5	5	5
Note, 977 is the number of haplotypes derived from			84 111 60 60				
nodes A-F. However, 1,048 total haplotypes were confidently reconstructed.							

Table A5. Barcoded Primers.

Genes		Primers F	Primers R	Amplicon Size (excluding MIDs)	MgCl$_2$ Concentration
TLR2_1	MID2	acgctcgacaTCCTGCTCCATATTCCTACG	acgctcgacaTGACTGTGTTTGACATCATGG	816	1.0 X
TLR2_2	MID2	acgctcgacaCTCATTCATTTATGGCTGGC	acgctcgacaGACCTGAACCAGGAGGATG	668	1.0 X
TLR2_3	MID2	acgctcgacaAGATCACCTATGTCGGCAAC	acgctcgacaCATGGGTACAGTCATCAAACTC	681	1.0 X
TLR2_4	MID2	acgctcgacaAGCATCCATCAGTGAAATGAG	acgctcgacaGGTAAGAAGGAGGCATCTGG	774	1.0 X
TLR2_5	MID2	acgctcgacaAGTTTAACCCAGTGCCTTCC	acgctcgacaTGGAGTCAATGATGTTGTCG	730	1.0 X
TLR2_6	MID2	acgctcgacaCCTACTGGGTGGAGAACCTC	acgctcgacaACCACCAGACCAAGACTGAC	436	1.0 X
TLR6_1	MID6	atatcgcgagATTGAGAGTAATCAGCCAAT	atatcgcgagGTAAGGTTGGTCCTCCAGTG	876	1.0 X
TLR6_2	MID6	atatcgcgagACTACCCATTGCTCACTTGC	atatcgcgagCTATACTCCCAACCCAAGAGC	805	1.0 X
TLR6_3	MID6	atatcgcgagGACACACGCTTTATACACATGC	atatcgcgagCACTGACACACCATCCTGAG	845	1.0 X
TLR6_4	MID6	atatcgcgagGCCAAGTATCCAGTGACGTG	atatcgcgagAATGGTGTTCTGTGGAATGG	604	1.0 X
TLR3_05	MID3	agacgcactcGTCGCCATTTCCTTCTCC	agacgcactcCCACAAACTCTCCCCTTCC	701	1.0 X
TLR3_06	MID3	agacgcactcATTGGAGGCAGGTTCTTCAC	agacgcactcATCTCATTGTGTTGGAGGTTC	598	1.0 X
TLR3_07	MID3	agacgcactcGTCAAAGTCTCCCTTGGTTG	agacgcactcGCTGACAAGAAAAAGGTGGT	669	1.0 X
TLR3_08	MID3	agacgcactcGGGATGAAAAAGTGTCGAGT	agacgcactcGTCTGTGCTTTGGGATGTTT	629	1.0 X
TLR3_09	MID3	agacgcactcTCAAAAGTAGCACGAAATGG	agacgcactcAGTCTTGGCATCAAAAATGG	528	1.0 X
TLR3_10	MID3	agacgcactcTGCTATTTTGCTGTCCAGTT	agacgcactcGGACCCTCCACTTCTTTTTG	527	1.0 X
TLR3_11	MID3	agacgcactcCCTTCACACATACTGCTTTGG	agacgcactcTCCCGATACTCTTCTTCTTGG	597	1.0 X
TLR3_12	MID3	agacgcactcCCTATAACGGAGTAAACCTAACCT	agacgcactcCTGTGTAAAACCACGATAAGCA	612	1.0 X
TLR3_13	MID3	agacgcactcAGTTTCAGGTGATTAGCAAAGG	agacgcactcCTCAATCTTTCCCAGCATCA	577	1.0 X

TLR3_14	MID3	agacgcactcCTGTGCTGTATTGCTTCTCTG	agacgcactcGTTTCCATCTGTCTTCTGCTCT	636	1.0 X
TLR3_15	MID3	agacgcactcTTCCTTCTCTCCTGCCTTCT	agacgcactcCACTACTCAGCACCCACATC	641	1.0 X
TLR3_16	MID3	agacgcactcGTCTGGGAGATCAGGGAAG	agacgcactcCCACTGAAAGGAAAAATCGT	701	1.0 X
TLR3_17	MID3	agacgcactcGCTCTTTTTATGGGCTTTCC	agacgcactcCCTTCAGCAACTCGTCATTT	635	1.0 X
TLR3_18	MID3	agacgcactcCCTGAAAAATGTGGACTGC	agacgcactcGTATTGGGGCGGAGTGTT	577	1.0 X
TLR3_19	MID3	agacgcactcCCACACCAACATCTCTGAAC	agacgcactcAAACTGGACACAGCCAAATC	590	1.0 X
TLR3_20	MID3	agacgcactcCAAAAGGTAGGTGAACACTATGAC	agacgcactcATATGGGACGGGCAGTTT	518	1.0 X
TLR3_21	MID3	agacgcactcGTAGCCATTCCCTTCTCCA	agacgcactcCAGCCCAACACTCTAAAATC	545	1.0 X
TLR3_22	MID3	agacgcactcACCTGGGTTTTAGTGACAAG	agacgcactcGCCTGAAATAGGGAGACATA	509	1.0 X
TLR3_23	MID3	agacgcactcGTCCAGAAATTCAGCACATT	agacgcactcAGGTGTACGTTTTACCCTTTCA	437	1.0 X
TLR7_01	MID7	cgtgtctctaCCCAATGTGTAGGGAAAATG	cgtgtctctaCACAGGGCAGAGTTTTAGGA	715	1.0 X
TLR7_02	MID7	cgtgtctctaTTTCAGGTGTTTCCAATGTG	cgtgtctctaGGATCAATCTGTAGGGGAGAA	851	1.0 X
TLR7_03	MID7	cgtgtctctaGAAATTGCCCTCGTTGTT	cgtgtctctaAGCCGATTGTTAGAGAAGTCC	892	1.0 X
TLR7_04	MID7	cgtgtctctaCAGGAAATAGCATTAGCCAGA	cgtgtctctaTAACCCACCAGACAAACCAC	822	1.0 X
TLR7_05	MID7	cgtgtctctaCCAGAAAACGTCCTCAACAA	cgtgtctctaAGTCACATTCGGCAAAGAAG	871	1.0 X
TLR7_06	MID7	cgtgtctctaAACAAACCCACAGGCTCAC	cgtgtctctaCAGGAGAGAAAGAGCAAGGA	669	1.0 X
TLR8_01	MID8	ctcgcgtgtcGCGTTTCCTTGAGTTATGCT	ctcgcgtgtcCTTCCGTCACATCTTTGTCC	583	1.0 X
TLR8_02	MID8	ctcgcgtgtcGCAGAATGTAATGGTCGTCG	ctcgcgtgtcCAAGGTACACAGGGAAATGG	659	1.0 X
TLR8_03	MID8	ctcgcgtgtcAGTGGAAACTGCCCGAGA	ctcgcgtgtcGCTTCAGGATGTGACTTTGG	601	1.0 X
TLR8_04	MID8	ctcgcgtgtcCAGAATATCACCCTTGGTCAG	ctcgcgtgtcAGCATTCCACAGAAGGTCAA	592	1.0 X
TLR8_05	MID8	ctcgcgtgtcTAACGCACCGTCTAGGATTT	ctcgcgtgtcTCTCCGAAGTCACAGGTACAG	625	1.0 X
TLR8_06	MID8	ctcgcgtgtcTGTTTTGGAACTAGGGGGTAA	ctcgcgtgtcCTTGCTTTGGTTGATGCTCT	526	1.0 X
TLR8_07	MID8	ctcgcgtgtcCCTGGAAGAGAGTGAGGACA	ctcgcgtgtcGGCTCTGAAGTGGATGCTAA	665	1.0 X
TLR9_01	MID11	tgatacgtctGTTTGTGCTCTGATGGTGCT	tgatacgtctCCCCTTCCTCTTTCTACTCC	413	1.0 X
TLR9_02	MID11	tgatacgtctCTTCACCTCTCCCCAGACTT	tgatacgtctCCGTGTTTCTCTCCATCACT	414	1.0 X
TLR9_03	MID11	tgatacgtctTTCTCACTTCCTCTGATCTCT	tgatacgtctTGCCTAGCTCTTTCATGCTC	565	1.0 X
TLR9_04	MID11	tgatacgtctATTCGTTCTGACCCACAGCA	tgatacgtctTTCTCTCTCCAGTGCCCATC	537	1.0 X

110

TLR9_05	MID11	tgatacgtctAGATTGCAGGTCTCAGGATG	tgatacgtctACAGGTGGACGAAGTCAGAG	499	1.0 X
TLR9_06	MID11	tgatacgtctCCAGCCTCTCCTTAATCTCC	tgatacgtctCGGAACCAATCTTTCTCTAGTT	718	1.0 X
TLR9_07	MID11	tgatacgtctCCTGACACCTTCAGTCACCT	tgatacgtctGCGGGTAAACATCTCTTGCT	651	1.0 X
TLR9_08	MID11	tgatacgtctCGTCAGCTCAAAGGACTTCA	tgatacgtctAGGGTGTGCAGATGGTTCTC	546	1.0 X
TLR9_09	MID11	tgatacgtctGGGAGACCTCTATCTCTGCTTT	tgatacgtctCGCTCACGTCTAGGATTTTC	428	1.0 X
TLR9_10	MID11	tgatacgtctCCTCCTGGTTCGGTTCCTTA	tgatacgtctCGGTTATAGAAGTGACGGTTG	837	1.0 X
TLR1_01	MID1	acgagtgcgtATGCCTGACATCCTCTCACT	acgagtgcgtAGTTCCAGACTCACTGTGGTG	470	1.0 X
TLR1_02	MID1	acgagtgcgtTCCAGTGTGCAGTCAATCAC	acgagtgcgtAGAACCTTGATCTGAGGAGGT	890	1.0 X
TLR1_03	MID1	acgagtgcgtTGACCCAGGAAATGAAGTCT	acgagtgcgtCCGTGTTAATGTATTTCTGCTG	1195	1.0 X
TLR5_01	MID5	atcagacacgTTTGGGAAACGGAGGATAAG	atcagacacgGCACCTTTGAGGCTGTGA	642	1.0 X
TLR5_02	MID5	atcagacacgGCCTGCTTTTGATACTTTGG	atcagacacgAGGTGTCCGCTATGTTCTCA	661	1.0 X
TLR5_03	MID5	atcagacacgTCCCTTACCTTCCAGCAGA	atcagacacgAAGTTGGGGAAAACATTAGG	563	1.0 X
TLR5_04	MID5	atcagacacgGGCAGATTAGAGGGGAAAGA	atcagacacgCCATCAAAGAAGCAGGAAGA	541	1.0 X
TLR5_05	MID5	atcagacacgTCACTCTCCCTTCTTCTCCA	atcagacacgCAGACACTTGTTCCAGTCCA	687	1.0 X
TLR5_06	MID5	atcagacacgCCTCCAAGGGAAAACACTCT	atcagacacgATTGGCTGTAAGTGGGATGT	700	1.0 X
TLR5_07	MID5	atcagacacgTTTTCTTCCAAGCATTCCTA	atcagacacgAGCCAGAGAGTTTGGGTACA	652	1.0 X
TLR5_08	MID5	atcagacacgGAAACCAGCTCCTCTCTCCT	atcagacacgATCTTTCTGCTGCTCCACAC	592	1.0 X
TLR5_09	MID5	atcagacacgAGACTTTGAATGGGTGCAGA	atcagacacgTGGTAACTGGCGGAAATAAA	541	1.0 X
TLR5_10	MID5	atcagacacgGGAGCAGTTTCCACTTATCG	atcagacacgATTCTCATGCCGGTTTCTTT	764	1.0 X
TLR5_10FixR	MID5		atcagacacgATTCTCATGCTGGCTTCTTT	764	1.0 X
TLR10_01	MID10	tctctatgcgCTGAGGTGAACCAGTGATAAAA	tctctatgcgATCGTCCCAGGATAAGTCAA	813	1.0 X
TLR10_02	MID10	tctctatgcgTGCCCATCTTAAACACAACA	tctctatgcgACCCAAAAACAGAATCAGCA	768	1.0 X
TLR10_03	MID10	tctctatgcgCCAGCAACACATCCCTGA	tctctatgcgAAAGTGGAGGCAGCAGAAG	623	1.0 X
TLR10_04	MID10	tctctatgcgATTGTGGTTGTCATGCTCGT	tctctatgcgAACCTCCAAACCCTTCATTC	655	1.0 X
TLR10_05	MID10	tctctatgcgTTTATTAGACACCAGAGGGACA	tctctatgcgGCGGATTCTTTGTGATTGAG	748	1.0 X
TLR10_06	MID10	tctctatgcgTATTGTTGGCTGCACTGAGA	tctctatgcgAGACGTGTGTTCTGGGAAAG	868	1.0 X
TLR4_1	MID4	agcactgtagCGGGGAGAGACGACACTACA	agcactgtagTGTTTGCAAATGAACCTAACCA	288	1.0 X

TLR4_2	MID4	agcactgtagTCTTTGCTCGTCCCAGTAGC	agcactgtagAAGTGAATGAAAAGGAGACCTCA	384	1.0 X
TLR4_3	MID4	agcactgtagGAAATTGGCATTCAGTGGTC	agcactgtagCCGTCAGTATCAAGGTGGAG	486	1.0 X
TLR4_4	MID4	agcactgtagCTTTGTTTCATCTGCCTTGC	agcactgtagTTGAGTAGGGGCATTTGATG	508	1.0 X
TLR4_5	MID4	agcactgtagGCCTTTTCTGGGCTATCAAG	agcactgtagTATCGTCCCCTGAGAATTTG	541	1.0 X
TLR4_6	MID4	agcactgtagGATCTTTCCTGGAGGGACTG	agcactgtagATCAAGGTAGCGGAGGTTTC	539	1.0 X
TLR4_7	MID4	agcactgtagCTGGATTTTCAGCATTCCAC	agcactgtagTCTGCACACATCATTTGCTC	554	1.0 X
TLR4_8	MID4	agcactgtagGGCCTCTAAGGAGCAAGAAC	agcactgtagTAACCTTACGGCTTTTGTGG	533	1.0 X
TLR4_9	MID4	agcactgtagCTTTCAGCTCTGCCTTCACT	agcactgtagGCGTACCACTGAATCACCA	535	1.0 X

APPENDIX B

Table B1. Normalization Protocols.

Amplicon	Size(BP)	Mean	Difference	Adjust	Adjust	Add ul
1-1	550	581.8972	31.8972	Yes	0.0548159	2.054816
1-2	554	581.8972	27.8972	Yes	0.0479418	3.808233
1-3	597	581.8972	-15.1028			4
1-4	587	581.8972	-5.1028			4
1-5	598	581.8972	-16.1028			4
1-6	573	581.8972	8.8972			4
1-7	528	581.8972	53.8972	Yes	0.0926232	3.629507
2-1	596	581.8972	-14.1028			4
2-2	580	581.8972	1.8972			4
2-3	564	581.8972	17.8972			4
2-4	554	581.8972	27.8972	Yes	0.0479418	3.808233
2-5	600	581.8972	-18.1028			4
2-6	550	581.8972	31.8972	Yes	0.0548159	3.780737
2-7	554	581.8972	27.8972	Yes	0.0479418	3.808233
2-8	561	581.8972	20.8972			4
3-20	573	581.8972	8.8972			4
3-21	577	581.8972	4.8972			4
3-22	594	581.8972	-12.1028			4
3-24	588	581.8972	-6.1028			4
3-25	576	581.8972	5.8972			4
3-26	572	581.8972	9.8972			4
3-27	583	581.8972	-1.1028			4
3-28	589	581.8972	-7.1028			4
3-29	572	581.8972	9.8972			4
3-30	553	581.8972	28.8972	Yes	0.0496603	3.801359
3-31	567	581.8972	14.8972			4
3-32	574	581.8972	7.8972			4
3-33	558	581.8972	23.8972	Yes	0.0410677	3.835729
3-34	596	581.8972	-14.1028			4
3-35	582	581.8972	-0.1028			4
3-36	568	581.8972	13.8972			4
3-37	560	581.8972	21.8972			4
3-38	579	581.8972	2.8972			4
3-39	579	581.8972	2.8972			4
4-1	562	581.8972	19.8972			4
4-2	570	581.8972	11.8972			4
4-3	587	581.8972	-5.1028			4
4-4	564	581.8972	17.8972			4
4-5	600	581.8972	-18.1028			4
4-6	597	581.8972	-15.1028			4
4-7	597	581.8972	-15.1028			4
4-8	591	581.8972	-9.1028			4
4-9	591	581.8972	-9.1028			4
4-10	586	581.8972	-4.1028			4
4-11	552	581.8972	29.8972	Yes	0.0513788	3.794485
4-12	572	581.8972	9.8972			4
4-17	593	581.8972	-11.1028			4
4-18	564	581.8972	17.8972			4
4-19	583	581.8972	-1.1028			4
4-20	580	581.8972	1.8972			4
4-21	595	581.8972	-13.1028			4
4-22	565	581.8972	16.8972			4
4-23	576	581.8972	5.8972			4
4-24	558	581.8972	23.8972	Yes	0.0410677	3.835729
6-1	553	581.8972	28.8972	Yes	0.0496603	3.801359

6-2	580	581.8972	1.8972			4
6-3	562	581.8972	19.8972			4
6-3M	803	581.8972	-221.1028	Yes	-0.379969	2.480125
6-4	976	581.8972	-394.1028	Yes	-0.677272	1.290911
7-1	558	581.8972	23.8972	Yes	0.0410677	3.835729
7-2	580	581.8972	1.8972			4
7-3	553	581.8972	28.8972	Yes	0.0496603	3.801359
7-4	569	581.8972	12.8972			4
7-5	565	581.8972	16.8972			4
7-20	593	581.8972	-11.1028			4
7-21	570	581.8972	11.8972			4
7-22	599	581.8972	-17.1028			4
7-23	553	581.8972	28.8972	Yes	0.0496603	3.801359
7-24	576	581.8972	5.8972			4
7-46	570	581.8972	11.8972			4
7-47	567	581.8972	14.8972			4
7-48	579	581.8972	2.8972			4
7-49	590	581.8972	-8.1028			4
7-50	567	581.8972	14.8972			4
7-51	582	581.8972	-0.1028			4
7-52	552	581.8972	29.8972	Yes	0.0513788	3.794485
7-53	577	581.8972	4.8972			4
7-54	580	581.8972	1.8972			4
8-11	598	581.8972	-16.1028			4
8-12	559	581.8972	22.8972			4
8-13	552	581.8972	29.8972	Yes	0.0513788	3.794485
8-14	561	581.8972	20.8972			4
8-15	557	581.8972	24.8972	Yes	0.0427863	3.828855
8-16	600	581.8972	-18.1028			4
8-17	620	581.8972	-38.1028	Yes	-0.06548	3.738079
8-18	569	581.8972	12.8972			4
8-19	589	581.8972	-7.1028			4
8-20	573	581.8972	8.8972			4
8-21	576	581.8972	5.8972			4
9-1	555	581.8972	26.8972	Yes	0.0462233	3.815107
9-2	570	581.8972	11.8972			4
9-3	595	581.8972	-13.1028			4
9-4	598	581.8972	-16.1028			4
9-5	586	581.8972	-4.1028			4
9-6	575	581.8972	6.8972			4
9-7	564	581.8972	17.8972			4
9-8	600	581.8972	-18.1028			4
9-9	578	581.8972	3.8972			4
9-10	604	581.8972	-22.1028			4
9-11	560	581.8972	21.8972			4
10-1	553	581.8972	28.8972	Yes	0.0496603	3.801359
10-2	598	581.8972	-16.1028			4
10-3	600	581.8972	-18.1028			4
10-4	593	581.8972	-11.1028			4
10-5	587	581.8972	-5.1028			4
10-6	597	581.8972	-15.1028			4
10-7	573	581.8972	8.8972			4
Average	581.8972					
Median	576					
Stand Dev	47.25262					
Half Stand Dev	23.62631					
	Edges					
Left	Right					
534.64458	629.1498					
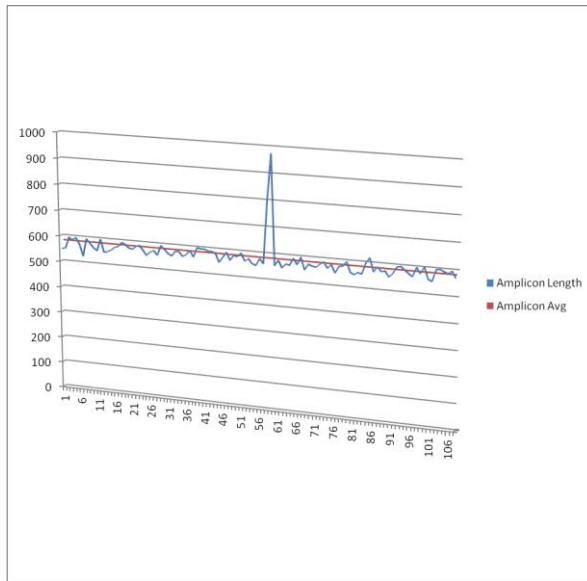
Figure B1. Graph of amplicon sizes

Table B2. Validated SNPs.

Gene_SNP or Indel	Variation	ECA	NonSynonymous	MAF
TLR1_G682T	G/T	3	X	0.099
TLR1_A707G	A/G	3	X	0.053
TLR1_A869G	A/G	3	X	0.255
TLR1_C1485G	C/G	3	X	0.036
TLR1_G1717A	G/A	3	X	0.078
TLR1_A2212G	A/G	3	X	0.047
TLR2_G157A	G/A	2		0.385
TLR2_T241C	T/C	2		0.047
TLR2_T305A	T/A	2		0.005
TLR2_A370G	A/G	2		0.135
TLR2_C533T	C/T	2		0.042
TLR2_G609C	G/C	2	X	0.048
TLR2_G1514C	G/C	2		0.453
TLR2_G1639A	G/A	2	X	0.395
TLR2_T2159C	T/C	2		0.042
TLR2_G2305A	G/A	2	X	0.172
TLR2_C2495T	C/T	2		0.057
TLR3_T7292C	T/C	27		0.016
TLR3_T7837G	T/G	27		0.074
TLR3_G7867A	G/A	27		0.042
TLR3_A7922C	A/C	27	X	0.073
TLR3_A8302T	A/T	27		0.174
TLR3_G8818A	G/A	27		0.182
TLR3_T9306C	T/C	27		0.375
TLR3_C9397A	C/A	27		0.089
TLR3_A10278T	A/T	27		0.016
TLR3_G10520A	G/A	27		0.021
TLR3_G10677C	G/C	27		0.448
TLR3_T11004A	T/A	27		0.032
TLR3_G11043C	G/C	27		0.073
TLR3_T11296C	T/C	27		0.189
TLR3_C11308T	C/T	27		0.026
TLR3_C11380T	C/T	27		0.12
TLR3_C11390T	C/T	27		0.016
TLR3_A11498C	A/C	27		0.214
TLR3_A11513G	A/G	27		0.099
TLR3_A11559G	A/G	27		0.438

TLR3_C11609T	C/T	27		0.01
TLR3_T11680C	T/C	27		0.193
TLR3_C11874T	C/T	27		0.255
TLR3_C12094A	C/A	27		0.12
TLR3_C12163T	C/T	27		0.078
TLR3_A12339G	A/G	27		0.12
TLR3_C12503A	C/A	27	X	0.279
TLR3_T13787A	T/A	27	X	0.005
TLR3_A14040G	A/G	27		0.016
TLR3_T14310G	T/G	27		0.031
TLR3_G14470T	G/T	27		0.026
TLR4_T525C	T/C	25		0.214
TLR4_C599T	C/T	25	X	0.026
TLR4_C658T	C/T	25	X	0.062
TLR4_A909C	A/C	25		0.021
TLR4_G1027T	G/T	25		0.214
TLR4_A1030T	A/T	25		0.005
TLR4_G1267A	G/A	25		0.219
TLR4_T1538C	T/C	25		0.07
TLR4_A1580C	A/C	25		0.074
TLR4_C1687T	C/T	25		0.021
TLR4_G1810T	G/T	25		0.078
TLR4_T1832C	T/C	25		0.026
TLR4_C1869T	C/T	25		0.391
TLR4_T1900C	T/C	25		0.216
TLR4_C1979T	C/T	25		0.026
TLR4_A2091G	A/G	25		0.083
TLR4_A2210C	A/C	25		0.005
TLR4_A2349G	A/G	25		0.106
TLR4_C2355T	C/T	25		0.031
TLR4_A2661C	A/C	25		0.216
TLR4_G2666T	G/T	25		0.078
TLR4_A2967G	A/G	25		0.214
TLR4_A3061G	A/G	25		0.214
TLR4_C3238T	C/T	25		0.149
TLR4_C3256A	C/A	25		0.016
TLR4_A3336G	A/G	25		0.151
TLR4_A3780G	A/G	25		0.032
TLR4_T4082G	T/G	25		0.337
TLR4_A4277C	A/C	25		0.037
TLR4_T4335C	T/C	25		0.052
TLR4_C4612T	C/T	25		0.347
TLR4_C4997G	C/G	25		0.109
TLR4_C5036G	C/G	25		0.349
TLR4_G5347A	G/A	25		0.121
TLR4_T6935C	T/C	25		0.347
TLR4_A6941G	A/G	25		0.214
TLR4_T6955C	T/C	25		0.333
TLR4_G7052A	G/A	25		0.198
TLR4_G7085T	G/T	25		0.344
TLR4_A7334G	A/G	25		0.344
TLR4_G7350A	G/A	25		0.079
TLR4_C7439G	C/G	25		0.026
TLR4_A7448G	A/G	25		0.344
TLR4_G7660A	G/A	25		0.34
TLR4_T7721C	T/C	25		0.344
TLR4_T7757A	T/A	25		0.016
TLR4_C7760T	C/T	25		0.344
TLR4_C7790T	C/T	25		0.036
TLR4_T8096A	T/A	25		0.021
TLR4_T8456G	T/G	25		0.344
TLR4_A8604G	A/G	25	X	0.344
TLR4_C8649A	C/A	25	X	0.353
TLR4_C8753T	C/T	25		0.116
TLR4_C8939T	C/T	25		0.203
TLR4_T9289C	T/C	25	X	0.405

TLR4_T9449C	T/C	25		0.344
TLR4_G9482A	G/A	25		0.385
TLR4_A9609G	A/G	25	X	0.347
TLR4_T9707C	T/C	25		0.104
TLR4_A9794G	A/G	25		0.036
TLR4_T9839C	T/C	25		0.38
TLR4_A10353T	A/T	25		0.385
TLR4_C10384T	C/T	25		0.021
TLR6_T1728C	T/C	3	X	0.214
TLR6_A2646G	A/G	3	X	0.037
TLR7_T827A	T/A	X		0.021
TLR7_G846A	G/A	X		0.156
TLR7_T966A	T/A	X		0.01
TLR7_A1197C	A/C	X		0.082
TLR7_A1207C	A/C	X		0.083
TLR7_A1453T	A/T	X		0.073
TLR7_C1907T	C/T	X		0.041
TLR7_G1908A	G/A	X		0.042
TLR7_C1922T	C/T	X		0.021
TLR7_G2007A	G/A	X		0.041
TLR7_A2014G	A/G	X		0.041
TLR7_A8436G	A/G	X		0.156
TLR7_G8533A	G/A	X		0.072
TLR7_C8651T	C/T	X		0.062
TLR7_C8681T	C/T	X		0.371
TLR7_C8698T	C/T	X		0.155
TLR7_C8730T	C/T	X		0.147
TLR7_A9370G	A/G	X		0.021
TLR7_C9499T	C/T	X		0.094
TLR7_C9995G	C/G	X		0.385
TLR7_G10099C	G/C	X		0.072
TLR7_T10294C	T/C	X		0.103
TLR7_G10436A	G/A	X		0.021
TLR7_A19828G	A/G	X		0.32
TLR7_G19863A	G/A	X		0.144
TLR7_A20133G	A/G	X		0.01
TLR7_G20283A	G/A	X		0.24
TLR7_G20781T	G/T	X	X	0.01
TLR7_T21348C	T/C	X		0.031
TLR7_C21672G	C/G	X		0.175
TLR7_C21916T	C/T	X		0.082
TLR8_T4817C	T/C	X		0.156
TLR8_G4946T	G/T	X		0.042
TLR8_C5437T	C/T	X		0.064
TLR8_T5591C	T/C	X		0.255
TLR8_G5604T	G/T	X		0.052
TLR8_C5724G	C/G	X		0.053
TLR8_A5773G	A/G	X		0.073
TLR8_A6148G	A/G	X	X	0.021
TLR8_T6151G	T/G	X	X	0.01
TLR8_T6537G	T/G	X	X	0.095
TLR8_C6740T	C/T	X		0.074
TLR8_A8140G	A/G	X	X	0.01
TLR8_A8516G	A/G	X		0.25
TLR8_T8786C	T/C	X		0.213
TLR8_G9005A	G/A	X		0.042
TLR8_A9056G	A/G	X		0.188
TLR8_T9278C	T/C	X		0.053
TLR9_C127G	C/G	16		0.021
TLR9_T462G	T/G	16		0.079
TLR9_C739T	C/T	16		0.083
TLR9_C852T	C/T	16		0.307
TLR9_A916C	A/C	16		0.224
TLR9_C1387T	C/T	16	X	0.121
TLR9_C1471T	C/T	16	X	0.036
TLR9_G3683A	G/A	16		0.068

TLR9_C3749G	C/G	16		0.01
TLR9_C3914T	C/T	16		0.307
TLR10_A349G	A/G	3	X	0.302
TLR10_A1011T	A/T	3	X	0.026
TLR10_T1065G	T/G	3	X	0.005
TLR10_G1474A	G/A	3	X	0.302
TLR10_T1690C	T/C	3		0.042
TLR10_C1807T	C/T	3		0.245
TLR10_T1909C	T/C	3	X	0.052
TLR10_A2111G	A/G	3	X	0.116

Table B3. Tag SNPs.

Gene	tagSNP	Total Alleles Captured
TLR1	TLR1_G682T	6 of 6 (100%) with these 5
	TLR1_A869G	
	TLR1_C1485G	
	TLR1_G1717A	
	TLR1_A2212G	
TLR2	TLR2_G157A	11 of 11 (100%) with these 8
	TLR2_T305A	
	TLR2_A370G	
	TLR2_C533T	
	TLR2_G1639A	
	TLR2_T2159C	
	TLR2_G2305A	
	TLR2_C2495T	
TLR3	TLR3_T7837G	31 of 31 (100%) with these 25
	TLR3_G7867A	
	TLR3_A7922C	
	TLR3_A8302T	
	TLR3_G8818A	
	TLR3_T9306C	
	TLR3_C9397A	
	TLR3_G10520A	
	TLR3_G10677C	
	TLR3_T11004A	
	TLR3_G11043C	
	TLR3_C11308T	
	TLR3_C11380T	
	TLR3_C11390T	
	TLR3_A11513G	
	TLR3_A11559G	
	TLR3_C11609T	
	TLR3_T11680C	
	TLR3_C11874T	
	TLR3_C12163T	
	TLR3_A12339G	
	TLR3_C12503A	
	TLR3_T13787A	
	TLR3_T14310G	
	TLR3_G14470T	
TLR4	TLR4_T525C	63 of 63 (100%) with these 62
	TLR4_C599T	
	TLR4_C658T	
	TLR4_A909C	
	TLR4_G1027T	
	TLR4_A1030T	
	TLR4_G1267A	
	TLR4_T1538C	
	TLR4_C1687T	
	TLR4_G1810T	
	TLR4_T1832C	
	TLR4_C1869T	
	TLR4_T1900C	

	TLR4_C1979T	
	TLR4_A2091G	
	TLR4_A2210C	
	TLR4_A2349G	
	TLR4_C2355T	
	TLR4_A2661C	
	TLR4_G2666T	
	TLR4_A2967G	
	TLR4_A3061G	
	TLR4_C3238T	
	TLR4_C3256A	
	TLR4_A3336G	
	TLR4_A3780G	
	TLR4_T4082G	
	TLR4_A4277C	
	TLR4_T4335C	
	TLR4_C4612T	
	TLR4_C4997G	
	TLR4_C5036G	
	TLR4_G5347A	
	TLR4_T6935C	
	TLR4_A6941G	
	TLR4_T6955C	
	TLR4_G7052A	
	TLR4_G7085T	
	TLR4_A7334G	
	TLR4_G7350A	
	TLR4_C7439G	
	TLR4_A7448G	
	TLR4_G7660A	
	TLR4_T7721C	
	TLR4_T7757A	
	TLR4_C7760T	
	TLR4_C7790T	
	TLR4_T8096A	
	TLR4_T8456G	
	TLR4_A8604G	
	TLR4_C8649A	
	TLR4_C8753T	
	TLR4_C8939T	
	TLR4_T9289C	
	TLR4_T9449C	
	TLR4_G9482A	
	TLR4_A9609G	
	TLR4_T9707C	
	TLR4_A9794G	
	TLR4_T9839C	
	TLR4_A10353T	
	TLR4_C10384T	
TLR6	TLR6_T1728C	2 of 2 (100%) with these 2
	TLR6_A2646G	
TLR7	TLR7_T827A	31 of 31 (100%) with these 25
	TLR7_G846A	
	TLR7_T966A	
	TLR7_A1207C	
	TLR7_C1907T	
	TLR7_G1908A	
	TLR7_C1922T	
	TLR7_C8651T	
	TLR7_C8681T	
	TLR7_C8698T	
	TLR7_C8730T	
	TLR7_A9370G	
	TLR7_C9499T	
	TLR7_C9995G	
	TLR7_G10099C	

| | TLR7_T10294C | |
	TLR7_G10436A	
	TLR7_A19828G	
	TLR7_G19863A	
	TLR7_A20133G	
	TLR7_G20283A	
	TLR7_G20781T	
	TLR7_T21348C	
	TLR7_C21672G	
	TLR7_C21916T	
TLR8	TLR8_T4817C	17 of 17 (100%) with these 16
	TLR8_G4946T	
	TLR8_C5437T	
	TLR8_T5591C	
	TLR8_G5604T	
	TLR8_C5724G	
	TLR8_A5773G	
	TLR8_A6148G	
	TLR8_T6151G	
	TLR8_T6537G	
	TLR8_A8140G	
	TLR8_A8516G	
	TLR8_T8786C	
	TLR8_G9005A	
	TLR8_A9056G	
	TLR8_T9278C	
TLR9	TLR9_C127G	10 of 10 (100%) with these 8
	TLR9_T462G	
	TLR9_C852T	
	TLR9_A916C	
	TLR9_C1387T	
	TLR9_C1471T	
	TLR9_G3683A	
	TLR9_C3749G	
TLR10	TLR10_A349G	8 of 8 (100%) with these 7
	TLR10_A1011T	
	TLR10_T1065G	
	TLR10_T1690C	
	TLR10_C1807T	
	TLR10_T1909C	
	TLR10_A2111G	
Totals	158 tags	179 Total variable sites
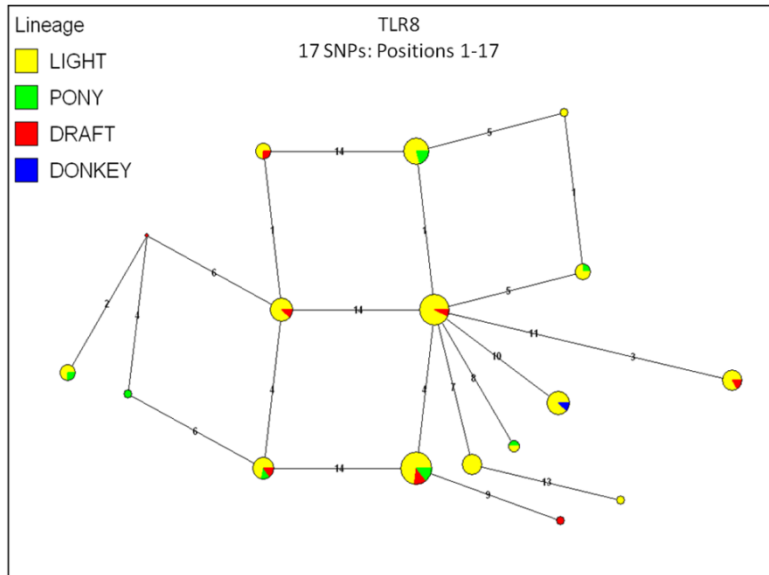
Figure B2. Median joining (MJ) haplotype network for equine *TLR8*. Because MJ networks require the absence of recombination [73], each network represents intragenic regions of elevated LD. Haplotypes predicted for light horses, ponies, draft horses, and donkeys are color coded. Numbers indicate SNP positions in numerical order (see Table B2 for SNP information). Node sizes are proportional to haplotype frequency, and all branch lengths are drawn to scale.

Table B4. Barcoded Primers.

Genes		Primers F	Primers R	Amplicon Size (Excluding MIDs)	MgCl2 Concentration
TLR1_01	MID1	acgagtgcgtCATACAGCCTACCCCTTCCT	acgagtgcgtAAAGATGATGGCAAAGTGGA	550	1.0 X
TLR1_02	MID1	acgagtgcgtATATTTCCCTGGCTTCAGGT	acgagtgcgtGTTCACAGTAGGGTGGCAAG	554	1.0 X
TLR1_03	MID1	acgagtgcgtTCGGACTTCTGACATCCAAT	acgagtgcgtTCCCAAGCTGTTTCAATGTT	597	1.0 X
TLR1_04	MID1	acgagtgcgtATTTTGGATGTGTCCCTCAG	acgagtgcgtAGAGACTTCATCTCCTTGGTCA	587	1.0 X
TLR1_05	MID1	acgagtgcgtCCTTGGCTTCAGAGATTTTG	acgagtgcgtTCCAGGAAGGTTGGTTAAAGA	598	1.0 X
TLR1_06	MID1	acgagtgcgtGAGCATCCCTAAACAAATCATG	acgagtgcgtGCAGAATCGTGCCCACTA	573	1.0 X
TLR1_07	MID1	acgagtgcgtTAGTGGGCACGATTCTGC	acgagtgcgtCGCGTTTAGATTCAGGTCAA	528	1.0 X
TLR2_1	MID2	acgctcgacaGGAGCATGGGACCTTTAACT	acgctcgacaCTTCCTTGGAGAGGCTGATT	596	1.0 X

TLR2_2	MID2	acgctcgacaATCACTGGACAATGCCACAT	acgctcgacaTCATACCGCTGGAGATTTGT	580	1.0 X
TLR2_3	MID2	acgctcgacaTTACAAAACACTCGGGGAAA	acgctcgacaACTGTGGAATACGCAACCTC	564	1.0 X
TLR2_4	MID2	acgctcgacaTCTGGAGTGTCAGAAGCAGA	acgctcgacaACTTCCAGTGTCTGGGGAAT	554	1.0 X
TLR2_5	MID2	acgctcgacaCCCTCCCTCCAAACCTTAAT	acgctcgacaTCAGAGACCGAGAGACGAGT	600	1.0 X
TLR2_6	MID2	acgctcgacaGCCTCCTTCTTACCCATGTT	acgctcgacaCCCGCTTATGAAGACACAAC	550	1.0 X
TLR2_7	MID2	acgctcgacaGTGGTACATGAAAATGATGTGG	acgctcgacaTAAAGACCACCAGCAAACCA	554	1.0 X
TLR2_8	MID2	acgctcgacaGAGCCACAAAACCATCTTTG	acgctcgacaGCCAACTGCTACAGCTAATTCA	561	1.0 X
TLR3_20	MID3	agacgcactcAGATTTGGTGTAAAGGTGGTTG	agacgcactcAAAGGGCAACAGTCTCAAAA	573	1.0 X
TLR3_21	MID3	agacgcactcGGGTGGAGGTGAAGAATGA	agacgcactcGCTCATTGTGCTGGAGGTC	577	1.0 X
TLR3_22	MID3	agacgcactcGCATCTCAAAACTAGAGCCAGA	agacgcactcTCTCCCCATTTAGGCAAATC	594	1.0 X
TLR3_24	MID3	agacgcactcAACTCTGCCTCTCTCCCTTC	agacgcactcTATTCTCGGCATCTGAGGTC	588	1.0 X
TLR3_25	MID3	agacgcactcAACCAAAGCCATTGTCAAAA	agacgcactcGTCATGCAACAGCCTTGTC	576	1.0 X
TLR3_26	MID3	agacgcactcTTCCCTTTTACCTGAGTGGA	agacgcactcTGTTAGAGTCTTGCCATCAAAA	572	1.0 X
TLR3_27	MID3	agacgcactcTGATCCTTGAAGTAAGAAACCA	agacgcactcGTCACCCAACTCCTTTCCTT	583	1.0 X
TLR3_28	MID3	agacgcactcCCTTGGCAATTCTTCTTTGA	agacgcactcACCCAGCCCAATGAAATAGT	589	1.0 X
TLR3_29	MID3	agacgcactcCCCTATCTGGGCTTTCTCTC	agacgcactcCCCAAATGCTAACACTGGTT	572	1.0 X
TLR3_30	MID3	agacgcactcAATCGCAAACCAAATCAGAA	agacgcactcAAGAGGCAACATAGCACAGC	553	1.0 X
TLR3_31	MID3	agacgcactcAACTGTATGGCAGGCACTGT	agacgcactcCAAGAGAACAGAAGGCAGGA	567	1.0 X
TLR3_32	MID3	agacgcactcGACATGGTGGAAATCAGAGC	agacgcactcTCTCCATCCCTCTACTGCAC	574	1.0 X
TLR3_33	MID3	agacgcactcAGAAAACCTTGGTGGAAAGC	agacgcactcCTGTTCAGAGAGAGGCCAAA	558	1.0 X
TLR3_34	MID3	agacgcactcCAGACTGTTGCGTTTTGGTT	agacgcactcCGCAAACTTGAAAAGGAGTT	596	1.0 X
TLR3_35	MID3	agacgcactcCATTTCTCTTGCTTCGCTTC	agacgcactcTTCTCAAGACCCTCCAACAG	582	1.0 X
TLR3_36	MID3	agacgcactcTTCGCCCTCTTCATAACTTG	agacgcactcAAAAGCATCACTGGGAAACC	568	1.0 X
TLR3_37	MID3	agacgcactcTGACGATCAGGTGTCTCTGA	agacgcactcAGTCCCTTTCTTCCAGACAAA	560	1.0 X
TLR3_38	MID3	agacgcactcAAGATAGGGACTGGGTCTGG	agacgcactcGTTCAAGATACAGCGCGATT	579	1.0 X
TLR3_39	MID3	agacgcactcGCATCAAAAGGAGCAGAAAA	agacgcactcATTGGGAAAATTACGCCTTT	579	1.0 X
TLR4_1	MID4	agcactgtagAAGAAAATTGAAGTCACCATCC	agcactgtagGCAGAGAGCAGCTTTTCAGA	562	1.0 X
TLR4_2	MID4	agcactgtagACAGAAAATGCCAGGATGAT	agcactgtagCTTCTCCCTGAGATTGAAAGG	570	1.0 X
TLR4_3	MID4	agcactgtagAGAGCTGTAGGAAGGCTGCT	agcactgtagGCACAGAGAGGAGGATGAGA	587	1.0 X
TLR4_4	MID4	agcactgtagTGGCAGGGTTAGAAACAAGA	agcactgtagTGATATTACCATAGGGCACCA	564	1.0 X
TLR4_5	MID4	agcactgtagTCCTGAGTCTCCGTTCTAACA	agcactgtagGGAGATGACATTGAAGCAGAA	600	1.0 X
TLR4_6	MID4	agcactgtagTATCCACCTTTCTCGTGCTC	agcactgtagGATTGGAGGAGGCTTCTGAG	597	1.0 X
TLR4_7	MID4	agcactgtagAGTTTCATCACTCCCACCAG	agcactgtagGGACCTCTTCCTTACCCTCTT	597	1.0 X
TLR4_8	MID4	agcactgtagATACACATGGAACAGCCACA	agcactgtagAGTGCTATATGCTGCCTTGG	591	1.0 X
TLR4_9	MID4	agcactgtagGATGGCCTGTGGGACTATCT	agcactgtagGTACAGCATTGGGGAGACTG	591	1.0 X
TLR4_10	MID4	agcactgtagCGAAGTACAATGAGCCAAGG	agcactgtagTGCCCTTTTATCTCCCTTCT	586	1.0 X

TLR4_11	MID4	agcactgtagGGATGGTTGAAAGAAGGTTG	agcactgtagGAGACTTGGAGATGGGAGGT	552	1.0 X
TLR4_12	MID4	agcactgtagACCCTACACAAGGTGAAATGTT	agcactgtagACTACCCACCACAGACAAGG	572	1.0 X
TLR4_17	MID4	agcactgtagACAGAGGCACCAGAACTCAG	agcactgtagAGCAGGCTTCTCTGAAAACA	593	1.0 X
TLR4_18	MID4	agcactgtagTCACCGTTCCTCCACATATC	agcactgtagAACAATCCCAGCTCTTCACA	564	1.0 X
TLR4_19	MID4	agcactgtagATCCCGAATCTTCAGAGCTT	agcactgtagAAGGCATCTGGTTGGATAAA	583	1.0 X
TLR4_20	MID4	agcactgtagTTGCGTGTGCTACATCAAAT	agcactgtagAACTCAAGCGATTTCTGCTG	580	1.0 X
TLR4_21	MID4	agcactgtagCACCTCTCTCAAAAGGTTGG	agcactgtagAATCTGGAGGGAATGGAGAG	595	1.0 X
TLR4_22	MID4	agcactgtagACCACCCTGGACCTTTCTAA	agcactgtagCTTCGTCCTGGCTTGAGTAG	565	1.0 X
TLR4_23	MID4	agcactgtagTCATGGTTTCTGTCATAGCAGT	agcactgtagGTCTGCTTTCTGCTGCATCT	576	1.0 X
TLR4_24	MID4	agcactgtagAGGAAGGTTTCCACAAAAGC	agcactgtagCCCCTGGAGGTTCTGTATTT	558	1.0 X
TLR6_1	MID6	atatcgcgagTTGGAACCATAATCCAATTCTC	atatcgcgagTATGGGTGGAAAACAAGCTG	553	1.0 X
TLR6_2	MID6	atatcgcgagAGAAGATTTCCTGCCATCCT	atatcgcgagGCTTTCAATGCCGTTTTAGA	580	1.0 X
TLR6_3	MID6	atatcgcgagCCATTAGACAGCCAACCCTA	atatcgcgagACCACTAGACTCTCAACCCAAG	562	1.0 X
TLR6_3M	MID6		atatcgcgagGGCTGATGGGTCGGAAA	803	1.0 X
TLR6_4	MID6	atatcgcgagTTTCCGACCCATCAGCC	atatcgcgagGCAGATAATGGAGGCACAAT	976	1.0 X
TLR7_01	MID7	cgtgtctctaAGGGGAGGAGAGAGAACTGA	cgtgtctctaAAGCTGGACAGAGAAAGTGC	558	1.0 X
TLR7_02	MID7	cgtgtctctaCTGATCTTGACGCCTCTCAT	cgtgtctctaGGTCTCTTTTCCCCTATTGC	580	1.0 X
TLR7_03	MID7	cgtgtctctaATGGTGTCCTCTGAACGAGT	cgtgtctctaTAAGAGAGCTTGGGTGATGG	553	1.0 X
TLR7_04	MID7	cgtgtctctaGCTAAAACTGGGCAGATGAA	cgtgtctctaATCCTATCATGCCATCCTCA	569	1.0 X
TLR7_05	MID7	cgtgtctctaTGGGGTACTCTCTTACAAAGGA	cgtgtctctaTCTGGAAGTGGAGTTTCCAT	565	1.0 X
TLR7_20	MID7	cgtgtctctaCAAACCCACAAATGGTTGTC	cgtgtctctaCTCAGTGAACCAAGCCTTTC	593	1.0 X
TLR7_21	MID7	cgtgtctctaACGGTTTGAAAGGGGAAAT	cgtgtctctaAAGGGGTTGAAGTGGGAAAAG	570	1.0 X
TLR7_22	MID7	cgtgtctctaTTGTGTTCTCACTGGGGTTT	cgtgtctctaAGGATGGGCACGATGTTAG	599	1.0 X
TLR7_23	MID7	cgtgtctctaTTGATGAGCGGTGTGTAGGT	cgtgtctctaTTTCCAGAAGTCTCCACGTC	553	1.0 X
TLR7_24	MID7	cgtgtctctaGAGCCTTGATTTATTCAGCAAA	cgtgtctctaTTTTCTGTAAAGGGCCAGATAG	576	1.0 X
TLR7_46	MID7	cgtgtctctaATGGGGATAATGGATCTCCT	cgtgtctctaCCAAGGAGTTTGGAAATTAGG	570	1.0 X
TLR7_47	MID7	cgtgtctctaGCAAAACAGAGGCAGTAAATG	cgtgtctctaAAAAGATATTGTTGGCCTCAAG	567	1.0 X
TLR7_48	MID7	cgtgtctctaTGCAGATTAAACCCAGAAGC	cgtgtctctaATCTTTGGGGCACATACTGA	579	1.0 X
TLR7_49	MID7	cgtgtctctaAGCTGCAAATTCTTGACCTAAG	cgtgtctctaAGAGCAGAAGCCAACTTCAC	590	1.0 X
TLR7_50	MID7	cgtgtctctaTCCTTGATCTTGGCACTAACTT	cgtgtctctaGCATGTGAGTAATTCCCTCTG	567	1.0 X
TLR7_51	MID7	cgtgtctctaCTCTACTCGACGGCTTTTGA	cgtgtctctaTCGTAACTGGAAAGCATCTTG	582	1.0 X
TLR7_52	MID7	cgtgtctctaCCAGGAGCCTCAAGAAACTA	cgtgtctctaCTGTAACCGCTGGGTCTTTA	552	1.0 X
TLR7_53	MID7	cgtgtctctaCCACAGCGAATCACCTCTAT	cgtgtctctaACCGTCTCTTTGAACACCTG	577	1.0 X
TLR7_54	MID7	cgtgtctctaCAGAAGTCCAAATTCCTCCA	cgtgtctctaTCATATTGACAGACCTTGAGCA	580	1.0 X
TLR8_11	MID8	ctcgcgtgtcCTGCCTACCACACCAGGTAA	ctcgcgtgtcGCTCGTCCTGTCAACTTCTG	598	1.0 X
TLR8_12	MID8	ctcgcgtgtcAAGAAAGTGAGGCACTCTGC	ctcgcgtgtcTCTCTTATTGGCATTTACCACA	559	1.0 X

TLR8_13	MID8	ctcgcgtgtcCCCGACACTTTGTTTGTTTT	ctcgcgtgtcTGAAATTAACAGTGCTTCACCA	552	1.0 X
TLR8_14	MID8	ctcgcgtgtcTCTTTTAGCACTGTGAAGCTGA	ctcgcgtgtcAGGAATGCCCCATCTGTAAT	561	1.0 X
TLR8_15	MID8	ctcgcgtgtcTCAAGGGCTGCAAAATCTTA	ctcgcgtgtcTACAGGGAAATGGTGCATTG	557	1.0 X
TLR8_16	MID8	ctcgcgtgtcCTGCCAAACTCCTTGAGAGA	ctcgcgtgtcATTGGGGAAATGTTGGAAAA	600	1.0 X
TLR8_17	MID8	ctcgcgtgtcCGTCTCTCCAGATATTGCACTT	ctcgcgtgtcTACCCCTGCTATTCGGAAAT	620	1.0 X
TLR8_18	MID8	ctcgcgtgtcCATGACATTGCCTGCTTAAA	ctcgcgtgtcAGGTCAAGCAAGTGGAGATG	569	1.0 X
TLR8_19	MID8	ctcgcgtgtcCCAGCATATCCCAGATGAAG	ctcgcgtgtcTAGGGCAGCCAACATAACTG	589	1.0 X
TLR8_20	MID8	ctcgcgtgtcATGTCATCTGTGCCAGTCCT	ctcgcgtgtcGCAAATACTGGGAATGCTGT	573	1.0 X
TLR8_21	MID8	ctcgcgtgtcTCATGCAGAGCATAAACCAA	ctcgcgtgtcGATCCTAACCCCAGGAGATG	576	1.0 X
TLR9_01	MID11	tgatacgtctGCACTGCCCCTAGTTCTAATC	tgatacgtctCTTCAGGGTTCAGTATCCTCAG	555	1.0 X
TLR9_02	MID11	tgatacgtctCTCTGGATCATCTCCCACTC	tgatacgtctGGCCATTTCTCTTTCCTTCT	570	1.0 X
TLR9_03	MID11	tgatacgtctAAAAGGAGAGGAAGGCTGGT	tgatacgtctGGACTTCAGGAACAGCCAGT	595	1.0 X
TLR9_04	MID11	tgatacgtctCACTCTCACCCAATCTCCAC	tgatacgtctCGCAGGGGTTCTTGTAGTAG	598	1.0 X
TLR9_05	MID11	tgatacgtctCTCGTGTCCCTGATCCTGA	tgatacgtctGAATGCCTTGGTTTTGGTG	586	1.0 X
TLR9_06	MID11	tgatacgtctGGCCTCGTGTTGAAGGATAG	tgatacgtctAGGTCCAAGGTGAAGCTGAG	575	1.0 X
TLR9_07	MID11	tgatacgtctCATCAGTGGAGCTGTGGAG	tgatacgtctCACATCTGGCTCAGGGAAT	564	1.0 X
TLR9_08	MID11	tgatacgtctTGGACCTCAGCTACAACAGC	tgatacgtctCTGCTAGGAAACCAAACCAG	600	1.0 X
TLR9_09	MID12	tgatacgtctCTAGACCTGTCCCAGAATCG	tgatacgtctACCAGCAATGAGAGACCAAA	578	1.0 X
TLR9_10	MID11	tgatacgtctCCTTTGTGGACTTCCTGCT	tgatacgtctCTCAGGATTACCAGCACCAC	604	1.0 X
TLR9_11	MID11	tgatacgtctCCTGAGCTATGATGCCTTTG	tgatacgtctATAGGCAGAGAGGCAAGGTC	560	1.0 X
TLR10_01	MID10	tctctatgcgGCTACCCAAAGGAGATGTGA	tctctatgcgACCACCACCCTGGATTTATC	553	1.0 X
TLR10_02	MID10	tctctatgcgAACCAGCAATAAAATCCTTGG	tctctatgcgTGTGTTTAAGATGGGCAAGC	598	1.0 X
TLR10_03	MID10	tctctatgcgTAGGTTTGAGTGGGGCAAA	tctctatgcgGACAAGTCGGGAACACCATA	600	1.0 X
TLR10_04	MID10	tctctatgcgTTCCACATCCAAAATGTGACT	tctctatgcgAGCTCTCGCAAAGACTTCAG	593	1.0 X
TLR10_05	MID10	tctctatgcgAAATGATGAAAATTGCTGGTG	tctctatgcgTAAGATACCAGGGCAGGTCA	587	1.0 X
TLR10_06	MID11	tctctatgcgTATTCAGAGGGCATGATGGT	tctctatgcgGCAGTAGAGTGGAATGGGTTC	597	1.0 X
TLR10_07	MID10	tctctatgcgAAGGCAACCCAAGGACAAC	tctctatgcgGAGAAATTGCAGACCCTTGA	573	1.0 X