Available online at www.sciencedirect.com
ScienceDirect
Preventive Veterinary Medicine 84 (2008) 121–134
PREVENTIVE VETERINARY MEDICINE
www.elsevier.com/locate/prevetmed

Assessing familial aggregation of paratuberculosis in beef cattle of unknown pedigree

Jason B. Osterstock [a,*], Geoffrey T. Fosgate [b], James N. Derr [c], Noah D. Cohen [a], Allen J. Roussel [a]

[a] *Department of Large Animal Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, United States*
[b] *Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, United States*
[c] *Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, United States*

Received 28 August 2007; received in revised form 26 November 2007; accepted 29 November 2007

Abstract

The objective of this study was to assess genetic similarity of beef cattle using microsatellite markers and to use this information to describe familial aggregation of paratuberculosis test results in Texas beef cattle. Paratuberculosis testing was performed on 2622 adult beef cattle using two commercially available serum ELISAs and radiometric fecal culture. Pedigree records were collected for registered purebred herds and herds with sufficiently detailed production records to identify parent-offspring pairs. Cases were defined as cattle with at least one positive paratuberculosis test result. Three controls were matched by herd of residence for each case. All parent-offspring pairs, cases, and controls were genotyped for 12 microsatellites. Bayesian analysis of allele frequency data was used to describe population substructure and assign individual cattle into groups of genetically similar cattle. The proportion of known parent-offspring pairs assigned to the same cluster was used to assess the validity of the approach to identify familial structure. Conditional logistic regression was used to describe the association between cluster assignment and paratuberculosis test-status matched by herd. Nine clusters of genetically similar individuals were identified and were supported by the proportion of parent-offspring pairs assigned to the same clusters. Increased odds of having at least one positive paratuberculosis test result were identified for two clusters compared to the cluster with the lowest proportion of positive paratuberculosis test results after conditioning on herd. The

* Corresponding author at: Department of Large Animal Clinical Sciences, Texas A&M University, 4475 TAMU, College Station, TX 77843-4475, United States. Tel.: +1 979 845 3541; fax: +1 979 847 8863.
E-mail address: josterstock@cvm.tamu.edu (J.B. Osterstock).

0167-5877/$ – see front matter
doi:10.1016/j.prevetmed.2007.11.008

results of this study demonstrate that population substructure can be used to describe familial aggregation of paratuberculosis test results in beef cattle of unknown pedigree.

Keywords: Beef-herd monitoring; Cattle-microbiological disease; ELISA; Genetics; Johne's disease; *Mycobacterium avium* subsp. *paratuberculosis*

1. Introduction

Paratuberculosis, commonly referred to as Johne's disease, is a chronic granulomatous enteritis of ruminants associated with *Mycobacterium avium* subsp. *paratuberculosis* (MAP) infection. The disease is associated with significant economic losses in the U.S. cattle industries (Stabel, 1998; Ott et al., 1999; Pence et al., 2003; Lombard et al., 2005; Losinger, 2005, 2006). Animals generally become infected during the first year of life, but do not develop clinical signs until later in life (Larsen et al., 1975). There is evidence to suggest that genetics play a role in disease resistance in dairy cattle populations. Heritability of paratuberculosis in dairy cattle has been estimated to be between 6 and 15% (Koets et al., 2000; Mortensen et al., 2004; Gonda et al., 2006), and associations between paratuberculosis status and polymorphism within bovine chromosome 20 have been described (Gonda et al., 2007). However, all of these studies were performed in dairy cattle for which pedigree data were available and some of these studies (Gonda et al., 2006, 2007) utilized a limited number of sires to identify associations between genotype and disease status.

Genetic influence on paratuberculosis status in beef cattle has not been well described. Evidence of increased risk of seropositivity in *Bos indicus* cattle breeds has been identified in two studies (Roussel et al., 2005; Elzo et al., 2006). We have shown that familial aggregation of paratuberculosis ELISA-status exists in Texas Longhorn cattle based on the association between ancestors and test-status of offspring and have identified a positive linear association between the antibody status of the dam and her offspring (Osterstock et al., 2007a).

Investigations of genetic contributions to infectious disease risk often begin with describing familial aggregation. Familial aggregation is generally defined as the increase in odds of disease among family members of affected individuals compared to those of the unaffected (Liang and Beaty, 2000). Describing this association can help to estimate the proportion of disease risk associated with genetic factors and begin to differentiate environmental and genetic components. Studies of genetics and infectious disease in cattle populations have generally been performed using either candidate gene approaches or animals of known pedigree. There are important limitations to these two approaches. Genetic resistance to infectious disease is likely to be polygenic with a few exceptions including resistance to colibacillosis in swine (Sellwood, 1979; Rapacz and Hasler-Rapacz, 1986). This suggests that candidate gene approaches are unlikely to sufficiently describe genetic contributions to disease risk. The efficiency of candidate gene studies may be limited if the genotypes of sampled individuals do not reflect the genetic distribution of target populations. Candidate gene studies are also subject to bias due to preconceived knowledge of pathophysiology of infectious diseases. For example, recognition of the role of cell-mediated immune responses in paratuberculosis in sheep motivated preliminary studies investigating candidate gene polymorphism and paratuberculosis status (Reddacliff et al., 2005). Significant associations between polymorphism in natural resistance associated protein 1 (formerly NRAMP1; presently referred to as SLC11a1) and major histocompatability complex

(MHC) loci were noted, but identification of these associations may be biased by selection of loci sampled rather than reflecting the importance of these loci in susceptibility. Family-based studies can help identify familial aggregation and begin to discriminate between genetic and environmental components of disease risk. However, populations of cattle suitable for family-based association studies are limited to registered purebreds and animals with extensive production records including herds that use artificial insemination or embryo transfer. It is likely that these subpopulations of cattle do not reflect the larger genetic pool of the commercial cattle industries, raising concern regarding selection bias in these studies and external validity. Additionally, differences in management of purebred animals or herds with extensive pedigree records may introduce selection bias if the management of these herds also influences risk of exposure to infectious agents. An alternative approach would be to describe distributions of disease in cattle by evaluating genetic similarity among cattle that are sampled based on disease risk rather than availability of pedigree records.

The objective of this case-control study was to compare the odds of being test-positive for paratuberculosis using antibody tests and bacteriologic culture of feces between groups of genetically similar beef cattle defined using Bayesian methods for describing population substructure. Additionally, we sought to validate this approach for describing familial aggregation by comparing the cluster results with known familial structure obtained from pedigree records.

2. Materials and methods

2.1. Animals and diagnostic tests

Two groups of beef cattle in central Texas were identified for paratuberculosis testing and genetic analysis. Texas Longhorn breeders within a 150 mile radius of College Station, TX were surveyed from a list obtained from 1 of 2 Texas Longhorn breed directories. The survey instrument was delivered by mail and solicited information regarding herd size, history of paratuberculosis, willingness to participate in the Texas Voluntary Johne's Disease Program, and willingness to participate in this research project. Of the herds willing to enroll in the state program and the research project, all herds with greater than 15 animals 2 years of age or older and all herds with a history of clinical paratuberculosis, animals with undifferentiated chronic diarrhea and weight loss, or paratuberculosis test-positive animals were selected for sampling. A second group of herds were identified for sampling based on diagnosis of paratuberculosis within the herd through admission of individual animals to the Texas Veterinary Medical Center or by referral from practicing veterinarians in Texas. These herds were selected independent of beef cattle breed or availability of pedigree records. Pedigree data, when available, was collected for all registered animals in both groups of herds from breed registry certificates. Pedigree data were used to identify parent-offspring pairs included among all cattle sampled.

Within all herds, all animals 2 years of age or older were selected for diagnostic testing. Fecal samples were collected from the rectum with a single-use, non-lubricated rectal sleeve. Blood samples were collected by coccygeal or jugular venipuncture. Whole blood was preserved on FTA® Classic cards (Whatman International, Newton Center, MA) for genotyping. The remaining blood sample was submitted for paratuberculosis testing using two commercially available ELISA test kits: ELISA-A (HerdChek®, IDEXX Laboratories Inc., Westbrook, ME) and ELISA-B (Parachek®, Prionics, Schlieren, Switzerland). Test results for ELISA-A were converted to S:P ratios by taking the difference between the sample optical density (OD) and the

mean of duplicate negative control ODs and dividing by the difference between the means of the positive and negative control ODs. Individual sample results were dichotomized into positive or negative using the S:P ratio cut-off of 0.25, as recommended by the manufacturer. Results from ELISA-B were classified as positive if the difference between the assay cut-off value and the sample OD was greater than 0. The cut-off value for this assay is determined by adding 0.1 to the mean OD of the duplicate negative controls on each run of the assay. Fecal samples were submitted for radiometric fecal culture in liquid medium as previously described (Collins et al., 1990). Briefly, the medium was supplemented with mycobactin J, egg yolk suspension, and antimicrobials. Fecal samples were decontaminated with 1.0% hexadecylpyridinium chloride and concentrated via filtration. The resulting filter membrane was placed into radiometric culture medium (BACTEC 12B medium, BD Diagnostic Systems, Franklin Lakes, NJ) and evaluated weekly for growth using an ionization detector (BACTEC 460, Johnston Laboratories, Towson, MD). A PCR assay for the IS900 gene insertion element was used to identify MAP when acid-fast organisms were cultured. Mycobacterial isolates negative for IS900 were classified as non-MAP *Mycobacterium* spp. and further characterization was not performed. ELISA-A and ELISA-B have been estimated to be 17–34% and 16–28% sensitive, respectively, and 95–97% and 96–99% specific, respectively, relative to fecal culture (Collins et al., 2005; Lombard et al., 2006; McKenna et al., 2005; Scott et al., 2007). Radiometric fecal culture has been reported to be 54% sensitive and is generally assumed to be 100% specific (Sockett et al., 1992). However, most test validity estimates are derived from studies in dairy cattle and sensitivity of both ELISAs and fecal culture may vary with stage of disease.

2.2. Genotyping

All parent-offspring pairs identified from breed registry certificates or production records were selected for genotyping regardless of paratuberculosis test-status for validation of the clustering method. Cases selected for genotyping were defined as all animals with at least one positive paratuberculosis test result; ELISA-A, ELISA-B, or fecal culture. For each case, three controls were matched by herd and randomly selected from those animals in the herd with ELISA-A S:P ratio $\leq$0.0. Parent-offspring pairs used to validate the clustering methods were not eligible for inclusion as cases or controls. DNA samples were obtained from whole blood stored on FTA® Classic cards collected during herd sampling.

A biopsy punch was used to harvest a 1.2 mm sample from blood samples stored on FTA® Classic cards for each animal selected for genotyping. Punches were placed in 200 μL wells in 96 well plates for processing. All plates included positive and negative control wells consisting of purified DNA from an animal of known genotype and PCR reagents only, respectively. Preparation of the punches was performed using a modification of previously described methods (Thacker et al., 1999). All samples were prepared for genotyping using three consecutive washes with 150 μL of FTA® wash (Whatman International, Newton Center, MA) with 15 min of incubation on a rocker panel at room temperature and 30 s of vortexing every 5 min during incubation for each wash. Samples were then rinsed using 150 μL of Tris EDTA (1 mM Tris, 0.1 mM EDTA, pH 7.4) with 15 min of incubation on a rocker panel at room temperature and 30 s of vortexing every 5 min during incubation. Punches were dried in a forced-air hood for 2.5–3 h.

A panel of microsatellites developed for parentage testing was used to genotype all parent-offspring pairs, cases, and controls (Schnabel et al., 2000). The panel consisted of two multiplexed PCR reactions with six microsatellites each (Table 1). Multiplex 1 conditions consisted of the following in a 5 μL reaction: 1.2 mm DNA punch, 0.375 units Taq polymerase

Table 1
Microsatellite loci, their respective fluorescent dyes, chromosomal position, and the number of alleles reported in cattle (Schnabel et al., 2000)

Marker	Label	Position	# Known alleles
Multiplex 1			
BM17132	FAM	19	11
BMS1862	VIC	24	13
BMS410	NED	12	12
BMS510	VIC	28	12
BMS527	FAM	1	12
RM372	VIC	8	8
Multiplex 2			
BM1225	NED	20	9
BM1706	FAM	16	10
BM1905	NED	23	11
BM2113	FAM	2	8
BM4440	NED	2	11
BM720	VIC	13	13

(Promega GoTaq®, Promega Corp., Madison, WI), 2.5 mM deoxynucleotide triphosphate (dNTP), 17.5 mM $MgCl_2$, 0.5 μL 10× MasterAmp (Epicentre Biotechnologies, Madison, WI), 1.5 μL 5× buffer (Promega 5× Colorless GoTaq® Flexi Buffer, Promega Corp., Madison, WI), and 0.4–1.35 μM reverse and fluorescently labeled forward primers. Multiplex 2 conditions differed only in primer concentrations; 0.4–0.9 μM reverse and fluorescently labeled forward primers. All PCR reactions were carried out in a GeneAmp PCR system 9700 thermocycler (Applied Biosystems, Foster City, CA). Thermocycler events included: 3 min 96 °C; 4 cycles of 20 s at 96 °C, 30 s at 58 °C, and 1.5 min at 65 °C; 1 min at 96 °C, 1 min at 54 °C, 20 min at 65 °C; maintained at 4 °C. One microliter of PCR product from each sample was treated with 10 μL deionized formamide and denatured for 4 min at 96 °C after addition 0.3 μL of internal size standard (MapMarker ROX, BioVentures, Murfreesboro, TN). The resulting products were genotyped on an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA). Genotype was determined using known alleles for cattle (Schnabel et al., 2000) and commercially available software (GeneMapper version 3.7, Applied Biosystems, Foster City, CA). Allele assignments were reviewed manually by a single investigator (JO) to confirm genotype and identify samples of insufficient quality that required repeated genotyping.

2.3. Data analysis

Cluster analysis was performed on all parent-offspring pairs, cases, and controls to define population substructure using commercially available software (Pritchard et al., 2000). Clusters were defined using a Bayesian clustering algorithm based on allele frequency. Cattle were assigned to their respective clusters probabilistically given their genotype and the allele distributions within clusters. The model for this analysis assigns a probability for the animal to have an individual allele given an assumed population of origin, the allele frequency in that population of origin, and the degree of admixture. The populations of origin and allele frequencies in the population of origin are unknown quantities and the degree of admixture can be estimated from the data after assigning the total number of clusters. The allele probabilities within each cluster were modeled assuming a Dirichlet distribution. This is a multivariate

multinomial distribution that allows the sum of allele probabilities for individual loci over all clusters and the sum of probabilities for all alleles at each locus within a cluster to equal 1. Markov chain Monte Carlo (MCMC) simulation techniques were used to sample from the probability of the population of origin for the allele, the allele frequency within that population, and the proportion of a given individual's genome that arose from that population given the genotypic data. The probability that an individual belonged to a given cluster could then be inferred based on the results of the sampling methods.

Clustering was performed using a burn-in of 20,000 iterations followed by a MCMC of 50,000 iterations. Analysis was performed for 1–22 clusters, corresponding to the number of herds, and replicated five times within each number of clusters (k) to assess stability of the model for the assigned cluster number. The post hoc estimate of model fit in this program ($L(K)$) is the average log likelihood of the data for all MCMC iterations within a given k less the variance of this mean. A regression line was plotted for replicates of models for each k using a fractional polynomial prediction of $L(K)$. Optimal number of clusters was determined by considering the maximum value of $L(K)$ for each k and by using the delta-k (Δk) statistic proposed by Evanno et al. (2005). This statistic is derived by measuring the second order rate of change in $L(K)$ between successive runs at increasing k. The optimal k is selected at the maximum absolute value of Δk. Each individual's probability of assignment to the clusters was calculated at the optimal k based on allele frequency of the respective clusters and the genotype of the individual.

Validity of the cluster assignment was assessed by evaluating the proportion of known parent-offspring pairs assigned to the same cluster at the optimal k and was compared to these proportions for other values of k using a two-sided test for homogeneity of proportions. A Chi-square test was performed to compare the number of parent-offspring pairs assigned to the same cluster with the number of pairs of individuals that would be assigned to the same cluster by chance assuming binomial sampling. Distribution of clusters within and among breeds and herds was also evaluated to ensure that analysis of population substructure discriminated between breeds and identified genetic differences beyond the level of herd where appropriate.

The odds of having at least one positive paratuberculosis test result among the two ELISAs and fecal culture were compared among clusters using conditional logistic regression of case and control samples conditioned on herd of residence. Conditional logistic regression was performed with commercially available software (Intercooled Stata version 9.2 for Windows, StataCorp. LP, College Station, TX) and a P-value <0.05 was used to assess statistical significance. Model fit was assessed using likelihood ratio tests to compare intercept-only models to those including a term for cluster.

3. Results

3.1. Sampling and diagnostic tests

Surveys were mailed to 762 Texas Longhorn breeders in central and coastal Texas. The total number of respondents was 147 (19.3%) from which 17 herds containing a total of 721 animals $\geq$2 years of age were selected for sampling. Of the 721 animals sampled, 715 had complete diagnostic test results including both ELISAs and fecal culture, 4 (0.6%) were fecal culture positive for MAP, 77 (10.7%) had other *Mycobacterium* spp. cultured from their feces, 36 (5.0%) were positive on ELISA-A, and 24 (3.4%) were positive on ELISA-B. Of those animals with negative fecal cultures, 24 (3.8%) were seropositive with ELISA-A and 12 (1.9%) were seropositive with ELISA-B.

Five additional beef cattle herds were sampled including 1901 animals ≥2 years of age. These included three crossbred herds of varying *B. indicus* × *B. taurus* proportions ($n = 747$), 1 herd with Brahman, Shorthorn, and crossbred cattle ($n = 381$, 44, and 523, respectively), and 1 herd with Angus cattle ($n = 206$). All herds had reported clinical cases of paratuberculosis within the previous 2 years. Within these herds, 1892 animals had complete diagnostic test results including 5 (0.3%) fecal culture positive for MAP, 39 (2.1%) fecal culture positive for other *Mycobacterium* spp., 59 (3.1%) positive on ELISA-A, and 62 (3.3%) positive on ELISA-B.

3.2. Genotyping and cluster analysis

Genotypes were established for 592 animals sampled for paratuberculosis testing. This included 299 animals in known parent-offspring pairs, 105 animals positive on at least one test for paratuberculosis, and 288 herd-matched, test-negative controls. Analysis of population substructure identified several potential optimal numbers of clusters based on the reflection of the regression line associated with a plot of $L(K)$ for replicates of the analysis over k assignments from 1 to 22 (Fig. 1) and the Δk statistic (Fig. 2). Cluster assignments for k equal to 9 was selected for analysis based on the proximity of the $L(K)$ regression curve to the transition to stability at that value of k and the similarity of Δk for k equal to 9 compared to 7 and 11 clusters.

Cluster assignment appeared to appropriately differentiate breeds and define genetic diversity beyond the level of herd (Tables 2 and 3). Of the known parent-offspring pairs that were genotyped, the use of nine clusters assigned both parent and offspring to the same cluster for 46 of 55 (83.4%) of sire-offspring pairs and 110 of 149 (73.8%) of dam-offspring pairs. The number of pairs assigned to the same cluster was significantly greater than would be expected by chance (7.3 and 19.7 for sire and dam-offspring pairs, respectively) for the number of individuals genotyped

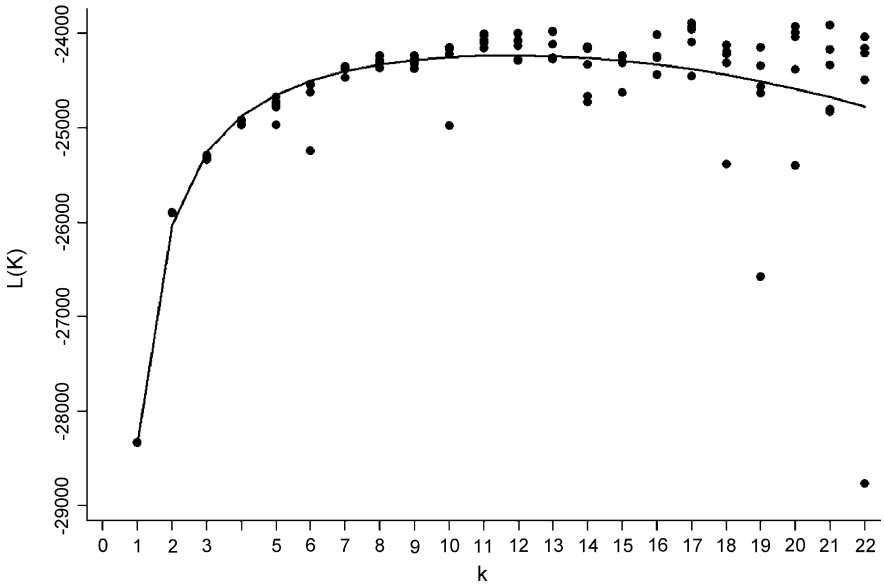

Fig. 1. Mean log likelihood less the variance of the mean ($L(K)$), number of assigned clusters (k), and regression line for replicates of a Bayesian model to assign individual beef cattle sampled for paratuberculosis testing to clusters of genetically similar individuals based on allele frequency for 12 microsatellite loci.

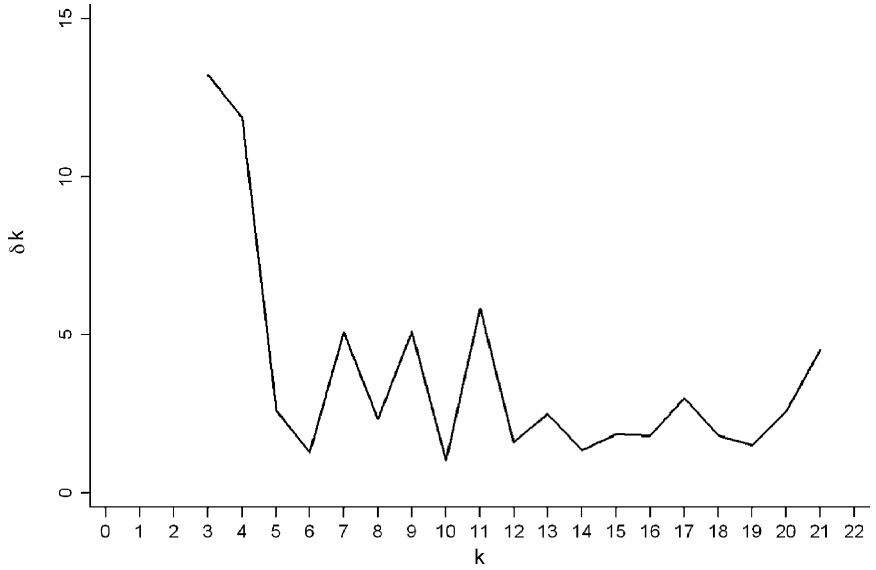

Fig. 2. Evanno et al.'s delta-k (Δk) (2005) over different numbers of assigned clusters (k) for identification of population substructure in beef cattle sampled for paratuberculosis testing. Larger values of Δk are indicative of improved model fit for corresponding levels of k.

and this number of clusters ($P < 0.001$). The proportion of parent-offspring pairs assigned to the same cluster appeared consistent for values of k equal to 4, 7, 9, and 11. Significant differences were not detected for the proportion of sire-offspring pairs or dam-offspring pairs assigned to the same cluster between any of the values of k evaluated.

3.3. Associations with paratuberculosis

Proportion of paratuberculosis test-positive cattle in each cluster assignment varied among the nine clusters (Table 4). For the conditional logistic regression models, the cluster with the lowest proportion of seropositive animals when excluding parent-offspring pairs was selected as the referent cluster. Inclusion of a term for cluster was associated with improved model fit (LR test P-value = 0.0004). Significant increases in odds of having at least one positive paratuberculosis test

Table 2
Distribution of breeds within clusters of genetically similar beef cattle defined on the basis of allele frequency for 12 microsatellites

Breed	Cluster									Total
	1	2	3	4	5	6	7	8	9	
Angus	4	0	0	2	3	0	0	0	0	9
Brahman	0	0	0	0	0	0	13	0	100	113
Longhorn	14	72	20	9	3	120	2	67	0	307
Shorthorn	0	0	2	0	10	0	0	0	0	12
Crossbred	25	0	10	38	30	1	37	2	8	151
Total	43	72	32	49	46	121	52	69	108	592

Table 3
Distribution of cluster assignments defined on the basis of allele frequency for 12 microsatellites within herds of residence

Herd	Breeds	Cluster									Paratuberculosis test-positive	Total
		1	2	3	4	5	6	7	8	9		
1	LH	0	9	1	1	0	19	0	9	0	1	39
2	LH	0	1	0	0	0	4	0	4	0	3	9
3	LH	0	0	0	0	0	7	0	6	0	2	13
4	LH	0	1	0	0	0	6	0	13	0	2	20
5	LH	0	2	0	0	0	2	0	0	0	0	4
6	LH	0	3	1	0	0	13	0	7	0	2	24
7	LH	5	0	9	1	1	9	1	5	0	9	31
8	LH	8	0	4	1	1	6	1	1	0	7	22
9	LH	1	13	3	4	0	2	0	0	0	4	23
10	LH	0	3	1	0	0	4	0	0	0	2	8
11	LH	0	1	0	0	0	9	0	2	0	0	12
12	BRAH, SH, XB	10	0	5	24	34	1	45	2	105	59	226
13	LH	0	0	0	0	0	15	0	2	0	1	17
14	LH	0	34	0	0	0	3	0	3	0	3	40
15	LH	0	0	0	2	0	4	0	3	0	3	9
16	LH	0	4	0	0	1	10	0	6	0	1	21
17	XB	2	0	0	1	0	0	0	0	0	1	3
18	XB	2	0	3	7	3	0	3	0	2	3	20
19	ANG	4	0	0	2	3	0	0	0	0	3	9
20	XB	11	0	4	6	3	0	2	0	1	7	27
21	LH	0	1	1	0	0	1	0	5	0	1	8
22	LH	0	0	0	0	0	6	0	1	0	1	7
Total		43	72	32	49	46	121	52	69	108	115	592

Breeds (LH: Texas Longhorn; BRAH: Brahman; SH: Shorthorn; XB: *Bos indicus* × *B. taurus* cross; ANG: Angus). Cattle were defined as paratuberculosis positive if samples were positive for MAP antibody on either commercial ELISA or microbial culture of feces for MAP.

Table 4
Numbers of paratuberculosis positive and negative beef cattle assigned to each cluster and proportion of positive cattle in each cluster for known parent-offspring pairs, cases, and controls

Paratuberculosis status	Cluster									Total
	1	2	3	4	5	6	7	8	9	
Parent-offspring pairs										
Negative	0	59	4	5	1	77	1	42	0	189
Positive	0	4	1	0	0	3	0	2	0	10
	0.00	0.06	0.20	0.00	0.00	0.04	0.00	0.05	0.00	0.05
Cases and controls										
Negative	37	3	19	36	38	28	32	21	74	288
Positive	6	6	8	8	7	13	19	4	34	105
	0.14	0.67	0.30	0.18	0.16	0.32	0.37	0.16	0.31	0.27

Cattle were defined as paratuberculosis positive if samples were positive for MAP antibody on either commercial ELISA or microbial culture of feces for MAP.

Table 5
Results of conditional logistic regression model comparing odds of being positive for MAP antibody on either commercial ELISA or microbial culture of feces for MAP matched by herd

Cluster	OR	95% CI		*P*-value
		Lower	Upper	
2	36.4	3.1	430.4	0.004
3	3.4	1.0	12.0	0.056
4	2.0	0.6	6.8	0.241
5	1.9	0.5	6.7	0.334
6	3.7	0.9	14.5	0.060
7	7.4	2.2	25.0	0.001
8	1.6	0.3	8.5	0.587
9	5.9	1.8	19.4	0.003

The referent group is the cluster with the lowest proportion of positive test results among the nine clusters excluding parent-offspring pairs.

result were identified for clusters 2 (OR 36.4; 95% CI 3.1–430.4), 7 (OR 7.4; 95% CI 2.2–25.0), and 9 (OR 5.9; 95% CI 1.8–19.4) compared to the cluster with the lowest proportion of positive test results (cluster 1) (Table 5). Cluster 7, containing a total of 49 animals, included 5 of the 9 animals fecal culture positive for MAP (1 Longhorn and 4 Brahman cattle). No other cluster contained more than 1 animal fecal culture positive for MAP.

4. Discussion

Identifying familial aggregation of disease is an important step in identifying genetic differences in susceptibility to infectious disease. Familial aggregation also helps to differentiate genetic and environmental influences on disease susceptibility and to identify populations with disparate disease risks for future study. Studies aimed at describing familial aggregation in cattle populations have traditionally relied upon animals of known pedigree. For some infectious diseases, the study of animals with pedigree information may introduce selection bias as other management factors associated with disease prevalence may also be associated with herds of pedigreed animals. Examples would include selection for dairy cattle from farms with detailed production records or for purebred beef cattle populations. These operations may also be more likely to purchase cattle from other operations, a risk factor for introduction of paratuberculosis into the herd (USDA, 2002). Alternatively, these herds may have more familiarity with paratuberculosis and may be more likely to have control programs in place.

This study identified increased odds of positive paratuberculosis results in some clusters of beef cattle defined on the basis of genetic similarity rather than pedigree data. Although insufficient numbers of animals with MAP positive fecal culture results were available for statistical analysis, 5 of 9 animals with positive fecal cultures were in the same cluster supporting the aggregation of paratuberculosis within this group of animals. The proportion of known parent-offspring pairs assigned to the same cluster indicated that performing cluster analysis using microsatellite data yields potentially similar results as would have been obtained from pedigree information while avoiding potential selection biases. The proportion of parent-offspring pairs assigned to the same cluster may have been improved by including additional microsatellites. Additionally, because the allele frequency clustering framework is dependent upon the allele frequencies in the entire population, some rare alleles may have had undue

influence on cluster assignment. Further, the probability that an individual belonged in a given cluster was similar for multiple clusters for some cattle genotyped in this study. This method does appear to be able to differentiate genetic similarity beyond the herd level and may be particularly useful in describing familial aggregation in cattle that are extensively traded among producers, as is observed in Texas Longhorn cattle, where genetic similarity may spread across herds.

The clustering approach employed here based on allele frequencies has been applied to studies of genetic differentiation in humans and chickens (Rosenberg et al., 2001, 2002). Results of these studies indicate that these methods more accurately reflect known population structure compared to genetic distance-based methods (Rosenberg et al., 2001). To our knowledge, this clustering framework has not been used to establish clusters of genetically similar individuals with the intent of comparing proportions of diseased animals between clusters. A potential pitfall to this approach is increase in the overall experiment-wise error by applying multiple statistical models in deriving the effect estimate. Additionally, this could introduce misclassification bias if the results of the clustering models yielded inaccurate assignment of individuals to cluster groups. Misclassification bias could also be introduced if the results of the genotyping included incorrectly assigned alleles.

The allele frequency-based cluster models used here, however, are not without limitation. Selection of the optimum value for k is somewhat subjective and model fit cannot be readily tested statistically between models for different levels of k. We used two criteria for selecting the optimal k and both methods yielded similar results. It is likely that some datasets evaluated using these methods would have multiple values for k that similarly fit with the data because certain clusters could be subdivided with increasing values of k. Another limitation of these models is lack of convergence for individual runs at a given k. MCMC sampling methods may not converge and may not sufficiently explore all potential modes causing inferences based on these models, specifically the proportion of an individual's genome that would be assigned to a given cluster, to be unstable. We attempted to reduce the likelihood of relying on spurious model results by replicating each model run five times for each value of k.

The clusters with significantly increased odds of positive paratuberculosis test results demonstrate some additional associations and features of this approach to describing familial aggregation. Cluster 2 was comprised exclusively of Texas Longhorn cattle. The cattle included in this cluster represent 10 different herds demonstrating the utility of this approach in identifying genetic structure across herds. This supports the dissemination of genetic factors associated with paratuberculosis test-status as environmental factors would be expected to play a smaller role in the odds of having a positive paratuberculosis test result with so many environmental exposures represented. Clusters 7 and 9 were comprised predominantly of Brahman and Brahman-cross cattle. Cluster 7 contained cattle from 5 herds including 2 Texas Longhorns, 13 Brahmans, and 37 crossbred animals with substantial proportions of *B. indicus* ancestry. Similarly, cluster 9 contained cattle from 3 herds including 100 Brahman cattle from a single herd and 8 animals from 3 herds with substantial *B. indicus* ancestry. The increased odds of having at least one positive paratuberculosis test result observed in these groups supports the findings in two other studies that have identified increased prevalence of paratuberculosis seropositivity in *B. indicus* breeds of cattle (Roussel et al., 2005; Elzo et al., 2006).

The results of this study should be interpreted with caution regarding impacts on paratuberculosis control programs. The majority of the animals classified as paratuberculosis test-positive had positive results for one or both serum ELISAs, but did not have MAP isolated from their feces using radiometric fecal culture methods. Additionally, 116 animals had non-

MAP *Mycobacterium* spp. isolated from their feces. Previous reports have found that environmental and experimental exposure to non-MAP *Mycobacterium* spp. is associated with false-positive serum ELISA results in Texas beef cattle (Osterstock et al., 2007b; Roussel et al., 2007). Therefore, the increased odds observed in some of the clusters in this study could reflect genetic differences associated with humoral responses to *Mycobacterium* spp. rather than infection with MAP. The influence of the findings of these two studies on the perceived increase in prevalence of paratuberculosis seropositivity in Brahman cattle is unclear, but may be associated with the geographical distributions in Texas of these cattle breeds and the *Mycobacterium* spp. associated with false-positive antibody responses.

Matching animals selected for genotyping by herd and removing a subset of animals (parent-offspring pairs) from the analysis introduces some limitations to this study. Animals with ELISA-A S:P ratios $\leq$0.0 were preferentially selected as controls to increase the likelihood of genotyping animals with both highly resistant and susceptible genotypes for positive paratuberculosis test results. The use of conditional logistic regression to control for the matching variable, in this case herd, creates data that becomes sparse once stratified. This causes imprecise estimates as evidenced by the wide confidence intervals for many of the effect estimates. However, animals within a herd may have similar genetic composition and environmental exposures prompting attempts to control for this potential confounder in the analysis. The effect of the selection bias imposed by this matching criterion would likely be bias toward the null, assuming individuals within a herd are more likely to be genetically similar. Larger sample sizes may help to eliminate the problem of sparse data after stratifying by herd; however, paratuberculosis prevalence in beef cattle is generally low (Thorne and Hardin, 1997; Dargatz et al., 2001; Hill et al., 2003; Pence et al., 2003; Roussel et al., 2005) and a consistent impediment to studies of the disease in these cattle. Removing parent-offspring pairs from the analysis contributed to this problem by reducing sample size since some cattle among these pairs were paratuberculosis test-positive. They were removed in this study to allow unbiased assessment of the validity of the method for describing familial aggregation, but this removal would likely be unnecessary in future studies using similar methods.

FTA cards have several advantages and disadvantages in studies of genetic association. The cards can be stored for long periods of time in room conditions with the DNA preserved on the filter. This allows DNA samples to be obtained during diagnostic test sample collection and stored pending test results and reduces the need for additional animal handling. This is particularly important in paratuberculosis research where delays in receiving diagnostic test results may be associated with degradation of DNA obtained and stored using other methods. Harvest of DNA from the FTA card filter is relatively easy and uses a very small piece of the filter, leaving a substantial portion of the sample available for future study. There are some challenges to using FTA cards to preserve DNA. Contamination of samples from a given animal by biological material from another animal may invalidate the sample. Drying and preservation of the sample on the filter may also be affected by environmental conditions, particularly high humidity. PCR conditions must be optimized for amplification of DNA preserved on FTA cards and amplification of large amounts of DNA (e.g. whole-genome) may not be possible from these samples.

The results of this study offer a novel approach to the assessment of familial aggregation of infectious disease in cattle populations that may be less subject to selection bias than study designs that rely on the identification of animals of known ancestry. We also have identified significant differences in the odds of having positive paratuberculosis test results among the clusters supporting a genetic basis for these differences in beef cattle.

5. Conclusions

Clustering of beef cattle based on microsatellite allele frequency data was used to define groups of genetically similar individuals and demonstrated that these methods can be used reliably in the absence of pedigree information to describe familial aggregation of infectious disease. Three clusters were identified that had significantly greater odds of having at least one positive paratuberculosis test result. One of these clusters (OR 36.4; 95% CI 3.1–430.4) was comprised exclusively of Texas Longhorn cattle. Two of these clusters (OR 7.4; 95% CI 2.2–25.0 and OR 5.9; 95% CI 1.8–19.4) contained large proportions of Brahman and *B. indicus* crossbred cattle. The results of this study demonstrate that familial aggregation of paratuberculosis test-status exists in beef cattle and support the hypothesis that *B. indicus* cattle breeds may be predisposed to MAP infection or seropositivity.

Acknowledgements

The authors wish to recognize the efforts of Dr. Natalie Halbert and Claire Kolenda for their technical assistance in this research project. This project was funded in part by USDA-APHIS-VS Award #05-9100-0995-CA and The Douglas A. Armstrong Memorial Scholarship, Academy of Veterinary Consultants and Bayer Animal Health, 2006.

References

Collins, M.T., Kenefick, K.B., Sockett, D.C., Lambrecht, R.S., McDonald, J., Jorgensen, J.B., 1990. Enhanced radiometric detection of *Mycobacterium paratuberculosis* by using filter-concentrated bovine fecal specimens. J. Clin. Microbiol. 28, 2514–2519.

Collins, M.T., Wells, S.J., Petrini, K.R., Collins, J.E., Schultz, R.D., Whitlock, R.H., 2005. Evaluation of five antibody detection tests for diagnosis of bovine paratuberculosis. Clin. Diagn. Lab. Immunol. 12, 685–692.

Dargatz, D.A., Byrum, B.A., Hennager, S.G., Barber, L.K., Kopral, C.A., Wagner, B.A., Wells, S.J., 2001. Prevalence of antibodies against *Mycobacterium avium* subsp. *paratuberculosis* among beef cow-calf herds. J. Am. Vet. Med. Assoc. 219, 497–501.

Elzo, M.A., Rae, D.O., Lanhart, S.E., Wasdin, J.G., Dixon, W.P., Jones, J.L., 2006. Factors associated with ELISA scores for paratuberculosis in an Angus-Brahman multibreed herd of beef cattle. J. Anim. Sci. 84, 41–48.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.

Gonda, M.G., Chang, Y.M., Shook, G.E., Collins, M.T., Kirkpatrick, B.W., 2006. Genetic variation of *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. J. Dairy Sci. 89, 1804–1812.

Gonda, M.G., Kirkpatrick, B.W., Shook, G.E., Collins, M.T., 2007. Identification of a QTL on BTA20 affecting susceptibility to *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. Anim. Genet. 38, 389–396.

Hill, B.B., West, M., Brock, K.V., 2003. An estimated prevalence of Johne's disease in a subpopulation of Alabama beef cattle. J. Vet. Diagn. Invest. 15, 21–25.

Koets, A.P., Adugna, G., Janss, L.L., van Weering, H.J., Kalis, C.H., Wentink, G.H., Rutten, V.P., Schukken, Y.H., 2000. Genetic variation of susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in dairy cattle. J. Dairy Sci. 83, 2702–2708.

Larsen, A.B., Merkal, R.S., Cutlip, R.C., 1975. Age of cattle as related to resistance to infection with *Mycobacterium paratuberculosis*. Am. J. Vet. Res. 36, 255–257.

Liang, K.Y., Beaty, T.H., 2000. Statistical designs for familial aggregation. Stat. Methods Med. Res. 9, 543–562.

Lombard, J.E., Garry, F.B., McCluskey, B.J., Wagner, B.A., 2005. Risk of removal and effects on milk production associated with paratuberculosis status in dairy cows. J. Am. Vet. Med. Assoc. 227, 1975–1981.

Lombard, J.E., Byrem, T.M., Wagner, B.A., McClusky, B.J., 2006. Comparison of milk and serum enzyme-linked immunosorbent assays for diagnosis of *Mycobacterium avium* subspecies *paratuberculosis* infection in dairy cattle. J. Vet. Diagn. Invest. 18, 448–458.

Losinger, W.C., 2005. Economic impact of reduced milk production associated with Johne's disease on dairy operations in the USA. J. Dairy Res. 72, 425–432.

Losinger, W.C., 2006. Economic impacts of reduced milk production associated with epidemiological risk factors for Johne's disease on dairy operations in the USA. J. Dairy Res. 73, 33–43.

McKenna, S.L., Keefe, G.P., Barkema, H.W., Sockett, D.C., 2005. Evaluation of three ELISAs for *Mycobacterium avium* subsp. *paratuberculosis* using tissue and fecal culture as comparison standards. Vet. Microbiol. 110, 105–111.

Mortensen, H., Nielsen, S.S., Berg, P., 2004. Genetic variation and heritability of the antibody response to *Mycobacterium avium* subspecies *paratuberculosis* in Danish Holstein cows. J. Dairy Sci. 87, 2108–2113.

Osterstock, J.B., Fosgate, G.T., Cohen, N.D., Derr, J.N., Manning, E.J.B., Collins, M.T., Roussel, A.J., 2007a. Familial associations with paratuberculosis ELISA results in Texas Longhorn cattle, Vet. Microbiol., doi:10.1016/j.vetmic.2007.10.027.

Osterstock, J.B., Fosgate, G.T., Norby, B., Manning, E.J., Collins, M.T., Roussel, A.J., 2007b. Contribution of environmental mycobacteria to false-positive serum ELISA results for paratuberculosis. J. Am. Vet. Med. Assoc. 230, 896–901.

Ott, S.L., Wells, S.J., Wagner, B.A., 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. Prev. Vet. Med. 40, 179–192.

Pence, M., Baldwin, C., Black III, C.C., 2003. The seroprevalence of Johne's disease in Georgia beef and dairy cull cattle. J. Vet. Diagn. Invest. 15, 475–477.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rapacz, J., Hasler-Rapacz, J., 1986. Polymorphism and inheritance of swine small intestinal receptors mediating adhesion of three serological variants of *Escherichia coli*-producing K88 pilus antigen. Anim. Genet. 17, 305–321.

Reddacliff, L.A., Beh, K., McGregor, H., Whittington, R.J., 2005. A preliminary study of possible genetic influences on the susceptibility of sheep to Johne's disease. Aust. Vet. J. 83, 435–441.

Rosenberg, N.A., Burke, T., Elo, K., Feldman, M.W., Freidlin, P.J., Groenen, M.A., Hillel, J., Maki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K., Weigend, S., 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159, 699–713.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W., 2002. Genetic structure of human populations. Science 298, 2381–2385.

Roussel, A.J., Libal, M.C., Whitlock, R.L., Hairgrove, T.B., Barling, K.S., Thompson, J.A., 2005. Prevalence of and risk factors for paratuberculosis in purebred beef cattle. J. Am. Vet. Med. Assoc. 226, 773–778.

Roussel, A.J., Fosgate, G.T., Manning, E.J., Collins, M.T., 2007. Association of fecal shedding of mycobacteria with high ELISA-determined seroprevalence for paratuberculosis in beef herds. J. Am. Vet. Med. Assoc. 230, 890–895.

Schnabel, R.D., Ward, T.J., Derr, J.N., 2000. Validation of 15 microsatellites for parentage testing in North American bison, Bison bison and domestic cattle. Anim. Genet. 31, 360–366.

Scott, H.M., Fosgate, G.T., Libal, M.C., Sneed, L.W., Erol, E., Angulo, A.B., Jordan, E.R., 2007. Field testing of an enhanced direct-fecal polymerase chain reaction procedure, bacterial culture of feces, and a serum enzyme-linked immunosorbent assay for detecting *Mycobacterium avium* subsp. *paratuberculosis* infection in adult dairy cattle. Am. J. Vet. Res. 68, 236–245.

Sellwood, R., 1979. *Escherichia coli* diarrhoea in pigs with or without the K88 receptor. Vet. Rec. 105, 228–230.

Sockett, D.C., Carr, D.J., Thomas, C.B., Collins, M.T., 1992. Evaluation of conventional and radiometric fecal culture and a commercial DNA probe for diagnosis of *Mycobacterium paratuberculosis* infections in cattle. Can. J. Vet. Res. 56, 148–153.

Stabel, J.R., 1998. Johne's disease: a hidden threat. J. Dairy Sci. 81, 283–288.

Thacker, C., Phillips, C.P., Syndercombe Court, D., 1999. Use of FTA cards in small volume PCR reactions. In: Sensabaugh, G.F., Lincoln, P.J., Olaisen, B. (Eds.), Proceedings of the 18th ISFH Congress. Elsevier Science B.V., pp. 473–475.

Thorne, J.G., Hardin, L.E., 1997. Estimated prevalence of paratuberculosis in Missouri, USA cattle. Prev. Vet. Med. 31, 51–57.

USDA, 2002. Uniform program standards for the voluntary bovine Johne's disease control program. US Department of Agriculture: Animal and Plant Health Inspections Service: Veterinary Services.