BMC Microbiology

Research article

Brucella 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs)

Betsy J Bricker*[1], Darla R Ewalt[2] and Shirley M Halling[1]

Address: [1]United States Department of Agriculture, Agricultural Research Service, National Animal Disease Center, 2300 Dayton Rd, Ames, IA, 50010, USA and [2]United States Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services, National Veterinary Services Laboratories, 1800 Dayton Rd, Ames, IA, 50010, USA

Email: Betsy J Bricker* - bbricker@nadc.ars.usda.gov; Darla R Ewalt - Darla.R.Ewalt@aphis.usda.gov; Shirley M Halling - shalling@nadc.ars.usda.gov

* Corresponding author

Published: 11 July 2003

BMC Microbiology 2003, **3**:15

Received: 11 April 2003
Accepted: 11 July 2003

This article is available from: http://www.biomedcentral.com/1471-2180/3/15

Abstract

Background: Currently, there are very few tools available for subtyping *Brucella* isolates for epidemiological trace-back. Subtyping is difficult because of the genetic homogeneity within the genus. Sequencing of the genomes from three *Brucella* species has facilitated the search for DNA sequence variability. Recently, hypervariability among short tandem repeat sequences has been exploited for strain-typing of several bacterial pathogens.

Results: An eight-base pair tandem repeat sequence was discovered in nine genomic loci of the *B. abortus* genome. Eight loci were hypervariable among the three *Brucella* species. A PCR-based method was developed to identify the number of repeat units (alleles) at each locus, generating strain-specific fingerprints. None of the loci exhibited species- or biovar-specific alleles. Sometimes, a species or biovar contained a specific allele at one or more loci, but the allele also occurred in other species or biovars. The technique successfully differentiated the type strains for all *Brucella* species and biovars, among unrelated *B. abortus* biovar 1 field isolates in cattle, and among *B. abortus* strains isolated from bison and elk. Isolates from the same herd or from short-term *in vitro* passage exhibited little or no variability in fingerprint pattern. Sometimes, isolates from an animal would have multiple alleles at a locus, possibly from mixed infections in enzootic areas, residual disease from incomplete depopulation of an infected herd or molecular evolution within the strain. Therefore, a mixed population or a pool of colonies from each animal and/or tissue was tested.

Conclusion: This paper describes a new method for fingerprinting *Brucella* isolates based on multi-locus characterization of a variable number, eight-base pair, tandem repeat. We have named this technique "HOOF-Prints" for Hypervariable Octameric Oligonucleotide Finger-Prints. The technique is highly discriminatory among *Brucella* species, among previously characterized *Brucella* strains, and among unrelated field isolates that could not be differentiated by classical methods. The method is rapid and the results are reproducible. HOOF-Printing will be most useful as a follow-up test after identification by established methods since we did not find species-specific or biovar-specific alleles. Nonetheless, this technology provides a significant advancement in brucellosis epidemiology, and consequently, will help to eliminate this disease worldwide.

Background

Brucellosis is a worldwide zoonotic disease caused by a number of host-adapted species of the gram-negative bacterial genus *Brucella*. In addition to the economic losses caused by reproductive failure in a number of important livestock animals, accidental transmission of the disease to humans can occur through animal husbandry, meat processing activities or ingestion of contaminated unpasteurized milk. The highest incidence of brucellosis is found in regions where local custom encourages the consumption of raw goat, camel or bovine milk, or of soft cheeses prepared from unpasteurized milk. Children in these regions are particularly vulnerable because of their elevated consumption of milk and dairy products.

Many countries have implemented eradication programs resulting in the reduction or elimination of the disease, but the disease remains enzootic in many regions of the world. In those countries where the disease has been eradicated or strictly controlled, continued surveillance is essential to preventing re-emergence of the disease. Once a new infection has been confirmed in a herd, it is critical to prevent further spread of the disease to other herds. It is equally important to determine by epidemiological trace-back analysis where the infection originated, how it was spread, and what measures are needed to prevent additional spread of the disease from this primary source.

Whenever possible, trace-back is confirmed by comparison of the outbreak strain with isolates obtained from the primary source. Identity is established by examining strain specific traits. In the case of *Brucella*, species are identified by the analysis of a large panel of traits composed of serology, growth requirements and biochemical phenotype [1]. These traits also permit additional subtyping of some species into biovars. Unfortunately, specific biovars tend to predominate in certain geographical areas. For example, in the USA, 85% of bovine infections involve *B. abortus* biovar 1.

A number of investigators have attempted to devise methods for genotyping *Brucella* strains. Published methods include enterobacterial repetitive intergenic consensus sequence-PCR (ERIC-PCR), and repetitive intergenic palindromic sequence-PCR (REP-PCR) [2,3]; random amplified polymorphic DNA-PCR (RAPD-PCR) or arbitrary primed-PCR (AP-PCR) [4,5]; and restriction fragment length polymorphism-PCR (RFLP-PCR) of the *omp*2 locus [6,7]. ERIC-PCR and RAPD-PCR are both affected by assay conditions and environmental effects during the amplification process [8–10]. Although the results are highly reproducible within a laboratory, laboratory-to-laboratory reproducibility has been problematic and thus makes the universal application of these methods unlikely (compare [2] and [3]). RFLP-PCR of the *omp*2 locus is not constrained by issues of reproducibility and has been useful for the differentiation of *Brucella* species [6] and for differentiation among isolates from marine mammal hosts [7]. However, as a tool for the epidemiology of brucellosis in livestock, this technique is limited by the low rate of natural sequence divergence within the locus at the biovar level.

Recently, the complete genomes from *B. melitensis* [11] and *B. suis* [12] were published and the genomic sequence from *B. abortus* is complete but not yet fully annotated [13]. These data have made it possible to search for regions of DNA sequence variability that might be useful as markers for diagnostic identification and differentiation. This paper describes the first application of microsatellite fingerprinting to *Brucella* isolates. Microsatellite fingerprinting exploits the occurrence of variable number tandem repeats (VNTRs) ranging from 1- to 10-bp per repeat, and 1 to >30 repeats per locus. These VNTRs have been observed throughout the prokaryotic and eukaryotic kingdoms [14,15]. The addition or deletion of repeat units within a locus is believed to be caused by slipped-strand nucleotide mispairing, and/or unequal crossover events which increase the rate of mutation at these loci [16]. The variations in the number of repeats give rise to molecular alleles for a given locus and aid in the characterization of the individual. When multiple loci are analyzed, the resulting fingerprint can be highly discriminating or even unique. Though commonly used for forensic identification in humans, the technique has recently been used for subtyping a number of pathogenic bacteria [17–22].

We have employed microsatellite fingerprinting for genotyping *Brucella* species based on the number of tandem repeats of the DNA sequence "AGGGCAGT" at eight loci in the genome. A PCR-based method was developed to identify the specific alleles at each of the repeat loci. The technique is compatible with high-throughput, automated, detection technologies, as well as traditional, less costly detection methods. We have named this technique "HOOF-Prints" for Hypervariable Octameric Oligonucleotide Finger-Prints. The type strains of all the classical *Brucella* species and biovars were surveyed to assess the discriminating power of this fingerprint technique. Examples from naturally infected cattle and wildlife herds were also examined to assess the level of divergence among and within outbreaks.

Results

Identification and characterization of the tandem repeat loci

Multiple tandem repeats of the octameric sequence "AGGGCAGT" were initially discovered when searching sequence files generated from the shotgun sequencing of

Table 1: Locations of the nine VNTR loci within the published genome sequences of *B. melitensis* and *B. suis*.

VNTR	*Brucella melitensis*				*Brucella suis*			
Locus	Acc. No[α]	Region	Location	Strand	Acc. No.	Region	Location	Strand
1	AE009669	Chrom II; Sect. 31	bp-4619	(-)	AE014592	Chrom II; Sect. 88	bp-2372	(+)
	NC-003318	Chrom II CDS	bp-322922		NC-004311	Chrom II CDS	bp-976069	
2	AE009582	Chrom I; Sect. 139	bp-6689	(+)	AE014357	Chrom I; Sect. 43	bp-6142	(-)
	NC-003317	Chrom I CDS	bp-1505161		NC-004310	Chrom I CDS	bp-477739	
3	AE009620	Chrom I; Sect. 177	bp-3018	(+)	AE014322	Chrom I; Sect. 8	bp-5823	(-)
	NC-003317	Chrom I CDS	bp-1915985		NC-004310	Chrom I CDS	bp-87085	
4	AE009669	Chrom II; Sect. 31	bp-10855	(-)	AE014591	Chrom II; Sect. 87	bp-6120	(+)
	NC-003318	Chrom II CDS	bp-329159		NC-004311	Chrom II CDS	bp-969847	
5	AE009622	Chrom I; Sect. 179	bp-5818	(+)	AE014320	Chrom I; Sect. 6	bp-4014	(-)
	NC-003317	Chrom I CDS	bp-1940493		NC-004310	Chrom I CDS	bp-62866	
6	AE009586	Chrom I; Sect. 143	bp-2632	(+)	AE014354	Chrom I; Sect. 40	bp-97	(-)
	NC-003317	Chrom I CDS	bp-1643997		NC-004310	Chrom I CDS	bp-438926	
7	AE009622	Chrom I; Sect. 179	bp-3883	(+)	AE014320	Chrom I; Sect. 6	bp-5972	(-)
	NC-003317	Chrom I CDS	bp-1938559		NC-004310	Chrom I CDS	bp-64824	
8	AE009499	Chrom I; Sect. 56	bp-166	(-)	AE014439	Chrom I; Sect. 125	bp-10292	(+)
	NC-003317	Chrom I CDS	bp-588659		NC-004310	Chrom I CDS	bp-1399544	
9	AE009575	Chrom I; Sect. 132	bp-11054	(+)	AE014364	Chrom I; Sect. 50	bp-108	(-)
	NC-003317	Chrom I CDS	bp-1434245		NC-004310	Chrom I CDS	bp-548878	

[α] Acc. No. is the GenBank accession number

Brucella abortus biovar 1, strain 9–941 for a subsequence (TAGGGC) of IS711. The B. abortus genome had nine chromosomal loci containing at least two complete copies of the octameric repeat. The repeat regions and approximately 200-bp of upstream and downstream flanking sequence were individually compared to the Brucella melitensis biovar 1 and Brucella suis biovar 1 published genome sequences. To facilitate comparisons among the three Brucella genomes, Table 1 indexes the coordinates of each VNTR locus in the B. melitensis and B. suis published genomes. To facilitate BLAST searches, each locus was mapped to both the complete chromosome sequence and to the matching chromosome section in each genome. The sequence location specified in Table 1 denotes the start of the first repeat unit at that locus. Since the published B. melitensis and B. suis genomes are numbered in opposite orientation, the relevant strand for each genome was also listed in Table 1.

The homologous sequences were aligned revealing variable numbers of repeats at eight of the nine loci (Fig. 1). Among the three sequenced *Brucella* genomes, the number of complete repeat units ranged from one copy (e.g. VNTR Locus-4 of *B. suis*) to fourteen copies (VNTR Locus-7 of *B. abortus*). A notable feature of the repeat loci was the significant conservation of the sequence immediately downstream of the repeat units in eight of the nine loci. Alignment of the downstream sequences of *B. abortus*, *B. melitensis* and *B. suis* (Fig. 2), identified a conserved sequence stretch of about 100-bp. These sequences had a higher than normal level of sequence divergence than is typically found among *Brucella* species, but a consensus sequence was evident (top line of Fig. 2). VNTR Locus-9 was the only locus that contained the same number of repeat units (3 complete units) in all three genomes. It was also the only locus lacking the conserved downstream sequence (data not shown).

The VNTR loci were found on both chromosomes of the published genomes and in most cases were distant from each other (Table 1). However, VNTR Locus-5 and VNTR Locus-7 were approximately 2-kb apart on the large chromosomes of the *B. abortus*, *B. melitensis* and *B. suis* sequenced genomes and were arranged in the same orientation. VNTR Locus-1 was about 6-kb away from VNTR Locus-4 on the small chromosomes of all three sequenced genomes. VNTR Locus-1 and VNTR Locus-4 were also arranged in the same orientation.

Development of the PCR assay

We wanted to examine the depth of variability of the tandem repeat loci among other *Brucella* species, and more importantly, whether variability also occurred among biovars and among independent isolates within a biovar. To rapidly characterize the VNTR loci from a bacterial isolate, eight parallel PCR assays were developed to enumerate the number of repeats at each locus. Amplification primers (Table 2) were designed to anneal to the sequences immediately upstream and downstream of the repeat units. The number of repeats at each locus were deduced from the size of the amplicon produced. The PCR templates consisted of methanol-fixed bacteria from a broad array of *Brucella* strains and isolates (listed in Table 3 and the Materials and Methods section).

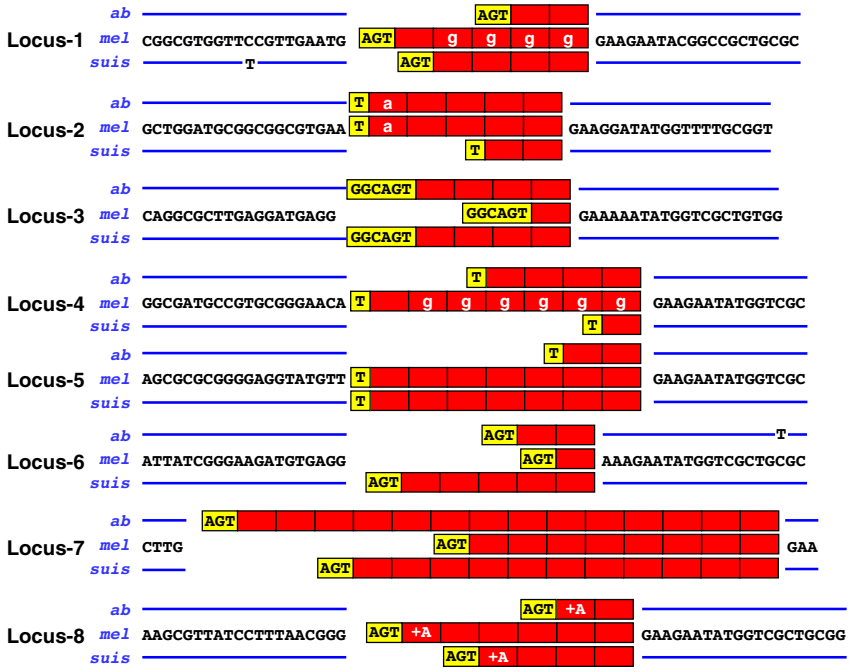

Figure 1

Repeat polymorphisms in VNTR loci aligned from the sequenced genomes of *B. abortus*, *B. melitensis* and *B. suis*. The sequences of the repeat region and flanking DNAs are shown for the eight variable loci. A red box designates a complete copy of the "AGGGCAGT" repeat; a red box containing a "g" designates the alternative repeat sequence "**g**GGGCAGT"; a red box containing an "a" designates the alternative repeat sequence "A**a**GGCAGT"; a red box containing a "+A" designates the alternative 9-bp repeat sequence "**a**AGGGCAGT"; and yellow boxes enclose partial repeat units. The blue lines indicate sequence identity. The individual sequences are labeled "*ab*" for *B. abortus* strain 9–941; "*mel*" for *B. melitensis* strain 16M; and "*suis*" for *B. suis* strain 1330.

Amplified products were sized by electrophoresis on 3% Metaphor or 4% agarose gels such that the 8-bp increase of each additional repeat unit was identified by comparison with a 25-bp mass ladder. Sample data of the amplicons generated from three loci of the *B. abortus* biovar type strains and two laboratory strains, are shown in Fig. 3A. Alternatively, the HOOF-Print procedure was adapted for high throughput, automated fluorescent DNA fragment analysis, by tagging each unique upstream primer with one of three fluorescent labels (Table 2). The amplicon sizes were determined by comparison with a co-migrating DNA mass ladder labeled with an alternative fluorescent tag. Data generated for VNTR Locus-1 by automated fluorescent DNA fragment analysis for the same isolates assayed in Fig. 3A, are shown in Fig. 3B. Detection of the fluorescent tagged amplicons was typically very sensitive with a high signal to noise ratio. Accurate sizing of the amplicons was essential for calculating the number of repeat units at each locus.

Identification of VNTR alleles

Alleles were defined by the number of tandem repeat units found at that specific locus. For simplicity, the allele names correspond to the estimated number of complete repeat units in the amplicon (e.g. Allele-1 has 1 complete repeat unit; Allele-6 has 6 complete repeat units). Some of

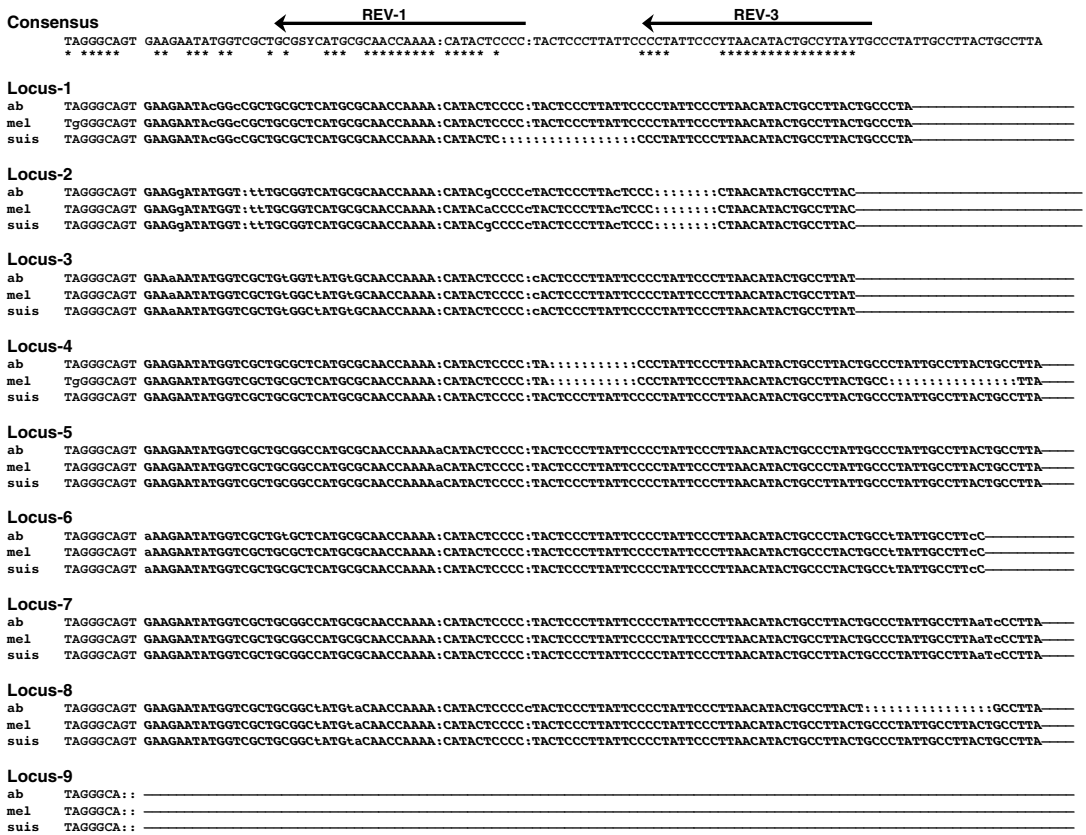

Figure 2

Alignment of DNA sequences downstream from the octameric repeat units. The flanking downstream sequences for each of the 9 VNTR loci from the three sequenced genomes were aligned by the Sequencher 4.0 software program. A consensus sequence, derived by alignment of VNTR Loci 1 through 8, is shown in the top line. The individual sequences are labeled "*ab*" for *B. abortus* strain 9–941; "*mel*" for *B. melitensis* strain 16M; and "*suis*" for *B. suis* strain 1330. Bold letters designate the last octameric repeat unit; lower case indicates divergence from the consensus sequence; colons represent missing nucleotides. An asterisk, "*", below the consensus sequence signifies identity for all *Brucella* species in all VNTR Loci 1 through 8. The two arrows above the consensus sequence mark the positions of the two reverse primers used for allele amplification. Y = C or T; R = A or G; S = G or C.

the VNTR loci were found to contain small deletions or insertions within the 3'-conserved sequence (Fig. 2). In VNTR Locus-1 and VNTR Locus-4, these modifications occurred within the amplified sequence. When the mutation affected a multiple of 8-bp, the amplicon still matched one of the hypothetical allele sizes (VNTR Locus-1 of *B. suis* biovar 1, Fig. 2). However, the published *B. suis* biovar 1 genome contained an 11-bp insertion within the amplified 3' sequence of Locus-4, when compared to the corresponding *B. abortus* and *B. melitensis* sequences (Fig. 2). This insertion resulted in an alternative ladder of hypothetical allele sizes that was offset by 11-bp.

In practice, the empirically determined product size was found to sometimes differ from the expected amplicon size by 1- or 2-bp, especially at higher allele numbers. In other cases, the observed amplicon size fell outside of the expected amplicon sizes. To determine the allele assignment, selected amplicons were examined by DNA sequencing (data not shown).

Comparison of alleles among the Brucella *Reference Strains for all species and biovars*

A panel of 19 pure *Brucella* cultures, composed of all of the *Brucella* species and biovar FAO/WHO Reference Strains,

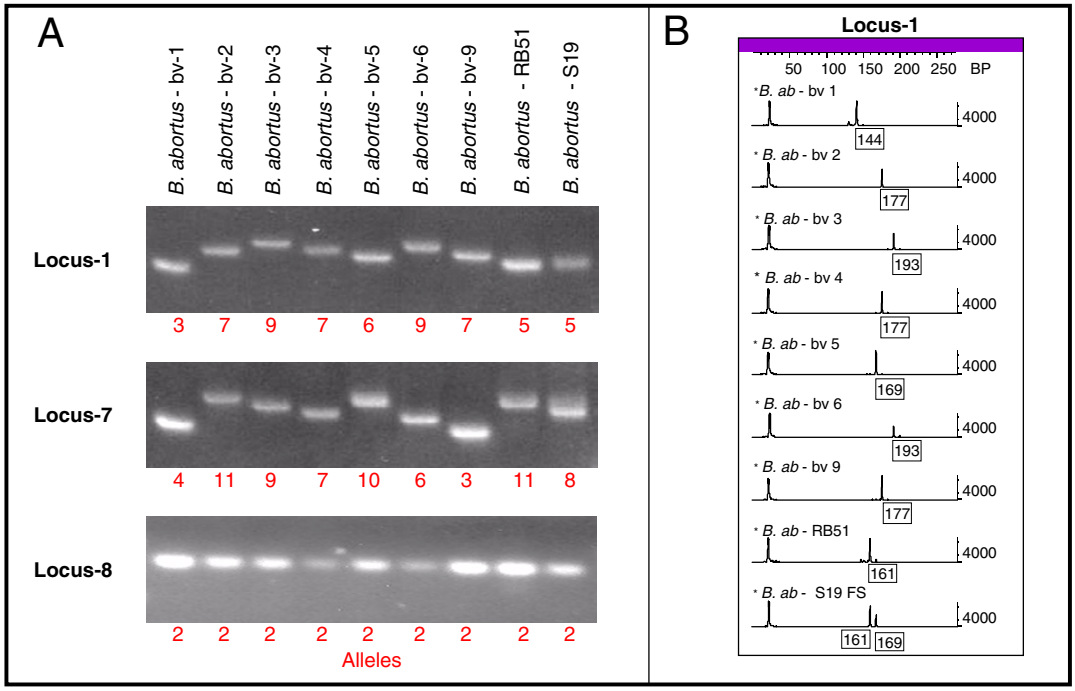

Figure 3

Analysis of allele variability by agarose gel electrophoresis and fluorescent tagged capillary electrophoresis. [A] Amplified fragments from three representative VNTR loci were resolved by electrophoresis in a horizontal gel composed of 3% Metaphor agarose. Sample identifications are given above their respective lanes. Calculated alleles (repeat units) are provided in red under each product. Product sizes range from 143–193 bp (Locus 1); 91–157 bp (Locus-7); and 86 bp (Locus-8). [B] Electropherograms generated from the capillary electrophoresis of fluorescently tagged products amplified from VNTR Locus-1 of the same strains assayed in Panel A. Product sizes (in bp) are shown in boxes below the major peaks. Sizes were calculated by the GeneScan software relative to the GeneScan 500-ROX size markers included in each sample as an internal standard. The vertical axis is the relative peak height of the detected fluorescent products. The first major peak in each electropherogram is the run-through of unincorporated primer.

was assayed by the HOOF-Print procedure. The amplicons obtained from each locus were sized and the numbers of repeat units were determined. The data were compiled in chart format for easy comparison (Fig. 4). The collective alleles amplified from a given isolate formed its genetic fingerprint. Each fingerprint was unique and each of the type strains could be easily differentiated. However, parts of the fingerprint patterns were similar among the related biovars of a single species. The VNTR Locus-8 locus, for example, contained two repeats in all *B. abortus* biovars. Similarly, the VNTR Locus-3 locus contained a single repeat in the three *B. melitensis* biovars.

In some cases, two or more products were amplified from a single locus [e.g.VNTR Locus-7 in *B. melitensis* biovar 1 (6 and 7 repeats) and biovar 2 (11 and 12 repeats), *B. canis* (10 and 11 repeats), and *B. suis* biovar 3 (5 and 6 repeats)]. These results were reproducible from cultures grown independently but originating from the same master seed. In these instances, the alleles would often, but not always, differ by a single 8-bp increment.

Table 2: PCR primer sequences

Primer name	5' to 3' sequence	Reverse Primer
LOCUS-1 Fwd	**HEX**·GGT GAT TGC CGC GTG GTT CCG TTG AAT GAG	REV-3
LOCUS-2 Fwd	**HEX**·CCC GCA TGA TCC GCG AAC AGC TGG ATG	REV-1
LOCUS-3 Fwd	**NED**·CAG GCG CTT GAG GAT GAG GCG GCA G	REV-3
LOCUS-4 Fwd	**6-FAM**·GCA GAA TTT TCG AGG CAT TCG GCG ATG	REV-3
LOCUS-5 Fwd	**6-FAM**·GTG CTC CAG GGC GCC GGG AGG TAT GTT TAG	REV-3
LOCUS-6 Fwd	**NED**·GCC GCA GGA AAG CAG GCG ATC TGG AGA TTA TC	REV-3
LOCUS-7 Fwd	**6-FAM**·CAG AGC CGT CGG TGG TTA CTT GAG TAG GGC AG	REV-1
LOCUS-8 Fwd	**NED**·GTG GGA AGC GTT ATC CTT TAA CGG GAG TAA GGG	REV-1
REV-1	GGG GAG TAT GTT TTG GTT GCG CAT GAC CGC	--
REV-3	GGG GGC ART ARG GCA GTA TGT TAA GGG AAT AGG G	--

HEX, NED and 6-FAM are fluorescent dyes covalently bound to the 5' end of the synthetic primer.

Table 3: *B. abortus* field strains used in this study.

Strain	Initial Isolation	Host Species
B. abortus fs 2018	Chino, CA	Bovine
B. abortus fs 2019	Chino, CA	Bovine
B. abortus fs 2045	Florida	Bovine
B. abortus fs 2046	Marathon, FL	Bovine
B. abortus fs 2047	Marathon, FL	Bovine
B. abortus fs 2146	Okeechobee, FL	Bovine
B. abortus fs 2147	Okeechobee, FL	Bovine
B. abortus fs 2148	Leesburg, FL	Bovine
B. abortus fs 2421	San Jacinto, CA	Bovine
B. abortus fs 2422	San Jacinto, CA	Bovine
B. abortus fs 2423	Arkansas	Bovine
B. abortus fs 2428	Huntsville, AL	Bovine
B. abortus fs 2429	Huntsville, AL	Bovine
B. abortus fs 2430	Huntsville, AL	Bovine
B. abortus fs 2431	Huntsville, AL	Bovine
B. abortus fs 2432	Weir, KS	Bovine
B. abortus fs 2385	Webster, FL	Bovine
B. abortus fs 8–1097	Madison Co. MT	Elk
B. abortus fs 7–1493	Laramie, WY	Elk
B. abortus fs 2–0993	Park Co., MT	Bison
B. abortus fs 5–1059	Park Co., MT	Bison
B. abortus fs 8–0999	Stanley Co., SD	Bison$^{\alpha}$

$^{\alpha}$ This bison was from a private herd.

As mentioned above, VNTR Locus-4 of the *B. suis* biovar 1 sequenced genome was 11-bp larger than the homologous regions of *B. abortus* biovar 1 and *B. melitensis* biovar 1. When the sizes of the VNTR Locus-4 fragments amplified from the panel of *Brucella* strains were compared to the size range of the normal and the alternative alleles, we observed that the products amplified from *B. canis*, *B. neotomae*, *B. ovis*, and *B. suis* biovars 2, 4 and 5, were consistent with the 11-bp offset allelic ladder of *B. suis* biovar 1 rather than the allelic ladder exhibited by the *B. melitensis* or *B. abortus* biovars (Fig. 4, starred alleles).

There was disagreement between the fragment sizes calculated from the published genomic sequences and the observed fragment sizes amplified from some of the VNTR loci in *B. melitensis* biovar 1 and *B. suis* biovar 1 (e.g. VNTR Locus-5). Furthermore, no amplification was detected for the VNTR Locus-1 or VNTR Locus-4 loci of *B. melitensis* biovar 1, even though the same type strains (strains 16M and 1330) were used for both studies.

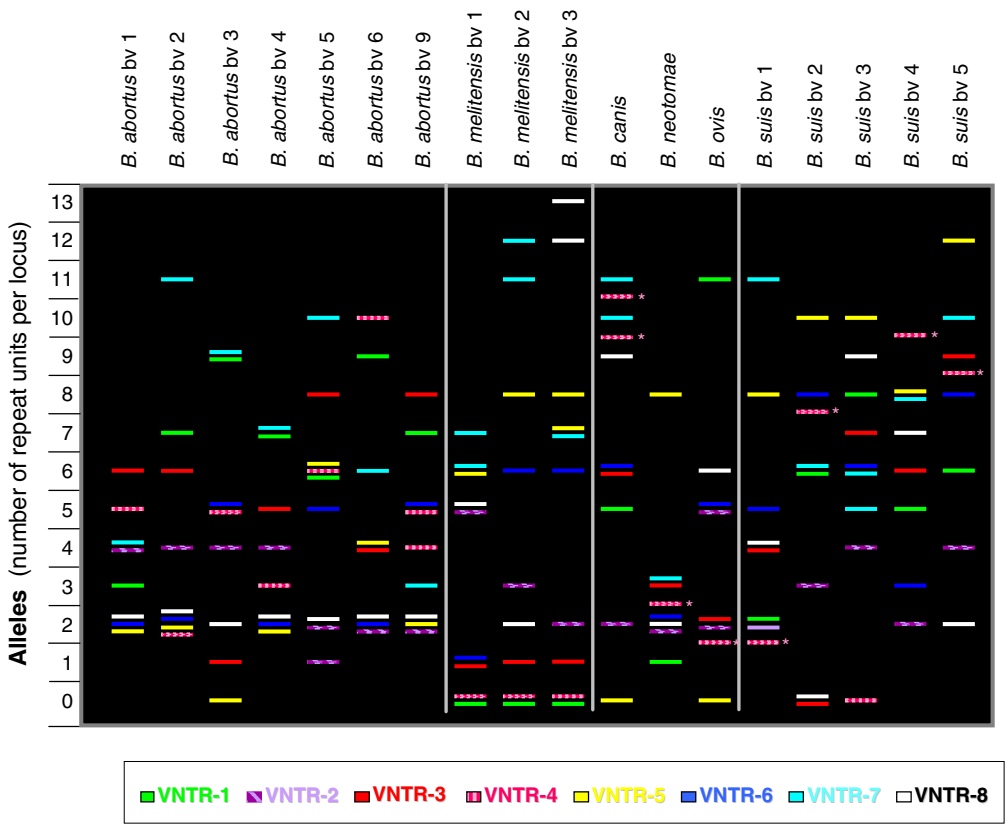

Figure 4

Multilocus allele analysis of *Brucella* species and biovar type strains. Graphical representation of the alleles generated by fluorescent tagged-PCR amplification of the genomic DNA of *Brucella* species and biovars. The allele number reflects the total number of repeats calculated for each locus. Each lane is the compilation of data from the independent amplification of each of the eight VNTR loci. The star next to the VNTR Locus-4 allele in some samples shows a match with the alternative allele ladder indicating an 11-bp insertion in the non-repeat region of the amplicon.

Comparison of alleles among selected field isolates of B. abortus

Most bovine brucellosis in the USA is caused by *B. abortus* biovar 1. For a typing method to be of value for epidemiological trace-back, it is essential to be able to subtype biovar 1 isolates. A small panel of *B. abortus* biovar 1-field strains, consisting of 17 bovine isolates, 2 elk isolates and 3 bison isolates, was tested to look at strain diversity in the U.S. Cattle field isolates were acquired from 10 herds in 5 states. The compiled fingerprints, along with the fingerprint of *B. abortus* reference strain 544, are shown in Fig. 5. The allele patterns of *Brucella* isolates cultured from multiple animals in the same herd are boxed together. The first observation was that while the fingerprint pattern within a herd appeared to be stable, the profile for each herd was unique. It was also apparent that the individual loci were mutating at different rates. For example, among the 17 cattle field isolates, Allele-2 was the only allele found at VNTR Locus-5 and Locus-8. Allele-2 was also found at Locus-6 in 16 of the 17 cattle isolates. At the same time, VNTR Locus-1 (7 alleles), Locus-3 (6 alleles), and Locus-7 (5 alleles), showed the highest variation. The fingerprint profiles of the wildlife isolates were also unique and easily distinguished from the cattle isolates. Nonetheless, these isolates shared the same alleles found for VNTR Locus-5 (Allele-2) and Locus-8 (also Allele-2) in cattle. The bison isolates shared common alleles at VNTR Locus-3 (Allele-3) and Locus-4 (Allele-11). However,

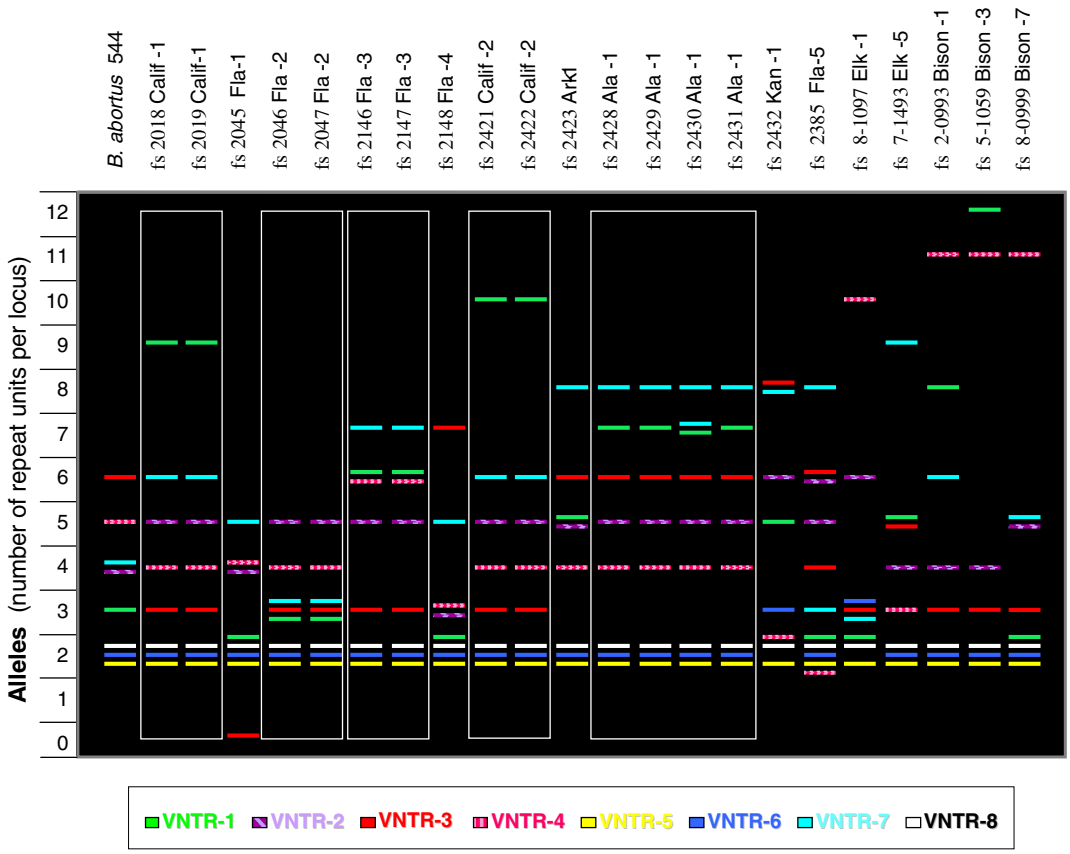

Figure 5
Multilocus allele analysis of *B. abortus* field strains. Graphical representation of the alleles generated by fluorescent tagged-PCR amplification of the genomic DNA of *Brucella abortus* biovar 1 field isolates. Each lane is the compilation of data from the independent amplification of each of the eight VNTR loci. Among cattle isolates, the samples are labeled with the state of origin; independent herds are indicated by an arbitrary number; and samples from the same herd are boxed. Each isolate was obtained from a single animal.

because of the very small sample size, it remains to be seen if this constitutes a true pattern.

Comparison of alleles among multiple production lots of commercially prepared B. abortus *vaccine strain RB51*

Since a significant degree of variation was found among laboratory strains and field isolates of Brucella, we wanted to look at the level of allele variation generated during limited in vitro passage of Brucella. For this purpose, we tested four non-consecutive production lots of commercially prepared RB51 vaccine. The fingerprints for these samples were also compared with the RB51 parental strain, 2308, with the B. abortus biovar 1 type strain, 544, and with the unrelated vaccine strain S19. The results, presented in tabular form (Table 4), demonstrated that limited in vitro passage of RB51 did not induce any allele variation, and that RB51 was distinct from the other B. abortus biovar 1 laboratory strains including the parental strain 2308.

Discussion

Historically, it has been very difficult to irrefutably trace natural *Brucella* transmission unless the strain involved happened to belong to a rare biovar. Furthermore, where total depopulation is not applied, it is difficult or impossible to differentiate a new exposure from a recrudescent infection in a recovered herd with the current tools available. Some techniques have been useful for the

Table 4: Comparison of fingerprints from multiple production lots of commercially prepared RB51 vaccine. Non-sequential production lots are compared with *B. abortus* biovar 1 laboratory strains including the RB51 parental strain, 2308.

Strain/Lot	Locus 1	Locus 2	Locus 3	Locus 4	Locus 5	Locus 6	Locus 7	Locus 8
544	3	4	6	5	2	2	4	2
S19[α]	5	4	4	2	2	2	8	2
2308[β]	4	4	6	2	2	2	10	2
RB51-1413	5	4	4	2	2	2	11	2
RB51-1474	5	4	4	2	2	2	11	2
RB51-1501	5	4	4	2	2	2	11	2
RB51-1555	5	4	4	2	2	2	11	2

[α] Strain 19 also amplified allele 6 in some cultures. [β] The fingerprint pattern of strain 2308 was derived from a single analysis instead of the usual duplicate analysis.

identification and/or differentiation of *Brucella* species and biovars [1,6,7,23,24], but few tools have been available for the epidemiological trace-back of new infections in livestock [1–5]. Except for biotyping, none of these more discriminating techniques has been universally accepted, probably because of concerns about laboratory-to-laboratory reproducibility. The HOOF-Print technique eliminates that concern because of the extreme specificity for the target loci.

The assay described in this study exploits hypervariable tandem repeats (VNTRs) composed of the octameric DNA sequence "AGGGCAGT". *Brucella* strains can be characterized by the number of repeats at each of eight chromosomal loci. Associated with the repeat loci is a conserved sequence, approximately 100-bp in length, downstream from the repeat region. The function of the conserved sequence is unknown; however, analysis of the secondary structure shows this region forms long stems with intermittent loop structures. The one repeat locus that lacks this sequence contains a stable segment consisting of three repeat units in *B. abortus*, *B. melitensis* and *B. suis*.

In other pathogenic bacteria, the hypermutability of VNTR loci has been exploited for strain typing [17–22]. To assess the level of variability associated with the *Brucella* VNTR loci, a set of eight parallel PCR assays were developed for amplification of the individual repeat regions. Strong conservation of the sequences upstream and downstream from the repeats allowed the design of very specific primers that generate extremely reproducible results. The primers anneal selectively to long sequences that are both unique and conserved. Because of the strength of the primer-target pairing and empirically optimized conditions, the reactions are robust and relatively tolerant of minor fluctuations and variations in the assay environment.

In anticipation of future multiplex development and high-throughput processing, each of the unique upstream amplification primers was tagged with one of three fluorescent compounds. The use of fluorescent dyes increased the detection level of amplified products substantially and may enhance the analysis of samples with very low levels of bacteria. However, some regions of the world experiencing high incidence of brucellosis may not have access to the expensive machinery and expertise necessary for fluorescent detection of amplicons. Therefore, the range of amplicon sizes selected for this method allows flexibility in the methods available for detection and differentiation of alleles, including less costly and more common technologies such as agarose gel electrophoresis.

The results show that the HOOF-Print assay is a highly discriminatory technique among *Brucella* species and biovars. The degree of hypervariability at the eight VNTR loci was surprising considering the profound genetic homogeneity of the *Brucella* genus that has made differentiation so difficult. Even more unexpected was the remarkable discrimination the HOOF-Print technique revealed among *B. abortus* biovar 1 field isolates. Nevertheless, isolates collected from multiple animals in the same herd had identical fingerprints. Examination of the fingerprint profiles obtained from non-consecutive production lots of vaccine strain RB51 supports that *Brucella* fingerprints evolve gradually. These characteristics make the HOOF-Print octameric repeats ideal for rapid strain typing.

Slip-strand mispairing is believed to be an important factor in generating tandem repeat polymorphisms [16]. Misalignment of DNA strands during chromosome replication is thought to lead to the addition or deletion of repeat units. Unequal crossover events are also believed to be involved in repeat expansion or contraction. The frequency of mutation events appears to be influenced by the number of repeats, the sequence flanking the repeats, and may involve multiple mechanisms [15]. Thus, VNTR-loci

typically mutate at differential rates. This differential rate of mutation at individual VNTR loci in *Brucella* is a fortuitous attribute for VNTR fingerprint analysis. It provides a glimpse into the evolution of the fingerprint profile and may provide some insight into interrelationships among strains. Our data supports this possibility. Certain loci mutate so slowly that particular alleles appear to become fixed in specific species and/or biovars of *Brucella* (e.g. Allele 2 for VNTR Locus-8 in all *Brucella abortus* isolates tested). Whether this will ultimately lead to species and/or biovar identification through fingerprint analysis will require the additional testing of a large number and wide variety of isolates. While some loci seem to mutate slowly, other loci appear to mutate frequently, enhancing the discriminatory power of technique. Currently, the individual rates of mutation and what factors influence those rates are unknown. Moreover, the affects of long term serial passage, both in culture and in animals, are also unknown. We plan to investigate these issues further.

During the isolation procedure for diagnostic analysis, multiple colonies from an individual tissue or animal are typically pooled for testing. We observed that in several instances pooled colonies from a single source contained multiple alleles at a single locus (for example, see the S19 electropherogram in Fig. 3B). At first this was perplexing, as it could indicate that certain loci may be too unstable to be valuable. However, further examination revealed that the multiple allele patterns were reproducible and isolate specific. This evokes several, non-exclusive explanations. In endemic areas, animals have been shown to carry multiple infectious strains simultaneously. In another scenario, incomplete depopulation of an infected herd could result in a dual infection during a subsequent outbreak. It is equally possible that the multiple alleles reflect a collection of accumulating subpopulations present in the host. In natural infections, animals typically receive a large dose of bacteria during exposure. These large doses could contain evolving subpopulations generated by the genetic instability within individual VNTR loci. Over time, it is predictable that the composite fingerprint of the resident population would gradually change. Regardless of the mechanism by which they come about, multiple alleles can occur. Therefore, during an outbreak or with a newly discovered infection involving *Brucella*, it will be essential to test multiple (or pooled) isolates and multiple animals to establish the complete representative fingerprint profile for epidemiological use.

Unexpectedly, different alleles were found at some loci than were predicted from the published genome sequences of *B. melitensis* and *B. suis*. This was despite using the same strains for our studies. It should be stressed, however, that little is known about the histories (i.e. number of passages, clonal selections, and culture conditions) of these strains between the time they were initially isolated from the host, disseminated to other laboratories, and prepared for these respective studies.

We anticipate that the testing of a larger number of isolates from various geographic areas will ultimately lead to the development of an international *Brucella* fingerprint database that can be searched for epidemiological use. A broad survey of the fingerprints derived from marine isolates will be an interesting complement to the data compiled from the classical terrestrial strains. The HOOF-Prints can be deposited in spreadsheet format or by numerical sequence of alleles (e.g. *B. abortus* strain 544: 3, 4, 6, 5, 2, 2, 4, 2) for rapid, easy, computerized identification or classification of isolates, independent of the testing laboratory involved. With a minimum amount of effort to standardize the process, a VNTR HOOF-Print database could become an invaluable resource for tracking and eliminating brucellosis outbreaks worldwide.

Conclusions

Currently, very few genetic markers have been identified for *Brucella* species that can be applied to epidemiological investigation of an outbreak. We have developed a new technique that has the power to definitively characterize and differentiate *Brucella* isolates based on multilocus VNTR fingerprinting. The assay is intended to complement the existing PCR and bacteriological diagnostic tests currently used for *Brucella* species and biovar identification. The existence of at least eight hypervariable loci exponentially increases the discriminative power of the assay for the characterization of closely related strains beyond the biovar level. Through the analysis of multiple loci, similarities as well as differences can be used to assess the genetic relatedness of field isolates. This, in turn, can provide insight into the evolution of a brucellosis outbreak and how to prevent further spread of the disease. Ultimately, we envision that this technique will be used to generate an international database of *Brucella* fingerprints that can be accessed and used by epidemiologists for control of the disease.

Methods

Bacteria used in this study

The bacterial strains used for this study included the complete array of *Brucella* type strains and/or FAO/WHO Reference Strains: *B. abortus* biovar 1, strain 544 (ATCC-23448); *B. abortus* biovar 1, strain S19 (isolates from stock culture and from infected cattle); *B. abortus* biovar 1, strain 2308; *B. abortus* biovar 1, strain RB51 (from the NADC Culture Collection and from commercially prepared vaccine production lots); *B. abortus* biovar 2, strain 86/8/59, (ATCC-23449); *B. abortus* biovar 3, strain Tulya, (ATCC-23450); *B. abortus* biovar 4, strain 292, (ATCC-23451); *B. abortus* biovar 5, strain B3196, (ATCC-23452); *B. abortus*

biovar 6, strain 870, (ATCC-23453); *B. abortus* biovar 9, strain C68, (ATCC-23455); *B. canis*, RM-6/66, strain (ATCC-23365); *B. melitensis* biovar 1, strain 16M, (ATCC-23456); *B. melitensis* biovar 2, strain 63/9, (ATCC-23457); *B. melitensis* biovar 3, strain Ether, (ATCC-23458); *B. neotomae*, strain 5K33, (ATCC-23459); *B. ovis*, strain 63/290, (ATCC-25840); *B. suis* biovar 1, strain 1330, (ATCC-23444); *B. suis* biovar 2, strain Thomsen, (ATCC-23445); *B. suis* biovar 3, strain 686, (ATCC-23446); *B. suis* biovar 4, strain 40, (ATCC-23447); *B. suis* biovar 5, strain 513, (ATCC-NA). The *B. abortus* field strains that were used to evaluate the technique and their origins are listed in Table 3.

Culturing and harvesting of bacteria for PCR analysis

Cultures were prepared from frozen aliquots (type strains) or from mixed colonies (field isolates) isolated on agar plates from tissue harvested from an infected animal. Cultures were streaked onto tryptose agar or trypticase-soy agar plates containing 5% serum, and incubated with 10% CO_2 at 37°C for 24–72 hrs [1]. Bacteria were harvested from the plate with saline (0.85 % sodium chloride) and were preserved by the addition of two volumes of 100% methanol. The preserved cells were stored at 4°C until needed. Immediately prior to PCR analysis, the methanol preserved cells were diluted in distilled water to a spectrophotometer OD_{600} reading of approximately 0.2 – 0.15 Absorbance units. The minimum dilution ratio was 1:10 to dilute the methanol concentration to ≤ 6.6%.

DNA accession numbers used in this study

The DNA sequences described in this report can be found in GenBank with the following accession numbers: Bru-VNTR-1 (clone 1) = AY099417; Bru-VNTR-2 (clone 3) = AY099419; Bru-VNTR-3 (clone 4) = AY099420; Bru-VNTR-4 (clone 5) = AY099421; Bru-VNTR-5 (clone 6) = AY099422; Bru-VNTR-6 (clone 7) = AY099423; Bru-VNTR-7 (clone 8) = AY099424; Bru-VNTR-8 (clone 9) = AY099425; Bru-VNTR-9 (clone 2) = AY099418; *B. melitensis* Chromosome I = NC 003317; *B. melitensis* Chromosome II = NC 003318; *B. suis* Chromosome I = NC 004310; *B. suis* Chromosome II = NC 004311.

PCR protocol

A set of eight PCR assays was developed to synthesize amplicons containing the tandem repeat sequences. Primer pairs (Table 2) were designed to hybridize within 50-bp upstream and downstream of each of the eight, hypervariable, repeat loci. Minimizing the sizes of the amplicons made the 8-bp size increments visually distinct on agarose gels (Fig. 3A) and clearly resolved by capillary electrophoresis (Fig. 3B). The forward primer was synthesized with one of three fluorescent dyes – HEX (green), NED (yellow), or 6-FAM (blue) (Table 2), covalently bound to the 5' end of the primer (Applied Biosystems, Foster City, CA). Because of the highly conserved nature of the downstream sequences, only two primers were needed for the downstream sequences of the VNTR loci. Several assay parameters were optimized for peak performance.

The PCR reaction mix consisted of 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.25 mM each of dNTPs, 2.5 mM $MgCl_2$, 0.2 μM each primer, and 0.6 units *Tsp* polymerase (Platinum GenoTYPE *Tsp* DNA Polymerase, Cat # 11448-032, Invitrogen, Carlsbad, CA). Cycling parameters were 94°C for 2.0 min; followed by 32 cycles of 94°C for 15 sec., 55°C for 20 sec., and 72°C for 1.5 min. Amplified products were stored at 4°C until needed. All amplifications were confirmed by gel electrophoresis (3% Metaphor agarose [Cat # 50181, BioWhittaker Molecular Applications, Rockland, ME] in 0.5X TBE [25]). Ethidium bromide-stained products that produced obvious, solitary bands after electrophoresis were processed for size analysis.

The amplified samples were diluted between 1:10 to 1:100 in water, depending on the estimated concentration of amplicon(s) in each sample. A 1-μl aliquot was sized on an ABI Prism 3100 Capillary Genetic Analyzer (Applied Biosystems, Foster City, CA) with virtual dye set D. GeneScan 500-ROX size marker (Cat # 401734; Applied Biosystems) was added to each sample to create a standard curve for each run. Each locus was amplified individually except for VNTR Locus-7. This locus gave better results when multiplexed with VNTR Locus-2 and VNTR Locus-8. Multiple runs of a single amplicon were found to be highly reproducible. Each fingerprint described was the result of at least two independent amplifications and analyses at each variable locus.

Based on the DNA sequence data, the expected amplicon sizes were calculated for the series of hypothetical products ranging from one to fifteen repeat units at each given locus. Each of these hypothetical products was considered a potential allele for the specific genetic locus.

DNA analysis programs

Comparative sequence searches were accomplished using "BLAST" (blast-n) [26] or "BLAST 2 sequences" [27]. Alignments were performed using Sequencher version 4.0 (Gene Codes Corp., Ann Arbor, MI) with manual editing. The raw capillary electrophoresis data produced on the ABI Prism 3100 Capillary Genetic Analyzer was analyzed using ABI Prism GeneScan and GenoTyper Analysis Software Version 3.7 (Applied Biosystems).

DNA sequencing of amplified products

For selected samples, the amplified DNA product from a single 100-μl reaction was purified with the QIAquick PCR Purification Kit (Cat. # 28104; Qiagen Inc., Valencia,

CA), from which 2-μl were processed with the ABI-PRISM® BigDye™ Terminators v3.0 Cycle Sequencing Kit (Cat. no. 4390242, Applied Biosystems) supplemented with a nested primer. After the unincorporated primers and nucleotides were removed, the sample was analyzed on an ABI-PRISM Model-377 DNA Sequencer.

List of abbreviations

ORF – Open Reading Frame; PCR – polymerase chain reaction; bp – base pair; kb – kilobase; ATCC – American Type Culture Collection; fs – field strain; VNTR – variable number tandem repeats.

Authors' Contributions

BB identified and characterized the tandem repeats, designed and developed the PCR, compiled and analyzed the data, defined the alleles, and performed the strain comparisons. BB also drafted the manuscript. DE isolated, cultured and processed the *Brucella* field isolates. She typed the isolates to species and biovar based on the classical phenotypic characteristics, and she edited the manuscript. SH suggested that we investigate microsatellite typing of *Brucella*, provided genomic sequence data for *B. abortus*, and edited the manuscript. All authors have read the manuscript and approved the final version.

Acknowledgements

We would like to thank Leslie Lincoln for her excellent technical assistance in this project.

References

1. Alton GG, Jones LM and Pietz DE: **Laboratory Techniques in Brucellosis** *Monograph Series Geneva, Switzerland: World Health Organization*; 1975.
2. Mercier E, Jumas-Bilak E, Allardet-Servent A, O'Callaghan D and Ramuz M: **Polymorphism in *Brucella* strains detected by studying distribution of two short repetitive DNA elements** *J Clin Microbiol* 1996, **34:**1299-1302.
3. Tcherneva E, Rijpens N, Naydensky C and Herman LM: **Repetitive element sequence based polymerase chain reaction for typing of *Brucella* strains** *Vet Microbiol* 1996, **51:**169-178.
4. Fekete A, Bantle JA, Halling SM and Stich RW: **Amplification fragment length polymorphism in *Brucella* strains by use of polymerase chain reaction with arbitrary primers** *J Bacteriol* 1992, **174:**7778-7783.
5. Tcherneva E, Rijpens N, Jersek B and Herman LM: **Differentiation of *Brucella* species by random amplified polymorphic DNA analysis** *J Appl Microbiol* 2000, **88:**69-80.
6. Cloeckaert A, Verger JM, Grayon M and Grepinet O: **Restriction site polymorphism of the genes encoding the major 25 kDa and 36 kDa outer-membrane proteins of *Brucella*** *Microbiol* 1995, **141:**2111-2121.
7. Cloeckaert A, Verger JM, Grayon M, Paquet JY, Garin-Bastuji B, Foster G and Godfroid J: **Classification of *Brucella* spp. isolated from marine mammals by DNA polymorphism at the *omp* 2 locus** *Microbes Infect* 2001, **3:**729-738.
8. Gao Z, Jackson KM and Leslie DE: **Pitfalls in the use of random amplified polymorphic DNA (RAPD) for fingerprinting of gram-negative organisms** *Pathology* 1996, **28:**173-177.
9. Grundmann HJ, Towner KJ, Dijkshoorn L, Gerner-Smidt P, Maher M, Seifert H and Vaneechoutte M: **Multicenter study using standardized protocols and reagents for evaluation of reproducibility of PCR-based fingerprinting of *Acinetobacter* spp** *Clin Microbiol* 1997, **35:**3071-3077.
10. Tyler KD, Wang G, Tyler SD and Johnson WM: **Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens** *J Clin Microbiol* 1997, **35:**339-346.
11. DelVecchio VG, Kapatral V, Redkar RJ, Patra G, Mujer C, Los T, Ivanova N, Anderson I, Bhattacharyya A, Lykidis A, Reznik G, Jablonski L, Larsen N, D'Souza M, Bernal A, Mazur M, Goltsman E, Selkov E, Elzer PH, Hagius S, O'Callaghan D, Letesson JJ, Haselkorn R, Kyrpides N and Overbeek R: **The genome sequence of the facultative intracellular pathogen *Brucella melitensis*** *Proc Natl Acad Sci USA* 2002, **99:**443-448.
12. Paulsen IT, Seshadri R, Nelson KE, Eisen JA, Heidelberg JF, Read TD, Dodson RJ, Umayam L, Brinkac LM, Beanan MJ, Daugherty SC, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Nelson WC, Ayodeji B, Kraul M, Shetty J, Malek J, Van Aken SE, Riedmuller S, Tettelin H, Gill SR, White O, Salzberg SL, Hoover DL, Lindler LE, Halling SM, Boyle SM and Fraser CM: **The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts** *Proc Natl Acad Sci USA* 2002, **99:**13148-13153.
13. **Microbial Genome Project – University of Minnesota: *Brucella abortus*** [http://www.cbc.umn.edu/ResearchProjects/AGAC/Pub_Brucella/Brucellahome.html]
14. Bennett P: **Demystified ... microsatellites** *Mol Pathol* 2000, **53:**177-183.
15. van Belkum A, van Leeuwen W, Scherer S and Verbrugh H: **Occurrence and structure-function relationship of pentameric short sequence repeats in microbial genomes** *Res Microbiol* 1999, **150:**617-626.
16. Levinson G and Gutman GA: **Slipped-strand mispairing: A major mechanism for DNA sequence evolution** *Mol Biol Evol* 1987, **43:**203-221.
17. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ and Hugh-Jones ME: **Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.*** *J Bacteriol* 2000, **182:**2928-2936.
18. Farlow J, Smith KL, Wong J, Abrams M, Lytle M and Keim P: ***Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis** *J Clin Microbiol* 2001, **39:**3186-3192.
19. van Belkum A, Scherer S, van Leeuwen W, Willemse D, van Alphen L and Verbrugh H: **Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*** *Infect Immun* 1997, **165:**5017-5027.
20. Smittipat N and Palittapongarnpim P: **Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*** *Tuber Lung Dis* 2000, **80:**69-74.
21. Coletta-Filho HD, Takita MA, de Souza AA, Aguilar-Vildoso CI and Machado MA: **Differentiation of strains of *Xylella fastidiosa* by a variable number of tandem repeat analysis** *Appl Environ Microbiol* 2001, **67:**4091-4095.
22. Klevytska AM, Price LB, Schupp JM, Worsham PL, Wong J and Keim P: **Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome** *J Clin Microbiol* 2001, **39:**3179-3185.
23. Bricker BJ and Halling SM: **Differentiation of *Brucella abortus* bv. 1, 2, and 4, *Brucella melitensis*, *Brucella ovis*, and *Brucella suis* bv. 1 by PCR** *J Clin Microbiol* 1994, **32:**2660-2666.
24. Vizcaino N, Verger JM, Grayon M, Zygmunt MS and Cloeckaert A: **DNA polymorphism at the omp-31 locus of *Brucella* spp.: evidence for a large deletion in *Brucella abortus*, and other species-specific markers** *Microbiology* 1997, **143:**2913-21.
25. Sambrook J, Fritsch EF and Maniatis T: *Molecular Cloning: a Laboratory Manual* 2nd edition. *Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press*; 1989.
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs** *Nucleic Acids Res* 1997, **25:**3389-3402.
27. Tatusova TA and Madden TL: **Blast 2 sequences – a tool for comparing protein and nucleotide sequences** *FEMS Microbiol Lett* 1999, **174:**242-250.