Infection, Genetics and Evolution 14 (2013) 214–222

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Multi-locus typing scheme for *Babesia bovis* and *Babesia bigemina* reveals high levels of genetic variability in strains from Northern Argentina

Eliana Guillemi [2], Paula Ruybal [1], Verónica Lia [1], Sergio González, Marisa Farber [1], Silvina E. Wilkowsky [1,*]

Instituto de Biotecnología, Centro de Investigaciones en Ciencias Veterinarias y Agronómicas, INTA Castelar, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 6 August 2012
Received in revised form 7 December 2012
Accepted 9 December 2012
Available online 31 December 2012

Keywords:
Babesia bovis
Babesia bigemina
MLST
Argentina
Phylogeny
Genetic diversity

ABSTRACT

Bovine babesiosis, caused by the protozoa *Babesia bovis* and *Babesia bigemina*, is a tick-borne disease distributed in tropical regions worldwide. Current control measures are based on the use of acaricides and live attenuated vaccines. The major economic impact of babesiosis lies in the cattle industry.

In order to gain insight into the extent of genetic diversity in populations of parasites in the field, we developed two MLST schemes for the molecular genotyping of *B. bigemina* and *B. bovis*. We have also developed a custom-designed bioinformatic pipeline to facilitate the automated processing of raw sequences and further diversity and phylogenetic analysis.

The overall MLST scheme exhibited the maximum discriminatory power (Simpson Index = 1) for *B. bovis* and a high level of discrimination for *B. bigemina* (Simpson Index = 0.9545). Genetic diversity was very high and infections with multiple genotypes were frequently found for both parasites in outbreak samples from the Northeast and Northwest of Argentina. Recombination events, which could have arisen from these multiple infections, were suggested by intra-*loci* linkage disequilibrium analysis and the lack of congruence in phylogenetic trees from individual genes.

The two MLST schemes developed here are a robust, objective and easily adoptable technology to analyze the genetic diversity and population structure of parasites of the genus *Babesia*.

1. Introduction

Bovine babesiosis is caused by the intraerythrocytic protozoan parasites *Babesia bovis* and *Babesia bigemina*. Both parasites are tick-transmitted and distributed in tropical regions worldwide. The major economic impact of babesiosis lies in the cattle industry. In Argentina, this disease is responsible for economic losses of 38.9 million dollars per year (Späth et al., 1994). Costs due to babesiosis are incurred not only from mortality, ill-thrift, abortions, loss of milk and meat production but also for draft power and control measures such as acaricide treatments, purchase of vaccines and therapeutics (Bock et al., 2004).

Live vaccines using attenuated strains are actually the only method to induce long-lasting immunity against babesiosis in cattle.

These vaccines consist of attenuated strains originated after serial passages of pathogenic strains in splenectomised calves (Shkap et al., 2007). The mechanism by which attenuation occurs is not fully understood, but current hypothesis support the selective enrichment of less virulent parasite subpopulations (Shkap et al., 2007) which may have distinct genetic and phenotypic characteristics.

Several methods and markers were developed to characterize the genetic diversity in both *Babesia* species (Hilpertshauser et al., 2007; Wilkowsky et al., 2008; Genis et al., 2009; Wilkowsky et al., 2009; Perez-Llaneza et al., 2010; Simuunza et al., 2011). However these methods are difficult to standardize or provide limited information on the phylogenetic relationships among strains. These limitations demonstrate the need of a more general approach to provide accurate, portable data that are appropriate for the epidemiological investigation of parasite strains and which also reflect their evolutionary and population biology. These studies will allow not only to understand the degree to which parasite populations could be modified by vaccination but also to determine the extent of genetic diversity in populations of parasites in the field. In order to gain insight into these subjects a new panel of polymorphic molecular markers distributed across the genomes of these parasites is essential. When identifying molecular markers for their use in population studies it is important to choose multiple *loci* that are selectively neutral. *Loci* under positive selection

* Corresponding author. Tel.: +54 11 46211447; fax: +54 11 4621 0199.
E-mail addresses: eguillemi@cnia.inta.gov.ar (E. Guillemi), pruybal@cnia.inta.gov.ar (P. Ruybal), vlia@cnia.inta.gov.ar (V. Lia), sgonzalez@cnia.inta.gov.ar (S. González), mfarber@cnia.inta.gov.ar (M. Farber), swilkowsky@cnia.inta.gov.ar (S.E. Wilkowsky).
[1] Researchers of CONICET (National Research Council).
[2] Doctoral fellow of CONICET.

1567-1348/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.meegid.2012.12.005

may give a distorted view of population structure and transmission dynamics, since selection rather than population history may determine the patterns of distribution of alleles within populations for these *loci*.

Multilocus sequence typing (MLST) was proposed in 1998 as a portable, universal, and definitive method for characterizing bacteria, being *Neisseria meningitidis* the species for which the first MLST scheme was developed. In addition to providing a standardized approach to data collection, by examining the nucleotide sequences of multiple *loci* encoding housekeeping genes or fragments of them, the information is compared at the DNA-sequence level. Besides, MLST data are made freely available over the Internet to ensure that a uniform nomenclature is readily available to all those interested in categorizing any given microorganism (Maiden, 2006).

We describe here the development of MLST schemes for *B. bigemina* and *B. bovis*, respectively, using sequences from 6 to 7 genes to type a set of reference and field strains of different geographic origin and phenotypic characteristics. The design of the MLST scheme was assisted by the availability of the completed *B. bovis* genome and the partially completed genome of *B. bigemina*. We have also developed a bioinformatic pipeline for the automated analysis of raw sequences and further diversity and phylogenetic analysis.

2. Materials and methods

2.1. Strains and genomic DNA isolation

Details of the strains and isolates used are described in Tables 1 and 2.

The highly virulent, tick-transmissible T2Bo (Texas) isolate is the published *B. bovis* reference genome. The *B. bigemina* Sanger reference strain is a virulent Australian isolate used for the genome project. *B. bovis* attenuated strain R1A (vaccine), virulent strain S2P and *B. bigemina* attenuated strains S1A (vaccine), S2A and virulent strains Mexico, S2P and S3P were *in vitro* cultured using a microaerophilus stationary phase system (Levi and Ristic, 1980). The *B. bovis* M2P, M3P and S7P virulent isolates, M1A attenuated strain and the *B. bigemina* B38, M1A (attenuated vaccine strain), M2P and M1P virulent strains, were amplified in splenectomized calves.

Blood samples from acute cases of the north region of Argentina (*B. bovis*: 35, 394, Perugorria, Sauce and Tomasito and *B. bigemina*: M30) were obtained from bovines that showed clinical signs of acute infection and were positive for *B. bovis or B. bigemina* by microscopic observation of Giemsa-stained blood smears. *Rhipicephalus microplus* adult ticks that were naturally attached to the bovine that undergone acute infection with *B. bigemina* M30 were removed and incubated at 28° C for oviposition. Resulting egg masses from individual replete females were placed individually into sterile containers (one egg mass per container) under the same incubation conditions and allowed to hatch. Subsequent groups of larvae representing progeny from a single female were kept separate from each other and incubated for 3 weeks at 28° C and 92.5% humidity. One gram of larvae (equivalent to approximately 20,000 organisms) were placed in skin patches over a calf with intact spleen (B32) and determined to be free of *B. bigemina* infection by enzyme-linked immunosorbent assay. When parasites could be detected in B32 calf by Giemsa-stained blood smears, blood was obtained by jugular venipuncture and used to inoculate splenectomized calf B38. When this calf reached the peak of parasitemia (determined by serial Giemsa-stained blood smears), jugular blood was obtained and processed for genomic DNA extraction. Genomic DNA from strains Uruguay and Brasil was kindly provided by Dr. María A. Solari from DILAVE Uruguay and Dr. João Ricardo Souza Martins, Instituto de Pesquisas Veterinárias Desidério Finamor/FEPAGRO, Brazil. The Mexico strain was provided by Dr. Carlos A. Vega y Murguía, CENID-PAVET, INIFAP, Mexico.

Genomic DNA isolation from reference strains from experimentally inoculated bovines or erythrocyte culture was performed from PBS-washed and packed infected erythrocytes. DNA was extracted by SDS lysis/proteinase K digestion, followed by phenol/chloroform extractions, and a standard ethanol precipitation (Sambrook et al., 1989). DNA extraction from field samples was performed by the method previously described by Higuchi (1989) which consists in two lysis steps and an overnight incubation with proteinase K.

All samples were analyzed by the reverse line blot hybridization assay using *Babesia/Theileria* catch-all probes and *B. bovis* or *B. bigemina* specific probes (Petrigh et al., 2008).

Table 1
B. bovis strains used in this study.

Isolate	Geographical location	Phenotype	ST
T2B	USA	Pathogenic	1
M1A	Corrientes, NE Argentina	Attenuated	2
M2P	Corrientes, NE Argentina	Pathogenic	3
S2P	Salta, NW Argentina	Pathogenic	4
M3P	Corrientes, NE Argentina	Pathogenic	5
R1A	Santa Fé, Argentina	Attenuated	6
35	Salta, NW Argentina	Pathogenic	7
394	Salta, NW Argentina	Pathogenic	8
Perugorria	Corrientes, NE Argentina	Pathogenic	9
S7P	Santa Fé, Argentina	Pathogenic	10
Tomasito	Corrientes, NE Argentina	Pathogenic	11
Sauce	Corrientes, NE Argentina	Pathogenic	12
Brasil	Brasil	Pathogenic	13
Uruguay	Uruguay	Pathogenic	14

Table 2
B. bigemina strains used in this study.

Isolate	Geographical location	Phenotype	ST
Sanger	Australia	Pathogenic	1
M1A	Corrientes, NE Argentina	Attenuated	2
M1P	Corrientes, NE Argentina	Pathogenic	3
M2P	Corrientes, NE Argentina	Pathogenic	4
S1A	Salta, NW Argentina	Attenuated	2
S2A	Salta, NW Argentina	Attenuated	2
S2P	Salta, NW Argentina	Pathogenic	5
S3P	Salta, NW Argentina	Pathogenic	6
Brasil	Brasil	Attenuated	7
Mexico	México	Pathogenic	8
M30	Salta, NW Argentina	Pathogenic	9
B38	Salta, NW Argentina	Pathogenic	10

2.2. Target loci

Twenty three *loci* were preselected for *B. bovis* and twenty one for *B. bigemina* using information available from genome strains. All preselected *loci* were single copy and homogeneously distributed through the four *B. bovis* chromosomes. In *B. bigemina*, *loci* were selected using the contigs available at the *B. bigemina* genome project (http://www.sanger.ac.uk/resources/downloads/protozoa/babesia-bigemina.html). The search was performed by Blast (tblastn) using *B. bovis* annotated genes as query. Primers were designed for all genes and amplifications were performed with DNA from 2 reference strains from each parasite. Only those PCR products with high sensitivity and specificity for the corresponding parasite were finally selected. Specificity controls included DNA from uninfected *Bos taurus, Anaplasma marginale* and the heterologous *Babesia* species (i.e. *B. bigemina* for *B. bovis* PCR and viceversa). Amplicons were sequenced in both strands and those *loci* showing a neutral pattern according to the ratio of mean non-synonymous

substitutions per non-synonymous site/mean synonymous substitution per synonymous site (d*N*/d*S* ratio) were selected for further analysis. The d*N*/d*S* ratio was calculated using the START2 program available from http://pubmlst.org/software/analysis/start2/ (Jolley et al., 2001).

Seven genes were finally chosen for *B. bovis*: *check* (Choline/ethanolamine kinase; XM_001610703.1), *dnaJ* (DnaJ domain containing protein; XM_001610032.1), *gpad* (G-patch domain containing protein; XM_001610843.1), *pkid* (Protein kinase domain containing protein; XM_001608647.1), *rcc* (Regulator of chromosome condensation (RCC1) domain containing protein; XM_001609092.1), *rip9* (Ribosomal protein L9, N-terminal domain containing protein; XM_001611709.1) and *rho4* (Rhomboid 4; XM_001610078.1). Details of the *B. bovis* target *loci* and their localization on the four chromosomes of strain T2Bo are presented in Table 3. Four out of the 7 *B. bovis* genes have introns (*check, gpad, pkid and rho4*) and the other 3 do not. All nucleotide polymorphism analysis was done on the exons.

In *B. bigemina*, 6 genes were finally selected. The sinteny between each contig containing the target *loci* and the 4 *B. bovis* chromosomes was confirmed using Promer (Kurtz et al., 2004). All *loci* were named according to its ortholog in the *B. bovis* genome: *cyp* (Papain family cysteine protease containing protein; XM_001612081.1), *dnaJ* (DnaJ domain containing protein; XM_001610032.1), *sbp3* (Spherical body protein 3; XM_001609020.1), *sbp4* (Spherical body protein 4; XM_001610418.1), *zfc* (Zinc finger C-x8-C-x5-C-x3-H type domain containing protein; XM_001611468.1) and *rcc* (Regulator of chromosome condensation (RCC1) domain containing protein; XM_001609092.1). Contigs 4137 and 4115 (*cyp* and *zfc* respectively) mapped against *B. bovis* chromosome 3; contigs 4145.0 and 3863.0 (*sbp3* and *rcc* respectively) mapped against *B. bovis* chromosome 1; contig 4182 (*dnaJ*) mapped against *B. bovis* chromosome 2 and contig 4098 (*sbp4*) mapped against *B. bovis* chromosome 4. With this analysis, we assume that the target genes are evenly distributed in the *B. bigemina* genome. Details of the *6 loci* are presented in Table 4. None of the *B. bigemina* genes have predicted introns.

2.3. PCR amplification and gene sequencing

Primers used to amplify and sequence the several gene targets are listed in Tables 3 and 4. PCR was performed in a 50 μl reaction mixture containing 0.4 μmol of each primer, 0.2 mM of each deoxyribonucleotide triphosphate (Promega, Madison, Wi. USA), 1.25 U of GoTaq DNA polymerase (Promega.), 10 μl of 5× PCR buffer and 200 ng genomic DNA. Amplification was carried out in a thermocycler (Bio-Rad MyCycler Thermal Cycler) with an initial 3 min denaturation at 94 °C, followed by 35 cycles, which consisted of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and elongation at 72 °C for 45 s, followed by a final extension step of 72 °C for 10 min. Field samples in which one-step amplification was negative, were subjected to a first round of PCR using Thermo Scientific Phusion® Blood Direct PCR Kit followed by a second step using the same primers and standard PCR reagents. Five microliters of each amplified product were electrophoresed in a 1% agarose gel stained with ethidium bromide. A molecular size marker (1 Kb Plus DNA Ladder, Invitrogen) was used to determine PCR product size. The remaining 45 μl of amplified product were purified by precipitation with 11.25 μl of 125 mM EDTA and 135 μl of absolute ethanol, centrifugation at 10,000*g*, precipitation with 70% ethanol and resuspension in pure water. Both strands of purified amplicons were sequenced on a Big Dye Terminator v3.1 kit from Applied Biosystems and analyzed on an ABI 3130XL genetic analyzer from the same supplier. Sequences were deposited in GenBank under accession numbers JX495180-JX495251 for *B. bigemina* and JX495252-JX495349 for *B. bovis*.

2.4. Development of the MLST-pipeline for sequence analysis and ST assignment

Editing and analysis of chromatogram traces were performed using MLST-pipeline, a web-based workflow for the analysis of Multilocus Sequence Typing schemes developed at the Bioinformatics Unit of the Biotechnology Institute at INTA (Perez Principi et al., 2010). The open source pipeline is available for UNIX-type Platforms. Briefly, raw trace files from both strands were named in a standardized way so as to identify the trace files that have to be assembled together (forward and reverse strand). Each base of the selected template region was confirmed by at least two chromatograms (forward and reverse); if there were ambiguities for a sequence, the sequence analysis was repeated. Base calling was performed using Phred (Richterich, 1998). Manual analysis on the electropherograms was also performed so as to detect double peaks that could be indicative of mixed infections. Only the predominant allele present at each locus within each infection was considered. This predominant allele was defined as the highest peak in electropherogram traces. This procedure results in unbiased estimation of allele frequencies within a population, if we assume the composition of PCR products is representative of the composition of templates (Anderson et al., 2000).

Assembling was carried out automatically using pairs of forward and reverse reads using CAP3 (Huang and Madan, 1999) and sequence trimming was performed to get FASTA formatted sequences. At this step a checkpoint report allows the user to identify which contigs are assembled and which are rejected.

MLST data were analyzed by the standard MLST approach (Maiden, 2006); for each gene, a number was attributed to each allelic variant, and the sequence type (ST) of a strain corresponded to the combination of the allele numbers of the seven or six genes. *B. bovis* and *B. bigemina* genome strains were selected as the reference strains for ST determination. The ST assignment was performed by in-house developed scripts where single nucleotide polymorphisms (SNPs) of a new allele were deduced by alignment and comparison to the reference sequences. At this step a check point report informed about the whole ST assignation process. The criteria used to define a true SNP in a sequence were two: that the SNP should be biallelic and it should be present in at least 2 of the isolates analyzed. Only nucleotide changes in both forward and reverse sequences were accepted.

The pipeline outputs were FASTA alignment files ready to use for downstream analysis (e.g. Phylogenetic analysis, e-BURST). The whole process generated two databases: a temporary one that held all the results generated by the pipeline during the user guided process and a permanent one that kept genotype information after input data was curated by the database administrator. The database can be accessed at http://inta.gov.ar/Babesia_MLST.

2.5. Genetic variability and DNA polymorphism analyses

The number of alleles was used to compare the levels of genetic diversity between *B. bovis* and *B. bigemina*. Given the differences in the sample sizes from each species (14 vs 12 isolates) the rarefaction procedure implemented in Fstat 2.9.3.2 software (Goudet, 1995) was used to obtain an adjusted measure of the number of alleles per locus. Differences between *B. bovis* and *B. bigemina* were tested for significance by a Wilcoxon test using Infostat (Di Rienzo et al., 2009).

Simpson's index of diversity was used to assess the discriminatory power of the typing method presented here, as described in Hunter and Gaston (1988). It was calculated using software available at http://insilico.ehu.es/mini_tools/discriminatory_power/index.php.

Table 3
List of loci used for *B. bovis* MLST, with primer details.

Gene	Function	Accession code	Chromosome	Primer sequence	Amplicon size (bp)
gpad	G-patch domain containing protein	XM_001610843.1	4	***gpad*-F**: 5′ ATTTAGGGAGAAGGTGCAGA 3′ ***gpad*-R**: 5′ TCACTTCTTTTCGGTTGAATGT 3′	563
check	Choline/ethanolamine kinase, putative	XM_001610703.1	4	***check*-F**: 5′ CCGAACTACCATCCGAAGAA 3′ ***check*-R**: 5′ TCTCCGATGATTGGTTCCAC 3′	612
dnaJ	DnaJ domain containing protein	XM_001610032.1	2	***dnaJ*-F**: 5′CGTTGATAGCGTGCTGACA 3′ ***dnaJ*-R**: 5′ATTCCTGTTGACGCTTGGTC 3′	542
pkid	Protein kinase domain containing protein	XM_001608647.1	1	***pkid*-F**: 5′ TGGCACGTTTTGGTTGTTTA 3' ***pkid*-R**: 5′ AAAGCGAATGACTTGGCAGT 3′	581
rcc	Regulator of chromosome condensation (RCC1) domain containing protein	XM_001609092.1	1	***rcc*-F**: 5′ CAAGGAGATGCAACGAAGGT 3′ ***rcc*-R**: 5′ CAAGACCATATCCGCCCTAA 3′	605
rip9	Ribosomal protein L9, N-terminal domain containing protein	XM_001611709.1	3	***rip9*-F**: 5′ GCCGTCAATATCGCAAACGC 3′ ***rip9*-R**: 5′ CTTCTGGCGCATGTGATCTA 3′	388
rho4	Rhomboid 4	XM_001610078.1	2	***rho4*-F**: 5′ CACTCATATCGTGGGCATTG 3′ ***rho4*-R**: 5′: TTATCCATTGGTGCCGTTTT 3′	609

Table 4
List of loci used for *B. bigemina* MLST with primer details. Gene names are of *B. bovis* orthologs.

Gene	Function*	Contig number*	Primer sequence	Amplicon size (bp)
cyp	Papain family cysteine protease containing protein	4137	***cyp*-F**: 5′ACGTCGGTCATGGAGAAGGC 3′ ***cyp*-R**: 5′AGCGATTGCGGACACGTGGA 3′	888
dnaJ	DnaJ domain containing protein	4182	***dnaJ*-F**: 5′TAGCGGGTCAAGTGAAGAGGA 3′ ***dnaJ*-R**: 5′AAGCGTCACTGAAGGCACCG 3′	563
sbp3	Spherical body protein 3	4145	***sbp3*-F**: 5′AGGGGTGTGGTCAAGTTGGT 3′ ***sbp3*-R**: 5′CTTCATCATCAGGGTCCGCG 3′	625
sbp4	Spherical body protein 4	4098	***sbp4*-F**: 5′GTGCGCGGTTTTGACAATGT 3′ ***sbp4*-R**: 5′TCGTCGAACTGCTCAAGGGT 3′	695
rcc	Regulator of chromosome condensation (RCC1) domain containing protein	3863.0	***rcc*-F**: 5′ GACTGGACATGACACCGTCAA 3′ ***rcc*-R**: 5′ TGCCGTTTCCTGTGAGTTGCTA 3′	665
zfc	Zinc finger C-x8-C-x5-C-x3-H type domain containing protein	4115.0	***zfc*-F**: 5′TCTTGTTCCTTGGCGATGGA 3′ ***zfc*-R**: 5′CGTGATTTCGGTGGAGCAGT 3′	627

* Contigs 4137 and 4115.0 mapped against *B. bovis* chromosome 3; contigs 4145 and 3863.0 mapped against *B. bovis* chromosome 1; contig 4182 mapped against *B. bovis* chromosome 2 and contig 4098 mapped against *B. bovis* chromosome 4.

DNA sequence polymorphism and all subsequent analyses were performed using DnaSP 5.00.02 package (Librado and Rozas, 2009). Nucleotide diversity, Pi (π), the average number of nucleotide differences per site between two sequences was calculated according to Nei (1987), using the Jukes and Cantor (1969) correction. Theta (Watterson's mutation parameter) was calculated for the whole sequence from S (Watterson, 1975). Eta (η) is the total number of mutations, and *S* is the number of segregating (polymorphic) sites d*N* (the number of non-synonymous substitutions per non-synonymous site) and d*S* (the number of synonymous substitutions per synonymous site) for any pair of sequences were calculated according to Nei and Gojobori (1986). Tajima's D test (Tajima, 1989) was used for testing the null hypothesis of neutrality (Kimura, 1983). The statistical significance of Tajima's D was calculated using a two tailed test and the coalescent algorithm implemented in DnaSP 5.00.02.

Linkage disequilibrium among the genes included in *B. bovis* and *B. bigemina* MLST schemes was assessed using the standardized Index of Association (I_{As}) as implemented in the software LIAN 3.0 (Haubold and Hudson, 2000). Linkage disequilibrium among polymorphic sites within genes was estimated based on the squared allele-frequency correlations (r^2) (Hill and Robertson, 1968), since the standardized disequilibrium coefficient ($\acute{D}$) is strongly affected by small sample sizes, resulting in highly erratic behaviour when comparing *loci* with low allele frequencies (Flint-Garcia et al., 2003). Significant pairwise associations were assessed by Fisher exact tests and corrected for multiple comparisons using Bonferroni procedures. Pairwise comparisons from all genes were pooled and plotted together for each species. The ZnS statistic (Kelly, 1997), which is the average of r^2 over all pairwise comparisons was computed using the DnaSP 5.00.02 package.

2.6. Phylogenetic analysis

Phylogenetic analysis was performed by constructing Maximum Likelihood (ML) trees from both individual *loci* and concatenated nucleotide sequences. Prior to ML analyses, the computer program ModelTest 3.06 (Posada and Crandall, 1998) was used to

select the most appropriate model of molecular evolution for each *loci* and also for the concatenated datasets according to the Akaike Information Criterion (AIC). Concatenated sequences of *B. bigemina* and *B. bovis* were analyzed under a Hasegawa–Kishino–Yano (HKY) model with gamma rate heterogeneity and invariant sites and under a General Time Reversible (GTR) model with gamma distributed-rates, respectively.

ModelTest analysis was conducted at the Phylemon Web server (Tarraga et al., 2007). ML tree searches were performed with MEGA 5 (Tamura et al., 2011). Node robustness was assessed with 500 bootstrap replicates.

2.7. Strain relationships based on allelic profiles

The Phyloviz program (http://goeburst.phyloviz.net/), which uses the same similar clustering rules as eBURST but provides a global optimal solution, was used to determine the relationships between STs (Francisco et al., 2012; Feil and Enright, 2004). Clonal complexes (CCs) were defined as STs that are linked through single locus variants (SLVs) and named on the basis of the predicted founder ST, which is the ST associated having the largest number of SLVs.

3. Results

3.1. Identification of isolated samples and assignment of haplotypes

The confirmation of the samples used in this study as *B. bovis* or *B. bigemina* was conducted using the reverse line blot hybridization assay with a custom designed membrane (Petrigh et al., 2008). Only samples that were positive for the *Babesia/Theileria* catch-all probes and uniquely positive to *B. bovis* or *B. bigemina* specific probes were chosen (data not shown).

MLST alleles and STs for *B. bovis* and *B. bigemina* isolates are presented in Supplementary Tables 1 and 2 respectively. Visual analysis of the electropherograms in field isolates of both parasites revealed the existence of more than one genotype in the same sample (Suppl. Fig. 3). All *B. bovis loci* showed double peaks in field samples. The locus which showed the highest frequency of these peaks was *rcc*. Double peaks were visualized as a major and a minor peak of a third of the former. Similarly, in *B. bigemina*, the field case M30 showed also double peaks in 4 out of 6 *loci* (*sbp4, zfc, rcc* and *cyp)*. With the exception of *cyp*, the remaining 3 loci showed a single peak after tick and calf passages which corresponded to the minor peak while *cyp*'s double peaks remained unchanged throughout the passages.

Possible errors due to DNA polymerase were discarded since the double peaks were coincident with SNPs at positions of high frequency in all isolates analyzed. Therefore, STs were constructed for each sample representing the combination of predominant alleles identified at each locus as already described (Anderson et al., 2000). While this approach does not capture all the diversity in each sample, it can be viewed as a representative sample of the genotypes in the population.

3.2. Allelic and genotypic variation in B. bovis and B. bigemina

Allele sizes for the genes included in the *B .bovis* MLST scheme varied between 388 bp (*rip9*) and 612 bp (*check*). In *B. bigemina*, sizes varied between 563 bp (*dnaJ*) and 888 bp (*cyp*). While the number of distinct alleles per gene in *B. bovis* was variable and ranged from 2 (rho 4) to 10 (rip9), in *B. bigemina*, a lower number of alleles was found. Half of the genes selected had between 1 (*dnaJ* and *sbp3*) and 2 (*sbp4*) alleles. The other 3 genes had either 7 (*cyp*) or 8 alelles (*zfc* and *rcc*). No statistical differences in genetic variability were detected when comparing the number of alleles per locus between *B. bovis* and *B. bigemina* (Wilcoxon test, $p > 0.05$).

This variation in the number of alleles from one locus to another did not, however, reflect the polymorphism in the target gene segment amplified. Polymorphic sites or S (i.e. number of mutated sites per *locus*) varied from 8 to 40 for *B. bovis* (Table 5) and from 0 to 15 for *B. bigemina* (Table 6). When weighting these data by considering the size of the amplified fragment, the resulting percentage of site polymorphism varied from 1.69% to 8.69% for *B. bovis* and 0% to 2.55% *for B. bigemina* (Tables 5 and 6, respectively). Again, *B. bigemina* displays lower levels of variation than *B. bovis*.

There was no correlation between the number of alleles and the number of polymorphic sites, or the percentage of polymorphisms in *B. bovis*, indicating that a very few number of strains carry most of the observed polymorphisms on a limited number of alleles. However, in *B. bigemina*, the number of polymorphic sites

Table 5
Assessment of DNA polymorphism, linkage disequilibrium and neutrality tests for the *B. bovis* strains used in this study.

Gene	Hp	*S*	η	Pa	η(s)	θ	π	d*N*/d*S*	Tajima's D	Rm	ZnS	PW	*F*	BP	Si
check	5	9	9	9	0	0.00534	0.00715	0.26666667	1.30333 ns	1	0.4781	36	21	3	0.8571
dnaJ	8	18	19	16	2	0.01171	0.01189	0.30379937	0.06365 ns	1	0.5119	136	66	0	0.8242
gpad	4	20	20	20	0	0.01553	0.02540	0.30379937	2.65649**	0	0.8429	190	171	0	0.7473
pkid	4	21	21	21	0	0.01295	0.02058	0.13710083	2.47210**	0	0.7607	210	169	0	0.6923
rcc	6	14	15	14	0	0.00828	0.00898	0.12890545	0.34935 ns	1	0.5364	78	46	1	0.8022
rip9	10	40	43	39	1	0.03885	0.05021	0.10555423	1.27460 ns	6	0.3439	666	209	0	0.9231
rho4	2	8	8	8	0	0.00547	0.00764	0.26666667	1.50029 ns	0	1	28	28	28	0.4396

HP: haplotypes.
S: polymorphic sites.
η: total number of mutations.
Pa: parsimony informative sites.
η(s): number of singletons.
θ: Watterson's mutation parameter (calculated from Eta).
π: Nucleotide diversity.
d*N*/d*S*: rate of non-synonymous vs. synonymous substitutions.
Tajima's D: Tajima's *D* test of neutrality.
Rm: minimal recombination events.
ZnS: ZnS statistic.
PW: number of pairwise comparisons.
F: number of significant associations tested by a Fisher exact test.
BP: number of significant associations tested by a Fisher exact test after Bonferroni procedure.
Si: Simpson's index.
** Significant at $0.001 < P < 0.01$.

Table 6
Assessment of DNA polymorphism, linkage disequilibrium and neutrality tests for the *B. bigemina* strains used in this study.

Gene	Hp	*S*	η	Pa	η(s)	θ	π	d*N*/d*S*	Tajima's D	Rm	ZnS	PW	*F*	BP	Si
cyp	7	6	6	5	1	0.00302	0.00326	0.48497324	0.32765 ns	2	0.1208	15	0	0	0.8939
dnaJ	1	0	0	0	0	ND	ND	ND	ND	ND	ND	ND	ND	ND	0
sbp3	1	0	0	0	0	ND	ND	ND	ND	ND	ND	ND	ND	ND	0
sbp4	2	1	1	1	0	0.00072	0.00114	0	1.46364 ns	3	0.3890	105	33	0	0.5303
zfc	8	9	10	9	0	0.00658	0.00770	3.59336609	0.75044 ns	5	0.1464	28	0	0	0.9242
rcc	8	15	15	15	0	0.00902	0.01259	0.22909991	1.46364 ns	0	0.3980	0	0	0	0.9242

HP: haplotypes.
S: polymorphic sites.
η: total number of mutations.
Pa: parsimony informative sites.
η(s): number of singletons.
θ: Watterson's mutation parameter (calculated from Eta).
π: Nucleotide diversity.
d*N*/d*S*: rate of non-synonymous vs. synonymous substitutions.
Tajima's D: Tajima's *D* test of neutrality.
Rm: minimal recombination events.
ZnS: ZnS statistic.
PW: number of pairwise comparisons.
F: number of significant associations tested by a Fisher exact test.
BP: number of significant associations tested by a Fisher exact test after Bonferroni procedure.
Si: Simpson's index.

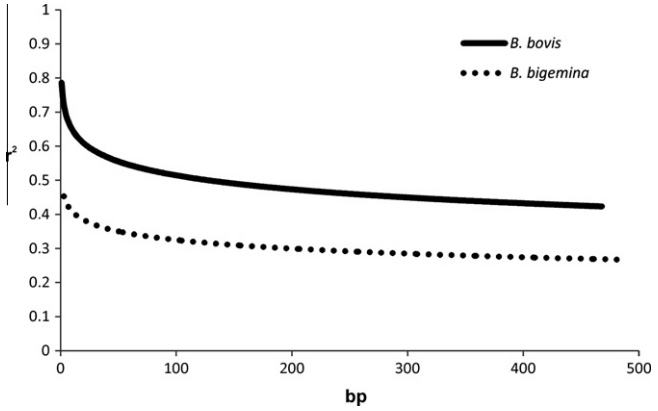

Fig. 1. Linkage disequilibrium (LD) decay over distance in *B. bovis* and *B. bigemina*. The squared correlations of allele frequencies (r^2) between polymorphic sites is plotted against the distance in base pairs.

correlates significantly with the number of alleles (Pearson coefficient = 0.91, $p < 0.01$).

Simpson's diversity indexes for each locus are shown in Table 5 and 6. They varied between 0.439 (*rho4*) and 0.93 (*rip9*) for *B. bovis* and between 0 (*dnaJ* and *sbp3*) and 0.9242 (*zfc* and *rcc*) for *B. bigemina*. Given that no ST was found more than once in the *B. bovis* samples, the overall MLST scheme exhibited the maximum discrimination power (Si = 1). The same level of discrimination (Si = 1) was achieved with only 4 *loci* (*check*, *dnaJ*, *rip9* and *rcc*). For *B. bigemin*a the discrimination power of the overall MLST scheme with 3 *loci* (*cyp*, *rcc* and *zfc*) was 0.9545 indicating that a set of 3 genes could be used instead of the complete MLST scheme.

3.3. Analysis of nucleotide variation

The nucleotide diversity indexes (π) obtained for *B. bovis* ranged from 0.00715 (*check)* to 0.05021 (*rip9)*. In *B. bigemina*, these indexes varied from 0.0011 (*sbp4)* to 0.012 (*rcc)*.

The proportion of nucleotide substitutions that altered the amino acid sequence (non-synonymous substitutions, (d*N*)) and the proportion of silent changes (synonymous substitutions, (d*S*)) were calculated for each gene. The d*N*/d*S* ratio varied from 0.1 and 0.3 for *B. bovis*. In *B. bigemina*, the d*N*/d*S* ratio of *dnaJ* and *sbp3* could not be calculated because only one allele was found. Of the remaining 4 genes, 3 of them had ratios between 0 and 0.48 and only one gene (*zfc*) had a value above 1.

The Tajima's *D* neutrality test was also conducted on the *loci* sequences. Tajima's *D* tests for all genes in *B. bigemina* and for five out of seven in *B. bovis* showed no statistical significance, indicating that a neutral model of sequence evolution could not be rejected. In *B. bovis, gpad* and *pkid* showed *D* values significantly larger than 0.

The I_{As} indexes among genes did not show significant departures from linkage equilibrium either for *B. bovis* (I_{As}: 0.0412, $p > 0.05$) or for *B. bigemina* (I_{As}: 0.0045, $p > 0.05$). Within genes, statistically significant LD ($p < 0.001$) was found for 25% of the total pairwise comparisons performed for *B. bovis*, whereas no significant deviations from independent segregation were obtained for *B. bigemina* at $p < 0.001$. Since all the regions studied here are <1kbp, this analysis reveals disequilibrium patterns at short distances. For *B. bovis, a* logarithmic trendline fit to the pairwise r^2 reaches a value of 0.425 at 450 bp, whereas for *B. bigemina* the estimate drops to 0.267 for the same distance (Fig. 1). Moreover, average r^2 for distances under 100 bp were markedly lower for *B. bigemina* than for *B. bovis*. In sum, the average magnitude of LD, measured by both ZnS (Tables 5 and 6) and the logarithmic trend line fit to pairwise r^2 was lower in *B. bigemina* than in *B. bovis*.

3.4. Phylogenetic analysis of MLST loci

Variation was observed in the seven genes examined for *B. bovis* isolates, resulting in a relatively high number of informative sites for phylogenetic analysis (Table 5). Most polymorphisms corresponded to nucleotide substitutions, with only one base deletion detected in the only intron of *pkid* locus.

ML trees from individual *loci* were characterized by weak branch supports, with only a few groups showing bootstrap values above 80%. Tree topologies were mostly incongruent and no clear groupings were detected at any level (data not shown). Combination of individual *loci* into a single concatenated matrix produced a moderate improvement on branch support values, and most of the groups retrieved in the resulting tree differed from those obtained in the individual gene analysis (Supplementary Figs. 1 and 2). No association was apparent between the inferred clades and the geographical origin of the samples, except for a pair of acute

cases that corresponded to the same outbreak in the province of Corrientes (Tomasito and Sauce).

Four out of the six genes analyzed for *B. bigemina* showed variation. For this parasite, individual *loci* were not used for phylogenetic analysis due to the low number of parsimony informative sites for each gene, and only the combined data matrix was considered for phylogenetic inference.

As observed for *B. bovis*, the resulting ML tree showed low branch support. The three experimentally attenuated Argentinean strains (M1A, S1A and S2A) clustered together, given that they share the same ST (Suppl. Fig. 2). No other pattern of association was found for *B. bigemina*.

3.5. Strain relationships based on allelic profiles

Relationships based on STs can be more reliable than nucleotide-based phylogenies if homologous recombination occurs, as import of a single divergent allele would strongly affect the phylogenetic position of the recipient strain. A clonal complex (CC) can be defined as a group of profiles in which each profile differs by no more than one gene from at least one other profile of the group (Single locus variant, SLV) (Feil and Enright, 2004). STs with a single gene difference in the gene set are very likely to share a common ancestor. GoeBURST analysis of allelic profiles of *B. bovis* revealed 1 CC containing STs 11 and 12 (data not shown) that corresponds to Tomasito and Sauce isolates, which were clinical cases isolated from the same province of Corrientes. These two STs were included within the same clade in the nucleotide-based phylogeny (Suppl. Fig. 1). The remaining 12 STs were singletons. No CC could be defined for *B. bigemina*.

4. Discussion

Genetic markers are invaluable tools to investigate pathogen diversity and population structure and have an enormous impact in the design and implementation of control strategies.

We developed a MLST scheme based on a set of housekeeping genes for evolutionary analysis and strain typing in *B. bovis and B. bigemina*. This is the first MLST scheme developed on *Babesia* species and required a thorough analysis of the genomic data available at public databases. This information facilitated gene choice and primer design. The selected genes have been confirmed to be ubiquitous within the *Babesia spp*. population and are likely to be predominantly under neutral selection.

A general tendency to d*N*/d*S* values below 1 has been observed in most protein coding regions of a wide variety of organisms, suggesting that some degree of purifying selection might be a common feature of protein evolution (Messier and Stewart, 1997; Jordan et al., 2002; Kitami and Nadeau, 2002). The genes included in our MLST schemes seem to be no exception to this pattern; d*N*/d*S* values obtained here were below 1 for all *loci* in *B. bovis* and for 5 out of 6 *loci* in *B. bigemina*. Moreover, Tajima's *D* test failed to detect deviations from neutrality in 11 of the 13 genes studied. We therefore argue that the MLST genes proposed here fulfill all the criteria suggested for large-scale typing (Maiden, 2006) and form a representative sample of the core genomes of both *Babesia* species.

The MLST strategy implies the generation and analysis of a large amount of data, therefore, a custom-designed bioinformatic pipeline named "MLST-Pipeline" was developed to automate DNA sequence editing and analysis and to significantly reduce the time required for processing data. Another advantage of the "MLST-Pipeline" developed here is the use of free software which makes it available to low-income research or health institutions. The "MLST-Pipeline" is fully adaptable to other MLST schemes of any organism.

Superimposed double nucleotide peaks on the sequence electropherograms were visualized in both *Babesia* species. In our approach, base-calling was based on peak height, as it was demonstrated that peak height on a pherogram is a product of the true proportions of parasite clones (Ford and Schall, 2011). Although TA-cloning could be an alternate way to assess allele frequencies for each gene, the idea behind the MLST typing is a direct and readily automated technology that can be scalable to high numbers of samples.

The double peaks found in this study can only be attributed to mixed genotype infections since both *Babesia* species are haploid and all MLST *loci* selected were single copy. In Argentina, both *Babesia spp.* are transmitted by the one-host tick *Rhipicephalus microplus*. Therefore, coinfection is highly likely to occur, especially in regions of high tick transmission from where many of the studied samples came from. The *B. bigemina* M30 isolate is an example of the heterogeneity of field strains and supports the idea that population dynamics can be altered by passages through different hosts (i.e. bovines and ticks).

Berens et al. (2007) have previously demonstrated that both bovines and ticks can support virulent *B. bovis* coinfection through all phases of the hemoparasite's life cycle. More recently, Lau et al. (2010) also found that multiple genotypes can coexist in a single bovine and that the predominant genotype at one time point was not the same 6 months later. This could be result of super-infection due to exposure to tick population infected with one or more different genotypes. The work by Simuunza et al. (2011) also revealed that most field isolates of *B. bovis* from Turkey and Zambia contain a mixture of mixed genotypes. Mixed infections were also reported for different species of the phylogenetically related parasites *Plasmodium* (Bonizzoni et al., 2009) and *Theileria* (Oura et al., 2005).

In both parasites, a high level of nucleotide diversity was found, with an ST/strain ratio of 1 for *B. bovis* and 0.83 for *B. bigemina*. Similar results were obtained in a previous study on *B. bovis*, in which unique multilocus genotypes were obtained and recombination was strongly suggested (Simuunza et al., 2011). Even though the number of Argentinean samples studied is low, the high number of genotypes found, along with the finding of more than one genotype in the same bovine sample, indicates that in regions of high transmission, ticks may be feeding from a mixture of genotypes which will have the possibility to recombine within the midgut of the vector resulting in genetic exchange. In this sense, recombination tests were positive for 4 *loci* in *B. bovis* and 3 in *B. bigemina*, which supports the hypothesis of genetic exchange during the tick stages of both parasites. Consistent with this hypothesis, linkage disequilibrium among genes was not detected for either parasite indicating the absence of clonal structure and a strong influence of the sexual stage in the distribution of genetic variation. In addition, the imprint of recombination was also made apparent by the rapid decay of LD among segregant sites within genes, particularly for *B. bigemina*. Despite the short-range LD that we were able to asses, the logarithmic trend line predicted a decay of LD (i.e. $r^2 \sim 0.1$) within 41 kbp for *B. bigemina* and within 111 kbp for *B. bovis*.

Incongruence among the phylogenetic trees derived from the individual *loci* that compose MLST schemes has been interpreted as evidence of recombination in different microorganisms (Salerno et al., 2007; Feil and Spratt, 2001; Yeo et al,. 2011). Although lack of congruence was clearly the case for the individual trees of *B. bovis*, the low bootstrap values obtained for most groupings suggest that other factors such as insufficient or contradictory phylogenetic signal could also be responsible for this pattern.

As genetic recombination events tend to blur the phylogenetic signal, strain relationship studies based on allelic profiles or STs

represent a better approach to analyze relations among samples. GoeBURST was employed to divide the two set of strains into clonal complexes by comparison of their allelic profiles. This analysis indicated that only two *B. bovis* isolates were closely related establishing a single clonal complex (ST11 and ST12) within the collection studied. This CC was built as a consequence of both isolates sharing 6 out of 7 *loci* (SLV). Interestingly, the two samples that belong to this CC (Tomasito and Sauce) had a common geographic and epidemiologic origin, as they correspond to isolates from a babesiosis outbreak in the Corrientes province, Argentina. This suggests that the lack of a more general population structure was probably due to the fact that the majority of isolates corresponded to unrelated samples. In other words, the obtained CC could be regarded as an indication of the power of this MLST scheme to track local disease outbreaks and how they relate to each other. Two *B. bigemina* samples, M30 and B38, were closely related since B38 was derived from two bovine and one tick passages of M30. However, they differed in 4 out of 6 *loci*. This could be due to population selection within the host, the vector and/or the transmission process. The 3 Argentinian *B. bigemina* vaccine strains studied here (M1A, S1A and S2A) showed the same ST and therefore clustered together in the phylogenetic analysis although they were attenuated in independent procedures from different virulent isolates. It remains to be determined if this fact could be related to the results published by Lau et al. (2011) for *B. bovis* in which genetic diversity was significantly reduced in experimentally attenuated strains from diverse geographical backgrounds.

The proposed MLST scheme has a high discriminatory power even in samples from close geographical distance (i.e. same ranch). The overall Si for the entire scheme was 1 for *B. bovis* and 0.9545 *for B. bigemina*, well above the 90% which has been suggested as the minimal value required for reliable typing (Hunter and Gaston, 1988). The small sample sizes and the widespread geographical origin of the samples studied here could lead to the overestimation of Si indexes. However, even if the Si indexes were lower in a larger sample, it is likely they would still be above the accepted 90%, given the high levels of genetic variation detected in both parasites.

The MLST method described here is based on seven and six genes for *B. bovis* and *B. bigemina* respectively. Although most MLST schemes are based on seven genes, there is no particular reason for choosing this number of *loci*, especially if a lower number provides enough discrimination. Our analysis of allelic diversity indicated that 4 genes were enough to discriminate all *B. bovis* strains and 3 genes gave equal discriminatory power than 6 for *B. bigemina*. Further analyses using a larger set of samples will confirm if a smaller number of MLST *loci* could be used instead of the complete scheme presented in this study.

In summary, the MLST scheme developed here is a robust, objective and easily adoptable technology to analyze various aspects of the genetic diversity and population structure of parasites of the genus *Babesia*. Definitive conclusions about the prevalence or absence of certain genotypes in defined geographic regions will be achieved by analyzing a larger number of samples. The potential application of these markers may include worldwide comparison of strain stocks, subunit vaccine design, characterization of parasite lines used for live vaccination, investigation of the basis of live vaccine breakdowns and comparisons among livestock strains and those found in wild animal reservoirs.

Acknowledgments

The technical assistance of P. Zimmer, N. Sarmiento, R. Neumann and D. Perez Principi is gratefully acknowledged. The authors would like to especially thank Dr. Ignacio Echaide for his permanent scientific input and for providing us with the reference strains. Financial support was obtained by ANPCyT-PICT 1634 and INTA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2012.12.005.

References

Anderson, T.J., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P., 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. 17, 1467–1482.

Berens, S.J., Brayton, K.A., McElwain, T.F., 2007. Coinfection with antigenically and genetically distinct virulent strains of *Babesia bovis* is maintained through all phases of the parasite life cycle. Infect. Immun. 75, 5769–5776.

Bock, R., Jackson, L., de Vos, A., Jorgensen, W., 2004. Babesiosis of cattle. Parasitology. 129 (Suppl), S247–S269.

Bonizzoni, M., Afrane, Y., Baliraine, F.N., Amenya, D.A., Githeko, A.K., Yan, G., 2009. Genetic structure of *Plasmodium falciparum* populations between lowland and highland sites and antimalarial drug resistance in Western Kenya. Infect. Genet. Evol. 9, 806–812.

Di Rienzo, J.A., Casanoves, F., Balzarini, M.G., Gonzalez, L., Tablada, M., Robledo, C.W., 2009. InfoStat versión 2009. GrupoInfoStat FCA, Universidad Nacional de Córdoba, Argentina.

Feil, E.J., Enright, M.C., 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. 7, 308–313.

Feil, E.J., Spratt, B.G., 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. 55, 561–590.

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S., 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant. Biol. 54, 357–374.

Ford, A.F., Schall, J.J., 2011. Relative clonal proportions over time in mixed-genotype infections of the lizard malaria parasite *Plasmodium mexicanum*. Int. J. Parasitol. 41, 731–738.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., Carriço, J.A., 2012. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13, 87–97.

Genis, A.D., Perez, J., Mosqueda, J.J., Alvarez, A., Camacho, M., Mde Muñoz, L., Rojas, C., Figueroa, J.V., 2009. Infect. Evol. 6, 1102–1107.

Goudet, J., 1995. Fstat version 1.2: a computer program to calculate F-statistics. Heredity 86, 485–486.

Haubold, B., Hudson, R.R., 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis. Bioinformatics 16, 847–848.

Higuchi, R., 1989. Rapid, efficient DNA extraction for PCR from cells or blood. Amplifications 2, 1–3.

Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231.

Hilpertshauser, H., Deplazes, P., Meli, M.L., Hofmann-Lehmann, R., Lutz, H., Mathis, A., 2007. Genotyping of *Babesia bigemina* from cattle from a non-endemic area (Switzerland). Vet. Parasitol. 145, 59–64.

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868–877.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26, 2465–2466.

Jolley, K.A., Feil, E.J., Chan, M.S., Maiden, M.C.J., 2001. Sequence type analysis and recombinational test (START). Bioinformatics 17, 1230–1231.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. 12, 962–968.

Evolution of protein molecules, 1969. In: Munro, H.N. (Ed.). Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132.

Kelly, J.K., 1997. A test of neutrality based on interlocus associations. Genetics 146, 1197–1206.

Kitami, T., Nadeau, J.H., 2002. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. Nat. Genet. 32, 191–194.

Kimura, M., 1983. Rare variant alleles in the light of the neutral theory. Mol. Biol. Evol. 1, 84–93.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12.

Lau, A.O., Cereceres, K., Palmer, G.H., Fretwell, D.L., Pedroni, M.J., Mosqueda, J., McElwain, T.F., 2010. Genotypic diversity of merozoite surface antigen 1 of *Babesia bovis* within an endemic population. Mol. Biochem. Parasitol. 172, 107–112.

Lau, A.O., Kalyanaraman, A.K., Echaide, I., Palmer, G.H., Bock, R., Pedroni, M.J., Rameshkumar, M., Ferreira, M.B., Fletcher, T.I., Mc Elwain, T.F., 2011. Attenuation of virulence in an Apicomplexan hemoparasite results in reduced genome diversity at the population level. BMC Genomics 12, 410–423.

Levi, M.G., Ristic, M., 1980. *Babesia bovis*: continuous cultivation in a microaerophilous stationary phase culture. Science 207, 1218–1220.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Maiden, M.C.J., 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60, 561–588.
Messier, W., Stewart, C.B., 1997. Episodic adaptive evolution of primate lysozymes. Nature 385, 151–154.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Oura, C.A., Asiimwe, B.B., Weir, W., Lubega, G.W., Tait, A., 2005. Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. Mol. Biochem. Parasitol. 140, 229–239.
Perez-Llaneza, A., Caballero, M., Baravalle, E., Mesplet, M., Mosqueda, J., Suarez, C.E., Echaide, I., Katzer, F., Pacheco, G.M., Florin-Christensen, M., Schnittger, L., 2010. Development of a tandem repeat-based multilocus typing system distinguishing *Babesia bovis* geographic isolates. Vet. Parasitol. 167, 196–204.
Perez Principi D., Delfino S., Guillemi E., Ruybal P., Wilkowsky S., Farber M., 2010. MSLT-pipeline: a web-based workflow for the analysis of Multilocus Sequence Typing schemes. In: Proceedings of the First International Conference on Bioinformatics (SOIBIO) Sociedad Iberoamericana de Informática, Chile.
Petrigh, R., Ruybal, P., Thompson, C., Neumann, R., Moretta, R., Wilkowsky, S., Draghi, G., Echaide, I., Torioni de Echaide, S., Farber, M., 2008. Improved molecular tools for detection of *Babesia bigemina*. Ann. N.Y Acad. Sci. 1149, 155–157.
Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Richterich, P., 1998. Estimation of errors in "raw" DNA sequences: a validation study. Genome Res. 8, 251–259.
Salerno, A., Delétoile, A., Lefevre, M., Ciznar, I., Krovacek, K., Grimont, P., Brisse, S., 2007. Recombining population structure of *Plesiomonas shigelloides* (Enterobacteriaceae) revealed by multilocus sequence typing. J. Bacteriol. 189, 7808–7818.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, New York.
Shkap, V., de Vos, A.J., Zweygarth, E., Jongejan, F., 2007. Attenuated vaccines for tropical theileriosis, babesiosis and heartwater: the continuing necessity. Trends Parasitol. 23, 420–426.
Simuunza, M., Bilgic, H., Karagenc, T., Syakalima, M., Shiels, B., Tait, A., Weir, W., 2011. Population genetic analysis and sub-structuring in *Babesia bovis*. Mol. Biochem. Parasitol. 177, 106–115.
Späth, E.J.A., Guglielmone, A.A., Signorini, A.R., Mangold, A.J., 1994. Estimación de las pérdidas económicas directas producidas por la garrapata *Boophilus microplus* y las enfermedades asociadas en la Argentina. Terios 23, 116–119.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tarraga, J., Medina, I., Arbiza, L., Huerta-Cepas, J., Gabaldón, T., Dopazo, J., Dopazo, H., 2007. Phylemon: a suite of web tools for molecular evolution, phylogenetics and phylogenomics. Nuc. Acids Res. 35, W38–W42.
Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.
Wilkowsky, S.E., Moretta, R., Mosqueda, J., Gil, G., Echaide, I., Lía, V., Falcón, A., Florin, Christensen.M., Farber, M., 2009. A new set of molecular markers for the genotyping of *Babesia bovis* isolates. Vet. Parasitol. 161, 9–18.
Wilkowsky, S., Farber, M., Gil, G., Echaide, I., Mosqueda, J., Alcaraz, E., Suarez, C.E., Florin-Christensen, M., 2008. Molecular characterization of *Babesia bovis* strains using PCR restriction fragment length polymorphism analysis of the msa2-a/b genes. Ann. N.Y. Acad. Sci. 1149, 141–144.
Yeo, M., Mauricio, I.L., Messenger, L.A., Lewis, M.D., Llewellyn, M.S., Acosta, N., Bhattacharyya, T., Diosque, P., Carrasco, H.J., Miles, M.A., 2011. Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl. Trop. Dis. 5, e1049.