Infection, Genetics and Evolution 30 (2015) 186–194

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Development of a Multilocus Sequence Typing scheme for the study of *Anaplasma marginale* population structure over space and time

Eliana C. Guillemi [a,*], Paula Ruybal [a], Verónica Lia [a], Sergio Gonzalez [a], Sergio Lew [b], Patricia Zimmer [c], Ludmila Lopez Arias [a], Jose L. Rodriguez [d], Sonia Y. Rodriguez [d], Roger Frutos [e], Silvina E. Wilkowsky [a], Marisa D. Farber [a]

[a] *Instituto de Biotecnologia, Centro de Investigaciones en Ciencias Veterinarias y Agronómicas, INTA, Buenos Aires, Argentina*
[b] *Instituto de Ingeniería Biomédica, Universidad de Buenos Aires, Buenos Aires, Argentina*
[c] *Estación Experimental Agropecuaria Mercedes, INTA, Corrientes, Argentina*
[d] *CORPOICA, Bogotá, Colombia*
[e] *Cirad, UMR 17, Cirad-Ird, TA-A17/G, Campus International de Baillarguet, Montpellier, France*

ARTICLE INFO

Article history:
Received 8 October 2014
Received in revised form 21 December 2014
Accepted 22 December 2014
Available online 27 December 2014

Keywords:
Anaplasma marginale
MLST
Tick-borne diseases
Population structure
Molecular epidemiology

ABSTRACT

Bovine Anaplasmosis caused by *Anaplasma marginale* is a worldwide disease prevalent in tropical and subtropical regions where *Rhipicephalus microplus* is considered the most significant biological vector. Molecular markers previously applied for *A. marginale* typing are efficient for isolate discrimination but they are not a suitable tool for studying population structure and dynamics. Here we report the development of an MLST scheme based on the study of seven genes: *dnaA*, *ftsZ*, *groEl*, *lipA*, *recA*, *secY* and *sucB*. Five annotated genomes (Saint Maries, Florida, Mississippi, Puerto Rico and Virginia) and 53 bovine blood samples from different world regions were analyzed. High nucleotide diversity and a large proportion of synonymous substitutions, indicative of negative selection resulted from DnaSP 5.00.02 package application. Recombination events were detected in almost all genes, this evidence together with the coexistence of more than one *A. marginale* strain in the same sample might suggest the superinfection phenomena as a potential source of variation. The allelic profile analysis performed through GoeBURST shown two main CC that did not support geography. In addition, the AMOVA test confirmed the occurrence of at least two main genetically divergent groups. The composition of the emergent groups reflected the impact of both historical and environmental traits on *A. marginale* population structure. Finally, a web-based platform "Galaxy MLST-Pipeline" was developed to automate DNA sequence editing and data analysis that together with the Data Base are freely available to users.

The *A. marginale* MLST scheme developed here is a valuable tool with a high discrimination power, besides PCR based strategies are still the better choice for epidemiological intracellular pathogens studies. Finally, the allelic profile describe herein would contribute to uncover the mechanisms in how intracellular pathogens challenge virulence paradigm.

1. Introduction

Bovine Anaplasmosis is a tick borne disease caused by the Gram negative bacterium *Anaplasma marginale*, an obligate intracellular parasite of bovine erythrocytes that causes a moderate to severe hemolitic anemia, jaundice and hemoglobinuria without hemoglobinemia (Kocan et al., 2003). *A. marginale* belongs to the phylum Proteobacteria, alpha Proteobacteria class, order Rickettsiales, Anaplasmataceae family.

Rhipicephalus microplus is considered the most important biological vector for *A. marginale* in tropical and subtropical regions of the world (de la Fuente et al., 2007). Since *R. microplus* eradication, tick transmission of *A. marginale* in the United States is mediated by *Dermacentor andersoni* and *Dermacentor variabilis* (Futse et al., 2003). Additionally, other hematophagous insects and the use of infected blood fomites could cause mechanical transmission (Kocan et al., 2003). The economic losses generated by this disease are not only associated with morbidity and mortality in cattle, but also with a lower weight gain rate, lower milk production, abortions and treatment costs. Among control measures to prevent severe morbidity and mortality due to anaplasmosis, *A. marginale*

* Corresponding author.
E-mail address: guillemi.eliana@inta.gob.ar (E.C. Guillemi).

http://dx.doi.org/10.1016/j.meegid.2014.12.027
1567-1348/

subsp. centrale as live attenuated vaccine is currently being used in tropical and subtropical countries in Africa Asia, Australia, and Latin America (Bock and De Vos, 2001).

It is relevant to achieve accurate methods for genotyping and characterizing strains as well as for studying the structure and dynamics of *A. marginale* populations in the field. Several methods and markers have been developed to characterize the genetic diversity of *A. marginale.* These are mostly focused on the Major Surface Proteins (MSPs) MSP4 and MSP1a (Almazán et al., 2008; de la Fuente et al., 2001, 2002, 2004, 2007; Lew et al., 2002; Mtshali et al., 2007; Palmer et al., 2001; Ruybal et al., 2009; Vidotto et al., 2006) and are useful for discriminating isolates. However, when the genetic population structure is under study, it is important to choose multiple loci that are selectively neutral. In fact, loci under positive selection may give a distorted view of the genetic population structure and its transmission dynamics, since selection rather than population history may determine the patterns of distribution of alleles within populations for these loci.

Previously published data (Ruybal et al., 2009) analyzing the *msp1a* marker refers to a wide genetic diversity of *A. marginale* population in Argentina. In this study, the authors demonstrated that the genetic population diversity was higher for tick-infested regions than for tick-free areas. Moreover, Estrada-Peña et al. (2009) studied the variability of MSP1a sequence worldwide and reported that this molecular marker is associated to the world ecological regions; therefore, the evolution of *A. marginale* may be linked to ecological traits affecting tick vector performance.

In 1998, Multilocus Sequence Typing (MLST) was first proposed for the characterization of isolates of the human pathogen *Neisseria meningitidis* (Maiden et al., 1998). This tool enables genotypic characterization of isolates and the study of the global dispersion of some new variants of pathogens (Mayer et al., 2002). In addition to these epidemiological studies of medical interest, the data obtained by MLST strategy apply to evolutionary and population studies (Jolley et al., 2000). In fact, it can be employed to estimate the frequency of recombination events and mutations and to investigate evolutionary relationships between organisms belonging to the same genus (Godoy et al., 2003).

We report here the development of the first MLST scheme for *A. marginale* and its application for population structure studies. In this scheme, 7 genes are employed for the discrimination of even very closely related strains. The design of the MLST scheme was assisted by the availability of the complete *A. marginale* genome. We have also developed a bioinformatic pipeline for the automated analysis of raw sequences and further diversity and phylogenetic analysis.

The MLST scheme developed in this work was applied for the study of 58 isolates from different world regions. Taking into account the results previously published by Ruybal et al. (2009) and Estrada-Peña et al. (2009), we hypothesized that geographically related isolates will tend to have a more similar genotype composition compared to the geographically distant isolates.

2. Materials and methods

2.1. Strains and genomic DNA isolation

A total of 58 *A. marginale* strains were analyzed. Five of them came from annotated genomes (Saint Maries, Florida, Mississippi, Puerto Rico and Virginia) and the other 53 were collected from countries in North and South America, Europe, and Africa (Table 1). Field samples were detected as positive for *A. marginale* by microscopic observation of Giemsa-stained blood smears and by PCR amplification of the *msp5* gene. Some of the field samples were from the same geographic region (Argentina provinces), even from the same ranch. Additionally, four of them came from outbreaks (Table 1). The genomic DNA extraction was performed by phenol/chloroform method and a standard ethanol precipitation (Sambrook et al., 1989) from PBS-washed and packed infected erythrocytes.

2.2. Target loci

The search was performed using genes applied for other MLST schemes in related microorganisms (Adakal et al., 2009), which have been previously described in the literature (Baldo et al., 2006; Jacobson et al., 2008; Vitorino et al., 2007) as reference. Those genes were single copy and encoded conserved proteins. Fourteen candidate genes were pre-selected. Primers were designed using Primer 3 program (http://frodo.wi.mit.edu/primer3/) for the first approach and then parameters were adjusted manually by IDT OligoAnalyzer tool from Integrated DNA Technologies (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/).

Specificity was initially evaluated *in silico* by BLASTn (http://blast.ncbi.nlm.nih.gov/) against *Bos taurus*, *Babesia bovis* and *B. bigemina* database, since DNA from these organisms can be found as contaminant in blood samples from field cattle. Later, specificity was corroborated experimentally using DNA from uninfected *B. taurus* and from *Babesia* species. Two reference strains from *A. marginale* (Mercedes and Salta) were amplified with candidate primers and only those PCR products with high sensitivity and specificity were finally selected.

Amplicons were sequenced from both strands and loci showing a neutral pattern were selected for further analysis. The degree of selection operating on the target genes was determined according to the ratio of mean non-synonymous substitutions per non-synonymous site/mean synonymous substitution per synonymous site (*dN*/*dS* ratio). The *dN*/*dS* ratio was calculated using the START2 program available from http://pubmlst.org/software/analysis/start2/ (Jolley et al., 2001).

Seven genes homogeneously distributed through the genome (Table 2 and Supplementary Fig. 1) were finally chosen: *dnaA* (DnaA chromosomal replication initiation protein; AM430), *ftsz* (cell division protein FtsZ; AM1261), *groEl* (Chaperonin GroEL; AM944), *lipA* (lipoyl synthase; AM820), *recA* (RecA recombination protein; AM085), *secY* (preprotein translocase subunit SecY; AM892) and *sucB* (dihydrolipoamide acetyltransferase component; AM1087).

2.3. PCR amplification and gene sequencing

The primers used to amplify and sequence the seven target genes are listed in Table 2. 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 of 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. Five microliters of each amplified product were analyzed by electrophoresis in 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 the amplified products 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 the 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