Veterinary Parasitology 167 (2010) 187–195

Contents lists available at ScienceDirect

Veterinary Parasitology

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

Mining the genetic diversity of *Ehrlichia ruminantium* using *map* genes family

Modestine Raliniaina [a,b,1], Damien F. Meyer [b,1,*], Valérie Pinarello [b], Christian Sheikboudou [b], Loic Emboulé [b], Yane Kandassamy [b], Hassane Adakal [d], Frédéric Stachurski [c,d], Dominique Martinez [c], Thierry Lefrançois [b], Nathalie Vachiéry [b,*]

[a] *FOFIFA, DRZV, Antananarivo, Madagascar*
[b] *CIRAD, UMR15 CIRAD/INRA, Site de Duclos, Prise d'Eau 97170 Petit-Bourg, Guadeloupe*
[c] *CIRAD, UMR15 CIRAD/INRA, Montpellier, France*
[d] *CIRDES, Bobo-Dioulasso, Burkina Faso*

ARTICLE INFO

Keywords:
Ehrlichia ruminantium
Heartwater
Major antigenic proteins (*map*)
Genetic diversity

ABSTRACT

Understanding bacterial genetic diversity is crucial to comprehend pathogenesis. *Ehrlichia ruminantium* (*E. ruminantium*), a tick-transmitted intracellular bacterial pathogen, causes heartwater disease in ruminants. This model rickettsia, whose genome has been recently sequenced, is restricted to neutrophils and reticulo-endothelial cells of its mammalian host and to the midgut and salivary glands of its vector tick. *E. ruminantium* harbors a multigene family encoding for 16 outer membrane proteins including MAP1, a major antigenic protein. All the 16 *map* paralogs are expressed in bovine endothelial cells and some are specifically translated in the tick or in the mammalian host.

In this study, we carried out phylogenetic analyses of *E. ruminantium* using sequences of 6 MAP proteins, MAP1, MAP1-2, MAP1-6, MAP1-5, MAP1+1 and MAP1-14, localized either in the center or at the borders of the *map* genes cluster.

We show that (i) *map1* gene is a good tool to characterize the genetic diversity among Africa, Caribbean islands and Madagascar strains including new emerging isolates of *E. ruminantium*; (ii) the different *map* paralogs define different genotypes showing divergent evolution; (iii) there is no correlation between all MAP genotypes and the geographic origins of the strains; (iv) The genetic diversity revealed by MAP proteins is conserved whatever is the scale of strains sampling (village, region, continent) and thus was not related to the different timing of strains introduction, *i.e.* continuous introduction of strains versus punctual introduction (Africa versus Caribbean islands).

These results provide therefore a significant advance towards the management of *E. ruminantium* diversity. The differential evolution of these paralogs suggests specific roles of these proteins in host–vector–pathogen interactions that could be crucial for developing broad-spectrum vaccines.

1. Introduction

Ehrlichia ruminantium, a member of the order Rickettsiales, is the etiologic agent of heartwater in ruminants including cattle, sheep, goats and a variety of wild ruminants (Peter et al., 2002). This wide host range might be broader with a recent report of three fatal human cases of suspected ehrlichiosis involving unrelated individuals

* Corresponding authors. Tel.: +590 590 25 54 42; fax: +590 590 94 03 96.
E-mail addresses: damien.meyer@cirad.fr (D.F. Meyer), nathalie.vachiery@cirad.fr (N. Vachiéry).
[1] These authors contributed equally to this work.

0304-4017/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.vetpar.2009.09.020

where *E. ruminantium* DNA was detected in patients (Allsopp et al., 2005).

E. ruminantium is transmitted by ticks of the genus *Amblyomma*, *Amblyomma variegatum* and *hebraeum* in Africa. This deadly Gram-negative intracellular α-proteobacterium is prevalent in the warmer and wetter parts of the world in both hemispheres such as sub-Saharian Africa, Caribbean Islands and in the Indian Ocean Islands of the Comoros, Madagascar, Mascarene, Reunion and Mauritius (Provost and Bezuidenhout, 1987). Heartwater is now also threatening livestock on the American mainland because of the presence of suitable indigenous vectors or adaptation of exotic ticks (Barré et al., 1987; Burridge et al., 2000, 2002).

E. ruminantium presents a broad genetic diversity (Allsopp et al., 1999, 2001; Martinez et al., 2004) which constitutes the main limitation for the development of an efficient vaccine. Moreover, little genetic information is available to discriminate the immunogenic variants of *E. ruminantium*. Up to now, genetic characterization was based preferentially on largely conserved genes such as *pCS20*, *groESL* (Allsopp et al., 2003). A better genetic characterization of representative *E. ruminantium* strains that circulate in the field, using other target genes including polymorphic genes, is necessary. To date, the genomes of three virulent strains of *E. ruminantium* strains have been published (Collins et al., 2005; Frutos et al., 2006, 2007) and three other virulent and attenuated strains are under sequencing by a French consortium Genoscope/CIRAD, making this bacterium a good model for genomic studies in Rickettsiales.

Studies on bacteria from the family *Anaplasmataceae* revealed an important role for the bacterial outer membrane proteins (OMPs) in the stimulation of the host immune response and protection of the host from infection (Lopez et al., 2005; Ohashi et al., 1998). The outer membrane protein MAP1 (Major Antigenic Protein 1) triggers a dominant antibody response of animals infected with *E. ruminantium* (van Vliet et al., 1994). The locus containing *map1* has been characterized and consists of 16 *map1* paralogs, defining thus the *map* genes cluster of *E. ruminantium* (van Heerden et al., 2004). The expression and translation of the *map* genes was studied both in host and tick cells (Bekker et al., 2005; Postigo et al., 2007).

The aim of this study was to characterize the genetic diversity of *E. ruminantium* strains from Caribbean islands, Africa and Madagascar (Indian Ocean island) in order to evaluate the importance of *E. ruminantium* strains diversity in various areas where pathogen introduction and mix were different. Africa is thought to be the original continent of *E. ruminantium* and the birthplace of its diversity (Uilenberg et al., 1984). In Guadeloupe, *E. ruminantium* was introduced from Africa with livestock during the slavery period more than 200 years ago (Uilenberg et al., 1984). None new introduction has been observed since 150 years at least. Madagascar represents an intermediate situation between Africa and Guadeloupe, since heartwater could have been introduced during the migrations of domestic livestock around 1200 years ago and new strains of *E. ruminantium* could have been introduced by recent importations of ameliorated breeds from Africa.

In terms of molecular phylogeny, three major questions arise: (1) Is there a similar or different evolution of the *map* multigene family? (2) Is the genetic diversity of *E. ruminantium* strains similar at the village scale, between regions and between continents? (3) Does a phylogenetic analysis based on several polymorphic genes such as *map* genes allow the identification of geographic or antigenic groups?

The subsequent question concerning evolutionary genetics is whether differences in genetic diversity of *E. ruminantium* could be associated with the various history of pathogen introduction.

In this context, this study presents the phylogenetic analyses of *E. ruminantium* we performed using the *map1* multigenic family that comprise 16 paralogs (Bekker et al., 2005). To achieve our goal, we analyzed amino acid sequences of 6 MAP proteins, MAP1, MAP1-2, MAP1-6, MAP1-5, MAP1+1 and MAP1-14, localized along the *map* genes cluster (Bekker et al., 2005).

2. Materials and methods

2.1. Bacterial strains and growth conditions

E. ruminantium isolates used in this study, their source and their geographic origins are presented in Table 1. Our reference strain Gardel of *E. ruminantium* (Erga) was isolated on the island of Guadeloupe in 1982 from a goat injected with a homogenate of a female *A. variegatum* tick collected from cows (Uilenberg et al., 1985). For cultivated strains, *E. ruminantium* was multiplied successively in bovine umbilical endothelial cells and in bovine aorta endothelial cells grown in Glasgow minimal essential medium complemented with fetal calf serum, tryptose-phosphate broth, and antibiotics (Bezuidenhout et al., 1985) at 37 °C, 5% CO_2, with a weekly passage on fresh cells (Martinez et al., 1990). DNA was extracted as described below.

2.2. DNA extraction, PCR amplifications and sequencing

The QIAamp DNA mini kit (QIAGEN GmbH, Hilden, Germany) was used to extract genomic DNA from blood, brain, tick cells or cell culture derived *E. ruminantium* isolates according to manufacturer's instructions. Elementary bodies were purified from culture supernatant, as previously described (Martinez et al., 1994), resuspended in 350 μl of phosphate-buffered saline containing 0.36 μg/ml of DNase to remove contaminating host cell DNA, and incubated for 90 min at 37 °C prior to the addition of 25 mM EDTA (Martinez et al., 1990). Extraction of DNA from elementary bodies was done as previously described (Perez et al., 1997).

Primer pairs for amplification of *map1*, *map1+1*, *map1-2*, *map1-5*, *map1-6* and *map1-14* gene coding regions were designed using published sequences. Primer sequences, melting and annealing temperatures used in this study are given in Table 2. A nested PCR was used for *map1* gene as described in Martinez et al. (2004) using modified primers listed in Table 2. The second phase was done using 1 μl of PCR product from the first phase as matrix. Amplification reactions (25 μl) contained 1 μl of each primer (50 ng/μl), 2 μl of 2.5 mM dNTP mix, 1.5 μl of 50 mM $MgCl_2$, 2.5 μl of 10× proprietary buffer supplied with the enzyme, 0.1 μl of

Table 1
E. ruminantium isolates used in this study for *map1*, *map1+1*, *map1-2*, *map1-5*, *map1-6* and *map1-14* genes amplification.

Isolate	Source	Geographic origin	*map1*	*map1-2*	*map1-5*	*map1-6*	*map1+1*	*map1-14*
Caribbean, Car								
AntGeorges	Tick	Antigua	+					
Antigua	Reference	Antigua	+					
BlondeCp7	Cell culture p7	Guadeloupe	+		+	+		+
Gardel	Reference	Guadeloupe	+	+	+	+	+	+
Gardelp230	Cell culture p230	Guadeloupe	+					
Gu509797M1	Tick	Guadeloupe	+					
Gu537782F1	Tick	Guadeloupe	+					
Gu5697Cer	Brain	Guadeloupe	+					
Gu6001F1	Tick	Guadeloupe	+	+	+	+	+	+
Gu665394M1	Tick	Guadeloupe	+					
GuBM1T	Tick	Guadeloupe	+					
GuBM2T	Tick	Guadeloupe	+					
GuBM601T	Tick	Guadeloupe	+					
Mg01503F1	Tick	Marie-Galante	+					
Mg041530F2	Tick	Marie-Galante	+					
Mg0901M1	Tick	Marie-Galante	+					
Mg12861M2	Tick	Marie-Galante	+					
Mg3524106F1	Tick	Marie-Galante	+					
Mg4410M1	Tick	Marie-Galante	+					
Mg83834562	Tick	Marie-Galante	+					
Western Africa, WA								
Ban033F1	Tick	Banankélédaga, Burkina Faso	+					
Ban112Cp7	Cell culture p7	Banankélédaga, Burkina Faso	+	+	+	+	+	+
Ban181T	Tick	Banankélédaga, Burkina Faso	+					
Ban455Cer	Brain	Banankélédaga, Burkina Faso	+					
Bank421Cer	Tick	Banankélédaga, Burkina Faso	+		+	+		
Bek070T1	Tick	Bekuy, Burkina Faso	+					
Bek242T	Tick	Bekuy, Burkina Faso	+					
Bek255Cp5	Cell culture p5	Bekuy, Burkina Faso	+	+	+	+	+	+
Bek282Cer	Brain	Bekuy, Burkina Faso	+					
Bek313Cer	Brain	Bekuy, Burkina Faso	+					
Bek320T	Tick	Bekuy, Burkina Faso	+					
Bek448FC2	Tick*	Bekuy, Burkina Faso	+					
Bek448FC3	Tick*	Bekuy, Burkina Faso	+					
Bek448FC4	Tick*	Bekuy, Burkina Faso	+					
Bek448MC3	Tick*	Bekuy, Burkina Faso	+					
Bek448MC4	Tick*	Bekuy, Burkina Faso	+					
Bek475T	Tick	Bekuy, Burkina Faso	+					
Burkinafa	Reference	Banankélédaga, Burkina Faso	+					
CamerCp9	Reference and cell culture p9	Cameroun	+	+	+	+	+	+
Lamb065SCer	Brain	Lamba, Burkina Faso	+					
Lamb076T	Tick	Lamba, Burkina Faso	+					
Lamb107T	Tick	Lamba, Burkina Faso	+					
Lamb194bCer	Brain	Lamba, Burkina Faso	+					
Lamb229T	Tick	Lamba, Burkina Faso	+					
Lamb350Cer	Brain	Lamba, Burkina Faso	+					
Lamb465T	Tick	Lamba, Burkina Faso	+					
Lamb479Cp14	Tick and cell culture p14	Lamba, Burkina Faso	+	+	+	+	+	+
Mali	Reference	Mali						
Pokoase	Reference	Ghana	+					
Sara188T2	Tick	Sara, Burkina Faso	+					
Sara292F1	Tick	Sara, Burkina Faso	+					
Sara371T	Tick	Sara, Burkina Faso	+					
Sara409Cer	Brain	Sara, Burkina Faso	+					
Sara445Cer	Tick	Sara, Burkina Faso	+		+	+	+	+
Sara446Cer	Brain	Sara, Burkina Faso	+					
Sara450T	Tick	Sara, Burkina Faso	+					
SenDiaksoa	Tick	Diaksao, Senegal	+					
SenDioumb	Tick	Diambalo, Senegal	+					
Senegal	Reference	Senegal	+	+	+	+	+	+
SenKDB	Tick	Keur Mbir Dao, Senegal	+					
SenM10T	Tick	Niagues, Senegal	+					
SenM12T	Tick	Niagues, Senegal	+					
SenM16T	Tick	Niagues, Senegal	+					
SenM1T	Tick	Niagues, Senegal	+					

Table 1 (*Continued*)

Isolate	Source	Geographic origin	*map1*	*map1-2*	*map1-5*	*map1-6*	*map1+1*	*map1-14*
Southern Africa, SA								
Ball3	Reference and blood	South Africa	+					
Blaaukrans	Reference	South Africa	+					
Ebostawana	Reference	Bostwana	+					
Ludlow	Reference	South Africa	+					
LutaleCp2	Reference and cell culture p2	Zambia	+	+	+	+	+	+
Mara87_7	Reference	South Africa	+					
MaraCp1	Tissue culture p1	South Africa	+	+	+	+	+	+
Kwanyanga	Reference	South Africa	+					
Omatjenne	Reference	Namibia	+					
UmpalaCp2	Reference and cell culture p2	Mozambique	+	+	+	+	+	+
Welgevonden	Reference	South Africa	+	+	+	+	+	+
Eastern Africa, EA								
Umbanein	Reference	Sudan	+					
Kiswani	Reference	Kenya	+					
Indian Ocean, IO								
MadKJSF2	Tick	Kiandjasoa, Madagascar	+					
MadPierP1T	Tick	Ampitilina, Madagascar	+					
MadRZFM1	Tick	Andranonahoatra, Madagascar	+					

Reference strains are shaded in gray. (+) The strain was amplified for the corresponding *map* gene; (p) passage; (*) clones of *map1* PCR product isolated from a tick Bekuy 448 containing a mix of *E. ruminantium* strains. Geographic notations: WA, Western Africa; EA, Eastern Africa; SA, Southern Africa; Car, Caribbean (West Indies).

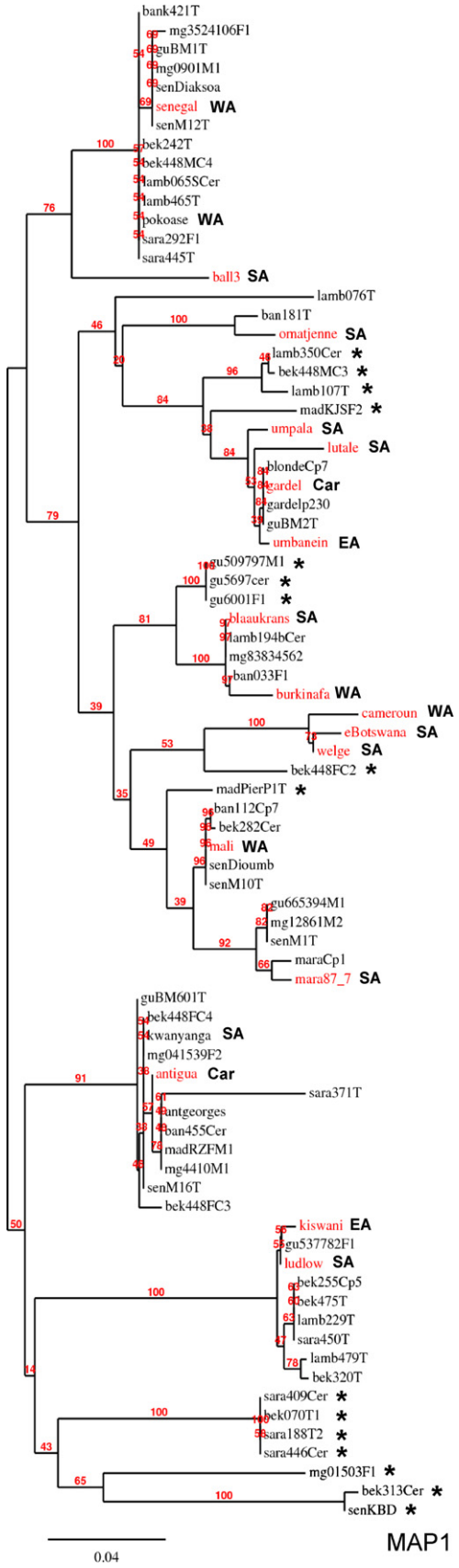
5 U/μl Taq polymerase (EurobioTaq ADN polymerase, Eurobio) and 1 μl of DNA. Reactions conditions were as follows: initial denaturation 3 min at 95 °C, followed by 35 cycles of 50 s denaturation at 95 °C; 50 s annealing at the appropriate temperature; 50 s extension at 72 °C, with a final extension 10 min at 72 °C; hold at 4 °C. All reactions were performed in an icycler thermal cycler (BioRad). PCR products were visualized after electrophoresis on 1.5 % agarose gels. Single amplicons of the expected size around 700–800 bp were purified and automatically sequenced (200 ng of DNA) using the gold standard ABI 3730*xl* DNA Analyzer platform (Applied Biosystems) by COGENICS corporation. DNA and protein sequences were analyzed using the Vector NTI package. Homology searches were performed using the NCBI BLAST server.

2.3. Phylogenetic analyses

Amino acid sequences for MAP1, MAP1+1, MAP1-2, MAP1-5, MAP1-6 and MAP1-14 were aligned and phylogenetic trees were reconstructed by the neighbor-joining method. The Phylogeny.fr platform was used to carry out the multiple alignments of amino acid sequences, the phylogenetic reconstruction and the graphical representation of the inferred tree (Dereeper et al., 2008). The 'A la Carte' mode which offers more flexibility and sophistication allowed us to select and set up all the required steps for our analysis. The MUSCLE program (Edgar, 2004a,b) was used for multiple alignments of the nucleotide sequences. The unmatched ends were deleted to obtain a homogeneous matrix of characters and thus increase the reliability of the tree

Table 2
Primers used for the amplification of *map1*, *map1+1*, *map1-2*, *map1-5*, *map1-6* and *map1-14* genes.

Gene	Primer	Primer sequence	Annealing temperature (°C)
map1			
First phase	NT	5′-ATGAATTGCAAGAAAATTTTTATAACAAG-3′	50
	CT1bis	5′-TTAAAATACAAACCTTCCTCCAATTTC-3′	
Second phase	LP	5′-CCTGGTGTGTCCTTTTCTGA-3′	55
	CT2bis	5′-CCTTCCTCCAATTTCTATACCGAAATGGC-3′	
map1+1	mapX11F2	5′-GGTGTTACATTAAGTGCATTGTTG-3′	40
	EcoMapX11GR*	5′-(GGAATTC)CTATAAGAAAAATCTAATCCCAAC-3′	
map1-2	map1-2F2	5′-CTCATTGCAACAGGTATAGTATTAA-3′	45
	KpnMap1-2GR*	5′-(GGGGTACC)TTATAACTTTAATCCAAATCTTATTCC-3′	
map1-5	mapX8F2	5′-GGTAGTGAAATTGGATTAAG-3′	48
	mapX8R2	5′-CCTCATTTAATTGCAATTTATAGTC-3′	
map1-6	mapX7F2	5′-CAGTTAACATATTTATAGGAGC-3′	45
	mapX7R2	5′-CTTAATCCAATTTCACTACC-3′	
map1-14	mapX12F2	5′-TGGCACATAATGCATTTTCATC-3′	42
	mapX12R2	5′-GTAAGCAAAAGCTTAAACTC-3′	



obtained. Sequence alignments as obtained by MUSCLE alignment program were given to PROTDIST program in the Phylip format. The phylogenies were constructed and the trees built by the BioNJ method (Gascuel, 1997) with the Jones–Taylor–Thornton (JTT) distance measure matrix (Jones et al., 1992) as implemented in the PHYLIP PROTDIST program (Felsenstein, 1989). PROTDIST computed a distance matrix from a sequence alignment using the JTT model for amino acids substitution. This matrix is an empirical one that scales probabilities of change from one amino acid to another, assuming that the total change between the two amino acid sequences is 1%. The robustness of the tree's topology and the reliability of each clade were assessed by bootstrap analysis of the neighbor-joining method using SEQBOOT, PROTDIST and NEIGHBOR programs based on 1000 repeated samplings. Finally, the bootstrap trees were extracted using the CONSENSE program and the resulting tree can be drawn using NJPLOT program (Perriere and Gouy, 1996; Persson, 2000). The bootstrap procedure is a sound and accurate way to obtain robust phylogenetic trees, and its use is greatly facilitated by the speed of distance methods.

2.4. Determination of clusters of genotypes

For each MAP protein, *E. ruminantium* strains were clustered when they were supported by a minimal bootstrap value of 60, thus defining various genotypes.

3. Results

We completed PCR amplification and sequencing of target genes for strains from African and Caribbean areas in order to generate phylogenetic trees with the neighbor-joining method. The comparison of whole-protein amino acids sequences, using alignments generated by MUSCLE (Edgar, 2004a,b), showed that these proteins are well conserved between each other and among *E. ruminantium* strains.

3.1. Genetic characterization of E. ruminantium using map1 gene

The phylogenetic tree generated with MAP1 sequences of 80 strains of *E. ruminantium* clearly shows that *E. ruminantium* strains can be subdivided into several major groups that do not correspond to their geographic origin. (Fig. 1; Table 1). We observed the same MAP1 genotypes for isolates from Guadeloupe, Burkina Faso, Senegal, *e.g.* the MAP1 genotype Antigua. The MAP1 tree topology obtained by BioNJ analysis is in accordance with Maximum Likelihood phylogenetic tree previously obtained with

Fig. 1. Neighbor-joining (NJ) tree with bootstrap values of MAP1 sequences from 80 strains of *E. ruminantium*. Phylogenies were created by neighbor-joining analysis (BIONJ), using distance matrix (PROTDIST) with JTT substitution model. Bootstrap values over 50 based on 1000 replicates are considered for each node and the branch length index is indicated below the phylogeny (0.04 substitutions/site). Notations indicate geographical origins: WA, Western Africa; EA, Eastern Africa; SA, Southern Africa; and Car, Caribbean (French Antilles). New isolates that do not correspond to any known genotype are indicated by an asterisk.

fewer strains (Allsopp et al., 2001). Thus, we found the same clades of reference strains as follows for isolates from different geographical areas: Senegal/Pokoase, Omatjenne, Gardel/Lutale/Umbaneim/Umpala, Burkina Faso/Blaaukrans, Cameroun/E. Bostwana/Welgevonden, Mali/Mara, Antigua/Kwanyanga and Kiswani/Ludlow (Fig. 1). Some strains formed a unique group (without a reference strain), *e.g.* Guadeloupe strains Gu509797M1, Gu5697Cer and Gu6001F1 or two out of three strains from Madagascar island, MadKJSF2 and MadPierP1T. We observed several strains that belong to the same clade, which originate from different areas, *e.g.* SenKBD and Bek313 from Senegal and Burkina Faso (Bekuy village) or Sara409Cer, Bek070T1, Sara188T2, Sara446Cer from Sara and Bekuy villages. The very high bootstrap value indicates that this clustering is not due to a random phenomenon. These new strains correspond to new genotypes when analyzed by RFLP using *map1* gene (data not shown).

We then characterized the polymorphisms of these new genotypes in comparison with known MAP1 genotypes in order to point out the generation of the diversity. MAP1 amino acids sequences present three hypervariable regions, each about 10–15 amino acids long (Allsopp et al., 2001). The polymorphism between each group of unknown strains and the closest reference strains is mainly due to amino acid substitutions or deletions in the first polymorphic region whereas the polymorphism inside these groups of unknown strains is localized all along the protein (data not shown).

3.2. Genetic characterization of *E. ruminantium* using other map genes

The MAP proteins were used to characterize 11 strains for MAP1-2 (Fig. 2A), 15 strains for MAP1-5 (Fig. 2B), 14 strains for MAP1-6 (Fig. 2C), 12 strains for MAP1+1 (Fig. 2D), and 13 strains for MAP1-14 (Fig. 2E). The first

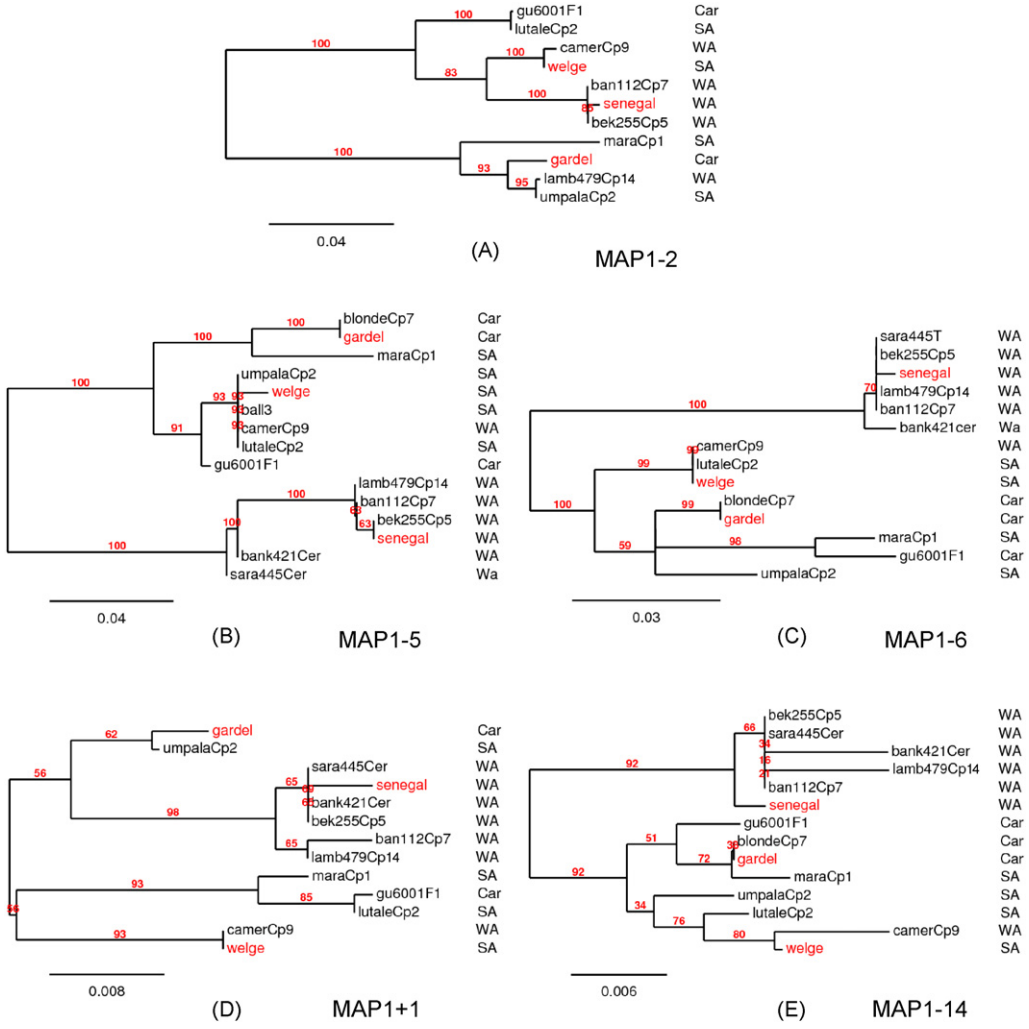


Fig. 2. Neighbor joining (NJ) trees with bootstrap values of MAP sequences from strains of *E. ruminantium*. Phylogenies were created by Neighbor joining analysis (BIONJ), using distance matrix (PROTDIST) with JTT substitution model. Bootstrap values over 50 based on 1000 replicates are represented for each node and the branch length index is indicated below the phylogeny (number of substitutions/site). (A) MAP1-2 sequences from 11 strains; (B) MAP1-5 sequences from 15 strains; (C) MAP1-6 sequences from 14 strains; (D) MAP1+1 sequences from 12 strains; (E) MAP1-14 sequences from 13 strains. Notations indicate geographical origins: WA, Western Africa; SA, Southern Africa; Car, Caribbean (French Antilles).

result to note for the phylogenic trees of MAP1+1, MAP1-2, MAP1-5, MAP1-6 and MAP1-14 amino acid sequences is that their topologies are clearly different (Fig. 2). Indeed, one single strain could share different genotypes depending on the selected MAP. This was observed for Mara, Guad6001, Lamba479, Lutale and Umpala. This result demonstrates that *map* genes are hypervariable and that there is no conservation of one given strain genotype among *map* paralogs.

All the strains of the Senegal genotype defined by the MAP1 phylogenetic tree have also the Senegal genotype when analyzed with MAP1+1, MAP1-2, MAP1-5, MAP1-6 and MAP1-14 and are clustered together (Fig. 2). All these strains originated from Western Africa. Inversely, Ban 112Cp7 and Bekuy 255Cp5 which were respectively Mali and Kiswani MAP1 genotypes belong to Senegal genotype for other *MAP* genes.

With the exception of MAP1-2 tree, the Senegal cluster shows a greater genetic distance when compared to the Gardel or Welgevonden cluster indicating a relative independence in the evolution of the corresponding strains. Moreover, Welgevonden and Gardel clusters are more closely related to each other.

Strains originated from Caribbean were often grouped with strains from Southern Africa, especially for trees based on MAP1-5, MAP1-6 and MAP1+1 sequences.

We identified new genotypes, of which the corresponding strains did not group with any reference strains: Gu6001F1/LutaleCp2 with MAP1-2 sequences (Fig. 2A), Gu6001F1 with MAP1-5 (Fig. 2B), Umpala with MAP1-6 (Fig. 2C) and Lamb479Cp14/BlondeCp7 with MAP1-14 sequences (Fig. 2E).

4. Discussion

E. ruminantium has a *map1* multigene family that consists of 16 copies of homologous genes encoding 28–30 kDa outer membrane proteins. The *map1*-related genes are highly diversified among strains of *E. ruminantium* (Yu et al., 2007). We provide evidences that *map* paralogs have different discriminatory abilities for genotyping of *E. ruminantium* strains, due to their high genetic variability. Thus, for one given strain, each MAP protein could define a different genotype, whatever the position of the *map* paralogs inside the *map* cluster. However, MAP1 sequences have a better resolution than other MAP sequences and therefore constitute a gold marker to characterize the genetic diversity of *E. ruminantium*. A limited sequence of MAP1 (around 750 pb) was sufficient to type *E. ruminantium* strains accurately. The tendency is then to characterize genetic diversity using polymorphic markers such as MAP1 rather than other genes previously used such as *pCS20* and *groESL* genes that show very little variation among *E. ruminantium* isolates (Allsopp et al., 2003). Moreover, our data also indicate that *map1* gene is a robust tool to identify new emerging isolates of *E. ruminantium*. The origin of the Caribbean strains of *E. ruminantium* could be further assessed using MLST approach.

The genetic diversity described using this MAP sequences phylogenetic analysis is conserved independently of the geographic scale considered. Particularly, it is possible to find different MAP1 genotypes in the same village, and the same genotypes could also be found at the region and continent level. Beside these results we were not able to find a significant correlation between MAP genotypes and the geographic origins of the strains. This absence of link between geographical origin and *map1* genotypes reinforce previous data obtained with a smaller sampling of *E. ruminantium* strains (Allsopp et al., 2001; Martinez et al., 2004).

Using MAP markers, our results demonstrated that there was no influence of the timing of introduction of *E. ruminantium* strains (continuous or punctual introduction) on the evolution of *map* multigene family. Same genotypes are present in Africa and in the Caribbean islands where strains introduction was punctual and stopped more than 150 year ago. This could be due to the neutral selective pressure of the target MAP proteins. For Madagascar strains, 2 out of 3 isolates had new MAP1 genotypes. This could be due to a divergent evolution of Madagascar strains compared to African strains. However, further genetic characterization on larger *E. ruminantium* strains from Madagascar and other Indian Ocean islands is necessary.

The genetic variability among *map* paralogs is surprising for a multigene family that is organized. This could be related to the function of the different MAP proteins. Previous work has shown that *map* genes of *E. ruminantium* are organized in a particular operon in which genes are differentially regulated, depending on host and vector cell environment (Bekker et al., 2005). For instance, *map1-1* and *map1+1* genes are expressed in vectoring ticks, non-vectoring ticks and also in bovine endothelial cells whereas two other paralogs (*map1-11* and *map1-3*) were never found to be transcribed in Gardel and Welgevonden strains (Bekker et al., 2005). Furthermore, four *map* paralogs are transcribed in non-vector ticks and endothelial cells but not in vectoring ticks (Bekker et al., 2005), and out of the 16 *map* paralogs examined in another transcription analysis, only transcripts from *map1* and *map1-1* were detected in ticks (Postigo et al., 2007). At the protein levels similar observations have been made of differential expression of MAP1 family proteins *in vitro* in mammalian cells and tick cell cultures (Postigo et al., 2008). Differential expression could be associated with different selection pressure (evolutionary forces) leading to genetic variability among the MAP family members.

Comparative analyses of mutational trends between Gardel strain and Welgevonden strain of *E. ruminantium* revealed that only *map1* and *map1-13* genes display significant selective pressure toward synonymous substitutions (Frutos et al., 2006). These data indicate that *map* paralogs should be conserved in their original form to maintain the proteins function. Overall, it is very likely that the *map* genes cluster could be the result of gene duplications from an ancestral gene where copies could have been maintained and are under divergence driven by progressive accumulation of point mutations, according to Ohno's model (Bershtein and Tawfik, 2008; Ohno, 1970). The homologous proteins created in this way by duplication and divergence form a family whose members can

have similar functions or not. Thus each MAP protein could have a peculiar role during the interaction of *E. ruminantium* with its vector or its host. This is in accordance with the permanent ongoing process of genome plasticity shown for *E. ruminantium* (Frutos et al., 2006). This is also supported by numerous theoretical, computational, and experimental works that indicate that mutations of adaptive potential can occur under "neutrality" (Amitai et al., 2007; Nei, 2007)—namely, under selection to maintain the original function and structure. Phenotypic changes under neutrality support the notion that neutral sequence changes (non-adaptative) in a current context could facilitate adaptation under changing conditions, by both providing an immediate advantage, and reducing the number of mutations required for the divergence of new functions (Amitai et al., 2007).

The understanding of the evolutionary mechanisms of such multigene families could have an obvious implication in the adaptation of *E. ruminantium* to its vector tick and/or to its ruminant host. Therefore, extensive biodiversity and phylogeography analyses would help to understand the genetic and geographic structure of *E. ruminantium*.

Conflict of interest statement

None.

Acknowledgement

The research described in this manuscript was supported by the European Union grant Inco-Dev #FP6-003713: "Epigenevac".

References

Allsopp, M.T., Dorfling, C.M., Maillard, J.C., Bensaid, A., Haydon, D.T., van Heerden, H., Allsopp, B.A., 2001. *Ehrlichia ruminantium* major antigenic protein gene (*map1*) variants are not geographically constrained and show no evidence of having evolved under positive selection pressure. J. Clin. Microbiol. 39, 4200–4203.

Allsopp, M.T., Hattingh, C.M., Vogel, S.W., Allsopp, B.A., 1999. Evaluation of 16S, map1 and pCS20 probes for detection of *Cowdria* and *Ehrlichia* species. Epidemiol. Infect. 122, 323–328.

Allsopp, M.T., Van Heerden, H., Steyn, H.C., Allsopp, B.A., 2003. Phylogenetic relationships among *Ehrlichia ruminantium* isolates. Ann. N. Y. Acad. Sci. 990, 685–691.

Allsopp, M.T.E.P., Louw, M., Meyer, E.C., 2005. *Ehrlichia ruminantium*: An emerging human pathogen? Ann. N. Y. Acad. Sci. 1063, 358–360.

Amitai, G., Gupta, R.D., Tawfik, D.S., 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. HFSP J. 1, 67–78.

Barré, N., Uilenberg, G., Morel, P.C., Camus, E., 1987. Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic *Amblyomma* ticks. Onderstepoort J. Vet. Res. 54, 405–417.

Bekker, C.P., Postigo, M., Taoufik, A., Bell-Sakyi, L., Ferraz, C., Martinez, D., Jongejan, F., 2005. Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. J. Bacteriol. 187, 4782–4791.

Bershtein, S., Tawfik, D.S., 2008. Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol. Biol. Evol. 25, 2311–2318.

Bezuidenhout, J.D., Paterson, C.L., Barnard, B.J., 1985. In vitro cultivation of *Cowdria ruminantium*. Onderstepoort J. Vet. Res. 52, 113–120.

Burridge, M.J., Simmons, L.A., Allan, S.A., 2000. Introduction of potential heartwater vectors and other exotic ticks into Florida on imported reptiles. J. Parasitol. 86, 700–704.

Burridge, M.J., Simmons, L.A., Peter, T.F., Mahan, S.M., 2002. Increasing risks of introduction of heartwater onto the American mainland associated with animal movements. Ann. N. Y. Acad. Sci. 969, 269–274.

Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., Faber, F.E., van Heerden, H., Josemans, A., van Kleef, M., Steyn, H.C., van Strijp, M.F., Zweygarth, E., Jongejan, F., Maillard, J.C., Berthier, D., Botha, M., Joubert, F., Corton, C.H., Thomson, N.R., Allsopp, M.T., Allsopp, B.A., 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. Proc. Natl. Acad. Sci. U.S.A. 102, 838–843.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucl. Acids Res. 36, W465–469.

Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113.

Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Felsenstein, J., 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Frutos, R., Viari, A., Ferraz, C., Morgat, A., Eychenie, S., Kandassamy, Y., Chantal, I., Bensaid, A., Coissac, E., Vachiery, N., Demaille, J., Martinez, D., 2006. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. J. Bacteriol. 188, 2533–2542.

Frutos, R., Viari, A., Vachiery, N., Boyer, F., Martinez, D., 2007. *Ehrlichia ruminantium*: genomic and evolutionary features. Trends Parasitol. 23, 414–419.

Gascuel, O., 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14, 685–695.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Lopez, J.E., Siems, W.F., Palmer, G.H., Brayton, K.A., McGuire, T.C., Norimine, J., Brown, W.C., 2005. Identification of novel antigenic proteins in a complex *Anaplasma marginale* outer membrane immunogen by mass spectrometry and genomic mapping. Infect. Immun. 73, 8109–8118.

Martinez, D., Maillard, J.C., Coisne, S., Sheikboudou, C., Bensaid, A., 1994. Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. Vet. Immunol. Immunopathol. 41, 153–163.

Martinez, D., Swinkels, J., Camus, E., Jongejan, F., 1990. Comparison between 3 antigens for the serodiagnosis of heartwater disease by indirect immunofluorescence. Rev. Elev. Med. Vet. Pays. Trop. 43, 159–166.

Martinez, D., Vachiery, N., Stachurski, F., Kandassamy, Y., Raliniaina, M., Aprelon, R., Gueye, A., 2004. Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. Ann. N. Y. Acad. Sci. 1026, 106–113.

Nei, M., 2007. The new mutation theory of phenotypic evolution. Proc. Natl. Acad. Sci. U.S.A. 104, 12235–12242.

Ohashi, N., Zhi, N., Zhang, Y., Rikihisa, Y., 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. Infect. Immun. 66, 132–139.

Ohno, S., 1970. Evolution by Gene Duplication. Allen & Unwin, London.

Perez, J.M., Martinez, D., Debus, A., Sheikboudou, C., Bensaid, A., 1997. Detection of genomic polymorphisms among isolates of the intracellular bacterium *Cowdria ruminantium* by random amplified polymorphic DNA and Southern blotting. FEMS Microbiol. Lett. 154, 73–79.

Perriere, G., Gouy, M., 1996. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78, 364–369.

Persson, B., 2000. Bioinformatics in protein analysis. EXS 88, 215–231.

Peter, T.F., Burridge, M.J., Mahan, S.M., 2002. *Ehrlichia ruminantium* infection (heartwater) in wild animals. Trends Parasitol. 18, 214–218.

Postigo, M., Taoufik, A., Bell-Sakyi, L., Bekker, C.P., de Vries, E., Morrison, W.I., Jongejan, F., 2008. Host cell-specific protein expression in vitro in *Ehrlichia ruminantium*. Vet. Microbiol. 128, 136–147.

Postigo, M., Taoufik, A., Bell-Sakyi, L., de Vries, E., Morrison, W.I., Jongejan, F., 2007. Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. Vet. Microbiol. 122, 298–305.

Provost, A., Bezuidenhout, J.D., 1987. The historical background and global importance of heartwater. Onderstepoort J. Vet. Res. 54, 165–169.

Uilenberg, G., Barre, N., Camus, E., Burridge, M.J., Garris, G.I., 1984. Heartwater in the Caribbean. Prevent. Vet. Med. 2, 255–267.

Uilenberg, G., Camus, E., Barre, N., 1985. A strain of *Cowdria ruminantium* isolated in Guadeloupe (French West Indies). Rev. Elev. Med. Vet. Pays Trop. 38, 34–42.

van Heerden, H., Collins, N.E., Brayton, K.A., Rademeyer, C., Allsopp, B.A., 2004. Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium*. Gene 330, 159–168.

van Vliet, A.H., Jongejan, F., van Kleef, M., van der Zeijst, B.A., 1994. Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. Infect. Immun. 62, 1451–1456.

Yu, X.J., McBride, J.W., Walker, D.H., 2007. Restriction and expansion of *Ehrlichia* strain diversity. Vet. Parasitol. 143, 337–346.