Journal of Microbiological Methods 88 (2012) 205–211

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

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

A new typing technique for the Rickettsiales *Ehrlichia ruminantium*: Multiple-locus variable number tandem repeat analysis

Héloïse Pilet [a,1], Nathalie Vachiéry [b], Moez Berrich [a], Rim Bouchouicha [a,2], Benoît Durand [c], Ludovic Pruneau [b], Valérie Pinarello [b], Angélique Saldana [b], Catherine Carasco-Lacombe [b], Thierry Lefrançois [b], Damien F. Meyer [b], Dominique Martinez [b], Henri-Jean Boulouis [a], Nadia Haddad [a,*]

[a] UPE, Ecole nationale vétérinaire d'Alfort, UMR BIPAR, ENVA, ANSES, UPEC, USC INRA, 23, rue du Gl de Gaulle - 94703 Maisons-Alfort, France
[b] CIRAD, UMR CMAEE, F- 97170 Petit-Bourg, Guadeloupe, France
[c] ANSES /Laboratoire de santé animale, Unité d'épidémiologie, 23, rue du Gl de Gaulle - 94703 Maisons-Alfort, France

ARTICLE INFO

Article history:
Received 5 October 2011
Received in revised form 18 November 2011
Accepted 18 November 2011
Available online 29 November 2011

Keywords:
Molecular typing
Epidemiology
MLVA
VNTR
Ehrlichia ruminantium

ABSTRACT

Ehrlichia ruminantium (*ER*) is a member of the order Rickettsiales transmitted by *Amblyomma* ticks. This obligatory intracellular bacterium is the causative agent of a fatal disease in ruminants, named heartwater. It represents a constraint on breeding development in sub-Saharan Africa and in the Caribbean. The genetic diversity of the strains of *ER*, which could be a limiting factor to obtain effective vaccines, needs to be better characterized. For this purpose, we developed a molecular typing technique based on the polymorphism of variable number tandem repeat (VNTR) sequences, MLVA (multiple locus VNTR analysis).

Eight (out of 21) VNTR candidates were validated using 17 samples representing a panel of *ER* strains from different geographical origins from West, South Africa, and Caribbean areas and in *ER* infected ticks and goat tissues. This result demonstrated the ability of these VNTRs to type a wide range of strains. The stability of the selected VNTR markers was very good, at the time scale needed for epidemiological purposes: in particular, no difference in the VNTR profiles was observed between virulent and attenuated strains (for Gardel and Senegal strains) and between strains (Gardel and Blonde strains) isolated in the same area 19 years apart. We validated the strong discriminatory power of MLVA for *ER* and found a high level of polymorphism between the available strains, with 10 different profiles out of 13 *ER* strains.

The MLVA scheme described in this study is a rapid and efficient molecular typing tool for *ER*, which allows rapid and direct typing of this intracellular pathogen without preliminary culture and gives reliable results that can be used for further epidemiological studies.

1. Introduction

Ehrlichia ruminantium (*ER*) is the causative agent of heartwater, a major tick-borne disease of livestock in Africa that has expanded to many regions where competent species of *Amblyomma* vectors are present (Allsopp, 2010). *A. variegatum* is the most widely distributed in Africa. *ER* and its vector *A. variegatum* were introduced into the Caribbean between 1733 and 1830 with cattle brought from West Africa (Maillard and Maillard, 1998; Molia et al., 2008; Uilenberg et al., 1984). *ER* is now susceptible to invade the American mainland (Barre et al., 1987). The economic impact of this disease in terms of mortality, cost of treatment with acaricides, cost of vaccination, reduced livestock productivity and cattle exchange prohibition is important. Recently, *ER* DNA has been associated with fatal diseases in three human patients (Allsopp et al., 2005), suggesting that this bacterium or at least some strains within this species could be zoonotic, like other *Anaplasmataceae*.

Initially, different strains have been defined according to their geographic origins. The variable efficiency of available "vaccines", either commercial or experimental (Mahan et al., 2001; 1998; Marcelino et al., 2007; Martinez et al., 1994; Vachiery et al., 2006), towards "heterologous" strains is apparently due to the genetic and antigenic diversity of *ER* (Adakal et al., 2010a,b; Faburay et al., 2007a; Frutos et al., 2006). However, the lack of accurate typing tools and the fact that *ER* is very difficult to cultivate, being strictly intracellular, made it difficult to explore in depth the genetic diversity of *ER*. It is only recently that this has become possible, when molecular methods were made available.

* Corresponding author. Fax: +33 1 43 96 71 31.
E-mail address: nhaddad@vet-alfort.fr (N. Haddad).
[1] Present address: NewVectys Bâtiment CERVI 5e étage, Hopital de la Pitié Salpétrière, 83, bd de l'Hopital, 75013 Paris, France.
[2] Present address: GPMS, Institut de Génétique et Microbiologie, Bât 400, Université Paris-Sud 91405 Orsay cedex, France.

0167-7012/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2011.11.011

Since now, two nested PCR tests (Molia et al., 2008; Martinez et al., 2004) have been developed that allow detection and direct typing of *ER* on both mammalian hosts (cattle, sheep and goat) and vectors (*Amblyomma* ticks). They are based on the amplification of *pCS20*, which is strongly conserved between strains, and of segments belonging to the *map1* (Major Antigenic Protein 1) gene (Adakal et al., 2010a,b; Allsopp et al., 2001; Faburay et al., 2007a,b; Raliniaina et al., 2010). To date, strain typing has been based more particularly on *map1* gene sequence which displays three hyper variable regions (Faburay et al., 2007a; Raliniaina et al., 2010). It seems to be a good discriminatory tool; however, there is no relationship between *map1* genotypes and geographical origins or evolution of ancestral founders. Thus, there was a need to explore other genotyping methods using multi-locus approaches that could be complementary and more informative concerning the *ER* population structure and heterogeneity. Multi-locus sequence typing (MLST) analysis was developed for *ER* and was conducted at a regional level in Burkina Faso during two successive years showing the presence of two distinct populations with low genetic diversity, one in stasis and the other in clonal expansion pointing out one ancestral founder (Adakal et al., 2010a). Nevertheless, MLST, which is useful for evolutionary investigations on bacterial populations, is generally considered to display limited discriminatory power, being based on the restricted polymorphism of housekeeping genes. MLST gives a global picture of bacterial population structuration but does not allow differentiating widely within main clusters of strains.

Polymorphic Variable Number Tandem Repeats (VNTRs) represent good candidates for the development of a discriminatory, more user-friendly and less costly typing technique. The combination of different VNTRs can reveal a good level of polymorphism (Vergnaud and Denoeud, 2000). This principle has been the basis for the development of Multiple-locus VNTR analysis (MLVA). In many bacteria, including those displaying low genetic heterogeneity using other techniques, MLVA has proved to be a reliable typing method by discriminatory, simple and transferable, with an excellent level of stability for the markers (Lindstedt, 2005). We have successfully developed MLVA for another α-proteobacterium, *Bartonella henselae* (Bouchouicha et al., 2009; Monteil et al., 2007). We have been encouraged to develop this technique for *ER*, which is very rich in VNTR (up to 8.5% of the genome) (Collins et al., 2005; Frutos et al., 2006). Majorities of *ER* VNTRs are intergenic and are distributed in two categories: those which length is less than 25 bp and those between 100 and 300 bp, which represent 75% of the VNTR.

The availability of the complete genome sequences of three strains of *ER*, including Gardel strain, ERGA, isolated in Guadeloupe (1,499,920 bases encoding 950 genes) (Collins et al., 2005; Frutos et al., 2006), and two Welgevonden strains (initially isolated in South Africa), ERWO (Collins et al., 2005; Frutos et al., 2006) and ERWE (Frutos et al., 2006) (1,516,355-bp encoding 888 proteins *vs* 1,512,977-bp encoding 958 proteins respectively), makes it potentially easy to develop a MLVA technique, by identifying VNTRs that are at least dimorphic. On the whole, the gene organization of the Gardel and Welgevonden strains is similar (Frutos et al., 2006), but many intergenic segments and genes are mutated (Vachiery et al., 2008b).

In this study, we describe for the first time the development of MLVA based on such VNTR structures, in order to characterize the diversity of *ER* and the spatial distribution of this diversity. For application on epidemiological and surveillance studies, MLVA has to be able to amplify genetically different *ER* strains from different geographical origins and has to be used on field samples, including both clinical samples from mammalian hosts (blood or tissue samples) and collected ticks. Consequently, for *ER*, MLVA was developed in two steps, with a first development based on DNA from *ER* strains isolated in different countries and purified from infected cell culture, and a second step on DNA from experimentally infected ticks and goats.

2. Methods

2.1. Biological samples

2.1.1. DNA of ER strains from in vitro culture

ER strains were isolated from blood of ill ruminants in different geographical areas and then cultured *in vitro* in bovine endothelial cells first in the regional laboratories and then at CIRAD. Bovine aorta endothelial cells used to produce *ER* strains were primary cell lines isolated at CIRAD laboratory from Creole cattle. *ER* DNA was extracted from infected bovine aorta endothelial cells following the protocol previously described (Frutos et al., 2006, Martinez et al., 1994). Briefly, after high speed centrifugation (20,000×*g* during 30 min), the pellet of elementary bodies was resuspended in 350 µl of saline phosphate buffer. 150 µl of DNase (1 µg/ml) was added to remove the contaminant bovine DNA from host cells and the samples were incubated at 37 °C for 90 min. The treatment was stopped by adding 25 mM of EDTA. Whole bacterial DNA was obtained using QIAamp extraction kit (Qiagen, France).

Seventeen DNA samples corresponding to 13 strains of *ER* from various geographical origins (including Gardel and Senegal at 2 different passages on bovine aortic endothelial cells), were tested by PCR targeting VNTR (Table 1). These 13 strains included two of the three strains sequenced previously: ERGA (*ER* Gardel) NC_006831 (GenBank Accession, CR925677) and ERWE (*ER* Welgevonden) NC_006832 (GenBank Accession, CR925678) (Frutos et al., 2006). ERWE is originated from ERWO, initially isolated in South Africa in 1985 but maintained in a different cell environment after reception in Guadeloupe in 1988. Even if ERWO and ERWE are very close, they differ from each other by more than 3 kb as a result of having different numbers of tandem repeats and genes (Frutos et al., 2006). The CTVM Gardel strain was originated from Gardel strain isolated in Guadeloupe in 1982 and transferred and cultivated in 1993 at the Centre for Tropical Veterinary Medicine (CTVM, University of Edimburgh, Scotland) (Bekker et al., 2005). It is characterized by a recombination between *map1–2* and *map1–3* genes and complete deletion of *map1–2*. Blonde strain was also isolated in Guadeloupe in 2001 from infected ticks collected on a sick goat and engorged on naive animals and then cultured *in vitro*. Virulent Senegal, Lutale and Umpala strains were kindly provided by Pr Frans Jongejan from Utrecht University and Sankat 430 by Dr Lesley Bell Sakyi from the CTVM. Bekuy 255, Banan 112, Bankouma 421 and Cameroon strains were provided by Dr Frederic Stachurski from CIRDES (Burkina Faso) (Bekker et al., 2005). Senegal and Gardel strains were attenuated *in vitro* after successive passages in bovine endothelial cells at CIRAD laboratory. The

Table 1
DNA samples of *ER* strains from *in vitro* culture used to test VNTR.

Isolate/strain	Name	Passages	Virulence[a]	Origin	Geographical area
Gardel	ERGA p18	18	+	Guadeloupe	Caribbean
Gardel	ERGA p27	27	+		
Gardel	ERGA p40	40	+		
Gardel	ERGA p237	237	−		
Gardel	CTVM p5	5	+/−		
Blonde	BLONDE p8	8	+		
Bekuy 255	ERBE p9	9	+	Burkina Faso	West Africa
Bankouma 421	ERBAK p4	4	+	Burkina Faso	
Banan 112	ERBA p7	7	+	Burkina Faso	
Cameroon	ERCA p9	9	+	Cameroon	
Sankat 430	ERSA p4	4	+	Ghana	
Senegal	ERSE p5	5	+	Senegal	
Senegal	ERSE p64	64	−	Senegal	
Lutale	ERLU p2	2	+	Zambia	South Africa
Mara	ERMA p1	1	+	South Africa	
Umpala	ERUM p2	2	+	Mozambique	
Welgevonden	ERWE p11	11	+	South Africa	

[a] Virulence *in vivo*.

attenuation of Gardel passage 237 and Senegal passage 64 was verified on goats. Naïve goats were infected experimentally with 9.10^4 (2 goats) and 9.10^5 (1 goat) viable elementary bodies of Gardel passage 230 which corresponded respectively to 1 lethal dose and 10 lethal doses for the virulent strain. Infection induced transient clinical symptoms such as fever and loss of appetite but no mortality occurred (Vachiery et al., 2006). The same protocol was performed using Senegal passage 64 with 2 goats infected with 9.10^4 and one goat with 9.10^5 viable elementary bodies. Animals recovered without any treatment and were slightly ill during incubation period.

2.1.2. DNA from ER infected ticks and goats

ER DNA was extracted from *A. variegatum* ticks, brain and lung from Creole goats experimentally infected with Gardel (tick samples 0517F3 and 0517M4 and goat 0046 brain and 9833 brain and lung) or infected with Bekuy 255 (tick sample 0340M1). Briefly, goats were inoculated intravenously using calibrated *ER* from infected cells as described previously (Vachiery et al., 2006). Five to 6 days after infection, nymphs were engorged on the animals in order to feed during hyperthermia around day 10. After tick moulting, DNA was extracted from ticks using DNA QiaAmp minikit (Qiagen, France) according to the manufacturer's instructions and the positivity for *ER* was confirmed using *pCS20* nested PCR (Molia et al., 2008). Organs (brain and lung) were collected after death of goats experimentally infected and DNA was extracted using DNA QiaAmp minikit (Qiagen, France). Uninfected controls for both types of samples, ticks (0622M1) and goat organs (0543: brain and 0327: brain and lung), were tested.

Ethics statement: Animal experiments were conducted according to internationally approved OIE standards, under authorizations set forth by the director of the veterinary services of Guadeloupe on behalf of the Prefect of Guadeloupe on August 2006 (authorization number: A971-18-01).

2.2. Computer analysis of repetitive DNA sequences for use as VNTR candidates

The genomic DNA sequences of two *ER* reference strains ERGA (*ER* Gardel), NC_006831, and ERWO (*Ehrlichia ruminantium* str. Welgevonden, GenBank Accession CR767821.1, Reference Sequence NC_005295.2) (Collins et al., 2005; Frutos et al., 2006) were screened for repetitive DNA sequences using the tandem repeats database developed by Le Flèche et al. (2001). The following criteria were considered for application to the selection of VNTR candidates: 1/ Different numbers of repeats between the two reference strains ERGA and ERWO; 2/ Absence of homology between a candidate sequence and its flanking regions; 3/ Repeat unit lengths between 9 and 300 bp; 4/ Percent matches of tandem repeats preferably higher than 75%. Predicted PCR product sizes for the *ER* reference strains were also deduced using the minisatellite website (http://minisatellites.u-psud.fr). BLASTN analysis of the repeat sequences allowed exclusion of the repeats present in other available genomes, especially from other *Anaplasmataceae* in order to avoid any cross reaction. The percent matches of repeat units were also provided for each reference strain ERGA and ERWO by the database developed by Le Flèche et al. (2001). After VNTR selection and validation on *ER* strains, homology between ERWO and ERWE for each validated VNTR was also measured by multi align software using NCBI database.

2.3. MLVA development

2.3.1. Primer design for the VNTR candidates

The tandem repeats database described by Le Flèche et al. (2001) also provides the flanking sequences for each VNTR candidate, both 450 bp upstream and 450 bp downstream of the repeat motif. BLASTN analysis of these flanking sequences was used for designing the forward and reverse primers. Primers were designed according to the following criteria: 1/ GC% similar for all primers in order to use the same PCR conditions for all VNTR; 2/ GC% close to 50%; 3/ Absence of annealing with other regions of *ER* genome, with other pathogen genomes and with bovine genome. The VNTR candidates were known as RUs (*RUminantium*), and each selected VNTR was designated RU followed by a number. The location of each RU in the *ER* genome is described as the 'locus'. An 'allele' corresponds to a given number of repeat units for a given RU or locus. The VNTR name, basic unit (BU) length, % GC, % conservation between ERGA and ERWO, primer sequences, number of BU for ERGA and expected size of amplicons on Gardel strain are shown in Table 2.

2.3.2. VNTR candidate amplification, analysis and selection

VNTR amplification of *ER* DNA samples was conducted in a volume of 25 µl containing 1 µl of purified DNA, 1× TaKaRa buffer, 2.5 mM of each dNTP, 0.8 µM of forward and reverse primers, and 1 U TaKaRa Ex Taq™ polymerase (TAKARA Bio Inc., Japan). After the initial denaturation 3 min at 94 °C, the 35 cycles PCR run included 3 different steps: denaturation at 94 °C for 30 s, annealing at 50 °C for 1 min and extension at 72 °C for 1 min and 30 s. It was followed by final extension 10 min at 72 °C. Taq polymerase (Eurobio, France) at 0.5 U per mix reaction was used too.

For infected or naïve ticks and organs, 1–2 µl of DNA was used and the following polymerases were tested with the same PCR conditions as described above: TaKaRa Ex Taq™ (TAKARA Bio Inc, Japan), Pwo SuperYield DNA polymerase (Roche, France), Taq polymerase (Eurobio, France) and GoTaq® Hot Start Polymerase (Promega, France).

PCR products were separated by gel electrophoresis in 1–2% agarose gels, and stained with ethidium bromide. Long gels (26×40 cm, CBS Scientific® model SGU-2640T-02) and migration times of 2 h at 260 V/cm (Generator PS608 (600 V–800 mA–300 W), Apelex ®, France) were used, in order to allow an accurate measure of the band size. Different molecular ladders were used (from 50 bp to 1 kb). These procedures permitted to differentiate two alleles even with 9pb difference. For a given RU, the expected PCR product length for ERGA strain was calculated considering the unit length, the number of units in the ERGA strain locus, and the length of the flanking sequences separating primer hybridization sites from RU regions. The estimated size range was determined for the different alleles.

2.3.3. Data analysis

For evaluating the discriminatory power of the selected RU, the Hunter and Gaston, (1988) discrimination index (DI) was used, as recommended by the European Society of Clinical Microbiology and Infectious Diseases Study Group on Epidemiological Markers (Struelens, 1996). This index measures the probability that two isolates or strains, randomly chosen, will have different types (Struelens, 1996). It is defined by:

$$DI = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} \left(n_j - 1\right)$$

N: number of isolates or strains; S: total number of alleles; and n_j: number of isolates or strains with the allele j. Polymorphism is considered high when this index is higher than 95% (Felsenstein, 1989).

Clustering analysis was done using a phenetic approach, since the comparison of strains was based on small genomic sequences. The distance matrix was constructed by counting the number of different loci between isolates. With this method, the character states are considered to be unordered and, for a given RU, the same weight is given to a small or a large difference of the number of repeats. Several distance measures may be used when building trees using neighbour-joining algorithm (or other distance-based methods such as UPGMA). However, distances based upon the number of repeats (Manhattan distance, Euclidian distance, Minkowski distance, etc.)

Table 2
Characteristics and primers of *ER* targeted VNTR. B.U.: basic unit; between brackets: position in the genome of ERGA reference strain.

VNTR name	B.U. length (pb)	GC content (%)	Conservation[a] (%)	Primer sequences (5′→3′)	Theorical amplicon length in ERGA (pb)	Nb of UB for ERGA	Nb of alleles/locus	Allele size range (units)	Intragenic Zone
RU-1 (106508)	174	26	94	F CATGATGGTATGACAGC R CACACAGTACAATAGGC	699	2.0	3	0.3–4	+ (82 bp from the end of the gene, *i.e.* 5%)
RU-2 (157452)	153	25	94	F GCAACTTCTGCTACTGT R GCATTCAGATGGAGATG	1107	3.0	3	1–5	+ (16 bp from the end of the gene, *i.e.* 3.5%)
RU-3 (415341)	155	26	95	F CCATCATTACCACCACA R GCAGATAGCCAAGCAAT	1157	4.3	3	2–4	+
RU-4 (453231)	152	20	89	F CTCTTCTATGGTTTGTGAG R GACTCTGAACAATGAACTG	1071	2.8	4	1–4	
RU-5 (465330)	10	8	53	F AAGACACCAATGTGAGC R CCATTACCTGATATTACAC	740	10.8	3	8–11	
RU-6 (506874)	180	20	96	F AGATAAGCCTTTGTTGTGTT R TTACTCAATAATACCCTCATG	728	2.0	3	1–3	
RU-7 (547418)	134	23	92	F CATTTCTTGATGAGTTGG R GACTTGCTATAAATCGC	925	4.0	3	1–7	
RU-8 (600383)	12	32	84	F ATGAGAGATAGTAGGAAG R CAGATGATAACCCTCG	331	14.3	6	13–35	+
RU-10 (840218)	9	44	99	F CTATCTCACCTGGAAGT R CACAAGAACTTACCGAG	1053	27.0	5	9–73	
RU-11 (852733)	178	20	86	F GTACAGGAAAGCGTAAC R ACACTCTCTGCATCCT	1146	3.1	4	2–5	
RU-12 (867581)	15	31	88	F CAGTATGAACATCAGGT R CTGGATGTGGGATCAAT	555	4.4	5	5–14	+
RU-13 (975787)	142	22	93	F CTCTCATCTACCCATAC R GCAGTGTCATCAGTATC	1260	3.4	3	2–10	
RU-14 (1065917)	142	25	88	F TCACTTACAGCACCAC R ACATTTCTTTATCGAGTAGG	970	2.8	4	0.5–5	
RU-15 (1071300)	124	23	93	F TACCACTCTCTGATATTC R GATTGACACTATAACCCT	779	4.0	3	3–5	+
RU-16 (1200221)	136	24	86	F CCATTACTAAGAAAGGGA R CCATAGATTACACCTACA	544	2.8	4	2–8	
RU-17 (1280006)	203	21	89	F CTATCAAGACAGGAAGC R GAACTACGATGATGGAG	1349	4.0	2	1–2	+ (8 bp from the end of the gene, *i.e.* 1.7%)
RU-18 (1313779)	155	18	87	F CTAGAAAAGACCCACTC R CACCACACTTTCAGTTC	621	2.3	4	1–5	
RU-19 (1318297)	127	25	92	F CATTATCAGGTAAACACG R CTGCTATTGAAGTTGAGT	785	3.0	4	2–4	
RU-20 (1433899)	146	22	93	F CAAGTATTCTACGCAGAT R CCTGTAGTGGCTTATTAT	767	2.1	4	1–6	
RU-21 (1439261)	12	56	76	F CCTTATCAAAAGATGAAGAG R CTTGACCTATGCCAATG	245	4.7	4	3–7	+
RU-22 (1459970)	24	43	84	F GGATTACCTGAAGTAGG R TCAGCGGAATACAGTAC	471	5.3	4	4–6	+

[a] Conservation between B.U/sequences.

should only be used when a consensus exists about the marker evolution model. Such a consensus does not exist in our case. Therefore the most conservative distance (the categorical distance) was chosen, which is solely based on the number of differences between markers. This distance is widely used in VNTR-based studies e.g. (Asgharzadeh et al., 2011; Li et al., 2009; Skuce et al., 2002). It is also the default distance proposed for neighbour-joining analysis of VNTR data in common software such as Paup ("Uncorrected-P" distance) or in online resources such as MIRU-VNTRplus (Allix-Béguec et al., 2008). Neighbor-joining (NJ) cluster analysis was then performed using PHYLIP (Felsenstein, 1989). Two dendrograms were built, one with 16 VNTR selected initially, and the second with the 8 VNTR selected during the second step.

3. Results

3.1. Development of MLVA using ER DNA from in vitro cultures

From the two sequences of the reference *ER* strains, ERGA and ERWO, 21 RU candidates fit with the selected criteria, and were at least dimorphic (Table 2). The percentage of identity between repeated units varied from 76% to 99%, except for RU-5 (53%). The bimodal distribution of the candidates according to their sizes (6, *i.e.* 28.5% with sizes from 9 to 24 bp *vs* 15 *i.e.* 71.5% with sizes between 124 and 203 bp) almost strictly reproduces the proportion of these VNTR within *ER* genome (Frutos et al., 2006). These 21 RU candidates were considered for further testing of their polymorphism and of their ability to distinguish strains from different geographical origins, using a series of 13 strains from different regions (6 West African strains: Bekuy 255, Banan 112, Bankouma 421, Cameroon, Sankat 430 and Senegal, 3 Caribbean isolates; Gardel, Gardel CTVM and Blonde, and 4 South African isolates: Welgevonden, Umpala, Mara and Lutale). For RU-1, 5, 17, 18, and 20, we observed a lack of amplification for at least two strains, except for Gardel, Welgevonden and Cameroon strains which were systematically amplified (data not shown). In the case of the Senegal strains, RU-5 and 20 were amplified only with the virulent strains; conversely amplification was observed only for RU-1 and 17, on the attenuated strain. There was amplification of the majority of *ER* strains using RU-2 to 4, 6 to 8, 10 to 16, RU-19, RU-21 and 22 (Table 3).

3.1.1. VNTR polymorphism and discriminatory potential

The 16 VNTR, RU-2 to 4, 6 to 8, 10 to 16, RU-19, RU-21 and RU-22, that amplified all tested strains (apart RU-8 for Bankouma 421) showed similar polymorphisms (3–4 alleles), apart RU-8, 10 and 12,

which were more polymorphic with 6, 5 and 5 alleles respectively (Table 3). Distinct and clear amplification products were obtained for these 16 VNTR. RU-3 and RU-13 profiles were identical for the 13 strains whereas all the other VNTR allowed differentiating them (Table 3). Altogether, the 16 VNTR allowed to distinguish ten different profiles among these 13 strains.

For the determination of the DI, 12 different strains were taken into consideration, in order to avoid the inclusion of artificial clusters of strains linked to laboratory conditions (Gardel and CTVM strains) and attenuation phenomena (Gardel and Senegal strains). The global genetic DI value based on the number of alleles and on their frequency was 0.97.

A dendrogram was built using NJ method on the 16 VNTRs (data not shown). Nine strains corresponding to Western Africa on the one hand and Southern Africa on the other hand are grouped by geographical origin in the dendrogram, like 1/ ERSE p5 from Senegal, Sankat 430 (ERSA p4) from Ghana and the 3 strains from Burkina Faso, or 2/ Mara (ERMA p1) from South Africa, Lutale (ERLU p2) from Zambia and Umpala (ERUM p2) from Mozambique, whereas two strains originating from different areas are identical in the dendrogram (Welgevonden and Cameroon strains).

3.1.2. Stability

For the 16 VNTRs candidates, no variations of size were observed both for Gardel and for Senegal strains, between virulent and attenuated strains whatever the number of passages on bovine endothelial cells (until 237 for Gardel and 64 for Senegal) (Table 3). In addition, the Gardel CTVM and Blonde strains did not differ from original Gardel (ERGA p18) regarding the VNTRs. Moreover, there was no difference in the number of repetitions between ERWO and ERWE strains by *in silico* analysis, except for RU-10 and RU-13 (data not shown). For RU-10, there were 8 and 22 basic units for ERWO and ERWE respectively. The predicted number of repetitions for ERWE was confirmed for the 15/16 VNTRs including RU-10, by PCR amplification on ERWE DNA (Table 3). For RU-13, a difference in the number of repetitions between ERWE and ERWO, 9 *vs* 13 units, was anticipated by *in silico* analysis. After PCR amplification on ERWE DNA, we obtained a PCR amplicon corresponding to 10 basic units instead of 9 on RU-13. For Gardel strain, the number of predicted repetitions was confirmed by PCR amplification for all the VNTRs except for RU-7, which contained 5 instead of 4 basic units (Table 3).

3.2. Development of MLVA using ER DNA from infected ticks and goats

Experimentally infected samples, ticks and organs mimicking field samples, were tested for the 16 VNTRs. The specificity of the 16 pairs of primers (Table 2) was simultaneously checked on uninfected tick (0622M1) and uninfected goat samples: brain from goat 0543 and brain and lung from goat 0327. Satisfactory amplification for infected DNA samples was obtained both with GoTaq® Hot Start and Eurobio Taq Polymerases for ticks 0517F3 and M4 and 0046 brain. PCR signals obtained on experimentally infected ticks and organs (*i.e.* 0340M1, 9833 lung and brain) were weaker than those from *in vitro ER* DNAs.

Faint multi-bands or distinct amplicons were obtained also on uninfected ticks and organs for RU-3, 4, 8, 13, 15, 16, 19 and 22, traducing unspecific hybridizations of primers (data not shown). Thus, these VNTRs could not be used in these conditions on field samples. *ER* was detected from all infected ticks and organs for RU-6, 10, 12 and RU-21 with no signal for negative samples. The expected VNTR sizes were obtained from experimentally infected ticks and goats infected with Gardel and Bekuy 255 strains. Successful detection was observed on 2 out of 3 infected ticks and the 3 organs for RU-2, 7 and RU-14. For RU-11, both infected ticks and goat organs were not amplified, indicating a lower efficiency of these PCRs.

In summary, the optimal VNTRs were RU-2, 6, 7, 10, 11, 12, 14 and RU-21, when taking into consideration both signal sensitivity and specificity. These 8 VNTRs were as discriminatory as the 16 VNTRs, and they allowed obtaining the same distribution of the strains in the dendrogram (Fig. 1 for 8 VNTR, data not shown for 16 VNTR) and the same DI value (0.97).

4. Discussion

The main difficulty at the early stage of the development of the MLVA technique for *ER* was related to its very low genome GC content (27.5%) (Frutos et al., 2006) as the optimal conditions for designed primers was around 50% of GC. The development on *ER* strains was successful for 16 out of 21 selected VNTRs. The positioning of these 16 VNTRs all along the genome with a bimodal distribution of their size strongly suggests that they are representative of the population of VNTR of *ER* genome (Frutos et al., 2006).

The lack of amplification for the 5 remaining VNTRs was probably due to important mutations or deletions in flanking zones, including or not VNTRs, for unamplified *ER* strains rather than default of primer

Table 3
VNTR profiles for 17 *in vitro ER* samples (the number of B.U. per VNTR is indicated for each strain). Negative values in the case of RU-12 result from truncated flanking sequences with probably complete deletion of the VNTR. (1) Same number of basic units as theoretical number of units except for RU-7. (2) Same number of basic units as theoretical number of units except for RU-13.

Name	RU-2	RU-3	RU-4	RU-6	RU-7	RU-8	RU-10	RU-11	RU-12	RU-13	RU-14	RU-15	RU-16	RU-19	RU-21	RU-22
ERGA p18[1]	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
ERGA p27[1]	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
ERGA p40[1]	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
ERGAp237[1]	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
BLONDE p8	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
CTVM p5	3	4	3	2	5	13	30	3	4	3	3	4	3	3	5	4
ERBE p9	1	2	1	1	2	30	10	2	2	2	3[a]	3	4	1	3	2
ERBAK p4	1	2	1	1	2	NA	22	2	−2	2	0.5	3	4	2	3	2
ERBA p7	1	2	1	1	2	30	10	2	2	2	3	3	4	2	3	2
ERCA p9	5	3	4	3	8	20	22	4	2	10	5	5	8	4	4	6
ERSA p4	1	2	1	1	2	36	22	2	2	2	3	3	2	2	3	2
ERSE p5	1	2	1	1	2	36	21	2	−2	2	1	3	2	2	3	2
ERSE p64	1	2	1	1	2	36	21	2	−2	2	1	3	2	2	3	2
ERLU p2	5	3	2	3	5	20	30	3	4	10	5	3	2	4	4	6
ERMA p1	3	2	2	2	5	17	21	5	−5	2	3	3	3	2	3	6
ERUM p2	3	2	3	3	5	37	14	3	16	2	3	3	3	2	7	4
ERWE p11[2]	5	3	4	3	8	20	22	4	2	10	5	5	8	4	4	6
Nb of alleles	3	3	4	3	3	6	5	4	5	3	4	3	4	4	4	4

p: number of passages in culture.
NA: non amplified.
[a] Corresponding to 3.4 with an allelic profile different from Gardel cluster.

design. For Umpala strain, depending on the MAP proteins, the tree topology varies considerably: Umpala belongs to Gardel cluster (for MAP1–2, MAP1 and MAP1 + 1), to Welgevonden cluster (for MAP1–5) or appears different from both Gardel and Welgevonden clusters (for MAP1–6 and MAP1–14) (Raliniaina et al., 2010). On the other hand, the sequences of *pCS20* fragment or *map* multigenic family for Senegal and Mara strains were shown to differ widely from those of Gardel and Welgevonden strains (Allsopp et al., 2003; Raliniaina et al., 2010). These differences could explain the default of hybridization of some VNTR primers based on Gardel and Welgevonden sequences for Umpala, Senegal and Mara strains. Interestingly, either virulent or attenuated Senegal strains could not be amplified with RU-1, RU-5, 17 and 20, which suggests a lack of stability of these VNTRs and modifications of flanking zones along attenuation process. Apart RU-1 and RU-17, all are intergenic, thus *a priori* less submitted than intragenic VNTRs to counter-selection of mutations. In addition, a very limited proportion of each of these VNTR was located within the extreme 3′ end of the corresponding gene sequence, *i.e.* 5% of RU-1 and 1.7% of RU-17. Thus, a polymorphism of these VNTRs has probably no impact on the functionality of the corresponding proteins: bifunctional GMP synthase/glutamine aminotransferase and chaperon protein dnaJ respectively.

Intergenic VNTRs, and particularly those with larger size, are considered to be or to have been involved into expansion/contraction genomic events and thus in the genome plasticity (Frutos et al., 2006), in addition to recombinational events (Hughes and French, 2007). Even if evolutionary potential of VNTRs is needed in order to generate the diversity required for typing, a VNTR profile has to be stable enough for allowing epidemiological studies at least at the scale of an outbreak, and even for evolutionary studies, if the results are correlated with techniques such as MLST. The divergent profiles of the two Welgevonden strains, ERWO and ERWE, which differ by more than 3 kb, are illustrative of the evolutionary potential of the VNTR when *ER* is maintained in different *in vivo* and/or *in vitro* environments. Nevertheless, for the 16 selected VNTRs, differences between ERWO and ERWE defined by *in silico* analysis and confirmed by PCR for ERWE were limited to 2 VNTRs only. Stability of the tested VNTRs, at least in stable environments, is also suggested by the absence of any modification of the number of repeated sequences for Gardel and Senegal strains respectively despite a high number of passages *in vitro*. There is no involvement of the polymorphic VNTR tested in virulence, contrary to previous studies that have shown or suggested associations between repetitive sequences and pathogenicity, zoonotic potential and/or phase variation (Bouchouicha et al., 2009; Harper, et al., 2008; Parkhill et al., 2000; van Belkum, 1999). The same VNTR profile was observed for the CTVM and original Gardel strains despite important genomic modifications that occurred elsewhere in the genome, in particular in *map* genes and between Blonde and Gardel strains, which originate from strains isolated at 19 years' intervals in Guadeloupe.

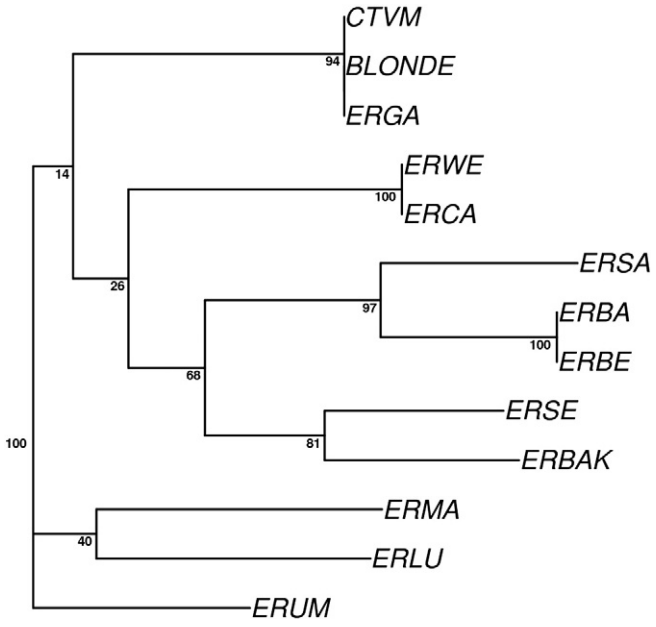

Fig. 1. Dendrogram built using 8 different VNTRs for 13 *ER* strains. Blonde = Blonde strain; CTVM = Gardel strain cultivated at the CTVM; ERBA = Banan112; ERBAK = Bankouma 421; ERBE = Bekuy strain; ERCA = Cameroon strain; ERGA = Gardel strain; ERLU = Lutale strain; ERMA = Mara strain; ERSA = Sankat 430 strain; ERSE = Senegal strain; ERUM = Umpala strain; ERWE = Welgevonden strain.

The ability of developed VNTRs to type *ER* from experimentally infected ticks and goats mimicking field samples was essential in order to use this method for further molecular epidemiological studies. Eight VNTRs RU-2, 6, 7, 10, 11, 12, 14 and RU-21 offered promising results on experimentally infected samples. These VNTRs, except RU-11, could be used immediately for the characterization of strains within host organs during epidemiological studies.

The high DI value reflects the good potential of the MLVA technique for diversity evaluation of *ER* populations. Our results reinforce the assertion of a great diversity among *ER* isolates (Adakal et al., 2010b; Martinez et al., 2004; Raliniaina et al., 2010; Vachiéry et al., 2008a). The topology of dendrograms was different between *map1* and MLVA, *i.e.* Lutale and Umpala belong to the Gardel *map1* cluster whereas they were widely different using MLVA. Thus, the MLVA technique appeared more discriminatory than genotyping based on *map-1* gene.

Our results are in accordance with those obtained using MLST, *pCS20* or *groESL*, as the Welgevonden, Senegal/Sankat 430 and Gardel strains belong to 3 different clusters, which suggests that MLVA is robust for studying strain relationships (Allsopp et al., 2003; van Heerden et al., 2004).

Our preliminary results suggested that MLVA permitted to associate MLVA profiles and geographical clusters from Southern Africa, from Caribbean region and from Western Africa. However, Cameroon and Welgevonden strains had the same VNTR profiles even though they were geographically distant. It has to be underlined that Cameroon and Welgevonden strains belong to the same cluster for 6 polymorphic MAP proteins (Raliniaina et al., 2010). More field samples or *ER* isolates from the field will be required in order to confirm that MLVA can be used as a geographical genetic marker, as a guide for the choice of vaccine strains for regional vaccines' design and for traceability studies, especially if its zoonotic potential is confirmed.

5. Conclusions

MLVA will be suitable, quick and easier than MLST, which implied a sequencing step, for studying *ER* strains in regional laboratories. An optimization of the *ER* MLVA method considering the 16 VNTRs will be done by development of nested PCR in order to improve the sensitivity of *ER* typing in *Amblyomma* ticks and to do molecular epidemiological studies. In addition, further comparative studies with MLST using more isolates from various origins will be helpful for substantiating the limits of the use of MLVA for *ER*, which already appears as a very promising technique for studying *ER* diversity at local and more global scales.

More largely this is the first study on the development of MLVA approach for a *Rickettsiales*. It demonstrated that it could be developed and used for other *Rickettsiales* that could have similar genetic characteristics, like *Rickettsia prowazekii* and *R. typhi*.

Acknowledgements

- GIP INRA-CIRAD funding, Project " Development of *Ehrlichia ruminantium* molecular typing by Multiple Locus VNTR Analysis (MLVA)"

- CRVOI funding (Centre de recherche et de veille sur les maladies émergentes dans l'océan indien), Project DIGEERAV-OI (PRAO/ AlqD/ CRVOI/08/05).
- European project, FEDER 2007–2013, FED 1/1.4-30305, "Risque en santé animale et végétale".

References

Adakal, H., Gavotte, L., Stachurski, F., Konkobo, M., Henri, H., Zoungrana, S., 2010a. Clonal origin of emerging populations of *Ehrlichia ruminantium* in Burkina Faso. Infect. Genet. Evol. 10, 903–912.

Adakal, H., Stachurski, F., Konkobo, M., Zoungrana, S., Meyer, D.F., Pinarello, V., et al., 2010b. Efficiency of inactivated vaccines against heartwater in Burkina Faso: impact of *Ehrlichia ruminantium* genetic diversity. Vaccine 28, 4573–4580.

Allix-Beguec, C., Harmsen, D., Weniger, T., Supply, P., Niemann, S., 2008. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J. Clin. Microbiol. 46, 2692–2699.

Allsopp, B.A., 2010. Natural history of *Ehrlichia ruminantium*. Vet. Parasitol. 167, 123–135.

Allsopp, M.T.E.P., Dorfling, C.M., Maillard, J.C., Bensaid, A., Haydon, D.T., van Heerden, H., et al., 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., 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., Louw, M., Meyer, E.C., 2005. *Ehrlichia ruminantium*: an emerging human pathogen? Ann. N. Y. Acad. Sci. 1063, 358–360.

Asgharzadeh, M., Kafil, H.S., Roudsary, A.A., Hanifi, G.R., 2011. Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. Infect. Genet. Evol. 11, 124–131.

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

Bouchouicha, R., Durand, B., Monteil, M., Chomel, B.B., Berrich, M., Arvand, M., et al., 2009. Molecular epidemiology of feline and human *Bartonella henselae* isolates. Emerg. Infect. Dis. 15, 813–816.

Collins, N.E., Liebenberg, J., De Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., et al., 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.

Faburay, B., Geysen, D., Munstermann, S., Bell-Sakyi, L., Jongejan, F., 2007a. Longitudinal monitoring of *Ehrlichia ruminantium* infection in Gambian lambs and kids by pCS20 PCR and MAP1-B ELISA. BMC Infect. Dis. 7, 85.

Faburay, B., Geysen, D., Ceesay, A., Marcelino, I., Alves, P.M., Taoufik, A., et al., 2007b. Immunisation of sheep against heartwater in The Gambia using inactivated and attenuated *Ehrlichia ruminantium* vaccines. Vaccine 25, 7939–7947.

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., et al., 2006. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. J. Bacteriol. 188, 2533–2542.

Harper, K.N., Liu, H., Ocampo, P.S., Steiner, B.M., Martin, A., Levert, K., et al., 2008. The sequence of the acidic repeat protein (arp) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong positive selection in the subspecies that causes syphilis. FEMS Immunol. Med. Microbiol. 53, 322–332.

Hughes, A.L., French, J.O., 2007. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. Gene 387, 31–37.

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.

Le Fleche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., et al., 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol. 1, 2.

Li, Y., Cui, Y., Hauck, Y., Platonov, M.E., Dai, E., Song, Y., et al., 2009. Genotyping and phylogenetic analysis of *Yersinia pestis* by MLVA: insights into the worldwide expansion of Central Asia plague foci. PLoS One 4, e6000.

Lindstedt, B.A., 2005. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis 26, 2567–2582.

Mahan, S.M., Kumbula, D., Burridge, M.J., Barbet, A.F., 1998. The inactivated *Cowdria ruminantium* vaccine for heartwater protects against heterologous strains and against laboratory and field tick challenge. Vaccine 16, 1203–1211.

Mahan, S.M., Smith, G.E., Kumbula, D., Burridge, M.J., Barbet, A.F., 2001. Reduction in mortality from heartwater in cattle, sheep and goats exposed to field challenge using an inactivated vaccine. Vet. Parasitol. 97, 295–308.

Maillard, J.C., Maillard, N., 1998. Historique du peuplement bovin et de l'introduction de la tique *Amblyomma variegatum* dans les îles françaises des Antilles: Synthèse bibliographique. Ethnozootechnie 1, 19–36.

Marcelino, I., Vachiéry, N., Amaral, A.I., Roldao, A., Lefrançois, T., Carrondo, M.J.T., et al., 2007. Effect of the purification process and the storage conditions on the efficacy of an inactivated vaccine against heartwater. Vaccine 25, 4903–4913.

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

Molia, S., Frebling, M., Vachiery, N., Pinarello, V., Petitclerc, M., Rousteau, A., et al., 2008. *Amblyomma variegatum* in cattle in Marie Galante, French Antilles: prevalence, control measures, and infection by *Ehrlichia ruminantium*. Vet. Parasitol. 153, 338–346.

Monteil, M., Durand, B., Bouchouicha, R., Petit, E., Chomel, B., Arvand, M., et al., 2007. Development of discriminatory multiple-locus variable number tandem repeat analysis for *Bartonella henselae*. Microbiology 153, 1141–1148.

Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., et al., 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 403, 665–668.

Raliniaina, M., Meyer, D.F., Pinarello, V., Sheikboudou, C., Emboule, L., Kandassamy, Y., et al., 2010. Mining the genetic diversity of *Ehrlichia ruminantium* using map genes family. Vet. Parasitol. 167, 187–195.

Skuce, R.A., McCorry, T.P., McCarroll, J.F., Roring, S.M., Scott, A.N., Brittain, D., et al., 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology 148, 519–528.

Struelens, M.J., 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin. Microbiol. Infect. 2, 2–11.

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

Vachiery, N., Lefrançois, T., Esteves, I., Molia, S., Sheikboudou, C., Martinez, D., 2006. Optimisation of the inactivated vaccine dose against heartwater and *in vitro* quantification of *Ehrlichia ruminantium* challenge material. Vaccine 24, 4747–4756.

Vachiéry, N., Jeffery, H., Pegram, R., Aprelon, R., Pinarello, V., Kandassamy, Y., et al., 2008a. *Amblyomma variegatum* ticks and heartwater on three Caribbean Islands. Ann. N. Y. Acad. Sci. 1149, 191–195.

Vachiery, N., Maganga, G., Lefrançois, T., Kandassamy, Y., Stachurski, F., Adakal, H., et al., 2008b. Differential strain-specific diagnosis of the heartwater agent: *Ehrlichia ruminantium*. Infect. Genet. Evol. 8, 459–466.

van Belkum, A., 1999. Short sequence repeats in microbial pathogenesis and evolution. Cell. Mol. Life Sci. 56, 729–734.

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.

Vergnaud, G., Denoeud, F., 2000. Minisatellites: mutability and genome architecture. Genome Res. 10, 899–907.