



## MLST scheme of *Ehrlichia ruminantium*: Genomic stasis and recombination in strains from Burkina-Faso

Hassane Adakal<sup>a,b</sup>, Damien F. Meyer<sup>c</sup>, Catherine Carasco-Lacombe<sup>c</sup>, Valérie Pinarello<sup>c</sup>, Florian Allègre<sup>c</sup>, Karine Huber<sup>b</sup>, Frederic Stachurski<sup>b</sup>, Serge Morand<sup>d</sup>, Dominique Martinez<sup>b</sup>, Thierry Lefrançois<sup>c</sup>, Nathalie Vachery<sup>c,1,\*</sup>, Roger Frutos<sup>e,1,\*\*</sup>

<sup>a</sup> CIRDES Laboratoire de Biotechnologie (URBIO), 559, 3-51 Avenue du Gouverneur Louveau, B.P.O. 454, Bobo-Dioulasso, Burkina-Faso

<sup>b</sup> Cirad, UMR15, Cirad-Inra, TA-A15/G, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

<sup>c</sup> Cirad, Site de Duclos, 97170 Prise d'eau, Guadeloupe, France

<sup>d</sup> CNRS, ISEM, UMR 5554, CNRS-UM2-IRD, Université Montpellier 2 CC065, Place E. Bataillon, 34095 Montpellier Cedex 5, France

<sup>e</sup> Cirad, UMR 17, Cirad-Ird, TA-A17/G, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

### ARTICLE INFO

#### Article history:

Received 10 June 2009

Received in revised form 9 August 2009

Accepted 18 August 2009

Available online 25 August 2009

#### Keywords:

*Ehrlichia ruminantium*

MLST

Phylogeny

Recombination

Genomic stasis

Polymorphism

Genetic diversity

### ABSTRACT

Heartwater, caused by the intracellular bacterium *Ehrlichia ruminantium*, is a major tick-borne disease of livestock in Africa also introduced in the Caribbean. The main problem encountered with the control of this disease is the lack of efficient vaccine in the field. This is thought to be related to the high genetic diversity of strains circulating in a same area. A set of eight circulating strains was isolated from a herd of cows in a small locality in Burkina-Faso and analyzed along with two reference strains, i.e. ERGA and ERWO, for which full-length genome was available. A MLST analysis was developed based on the genes *gltA*, *groEL*, *lepA*, *lipA*, *lipB*, *secY*, *sodB* and *sucA*. Phylogeny analysis was conducted both on concatenated MLST loci and on each individual locus. This showed differing phylogenies for each individual target gene. Most of the recorded polymorphism was borne by three strains: 331, 469 and 623. The neutrality hypothesis could not be rejected. Recombination and linkage disequilibrium were shown to have occurred. A core of seven strains displayed little polymorphism and signs of most likely ancient recombination events. The two reference strains, one from the Caribbean separated from west African strains three centuries ago and another one isolated in South Africa, were very closely related to the core strains whereas the three differing strains displayed recombination and most of the parcimony informative sites. These data suggest that some strains are in genomic stasis, as expected for intracellular parasites, while others emerge in the same area with DNA polymorphism. This work also shows that the MLST scheme developed can discriminate between these two kinds of strains.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Cowdriosis or heartwater is a tick-borne disease transmitted by *Amblyomma* ticks to domestic and wild ruminants. The bacterial causative agent, *Ehrlichia ruminantium*, is endemic in Sub-Saharan Africa, in Madagascar, and in some Caribbean islands from where it represents a threat to the American mainland (Barré et al., 1987). In endemic areas, heartwater represents a significant obstacle to improvement of livestock production by the introduction of exotic breeds. Several types of vaccines have been developed but they displayed limited efficiency in the field (Collins et al., 2003a,b;

Jongejan, 1991; Mahan et al., 2003; Martinez et al., 1994, 1996; Zwegarth et al., 2005). This is thought to be due to a wide diversity of *E. ruminantium* which was shown to impact cross-protection features (Du Plessis et al., 1989; Adakal et al., 2004) and to their simultaneous circulation in a same region. Several methods and markers were developed to characterize genetic diversity. Out of these, the *map1* gene, encoding the Major surface protein MAP1 used in ELISA tests (Van Vliet et al., 1995), displays a high level of sequence polymorphism between strains (Reddy et al., 1996; Allsopp et al., 2001). However, although *map1* can characterize genetic diversity of strains in field-collected strains (Martinez et al., 2004; Faburay et al., 2007; Raliniaina et al., in press), no correlation could be found with cross-protection between strains. Recombination has been reported in *E. ruminantium* (Hughes and French, 2007; Allsopp and Allsopp, 2007) which could explain at least part of the genetic diversity. *E. ruminantium* is also displaying an active mechanism of genome size plasticity (Frutos et al., 2006,

\* Corresponding author.

\*\* Corresponding author. Tel.: +33 4 64 59 39 62; fax: +33 4 67 59 39 60.

E-mail address: [roger.frutos@cirad.fr](mailto:roger.frutos@cirad.fr) (R. Frutos).

<sup>1</sup> Co-last authors.

2007). On the other hand, *E. ruminantium* is an obligate intracellular parasite which is expected to display genomic stasis after adaptation to intracellular parasitism (Tamas et al., 2002). The *E. ruminantium* genetic diversity is most likely rather complex and there is thus a real need to develop accurate methods for genotyping and characterizing strains as well as for monitoring the structure and dynamics of *E. ruminantium* populations in the field.

We report here the development of a Multilocus Sequence Typing (MLST) scheme for *E. ruminantium* to discriminate even very closely related strains. We also report that strains which are most likely in genomic stasis, but with ancient recombination events, are co-circulating in the same area along with more recently emerged strains bearing most of the nucleotidic polymorphism.

## 2. Materials and methods

### 2.1. Field collection of ticks

*Amblyomma variegatum* ticks were collected in three villages of Bekuy (3.91°W, 11.60°N), Sara (3.83°W, 11.71°N) and Lamba (4.05°W, 11.66°N). Vegetation in the area consists of bushy and woody savannah and in few places, sparse forest. Visible ticks were collected from animals and conserved in 70% ethanol for further species identification. Flat *A. variegatum* ticks were also stored in alcohol while engorged females were allowed to molt, in controlled conditions (27–29 °C temperature and 85–90% relative humidity) to the following stages which were used for DNA extraction.

### 2.2. *E. ruminantium* detection in ticks

DNA was extracted from individual ticks using the Qiamp DNA extraction kit according to the supplier (Qiagen). Detection of *E. ruminantium* in ticks was conducted by targeting a fragment of *pCS20* gene using a semi-nested PCR as previously described (Waghela et al., 1991; Mahan et al., 1992; Peter et al., 1995; Mahan et al., 1998; Martinez et al., 2004). The semi-nested PCR was slightly modified by the use of primers including universal nucleotides (Molia et al., 2008). Universal primers AB128' (ACTAGTAGAAATTGCACAATCYAT) and AB130' (RCTDGCWGCTT-TYTGTTTCAGCTAK) were used as external primers whereas AB128' and AB129' (TGATAACTTGGWGRRGDARTCCTT) were used as internal primers (Molia et al., 2008). PCR reaction was carried out in a 25 µl volume containing 1 µl of DNA, buffer (Qiagen) (MgCl<sub>2</sub> at 1.5 mM final), 200 mM dATP, 200 mM dGTP 200 mM dCTP and 200 mM dTTP (Eurogentec), 0.4 mM each of the corresponding forward and reverse primers, 1 U/µl Taq polymerase (Qiagen). The first round of semi-nested PCR comprised an initial denaturation step of 3 min at 94 °C followed by 35 amplification cycles of 50 s denaturation at 94 °C, 50 s annealing at 50 °C and 50 s extension at 72 °C with a final extension step of 10 min at 72 °C. The same program was applied to the second round of the semi-nested PCR with an annealing temperature of 55 °C.

### 2.3. Target loci

Selection of MLST loci was based both on comparative genomic analysis of *E. ruminantium* (Collins et al., 2005; Frutos et al., 2006) and reports of orthologous genes used for MLST analysis on *E. ruminantium* or other pathogens (Allsopp et al., 2003; Allsopp and Allsopp, 2007; Harbottle et al., 2006; Inokuma et al., 2001; McCombie et al., 2006; Ahmed et al., 2006). These genes: *gltA*, *groEL*, *lepA*, *lipA*, *lipB*, *secY*, *sodB* and *sucA* are presented in Supplementary Table 1. Three reference genes were added for polymorphism analysis and evolutionary patterns based on their

previously reported substitution biases as shown by a Fisher exact test (Frutos et al., 2006). *map1* and *clpB* display a statistically significant excess of synonymous substitutions whereas CDS8580 displays a statistically significant excess of nonsynonymous substitutions (Supplementary Table 1). The distribution of the target loci on the genome of *E. ruminantium* strain Gardel is shown in Supplementary Fig. 1. Orthologous genes from the closely related species *Ehrlichia canis* strains Jake (NC\_007354), *Ehrlichia chaffeensis* strain Arkansas (NC\_007799) and from *Anaplasma marginale* strain Ste Maries (NC\_004842) were used for outgroup rooting.

### 2.4. Primers and PCR reactions

Primers were designed using Primer 3 (Rozen and Skaletsky, 2000). To overcome the problem of potential low infection levels of *E. ruminantium* in ticks, each MLST locus was amplified by nested PCR using a double pair of primers (Supplementary Table 1). In most cases, one of the external primers was used for the second PCR in a semi-nested PCR (Supplementary Table 1). PCR amplifications were performed in 50 µl reaction buffer (Qiagen) (MgCl<sub>2</sub> at 1.5 mM final) with 200 mM dATP, 200 mM dGTP 200 mM dCTP and 200 mM dTTP (Eurogentec), 0.4 mM each of the corresponding forward and reverse primers, 1 U/µl Taq polymerase (Qiagen). The first round PCR program consisted of an initial denaturation step of 3 min at 94 °C followed by 15 cycles of 50 s denaturation at 94 °C, 50 s annealing at the relevant temperature according to each gene (Supplementary Table 1) and 50 s elongation at 72 °C followed by a final extension step of 10 min at 72 °C. 40 cycles were performed for the second round under same conditions. The primer pairs of the second round of each nested-PCR were used for sequencing of the corresponding PCR products (sequenced by Cogenics, Meylan, France).

### 2.5. Sequence alignment, phylogenetic trees and statistics

Consensus sequences were obtained using Nucmer (Kurtz et al., 2004) and Contig Aligner ([http://nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/contig\\_aligner/](http://nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/contig_aligner/)) and Multiple sequence alignments were conducted with ClustalX2.0.3 (Thompson et al., 1997). Whenever relevant, gap regions were eliminated to perform alignments for phylogenetic purpose. Verification of reading frames was performed with ExPasy Translate tool (<http://www.expasy.ch/tools/dna.html>). Phylogenetic trees were constructed using FigTree 1.2.2 (<http://tree.bio.ed.ac.uk/>). Similarity and difference matrices were constructed from ClustalX2 alignments using BioEdit 7.0.9.0 (Hall, 1999). DNA sequence polymorphism and all subsequent tests were investigated using several functions from the DnaSP5.00.02 package (Librado and Rozas, 2009). Haplotypes (alleles) were calculated according to Nei (1987). Nucleotide diversity,  $\pi$  ( $\pi$ ), the average number of nucleotide differences per site between two sequences was calculated according to Nei (1987) and using the Jukes and Cantor (1969) correction. Theta (Watterson's mutation parameter) was calculated from  $\eta$  ( $\eta$ ) or from  $S$  (Watterson, 1975).  $\eta$  ( $\eta$ ) is the total number of mutations, and  $S$  is the number of segregating (polymorphic) sites.  $K_a$  (the number of nonsynonymous substitutions per nonsynonymous site), and  $K_s$  (the number of synonymous substitutions per synonymous site) for any pair of sequences were calculated according to Nei and Gojobori (1986). Tajima's D test (Tajima, 1989) was used for testing the hypothesis that all mutations are selectively neutral (Kimura, 1983). The D test is based on the differences between the number of segregating sites and the average number of nucleotide differences. The confidence limit of D is assessed according to a two-tailed test (Tajima, 1989). Other tests of neutrality are the

statistical tests  $D^*$  and  $F^*$  proposed by Fu and Li (1993) for testing the hypothesis that all mutations are selectively neutral. The  $D^*$  test statistic is based on the differences between  $h_s$ , the number of singletons (mutations appearing only once among the sequences), and  $\eta$ , the total number of mutations (Fu and Li, 1993). The  $F^*$  test statistic is based on the differences between  $h_s$ , the number of singletons (mutations appearing only once), and  $k$ , the average number of nucleotide differences between pairs of sequences (Fu and Li, 1993). The statistical significance of the Fu and Li's  $F^*$  and  $D^*$  tests was also calculated using a two-tailed test. A last test of neutrality was conducted using Fu's  $F_s$  statistic (Fu, 1997). Fu's  $F_s$  test (Fu, 1997) is based on the haplotype (gene) frequency distribution to determine the value of  $\theta$  (Ewens, 1972). ZnS statistics (Kelly, 1997) is the average of  $R^2$  (Hill and Robertson, 1968) over all pairwise comparisons. It reflects excess of linkage disequilibrium compared with that expected under neutrality. Wall's  $B$  statistic (Wall, 1999) is the proportion of congruent pairs of adjacent segregating sites (with common phylogeny), whereas Wall's  $Q$  statistics (Wall, 1999) also considers the length of the corresponding region. Significant pairwise associations were assessed by  $\chi^2$ , Fisher exact test and Bonferroni procedure.

### 3. Results

#### 3.1. Identification of isolated strains

The confirmation of the isolated strains as *E. ruminantium* was conducted using a three-way approach. First of all a pCS20 PCR test specific for *E. ruminantium* (Waghela et al., 1991; Mahan et al., 1992, 1998; Peter et al., 1995; Martinez et al., 2004) was conducted and only positive samples were retained for MLST analysis. The CDS 8580, a gene of unknown function which might in fact be a pseudogene (Frutos et al., 2006) was used as secondary control. This CDS is only found in *E. ruminantium* and is absent in the published genomes of *E. canis*, *E. chaffeensis* and *A. marginale*. CDS 8580 was present in all the strains analyzed here with a high degree of conservancy. This confirmed that the strains analyzed here were indeed *E. ruminantium* strains. Finally, the *gltA* gene in *E. ruminantium* displayed two small deletions of three nucleotides each when compared to *gltA* sequences from *E. canis* and *E. chaffeensis*. These deletions were located at position 456–459 and 474–477 with respect to the *gltA* start codon in *E. canis* and *E. chaffeensis*. All the strains analyzed here displayed these deletions. Furthermore, the nucleotide sequence 5'-GTGAGTTAGAGTGT-3' located between these two small deletions was 100% conserved among all the strains studied here whereas it did not match the corresponding region in *E. canis* and *E. chaffeensis*. This further

confirmed that all the strains studied in this work were *E. ruminantium* strains.

#### 3.2. Identification of haplotypes (alleles) and sequence types (ST)

*E. ruminantium* strain 2 was selected as the reference strain for ST determination. As a consequence strain 2 carried the haplotype (allele) 1 for all the loci considered. STs and loci polymorphism are presented in Table 1. Haplotypes sequences are shown in supplementary material (Supplementary Table 2). Sequences were deposited in GenBank and accession numbers are displayed in Supplementary Table 2. The overall number of haplotypes identified among the 10 strains analyzed were 3, 5, 4, 3, 8, 6, 3 and 7 for *gltA*, *groEL*, *lepA*, *lipA*, *lipB*, *secY*, *sodB* and *sucA*, respectively (Table 1). This variation in the number of alleles from one locus to another did not, however, reflect the polymorphism in the target gene segment amplified. Polymorphic sites or S (i.e. number of mutated sites per locus) varied from 4 to 45 (Table 1). When weighting these data by considering the size of the amplified fragment, the resulting percentage of site polymorphism varied from 0.96% to 9.84% (Table 1). However, 6 out of 8 loci displayed percentage of polymorphism below 4% with only *groEL* displaying 9.84% variability. There was no linear correlation between the number of haplotypes and the number of polymorphic sites, or the percentage of polymorphisms, indicating thus that a very few number of strains carry most of the observed polymorphism on a limited number of haplotypes. When considering both the sequence identity matrix and the sequence difference matrix (Supplementary Table 3), three strains appeared clearly more divergent: strains 331 and 469 on one hand and strain 623 on the other hand.

#### 3.3. Phylogenetic analysis from concatenated sequences

When using midpoint rooting (Fig. 1a), the group comprising strains 331 and 469 was separated from 1062 with a node bootstrap value of 1000. Strain 1062 branches separately and associates to a bootstrap value of 873. When considering the closely related species *E. chaffeensis* (Fig. 1b), *E. canis* (Fig. 1c) and the Rickettsial *A. marginale* (Fig. 1d) as an outgroup, the topology was slightly different with no first order significant bootstrap values excepted for the strains 331 and 469 which consistently branch separately with a bootstrap of 1000.

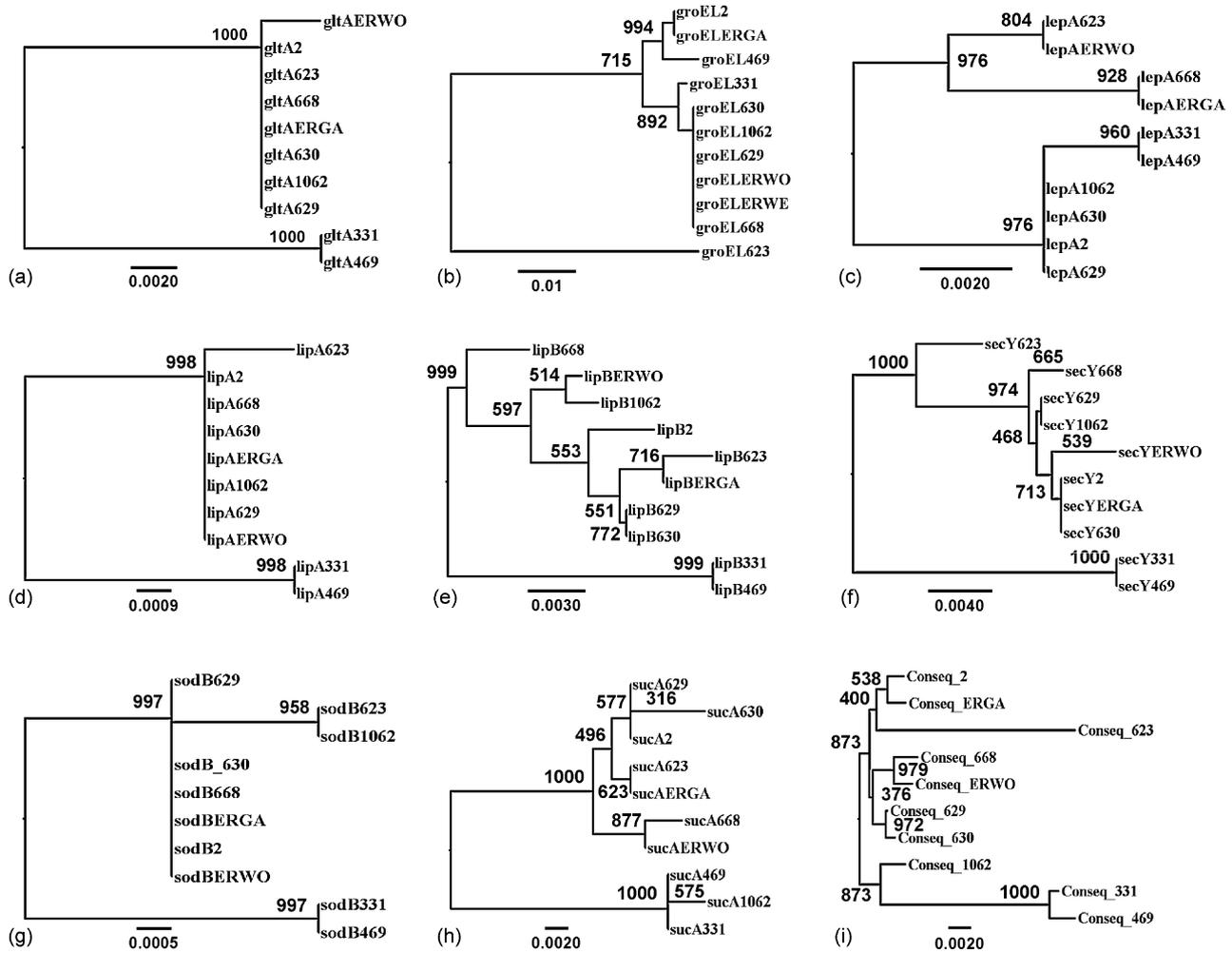
#### 3.4. Phylogeny of individual MLST loci

To further clarify the role of the various haplotypes and loci in the overall diversity displayed by the concatenated sequences, a

**Table 1**  
Summary of STs for the *E. ruminantium* strains considered.

Strains	MLST loci							
	<i>gltA</i>	<i>groEL</i>	<i>lepA</i>	<i>lipA</i>	<i>lipB</i>	<i>secY</i>	<i>sodB</i>	<i>sucA</i>
Strain 2	1	1	1	1	1	1	1	1
Strain 331	3	4	4	3	6	6	3	5
Strain 469	3	5	4	3	6	6	3	5
Strain 623	1	2	2	2	2	2	2	2
Strain 629	1	3	1	1	7	5	1	1
Strain 630	1	3	1	1	7	1	1	6
Strain 668	1	3	3	1	3	3	1	3
Strain 1062	1	3	1	1	8	5	2	7
ERGA	1	1	3	1	4	1	1	2
ERWO	2	4	2	1	5	4	1	4
N Haplotypes	3	5	4	3	8	6	3	7
S	10	44	7	6	13	25	4	16
Polymorphism (%)	2.53	9.84	1.47	1.76	3.63	4.26	0.96	3.99
Size (bp)	395	447	475	341	358	587	415	401





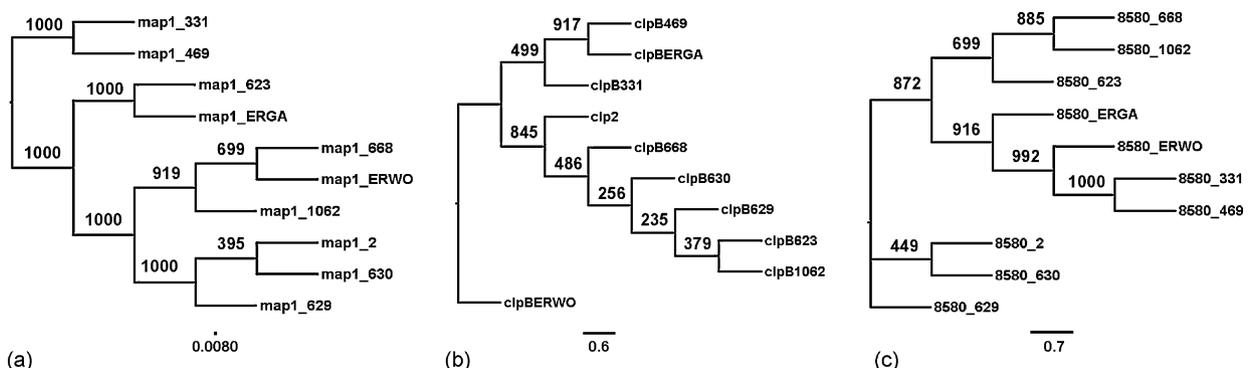
**Fig. 2.** Neighbor-joining trees of the individual MLST loci sequences: (a) *gtaA*; (b) *groEL*; (c) *lepA*; (d) *lipA*; (e) *lipB*; (f) *secY*; (g) *sodB*; (h) *sucA*; (i) concatenated sequences (identical to Fig. 1a). All trees are midpoint rooted. Midpoint and outgroup rooting using *E. canis* str. Jake, *E. chaffeensis* str. Arkansas or *A. marginale* str. St Maries orthologous sequences yielded the exact same trees. Bootstrap values were calculated for nodes.

hitchhiking – population expansion). Data obtained from Fu's  $F_s$  test did not correlate with those from other tests (Table 2). These discrepancies and lack of significance with all the neutrality tests applied, added to the previous observation that the variability was borne by few strains on few loci, led to the implementation of the same neutrality tests while excluding in turn two sets of strains (Table 2). The first set to be excluded was the reference strains ERGA and ERWO. The second set of strains to be excluded was

strains 331, 469 and 623 which carried most of the genetic variability (Supplemental Table 3, Fig. 2).

### 3.6. Assessment of the presence of recombination events

The various tests mentioned above are specifically susceptible to recombination which could easily blur the whole figure. Furthermore, this could also explain the contradictory tree



**Fig. 3.** Neighbor-joining trees of the individual reference genes: (a) *map1*; (b) *clpB*; (c) CDS8580. All trees are midpoint rooted. Midpoint and outgroup rooting using *E. canis* str. Jake, *E. chaffeensis* str. Arkansas or *A. marginale* str. St Maries orthologous sequences yielded the exact same trees. Bootstrap values were calculated for nodes.

**Table 2**  
Assessment of selective pressure on three subsets of *E. ruminantium* strains.

	Full set (10 strains)				$\Delta$ ERGA–ERWO (8 strains)				$\Delta$ 331/469/623 (7 strains)				
	Tajima's D	F*	D*	Fu's Fs	Tajima's D	F*	D*	Fu's Fs	Tajima's D	F*	D*	D*	Fu's Fs
gltA	-0.1691 <sup>NS</sup>	0.8713 <sup>NS</sup>	1.0846 <sup>NS</sup>	3.493	0.5448 <sup>NS</sup>	1.4025 <sup>NS</sup>	1.48571 <sup>*</sup>	5.660	-1.00623 <sup>NS</sup>	-1.10146 <sup>NS</sup>	-1.04881 <sup>NS</sup>		-0.095
groEL	-1.4919 <sup>NS</sup>	-1.8088 <sup>NS</sup>	-1.9533 <sup>NS</sup>	4.152	-1.39245 <sup>NS</sup>	-1.44979 <sup>NS</sup>	-1.60206 <sup>NS</sup>	3.162	0.86284 <sup>NS</sup>	1.47500 <sup>NS</sup>	1.49166 <sup>*</sup>		4.745
lepA	0.9377 <sup>NS</sup>	1.4291 <sup>NS</sup>	1.3821 <sup>NS</sup>	1.567	-0.41500 <sup>NS</sup>	-0.07272 <sup>NS</sup>	0.03986 <sup>NS</sup>	0.583	1.04431 <sup>NS</sup>	1.05865 <sup>NS</sup>	0.95066 <sup>NS</sup>		2.167
lipA	-0.2792 <sup>NS</sup>	0.5790 <sup>NS</sup>	0.7749 <sup>NS</sup>	1.955	0.15875 <sup>NS</sup>	0.78299 <sup>NS</sup>	0.87595 <sup>NS</sup>	1.981	-0.38255 <sup>NS</sup>	-0.41257 <sup>NS</sup>	-0.36923 <sup>NS</sup>		4.780
lipB	0.5327 <sup>NS</sup>	0.9210 <sup>NS</sup>	0.9121 <sup>NS</sup>	-1.917	0.67142 <sup>NS</sup>	1.02698 <sup>NS</sup>	1.00248 <sup>NS</sup>	-0.221	0.34514 <sup>NS</sup>	0.21071 <sup>NS</sup>	0.14917 <sup>NS</sup>		-2.200
secY	0.0637 <sup>NS</sup>	0.6745 <sup>NS</sup>	0.7753 <sup>NS</sup>	1.836	0.58311 <sup>NS</sup>	1.11204 <sup>NS</sup>	1.12603 <sup>NS</sup>	2.355	-0.33025 <sup>NS</sup>	-0.39280 <sup>NS</sup>	-0.37037 <sup>NS</sup>		-0.226
sodB	0.0224 <sup>NS</sup>	1.0547 <sup>NS</sup>	1.2391 <sup>NS</sup>	1.194	0.48523 <sup>NS</sup>	1.23248 <sup>NS</sup>	1.31251 <sup>NS</sup>	1.243	-1.00623 <sup>NS</sup>	-1.10146 <sup>NS</sup>	-1.04881 <sup>NS</sup>		-0.095
sucA	0.6637 <sup>NS</sup>	0.6564 <sup>NS</sup>	0.5558 <sup>NS</sup>	-0.132	0.86894 <sup>NS</sup>	0.61328 <sup>NS</sup>	0.45537 <sup>NS</sup>	0.265	-1.09826 <sup>NS</sup>	-1.23869 <sup>NS</sup>	-1.13318 <sup>NS</sup>		-1.037
clpB	-0.9635 <sup>NS</sup>	-1.1435 <sup>NS</sup>	-1.2372 <sup>NS</sup>	0.765	-1.21921 <sup>NS</sup>	-1.27919 <sup>NS</sup>	-1.13595 <sup>NS</sup>	0.702	-1.10893 <sup>NS</sup>	-1.16817 <sup>NS</sup>	-1.05316 <sup>NS</sup>		0.812
map1	0.4800 <sup>NS</sup>	1.3210 <sup>NS</sup>	1.3988 <sup>NS</sup>	6.052	0.18612 <sup>NS</sup>	0.50172 <sup>NS</sup>	0.52969 <sup>NS</sup>	8.104	0.48009 <sup>NS</sup>	1.32105 <sup>NS</sup>	1.39881 <sup>*</sup>		6.052
CDS8580	-0.0967 <sup>NS</sup>	0.3926 <sup>NS</sup>	0.4957 <sup>NS</sup>	1.892	0.66954 <sup>NS</sup>	1.45663 <sup>NS</sup>	1.50216 <sup>*</sup>	3.169	-1.01530 <sup>NS</sup>	-1.11759 <sup>NS</sup>	-1.01143 <sup>NS</sup>		3.065

NS: statistically not significant.

\* Statistically significant at  $P < 0.05$ .

topologies observed in Figs. 2 and 3. The putative presence of recombination events was therefore analyzed on the same three sets of strains: the full set (all 10 strains), after exclusion of ERGA and ERWO ( $\Delta$  ERGA–ERWO) and after exclusion of strains 331, 469 and 623 ( $\Delta$  331–469–623). These data are summarized in Table 3. Putative recombination was assessed both by estimating the minimum number of recombination, Rm, for each locus according to the four gametes model (Hudson and Kaplan, 1985). The presence of codons with multiple evolutionary pathways was also detected. When analyzing the full set of sequences, which includes the reference strains ERGA and ERWO as well as the strains 331, 469 and 623, some elements immediately appeared. At first, several recombination events were detected among the MLST loci as well as among the reference loci. The loci bearing the highest number of recombination events were *map1* and CDS8580 (Table 3). Considering the second set of strains, excluding ERGA and ERWO, data were almost the same as with the whole set of loci for the MLST targets, the only difference being a slight increase of  $\theta$  (theta). A difference however appeared in the case of *map1* and CDS8580 with a decrease of Rm. As expected, the removal of strains 331/469/623 eliminated almost all polymorphism. Rm and MEP were also affected but differently depending on the locus (Table 3). In the case of *groEL*, Rm became null indicating that these strains bore the detected recombination events. The maximum effect was observed with *lipA* for which all polymorphism was eliminated. *sucA* also displayed a decrease of Pa whereas S and  $\eta$  remained the same, which was indicative of an excess of singletons and an expansion of population. Some recombination events were detected in this set indicating that if most of the detected recombination events were associated with the three strains removed, i.e. 331, 469 and 623, other events were associated with the remaining strains. *map1* was highly affected by recombination and half of these detected events were present in the subset comprising strains 331, 469 and 623.

### 3.7. Assessment of linkage disequilibrium

Linkage disequilibrium is the non-random association of two sites and is therefore expected to be affected by recombination. Table 4 presents the same subsets of strains as in Tables 2 and 3, but with an assessment of linkage disequilibrium. Significant associations are found but following a differing pattern depending upon strains and loci (Table 4). When the whole set of strains is considered, variation in congruence between loci seems to follow the distribution of putative recombination (Table 3) the lowest level of congruence being associated to the loci for which recombination was detected. Congruence seems to increase for *groEL* and *lipA* when strains 331, 469 and 623 are removed. Congruence decreases for *secY* and *lipB* when these same three

strains are removed confirming that for these loci polymorphism is on the remaining strain. Linkage disequilibrium seems to be more evenly distributed for *sucA* and the three reference loci, i.e. *clpB*, *map1* and CDS8580 (Table 4).

## 4. Discussion

The data reported here on the development of a MLST scheme for *E. ruminantium* is a step forward to precise genotyping of this bacterium. This MLST scheme is capable of clearly discriminating between even closely related strains, such as for instance strains 331 and 469. The use of MLST coupled with the results of cross-protection assays will allow to define protective groups of strains and to identify putative vaccinal strain. PCR-based method are efficient for detecting *E. ruminantium* in hosts and ticks (Peter et al., 1995) and to characterize isolates of *E. ruminantium* from related species (Allsopp et al., 1997; Bekker et al., 2002; Vachieri et al., 2008; Raliniaina et al., in press) but routine strain-specific diagnosis is not yet achieved. This work is the first MLST scheme developed on *E. ruminantium* although preliminary investigations have been previously reported (Allsopp and Allsopp, 2007).

A critical point in developing this MLST approach was not only to obtain accurate genotyping markers but also to address issues such as genome plasticity, recombination and genetic background of the circulating strains. These phenomena are clearly affecting strains evolution and pathogenicity and must be addressed more in depth. *E. ruminantium* has been reported to display a dynamic mechanism of genome size plasticity (Frutos et al., 2006, 2007). Similarly, recombination has been reported and is suspected to be involved in genetic diversity (Hughes and French, 2007; Allsopp and Allsopp, 2007; Bekker et al., 2005). However, recombination in *E. ruminantium* has never been studied on a relevant set of strains. Hughes and French (2007) have addressed it from a statistical and evolutionary standpoint but only on the three published genomes (Collins et al., 2005; Frutos et al., 2006). Allsopp and Allsopp (2007) have reported massive intergenome recombination events in *E. ruminantium* strains in the field, but solely based upon DNA alignments and without statistics on DNA polymorphism and evolutionary trends. This work was therefore designed to specifically address both fine genotyping and DNA polymorphism statistical analysis on a well defined sampling allowing for identification of evolutionary trends.

The strain ERWO isolated from South Africa was reported to be highly divergent from strain ERGA and from strains from Western Africa (Allsopp and Allsopp, 2007). A lack of recombination between Western and South-Eastern African strains was also reported (Allsopp and Allsopp, 2007) and that ERGA was clearly different and representative of ancestral line. The data reported here are in contradiction with this conclusion. Strains closely

**Table 3**  
Assessment of recombination events in three subsets of *E. ruminantium* strains.

Hp	Full set (10 strains)							Δ ERGA-ERWO (8 strains)							Δ 331/469/623 (7 strains)									
	Pa	S	η	θ	π	Rm	MEP	Hp	Pa	S	η	θ	π	Rm	MEP	Hp	Pa	S	η	θ	π	Rm	MEP	
<i>gltA</i>	3	10	10	10	3.400	0.00861	0	0	2	9	9	9	3.857	0.00976	0	0	2	0	1	1	0.286	0.00072	0	0
<i>groEL</i>	5	8	44	47	11.533	0.02580	3	3	5	8	44	47	13.429	0.03004	3	3	2	7	7	7	11.533	0.00746	0	0
<i>lepA</i>	4	7	7	7	3.022	0.00636	0	0	4	4	7	7	2.464	0.00519	0	0	3	5	6	6	2.952	0.00622	0	0
<i>lipA</i>	3	5	6	6	1.978	0.00580	0	0	3	5	6	6	2.393	0.00702	0	0	1	0	0	0	0.00000	0.00000	0	0
<i>lipB</i>	8	11	13	13	5.133	0.01434	3	1	6	11	13	13	5.679	0.01586	3	1	6	4	7	7	3.048	0.00851	1	0
<i>secY</i>	6	20	25	25	8.956	0.01526	2	0	5	20	23	23	9.857	0.01679	2	0	4	2	5	5	1.905	0.00324	1	0
<i>sodB</i>	3	4	4	4	1.422	0.00343	0	0	3	4	4	4	1.714	0.00413	0	0	2	0	1	1	0.286	0.00069	0	0
<i>sucA</i>	7	12	16	16	6.467	0.01613	1	0	6	11	16	16	7.214	0.01799	1	0	6	3	16	16	5.238	0.01306	1	0
<i>clpB</i>	5	4	13	13	3.622	0.00583	1	0	4	2	9	9	2.607	0.00420	0	0	4	2	10	10	3.238	0.00521	0	0
<i>map1</i>	7	141	149	159	61.622	0.09524	21	3	5	109	147	156	62.214	0.09616	18	3	5	42	97	100	38.143	0.05895	10	5
<i>CDS 8580</i>	7	32	40	44	15.244	0.05863	7	1	5	30	31	31	13.464	0.05179	1	0	4	6	25	26	8.714	0.05863	3	0

HP: haplotypes; PA: parcimony informative sites; S: polymorphic sites; η: total number of mutations; θ: Watterson's mutation parameter; π: nucleotide diversity; Rm: minimal recombination events; MEP: codon with multiple evolutionary pathway.

**Table 4**  
Assessment of linkage disequilibrium in three subsets of *E. ruminantium* strains.

	Full set (10 strains)						Δ ERGA-ERWO (8 strains)						Δ 331/469/623 (7 strains)					
	ZnS	B	Q	PW	χ <sup>2</sup>	F	ZnS	B	Q	PW	χ <sup>2</sup>	F	ZnS	B	Q	PW	χ <sup>2</sup>	F
<i>gltA</i>	0.8056	0.7778	0.8000	45	36	36	1.0000	1.0000	1.0000	36	36	36	NA	NA	NA	0	0	0
<i>groEL</i>	0.7372	0.6750	0.6829	820	600	4	0.7461	0.6750	0.6829	820	567	1	1.0000	1.0000	1.0000	21	21	21
<i>lepA</i>	0.3899	0.1667	0.2857	21	4*	4	0.3988	0.1667	0.2857	21	4	3	0.5333	0.4000	0.5000	1	4	4
<i>lipA</i>	0.6759	0.6000	0.6667	15	10*	10	0.6825	0.6000	0.6667	15	10	10	NA	NA	NA	0	0	0
<i>lipB</i>	0.3061	0.0833	0.1538	78	25	12	0.3217	0.0833	0.1538	78	24	11	0.2343	0.3333	0.5714	21	4	1
<i>secY</i>	0.4666	0.3750	0.4400	300	155	88	0.5350	0.4091	0.4783	253	154	88	0.2786	0.2500	0.4000	10	1	0
<i>sodB</i>	0.5313	0.3333	0.5000	6	3*	3	0.5556	0.3333	0.5000	6	3*	3	NA	NA	NA	0	0	0
<i>sucA</i>	0.4276	0.2000	0.2500	120	56	36	0.4656	0.2000	0.2500	120	56	36	0.4656	0.2000	0.2500	120	46	0
<i>clpB</i>	0.2221	0.0833	0.1538	78	14	2	0.3001	0.1250	0.2222	36	6	1	0.3294	0.0000	0.0000	45	9	1
<i>map1</i>	0.3327	0.2391	0.2734	9591	2248	1862	0.3659	0.2628	0.2971	9453	3145	1576	0.3885	0.3763	0.4468	4371	1206	111
<i>CDS 8580</i>	0.3894	0.1429	0.1667	630	277	121	0.8252	0.7667	0.7742	465	352	351	0.4512	0.3478	0.4167	276	91	3

ZnS: ZnS statistic (Kelly, 1997). ZnS is the average of R<sup>2</sup> (Hill and Robertson, 1968) over all pairwise comparisons. B: Wall's B statistics (Wall, 1999); Q: Wall's Q statistics (Wall, 1999); PW: number of pairwise comparisons; χ<sup>2</sup>: number of significant associations tested by chi-square test; F: number of significant associations tested by a fisher exact test.

\* Significant associations confirmed by the Bonferroni procedure.

related to both ERGA and ERWO are currently circulating in Burkina-Faso which contradicts the previous statement that ERGA is a unique genotype accidentally brought to the Caribbean (Allsopp and Allsopp, 2007). It is also not possible to confirm that the South-Eastern African continent is the centre of diversity of *E. ruminantium* (Allsopp and Allsopp, 2007). The higher diversity observed in this region when compared to Western Africa might simply be due to a biased sampling.

Based on nucleotide substitution rates, ERGA and ERWO were reported to have diverged from a common ancestor between 2 and 4 million years ago (Hughes and French, 2007). The rationale behind this was a molecular clock associated to the *Ka/Ks* ratio and the assumption that accumulation and fixation of mutations through time is representative of the segregating polymorphisms and genetic distances. However, the work described here brings elements to modulate this conclusion.

Both ERGA and ERWO are not only closely related to strains currently circulating in Burkina-Faso but remained unchanged, especially in the case of ERGA, even through spatial and temporal isolation. This is in accordance with the paradigm of genomic stasis which characterizes obligate intracellular bacteria. Indeed, following initial genome size reduction and massive loss of genes as part of the adaptation to the intracellular lifestyle (Moran and Mira, 2001; Moran and Plague, 2004; Sallstrom and Andersson, 2005), the genome of cell parasites remain highly stable over long periods of time (Tamas et al., 2002; Klasson and Andersson, 2004). This seems to be the case also with *E. ruminantium*. The differences observed between ERGA and ERWO might not thus be due to regular accumulation over time but rather a consequence of the divergence process. Recombinations are most likely to have occurred at that time and to have blurred the phylogeny as observed in this work. Accumulation of mutations may also occur through pseudogenes decay. Pseudogenes content is very high in *E. ruminantium* (Collins et al., 2005; Frutos et al., 2006, 2007) and active creation (Collins et al., 2005) as well as decay (Frutos et al., 2007) have been demonstrated. Straight molecular clock cannot be used in the case of *E. ruminantium* and the reported dates of separation should be reconsidered. Furthermore, the *Ka/Ks* ratio used for establishing molecular clock and evolutionary trends is not a reliable marker of evolution at the infraspecific level. Indeed, the relationship between selection and *Ka/Ks* does not follow a monotonic function and it is therefore impossible to infer selection pressures, positive or negative, from *Ka/Ks* (Kryazhimskiy and Plotkin, 2008).

What is however clear from this work is that there is currently, at least in the sampled part of Burkina-Faso, simultaneous circulation of strains apparently in genomic stasis and others in a diverging state which bears most the observed diversity. Recombination is recorded in all groups and is a still active mechanism and most likely an important factor in *E. ruminantium* evolution. They have obviously accumulated mutations and undergone recombination, but it is not possible to say what triggered this active evolutionary state when compared to the other strains in genomic stasis. This might be also considered along with the previously reported active mechanism of genome size plasticity, although nothing is known either about it (Frutos et al., 2006, 2007). This work nevertheless suggests that the evolution of *E. ruminantium* is complex with apparently contradictory phenomena involved, i.e. genomic stasis and recombination, being present in co-circulating strains. However, this work is based on a limited number of strains and must be confirmed on a larger set of strains in the same sampling area to draw more definitive conclusions. Two further steps to consider will be to sample in the same region over several years and to extend it to other geographic regions.

## Acknowledgements

This work was supported by EC FP6 INCO-DEV grant "Epigenevac" FP6-003713. H. Adakal was supported by Agence Universitaire pour la Francophonie (AUF) mobility grant program and by mobility grants from CIRAD. The authors are grateful to B. Godelle (ISEM-UM2) for fruitful discussion and comments.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2009.08.003.

## References

- Adakal, H., Bada-Alambedji, R., Stachurski, F., 2004. Etude de la variabilité antigénique d'*Ehrlichia ruminantium* dans la région sud-ouest du Burkina Faso en vue de l'application d'une vaccination. *Revue Africaine de Santé et de Productions Animales* 2, 132–137.
- Ahmed, N., Devi, S.M., Valverde, M. de L., Vijayachar, P., Machang'u, R.S., Ellis, W.A., Hartskeerl, R.A., 2006. Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. *Ann. Clin. Microbiol. Antimicrob.* 5, 28 doi:10.1186/1476-0711-5.
- Allsopp, M.T.E.P., Allsopp, B.A., 2007. Extensive genetic recombination occurs in the field between different genotypes of *Ehrlichia ruminantium*. *Vet. Microbiol.* 124, 58–65.
- Allsopp, M., Visser, E.S., du Plessis, J.L., Vogel, S.W., Allsopp, B.A., 1997. Different organisms associated with heartwater as shown by analysis of 16S ribosomal RNA gene sequences. *Vet. Parasitol.* 71, 283–300.
- Allsopp, M.T.E.P., Dorfling, C.M., Maillard, J.C., Bensaïd, 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., Van Heerden, H., Steyn, H.C., Allsopp, B.A., 2003. Phylogenetic relationships among *Ehrlichia ruminantium* isolates. *Ann. N.Y. Acad. Sci.* 990, 685–691.
- 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, 406–416.
- Bekker, C.P.J., De Vos, S., Taoufik, A., Sparagano, O.A.E., Jongejan, F., 2002. Simultaneous detection of *Anaplasma* and *Ehrlichia* species in ruminants and detection of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks by reverse line blot hybridization. *Vet. Microbiol.* 89, 223–238.
- Bekker, C.P.J., 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.
- 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 Striip, M.F., Zweggarth, E., Jongejan, F., Maillard, J.C., Berthier, D., Botha, M., Joubert, F., Corton, C.H., Thomson, N.R., Allsopp, T.M., Allsopp, B.A., 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable number. *Proc. Natl. Acad. Sci. U.S.A.* 102, 838–843.
- Collins, N.E., Pretorius, A., van Kleef, M., Brayton, K.A., Allsopp, T.M., Zweggarth, E., Allsopp, B.A., 2003b. Development of improved attenuated and nucleic acid vaccines for heartwater. *Dev. Biol.* 114, 121–136.
- Collins, N.E., Pretorius, A., Van Kleef, M., Brayton, K.A., Zweggarth, E., Allsopp, B.A., 2003a. Development of improved vaccines for heartwater. *Ann. N.Y. Acad. Sci.* 990, 474–484.
- Du Plessis, J.L., Van Gas, L., Olivier, J.A., Bezuidenhout, J.D., 1989. The heterogeneity of *Cowdria ruminantium* stocks: cross-immunity and serology in sheep and pathogenicity to mice. *Onderstepoort. J. Vet. Res.* 56, 195–201.
- Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3, 87–112.
- Faburay, B., Jongejan, F., Taoufik, A., Ceesay, A., Geysen, D., 2007. Genetic diversity of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks and small ruminants in The Gambia determined by restriction fragment profile analysis. *Vet. Microbiol.* 126, 189–199.
- Frutos, R., Viari, A., Ferraz, C., Morgat, A., Eychenié, S., Kandassamy, Y., Chantal, I., Bensaïd, A., Coissac, E., Vachieri, 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., Vachieri, N., Boyer, F., Martinez, D., 2007. *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends Parasitol.* 23, 414–419.
- Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* 41, 95–98.

- Harbottle, H., White, D.G., McDermott, P.F., Walker, R.D., Zhao, S., 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. *J. Clin. Microbiol.* 44, 2449–2457.
- Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231.
- Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*. 111, 147–164.
- Hughes, A.L., French, J.O., 2007. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. *Gene* 387, 31–37.
- Inokuma, H., Brouqui, P., Drancourt, M., Raoult, D., 2001. Citrate synthase gene sequence: a new tool for phylogenetic analysis and identification of *Ehrlichia*. *J. Clin. Microbiol.* 39, 3031–3039.
- Jongejan, F., 1991. Protective immunity to heartwater (*Cowdria ruminantium* infection) is acquired after vaccination with in vitro-attenuated *Rickettsia*. *Infect. Immun.* 59, 729–731.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21–132.
- Kelly, J.K., 1997. A test of neutrality based on interlocus associations. *Genetics* 146, 1197–1206.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA, USA.
- Klasson, L., Andersson, S.G.E., 2004. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* 12, 37–43.
- Kryazhimskiy, S., Plotkin, J.B., 2008. The population genetics of dN/dS. *PLoS Genet.* 4, e1000304.
- Kurtz, A., Phillippy, A.L., Delcher, M., Smoot, M., Shumway, C., Antonescu, S.L., Salzberg, 2004. Fast algorithms for large-scale genome alignment and comparison. *Genome Biol.* 5, R12.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Mahan, S.M., Peter, T.F., Simbi, B.H., Burridge, M.J., 1998. PCR detection of *Cowdria ruminantium* infection in ticks and animals from heartwater-endemic regions of Zimbabwe. *Trop. Vet. Med.* 849, 85–87.
- Mahan, S.M., Barbet, A.F., Burridge, M.J., 2003. Development of improved vaccines for heartwater. *Dev. Biol.* 114, 137–145.
- Mahan, S.M., Waghead, S.D., McGuire, T.C., Rurangirwa, F.R., Wassink, L.A., Barbet, A.F., 1992. A cloned DNA probe for *Cowdria ruminantium* hybridizes with eight heartwater strains and detects infected sheep. *J. Clin. Microbiol.* 30, 981–986.
- Martinez, D., Vachieri, N., Stachurski, F., Kandassamy, Y., Raliniaina, M., Aprelon, R., Gueyve, A., 2004. Nested PCR for the detection and genotyping of *Ehrlichia ruminantium*. Use in genetic diversity analysis. *Ann. N.Y. Acad. Sci.* 1026, 106–113.
- Martinez, D., Maillard, J.C., Coisne, S., Sheikboudou, C., Bensaïd, 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., Perez, J.M., Sheikboudou, C., Debus, A., Bensaïd, A., 1996. Comparative efficacy of Freund's and Montanide ISA50 adjuvants for the immunisation of goats against heartwater with inactivated *Cowdria ruminantium*. *Vet. Parasitol.* 67, 175–184.
- McCombie, R.L., Finkelstein, R.A., Wood, D.E., 2006. Multilocus sequence typing of historical *Burkholderia pseudomallei* isolates collected in Southeast Asia from 1964 to 1967 provides insight into the epidemiology of melioidosis. *J. Clin. Microbiol.* 44, 2951–2962.
- Molia, S., Frebling, M., Vachieri, N., Pinarello, V., Petitclerc, M., Rousteau, A., Martinez, D., Lefrançois, T., 2008. *Amblyomma variegatum* in cattle in Marie Galante, French Antilles: prevalence, control measures, and infection by *Ehrlichia ruminantium*. *Vet. Parasitol.* 153, 338–346.
- Moran, N.A., Mira, A., 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* 2, 0054.1–0054.12.
- Moran, N.A., Plague, G.R., 2004. Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.* 14, 627–633.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Peter, T.F., Deem, S.L., Barbet, A.F., Norval, R.A., Kimby, B.H., Kelly, P.J., Mahan, S.M., 1995. Development and evaluation of PCR assay for detection of low levels of *Cowdria ruminantium* infection in *Amblyomma* ticks not detected by DNA probe. *J. Clin. Microbiol.* 33, 166–172.
- Raliniaina, M., Meyer, D., Pinarello, V., Sheikboudou, C., Emboulé, L., Kandassamy, Y., Adakal, H., Stachurski, F., Martinez, D., Lefrançois, T., Vachieri, N., in press. Mining the genetic diversity of *Ehrlichia ruminantium* using *map* genes family. *Vet. Parasitol.*
- Reddy, G., Sulsona, C.R., Harrison, R.H., Mahan, S.M., Burridge, M.J., Barbet, A.F., 1996. Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. *Clin. Diagn. Lab. Immunol.* 3, 417–422.
- Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.
- Sallstrom, B., Andersson, S.G., 2005. Genome reduction in the alpha-Proteobacteria. *Curr. Opin. Microbiol.* 8, 579–585.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A., Andersson, S.G.E., 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296, 2376–2379.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Vachieri, N., Maganga, G., Lefrançois, T., Kandassamy, Y., Stachurski, F., Adakal, H., Ferraz, C., Morgat, A., Bensaïd, A., Coissac, E., Boyer, F., Demaille, J., Viari, A., Martinez, D., Frutos, R., 2008. Differential strain-specific diagnosis of the heartwater agent: *Ehrlichia ruminantium*. *Infect. Genet. Evol.* 2008 8, 459–466.
- Van Vliet, A.H., van der Zeijst, B.A., Camus, E., Mahan, S.M., Martinez, D., Jongejan, F., 1995. Use of a specific immunogenic region on the *Cowdria ruminantium* MAP1 protein in a serological assay. *J. Clin. Microbiol.* 33, 2405–2410.
- Waghead, S.D., Rurangirwa, F.R., Mahan, S.M., Yunker, C.E., Crawford, T.B., Barbet, A.F., Burridge, M.J., McGuire, T.C., 1991. A cloned DNA probe identifies *Cowdria ruminantium* in *Amblyomma variegatum* ticks. *J. Clin. Microbiol.* 29, 2571–2577.
- Wall, J.D., 1999. Recombination and the power of statistical tests of neutrality. *Genet. Res.* 74, 65–69.
- Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* 7, 256–276.
- Zweygarth, E., Josemans, A.L., Van Strijp, M.F., Lopez-Rebollar, L., Van Kleef, M., Allsopp, B.A., 2005. An attenuated *Ehrlichia ruminantium* (Welgevonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* 23, 1695–1702.