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

Hassane Adakal, Damien F. Meyer, Catherine Carasco-Lacombe, Valérie Pinarelo, Florian Allègre, Karine Huber, Frederic Stachurski, Serge Morand, Dominique Martinez, Thierry Lefrançois, Nathalie Vachiery, Roger Frutos.

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 Subsaharan Africa, in Madagascar, and in some Caribbean islands from where it represents a threat to the American mainland (Barre 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; Zweygarth 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 genetic diversity of strains in field-collected strains (Martinez et al., 2001). However, although *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, ...
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' (ACTAGTAGAATTGGCAACATCAT) and AB130' (RCTDGCGGTCT-TYTGTTCAGCTAK) were used as external primers whereas AB128' and AB129' (TGATACTTGWGGRCGRDARTCCTT) 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) (MgCl2 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 of PCR amplifications were performed in 50 μl reaction buffer (Qiagen) (MgCl2 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).

2.4. 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) (MgCl2 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/) 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 (θ), 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 (η) or from S (Watterson, 1975). Eta (η) is the total number of mutations, and S is the number of segregating (polymorphic) sites. Kd (the number of nonsynonymous substitutions per nonsynonymous site) and Ks (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...
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 a 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 *E. canis* and *E. chaffeensis* sequences. 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. Haplotype 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 1000. Strain 1062 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
similar phylogenetic analysis was conducted on each locus considered separately. For each locus, trees were constructed with both midpoint rooting and outgroup rooting using the relevant orthologous genes from *E. chaffeensis*, *E. canis* or *A. marginale* as an outgroup. In contrast to distributions obtained with concatenated sequences, identical results were obtained for each locus with both midpoint and outgroup rooting. Therefore, only midpoint rooting distribution was shown in Fig. 2. Tree topologies were different to those observed for the concatenated sequences. Furthermore, different topologies were observed for each locus (Fig. 2) with the exception of *gltA*, *lipA* and *sodB*. Although no common phylogenetic pathway could be inferred from these individual locus alignments, some conserved features were still present, the main one being the recurrent clustering of strains 331 and 469 apart from all the other strains. The only exception to this trend was the *groEL* locus for which strains 331 and 469 were separated within different clusters. Interestingly, *groEL* displayed a trichotomy in which strains 331 and 469 occupied a different branch whereas the third corresponded to strain 623, the third most diverging strain (Fig. 2). The latter also branched separately in most of the single-locus trees, although to a lower extent when compared to strains 331 and 469. To further investigate this aspect, MLST loci were compared to three other loci which have been previously shown through whole-genome comparative genomic analysis to display a statistically significant bias of synonymous/nonsynonymous substitutions (Frutos et al., 2006). Two loci, the major antigenic protein gene *map1* and the chaperone gene *clpB* were characterized by a distorted excess of synonymous substitutions (purifying selection) whereas a third one, CDS8580, was characterized by a significant bias towards nonsynonymous substitutions (positive selection). *map1* and CDS8580, displayed a highly structured clustering characterized by high bootstrap values (Fig. 3). Interestingly, these two loci also discriminated the 331/469 strains as a separate clade. *clpB*, behaved in a similar way as the MLST loci although yielding a different topology (Fig. 3).

3.5. DNA sequence polymorphism and selective pressure

To investigate this diversity in the topology of phylogenetic trees, the different sequences were tested for polymorphism and selective pressure. Comparative analysis of the DNA sequence polymorphism parameters is presented in Supplementary Table 4. *Ka/Ks* (dn/ds) ratio for the MLST loci was rather low with values ranging from 0.034 to 0.139. The reference loci *map1* and *clpB* display *Ka/Ks* ratio of 0.146 and 0.048, respectively. CDS8580 displayed a very high *Ka/Ks* value of 1.491. Several neutrality tests, i.e. Tajima’s D, Fu and Li’s F* and D* and Fu’s Fs, were also conducted on the loci sequences (Table 2). The most important feature with the Tajima’s D figures was that all tests, for all the loci considered, had no statistical significance indicating thus that the hypothesis of neutral evolution could not be rejected in any of the cases. Fu and Li’s *D* and F* neutrality tests were also applied. Interestingly, the only two loci displaying negative *F* and *D* data were again the chaperone *groEL* and *clpB*, but even more interestingly, none of these tests was statistically significant, indicating again, that the hypothesis of neutral evolution could not be rejected. A third kind of neutrality test, was therefore applied, Fu’s Fs which is based on haplotype frequency distribution for a given value of θ derived from the average number of pairwise nucleotide differences. Fu’s Fs was negative in the presence of an excess of rare alleles (excess of recent mutations – genetic
hitchhiking – population expansion). Data obtained from Fu's Fs
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

\begin{figure}
\centering
\includegraphics[width=\textwidth]{tree1}
\caption{Neighbor-joining trees of the individual MLST loci sequences: (a) gltA; (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.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{tree2}
\caption{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.}
\end{figure}
Table 2
Assessment of selective pressure on three subsets of E. ruminantium strains.

<table>
<thead>
<tr>
<th>locus</th>
<th>Tajima’s D</th>
<th>F*</th>
<th>D*</th>
<th>Fu’s Fs</th>
</tr>
</thead>
<tbody>
<tr>
<td>gltA</td>
<td>-0.1691NS</td>
<td>0.8713NS</td>
<td>1.0846NS</td>
<td>3.493</td>
</tr>
<tr>
<td>groEL</td>
<td>-1.4919NS</td>
<td>-1.8088NS</td>
<td>-1.9533NS</td>
<td>4.152</td>
</tr>
<tr>
<td>lipA</td>
<td>0.9377NS</td>
<td>1.4291NS</td>
<td>1.3821NS</td>
<td>1.567</td>
</tr>
<tr>
<td>lipB</td>
<td>-0.2792NS</td>
<td>0.5790NS</td>
<td>0.7749NS</td>
<td>1.955</td>
</tr>
<tr>
<td>secY</td>
<td>0.6037NS</td>
<td>0.6745NS</td>
<td>0.7753NS</td>
<td>1.836</td>
</tr>
<tr>
<td>sodB</td>
<td>0.0224NS</td>
<td>1.0547NS</td>
<td>1.293NS</td>
<td>1.194</td>
</tr>
<tr>
<td>sucA</td>
<td>0.6637NS</td>
<td>0.6564NS</td>
<td>0.5558NS</td>
<td>0.132</td>
</tr>
<tr>
<td>clpB</td>
<td>-0.9635NS</td>
<td>-1.1435NS</td>
<td>-1.2372NS</td>
<td>0.765</td>
</tr>
<tr>
<td>map1</td>
<td>0.4800NS</td>
<td>1.3210NS</td>
<td>1.3988NS</td>
<td>6.052</td>
</tr>
<tr>
<td>CDS8580</td>
<td>-0.0967NS</td>
<td>0.3926NS</td>
<td>0.4957NS</td>
<td>1.892</td>
</tr>
</tbody>
</table>

NS: statistically not significant.
*Statistically significant at P < 0.05.

Table 3 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; Vachiery 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 et al., 1997; Bekker et al., 2002). 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.

<table>
<thead>
<tr>
<th>Hp</th>
<th>Δ ERGA-ERWO (8 strains)</th>
<th>Δ 331/469/623 (7 strains)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pa S η θ π RM MEP</td>
<td>Pa S η θ π RM MEP</td>
</tr>
<tr>
<td></td>
<td>Full set (10 strains)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gltA 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>groEL 5 8 44 47 11.531 0.02580 3 3 5 8 44 47 13.429 0.03004 3 3 2 7 7 7 11.333 0.00746 0 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lipA 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lipB 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>secY 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sodB 3 4 4 4 1.422 0.00636 0 0 3 4 4 4 1.714 0.00413 0 0 2 0 1 1 0.286 0.00069 0 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sucA 7 12 16 16 6.467 0.01613 0 0 6 11 16 16 7.214 0.01799 0 0 6 3 16 16 5.238 0.01306 1 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>clpB 5 4 13 13 3.622 0.00583 1 0 4 2 9 9 2.607 0.00420 0 0 4 2 5 5 1.905 0.00324 1 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>map1 7 141 149 159 61.622 0.00583 7 1 5 30 31 31 13.464 0.05179 1 0 4 6 25 26 8.714 0.05863 3 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDS 8580 7 32 40 44 15.244 0.00583 7 1 5 30 31 31 13.464 0.05179 1 0 4 6 25 26 8.714 0.05863 3 0</td>
<td></td>
</tr>
</tbody>
</table>

HP: haplotypes; PA: parsimony 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.

<table>
<thead>
<tr>
<th>Hp</th>
<th>Δ ERGA-ERWO (8 strains)</th>
<th>Δ 331/469/623 (7 strains)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ZnS B Q PW χ² F</td>
<td>ZnS B Q PW χ² F</td>
</tr>
<tr>
<td></td>
<td>Full set (10 strains)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gltA 0.8056 0.7778 0.8000 45 36 36 1.0000 1.0000 1.0000 36 36 36 NA NA NA 0 0 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>groEL 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lipA 0.3899 0.1667 0.2857 21 4’ 4 0.3988 0.1667 0.2857 21 4 4 0.5333 0.4000 0.5000 1 4 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lipB 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>secY 0.4666 0.3750 0.4400 300 155 88 0.5556 0.3333 0.5000 6 3’ 3 0.2786 0.2500 0.4000 10 1 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sodB 0.5313 0.3333 0.5000 6 3 3 0.5556 0.3333 0.5000 6 3 3 0.2343 0.3333 0.5714 21 4 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sucA 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>clpB 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>map1 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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDS 8580 0.3894 0.1429 0.1667 630 277 121 0.8252 0.7667 0.7724 465 352 351 0.4512 0.3478 0.4167 276 91 3</td>
<td></td>
</tr>
</tbody>
</table>

ZnS: ZnS statistic (Kelly, 1997). ZnS is the average of $R^2$ (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; $\chi^2$: 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 “Epigenecav” 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**


