Fine-Scale Phylogeographic Structure of \textit{Borrelia lusitaniae} Revealed by Multilocus Sequence Typing

Liliana R. Vitorino$^{1,2,3}$, Gabriele Margos$^2$, Edward J. Feil$^2$, Margarida Collares-Pereira$^3$, Libia Zé-Zé$^{1,4}$, Klaus Kurtenbach$^2$

1 Departamento de Biologia Vegetal/Centro de Genética e Biologia Molecular, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa, Portugal, 2 Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom, 3 Unidade de Leptospirose e Borreliose de Lyme, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal, 4 Centro de Estudos de Vectores e Doenças Infecciosas, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisboa, Portugal

Abstract

\textit{Borrelia lusitaniae} is an Old World species of the Lyme borreliosis (LB) group of tick-borne spirochetes and prevails mainly in countries around the Mediterranean Basin. Lizards of the family \textit{Lacertidae} have been identified as reservoir hosts of \textit{B. lusitaniae}. These reptiles are highly structured geographically, indicating limited migration. In order to examine whether host geographic structure shapes the evolution and epidemiology of \textit{B. lusitaniae}, we analyzed the phylogeographic population structure of this tick-borne bacterium using a recently developed multilocus sequence typing (MLST) scheme based on chromosomal housekeeping genes. A total of 2,099 questing nymphal and adult \textit{Ixodes ricinus} ticks were collected in two climatically different regions of Portugal, being \approx 130 km apart. All ticks were screened for spirochetes by direct PCR. Attempts to isolate strains yielded 16 cultures of \textit{B. lusitaniae} in total. Uncontaminated cultures as well as infected ticks were included in this study. The results using MLST show that the regional \textit{B. lusitaniae} populations constitute genetically distinct populations. In contrast, no clear phylogeographic signals were detected in sequences of the commonly used molecular markers \textit{ospA} and \textit{ospC}. The pronounced population structure of \textit{B. lusitaniae} over a short geographic distance as captured by MLST of the housekeeping genes suggests that the migration rates of \textit{B. lusitaniae} are rather low, most likely because the distribution of mediterranean lizard populations is highly parapatric. The study underlines the importance of vertebrate hosts in the geographic spread of tick-borne microorganisms.

Introduction

Lyme borreliosis (LB) is a complex tick-borne zoonosis and the most frequent vector-borne disease of humans in the temperate zone of both the New and Old World. It is known as the tick-borne disease. While all these four species occur in central and northern parts of Portugal and Spain, \textit{B. lusitaniae} is the sole species of the LB group in southern Portugal and North Africa [12–14]. Lizards of the family \textit{Lacertidae} are now believed to be important reservoir hosts of \textit{B. lusitaniae} [15,16]. These reptiles are known to be highly structured phylogeographically, suggesting limited migration between populations from different localities [17–20]. This is likely to have implications for the evolution and epidemiology of \textit{B. lusitaniae}.

On the Iberian Peninsula several species of LB group spirochetes have been detected in \textit{Ixodes ricinus} ticks, mainly \textit{B. garinii}, \textit{B. afzelii}, \textit{B. valaisiana} and \textit{B. lusitaniae} [6–9]. \textit{B. garinii} and \textit{B. afzelii} are known to be pathogenic in humans. \textit{B. lusitaniae} has been shown to be pathogenic in laboratory mice [10] and has also been isolated from human patients [11].

While all these four species occur in central and northern parts of Portugal and Spain, \textit{B. lusitaniae} is the sole species of the LB group in southern Portugal and North Africa [12–14]. Lizards of the family \textit{Lacertidae} are now believed to be important reservoir hosts of \textit{B. lusitaniae} [15,16]. These reptiles are known to be highly structured phylogeographically, suggesting limited migration between populations from different localities [17–20]. This is likely to have implications for the evolution and epidemiology of \textit{B. lusitaniae}.

\textit{B. lusitaniae} spirochetes have commonly been typed using single loci, such as different intergenic spacer regions (IGS) [21,22] or the genes encoding the outer surface proteins \textit{A (ospA)} [23] and \textit{C (ospC)} [24]. However, single-locus approaches have drawbacks in terms of inferring evolutionary relationships among the microbial populations [25,26]. Multilocus sequence typing (MLST) [27] or multilocus sequence analysis (MLSA) [28] based on housekeeping genes are considered to
be the most powerful genotyping tools in studies of the population biology of microbial organisms. In order to infer possible processes that shape the evolution and epidemiology of *B. lusitaniae* at a finer geographic scale in Portugal, we evaluated whether this bacterium is structured phylogeographically. For this, we applied a recently developed MLST scheme based on chromosomal housekeeping genes of *B. burgdorferi* [29] to samples of *B. lusitaniae* obtained from two regions of Portugal, Mafra and Grândola (Figure S1).

In addition to MLST of the core genome, we analyzed the 5S–23S IGS, *ospA* and *ospC* of the *B. lusitaniae* samples. While phylogenetic analyses of *ospA* and *ospC* did not provide signals of geographic structuring of *B. lusitaniae*, the results obtained using MLST revealed that the *B. lusitaniae* populations from these two regions constitute genetically distinct subpopulations. This analysis, therefore, confirms the increased utility of multiple housekeeping genes for studies of the geographic population structure of LB group spirochetes and suggests an association between the population structure of the bacteria and that of their vertebrate hosts.

**Results**

Based on sequence analyses of multiple housekeeping genes (i.e. *clpA, clpX, nifS, pepX, pyrG, recG* and *rplB*) of the *B. lusitaniae* samples analyzed in this study (Table 1), 13 sequence types (STs) were defined by MLST, and no ST was observed in more than two samples (Table 2). Among the housekeeping genes, the highest sequence diversity was noted in *clpA, pepX* and *rplB*, which also revealed high numbers of alleles (Table 3). The *nifS* gene was the least polymorphic of the housekeeping genes analyzed, with a percentage of variable sites of 1.06, the lowest number of alleles and also the lowest level of nucleotide diversity per site (Table 3). The average ratios of non-synonymous and synonymous substitutions (dN/dS) of the housekeeping genes and *ospA* were <1, indicating that they are nearly neutral under or purifying selection (Table 3). The MLST data have been submitted to the MLST website hosted at Imperial College London, United Kingdom (www.mlst.net), and can be accessed via strain ID or ST. For *ospC*, the dN/dS ratio was >1, suggesting that the gene encoding this outer surface protein is under some level of positive selection [24,30].

The MLST tree generated in this study discriminates the Mafra samples of *B. lusitaniae* from the Grândola samples (Figure 1). The human isolate PoHL1 clusters together with samples from Mafra in 100% of the trees drawn from the posterior probability. Interestingly, PoTiBmP220, a strain detected in Mafra, arises from the branch representing the Grândola samples (see below). Signals of phylogenetic structuring were also found for the individual housekeeping genes and the IGS, but the intrapopulation phylogenies were less resolved (Figures S2, S3, S4, S5, S6, S7, S8, Figure 2). In contrast, the phylogenetic trees of *ospA* and *ospC* showed no clear signals of geographic structuring of the *B. lusitaniae* samples (Figures 3 and 4). For *ospC*, the lack of geographic structuring may be related to balancing selection and/or recombination. Consistent with this, apart from signatures of positive selection, several recombination events were detected in the *ospC* sequences using the RDP suite of programs. Recipient and donor strains, position in the alignment and P-values for the individual methods are shown in Table 4. Recombination events will influence the tree topology and may lead to the polytomies that are observed in the *ospC* tree (Figure 4). No recombination events were detected in *ospA* using RDP. (The sequences of the IGS, *ospA* and *ospC* have been deposited in the GenBank database under accession numbers EF179549 to EF179604.)

### Table 1. *B. lusitaniae* samples analyzed in this study.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Sample</th>
<th>Date of collection</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mafra</td>
<td>PoTIbL37</td>
<td>April 1999</td>
<td>[9]</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfP109</td>
<td>May 2004</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfP220</td>
<td>April 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfJ2</td>
<td>January 2001</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfJ50</td>
<td>January 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfP147</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbMfP364</td>
<td>December 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td>Grândola</td>
<td>PoTIbGr41</td>
<td>November 2002</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr82</td>
<td>November 2002</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr128</td>
<td>February 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr130</td>
<td>February 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr131</td>
<td>February 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr136</td>
<td>February 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr143</td>
<td>February 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr209</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr210</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr211</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr212</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr213</td>
<td>March 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr288</td>
<td>April 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr293</td>
<td>April 2003</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>PoTIbGr409</td>
<td>May 2003</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

* B. lusitaniae strains successfully cultured in BSKII medium.
* Samples excluded from the study due to multiple infections with different *Borrelia* strains.
* Sample excluded from the *ospC* analysis since no *ospC* sequence was obtained.

All the strains were detected in or isolated from *I. ricinus* ticks, except for one isolate PoHL1 that was obtained from a human skin biopsy. doi:10.1371/journal.pone.0004002.t001

An analysis of the pairwise divergences between the samples at the housekeeping genes and *ospC* also illustrates the difference between these loci. The distribution of pairwise differences of the concatenated housekeeping genes is bimodal (Figure 5A), with the peak at the lower distances representing intra-population comparisons and the peak at high distances representing inter-population comparisons. In contrast, although two peaks are still discernable in the distribution for *ospC*, these peaks are much less distinct (Figure 5B). When the same analysis was carried out for *ospA*, the distribution was notably bimodal (Figure S10A). However, this was predominantly due to the inclusion of two highly diverged strains at this locus, the human-derived strain PoHL1 and the tick isolate PoTIbL37, as the peak corresponding to large distances reflects comparisons involving one of these isolates. When these isolates were removed, the distribution was no longer bimodal (Figure S10B).

Although intragenic recombination was not detected in the individual housekeeping genes using the RDP suite of programs, a putative recombination event corresponding to the region of *clpX* was detected with RDP and Bootscan when the housekeeping gene sequences were concatenated (p = 0.019). Indeed, in the allelic profiles *clpX* allele 19 (Table 2) was found to be the only allele shared between ST69 from Mafra and ST64 and ST68 from Grândola (for alleles of strain PoTIbMfP220, see below). The alignment of the polymorphic sites of the concatenated house-
keeping genes demonstrated that clpX of ST64 (PoTiBGr136) and ST69 (e.g. PoHL1) had obviously recombined (Figure 6). To further investigate this, an analysis of the STs (as defined in Table 2) using ClonalFrame software confirmed a single recombination event for clpX on the branch above node D (Figures S11 and S12). At node D STLus3 and STLus4 are split, which correspond to samples PoTiBmfJ2, PoTiBmfJ50, and PoTiBmfP220 (Table 2). This finding indicates that the housekeeping genes were found to range from 0.0132 to 0.0137 (Table 5). This is lower than the genetic distance between the two regions, despite the pronounced phylogeographic signals suggesting that the two regional populations represent genetically distinct [29]. We have, furthermore, provided evidence that B. burgdorferi originated in Europe and not in North America [29], thereby supporting the rationale for considering the diverse B. lusitaniae populations analyzed in this study as conspecific.

### Discussion

MLST and MLSA are the most powerful tools for analyzing the evolution and population biology of microbial populations [28,31]. Most MLST/MLSA schemes used so far have been applied to directly transmitted pathogens. Because the majority of indirectly transmitted zoonotic microparasites are maintained by wildlife and vectors, such as ticks or mosquitoes, environmental factors are particularly important in shaping their evolution and spread. As this may result in geographic structuring, genotyping methods of vector-borne microbial organisms should have the power to capture phylogeographic structure and to infer species trees. Using a novel MLST scheme, we have recently demonstrated that North American and European populations of LB group spirochetes are genetically distinct [29]. We have, furthermore, provided evidence that B. burgdorferi originated in Europe and not in North America [29]. Here we analyzed B. lusitaniae samples from two climatically different regions of Portugal (Figure S1) using this MLST scheme, suggesting that the two regional B. lusitaniae populations represent genetically distinct lineages. In contrast, no robust phylogeographic signals were observed for ospA and ospC.

The numbers of B. lusitaniae strains that were successfully cultured differed remarkably between the two regions, despite the number of new alleles found. STs 60–69 are based on eight housekeeping genes including uvrA and were defined according to the MLST website, www.mlst.net. Allele numbers of uvrA for STs 60–69 can be found in the website under strain ID. For the five samples where no uvrA data were available, alleles for the seven remaining housekeeping genes were assigned allele numbers according to the website, however, STs were arbitrarily labelled Lus 1-4. Alleles of the IGS, ospA and ospC were assigned numbers in the order new alleles were found.

### Table 2. Allelic profiles and STs of B. lusitaniae.

<table>
<thead>
<tr>
<th>B. lusitaniae samples</th>
<th>clpA</th>
<th>clpX</th>
<th>nifS</th>
<th>pepX</th>
<th>pyrG</th>
<th>recG</th>
<th>rplB</th>
<th>ST</th>
<th>IGS</th>
<th>ospA</th>
<th>ospC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoTiBGr41</td>
<td>26</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>13</td>
<td>22</td>
<td>13</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PoTiBGr82</td>
<td>27</td>
<td>16</td>
<td>18</td>
<td>22</td>
<td>13</td>
<td>23</td>
<td>14</td>
<td>61</td>
<td>2</td>
<td>2</td>
<td>ND</td>
</tr>
<tr>
<td>PoTiBGr130</td>
<td>27</td>
<td>17</td>
<td>18</td>
<td>22</td>
<td>13</td>
<td>23</td>
<td>14</td>
<td>62</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>PoTiBGr131</td>
<td>28</td>
<td>18</td>
<td>18</td>
<td>21</td>
<td>14</td>
<td>22</td>
<td>13</td>
<td>63</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>PoTiBGr136</td>
<td>29</td>
<td>19</td>
<td>18</td>
<td>23</td>
<td>14</td>
<td>24</td>
<td>15</td>
<td>64</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>PoTiBGr409</td>
<td>29</td>
<td>19</td>
<td>18</td>
<td>23</td>
<td>14</td>
<td>24</td>
<td>15</td>
<td>64</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>PoTiBGr143</td>
<td>30</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>13</td>
<td>23</td>
<td>16</td>
<td>65</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>PoTiBGr211</td>
<td>30</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>13</td>
<td>23</td>
<td>16</td>
<td>65</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>PoTiBGr209</td>
<td>26</td>
<td>20</td>
<td>18</td>
<td>22</td>
<td>15</td>
<td>22</td>
<td>17</td>
<td>66</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>PoTiBGr213</td>
<td>26</td>
<td>20</td>
<td>18</td>
<td>22</td>
<td>15</td>
<td>22</td>
<td>17</td>
<td>66</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>PoTiBGr288</td>
<td>28</td>
<td>15</td>
<td>18</td>
<td>24</td>
<td>16</td>
<td>22</td>
<td>18</td>
<td>67</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>PoTiBGr293</td>
<td>26</td>
<td>19</td>
<td>18</td>
<td>22</td>
<td>17</td>
<td>23</td>
<td>19</td>
<td>68</td>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>PoHL1</td>
<td>31</td>
<td>19</td>
<td>19</td>
<td>25</td>
<td>18</td>
<td>25</td>
<td>20</td>
<td>69</td>
<td>10</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>PoTiBL37</td>
<td>31</td>
<td>19</td>
<td>19</td>
<td>25</td>
<td>18</td>
<td>25</td>
<td>20</td>
<td>69</td>
<td>11</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>PotiBmfp147</td>
<td>31</td>
<td>19</td>
<td>19</td>
<td>25</td>
<td>18</td>
<td>25</td>
<td>20</td>
<td>69</td>
<td>11</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>PotiBmfp220</td>
<td>32</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>14</td>
<td>26</td>
<td>14</td>
<td>Lus1</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>PotiBmfp220</td>
<td>33</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>14</td>
<td>26</td>
<td>14</td>
<td>Lus2</td>
<td>12</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>PotiBmfp50</td>
<td>33</td>
<td>21</td>
<td>20</td>
<td>26</td>
<td>18</td>
<td>25</td>
<td>20</td>
<td>Lus3</td>
<td>10</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>PotiBmfp364</td>
<td>34</td>
<td>22</td>
<td>21</td>
<td>27</td>
<td>19</td>
<td>25</td>
<td>21</td>
<td>Lus4</td>
<td>10</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>

ND: not determined.
doi:10.1371/journal.pone.0004002.t002
Table 3. Loci used for MLST and single locus typing.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Product</th>
<th>Primer sequence (5’–3’)</th>
<th>Amplicon size (bp)</th>
<th>MLST fragment size (bp)</th>
<th>G+C %</th>
<th>No. of alleles</th>
<th>dN/dS ratio</th>
<th>% VI</th>
<th>N/A</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>clpA</td>
<td>Clp protease subunit A</td>
<td>clpAF1240: GATAGATTCTTCCAGACAAAG</td>
<td>864</td>
<td>579</td>
<td>25.7</td>
<td>9</td>
<td>0.308</td>
<td>2.94</td>
<td>0.01122</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>clpAr2104: CAAAAAACATCTTATTTTCATTTTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clpX</td>
<td>Clp protease subunit X</td>
<td>clpXF403: AATGTGCCATTGCAATAGC</td>
<td>721</td>
<td>624</td>
<td>30.7</td>
<td>8</td>
<td>0.026</td>
<td>1.92</td>
<td>0.00529</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>clpXr1124: TTAAAGAGCTCTAAATAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nifS</td>
<td>Aminotransferase</td>
<td>nifSF1: ATGGAATTACCAAAATTTAAG</td>
<td>696</td>
<td>564</td>
<td>27.1</td>
<td>4</td>
<td>0.079</td>
<td>1.06</td>
<td>0.00446</td>
<td></td>
</tr>
<tr>
<td>pepX</td>
<td>Dipeptidyl aminopeptidase</td>
<td>pepXF449: TATCCAAACCTGGCAATCC</td>
<td>723</td>
<td>570</td>
<td>28.0</td>
<td>7</td>
<td>0.15</td>
<td>2.45</td>
<td>0.0080</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pepXR1172: GTCCATGCTATAGTTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyrG</td>
<td>CTP synthase</td>
<td>pyrGF448: GATATGAAATAATATTTATTG</td>
<td>742</td>
<td>603</td>
<td>31.2</td>
<td>7</td>
<td>0</td>
<td>1.49</td>
<td>0.00549</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pyrGR1190: CAACATTACGGGAAATTTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>recG2</td>
<td>DNA recombinase</td>
<td>recGF917: CTTTATAGGAGTTGGATATC</td>
<td>777</td>
<td>651</td>
<td>30.6</td>
<td>5</td>
<td>0.06</td>
<td>1.38</td>
<td>0.00353</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>recGR1649: GAAGTCCAAGAGCTCAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rplB</td>
<td>50S ribosomal protein L2</td>
<td>rplBF2: TGGGTATAGCATTTTATGC</td>
<td>758</td>
<td>624</td>
<td>36.1</td>
<td>9</td>
<td>0.085</td>
<td>2.40</td>
<td>0.00757</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rplBR760: GCTGTCACAAAGGAGACA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ospA</td>
<td>Outer surface protein A</td>
<td>Fw1: GACACTGCCTTCTGGATAGC</td>
<td>302</td>
<td>288</td>
<td>37.4</td>
<td>10</td>
<td>0.69</td>
<td>14.9</td>
<td>0.03050</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rv1: CITTCTTTTCCTTCCTTTGT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fw2: AAAAAAGAGCCGAAATACGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rv2: ATCAAGCTGTGCTATTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ospC</td>
<td>Outer surface protein C</td>
<td>Fw1: TGAAGAGAAATACCATAGTGGCA</td>
<td>470</td>
<td>376–385</td>
<td>35.7</td>
<td>14</td>
<td>2.24</td>
<td>36.6</td>
<td>0.14230</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rv1: TTTTTGAGATTGTCTGCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fw2: TGGCTGTGAAGAGGTTGAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rv2: GCCACACAGGGACTGTATAGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGS</td>
<td>5S–23S</td>
<td>N/A</td>
<td>Reference 21</td>
<td>225</td>
<td>151–183</td>
<td>18.6</td>
<td>12</td>
<td>N/A</td>
<td>9.23</td>
<td>0.01806</td>
</tr>
</tbody>
</table>

N/A: not applicable.
doi:10.1371/journal.pone.0004002.t003
fact that isolation attempts were made by the same person and method. While the bacterial populations may have subtle differences in metabolic requirements, previous studies have demonstrated that the infection prevalence of B. lusitaniae in ticks from Mafra is orders of magnitude lower than in those from Grândola [9,12]. This may explain the disparity in the number of isolates obtained. We, therefore, included infected ticks, from which the spirochetal genes of interest were amplified directly without prior culturing.

Although both Network analysis and allelic admixture analysis (ClonalFrame) indicated recombination events at one housekeeping gene (clpA), the overall ratio of recombination to mutation was very low, suggesting that the linear chromosome of B. lusitaniae is relatively clonal.

In another study the heterogeneity of B. lusitaniae was examined at a broader geographic scale compared with our study using ospA as marker [32], suggesting the existence of two major lineages in the Mediterranean Basin. According to that study, Italian and German strains form a 'European' lineage and the Portuguese strains PotiB1-3 and North African strains an 'African' lineage, with the Portuguese patient isolate PoHL1 being an exception as it was placed in the European clade. However, our findings indicate that the commonly used molecular marker ospA is not suitable for phylogeographic analyses of B. lusitaniae at a smaller geographic scale (Figure 3, Figure S9). The reasons for the lack of clear geographic signals contained in the ospA sequences remain unknown, since no recombination events were detected for this gene.

ospC is another popular molecular marker of LB group spirochetes [24,30,33]. As for ospA, however, analyses of ospC did not reveal signals of geographic structure of B. lusitaniae at a small geographic scale. Recombination and balancing selection are possible processes that homogenize the spatial frequency distribution of ospC alleles of B. lusitaniae, but either of these processes may generate a uniform geographic structure. As recombination has been detected and the dN/dS ratios of ospC were >1, we hypothesize that both processes shape the population structure of this gene.

The lack of geographic structuring observed at ospA and ospC may allow to draw the conclusion that the two B. lusitaniae populations analyzed in this study are spatially mixed and that this bacterial species is not structured phylogeographically at the scale analyzed. On the other hand, the fact that the two populations from Mafra and Grândola do not share STs strongly indicates that the two populations presently do not, or very rarely, migrate between the two regions. It is possible that the patterns seen at the outer surface protein genes reflect ancient events that arose in a continuously distributed ancestral population (discussed below).

Given that bacterial housekeeping genes typically evolve very slowly (rates of synonymous substitution per site and year $\sim 10^{-7}$), it is likely that the bacterial populations have been separated from each other for a long time. It is possible that the two geographic clades of B. lusitaniae as demarcated by MLST represent diverged descendants of a common ancestral population prevailing during past glacial maxima. Given that Mafra and Grândola are only $\sim$130 km apart, isolation of these B. lusitaniae populations by distance alone is an unlikely explanation for the observed genetic divergence. It is plausible to assume that these local populations diverged through vicariance, because climate change after the last Ice Ages has generated ecological barriers between Mafra and Grândola. There is palaeobotanic evidence that during the last glacial maximum most of Portugal was covered by temperate mixed forests [34], whereas the present day climate and vegetation of southern Portugal, but not that of Mafra, resembles that of the African Maghreb [9]. Postglacial ecological differences between Mafra and Grândola, including those imposed by more recent human activities, are likely to have shaped the population structure and biogeographic patterns of vertebrate host communities, in particular reptilian populations. Furthermore, the river Tejo is likely to act as firm present-day barrier to migration of terrestrial reptiles between Mafra and Grândola (Figure S1). A number of studies have, in fact, revealed that the reptilian populations in the Mediterranean Basin are highly structured genetically and that their distribution is parapatric [17–20]. Because lizards are now considered important (if not the exclusive) reservoir hosts of B. lusitaniae [14–16], their limited dispersal will affect the migration rates of B. lusitaniae, resulting in the observed fine-scale geographic structure of this tick-borne bacterium. Although I. ricinus ticks infected with B. lusitaniae may be dispersed rapidly over long distances when feeding on highly mobile hosts, such as migratory birds, this is unlikely to be an important process in the effective dispersal of B. lusitaniae. Feeding tick larvae apparently do not acquire B. lusitaniae from vertebrate species other than lizards. On the other hand, B. lusitaniae-infected nymphs that feed on long-distance migrants will give rise to questing adult ticks that subsequently feed on larger animals, such as deer, which are not reservoir competent for any of the species of the LB group of spirochetes [3,4]. Thus, only larvae and nymphs

that feed on lizards will maintain the cycles of *B. lusitaniae*. In other words, the migration rates of *B. lusitaniae* are determined by those of lizards.

In central and northern parts of the Iberian Peninsula *B. garinii*, *B. valaisiana* and *B. afzelii* have been recorded in addition to *B. lusitaniae* [7–9], a pattern of species richness that is similar to that recorded for Central Europe [3,35–37]. The presence of these species strongly suggests that rodents and birds are also involved in the ecology of LB in central and northern Portugal [4]. In contrast, *B. lusitaniae* is the only species of the LB group in southern Portugal and North Africa [12–14]. It is interesting to note that the infection prevalence of *B. lusitaniae* in southern Portugal and North Africa was found to be much higher than the overall infection prevalence of all species of the LB group taken together in other parts of the world, including the Mafra region of Portugal [3,9,12–14,37]. This might indicate the operation of the ‘dilution effect’ in central and northern Portugal due to a more diverse vertebrate community which *I. ricinus* ticks parasitize [38]. The apparent lack of *B. garinii*, *B. valaisiana* and *B. afzelii*, but high infection prevalence of *B. lusitaniae*, in southern Portugal and North Africa suggests that immature *I. ricinus* ticks in that region feed mainly on reptilian hosts, allowing for considerable amplification of *B. lusitaniae*.

Taken together, the study strongly supports the idea that levels and patterns of host specialization of vector-borne microparasites affect their emergence and geographic spread. Population and landscape genetic studies of other vector-borne systems are needed to test the generality of this idea.

**Materials and Methods**

**Tick collection and habitat description**

Questing *I. ricinus* ticks were collected between 2001 and 2004 by blanket dragging in sylvatic habitats in Mafra (Estremadura region, ~25 km north of Lisbon, Portugal; 1,598 nymphal ticks, 413 adult ticks) and in Grândola (Alentejo region, ~100 km south of Lisbon; 88 adult ticks [40 male, 48 female ticks] (Figure S1). The climate in Mafra is temperate and humid, influenced by the Atlantic. The dense woodland habitats consist of deciduous oaks (*Quercus faginea*), eucalyptus, pine and chestnuts with well developed herbage layers. The Mafra site was located inside a park that was

![Bayesian phylogenetic tree of *B. lusitaniae* strains based on the IGS.](https://www.plosone.org/figure?cid=17)

**Figure 2. Bayesian phylogenetic tree of *B. lusitaniae* strains based on the IGS.** The tree was rooted with *B. burgdorferi* strain B31. Posterior probabilities values are indicated to provide branch support. The scale bar represents 5% sequence divergence. *B. lusitaniae* samples derived from Grândola are highlighted.

doi:10.1371/journal.pone.0004002.g002
created in the 18th century and served as leisure and hunting site for the Portuguese royalty. Movements of large animals into and out of the park are restricted. The climate in Grândola is more Mediterranean and drier than that in Mafra. Cork oaks (Q. suber) are common [9].

Screening for spirochetes and culturing

After decontamination, the ticks were cut into two halves under aseptic conditions. One half was inoculated into BSK-II media to obtain isolates. The remaining halves of the ticks were analyzed for infection by direct PCR targeting the 5S-23S IGS of the spirochetes [21]. The samples were assigned to Borrelia species using restriction fragment length polymorphism as described previously [2]. One B. lusitaniae culture was obtained from the 2,011 ticks collected in Mafra, and 15 B. lusitaniae cultures from the 88 adult ticks collected in Grândola. While in Grândola B. lusitaniae is relatively abundant and the sole LB species found [12], its prevalence in Mafra is very rare which is reflected in the limited number of isolates obtained from this region [9]. Only uncontaminated cultures and a subset of ticks found to be infected with B. lusitaniae were included in this study (Table 1). Isolation of LB spirochetes is carried out in liquid media, and cloning procedures using subsurface plating on solid media are difficult to perform and not carried out routinely. Mixed infections were, therefore, excluded at the stage of sequence analysis (see below).

PCR and sequencing

MLST was performed on cultured isolates of B. lusitaniae and directly on some infections in ticks without prior isolation of the spirochetes. The original MLST scheme developed for B. burgdorferi by Margos and colleagues [29] comprised eight housekeeping genes, i.e. clpA, clpX, nifS, pepX, pyrG, recG, rplB and uvrA. For five tick-derived B. lusitaniae samples, uvrA could not be amplified using a single pair of PCR primers and, therefore, most analyses in this study were carried out without uvrA. In addition, ospA, ospC and the 5S-23 IGS were amplified and sequenced. The PCR primers used in this study are shown in Table 3.

For DNA preparation, cultured Borrelia strains (1×10^7 spirochetes) were centrifuged at 13,000 rpm for 20 min, resuspended in 1 ml of PBS buffer and heated to 100°C for 10 min. For PCR amplification of the housekeeping genes, a 1/1000 dilution of these preparations was used as DNA template. PCR reactions

---

**Figure 3. Bayesian phylogenetic tree of B. lusitaniae strains based on ospA.** The tree was rooted with B. burgdorferi strain B31. Posterior probabilities values are indicated to provide branch support. The scale bar represents 1% sequence divergence. B. lusitaniae samples derived from Grândola are highlighted.
doi:10.1371/journal.pone.0004002.g003
were performed in 50 μl volume of 1× reaction mix (BioMix Red, BIOLINE, United Kingdom), 25 pmol of each primer and 2.5 μl of template DNA. The amplification conditions were as follows: 2 min of initial denaturation at 95°C, then 40 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. The amplification was completed by a final step of 5 min at 72°C to allow complete extension of all PCR products.

PCR amplification of the housekeeping genes from tick lysates as templates were performed using HotstarTaq Mastermix (Qiagen, Germany) under the following conditions: initial 2 min of initial denaturation at 95°C, then 40 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. The amplification was completed by a final step of 5 min at 72°C to allow complete extension of all PCR products.

**Table 4.** Recombination at ospC of *B. lusitaniae* samples.

<table>
<thead>
<tr>
<th>Recipient</th>
<th>Donor</th>
<th>Position</th>
<th>Gscale</th>
<th>Method</th>
<th>Av P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoTimfJ2 (PotimfP147)*</td>
<td>PoTiGr130 (PotimfP147)*</td>
<td>40–178</td>
<td>0</td>
<td>GENECONV</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MaxChi</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chimaera</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SiScan</td>
<td>0.00062</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3SEQ</td>
<td>0.0054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LARD</td>
<td>0.28</td>
</tr>
<tr>
<td>PoHL1/PoTiB37</td>
<td>PoTimfJ2</td>
<td>271–368</td>
<td>5</td>
<td>GENECONV</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MaxChi</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chimaera</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3SEQ</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LARD</td>
<td>0.1</td>
</tr>
</tbody>
</table>

*In this analysis the status of strain PoTimfP147 was uncertain and designated as ‘minor parent’ (donor) or maybe ‘daughter’ (recipient).

Figure 4. Bayesian phylogenetic tree of *B. lusitaniae* strains based on ospC. The tree was rooted with *B. burgdorferi* strain B31. Posterior probabilities values are indicated to provide branch support. The scale bar represents 5% sequence divergence. *B. lusitaniae* samples derived from Grândola are highlighted.

doi:10.1371/journal.pone.0004002.g004
Figure 5. Distribution of pairwise genetic distance for *B. lusitaniae* housekeeping genes (A) and *ospC* (B).
doi:10.1371/journal.pone.0004002.g005

Figure 6. Variable sites for each housekeeping gene shown for ST64, ST69, STLus2, STLus3 and STLus4. The numbers above the sequences refer to the position in the concatenated alignment. Regions corresponding to the individual genes are separated by a line and gene names are given on the top. The likely recombination event between ST64 and ST69 in *clpX* is indicated by an arrow.
doi:10.1371/journal.pone.0004002.g006
denaturation at 95°C for 15 min, 10 cycles of 95°C for 30 s, 52°C for 30 s, 72°C for 1 min, and then 30 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 1 min, and a final elongation step of 72°C for 5 min.

In parallel with MLST of the core genome, the IGS, ospA and ospC were analyzed. To amplify the latter two genes, two sets of primers were designed for a nested PCR approach (Table 3). Amplification of the IGS and the ospC gene was carried out in a 50 μl reaction mixture containing 1 pmol of each primer, 200 μM (each) dATP, dGTP, dCTP and dTTP (Invitrogen, United States), 1.75 U of Taq polymerase (Invitrogen, United States), 2 mM MgCl₂, 0.5× BSA, 1× Taq buffer. To amplify the IGS, we used the primers and the PCR conditions described elsewhere [21]. For ospA and ospC, the first round of amplification was carried out using a touchdown protocol; after an initial denaturation step of 95°C for 5 min, 2 cycles of 95°C for 1 min, 64°C for 1 min, 72°C for 1 min were run, followed by decreasing the annealing temperature by 1°C per 2 cycles until reaching an annealing temperature of 55°C, used for the next 17 cycles. The final extension was set at 72°C for 5 min. A dilution of 1/100 of the first PCR product was used for the second set of PCR cycles. An initial cycle of denaturation for 5 min at 95°C was followed by 35 cycles of 95°C for 1 min, 50°C for 1 min and 72°C for 1 min and a final elongation step at 72°C for 5 min.

The amplified products were purified and sequenced. The DNA sequences were analyzed using the software package DNASTAR Lasergene 7 (DNASTAR Inc., United States). Samples providing ambiguous sequences were re-amplified and/or re-sequenced. Mixed infections in samples were readily revealed by analyses of the electropherograms of the housekeeping genes and excluded from this study.

For some strains, sequencing of ospA and/or ospC directly from PCR products was difficult. Therefore, these PCR products were cloned into a T-vector (pGEM-T, Promega, United Kingdom). Thereafter, several clones were sequenced using the universal T7 and SP6 primers. For the strain PoTiBGr82, we could not clone the ospC fragment, thus, no sequence data of this locus is available for this isolate.

### Table 5. Pairwise genetic distance among of housekeeping genes of selected *B. lusitaniae* samples and *B. burgdorferi* strains B31, NE49 and Z41293.

<table>
<thead>
<tr>
<th>Strain</th>
<th>B31</th>
<th>NE49</th>
<th>Z41293</th>
<th>PoTiBL37</th>
<th>PoTimfP364</th>
<th>PoTimfP220</th>
</tr>
</thead>
<tbody>
<tr>
<td>B31</td>
<td></td>
<td>0.024*</td>
<td>0.021*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NE49</td>
<td>0.0166(0.0183)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z41293</td>
<td>0.0152(0.0170)</td>
<td>0.0137</td>
<td>0.0132</td>
<td>0.0024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PoTiBGr41</td>
<td>0.0780</td>
<td></td>
<td></td>
<td>0.0137</td>
<td>0.0132</td>
<td>0.0024</td>
</tr>
<tr>
<td>PoTiBL37</td>
<td>0.0798</td>
<td></td>
<td>0.0048</td>
<td></td>
<td>0.0113</td>
<td></td>
</tr>
<tr>
<td>PoTimfP364</td>
<td>0.0800</td>
<td></td>
<td></td>
<td>0.0048</td>
<td></td>
<td>0.0132</td>
</tr>
</tbody>
</table>

*pairwise genetic distance for multiple genes as determined by Postic and colleagues [52]. The values in brackets are based on eight housekeeping genes, including *uvrA*. The pairwise genetic distance among samples from Mafra ranged from 0.0002–0.0036, except for PoTimfP220, whereas these values ranged from 0.0–0.0048 for samples from Grândola.
Analysis of MLST data

G+C content, percentage of variable sites (VI) and average number of nucleotide differences per site (½) were calculated for each locus using DnaSP version 4.0 [39]. Average dN/dS ratios were estimated using the modified Nei-Gojobori/Jukes-Cantor method in MEGA 4 [40]. MEGA 4 was also used to determine pairwise genetic distances. This approach was used to help calibrate the threshold levels of sequence divergence used to delineate species. In addition, the distance matrices based on the concatenated housekeeping gene, ospC and ospA sequences obtained with MEGA were transferred into Minitab Statistical Software® (Minitab Inc., State College PA, U.S.A.) to generate histograms of the frequency of genetic distances in *B. lusitaniae*.

Sequences of the housekeeping genes were assigned allele numbers. For those samples for which eight housekeeping genes could be amplified and sequenced, STs were defined according to the MLST website hosted at Imperial College London, United Kingdom (www.mlst.net). For the five samples for which no **all** sequence information was obtained, sequences of the remaining seven genes were also assigned allele numbers according to the website, but STs were arbitrarily labelled as Lus1-4, because the *Borrelia* MLST website can only define STs if eight housekeeping genes are available.

Phylogenies were inferred for the concatenated sequences of the housekeeping genes and, individually, for the IGS, *ospA* and *ospC*. All alignments were made using MEGA 4. Phylogenetic trees were constructed using MrBayes software version 3.0b4 [41]. Sequences of the North American *B. burgdorferi* strain B31 were used to root the trees. Hierarchical likelihood ratio tests were conducted using MrModeltest (http://www.abo.se/~nylander) to provide the evolutionary models used in the Bayesian analysis. The models selected were GTR for *recG* and IGS, GTR+I for *clpX*, *pepX* and *ospA*, GTR+G for *clpX*, *pyrG* and *ospC*, and HKY+I for *nifS* and *rplB*. For the analysis of the concatenated sequences of the housekeeping genes, the GTR+G+I model was used. Each MrBayes analysis consisted of 2×10⁶ generations or until the standard deviation of split frequencies was <0.01 from a random starting tree and four Markov chains (with default heating values) sampled every 500 generations. To prevent reaching only apparent stationarity, two separate runs were made for each analysis. The first 1,000 sampled trees were discarded, resulting in a set of 3,000 analyzed trees sampled after stationarity.

Detection of recombination in sequence data of *B. lusitaniae*

Sequences of *B. lusitaniae* (housekeeping genes, *ospA* and *ospC*) were tested for putative recombination events using Recombination Detection Program, version 3 (RDP3) [42]. The housekeeping genes were tested individually and as concatenated sequences.

In the RDP suite of programs a number of different methods are implemented and can be used simultaneously. The methods chosen for recombination detection in *B. lusitaniae* sequences included RDP [42], GENECONV [43], Maximum Chi Square (MaxChi) [44,45], Chimaera [45], Sister Scanning (SiScan) [46], and 3SEQ [47] which constitute the most powerful methods currently available. Likelihood Assisted Recombination Detection (LARD) [48] was used to confirm potential recombination events detected by other methods.

To test *B. lusitaniae* sequences, the general settings were as follows: the highest acceptable P-value was set to 0.05 with Bonferroni corrections. In RDP the window size was set to 30 and the setting ‘internal and external references’ was chosen as recommended for small datasets (RDP3 Instruction Manual). In MaxChi the ‘variable site per window’ was set to 70, and ‘strip gaps’ switched on. In Chimaera the ‘variable sites per window’ was set to 60; and in SiScan the window size was set to 150 with a step size of 40. Two different analyses were done with identical settings for these programs. For GENECONV, one analysis was done with GSCALE set to 0, while in the second analysis GSCALE was set to 5 (which apparently is better to detect more ancient recombination events). Recombination events that were detected by more than two methods were confirmed with LARD, and P-values are given in Table 4.

ClonalFrame is a model-based method which was developed specifically for the analysis of multilocus sequence typing data to infer the clonal relationship of bacteria. The method allows to infer the chromosomal position of recombination events, to estimate the degree of relatedness of bacterial strains at different timescales and to reveal information on when strains last shared a common ancestor [49,50]. To run ClonalFrame, an input file was created containing the sequences of STs for each individual housekeeping gene. Because ClonalFrame cannot estimate the value for ω, Watterson’s θ (per sequence) was determined in DnaSP [39] using the concatenated housekeeping gene sequences. The concatenated housekeeping gene sequences were used in the Splitstree software package [51] to perform a network analysis using SplitDecomposition.

Supporting Information

Figure S1 Map of Portugal showing the sampling sites Mafra and Grândola.
Found at: doi:10.1371/journal.pone.0004002.s001 (0.08 MB PPT)

Figure S2 Bayesian phylogenetic inference for *clpA* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s002 (0.06 MB PPT)

Figure S3 Bayesian phylogenetic inference for *clpX* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s003 (0.06 MB PPT)

Figure S4 Bayesian phylogenetic inference for *nifS* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s004 (0.06 MB PPT)

Figure S5 Bayesian phylogenetic inference for *pepX* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s005 (0.06 MB PPT)

Figure S6 Bayesian phylogenetic inference for *pyrG* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s006 (0.06 MB PPT)

Figure S7 Bayesian phylogenetic inference for *recG* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s007 (0.06 MB PPT)

Figure S8 Bayesian phylogenetic inferences for *rplB* of *B. lusitaniae*.
Found at: doi:10.1371/journal.pone.0004002.s008 (0.06 MB PPT)

Figure S9 Bayesian phylogenetic inference for *ospA* of *B. lusitaniae*, including samples from Italy and the Portuguese strains Poti B1-3. The figure shows that the Portuguese human isolate PoHL1 clusters with Italian samples (ITA01, ITA02; ‘European’ lineage), whereas samples from Mafra and Grândola cluster together with strains PotiB1-3 (‘African’ lineage) [32]. Using MLST the Portuguese human isolate PoHL1 clusters with samples from Mafra.
Found at: doi:10.1371/journal.pone.0004002.s009 (0.07 MB PPT)

Figure S10 Distribution of pairwise genetic distances at *ospA* of *B. lusitaniae*. The distribution of all samples included shows a
bimodal distribution (A). Upon removal of strains PoHL1 and PoTiBL37, this distribution was not bimodal anymore, indicating that ospA does not clearly separate the regional B. lusitaniae populations (B).

Found at: doi:10.1371/journal.pone.0004002.s010 (0.10 MB PPT)

**Figure S11** Analysis of MLST sequences with ClonalFrame software. The figure shows the inferred genealogy of STs. The numbers of STs correspond to numbers as shown in Table 2. The nodes are labelled with letters A to L. Found at: doi:10.1371/journal.pone.0004002.s011 (0.08 MB PPT)

**Figure S12** Probability plots for recombination. Each diagram corresponds to likely substitution/recombination events inferred on the branches above nodes C to L in Figure S11 (e.g. events on the branch above node C). For nodes A and B no diagram was obtained. The columns in each diagram correspond to each of the seven housekeeping gene fragments. The scale on the y axis ranging from 0 to 1 refers to the probability of recombination. The height of the red lines represents the inferred probability for recombination. Only for clpX on the branch above node D a recombination event was inferred. The black crosses indicate inferred substitutions, their intensity being proportional to their probability.

Found at: doi:10.1371/journal.pone.0004002.s012 (0.25 MB PPT)

**Acknowledgments**

We thank A. Quaresma for collecting ticks and D. M. Aanensen for curating the *Borrelia* MLST website hosted at Imperial College London, United Kingdom (www.mlst.net).

**Author Contributions**

Conceived and designed the experiments: KK LRV MCP LZZ. Performed the experiments: LRV. Analyzed the data: KK LRV GM MF MCP LZZ. Contributed reagents/materials/analysis tools: KK. Wrote the paper: KK LRV GM MF MCP LZZ.