Genotyping reveals a wide heterogeneity of
Tropheryma whipplei

Wenjun Li,1 Florence Fenollar,1 Jean-Marc Rolain,1 Pierre-Edouard Fournier,1 Gerhard E. Feurle,2 Christian Müller,3 Verena Moos,2 Thomas Marth,4 Martin Altwegg,5 Romana C. Calligaris-Maibach,6 Thomas Schneider,7 Federico Biagi,8 Bernard La Scola1 and Didier Raoult1

Correspondence
Didier Raoult
didier.raoult@gmail.com

1Unité des Rickettsies, IFR 48, CNRS UMR 6020, Faculté de Médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
2DRK Krankenhaus Neuwied, Germany
3Universitätsklinik für Innere Medizin IV, Klinische Abteilung Gastroenterologie und Hepatologie, Medizinische Universität, Wien, Austria
4Department of Internal Medicine, St Josef Hospital, Zel, Germany
5Bio-Analytica AG, Maihofstrasse 95a, 6000 Luzern 6, Switzerland
6Department of Medical Microbiology, University of Zurich, Zurich, Switzerland
7Department of Infectious Diseases, Charité, Campus Benjamin Franklin, Berlin, Germany
8Gastroenterology Unit, Fondazione IRCCS, Policlinico San Matteo, University of Pavia, Italy

Received 30 July 2007
Revised 25 October 2007
Accepted 25 October 2007

Tropheryma whipplei, the causative agent of Whipple’s disease, is associated with various clinical manifestations as well as an asymptomatic carrier status, and it exhibits genetic heterogeneity. However, relationships that may exist between environmental and clinical strains are unknown. Herein, we developed an efficient genotyping system based on four highly variable genomic sequences (HVGSs) selected on the basis of genome comparison. We analysed 39 samples from 39 patients with Whipple’s disease and 10 samples from 10 asymptomatic carriers. Twenty-six classic gastrointestinal Whipple’s disease associated with additional manifestations, six relapses of classic Whipple’s disease (three gastrointestinal and three neurological relapses), and seven isolated infections due to T. whipplei without digestive involvement (five endocarditis, one spondylodiscitis and one neurological infection) were included in the study. We identified 24 HVGS genotypes among 39 T. whipplei DNA samples from the patients and 10 T. whipplei DNA samples from the asymptomatic carriers. No significant correlation between HVGS genotypes and clinical manifestations of Whipple’s disease, or asymptomatic carriers, was found for the 49 samples tested. Our observations revealed a high genetic diversity of T. whipplei strains that is apparently independent of geographical distribution and unrelated to bacterial pathogenicity. Genotyping in Whipple’s disease may, however, be useful in epidemiological studies.

INTRODUCTION

Infection with Tropheryma whipplei causes a variety of clinical manifestations, such as classic Whipple’s disease marked by histological lesions in the gastrointestinal tract associated with diarrhoea, lymphadenopathy, arthritis, blood culture-negative endocarditis and isolated neurological infection without these typical histological lesions (Dutly & Altwegg, 2001; Fenollar et al., 2007a; Fenollar & Raoult, 2001; Marth & Raoult, 2003). In addition, T. whipplei DNA has been detected in the environment and in
asymptomatic carriers (Dutly et al., 2000; Ehrbar et al., 1999; Maiwald et al., 1998; Street et al., 1999). However, the prevalence of T. whipplei in the environment is still controversial because cultivation of T. whipplei directly from environmental samples is extremely difficult. The recent culture of T. whipplei from stools suggests a faeco-oral transmission of the disease (Raoult et al., 2006).

Several questions remain about the pathogenicity of T. whipplei. In theory, some strains may exhibit specific pathogenic virulence leading to classic Whipple’s disease, others may cause endocarditis or central nervous system disease, and other strains may be apathogenic. Finally, it is not clear whether a differential geographical distribution exists among T. whipplei strains.

Completion of the genome sequences for two T. whipplei strains, TW08/27 and Twist, may allow a rational selection of appropriate genomic sequences for typing this bacterium (Bentley et al., 2003; Raoult et al., 2003). In this study, we selected four highly variable genomic sequences (HVGSs) in the aligned genome sequences of strains TW08/27 and Twist. Indeed, we have previously demonstrated that the greatest discriminatory power resides in the most variable regions. Twist and TW08/27. To this end, BLAST2 (http://www.ncbi.nlm.nih.gov/blast/blast2.cgi) was used to align the genomic sequences of T. whipplei strains Twist (Raoult et al., 2003) (GenBank accession no. AE014184) and TW08/27 (Bentley et al., 2003) (GenBank accession no. BX072543). Then, genomic sequences with low homology, with sizes ranging from 300 to 500 bp and flanked by conserved sequences, were selected and used as genotyping markers in this study.

**METHODS**

**Patients.** Forty-nine samples were analysed in this study, including 39 specimens taken from 39 patients with a diagnosis of Whipple’s disease on the basis of clinical manifestations, histological analysis and PCR assays, as previously reported (Fenollar et al., 2002, 2004, 2007a; Lepidi et al., 2003). The presentations linked to T. whipplei infections were defined as gastrointestinal or classic Whipple’s disease characterized by positive periodic acid-Schiff (PAS)-staining lesions on small-bowel biopsies associated with additional manifestations, and as isolated infections characterized by the lack of the histological lesions on small-bowel biopsies, such as blood culture-negative endocarditis, isolated spondylodiscitis or isolated neurological infection. Relapses of classic Whipple’s disease were divided into gastrointestinal and neurological relapses, depending on the clinical manifestations and the involved organs. Ten samples from 10 asymptomatic carriers were also included, corresponding to healthy people or patients without clinical manifestations of Whipple’s disease and/or an excluded diagnosis of Whipple’s disease based on the analysis of small-bowel biopsies, as previously reported (Ehrbar et al., 1999; Fenollar et al., 2007b; Schoniger-Hekele et al., 2007).

**Study design.** Detection of T. whipplei in specimens was carried out using PCR targeting the 16S–23S rRNA internal transcribed spacer (ITS), the rpoB gene and/or a repeated sequence, as described in previous studies (Drancourt et al., 2001; Fenollar et al., 2002, 2004). Each T. whipplei strain detected in a human specimen was given a unique code containing ordinal information, i.e. country where the sample was collected, clinical manifestation, and organ or tissue from which the specimen was taken (Fig. 1 and Supplementary Table S1). For example, FrDDb1 stands for one T. whipplei strain detected in the duodenal biopsy (Db) of a French patient (Fr) with classic Whipple’s disease (D = digestive).

**DNA preparation.** Total genomic DNA was extracted from each specimen using the QIAamp Tissue kit (Qiagen), as recommended by the manufacturer.

**Selection of HVGSs.** Four fragments were selected by identifying the most variable sequence fragments between T. whipplei strains Twist and TW08/27. To this end, the BLAST2 program (http://www.ncbi.nlm.nih.gov/blast/blast2.cgi) was used to align the genomic sequences of T. whipplei strains Twist (Raoult et al., 2003) (GenBank accession no. AE014184) and TW08/27 (Bentley et al., 2003) (GenBank accession no. BX072543). Then, genomic sequences with low homology, with sizes ranging from 300 to 500 bp and flanked by conserved sequences, were selected and used as genotyping markers in this study.

**Design of primers, PCR amplification and sequencing.** Amplifications of the ITS, the rpoB gene and/or repeated sequences were performed using the previously described conditions and the tsw3/tsw4r (Fenollar et al., 2002), TWRPOB.F/TWRPOB.R (Drancourt et al., 2001) and 5303F/5303R primer pairs (Fenollar et al., 2004), respectively. Primers for amplifying and sequencing the four HVGSs were selected using the Primer 3.0 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). All primers were obtained from Eurogentec and are shown in Table 1. PCR reactions were carried out in a PTC-200 automated thermal cycler (MJ Research). Five microlitres of 3 ng μl⁻¹ of each DNA solution were amplified in a 25 μl reaction mixture containing 50 pM each primer; 200 μM of each dATP, dCTP, dGTP and dTTP (Invitrogen); 1.5 U Hotstar Taq DNA polymerase (Qiagen); 2.5 μl 10 × PCR buffer; and 1 μl 25 mM MgCl₂. The following conditions were used for amplification: initial heat activation at 95 °C for 15 min was followed by 40 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 55 °C, and extension for 1 min at 72 °C. Amplification was completed by holding the reaction mixture for 5 min at 72 °C to allow complete extension of the PCR products. PCR products were purified by using the MultiScreen PCR filter plate (Millipore), as recommended by the manufacturer. Amplicons were sequenced in both directions using BigDye 1.1 chemistry (Applied Biosystems) on an ABI 3130xl automated sequencer (Applied Biosystems), as recommended by the manufacturer. Sterile water was used as a negative control in each PCR assay. All sequences were checked twice in both directions to ensure the reliability of the typing method.

**Sequence analysis and phylogenetic analysis.** Nucleotide sequences were edited using the Autoassembler package (Perkin-Elmer). For each HVGS, a sequence type was defined as a sequence exhibiting unique mutation(s). HVGS genotypes were defined as unique combinations of the four HVGS types. The genotypic distribution of T. whipplei strains associated with four clinical manifestations was analysed to evaluate the correlation between HVGS genotypes and distinct clinical manifestations of Whipple’s disease.

Multiple alignment of sequences was carried out using the CLUSTAL W software (Thompson et al., 1994). For phylogenetic analysis, sequences of the four HVGSs were concatenated. Phylogenetic relationships among T. whipplei genotypes were inferred using the unweighted pair group method with arithmetic mean (UPGMA), neighbour-joining and maximum-parsimony methods within the MEGA 3.1 software (Kumar et al., 2001). All different HVGS sequence types were deposited in GenBank (Table 1).
RESULTS

Patients

The presentations linked to Whipple’s disease were: 26 classic gastrointestinal Whipple’s disease associated with additional manifestations; six relapses of classic Whipple’s disease, including three gastrointestinal and three neurological relapses; and seven isolated infections due to *T. whipplei*, including five endocarditis, one spondylodiscitis and one neurological infection (Fig. 1, Supplementary Table S1). From the 10 patients without Whipple’s disease, there were five stool samples taken from five Austrians, and four gastric juice samples and one duodenal biopsy taken from five Swiss (Fig. 1, Supplementary Table S1).

Genotyping *T. whipplei* based on four HVGSs

The four selected regions exhibited 92.51, 97.48, 93.42 and 89.22 % identity between *T. whipplei* strains Twist and TW08/27, contained 14, 8, 10 and 11 variable positions, and classified the 49 strains in 8, 8, 6 and 4 sequence types, respectively (Tables 1 and 2). By combining sequence types of the four sequences, the 49 tested *T. whipplei* strains were classified into 24 HVGS genotypes (Table 1). Of these, genotype 1, which was identified in 13 *T. whipplei* strains (26.5 % of the 49 tested *T. whipplei* DNA samples) was the most common in this study (Fig. 1). Genotypes 3, 8, 2, 16 and 19 included six, four, three, three and two *T. whipplei* DNA samples, respectively, which together accounted for 36.7 % of the 49 tested *T. whipplei* strains (18/49) (Fig. 1). Each of the remaining 18 genotypes contained only one strain each (Fig. 1).

When stratifying genotypes according to clinical manifestations, 18 genotypes (1–6, 8–16, 19, 21 and 23) were identified among the 32 *T. whipplei* DNA samples detected in patients with classic Whipple’s disease; five genotypes (3, 7, 8, 20 and 24) were identified among the five strains detected in patients with endocarditis; the single strain detected in a patient with spondylodiscitis was of genotype 2, which was also shared by another two strains from patients with classic Whipple’s disease; the 10 strains detected in patients without Whipple’s disease were classified into seven genotypes (1, 8, 16–19 and 22) (Fig. 1).

Phylogenetic analysis of 24 HVGS genotypes

Phylogenetic trees obtained from concatenation of the four HVGS sequences using the neighbour-joining and
maximum-parsimony methods showed similar phylogenetic classifications to the UPGMA method. The 24 HVGS genotypes were grouped into three clusters and one single branch linked with cluster 3 (Fig. 1, Supplementary Table S1). Cluster 1 contained 12 HVGS genotypes from 21 T. whipplei strains, cluster 2 contained seven HVGS genotypes

### Table 1. Genomic information for the four HVGSs, primers used for amplification and sequencing of these HVGSs, and the reference numbers of the nucleotide sequences deposited in GenBank

Nucleotide sequences from the HVGS 1 marker were deposited in GenBank under reference numbers EF363910, EF363911, EF363912, EF363913, EF363914, EF363915, EF363916 and EF363917 for genotypes 1, 2, 3, 4, 5, 6, 7 and 8, respectively. Nucleotide sequences from the HVGS 2 marker were deposited in GenBank under reference numbers EF363917, EF363918, EF363919, EF363920, EF363921, EF363922, EF363923, EF363924 and EF363925 for genotypes 1, 2, 3, 4, 5, 6, 7 and 8, respectively. Nucleotide sequences from the HVGS 3 marker were deposited in GenBank under reference numbers EF363926, EF363927, EF363928, EF363929, EF363930 and EF363931 for genotypes 1, 2, 3, 4, 5 and 6, respectively. Nucleotide sequences from the HVGS 4 marker were deposited in GenBank under reference numbers EF363932, EF363933, EF363934 and EF363935 for genotypes 1, 2, 3 and 4, respectively.

<table>
<thead>
<tr>
<th>HVGS</th>
<th>Content of HVGS</th>
<th>Genotype Size (bp)</th>
<th>Position on Twist genome</th>
<th>Position on TW08/27 genome</th>
<th>Identity between TW08/27 and Twist</th>
<th>Forward primer sequence (5’–3’)</th>
<th>Reverse primer sequence (5’–3’)</th>
</tr>
</thead>
</table>
| HVGS 1 | TWT133 and intergenic spacer (TWT133–TWT134) | 8 | 227 | 154 935–155 161 | 154 333–154 553 | 92.51 % | GCTGCGCGAAGTAATTG | AGATACATGCGGAGATAC
| HVGS 2 | proS/prolyl-tRNA synthetase (TW183) | 8 | 318 | 233 956–234 273 | 661 022–661 339 | 97.48 % | GCCTTGACTATGACATAAATAAC | CGTGACTAATCATGCGGACAC
| HVGS 3 | secA-hp (TWT131) intergenic spacer | 6 | 150 | 152 732–152 881 | 152 147–152 297 | 93.42 % | TTTGCATAGCGTTCTGTAG | AGACCTCACTGTATACGG
| HVGS 4 | TW183 and intergenic spacer (TWT183–TWT184) | 4 | 162 | 234 449–234 610 | 660 680–660 846 | 89.22 % | CGGATTCGACGAAATGTC | ATACACGAAGCTGGATATGC

### Table 2. Polymorphism characteristics of the four HVGSs

The numbers in the HVGS polymorphism column show the position of each variable nucleotide with reference to the Twist strain. The allele before the number is that within the Twist strain and the alleles after the number are possible variable nucleotides within other T. whipplei strains. Inser., insertion; dele, deletion.

<table>
<thead>
<tr>
<th>HVGS</th>
<th>Number of nucleotide variations</th>
<th>Number of genotypes</th>
<th>HVGS polymorphism, with reference to the Twist strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVGS 1</td>
<td>14</td>
<td>8</td>
<td>GGGGAT11-16 dele C124T G24T G156A C122T</td>
</tr>
<tr>
<td>HVGS 2</td>
<td>8</td>
<td>8</td>
<td>GCGGAT8 C29T G156A C140T C122C</td>
</tr>
<tr>
<td>HVGS 3</td>
<td>10</td>
<td>6</td>
<td>G19 T20 T103A G103A T20C C96T CT45-55TC A1666 A1666</td>
</tr>
<tr>
<td>HVGS 4</td>
<td>11</td>
<td>4</td>
<td>G146A G146A G146A G146A</td>
</tr>
<tr>
<td>Total (four HVGSs)</td>
<td>43</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

W. Li and others
from 23 T. whipplei strains, and cluster 3 contained four HVGS genotypes from five T. whipplei strains (Fig. 1). Genotype 24, identified in strain Twist and isolated from the cardiac valve of a Canadian patient with endocarditis, grouped with cluster 3 but with a low bootstrap value (Fig. 1). The phylogenetic classification of T. whipplei strains associated with distinct clinical manifestations was also analysed to evaluate the correlation between HVGS genotypes and distinct clinical forms (Fig. 1, Supplementary Table S1). Cluster 1 contained 14, three, one, one and two T. whipplei strains detected in patients with classic Whipple’s disease, endocarditis, spondylodiscitis, neurological relapse, and patients without Whipple’s disease, respectively (Fig. 1). Nine, one, one, two, two and seven strains detected in patients with classic Whipple’s disease, neurological disorder, endocarditis, gastrointestinal relapse, neurological relapse, and individuals without Whipple’s disease, respectively, were grouped into cluster 2 (Fig. 1). Cluster 3 contained three, one and one strains from patients with classic Whipple’s disease, a patient with gastrointestinal relapse and a patient without Whipple’s disease, respectively (Fig. 1). Thus, in this study, no significant correlation between HVGS genotypes and distinct manifestations of Whipple’s disease was found on the basis of genotypic and phylogenetic analysis.

**DISCUSSION**

In this study, we found the combination of four HVGSs, selected following comparison of the complete genomes of two T. whipplei strains, useful for genotyping 49 clinical strains of this bacterium. Prior to our study, because of the fastidious nature of T. whipplei, several targets, such as 16S rDNA, 23S rDNA, hsp65, atpD, tuf, groEL, rpoB and rnpD and 16S–23S rDNA ITS, had been studied to assess the phylogenetic organization of T. whipplei (Drancourt et al., 2001; Hinrikson et al., 1999b, 2000b; Maiwald et al., 1996, 2000, 2003; Morgenegg et al., 2000; Relman et al., 1992; Wilson et al., 1991). Most of these DNA sequences, however, are more conserved and reveal a limited genetic diversity among T. whipplei strains, although they contribute to some extent to genus and species classification. From sequences of domain III of 23S rDNA, two sequence types were identified among 50 T. whipplei strains detected in human specimens by Hinrikson et al. (2000a, b). Using a 620 bp fragment of the hsp65 gene, four genetic variants were identified among eight T. whipplei strains, but these strains had been selected from representative strains of three ITS genotypes (Morgenegg et al., 2000). ITS, currently the first-used molecular marker for typing T. whipplei infection, has so far allowed the differentiation of seven genotypes (Geissdorfer et al., 2001a; Hinrikson et al., 1999a, b; Maiwald et al., 2000). In a first study, ITS was found to be identical in nine Swiss strains (Hinrikson et al., 1999a). In a second study, three ITS sequence types were identified among 38 strains from 28 Swiss patients with Whipple’s disease (Hinrikson et al., 1999b). Subsequently, five ITS sequence types were identified in 43 patients with Whipple’s disease from the USA (11 patients), Germany (28 patients), Switzerland (three patients) and Austria (one patient) (Maiwald et al., 2000). More recently, a new ITS sequence type, type 7, was described in a German patient with endocarditis (Geissdorfer et al., 2001a). Combination of the four HVGSs allowed the identification of 24 HVGS genotypes among the 49 studied T. whipplei strains. The degree of sequence identity of HVGSs between strains Twist and TW08/27 (89.2–97.5 %) was lower than that of ITS (99.32 %) (Geissdorfer et al., 2001b; Hinrikson et al., 1999b). Thus, compared with previously described DNA sequence-based typing methods, our HVGS-based typing method was more discriminatory.

Recently, another typing strategy that combined three variable number of tandem repeats (VNTRs) and ITS discriminated 11 T. whipplei DNAs detected in patients with Whipple’s disease. Each of the four markers (three VNTRs and ITS) identified two, three, six and two types among the 11 T. whipplei DNAs, respectively (Maiwald et al., 2000). However, DNA size-based typing methods are not as discriminatory as sequence-based typing methods, because size variations may result from nucleotide insertions or deletions other than at VNTR sites. Phylogenetic analysis combining the two methods is difficult. In addition, the Maiwald et al. (2000) study was based on a limited number of strains.

Although the study of additional strains may be necessary, our results suggest that asymptomatic carriage and distinct clinical manifestations of Whipple’s disease are not related to the genetic diversity of T. whipplei but rather to as yet unknown patient characteristics such as a deficient immune function, as proposed previously (Marth et al., 1994, 1997).

Although Whipple’s disease is a rare systemic infectious disease, and mainly endemic in Europe and America, the relationships between genetic diversity and geographical distribution of the causative agent remain unclear. From ITS sequences, it appears that two common sequence types (types 1 and 2) are observed, with similar frequencies in patients from both the USA and Europe (Hinrikson et al., 1999b; Maiwald et al., 2000). Only one T. whipplei strain from outside Europe, Twist, was included in our study. Strain TW08/27, isolated from cerebrospinal fluid of a German patient with a neurological relapse of classic Whipple’s disease, had the most frequent HVGS genotype, type 1, and grouped into cluster 2 with 22 European T. whipplei strains (Fig. 1). Strain Twist, isolated from a cardiac valve of a Canadian patient with endocarditis due to T. whipplei, exhibited a specific HVGS genotype, type 24, and grouped with cluster 3 with an extremely low bootstrap value (Fig. 1). We could not draw any conclusion about the geographical distribution of T. whipplei. Ten HVGS genotypes were identified among 11 Swiss T. whipplei strains, in contrast to 15, two and three HVGS genotypes identified among 25 French, five Austrian and
seven German T. whipplei strains, respectively. The genetic diversity of Swiss T. whipplei strains seems to be higher than that observed in French, Austrian and German strains. However, more T. whipplei strains originating from different continents should be tested using this typing method to study the geographical distribution of T. whipplei. The high variability provides further evidence that positive PCR results in asymptomatic carriers are not, as was previously suggested (Dutly & Altwegg, 2001; Fenollar et al., 2007a, b), due to carry-over contamination.

In conclusion, we identified four HVGSs based upon genomic comparison of T. whipplei strains TW08/27 and Twist. The four HVGSs classified 49 T. whipplei strains detected in patients with or without manifestations of Whipple’s disease into 24 genotypes. However, no significant correlation between HVGS types and clinical manifestations of Whipple’s disease or geographical distribution of T. whipplei was found. Our observations revealed a high genetic diversity of T. whipplei strains that apparently did not determine microbial pathogenicity and geographical distribution. Genotyping of T. whipplei may, however, be useful for epidemiological investigations such as the discrimination of recurrent Whipple’s disease from reinfection with another strain.

ACKNOWLEDGEMENTS

We thank Melanie Ihrlg for reviewing the manuscript. This project was funded by the 5th Framework Programme of the European Commission (ref. QLG1-CT-2002-01049).

REFERENCES


Edited by: K. E. Weaver