Organization, Structure, and Variability of the rRNA Operon of the Whipple’s Disease Bacterium (Tropheryma whippelii)

MATTHIAS MAIWALD,1 AXEL VON HERBAY,2 PAUL W. LEPP,1 AND DAVID A. RELMAN1,3,4*

Departments of Microbiology and Immunology1 and Medicine,3 Stanford University School of Medicine, Stanford, California 94305; Veterans Affairs Palo Alto Health Care System, Palo Alto, California 94304; and Pathologisches Institut der Universität, 69120 Heidelberg, Germany2

Received 8 December 1999/Accepted 12 March 2000

Whipple’s disease is a systemic disorder associated with a cultivation-resistant, poorly characterized actinomycete, Tropheryma whippelii. We determined a nearly complete rRNA operon sequence of T. whippelii from specimens of 3 patients with Whipple’s disease, as well as partial operon sequences from 43 patients. Variability was observed in the 16S-23S rRNA spacer sequences, leading to the description of five distinct sequence types. One specimen contained two spacer sequence types, raising the possibility of a double infection. Secondary structure models for the primary rRNA transcript and mature rRNAs revealed rare or unique features.

Whipple’s disease was described in 1907 (as intestinal lipodystrophy) and is a multisystem disorder of humans involving the intestinal tract as well as various other organs (3). A constant feature of the disease is the presence in affected tissues of uniform bacteria that are approximately 0.2 by 1.5 to 2.5 μm. These bacteria have a characteristic morphology when viewed with electron microscopy (25). However, numerous attempts to cultivate this bacterium have failed (3). In 1997, propagation in human macrophage cell cultures was reported (24); and recently, propagation in human fibroblasts (20), but these findings have not yet been confirmed by other groups. Thus, this organism remains poorly characterized.

Molecular data from the Whipple’s disease bacterium became available through broad-range 16S ribosomal DNA (rDNA) PCR and sequence analysis (21, 31). A phylogenetic assessment based on 1,321 bp of 16S rDNA established the bacterium (Tropheryma whippelii) as an actinomycete (21). Taxon-specific 16S rDNA primers have since been used to detect the bacterium in additional patients with Whipple’s disease (19, 21, 28, 29), as well as in sewage effluent (16). The determination of a nearly complete 16S rDNA sequence and the 16S-23S ribosomal RNA intergenic spacer sequence has provided additional information (16). A reassessment of its phylogeny revealed a position between the actinomycetes with group B peptidoglycan and the family Cellulomonadaceae.

The 16S-23S rDNA spacer has been used for strain differentiation in a variety of bacterial species (4). Two recent studies, both from the same group in Switzerland, addressed the potential variability of the 16S-23S ribosomal intergenic spacer of T. whippelii (5, 6). In the first study, the spacer region was found to be homogenous in specimens from 9 Swiss individuals; in the second study, three types of intergenic spacers were detected in specimens from 28 individuals. In contrast to the 16S-23S rDNA spacer, little information has been published on the intraspecies variability of the 23S-5S rDNA spacer and of regions immediately upstream and downstream of the rRNA operon.

Because of the paucity of information concerning T. whippelii and the usefulness of the rRNA operon as a source of data for microbial characterization, we determined and analyzed a nearly complete rRNA operon sequence directly from tissues from several patients with Whipple’s disease. In addition, we analyzed partial operon sequences from 43 patients.

Patients and clinical specimens. A total of 56 clinical specimens from 43 patients with Whipple’s disease who originated from the United States (11 patients), Germany (28 patients), Switzerland (3 patients), and Austria (1 patient) were examined. The specimens comprised 37 intestinal biopsy specimens, 10 cerebrospinal fluid (CSF) samples, 2 brain tissue specimens, 2 lymph node tissue specimens, 1 tissue specimen associated with prostatic hip material, 1 colon tissue specimen, 1 ocular lens capsule specimen, 1 blood sample, and 1 synovial fluid specimen. All patients, except three patients from the United States, had Whipple’s disease confirmed by histology. These three patients had positive PCR results by CSF analysis, and information on histology was not available. The specimens were prepared for PCR as previously described (20, 29, 30).

Previously uncharacterized regions of the rRNA operon were amplified by PCR with primers designed from known T. whippelii sequences in combination with primers designed from highly conserved regions of other bacteria (Table 1). To obtain sequence material downstream of the 5S rRNA gene, the restriction site PCR method of Sarkar et al. (23) was used (Table 1). PCRs were performed according to standard protocols using AmpliTaq Gold polymerase (PE Applied Biosystems, Foster City, Calif.). Sequences were determined either directly from PCR products or after cloning, as described previously (15).

Primary rRNA operon sequence and phylogeny. A nearly complete rRNA operon sequence, spanning 5,747 bp, was assembled from PCR products from the intestinal biopsy specimens of a patient with Whipple’s disease (patient 1 [17]). Sequences from at least three independent clones of each PCR product (Table 1) were determined. The sequence begins in the box C element, which is 50 to 380 bp downstream of the start of transcription in other actinomycetes (7, 9, 10, 18), and extends from 162 bp upstream of the 16S rRNA gene to 493 bp downstream of the 5S rRNA gene. The arrangement of rRNA
TABLE 1. PCR primers and amplified regions within the rRNA operon sequence of T. whippelii

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>Region amplified</th>
<th>Source and/or reference(s) for primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>16Spro1 (5'-TTGAGAATCTAABAGYTG-3') and tw318r (5'-CGAAATTATCCAAAGTTAG-3')</td>
<td>1–298</td>
<td>This work</td>
</tr>
<tr>
<td>fd1mod (5'-AGAGTTTGTATCGYTTYAG-3') and whip2 (5'-ATTGCTCCACCTTGGGA-3')</td>
<td>193–1400</td>
<td>21, 28, 30</td>
</tr>
<tr>
<td>whip1 (5'-AGAGATACGCCCGGGCAA-3') and 23S rDNA primer 2 (5'-GGTACCTTAGTGTTCAGTC-3')</td>
<td>1167–2182</td>
<td>12, 21, 28</td>
</tr>
<tr>
<td>tw1581f (5'-GTGCACTTAACTTGTGGGAGA-3') and tw2015r (5'-GCATCCACATTGTGCTTAAA-3')</td>
<td>1602–1993</td>
<td>This work</td>
</tr>
<tr>
<td>tw1974f (5'-GTATTTTGTGATTCAGCCTAC-3') and ms37ar (5'-CTGCTTCTAAAGCCACATCCT-3')</td>
<td>1994–3136</td>
<td>This work; 13</td>
</tr>
<tr>
<td>tw3002f (5'-TGCCGAGTATCGAGGCA-3') and ms38a (5'-GACAAGGAATTTCGCTACCTTA-3')</td>
<td>3022–4102</td>
<td>This work; 13</td>
</tr>
<tr>
<td>tw3887f (5'-GGCGGACCTTGGGCCCTTG-3') and co5189r (5'-GCTTCCGGGTTGCAAGATG-3')</td>
<td>3904–5171</td>
<td>This work</td>
</tr>
<tr>
<td>tw5104f (5'-CTTGATGTGCGGCCCTTTGC-3') and tw5745r (5'-AAGATCCACCTCGACATCG-3')</td>
<td>5124–5722</td>
<td>This work</td>
</tr>
<tr>
<td>tw5068f (5'-ATGTCAGACCGAAACATTGTGG-3') or tw5104f and Nhe1-RSO (5'-AATACGACTCCTATAGG-3')</td>
<td>5124–5747</td>
<td>This work; 23</td>
</tr>
</tbody>
</table>

*In sequence AF190866. Primers are excluded.

genes is in the order 5'-16S-23S-5S-3'. tRNA genes were not identified.

Phylogenetic analyses were performed as described previously (14) for 16S, 23S, and 5S rRNA, utilizing 1,365, 3,033, and 120 masked nucleotide positions, respectively. Results of 16S rRNA phylogeny were consistent with those of a previous analysis (15). Both 23S rRNA- and 5S rRNA-based phylogenies consistently placed T. whippelii within the Actinobacteria, while a more detailed resolution of relationships between this organism and the actinomycetes with group B peptidoglycan was not possible, due to a lack of a sufficient number of related sequences in both 23S rRNA (2) and 5S rDNA (27) databases.

Variance of rRNA operon sequences. Two additional nearly complete operon sequences of 5,722 bp each, ending with primer tw5745r (Table 1), were determined from a postmortem brain tissue specimen (patient 2) and from a CSF sample (patient 3). The sequence in the sample from patient 2 differed from that of patient 1 at 1 position within the 16S rRNA leader region, 1 position within the 16S rRNA gene, and 10 positions clustered within a region downstream of the 5S rRNA gene (between nucleotides 5322 and 5513). The sequence in the sample from patient 3 had one difference from that of patient 1 within the 16S rRNA gene, two in the 16S-23S ribosomal RNA intergenic spacer, and two in the 23S rRNA gene.

Several operon regions were sequenced from additional patient specimens. These included the 16S rRNA gene leader region (2 specimens from 2 patients; the region between primers 16Spro1 and tw318r) (Table 1), the 16S-23S ribosomal intergenic spacer (53 specimens from 40 patients; primers tw1581f and tw2015r), the region of the insertion in the 23S rRNA gene (1 specimen from 1 patient; primers tw3002f and ms38a), and the 23S-5S ribosomal intergenic spacer (2 specimens from 2 patients; primers tw3887f and co5189r). No differences between these additional sequences and that determined from patient 1 were found in the 16S rRNA gene leader region, the 23S rRNA gene region containing the insertion (positions 3022 to 4102), and the 23S-5S intergenic spacer region.

In previous work, three 16S-23S rDNA spacer sequence types were identified (6, 15). In our study, five types of 16S-23S ribosomal RNA intergenic spacer sequences were found. Type 1 was defined as the original sequence determined for this spacer region (15); it corresponded to the sequence of patients 1 and 2 as described above and was found in a total of 16 specimens from 14 patients. Type 2 (6) was most frequently detected; it corresponded to that of patient 3 and was found in 36 specimens from 26 patients. Type 3, as defined by Hinrikson et al. (6), was not found in any specimen. Types 4 through 6 have not been previously described and were each found in single patients (designated patients 4 to 6). There were a total of 10 variable nucleotide positions and two to six differences with the type 1 sequence. The differences, with reference to type 1 and numbered according to this type's 294-bp spacer sequence, were as follows: type 2, T56C and Δ86; type 4, A94G, C98T, G115A, and G200A; type 5, A94G, G115A, and G200A; type 6, T56C, Δ86, C124T, C143T, T148C, and A157G. To confirm the existence of distinct spacer types by an independent method, PCR products (primers tw1581f and tw2015r) from each of these types were subjected to digestion with the restriction endonucleases HaeIII, HindIII, MseI, MnlI, XmnI, and TaqI. Distinct and expected band patterns were obtained for each spacer type.

There were 11 patients from whom more than one specimen was examined. This comprised five patients from Germany and one from Switzerland with more than one intestinal biopsy specimen, four patients from Germany with intestinal biopsy and CSF specimens, and one patient from the United States (patient 6) with an intestinal biopsy and a blood specimen.
Multiple clones from individual PCR products (range, 2 to 10 clones) were sequenced for 19 patients. In all instances (except for patient 5, whose case is discussed below), 16S-23S spacer types were consistent among different clinical samples from the same patient and between different clones of the same PCR product. This is consistent with the hypothesis that a single bacterial clone undergoes systemic dissemination (29) and is further supported by the results from patient 3, who had an intestinal biopsy taken in 1987 and two CSF specimens taken in 1998 and 1999, all with identical spacer types, and by the results of patient 6, whose samples from different sites yielded the same, unique type 6 spacer sequence. Nucleotide differences observed elsewhere in the operon sequence did not correlate with 16S-23S spacer type.

Geographic analysis of 16S-23S spacer types showed that 10 of 28 German patients had type 1, 17 had type 2, and 1 had a mixed type 2 and 5 infection (discussed below). Of 11 patients from the United States, 3 had type 1, 7 had type 2, and 1 had type 6. Of three patients from Switzerland, two had type 2, and one had type 1. The single patient from Austria had a type 4 sequence. While the number of specimens from each area is relatively small, it appears clear that the two common types (types 1 and 2) are observed at similar frequencies in patients from both the United States and Europe, suggesting that a differential distribution of T. whippelii between the Old World and the New World, such as has been described for Borrelia burgdorferi sensu lato and its genospecies (1), is unlikely.

The five 16S-23S rRNA spacer types found in our 43 patients expand our understanding of T. whippelii sequence diversity (6). Whipple’s disease was confirmed by histological methods in all but 3 of our 43 patients. We found a predominance of spacer type 2 (65%), whereas the Swiss study (6) described a predominance of type 1. The previously reported uniformity of spacer type 1 in Swiss individuals (5) could not be confirmed.

**Sequence variability within a single host.** In the case of patient 5, four clones of the PCR-amplified 16S-23S spacer were sequenced: three yielded an identical but novel type 5 sequence, and one contained a type 2 sequence. Direct sequencing and restriction endonuclease analysis of the PCR product revealed expected sequence ambiguities correspond-
FIG. 2. Variable and distinguishing structures of *T. whippelii* rRNAs. (A) 16S rRNA; (B) 5' half of 23S rRNA; (C) 3' half of 23S rRNA. Helix 33 and 23S rRNA helices A1 and I3 are considered unique to *T. whippelii* (see text); the others have helix length and bulge variabilities relative to the structures of other actinomycetes (http://www.rno.icmb.utexas.edu/). Arrows, with positions in mature rRNAs, indicate unusual nucleotides. There are two possible structures (a and b) for helix 6 in the 16S rRNA. The 23S rRNA insertion of the high-G+C-content gram-positive bacteria is located at positions 3522 to 3601 of the sequence AF190686.
ing to differences between sequence types 2 and 5. These findings suggest that both sequence types were present in the patient’s specimen and raise two possibilities: either the bacterium has more than one rRNA operon, or the patient was infected by two different bacterial strains. While it is not possible to draw a firm conclusion between these two alternatives, there are two findings consistent with the notion of a single rRNA operon in *T. whippelii*. First, there was concordance of rRNA spacer type between different clinical samples and clones of PCR products from specimens from each of the other patients (discussed above). Second, amplification of genomic DNA flanking the 3′ end of the operon by restriction site PCR (24) from the specimen of patient 1 generated two fragments in the same reaction, yielding 491- and 677-bp sequences, respectively; the sequence of the 491-bp fragment extended 206 bp downstream of the 5S rRNA gene and was identical to the sequence of the 677-bp PCR product in the region of overlap. Thus, this initial analysis failed to find variability in the genome adjacent to, but outside of, the rRNA operon. For comparison, in the *Mycobacterium tuberculosis* genome (accession no. Z73902), a non-rRNA gene begins 120 bp downstream of the 5S rRNA. Thus, the second explanation, i.e., a dual infection by different *T. whippelii* strains, appears more likely.

**Structural and regulatory features of the primary RNA transcript.** A secondary structure model of the primary rRNA transcript was constructed by visual comparison to published structures (7–11, 18), with some assistance from the Mfold program (Genetics Computer Group, Madison, Wis.). The proposed model is shown in Fig. 1. This model is consistent with published models for other members of the class Actinobacteria (7–11, 19) and shows two stems, one bracketing pre-16S rRNA and consisting of the 16S rRNA leader region and the 5′ portion of the 16S-23S rRNA spacer (spacer 1a), and another bracketing pre-23S rRNA and consisting of the 3′ portion of the 16S-23S rRNA spacer (spacer 1b) and the 23S-5S rRNA spacer (spacel 2). The 5′ end of the model begins with primer 16Spro1, which is located within and between the box A and box C elements of the operon leader region. The polymorphic positions within the 16S-23S spacer were mapped on the secondary structure model (Fig. 1). According to the model, all nucleotide differences between 16S-23S spacer types either are located in loops, where they are not involved in base pairing, or create tolerated changes in paired positions that preserve the structure.

**Structural features associated with mature tRNAs.** Putative secondary structures of 16S, 23S, and 5S rRNAs were inferred by manual alignment to published structures (18; http://www.rma.icmb.utexas.edu/). The proposed model for 5S rRNA (data not shown) was identical to a previously-published model for *Streptomyces ambofaciens* (18), except for a paired set of nucleotides in *T. whippelii* 5S rRNA (G20 and C63). Several regions in the *T. whippelii* predicted 16S and 23S rRNA secondary structures were found to differ from those of other actinomycetes (http://www.rna.icmb.utexas.edu/) (Fig. 2). Most of these variable regions involved differences in helix length and unpaired or unilateral bulges along helix stems. Three features of the 16S rRNA and 23S rRNA, however, were found to be unique to *T. whippelii*, after comparison with a larger set of structures and sequences (10,073 16S RNA and 503 23S rRNA bacterial sequences). One of these features is the absence of an otherwise conserved pairing between 16S rRNA positions 945 and 954 in helix 33 (Fig. 2). The second was observed in the vicinity of the stem formed by the 5′ and 3′ ends of mature 23S rRNA (Fig. 2). Here, the central loop of 23S rRNA consists of eight nucleotides, whereas six are present in nearly all other bacterial 23S rRNAs. The third feature was located in helix 13 of 23S rRNA; it consisted of two noncanonical G-A pairings at the end of the helix, closed by a pentaloop. Except for *Micrococcus luteus*, which has the same constellation, all other bacteria in the database have a shorter helix without G-A pairings, closed by a tetraloop. G-A pairings are only observed in archaea and eukaryotes.

The insertion sequences of high-G+C-content gram-positive bacteria in the 23S rRNA determined by Roller et al. (22) were visually examined for possible conserved base pairings and covariation at paired positions, and the *T. whippelii* insertion sequence was aligned to these sequences with the ARB program package (26). The insertion of *T. whippelii* (80 bp) was quite different from the insertions of other actinomycetes (86 to 116 bp) with respect to both length and primary sequence. It had similarity values of only 25 to 43% with those of other actinomycetes, while the other previously determined insertion sequences (22) had values of 42 to 81% among each other. This significant divergence from the other actinomycetes seems intriguing and suggests an unusual evolutionary history for this organism.

**Conclusions.** The nearly complete *T. whippelii* rRNA operon sequence is currently the largest body of molecular data for this medically important organism, an organism that has otherwise resisted characterization. The sequence will provide targets for new diagnostic primers and probes. Several features are particularly noteworthy, such as (i) the very low degree of nucleotide similarity of the 23S rRNA insertion with those of other actinomycetes and (ii) rare or unique rRNA secondary structures, which support a relatively distant relationship of *T. whippelii* to the cultivated members of the Actinobacteria. The possibility of dual infection of a host by two *T. whippelii* strains is raised for the first time. Our identification of new 16S-23S rRNA intergenic spacer types and downstream genomic polymorphisms may facilitate further analysis of the ecology of this organism and the epidemiology of Whipple’s disease.

**Nucleotide sequence accession numbers.** The rRNA operon sequence determined from patient 1 has been deposited in the GenBank and EMBL databases under the accession no. AF190686. The operon sequences of patients 2 and 3 and the 16S-23S ribosomal RNA intergenic spacer types 4 to 6 of patients 4, 5, and 6 have been assigned accession no. AF190687 through AF190691. The secondary structures of 16S rRNA and 23S rRNA have been deposited at the rRNA secondary structure website (http://www.rna.icmb.utexas.edu/).

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Ma 1663/3-1) to M.M. and by grants from the Lucille P. Markey Charitable Trust and the Donald and Delia Baxter Foundation to D.A.R.

We thank D. H. Persing, P. S. Mitchell, and J. N. Thorvilson (Rochester, Minn.) for providing specimens from 11 patients diagnosed at the Mayo Clinic; M. Schneemann (Zürich, Switzerland) for providing specimens from 2 Swiss patients; I. J. McLaughlin (Applied Biosystems) for technical advice and support, R. R. Gutell (Austin, Tex.) for assistance with rRNA secondary structure analysis, and T. M. Schmidt (Michigan State University, East Lansing, Mich.) for advice regarding rRNA sequence analysis.

REFERENCES
4. Gürrieri, V., and V. A. Stanisich. 1996. New approaches to typing and iden-