INFECTION AND IMMUNITY, Dec. 2007, p. 5859–5866 Vol. 75, No. 12
0019-9567/07/$08.00+0 doi:10.1128/IAI.00709-07

Genome Differences between *Treponema pallidum* subsp. *pallidum* Strain Nichols and *T. paraluiscuniculi* Strain Cuniculi A▽†

Michal Strouhal,[1] David Šmajs,[1,2]* Petra Matějková,[1,2] Erica Sodergren,[2] Anita G. Amin,[2] Jerrilyn K. Howell,[3] Steven J. Norris,[3] and George M. Weinstock[2]

Department of Biology, Faculty of Medicine, Masaryk University, Kamenice 5, Building A6, 625 00 Brno, Czech Republic[1]; Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Alkek N1619, Houston, Texas 77030[2]; and Department of Pathology and Laboratory Medicine, University of Texas-Houston Medical School, 6431 Fannin Street, Houston, Texas 77030[3]

Received 25 May 2007/Returned for modification 20 July 2007/Accepted 8 September 2007

The genome of *Treponema paraluiscuniculi* strain Cuniculi A was compared to the genome of the syphilis spirochete *Treponema pallidum* subsp. *pallidum* strain Nichols using DNA microarray hybridization, whole-genome fingerprinting, and DNA sequencing. A DNA microarray of *T. pallidum* subsp. *pallidum* Nichols containing all 1,039 predicted open reading frame PCR products was used to identify deletions and major sequence changes in the Cuniculi A genome. Using these approaches, deletions, insertions, and prominent sequence changes were found in 38 gene homologs and six intergenic regions of the Cuniculi A genome when it was compared to the genome of *T. pallidum* subsp. *pallidum* Nichols. Most of the observed differences were localized in *tpr* loci and the vicinity of these loci. In addition, 14 other genes were found to contain frameshift mutations resulting in major changes in protein sequences. Analysis of restriction target sites representing 0.34% of the total genome length and DNA sequencing of three PCR products (0.46% of the total genome length) amplified from Cuniculi A chromosomal regions and comparison to the Nichols genome revealed a sequence similarity of 98.6 to 99.3%. These results are consistent with a close genetic relationship among the *T. pallidum* strains and subspecies and a strong, but relatively divergent connection between the human and rabbit pathogens.

The genus *Treponema* comprises five noncultivable species and subspecies showing various degrees of invasiveness and pathogenicity to humans (18). *Treponema pallidum* subsp. *pallidum* is the highly invasive causative agent of syphilis and can cause infection of the central nervous system, cardiovascular system, and almost any other tissue. *T. pallidum* subsp. *pertenue* and *T. pallidum* subsp. *endemicum* are moderately invasive pathogens that cause yaws and endemic syphilis (bejel), respectively; they cause lesions in skin and bone but rarely affect other internal organs. *Treponema carateum* is the causative agent of the noninvasive human disease pinta (25), and *Treponema paraluiscuniculi* is not infectious to humans (10) but causes venereal spirochetosis in rabbits.

The *T. pallidum* subspecies and *T. paraluiscuniculi* cannot be distinguished by morphology, protein content, or physiology (12, 17), suggesting that they are closely related. Serum from rabbits infected with *T. paraluiscuniculi* cross-reacted with 21 of 22 proteins recognized by rabbit antibodies raised against *T. pallidum* subsp. *pallidum* (1). However, in rabbits (which are susceptible to both *T. pallidum* subsp. *pallidum* and *T. paraluiscuniculi* infection) there is no immunological cross-protection against these species (23, 25). In addition to a lack of cross-immunity, these bacterial species differ in their host specificity and the clinical manifestations of the diseases that they cause. Human syphilis is a sexually transmitted disease characterized by infection of a wide spectrum of tissues and organs, multiple stages, persistent infection for years to decades, and variouos clinical manifestations (18), whereas rabbit venereal spirochetosis is characterized by genital lesions (17). These findings suggest that there are important differences between the two species in terms of antigens and virulence factor expression. Genetic differences between *T. pallidum* subsp. *pallidum* and *T. paraluiscuniculi* must account for the observed differences in immunity, host specificity, and clinical manifestations.

Neither *T. pallidum* subsp. *pallidum* nor *T. paraluiscuniculi* has been cultured continuously in vitro, and this fact prevents the use of common molecular genetic approaches to study these pathogens. Sequencing and in silico analysis of the *T. pallidum* subsp. *pallidum* Nichols genome (8, 26) allowed comparison of these genomes by use of comparative genomics methods.

In this study we compared the genomes of *T. pallidum* subsp. *pallidum* Nichols and *T paraluiscuniculi* Cuniculi A using DNA microarray hybridization, whole-genome fingerprinting (WGF), and sequencing of chromosomal regions.

MATERIALS AND METHODS

Isolation of *T. pallidum* and *T. paraluiscuniculi* chromosomal DNA. *T. pallidum* subsp. *pallidum* Nichols and *T. paraluiscuniculi* Cuniculi A were maintained by rabbit inoculation and purified by Hypaque gradient centrifugation as described previously (2, 8). Chromosomal DNA was prepared as described by Fraser et al. (8).

DNA labeling, microarray hybridization, and data analysis. Preparations of *T. pallidum* subsp. *pallidum* and *T. paraluiscuniculi* chromosomal DNA (0.25 to 0.75 μg) were labeled fluorescently using the Klenow enzyme (New England Biolabs, Ipswich, MA) and random nonamers with a CyScribe First-Strand cDNA labeling kit (Amersham Pharmacia Biotech, Piscataway, NJ) according to the protocol described previously (24). Microarrays containing PCR products representing the 1,039 *T. pallidum* subsp. *pallidum* Nichols open reading frames (ORFs) were

* Corresponding author. Mailing address: Department of Biology, Building A6, Faculty of Medicine, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic. Phone: 420 549 497 496. Fax: 420 549 491 327. E-mail: dsmajs@med.muni.cz.

† Supplemental material for this article may be found at http://iai.asm.org/.

▽ Published ahead of print on 24 September 2007.

TABLE 1. DNA microarray-predicted deletions and sequence changes in the *T. paraluiscuniculi* Cuniculi A genome: *T. paraluiscuniculi* ORFs with the lowest $ASR_{Cuniculi\ A/Nichols}$

ORF	$ASR_{Cuniculi\ A/Nichols}$	Predicted protein function (gene)	Family	Predicted ORF length (bp) based on the Nichols genome
TP0136	0.14	Hypothetical protein	PGF15	1,488
TP0127	0.14	Hypothetical protein	PGF14	690
TP0137	0.17	Hypothetical protein	PGF15	138
TP0315	0.19	Hypothetical protein	PGF14	648
TP0619	0.23	Hypothetical protein	PGF14	813
TP0134	0.28	Hypothetical protein	PGF15	376
TP0617	0.46	Hypothetical protein	PGF14	279
TP0462	0.47	Hypothetical protein	PGF15	843
TP0618	0.48	Hypothetical protein	PGF14	360
TP0117	0.49	TprC (*tprC*)	PGF2	1,797
TP0128	0.52	Hypothetical protein		348
TP0135	0.56	Hypothetical protein	PGF15	942
TP0131	0.58	TprD (*tprD*)	PGF2	1,797
TP0621	0.58	TprJ (*tprJ*)	PGF2	2,277
TP0316	0.6	TprF, authentic frameshift (*tprF*)	PGF2	1,195
TP0620	0.61	TprI (*tprI*)	PGF2	1,830
TP0897	0.63	TprK (*tprK*)	PGF2	1,518
TP0129	0.66	Hypothetical protein		477
TP1031	0.66	TprL (*tprL*)	PGF2	1,545
TP0317	0.68	TprG (*tprG*)	PGF2	2,271
TP0970	0.68	Hypothetical protein		117
TP0896	0.7	Hypothetical protein		153

prepared as described by Šmajs et al. (24). The pretreated slides (24) were hybridized simultaneously with labeled DNA using the CyScribe First-Strand cDNA labeling kit (Amersham Pharmacia Biotech). Quantitation of hybridization, exclusion of outliers, and data normalization were performed using the TIGR Spotfinder and TIGR MIDAS software (21). Combining the results of four independent experiments, including dye swapping in two separate hybridizations, yielded 12 possible values for each gene. From these data points, average signal ratios (ASR) and standard deviations were calculated. From these data, a set of genes with mean ASR of labeled *T. paraluiscuniculi* Cuniculi A DNA to labeled *T. pallidum* subsp. *pallidum* Nichols DNA ($ASR_{Cuniculi\ A/Nichols}$) less than 0.7 (average $\log_2$ ratio less than −0.51) was derived. This set comprised 22 genes that are likely to contain deletions and/or major sequence changes. No genes with a mean ASR greater than 1.43 (average $\log_2$ ratio greater than 0.51) were identified.

WGF. WGF was performed as described previously (27). The chromosomal DNA was amplified in 97 overlapping regions with a median length of 12,307 bp (range, 1,778 to 24,758 bp) using a GeneAmp XL PCR kit (Applied Biosystems, Foster City, CA). The primer pairs used for these amplifications are shown in Table S1 in the supplemental material. Each PCR product was digested with BamHI, EcoRI, or HindIII or combinations of these enzymes. To thoroughly assess the possible presence of deletions and insertions in restriction fragments, additional digestions were performed as needed to reduce the length of each restriction fragment to ≤4 kb. This was achieved by additional digestion with AccI, ClaI, EcoRV, KpnI, MluI, NcoI, NheI, RsrII, SacI, SpeI, XbaI, or XhoI (NEB) or combinations of these enzymes. The resulting fingerprints for *T. pallidum* subsp. *pallidum* Nichols were compared to those for the *T. paraluiscuniculi* Cuniculi A genome.

PCR amplification and DNA sequencing. Standard methods were used for PCR amplification from a chromosomal DNA template and agarose gel electrophoresis (22). For sequencing of PCR products, XL PCR was used to minimize the number of PCR errors. Oligonucleotide primers were designed with Primer3 software (20). The resulting PCR products were purified using a QIAquick PCR purification kit (QIAGEN) and were sequenced using a *Taq* DyeDeoxy terminator cycle sequencing kit (Applied Biosystems). Complete sequences of amplified regions were finished using specifically designed synthetic oligonucleotides as primers. Computer-assisted sequence analysis was performed using the LASERGENE program package (DNASTAR, Madison, WI). Three XL PCR products comprising regions TPI12, TPI25A, and TPI25B (see Table S1 in the supplemental material) were purified and subjected to mechanical shearing to obtain smaller fragments (0.5 to 1 kb) that were cloned into the pUC18 vector. The resulting recombinant plasmids (96 plasmids for each XL PCR product) of the small insert library were isolated and sequenced using forward and reverse pUC18 primers to obtain multiple coverage (i.e., 2 × 96 sequencing reactions per XL PCR product).

Nucleotide sequence accession numbers. The nucleotide sequences reported in this study have been deposited in the GenBank database under accession numbers EF057750, EF137736 to EF137743, and EF419245 to EF419253.

RESULTS

DNA microarray genome comparison. In this analysis, we used a *T. pallidum* subsp. *pallidum* Nichols DNA microarray containing PCR products corresponding to all 1,039 predicted ORFs (24). *T. pallidum* subsp. *pallidum* Nichols and *T. paraluiscuniculi* Cuniculi A chromosomal DNA were labeled with the Cy3 and Cy5 dyes and hybridized simultaneously on individual arrays. The hybridizations were performed four times, including dye swapping in two hybridizations, resulting in a total of 12 hybridizations for each ORF. Hybridization of labeled *T. pallidum* subsp. *pallidum* Nichols DNA probes yielded a detectable signal in ≤6 of the 12 reactions for 11 ORFs (TP0161, TP0224, TP0490, TP0573, TP0645, TP0753, TP0777, TP0795, TP0818, TP0932, and TP1032), and therefore analysis of these ORFs was not performed. All of these ORFs are relatively short (93, 105, 189, 93, 177, 285, 225, 159, 153, 93, and 432 bp, respectively) and code for a conserved hypothetical protein (TP0490) or hypothetical proteins (TP0161, TP0224, TP0573, TP0645, TP0753, TP0777, TP0795, TP0818, TP0932, and TP1032). Thus, data were calculated for 1,028 of 1,039 genes (99%) by determining the $ASR_{Cuniculi\ A/Nichols}$ representing the average, normalized ratio of *T. paraluiscuniculi* Cuniculi A DNA fluorescent signals to *T. pallidum* subsp. *pallidum* Nichols DNA fluorescent signals for replicate spots on each microarray and in replicate experiments. A value of 1.0 corresponded to the mean signal for all genes of the array.

The results of the DNA microarray hybridizations are shown in Table 1. Use of the Cuniculi A probe yielded significantly

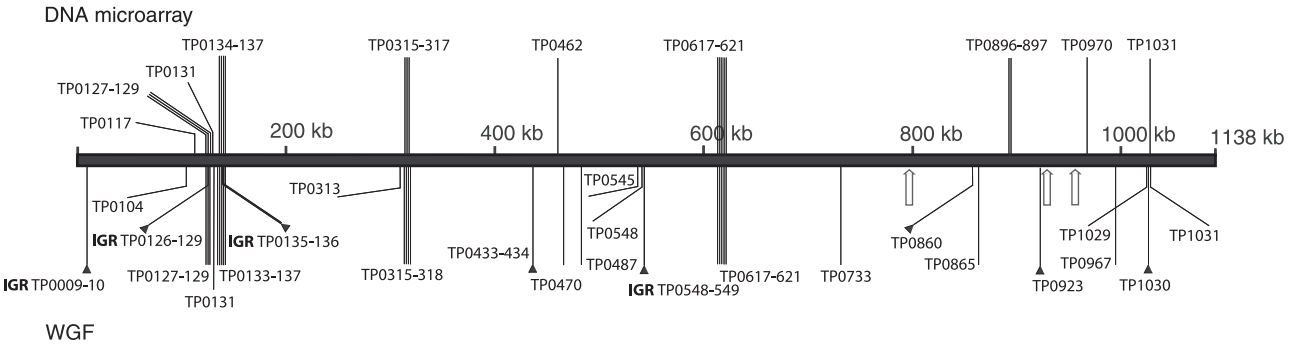

FIG. 1. Schematic representation of genes detected by lower-microarray-signal (DNA microarray) analysis, indels detected by WGF, and genes selected for sequencing in the *T. paraluiscuniculi* Cuniculi A genome. Deletions and sequentially diverse genes are indicated by vertical bars, and insertions are indicated by vertical lines ending with triangles. Open arrows indicate chromosomal regions that were sequenced in the Cuniculi A genome.

lower signals for 22 genes, with $ASR_{Cuniculi\ A/Nichols}$ ranging from 0.14 to 0.7 (Table 1). These genes were not randomly distributed throughout the genome (Fig. 1) and tended to be clustered in regions containing *tpr* genes and genes in the vicinity of *tpr* genes. All but four putative genes (TP0128, TP0129, TP0896, and TP0970) belonged to paralogous gene family 2 (PGF2), PGF14, and PGF15. PGF2 represents *tpr* genes encoding Tpr proteins, which are *T. pallidum*-specific proteins of unknown function with similarity to the *Treponema denticola* membrane protein Msp (7). Eight *tpr* genes (*tprC*, *tprD*, *tprF*, *tprG*, *tprI*, *tprJ*, *tprK* and *tprL*) had significantly lower $ASR_{Cuniculi\ A/Nichols}$ values, indicating that there were deletions or sequence diversity in the *T. paraluiscuniculi* Cuniculi A genes. In contrast, signals for the *tprA*, *tprB*, *tprE*, and *tprH* genes were similar in the two genomes, indicating that these genes are present in the Cuniculi A genome and that the Cuniculi A genes are highly homologous to their Nichols counterparts. Genes belonging to PGF14 and PGF15 encode hypothetical proteins with unknown functions encoded by genes in the vicinity of *tpr* genes. The $ASR_{Cuniculi\ A/Nichols}$ values for PGF14 and PGF15 are considerably lower (range, 0.14 to 0.56) than the $ASR_{Cuniculi\ A/Nichols}$ values for *tpr* genes (0.49 to 0.68). The average lengths of the *tpr* genes and the PGF14 and PGF15 genes listed in Table 1 were 1,779 bp and 658 bp, respectively, so the PCR products used in the microarray were longer for the *tpr* genes. Therefore, this approach may be less sensitive to detection of sequence changes when larger genes are examined.

WGF of Nichols and Cuniculi A genomes. The genomes of *T. pallidum* subsp. *pallidum* strain Nichols and *T. paraluiscuniculi* strain Cuniculi A were analyzed using 97 overlapping XL PCR amplicons covering the entire treponemal genomes. Restriction mapping of 97 XL PCR products revealed 20 chromosomal regions (in 18 TPI intervals [Table 2]) with detectable insertions or deletions (indels) in the Cuniculi A genome (Table 2). Subsequent sequencing of heterologous parts of these TPI intervals revealed 10 deletions ranging from 15 to 2,609 bp and eight insertions ranging from 41 to 885 bp, as well as two regions (TPI12-13 and TPI66A) with both deletions and insertions. In addition, eight genes were found to contain small indels and multiple single nucleotide polymorphisms (SNPs) (TP0133, TP0136, TP0137, TP0315, TP0548, TP0617, TP0618, and TP1031 [Table 3]), and three *tpr* genes contained multiple SNPs and single nucleotide indels that resulted in frameshift gene mutations (TP0131, TP0313, and TP0621 [Table 3]). The largest deletions comprised TP0127 to TP0129, TP0133 to TP0135, *tprF* (TP0316) and *tprG* (TP0317), and TP0619 and *tprI* (TP0620). Deletions in the *tprF* and *tprG* genes (Table 2) and a frameshift mutation in *tprG* (Table 3) resulted in a new ORF encoding a 429-amino-acid TprI-like protein. A similar finding in the *T. paraluiscuniculi* Cuniculi A genome was described by Giacani et al. (9). The largest insertion was localized in the intergenic region (IGR) between the TP0126 and TP0129 loci (the TP0127 and TP0128 genes are missing in the Cuniculi A genome) and included 1,451 bp containing three ORFs showing similarity to the *tprK* sequence. Another large insertion (885 bp) comprising 14 complete repetitions (length, 60 bp) and one incomplete repetition (45 bp) was found in overlapping parts of the TP0433 and TP0434 genes. The expansion of repetitive sequences was shown by gel electrophoresis and draft sequencing of tandem repeats (data not shown). Because of the tandem repeats, this region was not sequenced completely, and the number of repeats was estimated based on gel electrophoresis. The described insertion resulted in fusion of the two genes (TP0433 and TP0434) to form a single acidic repeat protein gene (*arp*) (19). Except for deletions of TP0104 (encoding a 5′ nucleotidase), the *tprF*, *tprG*, and *tprI* genes, and TP0545 (encoding MglB-1, a methylgalactoside ABC transporter, periplasmic galactose-binding protein), all other indels were localized in genes encoding hypothetical proteins or in intergenic regions.

Taken together, the indels identified resulted in complete deletion of five genes (TP0127, TP0128, TP0134, TP0619, and TP0620), 13 partial gene deletions (TP0104, TP0129, TP0135, TP0136, TP0315, TP0316, TP0317, TP0470, TP0545, TP0733, TP0865, TP0967, and TP1029), seven gene elongations (TP0133, TP0487, TP0548, TP0616, TP0860, TP0923, and TP1031), two gene fusions (TP0433-TP0434 and TP0617-TP0618), two deletions in IGR (TP0135-TP0136 and TP0545-TP0546), and four insertions in IGR (TP0009-TP0010, TP0126-TP0129, TP0135-TP0136, and TP0548-TP0549). In addition, 14 frameshifts were identified in TP0126, TP0131, TP0132, TP0309, TP0310, TP0311, TP0313, TP0317, TP0318, TP0487, TP0621, TP0922, TP0968, and

TABLE 2. Comparison of the *T. pallidum* subsp. *pallidum* Nichols and *T. paraluiscuniculi* Cuniculi A genomes using WGF and subsequent DNA sequencing of identified diverse regions: prominent insertions and deletions

XL PCR-amplified region (TPI interval)	Detected indel[a]	Affected IGR or gene(s)	Protein or gene change[b]	Putative function, biological relevance, and/or BLAST hit[b]	GenBank accession no.[a]
TPI2	Insertion (41 bp) (10208–10209)	IGR TP0009-TP0010		*tprA* sequence (100% identity; nt 421 to 381), *tprB* sequence (100% identity; nt 216 to 256)	EF137737 (9630–10791)
TPI8	Deletion (153 bp) (115478–115630)	TP0104	51-aa deletion (aa 541 to 591) at C terminus	UshA (5′ nucleotidase), putative substrate binding site[c]	EF057750 (115191–115911)
TPI12-13	Deletion (228 bp) and insertion (1,451 bp) between coordinates 148296 and 150385 (411 nt in this region remained as TP0129 sequence)	TP0127, TP0128 TP0129 IGR TP0126-TP0129	Genes completely deleted Gene partially deleted	Hypothetical proteins Hypothetical protein TP0129, ORF encoding 136 aa remained Insertion (1,451 bp) of *tprK*-like sequence with three ORFs encoding 121, 155, and 184 aa	EF137743 (145689–159352)
	Deletion (1,063 bp) between coordinates 153136 and 156388 (2,190 nt in this region remained)	TP0134 TP0135	Gene completely deleted Gene partially deleted	Hypothetical protein Hypothetical TP0135-like sequence with two ORFs encoding 228 and 103 aa and additional hypothetical TP0126-like ORF encoding 292 aa	
	Deletion (63 bp) (156459–156521)	IGR TP0135-TP0136			
	Insertion (97 bp) (156750–156751)	IGR TP0135-TP0136		No BLAST hit	
TPI25A-25B	Deletion (1,660 bp) between coordinates 331139 and 334663 (1,864 nt in this region remained)	TP0316, TP0317	Genes partially deleted	TprF and TprG proteins; additional ORF encoding 429 aa with 71% aa identity to TprI; nonfunctional *tprG* and *tprI*; *tprG*/*tprI* hybrid[d]	EF137742 (325013–337122)
TPI32B	Insertion (885 bp) (461015–461517)[e]	TP0433, TP0434	14 complete repetitions plus 45 nt of incomplete repetition inserted	Arp protein, fusion of TP0433 and TP0434 genes to *arp* gene[f]	EF137736 (460718–464495)
TPI34	Deletion (264 bp) (497265–497688)[g]	TP0470	Gene partially deleted, 11 tandem repetitions deleted	Conserved hypothetical protein; 88 aa (aa 246 to 333) deleted; no protein domain was identified in the deleted region	EF137740 (496637–498283)
TPI38	Insertion (66 bp) (519663–519664)	TP0487	22-aa elongation	Hypothetical TP0487-like protein	EF419245 (519481–521656)
TPI41-42A	Deletion (103 bp) (588576–588678)	TP0545, IGR TP0545-TP0546	11 aa of signal peptide deleted from TP0545 (*mglB-1*)	Methylgalactoside ABC transporter; periplasmic galactose-binding protein	EF137738 (587756–594311)
	Insertion (52 bp) (593023–593024)	IGR TP0548-TP0549		No BLAST hit	
TPI48	Deletion (2,609 bp) between coordinates 669739 and 672944 (597 nt in this region remained as TP0618-like sequence)	TP0619, TP0620	Genes completely deleted	Hypothetical protein TP0619 and TprI	EF137741 (667318–678691)
TPI57A	Deletion (21 bp) between coordinates 798596 and 798646 (30 nt in this region remained)	TP0733	Gene partially deleted, 7 aa deleted	Hypothetical protein	EF419246 (797469–800801)
TPI65	Insertion (30 bp) (936629–936630)	TP0860	10-aa elongation	Hypothetical protein	EF419247 (936435–936924)

Continued on following page

TABLE 2—*Continued*

XL PCR-amplified region (TPI interval)	Detected indel[a]	Affected IGR or gene(s)	Protein or gene change[b]	Putative function, biological relevance, and/or BLAST hit[b]	GenBank accession no.[a]
TPI66A	Deletion (12 bp) (944598–944609) Insertion (18 bp) (944541–944542) Deletion (21 bp) (944086–944106)	TP0865	Gene partially deleted, 5 aa deleted	Hypothetical protein	EF419248 (943859–945703)
TPI68	Insertion (330 bp) between coordinates 1001719 and 1002228 (510 nt in this region remained)	TP0923	105-aa elongation	No BLAST hit, hypothetical protein	EF419249 (999546–1002637)
TPI71	Deletion (15 bp) between coordinates 1050290 and 1050319 (15 nt in this region remained)	TP0967	Gene partially deleted, 5 aa deleted	Hypothetical protein	EF419250 (1049862–1051795)
TPI77	Deletion (18 bp) (1123530–1123547) Insertion (67 bp) (1124102–1124103)	TP1029 TP1030	Gene partially deleted, 6 aa deleted	Hypothetical protein No BLAST hit, hypothetical protein	EF137739 (1122770–1125482)

[a] The numbers in parentheses are the coordinates in the strain Nichols genome.
[b] aa, amino acid.
[c] See reference 8.
[d] See reference 9.
[e] Insertion into the chromosomal region containing tandem repetitions (length, 60 nt) (i.e., increased number of tandem repetitions).
[f] See reference 19.
[g] Deletion in the chromosomal region containing tandem repetitions (length, 24 nt) (i.e., decreased number of tandem repetitions).

TP1030 leading to premature termination of 10 genes, elongation of three genes, and a reading frame change in one gene (Table 3).

Hypothetical proteins were characterized by searching the InterPro and Pfam databases and constructing hydrophobicity plots. Of the 25 hypothetical genes described in this study (Tables 1 to 3), 17 were completely sequenced in both the Cuniculi A and Nichols strains. The corresponding 17 hypothetical proteins were analyzed to predict cellular localization. Signal sequences were predicted for six Nichols proteins (encoded by TP0133, TP0134, TP0135, TP0136, TP0548, and TP0733) and five Cuniculi A proteins (encoded by TP0315, TP0470, TP0548, fused genes TP0617 and TP0618, and TP0733). In both strains, transmembrane regions were predicted for TP0733. Three putative protein domains (UPF0164, TPR, and DbpA) were found in five hypothetical proteins (encoded by TP0470, TP0548, TP0860, TP0865, and TP1029). No differences between the Nichols and Cuniculi A strains were found in domain distribution.

DNA sequencing of Cuniculi A chromosomal regions. To determine the level of sequence identity between the Nichols and Cuniculi A genomes, three chromosomal regions that also included IGR were sequenced. Analysis of these regions (5,289 bp; 0.46% of the genome), comprising genes TP0798 to TP0800 (accession number EF419251), TP0933 and TP0934 (accession number EF419252), and TP0961 and TP0962 (accession number EF419253), revealed 37 SNPs that resulted in 11 amino acid substitutions in the corresponding proteins. The average density of SNPs represented one nucleotide change per 143 bp (99.3% identity).

DISCUSSION

Using both DNA microarray and WGF approaches, deletions, insertions, and prominent sequence changes in 38 *T. paraluiscuniculi* Cuniculi A gene homologs were found when genes of this strain were compared to annotated genes of *T. pallidum* subsp. *pallidum* Nichols (8). In addition, 14 genes were found to contain frameshift mutations suggesting inactivation or changed functions of the genes. DNA microarray hybridization of labeled chromosomal DNA revealed 22 ORFs (predicted genes) with a lower signal for the *T. paraluiscuniculi* Cuniculi A genome than for the Nichols genome, indicating possible deletions or sequence diversity of the corresponding chromosomal loci. An alternative approach, WGF with subsequent sequencing, revealed 20 chromosomal regions (in 18 TPI intervals) with indels larger than 10 bp in the coding regions (22 genes with detectable indels), and six indels were found in the intergenic regions of the *T. paraluiscuniculi* Cuniculi A genome. An additional 11 genes were found to contain small indels and multiple SNPs (TP0131, TP0133, TP0136, TP0137, TP0313, TP0315, TP0548, TP0617, TP0618, TP0621, and TP1031). Seventeen of the gene deletions or sequentially diverse genes (i.e., genes with multiple nucleotide changes) were detected both by DNA microarray analysis (17 of 22 gene deletions and sequentially diverse genes) and by WGF (17 of 27 gene deletions and sequentially diverse genes) (Fig. 1). Compared to the DNA microarray approach, WGF detected 10 additional deletions or sequentially diverse genes (TP0104, TP0133, TP0313, TP0470, TP0545, TP0548, TP0733, TP0865, TP0967, and TP1029). With the exception of TP0133 and

TABLE 3. Comparison of the *T. pallidum* subsp. *pallidum* Nichols and *T. paraluiscuniculi* Cuniculi A genomes using WGF and subsequent DNA sequencing of identified diverse regions: major sequence changes (multiple SNPs) and frameshifts

Affected gene(s)	Detected multiple SNPs or frameshift[a]	Protein change[b]
TP0126	Deletion (1 nt) resulting in frameshift at nt 148 (148340)	Truncated hypothetical protein TP0126 (227 aa)
TP0131 (*tprD*)	Insertion (1 nt) resulting in frameshift between nt 438 and 439 (152459–152460) Multiple SNPs	Two hypothetical proteins, TprD2a and TprD2b (196 and 463 aa, respectively) Nonfunctional *tprD2*[c]
TP0132	Insertion (1 nt) resulting in frameshift between nt 28 and 29 (153123–153124)	Truncated hypothetical protein TP0132 (64 aa)
TP0133	Multiple SNPs and small indels	Hypothetical TP0133-like protein (420 aa); 74.7% aa identity to Nichols TP0133 homolog (410 aa); 10-aa elongation at N terminus
TP0136	Multiple SNPs and small indels	Hypothetical protein (437 aa); 42.7% aa identity to the Nichols homolog TP0136 (495 aa); 74.6% identity the Nichols homolog TP0133 (410 aa); 96.1% identity to the Cuniculi A homolog TP0133 (420 aa)
TP0137	Multiple SNPs and small indels	Hypothetical protein TP0137 (45 aa)
TP0309	Deletion (2 nt) resulting in frameshift at nt 207 and 208 (325811–325812)	Amino acid ABC transporter; periplasmic binding protein; two truncated TP0309 proteins (107 and 288 aa)
TP0310	Insertion (1 nt) resulting in frameshift between nt 375 and 376 (326766–326767)	49-aa elongation of hypothetical protein TP0310 at C terminus (176 aa)
TP0311	Insertion (1 nt) and deletion (1 nt) between coordinates 326766 and 326767 and at coordinate 326872, respectively	Hypothetical protein TP0311 (47 aa); 29.2% aa identity to the Nichols homolog TP0311 (47 aa)
TP0313 (*tprE*)	Insertion (1 nt) resulting in frameshift between nt 652 and 653 (328636–328637) Multiple SNPs	Two hypothetical proteins, TprEa and TprEb (260 and 560 aa, respectively) Nonfunctional *tprG1*; *tprG*/*J* hybrid[d]
TP0315	Multiple SNPs and small indels	Truncated hypothetical protein TP0315 (120 aa)
TP0317 (*tprG*)[e]	Insertion (1 nt) resulting in frameshift between nt 652 and 653 (334011–334012) Deletion[e]	Nonfunctional *tprG* and *tprI*; *tprG*/*I* hybrid[c]
TP0318	Deletion (2 nt) resulting in frameshift at nt 121 and 122 (334673–334674)	Truncated hypothetical protein TP0318 (40 aa)
TP0487[e]	Insertion (1 nt) resulting in frameshift between 520371–520372 Insertion[e]	10-aa elongation of hypothetical protein TP0487 at N terminus (535 aa)
TP0548	Multiple SNPs and small indels	Hypothetical protein TP0548 (437 aa); 82.8% aa identity to Nichols homolog (434 aa)
TP0616	$G_{668671} \rightarrow A$ ($M_1 \rightarrow I$), missing start codon at coordinates 668669 to 668671	29-aa elongation of hypothetical protein TP0616 at N terminus (283 aa)
TP0617, TP0618	Multiple SNPs and small indels	Fusion of genes TP0617 and TP0618 (254 aa)
TP0621 (*tprJ*)	Insertion (1 nt) resulting in frameshift between nt 652 and 653 (674569–674570) Multiple SNPs	Two hypothetical proteins, TprJa and TprJb (260 and 580 aa, respectively) Nonfunctional *tprG2* allele; *tprG*/*J* hybrid[d]
TP0922	Insertion (1 nt) and insertion (2 nt) between coordinates 1001601 and 1001602 and between coordinates 1000784 and 1000785, respectively	Hypothetical protein (350 aa) related to TDE0306 protein encoded in *T. denticola* genome[f]
TP0968	Insertion (1 nt) resulting in frameshift between coordinates 1051760 and 1051761	Possible truncated hypothetical protein TP0968
TP1030[e]	Insertion (1 nt) resulting in frameshift between nt 138 and 139 (1124188–1124189) Insertion[e]	Truncated hypothetical protein TP1030 (51 aa)
TP1031 (*tprL*)	Multiple SNPs	98-aa elongation of TprL protein at N terminus (possible fusion of the TP1030 and TP1031 genes)

[a] The numbers in parentheses are the coordinates in the strain Nichols genome.
[b] aa, amino acids.
[c] See reference 9.
[d] See references 9 and 11.
[e] See Table 2.
[f] See reference 7.

TP0470, relatively small deletions (i.e., deletions ranging from 0.74 to 8.58% of the total gene length) were detected by WGF and missed by DNA microarray hybridization. It is likely that such deletions cannot be detected by DNA microarray hybridization under the conditions used. The sequences present in the TP0136 locus in the Cuniculi A genome are similar to the Nichols TP0133 sequences. Therefore, this false-negative result for TP0133 obtained with the Cuniculi A DNA microarray was likely due to DNA cross-hybridization of labeled TP0136 DNA. The deleted region of TP0470 (23.8% of the gene length) comprises a chromosomal region containing tandem repetitions (length, 24 nucleotides [nt]); i.e., the deletion resulted in a decreased number of tandem repetitions. This region also showed interstrain genetic heterogeneity within *T. pallidum* strains (data not shown). In this locus, PCR products of variable lengths were also observed after amplification from the Cuniculi A DNA, suggesting possible intrastrain heterogeneity or PCR artifacts. Populations of spirochetes containing different numbers of tandem repetitions may distort the results of DNA microarray hybridization analyses.

Compared to the WGF results, the DNA microarray approach identified five additional genes (TP0117, TP0462, TP0896, TP0897, and TP0970) with lower hybridization signals. Two of these genes belonged to PGF2, one belonged to PGF15, and two were unique (TP0896 and TP0970). Sequence diversity of these genes was identified as the reason for the lower hybridization signals on the DNA microarray. In these genes, the sequence diversity was dispersed throughout the entire genes and thus had the potential to affect hybridization to a DNA microarray (P. Matějková, unpublished results). DNA microarray and WGF approaches thus represent complementary methods; DNA microarray analysis allows selective detection of diverse chromosomal regions, and WGF allows selective identification of insertions within the genes and indels in intergenic regions.

Several of the observed indels and sequence changes were identified in the family of *tpr* genes (in 8 of 12 *tpr* genes). The *T. pallidum* repeat (*tpr*) genes encode paralogous proteins with sequence similarity to the major outer sheath protein (Msp) of *T. denticola* (7). The *tpr* genes are specific for *T. pallidum* and *T. paraluiscuniculi*, and several of them show heterogeneity both within and between the *T. pallidum* subspecies and strains examined (3, 4, 5). It is believed that the Tpr proteins are involved in pathogenesis and/or immune evasion. The TprK protein was found to induce a strong humoral and cellular immune response (3, 14, 15), and variable regions of TprK are responsible for the specificity of the antibody response (16). Moreover, sequences of variable regions of TprK change during infection and passage of *T. pallidum* subsp. *pallidum* strains (6) by a gene conversion mechanism with donor sites in the vicinity of *tpr* genes (e.g., in TP0137 and in TP0126 to TP0130). Thus, some of the observed genetic differences in the *tprK* locus of the *T. paraluiscuniculi* genome may also be due to this gene conversion mechanism. In addition, three new ORFs in the *T. paraluiscuniculi* genome with *tprK*-like sequences were identified.

With the exception of *tpr* genes, the TP0104 gene (5′ nucleotidase), and the TP0545 gene (periplasmic galactose-binding protein), all other detected indels or sequence changes were localized in the genes encoding a conserved hypothetical protein (TP0470) or hypothetical proteins. The average transcription rate of these genes in *T. pallidum* subsp. *pallidum* cultivated in rabbit testes is considerably higher (1.74) than the average transcription rate of all genes of *T. pallidum* subsp. *pallidum* Nichols (1.0) (24). In addition, 8 of 29 (27.6%) of the proteins encoded by these genes were found to be recognized by serum antibodies derived from rabbits 84 days after infection with the Nichols strain (13). Both of these findings indicate that several of the putative genes identified are transcribed and translated and suggest that these *T. pallidum* subsp. *pallidum* genes are important during infection of rabbits. Most of these genes (17 of 29) were localized in the vicinity of *tpr* genes. Insertions identified in the *T. paraluiscuniculi* genome indicated that the sequences were *tprK*-like, *tprA* or *tprB* sequences, or unique sequences with no homologous sequences identified by the BLAST search. Deletion of the signal sequence peptide in MglB-1 encoded by TP0545 in the Cuniculi A strain may result in aborted export of this protein to the periplasm.

Seventeen hypothetical proteins were analyzed to predict cellular localization. Signal sequences were predicted in six and five proteins encoded in the Nichols and Cuniculi A genomes, respectively. Except for two hypothetical proteins (TP0548 and TP0733), signal sequences were predicted for different Nichols and Cuniculi A proteins. Possible localization of these proteins outside the cytoplasm may contribute to the different host ranges and pathogenicities of the Nichols and Cuniculi A strains.

A portion of the TPI12 region of the *T. paraluiscuniculi* genome sequenced in this study was nearly identical to a previously sequenced 2,792-nt region (accession number AY685237) comprising a nonfunctional *tprD2* gene (9). Differences in 9 nt were found. Other regions of near identity with previously sequenced regions (9) were found in TPI2 and the accession number AY685232 sequence (*tprA*, 1,003 nt), in TPI2 and the accession number AY685233 sequence (*tprB*, 838 nt), in TPI25A and the accession number AY685239 sequence (nonfunctional *tprG1*, 3,255 nt), in TPI25B and the accession number AY685238 sequence (nonfunctional *tprG* and *tprI*, 2,449 nt), in TPI48 and the accession number AY685240 sequence (nonfunctional *tprG2*, 3,018 nt), and in TPI77 and the accession number AY685235 sequence (*tprL*, 1,331 nt). Within these regions, two, zero, seven, four, nine, and three nucleotide differences were found, respectively. These results could reflect differences accumulated in the Cuniculi A genome during independent cultivation in different laboratories; they potentially could also be due to PCR errors. It was previously shown that in the *tprK* locus (TP0897), sequence changes occurred during infection and passage of *T. pallidum* subsp. *pallidum* strain Chicago (6).

Altogether, 639 target restriction sites (representing 3.8 kb of the genomic sequence or 0.34% of the Nichols genome) in the Cuniculi A genome were analyzed with three enzymes (BamHI, HindIII, and EcoRI). Assuming that the majority of additional or missing restriction target sites were due to single nucleotide changes, the sequence similarity of the Cuniculi A and Nichols genomes could be predicted to be 98.6%. Sequencing of three chromosomal regions representing 0.46% of the Cuniculi A genome revealed a sequence identity of 99.3%. However, the latter result is a rather high estimate of the sequence identity because the value could be distorted by a

number of factors, including nonrandom distribution of sequenced DNA and the fact that the sequentially divergent regions in the Cuniculi A strain appear to be localized in certain chromosome regions.

The data presented indicate that the genomes of *T. pallidum* subsp. *pallidum* and *T. paraluiscuniculi* are very closely related and that most of the observed differences are localized in *tpr* loci and in the vicinity of these loci, suggesting their possible role in the host range and pathogenicity of *T. pallidum* subsp. *pallidum*. The high degree of sequence similarity of the genomes tested could be used for planning an optimal genome sequencing strategy. In further studies, the high level of relatedness of the *T. pallidum* subsp. *pallidum*, *T. pallidum* subsp. *pertenue*, and *T. paraluiscuniculi* genomes could be used for identifying and deciphering *T. pallidum* subsp. *pallidum* virulence determinants.

ACKNOWLEDGMENTS

We thank S. Lukehart for providing the *T. paraluiscuniculi* Cuniculi A strain.

This work was supported by Public Health Service grants to G.M.W. (grants R01 DE12488 and R01 DE13759) and S.J.N. (grants R01 AI49252 and R03 AI69107) and by grants 310/04/0021 and 310/07/0321 from the Grant Agency of the Czech Republic, grant NR8967-4/2006 from the Ministry of Health of the Czech Republic, and grant VZ MSM0021622415 from the Ministry of Education of the Czech Republic to D.S.

REFERENCES

1. **Baker-Zander, S. A., and S. A. Lukehart.** 1984. Antigenic cross-reactivity between *Treponema pallidum* and other pathogenic members of the family *Spirochaetaceae*. Infect. Immun. **46:**116–121.
2. **Baseman, J. B., J. C. Nichols, O. Rumpp, and N. S. Hayes.** 1974. Purification of *Treponema pallidum* from infected rabbit tissue: resolution into two treponemal populations. Infect. Immun. **10:**1062–1067.
3. **Centurion-Lara, A., C. Castro, L. Barrett, C. Cameron, M. Mostowfi, W. C. Van Voorhis, and S. A. Lukehart.** 1999. Treponema pallidum major sheath protein homologue TprK is a target of opsonic antibody and the protective immune response. J. Exp. Med. **189:**647–656.
4. **Centurion-Lara, A., C. Godornes, C. Castro, W. C. Van Voorhis, and S. A. Lukehart.** 2000. The *tprK* gene is heterogeneous among *Treponema pallidum* strains and has multiple alleles. Infect. Immun. **68:**824–831.
5. **Centurion-Lara, A., E. S. Sun, L. K. Barrett, C. Castro, S. A. Lukehart, and W. C. Van Voorhis.** 2000. Multiple alleles of *Treponema pallidum* repeat gene D in *Treponema pallidum* isolates. J. Bacteriol. **182:**2332–2335.
6. **Centurion-Lara, A., R. E. LaFond, K. Hevner, C. Godornes, B. J. Molini, W. C. Van Voorhis, and S. A. Lukehart.** 2004. Gene conversion: a mechanism for generation of heterogeneity in the tprK gene of Treponema pallidum during infection. Mol. Microbiol. **52:**1579–1596.
7. **Fenno, J. C., K. H. Muller, and B. C. McBride.** 1996. Sequence analysis, expression, and binding activity of recombinant major outer sheath protein (Msp) of *Treponema denticola*. J. Bacteriol. **178:**2489–2497.
8. **Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, C. Fujii, S. Garland, B. Hatch, K. Horst, K. Roberts, M. Sandusky, J. Weidman, H. O. Smith, and J. C. Venter.** 1998. Complete genome sequence of Treponema pallidum, the syphilis spirochete. Science **281:**375–388.
9. **Giacani, L., E. S. Sun, K. Hevner, B. J. Molini, W. C. Van Voorhis, S. A. Lukehart, and A. Centurion-Lara.** 2004. Tpr homologs in *Treponema paraluiscuniculi* Cuniculi A strain. Infect. Immun. **72:**6561–6576.
10. **Graves, S., and J. Downes.** 1981. Experimental infection of man with rabbit-virulent Treponema paraluis-cuniculi. Br. J. Vener. Dis. **57:**7–10.
11. **Gray, R., C. Mulligan, B. Molini, E. S. Sun, L. Giacani, C. Godornes, A. Kitchen, S. A. Lukehart, and A. Centurion-Lara.** 2006. Molecular evolution of the tprC, D, I, K, G, and J. genes in the pathogenic genus Treponema. Mol. Biol. Evol. **23:**2220–2233.
12. **Hovind-Hougen, K., A. Birch-Andersen, and H. J. Jensen.** 1973. Electron microscopy of Treponema cuniculi. Acta Pathol. Microbiol. Scand. Sect. B Microbiol. Immunol. **81:**15–28.
13. **McKevitt, M., M. B. Brinkman, M. McLoughlin, C. Perez, J. K. Howell, G. M. Weinstock, S. J. Norris, and T. Palzkill.** 2005. Genome scale identification of *Treponema pallidum* antigens. Infect. Immun. **73:**4445–4450.
14. **Morgan, C. A., S. A. Lukehart, and W. C. Van Voorhis.** 2002a. Immunization with the N-terminal portion of *Treponema pallidum* repeat protein K attenuates syphilitic lesion development in the rabbit model. Infect. Immun. **70:**6811–6816.
15. **Morgan, C. A., B. J. Molini, S. A. Lukehart, and W. C. Van Voorhis.** 2002b. Segregation of B and T cell epitopes of Treponema pallidum repeat protein K to variable and conserved regions during experimental syphilis infection. J. Immunol. **169:**952–957.
16. **Morgan, C. A., S. A. Lukehart, and W. C. Van Voorhis.** 2003. Protection against syphilis correlates with specificity of antibodies to the variable regions of *Treponema pallidum* repeat protein K. Infect. Immun. **71:**5605–5612.
17. **Norris, S. J., D. L. Cox, and G. M. Weinstock.** 2001. Biology of Treponema pallidum: correlation of functional activities with genome sequence data. J. Mol. Microbiol. Biotechnol. **3:**37–62.
18. **Norris, S. J., V. Pope, R. E. Johnson, and S. A. Larsen.** 2003. Treponema and other human host-associated spirochetes, p. 955–971. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, J. H. Jorgensen, and R. H. Yolken (ed.), Manual of clinical microbiology, 8th ed. ASM Press, Washington, DC.
19. **Pillay, A., H. Liu, C. Y. Chen, B. Holloway, A. W. Sturm, B. Steiner, and S. A. Morse.** 1998. Molecular subtyping of Treponema pallidum subspecies pallidum. Sex. Transm. Dis. **25:**408–414.
20. **Rozen, S., and H. J. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365–386. *In* S. Krawetz and S. Misener (ed.), Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ.
21. **Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush.** 2003. TM4: a free, open-source system for microarray data management and analysis. BioTechniques **34:**374–378.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23. **Schell, R. F., A. A. Azadegan, S. G. Nitskansky, and J. L. LeFrock.** 1982. Acquired resistance of hamsters to challenge with homologous and heterologous virulent treponemes. Infect. Immun. **37:**617–621.
24. **Šmajs, D., M. McKevitt, J. K. Howell, S. J. Norris, W. W. Cai, T. Palzkill, and G. M. Weinstock.** 2005. Transcriptome of *Treponema pallidum*: gene expression profile during experimental rabbit infection. J. Bacteriol. **187:**1866–1874.
25. **Turner, T. B., and D. H. Hollander.** 1957. Biology of the treponematoses. World Health Organization, Geneva, Switzerland.
26. **Weinstock, G. M., J. M. Hardham, M. P. McLeod, E. Sodergren, and S. J. Norris.** 1998. The genome of Treponema pallidum: new light on the agent of syphilis. FEMS Microbiol. Rev. **22:**323–332.
27. **Weinstock, G. M., S. J. Norris, E. Sodergren, and D. Šmajs.** 2000. Identification of virulence genes in silico: infectious disease genomics, p. 251–261. *In* K. A. Brogden, J. A. Roth, T. B. Stanton, C. A. Bolin, F. C. Minion, and M. J. Wannemuehler (ed.), Virulence mechanisms of bacterial pathogens, 3rd ed. ASM Press, Washington, DC.

Editor: A. Camilli