JB
Journals.ASM.org

Comparative Investigation of the Genomic Regions Involved in Antigenic Variation of the TprK Antigen among Treponemal Species, Subspecies, and Strains

Lorenzo Giacani,[a] Stephanie L. Brandt,[a]* Maritza Puray-Chavez,[a]* Tara Brinck Reid,[a] Charmie Godornes,[a] Barbara J. Molini,[a] Martin Benzler,[b] Jörg S. Hartig,[b] Sheila A. Lukehart,[a,c] and Arturo Centurion-Lara[a]

Departments of Medicine[a] and Global Health,[c] University of Washington, Seattle, Washington, USA, and Department of Chemistry and Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Konstanz, Germany[b]

Although the three *Treponema pallidum* subspecies (*T. pallidum* subsp. *pallidum*, *T. pallidum* subsp. *pertenue*, and *T. pallidum* subsp. *endemicum*), *Treponema paraluiscuniculi*, and the unclassified Fribourg-Blanc treponeme cause clinically distinct diseases, these pathogens are genetically and antigenically highly related and are able to cause persistent infection. Recent evidence suggests that the putative surface-exposed variable antigen TprK plays an important role in both treponemal immune evasion and persistence. *tprK* heterogeneity is generated by nonreciprocal gene conversion between the *tprK* expression site and donor sites. Although each of the above-mentioned species and subspecies has a functional *tprK* antigenic variation system, it is still unclear why the level of expression and the rate at which *tprK* diversifies during infection can differ significantly among isolates. To identify genomic differences that might affect the generation and expression of TprK variants among these pathogens, we performed comparative sequence analysis of the donor sites, as well as the *tprK* expression sites, among eight *T. pallidum* subsp. *pallidum* isolates (Nichols Gen, Nichols Sea, Chicago, Sea81-4, Dal-1, Street14, UW104, and UW126), three *T. pallidum* subsp. *pertenue* isolates (Gauthier, CDC2, and Samoa D), one *T. pallidum* subsp. *endemicum* isolate (Iraq B), the unclassified Fribourg-Blanc isolate, and the Cuniculi A strain of *T. paraluiscuniculi*. Synteny and sequence conservation, as well as deletions and insertions, were found in the regions harboring the donor sites. These data suggest that the *tprK* recombination system is harbored within dynamic genomic regions and that genomic differences might be an important key to explain discrepancies in generation and expression of *tprK* variants among these *Treponema* isolates.

The pathogenic, noncultivable treponemes include *Treponema pallidum* with its three subspecies (*T. pallidum* subsp. *pallidum*, *T. pallidum* subsp. *pertenue*, and *T. pallidum* subsp. *endemicum*), *Treponema carateum*, *Treponema paraluiscuniculi*, and the unclassified Fribourg-Blanc treponeme. *T. pallidum* subsp. *pallidum* and *T. paraluiscuniculi* are the etiologic agents of human and rabbit venereal syphilis, respectively, while *T. pallidum* subsp. *pertenue*, *T. pallidum* subsp. *endemicum*, and *T. carateum* cause the endemic treponematoses known as yaws, bejel, and pinta (10, 18). Although regarded as distinct syndromes, these treponematoses share many noteworthy similarities, particularly evident when the clinical manifestations of the diseases are compared. Generally acquired during sexual intercourse, human syphilis alternates episodes of active clinical disease with periods of asymptomatic, latent infection (61). *T. pallidum* multiplication at the inoculation site induces the appearance of a primary lesion (chancre) (37) that persists for several weeks and then spontaneously heals following treponemal immune clearance. Proliferation of disseminated *T. pallidum* cells may cause the systemic manifestations of secondary syphilis, such as mucocutaneous lesions and generalized lymphadenopathy (5), that also spontaneously resolve. Following a latency stage, syphilis progression into the tertiary stage may result in gummatous disease, cardiovascular syphilis, general paresis, or tabes dorsalis (27, 28, 54).

Like venereal syphilis, yaws, bejel, and pinta have relapsing courses and prominent cutaneous manifestations but are generally acquired during childhood via skin-to-skin and skin-to-mucous membrane contact (transmission routes that are also similar to those of *T. pallidum* subsp. *pallidum*) (2, 49, 50). It has been stated that central nervous system (CNS) and placental invasion do not occur in these infections, although this statement has been disputed (51). The Fribourg-Blanc treponeme was isolated from an African baboon with no signs of clinical infection. Little is known about the pathogenesis of the infection (20); however, experimental inoculations showed that this organism is able to produce an active infection in humans (58, 59). Rabbit venereal syphilis (caused by *T. paraluiscuniculi*) is a milder infection characterized by slightly elevated, scaly patches or eroded sores in the genitals, the anus, and, less frequently, the eyelids, lips, and paws of the animals (15). Development of a systemic secondary stage preceding latency (15, 57) in *T. paraluiscuniculi*-infected rabbits has never been properly investigated.

These pathogens are morphologically indistinguishable, as well as antigenically and genetically very similar. Antigenic profile studies showed only minimal differences in antibody reactivity to *T. pallidum* subsp. *pallidum*, *T. pallidum* subsp. *pertenue*, Fribourg-Blanc, and *T. paraluiscuniculi* lysates (4, 41, 47, 60). More

Received 17 May 2012 **Accepted** 29 May 2012
Published ahead of print 1 June 2012
Address correspondence to Lorenzo Giacani, giacal@u.washington.edu.
* Present address: Stephanie L. Brandt, Seattle Genetics, Bothell, Washington, USA; Maritza Puray-Chavez, Universidad Nacional Mayor de San Marcos, Lima, Peru.
Supplemental material for this article may be found at http://jb.asm.org/.

doi:10.1128/JB.00863-12

recently, several comparative genomics studies confirmed the high level of relatedness among these treponemal strains (22, 42, 43, 56).

Given the remarkable relatedness shared by these pathogens and their ability to escape a strong immune response and persist in the host, it is plausible that similar mechanisms are employed by these spirochetes to ensure their survival during infection. Antigenic variation likely plays an important role. In *T. pallidum* subsp. *pallidum*, we have demonstrated that the putative outer membrane protein (OMP) TprK undergoes extensive antigenic variation in seven variable (V) regions, called V1 to V7, separated by conserved amino acid sequences (12, 36). Although the location of TprK in the pathogen's OM has been questioned (13, 29), several lines of evidence support TprK as being surface exposed. In rabbits infected with the Nichols strain, TprK variable regions are targets of a strong humoral response (44, 45) and antibodies against TprK were shown to be opsonic and enhance treponemal phagocytosis by macrophages (11). These data indeed agree with current three-dimensional (3D) models that show TprK as a beta-barrel structure with the protein's V regions protruding toward the extracellular milieu (A. Centurion-Lara, unpublished data). Furthermore, immunization of rabbits with recombinant TprK significantly attenuated lesion development following infectious challenge, although it did not provide sterile immunity (11). The role of antigenic variation of TprK in immune evasion during syphilis infection is also supported by the evidence that immunization with a synthetic TprK V6 peptide induces positive selection of new V6 variants upon challenge with the parental isolate (25). Additionally, immunosuppression in the rabbit model results in a significantly lower proportion of V6 variants identified during infection, providing strong evidence that the immune response positively selects for treponemes expressing variant V regions (25).

Using the *T. pallidum* subsp. *pallidum* Chicago strain as a model, we have previously shown that generation of *tprK* variants occurs during experimental infection by nonreciprocal segmental gene conversion, namely, the unidirectional transfer of genetic information from donor sites (DSs), located approximately 130 kbp downstream of the single *tprK* expression site, into the *tprK* expression locus (12). In the Chicago strain genome, a total of 52 *tprK* DSs can be identified; of these, 50 are located in the 3′-flanking region of TP0131, and 2 are in the TP0131 5′-flanking region (12). A large proportion of DSs also show the unique 4-bp terminal repeat sequences defining individual *tprK* V regions. A variable number of DSs are available for each V region: 5, 6, 10, 4, 6, 13, and 8 DSs for V1, V2, V3, V4, V5, V6, and V7, respectively, in the Chicago strain (12).

LaFond et al. (36) showed that generation of *tprK* variants occurs at different rates in the Chicago versus the Nichols isolates of *T. pallidum* subsp. *pallidum*. *tprK* diversifies rapidly in the Chicago strain, while diversification is considerably slower in the Nichols strain of *T. pallidum* subsp. *pallidum*. The rates of diversification of *tprK* have not been yet experimentally compared with respect to other *T. pallidum* subsp. *pallidum* strains among yaws and bejel isolates or in the Fribourg-Blanc isolate, though sequence heterogeneity in the *tprK* gene in these isolates suggests a higher baseline diversification rate than in the Nichols isolate of *T. pallidum* subsp. *pallidum*.

Because of the apparent importance of TprK in treponemal pathogenesis, we have compared *tprK* donor and expression sites among these spirochetes. We found sequence diversity of DSs among isolates, as well as the presence of A/T-rich sequences of various lengths, resembling an upstream regulatory (UP) element, in the *tprK* promoter region. Furthermore, putative *cis*-acting DNA elements, able to form guanine quadruplexes (G4) and also exhibiting sequence variability among treponemal species and subspecies, were found in proximity to the *tprK* expression site and within DSs. Similar sequences were shown to be required for PilE antigenic variation by *Neisseria gonorrhoeae* (8, 9). Our study lays the foundation for further investigations that will elucidate how such differences among strains, subspecies, and species could affect the generation and expression of TprK variants during treponemal infections.

MATERIALS AND METHODS

Treponemal strain propagation and nucleic acid extraction. Eight *T. pallidum* subsp. *pallidum* strains (Nichols Sea, Nichols Gen, Sea81-4, Chicago, Dal-1, Street14, UW126, and UW104) were used in this study. With the exception of the Nichols Gen, Dal-1, and Street14 strains, whose genome sequences are already available (GenBank accession numbers AE000520.1, CP000805.1, and CP003115.1, respectively), all other strains were propagated in our laboratory for sequencing purposes. The Chicago strain genome was previously obtained in our laboratory (22) (GenBank accession number CP001752.1). The Seattle Nichols strain was supplied by James N. Miller (University of California, Los Angeles, CA) in 1979; the Sea 81-4 strain was isolated in Seattle from a primary lesion in 1981; the Chicago strain was initially supplied by Paul Hardy and Ellen Nell (Johns Hopkins University, Baltimore, MD), along with the Cuniculi A, Samoa D, Iraq B, and Fribourg-Blanc strains. Peter Perine (Centers for Disease Control and Prevention, Atlanta, GA) initially provided the Gauthier strain of *T. pallidum* subsp. *pertenue*. All *T. pallidum* subspecies, *T. paraluiscuniculi*, and the Fribourg-Blanc treponeme were propagated in New Zealand White rabbits by intratesticular inoculation as previously described (39). *T. carateum* is unable to multiply in rabbits, and no strains or samples of this pathogen are available. Animal care was provided in accordance with the procedures outlined in the Guide for the Care and Use of Laboratory Animals under protocols approved by the University of Washington Institutional Animal Care and Use Committee (IACUC). Before infection, each rabbit had been serologically tested to rule out a naturally occurring infection with *T. paraluiscuniculi*. Treponemes were extracted from infected rabbit testes at peak orchitis. Collected organisms were separated from host cellular gross debris by low-speed centrifugation (250 × *g* for 10 min at room temperature); the supernatants were spun in a microcentrifuge for 30 min at 12,000 × *g* at 4°C. The pellet was suspended in 1 ml of 1× lysis buffer (10 mM Tris, pH 8.0, 0.1 M EDTA, 0.5% sodium dodecyl sulfate) for DNA extraction or in 400 μl of Ultraspec buffer (Biotecx Laboratories, Inc., Houston, TX) for RNA isolation. DNA extraction was performed as previously described with the QIAamp DNA minikit (Qiagen Inc., Chatsworth, CA) (21), taking careful precautions to prevent cross-contamination between samples. RNA extraction was performed according to the manufacturer's instructions followed by treatment with TURBO-DNase (Ambion, Austin, TX). Absence of residual DNA was assessed by PCR as previously described (21). DNase-treated RNA was stored at −80°C until use.

PCR amplification, cloning, sequencing, and sequence analysis of *tprK* open reading frames (ORFs) and donor regions. The *T. pallidum* subsp. *pallidum* strain Nichols genome sequence (19) was used to design primers (Table 1) targeting the 5′ and 3′-flanking regions of the TP0131 gene, where the *tprK* donor sites are located. PCR amplification for the TP0131 5′-flanking region using genomic DNA as a template was performed in a 50-μl final volume containing 200 μM deoxynucleoside triphosphates (dNTPs), 1.5 mM $MgCl_2$, and 2.5 U of GoTaq DNA polymerase (Promega, Madison, WI). The cycling conditions were as follows: initial denaturation at 95°C for 5 min, followed by 95°C for 1 min, 60°C

TABLE 1 Primers used in this study

Primer	Sequence	Use[a]	Target
5DF-S	5′-CAGTCCAAGAAGCGGAAAAG	1, 2	TP0131 5′-flanking region
5DF-As	5′-AAACGCCAGCAGTTCCAGCG	1, 2	TP0131 5′-flanking region
TP0136Ins-S	5′-ACACAACGGCTGCGAATACT	1	TP0131 5′-flanking region
TP0136Ins-As	5′-ACCGACCGTGCCCCATACT	1	TP0131 5′-flanking region
TPrflp-S	5′-TTTTGTGTGTGGGGAGGAGT	1, 2	TP0131 5′-flanking region
TPrflpAs	5′-ATGGCTGGGGGTAGGGCTTGC	1, 2	TP0131 5′-flanking region
3DF-S	5′-ATTGCCGAGAGCATCTGGT	1, 2	TP0131 3′-flanking region
3DF-As	5′-CCGCTCTCCTTCCCATAAGA	1, 2	TP0131 3′-flanking region
3DF-S1	5′-GCCTCATACCGACTGGGAG	1, 2	TP0131 3′-flanking region
3DF-S2	5′-CCTGCTCCTGGTACGAAGT	1, 2	TP0131 3′-flanking region
3DF-S3	5′-CAGCAGGAACTCCCAGCACGT	2	TP0131 3′-flanking region
3DF-As1	5′-GGGCGGACGTCCCACACC	2	TP0131 3′-flanking region
3′DF-1-2-AS	5′-ATTGCGTCTTTTGCGCGCG	2	TP0131 3′-flanking region
3′DF-1-3-AS	5′-AAACGCGGCGCTGTGGTTT	2	TP0131 3′-flanking region
3′DF-1-4-AS	5′-GGGAGTTGGGCTCGGGC	2	TP0131 3′-flanking region
3′DF-1-5-AS	5′-CTCGGGCTTGGGTGAAGG	2	TP0131 3′-flanking region
3′DF-1-6-AS	5′-ACACCCTGGGTGAAAAGACT	2	TP0131 3′-flanking region
3′DF-1-7-AS	5′-GTACGTGCTGGGAGTTCCT	2	TP0131 3′-flanking region
3′DF-1-8-AS	5′-TGTGCGGTCATGGCCATGT	2	TP0131 3′-flanking region
3′DF-1-9-S	5′-CGCAGGCCAGCAATGTATTT	2	TP0131 3′-flanking region
3′DF-1-10-AS	5′-TTTGGTTTGCCCTCCCAG	2	TP0131 3′-flanking region
3′DF-1-11-S	5′-CAAGCAGGGAGGACGCGT	2	TP0131 3′-flanking region
3′DF-1-12-AS	5′-CTTCGGGAATCGAGTGTAGT	2	TP0131 3′-flanking region
3′DF-1-13-S	5′-TCGTACTGCGCGTCATCGA	2	TP0131 3′-flanking region
3′DF-1-14-S	5′-TGGACTATCCATTCCCCATC	2	TP0131 3′-flanking region
3′DF-1-15-AS	5′-GTTGGGCTGGGGGAAATAG	2	TP0131 3′-flanking region
3′DF-1-16-S	5′-TATCGCCCCAACGCACGC	2	TP0131 3′-flanking region
3′DF-1-17-S	5′-GACGGGGAGGGCAAGTCA	2	TP0131 3′-flanking region
3′DF-1-18-S	5′-TACATGCCCGTCCATTGGAA	2	TP0131 3′-flanking region
3′DF-1-19-AS	5′-CATGGCATTGGTGAGAAAGA	2	TP0131 3′-flanking region
3′DF-1-20-S	5′-CAAAGACAGTCAGGGCAAGG	2	TP0131 3′-flanking region
3′DF-1-21-AS	5′-TGTCCTGTTCTTGTGCTCCA	2	TP0131 3′-flanking region
3′DF-1-22-S	5′-TCTACTTCTCGGCTCGCAGT	2	TP0131 3′-flanking region
3′DF-1-23-AS	5′-TATTGAGTGTGCGCTTACGG	2	TP0131 3′-flanking region
3′DF-2-S	5′-ACAACGAAGTTCCATAGGAC	2	TP0131 3′-flanking region
3′DF-2-AS	5′-TTTAAAGGGTTCGTTCTCGC	2	TP0131 3′-flanking region
3′DF-2-AS-1	5′-ATCCTCTTACGCGTCTCGTC	2	TP0131 3′-flanking region
3′DF-2-AS-2	5′-CACTCCCGCGCATATTCCC	2	TP0131 3′-flanking region
3′DF-2-AS-3	5′-GCCGAAAGGCGTATAGTACT	2	TP0131 3′-flanking region
3′DF-2-AS-4	5′-TGCAAGGTAATAGTACGGAGT	2	TP0131 3′-flanking region
3′DF-2-AS-5	5′-TTGGATGGCGCATCGATTAC	2	TP0131 3′-flanking region
3′DF-2-AS-6	5′-GCACGGCGCTGTTCAGTAA	2	TP0131 3′-flanking region
KI-S	5′-GTGGTGTATCAGCGGGTAGG	1, 2	*tprK* ORF
KI-As	5′-GACATGCCCCTACGAATTG	1, 2	*tprK* ORF
K-RFLP-S	5′-TGGCGTGCAGGAATACATTA	2	*tprK* V3–V7
K-RFLP-As	5′-TACCCCACACTCGTAATACCC	2	*tprK* V3–V7
K-RT-PCR-S	5′-AGTTTGCGTCTAACACCGACTG	2	*tprK* V2–V5
K-RT-PCR-As	5′-TCGCATGGCCATGTTGAGAAAT	2	*tprK* V2–V5
9V-As	5′-CCTCAAGGAAAGAAGTATCAGG	2	*tprK* V1–V5
K-RACE-1	5′-AGCACACAGACCCCAAAGCTTC	3	*tprK* V3–V4
K-RACE-2	5′-TAATGTATTCCTGCACGCCATC	4	*tprK* V4–V3
K-RACE-3	5′-CAAACTTAAACGCAATATCAAA	4	*tprK* V2–V1

[a] 1, amplification (sequence analysis); 2, sequencing; 3, reverse transcription (5′-RACE protocol); 4, amplification (5′-RACE protocol).

for 1 min, and 72°C for 2 min for 45 cycles in total, with a final extension at 72°C for 10 min. The TP0131 3′-flanking region was amplified using a TaKaRa Bio Inc. (Shiga, Japan) LA PCR kit, preferred for targets of >3 kbp. Amplification was performed in a 50-μl final volume containing 200 μM deoxynucleoside triphosphates, 1.5 mM $MgCl_2$, and 2.5 U of *Taq* polymerase. The cycling conditions were as follows: initial denaturation at 95°C for 10 min, followed by 95°C for 1 min, 60°C for 1 min, and 72°C for 3 min for 45 cycles in total, with a final extension at 72°C for 10 min.

Short amplicons from the TP0131 5′-flanking region and long amplicons from the TP0131 3′-flanking region were subsequently cloned into the pCR2.1-TOPO and pCR-XL-TOPO cloning vectors (Invitrogen, Carlsbad, CA), respectively, according to the manufacturer's instructions. Plasmid DNA from colonies containing inserts was extracted with the QIAprep spin miniprep kit (Qiagen), and at least two clones, each one from a different amplification reaction, were sequenced in both directions using a primer-walking approach (primers are listed in Table 1). Sequenc-

ing reactions and analyses were performed by the University of Washington Sequencing Facility. Sequences were assembled using the BioEdit sequence alignment editor (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and then aligned using the MAFFT alignment program (http://mafft.cbrc.jp/alignment/software/). Similarly, the *tprK* gene was amplified from all strains described above using *tprK* flanking primers and was sequenced using appropriate internal primers (Table 1). Amplification conditions for the *tprK* gene were previously described (26).

Prediction of G-quadruplex-forming sequence (G4FS) location in *T. pallidum* subsp. *pallidum* (Nichols Gen strain) genome and *Borrelia burgdorferi* (B31 strain) lp28-1. Composition and distribution of putative quadruplex-forming G-rich sequences in the full-length *T. pallidum* subsp. *pallidum* (Nichols strain) genome (GenBank accession number AE000520.1) and full-length *Borrelia burgdorferi* (B31) lp28-1 plasmid (GenBank accession number AE000794.2) were identified using QGRS (Quadruplex-forming G-Rich Sequences) Mapper (http://miracle.igib.res.in/quadfinder/gquadruplex.html) and QuadFinder (http://bioinformatics.ramapo.edu/QGRS/analyze.php). Parameters for QGRS Mapper were set so that the maximum length of the putative G-quadruplex was 30 nucleotides (nt) with a minimum of 2 G groups. Loop size was set from 0 to a maximum of 15 nt. Parameters for QuadFinder were $3 \leq$ G-Stretch ≤ 5 and $1 \leq$ N-Stretch ≤ 7. Hypothetical G4FS sequences identified by both programs and with G-scores (provided only by QGRS Mapper) equal to or greater than 29 were analyzed in this study. The scoring method uses previously published principles (3, 33, 34), in which G4FS is assigned a score based on a system which rewards an arrangement of G's that is likely to form a unimolecular quadruplex. For calculation of G-scores, the size and distribution of the gaps in the predicted QGFS are taken into account. Genomic coordinates of the putative G4FS, provided by both QGRS Mapper and QuadFinder, were used to determine G4FS location within ORFs or intergenic regions using the annotation of the *T. pallidum* subsp. *pallidum* strain Nichols genome or *B. burgdorferi* B31 plasmid lp28-1 available in GenBank.

Circular dichroism. DNA oligonucleotides for circular dichroism (CD) were obtained from MWG Biotech (Ebersberg, Germany). CD spectra were obtained as described by Joachimi et al. (32). Briefly, oligonucleotides were resuspended at 5 μM final concentration in diethylpyrocarbonate-treated water buffered with 50 mM Tris-HCl (pH 7.5) and optional addition of 100 mM KCl. Oligonucleotides were denatured by heating to 95°C for 5 min, and renaturation was allowed by slow cooling to a final temperature of 20°C over a period of 16 h. CD spectra were recorded on a Jasco 715 spectrometer in 1-cm-path cuvettes, with 0.5-nm resolution, 1.0-nm bandwidth, and speed of 20 nm/min at 20°C. Each spectrum was accumulated three times and averaged. Buffer alone was used as a negative control.

5′-RACE of *tprK* gene. The 5′ rapid amplification of cDNA ends (5′-RACE) system (Invitrogen) was used to determine transitional start sites (TSSs) in the 5′-flanking regions of *tprK*. 5′-RACE analysis was performed on *T. pallidum* subsp. *pallidum* (Nichols Sea strain) and *T. paraluiscuniculi* (Cuniculi A strain) total RNA following the manufacturer's instructions, except that the SuperScript system for cDNA synthesis was replaced by the ThermoScript reverse transcription (RT) kit (Invitrogen) to increase yield due to the high GC content (52.8%) of treponemal transcripts. A 2.5-pmol sample of gene-specific primer (K-RACE-1; Table 1) and 2 μg of sample RNA were used in each reaction. After the reaction was terminated, 1 μl of RNase H was added to the tube and the cDNA was incubated for 20 min at 37°C. All PCR amplification reactions were performed using 50 μl of dC-tailed cDNA in a 50-μl final volume containing 200 μM each dNTP, 20 mM Tris-HCl (pH 8.4), 1.5 mM $MgCl_2$, 50 mM KCl, 400 nM each primer, and 2.5 U of GoTaq DNA polymerase (Promega). For nested PCR, 1 μl of the original amplicon (for a 1/500 dilution) was used for the second amplification. Cycling parameters were as follows: initial denaturation for 2 min at 94°C, followed by 1 min at 94°C, annealing for 1 min at 63°C, and extension for 1 min at 72°C, for a total of 45 cycles. The final extension was 10 min at 72°C. PCR products were separated in 2% agarose gels, gel purified, cloned, and sequenced as described above.

TABLE 2 GenBank accession numbers

Species and strain	Sequence	Accession no.
T. pallidum subsp. *pallidum*		
Nichols Gen	TP0131 5′-flanking region	AE000520.1
	TP0131 3′-flanking region	
Nichols Sea	TP0131 5′-flanking region	JX025054
	TP0131 3′-flanking region	JX025071
Chicago	TP0131 5′-flanking region	CP001752.1
	TP0131 3′-flanking region	
Sea81-4	TP0131 5′-flanking region	JX025057
	TP0131 3′-flanking region	JX025068
UW104	TP0131 5′-flanking region	JX025056
	TP0131 3′-flanking region	JX025069
UW126	TP0131 5′-flanking region	JX025055
	TP0131 3′-flanking region	JX025070
Street14	TP0131 5′-flanking region	CP000805.1
	TP0131 3′-flanking region	
Dal-1	TP0131 5′-flanking region	CP003115.1
	TP0131 3′-flanking region	
T. pallidum subsp. *pertenue*		
Gauthier	TP0131 5′-flanking region	JX025058
	TP0131 3′-flanking region	JX025064
CDC2	TP0131 5′-flanking region	CP002375.1
	TP0131 3′-flanking region	
Samoa D	TP0131 5′-flanking region	JX025060
	TP0131 3′-flanking region	JX025066
T. pallidum subsp. *endemicum*		
Iraq B	TP0131 5′-flanking region	JX025059
	TP0131 3′-flanking region	JX025065
Unclassified		
Fribourg-Blanc	TP0131 5′-flanking region	JX025061
	TP0131 3′-flanking region	JX025067
T. paraluiscuniculi		
Cuniculi A	TP0131 5′-flanking region	JX025062
	TP0131 3′-flanking region	JX025063

Nucleotide sequence accession numbers. All new sequences obtained in this study were deposited in GenBank. Accession numbers are reported in Table 2.

RESULTS

***tprK* donor sequences in the TP0131 (*tprD*) 3′-flanking region contain similarities and unique elements among treponemal species, subspecies, and strains.** All species, subspecies, and strains of pathogenic treponemes analyzed in this study (Table 3) show similar architecture in the TP0131 3′-flanking region, which contains the larger group of DSs, contained within the ORFs annotated in the published Nichols genome (Nichols Gen) as TP0126 to TP0130 (with the exception of TP0127, which contains no DSs), as well as within the intergenic regions that separate these ORFs (schematically represented in Fig. 1; chromosomal location 148374 to 151104 in the published Nichols genome [19]). According to our analysis, the genomes of the Fribourg-Blanc isolate and the human treponemes, including the Nichols strain of *T. pallidum* subsp. *pallidum* propagated in our laboratory (identified as Nichols Sea), harbor 3′ donor regions of comparable sizes, rang-

TABLE 3 Isolates sequenced in this study and corresponding sizes of DS regions[a]

Species and strain	Length (bp) of genomic region hosting *tprK* DSs	
	TP0131(*tprD*) 3′-flanking region	TP0131(*tprD*) 5′-flanking region
T. pallidum subsp. *pallidum*		
Nichols Gen[b]	2,730	239
Nichols Sea[b]	3,931	239
Chicago	3,931	239
Sea81-4	3,938	239
UW104	3,988	239
UW126	3,988	239
Dal-1[d]	3,931	239
Street14[d]	3,988	239
T. pallidum subsp. *pertenue*		
Gauthier	3,977[c]	239
Samoa D	3,976	239
CDC2[d]	3,977[c]	239
T. pallidum subsp. *endemicum*		
Iraq B	3,975	239
Unclassified		
Fribourg-Blanc	3,976	239
T. paraluiscuniculi		
Cuniculi A	2,490	400

[a] These regions also include TP0127 (687 bp), although this ORF does not contain DSs (Fig. 1).
[b] Nichols Gen refers to the published Nichols strain sequence (19); Nichols Sea is the Nichols strain currently propagated in our laboratory.
[c] This 1-nt difference between Gauthier/CDC2 and Samoa D 3′-flanking regions is due to a indel within a varying homopolymeric G tract.
[d] Sequences from Dal-1, Street14, and CDC2s were obtained from GenBank (accession numbers CP000805.1, CP003115.1, and CP002375.1, respectively).

ing from 3,931 to 3,988 bp (Table 3). A deletion of 1,204 bp was identified in the TP0126-TP0127 intergenic spacer in the Nichols strain originally used for the first *T. pallidum* subsp. *pallidum* genome project (Nichols Gen) (19) (Fig. 1 and Table 3). At the beginning and at the end of this deletion, there are two direct repeats of 24 bp (AATGTATTTCAGGGTGTCTTTCTC), suggesting a loop-out mechanism for this deletion. The deletion decreases the overall number of DSs in the Nichols Gen isolate to 34, compared to the 54 DSs described in the Nichols strain propagated in our laboratory (Nichols Sea) (Table 4). Similarly, in the *T. paraluiscuniculi* Cuniculi A strain genome, a 1,441-bp deletion of portions of TP0127 and TP0128 reduces these DS-containing regions to only 2,490 bp (Table 3 and Fig. 1), eliminating 20 DSs (Table 4) that in *T. pallidum* strains overlap the TP0128 sequence and the TP0128-TP0127 intergenic spacer (Fig. 1) (12).

Among the syphilis-causing *pallidum* strains, there is sequence conservation in the TP0130, TP0129, TP0128, and TP0127 putative ORFs, but single nucleotide polymorphisms (SNPs) and indels occur in the TP0128-TP0127 and the TP0127-TP0126 intergenic spacers (Fig. 1). In the UW126 and UW104 strains, we identified an insertion of 51 bp located in the TP0127-TP0126 intergenic spacer modifying an already existing donor site for *tprK* V7; this insertion codes for the peptide sequence GAVPGAVGAV PGAAA (40). TP0130, TP0129, TP0128, and TP0127 putative ORFs are also conserved in non-syphilis-causing strains (*T. pallidum* subsp. *pertenue*, *T. pallidum* subsp. *endemicum*, and the Fribourg-Blanc isolate).

The most significant change in the sequences of the non-syphilis-causing strains (Gauthier, Iraq B, Samoa D, CDC2, and the Fribourg-Blanc treponeme) is a deletion of 30 bp within the TP0128-TP0127 intergenic region that does not affect any DS identified to date and an insertion of 90 bp (Fig. 1) within the TP0127-TP0126 spacer containing one new DS of 33 nt for V6 (CAAGCCCTACCCCCAGCCATTTACTTCCCCCTG) and a small donor site for V7 (CTGCTCCTG). Two more insertions, although of limited length and not containing any recognizable DS, are located in the TP0130-TP0129 intergenic region (Fig. 1)

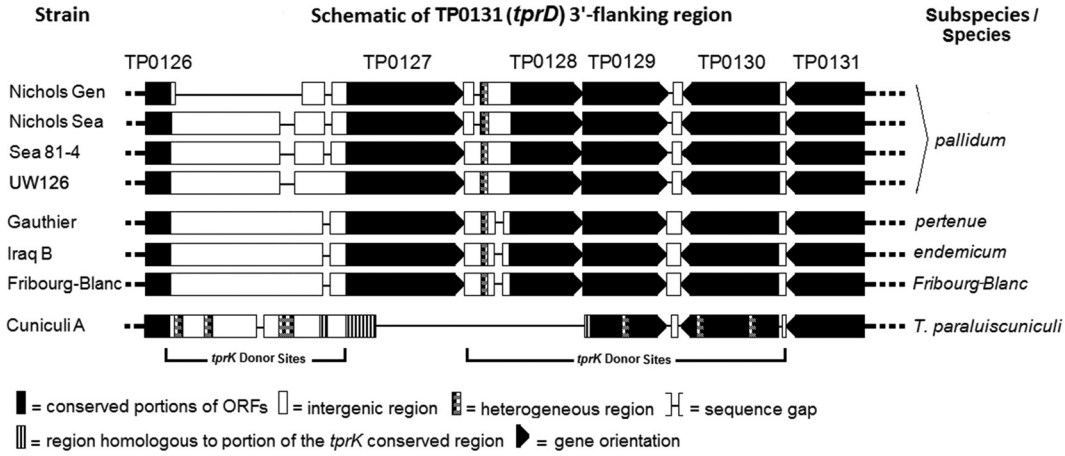


FIG 1 Schematic representation of the TP0131(*tprD*) 3′-flanking genomic region containing *tprK* DSs of the following treponemal isolates: *T. pallidum* subsp. *pallidum* (Nichols Gen, Nichols Sea, Seattle 81-4, and UW126 strains), *T. pallidum* subsp. *pertenue* (Gauthier, CDC2, and Samoa D strains), *T. pallidum* subsp. *endemicum* (Iraq B strain), the Fribourg-Blanc (simian) isolate, and *T. paraluiscuniculi* (Cuniculi A strain). The TP0131 3′-flanking regions are identical between *T. pallidum* subsp. *pallidum* Nichols Sea, Chicago, and Dal-1 strains, between UW126, UW104, and Street14 strains, and between *T. pallidum* subsp. *pertenue* Samoa D, Gauthier, and CDC2 strains. Lengths of the ORFs and intergenic spacers are not in proportion.

TABLE 4 Total number of donor sites and distribution by V region

Species and strain	Total no. of DSs		Total no. of DSs per *tprK* V region						
	TP0131 (*tprD*) 3′-flanking region	TP0131 (*tprD*) 5′-flanking region	V1	V2	V3	V4	V5	V6	V7
T. pallidum subsp. *pallidum*									
Nichols Gen	32	2	3	3	6	3	4	9	6
Nichols Sea	50	4	5	6	11	4	6	13	9
Chicago	50	2	5	6	10	4	6	13	8
Sea81-4	50	2	5	6	10	4	6	13	8
UW104	50	2	5	6	10	4	6	13	8
UW126	50	2	5	6	10	4	6	13	8
Street14[a]	50	2	5	6	10	4	6	13	8
Dal-1[a]	50	4	5	6	11	4	6	13	9
T. pallidum subsp. *pertenue*									
Gauthier	52	2	5	6	10	4	6	14	9
Samoa D	52	2	5	6	10	4	6	14	9
CDC2[a]	52	2	5	6	10	4	6	14	9
T. pallidum subsp. *endemicum*									
Iraq B	52	2	5	6	10	4	6	14	9
Unclassified									
Fribourg-Blanc	52	2	5	6	10	4	6	14	9
T. paraluiscuniculi									
Cuniculi A	31	5	3	4	8	5	6	5	5

[a] Sequences from these strains were obtained from GenBank (accession numbers CP000805.1, CP003115.1, and CP002375.1 for Dal-1, Street14, and CDC2, respectively).

and in the TP0128-TP0127 intergenic region (Fig. 1) of the non-syphilis-causing strains.

In *T. paraluiscuniculi* (Cuniculi A strain), the 3′ donor region contains a deletion of 1,441 bp encompassing 96 bp of the 5′ end of the TP0129 open reading frame (ORF), the entire TP0128 ORF, and a large portion of TP0127 (Fig. 1). Interestingly, the 3′ donor region of *T. paraluiscuniculi* Cuniculi A contains 356 bp that correspond to the V3-V5 regions and intervening conserved regions present in the *tprK* expression site of this strain. There is no equivalent to this region in the donor sites of any of the other treponemal strains examined to date. The longest DNA region in the human and rabbit treponemes and in the Fribourg-Blanc isolate that contains a sequence corresponding to the *tprK* ORF conserved portion is only 148 bases that belong to the 3′ end of the gene located immediately downstream of one DS for V7 (44 bp long). The V7 donor site and the conserved *tprK* region are arranged in the same order as they occur in the *tprK* expression site. A full-length sequence alignment for the TP0131 3′-flanking regions of all *T. pallidum* subspecies, *T. paraluiscuniculi*, and the Fribourg-Blanc isolate is provided in Fig. S1 in the supplemental material. The sequences of all DSs identified in this study are reported in Table 5.

***tprK* donor sequences in the TP0131 5′-flanking region are highly conserved among treponemal species and subspecies.** Sequence comparison of the DSs localized in the TP0131 5′-flanking region (chromosomal location 158280 to 158518 in the Nichols Gen strain genome) shows nearly complete sequence identity among all syphilis- and non-syphilis-causing strains, as well as the Fribourg-Blanc treponeme. In contrast, this genomic region is unique in *T. paraluiscuniculi* in that it contains an insertion of 177 bp (Fig. 2). This insertion has homology with DNA sequences in the TP0131 3′ donor region and contains one duplicate donor cassette for V3 and one for V5 (Fig. 2, in green and pink, respectively). Except for the presence of 22 SNPs in the Cuniculi A strain, and 4 SNPs in the *T. pallidum* subsp. *pertenue* strains Gauthier and Samoa D (Fig. 2), as well as in the Fribourg-Blanc treponeme, the DNA sequences flanking the Cuniculi A insertion are identical in all strains. These flanking sequences carry a DS for V6 and one for V7 (Fig. 2, dark and light orange, respectively) and are preceded by the ORF homologous to TP0136 (19), reported to be a surface-exposed, fibronectin-binding protein (7). Additional donor sites were recently identified in Nichols Sea as a 125-bp insertion, containing one DS for V3 (green) and one for V7 (orange) (Fig. 3A). This insertion precisely reproduces a sequence contained in the Nichols Sea TP0131 3′ donor region (see Fig. S1 in the supplemental material, positions 2825 to 3000) that is absent in the Nichols Gen strain. Amplification of the corresponding genomic region in the other strains used in this work (performed with primers TP0136Ins-S and TP0136Ins-As; Table 1) initially suggested that the TP0136 insertion was unique to the Nichols Sea strain (Fig. 3B). Recently, however, Mikalová et al. reported a similar insertion in the *T. pallidum* subsp. *pallidum* Dal-1 genome (43). Additionally, this TP0136 insertion is found in the Nichols Farmington, Nichols Dallas, and Nichols UCLA strains while not found in the Nichols CDC, Street14, or Nichols Houston strain (T. B. Reid, unpublished data). Lack of amplification when Cuniculi A DNA is used (Fig. 3B) is due to the marked sequence diversity at the primer binding site between the rabbit treponeme and the Nichols Sea TP0136 ORFs (19).

***tprK* expression site.** DNA alignments of *tprK* sequences from

TABLE 5 Sequence of all donor sites for *tprK* variable regions in treponemal species and subspecies studied here

Region and site	Sequence	Strain(s)[d]
V1 DSs		
DS7	GGGCATTGCATCTGAAACTGGTGGCGCCGGAGCCCTCA	All strains
DS15	GGGGGCATTGCATCCGTGGTGGCGCCATCAAGCA	Nichols Gen, UW104, UW126, Street14, Sea81-4
	GGGGCATTGCATCCGTGGTGGCGCCATCAAGCA	Nichols Sea, Chicago, Dal-1
	GGGGGCATTGCATCCGATGGTGGCACCATCAAGCA	Cuniculi A, Gauthier, Iraq B, CDC2, Samoa D, Fribourg-Blanc
DS20	GGGGCTTGCATCTGAAAAAAATGGTGGCGCCCAACCCCTCAAGCA[c]	Nichols Sea, Chicago, Dal-1
	GGGGCATTGCATCTGAAAAAAATGGTGGCGCCCAACCCCTCAAGCA[c]	Gauthier, Iraq B, Samoa D, Fribourg-Blanc, Sea81-4, CDC2, UW104, UW126, Street14, Nichols Gen
DS29	CTGGGGCATTGCATATGAAAATGGTGGCGCCCAACCCCTCAAGCA[b,c]	All strains
DS37	GGGGCATTGCATCCGAAGATGGTAGCGCCGGAAACCTCAAGCATGGA[b]	All strains
	GGGCATTGCATCCGAAGATGGTAGCGCCGGAAACCTCAAGCATGGA[b]	Cuniculi A
V2 DSs		
DS8	ACCGACTGGGAGGGCAAAGACAGTCAGGGCAAGGCCCCAGCAGGAACTTCCAGCAC	Cuniculi A
	ACCGACTGGGAGGGCAAAGACAATCAGGGCAAGGCCCCAGCAGGAACTCCCAGCAC	Iraq B
	ACCGACTGGGAGGGCAAAGACAGTCAGGGCAAGGCCCCAGCAGGAACTCCCAGCACGTA	All others
DS16	ACTGGGAGGGCAAAGACAGCAAGGGCGTCGCCCAAGCAGGAGCAAACCACAGCA	Gauthier, CDC2
	ACTGGGAGGGCAAAGACAGCAAGGGCGTCGTCCAAGCAGGAGCAAACCACAGCA	All others
DS21	CTAACACCGACTGGGAGGGCAAACCAAACGGCAACGTCCCAGCAGGAGTAACCCCCAGCA[c]	All strains
DS30	ACACCGACTGGGAGGGCAAGTCAAACACGGGCGCCCCAGCAGCAGG[b,c]	All strains
DS38	GGGAGGGCAAGTCAAACACGGGCGCCGCCCGAGCAGGAAGAAACCACCGC[b]	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	GGGAGGGCAAGTCAAACACGGGCGCCGCCCGAGCAGGAAGAAACCACAGC[b]	All others
DS45	ACATCAACTTCCCGGTGTATGGGGGTGTCTTGCACGCATCGCAGGCT[b]	All strains
V3 DSs		
DS3.1[a]	TACAATTCCACGCTTTCTGGCGACTATGCCCTACCCCGAGCCCCAGCCCCAATCCTA	Cuniculi A
DS3	TAGCGGCTATGCCGCAGCCCCAGCCATCAACTTCCCGGT	Cuniculi A
	TAGCAGCTATGCCACAGCCCGAGCCGGAGCCGACATCAACTTCCCGGT	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	TAGCGGCTATGCCACAGCCCGAGCCGGAGCCGACATCAACTTCCCGGT	All others
DS10	CCCGAGCCCTACCCGCAGCCGCAGCCGCAGTCAACAACGACATCTTAT	Cuniculi A
	CCCGAGCCCTACCCGCAGCCGCAGCCGCAGCCGCAGTCAACAACGACATCTTAT	All others
DS12	CCCCAAGCCCAAGCCGCAGCCAACATCAACTTCCCGGTATGGA	Cuniculi A
	CCCCAAGCCCCAGCCGCAGCCAACATCAACTTCCCGGTATGGA	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	CCCCAAGCCCCAGCCGCAGCCAACATCAACTTCCCGGTATGGG	All others
DS18	GCCCGAGCCGCAGCCCCAGCCGGAGCCAACGACATCTCAT	Cuniculi A
	GCCCGAGCCCGAGCCGGAGCCGCAGTGCCAGCCGCAGCCGACGACATCTTAT	All others
DS24	AACTCCACACTGTCTAGCGGCTATGCCCAAGCAGCCGGAGCCGCAGCCCA[c]	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc, Sea81-4
	AACTCCACACTGTCTAGCGGCTATGCCCAAGCAGCCGGAGCCGCAGCCGGAGTCCCAG CCCTACCCGGAGCCGCAGCCCA[c]	All others
DS26	AGCCCCAGCCCCAGCCCCAGCCCCAGCCAACAACGCCATCTTATGGGA	Nichols Gen, Nichols Sea, Chicago, Dal-1
	AGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCAACAACGCCATCTTATGGGA	Sea81-4
	GCCCTGGCCCGAGCCCGACCCCCAGCCCCAGCCAACAACGCCATCTTATGGGA	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc, UW104, UW126
	AGCCCTGGCCGGAGCCCTACCCCGAGCCCCAGACATCTTATGGGA	Cuniculi A
DS32	CCCCAGCCCAACCCCCAGCCAACATCTTATGGG[b,c]	All strains
DS34	GCCCAAGCCCGAGCCCTGGCAGCCGG[b,c]	All strains
DS41	GCCCTACCCCCAGCCGCAGCCCCAGCCAATGACATCTTATGGG[b]	Cuniculi A

(Continued on following page)

TABLE 5 (Continued)

Region and site	Sequence	Strain(s)[d]
	GCCCAACCCCCAGCCGTAGCCCCAGCCAATGACATCTTATGGGA[b]	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	GCCCAACTCCCAGCCGTAGCCCCAGCCAATGACATCTTATGGGA[b]	All others
DS43	TCCACGCTTTCTGGCGACTATGCCCGACCCGGAGCCGCAGCCG[b]	Cuniculi A
	TCCACACTTTCTGGCGACTATGCCCGAGCCGCAGCCGCAGCCG[b]	Sea81-4
	TCCACGCTTTCTGGCGACTATGCCCGAGCCGCAGCCGCAGCCG[b]	All others
DS55[e]	TGCTGCTCCTACGAAGTGGAAGGCAGGATATTGTGGGTAT	Nichols Sea, Dal-1
V4 DSs		
DS6	TGGGGCATAAGAAAAACGGAGCGAATGGCGACATAGGCGCAGA	All strains
DS14	CGTGGGGCGTAAGAAAGACGGAGCGCAGGGCAATGGCGTAGGCGC	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	CGTGGGGCGTAAGAAAGACGGAGCGCAGGGCACCGTAGGCGC	All others
DS36	CAGACGTGGGGCATAAGAAAGAGAATGCAGCGAACGTCAATGGCACCGT[b]	All strains
DS36.1	CAGACGTGGGGCATAAGAGAAACGTCCGAGCCCAAGCCCAAGCCGCAGATGCGTT	Cuniculi A
DS47	CGTGGGGCATAAGAAAAACGCAGCTCCCGATGGCATAGGCGCCTCACGCGC[b]	All strains
V5 DSs		
DS3.2[a]	TTCCCGGTGTGGGACGTCCGCCCGCATCGAAGGCCAGCAATGTGTTTAAAGATGTCTTTCTC ACCGATACCACACCCATGCGGACGCACGACGGTGCAG	Cuniculi A
DS5	CGCATCGAAGGCCAGCAATGTGTTTATAGACGTCTTTCTCACCAATGCCATGGACATGCAG ACGCACGACTG	Cuniculi A
	CGCATCGGAGGCCAGCAATGTGTTTAAAGACGTCTTTCTCACCAATGCCATGGACATGCAG ACGCACGACTG	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	CGCATCGAAGGCCAGCAATGTGTTTAAAGACGTCTTTCTCACCAATGCCATGGACATGCAG ACGCACGACTG	All others
DS13	CGCAGGCCAGCAATGTATTTCAGGGAGTATTTCTCAACATGGCCATGACCGCACACGACTG	All strains
DS13.1	TATTTCGCATCGAAGGCCAGCAATGTGTTTAAAGATGTCTTTCTCGCTAGAAACATGGACAT GCAGACGCACGACTGTGCTACTTATATCAA	Cuniculi A
DS28	AGCAATGTATTTCAGGGTGTCTTTCTCACCACACCCATGCAGAAGGACGACTG[b]	All strains
DS35	TATTCGGGGGAGTATTTCTCACCAATAACATGCTGCAGCACGACTG[b]	All strains
DS46	AGCAATGTATTCGAGGGAGTATTTCTCACCACACCCATGCGGACGCACGACAT[b]	Cuniculi A
	AGTAATGTATTTCAGGGTGTCTTTCTCACCGATACCACACCCATGCGGACGCACGACAT[b]	All others
DS48	CGCATCGAAGGCCAGCAATGTATTTCAGGGTGTCTTTCTCGCTCGAAACATAGCCATGCG AGAGCACGACTG[b,c]	Samoa D
	CGCATCGAAGGCCAGCAATGTATTTCAGGGTGTCTTTCTCGCTAGAAACATAGCCATGCGA GAGCACGACTG[b,c]	Gauthier, Iraq B, CDC2, Fribourg-Blanc
	CGCATCGAAGGCCAGCAATGTATTCGAGGGAGTATTTCTCGCTAGAAACATAGCCATGCGA GAGCACGACTG[b,c]	All others
V6 DSs		
DS1[a]	AAGCCCTACCCCCAGCCATTTACTTCCC	All strains
DS4	CCCCAGCCATCAACTTCCCGGTGT	Cuniculi A
	AGCCCGAGCCGGAGCCGACATCAACTTCCCGGTGT	All others
DS11[f]	CGTACATGCCCGTCCACTACAAGGTCCTGACCGGACCCCAAGCCCAAGCCGCAG[c]	Cuniculi A
	CGTACATGCCCGTCCACTACACGGTCCTGACCGGACCCCAAGCCCCAGCCGCAG[c]	All strains
DS19[f]	CGTACATGCCTGTCCATTACAAAGTCCTAAAAGCCCACGCCCGAGCCCCAGCCGACATCCA CTTCCCGGTGT[c]	All strains
DS23[f]	GCCCGAGCCGTACCCCCAGCCCGAGTTGACATCTACTTCTCGG[c]	All strains
DS25[f]	CGTACGTACATGCCCGTCCATTGGAAAGCCCCAGCCCCAGCCCCAGCCCCAGCCAACAACGC CATCTTATGGGA	Nichols Gen, Nichols Sea, Chicago, Dal-1
	CGTACGTACATGCCCGTCCATTGGAAAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCC AACAACGCCATCTTATGGGA	Sea81-4
	CGTACGTACATGCCCGTCCATTGGAAAGCCCTGGCCCGAGCCCGACCCCCAGCCCCAGCC AACAACGCCATCTTATGGGA	Gauthier, Iraq B, Samoa D, CDC2, Street14, Fribourg-Blanc
	TGTACATGCCCGTCCATTGGAAAGCCCTGGCCGGAGCCCTACCCCGAGCC	Cuniculi A
DS33	TGCCCGTCCACGCCCAAGCCCGAGCCCTGGCAGCCGGAGTCCCACCCGGAGCCCCTG ATA[b,c]	All strains
DS40[f]	GAGAACGGCATGCCCGTCCATTGGAACGTCAGCGTACAGTCCCACGCCCGAGCCCTACCC CCAGCCGCAGCCCCAGCCAATGACATC[b]	Cuniculi A

(Continued on following page)

TABLE 5 (Continued)

Region and site	Sequence	Strain(s)[d]
	CATGCCCGTCTATTACTTCGCAGCCCGAGCCCAACCCCCAGCCGTAGCCCCAGCCAATGACATCT[b]	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	CATGCCCGTCTATTACTTCGCAGCCCGAGCCCAACTCCCAGCCGTAGCCCCAGCCAATGACATCT[b]	All others
DS42[f]	CGTACATGCCCGTCCATTGGAACGCCTTCACCCAAGCCCGAGCCCTGCCCGGAGCCCCAG TCCC AGCCATTTACTTCCCGGT[b]	Cuniculi A
	CGTACATGCCCGTCCATTGGAACGCCTTCACCCAAGCCCGAGCCCTGCCCGGAGCCCCAG TCCC AGCCATCTACTTCCCGGT[b]	All others
DS44[f]	GCCCGAGCCGCAGCCGCAGCCGGGGCTGGAGTCGACATCAACTTCCCGGTGTATGG[b,c]	All strains
DS50	AAAACCCGAGCCCTACCCGCAGCCGCAGCCGCAGCCGCAGTCAACAACGACATCT[c]	All strains
DS51	AACTCCACACTGTCTAGCGGCTATGCCCAAGCAGCCGGAGCCGCAGCCCA[c]	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
	AACTCCACACTGTCTAGCGGCTATGCCCAAGCAGCCGGAGCCGCAGCCGGAGTCCCAGCCC TACCCGGAGCCGCAGCCCA[c]	All others
DS52	CCCGAGCCCTACCCGCAGCCGCAGCCGCAGCCGCAGTCAACAACGACATCTTAT[c]	All strains
DS53	AAGCCCTACCCCCAGCCATTTACTTCCC	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
V7 DSs		
DS2[a]	CCCCCTGCTGCTCCTGGTACGAAGTGGAGCAAGGAATATTGTGGGTATTACG	Gauthier
	CCCCCTGCTGCTCCTGCTACGAAGTGGAGCAAGGAATATTGTGGGTATTACG	All others
DS2.1[a]	CCCGGTTGTAG	Cuniculi A
DS2.2[a]	GACGTGGAGCAAGGAATATTGTGGGTATTACGA	Cuniculi A
DS9	GTACGGCGGTACGAACAAAAAGGCCACGCCCCCTGCTGCTCCTGCTGTTCCTACGAAGTGGA AGGCAGAATATTG	Cuniculi A
	GTACGGCGGTACGAACAAAAAGGCCACGCCCCCTGCTGCTCCTGCTGCTCCTACGAAGTGGA AGGCAGAATATTG	All strains
DS17	GTACAGGCGGTACGAACAAAAAAGCTGCTGCTGCAGCCCCTGCTCCTGGTACGAAGTGGAGCA	Cuniculi A
	GTACGGCGGTACGAACAAAAAAGCTGCTGCTGCAGCCCCTGCTCCTGGTACGAAGTGGAGCA	All strains
DS22	GTACGGCGGTACGAACAAGCAAGCTGCTGCGGTCTGCTCTTACGAAG[c]	All strains
DS27	CCCCTGCTGCTGCAGTTCCTGGTACGAAGT[c]	All strains
DS31	GCGGTACGAACAAGAAAAACGATGCTGCTCCTGGTGCAGTTCCTGGGGCGGTCCCTGGT GCAG TTCCTGGTGCTGCTGCTCTTACGAAGTGGAAGGCAGGATATTGTGGGTAT[b,c]	UW126, Street14
	GCGGTACGAACAAGAAAAACGATGCTGCTCCTGGTGCGGTTCCTGGGGCGGTCCCTGGT GCAG TTCCTGGTGCTGCTGCTCTTACGAAGTGGAAGGCAGGATATTGTGGGTAT[b,c]	UW104
	GCGGTACGAACAAGAAAAACGATGCTGCTCCTACGAAGTGGAAGGCAGGATATTGTGG GTAT[c]	All others
DS39	TGCTCCTGCTCTTACGAAGTGGAGCAAAGGATATTGTGGGTATTACG[c]	Cuniculi A
	TGCTCCTGGTACGAAGTGGAGCAAGGGATATTGTGGGTATTACG[c]	Fribourg-Blanc
	TGCTCCTGGTACGAAGTGGAGCAAGGAATATTGTGGGTATTACG[c]	Gauthier, Iraq B, Samoa D, CDC2
	TGCTCCTGCTACGAAGTGGAAGGCAGAATATTGTGGGTATTACG[c]	All others
DS49	GTACGGCGGTACGAACAAGCAAGCTGCTGCGGT[c]	All strains
DS54	CCCCCTGCTGCTCCTGGTACGAAGTGGAGCAAGGAATATTGTGGGTATTACGA	Gauthier, Iraq B, Samoa D, CDC2, Fribourg-Blanc
DS56[e]	CCCCCAGCCCAACCCCCAGCCAA	Nichols Sea, Dal-1

[a] DSs located in the TP0131 (*tprD*) 5′-flanking region.
[b] DSs absent in the original *T. pallidum* subsp. *pallidum* Nichols strain genome sequence (19).
[c] DSs absent in *T. paraluiscuniculi* (Cuniculi A strain).
[d] Nichols Gen, published Nichols strain sequence (19); Nichols Sea, Nichols strain propagated in our laboratory.
[e] These DSs are located in the Tp0136 insertion.
[f] These DSs are also used to generate variability in V3.

the *T. pallidum pallidum*, *T. pallidum pertenue*, and *T. pallidum endemicum* subspecies, as well as the Fribourg-Blanc and Cuniculi A isolates, reveal that the *tprK* expression site of these strains is characterized by a high degree of sequence identity of the conserved regions, with similar location and sequence anatomy of V regions (12). In the *T. pallidum* subsp. *endemicum* isolates, the conserved region between V1 and V2 contains an extra CGC (Ala) codon (Fig. 4A). In *T. paraluiscuniculi* TprK, a unique sequence (SGDPYTHLLTGLNAGVEARV) previously identified between V5 and V6 (26) is identical to a conserved region of subfamily I

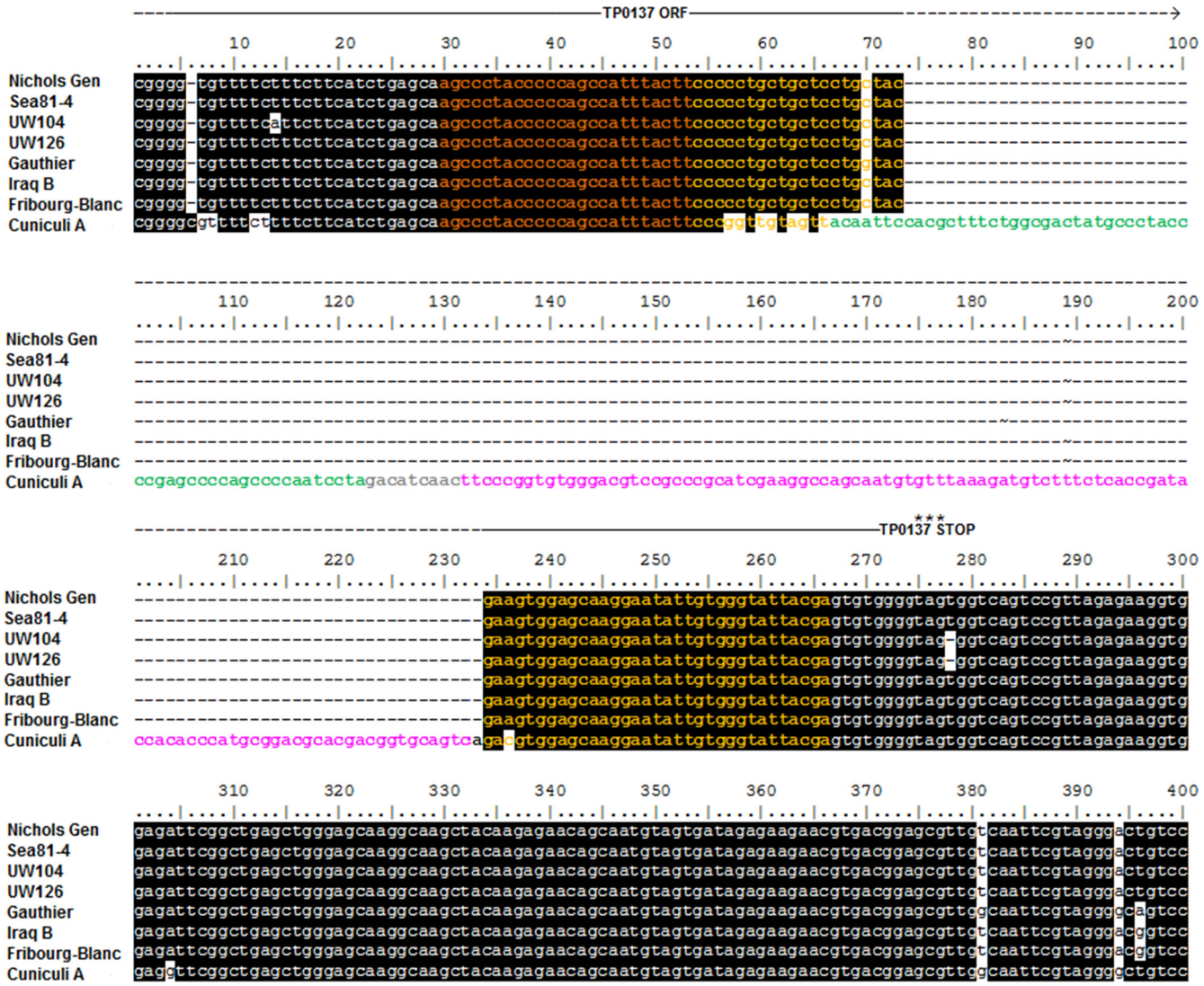


FIG 2 Alignment of *tprK* DSs located in the TP0131 (*tprD*) 5′-flanking region of treponemal strains. *T. paraluiscuniculi* (Cuniculi A strain) shows a newly identified insertion containing one donor site for V3 (green) and one donor site for V5 (pink). Three DSs, one for V6 and two for V7, are highlighted in dark and light orange, respectively. Shaded areas indicate regions of sequence identity. Nichols Sea and Dal-1 sequences are identical to the sequence shown for Nichols Gen. Street14 sequence is identical to the sequence shown for UW126. Samoa D sequence is identical to the sequence from the Iraq B strain. CDC2 sequence is identical to the sequence shown for Gauthier. The TP0137 ORF, as annotated in the Nichols Gen (19), is indicated above the alignment.

and II Tpr antigens (Fig. 4A, sequence substitution). Sequence variability, though limited, is also present in the *tprK* promoter region (Fig. 4B) when *T. paraluiscuniculi* and *T. pallidum* are compared. More specifically, diversity spans from an A/T-rich sequence located 22 nt upstream of the putative *tprK* −35 sigma70 (σ^{70}) signature, the −35 and −10 σ^{70} sequences, to the spacer between the putative −10 σ^{70} sequence and *tprK* start codon, the region which contains the experimentally identified transcriptional start site (TSS) (Fig. 4B). In prokaryotes, A/T-rich sequences located in close proximity to the −35 σ^{70} hexamer are also known as UP elements. These elements are sequences that increase transcription by interacting with the C-terminal domain of the RNA polymerase α-subunit (16, 17, 52, 53). Recently, intrastrain variability in the number of A's of this putative UP element was identified by T. B. Reid (unpublished) in the Chicago (7 to 9 A's) and Nichols (7 to 8 A's) strains of *T. pallidum* subsp. *pallidum* and in *T. paraluiscuniculi*, Cuniculi A strain (9 to 10 A's) (L. Giacani, unpublished data). Equally interesting is that a typical ribosomal binding site (RBS) (-GGAG-) is not identifiable in *T. paraluiscuniculi* as opposed to the remaining treponemal isolates (Fig. 4B). Despite such sequence differences, there is sequence homology of the putative −35 and −10 σ^{70} sequences from these isolates with the *E. coli* consensus sequences. The fact that TprK is still expressed in both these isolates despite sequence diversity in the promoter region is supported by the experimentally identified TSSs, the presence of *tprK* mRNA during infection (Giacani, unpublished), and the fact that antibodies and T-cell responses are generated against TprK during experimental infection with the Cuniculi A strain of *T. paraluiscuniculi* (26).

G4FS identification and circular dichroism spectra. Forty-five putative G4FS with a G-score of ≥29 (approximately 1% of the putative G4FS found in total by QuadFinder and QGRS

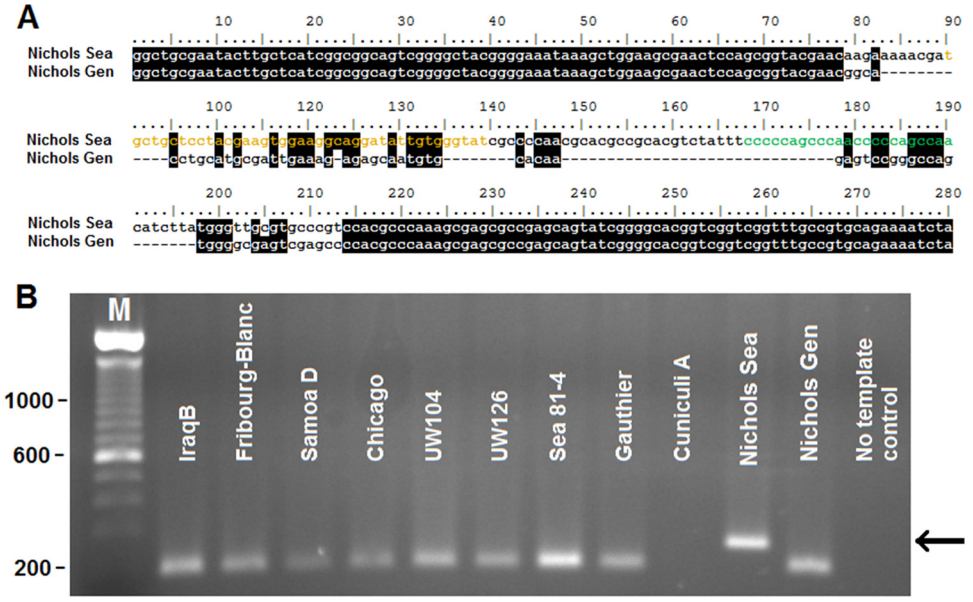


FIG 3 (A) In the Nichols Sea strain of *T. pallidum* subsp. *pallidum*, a 125-bp-long replacement sequence can be found in the TP0131 5′-flanking region, within TP0136, containing one DS for V3 (green) and one for V7 (orange). (B) PCR amplification of the genomic region encompassing the insertion shows an amplicon of higher size only when Nichols Sea genomic DNA is used as the template. Lane M, size markers (sizes are in bp). Arrow indicates amplicon position.

Finder) were identified in the *T. pallidum* subsp. *pallidum* strain Nichols genome (19) (Table 6 and Fig. 5A), almost evenly distributed across the genome, with the exception of the sequence between coordinates 550000 and 733330 (Fig. 5A), where no putative G4FS were detected. Five G4FS are clustered in the large DS-containing genomic region surrounding Tp0131. In more detail, two G4FS (with G-scores of 34 and 63) were identified within the TP0126-TP0127 intergenic region (Table 6 and Fig. 5B), while

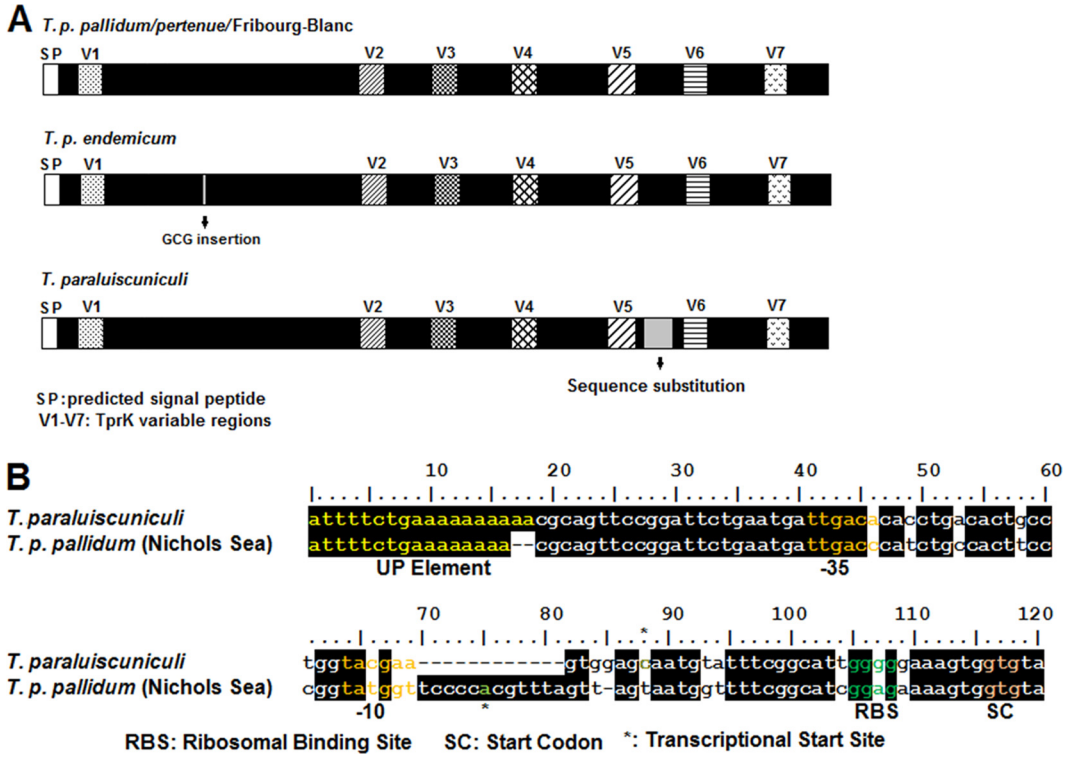


FIG 4 (A) Schematic representation of the *tprK* ORF in *T. pallidum* subspecies and *T. paraluiscuniculi* showing sequence conservation and heterogeneity within the conserved regions. (B) Sequence diversity within the *tprK* promoter region between *T. pallidum* isolates and *T. paraluiscuniculi*. RBS, ribosomal binding site (green); SC, *tprK* start codon (orange). Yellow, UP element; orange, −10 and −35 σ^{70} putative consensus sequences. *, experimentally determined transcriptional start sites for *tprK*.

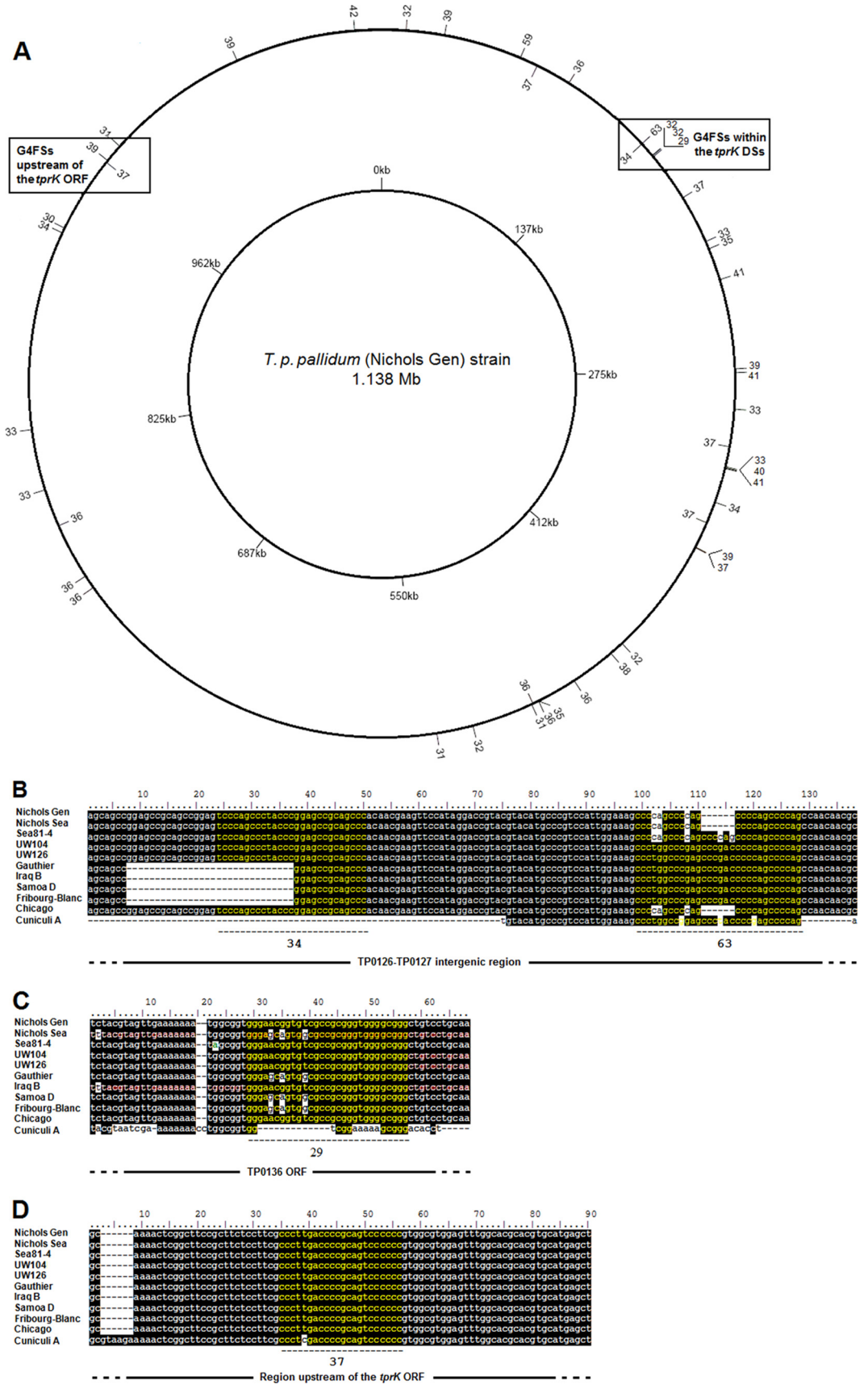
TABLE 6 G4FS identified in the Nichols strain genome of *T. pallidum* subsp. *pallidum* (19)

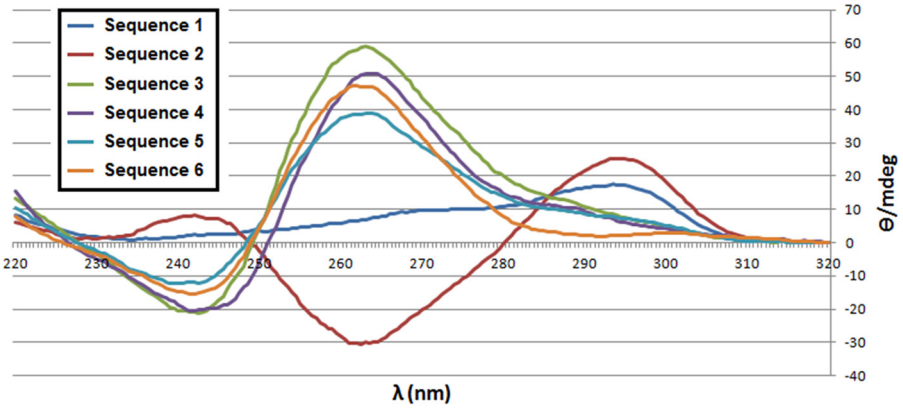
Position	Location	Length (bp)	Sequence[a]	G-score	Strand
12379[b]	TP0012	29	GGGAGCCTTCAATAAGGGGGAGGGAAGGG	32	+
32161	TP0025	17	GGGATAGGGGGGATGGG	39	+
72671	TP0067	25	GGGGGGGGGTATGGGGCTTATGGGG	59	+
82202	TP0074	30	GGGGAGTTGGGAACCCTTTTGGGAATTGGG	37	−
100635	TP0091	30	GGGCGGGGGTACCCCGGGTGGCACATCGGG	36	+
149328[c]	IGR104	22	GGGGCTGGGGCTGGGGCTGGGG	63	+
149398[c]	IGR104	26	GGGCTGCGGCTCCGGGTAGGGCTGGG	34	−
157297[c]	TP0136	29	GGGAACTGCGGTTGCGGGGGAGGGGGGGG	32	+
157393[c]	TP0136	29	GGGAACTGCGGTTGCGGGGGAGGGGGGGG	32	+
158101[c]	TP0136	29	GGGAACGGTGTCGCCGCGGGTGGGGCGGG	29	+
183960	TP0162	20	GGGTAAACGGGGGGAGTGGG	37	+
209260	TP0196	27	GGGAGCTGTAAGATGGGTCGGGGTGGG	33	+
213318	TP0207	27	GGGTCGTGGGTCCTCCTCTGGGCGGGG	35	+
230028	IGR175	20	GGGGGAGGGAAGGGTGTGGG	41	+
276466	TP0264	24	GGGTGCACAGGGTGAGGGAGGGGG	39	+
278453	IGR211	19	GGGAAGGGGGAGGGAAGGG	41	+
297175	TP0282	24	GGGGGGCTAGGGCTAAGTGCAGGG	33	+
316621	TP0303	29	GGGTAGAGGGAACAGGGATTCGTACCGGG	37	−
327048[d]	TP0312	25	GGGGGGAATAGTGGTGGGGATAGGG	33	+
327636[d]	TP0312	24	GGGAATGGGGATAGGGTTGCTGGG	40	+
327948[d]	IGR236	16	GGGCCGGGGGGGGGGG	41	+
346017	TP0326	27	GGGTGAACGGGGGGGTGGACTTTCGGG	34	+
357406	TP0334	27	GGGATTCGGGTGTGGGCATCTAACGGG	37	−
371115	TP0345	30	GGGAGGGCACCCCTTGCGGGGCACGCCGGG	33	+
371369	TP0346	27	GGGTATGGGCGCTTTGGGATATATGGG	39	+
433549	TP0408	26	GGGGGGCTGTGGGTCGGTCAAACGGG	32	+
441155	TP0414	25	GGGTGACGGGGATATTCGGGATGGG	38	+
464653	TP0438	19	GGGCGGGGGGTGTGCCGGG	36	+
485699	TP0456	30	GGGTGAGGGCGGTGCATTGGGCGTTCCGGG	36	+
485881	TP0456	20	GGGTGGGGGGAACTGTTGGG	35	+
488893	TP0458	26	GGGTGAACAGAGGGTGCAGGGGAGGG	36	−
489727	IGR335	24	GGGGGGAGGGTGACTGAAGGTGGG	31	+
521840	TP0488	26	GGGATTTTTCTGTAGGGGGGGTGGGG	32	+
540710	IGR364	24	GGGGGGAGGGTGTCGGAGGACGGG	31	+
742540	IGR483	30	GGGCGGGGTATGCGCAGGGTGCACAGCGGG	36	+
749636	TP0685	30	GGGAAAGGGTGCCGTGGTGGGGCTTATGGG	36	+
778932	TP0711	30	GGGCATCGGGTGCGGGGCAGGAGACTCGGG	36	−
798620	TP0733	23	GGGGCGGGGGGTTTCACCTCGGG	33	+
830276	IGR549	22	GGGCCTTTTTGAGGGCGGGGGG	33	+
933620	TP0856	30	GGGTACCGACGTGGGTCTGCAATGGGTGGG	34	+
936331	TP0859	27	GGGTATCGCCACACCCGGGGGGGGGGG	30	+
976267[e]	TP0898	22	GGGGGGACTGCGGGGTCAAGGG	37	−
977324[e]	TP0898	16	GGGGGGTGCGGGTGGG	39	+
986979	TP0905	25	GGGGGGCAGGGCGTCTAGGCAAGGG	31	+
1061766	TP0978	24	GGGATTATCGGGGGAGGGATAGGG	39	+
1123979[f]	TP1030	15	GGGCGGGGGGGGGGG	42	+

[a] Putative G residues involved in formation of the G quadruplex are underlined.
[b] Putative G4FS located downstream of *tprB*.
[c] Putative G4FS located within the *tprD* (TP0131) 3′-flanking region (containing *tprK* DSs). These sequences were tested for ability to form G4 using circular dichroism.
[d] Putative G4FS located upstream of *tprE*.
[e] Putative G4FS located upstream of *tprK* expression site. These sequences were tested for ability to form G4 using circular dichroism.
[f] Putative G4FS located upstream of *tprL*.

three G4FS were found 145 bp (with a G-score of 29) (Fig. 5C) and 855,953 bp (both with a G-score of 32) upstream of the TP0131 5′-flanking region with *tprK* DSs (Table 6). A putative G4FS (G-score of 37) was also found 493 bp upstream of the annotated *tprK* start codon (Table 6 and Fig. 5D). Different degrees of sequence length and variability were found in these hypothetical G4FS among treponemal species and subspecies, in particular for the sequences located within the TP0131 3′-flanking region (Fig. 5B and C). A total of 15 putative G4Fs with G-scores between 9 and 19 were found in the *B. burgdorferi* (B31) lp28-1 plasmid, 12 of which lie between coordinates 26997 and 28073 (see Table S1 in the supplemental material).

CD analysis of selected putative G4Fs in proximity to the *tprK* expression and donor sites (marked with an asterisk in Table 6 and reported in Fig. 6) showed that all but one of the analyzed sequences fold into well-defined quadruplexes. CD spectra are com-


A
G4FSs upstream of the tprK ORF
G4FSs within the tprK DSs
T. p. pallidum (Nichols Gen) strain
1.138 Mb
0kb
137kb
275kb
412kb
550kb
687kb
825kb
962kb
B
Nichols Gen
Nichols Sea
Sea81-4
UW104
UW126
Gauthier
Iraq B
Samoa D
Fribourg-Blanc
Chicago
Cuniculi A
34
63
TP0126-TP0127 intergenic region
C
29
TP0136 ORF
D
37
Region upstream of the tprK ORF



Sequence		Location
1	GGGGCTGGGGCTGGGGCTGGGG	149328 IGR104 (*tprK* DS region)
2	GGGCTGCGGCTCCGGGTAGGGCTGGG	149398 IGR104 (*tprK* DS region)
3	GGGAACTGCGGTTGCGGGGGAGGGGGGGG	157297 TP0136 (*tprK* DS region)
4	GGGAACGGTGTCGCCGCGGGTGGGGCGGG	158101 TP0136 (*tprK* DS region)
5	GGGGGGACTGCGGGGTCAAGGG	976267 TP0898 (upstream of *tprK*)
6	GGGGGGTGCGGGTGGG	977324 TP0898 (upstream of *tprK*)

FIG 6 Circular dichroism (CD) spectra of putative G4Fs in proximity to the *tprK* expression and donor sites. Most of the sequences fold into well-defined quadruplexes, with spectra compatible with G4 with all-parallel strand orientation for sequences 3 to 6. Sequence 1 failed to fold into a stable G4, despite its very elevated G-score. Sequence 2 folded into a structure that resembles a partly folded G4 with antiparallel strand orientation. Buffer alone was used as a negative control (data not plotted).

patible with G4 structures with all-parallel strand orientation for sequences 3 and 4 (located within the region harboring the *tprK* donor sites) and for sequences 5 and 6 (located in proximity to the *tprK* expression site). Interestingly, both sequence 5 and sequence 6 fold into G4s despite a missing nucleotide in the first loop, in theory required for a perfect three-stack G4 (32). Sequence 1 failed to fold into a G4, despite its high G-score. Sequence 2 generated spectra compatible with a partly folded G4 with antiparallel strand orientation (32). CD spectra are shown in Fig. 6.

DISCUSSION

The TprK antigen, a putative OMP and virulence factor, has been shown to undergo antigenic variation (12) by segmental gene conversion during experimental infection in rabbits (25). This antigenic variation system has two components: an expression site that harbors the *tprK* ORF and donor sites, located at both flanking sides of TP0131, that are copied into the seven *tprK* V regions. Given the importance of the TprK antigen during syphilis and, likely, during other treponemal infections, it is striking that *Treponema* isolates like the Chicago strain of *T. pallidum* subsp. *pallidum* can diversify the sequence of the *tprK* gene (before the appearance of TprK-specific antibodies) at a significantly high baseline rate (35, 36), while in other equally virulent syphilis strains, like Nichols, the *tprK* gene remains virtually clonal during passage at 10-day intervals and varies its sequence only after onset of an adaptive immune response against the initial TprK variant (36). Because no differences were found between the annotated recombinases in these isolates (data not shown), to investigate the reasons behind such dissimilar rates of variation, we comparatively analyzed the anatomy of the *tprK* expression site and DS regions among treponemal species and subspecies.

There is a very high degree of synteny and sequence conservation (99.57%) among the *T. pallidum* subspecies and the Fribourg-Blanc isolate in the genomic regions where the *tprK* DSs are located. This degree of sequence identity suggests a very close evolutionary relationship among these isolates. The ~1.2-kb deletion observed in the Nichols Gen strain is known today to be carried by a subpopulation of this strain of *T. pallidum* subsp. *pallidum* (56) and was likely caused by the loss of genetic material between two direct repeats (highlighted in Fig. S1 in the supplemental material), which may undergo excision due to slipped-strand mispair-

FIG 5 Location of the putative G4FS in the *T. pallidum* subsp. *pallidum* (Nichols Gen strain) genome (19). (A) G4FS are mapped using the G-score (ranging from 29 to 63, shown in the outer ring) in the *T. pallidum* subsp. *pallidum* Nichols Gen strain. Except for the genomic region encompassed approximately between nt 550000 and 733330, G4FS appear to be evenly distributed throughout the genome. G4FS in proximity to the *tprK* DSs and *tprK* gene are indicated. (B) Two of the putative G4FS (yellow) located in the TP0131 3′-flanking region (TP0128-TP0127 intergenic spacer) in *T. pallidum* subspecies and *T. paraluiscuniculi* showing sequence diversity among these strains. For each putative G4FS the G-score is indicated below the sequences. The Street14 sequence is identical to the UW126 strain sequence, the Dal-1 sequence is identical to the Chicago strain sequence, and the CDC2 sequence is identical to the Gauthier sequence. (C) One of the putative G4FS (yellow) located in the TP0131 5′-flanking region (within the TP0136 ORF) in *T. pallidum* subspecies and *T. paraluiscuniculi* showing sequence diversity among these strains. The G4FS G-score is indicated below the sequence. The Street14 sequence is identical to the UW104 strain sequence, the Dal-1 sequence is identical to the Chicago strain sequence, and the CDC2 sequence is identical to the Gauthier sequence. (D) Putative G4FS located upstream of the *tprK* ORF in *T. pallidum* subspecies and *T. paraluiscuniculi* showing sequence diversity between the rabbit treponeme and the rest of the strains. Street14, Dal-1, and CDC2 sequences are identical to Nichols Gen sequences.

ing during DNA replication (38, 62). In the Nichols Gen strain, the implications of the loss of ~30% of its donor sites due to the ~1.1-kb deletion for the physiology of this spirochete and disease progression are currently unclear. Even when a full repertoire of DSs is present (such as in the Nichols Sea strain genome), diversification of the TprK antigen in the different Nichols lineages appears to require an adaptive host immune response against the initial TprK variant (36). Since its isolation in 1912 from the cerebrospinal fluid (CSF) of a patient with secondary syphilis (46), the Nichols strain of *T. pallidum* subsp. *pallidum* has been continuously propagated in rabbits, and the hypothesis that slower *tprK* variability in the Nichols strain could reflect its adaptation to rapid passage in immunologically naive rabbits cannot be excluded.

Variability in the TP0131 3′-flanking region also affects the ORF annotation process for the sequences encompassing TP0126-TP0127. This topic was also addressed by Mikalová et al. (43), who predicted a TP126 ORF of 1,374 nt (458 amino acids [aa]) in Samoa D but an ORF of only 669 nt (223 aa) in Gauthier and Fribourg-Blanc. According to this annotation, TP0126 sequence would encompass several *tprK* DSs in Samoa D but none in Gauthier or Fribourg-Blanc. To further complicate this scenario, in the original *T. pallidum* subsp. *pallidum* genome project (19), in the TP0131 −3′-flanking region (Fig. 1, Nichols Gen strain), TP0126 was annotated as a 291-aa putative protein, an intermediate size between those of Samoa D and Gauthier/Fribourg-Blanc TP0126 (43). Based on our study, however, we propose that two elements contribute to the correct annotation of this genomic region: namely, the presence of variable homopolymeric G-tracts (poly-G) both upstream of TP0126 and within the TP0127 ORF (Fig. 7A to C). With regard to TP0126, a poly-G of 10 G residues (like the one reported for Samoa D) (43) extends the TP0126 coding sequence to 1,374 nt, while a poly-G of only nine residues induces the identification of an alternative start for TP0126, located 705 nt downstream of the one identified by Mikalová et al. in the Samoa D strain. The intrastrain variability of this poly-G prompted us to identify the most likely start codon for TP0126. Initial prediction of promoter elements for TP0126 (Fig. 7A) supports our hypothesis that, in all these genomes, TP0126 is actually a 223-aa protein, encoded by an ORF that does not encompass any *tprK* DS. This shorter TP0126 is predicted to carry a cleavable signal peptide, and it shows structural homology (data not shown) with *Escherichia coli* OmpW, an OMP involved in transporting small hydrophobic molecules from the extracellular milieu into the periplasm (1, 30). Also, because the poly-G sequence is located between putative classical σ^{70} signatures, we postulate that TP0126 might undergo phase variation, similar to that described for *Nesseria menigitidis porA* (63), where the poly-G is also located between the −10 and −35 σ signatures. Changes in the poly-G length in the *porA* promoter regulate gene transcription by varying the optimal spacing of 17 nucleotides between the consensus sequences. Preliminary results (not shown) using a reporter gene system (23), in which our identified TP0126 promoter drives green fluorescent protein (GFP) transcription, showed that poly-G sequences of various lengths induced different levels of fluorescence reflecting modulation of transcription. This reinforces the hypothesis that, in all isolates studied here, TP0126 might indeed be a shorter ORF than predicted for Samoa D (43) or the Nichols Gen strain (19). While the poly-G associated with TP0126 is located in the putative promoter of this shorter ORF, the poly-G tract associated with TP0127 is located within its coding sequence. Depending upon the number of G's present within the TP0127 ORF, two alternative TP0127 proteins of comparable size but different primary structure can be predicted, as well as a third variant characterized by a premature stop codon that truncates the protein (Fig. 7B). Previously, poly-G tracts were shown in *T. pallidum* subsp. *pallidum* to control transcription of a subset of *tpr* genes (23); however, in the case of TP0127, variation in the poly-G length might also induce the synthesis of different proteins from the same genomic locus. Interestingly, when the sequence unique to the TP0127.3 variant (Fig. 7D) is analyzed for structural homology using the LOMETS metaserver (zhanglab.ccmb.med.umich.edu/LOMETS), a MutS1 homolog is identified. MutS1 homologs are involved in preventing recombination between partially homologous DNA sequences, suggesting a possible role for this DS-neighboring ORF in the generation of *tprK* variability.

Upstream of the *tprK* ORF, where the gene promoter is located, sequence heterogeneity might influence the level of *tprK* expression, due to (i) the presence of a variable A/T-rich element, similar to an UP element, upstream of the *tprK* −35 σ^{70} signature and (ii) the absence of a typical RBS in *T. paraluiscuniculi*, compared to the other treponemal strains analyzed here. Preliminary quantitative amplification studies targeting *tprK* mRNA in Nichols Sea, Chicago, Sea81-4, Gauthier, Samoa D, Iraq B, and Cuniculi A strains (24; Giacani, unpublished) show differing levels of expression of the *tprK* gene in these isolates, which warrants further exploration of the role of sequence diversity in the *tprK* promoter region among strains.

It is particularly noteworthy that G-rich sequences able to fold into guanine quadruplex (G4) motifs are found upstream of the *tprK* expression site and within the genomic regions where the *tprK* DSs are located. Although G4FS (31, 65) have been implicated in many biological processes in eukaryotes, such as mRNA stability (55), transcription pausing (64), Fragile X mental retardation protein binding (14), and translation initiation and repression (6), Cahoon and Seifert (8, 9) recently reported that a G4FS located upstream of the *pilE* gene, in the intergenic region preceding the ORF, is required for pilin antigenic variation in *N. gonorrhoeae*. Specifically, this putative *N. gonorrhoeae* G4FS is hypothesized to be a specialized sequence that favors recombination initiation and gene conversion between *pilE* and its donor sites. The role of G4FS in the generation of *tprK* variants is still hypothetical, and it awaits further experimental evidence. It is intriguing that, in treponemal isolates, the location of these G4FS mirrors that reported for *N. gonorrhoeae*, with the G-rich sequence of the G4FS in proximity to the ORF start codon and in the opposite strand with respect to the expression site (8). The finding of G4FS within the genomic region containing *tprK* DSs is novel, however, in that there is no parallel finding for *N. gonorrhoeae*, in which the presence of G4FS near the *pilE* donor sites was not reported (8). The presence of sequence variability in the DNA regions where these treponemal G4FS are located suggests a possible role for these sequences in the generation of *tprK* variants in pathogenic treponemes. Interestingly, when G4FS are predicted in *Borrelia burgdorferi* B31 linear plasmid 28-1 (lp28-1), which contains donor sequences and an expression site for the vari-

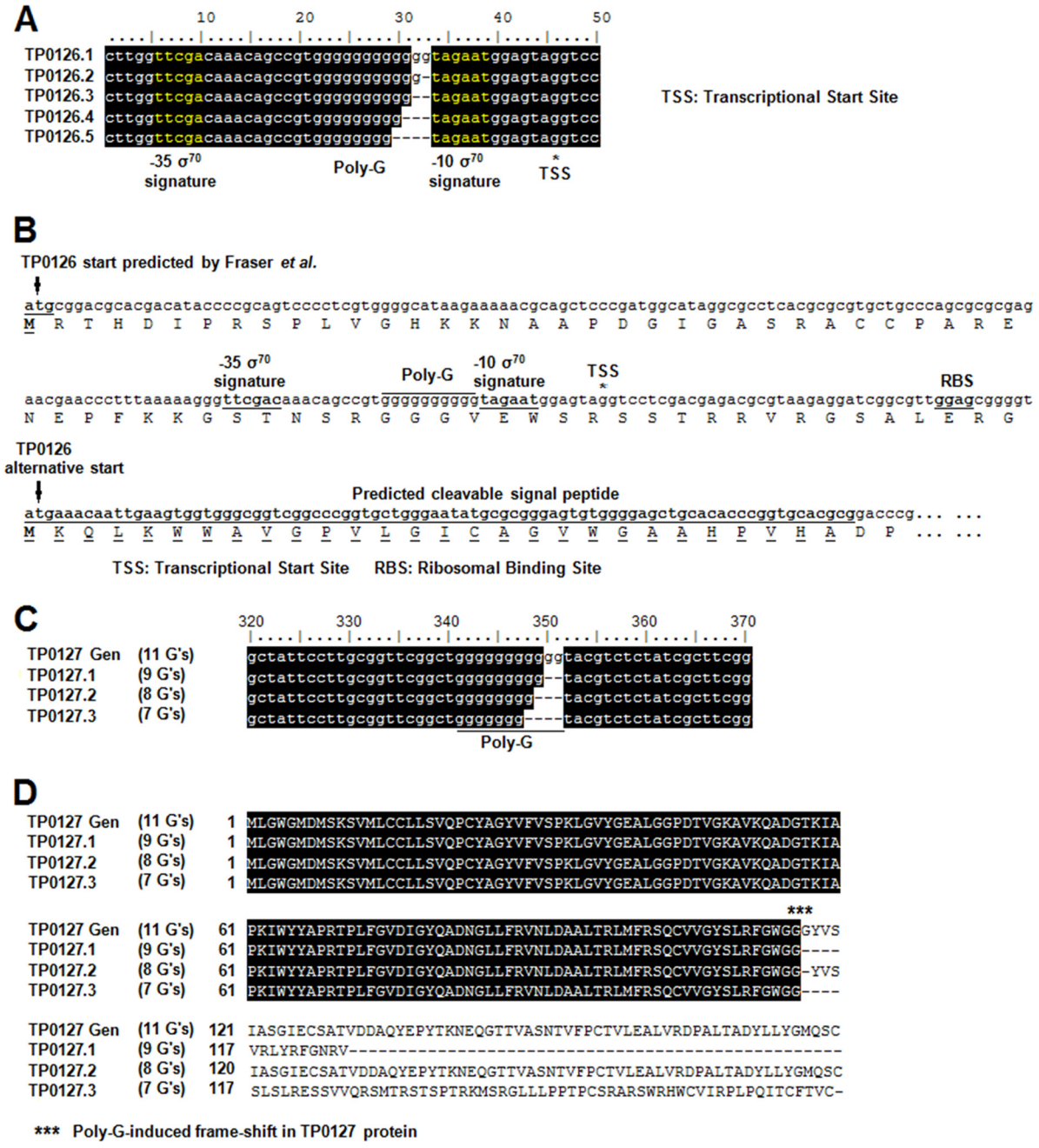


FIG 7 Length variability of poly-G sequences in the promoter region of TP0126 and predicted effect in translation of TP0127. (A) Variability in the number of G residues in the poly-G upstream of TP0126 ORF. Sequences obtained from amplification of the TP0126 ORF and its newly predicted promoter regions with template DNA from the *T. pallidum* subsp. *pallidum* Chicago strain. Putative −10 and −35 σ^{70} signatures are highlighted in yellow. (B) Comparison of TP0126 start codon identified by Fraser et al. (19) and the alternative TP0126 start downstream of the poly-G sequence. Predicted σ^{70} signatures, transcriptional start site (TSS), ribosomal binding site (RBS), and putative signal peptide for the alternative ORF are indicated. (C) Variability in the poly-G within the TP0127 ORF. TP0127 Gen refers to the TP0127 ORF identified by Fraser et al. (19); alternative poly-G lengths were isolated from the *T. pallidum* subsp. *pallidum* Chicago strain. (D) Effect of different poly-G length on TP0127 predicted proteins. Nine G residues (TP0127.1) cause an early TP0127 truncation; 11 or 8 residues are associated with a 180- or 179-aa-long peptide (TP0127 Gen and TP0127.2, respectively), while 7 residues (TP0127.3) result in a 175-aa-long protein with a different amino acid sequence than TP0127.2.

able surface-exposed lipoprotein VlsE (48), 12 of 15 predicted G4FS are located within the *vlsE* expression sites (nt 26997 to 28073; data not shown). Even though the location of these G4FS differs from the one described in both *N. gonorrhoeae* and *T. pallidum* subsp. *pallidum*, the distribution of G4FS on lp28-1 is suggestive of the involvement of these sequences also in the generation of VlsE variants in the Lyme disease spirochete.

Our findings set the stage for future studies in which many aspects of the *tprK* antigenic variation system may be explored, including DS usage among species and subspecies and the roles of

SNPs, G4FS, and small sequence changes in adaptation and pathogenicity.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI063940 and AI42143 (to S.A.L., A.C.-L., and L.G.) and AI096129 (to L.G.) and the 2010 ASTDA Developmental Award (to L.G.).

We are grateful to Christina Marra for providing us with the *T. pallidum* subsp. *pallidum* UW104 and UW126 strains.

REFERENCES

1. **Albrecht R, Zeth K, Soding J, Lupas A, Linke D.** 2006. Expression, crystallization and preliminary X-ray crystallographic studies of the outer membrane protein OmpW from *Escherichia coli*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **62**:415–418.
2. **Antal GM, Lukehart SA, Meheus AZ.** 2002. The endemic treponematoses. Microbes Infect. **4**:83–94.
3. **Bagga PS.** 2008. Bioinformatics approaches for studying untranslated regions of mRNAs. Methods Mol. Biol. **419**:1–21.
4. **Baker-Zander SA, Lukehart SA.** 1984. Antigenic cross-reactivity between *Treponema pallidum* and other pathogenic members of the family Spirochaetaceae. Infect. Immun. **46**:116–121.
5. **Baughn RE, Musher DM.** 2005. Secondary syphilitic lesions. Clin. Microbiol. Rev. **18**:205–216.
6. **Bonnal S, et al.** 2003. A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. J. Biol. Chem. **278**:39330–39336.
7. **Brinkman MB, et al.** 2008. A novel *Treponema pallidum* antigen, TP0136, is an outer membrane protein that binds human fibronectin. Infect. Immun. **76**:1848–1857.
8. **Cahoon LA, Seifert HS.** 2009. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. Science **325**:764–767.
9. **Cahoon LA, Seifert HS.** 2011. Focusing homologous recombination: pilin antigenic variation in the pathogenic *Neisseria*. Mol. Microbiol. **81**:1136–1143.
10. **Castro LG.** 1994. Nonvenereal treponematosis. J. Am. Acad. Dermatol. **31**:1075–1076.
11. **Centurion-Lara A, et al.** 1999. *Treponema pallidum* major sheath protein homologue Tpr K is a target of opsonic antibody and the protective immune response. J. Exp. Med. **189**:647–656.
12. **Centurion-Lara A, et al.** 2004. Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. Mol. Microbiol. **52**:1579–1596.
13. **Cox DL, et al.** 2010. Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of *Treponema pallidum*. Infect. Immun. **78**:5178–5194.
14. **Darnell JC, et al.** 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell **107**:489–499.
15. **DiGiacomo RF, et al.** 1984. Clinical course and treatment of venereal spirochaetosis in New Zealand White rabbits. Br. J. Vener. Dis. **60**:214–218.
16. **Estrem ST, Gaal T, Ross W, Gourse RL.** 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. U. S. A. **95**:9761–9766.
17. **Estrem ST, et al.** 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev. **13**:2134–2147.
18. **Farnsworth N, Rosen T.** 2006. Endemic treponematosis: review and update. Clin. Dermatol. **24**:181–190.
19. **Fraser CM, et al.** 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science **281**:375–388.
20. **Fribourg-Blanc A, Niel G, Mollaret HH.** 1963. Note sur quelques aspects immunologiques du cynocephale africain. Bull. Soc. Pathol. Exot. **56**:474–485.
21. **Giacani L, Hevner K, Centurion-Lara A.** 2005. Gene organization and transcriptional analysis of the *tprJ*, *tprI*, *tprG*, and *tprF* loci in *Treponema pallidum* strains Nichols and Sea 81-4. J. Bacteriol. **187**:6084–6093.
22. **Giacani L, et al.** 2010. Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. J. Bacteriol. **192**:2645–2646.
23. **Giacani L, Lukehart S, Centurion-Lara A.** 2007. Length of guanosine homopolymeric repeats modulates promoter activity of Subfamily II *tpr* genes of *Treponema pallidum* ssp. *pallidum*. FEMS Immunol. Med. Microbiol. **51**:289–301.
24. **Giacani L, et al.** 2007. Quantitative analysis of *tpr* gene expression in *Treponema pallidum* isolates: differences among isolates and correlation with T-cell responsiveness in experimental syphilis. Infect. Immun. **75**:104–112.
25. **Giacani L, et al.** 2010. Antigenic variation in *Treponema pallidum*: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J. Immunol. **184**:3822–3829.
26. **Giacani L, et al.** 2004. Tpr homologs in *Treponema paraluiscuniculi* Cuniculi A strain. Infect. Immun. **72**:6561–6576.
27. **Gjestland T.** 1955. The Oslo study of untreated syphilis. Acta Derm. Venereol. Suppl. **35**(Suppl. 34):3–368.
28. **Golden MR, Marra CM, Holmes KK.** 2003. Update on syphilis: resurgence of an old problem. JAMA **290**:1510–1514.
29. **Hazlett KR, et al.** 2001. The TprK protein of *Treponema pallidum* is periplasmic and is not a target of opsonic antibody or protective immunity. J. Exp. Med. **193**:1015–1026.
30. **Hong H, Patel DR, Tamm LK, van den Berg B.** 2006. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. J. Biol. Chem. **281**:7568–7577.
31. **Huppert JL.** 2010. Structure, location and interactions of G-quadruplexes. FEBS J. **277**:3452–3458.
32. **Joachimi A, Benz A, Hartig JS.** 2009. A comparison of DNA and RNA quadruplex structures and stabilities. Bioorg. Med. Chem. **17**:6811–6815.
33. **Kikin O, D'Antonio L, Bagga PS.** 2006. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. **34**:W676–W682.
34. **Kostadinov R, et al.** 2006. GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. Nucleic Acids Res. **34**:D119–D124.
35. **LaFond RE, et al.** 2003. Sequence diversity of *Treponema pallidum* subsp. *pallidum tprK* in human syphilis lesions and rabbit-propagated isolates. J. Bacteriol. **185**:6262–6268.
36. **LaFond RE, Centurion-Lara A, Godornes C, Van Voorhis WC, Lukehart SA.** 2006. TprK sequence diversity accumulates during infection of rabbits with *Treponema pallidum* subsp. *pallidum* Nichols strain. Infect. Immun. **74**:1896–1906.
37. **LaFond RE, Lukehart SA.** 2006. Biological basis for syphilis. Clin. Microbiol. Rev. **19**:29–49.
38. **Levinson G, Gutman GA.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.
39. **Lukehart SA, Baker-Zander SA, Sell S.** 1980. Characterization of lymphocyte responsiveness in early experimental syphilis. I. In vitro response to mitogens and *Treponema pallidum* antigens. J. Immunol. **124**:454–460.
40. **Marra CM, et al.** 2006. Antibiotic selection may contribute to increases in macrolide-resistant *Treponema pallidum*. J. Infect. Dis. **194**:1771–1773.
41. **Martin PM, Cockayne A, Georges AJ, Penn CW.** 1990. Immune response to *Treponema pertenue* and *Treponema pallidum* Nichols in patients with yaws. Res. Microbiol. **141**:181–186.
42. **Matejkova P, et al.** 2008. Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. BMC Microbiol. **8**:76.
43. **Mikalová L, et al.** 2010. Genome analysis of *Treponema pallidum* subsp. *pallidum* and subsp. *pertenue* strains: most of the genetic differences are localized in six regions. PLoS One **5**:e15713. doi:10.1371/journal.pone.0015713.
44. **Morgan CA, Lukehart SA, Van Voorhis WC.** 2003. Protection against syphilis correlates with specificity of antibodies to the variable regions of *Treponema pallidum* repeat protein K. Infect. Immun. **71**:5605–5612.
45. **Morgan CA, Molini BJ, Lukehart SA, Van Voorhis WC.** 2002. 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.
46. **Nichols HJ, Hough WH.** 1913. Demonstration of *Spirochaeta pallida* in the cerebrospinal fluid. JAMA **60**:108–110.
47. **Noordhoek GT, et al.** 1990. A new attempt to distinguish serologically the subspecies of *Treponema pallidum* causing syphilis and yaws. J. Clin. Microbiol. **28**:1600–1607.
48. **Norris SJ.** 2006. Antigenic variation with a twist—the *Borrelia* story. Mol. Microbiol. **60**:1319–1322.

49. **Perine PL.** 1998. Endemic treponematoses, p 1033–1036. *In* Isselbacher KJ, et al. (ed), Harrison's principles of internal medicine, 14th ed. McGraw Hill, Inc., New York, NY.
50. **Perine PL, et al.** 1984. Handbook of endemic treponematoses: yaws, endemic syphilis, and pinta. World Health Organization, Geneva, Switzerland.
51. **Roman GC, Roman LN.** 1986. Occurrence of congenital, cardiovascular, visceral, neurological, and neuro-ophthalmologic complications in late yaws: a theme for future research. Rev. Infect. Dis. **8**:760–770.
52. **Ross W, Aiyar SE, Salomon J, Gourse RL.** 1998. *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J. Bacteriol. **180**:5375–5383.
53. **Ross W, Ernst A, Gourse RL.** 2001. Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. Genes Dev. **15**:491–506.
54. **Simon, R. P.** 1985. Neurosyphilis. Arch. Neurol. **42**:606–613.
55. **Simonsson T.** 2001. G-quadruplex DNA structures—variations on a theme. Biol. Chem. **382**:621–628.
56. **Smajs D, et al.** 2011. Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. PLoS One **6**:e20415. doi:10.1371/journal.pone.0020415.
57. **Small JD, Newman B.** 1972. Venereal spirochetosis of rabbits (rabbit syphilis) due to *Treponema cuniculi*: a clinical, serological, and histopathological study. Lab. Anim. Sci. **22**:77–89.
58. **Smith JL, et al.** 1971. Neuro-ophthalmological study of late yaws and pinta. II. The Caracas Project. Br. J. Vener. Dis. **47**:226–251.
59. **Smith JL, Pesetsky BR.** 1967. The current status of *Treponema cuniculi*. Review of the literature. Br. J. Vener. Dis. **43**:117–127.
60. **Thornburg RW, Baseman JB.** 1983. Comparison of major protein antigens and protein profiles of *Treponema pallidum* and *Treponema pertenue*. Infect. Immun. **42**:623–627.
61. **Turner TB, Hollander DH.** 1957. Biology of the treponematoses. World Health Organization, Geneva, Switzerland.
62. **van Belkum A, Scherer S, van Alphen L, Verbrugh H.** 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62**:275–293.
63. **van der Ende A, Hopman CT, Dankert J.** 2000. Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. Infect. Immun. **68**:6685–6690.
64. **Yonaha M, Proudfoot NJ.** 1999. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Mol. Cell **3**:593–600.
65. **Zhou W, Brand NJ, Ying L.** 2011. G-quadruplexes-novel mediators of gene function. J. Cardiovasc. Transl. Res. **4**:256–270.