A segment for publication info and copyright.

Gene Organization and Transcriptional Analysis of the *tprJ*, *tprI*, *tprG*, and *tprF* Loci in *Treponema pallidum* Strains Nichols and Sea 81-4†

Lorenzo Giacani,[1] Karin Hevner,[1] and Arturo Centurion-Lara[1,2]*

Departments of Medicine[1] and Pathobiology,[2] University of Washington, Seattle, Washington

The *tpr* gene family of *Treponema pallidum* subsp. *pallidum*, the causative agent of syphilis, has recently become the focus of intensive investigation. TprF and TprI sequences are highly conserved among different isolates and are the targets of strong humoral and cellular immune responses of the host, and immunization with a recombinant peptide from the amino terminus of these antigens has been shown to alter significantly lesion development following homologous challenge. This indicates that these antigens are expressed during infection and strongly suggests a key functionality. *tprF* and *tprI* are located immediately downstream of the *tprG* and *tprJ* genes, respectively, separated by very short intergenic spacers (55 nucleotides for *G-F* and 56 nucleotides for *J-I*). Preliminary analysis using gene-specific primers failed to amplify *tprJ* in the Sea 81-4 isolate. In this study, sequence and transcriptional analysis of these loci showed a similar gene organization in the Nichols and Sea 81-4 strains, a complex pattern of transcription, and the presence of G homopolymeric repeats of variable lengths upstream of the *tprF*, *tprI*, *tprG*, and *tprJ* transcriptional start sites. However, distinctive features were also identified in the Sea 81-4 isolate, including a *tprG*-like open reading frame in the *tprJ* locus, a frameshift and a premature termination in the *tprG* coding sequence, a longer *tprG-tprF* intergenic spacer, and absence of cotranscription of the *tprG-tprF* genes.

Treponema pallidum subsp. *pallidum*, the etiologic agent of venereal syphilis, is highly pathogenic and able to cause a chronic, lifelong infection in the infected individual despite a vigorous humoral and cellular immune response (22). The Nichols strain genome sequence revealed 48 open reading frames (ORFs) coding for potential virulence factors (8). Among them, the *tpr* gene family consists of 12 paralogs with homology to the major sheath protein of *Treponema denticola* and accounts for approximately 2% of the small 1.138-Mb *T. pallidum* genome. According to their predicted amino acid homologies, the Tprs are divided into three subfamilies: subfamily I (*tprC, -D, -F,* and *-I*), subfamily II (*tprE, -G,* and *-J*), and subfamily III (*tprA, -B, -H, -K,* and *-L*) (1). Subfamily I and subfamily II Tprs, respectively, show common amino and carboxyl termini flanking central domains that differ in sequence and length. Only limited homology among its members characterizes subfamily III Tprs (1). TprF is unique in that it contains a deletion of the central nonconserved region and a portion of the C-terminal conserved region, resulting in a frameshift and premature termination. Analysis of the *tprF* and *tprI* ORFs in the genome sequence of the Nichols strain revealed that they are located immediately downstream of the *tprG* and *tprJ* genes separated by very short intergenic spacers (55 nucleotides for *G-F* and 56 nucleotides for *J-I*). Preliminary analysis using gene-specific primers showed that the *tpr* genes are conserved among syphilis isolates; however, *tprJ*-specific primers failed to demonstrate the presence of this gene in the

Sea 81-4 strain, suggesting either the absence of *tprJ* in this strain or sequence differences in the primer binding sites.

In seven *T. pallidum* subsp. *pallidum* isolates examined to date (20), *tprF* and *tprI* are invariant, suggesting a key functionality of these genes. TprI and TprF are predicted to have NH$_2$-terminal cleavable signal peptides and a putative outer membrane location. Immunization with a recombinant peptide comprising the conserved amino terminus of TprF/TprI altered significantly lesion development following intradermal homologous challenge, although it did not prevent infection (20). Additionally, antibody and T-cell responses to TprF/TprI during infection with different syphilis isolates have been recently studied in the rabbit model (12, 20). Overall, the TprF/TprI NH$_2$-terminal conserved region elicits a marked antibody response during infection with the *T. pallidum* subsp. *pallidum* Chicago isolate and also strong T-cell responses during infection with different *T. pallidum* subsp. *pallidum* isolates (Nichols, Bal-3, Chicago, and Sea 81-4 strains) (20). Because the sequences of TprF and TprI are identical, differential reactivities cannot be explained by differences in inherent antigenicity or sequence heterogeneity in the antigens, suggesting modulation of *tpr* gene expression during infection. This is supported by recent microarray and real-time PCR studies (18). Our own studies using real-time PCR also support the hypothesis of differential transcription of the *tpr* genes among syphilis strains (unpublished data). Taken together, these data strongly suggest that these antigens are playing a significant role in the immune response to syphilis and likely in syphilis pathogenesis.

Gene regulation in *T. pallidum* has been little studied, and the mechanisms and regulatory elements of gene expression of the *tpr* genes remain to be defined. Because of the apparent importance of the TprF/TprI antigens in the immune response during syphilis infection and because strain-specific differences in neuroinvasion and clinical phenotype (21) between the Sea

* Corresponding author. Mailing address: Department of Medicine, Box 359779, Harborview Medical Center, 325 Ninth Ave., Seattle, WA 98104-2499. Phone: (206) 341-5364. Fax: (206) 341-5363. E-mail: acentur@u.washington.edu.

81-4 and Nichols strains have been reported (strain Sea 81-4 induces higher pleocytosis in cerebrospinal fluid than Nichols, fewer indurated skin lesions after intravenous inoculation, and lower Venereal Disease Research Laboratory [VDRL] titers), we examined the sequence organization, promoter regions, and transcription patterns with respect to *tprG* and *tprJ* in the Nichols and Sea 81-4 isolates in a further attempt to understand the role of the *tpr* gene family during infection. We have identified in this study distinctive features in the Sea 81-4 isolate, such as a unique sequence composition of the *tprJ* locus (which contains a *tprG*-like ORF), a frameshift and premature termination of the ORF in the *tprG* locus, a longer *tprG*-*tprF* intergenic spacer, and a different pattern of transcription than the Nichols isolate.

MATERIALS AND METHODS

***T. pallidum* strain propagation and nucleic acid extraction.** *T. pallidum* subsp. *pallidum* Nichols and Sea 81-4 strains were propagated in New Zealand White rabbits as reported elsewhere (14). The Sea 81-4 strain was isolated in 1981 from the primary lesion of a patient attending the STD Clinic at Harborview Medical Center, Seattle, Wash. The Nichols strain, originally isolated from cerebrospinal fluid in 1912 (16), was provided by James N. Miller (University of California, Los Angeles) in 1979. Treponemes were extracted from infected rabbit testes as described elsewhere (3). Briefly, organisms were separated from host cellular debris by low-speed centrifugation and, then, supernatants were spun in a microcentrifuge for 30 min at 12,000 rpm at 4°C and the pellets were resuspended in 200 µl 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 (2) using the QIAamp DNA Mini kit (QIAGEN Inc., Chatsworth, CA). RNA extraction was performed according to the manufacturer's instructions followed by treatment with DNase I (Invitrogen, Carlsbad, CA). DNase-treated RNA was stored at −80°C until use. Every extraction was performed taking careful precautions to prevent cross-contamination.

PCR amplification of Sea 81-4 *tpr* genes, cloning, sequencing, and sequence analysis. Primers were designed upstream of *tprJ* and *tprG* as well as downstream of *tprI* and *tprF*, based on the Nichols sequences. They were used to amplify the corresponding regions spanning the *tprJ/tprI* and the *tprG/tprF* ORFs from the Sea 81-4 strain (Table 1). The reactions were performed using the LA-PCR long PCR kit (Takara Bio Inc., Shiga, Japan) in a 100-µl final volume containing 200 µM deoxynucleoside triphosphates (dNTPs), 2.5 mM MgCl₂, and 0.05 U/µl of LA-*Taq*. The cycling conditions were denaturation at 94°C for 1 min, followed by 35 cycles of 98°C for 20 s and 68°C for 5 min. Final extension was 72°C for 10 min. The products were separated in 1% agarose gels, purified using the QIAquick gel extraction kit (QIAGEN), and cloned into the TOPO-XL cloning vector (Invitrogen). Plasmid DNA from colonies containing inserts was extracted using the QIAGEN Plasmid Mini kit (QIAGEN) and sequenced in both directions with the Applied Biosystems dye terminator sequencing kit (Perkin-Elmer, Foster City, CA) using the primer walking approach. Sequences were aligned using the Multiple Alignment program (http://searchlauncher.bcm.tmc.edu/multi-align /multi-align.html), and shading of identical bases was done with the Boxshade 3.21 program (http://www.ch.embnet.org/software/BOX_form.html). Open reading frames were determined using the NCBI ORF finder (http://www.ncbi.nlm .nih.gov/gorf/gorf.html) and by sequence comparison with the Nichols *tpr* loci.

RT-PCR analysis. All primers used in this study are listed in Table 1. Three sets of primers were used to determine the presence of mRNA spanning the intergenic spacers between *tprJ* and *tprI* in the Nichols strain (JI1 plus JIGF1, JI2 plus JIGF1, and JI2 plus JIGF2) as well as in the Sea 81-4 strain (JI1 plus JIGF1, JI1 plus JIGF2, and JI2 plus JIGF2), two sets of primers for the Nichols *tprG*-*tprF* intergenic region (GF1 plus JIGF1 and GF2 plus JIGF2), and two sets for the Sea 81-4 *tprG*-*tprF* spacer (sGF1 plus JIGF1 and sGF2 plus JIGF2). To investigate the presence of transcript upstream of *tprJ* and *tprG* in the Nichols and Sea 81-4 strains and as a first approach to localize possible transcriptional start sites (TSSs), a total of six primer combinations were used in the 5′-flanking regions of *tprJ* (PJ1 plus PJG, PJ2 plus PJG, and PJ3 plus PJG) and *tprG* (PG1 plus PJG, PG2 plus PJG, and PG3 plus PJG). The PJG antisense primer sequence is shared by both *tprJ* and *tprG*. Specific primers targeting the central regions of *tprJ*, *tprI*, *tprG*, and *tprF* were used to assess the presence of gene-specific mRNA (listed as reverse transcription-PCR [RT-PCR] control primers

in Table 1). Primers that amplify *tp47* were used as a positive control. Primer sequences and optimal annealing temperatures for each primer set are listed in Table 1; the primer locations are shown schematically in Fig. 3, below.

cDNA synthesis from DNase I-treated *T. pallidum* RNA was performed as described elsewhere (1). The same primers were used in both strains, except when otherwise specified. PCR amplifications were performed in a 50-µl final volume, containing 200 µM dNTPs, 50 mM Tris-HCl (pH 9.0 at 20°C), 1.5 mM MgCl₂, 20 mM NH₄SO₄, and 2.5 U of *Taq* polymerase (Promega, Madison, WI). The cycling conditions were denaturation at 95°C for 2 min, followed by 95°C for 1 min, annealing (at temperatures reported in Table 1) for 1 min, and extension at 72°C for 1 min for 45 cycles. Final extension was 72°C for 10 min. DNase I-treated RNA, DNase I-untreated RNA, and DNA were used as amplification controls. All PCR products were separated in 2% agarose gels.

In order to assess sequence identity and primer specificity, the amplicons relative to the promoter regions and intergenic spacers from the Sea 81-4 and Nichols strains obtained from both cDNA (when a positive signal could be retrieved) and DNA (used as positive controls in the reactions) were cloned into the TOPO-TA cloning vector (Invitrogen). Plasmids were purified from the clones containing inserts using the QIAGEN Plasmid Mini kit (QIAGEN). Sequencing was performed as described above. Cloned amplicons from the promoter regions in both strains were generated with the primers PJ2 and PG2 in combination with PGJ (Table 1). The products cloned from the *tprJ-tprI* intergenic spacer were obtained with the JI2 and JIGF1 primers (Table 1), also common to both strains. In contrast, cloned amplicons from the *tprG-tprF* intergenic region were obtained with specific primers GF 1 and JIGF 1 for the Nichols strain and sGF 1 with JIGF 1 for the Sea 81-4 strain (Table 1). An average of seven different clones were sequenced and analyzed for each region.

IVT assay. Initial computer analysis (using the Neural Network Promoter Prediction program for prokaryotes, available at http://www.fruitfly.org/seq _tools/promoter.html) predicted the presence of putative sigma 70-like promoters in the 5′-flanking regions of *tprG*, *tprJ*, and *tprF/I* (sequences are identical for *tprF* and *tprI* in this region) in both strains. The regions upstream of these four *tpr* genes were amplified only from the Nichols strain, because they are identical in both strains (primers in Table 1). They were initially cloned into the TOPO-TA vector. The inserts were sequenced to verify sequence identity and the absence of mutations. The inserts were subsequently excised using the EcoRI and PstI restriction sites and subcloned into the pSA509 vector (kindly provided by D. J. Jin, Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Md.), specifically designed for in vitro transcription (IVT) assays (4). Plasmid-containing inserts were purified using the QIAgen Plasmid Mini kit (QIAGEN) and brought to a final concentration of 1 µg/ml with diethyl pyrocarbonate (DEPC)-treated water.

One microgram of pSA509 plasmid carrying the putative promoter sequence of *tprJ*, *tprG*, or *tprF/I* was then mixed with 2 µl of 5× transcription buffer (0.2 M Tris-HCl, pH 7.5, 0.75 M KCl, 50 mM MgCl₂, 0.05% Triton X-100, 0.5 M dithiothreitol), 1 U of *Escherichia coli* RNA polymerase holoenzyme (Epicenter, Madison, WI), and DEPC-treated water to a final volume of 9 µl. After preincubation at 37°C for 15 min, 1 µl of 10× biotin RNA labeling mix (10 mM ATP, 10 mM CTP, 10 mM GTP, 6.5 mM UTP, 3.5 mM biotin-16-UTP, pH 7.5 [F. Hoffmann-La Roche Ltd., Basel, Switzerland]) was added to the reaction mixture and then incubated for 10 min at 37°C. The products obtained from the IVT assay were collected by brief centrifugation and treated with molecular-grade DNase I (Invitrogen). Five microliters of 10× DNase I buffer, 5 U of DNase I, and DEPC-treated water to a 50-µl final volume were added to the reaction mixture and then incubated at room temperature for 15 min. The reaction was stopped by adding 5 µl of 25 mM EDTA to the mixture followed by a 10-min incubation at 65°C. The IVT products were then purified by phenol-chloroform extraction (15) and analyzed by using a 5-′ rapid amplification of cDNA ends (5′-RACE) system.

5′-RACE. The 5′-RACE system (Invitrogen) was used to determine TSSs in the 5′-flanking regions of *tprF*, *tprI*, *tprJ*, and *tprG*. 5′-RACE analysis was performed on two different templates: *T. pallidum* total RNA (*tprJ* and *tprG* as well as *tprF* and *tprI* promoters of the Nichols and Sea 81-4 isolates [Table 2]) and RNA obtained from the IVT assays (*tprF/I*, *tprG*, and *tprJ* promoters of both isolates [Table 2]) as described above. 5′-RACE analysis was performed following the manufacturer's instructions on RNA obtained from IVT assays. When 5′-RACE was applied to total RNA, we followed the manufacturer's protocol except that the SuperScript system for cDNA synthesis was replaced by the ThermoScript RT kit (Invitrogen) to increase yield due to the high GC content (52.8%) of the *T. pallidum* genome. A 2.5-pmol aliquot of gene-specific primer and 2 µg of sample RNA were used in each reaction mixture. 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. Primers used in this set of experiments are listed

TABLE 1. Primers used in this study

Purpose and primer name	5′ to 3′ sequence	Product size (bp)[a], primer combination	Annealing temp (°C)
Cloning			
TPRJ-L	CGAGTGAGGCTCATCAAGAA	5,205, TPRJ-L with I-R	68
I-R	AGTAAGCCCTGCCCAAGAAC		
PG 1	CCCTGCGTTTCCCATCTG	4,233, PG1 with TPRF-R	68
TPRF-R	GTACTACCTTCCCCCGGTCT		
RT-PCR (controls)			
JS	TCTTCACACCCCGCAGGGAA	364, JS with JAs	65
JAs	CGTTATTTCCGTTCGCATCATC		
IS[b]	GACCCTGCCGATGCAGGTAAT	336, IS with IAs	65
IAs	TAAGCACGATGTCCGACTGACT		
GS	GAAGGTGTTCATTACCGACCCT	359, GS with GAs	65
GAs	TTGTAGCCTCAGCCGTAAGCTT		
FS[b]	GACCCTGCCGATGCAGGTAAT	266, FS with Fas	65
FAs	TCAGCAAGCACCCCTGTTC		
tp47S	CGTGTGGTATCAACTATGG	310, tp47S with tp47As	63
tp47As	TCAACCGTGTACTCAGTGC		
JI 1	CCTTCCGGCGGTTCCTC	558, (N.), 559 (S.), JI 1 with JIGF 1; 504 (N), 505 (S.), JI 1 with JIGF 2	60
RT-PCR (intergenic spacers)			
JI 2	TCCTCAGGGCACATTGGCCT	546 (N), 547 (S.), JI 2 with JIGF 1; 492 (N.), 493 (S.), JI 2 with JIGF 2	63
GF 1 (Nichols specific)	GTGCGTGGCATTCAGGAAAA	562, GF 1 with JIGF 1	60
		508, GF 1 with JIGF 2	60
sGF 1 (Sea 81-4 specific)	GACGCTATACAAAAATAATAACG	604, sGF 1 with JIGF 1	62
sGF 2 (Sea 81-4 specific)	CCGGCAGTTCCTCAGGAAAA	507, sGF 2 with JIGF 2	62
JIGF 1	GCCAGTGCGTGGATTCTT		
JIGF 2	ACTGACCTGCGGAGTGAGTA		
PJ 1	AAGTTTGCTTTCAGATCCGC	344 (N.), 343 (S.), PJ 1 with PJG	60
RT-PCR (5′-flanking regions)			
PJ 2	AAAGAAAAAGGATTTCCGCA	284 (N.), 283 (S.), PJ 2 with PJG	
PJ 3	GAACAATGCGCAGCACCG	224 (N.), 223 (S.), PJ 3 with PJG	60
PG 1	CCCTGCGTTTCCCATCTG	365 (N.), 368 (S.), PG 1 with PJG	60
PG 2	CGCGTACCCACTTCTCTCTC	287, PG 2 with PJG	60
PG 3	AAACATGCGCAAAATAAGGG	242, PG 3 with PJG	60
PJG	GTGGCAGAGCCAGTTAGCTT		
5′-RACE			
J/G RACE-1[c]	CATACAGTGCCGGGTGCT	cDNA synthesis only	
J/G-RACE-2	GCATAACCAGGGGAAAGGAT	395, J/G-RACE with RACE AAP primer	63
F/I-RACE-1[c]	GACGCAAGCTCTACTGCCA	cDNA synthesis only	
F/I-RACE-2	TTCAGCTGCAGCTGTGCC	Expected size, 337	63
F/I-RACE-3	GCGGGTGTGGGTGTGCT	293, F3-RACE with RACE AUA primer	63
IVT			
PJ 1	AAGTTTGCTTTCAGATCCGC	351, PJ 1 with JG-IVT-rev	60
PG 1	CCCTGCGTTTCCCATCTG	372, PJ 1 with JG-IVT-rev	60
JG-IVT-rev	CTGCAGTGTGGCAGAGCCAG		
E/I 1	ACAGCTGCGTGCTGGTATTT	520 (N.), 521 (S.), E 1 with E-IVT-rev	60
E/I-IVT-rev	CTGCAGAGACGCAAGCTCTACTGC		

[a] (N.) indicates product size in the Nichols strain; (S.) indicates product size in the Sea 81-4 strain.
[b] IS and FS primers are identical.
[c] J/I-RACE-1 is specific for both *tprJ* and *tprI*, and G/F-RACE-1 is specific for both *tprG* and *tprF*.

in Table 1. Primers J/G-RACE 1 and F/I-RACE 1, used for initial retrotranscription from total *T. pallidum* RNA, are *tprJ/G* and *tprI/F* specific, respectively, due to sequence identity between these pairs of genes in the region where the primer was designed. All PCR amplification reactions were performed using 5 μl of dC-tailed cDNA in a 50-μl final volume containing 200 μM of each dNTP, 20 mM Tris-HCl (pH 8.4), 1.5 mM MgCl₂, 50 mM KCl, 400 nM of each primer, and 2.5 U of *Taq* DNA polymerase (Invitrogen). For nested PCR, 1 μl of the original amplicon of a 1/500 dilution was used. Cycling parameters were 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. Final extension was 10 min at 72°C. Thirty-five cycles were required for amplification of IVT-generated transcripts, 45 cycles for when

5′-RACE was used on total RNA, and 35 cycles for nested PCR to identify the Sea 81-4 *tprF* TSS. PCR products were separated in 2% agarose gels, gel purified, cloned, and sequenced as described above.

RESULTS

Sequence organization of the Sea 81-4 *tprJ/I* and *tprG/F* ORFs. Long-range PCR amplification using primers (Table 1) that bind upstream of the start codons of *tprJ* and *tprG* in

TABLE 2. 5′-RACE and 5′-RACE–IVT reactions.

| Gene(s)	5′ RACE on total RNA	5′-RACE on IVT RNA
tprJ	Performed on Nichols and Sea 81-4 strains; primer J/G-RACE-1 is shared by both *tprJ* and *tprG;* it is located at position +282 in the ORF.	Performed on the *tprJ* putative promoter sequence cloned from the Nichols strain[a] with the J/G-RACE-1 primer
tprG	Same as for *tprJ*	Performed on the *tprG* putative promoter sequence cloned from the Nichols strain[a] with the J/G-RACE-1 primer
tprI/*tprF*	Performed on Nichols and Sea 81-4 strains; primer F/I-RACE-1 is shared by both *tprF* and *tprI;* it is located at position +295 in the ORF.	Performed on the *tprF/I*[b] putative promoter sequence cloned from the Nichols strain[a] with the F/I-RACE-1 primer

[a] This sequence is identical in both the Sea 81-4 and Nichols strains of *T. pallidum.*
[b] *tprI* and *tprF* share an identical sequence in this region.

combination with primers binding the regions downstream of the stop codons of *tprI* and *tprF*, respectively, yielded in the Sea 81-4 isolate two amplicons of 5,205 and 4,903 bp in length (Fig. 1; see also Fig. S1A and S1B in the supplemental material).

After cloning and sequencing these amplicons, comparative analysis of the Sea 81-4 sequences with the corresponding regions of the Nichols isolate showed an overall similar gene organization with two significant differences: (i) the *tprJ* locus
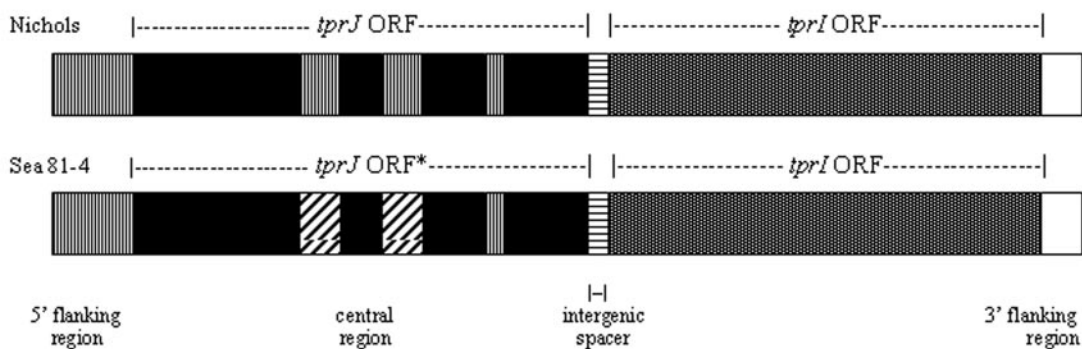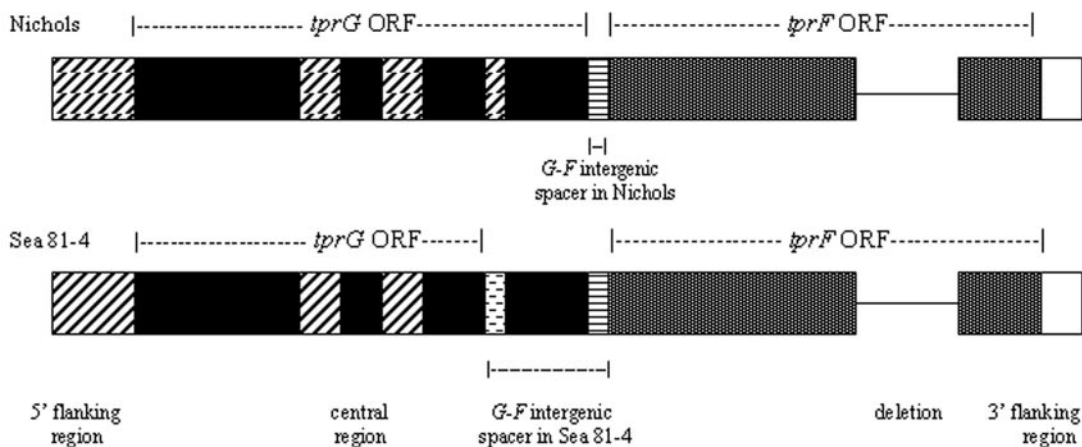
FIG. 1. Schematic representation of the *tprF*, *tprG*, *tprI*, and *tprJ* loci in the Nichols and Sea 81-4 strains. Similar patterns indicate sequence identity. The Sea 81-4 strain *tprG*-like ORF central region is identical to the Sea 81-4 and Nichols *tprG* central regions. The Sea 81-4 TprG is shorter than its homolog in Nichols due to a premature stop generating a longer *tprG-tprF* spacer. *tprI* and *tprF* loci are virtually identical in both isolates. A few base pair changes are described in the text.

FIG. 2. Alignment of the deduced protein sequences of the Sea 81-4 *tprJ* and Nichols *tprG* and *tprJ* ORFs. Identical residues are indicated by black shading, and nonsynonymous changes are indicated in gray or no shading. The *tprJ* locus in Sea 81-4 contains a *tprG*-like ORF. The Sea 81-4 TprJ central region (amino acid positions 326 to 502) has been replaced by the corresponding Nichols/Sea 81-4 TprG sequence; the COOH terminus is identical to the Nichols TprJ (amino acid positions 627 to 649).

in Sea 81-4 contains a *tprG*-like ORF, almost identical to *tprG* except for a *tprJ* nucleotide signature of 49 bases near the 3′ end (Fig. 2; see also Fig. S1A and S2 in the supplemental material); and (ii) the predicted TprG amino acid sequence in Sea 81-4 is 124 amino acids shorter than its homolog in Nichols

(632 amino acids instead of 756 amino acids), due to an extra adenosine in the 3′ end (see Fig. S1A in the supplemental material), which introduces a frameshift in the coding sequence and an earlier stop codon. Due to this C-terminal truncation of the TprG protein, the Sea 81-4 *tprG-tprF* inter-

genic spacer is 422 nucleotides long, instead of 56 bp as in Nichols (see Fig. S1B in the supplemental material). No other sequence changes were observed between the Sea 81-4 and Nichols TprG ORFs, except for two single amino acid changes upstream of the frameshift (see Fig. S3 in the supplemental material). The TprF sequences between the two strains show two amino acid changes upstream of the central nonconserved region in Sea 81-4 (see Fig. S4 in the supplemental material). A trinucleotide insertion and a single base change differentiated the *tprG* 5′-flanking regions of Sea 81-4 from the Nichols strain sequence, and a single nucleotide change was seen in the 3′-flanking region of *tprF* (see Fig. S1B in the supplemental material).

Transcriptional analysis of the *tpr* 5′-flanking and intergenic regions. RT-PCR analysis was performed to determine whether *tprJ* and *tprI* as well as *tprG* and *tprF* are cotranscribed and to roughly define the boundaries of the 5′ ends of the *tprJ* and *tprG* putative promoter regions. cDNA from both *T. pallidum* strains consistently gave PCR amplicons for the *tprJ-I* intergenic spacer with all primer sets used, indicating that these two genes are cotranscribed in both strains (Fig. 3A). In contrast, transcript spanning the *tprG-tprF* intergenic region was detected only in the Nichols strain; cDNA from the Sea 81-4 isolate gave no signal (Fig. 3B) with any of the relevant primer sets used. However, a positive RT-PCR using the F-S- and F-As-specific primers, which target a specific central region of *tprF*, indicated transcription of *tprF* and, by inference, the presence of a promoter in its 5′-flanking region able to drive transcription of this gene (Fig. 3B). Negative controls (DNase I-treated RNA and H_2O) always gave no amplification, ruling out DNA contamination.

None of our primer combinations gave a positive signal from cDNA when the regions upstream of *tprG* and *tprJ* ORFs containing the putative promoter regions were analyzed by RT-PCR. This suggests the presence of a TSS within the first 115 bp upstream of the *tprG* putative start codon in both strains and within the first 99 bp and 98 bp upstream of the predicted *tprJ* start codon in the Nichols and Sea 81-4 isolates, respectively (Fig. 3B).

Specific mRNA for *tprG* was consistently detected in cDNA from the Nichols and Sea 81-4 strains and for *tprJ* in cDNA from the Nichols isolate. No transcript was detected for *tprJ* in the Sea 81-4 strain (Table 1) due to lack of the target sequence in this locus (the *tprJ* locus in the Sea 81-4 strain contains a *tprG*-like ORF instead) or elsewhere in the chromosome. Transcript for the *tp47* gene was always present in both strains, indicating the presence of *T. pallidum* RNA in all samples. DNase I-untreated RNA and DNA always gave positive amplification. DNase I-treated RNA controls without the reverse transcription step were always negative (data not shown).

Identification of transcriptional start sites and homopolymeric repeats of variable lengths in 5′-flanking regions. To define the TSSs of the *tprJ* and *tprG* genes in the Nichols and Sea 81-4 strains, we used 5′-RACE alone on total *T. pallidum* RNA; results were then confirmed using IVT followed by 5′-RACE assays on cloned promoter regions of these genes. TSSs for *tprG* and *tprJ* were identified 9 bases upstream of the start codons described in the Nichols genome sequence (Fig. 4A and B).

We used the same approach to determine the presence of alternative TSSs upstream of *tprF* and *tprI* in both isolates. Alternative *tprF* and *tprI* TSSs were found in both strains, 29 bases upstream of their predicted start codon (Fig. 4C). Remarkably, all TSSs identified are located immediately downstream of G polynucleotide repeats.

Initial alignments of the promoter regions of all four genes in the published Nichols genome sequence revealed the presence of G polynucleotide repeats. Sequencing of the cloned amplicons from the transcriptional analysis not only confirmed the specificity of the amplifications but also revealed variability in length of the G homopolymeric tracts in clones from the Sea 81-4 and Nichols strains. The lengths of the *tprJ* and *tprG* G repeats in the Sea 81-4 strain vary from 7 to 11 nucleotides and from 8 to 11 nucleotides, respectively. In the Nichols isolate these repeats contain 9 to 11 nucleotides in both the *tprJ* and *tprG* promoter regions. The upstream regions of *tprF* and *tprI* (intergenic spacers) contain repeats that are 8 to 10 nucleotides long (*tprJ-tprI* intergenic regions) and 8 to 9 nucleotides long (*tprG-tprF* spacers) in both strains.

DISCUSSION

The organization of the *tprJ-tprI* and *tprG-tprF* loci in the Sea 81-4 strain is similar to the gene organization reported for the Nichols strain. The Sea 81-4 *tprF* and *tprI* genes show only a few single-nucleotide differences compared to their homologs in the Nichols strain and six other syphilis isolates (20), suggesting functional conservation and, although still undetermined, an important role common to all *T. pallidum* strains. The presence of the *tprG*-like ORF in the *tprJ* locus of the Sea 81-4 strain is intriguing in that this is the only syphilis isolate known to date to carry such a gene in this locus; at least seven other syphilis isolates appear to have the Nichols *tprJ* allele as determined by PCR analysis with *tprJ*-specific primers (data not shown). However, the Sea 81-4 *tprJ* ORF is nearly identical to the *tprJ* locus ORF in the Gauthier isolate of *Treponema pallidum* subsp. *pertenue* (>98% identity) (19) and the *Treponema paraluiscuniculi tprG1* and *tprG2* genes (>96% identity) (9). The high degree of sequence conservation of this *tprG*-like ORF among species suggests an ancient origin of this gene, and the presence of syphilis isolates with *tpr* genes such as the Sea 81-4 *tprG*-like ORF in other subspecies or species suggests different evolutionary pathways among syphilis strains from a common predecessor. It is unlikely that the Sea 81-4 isolate could mutate the *tprG*-like sequence to generate a Nichols *tprJ*-like sequence (the Nichols *tprG* and *tprJ* sequences differ in 605 nucleotides) unless lateral gene transfer between an isolate carrying the *tprJ* ORF and the Sea 81-4 strain occurred. An alternative explanation for this unique sequence composition may be geographical distribution, although the absence of this *tprG*-like ORF in other strains isolated from the same region (Seattle) in the same year (1981) speaks against this possibility.

Because *T. pallidum* cannot be cultured in vitro continuously, only limiting amounts of RNA can be obtained. For this reason, we chose the 5′-RACE assays alone and in combination with IVT assays to identify the TSSs of *tprF*, *tprI*, *tprG*, and *tprJ*. Sometimes, different assays may result in the identification of slightly different transcriptional start sites, usually within a few bases of each other. In our case, both approaches identi-
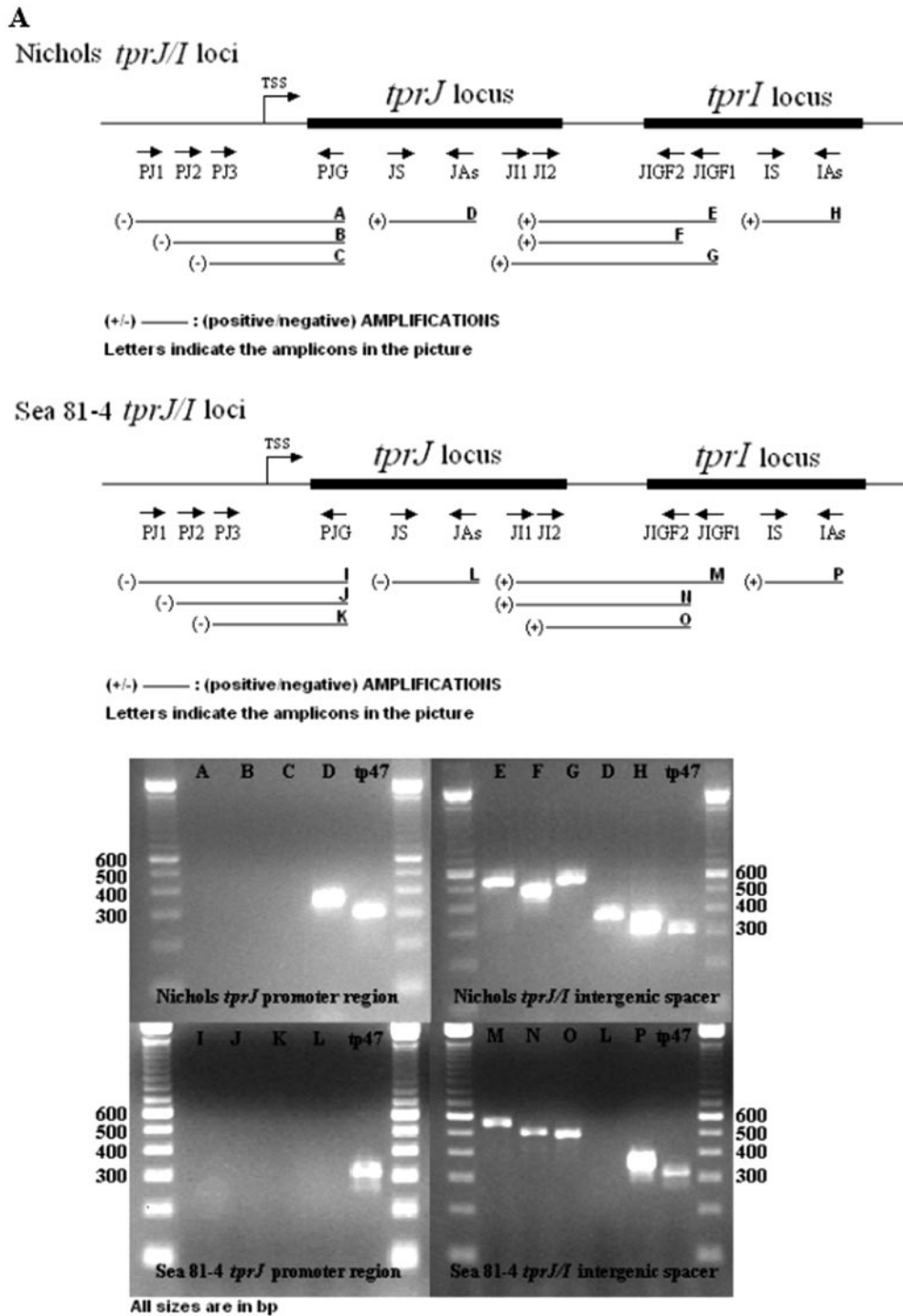
FIG. 3. Transcriptional analysis of Nichols and Sea 81-4 *tprJ-tprI* and *tprG-tprF* loci. Primer positions used for the RT-PCR analysis are schematically shown along with the amplification result (+ or −) for each reaction. Expected amplicons are named A through P. Pictures show only results obtained from cDNA. DNase I-treated RNA, DNase I-untreated RNA, and DNA were used as negative or positive controls (not shown). (A) The *tprJ-tprI* intergenic spacer is shown to be transcribed in both strains, indicating cotranscription of these two genes. No amplification of the *tprJ* 5′-flanking region was detectable in either strain, due to lack of transcript encompassing this region. (B) *tprG* and *tprF* are shown to be cotranscribed only in the Nichols strain. No amplification of the Sea 81-4 *tprG-F* intergenic spacer was seen. *tprG* 5′-flanking region amplification in both strains was also negative. In both panels A and B, lengths of ORFs, intergenic regions, flanking regions, and amplicons are not reported on scale, to facilitate comprehension of primer positions.
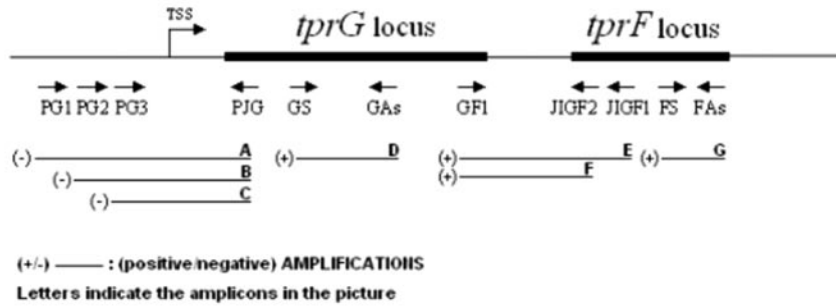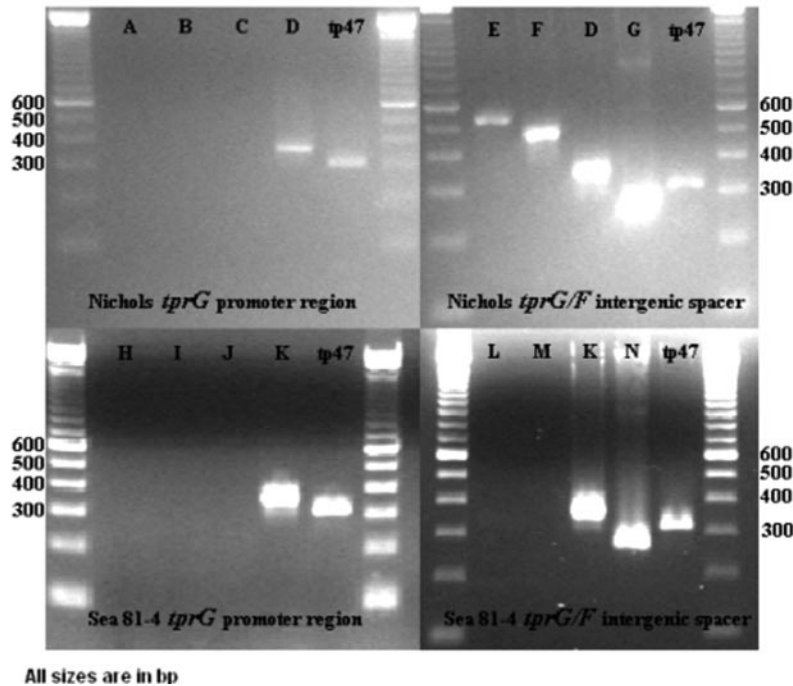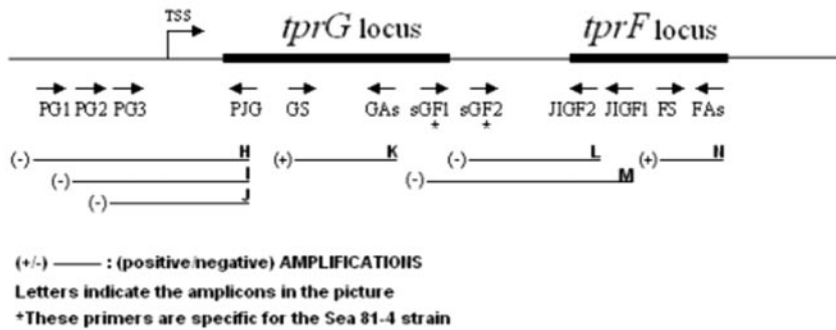
FIG. 3—*Continued.*

fied the same transcriptional starts in both the Nichols and the Sea 81-4 strains. The positions of all four TSSs are similar in that they are located immediately downstream of a G homopolymeric repeat. This unique architecture has been previ-

ously described in *T. pallidum* by Weigel et al. (24) for the *tp47* gene. The *tp47* transcriptional start site is also located immediately downstream of a G homopolymeric repeat, as determined by both primer extension and with the *E. coli* transcrip-

A. Nichols/Sea 81-4 *tprJ* promoter region *

```
                            +1
ACTAGACAGGGCCggggggggggggTGAGGTAGCATGGGGTGCATGCGGTGG
                              Met G   C   M   R   W
```

```
GggagTGTGCTGTGTGTGGTGGTGGGGGTAGgagCGAGCGGGGGAGTGCTC
 G  S  V  L  C  V  V  V  G  V  G  A  S  G  G  V  L
```

B. Nichols/Sea 81-4 *tprG* promoter region *

```
                            +1
ACTAGACAGGGCCggggggggggggTGAGGTAGCATGGGGTGCATGCGGTGG
                              Met G   C   M   R   W
```

```
GggagTGTGCTGTGTGTGGTGGTGGGGGTAggagCGAGCGGGGGAGTGCTC
 G  S  V  L  C  V  V  V  G  V  G  A  S  G  G  V  L
```

C. Nichols/Sea 81-4 *tprF/tprI* promoter regions *

```
                         +1
CGGCTAGACGGGTgggggggggggTGAGGTAGAAGTGAGAGGAGGGggagTGA
```

```
GTGGGCAGGCAGGTGATGCAAGCGGGGGTACTTGCGGGCATGGTATGTGCT
Met G  R  Q  V  M  Q  A  G  V  L  A  G  M  V  C  A
```

FIG. 4. *tprJ*, *tprG*, and *tprF* and *tprI* TSSs identified using IVT in combination with 5′-RACE and 5′-RACE applied to RNA extracted from both strains. +1, transcriptional start site; Met, translational start according to the Nichols genome sequence (8). Bold lowercase sequences represent the homopolymeric repeats. Boxed codons represent alternative predicted translational start sites. Bold italic lowercase sequences represent putative ribosomal binding sites. *, The number of Gs in the homopolymeric repeat preceding the TSSs varied in the different clones.

tion machinery in vivo. It is of interest that the *tpr* TSSs are all Ts, while in *E. coli*, the majority of experimentally demonstrated TSSs are usually Gs or As. However, several other *E. coli* genes have been also shown to have TSSs containing pyrimidines (13). To date, it is unknown what the biological significance of this phenomenon is and what proportion of TSSs in *T. pallidum* start with purines or pyrimidines. In *T. pallidum*, only a few TSSs have been deduced experimentally to date (10, 11, 24).

A predicted sigma 70 ORF is described in the published Nichols genome sequence. The ability of the 5′-flanking regions of *tprJ*, *tprI*, *tprG*, and *tprF* to drive transcription using the *E. coli* sigma 70 holoenzyme indicates the presence of sigma 70-like promoters within these regions. However, the absence of typical *E. coli* −35 and −10 consensus sequences is interesting. This finding may represent unique promoters in *T. pallidum*. Lack of −35 and −10 typical *E. coli* sigma 70 signatures has been already reported in the promoter region of the *tp47* gene of *T. pallidum* (24), indicating that this is not unique to the *tpr* genes.

RT-PCR analysis showed cotranscription of *tprG* and *tprF* as well as of *tprJ* and *tprI* in the Nichols strain. Of interest is the observation that 5′-RACE plus IVT performed on the *tprF/I* 5′-flanking region also showed the ability of these regions to drive transcription. Similarly, the transcription of *tprF* in the

Sea 81-4 strain independently from *tprG* is interesting, indicating the presence of a TSS able to generate suboperonic transcripts for *tprF/I* in both strains. Suboperonic transcription in a rather larger polycistronic transcript has been already described in bacterial organisms, including spirochetes (17, 6, 7). This takes place when one or more genes from the largest polycistronic transcript are transcribed independently from the others to fulfill a different function, providing an organism with the ability to respond to particular needs or host signals.

Homopolymeric repeats in promoter regions have been shown to regulate virulence gene expression in several bacterial pathogens. Two classic examples are the variable expression of the PorA antigen of *Neisseria meningitidis* (23) and the Vlp variable lipoproteins of *Mycoplasma hyorhinis* (5). These organisms control transcription of these genes by altering the length of the poly(G) and poly(A) repeats in their promoters, respectively. The homopolymeric G repeats located immediately upstream of the TSSs of the *tprF*, *tprI*, *tprG*, and *tprJ* loci vary in length in individual bacteria and between the Nichols and Sea 81-4 strains in DNA and mRNA samples and are likely to be involved in regulation of the transcription levels of these genes by modulating promoter strength and inducing differential expression of these genes among *T. pallidum* strains over the course of infection. The difference in the number of G residues is not due to PCR or sequencing artifacts. This is

supported by the fact that homopolymeric repeats in the promoter region of the Nichols *tprK* (another *tpr* homolog) always show the same number of nucleotides in clones obtained from different DNA batches, PCRs, and sequencing reactions.

There are several issues that still remain unanswered, including promoter strength, levels of transcription over time, transcript length, correlation between levels of mRNA and protein synthesis, and patterns of expression in response to microenvironmental changes during infection. Nevertheless, the differences in sequence organization and cotranscription in these paralogous genes in the Nichols and Sea 81-4 strains suggest that they may be useful models for the study of gene regulation in this bacterium.

REFERENCES

1. **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.
2. **Centurion-Lara, A., C. Castro, R. Castillo, J. M. Shaffer, W. C. Van Voorhis, and S. A. Lukehart.** 1998. The flanking region sequences of the 15-kDa lipoprotein gene differentiate pathogenic treponemes. J. Infect. Dis. **177:**1036–1040.
3. **Centurion-Lara, A., C. Castro, W. C. Van Voorhis, and S. A. Lukehart.** 1996. Two 16S-23S ribosomal DNA intergenic regions in different *Treponema pallidum* subspecies contain tRNA genes. FEMS Microbiol. Lett. **143:**235–240.
4. **Choy, H. E., and S. Adhya.** 1992. Control of gal transcription through DNA looping: inhibition of the initial transcribing complex. Proc. Natl. Acad. Sci. USA **89:**11264–11268.
5. **Citti, C., and K. S. Wise.** 1995. *Mycoplasma hyorhinis vlp* gene transcription: critical role in phase variation and expression of surface lipoproteins. Mol. Microbiol. **18:**649–660.
6. **Cunningham, L., and J. R. Guest.** 1998. Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. Microbiology **144:**2113–2123.
7. **Dobrikova, E. Y., J. Bugrysheva, and F. C. Cabello.** 2001. Two independent transcriptional units control the complex and simultaneous expression of the *bmp* paralogous chromosomal gene family in *Borrelia burgdorferi*. Mol. Microbiol. **39:**370–378.
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, J. C. Venter, et al. 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. **Hazlett, K. R. O., T. J. Sellati, T. T. Nguyen, D. L. Cox, M. L. Clawson, M. J. Caimano, and J. D. Radolf.** 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.
11. **Isaacs, R. D., and J. D. Radolf.** 1990. Expression in *Escherichia coli* of the 37-kilodalton endoflagellar sheath protein of *Treponema pallidum* by use of the polymerase chain reaction and a T7 expression system. Infect. Immun. **58:**2025–2034.
12. **Leader, B. T., K. Hevner, B. J. Molini, L. K. Barrett, W. C. Van Voorhis, and S. A. Lukehart.** 2003. Antibody responses elicited against the *Treponema pallidum* repeat proteins differ during infection with different isolates of *Treponema pallidum* subsp. *pallidum*. Infect. Immun. **71:**6054–6057.
13. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:**1507–1516.
14. **Lukehart, S. A., S. A. Baker-Zander, and S. Sell.** 1980. Characterization of lymphocyte responsiveness in early experimental syphilis. I. In vitro response to mitogens and *Treponema pallidum* antigens. J. Immunol. **124:**454–460.
15. **Maniatis, T., J. Sambrook, and E. F. Fritsch.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
16. **Nichols, H. J., and W. H. Hough.** 1913. Demonstration of *Spirochaeta pallida* in the cerebrospinal fluid. JAMA **60:**108–110.
17. **Schmid, M. B., and J. R. Roth.** 1983. Internal promoters of the *his* operon in *Salmonella typhimurium*. J. Bacteriol. **153:**1114–1119.
18. **Smajs, 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.
19. **Stamm, L. V., S. R. Greene, H. L. Bergen, J. M. Hardham, and N. Y. Barnes.** 1998. Identification and sequence analysis of *Treponema pallidum tprJ*, a member of a polymorphic multigene family. FEMS Microbiol. Lett. **169:**155–163.
20. **Sun, E. S., B. J. Molini, L. K. Barrett, A. Centurion-Lara, S. A. Lukehart, and W. C. Van Voorhis.** 2004. Subfamily I *Treponema pallidum* repeat protein family: sequence variation and immunity. Microbes Infect. **6:**725–737.
21. **Tantalo, L. C., S. A. Lukehart, and C. M. Marra.** 2005. *Treponema pallidum* strain-specific differences in neuroinvasion and clinical phenotype in a rabbit model. Infect. Dis. **191:**75–80.
22. **Turner, T. B., and D. H. Hollander.** 1957. Biology of the treponematoses. World Health Organization, Geneva, Switzerland.
23. **van der Ende, A., C. T. Hopman, S. Zaat, B. B. Essink, B. Berkhout, and J. Dankert.** 1995. Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the −10 and −35 regions of the promoter. J. Bacteriol. **177:**2475–2480.
24. **Weigel, L. M., M. E. Brandt, and M. V. Norgard.** 1992. Analysis of the N-terminal region of the 47-kilodalton integral membrane lipoprotein of *Treponema pallidum*. Infect. Immun. **60:**1568–1576.