Transcription of TP0126, *Treponema pallidum* Putative OmpW Homolog, Is Regulated by the Length of a Homopolymeric Guanosine Repeat

Lorenzo Giacani,[a,b] Stephanie L. Brandt,[a*] Wujian Ke,[a,c] Tara B. Reid,[a*] Barbara J. Molini,[a] Stefanie Iverson-Cabral,[a] Giulia Ciccarese,[d] Francesco Drago,[d] Sheila A. Lukehart,[a,b] Arturo Centurion-Lara[a*]

Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington, USA[a]; Department of Global Health, University of Washington, Seattle, Washington, USA[b]; Graduate School, Southern Medical University, Guangzhou, People's Republic of China, and Division of STD, Guangdong Provincial Center for STI & Skin Diseases Control and Prevention, Guangzhou, People's Republic of China[c]; Department of Dermatology, IRCCS Azienda Universitaria Ospedaliera San Martino-IST, Genoa, Italy[d]

An effective mechanism for introduction of phenotypic diversity within a bacterial population exploits changes in the length of repetitive DNA elements located within gene promoters. This phenomenon, known as phase variation, causes rapid activation or silencing of gene expression and fosters bacterial adaptation to new or changing environments. Phase variation often occurs in surface-exposed proteins, and in *Treponema pallidum* subsp. *pallidum*, the syphilis agent, it was reported to affect transcription of three putative outer membrane protein (OMP)-encoding genes. When the *T. pallidum* subsp. *pallidum* Nichols strain genome was initially annotated, the TP0126 open reading frame was predicted to include a poly(G) tract and did not appear to have a predicted signal sequence that might suggest the possibility of its being an OMP. Here we show that the initial annotation was incorrect, that this poly(G) is instead located within the TP0126 promoter, and that it varies in length *in vivo* during experimental syphilis. Additionally, we show that TP0126 transcription is affected by changes in the poly(G) length consistent with regulation by phase variation. *In silico* analysis of the TP0126 open reading frame based on the experimentally identified transcriptional start site shortens this hypothetical protein by 69 amino acids, reveals a predicted cleavable signal peptide, and suggests structural homology with the OmpW family of porins. Circular dichroism of recombinant TP0126 supports structural homology to OmpW. Together with the evidence that TP0126 is fully conserved among *T. pallidum* subspecies and strains, these data suggest an important role for TP0126 in *T. pallidum* biology and syphilis pathogenesis.

S yphilis is a chronic sexually transmitted infection caused by the spirochete *Treponema pallidum* subsp. *pallidum*, a Gram-negative obligate human pathogen. Despite being perceived by the general public to be a disease of the past, syphilis continues to be a significant burden for global health, with an estimated prevalence of 36 million cases worldwide and an incidence that exceeds 11 million new cases every year (1). Although most syphilis cases occur in Latin America, sub-Saharan Africa, and Southeast Asia (2), including China (3), a worrying resurgence of syphilis has recently been observed in many developed countries, including the United States, Canada, Australia, and several European nations (4–8). Congenital syphilis remains a leading cause of stillbirths and perinatal deaths among neonates in developing areas (9, 10), and patients with syphilis are at increased risk for transmission and acquisition of HIV (11, 12).

T. pallidum subsp. *pallidum* is an extremely successful pathogen, characterized by the ability to invade and infect virtually any organ and elude the host immune response, resulting in pathogen persistence for decades in infected individuals in the absence of treatment. Phase variation is one of the means evolved by pathogenic bacteria to rapidly create phenotypic diversity within a population. When this process influences expression of surface antigens, for example, it can facilitate evasion of the host immune response (13), influence the affinity of a pathogen for different host anatomical niches, or foster adaptation to diverse or rapidly changing microenvironments (14, 15). Variable expression of opacity (Opa) proteins in *Neisseria meningitidis*, for example, is reported to change the pathogen's tropism for human epithelium,

endothelium, and phagocytic cells (16). At the genetic level, one of the mechanisms responsible for phase variation is the expansion and contraction of repetitive nucleotide sequences, such as homopolymeric tracts [e.g., poly(G) or poly(C) repeats], a phenomenon resulting from slipped-strand mispairing of DNA strands during replication (17, 18). When such homopolymeric repeats of various lengths are located in a gene promoter between the -35 and -10 binding sites for the σ subunit of the RNA polymerase (RNAP) holoenzyme (14, 19), they can modulate promoter activity by increasing or reducing the relative distance between the -35

TABLE 1 Treponemal strains used in this study

Species	Strain name	Source	Location	Yr of isolation
T. pallidum subsp. *pallidum*	Nichols (Seattle and Houston lineages)[a]	Cerebrospinal fluid	Washington, DC	1912
	Sea81-4[b]	Primary chancre	Seattle, WA	1980
	Bal3[c]	Blood, congenital	Baltimore, MD	Circa 1973
	MexicoA[c]	Primary chancre	Mexico	1953
	Chicago[c]	Primary chancre	Chicago, IL	1951
T. pallidum subsp. *pertenue*	SamoaD[c]	Skin	Western Samoa	1953
T. pallidum subsp. *endemicum*	BosniaA[c]	Penile ulcer	Bosnia	1950
	IraqB[c]	Oral mucous patches	Iraq	1951
T. paraluiscuniculi	CuniculiA[c]	Unknown	Baltimore, MD	Unknown

[a] Originally provided by James N. Miller, University of California, Los Angeles, CA, to Sheila Lukehart (Seattle lineage) or to Steven J. Norris (Houston lineage). Nichols Houston was kindly provided by Steven J. Norris and was used to derive the Nichols Houston O clonal isolate.
[b] Strain isolated in Seattle, WA, by Sheila A. Lukehart, University of Washington, Seattle, WA.
[c] Strains provided by Paul Hardy and Ellen Nell, Johns Hopkins University, Baltimore, MD.

and −10 binding sites. Because of this accordion-like mechanism, gene transcription occurs when the distance that separates the −35 and −10 binding sites is close to 17 nucleotides (nt), which is optimal for RNAP binding. Vice versa, a nucleotide spacer between the −35 and −10 sequences that greatly exceeds or falls below 17 nt is associated with a reduction in the level of transcription or gene silencing.

Antigenic variation is one of the specialized mechanisms evolved by the syphilis agent to foster immune evasion and persistence in the host (20–22). We have previously shown that transcription of the *T. pallidum tprF*, *tprI*, *tprE*, and *tprJ* genes is modulated by promoter-associated poly(G) sequences of various lengths (23) and may lead to phase variation in these antigens. In addition, the abundance of similar homopolymeric tracts in the *T. pallidum* genome suggested that the use of these elements to modulate gene transcription might be widespread in this pathogen. In this study, we focused on the poly(G) tract associated with the *T. pallidum* TP0126 gene, originally annotated as residing within the coding sequence for a hypothetical protein of unknown function (24). Our data show that the TP0126-associated poly(G) is located within the gene promoter and that this element varies in length *in vivo* during experimental infection in the rabbit model of syphilis and is involved in transcriptional regulation of TP0126 with a mechanism consistent with phase variation. In addition, experimental identification of the TP0126 transcriptional start site (TSS) supports the suggestion that the open reading frame (ORF) is 69 amino acids (aa) shorter than originally annotated and that the encoded TP0126 protein harbors an NH_2-terminal cleavable signal peptide commonly employed by Gram-negative bacteria to sort surface-exposed antigen. Evidence that phase variation is often (although not exclusively) reported to affect expression of surface antigens in bacterial pathogens (14, 19) prompted us to begin investigating whether TP0126 could be a newly identified putative surface antigen. We provide *in silico* evidence that identifies the shorter TP0126 ORF to be a putative homolog of OmpW, an outer membrane porin likely involved in transporting hydrophobic molecules into the outer membrane (25, 26). Circular dichroism (CD) analysis of recombinant TP0126 showed a β-sheet component compatible with structural homology between TP0126 and OmpW. Altogether, these findings suggest that regulation by phase variation is more widespread in *T. pallidum* than currently

reported and that TP0126's possible localization on the *T. pallidum* surface and its role in the biology of this spirochete and syphilis pathogenesis are worthy of further investigation.

MATERIALS AND METHODS

Ethics statement. New Zealand White rabbits were used for treponemal strain propagation and experimental infections. Animal care was provided in full accordance with the guidelines in the *Guide for the Care and Use of Laboratory Animals* (27), and experimental procedures were conducted under protocols approved by the University of Washington Institutional Animal Care and Use Committee (IACUC). The protocol number assigned by the IACUC committee that approved this study is 4243.01. Deidentified sera from human syphilis patients were used in this study, and this research has been determined by the University of Washington Human Subjects Division not to meet the federal regulatory definition of human subjects research.

Strain propagation, clonal isolate derivation, sample collection, and nucleic acid extraction. Five *T. pallidum* subsp. *pallidum* strains (Bal3, Chicago, Nichols Seattle, MexicoA, and Sea81-4), two *T. pallidum* subsp. *endemicum* strains (BosniaA and IraqB), *T. pallidum* subsp. *pertenue* strain SamoaD, and *Treponema paraluiscuniculi* strain CuniculiA were propagated in New Zealand White rabbits by means of intratesticular (i.t.) inoculation as previously described (28). Briefly, spirochetes were extracted from infected rabbit testicles at peak orchitis following injection of a maximum of 5×10^7 bacterial cells per testicle and placed in sterile saline. Treponemes were collected in sterile 15-ml tubes, taking the necessary precautions to avoid cross-contamination between samples in case of contemporary harvests. Bacteria were then immediately spun in an Eppendorf 5810R centrifuge (Eppendorf, Hauppauge, NY) at 1,000 rpm (equivalent to ~250 × *g*) for 10 min to remove rabbit tissue debris, followed by transfer of the supernatant to 1.7-ml sterile microcentrifuge tubes and centrifugation at 12,000 × *g* for 30 min at 4°C (28) to pellet the treponemes. The pellets were subsequently resuspended in 200 μl of 1× lysis buffer (10 mM Tris, pH 8.0, 0.1 M EDTA, 0.5% sodium dodecyl sulfate [SDS]) if they were meant to be used for DNA isolation or in 400 μl of phenol-based UltraSpec buffer (Biotex Inc., TX) prior to total RNA isolation. Table 1 reports the different geographical regions, years, and anatomical sources of isolation of the strains studied here.

A clonal Nichols strain of *T. pallidum* subsp. *pallidum* (Nichols Houston O) was obtained as previously described (28, 29). This technique, originally developed to study the antigenic variation of the TprK protein, was previously shown to be an effective method to obtain treponemal isolates isogenic for the *tprK* gene from a heterogeneous population of cells expressing different TprK variants (22, 29–31). Briefly, to obtain the

Nichols Houston O strain, a naive rabbit was injected intravenously (i.v.) with 10^8 *T. pallidum* subsp. *pallidum* (Nichols Houston strain) cells through the marginal ear vein, and treponemes were allowed to disseminate and form isolated skin lesions visible on the rabbit shaven back. Clonality is achieved because each discrete skin lesion is believed to be seeded by a single treponemal cell that carries a unique *tprK* sequence (29, 30, 32). Biopsy specimens of isolated skin lesions were minced in 1 ml of normal rabbit serum (NRS). Following homogenization, approximately half of the treponemal suspension obtained was injected into a naive recipient rabbit to propagate the clone. After multiplication of the clonal isolate within the rabbit, treponemes were harvested again and used to infect three naive rabbits intradermally (i.d.; 10^6 *T. pallidum* cells per site) in multiple sites on their shaven backs. The clonality of the treponemal inoculum was assessed by sequencing as previously reported (32) and by fluorescent fragment length analysis (FLA; see below) of the TP0126-associated poly(G) tract. Biopsy specimens from lesions appearing at the intradermal injection sites were collected from all infected rabbits weekly for 3 weeks and minced in 1× lysis buffer for DNA extraction to evaluate the variation of the poly(G) tract upstream of TP0126 using FLA.

DNA was extracted using a DNA minikit (Qiagen Inc., Chatsworth, CA) according to the manufacturer's instructions, with the exception that 50 μl of proteinase K (from a 100-mg/ml stock solution) was added and the sample was incubated overnight at 56°C. After the final elution step in 200 μl of molecular-grade H_2O, DNA was stored at −20°C until needed for amplification. The protocols used for isolation of total RNA from the samples in UltraSpec buffer and DNase I treatment to obtain DNA-free RNA samples prior to reverse transcription were previously reported in detail (33).

Comparative study of TP0126 sequence and fluorescent fragment length analysis of the TP0126-associated poly(G) repeat. The full TP0126 ORF (as originally annotated in the Nichols genome) (24) was amplified using DNA extracted from all isolates studied here (with the exception of the Nichols Houston lineage, whose genome was sequenced twice and whose TP0126 sequence is already available [24, 34]) and sequenced to evaluate genetic diversity. Primers and amplicon sizes are listed in Table 2. For the TP0126 full-length ORF plus 204 and 106 nucleotides in the 5′ and 3′ flanking regions, respectively, amplifications were performed in a 50-μl final volume using 2 U of AccuPrime *Pfx* polymerase (Life Technologies, Carlsbad, CA) and 100 ng of DNA template in each reaction mixture. The mix was also supplied with primers, $MgSO_4$, and deoxynucleoside triphosphates (dNTPs) at final concentrations of 300 nM each, 1 mM, and 300 μM, respectively. Amplifications were carried out for 45 cycles with denaturation (94°C), annealing (60°C), and extension (68°C) times of 30 s, 30 s, and 1 min, respectively. The initial denaturation (94°C) and final extension (68°C) steps were 10 min each. For each strain, two independent amplifications were performed using the same template DNA. Subsequently, amplicons were cleaned of unincorporated primers and dNTPs using the ExoSAP-IT reagent (Affymetrix, Santa Clara, CA). Following spectrophotometric quantification using an ND-1000 instrument (NanoDrop Technologies, Wilmington, DE), 40 ng of template was mixed with 25 pmol of one (of four) sequencing primers (Table 2) designed to ensure complete coverage of both strands. Sanger sequencing was outsourced (Genewiz, Seattle, WA), but the results were analyzed in the laboratory using Bioedit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html). A smaller DNA fragment containing the TP0126-associated poly(G) repeat was also amplified for FLA. Amplification was performed using a 6-carboxyfluorescein (FAM)-labeled sense primer and an unlabeled antisense primer. A pig-tailed sequence (GTTTCTT; Table 2) was added to the 5′ end of the antisense primer to ensure uniform addition of nontemplated A overhangs to all amplicons (35). Amplifications were performed in a 50-μl final volume using 0.2 unit of GoTaq polymerase (Promega Inc., Madison, WI) with 100 ng of DNA template in each reaction mixture and primers, $MgCl_2$, and dNTPs at final concentrations of 200 nM, 1.5 mM, and 200 μM, respectively. Initial denaturation (94°C) was for 10 min, while final extension (72°C) was for

30 min. Denaturation (94°C), annealing (60°C), and extension (72°C) were carried out for 30 s each for a total of 45 cycles. Amplification products were purified using a QIAquick PCR purification kit (Qiagen). Concentrations were measured spectrophotometrically, and all samples were diluted to a 0.2-ng/μl final concentration. One microliter of each sample was mixed with 15.4 μl of Hi-Di formamide and 0.1 μl of an HD400 carboxy-X-rhodamine (ROX)-labeled DNA size marker (both reagents were from Life Technologies). Samples were transferred to a 96-well plate and denatured by incubation at 94°C for 2 min, chilled on ice for 1 min, and loaded on an ABI 3730xl DNA analyzer (Life Technologies) for separation by capillary electrophoresis. Electropherograms were analyzed using the GeneMapper (version 4.0) software package (Life Technologies); data on amplicon length (determined by comparison to the length of the ROX-labeled marker) and intensity (measured as the area under a peak) were collected to evaluate the relative amounts of amplicons with poly(G) regions of different lengths within each sample. For each strain, FLA was performed in triplicate using three independent amplification products with the same template DNA. For data analysis, the sum of the area underneath all peaks generated by amplicons with the same number of G residues was divided by the total area underneath all peaks. Data analysis was performed using analysis of variance (ANOVA), with significance being set at a P value of <0.05, and the Bonferroni correction was applied for multiple comparisons. Analysis was performed using Prism software (version 5; GraphPad Software, La Jolla, CA). Amplification of the *tprK* gene to assess the clonality of the Nichols Houston O strain was performed as reported previously (36). The primers used are listed in Table 2.

TP0126 transcriptional start site identification. A 5′ rapid amplification of cDNA ends (RACE) kit (Life Technologies) was used to determine the TP0126 gene TSS (position +1) and, hence, the location of the −10 and −35 sequences of the TP0126 promoter. 5′ RACE was performed on total RNA from the *T. pallidum* subsp. *pallidum* Nichols Seattle and Chicago strains following the kit manufacturer's instructions. For each strain the procedure was carried on in duplicate using the same template RNA. Briefly, for the initial reverse transcription step, 1 μg of sample RNA and 2.5 pmol of a first TP0126-specific antisense primer were used (Table 2). Subsequent to reverse transcription, 1 μl of an RNase H-T1 mix was added to the tube and the cDNA was incubated for 20 min at 37°C prior to cDNA purification using the columns and buffers provided with the 5′ RACE kit. Prior to amplification, dC tailing of purified cDNA was performed according to the instructions provided with the kit. All PCRs were performed using 5 μl of dC-tailed cDNA in a 50-μl final volume containing 2.5 units of GoTaq polymerase (Promega), 200 μM each dNTP, 1.5 mM $MgCl_2$, and 400 nM each primer (a second TP0126-specific antisense primer annealing upstream of the one used for first-strand synthesis and the abridged anchor primer [AAP] provided with the kit; Table 2). It was not necessary to perform a nested amplification step. Cycling parameters included initial denaturation (94°C) and final extension (72°C) for 10 min each. Denaturation (94°C), annealing (60°C), and extension (72°C) steps were carried out for 1 min each for a total of 45 cycles. PCR products were purified from agarose gels using the QIAquick gel extraction kit (Qiagen) and cloned into the pCRII-TOPO TA vector (Life Technologies) according to the instructions of the manufacturer. Plasmid DNA was extracted from at least 10 insert-containing colonies per cloning reaction using a plasmid minikit (Qiagen) and sequenced with vector-specific sense and antisense primers. Sequence data were analyzed using the Bioedit software. Following TSS identification, the region upstream of the +1 nucleotide was analyzed using the BProm program, a bacterial promoter recognition program that uses an algorithm trained with recognized σ^{70} signatures, to confirm the initial predictions of the location of the −35 and −10 sequences and the likelihood that they are binding sites for the σ^{70} subunit of the RNA polymerase holoenzyme.

TP0126 promoter-GFP reporter assay. TP0126 promoters containing poly(G) sequences of different lengths were available from a plasmid library previously generated by amplifying, cloning, and sequencing the TP0126 region containing the poly(G) repeat from syphilis-causing

TABLE 2 Primers used in this study

Expt	Primer name[a]	Sequence (5′ → 3′)	Amplicon size (bp)
TP0126 amplification and sequencing for analysis of genomic diversity among strains	TP0126-S1	ATGCGGACGCACGACATACC	924[b]
	TP0126-As1	GAGCGGCACGTCCGCGCAA	
	TP0126-S2	CGCGGACCCGTGGGACACT	NA[c]
	TP0126-S3	GGCTGGCTTCAATTATCAGC	NA
	TP0126-As2	TACCGTCAGTGGCAAGCGGA	NA
	TP0126-As3	CGCAGTAGTGTCCCACGGGT	NA
FLA to evaluate poly(G) variability and TP0126 5′ RACE to determine TP0126 TSS	FAM-TP0126-S	FAM-ACGACATACCCCGCAGTC	275
	Pig-TP0126-As	GTTTCTTGTGCAGCTCCCCACACTC	
	RT-TP0126-As	GCCTTGAGCACAAGACCGTA	NA
	RACE-TP0126-As	GTGCAGCTCCCCACACTCCCG	154
	Abridged anchor primer (AAP)	GGCCACGCGTCGACTAGTAC(G)$_{14}$	
Expression of recombinant TP0126 and TP0493 (σ^{70}) proteins for ELISA and IVT assay	NT-TP0126-S	TGGGACACTACTGCGGCGG	672
	NT-TP0126-As	TCAGAGGTGGTAGCGAACGT	
	CT-TP0493-S	ATGATGGAGCTGTCACGTAC	1,833
	CT-TP0493-As	ACTGTCAAGGAAATCCTTCA	
	TP0493-S2	ATCTGCGTGATATCGGCCAA	NA
	TP0493-S3	AAAAGCAACTTATGATTGCT	NA
	TP0493-S4	ATAGAGCAGATAAATAAAGTG	NA
	TP0493-As2	TATCTGCTCTATCATGTGTAC	NA
	TP0493-As3	ATCTATCTCCTCTAATTGCA	NA
	TP0493-As4	TTCCGCGCGTTTGACAGCAA	NA
Real-time quantification of TP0126, GFP, and TP0574 messages	GFPrt-S	CTACCTGTTCCATGGCCAAC	263
	GFPrt-As	TGTGTCCGAGAATGTTTCCAT	
	TP0126rt-S	CGCAAGTGTGCAGTACCATT	201
	TP0126rt-As	GCCTTGAGCACAAGACCGTA	
	TP0574rt-S	CGTGTGGTATCAACTATGG	313
	TP0574rt-AS	TCAACCGTGTACTCAGTGC	
GFP transcription assay (in *E. coli* and *in vitro*) to evaluate poly(G) effect on TP0126 transcription	Prom-TP0126-S	GGTTCGACAAACAGCCGTGG	85–89[d]
	Prom-TP0126-As	CATACCCCGCTCCAACGCCG	
	Nopro-TP0574-S	CGTGTGGTATCAACTATGG	313
	Nopro-TP0574-As	TCAACCGTGTACTCAGTGC	
	Lac-S	CGCAACGCAATTAATGTGAGT	124
	Lac-As	CATAGCTGTTTCCTGTGTGAA	
Amplification of a TP0126 promoter with 7 G residues in poly(G) tract	TP0126-7G-PromS	GGTTCGACAAACAGCCGTGGGGGGGGTAGA	84
	Prom-TP0126-As	CATACCCCGCTCCAACGCCG	
tprK amplification	tprK-S	TCCCCCAGTTGCAGCACTAT	2,133
	tprK-As	TCGCGGTAGTCAACAATACCA	

[a] RT-TP0126-As and TP0126rt-AS are identical. TP0574rt-S is identical to Nopro-TP0574-S, and TP0574rt-As is identical to Nopro-TP0574-As.
[b] With a poly(G) sequence of 9 G repeats.
[c] NA, not applicable, which indicates sequencing primers or primers used for specific reverse transcription reactions (RT-TP0126-As).
[d] Depending on the number of G residues.

strains. From this library, TP0126 promoter regions containing poly(G) regions of 8 to 12 nucleotides were amplified again with primers (Table 2) designed to place the green fluorescent protein (GFP) reporter gene of the pGlow-TOPO TA vector (Life Technologies) under the control of the experimentally identified TP0126 promoter and fuse the TP0126 ATG with the GFP ORF. No promoter regions with 7 or fewer G residues in the poly(G) tract were found in the library. Thus, to obtain a TP0126 promoter with 7 G residues, a primer (Table 2) was designed to obtain an amplicon with the poly(G) tract of the desired length so that it could be cloned into the pGlow-TOPO TA vector. Nichols Seattle DNA was used as the template for this amplification. The chemistry of the amplification was identical to that reported for the TP0126 full-length ORF (see above). Initial denaturation (94°C) and final extension (72°C) were 10 min each,

while denaturation (94°C), annealing (60°C), and extension (72°C) were carried out for 30 s each for a total of 45 cycles. Amplicons included 18 nt upstream of the poly(G) tract, predicted to harbor the −35 sequence of the TP0126 promoter according to the 5′ RACE and BProm analyses, and 59 nt downstream of the poly(G), predicted to contain the −10 consensus sequence and a putative ribosome binding site (RBS; GGAG) located 6 nt upstream of the TP0126 ATG. Amplicons were gel purified to eliminate the plasmid template and cloned into the pGlow-TOPO TA vector according to the manufacturer's instructions. With the exception of the start codon, no other TP0126 codons were present in the constructs. Expression of GFP from these constructs resulted in the addition of 9 extra amino acids to the actual GFP peptide, encoded by TP0126 ATG, and 8 additional codons already present in the vector sequence. In total, six

different constructs were obtained for the TP0126 promoter, with the poly(G) repeats in these constructs being 7 to 12 nt long. A construct containing the *lac* promoter upstream of the GFP gene was used as a positive control (the primers are listed in Table 2). As a negative control to determine background fluorescence, the TP0574 ORF fragment amplified for message quantification purposes (see below for a description of the quantification of the TP0126 message) and not predicted to harbor a promoter or a ribosome binding site was inserted upstream of the GFP-coding sequence in the pGlow-TOPO TA vector. All constructs were sequenced on both strands to verify sequence accuracy and the correct insert orientation. The constructs were then used to transform *Escherichia coli* TOP10 cells (Life Technologies), which do not carry the *lacI* repressor gene.

For GFP fluorescence measurements, cells transformed with the various constructs were inoculated from a plate into 4 ml of LB-ampicillin (100 μg/ml) broth and grown at 37°C for 4 h. The optical density at 600 nm (OD_{600}) of all cultures was then measured using a biophotometer (Eppendorf), and cultures were diluted to identical optical densities (0.5 absorbance units [AU]). Subsequently, the OD_{600} and fluorescence were recorded in parallel until the cultures reached an OD_{600} of 2 AU. For fluorescence readings, 400 μl of culture was spun for 4 min at 12,000 × *g* and resuspended in an equal volume of phosphate-buffered saline (PBS). Cells were then divided into four wells (100 μl/well) of a black OptiPlate-96F plate (PerkinElmer, Boston, MA) for top fluorescence reading. The excitation and emission wavelengths were 405 and 505 nm, respectively, and readings were performed in a Fusion universal microplate analyzer (Packard, Meriden, CT). The reported data represent the fluorescence (expressed in arbitrary units) normalized to the optical density of the culture. Background values obtained from each experiment (using *E. coli* cells transformed with the reporter vector containing the TP0574 ORF fragment) were subtracted from the sample values. Differences in the levels of fluorescence between cultures were compared using ANOVA, with significance being set at a *P* value of <0.05, and the Bonferroni correction was applied for multiple comparisons. The experiment was repeated twice to ensure reproducibility.

TP0126 IVT assay. An *in vitro* transcription (IVT) assay was developed to further confirm the results of the GFP reporter assay and to investigate whether the *T. pallidum* σ⁷⁰ factor initiates TP0126 transcription, as predicted by BProm. To obtain templates for the IVT assay, DNA fragments containing variants of the TP0126 promoter [with 7- to 12-nt-long poly(G) regions] followed by the GFP gene were excised from the pGlow-TOPO TA vectors used for the GFP reporter assay (see above). Each of the vectors was incubated with 10 U of ZraI overnight at 37°C, purified using the QIAquick PCR purification kit (Qiagen), and then incubated with the NotI restriction enzyme (both enzymes were from New England Biolabs, Ipswich, MA) overnight at 37°C. As positive and negative controls, respectively, a *lac* promoter-GFP fragment and the TP0574-GFP fragment (the no-promoter control) were also excised. Excised fragments were separated from the vector through gel electrophoresis, purified using the QIAquick gel extraction kit (Qiagen), and quantitated spectrophotometrically. Additional reagents included *T. pallidum* recombinant σ⁷⁰ (see below for the expression protocol), *E. coli* core RNAP, and *E. coli* RNAP-σ⁷⁰ holoenzyme (both the core RNAP and the RNAP-σ⁷⁰ holoenzyme were from Epicenter, Madison, WI). For the actual assay, performed in a 100-μl final volume, 10 pmol of each DNA template was mixed with 20 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), 10 μl of 10× nucleoside triphosphate mix (20 mM [each] ATP, CTP, GTP, and UTP), and 5 μl of 0.1 U/μl of inorganic pyrophosphatase (New England BioLabs) to prevent precipitation of inorganic pyrophosphate. With regard to the transcriptional machinery (core RNAP and σ subunit), each template DNA was tested under four different conditions: (i) with no RNAP in the presence of *T. pallidum* σ⁷⁰ (negative control), (ii) with *E. coli* core RNAP alone with no σ factors (to determine the level of nonspecific transcription), (iii) with *E. coli* core RNAP plus *E. coli* σ⁷⁰, and (iv)

with *E. coli* core RNAP plus *T. pallidum* σ⁷⁰. The reaction mixtures were incubated at 37°C for 2 h. Following incubation, the products were briefly centrifuged and treated with 10 U of Turbo DNase I (Life Technologies) according to the protocol provided by the manufacturer. Incubation was prolonged to a total of 1 h prior to termination using the inactivation reagent provided with the kit. The IVT assay products were then purified by phenol-chloroform according to standard protocols (37) and reverse transcribed using a SuperScript III first-strand synthesis kit (Life Technologies) and a GFP-specific antisense primer according to the manufacturer's instructions. Reverse transcription products were analyzed by quantitative PCR (qPCR) using GFP-specific primers (Table 2) annealing downstream of the one used for reverse transcription. Amplification and data collection were carried on using a ViiA-7 real-time PCR system and Power SYBR green PCR master mix (Life Technologies). A GFP standard was created as described below for the TP0126 standard. Triplicate amplifications were performed for each IVT assay using 3 μl of cDNA. Primer concentration and cycling conditions were as described above for the TP0126 and TP0574 genes. Results were analyzed using ViiA-7 real-time PCR software and reported as the number of copies of the GFP message per microliter of sample. Significance was evaluated using ANOVA, and the Bonferroni correction was applied for multiple comparisons. Significance was set at a *P* value of <0.05.

Quantification of TP0126 message. The TP0126 message levels at peak orchitis, the time when bacterial harvest occurred, were analyzed using a relative quantification protocol with external standards. This approach normalizes the amount of message from a target gene to the amount of mRNA of a reference gene present in the same sample (in this case, TP0574, encoding the *T. pallidum* 47-kDa lipoprotein, an antigen conserved in all treponemal species, subspecies, and strains used in this study). To obtain the TP0126 standards, a fragment of the TP0126 ORF of 201 nt was amplified from Nichols Seattle DNA and cloned into the pCRII-TOPO TA vector (Life Technologies). Primers are reported in Table 2. Amplification was performed in a 50-μl final volume using 0.2 units of GoTaq polymerase (Promega) with 100 ng of DNA template. Final primer, MgCl₂, and dNTP concentrations were 200 nM, 1.5 mM, and 200 μM, respectively. Initial denaturation (94°C) and final extension (72°C) were for 10 min each. Denaturation (94°C), annealing (60°C), and extension (72°C) were carried out for 30 s each for a total of 45 cycles. Following cloning, the construct was sequenced to exclude the possibility of amplification errors and linearized using the vector's EcoRV site. EcoRV (Promega) digestion was performed for 12 h at 37°C. The digested plasmid was subsequently purified with the QIAquick PCR purification kit (Qiagen), and the concentration was measured spectrophotometrically to calculate the plasmid copy number per microliter. A standard curve was generated by serially diluting (10-fold) the plasmid over a range from 10⁶ to 10⁰ copies/μl. To achieve a reliable standard curve, a plasmid standard was amplified in four replicates for each dilution point over the complete range. Amplifications and data collection were carried out using the ViiA-7 real-time PCR system and Power SYBR green PCR master mix (Life Technologies). The same TP0126 primers used for construction of the plasmid standard were used for message quantification. To obtain the template for these amplifications, total RNA from each treponemal strain was reverse transcribed using a SuperScript III first-strand synthesis kit (Life Technologies) according to the manufacturer's instructions and by using the maximum volume of RNA allowed. Prior to quantitative PCR, reverse-transcribed samples, along with DNase I-treated and untreated samples, were checked by qualitative PCR as previously described (33) to ensure that undigested residual DNA would not affect quantification. Construction of the TP0574 standard was also previously described in detail (33). Triplicate amplifications were performed for both the TP0126 and TP0574 genes for each strain analyzed. Three microliters of cDNA template was used in each reaction. Primers were used at 500 nM each. Cycling conditions were initial denaturation (94°C) for 10 min, followed by denaturation (94°C) for 15 s and annealing-extension (60°C) for 1 min. Amplification was followed by a standard melting curve generation step.

Results were analyzed using the ViiA-7 real-time PCR software. Data analysis was performed using ANOVA, with significance being set at a P value of <0.05, and the Bonferroni correction was applied for multiple comparisons.

Expression of recombinant TP0126 and *T. pallidum* σ⁷⁰. Recombinant TP0126, devoid of its putative signal peptide, was expressed following ORF cloning into the pEXP-5-NT TOPO vector (Life Technologies), which added to the TP0126 coding sequence an additional 22 NH_2-terminal amino acids, including the 6×His tag. No additional residues were added at the COOH terminus. Amplification of the TP0126 ORF from Nichols Seattle strain DNA was performed in our laboratory, as was ORF cloning into the expression vector according to the manufacturer's instructions and confirmation of the correct sequence and orientation for expression. Briefly, the TP0126 gene was amplified in a 50-µl final volume using 0.2 unit of GoTaq polymerase (Promega) with approximately 100 ng of DNA template. Primer (Table 2), $MgCl_2$, and dNTP final concentrations were 200 nM, 1.5 mM, and 200 µM, respectively. Initial denaturation and final extension (72°C) were for 10 min each. Denaturation (94°C), annealing (60°C), and extension (72°C) were carried out for 1 min each for a total of 45 cycles. Protein expression and purification services were purchased from ProMab Biotechnologies, Inc. (Richmond, CA). According to their protocol, a starter culture of *E. coli* Rosetta cells (Merck KGaA, Darmstadt, Germany) carrying the recombinant plasmid was prepared by inoculating 2.0 ml of LB-ampicillin (100 µg/ml) broth with a single *E. coli* colony carrying the TP0126 transgene and allowed to grow at 37°C in a shaking incubator at 250 rpm until the OD_{600} reached 1.0 AU. The starter culture was then diluted to 20 ml in fresh broth. When the new culture again reached 1.0 AU, it was then diluted into 2 liters of fresh broth and split in two 4-liter flasks. When the culture reached an OD_{600} of 0.6 AU, it was induced with IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 1 mM. The cultures were then incubated for 5 h at 30°C and subsequently harvested by centrifugation at $3,500 \times g$ for 15 min at 4°C, resuspended in 100 ml of PBS, and centrifuged again. Recombinant TP0126 was purified under denaturing conditions. Following pellet resuspension in 5 ml of binding buffer (50 mM NaH_2PO_4, 8 M urea, 10 mM imidazole, pH 8.0) per g of culture weight, the suspension was kept on ice and sonicated with 100 pulses of 6 s each, with each pulse being separated by 10-s intervals. Insoluble components were precipitated by centrifugation at 18,000 rpm for 30 min at 4°C, and the cell lysate was saved. For affinity chromatography, 2.0 ml of nickel-agarose was packaged into a column 2.5 by 10 cm and washed with 3 column volumes of molecular-grade H_2O and 6 column volumes of binding buffer. Cell lysate was then loaded, and the flow was adjusted to 1 ml/min. Unbound proteins were washed using 10 bed volumes of binding buffer, followed by 6 column volumes of wash buffer (50 mM NaH_2PO_4, 8 M urea, 20 mM imidazole, pH 8.0). Washing continued until the A_{280} of the flowthrough was <0.01 AU. Recombinant TP0126 was eluted with 2 ml of elute buffer (50 mM NaH_2PO_4, 8 M urea, 300 mM imidazole, pH 8.0). This elution step was repeated three times in total, and eluates were analyzed for purity and yield by SDS-PAGE. For CD, recombinant TP0126 was refolded using a ProFoldin refolding column (Hudson, MA) containing lysophosphatidylcholine (~5 mM), arginine (~150 mM), glycerol (~10%), dodecyl maltoside (0.7 mM), and Tris-HCl (0.1 mM), pH 7.5, according to the manufacturer's instructions. Refolded TP0126 was dialyzed against PBS using a 10-kDa-molecular-mass-cutoff Side-A-Lyzer dialysis cassette (Pierce Biotechnology Inc., Rockford, IL), and the concentration was evaluated using a micro-bicinchoninic acid protein assay kit (also from Pierce). CD spectra (190 to 260 nm) were acquired in triplicate at room temperature using 0.5 mg/ml of recombinant, refolded TP0126 in a Jasco-1500 high-performance CD spectrometer. CD spectra were analyzed using the online platform Dichroweb (http://dichroweb.cryst.bbk.ac.uk /html/home.shtml) and the spectra from buffer alone for background subtraction.

Recombinant *T. pallidum* σ⁷⁰ (encoded by the TP0493 gene) was expressed as a soluble recombinant protein in *E. coli* using the same protocol for expression and purification previously used for another *T. pallidum* σ factor (σ²⁴, encoded by the TP0092 gene) and the *T. pallidum* cyclic AMP receptor protein (CRP) homolog, encoded by TP0262 (38, 39). The TP0493 gene was amplified from strain Nichols DNA as described above for the TP0126 gene. The resulting amplicon was directly cloned into the pEXP5-CT/TOPO expression vector (Life Technologies) according to the manufacturer's instructions. A suitable plasmid was selected after assessment of the correct orientation and the accuracy of the insert sequence. The pEXP5-CT/TOPO vector allows expression of a recombinant protein without any accessory fusion sequence, with the exception of a carboxyl-terminal 6×His tag preceded by 2 amino acid residues (Lys-Gly).

Humoral response against TP0126 during experimental syphilis and human infection. Purified recombinant TP0126 in PBS containing 0.1% sodium azide and recombinant TP0574 (the 47-kDa lipoprotein), purchased from ViroStat (Portland, ME) and used as a positive-control antigen, were used to coat the wells of a 96-well enzyme-linked immunosorbent assay (ELISA) EIA II Plus microplate (ICN, Irvine, CA). The plates, containing 0.5 µg/well of TP0126 or TP0574 protein in 100 µl, were incubated at 37°C for 2 h and subsequently at 4°C overnight to induce antigen binding to the test wells. The wells were then washed three times with PBS, blocked by incubation for 1 h at room temperature with 200 µl of 3% nonfat milk–PBS per well, and washed again. For each of the 9 strains studied here (Bal3, Chicago, Nichols Seattle, MexicoA, Sea81-4, BosniaA, IraqB, SamoaD, and CuniculiA), sera from three infected rabbits were collected at approximately 4 and 12 weeks after i.t. inoculation and pooled. Pooled sera from three uninfected rabbits were used as a negative control. To remove antibodies against *E. coli* antigens, all rabbit sera to be tested were adsorbed against an *E. coli* (Rosetta strain) lysate. Ten microliters of each adsorbed pool was diluted 1:20 in 1% nonfat milk–PBS, and 100 µl was dispensed into the wells. Sera were incubated overnight at room temperature. The wells were then washed three times with PBS plus 0.05% Tween 20 (Sigma Inc., St. Louis, MO). As a procedural control, a mouse anti-6×His antibody (Sigma) was used to ensure TP0126 binding to the plate. Sera from syphilis patients at different stages (three patients with early latent syphilis and five patients with late latent syphilis), as well as 12 serum samples from healthy controls, were tested at a 1:20 dilution. One hundred microliters of secondary antibody (alkaline phosphatase-conjugated goat anti-rabbit IgG, goat anti-mouse IgG, or goat anti-human IgG; all from Sigma) diluted 1:2,000 in 1% nonfat milk–PBS was then added to each well, and the plates were incubated for an additional 3 h at room temperature before repeating the washing step. After addition of 50 µl of 1 mg/ml *para*-nitrophenylphosphate (Sigma) to each well, the plates were developed for 45 min and read at 405 nm on a Multiskan MC plate reader (Titertek, Huntsville, AL). The mean of the background readings (obtained from uninfected rabbit sera or healthy human blood donors) was subtracted from the mean for triplicate experimental wells for each antigen. Results were analyzed using Student's *t* test, with significance being set at a P value of <0.05.

***In silico* analysis of TP0126 hypothetical protein.** The sequence and structural homology of the predicted protein encoded by the shorter TP0126 ORF to other bacterial proteins was investigated using the BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), I-TASSER (http://zhanglab.ccmb .med.umich.edu/I-TASSER/), and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2 /html/page.cgi?id=index) programs. The presence of a cleavable signal peptide was predicted by use of the SignalP (version 4.1; http://www.cbs .dtu.dk/services/SignalP/), LipoP (version 1.0; http://www.cbs.dtu.dk /services/LipoP/), and PrediSi (http://www.predisi.de/) programs.

RESULTS

Comparative study of TP0126 sequence, TP0126 transcriptional start site identification, and analysis of TP0126-associated poly(G) repeat variability *in vivo*. The TP0126 ORF was sequenced to evaluate the genomic diversity among six *T. pallidum* subsp. *pallidum* strains (Bal3, Chicago, MexicoA, Nichols Seattle, UW249, Sea81-4), two *T. pallidum* subsp. *endemicum*

strains (BosniaA and IraqB), *T. pallidum* subsp. *pertenue* strain SamoaD, and *T. paraluiscuniculi* strain CuniculiA. The Nichols Houston TP0126 sequence was already available (24, 34). *T. pallidum* subsp. *endemicum* and *T. pallidum* subsp. *pertenue* cause nonvenereal infections (bejel and yaws, respectively) (40); *T. paraluiscuniculi* is the agent of rabbit venereal syphilis and is reportedly not infectious to humans, despite the high level of genetic identity with human-pathogenic treponemes (36, 41, 42).

Nucleotide sequence comparison of the TP0126 ORF (as initially annotated in the Nichols strain) (24) displayed very limited sequence diversity in the gene ORF among the analyzed strains (Fig. 1A). However, upon translation of the annotated sequences, it appeared that all but three ORFs (from the Nichols Seattle, Nichols Houston, and BosniaA strains) carried a predicted frameshift and a premature stop codon due to a poly(G) repeat of 9 G residues (versus 10 G residues in the Nichols Seattle, Nichols Houston, and BosniaA strains; Fig. 1A). This prompted us to identify the actual TP0126 transcriptional start site (TSS) by 5′ rapid amplification of cDNA ends (RACE) with RNA from the Nichols Seattle and Chicago strains of *T. pallidum* subsp. *pallidum* as the template. 5′ RACE results showed that the TP0126 TSS is located 12 nt downstream of the poly(G) repeat, which, therefore, is not part of the coding sequence, as originally annotated (24). Using the verified TSS position, −35 (TTGCAC) and −10 (TAGGAT) σ^{70} signature sequences were located (Fig. 1A) upstream and downstream of the poly(G) repeat, respectively, as is often seen for genes undergoing phase variation (14, 19). The location of the reported −35 and −10 sequences was also supported by the bacterial promoter identification program BProm (see Materials and Methods). BProm also predicted the dependence of the TP0126 promoter on the σ^{70} transcription factor.

Sequence translation from the first start codon (ATG) downstream of the experimentally identified TSS revealed a highly conserved putative protein in all isolates, with the exception of 1 amino acid difference at position 112 in *T. paraluiscuniculi* TP0126 (Fig. 1B), due to an A-to-G transition at position 538 in the originally annotated DNA sequence (Fig. 1A). The choice of this new start codon for TP0126 was supported by the presence of a putative ribosome binding site (RBS) 6 nucleotides upstream (Fig. 1A). Within this corrected TP0126 ORF, a single synonymous nucleotide change was seen at position 383 of the sequence in Fig. 1A, where a T is present in all syphilis-causing isolates (with the exception of MexicoA), while all other subspecies and species tested harbor a G in this position. Another synonymous change (A to G) is present at position 641 of the *T. paraluiscuniculi* sequence (Fig. 1A).

***In vivo* variation of length in the TP0126-associated poly(G) tract.** To corroborate the hypothesis of a possible role for poly(G) in the transcriptional regulation of the TP0126 gene, we investigated whether this repeat varied in length *in vivo* within each strain studied here by monitoring changes in the poly(G) region in the clonal Nichols Houston O isolate over time. For this purpose, we used a fluorescent fragment length analysis (FLA) method based on amplification of the poly(G) repeat with a fluorescent primer and subsequent amplicon separation by capillary electrophoresis in a genetic analyzer. The results are shown in Fig. 2. Although the clonal Nichols Houston O isolate used to inoculate rabbits (Fig. 2A, Nichols HO inoculum) was found to carry a clonal *tprK* gene (data not shown), this strain showed only a nearly clonal TP0126-associated poly(G), with no 8-G-residue tracts being identified

(94.1% of promoters had 9 G residues, and 5.9% had 10 G residues). Promoters with 8 G residues, however, were detected in all subsequent samples from lesions developing in rabbits infected with this inoculum. Additionally, significant variability in the proportion of promoters with 9 and 10 G residues was noted in these samples (Fig. 2A). These results demonstrate that the length of the TP0126-associated poly(G) region varies *in vivo* within an isolate. When different *T. pallidum* strains obtained directly from infected rabbit tissue at peak orchitis were analyzed, the TP0126-associated poly(G) region was found to be variable and range from 8 to 11 nt in length, although not all lengths were detected in every strain (Fig. 2B). The Sea81-4 isolate carried a significantly higher proportion of promoters with 8 G residues than all remaining strains. No significant differences were found among the remaining strains (Bal3, CuniculiA, IraqB, SamoaD, Chicago, and MexicoA), where promoters carrying 8 G residues could be detected.

Role of poly(G) length in transcription. To investigate whether poly(G) repeats of different lengths would affect the activity of the TP0126 promoter, we adopted an *E. coli*-based heterologous system that allows monitoring of expression of a vector-encoded GFP reporter gene previously placed under the control of an exogenous promoter. This approach was previously used by our group to evaluate the role of similar poly(G) repeats in transcription of the *tprF*, *tprI*, *tprE*, and *tprJ* genes, which also encode *T. pallidum* putative outer membrane proteins (OMPs) (23). For this study, six different TP0126 promoters [with poly(G) regions ranging in length from 7 to 12 nt] were tested along with positive and negative controls (the *lac* promoter and a promoterless reporter vector, respectively). The results (Fig. 3A) showed that a maximal GFP fluorescence signal was detected when the TP0126 promoter carried a poly(G) region of 7 residues. Eight G residues also induced an elevated fluorescence signal, although it was significantly lower than that with the promoter with 7 G residues. When ≥9 G residues were present in the promoter poly(G), the fluorescence signal was reduced by >80% in comparison to that of the promoter with 7 or 8 G residues; a signal was undetectable with 12 residues (Fig. 3A). This result strongly supports the hypothesis that TP0126 expression is controlled by phase variation. As mentioned above, the optimal length for the −35/−10 spacer is 17 nt. In the TP0126 promoter, a poly(G) of 7 residues brings the length of the spacer to exactly 17 nucleotides, which appears to be optimal for transcription. A spacer of 18 nt, however, still allows transcription. It is interesting to note that despite repeated amplification and cloning attempts, a poly(G) repeat with 7 G residues could not be isolated directly from bacterial DNA and instead needed to be engineered using a synthetic primer (Table 2), which suggests that TP0126 overexpression may be detrimental to the microorganism *in vivo*.

In addition to the use of our heterologous system, we also devised an *in vitro* transcription (IVT) assay to confirm the influence of poly(G) regions of different lengths on the activity of the TP0126 promoter and also to investigate whether *T. pallidum* housekeeping factor σ^{70} would initiate TP0126 transcription, as predicted by BProm. To do so, the same test and control constructs used in our *E. coli*-based heterologous system were mixed *in vitro* with RNAP core enzyme saturated with either the recombinant *T. pallidum* or *E. coli* σ^{70} protein to carry out transcription. Assay output was represented by the number of copies of the GFP transcript per microliter of reverse-transcribed mRNA generated during the IVT assay reaction, measured by real-time qPCR. Procedural controls

A

Start Codon reported in the
Nichols genome sequence

```
Chicago        1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
MexicoA        1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
NicholsSeattle 1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
IraqB          1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
BosniaA        1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
CuniculiA      1 ATGCGGACGCACGACATACCCCGCAGTCCCCTCGTGGGGCATAAGAAAAACGCAGC....
```

```
                      -35                            -10        +1
                      ------                         ------     TSS
Chicago        121 GGTTCGACAAACAGCCGTGGGGGGGGG-TAGAATGGAGTAGGTCCTCGACGAGACGCGTA
MexicoA        121 GGTTCGACAAACAGCCGTGGGGGGGGG-TAGAATGGAGTAGGTCCTCGACGAGACGCGTA
NicholsSeattle 121 GGTTCGACAAACAGCCGTGGGGGGGGGGTAGAATGGAGTAGGTCCTCGACGAGACGCGTA
IraqB          121 GGTTCGACAAACAGCCGTGGGGGGGGGGTAGAATGGAGTAGGTCCTCGACGAGACGCGTA
BosniaA        121 GGTTCGACAAACCGCCGTGGGGGGGGGGTAGAATGGAGTAGGTCCTCGACGAGACGCGTA
CuniculiA      121 GGTTCGACAAACAGCCGTGGGGGGGGG-TAGAATAGAGTGGGTCCTCGGCGAGATGCGTA
```

```
                          Newly identified
                    RBS    start codon
Chicago        180 AGAGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
MexicoA        180 AGAGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
NicholsSeattle 181 AGAGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
IraqB          180 AGAGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
BosniaA        181 AGAGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
CuniculiA      180 AGGGATCGGCGTTGGAGCGGGGTATGAAACAATTGAAGTGGTGGGCGGTCGGCCC....
```

```
Chicago        360 CAAGTGGATTGGTGTAACAGCGGTCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
MexicoA        360 CAAGTGGATTGGTGTAACAGCGGCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
NicholsSeattle 361 CAAGTGGATTGGTGTAACAGCGGTCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
IraqB          360 CAAGTGGATTGGTGTAACAGCGGGCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
BosniaA        361 CAAGTGGATTGGTGTAACAGCGGGCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
CuniculiA      360 CAAGTGGATTGGTGTAACAGCGGGCGCGGGGACGACCGTAATGCCAACGCGCAAAC....
```

```
Chicago        484 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAAT....
MexicoA        484 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAAT....
NicholsSeattle 485 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAAT....
IraqB          484 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAAT....
BosniaA        485 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAAT....
CuniculiA      484 GCCTGTCGCTCGGGATTGACGCAAGTGTGCAGTACCATTGCAGCTATCCAAACAGT....
```

```
Chicago        604 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTAGGGGCTGGCTTCAA....
MexicoA        604 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTAGGGGCTGGCTTCAA....
NicholsSeattle 605 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTAGGGGCTGGCTTCAA....
IraqB          604 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTAGGGGCTGGCTTCAA....
BosniaA        605 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTAGGGGCTGGCTTCAA....
CuniculiA      604 TACACGGTTGCCTTCTGGCGCATCCGCTTGCCACTGACGGTGGGGGCTGGCTTCAA....
```

```
Chicago        844 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
MexicoA        844 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
NicholsSeattle 845 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
IraqB          844 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
BosniaA        845 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
CuniculiA      845 ATAGCCGCGGGTGTGGACGTTCGCTACCACCTCTGA
```

B

```
NicholsSeattle   1 MKQLKWWAVGPVLGICAGVWGAAHPVHADPWDTTAAGRSTIRLSAMGAVPLFQVDWCNSG
CuniculiA        1 MKQLKWWAVGPVLGICAGVWGAAHPVHADPWDTTAAGRSTIRLSAMGAVPLFQVDWCNSG

NicholsSeattle  61 RGDDRNANAQTNGHKYIYPAFSAALGFEHFVCRGLSLGIDASVQYHCSYPNNTYSPTTPY
CuniculiA       61 RGDDRNANAQTNGHKYIYPAFSAALGFEHFVCRGLSLGIDASVQYHCSYPNSTYSPTTPY

NicholsSeattle 121 YYLAIPVALTAGYTVAFWRIRLPLTVGAGFNYQHYYTSTYYGLVLKAAAGCYFQLTEHWS
CuniculiA      121 YYLAIPVALTAGYTVAFWRIRLPLTVGAGFNYQHYYTSTYYGLVLKAAAGCYFQLTEHWS

NicholsSeattle 181 LGVSATYSGVPRSCEKIIEEDRQQTNTRTAQFIAAGVDVRYHL
CuniculiA      181 LGVSATYSGVPRSCEKIIEEDRQQTNTRTAQFIAAGVDVRYHL
```
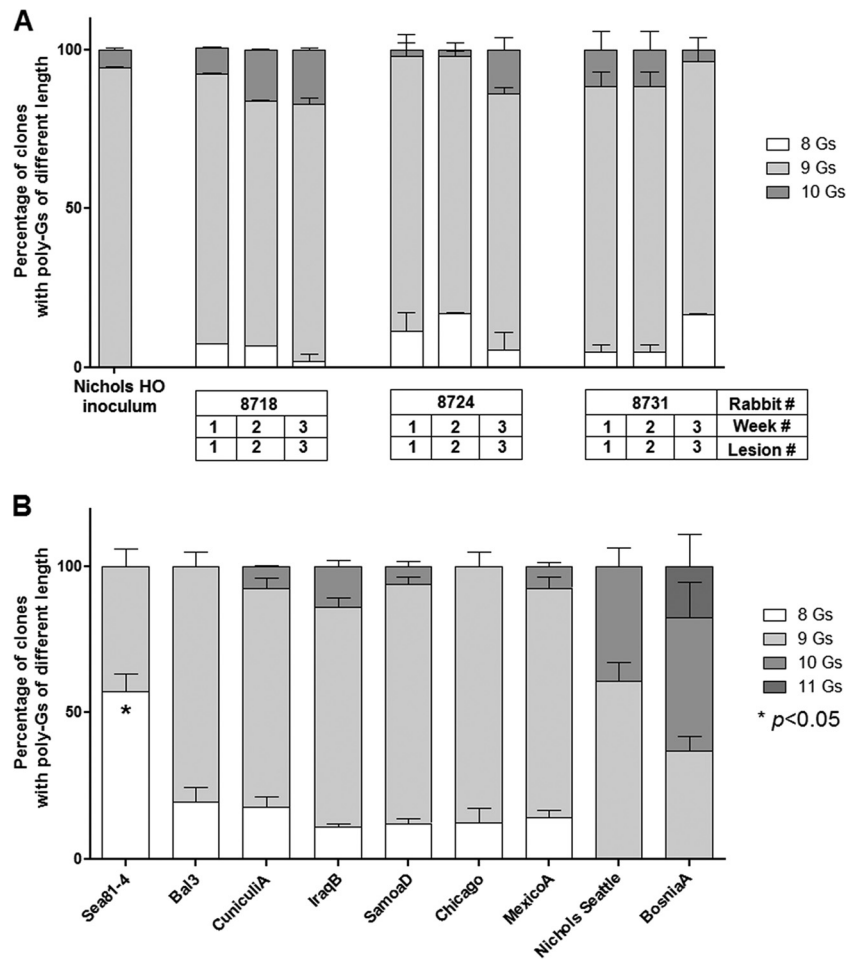
Infection and Immunity

FIG 2 Analysis of the TP0126-associated poly(G) region *in vivo* during experimental syphilis. (A) Results of FLA of the TP0126-associated poly(G) region are shown for the clonal strain Nichols Houston O (HO) inoculum and treponemes found in lesions after i.d. infection with that inoculum. Data are shown for lesions collected from three rabbits over a 3-week infection. (B) Results of FLA of the TP0126-associated poly(G) region for nine *T. pallidum* strains harvested at peak orchitis. *, statistically significant difference (*P* < 0.05) between the proportion of TP0126 promoters with 8 G residues in the Sea81-4 strain compared to the proportion in all remaining strains. The proportion of TP0126 promoters with 8 G residues was not found to be significantly different among the other *T. pallidum* isolates.

included IVT assay reactions for all promoters in the absence of core RNAP and with core RNAP but in the absence of the σ sub-unit. The results (Fig. 3B) were very similar to those obtained in *E. coli*, with the highest level of transcription being seen for the TP0126 promoter carrying a poly(G) of 7 residues and decreasing transcriptional activity being associated with promoters contain-ing longer poly(G) regions. As expected from the GFP assay con-ducted in *E. coli*, both the *E. coli* and *T. pallidum* σ^{70} proteins were able to initiate transcription from the TP0126 promoter at similar

levels. Altogether, these results suggest that variability in the poly(G) tract within the TP0126 promoter modulates expression of this gene at the transcriptional level and that the housekeeping factor σ^{70} is involved in TP0126 transcription. These results con-firmed that the use of *E. coli* σ^{70} as a surrogate for *T. pallidum* σ^{70} is valid, as also suggested by previous IVT experiments with *T. pallidum* promoters and the *E. coli* transcriptional machinery per-formed by our group (43).

Quantification of TP0126 message *in vivo*. Differential mRNA

FIG 1 Comparative analysis of the TP0126 ORF as originally annotated in the *T. pallidum* subsp. *pallidum* Nichols strain (24) (A) and of the TP0126 hypothetical protein (B). Only sequences harboring diversity are shown in these alignments. Sequences from syphilis-causing strains (Chicago, MexicoA, and Nichols Seattle) are on top, sequences from endemic treponematosis-causing strains (IraqB and BosniaA) are in the middle, and the sequence of the *T. paraluiscuniculi* CuniculiA strain sequence is at the bottom. Nucleotide differences are shown by a white or a gray background, in contrast to a black background, which identifies conserved residues. Regulatory elements, including the poly(G) region, putative −35 and −10 binding sites, transcriptional start site (experimentally identified), predicted ribosome binding site, and start and stop codons are highlighted in orange. In panel A, omitted sequences for syphilis-causing strains (Bal3 and Sea81-4) are identical to the Chicago strain sequence. The Nichols Houston strain sequence is identical to the Nichols Seattle strain sequence. The *T. pallidum* subsp. *pertenue* SamoaD sequence (omitted) is identical to the *T. pallidum* subsp. *endemicum* IraqB sequence. Portions of the sequences identical in all strains were omitted and are indicated by ellipses. In panel B, predicted amino acid sequences for the putative TP0126 protein from the Chicago, MexicoA, Bal3, Sea81-4, Nichols Houston, IraqB, BosniaA, and SamoaD strains are identical to the Nichols Seattle sequence shown. The first 28 aa of the putative TP0126 protein are predicted to be a cleavable signal peptide (see "*In silico* analysis of TP0126 hypothetical protein and CD" in Results section).
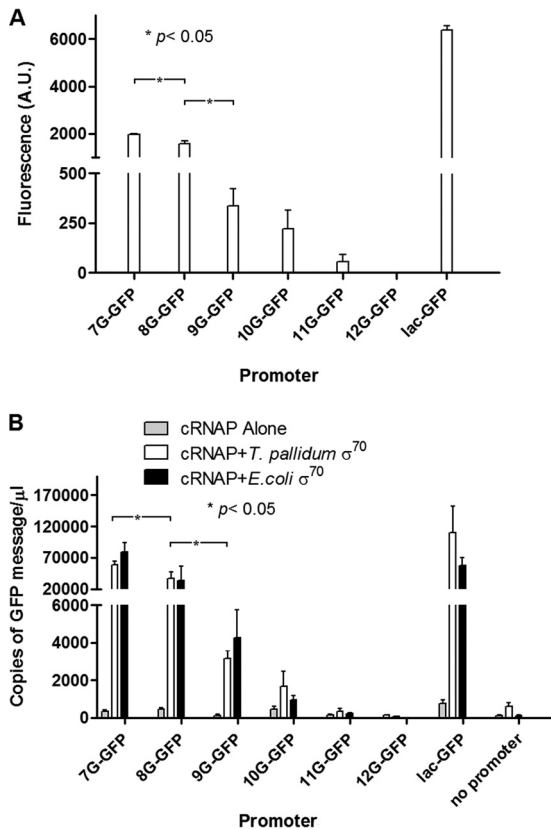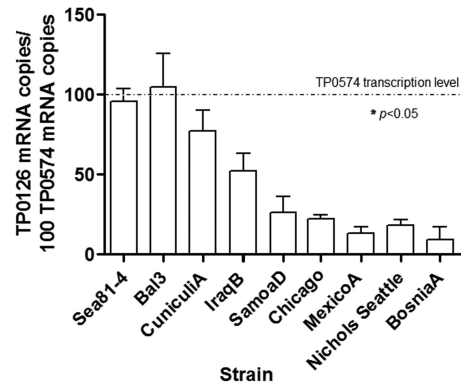
FIG 4 Message quantification for TP0126 in treponemal isolates. The number of TP0126 gene mRNA copies is normalized to 100 copies of the mRNA of the TP0574 gene (encoding the 47-kDa lipoprotein). In the Sea81-4, Bal3, and CuniculiA strains, TP0126 expression was significantly higher than that in all other strains, with the exception of CuniculiA and IraqB TP0126 message levels. TP0126 expression in IraqB was found to be higher than that in MexicoA, Nichols Seattle, and BosniaA but not SamoaD and Chicago. MexicoA, Nichols Seattle, and BosniaA exhibited similar transcription levels for TP0126. Transcription levels roughly matched the percentage of poly(G) repeats carrying 8 G residues (reported in Fig. 2B) in these strains.

FIG 3 Analysis of effect of poly(G) length (7 to 12 nt) on TP0126 transcription. (A) Fluorescence induced in *E. coli* TOP10 cells transformed with a pGlow-TOPO TA vector where GFP transcription is under the control of TP0126 promoters with poly(G) regions of different lengths. A *lac* promoter-GFP construct was used as a positive control. The *lac* promoter is recognized by σ^{70}, and the *E. coli* strain used for this assay does not carry the gene that encodes the LacI repressor. The background fluorescence collected from *E. coli* cells transformed with a pGlow-TOPO TA vector that carries a fragment of the *T. pallidum* TP0574 ORF devoid of a promoter was subtracted from the values for the other samples. *, statistically significant difference ($P < 0.05$). Signals from 9 G-, 10 G-, and 11 G-GFP constructs were not found to be significantly different. A signal from the 12 G-GFP construct was undetectable. (B) Results of the IVT assay using the same constructs employed for the TP0126 promoter-GFP assay in *E. coli*. Three different conditions are shown: transcription driven by cRNAP alone, cRNAP plus *T. pallidum* σ^{70}, or cRNAP plus *E. coli* σ^{70}. The data output is the number of copies of the GFP transcript per microliter measured by real-time qPCR. There was no transcription in the absence of cRNAP (data not shown). *, statistically significant difference ($P < 0.05$) when signals from different constructs are compared. The message levels (white bars) induced by the 9 G-, 10 G-, 11 G-, and 12 G-GFP constructs were not found to be significantly different.

levels among strains leading to differential expression of an antigen can be the outcome of several factors, including transcriptional control based on randomly changing *cis*-acting elements, such as homopolymeric repeats in promoter regions. To examine whether this occurs in *T. pallidum*, a real-time qPCR assay was developed to quantitate the TP0126 message in treponemal strains at the time of harvest (peak orchitis) from the rabbit host. This approach normalizes the TP0126 message level to that of the TP0574 gene (encoding the *T. pallidum* 47-kDa antigen, a highly expressed periplasmic lipoprotein), as previously described (33). Quantification data revealed that TP0126 is variably expressed in the isolates studied here (Fig. 4). High levels of expression of

TP0126 that were not significantly different among the Sea81-4, Bal3, and CuniculiA strains and compatible with those of the 47-kDa antigen were seen. The rank order of strains from the highest to the lowest transcription levels (Fig. 4) mirrors in part the rank order of the strains for the proportion of poly(G) repeats carrying 8 G residues (reported in Fig. 2A), suggesting that strains carrying a higher proportion of 8-G-residue tracts have higher levels of expression of TP0126.

Humoral response against TP0126 during experimental and human syphilis spirochete infection. The possibility of variable degrees of humoral immunity to TP0126 in hosts infected with *Treponema* isolates was evaluated by ELISA using recombinant TP0126 and sera obtained from infected rabbits at 4 and 12 weeks postinfection. The results of the TP0126 ELISA (Fig. 5A) support the suggestion that infection with different *T. pallidum* isolates leads to differential levels of production of TP0126-specific antibodies. Most of the serum samples reacted positively to the TP0574 antigen (Fig. 5B). The exceptions were the sera from rabbits infected with the Sea81-4 and CuniculiA strains collected at week 4, which showed reactivity to the antigen significantly lower than that of sera obtained at the same time point from rabbits infected with the other strains.

Analysis of sera from human patients with early and late latent syphilis to assess the presence of a humoral response against TP0126 revealed that all serum samples had a weak but detectable humoral response against TP0126 (Fig. 6). All human serum samples, however, reacted strongly against the *T. pallidum* TP0574 antigen, used as a control.

***In silico* analysis of TP0126 hypothetical protein and CD.** The presence of a putative cleavable signal peptide and the structural homology of the TP0126 protein to other proteins was investigated using a variety of *in silico* tools, including the BLAST, I-TASSER, Phyre2, SignalP, PrediSi, and LipoP programs. All signal peptide prediction programs agreed on the presence of a signal peptide with a cleavage site between residues 28 and 29. Although BLAST searches were unable to identify a TP0126 ortholog in
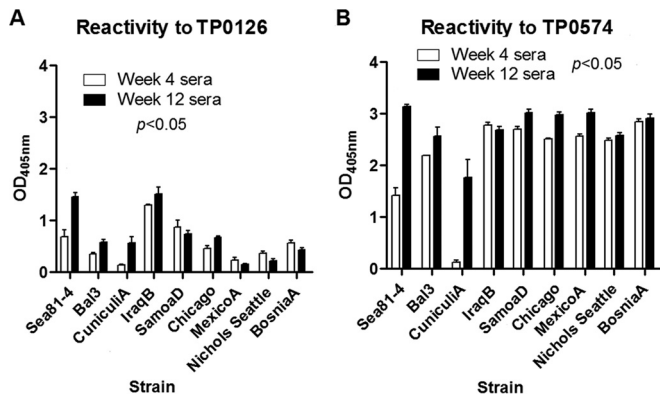
FIG 5 Humoral reactivity to the TP0126 antigen (A) and the TP0574 antigen (B) of sera from rabbits infected with different treponemal strains. Sera were collected at 4 weeks and 12 weeks after rabbit i.t. infection. A mouse anti-6×His monoclonal antiserum was used to confirm TP0126 binding to the wells of the ELISA plate; the OD_{405} measured using anti-6×His antibody was 3.64 AU. The optical density from sera from uninfected rabbits was used for background subtraction.
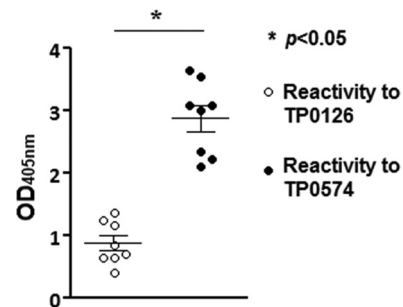

FIG 6 Humoral reactivity of sera from patients with syphilis at different stages (primary, secondary, and latent) to the TP0126 protein and the TP0574 antigen (the 47-kDa lipoprotein). A mouse anti-6×His monoclonal antiserum was used to confirm TP0126 binding to the wells of the ELISA plate. The optical density for sera from uninfected patients was used for background subtraction. Significance was evaluated using Student's t test, with significance being set at a P value of ≤0.05. *, statistically significant difference between reactivity to the TP0126 and TP0574 antigens for each group of sera.

other prokaryotes on the basis of sequence homology, both I-TASSER and Phyre2 identified TP0126 to be a structural homolog of *E. coli* OmpW and *Pseudomonas aeruginosa* OprG, also a member of the OmpW family of proteins. Secondary structure and disorder prediction by I-TASSER (Fig. 7A) reports an α-helix component of 3.5% (when a protein sequence lacking the signal peptide is used for analysis), a β-barrel component of 49.2%, and a random coil (RC) component of 47.3%. A similar prediction by Phyre2, however, reported component proportions of 14% α helix, 70% β barrel, and 16% RC. The I-TASSER confidence (C) score and TM score were −2.52 and 0.42 ± 0.14, respectively. The C score estimates the quality of the predicted models and typically ranges from −5 to +2, with a higher C-score value signifying a model with a high confidence and vice versa. The TM score provides a scale for measuring the similarity between two structures; a TM score of >0.5 indicates a model of correct topology, and a TM score of <0.17 indicates random similarity. Due to the moderate confidence for this model expressed by I-TASSER analysis, re-folded recombinant TP0126 (without the signal peptide) was used to produce circular dichroism (CD) spectra and to gain preliminary experimental evidence for the TP0126 structure. Analysis of the CD spectra (Fig. 7B) with the Dichroweb program revealed proportions of β-barrel (43%), α-helix (30%), and random coil (27%) components that are compatible with a β-barrel OMP.

DISCUSSION

Although syphilis can easily be diagnosed with traditional and modern tests (44–46) and effectively treated with penicillin (47), the continuing high global prevalence of this infection is a strong argument for the need to deepen our knowledge of syphilis pathogenesis. Such research efforts will allow syphilis investigators to understand what mechanisms make *T. pallidum* such a successful pathogen able to invade virtually every bodily organ, including the placenta, and persist for the lifetime of the infected host, in spite of a robust immune response (48). The knowledge accumulated in the last decade on *T. pallidum* variable antigen TprK, for example, supports a pivotal role for antigenic variation in immune evasion and pathogen persistence (29, 30, 32).

Phase variation is a mechanism that allows reversible and quick on/off switching of gene expression at the transcriptional or translational level and is generally mediated by modification of DNA repeats of various lengths located within a gene promoter or ORF. In other pathogens, such as *N. meningitidis* or *Helicobacter pylori*, phase variation is strongly associated with expression of genes involved in adaptation to different environmental conditions, including different sites of infection, but also immune evasion and persistence (14, 15, 49). Whether phase variation has a role in fostering the ability of *T. pallidum* to avoid immune clearance, persist, and perhaps adapt to the diverse microenvironments in the host is still unclear. However, evidence that transcription of at least three *T. pallidum* repeat (*tpr*) genes encoding putative surface-exposed virulence factors (24, 50) is controlled by phase variation (23) strongly suggests that this mechanism is also important in syphilis pathogenesis. A preliminary search for poly(G) sequences (of ≥8 residues) in the Nichols strain genome (24) leads to the identification of 20 such elements that are located upstream or within *T. pallidum* ORFs (Table 3) and that have been studied only partially or not at all (Table 3). Preliminary analysis of the poly(G) region located within the TP0127 ORF, for example, showed that this poly(G) tract is variable in length within isolates, and depending upon the number of G residues present, two alternative TP0127 proteins of comparable size but different primary structure are predicted, as is a third variant with a premature stop codon that truncates the protein (31). The identification of variable sequence repeats and the analysis of their possible functional significance based upon their sequence context, for example, allowed Saunders et al. (51) to identify 52 new potentially phase-variable genes in *N. meningitidis*, in addition to those previously recognized. Such findings support the suggestion that in *N. meningitidis* the role of phase variation in mediating bacterium-host interactions is likely greater than that which has been appreciated thus far. The expression of these genes in different combinations could generate an impressive number of bacterial phenotypic variants, allowing overall better pathogen fitness in the arena represented by the infected host. Hence, the identification of homopolymeric sequences of various lengths in the *T. pallidum* genome could provide a simple approach to single out genes with an important role in virulence.
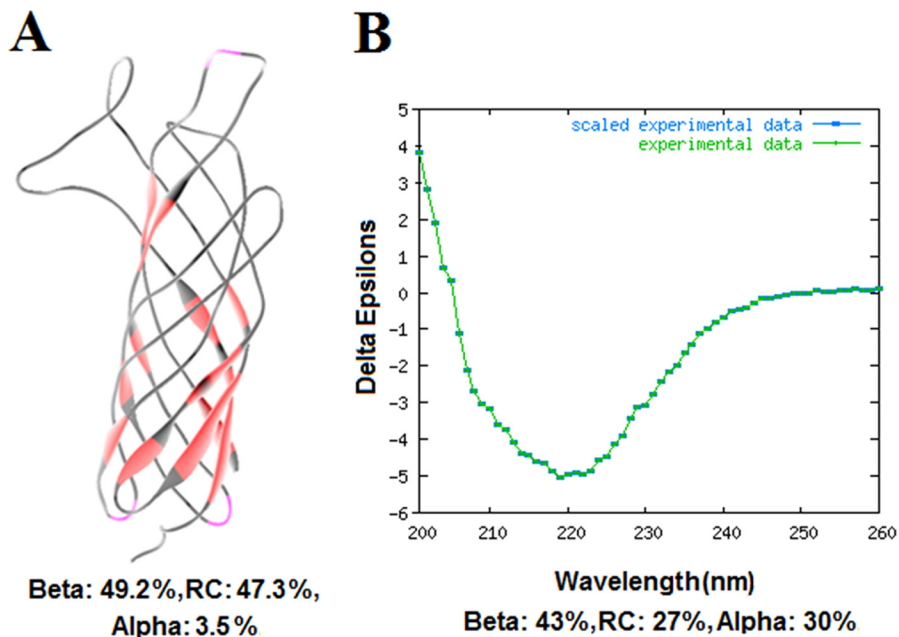
FIG 7 TP0126 model by I-TASSER (A) and CD spectra of recombinant TP0126 (B). The I-TASSER and Phyr2 predictions of the TP0126 structure are based on structural homology with the *E. coli* OmpW porin, a small β-barrel protein likely involved in the transport of molecules with a hydrophobic character into the outer membrane. The proportions of the β-barrel, random coil (RC), and α-helix components based on I-TASSER prediction are 49.2%, 3.5%, and 47.3%, respectively. Phyr2 predicted higher proportions of the β-barrel and α-helix components (70% and 14%, respectively) but a lower proportion of the RC component (16%). The CD spectra of recombinant TP0126 expressed without the putative cleavable signal peptide (but with 22 additional plasmid-encoded NH$_2$-terminal amino acids), however, are consistent with proportions of β-barrel, RC, and α-helix components of 43%, 27%, and 30%, respectively.

TABLE 3 *T. pallidum* genes with associated poly(G) tracts[a]

Gene no.	Putative gene function	Poly(G) location[b]
TP0026	Flagellar motor switch protein	−100
TP0041-TP0042[c]	Hypothetical proteins	−31
TP0059	Hypothetical protein	+153
TP0067	Conserved hypothetical protein	+24
TP0084	Hypothetical protein	−48
TP0107	Liposaccharide biosynthesis protein	−307
TP0127	Hypothetical protein	+341
TP0136	Fibronectin-binding protein	+532
TP0193	30S ribosomal protein S19	−9
TP0216	Heat shock protein 70	+60
TP0230	Primosomal protein N	+942
TP0257	Protein D, glycerophosphodiester phosphodiesterase	+29
TP0279	30S ribosomal protein S1/cytidylate kinase	+7
TP0347	Hypothetical protein	+99
TP0379	Preprotein translocase subunit	−70
TP0479	Hypothetical protein	+95
TP0497	Hypothetical protein	+478
TP0697	Hypothetical protein	+27
TP0798	MetG, methionyl-tRNA synthetase	+826
TP0860	Hypothetical protein	+78

[a] Only poly(G) tracts of ≥8 residues are listed. A search was performed using the *T. pallidum* subsp. *pallidum* Nichols genome (21).
[b] Location indicates the position of the first poly(G) residue in relation to the gene annotated start codon. A negative value indicates that the poly(G) tract is located upstream of the gene ATG. A positive value indicates that the poly(G) tract is within the gene ORF.
[c] TP0041 is a 40-aa-long ORF and might not be real.

The well-documented role of phase variation in controlling expression of surface-exposed proteins involved in the host-pathogen interaction promoted us to investigate whether TP0126 might also be a putative *T. pallidum* OMP not previously identified (52, 53). As reported here, *in silico* analysis and CD spectrum data indicate that TP0126 exhibits structural characteristics consistent with a β-barrel OMP, supporting our hypothesis that TP0126 is a putative surface antigen. We are currently utilizing additional approaches to further substantiate these findings, including investigating whether anti-TP0126 sera can opsonize *T. pallidum* in phagocytosis assays and whether the protein can be detected on the *T. pallidum* surface using a gold-labeled anti-TP0126 serum and electron microscopy, as previously reported for the TprI antigen (54). In addition, experiments identifying the location of B- and T-cell epitopes on TP0126 will provide additional support for our current structural model of the protein (not shown) and its localization in the pathogen's outer membrane. Similar analyses have been used previously in our work on the TprK protein, also predicted to have a β-barrel structure similar to that of TP0126, but with a higher number of antiparallel β strands (14 instead of 8) and predicted surface loops (seven instead of four). For TprK, sera obtained from syphilis spirochete-infected rabbits contain antibodies that react only with TprK's putative surface loops, while T-cell epitopes are confined to amino acid sequences that form the β-strand scaffold (55).

If it is demonstrated that TP0126 is a bona fide OMP, its very high sequence conservation among *T. pallidum* isolates and subspecies makes it an attractive vaccine candidate for syphilis to aid public health control initiatives targeting this disease (56, 57). Historically, the identification of *T. pallidum* OMPs has been difficult (58, 59), and ongoing research efforts are using the reverse vaccin-

ology approach (60, 61) to identify novel vaccine candidates for syphilis: this approach includes the use of *in silico* analysis of available *T. pallidum* genomes to identify hypothetical proteins containing sequence or structural homology to known bacterial OMPs (53, 54); functional activities characteristic of OMPs, such as attachment of *T. pallidum* putative surface antigens to host components (62–65); or the ability of monospecific antibodies to opsonize viable *T. pallidum* cells (20, 66). Evidence, reported here, that during natural infection in humans TP0126 is a weak target for the humoral immune response might simply reflect the limited immunogenicity of this antigen *in vivo*, possibly due to low expression, but it certainly supports the suggestion that, like in rabbits, this antigen is expressed by the pathogen during natural infection and could therefore represent a target on the pathogen's surface. Future protection and opsonophagocytosis studies using a specific anti-TP0126 antiserum will elucidate whether immunity against this protein might facilitate clearance of *T. pallidum* from early lesions, while T-cell proliferation assays will be used to investigate whether TP0126 is also a target of the host cellular immune response.

The characterization of the poly(G) tract within the TP0126 promoter and its involvement in transcriptional regulation and phase variation not only call for additional work to characterize the TP0126 location, structure, and function but also highlight the possibility that ORF annotation in *T. pallidum* by computational gene finding might be subject to errors. Using TP0126 as an example, the original annotation suggested that the hypothetical protein was much larger than it likely is. Thus, despite the important contributions provided by gene annotation pipelines, it is necessary to corroborate *in silico* predictions with transcriptional data. Again, in the case of TP0126, this approach led to the identification of the gene's putative start codon and to the prediction of a cleavable signal peptide previously hidden by the 68 additional NH_2-terminal residues in the original TP0126 annotation (24). Annotation issues can be partially addressed by comparative analysis of available *T. pallidum* genomes. The Nichols strain of *T. pallidum* subsp. *pallidum*, for instance, was sequenced twice (34) due to the high number of sequencing and annotation errors found in the first genome sequence published in the late 1990s (24, 67). Comparison of the TP0126 ORF between the new and old genomes reveals that in the resequenced Nichols strain, due to the presence of fewer residues within the poly(G) repeat, the TP0126 putative sequence is identical to the one proposed here. Such information was unavailable at the time that our work on TP0126 was under way. This difference in the length of the poly(G) tract between the two sequencing efforts likely reflects the most common population in each bacterial harvest.

Our data show that TP0126 transcription can be initiated *in vitro* by the housekeeping factor σ^{70}, suggestive of an important function for this protein *in vivo* in *T. pallidum*. This conclusion is also supported by the complete sequence conservation of this antigen among *T. pallidum* isolates (a single nonsynonymous nucleotide change was identified in the ORF of the related species *T. paraluiscuniculi*). Currently, the only known OmpW function is to serve as a receptor for colicin S4 (68), and the possibility that TP0126 serves a similar function in *T. pallidum* today appears to be extremely unlikely. Recent data on the function of OmpW homologs in other organisms suggest that members of this protein family may be involved in bacterial adaptation to various environmental stresses. Expression of OmpW from *Vibrio cholerae* was found to be affected by culture conditions, such as oxygen and nutrient availability, temperature, and salinity (69), while in *Salmonella enterica* serovar Typhimurium, OmpW expression has been associated with drug resistance (70, 71), perhaps suggesting an efflux pump mechanism that might also function in osmoregulation. Further characterization of the TP0126 protein will perhaps highlight functional similarities between the roles of OmpW in other pathogens and the role of TP0126 in *T. pallidum* biology. The work described above, however, indicates the possibility of a multifaceted role for this protein in helping pathogens deal with environmental stresses and, possibly, nutrient acquisition. In this sense, the crystal structures of *E. coli* OmpW (25) and the *P. aeruginosa* OmpW homolog OprG (26) are partially illuminating. According to the results of Hong et al. (25), who elucidated the *E. coli* OmpW structure, the most unusual and interesting feature of the OmpW channel is the hydrophobic character of the β-barrel interior, considering that porins typically have hydrophilic channels filled with water (72–74). The OmpW channel, therefore, would be implicated in the transport of molecules with a hydrophobic character yet to be identified. The *E. coli* OmpW crystal structure also suggests the hypothesis that molecules able to access the OmpW channel would be directly delivered into the hydrophobic moiety of the outer leaflet of the outer membrane, due to a putative channel exit embedded into the membrane (25). This peculiarity would allow the porin substrate to avoid the hydrophilic periplasmic space. Similar findings were later reported for the *P. aeruginosa* OprG by Touw et al. (26). Whether TP0126 might be the transporter that mediates the acquisition of fatty acids (one of the many macromolecules that the syphilis agent is unable to synthesize but must scavenge from the host) in *T. pallidum* has yet to be shown, but it certainly appears to be an intriguing working hypothesis.

ACKNOWLEDGMENTS

REFERENCES

1. **World Health Organization.** 2011. Prevalence and incidence of selected sexually transmitted infections: *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, syphilis and *Trichomonas vaginalis*. Methods and results used by WHO to generate 2005 estimates. World Health Organization, Geneva, Switzerland.
2. **World Health Organization.** 2001. Global prevalence and incidence of selected curable sexually transmitted infections: overview and estimates. World Health Organization, Geneva, Switzerland.
3. **Tucker JD, Chen XS, Peeling RW.** 2010. Syphilis and social upheaval in China. N Engl J Med **362:**1658–1661. http://dx.doi.org/10.1056/NEJMp0911149.
4. **Anonymous.** 2011. Syphilis on the rise in the USA. Lancet **378:**542. http://dx.doi.org/10.1016/S0140-6736(11)61274-9.
5. **Anonymous.** 2007. 2004 Canadian sexually transmitted infections surveillance report. Can Commun Dis Rep **33**(Suppl 1)**:**1–69.
6. **Botham SJ, Ressler KA, Bourne C, Ferson MJ.** 2006. Epidemic infectious syphilis in inner Sydney—strengthening enhanced surveillance. Aust N Z

J Public Health **30**:529–533. http://dx.doi.org/10.1111/j.1467-842X.2006
.tb00781.x.

7. **Herbert LJ, Middleton SI.** 2012. An estimate of syphilis incidence in Eastern Europe. J Glob Health **2**:010402. http://dx.doi.org/10.7189/jogh .02.010402.

8. **Simms I, Wallace L, Thomas D, Emmett L, Shankar A, Vinson M, Padfield S, Andrady U, Whiteside C, Williams C, Midgley C, Johnman C, McLellan A, Currie A, Logan J, Leslie G, Licence K, Hughes G.** 2014. Recent outbreaks of infectious syphilis, United Kingdom, January 2012 to April 2014. Euro Surveill **19**(24):pii=20833. http://www.eurosurveillance .org/ViewArticle.aspx?ArticleId=20833.

9. **Gerbase AC, Rowley JT, Heymann DH, Berkely SF, Piot P.** 1998. Global prevalence and incidence estimates of selected curable STDs. Sex Transm Infect **74**:S12–S16.

10. **Goldenberg RL, Thompson C.** 2003. The infectious origins of stillbirth. Am J Obstet Gynecol **189**:861–873. http://dx.doi.org/10.1067/S0002-9378 (03)00470-8.

11. **Douglas JM, Jr.** 2009. Penicillin treatment of syphilis: clearing away the shadow on the land. JAMA **301**:769–771. http://dx.doi.org/10.1001/jama .2009.143.

12. **White RG, Orroth KK, Korenromp EL, Bakker R, Wambura M, Sewankambo NK, Gray RH, Kamali A, Whitworth JA, Grosskurth H, Habbema JD, Hayes RJ.** 2004. Can population differences explain the contrasting results of the Mwanza, Rakai, and Masaka HIV/sexually transmitted disease intervention trials?: a modeling study. J Acquir Immune Defic Syndr **37**:1500–1513. http://dx.doi.org/10.1097/01.qai.0000127062 .94627.31.

13. **Hallet B.** 2001. Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. Curr Opin Microbiol **4**:570–581. http://dx.doi .org/10.1016/S1369-5274(00)00253-8.

14. **van der Woude MW, Baumler AJ.** 2004. Phase and antigenic variation in bacteria. Clin Microbiol Rev **17**:581–611. http://dx.doi.org/10.1128/CMR .17.3.581-611.2004.

15. **Salaun L, Snyder LA, Saunders NJ.** 2003. Adaptation by phase variation in pathogenic bacteria. Adv Appl Microbiol **52**:263–301. http://dx.doi.org /10.1016/S0065-2164(03)01011-6.

16. **Hauck CR, Meyer TF.** 2003. 'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication. Curr Opin Microbiol **6**:43–49. http: //dx.doi.org/10.1016/S1369-5274(03)00004-3.

17. **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.

18. **Levinson G, Gutman GA.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol **4**:203–221.

19. **Wisniewski-Dye F, Vial L.** 2008. Phase and antigenic variation mediated by genome modifications. Antonie Van Leeuwenhoek **94**:493–515. http: //dx.doi.org/10.1007/s10482-008-9267-6.

20. **Centurion-Lara A, Castro C, Barrett L, Cameron C, Mostowfi M, Van Voorhis WC, Lukehart SA.** 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. http://dx.doi.org/10.1084 /jem.189.4.647.

21. **Centurion-Lara A, Godornes C, Castro C, Van Voorhis WC, Lukehart SA.** 2000. The *tprK* gene is heterogeneous among *Treponema pallidum* strains and has multiple alleles. Infect Immun **68**:824–831. http://dx.doi .org/10.1128/IAI.68.2.824-831.2000.

22. **Centurion-Lara A, LaFond RE, Hevner K, Godornes C, Molini BJ, Van Voorhis WC, Lukehart SA.** 2004. Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. Mol Microbiol **52**:1579–1596. http://dx.doi.org/10.1111/j .1365-2958.2004.04086.x.

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* subsp. *pallidum*. FEMS Immunol Med Microbiol **51**:289–301. http://dx.doi.org/10.1111/j.1574-695X.2007.00303.x.

24. **Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey EK, Clayton R, Ketchum KA, Sodergren E, Hardham JM, McLeod MP, Salzberg S, Peterson J, Khalak H, Richardson D, Howell JK, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton MD, Venter JC.** 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science **281**:375–388. http: //dx.doi.org/10.1126/science.281.5375.375.

25. **Hong H, Patel DR, Tamm LK, van den Berg B.** 2006. The outer

26. membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. J Biol Chem **281**:7568–7577. http://dx.doi.org/10 .1074/jbc.M512365200.

26. **Touw DS, Patel DR, van den Berg B.** 2010. The crystal structure of OprG from *Pseudomonas aeruginosa*, a potential channel for transport of hydrophobic molecules across the outer membrane. PLoS One **5**:e15016. http: //dx.doi.org/10.1371/journal.pone.0015016.

27. **National Research Council.** 2011. Guide for the care and use of laboratory animals, 8th ed. National Academies Press, Washington, DC.

28. **Lukehart SA, Marra CM.** 2007. Isolation and laboratory maintenance of *Treponema pallidum*. Curr Protoc Microbiol Chapter 12:Unit 12A.1. http: //dx.doi.org/10.1002/9780471729259.mc12a01s7.

29. **LaFond RE, Molini BJ, Van Voorhis WC, Lukehart SA.** 2006. Antigenic variation of TprK V regions abrogates specific antibody binding in syphilis. Infect Immun **74**:6244–6251. http://dx.doi.org/10.1128/IAI.00827-06.

30. **Giacani L, Molini BJ, Kim EY, Godornes BC, Leader BT, Tantalo LC, Centurion-Lara A, Lukehart SA.** 2010. Antigenic variation in *Treponema pallidum*: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J Immunol **184**:3822–3829. http: //dx.doi.org/10.4049/jimmunol.0902788.

31. **Giacani L, Brandt SL, Puray-Chavez M, Brinck Reid T, Godornes C, Molini BJ, Benzler M, Hartig JS, Lukehart SA, Centurion-Lara A.** 2012. Comparative investigation of the genomic regions involved in antigenic variation of the TprK antigen among treponemal species, subspecies, and strains. J Bacteriol **194**:4208–4225. http://dx.doi.org/10.1128/JB.00863-12.

32. **Reid TB, Molini BJ, Fernandez MC, Lukehart SA.** 2014. Antigenic variation of TprK facilitates development of secondary syphilis. Infect Immun **82**:4959–4967. http://dx.doi.org/10.1128/IAI.02236-14.

33. **Giacani L, Molini B, Godornes C, Barrett L, Van Voorhis WC, Centurion-Lara A, Lukehart SA.** 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. http://dx.doi.org/10.1128/IAI.01124-06.

34. **Petrosova H, Pospisilova P, Strouhal M, Cejkova D, Zobanikova M, Mikalova L, Sodergren E, Weinstock GM, Smajs D.** 2013. Resequencing of *Treponema pallidum* subsp. *pallidum* strains Nichols and SS14: correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters. PLoS One **8**:e74319. http://dx.doi.org/10.1371 /journal.pone.0074319.

35. **Brownstein MJ, Carpten JD, Smith JR.** 1996. Modulation of nontemplated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques **20**:1004–1006, 1008–1010.

36. **Giacani L, Sun ES, Hevner K, Molini BJ, Van Voorhis WC, Lukehart SA, Centurion-Lara A.** 2004. Tpr homologs in *Treponema paraluiscuniculi* Cuniculi A strain. Infect Immun **72**:6561–6576. http://dx.doi.org /10.1128/IAI.72.11.6561-6576.2004.

37. **Maniatis T, Sambrook J, Fritsch EF.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

38. **Giacani L, Godornes C, Puray-Chavez M, Guerra-Giraldez C, Tompa M, Lukehart SA, Centurion-Lara A.** 2009. TP0262 is a modulator of promoter activity of *tpr* subfamily II genes of *Treponema pallidum* ssp. *pallidum*. Mol Microbiol **72**:1087–1099. http://dx.doi.org/10.1111/j.1365 -2958.2009.06712.x.

39. **Giacani L, Denisenko O, Tompa M, Centurion A.** 2013. Identification of the *Treponema pallidum* subsp. *pallidum* TP0092 (σE) regulon: implications for pathogen persistence in the host and syphilis pathogenesis. J Bacteriol **195**:896–907. http://dx.doi.org/10.1128/JB.01973-12.

40. **Giacani L, Lukehart SA.** 2014. The endemic treponematoses. Clin Microbiol Rev **27**:89–115. http://dx.doi.org/10.1128/CMR.00070-13.

41. **Graves S, Downes J.** 1981. Experimental infection of man with rabbit-virulent *Treponema paraluis-cuniculi*. Br J Vener Dis **57**:7–10.

42. **Smajs D, Zobanikova M, Strouhal M, Cejkova D, Dugan-Rocha S, Pospisilova P, Norris SJ, Albert T, Qin X, Hallsworth-Pepin K, Buhay C, Muzny DM, Chen L, Gibbs RA, Weinstock GM.** 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. http://dx.doi.org/10.1371/journal.pone.0020415.

43. **Giacani L, Hevner K, Centurion-Lara A.** 2005. Gene organization and transcriptional analysis of the *tprJ*, *tprI*, *tprG* and *tprF* loci in the Nichols and Sea 81-4 *Treponema pallidum* isolates. J Bacteriol **187**:6084–6093. http://dx.doi.org/10.1128/JB.187.17.6084-6093.2005.

44. **Larsen SA.** 1999. Manual of tests for syphilis. American Public Health Association, Washington, DC.

45. **Castro AR, Esfandiari J, Kumar S, Ashton M, Kikkert SE, Park MM, Ballard RC.** 2010. Novel point-of-care test for simultaneous detection of nontreponemal and treponemal antibodies in patients with syphilis. J Clin Microbiol **48**:4615–4619. http://dx.doi.org/10.1128/JCM.00624-10.

46. **Castro AR, Mody HC, Parab SY, Patel MT, Kikkert SE, Park MM, Ballard RC.** 2010. An immunofiltration device for the simultaneous detection of non-treponemal and treponemal antibodies in patients with syphilis. Sex Transm Infect **86**:532–536. http://dx.doi.org/10.1136/sti.2010.042937.

47. **Workowski KA, Berman SM.** 2011. Centers for Disease Control and Prevention sexually transmitted disease treatment guidelines. Clin Infect Dis **53**(Suppl 3)**:**S59–S63. http://dx.doi.org/10.1093/cid/cir694.

48. **LaFond RE, Lukehart SA.** 2006. Biological basis for syphilis. Clin Microbiol Rev **19**:29–49. http://dx.doi.org/10.1128/CMR.19.1.29-49.2006.

49. **Saunders NJ.** 2003. Evasion of antibody responses: bacterial phase variation, p 103–124. *In* Henderson B, Oyston PCF (ed), Bacterial evasion of host immune responses. Cambridge University Press, Cambridge, United Kingdom.

50. **Smajs D, Norris SJ, Weinstock GM.** 2012. Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. Infect Genet Evol **12**:191–202. http://dx.doi.org/10.1016/j.meegid.2011.12.001.

51. **Saunders NJ, Jeffries AC, Peden JF, Hood DW, Tettelin H, Rappuoli R, Moxon ER.** 2000. Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. Mol Microbiol **37**:207–215. http://dx.doi.org/10.1046/j.1365-2958.2000.02000.x.

52. **Cameron CE, Lukehart SA.** 2014. Current status of syphilis vaccine development: need, challenges, prospects. Vaccine **32**:1602–1609. http://dx.doi.org/10.1016/j.vaccine.2013.09.053.

53. **Cox DL, Luthra A, Dunham-Ems S, Desrosiers DC, Salazar JC, Caimano MJ, Radolf JD.** 2010. Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of *Treponema pallidum*. Infect Immun **78**:5178–5194. http://dx.doi.org/10.1128/IAI.00834-10.

54. **Giacani L, Sambri V, Marangoni A, Cavrini F, Storni E, Donati M, Corona S, Lanzarini P, Cevenini R.** 2005. Immunological evaluation and cellular location analysis of the TprI antigen of *Treponema pallidum* subsp. *pallidum*. Infect Immun **73**:3817–3822. http://dx.doi.org/10.1128/IAI.73.6.3817-3822.2005.

55. **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. http://dx.doi.org/10.4049/jimmunol.169.2.952.

56. **Centers for Disease Control and Prevention.** 1999. The national plan to eliminate syphilis from the United States. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA.

57. **World Health Organization.** 2007. The global elimination of congenital syphilis: rationale and strategy for action. World Health Organization, Geneva, Switzerland.

58. **Radolf JD, Norgard MV, Schulz WW.** 1989. Outer membrane ultrastructure explains the limited antigenicity of virulent *Treponema pallidum*. Proc Natl Acad Sci U S A **86**:2051–2055. http://dx.doi.org/10.1073/pnas.86.6.2051.

59. **Radolf JD.** 1994. Role of outer membrane architecture in immune evasion by *Treponema pallidum* and *Borrelia burgdorferi*. Trends Microbiol **2**:307–311. http://dx.doi.org/10.1016/0966-842X(94)90446-4.

60. **Rappuoli R.** 2001. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine **19**:2688–2691. http://dx.doi.org/10.1016/S0264-410X(00)00554-5.

61. **Donati C, Rappuoli R.** 2013. Reverse vaccinology in the 21st century: improvements over the original design. Ann N Y Acad Sci **1285**:115–132. http://dx.doi.org/10.1111/nyas.12046.

62. **Brinkman MB, McGill MA, Pettersson J, Rogers A, Matejkova P, Smajs D, Weinstock GM, Norris SJ, Palzkill T.** 2008. A novel *Treponema pallidum* antigen, TP0136, is an outer membrane protein that binds human fibronectin. Infect Immun **76**:1848–1857. http://dx.doi.org/10.1128/IAI.01424-07.

63. **Cameron CE, Brown EL, Kuroiwa JM, Schnapp LM, Brouwer NL.** 2004. *Treponema pallidum* fibronectin-binding proteins. J Bacteriol **186**:7019–7022. http://dx.doi.org/10.1128/JB.186.20.7019-7022.2004.

64. **Cameron CE.** 2003. Identification of a *Treponema pallidum* laminin-binding protein. Infect Immun **71**:2525–2533. http://dx.doi.org/10.1128/IAI.71.5.2525-2533.2003.

65. **Houston S, Hof R, Francescutti T, Hawkes A, Boulanger MJ, Cameron CE.** 2011. Bifunctional role of the *Treponema pallidum* extracellular matrix binding adhesin Tp0751. Infect Immun **79**:1386–1398. http://dx.doi.org/10.1128/IAI.01083-10.

66. **Cameron CE, Lukehart SA, Castro C, Molini B, Godornes C, Van Voorhis WC.** 2000. Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies *pallidum* Tp92. J Infect Dis **181**:1401–1413. http://dx.doi.org/10.1086/315399.

67. **Giacani L, Chattopadhyay S, Centurion-Lara A, Jeffrey BM, Le HT, Molini BJ, Lukehart SA, Sokurenko EV, Rockey DD.** 2012. Footprint of positive selection in *Treponema pallidum* subsp. *pallidum* genome sequences suggests adaptive microevolution of the syphilis pathogen. PLoS Negl Trop Dis **6**:e1698. http://dx.doi.org/10.1371/journal.pntd.0001698.

68. **Pilsl H, Smajs D, Braun V.** 1999. Characterization of colicin S4 and its receptor, OmpW, a minor protein of the *Escherichia coli* outer membrane. J Bacteriol **181**:3578–3581.

69. **Nandi B, Nandy RK, Sarkar A, Ghose AC.** 2005. Structural features, properties and regulation of the outer-membrane protein W (OmpW) of *Vibrio cholerae*. Microbiology **151**:2975–2986. http://dx.doi.org/10.1099/mic.0.27995-0.

70. **Hu WS, Li PC, Cheng CY.** 2005. Correlation between ceftriaxone resistance of *Salmonella enterica* serovar Typhimurium and expression of outer membrane proteins OmpW and Ail/OmpX-like protein, which are regulated by BaeR of a two-component system. Antimicrob Agents Chemother **49**:3955–3958. http://dx.doi.org/10.1128/AAC.49.9.3955-3958.2005.

71. **Gil F, Ipinza F, Fuentes J, Fumeron R, Villarreal JM, Aspee A, Mora GC, Vasquez CC, Saavedra C.** 2007. The *ompW* (porin) gene mediates methyl viologen (paraquat) efflux in *Salmonella enterica* serovar Typhimurium. Res Microbiol **158**:529–536. http://dx.doi.org/10.1016/j.resmic.2007.05.004.

72. **Galdiero S, Falanga A, Cantisani M, Tarallo R, Della Pepa ME, D'Oriano V, Galdiero M.** 2012. Microbe-host interactions: structure and role of Gram-negative bacterial porins. Curr Protein Pept Sci **13**:843–854. http://dx.doi.org/10.2174/138920312804871120.

73. **Schirmer T.** 1998. General and specific porins from bacterial outer membranes. J Struct Biol **121**:101–109. http://dx.doi.org/10.1006/jsbi.1997.3946.

74. **Bishop RE.** 2008. Structural biology of membrane-intrinsic beta-barrel enzymes: sentinels of the bacterial outer membrane. Biochim Biophys Acta **1778**:1881–1896. http://dx.doi.org/10.1016/j.bbamem.2007.07.021.