Fine Analysis of Genetic Diversity of the \textit{tpr} Gene Family among Treponemal Species, Subspecies and Strains

Arturo Centurion-Lara$^{1,4}$, Lorenzo Giacani$^1$, Charmie Godornes$^1$, Barbara J. Molini$^1$, Tara Brinck Reid$^2$, Sheila A. Lukehart$^{1,3}$

1 Department of Medicine, University of Washington, Seattle, Washington, United States of America, 2 Pathobiology Graduate Program, University of Washington, Seattle, Washington, United States of America, 3 Department of Global Health, University of Washington, Seattle, Washington, United States of America

Abstract

\textbf{Background:} The pathogenic non-cultivable treponemes include three subspecies of \textit{Treponema pallidum} (\textit{pallidum, pertenue, endemicum}), \textit{T. carateum}, \textit{T. paraluiscuniculi}, and the unclassified Fribourg-Blanc treponeme (Simian isolate). These treponemes are morphologically indistinguishable and antigenically and genetically highly similar, yet cross-immunity is variable or non-existent. Although all of these organisms cause chronic, multistage skin and systemic disease, they have historically been classified by mode of transmission, clinical presentations and host ranges. Whole genome studies underscore the high degree of sequence identity among species, subspecies and strains, pinpointing a limited number of genomic regions for variation. Many of these “hot spots” include members of the \textit{tpr} gene family, composed of 12 paralogs encoding candidate virulence factors. We hypothesize that the distinct clinical presentations, host specificity, and variable cross-immunity might reside on virulence factors such as the \textit{tpr} genes.

\textbf{Methodology/Principal Findings:} Sequence analysis of 11 \textit{tpr} loci (excluding \textit{tprK}) from 12 strains demonstrated an impressive heterogeneity, including SNPs, indels, chimeric genes, truncated gene products and large deletions. Comparative analyses of sequences and 3D models of predicted proteins in Subfamily I highlight the striking co-localization of discrete variable regions with predicted surface-exposed loops. A hallmark of Subfamily II is the presence of chimeric genes in the \textit{tprG} and \textit{J} loci. Diversity in Subfamily III is limited to \textit{tprA} and \textit{tprL}.

\textbf{Conclusions/Significance:} An impressive sequence variability was found in \textit{tpr} sequences among the Treponema isolates examined in this study, with most of the variation being consistent within subspecies or species, or between syphilis vs. non-syphilis strains. Variability was seen in the \textit{pallidum} subspecies, which can be divided into 5 genogroups. These findings support a genetic basis for the classification of these organisms into their respective subspecies and species. Future functional studies will determine whether the identified genetic differences relate to cross-immunity, clinical differences, or host ranges.

Introduction

Non-cultivable pathogenic treponemes include three subspecies of \textit{Treponema pallidum}: \textit{T. pallidum} subsp. \textit{pallidum} (\textit{T. p. pallidum}), \textit{T. pallidum} subsp. \textit{pertenue} (\textit{T. p. pertenue}) and \textit{T. pallidum} subsp. \textit{endemicum} (\textit{T. p. endemicum}). These subspecies are human pathogens and cause venereal syphilis, yaws and bejel, respectively. Other very closely related species or isolates are \textit{Treponema paraluiscuniculi} and the Fribourg-Blanc or Simian treponeme. \textit{T. paraluiscuniculi} causes venereal syphilis in rabbits and is reportedly not infectious for humans [1,2]. The unclassified Simian treponeme was isolated from a baboon, causes a yaws-like disease in non-human primates, and is able to cause active infections in humans [3–5]. All of these organisms can be propagated in rabbits and cause disease following experimental inoculation of rabbits. \textit{Treponema carateum} causes the human disease, pinta, but no strains of this organism are available.

The infections caused by \textit{T. pallidum} organisms are characterized by chronic infection with distinct early and late clinical manifestations. Syphilis, usually a sexually transmitted infection, is a highly invasive process and can involve virtually any organ or system including the central nervous system. In pregnant women, early syphilis infection often results in transmission to the fetus. Each year, approximately twelve million new cases of syphilis are estimated to occur globally [6,7]. Yaws and bejel affect approximately 3 million people worldwide and are transmitted by non-sexual direct contact, usually during childhood and largely affecting people living in remote villages in developing countries. Yaws and bejel have predominantly skin or mucous membrane and osseous manifestations [8–10], with tissue destruction late in infection. Pinta causes significant skin discoloration in the late
Pathogenic treponemes include three subspecies of Treponema pallidum (pallidum, pertenue, endemicum), T. carateum, T. paraluiscuniculi, and the unclassified Fribourg-Blanc treponeme. Although they share morphology and have very similar antigenic profiles, they have traditionally been distinguished by mode of transmission, host specificity and the clinical manifestations that they cause. The molecular basis for these disease characteristics is not known. Comparative genomics has revealed that sequences differ among species and subspecies are found in very localized regions of the chromosome. Many of these regions of sequence variation are found in the tpr genes, which encode a family of twelve candidate virulence factors, many of which are predicted to be outer membrane proteins. Most of the tpr-specific sequence changes are consistent within subspecies or species, supporting the historical classification of these organisms into separate subspecies and species. Functional studies are needed to determine whether any of the tpr gene differences are related to differences in host range, immunity, or clinical manifestations.

The molecular basis for host specificity and the different clinical manifestations caused by these treponemes is not known. These organisms are morphologically identical [1,3,14–17] with very similar antigenic composition [18–23], stressed by the fact that, to date, infection-induced antibody or cellular immune responses cannot distinguish species, subspecies or strains. Protective immunity is induced only by long-term infection and is subspecies-specific [24]. In cross-immunity experiments [1] in which initial infections in the rabbit model lasted at least 3 months, three scenarios are observed: 1) inoculation with a particular strain results in complete protection against re-infection with the homologous strain, 2) protection against re-infection with another strain of the same subspecies is variable or non-existent, and 3) protection against challenge with other species or subspecies is absent. These cross-immunity observations are in concordance with inoculation studies in humans conducted by Magnuson et al. [25]. Subjects with treated late latent syphilis challenged with the Nichols strain had either of two outcomes: 1) those that did not develop either clinical signs or serological evidence of re-infection, indicating immunity; and 2) those that had increases in serological titers and/or development of darkfield positive lesions after inoculation, interpreted as active reinfection with the challenge strain. Although there was no evidence for waning immunity in the subjects who were susceptible to reinfection, this is a possible explanation. However, the lack of cross-immunity among highly similar species/subspecies may also reflect differences in a set of immunologically “inconspicuous” epitopes, underlying immunodominant, but not protective, antigens such as Tp47 (TP0574). These immunodominant antigens may act as decoy systems as described for other bacterial pathogens [26].

Recent comparative analyses of whole genome sequences [27–30] (Giacani et al., unpublished) reported <0.1% sequence differences among T. p. pallidum strains [29]; <0.2% between T. p. pertenue and T. p. pallidum subspecies [29]; and <1.2% between T. paraluiscuniculi and the human treponemes [31,32]. Sequence diversity is primarily localized to six hot spots [29], which include regions encoding several members of the tpr gene family. The Tpr proteins represent candidate virulence factors, and have been the focus of intense research for the last decade. As a consequence, the distinct clinical presentations, host specificities, and variable cross-immunity studies suggest that foci of sequence diversity, including the tpr genes, may be the basis for explaining the differences described above for the treponemal infections.

Sequence homology divides the tpr family, a group of twelve paralogs, into three subfamilies: Subfamily I (tpr C, D, F and I), Subfamily II (tprE, G and J) and Subfamily III (tprA, B, H, K and L). As we progressively gain a better understanding of this gene family, an essential role for many of these genes is more apparent. Several studies show that the Tpr antigens are expressed during infection and are able to elicit marked antibody and cellular immune responses in the infected host [33–40]. Of the encoded Tpr antigens, TprA, B, C, D, E, F, I, J and K have been predicted to be outer membrane proteins (OMP) [33,40,41]. Opsonization and/or vaccine studies with these proteins support surface exposure [33,37,42–44] and both antigenic variation (TprK) [45] and phase variation (TprE, G, J) [46] mechanisms have been identified in tpr members. Yet, the high invasiveness and ability of T. pallidum to persist for decades in the host suggest that this spirochete may rely not only on antigenic and phase variation for survival. To influence infection outcomes, T. pallidum may also employ other strategies including genetic drift, genetic shift and or pathoadaptive point mutations, which can arise either during long term evolution or rapidly during a single infection. An important body of evidence has accumulated showing genetic variation in specific regions of the T. pallidum genome among subspecies and among strains [47–56]. The present study demonstrates significant sequence diversity in the tpr gene family, which can have important implications in understanding evolution of these organisms, as well as cross-immunity, strain typing and vaccine design.

Materials and Methods

Ethics statement

No investigations were undertaken using humans or human samples in this study. New Zealand white rabbits were used for strain propagation. Animal care was provided in accordance with the procedures outlined in the Guide for the Care and Use of Laboratory Animals, and all work was conducted under protocols approved by the University of Washington Institutional Animal Care and Use Committee.

Bacterial strains and DNA extraction

T. pallidum subspecies, T. paraluiscuniculi, and the Fribourg-Blanc treponeme were propagated in New Zealand white rabbits by intratesticular inoculation as previously described [57]. DNA was extracted for PCR amplification from the following isolates: T. pallidum subsp. pallidum (Sea 81-4, Mexico A, Bal 3), T. pallidum subsp. pertenue (Gauthier, CDC2, Samoa D), T. pallidum subsp. endemicum (Iraq B, Bosnia A), the Fribourg-Blanc treponeme (Simian isolate) and T. paraluiscuniculi (Cuniculi A). These strains were selected to represent different species/subspecies, geograph-
ical regions of origin, years of isolation, and anatomical sources (Table 1). The sequences of the tpr genes for the T. p. pallidum Nichols and Street 14 strains were downloaded from their corresponding genome sequences, GenBank accession numbers NC_000919.1 and NC_010741.1, respectively [27,28]. Although we determined tpr sequences for a number of other strains of T. pallidum subsp. pallidum, only strains defining the 5 identified genogroups of T. p. pallidum are included in this manuscript. To ensure that the correct strain was propagated and extracted, only one strain of treponeme was handled at any time during the propagation and freezing process, and rabbit ear tags as well as labels on tubes were double-checked. Bacteria were extracted from infected rabbit testes in sterile saline, collected in sterile 1.7-ml microcentrifuge tubes, taking precautions to prevent cross-contamination between samples, and spun immediately in a microcentrifuge at 1,000 g for 10 minutes to remove rabbit debris, followed by centrifugation of the supernatant at 12,000 g for 30 min at 4 °C [57]. Pellets were resuspended in 200 µl of 1X lysis buffer (10 mM Tris [pH 8.0], 0.1 M EDTA, 0.5% sodium dodecyl sulfate), and DNA was extracted with the QiaGen (Chatsworth, Calif.) kit for genomic DNA extraction as described in the manufacturer’s instructions, but adding 50 ul of proteinase K (100 mg/ml stock solution) and incubating the sample for 2 h at 56 °C. After the final elution step in 200 µl of H2O2, DNA was used for analysis by PCR and sequencing.

PCR amplification, cloning, sequencing, sequence analysis and 3D models

The Nichols T. pallidum genome sequence [27] was used to design primers in the 5’ and 3’ flanking regions of the tpr genes to amplify the corresponding DNA regions from genomic DNA of the 10 treponemal strains. Table S1 lists the primers used for amplification and sequencing. Using genomic DNA as a template, whole ORF amplifications were performed in a 50-µl final volume containing 200 µM deoxynucleoside triphosphates, 1.5 mM MgCl2, and 2.5 U of GoTaq DNA polymerase (Promega, USA). Primers were designed in the 5’ and 3’ flanking regions of the tpr genes to amplify the corresponding DNA regions from genomic DNA of the 10 treponemal strains. Table S1 lists the primers used for amplification and sequencing. Using genomic DNA as a template, whole ORF amplifications were performed in a 50-µl final volume containing 200 µM deoxynucleoside triphosphates, 1.5 mM MgCl2, and 2.5 U of GoTaq DNA polymerase (Promega, USA). For larger amplicons such as the tprJ-I and tprK loci, long PCR was performed using the LongAmp Taq Kit as instructed by the manufacturer (New England Biolabs, USA). The products were cloned into the pCRII-TOPO or TOPO-XL (long amplicons) cloning vectors (Invitrogen, USA) according to the manufacturer’s instructions. Plasmid DNA was extracted by using the QiaGen Plasmid Minikit (QiaGen, USA), and two to ten clones for each strain were sequenced with the Applied Biotools dye terminator sequencing kit (Perkin-Elmer, USA). Consensus sequences were obtained with the CAP sequence assembly program [58] and ORFs from each strain at each locus were aligned using the MAFFT alignment program [59]. GenBank accession numbers are listed in Table S2. Structural homologs were identified using the 3D jury approach [60]. Structural (3D) models for TprC, TprD, and TprI were generated using the TMBpro algorithm [61]. The orientation of the predicted loops in the TMBpro models, surface exposed vs. periplasmic, was determined as previously described by Randall et al. [61]. Signal peptide predictions were performed using the Predisi algorithm [62].

Results

The tpr loci defined in the original Nichols genome sequence [27] were used to determine corresponding genes in 11 additional treponemal strains, including four T. p. pallidum, three T. p. pertenue, two T. p. endemicum, one T. paraluiscuniculi, and the Fribourg-Blanc strains (Table 1). TprK is excluded from this analysis because of the already extensive work that has been done on this gene [33,34,42,43,45,53,63–67]. Sequence analyses of the tpr loci from these strains identified significant heterogeneity within and among pallidum subspecies, the Fribourg-Blanc isolate and T. paraluiscuniculi. Figure 1 summarizes our findings. While Subfamilies I and II display a wide range of changes, diversity in Subfamily III is limited largely to the tprC and tprL loci. The observed changes are quite diverse, including SNPs of synonymous and non-synonymous character, indels, deletions of entire ORFs, chimeras, and alleles with large unique regions. Readers interested in the entire spectrum of sequence modifications identified in this study are referred to DNA and amino acid sequence alignments for each

Table 1. Treponemal strains reported in this study.

| Species, subspecies | Strain name | Source | Location | Year of isolation |
|---------------------|-------------|--------|----------|-------------------|
| T. p. pallidum      | Nichols     | Cerebrospinal fluid | Washington DC | 1912              |
|                     | Sea 81-4    | Primary chancre | Seattle     | 1980              |
|                     | Bal 3c      | Blood, congenital | Baltimore   | Unknown           |
|                     | Mexico A    | Primary chancre | Mexico      | 1953              |
|                     | Street 14d  | Skin    | Atlanta   | 1977              |
| T. p. pertenue      | Gauthier    | Skin    | Brazzaville | Early 1960’s      |
|                     | Samoa D    | Skin    | Western Samoa | 1953             |
|                     | CDCZ       | Skin    | Ghana     | 1980              |
| T. p. endemicum     | Bosnia A   | Penile ulcer | Bosnia     | 1950              |
|                     | Iraq B     | Oral mucous patches | Iraq | 1951         |
| Simian species      | Fribourg-Blanc | Lymph node | Guinea | 1966             |
| T. paraluiscuniculi | Cuniculi A  | ?       | Baltimore | Unknown           |

aOriginally provided by James N. Miller, University of California, Los Angeles, CA.
bStrain isolated in Seattle by Sheila A. Lukehart, University of Washington, Seattle, WA.
cStrains provided by Paul Hardy and Ellen Nell, Johns Hopkins University, Baltimore, MD.
dProvided by Sandra A. Larsen, Center for Disease Control and Prevention, Atlanta, GA.
eProvided by Peter Perine, Centers for Disease Control and Prevention, Atlanta, GA.
fProvided by Robert George and Victoria Pope, Center for Disease Control and Prevention, Atlanta, GA.

doi:10.1371/journal.pntd.0002222.t001
locus appended as Figure S1. For practical and comparison purposes, the tpr loci annotated in the Nichols genome sequence [27] will be considered as the reference ORFs.

It is noteworthy that many of the sequence changes divide the *T. pallidum* strains cleanly by subspecies or species: examples include no tprI ORF in *endemicum* strains; and tprG/J chimeras in both *tprG* and *tprJ* loci in *pertenue* strains while these two loci contain GI and GJ chimeras, respectively, in *endemicum* strains. In some cases, the sequences clearly divide syphilis vs. non-syphilis *T. pallidum* subspecies: as an example, the absolute conservation of tprF (with frameshift) and intact tprI sequences in the syphilis strains, while tprI-like sequences are found in these loci in non-syphilis strains. Within subspecies, the syphilis strains demonstrated the most heterogeneity, being divided into five genotypes.

### SUBFAMILY I: tprC and tprD loci and their encoded proteins

Subfamily I tprS include the tprC, D, F, and I loci. Initial examination of deduced protein alignments from the Nichols strain showed a significantly high degree of sequence conservation within Subfamily I at the amino and carboxyl termini, with central unique regions [33]; however, discrete heterogeneity was later evident in the amino and carboxyl regions when additional strains were analyzed. The Nichols TprC/D proteins reportedly have porin activity and an OM localization [68]. Although not yet experimentally demonstrated, TprF and I are also predicted to have a cleavable signal peptide and to be surface exposed [33,40,41,44].

| Species or subspecies          | Subgroup (Example Strain) | tprC locus | tprD locus | tprF locus | tprI locus | tprE locus | tprG locus | tprJ locus | tprA locus | tprB locus | tprH locus | tprL locus |
|-------------------------------|---------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| **T. p. pallidum**            |                           | tprC*      | tprD*      | tprF       | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Nichols                       |                           | tprC       | tprD2      | tprF       | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Bal 3                         |                           | tprC like  | tprD like  | tprF like  | tprI like  | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Sea 81-4                      |                           | tprC like  | tprD2      | tprF       | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Mexico A                      |                           | tprC like  | tprD2      | tprF       | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Street 14                     |                           | tprC like  | tprD2      | tprF       | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| **T. p. pertenue**            |                           | tprC like  | tprD like  | tprI * like| tprI * like| tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Gauthier                      |                           | tprC like  | tprD2      | tprF1      | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Samoa A                       |                           | tprC like  | tprD2      | tprF1      | tprI       | tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| CDC2                          |                           | tprC like  | tprD like  | tprI * like| tprI * like| tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| **T. p. endemicum**           |                           | tprC like  | tprD2      | No ORF     | tprF1      | tprI * like| tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Iraq B                        |                           | tprC like  | tprD2      | No ORF     | tprF1      | tprI       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Bosnia A                      |                           | tprC like  | tprD2      | No ORF     | tprF1      | tprI       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| **Simian**                    |                           | tprC like  | tprD2      | tprI * like| tprI * like| tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Fribourg-Blanc                |                           | tprC like  | tprD2      | tprI * like| tprI * like| tprE       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| **T. paraluciscuniculi**      |                           | tprC like  | tprD2       | No ORF     | tprF1      | tprI       | tprG/J     | tprA        | tprB       | tprH       | tprL       |
| Cuniculi A                    |                           | tprC like  | tprD2       | No ORF     | tprF1      | tprI       | tprG/J     | tprA        | tprB       | tprH       | tprL       |

Figure 1. *tpr* alleles among treponemal species, subspecies and strains. *tprD2*: A *tprD* allele which contains a 330-bp unique central region and three smaller heterogeneous regions at the 3’ end. *tprC-like* and *tprD-like*: similar to *tprC* or *tprD*, respectively, with small sequence differences in the N-terminal unique regions (DVRs). *tprG/J*: A chimera where the 3’ end contains *tprJ* signatures. *tprG/I*: A chimera where the 5’ end is homologous to *tprG*, and the central and 3’ regions are homologous to the corresponding regions of *tprI*. *tprA-like*, *tprI-like* and *tprL-like* contain small sequence differences. *tprE-like* and *tprH-like* contains small sequence differences that segregate syphilis from non-syphilis treponemes. *tpr2*: A unique *tpr*-allele in *T. p. pertenue* and Fribourg-Blanc strain. * indicates that *tprC* and *tprD* in the Nichols strain and that *tprI*-like sequences in the *tprF* and *tprI* loci are also identical in the *pertenue* subspecies and the Fribourg-Blanc treponeme.
doi:10.1371/journal.pntd.0002222.g001
the non-syphilis treponemes carries \( \text{tprC/D} \) ORFs identical to the Nichols strain. As previously reported by our group, \( \text{tprD} \) has four unique regions that differentiate it from Nichols \( \text{tprD} \) and the \( \text{tprD} \)-like sequences: a 330-bp central region and three smaller regions toward the end of the open reading frame (Figure 2) [56].

The \( \text{tprC} \) locus of the \( \text{tprD} \)-containing Bal 3, Sea 81-4, Street 14, Mexico A, Samoa D, Bosnia A, Iraq B, Fribourg-Blanc strains contains \( \text{tprC} \)-like ORFs, with small sequence changes compared to the Nichols \( \text{tprC} \) [37]. Overall, the sequence homology among \( \text{tprC} \) alleles is >95%. All \( \text{pertenue} \), \( \text{endemicum} \) and the Fribourg-Blanc strains also have \( \text{tprC} \)-like sequences. As previously reported for \( \text{T. paraluiscuniculi} \), Cuniculi A strain [31,32,36], the \( \text{tprC} \) and \( \text{D} \) loci are occupied by two truncated \( \text{tprD} \) variants.

In both \( \text{tprC} \) and \( \text{tprD} \), sequence variation does not occur randomly, but rather is found in discrete variable regions (DVRs); the same background color indicates sequence identity. Vertical blue lines indicate discrete variable regions (DVRs), which contain mutations in great majority of non-synonymous character. Green color indicates unique \( \text{tprD} \) signatures. Light blue background indicates predicted untranslated regions of the ORFs in the \( \text{tprD} \) alleles due to a single nucleotide insertion, frameshifting and a premature stop.

doi:10.1371/journal.pntd.0002222.g002

**SUBFAMILY I: \( \text{tpr F} \) and \( \text{I} \) loci and their encoded proteins**

In the Nichols genome, \( \text{tprF} \) and \( \text{tprI} \) loci are 1107 and 1827 nucleotides long, respectively. Their sequences are identical except that \( \text{tprF} \) is a truncated version of \( \text{tprI} \) due to a 720 nucleotide deletion (spanning the central and most of the 3’ region) in \( \text{tprF} \), resulting in a shorter ORF, frameshifting and a premature termination [27]. In \( \text{T. p. pallidum} \) strains, \( \text{tprF} \) genes are identical in all isolates sequenced to date (Figure 1 and Supplemental Figures 1 and 2.3 in Figure S1 and S2). In contrast to the syphilis strains, the \( \text{pertenue} \) and Fribourg-Blanc isolates have a full length (not frameshifted) duplicated \( \text{tprI} \) gene at the \( \text{tprF} \) locus.
Figure 3. Structural models of TprC/D and TprI. Non-templated 3D models generated for the mature Nichols TprC/D and Nichols TprI peptides using the TMBpro algorithm [61] suggest a typical β-barrel structure. DVR, discrete variable regions. EL, external loops. Variable regions, DVR1–DVR7 for TprC/D and DVR1–DVR9 for TprI, as defined by protein sequence alignments (Figure S1 and S2) are indicated by red color (loops, font and arrows). Note that each DVR co-localizes with a predicted EL. Orientation of the structure was determined as specified by Randall et al [61]. Proposed conserved and variable surface exposed loops are highlighted in blue and red, respectively, and proposed periplasmic exposed regions of the proteins are in purple.

doi:10.1371/journal.pntd.0002222.g003
Interestingly, however, the tprF locus is deleted in the endemicum strains Iraq B and Bosnia A, and in T. paraluiscuniculi. tprI loci are virtually identical to each other in T. p. pallidum strains except for the presence of a few synonymous SNPs in the 5’ and central regions reported in Street 14 [28]. In contrast, however, tprF or tprI ORFs are absent in the rabbit pathogen T. paraluiscuniculi. [36].

For a more detailed analysis of the polymorphism observed in the tprF and tprI loci, a sequence alignment was generated including all genuine (not truncated or replaced) tprF and tprI loci from 11 strains (Supplemental Figure 2.3 in Figure S2). TprF and TprI in syphilis and non-syphilis organisms display DVR patterns resembling the heterogeneity observed in TprC and TprD above, though to a lesser extent (Supplemental Figure 2.3 in Figure S2). Changes are clustered in 9 DVRs spread throughout the protein sequences. Deduced TprF and TprI proteins are also predicted to be outer membrane proteins [33,40,41,68]. Structural predictions also suggest that TprF and TprI are homologs of transport porins with OM localization, and 3D predictions of TprF/TprI peptides without signal peptides (Figure 3 bottom panel, and Supplemental Figure 2.5 in Figure S2), again suggesting an important role for these variable regions during infection.

SUBFAMILY II: tprE, G and J loci and their encoded proteins

The Subfamily II genes include tpr E, G, and J, which code for proteins nearly 800 amino acids in length with highly conserved amino termini, unique central regions and carboxyl ends with small unique gene-specific signatures [33]. tprE shows very limited sequence variation among strains and subspecies, however, the observed changes clearly segregate syphilis from non-syphilis treponemes and the Fribourg-Blanc strain (Supplemental Figure 1.4 in Figure S1). T. paraluiscuniculi has a tprGJ chimera (predicted truncation) in the tprE locus [36].

In contrast, the tprG locus is more diverse in its gene sequence, in that five different groups of ORFs can be found (Figure 1, Figure 4 and Supplemental Figure 1.6.1 and 1.6.2 in Figure S1): 1) tprG sequences as described in the Nichols genome (Bal 3, Street 14); 2) a truncated tprG due to two single and one 3-nucleotide insertions (position range 1885–1956), frameshifting, and a premature stop at its 3’ end (Sea81-4); 3) a tprGJ chimera, in which the 3’ end of tprG has been replaced by the corresponding

![Figure 4. tprG and tprJ loci. Top: proteins encoded in the tprG locus. Bottom, proteins encoded in the tprJ locus. Regions of the same color indicate sequence identity among gene products. Five different variants can be identified among the 12 strains analyzed in the tprG locus. Nichols, Bal3 and Street 14 encode the Nichols reference TprG. Sea81-4encodes a truncated TprG. The TprGJ chimera is found in Mexico A, Samoa D, Gauthier and CDC2; the TprGI chimera is found in Iraq Band Bosnia A, and a truncated TprGJ chimera is found in Cuniculi A and Fribourg-Blanc. At the tprJ locus, Nichols, Bal3, Mexico and Street 14 contain the Nichols TprJ. Sea81-4 and all non-syphilis strains carry theTprGJ hybrid. Green, signal peptide. doi:10.1371/journal.pntd.0002222.g004](image-url)
region of \textit{tprJ} as evidenced by the presence of a \textit{tprJ}-specific signature \texttt{TAAACGGGACACCCTCCTCCTCCGCGGTTCCTGAGGGACATTTGGCCT} near the 3’ end of \textit{tprJ} (Mexico A and all \textit{T. p. pertenue} strains); 4) a \textit{tprGI} chimera in which the 5’ end of the ORF is homologous to the corresponding region of \textit{tprG}, and its central and 3’ regions of the gene are homologous to the corresponding regions of \textit{tprI} (all \textit{T. p. endemicum} strains); and 5) a truncated \textit{tprGI} chimera due to a single nucleotide insertion (\textit{T. paraluiscuniculi} and the Fribourg-Blanc strain).

While four \textit{T. p. pallidum} strains have the reference Nichols \textit{tprJ} sequence, the \textit{T. p. pallidum} Sea 81-4 strain and all non-\textit{pallidum} treponemes studied to date contain a \textit{tprGJ} chimera in the \textit{tprJ} locus (Figure 1 and Figure 4). The rabbit pathogen, however, contains a \textit{tprGJ} chimera that codes for a truncated protein due to an insertion in its 5’ end [36].

\textbf{SUBFAMILY III: \textit{tprA}, \textit{B}, \textit{H}, and \textit{L} loci}

Subfamily III \textit{tprs} show a reduced degree of homology among family members, compared to Subfamilies I and II, with only small regions of sequence identity scattered throughout the coding sequences [27,33]. This is contrasted by a lower level of sequence heterogeneity at each locus among strains, subspecies, and species. \textit{tprB} shows no variation among all strains (Supplemental Figure 1.2 in Figure S1). Among strains and subspecies, the \textit{tprH} locus also contains highly homologous sequences, with only a few point mutations, of which 3 SNPs consistently distinguish syphilis vs. non-syphilis organisms (Supplemental Figure 1.7 in Figure S1).

In the \textit{tprA} locus, at positions 706 to 711, there is a short region containing either three or four CT dinucleotide repeats. Strains containing only three CT repeats carry a gene that codes for a truncated protein due to a frameshift leading to a premature stop (Nichols, Mexico A, Street 14 and Bal 3). In contrast, strains carrying \textit{tprA} genes with four CT repeats (the syphilis Sea 81-4 and all non-syphilis isolates) have no predicted frameshift and generate a sequence encoding a full length TprA product (Figure 5).

\textit{tprL} (\textit{tp1031}) shows major changes among strains and subspecies. Re-analysis of this region in the all of the \textit{endemicum} and \textit{pallidum} strains and 8 additional syphilis strains (Brinck Reid et al., unpublished) revealed a larger putative \textit{tprL} ORF coding for a protein sequence of 602 amino acids, compared to 514 amino acids as previously reported for the Nichols and Street 14 strains [27,28]. In this extended ORF (Figure 6), an alternative start codon (CTG) was identified with a typical ribosomal binding site (RBS, GGAGG). Furthermore, beginning at position −31, a 15 to 17 nucleotide poly-G tract flanked by −10 and −35 σ70 signatures (TAGACA and TGTTGT) is evident (Figure 6). Unlike the TprL product annotated in the Nichols genome sequence, the extended TprL is predicted to have a putative OM localization, with a predicted cleavable signal peptide (cleavage between positions 25 and 26, VFS-EQ). Compared to \textit{T. p. pallidum} and \textit{T. p. endemicum} sequences, our analysis revealed a gene fusion in the \textit{T. p. pertenue} and Fribourg-Blanc strains caused by a deletion of 278 nucleotides (Figure 6), encompassing the 5’ end and central regions of the \textit{tp1030} ORF and a small fragment of the 5’ end of \textit{tprL} including its start codon. This deletion creates a hybrid sequence (\textit{tp1030} and \textit{tprL}, here called \textit{tprL1}) of 1668 bp with the start codon (ATG) in the plus strand of \textit{tp1030} (the \textit{tp1030} coding sequence is located on the minus strand of the chromosome) in frame with the rest of \textit{tprL} (\textit{tp1031}). As a consequence, the first 130 nucleotides of this new
Pertenue tprL1 (Figure 1, Figure 6 and Supplemental Figure 1.9 in Figure S1) are unique, not found in T. p. pallidum or endemicum tprL. The new extended TprL (in T. p. pallidum and T. p. endemicum and T. paraluiscuniculi) and the newly predicted TprL1 proteins (T. p. pertenue and the Fribourg-Blanc treponeme) are 602 and 556 amino acids long, respectively. Because the first 44 amino acids of TprL1 are encoded by the plus strand, this region is unique to the yaws and simian strains, with no homologous peptide in the pallidum and endemicum proteins (Supplemental Fig. 1.9 Figure S1). This unique peptide sequence is also not found elsewhere in the chromosome. Unlike the newly predicted extended TprL, TprL1 does not have a predicted signal peptide (Figure 6). This raises the possibility that the pallidum and endemicum subspecies may have an OM-localized TprL, while this would be predicted to be absent in the pertenue subspecies.

**Discussion**

The 12 treponemal isolates from the three T. pallidum subspecies (pallidum, pertenue and endemicum), the Fribourg-Blanc treponeme, and T. paraluiscuniculi show pleomorphic genetic changes in the tpr family characterized by SNPs, indels, chimeric sequences, and even absence of entire ORFs. Initial comparisons of the currently available full genome sequences of the Nichols, Chicago C, Sea81-4 and Street 14 syphilis strains revealed a high degree of sequence identity and a remarkable conservation of their genome organization [30] (and Giacani et al., unpublished). The study by Mikalova et al. [29] confirmed these observations, reporting clustering of sequence divergence in only a handful of distinct genomic regions among syphilis and non-syphilis strains, similar to those identified previously by Weinstock and colleagues.
Many of the hot spots of diversity are located in genes encoding members of the Tpr antigen family. The present study, however, provides a detailed description of sequence diversity within this paralog family and uncovers a rich number of sequence modifications among species, subspecies and strains. Importantly, our analyses also indicate some alternative genes or modified loci.

It is striking that much of the sequence diversity identified in the tpr genes segregates the strains into the same subspecies and species groups that were originally defined according to their modes of transmission, their natural hosts, and the diseases they cause. This is most effectively seen in the colored blocks in Figure 1. Given that the tpr loci represent the primary regions comprising the extremely low genomic diversity among the T. pallidum subspecies, it is likely that the proteins encoded by these variant genes play a major role in the differing pathogenesis of syphilis vs. yaws vs. endemic syphilis. Assigning a definitive role for individual proteins or combinations of proteins in determining clinical outcomes, however, awaits the determination of the functions of the Tpr proteins and the ability to genetically manipulate these genes within the organism. To inform studies of possible location and function, computational and immunological studies can provide clues for individual gene products.

Several arguments emphasize a key role for TprC and TprD during syphilis infection: 1) they are the targets of strong antibody and cellular immune responses [35,37,40,56]; 2) immunization with recombinant TprC/D induces partial protection against infectious challenge [37]; 3) their surface exposure is supported by opsonophagocytosis assays [68] (Lukehart et al., unpublished); 4) TprC and D show sequence diversity among strains [37] (and this study); and 5) 3D models predict a typical β-barrel structure with surface-exposed loops that contain each of the regions where sequence diversity is localized (this study). It is highly unlikely that the co-location of sequence diversity and predicted surface-exposed loops is coincidental. A recent study by Anand et al. [68] proposes an alternative model for TprC and TprF, suggesting that the amino terminus of these two proteins is localized in the periplasmic space. However, experimental evidence argues against this model. Recombinant amino terminal TprF/I peptide induces partial protection against homologous challenge in immunization experiments in the rabbit model [37] and elicits opsonizing antibodies upon immobilization (Lukehart et al., unpublished), observations supportive of surface exposure. However, the TprC and D sequence diversity localized in the exposed DVR identified among subspecies in the present study may contribute to the variable degree of cross-protection observed among T. pallidum strains and subspecies in infection-induced immunity. In this context, it is possible that sequence differences in the DVRS of TprC and D could lead to subspecies- or strain-specific surface-exposed epitopes that are critical to opsonic function or other mechanisms of protection. Studies are ongoing to test this hypothesis. A recognized example of functionally important strain-specific epitopes is loop 5 of the OMP P2 protein of non-typeable Haemophilus influenzae, which is associated with elicitation of bactericidal antibodies and protective immunity [70]. An alternative, or complementary, function of variable surface-exposed loops (e.g. DVR) could be that of providing steric hindrance to prevent the immune system from recognizing conserved external loops on the antigen, which are perhaps essential for correct protein structure or function. It is noteworthy that TprC and TprD are each predicted by 3D analysis to contain 4 conserved external loops.

During natural human infection and experimental infection of rabbits [37,56], antibodies are made against TprC/D and TprD2. In addition to TprC and D, the TprD2 variant is also predicted to have surface exposure [37,40,56], and is found in both syphilis and non-syphilis treponemes (Figure 1). The regions unique to TprD2 also contain predicted external loops, thus adding another layer of complexity to the already existing set of predicted loops for TprC and D (not shown). Our structural predictions of TprC/D showing co-localization of external loops with DVRS is strong support for our hypothesis that antigenic differences in surface exposed loops of TprC and D have functional significance in immunity to the T. pallidum subspecies, and may be determinants of cross-immunity among subspecies and strains.

Of interest is the observation that the CDC2 strain maintained in Seattle (originally obtained in 2005 from Rob George and Virginia Pope from the Centers for Disease Control in Atlanta, GA) contains a tprD-like allele while the corresponding sequence reported by Mikalova et al. [29] contains a tprD2 sequence. Re-sequencing of the tprD locus of this strain using our original frozen stocks confirmed that the CDC2 strain indeed contains a tprD-like allele. Also, we have sequenced the tprD locus of the pertenue CDC1 strain, isolated in a neighboring village in Africa from where the CDC2 strain was obtained, and found that the CDC1 strain also contains a tprD-like gene. It may demand a significant effort to identify the source of discrepancy between our data and that of Mikalova et al., perhaps requiring the analysis of the two CDC2 lineages over the last several years.

In contrast to syphilis treponemes, the tprF and I loci in T. p. pertenue and the Fribourg-Blanc treponemes each contain identical full-length ORFs. Although their coding sequences are identical within each location, tprF and tprI are located in separate tprG-F and tprJ-I operons, respectively, and their expression may be differentially modulated. The number of G residues in a polyG string in their promoters controls phase variation of these operons [46], and the binding of TpCRP (Tp0262) to the promoters was shown to either increase (tprJ) or decrease (tprG) transcription of the operon [71]. The implications of a “double dose” of tprF in the non-pallidum strains might be reflected in the total amount of message made in tissue specific locations or in differential expression over time during infection. Preliminary studies of antibody reactivity in rabbits infected with T. p. pertenue Gauthier strain demonstrate high levels of antibody to TprI, consistent with high (or double) expression of the protein (Lukehart et al., unpublished). The strong resemblance of the TprF/F 3D predictions to the TprC/D structural models, and the co-localization of DVRS and external loops suggest analogous roles at the microbe-host interface.

T. pallidum tprGI chimeras were identified by Giacani et al. [36] in T. paraluiscuniculi and also present in the whole genome sequences later reported by Strouhal et al. [31] and Smajs et al. [32], whose unique sequence composition was also recognized by these authors. Our analysis shows that, in all strains of T. p. pertenue, T. p. endemicum and the Fribourg-Blanc treponeme, the G and J loci are occupied by either tprGJ or tprGJ chimeric genes. In contrast, the Nichols reference tprG and tprJ genes are frequently found in syphilis isolates, but not in any pertenue, endemicum or the Fribourg-Blanc strains tested to date. Only the T. p. pallidum Mexico A and Seattle 81-I4 strains carry the GJ chimeric gene in the tprG and tprJ loci, respectively. Of interest is the presence of three truncated chimeras encoded by the tpr E, G, and J loci in T. paraluiscuniculi. This, in addition to predicted truncations or absences of Subfamily I Tprs (Figure 1), is perhaps related to the inability of T. paraluiscuniculi to infect humans, although further study is needed to explore this issue more thoroughly.

One might wonder whether the tpr chimeras identified in this study are artificial, due to “jumping” between highly similar
sequences during PCR amplification [72–74]. In our study, tpr chimeras are unlikely to be artifacts for two reasons: 1) independent PCR amplifications of treponemal DNA obtained from different strain harvests rendered identical sequences, and 2) published sequences obtained by multiple sequencing approaches also show the same chimeras [31,32,36,75].

With the exception of TprK, little is known about the other members of Subfamily II tprs (tprA, tprB, tprH, and tprL). TprA, B and L are predicted to be OMPs [40,41], and Tpr B induces antibodies that promote opsonophagocytosis [Lakehart et al., unpublished]. Sequence conservation of tprB and tprH across species, subspecies, and strains suggests a required function for these proteins in the biology of T. pallidum. Nucleotide repeats, whether in regulatory or coding regions, are frequently associated with modulation of gene expression in an ON-OFF manner. The structure of the promoter region of the newly proposed extended tprL ORF is highly reminiscent of modulation of gene expression by single nucleotide repeats in the promoters of porA and opc loci of Neisseria meningitidis [76–78]. One could argue that predictions of an extended tprL ORF may lack accuracy because of the assumption of CTG as start codon, an underrepresented start codon in T. pallidum. The variable lengths resembling classic bacterial phase variation systems. In tprA, the variable number of CT dinucleotide repeats creates frameshifting and premature termination, dividing strains carrying tprA genes coding for full length product from those encoding predicted truncated products (Figure 5 and Supp 1). This is another mechanism for possible phase variation.

Our analysis of the tpr gene sequences is based on an approach of targeted PCR amplification, cloning, and sequencing a number of clones to obtain consensus sequences. The tpr ORF sequences appear to be unchanging within a given strain during infection. However, limited information at the population level invites speculation about the possible presence of genetically distinct subpopulations within isolates. Smajs et al. [28,79] reported that at least two subpopulations are present within the Nichols strain as defined by a ~1 Kb deletion in the flanking region of tpo131. Our approach could have overlooked underrepresented variant organisms within isolates and, if intrastrain variation indeed exists, our findings might then reflect amplification of the most predominant subpopulation. Small mutational changes, even SNPs, in coding or non-coding regions can affect transcription, translation, or folding of the protein themselves, of neighboring genes, or those at more distant sites [80–83]. This could explain, for example, some of the differences in transcription observed among treponemal strains [40]. On the other hand, the now standard use of template-based assembly of short stretches of sequence generated by newer sequencing technologies can overlook the existence of hybrid genes or missing ORFs, whereas our individual-ORF sequencing approach can clearly identify these variations. Renewed efforts to address all of the above questions may be effectively resolved using next generation approaches such as deep sequencing of targeted regions, single cell isolation, or whole transcriptome sequencing.

How might knowledge of tpr sequence diversity be translated into tools that are relevant to persons who are infected with one of the pathogenic treponemes? The geographical distribution of yaws and syphilis is not as distinct as decades ago, and travel or migration can serve to transport an infection between urban and rural settings, complicating diagnosis. Because of the re-emergence of yaws over the past 20 years [84], etiological differentiation of yaws vs. syphilis infections is desirable, and a practical approach for diagnosis is needed. The overall reported genetic variability between syphilis and yaws treponemes (0.2%) makes these organisms almost genetically indistinguishable, and existing serological tests fail to differentiate the infections. Several small signatures that differentiate the distinct species/subspecies have already been identified in several genes [47–49,31,32,95]. The unique sequence composition of TprL described here in pertenue vs. pallidum strains reveals a possible 90 amino acid sequence unique to non-yaws treponemes, which includes a 23 amino acid predicted signal peptide, as well as a 44 amino acid peptide specific to T. p. pertenue. Given that Giacani et al. [40] showed that the tprL ORF is actively transcribed in both syphilis and yaws treponemes during experimental infection, our findings could facilitate the development of targeted serological screening for differentiating these two infections.

Treponemal infections are chronic, yet only a minority of infected persons develops the severe late manifestations of disease. Is it possible that small genetic markers in the infecting could predict clinical outcome? We previously showed that rabbits infected intravenously with the Sea 81-4 strain had higher levels of cerebrospinal fluid (CSF) inflammation, compared to other infecting strains, while animals infected with Bal7 had more severe skin disease [86]. Our recent work in humans supports the hypothesis that disease outcome may be related to genetically defined strain types [55]. Subfamily II tprs and the asp genes were first utilized for strain typing purposes by Pillay et al. [50], although they were not able to correlate strain type with clinical outcome. Using an enhanced strain typing system developed by Marra et al. [55], which includes the targets initially described by Pillay et. al. [50] and the tph/psbB gene, we demonstrated that patients infected by 164/1/554/1 type strains were significantly more likely to have neurosyphilis [55]. Four of the pallidum strains shown to represent different genotypes in this report (Nichols, Street 14, Mexico A and Sea 81-4) fall into four different molecular types using the enhanced typing system. The correlation supports the possibility that sequence changes in the tpr genes may be related to specific disease manifestations.

It is noteworthy that T. paraluiscuniculi causes a very mild infection in its natural host, compared to syphilis, and is unable to infect humans [2,12,87]. One possible explanation for mild natural infection and the failure to infect other hosts is the dearth of functional Tpr proteins in this organism: there are seven ORFs of the same type), although it resembles T. p. pertenue. In contrast, all T. pallidum subspecies and the Fribourg-Blanc treponemes, which have fuller Tpr repertoires, can multiply in more than one vertebrate host and can cause infection in humans. The Fribourg-Blanc treponeme, isolated from non-human primates from a yaws-endemic region in Africa [3,4], resembles very closely the tpr repertoire of yaws strains (10 out of 12 ORFs are of the same type), although it resembles T. p. endemicum at the G locus, implying shared evolutionary pathways, as previously proposed [29,65,88], as well as common strategies of interaction between microbes and their host.

Although the clinical outcome of infection is likely dependent upon several factors, including individual host immunity, inoculum size, and route of infection, sequence changes in the tpr genes could determine differences in antigenicity or function, resulting in different adaptive strategies and differences in pathogenicity. While the distribution of tpr gene variants among the 12 isolates studied here appears, in most cases, to be clustered by subspecies, some isolates in the T. p. pallidum group share tpr variants that are otherwise restricted to non-syphilis organisms. For example, Sea 81-4 contains four tpr ORFs present in the endemicum subgroups (Figure 1), and Mexico A contains the tprG chimera in the tprG locus. The recent demonstration of syphile-like genital lesions and purported sexual transmission of a yaws-like treponeme in wild baboons [13] suggests that pathogenicity and mode of transmission may not, however, be completely hard-wired in the genome. The sharing of some tpr variants among individual pallidum strains and the non-pallidum strains confounds the concept of a purely
genetic basis for the nature of the disease. These findings again raise the 1960’s nature vs. nurture controversy between Hudson and Hackett with regard to the biological or environmental/epidemiological basis for the differing clinical manifestations seen among the treponematoses [99,100]. Based upon tpr sequencing, there is genetic heterogeneity (five genogroups) within the *T. pallidum* subspecies, as well as some overlap among subspecies and species. Rather than having discrete organisms for each treponemal disease, there may in fact be a genetic continuum of the pathogenic *Treponema*, individual components of which affect pathogenesis in an individual host in concert with social or environmental factors that influence routes of transmission and disease manifestations. Finding the answer to this question will depend upon the ability to genetically manipulate *T. pallidum* so that the effects of individual genes can be definitively assessed.

Supporting Information

**Figure S1** Predicted full length amino acid and DNA sequence alignments of the *tpr* gene family by locus. For the ORFs containing indels resulting in frameshifts and truncated proteins, only the encoded amino acid sequences before the premature stop codon are shown. In *tprD* locus, *tprD2* alleles were excluded for clarity purposes only. In *tprG* locus, because of the significant dissimilarities among sequences, alignments are separated in two groups: *GJ* and *Gji* chimeras. Red font, *T. p. pallidum* subspecies; blue *T. p. pertenue*, bright green, *T. p. endemicum*; yellow, the Simian treponeme; and pink, *T. paraluiscuniculi*. (DOCX)

**Figure S2** Alignment of amino acid sequences of the predicted protein sequences encoded at the *tprD* (2.1), *tprC* (2.2), and *tprF* and *tprI* (2.3) loci. The TprD2 truncated proteins encoded by the *tprD/C/D* loci in *T. paraluiscuniculi*, as well as by the *tprF* locus in *T. p. pallidum* strains are not included in the alignment for clarity purposes. Also, no *T. paraluiscuniculi* *tprF* and *tprI* ORFs are included because the *tprF* or *tprI* coding sequences are absent in the rabbit pathogen. DVR: Discrete variable regions. EL: External loops predicted by 3D models. SP: Predicted signal peptide. The last letter on the left column (strain name) indicates the locus where the predicted protein sequence is encoded. Red font, *T. p. pallidum* subspecies; blue *T. p. pertenue*, brilliant green, *T. p. endemicum*; and yellow, the Simian treponeme. (DOCX)

**Table S1** Primers used for amplification (1) or sequencing (2).

(DOCX)

**Table S2** GenBank accession numbers per locus and strain.

(DOCX)

**Author Contributions**

Conceived and designed the experiments: ACL SAL. Performed the experiments: LG CG BJM TBR. Analyzed the data: ACL SAL LG. Wrote the paper: ACL SAL LG.

**References**

1. Turner TB, Hollander DH (1957) Biology of the Treponematoses. Geneva: World Health Organization.

2. Graves S, Downes J (1981) Experimental infection of man with rabbit-virulent *Treponema pallidum subsp. carคลidum*. Br J Vener Dis 57: 7–10.

3. Fribourg-Blanc A, Mollaret HH, Niel G (1966) [Serologic and microscopic confirmation of treponemosis in Guinea baboons]. Bull Soc Pathol Exot Filiales 59: 54–56.

4. Fribourg-Blanc A, Mollaret HH (1969) Natural treponematoses of the African primate. Primates Med 3: 113–121.

5. Smith JL, David NJ, Indjin S, Israel GW, Levine BM, et al. (1971) Neuro-ophthalmological study of late yaws and pinta. II: The Caracas Project. Brit J Vener Dis 47: 226–251.

6. Gerbase AC, Rowley JT, Heymann DH, Berkley SF, Post P (1998) Global prevalence and incidence estimates of selected curable STDS. Sex Transm Infect 74: S12–16.

7. World Health Organization DoKHaR (2007) The global elimination of congenital syphilis: rationale and strategy for action. http://www.who.int/reproductivehealth/publications/ris/9789241595830/en/

8. Meheus A, Antal GM (1992) The endemic treponematoses: not yet eradicated. World Health Stat Q 45: 228–237.

9. Antal GM, Lukehart SA, Meheus AZ (2002) The endemic treponematoses. Microbes Infect 4: 83–94.

10. Asiedu K, Amouzou B, Dharwal A, Karam M, Lobis D, et al. (2000) Yaws eradication: past efforts and future perspectives. Bull World Health Organ 84: 499–499A.

11. Roman GC, Roman LN (1986) Occurrence of congenital, cardiovascular, visceral, neurologic, and neuro-ophthalmologic complications in late yaws: a theme for future research. Rev Infect Dis 8: 760–770.

12. Di Giacomo RF, Lukehart SA, Talburt CD, Baker-Zander SA, Coudron J, et al. (1984) Clinical course and treatment of venereal spirochaetosis in New Zealand white rabbits. Br J Vener Dis 60: 214–218.

13. Knauf S, Batamuzi EK, Mlengeya T, Kilewo M, Lejora IA, et al. (2012) Footprint of positive selection in *Treponema pallidum* subspp. pallidum and *T. p. endemicum*. PLoS Negl Trop Dis 6: e15713.

14. Ocvirkov NM, Deletorskij VV (1970) *Treponema pertenue* under the electron microscope. Br J Vener Dis 46: 349–359.

15. Hovind-Hougen K (1976) Determination by Means of Electron Microscopy of *Treponema pertenue* and T. paraluiscuniculi chimeras. Red font, *T. p. pallidum* subspecies; blue *T. p. pertenue*, bright green, *T. p. endemicum*; and yellow, the Simian treponeme. (DOCX)

16. Sepetjian M, Guerra P, Sahaseta D, Thivolet J, Monier JC (1969) Contribution a l’étude du treponeme isole du singe par A. Bull World Health Organ 40: 141–151.

17. Frasch CM, Norris SJ, Weinstock GM, White O, Sutton GG, et al. (1998) Genome analysis of *Treponema pallidum* subsp. pallidum strains: most of the genetic differences are localized in six regions. PLoS One 5: e15713.

18. Sepetjian M, Guerra PT, Sahaseta D, Thivolet J, Monier JC (1969) Contribution a l’étude du treponeme isole du singe par A. Bull World Health Organ 40: 141–151.

19. Thornburg RW, Bauman JB (1983) Comparison of major protein antigens and protein profiles of *Treponema pallidum* and *T. p. endemicum*. Infect Immun 42: 623–627.

20. Baker-Zander SA, Lukehart SA (1983) Molecular basis of immunological cross-reactivity between *Treponema pallidum* and *T. p. pertenue*. Infect Immun 42: 654–656.

21. Baker-Zander SA, Lukehart SA (1984) Antigenic cross-reactivity between *Treponema pallidum* and other pathogenic members of the family Spirochaetaeae. Infect Immun 46: 116–121.

22. Mein P, Cockayne A, Georgs AF, Penn CW (1990) Immune response to *Treponema pertenue* and *T. p. endemicum* Nichols in patients with yaws. Res Microbiol 141: 181–186.

23. Martin MR, Cockayne A, Schouls LM, Melcon RH, Stole E, et al. (1990) A new attempt to distinguish serologically the subspecies of *Treponema pallidum* causing syphilis and yaws. J Clin Microbiol 28: 1600–1607.

24. Magnuson HJ, Rosenau BJ (1948) The rate of development and degree of acquired immunity in experimental syphilis. Am J Syph Gonorreae Vener Dis 32: 410–436.

25. Magnuson HJ, Thomas EW, Okamsky S, Kaplan BI, DeMello L, et al. (1956) Inoculation syphilis in human volunteers. Medicine 35: 33–82.

26. Wines BD, Ramsland PA, Trist HM, Garman S, Brink R, et al. (2011) Interaction of human, rat, and mouse immunoglobulin A (IgA) with *Staphylococcal superantigenic* 7 (SSEA7) decay protein and leukocyte IgA receptor. J Biol Chem 286: 33118–33124.

27. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, et al. (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science 201: 373–388.

28. Matejko P, Strouhal M, Smajis D, Norris SJ, Palkl T, et al. (2008) Complete genome sequence of *Treponema pallidum* subspp. pallidum strain SS14 determined with oligonucleotide arrays. BMC Microbiol 8: 76.

29. Mikalo A, Strouhal M, Cekavka D, Zabinkova M, Popisilova P, et al. (2010) Genome analysis of *Treponema pallidum* subsp. pallidum and subsp. pertenue: parasites mostly of the genetic differences are localized in six regions. PLoS One 5: e15713.

30. Giacani L, Chattopadhyay S, Centurion-Lara A, Jeffrey BM, Le HT, et al. (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.

PLOS Neglected Tropical Diseases | www.plosntds.org 12 May 2013 | Volume 7 | Issue 5 | e2222
31. Strouhal M, Smajs D, Matejkova P, Sodergren E, Amin AG, et al. (2007) Genome differences between Treponema pallidum subsp. pallidum strain Nichols and T. pallidum subsp. carculii strain Cuniculi. A Infect Immun 75: 3859–3866.
32. Smajs D, Zobansilova K, Strouhal M, Cechova D, Dugan-Rocha S, et al. (2011) Complete genome sequence of Treponema pallidum subsp. carculii A: the loss of infectivity to humans is associated with genomic decay. PLoS One 6: e20413.
33. Centurion-Lara A, Castro C, Barrett L, Cameron C, Mostowfi M, et al. (1999) Treponema pallidum major sheath protein homologue Tpr K is a target of osnogenic selection likely to influence spirochete morphology and the protective immune response. J Exp Med 189: 647–657.
34. Morgan CA, Molini BJ, Lukehart SA, Van Voorhis WC (2008) Segregation of B and T cell epitopes of Treponema pallidum repeat protein K to variable and conserved regions during experimental syphilis infection. J Immunol 188: 952–957.
35. Leader BT, Hevner K, Molini BJ, Barrett LK, Van Voorhis WC, et al. (2003) Antibody responses elicited against the Treponema pallidum repeat proteins differ during infection with different isolates of Treponema pallidum subsp. pallidum. Infect Immun 71: 6054–6057.
36. Gicanti L, Sun ES, Hevner K, Molini BJ, Van Voorhis WC, et al. (2004) Tpr homology in Treponema pallidum carculii Cuniculi A strain. Infect Immun 72: 6561–6567.
37. Sun ES, Molini BJ, Barrett LK, Centurion-Lara A, Lukehart SA, et al. (2004) Sequence variation and association with neurosyphilis. J Infect Dis 202: 1380–1388.
38. Treponema pallidum multiple alleles. Infect Immun 68: 824–831.
39. Smajs D, McKevitt M, Howell JK, Norris SJ, Cai WW, et al. (2005) Complete genome sequence of Treponema pallidum subsp. pallidum strain Nichols. Genome Res 15: 618–638.
41. Weinstock GM, Hardham JM, McLeod MP, Sodergren EJ, Norris SJ (1998) The genome of Treponema pallidum: new light on the agent of syphilis. FEMS Microbiol Rev 22: 323–332.
42. Haase EM, Camparguari AA, Sarwar J, Shero M, Wirth M, et al. (1991) Strain-specific and immunodominant surface epitopes of the P2 porin protein of nontypeable Haemophilus influenzae. Infect Immun 59: 1278–1284.
43. Giacani L, Gicanti L, Molini BJ, Godornes C, Barrett LK, Van Voorhis WC, et al. (2007) Quantitative analysis of tpr gene expression in Treponema pallidum isolates identifies differences among isolates and correlation with T-cell responsiveness in experimental syphilis. Infect Immun 75: 101–112.
44. Centurion-Lara A, Castro C, Barrett L, Cameron C, Mostowfi M, et al. (2000) Subtyping of Treponema pallidum subsp. carculii strains. Infect Immun 68: 5187–5191.
45. Giacani L, Lukehart S, Centurion-Lara A (2007) Length of guanosine repeat protein K attenuates syphilitic lesion development in the rabbit model. Infect Immun 75: 6054–6057.
46. Giacani L, Lukehart S, Centurion-Lara A (2007) TprC/D (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of Treponema pallidum, has a bipartite domain structure. J Bacteriol 194: 2321–2333.
47. Weinstock GM, Hardham JM, McLeod MP, Sodergren EJ, Norris SJ (1998) The genome of Treponema pallidum: new light on the agent of syphilis. FEMS Microbiol Rev 22: 323–332.
48. Sauna ZE, Calcagno AM, Kim IW, Kimchi-Sarfaty C, Oh JM, et al. (2007) Molecular evolution of the tprC, D, I, K, and J genes in the pathogenic genus Treponema. Mol Biol Evol 24: 2220–2233.
49. Giacani L, Brandt SL, Puray-Chavez M, Bruck Reil T, Godornes C, et al. (2004) 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.
50. Centurion-Lara A, LaFond RE, Hevner K, Godornes C, Molini BJ, et al. (2004) Genomic conversion: a mechanism for generation of heterogeneity in the tprK gene of Treponema pallidum during infection. Molecular Microbiology 52: 1537–1536.
51. Anand A, Luthra A, Dunham-Ems S, Caimano MJ, Karanjan C, et al. (2012) Tpr/C/D (Tp0117/131), a trimeric, pore-forming outer membrane protein of Treponema pallidum, has a bipartite domain structure. J Bacteriol 194: 2321–2333.
86. Tantalo LC, Lakehart SA, Marra CM (2005) Treponema pallidum strain-specific differences in neuroinvasion and clinical phenotype in a rabbit model. J Infect Dis 191: 75–80.
87. Small JD, Newman B (1972) Venereal spirochetosis of rabbits (rabbit syphilis) due to Treponema caridae: a clinical, serological, and histopathological study. Lab Anim Sci 22: 77–89.
88. 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.
89. Hackett CJ (1963) On the Origin of the Human Treponematoses (Pinta, Yaws, Endemic Syphilis and Venereal Syphilis). Bull World Health Organ 29: 7–41.
90. Hudson EH (1965) Treponematosis in perspective. Bull World Health Org 32: 733–748.