

# Whole Genome Sequence of the *Treponema* Fribourg-Blanc: Unspecified Simian Isolate Is Highly Similar to the Yaws Subspecies

Marie Zobaníková<sup>1</sup>, Michal Strouhal<sup>1,2</sup>, Lenka Mikalová<sup>1</sup>, Darina Čejková<sup>1,2</sup>, Lenka Ambrožová<sup>1</sup>, Petra Pospíšilová<sup>1</sup>, Lucinda L. Fulton<sup>2</sup>, Lei Chen<sup>2</sup>, Erica Sodergren<sup>2</sup>, George M. Weinstock<sup>2</sup>, David Šmajs<sup>1\*</sup>

**1** Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic, **2** The Genome Institute, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, United States of America

## Abstract

**Background:** Unclassified simian strain *Treponema* Fribourg-Blanc was isolated in 1966 from baboons (*Papio cynocephalus*) in West Africa. This strain was morphologically indistinguishable from *T. pallidum* ssp. *pallidum* or ssp. *pertenue* strains, and it was shown to cause human infections.

**Methodology/Principal Findings:** To precisely define genetic differences between *Treponema* Fribourg-Blanc (unclassified simian isolate, FB) and *T. pallidum* ssp. *pertenue* strains (TPE), a high quality sequence of the whole Fribourg-Blanc genome was determined with 454-pyrosequencing and Illumina sequencing platforms. Combined average coverage of both methods was greater than 500×. Restriction target sites (n = 1,773), identified *in silico*, of selected restriction enzymes within the Fribourg-Blanc genome were verified experimentally and no discrepancies were found. When compared to the other three sequenced TPE genomes (Samoa D, CDC-2, Gauthier), no major genome rearrangements were found. The Fribourg-Blanc genome clustered with other TPE strains (especially with the TPE CDC-2 strain), while *T. pallidum* ssp. *pallidum* strains clustered separately as well as the genome of *T. paraluiscuniculi* strain Cuniculi A. Within coding regions, 6 deletions, 5 insertions and 117 substitutions differentiated Fribourg-Blanc from other TPE genomes.

**Conclusions/Significance:** The Fribourg-Blanc genome showed similar genetic characteristics as other TPE strains. Therefore, we propose to rename the unclassified simian isolate to *Treponema pallidum* ssp. *pertenue* strain Fribourg-Blanc. Since the Fribourg-Blanc strain was shown to cause experimental infection in human hosts, non-human primates could serve as possible reservoirs of TPE strains. This could considerably complicate recent efforts to eradicate yaws. Genetic differences specific for Fribourg-Blanc could then contribute for identification of cases of animal-derived yaws infections.

**Citation:** Zobaníková M, Strouhal M, Mikalová L, Čejková D, Ambrožová L, et al. (2013) Whole Genome Sequence of the *Treponema* Fribourg-Blanc: Unspecified Simian Isolate Is Highly Similar to the Yaws Subspecies. PLoS Negl Trop Dis 7(4): e2172. doi:10.1371/journal.pntd.0002172

**Editor:** Mathieu Picardeau, Institut Pasteur, France

**Received:** January 3, 2013; **Accepted:** March 7, 2013; **Published:** April 18, 2013

**Copyright:** © 2013 Zobaníková et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by a grant from the Ministry of Health of the Czech Republic (NT11159-5/2010), and by the Grant Agency of the Czech Republic (P302/12/0574) to DŠ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: dsmajs@med.muni.cz

## Introduction

*Treponema* Fribourg-Blanc was isolated in 1966 from baboons (*Papio cynocephalus*) in West Africa [1,2]. This strain was morphologically indistinguishable from *T. pallidum* ssp. *pallidum* (TPA) or ssp. *pertenue* (TPE) strains and the ability to cause human infection was experimentally verified [3]. In baboons, enlarged lymphatic nodes with no specific clinical signs were observed [2]. Several other cases of primate treponematoses have been described [4–9] either without clinical signs or with symptoms of yaws. Skin samples taken from baboons in the Gombe National Park revealed a yaws-like infection that appeared to be transmitted via sexual contact [10]. Furthermore, in a field survey in 2007 at Lake Manyara National Park in Tanzania, several olive baboons (*Papio hamadryas anubis*) showed severe ulcerations strictly localized to the anogenital regions [11]. Similar lesions were found also in wild baboons living in other Tanzanian National Parks and in the

Ngorongoro Conservation Area (Tanzania) [12]. Although this clinical manifestation suggested a disease similar to human syphilis infections, a genetic analysis of the causative agent showed higher genetic similarity to human yaws-causing strains than to syphilis-causing strains [11,12].

The causative agent of yaws, *Treponema pallidum* ssp. *pertenue* [13], predominantly causes infections in tropical regions of Africa, Asia, Oceania and South America with an estimated prevalence of 2 million cases worldwide [14]. Three TPE strains were recently sequenced [15] and the observed genetic difference from syphilis-causing strains of *T. pallidum* ssp. *pallidum* was lower than 0.2% of the genome sequence. In humans, yaws is a multi-stage disease, transmitted through direct skin contact from an infected patient to a recipient. It is characterized by skin nodules and ulcerations, joint and soft tissue destruction and bone changes. Although some reports have described infection of the central nervous system, cardiovascular system and fetus during yaws infection [16], there is

## Author Summary

A bacterial strain isolated in 1966 from baboons (*Papio cynocephalus*) in West Africa was preliminarily characterized as unclassified simian strain *Treponema* Fribourg-Blanc (FB). This strain was morphologically identical to *T. pallidum* ssp. *pallidum* (TPA, agent of syphilis) or ssp. *pertenue* (TPE, agent of yaws). In this study, we completed a high quality whole genome sequence of simian isolate *Treponema* Fribourg-Blanc and compared it to known genome sequences of *Treponema pallidum* strains. No major differences in the gene order of the FB genome were found when compared to all known genomes of *Treponema pallidum* subspecies. Moreover, the FB genome clustered with other TPE strains, while *T. pallidum* ssp. *pallidum* strains clustered separately. In general, the FB genome showed similar genetic characteristics to other TPE strains. Therefore, we proposed that the simian isolate Fribourg-Blanc be classified as a bacterial strain belonging to *Treponema pallidum* ssp. *pertenue*. It appears that, except for humans, the reservoir of yaws-causing treponemes may also include free-living primates, especially in Africa.

not enough experimental data to clearly prove the ability of TPE strains to invade the CNS or cause congenital infection. It is generally believed that humans are the primary reservoir of yaws. Since transmission requires direct contact with the causative agent of yaws, a risk of contamination between human and other primates could exist in regions where yaws and other primate treponemal infections occur simultaneously [10].

Several previous genetic studies have described partial FB sequences [17–26] and some of them predicted that FB strains were closely related to TPE strains [18,19,21,22,24]. Prior to this work, about 55 kbp (4.83%) of the FB genome sequence had been determined. In this communication, we compare the complete genome sequence of the simian isolate Fribourg-Blanc to three TPE strains (Samoa D, CDC-2, and Gauthier) and to five TPA strains (Nichols, DAL-1, Chicago, SS14, Mexico A). Based on the low genetic variability between Fribourg-Blanc and the 3 TPE strains, the Fribourg-Blanc bacterial strain is a *Treponema pallidum* ssp. *pertenue* strain.

## Materials and Methods

### Amplification and isolation of Fribourg-Blanc DNA

A sample containing extracted Fribourg-Blanc treponemes (from infected rabbit tissue) was obtained from D. L. Cox, CDC, Atlanta, GA, USA. The sample contained  $5 \times 10^6$  cells per ml and the DNA was amplified in one step directly from frozen cells ( $5 \times 10^3$  cells) with the whole genome amplification procedure (REPLI-g kit, QIAGEN, Valencia, CA, USA). Amplification resulted in 413 ng of DNA per  $\mu$ l, (30  $\mu$ l in total); however, both treponemal and rabbit DNA was present in the amplified DNA. Therefore, Fribourg-Blanc DNA was repeatedly amplified using the pooled segment genome sequencing (PSGS) method described previously [15]. Briefly, the genomic Fribourg-Blanc DNA was amplified with 134 specific primer pairs as overlapping PCR products (Table S1). To enable sequencing of paralogous genes, PCR products were separated into four pools (pool 1–4) and mixed in equimolar amounts. For 454-pyrosequencing, PCR products of these pools were labeled with multiplex identifier (MID) adapters and sequenced as four different samples. However, only one

sequencing mixture was prepared for Illumina because MID adapters were not available.

### DNA sequencing and assembly of the Fribourg-Blanc genome

Whole genome DNA sequencing used a Roche/Genome Sequencer FLX System platform (454 Life Sciences, Branford, CT, USA) combined with the Illumina/Solexa Genome Analyzer Ix approach (Illumina, San Diego, CA, USA). Sequencing was performed at The Genome Institute, Washington University School of Medicine (St. Louis, MO, USA). 454 reads were assembled using a Newbler assembler while Illumina reads were assembled using Velvet [27]. 454-pyrosequencing and Illumina sequencing resulted in average read lengths of 230 bp and 35 bp and the total average coverage of  $70\times$  and  $465\times$ , respectively. Assembled contigs obtained from both methods were aligned to the reference genome TPE CDC-2 using Lasergene software (DNASTAR, Madison, WI, USA). All gaps in the genome sequence and all discrepancies between contig sequences obtained using both methods were resolved using Sanger sequencing. Altogether, 85 genomic regions of the Fribourg-Blanc genome were amplified and Sanger sequenced.

In addition, several genomic regions were amplified with specific primers as *Treponema pallidum* intervals (TPI) using a GeneAmp XL PCR Kit (Applied Biosystems, Foster City, CA, USA) [28]. These intervals contained following paralogous genes: *tpnC* (TPI11), *tpnD* (TPI12), *tpnE* (TPI25A), *tpnF* and *tpnG* (TPI25B), *tpnI* and *tpnJ* (TPI48), and *tpnL* (TPI78). XL PCR products were purified using a QIAquick PCR Purification Kit (QIAGEN) and sequenced with a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using internal primers. The TPI71A (Table S1) region, which was not included in any pool, was sequenced similarly. The *tpnK* (TPFB\_0897), *arp* (TPFB\_0433), and TPFB\_0470 genes were amplified and cloned into pCR 2.1-TOPO (Invitrogen, Carlsbad, CA, USA) and five independent clones for TPFB\_0433, eight for TPFB\_0470 or ten clones for *tpnK* were sequenced.

A total of 11 genomic regions (in genes TPFB\_0012, TPFB\_0040, TPFB\_0067, TPFB\_0179, TPFB\_0279, TPFB\_0347, TPFB\_0859, TPFB\_0865 and in intergenic regions (IGR) TPFB\_0347–0348, IGR TPFB\_0379–0380, IGR TPFB\_0381–0382), containing homopolymeric (G or C) stretches were amplified with *Pfu* polymerase (Fermentas Inc., Glen Burnie, MD, USA) as follows: 5  $\mu$ l of  $10\times$  *Pfu* buffer with 20 mM  $MgSO_4$ , 1  $\mu$ l of dNTP mix (each nucleotide of 10 mM concentration), 1  $\mu$ l of DNA (1–5 ng/ $\mu$ l), 0.5  $\mu$ l of forward primer and 0.5  $\mu$ l of reverse primer, 41  $\mu$ l of water for PCR, and 1  $\mu$ l (2.5 U) of *Pfu* DNA polymerase. The cycling conditions were: 94°C for 1 minute; 30 cycles: 94°C for 1 minute, 60°C for 30 s, 72°C for 1 minute; 72°C for 10 minutes. To facilitate the subsequent cloning of these PCR products into a pCR 2.1-TOPO vector (Invitrogen), 0.2  $\mu$ l of *Taq* polymerase was added to the mixture and incubation at 72°C for 10 minutes followed. Plasmid DNA was isolated using a QIAGEN Plasmid Mini Kit (QIAGEN) and sequenced with universal primers from a TOPO TA Cloning Kit. At least five independent clones were sequenced.

### Whole genome fingerprinting (WGF)

To verify final genome assemblies, whole genome fingerprints of three enzymes including *Bam*H I, *Eco*R I and *Hind* III [24,28] were compared to the *in silico* restriction enzyme analysis of the sequenced Fribourg-Blanc genome. The average error rate of WGF for *Treponema paraluiscuniculi* strain Cuniculi A was previously

calculated [29] and corresponded to 27.9 bp (1.6% of the average fragment length) with a variation range between 0 and 132 bp.

### Gene identification, annotation and classification

The final whole genome sequence was assembled from 454-pyrosequencing and Illumina contigs and Sanger sequenced regions comprising the *tp*r genes, repetitive DNA regions (e.g. TPFB\_0433, TPFB\_0470), regions containing homopolymers, gaps between contigs and discrepant regions between 454 and Illumina contigs. The Geneious software v5.6.5 [30] was used for gene annotation based on the recent annotation of the CDC-2 genome [15]. Gene TPFB\_0897, coding for TprK protein, showed intrastrain variability and therefore nucleotides in variable regions were replaced with Ns in the complete genome sequence. Genes were tagged with the TPFB\_ prefix. In Fribourg-Blanc, the original locus tag numbering corresponds to the tag numbering of orthologous genes annotated in the TPE CDC-2 genome. For proteins with unpredicted functions, a gene size limit of 150 bp was applied. TPE genes were classified into seven groups according to their probable function as described previously, i.e. genes involved in general metabolism; in cell processes and cell structure; in DNA replication, repair, recombination; in regulation, transcription and translation; in transport; in virulence; and genes of unknown function [15].

### Comparisons of whole genome sequences

Whole genome nucleotide alignments of five TPA strains, three TPE strains [15], *Treponema paraluisuniculi* Cuniculi A strain (CP002103.1, [29]) and the Fribourg-Blanc isolate (CP003902.1) were used for determination of genetic relatedness using several approaches including calculation of nucleotide diversity ( $\pi$ ), calculation of nucleotide divergence ( $d_A$ ) and construction of a phylogenetic tree. TPA strains included Nichols (resequenced genome; unpublished data), DAL-1 (CP03115.1, [31]), SS14 (resequenced genome; unpublished data), Chicago (CP001752.1, [32]), and Mexico A (CP003064.1, [33]) while TPE strains included Samoa D (CP002374.1), CDC-2 (CP002375.1), and Gauthier (CP002376.1). Whole genome alignments were constructed using Geneious software and SeqMan software (DNASTAR, Madison, WI, USA). Nucleotide changes among studied whole genome sequences were analyzed using *DnaSP* software, version 5.10 [34]. An unrooted phylogenetic tree was constructed from whole genome sequence alignments using the Maximum Parsimony method and MEGA software [35].

### Nucleotide sequence accession numbers

The complete genome sequence of the Fribourg-Blanc isolate was deposited in the GenBank under the accession number CP003902.1.

## Results

### Whole genome sequencing, genome annotation, and genomic parameters

The FB genome was determined using two independent whole genome sequencing methods (454-pyrosequencing, Illumina) with a total combined average coverage greater than 500 $\times$ . The Sanger sequencing method was used for finishing the complete genome sequence and for additional sequencing including paralogous, repetitive and intrastrain variable chromosomal regions.

The Fribourg-Blanc genome was annotated according to the sequence of the CDC-2 genome [15]. The gene names were denoted with the TPFB\_ prefix (*Treponema pallidum* Fribourg-Blanc). The FB genome was most similar to TPE strains. The

summarized genomic features of the Fribourg-Blanc simian isolate (and other completely sequenced TPE strains) are shown in Table 1. The Fribourg-Blanc genome (1,140,481 bp) was 737–1151 bp longer than other TPE strains. No major genome rearrangements were found compared to the other 3 TPE genomes. Altogether, 1122 genes were annotated in the Fribourg-Blanc genome including 54 untranslated genes encoding rRNA, tRNA and other ncRNA (a short bacterial RNA molecules that are not translated into a protein). Compared to the other TPE genomes, genes TPFB\_0012 and TPFB\_0896 (both encoding hypothetical proteins) contained a 1 bp deletion (frameshift mutation) and nonsense mutation, respectively. Therefore, these two genes were not annotated in the FB genome. TPFB\_0304 (encoding treponemal conserved hypothetical protein) was not annotated because of nucleotide change in the stop codon followed by fusion with the TPFB\_0303 (encoding DNA mismatch repair protein MutL). The average and median gene lengths of the Fribourg-Blanc genome were calculated as 983 bp and 831 bp, respectively. The intergenic regions covered 53 kb and represented 4.63% of total FB genome length, which is similar to the length of these regions in other TPE strains. A total of 640 genes (57.0%) encoded proteins with predicted function, 139 genes encoded treponemal conserved hypothetical proteins (TCHP, 12.4%), 141 genes encoded conserved hypothetical proteins (CHP, 12.6%), 145 genes encoded hypothetical proteins (HP, 12.9%) and 3 genes (0.3%) were annotated as pseudogenes. When compared to the Nichols genome (AE000520.1), 9 additional genes (orthologous to TP0129, TP0132, TP0180, TP0266, TP0318, TP0370, TP0532, TP0671 and TP1030) can be considered as pseudogenes in the Fribourg-Blanc genome (the same genes were also considered pseudogenes in other TPE strains). When compared to TPE strains, 2 additional genes (orthologous to TPE\_0012, TPE\_0896; Table 2) can be considered pseudogenes in the FB genome.

### Whole genome fingerprinting

The *in silico* identified restriction target sites (RTS) within the FB genome were compared with experimental restriction digest patterns of individual TPI regions covering the entire TPE genome [24]. Altogether, 1,773 RTSs representing more than 10.6 kb of analyzed sequence were experimentally tested [24]. Since no discrepancies between *in silico* and experimental RTS analyses of the FB genome were found, the estimated sequencing error rate for the FB genome was therefore  $10^{-4}$  or less.

### Sequence relatedness of the FB genome to other pathogenic treponemal genomes

Sequence relatedness of the FB genome to other TP genomes based on available whole genome sequences is shown in Figure 1. The FB genome clustered with other TPE strains (especially with the TPE CDC-2 strain), while TPA strains, as well as the genome of the *T. paraluisuniculi* (TPc) strain Cuniculi A, each clustered separately. Calculated nucleotide diversity among currently sequenced *T. pallidum* and *T. paraluisuniculi* strains are shown in Table 3. Detailed characterization of nucleotide diversity between TPE strains and the FB isolate is shown in Table 4. The FB genome was found to be 99.97% identical to other TPE genomes. The lowest calculated nucleotide diversity ( $\pi$ )  $\pm$  standard deviation among TPE strains and the FB isolate was found between the Fribourg-Blanc and CDC-2 strain ( $0.00016 \pm 0.00008$ ), which is identical to nucleotide diversity between Samoa D and CDC-2 genomes. In contrast, the highest calculated nucleotide diversity was found between the Fribourg-Blanc and Gauthier strain ( $0.00044 \pm 0.00022$ ), which was similar to the difference between the Samoa D and and Gauthier strains ( $0.00044 \pm 0.00022$ ). For

**Table 1.** Summary of genomic features of the FB genome and three *T. pallidum* ssp. *pertenue* strains (Samoa D, CDC-2 and Gauthier).

Genome parameter	Fribourg-Blanc isolate	Samoa D	CDC-2	Gauthier
GeneBank accession number	CP003902.1	CP002374.1	CP002375.1	CP002376.1
Genome size	1,140,481 bp	1,139,330 bp	1,139,744 bp	1,139,417 bp
G+C content	52.80%	52.80%	52.80%	52.80%
No. of fused genes <sup>a</sup>	25 (52 corresponding genes in the Nichols genome)	25 (52 corresponding genes in the Nichols genome)	24 (50 corresponding genes in the Nichols genome)	24 (50 corresponding genes in the Nichols genome)
Sum of the intergenic region lengths (% of the genome length)	52,785 bp (4.63 %)	52,844 bp (4.64%)	52,963 bp (4.65%)	53,300 bp (4.68%)
Average/median gene length	982.6/831.0 bp	980.3/831.0 bp	980.4/831.0 bp	979.3/831.0 bp
No. of predicted protein-encoding genes	1065	1068	1068	1068
No. of genes encoded on plus/minus DNA strand	599/523	600/525	600/525	600/525
No. of genes coding for proteins with predicted function	640	640	640	640
No. of genes coding for treponemal conserved hypothetical proteins	139	140	140	140
No. of genes coding for conserved hypothetical proteins	141	141	141	141
No. of genes coding for hypothetical proteins	145	147	147	147
No. of annotated pseudogenes (no. of all pseudogenes compared to Nichols sequence <sup>b</sup> )	3 (14)	3 (12)	3 (12)	3 (12)
No. of tRNA loci	45	45	45	45
No. of rRNA loci	6 (2 operons)	6 (2 operons)	6 (2 operons)	6 (2 operons)
No. of ncRNAs	3	3	3	3

<sup>a</sup>Number of genes in a particular genome which sequence include at least 2 genes predicted in the Nichols genome AE000520.1. doi:10.1371/journal.pntd.0002172.t001

comparison, calculated  $\pi$  values between Fribourg-Blanc and TPA strains were one order of magnitude higher than  $\pi$  values between Fribourg-Blanc and TPE strains (Table 3).

### Genome differences specific for the FB genome

To define genome differences specific to the FB genome, the whole genome sequence of this strain was compared to the available genome sequences of TPE strains [15]. In coding regions, 6 deletions, 5 insertions and 117 substitutions differentiated FB from TPE genomes (Table 5, Table S2). Frameshift mutations (three deletions and two insertions) resulted in an omitted annotation of TPFB\_0012 (encoding hypothetical protein), in gene truncation (TPFB\_0040, *mcp* coding for methyl-accepted chemotaxis protein; TPFB\_0347 encoding hypothetical membrane protein; TPFB\_0484, encoding conserved hypothetical protein) or in gene elongation (TPFB\_0461a, encoding hypothetical protein). Other major changes were located in genes TPFB\_0548 (containing 42-bp deletion), TPFB\_0303 fused with TPFB\_0304, TPFB\_0126b (truncated as a result of a start codon mutation), TPFB\_0896 (not annotated because of a nonsense mutation), TPFB\_0433 encoding acidic repeat protein Arp (containing 15 tandem repeat units of 60-bp compared to 12, 4, and 10 repeat units in Samoa D, CDC-2, and Gauthier, respectively) and TPFB\_0470 (containing 22 tandem repeat units of 24-bp, compared to 12, 37, and 25 tandem repeat units in Samoa D, CDC-2, and Gauthier, respectively). A set of 117 substitutions resulted in one nonsense mutation, one mutation

affecting the start codon, one mutation affecting the stop codon and in 88 nonsynonymous mutations (82 nonconserved). Most of the changes were found in *tprC* (TPFB\_0117) and *tprI* (TPFB\_0620) genes. Mutations causing changes larger than 5 amino acid replacements or protein truncations and elongations are listed in Table 2.

### Discussion

The complete genome sequence of the simian isolate Fribourg-Blanc (FB) was determined and compared to five syphilis-causing (TPA) and three human yaws-causing *T. pallidum* ssp. *pertenue* (TPE) strains. Previous reports have shown that the FB strain (isolated from *Papio cynocephalus* in 1966 in West Africa) was morphologically indistinguishable from other TPA or TPE strains [2]. Moreover, the ability of FB strain to attach to mammalian cells was similar to TPE but different from TPA strains [36]. In addition to these studies, several other genetic studies showed a close relationship between FB strain and TPE strains of human origin [18,19,21,22,24].

Experimental human and monkey (monkeys of the genus *Macacus*) infection with the FB strain resulted in symptoms similar to yaws [3,37]. Conversely, primate infection with TPE strains of human origin resulted in detectable lesions in at least a subset of infected monkeys of the genus *Macacus* and *Semnopithecus* [38]. In addition, several other reports demonstrated that TPE strains of human origin can experimentally infect monkeys [39,40]. These

**Table 2.** Mutations causing gene changes resulting in protein truncations and elongations in comparison with TPE strains.

Gene (predicted protein function)	Nucleotide change	Coordinates of change in the FB genome (CP003902.1)	Result of nucleotide change
TPFB_0012 (HP)	1 bp deletion <sup>a</sup>	12479–12487	gene shortened by 47 bp to 129 bp, gene was not annotated in the FB genome
TPFB_0040, <i>mcp</i> (methyl-accepting chemotaxis protein)	5 bp insertion <sup>a</sup>	49359–49373	gene shortened by 17 bp to 2433 bp
TPFB_0126b (HP)	3 bp substitution in the start codon	148982–148984	gene shortened by 42 bp to 366 bp
TPFB_0303 (TCHP)	1 bp substitution in the stop codon	319012	gene fusion of genes ortologous to TPE_0303 and TPE_0304, gene was annotated as TPFB_0303 (5076 bp)
TPFB_0347 (HMP)	2 bp insertion <sup>a</sup>	373747–373761	gene shortened by 35 bp to 711 bp
TPFB_0433, <i>arp</i> (Arp protein)	15 tandem repeat units, one unit was 60 bp long	462777–463676	Samoa D, CDC-2, and Gauthier, contains 12, 4, and 10 repeat units, respectively
TPFB_0461a (HP)	1 bp deletion <sup>a</sup>	493013–493022	gene elongation by 61 bp to 243 bp
TPFB_0470 (CHP)	22 tandem repeat units, one unit 24 bp long	499435–499962	Samoa D, CDC-2, and Gauthier, contains 12, 37, and 25 repeat units, respectively
TPFB_0484 (CHP)	1 bp deletion <sup>a</sup>	517701–517708	gene shortened by 309 bp to 1707 bp
TPFB_0548 (TCHP)	42 bp deletion	594092–594093	gene shortened by 42 bp to 1257 bp
TPFB_0896 (HP)	2 bp substitution leading to nonsense mutation	977039, 977041	gene shortened by 99 bp to 54 bp, gene was not annotated in the FB genome

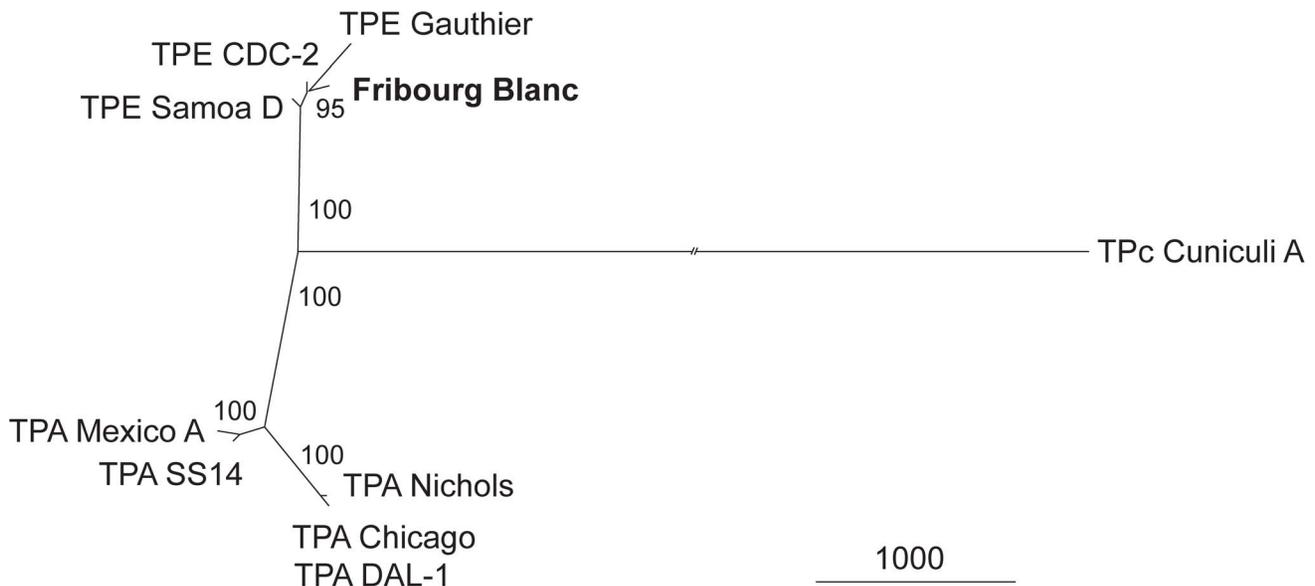
HP- hypothetical protein, CHP – conserved hypothetical protein, TCHP – treponemal conserved hypothetical protein, HMP – hypothetical membrane protein.

<sup>a</sup>changes in homopolymeric regions.

doi:10.1371/journal.pntd.0002172.t002

data indicate that both TPE and FB strains have overlapping or identical host range suggesting a close relationship among these strains.

As shown by this study, the genome of the FB simian isolate [1,2] was very similar to TPE strains of human origin. Although the FB genome was most closely related to the African TPE

**Figure 1.** An unrooted tree constructed from whole genome sequence alignments of 10 complete genome nucleotide sequences.

An unrooted tree constructed from whole genome sequence alignments using the Maximum Parsimony method and MEGA software [34]. The bar scale corresponds to 1000 nt changes. Bootstrap values based on 1,000 replications are shown next to the branches. All positions containing deletions in at least one genome sequence were omitted from further analysis. The analysis comprised 10 complete genome nucleotide sequences including 5 strains of TPA (*Treponema pallidum* ssp. *pallidum*), 3 strains of TPE (*Treponema pallidum* ssp. *pertenuis*), one TPC (*Treponema paraluis-cuniculi*) strain and the FB strain. There were a total of 1,129,016 nucleotide positions aligned in the final dataset. Note the clustering of the FB genome with other TPE strains. The branch of TPC was shortened (//).

doi:10.1371/journal.pntd.0002172.g001

**Table 3.** Calculated nucleotide diversity ( $\pi \pm$  standard deviation) between FB isolate and individual TPA strains, TPE strains and the Cuniculi A strain.

Strain	Nucleotide diversity
TPA Nichols	0.00206 $\pm$ 0.00103
TPA DAL-1	0.00209 $\pm$ 0.00104
TPA Chicago	0.00203 $\pm$ 0.00102
TPA SS14	0.00180 $\pm$ 0.00090
TPA Mexico A	0.00172 $\pm$ 0.00086
TPE Samoa D	0.00023 $\pm$ 0.00012
TPE CDC-2	0.00016 $\pm$ 0.00008
TPE Gauthier	0.00044 $\pm$ 0.00022
TPc Cuniculi A	0.01044 $\pm$ 0.00518

doi:10.1371/journal.pntd.0002172.t003

CDC-2 strain (isolated in Akorabo, Ghana in 1980 [41]), its relatedness to another TPE strain of African origin, strain Gauthier (isolated in Brazzaville, Congo in 1960 [42]) was lower compared to the TPE Samoa D strain (isolated in Western Samoa in 1953 [43]). Thus, the Gauthier strain was the most distinct among the TPE strains. Compared to TPA strains (Nichols, SS14, DAL-1, Chicago, Mexico A), the calculated nucleotide diversity between individual TPE strains and the FB isolate was one order of magnitude lower than between TPA strains and the FB isolate. These data suggest that the FB isolate is in fact another TPE strain.

Several previous studies described partial FB sequences [17–26]. Altogether, 38 Fribourg-Blanc sequences comprising 55066 bp (4.83% of the FB genome sequence) were found when searching databases. Altogether, 7 nucleotide discrepancies in our genomic sequence were identified, and most of them were located in *tpr* genes or their vicinity ( $n=4$ ) and in homopolymeric regions ( $n=2$ ). Analysis of individual sequencing reads in these regions in Illumina, 454-pyrosequencing and Sanger raw data (except for homopolymeric regions where 454 pyrosequencing reads were not considered relevant) supported the sequences presented by our research. Besides differences in *tpr* regions and in homopolymeric regions that are likely results of intrastrain heterogeneity [31,44–46], a single remaining difference was found in the gene TPFB\_1038 (*tpF-1*, GenBank acc. no. EU102242, [22]). This difference may represent a genetic difference between different passages of the FB strain, locus with intrastrain heterogeneity or sequencing error.

Specific changes (deletions, insertions, and substitutions) comprising 185 nucleotides in 68 genes differentiated the FB strain from other TPE strains. Major genetic changes between FB and TPE genomes resulting in protein truncations or elongations were

located in 9 genes. These genes encoded hypothetical proteins with the exception of TPFB\_0040 (encoding methyl-accepting chemotaxis protein, M<sub>cp</sub>). Moreover, the genome of the FB strain contained a different number of tandem repeat units in genes TPFB\_0433 (encoding the acidic repeat protein, Arp) and TPFB\_0470 (encoding a conserved hypothetical protein) compared to orthologous genes in individual TPE strains. The number and sequence of 60-bp tandem repeat units within the *arp* gene, in the FB genome, revealed the same pattern as previously described for this strain [23]. Variability in the number of tandem repeat units in genes orthologous to TPFB\_0470 was also described in TPE and TPA strains [15,24]. A relatively high expression rate of the TP0470 gene, a TPFB\_0470 ortholog, in the Nichols genome during experimental rabbit infection was found [47]. This fact together with the variable number of tandem repeat units in this gene indicates that this gene may be involved in pathogen-host interactions. In bacterial pathogens, highly synthesized proteins with a variable number of tandem repeats are often involved in interaction with the host, e.g. an abundant outer membrane protein, secretin PilQ, of *Neisseria meningitidis* contains four to seven octapeptide copies and is a potential vaccine candidate for serogroup B of *N. meningitidis* [48]. In all genes with affected length in the FB genome (except for TPFB\_0484 and TPFB\_0347), major sequence differences were also found in the orthologous genes in TPE, TPA, and Cuniculi A genomes [29]. Moreover, the TPFB\_0304 gene was also found variable among orthologous genes in TPE strains [15]. In addition, the highest number of substitutions between the FB and TPE genomes was located in *tpr* genes (*tprC*, *tprI*, *tprF*), which are known to be variable among *T. pallidum* subspecies. Therefore, genetic differences specific for the FB genome appears to be predominantly localized in the variable genomic regions thus suggesting that the observed differences between the FB genomes and other TPE genomes likely do not result in considerable changes in the host range and pathogenicity. In the FB genome, 117 substitutions resulted (except for one nonsense, one start codon and one stop codon mutation) in 88 (75.2%) nonsynonymous mutations (82 nonconserved, 75.1%). A similar percentage of nonsynonymous mutations were also found in other strain-specific changes in TPE genomes (ranging from 66.1 to 80.5%). In this respect, the FB strain is very similar to other TPE strains.

TPE strains appears to be the most ancient treponemes based on the skeletal changes typical for yaws that were identified in bones dated to 1.6 million years ago [49,50]. Based on whole genome alignments, the FB genome clearly clustered with other TPE strains. Moreover, TPE strains including Fribourg-Blanc clustered separately from TPA strains. Interestingly, almost the same number of nucleotide changes evolutionary separated the *Treponema paraluis-cuniculi* Cuniculi A strain from TPE and TPA strains. Assuming that the TPE strains represent ancestral strains, and since it has been suggested that the Cuniculi A genome evolved by genome decay [29], it is possible that this rabbit

**Table 4.** Calculated nucleotide diversity ( $\pi \pm$  standard deviation) between individual TPE strains and the FB isolate.

Fribourg-Blanc	Gauthier	CDC-2	Samoa D
0.00023 $\pm$ 0.00012	0.00044 $\pm$ 0.00022	0.00016 $\pm$ 0.00008	***
0.00016 $\pm$ 0.00008	0.00037 $\pm$ 0.00018	***	Samoa D
0.00044 $\pm$ 0.00022	***		CDC-2
***			Gauthier
			Fribourg-Blanc

doi:10.1371/journal.pntd.0002172.t004

**Table 5.** Genome differences specific for the FB genome (comprising 630 nucleotides).

Nucleotide difference	Non-coding sequences (IGR) (altogether 445 bp)		Coding sequences (altogether 185 bp)		
	Number of changes	Number of mutated nucleotides	Number of changes	Number of mutated nucleotides (relevant protein change)	The most affected gene(s)
deletion	3	2 × single bp, 1 × 2 bp, altogether 4 bp	6	3 × single bp, altogether 3 bp (protein truncation or elongation); 3, 6 and 42 bp, altogether 51 bp; (protein shortening)	TPFB_0012, 0461a, 0484; TPFB_0370, 0548, 0859
insertion	3	2 × 2 bp, 1 × 430 bp (TPFB_0696–TPFB_0697), altogether 434 bp	5	1 × 2 bp, 1 × 5 bp, altogether 7 bp (protein truncation); 3 × 3 bp, altogether 9 bp; (protein elongation)	TPFB_0347, 0040; TPFB_0179, 0279, 0462
substitution	7	7 × single bp, altogether 7 bp	117	117 bp (synonymous or nonsynonymous mutations)	TPFB_0117, 0126a, 0316, 0324, 0488, 0620, 0865, 0968

Genome differences specific for the FB genome (comprising 630 nucleotides) in both non-coding (intergenic regions, IGR) and coding regions when compared to TPE strains (Samoa D, CDC-2, Gauthier).

The *tpkK* gene was excluded from this list of changes because of high sequence diversity within TPE strains.

doi:10.1371/journal.pntd.0002172.t005

pathogen evolved sometime during the early evolution of TPA strains (Figure 1) while the FB strain evolved along a similar path as TPE strains; potentially as a result of the close relatedness of its hosts with humans. However, there are several evolutionary scenarios explaining genetic similarity of the Fribourg-Blanc and TPE strains including i) hypothesis that TPE was acquired by humans from nonhuman primates, ii) hypothesis that the Fribourg-Blanc and TPE exchanged regularly between humans and other primates and even iii) possibility that the Fribourg-Blanc represent adaptation of TPE to nonhuman primates. An additional sequence information from other nonhuman primate isolates will be needed to address this question. Moreover, such studies on treponemes isolated from nonhuman primates could help to clarify if TPA evolved from TPE.

Several molecular genetic studies previously suggested that the Fribourg-Blanc strain was very closely related or identical to *T. pallidum* ssp. *pertenue* [18,19,21,22,23,24]. A principal finding of this work was the demonstration that the FB genome has similar genetic characteristics as other TPE strains and that the differences specific to the FB genome are similar to those differentiating other TPE strains and located mainly in variable genomic loci. From the results mentioned above, we can infer that the unclassified simian isolate Fribourg-Blanc belongs to the *Treponema pallidum* ssp. *pertenue*. The FB strain was shown to cause experimental infection in human hosts and TPE strains can infect primates. Although the human and animal diseases may be epidemiologically independent, it is likely that a reservoir for yaws exists among primate populations and/or humans serve as a reservoir for baboon infection, especially in Africa. This could considerably complicate

recent efforts to eradicate yaws [51]. However, further sequence data on treponemes isolated from nonhuman primates will reveal if these treponemes show molecular signatures similar to FB or human TPE strains. Nevertheless, knowledge of specific FB genetic changes could be useful to the epidemiological aspect of yaws eradication.

## Supporting Information

**Table S1** List of primers used for amplification of TP intervals for Fribourg-Blanc strain (primer coordinates, primer sequence, primer length, and TP interval size according to the Nichols strain, GenBank # AE000520.1).

(XLS)

**Table S2** List of specific differences between the Fribourg-Blanc (FB) genome and genomes of 3 strains of *Treponema pallidum* ssp. *pertenue* (TPE; Samoa D, CDC-2 and Gauthier).

(XLS)

## Acknowledgments

The authors thank Dr. David Cox (CDC, Atlanta, GA, USA) for providing the Fribourg-Blanc strain.

## Author Contributions

Conceived and designed the experiments: MS PP ES GMW DS. Performed the experiments: MZ MS LM LA PP LLF. Analyzed the data: MZ MS LM DC LA PP LC. Contributed reagents/materials/analysis tools: MZ MS LM DC LA PP LC. Wrote the paper: MZ MS LM DC DS.

## References

- Fribourg-Blanc A, Mollaret HH, Niel G (1966) Serologic and microscopic confirmation of treponemosis in Guinea baboons. *Bull Soc Pathol Exot Filiales* 59: 54–59.
- Fribourg-Blanc A, Mollaret HH (1969) Natural treponematoses of the African primate. *Primates Med* 3: 113–121.
- Smith JL, David NJ, Indgin S, Israel CW, Levine BM, et al. (1971) Neuro-ophthalmological study of late yaws and pinta. II. The Caracas project. *Br J Vener Dis* 47: 226–251.
- Cousins D (1984) Notes on the occurrence of skin infections in gorillas (*Gorilla gorilla*). *Zool Garten NF Jena* 54: 333–338.
- Cousins D (2008) Possible goundou in gorillas. *Gorilla Journal* 37: 22–24. Electronic article.
- Felsenfeld O, Wolf RH (1971) Serological reactions with treponemal antigens in nonhuman primates and the natural history of treponematoses in man. *Folia Primatol (Basel)* 16: 294–305.
- Levero F, Gatti S, Gautier-Hion A, Menard N (2007) Yaws disease in a wild gorilla population and its impact on the reproductive status of males. *Am J Phys Anthropol* 132: 568–575.
- Lovell NC, Jurmain R, Kilgore L (2000) Skeletal evidence of probable treponemal infection in free-ranging African apes. *Primates* 41: 275–290.
- Meder A (1994) Causes of death and diseases of gorillas in the wild. *Gorilla J* 2: 19–20.
- Wallis J, Lee DR (1999) Primate conservation: The prevention of disease transmission. *Int J Primatol* 20: 803–826.

11. Knauf S, Batamuzi EK, Mlengeya T, Kilewo M, Lejora IA, et al. (2012) *Treponema* infection associated with genital ulceration in wild baboons. *Vet Pathol* 49: 292–303.
12. Harper KN, Fyumagwa RD, Hoare R, Wambura PN, Coppenhaver DH, et al. (2012) *Treponema pallidum* infection in the wild baboons of East Africa: Distribution and genetic characterization of the strains responsible. *PLoS One* 7: e50882.
13. Castellani A (1905) Further observations on parangi (Yaws). *Brit Med Jour*: 1330–1331.
14. WHO (1998) The World Health Report 1998-life in the 21 century: a vision for all. World Health Organization. pp 132.
15. Čejková D, Zbaníková M, Chen L, Pospíšilová P, Strouhal M, et al. (2012) Whole genome sequences of three *Treponema pallidum* ssp. *pertenue* strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. *PLoS Negl Trop Dis* 6: e1471.
16. 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.
17. Centurion-Lara A, Arroll T, Castillo R, Shaffer JM, Castro C, et al. (1997) Conservation of the 15-kilodalton lipoprotein among *Treponema pallidum* subspecies and strains and other pathogenic treponemes: genetic and antigenic analyses. *Infect Immun* 65: 1440–1444.
18. Centurion-Lara A, Castro C, Castillo R, Shaffer JM, Van Voorhis WC, et al. (1998) The flanking region sequences of the 15-kDa lipoprotein gene differentiate pathogenic treponemes. *J Infect Dis* 177: 1036–1040.
19. Cameron CE, Castro C, Lukehart SA, Van Voorhis WC (1999) Sequence conservation of glycerophosphodiester phosphodiesterase among *Treponema pallidum* strains. *Infect Immun* 67: 3168–3170.
20. Cameron CE, Lukehart SA, Castro C, Molini B, Godornes C, et al. (2000) Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies *pallidum* Tp92. *J Infect Dis* 181: 1401–1413.
21. Gray RR, Mulligan CJ, Molini BJ, Sun ES, Giacani L, et al. (2006) Molecular evolution of the *tprC*, *D*, *I*, *K*, *G*, and *J* genes in the pathogenic genus *Treponema*. *Mol Biol Evol* 23: 2220–2233.
22. Harper KN, Ocampo PS, Steiner BM, George RW, Silverman MS, et al. (2008) On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl Trop Dis* 2: e148.
23. Harper KN, Liu H, Ocampo PS, Steiner BM, Martin A, et al. (2008) The sequence of the acidic repeat protein (*arp*) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong positive selection in the subspecies that causes syphilis. *FEMS Immunol Med Microbiol* 53: 322–332.
24. Mikalová L, Strouhal M, Čejková D, Zbaníková M, Pospíšilová P, et al. (2010) Genome analysis of *Treponema pallidum* subsp. *pallidum* and subsp. *pertenue* strains: most of the genetic differences are localized in six regions. *PLoS One* 5: e15713.
25. Čejková D, Zbaníková M, Pospíšilová P, Strouhal M, Mikalová L, et al. (2012) Structure of *rm* operons in pathogenic non-cultivable treponemes: sequence but not genomic position of intergenic spacers correlates with classification of *Treponema pallidum* and *T. paraluiscuniculi* strains. *J Med Microbiol* 62: 196–207.
26. Giacani L, Brandt SL, Puray-Chavez M, Reid TB, Godornes C, et al. (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.
27. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.
28. Strouhal M, Šmajs D, Matějková P, Sodergren E, Amin AG, et al. (2007) Genome differences between *Treponema pallidum* subsp. *pallidum* strain Nichols and *T. paraluiscuniculi* strain Cuniculi A. *Infect Immun* 75: 5859–5866.
29. Šmajs D, Zbaníková M, Strouhal M, Čejková D, Dugan-Rocha S, et al. (2011) Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. *PLoS One* 6: e20415.
30. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, et al. (2012) Geneious v5.6.5. Available: <http://www.geneious.com>.
31. Zbaníková M, Mikolka P, Čejková D, Pospíšilová P, Chen L, et al. (2012) Complete genome sequence of *Treponema pallidum* strain DAL-1. *Stand Genomic Sci* 7: 2615838.
32. Giacani L, Jeffrey BM, Molini BJ, Le HT, Lukehart SA, et al. (2010) Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. *J Bacteriol* 192: 2645–2646.
33. Pětrošová H, Zbaníková M, Čejková D, Mikalová L, Pospíšilová P, et al. (2012) Whole genome sequence of *Treponema pallidum* ssp. *pallidum*, Strain Mexico A, suggests recombination between Yaws and Syphilis strains. *PLoS Negl Trop Dis* 6: e1832.
34. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using Maximum likelihood, Evolutionary distance, and Maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
36. Fieldsteel AH, Stout JG, Becker FA (1979) Comparative behavior of virulent strains of *Treponema pallidum* and *Treponema pertenue* in gradient cultures of various mammalian cells. *Infect Immun* 24: 337–345.
37. Sepečtjan M, Guerraz FT, Salussola D, Thivolet J, Monier JC (1969) Contribution to the study of the treponeme isolated from monkeys by A. Fribourg-Blanc. *Bull World Health Organ* 40: 141–151.
38. Castellani A (1907) Experimental Investigations on Framboesia Tropica (Yaws). *J Hyg (Lond)* 7: 558–569.
39. Schöbl O (1928) Experimental yaws in philippine monkeys and a critical consideration of our knowledge concerning framboesia tropica in the light of recent experimental evidence. *Philippine J Sci* 35: 209.
40. Nichols HJ (1910) Experimental Yaws in the Monkey and Rabbit. *J Exp Med* 12: 616–622.
41. Liska SL, Perine PL, Hunter EF, Crawford JA, Feeley JC (1982) Isolation and transportation of *Treponema pertenue* in Golden-Hamsters. *Current Microbiology* 7: 41–43.
42. Gastinel P, Vaisman A, Hamelin A, Dunoyer F (1963) Study of a recently isolated strain of *Treponema pertenue*. *Ann Dermatol Syphiligr (Paris)* 90: 155–161.
43. Turner TB, Hollander DH (1957) Biology of the treponematoses based on studies carried out at the International Treponematoses Laboratory Center of the Johns Hopkins University under the auspices of the World Health Organization. *Monogr Ser World Health Organ*: 3–266.
44. Giacani L, Molini B, Godornes C, Barrett L, Van Voorhis W, et al. (2007) Quantitative analysis of *tpr* gene expression in *Treponema pallidum* isolates: Differences among isolates and correlation with T-cell responsiveness in experimental syphilis. *Infect Immun* 75: 104–112.
45. Matějková P, Strouhal M, Šmajs D, Norris SJ, Palzkill T, et al. (2008) Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. *BMC Microbiol* 8: 76.
46. 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.
47. Šmajs D, McKeivitt M, Howell JK, Norris SJ, Cai WW, et al. (2005) Transcriptome of *Treponema pallidum*: gene expression profile during experimental rabbit infection. *J Bacteriol* 187: 1866–1874.
48. Haghi F, Peerayeh SN, Siadat SD, Zeighami H (2012) Recombinant outer membrane secretin PilQ(406–770) as a vaccine candidate for serogroup B *Neisseria meningitidis*. *Vaccine* 30: 1710–1714.
49. Rothschild BM, Hershkovitz I, Rothschild C (1995) Origin of yaws in the Pleistocene. *Nature* 378: 343–344.
50. de Melo FL, de Mello JC, Fraga AM, Nunes K, Eggers S (2010) Syphilis at the crossroad of phylogenetics and paleopathology. *PLoS Negl Trop Dis* 4: e575.
51. Maurice J (2012) WHO plans new yaws eradication campaign. *Lancet* 379: 1377–1378.