

Whole Genome Analyses of Treponemes: New Targets for Strain- and Subspecies- Specific Molecular Diagnostics

David Šmajš¹, Lenka Mikalová¹, Darina Čejková¹,
Michal Strouhal¹, Marie Zobaníková¹, Petra Pospíšilová¹,
Steven J. Norris² and George M. Weinstock³

¹Masaryk University Czech Republic

²University of Texas-Houston Medical School

³Washington University School of Medicine
USA

1. Introduction

The genus *Treponema* comprises several human uncultivable pathogens including *Treponema pallidum* subspecies *pallidum* (TPA, the causative agent of the sexually transmitted syphilis), *Treponema pallidum* subspecies *pertenue* (TPE, causative agent of yaws), *Treponema pallidum* subspecies *endemicum* (TEN, causing endemic syphilis), and *Treponema carateum* causing pinta. Additionally, the rabbit pathogen *Treponema paraluis-cuniculi* (TPC) is very similar to syphilis treponeme but is not pathogenic to humans. Other pathogenic treponemes (e.g. *Treponema denticola* and *T. vincentii*) differ from the others by having considerably larger genomes (MacDougall & Girons, 1995; Seshadri *et al.*, 2004). Moreover, these treponemes can be cultivated under *in vitro* conditions. The infections caused by human uncultivable pathogenic treponemes can be classified according to their invasivity, from the most invasive bacterium causing venereal syphilis to *Treponema carateum* (pinta), which is a non-invasive spirochete causing local dermal lesions (Antal *et al.*, 2002). Strains of non-venereal treponemes including *Treponema pallidum* subspecies *pertenue* and *endemicum* are considered moderately invasive.

The whole genome analyses of treponemes started with the completion of the whole genome sequence of *T. pallidum*, by Nichols strain in 1998 (Fraser *et al.*, 1998). Since then, a number of genome studies have been performed (e.g. Brinkmann *et al.*, 2006; Giacani *et al.*, 2010; Harper *et al.*, 2008a; Matějková *et al.*, 2008; McKeivitt *et al.*, 2003; McKeivitt *et al.*, 2005; Mikalová *et al.*, 2010; Šmajš *et al.*, 2005; Strouhal *et al.*, 2007; Titz *et al.*, 2008). The genomic data has provided new opportunities to study pathogenic treponemes and increase our understanding of these unique pathogens.

Serological tests are considered standard laboratory methods for the diagnosis of syphilis since direct diagnostic methods are limited by the fact that the *T. pallidum* treponemes cannot be cultured continuously under *in vitro* conditions. The rabbit infectivity test (RIT) is the gold standard for demonstrating *T. pallidum* infection, but is impractical for clinical use

because of high costs and delayed test results. Microscopic identification of treponemes in clinical samples, particularly in combination with direct fluorescent antibody tests using anti-*T. pallidum* antibodies, is highly specific for treponemal infections; however, it does not distinguish between pathogenic treponemal species and requires technical expertise that is not commonly available in a clinical setting. Observed clinical manifestations, history, and serology have thus been the standard procedure for diagnosis of treponemal infections for the past century. However, the available serologic tests have several important limitations: i) the antibody response to treponemal infections is often not detectable during the first 1-3 weeks of infection, ii) routine treponemal tests for syphilis, which detect IgG antibodies, will be positive if the patient has a previous history of syphilitic infections, iii) diagnosis of congenital syphilis can be confused by transferred antibodies from the mother, and iv) serological tests in patients with a risk of endemic treponematoses cannot distinguish between these infections and syphilis.

In the last years, there has been an increasing effort to apply PCR techniques for direct diagnosis of syphilis (for review see Šmajs *et al.*, 2006). PCR detection of treponemal DNA is a direct method with sensitivity as low as a few copies of the treponemal chromosome per PCR reaction. Moreover, PCR detection of treponemal 16S rRNA, present in many copies per single treponemal genome increases the sensitivity to 10^{-2} - 10^{-3} genome equivalents (Centurion-Lara *et al.*, 1997). However, the relatively low numbers of treponemes in whole blood put limitations on PCR diagnosis of syphilis from blood samples.

2. Whole genome analyses of uncultivable treponemes

2.1 Whole genome fingerprinting

The genomes of nine uncultivable treponemes including *T. p. pallidum* strains (Nichols, SS14, DAL-1 and Mexico A), *T. p. pertenue* strains (Samoa D, CDC-2 and Gauthier), *Treponema paraluis-cuniculi* Cuniculi A strain, and the Fribourg-Blanc simian isolate, were studied using the whole genome fingerprinting technique (WGF, Mikalová *et al.*, 2010; Strouhal *et al.*, 2007; Weinstock *et al.*, 2000). More than 130 individual amplicons covering the entire genome were digested with a set of several restriction endonucleases and the resulting restriction fragments were visualized using gel electrophoresis. WGF was used to estimate the genome size, genome structure and the sequentially diverse chromosomal regions (Table 1).

The observed differences, in the presence of restriction target sites, grouped *T. p. pallidum* strains into a separate cluster compared to *T. p. pertenue* strains. The Fribourg-Blanc isolate, although more distantly separated, was clustered with TPE strains (Fig. 1). Analysis of the *tpnC* and *tpnI* gene phylogeny (Gray *et al.*, 2006) revealed similar close relationships between the Fribourg-Blanc treponemes and *T. p. pertenue* strains. The Fribourg-Blanc isolate is infectious to humans and is able to cause symptoms of yaws (Smith, 1971; Smith *et al.*, 1971). Although the genome analysis of *T. pallidum* ssp. *endemicum* strain Bosnia A (Grin, 1952) has not yet been completed, the preliminary analysis of more than a quarter of the genome (25.6%) has revealed a relatedness among the TEN Bosnia A strain, TPE strains, and the Fribourg-Blanc treponeme (Fig. 1, panel B). The observed relatedness between the TPE and the Fribourg-Blanc strains suggests a possible common origin of these strains and potentially indicates treponemal strain transmission between humans and African primates. Since yaws and simian treponemal infections occur in overlapping geographic territories, human treponemal pathogens may have originated in Africa (Livingstone, 1991). The observed restriction target site diversity among TPA strains indicates the presence of two

separated groups (the Nichols group and the SS14 group) of TPA strains that coexist in the human population.

Strain name	Species/ Subspecies*	Place and year of isolation	Revealed genome size (kb)	Revealed genome sequence identity with Nichols (%)	Reference
Nichols	TPA	Washington, DC; 1912	1139.6†	100	Mikalová <i>et al.</i> , 2010; Nichols & Hough, 1913
DAL-1	TPA	Dallas; 1991	1139.9	99.98	Mikalová <i>et al.</i> , 2010; Wendel <i>et al.</i> , 1991
SS14	TPA	Atlanta; 1977	1139.5	99.92	Mikalová <i>et al.</i> , 2010; Stamm <i>et al.</i> , 1983
Mexico A	TPA	Mexico; 1953	1140.0	99.93	Turner & Hollander, 1957; Mikalová <i>et al.</i> , 2010
Samoa D	TPE	Western Samoa; 1953	1139.3	99.64	Turner & Hollander, 1957; Mikalová <i>et al.</i> , 2010
CDC-2	TPE	Akorabo, Ghana; 1980	1139.7	99.63	Liska <i>et al.</i> , 1982; Mikalová <i>et al.</i> , 2010
Gauthier	TPE	Congo; 1960	1139.4	99.64	Gastinel <i>et al.</i> , 1963; Mikalová <i>et al.</i> , 2010
Fribourg-Blanc	?	Guinea; 1966	1140.4	99.57	Fribourg-Blanc & Mollaret, 1969; Mikalová <i>et al.</i> , 2010
Cuniculi A	TPC	?	1133.4	98.21	Strouhal <i>et al.</i> , 2007

* TPA = *T. pallidum* subsp. *pallidum*; TPE = *T. pallidum* subsp. *pertenue*; TPC = *T. paraluisuniculi*.
†The 1.2 kb *tprK*-like insertion in part of Nichols population (Šmajš *et al.*, 2002) was added to the previously published genome sequence (Fraser *et al.*, 1998).

Table 1. Genome size and revealed genome sequence identity with the Nichols genome of *T. p. pallidum*, *T. p. pertenue*, *T. paraluisuniculi* and the Fribourg-Blanc strains.

The WGF technique also identified genomic regions showing variability in most investigated strains including the intergenic region between genes TP0126 and TP0127, and in the *arp*, TP0470, and TP0967 genes. Among the investigated TPA and TPE genomes, the *tprK*-like sequence inserted between the TP0126 and TP0127 genes was found in three different versions (Mikalová *et al.*, 2010). In the Nichols genome, this insertion was found only in part of the treponemal population (Šmajš *et al.*, 2002). With regard to the *arp* gene

(Pillay *et al.*, 1998), a variable number of tandem repetitions were found in the tested genomes. Based on amino acid variations, previously published papers (Harper *et al.*, 2008b; Liu *et al.*, 2007) classified the TPA and TPE Arp repeat motifs into 4 types (I, II, III, II/III), and the variability in repeat sequence types correlated with the sexual transmission strategy (Harper *et al.*, 2008b). The differences among tested strains were also found in a number of 24 bp tandem repeats of TP0470, a gene encoding a hypothetical protein, and in indels present in the hypothetical TP0967 gene (Mikalová *et al.*, 2010). As with the *tprK*-like insertion between the TP0126 and TP0127 genes, the number of 24 bp repetitions in TP0470 was reported to vary within individual bacterial isolates (Marra *et al.*, 2010).

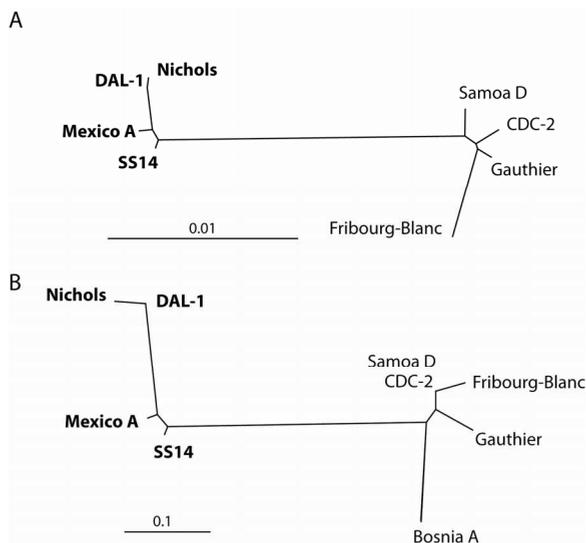


Fig. 1. Unrooted trees constructed from restriction target site data of the analyzed TPA and TPE genomes. Panel A: unrooted tree constructed from whole genome analyses. The Fribourg-Blanc isolate clusters with TPE strains. Panel B: unrooted tree constructed from 25.6% of the tested genomes. The Bosnia A strain clusters with TPE genomes, indicating a close relationship between this *T. pallidum* ssp. endemicum strain and TPE strains. The bar scale corresponds to 0.01 and 0.1 restriction target site (RTS) changes per RTS, respectively. TPA strains are shown in bold.

2.2 Whole genome sequencing

Historically, syphilis and yaws treponemes were considered to be separate species (based on differences in clinical manifestations of the corresponding diseases), but since 1984 they have been classified as subspecies (Smibert, 1984) based on DNA hybridization experiments (Miao and Fieldsteel, 1980).

The WGF technique revealed high sequence relatedness among all investigated genomes, with the most divergent, *T. paraluiscuniculi*, genome differing in less than 2% of the genome sequence (Strouhal *et al.*, 2007). These data indicated that complete, high-quality sequences were required for treponeme genome comparisons. The list of sequenced treponemal genomes and the status of sequencing is shown in Table 2.

With the exception of the Nichols and Chicago genomes, whole genome DNA sequencing has been performed using a combination of several approaches including comparative genome sequencing (CGS, Matějková *et al.*, 2008), 454 pyrosequencing (Margulies *et al.*, 2005) and the Solexa/Illumina method (Bennett, 2004). Isolated genomic DNA of most of the sequenced strains was amplified before genomic DNA sequencing. All discrepancies in CGS, 454 and Solexa/Illumina sequences were resequenced, using the dideoxyterminator sequencing method, until a final consensus sequence was obtained.

Treponemal strain	Treponeme	Sequencing method	Reference/GenBank accession number
Nichols	TPA	DDT	Fraser <i>et al.</i> , 1998
SS14	TPA	CGS	Matějková <i>et al.</i> , 2008
Chicago	TPA	Illumina	Giacani <i>et al.</i> , 2010
DAL-1	TPA	454, Illumina	unpublished
Mexico A	TPA	Illumina	unpublished
Samoa D	TPE	CGS, 454, Illumina	CP002374
CDC-2	TPE	454, Illumina	CP002375
Gauthier	TPE	454, Illumina	CP002376
Fribourg-Blanc	?	454, Illumina	unpublished
Cuniculi A	TPC	CGS, 454, Illumina	CP002103
Bosnia A	TEN	454, Illumina	unpublished

Table 2. Whole genome sequencing of uncultivable treponemal strains

Sequencing of the *T. paraluisuniculi* genome revealed 99.16% sequence identity (Šmajš *et al.*, 2011) of the conserved regions of the Nichols and Cuniculi A genomes. The identity between TPA and TPE genomes, greater than 99.8%, was found during sequencing of three TPE genomes (Čejková *et al.*, unpublished data). In all sequenced genomes, no major genome rearrangements were found. Despite the different clinical manifestations and host specificities, a nearly identical gene order was found in TPA, TPE, and *T. paraluisuniculi* strains, further establishing the close genetic relationship between these treponemal pathogens. The accuracy of genome assemblies and the sequencing error rate were estimated using the WGF approach and revealed high quality genome sequences with an error rate less than 10^{-4} . All investigated TPE strains were very similar in genome size with only 414 bp difference between the largest, CDC-2, and the smallest, Samoa D, genome. Nucleotide diversity (π) among sequenced TPE genomes was quite low (0.00032). In contrast, the nucleotide divergence (d_A) between TPA and TPE genomes was 3.6 - 4.6 times higher than the observed nucleotide diversity among each subspecies. These data indicate a significant evolutionary relationship between yaws and syphilis strains. Sequencing of additional TPA and TPE strains in the future will result in decreased numbers of genetic differences relevant to clinical manifestations of yaws and syphilis treponemes.

Altogether, 13 pseudogenes were found in the TPE genomes. In addition to pseudogenes, the genetic changes were analyzed in 970 similarly annotated protein-coding genes in both TPE and TPA strains. Compared to TPA strains, 70.4% of TPE genes encoded either identical proteins or identical proteins with strain specific differences; 194 (19.7%) genes encoded proteins with 1 amino acid substitution found in all tested TPE strains, 63 (6.4%) genes

encoded proteins with 2 to 5 amino acid changes, and only 34 (3.5%) genes encoded proteins with 6 or more amino acid replacements or other major protein changes.

Major sequence changes between TPA and TPE treponemes were found in the sequence for ethanolaminephosphotransferase, which is a pseudogene in TPE strains (TPE_0671), and in additional 12 genes with predicted functions including 8 *tpr* genes. The Tpr proteins are heterogenous proteins considered as potential virulence factors involved in pathogenesis and/or immune evasion (Giacani *et al.*, 2004; Gray *et al.*, 2006), and in inducing an antibody response during treponemal infection (Centurion-Lara *et al.*, 1999, 2000a, 2000b). Other differences between TPA and TPE involved a Mcp (methyl-accepting chemotaxis protein, TPE_0488), three treponemal antigens (TPE_0136; (Brinkman *et al.*, 2008), the Tp92 – outer membrane protein (TPE_0326; Cameron *et al.*, 2000), the Arp protein (TPE_0433, Pillay *et al.*, 1998)) and an elongated RecQ protein (TPE_0103) (Čejková *et al.*, unpublished results).

3. Targets for TPA strain-specific molecular diagnostics

3.1 Multilocus analyses of treponemal strains

The whole genome analyses of treponemal genomes revealed chromosomal regions with accumulated genetic diversity between TPA and TPE strains (Fig. 2).

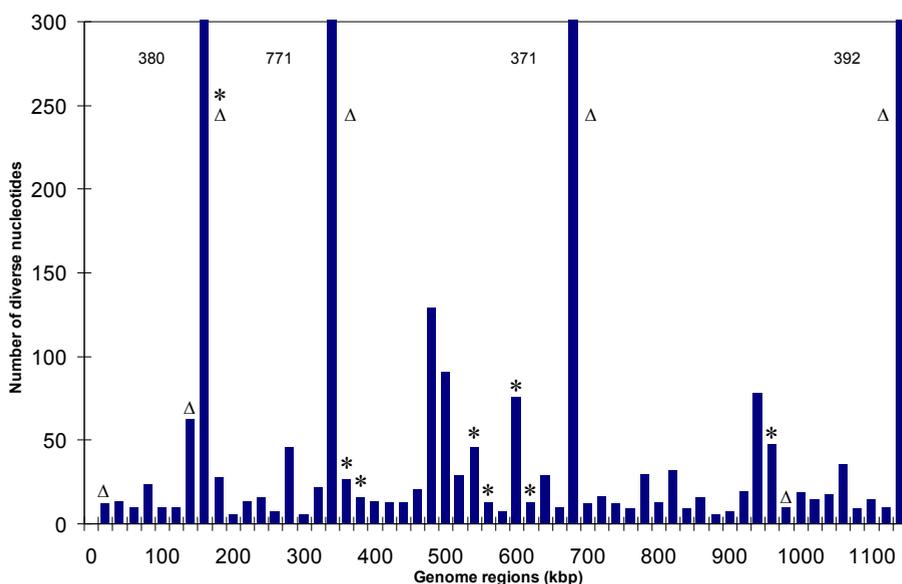


Fig. 2. Plot of numbers of nucleotide changes in 20 kb intervals between TPA and TPE strains along the treponemal chromosome. The exact number of nucleotide changes in the four most diverse regions is shown next to each column. Positions of selected chromosomal loci (see Table 3 and 4) are shown by asterisks. Positions of *tpr* genes in the treponemal genomes are shown with triangles (Δ).

In these regions, we determined the most genetically diverse genes and analyzed them in a set of TPA strains. In addition to four TPA strains (Nichols, SS14, DAL-1 and Mexico A),

TPA strains Grady (Atlanta, 1980), MN-3 (Minnesota, unknown), Philadelphia-1 (Philadelphia, 1988), Philadelphia-2 (Philadelphia, unknown), Bal-73-01 (Baltimore, 1973) were analyzed. All these strains were kindly provided by D. L. Cox, CDC, Atlanta, GA. The results of this analysis are summarized in the Tables 3 and 4. The greatest observed nucleotide difference was found in the TP0136 locus, followed by TP0548, TP0326 and TP0488. Interestingly, all investigated strains split into two subclusters containing either the Nichols or the SS14 strain (see Fig. 3).

Strain	Differences in the nucleotide sequences, %, (difference in number of nt)			
	TP0136	TP0326	TP0488	TP0548
Nichols	0	0	0	0
SS14	4.4	0.5	0.2	3.4
Mexico A	4.5	0.4	1.1	3.7
DAL-1	4.0	0 (0 nt)	0 (0 nt)	0 (0 nt)
Grady	4.4	0.5	0.1	3.4
MN-3	0.1	0 (1 nt)	0.1	0.8
Philadelphia 1	4.4	0.5	0.1	3.4
Philadelphia 2	0.1	0 (1 nt)	0.1	0.8
Bal-73-01	0 (0 nt)	0 (0 nt)	0 (1 nt)	0 (0 nt)

Table 3. Analysis of 4 chromosomal loci as potential targets for PCR detection and typing of clinical treponemal samples. The table shows the percentage of nucleotide differences compared to sequences present in the Nichols genome (Fraser *et al.*, 1998).

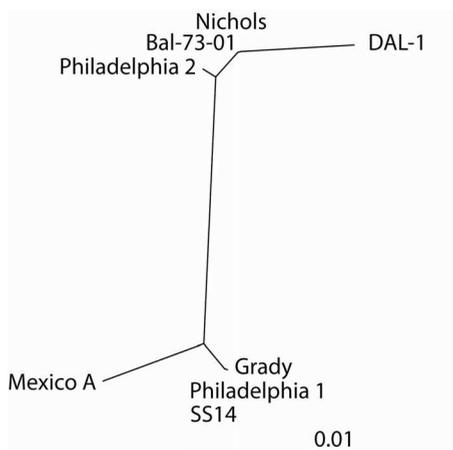


Fig. 3. The unrooted tree constructed from the nucleotide region resulting from concatenation of TP0136, TP0326, TP0488, and TP0548 loci sequenced in several TPA strains (of lengths ranging from 8342 to 8412 nucleotides). The bar scale corresponds to 0.01 nt changes per site. Please note that the TPA strains subcluster into two groups, one associated with the Nichols and the second associated with the SS14 strain.

In addition to TP0136, TP0548, TP0326 and TP0488 loci, we also tested other candidate chromosomal regions including TP0346, TP0515, TP0558, and TP0868 genes. The observed nucleotide diversity of these loci is shown in Table 4. The nucleotide diversity was considerably smaller among the latter group of genes (TP0346, TP0515, TP0558, and TP0868) indicating that their potential for testing of clinical TPA samples is lower than that of loci shown in Table 3. As in the previous case, Nichols-like strains were extremely similar to each other and distinct from SS14-like strains, reflecting different evolutionary relationships between these strains.

Strain	Number of detected nucleotide changes			
	TP0346	TP0515	TP0558	TP0868
Nichols	0	0	0	0
SS14	nd*	10	5	9
Mexico A	nd	nd	nd	2
DAL-1	0	0	0	0
Grady	2	nd	5	9
MN-3	0	nd	nd	0
Philadelphia 1	2	10	5	2
Philadelphia 2	nd	nd	nd	0
Bal-73-01	0	0	0	0

*nd, not determined

Table 4. Analysis of chromosomal loci as potential targets for PCR detection and typing of treponemes. The table shows the number of detected nucleotide changes compared to sequences present in the Nichols genome (Fraser *et al.*, 1998).

3.2 PCR analyses of clinical samples

Chromosomal genes including TP0136, TP0326, TP0488, TP0548, and TP0868, previously sequenced in TPA type strains, were used as DNA amplification targets from clinical samples containing treponemal DNA. The tested clinical samples were collected in the Czech Republic between years 2004 and 2010. The detected numbers of nucleotide changes were considerably lower than among the tested TPA type strains, indicating a genetic homogeneity of syphilis-causing strains in the Czech Republic. At locus TP0868, no diversity was observed among 5 tested clinical samples (see Table 5) and all TP0868 gene sequences were identical to the SS14 sequence. Interestingly, only strains identical or very similar to the SS14 strain were found among all investigated clinical strains (taken from 91 patients with sequenced treponemal DNA; unpublished results). Although not all gene sequences in the investigated strains were determined, the number of identified unique sequences (shown as unique a – d; Table 5) correlated with the number of nucleotide changes observed in Table 3, indicating that the most variable chromosomal regions identified during whole genome analyses were also the most variable among sequentially related clinical strains, i.e. causing syphilis, from particular geographic areas.

Clinical sample	TP0136	TP0548	TP0326	TP0488	TP0868	23S rDNA	No. of repetitions in the <i>arp</i> gene	<i>tprEGJ</i> genes
D-151	unique a*	SS14-like	nd**	nd	SS14-like	sensitive***	8	d****
3Z	unique b	SS14-like	nd	nd	nd	sensitive	14	d
2K	SS14-like	unique a	nd	nd	nd	A2058G	14	d
G-269	SS14-like	unique b	nd	nd	SS14-like	sensitive	14	p
RL73	nd	unique c	nd	nd	nd	A2059G	nd	d
1Z	nd	unique d	nd	nd	nd	A2058G	14	d
50AZ	SS14-like	nd	unique a	unique a	SS14-like	A2059G	15	d
11379	SS14-like	nd	SS14-like	unique a	SS14-like	sensitive	11	d
24K	SS14-like	SS14-like	nd	nd	nd	sensitive	14	d
40K	SS14-like	SS14-like	nd	nd	nd	A2058G	14	d
8284	SS14-like	SS14-like	nd	nd	nd	A2059G	15	d
24AZ	SS14-like	SS14-like	nd	nd	nd	A2059G	12	d
37K	nd	nd	nd	nd	nd	A2058G	12	d
No. of different genotypes	3	5	2	1	1	3	5	2

*unique a, sequencing of this locus revealed a sequence different from the SS14 sequence. Different sequences at a particular locus are denoted with letters a-d.

**nd, not determined

***sensitive, unmutated 23S rRNA resulting in susceptibility of TPA strains to macrolide antibiotics; A2058G or A2059G mutations result in resistance to macrolides.

****d, restriction patterns according to Pillay et al., (1998)

Table 5. Analysis of chromosomal loci as potential targets for PCR detection and typing of treponemes. Data for 13 selected samples isolated between the 2004 and 2010 from patients in Czech Republic are shown.

The data shown in Table 5 indicate numbers of different genotypes identified by sequencing analysis of several treponemal chromosomal loci. The greatest numbers of genotypes were found for the locus TP0548 (5) and for genetic variants detected by amplification of the *arp* gene (5). The use of the TP0548 locus in molecular typing of syphilis was first described by Flasarová *et al.* (2006) and this locus was recently incorporated to the enhanced molecular typing system of *Treponema pallidum* (Marra *et al.*, 2010). Analysis of TP0136 revealed 3 genotypes. Although the loci TP0326, TP0488 and TP0868 were sequenced in only a few of the presented isolates, one to two different genotypes were identified. Moreover, the unique sequences for the investigated TP loci appear to vary independently with each other and also vary independently with the number of repetitions in the *arp* gene and restriction profile of *tprEGJ* genes. This finding indicates the potential of using the above stated chromosomal loci in a detailed genetic identification of clinical samples. Although the typing of 23S rDNA locus revealed three genotypes, one encoding sensitivity to macrolide antibiotics and two coding for macrolide resistance (A2058G and A2059G mutations, respectively; Lukehart *et al.*, 2004; Matějková *et al.*, 2009; Stamm & Bergen, 2000), their use in molecular typing is probably limited by the fact that these mutations can be selected by use of macrolide antibiotics in a population. In the Czech Republic, more than 35% of clinical samples were found to contain a mutation encoding resistance to macrolide antibiotics (Flasarová *et al.*, unpublished results). In contrast, screening of 23S rDNA for A2058G in syphilitic samples taken from patients in Madagascar revealed no such mutation in 141 samples (Van Damme *et al.*, 2009). The recently improved CDC typing system (Marra *et al.*, 2010) relies on detection of the number of repetitions in the *arp* gene, on amplification and restriction digest analysis of *tprEGJ* genes and on sequencing of a part of the TP0548 gene.

4. Chromosomal targets for detection of non-TPA strains

4.1 Chromosomal targets for detection of TPE strains

Endemic treponematoses (caused by *T. pallidum* subsp. *pertenue*, *T. pallidum* subsp. *endemicum* and *T. carateum*) are estimated to currently affect more than 2.5 million people worldwide (Antal *et al.*, 2002). In the previous century, the number of yaws cases decreased from 50 million to a few million. In recent years, yaws has re-emerged in several rural populations in Africa, Asia and South America and a new effort to eradicate this disease has recently been undertaken (Asiedu *et al.*, 2008). Since single-dose penicillin is both cheap and available and no other disease reservoirs (besides humans and primates) are known, the chances for yaws eradication are relatively good. Molecular diagnosis of yaws treponemes in this situation is of fundamental importance. In the last two decades, several subtle genetic differences between TPA and TPE strains were published (Walker *et al.*, 1995; Centurion-Lara *et al.*, 1998; Centurion-Lara *et al.*, 2006). The most prominent indels common for all investigated TPE strains are shown in Table 6.

The regions listed in Table 6 and several additional regions (Mikalová *et al.*, 2010) need to be tested for other TPE strains before selecting the most suitable target for a molecular diagnosis of the yaws causing strains. Interestingly, all these indels are also found in the Fribourg-Blanc genome. However, the Fribourg-Blanc isolate can be differentiated based on the presence of specific indels (Mikalová *et al.*, 2010) as well as other individual TPE strains (Mikalová *et al.*, 2010).

Intergenic region (IGR) or gene(s)	Detected indel (bp)	GenBank accession no.
TP0266 (278334-278366)*	deletion (33 bp)	HM165228 Samoa D
		HM165229 Gauthier
		HM165230 CDC-2
		HM165231 Fribourg-Blanc
		HM585230 Samoa D
TP0316 (331265-331266)	insertion (635 bp)	HM585231 Gauthier
		HM585232 CDC-2
		HM585233 Fribourg-Blanc
		HM245777 Samoa D
IGR TP0548-TP0549 (593023-593024)	insertion (52 bp)	HM243496 Gauthier
		HM243495 CDC-2
		HM585227 Fribourg-Blanc
		HM623430 Samoa D
TP1030-TP1031 (1123987-1124363)	deletion (377 bp)	HM585235 Gauthier
		HM585236 CDC-2
		HM585254 Fribourg-Blanc

*The coordinates refer to the published Nichols genome (Fraser et al., 1998)

Table 6. Most prominent indels identified in all investigated TPE strains (Samoa D, CDC-2, and Gauthier). All these changes were also found in the Fribourg-Blanc genome (Mikalová et al., 2010).

4.2 Chromosomal targets for detection of TEN strains

The ongoing whole genome sequencing project of *T. p. endemicum* strain Bosnia A has already identified indels in at least 4 regions (ranging between 13 and ~60 bp), which can be used to differentiate the bejel treponeme from both the *T. p. pallidum* and *T. p. pertenue* strains (Table 7).

Intergenic region (IGR) or gene(s)	Detected indel (bp)	Coordinates in the Samoa D genome* (GenBank accession no. CP002374)
IGR TP0085-TP0086	deletion (13 bp)	94986-94998
TP0136	insertion (~0,06 kbp)	158205-158206
TP0326	insertion (15 bp)	348027-348028
TP0865	deletion (~25 bp)	945694-945718

*positions of detected indels in the TEN Bosnia A genome are shown as coordinates thereof in the Samoa D genome (GenBank accession no. CP002374)

Table 7. Indels identified in the TEN Bosnia A strain when compared to other investigated TPA and TPE strains.

5. Conclusions

The genomes of 9 pathogenic treponemes including *T. p. pallidum* strains (Nichols, SS14, DAL-1 and Mexico A), *T. p. pertenue* strains (Samoa D, CDC-2 and Gauthier), the Fribourg-Blanc isolate and *T. p. endemicum* (Bosnia A) were analyzed using several approaches

including whole genome fingerprinting and whole genome sequencing. Genome analyses revealed several important chromosomal loci suitable for diagnostic purposes including: i) syphilis-causing treponemes and their molecular typing, ii) yaws treponemes, and iii) bejel treponemes.

A sequencing-based typing scheme using simultaneous analysis of 3 loci (TP0136, TP0548 and 23S rDNA genes) in the *T. p. pallidum* genome was also evaluated. In addition, amplification of 23S rDNA locus and its subsequent restriction target analysis was used to detect mutations leading to macrolide resistance. The unique sequences in the investigated TP loci appear to combine independently with each other and also combine independently with the number of repetitions in the *arp* gene and restriction profiles of *tprEGJ* genes. Several genomic regions were found to differ between *T. p. pallidum* and *T. p. pertenue* strains and comprised indels ranging from 33 bp in the TP0266 gene to 635 bp in the *tprF* gene (TP0316). In all cases, the Fribourg-Blanc simian isolate showed changes similar to *T. p. pertenue* strains suggesting a close relationship to the *pertenue* subspecies. A partial genome analysis of *T. p. endemicum* strain, Bosnia A, showed, that this strain clustered with TPE strains, though more distantly than that of the Fribourg-Blanc isolate. The Bosnia A genome contained indels in at least 4 regions (ranging between 13 and ~60 bp) that can be used to differentiate the bejel treponeme from both *T. p. pallidum* and *T. p. pertenue* strains.

6. Acknowledgements

The authors thank Dr. David Cox for providing several TPA and TPE type strains. The TEN Bosnia A strain was kindly provided by S. Bruisten. This work was supported by grants from the U.S. Public Health Service to G.M.W. (R01 DE12488 and R01 DE13759), and S.J.N. (R01 AI49252 and R03 AI69107) and by grants of the Grant Agency of the Czech Republic (310/07/0321), of the Ministry of Health of the Czech Republic (NT11159-5/2010), and the Ministry of Education of the Czech Republic (VZ MSM0021622415) to D.S.

7. References

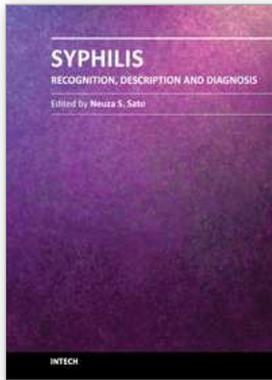
- Antal, G.M., Lukehart, S.A. & Meheus, A.Z. (2002). The endemic treponematoses. *Microbes and Infection*, Vol.4, No.1, (January 2002), pp. 83-94, ISSN 1286-4579
- Asiedu, K., Amouzou, B. & Dhariwal, A. (2008). Yaws eradication: past efforts and future perspectives. *Bulletin of the World Health Organization*, Vol.86, No.7, (July 2008), pp. 499, ISSN 0042-9686
- Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics*, Vol.5, No.4, (June 2004), pp. 433-438, ISSN 1462-2416
- Brinkman, M.B., McKeivitt, M. & McLoughlin, M. (2006). Reactivity of antibodies from syphilis patients to a protein array representing the *Treponema pallidum* proteome. *Journal of Clinical Microbiology*, Vol.44, No.3, (March 2006), pp. 888-891, ISSN 0095-1137
- Brinkman, M.B., McGill, M.A. & Pettersson, J.T. (2008). A novel *Treponema pallidum* antigen, TP0136, is an outer membrane protein that binds human fibronectin. *Infection and Immunity*, Vol.76, No.5, (May 2008), pp. 1848-1857, ISSN 0019-9567
- Cameron, C.E., Lukehart, S.A. & Castro, C. (2000). Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies *pallidum* Tp92. *The*

- Journal of Infectious Diseases*, Vol.181, No.4, (April 2000), pp. 1401-1413, ISSN 0022-1899
- Centurion-Lara, A., Castro, C. & Barrett, L. (1999). *Treponema pallidum* major sheath protein homologue TprK is a target of opsonic antibody and the protective immune response. *The Journal of Experimental Medicine*, Vol.189, No.4, (February 1999), pp.647-656, ISSN 0022-1007
- Centurion-Lara, A., Castro, C. & Castillo, R. (1998). The flanking region sequences of the 15-kDa lipoprotein gene differentiate pathogenic treponemes. *The Journal of Infectious Diseases*, Vol.177, No.4, (April 1998), pp. 1036-1040, ISSN 0022-1899
- Centurion-Lara, A., Castro, C. & Shaffer, J.M. (1997). Detection of *Treponema pallidum* by a sensitive reverse transcriptase PCR. *Journal of Clinical Microbiology*, Vol. 35, No. 6, (June 1997), pp. 1348-1352, ISSN 0095-1137
- Centurion-Lara, A., Godornes, C. & Castro, C. (2000b). The *tprK* gene is heterogeneous among *Treponema pallidum* strains and has multiple alleles. *Infection and Immunity*, Vol.68, No.2, (February 2000), pp. 824-831, ISSN 0019-9567
- Centurion-Lara, A., Molini, B.J. & Godornes, C. (2006). Molecular differentiation of *Treponema pallidum* subspecies. *Journal of Clinical Microbiology* Vol.44, No.9, (September 2006), pp. 3377-3380, ISSN 0095-1137
- Centurion-Lara, A., Sun, E.S. & Barrett, L.K. (2000a). Multiple alleles of *Treponema pallidum* repeat gene D in *Treponema pallidum* isolates. *Journal of Bacteriology*, Vol.182, No.8, (January 2000), pp. 2332-2335, ISSN 0021-9193
- Flasarová, M., Šmajš, D. & Matějková, P. (2006). Molecular detection and subtyping of *Treponema pallidum* subsp. *pallidum* in clinical specimens. *Epidemiologie, mikrobiologie, imunologie*, Vol.55, No.3, (August 2006), pp. 105-111, ISSN 1210-7913
- Fraser, C.M., Norris, S.J. & Weinstock, G.M. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science*, Vol.281, No.5375, (July 1998), pp. 375-388, ISSN 0036-8075
- Fribourg-Blanc, A., Mollaret, H.H. (1969). Natural treponematoses of the African primate. *Primates in medicine*, Vol.3, (n.d.), pp. 113-121
- Gastinel, P., Vaisman, A. & Hamelin, A. (1963). Study of a recently isolated strain of *Treponema pertenue*. *La Prophylaxie sanitaire et morale.*, Vol.35, (July 1963), pp. 182-188
- Giacani, L., Jeffrey, B.M. & Molini, B.J. (2010). Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. *Journal of Bacteriology*, Vol.192, No.10, (May 2010), pp. 2645-2646, ISSN 0021-9193
- Giacani, L., Sun, E.S. & Hevner, K. (2004). Tpr homologs in *Treponema paraluis-cuniculi* Cuniculi A strain. *Infection and Immunity*, Vol.72, No.11, (November 2004), pp. 6561-6576, ISSN 0019-9567
- Gray, R.R., Mulligan, C.J. & Molini, B.J. (2006). Molecular evolution of the *tprC*, D, I, K, G and J genes in the pathogenic genus *Treponema*. *Molecular Biology and Evolution*, Vol.23, No.11, (August 2006), pp. 2220-2233, ISSN 0737-4038
- Grin, E.I. (1952). Endemic syphilis in Bosnia; clinical and epidemiological observations on a successful mass-treatment campaign. *Bulletin of the World Health Organization*, Vol.7, No.1, (n.d.), pp. 1-74, ISSN 0042-9686
- Harper, K.N., Ocampo, P.S. & Steiner, B.M. (2008a). On the origin of the treponematoses: a phylogenetic approach. *PLoS Neglected Tropical Diseases*, Vol.2, No.1, (November 2007), pp. e148, ISSN 1935-2735

- Harper, K.N., Liu, H. & Ocampo, P.S. (2008b). 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 Immunology and Medical Microbiology*, Vol.53, No.3, (August 2008), pp. 322-332, ISSN 1574-695X
- Liska, S.L., Perine, P.L. & Hunter, E.F. (1982). Isolation and transportation of *Treponema pertenue* in golden hamsters. *Current Microbiology*, Vol.7, No.1, (n.d.), pp. 41-43, ISSN 0343-8651
- Liu, H., Rodes, B. & George, R. (2007). Molecular characterization and analysis of a gene encoding the acidic repeat protein (Arp) of *Treponema pallidum*. *Journal of Medical Microbiology*, Vol.56, No.6, (June 2007), pp. 715-721, ISSN 0022-2615
- Livingstone, F.B. (1991). On the origin of syphilis: an alternative hypothesis. *Current Anthropology*, Vol.32, No.5, (December 1991), pp. 587-590, ISSN 0011-3204
- Lukehart, S.A., Godornes, C. & Molini, B.J. (2004). Macrolide resistance in *Treponema pallidum* in the United States and Ireland. *New England Journal of Medicine*, Vol.351, No.2, (July 2004), pp. 154-158, ISSN 0028-4793
- MacDougall, J. & Saint Girons, I. (1995). Physical map of the *Treponema denticola* circular chromosome. *Journal of Bacteriology*, Vol.177, No.7, (April 1995), pp. 805-811, ISSN 0021-9193
- Margulies, M., Egholm, M. & Altman, W.E. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, Vol.437, No.7057, (September 2005), pp. 376-380, ISSN 0028-0836.
- Marra, Ch.M., Sahi, S.K. & Tantaló, L.C. (2010). Enhanced molecular typing of *Treponema pallidum*: geographical distribution of strain types and association with neurosyphilis. *Journal of Infectious Diseases*, Vol.202, No.9, (May 2010), pp. 1380-1388, ISSN 0022-1899
- Matějková, P., Strouhal, M. & Šmajš, D. (2008). Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. *BMC Microbiology*, Vol.8, No.1, (n.d.), pp. 76, ISSN 1471-2180
- Matějková, P., Flasarová, M. & Zákoucká, H. (2009). Macrolide treatment failure in a case of secondary syphilis: a novel A2059G mutation in the 23S rRNA gene of *Treponema pallidum* subsp. *pallidum*. *Journal of Medical Microbiology*, Vol.58, No.6, (June 2009), pp. 832-836, ISSN 0022-2615
- McKevitt, M., Patel, K. & Šmajš, D. (2003). Systematic cloning of *Treponema pallidum* open reading frames for protein expression and antigen discovery. *Genome Research*, Vol.13, No.7, (July 2003); pp. 1665-1674, ISSN 1088-9051
- McKevitt, M., Brinkman, M.B. & McLoughlin, M. (2005). Genome scale identification of *Treponema pallidum* antigens. *Infection and Immunity*, Vol.73, No.7, (July 2005), pp. 4445-4450, ISSN 0019-9567
- Miao, R.M. & Fieldsteel, A.H. (1980). Genetic relationship between *Treponema pallidum* and *Treponema pertenue*, two noncultivable human pathogens. *Journal of Bacteriology*, Vol. 141, No. 1, (n.d.), pp. 427-429, ISSN 0021-9193
- Mikalová, L., Strouhal, M. & Čejková, D. (2010). Genome analysis of *Treponema pallidum* subsp. *pallidum* and subsp. *pertenue* strains: Most of the genetic differences are localized in six regions. *PLoS One*, Vol.5, (December 2010), pp. e15713, ISSN 1932-6203

- Nichols, H.J. & Hough, W.H. (1913). Demonstration of *Spirochaeta pallida* in the cerebrospinal fluid. *Journal of the American Medical Association*, Vol.60, No.2, (n.d.), pp. 108-110, ISSN 0002-9955
- Pillay, A., Liu, H. & Chen, C.Y. (1998). Molecular subtyping of *Treponema pallidum* subspecies *pallidum*. *Sexually Transmitted Diseases*, Vol.25, No.8, (September 1998), pp. 408-414, ISSN 0148-5717
- Seshadri, R., Myers, G.S.A. & Tettelin, H. (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *PNAS*, Vol.101, No.15, (April 2004), pp. 5646-5651, ISSN 0027-8424
- Šmajš, D., McKeivitt, M. & Wang, L. (2002). BAC library of *T. pallidum* DNA in *E. coli*. *Genome Research*, Vol.12, No.3, (March 2002), pp. 515-522, ISSN 1088-9051
- Šmajš, D., McKeivitt, M. & Howell, J.K. (2005). Transcriptome of *Treponema pallidum*: gene expression profile during experimental rabbit infection. *Journal of Bacteriology*, Vol.187, No.5, (March 2005), pp. 1866-1874, ISSN 0021-9193
- Šmajš, D., Matějková, P. & Woznicová, V. (2006). Diagnosis of syphilis by polymerase chain reaction and molecular typing of *Treponema pallidum*. *Reviews in Medical Microbiology*, Vol.17, No.4, (December 2006), pp. 93-100, ISSN 0954-139X
- Šmajš, D., Zobaniková, M. & Strouhal, M. (2011). Complete genome sequence of *Treponema paraluis-cuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. *PLoS One*, Vol.6, No.5, (May 2011), pp. e20415, ISSN 1932-6203
- Smibert, R.M. (1984). Genus III: *Treponema* Schaudinn 1905, 1728^{AL}. In: *Bergey's Manual of Systematic Bacteriology*, Krieg, N.R. & Holt, J.G., (Ed.), pp. 49-57, Williams&Wilkins, ISBN 0683041088, Baltimore, MD, USA
- Smith, J.L. (1971). Neuro-ophthalmological study of late yaws. I. An introduction to yaws. *British Journal of Venereal Diseases and Genitourinary Medicine*, Vol.47, No.4, (January 1971), pp. 223-225
- Smith, J.L., David, N.J. & Indgin, S. (1971). Neuro-ophthalmological study of late yaws and pinta. II. The Caracas project. *British Journal of Venereal Diseases and Genitourinary Medicine*, Vol.47, No.4, (January 1971), pp. 226-251
- Stamm, L.V., Kerner, T.C. Jr & Bankaitis, V.A. (1983). Identification and preliminary characterization of *Treponema pallidum* protein antigens expressed in *Escherichia coli*. *Infection and Immunity*, Vol.41, No.2, (August 1983), pp. 709-721, ISSN 0019-9567
- Stamm, L.V. & Bergen, H.L. (2000). A point mutation associated with bacterial macrolide resistance is present in both 23S rRNA genes of an erythromycin-resistant *Treponema pallidum* clinical isolate. *Antimicrobial Agents and Chemotherapy*, Vol.44, No.3, (Marz 2000), pp. 806-807, ISSN 0066-4804
- Strouhal, M., Šmajš, D. & Matějková, P. (2007). Genome differences between *Treponema pallidum* subsp. *pallidum* strain Nichols and *T. paraluis-cuniculi* strain Cuniculi A. *Infection and Immunity*, Vol.75, No.12, (December 2007), pp. 5859-5866, ISSN 0019-9567
- Titz, B., Rajagopala, S.V. & Goll, J. (2008). The binary protein interactome of *Treponema pallidum*-the syphilis spirochete. *PLoS One*, Vol.3, No.5, (May 2008), pp. e2292, ISSN 1932-6203
- Turner, T.B. & Hollander, D.H. (1957). Biology of the treponematoses based on the studies carried out at the International Treponematoses Laboratory Center of the Johns

- Hopkins University under the auspices of the World Health Organization. *Monograph Series of the World Health Organization*, Vol.35, pp. 3-266, ISSN 0512-3038
- Van Damme, K., Behets, F. & Ravelomanana, N. (2009). Evaluation of azithromycin resistance in *Treponema pallidum* specimens from Madagascar. *Sexually Transmitted Diseases*, Vol.36, No.12, (December 2009), pp. 775-776, ISSN 0148-5717
- Walker, E.M., Howell, J.K. & You, Y. (1995). Physical map of the genome of *Treponema pallidum* subsp. *pallidum* (Nichols). *Journal of Bacteriology*, Vol.177, No.7, (April 1995), pp. 1797-1804, ISSN 0021-9193
- Weinstock, G.M., Norris, S.J. & Sodergren, E. (2000). Identification of virulence genes *in silico*: infectious disease genomics, In: *Virulence mechanisms of bacterial pathogens*, Brogden, K.A., Roth, J.A., Stanton, T.B., Bolin, C.A., Minion, F.C., Wannemuehler, M.J., (Ed.), 251-261, ASM Press, ISBN 1-55581-174-4, Washington, DC, USA
- Wendel, G.D., Jr, Sanchez, P.J. & Peters, M.T. (1991). Identification of *Treponema pallidum* in amniotic fluid and fetal blood from pregnancies complicated by congenital syphilis. *Obstetrics and Gynecology*, Vol.78, No.5, (November 1991), pp. 890-894, ISSN 0029-7844



Syphilis - Recognition, Description and Diagnosis

Edited by Dr. Neuza Satomi Sato

ISBN 978-953-307-554-9

Hard cover, 130 pages

Publisher InTech

Published online 21, November, 2011

Published in print edition November, 2011

Syphilis, a sexually transmitted disease was first described in 15th century, is caused by *Treponema pallidum* subsp. *pallidum* and occurs worldwide. This book is a collection of chapters presenting the novel knowledge about the *T. pallidum* and some historical and up to date information about venereal disease and syphilis. The collection of articles includes: immunological aspects recognition of *T. pallidum* by the pattern recognition receptors of the innate immune; the whole genome analysis of treponemes and new targets for its molecular diagnosis; some historical aspects of venereal diseases treatment; natural history of syphilis including clinical manifestation and epidemiology; a clinical aspects dealing with psychiatric manifestations of neurosyphilis; spatial and temporal patterns of primary syphilis and secondary syphilis described by the spatial and space-time scan statistics; a commonly used methods for laboratorial diagnosis, the serological response to treatment of syphilis and safety in blood transfusion. I hope this book will be useful for students and research fellows as well for the wide audience.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

David Šmajš, Lenka Mikalová, Darina Čejková, Michal Strouhal, Marie Zobaňíková, Petra Pospíšilová, Steven J. Norris and George M. Weinstock (2011). Whole Genome Analyses of Treponemes: New Targets for Strain and Subspecies-Specific Molecular Diagnostics, *Syphilis - Recognition, Description and Diagnosis*, Dr. Neuza Satomi Sato (Ed.), ISBN: 978-953-307-554-9, InTech, Available from:

<http://www.intechopen.com/books/syphilis-recognition-description-and-diagnosis/whole-genome-analyses-of-treponemes-new-targets-for-strain-and-subspecies-specific-molecular-diagnosis>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.