

RESEARCH ARTICLE

# Sequencing of *Treponema pallidum* subsp. *pallidum* from isolate UZ1974 using Anti-Treponemal Antibodies Enrichment: First complete whole genome sequence obtained directly from human clinical material

Linda Grillová<sup>1‡</sup>, Lorenzo Giacani<sup>2</sup>, Lenka Mikalová<sup>1</sup>, Michal Strouhal<sup>1</sup>, Radim Strnadel<sup>3</sup>, Christina Marra<sup>2</sup>, Arturo Centurion-Lara<sup>2</sup>, Lucy Poveda<sup>4</sup>, Giancarlo Russo<sup>4</sup>, Darina Čejková<sup>5</sup>, Vladimír Vašků<sup>6</sup>, Jan Oppelt<sup>7,8</sup>, David Šmajs<sup>1\*</sup>

**1** Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic, **2** Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, United States of America, **3** Department of Dermatovenereology, University Hospital Brno, Brno, Czech Republic, **4** Functional Genomics Center Zurich, University of Zurich, Zurich, Switzerland, **5** Department of Immunology, Veterinary Research Institute, Brno, Czech Republic, **6** 1<sup>st</sup> Dermatovenereological Clinic St. Anne's University Hospital Brno, Faculty of Medicine, Masaryk University, Brno, Czech Republic, **7** CEITEC-Central European Institute of Technology, Masaryk University, Brno, Czech Republic, **8** National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

‡ Current address: Biology of Spirochetes Unit, Institut Pasteur, Paris, France  
\* [dsmajs@med.muni.cz](mailto:dsmajs@med.muni.cz)



**OPEN ACCESS**

**Citation:** Grillová L, Giacani L, Mikalová L, Strouhal M, Strnadel R, Marra C, et al. (2018) Sequencing of *Treponema pallidum* subsp. *pallidum* from isolate UZ1974 using Anti-Treponemal Antibodies Enrichment: First complete whole genome sequence obtained directly from human clinical material. PLoS ONE 13(8): e0202619. <https://doi.org/10.1371/journal.pone.0202619>

**Editor:** Leonardo Mariño-Ramírez, National Institutes of Health, UNITED STATES

**Received:** June 15, 2018

**Accepted:** August 7, 2018

**Published:** August 21, 2018

**Copyright:** © 2018 Grillová et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The complete genome sequence of the UZ1974 sample was deposited in GenBank under the accession number CP028438. Raw data are available in SRA under the following accession number: SRP156463.

**Funding:** This work was supported by the Grant Agency of the Czech Republic (GA17-25455S) and by the Ministry of Health of the Czech Republic (17-31333A) to DS. Sequencing of the UW074B

## Abstract

*Treponema pallidum* subsp. *pallidum* (TPA) is the infectious agent of syphilis, a disease that infects more than 5 million people annually. Since TPA is an uncultivable bacterium, most of the information on TPA genetics comes from genome sequencing and molecular typing studies. This study presents the first complete TPA genome (without sequencing gaps) of clinical isolate (UZ1974), which was obtained directly from clinical material, without multiplication in rabbits. Whole genome sequencing was performed using a newly developed Anti-Treponemal Antibody Enrichment technique combined with previously reported Pooled Segment Genome Sequencing. We identified the UW074B genome, isolated from a sample previously propagated in rabbits, to be the closest relative of the UZ1974 genome and calculated the TPA mutation rate as  $2.8 \times 10^{-10}$  per site per generation.

## Introduction

*Treponema pallidum* subsp. *pallidum* (TPA) is the infectious agent of syphilis, a globally distributed, multi-stage, sexually transmitted disease with an annual incidence of more than 5.6 million cases, including about 350,000 cases of congenital syphilis [1, 2]. Since TPA cannot be continuously cultivated under *in vitro* conditions, most of the information on TPA genetics comes from genome sequencing studies and molecular typing studies [3–5].

strain was supported by grant R21AI094126 to LG. We acknowledge CF New Generation Sequencing Bioinformatics, supported by the CIISB research infrastructure (LM2015043 funded by the Ministry of Education, Youth and Sports of the Czech Republic), for their support with obtaining scientific data presented in this paper. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the program "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

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

The first complete genome sequence of TPA was published in 1998 [6] and since then, several other TPA genomes have been fully sequenced and analyzed ( $n = 6$ ) [7–12]. In all these cases, treponemal DNA was isolated from bacteria propagated in experimentally infected rabbits. The low number of completely sequenced TPA genomes reflects the limited number of available TPA strains propagated in rabbits as well as the limited number of TPA strains for which the treponemal DNA was purified in sufficient amounts for whole genome sequencing.

For many years, there have been attempts to obtain whole genome sequences of TPA directly from clinical samples, without treponemal replication in rabbits. These attempts were mainly motivated by the need to i) characterize TPA strains causing modern syphilis infections and ii) compare strains isolated directly from patients and strains propagated in rabbits to reveal any potential adaptation of TPA to the rabbit host. For years, whole genome sequencing of TPA from clinical material was hindered by the very low number of treponemes in clinical specimens and the massive contamination of human and other DNA that precluded efficient sequencing of TPA directly from clinical samples. There is approximately  $10^4$ -times less TPA DNA copies present in clinical samples isolated directly from patients ( $10^1$ – $10^4$  of TPA DNA copies per  $\mu\text{l}$  of sample) compared to samples propagated in rabbits [13].

This limitation was resolved with the introduction of the Pooled Segment Genome Sequencing (PSGS) approach [14–17], which allowed whole genome sequencing to be performed with a very small number of TPA DNA copies per sample ( $10^3$ – $10^4$  of TPA DNA copies per  $\mu\text{l}$  of sample). Briefly, this method is based on specific amplification of overlapping fragments of TPA DNA (average size of *Treponema pallidum* (TP) intervals = 10 kb), which together represent the whole genome. To overcome the mis-assembly of short reads generated by Next-Generation Sequencing (NGS), TP intervals were divided into four different pools that undergo NGS separately. However, since this technique is quite time-consuming, other techniques for culture-independent selective TPA DNA enrichments were developed. These techniques, introduced in 2016, were based on RNA baits or DNA microarray capture to selectively enrich TPA DNA directly from clinical samples. Since then, the number of sequenced TPA genomes has increased dramatically. In total, 43 draft TPA genomes with coverage greater than 80% have been determined [18, 19]. However, the TPA capture techniques failed to produce complete genome sequences since i) the baits are only available against known TPA sequences and ii) reference-guided or whole genome sequence *de novo* assemblies using relatively short sequencing reads produced by NGS platforms cannot cover treponemal paralogous regions and regions containing tandem repeats. Paralogous regions (including *tpr* genes), two copies of nearly identical RNA operons, and regions containing repetitive sequences, which represent approximately 2% of the length of TPA whole genomes, can not be determined using these approaches.

This study presents a new culture-independent method to sequence TPA directly from human clinical material. The method, designated Anti-Treponemal Antibody Enrichment (ATAE), is based on selective separation of TPA on the cellular level. In this work, ATAЕ is coupled with the previously developed PSGS approach. This study presents the first complete genome of TPA obtained directly from clinical material, without multiplication in rabbits.

## Materials and methods

### Clinical characteristics of the UZ1974 sample and ethics statement

The UZ1974 sample was collected on December 29<sup>th</sup>, 2014, from a male patient with a primary genital chancre (Department of Dermatovenerology, University Hospital Brno, Czech Republic). Treponemes from the sample were used for anti-treponemal antibody-based enrichment, DNA isolation, whole genome DNA amplification, and direct or PSGS whole genome

sequencing. This study was approved by the Ethics Committee of the Faculty of Medicine, Masaryk University (no. 25/2014); the patient signed an informed consent.

### Anti-Treponemal Antibodies Enrichment (ATAE)

**Enrichment of TPA cells.** TPA cells in the swab extract (PBS) were concentrated in the sample using polyclonal antibodies conjugated with biotin (Pierce™ *Treponema pallidum* Antibody PA1-73103; ThermoFisher Scientific, Waltham, MA, USA). Antibodies were bound to streptavidin coated magnetic beads (Dynabeads, CELLlection™ Biotin Binder Kit, ThermoFisher Scientific, Waltham, MA, USA). TPA cells were separated in the following steps: i) biotin-streptavidin binding (260 µl of beads, 240 µl of PBS, and 2 µl of anti-treponemal antibodies were mixed together and incubated at 4 °C overnight using end-over-end rotation); ii) removal of unbound antibodies (the mixture was twice washed with 500 µl of PBS); iii) incubation with swab extracts for one hour at room temperature (200 µl); iv) washing with PBS, and v) DNA isolation (QIAamp DNA Blood mini-kit, Qiagen, Hilden, Germany).

**Whole genome amplification (WGA).** WGA was carried out using Multiple Displacement Amplification with phi 29 polymerase (REPLI-g Single Cell Kit, Qiagen, Hilden, Germany). The WGA products were purified using QIAEX II beads according to the manufacturer's recommendations (Qiagen, Hilden, Germany).

### Quantification of TPA DNA and molecular typing

A nested PCR amplification (NPCR), with outer and inner primers targeting a single copy of the conserved treponemal gene encoding DNA polymerase (TP0105—*polA*), was used to quantify the number of copies of TPA DNA in the clinical sample as described previously [20]. This NPCR protocol was able to detect 1–10 molecules of TPA DNA [20]. Multi Locus Sequence Typing (MLST) was also performed. MLST determined the allelic profile with the three-letter code, where the first letter corresponds to the TP0136 allele, the second to the TP0548 allele, and the third to the TP0705 allele [21].

### Next-Generation Sequencing (NGS) and processing of sequencing data from enrichment

NGS was performed at the Functional Genomics Center Zurich, University of Zurich, Switzerland. Briefly, the NEB Next Ultra DNA Library Prep for Illumina (New England Biolabs, MA, USA) was used as described below. The samples were end-repaired, and adapters were ligated to the fragmented DNA samples. The samples were purified using Agencourt AMPure XP (Beckman Coulter Inc., Brea, CA, USA). Fragments containing adapters on both ends were selectively enriched with PCR. The quality and quantity of the resulting libraries were validated using Qubit® (1.0) Fluorometer and the TapeStation (Agilent, Waldbronn, Germany). The libraries were normalised to 10 nM and pooled equimolarly in Tris-Cl, pH 8.5 with 0.1% Tween 20. The resulting pool was sequenced on the Nextseq500 (Illumina, Inc, California, USA).

Quality of the raw reads was checked using FastQC [22]. The reads were pre-processed with the Cutadapt [23] and Fastx-toolkit [24]. First, the whole set of obtained reads was mapped to the human genome reference (hg38) and the human-matching reads were removed. Subsequently, the remaining reads were mapped to the TPA reference genome (GenBank Acc. No. CP004011.1). The mappings were performed using BWA MEM [25]. Mapping was post-processed using Samtools [26], Picard [27], and GATK [28]. Paralogous regions, regions containing repetitions and low-quality mappings were omitted from these analyses (mapping quality; MAPQ < 10). The overall mapping quality was checked using

Samtools [26], Samstat [29], and Qualimap [30]. Alignment-guided genome assembly (alignment consensus) was generated using Samtools [26].

### Pooled Segment Genome Sequencing (PSGS)

A WGA reaction diluted fifty times served as a template for TP intervals amplification during the PSGS phase as described previously [14, 31].

### Sequencing of TPI regions and processing of sequencing data from PSGS and *de novo* assembly

The amplified TP intervals ( $n = 279$ ) of the UZ1974 sample were NGS sequenced using the Illumina platform (NextSeq 500) at CEITEC (Brno, Czech Republic). Prior to NGS, the amplified TP intervals were labeled with multiplex identifier adapters and sequenced as four different samples to separate paralogous regions (Nextera™ XT DNA Sample Preparation Kit, Illumina Inc., Madison, WI, USA). The sequencing reads were trimmed (Trimmomatic, 0.32) [32], low quality bases were removed with a sliding window having a length of 4 nt, with an average quality of at least Phred = 17. When shorter than 50 bp, the sequencing reads were omitted from the analyses. Reads were analyzed with respect to four distinct pools and were *de novo* assembled using SeqMan NGen v4.1.0 software (DNASTAR, Madison, WI, USA) as well as mapped to the TPA reference genome (GenBank Acc. No. CP004011.1).

### Annotation of UZ1974 genome and nucleotide sequence accession number

For gene annotation, Geneious software v5.6.5 [33] was used. The *tprK* gene showed intra-strain variability and the corresponding nucleotides positions were denoted as “N” (coordinates: 975981–976013; 976114–976171; 976280–976336; 976402–976423; 976509–976534; 976656–976690; 977125–977156). The complete genome sequence of the UZ1974 sample was deposited in GenBank under the accession number CP028438. Raw data are available in SRA under the following accession number: SRP156463.

### Clinical characteristics and analyses of the UW074B sample

The UW074B sample was isolated from a syphilis-infected patient on July 1<sup>st</sup>, 2004, in Seattle, USA. The UW074B represented a human whole blood sample that was inoculated to rabbits and underwent two passages. New Zealand white rabbits were used for propagation of the UW074B strain and experimental infections. Animal care was provided in full accordance with the Guide for the Care and Use of Laboratory Animals and experimental procedures were conducted under protocol 2198.05 approved by the University of Washington Institutional Animal Care and Use Committee (IACUC).

### Extraction of UW074B DNA from rabbit tissue

Spirochetes were extracted in sterile saline from infected rabbit testicles and collected in 15 ml tubes. The suspensions were spun at 1,000 rpm for 10 minutes to remove rabbit tissue debris. The supernatant was transferred to microcentrifuge tubes and the bacteria was pelleted at 12,000 g for 30 min at 4 °C. Pellets were resuspended in 200 µl of lysis buffer (10 mM Tris pH 8.0, 0.1 M EDTA, 0.5% sodium dodecyl sulfate), and DNA extracted using a DNA Mini Kit (Qiagen Inc., Chatsworth, CA) according to the manufacturer’s instructions.

## NGS and bioinformatic analyses of UW074B sample

Extracted DNA was sequenced at Covance (Redmond, WA, USA) using the Illumina MiSeq platform. Quality of the raw reads was checked using FastQC [22]. The reads were pre-processed using Cutadapt [23]. First, the whole set of obtained reads was mapped using bbmap [34] to the rabbit genome reference (OryCun2.0) and the rabbit-matching reads were removed. Subsequently, the remaining reads were mapped to the TPA reference genome (GenBank Acc. No. CP004011.1). The mapping was performed using BWA MEM [25]. Mapping was post-processed using Samtools [26], Picard [27], and GATK [28]. Paralogous regions, regions containing repetitions, and low-quality mappings were omitted from these analyses (mapping quality/MAPQ < 10). Secondary, an improperly paired alignments were removed as well. The overall mapping quality was checked using Samtools [26] and Qualimap [30]. To avoid cross-mapping, several post-alignment filtering steps were added using Samtools [26], Picard [27], and NGSUtils/Bamutils [35]. The filtering kept only alignments with minimum 35 bp aligned length, a maximum of 5% mismatches of the mapped read length and/or a maximum of 5 mismatches, a maximum of 5% of soft-clipping, and 0% hard-clipping of the total read length, and a MAPQ  $\geq$  40. Alignment-guided genome assembly (alignment consensus) was generated using Samtools [26].

## Phylogenetic analyses

Maximum likelihood phylogenetic trees were generated using MEGA 6 with the Tamura Nei model and 1000 pseudorandom bootstrap replicates [36].

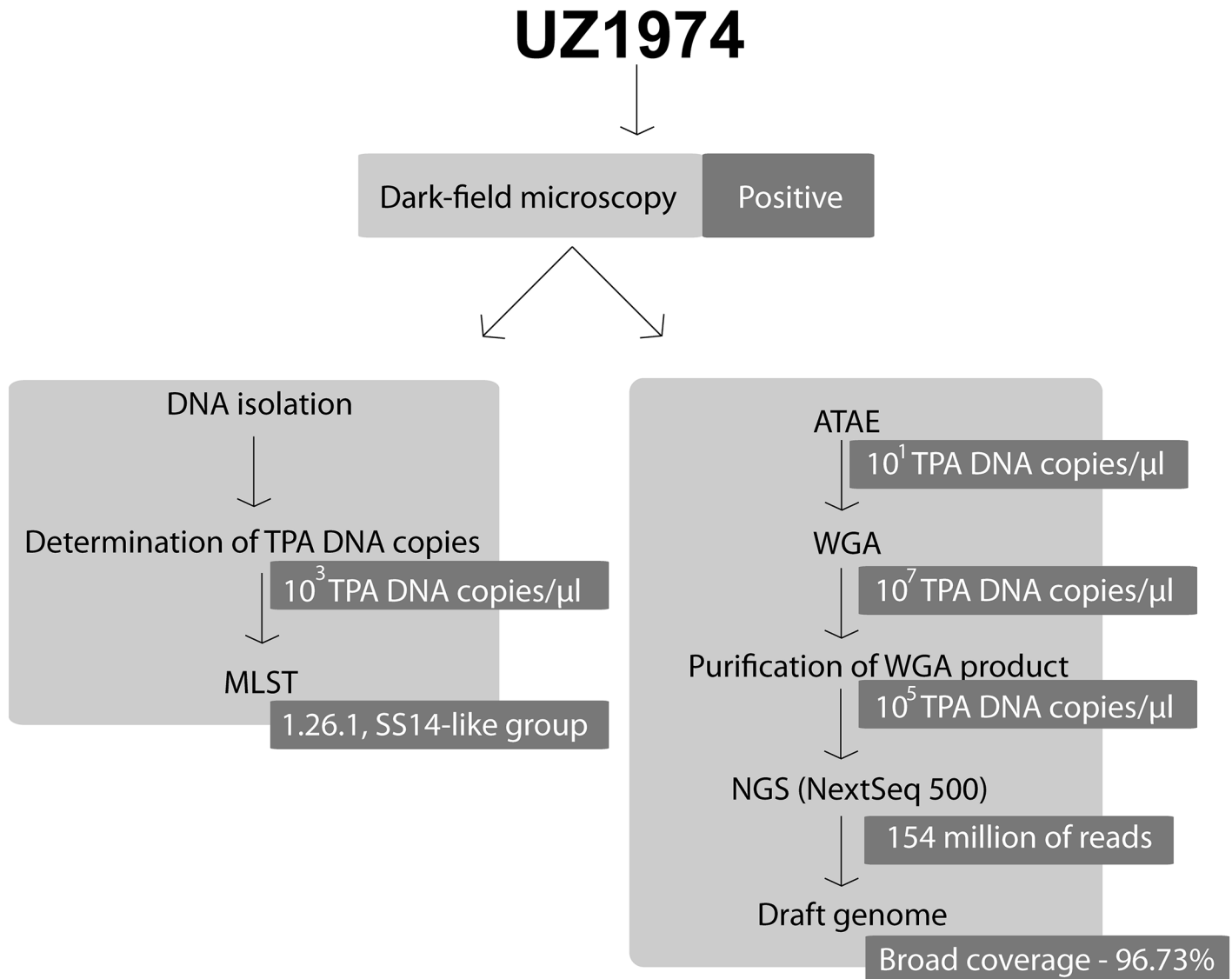
## Results

### Clinical characteristics, molecular typing, and number of TPA DNA copies in the UZ1974 sample

The primary chancre swab was taken from the genital region of a heterosexual patient (UZ1974) with primary syphilis who was infected by a sexual worker. The sample was collected at the Department of Dermatovenerology, University Hospital Brno, Czech Republic, in 2014. The swab extract was frozen in 10% glycerol at  $-80^{\circ}\text{C}$ . As revealed by molecular typing (MLST), the UZ1974 isolate belonged the SS14-like group of TPA strains (allelic profile 1.26.1) and contained an A2058G mutation in the 23S rRNA genes leading to macrolide resistance. The UZ1974 isolate was completely identical to the SS14 reference genome at the TP0136 locus (GenBank Acc. No. CP004011.1; allelic variant 1), contained three single nucleotide variants (SNVs) at the TP0548 locus (allelic variant 26) and two SNVs at TP0705 locus (allelic variant 1) compared to the SS14 reference genome (GenBank Acc. No. CP004011.1). The swab extract of the primary chancre was positive for dark-field microscopy suggesting that a relatively large number of treponemes were present in the sample. Prior to enrichment, we estimated  $10^3$  TPA DNA copies/ $\mu\text{l}$ , using established nested PCR protocol.

### Anti-Treponemal Antibodies Enrichment (ATAE)

The TPA cells present in the UZ1974 clinical sample were concentrated using polyclonal antibodies conjugated with biotin and bound to streptavidin coated magnetic beads (see [Material and methods](#) section). Following TPA enrichment, the total DNA was amplified with random primers and phi 29 polymerase (whole genome amplification; WGA). The WGA DNA products were then purified and sequenced using the Illumina platform (NextSeq 500). The workflow of ATAE and the whole DNA processing of clinical sample UZ1974 is shown in [Fig 1](#).



**Fig 1. Workflow of ATAE coupled with PSGS.** Dark-field microscopy, MLST, and determination of TPA DNA copies were performed on the UZ1974 clinical sample taken from a syphilis positive patient. TPA cells were concentrated in the sample using polyclonal antibodies conjugated with biotin, which were bound to streptavidin covered beads. Prior to NGS, whole genome amplification (WGA) was carried out using multiple displacement amplification using phi 29 polymerase; WGA products were purified using QIAEX II beads. The number of TPA DNA copies was monitored using the nested PCR protocol for *poIA* detection [20]. Using the BWA MEM algorithm, the whole set of obtained reads from NGS (Illumina NextSeq 500) was mapped to the human genome reference (hg38), removed, and the rest of the reads were mapped to the TPA reference genome (GenBank Acc. No. CP004011.1).

<https://doi.org/10.1371/journal.pone.0202619.g001>

As revealed by the pilot experiments done during ATAE development, the number of TPA DNA copies synthesized during WGA was directly dependent on the presence and concentration of contaminating (mostly human) DNA. Testing of WGA efficiency revealed that only a small amount of human DNA (3 ng) mixed with the positive control of TPA DNA (10 ng) decreased TPA amplification over 100 times (Grillová L., unpublished data). Moreover, in an unenriched UZ1974 clinical sample, the WGA procedure increased the number of TPA DNA copies by 2 orders of magnitude (to  $10^5$  TPA DNA copies/ $\mu$ l). The UZ1974 sample enriched by ATAE revealed  $10^1$  TPA DNA copies/ $\mu$ l before WGA and  $10^7$  DNA copies/ $\mu$ l after WGA. WGA therefore increased the number of copies by 6 orders of magnitude. After DNA

**Table 1. NGS statistics for the UZ1974 genome obtained using ATAES.**

NGS parameter	UZ1974
Total number of obtained reads after quality control	154 million
Reads mapped to the TPA SS14 genome reference (GenBank Acc. No. CP004011.1)	198,765
Average coverage depth	24.76x
Median coverage depth	9.55x
Broad coverage of TPA reference genome <sup>a</sup>	96.73%
Mean length of reads; range (bp)	142.54; 70–150

<sup>a</sup>The broad coverage calculated from positions with coverage  $\geq 3$ .

<https://doi.org/10.1371/journal.pone.0202619.t001>

purification of WGA products, we were able to prepare an enriched UZ1974 sample with a total TPA DNA amount of 0.1 ng/ $\mu$ l relative to a total DNA concentration of 180 ng/ $\mu$ l, i.e. the sample contained 1,800 times more contaminating DNA than TPA DNA.

A total of 154 million Illumina reads were obtained. Since the UZ1974 isolate was in the SS14-like group of TPA strains, the genome sequence of the SS14 strain (GenBank Acc. No. CP004011.1) was used as the reference sequence during the reference-guided approach. A total of 198,765 reads mapped to the TPA SS14 reference corresponded to an average genome coverage depth of 24.76x. Broad coverage for UZ1974 was 96.73%. Other statistical data are presented in [Table 1](#).

**PSGS.** In parallel, the genome of UZ1974 was amplified using PSGS, which was used to verify the ATAES sequencing results. Moreover, PSGS unequivocally determined the chromosomal paralogous regions and regions containing repetitive sequences. The average sequencing coverage depth for all TP intervals ( $n = 279$ ) was 1070.31x. Given that only 3.27% of the genome length was uncovered by ATAES in the UZ1974 genome, only 37 kbp had to be sequenced from the amplified intervals (TPI;  $n = 16$ ) to obtain a complete genome sequence.

## Analysis of UZ1974 genome sequence

The TPA UZ1974 genome was found to be closely related to the TPA SS14 genome. The TPA UZ1974 genome contained fourteen 60 bp-long repetitions in the TP0433 gene (i.e., *arp*; acidic repeat protein), which is the same number found in the TPA SS14 *arp* gene. In addition, similarly to SS14, the UZ1974 genome showed the same structure of RNA operons, i.e., the sequence of 16S-5S-23S rRNA genes were identical in both operons and both had the same order of spacer pattern encoding tRNA-Ile / tRNA-Ala [37], within the first and second *rrn* operon, respectively. The 23S rDNA sequence in both operons harbored the A2058G mutation encoding resistance to macrolide antibiotics. In contrast to the SS14 genome containing ten 24 bp-long repetitions in the TP0470 gene (coding for a tetratricopeptide repeat containing protein) [38], there were eight 24 bp-long repetitions in the UZ1974 genome.

Compared to the TPA SS14 genome (GenBank Acc. No. CP004011.1), the UZ1974 genome differed in 18 single nucleotide variants (SNVs); 17 of which were found in genes (or in annotated open reading frames) and one was found in the intergenic region ([Table 2](#)). All but one of the SNVs located in open reading frames resulted in amino acid replacements in the corresponding proteins ([Table 2](#)). The majority of amino acid replacements were found in genes predicted to code for virulence factors, outer membrane proteins, and metabolic functions. In addition to SNV differences, there were 16 length differences in homopolymeric tracts between the SS14 and UZ1974 genomes ([S1 Table](#)).

**Table 2. Identified SNVs between UZ1974 and SS14 genomes.**

SNV <sup>a</sup>	ORF (Gene)	Gene product	Functional category	Syn/Nonsyn <sup>b</sup>
A94901C	TPASS_20085	PTS family fructose porter component IIA	Transport	Syn
G135108C	TPASS_20117	Tpr protein C	Virulence	Nonsyn (P534A)
C174177T	TPASS_20151	putative NADH dehydrogenase (ubiquinone), subunit RnfD	General metabolism	Nonsyn (V260I)
T333559C	IGR <sup>c</sup>	NA <sup>d</sup>	NA <sup>d</sup>	NA <sup>d</sup>
G342703A	TPASS_20324	putative outer membrane protein	Unknown	Nonsyn (A540T)
T364888C	TPASS_20341	UDP-N-acetylmuramate—L-alanine ligase	General metabolism	Nonsyn (L64P)
A522907G	TPASS_20488	methyl-accepting chemotaxis protein	Cell processes	Nonsyn (D195G)
C556154T	TPASS_0515	putative outer membrane protein	Unknown	Nonsyn (R456C)
G593294A	TPASS_20548	putative outer membrane protein	Unknown	Nonsyn (G53R)
G593298A	TPASS_20548	putative outer membrane protein	Unknown	Nonsyn (G53E)
A593912G	TPASS_20548	putative outer membrane protein	Unknown	Nonsyn (K145E)
C674219T	TPASS_20620	Tpr protein I	Virulence	Nonsyn (H46R)
A674227C	TPASS_20620	Tpr protein I	Virulence	Nonsyn (V48G)
T674233C	TPASS_20620	Tpr protein I	Virulence	Nonsyn (E51K)
T760092C	TPASS_20691	segregation and condensation protein ScpA	Cell processes	Nonsyn (K30R)
C772846T	TPASS_20705	bifunctional membrane carboxypeptidase/penicillin-binding protein	General metabolism	Nonsyn (M625V)
T773095C	TPASS_20705	bifunctional membrane carboxypeptidase/penicillin-binding protein	General metabolism	Nonsyn (G708S)
T861444C	TPASS_20793	hypothetical protein	Unknown	Nonsyn (L311R)

<sup>a</sup>Coordinates according to the SS14 reference genome (GenBank Acc. No. CP004011.1).

<sup>b</sup>Synonymous/nonsynonymous amino acid replacement.

<sup>c</sup>IGR—Intergenic region.

<sup>d</sup>NA—Not applicable.

<https://doi.org/10.1371/journal.pone.0202619.t002>

### Comparison of the UZ1974 genome with the UW074B genomic sequence

A phylogenetic analysis of the whole genome sequence of the UZ1974 isolate (1,139,510 nt in length), with all available genome sequences from reference TPA strains and clinical isolates (n = 69; [S2 Table](#), [S1](#) and [S2 Figs](#)) [8–12, 18, 19, 39–41], revealed that UZ1974 and the draft genome sequence of TPA strain UW074B were closely related. To fully assess the genetic relatedness of UW074B, the genome was reassembled from the SRA data in the same way as the UZ1974 genome; the assembly covered 99.2% of the reference genome length (8,885 nt were not determined due to the paralogous character of the sequenced regions and/or due to the presence of repetitive sequences ([S3 Table](#)). A comparison of the complete genome sequences of both UZ1974 and UW074B revealed genetic difference at only one nucleotide position within the TP0548 gene (G vs. A; position 593,912 according to the TPA SS14 genome; CP004011.1). An additional genetic difference between the UZ1974 and UW074B genome sequences involved a 9-nt long repetition (TCCCTCCCC) in the TP0967 gene (between coordinates 1,051,840–1051,866; according to the TPA SS14 genome; CP004011.1). While the UZ1974 genome contained three such repetitions, the UW074B genome contained four. In addition, there were ten differences in the length of the homopolymeric tracts ([S4 Table](#)).

### TPA mutation rate derived from comparison of the UZ1974 and UW074B genomes

A single nucleotide difference detected in both analyzed genomes collected with 10.45 years between sample collection dates combined with the analyzed genome regions having a total length of 1,130,625 nt, corresponds to a mutation rate of  $8.46 \times 10^{-8}$  per nucleotide site per



year. Since sites with intrastrain heterogeneity do not represent fixed mutations, they were excluded from the estimation of the TPA mutation rate. Similarly, expansions or reductions in the number of repetitive sequence motifs were not considered as mutations. Considering the long doubling time of TPA, equal to about 30 hours [42, 43], one can assume that 91,584 hours (10.45 years) between the isolation of the two samples corresponded to 2,748 treponemal generations. Assuming 292 generations per year, the estimated mutation rate number corresponds to a TPA mutation rate of  $2.8 \times 10^{-10}$  per site per generation.

## Discussion

Although new techniques allowing culture-independent selective TPA DNA enrichment [18, 19] coupled with NGS were developed in 2016, these techniques have fundamental limitations since they are based on RNA or DNA baits derived from previous sequencing data and could therefore enrich only DNA that is complementary to the DNA sequences used in the microarray or bead capture. Any potentially novel treponemal sequences will remain undetected during DNA enrichment. Therefore, our intention was to develop a technique that would not show such a bias. One possible solution for this problem is to directly sequence all the DNA from the sample, however, contamination with human DNA and other microbial DNA precludes efficient sequencing of treponemal DNA and complicates genome assembly in the chromosomal regions conserved among microbial species. In this study, we developed an Anti-Treponemal Antibodies Enrichment (ATAE) method based on enrichment of TPA cells using polyclonal anti-treponemal antibodies. The number of TPA and human DNA copies before and after whole genome amplification showed, that the enrichment step of TPA is quite efficient even though significant amounts of TPA DNA was lost during this step.

We used ATAE on one clinical sample taken from a syphilis positive patient. The TPA UZ1974 isolate (allelic profile 1.26.1), belonged to the SS14-like group of TPA [21, 44, 45] and represents one of the most frequent allelic profiles found in the Czech Republic in recent years [46]. As with the SS14 strain, the TPA UZ4974 isolate harbored an A2058G mutation in both 23S rRNA genes resulting in resistance to macrolide antibiotics. According to Enhanced CDC-typing [47], UZ1974 was subtype “g” according to TP0548, representing the most frequent subtype found in Australia [48] and Europe (including the Czech Republic [46], Italy [49], Denmark [50], France [51], Ireland [47], and Switzerland [21]). At the same time, subtype “g” also belongs to the SS14- $\Omega$  (omega-cluster of SS14-like strains), which is currently spreading [18].

Despite ATAE being a useful technique, the enrichment was not as efficient as expected. Even though we tried several modifications of ATAE, we were unable to achieve a better treponemal to human cell ratio. Modifications to the ATAE protocol included i) different incubation times, ranging from 30 min to 1.5 hours at room temperature, ii) using monoclonal antibodies instead of polyclonal antibodies, iii) using a sepharose medium instead of magnetic beads, and iv) different numbers of washing steps. The ratio of TPA DNA to human DNA (HDNA) was about 1:1,800, which roughly corresponded to the human to TPA cell ratio, indicating that ATAE enriched the treponemal DNA 10-times compared to unenriched sample. As stated above, most of the total obtained reads belonged to the human genome and were excluded. The 43% of the remaining reads belonged to other bacteria (i.e., not to TPA), mostly to bacteria from the family *Prevotellaceae* that include bacteria isolated from many types of human material. When comparing ATAE enrichment efficiency to other available TPA culture-independent enrichments, including hybridization capture [18] and in-solution capture [19], ATAE had a similar or lower enrichment efficiency. Pinto and colleagues [19] were able to achieve a TPA/HDNA enrichment ratio of 1/1–1/100 while Arora and colleagues [18] were

able to reach a TPA/HDNA enrichment ratio of 1/10-1/1000. Another ATAE disadvantage is linked to the fact that TPA cells need to be intact during the enrichment step, therefore, clinical samples need to be processed shortly after sampling (hours after sampling). On the other hand, when using ATAE, there is no introduction of a DNA enrichment bias as a consequence of a sequence-specific enrichment protocol. Moreover, enrichment on the cellular level has the potential to be used for transcriptomic and proteomic studies.

Irrespective of the culture-independent enrichments method used, including hybridization capture, in-solution capture, and ATAE, there is another problem with paralogous genome regions and regions containing repetitions precluding finishing of complete genome sequences. Many TPA genomes determined in our lab were sequenced by the PSGS technique [14–17] based on sequencing of amplified overlapping fragments covering the entire TPA genome. This method is quite laborious and time-consuming, however, until now, the only method, which is able to overcome the mis-assembly of short reads generated by NGS and thus truly determine the paralogous regions. In this study, we combined this approach with newly developed ATAE technique. The ATAE was able to generate only draft genome. The missing regions and paralogous regions were in the end established (the gaps were filled) with the sequencing data generated by PSGS. This allowed us to obtain the first complete genome sequence isolated directly from human material.

A phylogenetic analysis of the UZ1974 whole genome sequence, with all available genome sequences from reference TPA strains and clinical isolates ( $n = 69$ ; S2 Table, S1 and S2 Figs), revealed that the UZ1974 and the TPA strain UW074B draft genome sequence were closely related. The mutation rate calculated from the UZ1974 and UW074B genomes corresponded to a TPA mutation rate of  $2.8 \times 10^{-10}$  per site per generation (assuming 292 generations per year), a number that is even lower than the recently estimated upper limit for the TPE mutation rate, i.e.,  $4.1 \times 10^{-10}$  per site per generation [31]. In our previous work on yaws treponemes isolated from Ghana, Africa (TPE strain Ghana-051 and TPE CDC 2575, isolated 7.25 years apart), we estimated an upper mutation rate limit of  $1.21 \times 10^{-7}$  per nucleotide site per year (genome size: 1,139,577 nt). Since both strains, TPE Ghana-051 and TPE CDC 2575, had the same consensus genome sequence, the upper limit of the mutation rate in yaws treponemes was estimated as  $4.1 \times 10^{-10}$  per site per generation [31]. In this study, the mutation rate estimation assumes that the TPA present in the UW074B sample was directly transmitted to other patients that led to infection of patient UZ1974. In reality, this is not the most probable scenario. Instead, it is more likely that both patients were infected by descendants of a common ancestor of the UW074B and UZ1974 strains and the evolutionary distance between treponemes in both samples was therefore longer than the one used for mutation rate estimation. It is therefore likely that the real mutation rate is even lower, making this estimation of TPA mutation rate ( $2.8 \times 10^{-10}$  per site per generation) probably close to the highest rate possible.

## Supporting information

**S1 Fig. Phylogeny for all TPA genome sequences available in GenBank.** Maximum likelihood phylogenetic tree generated in MEGA 6 for genome-wide variable positions ( $n = 419$ ) after excluding sites with missing data using all available TPA genomes ( $n = 69$ ) and the examined UZ1974 genome. Draft genomes used had a broad coverage of 90% or more. Repetitive and paralogous regions were not included in the analyses. (TIF)

**S2 Fig. Phylogeny for TPA genome sequences available in GenBank.** Maximum likelihood phylogenetic tree generated in MEGA 6 for genome-wide variable positions ( $n = 1081$ ) after excluding sites with missing data using available TPA genomes ( $n = 49$ ) and the examined

UZ1974 genome. Only draft genomes with broad coverage of 98.5% or more were used. Repetitive and paralogous regions were not included in the analyses.

(TIF)

**S1 Table. Differences in homopolymers found when comparing the UZ1974 isolate to the SS14 strain (GenBank Acc. No. CP004011.1).**

(DOCX)

**S2 Table. All available TPA genome sequences available in GenBank database.** This data was used for phylogenetic tree reconstructions in S1 and S2 Figs.

(DOCX)

**S3 Table. A list of chromosomal regions from the UW074B genome sequence that were excluded from further analyses due to unambiguous mapping of the sequencing reads.** Altogether, 8,885 nt out of the total genome length (0.8%) were not analyzed in the UW074B genome. In contrast to assembly of the UW074B genome, assembly of the UZ1974 genome sequence was based on PSGS, which allowed assembly of a complete genome sequence.

(DOCX)

**S4 Table. Differences in homopolymers found when comparing the UZ1974 isolate to the UW074B genome sequence.**

(DOCX)

## Acknowledgments

We thank Thomas Secrest (Secrest Editing, Ltd.) for his assistance with the English revision of the manuscript.

## Author Contributions

**Conceptualization:** Linda Grillová, David Šmajš.

**Data curation:** Linda Grillová, Lorenzo Giacani, Lenka Mikalová, Lucy Poveda, Giancarlo Russo, Darina Čejková, Jan Oppelt.

**Formal analysis:** Linda Grillová, Jan Oppelt.

**Funding acquisition:** Lorenzo Giacani, David Šmajš.

**Investigation:** Linda Grillová, Lenka Mikalová, Michal Strouhal, Christina Marra, Jan Oppelt.

**Methodology:** Linda Grillová, Jan Oppelt.

**Resources:** Lorenzo Giacani, Radim Strnadel, Christina Marra, Arturo Centurion-Lara, Vladimír Vašků.

**Software:** Jan Oppelt.

**Supervision:** David Šmajš.

**Visualization:** Linda Grillová, Jan Oppelt.

**Writing – original draft:** Linda Grillová.

**Writing – review & editing:** Lorenzo Giacani, Jan Oppelt, David Šmajš.

## References

1. World Health Organization. WHO guidelines for the treatment of *Treponema pallidum* (syphilis). 2016; <http://www.who.int/reproductivehealth/publications/rtis/syphilis-treatment-guidelines/en/>.
2. Peeling RW, Mabey D, Kamb ML, Chen XS, Radolf JD, Benyaken AS. Syphilis. Nat Rev Dis Primers. 2017; 3: 17073. <https://doi.org/10.1038/nrdp.2017.73> PMID: 29022569
3. Šmajš D, Norris SJ, Weinstock GM. Genetic diversity of *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. Infect Genet Evol. 2012; 12(2): 191–202. <https://doi.org/10.1016/j.meegid.2011.12.001> PMID: 22198325
4. Radolf JD, Deka RK, Anand A, Šmajš D, Norgard MV, Yang XF. *Treponema pallidum*, the syphilis spirochete: making a living as a stealth pathogen. Nat Rev Microbiol. 2016; 14(12): 744–759. <https://doi.org/10.1038/nrmicro.2016.141> PMID: 27721440
5. Šmajš D, Strouhal M, Knauf S. Genetics of human and animal uncultivable treponemal pathogens. Infect Genet Evol. 2018; 61: 92–107. <https://doi.org/10.1016/j.meegid.2018.03.015> PMID: 29578082
6. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, et al. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science. 1998; 281(5375): 375–388. PMID: 9665876
7. Matějková P, Strouhal M, Šmajš D, Norris SJ, Palzkill T, Petrosino JF, et al., Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. BMC Microbiol. 2008; 8:76. <https://doi.org/10.1186/1471-2180-8-76> PMID: 18482458
8. Giacani L, Jeffrey BM, Molini BJ, Le HT, Lukehart SA, Centurion-Lara A, et al. Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. J Bacteriol. 2010; 192(10): 2645–2646. <https://doi.org/10.1128/JB.00159-10> PMID: 20348263
9. Pětrošová H, Zabaníková M, Čejková D, Mikalová L, Pospíšilová P, Strouhal M, et al. Whole genome sequence of *Treponema pallidum* ssp. *pallidum*, strain Mexico A, suggests recombination between yaws and syphilis strains. PLoS Negl Trop Dis. 2012; 6(9): e1832. <https://doi.org/10.1371/journal.pntd.0001832> PMID: 23029591
10. Zabaníková M, Mikolka P, Čejková D, Pospíšilová P, Chen L, Strouhal M, et al. Complete genome sequence of *Treponema pallidum* strain DAL-1. Stand Genomic Sci. 2012; 7(1): 12–21. <https://doi.org/10.4056/sigs.2615838> PMID: 23449808
11. Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zabaníková M, Mikalová L, et al. Resequencing of *Treponema pallidum* ssp. *pallidum* strains Nichols and SS14: correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters. PLoS ONE. 2013; 8(9):e74319. <https://doi.org/10.1371/journal.pone.0074319> PMID: 24058545
12. Giacani L, Iverson-Cabral SL, King JCK, Molini BJ, Lukehart SA, Centurion-Lara A. Complete Genome Sequence of the *Treponema pallidum* subsp. *pallidum* Sea81-4 Strain. Genome Announc. 2014; 2(2).
13. Pinto M, Antelo M, Ferreira R, Azevedo J, Santo J, Borrego MJ, et al. A retrospective cross-sectional quantitative molecular approach in biological samples from patients with syphilis. Microb Pathog. 2017; 104:296–302. <https://doi.org/10.1016/j.micpath.2017.01.059> PMID: 28161356
14. Weinstock GM, Šmajš D, Hardham J, Norris SJ. From microbial genome sequence to applications. Res Microbiol. 2000; 151(2):151–8. PMID: 10865961
15. Čejková D, Zabaníková M, Chen L, Pospíšilová P, Strouhal M, Qin X, et al. 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. 2012; 6(1):e1471. <https://doi.org/10.1371/journal.pntd.0001471> PMID: 22292095
16. Zabaníková M, Strouhal M, Mikalová L, Čejková D, Ambrozova L, Pospíšilová P, et al. Whole genome sequence of the *Treponema* Fribourg-Blanc: unspecified simian isolate is highly similar to the yaws sub-species. PLoS Negl Trop Dis. 2013; 7(4):e2172. <https://doi.org/10.1371/journal.pntd.0002172> PMID: 23638193
17. Štaudová B, Strouhal M, Zabaníková M, Čejková D, Fulton LL, Chen L, et al. Whole genome sequence of the *Treponema pallidum* subsp. *endemicum* strain Bosnia A: the genome is related to yaws treponemes but contains few loci similar to syphilis treponemes. PLoS Negl Trop Dis. 2014; 8(11):e3261. <https://doi.org/10.1371/journal.pntd.0003261> PMID: 25375929
18. Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A, Herbig A, et al. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. Nat Microbiol. 2016; 2:16245. <https://doi.org/10.1038/nmicrobiol.2016.245> PMID: 27918528
19. Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. Nat Microbiol. 2016; 2: 16190. <https://doi.org/10.1038/nmicrobiol.2016.190> PMID: 27748767

20. Liu H, Rodes B, Chen CY, Steiner B. New tests for syphilis: rational design of a PCR method for detection of *Treponema pallidum* in clinical specimens using unique regions of the DNA polymerase I gene. *J Clin Microbiol*. 2001; 39(5):1941–1946. <https://doi.org/10.1128/JCM.39.5.1941-1946.2001> PMID: 11326018
21. Grillová L, Bawa T, Mikalová L, Gayet-Ageron A, Nieselt K, Strouhal M, et al. Molecular characterization of *Treponema pallidum* subsp. *pallidum* in Switzerland and France with a new multilocus sequence typing scheme. *PLoS One*. 2018; 13(7): e0200773. <https://doi.org/10.1371/journal.pone.0200773> PMID: 30059541
22. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data; 2014 [cited 2018 June 09]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
23. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17(1): 10.
24. Gordon A. FASTX Toolkit; 2010 [cited 2018 June 09]. [http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html).
25. Li H. Towards better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*. 2014; 30(20): 2843–2851. <https://doi.org/10.1093/bioinformatics/btu356> PMID: 24974202
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16): 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
27. Broad Institute. Picard Toolkit; 2015 [cited 2018 June 09]. <http://broadinstitute.github.io/picard/>.
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
29. Lassmann T, Hayashizaki Y, Daub CO. SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics*. 2011; 27(1): 130–131. <https://doi.org/10.1093/bioinformatics/btq614> PMID: 21088025
30. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics*. 2012; 28(20): 2678–2679. <https://doi.org/10.1093/bioinformatics/bts503> PMID: 22914218
31. Strouhal M, Mikalová L, Havlíčková P, Tenti P, Čejková D, Rychlík I, et al. Complete genome sequences of two strains of *Treponema pallidum* subsp. *pertenue* from Ghana, Africa: Identical genome sequences in samples isolated more than 7 years apart. *PLoS Negl Trop Dis* 2017; 11(9):e0005894. <https://doi.org/10.1371/journal.pntd.0005894> PMID: 28886021
32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
33. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analyses of sequence data. *Bioinformatics*. 2012; 28(12): 1647–9. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
34. Bushnell B. BBMap short read aligner; 2017 [cited 2018 June 09]. <http://sourceforge.net/projects/bbmap>
35. Breese MR, Liu Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*. 2013; 29(4): 494–496. <https://doi.org/10.1093/bioinformatics/bts731> PMID: 23314324
36. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30(12):2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
37. Čejková D, Zbaníková M, Pospíšilová P, Strouhal M, Mikalová L, Weinstock GM, et al. Structure of *rrn* operons in pathogenic non-cultivable treponemes: sequence but not genomic position of intergenic spacers correlates with classification of *Treponema pallidum* and *Treponema paraluis-cuniculi* strains. *J. Med Microbiol*. 2013; 62:196–207. <https://doi.org/10.1099/jmm.0.050658-0> PMID: 23082031
38. Naqvi AA, Shahbaaz M, Ahmad F, Hassan MI. Identification of functional candidates amongst hypothetical proteins of *Treponema pallidum* ssp. *pallidum*. *PLoS One*. 2015; 10(4): e0124177. <https://doi.org/10.1371/journal.pone.0124177> PMID: 25894582
39. Tong ML, Zhao Q, Liu LL, Zhu XZ, Gao K, Zhang HL, et al. Whole genome sequence of the *Treponema pallidum* subsp. *pallidum* strain Amoy: An Asian isolate highly similar to SS14. *PLoS One*. 2017; 12(8): e182768.
40. Sun J, Meng Z, Wu K, Liu B, Zhang S, Liu Y et al. Tracing the origin of *Treponema pallidum* in China using next-generation sequencing. *Oncotarget*. 2016; 7(28): 42904–42918. <https://doi.org/10.18632/oncotarget.10154> PMID: 27344187

41. Strouhal M, Oppelt J, Mikalová L, Arora N, Nieselt K, González-Candelas F, et al. Reanalysis of Chinese *Treponema pallidum* samples: all Chinese samples cluster with SS14-like group of syphilis-causing treponemes. *BMC Res Notes*. 2016; 11(1):16.
42. Cumberland MC, Turner TB. The rate of multiplication of *Treponema pallidum* in normal and immune rabbits. *Am J Syph Gonorrhoea Vener Dis*. 1949; 33(3): 201–202. PMID: [18121293](https://pubmed.ncbi.nlm.nih.gov/18121293/)
43. Magnuson HJ, Eagle H, Fleischman R. The minimal infectious inoculum of *Spirochaeta pallida* (Nichols strain) and a consideration of its rate of multiplication *in vivo*. *Am J Syph Gonorrhoea Vener Dis*. 1948; 32(1): 1–18. PMID: [18917621](https://pubmed.ncbi.nlm.nih.gov/18917621/)
44. Nechvátal L, Pětrošová H, Grillová L, Pospíšilová P, Mikalová L, Strnadel R, et al. Syphilis-causing strains belong to separate SS14-like or Nichols-like groups as defined by multilocus analysis of 19 *Treponema pallidum* strains. *Int J Med Microbiol*. 2014; 304(5–6):645–653. <https://doi.org/10.1016/j.ijmm.2014.04.007> PMID: [24841252](https://pubmed.ncbi.nlm.nih.gov/24841252/)
45. Šmajš D, Mikalová L, Strouhal M, Grillová L. Why Are There Two Genetically Distinct Syphilis-Causing Strains? *For Immunopathol Dis Therap*. 2016;
46. Grillová L, Pětrošová H, Mikalová L, Strnadel R, Dastychová E, Kuklová I, et al. Molecular typing of *Treponema pallidum* in the Czech Republic during 2011 to 2013: increased prevalence of identified genotypes and of isolates with macrolide resistance. *J Clin Microbiol*. 2014; 52: 3693–3700. <https://doi.org/10.1128/JCM.01292-14> PMID: [25100820](https://pubmed.ncbi.nlm.nih.gov/25100820/)
47. Marra CM, Sahi SK, Tantaló LC, Godornes C, Reid T, Behets F, et al. Enhanced Molecular Typing of *Treponema pallidum*: Geographical Distribution of Strain Types and Association with Neurosyphilis. *J Infect Dis*. 2010; 202: 1380–1388. <https://doi.org/10.1086/656533> PMID: [20868271](https://pubmed.ncbi.nlm.nih.gov/20868271/)
48. Read P, Tagg KA, Jeoffreys N, Guy RJ, Gilbert GL, Donovan B. *Treponema pallidum* Strain Types and Association with Macrolide Resistance in Sydney, Australia: New TP0548 Gene Types Identified. *J Clin Microbiol*. 2016; 54(8): 2172–2174. <https://doi.org/10.1128/JCM.00959-16> PMID: [27194693](https://pubmed.ncbi.nlm.nih.gov/27194693/)
49. Giacani L, Ciccarese G, Puga-Salazar C, Dal Conte I, Colli L, Cusini M, et al. Enhanced Molecular Typing of *Treponema pallidum* subsp. *pallidum* strains from Italian hospitals shows geographical differences in strain type heterogeneity, widespread resistance to macrolides, and lack of mutations associated with doxycycline resistance. *Sex Transm Dis*. 2017;
50. Salado-Rasmussen K, Cowan S, Gerstoft J, Larsen HK, Hoffmann S, Knudsen TB, et al. Molecular Typing of *Treponema pallidum* in Denmark: A Nationwide Study of Syphilis. *Acta Derm Venereol*. 2016; 96(2):202–206. <https://doi.org/10.2340/00015555-2190> PMID: [26122912](https://pubmed.ncbi.nlm.nih.gov/26122912/)
51. Grange PA, Allix-Beguec C, Chanal J, Benhaddou N, Gerhardt P, Morini JP, et al. Molecular subtyping of *Treponema pallidum* in Paris, France. *Sex Transm Dis*. 2013; 40(8):641–644. <https://doi.org/10.1097/OLQ.000000000000006> PMID: [23859911](https://pubmed.ncbi.nlm.nih.gov/23859911/)