



A public database for the new MLST scheme for *Treponema pallidum* subsp. *pallidum*: surveillance and epidemiology of the causative agent of syphilis.

Linda Grillova, Keith A. Jolley, David Šmajš, Mathieu Picardeau

► To cite this version:

Linda Grillova, Keith A. Jolley, David Šmajš, Mathieu Picardeau. A public database for the new MLST scheme for *Treponema pallidum* subsp. *pallidum*: surveillance and epidemiology of the causative agent of syphilis.. PeerJ, 2019, 6, pp.e6182. 10.7717/peerj.6182 . pasteur-02090367

HAL Id: pasteur-02090367

<https://pasteur.hal.science/pasteur-02090367>

Submitted on 4 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A public database for the new MLST scheme for *Treponema pallidum* subsp. *pallidum*: surveillance and epidemiology of the causative agent of syphilis

Linda Grillova¹, Keith Jolley², David Šmajs³ and Mathieu Picardeau¹

¹ Biology of Spirochetes Unit, Institut Pasteur, Paris, France

² Department of Zoology, University of Oxford, Oxford, UK

³ Department of Biology, Masaryk University, Brno, Czech Republic

ABSTRACT

Treponema pallidum subsp. *pallidum* is the causative agent of syphilis, a sexually transmitted disease with worldwide prevalence. Several different molecular typing schemes are currently available for this pathogen. To enable population biology studies of the syphilis agent and for epidemiological surveillance at the global scale, a harmonized typing tool needs to be introduced. Recently, we published a new multi-locus sequence typing (MLST) with the potential to significantly enhance the epidemiological data in several aspects (e.g., distinguishing genetically different clades of syphilis, subtyping inside these clades, and finally, distinguishing different subspecies of non-cultivable pathogenic treponemes). In this short report, we introduce the PubMLST database for treponemal DNA data storage and for assignments of allelic profiles and sequencing types. Moreover, we have summarized epidemiological data of all treponemal strains ($n = 358$) with available DNA sequences in typing loci and found several association between genetic groups and characteristics of patients. This study proposes the establishment of a single MLST of *T. p. pallidum* and encourages researchers and public health communities to use this PubMLST database as a universal tool for molecular typing studies of the syphilis pathogen.

Subjects Microbiology, Epidemiology, Infectious Diseases

Keywords *Treponema pallidum* subsp. *pallidum*, Molecular typing, PubMLST

INTRODUCTION

Syphilis is a bacterial infection caused by *Treponema pallidum* subsp. *pallidum*. Despite the availability of effective treatments, the annual worldwide incidence recently reached more than 5.6 million cases (*World Health Organization, 2016*).

The proper discrimination of strains of infectious pathogens is crucial for epidemiological and surveillance analyses, for description of population structure and dynamics and for the improvement of public health control strategies. More specifically, the association of specific strains of *T. p. pallidum* with different groups of patients can result in a better understanding of syphilis epidemiology. Association of macrolide resistance/sensitivity and allelic profiles are particularly important for a better

Submitted 10 October 2018
Accepted 29 November 2018
Published 9 January 2019

Corresponding author
Linda Grillova,
linda.grillova@pasteur.fr

Academic editor
Elliot Lefkowitz

Additional Information and
Declarations can be found on
page 9

DOI 10.7717/peerj.6182

© Copyright
2019 Grillova et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

understanding of the emergence of antibiotic-resistant strains. Finally, the molecular typing of syphilis-causing strains can help to determine syphilis diagnosis, especially in cases with atypical symptoms, thereby preventing syphilis-bejel confusion (Grange et al., 2016; Noda et al., 2018).

Multi-locus sequence typing (MLST) is typically based on amplification and sequencing of about seven housekeeping gene fragments (loci). For each locus, the different sequence variants are assigned as distinct alleles and, for each isolate, the combination of alleles defines the allelic profile and sequence type (ST). Allelic profiles and/or STs are used for definition of the strains and these can be clustered into clonal complexes based on shared alleles (Maiden et al., 2013). The traditional MLST approach is based on allele-based comparisons, where each allelic change is counted as a single event, regardless of the number of nucleotide changes involved. This helps to mitigate the effects of horizontal genetic transfer in organisms prone to recombination. The MLST data are commonly stored in publicly available and curated databases, such as PubMLST, which uses the Bacterial Isolate Genome Sequence Database (BIGSdb) platform (Jolley & Maiden, 2010). These databases can be used to automatically identify and assign new allele variants and STs, and provide tools to analyze the origin and phenotypic metadata of samples with the genotypic information of the typed pathogens. Several functions are available within this software to analyze and visualize epidemiological data (Jolley & Maiden, 2010; Maiden et al., 2013). BIGSdb is currently hosting over 100 different bacterial species and is one of the most widely used tools for epidemiological purposes (Jolley, Bray & Maiden, 2018).

Recently, we published a new MLST scheme of *T. p. pallidum* (Grillová et al., 2018a; Pospíšilová et al., 2018). The *T. p. pallidum* MLST scheme is somewhat unusual compared to typical MLST schemes reflecting the special characteristics of *T. p. pallidum*. The *T. p. pallidum* MLST uses three highly variable loci and performs sequence-based analyses, that is, it takes into account sequence differences among alleles.

- i) Even though some recent attempts to cultivate *T. p. pallidum* were successful using *T. p. pallidum* strains passed through cultures with rabbit epithelial cells (Edmondson, Hu & Norris, 2018), the in vitro cultivation of *T. p. pallidum* strains isolated directly from patients is still not possible. It is, therefore, necessary to perform MLST directly from clinical samples, which usually contain a very low number of *T. p. pallidum* DNA copies ($10-10^2$ *T. p. pallidum* DNA copies/ μ l) (Pinto et al., 2017). To ensure a high proportion of fully typed samples, the newly proposed MLST is based on three loci only, in comparison to the conventional seven-locus scheme used in other bacteria.
- ii) Multi-locus sequence typing schemes are usually based on several housekeeping genes. However, this approach is not suitable for monomorphic bacteria as *T. p. pallidum* (Šmajs, Norris & Weinstock, 2012; Radolf et al., 2016; Šmajs, Strouhal & Knauf, 2018). In this case, every single nucleotide variant (SNV) is important and valuable and the typing loci with the highest SNV density should be used to ensure sufficient variability to map the overall population diversity. However, the typing loci should still be stable over time. Therefore, the new MLST was designed on the principle of the

- sequence analyses using three highly variable loci—TP0136, TP0548 and TP0705 (Grillová *et al.*, 2018a). These three loci represented the lowest number of loci with the highest discrimination power possible among the candidate loci selected for typing.
- iii) Because of the monomorphic character of *T. p. pallidum* genome and the low number of typing loci, the “sequence-based analyses” approach is highly recommended. This approach, compared to the conventional “allele-based comparisons,” allows the researcher to distinguish alleles based on a number of nucleotide differences and permits interpretation of the data using maximum-likelihood phylogeny. Even though the *T. p. pallidum* MLST uses only three typing loci, the scheme has a high resolution (more than 30% of the resolution achieved with genome-wide data). This approach also allows distinction of *T. p. pallidum* from other treponemal subspecies and species, discriminates between the two *T. p. pallidum* clades (TP0136, TP0548) (Nechvátal *et al.*, 2014), and can be used for differentiation of strains within each of these clades (TP0136 for the Nichols clade, and TP0705 for the SS14 clade).

The purpose of this short report is to establish a single universal MLST scheme applicable to *T. p. pallidum* (Grillová *et al.*, 2018a) and introduce the PubMLST database for treponemal DNA data storage and ST assignments.

MATERIALS AND METHODS

Treponema pallidum subsp. *pallidum* database hosted at PubMLST

We have created a *T. p. pallidum* database hosted at PubMLST. PubMLST uses the BIGSdb platform for the storage and analysis of sequence data from bacterial isolates. The input format can be fasta files, which contain small contigs assembled from dideoxy chain termination sequencing of the typing loci, or complete or multiple contigs generated from next-generation sequencing platforms. When the new sequences are submitted to BIGSdb, the BLAST algorithm incorporated into BIGSdb identifies already known sequence variants or marks an unknown variant for curator verification. After verification by a curator, a novel allele number is assigned (Maiden *et al.*, 2013). Metadata of the patients may be submitted to the database as well, and includes some important characteristics of isolates, for example, isolate identification name/number; country where the strain was isolated; biological source of sample; year of isolation; and resistance to macrolide antibiotics. However, we encourage the researchers to upload as much information about patients and samples as available (e.g., age, sex and sexual orientation of the patients, stage of the disease, serological results etc.). The database can be accessed at <https://pubmlst.org/tpallidum/>. The database is primarily designed to be used as a tool for automatic ST assignments; for determination of new alleles; for storage of sample metadata; for identification of new associations between genetic types and metadata using various tools; and for user-friendly visualization of molecular typing data, particularly using GrapeTree (Zhou *et al.*, 2018) and Microreact (Argimón *et al.*, 2016) plugins. These plugins enable, for example, visualization of genomic epidemiology and phylogeography by showing the correlation of the STs and geographical areas in real time and visualization of phylogenetic data together with patient metadata.

However, this database can also serve as a repository for complete/draft genomes and gene-based comparative genomic data (using the Genome comparator tool). Sequences and metadata can be exported from the database in multiple formats. The database is overseen by the curators, who will check the submitted data and who will be available for any additional help (see <https://pubmlst.org/tpallidum/>).

Sequences submitted to PubMLST

All published whole genome and draft genome sequences of *T. p. pallidum* (Pětrošová et al., 2012, 2013; Giacani et al., 2010, 2014; Zbaníková et al., 2012; Tong et al., 2017; Arora et al., 2016; Sun et al., 2016; Pinto et al., 2016; Strouhal et al., 2018; Grillová et al., 2018b), data available at GenBank, as well as sequences obtained by MLST (Grillová et al., 2018a; Pospíšilová et al., 2018) were submitted to the database.

Phylogenetic analyses

The phylogenetic tree was generated using MEGA6 (Tamura et al., 2013) with the maximum likelihood bootstrap algorithm and Tamura Nei model based on concatenated sequences of typing loci of fully typed samples ($n = 286$). The tree was visualized using iTOL (Letunic & Bork, 2016), available as an external plugin of the BIGSdb.

Statistical analyses

Statistical analyses were performed using STATISTICA software v.12 (StatSoft, Tulsa, OK, USA) using data extracted from the two-field breakdown option. A Fisher's exact test was used to establish the correlations of patient's metadata and allelic variants and STs. Statistical significance was set at $p < 0.05$.

RESULTS

Treponema pallidum subsp. *pallidum* database

At the time of writing (September 2018), the databases contain 358 *T. p. pallidum* strains, mostly isolated from Europe (70.87%), followed by North America (26.33%), Asia (2.52%) and South America (0.28%) (Table 1). More than 90% of samples were clinically acquired and 31 strains were propagated in rabbits. A total of 71 samples represented whole/draft genomes and 287 samples were typed using MLST and yielded sequences only in typing loci. The samples were collected from 1912 to 2017. There were 272 samples collected from males (183 samples were collected from Men who have sex with men; MSM) and only four samples from females. Most of the samples were isolated from patients diagnosed with primary syphilis, followed by secondary syphilis, patients in the border of primary and secondary stage, and there were also three cases of congenital syphilis. The patients were between 0 and 71 years old. In addition, 21 samples were found negative by serology, but positive by PCR and all of them were isolated from patients with primary syphilis by genital, anal or throat swab (Table 1). Most of the samples (262, 73.39%) contained treponemes resistant to macrolide antibiotics caused by A2058G mutations (in 257 cases) and A2059G mutations (in five cases) in both treponemal 23S rRNA genes.

Fully typed *T. p. pallidum* strains ($n = 285$) were divided into 40 allelic profiles or strain types (ST1-ST40), and further into two clonal complexes, where 31 STs belong to the

Table 1 Clinical characteristics of the *T. p. pallidum* strains submitted to the BIGSdb.

Continent (n, %)		Clinical source (n, %)		Stage (n, %)	
Europe	253 (70.87)	Genital swab	180 (50.42)	Primary	119 (33.33)
North America	94 (26.33)	Anal swab	56 (15.69)	Primary/secondary	8 (2.24)
Asia	9 (2.52)	Throat swab	39 (10.92)	Secondary	40 (11.2)
South America	1 (0.28)	Skin lesion	17 (4.76)	Congenital	3 (0.84)
Country (n, %)		Blood	3 (0.84)	Unspecified	187 (52.38)
France	146 (40.90)	CSF	2 (0.56)	Macrolide resistance (n, %)	
Cuba	72 (20.17)	Amniotic fluid	1 (0.28)	Resistant	262 (73.39)
Switzerland	72 (20.17)	Tissue	1 (0.28)	Sensitive	62 (18.3)
Portugal	25 (7)	Other	1 (0.28)	Not done	33 (9.24)
USA	21 (5.88)	Unspecified	57 (15.97)	Mutations (n, %)	
China	9 (2.52)	Sex (n, %)		A2058G	257 (71.99)
Netherlands	5 (1.4)	Male	272 (76.19)	A2059G	5 (1.4)
Austria	3 (0.84)	Female	8 (2.24)	Unspecified	95 (26.61)
Czech Republic	2 (0.56)	Unspecified	77 (21.5)	Serology (n, %)	
Argentina	1 (0.280)	Sexual orientation (n, %)		Positive	182 (50.98)
Mexico	1 (0.28)	MSM	183 (51.26)	Negative	21 (5.88)
Source (n, %)		MSW	27 (7.56)	Unspecified	154 (43.14)
Clinically acquired	326 (91.32)	WSM	3 (0.84)		
Experimental animals	31 (8.68)	Unspecified	144 (40.34)		

Note:

CSF, cerebrospinal fluid; MSM, men who have sex with men; MSW, men who have sex with woman; WSM, woman who have sex with men.

clonal complex “SS14-like” and nine STs to the clonal complex “Nichols-like” (Fig. 1; Table 2). STs were divided based on the 137 variable sites in total present in the concatenated sequences of typing loci (2,584 bp), where 46 variable positions were found to be parsimony informative for distinction of the two clonal complexes. The majority of the samples were found to belong to the SS14-like clonal complex (92.3%), while only 7.7% of samples belonged to the Nichols-like clonal complex.

Association found using *T. p. pallidum* BIGSdb

We have found several association between genetic data and metadata of the samples including patient characteristics. Nichols-like strains were found to be associated with susceptibility to macrolides, and, on the other hand, SS14-like strains were associated with the presence of mutations leading to macrolide resistance ($p < 0.0001$) (Fig. 2A). The data showed that Nichols-like strains appear to be spreading predominantly among MSM patients (Fig. 2B). Moreover, Nichols-like strains were associated with an older population (35 years old and older, $p = 0.0104$) (Fig. 2C). Most of these association were already described previously (Grillová et al., 2014; Gallo Vaulet et al., 2017; Read et al., 2016). However, in the future, there is a need to verify these association by examination of a higher number of treponemal strains from different geographical areas and from different groups of patients to avoid sampling biases.

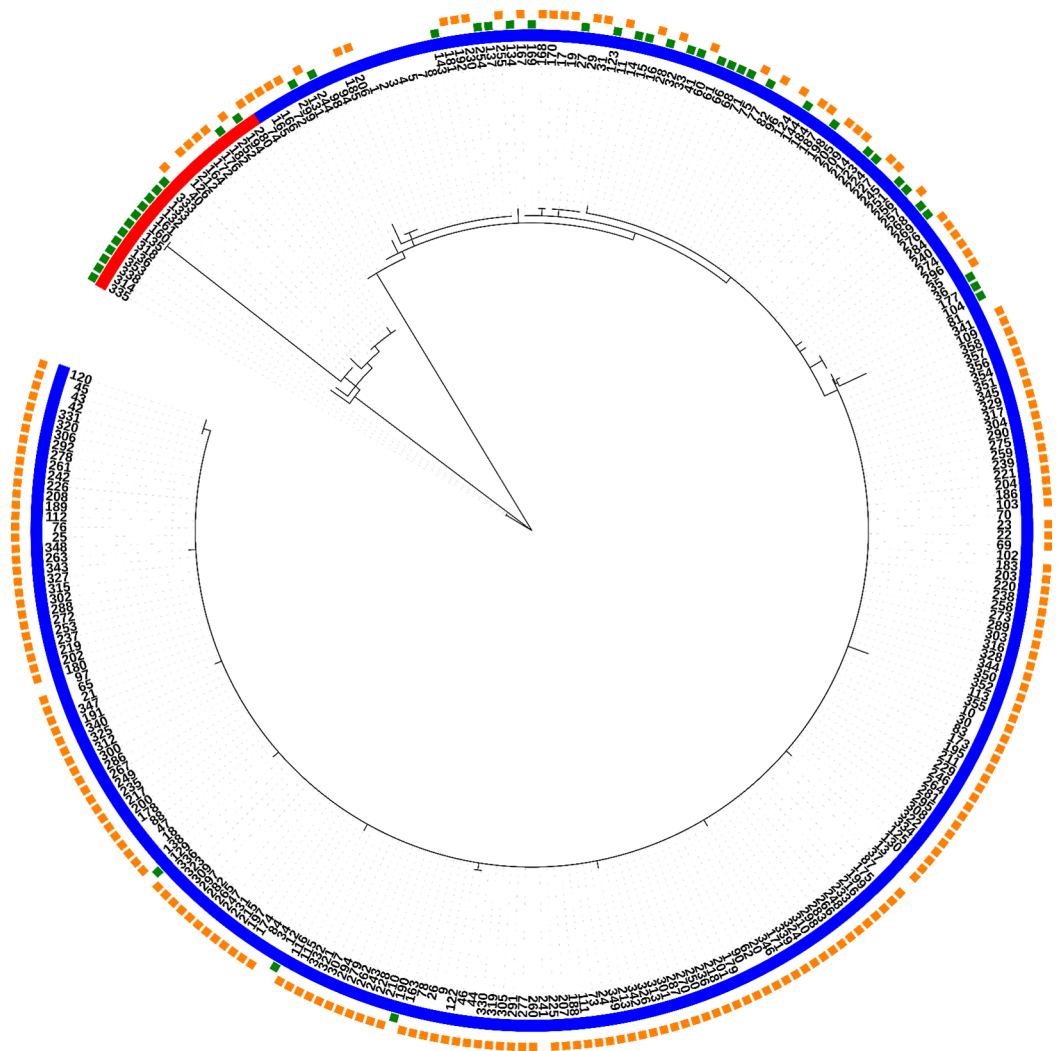


Figure 1 Phylogeny of all fully typed samples ($n = 286$) divided into 40 STs and two clonal complexes (Nichols-like and SS14-like), based on concatenated sequences of typing loci. Blue color represents the SS14-like clade and red color represents the Nichols-like clade. The green squares mark the macrolide sensitive strains and the orange squares mark the resistant strains.

Full-size DOI: 10.7717/peerj.6182/fig-1

ST1 was found to be the predominant sequencing type (allelic profile: 1.3.1, clonal complex: SS14-like). However, with the increasing number of newly submitted *T. p. pallidum* strains, the number of different STs also increased, suggesting that the diversity of the *T. p. pallidum* strains is still not fully revealed. Interestingly, five STs were found exclusively among the strains propagated in rabbits—ST20–ST24 ($n = 8$) and not among the strains isolated directly from patients. However, the remaining examined strains propagated in rabbits ($n = 23$) shared the same STs with the directly sequenced strains.

DISCUSSION

Since *T. p. pallidum* cannot be routinely cultivated in vitro, MLST needs to be performed directly from clinical samples. To obtain fully typed samples enabling assignment of ST,

Table 2 Sequencing types found among 286 *T. p. pallidum* fully typed strains submitted to the BIGSdb.

ST	TP0136	TP0548	TP0705	Clonal complex	No. of samples
1	1	3	1	SS14-like	169
2	1	1	1	SS14-like	37
3	1	1	8	SS14-like	13
4	7	1	9	SS14-like	5
5	2	1	2	SS14-like	3
6	3	2	3	Nichols-like	3
7	1	4	1	SS14-like	3
8	1	5	1	SS14-like	2
9	4	3	1	SS14-like	1
10	1	11	8	SS14-like	3
11	1	1	3	SS14-like	4
12	5	3	8	SS14-like	1
13	6	3	1	SS14-like	1
14	1	8	1	SS14-like	1
15	1	3	5	SS14-like	1
16	1	3	7	SS14-like	1
17	13	1	1	SS14-like	1
18	1	9	1	SS14-like	1
19	1	1	10	SS14-like	2
20	10	14	10	Nichols-like	3
21	1	13	10	SS14-like	1
22	11	14	10	Nichols-like	2
23	12	15	10	Nichols-like	1
24	6	1	8	SS14-like	1
25	1	26	1	SS14-like	1
26	9	7	3	Nichols-like	6
27	1	1	9	SS14-like	2
28	1	17	9	SS14-like	2
29	1	18	1	SS14-like	1
30	1	19	1	SS14-like	1
31	9	20	3	Nichols-like	1
32	14	3	1	SS14-like	1
33	1	1	11	SS14-like	1
34	1	22	12	SS14-like	1
35	1	23	1	SS14-like	1
36	1	1	13	SS14-like	1
37	15	7	3	Nichols-like	4
38	9	24	8	Nichols-like	1
39	9	25	3	Nichols-like	1
40	16	3	1	SS14-like	1

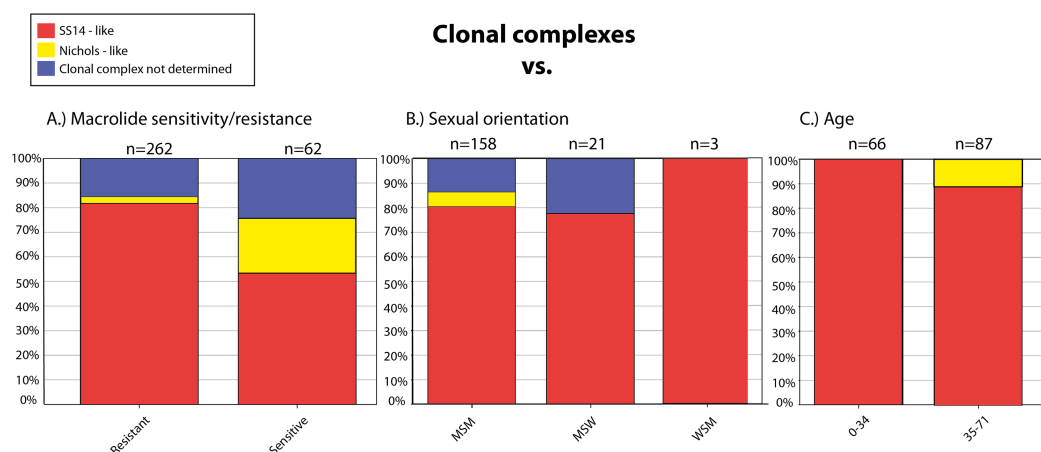


Figure 2 Clonal complexes associated with the macrolide sensitivity/resistance (A), sexual orientation (B), and age of the patients (C). The figure was created based on the BIGSdb output.
 Full-size DOI: 10.7717/peerj.6182/fig-2

the amplification efficiency of all three loci needs to be as high as possible. The amplification efficiency depends on several factors, including the type of material taken from patients, the time between sampling and DNA isolation, the type of DNA isolation protocol as well as the type of amplification protocol.

As reported in many studies, the highest concentration of the treponemal DNA is present in swabs taken from the primary chancre, followed by blood-derived samples, cerebrospinal fluid and others (Peng et al., 2011; Pinto et al., 2017). For example, the swabs taken from primary chancre represent the samples with the highest concentration of treponemal DNA and are the most suitable candidates for MLST. However, MLST profiles revealed from parallel samples taken from the same patient (e.g., primary chancre swab and whole blood) are also important with respect to the stability of the used typing loci. For instance, the typing loci used in the enhanced CDC-typing (based on restriction fragment length polymorphism analyses of the *tpoEGJ* genes and a number of repeats in the *arp* gene (Marra et al., 2010) were identified to be genetically unstable (Mikalová et al., 2013). This genetic instability was shown by revealing the different subtypes in samples isolated from swab and whole blood, both taken from the same patient. Loci selected for MLST (TP0136, TP0548 and TP0705) are considered to be genetically stable, showing the same allelic profiles in parallel samples as well as in samples collected from different rabbit passages (Grillová et al., 2018a).

Another important factor to obtain high-quality DNA is the time between sampling of patients and the DNA isolation. This time period should be as short as possible. If immediate DNA isolation is not possible, the swab extracts (PBS) should be stored at -20°C . The swab extracts should be centrifuged at the lowest possible speed for 10 min and the supernatant should be used for the subsequent isolation of DNA. Even though it is time consuming, nested PCR should be used for target loci amplification to achieve the highest amplification efficiency. Moreover, we recommend looking for the presence or absence of A2058G and A2059G mutations in 23S rDNA, which encodes the macrolide resistance (Grillová et al., 2018a).

Most of *T. p. pallidum* strains analyzed to date from experimentally infected rabbits belong to the Nichols-like group, while most clinical *T. p. pallidum* isolates obtained directly from patients belong to the SS14-like group (Šmajs *et al.*, 2016). We encourage researchers and public health communities to use MLST, which is able to distinguish these two clonal complexes. Moreover, we recommend submitting the sequencing data to PubMLST along with the metadata (including data from the serological tests). The expansion of the sample set in the PubMLST with syphilis isolates from different geographical areas can lead to a better understanding of the epidemiology of the two genetically distinct groups of syphilis treponemes.

CONCLUSIONS

Here, we present the PubMLST BIGSdb database of *T. p. pallidum*, which represents the first publicly available data storage of *T. p. pallidum* sequences connected to metadata of patients. The database identifies specific regions of uploaded sequencing data and their genetic variation in order to reveal the sequencing types of *T. p. pallidum* isolates based on the newly developed MLST (Grillová *et al.*, 2018a). We hope that this tool will open new opportunities in epidemiology of syphilis thereby allowing studies of sequencing types in different locations, tracking of syphilis infections and finding the association of particular strains with specific groups of patients. All these data should improve our understanding of syphilis epidemiology.

ACKNOWLEDGEMENT

We thank to Dr. Robert Anthony Gaultney for English editing.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Linda Grillova was funded by a PTR grant (PTR30-17) from the Institut Pasteur. This work was supported by the Grant Agency of the Ministry of Health of the Czech Republic (17-31333A) to David Šmajs. The development of PubMLST and BIGSdb has been supported by a Wellcome Trust Biomedical Resource Grant (104992). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Institut Pasteur: PTR grant: PTR30-17.

Grant Agency of the Ministry of Health of the Czech Republic: 17-31333A.

Wellcome Trust Biomedical Resource Grant: 104992.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Linda Grillova conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Keith Jolley conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- David Šmajš authored or reviewed drafts of the paper, approved the final draft.
- Mathieu Picardeau authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data are available at PubMLST: <https://pubmlst.org/tpallidum/>

REFERENCES

- Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial Genomics* 2:e000093 DOI 10.1099/mgen.0.000093.
- Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A, Herbig A, Strouhal M, Grillová L, Sánchez-Busó L, Kühnert D, Bos KI, Davis LR, Mikalová L, Bruisten S, Komericki P, French P, Grant PR, Pando MA, Valet LG, Fermepin MR, Martinez A, Centurion Lara A, Giacani L, Norris SJ, Šmajš D, Bosshard PP, González-Candelas F, Nieselt K, Krause J, Bagheri HC. 2016. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nature Microbiology* 2:16245 DOI 10.1038/nmicrobiol.2016.245.
- Edmondson DG, Hu B, Norris SJ. 2018. Long-term in vitro culture of the syphilis spirochete *Treponema pallidum* subsp. *pallidum*. *American Society for Microbiology* 9:e011153-18 DOI 10.1128/mBio.01153-18.
- Gallo Valet L, Grillová L, Mikalová L, Casco R, Rodríguez Fermepin M, Pando MA, Šmajš D. 2017. Molecular typing of *Treponema pallidum* isolates from Buenos Aires, Argentina: frequent Nichols-like isolates and low levels of macrolide resistance. *PLOS ONE* 12(2):e0172905 DOI 10.1371/journal.pone.0172905.
- Giacani L, Iverson-Cabral SL, King JCK, Molini BJ, Lukehart SA, Centurion-Lara A. 2014. Complete genome sequence of the *Treponema pallidum* subsp. *pallidum* Sea81-4 Strain. *Genome Announcements* 2(2):e00333-14 DOI 10.1128/genomeA.00333-14.
- Giacani L, Jeffrey BM, Molini BJ, Le HT, Lukehart SA, Centurion-Lara A, Rockey DD. 2010. Complete genome sequence and annotation of the *Treponema pallidum* subsp. *pallidum* Chicago strain. *Journal of Bacteriology* 192:2645–2646 DOI 10.1128/JB.00159-10.
- Grange PA, Mikalová L, Gaudin C, Strouhal M, Janier M, Benhaddou N, Šmajš D, Dupin N. 2016. *Treponema pallidum* 11qj subtype may correspond to a *Treponema pallidum* subsp. *Endemicum* strain. *Sexually Transmitted Diseases* 43(8):517–518 DOI 10.1097/OLQ.0000000000000474.
- Grillová L, Bawa T, Mikalová L, Gayet-Ageron A, Nieselt K, Strouhal M, Sednaoui P, Ferry T, Cavassini M, Lautenschlager S, Dutly F, Pla-Díaz M, Krützen M, González-Candelas F, Bagheri HC, Šmajš D, Arora N, Bosshard PP. 2018a. Molecular characterization of

- Treponema pallidum* subsp. *pallidum* in Switzerland and France with a new multilocus sequence typing scheme. *PLOS ONE* 13(7):e0200773 DOI 10.1371/journal.pone.0200773.
- Grillová L, Giacani L, Mikalová L, Strouhal M, Strnadel R, Marra C, Centurion-Lara A, Poveda L, Russo G, Čejková D, Vašků V, Oppelt J, Šmajš D. 2018b. 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* 3(8):e0202619 DOI 10.1371/journal.pone.0202619.
- Grillová L, Pětrošová H, Mikalová L, Strnadel R, Dastyčová E, Kuklová I, Kojanová M, Kreidlová M, Vaňousová D, Hercogová J, Procházka P, Zákoucká H, Krchňáková A, Vašků V, Šmajš D. 2014. Molecular typing of *Treponema pallidum* subsp. *pallidum* in the Czech Republic during 2011 to 2013: increased prevalence of identified genotypes and of isolates with macrolide resistance. *Journal of Clinical Microbiology* 52(10):3693–3700 DOI 10.1128/jcm.01292-14.
- Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: awaiting peer review]. *Wellcome Open Research* 3:124 DOI 10.12688/wellcomeopenres.14826.1.
- Jolley KA, Maiden MCJ. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595 DOI 10.1186/1471-2105-11-595.
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44(W1):W242–W245 DOI 10.1093/nar/gkw290.
- Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* 11:728–736 DOI 10.1038/nrmicro3093.
- Marra CM, Sahi SK, Tantaló LC, Godornes C, Reid T, Behets F, Rompalo A, Klausner JD, Yin YP, Mulcahy F, Golden MR, Centurion-Lara A, Lukehart SA. 2010. Enhanced molecular typing of *Treponema pallidum*: geographical distribution of strain types and association with neurosyphilis. *Journal of Infectious Diseases* 202(9):1380–1388 DOI 10.1086/656533.
- Mikalová L, Pospíšilová P, Woznicová V, Kuklová I, Zákoucká H, Šmajš D. 2013. Comparison of CDC and sequence-based molecular typing of syphilis treponemes: tpr and arp loci are variable in multiple samples from the same patient. *BMC Microbiology* 13:178 DOI 10.1186/1471-2180-13-178.
- Nechvátal L, Pětrošová H, Grillová L, Pospíšilová P, Mikalová L, Strnadel R, Kuklová I, Kojanová M, Kreidlová M, Vaňousová D, Procházka P, Zákoucká H, Krchňáková A, Šmajš D. 2014. Syphilis-causing strains belong to separate SS14-like or Nichols-like groups as defined by multilocus analysis of 19 *Treponema pallidum* strains. *International Journal of Medical Microbiology* 304(5–6):645–653 DOI 10.1016/j.ijmm.2014.04.007.
- Noda AA, Grillová L, Lienhard R, Blanco O, Rodríguez I, Šmajš D. 2018. Bejel in Cuba: molecular identification of *Treponema pallidum* subsp. *endemicum* in patients diagnosed with venereal syphilis. *Clinical Microbiology and Infection* 24(11):1210.e1–1210.e5 DOI 10.1016/j.cmi.2018.02.006.
- Peng R-R, Wang AL, Li J, Tucker JD, Yin Y-P, Chen X-S. 2011. Molecular typing of *Treponema pallidum*: a systematic review and meta-analysis. *PLOS Neglected Tropical Diseases* 5(11):e1273 DOI 10.1371/journal.pntd.0001273.
- Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zabaníková M, Mikalová L, Sodergren E, Weinstock GM, Šmajš D. 2013. Resequencing of *Treponema pallidum* ssp. *pallidum* strains

- Nichols and SS14: correction of sequencing errors resulted in increased separation of syphilis treponeme subclusters. *PLOS ONE* 8(9):e74319 DOI 10.1371/journal.pone.0074319.
- Pětrošová H, Zbaníková M, Čejková D, Mikalová L, Pospíšilová P, Strouhal M, Chen L, Qin X, Muzny DM, Weinstock GM, Šmajs D. 2012. Whole genome sequence of *Treponema pallidum* ssp. *pallidum*, strain Mexico A, suggests recombination between yaws and syphilis strains. *PLOS Neglected Tropical Diseases* 6(9):e1832 DOI 10.1371/journal.pntd.0001832.
- Pinto M, Antelo M, Ferreira R, Azevedo J, Santo I, Borrego MJ, Gomes JP. 2017. A retrospective cross-sectional quantitative molecular approach in biological samples from patients with syphilis. *Microbial Pathogenesis* 104:296–302 DOI 10.1016/j.micpath.2017.01.059.
- Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, Borrego MJ, Mendonça J, Carpinteiro D, Vieira L, Gomes JP. 2016. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nature Microbiology* 2(1):16190 DOI 10.1038/nmicrobiol.2016.190.
- Pospíšilová P, Grange PA, Grillová L, Mikalová L, Martinet P, Janier M, Vermersch A, Benhaddou N, Del Giudice P, Alcaraz I, Truchetet F, Dupin N, Šmajs D. 2018. Multi-locus sequence typing of *Treponema pallidum* subsp. *pallidum* present in clinical samples from France: infecting treponemes are genetically diverse and belong to 18 allelic profiles. *PLOS ONE* 13(7):e0201068 DOI 10.1371/journal.pone.0201068.
- Radolf JD, Deka RK, Anand A, Šmajs D, Norgard MV, Yang XF. 2016. *Treponema pallidum*, the syphilis spirochete: making a living as a stealth pathogen. *Nature Reviews Microbiology* 14(12):744–759 DOI 10.1038/nrmicro.2016.141.
- Read P, Tagg KA, Jeoffreys N, Guy RJ, Gilbert GL, Donovan B. 2016. *Treponema pallidum* strain types and association with macrolide resistance in Sydney, Australia: New TP0548 gene types identified. *Journal of Clinical Microbiology* 54(8):2172–2174 DOI 10.1128/JCM.00959-16.
- Šmajs D, Mikalova L, Strouhal M, Grillova L. 2016. Why are there two genetically distinct syphilis-causing strains? *Forum on Immunopathological Diseases and Therapeutics* 7:181–190 DOI 10.1615/ForumImmunDisTher.2017020184.
- Šmajs D, Norris SJ, Weinstock GM. 2012. Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. *Infection, Genetics and Evolution* 12(2):191–202 DOI 10.1016/j.meegid.2011.12.001.
- Šmajs D, Strouhal M, Knauf S. 2018. Genetics of human and animal uncultivable treponemal pathogens. *Infection, Genetics and Evolution* 61:92–107 DOI 10.1016/j.meegid.2018.03.015.
- Strouhal M, Oppelt J, Mikalová L, Arora N, Nieselt K, González-Candelas F, Šmajs D. 2018. Reanalysis of Chinese *Treponema pallidum* samples: all Chinese samples cluster with SS14-like group of syphilis-causing treponemes. *BMC Research Notes* 11:16 DOI 10.1186/s13104-017-3106-7.
- Sun J, Meng Z, Wu K, Liu B, Zhang S, Liu Y, Wang Y, Zheng H, Huang J, Zhou P. 2016. Tracing the origin of *Treponema pallidum* in China using next-generation sequencing. *Oncotarget* 7:42904–42918 DOI 10.18632/oncotarget.10154.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12):2725–2729 DOI 10.1093/molbev/mst197.
- Tong M-L, Zhao Q, Liu L-L, Zhu X-Z, Gao K, Zhang H-L, Lin L-R, Niu J-J, Ji Z-L, Yang T-C. 2017. Whole genome sequence of the *Treponema pallidum* subsp. *pallidum* strain Amoy: an Asian isolate highly similar to SS14. *PLOS ONE* 12:e0182768 DOI 10.1371/journal.pone.0182768.

- World Health Organization. 2016.** WHO guidelines for the treatment of *Treponema pallidum* (syphilis). Available at <http://www.who.int/reproductivehealth/publications/rtis/syphilis-treatment-guidelines/en/>.
- Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carriço JA, Achtman M. 2018.** GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Research* **28**(9):1395–1404 DOI [10.1101/gr.232397.117](https://doi.org/10.1101/gr.232397.117).
- Zobaníková M, Mikolka P, Cejková D, Pospíšilová P, Chen L, Strouhal M, Qin X, Weinstock GM, Šmajš D. 2012.** Complete genome sequence of *Treponema pallidum* strain DAL-1. *Standards in Genomic Sciences* **7**(1):12–21 DOI [10.4056/sigs.2615838](https://doi.org/10.4056/sigs.2615838).