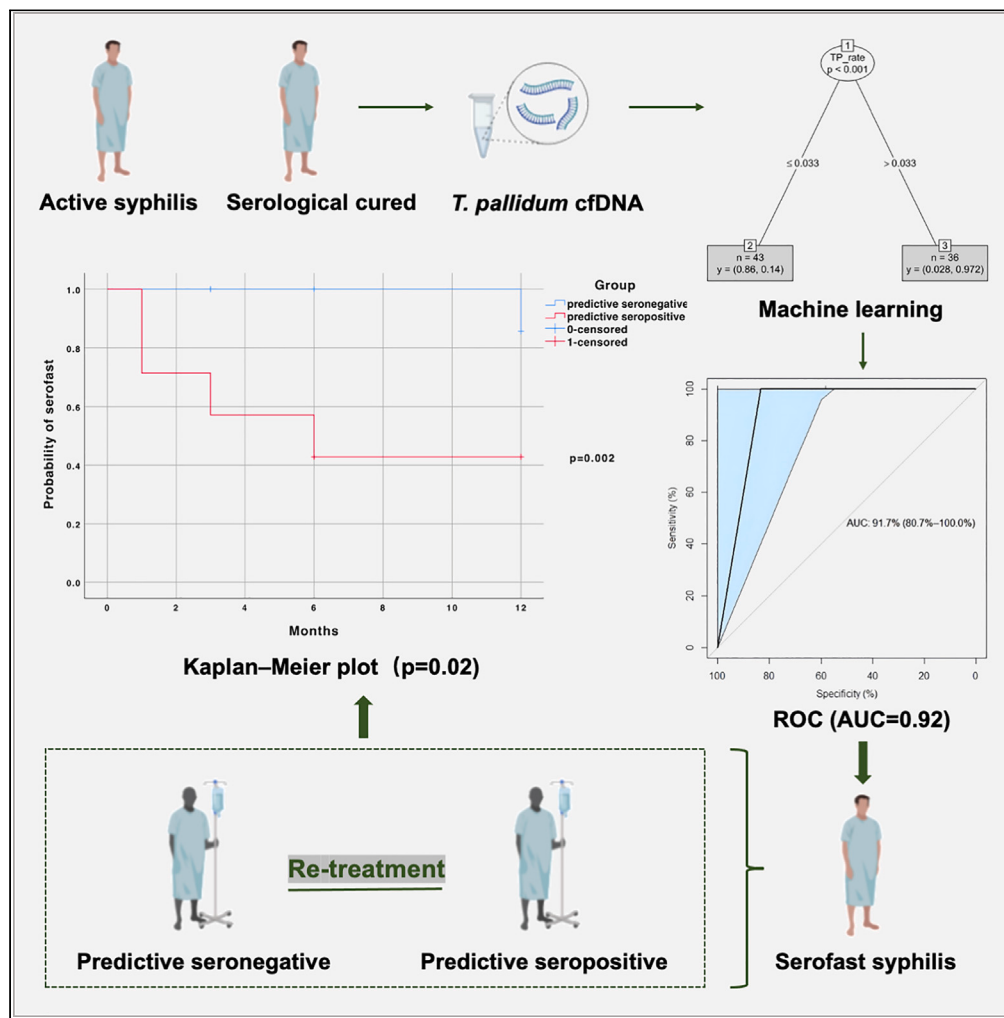


Article

Using circulating microbial cell-free DNA to identify persistent *Treponema pallidum* infection in serofast syphilis patients

Meng Yin Wu, Lu Chen, Li Cheng Liu, ..., Yi Jun Zou, Wei Jun Chen, Jun Li

lijun35@hotmail.com

Highlights

Serofast syphilis may indicate persistent *T. pallidum* infection

T. pallidum cfDNA can be used as an auxiliary diagnostic tool for syphilis

Retreatment is recommended for serofast patients with suspected persistent infection

Article

Using circulating microbial cell-free DNA to identify persistent *Treponema pallidum* infection in serofast syphilis patients

Meng Yin Wu,^{1,7} Lu Chen,^{2,7} Li Cheng Liu,^{2,7} Ming Juan Liu,^{1,7} Yan Feng Li,³ He Yi Zheng,¹ Ling Leng,⁴ Yi Jun Zou,⁵ Wei Jun Chen,⁶ and Jun Li^{1,8,*}

SUMMARY

The question of whether serofast status of syphilis patients indicates an ongoing low-grade *Treponema pallidum* (*T. pallidum*) infection remains unanswered. To address this, we developed a machine learning model to identify *T. pallidum* in cell-free DNA (cfDNA) using next-generation sequencing (NGS). Our findings showed that a TP_rate cut-off of 0.033 demonstrated superior diagnostic performance for syphilis, with a specificity of 92.3% and a sensitivity of 71.4% (AUROC = 0.92). This diagnosis model predicted that 20 out of 92 serofast patients had a persistent low-level infection. Based on these predictions, re-treatment was administered to these patients and its efficacy was evaluated. The results showed a statistically significant decrease in RPR titers in the prediction-positive group compared to the prediction-negative group after re-treatment ($p < 0.05$). These findings provide evidence for the existence of *T. pallidum* under serofast status and support the use of intensive treatment for serofast patients at higher risk in clinical practice.

INTRODUCTION

Syphilis is a contagious, sexually transmitted infection (STI) caused by *Treponema pallidum* (*T. pallidum*). Despite global health efforts, syphilis infection remains a significant health challenge worldwide. The World Health Organization estimated that 17.7 million individuals aged 15–49 are afflicted with syphilis globally, with approximately 5.6 million new cases each year, predominantly in low-income countries.¹ In recent decades, China has experienced a surge in syphilis cases, with an annual increase in incidence as high as 16.3% (Figure 1).² Syphilis infection is a relatively common cause of severe health complications, including adverse pregnancy outcomes and increased HIV transmission rates.

Effective management of syphilis involves early detection³ and appropriate treatment for cases with clinical and serological evidence of *T. pallidum* infection. Currently, benzathine penicillin is the recommended treatment for all stages of syphilis. However, about 15–58% of patients experienced persistent reactive serological test or fail to achieve serological cure,^{4,5} a condition known as the “serofast state”. Among patients with serofast status at 6 months, 73% still remained serofast at 12 months despite re-treatment with 2.4 million units benzathine penicillin.⁶ Furthermore, a recent study showed that 34.6% of asymptomatic syphilis patients with serofast status also have asymptomatic neurosyphilis,⁷ a rate significantly higher than that of the active syphilis group. Consequently, the serofast state is a major concern for both physicians and patients due to the potential risk of persistent low-level *T. pallidum* infection, which may progress to neurosyphilis and result in serious complications.

In recent years, plasma cell-free DNA (cfDNA) has been extensively investigated as a noninvasive method for detecting markers in the management of malignant tumors, as well as for diagnosing fetal chromosomal disorders and blood-borne and encephalitis infections.^{8–10} However, detecting *T. pallidum* cfDNA remains challenging due to its low concentrations in plasma samples of syphilis patients. In this study, we combined novel probe-capture with next-generation sequencing (NGS) and constructed a multi-dimensional machine learning model. We intended to establish a non-invasive approach with enhanced sensitivity for identifying serofast patients at risk of persistent low-level *T. pallidum* infection.

¹Department of Dermatology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

²Beijing Macro & Micro-test Bio-Tech Co., Ltd, Beijing 100083, China

³Department of Neurology, Peking Union Medical College Hospital, Beijing 100730, China

⁴State Key Laboratory of Complex Severe and Rare Diseases, Translational Medicine Center, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

⁵College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

⁶University of Chinese Academy of Sciences, Beijing 100049, China

⁷These authors contributed equally

⁸Lead contact

*Correspondence: lijun35@hotmail.com

<https://doi.org/10.1016/j.isci.2024.109399>



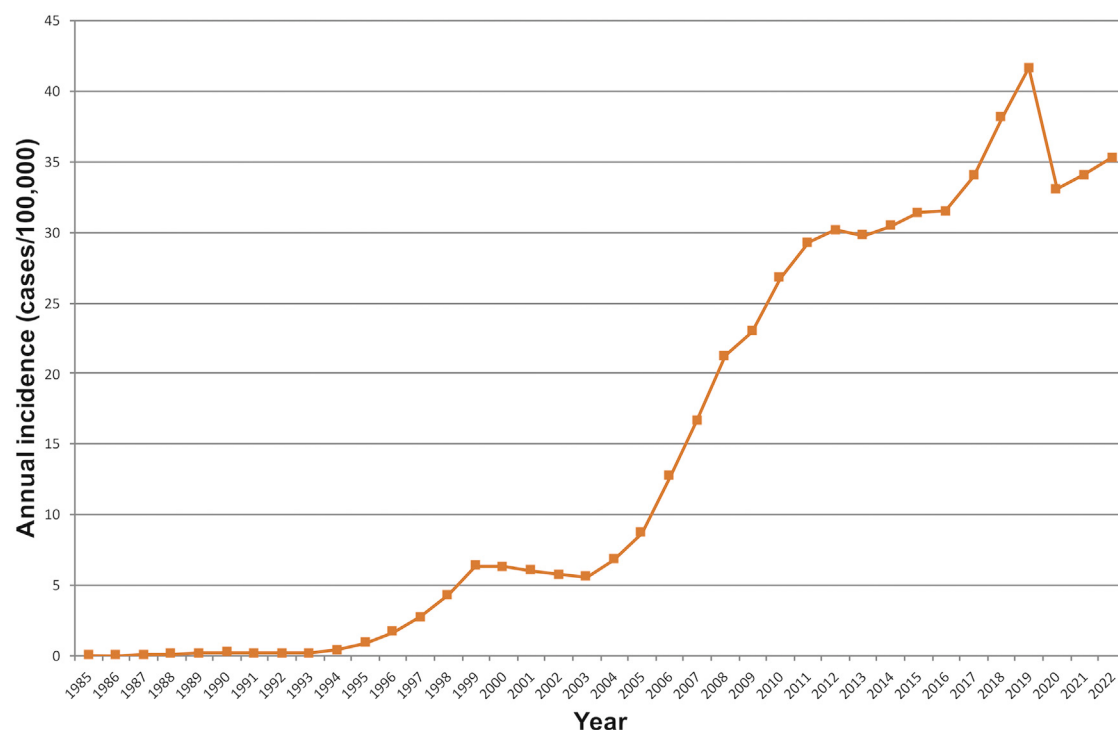


Figure 1. Reported incidence of syphilis cases in China from 1985 to 2022

RESULTS

In this study, 1620 eligibility assessments were performed during a two-year period (Figure 2). A total of 191 individuals met the inclusion criteria, including 48 active syphilis patients, 51 serologically cured individuals and 92 serofast cases (Table 1). 79 patients were recruited in the training cohort for training of machine learning models. 20 patients were set for the test cohort (Figure 2). The study population ranged

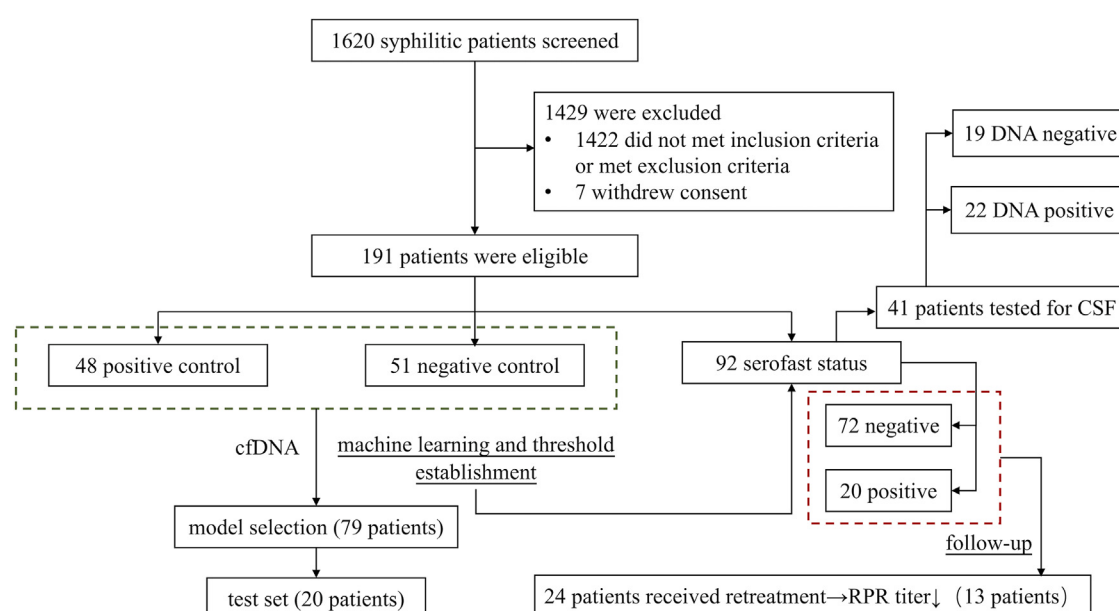


Figure 2. Enrolment and outcomes

The abbreviation cfDNA/DNA denotes T. Pallidum cell-free DNA/DNA; CSF denotes cerebrospinal fluid.

Table 1. Characteristics of syphilis patients at study entry

Characteristic	Untreated (N = 48)	Serological cure (N = 51)	Serofast (N = 92)
Sex			
Male	40	25	37
Female	8	26	55
Age (yr)			
Median	34.5	35.0	37.0
Range	17–84	18–70	20–84
Syphilis stage			
Early syphilis	44	37	42
Late syphilis	4	14	50
Infection status			
HBsAg	3/48	0/51	0/92
HCV	1/48	0/51	1/92
HIV	3/48	2/51	3/92

HBsAg, hepatitis B surface antigen; HCV, hepatitis C virus; HIV, human immunodeficiency virus.

in age from 17 to 84 years, with a mean age of 39.7 years and a median age of 35 years. Males constituted 53.4% (102 individuals) of the participants. The majority (64.4%, 123 individuals) had late syphilis, while 68 had early syphilis. Coinfection with other infectious diseases was observed in 13 cases, with 8 of these being HIV-infected (Table 1).

Based on the training cohort, we filtered out TP_rate as the optimal NGS read feature with the greatest information gain at the root node (Figure 3A). The model built on the TP_rate feature exhibited a prediction power to an area under the receiver operating characteristic curve (AUROC) of 0.92 (95% CI: 0.808–1.000) (Figure 3D). We chose the TP_rate cutoff value of 0.033 based on the 83.3% specificity and 100% sensitivity in the training cohort (Figures 3B and 3C). This cutoff value also yielded a specificity of 92.3% and a sensitivity of 71.4% when applied to the test cohort (Table S1). The cutoff value of TP_rate was also plotted onto the NGS reads data of our patients in the scatterplot (Figures S1–S4).

In the serofast cohort, the model identified 20 positive predictive samples and 72 negative predictive samples. We compared characteristics of the 92 participants and found a significant difference between the predictive positive group and the predictive negative group in terms of CSF proteins ($p = 0.039$). However, no statistically significant differences were observed in terms of sex, age, syphilis stage, baseline RPR titer, other CSF characteristics, treatment and follow-up time. Of the 24 patients who received re-treatment with benzylpenicillin or alternatives (ceftriaxone, doxycycline), 5 (21%) successfully achieved serological cure within 12 months. Only 1 patient exhibited serological cure in the predictive seronegative group, while a much higher proportion (4/5) in the predictive seropositive group achieved serological cure. Eight (33%) subjects exhibited a 2-fold decline in RPR titers, and 11 (46%) patients remained serofast at the end of the observation period. There was a statistically significant difference in serological responses after re-treatment between the two groups ($p = 0.049$). The median number of follow-up RPRs for the positive predictive group and the negative group was 2 (0–4) and 4 (2–16), respectively (Table 2). The comparison of the change in RPR titers between the two groups was also statistically significant ($p = 0.003$) (Table 3). Kaplan-Meier curves of the change in serological cure rates also showed a significant difference between the two groups during the 12-month follow-up after re-treatment ($p = 0.002$) (Figure 4).

DISCUSSION

The enigma of serofast status in syphilis management, where individuals continue to show positive serological test results despite being asymptomatic, remains unresolved. A subset of serofast patients exhibit abnormal changes in CSF without presenting any neurological symptoms.^{11,12} Some of these patients may eventually develop into symptomatic neurosyphilis. Previous studies have suggested that this phenomenon may indicate persistent low-level infections with *T. pallidum*.^{13–15} However, there is currently insufficient evidence to determine whether the serofast condition could contribute to persistent infection or invasion of the central nervous system by *T. pallidum*. The diagnostic methods currently available have limitations. The sensitivity of CSF non-treponemal tests for diagnosing neurosyphilis has been estimated to range from 30 to 70%.¹⁶ Additionally, pooled sensitivities as low as 31.2% in blood and 47.4% in CSF specimens have been reported for the detection of *T. pallidum* DNA by PCR.¹⁷ Consequently, routine laboratory CSF tests might fail to identify some patients with potential CNS invasion. And lumbar puncture, as an invasive and technically difficult procedure, is not highly recommended among HIV-negative serofast syphilis patients in the absence of neurological symptoms.^{18,19}

cfDNA is being studied as a biomarker for noninvasive identification of infectious pathogens. However, due to the low abundance of microbial cfDNA in the blood and its susceptibility to chemical damage,⁸ its clinical diagnostic application is challenging, especially for

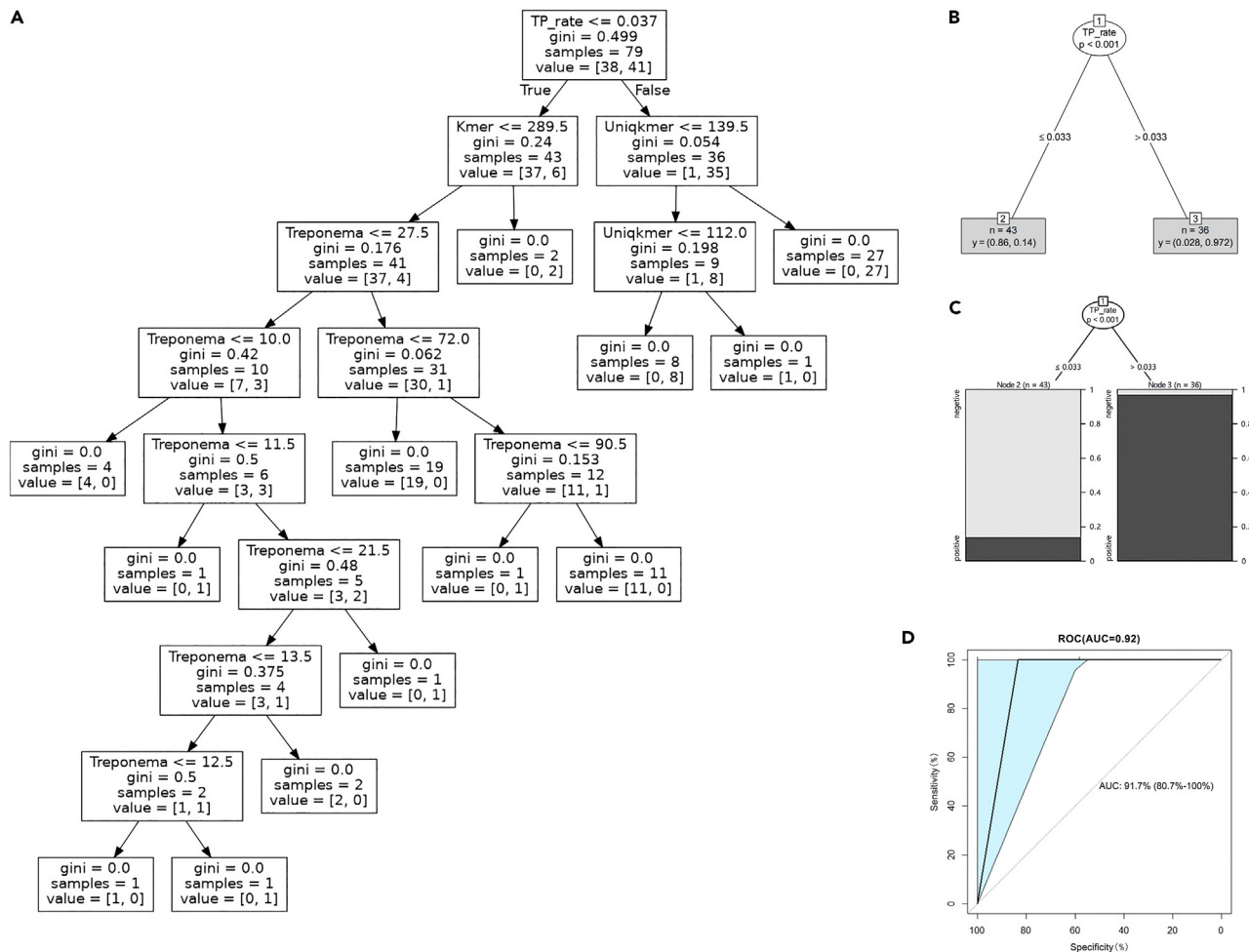


Figure 3. The syphilis diagnostic model achieved the best classification performance with the use of the TP_rate feature

(A) Decision tree for the diagnosis of syphilis.

(B and C) Conditional inference tree using TP_rate to differentiate active syphilis and serological cured syphilis. $TP_rate \leq 0.033$ led to node 2, which contained 43 patients identified as serological cured. $TP_rate > 0.033$ led to node 3, containing 36 active syphilis patients.

(D) Receiver operating characteristic curve for the diagnosis of syphilis. AUROC, area under the receiver operating characteristic curve.

asymptomatic patients. Therefore, we developed a novel probe-capture NGS method to increase the relative concentrations of DNA sequences of *T. pallidum*. However, the presence of detectable microbial DNA does not necessarily imply that the microbe is causing damage or is highly pathogenic. Thus, we investigated the prognostic value of five key cfDNA features (RPM, TP_rate , Kmer, Uniqkmer and Kmer_rate) with machine learning methods, and the feature with the greatest predictive power for syphilis infection status was found to be TP_rate . When using a cut-off value of 0.033, TP_rate demonstrated superior performance in the diagnosis of syphilis with a high specificity of 92.3% and a sensitivity of 71.4% (AUROC 0.92 [95%CI 0.81–1.00]).

To understand whether the serofast condition represents persistent infection, we applied this machine learning model in the serofast cohort to make predictions for the infectious status. It successfully identified 20 (21.7%) samples with predictive positive value, indicating the possibility of persistent infection. We divided the serofast cohort into two groups based on the predictive value. The comparison of characteristics and test results between the two groups showed no significant difference except for CSF proteins. The predictive negative group displayed a marginally elevated level of CSF proteins compared to predictive positive group (0.39 vs. 0.33, $p = 0.039$). This unexpected result may be attributed to the possibility that CSF may not be the only location where *T. pallidum* remains dormant in serofast patients. Prior research has indicated that *T. pallidum* can spread to various organs in humans,²⁰ and it has been observed that, up to 30% of ocular syphilis and 90% of otic syphilis cases exhibited a normal CSF examination.^{21,22}

To further verify this result, we administered re-treatment to both groups separately and monitored the trend of RPR titer decline during the follow-up period. In the predictive seropositive group, 57.1% achieved a serological cure within 12 months, compared to only 5.9% in the predictive seronegative group. The comparison for RPR titers after re-treatment between the two groups showed a significant difference. The trend of the change in serological cure rates by Kaplan-Meier analysis also indicates that there were statistically significant differences

Table 2. Characteristics of syphilis serofast patients

Characteristic	Positive (n = 20)	Negative (n = 72)	p value ^a
Sex			0.17
Male	5 (25%)	8 (11%)	
Female	15 (75%)	64 (89%)	
Age (years)	37.7 (9.7)	42.0 (16.5)	0.26
People who live with HIV/AIDS	0	3	
Syphilis stage			0.45
Early syphilis	11 (55%)	31 (43%)	
Late syphilis	9 (45%)	41 (57%)	
Baseline RPR titer	8 (2–8)	8 (2–8)	0.23
CSF			
Positive TPPA ^c	5/18 (28%)	23/52 (44%)	0.40
RPR titer	0.06 (0.23, n = 18)	0.22 (1.31, n = 52)	0.77
Positive FTA-IgG ^c	3/18 (17%)	15/52 (29%)	0.53
Positive FTA-IgM ^c	0 (n = 18)	0 (n = 52)	
Leukocytosis (cells/ μ L) ^b	2.22 (5.13; n = 18)	1.94 (6.53, n = 52)	0.75
Proteins (mg/L) ^b	0.33 (0.12; n = 18)	0.39 (0.13; n = 52)	0.039
Treatment			0.65
Penicillin	6 (60%)	12 (36%)	
Alternatives ^d	1 (10%)	5 (15%)	
Non-treat	3 (30%)	16 (49%)	
Follow-up time ^b	5 (3.83, n = 7)	8 (3.67, n = 17)	0.92
Serum RPR titer after re-treatment	2 (0–4)	4 (2–16)	0.049

Data are mean (SD), n (%), or median (25–75%), unless otherwise stated. CSF, cerebrospinal fluid.

^aAll tests were two-tailed and considered significant if $p < 0.05$. Fisher's exact tests or Student's t tests were done, as appropriate.

^bData are mean (SD; number of participants with data available).

^cData are n/N (%).

^dAlternative treatment includes ceftriaxone and doxycycline.

between the predictive seropositive group and the predictive negative group at each follow-up time point. However, the impact of age on re-treatment response should not be overlooked. Within the re-treatment cohort, three out of 24 participants were over 60 years old, and two of them showed no change in RPR titers. While both were in the predictive negative group, previous studies have reported a decreased likelihood of RPR titer reduction with older age, which may be attributed to the gradual decline of the immune system with age.^{5,23} The treatment response provided evidence supporting our assumption that some of the serofast patients were still potentially infected. This result also suggests that our probe-capture NGS method could be used to support clinical decision-making regarding the need of re-treatment for serofast syphilis patients.

To our knowledge, this study is the first to utilize the probe-capture NGS method to capture *T. pallidum* cfDNA in serofast syphilis patients. The integration of this method with machine learning provides a data-driven, highly interpretable tool that enhances the clinical interpretation of the infection status in patients suspected of having syphilis. This approach could potentially reduce the number of lumbar punctures and inform decisions about re-treatment in serofast patients.

Our study established a novel testing method for the diagnosis of syphilis and explored the potential reason for the serofast phenomenon. This is the first successfully attempt at detecting *T. pallidum* cfDNA in serofast syphilis patients, which verified the possibility of persistent

Table 3. Comparison of the decline in RPR titer after re-treatment between two groups

	Baseline RPR titer	Serum RPR titer after re-treatment	p value
Positive	8 (4–16)	2 (0–4)	0.027
Negative	8 (2–16)	4 (2–16)	0.017
			0.003

Data are median (25–75%).

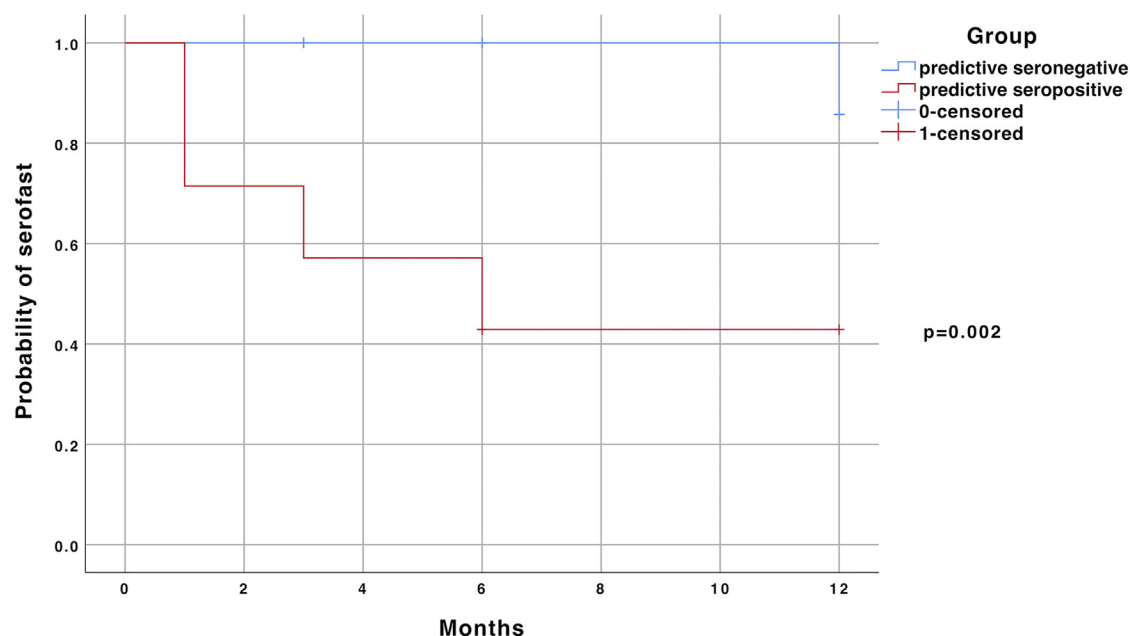


Figure 4. Kaplan-Meier plot of the probability of patients remaining in serofast status in the predictive seropositive group after re-treatment as compared with the predictive negative group ($p = 0.002$ by Mann-Whitney U-test)

infection. However, our findings could not fully elucidate the phenomenon of serofast syphilis. Further research is needed on the microbial information of *T. pallidum* among CSF samples from serofast syphilis patients.

Limitations of the study

However, there are several limitations to this study that should be acknowledged. The relatively small cohort size may have an impact on the model performance, potentially resulting in an underestimation of sensitivity. Additionally, the chemical damage of the cfDNA and contamination by leukocyte genomic DNA may lead to some false negatives. The use of combined data from participants treated with penicillin or other alternative therapies (ceftriaxone or doxycycline) may raise concerns about the possibility of treatment failure. Furthermore, although the positive predictive value can indicate the presence of *T. pallidum* in circulation, it cannot directly support a neurosyphilis diagnosis, potentially leading to over-treatment in clinical practice.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Study design and participants
- METHOD DETAILS
 - Sample collection
 - Design of target capture probes for *T. pallidum*
 - Library preparation and sequencing
 - Sequencing data pre-processing and statistics
 - Machine learning and threshold establishment
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109399>.

ACKNOWLEDGMENTS

The National High Level Hospital Clinical Research Funding (2022-PUMCH-B-092); CAMS Innovation Fund for Medical Sciences (CIFMS 2020-12M-C&T-B-048); the Infectious Diseases Special Project, Ministry of Health of China (2018ZX10732-401; 2018ZX10301101-004).

AUTHOR CONTRIBUTIONS

J.L. and W.C. conceived the study and participated in the study design. M.W., M.L., Y.L., and H.Z. collected the samples. L.C. and Y.Z. prepared the samples and performed next-generation sequencing. L.L. and M.W. conducted the data analysis and established the machine learning model. M.W., L.C., and M.L. wrote the manuscript. L.L. provided critical re to the manuscript. All authors read, revised, and approved the final manuscript for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 12, 2023

Revised: January 6, 2024

Accepted: February 28, 2024

Published: March 1, 2024

REFERENCES

- Peeling, R.W., Mabey, D., Kamb, M.L., Chen, X.S., Radolf, J.D., and Benzaken, A.S. (2017). Syphilis. *Nat. Rev. Dis. Primers* 3, 17073. <https://doi.org/10.1038/nrdp.2017.73>.
- Yang, S., Wu, J., Ding, C., Cui, Y., Zhou, Y., Li, Y., Deng, M., Wang, C., Xu, K., Ren, J., et al. (2017). Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study. *Lancet Infect. Dis.* 17, 716–725. [https://doi.org/10.1016/s1473-3099\(17\)30227-x](https://doi.org/10.1016/s1473-3099(17)30227-x).
- Mabey, D., Peeling, R.W., Ballard, R., Benzaken, A.S., Galbán, E., Changalucha, J., Everett, D., Balira, R., Fitzgerald, D., Joseph, P., et al. (2006). Prospective, multi-centre clinic-based evaluation of four rapid diagnostic tests for syphilis. *Sex. Transm. Infect.* 82, v13–v16. <https://doi.org/10.1136/sti.2006.022467>.
- Clement, M.E., Okeke, N.L., and Hicks, C.B. (2014). Treatment of syphilis: a systematic review. *JAMA* 312, 1905–1917. <https://doi.org/10.1001/jama.2014.13259>.
- Seña, A.C., Wolff, M., Martin, D.H., Behets, F., Van Damme, K., Leone, P., Langley, C., McNeil, L., and Hook, E.W. (2011). Predictors of serological cure and Serofast State after treatment in HIV-negative persons with early syphilis. *Clin. Infect. Dis.* 53, 1092–1099. <https://doi.org/10.1093/cid/cir671>.
- Seña, A.C., Wolff, M., Behets, F., Van Damme, K., Martin, D.H., Leone, P., McNeil, L., and Hook, E.W. (2013). Response to therapy following retreatment of serofast early syphilis patients with benzathine penicillin. *Clin. Infect. Dis.* 56, 420–422. <https://doi.org/10.1093/cid/cis918>.
- Cai, S.N., Long, J., Chen, C., Wan, G., and Lun, W.H. (2017). Incidence of asymptomatic neurosyphilis in serofast Chinese syphilis patients. *Sci. Rep.* 7, 15456. <https://doi.org/10.1038/s41598-017-15641-w>.
- Song, P., Wu, L.R., Yan, Y.H., Zhang, J.X., Chu, T., Kwong, L.N., Patel, A.A., and Zhang, D.Y. (2022). Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. *Nat. Biomed. Eng.* 6, 232–245. <https://doi.org/10.1038/s41551-021-00837-3>.
- Gu, W., Deng, X., Lee, M., Sucu, Y.D., Arevalo, S., Stryke, D., Federman, S., Gopez, A., Reyes, K., Zorn, K., et al. (2021). Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat. Med.* 27, 115–124. <https://doi.org/10.1038/s41591-020-1105-z>.
- Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S.S., Fang, L., Li, Z., Lin, L., Liu, R., et al. (2018). Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations, Patterns of Viral Infections, and Chinese Population History. *Cell* 175, 347–359.e14. <https://doi.org/10.1016/j.cell.2018.08.016>.
- Walter, T., Lebouche, B., Mialhes, P., Cotte, L., Roure, C., Schlienger, I., and Trepo, C. (2006). Symptomatic relapse of neurologic syphilis after benzathine penicillin G therapy for primary or secondary syphilis in HIV-infected patients. *Clin. Infect. Dis.* 43, 787–790. <https://doi.org/10.1086/507099>.
- Zhou, P., Gu, X., Lu, H., Guan, Z., and Qian, Y. (2012). Re-evaluation of serological criteria for early syphilis treatment efficacy: progression to neurosyphilis despite therapy. *Sex. Transm. Infect.* 88, 342–345. <https://doi.org/10.1136/sextrans-2011-050247>.
- Seña, A.C., Zhang, X.H., Li, T., Zheng, H.P., Yang, B., Yang, L.G., Salazar, J.C., Cohen, M.S., Moody, M.A., Radolf, J.D., and Tucker, J.D. (2015). A systematic review of syphilis serological treatment outcomes in HIV-infected and HIV-uninfected persons: rethinking the significance of serological non-responsiveness and the serofast state after therapy. *BMC Infect. Dis.* 15, 479. <https://doi.org/10.1186/s12879-015-1209-0>.
- Lin, L.R., Tong, M.L., Fu, Z.G., Dan, B., Zheng, W.H., Zhang, C.G., Yang, T.C., and Zhang, Z.Y. (2011). Evaluation of a colloidal gold immunochromatography assay in the detection of Treponema pallidum specific IgM antibody in syphilis serofast reaction patients: a serologic marker for the relapse and infection of syphilis. *Diagn. Microbiol. Infect. Dis.* 70, 10–16. <https://doi.org/10.1016/j.diagmicrobio.2010.11.015>.
- Lin, L.R., Zheng, W.H., Tong, M.L., Fu, Z.G., Liu, G.L., Fu, J.G., Zhang, D.W., Yang, T.C., and Liu, L.L. (2011). Further evaluation of the characteristics of Treponema pallidum-specific IgM antibody in syphilis serofast reaction patients. *Diagn. Microbiol. Infect. Dis.* 71, 201–207. <https://doi.org/10.1016/j.diagmicrobio.2011.07.005>.
- Davis, L.E., and Schmitt, J.W. (1989). Clinical significance of cerebrospinal fluid tests for neurosyphilis. *Ann. Neurol.* 25, 50–55. <https://doi.org/10.1002/ana.410250108>.
- Gayet-Ageron, A., Lautenschlager, S., Ninet, B., Perneger, T.V., and Combescur, C. (2013). Sensitivity, specificity and likelihood ratios of PCR in the diagnosis of syphilis: a systematic review and meta-analysis. *Sex. Transm. Infect.* 89, 251–256. <https://doi.org/10.1136/sextrans-2012-050622>.
- Hamill, M.M., Ghanem, K.G., and Tuddenham, S. (2023). State-of-the-Art Review: Neurosyphilis. *Clin. Infect. Dis.* ciad437. <https://doi.org/10.1093/cid/ciad437>.
- Zhang, X., Shahum, A., Yang, L.G., Xue, Y., Wang, L., Yang, B., Zheng, H., Chen, J.S., Radolf, J.D., and Sena, A.C. (2021). Outcomes From Re-Treatment and Cerebrospinal Fluid Analyses in Patients With Syphilis Who Had Serological Nonresponse or Lack of Seroreversion After Initial Therapy. *Sex. Transm. Dis.* 48, 443–450. <https://doi.org/10.1097/olq.0000000000001321>.
- Radolf, J.D., Deka, R.K., Anand, A., Šmajš, D., Norgard, M.V., and Yang, X.F. (2016). Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen. *Nat. Rev. Microbiol.* 14, 744–759. <https://doi.org/10.1038/nrmicro.2016.141>.
- Lapere, S., Mustak, H., and Steffen, J. (2019). Clinical Manifestations and Cerebrospinal Fluid Status in Ocular Syphilis. *Ocul. Immunol. Inflamm.* 27, 126–130. <https://doi.org/10.1080/09273948.2018.1521436>.

22. Yimtae, K., Srirompotong, S., and Lertsukprasert, K. (2007). Orosyphilis: a review of 85 cases Otolaryngology–head and neck surgery. *Otolaryngol. Head Neck Surg.* 136, 67–71. <https://doi.org/10.1016/j.otohns.2006.08.026>.
23. Horberg, M.A., Ranatunga, D.K., Quesenberry, C.P., Klein, D.B., and Silverberg, M.J. (2010). Syphilis epidemiology and clinical outcomes in HIV-infected and HIV-uninfected patients in Kaiser Permanente Northern California. *Sex. Transm. Dis.* 37, 53–58. <https://doi.org/10.1097/OLQ.0b013e3181b6f0cc>.
24. Janier, M., Unemo, M., Dupin, N., Tiplica, G.S., Potočník, M., and Patel, R. (2021). 2020 European guideline on the management of syphilis. *J. Eur. Acad. Dermatol. Venereol.* 35, 574–588. <https://doi.org/10.1111/jdv.16946>.
25. Workowski, K.A., Bachmann, L.H., Chan, P.A., Johnston, C.M., Muzny, C.A., Park, I., Reno, H., Zenilman, J.M., and Bolan, G.A. (2021). Sexually Transmitted Infections Treatment Guidelines, 2021. *MMWR Recomm. Rep.* 70, 1–187. <https://doi.org/10.15585/mmwr.rr7004a1>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
QiagenQIAamp Circulating Nucleic Acid Kit	Qiagen	Cat#55114
GAPDH-qPCR	Macro & micro- test bio-tech	N/A
KAPA HyperPrep Kits for Illumina	Kapa biosystems	Cat#KR0961
NimbleGen SeqCap Hybridization and Wash Kit	Roche	Cat#05634261001
Oligonucleotides		
GAPDH-DNA-F:TCAAGAAGGTGGTGA AGCAGG	This paper	N/A
GAPDH-DNA-R:CAGCGTCAAAGGTGG AGGAGT	This paper	N/A
GAPDH-DNA-P-VIC:5'VIC-CCTCAAGGG CATCCTGGGCTACACT-3'-BHQ1	This paper	N/A
Software and algorithms		
bwa	https://bio-bwa.sourceforge.net/	
fastp	https://github.com/OpenGene/fastp	0.20.1
hisat2	https://daehwankimlab.github.io/hisat2/	2.2.1 release
Kraken2	https://ccb.jhu.edu/software/kraken2/	2.0.7-beta
RefSeq database	https://ftp.ncbi.nih.gov/genomes/refseq	date:July 17, 2020
partykit package in R	https://partykit.r-forge.org/partykit/	V1.2.12
Python	https://www.python.org/	
pROC package in R	https://xrobin.github.io/pROC/	V1.18.0
SPSS	https://www.ibm.com/spss	V. 21.0

RESOURCE AVAILABILITY

Lead contact

Further information and requests for raw data and code should be directed to and will be fulfilled by the lead contact, Jun Li (lijun35@hotmail.com).

Materials availability

All materials reported in this paper will be shared by the [lead contact](#) upon request.

Data and code availability

Data that support the findings of this study have been deposited in Sequence Read Archive (SRA) at NCBI with the submission ID: SUB13467337 and are publicly available as of the date of publication.

The code for the sequencing data identification process and decision tree is available in the following GitHub repository: https://github.com/Karma0alpha/Code_TP.

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design and participants

All the syphilis cases in this study were outpatients who visited the Sexually Transmitted Infection Center of Peking Union Medical College Hospital (PUMCH) from November 2017 to April 2019 (see the [supplementary appendix](#)). This study was approved by the institutional review boards of the PUMCH. All participants gave informed consent for the use of their data for scientific researches.

Eligible participants were adults aged between 18 and 85 years. Laboratory test results and medical history were collected for case definition and confirmation of clinical stage. Patients with malignancy, pregnancy, and those with autoimmune diseases were excluded to minimize the probability of false positives. According to the case definitions, patients were categorized into three groups: untreated active syphilis (primary syphilis, secondary syphilis, tertiary syphilis and latent syphilis), serological cure and serofast status.

Primary syphilis is defined as clinical characterization by anal, genital, or oropharyngeal chancre and inguinal lymphadenopathy. Laboratory confirmation of *T. pallidum* in clinical specimens is required by rapid plasma reagin test (RPR test) and particle agglutination assays for antibodies to *T. pallidum* (TPPA), and/or fluorescent treponemal antibody absorption (FTA-ABS). A secondary syphilis case is defined as a clinically compatible patient characterized by cutaneous rash, mucosal lesions or lymphadenopathy with laboratory test results confirming syphilis. A latent syphilis case is defined as an asymptomatic case with a possible history of infection supported by a reactive RPR and reactive treponemal test. A case is classified as early latent syphilis (ELS) when the initial infection had occurred within a year. When the initial infection had occurred more than 12 months, a case is classified as late latent syphilis (LLS). When the date of initial infection cannot be established, latent syphilis is classified as latent of unknown duration (LUD). A tertiary syphilis is defined as syphilis acquired more than one years previously and has a history of primary, secondary, or latent syphilis with clinical manifestations involving the cardiovascular, ocular or central nervous system (CNS) and laboratory confirmation with positive non-treponemal tests, or CSF abnormalities characterized by higher than normal amounts of white blood cells (WBC) or protein.

As part of routine practice in our STI clinic, after treatment, all the syphilis patients were asked to periodically review their clinical symptoms and serum RPR titers every 3 months. In the early (primary, secondary, ELS) syphilis patients, at 6 months following treatment, a serological cure was defined as either a negative RPR or ≥ 2 dilution (4-fold) decrease in the RPR titer, if initial titers are positive at a 1:1 or 1:2 dilutions, as becoming non-reactive. In the late (LLS, LUD, tertiary) syphilis patients, at 12 months following treatment, a serological cure was defined as either a negative RPR or ≥ 2 dilution (4-fold) decrease in the RPR titer, if initial titers are positive at a 1:1 or 1:2 dilutions, as becoming non-reactive. Serofast status is defined as a < 4 -fold (2-fold dilution) decline in non-treponemal antibody titer at 6–12 months or as persistently low titer after standard treatment (Reinfection and potentially false-positive RPR tests including autoimmune diseases are excluded).²⁴ All serofast patients are recommended for lumbar puncture and CSF tests in case of asymptomatic neurosyphilis. Other neurological diseases are excluded based on clinical history and physical examination. We utilized cfDNA sequencing of *T. pallidum* among active syphilis group and serologically cured group for machine learning and threshold establishment. The constructed model was then applied in prediction for patients in the serofast status group.

Also, we administered re-treatment based on the predictive value of the machine learning model. Patients with a positive predictive value were treated as neurosyphilis cases (18–24 million units of aqueous crystalline procaine penicillin G per day, administered as 3–4 million units intravenously every 4 h for 10–14 days) followed by weekly injections of benzathine penicillin G (2.4 million units intramuscularly) for 3 weeks. Ceftriaxone 1–2 g daily IV for 10–14 days was used as an alternative treatment for patients with a positive reaction to the penicillin skin test. For those patients who were predicted to be negative, we repeated the treatment with 2.4 million units benzathine penicillin IM weekly for 3 weeks. Doxycycline 100 mg orally twice per day for 2 weeks or ceftriaxone 1.0 g intravenously per day for 10 days was used as an alternative therapy for early syphilis (doxycycline 100 mg orally twice per day for 4 weeks for late syphilis).²⁵ All participants were followed up every 3 months for serological tests. Treatment response was classified into three categories by serological change in RPR titer: (1) serological cure (RPR titer dropped > 4 -fold or became negative); (2) serological decline (RPR titer showed a 2-fold decline); (3) non-response (unchanged RPR titer).

All diagnoses were made by a dermatology specialist and retrospectively reviewed by MY Wu and J Li.

METHOD DETAILS

Sample collection

Blood (10 mL) was collected from each patient and was processed within 1 h of sample collection. Of the 10 mL blood, 4 mL was allocated in plain tubes for syphilitic testing using RPR and/or TPPA and/or FTA-ABS tests. The remaining 6 mL was placed in an anticoagulant tube, and centrifuged at 820 g for 10 min to separate the plasma from the peripheral blood cells. The plasma was then further centrifuged at 20,000 g for 10 min to pellet any remaining cells and stored at -80°C before testing. The cfDNA was extracted from 2 mL aliquots of plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions. The cfDNA was eluted into 50 μL AVE buffer and quantified through Qubit immediately or using -80°C stored plasma samples. As an internal control to verify the yield of DNA extraction, we quantified the amount of GAPDH DNA in each eluted DNA sample by Q-PCR (Macroµ Test), and estimated the yield of extraction by comparing the quantity.

CSF samples were collected according to standard sterile procedures. One mL CSF was placed in plain tubes for syphilitic and routine testing. RPR and/or TPPA and/or FTA-ABS tests were performed on the CSF samples.

Design of target capture probes for *T. pallidum*

We designed a probe pool of 92706 probes covering the whole genome of *T. pallidum* (GCF_000246755.1) and 3246 probes covering the human EFTUD2 gene. A k-mer Sliding window protocol was used to design probes. The probe length was 120 bp and the sliding window moves over target regions with 12 bp per step to get a 1-fold enrichment. Candidate probes were mapped with the bwa software to the *Homo sapiens* reference genome (hg19) and NCBI nucleotide databases (Nt databases), which the non-*T. pallidum* was excluded. The probe

sequences that could map to *Homo sapiens* (hg19) or other organisms were removed to get the specific probe set. This TP probe pool was synthesised by Twist Bioscience (USA).

Library preparation and sequencing

We first constructed a dual-indexed sequencing library from 10 ng of extracted cfDNA using KAPA HyperPrep Kits for Illumina (Kapa biosystems). The libraries were then enriched through hybridization using the biotinylated TP probe pool and the NimbleGen SeqCap Hybridization and Wash Kit (Roche). Following ligation-mediated PCR enrichment of the captured libraries, the PCR products were purified using beads to collect the fragments with an average length of 320 bp for sequencing. The resulting fragments were run on an Illumina Nextseq 500 sequencing instrument as single-end 75-bp reads, following the manufacturer's protocol.

Sequencing data pre-processing and statistics

Initially, the raw sequencing data was trimmed using fastp (version 0.20.1) to remove low-quality bases and adaptor sequences, resulting in clean reads. To remove host sequences, the clean reads were mapped to the human genome (GRCh38/hg38) using the sequence alignment software hisat2 (2.2.1 release). Microbial species in the unmapped reads were identified using Kraken2 (version 2.0.7-beta) and the RefSeq database: <https://ftp.ncbi.nih.gov/genomes/refseq> date: July 17, 2020, providing read counts for *T. pallidum* at the species level and *Treponema* at the genus level. From this sequencing data, additional statistics were derived for *T. pallidum* features factors, including the reads per million (RPM) by normalizing the sequencing depth among different samples, and the ratio of *T. pallidum* to *Treponema* (TP_rate). Additionally, metrics such as Kmer, Uniqkmer, and the ratio of Uniqkmer to Kmer (UK_rate) in Kraken2 were used in subsequent analysis workflows.

Machine learning and threshold establishment

Based on the five *T. pallidum* features factors (RPM, Kmer, Uniqk-mer, TP_rate, and UK_rate) datasets obtained from sequencing data processing and the clinical diagnosis results of the participating patients, we built diagnostic models and corresponding thresholds. First, we used Python to perform randomized stratified sampling on a dataset of 99 clinically confirmed specimens. Specifically, the dataset was randomly divided into a 79-case training cohort and a 20-case test cohort according to an 8:2 ratio, with each cohort having a similar proportion of clinically negative/positive specimens as the original dataset. Based on this, a decision tree was constructed for the training cohort using the partykit software package (V1.2.12) in R. The optimal feature factor with the largest information gain was filtered out at the root node as the optimal decision analysis indicator, and the conditional inference tree analysis was performed using the optimal TP_rate factor to obtain the negative/positive diagnostic thresholds and their accuracy for the training cohort prediction accuracy.

To further evaluate the above decision tree model, Receiver-operating characteristic curve analysis was performed to assess the performance of the test cohort data using the model. The AUROC values of each prediction model were compared using DeLong's method and the Bootstrap method: <https://github.com/UBrau/> (ModelPerformance). Scatterplots and boxplots of the diagnostic effect of using cutoff values for all syphilis samples were plotted using the scatterplot package and geom_boxplot package of R, respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using SPSS v. 21.0 (SPSS Inc., Chicago, IL). The characteristics were compared between groups using the Student's t-test or chi-square test for continuous or categorical variables, respectively. The decline in serum RPR titers after re-treatment were compared between the two groups using Mann-Whitney U-test.