

1 **Early syphilis in Guangzhou, China: presentation, molecular detection of *Treponema***
2 ***pallidum*, and genomic sequences in clinical specimens and isolates obtained by rabbit**
3 **infectivity testing**

4
5 Ligang Yang,^{1*} Xiaohui Zhang,^{1*} Wentao Chen,^{1*} Arlene C. Seña,^{2*} Heping Zheng,¹ Yinbo
6 Jiang,¹ Peizhen Zhao,¹ Rongyi Chen,¹ Liuyuan Wang,¹ Wujian ke,¹ Juan C. Salazar,^{3,4}
7 Jonathan B. Parr,² Joseph D. Tucker,² Kelly L. Hawley,^{3,4} Melissa J. Caimano,³ Christopher
8 M. Hennelly,² Farhang Aghakanian,² Feifei Zhang,¹ Jane S. Chen,² M. Anthony Moody,⁵
9 Justin D. Radolf,^{3,4#} Bin Yang^{1#}

10 ¹Dermatology Hospital, Southern Medical University, Guangzhou, China; ²Institute for
11 Global Health and Infectious Diseases, University of North Carolina at Chapel Hill, Chapel
12 Hill, North Carolina, USA; ³University of Connecticut School of Medicine, Farmington,
13 Connecticut, USA; ⁴Connecticut Children's Research Institute, Connecticut Children's,
14 Hartford, Connecticut, USA. ⁵Duke University, Durham, North Carolina, USA

15 ****Contributed equally as first authors***

16 ***#Corresponding authors:***

17 Justin D. Radolf, MD
18 Departments of Pediatrics and Medicine
19 UConn Health
20 263 Farmington Avenue
21 Farmington, CT 06030
22 Connecticut Children's

23 **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

282 Washington Street

24 Hartford, CT 06106

25 Email:jradolf@uchc.edu

26

27 Bin Yang, MD

28 Dermatology Hospital, Southern Medical University

29 Guangdong Provincial Center for Skin Diseases and STD Control

30 2 Lujing Rd, Yuexiu District

31 Guangzhou, China

32 Email:yangbin101@hotmail.com

33

34 Word count: 3607

35 Abstract: 250

36 Tables: 4

37 Figures: 3

38 Summary: We performed *Treponema pallidum* molecular detection and genome sequencing
39 from multiple specimens collected from early syphilis patients and isolates obtained by
40 rabbit inoculation. Our results support the use of whole genome sequencing from rabbit
41 isolates to inform syphilis vaccine development.

42 **Abstract**

43 **Background:** The global resurgence of syphilis requires novel prevention strategies.
44 Whole genome sequencing (WGS) of *Treponema pallidum* (*TPA*) using different specimen
45 types is essential for vaccine development.

46 **Methods:** Patients with primary (PS) and secondary (SS) syphilis were recruited in
47 Guangzhou, China. We collected ulcer exudates and blood from PS participants, and skin
48 biopsies and blood from SS participants for *TPA* *polA* polymerase chain reaction (PCR);
49 ulcer exudates and blood were also used to isolate *TPA* strains by rabbit infectivity testing
50 (RIT). *TPA* WGS was performed on 52 ulcer exudates and biopsy specimens and 25
51 matched rabbit isolates.

52 **Results:** We enrolled 18 PS and 51 SS participants from December 2019 to March 2022.
53 Among PS participants, *TPA* DNA was detected in 16 (89%) ulcer exudates and three (17%)
54 blood specimens. Among SS participants, *TPA* DNA was detected in 50 (98%) skin
55 biopsies and 27 (53%) blood specimens. *TPA* was isolated from 48 rabbits, with a 71%
56 (12/17) success rate from ulcer exudates and 69% (36/52) from SS bloods. Twenty-three
57 matched SS14 clade genomes were virtually identical, while two Nichols clade pairs had
58 discordant *tprK* sequences. Forty-two of 52 unique *TPA* genomes clustered in an SS14 East
59 Asia subgroup, while ten fell into two East Asian Nichols subgroups.

60 **Conclusions:** Our *TPA* detection rate was high from PS ulcer exudates and SS skin
61 biopsies and over 50% from SS whole blood, with RIT isolation in over two-thirds of
62 samples. Our results support the use of WGS from rabbit isolates to inform vaccine
63 development.

64 **Introduction**

65 During the past 20 years, syphilis, caused by the highly invasive and immunoevasive
66 spirochete, *Treponema pallidum subspecies pallidum* (*TPA*), has resurged as a global
67 public health problem [1-3]. In China, 57,196 new syphilis cases (12.7% of all cases in the
68 country) were reported in Guangdong Province in 2021 [4]. Although the province has
69 implemented syphilis intervention programs since 2000, the epidemic calls for novel
70 prevention and control strategies, including vaccine development.

71 Darkfield microscopy (DFM) has been used traditionally for *TPA* direct detection from
72 primary syphilis (PS) lesions, while detection of treponemal and non-treponemal antibodies
73 in serum is the mainstay for syphilis diagnosis [5, 6]. Several studies have demonstrated the
74 sensitivity of real-time *TPA* PCR on ulcer swabs and whole blood for diagnosis of PS and
75 secondary syphilis (SS) [7-10]. The inability to cultivate *TPA* from clinical samples
76 continues to hinder microbiologic diagnosis. Despite recent progress with *in vitro*
77 cultivation [11-13], rabbit infectivity testing (RIT) by intratesticular inoculation of rabbits
78 with clinical specimens remains the only means of recovering *TPA* strains [14, 15].

79 Whole-genome sequencing (WGS) of *TPA* genomes from geographically diverse
80 regions is essential for development of a syphilis vaccine with global efficacy. At present,
81 most *TPA* sequences have originated from Europe, North America and Australia; fewer
82 strains have been identified from Asia [16-19]. The majority of clinical specimens used for
83 WGS have been PS ulcer swabs with high *TPA* burdens [19-22]. Skin biopsies and whole
84 blood from SS patients are other potential specimens but have considerably lower *TPA*
85 burdens than ulcer exudates [5, 10, 23-25]. RIT allows strain amplification from clinical
86 specimens with low *TPA* burdens, but it is unclear whether the genomes of treponemes

87 obtained by rabbit infectivity testing accurately reflect those in the corresponding clinical
88 specimens.

89 In this study, we recruited patients with PS and SS from Guangzhou, China as part of
90 a parent study for syphilis vaccine development [26]. We collected ulcer exudates and
91 whole blood specimens from PS patients, and skin biopsy and whole blood specimens from
92 SS patients for quantitative PCR (qPCR) and RIT. WGS was performed on clinical
93 specimens with sufficient *TPA* DNA copy numbers and on corresponding rabbit isolated
94 strains. In addition to characterizing the clinical presentation of PS and SS in our patient
95 population, we aimed to (i) compare *TPA* qPCR results and RIT between specimen types,
96 (ii) assess the effect of rabbit passage on the genomic sequences of *TPA* strains isolated
97 from patients, and (iii) phylogenetically analyze *TPA* strains by WGS as a prelude for
98 syphilis vaccine development.

99

100 **Methods**

101 *Ethics statements*

102 This research was approved by the Medical Ethics Committee Dermatology Hospital of
103 Southern Medical University (GDDHLS-20181202). Rabbit experimentation was approved
104 by the Ethics Committee of Dermatology Hospital of Southern Medical University (approval
105 no. GDDHLS-20181202,12/12/2018) and South China Agricultural University Experimental
106 Animal Ethics Committee (approval no. 2020C004, 07/05/2020). All rabbits were housed
107 under approved biosafety conditions at South China Agricultural University. Their diet, care,
108 and maintenance conformed to institutional regulations.

109

110 *Patient recruitment*

111 A convenience sample of patients visiting the STI department of the Dermatology Hospital of
112 Southern Medical University (DHSMU) from December 2019 to March 2022 were recruited
113 for the study. Individuals aged ≥ 18 years and diagnosed with PS or SS were eligible for
114 enrollment [26]. Diagnosis of PS was based on the presence of one or more typical anogenital
115 ulcers with a positive DFM and/or reactive nontreponemal (Toluidine red unheated serum test,
116 TRUST) and treponemal (*Treponema pallidum* particle agglutination, TPPA) tests [5].
117 Diagnosis of SS was based on characteristic skin or mucosal lesions plus reactive TRUST
118 and TPPA tests [5, 27].

119 All study participants received standard care including physical examination and
120 treatment of STIs as indicated. For treatment of PS and SS, we provided 2.4 million units
121 (MU) of intramuscular benzathine penicillin G (BPG) weekly for 2 consecutive weeks (total
122 of 4.8 MU) in accordance with Chinese Center for Disease Control and Prevention treatment
123 guidelines [27]. Patients were asked to return within 90 days for a follow-up examination and
124 repeat TRUST titers; TPPA testing was repeated if the initial test was negative.

125 *Data and specimen collection*

126 A case report form (CRF) was used to collect demographic and clinical data, including sexual
127 orientation, sexual and medical histories [26]. Examination and laboratory findings were also
128 recorded in the CRF. For PS patients, we recorded the number and location of ulcers. For SS
129 patients, we recorded the type and location of rash, as well as other mucocutaneous
130 manifestations. For rashes, we estimated Body Surface Area of Involvement (BSA) scores, a
131 disease outcome predictor used in other skin diseases, such as psoriasis [28].

132 From PS participants, we collected ulcer exudates for DFM and research testing. The first
133 swab was used for DFM, and the residual of the first swab plus a second swab were placed
134 into 1 mL of TpCM-2 medium [13] for RIT [15]. The third and fourth swabs were placed in a
135 1.5 mL tube containing DNA/RNA shield (Zymo Research, Irvine, CA, USA; R1100–250)
136 for subsequent DNA extraction. Whole blood specimens were also collected from PS
137 participants for DNA extraction. From SS participants, two 4 mm punch skin biopsies were
138 obtained from rashes and stored in tubes containing DNA/RNA shield. Whole blood
139 specimens were also collected from SS participants for RIT and DNA extraction.

140

141 *Assessment of T. pallidum burdens by TaqMan qPCR*

142 DNA was extracted from participant samples using the DNeasy Blood & Tissue Kit
143 (QIAGEN, Hilden, Germany) according to manufacturer's instructions; 100 µL of
144 nuclease-free water was used to elute the DNA, followed by –80 °C storage. TaqMan PCR
145 for *TPA polA* was performed as previously described by Chen *et. al.* [29] using 1.25 µL of 10
146 µM primers, and measured on Real-Time PCR Instruments (BioRad, Hercules, CA, USA)

147

148 *Rabbit-infectivity testing*

149 RIT was conducted as previously described [15]. In short, adult New Zealand White rabbits
150 (males, 3 months of age, 2.5-3 kg) were pre-screened to confirm non-reactive syphilis
151 serologies at baseline. Genital ulcer exudates from participants diluted with 1 mL of TpCM-2
152 medium or 1ml of freshly collected whole blood was injected into each testis. The serologic
153 status of the rabbits was monitored weekly by *TPA* particle agglutination (TPPA) beginning

154 with the second month after inoculation. Seroconverted rabbits were sacrificed, and their
155 testes were aseptically removed for *TPA* isolation. Seronegative rabbits were euthanized after
156 3 months followed by passage of testicular extract to a second rabbit.

157

158 *Whole genome sequencing and bioinformatic analysis*

159 WGS was carried out on ulcer exudates from individuals with PS, skin biopsy specimens
160 from SS patients, and rabbit-passaged isolates using custom 120-nucleotide RNA
161 oligonucleotide baits obtained from Agilent Technologies (Santa Clara, CA, USA). *TPA*
162 enrichment was performed using the Sure Select XT Low Input kit at the University of
163 North Carolina at Chapel Hill (UNC) and the Sure Select XTHS2 kit at SMU, following
164 the previously established protocol [16]. The resulting *TPA*-enriched libraries were
165 sequenced using the Illumina MiSeq platform at UNC or the NovaSeq platform at SMU,
166 generating paired-end, 150 bp reads. The raw sequencing data, with residual human reads
167 removed, has been deposited at the Sequence Read Archive database (PRJNA815321).
168 Fifty-two *TPA* genomes from China included in this manuscript are also reported in
169 separate study by Seña *et al.* [26].

170 The sequencing reads and publicly available data were processed using a conservative
171 bioinformatic pipeline with minor modifications [16], which is available at
172 https://github.com/IDEELResearch/TPAallidum_genomics. Only sequences with at least
173 80% of their genome covered 3x were retained for variant calling. Variant calling was
174 performed with GATK HaplotypeCaller [30], using joint genotyping using GenotypeVCFs
175 module and hard filtering by VariantFiltration.

176 Comparison of paired *TPA* genomes (rabbit-passaged versus direct sequencing from
177 clinical samples) was conducted manually using a VCF file obtained from joint genotyping,
178 applying strict filtering parameters (depth \geq 8x, genotype quality \geq 20, and variant allele
179 frequency (VAF) \geq 0.85). Variants genes (*e.g.*, *tpvK*) were considered to be heterogeneous
180 if portion of reads supporting less frequent alleles were greater than 15% (*i.e.*, VAF
181 between 0.15 and 0.85). To generate a multiple alignment using the MAFFT (v7.520) [31],
182 we combined 52 Nichols- and SS14-clade consensus sequences generated in this study, 109
183 published *TPA* genomes and 2 outgroup genomes (*Treponema subsp. endemicum* Bosnia A
184 and *subsp. pertenue* Samoa D). The resulting alignment was then inputted into Gubbins
185 (v3.2.1),[32] which processed the alignment with default parameters for predicting and
186 masking homologous recombination regions. Phylogenetic tree was constructed under the
187 GTRGAMMA model with 1000 bootstraps replicates by Gubbins (v3.2.1), which then was
188 visualized using ggtree [33]. The genotypic resistance of macrolide was analyzed by
189 competitive mapping as previously described by Beale *et al.* [18, 34].

190

191 *Statistical analysis*

192 Descriptive analyses were conducted on participants' demographics and clinical
193 characteristics, *TPA* PCR, and RIT results. All data were analyzed using SPSS 22.0 (IBM®
194 SPSS® Statistics).

195

196 **Results**

197 *Sociodemographic data, clinical characteristics, and laboratory findings*

198 From December 2019 to March 2022, 154 participants met the screening criteria (Figure 1).
199 Of the 115 eligible persons, 69 individuals (18 PS and 51 SS) consented for enrollment.
200 Among enrolled participants, the median age was 27 (IQR: 22-32); 84% were male, and
201 33% were men who have sex with men (MSM) or men who have sex with both men and
202 women (MSMW) (Table 1). All participants but one were HIV-negative, 9% had a history
203 of prior syphilis, and 22% had a history of other STIs.

204 Seventeen of the 18 participants with PS were male, with the majority (61%)
205 presenting with solitary ulcers (Table 2, Figure 2A and 2B); one male participant presented
206 with a nearly healed chancre. Spirochetes were visualized by DFM in 12 of the 17 (71%)
207 patients with exudative lesions. The TRUST and TPPA tests were positive in 8 (45%) and
208 16 (89%) PS patients, respectively. *TPA* DNA was detected in 16 of 18 (89%) specimens
209 collected from chancres and 3 of 18 (17%) PS blood specimens. *TPA polA* copy numbers
210 ranged from 1.43 – 2250 copies/ μ l in swabs and 0.288- 37.99 copies/ μ l in whole blood
211 (Supplemental Table 1). All 12 DFM positive specimens were positive by *TPA* PCR
212 (Supplemental Table 1). *TPA* DNA was detected in ulcer exudates and whole blood in two
213 PS cases with negative TRUST and TPPA results (Supplemental Table 1).

214 Forty two of the 51 (82%) SS participants presented with a rash only; the other nine
215 presented with other manifestations, such as condyloma lata and alopecia (Figure 2 and
216 Table 3). BSA scores were <3% in 15 participants (30%), 3-10% in 10 (19%), and >10% in
217 26 (51%). All SS participants had reactive TRUST and TPPA tests. *TPA* DNA was
218 detected in 50 of 51 (98%) skin biopsies and 27 of 51 (53%) whole blood specimens. *TPA*

219 *polA* copy numbers ranged from 0.610 - 10,200 copies/μl in skin biopsies and 0.110 - 3,850
220 copies/μl in blood (Supplemental Table 2).

221

222 *Rabbit infectivity testing*

223 We inoculated New Zealand white rabbits with 17 ulcer exudates and 52 whole blood
224 specimens (one from the PS patient with a nearly healed chancre and 51 from SS patients;
225 Table 4). Fifty-one rabbits (74%) developed reactive TPPA tests, which included 14/17
226 (82%) rabbits inoculated with ulcer exudates and 37/52 (71%) inoculated with blood. The
227 median time to TPPA positivity following inoculation was 42 days for ulcer exudates
228 versus 50 days for whole blood. *TPA* strains were isolated from 48 of the 51 rabbits that
229 TPPA seroconverted – 12/17 (71%) from ulcer exudates and 36/52 (69%) from blood
230 (Table 4). One of two rabbits inoculated with PCR-negative exudates and 16 of 25 rabbits
231 inoculated with PCR-negative blood samples also yielded isolates (Table 4). Conversely,
232 RIT was negative in two of 15 PCR-positive ulcer exudates and seven of 27 PCR-positive
233 blood specimens. *TPA polA* copy numbers in the two PCR-positive RIT-negative ulcer
234 exudates were comparable to those in PCR-positive RIT-positive samples (Suppl. Table 1).
235 In contrast, *polA* copy numbers were at the low end in five of the seven PCR-positive
236 RIT-negative blood samples (Suppl. Table 2).

237

238 *Comparison of TPA genomic sequences from paired clinical samples and rabbit isolates*

239 We successfully generated a comprehensive set of 77 *TPA* genomes from 69 participants.

240 This dataset included 25 paired genomes, comprising clinical samples and their

241 corresponding rabbit-passaged isolates, as well as 27 unmatched genomes consisting of 23
242 rabbit-passaged isolates and 4 clinical samples. Supplementary Table 3 presents detailed
243 information regarding the samples subjected to WGS.

244 We compared *TPA* genomes from 25 clinical specimens and corresponding rabbit
245 isolated strains to ascertain whether WGS from rabbit-passaged isolates accurately reflect
246 those in the source clinical specimens. To ensure the reliability and rigor of the results, we
247 only included variants in the analysis that passed conservative filtering criteria.

248 Twenty three of the 25 genomes were identified as SS14-lineage, while the remaining
249 two were Nichols-like. Within the SS14-like group, we observed 6 discordant variants at 3
250 specific genomic sites when we applied the rigorous variant filtering criteria
251 (Supplementary table 4). However, with manual inspection, we found three additional
252 discrepant heterogenous variants (including 1 SNP and 2 indels) in *tprK* (Supplementary
253 table 5). Within the two paired Nichols-lineage genomes, we identified a total of 28
254 discrepant variations. Two of these discordant variants were found in one of the paired
255 genomes, while the remaining 26 discordant variants were present in the other
256 (Supplementary Table 6). All 28 discordant variants were found exclusively within the
257 *tprK* gene (Supplemental Table 6). Due to the complexity of the *tprK* region and the
258 limitations of short-read sequencing, we are unable to go deeper to potentially uncover
259 additional *tprK* variants. Overall, the results clearly showed that rabbit passage has minimal
260 effects on *TPA* genomic stability.

261

262 *Phylogenetic diversity of new TPA strains*

263 We constructed a phylogenetic tree to assess the diversity of the newly sequenced *TPA*
264 strains. For genomes from matched samples and rabbit isolates, we prioritized the use of
265 clinical samples. When clinical samples were unavailable, we utilized genomes from
266 rabbit-passaged isolates. The final analyzed dataset, therefore, consisted of 29 genomes
267 directly sequenced from clinical samples and 23 genomes from rabbit-passaged strains (7
268 from ulcer exudates and 16 from blood samples). We also included 109 previously
269 published *TPA* genomes, including 28 *TPA* strains from Japan and 22 others available from
270 China up to the present date. Additionally, our analysis included outgroup genomes from
271 *Treponema pallidum* subsp. *endemicum* Bosnia A and *Treponema pallidum* subsp.
272 *pertenue* Samoa D (Figure 3). The 52 newly sequenced Chinese strains fell into three
273 subgroups: 42 clustered in the SS14-Omega East Asia subgroup, eight were classified into
274 the Nichols C subgroup, and two were assigned to the Nichols E subgroup.

275 *TPA* strains harboring mutations in the 23S rRNA gene at positions A2058G or
276 A2059G are associated with genotypic resistance to macrolides [34, 35]. Forty-eight of the
277 newly sequenced Chinese *TPA* strains exhibited genotypic macrolide resistance,
278 predominantly through the A2058G allele mutation, while four strains were harbored
279 genotypic macrolide resistance with A2059 allele mutation (**Figure 3**).

280

281 **Discussion**

282 We enrolled participants who presented with a wide range of PS and SS clinical
283 manifestations, allowing for multiple specimen types for *TPA* molecular detection and
284 genome sequencing. In addition, we were able to adapt a traditional, highly specialized

285 syphilis research tool, RIT, for an entirely novel purpose of direct relevance to vaccine
286 development – obtaining isolates for WGS. The *TPA* genomic sequences from our site with
287 accompanying clinical data have provided new insights into the syphilis epidemic in
288 Guangzhou, paving the way for molecular epidemiological studies that can synergize with
289 future vaccine strategies.

290 We observed several interesting clinical findings in our study. Thirty-nine percent of our
291 PS participants presented with multiple ulcers in the anogenital area although single ulcers
292 are typical [36, 37]; over half of all PS participants had negative TRUST titers,
293 underscoring the importance of direct detection for diagnosis. We found the sensitivity of
294 DFM for PS in our study to be 71%, compared to the range of 70-100% in the literature [10,
295 38, 39]. Among our SS participants, we noted that over half of the SS participants
296 had >10% BSA scores suggesting widespread dermal involvement from *TPA*. Rashes on
297 the palms and soles are often indicative of a secondary syphilis diagnosis, reported in 40%
298 to 80% of cases in the literature [40, 41]; however, less than half of our participants had
299 these typical findings.

300 We found that the sensitivity of *TPA polA* PCR was 89% for chancres and 17% for
301 whole blood from PS participants, compared to 98% for skin biopsies and 53% for blood
302 from SS participants. Although our PCR sensitivity for whole blood was higher than in
303 some reports [10, 42, 43], the overall low detection rates are likely explained by the low
304 concentration of live *TPA* circulating in blood and the small volumes of specimens used for
305 PCR testing. We used PCR to detect *TPA* in SS biopsy specimens and achieved an
306 extraordinarily high positivity rate, supporting its use for diagnosis if it can be performed.

307 Our results support the importance of DFM and PCR to assist in the diagnosis and staging
308 of syphilis [5, 10, 43].

309 RIT has long been regarded as the gold standard for detection of small numbers of live
310 treponemes in clinical specimens [14, 15]. While a variety of specimens have been
311 successfully inoculated into rabbits over the years [14, 15], studies of acquired syphilis in
312 the antibiotic era have focused on cerebrospinal fluid to identify patients with
313 asymptomatic neurosyphilis [44, 45]. Our study, therefore, is unique in its emphasis on RIT
314 of ulcer exudates and whole blood. Interestingly, we isolated *TPA* in some rabbits
315 inoculated with PCR-negative blood, emphasizing the extraordinary sensitivity of RIT. The
316 differences in sensitivity between RIT and PCR may reflect the volume of whole blood
317 used for RIT versus *TPA* qPCR. On the other hand, we also noted that specimens with
318 detectable *TPA* DNA were not uniformly RIT-positive. Treponemal burdens in samples at
319 or near the minimal infectious inoculum of individual *TPA* strains may partially explain
320 these discordant results [46], while the presence of *TPA* DNA does not always reflect
321 viable treponemes [47]. Moreover, it is not widely appreciated that clinical strains do not
322 readily adapt to rabbits [48]. Indeed, as shown herein, amplification of the initial inoculum
323 by serial rabbit passage is typically required. We also noted that organisms may not be
324 recoverable from rabbits that clearly became infected based on conversion of their TPPA
325 tests.

326 A critical question in the use of RIT for WGS is whether rabbit passage induces
327 genomic alterations in the parental strain harbored by the inoculated clinical sample. We
328 addressed this issue by performing the first large scale comparison of genomic sequences

329 from matched clinical samples and rabbit isolates. The results clearly showed that rabbit
330 passage has minimal effects on *TPA* genomic stability. The genomes of all 23 recovered
331 SS14 clade strains were essentially identical to their clinical counterparts, while variation in
332 the two Nichols clade isolates was confined to their *tprK* loci. Evidence from WGS for
333 recombination in *TPA* strains [49, 50] implies that co-infection must occasionally occur,
334 presumably in genital ulcers where high treponemal burdens would maximize the
335 probability for genetic exchange. However, in all 25 matched cases, genomes in clinical
336 samples and rabbit isolates clearly derived from the same *TPA* strain. We anticipate that
337 future comparisons of this type will eventually identify co-infected individuals.

338 The newly sequenced *TPA* genomes were found to cluster within the SS14-Omega
339 Asia, Nichols C, and E subgroups. Interestingly, although three previously reported
340 Chinese SS14-clade *TPA* strains were found to cluster in the SS14-Omega subgroup [16],
341 the vast majority of Chinese *TPA* strains published to date belonged to the SS14-Omega
342 East Asian subgroup [51-53]. Furthermore, the Chinese *TPA* strains showed a high degree
343 of relatedness to publicly available Japanese *TPA* strains, both in the SS14-lineage and
344 Nichols-lineage strains. Notably, genotypic resistance to macrolides was observed in all
345 newly sequenced strains, indicating that azithromycin should not be considered as an
346 alternative drug for treating syphilis patients in China.

347 There were several limitations in our study which included our small sample size
348 enrolled from one site in Guangzhou. However, we encountered challenges with the
349 COVID-19 pandemic in China, and our population is representative of patients presenting
350 to a provincial hospital with a broad catchment area. We only inoculated one whole blood

351 specimen from a PS participant who had a healed ulcer, therefore limiting our assessment
352 of this sample type for RIT isolation. Based on the low *TPA* copy numbers in blood during
353 PS, however, we would not expect similar RIT isolation rates to whole blood from SS
354 participants. Our short-read sequencing and analysis approach prevents us from fully
355 resolving challenging but important loci such as *tprK*. However, our conservative variant
356 calling approach reduced ambiguity and allowed us to focus our analysis on
357 high-confidence discordant loci.

358 Our study serves as a reminder of the remarkable diversity of PS and SS clinical
359 manifestations, and the importance of sensitive detection methods. We showed that DFM,
360 which is no longer performed in most STI clinics, remains a valuable tool for rapid
361 diagnosis of exudative mucocutaneous lesions. We also confirmed previously reported high
362 qPCR sensitivities for ulcer exudates and skin biopsies. Although exclusively a research
363 tool, RIT provided a powerful means to obtain *TPA* strains for WGS from blood – a
364 compartment critical to the systemic disease process but heretofore inaccessible to genomic
365 investigation. The greater availability of WGS may open the door to genome-based studies
366 of local sexual networks that can assist in the development of novel epidemiologic
367 strategies for syphilis control in hyper-endemic areas. *TPA* outer membrane proteins are
368 widely considered the prime candidates for syphilis vaccine design [54]. Identification of
369 immunologically relevant outer membrane protein variants through WGS of *TPA* strains
370 from clinical specimens and RIT passaged isolates will greatly facilitate development of a
371 globally efficacious syphilis vaccine.

372

373 **Contributors**

374 ACS, JCS, KH conceptualized the overall project and the study, and HZ, JBP led genomic
375 sequencing efforts. All authors contributed to study design and implementation. LY and
376 BY led the clinical sites in enrollment, participant data and specimen collection. Both
377 contributed equally to this activity. XZ, RC, LW and WK participant in participant
378 enrollment, data and specimen collection. XZ, PZ FZ and JSC conducted data curation and
379 formal analysis of the clinical data. WC conducted data curation and formal analysis of the
380 genomic data. ACS, LY, JSC, JDT, JCS developed the case report forms. FZ assisted with
381 participant enrollment. WC, YJ, HZ conducted the rabbit-infectivity testing. WC and CMH
382 conducted the DNA extraction and *polA* PCR quantitation. FA, WC, HZ, CMH, MJC, JBP
383 contributed to the genome sequencing and phylogenetic analyses. LY, XZ, WC, ACS, JDR,
384 JSC prepared the tables and figures. XZ, PZ, FZ, MJC, ACS, JCS provided input into data
385 analyses. JDT, HZ, BY, JCS, JDR, JBP provided administration, resources and supervision
386 for the clinical or research laboratory sites. ACS, JDR, MAM contributed to the funding
387 acquisition from the NIH and BMGF. ACS, JSC, FA, and JBP verified the underlying data.
388 LY, XZ, WC prepared the manuscript, which was reviewed and edited by all authors. All
389 authors had full access to all the data in the study and had final responsibility for the
390 decision to submit for publication.

391

392 **Acknowledgments**

393 This project was funded by the US National Institutes of Health National Institute for
394 Allergy and Infectious Disease (NIAID) (U19AI144177 to JDR and MAM). This work also

395 was supported, in part, by the Bill & Melinda Gates Foundation (INV-036560 to ACS),
396 strategic research dollars from Connecticut Children's, NIAID (T32 T32AI007151 to FFA),
397 the National Nature Science Foundation of China (82220108006 to BY), the National
398 Institute of Virology and Bacteriology (Programme EXCELES, ID Project No.
399 LX22NPO5103, Funded by the European Union - Next Generation EU to DS). Under the
400 grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License
401 has already been assigned to the Author Accepted Manuscript version that might arise from
402 this submission. Fifty-two *TPA* genomes from China included in this manuscript are also
403 reported in separate study by Seña et al. [26].

404 The authors would like to thank all of the study participants and research staff in
405 Guangzhou, Cali, Lilongwe, and Chapel Hill. They would like to thank Myron Cohen for
406 his support and guidance as a prior member of the Scientific Review Committee for the
407 grant. They would also like to thank Andreea Waltmann and Fredrick Nindo for early
408 contributions to genomic sequencing methods development and data analysis pipelines,
409 respectively.

410

411 **Potential conflicts of interest**

412 All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form
413 for Disclosure of Potential Conflicts of Interest.

414

415 **Data availability**

416 The deidentified participant REDCap database will be made publicly available at the time

417 of publication. Requests to share the informed consent forms and database can be made by
418 emailing the corresponding authors at jradolf@uchc.edu and yangbin101@hotmail.com.
419 Raw sequencing data from this study with residual human reads removed are available
420 through the Sequence Read Archive (SRA, BioProject PRJNA815321). Data supporting
421 the findings of this study are available within the manuscript and Supplementary material.

Table 1. Sociodemographic Characteristics of 69 Patients Enrolled in Guangzhou, China

Characteristics	Primary syphilis n=18 (%)	Secondary syphilis n=51 (%)	Total n=69 (%)
Age (median/inter-quartile range)	27 (IQR 23-31)	27 (IQR 22-33)	27 (IQR 22-32)
Gender			
Women	1 (6)	10 (20)	11 (16)
Men	17 (94)	41 (80)	58 (84)
Sexual orientation			
Heterosexual	12 (67)	33 (65)	45 (65)
Men who have sex with men	4 (22)	9 (18)	13 (19)
Men who have sex with men and women	2 (11)	8 (16)	10 (14)
Declined to answer	0 (0)	1 (2)	1 (1)
Marital status			
Single	12 (67)	34 (67)	46 (67)
Married	6 (33)	12 (24)	18 (26)
Divorced/separated/widowed	0 (0)	5 (10)	5 (7)
Educational status			
Middle school or less	1 (6)	6 (12)	7 (10)
High school	2 (11)	15 (29)	17 (25)
Trade school	6 (33)	12 (24)	18 (26)
Undergraduate or Graduate	9 (50)	18 (35)	27 (39)
Age of first sexual encounter			
15~17 years	2 (11)	13 (25)	15 (22)
≥18 years	16 (89)	38 (75)	54 (78)
Reported sexual partners in past 3 months			
None	0 (0)	5 (10)	5 (7)
One	9 (50)	25 (49)	34 (49)
Two	3 (17)	9 (18)	12 (17)
Three or more	6 (33)	12 (24)	18 (26)
Reported sexual partners in past 12 months			
None	0 (0)	1 (2)	1 (1)
One	4 (22)	8 (16)	12 (17)
Two	3 (17)	14 (27)	17 (25)
Three or more	11 (61)	28 (55)	39 (57)
Commercial sex engagement	2 (11)	11(22)	13 (19)
HIV status at enrollment			
Positive	0 (0)	1 (2)	1 (1)
Negative	18 (100)	50 (98)	68(99)
History of syphilis	1 (6)	5 (10)	6 (9)
History of other sexually transmitted infections	4 (22)	11 (22)	15 (22)

Table 2. Clinical characteristics and laboratory findings for patients with primary syphilis

Characteristic	Number n=18 (%)
Location of genital ulcers	
Vulva	1 (6)
Penis	17 (94)
Number of genital ulcers	
1 ulcer	11 (61)
2 ulcers	2 (11)
3 or more ulcers	5 (28)
Darkfield microscopy ^a	
Positive	12 (71)
Negative	5 (29)
TPPA	
Positive	16 (89)
Negative	2 (11)
TRUST titer	
≥1:64	1 (6)
1:16 - 1:32	2 (11)
1:1 - 1:8	5 (28)
Negative	9 (50)
Unknown	1 (6)
TPPA+TRUST	
Both positive	9 (50)
Both negative	2 (11)
qPCR chancres ^b	
Positive	16 (89)
Negative	2 (11)
qPCR blood	
Positive	3 (17)
Negative	15 (83)

^a n=17; one patient did not undergo dark field microscopy because the ulcer was nearly healed.

^bqPCR was performed on 17 swabs and one biopsy.

Table 3. Clinical characteristics and laboratory findings among patients with secondary syphilis

Characteristic	Number n= 51 (%)
Manifestations	
Rash only	42 (82)
Rash and condyloma lata	1 (2)
Rash and mucous patches	1 (2)
Rash and alopecia	6 (12)
Condyloma lata only	1 (2)
Location of rash	
Trunk	39 (76)
Extremities	33 (65)
Palms and soles	23 (45)
Ano-genital area	15 (29)
Face	2 (4)
BSA score for rash	
<3%, mild	15 (30)
3%~10%, moderate	10 (19)
>10%, severe	26 (51)
Initial TRUST titer	
≥ 1:64	21 (41)
1:16 - 1:32	28 (55)
1:4 - 1:8	2 (4)
<1:4	0
<i>TPA</i> PCR of skin biopsy	
Positive	50 (98)
Negative	1 (2)
<i>TPA</i> PCR of blood	
Positive	27 (53)
Negative	24 (47)

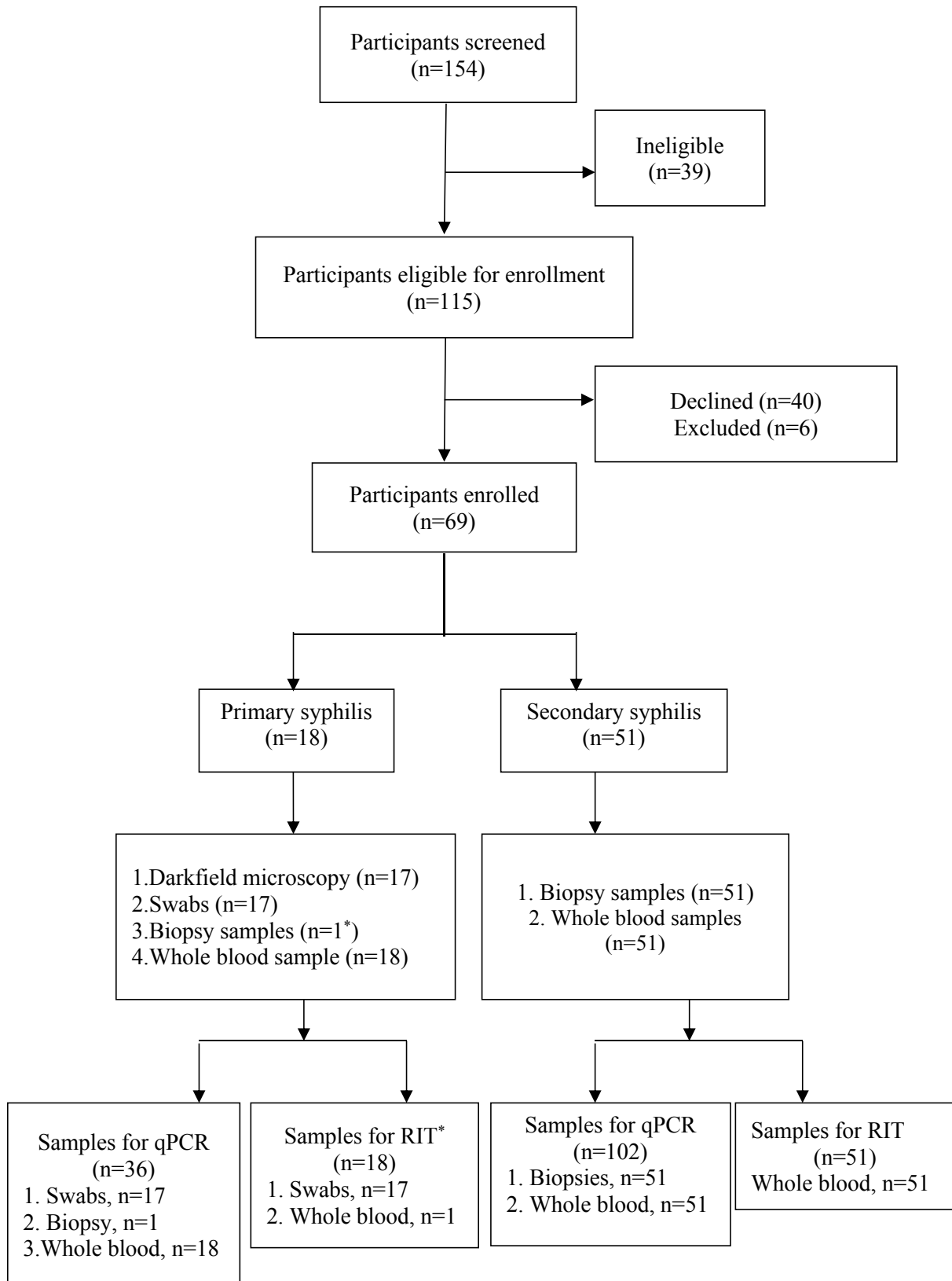
BSA: Body Surface Area.

Table 4. Comparison of *TPA* qPCR and rabbit infectivity testing (RIT) from 69 participants

	RIT (-) (%)	TPPA seroconverted (%)	<i>TPA</i> isolated (%)
Ulcer specimens (n=17)			
qPCR+, n=15	2 /15 (13)	13/15 (87)	11/15 (73)
qPCR-, n=2	1/2 (50)	1/2 (50)	1 /2(50)
Total	3/17 (18)	14/17 (82)	12/17 (71)
Blood specimens[#] (n=52)			
qPCR+, n=27	7/27 (26)	20/27 (74)	20/27 (74)
qPCR-, n=25	8/25 (32)	17/25 (68)	16/25 (64)
Total	15/52 (29)	37/52 (71)	36/52 (69)

[#]51 from SS patients and 1 from a patient with a nearly healed chancre.

Figure 1. Screening and enrollment algorithm of participants with primary and secondary syphilis.



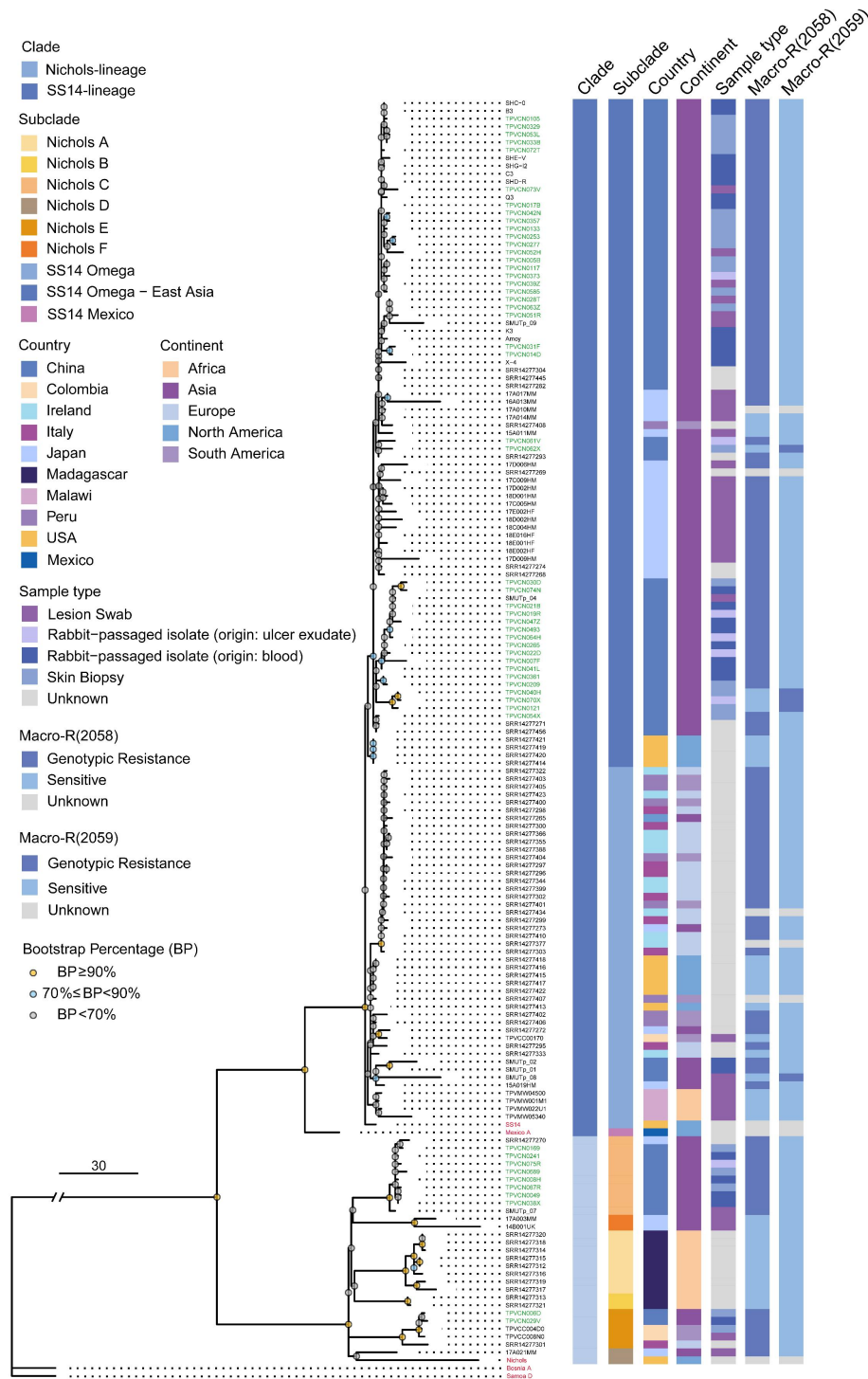
* 17 rabbits were inoculated with ulcer exudates, 1 rabbit was inoculated with blood as the chancre was almost healed. qPCR: qualitative polymerase chain reaction; RIT: rabbit infectivity testing

Figure 2. Representative clinical manifestations in study participants



A. Penile chancre. B. Chancre near scrotum. C scrotal maculopapular rash of secondary syphilis. D. perianal condyloma lata. E. Alopecia of secondary syphilis. F, G1, and G2. Psoriaform rash of secondary syphilis. H. Recently appeared rash of secondary syphilis (TRUST titer 1:4)

Figure 3. Phylogenetic diversity of new Chinese and published *Treponema pallidum* strains. The recombination-masked phylogeny tree was constructed using genetic sequences from 52 newly identified Chinese strains, 109 previously published strains [16, 17, 22, 52, 53, 55], and 2 outgroup strains (TPE and TEN) for comparison. The new Chinese *TPA* strains are highlighted in green on the tree, while five reference strains (SS14, Nichols, Mexico A, TPE, and TEN) are designated in red.



REFERENCES

1. Newman L, Rowley J, Vander Hoorn S, et al. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS One* **2015**; 10(12): e0143304.
2. Tao Y, Chen MY, Tucker JD, et al. A nationwide spatiotemporal analysis of syphilis over 21 years and implications for prevention and control in China. *Clin Infect Dis* **2020**; 70(1): 136-9.
3. Tang S, Shi L, Chen W, et al. Spatiotemporal distribution and sociodemographic and socioeconomic factors associated with primary and secondary syphilis in Guangdong, China, 2005-2017. *PLoS Negl Trop Dis* **2021**; 15(8): e0009621.
4. GD HC. The Health Commission of Guangdong Province announced the incidence of notifiable infectious diseases in the province for 2021 and January 2022. **2022**.
5. Sena AC, Pillay A, Radolf JD. *Treponema* and *Brachyspira*, human host-associated spirochetes. In: Carroll KC, Pfaller MA, Landry ML, et al. *Manual of Clinical Microbiology*. 12 ed. Washington, D.C.: ASM Press, **2019**:1083-108.
6. Ghanem KG, Ram S, Rice PA. The modern epidemic of syphilis. *New England Journal of Medicine* **2020**; 382(9): 845-54.
7. Gayet-Ageron A, Sednaoui P, Lautenschlager S, et al. Use of *Treponema pallidum* PCR in testing of ulcers for diagnosis of primary syphilis. *Emerg Infect Dis* **2015**; 21(1): 127-9.
8. Zhou L, Gong R, Lu X, Zhang Y, Tang J. Development of a multiplex real-Time PCR assay for the detection of *Treponema pallidum*, HCV, HIV-1, and HBV. *Jpn J Infect Dis* **2015**; 68(6): 481-7.
9. Wang C, Cheng Y, Liu B, et al. Sensitive detection of *Treponema pallidum* DNA from the whole blood of patients with syphilis by the nested PCR assay. *Emerg Microbes Infect* **2018**; 7(1): 83.
10. Theel ES, Katz SS, Pillay A. Molecular and direct detection tests for *Treponema pallidum* subspecies *pallidum*: a review of the literature, 1964–2017. *Clinical Infectious Diseases* **2020**; 71(Supplement_1): S4-S12.
11. Edmondson DG, Hu B, Norris SJ. Long-term *in vitro* culture of the syphilis spirochete *Treponema pallidum* subsp. *pallidum*. *mBio* **2018**; 9(3).
12. Edmondson DG, DeLay BD, Kowis LE, Norris SJ. Parameters affecting continuous *in vitro* culture of *Treponema pallidum* strains. *mBio* **2021**; 12(1).
13. Edmondson DG, Norris SJ. *In vitro* cultivation of the syphilis spirochete *Treponema pallidum*. *Curr Protoc* **2021**; 1(2): e44.
14. Turner TB, Hardy PH, Newman B. Infectivity tests in syphilis. *Br J Vener Dis* **1969**; 45(3): 183-95.
15. Lukehart SA, Marra CM. Isolation and laboratory maintenance of *Treponema pallidum*. *Curr Protoc Microbiol* **2007**; Chapter 12: Unit 12A 1.
16. Chen W, Smajs D, Hu Y, et al. Analysis of *Treponema pallidum* strains from China using improved methods for whole-genome sequencing from primary syphilis chancres. *J Infect Dis* **2021**; 223(5): 848-53.
17. Nishiki S, Lee K, Kanai M, Nakayama SI, Ohnishi M. Phylogenetic and genetic characterization of *Treponema pallidum* strains from syphilis patients in Japan by whole-genome sequence analysis from global perspectives. *Sci Rep* **2021**; 11(1): 3154.
18. Beale MA, Marks M, Cole MJ, et al. Global phylogeny of *Treponema pallidum* lineages reveals recent expansion and spread of contemporary syphilis. *Nature Microbiology* **2021**; 6(12): 1549-60.

19. Taouk ML, Tairaoa G, Pasricha S, et al. Characterisation of *Treponema pallidum* lineages within the contemporary syphilis outbreak in Australia: a genomic epidemiological analysis. *The Lancet Microbe* **2022**.
20. Pinto M, Borges V, Antelo M, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nat Microbiol* **2016**; 2: 16190.
21. Arora N, Schuenemann VJ, Jager G, et al. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nat Microbiol* **2016**; 2: 16245.
22. Lieberman NAP, Lin MJ, Xie H, et al. *Treponema pallidum* genome sequencing from six continents reveals variability in vaccine candidate genes and dominance of Nichols clade strains in Madagascar. *PLoS Negl Trop Dis* **2021**; 15(12): e0010063.
23. Cruz AR, Pillay A, Zuluaga AV, et al. Secondary syphilis in Cali, Colombia: new concepts in disease pathogenesis. *PLoS Negl Trop Dis* **2010**; 4(5): e690.
24. Tipple C, Hanna MO, Hill S, et al. Getting the measure of syphilis: qPCR to better understand early infection. *Sex Transm Infect* **2011**; 87(6): 479-85.
25. Cruz AR, Ramirez LG, Zuluaga AV, et al. Immune evasion and recognition of the syphilis spirochete in blood and skin of secondary syphilis patients: two immunologically distinct compartments. *PLoS Negl Trop Dis* **2012**; 6(7): e1717.
26. Sena AC, Matoga MM, Yang L, et al. Clinical and genomic diversity of *Treponema pallidum* subsp. *pallidum*: A global, multi-center study of early syphilis to inform vaccine research. *medRxiv* **2023**.
27. CCfDcaP NCfSC. Guidelines for diagnosis and treatment of syphilis, gonorrhoea, and genital *Chlamydia trachomatis* infection. *Chin J Dermatol* **2020**; 32: 168-79.
28. Ogdie A, Shin DB, Love TJ, Gelfand JM. Body surface area affected by psoriasis and the risk for psoriatic arthritis: a prospective population-based cohort study. *Rheumatology (Oxford)* **2022**; 61(5): 1877-84.
29. Chen CY, Chi KH, George RW, et al. Diagnosis of gastric syphilis by direct immunofluorescence staining and real-time PCR testing. *J Clin Microbiol* **2006**; 44(9): 3452-6.
30. Van der Auwera G, O'Connor B. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* 1st ed. Boston: O'Reilly, **2020**.
31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**; 30(4): 772-80.
32. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **2015**; 43(3): e15.
33. Yu G, Smith DK, Zhu H, Guan Y, Tsan-Yuk Lam T. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Meth Ecol Evol* **2017**; 8: 28-36.
34. Beale MA, Marks M, Sahi SK, et al. Genomic epidemiology of syphilis reveals independent emergence of macrolide resistance across multiple circulating lineages. *Nat Commun* **2019**; 10(1): 3255.
35. Lukehart SA, Godornes C, Molini BJ, et al. Macrolide resistance in *Treponema pallidum* in the United States and Ireland. *N Engl J Med* **2004**; 351(2): 154-8.
36. Towns JM, Leslie DE, Denham I, Azzato F, Fairley CK, Chen M. Painful and multiple anogenital lesions are common in men with *Treponema pallidum* PCR-positive primary syphilis without herpes simplex virus coinfection: a cross-sectional clinic-based study. *Sex Transm Infect* **2016**; 92(2): 110-5.
37. Radolf JD, Tramont EC, Salazar JC. Syphilis (*Treponema pallidum*). In: Mandell GL, Dolin R, Blaser MJ. *Mandell, Douglas and Bennett's Principles and Practice of*

- Infectious Diseases. 9 ed. Philadelphia: Churchill Livingstone Elsevier, **2019**:2865-92.
38. Larsen SA, Steiner BM, Rudolph AH. Laboratory diagnosis and interpretation of tests for syphilis. *Clin Microbiol Rev* **1995**; 8(1): 1-21.
 39. Wheeler HL, Agarwal S, Goh BT. Dark ground microscopy and treponemal serological tests in the diagnosis of early syphilis. *Sex Transm Infect* **2004**; 80(5): 411-4.
 40. Chapel TA. The signs and symptoms of secondary syphilis. *Sex Transm Dis* **1980**; 7(4): 161-4.
 41. Forrestel AK, Kovarik CL, Katz KA. Sexually acquired syphilis: Laboratory diagnosis, management, and prevention. *J Am Acad Dermatol* **2020**; 82(1): 17-28.
 42. Grange PA, Gressier L, Dion PL, et al. Evaluation of a PCR test for detection of *Treponema pallidum* in swabs and blood. *J Clin Microbiol* **2012**; 50(3): 546-52.
 43. Zhou C, Zhang X, Zhang W, Duan J, Zhao F. PCR detection for syphilis diagnosis: Status and prospects. *Journal of Clinical Laboratory Analysis* **2019**; 33(5).
 44. Lukehart SA, Hook EW, 3rd, Baker-Zander SA, Collier AC, Critchlow CW, Handsfield HH. Invasion of the central nervous system by *Treponema pallidum*: implications for diagnosis and treatment. *Ann Intern Med* **1988**; 109(11): 855-62.
 45. Rolfs RT, Joesoef MR, Hendershot EF, et al. A randomized trial of enhanced therapy for early syphilis in patients with and without human immunodeficiency virus infection. The Syphilis and HIV Study Group. *N Engl J Med* **1997**; 337(5): 307-14.
 46. 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 Gonorrhea Vener Dis* **1948**; 32(1): 1-18.
 47. Noordhoek GT, Wolters EC, de Jonge ME, van Embden JD. Detection by polymerase chain reaction of *Treponema pallidum* DNA in cerebrospinal fluid from neurosyphilis patients before and after antibiotic treatment. *J Clin Microbiol* **1991**; 29(9): 1976-84.
 48. Turner TB, Hollander DH. Biology of the Treponematoses. Geneva World Health Organization, **1957**.
 49. Grillova L, Oppelt J, Mikalova L, et al. Directly sequenced genomes of contemporary strains of syphilis reveal recombination-driven diversity in genes encoding predicted surface-exposed antigens. *Front Microbiol* **2019**; 10: 1691.
 50. Noda AA, Mendez M, Rodriguez I, Smajs D. Genetic Recombination in *Treponema pallidum*: Implications for diagnosis, epidemiology, and vaccine development. *Sex Transm Dis* **2022**; 49(1): e7-e10.
 51. Tong ML, Zhao Q, Liu LL, 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): e0182768.
 52. Liu D, Tong ML, Liu LL, Lin LR, Zhang HL, Yang TC. Characterisation of the novel clinical isolate X-4 containing a new tp0548 sequence-type. *Sex Transm Infect* **2021**; 97(2): 120-5.
 53. Sun J, Meng Z, Wu K, et al. Tracing the origin of *Treponema pallidum* in China using next-generation sequencing. *Oncotarget* **2016**.
 54. Hawley KL, Montezuma-Rusca JM, Delgado KN, et al. Structural modeling of the *Treponema pallidum* OMPEome: a roadmap for deconvolution of syphilis pathogenesis and development of a syphilis vaccine. *J Bacteriol* **2021**; 203(15): e0008221.
 55. Liu D, Chen R, He Y, et al. Longitudinal variations in the *tprK* gene of *Treponema pallidum* in an Amoy strain-infected rabbit model. *Microbiol Spectr* **2023**: e0106723.