

1 The Effect of Repeated Mass Drug Administration on the Transmission of Yaws: A Genomic 2 Epidemiology Study

3
4 Amber Barton DPhil¹, Petra Pospíšilová PhD², Camila G Beiras PhD^{3,4}, Lucy N. John MD⁵, Wendy
5 Houinei MPH⁵, Lorenzo Giacani PhD^{6,7}, David Šmajš PhD², Michael Marks PhD^{8,9,10}, Oriol Mitjà
6 PhD^{3,4}, Mathew A Beale PhD^{1*}, Nicholas R Thomson PhD^{1,8*}

- 7
8 1. Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton,
9 Cambridgeshire, UK
10 2. The Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
11 3. Fight Infectious Diseases Foundation, Hospital Universitari Germans Triás i Pujol, Badalona, Spain
12 4. Universitat Autònoma de Barcelona, Bellaterra, Spain
13 5. National Department of Health, Port Moresby, Papua New Guinea
14 6. Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA,
15 USA
16 7. Department of Global Health, University of Washington, Seattle, WA, USA
17 8. Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK
18 9. Hospital for Tropical Diseases, University College London Hospital, London, UK
19 10. Division of Infection and Immunity, University College London, UK
20 *co-senior authors

21 Abstract

22 **Background**

23 Yaws, a neglected tropical disease caused by *Treponema pallidum* subspecies *pertenue*, has evaded
24 eradication, in part due to a high proportion of asymptomatic cases. A cluster-randomised trial in a yaws
25 endemic area compared repeated mass drug administration (MDA) with one round of MDA followed
26 by targeted treatment. Repeated rounds of MDA reduced active and latent prevalence of yaws but led
27 to emergence and spread of azithromycin resistance to three children. Here we aimed to finely delineate
28 the dynamics of *T. pertenuae* sub-lineages over the course of this trial.

29 **Methods**

30 We performed whole genome sequencing directly on DNA from 263 swabs collected during this trial,
31 recovering 222 good-quality *T. pertenuae* genomes. We examined the phylogenetic relationships
32 between genomes linked to geospatial and patient metadata.

33 **Findings**

34 We identified 29 fine-scale sub-lineages of *T. pertenuae*, of which ten were eliminated by MDA, whilst
35 13 persisted in the control arm, one in the experimental arm, two in both study arms, and three were
36 first observed after commencing MDA. The two persistent sub-lineages had non-synonymous mutations
37 in penicillin binding proteins. One of these sub-lineages evolved macrolide resistance (N=3), and was
38 associated with lowered treponemal antibody levels ($p = 0.004$) and longer ulcer duration ($p = 0.015$).
39 Despite the study taking place within a relatively small geographic area (Namatanai District, in the
40 Island of New Ireland, Papua New Guinea) sub-lineages were geographically clustered, and older
41 children were more likely to share sub-lineages ($p = 6 \times 10^{-9}$).

42 **Interpretation**

43 Our findings show that the re-emergence of yaws after MDA was attributed to multiple sub-lineages.
44 The majority of these sub-lineages were detected in the population prior to MDA, and participants were
45 more likely to share sub-lineages within the same ward, suggesting that re-emergence was mostly driven
46 by local transmission. These findings could help inform future yaws elimination strategies.

47 **Funding**

48 European Research Council, European Union, Provincial Deputation of Barcelona, Barberà Solidaria

49 **Conclusions and Public Health Implications** Research that has not been certified by peer review and should not be used to guide clinical practice.

50 **Keywords** *Treponema pertenuae*, mass drug administration, yaws, genomics, epidemiology

51

52

53 Research in context

54 **Evidence before this study**

55 We searched PubMed on 24th July 2024 using the terms treponema * pertenuae OR yaws, genom* OR
56 sequencing, and "mass drug administration" OR azithromycin OR "mass treatment", without
57 restrictions for language or date. Two studies were previously published on the current cohort from the
58 Namatanai Province, Papua New Guinea, comparing three rounds of mass drug administration (MDA)
59 with one round of MDA followed by targeted treatment. These studies used multi-locus sequence-
60 typing and found that repeated MDA limited yaws to one sequence-type but resulted in three cases of
61 macrolide resistance. A separate study used whole genome sequencing to find that after a single round
62 of MDA on Lihir Island, Papua New Guinea, a rebound in yaws cases was caused by multiple sub-
63 lineages of the same MLST, but evolution of macrolide resistance only occurred once. No studies have
64 yet examined how repeated MDA affects the whole-genome diversity and evolution of yaws.

65 **Added value of this study**

66 Our findings show that re-emergence of yaws after MDA was caused by multiple sub-lineages, most of
67 which were already present in the population before MDA. In the group undergoing three rounds of
68 MDA there was still a small rebound in cases six months after the third round, caused by two
69 "persistent" sub-lineages with mutations in penicillin-binding proteins. One of these sub-lineages was
70 associated with lower treponemal antibody and developed macrolide resistance. Sub-lineages were
71 more likely to be shared between older participants and those in close geographical proximity.

72 **Implications of all the available evidence**

73 These data suggest that re-emergence is predominantly driven by cases missed by the initial round of
74 MDA rather than by importation of new cases. Much more efficient population suppression was
75 achieved using three rounds of MDA, and this is likely due to more comprehensively treating the
76 population and eliminating latent cases. Geographical clustering of sub-lineages suggests that
77 elimination by maintaining cases at a low enough prevalence to result in stochastic "fade out" could be
78 feasible and achieved by repeated mass drug administration. Transmission was found to be most
79 common amongst older children, and targeted approaches focusing on these groups may be beneficial.
80 However, ongoing surveillance for macrolide resistance will be needed to achieve eradication.

81

82 Introduction

83 The burden of the neglected tropical disease yaws, caused by infection with *Treponema pallidum*
84 subspecies *pertenuae* (*T. pertenuae*)¹, is highest in Papua New Guinea and the Solomon Islands².
85 Transmission is thought to be driven by skin-to-skin contact, with 75% of new cases in those aged
86 below 15 years¹. While primary lesions heal in 3-6 months, disseminated secondary lesions affect skin
87 and bones, and late-stage tertiary yaws can result in severe bone lesions¹. Without an eradication
88 campaign, yaws would cause an estimated 1.6 million disability-adjusted life years from 2015-2050³.

89 A worldwide elimination campaign led by WHO and UNICEF between 1952-1964⁴ successfully
90 reduced yaws 95% by treating clinically apparent cases with benzathine benzylpenicillin (BPG)
91 injections. However, due to logistical barriers in using BPG in remote locations, combined with a high
92 prevalence of asymptomatic latent cases, yaws has evaded eradication. In 2012 a randomised controlled
93 trial established that single-dose oral azithromycin had a 96% efficacy in treating yaws⁵, raising the
94 possibility of mass drug administration (MDA). This prompted WHO to launch the Morges Strategy
95 for yaws eradication, providing MDA in endemic communities followed by targeted treatment of
96 clinical cases at six-month intervals⁶.

97 In 2013-2016, a longitudinal study on Lihir Island, Papua New Guinea, found that the Morges strategy
98 only reduced yaws cases temporarily, before a recrudescence driven by three distinct sub-lineages of
99 multi-locus sequence-type (MLST) JG8/J_E11^{7,8}. Furthermore, five epidemiologically-linked cases with
100 the A2059G 23S rRNA macrolide resistance mutation were detected. 36-61% of cases identified during
101 targeted treatment rounds were due to residents missing the initial MDA, suggesting that multiple
102 rounds may be necessary to disrupt transmission. To evaluate this strategy, a 2018-2019 cluster-
103 randomised trial comprising 56,676 individuals was carried out in the Namatanai District of the Island
104 of New Ireland, Papua New Guinea, comparing a control group following the Morges strategy (one
105 round of MDA followed by two rounds of active case treatment every six months) with an experimental
106 group undergoing three rounds of MDA⁹. After 18 months, active cases were reduced by three-fold in
107 the control group, and 11-fold in the experimental group. However, three epidemiologically linked cases
108 harbouring the 23S rRNA A2058G macrolide resistance mutation were identified at the final timepoint
109 in the experimental group¹⁰.

110 Whilst MLST indicated that molecular diversity fell after MDA, 93% of cases belonged to a single
111 sequence type⁹. We hypothesised that genome-scale resolution would allow us to define sub-lineages
112 within the same MLST, allowing us to trace their distribution over the study. We therefore performed
113 direct whole genome sequencing (WGS) on samples from the Namatanai trial and confirmed that MDA
114 had a dramatic impact on *T. pertenuis* diversity, and that further rounds of MDA prevented the rebound
115 in diversity seen in the control group. We further used these genomic data to investigate the potential
116 fitness advantages of two sub-lineages which persisted in both arms throughout the study.

117

118 Methods

119

120 **Study population**

121 Study samples were collected during a cluster-randomised trial where 38 wards in three local-level
122 governments areas (LLGs) were randomised to either a control group following the Morges strategy or
123 an experimental group undergoing three rounds of MDA (appendix 1 p2)⁹. Estimated population
124 coverage ranged from 65% in the second treatment round to 99% in the follow-up survey. The trial
125 protocol was approved by the Medical Research Advisory Committee of the Papua New Guinea
126 National Department of Health. Informed oral consent was obtained from all trial participants, or for
127 children, the parents/guardians of trial participants.

128

129 **Whole genome sequencing and analysis**

130 From 297 PCR-positive swabs, 263 unique samples underwent library preparation using the previously
131 described pooled sequence-capture approach⁸ and whole genome sequencing (appendix 1 p2).
132 Sequencing reads were prefiltered, trimmed and down-sampled then mapped to Samoa D *T. pertenuis*
133 reference genome (appendix 1 p2). We detected recombinogenic regions (appendix 1 p6), created a
134 recombination-masked single nucleotide polymorphism (SNP) alignment including previously
135 sequenced contextual samples, and inferred a maximum-likelihood phylogeny (appendix 1 pp2-3).
136 Following ancestral reconstruction, samples were clustered into sub-lineages using rPinecone v0.1.0.
137 with a SNP threshold of 1¹¹. ARIBA version 2.14.6¹² was used to infer MLST according to the scheme
138 described by Medappa et. al¹⁰. Macrolide resistance alleles were inferred using the competitive mapping
139 approach previously described⁸.

140

141 **Pair-wise comparisons**

142 For every possible pair of samples where participant metadata was available, excluding comparisons
143 between the same sample, we identified whether pairs fell into the same sub-lineage, whether they were
144 from the same age group or sex, or whether they were from the same LLG, ward, or neighbouring wards
145 (appendix 1 p3). We then assessed whether pairs sharing the same sub-lineage were over-represented
146 in each of these categories using a chi-squared test. We calculated Bonferroni-corrected p-values to
147 account for multiple comparisons.

148

149 **Role of the funding source**

150 The funders of the study had no role in study design, data collection, data analysis, data interpretation,
151 or writing of the report. All authors had full access to all the data in the study.

152

153 Results

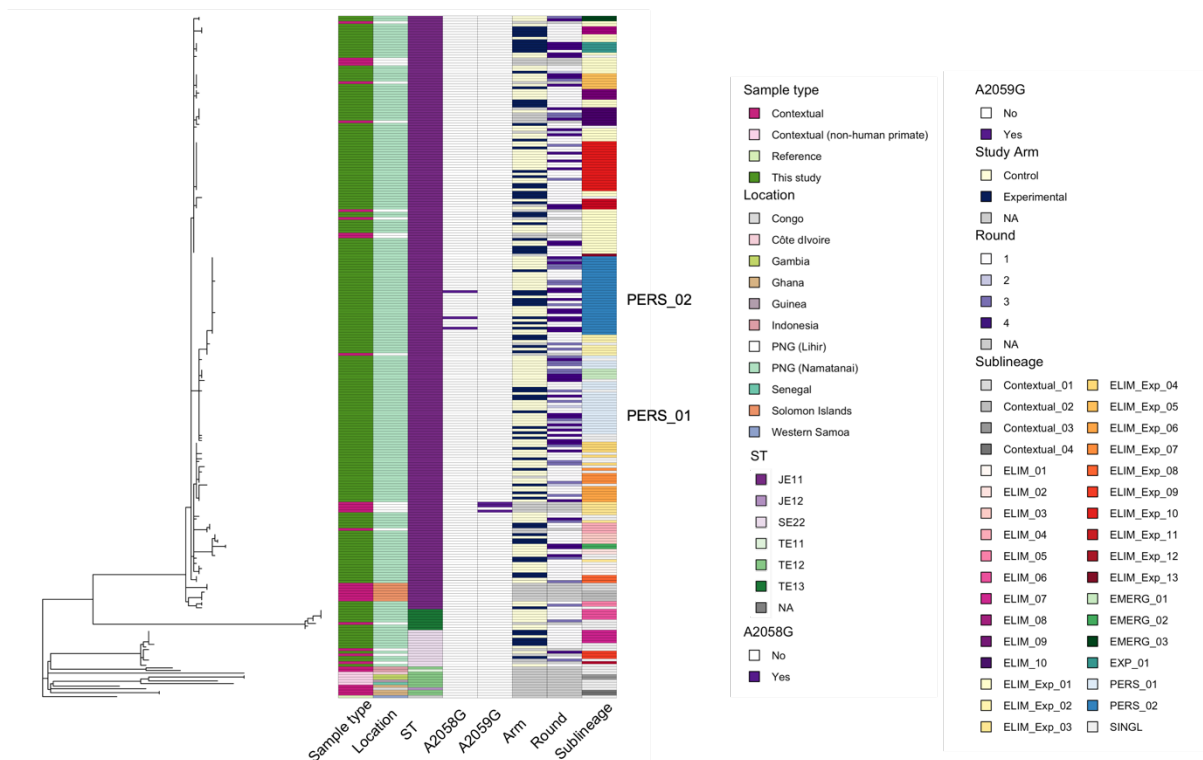
154 **Namatanai *T. pertenu* genomes are genetically homogenous but can be subdivided into distinct 155 sub-lineages**

156 After sequencing and filtering the genomes for quality, our dataset comprised 222 novel *T. pertenu*
157 genomes from Namatanai, and we contextualised these with 38 globally derived genomes, including 20
158 from nearby Lihir Island. Comparative analysis of genomes from Namatanai sequenced during this
159 study indicated very close genomic relationships (Figure 1), with a median pair-wise SNP distance of
160 5 (range 0-139). Namatanai genomes fell within three phylogenetic clades corresponding to the MLST
161 types J_E11 (n = 204), S_E22 (n = 11) and T_E13 (n = 7) described elsewhere^{9,10} and corresponding to types
162 JG8, SE7 and TD6 described previously^{7,8}. Our use of whole genome sequencing enabled us to finely
163 delineate these clades into 29 sub-lineages and 29 singletons with a distance of only 0-2 SNPs within
164 each sub-lineage. For the purposes of this study, we adopted a sub-lineage naming system based on
165 their properties within the trial ("PERS"- sub-lineage persisted in both arms throughout; "EXP"- sub-
166 lineage found in the experimental group only; "ELIM"- sub-lineage eliminated from all trial groups;
167 "ELIM_Exp"- sub-lineage eliminated from the experimental arm only; "EMERG"- sub-lineage first
168 detected after Round 1 MDA).

169

170 Our phylogenetic analysis using a globally derived genome collection indicated that genomes classified
171 as ST J_E11 from Papua New Guinea and the Solomon Islands were phylogenetically distinct from all
172 other *T. pertenu* genomes (Figure 1 showing all samples analysed; appendix 1 p7 with the predominant
173 Papua New Guinea J_E11 clade collapsed). However, analysis within J_E11 showed that although genomes
174 from the Solomon Islands were phylogenetically distinct, those from Lihir Island, Papua New Guinea⁸
175 co-occurred in the same sub-lineages as those in this study, suggesting the two islands share a *T.*
176 *pertenu* population. Moreover, although all *T. pertenu* strains harbouring macrolide resistance SNPs
177 were classified as J_E11 by MLST, the increased resolution afforded by WGS demonstrates that the sub-
178 lineage that developed macrolide resistance through a 23S A2059G mutation in Lihir (sub-lineage
179 ELIM_Exp_03 in Figure 1) was phylogenetically distinct from the sub-lineage which developed
180 resistance through the A2058G mutation in Namatanai (PERS_02).

181



182
183

184 Figure 1: Maximum likelihood whole genome sequencing phylogeny for samples collected in
185 Namatanai and contextual genomes. A heatmap indicates the sample type, location from which samples
186 were collected, MLST type, presence of macrolide-resistance conferring mutations 23S rRNA A2058G
187 and A2059G, study arm, study round, and sub-lineage as determined by rPinecone.

188

189 ***T. pertenuis* genetic diversity was reduced after mass drug administration but rebounded within**
190 **a year**

191 We examined temporal patterns of *T. pertenuis* genomic diversity, and observed a marked decline in
192 Shannon's diversity index in both study arms following the first round of MDA (from 2.9 to 0 in the
193 control arm, and from 2.8 to 1.1 in the experimental arm; appendix 1 p8). Despite targeted treatment of
194 clinical yaws cases, diversity in the control arm rebounded to near pre-MDA levels (2.7) within 12-18
195 months, in contrast to the experimental arm, where diversity remained low (1.0). To further delineate
196 these patterns, we examined temporal changes in frequency for the 29 sub-lineages comprising two or
197 more samples. Of these, ten sub-lineages observed at baseline were eliminated entirely during the
198 study. A total of sixteen sub-lineages were detected at baseline and persisted after MDA, in the control
199 arm (thirteen sub-lineages), the experimental arm (one sub-lineage) or both arms (two sub-lineages)
200 (Figure 2a). Three sub-lineages were not detected at baseline but were observed after MDA in the
201 control arm.

202

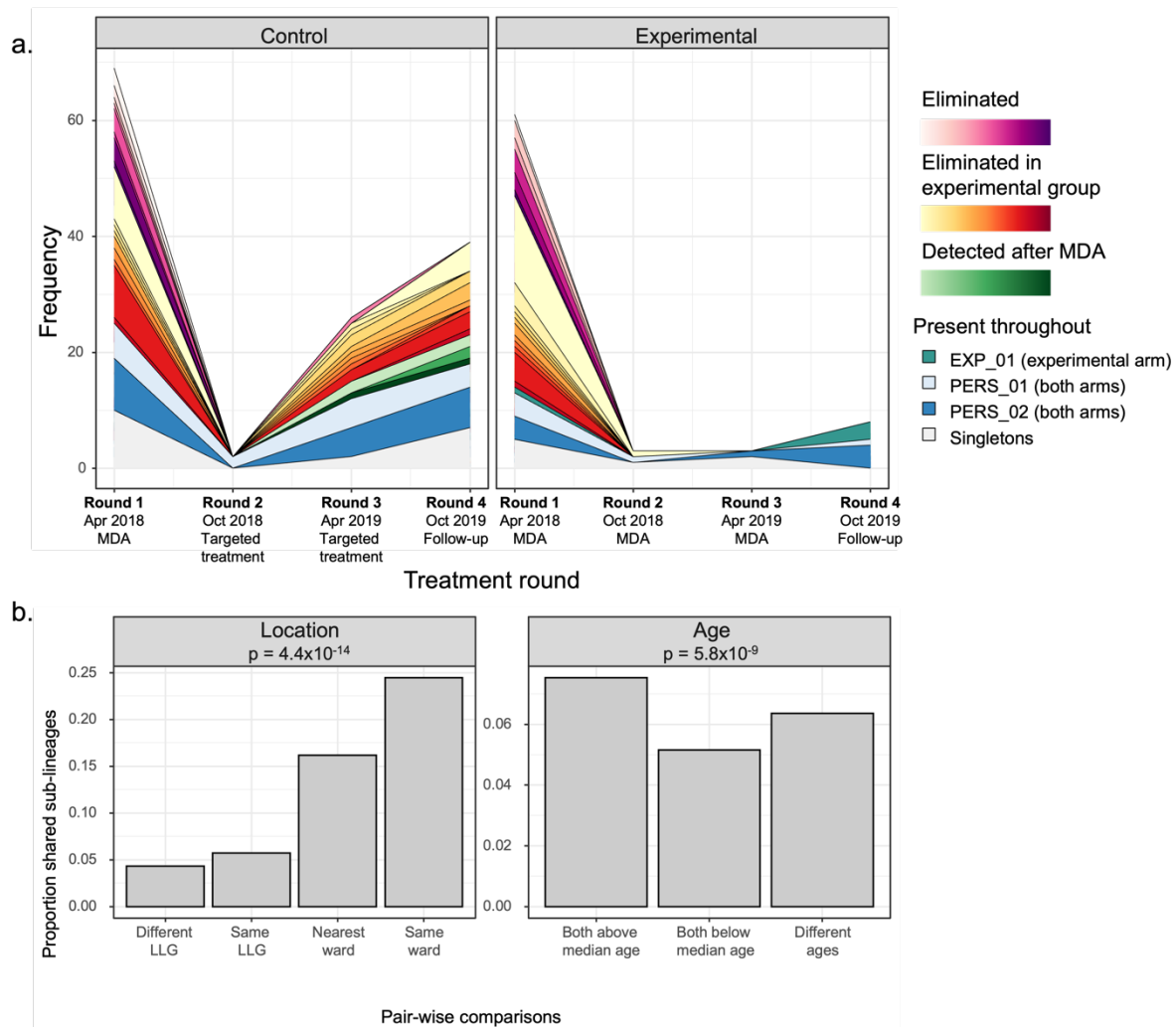
203 Next, we investigated whether recrudescence in the experimental arm was driven by individuals missed
204 during MDA, transmission from the control arm, or importation. At the end of the study, eight samples
205 representing three sub-lineages were present in the experimental group, of which two (PERS_01 and
206 PERS_02) had been detected in both arms throughout. The single PERS_01 sample left in the
207 experimental arm was both spatially and temporally closer to PERS_01 samples from the control arm
208 compared with the experimental arm (appendix 1 p9). Three of the four PERS_02 samples had
209 associated location metadata, and all three were in wards where PERS_02 had previously been detected
210 (appendix 1 p10). However, they were also less than 10km away from control wards containing

211 PERS_02 in the final round. Therefore, it was not possible to conclude whether PERS_02 persistence
212 was driven by transmission from the control arm or within the experimental arm. EXP_01, which was
213 only present in the experimental arm during the study, was detected in two wards in the final round,
214 including the same ward it was originally detected in at baseline (appendix 1 p11). None of the sub-
215 lineages present in the experimental arm in the final round had previously been found outside of the
216 Namatanai District.

217
218 We next used pairwise comparisons to investigate geographical clustering of sub-lineages and found
219 that pairs of samples from the same local-level government (LLG) were 1.3-fold more likely to share
220 the same sub-lineage compared to pairs from different LLGs (Figure 2b). Likewise, pairs of samples
221 from the same ward were 5.6-fold more likely to share the same sub-lineage compared to pairs from
222 different LLGs, whilst pairs from neighbouring wards were 3.7-fold more likely. Notably in both the
223 Namatanai and Sentral Niu Ailan LLGs, MDA had a dramatic effect on *T. pertenuis* diversity, followed
224 by a rebound in the control arm (appendix 1 p12). In contrast, *T. pertenuis* diversity in the Matalai LLG
225 was relatively low at the start of the study and remained relatively unaffected by MDA or targeted
226 treatment, with the same two persistent sub-lineages dominating at the end of the study. This was still
227 the case when the Namatanai and Sentral Niu Ailan LLG's sample size was adjusted to match that of
228 Matalai LLG, suggesting this pattern is not wholly attributed to the smaller population of the Matalai
229 LLG (appendix 1 p13).

230
231 Finally, we examined whether demographic factors were risk factors for transmission. Sex was not
232 significantly associated with sharing sub-lineages (appendix 1 p14; $p = 1$, chi-squared test). Since the
233 age of participants was unevenly distributed, we split participants into two equally sized groups, and
234 those above the median age for this study (aged 9-56 years) were more likely to share the same sub-
235 lineage with one another ($p = 6 \times 10^{-9}$, chi-squared test) compared to samples from younger participants
236 (aged 1-8 years).

237



238
239
240
241
242
243
244
245
246

Figure 2: a. Frequency of different sub-lineages each treatment round, stratified by study arm. Sub-lineages that were eliminated entirely by MDA are shown in pink; those that were only eliminated in the experimental arm in orange; those that emerged in 2019 in green; those that were persistent throughout the study in blue, and singletons in grey.

b. Proportion of pair-wise comparisons for which samples shared the same sub-lineage, stratified by sample location, study arm and age. Indicated p values are from a chi-squared test, Bonferroni-adjusted to account for each sample being compared to every other sample.

247 **Two persistent sub-lineages are characterised by non-synonymous mutations in penicillin-binding proteins**

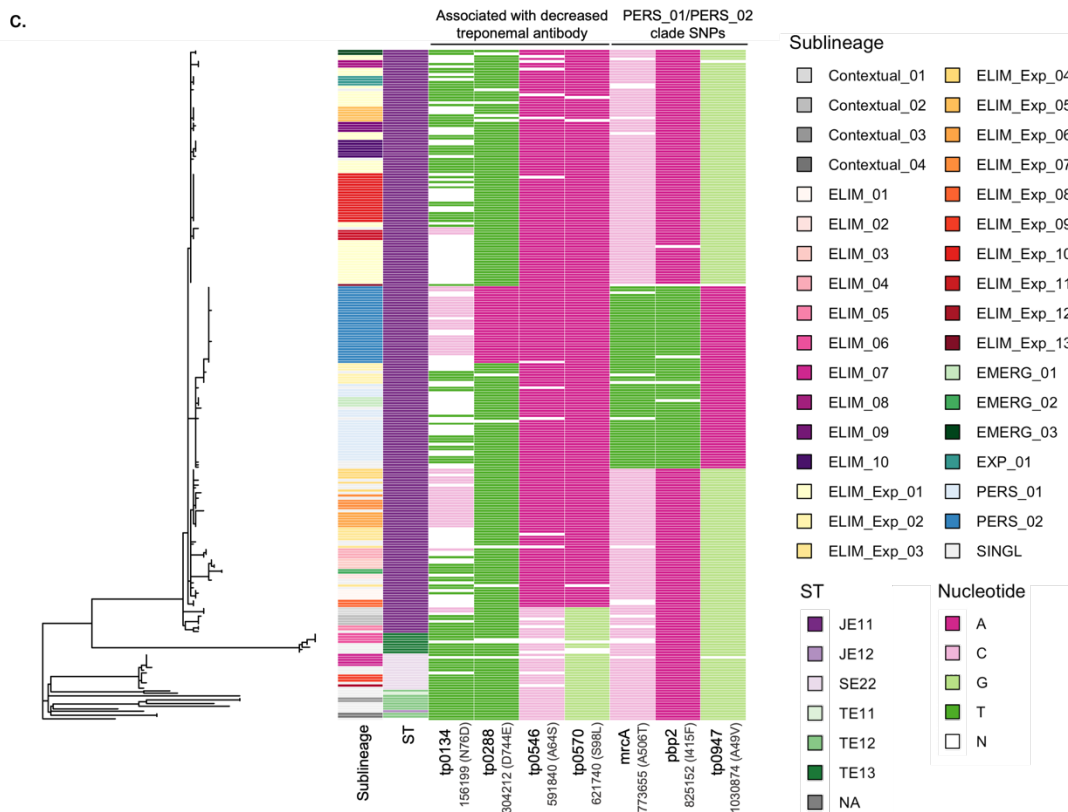
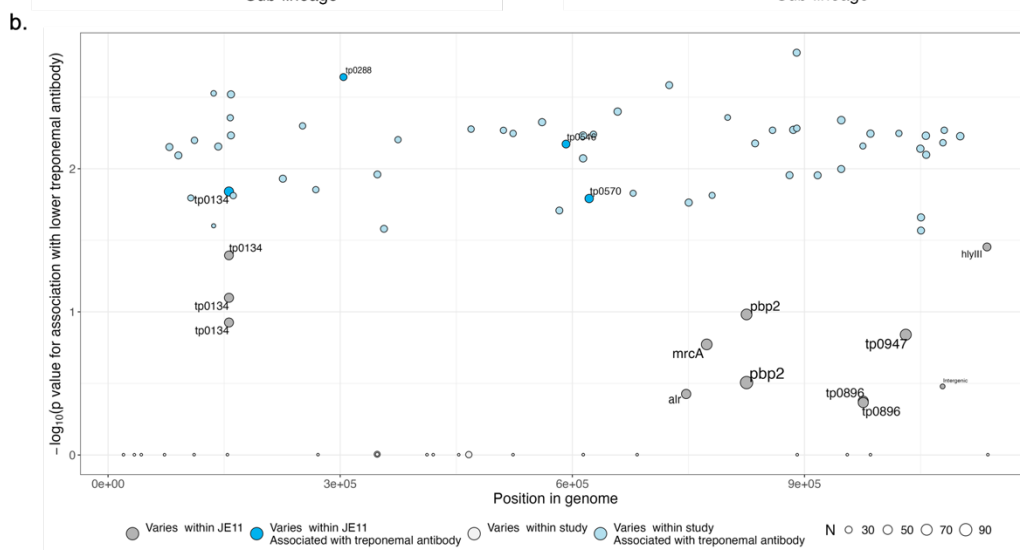
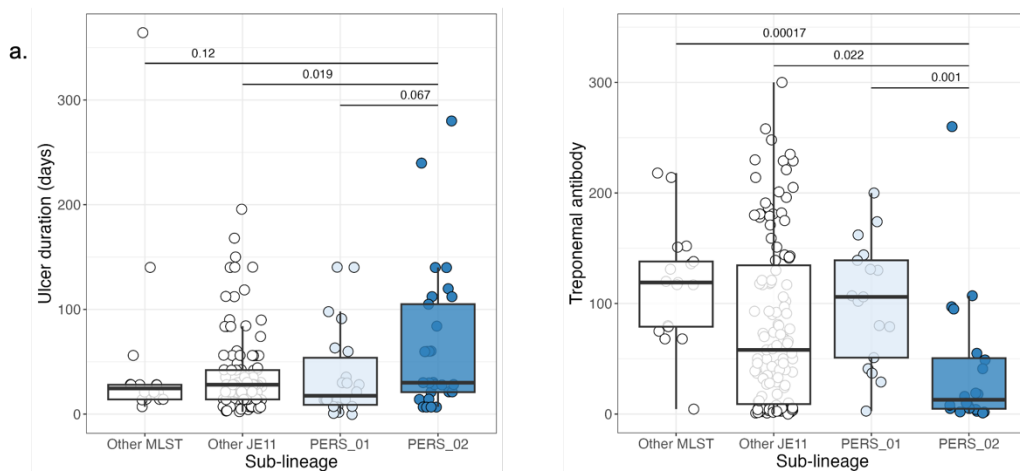
249 As described earlier, two phylogenetically related sub-lineages (PERS_01 and PERS_02) persisted throughout interventions in both arms, suggesting they might possess a fitness advantage. We looked for phenotypic differences that may explain this and found that participants infected with PERS_02 reported a longer ulcer duration and had lower optical densities for a treponemal antibody assay compared with other sub-lineages (appendix 1 p19; Figure 3a). There was no significant difference in non-treponemal antibody (appendix 1 p15). We then tested whether any SNPs, including those highlighted as part of potential recombination blocks, were associated with these phenotypes. Across 222 samples from the Namatanai study, 88 SNPs were found in at least 5% of genomes (appendix 1 pp19-21). While none of these SNPs were associated with a longer ulcer duration, 55 SNPs in 44 genes were associated with decreased treponemal antibody (one-sided Mann-Whitney test, false discovery rate (fdr) after Benjamin-Hochberg multiple-testing correction < 0.05 ; Figure 3b). Amongst these 44

260 genes, proteins with signal peptides (15/44, 34%) were over-represented relative to the rest of the
261 genome (142/1024, 14%) ($p = 0.0008$, Fisher's test).

262
263 Within the J_E11 MLST, 16 SNPs were present in at least 5% of genomes and occurred along the
264 ancestral phylogenetic branches differentiating J_E11 sub-lineages from one another. Four of these SNPs
265 were associated with decreased treponemal antibody (one-sided Mann-Whitney test, $fdr < 0.05$; Figures
266 3b and c), including a PERS_02-specific SNP causing a D744E mutation within the methyltransferase
267 domain of cytoplasmic protein *tp0288*, which is involved with cell envelope biogenesis (appendix 1
268 p16). *Tp0134*, a paralogue of the adhesin *tp0136*¹³, varied at three sites within J_E11 (G61S, D74E and
269 N76D), of which N76D was associated with lower treponemal antibody, and ¹³differed between
270 PERS_02 and PERS_01. Using the fine-scale whole-genome phylogeny previously generated in Beale
271 et al. 2021¹⁴ (appendix 1 p17) we found that in *Treponema pallidum* subspecies *pallidum*, the causative
272 agent of syphilis, *tp0134* varied at a site within the same loop (I78T) as *T. pertenuae*.

273
274 Of the 16 SNPs which differentiated J_E11 sub-lineages, three were specific to the clade encompassing
275 PERS_01 and PERS_02 (Figure 3d). This included two non-synonymous mutations in penicillin
276 binding proteins: I415F in the transpeptidase domain of *pbp2* (*tp0760*), and A506T in the transpeptidase
277 domain of *mrcA* (*tp0705*) (appendix 1 p18). Both sub-lineages also possessed a A49V mutation in the
278 cyclophilin-type peptidyl-prolyl cis-trans isomerase *tp0947*, an enzyme that catalyses an essential rate-
279 limiting step in protein folding. We examined whether these SNPs were also present in *T. pallidum*
280 (appendix 1 p17). I415F in *pbp2* was present in seven *T. pallidum* SS14-lineage samples in China, while
281 A506T *mrcA* was globally distributed, having arisen in multiple clades within the *T. pallidum* Nichols
282 (71/96 samples) and SS14 (30/407 samples) lineages. The two Nichols clades with A506T *mrcA* were
283 the only clades within the Nichols lineage to develop macrolide resistance. Another variant at this site,
284 A506V, was present in the globally distributed SS14 lineage (23/407). However, it is not known
285 whether these SNPs confer an observable phenotype.

286



288 Figure 3: a. Optical densities for a treponemal antibody assay and reported ulcer duration in days in
289 participants infected with PERS_01, PERS_02 and other sub-lineages. P values for a Mann-Whitney
290 test comparing PERS_02 with each other group is indicated.

291 b. Association of SNPs with lowered treponemal antibody compared with their position in the genome.
292 SNPs with a minor allele frequency > 5% were included. SNPs are coloured by whether they were
293 significantly associated with lowered treponemal antibody (one-tailed Mann-Whitney test, false
294 discovery rate < 0.05) and whether they varied within the J_E11 MLST. The size of each point reflects
295 the frequency of the SNP.

296 c. SNPs associated with lowered treponemal antibody that varied within J_E11, and SNPs that were
297 specific to the clade encompassing PERS_01 and PERS_02. For samples where no reads mapped to the
298 SNP locus, the nucleotide is denoted as “N”.

299

300 Discussion

301

302 Understanding sources of yaws transmission will be key for informing intervention strategies if WHO
303 is to achieve the target of yaws eradication by 2030¹⁵. In this study we used whole-genome sequencing
304 to elucidate the dynamics and transmission of yaws in the Namatanai district following two possible
305 MDA strategies. We observed that a single round of MDA only temporarily suppressed yaws diversity,
306 followed by recrudescence mostly driven by sub-lineages that had been previously detected in the study
307 population. Notably, two sub-lineages persisted in both arms of the study, of which one was associated
308 with lower treponemal antibody and went on to develop three cases of macrolide resistance.

309

310 Inferring transmission of *T. pallidum* is challenging due to its high level of genome conservation,
311 estimated to diverge by only one SNP every 6-8 years¹⁴. While MLST provides a useful tool for
312 discriminating highly divergent genetic groups of *T. pallidum* at a global level¹⁶, it cannot distinguish
313 sub-lineages at sufficient resolution to elucidate dynamics at a regional level. Previous research showed
314 that *T. pertenuis* in Namatanai was dominated by the J_E11 ST, particularly following the first MDA
315 round^{9,10}. Here, the increased resolution afforded by whole-genome sequencing allowed us to subdivide
316 samples considered identical by MLST into smaller sub-lineages with a distance of only 0-2 SNPs
317 within each sub-lineage. Notably, we used the core genome here, and resolving and appropriately
318 modelling incorporation of hypervariable regions using long-read or deep sequencing^{17,18} could enable
319 further delineation of recent transmission. To maximise the value of our dataset we included genomes
320 with up to 50% missing sites, reflecting the challenge of sequencing low-abundance pathogens. While
321 it is possible that we missed SNPs at low quality sites that would further delineate sub-lineages, because
322 our analysis accounts for missing sites along ancestral branches we do not expect this to substantially
323 impact our conclusions.

324

325 We found that yaws diversity dropped after the first round of MDA, remaining low in the experimental
326 arm following two further MDA rounds at six-month intervals, but rebounding over the same period in
327 the control arm. Previously it was observed that after the first round of MDA, multiple *T. pertenuis*
328 MLST types were still present in the control arm, indicating that a single MDA does not represent a
329 strong bottleneck¹⁰. Here we found that not only were multiple MLST groups present after MDA, but
330 also multiple J_E11 sub-lineages. 12/15 of these sub-lineages had been detected in the district before
331 MDA, indicating that recrudescence was likely driven by a substantial number of cases missed during
332 the initial round of MDA using the approach advocated by the Morges strategy. It is not yet clear
333 whether diversity in the experimental group would return to similarly high levels 18 months after MDA
334 cessation. However, repeated MDA successfully suppressed the *T. pertenuis* population, which

335 modelling suggests increases the likelihood of stochastic “fadeout”, where prevalence is low enough
336 that stochastic effects can result in spontaneous eradication¹⁹.

337

338 Defining closely related but distinct sub-lineages also allowed us to quantify sub-lineage sharing
339 between regions and sub-groups. Given the relatively small study region (covering a distance of
340 approximately 300km) and slow genetic substitution rate of *T. pertenuae*, we might have expected to
341 find strains fully mixed throughout the region. However we observed substantial geographic clustering,
342 which may be attributed to restricted travel options in these remote rural regions. This highly localised
343 transmission suggests that yaws elimination could be more attainable than anticipated, as targeted
344 treatment could be intensified in regions with clusters of cases, and clustering increases the likelihood
345 of stochastic fadeout¹⁹. Spatial heterogeneity within implementation units will need to be considered
346 when predicting prevalence. Furthermore, we observed sub-lineage sharing between neighbouring
347 wards, indicating that implementation units should encompass at-least sub-districts in spatial scale. We
348 also found that children aged nine or older, were more likely to share sub-lineages, which may reflect
349 transmission chains within primary schools.

350

351 Finally, we investigated factors contributing to the success of the two sub-lineages that not only
352 persisted but expanded in both study arms following MDA. These sub-lineages were both common at
353 the beginning of the study, and harboured I415F and A506T mutations in penicillin-binding proteins
354 *pbp2* and *mrcA* respectively. Analysing previous data, we found that the A506T *mrcA* variant also arose
355 in both major syphilis lineages, while the I415F *pbp2* variant was present in syphilis samples from
356 China²⁰. To our knowledge this is the first time these variants have been reported in yaws. While it is
357 possible these variants arose and expanded in both *T. pallidum* subspecies *pallidum* and *pertenuae* due
358 to random genetic drift, a more likely explanation is positive selection for these mutations. It is unclear
359 whether they confer increased resistance to sub-therapeutic doses of penicillin, or some other fitness
360 advantage, but if so, this may explain why these two sub-lineages were common at the beginning of the
361 trial. A case-control study is currently underway to investigate whether these variants are associated
362 with treatment failure in syphilis²¹. Another possibility is that because these proteins are associated
363 with cell envelope biogenesis, mutations could reduce growth rate *in vivo*. Not only could this extend
364 the latency period, allowing *T. pertenuae* to avoid targeted antibiotic treatment, but it could also enhance
365 tolerance to antibiotics that interfere with growth²². Future confirmation of this association could inform
366 decisions on whether to intensify treatment in regions where persistent sub-lineages are circulating.

367

368 Ultimately, only three cases in one of these two sub-lineages developed macrolide resistance over the
369 course of the study (PERS_02). Those infected with this sub-lineage exhibited a longer ulcer duration
370 and lower treponemal antibody levels. Proteins with signal peptides were over-represented amongst
371 those with variants associated with reduced antibody, and PERS_02 possessed variants in membrane
372 lipoprotein *tp0134*, a paralogue of the adhesin *tp0136*¹³. As periplasmic proteins are thought to be *T.*
373 *pallidum*'s principal immunogens after killing by phagocytes²³, variation in these proteins could prevent
374 recognition. PERS_02 also harboured a mutation in a cytoplasmic protein involved in cell envelope
375 biogenesis, which could slow bacterial growth and prevent *T. pallidum* reaching the critical capacity to
376 activate immunity²³. Previous studies have found that pathogens exhibit enhanced evolutionary rates
377 within hosts compared to rates inferred at the population level^{24–26}, suggesting that prolonged infections
378 may increase the likelihood of *de novo* mutations such as those causing macrolide resistance. Prolonged
379 infections also increase the likelihood of exposure to sub-therapeutic doses of azithromycin, creating a
380 selection pressure for resistance.

381

382 Our findings here and elsewhere indicate that repeated MDA is more effective in reducing yaws cases
383 than the single round MDA advocated in the Morges strategy⁹. However, this must be balanced against
384 increased cost of conducting additional rounds of MDA. We observed marked geographic clustering
385 even within a small area, suggesting that yaws elimination through targeting stochastic fade-out could
386 be achievable. Furthermore, increased strain sharing between older children suggests that future
387 campaigns targeting schools could offer a cost-effective solution for interrupting transmission. While
388 three rounds of MDA did not achieve total elimination, repeated MDA did prevent resurgence in the
389 diversity of yaws, and a strategy of long-term recurrent MDA on all neighbouring wards could be
390 effective in maintaining prevalence at a sufficiently low level to achieve elimination. However, the
391 potential selection pressure that repeated azithromycin MDA imposes towards macrolide resistance,
392 both on *T. pertenuis* and the microbiota, represents a reason for caution. Surveillance for macrolide
393 resistance and treatment failure will be necessary for eradication. It is likely that a multipronged strategy
394 involving MDA, targeted penicillin treatment and preventative interventions will be necessary to finally
395 achieve eradication.

396

397 **Contributors**

398 AB and MAB carried out the data analysis. AB prepared the figures and wrote the original draft of the
399 manuscript. LNJ and WH carried out the initial clinical study with support from CG-B and OM. PP and
400 DS carried out laboratory work. AB, CG-B, MAB, NRT, OM, LG and MM interpreted the data. All
401 authors reviewed and approved the final manuscript.

402

403 **Data sharing**

404 Raw whole-genome sequencing data is deposited in the European Nucleotide Archive. Accession
405 numbers and further data included in this analysis are provided in appendix 2.

406

407 **Declaration of interests**

408 The authors have no conflicts of interest to declare.

409

410 **Acknowledgements**

411 The cluster randomised trial was funded by Fundació “la Caixa”, Diputació de Barcelona, and Fundació
412 Barberà Solidària. AB, MAB and NRT were supported by Wellcome funding to the Sanger Institute
413 (220540/Z/20/A), and OM by the European Research Council under Grant Agreement number 850450.
414 The work was partially financed by funds from the National Institute of Virology and Bacteriology
415 project [Programme EXCELES, ID Project No. LX22NPO5103] funded by the European Union - Next
416 Generation EU to DS.

417

418 The authors acknowledge support from the Sequencing teams, and Olivier Seret and the Parasites and
419 Microbes Informatics team at the Wellcome Sanger Institute. This research was funded in whole, or in
420 part, by the Wellcome Trust (220540/Z/20/A). For the purpose of open access, the authors have applied
421 a CC-BY public copyright licence to any author-accepted manuscript version arising from this
422 submission.

423

424

425 References

426

427 1 Mitjà O, Asiedu K, Mabey D. Yaws. *The Lancet* 2013; **381**: 763–73.

- 428 2 World Health Organization. Yaws (Endemic treponematoses).
429 [https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/yaws-](https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/yaws-(endemic-treponematoses))
430 [-\(endemic-treponematoses\)](https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/yaws-(endemic-treponematoses)).
- 431 3 Fitzpatrick C, Asiedu K, Jannin J. Where the Road Ends, Yaws Begins? The Cost-
432 effectiveness of Eradication versus More Roads. *PLoS Negl Trop Dis* 2014; **8**.
433 DOI:10.1371/journal.pntd.0003165.
- 434 4 Maurice J. WHO plans new yaws eradication campaign. Elsevier Ltd, 2012
435 DOI:10.1016/S0140-6736(12)60581-9.
- 436 5 Mitjà O, Hays R, Ipai A, *et al*. Single-dose azithromycin versus benzathine
437 benzylpenicillin for treatment of yaws in children in Papua New Guinea: An open-
438 label, non-inferiority, randomised trial. *The Lancet* 2012; **379**: 342–7.
- 439 6 World Health Organization. Summary report of a consultation on the eradication of
440 yaws. 2012.
- 441 7 Mitjà O, Godornes C, Houinei W, *et al*. Re-emergence of yaws after single mass
442 azithromycin treatment followed by targeted treatment: a longitudinal study. *The*
443 *Lancet* 2018; **391**: 1599–607.
- 444 8 Beale MA, Noguera-Julian M, Godornes C, *et al*. Yaws re-emergence and bacterial
445 drug resistance selection after mass administration of azithromycin: a genomic
446 epidemiology investigation. *Lancet Microbe* 2020; **1**: e263–71.
- 447 9 John LN, Beiras CG, Houinei W, *et al*. Trial of Three Rounds of Mass Azithromycin
448 Administration for Yaws Eradication. *New England Journal of Medicine* 2022; **386**: 47–
449 56.
- 450 10 Medappa M, Pospíšilová P, Madruga MPM, *et al*. Low genetic diversity of *Treponema*
451 *pallidum* ssp. *pertenuis* (TPE) isolated from patients' ulcers in Namatanai District of
452 Papua New Guinea: Local human population is infected by three TPE genotypes. *PLoS*
453 *Negl Trop Dis* 2024; **18**: e0011831.
- 454 11 Wailan AM, Coll F, Heinz E, *et al*. rPinecone: Define sub-lineages of a clonal expansion
455 via a phylogenetic tree. *Microb Genom* 2019; **5**: 1–9.
- 456 12 Hunt M, Mather AE, Sánchez-Busó L, *et al*. ARIBA: Rapid antimicrobial resistance
457 genotyping directly from sequencing reads. *Microb Genom* 2017; **3**: 1–11.
- 458 13 Djokic V, Giacani L, Parveen N. Analysis of host cell binding specificity mediated by the
459 tp0136 adhesin of the syphilis agent *treponema pallidum* subsp. *Pallidum*. *PLoS Negl*
460 *Trop Dis* 2019; **13**. DOI:10.1371/journal.pntd.0007401.
- 461 14 Beale MA, Marks M, Cole MJ, *et al*. Global phylogeny of *Treponema pallidum* lineages
462 reveals recent expansion and spread of contemporary syphilis. *Nat Microbiol* 2021; **6**:
463 1549–60.
- 464 15 World Health Organization. Ending the neglect to attain the sustainable development
465 goals: a road map for neglected tropical diseases 2021–2030. 2020.
- 466 16 Grillova L, Jolley K, Šmajš D, Picardeau M. A public database for the new MLST
467 scheme for *Treponema pallidum* subsp. *Pallidum*: Surveillance and epidemiology of
468 the causative agent of syphilis. *PeerJ* 2019; **2019**: 1–13.
- 469 17 Addetia A, Lin MJ, Phung Q, *et al*. Estimation of Full-Length *tprK* Diversity in
470 *Treponema pallidum* subsp. *pallidum*. 2020. DOI:10.1128/mBio.
- 471 18 Lieberman NAP, Avendaño CC, Mohamed Bakhsh SAK, *et al*. Genomic Epidemiology
472 of *Treponema pallidum* and Circulation of Strains With Diminished *tprK* Antigen
473 Variation Capability in Seattle, 2021–2022. *Journal of Infectious Diseases* 2024; **229**:
474 866–75.

- 475 19 Mooring EQ, Marks M, Mitjà O, *et al.* Programmatic goals and spatial epidemiology
476 influence the merit of targeted versus of population-wide interventions for yaws
477 eradication. *bioRxiv* 2019. DOI:10.1101/640326.
- 478 20 Sun J, Meng Z, Wu K, *et al.* Tracing the origin of *Treponema pallidum* in China using
479 nextgeneration sequencing. *Oncotarget* 2016; **7**: 42904–18.
- 480 21 Mi HF, Shen X, Chen XQ, Zhang XL, Ke WJ, Xiao Y. Association between treatment
481 failure in patients with early syphilis and penicillin resistance-related gene mutations
482 of *Treponema pallidum*: Protocol for a multicentre nested case–control study. *Front*
483 *Med (Lausanne)* 2023; **10**. DOI:10.3389/fmed.2023.1131921.
- 484 22 Kaldalu N, Tenson T. Slow growth causes bacterial persistence. *Sci Signal* 2019; **12**.
485 DOI:10.1126/scisignal.aay1167.
- 486 23 Radolf JD, Deka RK, Anand A, Šmajs D, Norgard M V., Yang XF. *Treponema pallidum*,
487 the syphilis spirochete: Making a living as a stealth pathogen. *Nat Rev Microbiol*.
488 2016; **14**: 744–59.
- 489 24 Culyba MJ, van Tyne D. Bacterial evolution during human infection: Adapt and live or
490 adapt and die. *PLoS Pathog* 2021; **17**: 1–19.
- 491 25 Chaguza C, Hahn AM, Petrone ME, *et al.* Accelerated SARS-CoV-2 intrahost evolution
492 leading to distinct genotypes during chronic infection. *Cell Rep Med* 2023; **4**.
493 DOI:10.1016/j.xcrm.2023.100943.
- 494 26 Eldholm V, Norheim G, von der Lippe B, *et al.* Evolution of extensively drug-resistant
495 *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome*
496 *Biol* 2014; **15**: 490.
497
498