

1 Unravelling *Chlamydia trachomatis* Diversity in

2 Amhara, Ethiopia: MLVA-*ompA* Sequencing as a

3 Molecular Typing Tool for Trachoma

4 Anna Harte¹, Ehsan Ghasemian¹, Harry Pickering¹, Joanna Houghton¹, Ambahun Chernet², Eshetu
5 Sata², Gizachew Yismaw³, Taye Zeru³, Zerihun Tadesse², E. Kelly Callahan⁴, Scott D. Nash⁴, Martin
6 Holland¹.

7
8 ¹The London School of Hygiene and Tropical Medicine, London, UK

9 ²The Carter Center, Addis Ababa, Ethiopia

10 ³The Amhara Regional Health Bureau, Bahir Dar, Ethiopia

11 ⁴The Carter Center, Atlanta, GA, USA

12

13 Abstract

14 Trachoma is the leading infectious cause of blindness worldwide and is now largely confined
15 to around 40 low- and middle-income countries. It is caused by *Chlamydia trachomatis* (Ct),
16 a contagious intracellular bacterium. The World Health Organization recommends mass
17 drug administration (MDA) with azithromycin for treatment and control of ocular Ct
18 infections. To understand the molecular epidemiology of trachoma, especially in the context
19 of MDA and transmission dynamics, the identification of Ct genotypes is a necessity. While
20 many studies have used the Ct major outer membrane protein (*ompA*) for genotyping, it has
21 limitations.

22 Our study applies a novel typing system, Multiple Loci Variable Number Tandem Repeat
23 Analysis combined with *ompA* (MLVA-*ompA*). Ocular swabs were collected post-MDA from
24 four trachoma-endemic zones in Ethiopia between 2011-2017. DNA from 300 children with
25 high Ct polymerase chain reaction (PCR) loads was typed using MLVA-*ompA*, utilizing three
26 variable number tandem repeat (VNTR) loci within the Ct genome.

27

28 Results show that MLVA-*ompA* exhibited high discriminatory power (0.981) surpassing the
29 recommended threshold for epidemiological studies. We identified 87 MLVA-*ompA* variants
30 across 26 districts. No significant associations were found between variants and clinical signs
31 or chlamydial load. Notably, overall Ct diversity significantly decreased after additional MDA
32 rounds, with a higher proportion of serovar A post-MDA.

33

34 Despite challenges in sequencing one VNTR locus (CT1299), MLVA-*ompA* demonstrated
35 cost-effectiveness and efficiency relative to whole genome sequencing, providing valuable
36 information for trachoma control programs on local epidemiology. The findings suggest the
37 potential of MLVA-*ompA* as a reliable tool for typing ocular Ct and understanding
38 transmission dynamics, aiding in the development of targeted interventions for trachoma
39 control.

40

41 **Author Summary**

42 Trachoma is the leading infectious cause of blindness worldwide and is largely confined to
43 low- and middle-income countries. It is caused by *Chlamydia trachomatis* (Ct), a contagious
44 intracellular bacterium. The World Health Organization recommends mass drug
45 administration (MDA) with the antibiotic azithromycin for treatment of ocular Ct infections.

46 In most regions MDA is successfully reducing trachoma prevalence to the point where it is
47 no longer a public health issue, however in some places trachoma persists despite multiple
48 rounds of treatment. To investigate why trachoma persists, especially in the context of MDA
49 and transmission dynamics, the identification of Ct genotypes is necessary. Our study
50 applies a novel Ct typing system, which augments the standard method by adding three loci
51 with high mutation rates. Results show that the novel typing system was able to
52 discriminate between variants with greater resolution than the standard method, and was
53 both cost-effective and more efficient relative to the gold-standard of whole genome
54 sequencing. The findings suggest that this novel method is a reliable tool for typing ocular
55 Ct, which can aid in the development of targeted interventions for trachoma control.

56

57 **Introduction**

58

59 Trachoma is the leading infectious cause of blindness worldwide (1) and is found primarily in
60 about 40 low- and middle-income countries (LMIC). It is caused by *Chlamydia trachomatis*
61 (Ct), a contagious intracellular bacterium with 4 serotypes (A, B, Ba and C) that are typically
62 responsible for ocular infection in trachoma endemic countries. The World Health
63 Organization (WHO) recommends the use of the Surgery, Antibiotics, Facial Cleanliness, and
64 Environmental improvement (SAFE) strategy, including mass drug administration (MDA) of
65 the antibiotic azithromycin for controlling trachoma (1). Multiple rounds of MDA are usually
66 required which is determined by the prevalence of clinical signs within the district.
67 Reinfection can occur within months of treatment (2–4), and the clinical signs of disease
68 often persist in the population while ocular Ct infection is low or absent (5,6). The
69 surveillance of Ct strains is of interest for monitoring transmission dynamics, as well as

70 identifying the potential dominance of individual strains, which may indicate selection by
71 antibiotics or increased virulence, persistent infection and/or re-infection (7).

72

73 Molecular epidemiological studies of Ct infection within communities often focus on the
74 outer membrane protein A (*ompA*) gene, as this codes for the polymorphic major outer
75 membrane protein (MOMP), a key chlamydial antigen targeted by the host immune system
76 and one therefore theoretically subject to immune selection (8–11). However, previous
77 work on genital (D-K) and Lymphogranuloma venereum (LGV) serovars has demonstrated
78 that *ompA* genotyping alone has insufficient discriminatory power (12–15) for the reliable
79 identification of truly different variants within a population (16). Since recombination within
80 *ompA* has been demonstrated, this may also obscure true phylogenetic relationships (17).
81 Whole-Genome Sequencing (WGS) should offer the highest resolution for variant
82 identification; however, it remains prohibitive in terms of cost, equipment, high technical
83 expertise and large sample load required, placing this beyond the access and feasibility for
84 trachoma control programmes operating in LMIC. To address this, several polymerase chain
85 reaction (PCR)-amplicon sequence-based typing methods have been developed, where two
86 or more loci are amplified and then subjected to chain termination sequencing (Sanger
87 sequencing) for each individual sample.

88

89 These typing systems have been used to investigate Ct strain diversity, each with varying
90 levels of discriminatory power (7). Multi-locus sequence typing (MLST) and multiple loci
91 variable number of tandem repeat analysis (MLVA) are two types of PCR sequencing that
92 use sequence types (ST) to define different variants. Classic MLST is focused on sequence
93 variation in loci that are not considered to be under selection pressure and are stable

94 housekeeping genes (18,19) and is best suited to long-term and global epidemiology. Other
95 MLST studies have used highly variable loci instead, to improve variant resolution on a
96 smaller epidemiological scale (20–22). MLVA provides a further alternative and uses
97 variable number tandem repeats (VNTR), which have a higher rate of mutations (23) and
98 can therefore provide a useful identifier of novel strains for small scale, local
99 epidemiological studies (22). MLST methods have been applied to ocular serovars and in
100 trachoma epidemiology (20,24,25), however MLVA has not yet been used to investigate the
101 molecular epidemiology of trachoma.

102

103 In this study conjunctival DNA collected from children living in trachoma-endemic areas of
104 Amhara, Ethiopia, which were surveyed between 2011 and 2017, were used to sequence
105 three VNTR loci and *ompA* as outlined by Pedersen et al., (2008). The aim of this study was
106 to determine the efficacy of MLVA-*ompA* when applied to ocular samples and test the
107 discriminatory power (DP). We then investigated whether MLVA-*ompA* variants were
108 associated with clinical signs of trachoma and the impact of MDA on variants in the
109 population. We propose that MLVA-*ompA* typing could serve as a tool for the surveillance of
110 emerging Ct variants within the context of localized epidemiological studies. Additionally, it
111 can be employed to establish a profile of the evolutionary dynamics of Ct variants over time
112 within populations remaining endemic to trachoma.

113

114 **Methods**

115

116 **Study population**

117

118 The SAFE strategy was scaled up between 2007 and 2010 to reach all districts in Amhara.
119 Between 2011 and 2015 all districts were then surveyed to measure the impact of
120 approximately 5 years of SAFE interventions on trachoma prevalence (26,27). Impact
121 surveys use multi-stage sampling, and trained and certified graders use the WHO-simplified
122 grading system (28) to determine the presence of the trachoma clinical signs trachomatous
123 inflammation-follicular (TF), trachomatous inflammation-intense (TI) and trachomatous
124 scarring (TS) (27). As part of these surveys, conjunctival swabs were collected to estimate
125 population-based prevalence of Ct infection among children aged 1 to 5 years (26). Impact
126 surveys, including conjunctival swabbing, were repeated throughout the region between
127 2014 and 2021 after an additional 3 to 5 years of interventions (29,30). The samples used in
128 this study were a subset of those collected between 2011-2015 from 58 surveyed districts
129 which make up four zones (North Gondar, South Gondar, East Gojam and Waghemra) (29).
130 An additional subset of samples used in this study came from all 11 districts in South Gondar
131 zone, collected between 2014 and 2017 as part of the second round of surveys. Each
132 individual was given a unique ID code which was not known to anyone outside of the
133 research group.

134

135 **Ethical approval**

136

137 The ethical approvals associated with the collection and processing of these samples have
138 been described previously (27,29,31). Survey methods were reviewed and approved by the
139 Emory University Institutional Review Board (protocol 079–2006) as well as by the Amhara
140 Regional Health Bureau. Due to elevated levels of illiteracy within the community,
141 permission was granted for obtaining verbal consent or assent. Electronic recording of oral

142 consent or assent was implemented for all individuals, aligning with the principles outlined
143 in the Declaration of Helsinki. Participants had the option to conclude the examination at
144 any juncture without the need for providing an explanation.

145

146 **DNA extraction and sample preparation for sequencing**

147

148 The initial testing of samples was performed at the Amhara Public Health Institute in Bahir
149 Dar, Ethiopia. Swabs were randomised and pooled into batches of five samples per pool,
150 and tested for two highly conserved Ct plasmid targets using the Abbott Realtime assay on
151 the Abbott m2000 (Abbott Molecular Inc., Des Plaines, IL, USA) (26). For the subset of
152 districts included in this study, individual samples from positive pools were assayed again to
153 determine individual level infection. Ct load was determined by converting the Abbott
154 m2000 delta cycle to elementary body (EB) count using an EB standard curve, generated
155 using a standard set of EB titrations (29,32). There were 525 Ct positive samples collected
156 from these study districts. 300 samples with the highest Ct load were chosen for further
157 sequencing analysis. DNA was re-extracted from all 300 conjunctival swabs as previously
158 described (31). 99 samples with a sufficiently high Ct load were chosen for WGS, the results
159 of which have already been published, including the bioinformatic extraction of the *ompA*
160 region (31). For the remaining 201 samples, *ompA* was sequenced by Sanger sequencing,
161 however for the three VNTRs of interest, CT1291, CT1299 and CT1335, all 300 samples were
162 sequenced by Sanger sequencing Reference sequences from serovars A (A/2497: Acc.
163 FM872306; A/HAR13: NC_007429), B (B/Jali20/OT: Acc. FM872308; B/Tunis864:
164 ERR12253485), C (C/TW-3, Acc. NC_023060), D (D/UW3/CX, Acc. NC_000117), and L

165 (L3/404/LN: Acc. HE601955; L2/434/Bu: Acc. AM884176) were also sequenced for the three
166 VNTR regions to compare between published WGS sequences.

167

168 ***ompA* Sanger sequencing preparation**

169

170 Prior to sequencing, the extracted eye swab DNA underwent one or two rounds of PCR
171 using a nested PCR, following the procedure outlined in Andreasen et al. 2008, (11). For the
172 first round, 20 μ L of 5prime MasterMix (QuantaBio), 20 μ L of molecular grade water
173 (Corning, Manassas, VA20109 USA), 2.5 μ L each of forward and reverse primers *ompA*-87
174 and *ompA*-1163 and 5 μ L of DNA template were run using the cycling conditions as follows:
175 Initial denaturation of 2 min at 94 $^{\circ}$ C, then 35 cycles of 94 $^{\circ}$ C for 15 S and 62 $^{\circ}$ C for 75 S, and
176 a final elongation step of 72 $^{\circ}$ C for 10 min. This PCR resulted in a product of 1076 base pairs
177 (bp) in size. Samples were run on a gel, and if no band was present, nested PCR was
178 performed. For the second round of PCR, 10 μ L 5Prime HotMasterMix (QuantaBio), 1.25 μ L
179 *ompA*-87 and 1.25 μ L *ompA*-1059 primers and 11.5 μ L molecular grade water (Corning,
180 Manassas, VA20109 USA) were added to 1 μ L of a 1 in 200 dilution of the PCR product from
181 the first round. The final product size of the target sequence (*ompA*-inner) was 972 bp in
182 length. This product was then cleaned using a ratio of 0.8 X final volume of AMPure XP
183 magnetic beads (Beckman Coulter) following the manufacturer's instructions, quantified on
184 the Qubit 2.0 Fluorometer (Thermo Fisher) and diluted to a concentration of \sim 10 ng/ μ L.
185 Samples were sent to Source BioScience (Cambridge, UK) with *ompA*-97 forward primer
186 and/or *ompA*-1059 for samples sequenced in both directions.

187

188 **VNTR sequencing preparation**

189

190 All 300 samples were processed for the VNTRs, with a separate PCR reaction used to amplify
191 each VNTR region. DNA extracted from eight Ct reference samples were also sequenced for
192 comparison with existing genomic data. Samples were run in 25 µL reaction volumes, using
193 5 µL DNA template, 10 µL Accustart (Quantabio, Hilden, Germany), 1.25 µL each forward
194 and reverse primers, and 7.5 µL molecular grade water (Corning, Manassas, VA20109 USA).
195 Three different sources of primers were used, due to complications with the proximity of
196 some of the VNTR regions to the start of the sequencing, and the subsequent low-quality
197 bases associated with the first 30-50 bps. For primers sourced from Pedersen et al., (2008)
198 (33) and the new CT1299 forward primer, cycling conditions were 10 min held at 94 °C, then
199 40 cycles of 45 S at 94 °C, 20 S at 59 °C, 20 S at 72 °C, hold at 4 °C. For primers sourced from
200 Labiran et al., (34) the annealing temperature was 56 °C for CT1335 and 60 °C for CT1299.
201 Samples were sent to Source BioScience (Cambridge, UK) with 3.2 pmol of forward primer
202 for each VNTR. CT1299 was sequenced in both the forward and reverse direction for some
203 samples with low quality sequencing due to extended repeat base regions. Primers and
204 amplicon sizes for the three VNTR regions are listed in table 1. Samples that did not produce
205 sufficient sequencing quality to call the sequence type in the first round of sequencing for
206 CT1299 were re-run using a modified CT1299 forward primer to increase the distance
207 between the start of the sequence and the flanking and VNTR region (35). A total of 140
208 samples were tested using the new CT1299 forward primer.

209

210

211

212

213

214

215

216

217

218

219

220

221

222 **Table 1.** Forward and reverse primers used to amplify VNTR regions and *ompA* of *Chlamydia trachomatis* in
223 Ethiopian conjunctival samples collected between 2011-2017, for Sanger sequencing and the amplicon size
224 produced. VNTR: Variable number tandem repeat.

225

Target	Source	Forward primer	Reverse primer	Amplicon size
<i>ompA</i>	Andreasen et al., 2008(36)	TGAACCAAGCCTTATGATCGACGG- (<i>ompA</i> -87)	CGGAATTGTGCATTTACGTGAG- (<i>ompA</i> -1163)	1076
<i>ompA</i> -inner	Andreasen et al., 2008(36)	TGAACCAAGCCTTATGATCGACGG- (<i>ompA</i> -87)	GCAAGATTTTCTAGATTTTCATC- (<i>ompA</i> -1059)	972
CT1291	Pedersen et al., 2008(33)	GCCAAGAAAAACATGCTGGT	AGGATATTTCCCTCAGTTATTCG	225
CT1299	Labiran et al., 2012(34)	ATCGCTTAAGATTCTCGGAGG	AGGTTCTAGCTGAGCATGGG	342
CT1299 (new)	This study and Labiran et al., 2012(34)	GGAATTTCCATAGACGGTTGATA	AGGTTCTAGCTGAGCATGGG	381
CT1335	Labiran et al., 2012(34)	AAAGCGTCCTCTGGAAGGG	CCTTCTCCTAACAACTTACGC	208

226

227 **VNTR calling and MLVA-ompA**

228

229 Only VNTR sequences produced by Sanger sequencing were used in further analysis.
230 Sequences were aligned within each respective VNTR using MEGA-X (Version 11.0.13), and
231 the sequence type was recorded by manually counting the number of repeat base pairs as
232 well as the respective flanking regions, following the typing method outlined by Pedersen et
233 al., (33) and Wang et al., (37). Chromatograms were checked to ensure that base-calling was
234 accurate, and sequences were discarded from further analysis if a specific VNTR type could
235 not be assigned. References where WGS sequences were available on National Center for
236 Biotechnology Information (NCBI) from serovars A (A/2497; A/HAR13), B (B/Jali20/OT;
237 B/Tunis864), C (C/TW-3), D (D/UW3/CX), and L (L3/404/LN; L2/434/Bu) were used to
238 compare VNTR-amplicon Sanger sequencing against WGS derived VNTR sequence, alongside
239 comparison of published WGS data from a subset of samples from this study (31) by
240 calculating the kappa statistic.

241

242 ***ompA* serovar calling**

243

244 *ompA* sequences were extracted from WGS data generated from the 99 samples originally
245 analysed in Pickering et al., (31). All WGS sequences were aligned to *ompA* reference
246 sequences A/HAR13, B/Jali20 and C/TW-3 using bowtie2 (38), and samtools (39) was used to
247 extract the matching *ompA* sequences. The best alignments were selected based on the
248 minimum number of Ns present in the sequence. The 146 Sanger-derived *ompA* sequences,
249 were separated based on serovar, and aligned within each serovar group to a reference

250 sequence using MEGA-X; Serovar A (A/HAR13, Acc. DQ064279), serovar B (B/TW-5/OT, Acc.
251 M17342) and serovar Ba (VR-347-Ba, Acc. KP120856).

252

253 **Variant types and typeability**

254

255 Individual variant types were classified following Pedersen et al., (33), Wang et al., (37), and
256 Manning et al., (15) including both the VNTR and the associated flanking region. The
257 combination of all three VNTRs within a sample was used to assign MLVA type, with the
258 addition of each *ompA* type to create the MLVA-*ompA* variant. Variants are technically the
259 same as strains, however a variant is only referred to as a strain when it shows distinct
260 physical properties, such as the reference strains used for alignments. As we have not
261 investigated any potential physical properties of the variants identified in this study only
262 variant has been used. Typeability was calculated by dividing the number of samples that
263 were successfully sequenced for all three VNTRs and *ompA* (MLVA-*ompA*) by the total
264 number of samples available (300).

265

266 **Diversity and discriminatory power**

267

268 DP is defined as the ability to distinguish different variants, and is expressed as the
269 probability that two different variants will be placed into different categories. The level of
270 diversity was calculated for each genotyping system with and without associated VNTR-
271 flanking regions (individual VNTRs, MLVA, *ompA*, MLVA- *ompA*) using Simpson's index of
272 diversity following the recommended modification outlined by Hunter and Gaston (1988)
273 (40). Hutcheson's *t*-test was used to compare the Simpson diversity indices for pre- and

274 post-MDA time points for the South Gondar zone. Data was available from two different
275 time points for South Gondar because surveys were performed after 5 rounds of MDA and
276 8-10 rounds of MDA. Zonal diversity was compared between the two time points using
277 Hutcheson's *t*-test in the *vegan* package (41) in R, to determine if there was a significant
278 change in the level of diversity after extra rounds of MDA.

279

280 **Association analyses**

281

282 Association analyses were performed to investigate the relationship between individual
283 level TF and/or TI and the MLVA-*ompA* variants. Due to the high number of low frequency
284 variants, variants were stratified as; high (>4), medium (2-4) and low (1) frequency, and
285 treated as a categorical variable, with the high frequency group set as the reference. Mixed
286 effects logistic regression (package *lme4*, R, (42)) was performed for each clinical sign, with
287 the three variant categories as the independent variable, adjusting for age, gender and
288 district as a random effect.

289

290 Association analyses were also performed using village-level TI and TF as the independent
291 variable using generalised linear mixed effects models (Package *glmm*, R) (43), adjusting for
292 age, gender and district as a random effect. The same model was run using Ct load as the
293 independent variable. District-level Ct, TI and TF prevalence were also used as dependent
294 variables to test the relationship with district diversity, accounting for the number of
295 samples per district.

296

297 **Minimum spanning tree and phylogenetic tree analysis**

298

299 Minimum spanning trees (MSTs) were created using Phyloviz 2.0 software (44). Variants
300 based on the Pedersen et al., (33) typing system, updated to include the flanking regions by
301 Wang et al., (37), were used to create MSTs of all available variants, overlaid with data on
302 serovar, year of collection, and zone. For the phylogenetic tree, the three VNTR sequences
303 and flanking regions resulted from Sanger sequencing were concatenated in geneious
304 (Version 2023.2.1) for each sample. For sequences resulted from WGS, the three VNTR
305 sequences and flanking regions were extracted and concatenated for each sample.
306 Sequence alignments were generated using MAFFT (Version v7.490) (47,48) with a 200
307 PAM/K = 2 scoring matrix (45,46). PhyML was utilized to estimate maximum likelihood
308 phylogenies of aligned sequences with a GTR model of evolution and 100 bootstraps (47).

309 **Results**

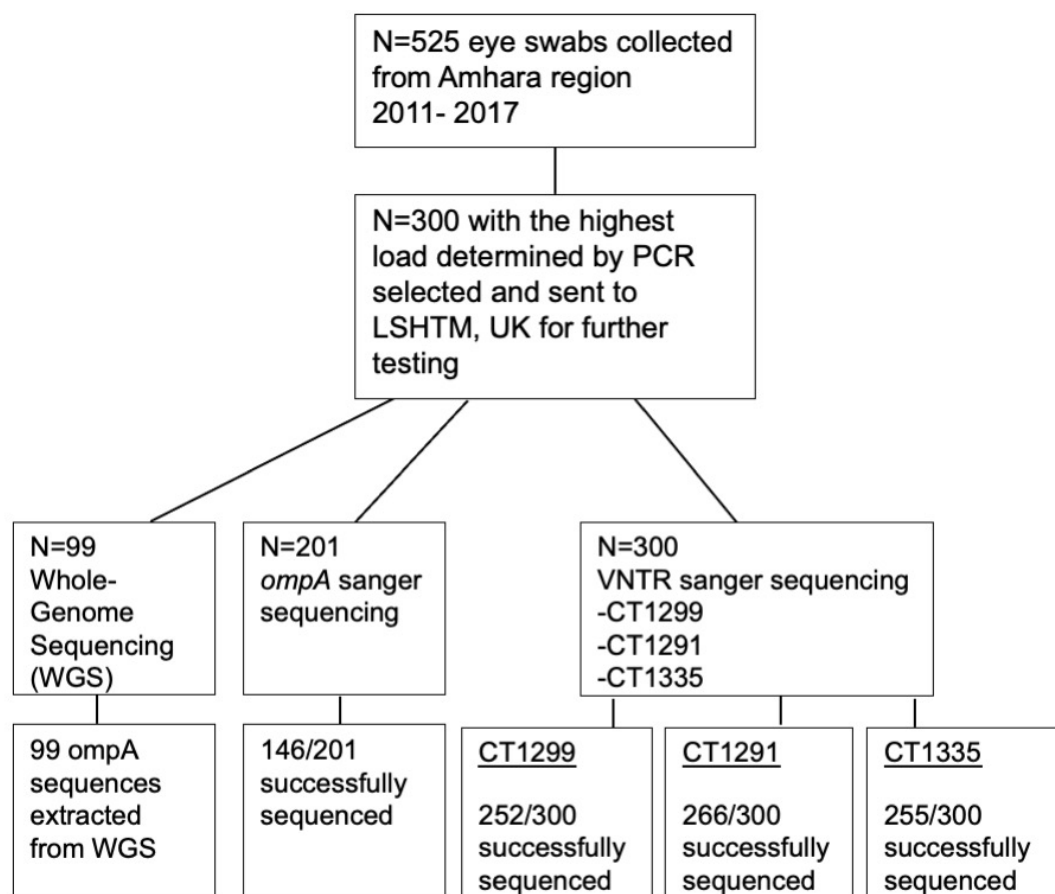
310

311 ***ompA* and VNTR Sequences**

312

313 *OmpA* serovar type derived from WGS was obtained for all 99 Ct WGS. A further 201
314 samples that were sequenced for *ompA* using Sanger sequencing generated 146 successful
315 sequences (72.6%). The 55 unsuccessful sequences were due to unreadable chromatograms
316 containing a large proportion of Ns, due to either poor quality DNA or overlapping peaks.
317 These could potentially represent mixed infection with multiple variants however these
318 results cannot be reliably resolved and were excluded. Data was missing in relatively equal
319 proportion over the six years that samples were collected (supplementary table 1). For the
320 three VNTRs generated by Sanger sequencing, CT1291 was successfully sequenced in
321 266/300 samples (88.6%), CT1299 was successful in 252/300 (84%) and CT1335 was

322 successful in 255/300 (85%). In total, 194/300 (65%) samples had sequencing results of a
323 high enough quality to call serovar and VNTR sequence types for all three VNTRs (Fig. 1).
324 Sequencing of *ompA* and the three VNTRs produced variable results, with CT1291 the most
325 successful in terms of typeability (88.7%) and CT1299 the least (84.0%) (Table 2, Fig. 2).
326 There were consistent difficulties with sequencing the repeat region of CT1299, with
327 samples exceeding 14Cs demonstrating evidence of polymerase slippage and PCR stutter
328 (23). Samples with greater than 14Cs in a row were therefore sequenced in both the
329 forward and reverse directions, however 10 CT1299 samples with 15-18 repeat Cs were not
330 able to be confidently identified, and were removed from the dataset.

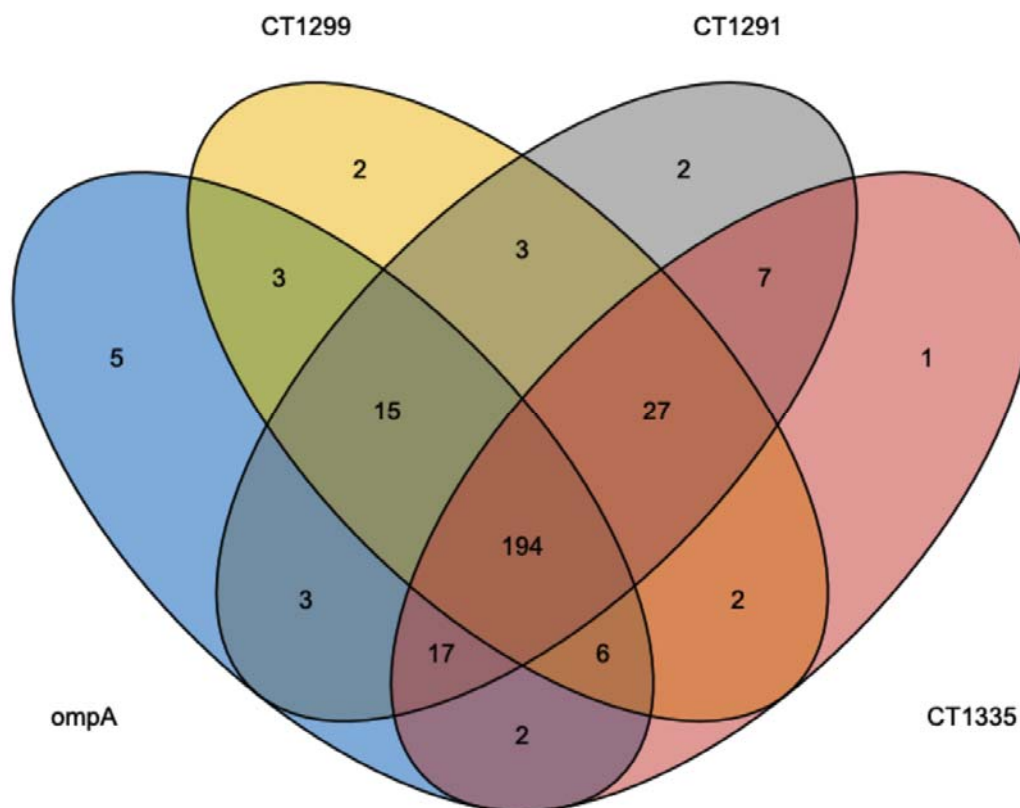


331

332 **Fig. 1.** Workflow diagram showing the process of conjunctival swab samples collected between
333 2011- 2017 from the Amhara region of Ethiopia, and the proportion successfully sequenced for each

334 VNTR and *ompA*. *OmpA*: outer membrane protein A. LSHTM: London School of Hygiene and Tropical

335 Medicine. VNTR: Variable number tandem repeat.



336

337 **Fig. 2.** Venn diagram illustrating the overlap in successful Sanger sequencing of each conjunctival

338 swab sample collected from Amhara, Ethiopia between 2011-2017 for the three VNTRs and *ompA*.

339 Nine samples were not able to be sequenced for any of the VNTRs or *ompA*. VNTR: Variable number

340 tandem repeat. *ompA*: outer membrane protein A.

341

342

343

344

345

346

347

348

349 **Table 2.** Typeability and discriminatory power for each VNTR separately, the VNTRs collectively
350 (MLVA), *ompA* and MLVA-*ompA*. Discriminatory power was determined following Hunter and
351 Gaston's modified Simpson's diversity index (Hunter and Gaston, 1987). Typeability is calculated by
352 the number of successful sequencing reactions divided by the total number of samples available. The
353 number of variants for the *ompA* sequencing refers to *ompA* serovars, as determined by homology
354 of each sequence within the BLAST-n NCBI database.

355

	CT1291	CT1299	CT1335	MLVA	<i>ompA</i> - genovar	<i>ompA</i> - variant type	MLVA- <i>ompA</i>
Total (300)	266	252	255	232	245	240	194
Typeability (%)	88.7	84.0	85.0	77.3	81.7	80.0	64.6
Number of sequence types	8	12	6	66	3	17	87
Discriminatory power	0.676	0.812	0.530	0.959	0.504	0.820	0.981

356

357

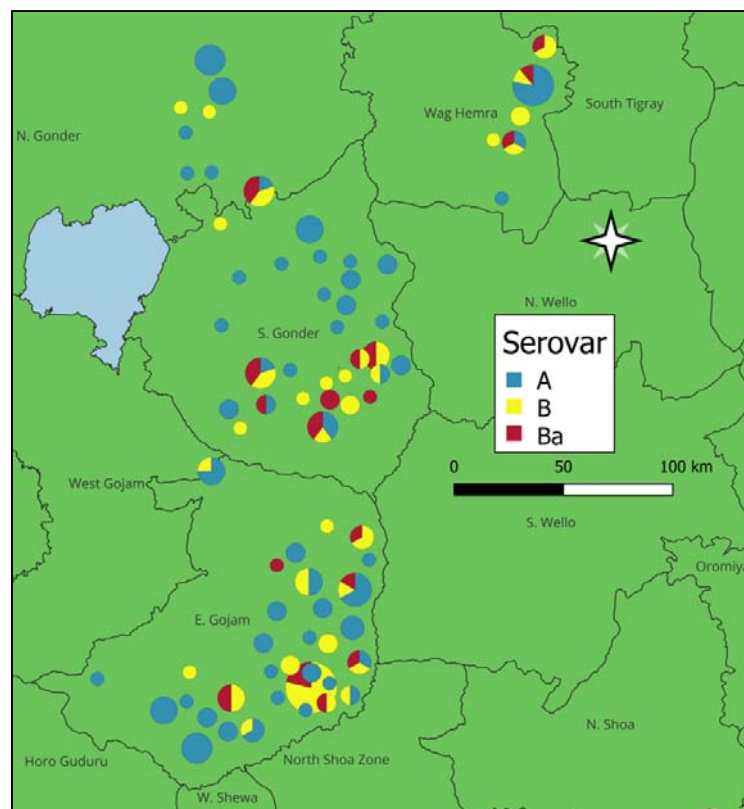
358 **VNTR and MLVA-*ompA* typeability**

359

360 The typeability of each VNTR and *ompA* individually were above 80%, however due to a
361 small proportion of dissimilarity between the samples that were not successful for each
362 round of sequencing (Fig. 2), the overall typeability of MLVA-*ompA* for this study was 64.6%.
363 There was a mix of serovars A, B and Ba across the entire Amhara region from samples
364 collected between 2011-2015 (Fig. 3). A total of 87 variants were found within this

365 population (DP=0.981). WGS sequences from reference serovar STs deposited in NCBI
366 matched our sample sequencing results for six out of eight reference genomes, with the
367 repeat C region in CT1299 resulting in a low quality read for C/TW-3 and a mismatch in
368 CT1299 region for B/Tunis864 (Table 3).

369



370

371 **Fig. 3.** Map of the Amhara region of Ethiopia showing the location and proportion of *Chlamydia*
372 *trachomatis* serovars A, B and Ba at the first sampling time points (2011 -2015) from conjunctival eye
373 swab samples collected after the first five rounds of mass drug administration. The size of the pie
374 chart is proportional to how many samples are represented, with larger pie charts indicating more
375 samples.

376

377

378

379

380

381 **Table 3.** *Chlamydia trachomatis* (Ct) reference strain sequencing results for each VNTR by Sanger
382 sequencing (“this study”) in comparison to VNTRs extracted from whole-genome sequences sourced
383 from NCBI (“reference”). Reference strain DNA was extracted from cells inoculated with bacteria
384 prior to the commencement of this study.

385

Serovar ID	CT1299		CT1291		CT1335		Reference Accession number
	This study	Reference	This study	Reference	This study	Reference	
L2/434/Bu	6C	6C	9C	9C	11T/8A	11T/8A	AM884176
B/Jali20	8C	8C	10C	10C	10T/8A	10T/8A	FM872308
B/Tunis864	13C	12C	7C	7C	10T/8A	10T/8A	ERR12253485
D/UW3/CX	12C	12C	10C	10C	10T/8A	10T/8A	NC_000117
C/TW-3	>17C	14C	9C	9C	10T/8A	10T/8A	NC_023060
A/HAR13	14C	14C	9C	9C	10T/8A	10T/8A	NC_007429
A/2497	12C	12C	8C	8C	10T/8A	10T/8A	FM872306
L3/404/LN	6C	6C	7C	7C	11T/8A	11T/8A	HE601955

386

387 CT1299 and CT1291 had the highest number of STs with 12 and 8 respectively, with CT1335
388 having 6 STs. Three new STs were recorded for CT1291, two new STs for CT1299 and four
389 new STs recorded in CT1335 (table 4). The frequency of the STs for each VNTR are listed in
390 Table 4a-c. Frequencies for both MLVA and MLVA-*ompA* show a relatively high proportion of
391 variant types are singletons since almost half of the MLVA-*ompA* variant types have a
392 frequency of one.

393

394

395 **Table 4.** VNTR sequence types (ST) and frequency table with Pedersen et al., (2008) and Wang et al.,
396 (2011) type codes for (a) CT1335, (b) CT1299 and (c) CT1291. Flanking regions forward (CT1291,
397 CT1335, CT1299) and after (CT1299) the repeat nucleotide regions are included. Novel STs are
398 identified with an asterisk *.

399 **A**

Pedersen VNTR STs	CT1335 forward flanking region and repeat sequence	Frequency
1a*	GAAAAGG- 9T/8A -GCTTTTGT	5
2	GAAAAAGG- 10T/7A -GCTTTTGT	17
2a*	GAAAAGG- 10T/7A -GCTTTTGT	11
3	GAAAAAGG- 10T/8A -GCTTTTGT	35
3a*	GAAAAGG- 10T/8A -GCTTTTGT	125
14*	GAAAAGG- 9T/7A -GCTTTTGT	1

400

401 **B**

Pedersen VNTR STs	CT1299 flanking region and repeat sequence	Frequency
2	TTATTCT- 8C -ATCAAA	1
3	TTATTCT- 9C -ATCAAA	19
4	TTATTCT- 10C -ATCAAA	64
4b*	TTATTCT- 10C - <u>T5C</u> -ATCAAA	1
5	TTATTCT- 11C -ATCAAA	36
6	TTATTCT- 12C -ATCAAA	11
7	TTATTCT- 13C -ATCAAA	21
8	TTATTCT- 14C -ATCAAA	40
8b*	TTATTCT- 14C - <u>T5C</u> -ATCAAA	1

402

403 **C**

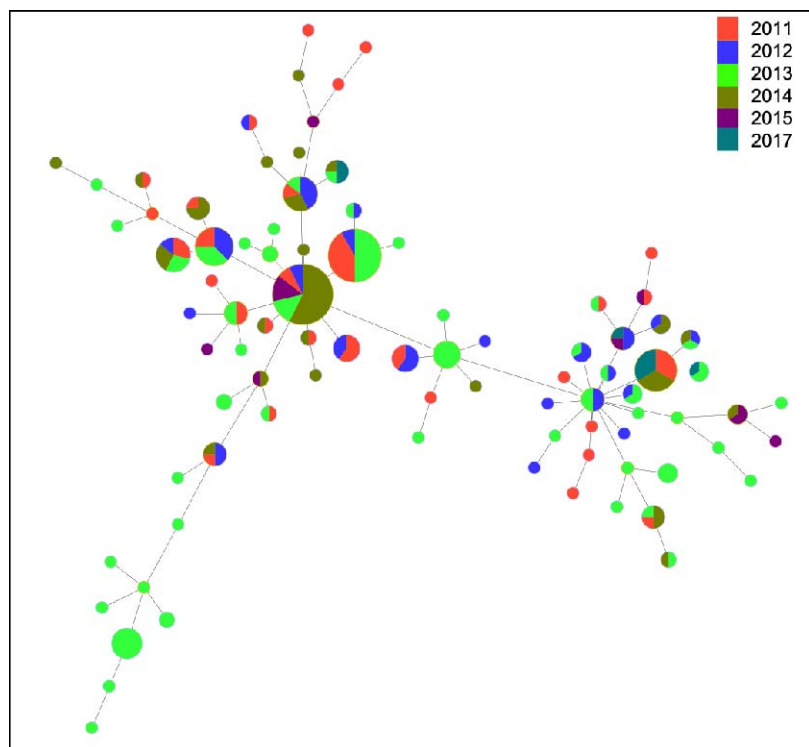
Pedersen VNTR STs	CT1291 forward flanking region and repeat sequence	Frequency
2	AAAATGGTCT- 8C -TATTG	28
2c*	AAAATGGT- 8C -TATTG	14
3	AAAATGGTCT- 9C -TATTG	99
3c*	AAAATGGT- 9C -TATTG	3
4	AAAATGGTCT- 10C -TATTG	38
5	AAAATGGTCT- 11C -TATTG	7
7	AAAATGGTCT- 12C -TATTG	1
8*	AAAATGGTCT- 7C -TATTG	4

404

405 **Minimum spanning trees**

406

407 The MLVA-*ompA* MST (Fig. 4) showed limited defined clustering when separated by year of
408 collection. There was a small cluster from 2013, however the most common variants,
409 represented by larger circles, were present in multiple years. There were also some defined
410 clusters by *ompA* serovar (supplementary Fig. 1), however the MSTs categorised by rounds
411 of MDA (supplementary Fig. 2) and zone (supplementary Fig. 3) showed limited clusters of
412 variants. There was a small (n= 10) defined cluster of variants collected in East Gojam in
413 2013 that were all *ompA* serovar B/Ba.



414

415

416 **Fig. 4.** A minimum spanning tree showing the sequence-types identified in the Amhara
417 region of Ethiopia between 2011-2017. Each MLVA-*ompA* sequence-type (n= 87) is coloured
418 by year of collection. Each circle represents a different sequence-type, with the size of circle
419 being directly proportional to the number of individuals who had that variant. Trees were
420 generated using PhyloViz 2.0.

421

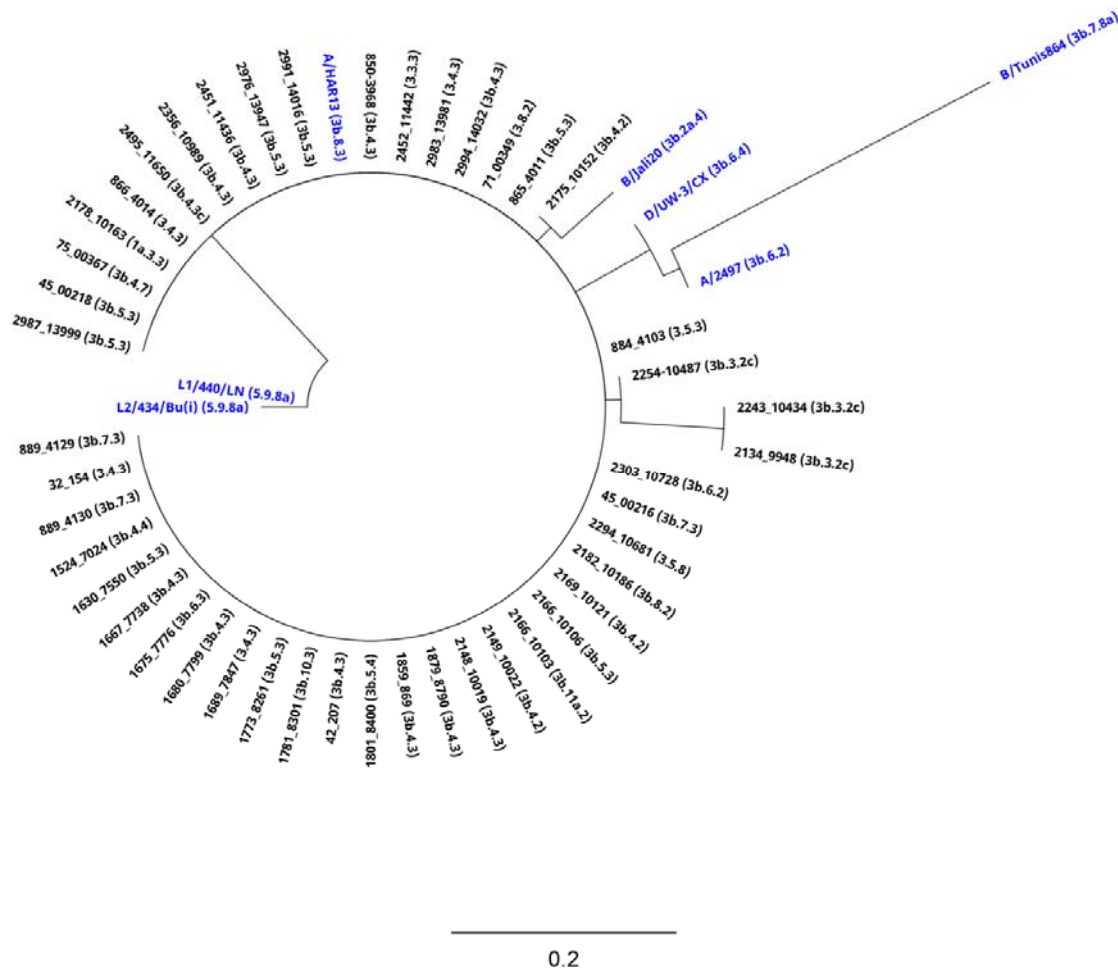
422 **Phylogenetic trees**

423

424 A total of 45 samples provided complete VNTR sequences along with their flanking regions
425 through WGS, which were subsequently utilized to construct a maximum likelihood
426 phylogenetic tree (Fig. 5). The remaining 44 samples that were unable to provide a complete
427 VNTR sequence were missing loci due to insufficient sequencing quality in either WGS or

428 sanger sequences or both (supplementary table 2). Simultaneously, the sequences obtained
429 from Sanger sequencing for these samples were employed for a parallel maximum
430 likelihood phylogenetic tree construction (Fig. 6). A comparative analysis between the tree
431 generated from Sanger sequencing and WGS data revealed a more intricate structure in the
432 Sanger data. In this context, Ethiopian sequences corresponding to each variant clustered
433 together, forming distinct branches separate from other variants. Conversely, the WGS-
434 derived sequences from Ethiopia exhibited a lack of differentiation based on assigned STs.
435 Notably, 41 out of the 45 sequences grouped closely with the reference strain A/HAR13,
436 suggesting a higher degree of homogeneity within this subset.

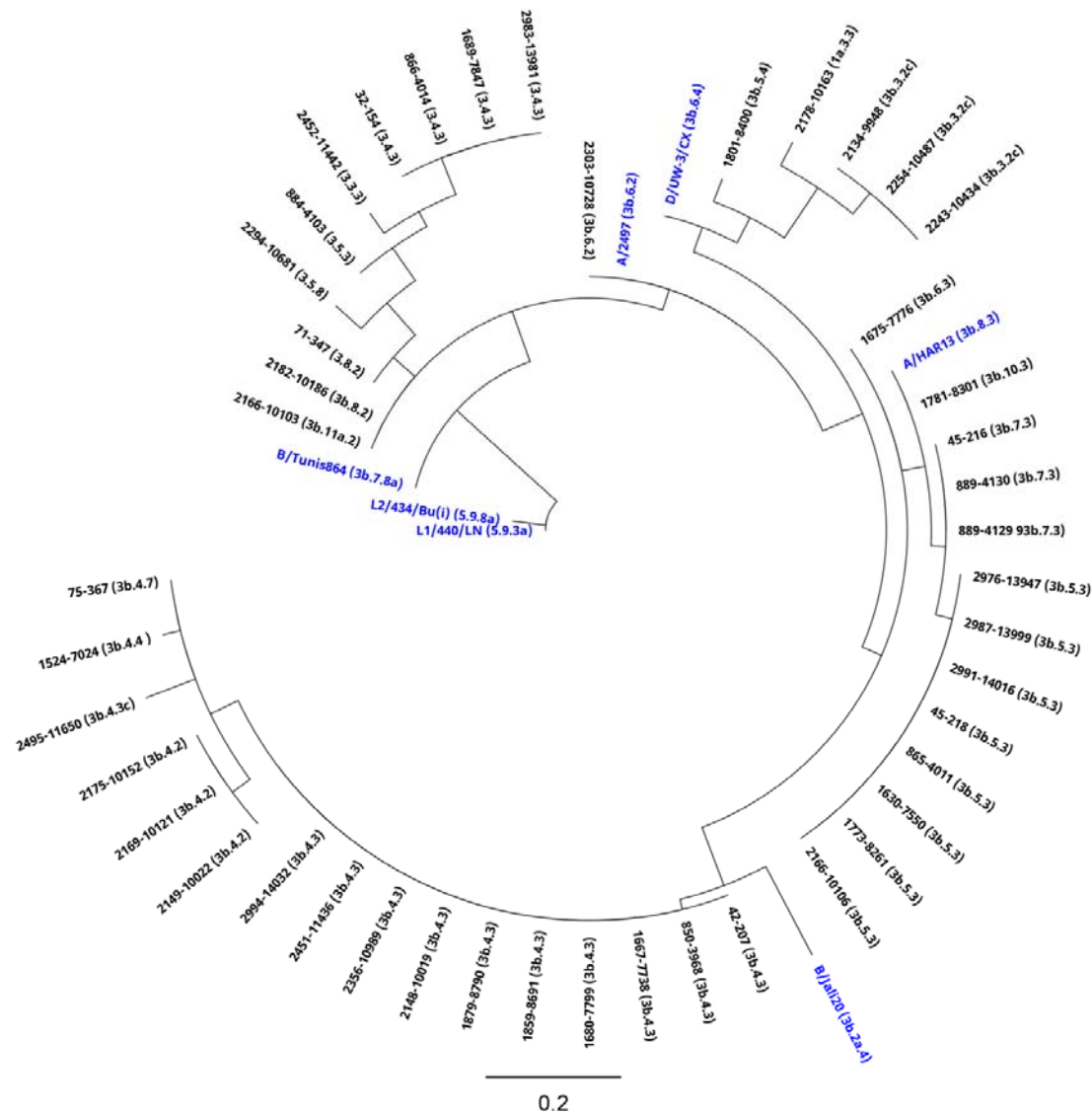
437



438

439 **Fig. 5.** Maximum likelihood phylogenetic reconstruction of multiple loci variable number of
440 tandem repeat analysis and flanking region loci bioinformatically extracted from Whole-
441 Genome Sequencing (WGS). The phylogenetic analysis encompasses three variable number
442 tandem repeat regions and associated flanking regions resulted from WGS of 45 *Chlamydia*
443 *trachomatis* (Ct) positive samples collected in Ethiopia and seven Ct reference strains.
444 Sequences were aligned using MAFFT, and a phylogenetic tree was reconstructed employing
445 PHYML (100 bootstraps). All reference strains are highlighted in blue.

446



447

448 **Fig. 6.** Maximum likelihood phylogenetic reconstruction of multiple loci variable number of
449 tandem repeat analysis and flanking regions loci-amplicons by Sanger sequencing. The
450 phylogenetic analysis encompasses three variable number tandem repeat regions and
451 associated flanking regions derived from Sanger sequencing of 45 *Chlamydia trachomatis*
452 (Ct) positive samples collected in Ethiopia and seven Ct reference strains. Sequences were
453 aligned using MAFFT, and a phylogenetic tree was reconstructed employing PHYML (100
454 bootstraps). All reference strains are highlighted in blue.

455

456 **WGS derived and PCR amplicon VNTR sequence comparisons**

457

458 The 99 samples that underwent WGS had fair consistency with Sanger sequencing results
459 for CT1335 (weighted kappa = 0.21), however poor consistency for CT1291 (weighted kappa
460 = 0.16). Only three CT1335 sequences were unable to be assigned using the WGS
461 sequences, which coincided with 3 samples where the VNTR type was 9T/8A, as opposed to
462 the most frequent 10T/8A. There was a discrepancy between the CT1335 flanking region
463 between WGS and Sanger sequences, with Sanger sequences identifying some samples
464 carrying the ST with a flanking sequence of four repeated Adenosine bases (GAAAAGG)
465 whereas WGS identified them as having five (GAAAAAGG) (supplementary table 2).

466

467 CT1291 sequences were unable to be assigned in 19 of the WGS sequences (19%), and were
468 inconsistent with the VNTR type called by Sanger sequencing in 45 individuals (45%). For the
469 CT1291 flanking region nine samples were unable to be called (9%), with three mismatches
470 between Sanger and WGS.

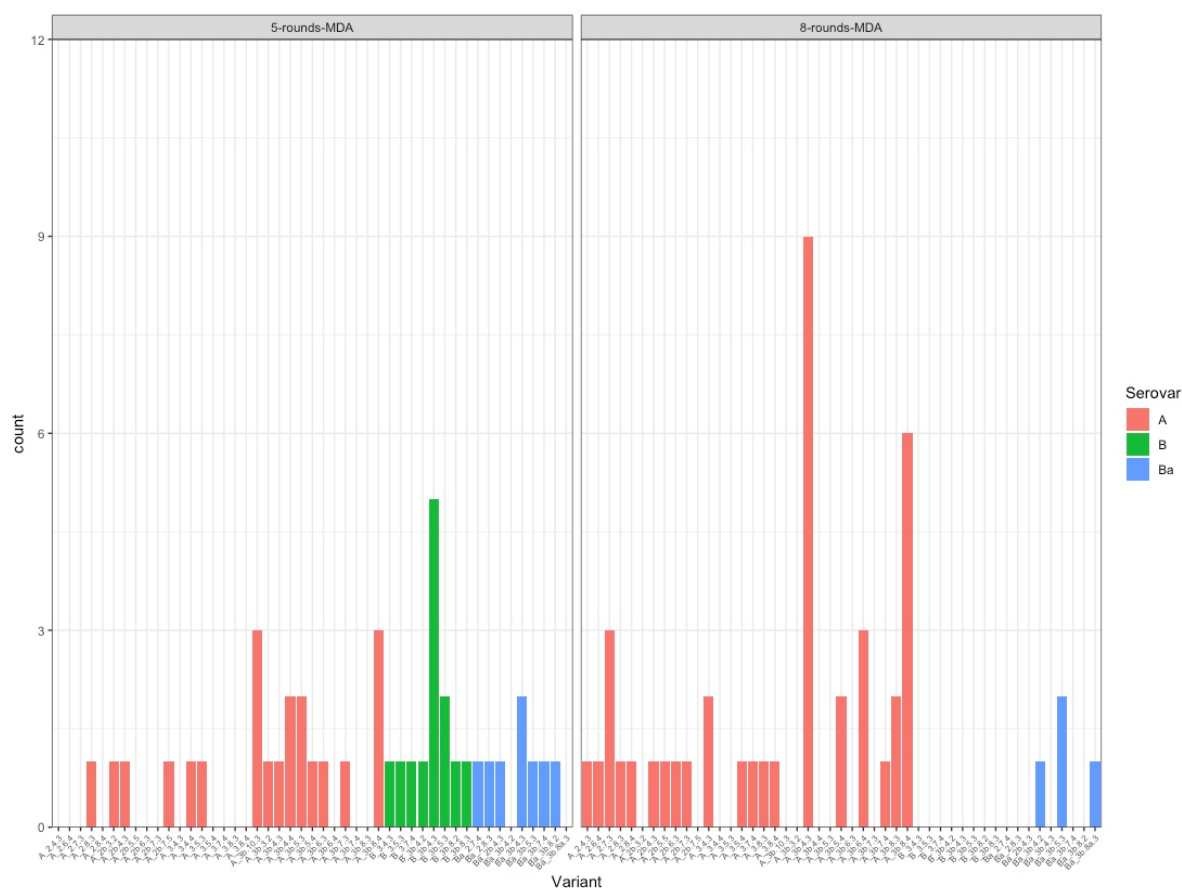
471

472 CT1299 VNTR sequences were unable to be called in 35 of the WGS sequences (35%), and in
473 the remaining 64 sequences, 61 were inconsistent with the VNTR type called by Sanger
474 sequencing (95%). The CT1299 flanking region did not show any variation in STs in Sanger
475 sequences, however in WGS sequences there were three sequences that had the flanking
476 sequence "ATTCT" repeated twice, which was not present in the Sanger sequence of the
477 same sample (supplementary table 2).

478

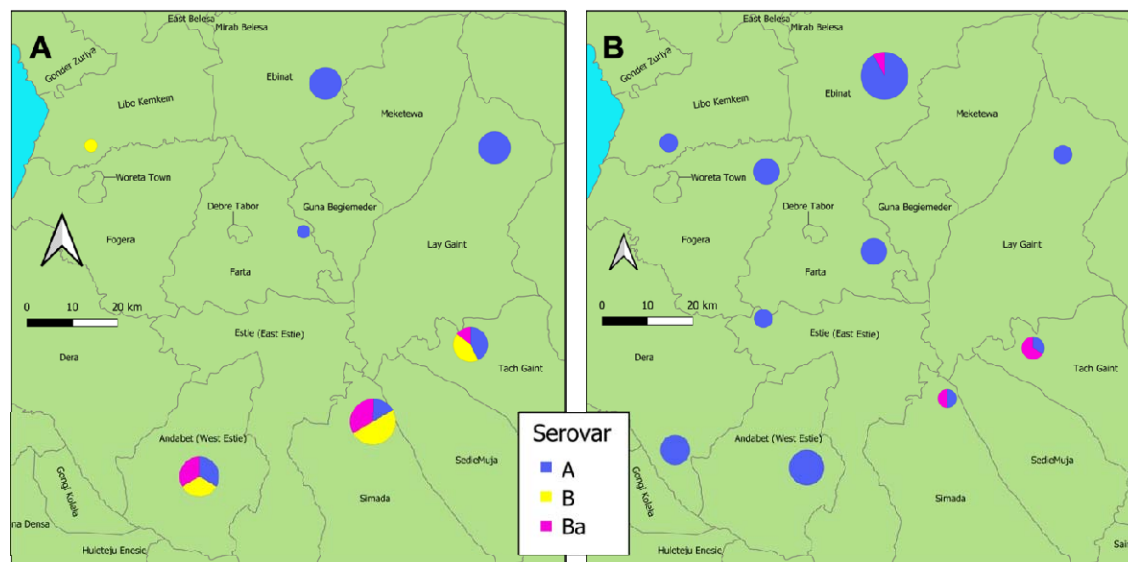
479 **Association analyses**

480 In the dataset available for analysis, 159 people had TF, 85 had TI, and 2 had TS. A total of
481 176/194 people (91%) had active trachoma (TF and/or TI). There were no significant
482 differences between the variant frequency groupings or serovar and either TI or TF clinical
483 signs, for both absence/presence analyses, village-level prevalence or Ct load (table 5).
484 There was a significant difference in MLVA-ompA diversity between 5 years of MDA and 8-
485 10 years of MDA in South Gondar ($p=0.006$), with less diversity after extra rounds of MDA
486 (Fig. 7). There was an observable ompA serovar switch over time (Fig. 8), with a higher
487 proportion of serovar A (91%) after the additional rounds of MDA (46%) (Fig. 7).



488
489 **Fig. 7.** Frequency plots for each *Chlamydia trachomatis* variant found in the South Gondar zone,
490 Ethiopia, 2011- 2017. The data has been split into two time points- the first, after five rounds of
491 MDA (n=43, total number of variants= 30), and the second after an additional three to five rounds of

492 MDA (n=44, total number of variants= 23). Bars are coloured by each respective serovar. MDA= mass
493 drug administration with azithromycin.
494



496 **Fig. 8:** Map of the South Gondar zone samples only (Amhara, Ethiopia), showing the range of
497 *Chlamydia trachomatis* serovars identified from conjunctival swabs collected between 2011-2014
498 after five rounds of MDA (A) and eight to ten rounds of MDA (B) collected between 2015-2017. The
499 size of the pie chart is proportional to how many samples are represented, with larger pie charts
500 indicating more samples.

501

502

503

504

505

506

507

508 **Table 5.** Association analyses results comparing the frequency category of the variant with the
509 absence or presence of TI or TF, the village level prevalence of TI or TF, and Ct load. All models were
510 adjusted for age and sex, with the “high” frequency group set as the reference level for all analyses.

511 TI= Trachomatous inflammation-intense, TF= Trachomatous inflammation-follicular.

512

Dependent variable	Variant frequency category	P value
Individual level TF and/or TI	High	Reference
	Medium	0.20
	Low	0.12
Village level TI	High	Reference
	Medium	0.90
	Low	0.91
Village level TF	High	Reference
	Medium	0.66
	Low	0.64
District level TI	High	Reference
	Medium	0.43
	Low	0.58
District level TF	High	Reference
	Medium	0.92
	Low	0.33
Individual Ct load	High	Reference
	Medium	0.48
	Low	0.37

513

514

515 Discussion

516

517 Understanding Ct transmission dynamics is an important part of trachoma control, as the
518 identification of Ct genotypes can provide information on the infection source, the presence
519 of repeat or persistent infections, and the impact of antibiotic treatment (48). Different
520 ocular genotypes of Ct may also elicit varying degrees of infection intensity and disease
521 severity (49). The majority of epidemiological studies have primarily focused on *ompA*,
522 however, this has been shown to be limited in terms of discriminatory power (7). The results
523 from this study have demonstrated that the use of MLVA-*ompA* can provide a sufficient
524 resolution for molecular epidemiology studies in trachoma since it exceeds the
525 recommended guideline value of DP= 0.95 (Van Belkum et al., 2007). MLVA-*ompA* detected
526 a reduction in Ct diversity with additional rounds of MDA which may mean that while the
527 prevalence of infection remain high in this part of Amhara, the reduction in pathogen
528 population diversity serves as an indicator of treatment impact, possibly indicating the start
529 of a decline in overall infection prevalence.

530

531 Application of MLVA by Pedersen et al. (2008) (33) found DP = 0.94, which is consistent with
532 other MLVA-*ompA* studies on genital serovars (50,51). For LGV serovars, analyses testing the
533 stability of VNTRs suggested MLVA to be a suitable typing method (34), however further
534 testing by the same group on a larger set of LGV clinical samples suggested that MLVA-
535 *ompA* could not reach the recommended 0.95 cut-off (15). Other typing systems such as a
536 combination of MLST and MLVA (22) and alternate MLVA targets (52) have been shown to
537 have a DP of between 0.6-0.99. The typing system with the highest DP of 0.99, MLST+MLVA,
538 requires 8 individual loci to be sequenced, double the number of this study, for a relatively

539 marginal gain in DP. In addition to this, it is possible that when the mutation rate of a
540 population is too high, a high level of DP provides too much resolution, and epidemiological
541 links may be obscured (22). In this study, the higher DP highlighted the difference between
542 the diversity indices of the two timepoints in South Gondar. The *ompA* serovar type
543 suggests a loss of diversity after extra rounds of MDA, with serovar A dominating in the
544 second time point relative to the first, and the enhanced resolution provided by MLVA-
545 *ompA* typing confirmed that the number of variants is significantly lower after extra rounds
546 of MDA. This would suggest that MDA has driven purifying selection, however whether or
547 not these STs persist in the population requires further monitoring. Surveys continue to
548 monitor trachoma in Amhara, and it is recommended that these newly collected samples
549 are MLVA-*ompA* typed to further understand the effect of MDA on diversity and the decline
550 in infection prevalence.

551

552 Studies have shown that particular Ct variants have greater virulence (53,54), with
553 identification of variants usually based on WGS. The relatively simpler and less expensive
554 method of MLVA-*ompA* can provide sufficient resolution, enabling the study of intervention
555 effects at the population level. We did not find significant associations between high,
556 medium or low frequency STs and the clinical signs of trachoma or Ct load at village or
557 district level. It is possible that due to the restricted number of samples included, there was
558 insufficient statistical power to identify a relationship, however it is more likely that the
559 regular rounds of MDA this population underwent has limited the likelihood of a single or
560 common variant to become established and cause a significant increase in clinical signs of
561 active trachoma. MSTs support this proposition as they demonstrate an homogenous

562 distribution of variants within zones and across time (i.e. a lack of evidence for clustering of
563 STs).

564

565 There are limitations to the MLVA-*ompA* typing method and our study. Firstly, mixed
566 infections are not able to be identified since sequencing of a mixed product cannot be
567 unequivocally interpreted, as mixed genotypes create mixed sequences (33). In addition to
568 this, in this study some samples were not able to be assigned for CT1291 and CT1299, with
569 issues associated with the repeat cytosine (C)-region. DNA polymerase slippage is a well-
570 known phenomenon where *taq* DNA polymerase disassociates from the template and re-
571 anneals at a matching base further along the strand, altering the original length of the
572 repeat region (55). This is part of the reason VNTRs have such high variation, as this happens
573 naturally during DNA replication. CT1335, which did not have any repeat Cs and had a
574 maximum number of 11 repetitions of T nucleotides, did not display this slippage. Whether
575 or not this is due to the number of repeats or cytosine itself driving polymerase slippage is
576 unknown. In other studies, however, repeat regions of 21 Cs have been recorded (56),
577 suggesting a methodological or DNA template issue. Short-read WGS struggles with repeat
578 regions, this is due to problems with alignment of short read sequences and the algorithms
579 used in next generation sequencing (NGS)/WGS pipelines (57,58). This is reflected in the
580 CT1291 and CT1299 VNTR sequences extracted from our WGS sequences from the samples
581 in this study. In addition, the maximum likelihood phylogenetic tree constructed through
582 WGS-derived sequences support this statement, wherein a significant portion of the
583 sequences were indistinguishable and clustered together, irrespective of their associated
584 STs. We found a high proportion of non-assigned bases within these two VNTR regions, as
585 well as inconsistent results with the Sanger sequencing VNTR results. One possible reason is

586 mapping ambiguity; When aligning short reads to a reference genome, the presence of
587 VNTRs can lead to mapping ambiguity. If the number of repeats differs between the sample
588 and the reference genome, it may be challenging to accurately map the reads to the correct
589 genomic location (59). As Sanger sequencing is often used to confirm the presence of single
590 nucleotide polymorphisms found using WGS, it is assumed that the Sanger derived VNTR
591 sequences in this study are the likely the more accurate (60).

592

593 The data from this study has shown that MLVA-*ompA* is a suitable technique for identifying
594 ocular Ct variants on a local scale. While there were some issues surrounding long repeat
595 regions, the typeability for all three VNTRs was approximately 65%. While the typeability of
596 *ompA* variants was above 80%, this was after a second round of nested PCR in ~33% of
597 samples, which involved further manipulation, financial costs and increased the risk of
598 contamination. NGS techniques are associated with different issues, such as the
599 requirement for more expensive equipment, more complicated analysis pipelines and
600 computing power. It is difficult to estimate the exact cost of WGS as genome size influences
601 the number of samples that can be run in parallel, and a bioinformatician is generally
602 required to analyse the data, a substantial part of overall costs (61). The cost of performing
603 WGS by Pickering et al (2022) (31) was estimated at £250 per sample without post
604 sequencing analysis, whereas the average cost of Sanger sequencing is estimated at ~£20
605 per sample for all four targets required for MLVA-*ompA*. Sanger sequencing also requires
606 more readily available technology than WGS (since PCR amplicons can be simply posted
607 service providers). Sanger sequencing of a single product generates less data that simpler to
608 interpret with minimal training required.

609

610 This dataset has provided a unique collection of samples in which to investigate the utility of
611 the MLVA-*ompA* method in a trachoma endemic region. Amhara is a region experiencing
612 persistently high levels of trachoma (30), and is one of few programs that has been regularly
613 monitoring Ct infection with ocular swabbing and PCR. The addition of Ct monitoring to
614 trachoma programs not only allows for the tracking of Ct prevalence, but also allows for a
615 deeper understanding of variant diversity, and the efficacy of MDA, and it is recommended
616 that other programs experiencing persistent or recrudescing trachoma include ocular Ct
617 testing. The application of this typing technique provides an accessible, affordable and
618 functional tool for tracking the spread and diversity of Ct variants over time and space,
619 which can provide important information on the efficacy of MDA, aiding trachoma
620 elimination programmes in resource limited settings.

621

622 **Acknowledgements**

623

624 Many thanks to the late Dr Julius Schachter and Jeanie Moncada for the kind gift of
625 B/Tunis864 original DNA. The authors thank Abbott for donation of the m2000 RealTime
626 molecular diagnostics system and consumables.

627

628 **References**

629

- 630 1. WHO. WHO Alliance for the Global Elimination of Trachoma: progress report on
631 elimination of trachoma, 2021. Weekly Epidemiological Record [Internet].
632 2022;97(31):353–364. Available from:
633 <https://apps.who.int/iris/bitstream/handle/10665/361290/WER9731-eng-fre.pdf>
- 634 2. Lietman T, Porco T, Dawson C, Blower S. Global elimination of trachoma: how
635 frequently should we administer mass chemotherapy? Nat Med. 1999;5(5):572–6.

- 636 3. Chidambaram JD, Alemayehu W, Melese M, Lakew T, Yi E, House J, et al. Effect of a
637 single mass antibiotic distribution on the prevalence of infectious trachoma. *JAMA*.
638 2006;295(10):1142–6.
- 639 4. Xiong T, Yue Y, Li WX, Choonara I, Qazi S, Chen HJ, et al. Effectiveness of azithromycin
640 mass drug administration on trachoma: a systematic review. *Chin Med J (Engl)*.
641 2021;134(24):2944–53.
- 642 5. Ramadhani AM, Derrick T, Macleod D, Holland MJ, Burton MJ. The relationship
643 between active trachoma and ocular *Chlamydia trachomatis* infection before and
644 after mass antibiotic treatment. *PLoS Negl Trop Dis*. 2016;10(10):e0005080.
- 645 6. Solomon AW, Burton MJ, Gower EW, Harding-Esch EM, Oldenburg CE, Taylor HR, et
646 al. Trachoma. *Nat Rev Dis Primers*. 2022;8(1):32.
- 647 7. Rawre J, Juyal D, Dhawan B. Molecular typing of *Chlamydia trachomatis*: An overview.
648 *Indian J Med Microbiol*. 2017;35(1):17–26.
- 649 8. Chin SA, Alemayehu W, Melese M, Lakew T, Cevallos V, Lietman TM, et al. Association
650 of *Chlamydia trachomatis* ompA genovar with trachoma phenotypes. *Eye [Internet]*.
651 2018;32(8):1411–20. Available from: <https://doi.org/10.1038/s41433-018-0069-4>
- 652 9. Pickering H, Teng A, Faal N, Joof H. Genome-wide profiling of humoral immunity and
653 pathogen genes under selection identifies immune evasion tactics of *Chlamydia*
654 *trachomatis* during ocular. 2017; Available from:
655 <https://www.nature.com/articles/s41598-017-09193-2>
- 656 10. Millman KL, Tavare S, Dean D. Recombination in the ompA gene but not the omcB
657 gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism,
658 immune surveillance, and persistence of the organism. *J Bacteriol*.
659 2001;183(20):5997–6008.
- 660 11. Andreasen AA, Burton MJ, Holland MJ, Polley S, Faal N, Mabey DCW, et al. *Chlamydia*
661 *trachomatis* ompA variants in trachoma: what do they tell us? *PLoS Negl Trop Dis*.
662 2008;2(9):e306.
- 663 12. Moller JK, Pedersen LN, Persson K. Comparison of Gen-probe transcription-mediated
664 amplification, Abbott PCR, and Roche PCR assays for detection of wild-type and
665 mutant plasmid strains of *Chlamydia trachomatis* in Sweden. *J Clin Microbiol*.
666 2008;46(12):3892–5.
- 667 13. Jurstrand M, Christerson L, Klint M, Fredlund H, Unemo M, Herrmann B.
668 Characterisation of *Chlamydia trachomatis* by ompA sequencing and multilocus
669 sequence typing in a Swedish county before and after identification of the new
670 variant. *Sex Transm Infect*. 2010;86(1):56–60.
- 671 14. Ikryannikova LN, Shkarupeta MM, Shitikov EA, Il'ina EN, Govorun VM. Comparative
672 evaluation of new typing schemes for urogenital *Chlamydia trachomatis* isolates.
673 *FEMS Immunol Med Microbiol*. 2010;59(2):188–96.
- 674 15. Manning C, O'Neill C, Clarke IN, Rebec M, Cliff PR, Marsh P. High-resolution
675 genotyping of Lymphogranuloma Venereum (LGV) strains of *Chlamydia trachomatis*
676 in London using multi-locus VNTR analysis-ompA genotyping (MLVA-ompA). *PLoS*
677 *One*. 2021;16(7):e0254233.
- 678 16. Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al.
679 Guidelines for the validation and application of typing methods for use in bacterial
680 epidemiology. *Clinical Microbiology and Infection*. 2007;13:1–46.
- 681 17. Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, et al.
682 Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies

- 683 phylogenetic relationships masked by current clinical typing. *Nat Genet.* 2012 Apr
684 11;44(4):413–9.
- 685 18. Miyoshi-Akiyama T, Hayakawa K, Ohmagari N, Shimojima M, Kirikae T. Multilocus
686 sequence typing (MLST) for characterization of *Enterobacter cloacae*. *PLoS One.*
687 2013;8(6):e66358.
- 688 19. Boonsilp S, Thaipadungpanit J, Amornchai P, Wuthiekanun V, Bailey MS, Holden MTG,
689 et al. A single multilocus sequence typing (MLST) scheme for seven pathogenic
690 *Leptospira* species. *PLoS Negl Trop Dis.* 2013;7(1):e1954.
- 691 20. Klint M, Fuxelius HH, Goldkuhl RR, Skarin H, Rutemark C, Andersson SG, et al. High-
692 resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence
693 analysis. *J Clin Microbiol.* 2007;45(5):1410–4.
- 694 21. Pilo S, Zizelski Valenci G, Rubinstein M, Pichadze L, Scharf Y, Dveyrin Z, et al. High-
695 resolution multilocus sequence typing for *Chlamydia trachomatis*: improved results
696 for clinical samples with low amounts of *C. trachomatis* DNA. *BMC Microbiol.*
697 2021;21(1):1–11.
- 698 22. Bom RJM, Christerson L, Schim van der Loeff MF, Coutinho RA, Herrmann B, Bruisten
699 SM. Evaluation of high-resolution typing methods for *Chlamydia trachomatis* in
700 samples from heterosexual couples. *J Clin Microbiol.* 2011;49(8):2844–53.
- 701 23. Fazekas AJ, Steeves R, Newmaster SG. Improving sequencing quality from PCR
702 products containing long mononucleotide repeats. *Biotechniques.* 2010;48(4):277–
703 85.
- 704 24. Dean D, Bruno WJ, Wan R, Gomes JP, Devignot S, Mehari T, et al. Predicting
705 phenotype and emerging strains among *Chlamydia trachomatis* infections. *Emerg*
706 *Infect Dis.* 2009;15(9):1385–94.
- 707 25. Harding-Esch EM, Holland MJ, Schémann JF, Sillah A, Sarr B, Christerson L, et al.
708 Impact of a single round of mass drug administration with azithromycin on active
709 trachoma and ocular *Chlamydia trachomatis* prevalence and circulating strains in The
710 Gambia and Senegal. *Parasit Vectors.* 2019;12(1):1–19.
- 711 26. Nash SD, Stewart AEP, Astale T, Sata E, Zerihun M, Gessese D, et al. Trachoma
712 prevalence remains below threshold in five districts after stopping mass drug
713 administration: results of five surveillance surveys within a hyperendemic setting in
714 Amhara, Ethiopia. *Trans R Soc Trop Med Hyg [Internet].* 2018;112(12):538–45.
715 Available from: <https://academic.oup.com/trstmh/article/112/12/538/5108528>
- 716 27. Stewart AEP, Zerihun M, Gessese D, Melak B, Sata E, Nute AW, et al. Progress to
717 eliminate trachoma as a public health problem in Amhara National Regional State,
718 Ethiopia: results of 152 population-based surveys. *Am J Trop Med Hyg.*
719 2019;101(6):1286.
- 720 28. Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR. A simple system for the
721 assessment of trachoma and its complications. *Bull World Health Organ [Internet].*
722 1987;65(4):477–83. Available from: <https://pubmed.ncbi.nlm.nih.gov/3500800>
- 723 29. Nash SD, Chernet A, Moncada J, Stewart AEP, Astale T, Sata E, et al. Ocular *Chlamydia*
724 *trachomatis* infection and infectious load among pre-school aged children within
725 trachoma hyperendemic districts receiving the SAFE strategy, Amhara region,
726 Ethiopia. *PLoS Negl Trop Dis.* 2020;14(5):e0008226.
- 727 30. Sata E, Nute AW, Astale T, Gessese D, Ayele Z, Zerihun M, et al. Twelve-year
728 longitudinal trends in trachoma prevalence among children aged 1–9 Years in

- 729 Amhara, Ethiopia, 2007–2019. *American Journal of Tropical Medicine and Hygiene*.
730 2021;104(4):1278–89.
- 731 31. Pickering H, Chernet A, Sata E, Zerihun M, Williams CA, Breuer J, et al. Genomics of
732 Ocular Chlamydia trachomatis after 5 years of SAFE interventions for trachoma in
733 Amhara, Ethiopia. *J Infect Dis*. 2022;225(6):994–1004.
- 734 32. Moncada J, Shayevich C, Philip SS, Lucic D, Schachter J. Detection of Chlamydia
735 trachomatis and Neisseria gonorrhoeae in rectal and oropharyngeal swabs and urine
736 specimens from men who have sex with men with abbott’s M2000 RealTime. *Sex
737 Transm Dis*. 2015;42(11):650–1.
- 738 33. Pedersen LN, Podenphant L, Moller JK. Highly discriminative genotyping of Chlamydia
739 trachomatis using omp1 and a set of variable number tandem repeats. *Clin Microbiol
740 Infect*. 2008;14(7):644–52.
- 741 34. Labiran C, Clarke IN, Cutcliffe LT, Wang Y, Skilton RJ, Persson K, et al. Genotyping
742 markers used for multi locus VNTR analysis with omp A (MLVA-omp A) and multi
743 sequence typing (MST) retain stability in Chlamydia trachomatis. *Front Cell Infect
744 Microbiol*. 2012;2:68.
- 745 35. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, et al. Guidelines for Sanger
746 sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic
747 Investigation*. 2020;32(6):767–75.
- 748 36. Andreasen AA, Burton MJ, Holland MJ, Polley S, Faal N, Mabey DC, et al. Chlamydia
749 trachomatis ompA Variants in Trachoma: What Do They Tell Us? *PLoS Negl Trop Dis*.
750 2008;2(9):e306.
- 751 37. Wang Y, Skilton RJ, Cutcliffe LT, Andrews E, Clarke IN, Marsh P. Evaluation of a high
752 resolution genotyping method for Chlamydia trachomatis using routine clinical
753 samples. *PLoS One*. 2011;6(2):e16971.
- 754 38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.
755 2012;9(4):357–9.
- 756 39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
757 SAMtools and BCftools. *Gigascience*. 2021;10(2):giab008.
- 758 40. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing
759 systems: an application of Simpson’s index of diversity. *J Clin Microbiol*.
760 1988;26(11):2465–6.
- 761 41. Oksanen J, Kindt R, Legendre P, O’Hara B, Stevens MH, Oksanen MJ, Suggests MA.
762 The vegan package. *Community ecology package*. 2007 Oct 3;10(631-637):719.
- 763 42. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
764 lme4. *J Stat Softw*. 2015;67(1):1–48.
- 765 43. Knudson C. An Introduction to Model-Fitting with the R package glmm. 2018.
- 766 44. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. PHYLOViZ 2.0:
767 providing scalable data integration and visualization for multiple phylogenetic
768 inference methods. *Bioinformatics*. 2017;33(1):128–9.
- 769 45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
770 improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
- 771 46. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple
772 sequence alignment based on fast Fourier transform. *Nucleic Acids Res*.
773 2002;30(14):3059–66.

- 774 47. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms
775 and methods to estimate maximum-likelihood phylogenies: assessing the
776 performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21.
- 777 48. Mabey DC, Solomon AW, Foster A. *Trachoma.* *Lancet.* 2003;362(9379):223–9.
- 778 49. Ghasemian E, Inic-Kanada A, Collingro A, Mejdoubi L, Alchalabi H, Keše D, et al.
779 Comparison of genovars and *Chlamydia trachomatis* infection loads in ocular samples
780 from children in two distinct cohorts in Sudan and Morocco. *PLoS Negl Trop Dis.*
781 2021;15(8):e0009655.
- 782 50. Qin X, Zheng H, Xue Y, Ren X, Yang B, Huang J, et al. Prevalence of *Chlamydia*
783 *trachomatis* genotypes in men who have sex with men and men who have sex with
784 women using multilocus *vntr* analysis-*omp a* typing in guangzhou, China. *PLoS One.*
785 2016;11(7):e0159658.
- 786 51. Satoh M, Ogawa M, Saijo M, Ando S. Multilocus VNTR analysis-*ompA* typing of
787 venereal isolates of *Chlamydia trachomatis* in Japan. *Journal of Infection and*
788 *Chemotherapy.* 2014;20(10):656–9.
- 789 52. Peuchant O, Le Roy C, Herrmann B, Clerc M, Bébéar C, de Barbeyrac B. MLVA
790 subtyping of genovar E *Chlamydia trachomatis* individualizes the Swedish variant and
791 anorectal isolates from men who have sex with men. *PLoS One.* 2012;7(2):e31538.
- 792 53. Borges V, Cordeiro D, Salas AI, Lodhia Z, Correia C, Isidro J, et al. *Chlamydia*
793 *trachomatis*: when the virulence-associated genome backbone imports a prevalence-
794 associated major antigen signature. *Microb Genom.* 2019;5(11).
- 795 54. Kari L, Whitmire WM, Carlson JH, Crane DD, Reveneau N, Nelson DE, et al. Pathogenic
796 diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is
797 affected by subtle genomic variations. *J Infect Dis.* 2008;197(3):449–56.
- 798 55. Viguera E, Canceill D, Ehrlich SD. Replication slippage involves DNA polymerase
799 pausing and dissociation. *EMBO J.* 2001;20(10):2587–95.
- 800 56. Labiran C, Rowen D, Clarke IN, Marsh P. Detailed molecular epidemiology of
801 *Chlamydia trachomatis* in the population of Southampton attending the genitourinary
802 medicine clinic in 2012-13 reveals the presence of long established genotypes and
803 transitory sexual networks. *PLoS One.* 2017;12(9):e0185059.
- 804 57. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. *Trends*
805 *in genetics.* 2008;24(3):142–9.
- 806 58. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O’huallachain M, et al. Performance
807 comparison of whole-genome sequencing platforms. *Nat Biotechnol.* 2012;30(1):78–
808 82.
- 809 59. Bakhtiari M, Shleizer-Burko S, Gymrek M, Bansal V, Bafna V. Targeted genotyping of
810 variable number tandem repeats with *advNTR*. *Genome Res.* 2018;28(11):1709–19.
- 811 60. GenScript. [https://www.genscript.com/gene-news/sanger-sequencing-vs-next-](https://www.genscript.com/gene-news/sanger-sequencing-vs-next-generation-sequencing.html)
812 [generation-sequencing.html](https://www.genscript.com/gene-news/sanger-sequencing-vs-next-generation-sequencing.html). 2021. Sanger Sequencing vs. Next-Generation
813 Sequencing (NGS).
- 814 61. Christensen KD, Dukhovny D, Siebert U, Green RC. Assessing the costs and cost-
815 effectiveness of genomic sequencing. *J Pers Med.* 2015;5(4):470–86.
- 816