

# Genomics of Ocular *Chlamydia trachomatis* after 5 years of SAFE interventions for trachoma in Amhara, Ethiopia

Harry Pickering,<sup>1</sup> Ambahun Chernet,<sup>2</sup> Eshetu Sata,<sup>2</sup> Mulat Zerihun,<sup>2</sup> Charlotte A. Williams,<sup>3</sup> Judith Breuer,<sup>3</sup> Andrew W. Nute,<sup>4</sup> Mahteme Haile,<sup>5</sup> Taye Zeru,<sup>5</sup> Zerihun Tadesse,<sup>2</sup> Robin L. Bailey,<sup>1</sup> E. Kelly Callahan,<sup>4</sup> Martin J. Holland<sup>1†</sup>, Scott D. Nash<sup>4†</sup>

1. Department of Clinical Research, London School of Hygiene & Tropical Medicine, UK
2. The Carter Center, Ethiopia
3. Division of Infection and Immunity, University College London, UK
4. The Carter Center, Atlanta, USA
5. Amhara Public Health Institute Bahir Dar, Ethiopia

† These authors contributed equally to this work

Corresponding author: Harry Pickering, [harrypickering@gmail.com](mailto:harrypickering@gmail.com)

**Summary:** Whole-genome sequencing of ocular *Chlamydia trachomatis* from Amhara, Ethiopia found no evidence of macrolide-resistance, despite five years of trachoma control interventions, including mass distribution of azithromycin. Polymorphisms in *ompA* were associated with prevalence of ocular infection and trachomatous disease.

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

To eliminate trachoma as a public health problem, the WHO recommends the SAFE (Surgery, Antibiotics, Facial cleanliness, and Environmental improvement) strategy. As part of the SAFE strategy in the Amhara Region, Ethiopia, the Trachoma Control Program distributed over 124 million doses of antibiotic between 2007 and 2015. Despite this, trachoma remained hyperendemic in many districts and a considerable level of *Chlamydia trachomatis* (*Ct*) infection was evident.

We utilised residual material from Abbott m2000 *Ct* diagnostic tests to sequence 99 ocular *Ct* samples from Amhara and investigated the role of *Ct* genomic variation in continued transmission of *Ct*.

Sequences were typical of ocular *Ct*, at the whole-genome level and in tissue tropism-associated genes. There was no evidence of macrolide-resistance in this population. Polymorphism around *ompA* gene was associated with village-level trachomatous inflammation-follicular prevalence. Greater *ompA* diversity at the district-level was associated with increased *Ct* infection prevalence.

We found no evidence for *Ct* genomic variation contributing to continued transmission of *Ct* after treatment, adding to evidence that azithromycin does not drive acquisition of macrolide resistance in *Ct*. Increased *Ct* infection in areas with more *ompA* variants requires longitudinal investigation to understand what impact this may have on treatment success and host immunity.

**Keywords:** trachoma, *Chlamydia trachomatis*, antimicrobial resistance, genomics, whole-genome sequencing, mass drug administration, azithromycin, neglected tropical disease

## Background

Trachoma is a blinding disease caused by *Chlamydia trachomatis* (*Ct*). To eliminate trachoma as a public health problem, the WHO recommends the SAFE (Surgery, Antibiotics, Facial cleanliness, and Environmental improvement) strategy [1]. This includes annual mass drug administration (MDA) of azithromycin to individuals aged  $\geq 6$  months and topical tetracycline eye ointment for pregnant women and children aged  $< 6$  months. The number of recommended years of interventions is based on prevalence of trachoma in a district [2]. For districts considered hyperendemic for trachoma, defined as a trachomatous inflammation-follicular (TF) prevalence of  $\geq 30\%$  among children aged 1-9 years, 5-7 years of SAFE are recommended followed by further population-based surveys to determine the impact of the interventions.

As part of the SAFE strategy in Amhara National Regional State, Ethiopia, the Trachoma Control Program distributed over 124 million doses of antibiotic from 2007-2015 [3]. Both administrative and self-reported coverage have demonstrated treatment coverage close to or above the WHO recommended threshold of 80% [3–5]. The program also provided health education and assisted in the construction of latrines as part of the F and E components of SAFE [3]. Despite an average of 5 years of these interventions, trachoma remained hyperendemic in many districts, with considerable levels of *Ct* infection [6].

Historically, *Ct* molecular epidemiology focused on *ompA*, [7] which encodes the major outer membrane protein. More recently multilocus sequence typing schemes have been used [8,9]. Since 2010, there has been a rapid expansion of *Ct* whole-genome sequencing (WGS), due to the ability to sequence directly from clinical samples [10,11]. Despite more than seven-hundred *Ct* genomes being sequenced [12–15], few studies have evaluated the role of genome-level variation in *Ct* transmission and outcomes of infection. Recent publications have begun to address these questions in *Ct* from trachoma-endemic settings [15,16]. WGS additionally allows monitoring of antimicrobial resistance in *Ct* [16–18], which is of critical importance as MDA with azithromycin is key for trachoma control, and is under consideration as an intervention for childhood mortality [19], neonatal sepsis [20], and malaria [21].

The Trachoma Control Program in Amhara has conducted multiple studies to better understand the epidemiology of trachoma in communities that have received approximately 5 years of annual MDA, yet still have significant levels of *Ct* infection and disease. This study sequenced 99 ocular *Ct* samples from Amhara to identify antimicrobial resistance alleles and investigate the role of genomic variation in the continued

transmission of *Ct*. We further explored the relationship between *Ct* genomic variation, ocular *Ct* infection prevalence, and trachomatous disease prevalence at the village and district-level.

## **Methods**

### *Study design and population*

Between 2007 and 2010 the SAFE strategy was scaled to reach all Amharan districts with interventions administered for 5 years. Methodology for these district-level surveys has been published previously [3]. Briefly, a multi-stage cluster randomized methodology was used, whereby clusters (villages) were selected using a population proportional to estimated size method, and within a cluster, a segmentation approach was used to randomly select 30-40 households [3].

After enumerating all residents, consented residents were examined for trachoma. Every-other cluster was chosen for swab collection prior to surveying a district, and during the house-to-house survey, the first 25 children aged 1-5 years with parental consent were swabbed for the presence of infection. If more than one child aged 1-5 years lived in a household, one child was randomly chosen by survey software.

### *Sample collection and processing*

Gloved graders swabbed the upper tarsal conjunctiva three times with a polyester-tipped swab, rotating 120 degrees along the swab's axis each time to collect a sufficient epithelial specimen [6]. Samples were transferred to the Amhara Public Health Institute (APHI) and stored at -20° C. Conjunctival swabs from each district were randomized and five samples were combined into each pool. Pools were processed with the RealTime PCR assay on the Abbott m2000 system at the APHI laboratory [6]. All individual samples from positive pools from North Gondar, South Gondar, East Gojam, and Waghemra were processed again to provide individual-level data [22]. Samples from these zones were prioritized owing to the persistent high trachoma prevalence. For positive individual samples, the PCR cycle threshold was converted to *Ct* elementary body (EB) equivalent concentration based on a calibration curve of known EB concentrations on the RealTime Assay [22].

Once *Ct* load was known for the positive individual samples, a total of 240 with the highest load were chosen for this project. Samples with sufficient *Ct* load, likely to obtain high quality full genome sequence data based on our previous studies, were re-extracted as described below [15,16,23].

### *Ct detection and sequencing preparation*

DNA was extracted from 800 µl residual material per sample from Abbott m2000 diagnostic tests using QIAamp mini DNA kit. Samples were quantified using a genome target by qPCR [24]. Samples with  $\geq 10$  genome copies per µl of DNA were considered for WGS.

### *Sequencing, processing, and analysis of Ct*

Sequencing was performed as previously described [15] except we utilised the SureSelectXT Low Input kit. Processing and analysis of sequenced reads was performed as previously described [16]. Briefly, raw reads were trimmed and filtered using Trimmomatic. Filtered reads were aligned to a reference genome (A/Har13) with Bowtie2, variants were called with SAMtools/BCFtools. Multiple genome and plasmid alignments were generated using progressiveMauve, multiple gene alignments were generated using MUSCLE. Phylogenies were computed using RAxML, predicted regions of recombination were masked using Gubbins. Domain structure of *tarP* and truncation of *trpA* were characterised as previously described [15]. ABRicate and the ResFinder database were used to identify antimicrobial resistance genes in the reference-assembled genomes and *de novo* assembled reads.

### *Genome-wide association analyses*

Genome-wide association analysis (GWAS) was performed to identify polymorphisms specific to this population of ocular *Ct* sequences through comparison of 99 Amharan *Ct* genomes to 213 previously sequenced samples from trachoma-endemic communities. Heterozygous calls and positions with greater than 25% missing data were removed. Polymorphisms were considered conserved in Amhara if the major allele frequency was  $> 0.8$  and rare in the representative ocular population if the same allele was at a frequency  $< 0.2$ . The final analysis included 116 single nucleotide polymorphisms (SNPs). A logistic regression was performed with each Amhara-specific site as the independent variable and origin of the sequence as the dependent variable (reference level; representative and comparator level; Amharan). P-values were Bonferroni corrected.

GWAS was performed to identify *Ct* polymorphisms associated with village-level clinical data. Heterozygous base calls were and positions with a minor allele frequency of less than 25% or greater than 25% missing data were removed. The final analysis included 681 SNPs. A linear regression was performed with each SNP as the independent variable and village-level *Ct* infection, TF, or TI prevalence as the dependent variable. District was included as a random-effect and with adjustment for age and gender. P-values were Bonferroni corrected. Additionally, a sliding-window approach was used to identify polymorphic regions of the genome. Windows of 10 kilobases were evaluated, with a step size of 1 kilobase. The final analysis included 907 polymorphic regions. A linear regression was performed with each polymorphic region collapsed into a pseudo-haplotype per sequence as the independent variable, including district as a random-effect and adjusted for age and gender. This model was compared to a model including only the covariates and random-effects by F-test. P-values were Bonferroni corrected.

#### *Inference of ompA sequences*

Complete sequences of *ompA* were obtained from whole-genome sequence data using the reference-based assembly method described above with one change. Each sample was assembled against four reference sequences (A/Har-13, B/Jali-20, C/TW-3, and D/UW-3) and the assembly with the highest coverage was used for downstream analyses. Serovar of *ompA* was assigned using maximum *blastn* homology against all published *Ct* sequences. Genotypes of *ompA* were manually determined using SeaView. Diversity of *ompA* genotypes was calculated as Simpson's D using *vegan* in R.

#### **Results**

Ocular swabs previously confirmed as positive for *Ct* DNA were selected for this study (n = 240), samples with sufficiently high concentration of *Ct* DNA after re-extraction were considered for WGS (n = 135). Of these, 99 were randomly selected for sequencing to match the complete dataset on age, gender, and zone of collection. The sequenced and complete samples were comparable (Table 1), except as expected a higher median load of infection in sequenced samples.

The Amharan *Ct* genomes formed two subclades within the T2 ocular clade (Figure 1). The two subclades were predominantly separated by *ompA* genotype, with 52 serovar A (SvA) and 47 serovar B (SvB) genomes. Focusing on genomes from ocular infections (SI Figure 2), the SvA Amharan genomes branch together independent from any previously sequenced *Ct*. The SvB Amharan genomes were split across two branches. One branch was most closely related to A/Har-13, isolated from Saudi Arabia. The second, smaller branch was most closely related to Ba/Apache-2 from the USA as well as recently sequenced ocular *Ct* from Solomon Islands.

Several *Ct* genes and genomic regions are hypothesised to be indicative of tissue tropism, with polymorphisms distinct to ocular, urogenital and LGV sequences. All Amharan *Ct* genomes had *tarP* domain structure typical of ocular sequences [25]. Similarly, all Amharan genomes had inactivating mutations in *trpA*, leading to a non-functional tryptophan synthase [26]. Polymorphic membrane proteins clustered phylogenetically with ocular isolates (SI Figure 2) [27]. There was minimal polymorphism in the *Ct* plasmid within the Amharan genomes and they were closely related to previously sequenced ocular isolates (SI Figure 3). There was no evidence for the presence of macrolide-resistance alleles in the assembled genomes or *de novo* assembled reads.

Amharan *Ct* genomes were compared to 213 previously sequenced samples from trachoma-endemic communities to identify genomic markers specific to Amhara.[12–16,23] Of 36,805 polymorphic sites (Figure 2a), 116 were conserved in Amhara (frequency  $\geq 0.8$ ) and rare in the representative ocular population (frequency  $\leq 0.2$ ). These were dispersed throughout the genome (Figure 2b). Fourteen genes harboured two such sites and five genes contained three sites, all of which have previously been identified as polymorphic in distinct populations of *Ct* (Figure 2c).

A GWAS was performed to identify polymorphism within the Amharan *Ct* genomes related to village-level prevalence of *Ct* infection. The analysis included 681 single nucleotide polymorphisms (SNPs) in 99 genomes. No SNPs were associated with village-level prevalence of infection (SI Figure 4). A secondary sliding-window approach was utilised to identify polymorphic regions of the genome associated with infection prevalence. The

analysis included 907 polymorphic regions in 99 genomes. No polymorphic regions were associated with village-level prevalence of infection (SI Figure 4).

No SNPs were associated with village-level prevalence of TF (Figure 3a). However, eight polymorphic regions from positions 774,000-791,000 were associated with TF prevalence (Figure 3b). SNPs in these regions were focused in CTA0743/*pbpB* (harbouring 29 SNPs), CTA0747/*sufD* (10 SNPs) and CTA0742/*ompA* (7 SNPs). All SNPs in *sufD* were synonymous, while 8/29 and 3/7 SNPs in *pbpB* and *ompA* were non-synonymous.

No SNPs or polymorphic regions were associated with village-level prevalence of TI (SI Figure 5).

As *ompA* variation was important in *Ct* phylogeny and heterogeneity in TF profiles, we further investigated the geographical distribution of *ompA* serovars and their relationship to levels of *Ct* infection and TF. SvA and SvB of *ompA* were distributed across all zones (Figure 4). Village-level *Ct* infection, TF and TI prevalence were not associated with *ompA* serovar ( $p = 0.860, 0.382$  and  $0.177$  respectively). We identified nine *ompA* types in this population (Table 2). Six were SvA, defined by nine non-synonymous polymorphisms. Three were SvB, defined by two non-synonymous polymorphisms. Four of nine types were present in all zones (A1, A3, A5 and B3), four were exclusive to East Gojam (A2, A4, A6 and B1) and one was found in East Gojam and North Gondar (B2) (SI Figure 6). Types A1 ( $n = 5$ ) and B1 ( $n = 6$ ) had a nucleotide predicted amino acid change in the surface-exposed, variable domain 1 (VD1), A2 ( $n = 2$ ) in VD2 and A4 ( $n = 1$ ) in VD4.

Most villages (55/61) had only one *ompA* type in this study, therefore we evaluated *ompA* diversity at the district-level, using Simpson's D and adjusting for number of genomes sampled per district. We used published *Ct* infection, TF and TI prevalence estimates [6,28] (Table 3). *Ct* infection and TI prevalence were significantly higher with increasing *ompA* diversity, a similar trend was found for TF prevalence. In a multivariate model, only *Ct* infection prevalence was associated with increasing *ompA* diversity.



## Discussion

This study sequenced *Ct* from ocular samples collected from districts in Amhara, Ethiopia which had received approximately 5 years of the SAFE strategy, as part of trachoma control efforts. We found sequences were typical of ocular *Ct*, at both the whole-genome level and in tropism-associated genes, yet phylogenetically distinct from most previously sequenced *Ct* genomes. There was no evidence of macrolide-resistance alleles in this ocular *Ct* population. Greater *ompA* diversity at the district-level was associated with increased *Ct* infection prevalence. A continued commitment to the implementation of the SAFE strategy with consideration of enhanced MDA accompanied by further longitudinal investigation is warranted in Amhara.

Almost 900 million doses of azithromycin have been distributed by trachoma control programmes since 1999 and in Amhara 15 million doses are administered annually [3]. Mass distribution of azithromycin is likely to become more common as evidence grows of off-target effects such as reducing infectious diseases [21,29], diarrheal diseases [30] and childhood mortality [20,31,32]. There is concern about the impact of these programmes on development of antimicrobial resistance in [18,33]. This is particularly true where community-wide treatment with azithromycin has been unable to eliminate trachoma as a public health problem within expected timelines [30,34]. It has been shown that treating communities with azithromycin can increase nasopharyngeal carriage of macrolide-resistant *Staphylococcus* [35] and *Streptococcus* [36] and alters the faecal microbiome [37,38], with reports of increased macrolide-resistant *Escherichia coli* [39]. This study, in agreement with previous work [16–18] found no evidence of macrolide-resistance in this *Ct* population. While encouraging, it does not rule out macrolide-resistance as a potential problem in these communities. Carriage of macrolide-resistant pathogens in the gut and nasopharynx may be impacted by antibiotic treatment. Additionally, presence of additional species of *Chlamydia* [40,41] and non-chlamydial bacteria [42–45] in the ocular niche have been associated with trachoma, therefore resistance in other bacteria may be important.

No *Ct* genomes in this study had acquired azithromycin resistance alleles, however, there may be other genomic factors which support *Ct* transmission after treatment. To explore this, we compared Amharan *Ct* genomes with previously sequenced *Ct* to find polymorphism specific to this population that could explain continued transmission. The few SNPs identified as specific to Amhara were dispersed across the genome in known

polymorphic genes, rather than being overrepresented in genes related to *Ct* survival. The typical nature of this *Ct* population was supported by phylogenetic clustering with other ocular *Ct* sequences, presence of a non-functional tryptophan synthase operon and tropism-associated polymorphism in *tarP* and the polymorphic membrane proteins. Similarly to recent studies from distinct trachoma-endemic communities [12–16,23], the *Ct* sequences in this population formed two closely related subclades within the ocular clade, primarily separated by *ompA* serovar. Evidence of phylogenetic clustering by country of collection and the similarity to *Ct* sequences collected over 50 years prior to this study suggests diversification in ocular *Ct* is slow and geography-related, rather than driven by treatment-derived selection pressure. A surprising finding in this study was that a subgroup of SvB *Ct* from Amhara were most closely related to a historical genome from USA (Ba/Apache-2) and recently collected genomes from Solomon Islands [23]. It is possible the origin of these genomes is unique within this population; however, it is more likely that this is further evidence of the slow diversification of *Ct*. In support of this, *ompA* SvB sequences were significantly less diverse than SvA in this study. Furthermore, all major branches of ocular *Ct* phylogeny studied here included samples collected decades apart from geographically disparate sites.

We identified several polymorphic regions associated with village-level TF prevalence. The polymorphisms were mostly frequently found in *ompA*, *pbpB* and *sufD*, all of which are known to be polymorphic. *OmpA* encodes the major outer membrane protein which is the primary target of host immune responses and is believed to function as an adhesin and/or porin [46]. The functions of *pbpB* and *sufD* in *Ct* are unknown, bacterial homologues of these genes function in peptidoglycan synthesis and response to oxidative stress respectively. It is plausible that genes hypothesised to be involved in immune evasion and response to stress could impact *Ct* survival and response to treatment.

We found approximately equal representation of SvA and SvB in this study. Both serovars were present in all districts and were not associated with village-level *Ct* infection, TF, or TI prevalence. However, *Ct* infection prevalence was increased in districts with greater *ompA* diversity. Our data agrees with a Nepalese study that found increased *ompA* diversity in villages to be associated with higher trachoma prevalence [47]. In contrast, a more recent study from Ethiopia found no association between *ompA* diversity and *Ct* infection levels [48]. It is known immunity to *Ct* is serovar-specific [49,50], therefore it is plausible that in villages with multiple serovar

in circulation, individuals are more likely to be exposed to a serovar they do not have protective immunity against. Presence of one or more *ompA* variants should not impact treatment success, however, it is possible that higher levels of *Ct* infection pre-treatment, driven by presence of multiple serovars, could increase the likelihood of low-level transmission persisting after treatment.

A potential limitation of this study was bias towards samples with higher *Ct* load. It is possible that relationships between *ompA* variation and *Ct* infection prevalence might have been different if lower load infections were included, particularly at the village-level, as the majority (34/61) were represented by one sequence. We have also not sequenced material from Abbott m2000 specimens previously, it is possible that long-term storage in this format and multiple freeze-thaw cycles may have impacted DNA quality or quantity. However, obtaining high quality genomes from all sequenced samples, with as low as 500 *Ct* genomes input, suggests quality was not an issue. Additionally, our sample size was restricted by both *Ct* load and the cost of sequencing. This is the largest collection of ocular *Ct* genomes from a single geographical population, but it is still possible that we may have missed some smaller effects in the genome-wide analyses due to limited statistical power. Lastly, while study villages were randomly selected using standard programmatic methods and subsampling was matched for zone of collection, the conclusions of this study may have been different if we had been able to sequence a larger and more geographically diverse population of *Ct*.

Despite approximately five years of azithromycin MDA, we found no evidence for *Ct* genomic variation contributing to continued transmission of *Ct*, adding to evidence that azithromycin MDA does not drive acquisition of macrolide-resistance alleles in *Ct*. This study demonstrates feasibility of WGS of low-load, residual material and highlights the added value of collecting ocular swabs as part of trachoma surveys. Collection and long-term storage of these samples has helped alleviate concerns of azithromycin resistance in Amharan *Ct*, while offering important insights into the relationship between *ompA* variation and *Ct* infection levels. Future longitudinal investigation will be needed to understand what impact *ompA* diversity may have on treatment success in Amhara and other trachoma-endemic regions.

## Declarations

### *Ethics approval and consent to participate*

Survey methods were approved by the Emory University Institutional Review Board (protocol 079-2006) as well as the Amhara Regional Health Bureau. Due to high illiteracy rate among the population, approval was obtained for oral consent/assent. Oral consent/assent was recorded electronically for all participants according to the principles of the Declaration of Helsinki. Respondents were allowed to terminate the examination at any point without explanation. Further permission for sample transfer and genomic sequencing of *Ct* was provided by Emory University Institutional Review Board and the Amhara Regional Health Bureau. Human DNA testing and genotyping was not conducted on these samples. Any unused biological material will be returned to Ethiopia or disposed as required.

Accepted Manuscript

### *Consent for publication*

Not applicable.

### *Availability of data and materials*

All sequence data are available from the European Bioinformatics Institute archives (PRJEB38668).

### *Competing interests*

We declare no competing interests.

### *Funding*

This work received financial support from the Coalition for Operational Research on Neglected Tropical Diseases (COR NTD), which is funded at The Task Force for Global Health by the Bill & Melinda Gates Foundation, the United Kingdom Department for International Development, and the United States Agency for International Development. Additional financial support was received from the International Trachoma Initiative. HP and MH were funded by EU Horizon 2020 (grant ID: 733373).

### *Authors' contributions*

HP, RLB, EKC, MJH and SDN contributed to study design. HP, CAW, AC, ES, MZ, ZT, EKC, and SDN contributed to data collection. HP, AWN, EKC, MJH and SDN contributed to data analysis. All authors interpreted the findings, contributed to writing the manuscript, and approved the final version.

### *Acknowledgements*

The authors would like to acknowledge the study participants and field team in Amhara, Ethiopia. The authors acknowledge the infrastructure support provided by the UCL/UCLH Biomedical Research Centre funded Pathogen Genomics Unit. We also thank Abbott for donation of the m2000 RealTime molecular diagnostics system and consumables.

## Figure legends

### Figure 1.

Maximum likelihood reconstruction of whole genome phylogeny of ocular *Chlamydia trachomatis* (Ct) sequences from Amhara, Ethiopia. Whole genome phylogeny of 99 Ct sequences from Amhara and 183 Ct clinical and reference strains. Amharan Ct sequences were mapped to Ct A/HAR-13 using Bowtie2. SNPs were called using SAMtools/BCFtools. Phylogenies were computed with RAxML from a variable sites alignment using a GTR + gamma model and are midpoint rooted. The scale-bar indicates pairwise distance. Ct sequences are coloured by origin of the sample (“Origin”) and *ompA* serovar (“Serovar”).

### Figure 2.

Single nucleotide polymorphisms on the *Chlamydia trachomatis* (Ct) genome specific to Amhara, Ethiopia. a) Single nucleotide polymorphisms (SNPs) conserved in Amhara, Ethiopia (allele frequency  $\geq 0.8$ ) and rare in other Ct sequences (allele frequency  $\leq 0.2$ ) were identified by comparing these Ct sequences (n = 99) to ocular genomes from other populations (n = 213). b) Logistic regression found SNPs specific to this Amharan population to be dispersed throughout the genome (n = 116). c) Five genes harboured three Amhara-specific SNPs, putative function was determined by reference to published literature.

### Figure 3.

Polymorphisms on the *Chlamydia trachomatis* (Ct) genome associated with village-level TF prevalence. a) No single nucleotide polymorphisms were significantly associated with village-level TF prevalence. b) Eight polymorphic regions from positions 774,000 to 791,000 were associated with village-level prevalence of TF.

### Figure 4.

Geographical distribution and similarity of *ompA* serovars. a) Four zones in Amhara, Ethiopia were represented in this study. Pie charts represent village-level Ct prevalence (pie diameter) and presence of *ompA* serovars A (red) and B (blue). Maps were generated using R package ggmap, shape files were obtained from Google Maps.

## References

1. Mariotti SP, Pararajasegaram R, Resnikoff S. Trachoma: looking forward to Global Elimination of Trachoma by 2020 (GET 2020). *Am. J. Trop. Med. Hyg.* 2003.
2. Emerson PM, Burton M, Solomon AW, Bailey R, Mabey D. The SAFE strategy for trachoma control: Using operational research for policy, planning and implementation. *Bull. World Health Organ.* 2006.
3. Stewart AEP, Zerihun M, Gessese D, et al. Progress to eliminate trachoma as a public health problem in Amhara National Regional State, Ethiopia: Results of 152 population-based surveys. *Am J Trop Med Hyg.* 2019; .
4. Astale T, Sata E, Zerihun M, et al. Population-based coverage survey results following the mass drug administration of azithromycin for the treatment of trachoma in Amhara, Ethiopia. *PLoS Negl Trop Dis.* 2018; .
5. Ebert CD, Astale T, Sata E, et al. Population coverage and factors associated with participation following a mass drug administration of azithromycin for trachoma elimination in Amhara, Ethiopia. *Trop Med Int Heal.* 2019; .
6. Nash SD, Stewart AEP, Zerihun M, et al. Ocular Chlamydia trachomatis infection under the SAFE strategy in Amhara, Ethiopia, 2011-2015. *Clin Infect Dis.* 2018; .
7. Geisler WM, Black CM, Bandea CI, Morrison SG. Chlamydia trachomatis OmpA genotyping as a tool for studying the natural history of genital chlamydial infection. *Sex Transm Infect.* 2008; .
8. Klint M, Fuxelius HH, Goldkuhl RR, et al. High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis. *J Clin Microbiol.* 2007; .
9. Pedersen LN, Pødenphant L, Møller JK. Highly discriminative genotyping of Chlamydia trachomatis using omp1 and a set of variable number tandem repeats. *Clin Microbiol Infect.* 2008; .
10. Seth-Smith HMB, Harris SR, Skilton RJ, et al. Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture. *Genome Res.* 2013; .
11. Christiansen MT, Brown AC, Kundu S, et al. Whole-genome enrichment and sequencing of Chlamydia trachomatis directly from clinical samples. *BMC Infect Dis.* 2014; .

12. Harris SR, Clarke IN, Seth-Smith HMB, et al. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet.* **2012**; .
13. Andersson P, Harris SR, Smith HMBS, et al. *Chlamydia trachomatis* from Australian Aboriginal people with trachoma are polyphyletic composed of multiple distinctive lineages. *Nat Commun.* **2016**; .
14. Hadfield J, Harris SR, Seth-Smith HMB, et al. Comprehensive global genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing and recent lineage expansion. *Genome Res.* **2017**; .
15. Last AR, Pickering H, Roberts C h., et al. Population-based analysis of ocular *Chlamydia trachomatis* in trachoma-endemic West African communities identifies genomic markers of disease severity. *Genome Med.* **2018**; .
16. Alkhidir AAI, Holland MJ, Elhag WI, et al. Whole-genome sequencing of ocular *Chlamydia trachomatis* isolates from Gadarif State, Sudan. *Parasites and Vectors.* **2019**; .
17. Hong KC, Schachter J, Moncada J, Zhou Z, House J, Lietman TM. Lack of macrolide resistance in *Chlamydia trachomatis* after mass azithromycin distributions for trachoma. *Emerg Infect Dis.* **2009**; .
18. O'Brien KS, Emerson P, Hooper PJ, et al. Antimicrobial resistance following mass azithromycin distribution for trachoma: a systematic review. *Lancet Infect. Dis.* 2019.
19. Keenan JD, Bailey RL, West SK, et al. Azithromycin to reduce childhood mortality in sub-Saharan Africa. *N Engl J Med.* **2018**; .
20. Bojang A, Baines SL, Donovan L, et al. Genomic investigation of *Staphylococcus aureus* recovered from Gambian women and newborns following an oral dose of intra-partum azithromycin. *J Antimicrob Chemother.* **2019**; .
21. Schachterle SE, Mtove G, Levens JP, et al. Short-term malaria reduction by single-dose azithromycin during mass drug administration for trachoma, Tanzania. *Emerg Infect Dis.* **2014**; .
22. Nash SD, Chernet A, Moncada J, et al. Ocular *Chlamydia trachomatis* infection and infectious load among pre-school aged children within trachoma hyperendemic districts receiving the SAFE strategy,



- Amhara region, Ethiopia. *PLoS Negl Trop Dis* [Internet]. Public Library of Science; **2020**; 14(5):e0008226. Available from: <https://doi.org/10.1371/journal.pntd.0008226>
23. Butcher RMR, Sokana O, Jack K, et al. Low Prevalence of Conjunctival Infection with *Chlamydia trachomatis* in a Treatment-Naïve Trachoma-Endemic Region of the Solomon Islands. *PLoS Negl Trop Dis*. **2016**; .
  24. Butcher R, Houghton J, Derrick T, et al. Reduced-cost *Chlamydia trachomatis*-specific multiplex real-time PCR diagnostic assay evaluated for ocular swabs and use by trachoma research programmes. *J Microbiol Methods*. **2017**; .
  25. Lutter EI, Bonner C, Holland MJ, et al. Phylogenetic analysis of *Chlamydia trachomatis* tarp and correlation with clinical phenotype. *Infect Immun*. **2010**; .
  26. Caldwell HD, Wood H, Crane D, et al. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. *J Clin Invest*. **2003**; .
  27. Gomes JP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, Dean D. Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: Evidence for serovar da recombination and correlation with tissue tropism. *J Bacteriol*. **2006**; .
  28. Nash SD, Stewart AEP, Astale T, et al. Trachoma prevalence remains below threshold in five districts after stopping mass drug administration: Results of five surveillance surveys within a hyperendemic setting in Amhara, Ethiopia. *Trans. R. Soc. Trop. Med. Hyg*. 2018.
  29. Harrison MA, Harding-Esch EM, Marks M, et al. Impact of mass drug administration of azithromycin for trachoma elimination on prevalence and azithromycin resistance of genital *Mycoplasma genitalium* infection. *Sex Transm Infect. England*; **2019**; 95(7):522–528.
  30. Coles CL, Seidman JC, Levens J, Mkocho H, Munoz B, West S. Association of mass treatment with azithromycin in trachoma-endemic communities with short-term reduced risk of diarrhea in young children. *Am J Trop Med Hyg*. **2011**; .
  31. Whitty CJM, Glasgow KW, Sadiq ST, Mabey DC, Bailey R. Impact of community-based mass treatment for trachoma with oral azithromycin on general morbidity in Gambian children. *Pediatr Infect*

- Dis J. **1999**; .
32. Keenan JD, Ayele B, Gebre T, et al. Childhood mortality in a cohort treated with mass azithromycin for trachoma. *Clin Infect Dis.* **2011**; .
  33. Mack I, Sharland M, Berkley JA, Klein N, Malhotra-Kumar S, Bielicki J. Antimicrobial Resistance Following Azithromycin Mass Drug Administration: Potential Surveillance Strategies to Assess Public Health Impact. *Clin Infect Dis.* **2020**; .
  34. Keenan JD, Tadesse Z, Gebresillase S, et al. Mass azithromycin distribution for hyperendemic trachoma following a cluster-randomized trial: A continuation study of randomly reassigned subclusters (TANA II). *PLoS Med.* **2018**; .
  35. Bojang E, Jafali J, Perreten V, et al. Short-term increase in prevalence of nasopharyngeal carriage of macrolide-resistant *Staphylococcus aureus* following mass drug administration with azithromycin for trachoma control. *BMC Microbiol.* **2017**; .
  36. Coles CL, Mabula K, Seidman JC, et al. Mass distribution of azithromycin for trachoma control is associated with increased risk of azithromycin-resistant *Streptococcus pneumoniae* carriage in young children 6 months after treatment. *Clin Infect Dis.* **2013**; .
  37. Parker EPK, Praharaaj I, John J, et al. Changes in the intestinal microbiota following the administration of azithromycin in a randomised placebo-controlled trial among infants in south India. *Sci Rep.* **2017**; .
  38. Doan T, Hinterwirth A, Worden L, et al. Gut microbiome alteration in MORDOR I: a community-randomized trial of mass azithromycin distribution. *Nat. Med.* 2019.
  39. J.C. Seidman, C.L. Coles, Levens J, Mkocho H, Munoz B, West SK. Increased resistance to azithromycin in *e. coli* following mass treatment for trachoma control in rural tanzania. *Am J Trop Med Hyg.* **2012**; 87(5):178.
  40. Dean D, Rothschild J, Ruettinger A, Kandel RP, Sachse K. Zoonotic *Chlamydiaceae* species associated with trachoma, Nepal. *Emerg Infect Dis.* **2013**; .
  41. Ghasemian E, Inic-Kanada A, Collingro A, et al. Detection of *Chlamydiaceae* and *Chlamydia*-like organisms on the ocular surface of children and adults from a trachoma-endemic region. *Sci Rep.* **2018**; .

42. Burton MJ, Hu VH, Massae P, et al. What Is causing active trachoma? The role of nonchlamydial bacterial pathogens in a low prevalence setting. *Investig Ophthalmol Vis Sci.* **2011**; .
43. Burr SE, Hart JD, Edwards T, et al. Association between Ocular Bacterial Carriage and Follicular Trachoma Following Mass Azithromycin Distribution in The Gambia. *PLoS Negl Trop Dis.* **2013**; .
44. Hu VH, Macleod D, Massae P, et al. Non-chlamydial bacterial infection and progression of conjunctival scarring in trachoma. *Investig Ophthalmol Vis Sci.* **2018**; .
45. Pickering H, Palmer CD, Houghton J, et al. Conjunctival Microbiome-Host Responses Are Associated With Impaired Epithelial Cell Health in Both Early and Late Stages of Trachoma. *Front Cell Infect Microbiol.* **2019**; .
46. Su H, Watkins NG, Zhang YX, Caldwell HD. Chlamydia trachomatis-host cell interactions: Role of the chlamydial major outer membrane protein as an adhesin. *Infect Immun.* **1990**; .
47. Zhang J, Lietman T, Olinger L, Miao Y, Stephens RS. Genetic diversity of Chlamydia trachomatis and the prevalence of trachoma. *Pediatr Infect Dis J.* **2004**; .
48. Chin SA, Morberg DP, Alemayehu W, et al. Diversity of chlamydia trachomatis in trachoma-hyperendemic communities treated with azithromycin. *Am J Epidemiol.* **2018**; .
49. Dawson C, Wood TR, Rose L, Hanna L. Experimental Inclusion Conjunctivitis in Man: III. Keratitis and Other Complications. *Arch Ophthalmol.* **1967**; .
50. Tarizzo ML, Nataf R, Nabli B. Experimental inoculation of thirteen volunteers with agent isolated from inclusion conjunctivitis. *Am J Ophthalmol.* **1967**; .

**Table 1. Demographic and trachoma characteristics of complete and sequenced samples, Amhara, Ethiopia, 2011-2015**

	Complete dataset (n = 240)	Sequenced dataset (n = 99)
<b>Median age, range (years)</b>	3 (1-5)	3 (1-5)
<b>Female; n (%)</b>	224 (52.6)	48 (48.5)
<b>Zone; n (%)</b>		
East Gojam	100 (41.7)	43 (43.4)
North Gondar	22 (9.2)	12 (12.1)
South Gondar	69 (28.8)	26 (26.3)
Waghemra	49 (20.4)	18 (18.2)
<b>Median cluster TF<sup>1</sup> prevalence, range (%)</b>	58.8 (13.5-90.7)	58.7 (15.3-90.7)
<b>Median cluster TI<sup>2</sup> prevalence, range (%)</b>	14.9 (0.0-51.4)	15.4 (0.0-51.4)
<b>Median cluster Ct prevalence, range (%)</b>	28.0 (4.0-100.0)	24.0 (4.0-100.0)
<b>Median load of infection, range<sup>3</sup></b>	368.9 (27.29-2.49x10 <sup>6</sup> )	1431.2 (213.67-1.26x10 <sup>6</sup> )

<sup>1</sup>trachomatous inflammation-follicular, <sup>2</sup>trachomatous inflammation-intense, <sup>3</sup>elementary bodies per swab

Accepted Manuscript

**Table 2. Description of nucleotide polymorphisms and amino acid changes in ompA of Amharan Ct sequences.**

<i>ompA</i> Type (number)	Nucleotide Position and Reference Nucleotide (Amino Acid Position and Reference Amino Acid)											
	Serovar A										Serovar B	
	272G (91S)	305C (102A)	433A (145T)	736A (246I)	940A (314K)	943C (315P)	946G (316V)	955A (319T)	956C (319T)	286A (96T)	1132G (378A)	
A1 (5)	A (Asp)			G (Ile)								
A2 (2)	G (Ala)			G (Ile)								
A3 (21)	G (Ile)											
A4 (1)	G (Glu)				G (Ala)		A (Ile)		G (Val)		T (Val)	
A5 <sup>ref</sup> (22)												
A6 (1)	T (Val)											
B1 (6)											G (Ala)	A (Thr)
B2 (6)											A (Thr)	
B3 <sup>ref</sup> (35)												

<sup>ref</sup>reference type per serovar to determine classify variants

**Table 3. Linear regression analysis of predictors of district-level *ompA* diversity.**

Variable <sup>1</sup>	Univariate			Multivariate		
	$\beta^2$	SE <sup>3</sup>	p-value	$\beta$	SE	p-value
<i>Ct</i> infection prevalence	1.134	0.238	0.69x10 <sup>-4</sup>	0.402	0.114	0.002
TF prevalence	0.005	0.004	0.159	-0.004	0.004	0.275
TI prevalence	0.033	0.012	0.008	0.016	0.013	0.260

<sup>1</sup>district-levels prevalence estimates, <sup>2</sup> $\beta$  = regression coefficient, <sup>3</sup>SE = standard error

Accepted Manuscript

Figure 1

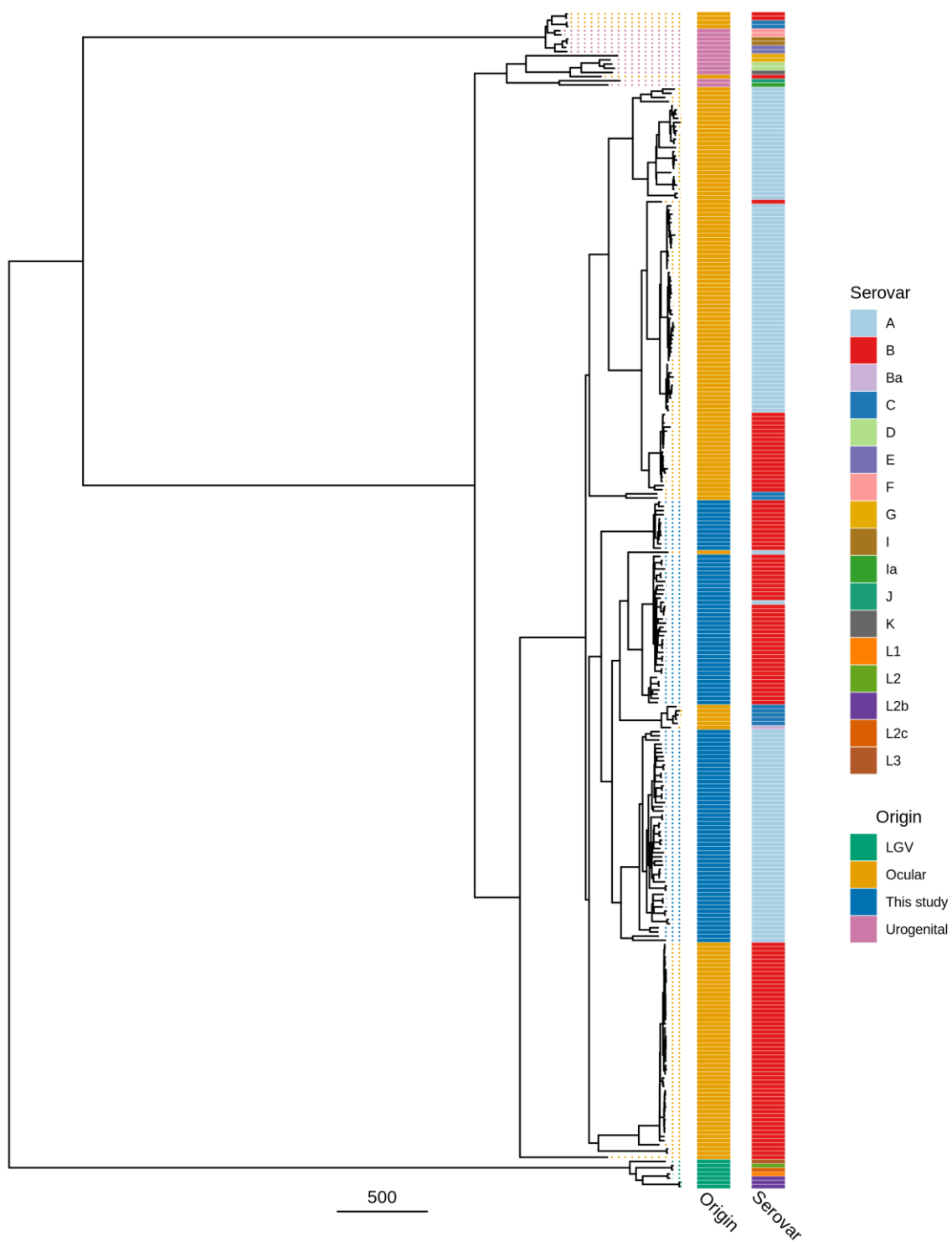
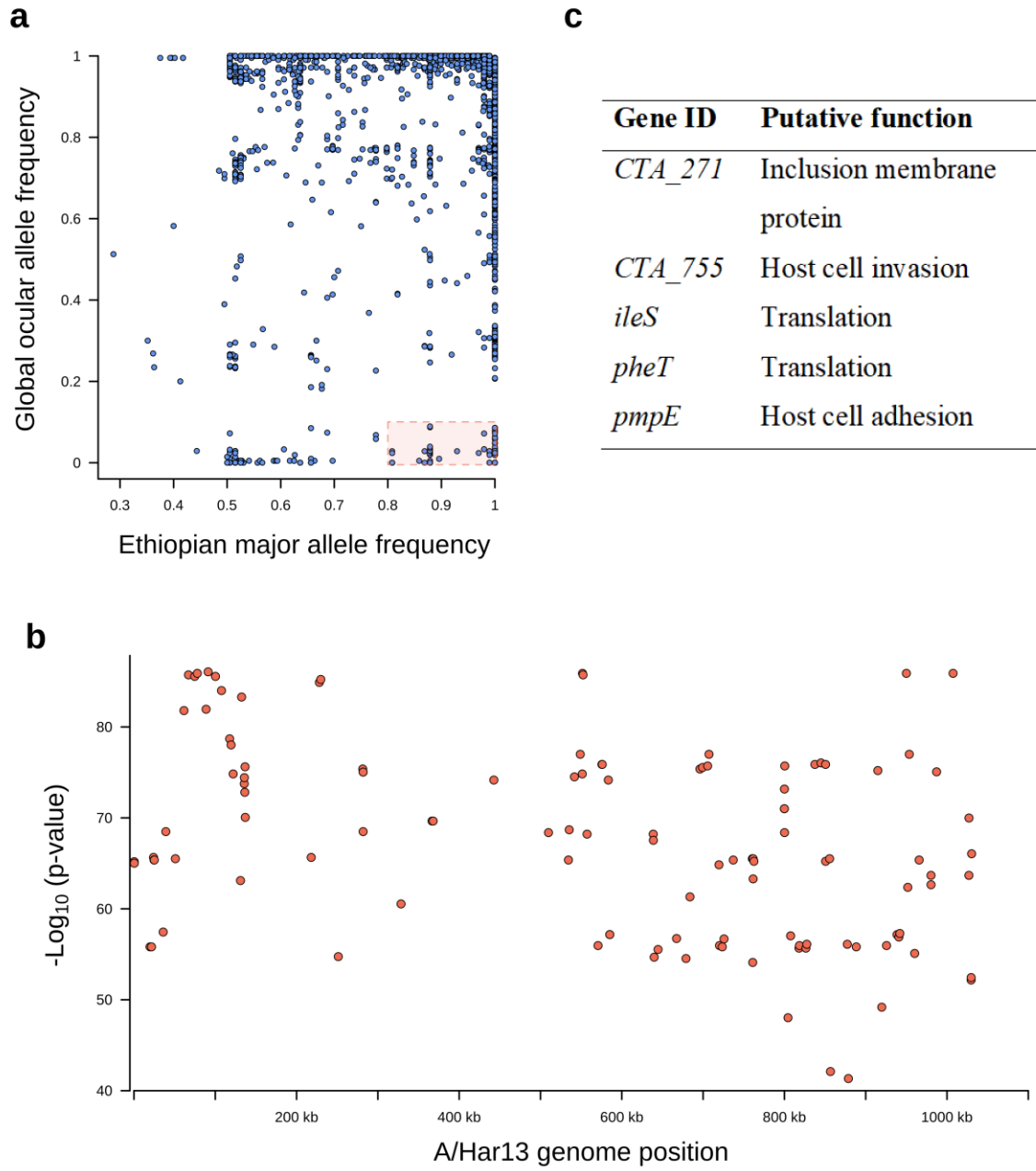


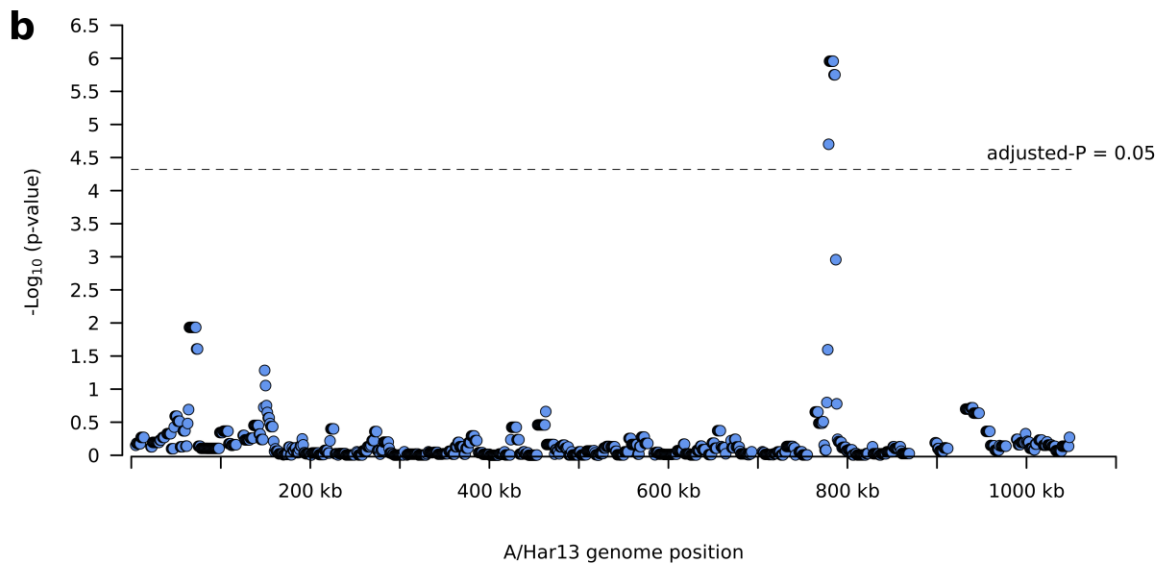
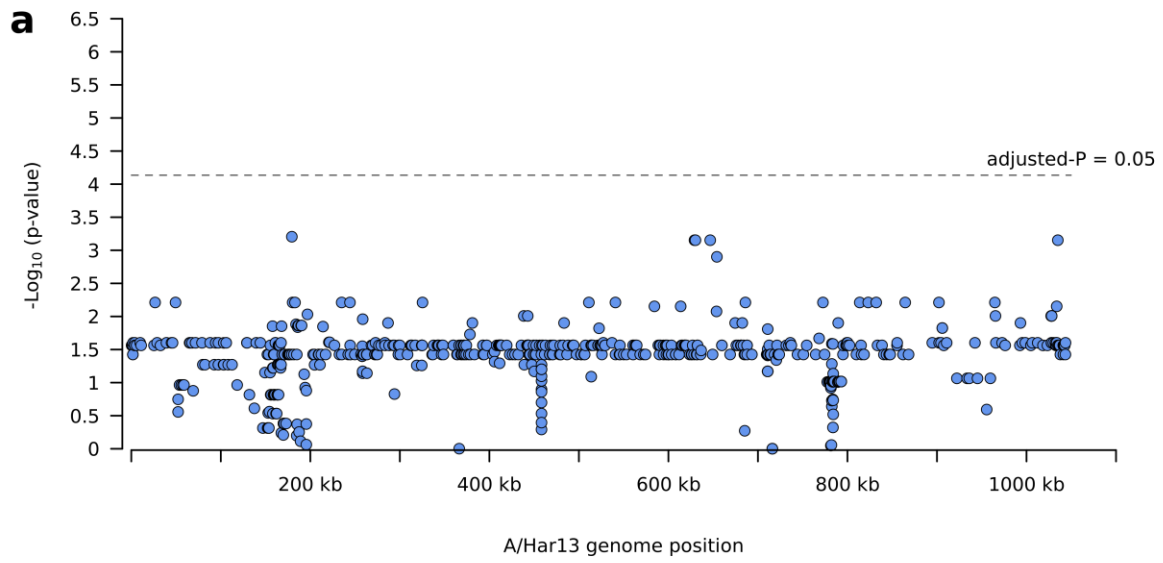
Figure 2



A



Figure 3



A

Accen

IScrip

Figure 4

