A study of the molecular and spatial determinants of ocular *Chlamydia trachomatis* infection on the trachoma-hyperendemic Bijagós Archipelago of Guinea Bissau, West Africa

Anna R. Last

2015

Clinical Research Department
Faculty of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine

Thesis submitted for the degree of
Doctor of Philosophy, University of London

The work contained in this thesis was supported by a Clinical PhD Training Fellowship grant from the Wellcome Trust (097330/Z/11/Z)

Declaration

I declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated within the thesis.

Signature Date: 5th September 2014

Anna R. Last

Abstract

Chlamydia trachomatis is the leading infectious cause of preventable blindness and the most common sexually transmitted bacterium worldwide. Trachoma presents an environment in which to investigate chlamydial pathogenicity, the conjunctivae serving as an accessible model with an objectively observable phenotype. The Bijagós Archipelago is a unique setting where trachoma is hyperendemic. The primary aims of this study were to use novel molecular, bioinformatic and geostatistical approaches in conjunction with population-based clinical and epidemiological metadata to investigate the micro-epidemiology of ocular *C. trachomatis* and active trachoma in this population.

The prevalence of trachoma and ocular *C. trachomatis* infection have been documented, and socio-environmental risk factors have been identified that may be important in the implementation of trachoma elimination activities in these communities.

A strong association was found between *C. trachomatis* ocular load (estimated using droplet digital PCR) and the level of conjunctival inflammation. Geostatistical analyses suggest that ocular *C. trachomatis* load may be important in transmission, as spatial clusters of high load infections were identified, whilst spatial clusters of low load infections were absent.

This study includes the first population-based pathogen genome-wide association scan (GWAS) for *C. trachomatis,* using high quality next generation whole genome sequence data obtained directly from clinical samples. The genome-wide associations with conjunctival inflammation (*incE*) and *C. trachomatis* load (*mutY* and *CTA_0271*) present genes involved in specific biological characteristics of *C. trachomatis,* the functions of which suggest that early interactions with host cells are important in *C. trachomatis* pathogenesis. Pathogen GWAS, applied in this context, is a powerful approach in the identification of multiple targets for further study in pathogenesis and directed study of potential vaccine candidates, allowing a greater understanding of association and interaction of genes on a genome-wide scale.

Following a single round of mass drug treatment with oral azithromycin (MDA) in these communities the prevalence of active trachoma and ocular *C. trachomatis* were significantly reduced. Individual and median loads of *C.*

trachomatis were reduced and the highest burden of disease and infection were concentrated in young children. Spatial clustering of infection identified using geostatistical tools was intensified following MDA, but the number of clusters of high load infections was reduced. The severity of conjunctival inflammation was reduced following MDA.

This study suggests that chlamydial load is important in disease pathogenesis and may be important in transmission of infection. Geospatial tools may be useful in the context of MDA to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission. The association between conjunctival inflammation and *C. trachomatis* load may reflect pathogen virulence. This is supported by the presence of genome-wide associations with *C. trachomatis* load and conjunctival inflammation identified by pathogen GWAS. Further epidemiological, *in vitro* and *in silico* studies are required to provide a more complete picture of the relationship between disease severity, bacterial load and chlamydial diversity in the context of transmission and elimination dynamics.

Preface

This thesis is presented as a 'Research Paper Style Thesis' in accordance with submission guidance provided by the London School of Hygiene and Tropical Medicine. The data chapters comprise papers that have been published, submitted for publication or that are in preparation for submission to particular peer-reviewed journals. The chapters prepared in this format are listed in bold italics in the ***Table of Contents*** on page 9 of the thesis. Publication details and acknowledgement of co-author contributions are included on the individual cover sheets for each paper. The remainder of the thesis is comprised of 'linking material' and includes a thorough introduction to the research study and study setting, a detailed section on field, clinical, laboratory and analytical methods employed and a summary of the research findings. All material within this thesis was written by Anna R. Last.

Acknowledgements

This work would not have been possible without the support and contribution of many people. I am grateful to my supervisors Robin Bailey and Martin Holland for their support, encouragement and enthusiasm for this study from the outset. It has been a privilege and inspiration to undertake this PhD as a member of the trachoma research group and benefit from the breadth of knowledge and experience of both my supervisors.

The LSHTM/Wellcome Trust Clinical PhD Fellowship in International Health provided me with incredible support and experience throughout this PhD. It has provided an ideal environment in which to progress with my future career in academic clinical medicine. It has also provided great opportunity to meet and collaborate with other academics and clinicians with a great breadth of interests and experience. I would like to thank the Wellcome Trust for supporting this study and David Mabey for encouraging my application for the Fellowship and subsequent mentorship.

I would like to thank other colleagues at LSHTM for their contributions towards this study. In particular Chrissy Roberts, who was instrumental in establishing the novel droplet digital PCR assay, used for the quantitation analyses in this study, and Taane Clark and Francesc Coll for their essential bioinformatics support in the analysis of the chlamydial whole genome sequence data. I would also like to thank Emma Harding-Esch, who although now works for Public Health England, has maintained an encouraging and enthusiastic interest throughout and has been a source of sound advice and provided thoughtful comments on many of the manuscripts resulting from this work. I would like to thank Susan Sheedy for her flexibility and efficiency in administering the Wellcome Trust grant through LSHTM.

Thanks are extended to Nick Thomson at the Wellcome Trust Sanger Institute for their collaboration on this project and sequencing of chlamydial strains for this study. Similarly I would like to thank Ian Clarke and Lesley Cutcliffe in the Department of Microbiology at the University of Southampton for their 'green fingers' with respect to the labour-intensive chlamydial cell culture carried out during pilot studies for this PhD.

The field studies on which this research is based would not have been possible without the supportive collaboration from colleagues in West Africa. I am indebted to my colleagues and field team on the Bijagos Archipelago, without whose enthusiasm,

interest and hard work, none of this would have been possible. It has been a pleasure to work in the most challenging field environment with the team. I am particularly thankful to my colleague and friend, Eunice Cassama, who has been part of this project from the beginning. Colleagues from the Programa Nacional de Saúde de Visão (Meno Nabicassa and Wilson Sa) have been incredibly supportive of the work we have undertaken.

I am grateful to colleagues at the MRC Unit (The Gambia), who assisted with vital infrastructure enabling this project to be conducted in such a remote location. Sarah Burr has been flexible and supportive in the organization of essential logistics through the MRC unit in Fajara.

I am hugely thankful for the support, friendship and welcome shown by colleagues at the Bandim Health Project in Bissau. Peter Aaby, Ane Fisker, Grethe Lemvik, Sofie Biering-Sorensen, Stine Byberg and Sanne Marie Thysen have been so helpful in providing insight into life in Bissau and advice on initial essential logistics in a challenging environment. I have great admiration for the work and longevity of the Bandim Health Project. I would also like to thank the survey participants and communities of the Bijagós Archipelago for endorsing and welcoming our study team and research.

Finally, I would not have been able to accomplish this without the love, encouragement, support and patience of my husband Luke Tapp.

Dedication

The work presented in this thesis is dedicated to my parents, Brian and Linden Last, and my husband Luke.

In memory of Margaret Catherine Last (1928-2013) and Boris (2011-2014).

Table of Contents

List of contributors to the research presented in this thesis

Name	Position/Institution	Contribution
Robin Bailey	Professor of Tropical Medicine, ITD, CTD, LSHTM	PhD Supervisor
Martin Holland	Reader in Immunology, ITD, CRD, LSHTM	PhD Co-Supervisor
Matthew Burton	Reader and Senior Wellcome Trust Research Fellow, in Clinical Science ITD, CRD, LSHTM	PhD Upgrading Examiner
Nick Dorrell	Senior Lecturer in Bacterial Pathogenesis, PMB, LSHTM	PhD Upgrading Examiner
David Mabey	Professor of Communicable Diseases, CRD, ITD, LSHTM	Director of Wellcome Trust Clinical PhD Fellowship Programme
Sarah Burr	Head of Chlamydial Laboratory, Disease and Elimination Theme, MRC The Gambia Unit, The Gambia	Molecular laboratory guidance and support. Logistic support through MRC The Gambia Unit.
Eunice Cassama	Ophthalmic Nurse, Programa Nacional de Saude de Visao, Ministerio de Saude Publica, Bissau, Guinea Bissau	Essential assistance with field survey work and administration of community mass drug treatment with azithromycin
Marito Gomes Cabral	Trainee Nurse, Programa Nacional de Saude de Visao, Ministerio de Saude Publica, Bissau, Guinea Bissau	Conjunctival digital photography.
Benedito Icossoboc	Field Team, Guinea Bissau	Essential field logistics and sensitization of communities
Eduardo Muscate	Field Team, Guinea Bissau	Essential field logistics and sensitization of communities
Ibraima Djassi	Field Team, Guinea Bissau	Essential field logistics.
Chrissy Roberts	Post-doctoral Researcher, ITD, CRD, LSHTM	Established the droplet digital PCR assay for detection and quantitation of *C. trachomatis* infections. Advised on analysis of GWAS data.
Taane Clark	Reader in Genetic Epidemiology and Statistical Genomics, PMB, LSHTM	Bioinformatics support for analysis of chlamydial whole genome sequence data.
Francesc Coll	PhD Student, PMB, LSHTM	Bioinformatics support for analysis of whole genome sequence data; significant contribution regarding development of scripts (and quality control algorithms) and use of appropriate software
Jody Phelan	PhD Student, PMB, LSHTM	Advice on recombination analyses for chlamydial sequence data (bioinformatics).
Helen Weiss	MRC Tropical Epidemiology Group, LSHTM	Statistical support for analysis of risk factor association data.
Neal Alexander	MRC Tropical Epidemiology Group, LSHTM	Statistical support for analysis of geospatial data.
Nick Thomson	Sulston Laboratories. Wellcome Trust Sanger Institute, Hinxton, UK	Development and use of DNA-bait sample enrichment technology to obtain chlamydial sequence data direct from clinical samples. Illumina sequencing conducted for samples included in this study. Bioinformatics support.
Ian Clarke	Department of Microbiology, University of Southampton, UK	Chlamydial culture of isolates from pilot studies.
Lesley Cutcliffe	Department of Microbiology, University of Southampton, UK	Chlamydial culture of isolates from pilot studies.

List of Tables presented in this thesis

Chapter 1

Chapter 5

Chapter 7

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Chapter 12

List of Figures presented in this thesis

Chapter 9

Chapter 10

Chapter 11

Chapter 1 Introduction: Trachoma and ocular Chlamydia trachomatis infection

1.1 Background

Infection with the obligate intracellular human pathogen *Chlamydia trachomatis* is the leading infectious cause of preventable blindness and the most common sexually transmitted bacterial infection, thus contributing to significant global morbidity and mortality. Trachoma is caused by ocular infection with *C. trachomatis* and manifests as distinct clinical syndromes beginning with an acute self-limiting keratoconjunctivitis then progressing to chronic inflammatory disease with subsequent conjunctival scarring and blinding sequelae. In 1997 the World Health Organization (WHO) launched a global effort to eliminate blinding trachoma through the implementation of the SAFE (Surgery for trichiasis, Antibiotics for active disease, Facial cleanliness and Environmental improvement) strategy to reduce transmission of infection.

1.2 Epidemiology and transmission

Trachoma is endemic in some of the poorest countries in the world, predominantly in poor rural areas of Africa and Asia, often in marginalised, displaced and remote populations. It is endemic in over 50 countries, with the highest prevalence of active disease and trichiasis in sub-Saharan Africa, in the savannah areas of East and Central Africa and in the Sahel of West Africa [1]. In these regions the distribution of disease is heterogeneous [2].

Ocular *C. trachomatis* is probably transmitted between individuals through direct spread from eye to eye during close contact, direct or indirect spread of infected nasal or ocular secretions on dirty fingers or cloths (fomites) and indirect passive transmission by eye-seeking flies. There is no known animal reservoir of *C. trachomatis* in endemic environments.

Many studies have suggested that the prevalence of disease is associated with environmental risk factors such as poor sanitation, access to water and latrine use [3, 4]. Eye seeking flies (*Musca sorbens*) have also been associated with disease as a passive vector [5,6] but significant disease exists in areas where fly populations are scarce and are therefore less likely to contribute to disease transmission [7]. *M. sorbens* breeds in human faeces and there may be overlap between fly populations and lack of latrine access or use [3].

Social risk factors such as migration events and crowded living conditions have also been shown to be important in transmission of *C. trachomatis* and the appearance of active disease [8,9].

Clustering of disease at the community, household and bedroom levels has been noted and is likely to reflect the dynamics of transmission between family members with prolonged close contact and the infectious nature of *C. trachomatis* [3,8-10]. Most transmission events occur at the household level with more gradual spread within the community [11]. Clustering of infection with *C. trachomatis* has also been demonstrated at household level [12].

The epidemiological dynamics of trachoma transmission are complex and not well understood. Trachoma endemicity is likely to be important, such that in hypoendemic populations disease may naturally decline [13-16] or disappear after a single round of community mass treatment with antibiotics [17,18]. In mesoendemic populations, disease prevalence may stabilize following community mass antibiotic treatment [19], and in hyperendemic populations disease and infection persist despite continued rounds of community mass treatment [20]. Mathematical models have been developed to attempt to capture these transmission dynamics and have focused on the effects of mass treatment on disease prevalence [21-24]. Others have examined household transmission [11,21] and the effect of repeated infection in the pathogenesis of trachoma [25]. Thus far it has not been possible to model the underlying mechanisms for positive feedback seen in trachoma transmission, where the hazard of a susceptible individual becoming infected (per case of infection) increases with increasing endemicity [26]. By incorporating individual-level factors such as socio-behavioural contact, *C. trachomatis* bacterial load, *C. trachomatis* strain diversity and superinfection, it may be possible to develop models that are able to account for these positive feedback effects, taking infection and disease endemicity into account. The use of *in silico* predictive models will complement the longitudinal study of transmission, which is time-consuming, expensive and labour-intensive, and allow for modeling of transmission within the context of treatment interventions.

1.3 Clinical features and natural history of disease

Trachoma is a progressive disease. Acute keratoconjunctivitis occurs most commonly in early childhood and is often referred to as ‘active disease’ or ‘active trachoma’. This stage of disease is characterised by a follicular conjunctivitis with inflammation. Follicles (subepithelial clusters of lymphoid cells) are easily visible on

the everted upper tarsal conjunctiva as circular yellow-white elevations of >0.5mm diameter. Papillary hypertrophy (engorgement of small vessels with surrounding oedema representing conjunctival inflammation) may also occur and obscure underlying deep tarsal vasculature if severe. Infected individuals may be asymptomatic or have only mild symptoms despite variable signs of disease on clinical examination. Repeated infection with *C. trachomatis* causes recurrent inflammatory episodes, the resolution of which may lead to upper tarsal conjunctival scarring. Scarring from follicle resolution at the corneal limbus results in the formation of shallow depressions known as 'Herbert's pits', which are the only pathognomonic sign of trachoma. As the scarring progresses, the conjunctiva becomes distorted, resulting in the inward turning of the lid margin (entropion) causing contact between the eyelashes and the globe of the eye (trichiasis). Corneal opacification is caused by corneal abrasion from inturned lashes, which leads to the blinding sequelae of trachoma. The direct abrasive effects of inturned lashes leads to an increased likelihood of secondary bacterial and fungal infections, compounding the patterns of chronic inflammation and scarring [27].

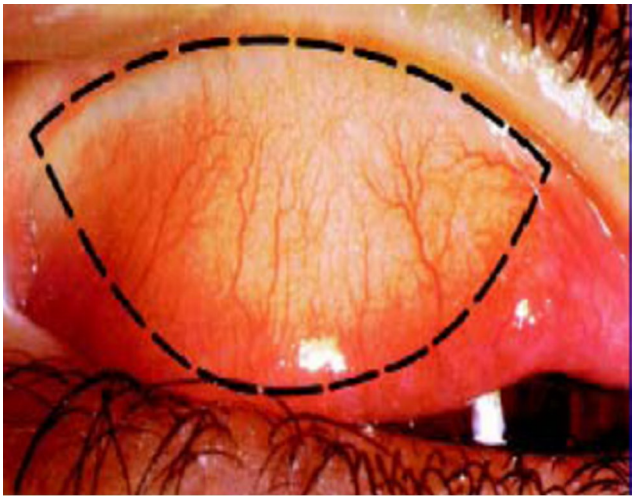
The majority of cases of active disease and ocular infection with *C. trachomatis* occur in childhood with a decline in prevalence as individuals enter adulthood [10]. Adult bacterial load with *C. trachomatis* is lower than in children and duration of infection and disease episodes are shorter, which may in part be due to an acquired immune response [21,28]. Conversely, the prevalence of scarring sequelae increases with age, reflecting cumulative insult from recurrent infection and inflammation. The pathogenesis of trachomatous scarring is complex, involving both host and pathogen factors. In areas where the prevalence of active disease is very high, scarring sequelae may be evident in younger age groups [29]. Recurrent infection and persistent inflammatory disease increase the likelihood of developing scarring sequelae [3,30].

1.4 Clinical diagnosis and grading systems

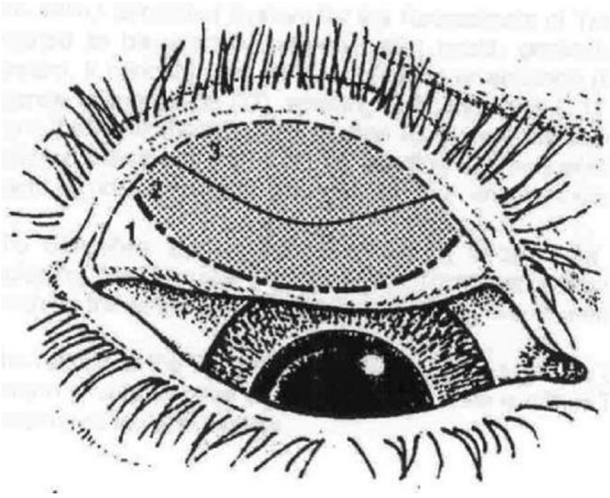
Examination of the eye for clinical signs of trachoma involves inspection of the lashes, the cornea and limbus and the upper tarsal conjunctiva (after everting the eyelid) (*Figure 1*). A binocular magnifying loupe (x2.5) and adequate lighting are required, but this is a simple examination that can be conducted in field conditions.

Figure 1. Zones of the upper tarsal conjunctiva used in trachoma grading

(A)



(B)



The entire central zone of the upper tarsal conjunctiva is examined when using the WHO simplified grading system (A) [31]. The upper tarsal conjunctiva is further divided into zones (1-3) used for scoring with the modified FPC (Follicles, Papillae, Conjunctival scarring) system (B) [33].

Grading systems have been employed to standardize the diagnosis of trachoma in field surveys and research studies. There are a number of different systems referred to in the literature [2]. The system currently used by trachoma control programmes is the WHO simplified grading system (*Table 1*) [31]. This system has shown to have good inter-observer reliability, which is not dependent on level of training of the examiner [32] and is easy to use and suitable for field studies, yielding useful information on the prevalence of active and scarring disease.

Table 1. The WHO Simplified Grading System [31]

Grade	Description
Trachomatous inflammation – follicular (TF)	Presence of ≥ 5 follicles (>0.5mm diameter) on the upper tarsal conjunctiva
Trachomatous inflammation – intense (TI)	Pronounced inflammatory thickening of the upper tarsal conjunctiva, causing 50% of the deep tarsal vessels to be obscured
Trachomatous scarring (TS)	Presence of scarring on the upper tarsal conjunctiva
Trachomatous trichiasis (TT)	At least one eyelash rubbing the globe
Corneal opacity (CO)	Easily visible central corneal opacity over the pupil

For research purposes, a more detailed system is often used such as the modified FPC (Follicles, Papillary hypertrophy, Conjunctival scarring) system developed by Dawson *et al.* [33] (*Figure 1* & *Table 2*) to 'describe more precisely the intensity of inflammatory trachoma and potentially disabling and disabling lesions'.

Table 2. The Modified FPC (Follicles, Papillary hypertrophy, Conjunctival scarring) Grading System [33]

Grade	Description
F	***Upper tarsal follicles***
F0	No follicles
F1	Follicles present but ≤ 5 in total in zones 2 and 3 together
F2	>5 follicles in zones 2 and 3 but less than 5 in zone 3
F3	≥5 follicles in each of zones 1, 2 and 3
P	***Papillary hypertrophy and diffuse infiltration***
P0	Absent; normal appearance
P1	Minimal: individual vascular tufts (papillae) prominent but deep subconjunctival tarsal vessels not obscured
P2	Moderate: more prominent papillae, normal vessels appear hazy, visible to naked eye
P3	Severe: conjunctiva thickened and opaque, more than 50% deep tarsal vessels obscured
C	***Conjunctival scarring***
C0	No scarring
C1	Mild: fine scattered scars on upper tarsal conjunctiva OR scars on other parts of conjunctiva
C2	Moderate: more severe scarring but without distortion of upper tarsus
C3	Severe: conjunctival scarring with distortion of upper tarsus
T/E	***Trichiasis and/or Entropion***
T/E0	No trichiasis or entropion
T/E1	Lashes deviated towards eye but not touching globe
T/E2	Lashes touching the globe but not rubbing the cornea
T/E3	Lashes constantly rubbing the cornea
CC	***Corneal scarring***
CC0	Absent
CC1	Minimal scarring or opacity but not involving the visual axis; clear central cornea
CC2	Moderate scarring or opacity involving the visual axis; papillary margin visible through opacity
CC3	Severe central scarring or opacity with papillary margin not visible through opacity

The WHO simplified and the modified FPC grading systems are often said to be directly comparable, allowing derivation of WHO simplified grades from FPC grades without further assessment of the patient. A comparison of the two systems is illustrated in *Table 3*. However, follicle size is not taken into account with the FPC system and the definitions of conjunctival scarring are not precise for either system, making the comparison for scarring grades more problematic.

Table 3 Comparison of the simplified and detailed grading systems for trachoma

Simplified (WHO simplified)	**Detailed (modified FPC)**	**Implication**
TF	F2 or F3	Presence of inflammatory trachoma
TI	P3	Severe intensity of inflammation
TS	C1*, C2 or C3	Presence of scarring trachoma
TT	T/E2 or T/E3	Potentially visually disabling lesion
CO	CC2 or CC3	Visually disabling lesion

*C1 may include scarring outside the central upper tarsal conjunctiva and is therefore not directly compatible

1.5 Trachoma control and elimination

In communities hyperendemic for trachoma a community mass treatment approach is advocated. The Global Alliance for the Elimination of Blinding Trachoma by the year 2020 (GET2020) was established by the WHO in 1997. The SAFE strategy is the focus of these control activities. Although the integration of all components of SAFE are thought to be important, most control programmes focus on surgery and antibiotic administration. Implementation of SAFE aims to eliminate blinding trachoma as a public health problem.

Surgery for trichiasis

Surgical correction of trichiasis may reduce the risk of corneal opacification and resultant blindness. Indications for and timing of surgery vary between programmes. The outcome of surgery is dependent on technique and operator, pre-operative disease severity and post-operative infection. Recurrence may occur in up to 40% of patients depending on duration of follow-up, though most recurrence is reported to occur within the first six months of surgery [34].

Antibiotics for infection

Antibiotics may be administered at the community and individual levels with the aim of reducing infection rates within communities or households, assuming that

transmission will be interrupted and subsequent scarring disease reduced. However, all trachoma prevalence estimates are based on clinical signs of trachoma rather than evidence of infection with *C. trachomatis*. The WHO currently recommends mass treatment with azithromycin on the basis of community prevalence of follicular trachoma (TF) in children aged 1-9 years of age (*Table 4*). Despite current WHO recommendations on frequency and duration of treatment, optimal parameters are uncertain. WHO recommended treatments are currently single dose oral azithromycin (20mg/kg up to a maximum dose of 1g) or tetracycline eye ointment (TEO) (applied twice daily for six weeks). This recommendation follows results from a large multi-centre study in three endemic countries where community mass treatment with azithromycin resulted in a marked reduction in the prevalence of chlamydial infection, sustained for 12 months' follow-up [35]. Other studies have found similar results [18]. Currently, 19 trachoma endemic countries have received over 225 million doses of azithromycin (Zithromax®) through the International Trachoma Initiative (ITI) as a philanthropic donation from Pfizer.

Table 4. WHO criteria for mass antibiotic treatment distribution [36]

Prevalence of TF in 1-9 year olds	**Recommendation**
District level ≥ 10%	Mass treat district annually for 3 years and then re-evaluate prevalence in district
District level < 10%	Community level assessment
Community level ≥ 10%	Mass treat community annually for 3 years and then re-evaluate prevalence in community
Community level ≥ 5% but < 10%	Target treatment to affected children and the members of their household
Community level < 5%	Antibiotic treatment not recommended

Azithromycin is not licensed for use in pregnancy. There is no evidence thus far suggesting that it is harmful. This is supported by other studies on the treatment of sexually transmitted diseases [37]. Azithromycin is not used in infants under the age of six months; an alternative six-week regime of TEO is recommended by the WHO.

Azithromycin is well tolerated and safe and has additionally been shown to have a beneficial effect on morbidity and mortality [38-40]. There does not appear to be any evidence of azithromycin resistance with respect to *C. trachomatis*. Transient development of resistance that is not felt to be clinically significant has been noted in *Streptococcus pneumoniae* [41,42].

Promotion of facial hygiene and environmental improvement

Promotion of facial cleanliness and hygiene and environmental improvement are thought to be important, but the evidence-base is lacking and the addition of these components to a control programme is labour-intensive and not always effective [5,43].

1.6 Ocular infection with Chlamydia trachomatis

1.6.1 Detection of infection with C. trachomatis

Trachoma control programmes currently rely on the clinical signs of disease for diagnosis and to decide when treatment should be initiated in accordance with local and WHO policy. In research studies it is often important to know the individual's infection status. There is no 'gold standard' in the diagnosis of infection with *C. trachomatis*. Cell culture of *C. trachomatis* is expensive, labour-intensive, technically challenging and lacks sensitivity [2]. Previously used tests include microscopy with Giemsa staining for inclusion bodies on a conjunctival smear and direct immunofluorescence with monoclonal antibodies to *C. trachomatis* antigens. Serology using enzyme-linked immunoassays (EIA) has moderate sensitivity, but lacks specificity due to cross-reactivity with other bacteria [2]. Nucleic acid amplification tests (NAAT) using a polymerase chain reaction (PCR) or transcription-mediated amplification (TMA) are the current tests of choice in diagnosing infection with *C. trachomatis*. NAAT are highly sensitive and specific, due their ability to amplify low levels of specific target DNA or RNA template, and are able to identify significantly more individuals with chlamydial infection within a trachoma endemic population than had previously been identified. Quantitative PCR (qPCR) has also been employed to measure bacterial load in endemic populations [18]. However, expense and complexity precludes these tests from use in the programmatic context at present, with the ultimate panacea being a simple, field-friendly NAAT-based point of care test.

1.6.2 Infection with C. trachomatis and the presence of clinical signs

C. trachomatis is thought to be the primary stimulant triggering inflammatory trachoma. However, *C. trachomatis* is not detected in all cases of active disease [3,44,45] even using highly sensitive NAAT. *C. trachomatis* may be detected in individuals at different stages of disease in hyperendemic communities, including individuals with no signs of clinical disease. In communities with low trachoma

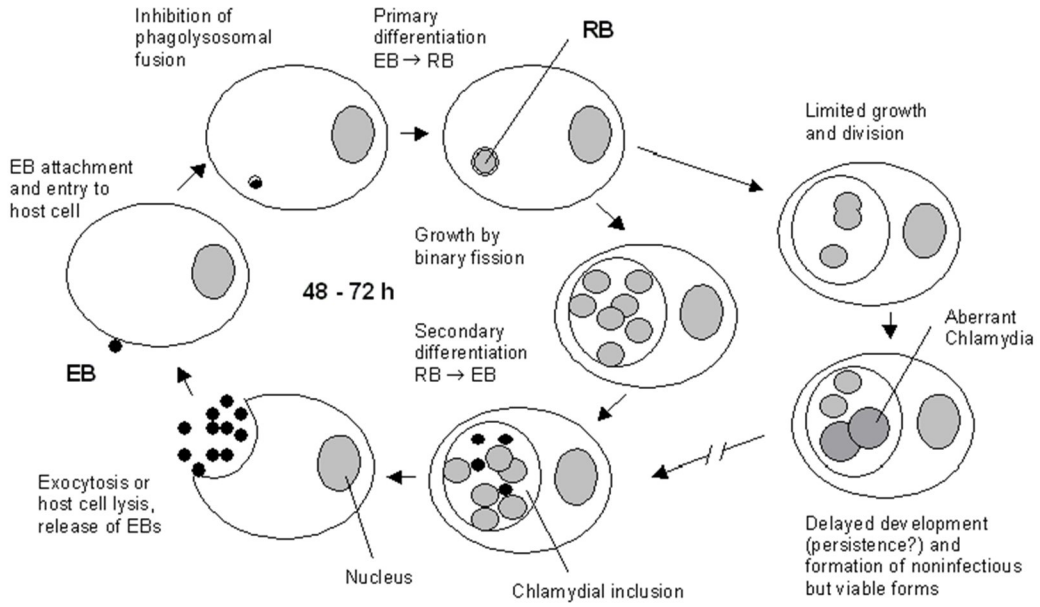
prevalence, *C. trachomatis* is only found in a minority of cases of active disease, and in some areas infection is not found at all. Individuals with intense trachomatous inflammation are more likely to be infected and have higher bacterial loads than those with follicular disease [2,3].

There may be an incubation period where infection is present but disease has not yet manifest. There is a lag time between clearance of infection and resolution of clinical signs [28], which may last for several weeks or longer, and may indicate the impact of the host immune response in the disease process. Age is also a determinant of how long disease and infection episodes persist, durations being longer in children. It is also hypothesized that a subclinical persistent infection may exist, where *C. trachomatis* does not replicate and thus may not induce clinical signs [46]. The clinical signs of conjunctival folliculitis and inflammation are not specific for trachoma and occur in conjunctivitis caused by a number of microorganisms including viruses, Moraxella spp. and other chlamydiae [47]. Detection of chlamydial infection by NAAT is also dependent on factors such as swab collection, storage and preparation and ultimately only tests for the presence of chlamydial DNA or RNA, thus a transient inoculation may also give a positive result that may not represent an established infection.

1.6.3 Characterisation of C. trachomatis

C. trachomatis is an obligate intracellular bacterium. During its development, *C. trachomatis* exists in two principal forms: the metabolically active intracellular reticulate bodies (RB) and the metabolically inactive extracellular elementary bodies (EB). The RB may also exist in a 'cryptic' state where it remains in a metabolically inactive arrested phase [48]. It is usually the EB which is the form facilitating transfer between host cells and chlamydiae. The *C. trachomatis* developmental cycle is shown in *Figure 2*.

Figure 2. Developmental Cycle of Chlamydiae [48]



C. trachomatis was first isolated in 1957 by Tang *et al*. [49], and was initially thought to be of viral origin. Extensive efforts have been made to characterise this important human pathogen into different types or strains. The identification of chlamydial strains is of vital importance to epidemiological studies of chlamydial infection, defining transmission, establishing test of cure parameters and in comparing the roles of persistent infection with reinfection. Currently there is lack of clarity in how to completely define chlamydial strains in a meaningful way that is comparable between studies. This has been due to the limited resolution of typing strategies that are not able to encompass the extent of full genomic diversity. The limited resolution of current single or multi-gene chlamydial sequence-based typing systems means that they are limited in their usefulness in studies focusing on the development of sequelae or strain-dependent virulence associations.

Two *C. trachomatis* biovars, trachoma and lymphogranuloma, (distinguishable based on physiological characteristics) have been described. Serotyping can further classify these biovars on the basis of possession of sequence-binding epitopes in their major outer membrane protein (MOMP) that are recognized by sera or monoclonal antibodies (mAbs). There are approximately 19 serovars (trachoma (A-C), sexually transmitted infection (D-K) and lymphogranuloma (L1-3)) [50]. Genetic sequence differences for the gene encoding MOMP (*ompA*) reflect the 15 MOMP serovars described originally by Wang and Grayston [51]. These strains

were further defined by epitope-mapping of *ompA* genes with a panel of mAbs [52]. This and other studies have resulted in *ompA* genotyping superceding serotyping due to the detail afforded by genotyping in the classification of strains that are unable to be serotypically characterised [53]. Subsequently, *ompA* genotyping data have expanded the classification of chlamydial strains to >20 genovars [54].

MOMP is a member of a family of genetically related surface-exposed outer membrane proteins. MOMP functions as a structural protein [55], porin [56-58] and cytoadhesion molecule [59-60]. The most widely studied and accepted function of MOMP is as a general diffusion porin [56]. Structural studies have been challenging due to the low solubility of MOMP, rendering structural details less clear [61,62]. Further knowledge of its structure will be crucial to understanding its role in the pathogenesis of chlamydial infection.

Since variable MOMP domains recognized by mAbs exist on the surface of the EB, immune pressure may drive its variability [63-66]. MOMP has also been suggested as a virulence factor with pathogenic roles including bacterial adhesion, invasion and intracellular survival, evasion of host defenses and stimulation of pro-inflammatory cytokines [65]. MOMP mosaics and natural recombination generate variability which may have pathogenic potential [53]. *OmpA* has been shown to be highly variable, and a hotspot for recombination [67]. In light of these phenomena it is not surprising that the analysis of *ompA* sequence data has failed to reveal convincing pathobiologic distinctions between genovars [66,68]. Occasional reports have suggested a link between disease severity and particular genovars [69], but these studies are limited in their methodology and sample size [54]. Strain typing using MOMP and *ompA* provides only a partial insight into strain variation and the pathobiology of *C. trachomatis* and thus other methods must be used to establish whether there are relationships between chlamydial strains, virulence and disease severity.

Other PCR-based genotyping methods with higher resolution have been developed: Multi-locus sequence typing (MLST) and Multiple loci variable number of tandem repeats (VNTR) analysis (MLVA). MLVA utilizes combined PCR and sequencing of *ompA* plus three VNTR loci to obtain higher diversity between strains of *C. trachomatis* [70]. In one study the resultant diversity showed a Simpson's index of diversity (D) of 0.94 (approaching the recommended index value for a discriminatory typing method (D=0.95) [71,72]) and the loci were found to be stable.

A classical MLST system involving seven housekeeping genes is probably most suited to analyse evolutionary changes and was not assessed for discriminatory capacity [73]. An alternative MLST uses sequence determination of five genomic targets plus *ompA* [74]. One study comparing current genotyping systems found the diversity index to be 0.82, 0.81, 0.95, 0.96 and 0.97 for *ompA*, classical MLST, alternative MLST, VNTR alone and MLVA respectively [75]. Another study found similar results with D being 0.78 for *ompA* and 0.95-0.99 for alternative MLST, MLVA and a combination of MLST+MLVA [76]. These systems are advantageous in that it is possible to conduct the tests on clinical specimens from routine NAAT diagnostics. It is less clear whether MLVA or MLST are sufficiently discriminatory in answering questions relating to disease severity and whether this level of discrimination would also be true for ocular isolates (as these studies refer to urogenital strains of *C. trachomatis*).

The first whole genome sequence of a serovar D *C. trachomatis* isolate was published by Stephens *et al.* in 1998, revealing a small (~1-Mb) highly conserved genome [77]. Since then much has been learned about chlamydial genetics, including the observation that current typing systems may be misleading when the underlying whole genome sequence is analysed [78]. This whole genome sequence (WGS) analysis by Harris *et al.* demonstrated that there is extensive recombination within the *ompA* gene, which can occur partially or completely both within and between biovars (also noted in other studies [67,79]) and that there is also evidence for genetic exchange and recombination within the cryptic plasmid. This evidence may explain why many studies have failed to find a significant association between disease severity and strain type [80,81]. However, strain-specific differences using WGS data related to *in vitro* virulence and *in vivo* clinical disease severity (in non-human primates) have been demonstrated [82]. This evidence of recombination in clinical strains has challenged our knowledge of chlamydial diversity and highlighted the implications of recombination in monitoring and epidemiological studies using other methods of strain typing.

WGS and comparative genomics have already been shown to be an effective strategy to discover loci associated with disease severity and virulence [78,79,81]. This is of particular importance as although genetic transformation of *C. trachomatis* has been achieved [83,84], working with chlamydiae in the laboratory remains inherently problematic, which severely limits definitive work on *in vitro* studies of

chlamydial virulence factors. Thus coupling WGS analysis with clinical phenotype and disease epidemiology may enable further characterization of *C. trachomatis*. Putative virulence factors identified include the polymorphic outer membrane protein (*pmp*), TARP and cytotoxin genes, chlamydial type III secretion systems (TTSS), plasticity zone mutants (such as *trpB*) and the cryptic plasmid [54].

1.7 Chapter 1 References

[1] World Health Organization (2012). Prevention of blindness and visual impairment (Priority Eye Diseases): Trachoma. https://www.who.int/blindness/cause/priority/en/index2.html.

[2] Solomon AW, Peeling RW, Foster A, Mabey DCW (2004). Diagnosis and Assessment of Trachoma. Clin Micro Rev 17(4):982-1011.

[3] Burton MJ, Holland MJ, Faal N et al. (2003). Which members of a community need antibiotics to control trachoma? Conjunctival *Chlamydia trachomatis* infection load in Gambian villages. Invest Ophthalmol Vis Sci 44:4215-4222.

[4] Emerson PM, Cairncross S, Bailey RL et al. (2000). Review of the evidence base for the 'F' and 'E' components of the SAFE strategy for trachoma control. Trop Med Int Health 5:515-527.

[5] Emerson PM, Lindsay SW, Alexander N et al. (2004). Role of flies and provision of latrines in trachoma control: cluster randomized controlled trial. Lancet 363:1093-1098.

[6] Jones BR (1975). The prevention of blindness from trachoma. Trans Ophthalmol Soc UK 95:16-33.

[7] Taylor HR, Velasco FM, Sommer A (1985). The ecology of trachoma: an epidemiological study in southern Mexico. Bull World Health Organ 63:559-567.

[8] Bailey R, Osmond C, Mabey DC et al. (1989). Analysis of the household distribution of trachoma in a Gambia village using a Monte Carlo simulation procedure. Int J Epidemiol 18:944-951.

[9] Burton MJ, Holland MJ, Makalo P et al. (2005). Re-emergence of *Chlamydia trachomatis* infection after mass antibiotic treatment of a trachoma-endemic Gambian community: a longitudinal study. Lancet 365:1321-8.

[10] West SK, Munoz B, Turner VM et al. (1991). The epidemiology of trachoma in central Tanzania. Int J Epidemiol 20:1088-1092.

[11] Blake IM, Burton MJ, Bailey RL et al. (2009). Estimating household and community transmission of ocular *Chlamydia trachomatis*. PLoS Negl Trop Dis 3:e401.

[12] Broman AT, Shum K, Munoz B et al. (2006). Spatial clustering of ocular chlamydial infection over time following treatment among households in a village in Tanzania. Invest Ophthalmol Vis Sci 47(1):99-104.

[13] Dolin PJ, Faal H, Johnson GJ et al. (1997). Reduction of trachoma in a sub-Saharan village in the absence of a disease control programme. Lancet 349(9064):1511-1512.

[14] Hoeschmann A, Metcalfe N, Kanjaloti S et al. (2001). Reduction of trachoma in the absence of antibiotic treatment: evidence from a population-based survey in Malawi. Ophth Epidemiol 8(2-3):145-153.

[15] Jha H, Chaudary J, Bhatta R et al. (2002). Disappearance of trachoma in Western Nepal. Clinical Infect Dis 35(6):765-8.

[16] Chidambaram JD, Alemeyehu W, Melese M et al. (2006). Effect of a single mass antibiotic distribution on the prevalence of infectious trachoma. JAMA 295(10):1142-1146.

[17] Gaynor BD, Miao Y, Cevallos V et al. (2003). Eliminating trachoma in areas with limited disease. Emerg Infect Dis 9(5):596-8.

[18] Solomon AW, Holland MJ, Alexander ND et al. (2004). Mass treatment with single dose azithromycin for trachoma. NEJM 351(19):1962-71.

[19] West SK, Munoz B, Mkocha H et al. (2005). Infection with *Chlamydia trachomatis* after mass treatment of a hyperendemic community in Tanzania: a longitudinal study. Lancet 366(9493):1296-1300.

[20] Lakew T, House J, Hong KC et al. (2009). Reduction and return of infectious trachoma in severely affected communities in Ethiopia. PLoS Negl Trop Dis 3(2):e376.

[21] Grassly N, Ward ME, Ferris S, Mabey DC, Bailey RL (2008). The natural history of trachoma infection and disease in a Gambian cohort with frequent follow-up. PLoS Negl Trop Dis 2:c341.

[22] Lee DC, Chidambaram JD, Porco TC, Lietman TM (2005). Seasonal effects in the elimination of trachoma. Am J Trop Med Hyg 72:468-70.

[23] Lietman T, Porco T, Dawson C, Blower S (1999). Global elimination of trachoma: how frequently should we administer mass chemotherapy? Nat Med 5:572-6.

[24] Ray KJ, Porco TC, Hong KC et al. (2007). A rationale for continuing mass antibiotic distributions for trachoma. BMC Infect Dis 7:91.

[25] Gambhir M, Basanez M, Burton MJ et al. (2009). The development of an age-structured model for trachoma transmission dynamics, pathogenesis and control. PLoS Negl Trop Dis 3(6):e462.

[26] Lietman TM, Gebre T, Ayele B et al. (2011). The epidemiological dynamics of trachoma may facilitate elimination. Epidemics 3(2):119-24.

[27] Hu VH, Massae P, Weiss HA et al. (2011). Bacterial infection in scarring trachoma. Invest Ophthalmol Vis Sci 52(5):2181-6.

[28] Bailey RL, Duong T, Carpenter R et al. (1999). The duration of human ocular *Chlamydia trachomatis* infection is age dependent. Epidemiol Infect 123:479-486.

[29] Ngondi J, Ole-Sempele F, Onsarigo A et al. (2006). Blinding trachoma in post-conflict southern Sudan. PLoS Medicine 3:e478.

[30] West SK, Munoz B, Nkocha H et al. (2001). Progression of active trachoma to scarring in a cohort of Tanzanian children. Ophthalmic Epidemiol 8:137-144.

[31] Thylefors B, Dawson CR, Jones BR et al. (1987). A simple system for the assessment of trachoma and its complications. Bull WHO 65:477-483.

[32] Taylor HR, West SK, Katala S, Foster A (1987). Trachoma: Evaluation of a new grading scheme in the United Republic of Tanzania. Bull WHO 65:485-8.

[33] Dawson CR, Jones BR, Tarizzo ML (1981). Guide to trachoma control in programs for the prevention of blindness. WHO, Geneva.

[34] Rajak SN, Makalo P, Sillah A et al. (2010). Trichiasis surgery in The Gambia: A 4-year prospective study. Invest Ophthalmol Vis Sci 51(10):4996-5001.

[35] Schachter J, West SK, Mabey D et al. (1999). Azithromycin in the control of trachoma. Lancet 354:630-635.

[36] WHO-ITI (2004). Joint Research Agenda Meeting for the Elimination of Blinding Trachoma. World Health Organization, Geneva.

[37] Gray RH, Wabwire-Mangen F, Kigozi G et al. (2001). Randomized trial of presumptive sexually transmitted disease therapy during pregnancy in Rakai, Uganda. Am J Obstet Gynecol 185:1209-1217.

[38] Whitty CJ, Glasgow KW, Sadiq ST et al. (1999). Impact of community-based mass treatment for trachoma with oral azithromycin on general morbidity in Gambian children. Pediatr Infect Dis *J* 18:955-958.

[39] Sadiq ST, Glasgow KW, Drakeley CJ et al. (1995). Effects of azithromycin on malariometric indices in The Gambia. Lancet 346:881-882.

[40] Bhosai SJ, Bailey RL, Gaynor BD, Lietman TM (2012). Trachoma: An update on prevention, diagnosis and treatment. Curr Opin Ophthalmol 23(4):288-95.

[41] Batt SL, Charalambous BM, Solomon AW et al. (2003). Impact of azithromycin administration for trachoma control on the carriage of antibiotic-resistant *Streptococcus pneumoniae*. Antimicrob Agents Chemother 47:2765-69.

[42] Maher MC, Alemayehu W, Lakew T et al. (2012). The fitness cost of antibiotic resistance in *Streptococcus pneumoniae*: insight from the field. PLoS ONE 7(9):e29407.

[43] West SK, Munoz B, Lynch M et al. (1995). Impact of face-washing on trachoma in Kongwa, Tanzania. Lancet 345:155-8.

[44] Lietman TM, Dawson CR, Osaki SY, Zegans ME (2000). Clinically active trachoma versus actual chlamydial infection. Med J Aust 172:93-4.

[45] Miller K, Schmidt G, Melese M et al. (2004). How reliable is the clinical exam in detecting ocular chlamydial infection? Ophthalmic Epidemiol 11:255-62.

[46] Schachter J, Moncada J, Dawson CR et al. (1988). Non-culture methods for diagnosing chlamydial infection in patients with trachoma: a clue to the pathogenesis of the disease? J Infect Dis 158:1347-1352.

[47] Harding-Esch EM, Edwards T, Sillah A et al. (2009). Active trachoma and ocular *Chlamydia trachomatis* infection in two Gambian regions: On course for elimination by 2020? PLoS Negl Trop Dis 3(12):e573.

[48] Beatty WL, Byrne GI, Morrison RP (1994). Repeated and persistent infection with Chlamydia and the development of chronic inflammation and disease. Trends Microbiol 2(3):94-8.

[49] Tang FF, Chang HL, Huang YT, Wang KC (1957). Studies on the etiology of trachoma with special reference to isolation of the virus in chick embryo. Chin Med J 75:429-447.

[50] Wang SP, Grayston JT (1970). Immunologic relationship between genital TRIC, lymphgranuloma venereum and related organisms in a new microtiter indirect immunofluorescence test. Am J Ophthalmol 70:367-74.

[51] Stephens RS, Sanchez-Pescador R, Wager E et al. (1987). Diversity of the major outer membrane proteins of *Chlamydia trachomatis*. J Bacteriol 169:3879-3885.

[52] Baehr W, Zhang YX, Joseph T et al. (1988). Mapping antigenic domains expressed by *Chlamydia trachomatis* major outer membrane protein genes. Proc Natl Acad Sci USA 85:4000-4.

[53] Millman K, Tavare S, Dean D (2001). Recombination in the MOMP gene but not the *omcB* gene of chlamydiae contributes to serovar-specific differences in tissue. J Bacteriol 183:5997-6008.

[54] Byrne GI (2010). *Chlamydia trachomatis* strains and virulence: Rethinking links to infection prevalence and disease severity. J Infect Dis 201(Suppl 2):s126-133.

[55] Hatch TP (1996). Disulfide cross-linked envelope proteins: The functional equivalent of peptidoglycan in chlamydiae? J Bacteriol 178:1-5.

[56] Bavoil P, Ohlin A, Schacter J (1984). Role of disulfide bonding in outer membrane structure and permeability in *Chlamydia trachomatis*. Infect Immun 44:479-485.

[57] Wyllie S, Ashley RH, Longbottom D, Herring AJ (1998). The major outer membrane protein of *Chlamydia psittaci* functions as a porin-like ion channel. Infect Immun 66:5202-5207.

[58] Jones HM, Kubo A, Stephens RS (2000). Design, expression and functional characterization of a synthetic gene encoding the *Chlamydia trachomatis* major outer membrane protein. Gene 258:173-181.

[59] Su H, Watkins N, Zhang YX, Caldwell HD (1990). *Chlamydia trachomatis* host cell interaction: Role of the chlamydial major outer membrane protein as an adhesin. Infect Immun 58:1017-1025.

[60] Swanson AF, Kuo CC (1994). Binding of the glycan of the major outer membrane protein of *Chlamydia trachomatis* to HeLa cells. Infect Immun 62:24-28.

[61] Rodriguez-Maranon MJ, Bush RM, Peterson LM et al. (2002). Prediction of the membrane-spanning Beta strands of the major outer membrane protein of Chlamydia. Protein Sci 11:1854-1861.

[62] Wang Y, Berg EA, Feng X et al. (2006). Identification of surface-exposed components of the major outer membrane protein of *Chlamydia trachomatis* serovar F. Protein Sci 15(1):122-134.

[63] Caldwell HD, Kromhout J, Schacter J (1981). Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis.* Infect Immun 31:1161-1176.

[64] Zhang YX, Steward SJ, Joseph T et al. (1987). Protective monoclonal antibodies recognize epitopes located on the major outer membrane protein of *Chlamydia trachomatis*. J Immunol 183:575-581.

[65] Confer AW, Ayalew S (2013). The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. Vet Microbiol 163(3-4):207-222.

[66] Stothard DR, Boguslawski G, Jones RB (1998). Phylogenetic analysis of the *Chlamydia trachomatis* major outer membrane protein and examination of potential pathogenic determinants. Infect Immun 66:3618-3625.

[67] Gomes JP, Bruno WJ, Nunes A et al. (2007). Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res 17:50-60.

[68] Pedersen LN, Herrmann B, Moller JK (2009). Typing *Chlamydia trachomatis*: from egg yolk to nanotechnology. FEMS Immunol Med Microbiol 55:120-30.

[69] Andreason AA, Burton MJ, Holland MJ et al. (2008*). Chlamydia trachomatis ompA* variants in trachoma: What do they tell us? PLoS Negl Trop Dis 2(9):e306.

[70] Pedersen LN, Podenphant L, Moller K (2008). Highly discriminative genotyping of *Chlamydia trachomatis* using *omp1* and a set of variable number tandem repeats. Clin Microbiol Infect 14:644-652.

[71] Van Belkum A, Tassios PT, Dijkshoorn L et al. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 13(Suppl 3):1-46.

[72] Christerson L, Herrmann B (2012). Guidelines for high resolution genotyping of *Chlamydia trachomatis* using multilocus sequence analysis. Methods Mol Biol 903:51-64.

[73] Pannekoek Y, Morelli G, Kusecek B et al. (2008). Multi locus sequence typing of Chlamydiales: Clonal groupings within the obligate intracellular bacterium *Chlamydia trachomatis*. BMC Microbiol 8:42.

[74] Klint M, Fuxelius HH, Goldkuhl RR et al. (2007). High-resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence analysis. J Clin Microbiol 45:1410-14.

[75] Ikryannikova LN, Shkarupeta MM, Shitikov EA et al. (2010). Comparative evaluation of new typing schemes for urogenital *Chlamydia trachomatis* isolates. FEMS Immunol Med Microbiol 59:188-196.

[76] Bom RJM, Christerson L, Schim van der Loeff MF et al. (2011). Evaluation of High-Resolution Typing Methods for *Chlamydia trachomatis* in Samples from Heterosexual Couples. J Clin Microbiol 49(8):2844-53.

[77] Stephens RS, Kalman S, Lammel C et al. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-9.

[78] Harris SR, Clarke IN, Seth-Smith HMB et al. (2012). Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nature Genet 44(4):413-20.

[79] Jeffrey BM, Suchland RJ, Quinn KL et al. (2010). Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun 78(6):2544-53.

[80] Millman K, Black CM, Johnson RE (2004). Population-based genetic and evolutionary analysis of *Chlamydia trachomatis* urogenital strain variation in the United States. J Bacteriol 186(8):2457-65.

[81] Lysén M, Osterlund A, Rubin CJ (2004). Characterisation of *ompA* genotypes by sequence analysis from all detected cases of *Chlamydia trachomatis* infections during 1 year of contact tracing in a Swedish county. J Clin Microbiol 42(4):1641-7.

[82] Kari L, Whitmire WM, Carlson JH et al. (2008). Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variants. J Infect Dis 197:449-56.

[83] Binet R, Maurelli AT (2009). Transformation and isolation of allelic exchange mutants of Chlamydia using DNA introduced by electroporation. Proc Natl Acad Sci USA 106:292-7.

[84] Wang Y, Kahane S, Cutcliffe LT et al. (2011). Development of a transformation system for *Chlamydia trachomatis*: Restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. *PLoS Pathog* 7(9):e1002258.

Chapter 2 A review of Chlamydia trachomatis virulence loci and their association with disease severity

2.1 Introduction

There is evidence of strain-dependent virulence associations in both urogenital and ocular *C. trachomatis* infection [1]. Some strain-dependent patterns have been demonstrated and studies are required to establish associations, and causality, between *C. trachomatis* strains and disease severity. Characterising *C. trachomatis* remains a challenge. Using current typing systems, *C. trachomatis* strains remain poorly defined with respect to their intracellular cell biology, virulence and ecological success. Now that it is possible to genetically manipulate chlamydia [2,3], perform mutagenesis studies [3,4,5] and obtain whole genome sequence data direct from clinical samples [6, Christiansen MT *et al.,* in press], it may be possible through laboratory and clinico-epidemiological studies to better define chlamydial strains and their association with disease severity.

2.2 The Major Outer Membrane Protein (MOMP)

The historical system of characterizing *C. trachomatis* by serological differences elicited by variable segments of MOMP and subsequent genotypic sequence differences for the gene encoding MOMP (*ompA*) is discussed in ***Chapter 1.6.3***. Several authors have suggested that immune pressure drives MOMP variation [7,8], and that MOMP variability reflects strain-related virulence [9]. Whole genome sequence analysis has shown *ompA* as a hotspot for recombination, and questioned its suitability as a marker for virulence [10]. Whilst some studies have described *ompA* variation in relation to ocular *C. trachomatis* load [11] the apparent inability of some hosts to clear infection caused by homotypic *ompA* genovars [12] and the suggestion that *ompA* is under purifying selection [11,13], it has not been possible to attribute pathobiological differences to *ompA* genovars [14,15].

2.3 Polymorphic outer membrane proteins (Pmp)

Chlamydial *pmp* genes were originally described by Longbottom *et al.* [16,17] and subsequently by Stephens *et al*. [18] based on whole genome sequence data related to large surface proteins in *Chlamydia psittaci* [19]. The nine-member family of *pmp* genes (*pmpA-pmpI*) in *C. trachomatis* all appear to be transcribed and encode proteins. The *pmp* gene family accounts for approximately 14% of the coding capacity for *C. trachomatis* [20,21].

Pmp proteins appear to fulfill several biological functions, but the full extent of their role is not entirely clear. Pmp A and D are highly conserved amongst *C. trachomatis* isolates, but others (Pmp E, F, H, I) are variable. Pmp D and pmpH are surface expressed [21,22]. These exported bacterial proteins have been shown to function as autotransporters within the Chlamydial outer membrane [21-23].

Pmps are involved in transmembrane transport as part of the chlamydial type five (autotransporter) secretion system, delivering effector proteins directly into specific host cell membranes [22, 24-26]. The N and C terminal domains of specific families share conservation between species, but the passenger domain provides unique function for these proteins, and they may be cleaved, uncleaved or surface expressed and therefore involved with the bacterial surface [27]. Several chlamydial Pmps (Pmp D, E, G, H) have been detected on the chlamydial elementary body (EB) surface [21]. Pmps B, D and H are strongly immunogenic and elicit a pro-inflammatory response [21]. Thus they have a putative role in virulence regarding modulation of inflammation and adherence to and invasion of host cells [23]. Pmp strain types have been shown to associate *C. trachomatis* isolates with pathobiotype and tissue tropism [28,29]. Currently data are limited to a small number of laboratory strains, but they are good candidates for further study of the virulence and pathogenesis of Chlamydial infection.

2.4 Chlamydial plasticity zone: Chlamydial cytotoxin and phospholipase D family

The chlamydial plasticity zone (PZ) is the site of extensive genomic variation between the chlamydial serovars [30,31]. The PZ contains a chlamydial toxin whose function is thought to involve GTPase inactivation and guanylate binding protein neutralization resulting in interferon gamma (IFN-g) resistance [32-24]. IFN-g stimulation of cells *in vitro* and *in vivo* induces transcription of genes that co-operate to eliminate intracellular bacteria, suggesting that inhibition of this mechanism may contribute to the pathogenesis of chlamydial disease [34]. Cytotoxin genomic variability has been associated with phenotypic variability relating to localization of GTPases to the inclusion membrane [34]. This genomic variability of the cytotoxin correlates with disease phenotypes such that genital strains have a single gene with a large deletion, LGV strains lack the gene and closely related *Chlamydia muridarum* has three copies [35,36]. This may explain the invasive nature of LGV strains versus the apparent restriction of urogenital and ocular strains to mucosal surfaces. Furthermore, cytotoxin gene polymorphisms have distinguished between ocular and

genital *C. trachomatis* isolates, where there may be either complete or partial cytotoxin genes present [32,37]. Further studies looking at associations within clinical strains have not yet been conducted, nor has the full detail of the structure and function of the protein been confirmed. Evidence to date suggests that there is unlikely to be differing toxic potential between strains, since there is little selective advantage for cytotoxin-bearing strains in humans.

The membrane attack complex/perforin (MACPF) protein (CT153) and members of the phospholipase-D (PLD) family are also encoded within the PZ [31]. The PZ PLDs are present in *C. trachomatis* [38] and other Chlamydiaceae [39-41]. There are limited data on the function of these proteins. They are known to modify lipids and membranes, may be secreted and localize to the inclusion membrane [42]. Reactivation of *C. trachomatis* from a gamma interferon-induced persistent state is blocked by primary alcohols, which are potent inhibitors of PLDs, implicating a role for the plasticity zone PLDs in this process [43].

2.5 Translocated Actin Recruiting Phosphoprotein (TARP)

TARP has been noted to have considerable sequence variation, a considerable proportion of which is in the form of a variable number of insertions/deletions (indels) between isolates [36,37,43]. *C. trachomatis* entry into a host cell results in reconfiguration of the host cell actin skeleton, mediated by TARP. It is suggested that TARP is involved in pathogen-directed phagocytosis (during the initial attachment phase of uptake of the elementary body) [44]. After phosphorylation TARP is thought to promote internalization via an actin recruiting mechanism [43-45]. In addition to the characteristic indels, there are several single nucleotide polymorphisms (SNPs) between *C. trachomatis* serovars, suggesting that intra-serovar recombination has a role in the diversity of these genes [10,46]. Given the variability between pathobiotypes it has been suggested that these genes may be associated with disease severity.

2.6 Inclusion membrane proteins (Incs)

Inclusion membrane proteins were originally identified prior to the advent of whole genome sequencing for chlamydia. High-resolution sequencing has provided significant information regarding *inc* genes and their proteins. Genomic analysis has shown that many Inc proteins are produced by each chlamydial species and several are produced early in development. Inc proteins may be important antigens in cell-mediated immune responses following chlamydial infection. IncA has been defined as

an effector protein, proposed to mediate inclusion fusion in *C. trachomatis* [47,48]. They also exhibit an innate ability to form vesicles, enter the membrane of intracellular compartments and are type III effector proteins [49]. The epidemiology of natural *incA* mutants has been studied suggesting that strains with an inability to express *incA* are associated with reduced virulence [50]. Since then further work has been conducted which is less conclusive [51]. Many more *incs* have now been identified. Inc proteins appear to be important in chlamydial biology, but their roles remain unclear [51,52]. Further investigation into their role in host-pathogen interaction is warranted.

2.7 Chlamydia Type III Secretion System (T3SS)

The T3SS provides Gram-negative bacteria with a specialized mechanism for arranging proteins for subsequent interaction with the host cell membrane, insertion of effector proteins into the cytosol and subsequent manipulation or modification of host cell behaviour. Early work on chlamydial T3SS [53] in combination with the advent of whole genome sequencing of chlamydia [18] demonstrated the T3SS secretion mechanism as a candidate system for a family of chlamydial proteins that have been shown to localize to the inclusion membrane during intracellular chlamydial growth [54]. This system is important in bacterial pathogenesis and is essential for full virulence of a number of clinically significant bacterial species including Yersinia spp., Salmonella spp., *Shigella flexneri*, *Pseudomonas aeruginosa*, *Escherichia coli*, Bordatella spp., Burkholderia spp. and Vibrio spp. [55]. In most of these pathogens the structural and effector genes are located on distinct pathogenicity islands [56]. In chlamydia, these genes are scattered across a number of operons and use the sigma factor for constitutive gene expression, suggesting that additional activators or repressors are required if components of the T3SS are developmentally regulated [57]. Putative structural components, effectors and chaperones have been identified [58-60]. Inc proteins, Pmps and TARP have been suggested as putative effector proteins. Given the intracellular existence of chlamydia, it seems that type three secreted products are likely to be important in chlamydial pathogenesis. Therefore some of these effectors and apparatus components may be suitable candidates for therapeutics or vaccines.

2.8 Chlamydial stress response proteins

Immunological studies have implicated host responses to a subset of chlamydial stress response proteins (chlamydial heat shock proteins (Hsp60 and

Hsp10) and GroEL/GroES homologues) with disease severity [61-65]. Heat shock proteins are specialized molecules performing different and essential roles in the cell including protein degradation, folding and trafficking. The chlamydial Hsp60 antigen shares nearly 50% sequence identity with the human homologue, therefore molecular mimicry has been suggested as being responsible for the severe inflammatory chlamydial disease found in individuals reacting against Hsp60 [61,65]. A study looking at women with urogenital *C. trachomatis* infection showed that in recurrent chlamydial infections, higher titres of antibodies to Hsp10 and Hsp60 were found [66].

Chlamydial GroEL is reported to signal through the toll like receptor (TLR) system, suggesting that it is also a potential mediator of inflammation during chlamydial disease. The *groE* operon (*groES/groEL*) remains highly conserved within and between chlamydial species [18] therefore association with disease phenotype seems unlikely due to lack of variability. It may be that the *groE* operon is important in pathogenesis and is implicated as a mediator of disease, but may not be differential in strain-dependent virulence. Gene expression and functional studies may be important to elucidate the function of these genes and whether or not they may be suitable potential vaccine candidates.

2.9 Chlamydial lipopolysaccharide and other glycolipids

Lipopolysaccharide (LPS) and lipooligosaccharides (LOS) are the primary lipid components of the outer cell membrane and are vital for cell viability in most Gram-negative bacteria. Studies blocking chlamydial LOS have shown that chlamydia remained viable and developed an inclusion but were not infectious [67]. In chlamydia cultured under these conditions, reticulate bodies accumulated, but failed to express essential proteins facilitating their transition to elementary bodies. This suggests an important interaction between LOS and an outer membrane developmental quality control system that facilitates transition from reticulate to elementary bodies [67]. Inhibiting LPS has been shown to reduce chlamydial infectivity suggesting a role for LPS in infectivity, although underlying mechanisms and details of host cell interaction are unknown [68]. Chlamydial LPS may act through TLRs and structural features may be associated with disease severity patterns, although it is not thought to be a LPS antagonist [69]. The functional role of genus-specific LPS is complex and not fully elucidated [70].

2.10 Chlamydial cryptic plasmid

The chlamydial plasmid is a small, highly conserved, non-conjugative, non-integrative DNA molecule that is virtually ubiquitous in *C. trachomatis*. There has been significant recent interest and progress in understanding the function of the plasmid in the context of disease pathogenesis. The 7.5-kb *C. trachomatis* plasmid has been shown to function as a virulence factor in animal models [71,72]. Phenotypic differences exist between plasmid-cured and wild-type *C. trachomatis* strains with respect to infectivity, glycogen accumulation, induction of inflammation and activation of TLR pathways [73,74]. Plasmid deletion mutagenesis studies showed that deletion of the plasmid-encoded *pgp4* gene results in an *in vitro* phenotype identical to that of a plasmid-free strain [5]. This supports bacterial transcriptome analysis showing a decrease in transcript levels of a subset of chromosomal genes in a naturally occurring plasmid-free strain of *C. trachomatis*, demonstrating that the plasmid is a transcriptional regulator of virulence-associated genes [75]. Maintenance of the plasmid at low copy number carries inherent risk during cell partition [76] but naturally occurring plasmid-free strains are rare [77-79]. A lower risk higher copy number system is metabolically expensive but may confer a 'fitness' advantage. Thus, the maintenance of 5-6 plasmids per genome [80] may maximise infectivity or intracellular survival whilst provoking minimal host immune response. The plasmid is clearly strongly selected for during successful infection or transmission of the pathogen within a host population [74].

Though there is convincing evidence that the chlamydial plasmid is a virulence factor [73-75,81-83], recent population-based data from a trachoma-endemic population suggest that plasmid copy number is not associated with disease severity and that additive gene dosage effects do not appear to correlate with pathogen virulence *in vivo* [80]. This supports *in vitro* work showing no association between plasmid copy number and tissue tropism [84]. Previous work *in vitro* and in animal models suggests that subtle genomic differences between chlamydial isolates are associated with differences in growth kinetics, immune responses and pathology [85,86].

The plasmid has also been investigated as a vaccine candidate. Recent studies have shown that plasmid deficient *C. trachomatis* strains are less virulent in ocular infection of macaques and functions as a live attenuated vaccine in that system [87]. Further epidemiological and *in vitro* studies using comparative pathogen genomics to

examine these associations are required to fully understand the role of the plasmid in human chlamydial disease.

2.11 Metabolic processes and chlamydial virulence

2.11.1 Tryptophan metabolism

Tryptophan metabolism has been shown to differentiate between ocular and urogenital strains of chlamydia, suggesting tissue-specific virulence association. Urogenital strains of *C. trachomatis* have evolved to survive under tryptophan-limited conditions through activation of tryptophan-synthase genes and acquisition of indole as a substrate from commensal bacterial flora [88]. This is thought to be achieved through functional partial tryptophan operons in urogenital strains [89]. Ocular strains present an enigma since they should be exquisitely sensitive to IFN-g-mediated tryptophan depletion as they do not appear to possess a functional tryptophan operon.

2.11.2 Iron and micronutrients

Hypoferraemia and transferrin receptor modulation is a known consequence of the inflammatory response [90]. Chlamydia respond to iron limitation in a regulated way, typical of many bacteria [91,92]. Iron in a form usable to the pathogen is often a limiting factor in any environment, thus to be able to effectively sequester iron may be important to chlamydial virulence. There is evidence that defining metabolic activity can give insight into biological processes relating to chlamydial infectivity [93].

Further studies on nutritional stress and other metabolic processes will emerge from a combination of transcriptomal and expression-based analyses in appropriate host-pathogen interaction models. These findings suggest that there may be rationale in suggesting that other metabolic pathways may be important in defining disease severity phenotypes amongst chlamydial strains. In combination with *in vitro* experiments, pathogen whole genome sequence analysis and corresponding clinical and epidemiological metadata, a more complete picture of metabolic pathways and their role may be elucidated.

2.12 Conclusion

Studying pathogen virulence factors at population level should enable the assignment of the relative importance of putative virulence associated genes and loci. This requires knowledge of host genetics, the microbial environment at the time of infection, host demographic data and longitudinal study as these factors will all impact on disease severity and the course of disease. To fully understand these

interactions, detailed host-pathogen data are required. Comparative bioinformatics analyses are powerful tools that can assist in investigating such complex data sets. Identification of pathogen virulence factors is crucial to understanding pathogenesis and transmission of infection, and will impact on future public health interventions such as surveillance following treatment with antibiotics and the development of a chlamydial vaccine.

2.13 Chapter 2 References

[1] Abdelsamed H, Peters J, Byrne GI (2013). Genetic variation in *Chlamydia trachomatis* and their hosts: impact on disease severity and tissue tropism. Future Microbiol 8(9):1129-46.

[2] Wang Y, Kahane S, Cutcliffe LT et al. (2011). Development of a transformation system for *Chlamydia trachomatis*: Restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. PLoS Pathog 7(9):e1002258.

[3] Wang Y, Kahane S, Cutcliffe LT et al. (2013). Genetic transformation of a clinical (genital tract) plasmid-free isolate of *Chlamydia trachomatis*: Engineering the plasmid as a cloning vector. PLoS ONE 8(3):e59195.

[4] Wang Y, Cutcliffe LT, Skilton RJ et al. (2013). Transformation of a plasmid-free genital tract isolate with a plasmid vector carrying a deletion in CDS6 revealed that this gene regulates inclusion phenotype. Pathogens Dis 67(2):100-103.

[5] Song L, Carlson JH, Whitmire VM et al. (2013). *Chlamydia trachomatis* plasmid-encoded *pgp4* is a transcriptional regulator of virulence associated genes. Infect Immun 81(3):636.

[6] Seth-Smith HMB, Harris S, Skilton RJ et al. (2013). Whole genome sequences of *Chlamydia trachomatis* direct from clinical samples without culture. Genome Res 23(5):855-866.

[7] Stothard DR, Boguslawski G, Jones RB (1998) Phylogenetic analysis of the *Chlamydia trachomatis* major outer membrane protein and eamination of potential pathogenic determinants. Infect Immun 66:3618-3625.

[8] Brunham RC, Plummer FA, Stephens RS (1993). Bacterial antigenic variation, host immune response and pathogen-host co-evolution. Infect Immun 61:2273-2276.

[9] Millman KL, Tavare S, Dean D (2001). Recombination in the MOMP gene but not the *omcB* gene of Chlamydiae contributes to serovar-specific differences in tissue. J Bacteriol 183:5997-6008.

[10] Harris SR, Clarke IN, Seth-Smith HMB et al. (2012). Whole genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic

relationships masked by current clinical typing. Nature Genet 44(4):413-s1.

[11] Andreasen AA, Burton MJ, Holland MJ et al. (2008). *Chlamydia trachomatis ompA* variants in trachoma: What do they tell us? PLoS Negl Trop Dis 2(9):e306.

[12] Hsieh YH, Bobo LD, Quinn TC, West SK (2001). Determinants of trachoma endemicity using Chlamydia trachomatis ompA DNA sequencing. Microbes Infect 3(6):447-58.

[13] Millman K, Black CM, Johnson RE et al. (2004). Population-based genetic and evolutionary analysis of Chlamydia trachomatis urogenital strain variation in the United States. J Bacteriol 186(8):2457-65.

[14] Jalal H, Verlande NQ, Kumar N et al. (2011). Genital chlamydial infection: Association between clinical features, organism genotype and load. J Med Microbiol 60(7):881-8.

[15] Fitch WM, Peterson EM, de la Maza LM (1993). Phylogenetic analysis of the outer membrane protein genes of Chlamydia and its implication for vaccine development. Mol Biol Evol 10:892-913.

[16] Longbottom D, Russell M, Jones GE et al. (1996). Identification of a multigene family coding for the 90kDa proteins of the ovine abortion subtype of *Chlamydia psittaci*. FEMS Microbiol Lett. 142(2-3):277-81.

[17] Longbottom D, Findlay J, Vreton E, Dunbar SM (1998). Immunoelecton microscopic localization of the OMP90 family on the outer membrane surface of *Chlamydia psittaci*. FEMS Microbiol Lett. 164(1):111-7.

[18] Stephens RS, Kalman S, Lammel C et al. (1998). Genomic sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-9.

[19] Longbottom D, Russell M, Dunbar SM et al. (1998). Molecular cloning and characterization of genes coding for the highly immunogenic cluster of 90 kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. Infect Immun 66:1317-1324.

[20] Rockey DD, Lenart J, Stephens RS (2000). Genome sequencing and our understanding of chlamydiae. Infect Immun 68:5473-5479.

[21] Tan C, Spitznagel JK, Shoe H-Z et al. (2006). The polymorphic membrane protein gene family of the chlamydiaciae. Chlamydia Genomics and

Pathogenesis. Ed: Bavoil PM, Wyrick PB. Horizon Bioscience; Norfolk, UK. pp 195-218.

[22] Swanson KA, Taylor LD, Frank SD et al. (2009). *Chlamydia trachomatis* polymorphic membrane protein D is an oligomeric autotransporter with a higher-order structure. Infect Immun 77:508-516.

[23] Werhl W, Brinkmann V, Jungblut PR et al. (2004). From the inside out – Processing of the Chlamydial autotransporter PmP D and its role in adhesion and activation of host cells. Mol Microbiol 51(2):319-334.

[24] Henderson IR, Lam AC (2001). Polymorphic proteins of Chlamydia spp. – Autotransporters beyond the Proteobacteria. Trends Microbiol 9(12):573-8.

[25] Vandahl BB, Pedersen AS, Gevaert K et al. (2002). The expression, processing and localization of polymorphic membrane proteins in *Chlamydia pneumoniae* strain CWL029. BMC Microbiol 2:36.

[26] Wheelhouse N, Sait M, Wilson K et al. (2012). Expression patterns of five polymorphic membrane proteins during the *Chlamydia abortus* developmental cycle. Vet Microbiol 160(3-4):525-529.

[27] Henderson IR, Navarro-Garcia F, Nataro JP (1998). The great escape: Structure and function of autotransporter proteins. Trends Microbiol 6:370-378.

[28] Stothard DR, Toth GA, Batteiger BE (2003). Polymorphic membrane protein H has evolved in parallel with the three disease-causing groups of Chlamydia trachomatis. Infect Immun 71:1200-1208.

[29] Gomes JP, Nunes A, Bruno WJ et al. (2006). 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 188:275-286.

[30] Read TD, Brunham RC, Shen C et al. (2000). Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28(6):1397-406.

[31] Read TD, Myers GS, Brunham RC et al. (2003). Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci GPIC*): examining the role of niche-specific genes in the evolution of the chlamydiaceae. Nucleic Acids Res 31(8):2134-47.

[32] Belland RJ, Scidmore MA, Crane DD et al. (2001). *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci 98(24):13984-9.

[33] Carlson JH, Hughes S, Hogan D (2004). Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. Infect Immun 72(12):7063-72.

[34] Tietzel I, El Haibi C, Carabeo RA (2009). Human granulocyte binding proteins potentiate the anti-chlamydial effects of Interferon gamma. PLoS ONE 4(8):e6499.

[35] Fan B, Van Der Pol B, Nelson DE (2008). Do chlamydial cytotoxins mediate IFN-g immune evasion. Proc Eur Soc Chlamydia Res 6:165-71.

[36] Thomson NR, Holden TF, Carder C et al. (2008). *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res 18(1):161-71.

[37] Carlson JH, Porcella SF, McClarty G (2005). Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and geniotropic strains. Infect Immun 73(10):6407-18.

[38] Nelson DE, Crane DD, Taylor LD et al. (2006). Inhibition of chlamydiae by primary alcohols correlates with strain specific complement of plasticity zone phospholipase D genes. Infect Immun 74(1):73-80.

[39] Sait M, Livingstone M, Clark EM et al. (2014). Genome sequencing and comparative analysis of three *Chlamydia pecorum* strains associated with different pathogenic outcomes. BMC Genomics 15:23.

[40] Ciervo A, Mancini F, Cassone A (2007). Transcription, expression, localization and immunoreactivity of *Chlamydophila pneumoniae* phospholipase D proteins. Microb Pathog 43(2-3):96-105.

[41] Kalman S, Mitchell W, Merathe R et al. (1999). Comparative genomes of *Chlamydia pneumoniae* and Chlamydia trachomatis. Nat Genet 21(4):385-9.

[42] Taylor LD, Nelson DE, Dorward DW et al. (2010). Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. 78(6):2691-99.

[43] Clifton DR, Fields KA, Grieshaber SS et al. (2004). A chlamydial type III translocated protein is tyrosin-phosphorylated at the site entry and

associated with recruitment of actin. Proc Natl Acad Sci USA 101:10166-71.

[44] Jewett TJ, Fischer ER, Mead DJ (2006). Chlamydial TARP is a bacterial nucleator of actic. Pro Natl Acad Sci USA 103(42):15599-604.

[45] Lane BJ, Mutchler C, Al Khodor S et al. (2008). Chlamydial entry involves TARP binding of guanine nucleotide exchange factors. PLoS Pathog 4:e1000014, 1-11.

[46] Gomes JP, Bruno WJ, Nunes A. (2007). Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res 17(1):50-60.

[47] Hackstadt T, Scidmore-Carlson MA, Shaw EI, Fischer ER (1999). *Chlamydia trachomatis* IncA protein is required for homotypic vesicle fusion. Cell Microbiol 1:119-130.

[48] Suchland RJ, ROckey DD, Bannatine JP, Stamm WE (2000). Isolates of *Chlamydia trachomatis* that occupy nonfusogenic inclusions lack IncA, a protein localized to the inclusion membrane. Infect Immun 68:360-367.

[49] Giesler WM, Suchland RH, Rockey DD, Stamm WE (2001). Epidemiology and clinical manifestations of unique *Chlamydia trachomatis* isolates that occupy nonfusogenic inclusions. J Infect Dis 184:879-884.

[50] Mital J, Millar NJ, Dorward DW et al. (2013). Roles for chlamydial inclusion membrane proteins in inclusion membrane structure and biogenesis. PLoS ONE 8(5):e63426.

[51] Pannekoek Y, van der Ende A, Eijk PP et al. (2001). Normal IncA expression and fusogenicity of *Chlamydia trachomatis* isolates with the *incA I47T* mutation. Infect Immun 69(7):4654-6.

[52] Scidmore-Carlson MA, Shaw EI, Dooley CA et al. (1999). Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. Mol Microbiol 33:753-765.

[53] Hsia R-C, Pannekoek Y, Ingerowski E et al. (1997). Type III secretion genes identify a putative virulence locus of chalmydia. Mol Microbiol 25:351-359.

[54] Rockey DD, Heinzen RA, Hackstadt T (1995). Cloning and characterization of a *Chlamydia psittaci* gene coding for a protein localized in the inclusion membrane of infected cells. Mol Microbiol 15:617-26.

[55] Hueck CJ (1998). Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. Microbiol Mol Biol Rev 62(2):379-433.

[56] Mescas J, Strauss EJ (1996). Molecular mechanisms of bacterial virulence: Type III secretion and pathogenicity islands. Emerg Infect Dis 2:271-88.

[57] Hefty SP, Stephens RS (2007). *Chlamydia trachomatis* is encoded on ten operons preceded by a sigma 70-like promoter element. J Bacteriol 189:198-206.

[58] Fields KA, Hackstadt T (2000). Evidence for the selection of *Chlamydia trachomatis* copN by a type three secretion mechanism. Mol Micirobiol 38:1048-1060.

[59] Betts HJ, Twiggs LE, Sal MS et al. (2008). Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of *Chlamydia trachomatis*. J Bacteriol 190:1680-1690.

[60] Fields KA, Fisher ER, Mead DJ, Hackstadt T (2005). Analysis of putative *Chlamydia trachomatis* chaperones Scc2 and Scc3 and their use in the identification of type III secretion substrates. J Bacteriol 187:6466-6478.

[61] Peeling RW, Mabey DC (1999). Heat shock protein expression and immunity in chlamydial infections. Infect Dis Obstet Gynaecol 7(1-2):7-9.

[62] Witkin SS, Jeremias J, Toth M, Ledger WJ (1993). Cell-mediated immune response to the recombinant 57 kDa heat shock protein of *Chlamydia trachomatis* in women with salpingitis. J Infect Dis 167:1379-83.

[63] Toye B, Laferriere C, Claman P et al. (1993). Association between antibody to the chlamydial heat shock protein and tubal infertility. J Infect Dis 168:1236-1240.

[64] LaVerda D, Albanese LN, Ruther PE et al. (2000). Seroactivity to *Chalmydia trahcomatis* HSP10 correlates with disease severity n women. Infect Immun 68:303-9.

[65] Kinnuen A, Molander P, Morrison R et al. (2002). Chlamydial heat shock protein 60-specific T cells in inflamed salpingeal tissue. Fertil Steril 77:162-172.

[66] Nguyen BD, Cunningham D, Liang X et al. (2011). Lipooligosaccharide is required for the generation of infectious elementary bodies in *Chlamydia trachomatis*. Proc Natl Acad Sci 108(25):10284-9.

[67] Agrawal T, Vats V, Salhan S, Mittal A (2007). Mucosal and peripheral immune responses to chlamydial heat shock proteins in women infected with Chlamydia trachomatis. Clin Exp Immunol 148(3):461-8.

[68] Fadel S, Eley A. (2008). Is lipopolysaccharide a factor in infectivity of Chlamydia trachomatis? J Med Microbiol 57(3):261-6.

[69] Erridge C, Pridmore A, Eley A et al. (2004). Lipopolysaccharides of *Bacteroides fragilis, Chlamydia trachomatis* and *Pseudomonas aeruginosa* signal via Toll-like receptor 2. J Med Microbiol 53:735-740.

[70] Muller-Leonnies S, Gronows S, Brade L et al. (2006). A monoclonal antibody against a carbohydrate epitope in lipopolysaccharide differentiates Chlamydophila psittaci from Chlamydophila pecorum, Chlamydia pneumoniae and Chlamydia trachomatis. Glycobiology 16(3):184-96.

[71] Kari L, Whitmire WM, Olivares-Zavaleta N, Goheen MM, et al. (2011). A live-attenuated chlamydial vaccine prevents against trachoma in non-human primates. J Exp Med 208(11):2217-23.

[72] Frazer LC, Darville T, Chandra-Kuntal K, et al. (2012). Plasmid-cured *Chlamydia caviae* activates TLR2-dependent signaling and retains virulence in the guinea pig model of genital tract infection. PLoS One **7**(1):e30747.

[73] O'Connell CM, AbdelRahman YM, Green E et al. (2011). Toll-like receptor 2 activation by *Chlamydia trachomatis* is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by *C. trachomatis* but not by *C. muridarum*. Infect Immun 79:1044-1056.

[74] Russell M, Darville T, Chandra-Kuntal K et al. (2011). Infectivity acts as *in vivo* selection for maintenance of the chlamydial cryptic plasmid. Infect Immun **79**:98–107.

[75] Carlson JH, Whitmire WM, Crane DD et al. (2008). The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect Immun 76:2273-83.

[76] Twigg AJ, Sheratt D (1980). Trans-complementable copy-number mutants of plasmid ColE1. Nature 283(5743):216-8.

[77] Thomas NS, Lusher M, Storey CC, Clarke IN (1997). Plasmid diversity in chlamydia. Microbiology 143(6):1847-54.

[78] An Q, Radcliffe G, Vassallo R et al. (1992). Infection with a plasmid-free variant Chlamydia related to *Chlamydia trachomatis* identified by using multiple assays for nucleic acid detection. J Clin Microbiol 30:2814-21.

[79] Peterson EM, Markoff BA, Schacter J, de la Maza LM (1990). The 7.5kb plasmid present in *Chlamydia trachomatis* is essential for the growth of the micro-organism. Plasmid 23:144-8.

[80] Last AR, Roberts Ch, Cassama E et al. (2014). Plasmid copy number and disease severity in naturally occurring ocular *Chlamydia trachomatis* infection. J Clin Micro 52(1):324-7.

[81] Nunes A, Borrego MJ, Gomes JP (2013). Genomic features beyond *C. trachomatis* phenotypes: what do we think we know? Infect Genet Evol 16:392-400.

[82] O'Connell CM, Ingalls RR, Andrews CW Jr et al. (2007). Plasmid-deficient *Chlamydia muridarum* fail to induce immune pathology and protect against oviduct disease. J Immunol 179:4027-34.

[83] Comanducci M, Manetti L, Bini A et al. (1994). Humoral immune response to plasmid protein *pgp-3* in patients with *Chlamydia trachomatis* infection. Infect Immun 62:5491-7.

[84] Ferreira R, Borges V, Nunes A et al. (2013). Assessment of the load and transcriptional dynamics of *Chlamydia trachomatis* plasmid according to strains' tissue tropism. Microbiol Res 168(6):333-9.

[85] Kari L, Whitmire WM, Carlson JH et al. (2008). Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis 197:449-56.

[86] Miyairi I, Laxton JD, Wang X et al. (2011). *Chalmydia psittaci* genetic variants differ in virulence by modulation of host immunity. J Infect Dis. 204(4):654-663.

[87] Kari L, Whitmire VM, Olivares-Zavaleta N et al. (2011). A live attenuated chlamydial vaccine protects against trachoma in non-human primates. J Exp Med 208(11):2217-23.

[88] Bryne GI, Lehmann L, Landry GJ (1986). Induction of tryptophan catabolism is the mechanism for gamma interferon-mediated inhibition of intracellular *Chlamydia psittaci* replication in T24 cells. Infect Immun 53:347-351.

[89] Caldwell HD, Wood H, Crane D et al. (2003). Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Investig 111:1757-1769.

[90] Tacchini L, Gammella E, De Ponti C et al. (2008). Role of HIF-1 and NF-kB transcription factors in the modulation of transferrin receptor by inflammatory and anti-inflammatory signals. J Biol Chem 283:20674-20686.

[91] Rau A, Wyllie S, Whitimore J et al. (2005). Identification of *Chlamydia trachomatis* genomic sequences recognized by chlamydial divalent cation-dependent regulator A (DcrA). J Bacteriol 187:443-448.

[92] Dill BD, Raulston JE (2007). Examination of an inducible expression system for limiting iron availability during *Chlamydia trachomatis* infection. Microbes Infect 9:947-953.

[93] Sixt BS, Siegl A, Muller C et al. (2013). Metabolic features of *Protochlamydia amoebophila* elementary bodies: A link between activity and infectivity in chlamydia. PLoS Pathog 9(8):e1003553.

Chapter 3 Study Rationale, Hypotheses, Aims and Objectives

3.1 Study Rationale

The Bijagós Archipelago of Guinea Bissau has a significant burden of both active and scarring trachoma, with some of the highest disease rates in the world. The reasons why disease is so prevalent and severe on these islands are unclear. Socio-behavioural, environmental and host-pathogen factors are likely to be important. Using the eye as an accessible clearly defined phenotype and whole genome sequence analysis in conjunction with next generation PCR technology, a unique opportunity existed to investigate the pathogenicity of ocular strains of *C. trachomatis* in this geographically remote treatment-naïve population. Population-based survey methodology was employed to investigate trachoma and ocular *C. trachomatis* infection in the context of spatial trends in disease, socio-environmental risk factor associations, clinical (including disease phenotype and bacterial load) and chlamydial whole genome sequence data.

3.2 Hypotheses

1. The Bijagós Archipelago is trachoma-hyperendemic.
2. There are socioeconomic and environmental risk factors associated with active trachoma on the study islands of the Bijagós Archipelago.
3. There are spatial processes underlying the distribution of active trachoma and ocular *C. trachomatis* infection.
4. There is spatial clustering of ocular *C. trachomatis* infection.
5. Higher bacterial loads of *C. trachomatis* are associated with more severe clinical disease in active trachoma.
6. There is an association between (virulence) plasmid copy number (per *C. trachomatis* genome) and clinical disease severity.
7. There are genes and loci present on the *C. trachomatis* plasmid and chromosome that are associated with ocular *C. trachomatis* load and disease severity in active trachoma.
8. A single round of community mass drug treatment with oral azithromycin will reduce the prevalence and severity of active disease and ocular *C. trachomatis* infection on the study islands.

3.3 Aims

1. To establish the prevalence and spatial distribution of trachoma and ocular *C. trachomatis* infection on the study islands.
2. To examine the relationship between active trachoma and ocular *C. trachomatis* infection and their socio-environmental risk factor associations.
3. To examine the spatial epidemiology of active trachoma and ocular *C. trachomatis* infection.
4. To examine spatial clustering of ocular *C. trachomatis* infection in relation to *C. trachomatis* bacterial load.
5. To examine the relationship between clinical disease severity in active trachoma and *C. trachomatis* bacterial load.
6. To examine the relationship between clinical disease severity, *C. trachomatis* ocular load and *C. trachomatis* genomic diversity.
7. To assess the impact of a single round of mass treatment with azithromycin in an endemic island setting in the context of the spatial and molecular epidemiology of trachoma and ocular *C. trachomatis* infection.

3.4 Specific objectives

1. To examine a population-based sample of individuals for clinical signs of trachoma and capture the clinical phenotype accurately using a detailed trachoma grading system and digital conjunctival photography.
2. To investigate the relationship between clinical phenotype and infection with *C. trachomatis* load using accurate estimates of bacterial load obtained using a quantitative next generation droplet digital PCR (ddPCR) using both plasmid and genomic targets.
3. To conduct a household survey to examine socio-environmental risk factors for disease at household and individual levels using mixed effects regression models.
4. To create prevalence maps for active trachoma and ocular infection with *C. trachomatis* using geographic information systems (GIS).
5. To examine the spatial structure of active trachoma and ocular *C. trachomatis* infection using measures of spatial autocorrelation and spatial regression models.

6. To examine spatial clustering of ocular *C. trachomatis* infection in relation to *C. trachomatis* bacterial load using local indicators of spatial association.
7. To obtain *C. trachomatis* whole genome sequence data directly from *C. trachomatis* ddPCR-positive conjunctival samples to investigate associations between clinical phenotype and infecting genotype of *C. trachomatis* using comparative genomics and pathogen genome-wide association scan analysis to identify genes and loci associated with disease severity on the *C. trachomatis* plasmid and chromosome.
8. To conduct a single round of community mass drug treatment with azithromycin on the study islands.
9. To evaluate the effect of a single round of mass drug treatment at one year on active trachoma and ocular *C. trachomatis* infection using geostatistical and molecular methods.

Chapter 4 Study Setting and Study Design

4.1 Study Setting

The study was conducted on the Bijagós Archipelago of Guinea Bissau, West Africa. The study was conducted on four islands (Bubaque, Canhabaque, Soga and Rubane) on the archipelago (*Figure 1*). These four islands comprise a total rural population of 5,613 (National Population Census, 2010, Instituto Nacional de Estatística, Guiné-Bissau). The Bijagós Archipelago has a total area of more than 10,000 km^2 and is situated approximately 60km off the coast of Guinea Bissau. It is possible to travel by boat between Bubaque and each of the other three study islands within one hour. The surface area covers approximately 900 km^2, of which 350 km^2 is covered with mangrove forest [1]. Altitude does not exceed 50m. The climate is humid and tropical, with a prolonged rainy season, during which average monthly rainfall is 400mm^3 [2]. The mean monthly temperature is 27.3^{0}C (25.1-29.2^{0}C), with peak temperatures occurring prior to the rainy season. Water sources are available within walking distance from all settlements throughout the year. Such climatic and environmental factors make these unique trachoma-endemic communities, differing greatly to communities inhabiting arid, dusty environments typical of trachoma-endemic regions.

There are 88 islands and islets of which only about 20 are permanently inhabited. The remainder are inhabited periodically for seasonal agriculture and traditional initiation ceremonies. Inhabitants of the island are predominantly fishermen and hunter-gatherers, leading a traditional lifestyle in villages within the forest (*Figure 2*).

Figure 1. The Bijagós Archipelago of Guinea Bissau

Figure 2. Communities on the Bijagós Archipelago

4.2 Study Design

A cross-sectional population-based household survey with geospatial representation on the four study islands was conducted to obtain prevalence estimates of infection with *C. trachomatis*, active and scarring trachoma reflecting the geographical setting, risk factor exposure, chlamydial genomic diversity and bacterial load. Mass drug administration of azithromycin was conducted in communities in accordance with WHO guidelines and evaluated with longitudinal follow-up of the treated cohort one year after treatment.

4.3 Sample size calculation and sampling strategy

The 2010 population census indicated a rural population of 5,613 and pilot work suggested a prevalence of TF (1-9 year-olds) of 25% on the four islands. Pilot data (*Appendix I*) were used to make the following calculations. A 1:5 random sample of households (with over-sampling in small communities to ensure geospatial representation) was expected to lead to an adequate sample of 1,500 people in all 38 villages (Bubaque (14 villages), Canhabaque (17 villages), Soga (5 villages) and Rubane (2 villages)). This is representative of a one stage (2^{nd} stage of cluster random sampling) probability sample design and satisfied desired criteria for population-based prevalence surveys [3,4].

We expected to find approximately 300 people with active trachoma and at least 225 with scarring sequelae. Pilot data (*Appendix I*) indicated the presence of infection in 60% of those with signs of active disease and 10% of those with sequelae. Thus we expected to find approximately 200 infections within our sample. Oversampling in small settlements aimed to increase the chance of finding an infection to represent that space-point for the spatial analysis. A minimum requirement of six households or 50 people per settlement was expected to yield five infections (or a 1/32 chance of failing to find a case of infection).

4.4 Statistical power calculation to detect risk factor associations

A sample of 1,500 individuals gave adequate power to detect associations of common risk factors with modest effect sizes and rare risk factors with large effect sizes using conservative correction (design effect of 4) for anticipated disease clustering. Power to detect a risk factor (ocular/nasal discharge) found in 20% of normal subjects and 33% of cases (OR (Odds Ratio) = 2, 95% confidence) is 0.92. Similarly, with 95% confidence, for a risk factor present in 5% of normal subjects with an OR of 3 in cases, power is 0.94 and a risk factor present in 3% of normal subjects with an OR of 5 in cases is 0.99.

Studies investigating the molecular diversity of *C. trachomatis* have focused on a limited number of genetic targets making it difficult to anticipate and quantify diversity in terms of allele frequency and strain distribution. Therefore, there was limited basis for calculating a sample size for this aspect of the work, particularly with the inherent challenges related to obtaining whole genome sequence data directly from clinical samples. To compensate for this, stringent p-value thresholds were

employed in the analysis of sequence data with relation to detecting associations with clinical phenotype (see ***Chapter 6 and Chapter 10***).

4.5 Precision for mapping

Pilot data (*Appendix I*) suggested that the prevalence of TF in 1-9 year olds ass >25% on all four study islands. The sample size calculation allowed the prevalence of TF in 1-9 year olds to be estimated at 25% on Rubane (±10%), Soga (±6%) and Bubaque and Canhabaque (±4%). This was deemed adequate to determine whether the islands would require community mass treatment with azithromycin as per WHO and national policy.

4.6 Participant inclusion and exclusion criteria

Inclusion Criteria

All individuals in the target population who consented or for whom consent was obtained were included in the study.

Exclusion Criteria

Individuals who declined to participate, were unable to give consent, or unable to participate in the examination or sample collection due to

- illness
- incapacity
- inability to communicate

were not included in the study.

4.7 Protocol for community mass treatment with oral azithromycin

Single dose directly observed mass treatment with oral azithromycin (20mg/kg (maximum 1g) for children over the age of 6 months) was administered to rural and semi-urban populations on the study islands according to WHO guidelines [5]. Individuals with known or suspected hypersensitivity to azithromycin, self-reported pregnant women and infants <6 months in age were offered tetracycline eye ointment to be applied for six weeks in accordance with national policy. A follow-up survey one year following treatment was conducted to evaluate the impact of treatment on pathogen diversity, clinical disease, risk factors and spatial distribution using the methodology described.

4.8 Chapter 4 References

[1] Van der Linde HA, Dankin MH (1998). Enhancing Sustainability: Resources for our future. SUI Technical Series Volume 1. Proceedings of a workshop held at the world conservation congress for the Sustainable Use Initiative (SUI). IUCN, The World Conservation Union, Montreal Canada.

[2] World Bank. From the Climatic Research Unit (CRU) at the University of East Anglia (UEA). (2013) Climate Change Knowledge Portal: Guinea Bissau. (http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisRegion=Africa&ThisCCode=GNB).

[3] Ngondi J, Reacher M, Matthews F et al. (2009). Trachoma survey methods: A literature review. Bull World Health Organ 87:143-51.

[4] Bennett S, Woods T, Liyanage WM, Smith DL (1991). A simplified general method for cluster-sample surveys of health in developing countries. World Health Stat Q 44:98-106.

[5] WHO-ITI (2004). Joint Research Agenda Meeting for the Elimination of Blinding Trachoma. World Health Organization, Geneva.

Chapter 5 Survey, Clinical and Laboratory Methods

5.1 Field team recruitment and training

A dedicated field team were recruited and trained at the start of the research study. They were recruited using a transparent and competitive interview process. They were trained in issues surrounding consent and confidentiality, sensitization and survey methodology. Specific training was given to field research assistants responsible for digital conjunctival photography and procurement and handling of conjunctival samples for laboratory analysis. Refresher training sessions were conducted throughout the study to ensure high quality field data collection.

5.2 Community sensitization

Discussions were held in each village with leaders and communities. Knowledge and opinions on trachoma and blindness (active disease, trichiasis, blindness) and to what extent they felt these were problems in the community were discussed. Information about trachoma, its transmission and treatment was given, using visual aids. Participation in the study was discussed and agreed at community level. Opportunity for questions to be answered was given at each stage of the process. The study was further explained to household heads and each adult participant using locally accepted infrastructure.

A '*Study Advocate*' (leader, health personnel, responsible adult) was identified in each village to aid community mobilisation. Broadcasts from the local radio station were used to inform about trachoma and publicise the study.

5.3 Household selection and census

Field assistants identified a responsible adult in the household to consent to census data collection. The adult was informed that their household may be included in the survey and that household members would be examined for clinical signs of trachoma and that conjunctival swabs would be taken. The concept of random selection was illustrated using local analogy. Household head lists were generated for each village (*Appendices II & III*) from which 1:5 households were randomly selected using a computerized random number generator (RANDBETWEEN, MS Office, Microsoft®Office). A responsible adult from each selected household was asked to provide the name, gender, and age for all individuals resident in the household using the nationally accepted definition of a household; those who are part of the same 'fugon' – i.e. they eat from the same cooking pot. Prior to examination days the field

assistants updated the census in the study villages to ensure correct details of the *de facto* population (i.e. those who slept in the household the night before).

5.4 Ethical approval

This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained from the Comitê Nacional de Ética e Saúde (Guinea Bissau), the LSHTM Ethics Committee (UK) and The Gambia Government/MRC Joint Ethics Committee (The Gambia). Ethical approval documentation is presented in *Appendix IV*.

5.5 Informed consent

Consent was obtained at the community level from the village leader. Household heads from selected households were asked for verbal consent for the household to participate in the risk factor survey after the study information leaflet's contents were explained in the appropriate local language. If the household head was absent, the most senior person identified by household members was requested to provide consent. For ocular examination, each adult participant was asked for written consent before participating in the study. For all children examined, written (or thumbprint) consent was obtained from the parent/guardian or household head and witnessed, with the witness also signing the consent form. If consent was not obtained, such households or individuals were excluded. See *Appendix V* for consent information and forms (English and Portuguese).

5.6 Risk Factor Survey

Geographical, demographic, socio-economic and environmental information was collected at the household and individual levels (*Appendix VI*). Household-level risk factor data were obtained using questionnaires administered to the household head and included the level of household head education, presence and use of latrines, water and sanitation and whether there had been a face washing health education programme in the community in the last year. Individual-level data on facial cleanliness (the presence of ocular and/or nasal discharge and whether or not there were flies on the face during examination) were collected at the time of examination (*Appendix VII*).

Demographic, socio-economic and environmental information was collected at household and individual levels. Household-level risk factor data were obtained using questionnaires administered to the household head or an appropriate responsible adult and included items on the level of education of the household head, their socio-

economic status, whether the household had been exposed to any health education or promotion within the community, household access to and use of latrines, access and use of water and measures of sanitation, waste and presence of flies in the environment. The questionnaire was supported through observational data collected on water use, latrine use and environmental sanitation. Household size (measured as number of members of all ages) and number of children under the age of 10 years within the household was recorded. The researcher (AL) was blinded to the household members' trachoma status at the time of collection of household level risk factor and geospatial data.

5.7 Geo-coding Data

To satisfy adequate geospatial representation at village-level, all 38 villages on the four study islands were included in the survey. Small villages may be over-represented by the minimum sampling criteria imposed: 1:5 households were sampled with a minimum number of households required per village (six) to ensure spatial representation of that settlement. In very small settlements with less than six households in total, all households were sampled. Data were geo-coded at household and village level using the Garmin eTrex H handheld Global Positioning Systems (GPS) unit (Garmin Ltd., UK).

Each Garmin eTrex H was coded for use during the surveys. Waypoints were marked with accuracy ≤ 10m. Waypoints were recorded on data forms using the device code, waypoint number and GPS coordinates (*Appendix VII*). Data were saved on the device and downloaded each day using Garmin MapSource® v6.16.3 (Garmin Ltd., UK). GPS coordinates were projected into UTM Zone 28N in MapSource® and subsequently merged with the main database using household codes.

5.8 Ocular Assessment

Following the household risk factor survey all individuals from study households were invited to attend for clinical examination and conjunctival sampling at a central examination point in the village. Individuals' age, sex and ethnic group and data on facial cleanliness (the presence of ocular and/or nasal discharge and whether or not there were flies on the face) were collected at the time of examination. The examination team performed up to two further visits to that village to maximise uptake of the examination. Reserve households were used if households were unable or unwilling to attend for examination.

A trained examiner (AL) assessed each participant using two grading systems (detailed in *Tables 1&2* in ***Chapter 1.4***). The WHO simplified [1] and modified FPC [2] grading systems were employed to assign a trachoma grade to the right and left upper tarsal conjunctivae of each consenting participant using adequate light and a 2.5x binocular magnifying loupe. Both methods were used in order that the data from this study should be comparable to data from other studies and with control programme parameters, and so that detailed information relating to disease severity was recorded to test the hypotheses for this study. In the modified FPC system, follicles (F), papillae (P), conjunctival scarring (C), trichiasis and entropion (T/E) and corneal opacification (CC) are each assigned a separate grade from 0-3. FPC grades of F2/3 or P3 equate to a diagnosis of active disease (TF or TI by the WHO simplified system).

The examiner (AL) was standardized annually (achieving Kappa scores of 0.85 using the WHO simplified grading system in the field and with conjunctival photographs) against an international expert in trachoma grading (RB), and conducted all conjunctival examinations and obtained all conjunctival swabs.

Ocular assessment for clinical signs of trachoma in the field

5.9 Conjunctival digital photography

Two digital photographs were taken of each left everted upper tarsal conjunctiva to record the clinical phenotype (*Appendix VIII*). These serve as a record of evidence for clinical phenotype at baseline that may be useful in future longitudinal studies. The WHO simplified grading and conjunctival digital photography methods are robust, reproducible and objective [3]. Both WHO and modified FPC grading systems have been used previously in similar settings [4,5].

5.10 Conjunctival sampling

Two specimens from the upper tarsal conjunctiva of the left eye of each participant were obtained with Dacron swabs (Fisher Scientific) using a validated,

well-tolerated standardised procedure involving sweeping the swab three times horizontally across the upper tarsal conjunctiva rotating the head of the swab by 1/3 with each pass [3] The first swab was stored in Copan Universal Transport Medium (UTM) (Copan Diagnostics Inc., USA) for chlamydial culture and whole genome sequencing. The second dry swab was collected for molecular diagnostics.

Obtaining samples and digital photographs of the conjunctiva in the field.

5.10.1 Maintenance of the cold chain

The swabs were kept on ice in the field and frozen to -80^{0}C within 8 hours of collection. Specimens were transported in batches in liquid nitrogen dry shippers from the field base to the MRC Laboratories in Fajara, The Gambia for continued storage at -80^{0}C. Following DNA extraction, sample extract was stored at 4^{0}C. Swabs collected in UTM for chlamydial culture were maintained at -80^{0}C throughout.

5.10.2 Specimen collection quality control (QC)

To avoid cross-contamination in the field, the examiner (AL) changed gloves before each participant was examined. A new packet containing two swabs was used for each participant. The swab was passed to the examiner by a field assistant in its newly opened packaging. The field assistant also wore gloves and providing that no used swabs were touched and there was no visible dirt on the gloves, they were changed after every 15 examinations. Swabs were stored in cryovials. Labels with unique swab identifiers were placed on each cryovial and the corresponding data forms (*Appendices VII, IX, X)*.

Control swabs (1 in every 100) were also taken to ensure both field and laboratory QC. These were swabs that were pre-marked and drawn at random from the swab dispenser in the field. Instead of passing the swab over the conjunctiva, when a control swab was taken, the field assistant would alert the examiner, who

would pass the swab 10cm in front of the eye ensuring that no contact was made between the swab tip and the participant. In all other respects, the control swab would be treated in the same way as any other, including labeling. The data form label was placed in a designated box on the form (*Appendices VII & X*). These swabs were processed in the same way as all other samples in the laboratory, with the laboratory operator masked to the fact that they were control swabs.

5.11 Laboratory processing of conjunctival specimens

5.11.1 DNA extraction

DNA was extracted from swabs by a single operator. After suspending the dry swab in sterile phosphate buffered saline (PBS), DNA was extracted from the dry swab using an adapted whole blood protocol on the QIAxtractor® (Qiagen, Crawley, UK) automated instrument (*Appendix XI*).

5.11.2 Detection of Chlamydia trachomatis

Qualitative detection of *C. trachomatis* DNA was performed using the validated Roche Amplicor® CT/NG PCR assay (Roche Molecular Systems, NJ USA) (*Appendix XII*). Quantitative detection was performed using a next generation droplet digital PCR assay that was evaluated against Amplicor during this study [6] (*Appendices XII-XIII*).

5.11.3 Quantitation of chlamydial load

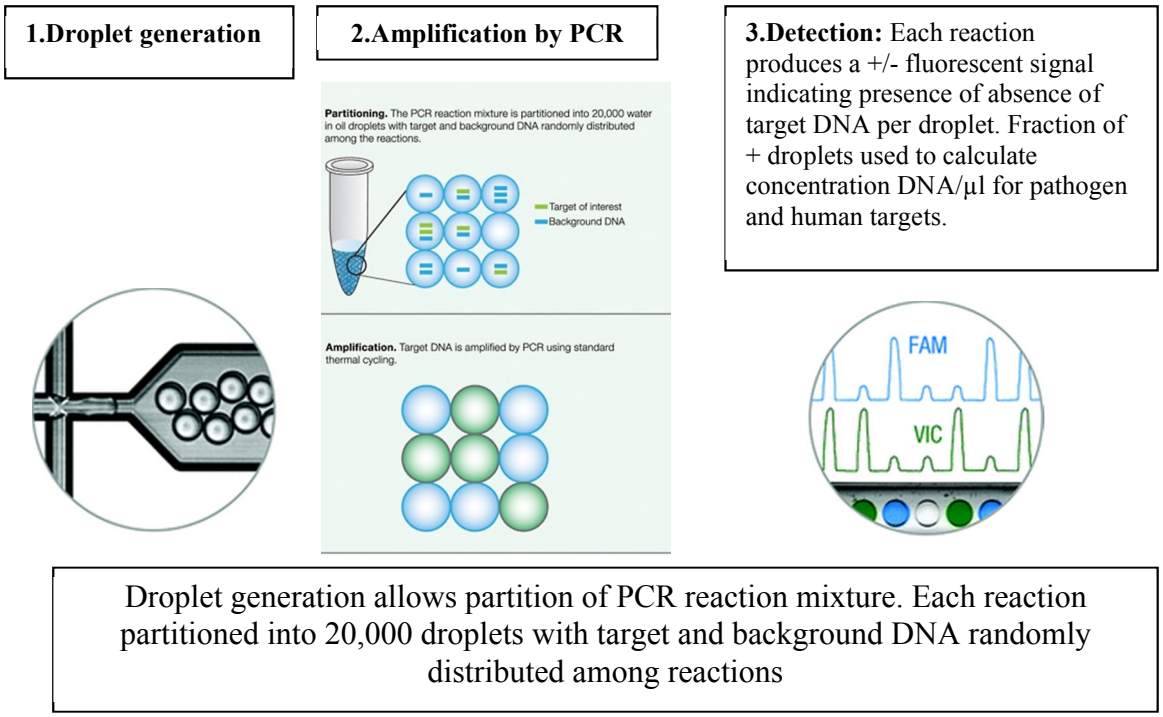
The Droplet Digital™ PCR (ddPCR™) system (Bio-Rad Laboratories, Hemel Hempstead, UK) was used to quantitate both the chlamydial plasmid and chromosome in each sample. ddPCR is considered the third generation of PCR technology, enabling the absolute quantitation of target DNA molecules, rather than a relative measure obtained with real-time PCR. Following generation of an oil/water emulsion droplet which partitions each sample into thousands of nanolitre-sized droplets, PCR is performed on a thermal cycler (C1000 Touch™). Each droplet contains PCR mastermix (11µl), nuclease-free water (3.85µl), forward and reverse primers and probes for the human DNA target (HURNASE-P), which acts as an endogenous control, and one of either the cryptic plasmid or the chromosomal target (*omcB*) (total 2.2µl of primer/probe supermix depending on chosen target). Each well containing droplets is inoculated with DNA extract (4.95µl). Four wells per plate serve as no template controls. After PCR, droplets from each sample are streamed

single file through the QX100™ droplet reader. Each droplet produces a fluorescent positive or negative signal indicating whether target DNA is present or not. Positive and negative droplets are counted in each sample and the concentration of target DNA is calculated (per µl) automatically by QX100™ software. *Table 1* shows the primer/probe sequences in the assay and *Figure 1* illustrates the process. Detection and quantitation methods have been published for use in this study [6-8].

Table 1. Primer and probe sequences for control and *C. trachomatis* targets using the ddPCR system

Molecular Target	**Nucleotide sequence and modification**
Internal Control: *Homo sapiens* RPP30 (encoding ribonuclease P/MRP 30kDa subunit)	
Forward (HURNASE-P-F)	5' AGA TTT GGA CCT GCG AGC G 3'
Reverse (HURNASE-P-R)	5' GAG CGG CTG TCT CCA CAA GT 3'
Probe (HURNASE_HEX_BHQ1)	5' FAM-TTC TGA CCT GAA GGC TCT GCG CG-BHQ1-3'
Target One: *C trachomatis* cryptic plasmid pLGV440 (circular; genomic DNA; 7500bp)	
Forward (Ct-Plasmid-F)	5' CAG CTT GTA GTC CTG CTT GAG AGA 3'
Reverse (Ct-Plasmid-R)	5' CAA GAG TAC ATC GTT CAA CGA AGA 3'
Probe (Ct-plasmid_FAM_BHQ1)	5' HEX-CCC CAC CAT TTT TCC GGA GCG A-BHQ1-3'
Probe (Ct-plasmid_HEX_BHQ1)	5' HEX-CCC CAC CAT TTT TCC GGA GCG A-BHQ1-3'
Target Two: *C. trachomatis* (Serovar A) *omcB* gene	
Forward (Ct-OMCB-F)	5' GAC ACC AAA CGC AAA GAC AAC AC 3'
Reverse (Ct-OMCB-R)	5' ACT CAT GAA CCG GAG CAA CCT 3'
Probe (Ct-OMCB-HEX-BHQ1)	5'FAM-CCA CAG CAA AGA GAC TCC CGT AGA CCG-BHQ1-3'

Figure 1. Droplet Digital PCR Workflow



Droplet generation allows partition of PCR reaction mixture. Each reaction partitioned into 20,000 droplets with target and background DNA randomly distributed among reactions

5.12 Whole genome sequencing of Chlamydia trachomatis

5.12.1 Pre-sequencing sample enrichment (SureSelect^{XT})

Whole genome sequence (WGS) data were obtained direct from clinical samples. DNA baits spanning the length of the *C. trachomatis* genome were compiled by SureDesign and synthesized by SureSelect^{XT} (Agilent Technologies, UK). *C. trachomatis* DNA extract from clinical samples was quantified and carrier human genomic DNA added to obtain a total of 3 µg input for library preparation. DNA was sheared using a Covaris E210 acoustic focusing unit. End-repair, non-templated addition of 3' –A adapter ligation, hybridisation, enrichment PCR and all post-reaction clean-up steps were performed according to the SureSelect^{XT} Illumina Paired-End Sequencing Library protocol (V1.4.1 Sept 2012). All recommended quality control measures were performed between steps.

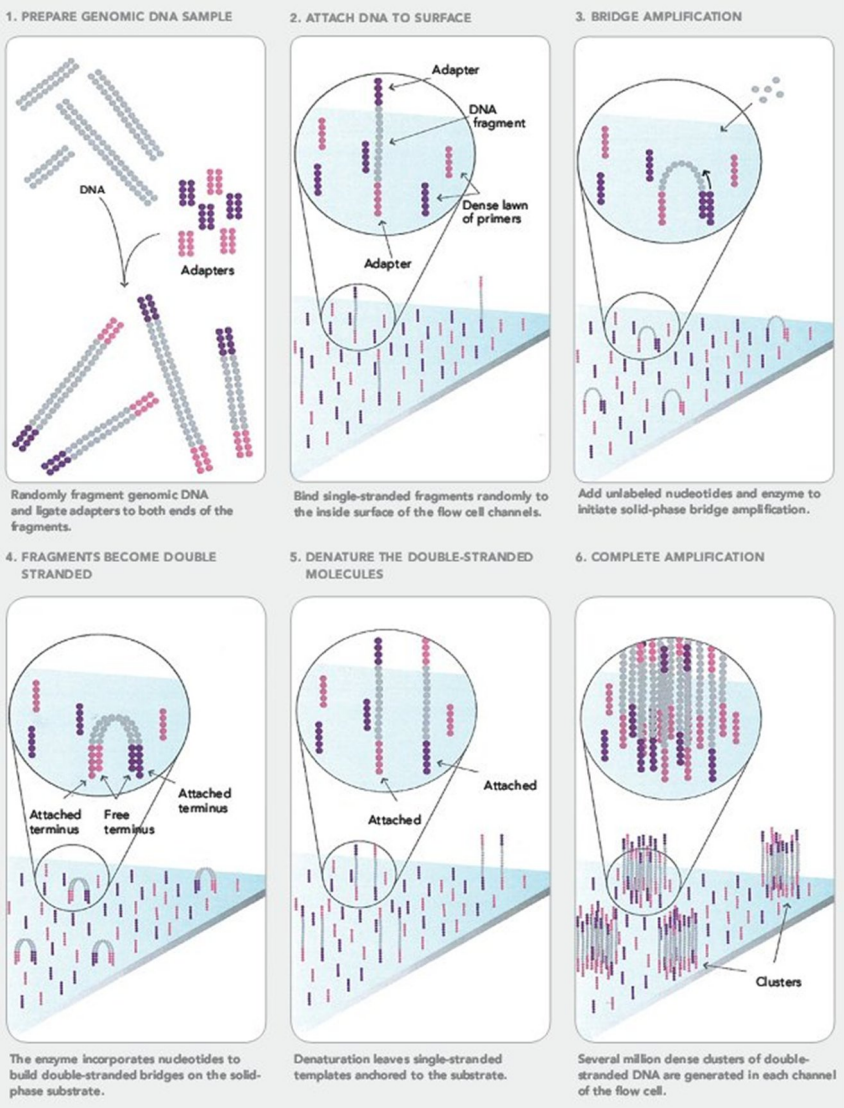
5.12.2 Chlamydial DNA sequencing

Pathogen whole genome sequencing was undertaken at the Wellcome Trust Sanger Institute.

5.12.2.1 Next generation Illumina sequencing

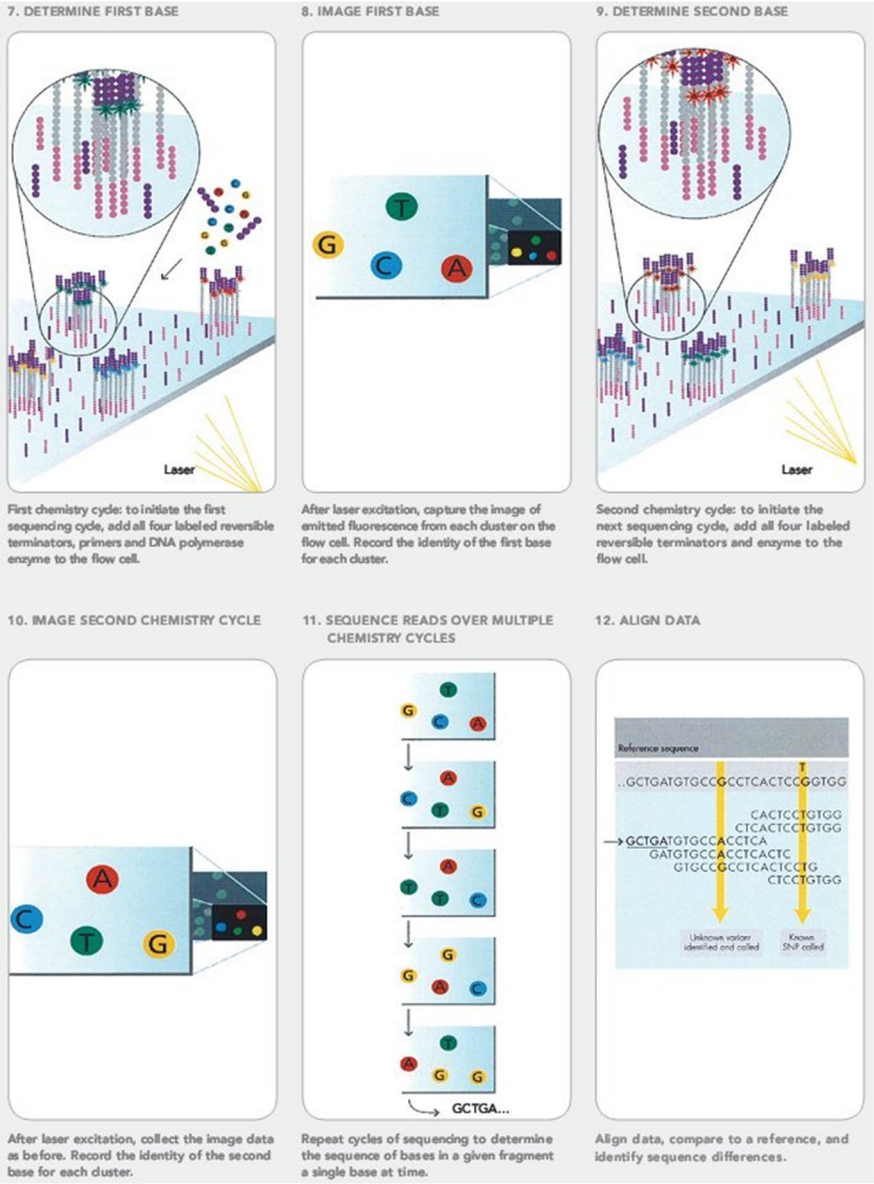
Illumina (Genome Analyzer IIx or HiSeq2000, Illumina Inc.) sequencing uses cluster generation followed by sequencing by synthesis methods using a reversible terminator-based method that enables detection of single bases as they are incorporated into growing DNA strands. A fluorescently-labelled terminator is imaged as each deoxyribonucleoside triphosphate (dNTP) is added and then cleaved to allow incorporation of the next base. Since all four reversible terminator-bound dNTPs are present during each sequencing cycle, natural competition minimises incorporation bias. The end result is true base-by-base sequencing (*Figures 2 & 3*).

Figure 2. Cluster generation by bridge amplification (Illumina)



Illumina utilizes a unique "bridged" amplification reaction that occurs on the surface of the flow cell.
The flow cell surface is coated with single stranded oligonucleotides that correspond to the sequences of the adapters ligated during sample preparation. Single-stranded, adapter-ligated fragments are bound to the surface of the flow cell exposed to reagents for polyermase-based extension. Priming occurs as the free/distal end of a ligated fragment "bridges" to a complementary oligo on the surface. (www.illumina.com)

Figure 3. Sequencing by Synthesis (Illumina)



A flow cell containing millions of unique clusters is loaded into the sequencer for automated cycles of extension and imaging. The first cycle of sequencing consists first of the incorporation of a single fluorescent nucleotide, followed by high resolution imaging of the entire flow cell. Any signal above background identifies the physical location of a cluster, and the fluorescent emission identifies which of the four bases was incorporated at that position. This cycle is repeated, one base at a time, generating a series of images each representing a single base extension at a specific cluster. The use of physical location to identify unique reads is a critical concept for all next generation sequencing systems. (www.illumina.com)

5.13 Chapter 5 References

[1] Dawson CR, Jones BR, Tarizzo ML (1981). Guide to trachoma control in programs for the prevention of blindness. Geneva: World Health Organization.

[2] Thylefors B, Dawson CR, Jones BR et al. (1987). A simple system for the assessment of trachoma and its complicatios. Bull World Health Organ 65(4):477-83.

[3] Stare D, Harding-Esch E, Munoz B et al. (2011). Design and baseline data of a randomised trial to evaluate coverage and frequency of mass treatment with azithromycin: The Partnership for the Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. Ophthalmic Epidemiol 18(1):20-9.

[4] Burton MJ, Holland MJ, Faal N et al. (2003). Which members of a community need antibiotics to control trachoma? Conjunctival Chlamydia trachomatis infection load in Gambian villages. Invest Ophthalmol Vis Sci 10(42):4215-22.

[5] Burton MJ, Holland MJ, Jeffries D et al. (2006). Conjunctival Chlamydial 16S Ribosomal RNA Expression in Trachoma: Is Chlamydial Metabolic Activity Required for Disease to Develop? Clin Infect Dis 42:463-70.

[6] Roberts CH, Last A, Molina-Gonzalez S et al. (2013). Development and evaluation of a next generation digital PCR diagnostic assay for ocular *Chlamydia trachomatis* infections. J Clin Microbiol 51(7):2195-2203.

[7] Last A, Roberts CH, Cassama E et al. (2014). Plasmid copy number and disease severity in naturally occurring ocular Chlamydia trachomatis infection. J Clin Microbiol 52(1):324-7.

[8] Butcher R, Last A, Herrera B et al. (2014). Evaluation of a *Chlamydia trachomatis* specific quantitative PCR assay by comparison to droplet digital PCR. Chlamydial Infections: Proceedings of the 13th International Symposium

on Human Chlamydial Infection. International Chlamydia Symposium, San Francisco, CA, USA. pp.217-220.

Chapter 6 Analytical Methods

6.1 Data cleaning

Demographic, socio-economic, clinical and laboratory data were double entered into a customised database (Microsoft Access 2007) and discrepancies resolved through reference to original data forms. Data were cleaned and analysed in STATA 13 (Stata Corporation, College Station, Texas USA) and the R statistical package v3.0.2 (The R Foundation for Statistical Computing (http://www.r-project.org).

6.2 Prevalence and socio-environmental risk factor association analysis

Random effects logistic regression models were used to assess the variability between villages and households assuming a three tier hierarchy to the data (at village, household and individual levels). Null models were used to examine the effect of cluster variables on the outcome using the likelihood ratio test (LRT). The log likelihood and the LRT were used to compare models.

Univariable associations with active trachoma (TF/TI) and infection with *C. trachomatis* were examined using two-level hierarchical random effects logistic regression, accounting for between-household variation. Covariates associated with active trachoma or *C. trachomatis* infection with $p<0.10$ (using the Wald test) were sequentially added to multivariable models after *a priori* adjustment for age and gender. Covariates were retained in the final model if the Wald p-value ≤ 0.05 unless otherwise specified. Further exploration of environmental predictor variables was conducted using logistic and hierarchical random effects logistic regression models as appropriate using the same criteria. *C. trachomatis* infection is on the causal pathway between several risk factors and active trachoma, therefore models with and without *C. trachomatis* infection were fitted. The model including *C. trachomatis* infection provides estimates of independent associations of other risk factors with active trachoma which are not mediated through *C. trachomatis* infection. Statistical significance was determined at the 5% level.

6.3 Disease severity and C. trachomatis infection association analysis

C. trachomatis load data were log(e) transformed due to the presence of significant negative skew. The geometric mean of load, respective standard error (SE) and 95% confidence intervals (CI) were calculated. An analysis of variance (ANOVA) with pair-wise comparisons was used to compare load across detailed

clinical phenotypes. Assessment of group differences and multiple comparisons were adjusted for using the Scheffé correction [1]. Associations between load and detailed clinical phenotype were examined using univariable and multivariable mixed effects linear and logistic regression models accounting for household-level clustering as described in section ***6.2***.

6.4 Geospatial analysis of disease and infection

Geocoded data were projected into UTM Zone 28N. ArcGIS 10.1 (ESRI Inc., USA) and the R statistical package v3.0.2 (The R Foundation for Statistical Computing (http://www.r-project.org) using spdep, automap and nlme packages) were used for geostatistical analyses [2-4].

6.4.1 Conceptualisation of spatial relationships

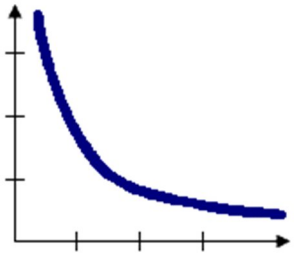
Spatial statistics integrate space and spatial relationships into statistical models. This requires adequate conceptualisation of spatial relationships within these models. There are a number of methods that can be used to conceptualise spatial relationships. These include inverse distance, fixed distance, zone of indifference, K nearest neighbour, and contiguity models. The choice of model is dependent on the nature of the observations in the data set.

The first three strategies are discussed as they are most relevant to the point data collected in this study. Other strategies are more suited to polygon data. The inverse distance models represent distance decay or impedance suggesting that the impact is reduced with increasing distance (*Figure 1(A)*). Fixed distance models are useful when fixed distances are known, such as travel-related distance. The fixed distance model imposes a sliding window across the data, outside which observations do not influence the model (*Figure 1(B)*). Neighbouring observations within this window are all weighted equally. The zone of indifference (*Figure 1(C)*) combines both inverse distance and fixed distance concepts, such that observations within a threshold distance are included as they would be within a fixed distance model. Beyond that critical threshold cut-off, the level of influence decays, as with the inversed distance model.

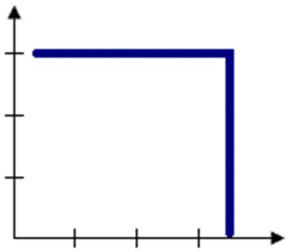
Fixed distance and zone of indifference methods are most appropriate with point data. The zone of indifference has the added benefit that it does not impose sharp boundaries on neighbouring relationships. The zone of indifference includes all observations as neighbours of each other.

Figure 1. Representation of the conceptualisation of key spatial relationships for use in geostatistical analyses.

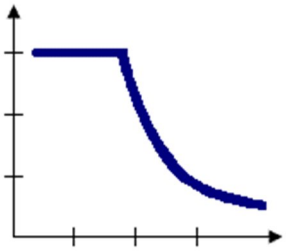
(A) Inverse distance (distance decay)



(B) Fixed distance window



(C) The zone of indifference



Where y = Impact/Weight and x = Distance

In the spatial analyses conducted, the zone of indifference was used to define adjacency. This method assumes that each observation has local influence that decreases with distance beyond a critical distance cut-off, resulting in an adapted model of impedance, or distance decay, such that all features have an impact on all other features, but this impact decreases with distance. This method is appropriate for point data and takes into account the extent of spatial autocorrelation in selecting a data-driven threshold cut-off of 1700m, derived from semivariogram analysis (see **Chapter 8**).

6.4.2 Spatial autocorrelation

Spatial analysis offers powerful methods to study the relationship between disease and infection. The spatial relationship between disease and infection defined over space is termed spatial dependence. This is measured as the existence of statistical association (dependence) between disease and infection associated with location. To explore spatial dependence measures of spatial autocorrelation are used.

Spatial autocorrelation assumes that near objects are more related than distant objects. In the context of infectious disease, this concept may be applied to explore the relationship between infection and disease where the spatial distribution can be used to define underlying processes such as exposure or transmission. Spatial regression models can be used to enhance current understanding of the epidemiology infectious disease.

Global spatial autocorrelation statistics are used to describe the overall spatial patterns in the data. Local indicators of spatial association take actual values of observations in the context of adjacent values allowing mapping and definition of clusters at exact locations. This allows for the examination of small-scale heterogeneity and identification of statistically significant clusters and outliers of similar observations in space.

6.4.2.1 Global Spatial Autocorrelation: Moran's I

The Moran's I statistic was used to evaluate global spatial autocorrelation in the distribution of the household-level prevalence of disease and infection. The values of Moran's I range between -1 and +1. A value close to -1 indicates negative autocorrelation (complete dispersion) whilst a value close to +1 indicates positive autocorrelation (clustering) of features. A value close to 0 suggests random arrangement. Z-scores and p-values are assigned to ascertain whether spatial autocorrelation is statistically significant [5]. Since the Moran's I statistic can be sensitive to skewed distributions, a permutation test was used to verify the results. This uses Monte Carlo simulations to generate a Moran's I sampling distribution [6].

Empirical semivariograms were constructed to obtain the distance range over which spatial autocorrelation occurs using the average squared distance between paired data values against the distance (or lag) separating the pairs to estimate the spatial covariance structure of the data [7].

6.4.2.2 Local Indicators of Spatial Association

In this analysis, 'low' values include both low values related to low load infections and null (zero) values related to uninfected individuals. To explore this, these analyses were conducted first on the whole data set, including uninfected individuals with a 'zero' value for infectious load, and then on infected individuals only.

6.4.2.2.1 Local (Anselin) Moran's I

Clustering and outlier analysis was conducted using the local (Anselin) Moran's I statistic [8]. A positive value for I indicates that a feature has neighbouring features with similarly high or low attributable values (a cluster). A negative value for I indicates that neighbouring features have dissimilar values and that this feature is an outlier. Cluster-Outlier analysis identifies spatial clusters of features with high or low values and spatial outliers by calculating Moran's I, Z-scores, p-values and a code to represent the cluster type. Cluster types are defined as clusters of high values (HH, High-High), clusters of low values (defined as above) (LL, Low-Low), a high value outlier surrounded by predominantly low values (HL, High-Low) and a low value outlier surrounded by predominantly high values (LH, Low-High).

*6.4.2.2.2 Getis-Ord Gi**

The Getis-Ord Gi* statistic was used to identify hot and cold spots for infection [9]. The Getis-Ord Gi* identifies statistically significant 'hotspots' containing similarly high values and statistically significant 'cold spots' containing similarly low values, taking into account the spatial dependence of neighbouring observations [10]. Z-scores and p-values are expressed for each Gi* value in comparison with the normal distribution of Gi* calculated by simulation. These values represent the statistical significance of the spatial clustering of values given the conceptualisation of spatial relationships and scale of analysis described. Statistically significant high z-scores show clustering of high load infections, whilst statistically significant low z-scores show clustering of low (or zero) load infections.

6.4.3 Geographically weighted spatial regression analysis

Geographically weighted spatial regression allows for modelling of spatially variable relationships, such that the statistical model can vary depending on location. Linear and logistic mixed effects regression models were applied to examine the

covariance structure and compare models with and without a spatial component. The log likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare models [11].

6.5 Chlamydia trachomatis whole genome sequence analysis

In order to obtain the highest quality data set for comparative genomics analysis, a data-driven approach is required. This is an iterative process and each stage of the pipeline can inform the next in terms of appropriate algorithms employed. Interrogation of informatics data can determine the most appropriate downstream analysis, therefore for this study we employed this data-driven approach resulting in the following pipeline for quality control and analysis to address the research hypotheses. Raw sequencing data in fastq format are available from the Wellcome Trust Sanger Institute through ftp://ftp.sanger.ac.uk/pub/pathogens/projects/LSHTM.

6.5.1 Raw Data Quality Assessment (FastQC)

FastQC is a quality control application for high throughput sequence data available from the Babraham Bioinformatics Institute (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw fastq files were used to obtain results from standard QC checks. FastQC provides detail on basic quality statistics of each fastq file, such as total sequence length, GC content (%), read length and filtered sequences. It summarizes per base sequence quality, sequence length distribution and information on duplicate and overrepresented sequences, adapter content and kmer content.

6.5.2 Raw Data Trimming (Trimmomatic)

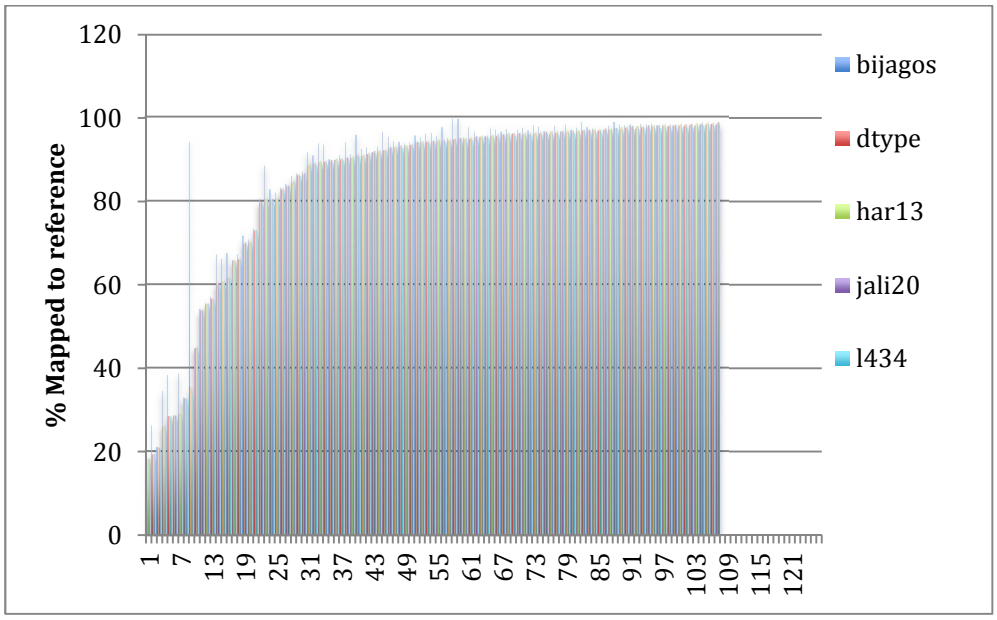
Trimmomatic is a flexible pre-processing trimming tool for paired-end sequence data. It includes a variety of processing steps for read trimming and filtering with the main algorithmic innovations related to identification of adapter sequences and quality filtering. This algorithm first removes technical sequences using a simple user-supplied technical sequence that is matched to the sequence data, and a 'palindromic' approach where short read lengths of adapter contamination are found at the ends of the reads. Quality filtering then employs standard sliding window and maximum information approaches. The latter ensures that the trimming process becomes increasingly stringent as it progresses through the reads, rather than applying a fixed quality threshold, it uses an adaptive approach incorporating a minimum length threshold, coverage (based on retained sequence length) and error rate (where

error probabilities from read quality socres are used to determine the accumulated likelihood of errors over the read). Trimmomatic performs well in challenging scenarios with poor quality sequence data [12].

6.5.3 Mapping to reference strains

Sequences were aligned to publically available reference strains *Chlamydia trachomatis A/HAR-13* (GenBank Accession Number NC_007429.1), *C. trachomatis B/Jali20/OT* (EMBL Accession Number FM872308), *C. trachomatis* L2 434 Bu (EMBL Accession Number AM884176), *C. trachomatis* D/13-96 (GenBank Accession Number NC_022119) and an assembled *C. trachomatis* sequence from pilot studies (Bijagos_86) using BWA [13]. Results from preliminary mapping to all reference strains was similar, therefore *C. trachomatis A/HAR-13* was chosen as the reference for subsequent analyses (*Figure 2*).

Figure 2. *C. trachomatis* whole genome sequences from study samples mapped to *C. trachomatis* reference strains using BWA



bijagos = assembled *C. trachomatis* genome from current clinical sample set, dtype = *C. trachomatis D/13-96,* har13 = *C. trachomatis A/HAR-13*, jali20 = *C. trachomatis B/Jali20/OT,* l434 = *C. trachomatis L2 434 Bu*

SAMTOOLS/BCFtools (SAMTOOLS) [14] and GATK [15] were used to call SNPs and small indels.

6.5.4 Phylogenetic analysis using RAxML

The best-scoring maximum likelihood phylogenetic trees were constructed based on SNPs across the genome using RAxML v7.4.2 [16] using all SNP sites spanning the genome.

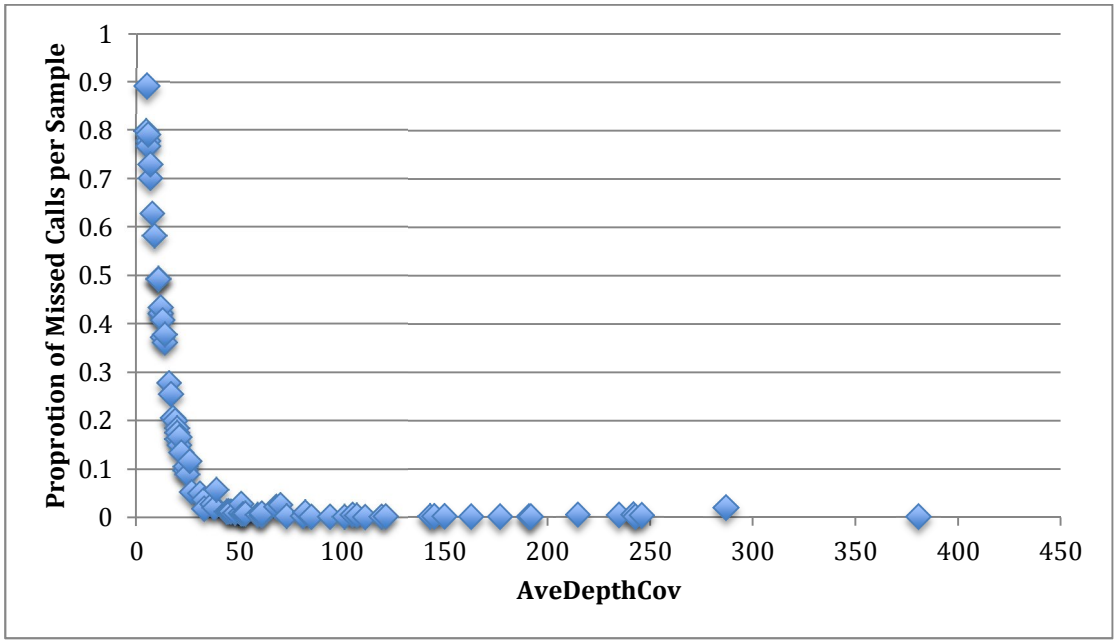
6.5.5 Recombination analysis using PhiPack and RAxML

Three compatibility-based recombination detection methods were applied to detect recombination hotspots using PhiPack [17]: the pairwise homoplasy index (Phi) [17], the maximum Chi2 [18] and the neighbour similarity score (NSS) [19] across the genome alignment. The confidence in the phylogenetic tree was examined by computing RAxML site-based likelihood scores across the genome [16]. Final phylogenetic trees were constructed adjusting for recombination [16].

6.5.6 Single Nucleotide Polymorphism (SNP) calling algorithms

Variants were selected as the intersection dataset between those obtained using both SNP callers and SNPs were further quality-filtered by mappability and uniqueness using the R statistical package v3.0.2 (The R Foundation for Statistical Computing, http://www.r-project.org) [20]. Non-unique SNPs with mappability values greater than one were filtered and SNP alleles were called using an alternative coverage-based approach where a missing call was assigned to a site if the total coverage was less than 20x depth or where one of the four nucleotides accounted for at least 80% total coverage [20]. There was a clear relationship between the average depth of coverage and proportion of missing calls, which provided a further quality filtering threshold (*Figure 3*). Based on these criteria, all sequences with greater than 10x average depth of coverage over the whole genome were retained.

Figure 3. Average depth of coverage of reads across the genome compared to the proportion of missing calls by SNP calling algorithm



6.5.7 Chlamydial diversity using Tajima's D

Tajima's D was calculated using the pegas and ape packages in R. Tajima's D tests the neutral mutation hypothesis by DNA polymorphism such that a negative Tajima's D implies an excess of low frequency polymorphisms suggesting population expansion and/or purifying selection and a positive Tajima's D indicates low levels of both high and low frequency polymorphisms indicating a decreasing population size and/or balancing selection. A negative Tajima's D suggests that rare alleles may be present at low frequency, indicating a recent population expansion following a bottleneck event. A positive Tajima's D suggests that there are multiple alleles present, some at low and some at high frequencies, and sudden population contraction causing balancing selection [21].

6.5.8 Chlamydial genome wide association scan (GWAS)

Genome wide association scans (GWAS) are powerful tools to assess the excess association between a SNP and defined phenotype. This method is confounded by population structure and linkage disequilibrium (LD) [22]. Therefore principal component analysis (PCA) was used to elucidate the population structure. Principal components from this were included in the GWAS models to adjust for this. Genomic inflation was corrected for with the occurrence of a polymorphism in the population of over 90% and a minor allele frequency of 5% in binary and linear models (3% in

ordinal models). To avoid the occurrence of false positive associations we used a conservative Bonferroni correction for multiple testing for genome-wide significance based on retaining 759 SNPs in the analysis. This resulted in a threshold p-value for genome-wide significance of 6.25xe-05.

SNP mutations were aggregated by coding and intergenic regions in an *n x m* matrix, where *n*=number of loci and *m*=number of sequences. In coding regions only non-synonymous mutations were aggregated.

Regression analyses were performed adjusting for age and population structure using the first three principal components. Logistic regression models were used to assess the association between disease phenotype (binary models for active trachoma and ordinal models for conjunctival inflammation) and SNPs across the *C. trachomatis* genome using MASS and glmm packages in R. Linear regression models with *C. trachomatis* load as a continuous outcome were used to assess the association between *C. trachomatis* ocular load and SNPs across the genome. The statistical significance of each variable was tested using the Wald test. The standard error for the log odds ratio was estimated from the model and used to calculate 95% confidence intervals for odds ratios (OR).

6.6 Chapter 6 References

[1] Ruxton GD, Beauchamp G (2008). Time for some a priori thinking about post hoc testing. Behav Ecol 19(3):690-3.

[2] Bivand R (2010). Spdep: Spatial dependence: Weighting schemes, statistics and models. R package version 0.5-4. (http://CRAN.Rproject.org/package=spdep).

[3] Hiemstra PH, Pebesma EJ, Twenhofel CJW, Heuvelink GBM (2009). Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. Computers & Geosciences 35(8):1711-1721.

[4] Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2014). *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-117 (http://CRAN.R-project.org/package=nlme).

[5] Moran PA (1950). Notes on continuous stochastic phenomena. Biometrika 37(1-2):17-23.

[6] Cliff AD, Ord JK (1981). Spatial Processes. Pion pp63-5.

[7] Diggle PJ, Tawn JA, Moyeed RA (1998). Model-based geostatistics. Royal Statistics Society. Applied Statistics 47(3):299-350.

[8] Anselin L (1995). Local indicators of spatial Association: LISA. Geographical Analysis 27(2):93-115.

[9] Getis A, Ord JK (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis 24(3).

[10] Ord JK, Getis A (1995). Local spatial autocorrelation statistics: Distributional issues and an application. Geographical Analysis 27(4).

[11] Zucchini W. An Introduction to Model Selection. J Mathematical Psychol 2000;44:41-61.

[12] Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-20.

[13] Langmead B, Trapnell C, Pop M et al. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.3.

[14] Li H, Handsaker B, Wysoker A et al. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

[15] McKenna A, Hanna M, Banks E et al. (2010). The Genome Analysis Tool Kit: A MapReduce framework for analysing next generation DNA sequencing data. Genome Res 20:1297-1303.

[16] Stamatakis A, Hoover P, Rougemont J (2008). A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 57:758-771.

[17] Bruen TC, Philippe H, Bryant D (2006). A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665-2681.

[18] Smith JM (1992). Analysing the mosaic structure of genes. J Mol Evol 34(2):126-9.

[19] Jakobsen IB, Easteal S (1996). A program for calculating and displaying compatibility matices as an aid in determining reticulate evolution in molecular sequences. Comput Appl Biosci 12(4):291-5.

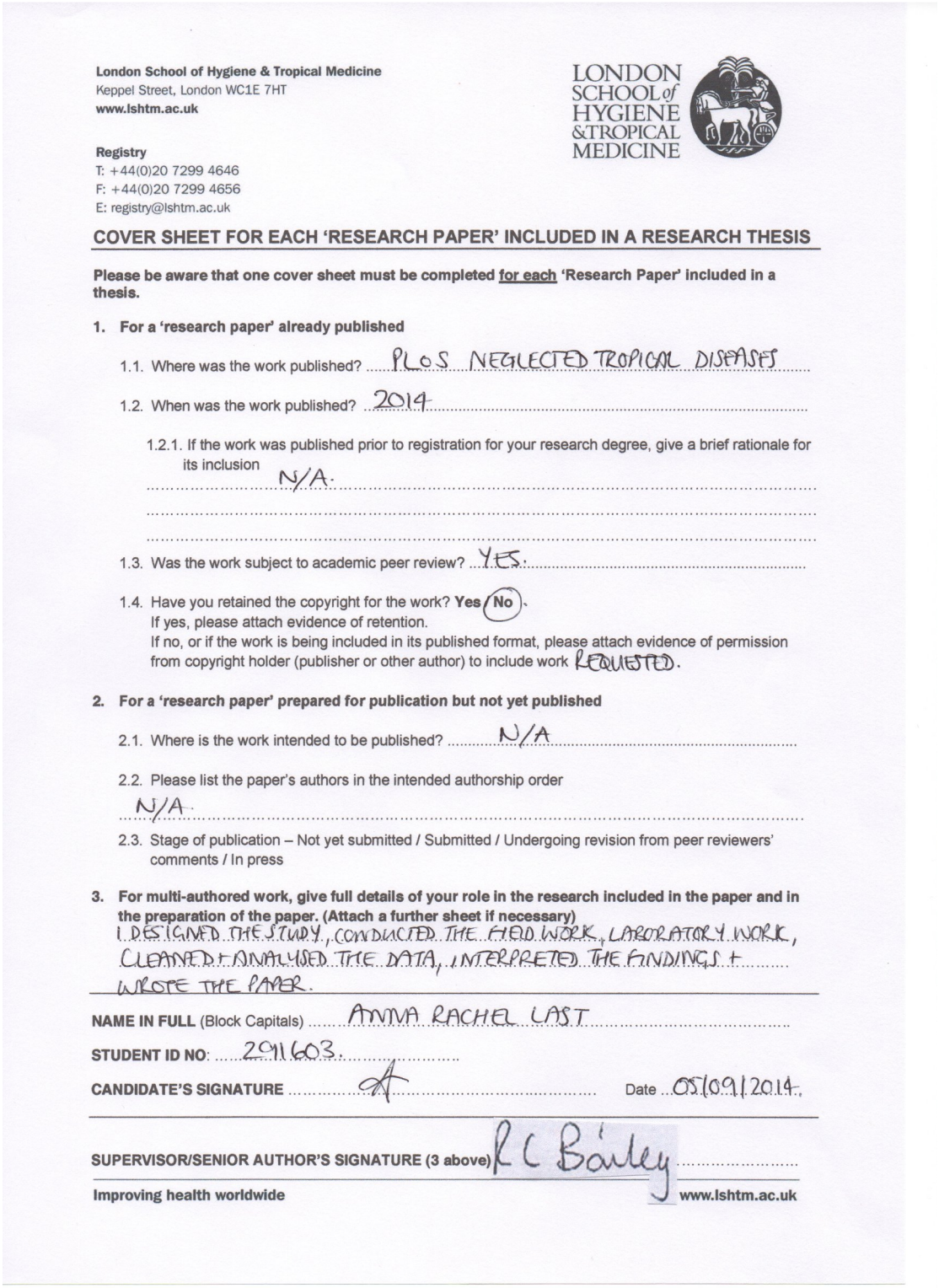
[20] Coll F, Preston M, Guerra-Assuncao JA et al. (2014). PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) 94(3):346-54.

[21] Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585-93.

[22] Thomas A, Abel HJ, Di Y et al. (2011). Effect of linkage disequilibrium on the identification of functional variants. Genet Epidemiol 35(Suppl 1):S115-9.

Chapter 7

Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. **For a 'research paper' already published**

1.1. Where was the work published? PLoS NEGLECTED TROPICAL DISEASES

1.2. When was the work published? 2014

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A.

1.3. Was the work subject to academic peer review? YES.

1.4. Have you retained the copyright for the work? **Yes** / **No** (No).
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work REQUESTED.

2. **For a 'research paper' prepared for publication but not yet published**

2.1. Where is the work intended to be published? N/A

2.2. Please list the paper's authors in the intended authorship order
N/A.

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. **For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)**
I DESIGNED THE STUDY, CONDUCTED THE FIELD WORK, LABORATORY WORK, CLEANED + ANALYSED THE DATA, INTERPRETED THE FINDINGS + WROTE THE PAPER.

NAME IN FULL (Block Capitals) ANNA RACHEL LAST

STUDENT ID NO: 291603.

CANDIDATE'S SIGNATURE A Date 05/09/2014.

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) R C Bailey

Improving health worldwide www.lshtm.ac.uk

Title

Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau

Short Title Risk Factors for Trachoma on the Bijagós Archipelago

Authors

Anna R Last[1#], Sarah E Burr[2], Helen A Weiss[3], Emma M Harding-Esch[1], Eunice Cassama[4], Meno Nabicassa[4], David C Mabey[1], Martin J Holland[1], Robin L Bailey[1]

[1]Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

[2]Disease Control and Elimination Theme, Medical Research Council Unit, The Gambia, Fajara, The Gambia

[3]MRC Tropical Epidemiology Group, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

[4]Programa Nacional de Saúde de Visão, Ministério de Saúde Publica, Bisssau, Guiné Bissau

[#]Corresponding Author

Abstract

Background Trachoma, caused by ocular infection with *Chlamydia trachomatis*, is hyperendemic on the Bijagós Archipelago of Guinea Bissau. An understanding of the risk factors associated with active trachoma and infection on these remote and isolated islands, which are atypical of trachoma-endemic environments described elsewhere, is crucial to the implementation of trachoma elimination strategies.

Methodology/Principal Findings A cross-sectional population-based trachoma prevalence survey was conducted on four islands. We conducted a questionnaire-based risk factor survey, examined participants for trachoma using the World Health Organization (WHO) simplified grading system and collected conjunctival swab samples for 1507 participants from 293 randomly selected households. DNA extracted from conjunctival swabs was tested using the Roche Amplicor CT/NG PCR assay. The prevalence of active (follicular and/or inflammatory) trachoma was 11% (167/1508) overall and 22% (136/618) in 1-9 year olds. The prevalence of *C. trachomatis* infection was 18% overall and 25% in 1-9 year olds. There were strong independent associations of active trachoma with ocular and nasal discharge, *C.*

trachomatis infection, young age, male gender and type of household water source. *C. trachomatis* infection was independently associated with young age, ocular discharge, type of household water source and the presence of flies around a latrine.
Conclusions/Significance In this remote island environment, household-level risk factors relating to fly populations, hygiene behaviours and water usage are likely to be important in the transmission of ocular *C. trachomatis* infection and the prevalence of active trachoma. This may be important in the implementation of environmental measures in trachoma control.

Author Summary
Trachoma, caused by ocular infection with *Chlamydia trachomatis*, is the leading infectious cause of blindness worldwide. The World Health Organization elimination strategy includes community mass treatment with oral antibiotics, education regarding hygiene and facial cleanliness and environmental improvements. Population-based trachoma prevalence surveys are essential to determine whether community interventions are required. Knowledge of risk factors associated with trachoma and *C. trachomatis* infection in a particular setting may help prioritise trachoma elimination activities. We conducted a trachoma prevalence survey to establish the prevalence of active (follicular and/or inflammatory) trachoma and *C. trachomatis* infection on the Bijagós Archipelago of Guinea Bissau. We also collected household risk factor data from survey participants. Active trachoma prevalence was 11% overall and 22% in children aged 1-9 years. *C. trachomatis* infection prevalence was 18% overall and 25% in children aged 1-9 years. Active trachoma and the presence of *C. trachomatis* infection were strongly correlated. Risk factors for disease and infection were similar. In this environment, measures of facial cleanliness (ocular and nasal discharge) and household-level risk factors relating to fly populations, hygiene behaviours and water usage are likely to be important in *C. trachomatis* transmission. This may have implications in the implementation of trachoma elimination activities.

Introduction

Trachoma is caused by ocular infection with *Chlamydia trachomatis* and is the leading infectious cause of blindness worldwide. It manifests as distinct clinical syndromes beginning with an acute self-limiting keratoconjunctivitis, which

following repeated episodes may progress to a more chronic inflammatory and immunofibrogenic process leading to conjunctival scarring and blinding sequelae. Trachoma is endemic in 50 countries, with 325 million people at risk of blinding disease [1]. Trachoma is responsible for visual impairment in 1.2 million people and 3% of blindness globally [1]. The highest prevalence of active trachoma (trachomatous inflammation-follicular (TF) and/or trachomatous inflammation-intense (TI)) is in sub-Saharan Africa and the distribution of disease is heterogeneous [2].

Ocular *C. trachomatis* is probably transmitted between individuals through direct spread from eye to eye during close contact, direct or indirect spread of infected nasal or ocular secretions on fingers or cloths (fomites) and indirect passive transmission by eye seeking flies. There is no known animal reservoir of *C. trachomatis* in endemic environments, the primary reservoir being young children.

Blinding trachoma is usually found in hot, arid, dusty regions. A recent systematic review examined studies reporting higher trachoma prevalence in savannah areas and areas of lower rainfall, and found weak but consistent evidence supporting anecdotal findings that trachoma is associated with semi-arid environments [3].

This study was conducted on the Bijagós Archipelago, a remote group of islands off the coast of Guinea Bissau with a total population estimated at 24,000 [4], where trachoma is hyperendemic. The climate and environment are not typical of trachoma-endemic areas. The islands are covered with subtropical forest and altitude does not exceed 50m. The climate is tropical, hot and humid. The islands are surrounded by mangroves and mudflats. There is significant rainfall (average 400mm/month) from May to November [5].

Many studies have suggested that the prevalence of trachoma is associated with environmental risk factors such as poor sanitation, access to water and latrine use [6,7]. Eye-seeking flies (*Musca sorbens*) have also been associated with trachoma as passive vectors [8] but significant disease exists in areas where fly populations are scarce and are therefore less likely to contribute to trachoma transmission [9]. *M. sorbens* preferentially breeds in human faeces and there may be association between

fly populations and lack of latrine access or use [6,8]. Social risk factors such as migration events and crowded living conditions have also been shown to be important in transmission of *C. trachomatis* and the appearance of active trachoma [10,11].

Clustering of disease at the community, household and bedroom levels has been noted and is likely to reflect the dynamics of transmission between family members with prolonged close contact [6,10-12]. Most transmission events have been shown to occur at the household level with more gradual spread within the community [13].

The World Health Organization (WHO) advocates the implementation of the SAFE strategy (Surgery for trichiasis, Antibiotics for active infection, Facial cleanliness to prevent disease transmission and Environmental improvement to increase access to water and sanitation) for trachoma elimination. The WHO recommends annual mass treatment of entire communities with oral azithromycin for three years if the prevalence of TF in 1-9 year olds within a district or community exceeds 10%. Mass antibiotic treatment aims to clear infection from communities such that transmission ceases to be a public health concern [14]. Following this, an assessment is made of A, F and E interventions and a decision is taken to continue or cease treatment [15].

Despite their inclusion in the SAFE strategy, local environmental factors are not well understood, though many are potentially modifiable risk factors for infection and disease. The relative importance of these risk factors is not clear and may differ between communities. Fewer studies have investigated risk factors for disease and infection simultaneously [16-19]. Understanding risk factors associated with trachoma and *C. trachomatis* infection may increase our understanding of disease and transmission dynamics allowing for optimization of community-specific interventions.

We examined household and individual-level risk factor associations with ocular *C. trachomatis* infection and active trachoma in this unique environment, where trachoma is a significant public health problem. Prior to these surveys, these communities were treatment-naïve and had not been exposed to any trachoma control interventions.

Methods

Ethical Approval

This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained from the Comitê Nacional de Ética e Saúde (Guinea Bissau), the LSHTM Ethics Committee (UK) and The Gambia Government/MRC Joint Ethics Committee (The Gambia). Verbal consent was obtained from community leaders. Written informed consent was obtained from all study participants or their guardians on their behalf if participants were children. A signature or thumbprint is considered an appropriate record of consent in this setting by the above ethical bodies.

Study Design and Population

We conducted a cross-sectional population-based trachoma prevalence survey on four islands of the Bijagós Archipelago of Guinea Bissau (Bubaque, Canhabaque, Soga and Rubane) in January 2012. Trachoma survey methodology has been described previously [20-22]. We randomly sampled one in five households, representing a one stage probability sample design satisfying desired criteria for population-based prevalence surveys [20,21]. A sample size of 1500 ensured adequate power with conservative correction (using a design effect of 4) to account for anticipated household clustering. The sample size provides over 90% power to detect an odds ratio (OR) of 2 associated with a risk factor found in 20% of subjects without disease or infection, or an OR of 3 for a risk factor present in 5% of subjects without disease or infection with 95% confidence. The sample size also provides good precision for an estimated TF prevalence of >25% in 1-9 year olds on the four islands of Bubaque and Canhabaque (±4%), Soga (±6%) and Rubane (±10%), which is adequate to determine whether these communities require mass drug treatment with azithromycin in line with WHO policy.

Household Census

A census of persons resident in randomly selected households was conducted prior to the household survey. Residency was defined as living within the household for longer than the preceding month or intending to stay resident in the household for longer than one month. This was updated to reflect the *de facto* population (those present in the household on the previous night) to limit absenteeism.

Household Risk Factor Survey

Demographic, socio-economic and environmental information was collected at household and individual levels. Household-level risk factor data were obtained using questionnaires administered to the household head or an appropriate responsible adult and included items on the level of education of the household head, their socio-economic status, whether the household had been exposed to any health education or promotion within the community, household access to and use of latrines, access and use of water and measures of sanitation, waste and presence of flies in the environment. The questionnaire was supported through observational data collected on water use, latrine use and environmental sanitation. Household size (measured as number of members of all ages) and number of children under the age of 10 years within the household was recorded. Researchers were masked to trachoma status of household members at the time of the household survey.

Trachoma Prevalence Survey

Following the household risk factor survey all individuals from study households were invited to attend for clinical examination and conjunctival sampling. Individuals' age, sex and ethnic group and data on facial cleanliness (the presence of ocular and/or nasal discharge and whether or not there were flies on the face) were collected at the time of examination.

Clinical Assessment

A single trained examiner assessed each participant using the WHO simplified grading system where TF (trachomatous inflammation – follicular) and/or TI (trachomatous inflammation – intense) constitute active trachoma and TS (trachomatous scarring), TT (trachomatous trichiasis) and CO (corneal opacity) are trachomatous sequelae which may lead to blindness [23]. A trachoma grade was assigned to the upper tarsal conjunctivae of each consenting participant using adequate light and a 2.5x binocular magnifying loupe.

Conjunctival Sampling

Two sequential samples were taken from the left upper tarsal conjunctiva of each participant with Dacron swabs (Fisher Scientific, UK) using a standardised procedure [24,25]. The first swab was collected into transport medium for other studies. The

second dry swab was collected into a microcentrifuge tube (Simport, Canada) and used in this study. Previous work using the Roche Amplicor CT/NG assay (Roche Molecular Systems, NJ USA) in a population-based study has shown that there was good agreement between first and second swabs with respect to *C. trachomatis* DNA positivity by PCR [26]. Swabs were kept on ice in the field and frozen to -80^{0}C within 8 hours of collection.

Measures were taken to avoid cross-contamination in the field. Control swabs (pre-marked swabs drawn at random from the swab dispenser and passed 10cm in front of the eye ensuring no contact between the swab tip and participant) were taken to ensure field and laboratory quality control.

Community Mass Treatment

After survey completion all communities on the study islands were treated with a single height-based dose of oral azithromycin in accordance with WHO and national protocols.

DNA Extraction

Each swab was suspended in 400µl sterile phosphate buffered saline (PBS) after thawing at room temperature. DNA was extracted from the swab/PBS suspension using an adapted whole blood protocol on the QIAxtractor (Qiagen, Crawley, UK) automated instrument and eluted into a final volume of 50µl DX Elution Buffer (Qiagen).

Detection of infection with *C. trachomatis*

C. trachomatis DNA was detected using the Roche Amplicor CT/NG assay (validated for use with ocular swabs [27]). Required reaction buffer conditions were obtained as described previously and used in the standard assay [28]. Positive and negative samples were assigned according to the manufacturer's instructions. In this study, *C. trachomatis* infection is defined as the presence of *C. trachomatis* DNA by Amplicor PCR.

Statistical Analyses

Data were double entered into a customised database (MS Access 2007). Discrepancies were resolved through reference to original data forms. Data were

further cleaned prior to analysis in STATA 13 (Stata Corporation, College Station, Texas USA).

Random effects logistic regression models were used to assess the variability between villages and households assuming a three tier hierarchy to the data (at village, household and individual levels). Null models were used to examine the effect of cluster variables on the outcome using the likelihood ratio test (LRT), which if significant, provided strong evidence that between-village and household variance was non-zero. The log likelihood and the LRT were used to compare models.

Univariable associations with active trachoma (TF/TI) and infection with *C. trachomatis* were examined using two-level hierarchical random effects logistic regression, accounting for between-household variation. Covariates associated with active trachoma or *C. trachomatis* infection with $p<0.10$ (using the Wald test) were sequentially added to the multivariable model after *a priori* adjustment for age and gender (as categorical variables). Covariates were retained in the final model if the Wald p-value ≤ 0.05 unless otherwise specified. Further exploration of environmental predictor variables was conducted using logistic and hierarchical random effects logistic regression models as appropriate using the same criteria. As *C. trachomatis* infection is on the causal pathway between several risk factors and active trachoma, models with and without *C. trachomatis* infection were fitted. The model including *C. trachomatis* infection provides estimates of independent associations of other risk factors with active trachoma which are not mediated through *C. trachomatis* infection.

All statistical analyses were carried out using STATA 13. Statistical significance was determined at the 5% level.

Results

Study Population Characteristics

From an estimated total rural population of 5,613 inhabitants on the four study islands [4], 1,511 individuals from 293 randomly selected households across 39 villages were enrolled. Of these, 1,508 had an ocular assessment and conjunctival swabs were obtained from 1,507. The median age of participants was 13 years (range 1 month -

88 years) and 57% were female. The majority of participants were of the Bijagós ethnic group (*Table 1*).

Prevalence of Active Trachoma and *C. trachomatis* Infection

The prevalence of active trachoma in 1-9 year olds was 22.0% (95% Confidence Interval (CI) 18.9-25.5%) (136/618). The prevalence of active trachoma was highest in children under the age of 5 years (27.3% (95% CI 23.1-31.9%) (113/416)). Overall, 11.1% (95% CI 9.4-12.6%) (167/1508) of the study population had active trachoma. The relationship between trachoma and infection is shown in *Table 2*. *C. trachomatis* DNA was detected in 18.0% overall (269/1507) and 25.4% of 1-9 year olds (157/618). All 15 (~1% of total) control swabs were negative for *C. trachomatis* DNA.

Multilevel Clustering of Active Trachoma and *C. trachomatis* Infection

Null models for both active trachoma and *C. trachomatis* infection adjusted for age and gender showed significant clustering at island, village and household levels. For active trachoma, the variance estimated due to between-household clustering was 1.11 (standard error (SE) 0.17, $p<0.0001$). The between-village clustering variance was 0.75 (SE 0.16, $p<0.0001$) and between-island clustering variance was 0.50 (SE 0.28, $p=0.0100$). For *C. trachomatis* infection, the variance estimated due to between-household clustering was 1.37 (SE 0.15, $p<0.0001$), between-village clustering was 0.89 (SE 0.14, $p<0.0001$) and between-island clustering was 0.40 (SE 0.18, $p=0.0005$). The clustering effect was strongest at household level and models adjusting for clustering at household level only were a better fit than those including adjustment for village and island clustering. Adjusting for clustering at household level significantly improved the model versus standard logistic regression analyses ($p<0.0001$). Two-level hierarchical regression models with adjustment for household level clustering are presented in this analysis.

Factors associated with active trachoma

Univariable associations with active trachoma are presented in *Table 3.* The final multivariable model showed that active trachoma was strongly independently associated with *C. trachomatis* infection (OR=11.2 (95% CI 6.9-18.1)), ocular (OR=2.0 (95% CI 1.0-4.0)) and nasal (OR=2.5 (95% CI 1.5-4.3)) discharge, male

gender (OR=1.9 (95% CI 1.2-2.9)) and being aged 0-5 years (OR=10.2 (95% CI 5.1-20.4) compared to being >15 years of age) (Model 2, *Table 4*). There was also a strong independent association between household water access and active trachoma, such that households with access only to a traditional natural spring as a water source had an increased risk of active trachoma compared to households with access to multiple water sources (OR=1.9 (95% CI 0.9-3.9)). The model without *C. trachomatis* infection shows stronger associations, indicating that some effect of these factors is mediated through *C. trachomatis* infection (*Table 4*). Comparison of the two models suggests that some of the effect of younger age and water source is partly mediated through *C. trachomatis* infection, but these remain independently associated with trachoma beyond this effect.

Factors associated with *C. trachomatis* infection

Univariable associations with *C. trachomatis* infection are presented in *Table 5*. In the final multivariable model *C. trachomatis* infection was strongly independently associated with being aged ≤10 years. The presence of ocular discharge (OR=2.3 (95% CI 1.3-4.4)) and household access only to a traditional natural spring (OR=6.6 (95% CI 2.8-15.2)) and or access to a single water source only (OR=3.9 (95% CI 1.9-8.0)) (rather than households who had access to multiple water sources) were strongly associated with infection (*Table 6*). The presence of flies around a latrine was also independently associated with infection (OR=2.1 (95% CI 1.1-3.8)). The presence of flies around a latrine were strongly associated with the presence of flies in the environment surrounding the household (OR=8.3 (95% CI 5.4-12.7), $p<0.0001$) and the presence of visible faeces within the latrine (OR=46.7 (95% CI 28.5-76.6), $p<0.0001$). There was no association between flies in the environment (OR=1.1 (95% CI 0.4-3.0), $p=0.91$) nor flies around the latrine (OR=0.5 (95% CI 0.1-2.6), $p=0.43$) with flies on the face at the time of examination.

Discussion

We have shown that there is clustering of both active trachoma and infection at household, village and island levels. We found that the most substantial clustering was present at household-level, thus the random effects introduced by this are adjusted for using two-level hierarchical mixed effects models. Clustering at household level has been described in other populations [10,13,29]. We found *C.*

trachomatis infection to be more strongly clustered at the household level than active trachoma, which has also been reported in Tanzanian communities [18].

Consistent with results from other studies, we found that being aged 0-5 years [10,16,29,30-34], the presence of ocular *C. trachomatis* infection [17] and the presence of ocular and nasal discharge [19,32,35,36] were strongly associated with active trachoma. We also found that being female was associated with a reduced odds of active trachoma, but not infection. The finding of an association between gender and active trachoma are contradictory in the literature [36-38]. In this analysis gender was included in all models due to *a priori* assumptions.

At the household level, the only household-level factor independently associated with both active trachoma and *C. trachomatis* infection was access to type and number of water sources, where household access to a traditional natural spring as the only water source was more strongly associated with active trachoma and *C. trachomatis* infection than access to another (single) source or multiple water sources of various types. This association is likely to be mediated through the presence of *C. trachomatis* infection. A similar association was shown in Ethiopia, where a piped water supply was associated with reduced risk of active trachoma compared to spring or river/lake water sources [37]. This may have been related to increased distance to the non-piped water sources [34] or to volume of water used [33]. In Bijagós communities, every household has access to a water source within 15 minutes walking distance. It is likely that this association may reflect more subtle hygiene and water use behaviours related to allocation of water collected for specific hygiene practices. A greater volume of water can be obtained with greater ease from an improved well or borehole than is possible from a natural spring. Although we found no statistically significant association between the volume of water stored in the household and active trachoma, it may be that the volume of water specifically allocated for hygiene practices is important, as shown in The Gambia [39], and that this finding is a proxy for water use allocation.

Children aged 0-5 years are at increased risk of ocular infection with *C. trachomatis*. In 6-10 year olds the risk is similarly elevated. In contrast, children aged 0-5 years are at much greater risk of active trachoma than 6-10 year olds. This supports suggestions

that young children are the main reservoir of infection [10,19,32] and demonstrates that this reservoir persists beyond pre-school age children in this population. The difference found in associations between age and active trachoma and infection may be explained by the presumed acquisition of immunity in childhood, which may result in reduced prevalence of clinically active trachoma despite the remaining strong association with infection. Duration [40] and intensity [41] of active trachoma and *C. trachomatis* infection have been shown to be age-dependent. It may also be that clinical signs present in 0-5 year olds are less specific for *C. trachomatis* infection and that other bacteria such as *Streptococcus pneumoniae* and *Haemophilus influenzae* contribute to the clinical phenotype, such that they induce a recall of the follicular response in children previously exposed to *C. trachomatis*, as hypothesised in cross-sectional studies showing an association between follicular trachoma and these bacteria in Tanzania and The Gambia [42,43]. Overall, there is a diminishing association with age, indicating that age may also represent a proxy for close contact and specific hygiene behaviours, acquired immunity and duration of infection. Whilst there were strong univariable associations between infection and ocular and nasal discharge, the association with nasal discharge is not significant in the multivariable model. Ocular discharge may more specific for *C. trachomatis* infection than nasal discharge, likely representing a consequence of *C. trachomatis* infection rather than a risk factor for it.

Latrine access was not associated with active trachoma or *C. trachomatis* infection, despite the fact that 60% of individuals do not have access to a latrine. Latrine access does not translate into latrine use [44]. We assessed latrine use through observation of the latrines in addition to collecting reported data. Although there is high reported use of latrines within a household with latrine access, there may be household members that do not use the latrine. In these populations children aged 1-9 years are often discouraged from using latrines due to caregiver concerns about their safety when using latrines (Thompson *et al*., unpublished data).

The presence of flies around a household pit latrine was strongly associated with infection in the final multivariable model, although there was no association between flies on the faces of individuals at the time of examination with either active trachoma or infection. The presence of flies on faces has been well-documented as a risk factor

for active trachoma in some environments [8,29,45] and there is evidence that the eye-seeking flies *Musca sorbens* (and possibly other domestic muscidae) are passive vectors in *C. trachomatis* transmission[45-48]. One study that examined fly density around used household pit latrines in The Gambia found that the majority of flies emerging from the latrine were *Chrysomya albiceps*, rather than *M. sorbens* [47]. We do not have data on the species of flies around pit latrines in Bijagós communities, however it is possible that there is ovipositioning by *M. sorbens* in or around latrines in these communities. We found a strong association between flies present in the environment and flies around the latrine, although there was no association with flies on the face at the time of examination. With decreasing age, there was a trend in association between flies on the face and flies around the latrine, but this association was non-significant.

M. sorbens have been shown to preferentially breed in human excreta, in addition to domestic animal faeces [46]. Although the presence of animal and human excreta within 15 metres of the household were not risk factors for active trachoma or *C. trachomatis* infection, they were ubiquitous in the villages. There was a strong association noted between flies around a latrine and the presence of visible faeces within a latrine. These findings suggest that flies may be important in the transmission of trachoma in this environment. Our study was conducted in the cooler dry season, possibly explaining the lack of association between flies on faces and in the environment and *C. trachomatis* infection in the regression analysis. Fly populations and density have been shown to have seasonality elsewhere [49,50].

The cross-sectional design of this study is a limitation in understanding determinants of disease and infection. There may also be some limitations in using stepwise regression methods such that models may be overfitted or unstable [51]. Risk factors are often difficult to define and despite the presence of statistically significant associations, it is likely that the relationship between socio-behavioural and environmental risk factors and disease and infection is complex and not fully explained by this analysis. It may be that intra- and inter-familial transmission occurs in these communities as a consequence of specific hygiene practices and complex socio-behavioural factors. These factors are often difficult to define and require longitudinal studies beyond the scope of the current study. Further epidemiological

studies examining the species and contribution of flies to trachoma transmission and water use and specific hygiene behaviours in relation to trachoma and *C. trachomatis* infection in this population may be of importance in successful trachoma control activities.

Conclusion

We have described individual and household-level risk factor associations with active trachoma and ocular infection with *C. trachomatis* on the Bijagós Archipelago to improve our understanding of the relationship between disease and infection in this remote treatment-naïve trachoma-hyperendemic population.

These data suggest that in this environment household-level risk factors relating to fly populations, hygiene behaviours and water usage are likely to be important in the transmission of ocular *C. trachomatis* infection. Education about cleanliness, sanitation and hygiene practices is likely to be important in reducing transmission of infection in these communities. Ensuring the provision of water sources which allow adequate water to be allocated for hygiene may assist this, and further studies examining specific hygiene practices may be useful. Reducing fly populations around the latrines where they exist may be of benefit. These findings may be important in the implementation of the F and E components of SAFE in this population.
In order to fully understand the factors associated with active trachoma and ocular *C. trachomatis* infection in these communities, further epidemiological studies examining transmission and clustering of *C. trachomatis* infection are required. These studies should focus on pathogen factors such as the role of infection intensity and strain diversity, and socio-behavioural factors such as specific hygiene behaviours.

Acknowledgements

Azithromycin was donated for treatment of trachoma on the islands by Pfizer Inc. through application via the International Trachoma Initiative. We extend our thanks to our colleagues in the Programa Nacional de Saúde de Visão in Bissau and to the communities of the Bijagós Archipelago, our field research team and all study participants in Guinea Bissau. We acknowledge the contribution from KM Thompson's unpublished data to our discussion.

Supporting Information Legends

Checklist S1: STROBE Checklist (*Appendix XIV*)

References

1. World Health Organization (2012). Prevention of blindness and visual impairment (Priority Eye Diseases): Trachoma. https://www.who.int/blindness/cause/priority/en/index2.html.
2. Solomon AW, Peeling RW, Foster A, Mabey DCW (2004). Diagnosis and Assessment of Trachoma. Clin Micro Rev 17(4):982-1011.
3. Ramesh A, Kovats S, Haslam D, Schmidt E, Gilbert CE (2013). The impact of climatic risk factors on prevalence, distribution, and severity of acute and chronic trachoma. PLoS Negl Trop Dis 7(11):e2513.
4. National Population Census (Bolama-Bijagós Region) (2010). Instituto Nacional de Estatística, Bissau, Guiné-Bissau.
5. World Bank (2013). From the Climatic Research Unit (CRU) at the University of East Anglia (UEA). Climate Change Knowledge Portal: Guinea Bissau. http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisRegion=Africa&ThisCCode=GNB
6. Burton MJ, Holland MJ, Faal N, Aryee EA, Alexander ND et al. (2003). Which members of a community need antibiotics to control trachoma? Conjunctival *Chlamydia trachomatis* infection load in Gambian villages. Invest Ophthalmol Vis Sci 44:4215-4222.
7. Emerson PM, Cairncross S, Bailey RL, Mabey DC (2000). Review of the evidence base for the 'F' and 'E' components of the SAFE strategy for trachoma control. Trop Med Int Health 5:515-527.
8. Emerson PM, Lindsay SW, Alexander N, Bah M, Dibba SM et al. (2004). Role of flies and provision of latrines in trachoma control: cluster randomized controlled trial. Lancet 363:1093-1098.
9. Taylor HR, Velasco FM, Sommer A (1985). The ecology of trachoma: an epidemiological study in southern Mexico. Bull World Health Organ 63:559-567.
10. Bailey R, Osmond C, Mabey DC, Whittle HC, Ward ME (1989). Analysis of the household distribution of trachoma in a Gambia village using a Monte Carlo simulation procedure. Int J Epidemiol 18:944-951.

11. Burton MJ, Holland MJ, Makalo P, Aryee EA, Alexander ND (2005). Re-emergence of *Chlamydia trachomatis* infection after mass antibiotic treatment of a trachoma-endemic Gambian community: a longitudinal study. Lancet 365:1321-8.
12. West SK, Munoz B, Turner VM, Mmbaga BB, Taylor HR (1991). The epidemiology of trachoma in central Tanzania. Int J Epidemiol 20:1088-1092.
13. Blake IM, Burton MJ, Bailey RL, Solomon AW, West SK (2009). Estimating household and community transmission of ocular *Chlamydia trachomatis*. PLoS Negl Trop Dis 3:e401.
14. Kuper H, Solomon AW, Buchan J, Zondervan M, Foster A et al. (2003) A critical review of the SAFT strategy for the prevention of blinding trachoma. Lancet Infect Dis 3:372-381.
15. WHO (2004). Report of the eighth meeting of the WHO Alliance for the Global Elimination of Blinding Trachoma, Geneva: Switzerland. http://www.who.int/blindness/publications/GET_8_Report.pdf
16. Abdou A, Nassirou B, Kadri B, Moussa F, Muñoz B et al. (2007) Prevalence and risk factors for trachoma and ocular Chlamydia trachomatis infection in Niger. Br J Ophthalmol 91:13-17.
17. West SK, Rapoza P, Muñoz B, Katala S, Taylor HR (1991). Epidemiology of ocular chlamydial infection in a trachoma-hyperendemic area. J Infect Dis 163:752-756.
18. Harding-Esch E, Edwards T, Mkocha H, Muñoz B, Holland MJ et al. (2010). Trachoma prevalence and associated risk factors in The Gambia and Tanzania: Baseline results of a cluster randomized controlled trial. PLoS Negl Trop Dis 4(11):e861.
19. Edwards T, Harding-Esch EM, Hailu G, Andreason A, Mabey DC et al. (2008). Risk factors for active trachoma and Chlamydia trachomatis infection in rural Ethiopia after mass treatment with azithromycin. Trop Med Int Health 13:556-565.
20. Ngondi J, Reacher M, Matthews F, Brayne C, Emerson P (2009). Trachoma survey methods: A literature review. Bull World Health Organ 87:143-51.
21. Bennett S, Woods T, Liyanage WM, Smith DL (1991). A simplified general method for cluster-sample surveys of health in developing countries. World Health Stat Q 44:98-106.

22. Wright HR, Vu H, Taylor HR (2005). How to assess the prevalence of trachoma. Br J Ophthalmol 89:526-7.
23. Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR (1987). A simple system for the assessment of trachoma and its complications. Bull World Health Organ 65(4):477-83.
24. Keenan JD, Lakew T, Alemayehu W, Melese M, Porco TC (2010). Clinical activity and polymerase chain reaction evidence of chlamydial infection after repeated mass antibiotic treatments for trachoma. Am J Trop Med Hyg 82(3):482-7.
25. Stare D, Harding-Esch E, Munoz B, Bailey R, Mabey D (2011). Design and baseline data of a randomised trial to evaluate coverage and frequency of mass treatment with azithromycin: The Partnership for the Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. Ophthalmic Epidemiol 18(1):20-9.
26. Michel CE, Roper KE, Divena MA, Lee HH, Taylor HR (2011). Correlation of clinical trachoma and infection in Aboriginal communities. PLoS Negl Trop Dis 5(3):e986.
27. Leslie DE, Azzalo F, Ryan N, Fyfe J (2003). An assessment of the Roche Amplicor *Chlamydia trachomatis/Neisseria gonorrhoeae* multiplex PCR assay in routine diagnostic use on a variety of specimen types. Commun Dis Intell Q Rep 27(3):373-9.
28. Holland MJ, Jeffries D, Pattison M, Korr G, Gall A (2010). Pathway focused arrays reveal increased matrix metalloproteinase-7 (Matrilysin) transcription in trachomatous trichiasis. Invest Ophthalmol Vis Sci 51(8):3893-3902.
29. Hagi M, Schemann JF, Mauny F, Momo G, Sacko D (2010). Active trachoma among children in Mali: Clustering and environmental risk factors. PLoS Negl Trop Dis 4(1):e583.
30. Schemann JF, Sacko D, Malvy D, Momo G, Traore L et al. (2002). Risk factors for trachoma in Mali. Int J Epidemiol 31:194-201.

31. Alemayehu W, Melese M, Fredlander E, Worku A, Courtright P et al. (2005). Active trachoma in children in central Ethiopia: association with altitude. Trans R Soc Trop Med Hyg 99:840-3.
32. Cumberland P, Hailu G, Todd J (2005). Active trachoma in children aged three to nine years in communities in rural Ethiopia: prevalence, indicators and risk factors. Trans R Soc Trop Med Hyg 99:120-7.
33. Polack S, Kuper H, Solomon AW, Massae PA, Abuelo C et al. (2006). The relationship between the prevalence of active trachoma, water availability and its use in a Tanzanian village. Trans R Soc Trop Med Hyg. 100:1075-1083.
34. Baggaley RF, Solomon AW, Kuper H, Polack S, Massae PA et al. (2006). Distance to water source and altitude in relation to active trachoma in Rombo district, Tanzania. Trop Med Int Health. 11:220-7.
35. Mesfin MM, de la Camera J, Tereke IG, Ananual G, Araya T et al. (2006). A community-based trachoma survey: Prevalence and risk factors in the Tigray region of northern Ethiopia. Ophthalmic Epidemiol 13:173-81.
36. Harding-Esch EM, Edwards T, Sillah A, Sarr-Sissoho I, Aryee EA et al. (2008). Risk factors for active trachoma in The Gambia. Trans R Soc Trop Med Hyg
37. Golovaty I, Jones L, Gelaye B, Tilahun M, Belete H et al. (2009). Access to water source, latrine facilities and other risk factors of active trachoma in Ankober, Ethiopia. 4(8):e6702.
38. Bejiga A, Alemayehu W (2001). Prevalence of trachoma and its determinants in Dalocha District, Central Ethiopia. Ophthalmic Epidemiol 8:119-125.

39. Bailey R, Downes B, Downes R, Mabey D (1991). Trachoma and water use: A case-control study in a Gambian village. Trans R Soc Trop Med Hyg. 85:824-8.

40. Bailey R, Duong T, Carpenter R, Whittle H, Mabey D (1999). The duration of human ocular *Chlamydia trachomatis* is age dependent. Epidemiol Infect 123(3):479-86.

41. Gambhir M, Basanex MG, Burton MJ, Solomon AW, Bailey RL et al. (2009). The development of an age-structured model for trachoma transmission dynamics, pathogenesis and control. PLoS Negl Trop Dis 3(6):e462.

42. Burton MJ, Hu VH, Massae P, Burr SE, Chevallier C et al. (2011). What is causing active trachoma? The role of non-Chlamydial pathogens in a low prevalence setting. Invest Ophthalmol Vis Sci. 52(8):6012-7.

43. Burr SE, Hart JD, Edwards T, Baldeh I, Bojang E et al. (2013). Association between ocular bacterial carriage and follicular trachoma following mass azithromycin distribution in The Gambia. Plos Negl Trop Dis 7(7):e2347.

44. O'loughlin R, Fentie G, Flannery B, Emerson PM. (2006). Follow-up of a low-cost latrine promotion programme in one district of Amhara, Ethiopia: characteristics of early adopters and non-adopters. Trop Med Int Health 11(9):1406-15.

45. Brechner RJ, West S, Lynch M (1992). Trachoma and flies. Individual vs environmental risk factors. Arch Ophthalmol 110:687-9.

46. Emerson PM, Bailey RL, Walraven GE, Lindsay SW (2001). Human and other faeces as breeding media of the trachoma vector Musca sorbens. Med Vet Entomol 15(3):314-20.

47. Emerson PM, Simms VM, Makalo P, Bailey RL (2005). Household pit latrines as a potential source of the fly Musca sorbens – a one year longitudinal study from The Gambia. Trop Med Int Health. 10(7):706-9.
48. Miller K, Pakpour N, Yi E, Melese M, Alemayehu W et al. (2004). Pesky trachoma suspect finally caught. Br J Ophthalmol 88(6):750-1.
49. Taye A, Alemayehu W, Melese M, Gayid A, Mekonnen Y et al. (2007). Seasonal and altitudinal variations in fly density and their association with the occurrence of trachoma in the Gurage zone of central Ethiopia. Ann Trop Med Parasitol 101(5):441-8.
50. Da Cruz L, Dadour IR, McAllister IL, Jackson A, Isaacs T. (2002). Seasonal variation in trachoma and bush flies in north-western Australian Aboriginal communities. Clin Experiment Ophthalmol 30(2):80-3.
51. Steyerberg EW, Eijkemans MJ, Habberna JD (1999). Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol 2(10):935-42.

Table 1. Study Population Characteristics

	N[b] (%)
Total rural population[a]	5,613
Study population	1,511
Female	869 (57.5)
Children <10 years	618 (40.9)
Bijagos ethnic group	1395 (92.3)
Islands	4
Villages	38
Households	293
Median number of members/household (IQR)	9 (24)
Median number of children < 10 years/household (IQR)	4 (12)
Access to water source all year	1475 (97.6)
Distance to water source < 15 minutes' walking time	1511 (100)
Access to a latrine	608 (40.4)
Access to private latrine	284 (18.8)

[a]Total rural population on Bubaque, Canhabaque, Soga and Rubane [4]

[b]N = total number and % = proportion. IQR = Interquartile range.

Table 2. The relationship between trachoma and presence of infection with *C. trachomatis* (*Ct*)

Age Group (N[a])	Clinical Status[b]	N[a](%)	*Ct* +[c] (%)	*Ct* –[c] (%)
All ages (1508)	No trachoma	999 (66.2)	131 (13.1)	868 (86.9)
	Active trachoma (TF and/or TI)	167 (11.1)	103 (61.7)	64 (38.3)
	TF	152 (10.1)	97 (63.8)	55 (36.2)
	TI	29 (2.0)	24 (82.8)	5 (17.2)
	TS	357 (23.7)	51 (14.3)	306 (85.7)
0-5 years (416)	No trachoma	298 (72.0)	39 (13.1)	259 (86.9)
	Active trachoma (TF and/or TI)	113 (27.3)	69 (61.1)	44 (38.9)
	TF	104 (25.1)	66 (63.5)	38 (36.5)
	TI	19 (4.6)	16 (84.2)	3 (15.8)
	TS	11 (2.7)	2 (18.2)	9 (81.8)
6-10 years (250)	No trachoma	220 (88.0)	44 (20.0)	176 (80.0)
	Active trachoma (TF and/or TI)	25 (10.0)	20 (80.0)	5 (20.0)
	TF	25 (10.0)	20 (80.0)	5 (20.0)
	TI	2 (0.8)	2 (100)	0 (0)
	TS	7 (2.8)	5 (71.4)	2 (28.6)

[a]N = total number and % = proportion.

[b]Using the WHO Simplified Grading System where TF=trachomatous inflammation-follicular, TI=trachomatous inflammation-intense, TS=trachomatous scarring [23]. Individuals may appear in more than one clinical category.

[c]*Ct* DNA detected (*Ct*+) or absent (*Ct*-) by Amplicor PCR.

Table 3. Multilevel univariable random effects logistic regression analysis of factors associated with active trachoma[a]

Variable	N (%)	cOR (95% CI)[b]	p-value[c]
Individual			
Age Group	1504		***<0.0001***
0-5 years	416	24.02 (12.92-44.65)	
6-10 years	250	6.26 (3.11-12.62)	
11-15 years	157	3.66 (1.60-8.39)	
>15 years	681	1.00 (baseline)	
Gender			
Female	854	1.00 (baseline)	***0.001***
Male	632	1.83 (1.28-2.61)	
Presence of *C. trachomatis* infection by PCR	258 (17.2)	11.19 (7.64-16.40)	***<0.0001***
Presence of ocular discharge	84 (5.6)	8.05 (4.62-14.02)	***<0.0001***
Presence of nasal discharge	261 (17.4)	7.90 (5.14-12.16)	***<0.0001***
Presence of flies on the face	21 (2.6)	2.188 (0.64-7.48)	*0.212*
Household			
Gender of household head (female)	509 (35.1)	0.99 (0.64-1.53)	*0.975*
Household size[d]		1.04 (0.99-1.09)	*0.068*
Number of children <10 years in household[d]		1.13 (1.02-1.24)	*0.019*
Socio-economic status			
Access to radio	747 (49.6)	0.94 (0.62-1.44)	*0.786*
Access to transport	227 (15.1)	1.20 (0.67-2.12)	*0.541*
Access to mobile phone	635 (42.2)	0.69 (0.45-1.06)	*0.090*
Relatives living in Bissau or abroad	1069 (71.0)	0.68 (0.43-1.07)	*0.090*
Access to savings at the end of the month	79 (5.3)	0.78 (0.28-2.13)	*0.623*
Educational level of household head			*0.298*
None	883 (58.7)	1.00 (baseline)	
Primary	466 (31.0)	1.31 (0.83-2.07)	
Secondary and above	156 (10.2)	0.76 (0.36-1.60)	
School attendance in children of school age	254 (62.4)	0.48 (0.30-0.76)	***0.002***
Health education received[e]	1054 (70.0)	0.75 (0.48-1.17)	*0.201*
Water access			
Access to water all year	1475 (97.5)	1.08 (0.23-4.99)	*0.923*
Access to water			***0.0002***
Multiple water sources	308 (20.5)	1.00 (baseline)	
Single water source other than natural spring	913 (60.7)	1.44 (0.87-2.54)	
Traditional natural spring only	284 (18.9)	3.37 (1.78-6.37)	
Water use within the household[e,f]			*0.359*
<1 'vasilha'	67 (4.5)	2.29 (0.40-12.93)	
1 'vasilha'	145 (10.0)	3.59 (0.85-15.13)	
>1 'vasilha'	1283 (85.8)	2.91 (0.79-10.78)	
Latrine access			
Access to latrine	608 (40.4)	1.01(0.66-1.55)	*0.951*
Access to private latrine (vs shared)	284 (18.8)	0.98 (0.75-1.29)	*0.903*
Latrine use[e]			*0.450*
Always	500 (85.3)	0.90 (0.57-1.42)	
Sometimes	51 (8.7)	1.00 (0.32-3.20)	
Never	35 (6.0)	2.53 (0.75-8.54)	
Latrine cleanliness[g]			*0.471*
Very Clean – 1	33 (6.0)	0.24 (0.03-2.30)	
2	80 (14.6)	0.49 (0.16-1.50)	
3	274 (50.0)	1.09 (0.63-1.89)	
4	64 (11.7)	1.24 (0.46-3.32)	
Very Dirty - 5	97 (17.7)	1.48 (0.68-3.25)	
Flies (≥20) present around latrine[g]	251 (16.7)	1.26 (0.74-2.1)	*0.391*
Faeces visible in latrine[g]	200 (13.3)	0.89 (0.48-1.65)	*0.703*

Environment[h]			
Flies (≥20) present	510 (32.6)	1.08 (0.69-1.69)	*0.726*
Faecal waste (human/animal)	1105 (73.4)	1.07 (0.66-1.73)	*0.797*
Domestic waste	1175 (78.1)	0.97 (0.58-1.65)	*0.922*
Animals present	1298 (86.3)	0.66 (0.38-1.17)	*0.159*

[a]Active trachoma defined as TF (inflammatory trachoma-follicular) and/or TI (inflammatory trachoma-intense) using the WHO simplified scoring system [23] [b]Unadjusted (crude) Odds Ratio (cOR) from two-level univariable mixed effects logistic regression analyses; CI=confidence interval [c]p-value for Wald test (Wald's Chi^2); significant associations (where $p \leq 0.05$) are highlighted in bold [d]Continuous numeric variables [e]reported by household head [f]a 'vasilha' is a vessel of capacity ~ 30 litres [g]researcher observed [h]researcher observed within 15m of the household

Table 4. Multilevel multivariable random effects logistic regression analysis of factors independently associated with active trachoma[a]

Variable	n	Model 1[b] aOR(95%CI)[d]	p-value[e]	Model 2[c] (Final Model) aOR (95% CI)[d]	p-value[e]
Individual					
Age Group	1504		***<0.0001***		***<0.0001***
0-5 years	416	13.66 (7.06-26.4)		10.22 (5.13-20.40)	
6-10 years	250	3.60 (1.70-7.59)		1.93 (0.88-4.24)	
11-15 years	157	3.08 (1.32-7.21)		2.41 (0.93-6.22)	
>15 years	681	1.00 (baseline)		1.00 (baseline)	
Gender					
Female	854	1.00 (baseline)	***0.034***	1.00 (baseline)	***0.005***
Male	632	1.58 (1.03-2.41)		1.89 (1.22-2.94)	
Presence of *C. trachomatis* infection	258	NA	NA	11.18 (6.9-18.1)	***<0.0001***
Presence of ocular discharge	84	2.71 (1.39-5.29)	***0.003***	2.04 (1.04-3.99)	***0.036***
Presence of nasal discharge	261	2.26 (1.35-3.77)	***0.002***	2.54 (1.51-4.26)	***0.001***
Household					
Access to water			***0.005***		***0.036***
Multiple water sources	308	1.00 (baseline)		1.00 (baseline)	
Single water source other than natural spring	913	1.42 (0.70-2.87)		0.90 (0.46-1.75)	
Traditional natural spring only	284	3.34 (1.50-7.45)		1.86 (0.89-3.89)	

[a]Active trachoma defined as TF (inflammatory trachoma-follicular) and/or TI (inflammatory trachoma-intense) using the WHO simplified scoring system [23]

[b]Model 1 shows the association of predictor variables with active trachoma.

[c]Model 2 includes the presence of *C. trachomatis* infection and demonstrates the effect of its inclusion on the predictor variables. Some of the association in Model 1 is mediated by *C. trachomatis* infection.

[d]Adjusted Odds Ratio (aOR) using multivariable two-level mixed effects logistic regression modelling; CI=confidence interval

[e]p-value for Wald test (Wald's Chi^2); significant associations (where $p \leq 0.05$) are highlighted in bold

Table 5. Multilevel univariable random effects logistic regression analysis of factors associated with ocular *C. trachomatis* infection

Variable		n	cOR (95% CI)[a]	*p-value*[b]
Individual				
Ethnicity		1497		***0.004***
	Bijagos	1395	5.05 (1.68-15.17)	
	Other	102	1.00 (baseline)	
Age Group		1504		***<0.0001***
	0-5 years	416	3.28 (2.24-4.80)	
	6-10 years	250	4.03 (2.59-6.29)	
	11-15 years	157	1.58 (0.90-2.78)	
	>15 years	681	1.00 (baseline)	
Gender				
	Female	854	1.00 (baseline)	*0.753*
	Male	632	1.05 (0.77-1.44)	
Presence of ocular discharge		84 (5.6)	4.17 (2.35-7.38)	***<0.0001***
Presence of nasal discharge		261 (17.4)	2.17 (1.48-3.20)	***<0.0001***
Presence of flies on the face		21 (2.6)	2.00 (0.57-7.10)	*0.282*
Household				
Gender of household head (female)		509 (35.1)	0.79 (0.48-1.29)	*0.343*
Household size[c]			1.04 (0.99-1.10)	*0.127*
Number of children <10 years in household[c]			1.09 (0.97-1.22)	*0.132*
Socio-economic status				
Access to radio		747 (49.6)	0.98 (0.61-1.58)	*0.945*
Access to transport		227 (15.1)	1.14 (0.58-2.24)	*0.699*
Access to mobile phone		635 (42.2)	0.57 (0.35-0.92)	*0.022*
Relatives living in Bissau or abroad		1069 (71.0)	0.88 (0.52-1.58)	*0.643*
Access to savings at the end of the month		79 (5.3)	0.69 (0.22-2.16)	*0.527*
Educational level of household head				*0.200*
	None	883 (58.7)	1.00 (baseline)	
	Primary	466 (31.0)	1.26 (0.75-2.12)	
	Secondary and above	156 (10.2)	0.56 (0.24-1.30)	
School attendance in children of school age		254 (62.4)	0.82 (0.56-1.19)	*0.289*
Health education received[d]		1054 (70.0)	0.78 (0.47-1.30)	*0.346*
Water access				
Access to water all year		1475 (97.5)	0.75 (0.15-3.91)	*0.734*
Access to water				***<0.0001***
	Multiple water sources	308 (20.5)	1.00 (baseline)	
	Single water source other than natural spring	913 (60.7)	3.69 (1.85-7.38)	
	Access to natural spring only	284 (18.9)	7.01 (3.11-15.81)	
Water use within the household[d,e]				*0.776*
	<1 'vasilha'	67 (4.5)	2.21 (0.47-10.43)	
	1 'vasilha'	145 (10.0)	1.41 (0.39-5.02)	
	>1 'vasilha'	1283 (85.8)	1.56 (0.54-4.53)	
Latrine access				
Access to latrine		608 (40.4)	1.21 (0.75-1.97)	*0.435*
Access to latrine (private vs shared)		284 (18.8)	1.03 (0.76-1.39)	*0.840*
Latrine use[d]				*0.321*
	Always	500 (85.3)	1.10 (0.67-1.80)	
	Sometimes	51 (8.7)	0.66 (0.17-2.58)	
	Never	35 (6.0)	3.73 (0.84-16.6)	
Latrine cleanliness[f]				*0.228*
	Very Clean – 1	33 (6.0)	0.55 (0.09-3.32)	
	2	80 (14.6)	0.58 (0.19-1.76)	
	3	274 (50.0)	1.08 (0.59-1.98)	
	4	64 (11.7)	1.11 (0.35-3.48)	
	Very Dirty - 5	97 (17.7)	2.63 (1.13-6.12)	

Flies (≥20) present around latrine[f]	251 (16.7)	1.90 (1.05-3.41)	***0.033***
Faeces visible in latrine[f]	200 (13.3)	1.15 (0.59-2.24)	*0.687*
Environment[g]			
Flies (≥20) present	510 (32.6)	1.37 (0.84-2.22)	*0.209*
Faecal waste (human/animal)	1105 (73.4)	0.90 (0.53-1.54)	*0.705*
Domestic waste	1175 (78.1)	0.65 (0.36-1.17)	*0.150*
Animals present	1298 (86.3)	0.49 (0.26-0.94)	***0.030***

[a]Unadjusted (crude) Odds Ratio (cOR) using two-level univariable mixed effects logistic regression; CI=confidence interval [b]p-value for Wald test (Wald's Chi^2); significant associations (where $p \leq 0.05$) are highlighted in bold [c]Continuous numeric variables [d]reported by household head [e]a 'vasilha' is a vessel of capacity ~ 30 litres [f]researcher observed [g]researcher observed within 15m of the household

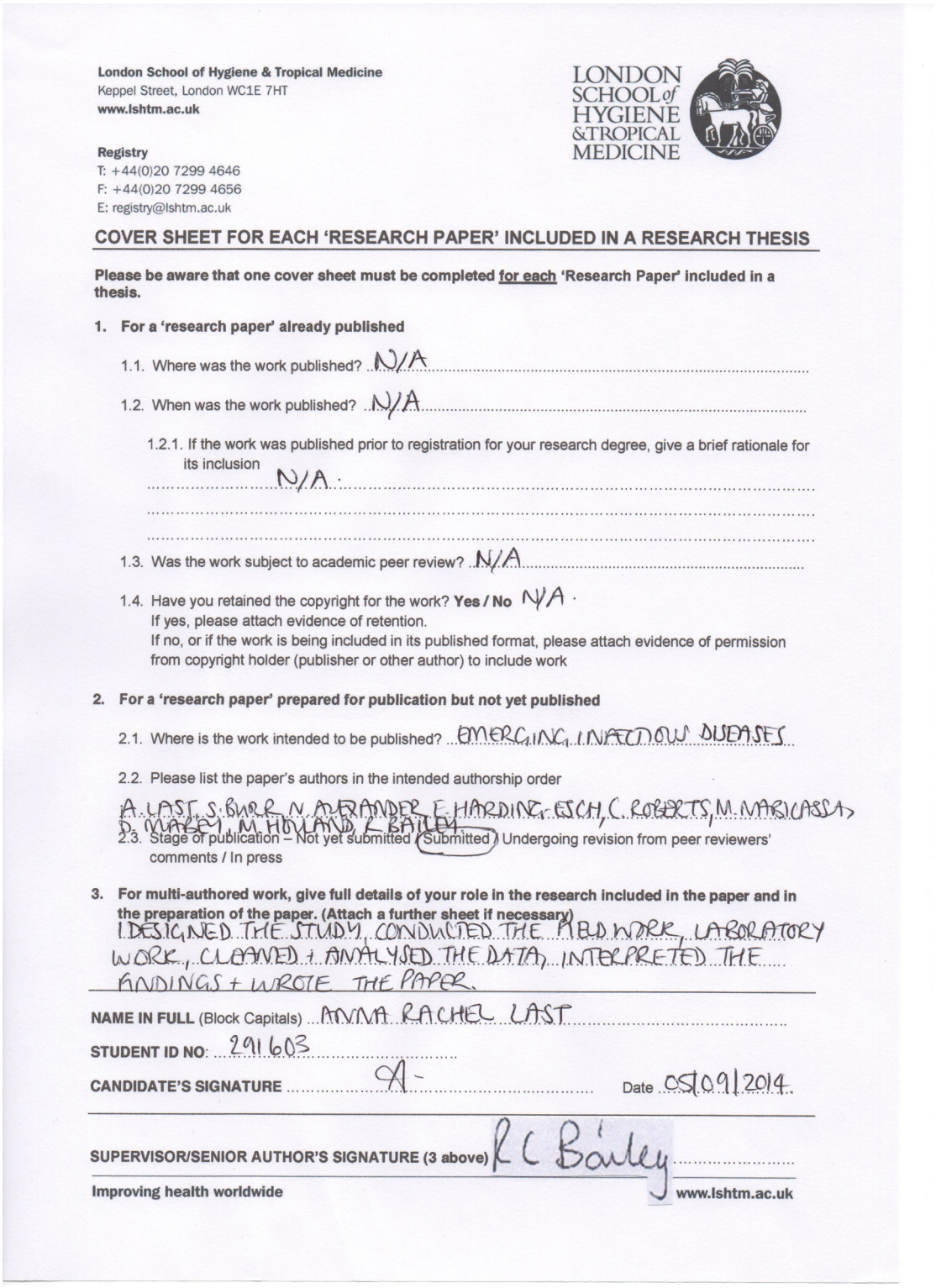
Table 6. Multilevel multivariable random effects logistic regression analysis of factors associated with ocular *C. trachomatis* infection

Variable		n	aOR (95% CI)[a]	p-value[b]
Individual				
Age Group		1504		***<0.0001***
	0-5 years	416	3.10 (1.04-4.70)	
	6-10 years	250	3.83 (2.38-6.16)	
	11-15 years	157	1.65 (0.92-2.96)	
	>15 years	681	1.00 (baseline)	
Gender				*0.377*
	Female	854	1.00 (baseline)	
	Male	632	0.86 (0.61-1.21)	
Presence of ocular discharge		84	2.33 (1.25-4.35)	***0.007***
Household				
Access to water				***<0.0001***
	Multiple water sources	308	1.00 (baseline)	
	Single water source other than natural spring	913	3.88 (1.88-8.01)	
	Traditional natural spring only	284	6.57 (2.83-15.23)	
Flies (≥20) present around latrine[c]		251	2.06 (1.10-3.84)	***0.023***

[a]Adjusted Odds Ratio (aOR) using two-level multivariable mixed effects logistic regression modelling; CI=confidence interval [b]p-value for Wald test (Wald's Chi^2); significant associations (where $p \leq 0.05$) are highlighted in bold [c]researcher observed

Chapter 8

Spatial clustering of high load ocular *Chlamydia trachomatis* infection and disease severity in trachoma-endemic communities on the Bijagós Archipelago, Guinea Bissau



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. **For a 'research paper' already published**

   1.1. Where was the work published? N/A

   1.2. When was the work published? N/A

   1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
   N/A.

   1.3. Was the work subject to academic peer review? N/A

   1.4. Have you retained the copyright for the work? **Yes / No** N/A.
   If yes, please attach evidence of retention.
   If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. **For a 'research paper' prepared for publication but not yet published**

   2.1. Where is the work intended to be published? EMERGING INFECTIOUS DISEASES

   2.2. Please list the paper's authors in the intended authorship order
   A. LAST, S. BURR, N. ALEXANDER, E. HARDING-ESCH, C. ROBERTS, M. MARCIASSA, D. MABEY, M. HOLLAND, R. BAILEY

   2.3. Stage of publication – Not yet submitted / (Submitted) / Undergoing revision from peer reviewers' comments / In press

3. **For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)**
   I DESIGNED THE STUDY, CONDUCTED THE FIELD WORK, LABORATORY WORK, CLEANED + ANALYSED THE DATA, INTERPRETED THE FINDINGS + WROTE THE PAPER.

NAME IN FULL (Block Capitals) ANNA RACHEL LAST

STUDENT ID NO: 291603

CANDIDATE'S SIGNATURE A- Date 05/09/2014

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) R C Bailey

Improving health worldwide **www.lshtm.ac.uk**

Title

Spatial clustering of high load ocular *Chlamydia trachomatis* infection and disease severity in trachoma-endemic communities on the Bijagós Archipelago, Guinea Bissau.

Authors and Affiliations

Anna Last[1#], Sarah Burr[2], Neal Alexander[3], Emma Harding-Esch[1], Chrissy h. Roberts[1], Meno Nabicassa[4], David Mabey[1], Martin Holland[1], Robin Bailey[1]

[1]Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, [2]Disease Control and Elimination Theme, Medical Research Council Unit The Gambia, P.O. Box 273 Banjul, Atlantic Boulevard, Fajara, The Gambia, [3]MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, [4]Programa Nacional de Saúde de Visão, Ministério de Saúde Publica, P.O. Box 50, Avenida de Unidade Africana, Bisssau, Guiné Bissau

Running Title

Spatial clustering of ocular *Chlamydia trachomatis*

Summary

Spatial clustering of high load ocular *Chlamydia trachomatis* infections and a strong association with severe inflammatory trachoma suggests the importance of chlamydial load in disease pathogenesis and transmission.

#CORRESPONDING AUTHOR Mailing Address: Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Telephone: +44 (0)207 927 2566 Email: anna.last@lshtm.ac.uk

Abstract

Trachoma is caused by ocular infection with *Chlamydia trachomatis*. A population-based geospatially representative survey conducted in a trachoma-hyperendemic treatment-naive population demonstrated that the highest *C. trachomatis* loads were associated with the most severe clinical signs of disease and that the strongest associations were with increasing inflammatory scores. Spatial clusters and hotspots, independent of age and gender, of individual infections with high bacterial loads were detected using local indicators of spatial association. There was no evidence of coldspots or clustering of low load infections, suggesting that high load infections may be important in transmission of chlamydial infection. These geospatial tools may be useful in trachoma surveillance in post-treatment settings to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission and may have wider application in monitoring concurrent neglected tropical diseases that are targeted for elimination.

Keywords

Chlamydia trachomatis; bacterial load; disease severity; spatial clustering; local indicators of spatial association; trachoma

Introduction

Chlamydia trachomatis is the leading infectious cause of blindness globally and the most common sexually transmitted bacterium. Trachoma is caused by infection with ocular strains of *C. trachomatis* and manifests as distinct clinical syndromes, beginning with an acute self-limiting kerato-conjunctivitis which may progress to chronic inflammatory disease with subsequent conjunctival scarring and blinding sequelae. A spectrum of disease severity exists, which can be graded objectively using validated grading systems (1,2).

The World Health Organization (WHO) advocates the implementation of the SAFE strategy (Surgery for trichiasis, Antibiotics for active infection, Facial cleanliness to prevent disease transmission and Environmental improvement to increase access to water and sanitation) for trachoma elimination. Mass Drug Administration (MDA) of azithromycin aims to clear infection from communities such that trachoma ceases to be a public health concern (3).

Understanding the spatial distribution of disease and infection in trachoma-endemic regions is increasingly recognized in national trachoma control programme planning, enabling the identification of at risk populations and prioritization of target areas for control and implementation and optimal scaling of SAFE (4-8). It may also be important in understanding transmission, transmission thresholds and the dynamics of spread and recovery from infection following intervention.

Spatial analysis offers powerful methods to study the relationship between active trachoma and ocular infection with *C. trachomatis*. The spatial relationship between disease and infection defined over space is termed spatial dependence. This is measured as the existence of statistical association (dependence) between disease and infection associated with location. To explore spatial dependence measures of spatial autocorrelation are used. Spatial autocorrelation assumes that near objects are more related than distant objects. In the context of infectious diseases, this concept may be applied to explore the relationship between infection and disease where the spatial distribution can be used to define underlying processes such as exposure or transmission. Spatial regression models can thus be used to understand the epidemiology of active trachoma and *C. trachomatis* infection.

Global spatial autocorrelation statistics are used to describe the overall spatial patterns in the data. Local indicators of spatial association take actual values of

observations in the context of adjacent values allowing mapping and definition of clusters at exact locations. This allows for the examination of small-scale heterogeneity and identification of statistically significant clusters and outliers of similar observations in space. In the context of global spatial patterns described above, this provides us with a greater understanding of the relationship between infection and disease.

In this study we apply geostatistical tools to investigate the fine-scale epidemiology of *C. trachomatis* infection and clinical signs of active trachoma in a population-based study conducted in a trachoma-hyperendemic treatment-naïve population from the Bijagós Archipelago of Guinea Bissau in West Africa. Although some risk factors have been identified in this population, the reasons why both active and scarring trachoma are highly prevalent and severe on these islands remain to be elucidated (9). Transmission dynamics, pathogen virulence and *C. trachomatis* load may be contributing factors. This is the first study to use local indicators of spatial association using individual-level *C. trachomatis* load data in the context of spatial dependence to investigate the relationship between infection and disease. This may be important in understanding disease pathogenesis and transmission dynamics, both of which are fundamental to successful trachoma elimination and surveillance strategies.

Methods

Ethical Statement

This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained from the Comitê Nacional de Ética e Saúde (Guinea Bissau), the LSHTM Ethics Committee (UK) and The Gambia Government/MRC Joint Ethics Committee (The Gambia). Written (thumbprint or signature) informed consent was obtained from all study participants or their guardians as appropriate.

After survey completion all communities on the study islands were treated with a single height-based dose of oral azithromycin in accordance with WHO and national protocols.

Study Area

The Bijagós Archipelago has a total area of more than 10,000 km^2 and lies between N $11^0$38'19.68" and N $10^0$51'40.32", W $16^0$29'38.40" and W $15^0$27'17.28". The surface area of exposed land covers approximately 900 km^2, of which 350 km^2 is covered with mangrove forest (10). Altitude does not exceed 50m. The climate is humid and tropical, with a prolonged rainy season from May-November, during which average monthly rainfall is 400mm^3 (11). The mean monthly temperature is 27.3^{0}C (25.1-29.2^{0}C), with peak temperatures occurring prior to the rainy season. There are 88 islands and islets of which only about 20 are permanently inhabited. The remainder are inhabited periodically for seasonal agriculture and traditional initiation ceremonies. The study was conducted on four islands of the archipelago (*Figure 1*). These four islands comprise a total rural population of 5,613 (National Population Census, 2010, Instituto Nacional de Estatística, Guiné-Bissau).

Study Design and Study Population

Trachoma survey methodology and this study population have been described previously (9,12-15). To satisfy adequate geospatial representation at village-level, we included all 38 villages on the four study islands and randomly sampled one in five households (with a minimum of five per village) from each. All were sampled if there were fewer than six households in the village. Small villages are thus over-represented by the minimum sampling criteria imposed. Data were geo-coded at household and village level using the Garmin eTrex H handheld Global Positioning Systems (GPS) unit (Garmin Ltd., UK).

Clinical Examination and Conjunctival Sampling

A single validated examiner assessed each participant using the WHO simplified and modified FPC grading systems (1,2). In the modified FPC system, follicles (F), papillae (P) and conjunctival scarring (C) are each assigned a separate grade from 0-3. FPC grades of F2/3 or P3 equate to a diagnosis of active trachoma (TF (follicular trachoma) or TI (inflammatory trachoma) by the WHO simplified system) and a grade of C2/3 (and in some cases C1) equates to a diagnosis of TS (trachomatous scarring). A trachoma grade was assigned to the upper tarsal conjunctivae of each consenting participant using adequate light and a 2.5x binocular magnifying loupe. Both methods were used in order that study data should be comparable to data used by trachoma control programmes and research studies requiring detailed information related to disease severity and have been used previously in similar settings (16,17).

Samples were taken from the left upper tarsal conjunctiva of each participant with Dacron swabs (Fisher Scientific, UK) using a well-tolerated standardized procedure described in previous studies (9,12,13,18,19).

Detection and Quantitation of C. trachomatis

DNA extraction and droplet digital PCR (ddPCR) (Bio-Rad Laboratories, Hemel Hempstead, UK) were conducted as described previously (9,12,13). We used *C. trachomatis* plasmid-based ddPCR to diagnose infection and a single-copy pathogen chromosomal gene (*omcB*) to estimate pathogen load in each plasmid-positive sample (12,13). Estimated quantities of *omcB* (*C. trachomatis* load) and plasmid are expressed as copies/swab.

Statistical Analysis

C. trachomatis quantitation data were processed as described previously (12,13). GPS data were downloaded into MapSource v16.16.3 (Garmin Ltd., UK). All data were double entered into a customised database (Microsoft Access 2007) and discrepancies resolved through reference to source documents. Data were cleaned and analysed in STATA 13 (Stata Corporation, College Station, Texas USA). Statistical significance was determined at the 5% level.

Mixed effects linear regression models of C. trachomatis load and Clinical Phenotype

C. trachomatis load data were log(e) transformed where indicated. The geometric mean of load, respective standard error (SE) and 95% confidence intervals (CI) were calculated. An analysis of variance (ANOVA) with pair-wise comparisons was used to compare load across detailed clinical phenotypes. Assessment of group differences and multiple comparisons were adjusted for using the Scheffé correction (20). Associations between load and detailed clinical phenotype were examined using univariable and multivariable mixed effects linear and logistic regression models accounting for household-level clustering detected in previous studies (9).

Geostatistical Analyses

Geocoded data were projected into UTM Zone 28N. ArcGIS 10.1 (ESRI Inc., USA) and the R statistical package v3.0.2 (The R Foundation for Statistical Computing, http://www.r-project.org using spdep, automap and nlme packages) were used for all geostatistical analyses (21-23). In the following analyses, the zone of indifference is used to define adjacency. This method assumes that each observation has local influence that decreases with distance beyond a critical distance cut-off, resulting in an adapted model of impedance, or distance decay, such that all features have an impact on all other features, but this impact decreases with distance. This

method is appropriate for point data and takes into account the extent of spatial autocorrelation in selecting data-driven threshold cut-offs (24).

Spatial Autocorrelation

The Moran's I statistic was used to evaluate global spatial autocorrelation in the distribution of the household-level prevalence of disease and infection. The values of Moran's I range between -1 and +1. A value close to -1 indicates negative autocorrelation (complete dispersion) whilst a value close to +1 indicates positive autocorrelation (clustering) of features. A value close to 0 suggests random arrangement. Z-scores and p-values are assigned to ascertain whether spatial autocorrelation is statistically significant (25). Since the Moran's I statistic can be sensitive to skewed distributions, a permutation test was used to verify the results. This uses Monte Carlo simulations to generate a Moran's I sampling distribution (26).

Empirical semivariograms were constructed to obtain the distance range over which spatial autocorrelation occurs using the average squared distance between paired data values against the distance (or lag) separating the pairs to estimate the spatial covariance structure of the data (27).

Linear and logistic mixed effects regression models were also applied to examine the covariance structure and compare models with and without a spatial component. The log likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare models (28).

Local Indicators of Spatial Association: Hotspot and Cluster-Outlier Analyses

The Getis-Ord Gi* statistic was used to identify hot and cold spots for infection (29). The Getis-Ord Gi* identifies statistically significant 'hotspots' containing similarly high values and statistically significant 'cold spots' containing similarly low values, taking into account the spatial dependence of adjacent

observations using the zone of indifference described previously (30). Z-scores and p-values are expressed for each Gi* value in comparison with the normal distribution of Gi* calculated by simulation. These values represent the statistical significance of the spatial clustering of values given the conceptualisation of spatial relationships and scale of analysis described. Statistically significant high z-scores show clustering of high load infections, whilst statistically significant low z-scores show clustering of low (or zero) load infections.

Clustering and outlier analysis was conducted using the local (Anselin) Moran's I statistic using the zone of indifference to define adjacency (31). A positive value for I indicates that a feature has adjacent features with similarly high or low attributable values (a cluster). A negative value for I indicates that adjacent features have dissimilar values and that this feature is an outlier. Cluster-Outlier analysis identifies spatial clusters of features with high or low values and spatial outliers by calculating Moran's I, Z-scores, p-values and a code to represent the cluster type. Cluster types are defined as clusters of high values (HH, High-High), clusters of low values (defined as above) (LL, Low-Low), a high value outlier surrounded by predominantly low values (HL, High-Low) and a low value outlier surrounded by predominantly high values (LH, Low-High).

In this analysis 'low' values include both low values related to low load infections and null (zero) values related to uninfected individuals. To explore this, these analyses were conducted first on the whole data set, including uninfected individuals with a 'zero' value for infectious load, and then on infected individuals only.

Results

Trachoma and *C. trachomatis* infection prevalence is hyperendemic in this population and is described elsewhere in detail (9,12,13). *C. trachomatis* load was skewed, with the majority of cases having low copy numbers (<1000 copies/swab). Log-(e) transformation removed the skew (skewness -0.106, *p=0.5454*) in these analyses.

C. trachomatis bacterial load and clinical phenotype in individuals with C. trachomatis infection

The geometric mean of estimated *omcB* copies/swab present in clinically normal conjunctivae (F0/P0/C0) was 294 copies/swab (95% C.I. 165-524). In clinically active trachoma it was 8562 copies/swab (95% C.I. 5412-13546). Significantly higher loads were detected in individuals with increasing F and P scores (*Table 1*). *C. trachomatis* load by age and clinical phenotype is shown in *Figure 2*. The majority of infections with high loads were in children under 10 years of age with active trachoma.

Spatial structure of active trachoma and C. trachomatis infection

Significant positive spatial autocorrelation was evident for *C. trachomatis* infection (Moran's I=0.19, *p<0.0001*) but not active trachoma (Moran's I=0.07, *p=0.0659*). Semi-variograms demonstrate that autocorrelation in infection is negligible in distances greater than 1719m (*Figure 3*). We found no evidence of spatial autocorrelation in the distribution of *C. trachomatis* load (Moran's I=0.05, *p=0.4464*).

C. trachomatis load was the strongest predictor of clinically active trachoma. In accordance with the semi-variograms, including spatial structure in multivariable

mixed effects regression analyses for active trachoma and *C. trachomatis* infection improves the fit of the models (*Table 2*).

Inflammatory and follicle scores were strongly associated with bacterial load. Inclusion of the spatial structure in models predicting *C. trachomatis* load improved the fit, but the effect of spatial dependence becomes undetectable when age and disease severity scores are included, thus resolving residual spatial variation (*Table 3*).

Hotspot analysis (Getis Ord Gi) of C. trachomatis infection (Figure 4)*

Intense clustering (statistically significant high z-scores), or hotspots, of high load infections was identified. There was no evidence of coldspots, or intense clustering of low load infections or zero-value uninfected individuals. There was no apparent difference between analyses conducted on the whole data set (infected and non-infected individuals) versus those conducted on infected individuals only.

Cluster and outlier analysis (local (Anselin) Moran's I) of C. trachomatis infection (Figure 5)

High load infections were clustered with other high load infections (HH clusters). Outliers where there was a single low load infection amidst predominantly high load infections (HL) were also demonstrated. There was no clustering of low load infections (or zero-values from uninfected individuals) (LL clusters) and there were no statistically significant outlying high loads surrounded by predominantly low loads (LH). Analyses were conducted on the whole data set and infections only as described above.

There were no individuals with infections below 10,000 *omcB* copies/swab noted within a HH cluster or as a HL outlier.

Discussion

The majority of ocular *C. trachomatis* infections occur in children, who have the highest loads and most severe active trachoma. However, infection predominantly occurs at low bacterial loads in the population overall. Similar findings have been observed in other studies (17,32,33), though we found a greater prevalence of infection across all age groups typical of hyperendemic settings. Almost half the individuals with quantifiable infection had a normal clinical phenotype, though the mean *C. trachomatis* load was significantly lower than in those with active trachoma, which is consistent with other studies in meso and hyperendemic settings (32,34).

In this population *C. trachomatis* load increases with disease severity (for both follicular (F) and inflammatory (P) scores), the strongest association being with increasing P-scores. The association between ocular *C. trachomatis* load and disease severity in trachoma is supported by findings from other studies (16,17,32,35). The association between *C. trachomatis* load and high F-scores is in part due to collinearity between F and P, where with high F-scores there is likely to be inflammation present. Inflammation has previously been found to be associated with high *C. trachomatis* loads and persistence of infection in children (36). This is consistent with the high prevalence of infection and the spectrum of load and phenotype observed in this age group but may also be associated with pathogen virulence if there are different *C. trachomatis* strains in circulation. Further analysis is underway to investigate associations between pathogen genotype, clinical phenotype and geospatial clustering of strains.

There is spatial dependence in the distribution of *C. trachomatis* infection, demonstrated by positive spatial autocorrelation. The distance over which this occurs (1719m) equates to that of village boundaries. We have shown that the highest burden

of infection (and load) is in children under 10 years of age. There are usually only one or two children of this age group within a household in these communties (9). These data are supportive of transmission occurring at village level in this population.

Although household and village-level clustering is evident in active trachoma and ocular *C. trachomatis* infection (9,37-39), the inclusion of this spatial structure in regression models for active trachoma and *C. trachomatis* infection further improves model fit and demonstrates underlying spatial processes in the relationship between infection and disease, such that cases of active trachoma may represent recent exposure to *C. trachomatis* infection. The effect of spatial dependence in infection is greater than in active trachoma, perhaps reflecting the complexity of the disease process, where host-pathogen interactions contribute.

C. trachomatis bacterial load is not globally spatially autocorrelated, likely reflecting the stronger influence of underlying social and biological rather than spatial processes. However, using local indicators of spatial association to examine fine-scale spatial clustering we have identified hotspots and clustering of high load infections. Spatial clusters, independent of age and gender, of individual infections with high bacterial loads (HH) and high load outliers that cluster with other low loads (HL) exist. Furthermore, there is no evidence of coldspots or clustering of low load infections (LL), suggesting that high load infections may be important in transmission of chlamydial infection. There appears to be a threshold of *C. trachomatis* load below which HH clustering (and being a HL outlier) does not occur. The same phenomena were observed in data including infected and uninfected individuals and data with infected individuals only suggesting that low load infection may represent a negligible risk of transmission. This supports the described Allee effect, which hypothesizes that

a reduction in chlamydial fitness due to reduced pathogen population size or density results in its disappearance from a population (40).

We did not find a threshold effect within hotspots demonstrated using Getis-Ord Gi*. This may be explained by differences in the underlying methodology, where the Getis-Ord Gi* includes the value of each feature examined in an average value in its analysis, whereas the local Moran's I statistic takes each feature but includes only adjacent features' values in the average value in its analysis. This is particularly evident where the scale of the analysis is small. The local Moran's I statistic may be more useful in this context as it provides greater detail on clustering and outliers, which may have application in disease surveillance.

There are a limited number of studies that have assessed the spatial distribution of trachoma (5,7,41) and ocular *C. trachomatis* infection (42,43). Broman *et al.* did not find clustering of active trachoma in children under the age of 10 years, but found clusters of households with high bacterial loads (defined as the household mean load) in treatment-naïve communities, supporting the findings in this study (42). Yohannan *et al*. examined binary household-level data using a K-function nearest neighbours analysis (43). The methodology for cluster detection in these studies was heterogeneous and measures of spatial dependence were not included in these analyses.

Longitudinal study is required to fully investigate the dynamics of *C. trachomatis* bacterial load in transmission. This study is limited in the assumption of load and clinical disease severity as a steady state. There are few data addressing stability of *C. trachomatis* load and disease phenotype (36). Bobo *et al.* conducted weekly surveys for three months in hyperendemic communities, finding that 62% of children had at least one infection during that time and of those 64% were persistently

infected and had higher mean *C. trachomatis* loads and more severe disease than those who were sporadically infected (36). Additionally, host conjunctival immune response and duration of infection add complexity to the interaction between *C. trachomatis* infection and disease severity (33,44).

Conclusion

This is the first study to use individual-level quantifiable *C. trachomatis* infections from a geospatially representative population-based sample to investigate spatial clustering of *C. trachomatis* infection. These data show that increasing *C. trachomatis* load is related to increasing disease severity in active trachoma, particularly with respect to inflammation. We have provided a global statistical measure for spatial autocorrelation in infection and disease and used local indicators of spatial association to describe the location and nature of the clusters in relation to *C. trachomatis* load.

These data suggest that high load *C. trachomatis* infections cluster spatially and may be important in transmission. *C. trachomatis* strain diversity may be associated with bacterial load, the spectrum of disease severity and the clustering of high load infections within these communities. Further epidemiological and *in vitro* studies are required to provide a more complete picture of the relationship between disease severity and chlamydial load.

We are currently developing mathematical models to investigate transmission dynamics of *C. trachomatis* and the effect of high load infections in these communities. Such models, in conjunction with geospatial tools may be useful as tools for post-MDA trachoma surveillance. Using these methods, it may be possible to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission and may improve our understanding of disease

pathogenesis and transmission. These methods may also have wider application in monitoring concurrent neglected tropical diseases that are targeted for elimination.

Aknowledgements

This work was supported by the Wellcome Trust (grant number 097330/Z/11/Z). We extend thanks to the study participants and field research team in Guinea Bissau.

References

(1) Dawson CR, Jones BR, Tarizzo ML. Guide to trachoma control in programs for the prevention of blindness. 1981;Geneva: World Health Organization.

(2) Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR. A simple system for the assessment of trachoma and its complicatios. Bull World Health Organ 1987;65(4):477-83.

(3) Kuper H, Solomon AW, Buchan J, Zondervan M, Foster A, Mabey D. A critical review of the SAFE strategy for the prevention of blinding trachoma. Lancet Infect Dis 2003;3:372-381.

(4) Smith JL, Sturrock HJ, Olives C, Solomon AW, Brooker SJ. Comparing the performance of cluster random sampling and integrated threshold mapping for targeting trachoma control, using computer simulation. PLoS Negl Trop Dis 2013;7(8):e2389.

(5) Smith JL, Flueckiger RL, Hooper PJ, Polack S, Cromwell EA, Palmer SL *et al*. The geographical distribution and burden of trachoma in Africa. PLoS Negl Trop Dis 2013;7(8):e2359.

(6) Smith JL, Haddad D, Polack S, Harding-Esch EM, Hooper PJ, Mabey DC *et al*. Mapping the global distribution of trachoma: why an updated atlas is needed. PLoS Negl Trop Dis 2011;5(6):e973.

(7) Clements AC, Kur LW, Gatpan G, Ngondi JM, Emerson PM, Lado M *et al*. Targeting trachoma control through risk mapping: the example of Southern Sudan. PLoS Negl Trop Dis 2010;4(8):e799.

(8) Baker MC, Mathieu E, Fleming FM, Deming M, King JD, Garba A *et al*. Mapping, monitoring and surveillance of neglected tropical diseases: towards a framework. Lancet 2010;375(9710):231-8.

(9) Last AR, Burr SE, Weiss HA, Harding-Esch EM, Cassama E, Nabicassa M *et al.* Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau. PLoS Negl Trop Dis 2014;8(6):e2900.

(10) Van der Linde HA, Dankin MH. Enhancing Sustainability: Resources for our future. SUI Technical Series Volume 1. Proceedings of a workshop held at the world conservation congress for the Sustainable Use Initiative (SUI). 1998: IUCN, The World Conservation Union, Montreal Canada.

(11) World Bank. From the Climatic Research Unit (CRU) at the University of East Anglia (UEA). 2013: Climate Change Knowledge Portal: Guinea Bissau. (http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisRegion=Africa&ThisCCode=GNB).

(12) Roberts CH, Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M *et al*. Development and evaluation of a next generation digital PCR diagnostic assay for ocular *Chlamydia trachomatis* infections. J Clin Microbiol 2013;51(7):2195-2203.

(13) Last A, Roberts CH, Cassama E, Nabicassa M, Molina-Gonzalez S, Burr SE *et al*. Plasmid copy number and disease severity in naturally occurring ocular Chlamydia trachomatis infection. J Clin Microbiol 2014;52(1):324-7.

(14) Harding-Esch E, Edwards T, Mkocha H, Muñoz B, Holland MJ, Burr SE *et al*. Trachoma prevalence and associated risk factors in The Gambia and Tanzania: Baseline results of a cluster randomized controlled trial. PLoS Negl Trop Dis 2010;4(11):e861.

(15) Edwards T, Harding-Esch EM, Hailu G, Andreason A, Mabey DC, Todd J *et al*. Risk factors for active trachoma and Chlamydia trachomatis infection in rural Ethiopia after mass treatment with azithromycin. Trop Med Int Health 2008;13:556-565.

(16) Burton MJ, Holland MJ, Faal N *et al*. Which members of a community need antibiotics to control trachoma? Conjunctival Chlamydia trachomatis infection load in Gambian villages. Invest Ophthalmol Vis Sci 2003;10(42):4215-22.

(17) Burton MJ, Holland MJ, Jeffries D, Mabey DC, Bailey RL. Conjunctival Chlamydial 16S Ribosomal RNA Expression in Trachoma: Is Chlamydial Metabolic Activity Required for Disease to Develop? Clin Infect Dis 2006;42:463-70.

(18) Keenan JD, Lakew T, Alemayehu W *et al*. Clinical activity and polymerase chain reaction evidence of chlamydial infection after repeated mass antibiotic treatments for trachoma. Am J Trop Med Hyg 2010;82(3):482-7.

(19) Stare D, Harding-Esch E, Munoz B *et al*. Design and baseline data of a randomised trial to evaluate coverage and frequency of mass treatment with azithromycin: The Partnership for the Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. Ophthalmic Epidemiol 2011;18(1):20-9.

(20) Ruxton GD, Beauchamp G. Time for some a priori thinking about post hoc testing. Behav Ecol 2008;19(3):690-3.

(21) Bivand R. Spdep: Spatial dependence: Weighting schemes, statistics and models. 2010: R package version 0.5-4. (http://CRAN.Rproject.org/package=spdep).

(22) Hiemstra PH, Pebesma EJ, Twenhofel CJW, Heuvelink GBM. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. Computers & Geosciences 2009;35(8):1711-1721.

(23) Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models.* 2014: R package version 3.1-117 (http://CRAN.R-project.org/package=nlme).

(24) Chang K. An introduction to geographic information systems (5th edition). 2010: New York:Thomas D. Timp pp327-40.

(25) Moran PA. Notes on continuous stochastic phenomena. Biometrika 1950;37(1-2):17-23.

(26) Cliff AD, Ord JK. Spatial Processes. 1981: Pion pp63-5.

(27) Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. Royal Statistics Society. Applied Statistics 1998;47(3):299-350.

(28) Zucchini W. An Introduction to Model Selection. J Mathematical Psychol 2000;44:41-61.

(29) Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geographical Analysis 1992;24(3).

(30) Ord JK, Getis A. Local spatial autocorrelation statistics: Distributional issues and an application. Geographical Analysis 1995;27(4).

(31) Anselin L. Local indicators of spatial Association: LISA. Geographical Analysis 1995;27(2):93-115.

(32) Solomon AW, Holland MJ, Burton MJ *et al*. Strategies for control of trachoma: observational study with quantitative PCR. Lancet 2003;362(9379):198-204.

(33) Faal N, Bailey RL, Jeffries D *et al.* Conjunctival FOXP3 expression in trachoma: Do regulatory T cells have a role in human ocular *Chlamydia trachomatis* infection? PLoS Med 2006;3(8):e266.

(34) West ES, Munoz B, Mkocha H *et al.* Mass treatment and the effect of the load in *Chlamydia trachomatis* infection in a trachoma-hyperendemic community. Invest Ophthalmol Vis Sci 2005;46(1):83-7.

(35) Michel CE, Roper G, Devona MA *et al*. Correlation of clinical trachoma and infection in Aboriginal communities. PLoS Negl Trop Dis 2011;5(3):e986.

(36) Bobo LD, Novak M, Munoz B *et al*. Severe disease in children with trachoma is associated with persistent Chlamydia trachomatis infection. J Infect Dis 1997;176:1524-30.

(37) Bailey R, Osmond C, Mabey DC, Whittle HC, Ward ME. Analysis of the household distribution of trachoma in a Gambian village using a Monte Carlo simulation procedure. Int J Epiemiol 1989;18(4):944-51.

(38) Blake IM, Burton MJ, Bailey RL, Solomon AW, West S, Munoz B *et al*. Estimating household and community transmission of ocular *Chlamydia trachomatis*. PLoS Negl Trop Dis 2009;3(3):e401.

(39) Hagi M, Schemann JF, Mauny F, Momo G, Sacko D, Traore L et al. Active trachoma among children in Mali: Clustering and environmental risk factors. PLoS Negl Trop Dis 2010;4(1):e583.

(40) Chidambaram JD, Lee DC, Porco TC, Lietman TM. Mass antibiotics for trachoma and the Allee effect. Lancet Inf Dis 2005;5(4):194-6.

(41) Polack SR, Solomon AW, Alexander NDE, Massae P, Safari S, Shao JF *et al.* The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma. Trans Royal Soc Med Hyg 2005;99:218-25.

(42) Broman AT, Shum K, Munoz B, Duncan DD, West SK. Spatial clustering of ocular chlamydial infection over time following treatment, among households in a village in Tanzania. Invest Ophth Vis Sci 2006;47(1):99-104.

(43) Yohannan J, He B, Wang J, Greene G, Schein Y, Mkocha H *et al.* Geospatial distribution and clustering of *Chlamydia trachomatis* in communities undergoing mass azithromycin treatment. Int Ophth Vis Sci 2014;55(7):4144-50.

(44) Gambhir M, Basanez MG, Burton MJ et al. The development of an age structured model for trachoma transmission dynamics, pathogenesis and control. PLoS Negl Trop Dis 2009;3(6):e462.

Table 1. Estimated *Chlamydia trachomatis* load (*omcB* copies/swab) and analysis of variance (ANOVA[a]) by detailed clinical phenotype in infected individuals

Clinical Phenotype	*n*[b]	*omcB* copies/swab (Geometric Mean)	95% C.I. (SE geometric mean)	*p-Value* (ANOVA)			Median	Min	Max
Normal (F0P0C0)	71	294	165, 524	*Baseline*			176	15	96333
Active Trachoma (TF and/or TI)	92	8562	5412, 13546	***p<0.00001***			14236	17	274835
Scarring Trachoma (TS)	19	928	280, 3074	*p=0.4069*			2142	17	49125
Follicular Score (F)									
0	91	438	251, 762	*Baseline*			227	15	202632
1	20	1288	448, 3697	*p=0.324*	*Baseline*		1710	34	96333
2	27	3212	1264, 8165	***p=0.002***	*p=0.624*	*Baseline*	3203	27	140693
3	46	19870	2832, 25927	***p<0.0001***	***p<0.0001***	***p=0.018***	22767	323	274835
Inflammatory Score (P)									
0	46	122	68, 218	*Baseline*			67	15	41059
1	70	1534	871, 2702	***p<0.0001***	*Baseline*		1469	16	202632
2	46	10413	5461, 19857	***p<0.0001***	***p<0.0001***	*Baseline*	18569	17	274835

	3	22	14053	5550, 35581	***p<0.0001***	***p<0.0001***	*p=0.964*	21864	34	158548
Scarring Score (C)										
	0	155	1902	1207, 2996	*Baseline*			2095	15	274835
	1	11	449	111, 1816	*p=0.438*	*Baseline*		204	34	11556
	2	9	990	192, 5111	*p=0.927*	*p=0.941*	*Baseline*	589	76	54651
	3	9	2475	253, 24192	*p=0.995*	*p=0.608*	*p=0.923*	7023	17	49125

[a]Scheffé correction used for multiple comparisons [b]*n*=number of individuals with quantifiable *C. trachomatis* bacterial load

Table 2. Multivariable mixed effects logistic regression analysis showing the effect of spatial dependence on clinically active trachoma and ocular *Chlamydia trachomatis* infection

[a]AIC=Akaike information criterion [b]With household or village as cluster variables [c]Including of spatial structure [d]defined as TF/TI using the WHO simplified grading system (2) [e]defined as log-(e) *omcB* copies/swab.

Model	Predictor Variables	*n*	*AIC*[a]
C. trachomatis infection			
No spatial[b]	age	1426	854.8
Spatial[c]			**801.8**
No spatial	age	1426	546.4
	active trachoma[d]	163	
Spatial			**495.6**
Active trachoma			
No Spatial	age	1426	697.9
Spatial			**659.0**
No Spatial	age	1426	389.5
	C. trachomatis infection	224	
Spatial			**362.3**
No Spatial	age	1426	251.2
	C. trachomatis load[e]	180	
Spatial			**232.7**

Table 3. Multivariable mixed effects linear regression analysis of factors predictive of *Chlamydia trachomatis* load (*omcB* copies/swab) in infected individuals

Model	Variable	*n*	OR[a]	95% C.I.	*p-Value*	*AIC*[b]	*BIC*[c]	*loglik*[d]
Null[e]								
						884.8	894.3	-439.4
Clustering								
Household[f]						884.3	893.8	-439.1
Village[f]						882.1	892.0	-438.2
Spatial[g]						799.7	834.8	-388.8
Final[h] *Including age and disease severity*								
Spatial						802.6	844.1	-388.3
No spatial						**797.7**	**829.6**	**-388.8**
Age	0-5 years	87	2.60	0.99, 6.87	**0.052**			
	6-10 years	45	0.70	0.24, 2.05	0.509			
	11-15 years	15	1.77	0.43, 7.20	0.427			
	>15 years	37	1.00	--	--			

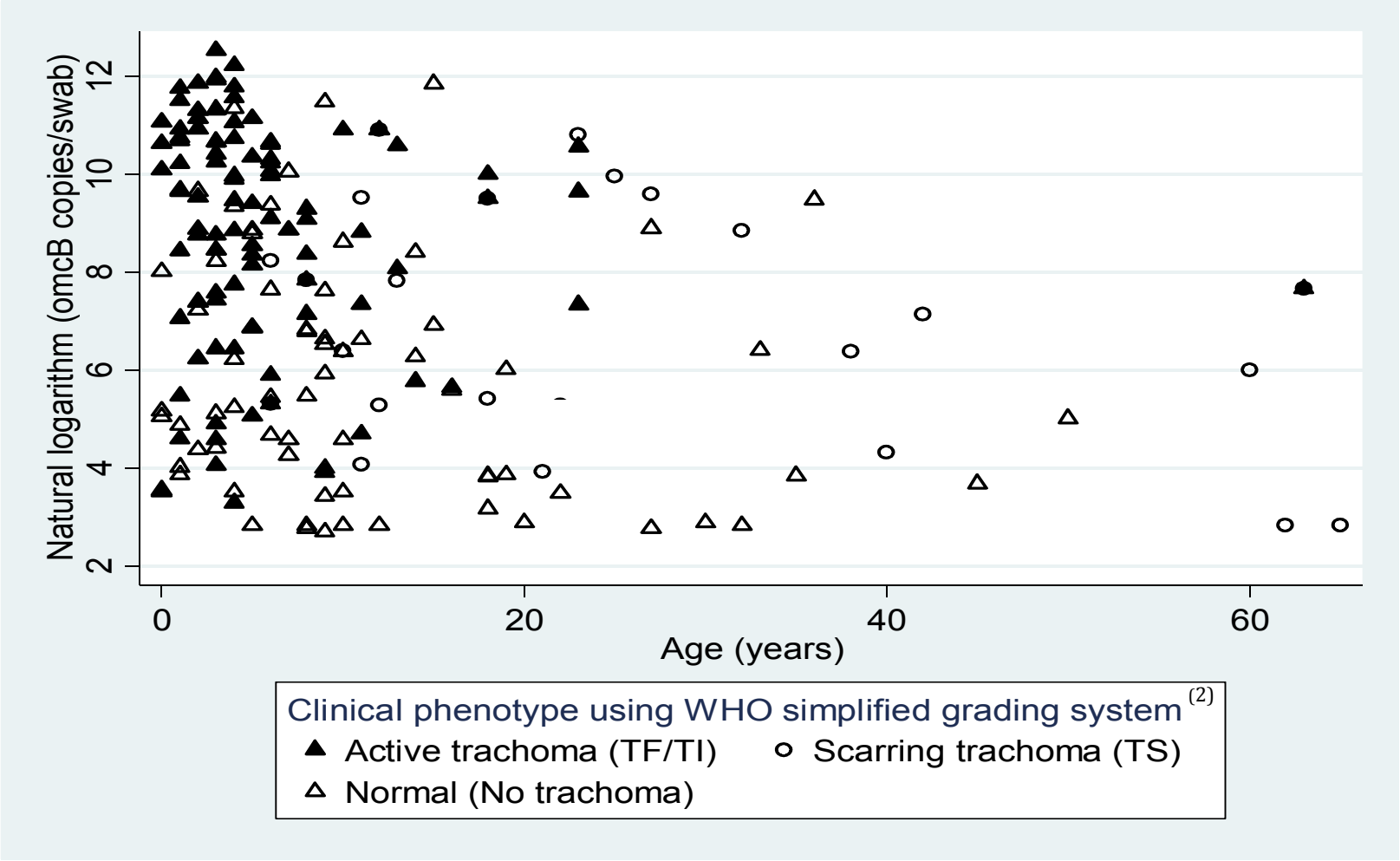
Disease Severity	Inflammatory Grade (P)								
		0	46	1.00	--	--			
		1	70	7.54	3.12, 18.20	**<0.0001**			
		2	46	22.8	8.15, 63.9	**<0.0001**			
		3	22	30.9	9.39, 101.50	**<0.0001**			
	Follicular Grade (F)								
		0	91	1.00	--	--			
		1	20	1.29	0.43, 3.84	0.649			
		2	27	2.20	0.82, 5.88	0.114			
		3	46	7.78	3.16, 19.15	**<0.0001**			

C. trachomatis load is defined as log-(e) *omcB* copies/swab. There was no evidence of heteroscedasticity of residuals (Breusch-Pagan/Cook Weisberg test for heteroscedasticity in the final model. (Chi2 = 0.44, *p=0.5079*)) [a]Exponentiated coefficient [b]AIC=Akaike information criterion [c]BIC=Bayesian information criterion [d]loglik=Log likelihoood [e]Null model with dummy cluster variable [f]Including household or village as cluster variables on outcome [g]Including spatial structure [h]Final model including covariates with and without adjustment for spatial structure

Figure 1. The Bijagós Archipelago, Guinea Bissau

The islands of Bubaque, Canhabaque, Rubane and Soga (circled in red) were included in the current study. © 2009 Ezilon.com

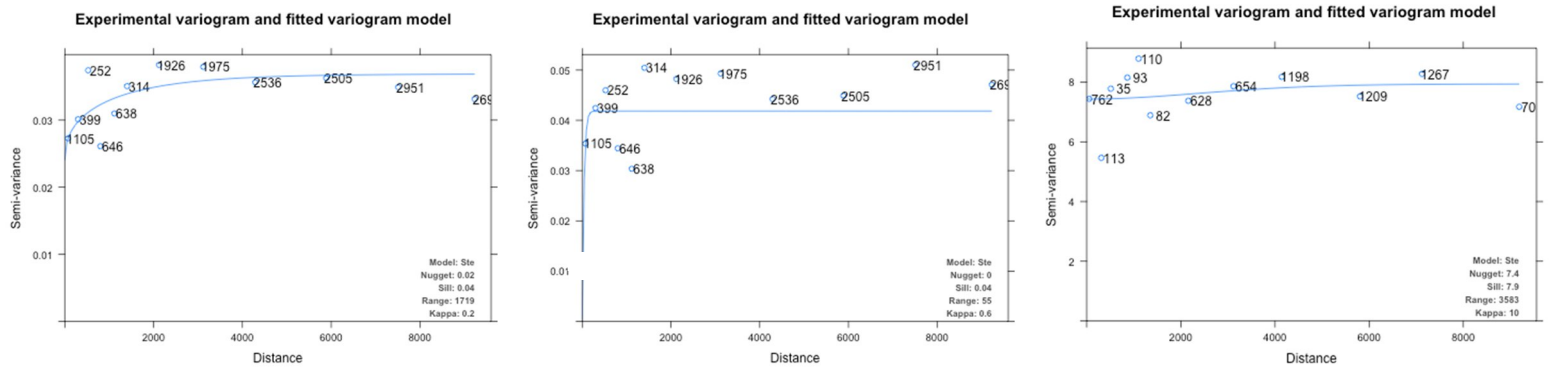
Figure 2. *Chlamydia trachomatis* load by age and clinical phenotype in infected individuals



y-axis shows the natural logarithmic scale of *C. trachomatis* bacterial load (*omcB* copies/swab)

Figure 3. Empirical semivariograms and fitted models for household prevalence of (a) ocular *Chlamydia trachomatis* infection and (b) active trachoma and the distribution of (c) ocular *Chlamydia trachomatis* bacterial load

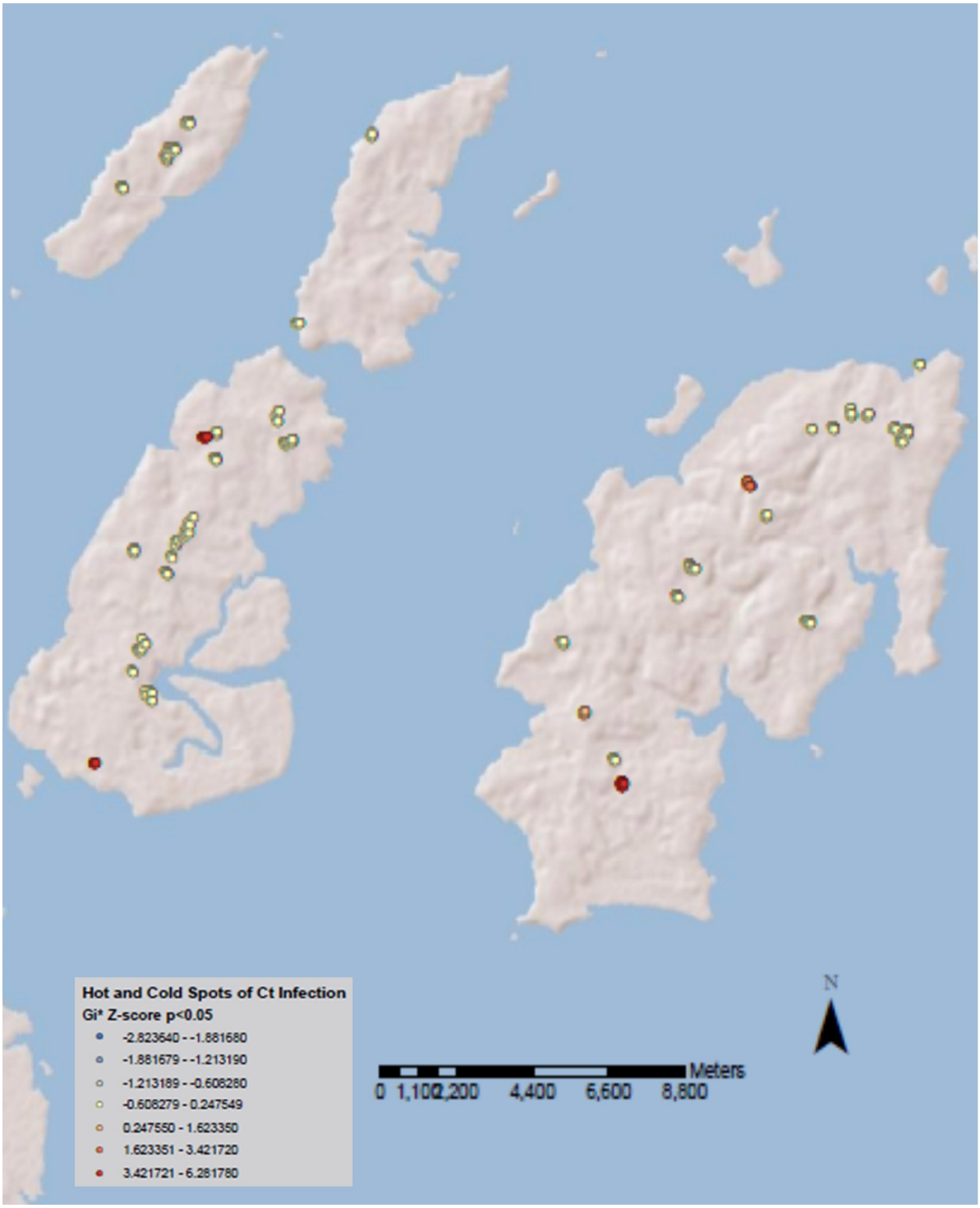
(a) (b) (c)



(a) Unadjusted household prevalence of *C. trachomatis* infection
(b) Household prevalence of active trachoma in 1-9 year olds
(c) Ocular *C. trachomatis* bacterial load

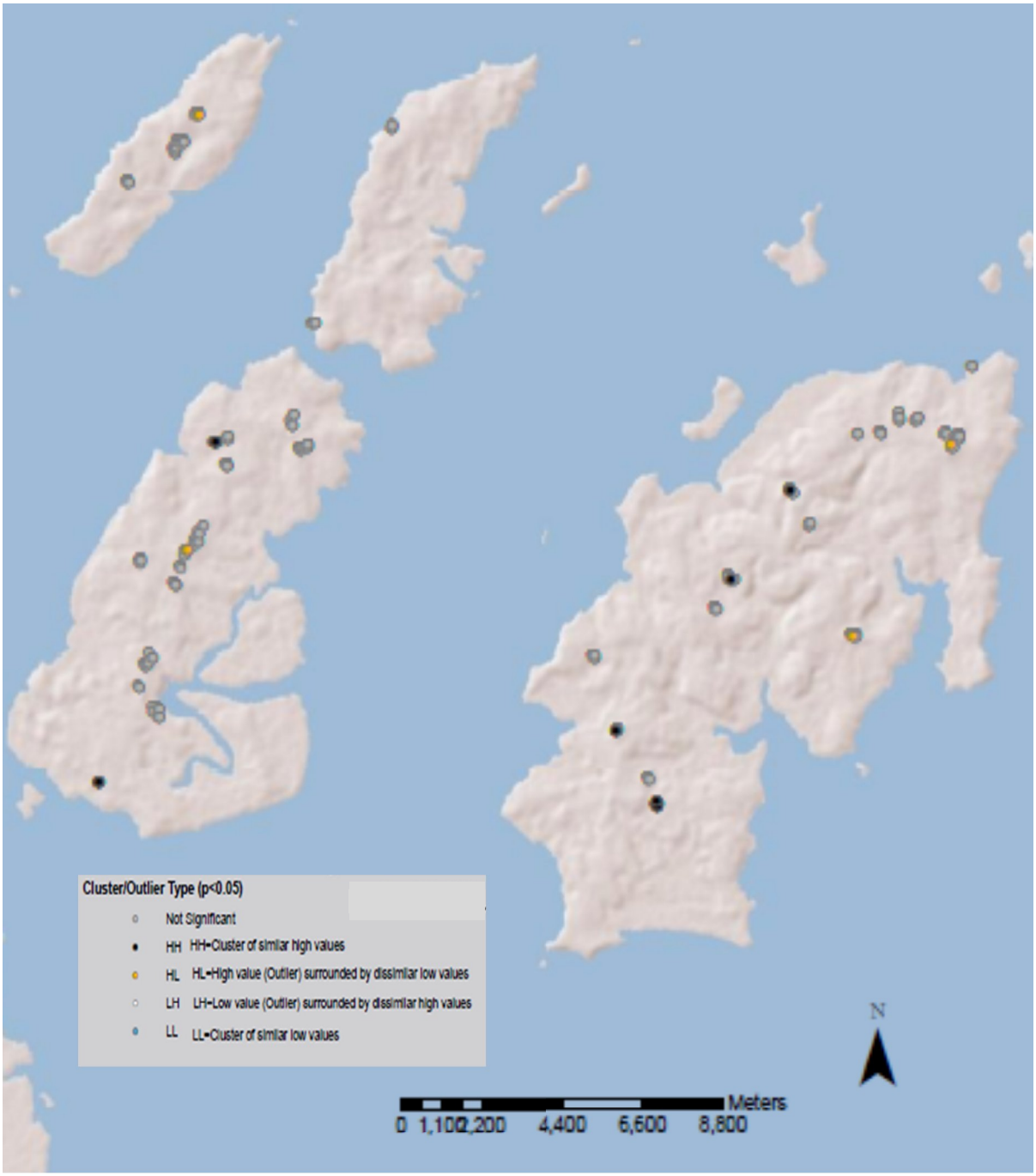
Prevalence data were log transformed (ln(ln+1)) due to significant negative skew. Active trachoma is defined as TF/TI by the WHO simplified grading system (2) (F2/3 or P3 by the modified FPC grading system (1)). Distance is indicated in metres. Model fit with the smallest residual sum of squares (Matern, M.Stein's parameterization (Ste)). All values of Kappa (smoothing parameter of the Matern model) tested. Nugget, sill, range and Kappa are all estimated from the data.

Figure 4. Hotspots of high load ocular *Chlamydia trachomatis* infections



C. trachomatis load was log transformed (ln(ln+1)) due to significant negative skew. Stastistically significant Getis-Ord Gi* Z-scores identify 'hotspots' containing similarly high values, taking into account the spatial dependence of neighbouring observations. There are no 'coldspots' (statistically significant clusters of similarly low values). Adjacency is defined by the zone of indifference.

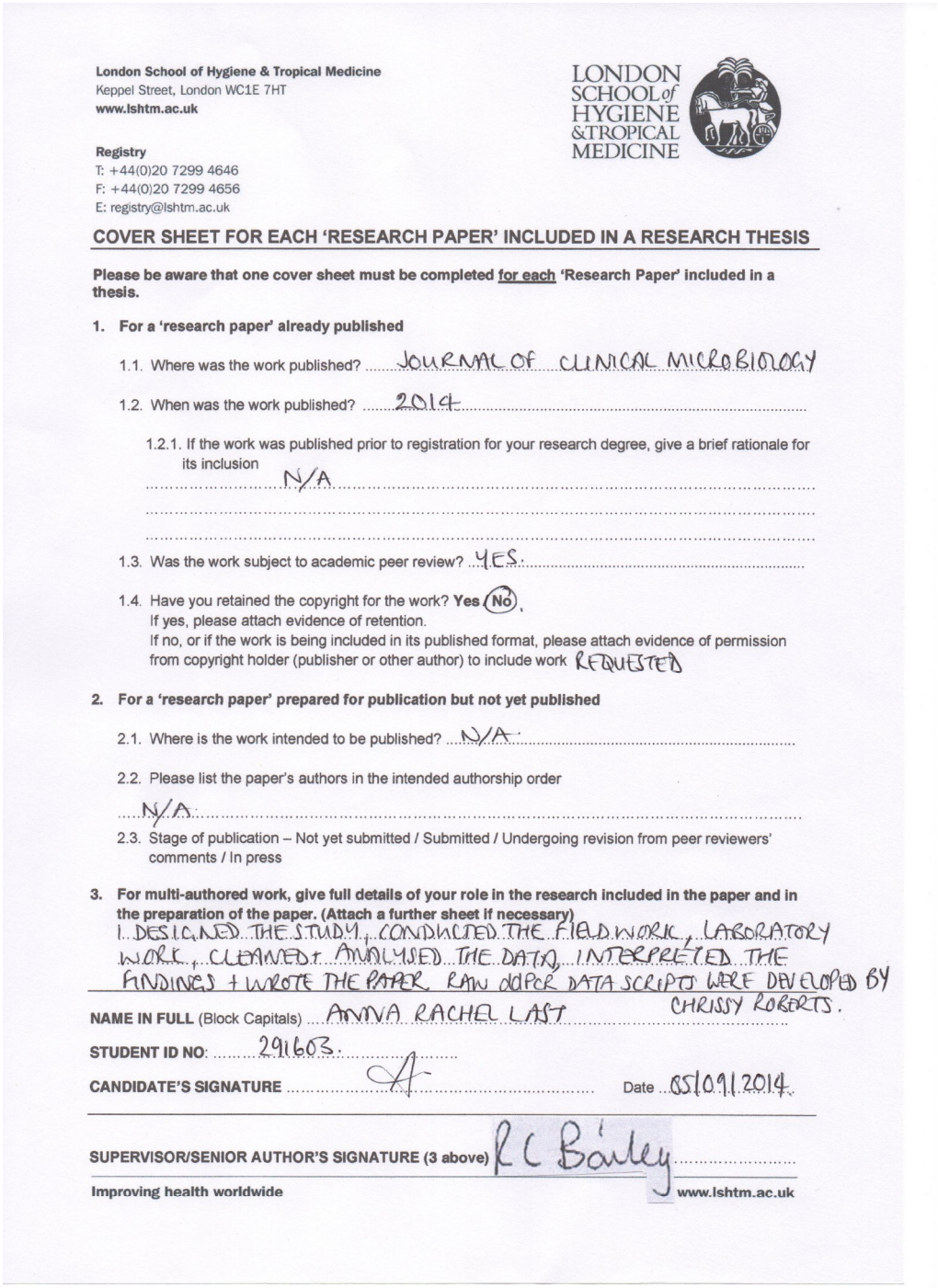
Figure 5. Clusters and outliers of high load ocular *Chlamydia trachomatis* infections



C. trachomatis load was log transformed (ln(ln+1)) due to significant negative skew. Statistically significant positive values for the Local Moran's I statistic indicate clustering with similarly high (H-H) or low (L-L) values. Negative values indicate that neighbouring observations have dissimilar values and that this observation is an outlier (H-L or L-H). H-H clusters and H-L outliers are observed. There are no L-L clusters. Observation values represent *C. trachomatis* load. Adjacency is defined by the zone of indifference.

Chapter 9

Plasmid copy number and disease severity in naturally occurring ocular *Chlamydia trachomatis* infection



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. **For a 'research paper' already published**

   1.1. Where was the work published? JOURNAL OF CLINICAL MICROBIOLOGY

   1.2. When was the work published? 2014

   1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
   N/A

   1.3. Was the work subject to academic peer review? YES.

   1.4. Have you retained the copyright for the work? **Yes** (No).
   If yes, please attach evidence of retention.
   If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work REQUESTED

2. **For a 'research paper' prepared for publication but not yet published**

   2.1. Where is the work intended to be published? N/A

   2.2. Please list the paper's authors in the intended authorship order
   N/A

   2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. **For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)**
   I DESIGNED THE STUDY, CONDUCTED THE FIELD WORK, LABORATORY WORK, CLEANED & ANALYSED THE DATA, INTERPRETED THE FINDINGS & WROTE THE PAPER. RAW QPCR DATA SCRIPTS WERE DEVELOPED BY CHRISSY ROBERTS.

NAME IN FULL (Block Capitals) ANNA RACHEL LAST

STUDENT ID NO: 291603.

CANDIDATE'S SIGNATURE ..... Date 05/09/2014.

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) R C Bailey

Improving health worldwide www.lshtm.ac.uk

TITLE

Plasmid copy number and disease severity in naturally occurring ocular *Chlamydia trachomatis* infection

RUNNING TITLE

Ocular *C. trachomatis* infection and plasmid copy number by ddPCR

Anna R Last[1#], Chrissy h Roberts[1], Eunice Cassama[2], Meno Nabicassa[2], Sandra Molina-Gonzalez[1], Sarah E Burr[1,3], David CW Mabey[1], Robin L Bailey[1], Martin J Holland[1]

[1]Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom

[2]Programa Nacional de Saúde de Visão, Ministério de Saúde Publica, P.O. Box 50, Avenida de Unidade Africana, Bissau, Guiné Bissau

[3]Disease Control and Elimination Theme, Medical Research Council Unit The Gambia, P.O. Box 273 Banjul, Atlantic Boulevard, Fajara, The Gambia

#CORRESPONDING AUTHOR Mailing Address: Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Telephone: +44 (0)207 927 2566 Email: anna.last@lshtm.ac.uk

Abstract

The *Chlamydia trachomatis (Ct)* plasmid is a virulence factor. Plasmid copy number, *Ct* load and disease severity were assessed in a trachoma-hyperendemic treatment-naïve population. Using droplet digital PCR plasmid copy number was found to be stable (median 5.34 (1-18)) and there were no associations with *Ct* load or disease severity.

KEYWORDS *Chlamydia trachomatis; trachoma; disease severity; plasmid copy number; ddPCR*

Trachoma is caused by infection with ocular strains of *Chlamydia trachomatis*. The 7.5-kb *C. trachomatis* plasmid has been shown to function as a virulence factor in animal models (1,2). Phenotypic differences exist between plasmid-cured and wild-type *C. trachomatis* strains with respect to infectivity, glycogen accumulation, induction of inflammation and activation of toll-like receptor pathways (3,4). Plasmid deletion mutagenesis studies showed that deletion of the plasmid-encoded *pgp4* gene results in an *in vitro* phenotype identical to that of a plasmid-free strain (5). This supports bacterial transcriptome analysis showing a decrease in transcript levels of a subset of chromosomal genes in a naturally occurring plasmid-free strain of *C. trachomatis*, demonstrating that the plasmid is a transcriptional regulator of virulence-associated genes (6).

There is little information in the literature relating plasmid copy number (per genome) to virulence (7-9). The mechanisms of plasmid virulence are not clearly defined, particularly in naturally occurring infections. We assessed plasmid copy number variation and its association with disease severity in ocular *C. trachomatis* infection from a trachoma-hyperendemic treatment-naïve population on the Bijagós Archipelago of Guinea Bissau.

Individuals from 300 randomly selected households from 38 villages on four islands were examined by a single trained examiner using the WHO simplified and modified FPC grading systems (10,11). In the FPC system, follicles (F), papillae (P) and conjunctival scarring (C) are separately scored 0-3. Active disease (TF (follicular trachoma) or TI (inflammatory trachoma) by the WHO simplified system) equates to F2/3 and P3 respectively. C2/3 (and in some cases C1) is equivalent to TS (trachomatous scarring). Both systems were used to provide detailed phenotypic

information and comparability with other studies. Individuals' age, sex and ethnicity were recorded.

Swabs were taken from the left upper tarsal conjunctiva of each participant using a validated procedure (12,13). Swabs were collected dry into microcentrifuge tubes (Simport, Canada), kept on ice in the field and frozen to -80^{0}C within 8 hours of collection. Measures were taken to avoid cross-contamination in the field and in the laboratory (13).

DNA extraction and droplet digital PCR (ddPCR) for detection of *C. trachomatis* plasmid were conducted as described previously (14). A second duplex assay was used to estimate plasmid and chromosome (*omcB*) target concentrations within the same reaction in plasmid positive samples. We used published primer-probe target sequences appropriate for quantitation of all genovars of *C. trachomatis* (7,14). We used a modified *omcB* probe to improve quenching efficiency and reduce background fluorescence (*Table 1*). Methods for mastermix preparation, droplet generation, thermal cycling conditions, droplet reading, target DNA concentration calculation and re-testing of saturated samples are described elsewhere (14). Estimated quantities of *omcB* and plasmid are expressed as copies/swab. *C. trachomatis* load refers to *omcB* copies/swab. Plasmid copy number (per genome) was calculated using the plasmid:genome ratio.

Raw quantitation data were processed as previously described (14). Geometric mean *omcB* load and linear and logistic regression analyses (with odds ratios (OR)) were conducted in STATA 12 (Stata Corporation, College Station, Texas USA) to examine associations between plasmid copy number, load and detailed clinical phenotype. *C. trachomatis* load and plasmid copy number data were log-(e) transformed and robust standard error used where indicated.

Of 1,511 individuals enrolled, 1,508 individuals consented to ocular assessment and 1,507 conjunctival swabs were obtained. The median age of participants was 13 years (1 month-88 years) and 57% were female. Most participants were of the Bijagós ethnic group. The prevalence of active trachoma (TF/TI) in 1-9 year olds was 21% (136/660) (95% CI 17.89-24.11%). Overall, 11% had clinically active trachoma (164/1508) (95% CI 9.42-12.58%). *C. trachomatis* plasmid DNA was detected in 16% overall (233/1507) (26% of 1-9 year olds). All samples were adequate by criteria described previously (14).

C. trachomatis load was estimated in 79% (184/233) of plasmid positive samples. In 21% of samples where plasmid load was very low, *omcB* was below the level of detection.

The geometric mean estimated *omcB* copies/swab varied by clinical phenotype (294 copies/swab (95% C.I. 165-524) in 73 with normal conjunctivae, 8562 copies/swab (95% C.I. 5412-13546) in 92 with active trachoma and 928 copies/swab (95% C.I. 280-2074) in 19 with scarring.

The median plasmid copy number was 5.34 (1-18.03) (*Figure 1*). Plasmid copy number was stable in infections across the four study islands (Kruskall Wallis H (Chi2)=4.5001(df=3,*p=0.2123*)). Plasmid copy number was not associated with the presence of active trachoma (OR 1.00, 95% CI 0.88-1.12, *p=0.960*), severity of inflammatory (OR 1.04, 95% CI 0.927-1.162, *p=0.515*) or follicular (OR 1.03, 95% CI 0.922-1.159, *p=0.572*) disease or *C. trachomatis* load (*Table 2*). At lower load the variance was highly heterogeneous (Levene's W0=55.3, df=2, *p<0.000000001*) (*Figure 2).*

The theoretical advantages of ddPCR are presented by Hindson *et al*. (15). These include nanolitre-sized droplet partitioning of the reaction, which promotes optimal

primer-template interaction conditions robust to variation in PCR efficiency, thus enabling accurate estimation of both plasmid and *omcB* copy numbers within the same reaction. We have discussed the precision and accuracy of our diagnostic ddPCR assay elsewhere (14).

There are a few published studies examining plasmid copy number in reference strains of *C. trachomatis* (7-9,16-17). Pickett *et al*. showed that across 12 *C. trachomatis* serovars the plasmid copy number was not significantly different but there were variations depending on growth phase and condition during *in vitro* culture (7). Seth-Smith *et al*. showed an increased plasmid copy number in ocular relative to urogenital strains (8). We demonstrate a stable plasmid copy number distribution in naturally occurring ocular *C. trachomatis* infection that does not vary with geographic location, clinical phenotype or *C. trachomatis* load. Our data show that ddPCR may have limitations in measuring plasmid copy number at very low load infections (<200 *omcB* copies/swab) where plasmid copy number variance is greatest. This observation may reflect a breakdown in the assumptions required to apply the Poisson distribution to accurately estimate load with ddPCR. Despite the caveats, our data suggest plasmid copy number stability in naturally occurring ocular *C. trachomatis* infection.

Maintenance of the plasmid at low copy number carries inherent risk during cell partition (18) but naturally occurring plasmid-free strains are rare (19-21). A lower risk higher copy number system is metabolically expensive but may confer a 'fitness' advantage. Thus, the maintenance of 5-6 plasmids per genome may maximise infectivity or intracellular survival whilst provoking minimal host immune response.

Though there is convincing evidence that the chlamydial plasmid is a virulence factor (3,4,6,22-24) our data suggest that plasmid copy number is not associated with disease severity and that additive gene dosage effects do not appear to correlate with

pathogen virulence *in vivo*. This supports *in vitro* work showing no association between plasmid copy number and tissue tropism (9). Previous work *in vitro* and in animal models suggests that subtle genomic differences between chlamydial isolates are associated with differences in growth kinetics, immune responses and pathology (25,26). Further epidemiological and *in vitro* studies using comparative pathogen genomics to examine these associations are required to fully understand the relationship between disease severity and chlamydial virulence.

Acknowledgements

This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained from the Comitê Nacional de Ética e Saúde (Guinea Bissau), the LSHTM Ethics Committee (UK) and the The Gambia Government/MRC Joint Ethics Committee (The Gambia). Written (thumbprint or signature) informed consent was obtained from all study participants or their guardians as appropriate. Following the survey all communities on the study islands were treated with azithromycin in line with WHO and national protocols.

This work was supported by the Wellcome Trust (grant number 097330/Z/11/Z). All reagents for the study were purchased at full list price but the authors are grateful to Bio-Rad Laboratories Inc. for the loan of the QX-100 platform free of charge for use in this study. We extend thanks to the study participants and field research team in Guinea Bissau.

All authors declare no conflict of interests.

References

[1] **Kari L, Whitmire WM, Olivares-Zavaleta N, Goheen MM, Taylor LD, Carlson JH, Sturdevant GL, Lu C, Bakios LE, Randall LB, Parnell MJ, Zhong G, Caldwell HD**. 2011. A live-attenuated chlamydial vaccine prevents against trachoma in non-human primates. J Exp Med **208**(11):2217-23.

[2] **Frazer LC, Darville T, Chandra-Kuntal K, Andrews CW Jr, Zurenski M, Mintus M, AbdelRahmen YM, Belland RJ, Ingalls RR, O'Connell CM**. 2012. Plasmid-cured *Chlamydia caviae* activates TLR2-dependent signaling and retains virulence in the guinea pig model of genital tract infection. PLoS One **7**(1):e30747.

[3] **O'Connell CM, AbdelRahman YM, Green E, Darville HK, Saira K, Smith B, Darville T, Scurlock AM, Meyer CR, Belland RJ**. 2011. Toll-like receptor 2 activation by *Chlamydia trachomatis* is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by *C. trachomatis* but not by *C. muridarum*. Infect Immun **79**:1044-1056.

[4] **Russell M, Darville T, Chandra-Kuntal K, Smith B, Andrews CW Jr, O'Connell CM**. 2011. Infectivity acts as *in vivo* selection for maintenance of the chlamydial cryptic plasmid. Infect Immun **79**:98–107.

[5] **Song L, Carlson JH, Whitmire WM, Kari L, Virtaneva K, Sturdevant DE, Watkins H, Zhou B, Sturdevant GL, Porcella SF, McClarty G, Caldwell HD**. 2013. *Chlamydia trachomatis* plasmid-encoded *pgp4* is a transcriptional regulator of virulence associated genes. Infect Immun **81**(3):636.

[6] **Carlson JH, Whitmire WM, Crane DD, Wicke L, Virtaneva K, Sturdevant DE, Kupko JJ 3rd, Porcella SF, Martinez-Orengo N, Heinzen RA, Kari L, Caldwell HD**. 2008. The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect Immun **76**:2273-83.

[7] **Pickett MA, Everson JS, Pead PJ, Clarke IN**. 2005. The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumonia*e (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. Microbiology **151**:893-903.

[8] **Seth-Smith HM, Harris SR, Persson K, Marsh P, Barron A, Bignall A, Bjartling C, Clarke L, Cutcliffe LT, Lambden PR, Lennard N, Lockey SJ, Quail MA, Salim O, Skilton RJ, Wang Y, Holland MJ, Parkhill J, Thomson NR, Clarke IN**. 2009. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. BMC Genomics **10**:239-49.

[9] **Ferreira R, Borges V, Nunes A, Borrego MJ, Gomes JP**. 2013. Assessment of the load and transcriptional dynamics of *Chlamydia trachomatis* plasmid according to strains' tissue tropism. Microbiol Res **168**(6):333-9.

[10] **Dawson CR, Jones BR, Tarizzo ML**. 1981. Guide to trachoma control in programs for the prevention of blindness. Geneva: World Health Organization.

[11] **Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR**. 1987. A simple system for the assessment of trachoma and its complicatios. Bull World Health Organ **65**(4):477-83.

[12] **Keenan JD, Lakew T, Alemayehu W, Melese M, Porco TC, Yi E, House JI, Zhou Z, Ray KJ, Acharya NR, Whitcher JP, Gaynor BD, Lietman TM**. 2010. Clinical activity and polymerase chain reaction evidence of chlamydial infection after repeated mass antibiotic treatments for trachoma. Am J Trop Med Hyg **82**(3):482-7.

[13] **Stare D, Harding-Esch E, Muñoz B, Bailey R, Mabey D, Holland M, Gaydos C, West S**. 2011. Design and baseline data of a randomised trial to evaluate coverage and frequency of mass treatment with azithromycin: The Partnership for the Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. Ophthalmic Epidemiol **18**(1):20-9.

[14] **Roberts CH, Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M, McCarthy E, Burr SE, Mabey DC, Bailey RL, Holland MJ**. 2013. Development and evaluation of a next generation digital PCR diagnostic assay for ocular *Chlamydia trachomatis* infections. J Clin Microbiol **51**(7):2195-2203.

[15] **Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M**. 2013. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods **10**(10):1003-5.

[16] **Palmer L and Falklow S**. 1986. A common plasmid of *Chlamydia trachomatis*. Plasmid **16**:52-62.

[17] **Tam JE, Davis CH, Thresher RJ, Wyrick P**. 1992. Location of the origin of replication for the 7.5-kb *Chlamydia trachomatis* plasmid. Plasmid **27**:231-6.

[18] **Twigg AJ and Sheratt D**. 1980. Trans-complementable copy-number mutants of plasmid ColE1. Nature **283**(5743):216-8.

[19] **Thomas NS, Lusher M, Storey CC, Clarke IN**. 1997. Plasmid diversity in chlamydia. Microbiology **143**(6):1847-54.

[20] **An Q, Radcliffe G, Vassallo R, Buxton G, O'Brien WJ, Pelletier DA, Weisberg WG, Klinger JD, Olive DM**. 1992. Infection with a plasmid-free variant Chlamydia related to *Chlamydia trachomatis* identified by using multiple assays for nucleic acid detection. J Clin Microbiol **30**:2814-21.

[21] **Peterson EM, Markoff BA, Schacter J, de la Maza LM**. 1990. The 7.5kb plasmid present in *Chlamydia trachomatis* is essential for the growth of the micro-organism. Plasmid **23**:144-8.

[22] **Nunes A, Borrego MJ, Gomes JP**. 2013. Genomic features beyond *C. trachomatis* phenotypes: what do we think we know? Infect Genet Evol **16**:392-400.

[23] **O'Connell CM, Ingalls RR, Andrews CW Jr, Scurlock AM, Darville T**. 2007. Plasmid-deficient *Chlamydia muridarum* fail to induce immune pathology and protect against oviduct disease. J Immunol **179**:4027-34.

[24] **Comanducci M, Manetti L, Bini A, Santucci A, Pallini V, Cevenini R, Sueur JM, Orfila J, Ratti G**. 1994. Humoral immune response to plasmid protein *pgp-3* in patients with *Chlamydia trachomatis* infection. Infect Immun **62**:5491-7.

[25] **Kari L, Whitmire WM, Carlson JH, Crane DD, Reveneau N, Nelson DE, Mabey DCW, Bailey RL, Holland MJ, McClarty G, Caldwell HD**. 2008. Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis **197**:449-56.

[26] **Miyairi I, Laxton JD, Wang X, Obert CA, Arva Tatireddigari VRR, van Rooijen N, Hatch TP, Byrne GI**. 2011. *Chalmydia psittaci* genetic variants differ in virulence by modulation of host immunity. J Infect Dis. **204**(4):654-663.

Tables and Figures

Table 1. Primer and probe sequences for control and *C. trachomatis* targets using the ddPCR system

Molecular Target	**Nucleotide Sequence and Modifications**
Internal Control ***Homo sapiens* RPP30 (encoding ribonuclease P/MRP 30kDa subunit)**[a]	
Forward (RPP30-F)	5' AGA TTT GGA CCT GCG AGC G 3'
Reverse (RPP30-R)	5' GAG CGG CTG TCT CCA CAA GT 3'
Probe (RPP30_HEX_BHQ1)[a]	5' HEX-TTC TGA CCT GAA GGC TCT GCG CG-BHQ1-3'
Target One ***C. trachomatis* cryptic plasmid pLGV440 (circular; genomic DNA; 7500bp)**	
Forward (Ct-plasmid-F)	5' CAG CTT GTA GTC CTG CTT GAG AGA 3'
Reverse (Ct-plasmid-R)	5' CAA GAG TAC ATC GTT CAA CGA AGA 3'
Probe (Ct-plasmid_FAM_BHQ1)[a,b]	5' 6FAM-CCC CAC CAT TTT TCC GGA GCG A-BHQ1-3'
Probe (Ct-plasmid_HEX_BHQ1)[c]	5' HEX-CCC CAC CAT TTT TCC GGA GCG A-BHQ1-3'
Target Two ***C. trachomatis* (Serovar A) *omcB* gene**	
Forward (Ct-*omcB*-F)	5' GAC ACC AAA CGC AAA GAC AAC AC 3'
Reverse (Ct-*omcB*-R)	5' ACT CAT GAA CCG GAG CAA CCT 3'
Probe (Ct-*omcB*-FAM-BHQ1)	5'6FAM-CCA CAG CAA AGA GAC TCC CGT AGA CCG-BHQ1-3'

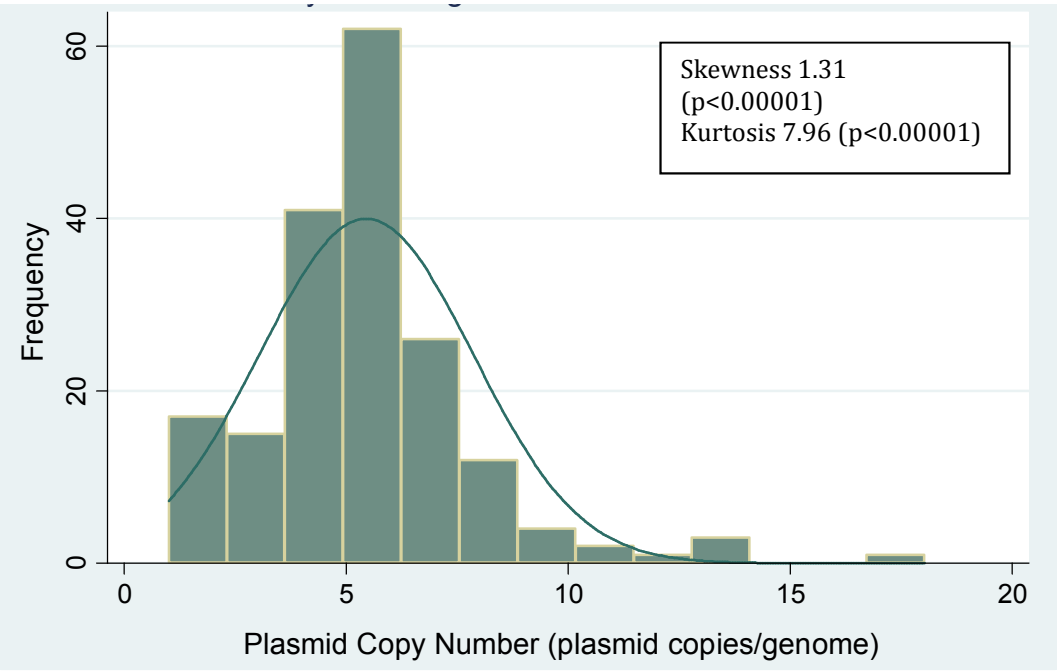
[a]MRP, mitochondrial RNA processing endoribonuclease; HEX, hexachlorofluorescein reporter; BHQ, black hole quencher; 6FAM, 6-carboxyfluorescein reporter
[b]*C. trachomatis* plasmid probe used in screening (first) assay
[c]*C.trachomatis* probe used in quantitative (second) assay

Table 2. The relationship between plasmid copy number and *C. trachomatis* load (*omcB* copies/swab)

***OmcB* (copies/swab)**	**n**	**Variance**	**Plasmid Copy Number[a] Min**	**Median**	**Max**
<100	41	19.8139	1	4.1514	18.0291
100-10000	82	2.7136	1	5.3421	9.2819
>10000	62	1.0814	3.6164	5.4261	8.3947

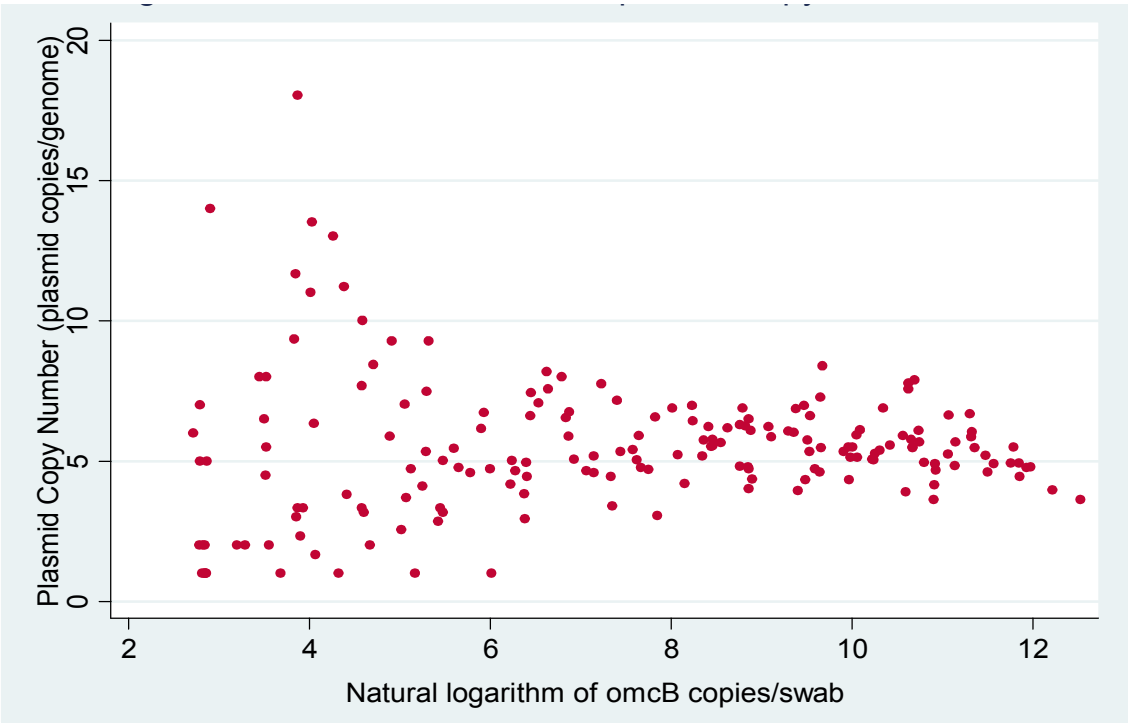
Kruskall Wallis H (Chi2) = 4.58, df=2, *p=0.10.*
[a]Plasmid Copy Number = plasmid copies/genome

Figure 1. Distribution of plasmid copy number variation in naturally occurring ocular infection with *C. trachomatis* within the study population



n	Min	Max	Mean	Standard Error(mean)	Median	25%	75%
184	1	18.03	5.45	0.177	5.34	4.39	6.46

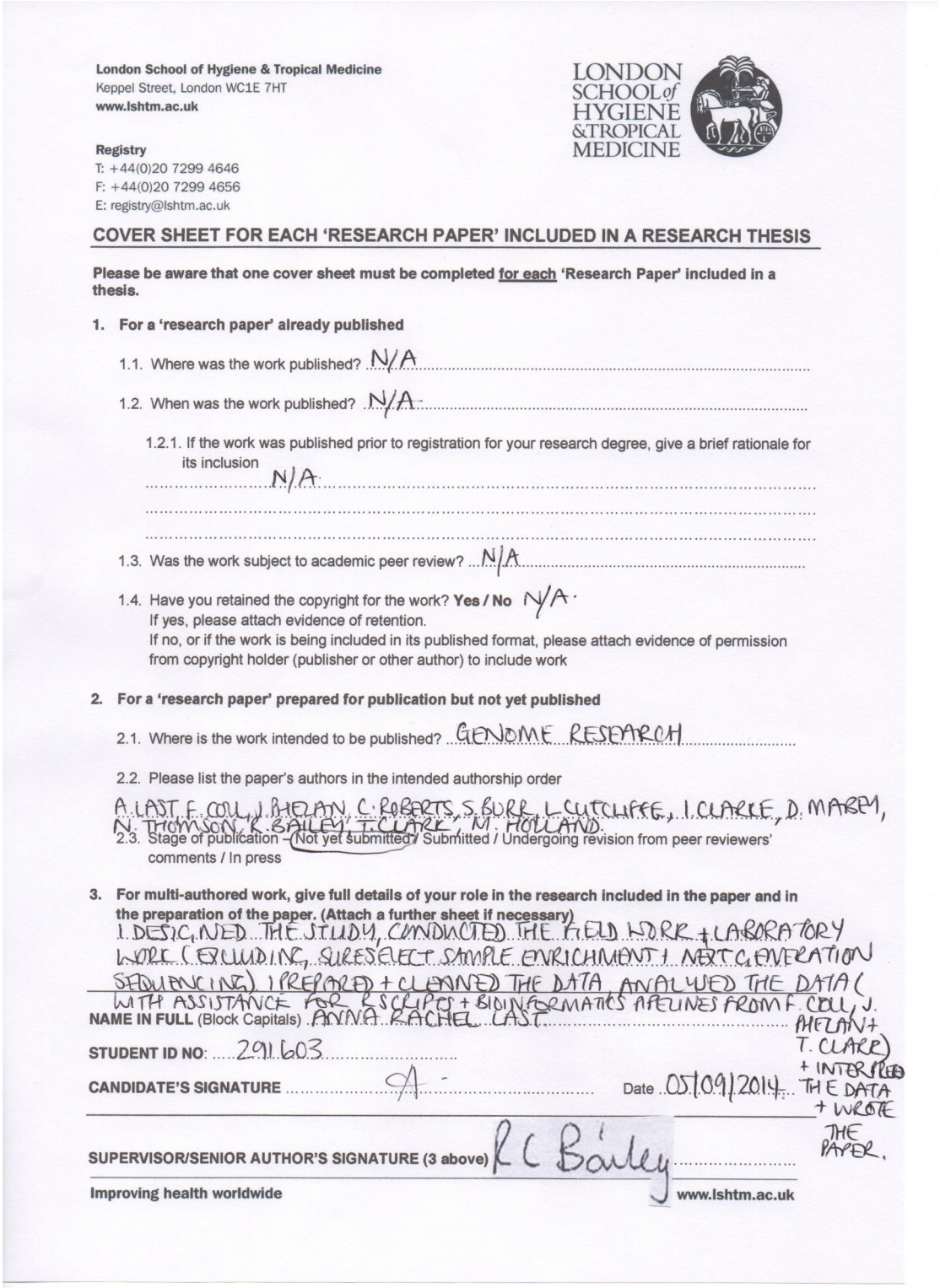
Figure 2. *C. trachomatis* load and plasmid copy number variation



y axis = Plasmid copy number per genome (*omcB* copies/plasmid)

Chapter 10

Identifying genes and loci associated with disease severity and ocular *Chlamydia trachomatis* load in trachoma using a pathogen Genome Wide Association Scan (GWAS) analysis



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. **For a 'research paper' already published**

   1.1. Where was the work published? N/A

   1.2. When was the work published? N/A

   1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
   N/A

   1.3. Was the work subject to academic peer review? N/A

   1.4. Have you retained the copyright for the work? **Yes / No** N/A
   If yes, please attach evidence of retention.
   If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. **For a 'research paper' prepared for publication but not yet published**

   2.1. Where is the work intended to be published? GENOME RESEARCH

   2.2. Please list the paper's authors in the intended authorship order
   A. LAST, F. COLL, J. PHELAN, C. ROBERTS, S. BURR, L. CUTCLIFFE, I. CLARKE, D. MABEY, N. THOMSON, R. BAILEY, T. CLARK, M. HOLLAND.

   2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. **For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)**
   I DESIGNED THE STUDY, CONDUCTED THE FIELD WORK + LABORATORY WORK (EXCLUDING SURESELECT SAMPLE ENRICHMENT + NEXT GENERATION SEQUENCING) I PREPARED + CLEANED THE DATA, ANALYSED THE DATA (WITH ASSISTANCE FOR R SCRIPTS + BIOINFORMATICS PIPELINES FROM F. COLL, J. PHELAN + T. CLARK) + INTERPRETED THE DATA + WROTE THE PAPER.

NAME IN FULL (Block Capitals) ANNA RACHEL LAST

STUDENT ID NO: 291603

CANDIDATE'S SIGNATURE A Date 05/09/2014

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) R C Bailey

Improving health worldwide www.lshtm.ac.uk

Identifying genes and loci associated with disease severity and ocular *Chlamydia trachomatis* load in trachoma using a pathogen Genome Wide Association Scan (GWAS) analysis

AR Last[1], F Coll[2], J Phelan[2], Ch Roberts[1], SE Burr[1,3], LT Cutcliffe[4], IN Clarke[4], DCW Mabey[1], NR Thomson[5], RL Bailey[1], TG Clark[2], MJ Holland[1]

[1]Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom, [2]Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom, [3]Disease Control and Elimination Theme, Medical Research Council Unit The Gambia, Fajara, The Gambia, [4]Molecular Microbiology Group, University of Southampton Medical School, Southampton, United Kingdom, [5]Pathogen Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom.

ABSTRACT

This is the first population-based genome-wide association scan (GWAS) for *Chlamydia trachomatis* performed on whole genome sequence data obtained directly from conjunctival samples taken from treatment-naïve trachoma-endemic communites in West Africa. Whole genome sequence data were obtained directly from clinical samples using novel DNA-bait sample enrichment. We analysed the sequence data in the context of detailed *in vivo* conjunctival phenotypes and ocular *C. trachomatis* load (*omcB* load by droplet digital PCR). We identified a total of 1050 SNPs in coding regions, representing greater than expected genomic diversity. GWAS was performed on 759 non-synonymous SNPs in coding regions. A number of associations were identified at genome-wide significance (*p=6.25xe-05*). The strongest association was between *incE* and conjunctival inflammation (OR=2.51, *p=5.16xe-08*). Two other loci (*mutY* and *CTA_0271)* were associated with *C. trachomatis* ocular load. These findings were supported by the presence of other synonymous SNPs in close proximity within these genes and loci. These data inform us about host-pathogen interactions that are likely to be important in *C. trachomatis* pathogenesis and may also be important vaccine and therapeutics targets.

KEYWORDS

Ocular *Chlamydia trachomatis*, bacterial load, conjunctival inflammation, genome wide association scan, virulence, pathogenesis

INTRODUCTION

Chlamydia trachomatis is an obligate intracellular human pathogen and is the leading infectious cause of blindness and the most common sexually transmitted bacterial infection contributing to significant global morbidity and mortality (Mabey 2003, Hu 2010). Chlamydial strains are differentiated into biovars based on pathobiological characteristics and serovars based on serospecificity for the major outer membrane protein (MOMP) encoded by the *ompA* gene (Rodriguez 1991). Serovars fall into biological groups associated with trachoma (A-C), sexually transmitted non-invasive disease (D-K) and lymphogranuloma venereum (LGV) (L1-3). Despite the diverse biological phenotypes (tissue tropism and disease outcome), *C. trachomatis* strains share near genomic synteny, suggesting that minor genetic changes determine pathogen-host and tissue-specific infection characteristics (Caldwell 2003, Jeffrey 2010, Nunes 2013).

Underlying host and pathogen factors responsible for such diverse clinical outcomes are unclear. Strain-specific differences have been described related to clinical presentation in trachoma (Hayes 1992, Bailey 1994, Kari 2008, Andreason 2010). Comprehensive genomic analysis of a limited pool of laboratory and clinical isolates of the major trachoma serotypes found subtle genetic differences relating to *in vitro* and *in vivo* pathogenic properties in the different isolates (Kari 2008). Whole genome sequencing has also been used to identify regions of likely recombination in recent clinical strains, demonstrating that whole-genome sequencing and comparative genomics is an effective approach for discovering variable loci in *C. trachomatis* that are associated with clinical presentation (Jeffrey 2010). Putative virulence factors identified through whole genome sequence analysis and subsequent *in vitro* and animal studies include the polymorphic outer membrane protein family (Stephens 1998, Longbottom 1998, Stothard 2003, Gomes 2006), type III secretion apparatus and effectors (Rockey 1995, Hsia 1997, Hefty 2007), genes involved in tryptophan (Caldwell 2003, Wood 2003, Carlson 2006) and glycogen (O'Connell 2011, Wang 2011) biosynthesis, members of the *incA* (Hackstadt 1999, Suchland 2000) and phospholipase D (Nelson 2006, Taylor 2010) families, genes from the heat shock protein family (LaVerda 2000, Kol 2000), the chlamydial cytotoxin (Carlson 2004) and plasmid-encoded genes implicated in the regulation of chromosomal virulence factors (Carlson 2008, Frazer 2012, Song 2013).

Basic research into *C. trachomatis* has long been hampered by significant technical barriers due to its obligate intracellular life cycle. Only recently has genetic manipulation of the chlamydial plasmid been possible, allowing for *in vitro* transformation and modification studies (Wang 2011, Wang 2013). Despite these advances there are limited genomic data available for current clinical strains, rendering it difficult to relate genetic changes to functional relevance and virulence factors *in vivo*.

In the current study we used a genome-wide approach to investigate associations with disease and chlamydial load in naturally occurring ocular *C. trachomatis* infections from treatment-naïve communities in a trachoma-endemic region in West Africa. We have previously published data from this population showing strong associations between ocular *C. trachomatis* load and disease severity, particularly in relation to inflammation (Last 2014a). We have also demonstrated spatial clustering of high load ocular *C. trachomatis* infections, suggesting that bacterial load may be important in pathogenesis and transmission (Last 2014a). This is the first study to use this approach to examine whole genome sequence data obtained directly from a cross-sectional collection of current clinical samples with *C. trachomatis* infection in the context of detailed clinical, epidemiological and spatial metadata.

METHODS

Survey and Sample collection

Survey, clinical examination and sample collection methods have been described previously (Last 2014a, Last 2014b, Last 2014c). Briefly, we conducted a cross-sectional population-based survey in trachoma-endemic communities on the Bijagós Archipelago of Guinea Bissau. The upper tarsal conjunctivae of each consenting participant were examined, two digital photographs were taken and a clinical grade was assigned using both the WHO simplified grading system (Thylefors 1981) and the modified FPC system (Dawson 1987). The modified FPC system allows detailed scoring of the conjunctiva for the presence of follicles (F), papillary hypertrophy (inflammation) (P) and conjunctival scarring (C), receiving a grade of 0-3 for each parameter. For this study we define the clinical phenotypes as active trachoma (where the conjunctiva receives a grade of follicular trachoma (TF) and/or intense inflammatory trachoma (TI) by the WHO grading system (F2 and/or P3 by the modified FPC system)) and as a scale of conjunctival inflammation (P0-P3). The

inflammatory phenotype (P score) was chosen as a phenotype on the basis of previous studies showing strong association between inflammation and *C. trachomatis* bacterial load (Burton 2006, Faal 2006, Last 2014a). Conjunctival swabs were obtained and DNA extracted and quantified using droplet digital PCR (ddPCR) as described previously (Roberts 2013, Last 2014c). *C. trachomatis* load refers to *omcB* (genomic) copies/swab.

Pre-sequencing sample enrichment

Whole genome sequence (WGS) data were obtained direct from clinical samples. DNA baits spanning the length of the *C. trachomatis* genome were compiled by SureDesign and synthesized by SureSelect^{XT} (Agilent Technologies, UK). *C. trachomatis* DNA extract from clinical samples was quantified and carrier human genomic DNA added to obtain a total of 3 μg input for library preparation. DNA was sheared using a Covaris E210 acoustic focusing unit. End-repair, non-templated addition of 3' –A adapter ligation, hybridisation, enrichment PCR and all post-reaction clean-up steps were performed according to the SureSelect^{XT} Illumina Paired-End Sequencing Library protocol (V1.4.1 Sept 2012). All recommended quality control measures were performed between steps.

Whole genome sequencing and sequence quality filtering

DNA was sequenced by the Wellcome Trust Sanger Institute using Illumina paired-end technology (Illumina GAII or HiSeq 2000). Sequences were aligned to the *Chlamydia trachomatis A/HAR-13* reference genome (GenBank Accession Number NC_007429.1) and plasmid (GenBank Accession Number NC_007430.1) using BWA (Langmead 2009). SAMTOOLS/BCFtools (SAMTOOLS) (Li 2009) and GATK (McKenna 2010) were used to call SNPs and small indels. Variants were selected as the intersection dataset between those obtained using both SNP callers and SNPs were further quality-filtered by mappability and uniqueness (Coll 2014). Non-unique SNPs with mappability values greater than one were filtered and SNP alleles were called using an alternative coverage-based approach where a missing call was assigned to a site if the total coverage was less than 20x depth or where one of the four nucleotides accounted for at least 80% total coverage (Coll 2014). There was a clear relationship between the average depth of coverage and proportion of missing calls, which provided a further quality filtering threshold. Based on this we retained all sequences with greater than 10x average depth of coverage over the whole genome.

Comparative genomics analyses

The best-scoring maximum likelihood phylogenetic trees were constructed using RAxML v7.4.2 (Stamatakis 2008) using all SNP sites spanning the genome. Recombination is known to occur in *C. trachomatis* (Jeffrey 2010, Harris 2012) and can be problematic in constructing phylogeny. We applied three compatibility-based recombination detection methods to detect recombination hotspots using PhiPack (Bruen 2006): the pairwise homoplasy index (Phi) (Bruen 2006), the maximum Chi2 (Smith 1992) and the neighbour similarity score (NSS) (Jakobsen 1996) across the genome alignment. We also examined the confidence in the phylogenetic tree by computing RAxML site-based likelihood scores (Stamatakis 2008). Final phylogenetic trees were constructed adjusting for recombination (Stamatakis 2008). Tajima's D was calculated to assess DNA polymorphism in the context of the neutral mutation hypothesis (Tajima 1989).

Genome wide association scan analyses

A genome wide association scan (GWAS) is a powerful tool to assess the excess association between a SNP and defined phenotype. In general this method is confounded by population structure and linkage disequilibrium (LD) (Thomas 2011). Therefore we used principal component analysis (PCA) to elucidate the population structure and included principal components to adjust for population structure. We also corrected for genomic inflation for the occurrence of a polymorphism in the population of over 90% and a minor allele frequency of 5% in binary and linear models (3% in ordinal models). To avoid the occurrence of false positive associations a conservative Bonferroni correction was applied for to adjust for multiple testing in genome-wide significance analysis.

SNP mutations were aggregated by coding and intergenic regions in an $n \times m$ matrix, where n=number of loci and m=number of sequences. In coding regions only non-synonymous mutations were aggregated.

Regression analyses were performed adjusting for age, population structure and LD, using the first three principal components. Logistic regression models were used to assess the association between disease phenotype (binary models for active trachoma and ordinal models for conjunctival inflammation) and SNPs across the *C. trachomatis* genome using MASS and glmm packages in the R statistical package v3.0.2 (The R Foundation for Statistical Computing, http://www.r-project.org). Linear regression models with *C. trachomatis* load as a continuous outcome were used to

assess the association between *C. trachomatis* ocular load and SNPs across the genome. The statistical significance of each variable was tested using the Wald test. The standard error for the log odds ratio was estimated from the model and used to calculate 95% confidence intervals for odds ratios (OR).

RESULTS

From a total of 184 *C. trachomatis* infections quantifiable by ddPCR, we were able to obtain whole genome sequence data for 126 using SureSelect^{XT}. We retained 83 high quality sequences for the final analyses following application of stringent filtering criteria. There were no differences between the 83 retained and the 43 excluded with respect to demographic characteristics, bacterial load, disease severity scores or geographical location.

Phylogenetic Reconstruction and Population Structure

This collection of population-based ocular *C. trachomatis* whole genome sequences reveals substantial genetic diversity compared to the global chlamydial diversity previously published (Harris 2012). Tajima's D for the sequence data set was -2.50 ($p=0.01$) suggesting an excess of low frequency polymorphisms indicating population expansion and/or purifying selection. After stringent filtering, we found 759 and 10 unique non-synonymous SNP sites in non-repetitive coding regions of the genome and plasmid respectively, relative to the reference strain *C. trachomatis A/HAR-13*. Recombination was extensive and hotspots across the genome are shown in *Figure 1(A)*. It was not possible to remove all recombinant SNPs for phylogenetic reconstruction, but removal of SNPs in regions were there was most likely to be greatest recombination (using Phi) had no effect on phylogenetic relationships. Additionally, a site-wise log likelihood plot demonstrated that there was no significant lack of confidence in the tree construction due to recombination (*Figure 1(B)*).

Figure 2 shows radial cladograms for the plasmid and genome for all sequences in the final data set with published *C. trachomatis* reference strains, rooted on *Chlamydia muridarum Nigg* (GenBank Accession Numbers NC_002182.1 (plasmid) and NC_002620.1 (genome)). *Figure 3* shows the genomic diversity of these data in the context of *ompA* genotyping. There does appear to be clustering by serovar (A or B) but SNP phylogeny demonstrates greater diversity than that defined by *ompA* genotyping that may be useful in strain classification. Furthermore, whole genome sequence typing has shown that despite conforming to ocular *ompA*

genotypes, there are a number of sequences whose genomes cluster with urogenital strains (*Figure 2(Aii)*). Principal component analysis demonstrated that the first three principal components captured the majority of this structural variance, and were included in the final GWAS (*Figure 4*). Polymorphic loci on the *C. trachomatis* genome are shown in *Figure 5*.

Identifying genes and loci associated with disease severity

We conducted two GWAS analyses on the 759 unique non-synonymous SNPs in coding areas found in the 83 ocular *C. trachomatis* sequences (*Figure 6*). A critical p-value threshold was defined as 0.05/759 = 6.25xe-05. Using conjunctival inflammatory phenotype (P score) in ordinal logistic regression analysis we identified a single gene (*incE*) that was strongly associated with papillary hypertrophy (Odds Ratio (OR)=2.51, *p=5.16xe-08*) (*Figure 6(B)*). This finding was supported by the presence of other synonymous and non-synonymous SNPs within *incE* and at adjacent positions on the chromosome (*Figure 7*). We also identified a negative association between level of conjunctival inflammation and the locus *CTA_0165* (OR=0.44, *p=1.66xe-12*). Further genes and loci were associated with clinical phenotype but did not reach this threshold for genome-wide significance. Additional genes and loci that reached significance at the 1% level are detailed in *Table 1*.

tolB and *rpoA* were also strongly associated with conjunctival inflammation but did not reach genome-wide significance (*p=0.0017* and *p=0.0062* respectively) (*Table 1*).

Identifying genes and loci associated with C. trachomatis ocular load

On the same set of whole genome sequences we conducted a linear GWAS analysis with *C. trachomatis* load (log-(e) *omcB* copies/swab) as the outcome (*Figure 8*). We found that *mutY* and *CTA_0271* were strongly associated with *C. trachomatis* load (*p=2.10xe-05* and *p=4.68xe-05* respectively). *pmpE*, *pmpF* and *pmpB* were strongly associated, almost reaching genome-wide significance (*p=0.0002*, *p=0.0004* and *p=0.0022* respectively). Additionally, association with *tyrP* and *tarP* were approaching significance (*p=0.0036* and *p=0.0037* respectively) (*Table 1*). Other genes and loci of interest that did not reach genome-wide significance are included at the 1% level in *Table 1*. The same GWAS analyses were conducted for the plasmid but no associations were found.

DISCUSSION

This is the first population-based GWAS analysis conducted for *C. trachomatis*. We have used high quality sequence data obtained direct from clinical samples in conjunction with an objective detailed clinical phenotype and accurate estimation of *C. trachomatis* ocular load. We investigated *C. trachomatis* genotypic associations with disease severity and bacterial load in a treatment-naïve population-based sample collection.

In this population *incE* had the strongest positive association with conjunctival inflammation. There were non-synonymous SNPs up and downstream to *incE* (in genes *clpB*, *araD*, *efp_1* and loci *CTA_0121* and *CTA_0127*) and within the same operon (*incD*, *incF* and *incG*), however these did not reach genome-wide significance. Further analysis of SNPs within the operon (presented as a regional view in *Figure 7*) showed a frequent non-synonymous SNP at position 136633 (TG) surrounded by other supporting synonymous SNPs forming a peak in the regional manhattan plot.

IncE is an inclusion membrane protein. There has been increasing interest in the role of this family of proteins since the characterization of *incD-G*. A study using RT-PCR showed that these four inclusion membrane proteins (IncD - IncG) are co-transcribed within the first two hours of internalization of *C. trachomatis*, suggesting that they are required for modification of the nascent inclusion in early chlamydial infection (Scidmore-Carlson 1999). Soon after development of inclusions, homotypic fusion occurs, allowing the fused inclusions to escape degradation through host endolysosomal pathways and promote fusion with each other to sustain higher bacterial loads, thus promoting infection. The role of the Inc family in naturally occurring fusing and non-fusing variants has been described (Suchland 2000). Serotyping studies showed that strains lacking the IncA serotype were non-fusogenic. Subsequent genotyping of *incA* showed that there were two non-synonymous SNPs within *incA* rendering it non-fusogenic (Suchland 2000, Fields 2002). Others reported that non-fusogenic strains unable to express *incA* had a greater tendency to cause subclinical infection in natural human genital tract infection (Giesler 2001).

Inclusion membrane proteins are type III effector proteins, which are secreted from within the parasitophorous vesicle and insert in the inclusion membrane. They are exposed on the cytosolic face of the inclusion where they can interact with host cells. They are expressed as a family and individually have an inherent property to

form vesicular structures. They appear to be important in inclusion membrane structure and biogenesis and are thought to play a role in virulence and pathogenesis (Carlson 2005, Mital 2013).

The locus *CTA_0165* was negatively associated with conjunctival inflammation at genome-wide significance level. This locus codes for a phosphatidylcholine-hydrolysing phospoholipase D, which is part of the chlamydial phospholipase D (PLD) family. The PLD loci and chlamydial cytotoxin are located within the plasticity zone (PZ). This is one of the most polymorphic regions in the genome. In the current study, other PLD loci were also highly polymorphic (*CTA_0163*, *CTA_0164* and *CTA_0167*) but did not reach genome-wide significance in the association analysis. The PZ PLDs are transcribed during normal and persistent infection and corresponding PLD proteins are located on the inclusion membrane of reticulate bodies. There is some evidence from *in vitro* studies that these PLDs may be involved in reactivation of *C. trachomatis* from a gamma interferon-induced persistent state (Nelson 2006). The functions of the PZ PLDs are not well defined, but they may be important, strain-specific pathogenic factors *in vivo* (Nelson 2006, Taylor 2010).

tolB and *rpoA* were approaching genome-wide significance in their association with conjunctival inflammation. *tolB* encodes a translocation protein and is located close to *ompA*. In *Pseudomonas aeruginosa* TolB is a periplasmic component of the TolB-Pal trans-envelope protein complex of Gram-negative bacteria. TolB is essential for *P. aeruginosa* growth *in vitro* and *in vivo* and TolB-deplete cells are defective in cell envelope integrity, resistance to human serum and some antibiotics, and in the ability to establish infection (Lo Sciuto 2014). There may be a role for TolB in growth and pathogenicity. *rpoA* has been shown to be strongly upregulated during chlamydial persistence in *Chlamydia pneumoniae* infections (Bunk 2010). RpoA, along with certain *pmp*, *ompA, porB* and *groEL*-encoded antigens, was suggested as an immunodominant marker in persistent *C. pneumoniae* infections (Bunk 2008).

mutY was associated with *C. trachomatis* bacterial load at genome-wide significance. *mutY* is described extensively in studies on *Escherichia coli* as a mutator locus involved in GC to TA transversions (Nghiem 1988). *mutY* is normally involved in mismatch corrections, removing adenines from mismatch situations and maintaining genomic stability (Manuel 2004). *E. coli* carrying *mutY* mutations lack the mismatch correction glycosylase. It is also thought to promote tryptophan-

independent mutants, with small in-frame deletions and transversions in the *trpA23* operon in other chlamydial species (Bridges 1997). The presence of *mutY* mutations has been documented in a hypervirulent *C. trachomatis* strain causing LGV (Somboonna 2011).

CTA_0271 encodes a 12kDa hypothetical membrane-associated protein of 116 amino acids in length from position 281,737-282,087. It is flanked by *glgP* (*CTA_0270*) and *yidC* (*CTA_0273*). Though neither of these loci reached genome-wide significance, an association was demonstated with *C. trachomatis* ocular load. *glgP* is involved in carbohydrate and glycogen metabolism and encodes GlP, an antigen encoded on the chlamydial genome but transcriptionally regulated by the plasmid (Li 2012). In mice, these antigens are thought to be associated with *C. muridarum* pathogenesis. Intravaginal GlP in mouse models produced protective immunity in terms of decreased vaginal shedding of *C. muridarum* 14 days after introduction (Li 2012). There was good correlation of protection with *C. muridarum*-specific antibody and a Th1 dominant T cell response, suggesting that GlP may be considered as part of a multi-subunit vaccine (Li 2012).

yidC is thought to be required for the maintenance of the translocation-competent state of certain autotransporters in the periplasm in *E. coli* (Jong 2010). Autotransporters are an important family of virulence factors secreted by Gram-negative bacteria. They translocate across the membrane and are eventually secreted into the extracellular environment. The role of two *yidC* insertases were examined in *Streptococcus mutans*, showing that *yidC* contributes to cell surface biogenesis and protein secretion in *S. mutans* (Palmer 2012).

Although below the genome-wide significance threshold, *pmp E, pmpF, pmpB, tyrP* and *tarP* were also associated with *C. trachomatis* load. The Pmp family have been linked to chlamydial virulence previously (Stothard 2003, Gomes 2006). They are surface-expressed, exhibit autotransporter characteristics and are involved in transmembrane transport as type III secretion effectors (Tan 2006, Swanson 2009). Specifically *pmpE* and *pmpF* are highly variable and are detected on the chlamydial elementary body surface (Tan 2006). PmpB (and PmpD and H) are strongly immunogenic and elicit a pro-inflammatory cytokine response (Tan 2006). They are strong candidates for a putative role in virulence via modulation of inflammation and adherence to host cells. In mouse models, PmpE-H have been identified as CD4 T cell vaccine candidates (Yu 2014). In *C. muridarum* genital tract infection, CD4 T cells

are the foundation for protective immunity, with antibodies playing a secondary role. In mice, each of these proteins was found to individually confer protection, measured by accelerated clearance of *C. muridarum* infection from the murine genital tract (Yu 2014). Combining multiple outer membrane components (PmpE-H + the major outer membrane protein (MOMP)) provided broader immunogenicity against chlamydial infection (Yu 2014). In our analysis, *ompA* (encoding MOMP) was polymorphic but not associated with inflammation or ocular *C. trachomatis* load.

tyrP is linked to tyrosine phosphorylation, occurring early in pathogen-host interactions (Clausen 1997). This species-specific mechanism of pathogen entry to host cells is thought to involve specific ligands and host cell co-receptors. The phenomenon was observed in *C. trachomatis* cell culture using McCoy cells and is hypothesized as a species-specific mechanism in the pathogenesis of chlamydial infection (Clausen 1997).

tarP is highly polymorphic between chlamydia strains (Carlson 2005, Thomson 2008). TarP is thought to be involved in pathogen-directed phagocytosis during the initial attachment phase of uptake of the chlamydial elementary body (Jewett 2006). Intra-serovar recombination is thought to play a role in the diversity of these genes (Gomes 2007, Harris 2012). Given the diversity in these genes between pathobiotypes, associations with ocular *C. trachomatis* load are not surprising and they may be associated with chlamydial virulence.

Although there were a number of polymorphic genes and loci identified, they did not reach the stringent genome-wide significance level stipulated using GWAS. This GWAS approach has been shown to be a robust approach for discovering novel loci in other pathogens (Laabei 2014, F Coll [unpublished data]). The current study has identified strong associations with putative virulence genes and loci, contextualizing these associations in currently circulating *C. trachomatis* strains. The data are validated in the findings of putative disease severity associations with *inc, pmp* and PLD genes. The chlamydial GWAS provides compelling evidence that these are important and relevant associations, and we have been able to associate them with detailed clinical phenotype and an accurate estimate of *C. trachomatis* load *in vivo*. We have also found associations with novel loci that are linked to virulence in other Gram-negative bacteria (*mutY, tolB, rpoA* and *yidC*) and a locus for a hypothetical membrane protein whose function is currently unclear (*CTA_0271*), suggesting these as putative virulence factors in *C. trachomatis*. Additionally, the results from this

study support findings of virulence-associated genes identified in murine models (*glgP* and *pmpE-H*).

Interestingly, there were no genome-wide associations with disease severity or *C. trachomatis* load found in genes or loci on the plasmid. However there were associations with genes on the chromosome that are thought to be regulated by plasmid-encoded genes (such as *glgP*). It is clear that the plasmid has an important role in chlamydial pathogenesis, but the mechanisms behind this remain unclear.

We were able to obtain high quality whole genome sequence data from samples with loads above 1000 *omcB* copies/swab (126/184 (68%) samples with quantifiable *C. trachomatis* DNA by ddPCR) from this population-based study. We further applied stringent quality control filters on WGS data obtained and retained 83 samples in the final analysis. Despite these limitations, we have generated a large archive of ocular *C. trachomatis* WGS data with corresponding detailed clinical and epidemiological metadata. These data are unique in their completeness in the global collection of *C. trachomatis* WGS data.

These samples were collected from a spatially representative population-based survey in treatment-naïve trachoma-endemic communities, thus offering further opportunity for association analysis at genome scale resolution with spatial phylodynamics. We identified a number of genes associated with ocular clinical phenotype and *C. trachomatis* bacterial load from this trachoma-endemic population in West Africa and intend to test these associations in other populations and replication cohorts. Through validating these findings in other populations we may be able to address limitations of the statistical power of GWAS analysis in which it is necessary to adjust for population structure and LD (Type I error) by using PCA (which may increase Type II error). We have attempted to counteract this by using a conservative p-value threshold for genome wide significance by applying a Bonferroni correction for multiple comparisons.

The associations with conjunctival inflammation and *C. trachomatis* load detailed in this study highlight genes involved with specific biological characteristics of *C. trachomatis* and its early interactions with host cells. These data inform us about these early interactions that are likely to be important in *C. trachomatis* pathogenesis and may also be important putative vaccine candidates. Pathogen GWAS, particularly in conjunction with a detailed *in vivo* phenotype and accurate estimation of *C. trachomatis* load, is a powerful approach to identifying multiple targets for further

study in pathogenesis and directed study of potential vaccine candidates, allowing a greater understanding of association and interaction of genes on a genome-wide scale.

DATA ACCESS

All sequence data are available from the Wellcome Trust Sanger Institute (ftp://ftp.sanger.ac.uk/pub/project/pathogens/Chlamydia/LSHTM)

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust (grant number 097330/Z/11/Z). We extend thanks to colleagues at the Programa Nacional de Saúde de Visão in the Ministério de Saúde Publica in Bissau, the study participants and dedicated field research team in Guinea Bissau. All communities received treatment for endemic trachoma in accordance with WHO and national policies following the survey.

DISCLOSURE DECLARATION

All authors declare no conflicts of interest.

REFERENCES

ANDREASEN AA, Burton MJ, Holland MJ et al. (2008). *Chlamydia trachomatis* ompA variants in trachoma: what do they tell us? PLoS Negl Trop Dis 2(9):e306.

BAILEY RL, Hayes L, Pickett M et al. (1994). Molecular epidemiology of trachoma in a Gambian village. Br J Ophthalmol 78:813-17.

BRIDGES BA, Timms AR (1997). Mutation in *Escherichia coli* under starvation conditions: A new pathway leading to small deletions in strains defective in mismatch correction. EMBO J 16(11):3349-56.

BRUEN TC, Philippe H, Bryant D (2006). A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665-2681.

BUNK S, Susnea I, Rupp J et al. (2008). Immunoproteomic identification and serological responses to novel *Chlamydia pneumoniae* antigens that are associated with persistent *C. pneumoniae* infections. J Immunol 180(8):5490-8.

BUNK S, Schaffert H, Schmid B et al. (2010). *Chlamydia pneumoniae*-induced memory CD4 T cell activation in human peripheral blood correlates with distinct antibody response patterns. Clin Vaccine Immunol 17(5):705-12.

BURTON MJ, Holland MJ, Jeffries D et al. (2006). Conjunctival Chlamydial 16S Ribosomal RNA Expression in Trachoma: Is Chlamydial Metabolic Activity Required for Disease to Develop? Clin Infect Dis 42:463-70.

CALDWELL HD, Wood H, Crane D et al. (2003). Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiates between genital and ocular isolates: Implications in pathogenesis and infection tropism. J Clin Invest 111:1757-69.

CARLSON JH, Hughes S, Hogan D (2004). Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. Infect Immun 72(12):7063-72.

CARLSON JH, Porcella SF, McClarty G (2005). Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and geniotropic strains. Infect Immun 73(10):6407-18.

CARLSON JH, Wood H, Roshick C et al. (2006). *In vivo* and *in vitro* studies of *Chlamydia trachomatis* TrpR:DNA interaction. Mol Microbiol 59(6):1678-91.

CARLSON JH, Whitmire WM, Crane DD et al. (2008). The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect Immun 76:2273-83.

CLAUSEN JD, Christiansen G, Holst HU, Birkelund S (1997). *Chlamydia trachomatis* uses host cell microtubule networks during early events of infection. Mol Microbiol 25(3):441-9.

COLL F, Preston M, Guerra-Assuncao JA et al. (2014). PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) 94(3):346-54.

DAWSON CR, Jones BR, Tarizzo ML (1981). Guide to trachoma control in programs for the prevention of blindness. Geneva: World Health Organization.

FAAL N, Bailey RL, Jeffries D et al. (2006). Conjunctival FOXP3 expression in trachoma: Do regulatory T cells have a role in human ocular *Chlamydia trachomatis* infection? PLoS Med 3(8):e266.

FIELDS KA, Fischer E, Hackstadt T (2002). Inhibition of fusion of *Chlamydia trachomatis* inclusions at 32 degrees C correlates with restricted export of IncA. Infect Immun 70(7):3816-23.

FRAZER LC, Darville T, Chandra-Kuntal K et al. (2012). Plasmid-cured *Chlamydia caviae* activates TLR2-dependent signaling and retains virulence in the guinea pig model of genital tract infection. PLoS One 7(1):e30747.

GIESLER WM, Suchland RH, Rockey DD, Stamm WE (2001). Epidemiology and clinical manifestations of unique *Chlamydia trachomatis* isolates that occupy nonfusogenic inclusions. J Infect Dis 184:879-884.

GOMES JP, Nunes A, Bruno WJ et al. (2006). 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 188:275-286.

GOMES JP, Bruno WJ, Nunes A (2007). Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res 17(1):50-60.

HACKSTADT T, Scidmore-Carlson MA, Shaw EI, Fischer ER (1999). *Chlamydia trachomatis* IncA protein is required for homotypic vesicle fusion. Cell Microbiol 1:119-130.

HARRIS SR, Clarke IN, Seth-Smith HMB et al. (2012). Whole genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nature Genet 44(4):413-s1.

HAYES LJ, Bailey RL, Mabey DC et al. (1992). Genotyping of *Chlamydia trachomatis* from a trachoma-endemic village in The Gambia by a nested polymerase chain reaction: identification of strain variants. J Infect Dis 166(5):1173-7.

HEFTY SP, Stephens RS (2007). *Chlamydia trachomatis* is encoded on ten operons preceded by a sigma 70-like promoter element. J Bacteriol 189:198-206.

HENDERSON IR, Lam AC (2001). Polymorphic proteins of chlamydia spp. – Autotransporters beyond the Proteobacteria. Trends Microbiol 9(12):573-8.

HSIA R-C, Pannekoek Y, Ingerowski E et al. (1997). Type III secretion genes identify a putative virulence locus of chalmydia. Mol Microbiol 25:351-359.

HU VH, Harding-Esch EM, Burton MJ et al. (2010). Epidemiology and control of trachoma: systematic review. Trop Med Int Health 15(6):673-91.

JAKOBSEN IB, Easteal S (1996). A program for calculating and displaying compatibility matices as an aid in determining reticulate evolution in molecular sequences. Comput Appl Biosci 12(4):291-5.

JEFFREY BM, Suchland RJ, Quinn KL et al. (2010). Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun 78:2544-2553.

JEWETT TJ, Fischer ER, Mead DJ (2006). Chlamydial TARP is a bacterial nucleator of actin. Pro Natl Acad Sci USA 103(42):15599-604.

JONG WS, ten Hagen-Jongman CM, Ruijter E et al. (2010). *YidC* is involved in the biogenesis of the secreted autotransporter haemoglobin protease. J Biol Chem 285(51):39682-90.

KARI L, Whitmire WM, Carlson JH et al. (2008). Pathogenic diversity among *Chlamydia trachomatis* ocular strains in non-human primates is affected by subtle genomic variations. J Infect Dis 197:449-56.

KOL A, Lichtman AH, Finberg RW et al. (2000). Cutting edge: Heat shock protein (HSP) 60 activates the innate immune response: CD14 is an essential receptor for HSP60 activation of mononuclear cells. J Immunol 164(1):13-17.

LAABEI M, Recker M, Rudkin JK et al. (2014). Predicting the virulence of MRSA from its genome sequence. Genome Res 24(5):839-49.

LANGMEAD B, Trapnell C, Pop M et al. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.3.

LAST AR, Burr SE, Alexander ND, Harding-Esch E, Roberts Ch et al. (2014a). Spatial clustering of high load ocular *Chlamydia trachomatis* infection and disease

severity in trachoma-endemic communities on the Bijagós Archipelago, Guinea Bissau. Emerg Infect Dis [under review].

LAST AR, Burr SE, Weiss HA et al. (2014b). Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau. PLoS Negl Trop Dis 8(6):e2900.

LAST A, Roberts CH, Cassama E et al. (2014c). Plasmid copy number and disease severity in naturally occurring ocular *Chlamydia trachomatis* infection. J Clin Microbiol 52(1):324-7.

LAVERDA D, Albanese LN, Ruther PE et al. (2000). Seroactivity to *Chlamydia trahcomatis* HSP10 correlates with disease severity in women. Infect Immun 68:303-9.

LI H, Handsaker B, Wysoker A et al. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

LI Z, Lu C, Peng B et al. (2012). Induction of protective immunity against *Chlamydia muridarum* intravaginal infection with a chlamydial glycogen phosphorylase. PLoS ONE 7(3):e32997.

LO SCIUTO A, Fernandez Pinar R, Bertuccini L et al. (2014). The periplasmic protein TolB as a potential drug target in *Pseudomonas aeruginosa*. PLoS ONE 9(8):e103784.

LONGBOTTOM D, Russell M, Dunbar SM et al. (1998). Molecular cloning and characterization of genes coding for the highly immunogenic cluster of 90 kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. Infect Immun 66:1317-1324.

MABEY DC, Solomon AW, Foster A (2003). Trachoma. *Lancet* 362:223-229.

MANUEL RC, Hitomi K, Arvai AS et al. (2004). Reaction intermediates in the catalytic mechanism of Escherichia coli MutY DNA glycosylase. J Biol Chem 279(45):4693-9.

McKENNA A, Hanna M, Banks E et al. (2010). The Genome Analysis Tool Kit: A MapReduce framework for analysing next generation DNA sequencing data. Genome Res 20:1297-1303.

MITAL J, Millar NJ, Dorward DW et al. (2013). Roles for chlamydial inclusion membrane proteins in inclusion membrane structure and biogenesis. PLoS ONE 8(5):e63426.

NELSON DE, Crane DD, Taylor LD et al. (2006). Inhibition of chlamydiae by primary alcohols correlates with strain specific complement of plasticity zone phospholipase D genes. Infect Immun 74(1):73-80.

NGHIEM Y, Cabrera M, Cupples CG, Miller JH (1988). The *mutY* gene: A mutator locus in *Escherichia coli* that generates GC ----- TA transversions. Proc Natl Acad Sci USA 85(8):2709-13.

NUNES A, Borrego MJ, Gomes JP (2013). Genomic features beyond *C. trachomatis* phenotypes: what do we think we know? Infect Genet Evol 16:392-400.

O'CONNELL CM, AbdelRahman YM, Green E et al. (2011). Toll-like receptor 2 activation by *Chlamydia trachomatis* is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by *C. trachomatis* but not by *C. muridarum*. Infect Immun 79:1044-1056.

PALMER SR, Crowley PJ, Oli MW et al. (2012). YidC1 and YidC2 are functionally distinct proteins involved in protein secretion, biofilm formation and cariogenicity of *Streptococcus mutans.* Microbiology 158(7):1702-12.

ROCKEY DD, Heinzen RA, Hackstadt T (1995). Cloning and characterization of a *Chlamydia psittaci* gene coding for a protein localized in the inclusion membrane of infected cells. Mol Microbiol 15:617-26.

ROBERTS Ch, Last A, Molina-Gonzalez S et al. (2013). Development and evaluation of a next generation digital PCR diagnostic assay for ocular *Chlamydia trachomatis* infections. J Clin Microbiol 51(7):2195-2203.

RODRIGUEZ P, Vekris A, de Barbeyrac B et al. (1991). Typing of *Chlamydia trachomatis* by restriction endonuclease analysis of the amplified major outer membrane protein gene. J Clin Microbiol 29:1132-1136.

SCIDMORE-CARLSON MA, Shaw EI, Dooley CA et al. (1999). Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. Mol Microbiol 33:753-765.

SMITH JM (1992). Analysing the mosaic structure of genes. J Mol Evol 34(2):126-9.

SOMBOONNA N, Wan R, Ojcius DM et al. (2011). Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L2) and D lineages. mBio 2(3):e00045-11.

SONG L, Carlson JH, Whitmire VM et al. (2013). *Chlamydia trachomatis* plasmid-encoded *pgp4* is a transcriptional regulator of virulence associated genes. Infect Immun 81(3):636.

STAMATAKIS A, Hoover P, Rougemont J (2008). A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 57:758-771.

STEPHENS RS, Kalman S, Lammel C et al. (1998). Genomic sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-9.

STOTHARD DR, Toth GA, Batteiger BE (2003). Polymorphic membrane protein H has evolved in parallel with the three disease-causing groups of *Chlamydia trachomatis*. Infect Immun 71:1200-1208.

SUCHLAND RJ, ROckey DD, Bannatine JP, Stamm WE (2000). Isolates of *Chlamydia trachomatis* that occupy nonfusogenic inclusions lack IncA, a protein localized to the inclusion membrane. Infect Immun 68:360-367.

SWANSON KA, Taylor LD, Frank SD et al. (2009) *Chlamydia trachomatis* polymorphic membrane protein D is an oligomeric autotransporter with a higher-order structure. Infect Immun 77:508-516.

TAJIMA F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585-93.

TAN C, Spitznagel JK, Shoe H-Z et al. (2006). The polymorphic membrane protein gene family of the chlamydiaciae. Chlamydia Genomics and Pathogenesis. Ed: Bavoil PM, Wyrick PB. Horizon Bioscience; Norfolk, UK. pp 195-218.

TAYLOR LD, Nelson DE, Dorward DW, Caldwell HD (2010). Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. Infect Immun 78(6):2691-9.

THOMAS A, Abel HJ, Di Y et al. (2011). Effect of linkage disequilibrium on the identification of functional variants. Genet Epidemiol 35(Suppl 1):S115-9.

THOMSON NR, Holden TF, Carder C et al. (2008). *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res 18(1):161-71.

THYLEFORS B, Dawson CR, Jones BR et al. (1987). A simple system for the assessment of trachoma and its complicatios. Bull World Health Organ 65(4):477-83.

WANG Y, Kahane S, Cutcliffe LT et al. (2011) Development of a transformation system for *Chlamydia trachomatis*: restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. PLoS Pathog 7(9):e1002258.

WANG Y, Cutcliffe LT, Skilton RJ et al. (2013). Transformation of a plasmid-free genital tract isolate with a plasmid vector carrying a deletion in CDS6 revealed that this gene regulates inclusion phenotype. Pathogens Dis 67(2):100-103.

WOOD H, Fehiner-Gardner C, Berry J et al. (2003). Regulation of tryptophan synthase gene expression in *Chlamydia trachomatis*. Mol Microbiol 49(5):1347-59.

YU H, Karunakoran KP, Jiang X, Brunham RC (2014). Evaluation of a multi-subunit recombinant Pmp and MOMP T cell vaccine against *Chlamydia muridarum* genital infection in three strains of mice. Vaccine 32(36):4672-80.

Figures

Figure Legends

Figure 1. Statistically significant hotspots of recombination using (A) the pairwise homoplasy index (Phi) (B) the site-wise log likelihood support for the best-scoring maximum likelihood tree are demonstrated along the *Chlamydia trachomatis* genome
Figure 2. (Ai) RAxML maximum likelihood phylogenetic tree including all sequences retained in the analysis after quality filtering (*n=83*, colour-coded blue). Radial phylogram representation of the best-scoring maximum likelihood tree using RAxML software. Sequences are shown with published *C. trachomatis* reference strains (black), mapped to reference strain *C. trachomatis A/HAR-13* (red) and rooted on *C. muridarum Nigg* (green). (Aii) Ocular *C. trachomatis* sequences (blue) clustering with urogenital strains (black). Radial cladograms produced representing the best-scoring maximum likelihood trees using RAxML software for (B) *C. trachomatis* genome (C) *C. trachomatis* plasmid.
Figure 3. RAxML maximum likelihood phylogenetic tree including all sequences retained in the analysis after quality filtering (*n=83*). Radial phylogram representation of the best-scoring maximum likelihood tree using RAxML software. Sequences are colour-coded by *ompA* genotype showing clustering of A (A/HAR-13 (red), A7249 (orange)) and B (B/JALI-20 (blue), B/TZ1A828 (purple)) ocular serovars..

Figure 4. The 83 ocular *C. trachomatis* sequences are plotted against principal components (PC) 1 – 4. Sequences are colour-coded by *ompA* genotype. PC 1-3 best capture this clustering by genotype, identifying the underlying population structure.
Figure 5. Highly polymorphic loci across the *C. trachomatis* genome are represented by SNP density per kb.
Figure 6. Locus GWAS for non-synonymous SNPs in coding regions for (A) Active trachoma (TF/TI) and (B) Conjunctival inflammation (P score; P0-3) in trachoma using binary and ordinal logistic regression models adjusting for age, population structure and linkage disequilibrium with Bonferroni correction for multiple testing (*p=5xe-05*). A $\log_{10}$ *p-value* of 2 is equivalent to a *p-value* of 0.01. Genes with an association significant at the 1% level are labeled.
Figure 7. Frequency of synonymous (black) and non-synonymous (red) SNPs shown across the *incE* and *incDEFG* operon (chromosomal position shown on x axis) in 83 isolates (y axis).
Figure 8. Locus GWAS for non-synonymous SNPs in coding regions for ocular *C. trachomatis* load (defined as log-(e) *omcB* copies/swab) linear regression models adjusting for age, population structure and linkage disequilibrium with Bonferroni correction for multiple testing (*p=6.25xe-05*). A $\log_{10}$ *p-value* of 2 is equivalent to a *p-value* of 0.01. Genes with an association significant at the 1% level are labeled.

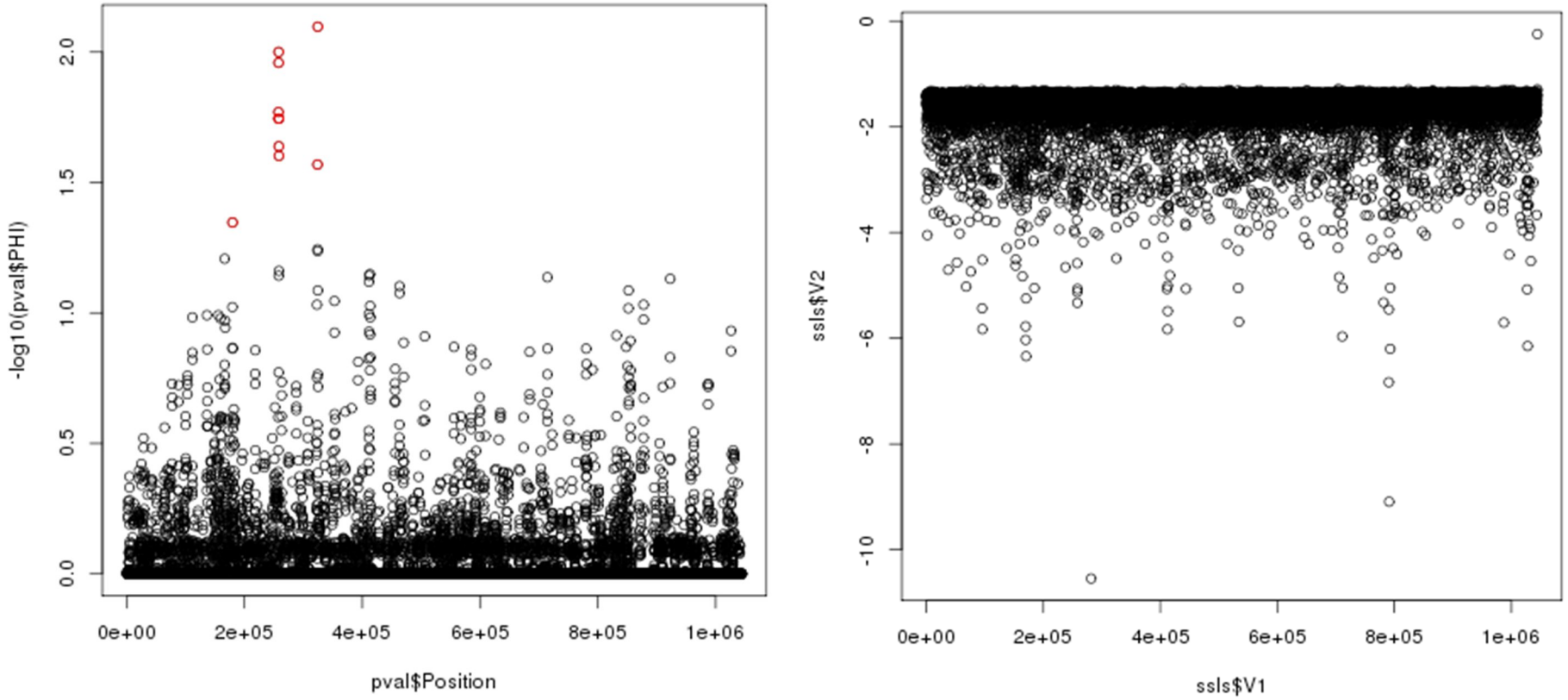
Tables

Table 1. Genes and loci associated with disease severity and ocular *Chlamydia trachomatis* load in trachoma

Table Legends

Table 1. Active trachoma is defined as TF/TI by the WHO simplified grading system (Thylefors 1981). Conjunctival inflammation is defined by a P score (0-3) using the modified FPC grading system (Dawson 1987). *C. trachomatis* ocular load is defined as the log-(e) *omcB* copies/swab. Genome-wide significance using a Bonferroni correction is defined as *p=6.25xe-05*. Represented here are associated genes and loci significant at the 1% level to account for false discovery rates. *Odds Ratio=0.03 suggesting a negative association.

Figure 1. Recombination hotspots on the *Chlamydia trachomatis* genome

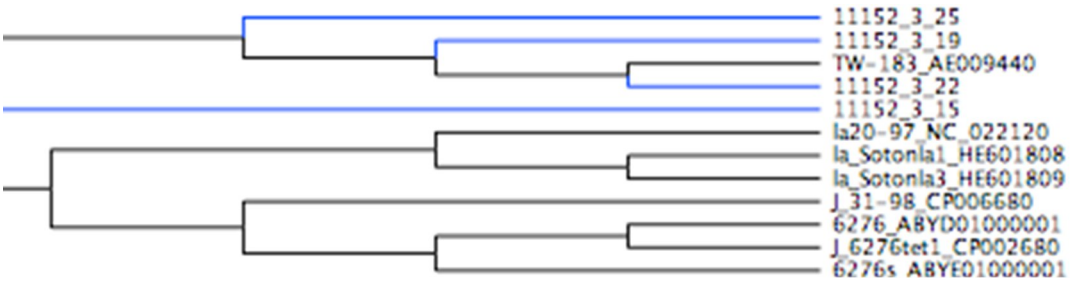


(A) (B)

Statistically significant hotspots of recombination using (A) the pairwise homoplasy index (Phi) and (B) the site-wise log likelihood support for the best-scoring maximum likelihood tree are demonstrated along the *Chlamydia trachomatis* genome.

Figure 2. Maximum likelihood phylogenetic trees showing population structure of ocular *Chlamydia trachomatis* sequences included in the analysis

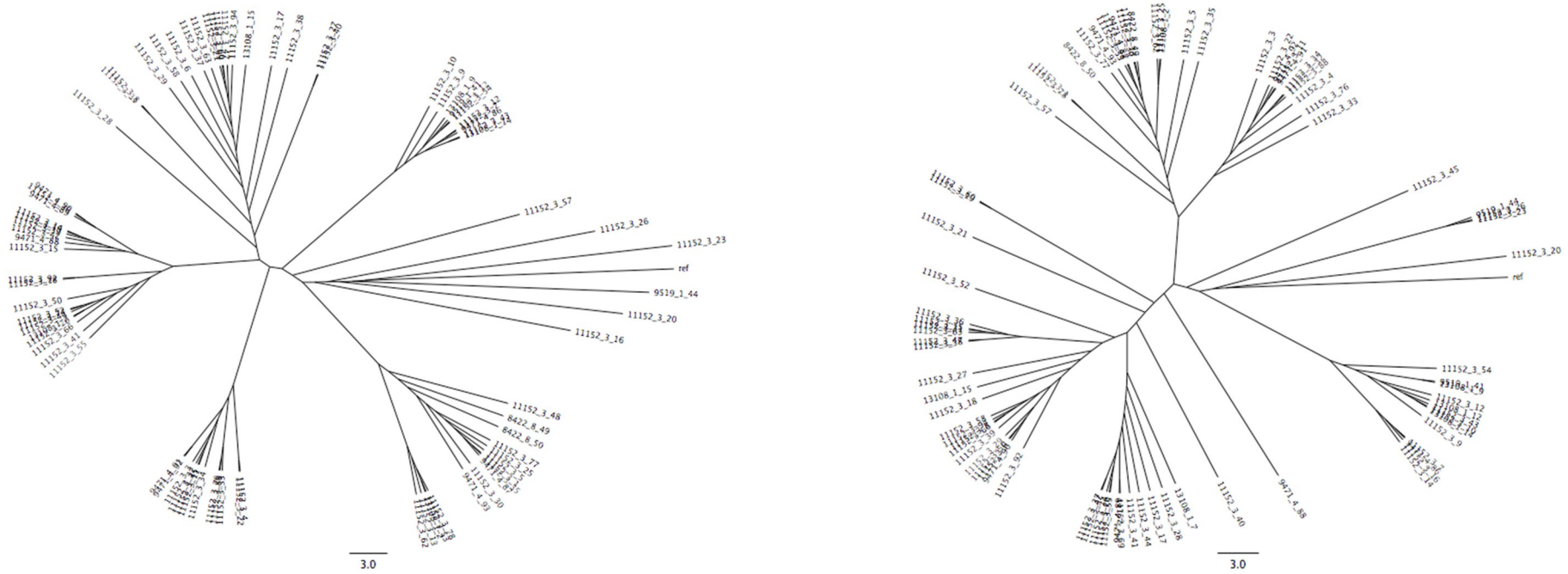
(Ai)



(Aii)

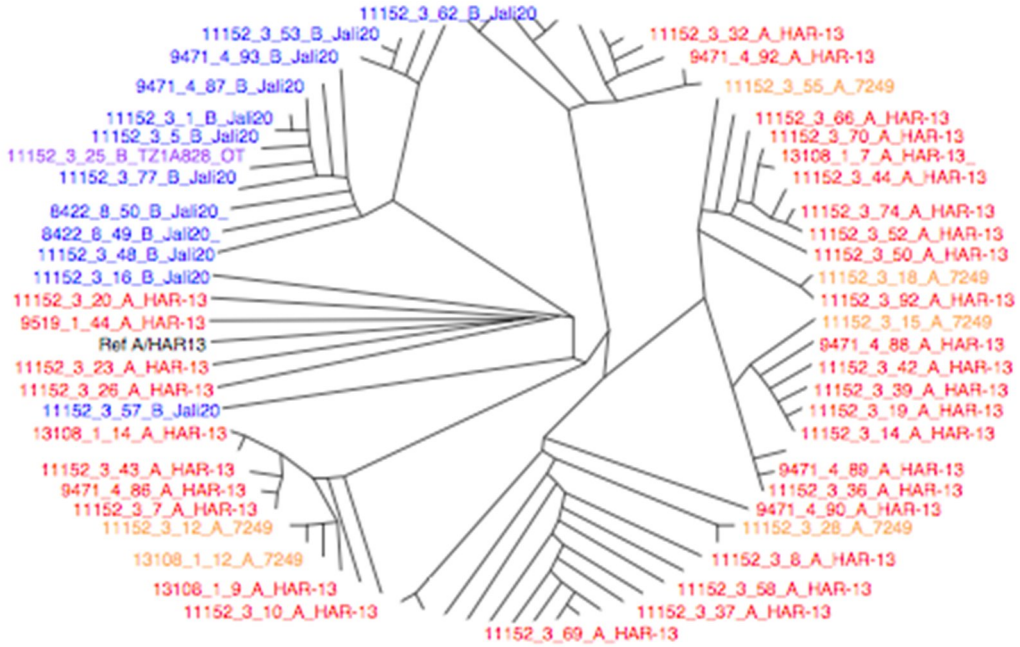
(B)

(C)



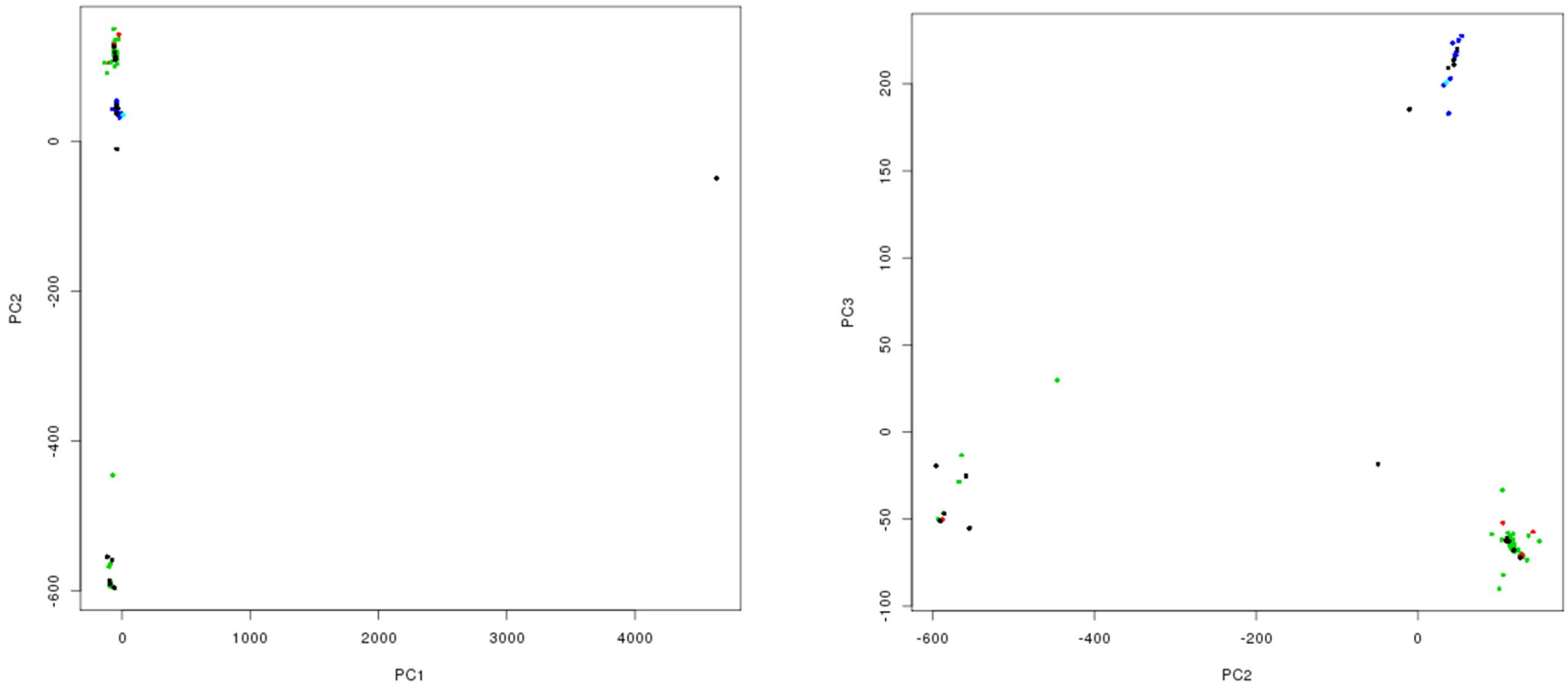
(Ai) RAxML maximum likelihood phylogenetic tree including all sequences retained in the analysis after quality filtering (*n=83*, colour-coded blue). Radial phylogram representation of the best-scoring maximum likelihood tree using RAxML software. Sequences are shown with published *C. trachomatis* reference strains (black), mapped to reference strain *C. trachomatis A/HAR-13* (red) and rooted on *C. muridarum Nigg* (green). (Aii) Ocular *C. trachomatis* sequences (blue) clustering with urogenital strains (black). Radial cladograms produced representing the best-scoring maximum likelihood trees using RAxML software for (B) *C. trachomatis* genome (C) *C. trachomatis* plasmid.

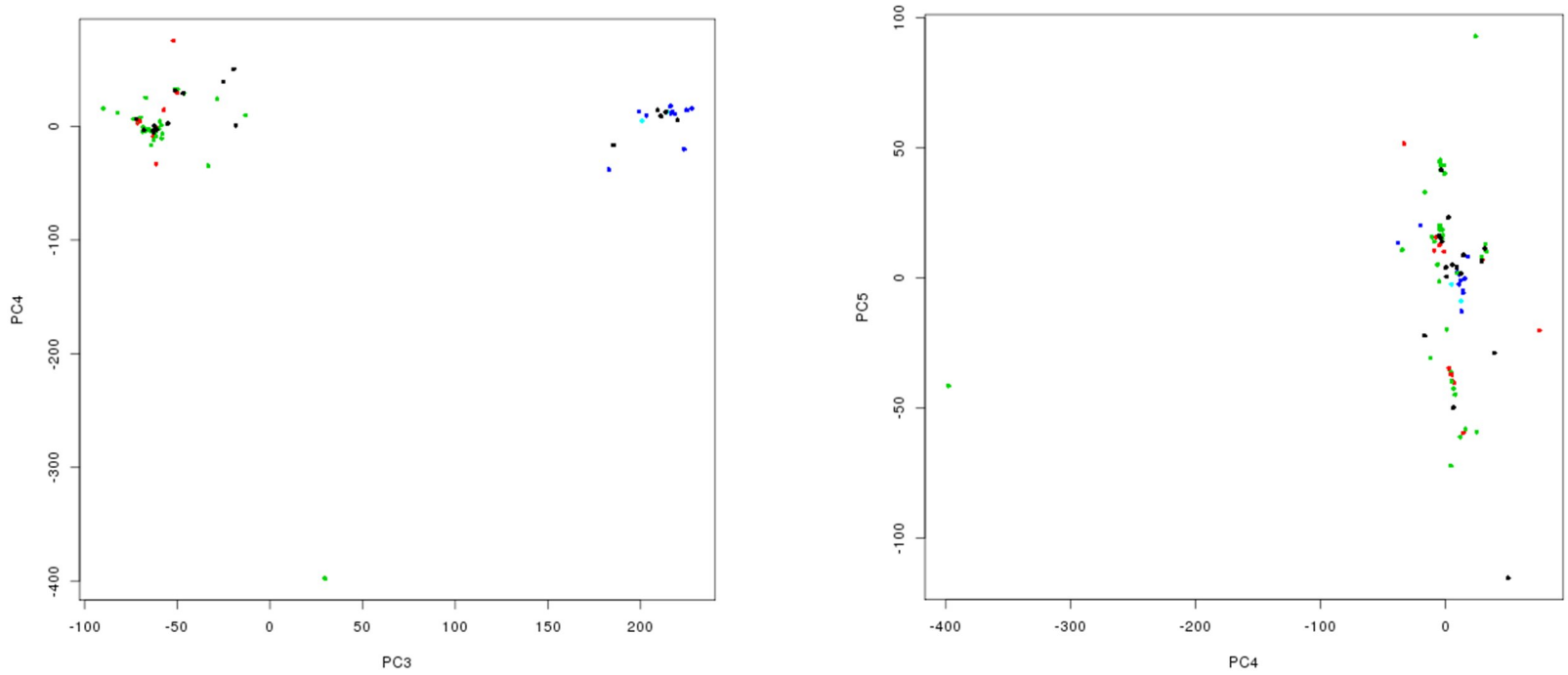
Figure 3. Maximum likelihood phylogenetic tree showing serovar clustering of ocular *Chlamydia trachomatis*



RAxML maximum likelihood phylogenetic tree including all sequences retained in the analysis after quality filtering (*n=83*). Radial cladogram representation of the best-scoring maximum likelihood tree using RAxML software. Sequences are colour-coded by *ompA* genotype showing clustering of A (A/HAR-13 (red), A7249 (orange)) and B (B/JALI-20 (blue), B/TZ1A828 (purple)) ocular serovars.

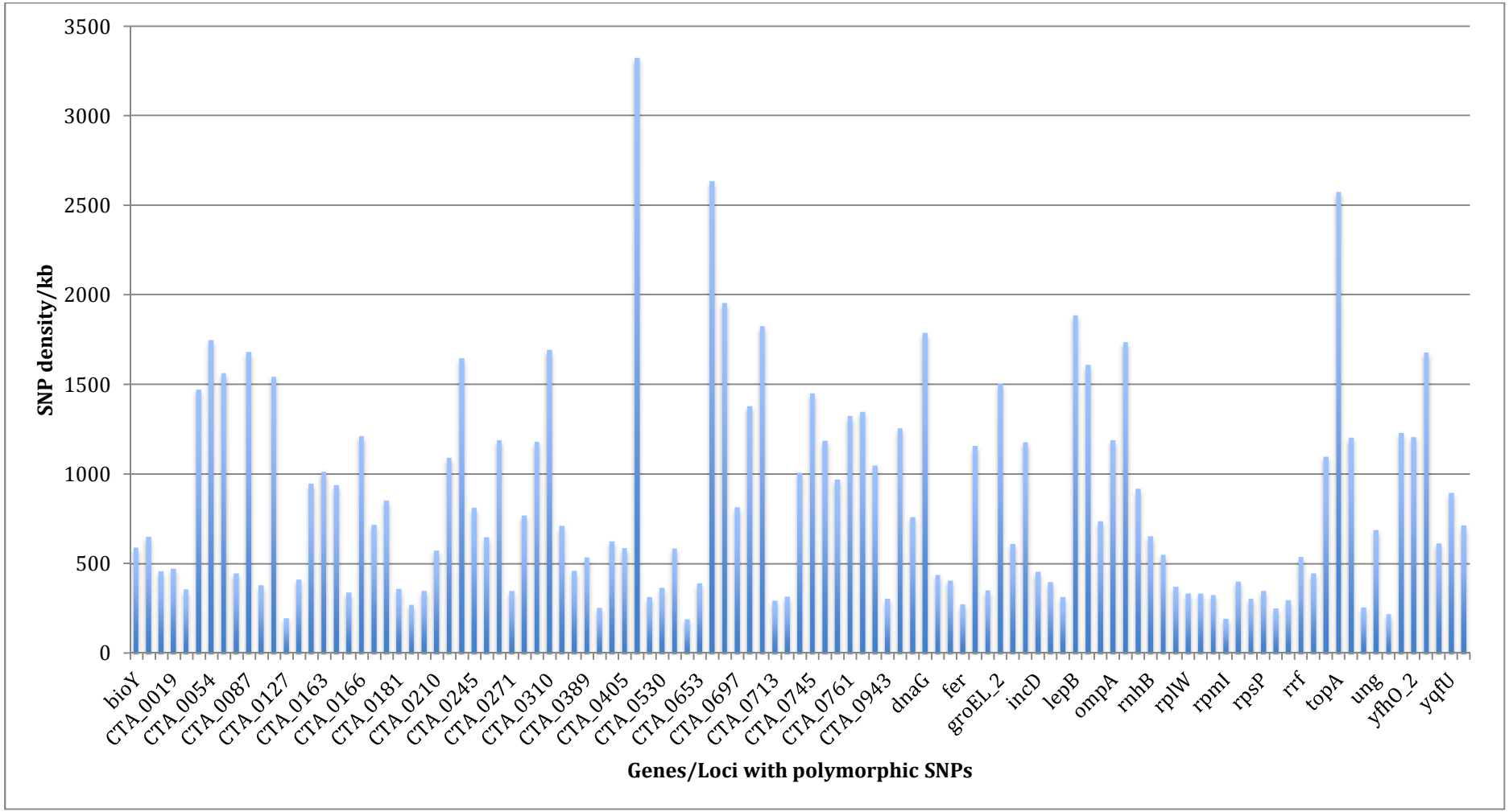
Figure 4. Principal Component Analysis of ocular *Chlamydia trachomatis* sequences





The 83 ocular *C. trachomatis* sequences are plotted against principal components (PC) 1 – 4. Sequences are colour-coded by *ompA* genotype. PC 1-3 best capture this clustering by genotype, identifying the underlying population structure.

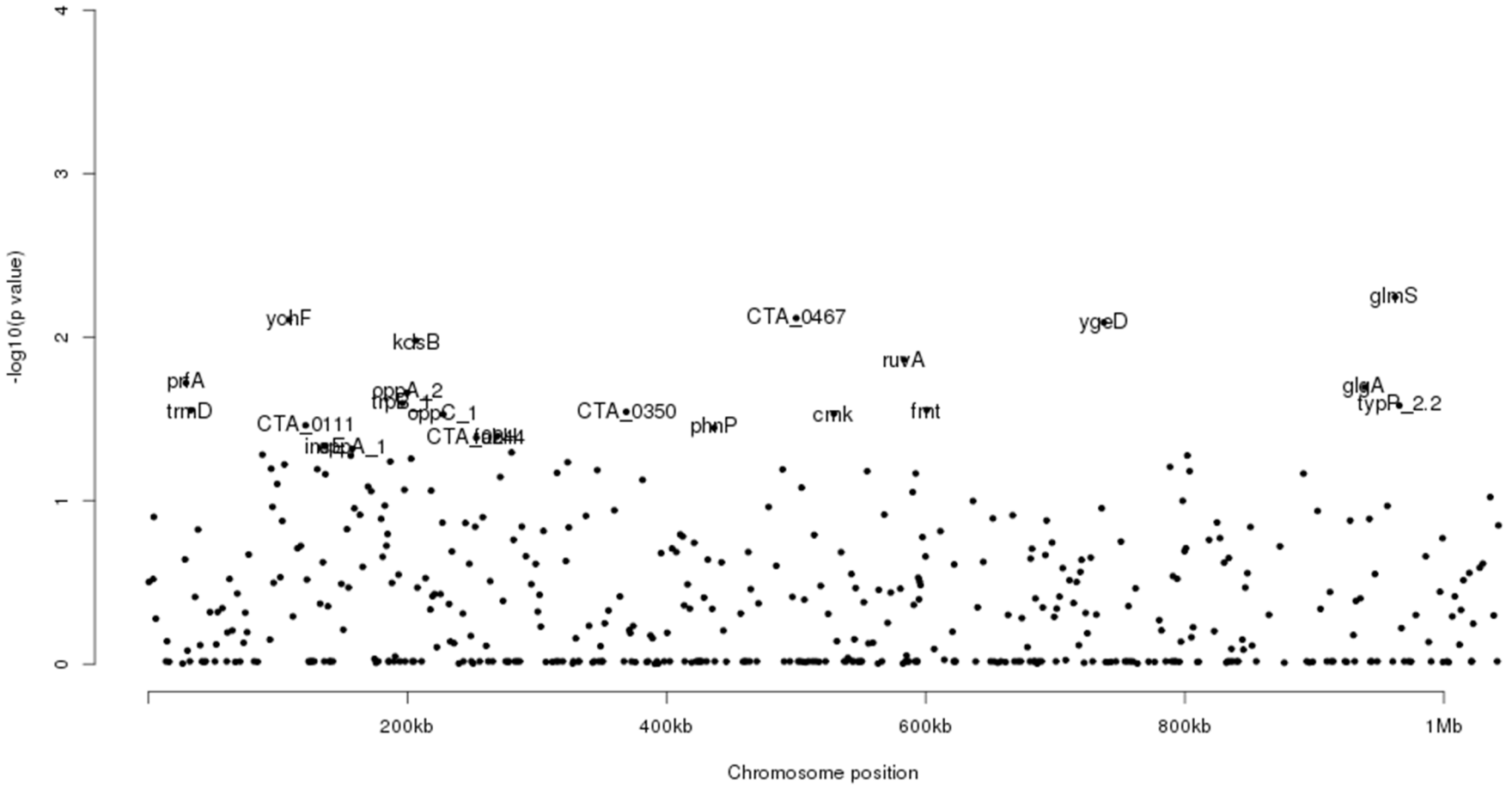
Figure 5. Polymorphic genes and loci on the *Chlamydia trachomatis* genome



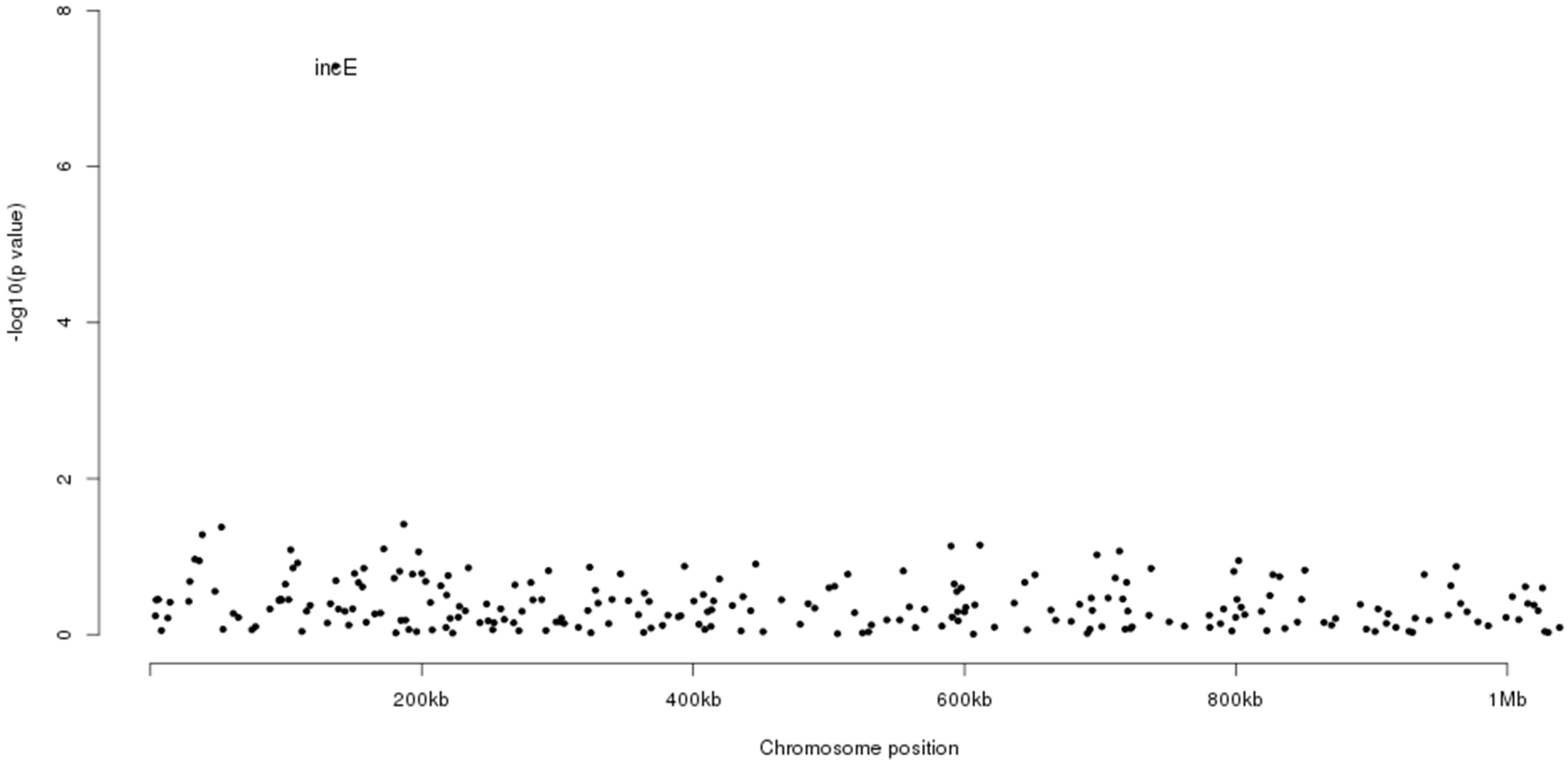
Highly polymorphic loci across the *C. trachomatis* genome are represented by SNP density per kb.

Figure 6. Genome-wide association scan to identify genes and loci of *Chlamydia trachomatis* associated with clinical disease phenotype in trachoma

(A) Active trachoma

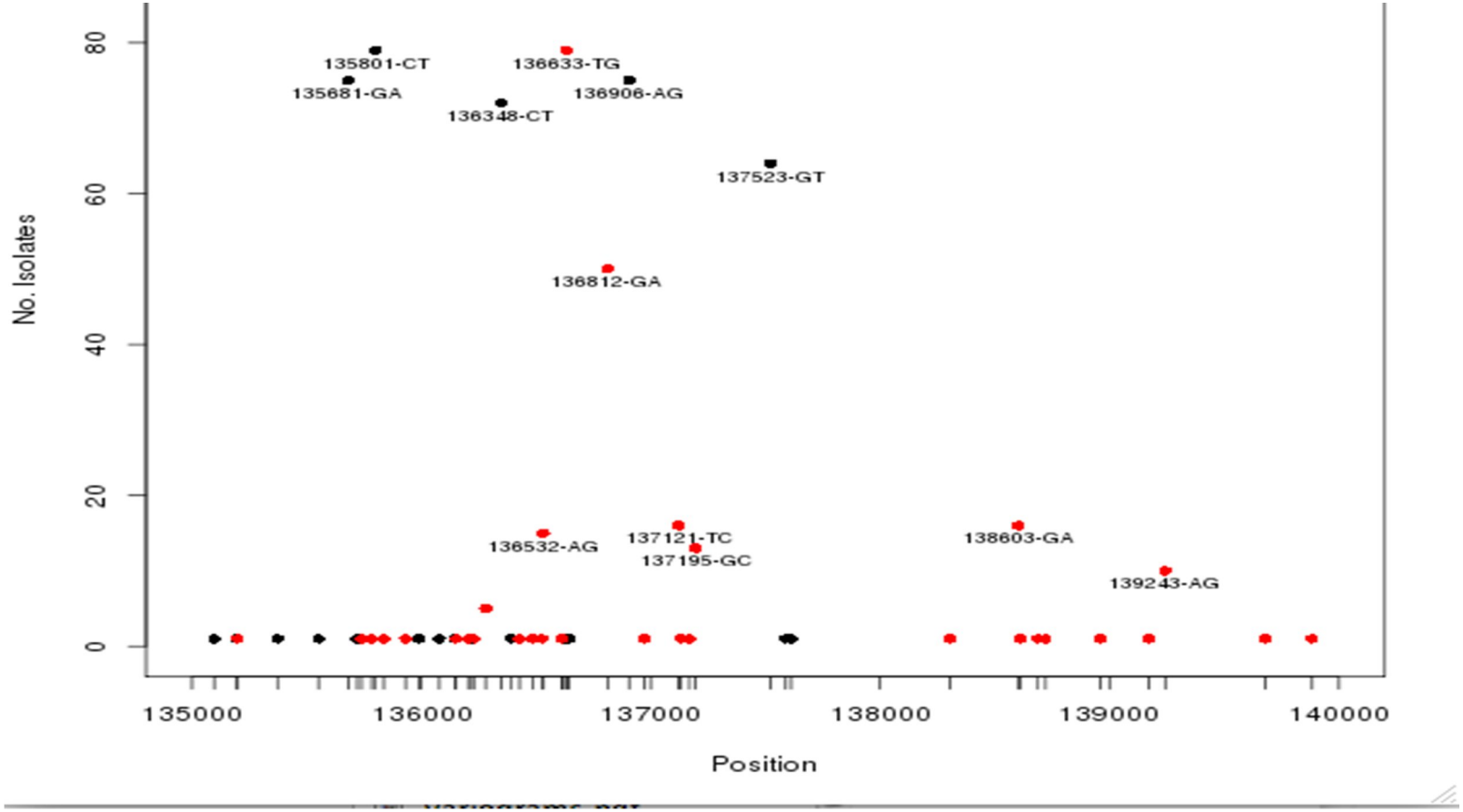


(B) Conjunctival inflammation



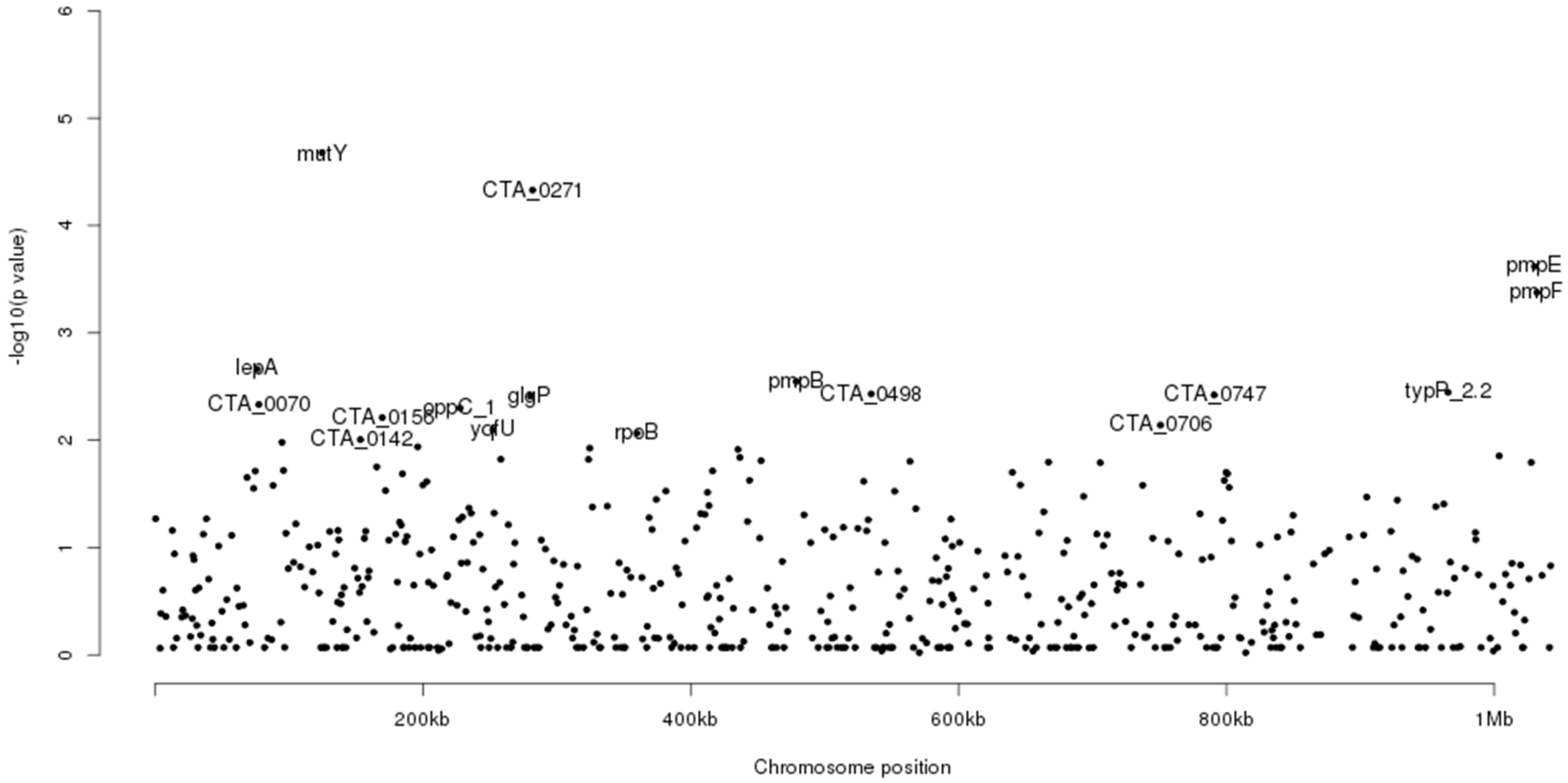
Locus GWAS for non-synonymous SNPs in coding regions for (A) Active trachoma (TF/TI) and (B) Conjunctival inflammation (P score; P0-3) in trachoma using binary and ordinal logistic regression models adjusting for age, population structure and linkage disequilibrium with Bonferroni correction for multiple testing (*p=5xe-05*). A $\log_{10}$ *p-value* of 2 is equivalent to a *p-value* of 0.01. Genes with an association significant at the 1% level are labeled.

Figure 7. Supporting SNPs across *incE* and *incDEFG* operon



Frequency of synonymous (black) and non-synonymous (red) SNPs shown across the *incE* and *incDEFG* operon (chromosomal position shown on x axis) in 83 isolates (y axis).

Figure 8. Genome-wide association scan to identify genes and loci associated with ocular *Chlamydia trachomatis* bacterial load in trachoma



Locus GWAS for non-synonymous SNPs in coding regions for ocular *C. trachomatis* load (defined as log-(e) *omcB* copies/swab) linear regression models adjusting for age, population structure and linkage disequilibrium with Bonferroni correction for multiple testing (*p=5xe-05*). A $\log_{10}$ *p-value* of 2 is equivalent to a *p-value* of 0.01. Genes with an association significant at the 1% level are labeled.

Table 1. Genes and loci associated with disease severity and ocular *Chlamydia trachomatis* load in trachoma

Clinical Phenotype	GWAS p-value	Gene	Locus	Size (Nº Amino Acids)	Biological function	Reference
Active Trachoma						
	0.0057	*glmS*	CTA_0889	606	Glutamine-fructose-6-phosphate aminotransferase involved in carbohydrate biosynthesis and glutamine metabolism	Carlson 2005 Stephens 1998
	0.0077		CTA_0465	71	Hypothetical membrane-associated protein	Carlson 2005
	0.0078	*ychF*	CTA_0097	366	Ribosome-binding ATPase involved in ATP and ribosomal binding	Carlson 2005
	0.0081	*ygeD*	CTA_0696	559	Transmembrane transport protein involved in macrolide efflux	Carlson 2005 Stephens 1998
	0.0105	*kdsB*	CTA_0200	254	3-deoxy-manno-octulosonate cytidyltransferase (CMP-2-keto-3-deoxyoctulosonic acid synthase) aka CKS (CMP-KDO synthase). Involved in lipopolysaccharide biosynthesis (activates KDO (a required 8-carbon sugar) for incorporation into bacterial LPS in Gram-negative bacteria).	Thomson 2008 Carlson 2005 Tipples 1995
	0.0134*	*ftsH*	CTA_0917	913	ATP-dependent metalloprotease involved in quality control of integral membrane proteins.	Thomson 2008 Carlson 2005
	0.0137	*ruvA*	CTA_0549	200	Holliday junction ATP-dependent DNA helicase. Part of the ruvA-ruvB complex involved in renaturing cruciform structure in supercoiled DNA. May promote strand exchange reactions in	Carlson 2005

					homologous recombination. Involved in DNA repair/recombination and SOS response.	
Conjunctival Inflammation						
	0.0017	*tolB*	CTA_0650	431	Periplasmic protein transport. Involved in TonB-independent protein uptake.	Carlson 2005
	0.0062	*rpoA*	CTA_0556	377	DNA-directed RNA polymerase subunit alpha. Involved in DNA-dependent RNA-polymerase transcription.	Carlson 2005
***C. trachomatis* ocular load**						
	0.0002	*pmpE*	CTA_0949	962	Polymorphic outer membrane protein. Autotransporter protein of the Type V Secretion System (TVSS) expressed in the elementary body phase of chlamydial development.	Carlson 2005 Henderson & Lam 2001 Stephens 1998
	0.0004	*pmpF*	CTA_0950	1034	Polymorphic outer membrane protein (as above)	Carlson 2005 Stephens 1998
	0.0022	*pmpB*	CTA_0448	1751	Polymorphic outer membrane protein (as above)	Carlson 2005 Stephens 1999
	0.0028	*lepA*	CTA_0069	602	Elongation factor (Ribosomal back-translocase LepA; GTP-dependent binding to ribosomes). Required for protein biosynthesis under stress conditions. May act as a fidelity factor for translation.	Carlson 2005
	0.0036	*tyrP_2.2*	CTA_0892	313	Tyrosine-specific transport protein. Involved in amino acid transmembrane transport.	Carlson 2005
	0.0037	*tarP*	CTA_0498	1106	Translocation actin recruiting protein (Tarp). Initiates signaling events regulating actin	Thomson 2008 Carlson 2005 Clifton 2004

					recruitment resulting in internalization. Secreted via chlamydial T3SS, where phosphorylated Tarp is exposed on the inclusion membrane on the internalized elementary body for several hours after entry. Proposed role in virulence and pathogenesis.	Stephens 1998
	0.0038		CTA_0747	395	ABC-transporter associated protein. Involved in iron-sulphur cluster assembly.	Carlson 2005
	0.0039	*glgP*	CTA_0270	814	Allosteric phosphorylase in carbohydrate metabolism. Glycosyltransferase involved in glycogen metabolism.	Seth-Smith 2009 Carlson 2005
	0.0046		CTA_0070	526	ADP/ATP carrier protein. Membrane component.	Carlson 2005
	0.0046	*oppC_1*	CTA_0218	281	Binding protein in a dependent membrane transport system.	Carlson 2005
	0.0062		CTA_0156	1449	Hypothetical membrane-spanning protein.	Carlson 2005
	0.0072		CTA_0706	608	Hypothetical exported protein.	Carlson 2005
	0.0080	*yqfU*	CTA_0241	298	Hypothetical membrane-spanning protein	Carlson 2005
	0.0086	*rpoB*	CTA_0337	1252	DNA-directed RNA polymerase subunit beta involved in DNA transcription	Carlson 2005
	0.0099		CTA_0142	300	Uncharacterised protein.	Carlson 2005
	0.0105		CTA_0086	98	Hypothetical membrane-spanning protein.	Carlson 2005
	0.0115	*trpB_1*	CTA_0185	315	Tryptophan synthetase beta. Involved in tryptophan biosynthesis.	Carlson 2005
	0.0118		CTA_0310	563	Uncharacterised protein.	Carlson 2005

	0.0122	*hlfX*	CTA_0413	447	GTPase associated with ribosomal subunit; postulated role in protein synthesis or ribosomal biogenesis.	Carlson 2005
	0.0140	*map*	CTA_0928	291	Methionine aminopeptidase. Removes N-terminal methionine from nascent proteins.	Carlson 2005
	0.0145	*phnP*	CTA_0414	266	Metal-dependent hydrolase	Carlson 2005

Active trachoma is defined as TF/TI by the WHO simplified grading system (Thylefors 1981). Conjunctival inflammation is defined by a P score (0-3) using the modified FPC grading system (Dawson 1987). *C. trachomatis* ocular load is defined as the log-(e) *omcB* copies/swab. Genome-wide significance using a Bonferroni correction is defined as *p=6.25xe-05*. Represented here are associated genes and loci significant at the 1% level to account for false discovery rates. *Odds Ratio=0.03 suggesting a mildly protective association.

Chapter 11

The impact of a single round of community mass drug treatment with azithromycin on disease severity and bacterial load in ocular *Chlamydia trachomatis* infection on the Bijagós Archipelago of Guinea Bissau

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed for each 'Research Paper' included in a thesis.

1. **For a 'research paper' already published**

   1.1. Where was the work published? N/A

   1.2. When was the work published? N/A

   1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
   N/A

   1.3. Was the work subject to academic peer review? N/A

   1.4. Have you retained the copyright for the work? **Yes / No** N/A
   If yes, please attach evidence of retention.
   If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. **For a 'research paper' prepared for publication but not yet published**

   2.1. Where is the work intended to be published? CLINICAL INFECTIOUS DISEASES

   2.2. Please list the paper's authors in the intended authorship order
   A.LAST, S.BURR, E.HARDING-ESCH, E.CASSAMA, M.NABICASSA, D.MABEY, M.HOLLAND, R.BAILEY

   2.3. Stage of publication – (Not yet submitted) / Submitted / Undergoing revision from peer reviewers' comments / In press

3. **For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)**
   I DESIGNED THE STUDY, CONDUCTED THE FIELD WORK, LABORATORY WORK, CLEANED + ANALYSED THE DATA, INTERPRETED THE FINDINGS + WROTE THE PAPER.

NAME IN FULL (Block Capitals) ANNA RACHEL LAST

STUDENT ID NO: 291603

CANDIDATE'S SIGNATURE A. Date 05/09/2014

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) R C Bailey

TITLE

The impact of a single round of community mass drug treatment with azithromycin on disease severity and bacterial load in ocular *Chlamydia trachomatis* infection on the Bijagós Archipelago of Guinea Bissau

RUNNING HEAD

Disease severity and *C. trachomatis* bacterial load and MDA

Anna R. Last[1#], Sarah E. Burr[2], Emma Harding-Esch[1], Eunice Cassama[3], Meno Nabicassa[3], David C.W. Mabey[1], Martin J. Holland[1], Robin L. Bailey[1]

[1]Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, [2]Disease Control and Elimination Theme, Medical Research Council Unit The Gambia, P.O. Box 273 Banjul, Atlantic Boulevard, Fajara, The Gambia, [3]Programa Nacional de Saúde de Visão, Ministério de Saúde Publica, P.O. Box 50, Avenida de Unidade Africana, Bisssau, Guiné Bissau

#CORRESPONDING AUTHOR Mailing Address: Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Telephone: +44 (0)207 927 2566 Email: anna.last@lshtm.ac.uk

Footnotes

(1) There are no commercial or other associations existing for any of the authors of this manuscript that may pose a conflict of interest.

(2) This work was supported by the Wellcome Trust [grant number 097330/Z/11/Z].

(3) Correspondence and requests for reprints should be addressed to the corresponding author: Dr Anna R. Last, Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.

Telephone: +44 (0)207 927 2566 Email: anna.last@lshtm.ac.uk

ABSTRACT

Background

Trachoma is caused by ocular infection with *Chlamydia trachomatis (Ct)*. The Bijagós Archipelago of Guinea Bissau is trachoma-hyperendemic. The WHO recommends three annual rounds of community mass drug administration (MDA) of azithromycin if the prevalence of follicular trachoma (TF) in 1-9 year olds > 10% at district level to achieve an elimination target of TF < 5%. Prior to this survey these communities were treatment-naïve and there had been no trachoma control activities.

Methods

We conducted a population-based trachoma survey on four islands. Left and right upper tarsal conjunctivae of each participant were assessed for trachoma using the WHO simplified and modified FPC grading systems. Swabs were taken from the left upper tarsal conjunctiva. After the survey mass MDA was undertaken in all communities on the study islands in line with WHO and national policy. We used a droplet digital PCR assay to diagnose and quantitate *Ct* infection. Repeat ocular examination and conjunctival swabs were obtained on the treated cohort one year following MDA.

Results

At baseline TF prevalence in 1-9 year olds was 22.0% (136/618). The overall prevalence of infection was 18.6% (25.4% in 1-9 year olds). Treatment coverage was estimated at 70%. At one year, TF in 1-9 year olds was significantly reduced (7.4% (29/394), $p<0.001$). Overall prevalence of infection was significantly reduced to 3.3% (34/1029), $p<0.001$) (6.6% in 1-9 year olds, $p<0.001$)). The median ocular *Ct* load had reduced from 2038 to 384 genomic copies/swab ($p<0.001$). There was greater clustering of infection following MDA, demonstrated by local indicators of spatial association, though the number of clusters of high load infections was reduced. Inflammatory disease was less severe following MDA, and most infection was present in children under the age of 10 years with active trachoma.

Conclusions

Despite initial high rates of disease and infection and lower than recommended MDA coverage, elimination of trachoma may be achieved with fewer than three rounds of MDA in this setting. The reduction in *Ct* load, disease severity and presence of high load *Ct* infection clusters following MDA inform about *Ct* infection transmission dynamics and suggest that transmission may be coming under control. These data may be useful in understanding the epidemiology of trachoma in endemic communities and have subsequent implications for the national trachoma control programme in Guinea Bissau.

KEYWORDS: *Chlamydia trachomatis*, bacterial load, trachoma, disease severity, community mass treatment

Chlamydia trachomatis is the leading infectious cause of blindness globally (1-3). Trachoma is caused by infection with ocular strains of *C. trachomatis* and manifests as distinct clinical syndromes, beginning with an acute self-limiting kerato-conjunctivitis which may progress to chronic inflammatory disease with subsequent conjunctival scarring and blinding sequelae.

The World Health Organization (WHO) advocates the implementation of the SAFE strategy (Surgery for trichiasis, Antibiotics for active infection, Facial cleanliness to prevent disease transmission and Environmental improvement to increase access to water and sanitation) for trachoma elimination. Mass Drug Administration (MDA) of azithromycin aims to clear infection from communities such that trachoma ceases to be a public health concern (4). The WHO currently recommends three annual rounds of MDA if the prevalence of follicular trachoma (TF) in 1-9 year olds is > 10% (at district level) to achieve a reduction in TF to < 5% (4). Following three rounds of MDA an impact survey is recommended, based on which decisions about continuing treatment or conducting surveillance are made. There are conflicting data from trachoma-endemic communities on the optimal duration and mode of administration required to achieve this target.

There are modest data from countries with hypo and mesoendemic trachoma suggesting that between one and three rounds of single dose MDA are sufficient. Oral azithromycin, even as a single dose delivered as a mass administration to communities, has significantly reduced the burden of active disease, and in some populations has eliminated infection with *C trachomatis* entirely (5). Often, disease (and infection) returns to communities, despite mass treatments, necessitating continued surveillance and repeat treatments. The evidence-base relating to optimal frequency of mass drug administration that will be effective in eliminating trachoma is not fully understood and may vary between settings (6).

In trachoma-hyperendemic populations, despite high coverage of MDA under research study conditions, levels of infection and disease, although reduced, can persist or re-emerge (7). Longitudinal studies in Tanzania suggest that with present WHO protocols, hyperendemic countries may need yearly mass treatment for over ten years, which has significant economic and logistic impact on national trachoma programmes, non-government organisations and pharmaceutical donors of azithromycin (8). Good coverage without elimination of disease at 18 months and re-emergence of disease within families has been demonstrated in other studies (9).

Studies in Ethiopia suggest that prevalence of disease and infection can be reduced with multiple annual mass treatments, but that on cessation of treatment, disease and infection re-emerge. In some regions where annual treatment appears to have had little impact, even biannual treatment that has been implemented under research study conditions in this setting has failed to reduce TF prevalence in 1-9 year olds to under 5% (10).

The dynamics of trachoma transmission, particularly in the context of MDA, are complex and not well understood. Trachoma endemicity is likely to be important, such that in hypoendemic populations disease can spontaneously disappear (11-14) or disappear after a single round of community mass treatment with antibiotics (5,15). In mesoendemic populations disease prevalence may stabilize following community mass antibiotic treatment (9), and in hyperendemic populations disease and infection persist despite continued rounds of community mass treatment (16).

At baseline (prior to MDA) trachoma was endemic on the remote Bijagós Archipelago of Guinea Bissau. In this population we found a strong association between *C. trachomatis* bacterial load and disease severity. We demonstrated spatial clustering of high load infections (17). Following the initial survey we sought to assess the impact of a single round of MDA on the prevalence of active trachoma and ocular *C. trachomatis* infection, with particular emphasis on the effect on *C. trachomatis* load, disease severity and spatial clustering of *C. trachomatis* infection one year following MDA.

Methods

Ethical Statement

This study was conducted in accordance with the declaration of Helsinki. Ethical approval was obtained from the Comitê Nacional de Ética e Saúde (Guinea Bissau), the LSHTM Ethics Committee (UK) and The Gambia Government/MRC Joint Ethics Committee (The Gambia). Written (thumbprint or signature) informed consent was obtained from all study participants or their guardians as appropriate. Following this study all communities on the study islands received further treatment with oral azithromycin in accordance with WHO and national protocols.

Study Design and Study Population

Trachoma survey methodology and this study population have been described previously (17-22). Briefly, we used first stage cluster random sampling with geospatial representation at village-level to randomly select households for inclusion in the survey at baseline. Data were geo-coded at household and village level as described previously (17). One year following MDA we sought to follow up members of all households enrolled at baseline.

Clinical Examination and Conjunctival Sampling

Clinical examination and conjunctival sampling were conducted at baseline and one year following treatment. A single validated examiner assessed each participant using the WHO simplified and modified FPC grading systems as described previously (4,17,20,23). In the modified FPC system, follicles (F), papillary hypertrophy (inflammation) (P) and conjunctival scarring (C) are each assigned a separate grade from 0-3. FPC grades of F2/3 or P3 equate to a diagnosis of active trachoma (TF (Trachoma-Follicular) or TI (Trachoma-Intense Inflammatory) by the WHO simplified system) and a grade of C2/3 (and in some cases C1) equates to a diagnosis of TS (trachomatous scarring). Clinical grading of the upper tarsal conjunctivae was conducted in the field as described previously (17-20). Samples were taken from the left upper tarsal conjunctiva of each participant using a well-tolerated standardized procedure described in previous studies (17-20).

Community Mass Treatment

A single height-based dose of oral azithromycin was offered to all individuals in all communities participating in the study in accordance with WHO and national policy. Alternative treatment with tetracycline eye ointment was offered if there were contraindications to treatment with azithromycin.

Detection and Quantitation of C. trachomatis

DNA extraction and droplet digital PCR (ddPCR) (Bio-Rad Laboratories, Hemel Hempstead, UK) were conducted as described previously (17,19,20). *C. trachomatis* plasmid-based ddPCR was used to diagnose infection and a single-copy pathogen chromosomal gene (*omcB*) to estimate pathogen load in each plasmid-positive sample (19,20). Estimated quantities of *omcB* (*C. trachomatis* load) are expressed as copies/swab.

Statistical Analysis

C. trachomatis quantitation data were processed as described previously (19,20). Data were double entered into a customised database (Microsoft Access 2007) and discrepancies resolved through source documents. Data were cleaned and analysed in STATA 13 (Stata Corporation, College Station, Texas USA). Statistical significance was determined at the 5% level.

We examined trachoma and *C. trachomatis* infection prevalence data at baseline and follow-up using a Chi2 test of proportions. *C. trachomatis* load data were log-(e) transformed where indicated. Median load comparisons were made between baseline and follow-up using Kruskall-Wallis Chi2. Associations between load and detailed clinical phenotype were examined using multivariable mixed effects linear and logistic regression models accounting for clustering detected in previous studies (17,18).

Geocoded data were projected into UTM Zone 28N and analysed in ArcGIS 10.1 (ESRI Inc., USA) as described previously (17). Briefly, Moran's I was calculated at baseline and follow-up to evaluate the effect of MDA on the global spatial distribution of active trachoma and *C. trachomatis* infection. A local indicator of spatial association (local (Anselin) Moran's I) was used to identify clusters and

outliers of *C. trachomatis* infection by load at baseline and follow-up. This method detects statistically significant clusters or outliers related to *C. trachomatis* load based on the calculation of z-scores for the distribution. Cluster types identified (relative to their z-score and p-value) include H-H (high loads associated with other high loads), L-L (low loads associated with other low loads), H-L (a high load outlier associated with other predominantly lower loads) and L-H a low load outlier associated with other predominantly high loads). The methods underlying these geostatistical tools and their application are discussed fully elsewhere (17).

Results

Prevalence of trachoma and C. trachomatis infection

Participant enrolment and follow-up are illustrated in *Figure 1*. MDA was conducted following the baseline survey and was estimated at 70% overall across the study islands. Characteristics at baseline and follow-up were not significantly different, nor were population-based estimates of scarring trachoma (TS) and trachomatous trichiasis (TT) (*Table 1*). The prevalence active trachoma (TF/TI) and ocular *C. trachomatis* infection was statistically significantly reduced following treatment (*Table 2*). The prevalence of TF in 1-9 year olds was reduced from 22.0% (95% CI (Confidence Interval) 18.9-25.5) (136/618) to 7.4% (95% CI 4.8-9.9) (29/394) ($p<0.001$). The prevalence of TI in this age group was also reduced (from 2.9% (95% CI 1.4-4.1) (18/618) to 1.5% (95% CI 0.3-2.7) (6/394)). The prevalence of ocular *C. trachomatis* infection was reduced in the population from 18.6% (280/1502) to 3.3% (34/1029) ($p<0.001$) and in 1-9 year olds from 25.4% (157/618)) to 6.6% (26/395) ($p<0.001$).

Clustering of active trachoma and C. trachomatis infection

There was some evidence for increased clustering of active trachoma and *C. trachomatis* infection at village-level at follow-up. There was also increased clustering of infection at household level at follow-up (*Table 3*). The Moran's I for *C. trachomatis* infection at baseline was 0.06 (z=2.10, *p=0.0353*) and 0.27 (z=3.85, *p=0.0001*) at follow-up indicating increased clustering following MDA.

C. trachomatis ocular load and disease severity

Median estimated load of *C. trachomatis* infection was significantly reduced from 2038 *omcB* copies/swab to 348 *omcB* copies/swab (Chi2 = 6.21, *p=0.0127*) (*Figure 2*). At follow-up almost all infections occurred in children under the age of 10 years, with 59% (20/34) occurring in children aged 0-5 years. At baseline a greater proportion of individuals with TS or normal conjunctivae were infected, indicating that infection was more widely spread in the population (*Figure 3*).

Disease severity with infection was also reduced from baseline to follow-up, most markedly with respect to conjunctival inflammation. There was a shift from higher proportions of P2 and P3 scoring disease at baseline to greater proportions of P0 and P1 scoring disease at follow-up (*Figure 4*). Conversely, the proportions of high scoring follicular disease (F2 and F3) were greater in the infected cohort at follow-up (*Figure 4*). The association between *C. trachomatis* load and disease severity is demonstrated in age-adjusted mixed effects linear regression models accounting for household clustering (*Table 4*). A strong association remains between *C. trachomatis* load and conjunctival inflammation (P score), though the association is weaker at follow-up. The association with F3 scores is maintained and is stronger at follow-up.

Age-adjusted mixed effects logistic regression analysis accounting for household clustering examining associations between disease severity and *C. trachomatis* infection showed that P scores were the strongest predictors at baseline. At follow-up there was stronger association between F scores and infection, although an association between infection and P scores remains evident (*Table 5*). Interestingly, there was a very strong association between F3 and the presence of infection at follow-up, independent of effects of collinearity with P scores.

Spatial clustering of high load C. trachomatis infections

Maps were generated using the local Moran's I statistic. These demonstrated clustering of *C. trachomatis* infection by load and found that at baseline there were a larger number of clusters of high load infections (H-H clusters) than at follow-up. High load outliers (H-L) were not present at follow-up. The H-H clusters present at follow-up were at different locations compared to baseline. One H-H cluster at follow-up was present in a location where there was a H-L outlier prior to MDA. A second H-H cluster was located where there was previously no clustering related to bacterial load. Clustering of low load (L-L clusters) infections was not evident at either time point. Cluster-outlier maps at baseline and follow-up are presented in *Figure 5*. The threshold *C. trachomatis* load for membership of a H-H cluster is ~10,000 *omcB* copies/swab irrespective of time point.

Discussion

These data suggest that in this island population, despite initial high rates of disease and infection and relatively low MDA coverage, reduction of TF to <5% in 1-9 year olds may be achievable with fewer than three rounds of MDA. The WHO recommends that MDA aim for 100% coverage to be considered adequate for trachoma control programmes (4). However, due to the significant logistic challenges

that exist in this remote area it was only possible to achieve an estimated 70% coverage.

Despite this, following MDA there was a significant reduction in the prevalence of both disease and infection following MDA. At both time points active trachoma and *C. trachomatis* infection (particularly those with the highest bacterial loads) were most prevalent in children under 10 years of age. Following MDA *C. trachomatis* infection has virtually disappeared in adults, and is reduced in those with scarred or normal conjunctivae. The ubiquitous presence of *C. trachomatis* infection across all age groups and clinical phenotypes at baseline is likely to represent the distribution typical of chronic hyperendemicity of trachoma before MDA (24). The reduction in the prevalence of infection and change in distribution by age and disease suggests a shift in the epidemiology of ocular *C. trachomatis* infection and may mark the beginning of control of transmission in these communities.

The median ocular *C. trachomatis* load was significantly lower in the population following MDA. This is consistent with findings from other studies suggesting that from 2-12 months following MDA the prevalence and load of infections remain low (9).

The number of clusters of high load infections detected using local spatial statistics was reduced and there was an absence of high load outlying infections amidst other low load infections after MDA. This supports the previous suggestion of bacterial load being important in transmission of infection and its maintenance in the population (14,17). This phenomenon may be due to reduced chlamydial diversity in the population following MDA. The role of chlamydial strain diversity in transmission is unclear but greater diversity is likely to represent more successful

transmission. There is some evidence that following MDA the number of strains by *ompA* genotyping was substantially reduced (25).

At both time points the threshold value for membership in a H-H cluster is ~10,000 *omcB* copies/swab, suggesting that there may be a threshold load important in transmission. The change from a H-L outlier prior to MDA to a H-H cluster following treatment suggests that there may be ongoing transmission at this location. Additionally a new H-H cluster is apparent where there was previously none. This particular location has a mobile population, being populated by fishermen and their families from Guinea Conakry, Sierra Leone and other islands on the archipelago. There are limited amenities in this settlement and it is possible that infection has been reintroduced. Introduction of infection following migration events has been documented in The Gambia (7). These data provide insight into the dynamics of transmission and the micro-epidemiology of *C. trachomatis* infection in these populations and support findings suggesting the importance of *C. trachomatis* bacterial load in transmission (14,17).

Clustering of disease and infection was more apparent following MDA. The strongest clustering was present at village level following MDA. This may be relevant in the context of previous spatial analyses conducted in this population suggesting that the village may be an important unit of transmission in these communities (17). An increase in clustering and size of clusters of infection in treated communities has been described (26).

Additionally, in this study were able to use the detailed conjunctival grading system to investigate associations between infection, bacterial load and disease severity following MDA. A strong association between *C. trachomatis* load and inflammatory trachoma has been described previously (17,24,27). Associations

between *C. trachomatis* genotype, using pathogen genome-wide association scans, and inflammatory phenotype have been shown (Last *et al*., in preparation (***Chapter 10***). The reduced association between infection and P score following MDA may reflect the decreased burden of circulating infection and decreased infection loads. It is likely that repeated episodes of infection are reduced following MDA due to a decrease in circulating *C. trachomatis* and subsequently reduced transmission. Moreover, in populations undergoing MDA there is evidence that clinical signs of trachoma become less specific for *C. trachomatis* infection (28), suggesting that only the more severe phenotypes remain predictive of *C. trachomatis* infection.

Although a single grader standardised against a trachoma-grading expert using the WHO simplified grading system was employed at both time points to minimise inter-observer bias, this may not eliminate all bias with respect to the use of the non-standardised modified FPC grading system and the associations noted between phenotype and bacterial load. However, the concurrent use of the WHO simplified system supports these associations and this study supports data from other studies (17,24,27).

There was some loss to follow-up during the course of this study due to the fluctuating nature of these communities. The characteristics of the population in terms of age distribution, gender and chronic sequelae (trachomatous trichiasis) suggest that the follow-up cohort is adequately representative of the baseline cohort. However, the prevalence of conjunctival scarring in the cohort at follow-up was higher, indicating that there may be selective drop out of less severely affected individuals. This may affect the overall prevalence of *C. trachomatis* infection found at follow-up, as the prevalence of infection in conjunctival scarring may be higher than in those with no clinical signs of trachoma (18). There may also be responder bias within the follow-

up cohort. Further longitudinal study is underway in a larger randomly selected population to verify prevalence data from this study

We are currently developing mathematical models to investigate transmission dynamics of *C. trachomatis* and the effect of high load infections in these communities. Such models, in conjunction with ongoing work on pathogen genomics to define *C. trachomatis* strain diversity and geospatial analysis may improve our understanding of disease pathogenesis and transmission and may be useful in trachoma surveillance in post-MDA settings to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission.

In summary, through investigating the micro-epidemiology of *C. trachomatis* infection and its relationship with bacterial load and disease severity, these data suggest that MDA is having an impact on transmission of ocular *C. trachomatis* in these communities. However, further monitoring is required, as our geospatial analyses suggest that that there may be on-going transmission and risk of reintroduction of infection to communities despite MDA. Further longitudinal study, utilising mathematical models, high resolution chlamydial genotyping and geospatial analysis, is necessary to provide a more complete picture of the relationship between disease severity, chlamydial load, transmission and elimination thresholds in communities undergoing MDA.

Acknowledgements

Azithromycin was donated for treatment of trachoma on the islands by Pfizer Inc. through application via the International Trachoma Initiative. We extend our thanks to our colleagues in the Programa Nacional de Saúde de Visão in Bissau and to the communities of the Bijagós Archipelago, our field research team and all study participants in Guinea Bissau.

References

(1) Resnikoff S, Pascolini D, Etya'ale D et al (2004). Global data on visual impairment in the year 2002. Bull World Health Organ *82*:844-851.

(2) Mabey DC, Solomon AW, Foster A (2003). Trachoma. Lancet 362:223-229.

(3) Hu VH, Harding-Esch EM, Burton MJ et al. (2010). Epidemiology and control of trachoma: systematic review. Trop Med Int Health 15(6):673-91.

(4) Dawson CR, Jones BR, Tarizzo ML (1981). Guide to trachoma control in programs for the prevention of blindness. Geneva: World Health Organization.

(5) Solomon AW, Holland MJ, Alexander ND et al. (2004). Mass treatment with single dose azithromycin for trachoma. New Engl J Med 351:1962-1971.

(6) Lietman T, Porco T, Dawson C and Blower S (1999). Global elimination of trachoma: How frequently should we administer mass chemotherapy? Nature Med 5(5):572-6.

(7) Burton MJ, Holland MJ, Makalo P et al. (2005). Re-emergence of *Chlamydia trachomatis* infection after mass antibiotic treatment of a trachoma-endemic Gambian community: a longitudinal study. Lancet 365:1321-8.

(8) West SK, Munoz B, Mkocha H et al. (2011). Number of years of annual mass treatment with azithromycin needed to control trachoma in hyperendemic communities in Tanzania. J Infect Dis 204:268-73.

(9) West SK, Munoz B, Mkocha H et al. (2005). Infection with *Chlamydia trachomatis* after mass treatment of a trachoma hyperendemic community in Tanzania: a longitudinal study. Lancet 366:1296-1300.

(10) Lakew T, Alemayehu W, Melese M et al. (2009). Importance of coverage and endemicity on the return of infectious trachoma after a single mass antibiotic distribution. PLoS Negl Trop Dis 3(8):e507.

(11) Dolin PJ, Faal H, Johnson GJ et al. (1997). Reduction of trachoma in a sub-Saharan village in the absence of a disease control programme. Lancet 349(9064):1511-1512.

(12) Hoeschmann A, Metcalfe N, Kanjaloti S et al. (2001). Reduction of trachoma in the absence of antibiotic treatment: evidence from a population-based survey in Malawi. Ophth Epidemiol 8(2-3):145-153.

(13) Jha H, Chaudary J, Bhatta R et al. (2002). Disappearance of trachoma in Western Nepal. Clinical Infect Dis 35(6):765-8.

(14) Chidambaram JD, Alemeyehu W, Melese M et al. (2006). Effect of a single mass antibiotic distribution on the prevalence of infectious trachoma. JAMA 295(10):1142-1146.

(15) Gaynor BD, Miao Y, Cevallos V et al. (2003). Eliminating trachoma in areas with limited disease. Emerg Infect Dis 9(5):596-8.

(16) Lakew T, House J, Hong KC et al. (2009). Reduction and return of infectious trachoma in severely affected communities in Ethiopia. PLoS Negl Trop Dis 3(2):e376.

(17) Last AR, Burr SE, Alexander ND et al. (2014). Spatial clustering of high load ocular *Chlamydia trachomatis* infection and disease severity in trachoma-endemic communities on the Bijagós Archipelago, Guinea Bissau. Emerg Infect Dis [under review].

(18) Last AR, Burr SE, Weiss HA et al. (2014). Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau. PLoS Negl Trop Dis 8(6):e2900.

(19) Roberts Ch, Last A, Molina-Gonzalez S et al. (2013). Development and evaluation of a next generation digital PCR diagnostic assay for ocular *Chlamydia trachomatis* infections. J Clin Microbiol 51(7):2195-2203.

(20) Last A, Roberts CH, Cassama E et al. (2014). Plasmid copy number and disease severity in naturally occurring ocular *Chlamydia trachomatis* infection. J Clin Microbiol 52(1):324-7.

(21) Harding-Esch E, Edwards T, Mkocha H et al. (2010). Trachoma prevalence and associated risk factors in The Gambia and Tanzania: Baseline results of a cluster randomized controlled trial. PLoS Negl Trop Dis 4(11):e861.

(22) Edwards T, Harding-Esch EM, Hailu G et al. (2008). Risk factors for active trachoma and *Chlamydia trachomatis* infection in rural Ethiopia after mass treatment with azithromycin. Trop Med Int Health 13:556-565.

(23) Thylefors B, Dawson CR, Jones BR et al. (1987). A simple system for the assessment of trachoma and its complications. Bull World Health Organ 65(4):477-83.

(24) Burton MJ, Holland MJ, Faal N et al. (2003). Which members of a community need antibiotics to control trachoma? Conjunctival *Chlamydia trachomatis* infection load in Gambian villages. Invest Ophthalmol Vis Sci 44:4215-4222.

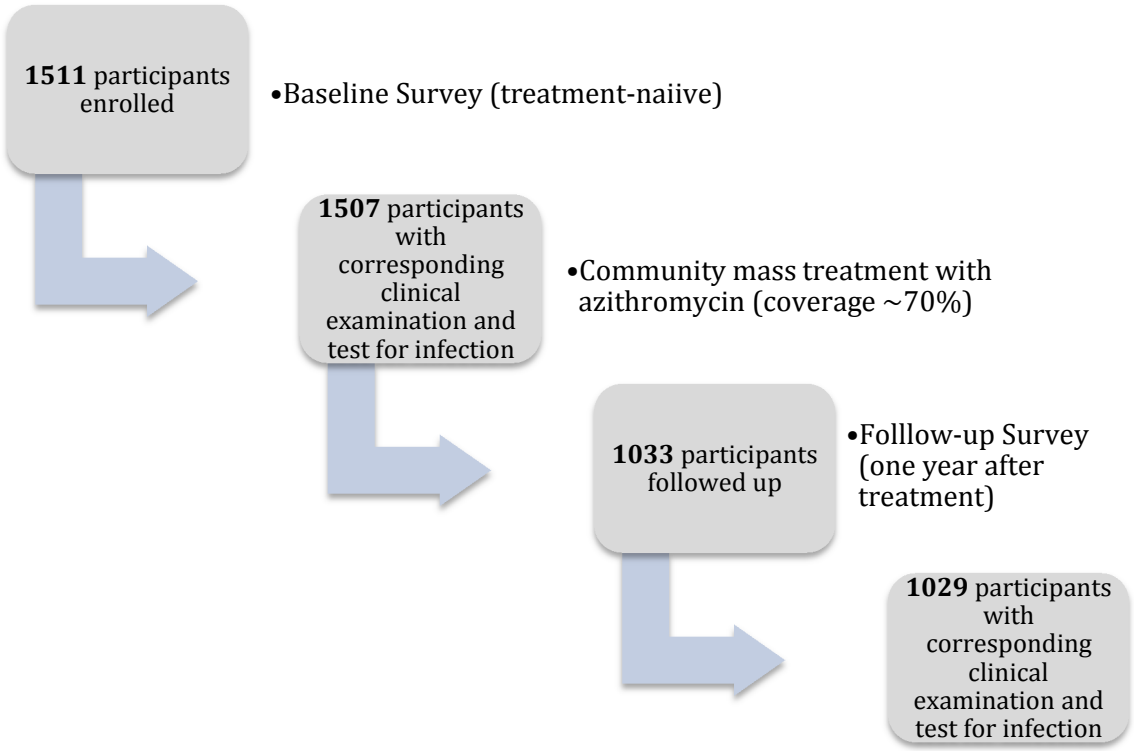
(25) Andreasen AA, Burton MJ, Holland MJ et al. (2008) *Chlamydia trachomatis ompA* variants in trachoma: what do they tell us? PLoS Negl Trop Dis 2(9):e306.

(26) Yohannan J, He B, Wang J et al. (2014). Geospatial Distribution and Clustering of *Chlamydia trachomatis* in Communities Undergoing Mass Azithromycin Treatment. Invest Ophthalmol Vis Sci 55(7):4144-50.

(27) Bobo LD, Novak N, Munoz B et al. (1997). Severe disease in children with trachoma is associated with persistent *Chlamydia trachomatis* infection. J Infect Dis 176(6):1524-30.

(28) Lee JS, Munoz BE, Mkocha H et al. (2014). The effect of multiple rounds of mass drug administration on the association between ocular *Chlamydia trachomatis* infection and follicular trachoma in preschool-aged children. PLoS Negl Trop Dis 8(4):e2761.

Figure 1. Enrollment of participants at baseline and follow-up one year after community mass treatment for trachoma control



Community mass treatment was distributed in accordance with WHO and national trachoma control policies. Zithromax® was donated by Pfizer Inc. through the International Trachoma Initiative.

Table 1. Study population characteristics at baseline and follow-up at one year

	Baseline	***n*[b]**	**Follow-up**	***n*[b]**
Median age (years) [IQR[a]]	13 [5-32]	1507	12 [5-35]	1033
Female	58%	869	63%	648
Age Group				
0-5 years	28%	416	28%	288
6-10 years	16%	250	17%	180
11-15 years	11%	157	10%	113
>15 years	45%	684	45%	452
Conjunctival Scarring (TS[c])				
Overall (prevalence) [95% CI[d]]	23.8 [21.6-25.9]	357/1502	30.5[27.7-33.3]	313/1026
By Age Group				
0-5 years	2.7[1.1-4.2]	11/414	2.1[0.4-3.7]	6/288
6-10 years	2.8[0.8-4.8]	7/250	5.6[2.2-9.1]	10/177
11-15 years	11.5[6.5-16.4]	18/157	18.9[11.6-26.2]	21/111
>15 years	47.2[43.5-51.0]	321/680	61.3[56.8-65.8]	276/450
Trachomatous Trichiasis ($TT_{>15}$[c])	3.5[2.1-4.9]	24/680	5.1[3.1-7.1]	23/450

[a]IQR = Inter Quartile Range [b]Denominator where indicated [c]WHO simplified grading system (22); $TT_{>15}$ reflects TT in those over the age of 15 years [d]CI = Confidence Interval

Table 2. The effect of community mass treatment with azithromycin on the prevalence of active trachoma and ocular *C. trachomatis* infection

Prevalence (%) [95% CI[a]]	**Baseline**	***n***	**Follow-up**	***n***
$TF_{1\text{-}9}$[b]	22.0[18.9-25.5]	136/618	7.4*[4.8-9.9]	29/394
$TI_{1\text{-}9}$	2.9[1.4-4.1]	18/618	1.5[0.3-2.7]	6/394
Ct_{all}[c]	18.6[16.7-20.6]	280/1502	3.3*[2.2-4.4]	34/1029
$Ct_{1\text{-}9}$	25.4[22.0-28.8]	157/618	6.6*[4.1-9.0]	26/395

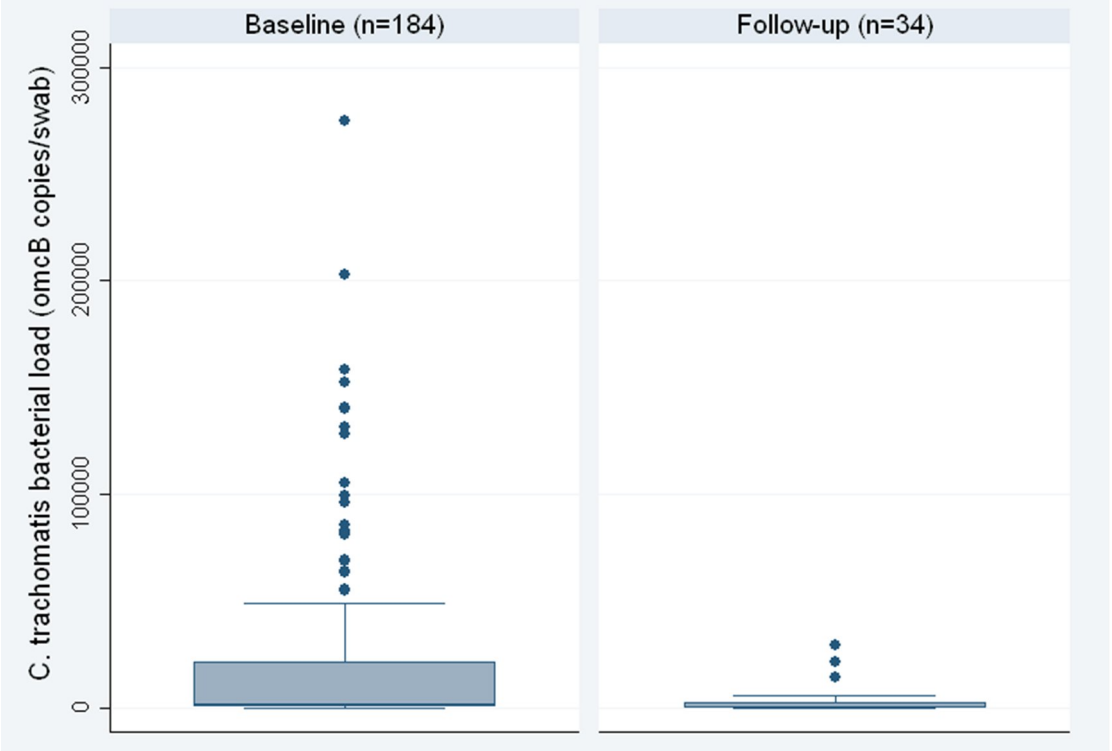
[a]CI = Confidence Interval [b]$TF_{1\text{-}9}$ = TF (Trachoma-Follicular) in 1-9 year olds; $TI_{1\text{-}9}$ = TI (Trachoma-Intense Inflammatory) in 1-9 year olds [c]*Ct = C. trachomatis* infection in the population overall and in 1-9 year olds *$p<0.001$ (Chi2)

Table 3. Clustering of active trachoma and *C. trachomatis* infection

	Active trachoma		*Ct* infection	
Cluster Level	$CE_{Baseline}$[a] [95% CI s.e.]	$CE_{Follow\text{-}up}$ [95% CI s.e.]	$CE_{Baseline}$[a] [95% CI s.e.]	$CE_{Follow\text{-}up}$ [95% CI s.e.]
Household	1.08[0.79-1.48]	0.89[0.85-0.95]	1.35[1.08-1.68]	1.52[0.85-2.74]
Village	0.76[0.50-1.14]	1.05[0.56-1.96]	0.89[0.65-1.21]	1.04[0.56-1.96]
Island	0.47[0.15-1.49]	0.34[0.08-1.37]	0.42[0.17-1.00]	0.19[0.01-5.76]

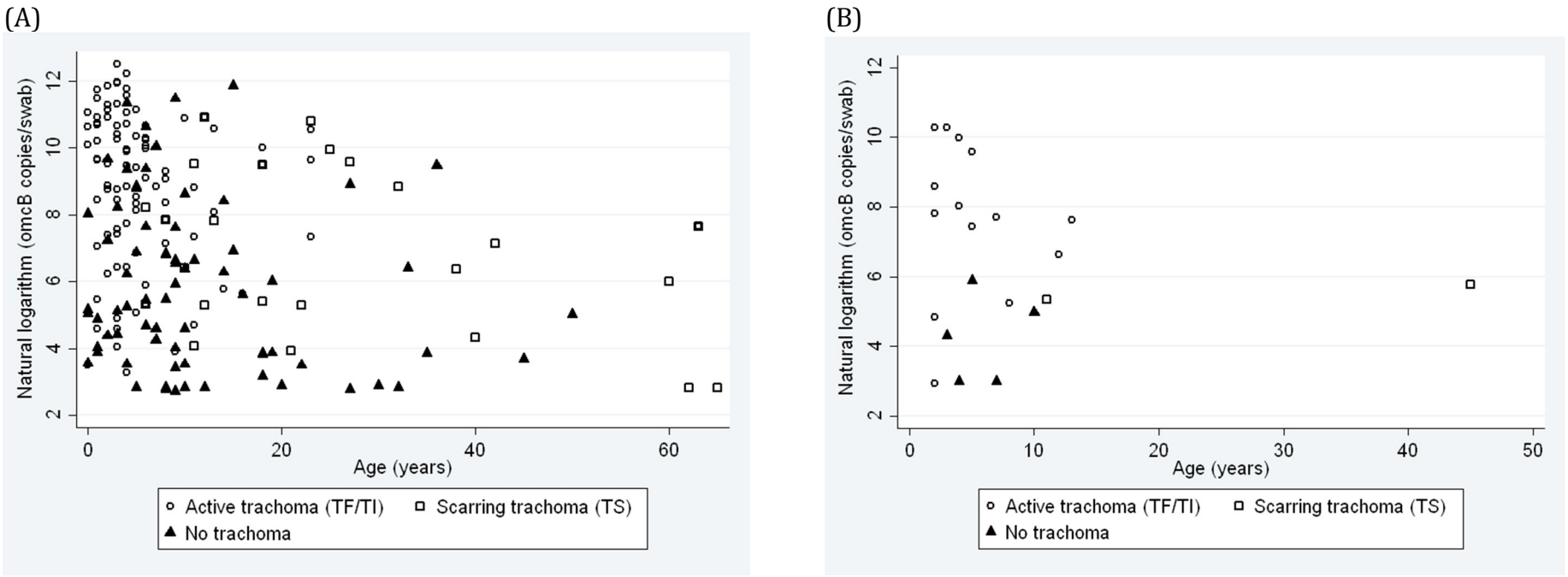
[a]CE = Cluster Estimates obtained from age-adjusted mixed effects regression models for active trachoma and *Ct* (*C. trachomatis*) infection at baseline and follow-up. 95% CI (Confidence Intervals) of the s.e. (standard error) are quoted. Mixed effects models including all three cluster levels showed household to have the strongest effect. All cluster estimates were significant at the 1% level (Wald Chi2).

Figure 2. Reduction in median *C. trachomatis* load in ocular *C. trachomatis* infection following community mass treatment with azithromycin



Box-whisker plot showing the median *C. trachomatis* load (*omcB* copies/swab) from individuals with conjunctival infection at baseline and follow-up.

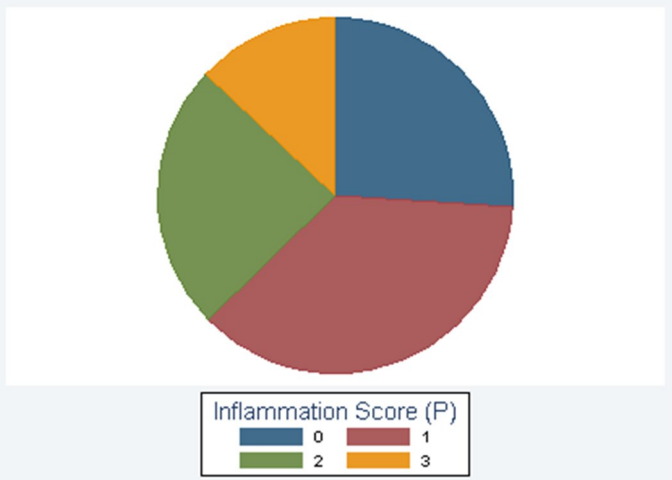
Figure 3. The effect of community mass treatment with azithromycin on ocular *C. trachomatis* load by age and clinical phenotype



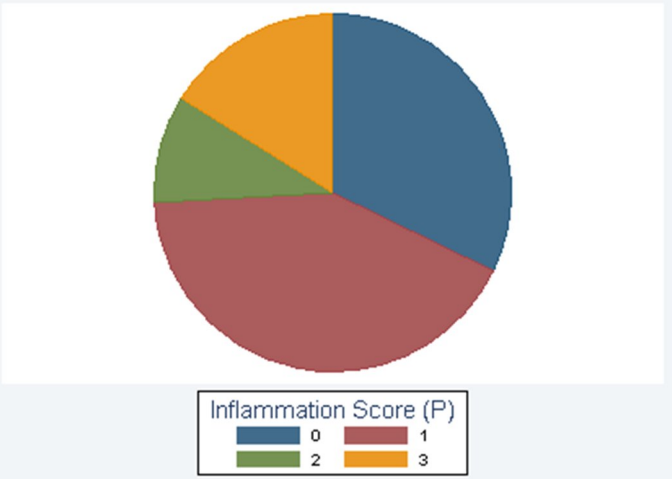
(A) Baseline (B) Follow-up. Clinical phenotype defined using the WHO simplified grading system (22).

Figure 4. The effect of community mass treatment with azithromycin on disease severity in individuals with ocular *C. trachomatis* infection

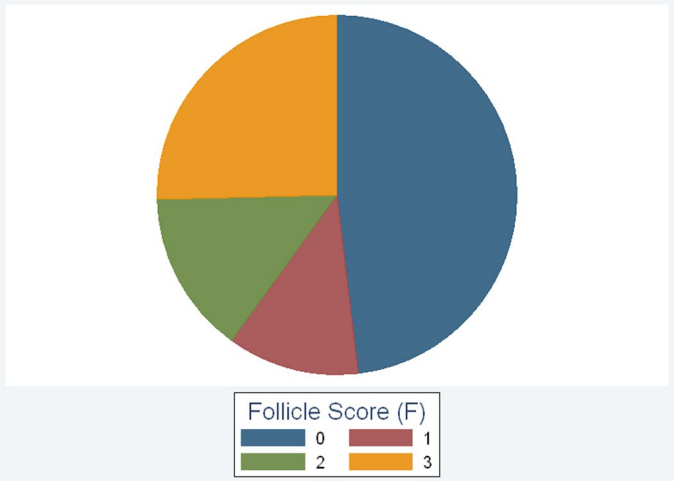
(a) n=184



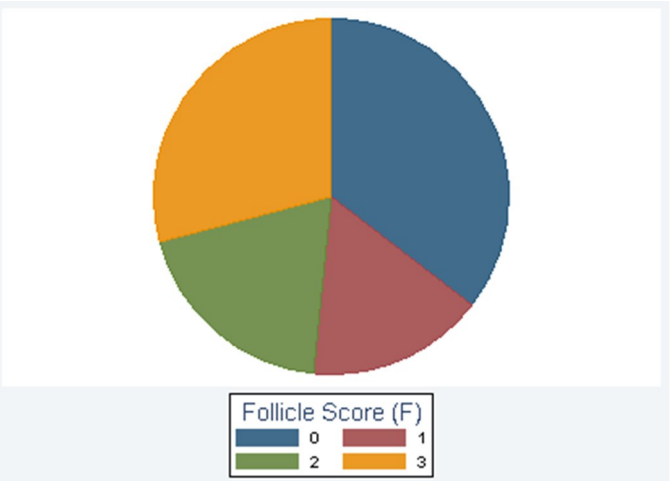
(b) n=34



(c) n=184



(d) n=34



Proportion of infected individuals with conjunctival inflammation (P score) at (a) baseline and (b) follow-up. Proportion of infected individuals with conjunctival follicles (F score) at (c) baseline and (d) follow-up. F and P scores according to the modified FPC grading system (4).

Table 4. The effect of community mass treatment with azithromycin on associations between disease severity and *C. trachomatis* bacterial load in individuals with ocular *C. trachomatis* infection

	Baseline	***(n=184)***		***Follow-up***	***(n=34)***	
Clinical Phenotype	***n***	**OR_{adj} [95% CI]**	***p-value***	***n***	**OR_{adj} [95% CI]**	***p-value***
F Score						
F0	89	--		11	--	
F1	22	1.67 [0.48-5.78]	0.417	5	1.06 [0.20-5.53]	0.948
F2	27	1.84 [0.60-5.68]	0.287	6	1.13 [0.12-10.28]	0.916
F3	42	6.16 [1.97-19.26]	**0.002**	9	9.60 [1.76-52.44]	**0.009**
P Score						
P0	48	--		10	--	
P1	68	6.15 [2.47-15.31]	**<0.001**	13	1.82 [0.50-6.60]	0.359
P2	45	21.74 [6.82-69.32]	**<0.001**	3	1.30 [0.17-9.64]	0.799
P3	24	27.61 [6.81-111.80]	**<0.001**	5	18.10 [3.35-97.74]	**0.001**

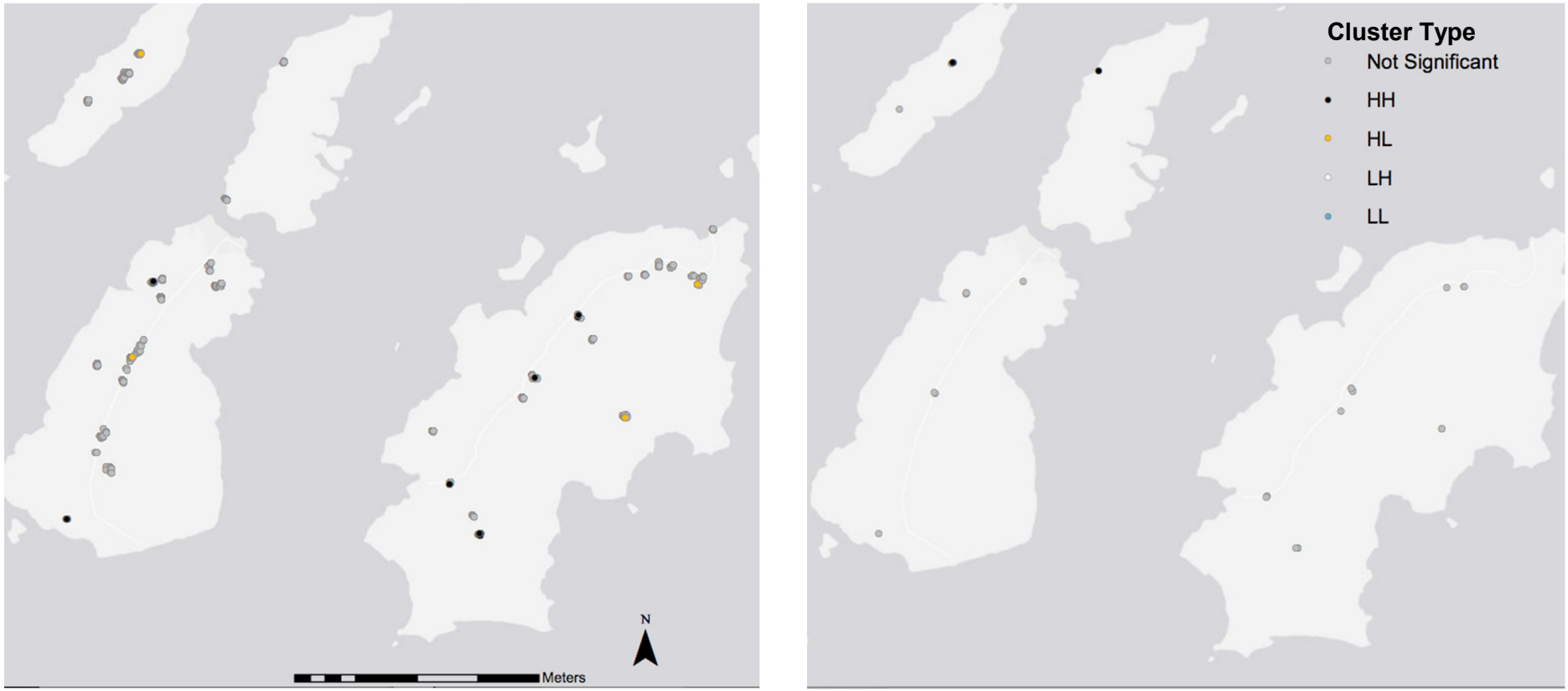
Age-adjusted multivariable mixed effects linear regression analysis accounting for household clustering with log-(e)-transformed *C. trachomatis* load (*omcB* copies/swab). F and P scores according to the modified FPC grading system (4). Individuals may appear in multiple clinical grading categories. p-value is for the Wald Chi2.

Table 5. The effect of community mass treatment with azithromycin on associations between clinical phenotype and the presence of ocular *C. trachomatis* infection

	Baseline	***(n=1507)***		***Follow-up***	***(n=1029)***	
Clinical Phenotype	***n***	**OR_{adj} [95% CI]**	***p-value***	***n***	**OR_{adj} [95% CI]**	***p-value***
Model 1						
TF	152	8.27 [5.07-13.48]	**<0.001**	133	40.43 [9.73-168.02]	**<0.001**
TI	27	9.72 [2.74-34.44]	**<0.001**	19	2.48 [0.30-20.47]	0.288
Model 2						
F Score						
F0	1268	--		950	--	
F1	78	1.87 [0.95-3.68]	0.068	40	4.82 [1.31-17.72]	**0.018**
F2	83	0.82 [0.41-1.66]	0.587	20	5.82 [1.06-32.14]	**0.043**
F3	59	3.76 [1.42-9.92]	**0.007**	12	49.31 [5.05-481.01]	**<0.001**
P Score						
P0	1200	--		882	--	
P1	189	5.82 [3.51-9.64]	**<0.001**	113	3.56 [1.14-11.10]	**0.028**
P2	68	15.84 [7.12-35.24]	**<0.001**	18	8.87 [1.33-59.27]	**0.024**
P3	29	30.37 [8.48-108.84]	**<0.001**	9	4.54 [0.41-49.39]	0.214

Age-adjusted multivariable mixed effects logistic regression analysis accounting for household clustering. Two models examining clinical phenotype were used: Model 1 summarised using TF (Trachoma-Follicular) and TI (Trachoma-Intense Inflammatory) according to the WHO simplified grading system (22) and Model 2 using more detailed F and P scores according to the modified FPC grading system (4). Individuals may appear in multiple clinical grading categories. p-value is for the Wald Chi2.

Figure 5. Cluster-Outlier maps showing the effect of community mass treatment on clusters of high load ocular *C. trachomatis* infections



(A) (B)

C. trachomatis load was log transformed (ln(ln+1)) due to significant negative skew. Statistically significant positive values for the Local Moran's I statistic indicate clustering with similarly high (H-H) or low (L-L) values. Negative values indicate that adjacent observations have dissimilar values and that this observation is an outlier (H-L (a high load outlier) or L-H (a low load outlier)). Maps are presented at (A) baseline and (B) follow-up after MDA. H-H clusters are observed at both time points. H-L outliers are observed at baseline only. There are no L-L clusters. Observation values represent *C. trachomatis* load.

Chapter 12 Study Summary, Limitations and Conclusions

12.1 Summary of Study Findings

12.1.1 Trachoma is hyperendemic on the Bijagós Archipelago (Chapters 7 and 11).

Table 1 summarises trachoma and ocular *C. trachomatis* infection prevalence data from the baseline survey using WHO indicators.

Table1. Prevalence of trachoma and ocular *C. trachomatis* infection in treatment-naïve communities on the study islands of the Bijagós Archipelago

	Prevalence (%) [95% Confidence Interval]	***n***
Follicular Trachoma (TF_{1-9}[b])	22.0[18.9-25.5]	136/618
Inflammatory Trachoma (TI_{1-9}[b])	2.9[1.4-4.1]	18/618
C. trachomatis infection$_{all}$[c]	18.6[16.7-20.6]	280/1502
C. trachomatis infection$_{1-9}$[c]	25.4[22.0-28.8]]	157/618
Conjunctival Scarring (TS[b])		
Overall	23.8 [21.6-25.9]	357/1502
By Age Group		
0-5 years	2.7[1.1-4.2]	11/414
6-10 years	2.8[0.8-4.8]	7/250
11-15 years	11.5[6.5-16.4]	18/157
>15 years	47.2[43.5-51.0]	321/680
Trachomatous Trichiasis ($TT_{>15}$[b])	3.5[2.1-4.9]	24/680

From the baseline survey of 1507 individuals with complete clinical examination and infection data. [b]Prevalence data presented according to the WHO simplified grading system [1]. TF_{1-9} = Trachoma-Follicular (TF) in 1-9 year olds; TI_{1-9} = Trachoma-Intense Inflammatory (TI) in 1-9 year olds; TS = Trachomatous scarring; $TT_{>15}$ = Trachomatous Trichiasis in those >15 years of age. [c]*C. trachomatis* DNA positivity by ddPCR in $_{all}$ (the population overall) and $_{1-9}$ (1-9 year olds).

12.1.2 Water access reflecting specific hygiene behaviours and the presence of flies around latrines are associated with the presence of ocular C. trachomatis infection in communities on the Bijagós Archipelago (Chapter 7).

Knowledge of risk factors associated with trachoma and *C. trachomatis* infection in a particular setting may help prioritise trachoma elimination activities. In this study, household-level clustering was adjusted for in all models examining risk factor associations. Active trachoma and the presence of *C. trachomatis* infection were strongly correlated. Risk factors for disease and infection were similar. In this environment, individual measures of facial cleanliness (ocular and nasal discharge) and household-level risk factors relating to fly populations, hygiene behaviours and water usage are likely to be important in *C. trachomatis* transmission.

There was a strong association between type of water source and both active trachoma and *C. trachomatis* infection. The association with active trachoma is likely to be mediated through *C. trachomatis* infection. In Bijagós communities, every household has access to a water source within 15 minutes walking distance. It is therefore likely that this association may reflect more subtle hygiene and water use behaviours related to allocation of water collected for specific hygiene practices.

Latrine access was not associated with active trachoma or *C. trachomatis* infection, despite the fact that 60% of individuals did not have access to a latrine. However, the presence of flies around a household pit latrine was strongly associated with infection in final multivariable models. There was a strong association between flies around a latrine and the presence of visible faeces within a latrine. These findings suggest that flies may be an important risk factor in the transmission of trachoma in this environment.

12.1.3 There is spatial clustering of high load ocular C. trachomatis infection (Chapter 8).

Spatial clusters and hotspots, independent of age and gender, of individual infections with high bacterial loads were detected using local indicators of spatial association. There was no evidence of coldspots or clustering of low load infections, suggesting that high load infections may be important in transmission of chlamydial infection.

These data suggest that there may be a threshold of *C. trachomatis* load below which clustering of high loads does not occur. Equally, these data suggest that having

a low load infection may be similar to being uninfected with respect to transmission, since those with no infection behaved similarly to those with low load infections in terms of clustering. This supports the proposed Allee effect, which hypothesizes that a reduction in chlamydial fitness due to reduced pathogen population size or density results in its disappearance from a population [2].

Measures of spatial dependence in the distribution of *C. trachomatis* infection, supported by household demographics, suggest that village-level transmission is important in this population.

12.1.4 High load ocular C. trachomatis infections were associated with the most severe inflammatory disease in active trachoma (Chapter 8).

The highest *C. trachomatis* loads were associated with the most severe clinical signs of disease and the strongest associations were with increasing inflammatory scores. The majority of ocular *C. trachomatis* infections occur in children, who have the highest loads and most severe active trachoma.

In this population *C. trachomatis* load increases with disease severity (for both follicular (F) and inflammatory (P) scores), the strongest association being with increasing P scores. The association between *C. trachomatis* load and high F scores is in part due to collinearity between F and P, such that individuals with high F scores are likely also to have some degree of inflammation present. Inflammation has previously been found to be associated with high *C. trachomatis* loads and persistence of infection in children [3].

12.1.5 The (virulence) plasmid copy number (per genome) is stable in naturally occurring ocular C. trachomatis infection and there is no association with disease severity or ocular C. trachomatis load (Chapter 9).

The *C. trachomatis* plasmid is a virulence factor. This is the first study to document plasmid copy number variation in naturally occurring ocular *C. trachomatis* infection, demonstrating its stability and that it does not vary with geographic location, clinical phenotype or *C. trachomatis* load.

Nanolitre-sized droplet partitioning of the ddPCR reaction, promotes optimal primer-template interaction conditions which are robust to variation in PCR efficiency, and enabled the accurate estimation of both plasmid and genome (*omcB*) copy numbers within the same reaction.

This study found the plasmid copy number to be stable, with a median copy number per genome of 5.34 (range 1-18 copies/genome) and there were no associations with *C. trachomatis* bacterial load or disease severity.

There is clearly a selective advantage in maintaining this extra-chromosomal element, and though there is convincing evidence that the chlamydial plasmid is a virulence factor [4-9], these data suggest that plasmid copy number is not associated with disease severity and that additive gene dosage effects do not appear to correlate with pathogen virulence *in vivo*.

12.1.6 Naturally occurring ocular C. trachomatis genomic diversity was greater than anticipated in the context of global C. trachomatis diversity (Chapter 10).

In the context of the global collection of published laboratory reference and clinical urogenital, lymphogranuloma venereum and ocular *C. trachomatis* strains [10], this study shows greater than anticipated phylogenetic diversity amongst these *C. trachomatis* infections. This is the first population-based prospective collection of naturally occurring ocular *C. trachomatis* strains obtained direct from clinical samples. Whole genome sequence analysis has demonstrated the extent of this diversity, summarized in phylogenetic trees in ***Chapter 9***. The Tajima's D at whole genome sequence level (D=-2.50, *p=0.01*) suggests an excess of low frequency polymorphisms, indicating population expansion and/or purifying selection [11]. This is supported by evidence of widespread recombination demonstrated using the pairwise homoplasy index and previous studies showing evidence of purifying selection using *ompA* genotyping [12,13].

12.1.7 incE was associated with conjunctival inflammation at a genome-wide significance level (Chapter 10).

In this population, the *C. trachomatis incE* gene was strongly associated with conjunctival inflammation at genome-wide significance level using pathogen genome wide association scan (GWAS) with ordinal logistic regression analysis. There were supporting synonymous and non-synonymous SNPs within the *incE-D* operon. IncE is part of the inclusion membrane protein family. Inclusion membrane proteins are type III effector proteins, secreted from within the parasitophorous vesicle and inserted in the inclusion membrane. They are thus exposed on the surface of the inclusion, where they may interact with host cells. They appear to be important to

inclusion membrane structure and biogenesis, and are thought to play a role in virulence and pathogenesis [14,15].

12.1.8 Novel genes and loci (mutY and CTA_0271) were associated with C. trachomatis ocular load at a genome-wide significance level (Chapter 10).

Linear GWAS models showed that the strongest associations at genome-wide significance with *C. trachomatis* ocular load were with *mutY* and *CTA_0271*. *mutY* has been identified in a hypervirulent *C. trachomatis* strain causing lymphogranuloma venereum [16] and in other chlamydial species causing tryptophan-independent mutants [17]. In *E. coli* this gene has been extensively described as a mutator locus involved in mismatch corrections and the maintenance of genomic stability. *CTA_0271* is a novel loci coding for a hypothetical membrane protein.

Other genes (*pmpE, pmpF, pmpB, tyrP and tarP*) were also associated with *C. trachomatis* load, but below the stringent threshold set for genome-wide significance correcting for multiple testing. The polymorphic membrane proteins (*pmps*), *tyrP* (causing tyrosine phosphorylation) and *tarP* (involved in actin skeleton reconfiguration and pathogen-directed phagocytosis) are thought to be associated with chlamydial virulence [14,19,20].

12.1.9 There were no associations between genes and loci on the C. trachomatis plasmid and disease severity or ocular C. trachomatis load (Chapter 10).

Ordinal and linear GWAS regression models did not show any association between plasmid genes and loci with conjunctival inflammation or ocular *C. trachomatis* load. However there were associations with genes on the chromosome that may be regulated by plasmid-encoded genes (such as *glgP*) [21]. It is clear that the plasmid has an important role in chlamydial pathogenesis, but the mechanisms behind this remain unclear.

12.1.10 A single round of community mass treatment with oral azithromycin significantly reduces the prevalence of active trachoma and ocular C. trachomatis infection (Chapter 11).

There was a significant reduction in the prevalence of both disease and infection following community mass distribution of azithromycin (MDA). Most infection, in particular those with the highest loads, was concentrated in children

under the age of 10 years with active trachoma. Following MDA *C. trachomatis* infection virtually disappeared in adults, and is markedly reduced in those with scarred or normal conjunctivae. The baseline distribution of infection and disease is representative of hyperendemicity of trachoma in this population before MDA [22]. The reduction in the prevalence of infection and change in distribution by age and disease after MDA may mark the beginning of control of transmission in these communities.

12.1.11 Following a single round of community mass treatment with oral azithromycin there is a significant reduction in median ocular C. trachomatis load and conjunctival inflammation (Chapter 11).

The median ocular *C. trachomatis* load was significantly lower in the population following MDA. This is consistent with findings from other studies suggesting that following MDA the prevalence and load of infections can remain low [23].

This study used a detailed conjunctival grading system to investigate associations between infection, bacterial load and disease severity following MDA. Following MDA there was weaker association between infection and P score than at baseline. Clinical signs of trachoma become less specific for *C. trachomatis* infection once MDA has been initiated [24]. This may reflect the overall decrease in circulating *C. trachomatis* and probable reduction in repeated episodes of infection. This may represent reduced transmission within these communities.

12.1.12 Following a single round of community mass treatment with oral azithromycin there is increased spatial clustering of ocular C. trachomatis infection (Chapter 11).

Clustering of disease and infection was more apparent following MDA using mixed effects regression models and measures of global spatial autocorrelation. The strongest clustering was present at village level following MDA. This may be relevant in the context of previous spatial analyses conducted in this population suggesting that the village may be an important unit of transmission in these communities [25]. The phenomenon of increased clustering of infection in treated communities has been described previously [26].

The number and size of clusters of high load infections detected using local spatial statistics was reduced and there was an absence of high load outlying infections amidst other low load infections after MDA. These disparate remaining small clusters of high load infections following MDA may represent isolated islands of infection rather than the larger clusters of high load infection present at baseline. This supports the previous suggestion of bacterial load being important in transmission of infection and its maintenance in the population and supports they hypothesis that transmission is beginning to come under control with MDA [2,25].

12.2 Study Limitations

Many of the findings from this study were based on cross-sectional data. A cross-sectional study has inherent limitations, providing a 'snap-shot' view of the epidemiology of trachoma and *C. trachomatis* infection. Cross-sectional data is limited to demonstration of associations rather than determination of causality. It is also difficult to make inferences about transmission of *C. trachomatis* infection based on cross-sectional data, particularly as the study is limited in the assumption of load and clinical disease severity as a steady state. Longitudinal study is required to fully investigate the dynamics of *C. trachomatis* bacterial load in transmission, but they are often time-consuming, expensive and logistically challenging. Some of these issues are currently being addressed through the development of predictive mathematical models that can be modified and validated using follow-up data.

Although a single grader standardised against a trachoma-grading expert using the WHO simplified grading system was employed at each data collection time point to minimise inter-observer bias, this may not eliminate intra-observer bias with respect to the use of the non-standardised modified FPC grading system. However, concurrent use of the WHO simplified system supports the associations between detailed clinical phenotype and bacterial load. Moreover these data support findings from other studies using the modified FPC grading system. Conjunctival digital photographs taken will be independently graded using the modified FPC system to further validate these findings.

There was some loss to follow-up in the cohort following MDA. This was largely due to unavoidable losses due to mobility within the population. The follow-up cohort were demographically similar to the baseline cohort, therefore it was assumed that they were representative of the population. However, to confirm the post-treatment findings there has been a further impact survey conducted in these communities, with new random selection of households to remove any cohort bias and ensure an adequate sample size to detect a reduction in prevalence of disease and infection in the context of MDA.

For the GWAS analysis it was only possible to obtain high quality whole genome sequence data for 68% of samples with quantifiable *C. trachomatis* DNA by ddPCR. Further application of stringent quality control filters on sequence data obtained resulted in 83 (66% of sequenced samples (45% of total quantifiable *C. trachomatis* positive samples)) samples being retained in the final analysis. This is likely to reduce the power of the study to adequately detect less extreme associations with disease severity or *C. trachomatis* load. Despite these limitations, these data represent a unique and sizeable archive of ocular *C. trachomatis* whole genome sequence data with corresponding detailed clinic-epidemiological metadata. Future validation of these findings in other populations (described in ***Chapter 12***) may address the limitations of the statistical power of GWAS analysis. Additionally a conservative p-value threshold for genome wide significance was applied using a Bonferroni correction for multiple comparisons, to reduce the detection of 'false positive' associations.

12.3 Study Conclusions

The Bijagós Archipelago is a unique setting where trachoma is hyperendemic, and at the time of the baseline survey communities were treatment-naïve. This is the first study of its kind to be conducted in this remote and challenging environment. This study has established the prevalence of trachoma and ocular *C. trachomatis* infection, and socio-environmental risk factors have been identified that may be important in the implementation of trachoma elimination activities in these communities.

The primary aims of this study were to use novel molecular, bioinformatic and geostatistical approaches in conjunction with population-based clinical and

epidemiological metadata to investigate the micro-epidemiology of ocular *C. trachomatis* and active trachoma in this population.

C. trachomatis ocular load was estimated using a next generation ddPCR assay that was optimized and evaluated for this study. The accurate estimation of *C. trachomatis* load for individual infections allowed detailed analysis of load and disease severity data. There was a strong association between *C. trachomatis* ocular load and disease severity, in particular with the level of conjunctival inflammation. Geostatistical analysis suggests that ocular *C. trachomatis* load may be important in transmission, through demonstration of the presence of spatial clusters of high load infections and the absence of spatial clusters of low load infections.

Whole genome sequence data was obtained directly from clinical samples using cutting edge genome-wide DNA-bait sample enrichment technology for the first time in ocular *C. trachomatis* infection. Next generation Illumina sequence technology and stringent bioinformatics algorithms were applied to obtain high quality whole genome sequence data for comparative genomics analysis. This study includes the first pathogen GWAS for *C. trachomatis,* identifying putative genes and loci associated with disease severity and *C. trachomatis* ocular load. Pathogen GWAS affords a powerful analysis allowing for adjustment of population structure and linkage disequilibrium. The associations with conjunctival inflammation (*incE*) and *C. trachomatis* load (*mutY* and *CTA_0271*) highlight genes involved with specific biological characteristics of *C. trachomatis.* The functions of these molecules suggest that early interactions with host cells are important in *C. trachomatis* pathogenesis. This is consistent with results from host GWAS studies examining conjunctival scarring in trachoma which suggest that host genes and pathways involved in early interactions in intracellular parasitism are important determinants of later fibrotic conjunctival pathology [Ch Roberts, unpublished data]. Pathogen GWAS, particularly in conjunction with a detailed *in vivo* phenotype and accurate estimation of *C. trachomatis* load, is a powerful approach to identifying multiple targets for further study in pathogenesis and directed study of potential vaccine candidates, allowing a greater understanding of association and interaction of genes on a genome-wide scale.

As part of this study, the first mass distribution of azithromycin (MDA) campaign was conducted in conjunction with the Programa Nacional de Saúde de Visão. Following a single round of MDA the prevalence of active trachoma and ocular *C. trachomatis* were significantly reduced. Individual (and median) load of *C.*

trachomatis was reduced and the highest burden of disease and infection were concentrated in young children. Spatial clustering of infection identified using geostatistical tools was intensified following MDA, but clusters of high load infections were reduced. The absence of low load clusters was evident at baseline and following MDA. Disease severity was reduced following MDA, particularly in relation to conjunctival inflammation.

In summary, spatial clustering of high load ocular *C. trachomatis* infections and a strong association with severe inflammatory trachoma suggests the importance of chlamydial load in disease pathogenesis and transmission. Geospatial tools may be useful in trachoma surveillance in post-treatment settings to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission. The association between conjunctival inflammation and *C. trachomatis* load may reflect pathogen virulence, highlighted by the presence of associations with *C. trachomatis* load and conjunctival inflammation identified by pathogen GWAS. Further epidemiological and *in vitro* studies are required to provide a more complete picture of the relationship between disease severity and chlamydial load.

12.4 Chapter 12 References

[1] Thylefors B, Dawson CR, Jones BR et al. (1987). A simple system for the assessment of trachoma and its complicatios. Bull World Health Organ 65(4):477-83.

[2] Chidambaram JD, Lee DC, Porco TC, Lietman TM (2005). Mass antibiotics for trachoma and the Allee effect. Lancet Inf Dis 5(4):194-6.

[3] Bobo LD, Novak M, Munoz B et al. (1997). Severe disease in children with trachoma is associated with persistent Chlamydia trachomatis infection. J Infect Dis 176:1524-30.

[4] O'Connell CM, AbdelRahman YM, Green E et al. (2011). Toll-like receptor 2 activation by *Chlamydia trachomatis* is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by *C. trachomatis* but not by *C. muridarum*. Infect Immun 79:1044-1056.

[5] Russell M, Darville T, Chandra-Kuntal K et al. (2011). Infectivity acts as *in vivo* selection for maintenance of the chlamydial cryptic plasmid. Infect Immun 79:98–107.

[6] Carlson JH, Whitmire WM, Crane DD et al. (2008). The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect Immun 76:2273-83.

[7] Nunes A, Borrego MJ, Gomes JP (2013). Genomic features beyond *C. trachomatis* phenotypes: what do we think we know? Infect Genet Evol 16:392-400.

[8] O'Connell CM, Ingalls RR, Andrews CW Jr et al. (2007). Plasmid-deficient *Chlamydia muridarum* fail to induce immune pathology and protect against oviduct disease. J Immunol 179:4027-34.

[9] Comanducci M, Manetti L, Bini A et al. (1994). Humoral immune response to plasmid protein *pgp-3* in patients with *Chlamydia trachomatis* infection. Infect Immun 62:5491-7.

[10] Harris SR, Clarke IN, Seth-Smith HMB et al. (2012). Whole genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nature Genet 44(4):413-s1.

[11] Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585-93.

[12] Millman KL, Tavare S, Dean D (2001). Recombination in the MOMP gene but not the *omcB* gene of Chlamydiae contributes to serovar-specific differences in tissue. J Bacteriol 183:5997-6008.

[13] Andreasen AA, Burton MJ, Holland MJ et al. (2008). *Chlamydia trachomatis ompA* variants in trachoma: What do they tell us? PLoS Negl Trop Dis 2(9):e306.

[14] Carlson JH, Porcella SF, McClarty G (2005). Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and geniotropic strains. Infect Immun 73(10):6407-18.

[15] Mital J, Millar NJ, Dorward DW et al. (2013). Roles for chlamydial inclusion membrane proteins in inclusion membrane structure and biogenesis. PLoS ONE 8(5):e63426.

[16] Somboonna N, Wan R, Ojcius DM et al. (2011). Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L2) and D lineages. mBio 2(3):e00045-11.

[17] Bridges BA, Timms AR (1997). Mutation in *Escherichia coli* under starvation conditions: A new pathway leading to small deletions in strains defective in mismatch correction. EMBO J 16(11):3349-56.

[18] Clausen JD, Christiansen G, Holst HU, Birkelund S (1997). *Chlamydia trachomatis* uses host cell microtubule networks during early events of infection. Mol Microbiol 25(3):441-9.

[19] Tan C, Spitznagel JK, Shoe H-Z et al. (2006). The polymorphic membrane protein gene family of the chlamydiaciae. Chlamydia Genomics and

Pathogenesis. Ed: Bavoil PM, Wyrick PB. Horizon Bioscience; Norfolk, UK. pp 195-218.

[20] Thomson NR, Holden TF, Carder C et al. (2008). *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res 18(1):161-71.

[21] Li Z, Lu C, Peng B et al. (2012). Induction of protective immunity against *Chlamydia muridarum* intravaginal infection with a chlamydial glycogen phosphorylase. PLoS ONE 7(3):e32997.

[22] Burton MJ, Holland MJ, Faal N et al. (2003). Which members of a community need antibiotics to control trachoma? Conjunctival *Chlamydia trachomatis* infection load in Gambian villages. Invest Ophthalmol Vis Sci 44:4215-4222.

[23] West SK, Munoz B, Mkocha H et al. (2005). Infection with Chlamydia trachomatis after mass treatment of a trachoma hyperendemic community in Tanzania: a longitudinal study. Lancet 366:1296-1300.

[24] Lee JS, Munoz BE, Mkocha H et al. (2014). The effect of multiple rounds of mass drug administration on the association between ocular *Chlamydia trachomatis* infection and follicular trachoma in preschool-aged children. PLoS Negl Trop Dis 8(4):e2761.

[25] Last AR, Burr SE, Alexander ND et al. (2014). Spatial clustering of high load ocular *Chlamydia trachomatis* infection and disease severity in trachoma-endemic communities on the Bijagós Archipelago, Guinea Bissau. Emerg Infect Dis [under review].

[26] Yohannan J, He B, Wang J et al. (2014). Geospatial Distribution and Clustering of *Chlamydia trachomatis* in Communities Undergoing Mass Azithromycin Treatment. Invest Ophthalmol Vis Sci 55(7):4144-50.

Chapter 13 Future Studies

Work is ongoing to develop mathematical models (in collaboration with the University of California, San Francisco) to investigate the transmission dynamics of *C. trachomatis* and the effect of high load infections in these communities, based on the findings of the current study suggesting that *C. trachomatis* load is important in pathogenesis and transmission. Such models, in conjunction with geospatial tools may be useful as tools for trachoma surveillance in post-MDA settings to identify clusters of infection and thresholds of *C. trachomatis* bacterial load that may be important foci of transmission and may improve our understanding of disease pathogenesis and transmission. These methods may also have wider application in monitoring concurrent neglected tropical diseases that are targeted for elimination.

The associations with conjunctival inflammation and *C. trachomatis* ocular load evident on pathogen GWAS analysis warrant further study. Further validation of these findings will be established through sequence-typing over *incE* in a replication cohort in conjunction with detailed clinical and immunological metadata. A full examination of synonymous and non-synonymous SNPs in adjacent regions, including over the *incD-G* operon, will be conducted. There is ongoing work examining allele frequency and linkage disequilibrium across these SNPs in association with conjunctival inflammation. Operon-based analysis, gene interaction studies and epistasis can be examined *in silico* to explore these associations further. *In vitro* analysis expressing IncE from open reading frame clones will be conducted following results from this study.

Comparative genomics analysis will be undertaken looking for signatures of selection and towards the development of a high-resolution SNP-based typing tool, allowing for the meaningful definition of strain-dependent virulence and its subsequent use in studies into the pathogenesis and epidemiology of *C. trachomatis* infection.

Further epidemiological and *in vitro* studies using comparative pathogen genomics to examine these associations are required to fully understand the relationship between disease severity and chlamydial virulence and identify potential vaccine candidates. The molecular, geostatistical and bioinformatics tools employed in these study have further application in the study of *C. trachomatis* infection,

pathogenesis, transmission and disease surveillance in the context of MDA in trachoma-endemic communities.

Appendices

Appendix I. Preliminary Field Study (November 2011 & February 2012)

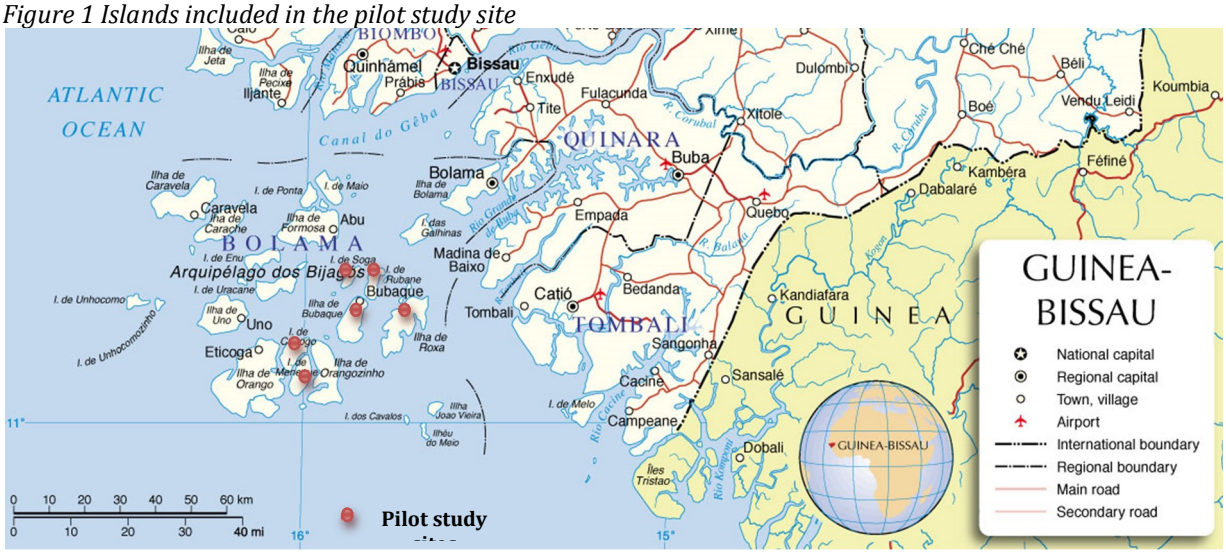
Aims

- To assess feasibility of the project both in the field and in the laboratory
- To establish trachoma endemicity
- To provide a rapid assessment of the patterns of active and scarring trachoma
- To demonstrate the presence of conjunctival infection with C trachomatis

Pilot Study Site

Islands of Bubaque, Canhabaque, Rubane, Soga, Meneque and Canogo of the Bijagos Archipelago of Guinea Bissau (*Figure 1*).

Figure 1 Islands included in the pilot study site



Pilot Study Sample Population

A convenience sample was drawn from village centres and schools in November 2010 and February 2011.

Methods

949 trachoma assessments were conducted, with conjunctival grades assigned using the WHO simplified grading system and conjunctival digital photographs taken on a convenience sample of all adults and children attending screening across 6 islands in 16 villages. Conjunctival swabs were taken from people with signs of active trachoma. Conjunctival sampling was well tolerated and accepted by participants in the survey. Optimum sample collection and storage methods in the field were established.

Samples were stored on ice in the field and transferred to a liquid nitrogen dewar within 8 hours. Samples were shipped in the dewar to the MRC Laboratories in The Gambia.

DNA was extracted from the conjunctival swabs using the QIAamp Mini DNA extraction kit (Qiagen). The Roche Amplicor® CT/NG PCR assay was used to detect presence of *C trachomatis* DNA.

Results

The age distribution of the sample is shown in *Figure 2*. There was abundant active disease (defined as the prevalence of TF and/or TI in children aged 1-9 years of age (prevalence 10-46%) with presence of *C trachomatis* infection in a significant proportion of cases (69.2%) with active disease (*Figures 3&4*). In those > 15 years of age, there were significant numbers of cases of conjunctival scarring (46%), trichiasis (8%) and corneal opacification (4%). Scarring was evident in children as young as 1 year, with several cases under the age of 10 years.

Figure 2 Age distribution within sample

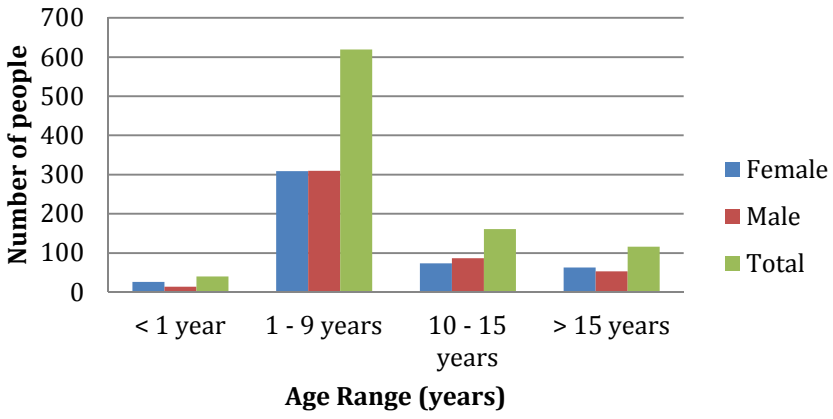


Figure 3 Number of cases of active trachoma with evidence of infection with C trachomatis (by PCR)

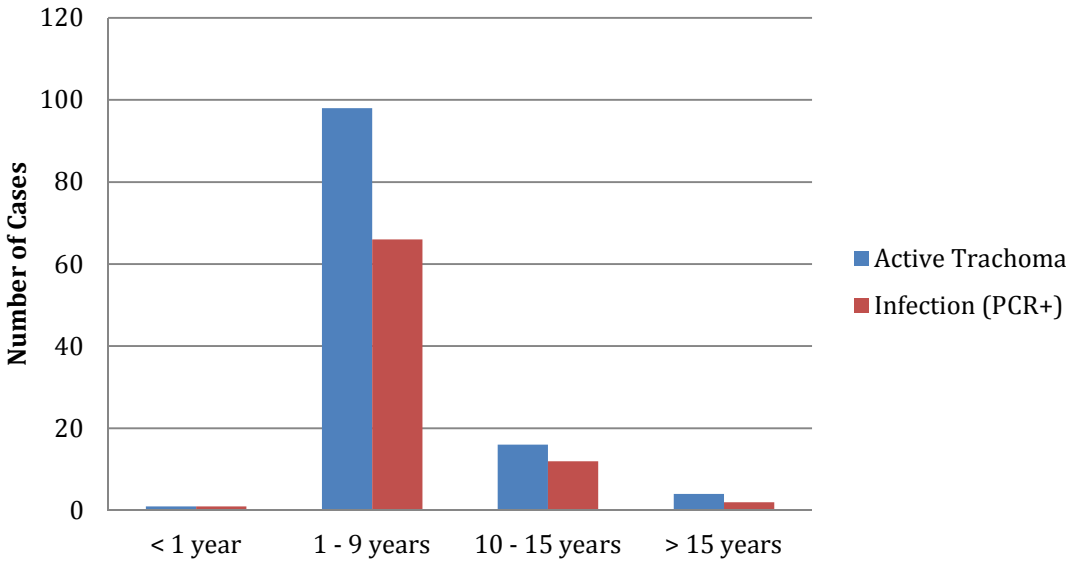
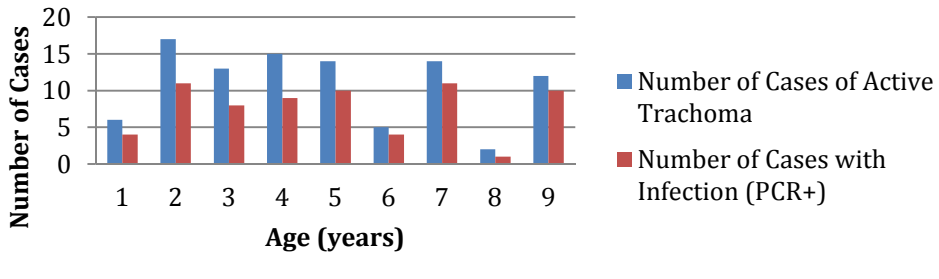


Figure 4 Cases of active trachoma with evidence of infection with C trachomatis by age (years)



There is also considerable variability in the number of cases of infection and the clinical severity of disease between age groups and location (*Figures 4&5*). This clustering does not appear to be related to ethnicity or gender and is not accounted for by typical risk factors for trachoma.

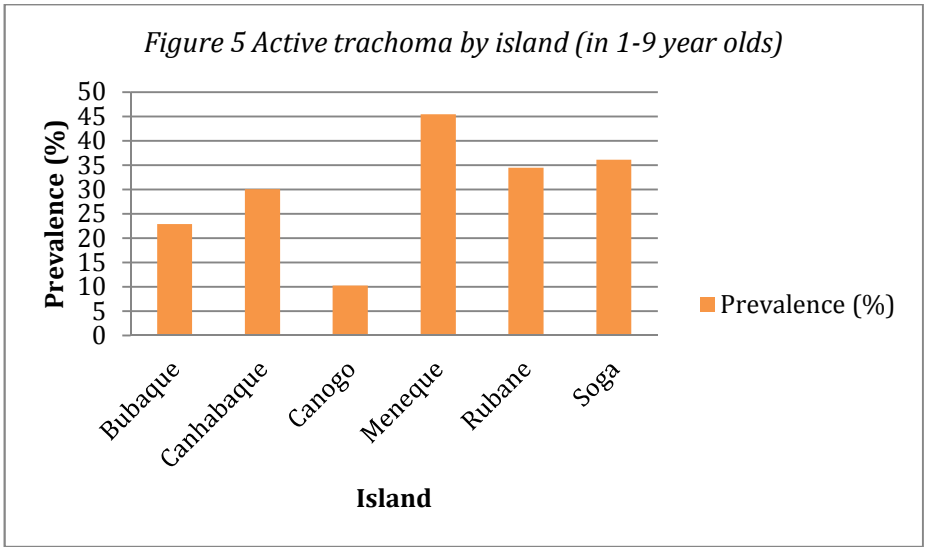


Figure 5 Active trachoma by island (in 1-9 year olds)

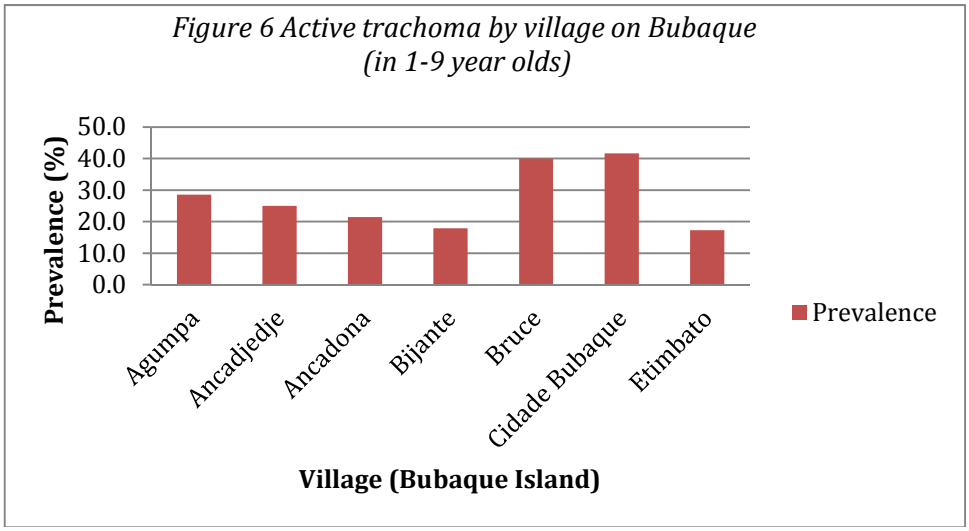


Figure 6 Active trachoma by village on Bubaque (in 1-9 year olds)

Conclusions

This pilot work demonstrates endemicity and significance of trachoma as a public health problem on the Bijagos Archipelago and served as an opportunity to refine field and laboratory protocols. It also highlights the variability in the clinical presentation of the disease by location thus confirming the rationale for the main study.

Appendix II. Household Head List

Formulário 2: Lista de Agregados Famíliares

Deixe em branco

Ilha _____ Código da Tabanca (TID)

Tabanca _____

Casa	Chefe de familia			Codigo da Casa (CID)
	Nome Próprio	Apelido	'Nome da Casa'	*Deixe em branco*
01				
02				
03				
04				
05				
06				
07				
08				
09				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				

39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				

86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				

Appendix III. Household Census Form

Formulário 3: Recenseamento

Tabanca _____ Chefe de Família _____
Código da tabanca (TID) __ __ Código da Casa (CID) __ __ / __ __ __
Há adultos na casa para completar esta ficha? Sim Não
Se Não, depois de 3 tentativas: **Desista**! Registre o nome da casa dos vizinhos e vá para a casa ao lado. Liste os nomes de todas as pessoas residentes no domicílio. Coloque o nome do chefe da família na primeira linha. Em seguida, o nome da mãe, seguido por seus filhos. ***Se houver uma criança com idade inferior a 10 anos, faça uma anotação do número de dois dígitos da mãe na última coluna. Notar gravidas com *.***

Número	Nome Próprio	Apelido	'Nome da Casa'	Sexo (M/F)	Idade (anos e mêses)	Data de Nacimento (dd/mm/aaaa)	Mãe (se < 10 anos)
01					ano mês		
02					ano mês		
03					ano mês		
04					ano mês		
05					ano mês		
06					ano mês		
07					ano mês		

08					ano mês		
09					ano mês		
10					ano mês		
11					ano mês		
12					ano mês		
13					ano mês		
14					ano mês		
15					ano mês		
16					ano mês		
17					ano mês		
18					ano mês		
19					ano mês		
20					ano mês		
21					ano mês		

22					ano mês		
23					ano mês		
24					ano mês		
25					ano mês		
26					ano mês		
27					ano mês		
28					ano mês		
29					ano mês		
30					ano mês		
31					ano mês		
32					ano mês		
33					ano mês		
34					ano mês		
35					ano mês		

36					ano mês		
37					ano mês		
38					ano mês		

Appendix IV. Ethical Approval Documents

LONDON SCHOOL OF HYGIENE
& TROPICAL MEDICINE

ETHICS COMMITTEE

APPROVAL FORM
Application number: 6026

Name of Principal Investigator	**Anna Last**
Faculty	**Infectious and Tropical Diseases**
Head of Faculty	**Professor Simon Croft**

Title: **Mapping Trachoma on the Bijagos Archipelago of Guinea Bissau, West Africa: A study of the pathogenicity of ocular strains of Chlamydia trachomatis using molecular and spatial determinants of trachoma**

This application is approved by the Committee.

Chair of the Ethics Committee .....

Date .....20 September 2011.....

Approval is dependent on local ethical approval having been received.

Any subsequent changes to the application must be submitted to the Committee via an E2 amendment form.

INSTITUTO NACIONAL DE SAÚDE PÚBLICA
Comité Nacional de Ética na Saúde

Nº Refª 135 /CNES/2011 Bissau, 30 de Setembro de 2011

Dra. Anna Last
Laboratórios MRC (Reino Unido), Fajara
Atlantic Boulevard
Caixa Postal 273, Banjul
Gâmbia

Projecto de Saúde de Bandim

BISSAU

ASSUNTO: Protocolo de Pesquisa Reformulado.

Com os melhores cumprimentos.

O Comité Nacional de Ética na Saúde reunido na sua sexta sessão ordinária do dia 30 de Setembro de 2011, ao analisar o Protocolo: "**Mapeamento do Tracoma no Arquipélago dos Bijagós de GNB usando técnicas especial e moleculares epidemiológicos**"- **Reformulado**, os membros presentes, decidiram aprovar a realização do estudo por se considerar haver cumprido com as exigências do CNES.

Com elevada consideração,

O Presidente,

Dr. Cunhate Na Bangna

A Instituto Nacional de Saúde Pública • Avenida Combatente da Liberdade de Pátria, Hospital "3 de Agosto" • Apartado: 861
Bissau, 1004 Bissau Codex • Administração TM: + 245 551 5150 • TM: (+ 245) 693 8511/590 4960.
E-mail: cunhate_nabangna@yahoo.com.br & djicoblama@hotmail.com.

Scientific Coordinating Committee
MRC Unit: The Gambia, Fajara
PO Box 273 Banjul, The Gambia
West Africa
Switchboard (+220) 4495442/6 Ext 2308
Fax (+220) 4495919/4496513
E-mail: scc@mrc.gm
Website (intranet): \\open.mrc.gm\home.asp

13 September 2011

Dr Anna Last
Faculty of Infectious and Tropical Diseases
Clinical Research Department
London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
UK

Dear Dr Last

SCC 1276v2, Mapping Trachoma on the Bijagos Archipelago of Guinea Bissau, West Africa: A study of the pathogenicity of ocular strains of Chlamydia trachomatis using molecular and spatial determinants of trachoma.

Thank you for resubmitting your revised proposal and response letter both dated 12 September 2011 incorporating the recommendations made by the SCC at its meeting held on 05 September 2011.

I am now pleased to approve your proposal which will be forwarded to the Ethics Committee for consideration at its meeting on 30 September 2011.

With best wishes

Yours sincerely

Professor Beate Kampmann
Chair, Scientific Coordinating Committee

Additional documents submitted for review:-

- CV – Anna Last
- Project Timeline

Appendix V. Consent Information and Forms

Information for Participants

Study: Mapping Trachoma on the Bijagos Archipelago of Guinea Bissau

Principal Investigator: Dr Anna Last

London School of Hygiene & Tropical Medicine (UK) and the Programa Nacional da Saúde Visao (PNSV) (Guinea Bissau)

What is trachoma and why is it important?

Trachoma is a common cause of blindness in many African countries, including Guinea Bissau. This disease is caused by infection (with a bacterium called chlamydia) which is passed from eye to eye. After people have had the infection many times, it can cause scarring of the upper eyelid, which pulls the eyelashes inwards, causing damage to the eye. This is very painful and in a significant number of cases results in blindness.

At certain stages of the disease it is possible to treat the infection using topical eye medication or antibiotics by mouth. In other studies conducted in Africa, we have discovered that treatment with an antibiotic called azithromycin works well to clear the infection even after a single dose. The World Health Organization (WHO) recommends that all communities where trachoma is a significant problem should be treated with this medication each year to eliminate trachoma. If the infection has already damaged your eyelids, you may need surgery to prevent blindness.

What is this study about?

We know that trachoma is a big problem on the Bijagos Islands. We do not know why there is such a big problem here. This study will help us to understand why this might be through identifying different strains of infection in relation to the disease in the community. The study will also help us learn how many people on the islands have infection, which will allow us to work with the Programa Nacional da Saúde Visao to offer you important treatment with antibiotics or surgery.

Who is conducting this study?

The London School of Hygiene & Tropical Medicine (UK) together with the Programa Nacional da Saúde Visao (Guinea Bissau).

Whose help do we need?

We need the help of approximately 2500 people living in communities on the islands.

What will we ask you to do?

We will conduct a census in your community. We will visit you in your community. If you agree to participate in the study, we will ask you some questions about yourself and where you live. We will then ask you if we can examine your eyes and take 2 samples (with cotton swabs) to test for the presence of infection. These swabs will be tested in the laboratory in The Gambia. We would also like to take a photo of each eye that we examine. We will return one year later and ask you to do the same again (eye examination, eye swabs and photos).

What are the benefits of participating in this study?

Trachoma is a huge problem on the Bijagos Islands. Trachoma can be treated effectively with the oral antibiotic, azithromycin. This treatment can cure the infection and prevent blindness caused by trachoma. Together with the PNSV

we will treat your communities with trachoma with antibiotics and provide surgery for those of you who need it to prevent blindness.

Are there any risks involved in participating in this study?

The eye examination is a little uncomfortable but will not harm your eye in any way. The examination takes only a few moments to complete. There are no risks associated with taking swab samples. This examination has been carried out in many countries without problem. Azithromycin is a safe medication and is used widely in Guinea Bissau and other African countries in the control of trachoma, as part of the work of the WHO and national eye care programmes to prevent blindness from this disease.

What tests will be done on the samples?

We will test the swabs in the laboratory in The Gambia and in the UK to improve our knowledge of the disease. The tests will show whether or not there is infection with chlamydia and will tell us how much infection is there.

What will happen to the records and photos we take of your eyes?

All the information that we collect is confidential. Only the organisers of the study will have access to them.

Do I have to take part in this study?

No, it is completely voluntary. It is your decision whether or not to participate. You may discontinue your participation in the study at any time without reason. This will not affect your access to health care or treatment for trachoma or any other disease in any way.

Thank you very much for considering your involvement in this important study. If you have any further questions, please ask us.

Dr Anna Last (Doctor and Study Coordinator) 00245 650 6554
Mrs Eunice Cassama (Ophthalmic Nurse) 00245 628 8282

Dr Meno Nabicassa, PNSV, Ministry of Public Health, Bissau
Dr Jose Nakutum, Regional Hospital ‘Marcelino Banca’, Bubaque

Informações para participantes
Estudo: Mapeamento do Tracoma no Arquipélago dos Bijagós na Guiné Bissau
Responsável: Dra Anna Last

Escola de Higiene e Medicina Tropical de Londres no Reino Unido com o Programa Nacional da Saúde da Visão (PNSV) da Guiné Bissau

O que é o tracoma e por que é importante?
O tracoma é uma causa comum de cegueira em muitos países Africanos, incluindo Guiné-Bissau. A doença é causada por uma infecção (por uma bactéria chamada clamídia) que se espalha de olho a olho. Depois que as pessoas tiveram a infecção muitas vezes, pode deixar cicatrizes na pálpebra superior, que puxa os cílios para dentro de forma a causar danos aos olhos. Isso é muito doloroso e muitas vezes leva à cegueira.

Em alguns estágios iniciais da doença, o tracoma pode ser tratado com pomada aplicada no olho, ou com comprimidos. Em estudos realizados em África, descobrimos que o comprimido antibiótico chamado azitromicina funciona bem até mesmo quando apenas uma dose é administrada. A Organização Mundial de Saúde (OMS) recomenda que todos os distritos onde o tracoma é comum tomem este medicamento a cada ano para eliminar tracoma . Se a infecção já prejudicou suas pálpebras, você pode precisar de cirurgia para prevenir a cegueira.

A respeito do que é este estudo?
Sabemos que o tracoma é um grande problema nas ilhas Bijagós. Nós não sabemos porque o problema é tão grande. A fim de compreender porque é que há um problema tão grande através da identificação de diferentes tipos de infecção com relação a doença de olho na comunidade. Nós também precisamos saber o número de pessoas infectadas na comunidade para que possamos trabalhar com o Projecto Nacional para a Saúde da Visão e poder dar a você um bom tratamento com comprimidos ou cirurgia.

Quem está fazendo este estudo?
Escola de Higiene e Medicina Tropical de Londres no Reino Unido juntamente com o Programa Nacional da Saúde da Visão (PNSV) da Guiné Bissau.

Da ajuda de quem precisamos?
Precisamos da ajuda de cerca de 2500 pessoas que vivem em comunidades nas ilhas.

O que vamos pedir-lhe para fazer?
Vamos fazer um recenseamento na sua comunidade. Vamos visitá-lo em sua comunidade. Se você concorda em participar do estudo, vamos fazer algumas perguntas sobre si mesmo e onde você vive.

Depois precisamos examinar seus olhos e tomar 2 amostras (com um cotonete) para testar a presença da infecção. As amostras serão testadas em laboratório na Gâmbia. Vamos tomar uma foto de cada olho que examinarmos.

Vamos voltar um ano mais tarde e pedir-lhe para fazer o mesmo novamente (exame oftalmológico, colheita de amostras e fotos)

Quais são os benefícios de participar deste estudo?

O tracoma é um grande problema nas ilhas Bijagós. O tracoma pode ser tratado eficazmente com o antibiótico azitromicina, um comprimido. Este tratamento pode curar a infecção e prevenir a cegueira por tracoma. Nós, juntamente com o PNSV, vamos a tratar comunidades com tracoma e fazer a cirurgia de triquíase para prevenir a cegueira.

Existem riscos causados por participar neste estudo?
O exame e a lavagem são apenas um pouco desconfortáveis mas não vão danificar o olho de maneira nenhuma. Vai demorar apenas poucos minutos. Não há riscos no processo de colheita de amostras. Já foi feita sem problemas em muitos países. Azitromicina é um medicamento seguro e é utilizada freqüentemente na Guiné-Bissau e outros países Africanos para o controle do tracoma, como parte do trabalho da OMS e programas nacionais para impedir a cegueira por esta doença.

Que testes serão feitos com as amostras?
Vamos testar as amostras em laboratório na Gâmbia e no Reino Unido para aumentar a nossa compreensão do tracoma. Os testes vão diagnosticar se há infecção por clamídia e verificar a quantidade e que tipo de clamídia está presente.

O que vamos fazer com os registos e as fotos que fazemos sobre seus olhos?
Todas as informações que coletamos serão confidenciais. Somente os organizadores do estudo terão acesso a elas.

Sou obrigado a participar deste estudo?
Não, a participação é completamente voluntária. É sua decisão participar deste estudo. Você pode negar-se a participar do estudo quando quiser sem dar explicação. E em caso de nao querer participar, isso nao afectara o seu relacionamento com o hospital ou seu tratamento de tracoma ou outras doenças de maneira nenhuma.

Muito obrigada por considerar estar envolvido neste trabalho importante. Se você tiver alguma dúvida, pergunte-nos.

Dra Anna Last (Médica e Coodenadora do Estudo) 00245 650 6554
Sra Eunice Cassama (Enfermeira ofthalmológica) 00245 628 8282

Dr Meno Nabicassa, Programa Nacional da Saúde da Visão, Ministério da Saúde Pública, Bissau

Dr Jose Nakutum, Hospital Regional da Bubaque ‘Marcelino Banca’

Form 5: Consent Form for Participants
Study: Mapping Trachoma on the Bijagos Archipelago of Guinea Bissau
London School of Hygiene & Tropical Medicine (UK) and the National Programme for Visual Health (PNSV) (Guinea Bissau)

Forename _____ Surname _____

HID _____ UID

Identification Sticker

I have read/had explained to me the information about the study. The Field Assistant has answered all my questions about the study.

I understand the information and will participate Yes No

Certification of Participant (or parent/guardian if child under 13 years of age)

Signature/Thumbprint _____

Name _____

Date _____

Certification of Field Assistant
I have explained the objectives of the study and I am satisfied that he/she understands and agrees to participate voluntarily.

Signature _____

Name _____

Date _____

Witness declaration if the participant is unable to read the consent information
I am witness to the study explanation and informed consent of the aforementioned participant .

Signature/Thumbprint _____

Name _____

Date _____

Formulário 5: Consentimento para Participantes

Estudo: Mapeamento do Tracoma no Arquipélago dos Bijagós na Guiné Bissau

Escola de Higiene e Medicina Tropical de Londres no Reino Unido com o Programa Nacional da Saúde da Visao (PNSV) da Guiné Bissau

Nome Próprio _____ Apelido _____

Número de identificação (UID) __ __ / __ __ __ / __ __ __ __

Eu li / recebi explicações sobre o estudo. O assistente respondeu todas as minhas perguntas sobre o estudo.

Eu entendo as informações e aceito participar Sim Não

Confirmação do Participante (Ou pai / responsável se a criança tiver menos de 13 anos)

Assinatura/Impressão digital _____

Nome _____

Data _____

Confimação do Assistente

Eu expliquei o objetivo do estudo e estou convencido de que ele/ela concorda em participar voluntariamente.

Assinatura _____

Nome _____

Data _____

Declaração de testemunha se o participante não puder ler as informações de consentimento

Eu sou testemunha das explicações fornecidas ao participante e do consentimento dado do participante acima mencionado para o estudo.

Assinatura/Impressão digital _____

Nome _____

Data _____

Appendix VI. Risk Factor Survey Questionnaire

Formulário 4: Inquérito Domiciliar

Tabanca _____ Código de Tabanca (TID) __ __
Chefe de Familia _____ Código de Casa (CID) __ __ / __ __ __
GPS Latitude _____Longitude _____

Há um adulto na casa para preencher esta ficha? Sim Não
Se não, depois de 3 tentativas: **Desista!** Vá para a próxima casa.
1.*O status sócio-econômico*
O grau de escolaridade do chefe de família?{1} Nenhuma{3} Elementar
{2} Primária {4} Secundária {5} Superior

A família escuta radio ≥ uma vez por semana {1} Sim {2} Não
A família possui / tem acesso a um meio de transporte? {1} Sim {2} Não

Detalhes _____
A família possui / tem acesso a um telefone móvel? {1} Sim {2} Não

Qual é a ocupação do chefe de família? {1} Agricultor {3} Sector Público
{2} Pescador {4} Sem trabalho
{5} Outra _____

Os seus filhos vão à escola? {1} Sim {2} Não
Se Sim, qual? Detalhes _____
Sobra dinheiro cada mês após a compra dos alimentos/essensiais?
{1} Sim {2} Não
Tem família na Bissau / no estrangeiro?
{1} Sim {2} Não
2.Latrinas
A família tem acesso a uma latrina?
{1} Sim {2} Não
Se Sim Marcá GPS: Lat _____Long _____
A família usa a essa latrina? {1} Sempre {2} Às vezes {3} Nunca

A latrina é comunitária ou privada? {1} Pública {2} Privada
A latrina está sendo usada? (**Observar**) {1} Sim {2} Não
Existe um caminho usada para a latrina? (**Observar**) {1} Sim {2} Não
Há moscas perto da latrina (≥2)? (**Observar**) {1} Sim {2} Não
Há fezes na latrina (visível com a lampada a mão)? (**Observar**)
{1} Sim {2} Não
A latrina está limpa? (**Observar**) {1} {2} {3} {4} {5}
MUITO LIMPA MUITO SUJA
3.Água
Distância da fonte de água regular Marcá GPS:
Lat _____Long _____
Quanto tempo demora para obter água (ida e volta)?

{1} menos que o tempo para cozinhar o arroz
{2} mais que o tempo para cozinhar o arroz
Tipo de fonte de água
{1} Furo {4} Torneira {7} Tanque {10} A água de superfície
{2} Poço {5} Rio {8} Fonte
{3} Bomba {6} Lago {9} Outro_____

A água é armazenada em casa? (**Observar**) {1} Sim {2} Não
Quantas vasilhas de água usa cada dia?
{1} Menos de Um {2} Um {3} Mais de Um
Detalhes _____
Há acesso à água durante o ano todo? {1} Sim {2} Não
Se Não, por quanto tempo? _____

4.Lixo
Há evidência de fezes humanas / animais a menos de 15m da casa? (**Observar**)
{1} Sim {2} Não
Há evidência de lixo solto a menos de 15m da casa? (**Observar**)
{1} Sim {2} Não
Há moscas no meio ambiente? (> 20 durante a entrevista - **Observar**)
{1} Sim {2} Não
Há animais que vivem perto da casa? (**Observar**)
{1} Sim {2} Não
Se Sim, quais? Detalhes _____

5.Educação da Saúde
Tem sido promovida alguma limpeza na localidade? {1} Sim {2} Não

6.Os residentes
Há alguns residentes no domicílio que passam > 1 mês longe de casa por ano?
{1} Sim {2} Não

Número de Residente*	Destinação e Razão	Duração (semanas, mêses, anos)	Número de vezes por ano

Quantas pessoas dormem em cada quarto?

Quarto	Adultos (mais de 15 anos)	Crianças menores de 10 anos	Crianças de 10 a 15 anos
1			
2			
3			
4			

Appendix VII. Ocular Assessment Forms

Formulário 6: Avaliação do tracoma (Primera Pesquisa)

Examinador/a _____ Assistente de Laboratório _____
Fotógrafo/a_____
TID __ __ CID __ __ / __ __ __
Número UID

__ __ / __ __ __ / __ __ __ __

Formulário de consentimento preenchido? **Sim** **Não**
Se **Não**, **DESISTA**! Preencha o formulário de consentimento.

Nome _____ Apelido___
_____

Idade	Ano	Mês	anos e mêses
Data de Nacimento			(dd/mm/aaaa)

Sexo Masculino / Feminino

Etnia {1} Bijagós {2} Balanta {3} Fula {4} Mandinka
{5} Manjaco {6} Papel {7} Falupe {8} Mancanha
{9} Mansoance {10} Nalu {11} Saracule {12} Outro

Local de nascimento _____
Local de infância _____
Passa mais de um mês longe de casa por ano? {1} Sim {2} Não
Vai à escola {1} Sim {2} Não
Qual? _____

Exame Clínico
Acuidade Visual

DIREITO	
ESQUERDA	

Classificação Facial {1} Nasal
{2} Ocular
{3} Moscas

Avaliação do Tracoma

DIREITO NF ☐ TF ☐ TIP ☐ TS ☐ TTC ☐ CO

AMOSTRA 'AIR' o AMOSTRA 'SPIKED'

Etiqueta Adesiva de AMOSTRA 1

Etiqueta Adesiva de AMOSTRA 2

Números dos fotos

1
2
3

ESQUERDA NF ☐ TF ☐ TIP ☐ TS ☐ TTC ☐ CO

AMOSTRA 'AIR' o AMOSTRA 'SPIKED'

Etiqueta Adesiva de AMOSTRA 1

Etiqueta Adesiva de AMOSTRA 2

Números

1
2
3

Formulário 6: Avaliação do tracoma (Segunda Pesquisa)

Examinador/a _____ Assistente de Laboratório _____
Fotógrafo/a_____
TID __ __ CID __ __ / __ __ __
Número UID

__ __ / __ __ __ / __ __ __ __

Formulário de consentimento preenchido? **Sim**
Não
Se **Não**, **DESISTA**! Preencha o formulário de consentimento.

Nome _____ Apelido___
_____

Idade

Ano	Mês

anos e mêses

Data de Nacimento (dd/mm/aaaa)

Sexo Masculino / Feminino

Exame Clínico
Acuidade Visual
DIREITO

ESQUERDA

Classificação Facial {1} Nasal
{2} Ocular
{3} Moscas

Avaliação do Tracoma

DIREITO N F TF TI P TS TT C CO

AMOSTRA 'AIR' o AMOSTRA 'SPIKED'

Etiqueta Adesiva de AMOSTRA 1

Etiqueta Adesiva de AMOSTRA 2

Números dos fotos

1
2
3

ESQUERDA N F TF TI P TS TT C CO

AMOSTRA 'AIR' o AMOSTRA 'SPIKED'

Etiqueta Adesiva de AMOSTRA 1

Etiqueta Adesiva de AMOSTRA 2

Números

1
2
3

Appendix VIII. Conjunctival Digital Photography Protocol

Body: Nikon D40/D60/D3000 Lens: VR AF S Micro Nikon 105mm 1:2.8G
SWM VR ED Micro IF 1:1 62 Requires 62mm diameter UV digital filter

General

- Make sure battery (and back up battery is charged)
- Make sure the filter lens protector is in place (and spare)
- Preserve batteries by switching camera off when not in use
- Check settings before use

Camera Settings

1. Insert battery and SD memory card
2. Turn camera on
3. Press 'Info' and check that battery is fully charged and that memory card is empty
4. Check that lens and filter are secure
5. Set to MANUAL FOCUS: 'M' on lens
6. Set to 'Full' on lens
7. Set lens to 1 1 0.314 (should show F57 on monitor)
8. Set the dial to aperture 'A' on the dial
9. Press 'Menu'. By pressing up and down arrow keys you can navigate the tabs on the left. Select the camera icon (SHOOTING MENU). Use the right arrow key and then up/down to select items from the shooting menu. Select IMAGE QUALITY -> OK -> JPEG Normal (select with arrow keys) -> OK. Select IMAGE SIZE -> OK -> Large -> OK. Select ISO SENSITIVITY -> 400 -> OK.
10. Exit by pressing 'Menu' until back to monitor ready to take picture.

Taking a Photo

1. Write the date and your name at the top of the photo log form
2. Write participant UID in space provided on photo log form
3. Make sure flash is on
4. Use the lens to focus to take picture of the UID
5. Then take 2 pictures of the left eye
   Focus in on the eye
   Make sure the conjunctiva is in the CENTRE
   Avoid glare/light artefact
   Try taking from different angles to avoid light artefact etc.
   Avoid gloved fingers
   Do not photograph the cornea
6. Repeat for the left eye
7. CHECK EACH PHOTO AS IT IS TAKEN. If it is not satisfactory delete using the trash can icon. Retake photos.

Appendix IX. Conjunctival Digital Photo Log Form

Formulário 7: Fotos

UID	Olho	Número de Foto (1) UID	Número de Foto (2) Olho	Número de Foto (3) Olho	Etiquetas Adesivas das Amostras
_ _ / _ _ _ / _ _ _ _	**Direito**				
	Esquerda				
_ _ / _ _ _ / _ _ _ _	**Direito**				
	Esquerda				
_ _ / _ _ _ / _ _ _ _	**Direito**				
	Esquerda				

Appendix X. Conjunctival Specimen Log Form

Formulário 8: Amostras

UID	Olho	Adesivas Amostra		UID	Olho	Adesivas Amostra
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	
_ _ / _ _ _ / _ _ _ _	Direito Esquerda			_ _ / _ _ _ / _ _ _ _	Direito Esquerda	

Appendix XI.

Standard Operating Procedure for DNA Extraction using the QIAxtractor Robot

Adapted from the MRC (The Gambia)

Standard Operating Procedure (SOP) for DNA extraction from biological samples using the QIAxtractor robot

Date: June 2010
Version: 1.02
Author: Alfred Ngwa (original SOP). Adapted by Alfred Ngwa and Anna Last for use with conjunctival swabs.
MRC The Gambia

Sample preparation for dry conjunctival swabs

- Prepare Phosphate Buffered Saline (PBS) (1xPBS)
- Prepare aliquots of 400µl PBS in 2.0 ml Simport tubes
- Label Simport tubes with PBS with unique identifier label that corresponds to original swab sample. ***Check that the numbers on the labels match.***
- Using a cocktail stick, carefully remove the swab from its original tube and place in the tube with PBS and close the tube. Retain the original tube. ***Check again that the labels match before inserting the swab into the tube with PBS.***
- Vortex swabs in PBS for 15 seconds at maximum speed.
- Using another cocktail stick, express the swab against the side of the tube and return it to its original tube. ***Check again that the labels are correct before inserting the swab into its original tube.***
- Record sample numbers processed and label the 10x10 box.
- Processed swab-PBS solution should be used immediately or stored at -70^{0}C until DNA extraction is required. Original swabs should be stored at -20^{0}C.

Pre-run digest and sample preparation (Adapted Whole Blood Protocol)

- Aliquoting samples into the round bottom block can be done in advance the night before and stored at 4^{0}C, but the digest should only be done on the day of extraction.
- If samples are frozen, thaw on ice and do a quick spin to collect sample at the bottom of tubes
- Remove all reagents from fridge and equilibrate to room temperature
    - Use DX Reagents and DX Plasticware (Qiagen)
- Prepare plate plan in 96-well format with sample IDs in lab book or on sample template (unique identifier labels can be used).
- Use 200µl of the swab-PBS solution for extraction on the QIAxtractor instrument
    - If nucleic acid yields are low, concentrating the sample by pelleting the swab rinse and resuspending in 200µl PBS may improve results
- Aliquot 200µl swab-PBS rinse into round bottom sample block according to plate plan. ***Check that the unique identifier numbers are correct.***
- Prepare the digest solution in a sterile 50ml falcon tube
    - 415µl DX Digest Enzyme
    - 30ml DXL
    - Mix gently by inverting 10 times
- Add 240µl of the digest solution to each well of the 96-well round bottom sample block containing samples. Be very careful not to contaminate adjacent columns of samples on the plate.
- Cover with a seal and incubate at 60^{0}C for one hour; gently shake to mix every 15 minutes.
- Centrifuge at 3000rcf for 5 minutes
- Transfer 220µl of supernatant to a clean lysis block (maintaining the same sample orientation as in the round bottom sample block).

Preparing the work space

- Whilst the sample digest is taking place, prepare the robot for the extraction run.
- Turn the PC on first. Then switch on the robot. Cancel through the vacuum extraction window.
- Wipe QIAxtractor components using in sequence: Milli-Q, 1% bleach and 70% ethanol in small quantities.
- Use the UV light to decontaminate for 15 minutes prior to use (lightbulb shortcut icon at top of computer screen.
- Observing sterile procedures set up the instrument deck
    - Load capture plate into position A1
    - Load elution plate (unopened)
    - Load lysis block into position B1
    - Maintain sample orientation for each block and label runs clearly
- Dispense reagent volumes into appropriate reagent troughs and cover securely with reusable lids
    - 108ml DX Wash (DXW) into 170ml reagent tub (position A1, reagent block R1)
    - 56ml DX Final Wash (DXF) into 170ml reagent tub (position B1, reagent block R1)
    - 60ml DX Binding Solution (DXB) (prepared by dissolving 0.6g DX Binding Additive in the DXB) into 270ml reagent tub (position c1)
    - Ensure that all reagent tubs are covered with appropriate lids securely

Starting a Run

- Start the control PC and click on 'Corbett' icon
- Switch on the QIAxtractor robot
- Double click 'Robotics 4' on the desktop
- Select 'Cancel' on the vacuum extraction wizard window
- Click on 'FILE' → 'OPEN' → 'MY DOCUMENTS' → 'MALARIA' → 'WHOLE BLOOD EXTRACTION PROTOCOL'
- In the program window click on 'Wizard' and select 'Vacuum DNA Extraction Wizard'
- Select columns to extract from and click 'Jump to End'
- Review protocol in right hand window and close the pop-up
- Load filter pipette tips
    - Right click tip box location and select 'all tips available'
- Click 'Control' and 'Start'
- Review check list (error reports in RED)
- Select all boxes and click 'OK' to start the run
- User intervention pause steps are added after each vacuum step. Take the required action and click 'OK' to continue run
- At the final user intervention pause (drying after DXF) dispense 8ml Elution Buffer (EB) into 70ml reagent tub, cover and load into position A2, reagent tub R1
- Remove the lid of the elution plate
- Close the hood and click 'OK' to continue
- At the end of the run, remove the elution plate, cover and centrifuge at 3000rcf for 1 minute
- Transfer the eluate into a pre-labelled 96-well V-bottomed sample storage plate, maintaining the same plate plan orientation. This results in approx 50µl eluate. Store at -20^{0}C in the long-term and 4^{0}C for short-term use
    - If more eluate is required, a manual elution using further EB on the elution plate and centrifugation, can be done

Post-Run Clean Up

- Dispose of plastic ware and liquid waste in accordance with laboratory guidelines for sample type and reagent hazard level
- Do not add bleach or acidic solution directly to solutions containing guanidine or extraction waste
- Thoroughly rinse carriages, waste sink and tip chute in cold tap water and allow to dry
- Separator tub and non-disposable tubs should be soaked in 1% bleach solution and rinse with Alpha-Q or Molecular Biology Grade RNAse-free water to ensure that they are DNA/RNA and DNA/RNAse free
- Clean with the UV light protocol
- Shut down PC and switch off robot

Yellow C f 01

Appendix XII. Standard operating procedure for: AMPLICOR CT/NG Test for the detection of *Chlamydia trachomatis* from ocular swabs

SOP Details:

Version: 2.0	**Date operational:**	**14 April 2009**
Author(s): Sarah Burr **Title:** Research Fellow, Trachoma	*Signature*	*Date*
		**16-07-2010**
Authorised by: Martin Holland **Title:** Senior Lecturer, Trachoma	*Signature*	*Date*
		16-07-2010
Ratifying Body:	Laboratory Management	
Ratifying Body Representative: Laboratory management committee	*Signature*	*Date*
Date to be reviewed:		
Current review changes: - Format adopted to MRC standard - Details on pooling of samples added - Section on responsibilities added		
Location:	Original hard copy with Laboratory Supervisor, Trachoma Hard copy with Laboratory technicians, Trachoma	
Confidential, unauthorised copying prohibited		

Review History:

First Operational:	13 February 2006			
Reviewed:	14 April 2009	16 Jul 2010		
Updated yes/no:	Yes	No		

Index of Contents

1. DEFINITIONS AND ABBREVIATIONS

Abbreviations

CT	*Chlamydia trachomatis*
MRC	Medical Research Council; representing Medical Research Council Laboratories (UK) Unit, The Gambia
MWP	Microwell Plate
NG	*Neisseria gonorrhoeae*
PRET	Partnership for the Rapid Elimination of Trachoma
SOP	Standard Operating Procedure

2. BACKGROUND

The AMPLICOR® CT/NG Test is an in vitro multiplex diagnostic test that can detect *Chlamydia trachomatis* (CT) and/or *Neisseria gonorrhoeae* (NG) in clinical specimens, including ocular swabs. This test utilizes polymerase chain reaction nucleic acid amplification and nucleic acid hybridization.

The AMPLICOR® CT/NG Test offers:

- High sensitivity without a loss of specificity;
- Increased sensitivity over traditional culture or DNA direct probe tests;
- An ideal screening application (high sensitivity in low prevalence populations, including asymptomatic and symptomatic patients); and
- Optional co-amplified internal control enabling inhibition control.

3. PURPOSE

This Standard Operating Procedure (SOP) describes the protocol for testing ocular swabs for the presence or absence

of *Chlamydia trachomatis* DNA using the commercially available AMPLICOR® CT/NG test.

4. SCOPE

This SOP applies to the testing of all ocular swabs collected in the framework of the Partnership for the Rapid Elimination of Trachoma (PRET) programme and to the testing of all internal and external quality control samples processed by the PRET team in The Gambia.

5. RESPONSIBILITIES

Laboratory Management is responsible for ensuring that time and resources are available to conduct the procedures as outlined in this SOP.

The Research Fellow is responsible for ensuring that all staff is properly trained in performing the test and that all staff are read and follow this SOP.

The Senior Lecturer may at any time perform a quality check on how the procedures are conducted, administered, and recorded.

6. HEALTH & SAFETY

All practices must be carried out in accordance with the Unit's health and safety regulations.

a. 6.1 Precautions

- Personal protective equipment (laboratory coat and gloves) is to be worn at all times.
- Safety cabinet is to be cleaned with 70% ethanol before work and with 2% hycolin followed by 70% ethanol after work.
- Guidelines for using Pre & Post PCR rooms are to be followed.

7. MATERIALS AND REAGENTS

Reagents are to be properly preserved and monitored for their purity. Store all reagents and plates at 2-8ºC, do not freeze.

The batch number and expiry date of all reagents used on a given day must be documented using form T4 Reagent batch and expiry log.

b. 7.1 CT/NG Specimen Preparation Kit

- CT/NG Urine Wash: Tris HCl Buffer, 300 mM Sodium Chloride, 0.1% Detergent
- CT/NG Lysis Buffer: Tris-HCL Buffer, 1% Solubilizer, 0.09% Sodium Azide.
- CT/NG Specimen Diluent: Tris HCL Buffer, 6mM Magnesium Chloride, 25 % Detergent, 0.05% Sodium azide.

c. 7.2 AMPLICOR® CT/NG Amplification Kit

- CT/NG Master Mix: Tris HCl Buffer, EDTA, 100 mM Potassium Chloride, Glycerol, 0.01% AmpliTaq®, 0.016% dUTP, 0.005% dATP, 0.005% dCTP, 0.005% dGTP, 0.01% Amperase, 0.00004% Biotinylated Primers SS01, SS02, CP24 and CP27, Sodium Azide.
- CT/NG Internal Control: Tris HCl Buffer, 0.001% Non-Infectious Plasmid containing *C. trachomatis* primer binding sequences and a unique probe binding region; 0.005% Poly rA RNA, EDTA, Amaranth dye, 0.05% Sodium azide.
- C. trachomatis Control: Tris HCl Buffer, 0.001% Non-infectious plasmid DNA containinf *C. trachomatis* sequences; 0.005% non-specific carrier DNA, 0.5% Detergent, EDTA, 0.05% Sodium Azide.
- N. gonorrhoeae Control: Tris HCl Buffer, Non infectious plasmid DNA containing *N. gonorrhoeae* sequences, 0.005% Non-specific carrier DNA, EDTA, 0.05% Sodium Azide.

d. 7.3 AMPLICOR® *Chlamydia trachomatis* Detection Kit

- CT MWP: Microwell Plate *Chlamydia trachomatis* coated with CT-Specific DNA Probe. 96 wells in one resealable pouch.
- [1] DN (Denaturation Solution): 1.6% Sodium Hydroxide, EDTA, Thymol Blue. **IRRITANT!**
- [2] CT/NG HYB (CT/NG Hybridization Solution): Sodium Phosphate Solution, 0.2% Solubilizer, 32.4% Sodium Thiocyanate. **HARMFUL!**
- [3] AV-HRP (Avidin-Horseradish Peroxidase Conjugate): Tris HCl Buffer, 0.001% Avidin-horseradish peroxidase

conjugate, Bovine Gamma Globulin, Emulsit 25, 0.1% Phenol1% ProClin® 150 preservative.

- [4A] SUB A (Substrate A): Citrate Solution, 0.01% Hydrogen Peroxide, 0.1% ProClin 150 preservative.
- [4B] SUB B (Substrate B): 0.1% TMB, 40% Dimethylformamide (DMF). **TOXIC!**
- [5] STOP (stop Reagent): 4.9% Sulfuric Acid.
- 10X WB (10X Wash Concentrate): 2% Phosphate Buffer, 9% Sodium Chloride, EDTA, 2% Detergent, 0.5% ProClin® 300 preservative.

8. PROCEDURES

All procedures are carried out in the Trachoma Lab (Whittle Building) and the Post PCR room as indicated below.

8.1 Principle of the Test

The AMPLICOR CT/NG test is based on PCR amplification of target DNA using biotinylated primers (CP24 & CP27). After PCR amplification, the amplicon is chemically denatured to form single stranded DNA by adding a Denaturation solution. Detection of the amplified product is done by ELISA, which is achieved by hybridization of the amplified products to MWP coated with oligonucleotide probes specific to the CT target (CP35) and detection of the probe-bound amplified products by colour formation.

An Internal Control (IC) is added in the Amplicor test as an option to check for inhibition of the test that may be due to certain substances present in the samples. The IC is noninfectious plasmid DNA and has a randomized internal sequence of similar length and base composition as the target sequence and in addition has a unique probe binding region that differentiates the IC amplification products from the target amplification products. The master mix used in the test contains biotinylated primers specific for the target and for the IC and therefore allows the IC to be co- amplified with the target DNA from samples. Detection is done on a separate MWP coated with oligonucleotide probes specific for IC (SK535). After completion of the ELISA, this results in colour formation in the MWP.

8.2 Specimen Preparation

8.2.1 DNA Extraction (to be done in the Safety Cabinet)

To prevent cross-contamination, always recap tubes and change gloves frequently.

- Using a sterile filter tip, pipette 500 µl AMPLICOR CT/NG lysis buffer into sterile DNA free 2 ml polypropylene screw cab tubes. Make sure tubes are properly closed to avoid any evaporation. Store tubes at 2-8ºC until use.
- Label the tubes containing the lysis buffer with the sample IDs to be extracted. Organize the samples in the same order in the racks as the labeled tubes with lysis buffer. This step can be done on the laboratory bench.
- Using a new, clean wooden stick, remove the specimen (swab) from the transport tube and place it into the tube containing lysis buffer. **Verify that the ID on the tube with lysis buffer is identical to the sample's ID.**
- Elute the DNA by vortexing the swab in lysis buffer for 10 seconds. Be sure to change gloves every 20 samples or whenever needed to avoid cross-contamination.
- Gently squeeze out the lysis buffer from swab by pressing it against the side of the tube. Replace the swab in the transport tube. Store the extracted swabs in a self-seal plastic bag. Label the bag with study number, date of extraction, and number of samples.
- Recap the tube containing eluted DNA and incubate for 15 min at room temperature.
- Change gloves and add 300 µl of AMPLICOR CT/NG Specimen Diluent to each lysed sample. Recap the tube.
- Mix by vortexing for 10 seconds. Incubate samples at room temperature for 5 min.
- Heat the prepared specimen, containing extracted DNA at 95°C for 10 min. Before proceeding to amplification, store extracted samples at 2-8°C for at least 12 hours to help in inactivating inhibitors. Samples should be tested within 7 days of extraction. Store DNA samples at -20°C if amplification will not be done within this period.

8.2.2 Preparation of controls (to be done in the Safety Cabinet)

Prepare NG controls first to avoid contamination with *C. trachomatis* DNA.

<u>Working NG Control</u>

- Pipette 1 ml of Specimen Diluent into a sterile 2 ml screw cap tube.

- Add 100 µl of NG positive control, recap and mix well by vortexing or pipetting. Label the tube as Working NG control.

Processed NG Control

- Pipette 250 µl of CT/NG lysis buffer into a sterile 2 ml tube. Add 250 µl of Working NG control, recap and mix by vortexing or pipetting.
- Incubate at room temperature for 10 min. Label the tube as Processed NG control.
- This control will be the negative control for the CT amplification and detection.

  Note: New working and processed controls must be prepared for each PCR set during a working day

Working CT Control

- Pipette 1 ml of Specimen Diluent into a sterile 2 ml tube screw cap tube.
- Add 100 µl of CT positive control, recap and mix well by vortexing or pipetting. Label tube as Working CT control.

Processed CT Control

- Pipette 250 µl of CT/NG lysis buffer into a sterile 2 ml tube. Add 250 µl of Working CT control, recap and mix by vortexing or pipetting.
- Incubate at room temperature for 10 min. Label the tube as CT processed control.
- This control will be the positive control for the CT amplification and detection.

  Note: New working and processed controls must be prepared for each PCR set during a working day.

8.2.3 Preparation of Master Mix (to be done on Pre-PCR bench)

- Using a sterile filter tip add 100 µl of CT/NG Internal control (Red) to CT/NG Master Mix (clear), recap the tube.
- Mix well by inverting 15 times. [3 master mix for each 96 samples].
- Aliquot 50 µl of master mix in the PCR strip tubes. Prepare only the required number of tubes. Label each strip of tubes. Any remaining amplification mix must be labeled with the date of preparation and stored at 2-8°C for further use, up to 7 days.

- Place the PCR tubes containing the master mix in a self-seal bag to be transferred to the area where the DNA samples are going to be loaded. In a separate bag, place the number of strip caps you will require to close the PCR tubes.

8.2.4 Pooling of samples (to be done in the safety cabinet)

As the expected prevalence of CT in The Gambia is low, samples are processed in pools.

- Label 92 sterile DNA-free 2 ml polypropylene screw cab tubes from 1 to 92.
- Arrange tubes containing extracted DNA samples into groups of five.
- Identify five samples making up a pool. Add 50 µl of each sample to the same 2 ml tube. Mix by vortexing or pipetting. Store pooled samples at 2-8°C. Pooled samples should be tested within 7 days of DNA extraction. Store DNA samples at -20°C if amplification will not be done within this period.
- Proceed to fill out document T1 Pool template indicating the arrangement of samples into pools of 5. THE TEMPLATE SHOULD REPRESENT THE EXACT ARRANGEMENT OF THE SPECIMENS IN THE 92 MICROTUBES. This will also represent the arrangement in further steps of amplification and in the 96 well plate used in the *Chlamydia trachomatis* detection.

8.2.5 Loading the DNA (to be done in the safety cabinet)

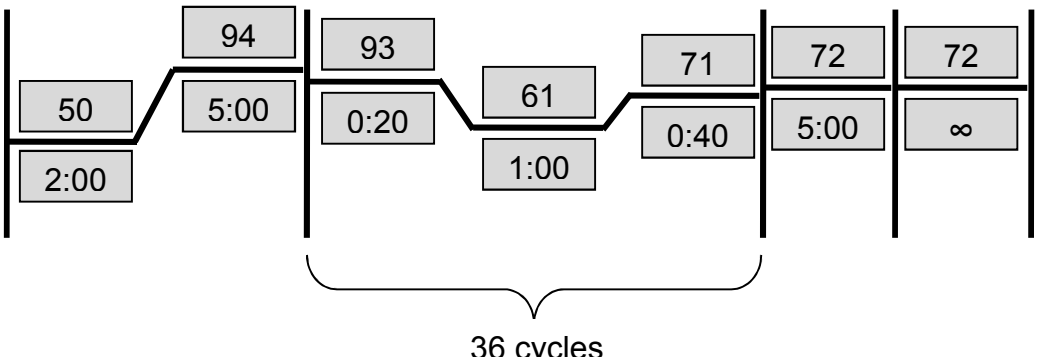
- Bring out the pooled DNA samples and organize them in a 96 position rack.
- If samples were stored at -20°C, they must be thawed. Samples are to be mixed thoroughly by vortexing and spun shortly to bring down any droplets on the cap or sides of tube. Make sure there is no precipitate present.
- Take **one** strip of tubes with master mix and transfer to a microtube rack.
- Add 50 µl of each pooled DNA sample to each tube containing master mix. Transfer the loaded strip to a new rack. Proceed loading **one** strip at a time. Change gloves whenever required.
- Once the samples have been loaded, proceed to load the controls in the microtubes containing the master mix. Load the Processed CT control first, change gloves and then load the Processed NG control. Ensure that both controls are in duplicate.

- Proceed to cap the strips one by one transferring them to the appropriate racks.

8.2.6 Amplification (to be done in Post PCR room)

Place the samples in the PCR system 9700 or CG1-96 thermal cycler for amplification and verify the Program conditions. Any change could interfere with the results.

Cycling conditions:



Hold program: hold at 72°C, not to exceed 24 hours.

8.2.7 Detection

Denaturation (to be done in Post PCR room)

All reagents and MWP are to be pre-warmed at room temperature for 10 min.

- Remove reaction tubes (one strip at a time) from thermocycler after amplification and remove the caps carefully to avoid aerosols.
- Immediately pipette 100 µl of Denaturation Solution DN (1) to the first row of reaction tubes using multichannel pipette with aerosol barrier tips. Repeat the process for each row using new aerosol barrier tips.
- Note: the Denaturation Solution must be added immediately after the amplified products are removed from the thermocycler and whilst still hot.
- Incubate at room temperature for 10 min to ensure full denaturation.
- Denatured amplicons can be kept at room temperature for a maximum of 1 hour, then proceed immediately to the hybridization step.

Hybridization (to be done in the laboratory)

- Remove the CT and IC MWP from the foil pouches. If less than 12 strips are required, remove the ones necessary for your test and place them in a clean frame. Label each strip with a permanent marker. Store the strips that are not going to be used in the pouch at 2-8ºC. Record the date the pouch was opened.
- Add 100 µl of Hybridization Solution CT/NG HYB (2) to each well in the CT and IC MWPs to be tested.
- Pipette 25 µl of the denatured amplicon to the appropriate wells in the CT and IC MWPs using a pipette with an aerosol barrier tip.
- Gently tap the plates 5 times until color changes from blue to light yellow to ensure adequate mixing had been done.
- Cover the CT and IC MWPs with clean MWP lids and incubate at 35-39°C for 1 hour.

Washing

- While the plates are incubating, prepare the Washing Buffer Working Solution by adding 1 volume of 10X Washing Buffer to 9 volumes of distilled water and mix well. Prepare at least 40 ml of working wash solution for each 8 well MWP strip. Working wash solution can be stored at 2-25°C in a clean closed plastic container up to 2 weeks.
- Empty the CT and IC plate contents and tap dry on 5 layers of paper towel. Fill the wells (400-450 µl), wait for 30 seconds, empty out contents and tap dry. Repeat process 4 more times. After all 5 washes, rinse the back of the plate with water to remove any excess wash buffer.

Conjugate

- Add 100 µl of Conjugate Solution AV-HRP (3) to each well in the CT and IC plate. Do not touch the walls of the wells.
- Cover the plates using clean MWP lids and incubate for 15 min at 35-39°C. If plates are not properly covered the content of the wells will evaporate.

Washing

- As described above. Make sure no traces of Conjugate Solution are left as it can lead to a false positive result. Adding an extra wash is recommended. Be sure to change gloves before adding the Substrate Solution.

Substrate

- Prepare the Substrate by mixing 2 ml of Substrate solution A (4A) SUB A with 0.5 ml of Substrate Solution B (4B)

SUB B as the working substrate for both CT and IC 8-well microwell plate strip (16 tests). For a full plate, prepare 12 ml (4A) Sub A to 3 ml (4B) Sub B. The reagent should be prepared no more than 15 min before use, stored at room temperature and in the dark.

- Add 100 µl of the Working substrate into each well.
- Keep the plate at room temperature (20-25°C) and in the dark for 10 min to allow color to develop.

Stop

- Stop the reaction by adding 100 µl of STOP Solution (5) to each well.
- Measure the absorbance at 450 nm using an ELISA plate reader.

8.3 Analysis of Results

A summary of results obtained for each PCR must be documented using form T3 Results summary sheet.

Validity of assay

- At least 2 replicates of the Working CT control and Working NG control should be included per test. The absorbance value of the Working NG control on the CT MWP (Negative control for the assay) should be less than 0.20 A_{450} units, and ideally around 0.06. IF THE RESULT IS ≥ 0.20 A_{450} UNITS THEN THE ASSAY IS NOT VALID AND MUST BE REPEATED. Fill out document T5 Failed run.
- The absorbance value of Working CT control on the CT MWP (Positive control for the assay) should be ≥2.0 A_{450} units. IF RESULT IS <2.0 A_{450} UNITS THEN THE ASSAY IS NOT VALID AND MUST BE REPEATED. Fill out document T5 Failed run.

Interpretation of results

CT Absorbance at 450 nm	IC Absorbance at 450 nm	Result Interpretation
< 0.2	≥ 0.2	*C. trachomatis DNA not detected.* *Sample **NEGATIVE** for C. trachomatis.*
≥ 0.8	ANY	*C. trachomatis DNA detected. Sample*

		POSITIVE *for C. trachomatis.*
≥0.2, < 0.8	≥0.2	EQUIVOCAL. Sample equivocal for *C. trachomatis.* Sample must be repeated neat (undiluted) in duplicate.
< 0.2	< 0.2	INHIBITORY. The amplification of DNA as not possible due to the presence of inhibitors. Sample must be repeated neat (undiluted). If sample is still inhibited it must be re-tested diluted (dilution 1:10). Dilution should be prepared in a fresh mixture v/v CT/NG lysis buffer and Specimen Diluent.

8.3.1 Negative samples

If a pool of samples is negative for *C. trachomatis* DNA according to the Interpretation of results (see above), all five samples making up that pool are considered negative.

8.3.2 Positive Specimens

If a pool of samples is positive for *C. trachomatis* DNA according to the Interpretation of results (see above), all five samples making up that pool must be repeated individually to identify the positive sample. Fill out document T2 Repeats template to indicate the arrangement of samples in the 96 well plate during the repeated test.

8.3.3 Equivocal Specimens

If a pool of samples is equivocal for *C. trachomatis* DNA according to the Interpretation of results (see above), all five samples making up that pool must be repeated individually to identify the equivocal sample. If a sample is equivocal when tested individually, it must be re-tested in duplicate. The sample should only be considered as positive if at least one of these repeats had an OD ≥ 0.8. If the repeats are equivocal or negative, the sample is considered negative. When repeating samples individually, fill out document T2

Repeats template to indicate the arrangement of samples in the 96 well plate.

8.3.4 Inhibitory Specimens

If a pool of samples is inhibited according to the Interpretation of results (see above), all five samples making up that pool must be repeated individually to identify the inhibited sample. If a sample is inhibited when tested individually, it must be re-tested diluted. When repeating samples individually, fill out document T2 Repeats template to indicate the arrangement of samples in the 96 well plate.

- Prepare a mixture v/v of CT/NG lysis buffer and Specimen Diluent.
- Prepare a dilution 1:10 from the neat sample (in a clean tube by mixing 10 µl of specimen DNA and 90 µl of mixture CT/NG lysis buffer and Specimen Diluent). Mix well.
- Re-test the samples following the standard CT/NG Amplification and Detection Protocols. If the result is still inhibited, re-test by preparing further dilutions (1:20, 1:50 and 1:100).

9. RECORDS

Records T1, T2, T3, T4 and T5 must be completed during the testing process as outlined above. If technician error renders a sample unusable, form T6 Void sample record must but be completed. If a sample is found to be ineligible, form T7 Ineligible sample record must be completed.

Attachments

T1 Pool Template
T2 Repeats Template
T3 Results summary sheet
T4 Reagent batch and expiry log
T5 Failed run record
T6 Void samples record
T7 Ineligible sample record

10. REFERENCES

AMPLICOR® *Chlamydia trachomatis/Neisseria gonorrhoeae* (CT/NG) Test User Manual. Copyright 2007, Roche Molecular Systems, Inc.

Appendix XIII. Standard Operating Procedure 7.01 Droplet Digital PCR

SOP 7.01
Droplet Digital PCR
Chlamydia Research programme
Version History

V.7.01 -13th July, 2012 Chrissy h. Roberts

SOP 7.01 Droplet Digital PCR

Summary

This protocol describes a method for droplet digital PCR of using the Biorad dx100 instrumentPCR products should be 22 µL, from which a 20 µL reaction will be used for droplet generationThe assay detects either the chlamydial plasmid or the chlamydial OMCB gene. Both implementationsuse the 30 kDa subunit of H.s. RNAseP gene as the endogenous control

Molecular targets

Endogenous control
gi|13937783|gb|BC006991.1| Homo sapiens ribonuclease P/MRP 30kDa subunit,
*Forward (HURNASE-P-F) 5'AGA TTT GGA CCT GCG AGC G 3'
*Reverse (HURNASE-P-R) 5'GAG CGG CTG TCT CCA CAA GT 3'
^Probe (HURNASE_HEX_BHQ1) 5'FAM-TTC TGA CCT GAA GGC TCT GCG CG-BHQ1-3'
AMPLICON and priming sites
agatttggacctgcgagcgggttctgacctgaaggctctgcgcggacttgtggagacagccgctc**********
****** ^^^^^^^^^^^^^^^^^^^^^^^^ **********
TARGET ONE : Chlamydia Trachomatis cryptic plasmid pLGV440 (X06707; SV 3; circular; genomic DNA; STD; PRO;
7500 BP.)
*Forward (Ct-Plasmid-F) 5'cagcttgtagtcctgcttgagaga3'
*Reverse (Ct-Plasmid-R) 5'caagagtacatcggtcaacgaaga3'
^Probe (Ct-plasmid-FAM-BHQ1) 5'HEX-ccccaccatttttccggagcga-BHQ1-3'
AMPLICON and priming sites
cagcttgtagtcctgcttgagagaacgtgcgggcgatttgccttaaccccaccatttttccggagcgagttacgaaga
caaaacctctt********** ^^^^^^^^^^^^^^^^^^^^^^^^ ****
cgttgaccgatgtactcttg**********
TARGET TWO : Chlamydia Trachomatis (Serovar A) OMCB gene.
*Forward (Ct-OMCB-F) 5'gacaccaaagcgaaagacaacac3'
*Reverse (Ct-OMCB-R) 5'actcatgaaccggagcaacct3'
^Probe (Ct-OMCB-HEX-BHQ1) 5'FAM-aagcaaaaaagcaagaaaaaaccacagcaaagag-BHQ1-3'

AMPLICON and priming sites

gacaccaaagcgaaagacaacacttctcataaaagcaaaaaagcaagaaaaaaccacagcaaagagactcccgt
agaccgtaaagaggt
********** ^^ ****
tgctccggttcatgagt
****

SOP 7.01 Droplet Digital PCR

Preparation of primers and probes

ALL STOCK SOLUTIONS should be at 100 µM

Primer/probe supermix A (CtPLASMID/HURNASEp)

2 µM Human RNAse P gene primers and probes
2 µM Chlamydia plasmid primers and probes

To prepare 1100 µL (enough for 500 reactions plus 10%)
volume
100 µM HURNASE-P-F 22 µL
100 µM HURNASE-P-R 22 µL
100 µM HURNASE_HEX_BHQ1 22 µL
100 µM Ct-Plasmid-F 22 µL
100 µM Ct-Plasmid-R 22 µL
100 µM Ct-Plasmid-FAM-BHQ1 22 µL
1X TE buffer (sterile/nuclease free) 968 µL

_____
total 1100 µL
Primer/probe supermix B (Ct_OMCB/HURNASEp)
2 µM Human RNAse P gene primers and probes
2 µM Chlamydia OMCB primers and probes
To prepare 1100 µL (enough for 500 reactions plus 10%)
volume
100 µM HURNASE-P-F 22 µL
100 µM HURNASE-P-R 22 µL
100 µM HURNASE_HEX_BHQ1 22 µL
100 µM Ct-OMCB-F 22 µL
100 µM Ct-OMCB-R 22 µL
100 µM Ct-OMCB-FAM-BHQ1 22 µL
1X TE buffer (sterile/nuclease free) 968 µL

_____
total 1100 µL

SOP 7.01 Droplet Digital PCR

PCR setup
WORK IN A PCR CLEAN & DNA FREE ENVIRONMENT

Each assay uses EITHER primer/probe supermix A OR primer/probe supermix B.
Composition per reaction

1X

2X ddPCR supermix 11 µL

Primer/probe supermix (2 µM each of six oligos) 2.2 µL
Nuclease free H2O 3.85 µL
SAMPLE DNA 4.95 µL
_____
total 22 µL
001 | To prepare sufficient ddPCR MASTERMIX to run one 96 well plate (100 reactions)
In A 2 mL, conical based microtube, add the following
100X
2X ddPCR supermix 1100 µL (2 X 550 µL)
Primer/probe supermix (2 µM each of six oligos) 220 µL
Nuclease free H2O 385 µL
_____
total 1705 µL
MOVE TO DNA LOADING/PCR CLEAN AREA
002 | To each well of a 96 well SKIRTED EPPENDORF twin-tec plate (in retainer), aliquot 17.05 µL ddPCR MASTERMIX
003 | Aliquot 4.95 µL of DNA from samples to each well of the plate.
004 | Seal with microamp cover film (ensure close seal) and vortex gently (1000 rpm) to mix reagents for 30 seconds
005 | Centrifugate at 1000 rcf for 1 minute to pool PCR mix at bottom of plate.
006 | Proceed to droplet generation

SOP 7.01 Droplet Digital PCR

Droplet generation

TURN ON THE HEAT SEALER
007 | Place a new, clean and EPPENDORF TWIN-TEC SEMI-SKIRTED 96 well plate in a retainer
008 | Carefully remove the microamp film, without disturbing the PCR mixes
009 | Place a new droplet generation (DG8) cartridge in the droplet cassette, the notch should face left (SEE FIGURE).

Close the cassette by applying gentle pressure to the short edges.

010 | Transfer 20 µL of each PCR mixture from row 1 in to each sample well of the cartridge (green on figure, marked
"sample" on the cassette).
Take care to avoid introducing air by putting pipette tips to the lowest point where the wall of the vessel meets the base.
Hold pipette at around 20 degree angle to wall, actuate slowly and raise tips from the base as the wells continue to fill
011 | Transfer 70 µL droplet generation oil to each oil well of the cartridge (red on figure, marked "oil" on the cassette")
012 | Attach a rubber gasket to the notched teeth on the cassette. Take care not to move the rubber across the wells as
this could lead to cross-contamination. Ensure that the four teeth are properly engaged with the gasket as failure to do thiswill prevent a good seal forming.
013 | Transfer the cassette to the droplet generation machine. Holding the cassette gently at the top and bottom, place on
the magnetised platform. A green light will indicate proper engagement.
014 | Activate droplet generation by pressing the button.
015 | Proceed to prepare the next cassette
016 | When the first droplets are complete, exchange the first cassette for the second in the droplet generator. Reactivate
the device
017 | Carefully transfer 45 µL of droplets from the droplet wells (blue on figure) to row 1 of the SEMI-SKIRTED plate.
Ensure that aspiration is performed slowly (around 10 s) when drawing and expelling droplets as failure to do this will leadto droplet instability
018 | Use a short length of orange lab-tape to temporarily seal the wells of row 1
019 | Prepare the droplet cassette for row 3, then transfer the droplets from row 2
020 | Continue this process until the plate is completed.
021 | Remove all orange lab-tape and replace with an easy-pierce, heat seal, ensuring that it is the right way up
022 | Place tray in the heat sealer, pull down the heat-block and press down again, hard for 5 seconds. Release, turn the
plate through 180 degrees, then repeat sealing for another 5 seconds.

023 | Put plate in Bio-Rad thermal cycler and run the program “DDPCR”

024 | When the cycle is complete, droplets can be stored for up to 48 hours, or counted immediately

DROPLET READING
If this is the first read of the day, start by performing a prime/flush/prime operation
025 | Transfer the plate to the special retainer and place this device in the droplet reader. A green light will indicate correct engagement with the platform. Close the door by pressing the button.
026 | Open the droplet reader software and load the appropriate template “plasmid” or “omcb”
027 | Select all wells on the plate and double click on the plate diagram to edit the details.
028 | For sample names, all wells should be labelled with the plate identifier number (1-16)
029 | Start the droplet count by pressing the run button

Appendix XIV. STROBE CHECKLIST

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-naïve Trachoma-hyperendemic Communities of the Bijagós Archipelago, Guinea Bissau (PNTD-D-13-02056)

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract ☑
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found ☑
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported ☑
Objectives	3	State specific objectives, including any prespecified hypotheses ☑
Methods		
Study design	4	Present key elements of study design early in the paper ☑
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection ☑
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants ☑
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable ☑
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group ☑
Bias	9	Describe any efforts to address potential sources of bias ☑
Study size	10	Explain how the study size was arrived at ☑
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why ☑
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding ☑
		(*b*) Describe any methods used to examine subgroups and interactions ☑
		(*c*) Explain how missing data were addressed ☑
		(*d*) If applicable, describe analytical methods taking account of sampling strategy ☑
		(*e*) Describe any sensitivity analyses **N/A**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed ☑
		(b) Give reasons for non-participation at each stage **N/A**
		(c) Consider use of a flow diagram **N/A**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential

		confounders ☑
		(b) Indicate number of participants with missing data for each variable of interest ☑
Outcome data	15*	Report numbers of outcome events or summary measures ☑
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included ☑
		(*b*) Report category boundaries when continuous variables were categorized ☑
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **N/A**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses ☑
Discussion		
Key results	18	Summarise key results with reference to study objectives ☑
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias ☑
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence ☑
Generalisability	21	Discuss the generalisability (external validity) of the study results ☑
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based ☑

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.