Spatial and genomic data to characterize endemic typhoid transmission

Jillian S. Gauld[1,2], Franziska Olgemoeller[3,4], Eva Heinz[12,3], Rose Nkhata[4], Sithembile Bilima[4], Alexander M. Wailan[6], Neil Kennedy[8,9], Jane Mallewa[10] Melita A. Gordon[4,5,10], Jonathan M. Read[2], Robert S. Heyderman[7], Nicholas R Thomson[6,11], Peter J. Diggle[2], Nicholas A. Feasey[3,4]

1. Institute for Disease Modeling, Bill and Melinda Gates Foundation, Seattle, USA
2. Centre for Health Informatics, Computing, and Statistics, Lancaster University, Lancaster, UK
3. Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK
4. Malawi-Liverpool Wellcome Programme, Blantyre, Malawi
5. Institute of Infection, Veterinary and Ecological Sciences, The University of Liverpool, Liverpool, UK
6. Wellcome Sanger Institute, Cambridge, UK
7. Division of Infection and Immunity, University College London, London, UK
8. Department of Paediatrics, University of Malawi the College of Medicine, Blantyre, Malawi
9. School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, UK
10. Adult Medicine, University of Malawi the College of Medicine, Blantyre, Malawi
11. Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine
12. Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK

Corresponding author: Jillian Gauld, jgauld@idmod.org, +18652992376, 500 5th Ave N, Seattle, WA 98109 USA

Summary: This study identified significant spatial heterogeneity of multi-drug resistant typhoid fever across Blantyre, Malawi, where the febrile disease has been endemic since 2011. Hydrological catchments were predictive of genomic patterns observed highlighting the role of environmental systems in ongoing transmission.

Abstract

Background

Diverse environmental exposures and risk factors have been implicated in the transmission of *Salmonella* Typhi, however, the dominant transmission pathways through the environment to susceptible humans remain unknown. Here, we utilize spatial, bacterial genomic, and hydrological data to refine our view of Typhoid transmission in an endemic setting.

Methods

546 patients presenting to Queen Elizabeth Central Hospital in Blantyre, Malawi with blood culture-confirmed typhoid fever between April 2015 and January 2017 were recruited to a cohort study. The households of a subset of these patients were geolocated, and 256 *S.* Typhi isolates were whole genome sequenced. Pairwise single nucleotide variant (SNV) distances were incorporated into a geostatistical modeling framework using multidimensional scaling.

Results

Typhoid fever was not evenly distributed across Blantyre, with estimated minimum incidence ranging across the city from less than 15 to over 100 cases/100,000/year. Pairwise SNV distance and physical household distances were significantly correlated (p=0.001). We evaluated the ability of river catchment to explain the spatial patterns of genomics observed, finding that it significantly improved the fit of the model (p=0.003). We also found spatial correlation at a smaller spatial scale, of households living <192 meters apart.

Conclusions

These findings reinforce the emerging view that hydrological systems play a key role in the transmission of typhoid fever. By combining genomic and spatial data, we show how multi-faceted data can be used to identify high incidence areas, understand the connections between them, and inform targeted environmental surveillance, all of which will be critical to shape local and regional typhoid control strategies.

Keywords: salmonella typhi; typhoid fever; genomics; spatial patterns; environmental transmission

Introduction

Water and sanitation improvements in the early 1900s led to effective elimination of typhoid fever in most high-income countries, without widespread use of either antibiotics or vaccines. However, typhoid remains a major cause of morbidity and mortality in low and middle-income countries, with an estimated 11 million cases occurring annually[1]. In March 2018, the typhoid conjugate vaccine (TCV) was recommended by WHO for control of typhoid, providing momentum for global initiatives to combat this disease[2]. Although this vaccine offers a high level of clinical protection[3], it does not completely prevent shedding of the disease[4]. Further, modeling studies have shown that elimination using a vaccine alone is unlikely[5]. Multi-faceted initiatives combining effective vaccines and detailed epidemiological surveillance, with water, sanitation and hygiene (WASH) interventions must therefore be developed[6].

Although provision of clean water and sanitation is a global priority and a Sustainable Development Goal, the pace of these changes have not aligned with TCV roll-out in typhoid endemic countries in Asia and Africa[7]. Targeted WASH strategies may support typhoid elimination on a more rapid timeline, however, targeting interventions is challenging, as typhoid transmission pathways do not appear to be consistent across locations[8,9].

Typhoid transmission can occur through two modes. First, “short-cycle” transmission is characterized by food or water contamination in close proximity. This transmission pathway is frequently linked to food handlers[10] and transmission within the household[11].

Next, “long-cycle” transmission occurs by contamination of and contact with the external environment[12]. Although a human-restricted pathogen, little is known about the behaviour of *Salmonella enterica* serovar Typhi in the environment between faecal excretion by one host and consumption by the next. Reservoirs for S. Typhi may exist in the environment, including amoebae, allowing for extended survival outside of the body[13]. Environmental mediators vary by location: the contamination of drinking water through stone taps in Kathmandu, Nepal has been hypothesized as a major driver of transmission in this setting[8], while in Santiago, Chile, produce contaminated with sewage was the key driver of transmission; drinking water was not considered a risk factor in this setting[9]. In Blantyre, Malawi, the use of river water for cooking and cleaning was identified as a risk factor for typhoid fever[14], providing further evidence for diverse environmental exposures to this pathogen.

Long-cycle transmission pathways are technically difficult to identify, partly because *S.* Typhi is challenging to isolate from environmental matrices[15]. In the absence of effective environmental surveillance for typhoid, we hypothesize that geolocation of cases and sequencing of their isolates will help elucidate our understanding of long cycle transmission pathways and environmental mediators of the disease..

In the current study, we investigate typhoid fever in Blantyre, Malawi, where a rapid emergence of multi-drug resistant *S*. Typhi occurred in 2011[16]. We utilize a geostatistical modeling framework to combine genomic information from isolates with geo-referenced hydrological, geographic, and demographic data. The current study’s cohort contained a

nested case-control study, which previously found a complex network of potential risk factors of typhoid fever related to both availability of WASH and the use of river water for cooking and cleaning, as well as social exposures such as daycare attendance[14]. Therefore, we specifically aimed to further investigate role of hydrological systems as a driver of transmission in this typhoid endemic setting.

Methods

Setting and case ascertainment

Queen Elizabeth Central Hospital (QECH) provides free secondary healthcare to the Blantyre urban area and surrounding district, as well as tertiary care to the southern region of Malawi. The Malawi-Liverpool Wellcome Programme (MLW) has been conducting sentinel surveillance of blood stream infections at QECH since 1998[17]. Patients living in Blantyre who were diagnosed with blood culture-confirmed typhoid fever between April 2015 and January 2017 at QECH were included in a prospective observational cohort. Age, residential area, HIV status, inpatient vs outpatient treatment, clinical presentation, complications and deaths were recorded from clinical case records and/or during patient interviews.

Geolocation and spatial analysis were restricted to patients residing in urban Blantyre. Residential location and location of any household water source used within three weeks prior to diagnosis were recorded using two methods. Starting in April 2015, the households of cases that were enrolled in the nested case-control study of children[14] were geolocated by a field team during household visits, using Garmin Etrex 30 GPS devices. Beginning in August 2015, the electronic PArticipant Locator application (ePAL), a tablet-based geolocation system,[18] was used to remotely geolocate the households for the remainder of the cohort.

Informed written consent was sought from adult participants and from the legal guardians of children.

Ethics statement

This study was approved by the University of Malawi, College of Medicine Research and Ethics Committee(P.08/14/1617), the Liverpool School of Tropical Medicine Research Ethics Committee(14.042), and the Lancaster University Faculty of Health and Medicine Ethics Committee(FHMREC17014).

Incidence mapping

We estimated the minimum incidence of typhoid fever associated with presentation to QECH for the enumeration areas (EAs) of Blantyre. Population estimates were derived from a 2016 census of Blantyre and surrounding areas that divided the city into 275 EAs[19]. This census included population structure by age bands (<5, 5-14, 15+ years of age), and the number of households. Numerous approaches to adjusting incidence of typhoid fever based on health care utilization, the probability of a diagnostic being performed, and its sensitivity have been proposed[20–22], however these were not available at the EA level. Thus we only adjusted for the study recruitment rates as a proportion of all diagnosed cases. Finally, a healthcare utilization survey was administered to a subset of controls from the nested case-control study[14].

All statistical analyses were conducted using R statistical software, version 3.5.1.[23] In order to estimate incidence across the city, a Poisson log-linear model with a spatial random effect was fitted using the PrevMap package[24]. Rates were estimated for each EA and age band, and estimated population size in each age band was included as an offset. The offset was weighted to account for the longer time period for geolocation of the case-control study

participants vs. the rest of the cohort. Covariate effects were explored, including distance from the centroid of each EA to QECH, average household size, population density, elevation at the centroid, and hydrological catchment[Supplementary Material 1]. The statistical model is further described in Supplementary Material 2.

Sequencing & Phylogenetic analysis

In order to investigate the genetic patterns of *S*. Typhi in this study, isolates of *S*. Typhi from the index cases were whole genome sequenced on Illumina HiSeq 10X machines generating 150 bp paired-end reads. Reads were mapped against the high-quality reference genome of *S.* Typhi 1036491 isolated in Blantyre, Malawi in 2012 (GCA_001367555.3). The pairwise single nucleotide variant (SNV) matrix was generated from this alignment.

The phylogeny was based on the same alignment and reconstructed using iq-tree[25] under the general time-reversible model, ascertainment correction, and with 1000 bootstrap replicates for branch support. The resulting tree was assessed for phylogenetic signal using tempest (v1.5.1) [26] and root-to-tip correlation was calculated. The phylogenetic tree was reconstructed into a maximum-likelihood joint ancestral reconstruction tree. Detailed protocols for genomic analyses are found in Supplementary Material 3, with accession numbers listed in Supplementary Material Table 2.

Spatio-genomic modeling

First, we tested for correlation between SNV distance and spatial distance. Next, the pairwise distance matrix of all absolute differences of SNVs was mapped to two dimensions using principal coordinates analysis (PCoA), a method of multidimensional scaling (MDS). PCoA takes as its input an m-by-m matrix of pairwise dissimilarities among m entities, in this case SNVs, and looks for a graphical representation of these m entities that best approximates this

complete set of dissimilarities by straight-line distances in a low-dimensional space, typically one- or two-dimensional for ease of visualization. We refer to the resulting principal coordinate axes generated from this analysis as genetic scores. The values of these scores themselves are not directly interpretable; rather, the closer together a set of SNVs are on the graphical representation, the stronger their genetic relatedness. We used a linear model with a spatial random effect to predict genetic score across the city. Finally, based on results of a previous risk factor study from a subset of this cohort that pointed to river water as a potential exposure,[14] we explored the ability of river catchment to predict genetic score.

Results

Characteristics of cohort

S. Typhi was isolated from 658 blood cultures between March 28, 2015 and January 12, 2017 (Figure 1), with an additional 2 isolates obtained from cerebrospinal fluid. 97%(641/660) of all isolates were multidrug-resistant to ampicillin, chloramphenicol and cotrimoxazole. Four patients were identified as having relapse or reinfection, with a second episode of *S.* Typhi bacteremia diagnosed 38 to 84 days after their first positive blood culture. 546 illness episodes were included in the cohort study and are further referred to as cases. 484 cases lived in the area of urban Blantyre and were therefore eligible for mapping. 314 cases consented to provide their household locations and 256 isolates were available to be sequenced (Figure 1).

The characteristics of the cohort are summarised in Table 1 and presented in detail in Supplementary Material Table 1. The median age was 11 years (IQR:6-19), and HIV seroprevalence was 10.7%(37/346). 73%(391/542) of patients were hospitalized, and hospital records were retrieved for 326. Case-fatality in the cohort was 1.5%(8/528).

Incidence mapping in Blantyre

297/658 cases both consented to provide their household locations and lived within EA bounds and hence were included in the geostatistical incidence model. We adjusted resulting incidence rates for the number of cases who originated from urban Blantyre but were not included in mapping(115) declined participation(58) or were lost upon follow-up(116). Estimated incidence in each EA is plotted in Figure 2. The model predicted the highest risk in the 5-14 age band, followed by the 15+ age band (Table 2). Of the evaluated covariates, average household size was a significant predictor of incidence (Table 2), with smaller household sizes indicative of a higher risk. Other tested covariates, including distance from hospital, did not significantly improve the model(Supplementary Material 2).

Incidence across Blantyre was geographically heterogeneous, with an overall estimated minimum incidence rate of 36 per 100,000 per year. Eight EAs were predicted to have an incidence greater than 100 per 100,000 whilst 13 EAs had an incidence rate of less than 15 per 100,000. Small-scale spatial correlation was identified in the region, with the range (φ) of this correlation estimated to be 318 meters (Table 2). This indicates a practical range of spatial correlation (>5%) reaching approximately 950 meters (Table 2).

Analysis of the healthcare utilization survey found no evidence that distance to hospital affects healthcare seeking behavior for individuals included in this study. We found no statistically significant difference between households that would or would not utilize QECH in the case of severe illness, based on the distance of their household to QECH(p=0.275).

Genomic epidemiology

Isolates of 256 patients were whole-genome sequenced, and all belonged to the H58 haplotype which has been identified globally, indicating sustained endemic transmission of this new strain in Blantyre. The isolates were highly clonal, but could still be resolved into six phylogenetic clades[Supplementary Material 3](Figure 3A). Root-to-tip correlation (0.07) indicated insufficient temporal signal to allow a temporal analysis[Supplementary Material 3].

Significant correlation between SNV distances between isolates and physical distance of household locations was observed in the data (p=0.001)[Supplementary Material 4]. The first two principal coordinates (PCs) resulting from the PCoA accounted for 38% of the variation in the SNV matrix (Figure 3B). Scores along primary axis (PC 1) were similar for the majority of the cohort, with the exception of 11 isolates whose genetic score was approximately -15. This cluster of similar genetic scores suggests a distinct genetic group (highlighted in Figure 3B), and were contained within clade 5 in the tree (Figure 3A). However, these isolates did not differ spatially or temporally compared to the the rest of the cohort[Supplementary Material 4]. These isolates may indicate infections from a chronic carrier of *S.* Typhi, or importations from another location with endemic H58.

PC 1 showed no evidence of spatial correlation[Supplementary Material 4], so although this axis reflects an aspect of genetic relatedness, it cannot be explored using a geostatistical framework. However, genetic scores were more evenly distributed along PC 2 (Figure 3B), indicating heterogeneity across this representation of the genetic signal. Further, there appeared to be spatial correlation of these scores approaching 2500 meters. This can be visualized with an empirical semivariogram (Figure 3C), which represents the variance of pairs of genetic score of isolates as a function of distance between them. This observation

was confirmed by statistical test[Supplementary Material 4]. We therefore fitted the linear geostatistical model to PC 2.

Blantyre has a complex river network (Figure 4A), with ten hydrological catchments containing the cases identified (Figure 4B). Using hydrological river catchment as a categorical predictor in the linear model significantly improved the model's fit to the genomic patterns observed compared to an intercept-only model(LL -301.9 vs. -289.4, p=0.003). Parameter estimates indicated similar genetic scores for individuals in catchments 2 and 8 (Table 3), distinct from the rest of the catchments. To confirm this observation, we conducted a contrast test to compare the mean coefficient values between catchments 2 and 8, and the rest of the river catchments. The difference was significantly different from zero as evaluated using a t-test ($p<0.001$).

Estimates of spatial correlation of the geostatistical model highlighted the multiple scales of spatio-genetic clustering. The range parameter (φ) indicated that the practical range of spatial correlation (>5%) was approximately 192 meters, indicating the model's spatial random effect was capturing short-distance spatial correlation. Although the city's geographical range spans approximately 20 kilometers, households in the cohort are clustered. 59% of the cohort has another cohort member within a distance of 192 meters, and 13 households that were geolocated reported more than one case. We conducted a sensitivity analysis using geolocated water sources instead of household locations. As the majority of individuals lived within close proximity of their water sources, the results were consistent with the findings using household location[Supplementary Material 4].

Discussion

Geolocating cases as a part of routine surveillance has become increasingly common, allowing for spatially informed disease control such as for the targeting Polio vaccines[29], and investigating hot-spots and transmission routes of Ebola and SARS-CoV-2[30]. Geospatial analyses for typhoid fever to-date have revealed the spatially heterogeneous nature of the disease at both municipal and national scales[31,32], but generalizable and epidemiologically relevant incidence covariates have yet to be identified[33]. Here, we demonstrated the utility of using statistical modeling to place multidimensional data sets in the context of a classical observational study, and in doing so provide deeper insight into transmission hot spots of *S*. Typhi. Initial mapping of the cohort revealed heterogeneity in incidence rates of typhoid across the city, and genomic analyses provided greater resolution of generalized endemic transmission of the disease, following a rapid emergence of a multi-drug resistant strain in 2011[16].

Smaller average household size was significantly predictive of incidence rate in the current study. The increased incidence with decreasing household size, after accounting for population density and controlling for differential age distributions of EAs, may be a consequence of younger families living in greater socioeconomic precarity, but requires further investigation. Other studies have found increased risk in lower elevation areas[34], which was not identified as a significant covariate in this study.

While reconstruction of phylogeny was successful in identifying discrete clades of *S*. Typhi, much granularity was lost by reducing the total genetic diversity present in 256 isolates to 6 categorical clades. Having excluded variable regions, we incorporated the full spectrum of remaining genetic variation by starting with all-against-all SNV distance followed by multidimensional scaling, which created a continuous variable for further analysis. Modeling

the principal coordinates extracted from the SNV-matrix enabled us to view a continuous representation of genetic relatedness spatially, as well as test the predictive power of spatial covariates. A significant correlation between spatial and genetic distance was subsequently found, showing typhoid fever patients living closer together were more likely to have *S.* Typhi isolates with closely related genomes. These findings support observations of generalized endemic transmission of typhoid fever, in contrast with a point source outbreak. Within this context, this study further demonstrated that in an endemic region like Malawi there is enough resolution, even within a single haplotype of *S.* Typhi, to infer likely hot-spots of transmission.

Blantyre is delineated and divided by a complex river network, and extrapolating from elevation maps, we can identify the numerous hydrological catchments across the city, areas within which any water or sewage will eventually converge to the same location. Adding river catchment to our spatial-genomic model resulted in a significantly improved fit. The importance of river catchment has been also been proposed in Fiji, where heterogeneity of disease incidence between hydrological sub-catchments was found[35]. The distinct differences in genomic patterns that appear to be delineated by hydrological catchments indicate that these hydrological zones may act as a unit of epidemiological mixing, which is plausible given the environmental component of long cycle typhoid fever transmission. This is important when thinking about how potential herd effects across a city may be interpreted after a vaccination event, or how to target environmental surveillance to fully characterize a geographic region.

Small-scale spatial correlation also existed in the model even accounting for river catchment, which could reflect the "short-cycle" pathway of typhoid transmission, and/or a common environmental exposure through attendance at school or daycare, as identified as a risk in our

nested case-control study[14]. Work to disentangle the underlying source of these exposures would help further elucidate these complex pathways.

There are limitations to this analysis. Our single tested spatial covariate in the spatial-genomic analysis, grounded in previous work, was hydrological catchment. We currently lack predictors reflecting broader social interactions such as school attendance or food-sharing amongst households, which might also have contributed to spatial clustering seen. We conducted a sensitivity test that revealed similar spatial-genomic patterns when using water source location vs. household location, however, this does not sufficiently address the heterogeneity in potential exposures outside of the home. Finally, we only recruited a subset of typhoid patients, those seeking care at hospital; analysis of bacterial genomes generated by active case-finding or community-based surveillance that detects milder or subclinical cases would provide even deeper insight into *S.* Typhi transmission.

Currently, typhoid conjugate vaccines are being introduced in areas with known typhoid transmission; however, parallel WASH interventions are likely to be necessary for typhoid elimination, and identifying intervention points for these measures remains a challenge. Here, we have paired spatial and genomic data to provide further evidence of the role that rivers may play in typhoid transmission. Previous work has indicated a risk with utilizing river water for cooking and cleaning[14]. When paired with the current study, these findings may be utilized by targeted WASH-based public health messaging and interventions to control typhoid through a focus on ensuring all water, not just drinking water, is safe for household use. The next step will be to extend and validate these findings by direct detection of *S.* Typhi in these systems. Therefore, the development of methods for sensitive detection of *S.* Typhi in the environment will be critical to support the planning of public health interventions to interrupt transmission in the future.

NOTES

Acknowledgements: We would like to thank the staff and patients of Queen Elizabeth Central Hospital that supported and or consented to be recruited to this work. We acknowledge expert informatics support from the Pathogen Informatics team at the Wellcome Sanger Institute.

Funding: This work was supported by Bill & Melinda Gates Foundation Investment OPP1128444 and Wellcome Programme Grant 206454. Institute for Disease Modeling is a research group within, and solely funded by, the Bill and Melinda Gates Foundation.

Potential Conflicts: JMR reports support from UKRI Wellcome, paid to their institution; reports personal fees/consulting fees from Centra Technology; reports receiving Payment for taught course to Ministry of Health, Saudi Arabia, paid to their institution; reports serving as Member of SPI-M-O subgroup of SAGE, working to advise UK government on coronavirus pandemic (unpaid), all outside the submitted work.

References

1. Stanaway JD, Reiner RC, Blacker BF, et al. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect Dis **2019**; 19:369–381.

2. World Health Organization. Typhoid vaccines: WHO position paper, March 2018 – Recommendations. Vaccine. 2019; 37:214–216.

3. Jin C, Gibani MM, Moore M, et al. Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. Lancet **2017**; 390:2472–2480.

4. Gibani MM, Voysey M, Jin C, et al. The impact of vaccination and prior exposure on stool shedding of salmonella typhi and salmonella paratyphi in 6 controlled human infection studies. Clin Infect Dis **2019**; 68:1265–1273.

5. Pitzer VE, Bowles CC, Baker S, et al. Predicting the Impact of Vaccination on the Transmission Dynamics of Typhoid in South Asia: A Mathematical Modeling Study. PLoS Negl Trop Dis **2014**; 8:40.

6. Stanaway JD, Atuhebwe PL, Luby SP, Crump JA. Assessing the feasibility of typhoid elimination. Clin Infect Dis **2020**; 71:S179–S184.

7. Soble A, Patel Z, Sosler S, Hampton L, Johnson H. Gavi support for typhoid conjugate vaccines: Moving from global investments to country introduction. Clin Infect Dis **2020**; 71:S160–S164.

8. Karkey A, Jombart T, Walker AW, et al. The Ecological Dynamics of Fecal

Contamination and Salmonella Typhi and Salmonella Paratyphi A in Municipal Kathmandu Drinking Water. PLoS Negl Trop Dis **2016**; 10.

9. Shuval HI. Investigation of typhoid fever and cholera transmission by raw wastewater irrigation in Santiago, Chile. Water Sci Technol **1993**; 27:167–174.

10. Hussein Gasem M, Dolmans WMVWMV, Keuter MM, Djokomoeljanto RR. Poor food hygiene and housing as risk factors for typhoid fever in Semarang, Indonesia. Trop Med Int Heal **2001**;

11. Vollaard AM, Ali S, Van Asten HAGH, et al. Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. J Am Med Assoc **2004**; 291:2607–2615.

12. Crump JA. Progress in Typhoid Fever Epidemiology. Clin Infect Dis **2019**; 68:S4–S9.

13. Douesnard-Malo F, Daigle F. Increased persistence of Salmonella enterica Serovar typhi in the presence of Acanthamoeba castellanii. Appl Environ Microbiol **2011**;

14. Gauld JS, Olgemoeller F, Nkhata R, et al. Domestic river water use and risk of typhoid fever: Results from a case-control study in Blantyre, Malawi. Clin Infect Dis **2020**; 70:1278–1284.

15. Nair S, Patel V, Hickey T, et al. Real-time PCR assay for differentiation of typhoidal and nontyphoidal Salmonella. J Clin Microbiol **2019**; 57.

16. Feasey NA, Gaskell K, Wong V, et al. Rapid Emergence of Multidrug Resistant, H58-Lineage Salmonella Typhi in Blantyre, Malawi. PLoS Negl Trop Dis **2015**; 9.

17. Musicha P, Cornick JE, Bar-Zeev N, et al. Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. Lancet Infect Dis **2017**; 17:1042–1052.

18. Harris RC. Informing development strategies for new tuberculosis vaccines: mathematical modelling and novel epidemiological tools. 2017. Available at: http://researchonline.lshtm.ac.uk/4648987/.

19. MacPherson P, Khundi M, Nliwasa M, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med **2019**; 17.

20. Crump JA, Youssef FG, Luby SP, et al. Estimating the incidence of typhoid fever and other febrile illnesses in developing countries. Emerg Infect Dis **2003**; 9:539–544.

21. Carey ME, MacWright WR, Im J, et al. The surveillance for enteric fever in Asia project (SEAP), severe typhoid fever surveillance in Africa (SETA), surveillance of enteric fever in India (SEFI), and strategic typhoid alliance across Africa and Asia (STRATAA) population-based enteric fever st. Clin Infect Dis **2020**; 71:S102–S110.

22. Antillon M, Saad NJ, Baker S, Pollard AJ, Pitzer VE. The Relationship between Blood Sample Volume and Diagnostic Sensitivity of Blood Culture for Typhoid and Paratyphoid Fever: A Systematic Review and Meta-Analysis. J Infect Dis **2018**; 218:S255–S267.

23. 3.5.1. RDCT. A Language and Environment for Statistical Computing. R Found Stat Comput **2018**; 2:https://www.R-project.org. Available at: http://www.r-project.org.

24. Giorgi E, Diggle PJ. PrevMap: An R package for prevalence mapping. J Stat Softw **2017**; 78.

25. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–274.

26. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol **2016**; 2:vew007.

27. Pupko T, Pe'er I, Shamir R, Graur D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol **2000**; 17:890–896.

28. Wailan AM, Coll F, Heinz E, et al. rPinecone: Define sub-lineages of a clonal expansion via a phylogenetic tree. Microb Genomics **2019**; 5:1–9.

29. Mercer LD, Safdar RM, Ahmed J, et al. Spatial model for risk prediction and sub-national prioritization to aid poliovirus eradication in Pakistan. BMC Med **2017**; 15.

30. Lau MSY, Dalziel BD, Funk S, et al. Spatial and temporal dynamics of superspreading events in the 2014-2015 West Africa Ebola epidemic. Proc Natl Acad Sci U S A **2017**; 114:2337–2342.

31. Osei FB, Stein A, Nyadanu SD. Spatial and temporal heterogeneities of district-level typhoid morbidities in Ghana: A requisite insight for informed public health response. PLoS One **2018**; 13:e0208006. Available at: http://dx.plos.org/10.1371/journal.pone.0208006. Accessed 11 February 2019.

32. Dewan AM, Corner R, Hashizume M, Ongee ET. Typhoid Fever and Its Association with Environmental Factors in the Dhaka Metropolitan Area of Bangladesh: A Spatial and Time-Series Approach. PLoS Negl Trop Dis **2013**; 7.

33. Antillón M, Warren JL, Crawford FW, et al. The burden of typhoid fever in low- and middle-income countries: A meta-regression approach. PLoS Negl Trop Dis **2017**; 11.

34. Akullian A, Ng'eno E, Matheson AI, et al. Environmental Transmission of Typhoid

Fever in an Urban Slum. PLoS Negl Trop Dis **2015**; 9.

35. Jenkins AP, Jupiter S, Mueller U, et al. Health at the Sub-catchment Scale: Typhoid and Its Environmental Determinants in Central Division, Fiji. Ecohealth **2016**; 13:633–651.

Tables

Table 1. Characteristics of the recruited cohort of typhoid fever patients presenting to Queen Elizabeth Central Hospital in Blantyre, Malawi.

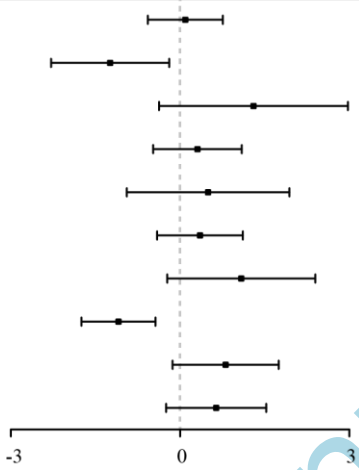
Characteristic	Value
Age, median years (range)	11 (6-19)
Female, n (%)	256/542 (47.2)
Malaria test positive, n (%)	7/533 (1.3)
Living in urban Blantyre, n (%)	484/542 (89)
Admitted, n (%)	391/542 (72.1)
Length of hospital stay, median (IQR)	4 (3-7)
Death, n (%)	8/520 (1.5)

Table 2. Parameter estimates for the geostatistical model of typhoid fever incidence in Blantyre, Malawi. This model includes the significant tested covariate of average household size, and is stratified by age band.

Parameter	**Estimate**	**Standard error**	**P value**
Intercept	-5.02	0.60	<0.001
Average household size	-0.95	0.14	<0.001
Age 5-14	1.08	0.04	<0.001
Age <5	0.75	0.04	<0.001
σ^2	0.45	0.11	..
φ	318	0.17	..
τ^2	0.21	0.23	..

Table 3. Estimated parameters for the geostatistical model of genetic scores, representing the genetic heterogeneity in *Salmonella* Typhi isolates. Hydrological catchment was included as a covariate in the model.

Catchment	Genomic score	95% CI
1	0.091	
2	−1.240	
3	1.301	
4	0.308	
5	0.496	
6	0.353	
7	1.084	
8	−1.092	
9	0.807	
10	0.640	



Parameter	Description	Estimate	Std. error
$sigma^2$	spatially correlated variance	4.116	1.106
phi	range of spatial correlation	40.496	1.119
tau^2	nugget (nonspatial) variance	0.165	1.859

*editable illustrator file attached separately

Figure legends

Figure 1. Consort chart outlining the process of recruiting individuals to the study, reasons for exclusion, as well as data availability for geographic and genomic data.

Figure 2. Estimated minimum annual incidence rate of typhoid fever for enumeration areas across the city of Blantyre, Malawi. The location of the recruiting hospital, Queen Elizabeth Central Hospital, is noted.

Figure 3. A. Joint ancestral state reconstruction tree based on whole-genome SNV phylogenetic analysis for the sequenced isolates of *S.* Typhi, showing the major clades of isolates determined via a root-to-tip directional approach. B. Further resolution of variation provided by decomposition of SNV matrix into the first two principal coordinates of the multidimensional scale, points colored by membership of major clades corresponding to the tree. C. Empirical semivariogram (proportional to one minus spatial correlation as a function of distance) of PC 2 of the SNV decomposition.

Figure 4. A. Major rivers of Blantyre with points indicating the approximate household locations of typhoid fever cases, B. Approximate household locations are colored by genetic score, and river catchmentsare outlined using polygons. .Catchments 2 and 8 are highlighted in yellow. Catchments not included in analysis are in grey. Precise locations of households are masked by randomization, and overlapping points have been jittered for visualization.