

RESEARCH ARTICLE

Spatial-temporal and phylogenetic analyses of epidemiologic data to help understand the modes of transmission of endemic typhoid fever in Samoa

Michael J. Sikorski^{1,2,3}, Jianguo Ma⁴, Tracy H. Hazen^{1,3}, Sachin N. Desai², Siaosi Tupua⁵, Susana Nimarota-Brown⁵, Michelle Sialeipata⁵, Savitra Rambocus⁶, Susan A. Ballard⁶, Mary Valcanis⁶, Robert E. Thomsen⁵, Roy M. Robins-Browne^{6,7}, Benjamin P. Howden⁶, Take K. Naseri⁵, Myron M. Levine^{2,3}, David A. Rasko^{1,3*}

1 Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **2** Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **3** Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **4** Department of Geographical Sciences, University of Maryland, College Park, Maryland, United States of America, **5** Ministry of Health, Government of Samoa, Apia, Samoa, **6** Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia, **7** Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

* drasko@som.umaryland.edu



OPEN ACCESS

Citation: Sikorski MJ, Ma J, Hazen TH, Desai SN, Tupua S, Nimarota-Brown S, et al. (2022) Spatial-temporal and phylogenetic analyses of epidemiologic data to help understand the modes of transmission of endemic typhoid fever in Samoa. *PLoS Negl Trop Dis* 16(10): e0010348. <https://doi.org/10.1371/journal.pntd.0010348>

Editor: Mariangela Bonizzoni, Università degli Studi di Pavia, ITALY

Received: March 22, 2022

Accepted: September 15, 2022

Published: October 17, 2022

Copyright: © 2022 Sikorski et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw sequence data have been submitted to GenBank and accession numbers of individual isolates are listed in [S1 Table](#). Assembled genomes are available on Figshare (doi: [10.6084/m9.figshare.18665686](https://doi.org/10.6084/m9.figshare.18665686)).

Funding: This work was supported by a grant [OPP1194582 / INV-000049] from the Bill & Melinda Gates Foundation (M.M.L.). Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has

Abstract

Salmonella enterica serovar Typhi (*S. Typhi*) is either widely distributed or proximally transmitted via fecally-contaminated food or water to cause typhoid fever. In Samoa, where endemic typhoid fever has persisted over decades despite water quality and sanitation improvements, the local patterns of *S. Typhi* circulation remain unclear. From April 2018–June 2020, epidemiologic data and GPS coordinates were collected during household investigations of 260 acute cases of typhoid fever, and 27 asymptomatic shedders of *S. Typhi* were detected among household contacts. Spatial and temporal distributions of cases were examined using Average Nearest Neighbor and space-time hotspot analyses. In rural regions, infections occurred in sporadic, focal clusters contrasting with persistent, less clustered cases in the Apia Urban Area. Restrictions to population movement during nationwide lockdowns in 2019–2020 were associated with marked reductions of cases. Phylogenetic analyses of isolates with whole genome sequences ($n = 186$) revealed one dominant genotype 3.5.4 ($n = 181/186$) that contains three Samoa-exclusive sub-lineages: 3.5.4.1, 3.5.4.2, and 3.5.4.3. Variables of patient sex, age, and geographic region were examined by phylogenetic groupings, and significant differences ($p < 0.05$) associated genetically-similar isolates in urban areas with working ages (20–49 year olds), and in rural areas with age groups typically at home ($< 5, 50+$). Isolates from asymptomatic shedders were among all three sub-lineages. Whole genome sequencing provided evidence of bacterial genetic similarity, which corroborated 10/12 putative epidemiologic linkages among cases and asymptomatic shedders, as well as 3/3 repeat positives (presumed relapses), with a median of one single

already been assigned to the Author Accepted Manuscript version that might arise from this submission. Research support was in part from federal funds from National Institutes of Health under National Institute of Allergy and Infectious Diseases grants F30AI156973 (M.J.S.) and U19AI110820 (D.A.R.), as well as National Institute of Diabetes and Digestive and Kidney Diseases training grant T32DK067872 (M.J.S.). M.M.L. is supported in part by the Simon and Bessie Grollman Distinguished Professorship at the University of Maryland School of Medicine. R.M.R.-B. reports grants from Bill & Melinda Gates Foundation, and non-financial support from the Government of Samoa during the conduct of the study. The funders had no role in the design and conduct of the study, collection, management, analysis, and interpretation of the data, preparation, review, or approval of the manuscript, and decision to submit the manuscript for publication.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: M.M.L. has a patent entitled "Broad spectrum vaccine against typhoidal and nontyphoidal *Salmonella* disease" (US 9,011,871 B2) issued. R.M.R.-B. reports non-financial support from the Government of Samoa during the conduct of the study.

nucleotide polymorphism difference. These findings highlight various patterns of typhoid transmission in Samoa that differ between urban and rural regions as well as genomic subtypes. Asymptomatic shedders, detectable only through household investigations, are likely an important reservoir and mobile agent of infection. This study advances a "Samoa *S. Typhi* framework" that supports current and future typhoid surveillance and control efforts in Samoa.

Author summary

Many typhoid endemic countries have evident transmission contributions from widely distributed contaminated water supplies and/or asymptomatic *S. Typhi* carriers who intermittently cause sporadic outbreaks. However, these patterns have not yet been examined in the island nation of Samoa, where typhoid has remained endemic for decades. In this study, we incorporated the discerning powers of spatial-temporal cluster analyses as well as phylogenetics of whole genome sequences (WGS) of *S. Typhi* isolates to examine detailed epidemiologic data collected through household investigations of culture-confirmed cases of typhoid fever occurring in Samoa from April 2018 through June 2020. We detected patterns consistent with both modes of transmission, varying between urban and rural regions, and we provided evidence of intra-household transmission of genetically similar isolates, thereby supporting a majority of putative epidemiologic linkages made during household investigations, and identifying important roles for asymptomatic shedders of *S. Typhi*. These findings advance our understanding of persistently endemic typhoid fever in Samoa and directly support the efforts of the Samoa Typhoid Fever Control Program of the Ministry of Health of Samoa.

Introduction

Typhoid fever is a potentially fatal human host-restricted infection caused by ingestion of *Salmonella enterica* serovar Typhi (*S. Typhi*) via contaminated food or water. Typhoid incidence, reported as cases per 100,000 people per year, is stratified into low (<10), moderate (10–100), high (>100–<500), and very high (≥ 500), with multiple different transmission mechanisms capable of maintaining typhoid fever at different incidence levels [1]. Widespread distribution of fecally-contaminated indirect sources (called "vehicles"), such as water supplies [2,3], crops irrigated with raw sewage [4,5], or processed foods [6,7], results in amplified, high and very high incidence transmission of *S. Typhi*. Sporadic, focal outbreaks of typhoid fever are typically associated with asymptomatic shedding of *S. Typhi* from a temporary (weeks to months) [8] or chronic (lifelong) carrier, such as "Typhoid Mary" [9]. These asymptomatic shedders inadvertently contaminate food that is consumed by contacts in close proximity, such as household members or attendees at gatherings (e.g. picnics and celebrations) [10].

Samoa, an island nation in the Polynesian sub-region of Oceania (the continental grouping of thousands of islands comprising 14 countries and 9 dependencies in the Central and South Pacific Ocean), has endured persistent endemic typhoid fever for decades despite improvements in drinking water quality, sanitation infrastructure, and economic development [11–14]. We have previously described endemic typhoid in Samoa by person, place, and time using

blood culture surveillance data from 2008–2019 [14] and characterized the population structure and evolutionary origins of a collection of *S. Typhi* genomes from Samoa spanning 1983–2020 [15]. However, it remains unclear to what extent endemic typhoid in Samoa is sustained by widespread, continuous distribution of an unknown fecally-contaminated vehicle (e.g., municipal water supplies) or focal, sporadic clusters caused by ingestion of food contaminated by temporary or chronic shedders. Differing in spatial and temporal distribution, these transmission cycles can be examined using geospatial point-pattern analyses [16,17]. Additionally, when paired with data from epidemiologic investigations, whole genome sequencing (WGS) analyses can correlate epidemiology with phylogeny, infer patterns of infection by genetic subtype, and quantify relatedness between isolates to evaluate suspected transmission events [18].

In Samoa, teams from the Samoa Typhoid Fever Control Program (STFCP) of the Ministry of Health (MoH) perform epidemiological investigations of the households of every culture-confirmed typhoid case to ascertain clinical status, environmental risks, prevalence of *S. Typhi* among household contacts, and evidence of epidemiologic association to other cases. In this study, we combined epidemiologic data collected through household investigations from April 2018 through June 2020 with geospatial point-pattern statistical tools and the high resolution discernment of WGS. We examined the spatial and temporal distribution of typhoid cases, as well as the sex, age, and geographic occurrences of infections by phylogenetic groupings. Putative epidemiologic linkages identified through case investigations and rare repeat positive isolates from the same individual were examined for genomic similarity and provided primary data to inform a phylogenetic threshold metric of relatedness among endemic *S. Typhi* strains. Our findings can help elucidate how *S. Typhi* circulates in Samoa and support the further integration of WGS into the epidemiologic surveillance and control efforts of typhoid control programs.

Methods

Ethics statement

This study received ethical clearance from the Health Research Committee of the MoH of Samoa, as well as The University of Maryland, Baltimore (UMB) Institutional Review Board protocol HP-00087489. Formal consent was not obtained because the project was a secondary analysis of pre-existing public health data obtained through routine public health activities performed by the Samoa Ministry of Health.

Study setting

This study took place from 2018–2020 in Samoa, a Pacific Island nation of area 2,842 km² (1,097 sq. mi) and ~200,000 population. Samoa comprises two populated islands, Upolu and Savaii, each with a central clinical laboratory and several peripheral health facilities (Fig 1). Typhoid fever is a notifiable infectious disease in Samoa that triggers epidemiologic investigation of any and all culture-confirmed cases [19,20]. Since 2018, the STFCP has strengthened epidemiologic surveillance for typhoid fever, improved clinical and environmental microbiology laboratory capacity, and initiated WGS at a collaborating regional reference laboratory in Melbourne, Australia. Typhoid Fever Epidemiologic “SWAT” Teams (equipped with specialized epidemiologic tools and tactics), one on each island, were trained to perform epidemiologic investigations of the household (and/or workplace or school) of every bacteriologically confirmed case of typhoid fever occurring anywhere in Samoa.

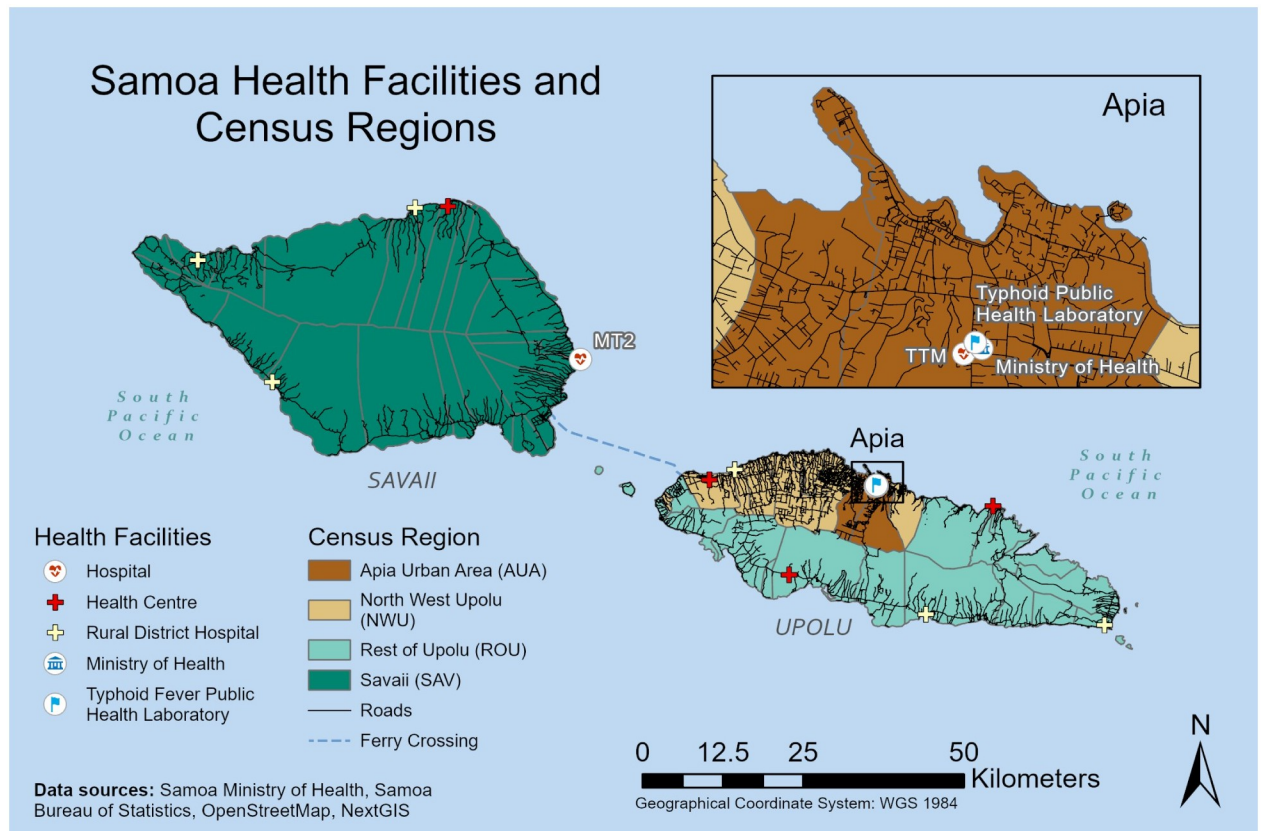


Fig 1. Map of Samoa with census regions, roads, and health facilities. Samoa is comprised of two primary islands, Upolu to the east and Savaii to the west. About three quarters of the population and the capital town of Apia are located on Upolu. The black lines indicate roads associated with populated areas, which are predominantly located around the periphery of the islands except for the urban areas where inland settlements are increasing. The islands are subdivided into four statistical census regions, 43 political districts (outlined in light grey), and >300 villages (not depicted). The four colored geographic areas differentiate the four census regions: Apia Urban Area (AUA), North West Upolu (NWU), Rest of Upolu (ROU), and Savaii (SAV). AUA is most dense with ~25% of the total population. Apia is the capital seat of government and includes the Ministry of Health headquarters, Samoa Typhoid Fever Public Health Laboratory, and Tupua Tamasese Meaole (TTM) national hospital, which processes the blood cultures for Upolu (insert). NWU is mixed peri-urban and rural in population density and separates Apia from the inter-island ferry (navy dashed line) and adjacent international airport (not shown). The Malietoa Tanumafili II (MT2) hospital processes blood cultures for Savaii. The health facility icons indicate the two central hospitals with clinical microbiology laboratories, TTM and MT2, and the 6 district hospitals and 4 peripheral health centers. Base layer map available from <https://planet.openstreetmap.org>.

<https://doi.org/10.1371/journal.pntd.0010348.g001>

Case ascertainment

Febrile patients who presented to an MoH health facility with clinical signs and symptoms of typhoid fever [21] triggered the collection of a blood culture. Patients from whom *S. Typhi* was isolated and confirmed by standard methods [22,23] were considered acute blood culture-confirmed cases of typhoid fever. Every such case was reported to the Typhoid SWAT Team on the incident island for expeditious investigation. On three occasions, the same individual presented to a health facility with culture-confirmed typhoid fever within ~1 month of a prior confirmed typhoid infection. These three repeat positive cases (presumed relapses) were counted as distinct cases but considered epidemiologically linked by person.

Case investigations and epidemiologic linkages

During household visits, the Typhoid SWAT teams evaluated environmental risk factors for typhoid transmission and ascertained the clinical wellbeing of the index case and their

household contacts. Global positioning system (GPS) coordinates of the home were collected using Garmin GPSMAP 64sc handheld navigators. The Typhoid SWAT teams also sought evidence of epidemiologic links to other recent typhoid cases occurring within two months in the same home/venue, in neighboring homes, or in contacts living elsewhere in Samoa. To detect temporary or chronic asymptomatic shedders of *S. Typhi* among household contacts, up to three rectal swabs collected 1–2 days apart from each contact were transported in Cary-Blair transport medium to the Typhoid Fever Public Health Laboratory in Apia and examined using standard bacteriologic methods [22,23]. Asymptomatic contacts from whom *S. Typhi* was isolated from stool cultures were referred for treatment [24]. These asymptomatic shedders were considered epidemiologically linked to the index case by person, place, and time.

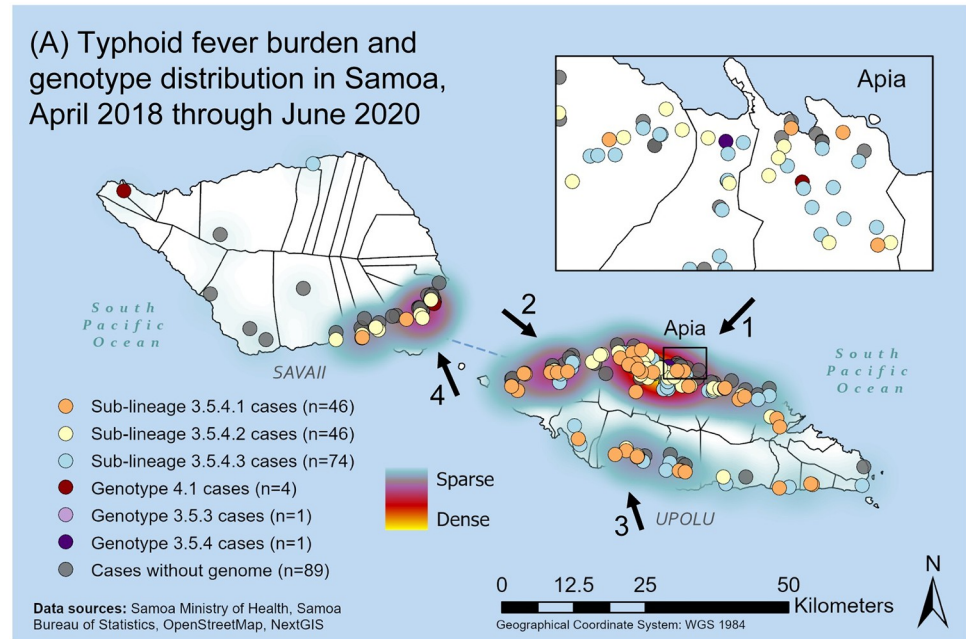
Spatial and temporal point pattern analyses

Using the spatial statistics toolbox in Esri ArcGIS Pro v2.9 [25], we performed two point pattern cluster analyses of case households ($n = 260$) to ascertain significant ($p < 0.05$) spatial and spatial-temporal clusters. First, we used the Average Nearest Neighbor (ANN) analysis to determine if the spatial distribution of points (i.e., case households) was clustered based on the average distances between them [26,27]. The resulting nearest neighbor ratio (R_n) ranged from zero (significant clustering) to 2.15 (uniform equidistant dispersion) and was supported by critical values (Z-score and p-value); an R_n equal to 1 indicated randomness or no detectable spatial pattern (Figs A and B in [S1 Text](#)). The ANN analysis was performed on household coordinates in aggregate and then by each of the four census regions (Fig 1 and Fig C-i in [S1 Text](#)), which separate the two Samoan islands into the Apia Urban Area (AUA; urban; ~625 persons per km²), North West Upolu (NWU; mixed peri-urban and rural; ~275 persons per km²), Rest of Upolu (ROU; rural; ~60 persons per km²) and Savaii (SAV; rural; ~26 persons per km²) [28].

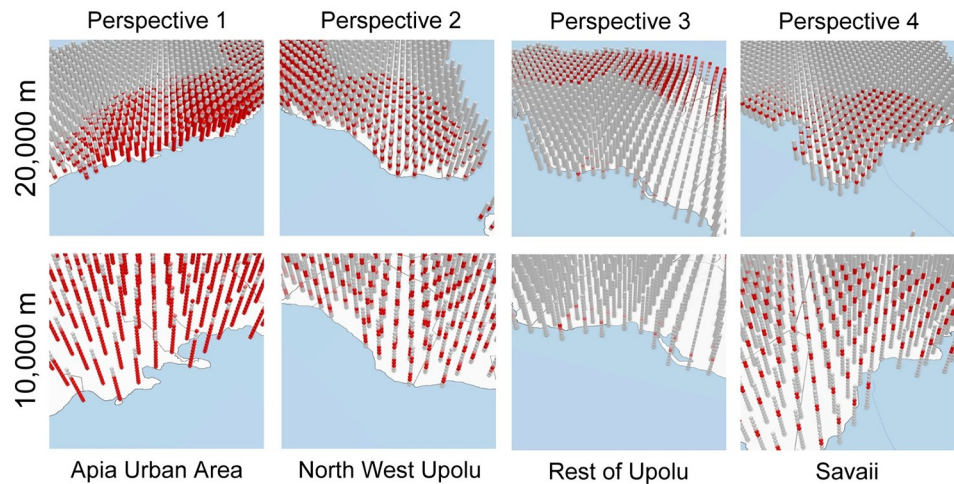
We also used the Space-Time Hotspot analysis tool to detect significant clustering of acute case households over time. Case household coordinates were binned into 1 km² x 2-month space-time cubes and visualized in three dimensions with red shading to indicate space-time clustering confidence levels (Fig 2). Detailed descriptions and methodologies for the ANN and Space-Time Hotspot tools are included in Figs A and B in [S1 Text](#).

Whole genome sequencing and phylogenetic analysis

All available Samoan *S. Typhi* blood isolates from acute cases and stool isolates from asymptomatic shedders were shipped to the Microbiological Diagnostic Unit Public Health Laboratory (University of Melbourne, Australia) for whole genome sequencing, as previously described [15]. Briefly, extracted DNA was sequenced on the Illumina NextSeq 500 with 150-cycle paired end chemistry. Genotypes were determined from reads mapped to the reference strain CT18 (GCA_000195995.1) using GenoTyphi v1.9.1 [29,30]. Genotype 3.5.4.1, 3.5.4.2, and 3.5.4.3 sub-lineages unique to Samoa were assigned based on canonical single nucleotide polymorphisms (SNPs) against CT18, as previously described [15]. To generate a locally referenced maximum-likelihood (ML) phylogeny, reads were assembled, and core genome SNPs were identified relative to the 2012 Samoan *S. Typhi* reference strain H12ESR00394-001A (GCA_001118185.2). After removal of recombination sites, a ML phylogeny was inferred using the generalized time-reversible (GTR) site substitution model with a Gamma rate distribution and the Lewis ascertainment bias correction (ASC_GTRGAMMA) and 100 bootstrap pseudo-replicates in RAxML v8.2.10 [31]. Pairwise SNP distances between genomes were calculated using snp-dists v0.8.2 [32] and compared among epidemiologically



(B) Space-time hotspot analysis visualized in 3D at two altitudes



(C) Method and legend

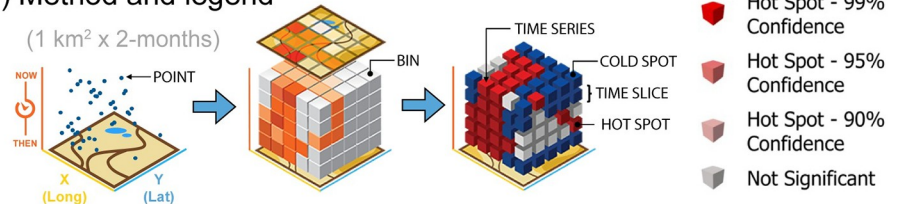


Fig 2. Spatial-temporal and genomic analysis of typhoid fever cases (burden) in Samoa from April 2018 through June 2020. (A) Using ArcGIS Pro v2.9, household GPS coordinates of acute cases of *S. Typhi* (n = 260) from April 27, 2018 to June 9, 2020 are plotted on a map of Samoa and colored according to *S. Typhi* genotype/sub-lineage. A point density heatmap is inferred from case coordinates. The four prominent hotspots are identified with arrows labeled 1–4, which also indicate aerial perspectives shown in the reconstructions in panel B. (B) Coordinates were binned into space-time cubes of 1km² x 2-months (see panel C and S1 Text for additional details) and tested for significant clustering by the emerging hotspot analysis implemented in ArcGIS Pro. The resulting 3D space-time hotspots were visualized from an aerial view of 20,000 meters and 10,000 meters focusing on the four typhoid hotspots identified in panel A. In the 3D renderings, each vertical column on the map represents a stack of cubes over 1 km² area and with a

height representing 2 months of time, totaling 13 time steps, or all 26 months of this study. Aggregation of points in both space and time generate a space-time hotspot, visible as a red cube or band of several adjacent cubes. The intensity of red shading indicates statistical confidence. (C) To generate space-time cubes, points were graphed in a 3-dimensional plane based on coordinates and time, and then binned into cubes of space and time dimensions. Counts within each cube were compared to nearby cubes to extrapolate hot and cold spots with statistical support (99%, 95%, 90%, not significant). Hotspots in the x-y plane occur within a time slice. Hotspots inferred from the z-direction fall within a time series. Adapted with permission from Esri (Redlands, CA: Environmental Systems Research Institute, Inc.). Base layer map available from <https://planet.openstreetmap.org>.

<https://doi.org/10.1371/journal.pntd.0010348.g002>

linked isolates. Detailed protocols are found in the [S1 Text](#), with accession numbers listed in [S1 Table](#) and assemblies available on Figshare (doi: [10.6084/m9.figshare.18665686](https://doi.org/10.6084/m9.figshare.18665686)) [15].

Epidemiologic analysis of genotypes/sub-lineages

Patient characteristics (sex, age group, and mean age) were assessed for differences between all cases and the subset of sequenced isolates using Pearson's chi-squared test of independence for the categorical variables and unpaired two-tailed T test of means for mean age. Phylogenetic groupings of *S. Typhi* were examined for associations with epidemiologic variables (infection status, sex, age group, census region, year) using Fisher's exact test. The ANN statistic was performed in ArcGIS Pro v2.9 on groupings of household coordinates by each genotype 3.5.4 sub-lineage (Fig C in [S1 Text](#)). The absolute and relative counts of phylogenetic groupings within each census region were plotted over 2-month time intervals using Microsoft Excel v16.58.

Statistical analyses

Spatial statistics were performed in ArcGIS Pro v2.9 using default tools. Epidemiologic comparisons were performed using the biostatistics package RVAideMemoire v0.9–81 [33] in R v4.1.1 [34]. Statistical significance was reported at $p < 0.05$ and indicated in the text, figure, and/or table legends, accordingly.

Results

Characteristics of the dataset

From April 2018–June 2020, the STFCP identified 260 blood culture-confirmed acute typhoid fever cases through a strengthened surveillance system and detected 27 asymptomatic *S. Typhi* fecal shedders through household investigations. We examined case household coordinates for spatial and spatial-temporal clustering, and then incorporated phylogenetic analyses of the sequenced subset of 186 Samoan *S. Typhi* isolates—172 from acute cases and 14 from asymptomatic shedders—among household contacts ([S1 Table](#)). All available isolates were sequenced in this study. Among the 186 whole genome sequenced isolates, three *S. Typhi* genotypes were identified: genotype 3.5.4 (181/186, 97.3%), genotype 4.1 (4/186, 2.2%), and genotype 3.5.3 (1/186; 0.5%). All but one genotype 3.5.4 isolates could be further categorized into three distinct sub-lineages denoted 3.5.4.1 ($n = 49$), 3.5.4.2 ($n = 51$), and 3.5.4.3 ($n = 80$) and differentiated from all other global *S. Typhi* genomes by single canonical SNPs ([S1 Table](#)), as previously described [15]. The sequences of isolates from acute cases showed no significant difference in patient characteristics (patient sex, mean age, or age group distribution) when compared to all reported cases ([Table 1](#)).

Table 1. Comparison of age and sex variables of acute cases associated with an *S. Typhi* whole genome sequence (WGS) versus acute cases missing WGS data.

Variable		All cases	Cases, WGS available	Cases, WGS missing	Significance
Total		260	172	88	-
Sex	Female, N (%)	112 (43.1%)	72 (41.9%)	40 (45.5%)	p = 0.560
	Male, N (%)	147 (56.5%)	99 (57.6%)	48 (54.5%)	
Age	Mean age (SD)	24.4 (16.7%)	24.1 (16.5%)	25.0 (17.3%)	p = 0.709
Age groups	0–4 years old, N (%)	21 (8.1%)	15 (8.7%)	6 (6.8%)	p = 0.864
	5–19 years old, N (%)	103 (39.6%)	66 (38.4%)	37 (42.0%)	
	20–59 years old, N (%)	111 (42.7%)	75 (43.6%)	36 (40.9%)	
	50+ years old, N (%)	25 (9.6%)	16 (9.3%)	9 (10.2%)	

Categorical variables (sex and age group) tested with Pearson's Chi-squared test of independence.

Continuous variable (mean age) tested with unpaired two-tailed T test of means.

Statistics were performed using the RVAideMemoire v0.9–81 [33] packages in R v4.1.1 [34].

<https://doi.org/10.1371/journal.pntd.0010348.t001>

Spatial clustering and spatial-temporal hotspot analyses of acute cases

When acute case households ($n = 260$) were mapped, areas with many typhoid cases (higher burden)—Apia, northwestern Upolu, southeastern Savaii, and Upolu's southern coast—were prominent among the four census regions (arrows in Fig 2A). We used the Average Nearest Neighbor (ANN) and Space-Time Hotspot analyses in ArcGIS Pro v2.9 to assess spatial and spatial-temporal clustering, respectively. As expected for a human host-restricted communicable disease that can be transmitted by short-cycle mode in which a carrier of *S. Typhi* (chronic or short-term) with fecally-soiled hands contaminates food or water consumed by close contacts, case households ($n = 260$) demonstrated significant clustering by ANN analysis ($R_n < 1$, $p < 0.005$). However, when grouped by census region (Fig C-i in S1 Text and Table 2), local differences emerged in the degree of clustering, i.e., R_n closer to 0, indicating more tightly grouped clusters. Case households in the AUA ($n = 42$), where population density is greatest and where residents receive putatively treated drinking water through a reticulated system,

Table 2. Results of the Average Nearest Neighbor (ANN) statistic performed on geographic and phylogenetic groupings of *S. Typhi* acute infections.

Group ^a	No. of points ^a	\bar{O} distance (meters)	\bar{E} distance (meters)	Nearest Neighbor Ratio (R_n)	ANN Z-score ^b	ANN P-value ^c	Fig C in S1 Text. Panel
All cases	260	676.07	2312.04	0.29	-21.83	$p < 0.005$	A
Cases AUA	42	204.87	331.34	0.62	-4.73	$p < 0.005$	A
Cases NWU	110	351.30	858.30	0.41	-11.85	$p < 0.005$	A
Cases ROU	55	607.40	2406.22	0.25	-10.61	$p < 0.005$	A
Cases SAV	53	1800.82	3007.99	0.60	-5.59	$p < 0.005$	A
Cases 3.5.4.1	46	1589.89	3230.86	0.49	-6.59	$p < 0.005$	B
Cases 3.5.4.2	46	1211.20	2990.57	0.41	-7.64	$p < 0.005$	C
Cases 3.5.4.3	74	1710.15	3304.99	0.52	-7.94	$p < 0.005$	D

ANN is a statistical comparison of the observed mean (\bar{O}) and expected mean (\bar{E}) distances between points yields a nearest neighbor ratio (R_n) further described by a probability (p-value) that the observed spatial pattern was created by a random process and standard deviations (z-score). The R_n ranges from 0 (indicating discrete clustering of points) to ~ 2.15 (indicating uniform dispersion of points equidistant from one another). An R_n value of 1 indicates complete spatial randomness.

^a Only household coordinates of acute cases were included to assess circulating *S. Typhi* detected through routine passive surveillance and to avoid biasing the point-pattern analyses to households where asymptomatic shedders were localized by active investigation.

^b All z-scores are small (< -1.96) corresponding to a confidence level above 95%.

^c All p-values are small (< 0.005) which indicate statistical significance.

<https://doi.org/10.1371/journal.pntd.0010348.t002>

were the least clustered of the census regions ($R_n = 0.62$). In contrast, case households in the rural ROU ($n = 55$), where districts each have independent (often untreated) water schemes, were the most clustered ($R_n = 0.25$) (Table 2). Notably, SAV ($n = 53$), which is more rural than ROU, demonstrated a similar degree of clustering ($R_n = 0.60$) to AUA (Table 2); however, there were ~ 7 case households remotely located in the north and west of Savaii far from any nearest neighbor (Fig C-i in S1 Text) that, if excluded, yielded an $R_n = 0.33$ for the remaining households in the south and southeast of Savaii where the vast majority of Savaii cases occurred.

To assess for temporal clustering by census region, a Space-Time Hotspot analysis was performed using a window of $1 \text{ km}^2 \times 2$ -months to bin case household coordinates ($n = 260$) and visualized in three-dimensions at two altitudes, 20,000 meters and 10,000 meters, to aid in visualization of each time series (Fig 2B). Temporal persistence of typhoid burden was detectable in AUA as sequential spatial-temporal hotspots, visible as contiguous red columns, ($p < 0.01$) associated with that geographic region (Fig 2B and 2C). This was in contrast with sporadic (short-lived) hotspots of infection evident in the rest of the country as identified by the discrete red bands ($p < 0.01$) between grey bars in those geographic regions (Fig 2B and 2C).

Phylogenetic analysis by genotype, person, place, and time

Subtyping of *S. Typhi* isolates facilitates elucidation of patterns of transmission based on the assumption that similar isolates form a network or chain of infection via a common vehicle and/or source. The presence of three genotype 3.5.4 sub-lineages in Samoa (Fig 3A) allowed the demarcation of useful groupings based on genetic similarity for subsequent, more detailed epidemiologic analyses of associations. The 14 sequenced isolates from asymptomatic shedders detected through household investigations were genotype 3.5.4 and showed no association with any particular sub-lineage (Table 3 and Fig 3A). Sub-lineage 3.5.4.1 was significantly ($p < 0.05$) associated with (i) age group (greater-than-expected frequency [GTEF] among ages 0–4 years and 50+; less-than-expected frequency [LTEF] in 5–49 year-olds), (ii) census region (GTEF in ROU, LTEF in AUA and SAV), and (iii) year (GTEF in 2019) (Table 3 and Fig 3A). Sub-lineage 3.5.4.2 was significantly ($p < 0.05$) associated with the patient's sex, being isolated from males more frequently than females (Table 3 and Fig 3A). Sub-lineage 3.5.4.3 was significantly ($p < 0.05$) associated with 20–49 year-olds and LTEF in the 5–19 and 50+ year old age groups, as well as census region (common to the four census regions but GTEF in the AUA and ROU) (Table 3 and Fig 3A).

Geospatially, case households associated with each genotype/sub-lineage were found in the four census regions (Table 3) and were overlapping (Fig 2A), indicating widespread circulation among the islands. Spatial clustering of each 3.5.4 sub-lineage was significant ($p < 0.005$) and intermediate ($0.4 < R_n < 0.6$) among the groupings tested by ANN (Table 2). Within each census region, we identified variation in the frequency as well as transient expansions and contracts in the proportions of each 3.5.4 sub-lineage over time (Fig 4). Total case counts peaked in July-August 2019 and then sharply declined in November-December 2019 and again in March-April 2020 (highlighted background colors in Fig 4), coinciding with three major national events.

Genetic distances between epidemiologic linkages

WGS analyses can also be used to verify epidemiologic linkages, thereby providing evidence of transmission based on the relatedness of *S. Typhi* isolates. We defined relatedness based on the number of core genome SNP differences between isolates, with fewer differences indicating

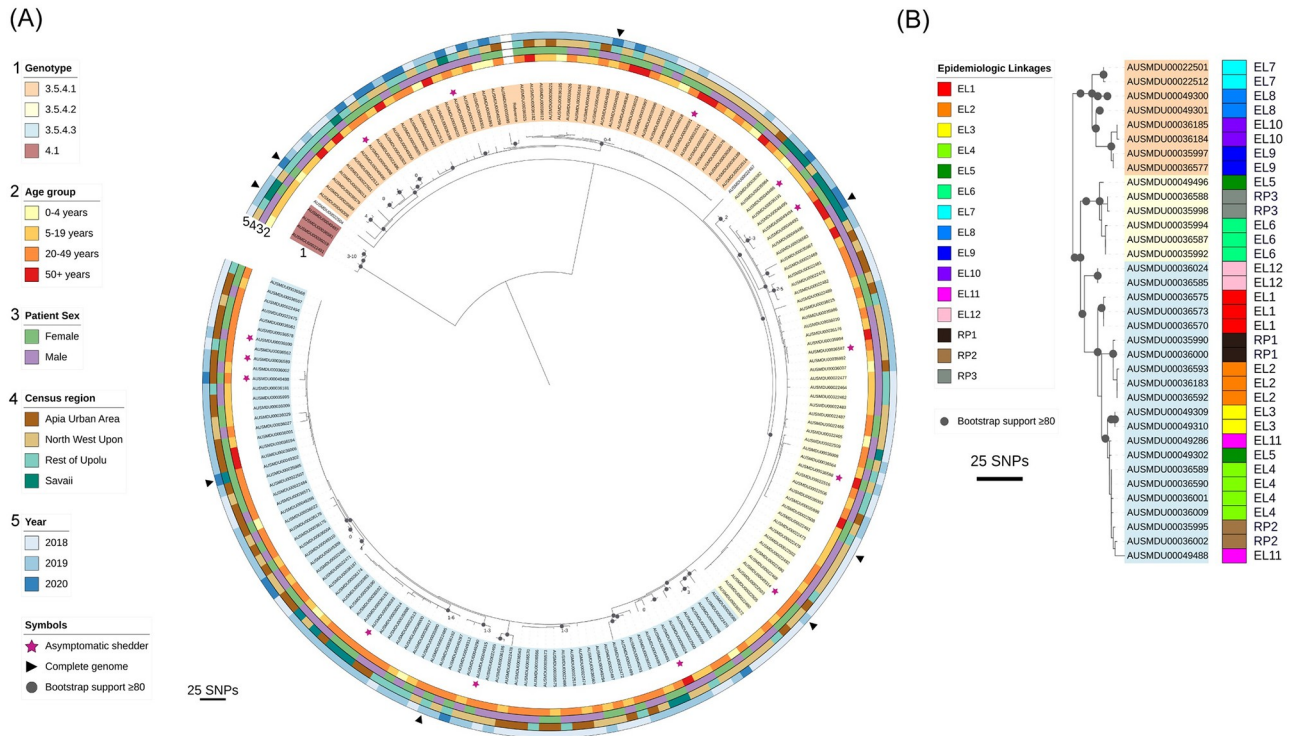


Fig 3. Maximum-likelihood phylogeny, associated epidemiologic variables, and epidemiologic linkages among *S. Typhi* from Samoa from April 2018 through June 2020. (A) Maximum-likelihood phylogeny from *de novo* assembled whole genome sequences of Samoan *S. Typhi* genotype 3.5.4 isolates from April 27, 2018 through June 9, 2020 (N = 186, S1 Table). Core genome single nucleotide polymorphisms (SNPs) were identified relative to the 2012 Samoa Reference genome H12ESR00394-001A (GCA_001118185.2). Genotype/sub-lineage is indicated by isolate label shading (ring 1). Two outlier isolates are not shaded. Asymptomatic shedders of *S. Typhi* are denoted with a pink star. The patient age group (ring 2), patient sex (ring 3), census region of residence (ring 4), and year of isolation (ring 5) are annotated in the circumscribing colored rings. (B) Epidemiologic linkages (EL) identified through case household investigations are compared against genomic relatedness of the isolates based on core genome phylogeny. The visible isolates were extracted from the tree in Panel A. In total 2/12 epidemiologic linkages are not supported by the genomic variation observed in the isolates from those samples (EL5/dark green and EL11/pink groups are not similar). Three repeat positive (RP) blood cultures isolated ~1 month apart are also compared, showing close genetic similarity.

<https://doi.org/10.1371/journal.pntd.0010348.g003>

closer relatedness. Our previous study has demonstrated that the mutation rate of the *S. Typhi* from Samoa is less than those observed for other geographic locations or genotypes [15]. Comparing the genomes of all 186 isolates in this study with the 2012 Samoa reference, the median number of SNP differences is 37 and ranged from 7 to 234, or median of 37 and range of 7 to 56 SNPs if the four outlying genotype 4.1 isolates are excluded (S2 Table). By genotype grouping, the median SNP differences were 15, 37, and 41 for sub-lineages 3.5.4.1, 3.5.4.2, and 3.5.4.3, respectively.

During household investigations, the Typhoid SWAT teams attempted to trace infections and link index cases by person, place, and/or time. In 12 instances, genomes were available for putatively linked infections in relation to person, place, and/or time (Fig 3B and Table 4). In six of these, an asymptomatic shedder of *S. Typhi* was identified. In nine epidemiologic linkages, the genotype sub-lineages matched and SNP distances between isolates were two or fewer, and eight of those epidemiologic linked isolates have one or fewer SNPs, indicating close relatedness (Fig 3B and Table 4). In two linkages, EL8 and EL11, the genotype sub-lineages matched, but the SNP distances between isolates were greater, with nine and 14 SNP differences, respectively (Table 2). In another (EL5), the isolates were of completely different *S. Typhi* sub-lineages and 28 SNPs apart (Table 4). Additionally, microbiological surveillance

Table 3. Comparison of epidemiologic variables relating to person, place, and time for each Samoa dominant *S. Typhi* genotype/sub-lineage.

	Genotype/Sub-lineage						TOTAL
	3.5.4.1	3.5.4.2	3.5.4.3	4.1	3.5.4	3.5.3	
Total	49 (26.3%)	51 (27.4%)	80 (43.0%)	4 (2.2%)	1 (0.5%)	1 (0.5%)	186
Status							
Acute Case	46 (93.9%)	46 (90.2%)	74 (92.5%)	4 (100%)	1 (100%)	1 (100%)	172
Asymptomatic Shedder	3 (6.1%)	5 (9.8%)	6 (7.5%)	0 (0%)	0 (0%)	0 (0%)	14
Significance	p = 1	p = 0.535	p = 1	p = 1	p = 1	p = 1	
Sex							
Female	24 (49%)	16 (31.4%)	40 (50%)	1 (25%)	0 (0%)	0 (0%)	81
Male	25 (51%)	35 (68.6%)	40 (50%)	3 (75%)	1 (100%)	1 (100%)	105
Significance	p = 0.404	p = 0.0470	p = 0.137	p = 0.634	p = 1	p = 1	
Age group							
0–4 years	7 (14.3%)	3 (5.9%)	4 (5.0%)	1 (25%)	0 (0%)	0 (0%)	15
5–19 years	17 (34.7%)	23 (45.1%)	23 (28.8%)	3 (75%)	0 (0%)	1 (100%)	67
20–49 years	15 (30.6%)	19 (37.3%)	50 (62.5%)	0 (0%)	1 (100%)	0 (0%)	85
50+ years	10 (20.4%)	6 (11.8%)	3 (3.8%)	0 (0%)	0 (0%)	0 (0%)	19
Significance	p = 0.004	p = 0.363	p = 0.0004	p = 0.093	p = 1	p = 0.543	
Census region							
AUA	4 (8.2%)	7 (14%)	23 (28.8%)	0 (0%)	1 (100%)	0 (0%)	35
NWU	20 (40.8%)	23 (44%)	35 (43.8%)	2 (50%)	0 (0%)	1 (100%)	81
ROU	22 (44.9%)	12 (24%)	13 (16.3%)	0 (0%)	0 (0%)	0 (0%)	47
SAV	3 (6.1%)	9 (18%)	9 (11.3%)	2 (50%)	0 (0%)	0 (0%)	23
Significance	p = 0.0013	p = 0.452	p = 0.0072	p = 0.127	p = 0.312	p = 1	
Year							
2018	15 (30.6%)	28 (54.9%)	32 (40.0%)	1 (25%)	1 (100%)	1 (100%)	78
2019	23 (46.9%)	20 (39.2%)	42 (52.5%)	3 (75%)	0 (0%)	0 (0%)	88
2020	11 (22.4%)	3 (5.9%)	6 (7.5%)	0 (0%)	0 (0%)	0 (0%)	20
Significance	p = 0.0083	p = 0.081	p = 0.328	p = 0.762	p = 0.527	p = 0.527	

Numbers in parentheses indicate the percentages of the number of times the genotype/sub-lineage was detected for each variable relative to the total number of the genotypes; P-values were calculated using Fisher's exact test using the RVAideMemoire v0.9–81 [33] packages in R v4.1.1 [34]. Bolded p-values are significant at $p < 0.05$.

<https://doi.org/10.1371/journal.pntd.0010348.t003>

during the study period identified three instances in which two positive *S. Typhi* cultures derived from the same individual were collected ~1 month apart (Table 4). In each instance, the genotype sub-lineages matched (Fig 3B) and the SNP distances between the first and second isolation were 1–3 SNPs, indicating the isolates were closely related (Table 4).

Discussion

Identifying the modes of transmission that maintain the endemicity of typhoid in Samoa has been a subject of interest for many decades [12,14]. Some insights were derived from our review of the Samoa typhoid burden from 2008–2019 [14]: 53–193 cases of blood culture-confirmed typhoid fever occurred annually without obvious seasonality, equating to a persistently moderate annual incidence. The greatest incidence and numbers of cases was observed among school age children and working age adults in the contiguous regions of NWU and AUA where piped treated water reaches most households, indicating that their daily life patterns put them at enhanced risk. Herein, we relate results of applying the potential discerning power of spatial-temporal cluster and phylogenetic analyses to typhoid infections in Samoa from April 2018 through June 2020.

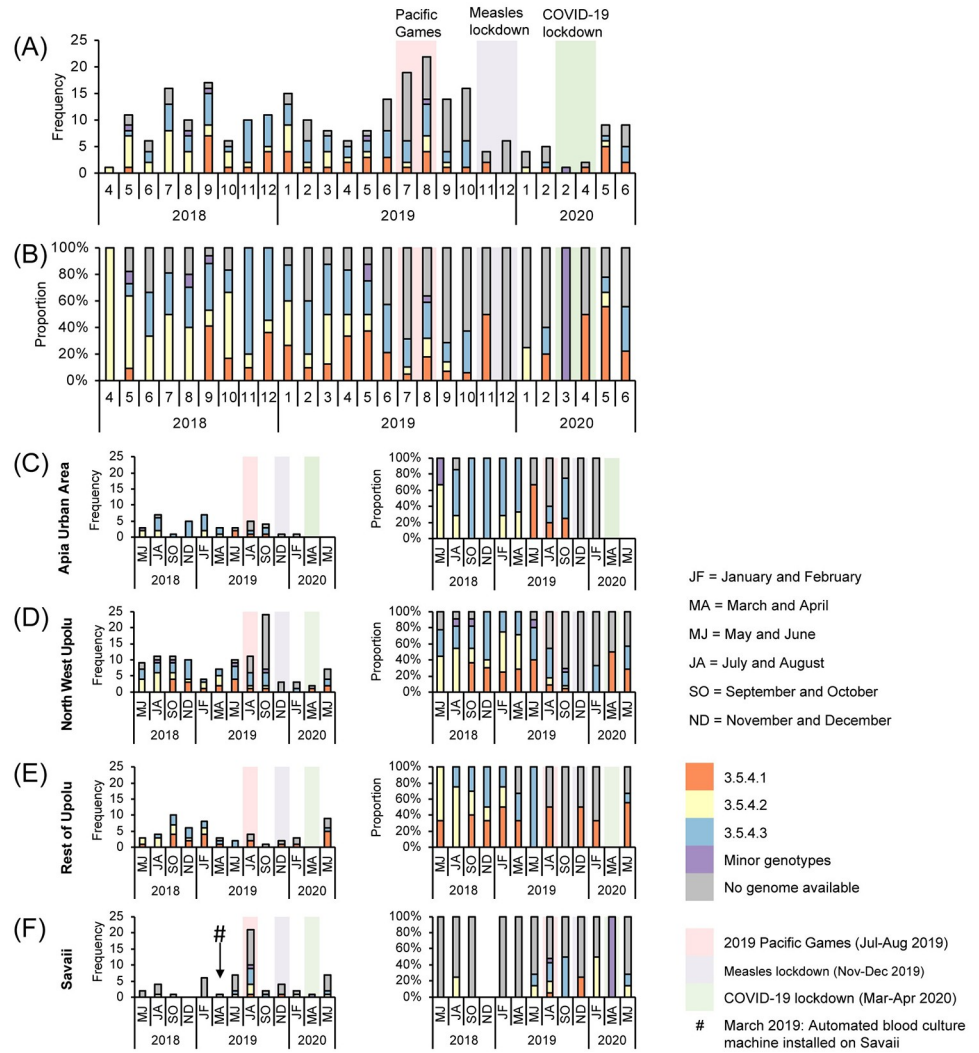


Fig 4. Bar graphs depicting the number and proportion of blood culture-confirmed cases of typhoid fever in Samoa over time from April 27, 2018 through June 9, 2020 by genotype/sub-lineage. (A-B) The frequency and proportion of acute cases (N = 260) occurring each month are colored by genotype/sub-lineage (n = 172) or greyed if no genome was available (n = 88). Genotype 3.5.4 sub-lineages 1–3, comprising a total of 166/172 sequenced isolates from acute cases, are identified as orange, yellow, and blue bars, respectively. Minor genotypes (purple bar) include genotype 4.1 (4/186), genotype 3.5.3 (1/186), and one genotype 3.5.4 outlier (1/186) that did not fall into one of the three sub-lineages. (C-F) The frequency and proportion of acute cases occurring over 2-month intervals in each census region. Three major events impacting population mobility—increases during the 2019 Pacific Games in Nov-Dec 2019, and restrictions during the measles lockdown from Nov-Dec 2019 [35] and COVID-19 lockdown from Mar-Apr 2020 [36]—are labeled as light red, light purple, and light green background colors, respectively, over the two-month time intervals affected.

<https://doi.org/10.1371/journal.pntd.0010348.g004>

The burden of typhoid fever was most concentrated in four locations (hotspots; arrows in Fig 2A) determined by point densities: Apia, northwestern Upolu, southeastern Savaii near the ferry terminal, and Upolu’s southern coast. We hypothesized that spatial and temporal differences among the underlying census regions would support either widespread, continuous typhoid circulation or focal, sporadic case clusters around households. Case households in both urban and rural regions showed significant clustering ($p < 0.005$, $R_n < 1$); however, the distances between cases were much smaller in rural ROU (lower R_n), indicating stronger foci of spatial clusters compared with more dispersed case clustering (greater R_n) in AUA (Table 2).

Table 4. Suspected epidemiologic linkages detected through active surveillance and associated genotype and pairwise SNP distance between isolates.

Linkage ID ^a	Year	Month	Day	Sex	Age	Status ^b	Relation to Index Case ^c	Epidemiologic Association	Strain ID	Genotype ^d	SNP Dist ^e
EL1	2018	12	3	F	13	Acute Case	Index	Index	AUSMDU00036575	3.5.4.3	-
	2018	12	4	M	38	Acute Case	Family; same HH	Person, place, time	AUSMDU00036570	3.5.4.3	1
	2018	12	12	F	20	Acute Case	Family; same HH	Person, place, time	AUSMDU00036573	3.5.4.3	1
EL2	2019	8	14	F	21	Acute Case	Index	Index	AUSMDU00036183	3.5.4.3	-
	2019	8	21	F	31	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036593	3.5.4.3	1
	2019	8	25	F	25	Acute Case	Family; different HH	Person, time	AUSMDU00036592	3.5.4.3	0
EL3	2020	6	3	M	40	Acute Case	Index	Index	AUSMDU00049310	3.5.4.3	-
	2020	6	9	M	11	Acute Case	Family; same HH	Person, place, time	AUSMDU00049309	3.5.4.3	0
EL4	2019	2	18	F	10	Acute Case	Index	Index	AUSMDU00036001	3.5.4.3	-
	2019	2	22	F	29	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036589	3.5.4.3	1
	2019	3	4	F	14	Acute Case	Family; same HH	Person, place, time	AUSMDU00036009	3.5.4.3	2
	2019	3	11	M	32	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036590	3.5.4.3	1
EL5	2020	5	5	M	67	Acute Case	Index	Index	AUSMDU00049496	3.5.4.2	-
	2020	5	29	F	21	Acute Case	Family; different HH	Person, time	AUSMDU00049302	3.5.4.3	28
EL6	2019	1	20	M	12	Acute Case	Index	Index	AUSMDU00035992	3.5.4.2	-
	2019	1	22	M	65	Acute Case	Family; same HH	Person, place, time	AUSMDU00035994	3.5.4.2	1
	2019	1	25	M	37	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036587	3.5.4.2	0
EL7	2018	9	14	M	9	Acute Case	Index	Index	AUSMDU00022501	3.5.4.1	-
	2018	9	17	F	52	Acute Case	Family; same HH	Person, place, time	AUSMDU00022512	3.5.4.1	1
EL8	2020	5	4	F	60	Acute Case	Index	Index	AUSMDU00049300	3.5.4.1	-
	2020	5	23	M	27	Acute Case	Caregiver; different HH	Person, place, time	AUSMDU00049301	3.5.4.1	14
EL9	2018	12	24	M	12	Acute Case	Index	Index	AUSMDU00036577	3.5.4.1	-
	2019	1	25	M	22	Acute Case	Gathering (funeral); diff HH	Person, place, time	AUSMDU00035997	3.5.4.1	0
EL10	2019	8	5	M	11	Acute Case	Index	Index	AUSMDU00036184	3.5.4.1	-
	2019	8	11	F	2	Acute Case	Family; different HH	Person, time	AUSMDU00036185	3.5.4.1	1
EL11	2019	10	11	F	3	Acute Case	Index	Index	AUSMDU00049286	3.5.4.3	-
	2020	4	3	F	28	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00049488	3.5.4.3	9
EL12	2019	6	13	M	16	Acute Case	Index	Index	AUSMDU00036024	3.5.4.3	-
	2019	6	25	M	65	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036585	3.5.4.3	0
RP1	2019	1	13	F	14	Acute Case	Index	Index	AUSMDU00035990	3.5.4.3	-
	2019	2	13	F	14	Acute Case	Self	Person	AUSMDU00036000	3.5.4.3	3
RP2	2019	1	23	M	16	Acute Case	Index	Index	AUSMDU00035995	3.5.4.3	-
	2019	2	18	M	16	Acute Case	Self	Person	AUSMDU00036002	3.5.4.3	3
RP3	2019	1	29	F	61	Acute Case	Index	Index	AUSMDU00035998	3.5.4.2	-
	2019	2	12	F	62	Asymptomatic Shedder	Self	Person	AUSMDU00036588	3.5.4.2	1

^a EL = Epidemiologic linkage; RP = repeat positive (culture); numbers are ascending based on date of isolate from index case

^b Acute cases are blood isolates; asymptomatic shedders are rectal swab stool isolates

^c The index case is the first culture-confirmed infection in the linkage; HH = Household

^d Genotypes were determined using GenoTyphi v1.9.1; sub-lineages were determined as previously described [15]

^e Single Nucleotide Polymorphism (SNP) pairwise distances calculated using snp-dists v0.8.2 [32].

<https://doi.org/10.1371/journal.pntd.0010348.t004>

In Savaii, the R_n was closer to the value for AUA, likely due to skewing from ~8 individual cases occurring in remote isolation distant from nearest neighbors (Fig C in [S1 Text](#)). Spatial-temporal hotspot analysis revealed consecutive, significant ($p < 0.01$) temporal hotspots in AUA and the adjacent peri-urban regions of NWU (Fig 4), which could represent a locus of persistent, amplified transmission and also a source of infection for persons from other parts of the islands who work, visit, or consume food prepared in and around Apia. Population distribution, the clustering of rural communities, access to health care facilities with blood culture capability, quality of water supplies and the integrity of the water distribution system (if any) may also play a role.

Contextualization of cases over time amid major national events that either greatly expanded or abruptly halted population mobility highlighted an association between population mobility and case frequency. Typhoid cases, all of which occurred in Samoan people, rose dramatically during the 2019 Pacific Games in July–August 2019 (light red background in Fig 4A) when there was an influx of athletes and supporters. Notably, we did not detect any non-Samoan genotypes during or after this period among the sequenced isolates (Fig 4A), suggesting that dissemination of an imported strain was less likely responsible for the observed increase in cases. Months later, gatherings and population mobility were limited during an intensive nationwide measles lockdown in November–December 2019 (light purple background in Fig 4A) [35], and during COVID-19 lockdowns in March–April 2020 (light green background in Fig 4A) [36]. During these periods of reduced population mobility, whether involuntary or out of concern of exposure to circulating infections (i.e., measles, COVID-19), the occurrence of typhoid diminished markedly (Fig 4A). These data highlight a potential role for community interactions and population movement in transmission, but do not rule out the contribution of changes in health seeking behavior or health care delivery during the lockdowns.

Sub-typing of clinically indistinguishable infections can help to ascertain unique networks of infection. Our previous WGS analyses of Samoan *S. Typhi* from 1983–2020 revealed that dominant Samoa-exclusive genotypes 3.5.4 and 3.5.3 can be divided into sub-lineages based on canonical SNPs [15]. Herein, we analyzed the 2018–2020 *S. Typhi* isolates by sub-lineages in relation to variables of person, place, and time (Table 3). Infections with sub-lineage 3.5.4.1 were more common in ROU and NWU among very young children and older adults who are typically home. Whereas, infections with sub-lineage 3.5.4.3 were more common in urban regions among 20–49 year olds, who represent the dominant workforce [28] that and may be exposed to *S. Typhi* outside the home. Infections with sub-lineage 3.5.4.2 were more common among men than women. These data suggest different groups of people in different regions have variable exposures to *S. Typhi*. Regrettably, because of the disruptions of COVID-19 and reallocation of resources, it was not possible to undertake a case/control study to verify whether adults and school age children who commute from NWU and ROU to work or school in Apia are at significantly increased risk for developing typhoid versus age- and village-matched persons who remain near home.

Spatial clustering by genotype 3.5.4 sub-lineages revealed variable degrees of clustering (Table 2) and considerable overlap among the household clusters where each sub-lineage was detected (Fig 2A), indicating that the sub-lineages are not geographically restricted to any area of Samoa. In fact, fluctuations and lapses in the absolute and relative proportions of the sub-lineages in each census region over time (Fig 4) indicated exchanges among the regions through a mobile agent (either human carrier or food/water vehicle) and/or a heterogeneous reservoir emitting distinct genotypes at different times. The water supplies are not interconnected between regions, making contaminated products or mobile human carriers more likely to be responsible. Evidence to support an important role for carriers are the 27 household

contacts of acute cases that were asymptotically shedding *S. Typhi*. The subset available for WGS analysis ($n = 14$) demonstrated similar phylogenetic distribution of the isolates among each of the three 3.5.4 sub-lineages (Table 3 and Fig 3) indicating roughly even contribution to the dominant networks of infection. Our experience suggests that asymptomatic shedders are an important and underestimated reservoir of *S. Typhi* in Samoa, detectable only through active surveillance and detailed case investigations.

We also utilized WGS to verify epidemiologic linkages implicated through the Typhoid SWAT team investigations. Related isolates based on limited genomic SNP differences would support these linkages. In our analysis, nearly all linkages were strongly supported by genetic similarity, with a median pairwise SNP distance of 1 SNP, and demonstrated likely within-household transmission events and contributions by asymptomatic shedders, compared to a median of 37 SNPs per genome when the collection of sequenced Samoan *S. Typhi* are compared [15]. These findings are consistent with previous typhoid studies in endemic regions where WGS was paired with household investigations in Jakarta, Kathmandu, and Kolkata [37–39] to support the occurrence of within-household transmission events.

There is not a consensus threshold for SNP differences that would associate *S. Typhi* infections in genomic analyses alone. We detected from 0 to 28 pairwise SNP differences among the 12 putative epidemiologic linkages and 3 instances of repeat positive cultures from the same individual collected ~1 month apart (Table 4). Relatedness criteria of 10 SNPs or less [40] would encompass 13/15 (86.7%) linkages, and 12/13 were within ≤ 4 SNPs, a suggested cutoff for *Salmonella enterica* [41]. Future studies to characterize *S. Typhi* population diversity in acute infections and asymptomatic carriers by sequencing and comparing multiple colonies obtained from stool cultures collected over time, as has been completed for other enteric pathogens [42], could aid in establishing appropriate thresholds for SNP-based relatedness among *S. Typhi* isolates for cluster identification.

This study has several limitations. Some *S. Typhi* isolates from cases and asymptomatic shedders were unavailable for sequencing, resulting in lost genomic data. The sequenced isolates were neither a random selection of a subset for WGS nor a deliberate non-random selection, but rather all the isolates available. Fortunately, all culture-confirmed cases had epidemiologic investigations that included recording the household GPS coordinates and permitting comparisons of the sequenced subset with the total collection of cases; no significant difference in patient sex, mean age, or age groupings was observed (Table 1). Another limitation is that we generally assumed household coordinates were a reasonable indicator of where an individual was infected, a necessary assumption for consistency in the spatial analyses. However, particularly for AUA and NWU cases, there were few epidemiologic investigations of workplaces of adults or schools of schoolchildren, venues where transmission may have occurred.

Sustained surveillance and institution of routine determination of WGS in Samoa could help guide interventions during the Consolidation Phase of the STFCP, which will focus on detection of asymptomatic carriers of *S. Typhi* and possible linked acute cases. Following completion of the mass vaccination of all Samoans 1–45 years of age [43], the genotype/sub-lineages of remaining sporadic cases can be compared to pre-vaccination campaign genotypes/sub-lineages. Findings in the current study build upon a functional “Samoan *S. Typhi* framework” that can inform public health decision making and future genomic epidemiology analyses in Samoa as a paradigm for the understudied region of Oceania.

Supporting information

S1 Text. Detailed methods. Whole genome sequencing, assembly, and mapping. Maximum-likelihood phylogeny. Spatial and temporal point pattern analyses. Average nearest neighbor

(ANN) analysis. Space-time hotspot analysis. Fig A. Range and interpretation of average nearest neighbor ratio (R_n) values. Fig B. Average nearest neighbor output summary diagram. Fig C. Maps of average nearest neighbor datasets.

(DOCX)

S1 Table. Samoa S. Typhi isolates analyzed in this study.

(XLSX)

S2 Table. Pairwise single nucleotide polymorphism (SNP) distances between Samoan S. Typhi genomes in this study.

(XLSX)

Acknowledgments

The authors wish to acknowledge the staff at the Ministry of Health of Samoa for their administrative and technical support of the Samoa Typhoid Fever Control Program, the staff at the Microbiological Diagnostic Unit Public Health Laboratory, Jane Han for her logistical support of the Samoa Typhoid Fever Surveillance Initiative, and Yuanyuan Liang of the Center for Vaccine Development and Global Health for her discussions of statistical methods.

Author Contributions

Conceptualization: Michael J. Sikorski, Myron M. Levine, David A. Rasko.

Data curation: Michael J. Sikorski, Sachin N. Desai, Siasoi Tupua, Michelle Sialeipata, Susan A. Ballard, Mary Valcanis.

Formal analysis: Michael J. Sikorski, Jianguo Ma, Tracy H. Hazen, David A. Rasko.

Funding acquisition: Robert E. Thomsen, Myron M. Levine, David A. Rasko.

Investigation: Michael J. Sikorski, Tracy H. Hazen, Sachin N. Desai, Siasoi Tupua, Susana Nimarota-Brown, Michelle Sialeipata, Savitra Rambocus, Susan A. Ballard, Mary Valcanis, Myron M. Levine, David A. Rasko.

Methodology: Michael J. Sikorski, Jianguo Ma, Tracy H. Hazen, Siasoi Tupua, Susana Nimarota-Brown, Michelle Sialeipata, Savitra Rambocus, Susan A. Ballard, Mary Valcanis, Myron M. Levine, David A. Rasko.

Project administration: Robert E. Thomsen, Roy M. Robins-Browne, Benjamin P. Howden, Take K. Naseri, Myron M. Levine, David A. Rasko.

Resources: Jianguo Ma, Siasoi Tupua, Susana Nimarota-Brown, Michelle Sialeipata, Savitra Rambocus, Susan A. Ballard, Mary Valcanis, Take K. Naseri.

Software: Michael J. Sikorski, Jianguo Ma.

Supervision: Jianguo Ma, Robert E. Thomsen, Roy M. Robins-Browne, Benjamin P. Howden, Take K. Naseri, Myron M. Levine, David A. Rasko.

Validation: Michael J. Sikorski, David A. Rasko.

Visualization: Michael J. Sikorski, Jianguo Ma, Tracy H. Hazen.

Writing – original draft: Michael J. Sikorski, Jianguo Ma, Myron M. Levine, David A. Rasko.

Writing – review & editing: Michael J. Sikorski, Jianguo Ma, Tracy H. Hazen, Sachin N. Desai, Siasoi Tupua, Susana Nimarota-Brown, Michelle Sialeipata, Savitra Rambocus,

Susan A. Ballard, Mary Valcanis, Robert E. Thomsen, Roy M. Robins-Browne, Benjamin P. Howden, Take K. Naseri, Myron M. Levine, David A. Rasko.

References

1. Crump JA. Progress in typhoid fever epidemiology. *Clin Infect Dis*. 2019; 68: S4–S9. <https://doi.org/10.1093/cid/ciy846> PMID: 30767000
2. Mermin JH, Villar R, Carpenter J, Roberts L, Samariddin A, Gasanova L, et al. A massive epidemic of multidrug-resistant typhoid fever in Tajikistan associated with consumption of municipal water. *J Infect Dis*. 1999; 179: 1416–1422. <https://doi.org/10.1086/314766> PMID: 10228063
3. Egoz N, Shihab S, Leitner L, Lucian M. An outbreak of typhoid fever due to contamination of the municipal water supply in northern Israel. *Isr J Med Sci*. 1988; 24: 640–3. PMID: 3215755
4. Sears SD, Ferreccio C, Levine MM, Cordano AM, Monreal J, Black RE, et al. The use of Moore swabs for isolation of *Salmonella typhi* from irrigation water in Santiago, Chile. *J Infect Dis*. 1984; 149: 640–642. <https://doi.org/10.1093/infdis/149.4.640> PMID: 6373964
5. Shuval HI. Investigation of typhoid fever and cholera transmission by raw wastewater irrigation in Santiago, Chile. *Water Sci Technol*. 1993; 27: 167–174. <https://doi.org/10.2166/wst.1993.0341>
6. Walker W. The Aberdeen typhoid outbreak of 1964. *Scott Med J*. 1965; 10: 466–479. <https://doi.org/10.1177/003693306501001203> PMID: 5858949
7. Taylor A, Santiago A, Gonzalez-Cortes A, Gangarosa EJ. Outbreak of typhoid fever in Trinidad in 1971 traced to a commercial ice cream product. *Am J Epidemiol*. 1974; 100: 150–157. <https://doi.org/10.1093/oxfordjournals.aje.a112017> PMID: 4604868
8. Srinivasan M, Sindhu KN, Giri S, Kumar N, Mohan VR, Grassly NC, et al. *Salmonella Typhi* shedding and household transmission by children with blood culture-confirmed typhoid fever in Vellore, South India. *J Infect Dis*. 2021; 224: S593–S600.
9. Soper GA. The curious career of Typhoid Mary. *Bull N Y Acad Med*. 1939; 15: 698–712. PMID: 19312127
10. Olsen SJ, Bleasdale SC, Magnano AR, Landrigan C, Holland BH, Tauxe RV., et al. Outbreaks of typhoid fever in the United States, 1960–99. *Epidemiol Infect*. 2003; 130: 13–21. <https://doi.org/10.1017/s0950268802007598> PMID: 12613741
11. World Health Organization. Typhoid and paratyphoid fevers from 1950 to 1954. *Bull World Health Organ*. 1955; 13: 173–191. PMID: 13260885
12. Cvjetanović B, Grab B, Uemura K. Epidemiological model of typhoid fever and its use in the planning and evaluation of antityphoid immunization and sanitation programmes. *Bull World Health Organ*. 1971; 45: 53–75. PMID: 5316853
13. Ministry of Natural Resources and Environment. National water and sanitation baseline survey. Apia, Samoa; 2015 [cited 2022 Mar 23]. <https://www.mnre.gov.ws/wp-content/uploads/2017/08/Final-Nationa-WASH-Baseline-Report.pdf>.
14. Sikorski MJ, Desai SN, Tupua S, Thomsen RE, Han J, Rambocus S, et al. Tenacious endemic typhoid fever in Samoa. *Clin Infect Dis*. 2020; 71: S120–S126. <https://doi.org/10.1093/cid/ciaa314> PMID: 32725232
15. Sikorski MJ, Hazen TH, Desai SN, Nimarota-Brown S, Tupua S, Sialeipata M, Rambocus S, Ingle DJ, Duchene S, Ballard SA, Valcanis M, Zufan S, Ma J, Sahl JW, Maes M, Dougan G, Thomsen RE, Robins-Browne RM, Howden BP, Naseri TK, Levine MM, Rasko DA. Persistence of Rare *Salmonella Typhi* Genotypes Susceptible to First-Line Antibiotics in the Remote Islands of Samoa. *mBio*. 2022 Sep 12: e0192022. Online ahead of print. <https://doi.org/10.1128/mbio.01920-22> PMID: 36094088
16. Gu H, Fan W, Liu K, Qin S, Li X, Jiang J, et al. Spatio-temporal variations of typhoid and paratyphoid fevers in Zhejiang Province, China from 2005 to 2015. *Sci Rep*. 2017; 7: 1–11. <https://doi.org/10.1038/s41598-017-05928-3> PMID: 28720886
17. Ismail K, Maiga G, Ssebugwaawo D, Nabende P, Mansourian A. Spatio-temporal trends and distribution patterns of typhoid disease in Uganda from 2012 to 2017. *Geospat Health*. 2020; 15: 326–336. <https://doi.org/10.4081/GH.2020.860> PMID: 33461278
18. Besser J, Carleton HA, Germer-Smith P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect*. 2018; 24: 335–341. <https://doi.org/10.1016/j.cmi.2017.10.013> PMID: 29074157
19. Parliament of Samoa. Consolidated Acts of Samoa 2010. Health Ordinance 1959, No.19. Pacific Islands Legal Information Institute; 2010 [cited 2022 Jan 15]. http://www.pacii.org/ws/legis/num_act/haa1969149.

20. Parliament of Samoa. Health Amendment Act 1969, No. 23. Pacific Islands Legal Information Institute; 1969 [cited 2022 Jan 15]. http://www.pacii.org/ws/legis/consol_act/ho1959122.
21. Basnyat B, Qamar FN, Rupali P, Ahmed T, Parry CM. Enteric fever. *BMJ*. 2021; 372: n437. <https://doi.org/10.1136/bmj.n437> PMID: 33637488
22. Perilla MJ, Bopp C, Wells J. *Salmonella* serotype Typhi: Identification and antimicrobial susceptibility testing. In: Manual for the laboratory identification and antimicrobial susceptibility testing of bacterial pathogens of public health importance in the developing world. Geneva: World Health Organization; 2003. pp. 103–162. <https://www.who.int/publications/i/item/WHO-CDS-CSR-RMD-2003.6>.
23. Strockbine NA, Bopp CA, Fields PI, Kaper JB, Nataro JP. *Escherichia*, *Shigella*, and *Salmonella*. In: Jorgensen JH, Pfaller MA, Carroll KC, editors. Manual of clinical microbiology. 11th ed. 11th ed. Washington, DC, USA: ASM Press; 2015. pp. 685–713.
24. Kalra SP, Naithani N, Mehta SR, Swamy AJ. Current trends in the management of typhoid fever. *Med J Armed Forces India*. 2003; 59: 130–135. [https://doi.org/10.1016/S0377-1237\(03\)80060-6](https://doi.org/10.1016/S0377-1237(03)80060-6) PMID: 27407487
25. Esri. ArcGIS Pro v2.9. Redlands, CA: Environmental Systems Research Institute, Inc.; 2021.
26. Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*. 1954; 35: 445–453. <https://doi.org/10.2307/1931034>
27. Pinder DA, Witherick ME. The principles, practice and pitfalls of nearest-neighbour analysis. *Geography*. 1972; 57: 277–288.
28. Samoa Bureau of Statistics. 2016 Census Briefs Nos. 1–4. Apia; 2016 [cited 2022 Jan 15]. <https://www.sbs.gov.ws/populationanddemography>.
29. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. *Nat Commun*. 2016; 7: 12827. <https://doi.org/10.1038/ncomms12827> PMID: 27703135
30. Dyson ZA, Holt KE. Five years of GenoTyphi: Updates to the global *Salmonella* Typhi genotyping framework. *J Infect Dis*. 2021; 224: S775–S780. <https://doi.org/10.1093/infdis/jiab414> PMID: 34453548
31. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
32. Seeman T. Snp-dists v0.8.2. GitHub; 2021. <https://github.com/tseemann/snp-dists>.
33. Hervé M. Package 'RVAideMemoire' v0.9–81. 2021 [cited 2022 Feb 14]. <https://cran.r-project.org/package=RVAideMemoire>.
34. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
35. Craig AT, Heywood AE, Worth H. Measles epidemic in Samoa and other Pacific islands. *Lancet Infect Dis*. 2020; 20: 273–275. [https://doi.org/10.1016/S1473-3099\(20\)30053-0](https://doi.org/10.1016/S1473-3099(20)30053-0) PMID: 32112752
36. UN.ESCAP. Asia-Pacific countries' COVID-19 policy responses: Samoa. 2021 [cited 2022 Jan 15]. https://www.unescap.org/sites/default/d8files/2021-03/Samoa_COVID%20Country%20profile%20230321.pdf.
37. Baker S, Holt K, Van De Vosse E, Roumagnac P, Whitehead S, King E, et al. High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban district of Jakarta, Indonesia. *J Clin Microbiol*. 2008; 46: 1741–1746. <https://doi.org/10.1128/JCM.02249-07> PMID: 18322069
38. Baker S, Holt KE, Clements ACA, Karkey A, Arjyal A, Boni MF, et al. Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol*. 2011; 1: 110008. <https://doi.org/10.1098/rsob.110008> PMID: 22645647
39. Holt KE, Dutta S, Manna B, Bhattacharya SK, Bhaduri B, Pickard DJ, et al. High-resolution genotyping of the endemic *Salmonella* Typhi population during a Vi (typhoid) vaccination trial in Kolkata. *PLoS Negl Trop Dis*. 2012; 6. <https://doi.org/10.1371/journal.pntd.0001490> PMID: 22303491
40. Burns LJ, Kovar LD, Angelo KM, Trees EK, Concepción-Acevedo J, McDermott MD, et al. Use of whole genome sequencing to complement characterisation of a typhoid fever outbreak among a Marshallese community: Oklahoma, 2015. *Epidemiol Infect*. 2019; 147: e11. <https://doi.org/10.1017/S0950268818002601> PMID: 30236166
41. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect*. 2018; 24: 350–354. <https://doi.org/10.1016/j.cmi.2017.12.016> PMID: 29309930

42. Sahl JW, Sistrunk JR, Fraser CM, Hine E, Baby N, Begum Y, et al. Examination of the enterotoxigenic *Escherichia coli* population structure during human infection. Gilmore MS, editor. MBio. 2015; 6: 1–6. <https://doi.org/10.1128/mBio.00501-15> PMID: 26060273
43. Pauulu U. New typhoid fever vaccine roll out. 2021 Aug 25 [cited 2022 Aug 25]. In: Samoa Observer [Newspaper]. Vaitele, Samoa. <https://www.samoobserver.ws/category/samoa/90045>.