

# Enteric fever cluster identification in South Africa using genomic surveillance of *Salmonella enterica* serovar Typhi

Anthony Marius Smith<sup>1,2,\*</sup>, Linda Kathleen Erasmus<sup>1</sup>, Nomsa Pauline Tau<sup>1</sup>, Shannon Lucrecia Smouse<sup>1</sup>, Hlengiwe Mimmy Ngomane<sup>1</sup>, Bolele Disenyeng<sup>1</sup>, Andrew Whitelaw<sup>3,4</sup>, Charlene Ann Lawrence<sup>5</sup>, Phuti Sekwadi<sup>1</sup> and Juno Thomas<sup>1</sup>

## Abstract

The National Institute for Communicable Diseases in South Africa participates in national laboratory-based surveillance for human isolates of *Salmonella* species. Laboratory analysis includes whole-genome sequencing (WGS) of isolates. We report on WGS-based surveillance of *Salmonella enterica* serovar Typhi (*Salmonella* Typhi) in South Africa from 2020 through 2021. We describe how WGS analysis identified clusters of enteric fever in the Western Cape Province of South Africa and describe the epidemiological investigations associated with these clusters. A total of 206 *Salmonella* Typhi isolates were received for analysis. Genomic DNA was isolated from bacteria and WGS was performed using Illumina NextSeq technology. WGS data were investigated using multiple bioinformatics tools, including those available at the Centre for Genomic Epidemiology, EnteroBase and Pathogenwatch. Core-genome multilocus sequence typing was used to investigate the phylogeny of isolates and identify clusters. Three major clusters of enteric fever were identified in the Western Cape Province; cluster one ( $n=11$  isolates), cluster two ( $n=13$  isolates), and cluster three ( $n=14$  isolates). To date, no likely source has been identified for any of the clusters. All isolates associated with the clusters, showed the same genotype (4.3.1.1.EA1) and resistome (antimicrobial resistance genes: *bla*<sub>TEM-1B</sub>, *catA1*, *sul1*, *sul2*, *dfrA7*). The implementation of genomic surveillance of *Salmonella* Typhi in South Africa has enabled rapid detection of clusters indicative of possible outbreaks. Cluster identification allows for targeted epidemiological investigations and a timely, coordinated public health response.

## DATA SUMMARY

Supplementary material can be found in the online version of this article. These supplementary material include: (1) NICD Enteric fever (typhoid and paratyphoid fever): Recommendations for diagnosis, management and public health response [1], Enteric fever case investigation form [2], Metadata for *Salmonella* Typhi isolates, Western Cape Province, South Africa, 2020–2021.

All sequencing data were uploaded to the public EnteroBase platform (<http://enterobase.warwick.ac.uk/species/index/senterica>) and so are freely available to access at the EnteroBase platform. In addition, sequencing data are deposited in the European Nucleotide Archive under the project accession number PRJEB39546.

## INTRODUCTION

*Salmonella enterica* remains a major cause of human disease worldwide [1, 3]. *Salmonella enterica* serovar Typhi (*Salmonella* Typhi), the causative agent of typhoid fever, is of particular concern. The World Health Organization estimates the global typhoid

Received 16 November 2022; Accepted 12 May 2023; Published 20 June 2023

**Author affiliations:** <sup>1</sup>Centre for Enteric Diseases, National Institute for Communicable Diseases, Division of the National Health Laboratory Service, Johannesburg, South Africa; <sup>2</sup>Department of Medical Microbiology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa; <sup>3</sup>Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; <sup>4</sup>National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa; <sup>5</sup>Communicable Disease Control, Service Priorities Coordination, Department of Health, Cape Town, South Africa.

\*Correspondence: Anthony Marius Smith, [anthonys@nicd.ac.za](mailto:anthonys@nicd.ac.za)

**Keywords:** *Salmonella* Typhi; enteric fever; whole-genome sequencing; genomic; WGS; surveillance; cluster; outbreak; South Africa; Africa.

**Abbreviations:** AMR, antimicrobial resistance; CED, centre for enteric diseases; cgMLST, core-genome multilocus sequence typing; MDR, multidrug resistance; NGS, next-generation sequencing; NICD, National Institute for Communicable Diseases; *Salmonella* Typhi, *Salmonella enterica* serovar Typhi; SNP, single nucleotide polymorphism; WGS, whole-genome sequencing.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary material is available with the online version of this article.

001044 © 2023 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

## Impact Statement

*Salmonella* Typhi, the causative bacterial agent of typhoid fever, remains endemic in many African countries. Active surveillance for typhoid fever and rapid identification/characterization of the causative agent is important to initiate appropriate public health responses, identify the source of the agent and control spread of the disease. Whole-genome sequencing (WGS) analysis of bacterial pathogens is the ultimate laboratory methodology. WGS data allows for full characterization of bacterial isolates, and importantly can be used to investigate the genetic relatedness of isolates, in order to investigate the source and spread of isolates. WGS is revolutionizing and transforming communicable diseases surveillance, allowing for more targeted investigations. WGS data can be used to identify clusters of cases in real-time and complement epidemiological investigation of outbreaks. Cluster identification allows for targeted epidemiological investigations and a timely, coordinated public health response. The National Institute for Communicable Diseases (NICD), South Africa participates in national laboratory-based surveillance for human isolates of *Salmonella* species. Laboratory analysis includes WGS of isolates. We report on WGS-based surveillance of *Salmonella* Typhi in South Africa from 2020 through 2021. We describe how WGS analysis identified clusters of enteric fever in the Western Cape Province of South Africa and describe the epidemiological investigations associated with these clusters.

fever disease burden at 11–20 million cases yearly, resulting in about 128000–161000 deaths per year [2]. Sub-Saharan Africa bears a substantial burden of disease, with most countries in this region reported to be endemic for typhoid fever [4–7].

In South Africa, the incidence rates of typhoid fever have decreased markedly over the last decades, but the disease remains endemic [8]. Over the past 15 years, an average of 97 laboratory-confirmed cases of enteric fever were reported per year [9, 10]. Outbreaks of typhoid fever have been documented in South Africa; the most recent large outbreak having occurred in 2005, in Delmas, Mpumalanga Province, where over 2900 cases were reported [11–13]. Typhoid fever remains a public health problem in the country, and it is important that a robust surveillance system is in place to initiate a timely public health response to every case, and to rapidly identify possible outbreaks. Enteric fever is a notifiable medical condition in South Africa, meaning that all laboratory-confirmed cases must be reported to the National Department of Health. This notification system is complemented by appropriate laboratory characterization of *Salmonella* Typhi isolates, including molecular subtyping of isolates to investigate their genetic relatedness [11, 14, 15]. The Centre for Enteric Diseases (CED) at the National Institute for Communicable Diseases (NICD) has recently integrated whole-genome sequencing (WGS) analysis of all *Salmonella* Typhi isolates into its surveillance system. This step to WGS was necessary to keep abreast with advances in microbiological methodologies and follow global trends in that many reference laboratories in global public institutions have transitioned to WGS as their primary methodology for molecular subtyping and characterization of bacterial pathogens [16–19]. WGS is revolutionizing and transforming communicable diseases surveillance, allowing for more targeted investigations. WGS data can be used to identify clusters of cases in real-time and complement epidemiological investigation of outbreaks.

We report on the use of WGS for real-time genomic surveillance of *Salmonella* Typhi in South Africa from 2020 through 2021. We describe how genomic analysis led to the identification of multiple clusters of enteric fever in the Western Cape Province and describe the epidemiological findings associated with the investigation of these clusters.

## METHODS

### Surveillance for *Salmonella* Typhi in South Africa

The NICD is the national public health institute of South Africa, providing reference microbiology, virology, epidemiology, surveillance and public health research and training to support the government's response to communicable disease threats. The CED at the NICD participates in national laboratory-based surveillance for human isolates of *Salmonella* species. All clinical isolates of *Salmonella* species are collected by voluntary submissions from >200 clinical microbiology laboratories across the country [9, 10]. Following a *Salmonella* species identification at a clinical microbiology laboratory, the isolate is received at the CED within 1 week. The CED then immediately commences with phenotypic characterization and WGS analysis (as described below), which is completed within 2–3 weeks. Notification of *Salmonella* Typhi cases through a notifiable medical conditions surveillance system enables audit and follow up of isolates that were not received. Isolates are accompanied by basic metadata (demographic data) including patient details, place of residence, and date of specimen collection.

### Epidemiologic case investigation

Enteric fever is a notifiable disease in South Africa, with standard guidelines for case investigation (Supplementary material, available in the online version of this article) [20]. Health authorities received notification directly through the national notifiable medical conditions surveillance system within 24 h of laboratory confirmation and initiated the case investigation. Clinical

and demographic details and underlying medical conditions were ascertained through patient interviews or abstracted from medical records with the use of a standardized case-investigation form (Supplementary material). Case patients (or next of kin or caregivers in cases in which the patient was a child, had died, or was too ill to respond) were interviewed to ascertain risk factors for infection with *Salmonella* Typhi, including travel, occupation, sources of water used in the home, exposure to water sources outside the home, sanitation and hygiene in the home, and sources of food consumed in the month prior to onset of illness. Faecal samples (stool or rectal swabs) were collected from household contacts where possible. Health authorities conducted an environmental health assessment at the patient's home (and when indicated, place of work) which informed further investigation of possible foodborne or waterborne sources of infection. This may have included sampling of food or water. Water samples were collected by environmental health authorities. At least one litre of water was collected per sampling site and stored at 2–8 °C during transport to the laboratory, and tested according to ISO (International Organization for Standardization) standard 19250 for the detection of *Salmonella*. Additionally, when clusters were identified on genomic analysis all case investigations were reviewed and additional investigations performed.

### Phenotypic characterization of bacteria

The CED received *Salmonella* Typhi on Dorset-Egg transport media [Diagnostic Media Products (DMP), National Health Laboratory Service, Johannesburg, South Africa] and sub-cultured onto 5% Blood Agar (DMP), to check for viability and purity. Cultures were identified using standard phenotypic microbiological identification and serotyping techniques, briefly described as follows. As required, bacterial colonies were identified using the VITEK-2 COMPACT 15 automated microbial identification system (bioMérieux, Marcy-l'Étoile, France). Serotyping was performed according to the White-Kauffmann-Le Minor Scheme [21]. Antimicrobial (ampicillin, chloramphenicol, ciprofloxacin, ceftriaxone, azithromycin) susceptibility testing was performed using the Etest method (bioMérieux). Interpretation of antimicrobial susceptibility data was done in accordance with the Clinical and Laboratory Standards Institute (CLSI) [22].

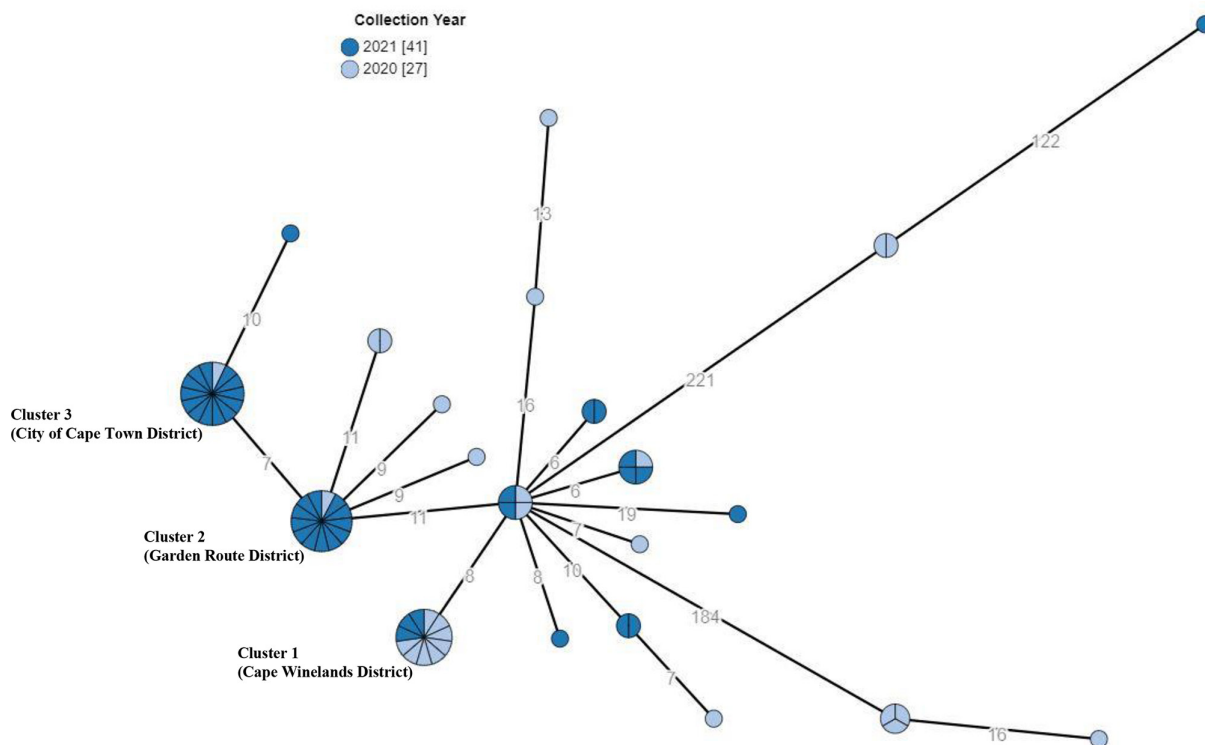
### Whole-genome sequencing (WGS) of bacteria

Genomic DNA was isolated from bacteria using an Invitrogen PureLink Microbiome DNA Purification Kit (Invitrogen, Waltham, Massachusetts, USA). WGS was performed using Illumina NextSeq (Illumina, San Diego, California, USA) next generation sequencing technology, with DNA libraries prepared using a Nextera DNA Flex Library Preparation Kit (Illumina), followed by 2×150bp paired-end sequencing runs with ~80 times coverage. All these methodologies associated with WGS, were performed using the protocols and standard operating procedures as described by the manufacturers of the kits and equipment.

### Analysis of WGS data

Illumina paired-end reads were analysed using the JEKESA bioinformatics pipeline version 1.0 (<https://github.com/stanikae/jekesa>), which incorporates multiple bioinformatics analysis tools, briefly described as follows. All underlying tools were set to default options, unless otherwise stated below. Quality control and read filtering of raw paired-end reads were performed using FastQC version 0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and TrimGalore version 0.6.2 (<https://github.com/FelixKrueger/TrimGalore>) set to a minimum Phred quality score of 30 and minimum read length of 50 bp. Species identification and closest reference detection were performed using BactInspector version 0.1.3 (<https://gitlab.com/antunderwood/bactinspector>). Contamination checks were performed using ConFindr version 0.7.4 [23] and Kraken2 version 2.0.8-beta [24]. *De novo* assembly was performed using SKESA version 2.3.0 [25], and the assemblies were optimized using Shovill version 1.1.0 (<https://github.com/tseemann/shovill>) with depth set to 100 and minimum contig length set to 200. Assembly metrics were assessed using QUAST version 5.0.2 [26]. The assembled genomes were further investigated using the following tools. Multilocus sequence typing (MLST) was performed using mlst version 2.19.0 (<https://github.com/tseemann/mlst>), based on PubMLST typing schemes (<https://pubmlst.org/>). Detection of antimicrobial resistance determinants were performed using ResFinder version 4.1 (<http://www.genomicepidemiology.org/services/>) [27]. *Salmonella* serovar prediction was performed using SeqSero2 version 1.1.0 (<http://denglab.info/SeqSero2>) [28] and SISTR version 1.1.2 ([https://github.com/phac-nml/sistr\\_cmd](https://github.com/phac-nml/sistr_cmd)) [29]. Plasmid DNA presence was investigated using PlasmidFinder version 2.1 (<http://www.genomicepidemiology.org/services/>). Information regarding *Salmonella* Typhi genotype [as called by single nucleotide polymorphism (SNP) analysis] was determined using the GenoTyphi Scheme [30] at the Pathogenwatch platform (<https://pathogen.watch/>).

Core-genome MLST (cgMLST) was used to investigate the phylogeny and genetic relatedness of isolates. This was made possible using the cgMLST scheme available at the EnteroBase platform (<http://enterobase.warwick.ac.uk/species/index/senterica>). Raw sequencing data (FastQ files for paired-end reads) were uploaded and investigated at EnteroBase using the cgMLST tool (using the 'cgMLST V2 +HierCC V1' scheme which incorporates analysis at 3002 genes). Further information regarding the operational tools of EnteroBase and tutorials on how to use the tools, can be found at the following URL links: <https://enterobase.readthedocs.io/en/latest/>, <https://bitbucket.org/enterobase/enterobase-web/wiki/GrapeTree>, <https://bitbucket.org/enterobase/enterobase-web/wiki/Tutorials>, <https://enterobase.readthedocs.io/en/latest/enterobase-tutorials/tutorials.html>. In addition, EnteroBase tools have previously been described by Zhou and coworkers [31]. The phylogeny and genetic relatedness of isolates was depicted using a



**Fig. 1.** Minimum spanning tree drawn using cgMLST data from *Salmonella* Typhi isolates ( $n=68$ ) sourced from the Western Cape Province of South Africa, 2020-2021. The circular nodes represent isolate(s). Isolates showing  $\leq 5$  allelic differences, are collapsed together into a single circular node. The larger the circular node, the more isolates which are reflected. The number of segments within a circular node are indicative of the number of isolates. The number values between adjacent nodes indicate the number of allele differences between connecting nodes (isolates). Three major clusters of isolates are indicated.

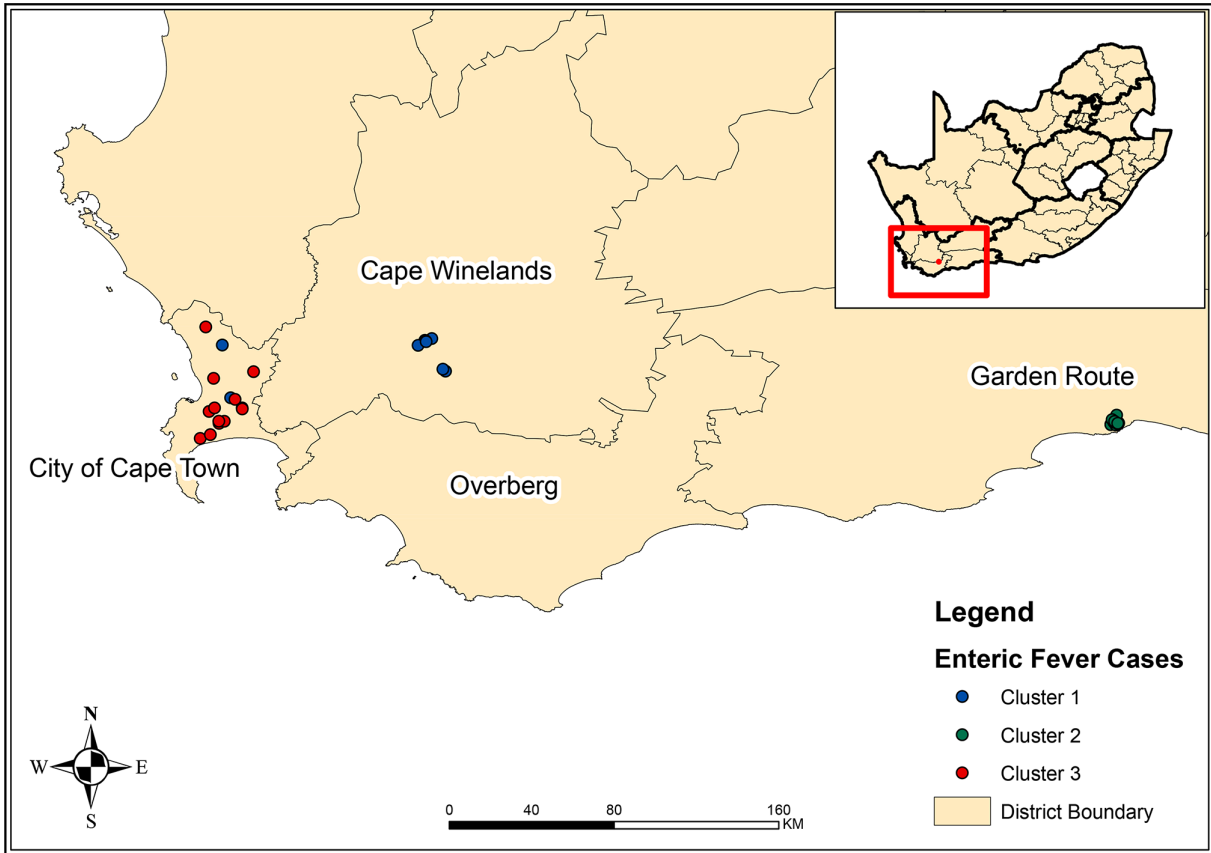
GrapeTree-generated minimum spanning tree using the 'MSTree V2' algorithm, tools which are all integrated within EnterBase [32]. Once the GrapeTree was produced, the settings/operators of the tool were set to 'collapse branches' at a value of '5', which resulted in isolates showing  $\leq 5$  allelic differences to collapse together into a 'cluster'. This was how a cluster of isolates was created and visualized. Our cluster definition was  $\geq 3$  isolates showing  $\leq 5$  allelic differences, as obtained by the above actions, following cgMLST analysis and generation of a GrapeTree. It is important to note that clusters were not defined by the cgMLST hierarchical cluster tool, i.e. not defined at the hierarchical cluster level 5 (HC5). In addition, clusters generated by our above described actions (visualization on a GrapeTree) are not always directly comparable to cluster number generated by HC5.

We further investigated genetic variation within clusters using a single nucleotide polymorphism (SNP) analysis of isolates using CSI Phylogeny version 1.4 (<https://cge.food.dtu.dk/services/CSIPhylogeny/>) [33].

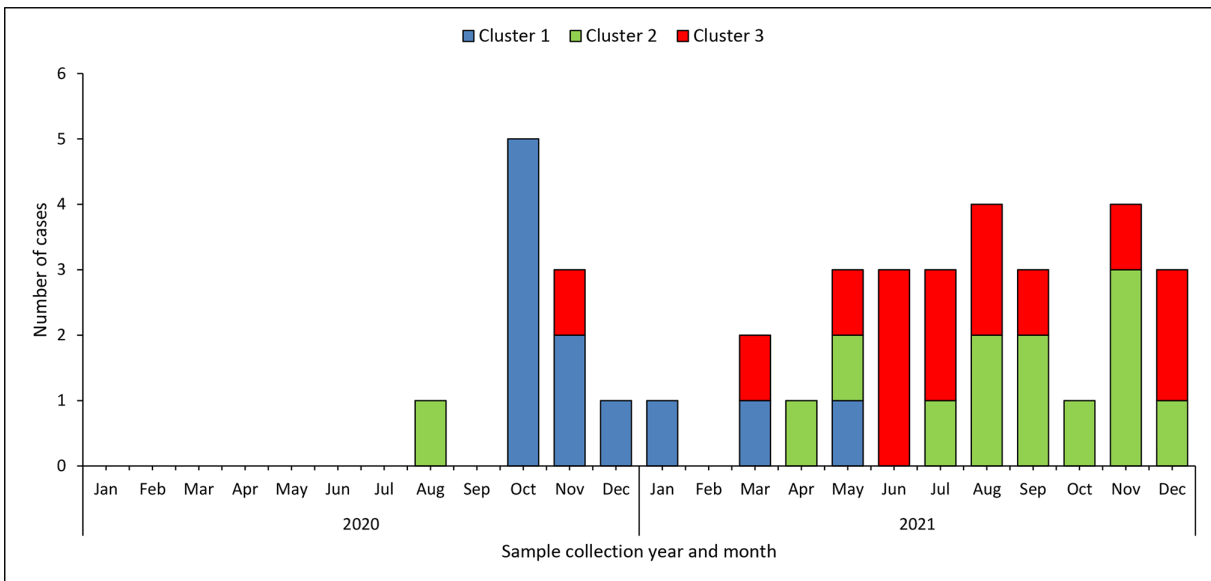
## RESULTS

We performed an analysis of WGS and cgMLST data of *Salmonella* Typhi isolates ( $n=68$ ) from the Western Cape Province of South Africa for 2020 through 2021. As per our cluster definition (described in the Methodology section), we identified three major clusters of isolates ( $\geq 11$  isolates per cluster) and three minor clusters of isolates ( $\leq 4$  isolates per cluster) (Fig. 1). The minor clusters were not further investigated due to human resource constraints. We simply do not have the resources to perform extended epidemiological investigations on every case. So, all our efforts were directed to focus on the three major clusters. These major clusters were identified in three districts of the Western Cape Province (Fig. 2). Cluster one was associated with Cape Winelands District (Breede Valley Sub-District) ( $n=11$  isolates), cluster two with Garden Route District (George Sub-District) ( $n=13$  isolates), and cluster three with the City of Cape Town District ( $n=14$  isolates) (Figs 1–3) (Supplementary material). Epidemiological investigations were conducted on these three major clusters.

Cluster one comprised 11 cases, identified from October 2020 through May 2021 (Figs 1–3). Ten cases were diagnosed on blood culture and one on stool culture. The majority of case patients (8/11; 73%) were 15 years of age or younger, and males predominated (8/11; 73%). Nine of the eleven case patients (9/11; 82%) lived in Cape Winelands District, and two lived in City



**Fig. 2.** Map of the Western Cape Province of South Africa showing the location of three major *Salmonella* Typhi clusters identified from 2020 through 2021.



**Fig. 3.** Epidemiological curve showing the month of sample collection in 2020–2021, for all isolates associated with three major *Salmonella* Typhi clusters identified in the Western Cape Province of South Africa.

**Table 1.** Genotypic properties of *Salmonella* Typhi isolates associated with three major clusters identified in the Western Cape Province of South Africa, 2020–2021

Cluster no.	Mlst*	cgMLST HC2 no.†	cgMLST HC5 no.†	cgMLST HC50 no.†	Genotype‡	Presence of acquired AMR genes	Mutations in the QRDRs of <i>gyrA</i> , <i>gyrB</i> , <i>parC</i> and <i>parE</i>	Plasmid presence
1	1	26478	26478	202	4.3.1.1.EA1	<i>bla</i> <sub>TEM-1B</sub> , <i>catA1</i> , <i>sul1</i> , <i>sul2</i> , <i>dfrA7</i>	None	IncQ1
2	1	256191 (n=12), 295929 (n=1)	202	202	4.3.1.1.EA1	<i>bla</i> <sub>TEM-1B</sub> , <i>catA1</i> , <i>sul1</i> , <i>sul2</i> , <i>dfrA7</i>	None	IncQ1
3	1	12503 (n=13), 221038 (n=1)	202	202	4.3.1.1.EA1	<i>bla</i> <sub>TEM-1B</sub> , <i>catA1</i> , <i>sul1</i> , <i>sul2</i> , <i>dfrA7</i>	None	IncQ1

\*As per assignment at PubMLST (<https://pubmlst.org/>).

†As per assignment at Enterobase (<http://enterobase.warwick.ac.uk/species/index/sentérica>).

‡As per assignment at Pathogenwatch (<https://pathogen.watch/>).

cgMLST, core-genome MLST; MLST, multilocus sequence typing; QRDRs, quinolone resistance-determining regions.

of Cape Town District but reported recent travel to Cape Winelands District. Four case-patients reported swimming in a river-water reservoir system in Cape Winelands District, so contamination of this surface water system was initially suspected to be the source of infection. However, *Salmonella* species was not isolated from multiple water samples collected from the river and water reservoir, and since most cases reported no exposure to this water system, it was concluded that this river-water reservoir system was unlikely to be a source of infection.

Cluster two comprised 13 cases. The first case was identified in August 2020, followed by an 8 month hiatus, after which 12 cases were identified from April 2021 through to December 2021 (Figs 1–3). Nine cases were diagnosed on blood culture and four on stool culture. The majority of case patients (11/13; 85%) were 15 years of age or younger, and males predominated (8/13; 62%). All case patients lived in the same George Sub-District; 8/13 lived within the same township and 2/8 lived in the same street. The case-patients lived in residential areas with inadequate sanitation and unhygienic conditions, including illegally dumped waste, pools of stagnant water from overflowing storm water drains, overflowing sewage works and man-made urinal stalls. Several samples of stagnant water pools and river water were tested, but *Salmonella* species was not detected. Case-patients of school-going age attended several different schools in the district, but through a school feeding programme all scholars in the district were served daily meals prepared by a single service provider. Attempts to investigate the service provider were unsuccessful and stool samples could not be collected from food handlers, so this possible common source of infection could not be confirmed or excluded.

Cluster three comprised 14 cases. The first case was identified in November 2020, followed by a 4 month hiatus, after which 13 cases were identified from March 2021 through to December 2021 (Figs 1–3). All cases were diagnosed on blood culture. Among case patients, males predominated (9/14; 64%); the age of patients ranged from 7 to 54 years (median age of 21 years), with only five cases (36%; 5/14) aged 15 years or younger. The cases were detected in the City of Cape Town District; case-patients resided in multiple townships in a large geographical area and were diagnosed at several public hospitals in the district. Investigations revealed no obvious epidemiological links between cases, and nothing to suggest exposure to a common source of infection. The widespread distribution of cases across the district could not inform targeted investigation with environmental sampling.

Epidemiological investigation of these clusters resulted in public health responses and activities, which included, the detection and treatment of asymptomatic carriers among the close contacts of cases, health communication to affected communities, and more frequent environmental sampling to monitor water quality.

All isolates associated with the three clusters showed similar phenotypic and genotypic properties. Genotypic properties for isolates are summarized in Table 1 and are briefly described as follows. All isolates were ST1, all showed genotype 4.3.1.1.EA1, and all showed a multidrug resistance (MDR) genotype with presence of multiple antimicrobial resistance (AMR) genes (*bla*<sub>TEM-1B</sub>, *catA1*, *sul1*, *sul2*, and *dfrA7*). Phenotypically, the isolates all showed resistance to ampicillin (MIC, 256 µg ml<sup>-1</sup>) and chloramphenicol (MIC, 256 µg ml<sup>-1</sup>); while all showed susceptibility to ciprofloxacin (MIC, ≤0.06 µg ml<sup>-1</sup>), ceftriaxone (MIC, ≤1 µg ml<sup>-1</sup>), and azithromycin (MIC, ≤16 µg ml<sup>-1</sup>).

## DISCUSSION

In recent years, next-generation sequencing (NGS) technology has rapidly advanced, resulting in shorter turnaround-times to WGS results and higher accuracy of sequencing data [19]. In parallel, costs of NGS have also decreased dramatically, making the technology more affordable and cost-effective for use in public health laboratories [18]. As a result, reference laboratories in many public institutions across the world have transitioned to WGS as the primary methodology for assessing relatedness of bacterial isolates for surveillance activities, cluster searches and outbreak investigations [16, 17]. The NICD is a national public health institute in South Africa and houses a Sequencing Core Facility which is well-established, well-equipped and well-resourced

to perform NGS in all applications of the technology. Many reference laboratories at the NICD have already implemented NGS and WGS for surveillance, outbreak investigation and research purposes. In particular, the CED at the NICD has implemented routine WGS analysis of all outbreak-prone enteric bacterial pathogens (including *Salmonella* species) into its laboratory-based surveillance system [34–36]. Our turnaround time, from receipt of a culture at the CED laboratory to completion of analysis of WGS data, is 2–3 weeks.

In South Africa, over the past 15 years, an average of 97 laboratory-confirmed cases of enteric fever were reported per year (range 66–140) [37]. In 2020 and 2021, 89 and 134 cases were reported per year respectively. Of the 223 cases in these 2 years, *Salmonella* Typhi isolates were received for 92% (206/223). Although the national number of cases of enteric fever in 2021 was within the expected range, at a provincial level there was a striking increase in cases in the Western Cape Province. From 2003 through 2020 an average of 21 cases (range 8–40) were reported annually in this province; in 2021, a total of 52 cases were reported, which was the highest number of annual cases since 2003 [37]. This increase in cases in Western Cape Province was from three districts (Cape Winelands, Garden Route and City of Cape Town) (Fig. 2). Using WGS analysis of *Salmonella* Typhi isolates, we identified three major enteric fever clusters in the Western Cape Province of South Africa from 2020 through 2021. A cluster indicates a possible common source of infection and could herald a larger outbreak. The relevant health departments were notified of the clusters, and a multi-stakeholder team was convened to investigate them. Unfortunately, no confirmed source(s) of infection were identified for any of the clusters. Contamination of municipal potable water was extremely unlikely to be the source of infection in any of the clusters, due to the demographics of the cases (including age profiles, places of residence, source(s) of drinking water and access to improved sanitation), and culture-negative environmental sampling. In particular, the size of the clusters (fewer than 15 cases in each), and the slow accumulation of cases over several months provided further evidence to exclude contamination of municipal potable water as the primary source of infection. It is very likely that there are complex chains of transmission within the respective communities, mostly due to the presence of unrecognised cases and carriers who serve as reservoirs of infection and lead to ongoing transmission. This makes it challenging to investigate and pinpoint source(s) and reservoir(s) of infection within the communities. The same difficulty in identifying the source(s) of genetically highly related *Salmonella* Typhi strains has been reported elsewhere [38].

All isolates associated with our three major enteric fever clusters showed genotype 4.3.1.1.EA1 (the H58 haplotype strain) [30, 39]. The cgMLST hierarchical cluster tool at the EnteroBase platform was used to further interrogate our isolates (Table 1). At HC50, where isolates are clustered at 50 allele differences; all cluster isolates showed HC50:202. The EnteroBase assignment HC50:202 is known to be indicative of the H58 haplotype, so all was in agreement with the SNP-defined assignment of genotype 4.3.1.1.EA1. Interrogation at the HC5 and HC2 levels provided more sensitive discrimination of our isolates. For cluster one isolates, at the HC2 level, all were assigned as HC2:26478. This HC2 assignment has only been reported for isolates sourced in South Africa (from 2020 onwards), with the exception of a single isolate sourced in England in 2014 (European Nucleotide Archive biosample accession number SAMN03475542). This English case had reported travel history to Zimbabwe, a country which neighbours South Africa (M. A. Chattaway, UK Health Security Agency, personal communication, 23 February 2023). For cluster two isolates, at the HC2 level, isolates were assigned either HC2:256191 ( $n=12$ ) or HC2:295929 ( $n=1$ ). The HC2:256191 assignment has only been reported for isolates sourced in South Africa (from 2020 onwards). For cluster three isolates, at the HC2 level, isolates were assigned either HC2:12503 ( $n=13$ ) or HC2:221038 ( $n=1$ ). The HC2:12503 assignment has previously been reported for isolates sourced in South Africa (from 2010 onwards) and Malawi isolates (sourced from 2011 onwards). We further investigated genetic variation within clusters using a SNP analysis of isolates. Cluster one isolates showed 0–1 SNP differences, cluster two isolates showed 0–6 SNP differences, and cluster three isolates showed 0–3 SNP differences. The level of SNP variation within clusters showed concordance/congruence to the level of allele variation determined using cgMLST. These concordance of results have previously been reported when investigating clusters and outbreaks of enteric bacterial pathogens [18, 40–45].

*Salmonella* Typhi genotype 4.3.1 is commonly associated with AMR, and well-established in several African countries, including Tanzania, Kenya, Malawi, Zambia, Zimbabwe and South Africa [30, 39, 46]. Genotype 4.3.1 is subdivided into three major lineages (4.3.1.1, 4.3.1.2, and 4.3.1.3). All isolates associated with our Western Cape Province clusters showed a sublineage of 4.3.1.1, named genotype 4.3.1.1.EA1 (genotype 4.3.1 lineage I sublineage East Africa I). All isolates associated with our Western Cape Province clusters showed MDR, but thankfully, all isolates showed phenotypic susceptibility to ciprofloxacin, ceftriaxone and azithromycin, which are the agents currently recommended for the treatment of typhoid fever in South Africa [20].

Limitations of our study included the following. Although enteric fever is a notifiable medical condition in South Africa (such that all laboratory-confirmed cases must be reported to the Department of Health), reported cases significantly underrepresent the true number of cases. The likelihood that enteric fever cases are identified and diagnosed depends on many factors, including the severity of illness, whether health workers are aware of the disease, and whether blood culture capacity is available. WGS data is only available on isolates submitted by diagnostic laboratories to the reference laboratory at the CED, NICD; in a few instances isolates were not sent to the CED. Limitations of the epidemiologic investigations for these clusters included the inability to trace some of the close contacts (usually because of incorrect addresses or contact numbers), and that stool samples from food handlers working for the school feeding programme service provider were not collected.

In conclusion, we describe the use of WGS for surveillance and cluster detection of enteric fever in South Africa. Cluster identification allowed for targeted epidemiological investigations and timely, coordinated public health responses. Unfortunately, no confirmed source(s) of infection were identified for any of the clusters. Nonetheless, the public health response and outbreak investigations raised awareness of enteric fever among health officials, health workers and communities in the relevant districts and helped to inform control measures. Following the public health response, there was a marked decline in the number of cases of enteric fever diagnosed in these districts; it is likely that interventions implemented during the outbreak response, including health communication and enhanced water quality monitoring, contributed to this decline.

#### Funding information

This study was made possible by support from the SEQAFRICA project which is funded by the Department of Health and Social Care's Fleming Fund using UK aid. The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health and Social Care or its Management Agent, Mott MacDonald.

#### Acknowledgement

We thank colleagues in the Western Cape Provincial Centre for Disease Control and Prevention Directorate and the Western Cape Department of Health for their assistance during the study.

We thank Jimmy Khosa for preparing the figure showing the map of the Western Cape Province of South Africa.

We thank all participants of the NICD GERMS-SA Laboratory Surveillance Network for submission of clinical isolates of *Salmonella* species to the NICD.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### Ethical statement

Ethical approval was obtained from the Human Research Ethics Committee of the University of the Witwatersrand (protocol reference number M1809107).

#### References

1. **GTaP C.** 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 (WHO).** Typhoid 31; 2018. <https://www.who.int/news-room/fact-sheets/detail/typhoid> [accessed 25 November 2021].
3. **Collaborators G-TSID.** The global burden of non-Typhoidal salmonella invasive disease: a systematic analysis for the global burden of disease study 2017. *Lancet Infect Dis* 2019;19:1312–1324.
4. **Kim JH, Mogasale V, Im J, Ramani E, Marks F.** Updated estimates of typhoid fever burden in sub-Saharan Africa. *Lancet Glob Health* 2017;5:e969.
5. **Pitzer VE, Meiring J, Martineau FP, Watson CH, Kang G, et al.** The invisible burden: diagnosing and combatting typhoid fever in Asia and Africa. *Clin Infect Dis* 2019;69:S395–S401.
6. **Marks F, von Kalckreuth V, Aaby P, Adu-Sarkodie Y, El Tayeb MA, et al.** Incidence of invasive salmonella disease in sub-Saharan Africa: a multicentre population-based surveillance study. *Lancet Glob Health* 2017;5:e310–e323.
7. **Kim CL, Cruz Espinoza LM, Vannice KS, Tadesse BT, Owusu-Dabo E, et al.** The burden of typhoid fever in sub-Saharan Africa: a perspective. *Res Rep Trop Med* 2022;13:1–9.
8. **Keddy KH, Smith AM, Sooka A, Tau NP, Ngomane HMP, et al.** The burden of typhoid fever in South Africa: the potential impact of selected interventions. *Am J Trop Med Hyg* 2018;99:55–63.
9. **GERMS-SA Annual Surveillance Review;** 2017. <https://www.nicd.ac.za/wp-content/uploads/2018/08/GERMS-SA-AR-2017-final.pdf> [accessed 1 September 2022].
10. **GERMS-SA Annual Surveillance Review;** 2018. <https://www.nicd.ac.za/wp-content/uploads/2019/11/GERMS-SA-AR-2018-Final.pdf> [accessed 1 September 2022].
11. **Smith AM, Keddy KH, Ismail H, Thomas J, van der Grÿp R, et al.** International collaboration tracks typhoid fever cases over two continents from South Africa to Australia. *J Med Microbiol* 2011;60:1405–1407.
12. **Makungo UB, Ramutshila TE, Mabotja MC, Thomas J, Lekalakala-Mokaba R, et al.** Epidemiological investigation of a typhoid fever outbreak in Sekhukhune District, Limpopo province, South Africa – 2017. *S Afr J Infect Dis* 2020;35:107.
13. **Keddy KH, Sooka A, Ismail H, Smith AM, Weber I, et al.** Molecular epidemiological investigation of a typhoid fever outbreak in South Africa, 2005: the relationship to a previous epidemic in 1993. *Epidemiol Infect* 2011;139:1239–1245.
14. **GERMS-SA Surveillance Network, Smith AM, Smouse SL, Tau NP, Bamford C, et al.** Laboratory-acquired infections of *Salmonella enterica* serotype Typhi in South Africa: phenotypic and genotypic analysis of isolates. *BMC Infect Dis* 2017;17.
15. **Tau NP, Smith AM, Wain JR, Tarupiwa A, Coulibaly KJ, et al.** Development and evaluation of a multiple-locus variable-number tandem-repeats analysis assay for subtyping *Salmonella* Typhi strains from sub-Saharan Africa. *J Med Microbiol* 2017;66:937–945.
16. **Ribot EM, Freeman M, Hise KB, Gerner-Smidt P.** PulseNet: entering the age of next-generation sequencing. *Foodborne Pathog Dis* 2019;16:451–456.
17. **Chattaway MA, Dallman TJ, Larkin L, Nair S, McCormick J, et al.** The transformation of reference microbiology methods and surveillance for *Salmonella* with the use of whole genome sequencing in England and Wales. *Front Public Health* 2019;7:317.
18. **Kubota KA, Wolfgang WJ, Baker DJ, Boxrud D, Turner L, et al.** PulseNet and the changing paradigm of laboratory-based surveillance for foodborne diseases. *Public Health Rep* 2019;134:225–285.
19. **Armstrong GL, MacCannell DR, Taylor J, Carleton HA, Neuhaus EB, et al.** Pathogen genomics in public health. *N Engl J Med* 2019;381:2569–2580.
20. **National Institute for Communicable Diseases (NICD), South Africa.** Enteric fever (typhoid and paratyphoid fever): Recommendations for diagnosis, management and public health response. Available; 2022. [https://www.nicd.ac.za/wp-content/uploads/2022/06/Enteric-Fever\\_NICD-recommendations\\_June-2022\\_final.pdf](https://www.nicd.ac.za/wp-content/uploads/2022/06/Enteric-Fever_NICD-recommendations_June-2022_final.pdf) [accessed 1 September 2022].
21. **Grimont PAD, Weill F-X.** Antigenic formulae of the *Salmonella* serovars. In: *WHO Collaborating Centre for Reference and Research on Salmonella*, 9th edition. Paris, France: Institut Pasteur, 2007. [https://www.pasteur.fr/sites/default/files/veng\\_0.pdf](https://www.pasteur.fr/sites/default/files/veng_0.pdf) [accessed 28 July 2020].
22. **CLSI.** *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Eighth Informational Supplement. CLSI Document M100-S28.* Wayne, Pennsylvania, USA: Clinical and Laboratory Standards Institute; 2018.



23. Low AJ, Koziol AG, Manninger PA, Blais B, Carrillo CD. ConFindr: rapid detection of intraspecies and cross-species contamination in bacterial whole-genome sequence data. *PeerJ* 2019;7:e6995.
24. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
25. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol* 2018;19:153.
26. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
27. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.
28. Zhang S, den Bakker HC, Li S, Chen J, Dinsmore BA, et al. SeqSero2: rapid and improved *Salmonella* serotype determination using whole-genome sequencing data. *Appl Environ Microbiol* 2019;85:23.
29. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, et al. The *Salmonella* In Silico Typing Resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 2016;11:e0147101.
30. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. *Nat Commun* 2016;7:12827.
31. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The Enterobase user's guide, with case studies on salmonella transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 2020;30:138–152.
32. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, et al. Grape-Tree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* 2018;28:1395–1404.
33. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O, Friedrich A. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One* 2014;9:e104984.
34. Smith AM, Weill F-X, Njamkepo E, Ngomane HM, Ramalwa N, et al. Emergence of *Vibrio cholerae* O1 Sequence Type 75, South Africa, 2018–2020. *Emerg Infect Dis* 2021;27:2927–2931.
35. Smith AM, Tau NP, Smouse SL, Allam M, Ismail A, et al. Outbreak of *Listeria monocytogenes* in South Africa, 2017–2018: laboratory activities and experiences associated with whole-genome sequencing analysis of isolates. *Foodborne Pathog Dis* 2019;16:524–530.
36. Smith AM, Tau NP, Ngomane HM, Sekwadi P, Ramalwa N, et al. Whole-genome sequencing to investigate two concurrent outbreaks of *Salmonella* Enteritidis in South Africa, 2018. *J Med Microbiol* 2020;69:1303–1307.
37. NICD Communicable Diseases Communique; 2022. <https://www.nicd.ac.za/wp-content/uploads/2022/04/NICD-Monthly-Communique-April.pdf> [accessed 27 February 2023].
38. Tiba-Casas MR, Sacchi CT, Gonçalves CR, Almeida EA, Soares FB, et al. Molecular analysis of clonally related *Salmonella* Typhi recovered from epidemiologically unrelated cases of typhoid fever, Brazil. *Int J Infect Dis* 2019;81:191–195.
39. Dyson ZA, Holt KE. Five years of genotypyphi: updates to the global *Salmonella* Typhi genotyping framework. *J Infect Dis* 2021;224:S775–S780.
40. Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, et al. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill* 2017;22:30544.
41. Chattaway MA, Painset A, Godbole G, Gharbia S, Jenkins C. Evaluation of genomic typing methods in the *Salmonella* reference laboratory in Public Health, England, 2012–2020. *Pathogens* 2012;12:223.
42. Coipan CE, Dallman TJ, Brown D, Hartman H, van der Voort M, et al. Concordance of SNP- and allele-based typing workflows in the context of a large-scale international *Salmonella enteritidis* outbreak investigation. *Microb Genom* 2020;6:e000318.
43. Jagadeesan B, Gerner-Smidt P, Allard MW, Leuillet S, Winkler A, et al. The use of next generation sequencing for improving food safety: translation into practice. *Food Microbiol* 2019;79:96–115.
44. Gerner-Smidt P, Besser J, Concepción-Acevedo J, Folster JP, Huffman J, et al. Whole genome sequencing: bridging one-health surveillance of foodborne diseases. *Front Public Health* 2019;7:365.
45. Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, et al. Implementation of Nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin Infect Dis* 2016;63:380–386.
46. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat Genet* 2015;47:632–639.

### Five reasons to publish your next article with a Microbiology Society journal

1. When you submit to our journals, you are supporting Society activities for your community.
2. Experience a fair, transparent process and critical, constructive review.
3. If you are at a Publish and Read institution, you'll enjoy the benefits of Open Access across our journal portfolio.
4. Author feedback says our Editors are 'thorough and fair' and 'patient and caring'.
5. Increase your reach and impact and share your research more widely.

Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org).