OXFORD UNIVERSITY PRESS | microLife

Salmonella enterica serovar Typhimurium ST313 sublineage 2.2 has emerged in Malawi with a characteristic gene expression signature and a fitness advantage

Journal:	*microLife*
Manuscript ID	Draft
Manuscript Type:	Research article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kumwenda, Benjamin; Kamuzu University of Health Sciences, School of Life Sciences and Health Professions; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences; Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre Canals, Rocío; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Predeus, Alexander; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Zhu, Xiaojun; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Kroger, Carsten; Trinity College Dublin School of Genetics and Microbiology, Department of Microbiology; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Pulford, Caisey; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Wenner, Nicolas; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Lacharme-Lora, Lizeth ; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Li, Yan; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Owen, Siân; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences Everett, Dean; Khalifa University, College of Medicine and Health Sciences Hokamp, Karsten; Trinity College Dublin School of Genetics and Microbiology, Department of Genetics Heyderman, Robert; University College London, Research Department of Infection, Division of Infection & Immunity; Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre Ashton, Philip M.; Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre Gordon, Melita; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences; Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre Msefula, Chisomo; Kamuzu University of Health Sciences, School of Life Sciences and Health Professions; Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre

	Hinton, Jay; University of Liverpool, Institute of Infection, Veterinary & Ecological Sciences
Keywords:	transcriptomics, Salmonella, evolution, AMR, genomics, sublineage

Salmonella enterica serovar Typhimurium ST313 sublineage 2.2 has emerged in Malawi with a characteristic gene expression signature and a fitness advantage

Benjamin Kumwenda[1,2,3,#], Rocío Canals[2,#], Alexander V. Predeus[2,#], Xiaojun Zhu[2], Carsten Kröger[2,4], Caisey Pulford[2], Nicolas Wenner[2], Lizeth Lacharme Lora[2], Yan Li[2], Siân V. Owen[2], Dean Everett[5], Karsten Hokamp[6], Robert S. Heyderman[3,7], Philip M. Ashton[3], Melita A. Gordon[2,3], Chisomo L. Msefula[1,3], Jay C. D. Hinton[2*].

[1]School of Life Sciences and Allied Health Professions, Kamuzu University of Health Sciences Blantyre, Blantyre, MALAWI

[2]Institute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Liverpool, UNITED KINGDOM,

[3]Malawi–Liverpool–Wellcome Programme, Blantyre, MALAWI.

[4]Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin, IRE.

[5]College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, UAE.

[6]Smurfit Institute of Genetics, School of Genetics and Microbiology, Trinity College Dublin, Dublin, IRE.

[7]Research Department of Infection, Division of Infection & Immunity, University College London, UK.

Key words: transcriptomics, comparative genomics, lineage evolution, gene expression, antibiotic resistance

[#]BK, RC and AP contributed equally to this work.

* **Corresponding author:** E-mail: jay.hinton@liverpool.ac.uk; Tel: +441517954573; Institute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Liverpool, United Kingdom.

Abstract

Invasive non-typhoidal *Salmonella* (iNTS) disease is a serious bloodstream infection that targets immune-compromised individuals, and causes significant mortality in sub-Saharan Africa. *Salmonella enterica* serovar Typhimurium ST313 causes the majority of iNTS in Malawi. We performed an intensive comparative genomic analysis of 608 *S.* Typhimurium ST313 isolates dating between 1996 and 2018 from Blantyre, Malawi. We discovered that following the arrival of the well-characterised *S.* Typhimurium ST313 lineage 2 in 1999, two multidrug-resistant variants emerged in Malawi in 2006 and 2008, designated sublineage 2.2 and 2.3 respectively. The majority of *S.* Typhimurium isolates from human bloodstream infections in Malawi now belong to sublineage 2.2 or 2.3. To understand the emergence of the prevalent ST313 sublineage 2.2, we studied two representative strains, D23580 (lineage 2) and D37712 (sublineage 2.2). The chromosome of ST313 lineage 2 and sublineage 2.2 only differed by 29 SNPs/small indels and a 3kb deletion of a Gifsy-2 prophage region including the *sseI* pseudogene. Lineage 2 and sublineage 2.2 had distinctive plasmid profiles. The transcriptome was investigated in 15 infection-relevant *in vitro* conditions and within macrophages. During growth in physiological conditions that do not usually trigger *S.* Typhimurium SPI2 gene expression, the SPI2 genes of D37712 were transcriptionally active. We identified down-regulation of flagellar genes in D37712 compared with D23580. Following phenotypic confirmation of transcriptomic differences, we discovered that sublineage 2.2 had increased fitness compared with lineage 2 during mixed-growth in minimal media. We speculate that this competitive advantage is contributing to the emergence of sublineage 2.2 in Malawi.

Introduction

Non-typhoidal *Salmonella* (NTS) is a key bacterial pathogen that threatens people across the world. Typhimurium and Enteritidis are the two serovars of *Salmonella enterica* responsible for the highest levels of self-limiting gastrointestinal disease in Europe, the USA and other high-income countries (Zhang *et al.*, 2003). In the industrialised world, NTS has been associated with intensive food production, animal husbandry, and global distribution systems (Majowicz *et al.*, 2010). The *S.* Typhimurium sequence types responsible for gastroenteritis globally include ST19, ST34 and monophasic 1,4,[5],12:i:- variants (Branchu *et al.,* 2018). The diarrhoeal NTS disease is termed dNTS, and is mainly foodborne, posing a significant burden to public health with approximately 153 million cases and 57,000 deaths per annum worldwide (Kirk *et al.*, 2015; Chirwa *et al.*, 2023).

In contrast, a lethal systemic disease called invasive non-typhoidal Salmonellosis (iNTS) has emerged in recent decades in low- and middle-income countries in sub-Saharan Africa. iNTS targets immunocompromised individuals such as adults with HIV, and children under five years of age with malaria, malnutrition or severe anaemia (Feasey *et al.*, 2012). In some countries of sub-Saharan Africa, *Salmonella* causes more cases of community-onset bloodstream infections than any other bacterial pathogen (Marchello *et al.*, 2019). In 2017, 535,000 cases of iNTS disease were estimated worldwide, with about 80% of cases and 77,000 deaths occurring in sub-Saharan Africa (Stanaway *et al.*, 2019)

Clinically, the treatment of iNTS is complicated by multi-drug (MDR) resistance which limits therapeutic options (Crump *et al.*, 2015). Widespread resistance of iNTS pathogens to first-line drugs such as chloramphenicol, ampicillin and cotrimoxazole has been seen in many countries (Kariuki *et al.*, 2006). This MDR phenotype may be one of the reasons the case fatality rate associated with iNTS is amongst the highest in comparison to any infectious disease (15%) (Marchello *et al.,* 2022). Resistance to second-line drugs such as ceftriaxone, ciprofloxacin and azithromycin has been reported in a few African countries (Tack *et al.*, 2020). Clearly, the challenge posed by MDR *Salmonella* must be addressed urgently (Gilchrist and MacLennan, 2019).

The African iNTS epidemic is mainly caused by two *Salmonella* pathovariants, *S.* Typhimurium sequence type 313 (ST313) and specific clades of *S.* Enteritidis (Kingsley *et al.*, 2009; Okoro *et al.*, 2012; Feasey *et al.*, 2016). *S.* Typhimurium ST313 is responsible for about two-thirds of clinical iNTS cases that have been reported in Africa (Gilchrist and MacLennan, 2019).

It is not certain how these pathogens are transmitted, but there is increasing evidence from case-control studies that ST313 strains are human-associated but not animal-associated within households (Post *et al.*, 2019; Koolman *et al.*, 2022). A recent summary concludes that the available data are consistent with iNTS disease being transmitted person-to-person (Chirwa *et al.*, 2023). Global efforts to combat iNTS infections are currently focused on

vaccine development, which has now progressed to Phase 1 clinical trials (Piccini and Montomoli, 2020; Skidmore *et al.,* 2023).

Since 1998, continuous sentinel surveillance for fever and bloodstream infections among adults and children has been undertaken at Queen Elizabeth Central Hospital (QECH). This tertiary referral hospital in Blantyre, Malawi, serves an urban population of about 920,000 with a high incidence of malaria, HIV and malnutrition (Musicha *et al.*, 2017). Following blood-culture of samples collected from patients of all ages presenting with fever, whole genome sequencing identified the ST313 variant of *S.* Typhimurium (Kingsley *et al.*, 2009). Phylogenetic analysis revealed that the chloramphenicol-sensitive ST313 lineage 1 was clonally-replaced in Malawi by the chloramphenicol-resistant lineage 2 (Okoro *et al.*, 2012). More recently, a ST313 sublineage II.1 (2.1) emerged from lineage 2 in Democratic Republic of Congo (DRC) in Central Africa. Sublineage 2.1 had altered phenotypic properties including biofilm formation and metabolic capacity and resistance to azithromycin (Van Puyvelde *et al.*, 2019). An elegant genomic analysis that provides insight regarding the diversity of S. Typhimurium ST19 clades in the context of ST313 lineage 2 clades is also available (Van Puyvelde *et al*., 2023).

The initial suggestion that ST313 lineage 2 was undergoing evolutionary change in East Africa came from a small study that identified several *S.* Typhimurium ST313 Malawian isolates, dated between 2006 and 2008, that differed from lineage 2 by 22 core-genome single nucleotide polymorphisms (SNPs) (Msefula *et al.*, 2012).

To examine the evolutionary trajectory of *S.* Typhimurium in Malawi at a large scale, we conducted a comparative genomic analysis study focused on 680 isolates dating between 1996 and 2018 (Pulford *et al.*, 2021). We previously reported that ST313 lineage 1 (L1) was replaced by lineage 2 (here designated L2.0), and discovered an antibiotic-sensitive lineage 3 (L3) that emerged in 2016 (Pulford *et al.*, 2021). We have now performed a more intensive phylogenetic analysis of the same collection of *S.* Typhimurium ST313 isolates, most of which caused bloodstream infections in Malawi over two decades. We discovered two novel sublineages named 2.2 (L2.2) and 2.3 (L2.3) that emerged 2006 - 2008, and have been replacing L2.0.

Here we present a comprehensive comparative genomic analysis of the most prevalent ST313 L2.2 sublineage, and report the results of a functional genomic approach that identified key phenotypic characteristics that distinguish L2.2 from L2.0.

Results

Identification of *S.* Typhimurium ST313 sublineages 2.2 and 2.3 in Malawi

The emergence of the ST313 lineage 2 genotype in Malawi in 2002 prompted us to hypothesise that subsequent evolution would select for variants with increased fitness,

leading to the clonal expansion of one or more sublineages by outcompeting previously dominant genotypes. We investigated this hypothesis by conducting a detailed core-gene SNP-based maximum likelihood (ML) phylogenetic analysis to investigate the population structure of *S.* Typhimurium ST313 L2.0 (Fig. S1). As well as identifying members of the antibiotic-sensitive lineage 3, reported previously (Pulford *et al.*, 2021), we discovered that ST313 L2 comprised three phylogenetically-distinct sublineages that differed by a total of 39 SNPs. The *S.* Typhimurium ST313 reference strain D23580 (Kingsley *et al.*, 2009) belongs to the ST313 L2.0 lineage (Fig. 1A). As ST313 sublineage L2.1 had been defined previously (Van Puyvelde *et al.*, 2019), the new sublineages which belonged to different hierBAPS level 2 clusters were designated L2.2 and L2.3 (Fig. 1A and Fig. S1). In total, we identified 151 L2.2 isolates, 74 L2.3 isolates and 350 L2.0 isolates.

In Blantyre, Malawi, *S.* Typhimurium ST313 L2.2 was first detected in 2006, and L2.3 was initially observed in 2008 (Fig. 1BC). Both L2.2 and L2.3 increased in prevalence at the Queen Elizabeth Central Hospital in Blantyre in subsequent years. By 2018, L2.2 and L2.3 had largely replaced L2.0 (Fig. 1BC). Our published Bayesian (BEAST) analysis (Pulford *et al.*, 2021) estimated that the Most Recent Common Ancestor (MRCA) of ST313 lineage 2 dates back to 1948 (95% HPD = 1929-1959).

To understand the accessory gene complement of L2.2 and L2.3, we compared the genomes of seven L2.2 isolates and four L2.3 isolates with 17 L2.0 isolates, ST313 L1 and ST19 and the results are shown in Fig. 1A and Table S1. *S.* Typhimurium strain D23580 is the representative strain of L2.0 (Kingsley *et al.*, 2009), for which we previously used long-read sequencing and other approaches to thoroughly characterise the chromosomal and plasmid complement (Canals *et al.*, 2019b).

Antimicrobial Resistance

MDR variants of *S.* Typhimurium with resistance to ampicillin and cotrimoxazole were detected at an early stage of the iNTS epidemic, from 1997 onwards (Gordon *et al.,* 2008). Multidrug-resistant variants of *S.* Typhimurium ST313 that were no longer susceptible to chloramphenicol, ampicillin and cotrimoxazole subsequently emerged in Malawi (Gordon *et al.,* 2008) and have been reported elsewhere in sub-Saharan Africa by the GEMS study (Kasumba *et al.*, 2021). The *S.* Typhimurium ST313 L2.0, L2.2 and L2.3 isolates shared the same MDR profiles (resistance to chloramphenicol, ampicillin and cotrimoxazole), and carried identical IS21-associated antimicrobial gene cassettes within the pSLT-BT plasmid (Fig. 2B).

Comparative genomics of *S.* Typhimurium ST313 sublineage 2.2

Because *S.* Typhimurium ST313 L2.2 was the predominant novel sublineage in Blantyre,

Malawi in 2018, we focused on L2.2 for the remainder of this study. We used the phylogeny (Fig. 1A) to select strain D37712 as a representative of L2.2. D37712 was isolated from the blood of an HIV-positive Malawian male child and has been deposited in the National Collection of Type Cultures as NCTC 14678. The draft genome sequence of D37712 was obtained in 2012 with Illumina technology, an assembly that comprised 27 individual contigs (Msefula *et al.*, 2012). To generate a reference-quality genome, we resequenced D37712 with both long-read PacBio and Illumina short-read technologies. Our hybrid strategy generated a complete genome assembly that included one circular chromosome and three plasmids (see Materials & Methods; GenBank CP060165, CP060166, CP060167 and CP060168). This high-quality genome sequence allowed us to conduct a detailed comparison between the genomes of L2.2 strain D37712 and L2.0 strain D23580 (accession number FN424405), summarised in Fig. 2 and Table S2.

Overall, the gene content of the two strains was largely equivalent. The D23580 annotation contains 4,823 protein-coding and pseudogenes and 287 small RNA (sRNA) genes that we identified previously (Canals *et al.,* 2019b), while D37712 contains 4,821 protein-coding and pseudogenes and the same 287 sRNAs. In total, the D37712 and D23580 genomes shared 4,729 orthologous protein-coding genes and pseudogenes. The 104 protein genes that differ are encoded by the pSLT ^{D37712}, pBT1^{D37712}, and pCol1B9^{D37712} plasmids.

Overview of D23580 and D37712 genomes

The chromosomes of D23580 and D37712 are 4,879,402 and 4,876,060 bp, respectively, about the same size as other S. Typhimurium genomes (Kingsley *et al.*, 2009; Branchu *et al.*, 2018). The D23580 and D37712 strains share an identical prophage profile, with both strains carrying five prophages (BTP1, Gifsy-2, ST64B, Gifsy-1, and BTP5) located at the same positions on the chromosome (Fig. 2A).

Comparison of D23580 and D37712 chromosomes

The detailed genomic comparison of D37712 with D23580 showed that the sizes of the two chromosomes varied by only 3,342 bp. Overall, the only differences between the genomes of the L2.0 and L2.2 strains were 26 chromosomal SNPs and small indels, plus one large deletion, and an inversion of the *hin* switch. In-depth annotation of the nucleotide variants identified 3 putative loss-of-function mutations (2 stop mutations, 1 frameshift insertion), 1 disruptive in-frame deletion, 4 synonymous mutations, 13 missense mutations, and 5 intergenic variants, summarised in Fig. 2A. None of the SNP differences that distinguished D37712 from D23580 were located within 150 nucleotides of a Transcriptional Start Site (Canals *et al.,* 2019b), and so would not be predicted to modulate gene expression.

The 3,358 bp-long deletion of a Gifsy-2 prophage-associated region that spanned the *sseI* pseudogene of D23580 (STMMW_10631) removed two coding sequences (STM1050-51; STMMW_10611-STMMW_10631), and substantially truncated the STM1049 (STMMW_10601) gene (Fig. 2E). The *sseI* gene encodes a cysteine hydrolase effector

protein that modulates the directional migration of dendritic cells during systemic infection (Brink *et al.*, 2018). In strain D23580, the insertion of an IS26 transposable element inactivated the *sseI* gene (Kingsley *et al.*, 2009), causing increased dendritic cell-mediated dissemination of strain D23580 during infection (Carden *et al.*, 2017). We used an independent PCR-based approach to confirm that the 3,358 bp deletion had removed the *sseI* gene from the chromosome of strain D37712 (Fig. S2).

Comparison of D23580 and D37712 plasmids

Here we put the genetic features of the representative strains for ST313 L2.0 and L2.2 into context with other isolates belonging to the Lineage 2 sublineages. ST313 L2.0 strain D23580 carries four plasmids, pSLT-BT, pBT1, pBT2 and pBT3 (Kingsley *et al.*, 2009). In contrast, ST313 L2.2 has a distinct plasmid complement (Fig. 1A, Fig. 2BCD). The plasmid profiles of D23580 and D37712 were confirmed by a combination of Illumina (short-read) and PacBio (long-read) sequencing (Materials and Methods).

In summary, strain D37712 carried the pSLT-BT, pBT2 and pCol1B9 plasmids as detailed below. Both D23580 and D37712 strains carried a variant of the pSLT-BT virulence plasmid (Kingsley *et al.*, 2009) that contains a Tn21-like transposable element with five antibiotic resistance genes. The D37712 version of pSLT-BT only differs from the pSLT-BT of D23580 in two important t ways (Fig. 2B). Firstly, the Tn21-like element is inserted in the opposite direction with regards to the rest of the plasmid, suggesting that the transposable element remains active. Secondly, three nucleotide variants were identified in the pSLT-BT carried by D37712, two deletions in noncoding regions, and one frameshift insertion that generates a pseudogene of *spvD.*

Plasmid pCol1B9 was of particular interest because it was absent from D23580, but was present in *S.* Typhimurium ST19 strain 4/74 (Richardson *et al.*, 2011; Fig. 1A). 4/74 is the parent of *S.* Typhimurium SL1344, a strain that has been used extensively for the study of *S.* Typhimurium pathogenesis and gene regulation in recent decades (Kröger *et al.*, 2012; Rankin & Taylor, 1966). Our new annotation of the pCol1B9-like plasmid identified 95 distinct protein-coding genes, while the previously published annotation of pCol1B9$^{4/74}$ assigned 101 protein-coding genes. Some of these represent annotation discrepancies, while others represent true genetic differences (Fig. S3).

Following careful examination, we identified 14 genes unique to pCol1B9^{D37712}, and 20 genes unique to pCol1B9$^{4/74}$. There were 81 genes carried by both plasmids. Interestingly, pCol1B9^{D37712} lacked the colicin toxin-antitoxin system that both gave pCol1B9 its name, and provides *Salmonella* with a competitive advantage in the gut (Nedialkova *et al.*, 2014). The pCol1B9^{D37712} plasmid carried a locus that was absent from pCol1B9$^{4/74}$, namely the *impC-umuCD* operon (Fig. S3) which encodes the error-prone DNA polymerase V responsible for the increased mutation rate linked to the SOS stress response in *E. coli* (Sikand *et al.*, 2021).

Comparison of pseudogene status of D23580 and D37712

Our comparative genomic analysis focused on the pseudogenes found in strains 4/74, D23580, and D37712 (Fig. 2F, Table S3). The pseudogenisation of several D23580 genes, compared with strain 4/74, have been linked to the invasive phenotype of African *Salmonella* ST313 (Kingsley *et al.*, 2009). We found that the pseudogene complement of D23580 was largely conserved in D37712, consistent with inheritance from a common ancestor. We have recently reported the role of the MacAB-TolC macrolide efflux pump in the virulence of *S.* Typhimurium ST313, and showed experimentally that *macB* was an inactive pseudogene in D23580 (Honeycutt *et al.*, 2020). Interestingly, the *macB* gene is functional in D37712. Compared with D23580, three additional D37712 genes were pseudogenised (*spvD, yadE,* and STMMW_42692, as detailed in Table S3). YadE is a predicted polysaccharide deacetylase lipoprotein. The functional impact of these pseudogenes on L2.2 remains to be established.

Overall the chromosomes of ST313 lineage 2 and sublineage 2.2 were highly-conserved and differed by just 29 SNPs/ small indels, and a 3kb deletion in the Gifsy-2 prophage region. The ST313 lineage 2 and sublineage 2.2 have distinct plasmid profiles.

Transcriptional landscape of S. Typhimurium ST313 sublineage L2.2

Previously, we characterized the primary transcriptome of two other *S.* Typhimurium strains, 4/74 and D23580, using a combination of multi-condition RNA-seq and differential RNA-seq (dRNA-seq) techniques (Canals *et al*., 2019b; Kröger *et al.*, 2013). To identify the transcriptional start sites (TSS) of strain D37712, we analysed a pooled sample containing RNA from 15 *in vitro* conditions by dRNA-seq and RNA-seq as detailed previously (Kröger *et al.*, 2013). The high similarity between the D23580 and D37712 chromosomes allowed us to map the curated set of TSS that were previously defined for D23580 (Hammarlöf *et al.,* 2018) onto a combined D37712/D23580 reference genome. To allow individual TSS to be examined in particular chromosomal or plasmid regions, data from both the dRNA-seq and pooled RNA-seq experiments can be visualised in our online genome browser (**http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data**).

Preliminary gene expression profiling of S. Typhimurium ST313 sublineage L2.2

Given the high level of similarity between the genomes of L2.2 and L2.0, we went on to identify differences at the transcriptional level. We performed a multi-condition RNA-seq-based transcriptomic analysis of gene expression profiles of L2.2 strain D37712 without biological replicates.

This comparative transcriptomic screen was based on our published approach (Canals *et al*., 2019b). Specifically, we used 15 individual infection-relevant *in vitro* conditions (Kröger *et al.*, 2013) and did intra-macrophage transcriptome profiling using the protocol previously established for *S.* Typhimurium ST19 (Srikumar *et al.*, 2015). The RNA-seq samples were mapped to a combined reference genome, which included the annotated D23580 chromosome (Canals *et al*., 2019b), as well as all the plasmids described earlier (pSLT-BT,

pBT1, pBT3 and pCol1B9; see Methods). The initial RNA-seq assessment (detailed in Methods) involved 2-4M non-rRNA/tRNA reads per sample, allowing gene signatures specific for each *in vitro* condition to be identified. Although single replicate RNA-seq experiments of this type cannot be used for statistically-robust differential gene expression analysis, they do provide a useful screening approach for identifying growth conditions to be used for follow-up experiments. The individual RNA-seq experiments showed broad condition-specific similarities in gene expression between strains 4/74, D37712, and D23580 (Fig. 3A). The gene expression values from each profiled condition are available as raw counts and TPMs in Tables S4 and S5.

To select the ideal environmental conditions to use for subsequent experiments, we assessed the expression profiles of known *Salmonella* pathogenicity islands which were broadly similar in strains D37712, and D23580. Although the expression profile of the SPI2 pathogenicity island was broadly similar between D37712, D23580 and 4/74 in most growth conditions, the SPI2 genes of D37712 were highly up-regulated in a single growth condition, NonSPI2 (Fig. 3B-C). NonSPI2 is a minimal medium with a neutral pH and a relatively high level of phosphate, in which *S.* Typhimurium does not usually express the SPI2 pathogenicity island (Löber *et al.*, 2006; Kröger *et al.*, 2013). This intriguing observation prompted us to perform a more discriminating set of transcriptomic experiments, as described below.

Differential gene expression analysis of *S.* Typhimurium D37712 versus D23580 in four *in vitro* conditions with multiple biological replicates

To define the transcriptional signature of strain D37712 more accurately, we generated RNA-seq data from D37712 grown in four *in vitro* conditions that stimulate expression of the majority of virulence genes: ESP, anaerobic growth, NonSPI2 and InSPI2, with multiple (3-4) biological replicates. The combination of acidity (pH 5.8) and low phosphate (0.4 mM Pi) in the InSPI2 media stimulates transcription of SPI2 genes in *S.* Typhimurium (Löber *et al.*, 2006; Kröger *et al.*, 2013). The NonSPI2 condition is based on the same PCN media recipe as InSPI2 media, but is neutral (pH 7.4), and contains higher levels of phosphate (25 mM Pi) (Löber *et al.*, 2006; Kröger *et al.*, 2013).

We compared the results with our published transcriptomic data for S. Typhimurium strains 4/74 and D23580 (Canals *et al.,* 2019b; Kröger *et al.*, 2013). Differential expression analysis with DEseq2, with conservative cut-offs (fold change ≥ 2, FDR ≤0.001), showed that the gene expression profiles of D37712 and D23580 were broadly similar, and shared key differences to the transcriptional profile of strain 4/74 under each of the four *in vitro* conditions (Fig. 4A). The differential expression results are summarized in Table S6.

We specifically investigated transcription of the *pgtE* gene, which encodes the outer-membrane protease previously linked to the ability of African *Salmonella* ST313 to resist human serum killing (Hammarlöf *et al.*, 2018). Compared to 4/74, the *pgtE* gene of both the

D23580 and D37712 strains showed a similar pattern of up-regulation by a factor of 7 to 18 across all conditions. This finding is consistent with the fact that D37712 carries the same T nucleotide in the -10 region of the *pgtE* promoter that is responsible for increased expression of the *pgtE* transcript in strain D23580 (Hammarlöf *et al.*, 2018).

There were no statistically-significant changes in expression of the majority (92%) of the 4,729 orthologous coding genes shared by D37712 and D23580. We identified a total of 364 genes that were differentially expressed in at least one growth condition between D37712 and D23580 as follows: ESP (69 differentially-expressed genes), anaerobic growth (214 differentially-expressed genes), NonSPI2 (88 differentially-expressed genes) and InSPI2 (17 differentially-expressed genes; Fig. 4B).

Overall, the differentially-expressed genes that distinguished D37712 from D23580 only showed expression differences in a single growth condition rather than across all conditions. The differentially expressed genes included flagellar genes (down-regulated), SPI2-associated genes(up-regulated), and genes involved in general and anaerobic metabolism (down-regulated).

SPI2 pathogenicity island genes play a key role in the intracellular replication of *S.* Typhimurium, and encode the type III secretion system that is responsible for translocation of key effector proteins into mammalian cells (Jennings *et al.*, 2017). The RNA-seq data showed that SPI2 genes were expressed at similarly high levels in both D37712 and D23580 strains following induction (InSPI2 media; Fig. 4B), and confirmed that the key SPI2 expression difference was only seen in strain D37712 under non-inducing growth conditions (NonSPI2 media). It is important to put this differential SPI2 expression into context. D37712 expresses SPI2 genes at about a 10-fold higher level than D23580 during growth in non-inducing NonSPI2 media, but the actual level of expression was 20-fold less than the level stimulated by growth in SPI2-inducing conditions (InSPI2 medium).

The up-regulation of *fljA* and *fljB* and the down-regulation of *fliC* in D37712, compared to D23580 in all four growth conditions likely reflects the opposite orientation of the *hin* switch in the D37712 genome compared to D23580. This type of *hin* inversion occurs frequently in *S.* Typhimurium (Johnson and Simon, 1985).

Another gene that was up-regulated in D37712 across all profiled conditions was the chromosomally-encoded $cysS^{chr}$, that encodes cysteine-tRNA synthetase. Previously, we reported that transcription of the $cysS^{chr}$ of strain D23580 was uniformly down-regulated compared to 4/74. This down-regulation was compensated by the presence of a pBT1 plasmid-encoded cysteine-tRNA synthetase (Canals *et al.,* 2019a). Accordingly, the increased expression of the chromosomal *cysS* gene in D37712 was consistent with the absence of the pBT1 plasmid. Our comparative transcriptomic analysis showed that expression levels of *cysS* were similar in D37712 and 4/74 under all growth conditions.

Numerous virulence genes and operons were differentially expressed between D23580 and D37712. The SPI-16-associated *gtrABCa* operon (STM0557, STM0558, STM0559) is responsible for adding glucose residues to the O-antigen subunits of LPS that enhance the long-term colonisation of the mammalian gastrointestinal tract by *S.* Typhimurium ST19 (Bogomolnaya *et al.*, 2008). We found that the *gtrABCa* genes were significantly up-regulated in several conditions in D37712, compared to both D23580 and 4/74.

The *spvABCD* operon of D37712 was up-regulated under non-SPI2-inducing growth conditions, compared to D23580. A signature pseudogene of ST313 L2.2 is the frameshift insertion in the *spvD* gene that generates a truncated version of the SpvD protein. The H199I mutation at position 199 and the associated 17 amino acid truncation is predicted to ablate the activity of the SpvD cysteine protease (Grabe *et al.*, 2016). SpvD negatively regulates the NF-B signaling pathway and promotes virulence of S. Typhimurium in mice. The functional consequences of the *spvD* variant of ST313 L2.2 strain D37712 and the up-regulation of expression of the *spvABCD* operon remain to be established experimentally.

The SalComD37712 community transcriptional data resource

To allow scientists to gain their own biological insights from analysis of this rich transcriptomic dataset, the transcriptomic and gene expression data generated in this study are presented online in a new community resource, SalComD37712. The data resource shows the expression levels of all D37712 coding and non-coding genes, including both chromosomal and plasmid-encoded transcripts. The SalComD37712 website complements our existing SalComD23580 (https://tinyurl.com/SalComD23580) resource, and adds an inter-strain comparison of gene expression profiles between D37712 and D23580 as well as normalized gene expression values (TPM), using an intuitive heat map-based approach. SalComD37712 included our published RNA-seq data (Canals *et al.*, 2019b), re-analysed with an updated bioinformatic pipeline and a combined reference genome (see Methods). This online resource facilitates the intuitive interrogation of transcriptomic data as described previously (Perez-Sepulveda and Hinton, 2018).

Additionally, we generated a unified genome-level browser that provides access to the *S.* Typhimurium L2.2 D37712 transcriptome, in the context of our previously published RNA-seq data for the L2.0 strain D23580 and the ST19 strain 4/74. This novel “combo” browser is available at http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data.

Identification of phenotypes that distinguish ST313 sublineage L2.2 from L2.0.

To explore the phenotypic impact of the transcriptomic signature of L2.2 (D37712), we performed a series of motility experiments, fluorescence-based gene expression experiments and mixed-growth assays.

D33712 showed a significantly decreased level of motility on NonSPI2 minimal media, compared with both the ST19 strain 4/74 and the L2 D23580 strain (Fig. 5A). This finding was consistent with the transcriptomic data, which showed down-regulation of D37712 flagellar genes compared with D23580 in the NonSPI2 condition (Fig. 4). In contrast, no differential expression of flagellar genes was seen between D33712 and D23580 in the InSPI2 growth condition (Fig. 4). The decreased motility phenotype may be linked to the inversion of the *hin* element detailed above. The flagella system encodes a distinct type III secretion apparatus responsible for the dual functions of bacterial motility and activation of the mammalian innate immune system via TLR5 (Lai *et al.*, 2013).

A key transcriptomic finding for strain D33712 was the expression of SPI2 genes during growth in an unusual environmental condition (NonSPI2) (Fig. 3B-C and Fig. 4B). NonSPI2 media differs from InSPI2 media by having a higher pH (pH7.4 versus pH5.8) and a higher level of phosphate (Löber *et al.*, 2006). This apparent differential expression of SPI2 genes at the transcriptomic level under non-inducing conditions led us to investigate the expression of SPI2 at a single cell level using fluorescence transcriptional fusions. First, we introduced an *ssaG*-GFP$^+$ transcriptional fusion into the chromosome of strains D33712 and D23580 (Methods; Table S8) to interrogate expression of the key SPI2 operon with flow cytometry. Figure 5B shows that in NonSPI2 media, the *ssaG* promoter was expressed at a 62% higher level in D33712 than in D23580 confirming the results of the transcriptomic analysis.

Because only a proportion of *S.* Typhimurium cells express certain pathogenicity island-encoded genes during *in vitro* growth (Ackermann *et al.*, 2008; Hautefort *et al.*, 2003), we determined whether the increased level of expression of SPI2 genes (Fig. 4B) was caused by a higher proportion of D33712 cells expressing SPI2 than D23580 cells. Using derivatives of the two strains that carried the *ssaG*-GFP$^+$ construct, we determined the numbers of fluorescent and non-fluorescent cells with flow cytometry (Methods). Under non-inducing conditions, slightly more D37712 cells expressed the *ssaG* SPI2 promoter than D23580 cells (65% vs 60%, respectively) (Fig. 5C). Although this small difference was statistically significant (t-test: $P<0.001$, $n=3$), it did not account for the 62% increased level of non-induced SPI2 expression seen in Fig. 5B.

SPI2 expression is controlled by a complex regulatory system that operates at both a negative and positive level, involving silencing via H-NS (Lucchini *et al.*, 2006), activation by SlyA and SsrB (Fass and Groisman, 2009; Walthers *et al.*, 2011) as well as input from OmpR and Fis under non-inducing conditions (Osborne and Coombes, 2011). The mechanistic basis of the aberrant SPI2 expression in strain D37712 is worthy of further study. Possible explanations include the incomplete silencing of SPI2 transcription or the partial activation of

the SPI2 virulence genes under non-inducing growth conditions by an unknown regulatory factor.

Increased fitness of *S.* Typhimurium ST313 sublineage L2.2 compared with L2.0 in minimal media.

It has become increasingly clear that distinct *Salmonella* pathovariants have evolved particular phenotypic properties that confer fitness advantages during infection of particular avian or mammalian hosts (Branchu *et al.*, 2018). Because *S.* Typhimurium ST313 L2.2 appeared to have displaced *S.* Typhimurium ST313 L2.0 in Malawi, we speculated that *S.* Typhimurium ST313 L2.2 might have the competitive edge in some situations. Accordingly, we determined bacterial fitness using a mixed-growth competition assay (Wiser and Lenski, 2015; Lian *et al.*, 2023). The competitive index was calculated in three different growth media using pair-wise combinations of strains D37712 and D23580. Two independent approaches were used to phenotypically distinguish the two strains, one based on antibiotic resistance (Fig. 5D) and the other based on fluorescent tagging (Fig. S5).

To confirm that strains engineered to be kanamycin-resistant or gentamicin-resistant did not impact on fitness (Methods), we first verified that the tagged variants of D37712 or D23580 did not confer a growth advantage in LB or NonSPI2 media (Fig. S7). Next, we used a mixed-growth assay to investigate fitness of *S.* Typhimurium ST313 L2.0 strain D23580 or *S.* Typhimurium ST313 L2.2 strain D37712 during growth in LB, or InSPI2 or NonSPI2 minimal media. The data show that both strains grew at similar levels following overnight mixed-growth in nutrient-rich LB media, but D37712 had a competitive advantage during mixed-growth in InSPI2 media (CI = 1.79; $P<0.05$) and a greater competitive edge in NonSPI2 media (CI = 2.20; $P<0.0001$).

We then used an independent fluorescence-based approach to assess the fitness of strains D23580 and D37712 during mixed-growth in NonSPI2 media. This time, the strains were engineered to carry either mScarlet or sGFP2 proteins and the mixed-growth experiments involved pair-wise comparisons of reciprocally-tagged strains. The flow cytometric data showed that in both cases D37712 had a significant competitive advantage in NonSPI2 media (Fig. S5 and S6).

This combination of antibiotic resistance-based and fluorescence-based competitive index experiments lead us to conclude that *S.* Typhimurium ST313 L2.2 strain D37712 had a clear fitness advantage over *S.* Typhimurium ST313 L2.0 strain D23580 during mixed-growth in two formulations of minimal media. The molecular basis of this fitness advantage remains to be established.

Perspective

Here, we report that *S.* Typhimurium ST313 L2.0 has been clonally replaced by the ST313 sublineages L2.2 and L2.3 as a cause of bloodstream infection in Blantyre, Malawi. In 2018, L2.2 represented the majority of the ST313 strains isolated from hospitalised patients in Malawi at the Queen Elizabeth Central Hospital. Our comparative genomic analysis of ST313 L2.3 identified 30 chromosomal alterations, one of which generated a deletion of the *sseI* effector gene.

Our RNA-seq-based analysis of ST313 L2.2 involved a detailed comparison versus ST313 L2.0 which revealed a key difference involving SPI2 expression. Following initially observations at the transcriptomic level in the ST313 L2 and L2.2 strains grown in a pH-neutral minimal medium (NonSPI2), the increased expression of SPI2 was confirmed at the single cell level using an *ssaG* transcriptional fusion.

A series of experiments showed that the ST313 L2.2 strain D37712 had a competitive advantage over L2 strain D23580 during mixed-growth in minimal media. We propose that this increased fitness of *S.* Typhimurium ST313 L2.2 has contributed to the replacement of ST313 L2.0 in Malawi in recent years.

Previously, we compared three virulence properties of the *S*. Typhimurium ST313 L2.0 D23580 and ST313 L2.2 D37712 strains. First, experiments involving Mucosal Invariant T (MAIT) cells showed that both D37712 and D23580 fail to elicit the high level of activation of MAIT cells that characterises infection by *S*. Typhimurium ST19 4/74 (Preciado-Llanes *et al.*, 2020). Second, the D37712 and D23580 strains stimulate similar levels of up-regulation of IL10 gene expression upon infection of human dendritic cells (Aulicino *et al.*, 2022). Third, we showed that both D37712 and D23580 express similarly high levels of the PgtE virulence factor that is responsible for the ability of *S*. Typhimurium ST313 to survive human serum-killing (Hammarlöf *et al.*, 2018). These findings lead us to conclude that the comparative genomic and transcriptomic differences that distinguish *S*. Typhimurium ST313 L2.0 strain D23580 from ST313 L2.2 D37712 (Fig. 4) do not modulate the ability of the pathogens to activate human MAIT cells or dendritic cells, or to influence the PgtE-mediated serum survival phenotype of *S.* Typhimurium ST313.

Ideally, the implications of the competitive advantage of ST313 L2.2 would be determined in the context of pathogenesis. However, we lack an informative infection model for *S.* Typhimurium ST313 (Lacharme-Lora *et al.,* 2019), and it is not yet possible to experimentally determine whether the improved fitness of L2.2 significantly enhances the success of ST313 during infection of humans.

Here we have investigated the intricate interplay of gene function that underpins the success of *S*. Typhimurium ST313 L2.2. Our hope is that our findings could contribute to future therapeutic or prophylactic strategies for combatting iNTS infections in the African setting.

Materials and methods

Bacterial strains

To investigate the evolutionary dynamics of *S.* Typhimurium ST313 L2 in Malawi over a 22 year period, we focused on the large collection of 8,000 *S.* Typhimurium isolates derived from bloodstream infection in hospitalised patients at the Queen Elizabeth Central Hospital, Blantyre, Malawi (Feasey *et al.,* 2015). The collection was assembled by the Malawi–Liverpool–Wellcome Trust Clinical Research Programme (MLW) between 1996 and 2018; the precise annual numbers of isolates are shown in Fig. 1C. A random sub-sampling strategy was used to select 608 isolates for whole-genome sequencing, which included 549 *S.* Typhimurium ST313 isolates (Pulford *et al.*, 2021).

The two *S.* Typhimurium ST313 strains that are the focus of this study are D23580 and D37712. D23580 was isolated from a Malawian 26-month-old child with malaria and anaemia in 2004. D37712 was isolated from the blood of an HIV-positive Malawian male child in 2006. These two African Salmonella strains have been deposited in the National Collection of Type Cultures (NCTC). The D23580 (lineage 2.0) strain is available as NCTC 14677. The ST313 sublineage 2.2 strain D37712 is available as NCTC 14678. All bacterial strains are detailed in Table S8.

Genome sequencing

The assembled genome and annotation of D23580 (Kingsley *et al.*, 2009; Canals *et al.*, 2019b) (L2.0) was obtained from the European Nucleotide Archive (ENA) repository (EMBL-EBI) under accession PRJEB28511 (https://www.ebi.ac.uk/ena/data/view/PRJEB28511). For genome sequencing of D37712 (L2.2), DNA was extracted using the Bioline mini kit, and quality was assessed using gel electrophoresis (0.5% agarose gel, at 30 volts for 18 h). The genome was generated by a combination of long read sequencing with a PacBio RS II and short-read sequening on an Illumina HiSeq machine at the Center for Genome Research, University of Liverpool, United Kingdom.

Sequence reads were quality checked using FastQC version 0.11.9 (Andrews, 2010) and MultiQC version 1.8 (Ewels *et al.*, 2016), trimmed using Trimmomatic (Bolger *et al.*, 2014). Hybrid assembly of the Illumina and PacBio sequence reads was done with Unicycler v0.4.7 (Wick *et al.*, 2017).

The assembled genome of S. Typhimurium SDT313 L2.2 strain D37712 was deposited in Genbank (GCA_014250335.1, assembly ASM1425033v1). Raw sequencing reads were deposited for both PacBio and Illumina, under BioProject ID PRJNA656698. Sequence Read Archive (SRA) database IDs are: SRR12444880 for Illumina and SRR12444881 for PacBio.

Comparative genomic analyses

To generate the data summarised in Fig. 1C, sequencing data of 29 *S*. Typhimurium ST313 strains (Msefula *et al.*, 2012) were downloaded from EMBL-EBI database (https://www.ebi.ac.uk, accession number ERA015722). Sequence reads were assembled using Unicycler v0.4.8 (Wick *et al.*, 2017). The quality of the assemblies was assessed by Quast v5.0.2 (Gurevich *et al.*, 2013). The N50 value of all assemblies was >20kb, and the number of contigs was <600.

To construct the phylogenetic tree (Fig. 1C), *Salmonella* Typhimurium strains D23580, D37712, LT2 (GCA_000006945.2), DT104 (GCA_000493675.1), 4/74 (GCA_000188735.1), and A130 (GCA_902500285.1) were added as contextual genomes. Roary was used to make the core gene alignment, construct the gene presence/absence matrix and identify orthologous genes (Page *et al.*, 2015). Phylogenetic trees were constructed using Randomized Accelerated Maximum Likelihood (RAxML) (Stamatakis *et al.*, 2005), and were visualised with the interactive Tree of Life online tool (iToL) (Letunic and Bork, 2006).

The assembled genome and annotation of *S*. Typhimurium ST19 representative strain 4/74 (Richardson *et al.*, 2011) were obtained from GenBank (Accession number GCF_000188735.1), while the raw sequencing data of 27 *S*. Typhimurium ST313 strains described in a previous study (Msefula *et al.*, 2012) were downloaded from EMBL-EBI database (https://www.ebi.ac.uk, accession number ERA015722). The raw reads were assembled using Unicycler v0.4.8 (Wick *et al.*, 2017). The quality of the assemblies was assessed by Quast v5.0.2 (Gurevich *et al.*, 2013). The N50 value of all assemblies was >20kb, and the number of contigs was <600.

To identify SNPs, Snippy v4.4.0 (https://github.com/tseemann/snippy) was used to map the raw reads against the 4/74 genome. To detect pseudogene-associated SNPs/indels in each sub-lineage, the SNPs/indels that caused nonsense or frameshifted mutations were filtered. The identifications and names of the disrupted genes were summarised, then the wild type gene sequences were extracted from the 4/74 genome. To validate the pseudogene-associated SNPs/indels, the wild type gene sequences were used to make a BLAST database with BLAST 2.9.0+ (Camacho *et al.*, 2009). The 29 genome assemblies were queried against the databases, using the BLASTn algorithm to confirm the nonsense and frameshifted mutations in all isolates.

Phylogenetic analysis of African *Salmonella* Typhimurium isolates dating from 1966 - 2018

To examine the overall population structure of *Salmonella* Typhimurium responsible for blood infection in Malawi (Fig. 1AB and Fig. S1), the raw reads of 707 published genome sequences were downloaded (Table S7). Trimmomatic v0.36 (Bolger, A. M., Lohse, 2014) was used to trim adapters and Seqtk v1.2-r94 (https://github.com/lh3/seqtk) was used to trim low-quality regions using the trimfq flag. Fastqc v0.11.5 (https://www. bioinformatics.babraham.ac.uk/projects/fastqc/) and multiqc v1.0 (http://multiqc. info) were used to pass sequence reads according to the following criteria: passed basic quality

statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 47% and 57%. Only high-quality reads were used in the downstream analysis. Sequence reads were aligned to the *S*. Typhimurium D23580 genome using Snippy v4.4.0 with parameter "- - mincov 5". The recombination sites of the alignment were removed by Gubbins (Croucher *et al.,* 2015), and the phylogenetic tree was built with Raxml-ng (Kozlov et al., 2019) using GTR_G models ad 100 bootstraps. The tree was rooted on *Salmonella* Typhi strain CT18 (GCA_000195995.1) as the outgroup. The tree was visualised with the interactive Tree of Life online tool (iToL) (Letunic and Bork, 2006). The sub-lineages were identified with rHierBAPS (Tonkin-Hill *et al.,* 2018). The stacked-area chart and the bar chart showing the percentage and number of isolates from each sub-lineage were made in MS Excel.

RNA purification and growth conditions

Initially, a screen of transcriptomic gene expression was performed without biological replicates. Total RNA was purified using TRIzol from *S*. Typhimurium D37712 grown in 15 different conditions as described previously (Kröger *et al.*, 2013). To generate statistically-robust gene expression profiles, total RNA was subsequently purified using TRIzol from *S*. Typhimurium D37712 grown in four *in vitro* growth conditions (ESP, anaerobic growth, NonSPI2, InSPI2) with three biological replicates as described previously (Kröger *et al.*, 2013). RNA was isolated from intra-macrophage D37712 following infection of RAW264.7 murine macrophages using our published protocol (Srikumar *et al.*, 2015).

RNA-seq of *S*. Typhimurium strain D37712 using Illumina technology

For transcriptomic analyses, cDNA samples were prepared from *S*. Typhimurium RNA by Vertis Biotechnologie AG (Freising, Germany). RNA was first treated with DNase and purified using the Agencourt RNAClean XP kit (Beckman Coulter Genomics). RNA samples were sheared using ultrasound, treated with antarctic phosphatase and re-phosphorylated with T4 polynucleotide kinase. RNA fragments were poly(A)-tailed using poly(A) polymerase and an RNA adapter was ligated to the 5'- phosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV reverse transcriptase. The resulting cDNA was PCR-amplified to about 10-20 ng/µl. The cDNA was purified using the Agencourt AMPure XP kit. The cDNA samples were pooled using equimolar amounts and size fractionated in the size range of 200-500 bp using preparative agarose gels. The cDNA pool was sequenced on an Illumina NextSeq 500 system using 75 bp read length.

For the biological replicates of the four growth conditions (ESP, anaerobic growth (abbreviated as NoO2), NonSPI2, and InSPI2) and the intra-macrophage RNA, cDNA samples were generated as above with some improvements in library preparation. First, after fragmentation with ultrasound, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. Second, first-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer, and, after purification, the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. Sequencing of the

cDNA was performed as described above. All raw sequencing reads were deposited to the Gene Expression Omnibus (GEO) database under accession GSE161403.

RNA-seq and dRNA-seq read processing and visualization

RNA-seq data from *S*. Typhimurium 4/74 and D23580 were extracted from previously published experiments (Kröger *et al.*, 2013; Srikumar *et al.*, 2015; Canals *et al.*, 2019b; GEO dataset GSE119724). A combined reference genome was generated that contained the D23580 chromosome plus plasmids pBT1, pBT2, pBT3, pSLT-BT (from D23580) and the D37712 plasmid pCol1B9^{D37712}. All reads were aligned and quantified using Bacpipe v0.8a (https://github.com/apredeus/multi-bacpipe). Briefly, basic read quality control was performed with FastQC v0.11.8. RNA-seq reads were aligned to the genome sequence using STAR v2.6.0c using "--alignIntronMin 20 --alignIntronMax 19 --outFilterMultimapNmax 20" options. A combined GFF file was generated by Bacpipe, where all features of interest were listed as a "gene", with each gene identified by a D37712 locus tag. Subsequently, read counting was done by featureCounts v1.6.4, using options "-O -M --fraction -t gene -g ID -s 1". For visualization, scaled gedGraph files were generated using bedtools genomecov with a scaling coefficient of 10^9/(number of aligned bases), separately for sense and antisense DNA strands. Bedgraph files were converted to bigWig using bedGraphToBigWig utility (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/). Coverage tracks, annotation, and genome sequence were visualized using JBrowse v1.16.6. Transcripts Per Million (TPM) were calculated for all samples and used as absolute expression values (Table S5). A conservative cut-off was used to distinguish between expressed (TPM >10) and not expressed (TPM ≤10), as we previously described (Kröger *et al.*, 2013). Relative expression values were calculated by dividing the TPM value for one condition in one strain by the TPM value for the same condition in a different strain. Before the calculation, all TPM values below 10 were set up to 10. A conservative fold-change cut-off of 3 was used to highlight differences in expression between strains.

Differential gene expression analysis with multiple biological replicates

For differential expression analysis of *S*. Typhimurium strains 4/74, D23580, and D37712, the raw counts (Table S4) from 3-5 biological replicates in four growth conditions were used (ESP, anaerobic growth (abbreviated as NoO2), NonSPI2, and InSPI2). Differential expression analysis was done using DESeq2 v1.24.0 with default settings. A gene was considered to be differentially expressed if the absolute value of its log2 fold change was at least 1 (i.e. fold change > 2), and adjusted p-value was< 0.001.

The SalComD37712 community data resource, and the associated Jbrowse genome browser

SalCom provides a user-friendly Web interface that allows the visualisation and compaison of gene expression values across multiple conditions and between strains. Particular genes can be selected through pre-defined lists of interest, such as all sRNAs or all genes belonging to a

specific pathogenicity island. The resulting heatmap-style display highlights expression differences, and provides access to the rich, manually curated annotation of strains D37712 and D23580. The actual values behind the display can be downloaded for further processing, and a link connects the current view to a genome browser interface.

Visualisation of all the RNA-seq and dRNA-seq (TSS) coverage tracks in JBrowse 1.16.6 shows sequence reads mapped against the combined reference genome described above. Overall, the genomic distance between strains 4/74 and D23580 (approximately 1000 SNPs, or ~1 SNP per 5000 nucleotides), and between D37712 and D23580 (approximately 30 SNPs, ~1 SNP per 150,000 nucleotides) allowed the alignment of RNA-seq reads to the simplified combined reference genome without significant loss of reads. The combined reference genome facilitated a direct comparison of gene coverage as well as transcriptional start sites. The unified browser is hosted at http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data.

Phenotypic and mixed competitive growth experiments

The swimming motility of S. Typhimurium strains D37712, D23580 and 4/74 was determined by a plate assay (Canals *et al.*, 2019b), which involved spotting 3 µL overnight culture onto 0.3% LB agar. Relative motility of the three strains was assessed by migration diameter after 4h and 8h of incubation at 37^{O}C.

Relative expression of the *ssaG* SPI2 promoter in strains D23580 and D37712 was measured at the single cell level via GFP fluorescence. Following the construction of a kanamycin-sensitive derivative of D23580 (strain JH4235), a P*ssaG::gfp*$^{+}$ transcriptional fusion was incorporated into the chromosome of JH4235 and D37712 by inserting the *gfp*$^{+}$ gene downstream of the *ssaG* gene, under the control of the P*ssaG* promoter. The P*ssaG::gfp*$^{+}$ D23580 derivative (JH4692), and the P*ssaG::gfp*$^{+}$ D37712 derivative (JH4693) are listed in Table S8.

The strains JH4692 and JH4693 were genome sequenced to confirm the integrity of the transcriptional fusions, and to verify that unintended nucleotide changes had not arisen. Following growth in 25 mL non-inducing NonSPI2 media in a 250 mL flask at 37°C with shaking at 220 rpm for approximately 8 hours until OD_{600}=0.3, fluorescence was determined with a BD FACSAria Flow Cytometer. The relative fluorescence of the two strains JH4692 and JH4693, and the numbers of individual fluorescent bacteria that expressed the P*ssaG::gfp*$^{+}$ promoter, were determined with FlowJo VX software.

The relative fitness of S. Typhimurium strains D37712 and D23580 was assessed in two independent mixed-growth experiments. First, kanamycin-resistant derivatives of each strain were constructed by inserting the *aph* kanamycin resistance gene into the chromosome at the intergenic region between the *STM4196* and *STM4197* genes, a region that we have previously shown to be transcriptionally silent (Canals *et al.*, 2019b). The strains were designated D23580::Km^{R} JH3794 and D37712::Km^{R}, JH4232. Mixed cultures of wild-type or

kanamycin-resistant derivatives of each strain were grown overnight in LB, InSPI2 and NonSPI2 media in a 250 mL flask at 37°C with shaking at 220 rpm. Following plating on LB agar or LB + kanamycin, colonies were counted and the ratio of bacterial strains was determined. To confirm that the insertion of kanamycin resistance at the intergenic region between *STM4196* and *STM4197* did not impact upon fitness, a mixed-growth experiment was done in both LB and NonSPI2 media (Fig. S7).

Second, to independently assess relative fitness, Tn*7*-based plasmids (Schlechter and Remus-Emsermann, 2019) were used to construct chromosomal sGFP2 and mScarlet derivatives of S. Typhimurium strains D23580 (sGFP2 derivative: JH4694; mScarlet derivative: JH4695) and D37712 (sGFP2 derivative: JH4696; mScarlet derivative: JH4697). The gene cassettes were inserted into the *S.* TyphimuriumTn*7* insertion site between the gene *STMMW_38451* and *glmS*. Mixed cultures of pairs of fluorescently-labelled strains were grown in NonSPI2 media at 37°C with shaking at 220 rpm for approximately 8 hours until OD_{600}=0.3. Levels of green and red fluorescence were determined with a BD FACSAria Flow Cytometer.

Figure Legends

Fig. 1. Emergence of *S.* Typhimurium ST313 sublineages L2.2 and L2.3 in Malawi. (A) Evolutionary dynamics of S. Typhimurium lineages in Blantyre, Malawi from 1996 to 2018. A maximum likelihood tree constructed with 1000 bootstraps using the GTRGAMMA model in RaxML rooted on ST19, LT2. The genomes of 549 *S*. Typhimurium ST313 isolates from bacteraemic patients at the Queen Elizabeth Hospital in Blantyre, Malawi were used for this analysis. The proportions of the five lineages/sublineages are shown. **(B)** The total number of isolates of each lineage/sublineage per year. **(C)** Phylogenetic comparison between representative strains of *S.* Typhimurium ST19 and four ST313 lineages/sublineages (L1, L2.0, L2.2, L2.3) showing the presence and absence of plasmids, prophages and the *spvD* pseudogene. The complete phylogenetic analysis of 707 *S.* Typhimurium genomes is shown in Fig.S1.

Fig. 2. Key genetic similarities and differences between the chromosome and plasmid profiles of D23580 (lineage 2) and D37712 (L2.2). (A) A comparison of the D23580 (L2.0) and D37712 (L2.2) chromosomes. The dots around the chromosome are different kinds of SNPs identified. Phages and *Salmonella* pathogenicity islands are shown in blue and red respectively. **(B)** Plasmid profile of D37712 versus D23580. The pSLT-BT virulence plasmid is present in both D37712 and D23580, and carries the Tn-21 transposable element; **(C)** pCol1B9 is present in D37712 and absent from D23580 (D) pBT3 is present in both D37712 and D23580. **(E)** Absence of *sseI* gene and the STM1050 coding sequence in L2.2 (D37712), as compared to *S.* Typhimurium ST19 4/74 and *S.* Typhimurium ST313 L2.0 (D23580). **(F)** List of pseudogenes in D37712 and D23580, with reference to 4/74. The colour blue means pseudogene/disrupted gene while grey indicates functional genes. *macB* is a pseudogene in D23580 (L2.0) but not in L2.2, while *spvD* is a pseudogene in L2.2 but not in L2.0. All L2.2 strains share similar pseudogenes.

Fig. 3. General comparison of expression profiles of strains 4/74, D23580, and D37712 under 17 different *in vitro* conditions. (A) Principal component analysis (PCA) plot of the individual RNA-seq samples, indicating the overall similarity in gene expression between the three strains. The 17 growth conditions have been defined previously (Kröger *et al.*, 2013). **(B)** Visualization of SPI-2 pathogenicity island expression with the Jbrowse genomic browser, at mid-exponential phase (MEP), InSPI2, and NonSPI2 *in vitro* conditions, which can be accessed here. **(C)** Boxplot visualization of SPI-2 gene expression at mid-exponential phase (MEP), InSPI2, and NonSPI2 *in vitro* conditions. The y-axis shows the combined log TPM values for 45 genes located in the SPI2 pathogenicity island, namely *ssaU, ssaT, ssaS, ssaR, ssaQ, ssaP, ssaO, ssaN, ssaV, ssaM, ssaL, ssaK, STnc1220, STM1410, ssaJ, ssaI, ssaH, ssaG, sseG, sseF, sscB, sseE, sseD, sseC, sscA, sseB, sseA, ssaE, ssaD, ssaC, ssaB, ssrA, ssrB, orf242, orf319, orf70, ttrR, ttrS, ttrC, ttrB, ttrA, orf408, orf245, orf32, and orf48*. The elevated expression of SPI-2 genes in strain D37712 cultured in NonSPI2 media is highlighted in a red box.

Fig. 4. Differential gene expression of *S.* Typhimurium 4/74, D37712, and D23580 under 4 in vitro conditions. (**A)** Boxplots indicating the number of differentially-expressed genes identified in the following *in vitro* growth conditions: early stationary phase, ESP; anaerobic growth, NoO2; SPI-2 inducing medium, InSPI2; SPI-2 non-inducing minimal medium, NonSPI2. Multiple (3 to 5) biological replicates were used for comparison. DESeq2 was used for differential analysis; only genes with |log2FC| ≥ 1 and with adjusted *p*-value ≤ 0.001 were retained. **(B)** Heatmap of the genes differentially expressed between D23580 and D37712. Functional groups and operons of interest are highlighted on the right of Panel B.

Fig. 5. Phenotypes that distinguish ST313 L2.2 from ST313 L2.0. (A) Swimming motility assay of strains D23589, D37712 and 4/74, with a representative plate shown on the left. Average migration diameters were measured after 4 and 8 hours. Each bar represents the mean of three biological replicates, with *error bars* showing standard deviation. Significant difference (***) indicates *P* value (*t* test) < 0.001. In Panels B & C, comparison of *ssaG* expression by flow cytometry using D23580 and D37712 derivatives containing a chromosomal *ssaG*-GFP$^+$ transcriptional fusion, strains SZS008 and SZS032, respectively. Cells were collected at 8 hours after inoculation in NonSPI2 media. Ten thousand events were acquired for each sample. **(B)** Mean fluorescent intensity signal of *ssaG*-GFP$^+$ for D23580 (SZS008, dark grey) and D37712 (SZS032, grey) grown in NonSPI2 media. Significant difference (***) indicates *P* value (*t* test) < 0.001. **(C)** The proportions of bacterial cells that expressed *ssaG*-GFP$^+$ during growth in NonSPI2 media was determined. Percentage of GFP-expressing (green) and non-fluorescent cells (white) for D23580 (SZS008) and D37712 (SZS032) is shown. Each bar represents the mean of three biological replicates, error bars show standard deviation. Significant difference (***) indicates *P* value (*t* test) < 0.001. **(D)** Relative fitness of wild-type D23580 and D37712 and their kanamycin-resistant derivatives. Bacterial numbers were determined after overnight culture of a 1:1 mixture (wild-type versus Km^R) in LB (left), InSPI2 (middle) and NonSPI2 (right) media. Each dot represents the log-transformed mean competitive index of three biological replicates with *error bars* representing 95% confidence interval from standard deviation. A log number higher than 0 reflects the increased fitness of kanamycin-resistant derivatives. *P* values were determined by *t* test (***: $P < 0.001$; **: $P < 0.01$; *: $P < 0.05$; ns: not significant).

Supporting information

Fig. S1, Maximum-likelihood phylogeny of 707 African *S.* Typhimurium isolates. All genome sequences have been published (Msefula *et al.,* 2012, Pulford *et al.,* 2021, Canals *et al.,* 2019b). Raw sequence reads were aligned to the *S.* Typhimurium D23580 genome (FN424405) using Snippy. The recombination sites of the alignment were removed by Gubbins, and the phylogenetic tree was built with Raxml-ng. The tree is rooted on *Salmonella* Typhi strain CT18 as the outgroup. The MLST sequence types, HierBAPS level 1 and level 2 clusters are shown in coloured concentric rings as indicated. The *S.* Typhimurium ST313 isolates are

categorised as Lineage 1, Lineage 2 or Lineage 3 according to HierBAPS level 1 clustering. ST313 Lineage 2 was then sub- divided into 3 sub-lineages according to HierBAPS level 2 clustering: ST313 L2.0, ST313 L2.2 and ST313 L2.3. The metadata and lineage designations of all the S. Typhimurium isolates are in Table S7.

Fig. S2. PCR-based confirmation of the deletion of the *sseI* gene from *S.* Typhimurium L2.2 D37712. Arrows from left to right show the forward strand while the left strand is shown by arrows from right to left. However, *sseI* gene in D23580 is a pseudogene with a SNP indicated as a red line.

Fig. S3. Genomic comparison of plasmids pCol1B9$^{4/74}$ and pCol1B9^{D37712} using Artemis Comparison Tool (ACT). Bottom panel details the differences observed in the most divergent regions, including colicin toxin-antitoxin system (in pCol1B9) and *impC-umuC-umuD* operon (in pCol1B9).

Fig. S4. RDAR Phenotypes of 4/74, D23580, D37712 and BKQZM9. The top panel shows the RDAR morphology assay and the bottom panel shows a complementary experiment that involves the induction of biofilm formation on 1% tryptone agar (MacKenzie *et al.,* 2019). Strain 4/74 was used as a RDAR-positive control, which has concentric rings and a wrinkled appearance (Pulford *et al.,* 2021). The S. Typhimurium ST313 L3 strain BKQZM9 is shown for comparative purposes.

Fig. S5. Competitive index analysis of D23580 and D37712 using fluorescently-tagged S. Typhimurium strains (A) Km^R-sGFP2 and Gm^R-mScarlet were inserted into the transposon Tn7 site of D23580 or D37712. Bent arrows represent promoters and directional arrows represent genes. **(B)** A 1:1 mix of Km^R-sGFP2 and Gm^R-mScarlet marked strain was inoculated in NonSPI2 media, followed by an overnight incubation in 37℃. The percentage of sGFP2 (green) and mScarlet (Red) -marked cells was determined by flow cytometry. Raw data are shown in Figure S7, 10,000 events were acquired for each sample. **(C)** Competitive index analysis of Km^R-sGFP2 and Gm^R-mScarlet marked strain. Bacterial numbers were determined by counting CFU for overnight culture of a 1:1 mixture in NonSPI2 media. Each dot represents a single biological replicate and the lane represents mean value. A competitive index of 1 indicates the equal fitness of two strains, while a number higher than 1 reflects an increased fitness of D37712.

Fig. S6. Raw flow cytometric data related to Fig. S5B. **(A)** JH4695 + JH4698 and **(B)** JH4696 + JH4697. A 1:1 mix of the Km^R-sGFP2 and Gm^R-mScarlet marked strains were inoculated in NonSPI2 media, followed by growth at 37℃ until OD_{600} = 0.3. The X-axis (labelled FITC) shows the GFP level and the Y-axis (labelled PE Yell-Grn) indicates the mScarlet level. Quadrant gates were used to separate four populations, and the black numbers indicate the percentage of events in each quadrant. In total, 10,000 events were acquired for each sample.

Fig. S7. The insertion of GFP-Km or RFP-Gm did not impact on fitness. A 1:1 mix of Km^R-sGFP2 and Gm^R-mScarlet marked strains were inoculated in LB or NonSPI2 media, followed by overnight incubation in 37℃. The competitive index (CI) was calculated using the formula $(CFU_{Gm})/(CFU_{Km})$. Each dot represents the CI from a single replicate and the horizontal bars indicate the mean of each dataset.

Supplementary data

Table S1: SNP and indel variants that differentiate L2.2 (strain D37712) and L2.3 (strain D49679).

Table S2: SNP and indel variants that differentiate L2.2 (strain D37712) and L2.0 (strain D23580).

Table S3: Pseudogenes carried by ST19 and ST313 L2.0 and L2.2 (strains 4/74, D23580 and D37712).

Table S4: Raw read counts for all processed RNA-seq samples shown in Figures 3 and 4 (strains 4/74, D23580, and D37712).

Table S5: TPM values for all processed RNA-seq samples shown in Figures 3 and 4 (strains 4/74, D23580, and D37712).

Table S6: DESeq2-based differential gene expression analysis for strains D23580 vs D37712 grown in four *in vitro* conditions.

Table S7: Metadata and lineage designations of the 708 S. Typhimurium isolates used to generate the maximum likelihood phylogeny (Fig. S1).

Table S8: Bacterial strains used in this study.

Acknowledgements

The authors thank Brian Coombes and Rob Kingsley for their constructive comments during the peer review process. We are grateful to present and former members of the Hinton laboratory for helpful discussions, and to Paul Loughnane for his expert technical assistance.

This work was supported by a Wellcome Trust Investigator award [grant numbers 106914/Z/15/Z and 222528/Z/21/Z] to J.C.D.H., and by the Malawi-Liverpool-Wellcome Research Centre Director's Fund. B.K. was funded by an AESA-RISE fellowship from the African Academy of Sciences [Grant Number: RPDF-18-04]. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

Conceptualization: B.K., R.H., M.A.G., C.L.M. and J.C.D.H.

Data curation: B.K., R.C., A.V.P., C.V.P., P.A.

Formal analysis: B.K., R.C., C.V.P., A.V.P., X.Z., C.K., S.V.O., Y.L., P.A., A.D. and J.C.D.

Funding acquisition: B.K., R.C., A.V.P., X.Z. and J.C.D.H.

Investigation: B.K., R.C.A., A.V.P., X.Z. and J.C.D.H.

Methodology: B.K., R.H., M.A.G., C.L.M. and J.C.D.H.

Project administration: B.K. and J.C.D.H.

Resources: B.K., R.H., M.A.G, C.L.M. and J.C.D.H.

Software: B.K. and A.V.P.

Supervision: M.A.G., C.L.G., C.L.M. and J.C.D.H.

Validation: B.K., R.C., A.V.P., X.Z. and J.C.D.H.

Visualization: B.K., R.C., Y.L., C.V.P., A.V.P. and J.C.D.H.

Writing original draft: B.K., R.C. and J.C.D.H

Writing reviews and editing: B.K., R.C., A.V.P., X.Z., C.K., S.V.O., A.D., R.H., M.G and J.C.D.H.

Equal contribution: Authors B.K., R.C. and A.V.P. made equal contributions to this work.

References

Ackermann M, Stecher B, Freed N, Songhet P, Hardt W, and Doebeli M (2008) Self-destructive cooperation mediated by phenotypic noise. 454:987–990.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aulicino A, Antanaviciute A, Frost J, Sousa Geros A, Mellado E, Attar M, Jagielowicz M, Hublitz P, Sinz J, Preciado-Llanes L, Napolitani G, Bowden R, Koohy H, Drakesmith H, and Simmons A (2022) Dual RNA sequencing reveals dendritic cell reprogramming in response to typhoidal Salmonella invasion. *Commun Biol* 5:111.

Bogomolnaya LM, Santiviago CA, Yang H-J, Baumler AJ, and Andrews-Polymenis HL (2008) 'Form variation' of the O12 antigen is critical for persistence of *Salmonella* Typhimurium in the murine intestine. *Mol Microbiol* 70:1105–1119.

Bolger AM, Lohse M, and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Branchu P, Bawn M, and Kingsley RA (2018) Genome Variation and Molecular Epidemiology of Salmonella enterica Serovar Typhimurium Pathovariants. *Infect Immun* 86:e00079-18.

Brink T, Leiss V, Siegert P, Jehle D, Ebner JK, Schwan C, Shymanets A, Wiese S, Nürnberg B, Hensel M, Aktories K, and Orth JHC (2018) Salmonella Typhimurium effector SseI inhibits chemotaxis and increases host cell survival by deamidation of heterotrimeric Gi proteins. *PLOS Pathog* 14:e1007248.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Canals R, Chaudhuri RR, Steiner RE, Owen SV, Quinones-Olvera N, Gordon MA, Baym M, Ibba M, and Hinton JCD (2019a) The fitness landscape of the African Salmonella Typhimurium ST313 strain D23580 reveals unique properties of the pBT1 plasmid. *PLOS Pathog* 15:e1007948.

Canals R, Hammarlöf DL, Kröger C, Owen SV, Fong WY, Lacharme-Lora L, Zhu X, Wenner N, Carden SE, Honeycutt J, Monack DM, Kingsley RA, Brownridge P, Chaudhuri RR, Rowe WPM, Predeus AV, Hokamp K, Gordon MA, and Hinton JCD (2019b) Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. *PLOS Biol* 17:e3000059.

Carden SE, Walker GT, Honeycutt J, Lugo K, Pham T, Jacobson A, Bouley D, Idoyaga J, Tsolis RM, and Monack D (2017) Pseudogenization of the Secreted

Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313 within Migratory Dendritic Cells. *Cell Host Microbe* 21:182–194.

Chirwa E, Dale, H, Gordon, MA, and Ashton, PM (2023) What is the Source of Infections Causing Invasive Nontyphoidal Salmonella Disease? *Open Forum Infect Diseaases* February 2023.

Crump JA, Sjölund-Karlsson M, Gordon MA, and Parry CM (2015) Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive Salmonella Infections. *Clin Microbiol Rev* 28:901–937.

Ewels P, Magnusson M, Lundin S, and Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

Fass E, and Groisman E (2009) Control of Salmonella pathogenicity island-2 gene expression. *Curr Opin Microbiol* 12:199–204.

Feasey NA, Dougan G, Kingsley RA, Heyderman RS, and Gordon MA (2012) Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *The Lancet* 379:2489–2499.

Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist L, Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C, Kariuki S, Maclennan CA, Onsare RS, Weill F-X, Le Hello S, Smith AM, McClelland M, Desai P, Parry CM, Cheesbrough J, French N, Campos J, Chabalgoity JA, Betancor L, Hopkins KL, Nair S, Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO, Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA, Parkhill J, Heyderman RS, Dougan G, Gordon MA, and Thomson NR (2016) Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat Genet* 48:1211–1217.

Gilchrist J, and MacLennan C (2019) Invasive Nontyphoidal Salmonella Disease in Africa. *Cell Mol Biol E Coli Salmonella Enterobact*, doi: doi:10.1128/ecosalplus.ESP-0007-2018.

Grabe GJ, Zhang Y, Przydacz M, Rolhion N, Yang Y, Pruneda JN, Komander D, Holden DW, and Hare SA (2016) The Salmonella Effector SpvD Is a Cysteine Hydrolase with a Serovar-specific Polymorphism Influencing Catalytic Activity, Suppression of Immune Responses, and Bacterial Virulence. *J Biol Chem* 291:25853–25863.

Gurevich A, Saveliev V, Vyahhi N, and Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075.

Hammarlöf DL, Kröger C, Owen SV, Canals R, Lacharme-Lora L, Wenner N, Schager AE, Wells TJ, Henderson IR, Wigley P, Hokamp K, Feasey NA, Gordon MA, and Hinton JCD (2018) Role of a single noncoding nucleotide in

the evolution of an epidemic African clade of *Salmonella*. *Proc Natl Acad Sci* 115.

Hautefort I, Proença MJ, and Hinton JCD (2003) Single-Copy Green Fluorescent Protein Gene FusionsAllow Accurate Measurement of *Salmonella* Gene Expression InVitro and during Infection of MammalianCells. *Appl Environ Microbiol* 69:7480–7491.

Honeycutt JD, Wenner N, Li Y, Brewer SM, Massis LM, Brubaker SW, Chairatana P, Owen SV, Canals R, Hinton JCD, and Monack DM (2020) Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive Salmonella isolates from Africa. *PLOS Pathog* 16:e1008763.

Jennings E, Thurston TLM, and Holden DW (2017) Salmonella SPI-2 Type III Secretion System Effectors: Molecular Mechanisms And Physiological Consequences. *Cell Host Microbe* 22:217–231.

Johnson R, and Simon M (1985) Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. *Cell* 41:781–791.

Kariuki S, Revathi G, Kariuki N, Kiiru J, Mwituria J, Muyodi J, Githinji JW, Kagendo D, Munyalo A, and Hart CA (2006) Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol* 55:585–591.

Kasumba IN, Pulford CV, Perez-Sepulveda BM, Sen S, Sayed N, Permala-Booth J, Livio S, Heavens D, Low R, Hall N, Roose A, Powell H, Farag T, Panchalingham S, Berkeley L, Nasrin D, Blackwelder WC, Wu Y, Tamboura B, Sanogo D, Onwuchekwa U, Sow SO, Ochieng JB, Omore R, Oundo JO, Breiman RF, Mintz ED, O'Reilly CE, Antonio M, Saha D, Hossain MJ, Mandomando I, Bassat Q, Alonso PL, Ramamurthy T, Sur D, Qureshi S, Zaidi AKM, Hossain A, Faruque ASG, Nataro JP, Kotloff KL, Levine MM, Hinton JCD, and Tennant SM (2021) Characteristics of *Salmonella* Recovered From Stools of Children Enrolled in the Global Enteric Multicenter Study. *Clin Infect Dis* 73:631–641.

Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, Marsh K, Achtman M, Molyneux ME, Cormican M, Parkhill J, MacLennan CA, Heyderman RS, and Dougan G (2009) Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* 19:2279–2287.

Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, Döpfer D, Fazil A, Fischer-Walker CL, Hald T, Hall AJ, Keddy KH, Lake RJ, Lanata CF, Torgerson PR, Havelaar AH, and Angulo FJ (2015) World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLOS Med* 12:e1001921.

Koolman L, Prakash R, Diness Y, Msefula C, Nyirenda TS, Olgemoeller F, Perez-Sepulveda B, Hinton JCD, Owen SV, Feasey NA, Ashton PM, and Gordon MA (2022) *Case-control investigation of invasive Salmonella disease in Africa – comparison of human, animal and household environmental isolates find no evidence of environmental or animal reservoirs of invasive strains*, Infectious Diseases (except HIV/AIDS).

Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K, and Hinton JCD (2013) An Infection-Relevant Transcriptomic Compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* 14:683–695.

Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hébrard M, Händler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J, and Hinton JCD (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci* 109.

Lai MA, Quarles EK, López-Yglesias AH, Zhao X, Hajjar AM, and Smith KD (2013) Innate Immune Detection of Flagellin Positively and Negatively Regulates Salmonella Infection. *PLoS ONE* 8:e72047.

Letunic I, and Bork P (2006) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:1283–1287.

Lian ZJ, Phan M, Hancock SJ, Nhu NTK, Paterson DL, and Schembri MA (2023) Genetic basis of I-complex plasmid stability and conjugation. *PLOS Genet* 19:e1010773.

Löber S, Jäckel D, Kaiser N, and Hensel M (2006) Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *Int J Med Microbiol* 296:435–447.

Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, and Hinton JCD (2006) H-NS Mediates the Silencing of Laterally Acquired Genes in Bacteria. *PLoS Pathog* 2:e81.

Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, and Hoekstra RM (2010) The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis. *Clin Infect Dis* 50:882–889.

Marchello C, Dale, A. P., Pisharody, S., Rubach, M. P., and Crump, J.A. (2019) A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. Antimicrobial Agents and Chemotherapy. 64.

Marchello CS, Dale AP, Pisharody S, Rubach MP, and Crump JA (2019) A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. *Antimicrob Agents Chemother* 64:e01974-19.

Msefula CL, Kingsley RA, Gordon MA, Molyneux E, Molyneux ME, MacLennan CA, Dougan G, and Heyderman RS (2012) Genotypic Homogeneity of Multidrug Resistant S. Typhimurium Infecting Distinct Adult and Childhood Susceptibility Groups in Blantyre, Malawi. *PLoS ONE* 7:e42085.

Musicha P, Cornick JE, Bar-Zeev N, French N, Masesa C, Denis B, Kennedy N, Mallewa J, Gordon MA, Msefula CL, Heyderman RS, Everett DB, and Feasey NA (2017) Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *Lancet Infect Dis* 17:1042–1052.

Nedialkova LP, Denzler R, Koeppel MB, Diehl M, Ring D, Wille T, Gerlach RG, and Stecher B (2014) Inflammation Fuels Colicin Ib-Dependent Competition of Salmonella Serovar Typhimurium and E. coli in Enterobacterial Blooms. *PLoS Pathog* 10:e1003844.

Okoro CK, Barquist L, Connor TR, Harris SR, Clare S, Stevens MP, Arends MJ, Hale C, Kane L, Pickard DJ, Hill J, Harcourt K, Parkhill J, Dougan G, and Kingsley RA (2015) Signatures of Adaptation in Human Invasive Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. *PLoS Negl Trop Dis* 9:e0003611.

Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, Kariuki S, Msefula CL, Gordon MA, de Pinna E, Wain J, Heyderman RS, Obaro S, Alonso PL, Mandomando I, MacLennan CA, Tapia MD, Levine MM, Tennant SM, Parkhill J, and Dougan G (2012) Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet* 44:1215–1221.

Owen SV, Wenner N, Canals R, Makumi A, Hammarlöf DL, Gordon MA, Aertsen A, Feasey NA, and Hinton JCD (2017) Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. *Front Microbiol* 8.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, and Parkhill J (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693.

Perez-Sepulveda BM, and Hinton JCD (2018) Functional Transcriptomics for Bacterial Gene Detectives. *Microbiol Spectr* 6:6.5.06.

Piccini G, and Montomoli E (2020) Pathogenic signature of invasive non-typhoidal *Salmonella* in Africa: implications for vaccine development. *Hum Vaccines Immunother* 16:2056–2071.

Post AS, Diallo SN, Guiraud I, Lompo P, Tahita MC, Maltha J, Van Puyvelde S, Mattheus W, Ley B, Thriemer K, Rouamba E, Derra K, Deborggraeve S, Tinto H, and Jacobs J (2019) Supporting evidence for a human reservoir of invasive non-Typhoidal Salmonella from household samples in Burkina Faso. *PLoS Negl Trop Dis* 13:e0007782.

Preciado-Llanes L, Aulicino A, Canals R, Moynihan PJ, Zhu X, Jambo N, Nyirenda TS, Kadwala I, Sousa Gerós A, Owen SV, Jambo KC, Kumwenda B, Veerapen N, Besra GS, Gordon MA, Hinton JCD, Napolitani G, Salio M, and Simmons A (2020) Evasion of MAIT cell recognition by the African *Salmonella* Typhimurium ST313 pathovar that causes invasive disease. *Proc Natl Acad Sci* 117:20717–20728.

Pulford CV, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ, Wenner N, Rodwell EV, Kumwenda B, Zhu X, Bennett RJ, Stenhouse GE, Malaka De Silva P, Webster HJ, Bengoechea JA, Dumigan A, Tran-Dien A, Prakash R, Banda HC, Alufandika L, Mautanga MP, Bowers-Barnard A, Beliavskaia AY, Predeus AV, Rowe WPM, Darby AC, Hall N, Weill F-X, Gordon MA, Feasey NA, Baker KS, and Hinton JCD (2021) Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. *Nat Microbiol* 6:327–338.

Rankin, J.D. and Taylor R.J. (1966) The estimation of doses of Salmonella typhimurium suitable for the experimental production of disease in calves. *Vec Rec* 78:706–7.

Richardson EJ, Limaye B, Inamdar H, Datta A, Manjari KS, Pullinger GD, Thomson NR, Joshi RR, Watson M, and Stevens MP (2011) Genome Sequences of Salmonella enterica Serovar Typhimurium, Choleraesuis, Dublin, and Gallinarum Strains of Well- Defined Virulence in Food-Producing Animals. *J Bacteriol* 193:3162–3163.

Schlechter R, and Remus-Emsermann M (2019) Delivering "Chromatic Bacteria" Fluorescent Protein Tags to Proteobacteria Using Conjugation. *BIO-Protoc* 9.

Sikand A, Jaszczur M, Bloom LB, Woodgate R, Cox MM, and Goodman MF (2021) The SOS Error-Prone DNA Polymerase V Mutasome and β-Sliding Clamp Acting in Concert on Undamaged DNA and during Translesion Synthesis. *Cells* 10:1083.

Skidmore PD, Canals R, and Ramasamy MN (2023). The iNTS-GMMA vaccine: a promising step in non-typhoidal Salmonella vaccine development. Expert Rev Vaccines (In Press) doi: 10.1080/14760584.2023.2270596.

Srikumar S, Kröger C, Hébrard M, Colgan A, Owen SV, Sivasankaran SK, Cameron ADS, Hokamp K, and Hinton JCD (2015) RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium. *PLOS Pathog* 11:e1005262.

Stamatakis A, Ludwig T, and Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.

Stanaway JD, Parisi A, Sarkar K, Blacker BF, Reiner RC, Hay SI, Nixon MR, Dolecek C, James SL, Mokdad AH, Abebe G, Ahmadian E, Alahdab F, Alemnew BTT, Alipour V, Allah Bakeshei F, Animut MD, Ansari F, Arabloo J,

Asfaw ET, Bagherzadeh M, Bassat Q, Belayneh YMM, Carvalho F, Daryani A, Demeke FM, Demis ABB, Dubey M, Duken EE, Dunachie SJ, Eftekhari A, Fernandes E, Fouladi Fard R, Gedefaw GA, Geta B, Gibney KB, Hasanzadeh A, Hoang CL, Kasaeian A, Khater A, Kidanemariam ZT, Lakew AM, Malekzadeh R, Melese A, Mengistu DT, Mestrovic T, Miazgowski B, Mohammad KA, Mohammadian M, Mohammadian-Hafshejani A, Nguyen CT, Nguyen LH, Nguyen SH, Nirayo YL, Olagunju AT, Olagunju TO, Pourjafar H, Qorbani M, Rabiee M, Rabiee N, Rafay A, Rezapour A, Samy AM, Sepanlou SG, Shaikh MA, Sharif M, Shigematsu M, Tessema B, Tran BX, Ullah I, Yimer EM, Zaidi Z, Murray CJL, and Crump JA (2019) The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis* 19:1312–1324.

Tack B, Phoba M-F, Barbé B, Kalonji LM, Hardy L, Van Puyvelde S, Ingelbeen B, Falay D, Ngonda D, van der Sande MAB, Deborggraeve S, Jacobs J, and Lunguya O (2020) Non-typhoidal Salmonella bloodstream infections in Kisantu, DR Congo: Emergence of O5-negative Salmonella Typhimurium and extensive drug resistance. *PLoS Negl Trop Dis* 14:e0008121.

Van Puyvelde S, Pickard D, Vandelannoote K, Heinz E, Barbé B, de Block T, Clare S, Coomber EL, Harcourt K, Sridhar S, Lees EA, Wheeler NE, Klemm EJ, Kuijpers L, Mbuyi Kalonji L, Phoba M-F, Falay D, Ngbonda D, Lunguya O, Jacobs J, Dougan G, and Deborggraeve S (2019) An African Salmonella Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. *Nat Commun* 10:4280.

Van Puyvelde S, de Block T, Sridhar S, Bawn M, Kingsley RA, Ingelbeen B, Beale MA, Barbé B, Jeon HJ, Mbuyi-Kalonji L, Phoba MF, Falay D, Martiny D, Vandenberg O, Affolabi D, Rutanga JP, Ceyssens PJ, Mattheus W, Cuypers WL, van der Sande MAB, Park SE, Kariuki S, Otieno K, Lusingu JPA, Mbwana JR, Adjei S, Sarfo A, Agyei SO, Asante KP, Otieno W, Otieno L, Tahita MC, Lompo P, Hoffman IF, Mvalo T, Msefula C, Hassan-Hanga F, Obaro S, Mackenzie G, Deborggraeve S, Feasey N, Marks F, MacLennan CA, Thomson NR, Jacobs J, Dougan G, Kariuki S, Lunguya O. (2023) A genomic appraisal of invasive Salmonella Typhimurium and associated antibiotic resistance in sub-Saharan Africa. Nat Commun. 14:6392.

Wick RR, Judd LM, Gorrie CL, and Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595.

Zhang S, Kingsley RA, Santos RL, Andrews-Polymenis H, Raffatellu M, Figueiredo J, Nunes J, Tsolis RM, Adams LG, and Bäumler AJ (2003) Molecular Pathogenesis of *Salmonella enterica* Serotype Typhimurium-Induced Diarrhea. *Infect Immun* 71:1–12.

Fig. 1

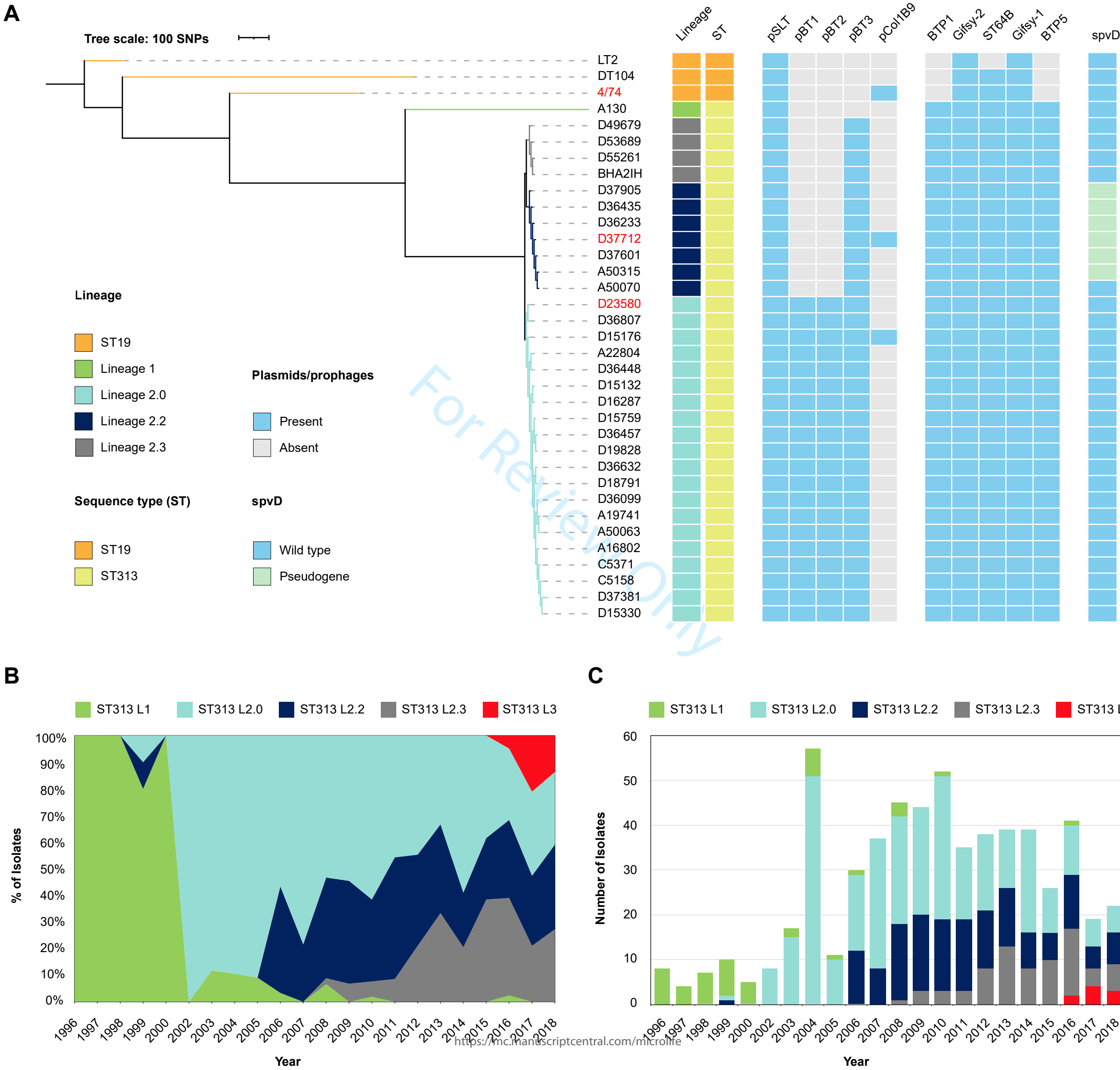
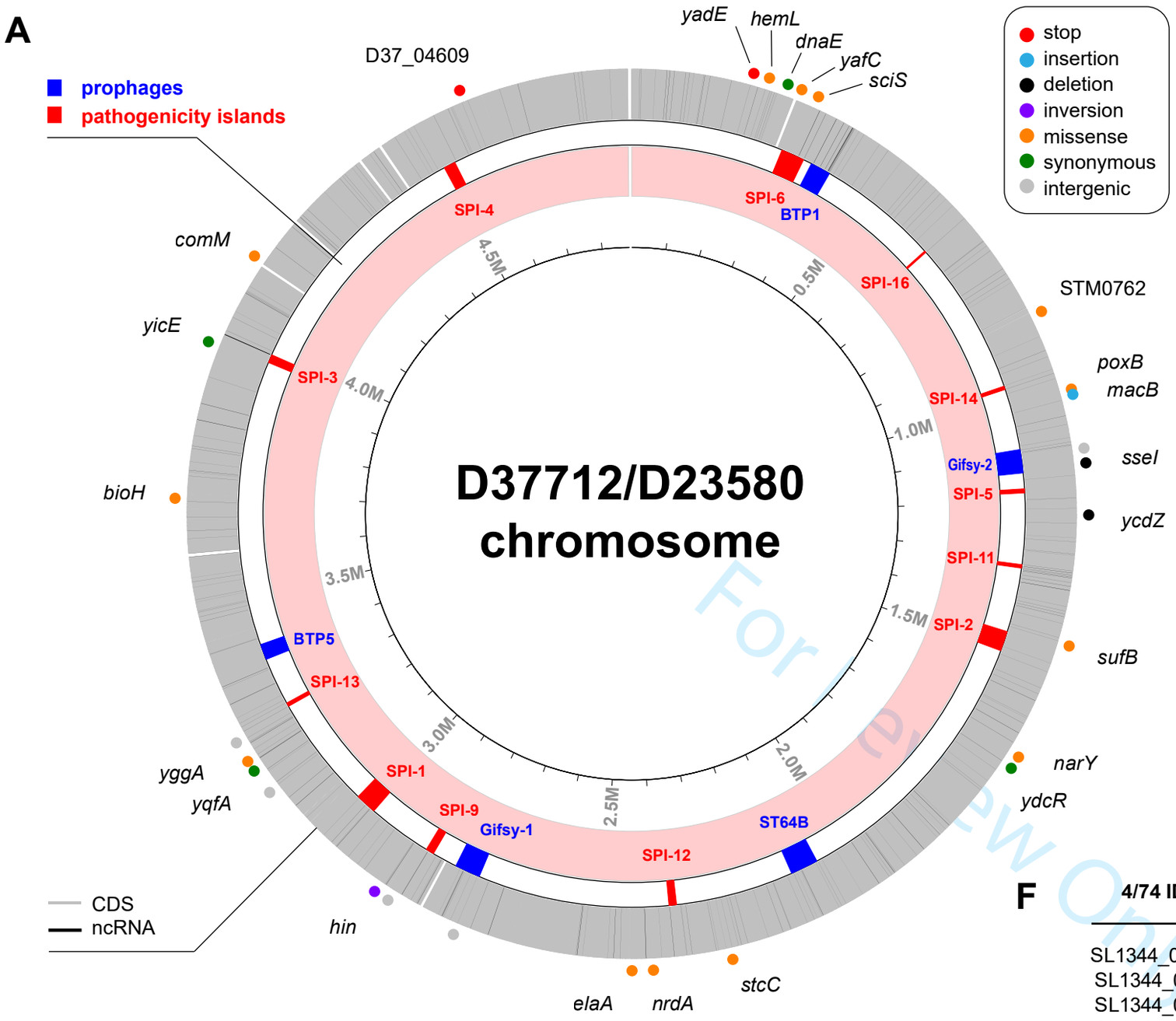
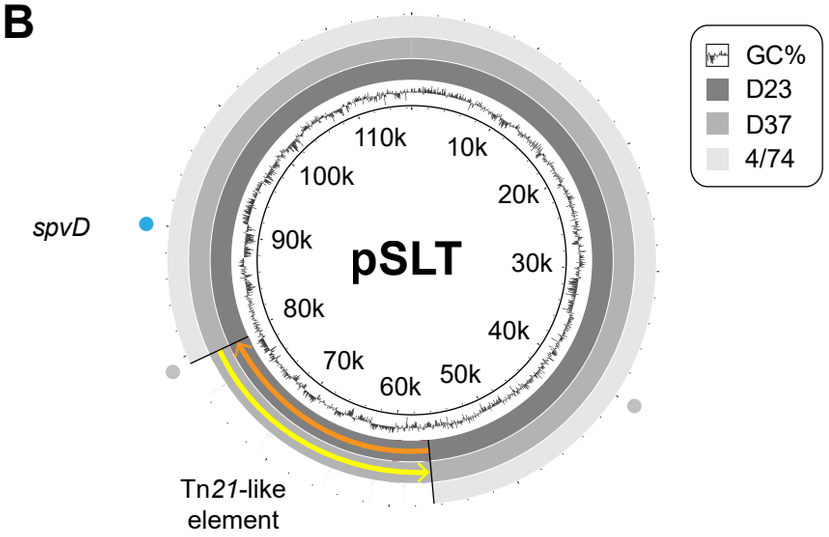
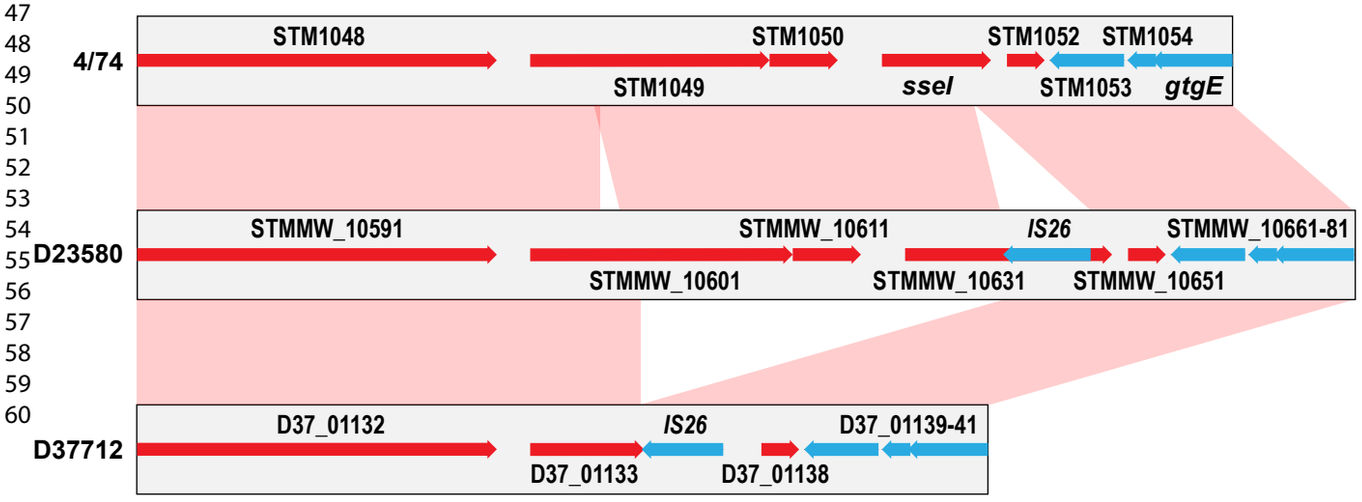


Fig. 2





E

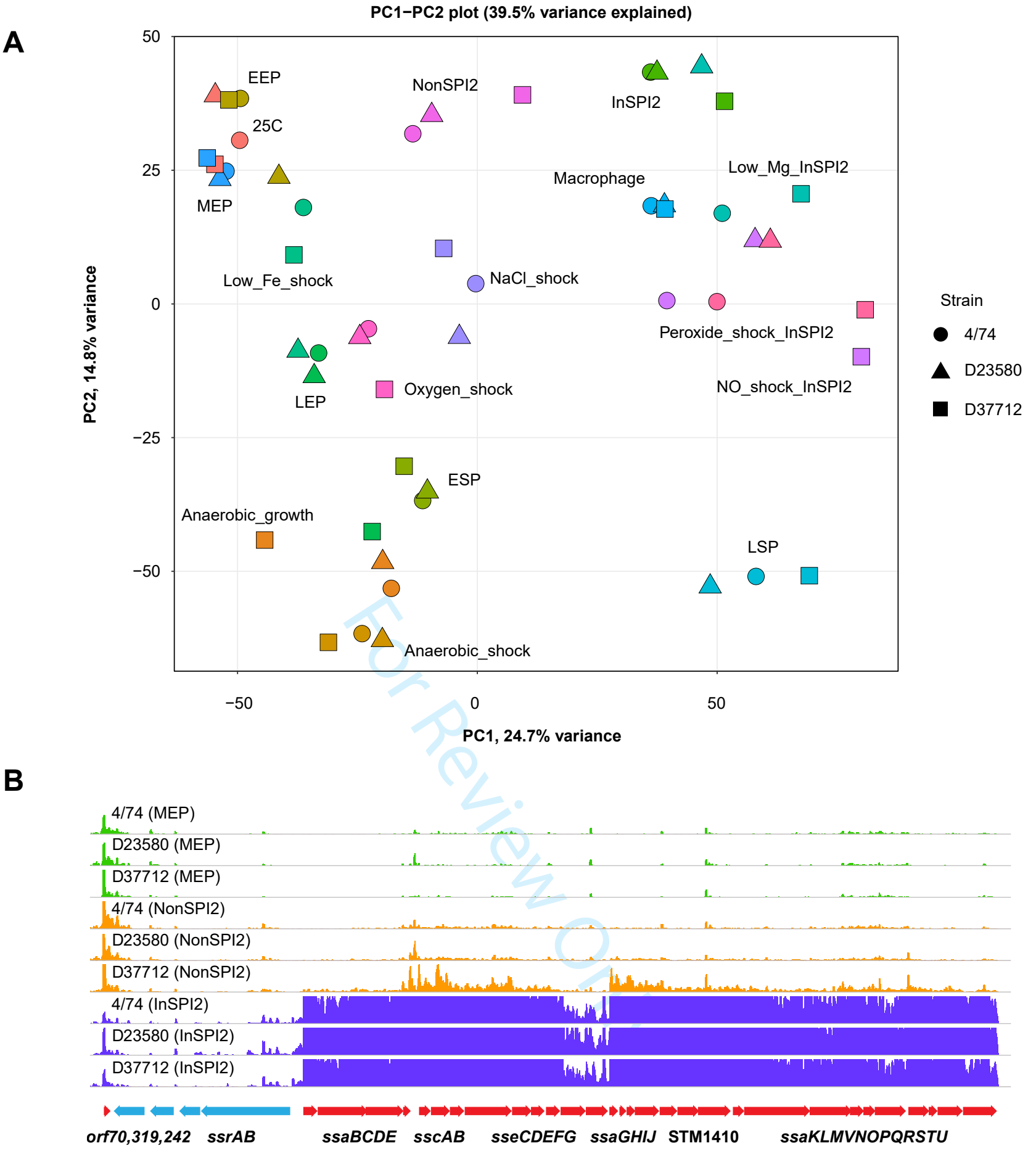


F

4/74 ID	Gene name	Product	4/74	D23580	D37712
SL1344_0066A	**STM0065**	putative viral protein			
SL1344_0180	***yadE***	polysaccharide deacetylase family protein			
SL1344_0708	**STM0726**	glycosyl transferase			
SL1344_0809	***ybiP***	phosphoethanolamine transferase			
SL1344_0881	***macB***	macrolide ABC transporter MacB			
SL1344_0962	**STM1023**	hypothetical protein			
SL1344_1446	***ydeE***	efflux MFS transporter YdeE			
SL1344_1567	**STM1637**	membrane protein			
SL1344_1873	**STM1940**	cell wall-associated hydrolase			
SL1344_2475	***ratB***	hypothetical protein			
SL1344_2659	***yfjG***	type II toxin-antitoxin system RatA family toxin			
SL1344_2911	***ygbE***	DUF3561 family protein			
SL1344_2990	**STM3012**	LacI family DNA-binding transcriptional regulator			
SL1344_3051	**STM3075**	sulfate/molybdate ABC transporter			
SL1344_3327	***ttdA***	L(+)-tartrate dehydratase subunit alpha			
SL1344_3589	***bcsG***	cellulose biosynthesis protein BcsG			
SL1344_3663	**STM3698**	MFS transporter			
SL1344_3711	**STM3745**	hypothetical protein			
SL1344_3733	***dgaE***	D-glucosaminate-6-phosphate ammonia lyase			
SL1344_3765	**STM3796A**	EamA family transporter			
SL1344_4131	**STM4196**	hypothetical protein			
SL1344_4223	***lpxO***	lipid A hydroxylase LpxO			
SL1344_4288	**STM4351**	putative transporter			
SLP1_0070	***spvD***	SPI-2 type III effector cysteine hydrolase SpvD			
SLP1_0103	**PSLT002**	phospholipase			

Wild type Pseudogene

Fig. 3


A
PC1−PC2 plot (39.5% variance explained)
PC2, 14.8% variance
PC1, 24.7% variance
EEP
25C
MEP
NonSPI2
InSPI2
Macrophage
Low_Mg_InSPI2
Low_Fe_shock
NaCl_shock
Peroxide_shock_InSPI2
LEP
Oxygen_shock
NO_shock_InSPI2
ESP
Anaerobic_growth
LSP
Anaerobic_shock
Strain
4/74
D23580
D37712
B
4/74 (MEP)
D23580 (MEP)
D37712 (MEP)
4/74 (NonSPI2)
D23580 (NonSPI2)
D37712 (NonSPI2)
4/74 (InSPI2)
D23580 (InSPI2)
D37712 (InSPI2)
orf70,319,242
ssrAB
ssaBCDE
sscAB
sseCDEFG
ssaGHIJ
STM1410
ssaKLMVNOPQRSTU

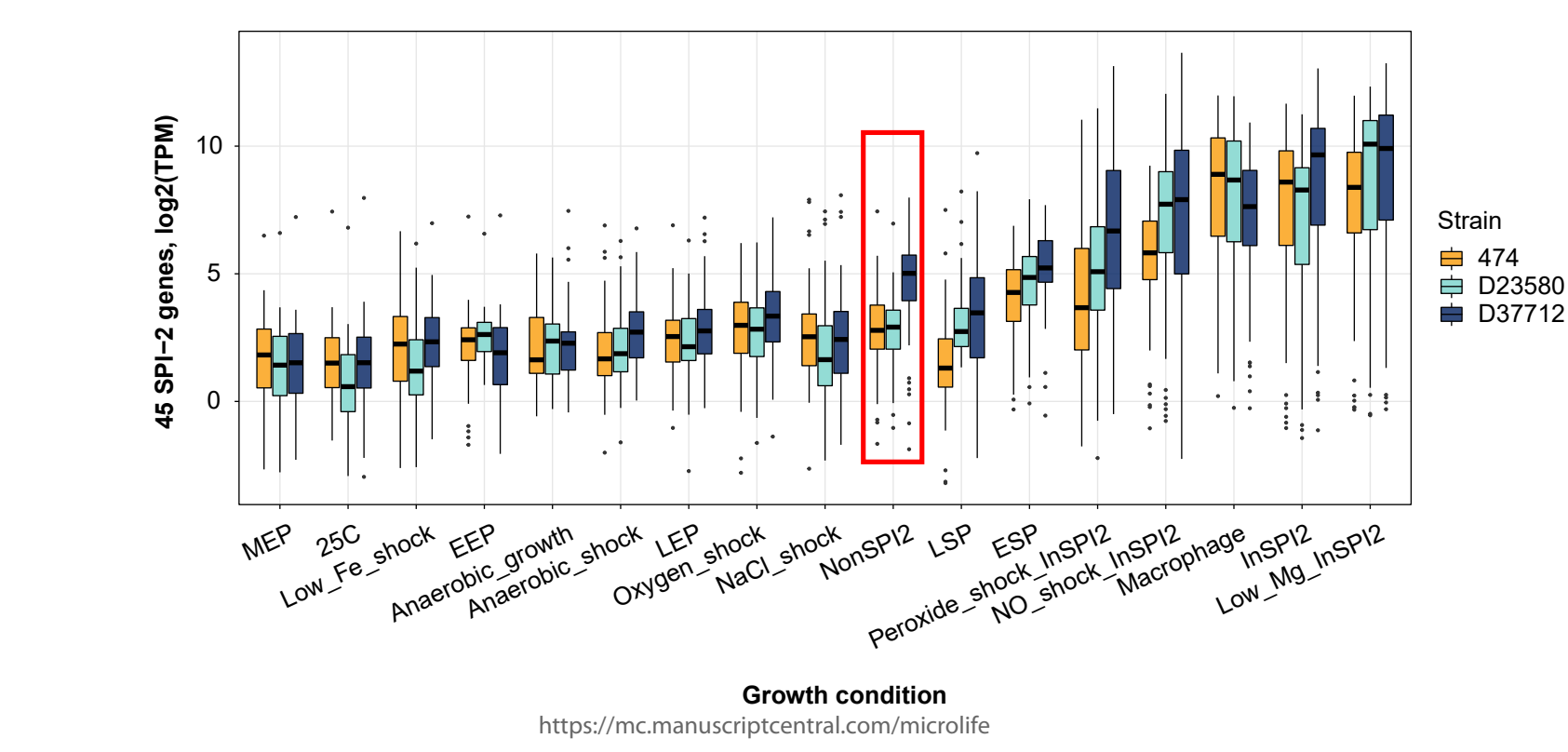

C
45 SPI−2 genes, log2(TPM)
Strain
474
D23580
D37712
MEP
25C
Low_Fe_shock
EEP
Anaerobic_growth
Anaerobic_shock
LEP
Oxygen_shock
NaCl_shock
NonSPI2
LSP
ESP
Peroxide_shock_InSPI2
NO_shock_InSPI2
Macrophage
InSPI2
Low_Mg_InSPI2
Growth condition

Fig. 4

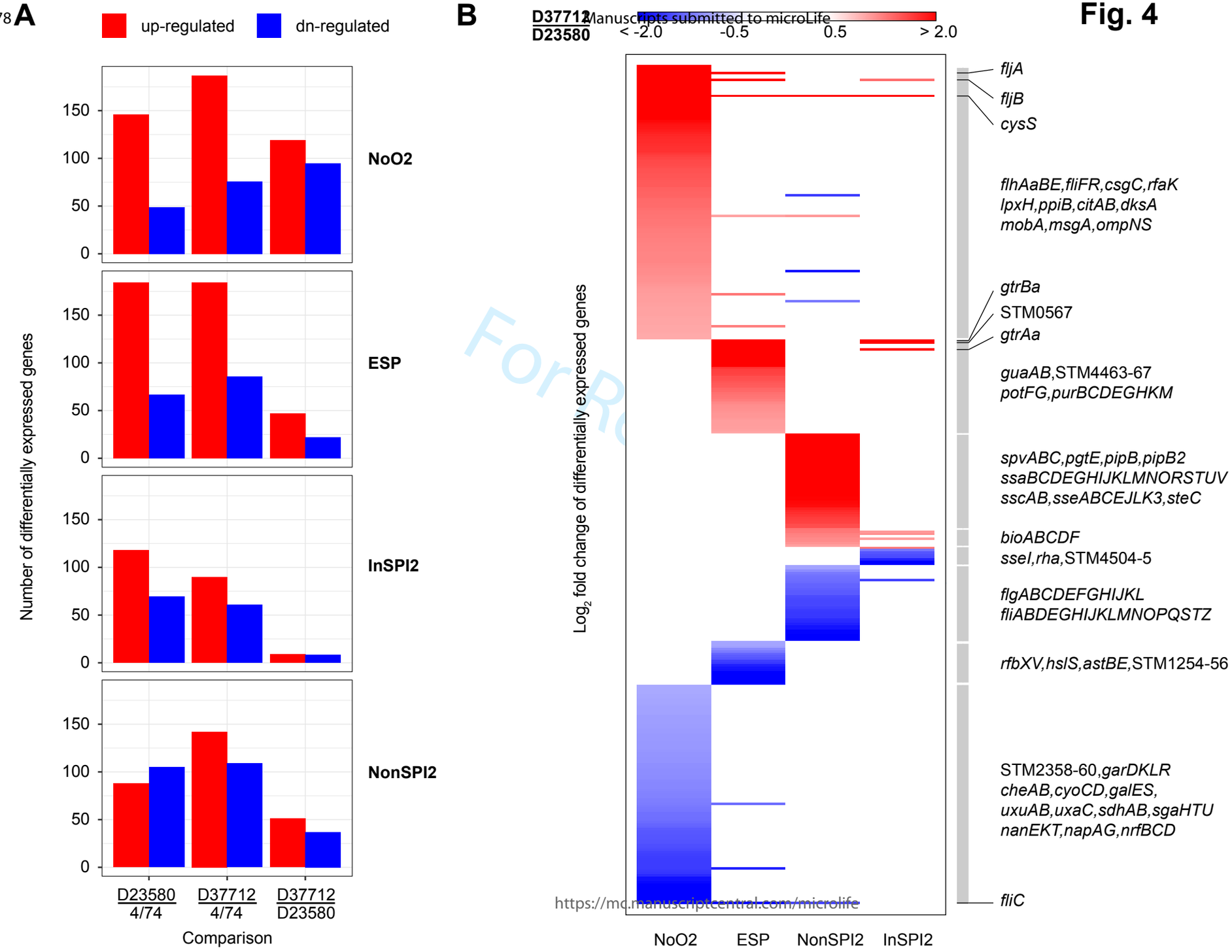


Fig. 5

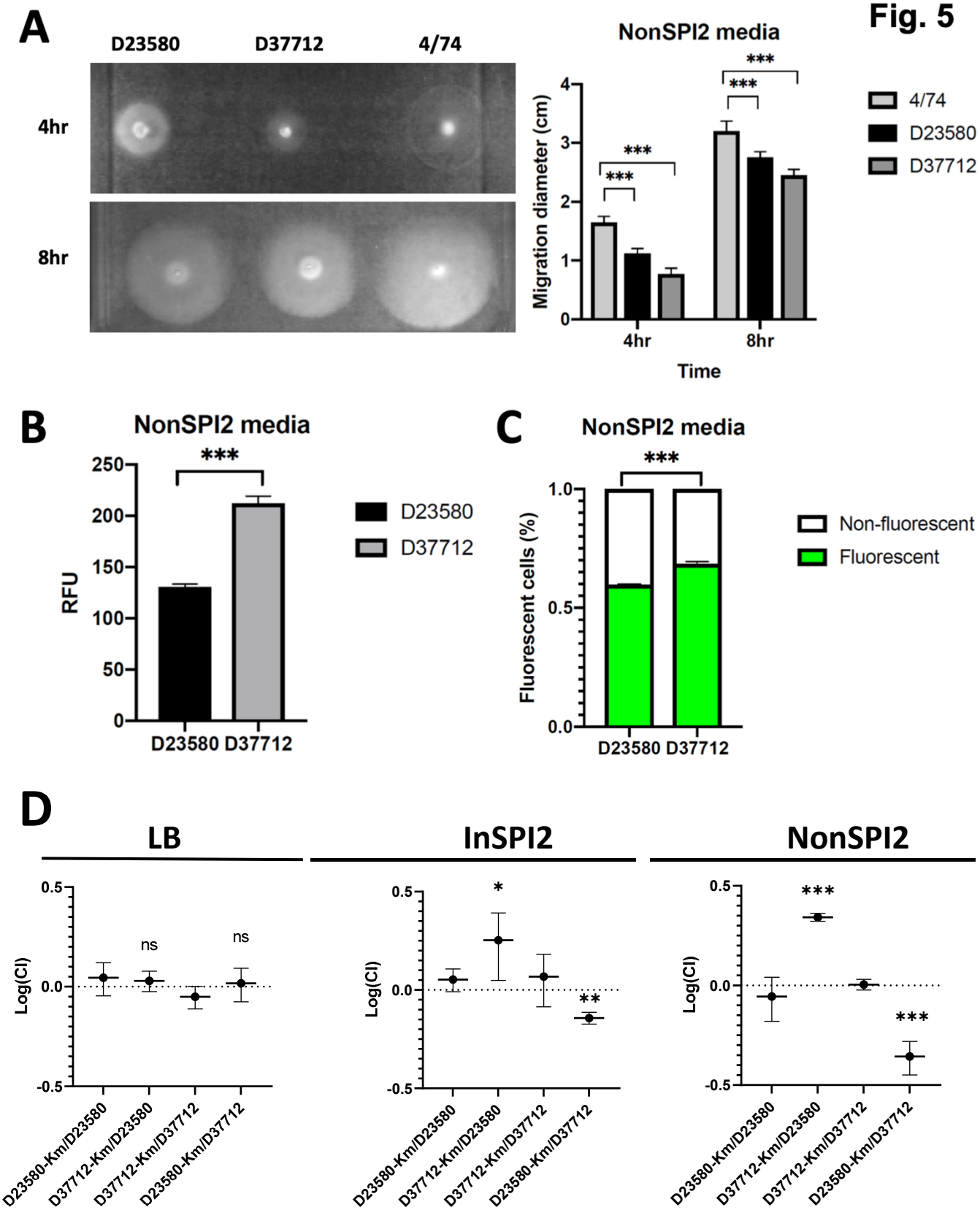

A
D23580
D37712
4/74
4hr
8hr
NonSPI2 media
Migration diameter (cm)
Time
4hr
8hr
4/74
D23580
D37712
B
NonSPI2 media
RFU
D23580 D37712
C
NonSPI2 media
Fluorescent cells (%)
Non-fluorescent
Fluorescent
D23580 D37712
D
LB
InSPI2
NonSPI2
Log(CI)
ns
D23580-Km/D23580
D37712-Km/D23580
D37712-Km/D37712
D23580-Km/D37712

Fig. S1

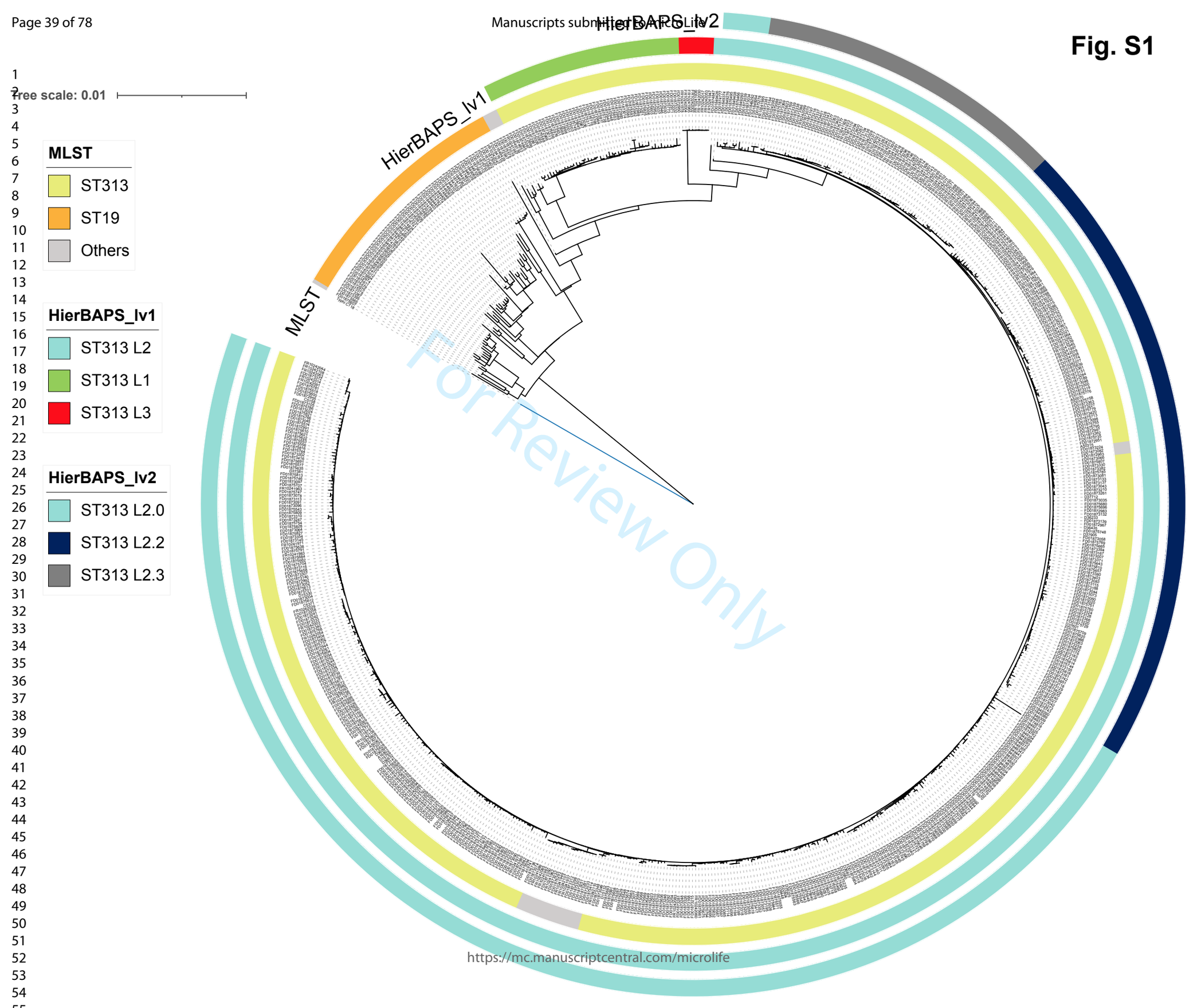


Fig. S2

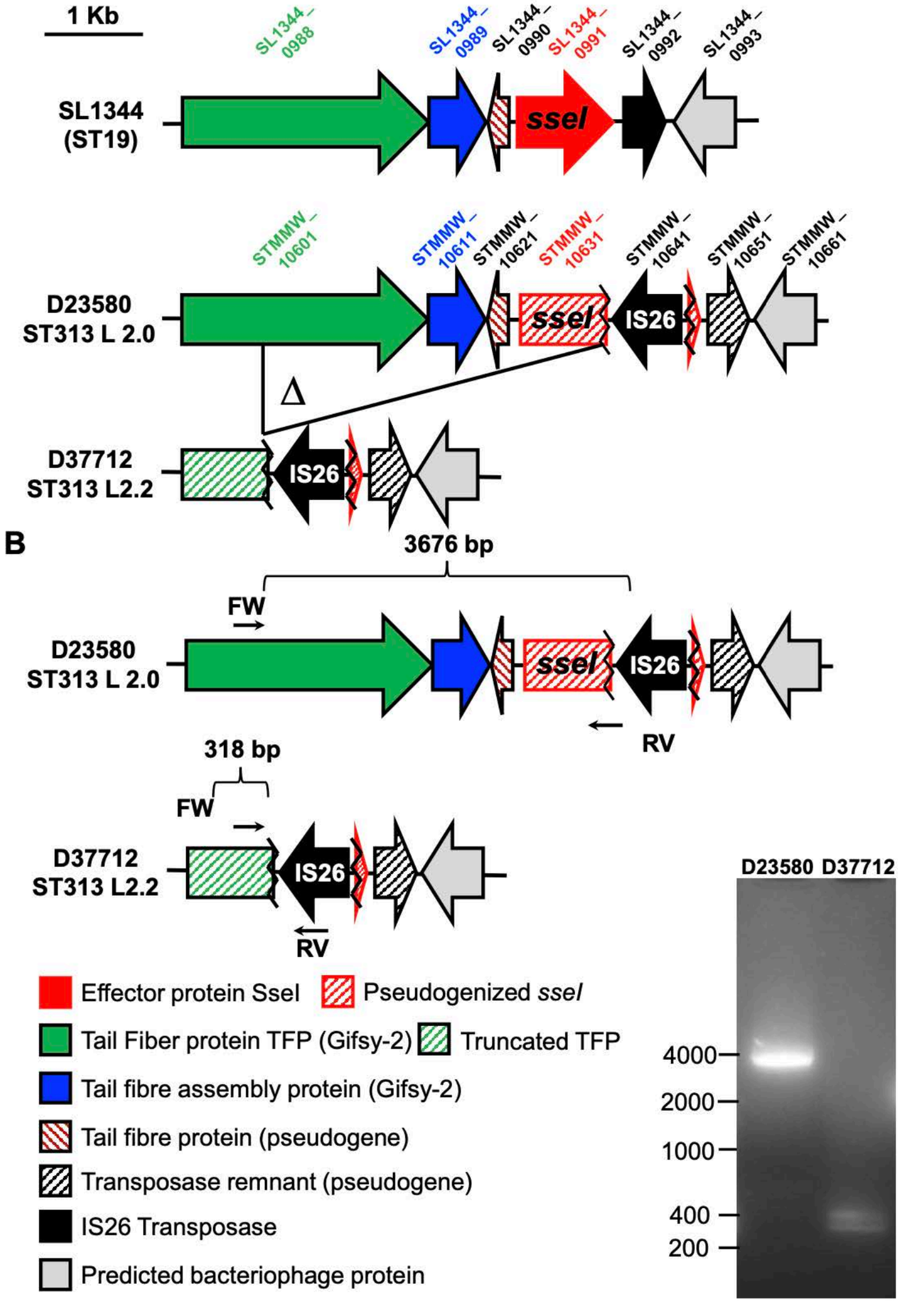

A
1 Kb
SL1344_0988
SL1344_0989
SL1344_0990
SL1344_0991
SL1344_0992
SL1344_0993
SL1344 (ST19)
sseI
STMMW_10601
STMMW_10611
STMMW_10621
STMMW_10631
STMMW_10641
STMMW_10651
STMMW_10661
D23580 ST313 L 2.0
IS26
Δ
D37712 ST313 L2.2
B
3676 bp
FW
RV
318 bp
Effector protein SseI
Pseudogenized sseI
Tail Fiber protein TFP (Gifsy-2)
Truncated TFP
Tail fibre assembly protein (Gifsy-2)
Tail fibre protein (pseudogene)
Transposase remnant (pseudogene)
IS26 Transposase
Predicted bacteriophage protein
D23580 D37712
4000
2000
1000
400
200

Fig. S3

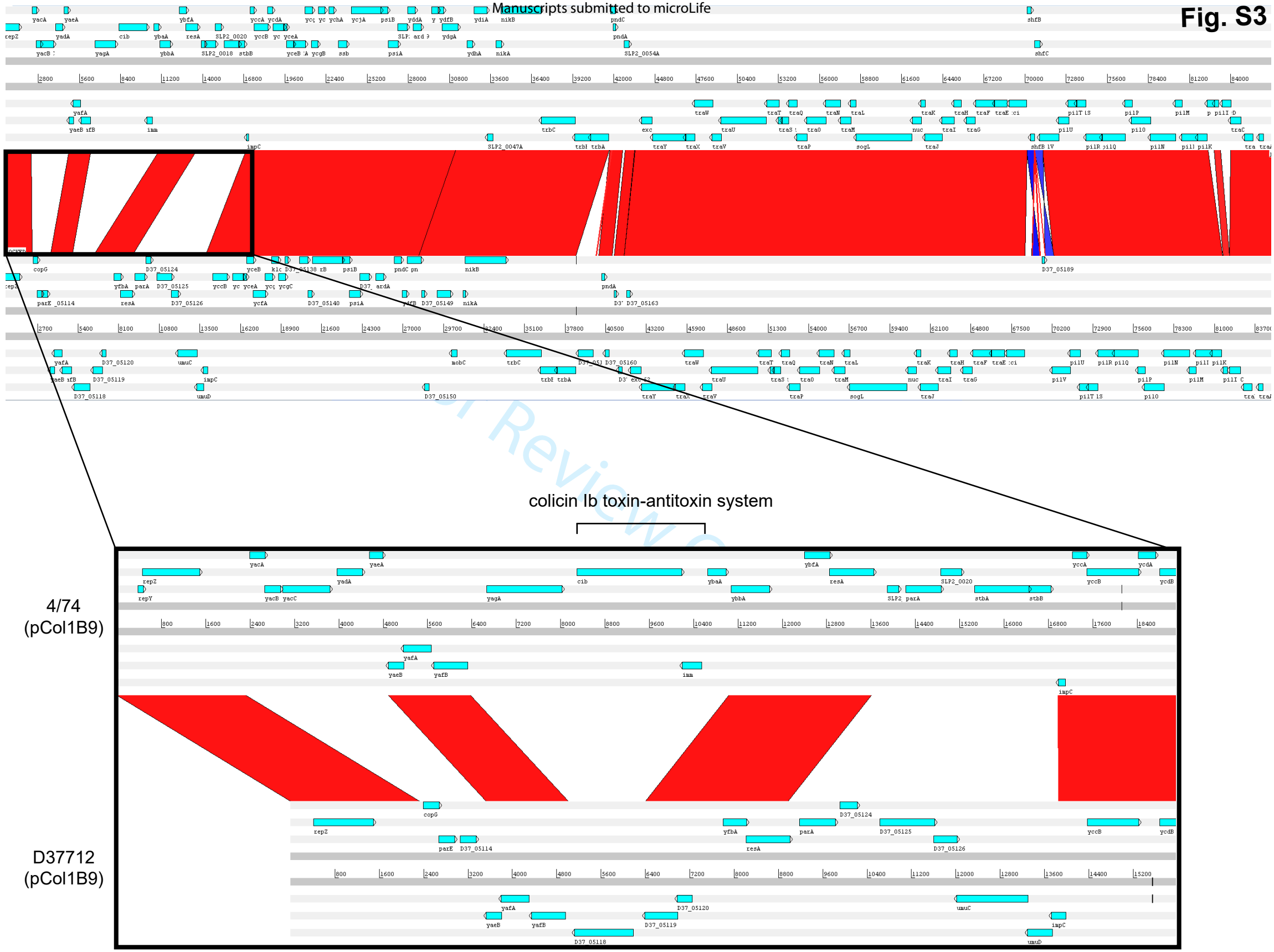

4/74 (pCol1B9)
D37712 (pCol1B9)
colicin Ib toxin-antitoxin system
4/74 (pCol1B9)
D37712 (pCol1B9)
impC-umuC-umuD operon

Fig. S4

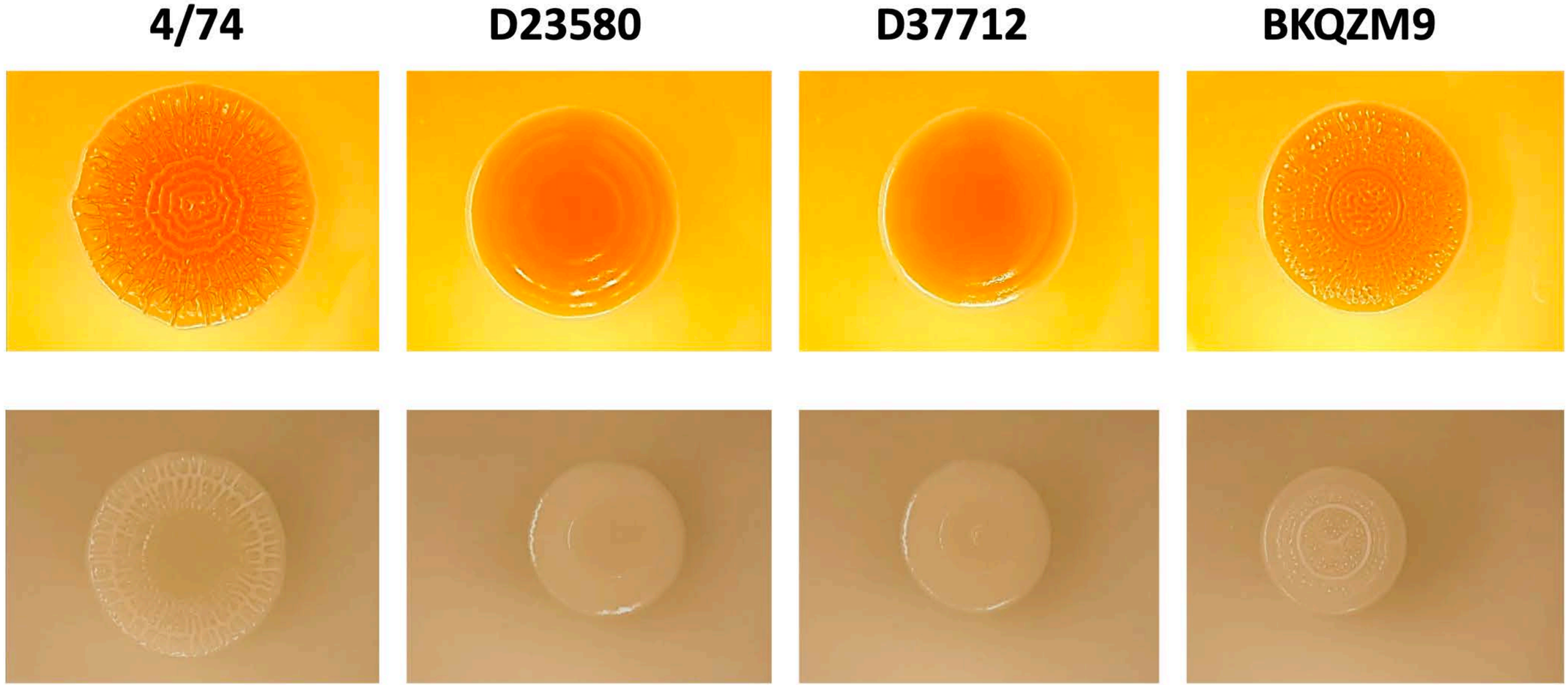
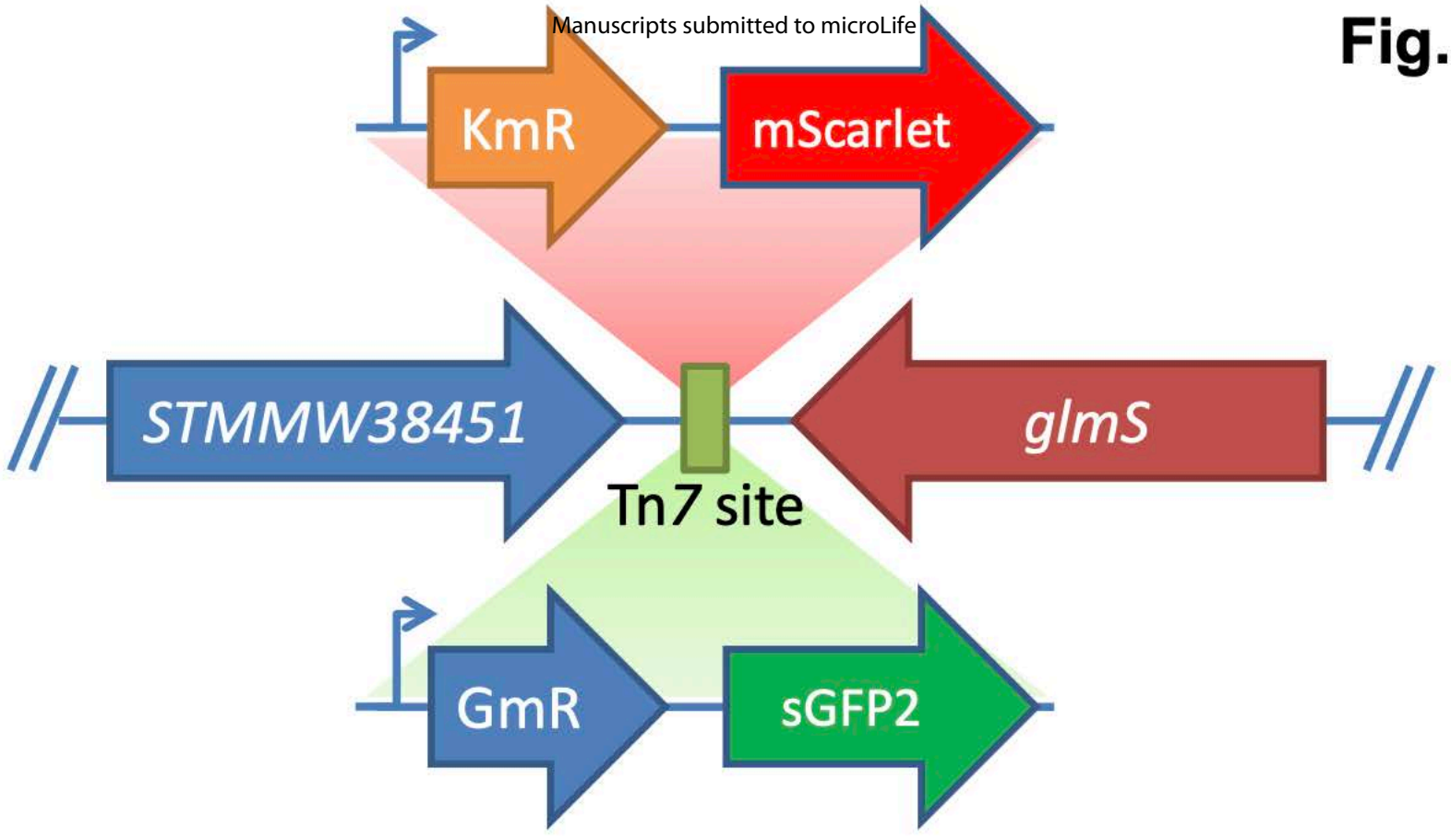
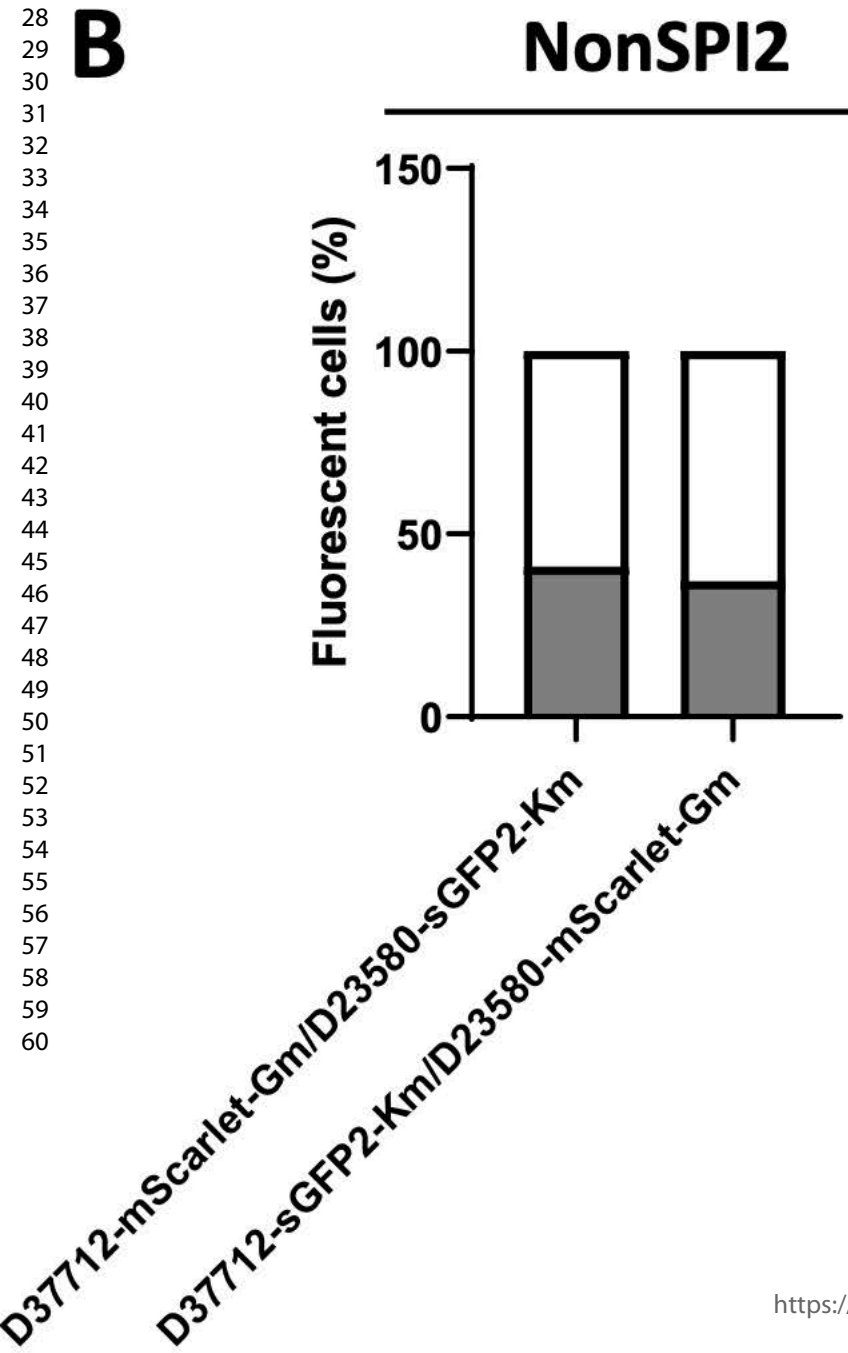


Fig. S5


A
KmR
mScarlet
STMMW38451
glmS
Tn7 site
GmR
sGFP2


B
NonSPI2
Fluorescent cells (%)
150
100
50
0
D37712
D23580
D37712-mScarlet-Gm/D23580-sGFP2-Km
D37712-sGFP2-Km/D23580-mScarlet-Gm

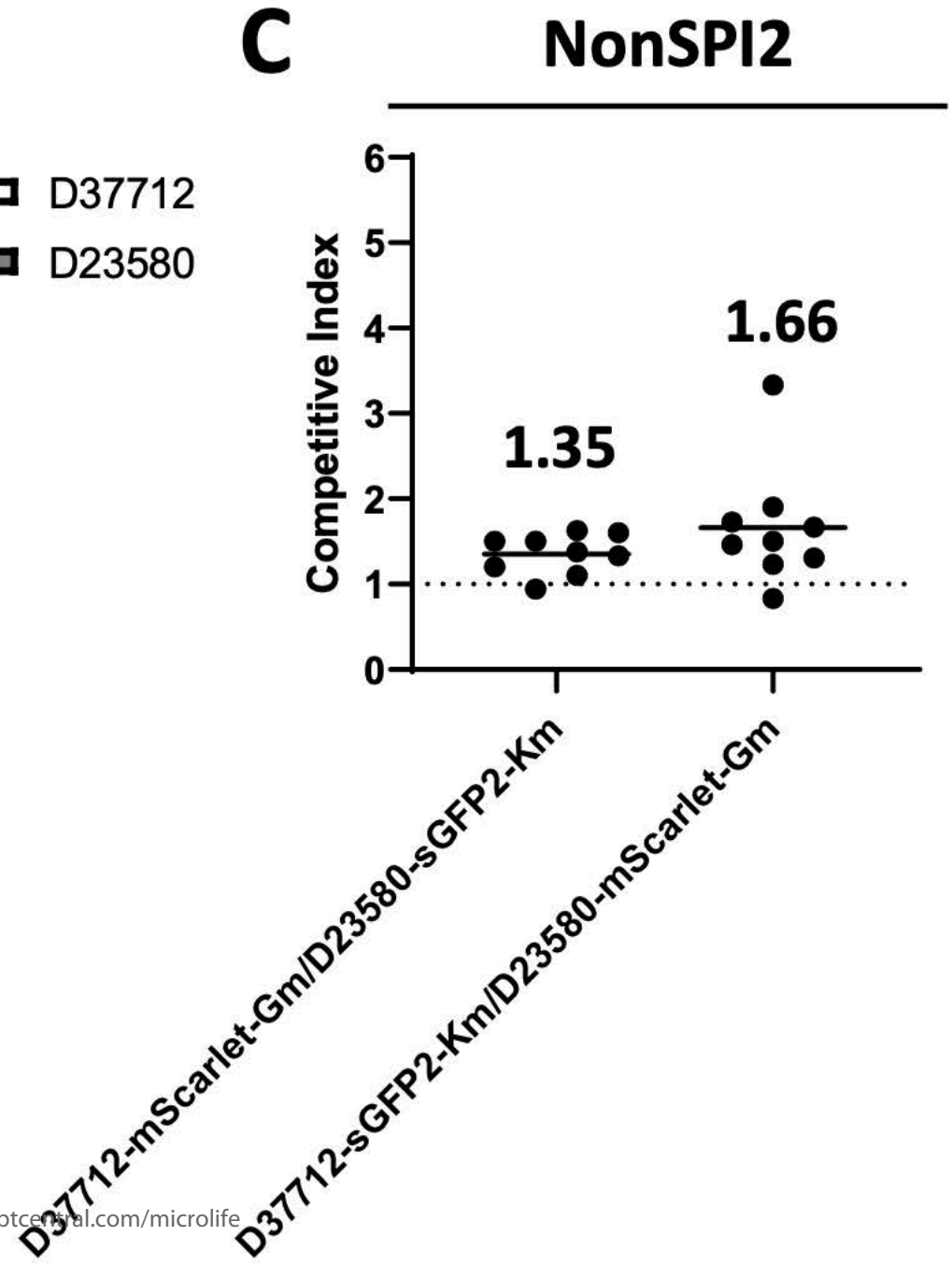

C
NonSPI2
Competitive Index
6
5
4
3
2
1
0
1.35
1.66
D37712-mScarlet-Gm/D23580-sGFP2-Km
D37712-sGFP2-Km/D23580-mScarlet-Gm

Fig. S6

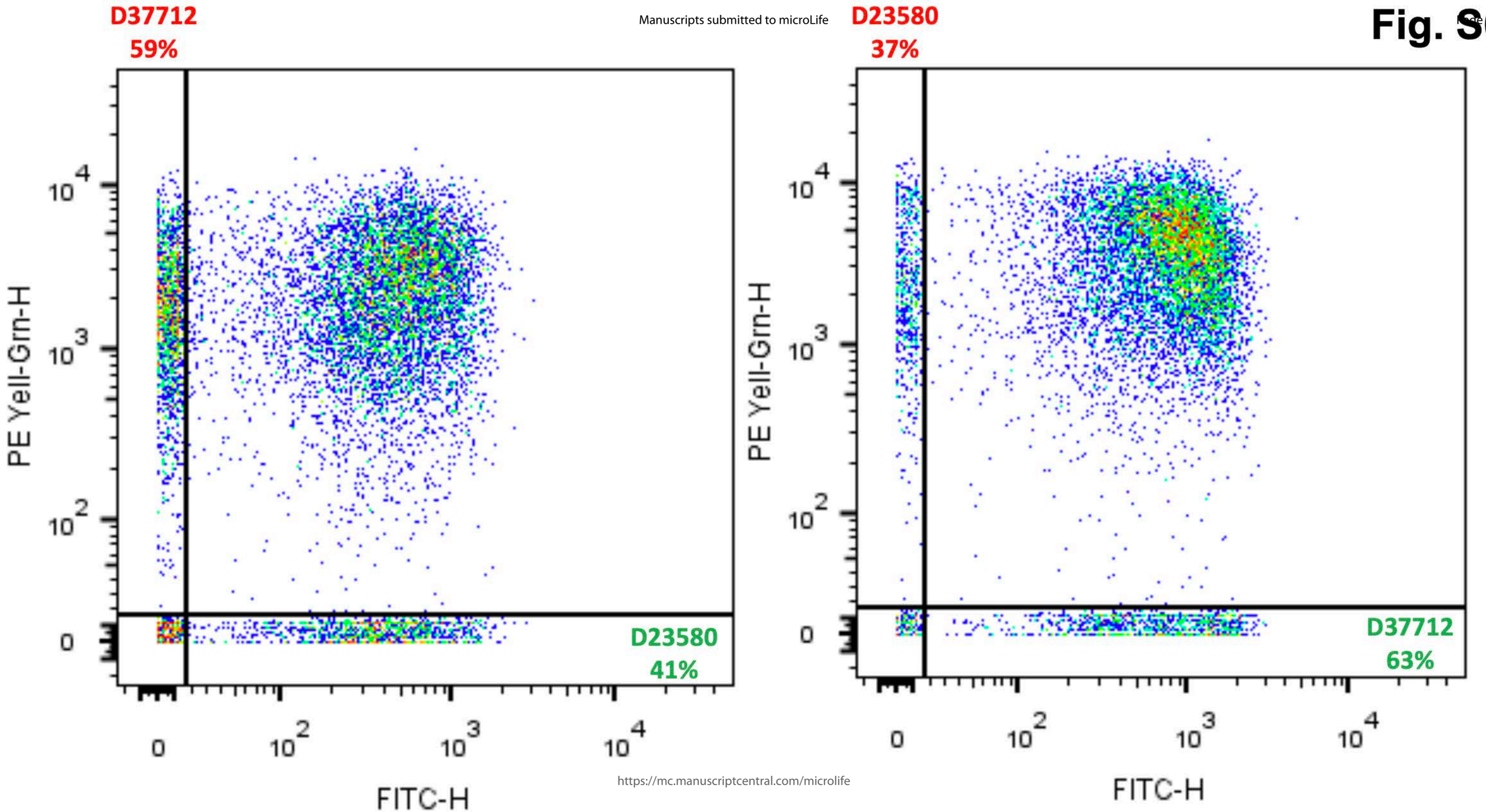
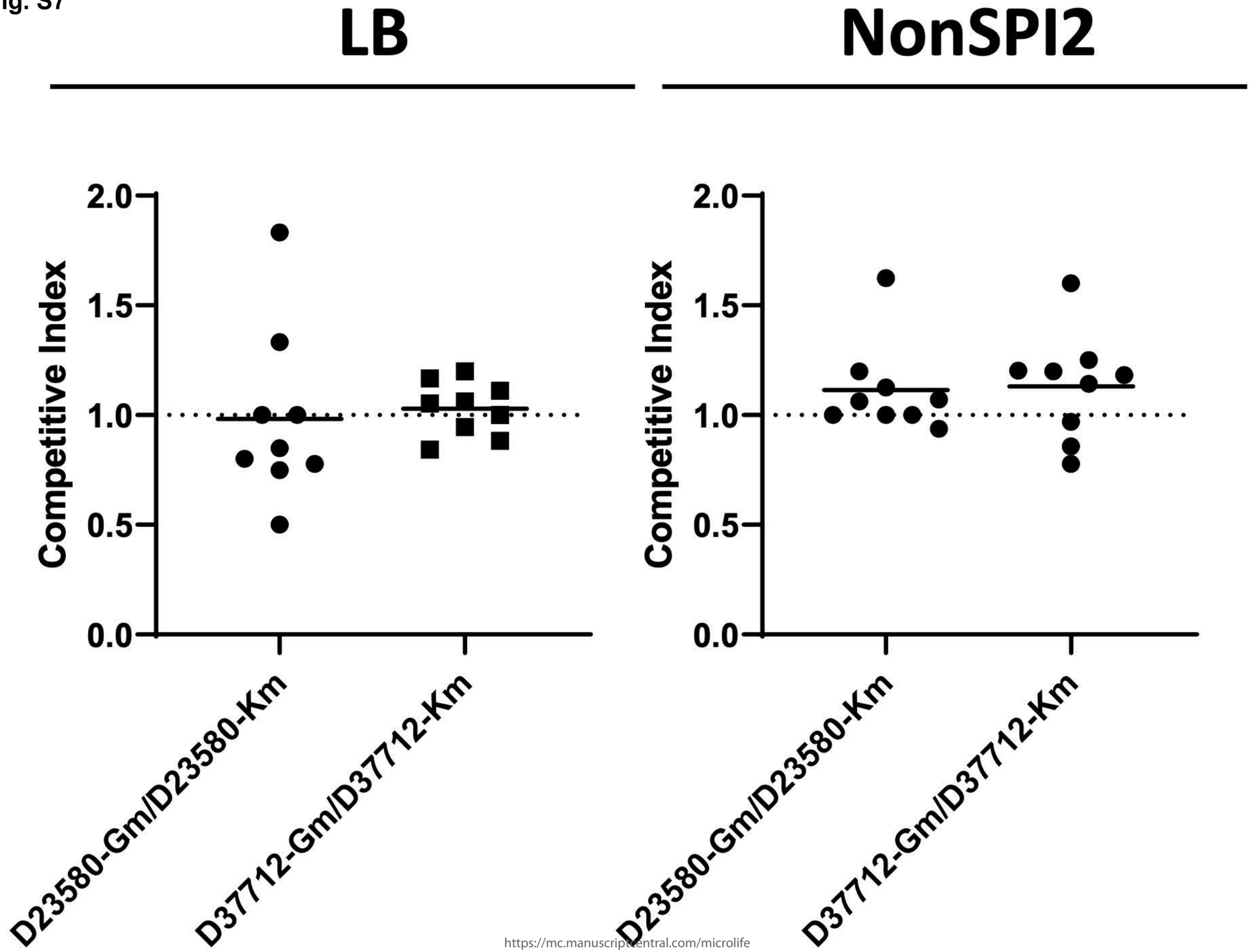

D37712
59%
D23580
41%
PE Yell-Grn-H
FITC-H
0
10^2
10^3
10^4
D23580
37%
D37712
63%

Fig. S7


LB
NonSPI2
Competitive Index
2.0
1.5
1.0
0.5
0.0
D23580-Gm/D23580-Km
D37712-Gm/D37712-Km

Salmonella enterica serovar Typhimurium ST313 sublineage 2.2 has emerged in Malawi with a characteristic gene expression signature and a fitness advantage

Benjamin Kumwenda[1,2,3,#], Rocío Canals[2,#], Alexander V. Predeus[2,#], Xiaojun Zhu[2], Carsten Kröger[2,4], Caisey Pulford[2], Nicolas Wenner[2], Lizeth Lacharme Lora[2], Yan Li[2], Siân V. Owen[2], Dean Everett[5], Karsten Hokamp[6], Robert S. Heyderman[3,7], Philip M. Ashton[3], Melita A. Gordon[2,3], Chisomo L. Msefula[1,3], Jay C. D. Hinton[2,*].

[1]School of Life Sciences and Allied Health Professions, Kamuzu University of Health Sciences Blantyre, Blantyre, MALAWI

[2]Institute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Liverpool, UNITED KINGDOM,

[3]Malawi–Liverpool–Wellcome Programme, Blantyre, MALAWI.

[4]Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin, IRE.

[5]College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, UAE.

[6]Smurfit Institute of Genetics, School of Genetics and Microbiology, Trinity College Dublin, Dublin, IRE.

[7]Research Department of Infection, Division of Infection & Immunity, University College London, UK.

Key words: transcriptomics, comparative genomics, lineage evolution, gene expression, antibiotic resistance

[#]BK, RC and AP contributed equally to this work.

* **Corresponding author:** E-mail: jay.hinton@liverpool.ac.uk; Tel: +441517954573; Institute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Liverpool, United Kingdom.

Abstract

Invasive non-typhoidal *Salmonella* (iNTS) disease is a serious bloodstream infection that targets immune-compromised individuals, and causes significant mortality in sub-Saharan Africa. *Salmonella enterica* serovar Typhimurium ST313 causes the majority of iNTS in Malawi. W, and we performed an intensive comparative genomic analysis of 608 *S.* Typhimurium ST313 isolates dating between 1996 and 2018 obtained from fever surveillance at the Queen Elizabeth Hospital, Blantyre, Malawi between 1996 and 2018. We discovered that following the upsurge arrival of the well-characterised *S.* Typhimurium ST313 lineage 2 from in 1999 onwards, two new multidrug-resistant variants sublineages designated 2.2 and 2.3, emerged in Malawi in 2006 and 2008, designated sublineage 2.2 and 2.3 respectively. The majority of *S.* Typhimurium isolates from human bloodstream infections in Malawi now belong to sublineage 2.2 or 2.3. To identify factors that characterisedunderstand the emergence of the prevalent ST313 sublineage 2.2, we performed genomic and functional analysis ofstudied two representative strains, D23580 (lineage 2) and D37712 (sublineage 2.2). Comparative genomic analysis showed that theThe chromosome of ST313 lineage 2 and sublineage 2.2 were broadly similar, only differingdiffered by 29 SNPs and /small indels and a a 3kb deletion in theof a Gifsy-2 prophage region that spannedincluding the *sseI* pseudogene. LineageL The of lineage 2 and sublineage 2.2 have uniquehad distinctive plasmid profiles that were verified by long read sequencing.. The transcriptome was was initially exploredinvestigated in 15 infection-relevant *in vitro* conditions and within macrophages. Differential gene expression was subsequently investigated in depth in the four most important *in vitro* During growth in physiological conditions. that do not usually trigger *S.* Typhimurium SPI2 gene expression, the SPI2 genes of D37712 were transcriptionally active. We identified up-regulation of SPI2 genes in non-inducing conditions, and down-regulation of flagellar genes in D37712, compared towith D23580. Following phenotypic confirmation of transcriptionaltranscriptomic differences, we discovered that sublineage 2.2 had increased fitness compared with lineage 2 during mixed-growth in minimal media. We speculate that this competitive advantage is contributing to the continuing presenceemergence of sublineage 2.2 in Malawi.

Introduction

Non-typhoidal *Salmonella* (NTS) is a ~~major~~key bacterial pathogen that threatens people across the world. Typhimurium and Enteritidis are the two serovars of *Salmonella enterica* ~~that cause~~responsible for the highest levels of self-limiting gastrointestinal disease in Europe, the USA and other high-income countries (Zhang *et al.*, 2003). In the industrialised world, NTS has ~~largely~~ been associated with intensive food production, animal husbandry, and global distribution systems (Majowicz *et al.*, 2010). ~~Globally, the most common sequence type of~~The *S.* Typhimurium ~~associated with~~sequence types responsible for gastroenteritis ~~is~~ globally include ST19~~. Diarrhoeal~~, ST34 and monophasic 1,4,[5],12:i:- variants (Branchu *et al.*, 2018). The diarrhoeal NTS disease ~~(~~is termed dNTS~~)~~, and is mainly foodborne ~~and poses~~, posing a significant burden to public health ~~globally, causing~~with approximately 153 million cases and 57,000 deaths per annum worldwide (Kirk *et al.*, 2015; Chirwa *et al.*, 2023).

In contrast, a lethal systemic disease called invasive non-typhoidal Salmonellosis (iNTS) has emerged in recent decades in low- and middle-income countries in sub-Saharan Africa. ~~Cases of~~ iNTS ~~are characterized by bloodstream infections of immune-compromised~~targets immunocompromised individuals such as adults with HIV, and children under five years of age~~, and HIV-positive adults. Anaemia~~ with malaria, malnutrition ~~and malaria are some of the major risk factors~~or severe anaemia (Feasey *et al.*, 2012). In some countries of sub-Saharan Africa, *Salmonella* causes more cases of community-onset bloodstream infections than any other bacterial pathogen (Marchello *et al.*, 2019). In 2017, 535,000 cases of iNTS disease were estimated worldwide, with about 80% of cases and 77,000 deaths occurring in sub-Saharan Africa (Stanaway *et al.*, 2019)

Clinically, the treatment of iNTS is complicated by multi-drug (MDR) resistance which limits therapeutic options (Crump *et al.*, 2015). Widespread resistance of iNTS pathogens to first-line drugs such as chloramphenicol, ampicillin and cotrimoxazole has been seen in many countries (Kariuki *et al.*, 2006). This MDR phenotype may be one of the reasons the case fatality rate associated with iNTS is amongst the highest in comparison to any infectious disease (15%) (Marchello *et al.,* 2022~~)~~). Resistance to second~~-~~line drugs such as ceftriaxone, ciprofloxacin and azithromycin has been reported in a few African countries (Tack *et al.*, 2020). Clearly, the ~~problem of~~challenge posed by MDR *Salmonella* must be addressed urgently (Gilchrist and MacLennan, 2019).

The African iNTS epidemic is mainly caused by two *Salmonella* pathovariants, *S.* Typhimurium sequence type 313 (ST313) and specific clades of *S.* Enteritidis (Kingsley *et al.*, 2009; Okoro *et al.*, 2012; Feasey *et al.*, 2016). *S.* Typhimurium ST313 is responsible for about two-thirds of clinical iNTS cases that have been reported in Africa (Gilchrist and MacLennan, 2019).

It is not certain how these pathogens are transmitted, but there is increasing evidence from case-control studies that ST313 strains are human-associated but not animal-associated

within households (Post *et al.*, 2019; Koolman *et al.*, 2022). A recent summary concludes that the available data are consistent with ~~the~~iNTS disease being transmitted person-to-person ~~transmission hypothesis for iNTS disease~~ (Chirwa *et al.*, 2023). Global efforts to combat iNTS infections are currently focused on vaccine development, which ~~is currently progressing~~has now progressed to Phase 1 clinical trials (Piccini and Montomoli, 2020; Skidmore *et al.*, 2023).

Since 1998, continuous sentinel surveillance for fever and bloodstream infections among adults and children has been undertaken at Queen Elizabeth Central Hospital (QECH). This tertiary referral hospital in Blantyre, Malawi, serves an urban population of about 920,000 with a high incidence of malaria, HIV and malnutrition (Musicha *et al.*, 2017). Following blood-culture of samples collected from patients of all ages presenting with fever, whole genome sequencing identified the ST313 variant of *S.* Typhimurium (Kingsley *et al.*, 2009). Phylogenetic analysis revealed that the chloramphenicol-sensitive ST313 lineage ~~1was~~1 was clonally-replaced in Malawi by the chloramphenicol-resistant lineage 2 (Okoro *et al.*, 2012). More recently, a ST313 sublineage II.1 (2.1) emerged from lineage 2 in Democratic Republic of Congo (DRC) in Central Africa. Sublineage 2.1 had altered phenotypic properties including biofilm formation and metabolic capacity and resistance to azithromycin (Van Puyvelde *et al.*, 2019). An elegant genomic analysis that provides insight regarding the diversity of S. Typhimurium ST19 clades in the context of ST313 lineage 2 clades is also available (Van Puyvelde *et al.*, 2023).

~~An~~The initial suggestion that ST313 lineage 2 was undergoing evolutionary change in East Africa came from a small study that identified ~~seven~~several *S.* Typhimurium ST313 Malawian isolates, dated between 2006 and 2008, that differed from lineage 2 by 22 core-genome single nucleotide polymorphisms (SNPs) (Msefula *et al.*, 2012).

To ~~begin to~~ examine the evolutionary trajectory of *S.* Typhimurium in Malawi at a large scale, we conducted a comparative genomic analysis study focused on 680 isolates dating between ~~1998~~1996 and 2018 (Pulford *et al.*, 2021). We previously ~~confirmed~~reported that ST313 lineage 1 (L1) was replaced by lineage 2 (here designated L2.0), and discovered an antibiotic-sensitive lineage 3 (L3) that emerged in 2016 (Pulford *et al.*, 2021).

We have now performed a more intensive phylogenetic analysis of the same collection of *S.* Typhimurium ST313 isolates, most of which caused bloodstream infections in Malawi over two decades. We discovered two novel sublineages named 2.2 (L2.2) and 2.3 (L2.3) that emerged 2006 - 2008, and have been replacing L2.0 ~~since 2006~~.

Here we present a comprehensive comparative genomic analysis of the most prevalent ST313 L2.2 sublineage, and report the results of a functional genomic approach that identified key phenotypic characteristics that distinguish L2.2 from L2.0.

Results

Identification of *S.* Typhimurium ST313 sublineages 2.2 and 2.3 in Malawi

~~The *S.* Typhimurium ST313 L2 (Lineage II) was originally identified as the major cause of iNTS cases across sub-Saharan Africa in the early 2000's (Kingsley *et al.*, 2009; Okoro *et al.*, 2012) (Okoro *et al.*, 2015). Subsequently, an azithromycin-resistant variant of *S.* Typhimurium ST313 was found in a single country, the Democratic Republic of Congo between 2008 and 2016, and was designated sublineage L2.1 (Van Puyvelde *et al.*, 2019).~~

The emergence of the ST313 lineage 2 genotype in Malawi in 2002 prompted us to hypothesise that subsequent evolution would select for variants with increased fitness, leading to the clonal expansion of one or more sublineages by outcompeting previously dominant genotypes. We investigated this hypothesis by conducting a detailed core-gene SNP-based maximum likelihood (ML) phylogenetic analysis to investigate the population structure of *S.* Typhimurium ST313 L2.0 (Fig. S1). As well as identifying members of the antibiotic-sensitive lineage 3,~~To investigate the evolutionary dynamics of *S.* Typhimurium ST313 L2 in Malawi over a 22 year period, we focused on the large collection of 8,000 *S.* Typhimurium isolates derived from bloodstream infection in hospitalised patients at the Queen Elizabeth Central Hospital, Blantyre, Malawi (Feasey *et al.*, 2015). The collection was assembled by the Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW) between 1996 and 2018; the precise annual numbers of isolates are shown in Fig 1B. A random sub-sampling strategy was used to select 608 isolates selected for whole-genome sequencing which included 549 *S.* Typhimurium ST313 isolates (Pulford *et al.*, 2021).~~

~~Here, we used a core-gene SNP-based maximum likelihood (ML) phylogenetic tree to investigate the population structure of *S.* Typhimurium ST313 L2.0 in more detail (Fig. S1). As well as identifying members of the antibiotic-sensitive lineage 3 that we~~ reported previously (Pulford *et al.*, 2021), we discovered that ST313 L2 ~~could be split into~~comprised three phylogenetically-distinct sublineages that differed by a total of 39 SNPs. The *S.* Typhimurium ST313~~ L2~~ reference strain D23580 (Kingsley *et al.*, 2009) ~~belonged~~belongs to the ~~first sublineage, which we have now designated as~~ ST313 L2.0 lineage (Fig~~ 1C~~. 1A). As ST313 sublineage L2.1 ~~has~~had been defined previously (Van Puyvelde *et al.*, 2019), the new sublineages ~~were designated as L2.2 and L2.3, and~~which belonged to different hierBAPS level 2 clusters were designated L2.2 and L2.3 (Fig~~ 1C~~. 1A and Fig. S1). ~~We~~In total, we identified 151 L2.2 isolates~~ and,~~ 74 L2.3 isolates~~, against a backdrop of~~ and 350 L2.0 isolates.

In Blantyre, Malawi, *S.* Typhimurium ST313 L2.2 was first detected in 2006, and L2.3 was initially observed in 2008 (Fig. ~~1A.~~1BC). Both L2.2 and L2.3 increased in prevalence at the Queen Elizabeth Central Hospital in Blantyre in subsequent years. By 2018, L2.2 and L2.3 had largely replaced L2.0 (Fig~~ 1A-B~~. 1BC). Our published Bayesian (BEAST) analysis (Pulford *et al.*, 2021) estimated that the Most Recent Common Ancestor (MRCA) of ST313 lineage 2 dates back to 1948 (95% HPD = 1929-1959).

To understand the accessory gene complement of L2.2 and L2.3, we compared the genomes of seven L2.2 isolates and four L2.3 isolates with 17 L2.0 isolates, ST313 L1 and ST19 ~~(~~and the results are shown in Fig~~ 1C,~~. 1A and Table S1~~).~~. *S.* Typhimurium strain D23580 is the representative strain of L2.0 (Kingsley *et al.*, 2009), for which we previously used long-read sequencing and other approaches to thoroughly characterise the ~~chromosome~~chromosomal and ~~the~~ plasmid complement (Canals *et al.*, 2019b).

Antimicrobial Resistance

~~AMR~~MDR variants of *S.* Typhimurium with resistance to ampicillin and cotrimoxazole were detected at an early stage of the iNTS epidemic, from 1997 onwards (Gordon *et al.,* 2008). Multidrug-resistant variants of *S.* Typhimurium ST313 that were no longer susceptible to chloramphenicol, ampicillin and cotrimoxazole subsequently emerged in Malawi (Gordon *et al.,* 2008) and have been reported elsewhere in sub-Saharan Africa by the GEMS study (Kasumba *et al.*, 2021). The *S.* Typhimurium ST313 L2.0, L2.2 and L2.3 isolates shared the same MDR profiles (resistance to chloramphenicol, ampicillin and cotrimoxazole), and carried identical IS21-~~AMR~~associated antimicrobial gene cassettes within the pSLT-BT plasmid~~.~~ (Fig. 2B).

Comparative genomics of *S.* Typhimurium ST313 sublineage 2.2

Because *S.* Typhimurium ST313 L2.2 was the predominant novel sublineage in Blantyre, Malawi in 2018, we focused on L2.2 for the remainder of this study. We used the phylogeny (Fig ~~1C.~~ 1A) to select strain D37712 as a representative ~~isolate~~ of L2.2. D37712 was isolated from the blood of an HIV-positive Malawian male child and has been deposited in the National Collection of Type Cultures ~~(NCTC). The initial~~as NCTC 14678. The draft genome sequence of D37712 was obtained in 2012 with Illumina technology, an assembly that comprised 27 individual contigs (Msefula *et al.*, 2012). To generate a reference-quality genome, we resequenced D37712 with both long-read PacBio and Illumina short-read technologies. Our hybrid strategy generated a complete genome assembly that included one circular chromosome and three plasmids (see Materials & Methods; GenBank CP060165, CP060166, CP060167 and CP060168). This high-quality genome sequence allowed us to conduct a detailed ~~comparative genomic analysis~~comparison between the genomes of L2.2 strain D37712 ~~with~~and L2.0 strain D23580 (accession number FN424405), summarised in Fig. 2 and Table S2.

Overall, the gene content of the two strains ~~contain a similar number of genes. The D37712 and D23580 genomes shared 5,016 orthologous genes, including 4,729 protein-coding genes and pseudogenes as well as the 287 small RNA (sRNA) genes that we identified previously.~~was largely equivalent. The D23580 annotation contains 4,823 protein-coding and

pseudogenes and 287 ~~sRNAs~~small RNA (sRNA) genes that we identified previously (Canals *et al.,* 2019b), while D37712 contains 4,821 protein-coding and pseudogenes and ~~287 sRNAs~~the same 287 sRNAs. In total, the D37712 and D23580 genomes shared 4,729 orthologous protein-coding genes and pseudogenes. The 104 protein genes that differ are encoded by the pSLT ^{D37712}, pBT1^{D37712}, and pCol1B9^{D37712} plasmids.

Overview of D23580 and D37712 genomes

The chromosomes of D23580 and D37712 are 4,879,402 and 4,876,060 bp, respectively, ~~and similar in~~about the same size ~~to~~as other S. Typhimurium genomes (Kingsley *et al.*, 2009; Branchu *et al.*, 2018). The D23580 and D37712 strains share ~~a similar~~an identical prophage profile, with both strains carrying five prophages (BTP1, Gifsy-2, ST64B, Gifsy-1, and BTP5) ~~which were located at the same positions on the chromosome. Previously, we have established that just one of these prophages, BTP1, is functional (Owen *et al.*, 2017). The BTP1 prophage of D23580 encodes the novel BstA phage defence system (Owen *et al.*, 2021) and a particularly high level of viable BTP1 phages is produced by spontaneous induction (Owen *et al.*, 2017).~~located at the same positions on the chromosome (Fig. 2A).

Comparison of D23580 and D37712 chromosomes

The detailed genomic comparison of D37712 with D23580 showed that the ~~two genomes were remarkably similar.~~sizes of the two chromosomes varied by only 3,342 bp. Overall, the only differences between the genomes of the L2.0 and L2.2 strains were 26 chromosomal SNPs and small indels, plus one large deletion, and an inversion of the *hin* switch. In-depth annotation of the nucleotide variants identified 3 putative loss-of-function mutations (2 stop mutations, 1 frameshift insertion), 1 disruptive in-frame deletion, 4 synonymous mutations, 13 missense mutations, and 5 intergenic variants, summarised in Fig ~~2A.~~. 2A. None of the SNP differences that distinguished D37712 from D23580 were located within 150 nucleotides of a Transcriptional Start Site (Canals *et al.,* 2019b), and so would not be predicted to modulate gene expression.

The 3,358 bp-long deletion of a Gifsy-2 prophage-associated region that spanned the *sseI* pseudogene of D23580 (STMMW_10631) removed two coding sequences (STM1050-51; STMMW_10611-STMMW_10631), and substantially truncated the STM1049 (STMMW_10601) gene (Fig. 2E). The *sseI* gene encodes a cysteine hydrolase effector protein that modulates the directional migration of dendritic cells during systemic infection (Brink *et al.*, 2018). In strain D23580, the insertion of ~~a~~an IS26 transposable element ~~IS15DEV~~ inactivated the *sseI* gene (Kingsley *et al.*, 2009), causing increased dendritic cell-mediated dissemination of strain D23580 during infection (Carden *et al.*, 2017). ~~To~~We used an independent PCR-based approach to confirm that the 3,358 bp deletion had removed the *sseI* gene from the chromosome of strain D37712~~, we used an independent PCR-based approach~~ (Fig. S2).

Comparison of D23580 and D37712 plasmids

Here we put the genetic features of the representative strains for ST313 L2.0 and L2.2 into context with other isolates belonging to the Lineage 2 sublineages. ST313 L2.0 strain D23580 carries four plasmids, pSLT-BT, pBT1, pBT2 and pBT3 (Kingsley *et al.*, 2009). In contrast, ST313 L2.2 ~~carried~~has a distinct plasmid complement (Fig ~~1C~~. 1A, Fig. 2BCD). The plasmid profiles of D23580 and D37712 were confirmed by a combination of Illumina (short-read) and PacBio (long-read) sequencing (Materials and Methods).

In summary, strain D37712 carried the pSLT-BT, pBT2 and pCol1B9 plasmids as detailed below. Both D23580 and D37712 strains ~~had~~carried a variant of the pSLT-BT virulence plasmid (Kingsley *et al.*, 2009) that contains a Tn21-like transposable element with five antibiotic resistance genes. The D37712 version of pSLT-BT ~~is similar to that~~only differs from the pSLT-BT of D23580~~, with~~ in two important ~~differences~~t ways (Fig. 2B). Firstly, the Tn21-like element is inserted in the opposite direction with regards to the rest of the plasmid, suggesting that the transposable element remains active. Secondly, three nucleotide variants were identified in the pSLT-BT ~~variant~~carried by D37712, two deletions in noncoding regions, and one frameshift insertion that generates a pseudogene of *spvD*. ~~The SpvD effector protein, a cysteine protease, is translocated by the SPI2 type 3 secretion system and suppresses the NF-κB-mediated pro-inflammatory immune response and contributes to virulence in mice (Grabe *et al.*, 2016).~~

Plasmid pCol1B9 was of particular interest because it was absent from D23580, but ~~is~~was present in *S.* Typhimurium ST19 strain 4/74 (Richardson *et al.*, 2011; Fig. 1A). 4/74 is the parent of ~~the~~ *S.* Typhimurium SL1344, a strain that has been used extensively for the study of *S.* Typhimurium pathogenesis and gene regulation ~~since 1986~~in recent decades (Kröger *et al.*, 2012; Rankin & Taylor, -1966). Our new annotation of the pCol1B9-like plasmid ~~included~~identified 95 distinct protein-coding genes, while the previously published annotation of pCol1B9$^{4/74}$ assigned 101 protein-coding genes. Some of these represent annotation discrepancies, while others represent true genetic differences (Fig. S3). ~~Upon~~

Following careful examination, we identified 14 genes ~~were~~ unique to pCol1B9^{D37712}, ~~while~~and 20 ~~were~~genes unique to pCol1B9$^{4/74}$. There were 81 genes carried by both plasmids. Interestingly, pCol1B9^{D37712} lacked the colicin toxin-antitoxin system that both gave pCol1B9 its name, and provides *Salmonella* with a competitive advantage in the gut (Nedialkova *et al.*, 2014). The pCol1B9^{D37712} plasmid carried a locus that was absent from pCol1B9$^{4/74}$, namely the *impC-umuCD* operon (Fig. S3) which encodes the error-prone DNA polymerase V responsible for the increased mutation rate linked to the SOS stress response in *E. coli* (Sikand *et al.*, 2021).

~~An 85 kb plasmid carried by D23580, pBT1, was previously shown by our laboratory to play an important role in *Salmonella* biology by encoding an orthologous *cysS* gene responsible for expressing the essential cysteinyl tRNA-synthetase enzyme (Canals *et al.*, 2019b). This pBT1 plasmid was completely absent from D37712, and from all isolates of sublineage L2.2 that were examined (Fig. 1C).~~

Comparison of pseudogene status of D23580 and D37712

Our comparative genomic analysis focused on the pseudogenes found in strains 4/74, D23580, and D37712 (Fig. 2F, Table S3). The pseudogenisation of several D23580 genes, compared with strain 4/74, have been linked to the invasive phenotype of African *Salmonella* ST313 (Kingsley *et al.*, 2009). We found that the pseudogene complement of D23580 was largely conserved in D37712~~.~~, consistent with inheritance from a common ancestor. We have recently reported the role of the MacAB-TolC macrolide efflux pump in the virulence of *S.* Typhimurium ST313, and showed experimentally that *macB* was an inactive pseudogene in D23580 (Honeycutt *et al.*, 2020). Interestingly, the *macB* gene is functional in D37712. Compared with D23580, three additional D37712 genes were pseudogenised (*spvD, yadE,* and STMMW_42692~~).~~, as detailed in Table S3). YadE is a predicted polysaccharide deacetylase lipoprotein. The functional impact of these pseudogenes on L2.2 remains to be established.

Overall the chromosomes of ST313 lineage 2 and sublineage 2.2 were highly-conserved and differed by just 29 SNPs/ small indels, and a 3kb deletion in the Gifsy-2 prophage region. The ST313 lineage 2 and sublineage 2.2 have distinct plasmid profiles.

Transcriptional landscape of S. Typhimurium ST313 sublineage L2.2

Previously, we characterized the primary transcriptome of two other *S.* Typhimurium strains, 4/74 and D23580, using a combination of multi-condition RNA-seq and differential RNA-seq (dRNA-seq) techniques (Canals *et al*., 2019b; Kröger *et al.*, 2013). To identify the transcriptional start sites (TSS) of strain D37712, we analysed a pooled sample containing RNA from 15 *in vitro* conditions by dRNA-seq and RNA-seq as detailed previously (Kröger *et al.*, 2013). The high similarity between the D23580 and D37712 chromosomes allowed us to map the curated set of TSS that were previously defined for D23580 (~~Hammarlof~~Hammarlöf *et al.,* 2018) onto a combined D37712/D23580 reference genome. To allow individual TSS to be examined in particular chromosomal or plasmid regions, data from both the dRNA-seq and pooled RNA-seq experiments can be visualised in our online genome browser (**http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data**).

Preliminary gene expression profiling of S. Typhimurium ST313 sublineage L2.2

Given the high level of similarity between the genomes of L2.2 and L2.0, we went on to identify differences at the transcriptional level. We performed a multi-condition RNA-seq-based transcriptomic analysis of gene expression profiles of L2.2 strain D37712 without biological replicates.

This comparative transcriptomic screen was based on our published approach (Canals *et al*., 2019b). Specifically, we used 15 individual infection-relevant *in vitro* conditions (Kröger *et al.*, 2013) and did intra-macrophage transcriptome profiling using the protocol previously

established for *S.* Typhimurium ST19 (Srikumar *et al.*, 2015). The RNA-seq samples were mapped to a combined reference genome, which included the annotated D23580 chromosome (Canals *et al.*, 2019b), as well as all the plasmids described earlier (pSLT-BT, pBT1, pBT3 and pCol1B9; see Methods). The initial RNA-seq assessment (detailed in Methods) involved 2-4M non-rRNA/tRNA reads per sample, allowing gene signatures specific for each *in vitro* condition to be identified. Although single replicate RNA-seq experiments of this type cannot be used for statistically-robust differential gene expression analysis, they do provide a useful screening approach for identifying growth conditions to be used for follow-up experiments. The individual RNA-seq experiments showed broad condition-specific similarities in gene expression between strains 4/74, D37712, and D23580 (Fig. 3A). The gene expression values from each profiled condition are available as raw counts and TPMs in Tables S4 and S5.

To select the ideal environmental conditions to use for subsequent experiments, we assessed the expression profiles of known *Salmonella* pathogenicity islands which were broadly similar in strains D37712, and D23580. Although the expression profile of the SPI2 pathogenicity island was broadly similar between D37712, D23580 and 4/74 in most growth conditions, the SPI2 genes of D37712 were highly up-regulated in a single growth condition, NonSPI2 (Fig. 3B-C). NonSPI2 is a minimal medium with a neutral pH and a relatively high level of phosphate, in which *S.* Typhimurium does not usually express the SPI2 pathogenicity island (Löber *et al.*, 2006; Kröger *et al.*, 2013). This intriguing observation prompted us to perform ~~thea~~ more discriminating set of transcriptomic experiments, as described below.

Differential gene expression analysis of *S.* Typhimurium D37712 versus D23580 in four *in vitro* conditions with multiple biological replicates

To define the transcriptional signature of strain D37712 more accurately, we generated RNA-seq data from D37712 grown in four *in vitro* conditions that stimulate expression of the majority of virulence genes: ESP, anaerobic growth, NonSPI2 and InSPI2, with multiple (3-4) biological replicates. The combination of acidity (pH 5.8) and low phosphate (0.4 mM Pi) in the InSPI2 media stimulates transcription of SPI2 genes in *S.* Typhimurium (Löber *et al.*, 2006; Kröger *et al.*, 2013). The NonSPI2 condition is based on the same PCN media recipe as InSPI2 media, but is neutral (pH 7.4), and contains higher levels of phosphate (25 mM Pi) (Löber *et al.*, 2006; Kröger *et al.*, 2013).

We compared the results with our published transcriptomic data for S. Typhimurium strains 4/74 and D23580 (Canals *et al.,* 2019b; Kröger *et al.*, 2013). Differential expression analysis with DEseq2, with conservative cut-offs (fold change ≥ 2, FDR ≤0.001), showed that the gene expression profiles of D37712 and D23580 were broadly similar, and shared key differences to the transcriptional profile of strain 4/74 under each of the four *in vitro* conditions (Fig. 4A). The differential expression results are summarized in Table S6.

We specifically investigated transcription of the *pgtE* gene, which encodes the outer-membrane protease previously linked to the ability of African *Salmonella* ST313 to resist human serum killing (Hammarlöf *et al.*, 2018). Compared to 4/74, the *pgtE* gene of both the D23580 and D37712 strains showed a similar pattern of up-regulation by a factor of 7 to 18 across all conditions. This finding is consistent with the fact that D37712 carries the same T nucleotide in the -10 region of the *pgtE* promoter that is responsible for increased expression of the *pgtE* transcript in strain D23580 (Hammarlöf *et al.*, 2018).

~~The~~ There were no statistically-significant changes in expression of the majority (92%) of the 4,729 orthologous coding genes ~~of both~~shared by D37712 and D23580 ~~were expressed at similar levels~~. We identified a total of 364 genes that were differentially expressed in at least one growth condition between D37712 and D23580 as follows: ESP (69 differentially-expressed genes), anaerobic growth (214 differentially-expressed genes), NonSPI2 (88 differentially-expressed genes) and InSPI2 (17 differentially-expressed genes; Fig. 4B).

Overall, the differentially -expressed genes that distinguished D37712 from D23580 ~~were seen~~only showed expression differences in a single growth condition ~~and~~rather than across all conditions. The differentially expressed genes included flagellar genes (down-regulated), SPI2-associated genes(up-regulated), and genes involved in general and anaerobic metabolism (down-regulated).

~~The~~ SPI2 pathogenicity island genes play a key role in the intracellular replication of *S.* Typhimurium, and encode the type III secretion system that is responsible for translocation of key effector proteins into mammalian cells (Jennings *et al.*, 2017). The RNA-seq data showed that SPI2 genes were expressed at similarly high levels in both D37712 and D23580 strains following induction (InSPI2 media; Fig. 4B), and confirmed that the key SPI2 expression difference was only seen in strain D37712 under non-inducing growth conditions (NonSPI2 media). It is important to put this differential SPI2 expression into context. D37712 expresses SPI2 genes at about a 10-fold higher level than D23580 during growth in non-inducing NonSPI2 media, but the actual level of expression was 20-fold less than the level stimulated by growth in SPI2-inducing conditions (InSPI2 medium).

The up-regulation of *fljA* and *fljB* and the down-regulation of *fliC* in D37712, compared to D23580 in all four growth conditions likely reflects the opposite orientation of the *hin* switch in the D37712 genome compared to D23580. This type of *hin* inversion occurs frequently in *S.* Typhimurium (Johnson and Simon, 1985).

Another gene that was up-regulated in D37712 across all profiled conditions was the chromosomally-encoded *cysS*^{chr}, that encodes cysteine-tRNA synthetase. Previously, we reported that transcription of the *cysS*^{chr} of strain D23580 was uniformly down-regulated compared to 4/74~~, a defect that~~. This down-regulation was compensated by the presence of a pBT1 plasmid-encoded cysteine-tRNA synthetase (Canals *et al.,* 2019a). ~~Increased~~Accordingly, the increased expression of the chromosomal *cysS* gene in D37712 was consistent with the absence of the pBT1 plasmid. Our comparative

~~transcripomic~~transcriptomic analysis showed that expression levels of *cysS* were similar in D37712 and 4/74 under all growth conditions.

Numerous virulence genes and operons were differentially expressed between D23580 and D37712. The SPI-16-associated *gtrABCa* operon (STM0557, STM0558, STM0559) is responsible for adding glucose residues to the O-antigen subunits of LPS that enhance the long-term colonisation of the mammalian gastrointestinal tract by *S.* Typhimurium ST19 (Bogomolnaya *et al.*, 2008). We found that the *gtrABCa* genes were significantly up-regulated in several conditions in D37712, compared to both D23580 and 4/74.

The *spvABCD* operon of D37712 was up-regulated under non-SPI2-inducing growth conditions, compared to D23580. A signature pseudogene of ST313 L2.2 is the frameshift insertion in the *spvD* gene that generates a truncated version of the SpvD protein. The H199I mutation at position 199 and the associated 17 amino acid truncation is predicted to ablate the activity of the SpvD cysteine protease (Grabe *et al.*, 2016). SpvD negatively regulates the NF-B signaling pathway and promotes virulence of S. Typhimurium in mice. The functional consequences of the *spvD* variant of ST313 L2.2 strain D37712 and the up-regulation of expression of the *spvABCD* operon remain to be established experimentally.

The SalComD37712 community transcriptional data resource

To allow scientists to gain their own biological insights from analysis of this rich transcriptomic dataset, the transcriptomic and gene expression data generated in this study are presented online in a new community resource, SalComD37712. The data resource shows the expression levels of all D37712 coding and non-coding genes, including both chromosomal and plasmid-encoded transcripts. The SalComD37712 website complements our existing SalComD23580 (https://tinyurl.com/SalComD23580) resource, and adds an inter-strain comparison of gene expression profiles between D37712 and D23580 as well as normalized gene expression values (TPM), using an intuitive heat map-based approach. SalComD37712 included our published RNA-seq data (Canals *et al.*, 2019b), re-analysed with an updated bioinformatic pipeline and a combined reference genome (see Methods). This online resource facilitates the intuitive interrogation of transcriptomic data as described previously (Perez-Sepulveda and Hinton, 2018).

Additionally, we generated a unified genome-level browser that provides access to the *S.* Typhimurium L2.2 D37712 transcriptome, in the context of our previously published RNA-seq data for the L2.0 strain D23580 and the ST19 strain 4/74. This novel “combo” browser is available at http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data.

Identification of phenotypes that distinguish ST313 sublineage L2.2 from L2.0.

To explore the phenotypic impact of the transcriptomic signature of L2.2 (D37712), we performed a series of motility experiments, fluorescence-based gene expression experiments and mixed-growth assays.

D33712 showed a significantly decreased level of motility on NonSPI2 minimal media, compared with both the ST19 strain 4/74 and the L2 D23580 strain (Fig. 5A). This finding was consistent with the transcriptomic data, which showed down-regulation of D37712 flagellar genes compared with D23580 in the NonSPI2 condition (Fig. 4). In contrast, no differential expression of flagellar genes was seen between D33712 and D23580 in the InSPI2 growth condition (Fig. 4). The decreased motility phenotype may be linked to the inversion of the *hin* element detailed above. The flagella system encodes a distinct type III secretion apparatus responsible for the dual functions of bacterial motility and activation of the mammalian innate immune system via TLR5 (Lai *et al.*, 2013).

A key transcriptomic finding for strain D33712 was the expression of SPI2 genes during growth in an unusual environmental condition (NonSPI2) (Fig. 3B-C and Fig. 4B). NonSPI2 media differs from InSPI2 media by having a higher pH (pH7.4 versus pH5.8) and a higher level of phosphate (Löber *et al.*, 2006). This apparent differential expression of SPI2 genes at the transcriptomic level under non-inducing conditions led us to investigate the expression of SPI2 at a single cell level using fluorescence transcriptional fusions. First, we introduced an *ssaG*-GFP$^+$ transcriptional fusion into the chromosome of strains D33712 and D23580 (Methods; Table S8) to interrogate expression of the key SPI2 operon with flow cytometry. Figure 5B shows that in NonSPI2 media, the *ssaG* promoter was expressed at a 62% higher level in D33712 than in D23580 confirming the results of the transcriptomic analysis.

Because only a proportion of *S.* Typhimurium cells express certain pathogenicity island-encoded genes during *in vitro* growth (Ackermann *et al.*, 2008; Hautefort *et al.*, 2003), we determined whether the increased level of expression of SPI2 genes (Fig. 4B) was caused by a higher proportion of D33712 cells expressing SPI2 than D23580 cells. Using derivatives of the two strains that carried the *ssaG*-GFP$^+$ construct, we determined the numbers of fluorescent and non-fluorescent cells with flow cytometry (Methods). Under non-inducing conditions, slightly more D37712 cells expressed the *ssaG* SPI2 promoter than D23580 cells (65% vs 60%, respectively) (Fig. 5C). ~~However,~~Although this small difference was statistically significant (t-test: P<0.001, n=3), it did not account for the 62% increased level of non-induced SPI2 expression seen in Fig. 5B.

SPI2 expression is controlled by a complex regulatory system that operates at both a negative and positive level, involving silencing via H-NS (Lucchini *et al.*, 2006), activation by SlyA and SsrB (Fass and Groisman, 2009; Walthers *et al.*, 2011) as well as input from OmpR and Fis under non-inducing conditions (Osborne and Coombes, 2011). The ~~reason for~~mechanistic basis of the aberrant SPI2 expression in strain D37712 is worthy of further study. Possible explanations include the incomplete silencing of SPI2 transcription or the

partial activation of the SPI2 virulence genes under non-inducing growth conditions. by an unknown regulatory factor.

Increased fitness of *S.* Typhimurium ST313 sublineage L2.2 compared with L2.0 in minimal media.

It has become increasingly clear that distinct *Salmonella* pathovariants have evolved particular phenotypic properties that confer fitness advantages during infection of particular avian or mammalian hosts (Branchu *et al.*, 2018). Because *S.* Typhimurium ST313 L2.2 appeared to have displaced *S.* Typhimurium ST313 L2.0 in Malawi, we speculated that *S.* Typhimurium ST313 L2.2 might have the competitive edge in some situations. Accordingly, we determined bacterial fitness using a mixed-growth competition assay (Wiser and Lenski, 2015; Lian *et al.*, 2023). The competitive index was calculated in three different growth media using pair-wise combinations of strains D37712 and D23580. Two independent approaches were used to phenotypically distinguish the two strains, one based on antibiotic resistance (Fig. 5D) and the other based on fluorescent tagging (Fig. S5).

To confirm that strains engineered to be kanamycin-resistant or gentamicin-resistant did not impact on fitness (Methods), we first verified that the tagged variants of D37712 or D23580 did not confer a growth advantage in LB or NonSPI2 media (Fig. S7). Next, we used a mixed-growth assay to investigate fitness of *S.* Typhimurium ST313 L2.0 strain D23580 or *S.* Typhimurium ST313 L2.2 strain D37712 during growth in LB, or InSPI2 or NonSPI2 minimal media. The data show that both strains grew at similar levels following overnight mixed-growth in nutrient-rich LB media, but D37712 had a competitive advantage during mixed-growth in InSPI2 media (CI = 1.79; $P<0.05$) and a greater competitive edge in NonSPI2 media (CI = 2.20; $P<0.0001$).

We then used an independent fluorescence-based approach to assess the fitness of strains D23580 and D37712 during mixed-growth in NonSPI2 media. This time, the strains were engineered to carry either mScarlet or sGFP2 proteins and the mixed-growth experiments involved pair-wise comparisons of reciprocally-tagged strains. The flow cytometric data showed that in both cases D37712 had a significant competitive advantage in NonSPI2 media (Fig. S5 and S6).

This combination of antibiotic resistance-based and fluorescence-based competitive index experiments lead us to conclude that *S.* Typhimurium ST313 L2.2 strain D37712 had a clear fitness advantage over *S.* Typhimurium ST313 L2.0 strain D23580 during mixed-growth in two formulations of minimal media. The molecular basis of this fitness advantage remains to be established.

Perspective

Here, we report that *S.* Typhimurium ST313 L2.0 has been clonally replaced by the ST313 sublineages L2.2 and L2.3 as a cause of bloodstream infection in Blantyre, Malawi. In 2018, L2.2 represented the majority of the ST313 strains isolated from hospitalised patients in Malawi at the Queen Elizabeth Central Hospital. Our comparative genomic analysis of ST313 L2.3 identified 30 chromosomal alterations, one of which generated a deletion of the *sseI* effector gene.

Our RNA-seq-based analysis of ST313 L2.2 involved a detailed comparison versus ST313 L2.0 which revealed a key difference involving SPI2 expression. Following initially observations at the transcriptomic level in the ST313 L2 and L2.2 strains grown in a pH-neutral minimal medium (NonSPI2), the increased expression of SPI2 was confirmed at the single cell level using an *ssaG* transcriptional fusion.

A series of experiments showed that the ST313 L2.2 strain D37712 had a competitive advantage over L2 strain D23580 during mixed-growth in minimal media. We propose that this increased fitness of *S.* Typhimurium ST313 L2.2 has contributed to the replacement of ST313 L2.0 in Malawi in recent years.

Previously, we compared three virulence properties of the *S*. Typhimurium ST313 L2.0 D23580 and ST313 L2.2 D37712 strains. First, experiments involving Mucosal Invariant T (MAIT) cells showed that both D37712 and D23580 fail to elicit the high level of activation of MAIT cells that characterises infection by *S*. Typhimurium ST19 4/74 (Preciado-Llanes *et al.*, 2020). Second, the D37712 and D23580 strains stimulate similar levels of up-regulation of IL10 gene expression upon infection of human dendritic cells (Aulicino *et al.*, 2022). Third, we showed that both D37712 and D23580 express similarly high levels of the PgtE virulence factor that is responsible for the ability of *S*. Typhimurium ST313 to survive human serum-killing (Hammarlöf *et al.*, 2018). These findings lead us to conclude that the comparative genomic and transcriptomic differences that distinguish *S*. Typhimurium ST313 L2.0 strain D23580 from ST313 L2.2 D37712 (Fig. 4) do not modulate the ability of the pathogens to activate human MAIT cells or dendritic cells, or to influence the PgtE-mediated serum survival phenotype of *S.* Typhimurium ST313.

Ideally, the implications of the competitive advantage of ST313 L2.2 would be determined in the context of pathogenesis. However, we lack an informative infection model for *S.* Typhimurium ST313 (Lacharme-Lora *et al.,* 2019), and it is not yet possible to experimentally determine whether the improved fitness of L2.2 significantly enhances the success of ST313 during infection of humans.

Here we have investigated the intricate interplay of gene function that ~~is~~ ~~underpinning~~underpins the success of *S*. Typhimurium ST313 L2.2. ~~We~~Our hope is that our findings ~~might~~could contribute to future therapeutic or prophylactic strategies for combatting iNTS infections in the African setting.

Materials and methods

Bacterial strains

To investigate the evolutionary dynamics of *S.* Typhimurium ST313 L2 in Malawi over a 22 year period, we focused on the large collection of 8,000 *S.* Typhimurium isolates derived from bloodstream infection in hospitalised patients at the Queen Elizabeth Central Hospital, Blantyre, Malawi (Feasey *et al.*, 2015). The collection was assembled by the Malawi–Liverpool–Wellcome Trust Clinical Research Programme (MLW) between 1996 and 2018; the precise annual numbers of isolates are shown in Fig. 1C. A random sub-sampling strategy was used to select 608 isolates for whole-genome sequencing, which included 549 *S.* Typhimurium ST313 isolates (Pulford *et al.*, 2021).

The two *S.* Typhimurium ST313 strains that are the focus of this study are D23580 and D37712. D23580 was isolated from a Malawian 26-month-old child with malaria and anaemia in 2004. D37712 was isolated from the blood of an HIV-positive Malawian male child in 2006. These two African Salmonella strains have been deposited in the National Collection of Type Cultures (NCTC). The D23580 (lineage 2.0) strain is available as NCTC 14677. The ST313 sublineage 2.2 strain D37712 is available as NCTC 14678. All bacterial strains are detailed in Table S8.

Genome sequencing

The assembled genome and annotation of D23580 (Kingsley *et al.*, 2009; Canals *et al.*, 2019b) (L2.0) was obtained from the European Nucleotide Archive (ENA) repository (EMBL-EBI) under accession PRJEB28511 (https://www.ebi.ac.uk/ena/data/view/PRJEB28511). For genome sequencing of D37712 (L2.2), DNA was extracted using the Bioline mini kit, and quality was assessed using gel electrophoresis (0.5% agarose gel, at 30 volts for 18 h). The genome was generated by a combination of long read sequencing with a PacBio RS II and short-read sequening on an Illumina HiSeq machine at the Center for Genome Research, University of Liverpool, United Kingdom.

Sequence reads were quality checked using FastQC version 0.11.9 (Andrews, 2010) and MultiQC version 1.8 (Ewels *et al.*, 2016), trimmed using Trimmomatic (Bolger *et al.*, 2014). Hybrid assembly of the Illumina and PacBio sequence reads was done with Unicycler v0.4.7 (Wick *et al.*, 2017).

The assembled genome of S. Typhimurium SDT313 L2.2 strain D37712 was deposited in Genbank (GCA_014250335.1, assembly ASM1425033v1). Raw sequencing reads were deposited for both PacBio and Illumina, under BioProject ID PRJNA656698. Sequence Read Archive (SRA) database IDs are: SRR12444880 for Illumina and SRR12444881 for PacBio.

Comparative genomic analyses

To generate the data summarised in Fig. 1C, sequencing data of 29 *S*. Typhimurium ST313 strains (Msefula *et al.*, 2012) were downloaded from EMBL-EBI database (https://www.ebi.ac.uk, accession number ERA015722). Sequence reads were assembled using Unicycler v0.4.8 (Wick *et al.*, 2017). The quality of the assemblies was assessed by Quast v5.0.2 (Gurevich *et al.*, 2013). The N50 value of all assemblies was >20kb, and the number of contigs was <600.

To construct the phylogenetic tree (Fig. 1C), *Salmonella* Typhimurium strains D23580, D37712, LT2 (GCA_000006945.2), DT104 (GCA_000493675.1), 4/74 (GCA_000188735.1), and A130 (GCA_902500285.1) were added as contextual genomes. Roary was used to make the core gene alignment, construct the gene presence/absence matrix and identify orthologous genes (Page *et al.*, 2015). Phylogenetic trees were constructed using Randomized Accelerated Maximum Likelihood (RAxML) (Stamatakis *et al.*, 2005), and were visualised with the interactive Tree of Life online tool (iToL) (Letunic and Bork, 2006).

The assembled genome and annotation of *S*. Typhimurium ST19 representative strain 4/74 (Richardson *et al.*, 2011) were obtained from GenBank (Accession number GCF_000188735.1), while the raw sequencing data of 27 *S*. Typhimurium ST313 strains described in a previous study (Msefula *et al.*, 2012) were downloaded from EMBL-EBI database (https://www.ebi.ac.uk, accession number ERA015722). The raw reads were assembled using Unicycler v0.4.8 (Wick *et al.*, 2017). The quality of the assemblies was assessed by Quast v5.0.2 (Gurevich *et al.*, 2013). The N50 value of all assemblies was >20kb, and the number of contigs was <600.

To identify SNPs, Snippy v4.4.0 (https://github.com/tseemann/snippy) was used to map the raw reads against the 4/74 genome. To detect pseudogene-associated SNPs/indels in each sub-lineage, the SNPs/indels that caused nonsense or frameshifted mutations were filtered. The identifications and names of the disrupted genes were summarised, then the wild type gene sequences were extracted from the 4/74 genome. To validate the pseudogene-associated SNPs/indels, the wild type gene sequences were used to make a BLAST database with BLAST 2.9.0+ (Camacho *et al.*, 2009). The 29 genome assemblies were queried against the databases, using the BLASTn algorithm to confirm the nonsense and frameshifted mutations in all isolates.

Phylogenetic analysis of African *Salmonella* Typhimurium isolates dating from 1966 - 2018

To examine the overall population structure of *Salmonella* Typhimurium responsible for blood infection in Malawi (Fig. 1AB and Fig. S1), the raw reads of 707 published genome sequences were downloaded (Table S7). Trimmomatic v0.36 (Bolger, A. M., Lohse, 2014) was used to trim adapters and Seqtk v1.2-r94 (https://github.com/lh3/seqtk) was used to trim low-quality regions using the trimfq flag. Fastqc v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and multiqc v1.0 (http://multiqc. info) were used to pass sequence reads according to the following criteria: passed basic quality

statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 47% and 57%. Only high-quality reads were used in the downstream analysis. Sequence reads were aligned to the *S*. Typhimurium D23580 genome using Snippy v4.4.0~~.~~ with parameter "- - mincov 5". The recombination sites of the alignment were removed by Gubbins (Croucher *et al.,* 2015), and the phylogenetic tree was built with Raxml-ng (Kozlov et al., 2019~~).~~) using GTR_G models ad 100 bootstraps. The tree was rooted on *Salmonella* Typhi strain CT18 (GCA_000195995.1) as the outgroup. The tree was visualised with the interactive Tree of Life online tool (iToL) (Letunic and Bork, 2006). The sub-lineages were identified with rHierBAPS (Tonkin-Hill *et al.,* 2018). The stacked-area chart and the bar chart showing the percentage and number of isolates from each sub-lineage were made in MS Excel.

RNA purification and growth conditions

Initially, a screen of transcriptomic gene expression was performed without biological replicates. Total RNA was purified using TRIzol from *S*. Typhimurium D37712 grown in 15 different conditions as described previously (Kröger *et al.*, 2013). To generate statistically-robust gene expression profiles, total RNA was subsequently purified using TRIzol from *S*. Typhimurium D37712 grown in four *in vitro* growth conditions (ESP, anaerobic growth, NonSPI2, InSPI2) with three biological replicates as described previously (Kröger *et al.*, 2013). RNA was isolated from intra-macrophage D37712 following infection of RAW264.7 murine macrophages using our published protocol (Srikumar *et al.*, 2015).

RNA-seq of *S*. Typhimurium strain D37712 using Illumina technology

For transcriptomic analyses, cDNA samples were prepared from *S*. Typhimurium RNA by Vertis Biotechnologie AG (Freising, Germany). RNA was first treated with DNase and purified using the Agencourt RNAClean XP kit (Beckman Coulter Genomics). RNA samples were sheared using ultrasound, treated with antarctic phosphatase and re-phosphorylated with T4 polynucleotide kinase. RNA fragments were poly(A)-tailed using poly(A) polymerase and an RNA adapter was ligated to the 5'- phosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV reverse transcriptase. The resulting cDNA was PCR-amplified to about 10-20 ng/µl. The cDNA was purified using the Agencourt AMPure XP kit. The cDNA samples were pooled using equimolar amounts and size fractionated in the size range of 200-500 bp using preparative agarose gels. The cDNA pool was sequenced on an Illumina NextSeq 500 system using 75 bp read length.

For the biological replicates of the four growth conditions (ESP, anaerobic growth (abbreviated as NoO2), NonSPI2, and InSPI2) and the intra-macrophage RNA, cDNA samples were generated as above with some improvements in library preparation. First, after fragmentation with ultrasound, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. Second, first-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer, and, after purification, the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. Sequencing of the

cDNA was performed as described above. All raw sequencing reads were deposited to the Gene Expression Omnibus (GEO) database under accession GSE161403.

RNA-seq and dRNA-seq read processing and visualization

RNA-seq data from *S*. Typhimurium 4/74 and D23580 were extracted from previously published experiments (Kröger *et al.*, 2013; Srikumar *et al.*, 2015; Canals *et al.*, 2019b; GEO dataset GSE119724). A combined reference genome was generated that contained the D23580 chromosome plus plasmids pBT1, pBT2, pBT3, pSLT-BT (from D23580) and the D37712 plasmid pCol1B9^{D37712}. All reads were aligned and quantified using Bacpipe v0.8a (https://github.com/apredeus/multi-bacpipe). Briefly, basic read quality control was performed with FastQC v0.11.8. RNA-seq reads were aligned to the genome sequence using STAR v2.6.0c using "--alignIntronMin 20 --alignIntronMax 19 --outFilterMultimapNmax 20" options. A combined GFF file was generated by Bacpipe, where all features of interest were listed as a "gene", with each gene identified by a D37712 locus tag. Subsequently, read counting was done by featureCounts v1.6.4, using options "-O -M --fraction -t gene -g ID -s 1". For visualization, scaled gedGraph files were generated using bedtools genomecov with a scaling coefficient of 10^9/(number of aligned bases), separately for sense and antisense DNA strands. Bedgraph files were converted to bigWig using bedGraphToBigWig utility (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/). Coverage tracks, annotation, and genome sequence were visualized using JBrowse v1.16.6. Transcripts Per Million (TPM) were calculated for all samples and used as absolute expression values (Table S5). A conservative cut-off was used to distinguish between expressed (TPM >10) and not expressed (TPM ≤10), as we previously described (Kröger *et al.*, 2013). Relative expression values were calculated by dividing the TPM value for one condition in one strain by the TPM value for the same condition in a different strain. Before the calculation, all TPM values below 10 were set up to 10. A conservative fold-change cut-off of 3 was used to highlight differences in expression between strains.

Differential gene expression analysis with multiple biological replicates

For differential expression analysis of *S*. Typhimurium strains 4/74, D23580, and D37712, the raw counts (Table S4) from 3-5 biological replicates in four growth conditions were used (ESP, anaerobic growth (abbreviated as NoO2), NonSPI2, and InSPI2). Differential expression analysis was done using DESeq2 v1.24.0 with default settings. A gene was considered to be differentially expressed if the absolute value of its log2 fold change was at least 1 (i.e. fold change > 2), and adjusted p-value was< 0.001.

The SalComD37712 community data resource, and the associated Jbrowse genome browser

SalCom provides a user-friendly Web interface that allows the visualisation and compaison of gene expression values across multiple conditions and between strains. Particular genes can be selected through pre-defined lists of interest, such as all sRNAs or all genes belonging to a

specific pathogenicity island. The resulting heatmap-style display highlights expression differences, and provides access to the rich, manually curated annotation of strains D37712 and D23580. The actual values behind the display can be downloaded for further processing, and a link connects the current view to a genome browser interface.

Visualisation of all the RNA-seq and dRNA-seq (TSS) coverage tracks in JBrowse 1.16.6 shows sequence reads mapped against the combined reference genome described above. Overall, the genomic distance between strains 4/74 and D23580 (approximately 1000 SNPs, or ~1 SNP per 5000 nucleotides), and between D37712 and D23580 (approximately 30 SNPs, ~1 SNP per 150,000 nucleotides) allowed the alignment of RNA-seq reads to the simplified combined reference genome without significant loss of reads. The combined reference genome facilitated a direct comparison of gene coverage as well as transcriptional start sites. The unified browser is hosted at http://hintonlab.com/jbrowse/index.html?data=Combo_D37/data.

Phenotypic and mixed competitive growth experiments

The swimming motility of S. Typhimurium strains D37712, D23580 and 4/74 was determined by a plate assay (Canals *et al.*, 2019b), which involved spotting 3 µL overnight culture onto 0.3% LB agar. Relative motility of the three strains was assessed by migration diameter after 4h and 8h of incubation at 37^{O}C.

Relative expression of the *ssaG* SPI2 promoter in strains D23580 and D37712 was measured at the single cell level via GFP fluorescence. Following the construction of a kanamycin-sensitive derivative of D23580 (strain JH4235), a P*ssaG::gfp*$^{+}$ transcriptional fusion was incorporated into the chromosome of JH4235 and D37712 by inserting the *gfp*$^{+}$ gene downstream of the *ssaG* gene, under the control of the P*ssaG* promoter. The P*ssaG::gfp*$^{+}$ D23580 derivative (JH4692), and the P*ssaG::gfp*$^{+}$ D37712 derivative (JH4693) are listed in Table S8.

The strains JH4692 and JH4693 were genome sequenced to confirm the integrity of the transcriptional fusions, and to verify that unintended nucleotide changes had not arisen. Following growth in 25 mL non-inducing NonSPI2 media in a 250 mL flask at 37°C with shaking at 220 rpm for approximately 8 hours until OD_{600}=0.3, fluorescence was determined with a BD FACSAria Flow Cytometer. The relative fluorescence of the two strains JH4692 and JH4693, and the numbers of individual fluorescent bacteria that expressed the P*ssaG::gfp*$^{+}$ promoter, were determined with FlowJo VX software.

The relative fitness of S. Typhimurium strains D37712 and D23580 was assessed in two independent mixed-growth experiments. First, kanamycin-resistant derivatives of each strain were constructed by inserting the *aph* kanamycin resistance gene into the chromosome at the intergenic region between the *STM4196* and *STM4197* genes, a region that we have previously shown to be transcriptionally silent (Canals *et al.*, 2019b). The strains were designated D23580::Km^{R} JH3794 and D37712::Km^{R}, JH4232. Mixed cultures of wild-type or

kanamycin-resistant derivatives of each strain were grown overnight in LB, InSPI2 and NonSPI2 media in a 250 mL flask at 37°C with shaking at 220 rpm. Following plating on LB agar or LB + kanamycin, colonies were counted and the ratio of bacterial strains was determined. To confirm that the insertion of kanamycin resistance at the intergenic region between *STM4196* and *STM4197* did not impact upon fitness, a mixed-growth experiment was done in both LB and NonSPI2 media (Fig. S7).

Second, to independently assess relative fitness, Tn*7*-based plasmids (Schlechter and Remus-Emsermann, 2019) were used to construct chromosomal sGFP2 and mScarlet derivatives of S. Typhimurium strains D23580 (sGFP2 derivative: JH4694; mScarlet derivative: JH4695) and D37712 (sGFP2 derivative: JH4696; mScarlet derivative: JH4697). The gene cassettes were inserted into the *S.* TyphimuriumTn*7* insertion site between the gene *STMMW_38451* and *glmS*. Mixed cultures of pairs of fluorescently-labelled strains were grown in NonSPI2 media at 37°C with shaking at 220 rpm for approximately 8 hours until OD_{600}=0.3. Levels of green and red fluorescence were determined with a BD FACSAria Flow Cytometer.

Figure Legends

Fig. 1. Emergence of *S.* Typhimurium ST313 sublineages L2.2 and L2.3 in Malawi. (A) Evolutionary dynamics of S. Typhimurium lineages in Blantyre, Malawi from 1996 to 2018. A maximum likelihood tree constructed with 1000 bootstraps using the GTRGAMMA model in RaxML rooted on ST19, LT2. The genomes of 549 *S.* Typhimurium ST313 isolates from bacteraemic patients at the Queen Elizabeth Hospital in Blantyre, Malawi were used for this analysis. The proportions of the five lineages/sublineages are shown. **(B)** The total number of isolates of each lineage/sublineage per year. **(C)** Phylogenetic comparison between representative strains of *S.* Typhimurium ST19 and four ST313 lineages/sublineages (L1, L2.0, L2.2, L2.3) showing the presence and absence of plasmids, prophages and the *spvD* pseudogene. The complete phylogenetic analysis of 707 *S.* Typhimurium genomes is shown in Fig.S1.

Fig. 2. Key genetic similarities and differences between the chromosome and plasmid profiles of D23580 (lineage 2) and D37712 (L2.2). (A) A comparison of the D23580 (L2.0) and D37712 (L2.2) chromosomes. The dots around the chromosome are different kinds of SNPs identified. Phages and *Salmonella* pathogenicity islands are shown in blue and red respectively. **(B)** Plasmid profile of D37712 versus D23580. The pSLT-BT virulence plasmid is present in both D37712 and D23580, and carries the Tn-21 transposable element; **(C)** pCol1B9 is present in D37712 and absent from D23580 (D) pBT3 is present in both D37712 and D23580. **(E)** Absence of *sseI* gene and the STM1050 coding sequence in L2.2 (D37712), as compared to *S.* Typhimurium ST19 4/74 and *S.* Typhimurium ST313 L2.0 (D23580). **(F)** List of pseudogenes in D37712 and D23580, with reference to 4/74. The colour blue means pseudogene/disrupted gene while grey indicates functional genes. *macB* is a pseudogene in D23580 (L2.0) but not in L2.2, while *spvD* is a pseudogene in L2.2 but not in L2.0. All L2.2 strains share similar pseudogenes.

Fig. 3. General comparison of expression profiles of strains 4/74, D23580, and D37712 under 17 different *in vitro* conditions. (A) Principal component analysis (PCA) plot of the individual RNA-seq samples, indicating the overall similarity in gene expression between the three strains. The 17 growth conditions have been defined previously (Kröger *et al.*, 2013). **(B)** Visualization of SPI-2 pathogenicity island expression with the Jbrowse genomic browser, ~~under~~at mid-exponential phase (MEP), InSPI2, and NonSPI2 *in vitro* conditions~~.~~, which can be accessed here. **(C)** Boxplot visualization of SPI-2 gene expression ~~under~~at mid-exponential phase (MEP), InSPI2, and NonSPI2 *in vitro* conditions. The y-axis shows the combined log TPM values for 45 genes located in the SPI2 pathogenicity island, namely *ssaU, ssaT, ssaS, ssaR, ssaQ, ssaP, ssaO, ssaN, ssaV, ssaM, ssaL, ssaK, STnc1220, STM1410, ssaJ, ssaI, ssaH, ssaG, sseG, sseF, sscB, sseE, sseD, sseC, sscA, sseB, sseA, ssaE, ssaD, ssaC, ssaB, ssrA, ssrB, orf242, orf319, orf70, ttrR, ttrS, ttrC, ttrB, ttrA, orf408, orf245, orf32, and orf48.* The elevated expression of SPI-2 genes in strain D37712 cultured ~~under~~in NonSPI2 ~~conditions~~media is highlighted in a red box.

Fig. 4. Differential gene expression of *S.* Typhimurium 4/74, D37712, and D23580 under 4 in vitro conditions. (**A)** Boxplots indicating the number of differentially-expressed genes identified in the following *in vitro* growth conditions: early stationary phase, ESP; anaerobic growth, NoO2; SPI-2 inducing medium, InSPI2; SPI-2 non-inducing minimal medium, NonSPI2. Multiple (3 to 5) biological replicates were used for comparison. DESeq2 was used for differential analysis; only genes with |log2FC| ≥ 1 and with adjusted *p*-value ≤ 0.001 were retained. **(B)** Heatmap of the genes differentially expressed between D23580 and D37712. Functional groups and operons of interest are highlighted on the right of Panel B.

Fig. 5. Phenotypes that distinguish ST313 L2.2 from ST313 L2.0. (A) Swimming motility assay of strains D23589, D37712 and 4/74, with a representative plate shown on the left. Average migration diameters were measured after 4 and 8 hours. Each bar represents the mean of three biological replicates, with *error bars* ~~representing~~showing standard deviation. Significant difference (***) indicates *P* value (*t* test) < 0.001. In Panels B & C, comparison of *ssaG* expression by flow cytometry using D23580 and D37712 derivatives containing a chromosomal *ssaG*-GFP$^+$ transcriptional fusion, strains SZS008 and SZS032, respectively. Cells were collected at 8 hours after inoculation in NonSPI2 media. Ten thousand events were acquired for each sample. **(B)** Mean fluorescent intensity signal of *ssaG*-GFP$^+$ for D23580 (SZS008, dark grey) and D37712 (SZS032, ~~,~~grey~~).~~) grown in NonSPI2 media. Significant difference (***) indicates *P* value (*t* test) < 0.001. ~~**(C)**~~**(C)** The proportions of bacterial cells that expressed *ssaG*-GFP$^+$ during growth in NonSPI2 media was determined. Percentage of ~~positive~~GFP-expressing (green) and ~~negative~~non-fluorescent cells (white) for ~~*ssaG* expression in each sample~~D23580 (SZS008) and D37712 (SZS032) is shown. Each bar represents the mean of three biological replicates, error bars show standard deviation. Significant difference (***) indicates *P* value (*t* test) < 0.001. **(D)** Relative fitness of wild-type D23580 and D37712 and their kanamycin~~--~~-resistant derivatives. Bacterial numbers were determined ~~by~~after overnight culture of a 1:1 mixture (wild-type versus Km^R) in ~~NonSPI2 (red~~LB (left), InSPI2 (~~blue~~middle) and ~~LB (black~~NonSPI2 (right) media. Each ~~bar~~dot represents the log-transformed mean competitive index of three biological replicates with *error bars* representing 95% confidence interval from standard ~~error. *P* values were determined by *t* test (***: *P* < 0.001; **: *P* < 0.01; *: *P* < 0.05; ns: no significance).~~deviation. A ~~competitive index of 1 indicates the equal fitness of two strains, while a~~log number higher than ~~1~~0 reflects the increased fitness of kanamycin-resistant derivatives. *P* values were determined by *t* test (***: *P* < 0.001; **: *P* < 0.01; *: *P* < 0.05; ns: not significant).

Supporting information

Fig. S1, Maximum-likelihood phylogeny of 707 African *S.* Typhimurium isolates. All genome sequences have been published (Msefula *et al.,* 2012, Pulford *et al.,* 2021, Canals *et al.,* 2019b). Raw sequence reads were aligned to the *S.* Typhimurium D23580 genome (FN424405) using Snippy. The recombination sites of the alignment were removed by Gubbins,

and the phylogenetic tree was built with Raxml-ng. The tree is rooted on *Salmonella* Typhi strain CT18 as the outgroup. The MLST sequence types, HierBAPS level 1 and level 2 clusters are shown in coloured concentric rings as indicated. The *S.* Typhimurium ST313 isolates are categorised as Lineage 1, Lineage 2 or Lineage 3 according to HierBAPS level 1 clustering. ST313 Lineage 2 was then sub- divided into 3 sub-lineages according to HierBAPS level 2 clustering: ST313 L2.0, ST313 L2.2 and ST313 L2.3. The metadata and lineage designations of all the S. Typhimurium isolates are in Table S7.

Fig. S2. PCR-based confirmation of the deletion of the *sseI* gene from *S.* Typhimurium L2.2 D37712. Arrows from left to right show the forward strand while the left strand is shown by arrows from right to left. However, *sseI* gene in D23580 is a pseudogene with a SNP indicated as a red line.

Fig. S3. Genomic comparison of plasmids pCol1B9$^{4/74}$ and pCol1B9^{D37712} using Artemis Comparison Tool (ACT). Bottom panel details the differences observed in the most divergent regions, including colicin toxin-antitoxin system (in pCol1B9) and *impC-umuC-umuD* operon (in pCol1B9).

Fig. S4. RDAR Phenotypes of 4/74, D23580, D37712 and BKQZM9. The top panel shows the RDAR morphology assay and the bottom panel shows a complementary experiment that involves the induction of biofilm formation on 1% tryptone agar (MacKenzie *et al.,* 2019). Strain 4/74 was used as a RDAR-positive control, which has concentric rings and a wrinkled appearance (Pulford *et al.,* 2021). The S. Typhimurium ST313 L3 strain BKQZM9 is shown for comparative purposes.

Fig. S5. Competitive index analysis of D23580 and D37712 using fluorescently-tagged S. Typhimurium strains (A) Km^R-sGFP2 and Gm^R-mScarlet were inserted into the transposon Tn7 site of D23580 or D37712. Bent arrows represent promoters and directional arrows represent genes. **(B)** A 1:1 mix of Km^R-sGFP2 and Gm^R-mScarlet marked strain was inoculated in NonSPI2 media, followed by an overnight incubation in 37℃. ~~Percentage~~The percentage of sGFP2 (green) and mScarlet (Red) -marked cells was ~~measured~~determined by flow cytometry. Raw data are shown in Figure S7, 10,000 events were acquired for each sample. **(C)** Competitive index analysis of Km^R-sGFP2 and Gm^R-mScarlet marked strain. Bacterial numbers were determined by counting CFU for overnight culture of a 1:1 mixture in NonSPI2 media. Each dot represents a single biological replicate and the lane represents mean value. A competitive index of 1 indicates the equal fitness of two strains, while a number higher than 1 reflects an increased fitness of D37712.

Fig. S6. Raw flow cytometric data related to Fig. S5B. **(A)** JH4695 + JH4698 and **(B)** JH4696 + JH4697. A 1:1 mix of the Km^R-sGFP2 and Gm^R-mScarlet marked strains were inoculated in NonSPI2 media, followed by growth at 37℃ until OD_{600} = 0.3. The X-axis (labelled FITC) shows the GFP level and the Y-axis (labelled PE Yell-Grn) indicates the

mScarlet level. Quadrant gates were used to separate four populations, and the black numbers indicate the percentage of events in each quadrant. In total, 10,000 events were acquired for each sample.

Fig. S7. The insertion of GFP-Km or RFP-Gm did not impact on fitness. A 1:1 mix of Km^R-sGFP2 and Gm^R-mScarlet marked strains were inoculated in LB or NonSPI2 media, followed by overnight incubation in 37℃. The competitive index (CI) was calculated using the formula $(CFU_{Gm})/(CFU_{Km})$. Each dot represents the CI from a single replicate and the horizontal bars indicate the mean of each dataset.

Supplementary data

Table S1: SNP and indel variants that differentiate L2.2 (strain D37712) and L2.3 (strain D49679).

Table S2: SNP and indel variants that differentiate L2.2 (strain D37712) and L2.0 (strain D23580).

Table S3: Pseudogenes carried by ST19 and ST313 L2.0 and L2.2 (strains 4/74, D23580 and D37712).

Table S4: Raw read counts for all processed RNA-seq samples shown in Figures 3 and 4 (strains 4/74, D23580, and D37712).

Table S5: TPM values for all processed RNA-seq samples shown in Figures 3 and 4 (strains 4/74, D23580, and D37712).

Table S6: ~~Differential~~DESeq2-based differential gene expression analysis ~~using DESeq2~~ for strains D23580 vs D37712 grown in four *in vitro* conditions.

Table S7: Metadata and lineage designations of the 708 S. Typhimurium isolates used to generate the maximum likelihood phylogeny (Fig. S1).

Table S8: Bacterial strains used in this study.

Acknowledgements

The authors thank Brian Coombes and Rob Kingsley for their constructive comments during the peer review process. We are grateful to present and former members of the Hinton laboratory for helpful discussions, and to Paul Loughnane for his expert technical assistance.

This work was supported by a Wellcome Trust Investigator award [grant numbers 106914/Z/15/Z and 222528/Z/21/Z] to J.C.D.H., and by the Malawi-Liverpool-Wellcome Research Centre Director's Fund. B.K. was funded by an AESA-RISE fellowship from the African Academy of Sciences [Grant Number: RPDF-18-04]. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

Conceptualization: B.K., ~~RH,~~R.H., M.A.G., C.L.M. and J.C.D.H.

Data curation: B.K., R.C., A.V.P., C.V.P., P.A.

Formal analysis: B.K., R.C., C.V.P., A.V.P., X.Z., C.K., S.V.O., Y.L., P.A., A.D. and J.C.D.

Funding acquisition: B.K., R.C., A.V.P., X.Z. and J.C.D.H.

Investigation: B.K., R.C.A., A.V.P., X.Z. and J.C.D.H.

Methodology: B.K., R.H., M.A.G., C.L.M. and J.C.D.H.

Project administration: B.K. and J.C.D.H.

Resources: B.K., R.H., M.A.G, C.L.M. and J.C.D.H.

Software: B.K. and A.V.P.

Supervision: M.A.G., C.L.G., C.L.M. and J.C.D.H.

Validation: B.K., R.C., A.V.P., X.Z. and J.C.D.H.

Visualization: B.K., R.C., Y.L., C.V.P., A.V.P. and J.C.D.H.

Writing original draft: B.K., R.C. and J.C.D.H

Writing reviews and editing: B.K., R.C., A.V.P., X.Z., C.K., S.V.O., A.D., R.H., M.G and J.C.D.H.

Equal contribution: Authors B.K., R.C. and A.V.P. made equal contributions to this work.

References

Ackermann M, Stecher B, Freed N, Songhet P, Hardt W, and Doebeli M (2008) Self-destructive cooperation mediated by phenotypic noise. 454:987–990.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aulicino A, Antanaviciute A, Frost J, Sousa Geros A, Mellado E, Attar M, Jagielowicz M, Hublitz P, Sinz J, Preciado-Llanes L, Napolitani G, Bowden R, Koohy H, Drakesmith H, and Simmons A (2022) Dual RNA sequencing reveals dendritic cell reprogramming in response to typhoidal Salmonella invasion. *Commun Biol* 5:111.

Bogomolnaya LM, Santiviago CA, Yang H-J, Baumler AJ, and Andrews-Polymenis HL (2008) 'Form variation' of the O12 antigen is critical for persistence of *Salmonella* Typhimurium in the murine intestine. *Mol Microbiol* 70:1105–1119.

Bolger AM, Lohse M, and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Branchu P, Bawn M, and Kingsley RA (2018) Genome Variation and Molecular Epidemiology of Salmonella enterica Serovar Typhimurium Pathovariants. *Infect Immun* 86:e00079-18.

Brink T, Leiss V, Siegert P, Jehle D, Ebner JK, Schwan C, Shymanets A, Wiese S, Nürnberg B, Hensel M, Aktories K, and Orth JHC (2018) Salmonella Typhimurium effector SseI inhibits chemotaxis and increases host cell survival by deamidation of heterotrimeric Gi proteins. *PLOS Pathog* 14:e1007248.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Canals R, Chaudhuri RR, Steiner RE, Owen SV, Quinones-Olvera N, Gordon MA, Baym M, Ibba M, and Hinton JCD (~~2019~~2019a) The fitness landscape of the African Salmonella Typhimurium ST313 strain D23580 reveals unique properties of the pBT1 plasmid. *PLOS Pathog* 15:e1007948.

Canals R, Hammarlöf DL, Kröger C, Owen SV, Fong WY, Lacharme-Lora L, Zhu X, Wenner N, Carden SE, Honeycutt J, Monack DM, Kingsley RA, Brownridge P, Chaudhuri RR, Rowe WPM, Predeus AV, Hokamp K, Gordon MA, and Hinton JCD (~~2019~~2019b) Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. *PLOS Biol* 17:e3000059.

Carden SE, Walker GT, Honeycutt J, Lugo K, Pham T, Jacobson A, Bouley D, Idoyaga J, Tsolis RM, and Monack D (2017) Pseudogenization of the Secreted

Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313 within Migratory Dendritic Cells. *Cell Host Microbe* 21:182–194.

Chirwa E, Dale, H, Gordon, MA, and Ashton, PM (2023) What is the Source of Infections Causing Invasive Nontyphoidal Salmonella Disease? *Open Forum Infect Diseaases* February 2023.

Crump JA, Sjölund-Karlsson M, Gordon MA, and Parry CM (2015) Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive Salmonella Infections. *Clin Microbiol Rev* 28:901–937.

Ewels P, Magnusson M, Lundin S, and Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

Fass E, and Groisman E (2009) Control of Salmonella pathogenicity island-2 gene expression. *Curr Opin Microbiol* 12:199–204.

Feasey NA, Dougan G, Kingsley RA, Heyderman RS, and Gordon MA (2012) Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *The Lancet* 379:2489–2499.

Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist L, Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C, Kariuki S, Maclennan CA, Onsare RS, Weill F-X, Le Hello S, Smith AM, McClelland M, Desai P, Parry CM, Cheesbrough J, French N, Campos J, Chabalgoity JA, Betancor L, Hopkins KL, Nair S, Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO, Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA, Parkhill J, Heyderman RS, Dougan G, Gordon MA, and Thomson NR (2016) Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat Genet* 48:1211–1217.

Gilchrist J, and MacLennan C (2019) Invasive Nontyphoidal Salmonella Disease in Africa. *Cell Mol Biol E Coli Salmonella Enterobact*, doi: doi:10.1128/ecosalplus.ESP-0007-2018.

Grabe GJ, Zhang Y, Przydacz M, Rolhion N, Yang Y, Pruneda JN, Komander D, Holden DW, and Hare SA (2016) The Salmonella Effector SpvD Is a Cysteine Hydrolase with a Serovar-specific Polymorphism Influencing Catalytic Activity, Suppression of Immune Responses, and Bacterial Virulence. *J Biol Chem* 291:25853–25863.

Gurevich A, Saveliev V, Vyahhi N, and Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075.

Hammarlöf DL, Kröger C, Owen SV, Canals R, Lacharme-Lora L, Wenner N, Schager AE, Wells TJ, Henderson IR, Wigley P, Hokamp K, Feasey NA, Gordon MA, and Hinton JCD (2018) Role of a single noncoding nucleotide in

the evolution of an epidemic African clade of *Salmonella*. *Proc Natl Acad Sci* 115.

Hautefort I, Proença MJ, and Hinton JCD (2003) Single-Copy Green Fluorescent Protein Gene FusionsAllow Accurate Measurement of *Salmonella* Gene Expression InVitro and during Infection of MammalianCells. *Appl Environ Microbiol* 69:7480–7491.

Honeycutt JD, Wenner N, Li Y, Brewer SM, Massis LM, Brubaker SW, Chairatana P, Owen SV, Canals R, Hinton JCD, and Monack DM (2020) Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive Salmonella isolates from Africa. *PLOS Pathog* 16:e1008763.

Jennings E, Thurston TLM, and Holden DW (2017) Salmonella SPI-2 Type III Secretion System Effectors: Molecular Mechanisms And Physiological Consequences. *Cell Host Microbe* 22:217–231.

Johnson R, and Simon M (1985) Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. *Cell* 41:781–791.

Kariuki S, Revathi G, Kariuki N, Kiiru J, Mwituria J, Muyodi J, Githinji JW, Kagendo D, Munyalo A, and Hart CA (2006) Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol* 55:585–591.

Kasumba IN, Pulford CV, Perez-Sepulveda BM, Sen S, Sayed N, Permala-Booth J, Livio S, Heavens D, Low R, Hall N, Roose A, Powell H, Farag T, Panchalingham S, Berkeley L, Nasrin D, Blackwelder WC, Wu Y, Tamboura B, Sanogo D, Onwuchekwa U, Sow SO, Ochieng JB, Omore R, Oundo JO, Breiman RF, Mintz ED, O'Reilly CE, Antonio M, Saha D, Hossain MJ, Mandomando I, Bassat Q, Alonso PL, Ramamurthy T, Sur D, Qureshi S, Zaidi AKM, Hossain A, Faruque ASG, Nataro JP, Kotloff KL, Levine MM, Hinton JCD, and Tennant SM (2021) Characteristics of *Salmonella* Recovered From Stools of Children Enrolled in the Global Enteric Multicenter Study. *Clin Infect Dis* 73:631–641.

Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, Marsh K, Achtman M, Molyneux ME, Cormican M, Parkhill J, MacLennan CA, Heyderman RS, and Dougan G (2009) Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* 19:2279–2287.

Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, Döpfer D, Fazil A, Fischer-Walker CL, Hald T, Hall AJ, Keddy KH, Lake RJ, Lanata CF, Torgerson PR, Havelaar AH, and Angulo FJ (2015) World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLOS Med* 12:e1001921.

Koolman L, Prakash R, Diness Y, Msefula C, Nyirenda TS, Olgemoeller F, Perez-Sepulveda B, Hinton JCD, Owen SV, Feasey NA, Ashton PM, and Gordon MA (2022) *Case-control investigation of invasive Salmonella disease in Africa – comparison of human, animal and household environmental isolates find no evidence of environmental or animal reservoirs of invasive strains*, Infectious Diseases (except HIV/AIDS).

Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K, and Hinton JCD (2013) An Infection-Relevant Transcriptomic Compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* 14:683–695.

Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hébrard M, Händler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J, and Hinton JCD (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci* 109.

Lai MA, Quarles EK, López-Yglesias AH, Zhao X, Hajjar AM, and Smith KD (2013) Innate Immune Detection of Flagellin Positively and Negatively Regulates Salmonella Infection. *PLoS ONE* 8:e72047.

Letunic I, and Bork P (2006) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:1283–1287.

Lian ZJ, Phan M, Hancock SJ, Nhu NTK, Paterson DL, and Schembri MA (2023) Genetic basis of I-complex plasmid stability and conjugation. *PLOS Genet* 19:e1010773.

Löber S, Jäckel D, Kaiser N, and Hensel M (2006) Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *Int J Med Microbiol* 296:435–447.

Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, and Hinton JCD (2006) H-NS Mediates the Silencing of Laterally Acquired Genes in Bacteria. *PLoS Pathog* 2:e81.

Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, and Hoekstra RM (2010) The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis. *Clin Infect Dis* 50:882–889.

Marchello C, Dale, A. P., Pisharody, S., Rubach, M. P., and Crump, J.A. (2019) A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. Antimicrobial Agents and Chemotherapy. 64.

Marchello CS, Dale AP, Pisharody S, Rubach MP, and Crump JA (2019) A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. *Antimicrob Agents Chemother* 64:e01974-19.

Msefula CL, Kingsley RA, Gordon MA, Molyneux E, Molyneux ME, MacLennan CA, Dougan G, and Heyderman RS (2012) Genotypic Homogeneity of Multidrug Resistant S. Typhimurium Infecting Distinct Adult and Childhood Susceptibility Groups in Blantyre, Malawi. *PLoS ONE* 7:e42085.

Musicha P, Cornick JE, Bar-Zeev N, French N, Masesa C, Denis B, Kennedy N, Mallewa J, Gordon MA, Msefula CL, Heyderman RS, Everett DB, and Feasey NA (2017) Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *Lancet Infect Dis* 17:1042–1052.

Nedialkova LP, Denzler R, Koeppel MB, Diehl M, Ring D, Wille T, Gerlach RG, and Stecher B (2014) Inflammation Fuels Colicin Ib-Dependent Competition of Salmonella Serovar Typhimurium and E. coli in Enterobacterial Blooms. *PLoS Pathog* 10:e1003844.

Okoro CK, Barquist L, Connor TR, Harris SR, Clare S, Stevens MP, Arends MJ, Hale C, Kane L, Pickard DJ, Hill J, Harcourt K, Parkhill J, Dougan G, and Kingsley RA (2015) Signatures of Adaptation in Human Invasive Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. *PLoS Negl Trop Dis* 9:e0003611.

Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, Kariuki S, Msefula CL, Gordon MA, de Pinna E, Wain J, Heyderman RS, Obaro S, Alonso PL, Mandomando I, MacLennan CA, Tapia MD, Levine MM, Tennant SM, Parkhill J, and Dougan G (2012) Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet* 44:1215–1221.

Owen SV, Wenner N, Canals R, Makumi A, Hammarlöf DL, Gordon MA, Aertsen A, Feasey NA, and Hinton JCD (2017) Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. *Front Microbiol* 8.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, and Parkhill J (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693.

Perez-Sepulveda BM, and Hinton JCD (2018) Functional Transcriptomics for Bacterial Gene Detectives. *Microbiol Spectr* 6:6.5.06.

Piccini G, and Montomoli E (2020) Pathogenic signature of invasive non-typhoidal *Salmonella* in Africa: implications for vaccine development. *Hum Vaccines Immunother* 16:2056–2071.

Post AS, Diallo SN, Guiraud I, Lompo P, Tahita MC, Maltha J, Van Puyvelde S, Mattheus W, Ley B, Thriemer K, Rouamba E, Derra K, Deborggraeve S, Tinto H, and Jacobs J (2019) Supporting evidence for a human reservoir of invasive non-Typhoidal Salmonella from household samples in Burkina Faso. *PLoS Negl Trop Dis* 13:e0007782.

Preciado-Llanes L, Aulicino A, Canals R, Moynihan PJ, Zhu X, Jambo N, Nyirenda TS, Kadwala I, Sousa Gerós A, Owen SV, Jambo KC, Kumwenda B, Veerapen N, Besra GS, Gordon MA, Hinton JCD, Napolitani G, Salio M, and Simmons A (2020) Evasion of MAIT cell recognition by the African *Salmonella* Typhimurium ST313 pathovar that causes invasive disease. *Proc Natl Acad Sci* 117:20717–20728.

Pulford CV, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ, Wenner N, Rodwell EV, Kumwenda B, Zhu X, Bennett RJ, Stenhouse GE, Malaka De Silva P, Webster HJ, Bengoechea JA, Dumigan A, Tran-Dien A, Prakash R, Banda HC, Alufandika L, Mautanga MP, Bowers-Barnard A, Beliavskaia AY, Predeus AV, Rowe WPM, Darby AC, Hall N, Weill F-X, Gordon MA, Feasey NA, Baker KS, and Hinton JCD (2021) Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. *Nat Microbiol* 6:327–338.

Rankin, J.D. and Taylor R.J. (1966) The estimation of doses of Salmonella typhimurium suitable for the experimental production of disease in calves. *Vec Rec* 78:706–7.

Richardson EJ, Limaye B, Inamdar H, Datta A, Manjari KS, Pullinger GD, Thomson NR, Joshi RR, Watson M, and Stevens MP (2011) Genome Sequences of Salmonella enterica Serovar Typhimurium, Choleraesuis, Dublin, and Gallinarum Strains of Well- Defined Virulence in Food-Producing Animals. *J Bacteriol* 193:3162–3163.

Schlechter R, and Remus-Emsermann M (2019) Delivering "Chromatic Bacteria" Fluorescent Protein Tags to Proteobacteria Using Conjugation. *BIO-Protoc* 9.

Sikand A, Jaszczur M, Bloom LB, Woodgate R, Cox MM, and Goodman MF (2021) The SOS Error-Prone DNA Polymerase V Mutasome and β-Sliding Clamp Acting in Concert on Undamaged DNA and during Translesion Synthesis. *Cells* 10:1083.

Skidmore PD, Canals R, and Ramasamy MN (2023). The iNTS-GMMA vaccine: a promising step in non-typhoidal Salmonella vaccine development. Expert Rev Vaccines (In Press) doi: 10.1080/14760584.2023.2270596.

Srikumar S, Kröger C, Hébrard M, Colgan A, Owen SV, Sivasankaran SK, Cameron ADS, Hokamp K, and Hinton JCD (2015) RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium. *PLOS Pathog* 11:e1005262.

Stamatakis A, Ludwig T, and Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.

Stanaway JD, Parisi A, Sarkar K, Blacker BF, Reiner RC, Hay SI, Nixon MR, Dolecek C, James SL, Mokdad AH, Abebe G, Ahmadian E, Alahdab F, Alemnew BTT, Alipour V, Allah Bakeshei F, Animut MD, Ansari F, Arabloo J,

Asfaw ET, Bagherzadeh M, Bassat Q, Belayneh YMM, Carvalho F, Daryani A, Demeke FM, Demis ABB, Dubey M, Duken EE, Dunachie SJ, Eftekhari A, Fernandes E, Fouladi Fard R, Gedefaw GA, Geta B, Gibney KB, Hasanzadeh A, Hoang CL, Kasaeian A, Khater A, Kidanemariam ZT, Lakew AM, Malekzadeh R, Melese A, Mengistu DT, Mestrovic T, Miazgowski B, Mohammad KA, Mohammadian M, Mohammadian-Hafshejani A, Nguyen CT, Nguyen LH, Nguyen SH, Nirayo YL, Olagunju AT, Olagunju TO, Pourjafar H, Qorbani M, Rabiee M, Rabiee N, Rafay A, Rezapour A, Samy AM, Sepanlou SG, Shaikh MA, Sharif M, Shigematsu M, Tessema B, Tran BX, Ullah I, Yimer EM, Zaidi Z, Murray CJL, and Crump JA (2019) The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis* 19:1312–1324.

Tack B, Phoba M-F, Barbé B, Kalonji LM, Hardy L, Van Puyvelde S, Ingelbeen B, Falay D, Ngonda D, van der Sande MAB, Deborggraeve S, Jacobs J, and Lunguya O (2020) Non-typhoidal Salmonella bloodstream infections in Kisantu, DR Congo: Emergence of O5-negative Salmonella Typhimurium and extensive drug resistance. *PLoS Negl Trop Dis* 14:e0008121.

Van Puyvelde S, Pickard D, Vandelannoote K, Heinz E, Barbé B, de Block T, Clare S, Coomber EL, Harcourt K, Sridhar S, Lees EA, Wheeler NE, Klemm EJ, Kuijpers L, Mbuyi Kalonji L, Phoba M-F, Falay D, Ngbonda D, Lunguya O, Jacobs J, Dougan G, and Deborggraeve S (2019) An African Salmonella Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. *Nat Commun* 10:4280.

Van Puyvelde S, de Block T, Sridhar S, Bawn M, Kingsley RA, Ingelbeen B, Beale MA, Barbé B, Jeon HJ, Mbuyi-Kalonji L, Phoba MF, Falay D, Martiny D, Vandenberg O, Affolabi D, Rutanga JP, Ceyssens PJ, Mattheus W, Cuypers WL, van der Sande MAB, Park SE, Kariuki S, Otieno K, Lusingu JPA, Mbwana JR, Adjei S, Sarfo A, Agyei SO, Asante KP, Otieno W, Otieno L, Tahita MC, Lompo P, Hoffman IF, Mvalo T, Msefula C, Hassan-Hanga F, Obaro S, Mackenzie G, Deborggraeve S, Feasey N, Marks F, MacLennan CA, Thomson NR, Jacobs J, Dougan G, Kariuki S, Lunguya O. (2023) A genomic appraisal of invasive Salmonella Typhimurium and associated antibiotic resistance in sub-Saharan Africa. Nat Commun. 14:6392.

Wick RR, Judd LM, Gorrie CL, and Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595.

Zhang S, Kingsley RA, Santos RL, Andrews-Polymenis H, Raffatellu M, Figueiredo J, Nunes J, Tsolis RM, Adams LG, and Bäumler AJ (2003) Molecular Pathogenesis of *Salmonella enterica* Serotype Typhimurium-Induced Diarrhea. *Infect Immun* 71:1–12.