Pan-genome and core genome insights into AMR-Positive *Salmonella* Lineages from African Food Chains

Muigano, M.N.[1] and Musila F.M.[2]

1. *Institute for Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000-00200, Nairobi, Kenya*
2. *Department of Biological and Life Sciences, Technical University of Kenya, P.O. Box 52428 - 00200, Nairobi, Kenya.*

Abstract

Salmonella enterica is a major public health concern in Africa because of its increasing antimicrobial resistance (AMR) to multiple antibiotics that threatens effective treatment and control. In this study, we analyzed 135 publicly available antimicrobial resistant *Salmonella* genomes from African food sources. Using a pan genomic analysis approach, we identified a pangenome of 15,269 genes, including a core genome of ~3.5 Mbp with 3,350 core genes (99–100% prevalence) and a substantial accessory genome comprising 183 soft-core, 1,441 shell, and 10,295 cloud genes. We found substantial genomic diversity among the AMR-positive *Salmonella enterica* isolates. AMR profiling revealed widespread distribution of resistance genes, with *aac(6')-laa_1* present in nearly all genomes, and *tet(A)*, *sul2*, and *fosA71* frequently co-occurring. Core virulence genes such as *invA*, *ssaL*, and *steA* were nearly universally conserved, suggesting essential roles in pathogenesis. Sporadic rare virulence genes like *papE*, *tcpC*, and *ybtP* were present in the genome pointing to lineage-specific patterns. Hierarchical clustering based on core genome multilocus sequence typing (cgMLST) revealed monophyletic clades for major serovars like *Enteritidis* and *Infantis*, while broader HC2000 groupings reflected shared evolutionary histories. Shannon diversity indices revealed geographical variation in AMR gene richness, with Burkina Faso, South Africa and Tunisia showing high intra-country diversity. Our analysis identified lineage-associated accessory genes, hence providing potential markers for

surveillance. This study offers valuable insights into the genomic architecture and AMR landscape of *Salmonella* in Africa and underscores the need for expanded genomic surveillance across diverse geographic and food production contexts.

1. Introduction

Salmonella infections are a major public health concern in the world and is implicated in a range of infections affecting both humans and animals. Every year, between 200 million and 1 billion *Salmonella* infections are reported around the world, with majority of them being foodborne-related (Lamichhane et al., 2024). Additionally, some estimates put cases of *Salmonella*-related deaths at over 150,000 (Lamichhane et al., 2024; Nazir et al., 2025). The economic burden is equally huge, with conservative estimates from 2012 putting the figure at $3.3 billion (Hoffmann et al., 2012). Several of its serovars are known for their zoonotic potential, with foodborne transmission representing a major route of infection.

The risk of *Salmonella* infection is higher in low-income countries such as those in Africa where hygiene and sanitation conditions may be inadequate (Ngogo et al., 2020). In many parts of the world including Africa, foodborne salmonellosis is made worse by the rising prevalence of antimicrobial-resistant (AMR) strains circulating in the food chain (Tack et al., 2020). The frequent recovery of AMR *Salmonella* from animal-derived food products has complicated treatment and control strategies (Salam et al., 2023). The problem is likely worse in resource-limited settings like those in many African countries.

AMR in *Salmonella* is often driven by a combination of mobile genetic elements such as plasmids, transposons, integrons, and resistance islands, alongside chromosomal mutations (Algarni et al., 2022; Partridge et al., 2018). These elements not only enable survival in antibiotic-exposed environments but also support the horizontal and vertical transfer of

resistance genes across bacterial species (Partridge et al., 2018). In addition, *Salmonella* species have intrinsic mechanisms for resisting antimicrobial therapy (Chaudhari et al., 2023). The presence of AMR genes in foodborne isolates suggests that these strains may act as important reservoirs of resistance, with implications for both animal and human health.

An analysis of large datasets of *Salmonella* genomes may offer useful insights into the AMR landscape. Advances in comparative genomics have made it possible to explore genetic diversity across bacterial populations in greater detail. One such approach is pangenome analysis, which allows researchers to examine both the conserved (core) and variable (accessory) gene content across multiple strains (Khan et al., 2023). This can offer insights into traits related to host adaptation, virulence, and antimicrobial resistance. Accessory genes, in particular, may include lineage-specific elements or mobile genetic factors that go undetected in single-genome analyses (Costa et al., 2020). Thus, in this study, we performed a pangenome analysis of AMR-positive *Salmonella enterica* strains isolated from food sources across different African regions. Publicly available genome assemblies were used to identify patterns in core and accessory gene content and to explore possible associations with antimicrobial resistance. By comparing strains from varied geographic locations, we aimed to uncover genetic markers or signatures that may be linked to resistance phenotypes. Our findings contribute to the growing body of genomic data on African *Salmonella* isolates and may inform future efforts in surveillance, diagnostics, and the management of foodborne infections.

2. Materials and Methods

2.1 Data collection and genome retrieval

In this study we retrieved a total of 526 *Salmonella enterica* genome assemblies from EnteroBase, a publicly accessible bacterial genome database and analysis platform

(https://enterobase.warwick.ac.uk/). The dataset was filtered to retain isolates of African origin, recovered exclusively from food-associated sources, including meat, milk, eggs, vegetables, and related processing environments (https://enterobase.warwick.ac.uk/species/*senterica*/search_strains?query=workspace:139979). To minimize the presence of temporal outliers, we restricted our search to assemblies collected after 2000. The search was conducted on 05 July 2025. To further refine the dataset, genome metadata were examined for antimicrobial resistance (AMR) profiles, resulting in the selection of 135 AMR-positive *Salmonella* genomes for downstream comparative analysis. Information regarding the isolates, including geographic distribution, food source, sequence type (ST), and inferred serovar, was extracted from EnteroBase records and are available in Supplementary Table 1.

2.2. Core Genome Multilocus Sequence Analysis

We used EnteroBase's in-built core genome multilocus sequence typing (cgMLST) scheme for initial high-resolution genotyping of the isolates (Accessed on July 07, 2025). We assigned clonal groups to the isolates using the Hierarchical Clustering of Core Genome Multi-Locus Sequence Typing (HierCC-cgMLST). We constructed a GrapeTree using the NINJA algorithm to visualize the phylogenetic structure and clustering of the strains based on allelic differences across core loci for all the 526 assemblies. This allowed for the preliminary identification of lineage patterns and inter-strain relatedness prior to whole-genome analysis.

2.3 Genome annotation and quality control

We downloaded the genome assemblies corresponding to the selected 135 AMR-positive isolates in FASTA format for further processing and annotation. The downloaded genome assemblies were subjected to structural and functional annotation using the Prokka v1.14.6 pipeline

(Seemann, 2014), within a Linux-based Ubuntu environment. Each genome was processed individually with Prokka's default bacterial gene prediction settings, and the output comprised GenBank (GBK), GFF3, and protein FASTA (FAA) files for each strain. The annotation workflow was implemented using a custom Bash script within a dedicated Conda environment. To ensure annotation consistency and compatibility with downstream pangenome analysis, all FASTA headers were pre-processed to remove complex characters and standardize sequence identifiers. Quality control of annotation outputs was confirmed by examining the number of annotated coding sequences (CDS), rRNA and tRNA genes, and checking for the presence of complete GFF files. A total of 135 GFF files were successfully generated and utilized for comparative genomic analysis.

2.4 Pangenome construction and core genome alignment

Comparative pangenome analysis was performed using the Roary pipeline (v3.13.0), a rapid large-scale tool for prokaryotic pangenome construction (Page et al., 2015). All 135 annotated .gff files were provided as input to Roary. The analysis was executed with the -e and --mafft options enabled to ensure accurate core gene alignment using MAFFT. The clustering identity threshold was set to 95%, and genes present in at least 99% of genomes were defined as part of the soft-core genome. Output files included the core gene alignment file (core_gene_alignment.aln), gene presence/absence matrix (gene_presence_absence.csv), and a variety of cluster statistics and supporting data.

2.5 Serovar Prediction and Phylogenetic reconstruction

Assemblies with unknown serovar composition within the AMR-positive *Salmonella* genomes were assigned using the *Salmonella* In Silico Typing Resource (SISTR) (Yoshida et al., 2016). SISTR integrates antigenic typing and core genome MLST (cgMLST) to assign serovar

predictions. In our study, we analyzed the cleaned genome assemblies using the command-line version of SISTR. The output included key metadata such as antigenic formula, serovar designation, and cgMLST information. Predicted serovars were then integrated with the phylogenetic tree and visualized in iTOL using color strip annotations, with rare serovars grouped under the category “Other” to improve interpretability.

Phylogenetic inference of the 135 *Salmonella* genomes was performed using the aligned core genome using FastTree (v2.1.11) (Price et al., 2010). The resulting phylogenetic tree was generated in Newick format and visualized using Interactive Tree of Life (iTOL v5) for downstream interpretation (Letunic and Bork, 2021). The tree topology was further integrated with metadata, including country of origin, ST, and AMR class, to explore potential associations between phylogeny and phenotypic traits.

2.6 Evaluating Genomic Plasticity with Heap’s Law Fitting

To assess the openness of the pangenome, we applied Heap’s Law, a power-law model originally adapted to microbial genomics by Tettelin and co-authors (Tettelin et al., 2008). This approach models the cumulative number of unique genes (P) discovered as more genomes (N) are added, using the equation

$$P = kN^{\gamma}$$

where k and γ are empirical constants derived through nonlinear curve fitting. The exponent γ serves as an indicator of pangenome dynamics with values closer to 1 implying high genomic fluidity and openness, while values near 0 suggest saturation and a closed gene repertoire. We used the criteria proposed by Hyun et al. (2022) to classify pangenomes as follows: $\gamma < 0.3$ (closed) and $\gamma > 0.3$ (intermediate to open).

2.7 Detection and analysis of antimicrobial resistance, plasmid, and virulence genes

To investigate the genetic basis of antimicrobial resistance, we screened all genome assemblies for acquired AMR genes using the ABRicate tool (v1.0.1) (https://github.com/tseemann/abricate) with the ResFinder database (Zankari et al., 2012). The search parameters included a minimum identity of 90% and minimum coverage of 80%. Detected genes were grouped according to resistance class (e.g., beta-lactams, aminoglycosides, fluoroquinolones, sulfonamides), and their distribution was cross-referenced with the accessory genome clusters obtained from Roary to identify potential resistance-associated genes and lineage-specific markers. The presence of resistance determinants was also evaluated in relation to phylogenetic clades derived from core genome alignment. Virulence-associated genes were identified using the VFDB database in ABRicate (Chen et al., 2016), while plasmid replicons were detected using the PlasmidFinder database (Carattoli et al., 2014).

2.8 Data analysis and visualization

Post-processing and downstream analyses were conducted in Python 3.10 using packages like pandas, seaborn, matplotlib, and scikit-learn. The gene presence/absence matrix generated by Roary (gene_presence_absence.csv) was used as input for exploratory analysis of genome-wide gene content variation (See Supplementary File 2). Heatmaps and binary matrices were constructed to visualize the distribution of accessory genes across isolates. Core genome phylogeny generated by FastTree was visualized using iTOL (Interactive Tree of Life) with metadata layers including country of origin, ST, and AMR class. The alignment file (core_gene_alignment.aln) was used as input to infer relationships among the strains and assess potential clustering by AMR gene profiles or food source. Additionally, presence/absence patterns of acquired AMR genes (as identified by ABRicate + ResFinder) were mapped to Roary accessory gene clusters to identify candidate lineage-specific or resistance-associated markers.

Genes unique to AMR-positive isolates or clades were shortlisted for further investigation. Data curation, figure generation, and final visualization were performed in Jupyter Notebook.

3. Results

3.1 Hierarchical Core Genome Phylogeny and Serovar-Level Clustering of *Salmonella* Isolates

To investigate the population structure of *Salmonella enterica* isolates from Africa, we analyzed phylogenetic relationships using cgMLST-based hierarchical clustering (HierCC) at two levels of resolution, along with *in silico* serotyping from SISTR1 (Figure 1). At the fine-scale HC100 level, closely related isolates clustered tightly, indicating close genetic similarity. Broader clustering at the HC2000 level was evident with distinct super-lineages encompassing multiple serovars. The SISTR1-based tree identified the most prevalent serovars as *Typhimurium* (n=38), *Enteritidis* (n=34), Kentucky (n=21), and Newport (n=30). Meanwhile, HC2000 clustering revealed dominance of super-lineages 2 (n=60), 12 (n=39), and 724 (n=27), whereas HC100 distinguished finer clades such as 724 (n=26), 924 (n=25), and 48 (n=22). Serovars like *Enteritidis* and *Infantis* formed tight monophyletic clusters across all methods, supporting clonal expansion. In contrast, serovars with high counts but spread across multiple HC100 clusters, such as *Typhimurium* and Newport, showed greater intra-serovar diversity. Furthermore, the grouping of diverse serovars within single HC2000 lineages highlights the presence of broader ancestral lineages circulating across the continent.

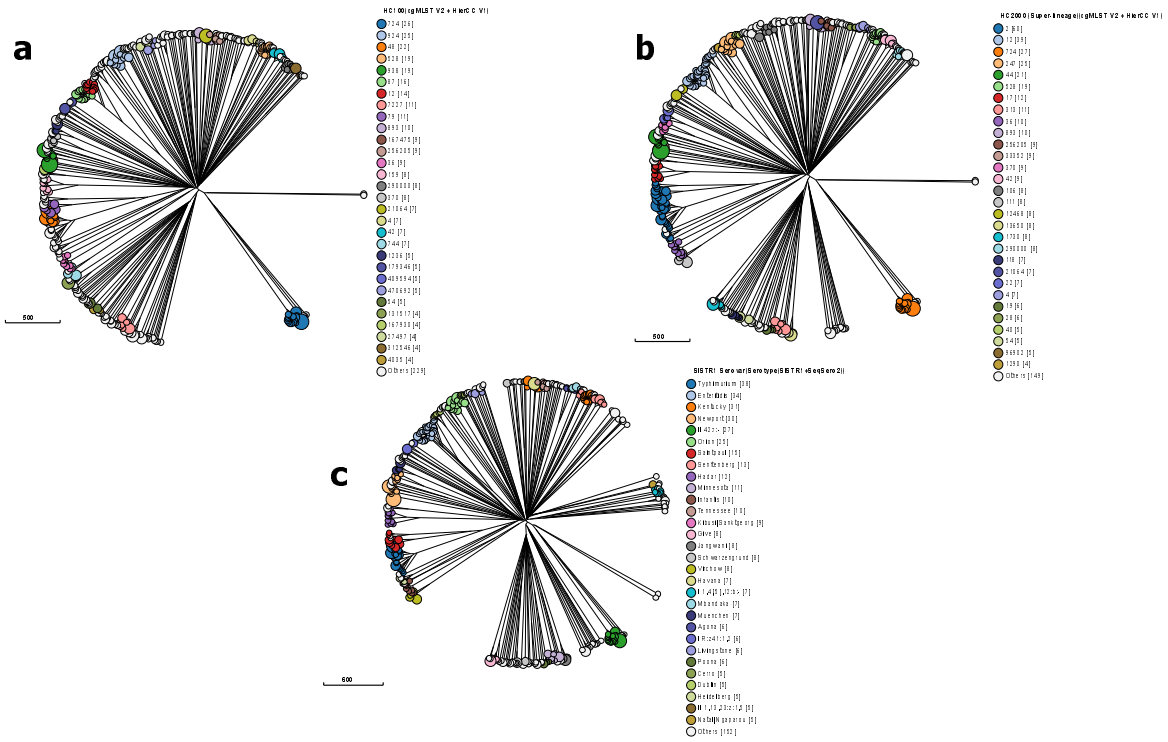

Figure 1: MLST structure of *Salmonella* in African food chains in Enterobase showing HC100 (cgMLST + HierCC) clustering (A), super-lineage hierarchical clustering (B), and SISTR serovar typing (C). Each node represents a single ST and those with the same color belong to the same serovar. GrapTree clustering was done with the NINJA NJ algorithm.

A total of 32 distinct *Salmonella* serovars were identified among the 135 AMR-positive genomes using the SISTR typing platform (Figure 2). These serovars represented a wide range of antigenic profiles and genetic diversity within the dataset. Key serovars included Kentucky, Enteritidis, Minnesota, Muenster, Schwarzengrund, and Newport, with Kentucky and its predicted variants constituting a majority of the isolates. Several rare or infrequently observed serovars were also detected but were grouped under the category "Other" for visualization clarity in the phylogenetic tree. This diversity suggests the presence of multiple evolutionary lineages and may indicate differing ecological niches or sources of infection. Figure 2 below shows an overlay of these serovar classifications onto the country and source type metadata. In general, four distinct phylogenetic clades are visible in the tree.

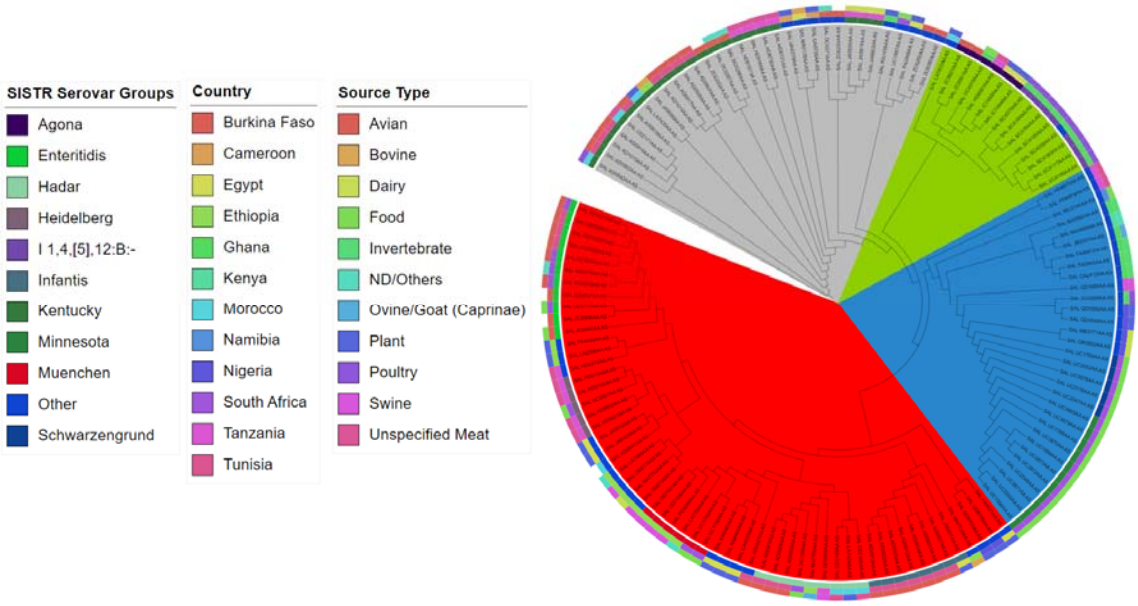

Figure 2: Core genome phylogeny of 135 AMR-positive *Salmonella enterica* isolates from African food chains. The tree was constructed on iTOL using cgMLST-based allele profiles. Metadata rings represent serovar (inner ring), country of isolation (middle ring), and source type (outer ring).

3.2 Pangenome analysis

Pan-genome analysis of 135 *Salmonella enterica* isolates from African food chains revealed a total of 15,269 unique gene clusters with an average gene length of 1,054.62 bp. Of these, 3,350 genes (21.9%), equivalent to ~3.53 Mbp, were classified as core genes as they were present in 99–100% of all genomes. This suggests a conserved set of functions essential across the *Salmonella* lineages in the African samples. An additional 183 genes (1.2%), which were found in 95–<99% of the isolates were identified as soft-core genes (Figure 3). The accessory genome comprised the majority of the gene clusters. It included 1,441 genes (9.4%) that formed the shell genome and which was shared by 15–<95% of isolates. The accessory genome also included a large proportion (10,295 genes; 67.4%) that fell within the cloud genome, and which were present in fewer than 15% of the genomes. This extensive cloud genome highlights substantial genetic diversity and plasticity across the sampled isolates. This could be due to lineage-specific

adaptations, horizontal gene transfer events, and differential environmental pressures across the continent.

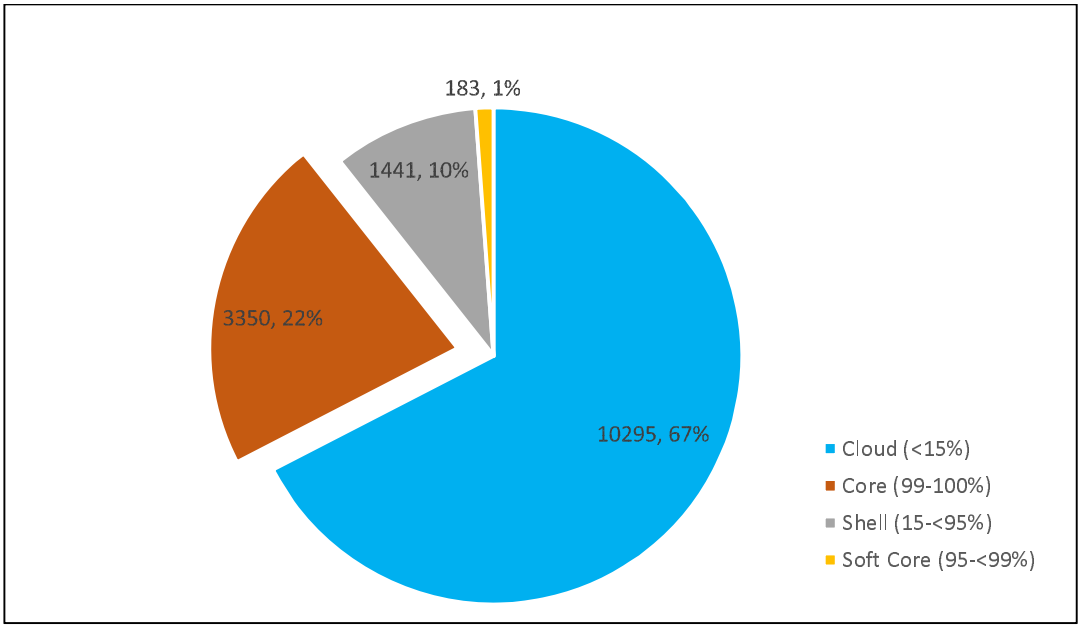

Figure 3: The distribution of genes found in the genomes of AMR *Salmonella* strains from food chains in Africa

The pan-genome curve shown in Figure 4 exhibited a steady increase in the cumulative number of genes as new *Salmonella* genomes were added, indicative of continuous gene acquisition across isolates. The growth curve showed a characteristic upward trajectory that began to gradually plateau as more genomes were included, suggesting that the majority of accessory genes had been captured within the dataset. Although the general trend was smooth, occasional minor sharp rises were observed, likely representing rare strain-specific or niche-associated genes introduced by a few individual genomes. The overall pattern of the curve aligns with an intermediate to closed pan-genome, as further supported by Heap's Law fitting (Tettelin et al., 2008). The fitted Heap's Law model yielded an exponent (γ) of 0.2844, indicating a moderately closed pan-genome. This suggests that the *Salmonella* genome from African food chains has little the capacity to acquire novel genes as additional genomes are sequenced. While new genes

continue to appear with additional genomes, the rate of gene discovery diminishes over time, which may be indicative of a reflection of a semi-diverse but not unlimited gene pool. Such a trend is characteristic of bacterial populations that engage in limited horizontal gene transfer and occupy ecologically overlapping environments, such as food chains across the African continent.

The core-genome curve displayed a progressive decline in the number of conserved genes shared across all genomes as the number of *Salmonella* assemblies increased. This trend reflects the expected reduction in universally shared genes as genomic diversity expands. The curve gradually flattened after approximately 65 genomes and this could be an indication of a stabilization of the core genome size. A single minor sharp drop was observed around that point, suggesting the inclusion of a genomically divergent strain that lacked a few otherwise conserved genes. This pattern is typical of bacterial populations with a relatively stable and conserved core genome, where most essential housekeeping genes are retained across strains. The eventual plateau indicates that adding more genomes is unlikely to significantly reduce the core gene set, an observation that is consistent with a closed or nearly closed core genome architecture. So, functional constraints on essential genes across food-associated *Salmonella* strains may exist and implies limited variability in the core functions necessary for survival and pathogenicity.

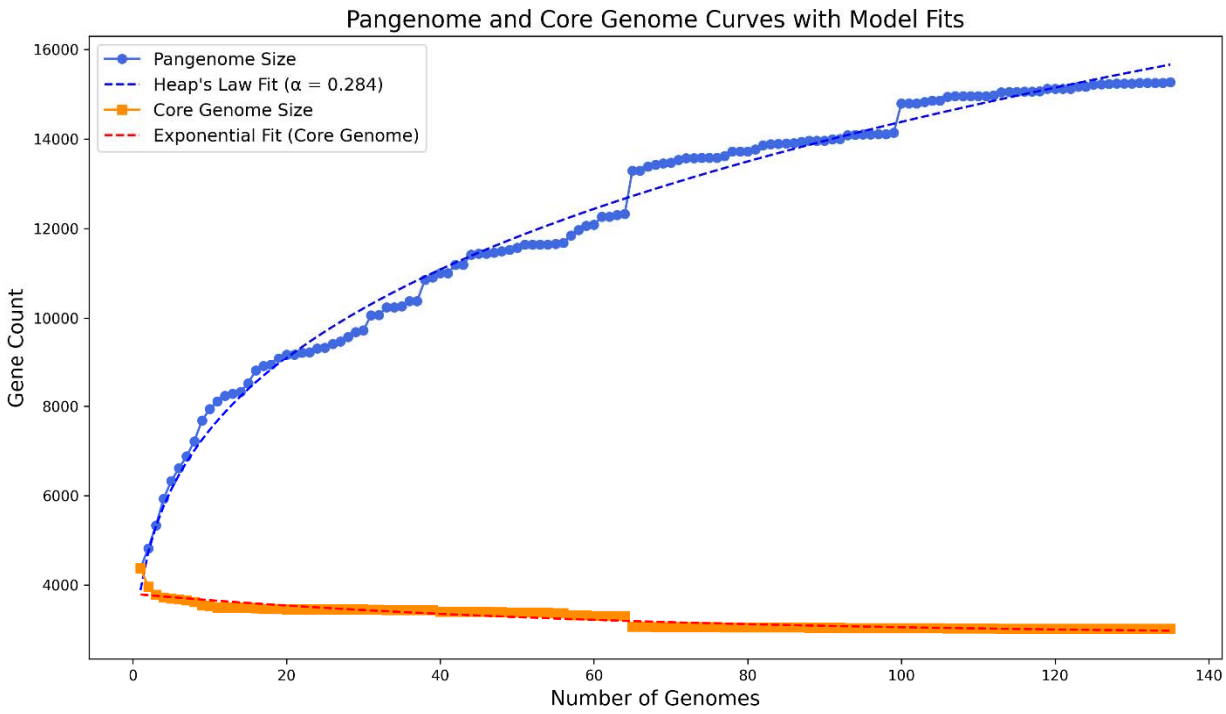

Figure 4: Pan-genome expansion dynamics of 135 AMR-positive *Salmonella enterica* isolates. The pan-genome accumulation curve (royal blue) shows the cumulative number of genes as genomes are sequentially added. Heap's Law model fit to the pan-genome (dotted blue line) illustrates an open pan-genome architecture ($\gamma < 1$), suggestive of ongoing horizontal gene acquisition and high genomic plasticity within the population. Core genome contraction curve (dark orange) of AMR-positive *Salmonella enterica* isolates from African food chains shows the decreasing trend in the number of shared (core) genes as more genomes are added. This indicates an increase in genetic diversity among the isolates. The plateauing pattern suggests a conserved core set of genes essential for the survival and maintenance of fundamental biological functions.

Figure 5 shows the gene frequency distribution of the 135 AMR-positive *Salmonella* genomes with a characteristic U-shaped profile, consistent with a dynamic bacterial pangenome. A prominent peak was observed at the far left of the histogram, indicating a high number of unique genes (n = 4,649) found in only one genome. These strain-specific genes may represent mobile genetic elements, recent horizontal acquisitions, or adaptations to niche-specific environments. As the number of genomes sharing a gene increased, the frequency of such genes decreased steadily, flattening off around the 40-genome mark, reflecting a large set of accessory genes (n =

7,604) that are variably distributed across the population. A second peak emerged at the far right of the histogram, representing core genes (n = 3,016) that are conserved across all 135 genomes. This bimodal distribution emphasizes the presence of both a stable, conserved genomic backbone and a flexible gene pool. This could point to the semi-closed nature of the *Salmonella* pangenome in AMR-associated isolates from African food chains.

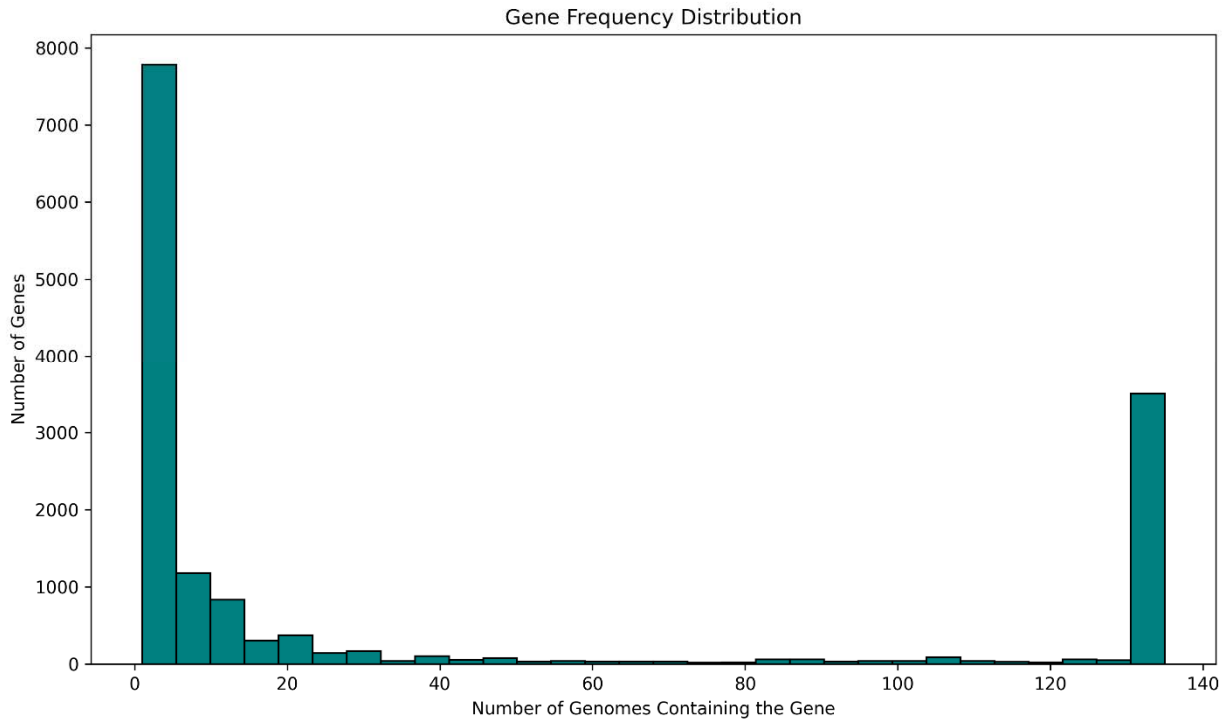

Figure 5: Gene frequency distribution among 135 *Salmonella enterica* genomes. The plot illustrates the distribution of gene presence across the analyzed genomes, with the x-axis representing the number of genomes in which a given gene is present and the y-axis representing the number of such genes. A bimodal distribution is observed, with peaks at the extreme left (representing rare and strain-specific genes) and right (ubiquitous, core genes). This points to the open nature of the *Salmonella* pan-genome and the coexistence of highly conserved and highly variable gene subsets.

3.3 AMR Resistance Genes

Figure 6 shows the frequencies of the top 20 AMR genes across various genomes. Among the 135 AMR-positive *Salmonella* genomes analyzed, the most common genes were *aac(6')-laa_1*

(n=134), *tet(A)_6* (n=59), *fosA71* (n=34), *sul2_2* (n=32), *alph(6)-ld_1* (n=30), *sul_5 (n=29)*, *aaph(3')-la_1* (n=18), *qnrB19_1* (n=18), *ant(3'')-la_1* (n=18), *blaTEM-1B_1* (n= 17). The most prevalent antimicrobial resistance gene was *aac(6')-laa_1*, detected in nearly all isolates (99%), suggesting its potential role as a conserved or chromosomally encoded resistance determinant. The AMR gene profile reveals a predominance of aminoglycoside resistance genes and widespread baseline resistance to this antibiotic class. The tetracycline resistance gene *tet(A)_6* was the second most frequent, found in 59 genomes (43.3%), followed by *fosA7_1* and *aph(6)-Id_1*, each present in at least 30 genomes. Sulfonamide resistance genes (*sul1_5* and *sul2_2*) were also commonly encountered, highlighting widespread resistance to sulfonamides. Notably, *blaTEM-1B_1*, a β-lactamase gene associated with penicillin resistance, was detected in 17 isolates, and *qnrB19_1*, conferring quinolone resistance, was present in 18 genomes. The presence of multiple aminoglycoside-modifying enzymes, including *ant(3'')-la_1* and *aph(3')-la_1*, further illustrates the diverse resistance gene landscape in *Salmonella* isolates circulating in the African food chain. Overall, our findings suggest that the co-occurrence of resistance to tetracyclines, fosfomycin, sulfonamides, and beta-lactams, demonstrating the multidrug-resistant nature of the *Salmonella* population under study.

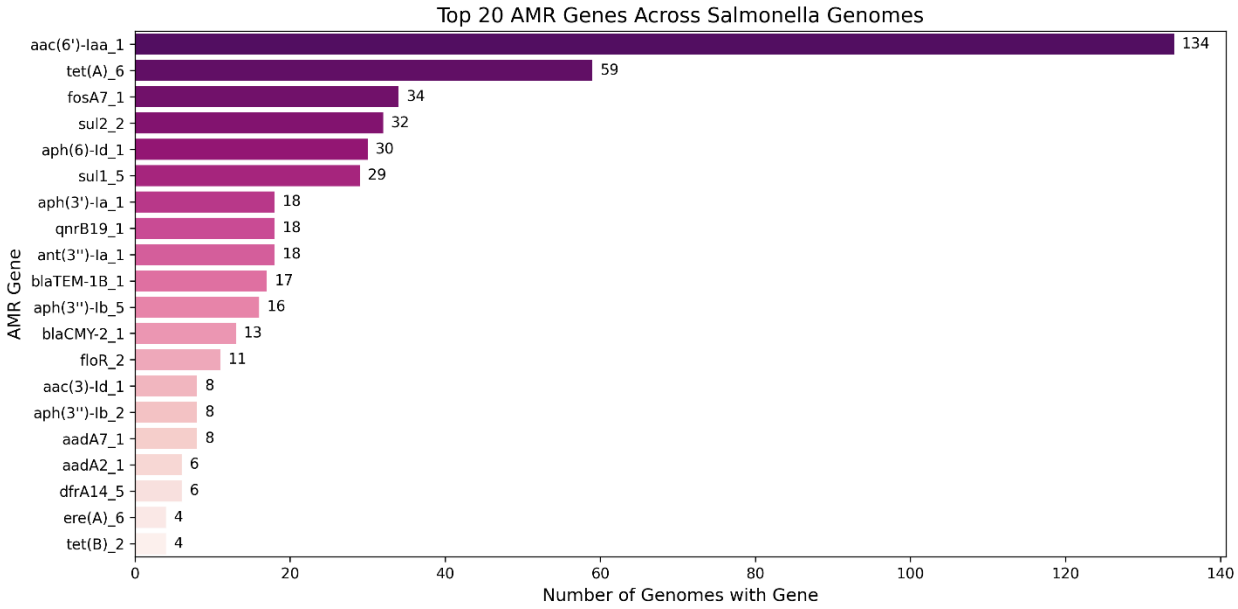

Figure 6: Distribution of the 20 Most Prevalent AMR Genes Among 135 AMR-Positive *Salmonella enterica* Genomes

Figure 7 below shows a binary heatmap depicting the presence and absence of AMR genes across *Salmonella* genomes, revealing distinct patterns of resistance distribution. The aminoglycoside resistance gene *aac(6')-laa_1* was universally present, appearing across all genomes. This is likely an indication of the potential chromosomal origin or lineage fixation of this gene. In contrast, other genes such as *tet(A)_6*, *fosA7_1*, *aph(6)-ld_1*, and *sul1_5* exhibited heterogeneous distribution, with clusters of genomes harboring multiple resistance determinants. The heatmap also highlighted varying degrees of AMR gene co-occurrence, particularly among tetracycline, sulfonamide, and aminoglycoside genes. This genomic diversity in resistance profiles is indicative of potential the role of horizontal gene transfer and niche adaptation in shaping the AMR landscape within foodborne *Salmonella* strains in Africa.

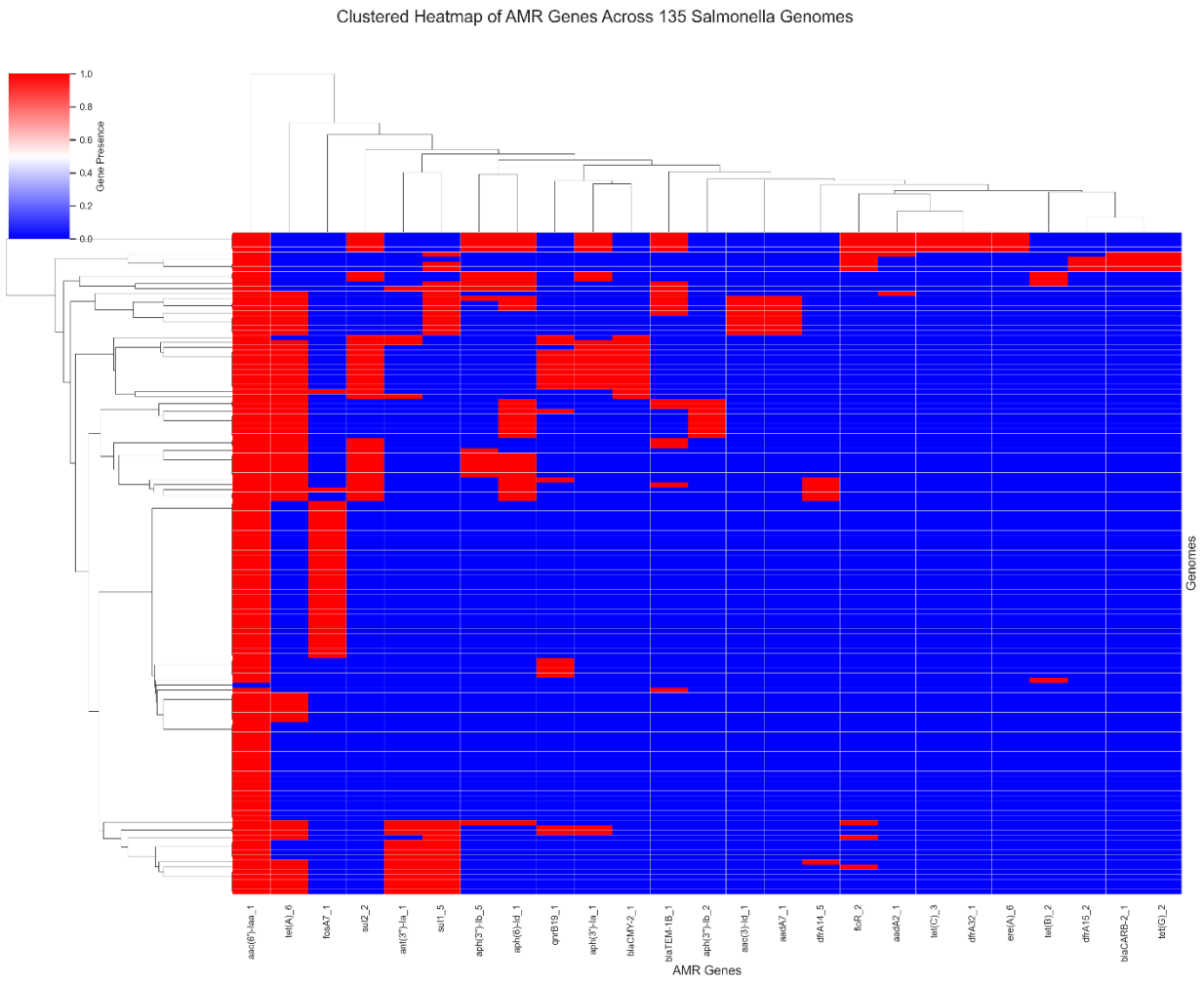

Figure 7: Presence–Absence Matrix of Antimicrobial Resistance Genes in AMR-Positive *Salmonella enterica* Genomes from African food chains

3.4 Structure and Distribution of AMR Genes

We used the Shannon diversity index to quantify AMR gene richness and evenness within individual genomes and across ecological and taxonomic strata. Fig. 8 below shows the distribution of these genes across various variables. It is evident that considerable inter-isolate variability exists whereby some genomes harbor highly diverse resistance repertoires. When categorized by source type (Fig. 8B), isolates from avian, poultry, and unspecified mean origins exhibited the highest AMR diversity, consistent with known antimicrobial practices in livestock

systems. Serovar-based analysis (Fig. 8C) showed that Heidelberg and Enteritidis carried more complex resistance profiles, which is likely due to intrinsic lineage traits or extrinsic ecological exposure. Regional variation (Fig. 8D) revealed that isolates from Burkina Faso, South Africa, Tanzania, and Tunisia had greater diversity and variation in AMR gene content. This could be an indication of a heterogeneous AMR gene pool, likely due to different sources of contamination, selective pressures, or diverse clonal lineages in these countries. These countries are potential regional selection hotspots. These findings highlight the utility of integrating ecological metadata with genomic surveillance to identify high-risk reservoir and inform AMR mitigation strategies.

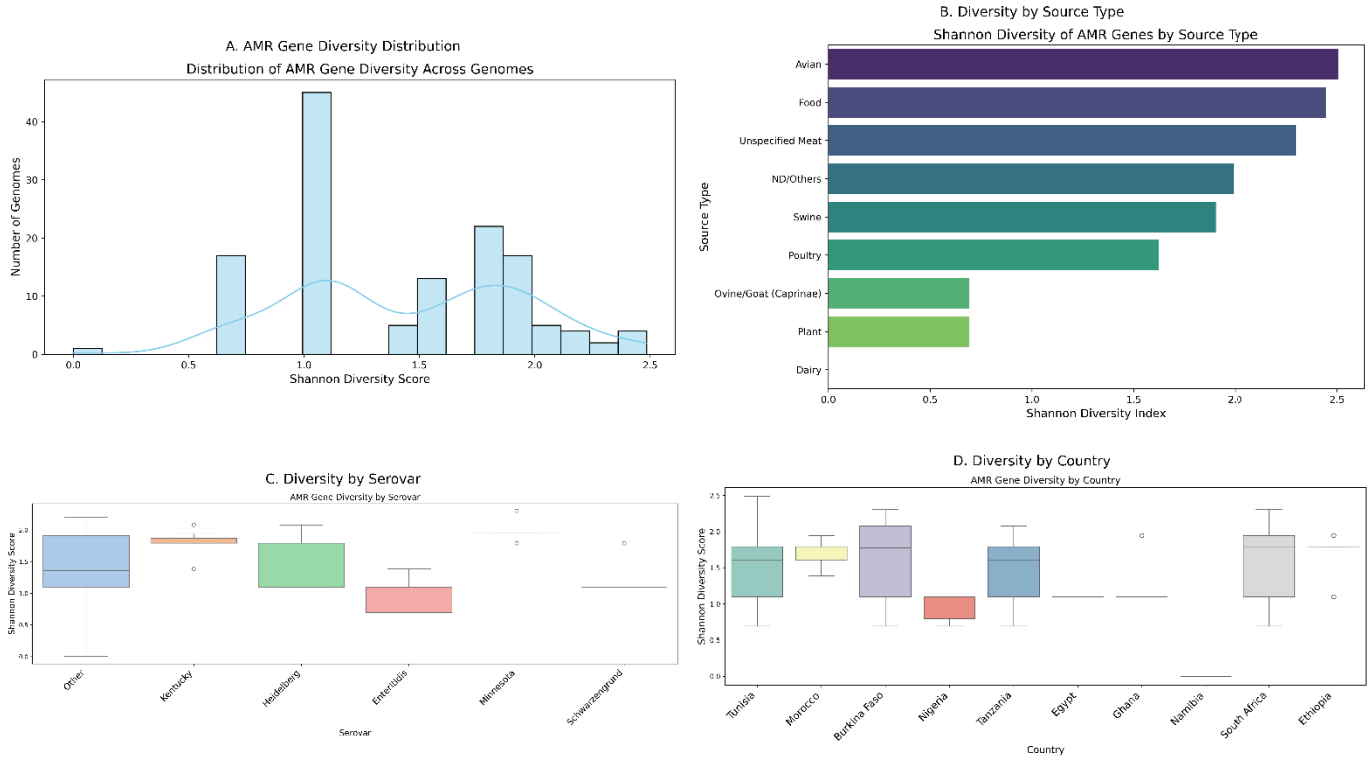

Figure 8. Shannon diversity of AMR genes across the variables (A) AMR gene diversity per isolate. (B) Distribution by source type. (C) Stratification by serovar, and (D) Regional variation by country of isolation.

We used NetworkX (https://github.com/networkx/networkx) for analysis of co-occurrence of antimicrobial resistance (AMR) genes whereby our results revealed structured and non-random patterns of association. These patterns could be suggestive of mobile genetic elements and

horizontal gene transfer mechanisms. Our findings point to the co-occurrence of aminoglycoside, beta-lactam, tetracycline, and quinolone resistance genes in *Salmonella* isolates. In the network view (Fig. 9-A), highly connected hubs such as *aac(6')-laa_1*, *tet(A)_6*, and *sul2_2* emerged as central nodes, often co-localizing within the same genome. These genes were linked to other major nodes include *ant(3")-la_1*, *qnrB19_1*, *blaCMY-2_1*, *aadA*, and *aph(3')-la_1*, *aph(3")-lb_5*, and *aph(6)-ld_1*. These associations suggest the presence of multidrug resistance cassettes and/or integrons. The corresponding matrix (Fig. 9-B) illustrates that several gene pairs frequently co-occurred across multiple isolates, reinforcing the hypothesis of coordinated dissemination. The dense connectivity among these genes depicts the modular nature of AMR gene acquisition and the potential role of mobile genetic elements, such as plasmids and transposons in facilitating their spread. The conserved nature of these resistance modules across diverse serovars and sources points to common selective pressures within African food production environments. This understanding may be important in identifying priority resistance loci for molecular surveillance and targeting potential vectors of horizontal gene transfer.

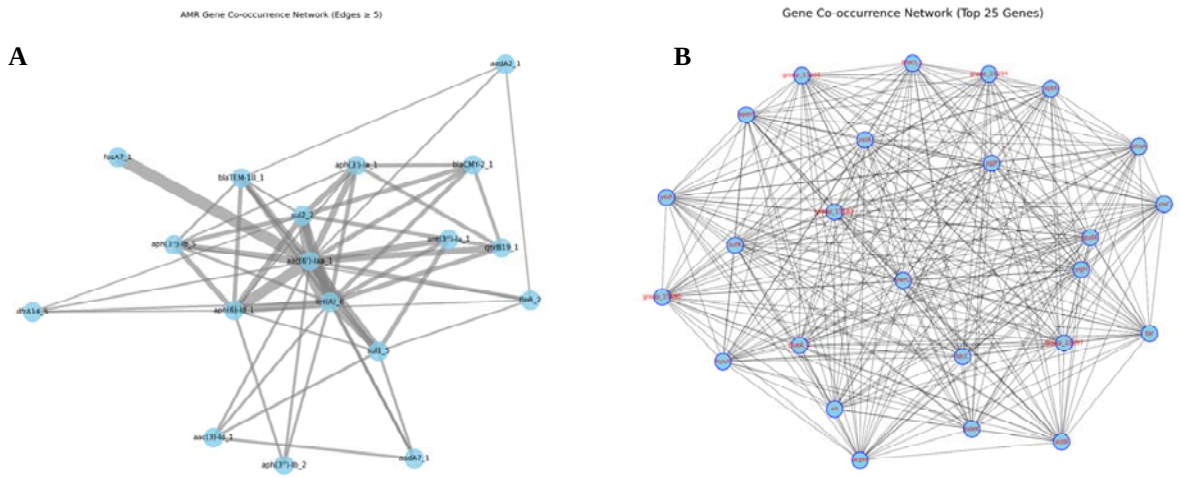

Figure 9. Co-occurrence patterns of AMR genes in *Salmonella enterica*. Fig.9(A) shows AMR gene co-occurrence network of interactions among frequently detected resistance genes. Fig. 9(B) is the Matrix view of the top 25 most frequent AMR gene co-occurrence patterns across isolates.

3.5 Plasmid burden and virulence gene presence

We analyzed plasmid burden and virulence gene content as components of the accessory genome with the results showing significant heterogeneity across serovars. Fig. 10 shows that isolates belonging to Muenchen, Heidelberg, Infantis, and *Typhimurium* serovars displayed higher plasmid counts than the rest. This could be indicative of enhanced genomic plasticity and a greater propensity for acquiring resistance elements. These serovars are often associated with broad host ranges and zoonotic transmission. As a result, plasmid-mediated traits likely facilitate adaptation to diverse ecological niches. Fig. 11 presents a heatmap of key virulence genes across various assemblies. Genes like *csgB*, *csgE*, *fimD*, *invA*, *invE*, *invH*, *mig-14*, *orgA*, *pipB2*, *sinP*, *sinH*, *sopD*, *spaO*, *spaR*, *SSaE*, *ssaI*, *ssaL*, *ssaR*, *ssaU*, *sscB*, *sseC*, and *steA* are nearly universally present in the *Salmonella* samples. It is likely that genes are part of the core virulence genome of *Salmonella enterica* among the foodborne serovars. Their universal presence suggests that the pathogenic potential is conserved even across genetically diverse lineages. These genes are likely components of the Type III Secretion Systems (T3SS) located on *Salmonella* Pathogenicity Islands (SPIs) such as SPI-1 and SPI-2, which are essential systems of host cell invasion, intracellular survival, and systemic spread as previously reported (Hansen-Wester and Hensel, 2001; Miki et al., 2009; Yu et al., 2010). Again *csg* and *film* cluster genes like *fimD* are well known mediators of biofilm production and related virulence in *Salmonella* (Musa et al., 2024). On the other hand, genes such as *papB*, *papE*, *papH*, *papK*, *pefC*, *rcK*, *tcpC*, *ybtP*, and *ybtT* were rare across samples and serovars. The rarity of these genes may point to specificity of ecological niches rather than accessory virulence. Nonetheless, the diversity in virulence profiles points to lineage-specific pathogenic mechanisms. Our results are suggestive of accessory genome's role in shaping virulence in *Salmonella*.

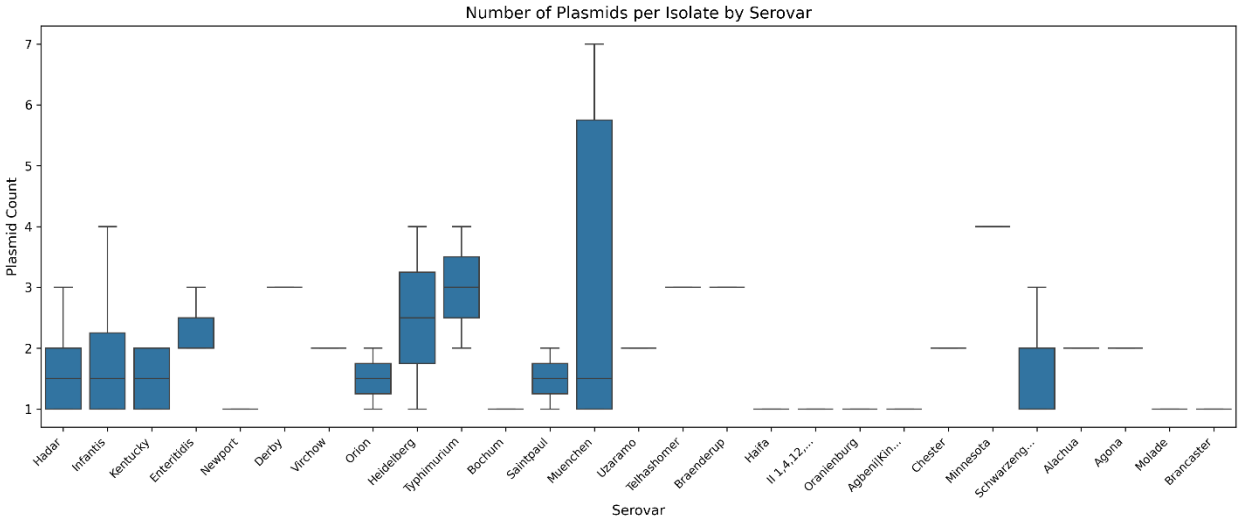

Figure 10: A boxplot showing the number of plasmids per isolate by serovar

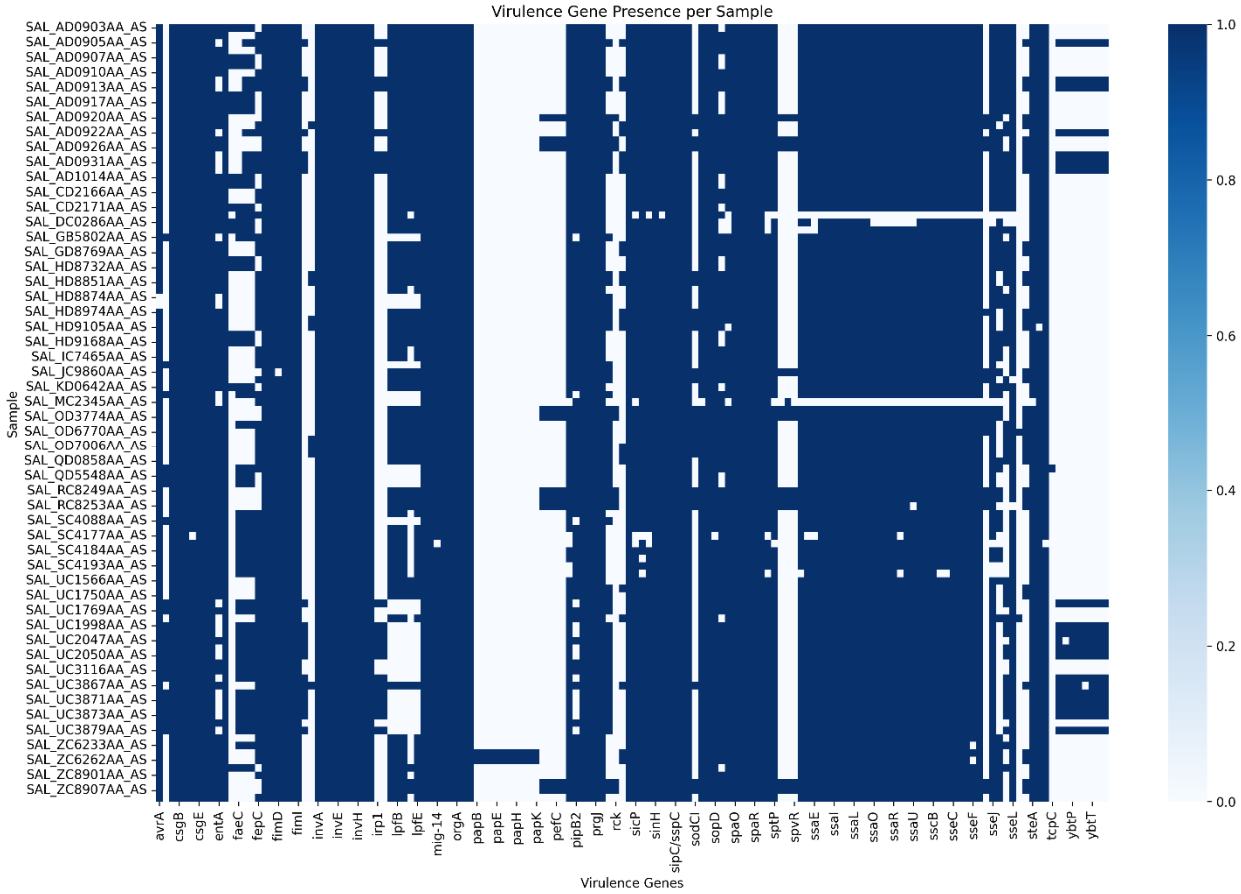

Figure 11: A heatmap depicting the distribution of virulence genes in AMR positive *Salmonella* genomes. The binary heatmap illustrates the presence (dark blue) and absence (light blue) of key

virulence genes across 135 *Salmonella enterica* genomes carrying antimicrobial resistance (AMR) determinants. Each row represents an individual genome while each column corresponds to a virulence gene.

4. Discussion

This study involved the analysis of genomic data of *Salmonella* isolates from food samples in Africa. Phylogenetic analysis using cgMLST-based HierCC clustering revealed tight, monophyletic clades at HC100, reflecting high genetic relatedness. Tight clustering at HC100 could be indicative of the presence of long-term endemic persistence (Zhou et al., 2020). The clustering also points to the possibility of the isolates sharing a high number of core genomes MLST alleles. This hypothesis is supported by the presence of a large number of core genome of 3,350 genomes present across the isolates. Broader clustering at HC2000 uncovered super-lineages that grouped multiple serovars, supporting the notion of shared ancestral backgrounds and highlighting the limitations of serotyping in capturing true evolutionary relationships. Lineage-specific gene association testing further identified genetic markers exclusive to certain serovars, opening avenues for targeted diagnostics and surveillance.

We also conducted a comprehensive pan-genomic and AMR landscape analysis of *Salmonella enterica* strains from African food chains. Our analysis focused on the distribution of resistance genes and their relationship to genetic lineage and phylogenetic structure. After analyzing 135 AMR-positive *Salmonella* genomes from cross the continent, we identified a total of 15,269 genes constituting the pangenome. Of these, 3,350 were core genes shared across nearly all isolates (99–100%), while the cloud genome representing rare or unique genes accounted for the largest proportion with 10,295 genes. These figures reflect a moderately closed pangenome structure with but with considerable genomic fluidity within the population, in line with earlier reports (Chand et al., 2020). The γ parameter of 0.2844 further supports the closed genome

nature of the *Salmonella* isolates from African food chains. Our findings are consistent with earlier reports of a closed genome ($\gamma = 0.235909$) in *Salmonella enterica* serovar (Coluzzi et al., 2025). Thus, the identification of new genomic regions in future will decline for each new genome (Laing et al., 2017; Tettelin et al., 2005).

AMR profiling revealed that the aminoglycoside resistance gene *aac(6')-laa_1* was nearly universal, occurring in 99% of isolates. This is consistent with an earlier study that reported the presence of aminoglycoside resistance genes in nearly all food isolates (Pettengill et al., 2020). Resistance to aminoglycosides occurs due to mutations that result in formation of aminoglycoside-modifying enzymes (AMEs) whose primary mechanism is the prevention of the antibiotic from binding the bacterial target (Kumar et al., 2025). Tetracycline (tetA), sulfonamide (sul2), and fosfomycin (fosA71) resistance genes were also widely distributed, indicating a multidrug-resistant (MDR) population structure. Previous studies in literature have shown the predominance of these genes in *Salmonella* isolates (Pavelquesi et al., 2021). Genes such as *qnrB19_1*, *ant(3")-la_1*, and *blaTEM-1B_1* were frequently co-localized in isolates, suggesting the presence of mobile genetic elements like plasmids or genomic islands that serve as vectors for AMR gene dissemination. This is further supported by the prominence of co-occurrence networks where hub genes such as *aac(6')-laa_1*, *tet(A)_6*, and *sul2_2* formed central nodes with multiple connections, indicating a potential for coordinated gene mobility and persistence.

We further established that core virulence determinants such as *invA*, *ssaL*, *ssaU*, *spaO*, and *steA* were consistently present across isolates and serovars. This could mean the presence of a conserved pathogenic backbone that is essential for *Salmonella* infection and intracellular survival. The near-universal presence of these genes highlights their evolutionary conservation and indispensability for host invasion, aligning with previous findings that place the Type III

secretion system (T3SS) and associated SPI-1/SPI-2 effectors at the core of *Salmonella* pathogenicity (Jennings et al., 2017; Miki et al., 2009). *Salmonella* pathogenicity island-1 (SPI-1) are present in *Salmonella* lineages and have the role of encoding a type III secretion system (TTSS-1) that is considered essential for invasion of intestinal epithelial cells, host adaptation and survival in the gastrointestinal tract (Lou et al., 2019; Phoebe Lostroh and Lee, 2001). Conversely, rare virulence genes such as papE, rcK, and tcpC were infrequently detected, reflecting their lineage-specific acquisition or possible roles in niche adaptation.

Diversity analyses using Shannon indices indicated significant variation in AMR gene richness across geographical locations. For example, the isolates from Tunisia showed high interquartile variability and extended error bars, which may be an indication of heterogeneous AMR acquisition. Previous studies from Tunisia have shown the presence of multi-drug resistance *Salmonella* (Oueslati et al., 2023). Similarly, higher diversity indices were noted for Burkina Faso, South Africa, Tanzania. This aligns with previous studies showing diverse *Salmonella* serotypes and antimicrobial resistance in sub-Saharan countries (Kilongosi Webale, 2024; Mwambene et al., 2025; Ramatla et al., 2021; Siourimè et al., 2017). In contrast, countries like Morocco displayed more stable profiles, indicating a potentially localized resistance gene pool. Such diversity underlines the importance of regional surveillance and suggests that resistance dynamics may not be uniformly distributed across Africa. Overall, genome analysis points to genomic complexity and a high AMR burden of *Salmonella* in African food chains. The detection of lineage-specific accessory genes, core virulence traits, and diverse AMR determinants supports the need for integrated genomic surveillance to guide antimicrobial stewardship and public health interventions.

5. Conclusion

This study provides a comprehensive analysis of AMR-positive *Salmonella enterica* genomes from African food chains. The findings show important trends in antimicrobial resistance, virulence, and lineage-specific gene content. While having a moderately closed pangenome, *Salmonella* isolates from African food chains show possibility of flexible gene acquisition. Hierarchical clustering confirmed the presence of serovar- and lineage-associated genes, suggesting potential targets for molecular surveillance and a conservation of core virulence determinants. Furthermore, country-level differences in AMR gene diversity highlight the impact of local selection pressures and the need for region-specific mitigation strategies. These findings show the value of genome-wide approaches in advancing our understanding of *Salmonella* epidemiology. The findings also attest to the need for the integration of genomic data into public health and food safety monitoring across Africa.

6. References

Algarni, S., Ricke, S.C., Foley, S.L., Han, J., 2022. The Dynamics of the Antimicrobial Resistance Mobilome of Salmonella enterica and Related Enteric Bacteria. Front. Microbiol. 13. https://doi.org/10.3389/fmicb.2022.859854

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. 58, 3895–3903. https://doi.org/10.1128/AAC.02412-14

Chand, Y., Alam, Md.A., Singh, S., 2020. Pan-genomic analysis of the species *Salmonella enterica*: Identification of core essential and putative essential genes. Gene Rep. 20, 100669. https://doi.org/10.1016/j.genrep.2020.100669

Chaudhari, R., Singh, K., Kodgire, P., 2023. Biochemical and molecular mechanisms of antibiotic resistance in *Salmonella* spp. Res. Microbiol. 174, 103985. https://doi.org/10.1016/j.resmic.2022.103985

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 44, D694-697. https://doi.org/10.1093/nar/gkv1239

Coluzzi, C., Piscon, B., Dérozier, S., Chiapello, H., Gal-Mor, O., 2025. Comparative genomics of Salmonella enterica serovars Paratyphi A, Typhi and Typhimurium reveals distinct profiles of their pangenome, mobile genetic elements, antimicrobial resistance and defense systems repertoire. Virulence 16, 2504658. https://doi.org/10.1080/21505594.2025.2504658

Costa, S.S., Guimarães, L.C., Silva, A., Soares, S.C., Baraúna, R.A., 2020. First Steps in the Analysis of Prokaryotic Pan-Genomes. Bioinforma. Biol. Insights 14, 1177932220938064. https://doi.org/10.1177/1177932220938064

Hansen-Wester, I., Hensel, M., 2001. Salmonella pathogenicity islands encoding type III secretion systems. Microbes Infect. 3, 549–559. https://doi.org/10.1016/S1286-4579(01)01411-3

Hoffmann, S., Batz, M.B., Morris, J.G., 2012. Annual Cost of Illness and Quality-Adjusted Life Year Losses in the United States Due to 14 Foodborne Pathogens. J. Food Prot. 75, 1292–1302. https://doi.org/10.4315/0362-028x.jfp-11-417

Hyun, J.C., Monk, J.M., Palsson, B.O., 2022. Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity. BMC Genomics 23. https://doi.org/10.1186/s12864-021-08223-8

Jennings, E., Thurston, T.L.M., Holden, D.W., 2017. Salmonella SPI-2 Type III Secretion System Effectors: Molecular Mechanisms And Physiological Consequences. Cell Host Microbe 22, 217–231. https://doi.org/10.1016/j.chom.2017.07.009

Khan, K., Jalal, K., Uddin, R., 2023. Pangenome diversification and resistance gene characterization in Salmonella Typhi prioritized RfaJ as a significant therapeutic marker. J. Genet. Eng. Biotechnol. 21, 125. https://doi.org/10.1186/s43141-023-00591-w

Kilongosi Webale, M., 2024. Genotyping virulence and resistance profiles in salmonella isolated from diarrheic children in Nairobi city, Kenya. Gastroenterol. Hepatol. Bed Bench 17, 430–437. https://doi.org/10.22037/ghfbb.v17i4.3026

Kumar, R., Adeyemi, N.O., Chattaraj, S., Alloun, W., Thamarsha, A.K.A.N.W.M.R.K., Anđelković, S., Mitra, D., Gautam, P., 2025. Antimicrobial resistance in *Salmonella*: One

Health perspective on global food safety challenges. Sci. One Health 4, 100117. https://doi.org/10.1016/j.soh.2025.100117
Laing, C.R., Whiteside, M.D., Gannon, V.P.J., 2017. Pan-genome Analyses of the Species Salmonella enterica, and Identification of Genomic Markers Predictive for Species, Subspecies, and Serovar. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.01345
Lamichhane, B., Mawad, A.M.M., Saleh, M., Kelley, W.G., Harrington, P.J., Lovestad, C.W., Amezcua, J., Sarhan, M.M., El Zowalaty, M.E., Ramadan, H., Morgan, M., Helmy, Y.A., 2024. Salmonellosis: An Overview of Epidemiology, Pathogenesis, and Innovative Approaches to Mitigate the Antimicrobial Resistant Infections. Antibiotics 13, 76. https://doi.org/10.3390/antibiotics13010076
Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296. https://doi.org/10.1093/nar/gkab301
Lou, L., Zhang, P., Piao, R., Wang, Y., 2019. Salmonella Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. Front. Cell. Infect. Microbiol. 9, 270. https://doi.org/10.3389/fcimb.2019.00270
Miki, T., Shibagaki, Y., Danbara, H., Okada, N., 2009. Functional Characterization of SsaE, a Novel Chaperone Protein of the Type III Secretion System Encoded by Salmonella Pathogenicity Island 2. J. Bacteriol. 191, 6843–6854. https://doi.org/10.1128/JB.00863-09
Musa, L., Toppi, V., Stefanetti, V., Spata, N., Rapi, M.C., Grilli, G., Addis, M.F., Di Giacinto, G., Franciosini, M.P., Casagrande Proietti, P., 2024. High Biofilm-Forming Multidrug-Resistant Salmonella Infantis Strains from the Poultry Production Chain. Antibiotics 13, 595. https://doi.org/10.3390/antibiotics13070595
Mwambene, S.J., Mtenga, A.B., Chengula, A.A., Hoza, A.S., 2025. Detection and antibiotic resistance of Salmonella isolates from selected poultry farms in Dar es Salaam, Tanzania. Access Microbiol. 7, 000879.v5. https://doi.org/10.1099/acmi.0.000879.v5
Nazir, J., Manzoor, T., Saleem, A., Gani, U., Bhat, S.S., Khan, S., Haq, Z., Jha, P., Ahmad, S.M., 2025. Combatting Salmonella: a focus on antimicrobial resistance and the need for effective vaccination. BMC Infect. Dis. 25, 84. https://doi.org/10.1186/s12879-025-10478-5
Ngogo, F.A., Joachim, A., Abade, A.M., Rumisha, S.F., Mizinduko, M.M., Majigo, M.V., 2020. Factors associated with Salmonella infection in patients with gastrointestinal complaints seeking health care at Regional Hospital in Southern Highland of Tanzania. BMC Infect. Dis. 20, 135. https://doi.org/10.1186/s12879-020-4849-7
Oueslati, W., Ridha Rjeibi, M., Benyedem, H., Jebali, M., Souissi, F., Selmi, R., El Asli, M.S., Barguellil, F., Ettriqui, A., 2023. Serotype Occurrence, Virulence Profiles, Antimicrobial Resistance and Molecular Characterization of Salmonella Isolated from Hospitalized Patients with Gastroenteritis in Great Tunisia between 2010 and 2020. Antibiotics 12, 526. https://doi.org/10.3390/antibiotics12030526
Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinforma. Oxf. Engl. 31, 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Partridge, S.R., Kwong, S.M., Firth, N., Jensen, S.O., 2018. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin. Microbiol. Rev. 31, e00088-17. https://doi.org/10.1128/CMR.00088-17

Pavelquesi, S.L.S., de Oliveira Ferreira, A.C.A., Rodrigues, A.R.M., de Souza Silva, C.M., Orsi, D.C., da Silva, I.C.R., 2021. Presence of Tetracycline and Sulfonamide Resistance Genes in Salmonella spp.: Literature Review. Antibiotics 10, 1314. https://doi.org/10.3390/antibiotics10111314

Pettengill, J.B., Tate, H., Gensheimer, K., Hsu, C.H., Ihrie, J., Markon, A.O., McDermott, P.F., Zhao, S., Strain, E., Bazaco, M.C., 2020. Distribution of Antimicrobial Resistance Genes Across *Salmonella enterica* Isolates from Animal and Nonanimal Foods. J. Food Prot. 83, 295–304. https://doi.org/10.4315/0362-028X.JFP-19-310

Phoebe Lostroh, C., Lee, C.A., 2001. The Salmonella pathogenicity island-1 type III secretion system. Microbes Infect. 3, 1281–1291. https://doi.org/10.1016/S1286-4579(01)01488-5

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE 5, e9490. https://doi.org/10.1371/journal.pone.0009490

Ramatla, T., Tawana, M., Onyiche, T.E., Lekota, K.E., Thekisoe, O., 2021. Prevalence of Antibiotic Resistance in Salmonella Serotypes Concurrently Isolated from the Environment, Animals, and Humans in South Africa: A Systematic Review and Meta-Analysis. Antibiotics 10, 1435. https://doi.org/10.3390/antibiotics10121435

Salam, Md.A., Al-Amin, Md.Y., Salam, M.T., Pawar, J.S., Akhter, N., Rabaan, A.A., Alqumber, M.A.A., 2023. Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. Healthcare 11, 1946. https://doi.org/10.3390/healthcare11131946

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Siourimè, S.N., Isidore, B.O.J., Oumar, T., Nestor, B.I.H., Yves, T., Nicolas, B., Aly, S., 2017. SEROTYPING AND ANTIMICROBIAL DRUG RESISTANCE OF SALMONELLA ISOLATED FROM LETTUCE AND HUMAN DIARRHEA SAMPLES IN BURKINA FASO. Afr. J. Infect. Dis. 11, 24–30. https://doi.org/10.21010/ajid.v11i2.4

Tack, B., Vanaenrode, J., Verbakel, J.Y., Toelen, J., Jacobs, J., 2020. Invasive non-typhoidal Salmonella infections in sub-Saharan Africa: a systematic review on antimicrobial resistance and treatment. BMC Med. 18, 212. https://doi.org/10.1186/s12916-020-01652-4

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M., 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. 102, 13950–13955. https://doi.org/10.1073/pnas.0506758102

Tettelin, H., Riley, D., Cattuto, C., Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Curr. Opin. Microbiol., Antimicrobials/Genomics 11, 472–477. https://doi.org/10.1016/j.mib.2008.09.006

Yoshida, C.E., Kruczkiewicz, P., Laing, C.R., Lingohr, E.J., Gannon, V.P.J., Nash, J.H.E., Taboada, E.N., 2016. The Salmonella In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft Salmonella Genome Assemblies. PloS One 11, e0147101. https://doi.org/10.1371/journal.pone.0147101

Yu, X.-J., McGourty, K., Liu, M., Unsworth, K.E., Holden, D.W., 2010. pH sensing by intracellular Salmonella induces effector translocation. Science 328, 1040–1043. https://doi.org/10.1126/science.1189000

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. https://doi.org/10.1093/jac/dks261

Zhou, Z., Alikhan, N.-F., Mohamed, K., Fan, Y., Achtman, M., 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 30, 138–152. https://doi.org/10.1101/gr.251678.119