Current Research in Food Science 10 (2025) 100957

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

A pan-genome perspective on the evolutionary dynamics of polyphyly, virulence, and antibiotic resistance in *Salmonella enterica* serovar Mbandaka highlights emerging threats to public health and food safety posed by cloud gene families

Nai-peng Kan [a,b,1], Zhiqiu Yin [c,*,1], Yu-feng Qiu [b,1], Enhui Zheng [b], Jianhui Chen [b], Jianzhong Huang [a,**], Yuhui Du [d,***]

[a] National Engineering Research Center of Industrial Microbiology and Fermentation Technology, College of Life Sciences, Fujian Normal University, Fuzhou, 350117, PR China
[b] Fujian Provincial Center for Disease Control and Prevention, Fuzhou, 350012, PR China
[c] Department of Clinical Laboratory, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510700, Guangdong, PR China
[d] MOE International Joint Research Laboratory on Synthetic Biology and Medicines, School of Biology and Biological Engineering, South China University of Technology, Guangzhou, 510006, PR China

ARTICLE INFO

Keywords:
Mbandaka
Foodborne pathogen
Evolutionary dynamic
Polyphyletic
Virulence
Antimicrobial resistance

ABSTRACT

Salmonella enterica serovar Mbandaka, a prevalent foodborne pathogen, poses a threat to public health but remains poorly understood. We have determined the phylogenomic tree, genetic diversity, virulence, and antimicrobial resistance (AMR) profiles on a large genomic scale to elucidate the evolutionary dynamics within the Mbandaka pan-genome. The polyphyletic nature of this serovar is characterized by two distinct phylogenetic groups and inter-serovar recombination boundaries, that potentially arising from recombination events at the H2-antigen loci. The open pan-genome exhibited a flexible gene repertoire, with numerous cloud gene families involved in virulence and AMR. Extensive gene gain and loss observed at the terminal nodes of the phylogenetic tree indicate that Mbandaka individuals have undergone frequent gene turnover. The resulting changes in virulence and AMR genes potentially pose emerging threats to public health. We explored serovar conversion due to recombination of H-antigen loci, inter-serovar divergences in gene gain and loss, prophage-mediated acquisition of virulence factors, and the role of incompatibility group plasmids in acquiring resistance determinants as key molecular mechanisms driving the pathogenicity and antibiotic resistance of Mbandaka. Our work contributes to a comprehensive understanding of the complex mechanisms of pathogenesis and the ongoing evolutionary arms race with current therapeutic approaches in serovar Mbandaka.

1. Introduction

After diverging from a common ancestor with *Escherichia coli*, *Salmonella* has evolved into an invasive pathogen that causes foodborne diseases in both cold- and warm-blooded animals, including humans (Schaechter and Group, 2004; Kamali et al., 2024). Serological-based subdivisions are determined by somatic O antigens and flagellin H antigens (phase-1 and phase-2 flagellin), as identified through reactions with specific antisera. According to the White-Kauffmann-Le Minor serotyping scheme, *Salmonella enterica* is divided into over 2600 serovars, with 1500 belonging to *Salmonella enterica* subsp. *enterica* (Valizadeh et al., 2022). Serotyping has traditionally been the basis for surveillance of *Salmonella* due to its wide application in classification, identification, and epidemiological investigation. Serovars have been shown to correlate with host range and pathogenesis; however, variations in host specificity and/or virulence within a single serovar

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: yzq7873728@126.com (Z. Yin), hjz@fjnu.edu.cn (J. Huang), duyuhui_107@163.com (Y. Du).
[1] Nai-peng Kan, Zhiqiu Yin, Yu-feng Qiu contributed equally to this paper as first authors. Author order was determined by contribution.

https://doi.org/10.1016/j.crfs.2024.100957
Received 24 September 2024; Received in revised form 29 November 2024; Accepted 12 December 2024
Available online 13 December 2024

necessitate further genetical subdivision. In particular, some serovars are polyphyletic, deriving from multiple independent ancestors (Alikhan et al., 2018). For example, the serovar Paratyphi B, known to be polyphyletic, can cause diseases that range from self-limiting gastroenteritis to severe systemic infections with differentiated host susceptibility (Pinna et al., 2016). Thus, subtyping and subsequent clustering of polyphyletic serovar isolates with different genotypes and phenotypes are essential for successful investigation and epidemic tracing. However, many polyphyletic serovars await systematically analysis based on extensive genomic data.

Serovar Mbandaka was first isolated from a human salmonellosis case in the Belgian Congo in 1948, with an antigenic formula of '6,7,14: z10:e,n,z15' (Hoszowski et al., 2016). In recent years, Mbandaka has emerged as a common cause of salmonellosis outbreaks. In Europe, it is currently classified as one of the top ten serovars responsible for human salmonellosis and ranks fifth among the most frequently isolated serovars from cattle and poultry (Hoszowski et al., 2016) ("The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-Borne Outbreaks in 2011 Has Been Published.," 2013). It is also prevalent in various other settings, including poultry feeds, foods (such as dairy products, fish and seafood, cereals, vegetables and fruits, pastries and sweets), and the environment (Hoszowski et al., 2016). Additionally, the United States Centers for Disease Control (CDC) has repeatedly reported Mbandaka as being associated with multistate outbreaks (Keaton et al., 2022). The widespread dissemination of the Mbandaka serovar in poultry farms and hospitals in China has been observed recently (Yan et al., 2021) (Li et al., 2020). Furthermore, the emergence of multidrug resistant strains of Mbandaka in various countries poses an increasing threat to human health (Li et al., 2020) (Ma et al., 2023). Given the ongoing threat of virulence and AMR to public health and food safety, there is a growing focus on genomic insights into this serovar (Hayward et al., 2013). The rapid increase in genome data has intensified efforts to understand the phylogeny, host adaptation, pathogenesis, and AMR profiles of Mbandaka (Brockhurst et al., 2019) (Hosseini et al., 2024). In a previous genomic survey, Antony et al. characterized the population structure, AMR, and virulence profiles of Mbandaka using over 400 genomes (Antony et al., 2020). Recent studies have revealed the genetic characteristics of the global prevalent sequence type 413 (ST413) of Mbandaka, including its multidrug resistance, conserved virulence, and bovine-related host adaptation (De Sousa Violante et al., 2023) (Benevides et al., 2024). Although Mbandaka has been the subject of several comparative studies, its inter-serovar phylogenetic relatedness, evolutionary dynamics, and genetic profiles of pathogenesis and AMR remain incompletely understood at the broad genomic level. Gains and losses of genes are major forces driving the evolution of virulence and antimicrobial resistance (AMR) in *Salmonella* (Porwollik and McClelland, 2003). Pan-genome analyses have proved powerful for characterizing genetic diversity and evolutionary dynamics in a certain pathogen, particularly the genetic innovations contributing to emerging virulence and AMR (Yin et al., 2022, 2024).

This would provide a novel, global overview that can be used to investigate serovar Mbandaka. In this study, we collected 1452 high-quality Mbandaka genomes to construct a phylogenomic overview that reveals the accurate genetic relatedness among Mbandaka strains. The work investigates the presence of inter-serovar recombination boundaries and explores how strains from multiple independent ancestors are carrying this serovar. Genetic diversity and evolutionary dynamics across serovars were characterized through pan-genome and gene gain and loss analyses. Virulence and antimicrobial resistance-related genetic profiles were also assessed.

2. Materials and methods

2.1. Isolation and genome sequencing

A *Salmonella* strain, KNP01, was isolated from the faecal sample of a healthy carrier at a public institution in Fuzhou, Fujian Province, China, in 2000. The isolate was serotyped using an agglutination assay with anti-serum obtained from the Chengdu Institute of Biological Products (CDIBP, China). The isolate was cultivated in Luria-Bertani broth at 37 °C for 12 h. Single colonies were inoculated into the LB medium and cultivated at 37 °C for 18 h. Antimicrobial susceptibility testing of strain KNP01 was performed by broth microdilution using dehydrated panels AST-N334 and AST-N335 (Biomerieux, Vitek2) according to standard protocols. The interpretative criteria were from the Clinical and Laboratory Standards Institute (CLSI) document M100-28th edition (Wayne, n.d.).

The genomic DNA of KNP01 was extracted from overnight cultures grown at 37 °C in Luria–Bertani broth under 180 rpm shaking conditions by using a TIANamp bacteria DNA kit (Tiangen Biotech, China) according to the manufacturer's protocol. The Qubit Broad Range assay kit (Invitrogen, United States) was used for quantification. Qualified genomic DNA was sequenced using the PacBio Sequel platform (SMRT sequencing) and Illumina NovaSeq 6000 sequencer (PE150 with 2 × 150-bp paired-end reads). A hybrid assembly was preformed that combined high quality Illumina short reads and filled in the gaps with PacBio long reads. Continuous PacBio long reads were used for de novo assembly using Falcon v0.3.0 (Chin et al., 2016). Raw data from the Illumina platform were filtered using FASTP v0.20.0 (S. Chen et al., 2018) according to the following standards: 1) remove reads with ≥10% unidentified nucleotides (N); 2) remove reads with ≥50 % bases with Phred quality scores ≤20; 3) remove reads aligned to the barcode adapter. After filtering, the resulting clean reads were used to correct the genome sequences to improve the quality of the assembly and to determine the final genome sequences using Pilon v1.23 (Walker et al., 2014). A segment of sequence was taken from the end and the beginning of the assembly, the two segments were spliced together, and the original sequencing reads were compared to the assembly to determine if the sequence looped or not, i.e., if there were gaps at the end. The KNP01 genome has been deposited in the National Center for Biotechnology Information (NCBI) RefSeq database (Accession number: CP113364.1). It was estimated to have 100.0% completeness with 0.08% contamination using checkM v1.0.13 software (Parks et al., 2015). The online interface of Type Strain Genome Server (TYGS) was utilized for accurate, genome-based taxonomy (Meier-Kolthoff and Göker, 2019). Functional assignment of the Clusters of Orthologous Genes (COG) was conducted using the eggNOG-mapper 2.1.9 software (Huerta-Cepas et al., 2017).

2.2. Genome collection and filtering

A search for public genome data of serovar Mbandaka was conducted on the online interface of the EnteroBase *Salmonella* Database (https://enterobase.warwick.ac.uk/species/index/senterica) (Alikhan et al., 2018) using "Mbandaka" as the keyword and filtering for coverage ≥100. All genomes collected were downloaded on June 21, 2023. Serotyping of these genomes was verified using both SeqSero v2.0 (Zhang et al., 2019) and SISTR v1.1.1 (Yoshida et al., 2016). Only strains with consistent serotyping results across both software tools were retained. Multilocus sequence typing (MLST: Achtman 7 Genes) were conducted using the mlst v2.23.0 software (https://github.com/tseemann/mlst). In total, 1449 genomes were collected for subsequent analyses. Additionally, 83 high-quality reference genomes of *S. enterica* were collected, which together represent 48 serovars and 71 MLST profiles (Cherchame et al., 2022). The final dataset for *S. enterica* (n = 1533) included the newly sequenced KNP01 genome, 83 high-quality reference genomes (including two Mbandaka genomes, CP019183.1

and CP022489.1) from the NCBI GenBank database, and 1449 Mbandaka genomes from the EnteroBase database (Table S1).

2.3. Pan-genome and phylogenetic analysis

Unified gene finding and re-annotation of all genomes were conducted using Prokka v1.14.5 software (Seemann, 2014). The resulting GFF files were analyzed with Panaroo v 1.3.4 using the default settings to determine orthologous groups of gene families within the pan-genome (Tonkin-Hill et al., 2020). Single-nucleotide polymorphisms (SNPs) from 1438 single-copy core gene families were extracted based on protein accession numbers and aligned using MAFFT v7.508 software (Katoh et al., 2019). To avoid phylogenetic confusion, we identified and excluded potential recombinational regions from the SNP set using ClonalFrameML v1.12 software (Didelot and Wilson, 2015). The final set of SNPs was analyzed using the maximum-likelihood (ML) method in IQ-TREE v 2.2.5 with the general time reversible (GTR) model and 1000 bootstrap replicates (Minh et al., 2020). Pairwise SNP distances were calculated using snp-dists v0.8.2 (https://github.com/tseemann/snp-dists). For the phylogenetic analysis of O-, H1-, and H2-antigens, core genes were extracted from the corresponding gene clusters within the serogroup C1 reference genomes (Yin et al., 2020). The extracted sequences were aligned using MAFFT v7.508 (Katoh and Standley, 2013), and phylogenetic trees were generated using MEGA v11 with the GTR model and 100 bootstrap replicates (Tamura et al., 2021).

2.4. Distinguishing recombination boundaries

To delineate recombination boundaries based on the capacity for gene exchange, as suggested by the Biological Species Concept (BSC), we measured the ratio of homoplasies/recombinations (h: those not vertically inherited from a common ancestor) to non-homoplasies/mutations (m: those due to new or vertically inherited mutations) using ConSpeciFix v1.3.0 (Bobay et al., 2018). To expedite the computational processing, these genomes were dereplicated using a 99.9% average nucleotide identity (ANI) cutoff using dRep v3.5.0 (Olm et al., 2017) with parameters (-pa 0.95 -sa 0.999 -cm larger). The dereplicated genomes subsequently served as input for ConSpeciFix. To assess population membership, we calculated the h/m ratios for the genome sets of Mbandaka-A (dereplicated genomes) and Mbandaka-A + -B (CP019183.1). When recombination boundaries are present between Mbandaka-A and Mbandaka-B, polymorphisms confined to the non-recombining genome(s) reduce the h/m ratios. The plots may show a sharp reduction in the *h/m* ratios, representing subsample iterations with and without the non-recombining genome(s), respectively.

2.5. Comparative genomic analysis

Mobile genetic elements (MGEs) were predicted using the online interface of VRprofile2 (Wang et al., 2022). The prophages were identified using the online interface of the PHAge search tool - Enhanced Release (PHASTER) (Wishart et al., 2023). The fastANI v1.33 (Jain et al., 2018) was used to calculate the average nucleotide identity (ANI) values. The Deeplasmid tool was used for distinguishing plasmids from assembled contigs or scaffolds (Andreopoulos et al., 2022). PlasmidFinder v.2.0 (Center for Genomic Epidemiology) was employed for screening plasmid replicons. AMR phenotype prediction were carried out using ResFinder v4.3.1 (Bortolaia et al., 2020), with a cutoff at 80% nucleotide identity and 60% nucleotide coverage. AMR genes were detected using Abricate v1.0.1 (https://github.com/tseemann/abricate) with the Comprehensive Antibiotic Resistance Database (CARD) v3.1.2 (Alcock et al., 2020).

2.6. Statistical analysis

The statistical analyses of homologous gene or gene clusters related to SPI-1 to SPI-22, prophages, fimbrial operons, and effectors were performed using the LS-BSR tool with default parameters (Sahl et al., 2014). The statistical analyses of pangenome comparisons were determinated using Panstripe v0.3.0 (Tonkin-Hill et al., 2023), which employs interaction terms to compare the relationship of the covariates with gene gain and loss between subgroup A and B pangenomes. A significant *P*-value for the "tip" term indicates that there is a different rate of gene exchange at the tips of the phylogeny compared to the internal branches. A significant p-value for the "core" term indicates that there is a significant association between the branch length of the core genome and the number of gene exchange events.

3. Results and discussion

3.1. Isolation and whole-genome sequencing of a human isolate

In this study, the strain KNP01 was isolated from the faecal sample of a public health staff member (a healthy carrier) in Fuzhou, Fujian Province, China, in 2000. Antimicrobial susceptibility testing showed that KNP01 was resistant to piperacillin and tetracycline, yet susceptible to other antibiotics tested (Table S2). The complete genome of KNP01 comprised one circular chromosome with no plasmid detected. It consists of 4,756,748 base pairs (bp) with a guanine-cytosine (GC) content of 52.2%, 4493 coding sequences (CDSs), 22 ribosomal RNA (rRNA) genes, and 84 transfer RNA (tRNA) genes (Fig. 1A). A total of 2927 (65.1%) CDSs were classified into COG categories (Fig. S1A). The genome-based taxonomic analysis using the TYGS assigned strain KNP01 to *S. enterica* subsp. *enterica* (Fig. S1B). *In silico* serovar prediction identified KNP01 as serovar Mbandaka with the antigenic formula '7: z10:e,n,z15' and '6,7,14:z10:e,n,z15'. The MLST (Achtman 7 Gene) profile of KNP01 was assigned as ST413 (*dnaN*: 70; *aroC*: 15; *hisD*: 78; *sucA*: 6; *purE*: 113; *thrA*: 68; *hemD*: 93), also corresponding to serovar Mbandaka. Although at least two thousand Mbandaka genomes are available as of June 21, 2023, only 17 complete genomes are present. Thus, unveiling the complete genome of the human isolate KNP01 could broaden our understanding of the genomic diversity within this serovar.

3.2. General information of serovar Mbandaka genomes

To deepen our understanding of the genomic features contributing to the foodborne illness and the spread of serovar Mbandaka, we collected 1452 serovar-checked genomes, including the KNP01 genome, two reference complete genomes (CP019183.1 and CP022489.1) from the NCBI GenBank database, and 1449 genomes from the EnteroBase database. The average genome size for serovar Mbandaka was 4778.7 ± 93.1 kb, containing 4470.6 ± 103.9 protein-coding genes. The genomes exhibited a minor variation in GC content (52.1 ± 0.1%) (Fig. 1B). In terms of the MLST (Achtman 7 Gene) profile, ST413 was the most prevalent sequence type (n = 1370; 94.4%), followed by ST1602 (n = 44; 3.0%), aligning with previous studies (Fig. S2A) (Antony et al., 2020) (De Sousa Violante et al., 2023). Compared to the report by Antony et al., which identified ST2238, ST2404, and ST2444 (Antony et al., 2020), our collection displayed greater STs diversity, encompassing additional STs such as ST1611, T2141, ST3016, ST3760, ST4754, ST6107, ST8708, ST9156, ST9414, and ST10539. These rare types were found in less than ten genomes each. The genome data of these diverse STs may facilitate an understanding of the genomic diversity of serovar Mbandaka. Over half of the isolates within this serovar were sourced from host-associated sources, with 419 (28.9%) from "Human" and 343 (23.6%) from "Animal" (Fig. S2B). Indeed, this serovar is widely recognized as a common causative agent of salmonellosis outbreaks in humans and poultry globally, especially within the European Union (Cheng et al., 2019) (Benevides et al., 2024). It is also

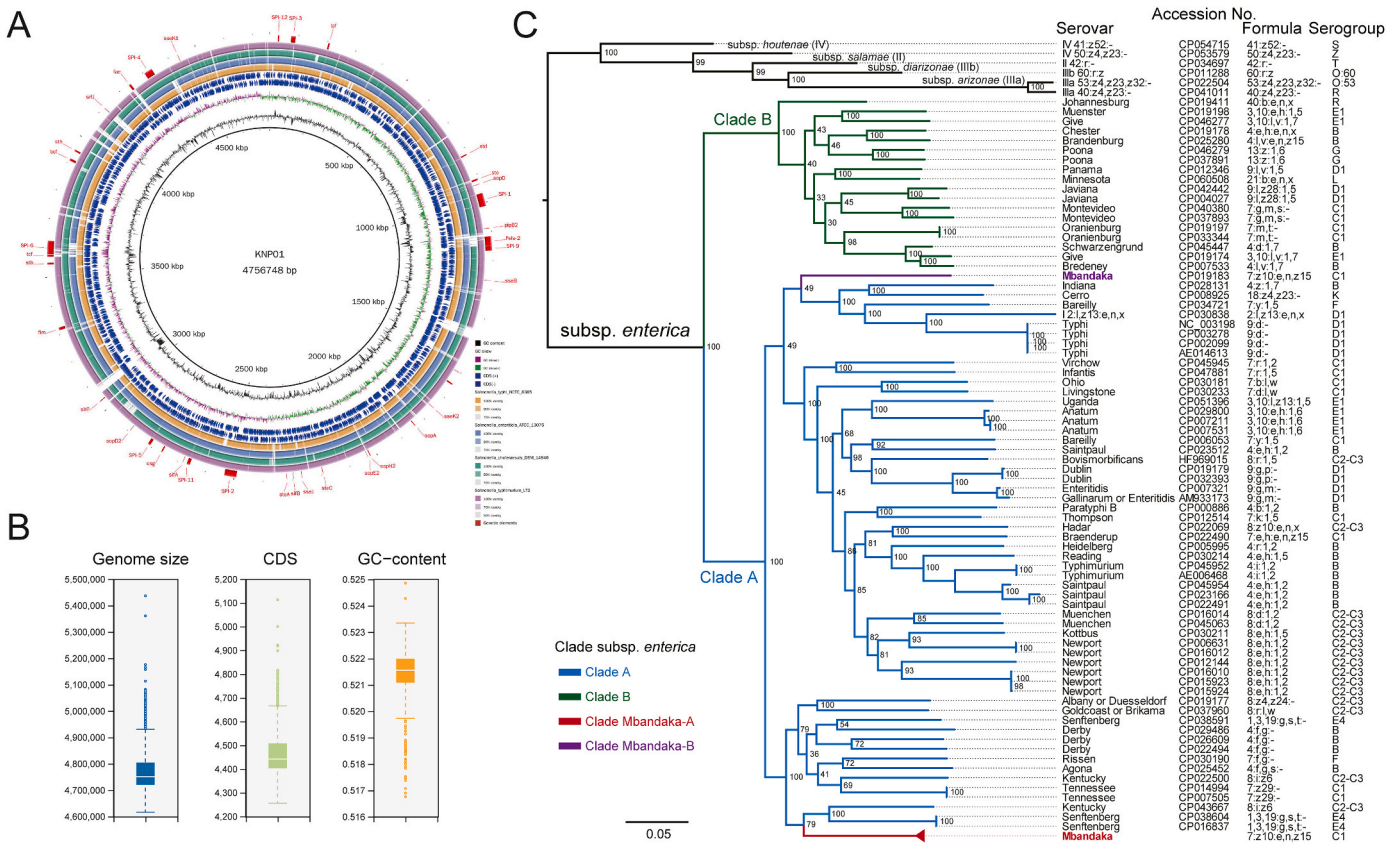

Fig. 1. A. Circular representation of the KNP01 genome. B. General genomic characteristics of 1452 Mbandaka genomes. The box plots illustrate the distribution of genome sizes (bp), the number of CDS, and GC content, respectively. C. Core genome phylogeny of *S. enterica*. The maximum likelihood (ML) tree was constructed based on single-nucleotide polymorphisms (SNPs) across 1438 single-copy core gene families shared by 1533 *S. enterica* genomes.

noteworthy that 221 (15.2%) and 113 (7.8%) were isolated from the "Environment" and "Food", respectively, suggesting that Mbandaka is frequently excreted by livestock into the environment (Achtman et al., 2020).

3.3. Phylogenomic analysis revealed the polyphyletic nature of serovar Mbandaka and intra-serovar recombination boundaries

To further investigate the genetic relatedness among Mbandaka members, we integrated these genomes with 81 reference genomes from other *S. enterica* serovars, creating a dataset of 1533 genomes for phylogenomic analysis. A ML phylogeny was constructed based on SNPs from the 1438 single-copy core gene families shared by all 1533 genomes. As shown in Fig. 1C, the topology of the core genome tree aligns with previous studies (Timme et al., 2013) (Worley et al., 2018), showing subsp. *enterica* deeply diverging from other subspecies into primary clades A and B. The Mbandaka members were categorized into two distinct evolutionary clades, Mbandaka-A and -B, both nested within primary clade A. Nearly all members grouped together to form a monophyletic clade (Mbandaka-A), positioned on a long primary branch adjacent to the members of serovars Kentucky and Senftenberg (Fig. 1C). Most genomes within Mbandaka-A clustered tightly with very short branch lengths, indicating minimal genetic diversity (Fig. S3). For the Mbandaka-B clade, a single strain (ATCC 51958) with the ST3016 MLST profile formed a deep clade near the members of serovars Indiana, Cerro, Typhi (Fig. 1C). These results underscore the polyphyletic nature of Mbandaka and clearly indicate that strains carrying this serovar do not share a recent common ancestor. This finding aligns with a previous phylogenetic analysis of *S. enterica* that identified polyphyletic nature in Mbandaka (Cherchame et al., 2022), even though this was not observed in several comparative genomic analyses (Antony et al., 2020; De Sousa Violante et al., 2023; Benevides et al., 2024).

A polyphyletic serovar, originating from multiple independent ancestors, may confound epidemiological investigations, as serotyping cannot accurately reflect the genetic differentiation among these strains (Alikhan et al., 2018) (Pinna et al., 2016). The BSC defines species boundaries based on reproductive isolation and the potential for gene exchange. Cobo-Simón et al. have shown that *Salmonella* serovars, while strictly clonal, experience sufficient levels of homologous recombination to establish species barriers, even amidst high nucleotide identity (Cobo-Simón et al., 2023). To determine whether recombination boundaries (as per the BSC) exist, we assessed the extent of homologous recombination within and between the Mbandaka clades using ConSpeciFix (Bobay et al., 2018). The ratio of homoplasies/recombinations (h) to non-homoplasies/mutations (m) polymorphisms across the entire set of core genes was calculated. Mbandaka-A displayed high h/m values (0.039 ± 0.017), suggesting prevalent recombination among its genomes. However, adding the ATCC 51958 genome, which represents Mbandaka-B, to the Mbandaka-A dataset significantly reduced the h/m values (0.009 ± 0.012) (Fig. 2A), indicating reproductive isolation between the Mbandaka-A and -B lineages and suggesting they should be regarded as distinct biological lineages. These findings indicate that recombination barriers exist between the polyphyletic lineages, even with a high degree of nucleotide identity and the same antigenic formula.

3.4. The polyphyletic nature of serovar Mbandaka appears to result from recombination events at the H2-antigen loci

Previous studies have suggested that the common serovars across distinct phylogenetic lineages may result from recombination at the genetic loci responsible for the serotype formula (O-, H1-, and H2-

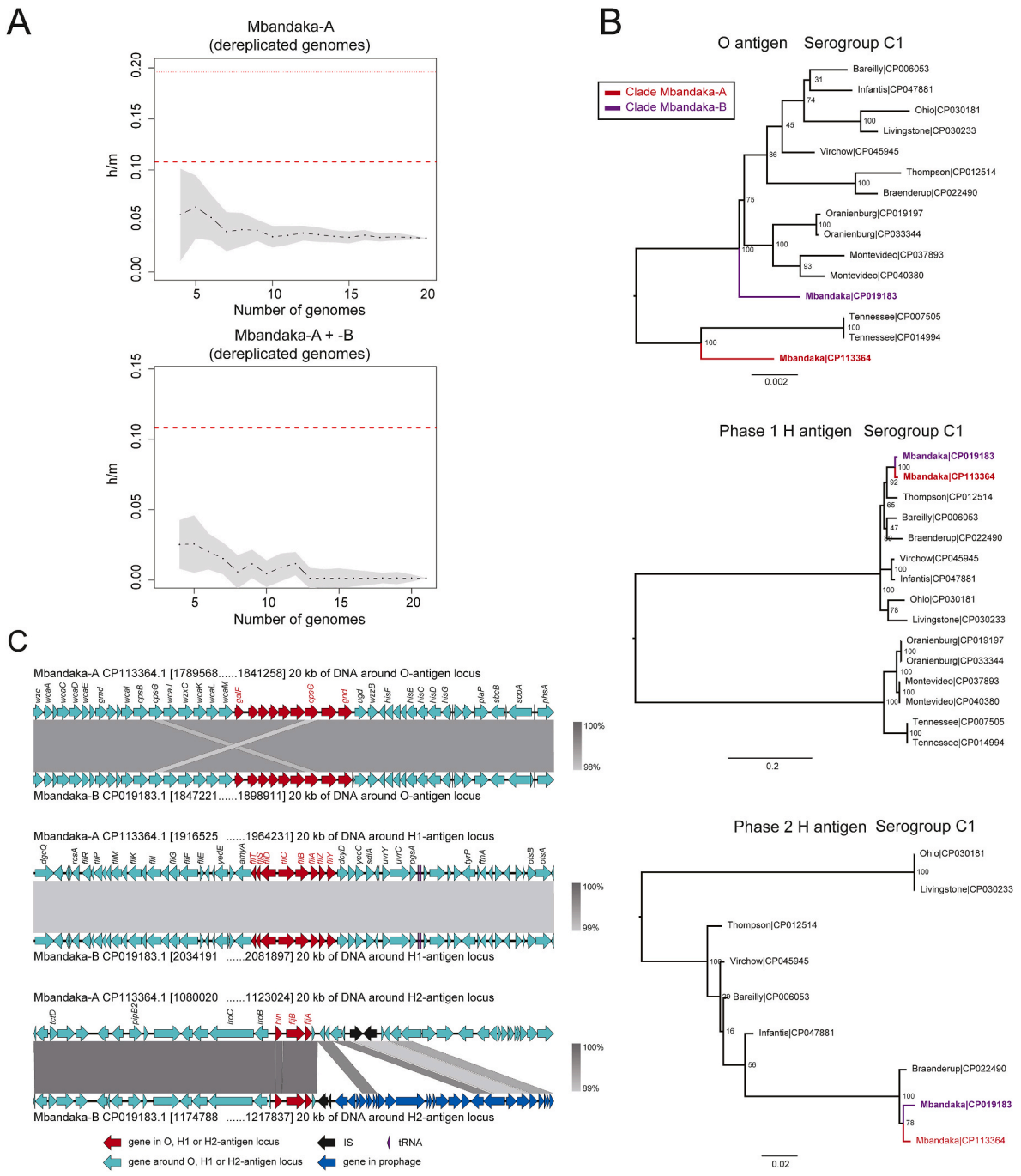

Fig. 2. A. ConSpeciFix output graphs display the homoplasies (h) and mutations (m) ratios for Mbandaka-A genomes (top) and for groups of Mbandaka genomes including the Mbandaka-B genome (bottom). For each subsample size of genomes on the X-axis, black dots represent the median h/m ratio, with grey areas indicating the standard deviation. Dashed and dotted lines indicate average and maximal h/m ratios, respectively. B. Phylogenetic analysis of O-, H1- and H2-antigens. ML trees were generated based on the nucleotide sequences of core genes within O-, H1- and H2-antigen loci of the reference genomes that belong to serogroup C1. The primary node values of the trees are the bootstrap values (100 replicates). C. Comparisons of the genetic organization of O-, H1- and H2-antigen loci between the Mbandaka-A (CP113364.1) and -B (CP019183.1) reference genomes. Genes are denoted by arrows and colored based on their functional classification, with grey shading indicating homologous regions.

antigen) (Pinna et al., 2016) (Yin et al., 2020). To explore the cause of the polyphyletic nature of serovar Mbandaka, we constructed phylogenies for the O-, H1-, and H2-antigen of Mbandaka-A, -B, and other C1 serovars, utilizing reference genomes (Fig. 2B). A striking similarity was observed in the topology of the O-antigen tree when compared with the core genome tree, while Mbandaka-A and -B demonstrated low relatedness in genetic aspects, implying a distinct evolutionary history for the O-antigen. High homology and a similar organizational structure were observed in the O-antigen loci and surrounding regions of Mbandaka-A and -B (Fig. 2C), suggesting that the O-antigen gene cluster shares an evolutionary history with the core genome. In contrast, Mbandaka-A and -B were found to cluster together in the H1- and H2-antigen trees, indicating common flagella loci (Fig. 2B). In the H1-antigen phylogeny, each individual clade of serovars exhibited relatively short branch lengths, except for two primary clades. The H1-antigen loci and surrounding regions of Mbandaka-A and -B also exhibited nearly identical

homology and an identical organizational structure (Fig. 2C). Meanwhile, no significant rearrangements or MGEs were observed. Therefore, the common H1-antigen with limited diversity may be due to selective pressures rather than recombination events. In the case of the H2-antigen, several insertion sequences (ISs) were found flanking the corresponding loci in both Mbandaka-A and -B (Fig. 2C). Additionally, the H2-antigen locus of Mbandaka-B (ATCC 51958) was found near a prophage region, designated as PHAGE_Salmon_Fels_2 (Fig. S4). Yates et al. recently reported that H2-antigen loci are located at the end of prophage regions, such as Salmon_SEN_8 and Salmon_Fels_2, in several serovars, suggesting horizontal transfer of this locus or integration of these prophages adjacent to *fljB* (Yates et al., 2024). Thus, the lack of diversity within the H2-antigen and the polyphyletic nature of Mbandaka may be attributed to recombination events at the H2-antigen loci, potentially mediated by MGEs.

Additionally, in this study, Mbandaka-B was a singleton, with no close relatives identified. However, two other strains, EA4368AA and WA9982AA, which were excluded from our collection due to low genome quality, share the same ST (ST3016) and are recorded in the EnteroBase database with the Mbandaka serovar. This suggests that Mbandaka-B is not merely a consequence of a sequencing or serotyping error. Distant evolutionary groups within polyphyletic serovars have been associated with divergences in pathogenesis, adaptation, and epidemicity (Pinna et al., 2016) (Sangal et al., 2010). Future studies should focus on the phenotypic differences corresponding to the two groups. Exploring virulence-related targets and quantifying key genomic differences between the two groups could provide a foundation for developing simple and rapid tests for clinical investigation and epidemic surveillance. Accordingly, it is essential to collect more diverse Mbandaka-B isolates, such as those belonging to ST3016, and to analyze

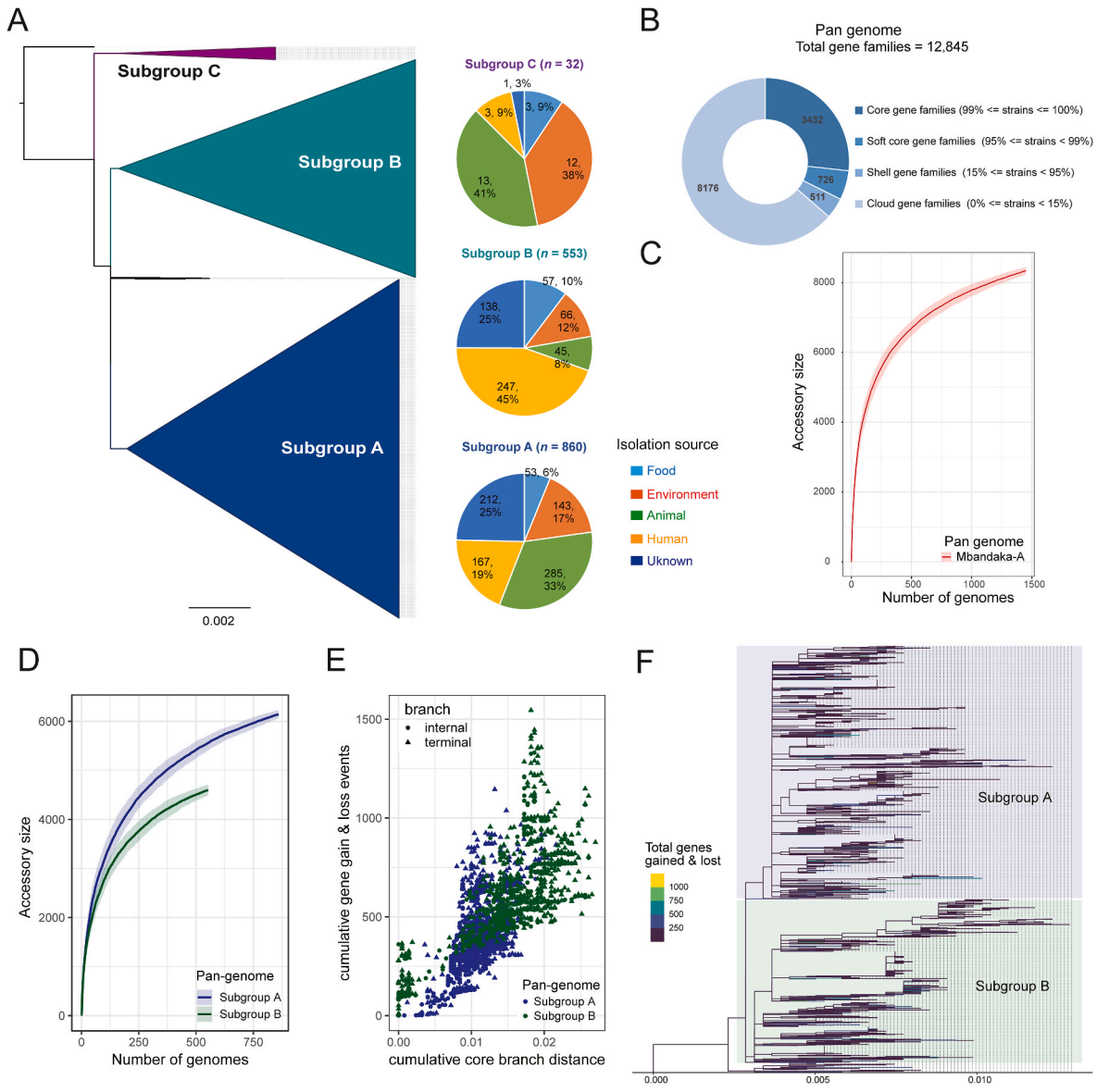

Fig. 3. A. Core genome phylogeny of Mbandaka-A. The ML tree was constructed based on SNPs across 2255 single-copy core gene families shared by 1451 Mbandaka-A genomes. The pie charts represent the percentage of isolation sources for members in subgroups A, B, and C, respectively. B. Pan-genome analysis of the Mbandaka-A genomes identified 12,845 gene families, with 3432 genes (29%) classified as core genes, present in ≥99% of the genomes. C. The accumulation curve of accessory genome size in the Mbandaka-A pan-genome inferred using Panaroo. The X-axis represents the number of genomes, while the Y-axis represents the number of genes. D. The accumulation curves of accessory genome size in the pan-genomes of subgroups A and B (Mbandaka-A). E. The cumulative number of gene gain and loss events versus the cumulative branch length starting from the root node of subgroups A and B in the Mbandaka-A core genome tree. F. Core genome phylogeny of Mbandaka-A with branches colored based on the number of gene gain and loss events inferred by maximum parsimony.

their genomic data for future research.

3.5. Phylogenomic analysis of Mbandaka-A demonstrated little diversity

Given that the representation of the vast majority of members, subsequent phylogenetic and pan-genomic analyses focused on Mbandaka-A. The core genome tree revealed that Mbandaka-A members formed a tight group with short branches (Fig. 3A). The majority of members were divided into three subgroups and six singletons, with the main branches between subgroups also being relatively short. Subgroups A and B represented two major subclades, comprising 553 and 860 strains, respectively, while subgroup C comprised 32 strains. The pairwise SNPs differences within the core gene families among Mbandaka-A members did not exceed 1345, with a mean of 39.6 ± 46.3. Members within subgroups A and B showed a mean SNP difference of 41.3 ± 17.4 and 29.3 ± 15.7 SNPs, respectively. The pairwise SNP difference between subgroups A and B was also small, averaging 42.8 ± 16.1. Additionally, Mbandaka-A members shared high whole-genome ANI values, exceeding 99.5%. The ANI values from comparisons between subgroups A and B were all above 99.9%. The ANI values between the subgroup A and B members were also greater than 99.9%. These results indicated that there were relatively little genetic divergences among most members of Mbandaka-A, even between subgroups.

In terms of isolation sources, the constituent strains of subgroups A and B showed a similar distribution patterns across “Food”, “Environment”, “Unknown” and host-associated sources (Fig. 3A). Within the host-associated sources, subgroup A members were more frequently isolated from “Animal”-related samples (subgroup A: $n = 285$; 33.1%; subgroup B: $n = 45$; 8.1%). In contrast, subgroup B showed a closer association with “Human”-related sources (subgroup B: $n = 247$; 44.7%; subgroup A: $n = 167$; 19.4%).

3.6. Pan-genome architecture of Mbandaka-A represented by cloud gene families

The emergence of pathogenic strains is primarily attributed to genetic innovation, particularly the acquisition of virulence and AMR genes through horizontal gene transfer (HGT) (Balasubramanian et al., 2022). Pan-genome analysis is a powerful approach for elucidating the potential genetic innovations within a bacterial species (Yin et al., 2019) (Yuan et al., 2020). The pan-genome analysis of Mbandaka-A revealed a pan-genome comprising 12,845 homologous gene families, including 3432 core, 726 soft-core, 511 shell, and 8176 cloud gene families (Fig. 3B). Cloud gene families, found in ≤15% of the analyzed genomes, made up the majority (63.7%) of the Mbandaka-A pan-genome. As more genomes are analyzed, the accessory size of the pan-genome shows a clear, linear upward trend (Fig. 3C), suggesting that Mbandaka-A possesses a gene pool capable of continuously acquiring exogenous genetic elements and has the potential to discover additional genes as new genomic sequences become available. Cloud gene families, associated with HGT, typically confer selective advantages to bacteria, such as niche adaptation, antibiotic resistance, and pathogenicity (Tettelin et al., 2008; Vernikos et al., 2015). Considering that *Salmonella* pathogenicity is largely due to the acquisition of more virulence factors (Worley et al., 2018) (Cobo-Simón et al., 2023), the flexibility of the Mbandaka-A pan-genome likely results in a diverse virulence gene repertoire, driven by the prevalence of horizontally acquired genes.

The generalist ecological lifestyle of the microbe is facilitated by a diverse accessory genome, shaped by gene gain and loss, and has important implications for the evolution of virulence and antibiotic resistance (Tonkin-Hill et al., 2023; Gladstone et al., 2021). We estimated the evolutionary dynamics of Mbandaka-A by estimating the rates of gene gain and loss within the Mbandaka-A pan-genome, using the core genome phylogeny and gene presence/absence matrices. The intensity of gene gain/loss, as mapped onto the phylogenetic tree, showed that Mbandaka-A experienced extensive gene gain/loss at its terminal nodes relative to its internal nodes (Fig. S5A). This suggests that the individuals have undergone frequent gene turnover, potentially accounting for the predominance of cloud gene families in the Mbandaka-A pan-genome.

3.7. The divergences in several characteristics between the subgroup A and B pan-genomes

For the major subgroups within Mbandaka-A, we constructed and compared the pan-genomes of subgroups A and B. Subgroup A pan-genome comprised 10,582 gene families, with 3779 (35.7%) core, 438 (4.1%) soft-core, 392 (3.7%) shell, and 5973 (56.4%) cloud gene families (Fig. 3C). For subgroup B, the pan-genome included 9088 gene families, consisting of 4066 (44.7%) core, 76 (0.8%) soft-core, 622 (6.8%) shell, and 4324 (47.6%) cloud gene families (Fig. S5B). Subgroup A had a larger pan-genome, attributed to a higher presence of cloud gene families. Furthermore, the upward trend of the pan-genome accumulation curve for subgroup A is steeper than that for subgroup B (Fig. 3D), indicating that greater accessory genome diversity in subgroup A. The pan-genome accumulation curve reflects the underlying population structure of the datasets; thus, the differences are likely driven by sampling biases, given that subgroup A encompasses a more diverse set of genomes from various poultry sources.

To further identify differences in pan-genome dynamics between subgroups A and B, we estimated the rates of gene gain and loss for the subgroup A and B nodes in the Mbandaka-A core genome tree using Panstripe. Subgroup A and B nodes exhibited similar rates of gene exchange, with a *P*-value of 0.659 for the “core” term in Panstripe output. However, the terminal nodes of subgroup B displayed a higher rate of gene exchange compared to those of subgroup A ($P = 0.0264$ for the “tip” term in Panstripe output) (Fig. 3E and F), suggesting that the number of genes involved in each gain and loss event in the tips of the phylogeny differs significantly between the two subgroups. This is typically attributed to variations in the gain and loss of highly mobile elements, which may not persist long enough to appear across multiple genomes (Tonkin-Hill et al., 2023). Subgroup B members were more closely associated with “Human”-related source, suggesting that the higher rate of gene exchange may reflect additional selection pressures within the human environment. It is evident that *S. enterica*, as an enteropathogen, can form persisters that survive antibiotic therapy in various host tissue and act as long-lived reservoirs of plasmid donors or recipients, facilitating the dissemination of promiscuous AMR plasmids in the gut (Bakkeren et al., 2019). Furthermore, the dynamics of gene gain and loss suggest that the more open pan-genome diversity of subgroup B is indicative of greater underlying population structure diversity and potential sampling bias within its genome dataset.

Overall, while small differences in SNP distances were observed between the subgroups A and B datasets, several divergences in pan-genome characteristics were noted, indicating distinct evolutionary dynamics within the accessory genomes of the two subgroups.

3.8. Virulence-related genotypic profile in Mbandaka shows a similar gene repertoire, but mobile virulence-related elements mediated by HGT cannot be ignored

In this study, we identified and characterized potential virulence-related genetic elements in the Mbandaka genomes, including *Salmonella* pathogenicity islands (SPIs), prophages, fimbrial operons, and type III secretion system (T3SS) effectors. Virulence factors encoded by SPI genes are believed to manipulate host cellular mechanisms and influence the host specificity of different *S. enterica* serovars (Eswarappa et al., 2008). The majority of Mbandaka genomes contained intact SPI-1 to −6, SPI-9, SPI-11, and two copies of SPI-12, which represent the core genome (Fig. 4). The intact SPI-13 and SPI-14 was detected in only one genome (SAL_JB2167AA_AA). The detection of SPI-1, SPI-2, SPI-4, SPI-5, SPI-13, and SPI-14 is consistent with the results of Antony et al.

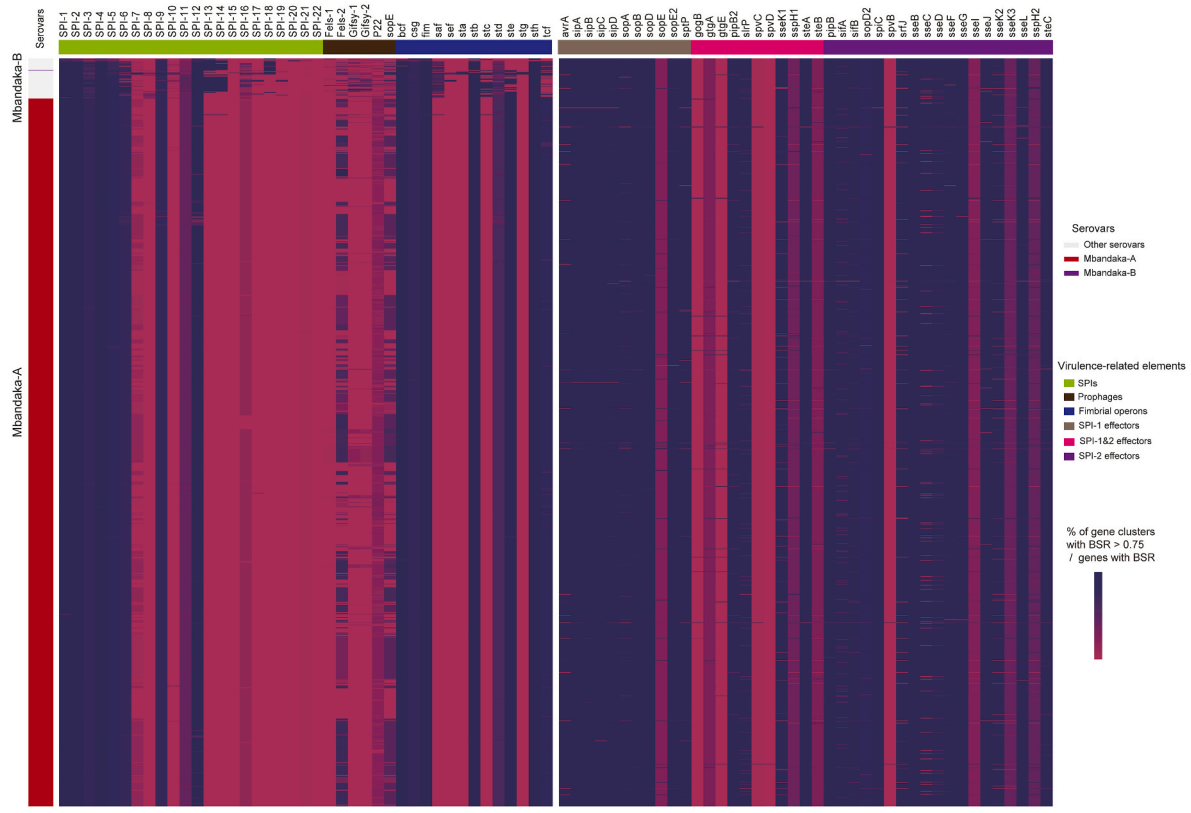

Fig. 4. Heatmap of the distribution of SPIs, prophage, fimbrial operons, and effectors in the *S. enterica* genomes. The color coding for gene clusters (SPIs, prophages, and fimbrial operons) is based on the percentage of genes in a cluster present in a genome, defined as the Blast score ratio (BSR) of the query gene >0.75. Effector colors reflect BSRs from genomic screens against the effector database.

(2020). SPI-3, SPI-6, SPI-9, SPI-11, and SPI-12 are in agreement with a previous comparative genomic analysis of two Mbandaka strains (Hayward et al., 2013). The intact SPI-18 was identified in Mbandaka-B, while no such detection was made in any Mbandaka-A members.

Prophages have been reported to influence the metabolic, virulence, and resistance characteristics of *Salmonella* (Yates et al., 2024). Incomplete Fels-2 and SopE Phi prophages were sporadically found in more than half of the Mbandaka genomes, indicating that these mobile elements are actively integrated in Mbandaka. The Fels-2 prophage carries the AMR gene *sdiA*, which encodes an antibiotic efflux pump, and plays a crucial role in the evolution of *Salmonella* surface antigens by facilitating the horizontal acquisition of the H2-antigen locus (Yates et al., 2024). The P2-like prophage SopE Phi carries the effector-coding gene *sopE*, which encodes a guanine nucleotide exchange factor that modulates host responses (Mirold et al., 2001). Furthermore, we identified eight fimbrial operons that are present in all Mbandaka-A genomes, including *bcf*, *csg*, *fim*, *stb*, *std*, *ste*, *sth*, and *tcf* (Fig. 4). The *saf* and *stc* operons were found in only one strain (JB2167AA), suggesting an individual HGT event. In contrast, in addition to also containing common *bcf*, *csg fim*, and *ste*, the Mbandaka-B genome uniquely contained *stc*, but *stb* and *tcf* were not detected.

SPI-1 and -2, as core components of the Mbandaka-A pan-genome, encode two T3SSs responsible for transporting a panel of effectors across bacterial and host cell membranes, facilitating bacterial entry and survival within host cells (e.g., intestinal epithelial cells and macrophages) (Sabbagh et al., 2010). In this study, we examined 41 effector-coding genes in the Mbandaka pan-genome, as previously reported in a comprehensive comparative genomic analysis of *S. enterica* (Yin et al., 2020). A total of 40 effector-coding genes were identified in the Mbandaka genomes, except for the absence of *gtgA* (Fig. 4). The majority (29 out of 40) were present as core components in the Mbandaka-A pan-genome. In contrast, the remaining *sopE*, *gogB*, *gtgE*, *spvC*, *spvD*, *sspH1*, *spvB*, *sseI*, and *sseK3* were sporadically found in ≤15% of the Mbandaka genomes, representing cloud gene families. These mobile effector-coding genes may be attribute to potential HGT events and are known to be associated with various mobile elements (Boyd et al., 2012). In addition to the previously mentioned *sopE* within the SopE Phi prophage, *gogB*, encoding anti-inflammatory effector, is located on the Gifsy-1 prophage (Coombes et al., 2005); *gtgE* and *sspH1* contribute to the full virulence of *S.* Typhimurium through the Gifsy-2 prophage (Ho et al., 2002; Miao et al., 1999); *sseI* within the Gifsy-3 prophage encodes an E3 ubiquitin ligase that is translocated into host cells to mediate long-term systemic infection (McLaughlin et al., 2009); sseK3 on the phage ST64B encodes an arginine N-acetylglucosamine transferase effector, essential for manipulating host cellular signal transduction (Esposito et al., 2018); *spvB*, *spvC*, and *spvD* encode the principal effectors responsible for the plasmid-mediated virulence of *S.* Typhimurium, including inducing apoptotic cell death in eukaryotic cells (Matsui et al., 2001). Furthermore, the Mbandaka-B genome uniquely contained *steB*, whereas *srfJ*, *sseF*, and *sseK2*, which is core components in the Mbandaka-A pan-genome, is not present.

Salmonella has developed strategies to resist the host's physical barriers and evade the immune response through its virulence factors (Wang et al., 2020). For serovar Mbandaka, SPI-2 and its secreted virulence effectors are likely to be implicated in immune escape. For instance, activation of SPI2 suppresses flagellin expression within host cells, thereby hindering the recognition of NLR family CARD domain containing 4 (NLRC4) (Reyes Ruiz et al., 2017). Effector SpiC, is critical for the ability to evade antigen presentation to T cells when *Salmonella* reside inside dendritic cells (DCs) (Tobar et al., 2006). SifA, SspH2, SlrP, PipB2, and SopD2 were important for the interference with antigen presentation in DCs, resulting in an inadequate activation of naive T cells (Halici et al., 2008). With the exception of *sspH2*, most of the genes encoding for these virulence factors are core components of the

Mbandaka pan-genome.

Taken together, the majority of identified SPIs, fimbrial operons, and effector-coding genes, representing the core genome, exhibited a similar virulence-related gene repertoire. This indicates that most Mbandaka strains have similar potential pathogenicity. Our analysis is consistent with the results of a previous comprehensive analysis, which reported that Mbandaka isolates displayed a similar ability to invade host cells and survive in an acidic environment, such as the stomach or intracellular phagosomes. However, it is worth noting that cloud virulence-related elements driven by HGT may encode mobile virulence factors. This could result in potential functional differences in pathogenicity among Mbandaka strains. Furthermore, our results highlight that prophages, acting as the key vehicles for mobile virulence-related elements, significantly contribute to the evolution of virulence, particularly in Mbandaka.

3.9. AMR-related genotypic profile of the Mbandaka pan-genome represented by cloud AMR genes

The extensive use of antimicrobial agents in livestock farming and food production has led to the emergence of globally resistant NTS strains, posing a significant threat to public health and food safety worldwide (Ahmadi et al., 2023; Yang et al., 2023). The WGS-based predicted AMR phenotypes were analyzed using ResFinder v4.3.1 (Table S3). The AMR genotypic profile of Mbandaka was predicted to be resistant to 54 antibiotics, but we did not observe any strains resistant to the remaining 37 antibiotics (Table 1). All Mbandaka genomes were uniformly resistant to tobramycin and amikacin. More than 5% of the Mbandaka genomes also showed resistance to streptomycin, tetracycline, doxycycline, sulfamethoxazole, minocycline, and spectinomycin. To further investigate the potential genetic mechanisms of resistance, we assessed the distribution of AMR-related genetic determinants in Mbandaka genomes (Fig. 5A). A total of 106 AMR genes (containing allelic variants) were identified, corresponding to resistance to 30 antibiotic classes, with primary resistances to fluoroquinolone, aminoglycoside, cephalosporin, penam, tetracycline, phenicol, and cephamycin (Fig. S6A and Table S4). Of these, 46 were present in more than 1444 Mbandaka genomes, representing core gene families (≥99%) (Fig. S6B). Most of the core AMR genes were also present in other *S. enterica* genomes, with the exception of *AAC(6′)-Iaa*. The remaining AMR genes ($n = 60$) were present in ≤15% of the Mbandaka genomes, representing cloud gene families. Among these, 15 were present in more than 10 genomes, and 45 were found in fewer than 10 genomes (Fig. 5A). The emergence of these scattered AMR genes may be attributed to HGT events. A total of 14.9% (217 out of 1452) Mbandaka genomes were found to harbor cloud AMR genes, averaging 1.7 ± 0.5. The number of cloud AMR genes per genome varied widely, from a minimum of zero to a maximum of 22. Eight strains were detected to contain at least 10 cloud AMR genes. Notably, two closely related strains, SM_F28R (SAL_BD0204AA_AA) and SM_B30R (SAL_BD0205AA_AA), exhibited the highest number of cloud AMR genes, totaling 22. These genes confer resistance to various antibiotics, including aminoglycosides (*AAC (3)-IId*, *APH(3′)-Ia*, *APH(6)-Id*, and *ANT(3″)-IIa*), beta-lactams (*NDM-1*, *CTX-M-55*, *TEM-1*, *LAP-2*, and *FONA-5*), diaminopyrimidines (*dfrA12* and *dfrA14*), fluoroquinolones (*qnrS1*), glycopeptide (*brp*), lincosamides (*linG*), macrolides (*mphA*, *mphE*, and *msrE*), phenicols (*floR*), rifamycins (*arr-2*), sulfonamides (*sul1* and *sul3*), and tetracyclines (*tet(A)*). The draft genomes of SM_F28R and SM_B30R are recorded in the EnteroBase *Salmonella* Database, and their complete genomes have been updated in the NCBI GenBank database. Further complete genome analysis revealed that the IncC plasmids pSM30_NDM_1 (CP138306.1) and pSM28_NDM_1 (CP138308.1) are responsible for carrying the majority of cloud AMR genes in these two strains. These plasmids include an aminoglycosides resistance gene (*AAC(3)-IId*), four beta-lactamases-coding genes (*NDM-1*, *CTX-M-55*, *TEM-1*, and *FONA-5*), a diaminopyrimidine resistance gene (*dfrA14*), a sulfonamide resistance gene (*sul1*), three

Table 1
Percentage of the Mbandaka genomes with AMR phenotype by WGS-based prediction tool ResFinder v4.3.1

Drug class	Antimicrobial	Number	Percentage
aminoglycoside	gentamicin	12	0.83%
aminoglycoside	tobramycin	1452	100.00%
aminoglycoside	streptomycin	142	9.78%
aminoglycoside	amikacin	1452	100.00%
aminoglycoside	isepamicin	0	0.00%
aminoglycoside	dibekacin	5	0.34%
aminoglycoside	kanamycin	8	0.55%
aminoglycoside	neomycin	8	0.55%
aminoglycoside	lividomycin	1	0.07%
aminoglycoside	paromomycin	1	0.07%
aminoglycoside	ribostamycin	1	0.07%
aminoglycoside	unknown aminoglycoside	1	0.07%
aminoglycoside	butiromycin	0	0.00%
aminoglycoside	butirosin	0	0.00%
aminoglycoside	hygromycin	0	0.00%
aminoglycoside	netilmicin	5	0.34%
aminoglycoside	apramycin	5	0.34%
aminoglycoside	sisomicin	5	0.34%
aminoglycoside	arbekacin	0	0.00%
aminoglycoside	kasugamycin	0	0.00%
aminoglycoside	astromicin	0	0.00%
aminoglycoside	fortimicin	0	0.00%
aminocyclitol	spectinomycin	73	5.03%
quinolone	fluoroquinolone	0	0.00%
quinolone	ciprofloxacin	29	2.00%
quinolone	unknown quinolone	0	0.00%
quinolone	nalidixic acid	0	0.00%
beta-lactam	amoxicillin	27	1.86%
beta-lactam	amoxicillin + clavulanic acid	6	0.41%
beta-lactam	ampicillin	27	1.86%
beta-lactam	ampicillin + clavulanic acid	6	0.41%
beta-lactam	cefepime	5	0.34%
beta-lactam	cefixime	3	0.21%
beta-lactam	cefotaxime	8	0.55%
beta-lactam	cefoxitin	6	0.41%
beta-lactam	ceftazidime	8	0.55%
beta-lactam	ertapenem	3	0.21%
beta-lactam	imipenem	3	0.21%
beta-lactam	meropenem	3	0.21%
beta-lactam	piperacillin	27	1.86%
beta-lactam	piperacillin + tazobactam	7	0.48%
beta-lactam	unknown beta-lactam	0	0.00%
beta-lactam	aztreonam	6	0.41%
beta-lactam	cefotaxime + clavulanic acid	0	0.00%
beta-lactam	temocillin	3	0.21%
beta-lactam	ticarcillin	27	1.86%
beta-lactam	ceftazidime + avibactam	3	0.21%
beta-lactam	penicillin	0	0.00%
beta-lactam	ceftriaxone	5	0.34%
beta-lactam	ticarcillin + clavulanic acid	3	0.21%
beta-lactam	cephalothin	19	1.31%
beta-lactam	piperacillin + clavulanic acid	0	0.00%
under_development	ceftiofur	0	0.00%
folate pathway antagonist	sulfamethoxazole	92	6.34%
folate pathway antagonist	trimethoprim	56	3.86%
fosfomycin	fosfomycin	0	0.00%
glycopeptide	vancomycin	0	0.00%
glycopeptide	teicoplanin	0	0.00%
glycopeptide	bleomycin	0	0.00%
lincosamide	lincomycin	8	0.55%
lincosamide	clindamycin	1	0.07%
streptogramin a	dalfopristin	0	0.00%
streptogramin a	pristinamycin iia	0	0.00%
streptogramin a	virginiamycin m	0	0.00%
streptogramin a	quinupristin + dalfopristin	0	0.00%
pleuromutilin	tiamulin	0	0.00%
macrolide	carbomycin	0	0.00%
macrolide	erythromycin	10	0.69%

(*continued on next page*)

Table 1 (*continued*)

Drug class	Antimicrobial	Number	Percentage
macrolide	azithromycin	7	0.48%
macrolide	oleandomycin	0	0.00%
macrolide	spiramycin	6	0.41%
macrolide	tylosin	0	0.00%
macrolide	telithromycin	6	0.41%
tetracycline	tetracycline	135	9.30%
tetracycline	doxycycline	135	9.30%
tetracycline	minocycline	79	5.44%
tetracycline	tigecycline	0	0.00%
streptogramin b	quinupristin	4	0.28%
streptogramin b	pristinamycin ia	4	0.28%
streptogramin b	virginiamycin s	4	0.28%
oxazolidinone	linezolid	0	0.00%
amphenicol	chloramphenicol	46	3.17%
amphenicol	florfenicol	13	0.90%
polymyxin	colistin	0	0.00%
steroid antibacterial	fusidic acid	0	0.00%
pseudomonic acid	mupirocin	0	0.00%
rifamycin	rifampicin	5	0.34%
nitroimidazole	metronidazole	0	0.00%
ionophores	narasin	0	0.00%
ionophores	salinomycin	0	0.00%
ionophores	maduramicin	0	0.00%

macrolide resistance genes (*mphA*, *mphE*, and *msrE*), and a bleomycin resistance gene (*brp*). Both SM_F28R and SM_B30R were isolated from faecal and blood samples of humans in China in December 2021. Additionally, previous studies have described multidrug resistant profiles in Mbandaka isolates from poultry farms and hospitals in China. These isolates are known to be responsible for spreading β-lactamase genes-harboring plasmids (Ma et al., 2023) (Li et al., 2020). The emergence of such multidrug resistant Mbandaka strains poses a significant threat to public health, underscoring the urgent need for meticulous monitoring and surveillance of this serovar.

3.10. Cloud AMR genes driven by plasmid-mediated HGT

Investigation of the dynamics of AMR genes is critical for the identification and verification of emerging multidrug resistance. The plasmid-mediated transmission of AMR genes has been reported to be involved in the development of antibiotic resistance in *Salmonella* isolates. Meanwhile, many serovars of *S. enterica* do not possess any plasmids (Rychlik et al., 2006). In this study, plasmid nucleotide sequences were detected in nearly half (44.4%, 644 out of 1452) of the Mbandaka genomes (Table S5) and were found dispersed in the core genome tree, not clustering into phylogenetic groups (Fig. 5A). This suggests that this serovar may be prone to acquiring plasmids. A total of 42 different plasmids were predicted using PlasmidFinder (Table S6). Col (pHAD28) was the most prevalent plasmid, detected in 112 (7.7%) genomes. Other predominant plasmids included IncHI2A (5.4%), IncHI2 (5.4%), Col440I (4.8%), and IncI1-I(Alpha) (4.3%) (Fig. 5B). Of the 217 Mbandaka genomes carrying cloud AMR genes, the vast majority (87.6%, 190 out of 217) contained plasmid sequences (Fig. 5C). In addition, a total of 46 AMR genes were found to be located in plasmid sequences. Among these, 42 were sporadically scattered throughout the Mbandaka genomes and are considered as cloud AMR genes, while only four were classified as core AMR genes. These results indicate that the transmission of AMR genes driven by plasmid-mediated HGT has promoted the development of antibiotic resistance in Mbandaka.

Although the most common replicon in the Mbandaka genomes was Col (pHAD28), several plasmid incompatibility groups, including IncHI2A, IncHI2, IncI1-I(Alpha), IncFIB(K), and IncQ1), were associated with cloud AMR genes (Fig. 5D and Table S6). Among these, IncHI2A and IncHI2 plasmids were particularly prevalent. Previous studies have demonstrated the presence of IncHI2 plasmids in multidrug resistant *Salmonella* isolates and their association with resistance to beta-lactams, quinolones, and colistin (W. Chen et al., 2016; Lima et al., 2019). However, IncHI2A and IncHI2 in the Mbandaka genomes commonly harbored *ANT(3″)-IIa*, *sul1*, and *cmlA1*, which correspond to resistance to aminoglycosides, sulfonamides, and chloramphenicols (Fig. 5D). For other incompatibility groups, IncI1-I(Alpha) plasmids were primarily linked to *ompK37* (beta-lactams) and *tet(B)* (tetracyclines); IncFIB(K) plasmids were mainly associated with *aadA2* (aminoglycosides), *sul1* (sulfonamides), *dfrA12* (diaminopyrimidines), and *tet(A)* (tetracyclines); IncQ1 plasmids carried AMR genes encoding resistance to aminoglycosides (*APH(3″)-Ib* and *APH(6)-Id*), sulfonamides (*sul2*), and diaminopyrimidines (*dfrA14*). Furthermore, the association between plasmids and AMR genes in Mbandaka genomes may be limited by the inclusion of a large number of draft genomes in our analyses. The plasmid recognition software could not differentiate all plasmid sequences in a mixture of contigs that derive from plasmids or chromosomes. As a result, our draft genomes may lack some critical information, especially concerning AMR genes located on plasmids. For example, the complete IncC plasmid in strains SM_F28R and SM_B30R were found to carry cloud AMR gene such as *NDM-1* and *CTX-M-55*, which were not identified in the draft genomes. Therefore, plasmids, as primary vehicles, likely play a more significant role in the transmission of AMR genes among Mbandaka strains than previously understood.

4. Conclusion

In conclusion, this study characterized the complete genome of a human isolate KNP01 and performed pan-genome and comparative genomic analyses of Mbandaka. It is clear, upon examining the data on a large genomic scale, that strains sharing serovar Mbandaka can be divided into two groups, each derived from independent ancestors. Our findings reveal the recombination events at the H2-antigen loci as the potential genomic basis for the confusion around this serovar. The polyphyletic nature of Mbandaka underscores the need for further investigation into potential differences in clinical phenotype and key genomic traits to fill the deficiencies of classical serotyping tests in clinical investigation and epidemic surveillance. The pan-genome analysis demonstrated an open gene pool with a high number of cloud gene families, reflecting the complexity of the genetic and evolutionary dynamics in Mbandaka. Notably, the extensive gene gain and loss observed at terminal nodes of the phylogenetic tree suggest frequent gene turnover in Mbandaka individuals, with implications for virulence and antibiotic resistance, potentially posing emerging threats to public health and food safety. The Mbandaka pan-genome contains a variety of virulence-related elements, including SPIs, prophage, fimbrial operons, and T3SS effectors. Cloud virulence-related elements, mainly driven by prophage-mediated HGT, may result in pathogenic differences and the emergence of novel disease phenotypes in Mbandaka individuals. Over one hundred AMR genes were identified in the pan-genome, with most scattered sporadically across Mbandaka genomes. Plasmid-mediated HGT, especially through incompatibility group plasmids (IncHI2A, IncHI2, IncI1-I(Alpha), IncFIB(K), and IncQ1), drives the transmission of cloud AMR genes among Mbandaka individuals, underscoring the role of these plasmids in the development of antibiotic resistance.

CRediT authorship contribution statement

Nai-peng Kan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Writing – original draft. **Zhiqiu Yin:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Yu-feng Qiu:** Data curation, Formal analysis, Methodology, Visualization. **Enhui Zheng:** Methodology, Resources, Validation. **Jianhui Chen:** Project administration, Resources, Supervision. **Jianzhong Huang:** Funding acquisition, Project administration, Resources, Writing – review & editing. **Yuhui Du:** Conceptualization, Investigation, Methodology, Writing – review & editing.

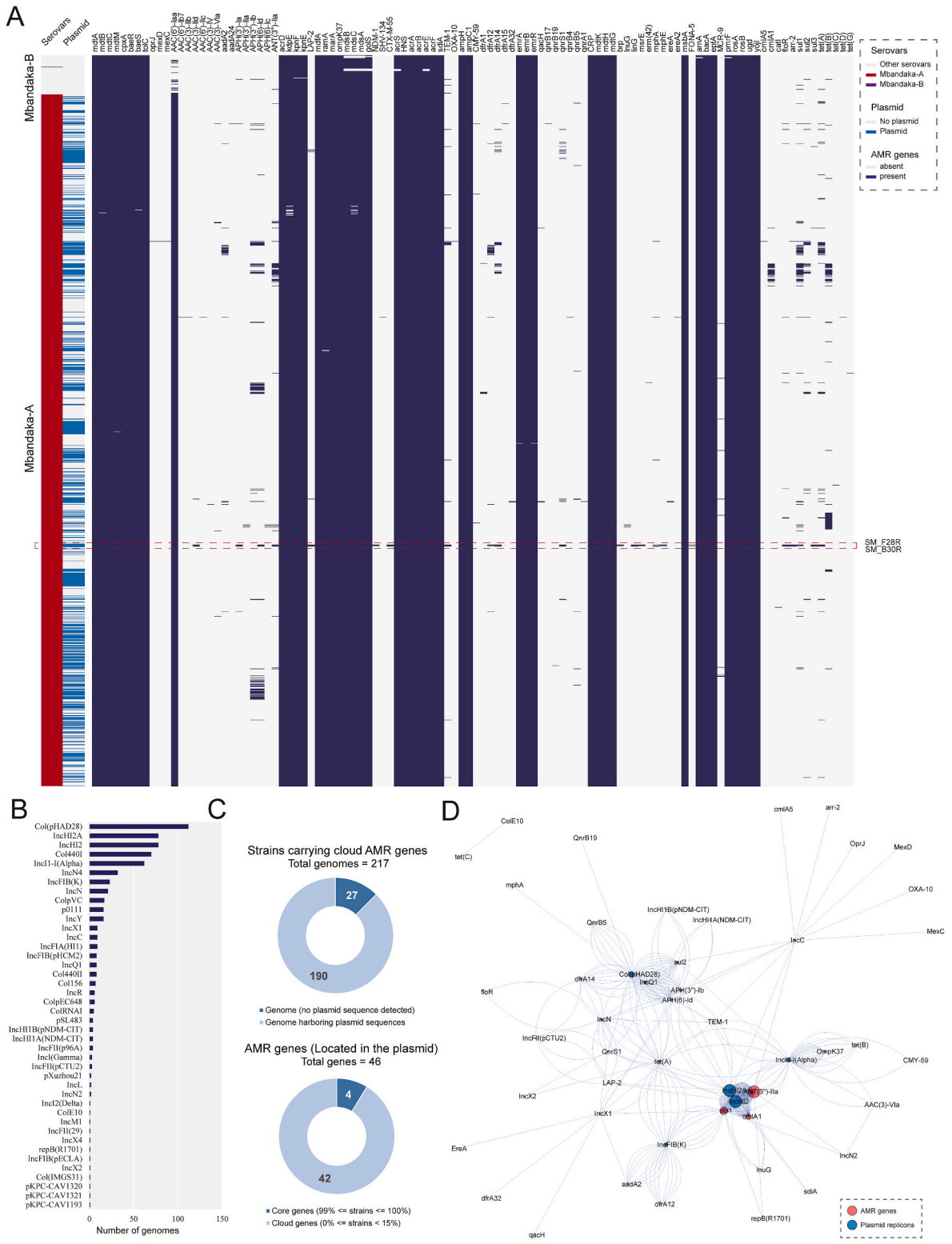

Fig. 5. Antimicrobial resistance (AMR)-related genetic profile. A. Heatmap of the distribution of AMR genes in the *S. enterica* genomes. Navy blocks indicate the presence of a gene, and grey blocks indicate absence. B. Distribution of the different types of plasmid replication in Mbandaka genomes. C. The top pie chart represents the percentage of strains with resistance genes containing plasmid sequences versus those without. The bottom pie chart represents the percentage of AMR genes that are located in the plasmid across different pan-genome components. D. Relationship network for plasmid replications and located AMR genes.

Ethical approval

No patients were recruited, and the data collected from human sample was anonymized and hence ethical approval and informed consent statements are not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by National Key Research and Development Program of China (No. 2022YFD1802104) and Fujian Provincial Health Technology Project (No. 2020QNA021). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2024.100957.

Data availability

Complete sequence of the KNP01 genome was deposited in the NCBI GenBank database (accession number CP113364.1).

References

Achtman, M., Zhou, Z., Alikhan, N.-F., Tyne, W., Parkhill, J., Cormican, M., Chiou, C.-S., Torpdahl, M., Litrup, E., Prendergast, D.M., Moore, J.E., Strain, S., Kornschober, C., Meinersmann, R., Uesbeck, A., Weill, F.-X., Coffey, A., Andrews-Polymenis, H., Curtiss Rd, R., Fanning, S., 2020. Genomic diversity of Salmonella enterica -The UoWUCC 10K genomes project. Wellcome Open Res. 5, 223. https://doi.org/10.12688/wellcomeopenres.16291.2.

Ahmadi, Z., Pakbin, B., Kazemi, M., Rahimi, Z., Mahmoudi, R., 2023. Genotyping and antibiotic susceptibility of Campylobacter species isolated from raw milk samples in Qazvin, Iran. BMC Res. Notes 16 (1), 314. https://doi.org/10.1186/s13104-023-06576-9.

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., et al., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48 (D1), D517–D525. https://doi.org/10.1093/nar/gkz935.

Alikhan, N.F., Zhou, Z., Sergeant, M.J., Achtman, M., 2018. A genomic overview of the population structure of Salmonella. PLoS Genet. 14 (4), e1007261. https://doi.org/10.1371/journal.pgen.1007261.

Andreopoulos, W.B., Geller, A.M., Lucke, M., Balewski, J., Clum, A., Ivanova, N.N., Levy, A., 2022. Deeplasmid: deep learning accurately separates plasmids from bacterial chromosomes. Nucleic Acids Res. 50 (3), e17. https://doi.org/10.1093/nar/gkab1115.

Antony, L., Fenske, G., Kaushik, R.S., Nagaraja, T.G., Thomas, M., Scaria, J., 2020. Population structure of Salmonella enterica serotype Mbandaka reveals similar virulence potential irrespective of source and phylogenomic stratification. F1000Res. 9, 1142. https://doi.org/10.12688/f1000research.25540.1.

Bakkeren, E., Huisman, J.S., Fattinger, S.A., Hausmann, A., Furter, M., Egli, A., Slack, E., Sellin, M.E., Bonhoeffer, S., Regoes, R.R., Diard, M., Hardt, W.-D., 2019. Salmonella persisters promote the spread of antibiotic resistance plasmids in the gut. Nat. 573 (7773), 276–280. https://doi.org/10.1038/s41586-019-1521-8.

Balasubramanian, D., López-Pérez, M., Grant, T.-A., Ogbunugafor, C.B., Almagro-Moreno, S., 2022. Molecular mechanisms and drivers of pathogen emergence. Trends Microbiol. 30 (9), 898–911. https://doi.org/10.1016/j.tim.2022.02.003.

Benevides, V.P., Saraiva, M.M.S., Nascimento, C.F., Delgado-Suárez, E.J., Oliveira, C.J.B., Silva, S.R., Miranda, V.F.O., Christensen, H., Olsen, J.E., Berchieri Junior, A., 2024. Genomic features and phylogenetic analysis of antimicrobial-resistant Salmonella Mbandaka ST413 strains. Microorgan. 12 (2). https://doi.org/10.3390/microorganisms12020312.

Bobay, L.-M., Ellis, B.S.-H., Ochman, H., 2018. ConSpeciFix: classifying prokaryotic species based on gene flow. Bioinf. 34 (21), 3738–3740. https://doi.org/10.1093/bioinformatics/bty400.

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Philippon, A., Allesoe, R. L., Rebelo, A.R., Florensa, A.F., Cattoir, V., Fagelhauer, L., Coppens, J., Xavier, B.B., Malhotra-kumar, S., Westh, H., Losch, S., Olkkola, S., Wieczorek, K., Amaro, A., et al., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 75 (12), 3491–3500. https://doi.org/10.1093/jac/dkaa345.

Boyd, E.F., Carpenter, M.R., Chowdhury, N., 2012. Mobile effector proteins on phage genomes. Bacteriophage 2 (3), 139–148. https://doi.org/10.4161/bact.21658.

Brockhurst, M.A., Harrison, E., Hall, J.P.J., Richards, T., McNally, A., MacLean, C., 2019. The ecology and evolution of pangenomes. Curr. Biol. 29 (20), R1094–R1103. https://doi.org/10.1016/j.cub.2019.08.012.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf. 34 (17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chen, W., Fang, T., Zhou, X., Zhang, D., Shi, X., Shi, C., 2016. IncHI2 plasmids are predominant in antibiotic-resistant Salmonella isolates. Front. Microbiol. 7, 1566. https://doi.org/10.3389/fmicb.2016.01566.

Cheng, R.A., Eade, C.R., Wiedmann, M., 2019. Embracing diversity: differences in virulence mechanisms, disease severity, and host adaptations contribute to the success of nontyphoidal Salmonella as a foodborne pathogen. Front. Microbiol. 10, 1368. https://doi.org/10.3389/fmicb.2019.01368.

Cherchame, E., Ilango, G., Noël, V., Cadel-Six, S., 2022. Polyphyly in widespread Salmonella enterica serovars and using genomic proximity to choose the best reference genome for bioinformatics analyses. Front. Public Health 10, 963188. https://doi.org/10.3389/fpubh.2022.963188.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J.R., Cantu, D., Rank, D.R., Schatz, M.C., 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13 (12), 1050–1054. https://doi.org/10.1038/nmeth.4035.

Cobo-Simón, M., Hart, R., Ochman, H., 2023. Gene flow and species boundaries of the genus Salmonella. mSyst. 8 (4), e0029223. https://doi.org/10.1128/msystems.00292-23.

Coombes, B.K., Wickham, M.E., Brown, N.F., Lemire, S., Bossi, L., Hsiao, W.W.L., Brinkman, F.S.L., Finlay, B.B., 2005. Genetic and molecular analysis of GogB, a phage-encoded type III-secreted substrate in Salmonella enterica serovar typhimurium with autonomous expression from its associated phage. J. Mol. Biol. 348 (4), 817–830. https://doi.org/10.1016/j.jmb.2005.03.024.

De Sousa Violante, M., Michel, V., Romero, K., Bonifait, L., Baugé, L., Perrin-Guyomard, A., Feurer, C., Radomski, N., Mallet, L., Mistou, M.-Y., Cadel-Six, S., 2023. Tell me if you prefer bovine or poultry sectors and I'll tell you who you are: characterization of Salmonella enterica subsp. enterica serovar Mbandaka in France. Front. Microbiol. 14, 1130891. https://doi.org/10.3389/fmicb.2023.1130891.

Didelot, X., Wilson, D.J., 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. 11 (2), 1–18. https://doi.org/10.1371/journal.pcbi.1004041.

Esposito, D., Günster, R.A., Martino, L., El Omari, K., Wagner, A., Thurston, T.L.M., Rittinger, K., 2018. Structural basis for the glycosyltransferase activity of the Salmonella effector SseK3. J. Biol. Chem. 293 (14), 5064–5078. https://doi.org/10.1074/jbc.RA118.001796.

Eswarappa, S.M., Janice, J., Nagarajan, A.G., Balasundaram, S.V., Karnam, G., Dixit, N. M., Chakravortty, D., 2008. Differentially evolved genes of Salmonella pathogenicity islands: insights into the mechanism of host specificity in Salmonella. PLoS One 3 (12), e3829. https://doi.org/10.1371/journal.pone.0003829.

Gladstone, R.A., McNally, A., Pöntinen, A.K., Tonkin-Hill, G., Lees, J.A., Skytén, K., Cléon, F., Christensen, M.O.K., Haldorsen, B.C., Bye, K.K., Gammelsrud, K.W., Hjetland, R., Kümmel, A., Larsen, H.E., Lindemann, P.C., Löhr, I.H., Marvik, Å., Nilsen, E., Noer, M.T., et al., 2021. Emergence and dissemination of antimicrobial resistance in Escherichia coli causing bloodstream infections in Norway in 2002-17: a nationwide, longitudinal, microbial population genomic study. Lancet. Microbe 2 (7), e331–e341. https://doi.org/10.1016/S2666-5247(21)00031-8.

Halici, S., Zenk, S.F., Jantsch, J., Hensel, M., 2008. Functional analysis of the Salmonella pathogenicity island 2-mediated inhibition of antigen presentation in dendritic cells. Infect. Immun. 76 (11), 4924–4933. https://doi.org/10.1128/IAI.00531-08.

Hayward, M.R., Jansen, V.A.A., Woodward, M.J., 2013. Comparative genomics of Salmonella enterica serovars Derby and Mbandaka, two prevalent serovars associated with different livestock species in the UK. BMC Genom. 14, 365. https://doi.org/10.1186/1471-2164-14-365.

Ho, T.D., Figueroa-Bossi, N., Wang, M., Uzzau, S., Bossi, L., Slauch, J.M., 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of Salmonella enterica serovar Typhimurium. J. Bacteriol. 184 (19), 5234–5239. https://doi.org/10.1128/JB.184.19.5234-5239.2002.

Hosseini, H., Mahmoudi, R., Pakbin, B., Manafi, L., Hosseini, S., Pilevar, Z., Brück, W.M., 2024. Effects of intrinsic and extrinsic growth factors on virulence gene expression of foodborne pathogens in vitro and in food model systems; a review. Food Sci. Nutr. 12 (9), 6093–6107. https://doi.org/10.1002/fsn3.4281.

Hoszowski, A., Zając, M., Lalak, A., Przemyk, P., Wasyl, D., 2016. Fifteen years of successful spread of Salmonella enterica serovar Mbandaka clone ST413 in Poland and its public health consequences. Ann. Agric. Environ. Med. : AAEM 23 (2), 237–241. https://doi.org/10.5604/12321966.1203883.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34 (8), 2115–2122. https://doi.org/10.1093/molbev/msx148.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9 (1), 5114. https://doi.org/10.1038/s41467-018-07641-9.

Kamali, A., Hosseini, H., Mahmoudi, R., Pakbin, B., Gheibi, N., Mortazavian, A.M., Shojaei, S., 2024. The sensory evaluation and antimicrobial efficacy of Lactobacillus acidophilus supernatant on Salmonella enteritidis in milk. Food Sci. Nutr. 12 (3), 1902–1910. https://doi.org/10.1002/fsn3.3883.

Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. 20 (4), 1160–1166. https://doi.org/10.1093/bib/bbx108.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7 : improvements in performance and usability article fast track. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Keaton, A.A., Schwensohn, C.A., Brandenburg, J.M., Pereira, E., Adcock, B., Tecle, S., Hinnenkamp, R., Havens, J., Bailey, K., Applegate, B., Whitney, P., Gibson, D., Manion, K., Griffin, M., Ritter, J., Biskupiak, C., Ajileye, K., Golwalkar, M., Gosciminski, M., et al., 2022. Multistate outbreak of Salmonella Mbandaka infections linked to sweetened puffed wheat cereal - United States, 2018. Epidemiol. Infect. 150, e135. https://doi.org/10.1017/S095026882200108X.

Li, W., Li, H., Zheng, S., Wang, Z., Sheng, H., Shi, C., Shi, X., Niu, Q., Yang, B., 2020. Prevalence, serotype, antibiotic susceptibility, and genotype of Salmonella in eggs from poultry farms and marketplaces in yangling, Shaanxi Province, China. Front. Microbiol. 11, 1482. https://doi.org/10.3389/fmicb.2020.01482.

Lima, T., Domingues, S., Da Silva, G.J., 2019. Plasmid-mediated colistin resistance in Salmonella enterica: a review. Microorg. 7 (2). https://doi.org/10.3390/microorganisms7020055.

Ma, W., Cui, X., Dong, X., Li, X., Liu, K., Wang, Y., Shi, X., Chen, L., Hao, M., 2023. Characterization of nontyphoidal Salmonella strains from a tertiary hospital in China: serotype diversity, multidrug resistance, and genetic insights. Front. Cell. Infect. Microbiol. 13, 1327092. https://doi.org/10.3389/fcimb.2023.1327092.

Matsui, H., Bacot, C.M., Garlington, W.A., Doyle, T.J., Roberts, S., Gulig, P.A., 2001. Virulence plasmid-borne spvB and spvC genes can replace the 90-kilobase plasmid in conferring virulence to Salmonella enterica serovar Typhimurium in subcutaneously inoculated mice. J. Bacteriol. 183 (15), 4652–4658. https://doi.org/10.1128/JB.183.15.4652-4658.2001.

McLaughlin, L.M., Govoni, G.R., Gerke, C., Gopinath, S., Peng, K., Laidlaw, G., Chien, Y.-H., Jeong, H.-W., Li, Z., Brown, M.D., Sacks, D.B., Monack, D., 2009. The Salmonella SPI2 effector SseI mediates long-term systemic infection by modulating host cell migration. PLoS Pathog. 5 (11), e1000671. https://doi.org/10.1371/journal.ppat.1000671.

Meier-Kolthoff, J.P., Göker, M., 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 10 (1), 2182. https://doi.org/10.1038/s41467-019-10210-3.

Miao, E.A., Scherer, C.A., Tsolis, R.M., Kingsley, R.A., Adams, L.G., Bäumler, A.J., Miller, S.I., 1999. Salmonella typhimurium leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. Mol. Microbiol. 34 (4), 850–864. https://doi.org/10.1046/j.1365-2958.1999.01651.x.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mirold, S., Rabsch, W., Tschäpe, H., Hardt, W.D., 2001. Transfer of the Salmonella type III effector sopE between unrelated phage families. J. Mol. Biol. 312 (1), 7–16. https://doi.org/10.1006/jmbi.2001.4950.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 11 (12), 2864–2868. https://doi.org/10.1038/ismej.2017.126.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25 (7), 1043–1055. https://doi.org/10.1101/gr.186072.114.

Pinna, E. De, Weill, F., Peters, T., 2016. What's in a name? Species-wide whole-genome sequencing resolves invasive and noninvasive lineages of Salmonella enterica serotype Paratyphi B. mBio 7 (4). https://doi.org/10.1128/mBio.00527-16.

Porwollik, S., McClelland, M., 2003. Lateral gene transfer in Salmonella. Microb. Infect. 5 (11), 977–989. https://doi.org/10.1016/s1286-4579(03)00186-2.

Reyes Ruiz, V.M., Ramirez, J., Naseer, N., Palacio, N.M., Siddarthan, I.J., Yan, B.M., Boyer, M.A., Pensinger, D.A., Sauer, J.-D., Shin, S., 2017. Broad detection of bacterial type III secretion system and flagellin proteins by the human NAIP/NLRC4 inflammasome. Proc. Natl. Acad. Sci. U. S. A 114 (50), 13242–13247. https://doi.org/10.1073/pnas.1710433114.

Rychlik, I., Gregorova, D., Hradecka, H., 2006. Distribution and function of plasmids in Salmonella enterica. Vet. Microbiol. 112 (1), 1–10. https://doi.org/10.1016/j.vetmic.2005.10.030.

Sabbagh, S.C., Forest, C.G., Lepage, C., Leclerc, J.M., Daigle, F., 2010. So similar, yet so different: uncovering distinctive features in the genomes of Salmonella enterica serovars Typhimurium and Typhi. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 305 (1), 1–13. https://doi.org/10.1111/j.1574-6968.2010.01904.x.

Sahl, J.W., Caporaso, J.G., Rasko, D.A., 2014. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2 (e332), 1–12. https://doi.org/10.7717/peerj.332.

Sangal, V., Harbottle, H., Mazzoni, C.J., Helmuth, R., Guerra, B., Didelot, X., Paglietti, B., Rabsch, W., Brisse, S., Weill, F.X., Roumagnac, P., Achtman, M., 2010. Evolution and population structure of Salmonella enterica serovar Newport. J. Bacteriol. 192 (24), 6465–6476. https://doi.org/10.1128/JB.00969-10.

Schaechter, M., Group, T.V.F.H., 2004. Escherichia coli and Salmonella 2000: the view from here. EcoSal Plus 1 (1). https://doi.org/10.1128/ecosalplus.1.4.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinf. 30 (14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38 (7), 3022–3027. https://doi.org/10.1093/molbev/msab120.

Tettelin, H., Riley, D., Cattuto, C., M. D, 2008. Comparative genomics: the bacterial pan-genome. Curr. Opin. Microbiol. 11 (5), 472–477. https://doi.org/10.1016/j.mib.2008.09.006.

Timme, R.E., Pettengill, J.B., Allard, M.W., Strain, E., Barrangou, R., Wehnes, C., Van Kessel, J.S., Karns, J.S., Musser, S.M., Brown, E.W., 2013. Phylogenetic diversity of the enteric pathogen Salmonella enterica subsp. enterica inferred from genome-wide reference-free SNP characters. Genom. Biol. Evolut. 5 (11), 2109–2123. https://doi.org/10.1093/gbe/evt159.

Tobar, J.A., Carreño, L.J., Bueno, S.M., González, P.A., Mora, J.E., Quezada, S.A., Kalergis, A.M., 2006. Virulent Salmonella enterica serovar typhimurium evades adaptive immunity by preventing dendritic cells from activating T cells. Infect. Immun. 74 (11), 6438–6448. https://doi.org/10.1128/IAI.00063-06.

Tonkin-Hill, G., Gladstone, R.A., Pöntinen, A.K., Arredondo-Alonso, S., Bentley, S.D., Corander, J., 2023. Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe. Genome Res. 33 (1), 129–140. https://doi.org/10.1101/gr.277340.122.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R.A., Frost, S.D.W., Corander, J., Bentley, S.D., Parkhill, J., 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 21 (1), 180. https://doi.org/10.1186/s13059-020-02090-4.

Valizadeh, S., Moosavy, M.-H., Ebrahimi, A., Basti, A.A., Mahmoudi, R., Khatibi, S.M.H., 2022. The effect of aqueous extract of saffron (crocus sativus L. Stigma) on the behavior of Salmonella typhimurium in A food model during storage at different temperatures. J. Nutrition Food Secur. 7 (1), 37–45. https://api.semanticscholar.org/CorpusID:246544387.

Vernikos, G., Medini, D., Riley, D.R., Tettelin, H., 2015. Ten years of pan-genome analyses. Curr. Opin. Microbiol. 23, 148–154. https://doi.org/10.1016/j.mib.2014.11.016.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9 (11), e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang, M., Goh, Y.-X., Tai, C., Wang, H., Deng, Z., Ou, H.-Y., 2022. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. Nucleic Acids Res. 50 (W1), W768–W773. https://doi.org/10.1093/nar/gkac321.

Wang, M., Qazi, I.H., Wang, L., Zhou, G., Han, H., 2020. Salmonella virulence and immune escape. Microorg. 8 (3), 407. https://doi.org/10.3390/microorganisms8030407.

Wayne, P. (n.d.). CLSI Document M100-S31. Clinical and Laboratory Standards Institute; Malvern, Pennsylvania, USA—The Clinical and Laboratory Standards Institute (CLSI) Has Published M100—Performance Standards for Antimicrobial Susceptibility Testing, 31st Edition.

Wishart, D.S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J.R., Stothard, P., Gautam, V., 2023. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 51 (W1), W443–W450. https://doi.org/10.1093/nar/gkad382.

Worley, J., Meng, J., Allard, M.W., Brown, E.W., Timme, R.E., 2018. Salmonella enterica phylogeny based on whole-genome sequencing reveals two new clades and novel patterns of horizontally acquired genetic elements. mBio 9 (6). https://doi.org/10.1128/mBio.02303-18.

Yan, S., Zhang, W., Li, C., Liu, X., Zhu, L., Chen, L., Yang, B., 2021. Serotyping, MLST, and core genome MLST analysis of Salmonella enterica from different sources in China during 2004-2019. Front. Microbiol. 12, 688614. https://doi.org/10.3389/fmicb.2021.688614.

Yang, C., Xiang, Y., Qiu, S., 2023. Resistance in Enteric Shigella and nontyphoidal Salmonella: emerging concepts. Curr. Opin. Infect. Dis. 36 (5), 360–365. https://doi.org/10.1097/QCO.0000000000000960.

Yates, C.R., Nguyen, A., Liao, J., Cheng, R.A., 2024. What's on a prophage: analysis of Salmonella spp. prophages identifies a diverse range of cargo with multiple virulence- and metabolism-associated functions. mSphere, e0003124. https://doi.org/10.1128/msphere.00031-24.

Yin, Z., Liang, J., Zhang, M., Chen, B., Yu, Z., Tian, X., Deng, X., Peng, L., 2024. Pan-genome insights into adaptive evolution of bacterial symbionts in mixed host-microbe symbioses represented by human gut microbiota Bacteroides cellulosilyticus. Sci. Total Environ. 927, 172251. https://doi.org/10.1016/j.scitotenv.2024.172251.

Yin, Z., Liu, J., Du, B., Ruan, H.-H., Huo, Y.-X., Du, Y., Qiao, J., 2020. Whole-genome-based survey for polyphyletic serovars of Salmonella enterica subsp. enterica provides new insights into public health surveillance. Int. J. Mol. Sci. 21 (15). https://doi.org/10.3390/ijms21155226.

Yin, Z., Liu, X., Qian, C., Sun, L., Pang, S., Liu, J., Li, W., Huang, W., Cui, S., Zhang, C., Song, W., Wang, D., Xie, Z., 2022. Pan-genome analysis of delftia tsuruhatensis reveals important traits concerning the genetic diversity, pathogenicity, and biotechnological properties of the species. Microbiol. Spectr. 11 (1), e0207221. https://doi.org/10.1128/spectrum.02072-21.

Yin, Z., Yuan, C., Du, Y., Yang, P., Qian, C., Wei, Y., Zhang, S., Huang, D., Liu, B., 2019. Comparative genomic analysis of the Hafnia genus reveals an explicit evolutionary relationship between the species alvei and paralvei and provides insights into pathogenicity. BMC Genom. 20 (1), 768. https://doi.org/10.1186/s12864-019-6123-1.

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 (1), e0147101. https://doi.org/10.1371/journal.pone.0147101.

Yuan, C., Wei, Y., Zhang, S., Cheng, J., Cheng, X., Qian, C., Wang, Y., Zhang, Y., Yin, Z., Chen, H., 2020. Comparative genomic analysis reveals genetic mechanisms of the variety of pathogenicity, antibiotic resistance, and environmental adaptation of providencia genus. Front. Microbiol. 11, 572642. https://doi.org/10.3389/fmicb.2020.572642.

Zhang, S., den Bakker, H.C., Li, S., Chen, J., Dinsmore, B.A., Lane, C., Lauer, A.C., Fields, P.I., Deng, X., 2019. SeqSero2: rapid and improved salmonella serotype determination using whole-genome sequencing data. Appl. Environ. Microbiol. 85 (23), 1–13. https://doi.org/10.1128/AEM.01746-19.