



## Article

# Hybrid Genomic Analysis of *Salmonella enterica* Serovar Enteritidis SE3 Isolated from Polluted Soil in Brazil

Danitza Xiomara Romero-Calle <sup>1,2,3</sup>, Francisnei Pedrosa-Silva <sup>4</sup> , Luiz Marcelo Ribeiro Tomé <sup>2</sup>, Thiago J. Sousa <sup>5</sup>, Leila Thaise Santana de Oliveira Santos <sup>3</sup>, Vasco Ariston de Carvalho Azevedo <sup>5</sup> , Bertram Brenig <sup>6</sup> , Raquel Guimarães Benevides <sup>1,3</sup>, Thiago M. Venancio <sup>4</sup> , Craig Billington <sup>7,\*</sup> and Aristóteles Góes-Neto <sup>1,2,3,5,\*</sup>

<sup>1</sup> Postgraduate Program in Biotechnology, State University of Feira de Santana (UEFS), Av. Transnordestina S/N, Feira de Santana 44036-900, BA, Brazil

<sup>2</sup> Molecular and Computational Biology of Fungi Laboratory, Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, MG, Brazil

<sup>3</sup> Department of Biological Sciences, Feira de Santana State University (UEFS), Feira de Santana 44036-900, BA, Brazil

<sup>4</sup> Laboratory of Chemistry, Function of Proteins and Peptides, Center for Biosciences and Biotechnology, Darcy Ribeiro North Fluminense State University (UENF), Campos dos Goytacazes 28013-602, RJ, Brazil

<sup>5</sup> Laboratory of Cellular and Molecular Genetics, Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte 31270-901, MG, Brazil

<sup>6</sup> Institute of Veterinary Medicine, Burckhardtweg, University of Göttingen, 37073 Göttingen, Germany

<sup>7</sup> Health & Environment Group, Institute of Environmental Sciences and Research, P.O. Box 29-181, Christchurch 8540, New Zealand

\* Correspondence: craig.billington@esr.cri.nz (C.B.); arigoesneto@icb.ufmg.br (A.G.-N.); Tel.: +64-3351-6019 (C.B.); +55-31-9413-0996 (A.G.-N.)



**Citation:** Romero-Calle, D.X.; Pedrosa-Silva, F.; Tomé, L.M.R.; Sousa, T.J.; de Oliveira Santos, L.T.S.; de Carvalho Azevedo, V.A.; Brenig, B.; Benevides, R.G.; Venancio, T.M.; Billington, C.; et al. Hybrid Genomic Analysis of *Salmonella enterica* Serovar Enteritidis SE3 Isolated from Polluted Soil in Brazil. *Microorganisms* **2023**, *11*, 111. <https://doi.org/10.3390/microorganisms11010111>

Academic Editor: Andrew Shelenkov

Received: 23 November 2022

Revised: 19 December 2022

Accepted: 27 December 2022

Published: 31 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** In Brazil, *Salmonella enterica* serovar Enteritidis is a significant health threat. *Salmonella enterica* serovar Enteritidis SE3 was isolated from soil at the Subaé River in Santo Amaro, Brazil, a region contaminated with heavy metals and organic waste. Illumina HiSeq and Oxford Nanopore Technologies MinION sequencing were used for de novo hybrid assembly of the *Salmonella* SE3 genome. This approach yielded 10 contigs with 99.98% identity with *S. enterica* serovar Enteritidis OLF-SE2-98984-6. Twelve *Salmonella* pathogenic islands, multiple virulence genes, multiple antimicrobial gene resistance genes, seven phage defense systems, seven prophages and a heavy metal resistance gene were encoded in the genome. Pangenome analysis of the *S. enterica* clade, including *Salmonella* SE3, revealed an open pangenome, with a core genome of 2137 genes. Our study showed the effectiveness of a hybrid sequence assembly approach for environmental *Salmonella* genome analysis using HiSeq and MinION data. This approach enabled the identification of key resistance and virulence genes, and these data are important to inform the control of *Salmonella* and heavy metal pollution in the Santo Amaro region of Brazil.

**Keywords:** whole genome sequencing; *Salmonella*; hybrid sequence assembly; heavy metal; antimicrobial resistance

## 1. Introduction

Salmonellosis, one of the primary causes of foodborne infections resulting from gram-negative enteropathogenic bacteria *Salmonella* spp., is a global threat to human health [1]. Typhoidal *Salmonella* causes enteric fever in humans, whereas non-typhoidal *Salmonella* (NTS) results in acute/chronic gastroenteritis. Annually, it is estimated that NTS is responsible for ~93.8 million infections and ~155,000 deaths [2].

NTS infections cause diarrhoea and a non-specific febrile illness that is clinically indistinguishable from other febrile illnesses [3]. *Salmonella enterica* subspecies *enterica* has more than 2600 serovars according to unique somatic (O) and flagellar (H) antigenic

formulae [4,5]. *S. enterica* sv. Typhimurium and *S. enterica* sv. Enteritidis are the main pathogens responsible for causing gastroenteritis in humans [6,7].

To prevent the occurrence of the main *Salmonella* serovars worldwide, several prevention and control measures are adopted in farms and food processing industries. In Brazil, *Salmonella* infection of flocks and transmission to poultry-derived food is a major transmission route for the pathogen. *Salmonella* is routinely managed on Brazilian farms by poultry vaccination and laboratory testing (Available online: [https://www.gov.br/agricultura/pt-br/assuntos/saude-animal-e-vegetal/saude-animal/programas-desau-de-animal/pnsa/2003\\_78.INconsolidada.pdf](https://www.gov.br/agricultura/pt-br/assuntos/saude-animal-e-vegetal/saude-animal/programas-desau-de-animal/pnsa/2003_78.INconsolidada.pdf) (accessed on 18 December 2022)). However, despite these measures several poultry diseases and foodborne *Salmonella* outbreaks have been reported in Brazil in recent decades [8].

Whole-genome sequencing (WGS) is useful in foodborne outbreak investigations and pathogen surveillance [9]. Illumina short-read sequencing technology has proven to be robust for characterizing pathogens of clinical care [10], but it is unable to resolve repetitive and GC-rich regions, thus producing unresolvable regions in the underlying genome assembly [11]. These unresolved regions impede completion of a whole-genome structure, which is crucial to determine if some genes are co-regulated or co-transmissible, and if they are located on the chromosome or plasmids [12]. Furthermore, the bias to identify key virulence genes during an outbreak investigation can also have negative impacts on public health assessment.

Nanopore sequencing technology can generate long reads to facilitate the completion of bacterial genome assemblies but can lack sequencing depth in some repetitive regions [13]. However, nanopore's long reads can span wide repetitive regions and help solve GC-rich regions, making it useful for resolving full-length genome sequences [14]. Nanopore sequencing technology exhibits lower read accuracy than Illumina sequencing which can produce systematic errors, as a result, it has only usually been applied as a complement to short-read sequencing for bacterial genome assembly [15]. Since the release of the MinION platform by Oxford Nanopore Technologies, nanopore chemistry, base-calling, and bioinformatic tools have been steadily improving and are now more able to produce accurate bacterial genome sequences independent of other sequencing technologies [16].

The combination of both short reads for base-calling accuracy and long reads for structural integrity has recently been developed as a hybrid assembly approach to close whole-genome assemblies, such as those found in the Unicycler and SPAdes pipelines [17,18]. Unicycler was specifically developed for hybrid assembly of bacterial genomes [18]. Unicycler generates a short-read assembly graph and then uses long-reads to build bridges to resolve all repeats in the genome, performs multiple rounds of short-read polishing and finally, it produces a complete genome assembly [14].

In this study, a hybrid genome assembly approach using MinION and HiSeq sequencing data was used to improve the assembly parameters and gene completeness, identification of virulence and antimicrobial resistance genes (ARG), genome phylogeny and pangenome in *Salmonella enterica* var. Enteritidis SE3 isolated from soil at the Subaé river in Santo Amaro, Brazil, a river polluted with organic waste and heavy metals.

## 2. Materials and Methods

Environmental soil samples were obtained from the Subaé river basin in Santo Amaro, Salvador de Bahia, Brazil. Approximately 100 g of soil sample was collected from river soil (12°31'46.77" S 38°44'1.24" W). The sample was transported in a refrigerated box (4–8 °C) to the laboratory where the analyses were undertaken immediately.

### 2.1. *Salmonella* Isolation

*Salmonella* was isolated according to the US Food and Drug Administration Bacteriological Analytical Manual (<https://www.fda.gov/media/79991/download> (accessed on 18 December 2022)). Briefly, 10 g or 10 mL of samples of each sample were pre-enriched in 100 mL lactose broth (supplier), at 37 °C for 24 h, 0.1 mL of pre-enriched culture was trans-

ferred to 10 mL enriched in Tetrathionate (TT) broth (HIMEDIA, Kennett Square, PA, USA) and incubated at 41 °C for 24 h. Broth cultures from the selective enrichment broth were plated on Xylose-Lysine-Desoxycholate (XLD) agar (HIMEDIA, Kennett Square, PA, USA), Bismuth sulfite agar (Acumedia Manufacturers Inc., San Bernardino, CA, USA) and Salmonella Shigella (SS) agar (HIMEDIA, PA, Kennett Square, USA). Colonies characteristic of *Salmonella* having a slightly transparent zone of reddish color and a black center for XLD, gray or brown-black colonies with or without metallic sheen for Bismuth Sulfite Agar, and beige colonies with black centers for SS agar were identified and picked. Then, the isolates were tested biochemically using the Triple Sugar Iron (TSI) test. *Salmonella* strains were confirmed when they showed good to excellent growth, pink colonies with black centers were detected, and the agar was red [19].

## 2.2. DNA Isolation

For bacteria, a single colony was enriched in 5 mL Luria Bertani (LB) broth, and 15 mL of enrichment broth was transferred to a centrifuge tube and centrifuged at 4000 rpm for 10 min. DNA from *Salmonella* strains was extracted and purified using the E.Z.N.A. Bacterial DNA Mini Kit (Omega Biotek, Norcross, GA, USA) following the instructions provided by the manufacturer. For phages, a crude lysate was centrifuged the lysate as described. DNA isolation from phages was carried out using the E.Z.N.A. Viral DNA Mini Isolation Kit (Omega Biotek, Norcross, GA, USA) following the instructions provided by the manufacturer. The quality and concentration of the bacteria and phage DNA was evaluated by Qubit Fluorometric Quantification (ThermoFisher Scientific, Waltham, MA, USA) and gel electrophoresis (1% of agarose gel, 80 V for 45 min in 1x TAE Buffer).

## 2.3. Amplification of 16S rRNA Gene

PCR amplification was performed using a Veriti™ 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA, USA), 16S gene Amplification PCR for the amplification of the 16S rRNA gene was carried out using universal primers 27F (5'-AGAGTTTGATCATGGCTCAG-3') as forward and 1492R (5'-GGTACCTTGTTACGACTT-3') as a reverse primer [20]. Approximately 10–100 ng of template was added to a reaction mix containing 10 µL Master Mix 2x (Qiagen, Germantown, MD, USA), 1 µL primer 27 F (10 µM), 1 µL primer 1492R (10 µM), and 1 µL reverse primer (10 µM). PCR was performed with the following cycling conditions: initial denaturation at 95 °C for 10 min, 35 cycles of denaturation at 95 °C for 1 min, annealing from 50 °C to 60 °C for 1 min, and extension at 72 °C for 1 min. A final extension was performed at 72 °C for 7 min. PCR products were visualized using GelRed (Biotium, San Francisco, CA, USA) on a 2% agarose gel which had been run at 80 V for 30 min. The separated PCR products were visualized under UV light and photographed.

## 2.4. 16S rRNA Gene Sequencing and Phylogenetic Analysis

The amplified 16S rRNA PCR products were purified and sequenced at Macrogen (Seoul, Republic of Korea) using the ABI 3100 sequencer with Big Dye Terminator Kit v.3.1. The same 16S rRNA primer sequences used for PCR were used for sequencing. The sequences were assembled and trimmed using Geneious Prime and submitted to the Greengenes database (<https://mcentral.org/expert-database/greengenes>, accessed on 18 December 2022). The sequences of this study and reference sequences were aligned with Clustal W, and the evolutionary history was inferred using the Neighbor Joining Method [21] and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates). There were a total of 1552 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [21].

## 2.5. Whole Genome Sequencing (WGS) by MinION and Illumina

Nanopore WGS sequencing was carried out at the Molecular and Computational Biology of Fungi Laboratory, Federal University of Minas Gerais (UFMG). The DNA library was prepared with ligation Sequencing kit (SQKRAD004, Oxford Nanopore Technolo-

gies, Oxford, UK) according to the manufacturer's instructions. Libraries were sequenced with qualified FLO-MIN106 flow cells (the initial bias voltage was  $-210$  mV and the active pores number around 516) for 36 h (basecalling function was used, the reads sequences was filtered using a `min_score = 9`) on a MinION (Oxford Nanopore Technologies, Oxford, UK) [22].

The quality of the sequencing was verified through the FastQC v0.11.9 program (<https://github.com/s-andrews/FastQC>, accessed on 18 December 2022). The Porechop v0.2.4 program [18] was used for the detection and elimination of the adapters, as well as for the demultiplexing of the Nanopore reads. Possible sequencing errors were treated with the Canu v2.1.1 monitor correction module [23]. The de novo assembly based on de Bruijn graphs of corrected sequences was carried out through the Flye v2.8.3 [24]. The contigs obtained using de novo assembly were subjected to a polishing (correction of raw contigs) with the Racon v1.4.22 program [25], which took the read mappings made with BWA v0.7.17 [26].

The Illumina sequencing library was prepared from genomic DNA [1  $\mu$ g] using the NEBNext Fast DNA Fragmentation and Library Preparation Kit (New England Biolabs, Ipswich, MA, USA) following the manufacturer's recommendations. The library quality was assessed using the Agilent 2100 Bioanalyzer equipment, and the paired-end DNA sequencing was carried out in the Illumina HiSeq 2500 platform. After sequencing, the raw read quality was assessed using the FastQC v0.11.5 software (<https://github.com/s-andrews/FastQC>, accessed on 15 January 2020).

## 2.6. Hybrid Genome Sequence Assembly

MinION long-reads were assembled using the Racon pipeline with default parameters [24] while Illumina short reads were assembled using the (i) SPAdes version: 3.15.3 [27], (ii) Unicycler [18] and (iii) Edena [28] software with default parameters. Hybrid assemblies using Illumina and MinION reads were performed using the software (i) MaSuRCA [29], and (ii) Unicycler. Genome quality and completeness for each assembly were evaluated using QUAST v4.6.0 [30], and BUSCO v4 (Benchmarking Universal Single-Copy Orthologs) [31]. BUSCO analyses were performed using the database bacteria odb\_10.

## 2.7. Serotype Identification

The identification of the serotype was carried out from the de novo contigs, using the SeqSero2 v1.2.1 program [32].

## 2.8. Gene Annotation

The annotation of genes for both the bacterial and plasmid genomes was performed through the predictor, based on hidden Markov models, Prokka v1.14.6 [33].

## 2.9. Genome Similarity Assessment

*Salmonella enterica* genomes (16,638) were downloaded from the NCBI Genbank database in July 2022. Genomes with more than 500 contigs were removed, and contigs smaller than 500 bp were removed from the remaining genomes. Genome quality was evaluated with CheckM v1.0.13 [34], using completeness and contamination score of  $\geq 90\%$  and  $\leq 10\%$ , respectively. Genome-distance estimation of genomes was performed with Mash v2.2.1 [35]. Near-identical redundant genomes were removed using in-house scripts to cluster genomes assemblies sharing pairwise Mash distances less than 0.005 ( $\sim 99.95\%$  average nucleotide identity (ANI)) and cluster representatives were chosen based on assembly N50. Further, the genome dataset was taxonomically verified using the Genome Taxonomy Database (GTDB). To investigate the genomic relatedness of the *S. enterica* SE3 strain and Genbank genomes, a genome-distance tree was built using a combination-distance matrix of Mash and ANI values, computed with Mash v2.2.1 and fastANI [36], respectively.

### 2.10. Pangenome Analysis

The *S. enterica* pangenome analysis was performed with Roary v.3.6, using 90% identity threshold to determine gene clusters [37]. The Heaps law model was used to estimate the pangenome openness. Core genes (present in up to 95% of the genomes) were aligned with MAFFT v.7.394 [38]. SNPs were extracted from the core-genome alignment using SNP-sites v.2.3.3 [39]. The phylogenetic tree was constructed using IQ-TREE [40], with ascertainment bias correction under the model GTR+ASC, and bootstrap support was performed using 1000 replicates. The resulting phylogenetic tree was visualized and rendered with iTOL v4 [41].

### 2.11. Mobile Genetic Element Identification and Annotation

Genomic islands were identified using Island Viewer software ([www.pathogenomics.sfu.ca/islandviewer/upload/](http://www.pathogenomics.sfu.ca/islandviewer/upload/) (accessed on 18 December 2022)) [42], virulomes were detected using VFanalyser/VFDB ([www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi](http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi) (accessed on 18 December 2022)) [43], resistomes were identified using ResFinder-4.1 (<https://cge.cbs.dtu.dk/cgi-bin/webface.fcgi?jobid=61358037000023BC9E7A4C58> (accessed on 18 December 2022)) [44], and CARD (<https://card.mcmaster.ca/> (accessed on 18 December 2022)) [45], Prophages were identified using Phaster ([www.phaster.ca](http://www.phaster.ca) (accessed on 18 December 2022)) [46], phage defense systems were detected using PAD-LOC (<https://padloc.otago.ac.nz/padloc/> (accessed on 18 December 2022)) [47] and DefenseFinder (<https://defense-finder.mdmparis-lab.com/> (accessed on 18 December 2022)) [48]. SPIFinder 2.0 was used to detect Pathogenic Islands (<https://cge.cbs.dtu.dk/services/SPIFinder/> (accessed on 18 December 2022)) [49]. BRIG was used to draw the chromosomal *Salmonella* genomes (<http://brig.sourceforge.net/> (accessed on 18 December 2022)) [50].

## 3. Results

### 3.1. *Salmonella* Isolation and Characterization

Presumptive *Salmonella* were isolated from soil at the Subaé River using *Salmonella* selective growth media. Isolates showed typical *Salmonella* characteristics: on XLD colonies had a slightly transparent zone of reddish color and a black center, on Bismuth Sulfite Agar there were gray or brown-black colonies with or without metallic sheen and in SS agar the colonies were beige with black centers. In biochemical tests, good growth was seen in TSI, with acid and gas reactions at depth, an alkaline surface (red) and presence of H<sub>2</sub>S.

### 3.2. Analysis of 16S rRNA

The presumptive *Salmonella* isolates were confirmed by 16S rRNA PCR amplification [50,51] and sequencing, followed by a sequence query of the Greengenes database. Analysis of the queries returned coverage of 100% and an E value of 0, with 99.91% identity to the same sequence, *Salmonella enterica* serovar Enteritidis (ID: MT621365.1).

### 3.3. Whole Genome Sequencing of *Salmonella* Isolate SE3

One of the *Salmonella* Enteritidis isolates, designated SE3, was sequenced by Illumina HiSeq and Oxford Nanopore MinION technologies. The number of reads from HiSeq sequencing was 15,997,283 and the number of reads from MinION sequencing was 13,326, after preprocessing. The MinION long reads had an average size of 5.1 kb, and the longest read was 28.8 kb (Table 1).

**Table 1.** Summary of the Illumina HiSeq and Oxford Nanopore MinION reads statistics after preprocessing step.

Sequence Data	HiSeq	MinION
Reads	15,997,283	13,326
Total read bases (bp)	7,999,481	67,978,671
Mean coverage (%)	51,185	13,590
Longest read (bp)	151	28,841
Mean read length (bp)	150	5101
GC %	52.00	52.18
Genome size (bp)	4,688,543	4,709,033

### 3.4. Genome Assembly

Six whole genome sequence assembly strategies, including hybrid and non-hybrid, were tested on the HiSeq and MinION sequencing data from *Salmonella* SE3 (Table 2). For Illumina HiSeq assembly, Unicycler had the best performance with 31 contigs, a total length of 4,683,367 bp, largest contig of 1,262,086 bp and N50 of 478,501 bp (Table 2). The Unicycler hybrid assembly had the best performance for genome assembly overall, with 10 contigs, total length of 4,713,463 bp, largest contig of 519,108 bp and N50 of 2,750,500 bp (Table 2) (Figure 1). When measuring genome completeness, Unicycler HiSeq and Unicycler hybrid assembly had the same result, with 98.4 % of the orthologous (complete) genes searched, 99.4 % were single-copy genes, 1.6 % genes were not identified or missing, and there were no identified single and fragmented genes (Table 3).

**Table 2.** Summary statistics for the assembled genome of *Salmonella* SE3 using reads from Illumina HiSeq and Oxford Nanopore Technologies MinION.

Assembly Method	Racon	Unicycler	Edena	SPAdes	Unicycler	MaSuRCA
Sequence data	MinION	HiSeq	HiSeq	HiSeq	Hybrid	Hybrid
Number of contigs	2	31	41	50	10	39
Number of contigs ( $\geq 0$ bp)	2	65	54	111	18	42
Number of contigs ( $\geq 50$ kb)	2	15	4,475,114	4,566,140	4	24
Largest contigs	4,671,311	1,262,086	488,615	1,276,166	2,750,500	519,108
Total length ( $\geq 50$ kb)	4,730,597	4,683,367	4,701,851	4,805,245	4,713,463	4,585,719
GC (%)	52.18	52.14	52.15	51.85	52.16	52.15
N50	4,671,311	478,501	181,604	491,607	2,750,500	246,991
L50	1	3	8	3	1	7

### 3.5. Completeness of the Genome Annotation

The genome of *Salmonella* SE3 was annotated using Prokka and rRNA, tRNA and gene coding sequences were successfully identified (Table 4 and Table S1). *Salmonella* SE3 showed ~99.9% ANI with *Salmonella enterica* subsp. *enterica* serovar Enteritidis OLF-SE2-98984-6.

### 3.6. Genomic Relatedness of *Salmonella* SE3

Available *S. enterica* genomes in the GenBank database ( $n = 16,638$ , July 2022) were downloaded but after filtering for CheckM quality, removing highly fragmented and near-identical redundant genomes (see methods for details), the remaining dataset was 1598 genomes. Further genomic identity analysis with a combined matrix of all Mash and fastANI pairwise distances between the genomes identified a further 159 genomes with incorrect taxonomic assignment which were excluded. The distance tree built with the combined matrix showed that the *Salmonella* SE3 genome was located within the properly classified cluster of *S. enterica* genomes (Figure 2A). The *S. enterica* dataset comprised 1439 *S. enterica* genomes sharing Mash distance values up to 0.03 (~97% fastANI identity) (Figure 2B).

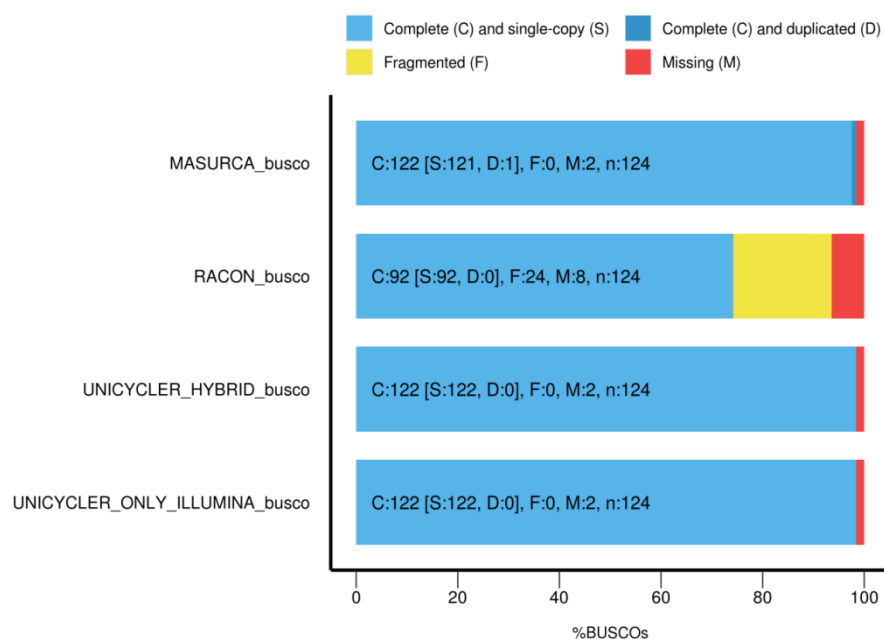


Figure 1. BUSCO completeness assessment of *Salmonella* SE3 genome.

Table 3. Completeness assessment of *Salmonella* SE3 assemblies using BUSCO software.

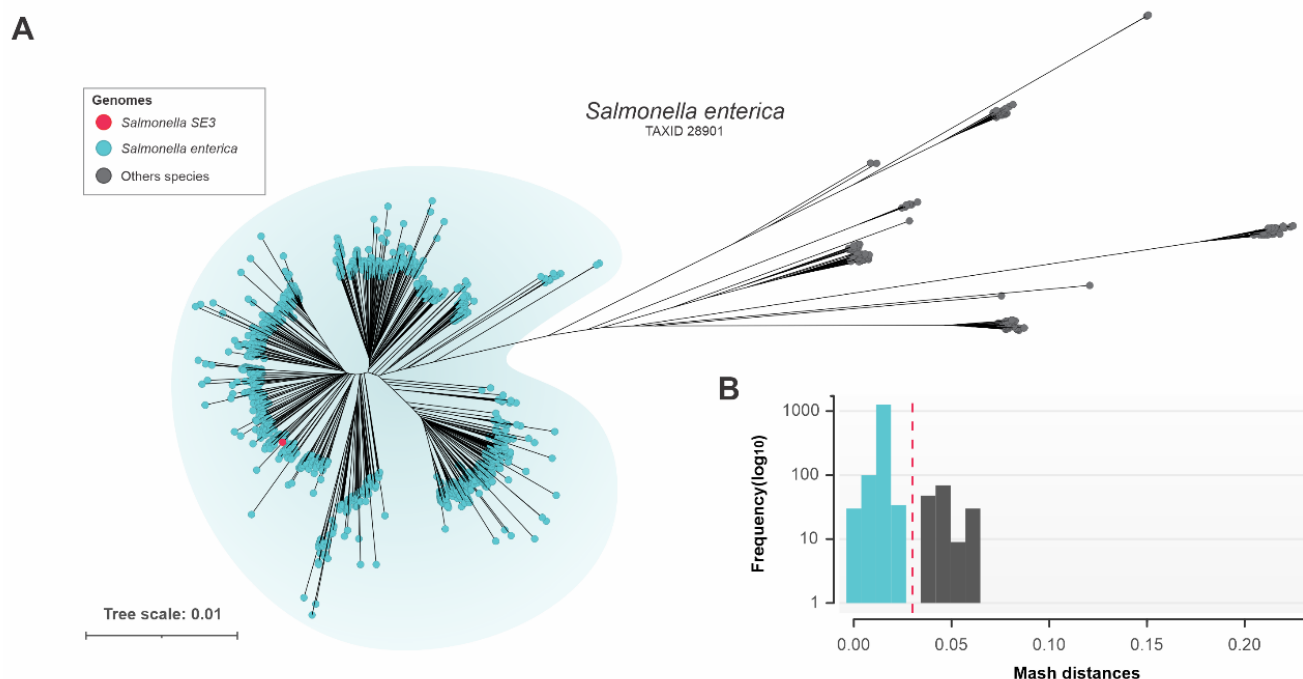
Assembly Method	Sequence Data	Complete (%)	Single Copy (%)	Duplicated (%)	Fragmented (%)	Missing (%)
Racon	MinION	74.2	74.2	0	19.4	6.4
Unicycler	HiSeq	98.4	98.4	0	0	1.64
Unicycler	Hybrid	98.4	98.4	0	0	1.64
MaSuRCA	Hybrid	98.4	97.6	0.8	0	1.64

Table 4. *Salmonella* SE3 genome features annotated by Prokka.

Annotated Genome	Features
rRNA	20
tRNA	87
Repeat region	2
CDS	4403
mRNA	1

### 3.7. Pangenome Analysis

The pangenome of 1439 *S. enterica* genomes is composed of 74,995 gene clusters, including a core genome (present in at least 95% of the genomes) of 2137 genes. The accessory genome comprises 3390 shell or shared genes (present from 15% to 95% of the genomes) and 69,352 cloud or singletons genes (present in up to 15% of the genomes) (Figure 3B). The Heaps Law estimate supports an open pangenome ( $\alpha = 0.52$ ) for *S. enterica*. (Figure 3A), indicating a high genetic diversity, and the capacity of this sympatric species to rapidly acquire exogenous DNA. We also performed a maximum-likelihood phylogenetic reconstruction using 292,004 SNPs extracted from core genes. This analysis revealed that *Salmonella* SE3 belongs to a monophyletic clade containing 23 *S. enterica* strains of serovar Enteritidis (Figure 3C).



**Figure 2.** Genome similarity of *Salmonella* SE3. (A) Distance tree of *Salmonella enterica* built using a combined matrix of all Mash and fastANI pairwise distances of *Salmonella* SE3 and 1598 genomes. Genomes classified by GTBD as *S. enterica* are shaded in blue. (B) Mash-distance values of *Salmonella* SE3 were calculated with 1598 *Salmonella* genomes. The maximum Mash-distance threshold (0.03) used to select genomes is represented by a dotted line.

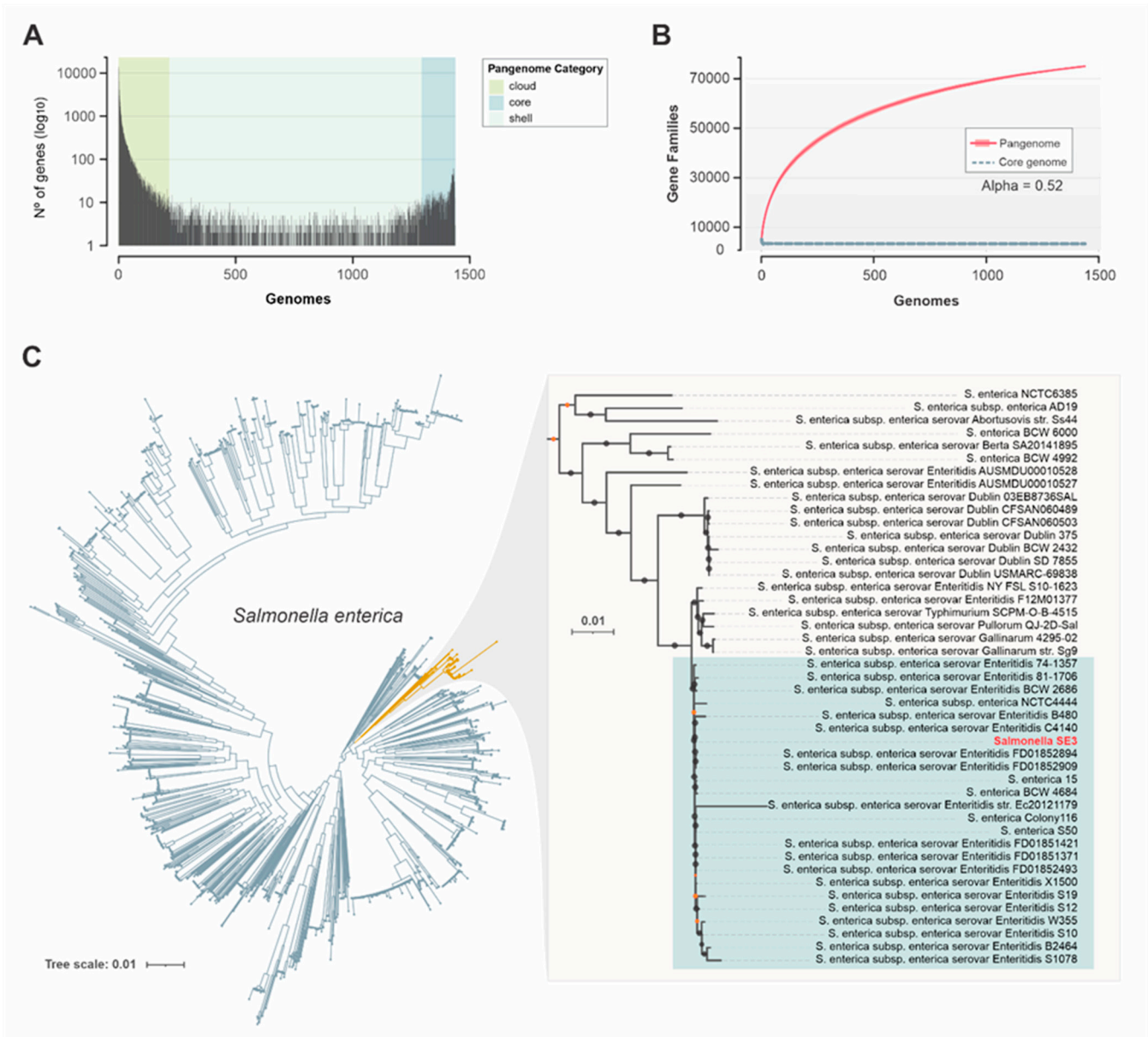
### 3.8. Genome Features

#### 3.8.1. Resistome Identification

Several resistance mechanisms were identified in *Salmonella* SE3 using the CARD database; resistance to aminoglycosides (alleles of *AAC(6′)-Iy*, *kdpE*, *baeR*), fluoroquinolones (alleles of *MdtK*, *emrB*, *emrR*, *sdiA*, *Escherichia coli* *acrA*, *acrB*, *rsmA*, *adeF*), macrolides (alleles of *Klebsiella pneumoniae* *KpnE*, *K. pneumoniae* *KpnF*, H-NS, CRP), monobactam (*golS*), nitroimidazole (*msbA*), tetracycline (*E. coli* *mdfA*), cephalosporin (*Haemophilus influenzae* PBP3 conferring resistance to beta-lactam antibiotics, *E. coli* EF-Tu mutants conferring resistance to Pulvomycin, *E. coli* *uhpT* with mutation conferring resistance to Fosfomycin, *E. coli* *glpT* with mutation conferring resistance to Fode novosfomycin), Figure 4.

According to their mechanism of resistance, the genes were classified as antibiotic efflux (*golS*, *baeR*, *MdtK*, *K. pneumoniae* *KpnE*, *K. pneumoniae* *KpnF*, H-NS, *sdiA*, *msbA*, *E. coli* *mdfA*, *kdpE*, *E. coli* *acrA*, *acrB*, *adeF*, CRP, *rsmA*, *emrB*, *emrR* and *marA*), antibiotic inactivation (*AAC(6′)-Iy*), antibiotic target alteration (*vanG*, *bacA*, *H. influenzae* PBP3 conferring resistance to beta-lactam antibiotics, *E. coli* *uhpT* with mutation conferring resistance to Fosfomycin, *E. coli* EF-Tu mutants conferring resistance to Pulvomycin, *E. coli* *glpT* with mutation conferring resistance to Fosfomycin, *E. coli* EF-Tu mutants conferring resistance to Pulvomycin, *pmrF*, *E. coli* *acrAB-tolC* with *marR* mutations conferring resistance to ciprofloxacin and tetracycline, *E. coli* *soxR* with mutation conferring antibiotic resistance and *E. coli* *soxS* with mutation conferring antibiotic resistance). Resfinder identified resistance against aminoglycosides: tobramycin (*aac(6′)-Iaa* (*aac(6′)-Iaa\_NC\_003197*) and amikacin (*aac(6′)-Iaa* (*aac(6′)-Iaa\_NC\_003197*)).



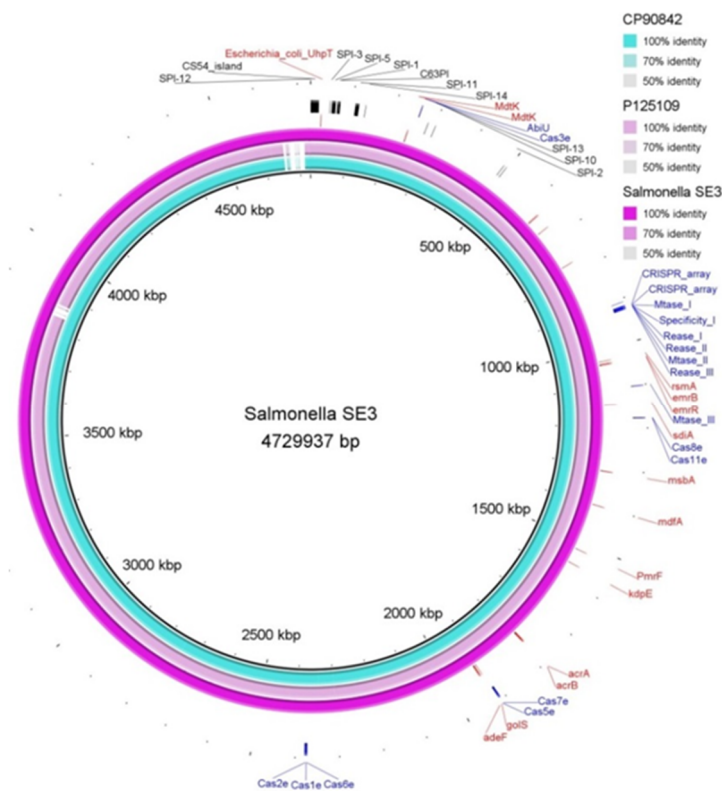


**Figure 3.** Pangenome of *Salmonella enterica* and phylogeny of *Salmonella* SE3. (A) Gene frequency of *S. enterica* pangenome. (B) Number of gene families in the *S. enterica* pangenome. The cumulative curve (in red) and an alpha value of the Heaps Law less than one (0.52) supports an open pangenome. (C) core-genome SNP tree of *Salmonella enterica* highlighting the phylogenetic group contained the *Salmonella* SE3 genome. The monophyletic clade containing the serovar Enteritidis of *S. enterica* is shaded in cool grey. Bootstrap values below and above 70% are represented by blue and dark-grey dots, respectively.

### 3.8.2. Viriome, Genomic Island and Pathogenic Island Identification

In total, 144 potential virulence genes were identified in *Salmonella* SE3 using VFanalyser/VFDB, some of the most important identified were *invA*, *sipA*, *sipB*, *sipC*, *fepA*, *sopA*, *sopB*, *sopD*, *sopE2*, *pefA*, *pefB*, *pefC*, *pefD* and *ssaB*. Genomic islands were detected using Island Viewer which uses three prediction methods: Integrated, IslandPath-DIMOB and SIGI-HMM. Twelve pathogenic islands were detected (Figure 4 and Table 5), and included virulence genes, secretion proteins, resistance genes, bacteriophage sequence regions, transposases and integrases. The gene *arsC*, encoding Arsenate reductase was identified in a genomic island. The *mdtK* gene (encoding multidrug resistance protein MdtK) was also

identified in the resistome analysis. Virulence genes identified using Island Viewer were very similar to those identified using VFanalyser/VFDB.



**Figure 4.** *Salmonella* SE3 antimicrobial resistance genes (red color), *Salmonella* Pathogenic Island (SP) (black color) and defense system (blue color) with two genomes of reference of *Salmonella* serovar Enteritidis (P125109 and CP9084.2).

### 3.9. Identification of Antiviral Defense Systems

Several antiviral defense system virulence genes were identified using PADLOC and DefenseFinder tools (Table 6). Both tools identified several systems: Cas type IE, CBASS type I, CRISPR array, restriction–modification (RM) RM type I, and RM type III. Similar antiviral systems and proteins were identified by PADLOC, except for AbiU and RM type II (Table 6 and Figure 4).

### 3.10. Prophage Identification

Of the prophages identified in *Salmonella* SE3 using PHASTER, two regions were intact, five regions were incomplete, and none were questionable (Table 7). Proteins were identified in the Gisfy and RE-2010 prophages including lysis, terminase, portal protein, protease, coat protein, tail shaft, attachment site, integrase, tail fiber and plate proteins.

**Table 5.** Pathogenicity islands identified in *Salmonella* SE3.

No	SPI	Identity	Query/Template Length	<i>Salmonella</i> Serotype	Insertion Site	Accession Number
1	SPI-1	99.7	2705/2705	Typhimurium SL1344	<i>fhlA/mutS</i>	AF148689
2	SPI-2	100	642/642	Gallinarum SGC_2	tRNA- <i>valV</i>	AY956827
3	SPI-3	99.05	738/738	Typhimurium 14028s	tRNA- <i>selC</i>	AJ000509
4	SPI-5	99.11	9069/9069	Typhimurium LT2	tRNA- <i>serT</i>	NC_003197
5	SPI-10	98.28	553/554	Gallinarum SGE_3	Unpublished	AY956839
6	SPI-11	98.54	9085/15686	Choleraesuis SC_B67	Gifsy-1	NC_006905
7	SPI-12	97.14	5766/11075	Choleraesuis SC_B67	tRNA- <i>pro</i>	NC_006905
8	SPI-13	100	341/341	Gallinarum SGA_10	tRNA- <i>pheV</i>	AY956834
9	SPI-14	99.8	501/501	Gallinarum SGA_8	Unpublished	AY956835
10	C63PI	99.12	4000/4000	Typhimurium SL1344	<i>fhlA</i>	AF128999
11	CS54	98.09	19669/25252	Typhimurium ATCC_14028	<i>xseA-yfgK</i>	AF140550
12	Unnamed	100	330/330	Enteritidis CMCC50041	–	JQ071613

**Table 6.** Antiviral defense systems of *Salmonella* SE3.

Number	System	Subtype	Tool	Reference
1	AbiU	AbiU	PADLOC	[46]
2	Cas type IE	Cas3e	PADLOC	[46]
3	Cas type IE	Cas8e	PADLOC	[46]
4	Cas type IE	Cas11e	PADLOC	[46]
5	Cas type IE	Cas7e	PADLOC	[46]
6	Cas type IE	Cas5e	PADLOC	[46]
7	Cas type IE	Cas6e	PADLOC	[46]
8	Cas type IE	Cas1e	PADLOC	[46]
9	Cas type IE	Cas2e	PADLOC	[46]
10	CBASS_type_I	Cyclase	PADLOC	[46]
11	CBASS_type_I	Effector	PADLOC	[46]
12	CRISPR array	CRISPR array	PADLOC	[46]
13	CRISPR array	CRISPR array	PADLOC	[46]
14	RM type I	Mtase I	PADLOC	[46]
15	RM type I	Specificity I	PADLOC	[46]
16	RM type I	Rease I	PADLOC	[46]
17	RM type II	Rease II	PADLOC	[46]
18	RM type II	Mtase II	PADLOC	[46]
19	RM type III	Rease III	PADLOC	[46]
20	RM type III	Mtase III	PADLOC	[46]
21	Cas Class1 subtype I E1	Cas3 I 5	DefenseFinder	[47]
22	Cas Class1 subtype I E1	Cas8e I E 1	DefenseFinder	[47]
23	Cas Class1 subtype I E1	Cas2gr11 I E 2	DefenseFinder	[47]
24	Cas Class1 subtype I E1	Cas7 I E 2	DefenseFinder	[47]
25	Cas Class1 subtype I E1	Cas5 I E 3	DefenseFinder	[47]
26	Cas Class1 subtype I E1	Cas6e I II IV V VI 1	DefenseFinder	[47]
27	Cas Class1 subtype I E1	Cas 1 I E 1	DefenseFinder	[47]
28	Cas Class1 subtype I E1	Cas2 I E 2	DefenseFinder	[47]
29	CBASS I 2	Cyclase SMODS	DefenseFinder	[47]
30	CBASS I 2	2TM Gros	DefenseFinder	[47]
31	RM Type III 2	Type III Reases	DefenseFinder	[47]
32	RM Type III 2	Type III Mtases	DefenseFinder	[47]
33	RM Type I 1	Type I S	DefenseFinder	[47]
34	RM Type I 1	Type I Mtases	DefenseFinder	[47]
35	RM Type I 1	Type I S	DefenseFinder	[47]
36	RM Type I 1	Type I Reases	DefenseFinder	[47]

Mtase = Methyltransferase I, Rease = restriction endonucleases.

**Table 7.** Prophage sequences annotated in *Salmonella* SE3 genome.

Completeness	Score	Proteins	Position	Best Match	Accession No.	GC (%)
Incomplete	60	27	805989–831780	<i>Shigella</i> phage POCJ13	NC_025434	48.7
Intact	150	40	1041034–1072153	<i>Salmonella</i> phage Gifsy-2	NC_010393	47.2
Incomplete	50	13	1276587–1286489	<i>Salmonella</i> phage Gifsy-2	NC_010393	46.7
Incomplete	30	9	1698977–1705339	<i>Shigella</i> phage POCJ13	NC_025434	45.6
Intact	150	49	1081056–1124788	<i>Salmonella</i> phage RE-2010	NC_019488	51.2
Incomplete	20	8	1435195–1442595	<i>Escherichia</i> phage 500465-2	NC_049343	53.2
Incomplete	40	9	29216–37324	<i>Salmonella</i> phage RE-2010	NC_019488	52.4

#### 4. Discussion

*Salmonella* SE3 was isolated from soil at the Subaé River in Santo Amaro, Brazil, a region contaminated with heavy metals and organic waste. The genome sequence of this isolate was determined using two sequencing technologies and six different bioinformatics strategies. Hybrid assembly showed the lowest number of contigs followed by MinION-alone assembly, with hybrid genome assembly resulting in a genome of 4.73 Mb, which was similar in size to that reported (4.68 Mb) for *Salmonella enterica* subsp. *enterica* serovar Enteritidis str. P125109 (NC\_011294.1) [52]. However, the GC content of the assembled genome (52.16%) was more similar to *Salmonella enterica* subsp. *enterica* serovar Enteritidis str. P125109 (NC\_011294.1) (52.17%) [52]. HiSeq assemblies have been traditionally considered the “gold standard” because MinION sequencing could introduce high numbers of errors and consequently may interfere with high-quality genome annotations due to reduced accuracy in gene prediction, producing a large number of misannotated genes [53,54]. However, the genome completeness of *Salmonella* SE3 with non-hybrid assembly and hybrid assembly were almost identical.

Phylogenetic analysis of the *Salmonella* SE3 genome revealed it was located within the properly classified cluster of *S. enterica*. During taxon analysis we identified 159 genomes with incorrect taxonomic classification, highlighting that it is important to confirm identity prior to undertaking phylogenetic analyses.

The pangenome analysis of *Salmonella* SE3, revealed the core genome was composed of 2137 genes and the accessory genome comprised 3390 shell genes and 69,352 cloud genes. This indicates *Salmonella* SE3 has an open pangenome with a diversity of unique genes. A study by Chand et al. [55] undertook a comparative genomic analysis of 44 genome sequences, representing 17 serovars of *S. enterica*, and concluded that the genus *Salmonella* displays an open pangenome, comprising a reservoir of 10,775 gene families. Of these 2847 constituted the core gene families, 4657 were dispensable or accessory gene families, and 3271 strain-specific gene families. Park et al. [56] constructed pangenomes of seven species to elucidate variations in the genetic contents of >27,000 genomes, as in our study, this work showed the pangenome of *Salmonella enterica* subsp. *enterica* was open. However, it is important to note that pangenome size is heavily influenced by the properties of the genomes used and variation would likely result in inconsistencies, and secondly, newly described genes are often included which results in open pangenomes [57].

The antimicrobial resistance gene profile of *Salmonella* SE3 identified genes potentially involved in resistance to aminoglycosides, fluoroquinolones, macrolides, a monobactam (*golS*), nitroimidazole (*msbA*), tetracycline and related drugs (*mdfA*), and cephalosporins. Other studies of *Salmonella* isolates from southern Brazil have also reported tetracycline (*mdfA*) and aminoglycoside (*aac(6′)-Iaa*) resistance genes, in addition to other genes such as *aac(3)-Iva*, *aph(3′)-Ib*, *aph(4)-Ia*, *aph(6)-Id*, *tet(34)* and *tet(A)* [57–61]. In the United States, additional antibiotic resistance mechanisms in *S. enterica* have been described [62], such as resistance to aminoglycosides (*aadA*, *aadB*, *aacC*, *aphA*, *strAB*),  $\beta$ -lactams (*bla*CMY-2, *PSE-1*, *TEM-1*), chloramphenicol (*cat1*, *cat2*, *cmlA*, *floR*), inhibitors of the folate pathway (*dfp*, *sul*), and tetracycline (*tetA*, *tetB*, *tetC*, *tetD*, *tetG*, and *tetR*), none of these resistance genes were detected in our study.

Ten *Salmonella* pathogenic islands were identified in *Salmonella* SE3 which is relatively high compared that reported for other *Salmonella* isolates. A *S. enterica* serovar Typhimurium isolate, ms202, from a patient in India possessed six *Salmonella* pathogenicity islands: SPI-1, SPI-2, SPI-3, SPI-4, SPI-5, and SPI-11 [63], but in our work, we did not identify SPI-4. The genes identified in SPI regions had similarity to known transporters, drug targets, and antibiotic-resistance genes, and in a subset of genomic islands, genes that facilitate the horizontal transfer of genes encoding numerous resistance and virulence factors of regions belonging to type III secretion systems (T3SS). Vilela et al. [64] analyzed six *Salmonella* Choleraesuis strains provided by the Brazilian *Salmonella* reference laboratory of the Oswaldo Cruz Foundation (FIOCRUZ-RJ), which receives *Salmonella* isolates from diverse isolation sources and regions of the country. Pathogenicity islands SPI-1, -2, -3, -4, -5, -9, -13, -14 and CS54 island were detected in five strains and SPI-11 in four strains. The majority of these SPI, with the exception of SPI 4 and SPI 11, were also detected in *Salmonella* SE3. SPI-1 and SPI-2 are known to be involved in the invasion of intestinal epithelial cells and survival and replication within phagocytic cells, respectively, through the formation of type 3 secretion systems, SPI-5 is associated with fluid secretion and inflammatory response and SPI-3, -4, -11, -13, -14 and CS54 are associated with *Salmonella* survival and adaptation to stresses within macrophages [65].

In total, 144 potential virulence genes were identified in *Salmonella* SE3. Some of these virulence genes are also found in other serovars of *Salmonella*. Borah et al. [66] investigated virulence genes in 88 *Salmonella* isolates recovered from humans and different species of animals. Among the 88 isolates, some virulence genes such *invA*, *sipA*, *sipB* and *sipC* were detected irrespective of the serovar, and these were also detected in *Salmonella* SE3. *fepA* was also present in a high percentage (64.7%) of isolates belonging to *Salmonella* serovars Enteritidis, Weltevreden, Typhi, Newport, Litchfield, Idikan and Typhimurium, as well as *Salmonella* SE3 and. Other virulence genes were present in varying percentages among the *Salmonella* serovars studied by Borah et al. [66] such as *sopB* (86.36%), *sopE2* (62.5%), *pefA* (79.54%) and *sefC* (51.14%); of these genes only *sefC* was not detected in *Salmonella* SE3. The virulence genes identified in *Salmonella* SE3 are involved in several different processes, such as the *invA* gene usually codes for a protein in the inner bacterial membrane that is responsible for the invasion of intestinal cells of the host [67,68]. The *fepA* gene encodes outer membrane receptor protein FepA, which participates in iron transport and plays a role in infection colonization in *Salmonella* [32]. T3SS-1 secretes proteins, termed effectors, across the inner and outer membranes of the bacterial cell. Some of the secreted effectors, including SipA, SipB and SipC are encoded by genes located on SPI1. The remaining effectors, including SopA, SopB, SopD, SopE and SopE2 are encoded by genes that are scattered around the *Salmonella* SE3 chromosome. Upon secretion the SipB, SipC, and SipD proteins are thought to form a complex in the eukaryotic membrane that is required for translocation of the remaining effectors into the host cell cytoplasm [69]. PefA is encoded by *Salmonella* SE3 and is the plasmid-encoded fimbrial major subunit antigen of *Salmonella* Typhimurium [70]. *Salmonella* plasmid-encoded fimbriae have been found to mediate adhesion to mouse intestinal epithelium [71].

The gene *arsC*, encoding arsenate reductase, was found in the genome of *Salmonella* SE3. Arsenate reductase is essential for arsenate resistance and transforms arsenate into arsenite which is extruded from the cell [72,73]. This is of interest as *Salmonella* SE3 was isolated from the soil of Subaé River where heavy metal concentrations were above reference values [74]. In addition, mussels (*Mytella charruana*) gathered from the same region also contained lead, arsenic and cadmium in concentrations above reference values [75]. Carvalho et al. [75] also determined the quality of soils in 39 households from nearby Santo Amaro City, and the Residential Investigation Value (RIV) was exceeded by Lead (23.1% of the samples), Cadmium (7.7%), Nickel (2.6%), Zinc (25.6%), Arsenic (2.6%), and Antimony (7.7%).

Several virus defence systems were detected in *Salmonella* SE3, including CRISPR-Cas type IE, CBASS type I, and RM type I and III systems. Similar antiviral systems and subtypes were identified by the PADLOC and DefenseFinder tools, except for AbiU and

RM type II which were only identified by PADLOC. Most bacteria, including *Salmonella*, possess multiple antiviral defence systems that protect against infection by phages and mobile genetic elements [47].

Seven prophages were detected in the *Salmonella* SE3 genome, two were intact, and five were incomplete. By comparison, in *S. enterica* Typhimurium ms202 nine prophages were detected, two were intact, five were incomplete and two were questionable [63]. Moreover, *Salmonella* SE3 had not only *Salmonella* prophage sequences (e.g. phage RE-2010) but also prophages annotated as belonging to closely related genera *Shigella* (phage POCJ13) and *Escherichia* (phage 500465-2), which may indicate horizontal gene transfer or polyvalent phages. A previous study has reported that phage populations in *S. enterica* contribute to horizontal gene transfer, including virulence and virulence-related genes within the sub-species [76–79]. Further studies on *Salmonella* phages may uncover the receptor-interaction mechanisms between phages and hosts which may lead to improving phage therapy as an option for the treatment or control of *Salmonella*.

## 5. Conclusions

Salmonellosis is a healthcare issue around the world, so genomic analysis of *Salmonella* isolates could be a key determinant for better control of salmonellosis. Our study showed the effectiveness of a hybrid sequence assembly approach for environmental *Salmonella* genome analysis using HiSeq and MinION data. *Salmonella* SE3 was determined to belong to a monophyletic clade containing 23 *S. enterica* strains of serovar Enteritidis. The hybrid genome assembly enabled mobile genetic elements, genomic islands, *Salmonella* Pathogenicity Islands, antiviral systems, antimicrobial resistance genes, virulence genes, and prophages to be identified in *Salmonella* SE3. Furthermore, a gene encoding heavy metal resistance, *arsC*, was detected. These data are important to inform the control of *Salmonella* and heavy metal pollution in the Santo Amaro region of Brazil.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms11010111/s1>, Supplementary Material (Table S1: Annotation of Prokka).

**Author Contributions:** Conceptualization, C.B., A.G.-N. and D.X.R.-C.; methodology, D.X.R.-C., C.B., F.P.-S., R.G.B., L.M.R.T. and L.T.S.d.O.S.; software, F.P.-S., L.M.R.T. and T.J.S.; validation, D.X.R.-C., F.P.-S., L.M.R.T. and T.J.S.; formal analysis, C.B., A.G.-N., D.X.R.-C. and T.M.V.; investigation, A.G.-N., C.B. and D.X.R.-C.; resources, A.G.-N., C.B., V.A.d.C.A. and B.B.; data curation, F.P.-S., L.M.R.T. and T.J.S.; writing—original draft preparation, D.X.R.-C.; writing—review and editing, C.B.; visualization, F.P.-S.; supervision, A.G.-N. and C.B.; project administration, A.G.-N., C.B., V.A.d.C.A. and B.B.; funding acquisition, A.G.-N., C.B., V.A.d.C.A. and B.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES)-Finance Code 001.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are grateful to Lucas, Gorete and Elinalva from State University of Feira de Santana (UEFS) for the donation of media growth culture. We are also grateful to Angel for introducing us to bioinformatics.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Hernández-Reyes, C.; Schikora, A. Salmonella, a Cross-Kingdom Pathogen Infecting Humans and Plants. *FEMS Microbiol. Lett.* **2013**, *343*, 1–7. [CrossRef] [PubMed]
2. Majowicz, S.E.; Musto, J.; Scallan, E.; Angulo, F.J.; Kirk, M.; O'Brien, S.J.; Jones, T.F.; Fazil, A.; Hoekstra, R.M.; International Collaboration on Enteric Disease “Burden of Illness” Studies. The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2010**, *50*, 882–889. [CrossRef] [PubMed]
3. The Global Burden of Non-Typhoidal Salmonella Invasive Disease: A Systematic Analysis for the Global Burden of Disease Study 2017—The Lancet Infectious Diseases. Available online: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(19\)30418-9/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(19)30418-9/fulltext) (accessed on 2 November 2022).
4. Das, S.; Ray, S.; Ryan, D.; Sahu, B.; Suar, M. Identification of a Novel Gene in ROD9 Island of *Salmonella* Enteritidis Involved in the Alteration of Virulence-Associated Genes Expression. *Virulence* **2018**, *9*, 348–362. [CrossRef] [PubMed]
5. Saleh, S.; Van Puyvelde, S.; Staes, A.; Timmerman, E.; Barbé, B.; Jacobs, J.; Gevaert, K.; Deborggraeve, S. *Salmonella* Typhi, Paratyphi A, Enteritidis and Typhimurium Core Proteomes Reveal Differentially Expressed Proteins Linked to the Cell Surface and Pathogenicity. *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007416. [CrossRef]
6. Rabsch, W.; Andrews, H.L.; Kingsley, R.A.; Prager, R.; Tschäpe, H.; Adams, L.G.; Bäuml, A.J. *Salmonella enterica* Serotype Typhimurium and Its Host-Adapted Variants. *Infect. Immun.* **2002**, *70*, 2249–2255. [CrossRef]
7. Carden, S.; Okoro, C.; Dougan, G.; Monack, D. Non-Typhoidal *Salmonella* Typhimurium ST313 Isolates That Cause Bacteremia in Humans Stimulate Less Inflammation than ST19 Isolates Associated with Gastroenteritis. *Pathog. Dis.* **2015**, *73*, ftu023. [CrossRef]
8. Kipper, D.; Mascitti, A.K.; De Carli, S.; Carneiro, A.M.; Streck, A.F.; Fonseca, A.S.K.; Ikuta, N.; Lunge, V.R. Emergence, Dissemination and Antimicrobial Resistance of the Main Poultry-Associated *Salmonella* Serovars in Brazil. *Vet. Sci.* **2022**, *9*, 405. [CrossRef]
9. Allard, M.W.; Strain, E.; Melka, D.; Bunning, K.; Musser, S.M.; Brown, E.W.; Timme, R. Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. *J. Clin. Microbiol.* **2016**, *54*, 1975–1983. [CrossRef]
10. Gilchrist, C.A.; Turner, S.D.; Riley, M.F.; Petri, W.A.; Hewlett, E.L. Whole-Genome Sequencing in Outbreak Analysis. *Clin. Microbiol. Rev.* **2015**, *28*, 541–563. [CrossRef]
11. Utturkar, S.M.; Klingeman, D.M.; Land, M.L.; Schadt, C.W.; Doktycz, M.J.; Pelletier, D.A.; Brown, S.D. Evaluation and Validation of de Novo and Hybrid Assembly Techniques to Derive High-Quality Genome Sequences. *Bioinform. Oxf. Engl.* **2014**, *30*, 2709–2716. [CrossRef]
12. Ashton, P.M.; Nair, S.; Dallman, T.; Rubino, S.; Rabsch, W.; Mwaigwisya, S.; Wain, J.; O’Grady, J. MinION Nanopore Sequencing Identifies the Position and Structure of a Bacterial Antibiotic Resistance Island. *Nat. Biotechnol.* **2015**, *33*, 296–300. [CrossRef]
13. Genome Assembly Using Nanopore-Guided Long and Error-Free DNA Reads | BMC Genomics | Full Text. Available online: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-1519-z> (accessed on 2 November 2022).
14. Genomic Analyses of Multidrug-Resistant *Salmonella* Indiana, Typhimurium, and Enteritidis Isolates Using MinION and MiSeq Sequencing Technologies | PLOS ONE. Available online: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0235641> (accessed on 2 November 2022).
15. Rang, F.J.; Kloosterman, W.P.; de Ridder, J. From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy. *Genome Biol.* **2018**, *19*, 90. [CrossRef] [PubMed]
16. The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/27887629/> (accessed on 2 November 2022).
17. Antipov, D.; Korobeynikov, A.; McLean, J.S.; Pevzner, P.A. HybridSPAdes: An Algorithm for Hybrid Assembly of Short and Long Reads. *Bioinformatics* **2016**, *32*, 1009–1015. [CrossRef] [PubMed]
18. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/28594827/> (accessed on 2 November 2022).
19. Asai, T.; Otagiri, Y.; Osumi, T.; Namimatsu, T.; Hirai, H.; Sato, S. Isolation of *Salmonella* from Diarrheic Feces of Pigs. *J. Vet. Med. Sci.* **2002**, *64*, 159–160. [CrossRef]
20. Senthilraj, R.; Prasad, G.S.; Janakiraman, K. Sequence-based identification of microbial contaminants in non-parenteral products. *Braz. J. Pharm.* **2016**, *52*, 329–336. [CrossRef]
21. Molecular Evolution and Phylogenetics: Nei, Masatoshi, Kumar, Sudhir + Free Shipping. Available online: <https://www.amazon.com/Molecular-Evolution-Phylogenetics-Masatoshi-Nei/dp/0195135857> (accessed on 2 November 2022).
22. Tomé, L.M.R.; da Silva, F.F.; Fonseca, P.L.C.; Mendes-Pereira, T.; Azevedo, V.A.D.C.; Brenig, B.; Góes-Neto, A. Hybrid Assembly Improves Genome Quality and Completeness of *Trametes villosa* CCMB561 and Reveals a Huge Potential for Lignocellulose Breakdown. *J. Fungi* **2022**, *8*, 142. [CrossRef]
23. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef]
24. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef]

25. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]
26. [PDF] Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM | Semantic Scholar. Available online: <https://www.semanticscholar.org/paper/Aligning-sequence-reads%2C-clone-sequences-and-with-Li/74574ee09030e8aad48fa349eb9b054e2f95ceb> (accessed on 2 November 2022).
27. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **2012**, *19*, 455–477. [CrossRef]
28. Hernandez, D.; François, P.; Farinelli, L.; Osterås, M.; Schrenzel, J. De Novo Bacterial Genome Sequencing: Millions of Very Short Reads Assembled on a Desktop Computer. *Genome Res.* **2008**, *18*, 802–809. [CrossRef]
29. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA Genome Assembler. *Bioinform. Oxf. Engl.* **2013**, *29*, 2669–2677. [CrossRef]
30. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUASt: Quality Assessment Tool for Genome Assemblies. *Bioinform. Oxf. Engl.* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]
31. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinform. Oxf. Engl.* **2015**, *31*, 3210–3212. [CrossRef] [PubMed]
32. Zhang, B.; Fan, Y.; Wang, M.; Lv, J.; Zhang, H.; Sun, L.; Du, H. Effect of RpoE on the Non-Coding RNA Expression Profiles of *Salmonella* Enterica Serovar Typhi under the Stress of Ampicillin. *Curr. Microbiol.* **2020**, *77*, 2405–2412. [CrossRef] [PubMed]
33. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinform. Oxf. Engl.* **2014**, *30*, 2068–2069. [CrossRef]
34. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef] [PubMed]
35. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast Genome and Metagenome Distance Estimation Using MinHash. *Genome Biol.* **2016**, *17*, 132. [CrossRef] [PubMed]
36. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries | Nature Communications. Available online: <https://www.nature.com/articles/s41467-018-07641-9> (accessed on 2 November 2022).
37. 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. Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis. *Bioinform. Oxf. Engl.* **2015**, *31*, 3691–3693. [CrossRef]
38. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
39. Page, A.J.; Taylor, B.; Delaney, A.J.; Soares, J.; Seemann, T.; Keane, J.A.; Harris, S.R. SNP-Sites: Rapid Efficient Extraction of SNPs from Multi-FASTA Alignments. *Microb. Genomics* **2016**, *2*, e000056. [CrossRef] [PubMed]
40. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
41. Interactive Tree of Life (ITOL) v4: Recent Updates and New Developments—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/30931475/> (accessed on 2 November 2022).
42. Bertelli, C.; Laird, M.R.; Williams, K.P.; Simon Fraser University Research Computing Group; Lau, B.Y.; Hoad, G.; Winsor, G.L.; Brinkman, F.S.L. IslandViewer 4: Expanded Prediction of Genomic Islands for Larger-Scale Datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [CrossRef]
43. Liu, B.; Zheng, D.; Zhou, S.; Chen, L.; Yang, J. VFDB 2022: A General Classification Scheme for Bacterial Virulence Factors. *Nucleic Acids Res.* **2022**, *50*, D912–D917. [CrossRef] [PubMed]
44. Alcock, B.P.; Raphenya, A.R.; Lau, T.T.Y.; Tsang, K.K.; Boucharde, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L.V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [CrossRef]
45. Jia, B.; Raphenya, A.R.; Alcock, B.; Waglechner, N.; Guo, P.; Tsang, K.K.; Lago, B.A.; Dave, B.M.; Pereira, S.; Sharma, A.N.; et al. CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2017**, *45*, D566–D573. [CrossRef]
46. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A Better, Faster Version of the PHAST Phage Search Tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [CrossRef]
47. PADLOC: A Web Server for the Identification of Antiviral Defence Systems in Microbial Genomes | Nucleic Acids Research | Oxford Academic. Available online: <https://academic.oup.com/nar/article/50/W1/W541/6593116?login=false> (accessed on 2 November 2022).
48. Tesson, F.; Hervé, A.; Mordret, E.; Touchon, M.; d’Humières, C.; Cury, J.; Bernheim, A. Systematic and Quantitative View of the Antiviral Arsenal of Prokaryotes. *Nat. Commun.* **2022**, *13*, 2561. [CrossRef]
49. Roer, L.; Hendriksen, R.S.; Leekitcharoenphon, P.; Lukjancenko, O.; Kaas, R.S.; Hasman, H.; Aarestrup, F.M. Is the Evolution of *Salmonella* Enterica Subsp. Enterica Linked to Restriction-Modification Systems? *mSystems* **2016**, *1*, e00009-16. [CrossRef]
50. Alikhan, N.-F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple Prokaryote Genome Comparisons. *BMC Genomics* **2011**, *12*, 402. [CrossRef]



51. dos Santos, H.R.M.; Argolo, C.S.; Argôlo-Filho, R.C.; Loguercio, L.L. A 16S rDNA PCR-Based Theoretical to Actual Delta Approach on Culturable Mock Communities Revealed Severe Losses of Diversity Information. *BMC Microbiol.* **2019**, *19*, 74. [[CrossRef](#)] [[PubMed](#)]
52. Vaid, R.K.; Thakur, Z.; Anand, T.; Kumar, S.; Tripathi, B.N. Comparative Genome Analysis of *Salmonella enterica* Serovar Gallinarum Biovars Pullorum and Gallinarum Decodes Strain Specific Genes. *PLoS ONE* **2021**, *16*, e0255612. [[CrossRef](#)] [[PubMed](#)]
53. González-Escalona, N.; Allard, M.A.; Brown, E.W.; Sharma, S.; Hoffmann, M. Nanopore Sequencing for Fast Determination of Plasmids, Phages, Virulence Markers, and Antimicrobial Resistance Genes in Shiga Toxin-Producing *Escherichia Coli*. *PloS ONE* **2019**, *14*, e0220494. [[CrossRef](#)]
54. Rapid, Multiplexed, Whole Genome and Plasmid Sequencing of Foodborne Pathogens Using Long-Read Nanopore Technology | Scientific Reports. Available online: <https://www.nature.com/articles/s41598-019-52424-x> (accessed on 2 November 2022).
55. Chand, Y.; Alam, M.A.; Singh, S. Pan-Genomic Analysis of the Species *Salmonella enterica*: Identification of Core Essential and Putative Essential Genes. *Gene Rep.* **2020**, *20*, 100669. [[CrossRef](#)]
56. Park, S.-C.; Lee, K.; Kim, Y.O.; Won, S.; Chun, J. Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size. *Front. Microbiol.* **2019**, *10*, 834. [[CrossRef](#)] [[PubMed](#)]
57. de Oliveira, F.A.; Brandelli, A.; Tondo, E.C. Antimicrobial Resistance in *Salmonella* Enteritidis from Foods Involved in Human Salmonellosis Outbreaks in Southern Brazil. *New Microbiol.* **2006**, *29*, 49–54. [[PubMed](#)]
58. Vaz, C.S.L.; Streck, A.F.; Michael, G.B.; Marks, F.S.; Rodrigues, D.P.; Dos Reis, E.M.F.; Cardoso, M.R.I.; Canal, C.W. Anti-microbial Resistance and Subtyping of *Salmonella enterica* Subspecies Enterica Serovar Enteritidis Isolated from Human Outbreaks and Poultry in Southern Brazil. *Poult. Sci.* **2010**, *89*, 1530–1536. [[CrossRef](#)]
59. Campioni, F.; Moratto Bergamini, A.M.; Falcão, J.P. Genetic Diversity, Virulence Genes and Antimicrobial Resistance of *Salmonella* Enteritidis Isolated from Food and Humans over a 24-Year Period in Brazil. *Food Microbiol.* **2012**, *32*, 254–264. [[CrossRef](#)]
60. Achtman, M.; Wain, J.; Weill, F.-X.; Nair, S.; Zhou, Z.; Sangal, V.; Krauland, M.G.; Hale, J.L.; Harbottle, H.; Uesbeck, A.; et al. Multilocus Sequence Typing as a Replacement for Serotyping in *Salmonella* Enterica. *PLoS Pathog.* **2012**, *8*, e1002776. [[CrossRef](#)]
61. Campioni, F.; Souza, R.A.; Martins, V.V.; Stehling, E.G.; Bergamini, A.M.M.; Falcão, J.P. Prevalence of GyrA Mutations in Nalidixic Acid-Resistant Strains of *Salmonella* Enteritidis Isolated from Humans, Food, Chickens, and the Farm Environment in Brazil. *Microb. Drug Resist. Larchmt. N* **2017**, *23*, 421–428. [[CrossRef](#)]
62. Frye, J.; Jackson, C. Genetic Mechanisms of Antimicrobial Resistance Identified in *Salmonella enterica*, *Escherichia coli*, and *Enterococcus* Spp. Isolated from U.S. Food Animals. *Front. Microbiol.* **2013**, *4*, 135.
63. Mohakud, N.K.; Panda, R.K.; Patra, S.D.; Sahu, B.R.; Ghosh, M.; Kushwaha, G.S.; Misra, N.; Suar, M. Genome Analysis and Virulence Gene Expression Profile of a Multi Drug Resistant *Salmonella enterica* Serovar Typhimurium Ms202. *Gut Pathog.* **2022**, *14*, 28. [[CrossRef](#)]
64. Vilela, F.P.; Rodrigues, D.D.P.; Ferreira, J.C.; Darini, A.L.D.C.; Allard, M.W.; Falcão, J.P. Genomic Characterization of *Salmonella enterica* Serovar Choleraesuis from Brazil Reveals a Swine Gallbladder Isolate Harboring Colistin Resistance Gene Mcr-1.1. *Braz. J. Microbiol. Publ. Braz. Soc. Microbiol.* **2022**, *53*, 1799–1806. [[CrossRef](#)] [[PubMed](#)]
65. Seribelli, A.A.; da Silva, P.; Frazão, M.R.; Kich, J.D.; Allard, M.W.; Falcão, J.P. Phylogenetic Relationship and Genomic Characterization of *Salmonella* Typhimurium Strains Isolated from Swine in Brazil. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2021**, *93*, 104977. [[CrossRef](#)]
66. Borah, P.; Dutta, R.; Das, L.; Hazarika, G.; Choudhury, M.; Deka, N.K.; Malakar, D.; Hussain, M.I.; Barkalita, L.M. Prevalence, Antimicrobial Resistance and Virulence Genes of *Salmonella* Serovars Isolated from Humans and Animals. *Vet. Res. Commun.* **2022**, *46*, 799–810. [[CrossRef](#)] [[PubMed](#)]
67. Sharma, I. Detection of InvA Gene in Isolated *Salmonella* from Marketed Poultry Meat by PCR Assay. *J. Food Process. Technol.* **2016**, *7*, 2. [[CrossRef](#)]
68. El-Sebay, N.A.; Shady, H.M.A.; El-Zeedy, S.A.E.-R.; Samy, A.A. InvA Gene Sequencing of *Salmonella* Typhimurium Isolated from Egyptian Poultry. *Asian J. Sci. Res.* **2017**, *10*, 194–202. [[CrossRef](#)]
69. Raffatellu, M.; Wilson, R.P.; Chessa, D.; Andrews-Polymenis, H.; Tran, Q.T.; Lawhon, S.; Khare, S.; Adams, L.G.; Bäumlner, A.J. SipA, SopA, SopB, SopD, and SopE2 Contribute to *Salmonella* Enterica Serotype Typhimurium Invasion of Epithelial Cells. *Infect. Immun.* **2005**, *73*, 146–154. [[CrossRef](#)]
70. Woodward, M.J.; Allen-Vercoe, E.; Redstone, J.S. Distribution, Gene Sequence and Expression in Vivo of the Plasmid Encoded Fimbrial Antigen of *Salmonella* Serotype Enteritidis. *Epidemiol. Infect.* **1996**, *117*, 17–28. [[CrossRef](#)]
71. Nicholson, B.; Low, D. DNA Methylation-Dependent Regulation of Pef Expression in *Salmonella* Typhimurium. *Mol. Microbiol.* **2000**, *35*, 728–742. [[CrossRef](#)]
72. Jackson, C.R.; Dugas, S.L. Phylogenetic Analysis of Bacterial and Archaeal ArsC Gene Sequences Suggests an Ancient, Common Origin for Arsenate Reductase. *BMC Evol. Biol.* **2003**, *3*, 18. [[CrossRef](#)]
73. Pei, R.; Zhang, L.; Duan, C.; Gao, M.; Feng, R.; Jia, Q.; Huang, Z. (Jacky) Investigation of Stress Response Genes in Antimicrobial Resistant Pathogens Sampled from Five Countries. *Processes* **2021**, *9*, 927. [[CrossRef](#)]
74. Andrade, M.F.D.; Moraes, L.R.S. Lead Contamination in Santo Amaro Defies Decades of Research and Delayed Reaction on the Part of the Public Authorities. *Ambiente Soc.* **2013**, *16*, 63–80. [[CrossRef](#)]
75. Carvalho, F.M.; Tavares, T.M.; Lins, L. Soil Contamination by a Lead Smelter in Brazil in the View of the Local Residents. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2166. [[CrossRef](#)]

76. Hardt, W.-D.; Urlaub, H.; Galán, J.E. A Substrate of the Centisome 63 Type III Protein Secretion System of *Salmonella* Typhimurium Is Encoded by a Cryptic Bacteriophage. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2574–2579. [[CrossRef](#)] [[PubMed](#)]
77. Figueroa-Bossi, N.; Uzzau, S.; Maloriol, D.; Bossi, L. Variable Assortment of Prophages Provides a Transferable Repertoire of Pathogenic Determinants in *Salmonella*. *Mol. Microbiol.* **2001**, *39*, 260–271. [[CrossRef](#)] [[PubMed](#)]
78. Switt, A.I.M.; Sulakvelidze, A.; Wiedmann, M.; Kropinski, A.M.; Wishart, D.S.; Poppe, C.; Liang, Y. Salmonella Phages and Prophages: Genomics, Taxonomy, and Applied Aspects. *Methods Mol. Biol. Clifton NJ* **2015**, *1225*, 237–287. [[CrossRef](#)]
79. Worley, J.; Meng, J.; Allard, M.W.; Brown, E.W.; Timme, R.E. *Salmonella enterica* Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel Patterns of Horizontally Acquired Genetic Elements. *mBio* **2018**, *9*, e02303-18. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.