Check for updates

*CORRESPONDENCE
Jörg Linde
✉ joerg.linde@fli.de

Oxford nanopore technologies—a valuable tool to generate whole-genome sequencing data for *in silico* serotyping and the detection of genetic markers in *Salmonella*

Christine Thomas[1,2], Ulrich Methner[1], Manja Marz[2] and
Jörg Linde[1]*

[1]Institute of Bacterial Infections and Zoonoses, Federal Research Institute for Animal Health, Friedrich-
Loeffler-Institute, Jena, Germany, [2]RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller
University Jena, Jena, Germany

Bacteria of the genus *Salmonella* pose a major risk to livestock, the food economy,
and public health. *Salmonella* infections are one of the leading causes of food
poisoning. The identification of serovars of *Salmonella* achieved by their diverse
surface antigens is essential to gain information on their epidemiological context.
Traditionally, slide agglutination has been used for serotyping. In recent years, whole-
genome sequencing (WGS) followed by *in silico* serotyping has been established as an
alternative method for serotyping and the detection of genetic markers for *Salmonella*.
Until now, WGS data generated with Illumina sequencing are used to validate *in silico*
serotyping methods. Oxford Nanopore Technologies (ONT) opens the possibility to
sequence ultra-long reads and has frequently been used for bacterial sequencing. In
this study, ONT sequencing data of 28 *Salmonella* strains of different serovars with
epidemiological relevance in humans, food, and animals were taken to investigate
the performance of the *in silico* serotyping tools SISTR and SeqSero2 compared
to traditional slide agglutination tests. Moreover, the detection of genetic markers
for resistance against antimicrobial agents, virulence, and plasmids was studied by
comparing WGS data based on ONT with WGS data based on Illumina. Based on the
ONT data from flow cell version R9.4.1, *in silico* serotyping achieved an accuracy of
96.4 and 92% for the tools SISTR and SeqSero2, respectively. Highly similar sets of
genetic markers comparing both sequencing technologies were identified. Taking
the ongoing improvement of basecalling and flow cells into account, ONT data can
be used for *Salmonella in silico* serotyping and genetic marker detection.

KEYWORDS

Salmonella, nanopore sequencing, whole-genome sequencing, *in silico* serotyping,
surveillance

1. Introduction

Salmonella are food-borne pathogens causing over 1.35 million infections every year in
the United States (1). In Europe, 60,050 human salmonellosis cases were registered in 2021
(2). These Gram-negative microorganisms can often be found in the intestinal tract of
animals and humans. Different species of farm animals, as well as pets, reptiles, and zoo

animals, might serve as reservoirs for *Salmonella*. The pathogen is transmitted to humans mainly *via* contaminated food, water, and by direct contact with infected hosts. The course of infection varies between asymptomatic intestinal colonization and severe systemic disease and is affected by factors of the host and the serovar involved (3).

The genus *Salmonella* comprises two species, namely *Salmonella enterica* and *Salmonella bongori*. The species can be further differentiated into serovars according to differences in their O and H antigen structure. Since serovars of *Salmonella* differ significantly in their virulence, pathogenesis, and host specificity, serotyping plays an important role in the surveillance and controlling of *Salmonella* infections and outbreaks (4).

The traditional method for the determination is serology-based serotyping according to the White–Kauffmann–Le Minor Scheme, which was first published in 1934 (5). Here, the liposaccharide O antigens, as well as the flagellar H1 and H2 antigens, are used in slide agglutination tests to determine the serovar. The O antigens are determined by lipopolysaccharides in the outer membrane of the bacterium. The H antigens may occur in two forms in the flagellum, together or alone, called H1 antigen and H2 antigen. The different variants of antigens are numbered, and the combination of the numbers reflects the antigenic formula of each serovar. The WHO collaborating center for reference and research on *Salmonella* updates regularly the serotyping scheme. Currently, there are over 2600 recognized serovars (6).

The rapid identification of *Salmonella* serovars is essential to gain information on monitoring, epidemiology, and intervention strategies. However, serology-based methods of serotyping are time-intensive and partly limited in their validity (7). *Salmonella* organisms that do not express the O and H antigens in full form, called rough strains, cannot be unequivocally determined by agglutination tests (8). Furthermore, there are strains of different serovars which do not enable the induction of the second H-phase, called monophasic variants. Therefore, the identification of the clear antigenic formula of rough strains and monophasic strains is limited with the traditional slide agglutination test (8).

In recent years, whole-genome sequencing (WGS) of bacteria has been increasingly applied as the costs and time investment of performing WGS and subsequent data analysis have been decreasing (9, 10). WGS can not only reveal information on genotypes of the pathogens but also enable the detection of genetic markers for antimicrobial resistance (AMR), resistance against disinfection agents, plasmids, or virulence, as well as *Salmonella* pathogenicity islands (SPIs) (11, 12). This information can be crucial in outbreak investigations and epidemiological studies and is helpful for effective control strategies. Additionally, WGS has been used for the identification of *Salmonella* serovars and might help overcome the limitations by classical slide agglutination using *Salmonella* antisera (7, 13–16). Tools for serotyping based on WGS data are SISTR developed by Yoshida et al. (14) and SeqSero2 published by Zhang et al. (15).

The tool SISTR analyzes the genetic variations within the O antigens which are encoded by the flippase (*wzx*) and polymerase (*wzy*) genes located in the *rfb* cluster region. To identify the H1 and H2 antigens, the genes *fliC* and *fljB* are analyzed. To achieve the identification, SISTR uses a database based on a *Salmonella* Genoserotyping Array (SGSA) developed earlier by the authors (17,

18). As WGS data can be incorrect due to sequencing errors, SISTR integrates a second step for the serovar prediction. In the first step, the antigenic formula is derived from the four genes representing the antigens which are also considered in the White–Kauffmann–Le Minor scheme. When this antigenic formula is not unique for one serovar due to incorrect or incomplete sequencing data, SISTR uses a specialized multilocus sequence typing scheme (330 loci; for details, see (14)) to identify the serovar in a second step. For this purpose, the phylogenetic clustering of the 330MLST loci is used to determine the single most likely serovar (14).

For SeqSero2, assembled genomes or raw reads can be used to identify *Salmonella* serovars. The tool has two variants. During the variant available for reads and assemblies, the algorithm maps k-mers generated from the input reads/assemblies to a database containing serovar determinants. The other variant is only applicable for read data. Here, all sequencing reads from a query genome were mapped to the serovar determinant database using BWA-MEM (19). Mapped reads are assembled into microassemblies which are then mapped with BLAST again to the database containing the serovar determinants. The mapping for both variants results in the antigenic formula from which the serovar can be inferred (15).

Recently, *in silico* serotyping has been validated with WGS datasets generated with Illumina devices (7, 13–16). Illumina sequencing data are characterized by short, but highly accurate reads. However, given the short reads, genome assembly based on Illumina is complicated, and assembled genomes are rarely completely contiguous. An alternative to this sequencing technique is nanopore sequencing developed by Oxford Nanopore Technologies (ONT). Here, ultra-long reads are generated. The extracted DNA or RNA passes through pores in a membrane. These pores measure the electric current of the electro-resistant membrane. Every nucleotide has a different impact on the current due to their specific nucleobases charge. These differences are afterward translated to nucleotides in a process called basecalling. However, reads from nanopore sequencing with R9.4.1 flow cells have lower accuracy scores compared to Illumina (20). Nevertheless, ONT is constantly improving the flow cells' performance and basecalling accuracy to achieve similar read accuracy as Illumina sequencing (21).

This study aimed to validate *in silico* serotyping based on WGS data generated with ONT. To this end, *Salmonella* strains of different serovars with epidemiological importance in humans, food, and farm animals (2) were selected to compare results from serotyping using agglutination tests and results from *in silico* serotyping-based sequence data from ONT with a R9.4.1 flow cell as well as Illumina. Moreover, the ability to detect genetic markers and plasmids with ONT and Illumina WGS data was compared.

2. Materials and methods

2.1. Materials

The dataset, comprised of 28 *Salmonella* strains of different serovars, was selected from the collection of bovine-derived *Salmonella* organisms at the National Reference Laboratory (NRL) for Salmonellosis in cattle in Germany. The NRL receives *Salmonella*

organisms isolated at cattle farms in Germany with suspected or confirmed outbreaks of salmonellosis from the investigation offices in the federal states for further characterization. The *Salmonella* organisms selected are of epidemiological relevance and present a wide range of serovars and include strains with special characteristics.

2.2. Serotyping

All *Salmonella* strains were serotyped using poly- and monovalent anti-O as well as anti-H sera (SIFIN, Germany) according to the White–Kauffmann–Le Minor scheme (6).

2.3. Sequencing and bioinformatics analysis

All 28 *Salmonella* strains were sequenced with an Illumina MiSeq and a GridION device (Figure 1). Long-read sequencing was performed using a GridION device from ONT. The genomic DNA of all 28 *Salmonella* strains was extracted with the QIAGEN® Genomic-tip 20/G Kit (QIAGEN, Germany) and the Genomic DNA Buffer Set (QIAGEN, Germany). Barcoding was done using the Ligation Sequencing Kit (SQK-LSK109) and the Native Barcoding Expansion Kit (EXP-NBD104). The sequencing was performed with a R9.4.1 flow cell (FLO-MIN106) according to the manufacturer's instructions.

For Illumina sequencing, the DNA was extracted and purified using the QIAGEN® Genomic-tip 20/G Kit (QIAGEN, Germany) and the Genomic DNA Buffer Set (QIAGEN, Germany). The concentration of the DNA was determined using the Qubit dsDNA BR assay kit (Invitrogen, United States). Sequencing libraries were created using the Nextera XT DNA Library Preparation Kit (Illumina Inc., United States). Paired-end sequencing with a length of 300 bps was performed on an Illumina MiSeq instrument according to the manufacturer's instructions (Illumina Inc., United States).

If no specific settings are mentioned, the tools utilized for the bioinformatic analysis were used in their respective default settings. Basecalling, trimming, and assembly of ONT data were performed with the pipeline mircoPIPE v 0.9 (22, 23). Guppy v 6.0.1 with the dna_r9.4.1_450bps_hac.cfg model was applied for basecalling and demultiplexing. The quality was checked with NanoPlot v 1.40.2 (24), adapter trimming was performed with Porechop v 0.2.3 (25), and filtering was performed with Japsa v 1.9-10a (26). The FASTQ files containing the trimmed long reads were afterwards assembled using Flye v 2.5 (27). In order to analyze the potential of ONT sequencing, assemblies were built with ONT data only, with ONT data combined with Illumina data and with Illumina data only. For long-read only assemblies (indicated with the ending _LR), the assemblies were only polished with the long reads itself using Racon v 1.4.9 (28) and Medaka v 0.10.0 (29), a polisher from ONT. For the assembly version combining ONT and Illumina data, these polished assemblies were further polished with NextPolish v 1.1.0 (30) and the corresponding short reads to receive the long-read assemblies with short-read polishing (indicated by the ending _LR_SR).

Illumina reads were assembled using the pipeline WGSBAC (v 2.2.0) (31). Here, the coverage was calculated, and the quality was accessed with FastQC (32). Quality and adapter trimming of Illumina reads, as well as, assembly was performed by Shovill v. 1.0.4 (33). These short-read only assemblies are indicated with the ending _SR throughout the study.

All different assembly versions (_LR, _LR_SR, and _SR) were further analyzed with the WGSBAC pipeline. Assembly quality was accessed with QUAST v 5.0.2 (34). Genome annotation was performed with Prokka v 1.13.3 (35), and the assemblies were checked for contamination with Kraken 2 v 0.10.6 (36) and Kraken2DB. The detection of genetic markers for virulence factors was performed with the tool ABRicate v0.8.10 (37) and the Virulence Factor Database (VFDB, version 27 March 2021) (38). Furthermore, AMRFinderPlus (39) was used to detect genetic markers for AMR (genes and mutations) and virulence. For *Salmonella* pathogenicity islands (SPIs), ABRicate together with a previously defined database (FLI, version 14 September 2020) was applied (40). Platon v 1.5.0 was used to detect plasmid-borne contigs (41). *In silico* serotyping was performed with two tools, namely, SISTR (14) and SeqSero2 (15). SeqSero2 was applied with the k-mers assembly-based mode.
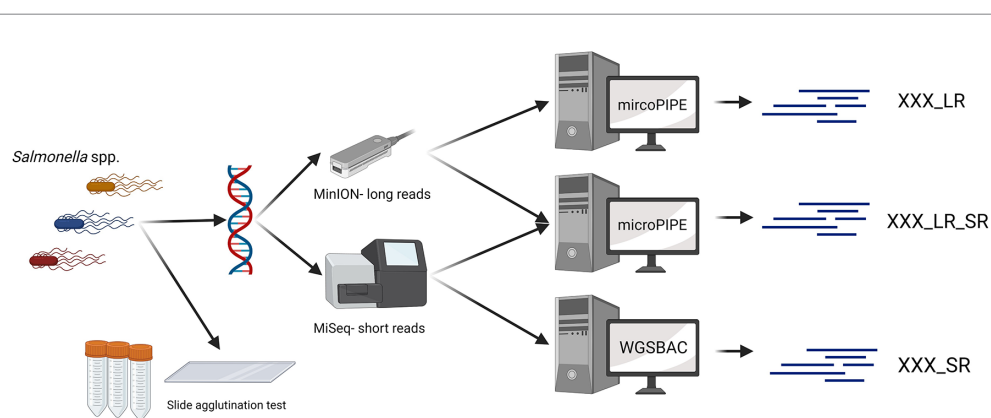

FIGURE 1
Data generation workflow. For 28 *Salmonella* strains, the serovar was determined with slide agglutination tests. Additionally, all *Salmonella* strains were sequenced with Oxford Nanopore Technologies and Illumina devices. Three different types of assemblies were built (_SR, _LR, _LR_SR). Graphic created with BioRender.com.

3. Results

3.1. Serotyping

A total of 28 *Salmonella* strains of different serovars were used to test *in silico* serotyping with long-read assemblies compared to slide agglutination tests. Furthermore, short-read-based assemblies were included for comparison. Therefore, the strains were serotyped by slide agglutination (Figure 1 and Table 1). Antigenic formulae of the serovars included in this study are shown in Supplementary Table S2. For three strains, no specific serovar could be determined with the agglutination test. First, no O antigen was detected with the agglutination test for strain 17 PM0072 (formula: O --; H g, p; --). This presumably rough strain resulted in no O antigens and H antigens g

and p. For strain 17 PM0296 (formula: O 4, 5; H i; --), the results of the agglutination test indicate a monophasic serovar. The third strain 16PM0121 (formula: O 9, 12; H l, v; --) led to erroneous results presumably due to errors in the synthesis of H2 flagellum.

3.2. Sequencing quality

DNA from 28 *Salmonella* strains was extracted and genome sequencing was performed with ONT and Illumina. The quality values for each strain and sequencing technique can be found in Supplementary Table S1. Illumina sequencing resulted in an average coverage of 108.86, with a minimum coverage of 30.62 and a maximum coverage of 208.15. ONT sequencing achieved an average

TABLE 1 Serotyping of 28 different *Salmonella* strains. Comparison of SISTR and SeqSero2 results and long-read only assemblies (_LR) with slide agglutination test results.

Strain	Agglutination test serovar	*in silico* serovar	
		SISTR	SeqSero2
16PM0089	Hadar	✓	✓
16PM0105	Gallinarum	✓	Gallinarum or Enteritidis[a]
16PM0121	O 9, 12 H l,v; -- s[b]	O --; H l,v; --[b]	Goettingen
16PM0122	Abony	✓	4:b:-[b]
16PM0124	III diarizonae	✓	✓
16PM0161	Tennessee	✓	✓
16PM0256	Goldcoast	✓	Goldcoast or Brikama[a]
16PM0296	Anatum	✓	✓
17PM0009	Enteritidis	✓	✓
17PM0296	O 4, 5 H i; --[b]	Typhimurium monophasic	Typhimurium monophasic
18PM0031	Choleraesuis	✓	Paratyphi C or Choleraesuis or Typhisuis[a]
18PM0045	Derby	✓	✓
18PM0092	Paratyphi B	✓[c]	✓
17PM0051	Kentucky	✓	✓
17PM0072	O -- H g, p[b]	Dublin	Dublin
18PM0004	Dublin	✓	✓
18PM0007	Coeln	✓	✓
17PM0282	Typhimurium	✓	✓
18PM0135	Kottbus	✓	✓
16PM0171	Livingstone	✓	✓
16PM0176	Stourbridge	✓	✓
17PM0054	Muenster	✓	✓
17PM0271	Indiana	✓	✓
16PM0056	Bovismorbificans	✓	✓
17PM0024	Mbandaka	✓	✓
17PM0167	Meleagridis	✓	✓
17PM0053	Agona	✓	✓
19PM0148	Infantis	✓	✓

[a]Correct serovar part of result.
[b]No serovar predicted.
[c]SISTR output: Paratyphi B var. Java.

coverage of 154.32, ranging from 45.21 to 500.38. The GC content ranged from 51.36% to 52.29%, which is close to the average GC content for the *Salmonella enterica* reference genome (NCBI RefSeq reference genome GCF_000006945.2) with 52% (42). The average read length for the ONT sequencing data was 5408.05, ranging from 2038.6 to 11337.8. For ONT data, the assemblies consist of one to five contigs, whereas the short-read only assemblies (_SR) consist of 26 to 212 contigs with an average of 53.36. Between 68 and 79% of the bases from Illumina sequencing reached Q30 (99.9% accuracy for basecalling). Nearly 6.5% to 15.1% of the ONT sequencing bases reached Q15 quality (96.8% accuracy for basecalling). At least 77.2% (short reads) and 88.96% (long reads) of the raw sequencing reads were classified as belonging to the genus *Salmonella*. The fraction of the reference genome of the short-read assemblies ranged from 67.6 to 98.7% and from 67.9 to 98.1% for the long-read assemblies (Supplementary Table S1).

3.3. *In silico* serotyping

Salmonella strains of different serovars serotyped by slide agglutination, listed in Table 1, were compared to long-read assemblies with two *in silico* serotyping tools. The results of *in silico* serotyping with short-read-based assemblies are included in Supplementary Table S2.

For ONT data (_LR), SISTR predicted 25 serovars completely matching the results of slide agglutination tests (Table 1). In addition, strain 17PM0072 (agglutination test formula: O --; H g, p; --) was assigned to the serovar *S*. Dublin for ONT data (_LR) and also for short-read-based assemblies (Supplementary Table S2). Strain 17PM0296 was identified as a monophasic variant of *S*. Typhimurium by SISTR. Again, the results of ONT are in accordance with Illumina regarding this strain (Supplementary Table S2). Strain 16PM0121 (through agglutination test O 9, 12 H l, v; --) was not assigned to a unique serovar by SISTR, neither with ONT nor with Illumina data. SISTR reported that only 279 belonging to the MLST330 for the long-read assembly (285 for _LR/_LR_SR) were found (which is below SISTR's internal quality cutoff of 297). The H1 antigen found by SISTR is not unique for a serovar.

For 3 out of 28 strains, SISTR produced a warning (Supplementary Table S2) when using ONT data (_LR). Warnings were raised for assemblies based on Illumina data (_SR) or the combination of both sequencing technologies (_LR_SR) for two and four strains. Most warnings concerned the amount of cgMLST330 loci identified in the assemblies.

SeqSero2 predicted 21 serovars matching the agglutination test for the long-read only assemblies (Table 1). For the presumably rough strain (agglutination test formula: O --; H g, p; --), SeqSero2 assigned

the serovar *S*. Dublin, coinciding with the SISTR results, independent of the sequencing technology (Supplementary Table S2). Similar to SISTR, SeqSero2 identified a monophasic variant of *S*. Typhimurium for strain 17PM0296, independent of the sequencing technology. Three strains resulted in multiple possible serovars when using SeqSero2 (for *S*. Gallinarum, *S*. Goldcoast, and *S*. Cholerasuis). Here, the antigenic formula found by SeqSero2 was not unique for one serovar. In cases where SeqSero2 reported multiple serovars, the serovar identified with the agglutination test was always included in the set of predicted serovars. A message indicated that the serovar should be differentiated with additional tests. Serovar *S*. Abony was not correctly predicted with the long-read assembly. The short-read assemblies resulted in similar false predictions (Supplementary Table S2). For strain 16PM0121, SeqSero2 predicted the serovar *S*. Goettingen, also for the short-read assemblies (Supplementary Table S2).

Overall, *in silico* serovar prediction based on ONT data was highly accurate when compared to the results of traditional slide agglutination. In fact, the tool SISTR achieved an accuracy of 96.4% (Table 2). For the tool SeqSero2, an accuracy of 92% was achieved, not taking results with multiple predicted serovars into account. The non-unique results of SeqSero2 imply further analysis steps, but the set of possible serovars remain significantly smaller with the correct serovar being part of the set. Since all tools independent of the sequencing technology assigned the presumably rough strain as *S*. Dublin, this was considered a correct result. Finally, *in silico* serovar prediction based on ONT data was comparable to results based on Illumina sequencing (Tables 1 and Supplementary Table S2).

3.4. Detection of genetic markers

3.4.1. Virulence factors

To compare the applicability of ONT data to detect genetic markers with the results based on Illumina data, genetic markers for virulence were determined for all assembly types based on short reads only (_SR), long reads only (_LR), and long-read assemblies polished with short reads (_LR_SR). Virulence factors were detected using ABRicate and the Virulence Factor Database (Supplementary Table S3) as well as using AMRFinderPlus (with option --plus). ABRicate reported the same set of virulence factors for 23 of 28 samples (82.1%) when comparing the different assembly types and sequencing technologies. For three strains (*S*. Bovismorbificans, *S*. Meleagridis, and *S*. Typhimurium monophasic), the same set of virulence factors were found by all technologies, but for long-read assembly (_LR and _LR_SR) assembly versions one additional factor. The *S*. Infantis assemblies led to the identification of 116 virulence factors for all assembly types. However, assemblies based on short reads (_SR and _LR_SR) resulted in one additional virulence factors (*entB*). The amount of identical virulence

TABLE 2 Overall results for SISTR and SeqSero2 for long-read only assemblies (_LR), short-read only assemblies (_SR), and long-read assemblies polished with short reads (_LR_SR).

Serotyping tool	Long-read assembly	Short-read assembly	Long-read assembly polished with short reads
SISTR	96.4% (26 + 1/28)[a]	92.8% (25 + 1/28)[a]	92.8% (25 + 1/28)[a]
SeqSero2	92% (22 + 1/25)[a,b]	88% (21 + 1/25)[a,b]	92% (22 + 1/25)[a,b]

[a]+1 is the rough mutant form determined as *S. Dublin*.
[b]The three results being not unique were excluded from the accuracy calculation.

factors was for *S.* Enteritidis 111 (up to six different) (Supplementary Table S3). AMRFinderPlus found genes *iroB* and *iroC* for 27 of 28 strains. *S. enterica* subsp. *diarizonae* (serovar O61; k,1,5,7 (*S.* Ill diarizonae)) was the only strain missing genes *iroB* and *iroC*. For *S.* Typhimurium, *S.* Derby, and *S.* Infantis, the long-read only assembly (_LR) included one gene less than the other two assembly types (missing *iroC, iroC,* and *iroB,* respectively) (Supplementary Table S3).

3.4.2. *Salmonella* pathogenicity islands

Next, the ability of ONT data to detect SPIs was compared to the results from Illumina data (Supplementary Table S4). For 16 strains (57%), the same sets of SPIs were detected by comparing both sequencing technologies (Supplementary Table S4). SPI 1 and SPI 9 were detected in all strains examined with all sequencing technologies. In all strains, except *S.* Abony and *S.* Ill diarizonae, SPI 2 was detected independent of the sequencing technology. The largest number of SPIs was found in *S.* Dublin (rough), *S.* Dublin, and *S.* Typhimurium monophasic (11 for short-read only assemblies and 12 for long-read assemblies). The *S.* Dublin (rough) and *S.* Dublin serovar harbored CS 54, SPI 1, SPI 2, SPI 4, SPI 5, SPI 9, SPI 12–SPI 17, and SPI19. The monophasic variant of *S.* Typhimurium did not contain SPI 17 and SPI 19 but SPI 3 and SPI 6 instead. The lowest number of SPIs was found in *S.* Ill diarizonae with the detection of SPI 1, SPI 18, and SPI 9. The assemblies based on the long-read data (_LR and _LR_SR) led to identical results for the detection of SPIs for all strains. In 12 of 28 cases (43%), the short-read only assembly (_SR) resulted in less SPIs than the long-read-based assemblies. In general, the genomic island CS54 as well as SPI 1–SPI 9 and SPI 11–SPI 19 were found in the WGS data of all *Salmonella* strains.

3.4.3. Plasmids

To compare the ability to detect and assemble plasmids, plasmid-borne contigs were determined with the tool Platon (Tables 3 and Supplementary Table S5). For the short-read only assemblies (_SR), the

TABLE 3 The number of plasmids found in 28 *Salmonella* strains. Comparison of Platon results with long-read only assemblies (_LR), short-read only assemblies (_SR), and long-read assemblies polished with short reads afterward (_LR_SR).

Strain	Serovar	Amount of plasmid-borne contigs		
		_SR	_LR	_LR_SR
16PM0089	Hadar	7	0	0
16PM0105	Gallinarum	4	1	1
16PM0121	–	4	2	2
16PM0122	Abony	4	1	1
16PM0124	III diarizonae	1	1	1
16PM0161	Tennessee	1	1	1
16PM0256	Goldcoast	0	0	0
16PM0296	Anatum	0	0	0
17PM0009	Enteritidis	2	1	1
17PM0296	Typhimurium monophasic	5	1	1
18PM0031	Choleraesuis	1	1	1
18PM0045	Derby	0	0	0
18PM0092	Paratyphi B	9	2	2
17PM0051	Kentucky	0	0	0
17PM0072	Dublin (rough)	1	1	1
18PM0004	Dublin	1	1	1
18PM0007	Coeln	0	0	0
17PM0282	Typhimurium	3	1	1
18PM0135	Kottbus	0	0	0
16PM0171	Livingstone	0	0	0
16PM0176	Stourbridge	0	0	0
17PM0054	Muenster	8	4	4
17PM0271	Indiana	0	0	0
16PM0056	Bovismorbificans	1	1	1
17PM0024	Mbandaka	1	1	1
17PM0167	Meleagridis	2	1	1
17PM0053	Agona	1	0	0
19PM0148	Infantis	3	1	1

TABLE 4 AMR genes found in 28 *Salmonella* strains. Comparison of AMRFinderPlus results with long-read only assemblies (_LR), short-read only assemblies (_SR), and long-read assemblies polished with short reads afterward (_LR_SR).

Strain	Serovar	Antimicrobial resistances[a]		
		_SR	_LR	_LR_SR
16PM0089	Hadar	AG/ST, TET, QUI	AG/ST, TET	AG/ST, TET
16PM0105	Gallinarum	–	–	–
16PM0121	–	AG/ST, SUL, BLA, TRM, QUI	AG/ST, SUL, BLA, TRM	AG/ST, SUL, BLA, TRM
16PM0122	Abony	–	–	–
16PM0124	III diarizonae	–	–	–
16PM0161	Tennessee	–	FOS	–
16PM0256	Goldcoast	–	–	–
16PM0296	Anatum	–	–	–
17PM0009	Enteritidis	–	QUI	–
17PM0296	Typhimurium monophasic	TET, SUL, AG/ST, BLA	TET, SUL, AG/ST, BLA	TET, SUL, AG/ST, BLA
18PM0031	Choleraesuis	–	–	–
18PM0045	Derby	FOS	FOS	FOS
18PM0092	Paratyphi B	AG/ST, BLA, TET, QUI/TCL, STR, TRM, PHE/CHL, SUL	AG/ST, BLA, TET, QUI/TCL, STR, TRM, PHE/CHL, SUL	AG/ST, BLA, TET, QUI/TCL, STR, TRM, PHE/CHL, SUL
17PM0051	Kentucky	–	–	
17PM0072	Dublin (rough)	–	–	–
18PM0004	Dublin	–	–	–
18PM0007	Coeln	–	–	–
17PM0282	Typhimurium	SUL, BLA, TET, PHE/CHL/FFC, AG/ST	SUL, BLA, TET, PHE/CHL/FFC, AG/ST	SUL, BLA, TET, PHE/CHL/FFC, AG/ST
18PM0135	Kottbus	–	–	–
16PM0171	Livingstone	–	–	–
16PM0176	Stourbridge	–	–	–
17PM0054	Muenster	–	–	–
17 PM0271	Indiana	–	–	–
16PM0056	Bovismorbificans	–	–	–
17PM0024	Mbandaka	–	–	–
17 PM0167	Meleagridis	FOS	FOS	FOS
17PM0053	Agona	FOS	FOS	FOS
19PM0148	Infantis	SUL, TET, AG/ST, QUI	SUL, TET, AG/ST, QUI	SUL, TET, AG/ST, QUI

[a]AG, aminoglycoside; ST, streptomycin; BLA, β-lactam; SUL, sulfonamide; TET, tetracycline; PHE, phenicol; CHL, chloramphenicol; FFC, florfenicol; FOS, fosfomycin QUI, quinolone; TCL, triclosan; MAC, macrolide; TRM, trimethoprim; STR, streptothricin.

number of possible plasmid-borne contigs ranged from zero to nine and was larger than for long-read assemblies (zero to four). This is caused by the discontinuousness of assemblies based on short reads (Supplementary Table S1). The long-read-based assemblies (_LR and _LR_SR) led to the same number of plasmid-borne sequences for each serovar. Platon did not find any plasmid-borne contig considering all assembly types for eight *Salmonella* serovars. For additional two strains (*S.* Hadar and *S.* Agona), Platon only characterized contigs of the short-read only assemblies as possible plasmids (7 and 1, respectively).

3.4.4. Antimicrobial resistance

Finally, ONT sequencing and Illumina sequencing were compared regarding the detection of genetic markers for AMR. AMR genes were determined with AMRFinderPlus (Table 4 and

Supplementary Table S6). In general, genetic markers for AMR against aminoglycoside, streptomycin, β-lactam, sulfonamide, tetracycline, fosfomycin, streptothricin, trimethoprim, phenicol, chloramphenicol, and quinolone were found across all strains. Assemblies of *S.* Paratyphi B contained the largest number of genetic markers for AMR (resistances against eight antibiotics classes). These include AMR for aminoglycoside/streptomycin, β-lactam, tetracycline, quinolone/triclosan, streptothricin, trimethoprim, sulfonamide, and phenicol/chloramphenicol. However, most strains (57.1%) did not result in any hit for an AMR gene with AMRFinderPlus. The different assembly types based on different sequencing technologies yielded the same sets AMR for 25 strain (89.2%). For *S.* Hadar and strain 16PM0121, in the short-read only assembly, additionally, gene *qnrB19* was found compared to

assemblies based on long reads. The long-read only assembly (_LR) of *S*. Enteritidis included additionally the *gyr_D87G* gene compared to the other assembly types. For *S*. Tennessee, the long-read only assembly included one gene more (*fosA7*).

4. Discussion

In this study, sequencing using ONT and Illumina of 28 *Salmonella* strains with serovars of epidemiological and zoonotic relevance was performed (2). Serotyping using two *in silico* tools based on respective genome assemblies was compared to the results from traditional slide agglutination tests. SISTR combined with long-read only assemblies achieved the highest accuracy at 96.4%, and SeqSero2 achieved 92% with the same assembly type. For short-read-based assemblies, SeqSero2 achieved accuracy scores of 92% and 88% accuracy for the _LR_SR version and the _SR version comparable to SISTR (both 92.8%). Considering additionally the three strains, SeqSero2 could not predict uniquely; in this study, SISTR and SeqSero2 did not achieve the same accuracy compared to Uelze et al. (13) when considering short-read assemblies (Supplementary Table S2). The predicted accuracy of 94% of SISTR combined with short-read assemblies as stated by Yoshida et al. (14) and Uelze et al. (13) was slightly improved in this study by the long-read only assemblies.

The slightly better performance of SISTR can be explained by the algorithm SISTR is based on. Both SeqSero2 and SISTR characterize the O and H antigens (14, 15) as an initial step. As the corresponding genes can include sequencing errors, SISTR is performing an additional step: here, a specialized MLST scheme consisting of 330 loci is used to phylogenetically cluster, and the serovar is derived from these clusters (14). Therefore, inconclusive results from the O and H antigen detection can be further identified as a unique serovar with the second step. SeqSero2 is only considering the information derived from the H and O antigens which may lead to non-unique results as seen in this study and previous studies (13, 16).

In this study, SeqSero2 was not able to correctly detect the O antigen of *S*. Abony. Uelze et al. stated that when using the Nextera XT DNA Library Preparation Kit, mapping-based tools such as SeqSero2 struggle with GC-biased sequencing data within the genetic locus of the O antigen, which might result in missing O antigen detection for SeqSero2 (13). Since the Nextera XT DNA Library Preparation Kit was used for sequencing, this result supports these findings. SISTR is not affected by this error source as it combines the identification of O and H antigens with a specialized MLST clustering approach. If the Nextera XT DNA Library Preparation Kit is chosen for sequencing, it is advisable to use SISTR for serotyping to avoid this error.

Strain 21PM0121 was not uniquely identifiable with the slide agglutination test due to the lack of detecting the H2 antigens resulting in the antigenic formula: O 9, 12 H l, v; −. SISTR was not able to identify a serovar. SeqSero2 reported the serovar *S*. Goettingen without any warnings. The H2 antigen reported by SISTR does correlate with the H2 antigen assigned to *S*. Goettingen (in combination with the H1 antigens). The results from the *in silico* serotyping tool indicate that the given strain could be the serovar *S*. Goettingen. This finding indicates an advantage for WGS data *in silico* serotyping compared to traditional agglutination tests as the complete antigenic formula can be determined.

Unlike the agglutination test method, the WGS approaches were able to assign a serovar for a rough strain of *S*. Dublin. Moreover, the two strains of serovar *S*. Dublin resulted in the exact same predictions for AMR and virulence genes as well as SPIs and the number of plasmids, supporting this characterization. Hence, the *in silico* serotyping tool with WGS data has an advantage to the agglutination test method, when considering rough strains of *Salmonella*. Moreover, *S*. Typhimurium and the monophasic variant of *S*. Typhimurium were identifiable and distinguishable with WGS data. This can be difficult with agglutination test methods as the antigens have to be expressed, which is not always the case in every cell condition leading to not decisive results (43). *In silico* serotyping tools are able to report the antigenic formula whereas the agglutination test can be prone to certain error sources (rough form, errors in synthesis, and monophasic variants).

In addition to serotyping, the WGS approach opens the opportunity to perform further bioinformatics analysis to study *Salmonella* strains in detail (44). Especially the detection of genetic markers for virulence and AMR as well as plasmid characterization and SPI detection are important steps in outbreak analysis.

In 23 of 28 (82.1%) *Salmonella* strains, the same virulence factors were determined for the Virulence Factor Database (VFDB) for all assembly types. For three isolates, the long-read assemblies included more factors, which indicates the potential of long-read assemblies for virulence factor prediction. Genes *iroB* and *iroC* were found in 27 of 28 strains, and they belong to the *iro* gene cluster (45). This cluster encodes for a mechanism to evade Lcn2-mediated host defense by mammals, which is blocking the iron uptake of the pathogens (46). The gene cluster is found in various *Salmonella* spp. and helps the organism to undergo this defense mechanism. Strain 21PM0124 is the only serovar of subspecies *S. enterica* subsp. *diarizonae* included in this study. All other strains are serovars belonging to the subspecies *S. enterica* subsp. *enterica*. The specific serovar IIIb 61:k:1,5, (7) is associated with sheep (47, 48). The absence of genes *iroB* and *iroC* in *S. enterica* subsp. *diarizonae* (serovar O61; k,1,5,7) confirms the findings of Uelze et al. (48). The authors indicate that this missing gene cluster could be induced by the adaption to the large intestine of sheep. Here, the pathogen is exposed to a high level of iron and thus does not need the acquisition of iron by salmochelin siderophore (48).

All *Salmonella* genomes encoded SPI 1 and SPI 9 which were identified in various serovars of *S. enterica* and *S. bongori* (49). SPI 1 is crucial for *Salmonella* to invade the epithelial cells of the hosts (50). Apart from genes expressed to help to dock to the host cell, expressed genes belonging to SPI also suppress early proinflammatory cytokine expression of the host's immune system (50). SPI 9 plays a role in adherence to eukaryotic cells (49). SPI 2 is important for the replication process and the systemic infections within macrophages (51, 52). 21PM0124, where SPI 2 was not detected, is of serovar O61; k,1,5,7. However, genes belonging to SPI 2 (*SlrP, SseF, SseG, SteC, SseJ*, and *SopD2*) were detected within 21PM0124. Opposite to this, other genes belonging to SPI 2 were not detected in the WGS data of strain 21PM0124, e.g., *SifA, PipB2, SteA, SseK1-2, SseL, SspH1-2, SpvB, SpvC*, and *SseL*. This indicates that fractions of SPI 2 are present in sheep-associated serovar but not the complete SPI 2. In general, the observation that short-read only assemblies resulted in lower detection

of SPIs (43%) compared to long-read-based assemblies supports the utilization of long-read technologies in the analysis of pathogens. As the detection of genetic markers is performed via sequence alignment to reference databases, the more discontiguous assemblies generated by short reads can lead to a smaller number of hits for those markers compared to assemblies based on ONT data.

For plasmid characterization, differences between data from the two sequencing methods were observed. The applied tool Platon classifies every contig as possible plasmid-borne or chromosomal. The higher amount of plasmid-borne contigs for short-read assemblies is in line with other studies (41, 53). Generally, Illumina-based assemblies contain a higher amount of contigs as the assembly process is more complex due to their short read length. Therefore, plasmids might be separated into multiple contigs using short-read-based assemblies. Here, long-read assemblies have a clear advantage as these reads allow the assembly of the entire plasmid. This explains why the short-read only assemblies in this study contain a larger number of plasmid-borne contigs compared to the long-read-based assemblies. For two strains, *S.* Hadar and *S.* Agona, the long-read-based assemblies (_LR, _LR_SR) resulted in zero plasmid-borne contigs. However, for the short-read only assemblies of those two strains, Platon found seven and one contigs as plasmid-borne contigs. For *S.* Hadar, six of the seven contigs were circular. The plasmid-borne contigs contained Inc. factors, oriT, replication, AMR, and mobilization genes. The only element found on the contig of *S.* Agona was a mobilization gene. As the long-read-based assemblies of these strains consisted only of one contig, no plasmid-born contigs were found. Platon assigns every contig with a size larger than 500 kb as a chromosome (41); therefore, assemblies consisting only of one contig automatically result in zero plasmid-borne contigs.

In this study, only subtle differences between the assembly types and sequencing technologies were observed regarding genetic markers for AMR. One example is missing genes in the long-read only assemblies (*S.* Hadar and O 9, 12 H l,v; −-). Here, the coverage of the gene found in the other two assemblies was identical, indicating that the gene was only part of the short-read data. For *S.* Tennessee, gene *fosA7* […] was only found in the _LR assembly, indicating that it was not covered by the short reads. The frequent appearance of aminoglycoside, tetracycline, sulfonamide, and β-lactam resistance genes was also approved by other studies (54, 55). In addition, the presence of AMR genes of phenicol, chloramphenicol, trimethoprim, quinolone, florfenicol, fosfomycin, and triclonsan was found, in line with other studies (56–62).

For serology-based serotyping, the gold-standard method of the past (6), no expensive equipment is needed (13). However, in theory, all different antisera have to be available in the laboratories for over 2600 identified serovars and must pass strict quality controls to prevent false-positives (13). In addition, the process can be time-consuming and labor-intensive as strains have to express all possible H antigens, involving several analysis steps (7, 13). WGS data not only allow the characterization of O and H antigens but additionally have the utility for the detection of genetic markers for AMR, virulence, and plasmids (63, 64). One advantage of ONT sequencing compared to Illumina sequencing is the generation of ultra-long reads. The length of the reads opens the possibility to create closed assemblies, whereas the short reads of Illumina may lead to discontiguous

assemblies (65). Therefore, ONT data allow the assembly of plasmids to closure (66). This can be helpful when analyzing AMR genes as their loci can be studied in more detail. Whether AMR genes are located on plasmids or chromosomes can have an impact on their ability to distribute and is, therefore, an important factor in surveillance (67, 68). Additionally, Illumina sequencing requires relatively high acquisition costs and high costs for sequencing low sample numbers (69). This cost difference can be crucial, especially for small laboratories with lower budgets and small sample numbers. This study shows that *in silico* serotyping and the detection of genetic markers are possible with data based on R9.4.1 flow cells, even though sequencing accuracy is comparably lower (70). Considering that ONT has recently introduced R10.4.1 flow cells with similar sequencing accuracy to Illumina sequencing for DNA (21, 71), ONT sequencing could be an alternative sequencing technique in bacterial genomics. While *in silico* serotyping based on Illumina WGS data is applied on a routine basis in some laboratories (72), ONT sequencing may become more integrated in public health laboratories. This study, in line with other studies in bacterial genomics, shows the potential of ONT sequencing for routine laboratories (73–76). Taking into account the hybrid approach of combining short reads and long reads, laboratories must have both technical devices and trained staff for both sequencing techniques leading to higher costs and more demand on staff. Since the accuracy with ONT alone or Illumina alone is sufficient enough for *in silico* serotyping and the detection of genetic markers, it is advisable to rely on one sequencing technique alone.

In this study, the preciseness of *in silico* serotyping with ONT-generated data was comparable to traditional serotyping. The ongoing development of ONT flow cells, which leads to improved sequencing accuracy, may enable the construction of bacterial genomes with long reads only (21, 71). Moreover, nanopore sequencing has advantages in plasmid characterization due to its simpler assembly process when compared to Illumina sequencing. Additionally, analysis results concerning virulence factors, AMR genes, and SPI detection based on ONT data are comparable to the results from Illumina WGS data. Thus, the results of this study indicate the applicability of ONT sequencing in *Salmonella in silico* serotyping, its ability to lead to more comprehensive surveillance of *Salmonella* outbreaks, and hence its possible use in routine laboratories.

Data availability statement

The data presented in the study are deposited in the European Nucleotide Archive (ENA) repository, accession number PRJEB59466. Data has been released already and is accessible via https://www.ebi.ac.uk/ena/browser/view/PRJEB59466.

Author contributions

JL and UM conceived and coordinated the study. CT performed bioinformatics analysis and drafted and wrote the manuscript. UM, JL, and MM critically read the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1178922/full#supplementary-material

SUPPLEMENTARY TABLE S1
Sequencing quality.

SUPPLEMENTARY TABLE S2
Serotyping.

SUPPLEMENTARY TABLE S3
Virulence factors.

SUPPLEMENTARY TABLE S4
Salmonella pathogenicity islands.

SUPPLEMENTARY TABLE S5
Plasmid.

SUPPLEMENTARY TABLE S6
Genetic markers for antimicrobial resistances.

References

1. CDC. *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA: U.S. Department of Health and Human Services (2019).

2. European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC). European food safety authority and European Centre for Disease Prevention and Control the European Union one health 2021 Zoonoses report. *EFSA J.* (2022) 19:e07666. doi: 10.2903/j.efsa.2021.6971

3. Fierer J, Guiney DG. Diverse virulence traits underlying different clinical outcomes of Salmonella infection. *J Clin Invest.* (2001) 107:775–80. doi: 10.1172/JCI12561

4. Olaimat AN, Holley RA. Factors influencing the microbial safety of fresh produce: a review. *Food Microbiol.* (2012) 32:1–19. doi: 10.1016/j.fm.2012.04.016

5. Microbiology SSotNCotISf. The genus salmonella lignieres, 1900. *J Hyg.* (1934) 34:333–50. doi: 10.1017/S0022172400034677

6. Grimont P, Weill FX. Antigenic Formulae of the Salmonella Serovars. *WHO Collaborating Centre for Reference and Research on Salmonella*, vol. 9. Geneva: WHO (2007).

7. Ibrahim GM, Morin PM. Salmonella serotyping using whole genome sequencing. *Front Microbiol.* (2018) 9:2993. doi: 10.3389/fmicb.2018.02993

8. Lindberg AA, Hellerqvist C. Rough mutants of *Salmonella typhimurium*: immunochemical and structural analysis of lipopolysaccharides from rfaH mutants. *Microbiology.* (1980) 116:25–32. doi: 10.1099/00221287-116-1-25

9. den Bakker HC, Moreno Switt AI, Cummings CA, Hoelzer K, Degoricija L, Rodriguez-Rivera LD, et al. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. enterica serovar Montevideo pulsed-field gel electrophoresis type. *Appl Environ Microbiol.* (2011) 77:8648–55. doi: 10.1128/AEM.06538-11

10. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS One.* (2014) 9:e87991. doi: 10.1371/journal.pone.0087991

11. Bale J, Meunier D, Weill F-X, DePinna E, Peters T, Nair S. Characterization of new Salmonella serovars by whole-genome sequencing and traditional typing techniques. *J Med Microbiol.* (2016) 65:1074–8. doi: 10.1099/jmm.0.000325

12. Bekal S, Berry C, Reimer AR, Van Domselaar G, Beaudry G, Fournier E, et al. Usefulness of high-quality core genome single-nucleotide variant analysis for subtyping the highly clonal and the most prevalent *Salmonella enterica* serovar Heidelberg clone in the context of outbreak investigations. *J Clin Microbiol.* (2016) 54:289–95. doi: 10.1128/JCM.02200-15

13. Uelze L, Borowiak M, Deneke C, Szabó I, Fischer J, Tausch SH, et al. Performance and accuracy of four open-source tools for in silico serotyping of Salmonella spp. based on whole-genome short-read sequencing data. *Appl Environ Microbiol.* (2020) 86:e02265–19. doi: 10.1128/AEM.02265-19

14. Yoshida C. Kruczkiewicz, Peter, Laing, CR, Lingohr, EJ, Gannon, VPJ, Nash, JHE, Taboada, EN the Salmonella in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft Salmonella genome assemblies. *PLoS One.* (2016) 11:e0147101. doi: 10.1371/journal.pone.0147101

15. Zhang S, den Bakker HC, Li S, Chen J, Dinsmore BA, Lane C, et al. SeqSero2: rapid and improved Salmonella serotype determination using whole-genome sequencing data. *Appl Environ Microbiol.* (2019) 85:e01746–19. doi: 10.1128/AEM.01746-19

16. Hendriksen RS, Pedersen SK, Leekitcharoenphon P, Malorny B, Borowiak M, Battisti A, et al. Final report of ENGAGE—establishing next generation sequencing ability for genomic analysis in Europe. *EFSA Support Publ.* (2018) 15:1431E. doi: 10.2903/sp.efsa.2018.EN-1431

17. Franklin K, Lingohr EJ, Yoshida C, Anjum M, Bodrossy L, Clark CG, et al. Rapid genoserotyping tool for classification of Salmonella serovars. *J Clin Microbiol.* (2011) 49:2954–65. doi: 10.1128/JCM.02347-10

18. Yoshida C, Lingohr EJ, Trognitz F, MacLaren N, Rosano A, Murphy SA, et al. Multi-laboratory evaluation of the rapid genoserotyping array (SGSA) for the identification of Salmonella serovars. *Diagn Microbiol Infect Dis.* (2014) 80:185–90. doi: 10.1016/j.diagmicrobio.2014.08.006

19. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* (2013). doi: 10.48550/arXiv.1303.3997

20. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol.* (2021) 82:801–11. doi: 10.1016/j.humimm.2021.02.012

21. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sörensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods.* (2022) 19:823–6. doi: 10.1038/s41592-022-01539-7

22. Murigneux V. *microPIPE: A Pipeline for High-Quality Bacterial Genome Construction Using ONT and Illumina Sequencing.* (2021). https://github.com/BeatsonLab-MicrobialGenomics/micropipe

23. Murigneux V, Roberts LW, Forde BM, Phan M-D, Nhu NTK, Irwin AD, et al. MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction. *BMC Genomics.* (2021) 22:474. doi: 10.1186/s12864-021-07767-z

24. de Coster W, D'Hert S, Schultz DT, Cruts M, van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* (2018) 34:2666–9. doi: 10.1093/bioinformatics/bty149

25. Wick R. *Porechop.* (2017). https://github.com/rrwick/Porechop

26. Nguyen SH. *Japsa - Just Another Java Paackage for Sequence Analysis.* (2019). https://github.com/mdcao/japsa

27. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long error-prone reads using repeat graphs. *Nat Biotechnol.* (2019) 37:540–6. doi: 10.1038/s41587-019-0072-8

28. Sovic I. *Racon - Consensus Module for Raw de novo DNA Assembly of Long Uncorrected Reads.* (2019). https://github.com/isovic/racon

29. Oxford Nanopore Technologies. *Medaka.* (2019). https://github.com/nanoporetech/medaka

30. Jiang Hu JF, Zongyi Sun SL. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics.* (2019) 36:2253–5. doi: 10.1093/bioinformatics/btz891

31. Linde J, Abdel-Glil M, García-Soto S. *WGSBAC: Modules for Genotyping and Characterisation of Bacterial Isolates Utilizing Whole-Genome-Sequencing Data*. (2019). https://gitlab.com/FLI_Bioinfo/WGSBAC

32. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Bioinformatics, Babraham Institute (2010).

33. Seemann T. *Shovill—Assemble Bacterial Isolate Genomes from Illumina Paired-End Reads*. (2020). https://github.com/tseemann/shovill

34. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. (2013) 29:1072–5. doi: 10.1093/bioinformatics/btt086

35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. (2014) 30:2068–9. doi: 10.1093/bioinformatics/btu153

36. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol*. (2019) 20:257–13. doi: 10.1186/s13059-019-1891-0

37. Seemann T. *ABRicate*. (2020). https://github.com/tseemann/abricate

38. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res*. (2016) 44:D694–7. doi: 10.1093/nar/gkv1239

39. Michael Feldgarden VB, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother*. (2019) 63:e0048319. doi: 10.1128/AAC.00483-19

40. García-Soto S, Abdel-Glil MY, Tomaso H, Linde J, Methner U. Emergence of multidrug-resistant *Salmonella enterica* subspecies enterica serovar infantis of multilocus sequence type 2283 in German broiler farms. *Front Microbiol*. (2020) 11:1741. doi: 10.3389/fmicb.2020.01741

41. Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A. Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. *Microb Genom*. (2020) 6:mgen000398. doi: 10.1099/mgen.0.000398

42. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. (2016) 44:D733–45. doi: 10.1093/nar/gkv1189

43. Barco L, Lettini AA, Ramon E, Longo A, Saccardin C, Pozza MAD, et al. A rapid and sensitive method to identify and differentiate *Salmonella enterica* serotype Typhimurium and *Salmonella enterica* serotype 4,[5], 12: i:-by combining traditional serotyping and multiplex polymerase chain reaction. *Foodborne Pathog Dis*. (2011) 8:741–3. doi: 10.1089/fpd.2010.0776

44. García-Soto S, Linde J, Methner U. Epidemiological analysis on the occurrence of *Salmonella enterica* subspecies enterica Serovar Dublin in the German Federal State Schleswig-Holstein Using Whole-Genome Sequencing. *Microorganisms*. (2023) 11:1. doi: 10.3390/microorganisms11010122

45. Campioni F, Vilela FP, Cao G, Kastanis G, dos Prazeres RD, Costa RG, et al. Whole genome sequencing analyses revealed that *Salmonella enterica* serovar Dublin strains from Brazil belonged to two predominant clades. *Sci Rep*. (2022) 12:10555. doi: 10.1038/s41598-022-14492-4

46. Fischbach MA, Lin H, Zhou L, Yu Y, Abergel RJ, Liu DR, et al. The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2. *Proc Natl Acad Sci*. (2006) 103:16502–7. doi: 10.1073/pnas.0604636103

47. Methner U, Moog U. Occurrence and characterisation of *Salmonella enterica* subspecies diarizonae serovar 61: k: 1, 5, (7) in sheep in the federal state of Thuringia, Germany. *BMC Vet Res*. (2018) 14:1–8. doi: 10.1186/s12917-018-1741-4

48. Uelze L, Borowiak M, Deneke C, Fischer J, Flieger A, Simon S, et al. Comparative genomics of *Salmonella enterica* subsp. diarizonae serovar 61:k:1,5,(7) reveals lineage-specific host adaptation of ST432. *Microb Genom*. (2021) 7:000604. doi: 10.1099/mgen.0.000604

49. Velásquez JC, Hidalgo AA, Villagra N, Santiviago CA, Mora GC, Fuentes JA. SPI-9 of *Salmonella enterica* serovar Typhi is constituted by an operon positively regulated by RpoS and contributes to adherence to epithelial cells in culture. *Microbiology*. (2016) 162:1367–78. doi: 10.1099/mic.0.000319

50. Raffatellu M, Wilson RP, Chessa D, Andrews-Polymenis H, Tran QT, Lawhon S, et al. SipA, SopA, SopB, SopD, and SopE2 contribute to *Salmonella enterica* serotype Typhimurium invasion of epithelial cells. *Infect Immun*. (2005) 73:146–54. doi: 10.1128/IAI.73.1.146-154.2005

51. Fass E, Groisman EA. Control of Salmonella pathogenicity island-2 gene expression. *Curr Opin Microbiol*. (2009) 12:199–204. doi: 10.1016/j.mib.2009.01.004

52. Jajere S. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World*. (2019) 12:504–21. doi: 10.14202/vetworld.2019.504-521

53. Juraschek K, Borowiak M, Tausch SH, Malorny B, Käsbohrer A, Otani S, et al. Outcome of different sequencing and assembly approaches on the detection of plasmids and localization of antimicrobial resistance genes in commensal *Escherichia coli*. *Microorganisms*. (2021) 9:598. doi: 10.3390/microorganisms9030598

54. Rodrigues GL, Panzenhagen P, Ferrari RG, Dos Santos A, Paschoalin VMS, Conte-Junior CA. Frequency of antimicrobial resistance genes in Salmonella from Brazil by in silico whole-genome sequencing analysis: an overview of the last four decades. *Front Microbiol*. (2020) 11:1864. doi: 10.3389/fmicb.2020.01864

55. Srednik ME, Lantz K, Hicks JA, Morningstar-Shaw BR, Mackie TA, Schlater LK. Antimicrobial resistance and genomic characterization of Salmonella Dublin isolates in cattle from the United States. *PLoS One*. (2021) 16:e0249617. doi: 10.1371/journal.pone.0249617

56. An R, Alshalchi S, Breimhurst P, Munoz-Aguayo J, Flores-Figueroa C, Vidovic S. Strong influence of livestock environments on the emergence and dissemination of distinct multidrug-resistant phenotypes among the population of non-typhoidal Salmonella. *PLoS One*. (2017) 12:e0179005. doi: 10.1371/journal.pone.0179005

57. Birošova L, Mikulašová M. Development of triclosan and antibiotic resistance in *Salmonella enterica* serovar Typhimurium. *J Med Microbiol*. (2009) 58:436–41. doi: 10.1099/jmm.0.003657-0

58. El-Tayeb MA, Ibrahim ASS, Al-Salamah AA, Almaary KS, Elbadawi YB. Prevalence, serotyping and antimicrobials resistance mechanism of *Salmonella enterica* isolated from clinical and environmental samples in Saudi Arabia. *Braz J Microbiol*. (2017) 48:499–508. doi: 10.1016/j.bjm.2016.09.021

59. Khatoon A, Malik HMT, Aurongzeb M, Raza SA, Karim A. Draft genome of a macrolide resistant XDR *Salmonella enterica* serovar Paratyphi a strain using a shotgun sequencing approach. *J Glob Antimicrob Resist*. (2019) 19:129–31. doi: 10.1016/j.jgar.2019.09.001

60. Lin D, Chen S. First detection of conjugative plasmid-borne fosfomycin resistance gene fosA3 in Salmonella isolates of food origin. *Antimicrob Agents Chemother*. (2015) 59:1381–3. doi: 10.1128/AAC.04750-14

61. Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, et al. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal *Salmonella enterica*. *Front Microbiol*. (2018) 9:592. doi: 10.3389/fmicb.2018.00592

62. Zhan Z, Xu X, Shen H, Gao Y, Zeng F, Qu X, et al. Rapid emergence of florfenicol-resistant invasive non-typhoidal salmonella in China: a potential threat to public health. *Am J Trop Med Hyg*. (2019) 101:1282–5. doi: 10.4269/ajtmh.19-0403

63. Gilchrist CA, Turner SD, Riley MF, Petri WA Jr, Hewlett EL. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev*. (2015) 28:541–63. doi: 10.1128/CMR.00075-13

64. Laing CR, Zhang Y, Thomas JE, Gannon VPJ. Everything at once: comparative analysis of the genomes of bacterial pathogens. *Vet Microbiol*. (2011) 153:13–26. doi: 10.1016/j.vetmic.2011.06.014

65. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. (2020) 38:701–7. doi: 10.1038/s41587-020-0422-6

66. Weber RE, Pietsch M, Fruhauf A, Pfeifer Y, Martin M, Luft D, et al. IS26-mediated transfer of Bla NDM-1 as the Main route of resistance transmission during a polyclonal, multispecies outbreak in a German hospital. *Front Microbiol*. (2019) 10:2817. doi: 10.3389/fmicb.2019.02817

67. Carattoli A. Plasmids and the spread of resistance. *Int J Med Microbiol*. (2013) 303:298–304. doi: 10.1016/j.ijmm.2013.02.001

68. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother*. (2018) 73:1121–37. doi: 10.1093/jac/dkx488

69. Wagner GE, Dabernig-Heinz J, Lipp M, Cabal A, Simantzik J, Kohl M, et al. Real-time Nanopore Q20+ sequencing enables extremely fast and accurate Core genome MLST typing and democratizes access to high-resolution bacterial pathogen surveillance. *J Clin Microbiol*. (2023) 61:e0163122. doi: 10.1128/jcm.01631-22

70. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One*. (2021) 16:e0257521. doi: 10.1371/journal.pone.0257521

71. Sanderson N, Kapel N, Rodger G, Webster H, Lipworth S, Street T, et al. Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flowcells and chemistries in bacterial genome reconstruction. *Microb Genom*. (2022) 9:mgen000910. doi: 10.1099/mgen.0.000910

72. Gand M, Mattheus W, Roosens N, Dierick K, Marchal K, Bertrand S, et al. A genoserotyping system for a fast and objective identification of Salmonella serotypes commonly isolated from poultry and pork food sectors in Belgium. *Food Microbiol*. (2020) 91:103534. doi: 10.1016/j.fm.2020.103534

73. Greig DR, Jenkins C, Gharbia S, Dallman TJ. Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. *Gigascience*. (2019) 8:giz104. doi: 10.1093/gigascience/giz104

74. Hall MB, Rabodoarivelo MS, Koch A, Dippenaar A, George S, Grobbelaar M, et al. Evaluation of Nanopore sequencing for *Mycobacterium tuberculosis* drug susceptibility testing and outbreak investigation: a genomic analysis. Lancet. *Microbe*. (2023) 4:e84–92. doi: 10.1016/S2666-5247(22)00301-9

75. Oude Munnink BB, Nieuwenhuijse D, Stein M, O'Toole Á, Haverkate M, Mollers M, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*. (2020) 26:1405–10. doi: 10.1038/s41591-020-0997-y

76. Smith C, Halse TA, Shea J, Modestil H, Fowler RC, Musser KA, et al. Assessing nanopore sequencing for clinical diagnostics: a comparison of next-generation sequencing (NGS) methods for *Mycobacterium tuberculosis*. *J Clin Microbiol*. (2020) 59:e00583–20. doi: 10.1128/JCM.00583-20