From the
Institute of Animal Hygiene and Veterinary Public Health
of the Faculty of Veterinary Medicine, Leipzig University

A Whole-Genome sequencing-based study of the emergent multidrug-resistant *Salmonella enterica* subspecies *enterica* serovar Infantis clones in German broiler farms

Inaugural-Dissertation
to obtain the degree of a
Doctor medicinae veterinariae (Dr. med. vet.)
from the Faculty of Veterinary Medicine
Leipzig University

Submitted by
Silvia García Soto
from Madrid

Leipzig, 2021

Mit der Genehmigung der Veterinärmedizinischen Fakultät der Universität Leipzig

Dekan: Herr Prof. Dr. Dr. Thomas Vahlenkamp

Betreuer: Herr Prof. Dr. Martin Pfeffer

Gutachter: Herr Prof. Dr. Martin Pfeffer, Institut für Tierhygiene und Öffentliches Veterinärwesen, Leipzig
Herr Dr. Torsten Semmler, Nachwuchsgruppe Mikrobielle Genomik, Robert-Koch-Institute, Berlin

Tag der Verteidigung: 14.09.2021

A mi Familia

“El camino es la meta”

„Der Weg ist das Ziel“

Proverb attributed to Kung Fu Tzu (551 AC -479 BC)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AMR	Antimicrobial-resistance
CARD	Comprehensive Antimicrobial Resistance Database
cgMLST	core-genome Multilocus Sequence Typing
ECDC	European Centre for Disease Prevention and Control
EFSA	European Food Safety Authority
ENA	European Nucleotide Archive
ESBL	Extended-spectrum beta-lactamase
ESI	Emergent *Salmonella* Infantis
ESIr	Emergent *Salmonella* Infantis resistance
ESIv	Emergent *Salmonella* Infantis virulence
EU	European Union
FBO	Food-Borne Outbreak
GIT	GNU Interactive Tools
GNU	GNU's Not Unix!
HTS	High-Throughput Sequencing
IBIZ	Institute of Bacterial Infections and Zoonoses
Ipf	Infantis plasmid-encoded fimbria
Klf	K88-like fimbria
KWS	Kauffmann-White Scheme
MCDA	Multiple Cross-Displacement Amplification
MDR	Multidrug-resistance
MGE	Mobile Genetic Elements
MIC	Minimum Inhibitory Concentration

MLST	Multilocus Sequence Typing
MLVA	Multiple-Locus Variable-number tandem repeat Analysis
MOL-PCR	Multiplex Oligonucleotide Ligation – Polymerase Chain Reaction
MOST	Metric oriented Sequender Typer
NGS	Next-Generation Sequencing
NTS	Non-typhoidal *Salmonella*
ONT	Oxford Nanopore Technologies
PCR	Polymerase Chain Reaction
pESI	plasmid of Emerging *Salmonella* Infantis
PFGE	Pulsed-Field Gel Electrophoresis
QAC	Quaternary Ammonium Compound
RKI	Robert Koch Institut
S.	*Salmonella*
SPI	*Salmonella* Pathogenicity Island
SOLiD	Sequencing by Oligo Ligation Detection
SRA	Sequence Read Archive
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
TGS	Third-Generation Sequencing
VFDB	Virulence Factor Database
WGS	Whole-Genome Sequencing
wgSNP	whole-genome Single Nucleotide Polymorphism

1. INTRODUCTION

Salmonella (*S.*) is a zoonotic pathogen which prevalence in the food chain supposes a considerable impact on public health. After campylobacteriosis (246,571 confirmed cases in 2018), salmonellosis remains the second most commonly reported gastrointestinal infection in humans in Europe (91,857 confirmed cases in 2018) (EFSA 2019a). Non-typhoidal salmonellosis occurs worldwide and is acquired through direct or indirect contact between humans and animals. Consumption of insufficiently cooked food that may carry *Salmonella enterica* (*S. enterica*) increases the risk of infection. The disease is mainly characterized by gastroenteritis, but its severity can vary from mild symptoms (self-limiting diarrhea) to invasive life-threatening extra-intestinal infection. *Salmonellosis* enteric disease is caused by a great variety of non-typhoidal *Salmonella* (NTS) serovars. From 2011 until 2016, the most-reported NTS serovars responsible for human cases were *S*. *enterica* serovar (*S.*) Enteritidis, *S*. Typhimurium (including monophasic *S*. Typhimurium), and *S*. Infantis (EFSA 2019a). Due to its relevance in public health, these serovars are considered as target serovars in the context of poultry production and National Control Programmes in poultry (e.g. Regulation (EC) No. 2160/2003) aimed at reducing its prevalence. Despite a decreasing trend in the prevalence of the target serovars observed during 2007-2018, *S*. Infantis became the most dominant serovar of all the serovars reported from broilers (36.5%) and broiler meat (56.7%) in 2018 (EFSA 2019a, b). The most frequent sources of contamination are broiler meat and derivate products but it is also found in pork and beef (RAJIC et al. 2005, LINDQVIST et al. 2007, EFSA 2019a).

In Germany, a total of 13,693 *Salmonella* human cases were reported in 2019 at the same level as previous years (13,592 in 2018) according to the Robert-Koch-Institut (RKI) (RKI 2019, 2020). The most-reported serovars were *S*. Enteritidis and *S*. Typhimurium of all reported salmonellosis infections in 2019 and, the consumption of contaminated raw eggs, pork, and derivate meat products were the most common sources (RKI 2020). However, compared with previous years, no changes in prevalence (less than 1%) of both serovars were observed (BfR 2020). In Germany, *S*. Infantis is mainly detected in broiler meat (37.6%), but also in pork (2.6%) (BfR 2020). In 2018 and unlike previous years, the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC) detected in Europe an increased level of multidrug-resistant (MDR) *S*. Infantis isolates recovered from broilers (60.3%) and broiler carcasses (74.2%) (EFSA 2020). The most frequently observed MDR pattern

consisted of resistance to ciprofloxacin, nalidixic acid, sulfamethoxazole, and tetracycline. However, variations of this MDR profile, including extended-spectrum beta-lactamase (ESBL)-producing and colistin-resistant *S*. Infantis were identified in broilers as well (FRANCO et al. 2015, CARFORA et al. 2018, EFSA 2020). Over the last 10 years, several European (EU) and non-EU countries revealed evidence of the rapid dissemination and clonal emergence of the MDR *S*. Infantis population (EFSA 2019a). The use of Whole-Genome-Sequencing (WGS) and bioinformatics tools in public-health laboratories' routine practice have enabled a prompt publication of studies on this topic. The most recent, include the report of the complete gap-free sequenced genome of the emergent "*S*. Infantis 119944" strain isolated in Israel, and comparative detailed genome analysis of the MDR *S*. Infantis clones and their plasmids emerged over the world (GYMOESE et al. 2019, ALBA et al. 2020, COHEN et al. 2020).

This Doctoral Thesis aimed to identify the potential genetic causes for the increased emergence of *S*. Infantis in German broiler production. For this purpose, first, we conducted a primer implementation, application, and evaluation of the performance of an in-house bioinformatics pipeline named WGSBAC for *Salmonella in silico* serotyping. Second, we apply the WGSBAC pipeline to characterize and compare *S*. Infantis strains collected during a 20-year distant period (from the 1990s to the 2010s) in different German broiler farms. Furthermore, we performed a genome comparison analysis between German *S*. Infantis and European *S*. Infantis genomes from public databases to study possible clonal relatedness. This study gives evidence of the occurrence and spread within two decades of an emergent MDR and virulent *S*. Infantis population of ST2283 in the German broiler production chain. The acquisition of a megaplasmid encoding resistance, virulence-associated determinants, and fitness mechanisms may explain this quick and worrying epidemiological event. The use of WGS and the application of a bioinformatics pipeline have been an effective approach in the production of accurate and reliable results for *Salmonella* serovar prediction and the characterization of *S*. Infantis German strains including the comparison with other *S*. Infantis EU and non-EU genomes.

2. LITERATURE OVERVIEW

2.1. The genus *Salmonella* and its epidemiological role as a foodborne pathogen

Salmonella is one of the most well-studied microorganisms and since its discovery in 1885 by the veterinarian surgeon Daniel Elmer Salmon and the pathologist Theobald Smith, it has taken the attention of many researchers (SALMON et al. 1886). The reason for this concern is its significant role as a foodborne pathogen. With approximately 153 million cases worldwide and 91,857 confirmed cases in Europe in 2018, it is the second most reported burden of foodborne disease after Campylobacter (246 571 cases in 2018) (EFSA 2019a).

The genus *Salmonella* is a gram-negative rod-shaped no spore-forming bacteria that is generally motile through multiple flagella, and that belongs to the family Enterobacteriaceae (GILL et al. 2018). The genus *Salmonella* consists of two species: *Salmonella bongori* and *Salmonella enterica*. *Salmonella enterica* (*S. enterica*) is further subdivided into six subspecies: *enterica* (I)*, salamae* (II)*, arizonae* (IIIa), *diarizonae* (IIIb)*, houtenae* (IV)*,* and *indica* (VI). *S. enterica* subspecies *enterica* accounts for approximately 99% of all clinical isolates from humans and warm-blooded animals (ACHTMAN et al. 2012). Depending on the disease syndrome, it is divided into two groups: Typhoidal *Salmonella* (invasive extra-intestinal infection characterized by high fever) and Non-Typhoidal *Salmonella* (mainly self-limiting gastroenteritis). Serological assays based on agglutination are used to make a further differentiation of *S. enterica* into serovars (GRIMONT et al. 2007, ACHTMAN et al. 2012). To date, more than 2,579 *Salmonella* serovars have been recognized and can be classified based on their host range into host-adapted, and non-host-adapted serovars (UZZAU et al. 2000). Host-adapted serovars that are associated almost exclusively with one host species (e.g. Paratyphi and Typhi for humans; Abortus equi for equine or Gallinarum for poultry) are referred as host-restricted serovars. Meanwhile, there are host-adapted serovars that are mainly associated with one species but may result in disease in others (e.g. Dublin for cattle but able to infect small ruminants, pigs, and humans or Choleraesuis for swine but can cause diseases in humans). The animals infected by these serovars manifest a carrier state (reservoir) as they excrete the pathogen without any clinical signs of infection (UZZAU et al. 2000). Meanwhile, although rare, the infection in humans deals with severe systemic disease and high mortality. On the other hand, non-adapted serovars such as *S.* Typhimurium and *S.* Enteritidis infect a wide range of hosts in which the clinical predominant manifestation is

associated with gastro-intestinal symptoms dealing with high morbidity but low mortality (UZZAU et al. 2000). *S*. Typhimurium and *S*. Enteritidis are included within the group of NTS serovars and cause non-typhoidal salmonellosis which clinical picture should not be confused with the invasive disease caused by serovar Typhi and serovars Paratyphi A, Paratyphi, Paratyphi C, and Sendai (CHENG et al. 2019). Most of the infections occurs due to direct contact by the consumption of contaminated food containing raw eggs or undercooked meat. Fresh products and already-to-eat food (lettuce, sprouts, etc...) are wining importance. Indirect transmission of the pathogen may occur between humans and infected animals or their environments (RABSCH et al. 2013).

During the last years, non-adapted *Salmonella* serovars have been reported by international authorities such as the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC) as causative agents of multi-country foodborne outbreaks (FBOs). *S*. Enteritidis was implicated in the majority of the FBOs through the consumption of contaminated eggs, egg products, bakery products, and mixed food (EFSA 2019a). Currently, the investigation of a FBO of *S.* Enteritidis linked to eggs and affecting 18 EU countries including Germany is ongoing and 656 cases have been confirmed. However, the source of contamination is not yet identified (ECDC-EFSA 2017, 2020a). Other serovars such as *S*. Typhimurium and *S*. Anatum were responsible recently for an FBO linked to the consumption of Brazil nuts (ECDC-EFSA 2020b). Additionaly, serovar Agona or serovar Poona were associated with FBOs linked to contaminated food matrixes such as infant formulas or ready-to-eat food (ECDC-EFSA 2018a, 2018b, 2019). According to the German Federal Office of Consumer Protection and Food Safety, in Germany, a total of 13,592 cases of salmonellosis were reported in 2018 (BVL 2019). Salmonellosis outbreaks investigated in Germany are large, regional, and have long been associated with serovars *S*. Enteritidis and *S*. Typhimurium in agreement with the situation in the EU (EFSA 2019a, UELZE et al. 2021). S. Enteritidis linked to eggs or egg-related products is responsible for the largest outbreaks (more than 300 cases), however smaller outbreaks associated with pasta, potato salads with mayonnaise, and bakery products have occurred as well during the last years (RKI 2019). On the other hand, *S*. Typhimurium is responsible for outbreaks with less number of cases (less than 100 cases) and is mostly linked with pork products (UELZE et al. 2021). *S*. Derby has been as well associated with the largest FBOs related to the consumption of pork products, like the one that occurred in 2013 that affected 145 elderly people (SIMON et al. 2018). Other rare serovars associated

with a variety of food sources were reported in 2017. Among them *S*. Kottbus and *S*. Agona were responsible and smoked ham, quail egg, milk products were associated (UELZE et al. 2021).

2.2. Most prevalent *Salmonella enterica* serovars in poultry production

In the context of poultry production, infections by the host-adapted serovar *S*. Gallinarum were of paramount importance at the beginning of intensive poultry production (the 1920s) (METHNER 2013). However, from the 1940s and onwards the ubiquitous *S*. Typhimurium rise in frequency and remained between 1950 and 1985 as a frequent serovar not only in poultry but also in beef, dairy products, and pork. During the mid-1980s and onwards, a sudden epidemiological change happened when *S*. Enteritidis rapidly increased in a very short time in animals and humans in numerous countries including Germany promoting a pandemic situation (METHNER 2013). During approximately ten years, *S*. Enteritidis was the most dominant serovar in laying hens and the cause of a global pandemic. The cause of the sudden *S*. Enteritidis increase remains unknown and is likely that more than one factor was implicated. Several hypotheses point out the role of mice as a reservoir in the poultry flocks and the lack of effective hygiene measures (HENZLER et al. 1992, WARD et al. 2000). The rapid spread was promoted by the consumption of eggs and other poultry products contaminated via vertical transmission from breeders through all the poultry chain (METHNER 2013). The urgent need for the establishment of control measures and high hygiene standards encouraged the development and application in EU state members of the Directive 2003/99/EC of the 17 November 2003 and the Regulation (EC) No 2160/2003 of 17 November 2003 (PARLIAMENT et al. 2003a, 2003b).

The Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents, aim to ensure the proper and effective measures to detect and control *Salmonella* to reduce their prevalence and the risk to public health (PARLIAMENT et al. 2003b). Furthermore, the Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC, regulates the monitoring of zoonoses and food-borne outbreaks caused by zoonotic agents (PARLIAMENT et al. 2003a). As a result of the Directive, *Salmonella* (*S*. Enteritidis and *S*. Typhimurium) prevalence in poultry dropped from approx.

20% to 2.7% in Germany and 3.5% in Europe in 2008 and reached in 2018 a very low prevalence of less than 1-2%.

In Germany, most of the recent human *Salmonella* infections are associated with *S*. Enteritidis (45%), *S*. Typhimurium (33%) followed far behind by *S*. Infantis (2.7%), and *S*. Derby (1.5%) (RKI 2019, 2020). In 2018, *S*. Typhimurium was the most prevalent serovar in cattle followed by the host-adapted *S*. Dublin and *S*. Enteritidis (METHNER. 2019). In pigs, *S*. Typhimurium was the most prevalent serovar in fattening pigs but second in sows were *S*. Derby was the dominant serovar (UELZE et al. 2021). Regarding poultry, *Salmonella* is monitored and controlled by the "Geflügel-Salmonellen-Verordnung (lfSalmoV)" as an implementation of the Regulation (EC) No 2160/2003 aiming at reducing the prevalence of *Salmonella* serovars with public health significance in breeding flocks, laying hens, broilers and turkeys (BUNDESMINISTERIUM 2014).

During the *S*. Enteritidis pandemic, the proportion of *S*. Typhimurium in humans decreased to 20% in contrast to the proportion of *S*. Enteritidis that increased from approx. 5% to over 40% and that reach a prevalence of 75% in 1992. However, in 2019 and according to the German National Salmonella Control program the prevalence of target serovars (*S*. Enteritidis and *S*. Typhimurium, including monophasic *S*. Typhimurium) was for all types of poultry (breeding flocks, broilers and fattening turkeys, laying hens and breeding turkeys) below the target value of 2 % (BfR 2020). The dominating serovars in broilers were *S.* Infantis (23.1%), *S.* Paratyphi B *d*-tartrate-fermenting (*d*T+) (17.3%), and *S*. Enteritidis (15.4%) (BVL 2018, 2019).

Despite a general decreasing trend of the target serovars of *Salmonella* in flocks due to the application of International and National Control Programs, the rapid emergence and widespread of multi-drug resistance *S*. Infantis population is concerning EU and non-EU countries (EFSA 2019a). Studies covering this topic have been published in Japan by ASAI et al. (2007), Hungary by NÓGRÁDY et al. (2007), Israel by GAL-MOR et al. (2010), Italy by FRANCO et al. (2015), CARFORA et al. (2018), and ALBA et al. (2020), Switzerland by HINDERMANN et al. (2017), Slovenia by PATE et al. (2019), Russia by BOGOMAZOVA et al. (2019), Turkey by ACAR et al. (2019), Peru by VALLEJOS-SANCHEZ et al. (2019), Ecuador by VINUEZA-BURGOS et al. (2019), Serbia by JOVCIC et al. (2020), the United States by TYSON et al. (2020), and Chile by LAPIERRE et al. (2020).

2.3. *Salmonella enterica* subspecies *enterica* serovar Infantis: a significant serovar in poultry production

Salmonella enterica subspecies *enterica* serovar Infantis (*S*. Infantis) belongs to the group of non-host-adapted *Salmonella* serovars that cause NTS. It occupies since 2016 the fourth in the list of most common serovars detected in humans in the EU after *S*. Enteritidis, *S*. Typhimurium, and monophasic *S*. Typhimurium (EFSA 2019b). Although *S*. Infantis is prevalent in pigs and cattle, broiler and broiler meat have been regularly identified as the most important sources for human infection (RAJIC et al. 2005, LINDQVIST et al. 2007, EFSA 2019a, EFSA 2020). During 2018, it was the dominant serovar reported in fowl accounting for 36.7% of all serotyped isolates, and unlike in previous years, *S*. Infantis was massively reported from broilers (36.5%) and broiler meat (56.7%)(EFSA 2019a). Studies performed by EFSA and the ECDC in 2018, detected an increased level of resistance over time in the EU. A multi-drug resistant (MDR) pattern consisting of resistance only to ciprofloxacin, nalidixic acid, sulfamethoxazole, and tetracycline was found among the multiresistant *S*. Infantis isolates recovered from broiler (60.3%) and broiler meat (74.2%) (EFSA 2020). Regarding fluoroquinolones (i.e. ciprofloxacin and nalidixic acid) it was detected in *S*. Infantis isolates recovered from broilers, fattening turkeys, and poultry meat (EFSA 2020). Besides, tigecycline-resistant *S*. Infantis accounted for 85.2% and 88.2% isolates from broiler and their carcasses (broiler meat). Meanwhile, resistance to colistin among *Salmonella* isolates recovered from broilers was more common in *S*. Enteritidis isolates (63.2%). Extended-spectrum beta-lactamase (ESBL)-producing *Salmonella* were identified in several serovars (including Infantis) from broilers more often than AmpC-producing *Salmonella*. Two European countries Italy and Hungary reported *S*. Infantis isolates with both ESBL and AmpC resistant phenotypes (FRANCO et al. 2015, SZMOLKA et al. 2018, EFSA 2020).

In Germany, *S*. Infantis is after *S*. Enteritidis and *S*. Typhimurium (including *S*. Typhimurium monophasic variant) the most common serovar involved in human salmonellosis reflecting the situation in the EU. In 2016 *S.* Infantis was by far the most reported serovar in chicken meat (61%) (BVL 2019). Regarding antimicrobial resistance, the percentage of sensitive isolates recovered from turkey meat (35.9%) and chicken meat (28.6%) were smaller in contrast to the percentage of sensitive isolates from pigs (52%). None of the isolates were resistant to 3rd generation cephalosporins (cefotaxime and ceftazidime) or carbapenemases (BVL 2019).

However, resistance to ciprofloxacin which is of particular importance as a treatment in humans was very common in broiler meat (62.3%). This confirms the results of previous investigations that reported significantly higher resistance rates in isolates from poultry compared with cattle and pigs (BVL 2019). It also reflects the more frequent use of this class of substances in poultry compared to cattle and pigs.

2.3.1. The detection and characterization of a conjugative megaplasmid (pESI) among the multidrug resistant emergent *S*. Infantis population

Israel published the first comparative analysis between non-emergent and emergent *S*. Infantis clones (AVIV et al. 2014). They studied *in vitro* the features of a conjugative megaplasmid named pESI (plasmid of Emerging *S.* Infantis) harbored by the Israeli emergent *S*. Infantis (ESI) clones. Compared with the parental plasmidless *S*. Infantis strains, the presence of the pESI plasmid contributed significantly to the antimicrobial resistance and high pathogenicity of the clones. Moreover, they proposed that the possession of this unique virulence-resistance pESI might promote the rapid spread of the MDR clones and potential replacement of the parental *S*. Infantis population (AVIV et al. 2014). Recently, the complete gap-free genome sequence of the Israeli emergent *S*. Infantis strain 119944 has been published by COHEN et al. (2020), and the genetic composition of its chromosome and pESI plasmid have been elucidated in detail improving the knowledge of its genetic composition. The genome of *S*. Infantis 119944 is composed of one chromosome of 4,725,957 bp and a 258,081 bp megaplasmid (pESI). The chromosome possesses ten known *Salmonella* pathogenicity islands (SPIs) (SPIs1-6, SPI-9, SPI-11, SPI-12, and CS54) and five chromosomal bacteriophages (COHEN et al. 2020). The pESI sequence revealed a megaplasmid of modular structure, incorporating different mobile genetic elements (MGEs) such as insertion sequence elements, transposases, and hypothetical proteins. Moreover, pESI is integrated by a conserved backbone comprising AMR, metal-resistance, and virulence-associated factors. These determinants define the MDR profile, enhance the pathogenicity, and contribute to the bacterial fitness of the ESI clones. Among the resistance factors, pESI contains genes mediating AMR to tetracyclines (*tetA*), sulfonamide (*sul1*), and trimethoprim resistance (*dfrA*). It also contains genes conferring resistance to quaternary ammonium compounds (QACs) (*qacEdelta1*) and tolerance to toxic compounds including mercury, due to the mercury operon (*merRDTPCA*). The virulence factors included in the pESI backbone comprise the

yersiniabactin-iron system operon, the Klf (K88-like fimbria), and the Ipf (Infantis plasmid-encoded fimbria) chaperon-usher fimbriae as previously described (AVIV et al. 2017). Furthermore, this virulence-resistance plasmid also carries different sets of toxin/antitoxin systems (e.g. CcdA/B, PemK/I, MazE/F, and VagCD) (COHEN et al. 2020).

The use of WGS and the accuracy of sophisticated bioinformatics tools have recently enabled the occurrence of detailed comparative genome analyses between ESI clones from distant locations (GYMOESE et al. 2019). Independently of the origin of the strains, a conserved chromosomal structure in contrast to the variable nature of the pESI has been described by COHEN et al. (2020). This variability may be the reason for the high genetic variation and phenotypic diversity existing among the different emergent MDR *S*. Infantis clones and their corresponding plasmids (pESI-like plasmids) (BOGOMAZOVA et al. 2019, ALBA et al. 2020, COHEN et al. 2020). Recent comparative genome analysis of the whole-, core- and accessory genomes of several *S*. Infantis strains and a variety of non-Infantis serovar strains revealed the high diversity within the accessory genome of *S*. Infantis. This finding suggests that the contribution to the rapid evolution of the ESI clones is more likely to be due to the variability in the accessory compared to the conserved chromosome (NAGY et al. 2020). Furthermore, ALBA et al. (2020) considered the pESI-like plasmids within ESI isolates in Europe as parasitic megaplasmids due to - in the words of the authors - their role to spread and “infect” different clonal lineages observed in Europe (ALBA et al. 2020).

2.3.2. The ongoing emergence and wide dissemination of the multidrug resistant *S*. Infantis population

In parallel with the rise of the incidence and rapid dissemination of MDR *S*. Infantis clones in broiler populations, over the last decade, several studies worldwide have been published to give evidence of this urgent public health concern. By the end of the 2000s, results of NÓGRÁDY et al. (2007) and NÓGRÁDY et al. (2008) showed the first evidence of the emergence, prevalence, and potential spread of MDR *S*. Infantis clones in broiler production in Hungary. Later, using phage typing and Pulsed-Field Gel Electrophoresis (PFGE) typing they observed closely related MDR clones in neighboring EU countries such as Germany, Italy, the United Kingdom, Poland, and Austria as well as in further non-EU countries as Japan and Israel (NÓGRÁDY et al. 2012). Years later and based on these previous investigations, other Hungarian studies focused their research on the sequencing and comparison of *S*. Infantis

genomes isolated from different sources (humans and broilers) during different decades to understand the rise, spread, and evolution of the Hungarian ESI clones (OLASZ et al. 2015, WILK et al. 2016, 2017). By this time, a Japanese group performed a phylogenetic and network analysis of *S*. Infantis strains from chicken meat, eggshells, environment, and human patients with and without symptoms (YOKOYAMA et al. 2014). They distinguished an evolutionary separation of the *S*. Infantis population in five genetic clusters, and in a further study YOKOYAMA et al. (2015) differentiated within the genetic clusters a novel subpopulation of strains carrying a megaplasmid with clear homology to the pESI plasmid recently reported in Israel (AVIV et al. 2014).

Meanwhile, FRANCO et al. (2015) proposed 2011 as the starting time point for the emergence of the MDR *S*. Infantis clones in Italy and detected among all (n=49) the isolates studied between 2011 and 2014 the presence of a conjugative plasmid of around 280-320Kb in size analogous to the Israeli pESI. However, the pESI-like plasmid from the Italian ESI strains carried the gene $bla_{\text{CTX-M-1}}$ leading to resistance to beta-lactamases (cephalosporin). The detection and characterization of the *mcr-1* gene mediating resistance to colistin among one ESI ESBL-positive clone completed the resistance profile of ESI in Italy (CARFORA et al. 2018).

Besides, in 2017, a study in Hungary revealed more details about the molecular epidemiology of the pESI-like plasmid found in the endemic *S*. Infantis strains from Hungary, and they detected as well genes for ESBL resistance ($bla_{\text{TEM-1)}}$ and *qnrS* for fluoroquinolone resistance (SZMOLKA et al. 2018). Moreover, they discussed the possibility of a switch in the epidemiology of *S*. Infantis isolates in poultry around the 1990s and the early 2000s. The same year, in Switzerland HINDERMANN et al. (2017) observed the occurrence of one ESBL-producing *S*. Infantis clone harboring the gene $bla_{\text{CTX-M-65}}$ rarely described within the European poultry industry. By this year, TATE et al. (2017) reported the existence of pESI-like positive *S*. Infantis in the United States for the first time. The analysis revealed the presence of the gene $bla_{\text{CTX-M-65}}$ among ESI clones reflecting the similarities to the Italian ESI clones. One year before, an outbreak in Ecuador as well revealed the presence of the $bla_{\text{CTX-M-65}}$ gene within *S*. Infantis strains, and in 2019 colleagues from Peru showed useful information to understand the spreading situation of the serovar in Peru (CARTELLE GESTAL et al. 2016, VALLEJOS-SANCHEZ et al. 2019).

More recently, colleagues performed a study in Slovenia in several farms with a focus on the virulence profile of the ESI clones to give insights into the mechanisms involved in the persistence of specific clones (PATE et al. 2019). They tested the biofilm-forming capacity of the MDR *S*. Infantis clones and concluded that the rapid dissemination of ESI clones in broiler production might be more related to the ineffective biosafety measures and disinfection practices than to its capacity to form biofilms. The implementation of WGS and bioinformatics pipelines in the routine practice of microbiology laboratories allowed the easy collection from public databases of complete sequenced genomes and the performance of comparative WGS-based international studies. GYMOESE et al. (2019) analyzed a collection of strains from different decades, sources, and locations providing more insights into the genetic composition and the global population structure of *S*. Infantis. They revealed the polyphyletic nature of the serovar Infantis and elucidated several lineages into the serovar population in which prophages might play an important role in the evolution of the serovar. The same year, a comparative meta-analysis performed in Russia using complete pESI-like sequences from *S*. Infantis isolates from Israel, Japan, Europe, the United States, and South America, provided detailed information about the conservative and variable components of pESI-like plasmids (BOGOMAZOVA et al. 2019). They detected chromosomal mutations in the gen *gyrA* (*gyrA*-S83Y and *gyrA*-D87Y) leading to resistance to fluoroquinolones and they propose that the presence of a conserved sequence of around 173 Kb, could have a major contribution to the global spread of the pESI- like plasmids. At the same time, a study in Turkey performed a single nucleotide polymorphism (SNP)-based phylogenetic analysis comparing strains from European and non-European countries revealing differences and similarities within the ESI clones (ACAR et al. 2019). Among the similarities, they observed similar AMR and virulence profiles among the Italian, Hungarian, and American strains. Interestingly, they found for the first time the ESBL-genes $bla_{\text{TEM-70}}$, $bla_{\text{TEM-148}}$, and $bla_{\text{TEM-198}}$ within Turkish ESI strains.

The latest studies on this topic have provided insights into the epidemiology of ESI populations focussing as well on the emergence of new AMR patterns ESBL and colistin. ALBA et al. (2020) used a chromosome and plasmid-based genotyping approach to elucidate the *S*. Infantis population in Europe confirming the heterogenicity of this serovar. Furthermore, they compared two variants of the pESI-like plasmids in terms of possession of $bla_{\text{CTX-M-1}}$ gene (the European isolates) or $bla_{\text{CTX-M-65}}$ gene (American isolates) thus updating the knowledge about the ESBL-positive ESI clones. Recently, the AMR genetic profiles and PFGE profiles were

determined to study the diffusion of ESI ESBL-positive in broiler meat production (PROIETTI et al. 2020). Meanwhile, JOVCIC et al. (2020) have focussed their research on the study of colistin-resistant *S*. Infantis. They observed reduced susceptibility to colistin in the majority of isolates and they reported for the first time in an *S*. Infantis isolate the detection of gene *fosA7* for resistance to fosfomycin and the gene *vgaA* for resistance to pleuromutilin.

Finally, KUREKCI et al. (2021) has provided updated insights into the pESI-like megaplasmid circulating in Turkey using ultimate sequencing technologies and hybrid assembly of the isolates. Among the ESI strains studied, they identified for the first time a single novel sequence type (ST7091) and one ESI isolate carrying the $bla_{\text{CMY-2}}$ gene mediating resistance to ceftazidime. In the United States, a very recent publication analyses the variation in gene content and the spread rate of a pESI-like plasmid carrying $bla_{\text{CTX-M-65}}$ gene (MCMILLAN et al. 2020).

In less than 10 years, several investigations focussed on the widely disseminated MDR and virulent *S*. Infantis population in EU and non-EU poultry farms. These studies have benefited from the rise in parallel with the most sophisticated WGS and bioinformatics approaches. These approaches warrant detailed knowledge to explain the causes behind the emergence and spread of this serovar.

2.4. Whole-genome sequencing and bioinformatics approaches in public health laboratories practice

WGS and bioinformatics analysis have substantially improved molecular diagnostics and foodborne pathogens surveillance (JAGADEESAN et al. 2019, HERNÁNDEZ et al. 2020). The application of WGS in microbiology laboratories has evolved rapidly over the last 40 years (CARRICO et al. 2018, KUMAR et al. 2019). The first DNA sequencing method ("first-generation" sequencing) was described in 1975 by Sanger and Coulson who elucidated the genome sequence of the bacteriophage ϕX174. The automated Sanger method deciphered the nucleotide sequences in single-stranded DNA using synthetic dideoxynucleosides and the enzyme DNA polymerase (SANGER et al. 1975). For almost 20 years, Sanger's method dominated the industry and still has many applications in projects involving sequenced data (SLATKO et al. 2018). In 1995, the nucleotide sequence of the genome of the bacterium *Haemophilus influenzae* was completed by FLEISCHMANN et al. (1995) due to the development of newer "second-generation" technologies or so-called Next-Generation

Sequencing (NGS). Four major leading NGS platforms can be mentioned. The platform 454 Genome Sequencer based on pyrosequencing was the first released technology in 2005 by 454 Life Sciences (now Roche) (MARGULIES et al. 2005). Only one year later, Solexa/Ilumina (now Illumina) sequencing platform appeared in the market and in 2007 applied Biosystems (now Life Technologies) released Sequencing by Oligo Ligation Detection (SOLiD). Later on, Ion Torrent technology (Ion Torrent Systems) appeared in 2011 using a novel approach based on a semiconductor chip that converts chemical information into sequencing information (SLATKO et al. 2018). NGS brought three major advances compared with Sanger sequencing. First, the need for a previous preparation of sequencing libraries, second, the performance of millions of sequencing reactions in parallel, and third, the ability to produce a faster large amount of data (VAN DIJK et al. 2014). The differences between them reside in the combination of strategies employed and the type of output generated (METZKER. 2010). Even though their numerous advantages, NGS platforms are not able to generate single molecules in real-time but relatively short reads (currently up to 300-500 bases). Therefore, under the names Third-Generation Sequencing (TGS) and Fourth-Generation Sequencing (FGS), real-time single-molecule sequencing technologies appeared. Pacific Biosciences (PacBio) leads currently the commercialization of TGS technologies and is considered the gold standard for the generation of contiguous and highly accurate reference genomes. It enables the sequencing of a single very long molecule (up to 30-50 kb) in real-time. On the other hand, nanopore-based DNA sequencing is being developed and emerged as competitive technology. It consists on the use of transmembrane proteins to produce pores that detect the DNA nucleotides and measure its current differences. The dominant platform is the portable sequencer MinION developed by Oxford Nanopore Technologies (ONT) (JAIN et al. 2016, SLATKO et al. 2018). While Illumina is currently the leading NGS platform for short-read sequencing (reads of 100-300 bp), MinION promises to be the dominant platform for long-read sequencing (reads from 10 to 50kb). To improve accuracy and compensate the drawbacks of short- and long-read sequencing approaches, PacBio or MinIOn assemblies are combined with Illumina data in so called “hybrid” assemblies. Besides, other novel technologies under development are currently being used in routine laboratory practice (e.g., *in situ* nucleic acid sequencing, microscopy-based sequencing and, whole mitochondrial genome sequencing) (KUMAR et al. 2019, ZASCAVAGE et al. 2019).

The applications of WGS and bioinformatics in public health laboratories provide detailed information of genetic treats of strains and samples (e.g., detection of antimicrobial resistance genes or virulence determinants, and the identification of plasmid replicons) (HENDRIKSEN et al. 2019, CARATTOLI et al. 2020), and the quick detection and tracking of outbreaks (e.g., serotyping, core-genome Multi Locus Sequence Typing (cgMLST), whole-genome Single Nucleotide Polymorphism (wgSNP) typing, phylogenetic analysis and, pangenome comparisons) (PETZOLD et al. 2017, QUAINOO et al. 2017). The possibilities are as wide as the informatics capabilities of the researcher and the computational resources of the laboratories. The definitions of the most used terms in a bioinformatics analysis are shown in Table 1.

In the case of *Salmonella*, research institutions as the Public Health England (PHE), the Sanger Institut, the Food and Drug Administration (FDA)) and the Centers for Disease Control and Prevention (CDC) with the global laboratory network PulseNet have been implementing WGS as the routine typing tool for *Salmonella* infections surveillance (WONG et al. 2016, RIBOT et al. 2016, ASHTON et al. 2016, NADON et al. 2017, WALDRAM et al. 2018, FELDGARDEN et al. 2019). Besides, under the name 10KSG (10000 *Salmonella* Genomes) consortium project, *S. enterica* data is being retrieved worldwide to understand the epidemiology, transmission, and virulence of NTS (ACHTMAN et al. 2020). The amount of data retrieved is accessible and freely available in public datasets. Currently, the number of publicly available *Salmonella* sequenced data is close to 300,000 genomes and is stored by the European Nucleotide Archive (ENA), the Sequence Read Archive (SRA), and Enterobase (ALIKHAN et al. 2018, ZHOU et al. 2020, PEREZ-SEPULVEDA et al. 2020) which comprises itself more than 287,101 *Salmonella* strains (in February 2021). However, the complete standardization of WGS in the food industry or veterinary and public health laboratories is far to be completely implemented. Furthermore, there are still limitations in terms of computing capacity and a lack of standard models for routine analysis that need to be considered (ECDC 2016).

Table 1. Definition of most used terms and parameters of a bioinformatics analysis (extracted from the ISO standard Microbiology of the food chain — Whole-genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance ISO/DIS 23418) (ISO 2020)

Annotation	The process of identifying genes and other features on genome assemblies
Assembly	The output from the process of aligning and merging sequencing reads into larger contiguous sequences (contigs)
Base-calling	The process of assigning nucleotides and quality scores to positions in sequencing reads
Contig	A contiguous stretch of DNA sequence that results from the assembly of smaller, overlapping DNA sequence reads
***de novo* assembly**	The process where the reads are assembled without the assistance of a reference genome as a template
Coverage	The average number of times each base in a genome is sequenced
Mapping	The use of software to align sequencing reads to reference sequences
Multi Locus Sequence Typing (MLST)	The method of genomic analysis in which nucleotide variants within predefined sets of loci, either core genome loci for cgMLST or whole-genome loci for wgMLST, are identified. For MLST analyses, reads are assembled or mapped. Target loci are identified, quality-filtered, and compared to a curated cgMLST or wgMLST database.
Sequencing library	The collection of genomic DNA fragments from a single isolate intended for determining genome sequence
N50	A measure to describe the quality of assembled genomes that are fragmented in contigs of different lengths. The N50 is defined as the minimum contig length needed to cover 50% of the genome. It means half of the genome sequence is in contigs larger than or equal to the N50 contig size.
Q30	A Phred-score of Q30 indicates that there is a 1 in 1000 chance that a base is incorrectly assigned (i.e. the base call is 99.9% accurate)
Pipeline	The software that combines other software to transfer data from file type A (e.g. raw sequencing data) via several steps to file type X (e.g. a phylogenetic tree)
Reads	The nucleotide sequence inferred from a fragment of DNA or RNA
Single Nucleotide Polymorphisms (SNPs) analyses	The process comprising the mapping of reads against a reference genome sequence and the study of the detected significant differences between mapped reads and the reference (SNP-calling). As an alternative, reads can also be assembled into contigs, and contigs from each sample can be aligned to each other to identify SNPs. Potential SNPs are quality-filtered to identify SNP positions. Those SNPs present within the conserved portion among all genomes are called cgSNPs. The matrix of the number of SNPs or cgSNPs differences between samples is used to create a phylogenetic tree.

2.4.1. Implementation and evaluation of a bioinformatics pipeline for the analysis of *Salmonella spp*. sequenced data

WGS has become the method of choice for the characterization and subtyping assays of *Salmonella spp*. and the list of bioinformatics tools and sophisticated pipelines is growing to simplify the automatic processing of sequenced data (CARRICO et al. 2018, TANG et al. 2019, BANERJI et al. 2020).

The last version of the in-house bioinformatics pipeline WGSBAC (v. 2.1.0) developed by the bioinformatics group at the Institute of Bacterial Infections and Zoonoses (IBIZ) is a public resource available online in https://gitlab.com/FLI_Bioinfo/WGSBAC. It is Linux-based and contains several modules including modified computer scripts, R language scripts, and public databases (Figure 1).

The pipeline WGSBAC takes as input raw short-reads (from Illumina) as well as already assembled data from own sequenced genomes (from Illumina or long-reads from MinION) or collected from public databases. The pipeline WGSBAC workflow is built based on several steps that any complete basic bioinformatics analysis should comprise: i) quality assessment and control of the sequenced data generated, ii) assembly and quality control of the assemblies iii) annotation, subtyping (sometimes including genoserotyping), detection of resistance genes, virulence genes and plasmid replicons, etc ...

When using WGSBAC to analyze genomic data, the first step, quality assessment and quality control (QA and QC) is performed by the software FASTQC (v. 0.11.7), that offers several statistics to provide an overview evaluation of the quality of the raw reads (ANDREWS 2018). WGSBAC uses as well an adapted script to calculate the theoretical coverage that can be defined as *LN/G*, where *L* is the read length, *N* is the number of reads and *G* is the haploid genome length (SIMS et al. 2014). For contamination detection, WGSBAC uses the software Kraken2 (v. 2.0.7_beta) and the database Kraken2DB (https://benlangmead.github.io/aws-indexes/k2) (WOOD et al. 2019). Kraken is a software for taxonomical classification of reads as well as assembled genomes and is used frequently in metagenomics projects.

Once the quality of the reads has been accurately checked, the second step of the bioinformatics analysis starts with the assembly. WGSBAC uses Shovill (v. 1.0.4) which is itself a bioinformatics pipeline developed by SEEMANN (2018) that uses the SPAdes assembler to

perform the *de novo* assembly on Illumina paired-end reads (BANKEVICH et al. 2012). Later on, the quality of the assemblies is assessed by the software QUAST (v. 5.0.2) that considers parameters such as the number of *contigs*, its average length in the draft genome, the size of the assembled genome, and the statistic N50 value (GUREVICH et al. 2013, CARRICO et al. 2018).

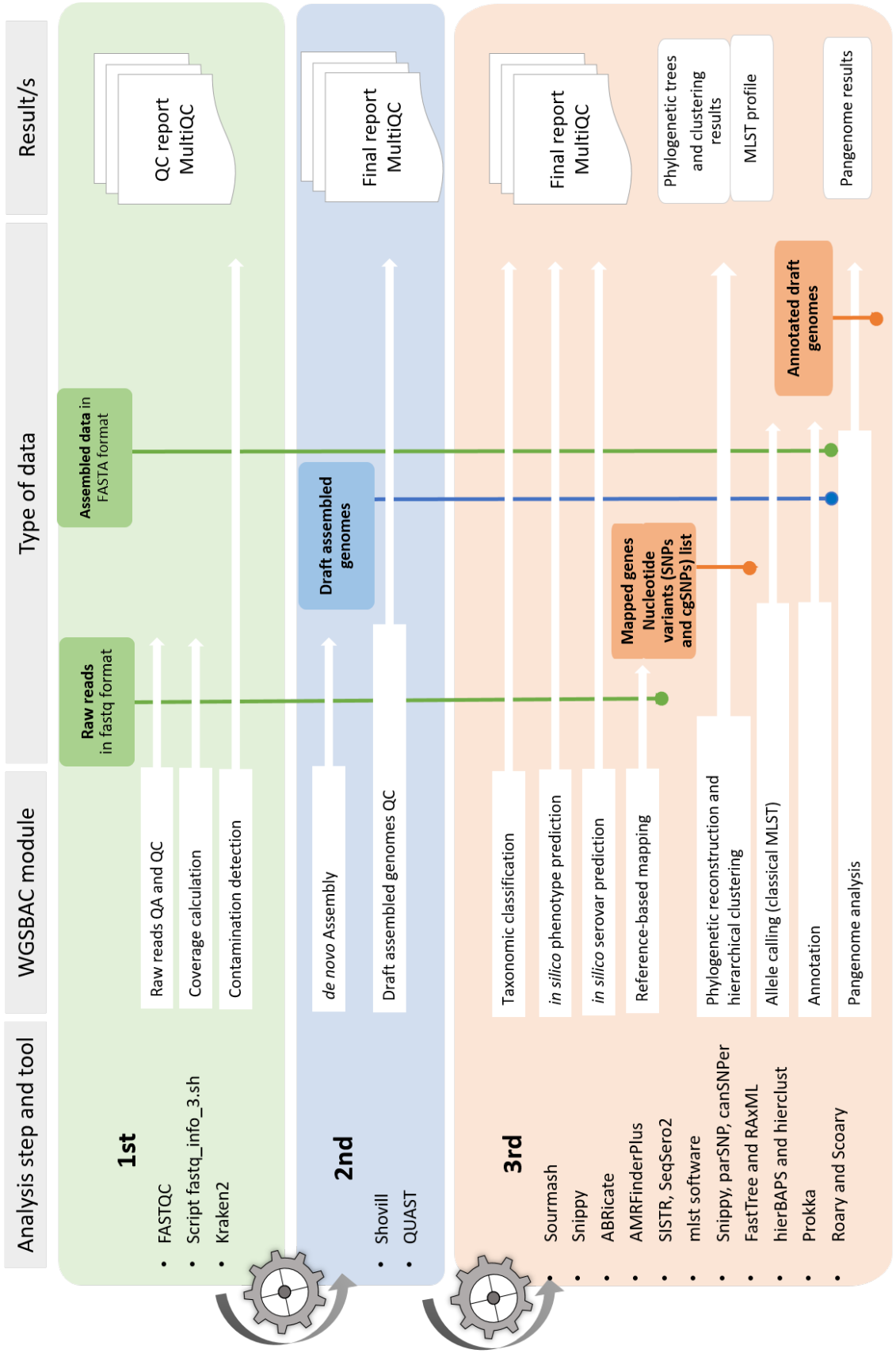
The third step of WGSBAC analysis comprises a variety of tools that will be set depending on the specific goal of the researcher study and that will reveal the biological meaning behind the genomes sequenced. Detection of AMR determinants and virulence factors is possible as well due to the screening of AMR or virulence genes as well as plasmid replicons identification. WGSBAC uses the tool ABRicate (v. 0.8.10) developed by SEEMANN (2015) and the databases ResFinder, the Comprehensive Antimicrobial Resistance Database (CARD), and NCBI for AMR detection and Virulence Factor Database (VFDB) for virulence gene detection (ZANKARI et al. 2012, FELDGARDEN et al. 2019, ALCOCK et al. 2020. PlasmidFinder is used to identify plasmid replicon genes and the tool AMRFinderPlus (v. 3.6.10) to complete the AMR genes screening with the detection of chromosomal point mutations dealing with resistance to antibiotics (FELDGARDEN et al. 2019, CARATTOLI et al. 2020). In the specific case of *Salmonella* genomes, WGSBAC includes three *in silico* serotyping tools that predict *Salmonella* serovars both on the reads and on the assemblies: SeqSero2 (v. 1.1.1) by ZHANG et al. (2019), and SISTR (v. 1.0.2) by YOSHIDA et al. (2016).

Furthermore, complete genotyping of the genomes is performed using classical Multilocus Sequence Typing (MLST). The corresponding sequence types (STs) will be assessed using the mlst software (v. 2.16.1) developed by SEEMANN (2014a) that employs the pubMLST database (JOLLEY et al. 2018). Reference base-mapping of the reads to a reference sequence or database of reference sequences is performed by Snippy (v. 4.3.6) (SEEMANN 2014b). It generates a list of the identified nucleotide variants (e.g., SNPs, cgSNPs, or short insertions/deletions) between the query genomes and the reference genomes. The nucleotide variants (SNPs and cgSNPs) list and a matrix of cgSNPs differences between samples are used for cgSNPs-based phylogenetic reconstruction. Phylogeny trees will be constructed by FastTree (v. 2.1.10) and/or RaxML (v. 8.2.12) (PRICE et al. 2009, STAMATAKIS. 2014). To complete the phylogenetic reconstruction of the strains, the generation of phylogenomic maps is possible due to the implementation of WGSBAC with the R package “ggplot”.

Additionally, software such as parSNP (v. 1.2) for core genome multi-alignment and canSNPer (v. 1.0.8) for hierarchical genotyping can be optionally set within the pipeline (TREANGEN et al. 2014, LÄRKERYD et al. 2014). Identification of subpopulations within the genomes is possible as WGSBAC includes an R implementation of hierarchical clustering through the hierBAPS algorithm (TONKIN-HILL et al. 2018). Furthermore, WGSBAC uses Prokka (v. 1.14.5) by SEEMANN (2014b) to annotate the genomes revealing the location and biological role of the genetic features encrypted in the DNA sequence. Later, the software Roary (v. 3.13.0) developed by PAGE et al. (2015) works on annotated genomes, generates a multiple alignment, and together with a script called Scoary, performs a pan-genome association analysis useful to make the association between presence/absence of genes and genetic traits related. Pan-genome analysis is interesting for identifying the core and accessory genes and to understand the evolution of the genomes.

Finally, to complement the bioinformatics analysis, external resources can be employed such as online tools with a user-friendly interface and curated public databases. For example, when using WGSBAC two software options are Ridom SeqSphere+ created by JUNEMANN et al. (2013) to infer phylogeny based on cgMLST and the web-based tool iTOL that helps in the visualization of phylogenetic trees (LETUNIC et al. 2019).

Figure 1. Workflow of WGSBAC (v. 2.1.0) with a focus on *Salmonella spp*. data analyses



3. PUBLICATION

García-Soto, S., Abdel-Glil, M., Tomaso, H., Linde, J., Methner, U. (2020)

Emergence of Multidrug-Resistant *Salmonella enterica* subespecies *enterica* serovar Infantis of Multilocus Sequence Type 2283 in German Broiler Farms

Frontiers Microbiology, Vol.11, Article 1741, 1-12.

Own contribution:

The following tasks and analysis were performed by myself:

- Application of the bioinformatics pipeline WGSBAC and performance of bioinformatics analysis (QA and QC of the data sequenced, assembly and QC of assemblies; reference-based mapping (of reads and assemblies), annotation, typing by classical MLST, cgMLST, subtyping including genoserotyping, SNPs analysis, phylogenetic analysis, and detection of resistance genes, virulence genes, and plasmid replicons)
- Searching and retrieving public datasets for the comparative genome analysis (ENA, SRA and, Enterobase)
- Genomic analysis of the complete sequence genome of the pESI-like plasmid (using WGSBAC pipeline and external software such as Ridom SeqSphere+, Geneious Prime, iTOL, and R programming language)
- Expression of results in tables and production of figures
- Upload of own sequenced dataset to online repositories (ENA)
- Writing of the manuscript and working on reviewers corrections

The tasks of the co-authors comprise conceiving and coordinating the study (PD Dr. Ulrich Methner and Dr. Jörg Linde). Performing serotyping, MIC determination, and writing the manuscript (PD Dr. Ulrich Methner). Developing, implementing, maintaining the bioinformatics pipeline, and writing the manuscript (Dr. Jörg Linde). Performing MinION sequencing and raw data analysis (Dr. Mostafa Abdel-Glil). Coordinating the study and critically reading the manuscript (PD Dr. Herbert Tomaso).

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 July 2020
doi: 10.3389/fmicb.2020.01741

Emergence of Multidrug-Resistant *Salmonella enterica* Subspecies *enterica* Serovar Infantis of Multilocus Sequence Type 2283 in German Broiler Farms

Silvia García-Soto, Mostafa Y. Abdel-Glil, Herbert Tomaso, Jörg Linde and Ulrich Methner**

Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institute, Jena, Germany

OPEN ACCESS

Edited by:
Daniel Yero, Autonomous University of Barcelona, Spain

Reviewed by:
Ohad Gal-Mor, Tel Aviv University, Israel
Béla Nagy, Institute for Veterinary Medical Research (MTA), Hungary

***Correspondence:**
Jörg Linde
joerg.linde@fli.de
Ulrich Methner
ulrich.methner@fli.de

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 March 2020*
Accepted: *03 July 2020*
Published: *17 July 2020*

Citation:
García-Soto S, Abdel-Glil MY, Tomaso H, Linde J and Methner U (2020) Emergence of Multidrug-Resistant Salmonella enterica Subspecies enterica Serovar Infantis of Multilocus Sequence Type 2283 in German Broiler Farms. Front. Microbiol. 11:1741. doi: 10.3389/fmicb.2020.01741

During the last decade, *Salmonella enterica* subspecies *enterica* serovar Infantis (*S*. Infantis) has become more prevalent across Europe with an increased capability to persist in broiler farms. In this study, we aimed to identify potential genetic causes for the increased emergence and longer persistence of *S*. Infantis in German poultry farms by high-throughput-sequencing. Broiler derived *S*. Infantis strains from two decades, the 1990s (n = 12) and the 2010s (n = 18), were examined phenotypically and genotypically to detect potential differences responsible for increased prevalence and persistence. *S*. Infantis organisms were characterized by serotyping and determining antimicrobial susceptibility using the microdilution method. Genotypic characteristics were analyzed by whole genome sequencing (WGS) to detect antimicrobial resistance and virulence genes as well as plasmids. To detect possible clonal relatedness within *S*. Infantis organisms, 17 accessible genomes from previous studies about emergent *S*. Infantis were downloaded and analyzed using complete genome sequence of SI119944 from Israel as reference. In contrast to the broiler derived antibiotic-sensitive *S*. Infantis strains from the 1990s, the majority of strains from the 2010s (15 out of 18) revealed a multidrug-resistance (MDR) phenotype that encodes for at least three antimicrobials families: aminoglycosides [*ant(3")-Ia*], sulfonamides (*sul1*), and tetracyclines [*tet(A)*]. Moreover, these MDR strains carry a virulence gene pattern missing in strains from the 1990s. It includes genes encoding for fimbriae clusters, the yersiniabactin siderophore, mercury and disinfectants resistance and toxin/antitoxin complexes. In depth genomic analysis confirmed that the 15 MDR strains from the 2010s carry a pESI-like megaplasmid with resistance and virulence gene patterns detected in the emerged *S*. Infantis strain SI119944 from Israel and clones inside and outside Europe. Genotyping analysis revealed two sequence types (STs) among the resistant strains from the 2010s, ST2283 (n = 13) and ST32 (n = 2). The sensitive strains from the 1990s, belong to sequence type ST32 (n = 10) and ST1032 (n = 2). Therefore,

this study confirms the emergence of a MDR *S*. Infantis pESI-like clone of ST2283 in German broiler farms with presumably high tendency of dissemination. Further studies on the epidemiology and control of *S*. Infantis in broilers are needed to prevent the transfer from poultry into the human food chain.

Keywords: *Salmonella* Infantis, broiler, emergence, multidrug resistance, pESI-like plasmid, whole genome sequencing

INTRODUCTION

Salmonella enterica subspecies *enterica* serovar Infantis (*S*. Infantis) belongs to the group of *Salmonella* serovars, which plays a major epidemiological role in humans and animals. In the European Union (EU), *S*. Infantis has been the third most common serovar in humans since 2006 with a relative share between 1% and 2% (EFSA, 2019). Although this serovar is prevalent also in pigs and cattle (Rajic et al., 2005; Lindqvist and Pelkonen, 2007), poultry especially broiler and their products have been identified as one of the most important sources of human infection with *S*. Infantis (EFSA, 2019, 2020). In 2018, *S*. Infantis was the most frequently reported serovar in fowl in the EU (EFSA, 2019), accounting for 36.7% of all *Salmonella* isolates. Moreover, unlike previous years, *S*. Infantis was not only detected in a few numbers of countries but widespread among most member states and massively reported from broilers (36.5% of all serotyped isolates) and broiler meat (56.7%).

During the last years, antimicrobial resistance has emerged in *S*. Infantis organisms from different animal sources and humans in various European countries (Nógrády et al., 2012; EFSA, 2020). Increasing incidence and dissemination of different multidrug-resistant (MDR) *S*. Infantis clones in broiler populations resulted in spreading of the organisms in the food chain and via poultry products to humans in countries such as Hungary (Olasz et al., 2015), Italy (Franco et al., 2015), Switzerland (Hindermann et al., 2017), Slovenia (Pate et al., 2019), and Russia (Bogomazova et al., 2019). Furthermore, observations are indicating a long persistence of *S*. Infantis in broiler farms and increased resistance against cleaning and disinfection procedures (Asai et al., 2007; Nógrády et al., 2007, 2008; Ross and Heuzenroeder, 2008; Pate et al., 2019). Thus, we were interested whether recent *S*. Infantis organisms gained new properties resulting in the modified characteristics.

There is evidence that the acquisition of a conjugative mega-plasmid provides the bacteria with new resistance properties (EFSA, 2020) but might also confer virulence-associated characteristics, higher resistance against heavy metals or disinfectants and fitness characteristics (Aviv et al., 2014).

In view of the increased prevalence of *S*. Infantis also in German broiler production in recent years (EFSA, 2020), the question raised on possible reasons. Therefore, this study aimed to characterize and compare *S*. Infantis strains originated from different broiler farms in Germany from 20-years distant decades, the 1990s and the 2010s. *S*. Infantis organisms isolated in different decades were phenotypically characterized by serotyping and determining the antimicrobial susceptibility. In this study, whole genome sequencing (WGS) and bioinformatics analysis were used to describe the genetic traits of *S*. Infantis strains from different German broiler farms collected from 20-years distant decades.

MATERIALS AND METHODS

Strain Selection

In this study, we analyzed a dataset consisting of 30 *S*. Infantis strains that cover a wide range of broiler farms in Germany (**Table 1**). Eighteen isolates were collected during the 2010s (time frame: 2014–2020) and 12 strains were isolated two decades earlier, in the 1990s (time frame: 1992–1998). *S*. Infantis strains were provided by the National Reference Laboratory for *Salmonella* at the German Federal Institute for Risk Assessment (BfR) or were received after request from regional diagnostic laboratories in different federal states in Germany. To compare the sequenced German *S*. Infantis strains with previously reported emergent *S*. Infantis clones, we searched for recent publications regarding *S*. Infantis. The criteria of selection of strains were source (broiler), region (central Europe) and, period of time when they were collected (between the 1990s and the 2010s). Thus, we downloaded sequence data of *S*. Infantis ($n = 17$) from Hungary (Olasz et al., 2015; Wilk et al., 2016, 2017) and Italy (Franco et al., 2015) (**Supplementary Table S1**). Beyond this criteria, we included strains from Israel where pESI was first studied (Aviv et al., 2014) and kept the strains 1326/28 (LN649235) from the United Kingdom and the non-broiler strain 335-3 (ATHK00000000) as representatives of the historical, or so-called "pre-emergent" strains. For comparison purposes, we included the recently published complete genome sequence of the *S*. Infantis strain 119944 harboring the pESI like megaplasmid from Israel (Cohen et al., 2020).

Serotyping and Antimicrobial Susceptibility Testing

All *Salmonella* isolates were serotyped using poly- and monovalent anti-O as well as anti-H sera (SIFIN, Germany) according to the Kauffmann–White scheme (Grimont and Weill, 2007). Antimicrobial susceptibility of the *S*. Infantis strains was assessed by determining the minimum inhibitory concentration (MIC) using the broth microdilution method with Sensititre^{TM} EUVSEC plates (Trek Diagnostic Systems Ltd., East Grinstead, United Kingdom). Epidemiological cut-off values were used according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST, 2018). Antimicrobial susceptibilities to sulfamethoxazole (SMX),

TABLE 1 | Epidemiological data and phenotypic AMR profile of *S.* Infantis strains used in this study.

Strain	Sample	Region of isolation	Year of isolation	Phenotypic AMR profile
19PM0346	2945	Bavaria farm 1	1992	SMX
19PM0348	2947	Bavaria farm 2	1992	SMX
19PM0349	2948	Bavaria farm 3	1992	SMX
19PM0350	2949	Lower Saxony farm 1	1992	SMX
19PM0351	2951	Lower Saxony farm 2	1994	SMX
20PM0240	3222	Mecklenburg-Western Pomerania	1994	SMX
20PM0243	3225	State of Hesse farm 2	1995	SMX
20PM0245	3227	Thuringia	1995	SMX
20PM0248	3230	State of Hesse farm 3	1996	SMX
20PM0252	3234	Baden-Wuerttemberg farm 2	1997	SMX
20PM0257	3239	Rhineland-Palatinate farm 2	1998	SMX
20PM0260	3242	Lower Saxony farm 4	1998	SMX
20PM0261	3243	Bavaria farm 4	2014	SMX-CIP-TET-NAL-TGC
20PM0263	3245	Lower Saxony farm 5	2014	SMX
20PM0267	3249	Saxony-Anhalt farm 1	2015	SMX-CIP-TET-NAL-TGC
20PM0268	3250	Bavaria farm 5	2015	SMX
20PM0270	3252	Saxony-Anhalt farm 2	2015	SMX-CIP-TET-NAL-TGC
20PM0271	3253	Brandenburg farm 1	2016	SMX-CIP-TET-NAL-TGC
20PM0273	3255	Lower Saxony farm 6	2016	SMX
20PM0275	3257	Bavaria farm 6	2016	SMX-CIP-TET-NAL-TGC
19PM0355	2954	Baden-Wuerttemberg farm 1	2017	SMX-CIP-TET-NAL-TGC
19PM0358	2957	Brandenburg farm 1	2018	SMX-CIP-TET-NAL-TGC
19PM0360	2959	Bavaria farm 4	2019	SMX-CIP-TET-NAL-TGC
19PM0148	2747	Bavaria farm 5	2019	SMX-CIP-TET-NAL-TGC
19PM0149	2748	Bavaria farm 6	2019	SMX-CIP-TET-NAL-TGC
19PM0150	2749	Bavaria farm 7	2019	SMX-CIP-TET-NAL-TGC
19PM0151	2750	Baden-Wuerttemberg farm 2	2019	SMX-CIP-TET-NAL-TGC
19PM0153	2752	Baden-Wuerttemberg farm 3	2019	SMX-CIP-TET-NAL-TGC
19PM0154	2753	Bavaria farm 8	2019	SMX-CIP-TET-NAL-TGC
20PM0045	3027	Bavaria farm 6	2020	SMX-CIP-TET-NAL-TGC

SMX, sulfamethoxazole; CIP, ciprofloxacin; TET, tetracycline; NAL, nalidixic acid; TGC, tigecycline.

trimethoprim (TMP), ciprofloxacin (CIP), tetracycline (TET), meropenem (MERO), azithromycin (AZI), nalidixic acid (NAL), cefotaxime (FOT), chloramphenicol (CHL), tigecycline (TGC), ceftazidime (TAZ), colistin (COL), ampicillin (AMP), and gentamicin (GEN) were examined.

Sequencing and Bioinformatics Analysis

For paired-end sequencing with Illumina, Genomic DNA of 30 *S.* Infantis strains 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 was performed on an Illumina MiSeq instrument according to the manufacturer's instructions (Illumina Inc., United States).

For long-read sequencing with MinION, a sequencing library was prepared using the Oxford Nanopore Technologies (ONT) 1D Ligation Sequencing Kit (SQK-LSK109) with the Native Barcoding Expansion Kit (EXP-NBD104) as recommended by the manufacturer. Raw FAST5 files were processed using Guppy toolkit (v. 3.4.1) (Oxford Nanopore Technologies). The Guppy command *guppy_basecaller* was used for basecalling and *guppy_barcoder* was used for demultiplexing. *De novo* assembly for long sequencing reads was performed using Flye (v. 2.6) (Kolmogorov et al., 2019). Assembly polishing was performed with four rounds by Racon (v. 1.4.3) (Vaser et al., 2017) and one final round with Medaka (v. 0.10.0). Finally, Pilon (v. 1.23) (Walker et al., 2014) was used to correct the final assembled data from Nanopore with Illumina reads using standard settings.

To analyze the sequencing data in a standardized manner, the Linux-based bioinformatics pipeline WGSBAC was used (v. 2.0.0)[1] (FLI_Bioinfo, 2020). Input for the pipeline was raw Illumina data and already assembled data (MinION). WGSBAC starts with quality control using FastQC (v. 0.11.7)[2] (Andrews, 2018). Next, it calculates the raw coverage by the number of reads multiplied with their average read length and divided by

[1]https://gitlab.com/FLI_Bioinfo/WGSBAC

[2]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

the genome size. WGSBAC performs assembly using Shovill (v. 1.0.4) (Seemann, 2018) an optimizer for SPAdes assembler (Bankevich et al., 2012).

The quality of the assembled genomes is then checked using QUAST (v. 5.0.2) (Gurevich et al., 2013). Genome annotation is made by Prokka (v. 1.14.5). In order to identify contamination, the pipeline uses Kraken 2 (v. 1.1) (Wood et al., 2019) and the database Kraken2DB to classify both reads and assemblies. For *in silico* serotyping, WGSBAC utilizes SISTR (v. 1.0.2) (Yoshida et al., 2016) on the assembled genomes.

For genotyping, WGSBAC uses classical multilocus sequence typing (MLST) on assembled genomes using mlst software (v. 2.16.1)[3] (Seemann, 2014a) that incorporates the PubMLST database[4] (Jolley et al., 2018). Furthermore, the pipeline includes mapping based SNP-typing using Snippy (v. 4.3.6)[5] (Seemann, 2014b) with standard settings and FastTree (v. 2.1.10) (Price et al., 2009) to calculate phylogenetic trees from SNPs. To infer phylogeny based on core genome multilocus sequence typing (cgMLST), we used the external software Ridom Seqsphere+ (v. 5.1.0) (Junemann et al., 2013) with default settings and the specific core genome scheme (cgMLST v2) for *Salmonella enterica* with 3,002 target loci developed by Enterobase (Alikhan et al., 2018).

For detection of antimicrobial resistance genes (AMR), virulence factors and plasmid replicon genes, WGSBAC uses Abricate (v. 0.8.10)[6] (Seemann, 2015) and the databases: ResFinder (Zankari et al., 2012) and NCBI (Feldgarden et al., 2019a), Virulence Factor Database (VFDB) (Chen et al., 2005) and PlasmidFinder (Carattoli et al., 2014), respectively. For the detection of point mutations in the gene *gyrA* leading to AMR, we used the software AMRFinderPlus (v. 3.6.10) (Feldgarden et al., 2019b).

For a deeper molecular characterization of the strains, we downloaded specific pESI119944-encoded gene sequences and created customized databases for Abricate (**Supplementary Table S2**). These databases include sequences of genes encoding for *Salmonella* pathogenicity islands (SPIs), Ipf and K88-like fimbrial clusters and pESI fitness determinants as the toxin-antitoxin (T/AT) system (CcdAB and PemK/MazF) and the mercury operon. Furthermore, for plasmid typing, allele sequences for incompatibility groups of plasmids IncI1 (five loci) and IncF RST (seven loci) were downloaded from the Plasmid PubMLST database[7] (Carattoli and Hasman, 2020). In order to complete the plasmid genomic characterization and to test the chimeric nature of pESI-like plasmids as described before (Aviv et al., 2014), we tested the detection of the gene sequence encoded for RepFIB replication protein A (*rep*B) and the oriV of IncP1 plasmids. We used the external software Geneious Prime (v. 2019.2.3)[8] to complete the plasmid strain annotation and for visualization of its main genomic features.

[3]https://github.com/tseemann/mlst

[4]https://pubmlst.org/salmonella

[5]https://github.com/tseemann/snippy

[6]https://github.com/tseemann/abricate

[7]https://pubmlst.org/plasmid/

[8]https://www.geneious.com

Two phylogenetic trees were constructed for the pESI-like positive strains using Snippy to study the plasmid SNP-based phylogeny. One tree using the plasmid sequence (CP047882) as reference and a second one using the chromosome sequence (CP047881) as reference of the complete genome sequence of SI119944 strain (Cohen et al., 2020). Trees were compared using the tanglegram function of the tool Dendroscope (v 3.5.9) (Huson and Scornavacca, 2012).

RESULTS

Serotyping and Antimicrobial Susceptibility Testing

All isolates were typed according to the Kauffmann–White scheme and revealed the complete antigenic formula (6, 7: r: 1, 5) for *S.* Infantis. As listed in **Table 1**, all *S.* Infantis strains from the 1990s were only resistant against SMX. Among the isolates from the 2010s, three strains were resistant to SMX and 15 were multidrug-resistant to SMX, CIP, TET, NAL, and TGC.

Genomic Features of Genomes of *S.* Infantis Strains

WGS of the 30 *S.* Infantis strains revealed general genomic characteristics and allowed genoserotyping of the strains (**Supplementary Table S3**). We sequenced an average of 1,439,617 reads per sample (range: 524,688–2,509,738). On average, assembled genomes consisted of 54 contigs (range: 37–91) with an average read-coverage of 68 fold (range: 24–128). The average genome size was 4.8 Mbp (range: 4.6–4.9 Mbp), GC content was 52.2% and N50 values average 257,721 (range: 90,139–445,475). Kraken2 on Illumina reads classified an average of 95.31% of reads as "*Salmonella*" on the genus level and an average of 94.42% of reads as "*Salmonella enterica*" at the species level. To confirm the serological serotyping, SISTR was used for *in silico* molecular typing and predicted serovar Infantis for all the strains included in the study corroborating the phenotypic findings.

Genotyping and Phylogeny of *S.* Infantis Strains

After assessing the general sequencing characteristics, classical MLST on the assembled genomes was carried out to get a broad overview of the *S.* Infantis genotypes (**Table 2**). Among the complete dataset, 15 out of 30 isolates examined belong to ST32. The remaining 15 belong to two single-locus variants of ST32, namely ST2283 (in the gene *sucA*) and ST1032 (in the gene *dnaN*). The majority of the strains from the 1990s belong to ST32 (n = 10) while two strains belong to ST1032. The majority of the strains from the 2010s belong to ST2283 (n = 13) while five strains belong to ST32. Within the dataset from the 2010s, only two strains that belong to ST32 revealed the same MDR pattern as *S.* Infantis strains with ST2283. For a deeper phylogenetic analysis of the 30 *S.* Infantis strains, a minimum spanning tree based on core genome MLST (cgMLST) and a phylogenetic tree based on single nucleotide polymorphisms (SNPs) were constructed. Although both approaches differ strongly, they produced similar results (**Figures 1A,B**). Both phylogenetic approaches group the

strains into two distinct decade- and ST-related groups except for samples 2748 and 3027 that belong to ST32 but they are grouped with the samples collected in the 2010s (**Figure 1B**) and three strains from the 2010s that are included within the group of the 1990s as they have ST32 (marked in red in **Figure 1B**). On average, the distance between the strains was 145 SNPs and 76 alleles.

Resistance and Virulence Genes Patterns

The analysis of the genome sequences revealed an AMR gene pattern specific for the 15 MDR strains from the 2010s. We named this pattern ESIr (Emergent S. Infantis resistance pattern) and it consists of the AMR genes: *ant(3")-Ia* (aminoglycoside resistance), *sul1* (sulfonamide resistance), *tet(A)* (TET resistance) and *qacEdelta1* [quaternary ammonium compound (QAC) and disinfectant resistance] (**Table 2**). Additionally, we detected a point mutation in the gene *gyrA* (*gyrA*-S83Y) leading to resistance against (fluoro)quinolones (**Supplementary Table S4**). The remaining three strains from the 2010s dataset did not reveal any of these resistance genes. Two strains from the 1990s, (2947 and 2949) carry the genes *ant(3")-Ia*, *sul1* and *qacEdelta1* as well as the gene *aac(3)-VIa* (aminoglycoside resistance) but not the gene *tet(A)*. The chromosomally encoded gene *aac(6')-Iaa* was detected among all the samples included in this study.

We detected a specific gene pattern of virulence genes among the 15 MDR strains from the 2010s (**Table 3**). We named this pattern ESIv (Emergent S. Infantis virulence) which consists of genes associated with the fimbrial clusters: K88-like fimbria (Klf) and the *S.* Infantis plasmid-encoded fimbria (Ipf) (Aviv et al., 2017). Moreover, the pattern includes genes encoding for the virulent yersiniabactin operon (*fyu*A, *irp*1, *irp*2, *ybt*A, *ybt*E, *ybt*P, *ybt*Q, *ybt*S, *ybt*T, *ybt*U, *ybt*X) and the mercury (*mer)* operon (*mer*R, *mer*T, *mer*P, *mer*C, *mer*A, *mer*D, *mer*E) conferring mercury resistance. Finally, ESIv pattern consists of the gene complex *ccd*A/B and the *pem*K/I family (T/AT system). Isolates from the 1990s did not show any of these genes. Apart from the ESIv pattern, we found among all the samples from the study the presence of ten SPIs: SPIs-1-6, SPI-9, SPIs-11-12, and CS-54 (**Supplementary Table S5**).

As previously reported (Bogomazova et al., 2019), similar resistance and virulence gene patterns have been determined by the presence of a pESI-like plasmid carried by emergent *S.* Infantis strains. Likely, in the case of the strains of this study, the presence of a pESI-like plasmid could explain this genetic profile. Therefore, we aimed at the genomic detection and further characterization of this pESI-like plasmid among our strains.

Detection, Genomic Characterization and Phylogeny of a pESI-Like Plasmid

First, we scanned for replicon sequences of plasmids (**Supplementary Table S6**). All the 15 MDR strains positive to ESIr and ESIv patterns, presented the replicon IncFIB (pN55391) (**Table 2**). A genomic in-depth analysis for the typification of the plasmids revealed that the 15 strains positive for the IncFIB(pN55391) replicon, had the profile: *ardA*2, *pilL*3, *sogS*9, *trbA*21 while *repI*1 was absent. However, they were positive for the RepFIB replication protein A (*rep*B). Besides, they were positive for the detection of the sequence of the plasmid RK2 (from *E. coli*) DNA transposon (Tn1723) insertion sites (M20134) revealing the chimeric nature of the pESI-like plasmid as described previously (Aviv et al., 2014, 2016).

The samples 2947 and 2949 from the 1990s were positive for replicon IncI1-I(Gamma) and had the complete IncI gene profile: *ardA*4, *pilL*1, *sogS*2, *trbA*13, *repI*1. Sample 2949 simultaneously carries IncFIC(FII) and IncFII(pSFO) replicons and two alleles of IncF RST: FII91 and FIC3 (**Supplementary Table S6**). Interestingly, sample 3255 from the 2010s, was negative for the presence of IncFIB (pN55391); however, it carried additionally three different replicons: IncFIC(FII), IncFII(S) and IncFII(SARC14) for IncF plasmids (**Supplementary Table S6**). Sample 3255 did not present any gene from the IncI1 scheme, nor *rep*B, but two genes from the IncF RST scheme: FIIS5 and FIC3. The remaining non-resistants trains from the dataset did not contain any gene for an incompatibility group of plasmids.

Second, to add further evidence that the 2010s prevalence of *S.* Infantis in German broilers may be due to the presence of a pESI-like megaplasmid, we downloaded and examined the complete genome sequence of the megaplasmid pESI119944 found for the first time in the strain SI119944 in Israel (Aviv et al., 2014). Indeed, the analysis of the complete sequence of this plasmid showed the presence of the replicon IncFIB(pN553391), the allele profile of an IncI, the origin of replication of an IncP and *rep*B (**Supplementary Table S6**). As shown in **Tables 2**, **3**, main resistance and virulence traits detected among the German strains positive for replicon IncFIB(pN553391), were also present in the Israelian strain SI119944. Third, to have an in-depth comparison of the pESI-plasmid found within the German strains, we performed Oxford Nanopore sequencing of the sample 2747 that represent the samples from the 2010s and meets the characteristics described above regarding resistance, virulence genes, and plasmid nature. The hybrid assembly of sample 2747 resulted in two closed contigs corresponding to the chromosome (4,678,881 bp length) and the pESI-like plasmid (278,542 bp length) that we designated as pESI2747. The N50 value of the genome was 4,678,881 bp and the GC content was 52.18%. Kraken2 gave a match of 100% for *Salmonella enterica*. **Figures 2A,B** show the main features of the alignment of pESI119944 from Israel and pESI2747 from Germany. The alignment shows a consensus sequence of ∼285,184 bp. This consensus sequence consists of a common fragment of ∼277,693 bp (∼97.37%) between both sequences and a non-common region of ∼7,301 bp (∼2.56%). The *repB* gene coding for RepFIB replication protein A is located in the positions 80,555 and 81,580 bps in pESI2747, while on pESI119944 it is located at the beginning of the plasmid. Moreover, the genes *ardA*, *pilL*, *sogS*, *trbA* were found as well in different locations along pESI2747 compared to pESI119944. The sequence of an origin of replication for IncP plasmid was found as well in both plasmid sequences between the *mer* operon and the resistance genes *tet*(A) and *tet*(R) as previously described (Aviv et al., 2014, 2016). In pESI2747, we found the ESIr and the ESIv pattern common for the 15 MDR strains from the 2010s. Resistance genes *ant(3")-la*, *sul1* and *qacEdelta1* were located together in a region

TABLE 2 | Sequence type (ST), resistance genes (ESIr pattern and other) and plasmid replicons detected in broiler-derived *S*. Infantis strains from Germany and comparison with the complete genome sequence of plasmid pESI detected in *S*. Infantis SI119944 strain from Israel.

Sample	Year of isolation	MLST (ST)	Resistance genes pattern (ESIr)	Other resistance genes	Plasmid replicon
2945	1992	32	–	*aac(6′)-Iaa*	–
2947	1992	32	–	*aac(3)VIa*, *ant(3″)-Ia*, *sul1, aac(6′)-Iaa*	IncI1-I(Gamma)
2948	1992	32	–	*aac(6′)-Iaa*	–
2949	1992	32	–	*aac(3)VIa*, *ant(3″)-Ia*, *sul1, aac(6′)-Iaa*	IncI1-I(Gamma), IncFIC(FII), IncFII(pSFO)
2951	1994	32	–	*aac(6′)-Iaa*	–
3222	1994	32	–	*aac(6′)-Iaa*	–
3225	1995	32	–	*aac(6′)-Iaa*	–
3227	1995	1032	–	*aac(6′)-Iaa*	–
3230	1996	1032	–	*aac(6′)-Iaa*	–
3234	1997	32	–	*aac(6′)-Iaa*	–
3239	1998	32	–	*aac(6′)-Iaa*	–
3242	1998	32	–	*aac(6′)-Iaa*	–
3243	2014	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
3245	2014	32	–	*aac(6′)-Iaa*	–
3249	2015	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
3250	2015	32	–	*aac(6′)-Iaa*	–
3252	2015	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
3253	2016	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
3255	2016	32	–	*aac(6′)-Iaa*	–
3257	2016	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2954	2017	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2957	2018	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2959	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2747	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2748	2019	32	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2749	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2750	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2752	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
2753	2019	2283	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
3027	2020	32	+	*aac(6′)-Iaa, gyrA*-S83Y	IncFIB(pN55391)
SI119944	**2008**	**32**	**+**	***aac(6′)-Iaa, dfrA14, gyrA*-D87Y**	**IncFIB(pN55391)**

flanked by an IS6 transposase IS26, Integrase/recombinase (int) (Uniprot: P62592)[9] and IS21 family transposase IS1326. Resistance genes to TET *tet(A)* and *tet(R)* were found as well together, having upstream the Tn3 family transposase TnAs1. In the non-common part, we found the TMP resistance encoding gene *dfrA14* only presented in the sequence of pESI119944. Regarding virulence, we found the k88-like fimbria (Klf) cluster on the sequences of both plasmids spanning a region of ∼8,000 bp and the Ipf cluster that occupies a region of 5,100 bp. Between them, we found the gene cluster *ccd*A-*ccd*B encoding for the toxin/antitoxin system and the *vagC* and *vapC* genes. The *pemK-pemI* is located in position 95,086. Moreover, we found the 11 genes of the yersiniabactin operon spanning a region of ∼29,000 bp.

To study, if all German strains potentially carrying pESI have a similar plasmid structure as pESI2747, we mapped the Illumina short reads to the sequence of pESI119944 and analyzed their coverage vector (**Supplementary Figure S1**). We found that the positive strains cover practically the complete sequence of pESI119944 except for a gap at the end of the reference sequence suppose to be the non-common part observed between our pESI-like plasmid and the pESI119944. In summary, we add evidence that multidrug-resistant strains in the 2010s carry a pESI-like megaplasmid.

Finally, we were interested if the plasmid evolves independently or if there has occurred a co-evolution together with the chromosomes. Therefore, we studied the plasmid-based SNP phylogeny as previously performed (Alba et al., 2020) for the 15 *S.* Infantis strains harboring the pESI-like plasmid (**Supplementary Figure S2**). In general the plasmid phylogeny seems to be similar to the chromosomal phylogeny.

Genomic Characteristics of *S.* Infantis Strains From Israel, Hungary, and Italy

To compare our findings with recently published results from the emergent *S.* Infantis clones, we downloaded a total of 17 genomes from studies performed in Italy (Franco et al., 2015), Hungary (Olasz et al., 2015; Wilk et al.,

[9]https://www.uniprot.org/

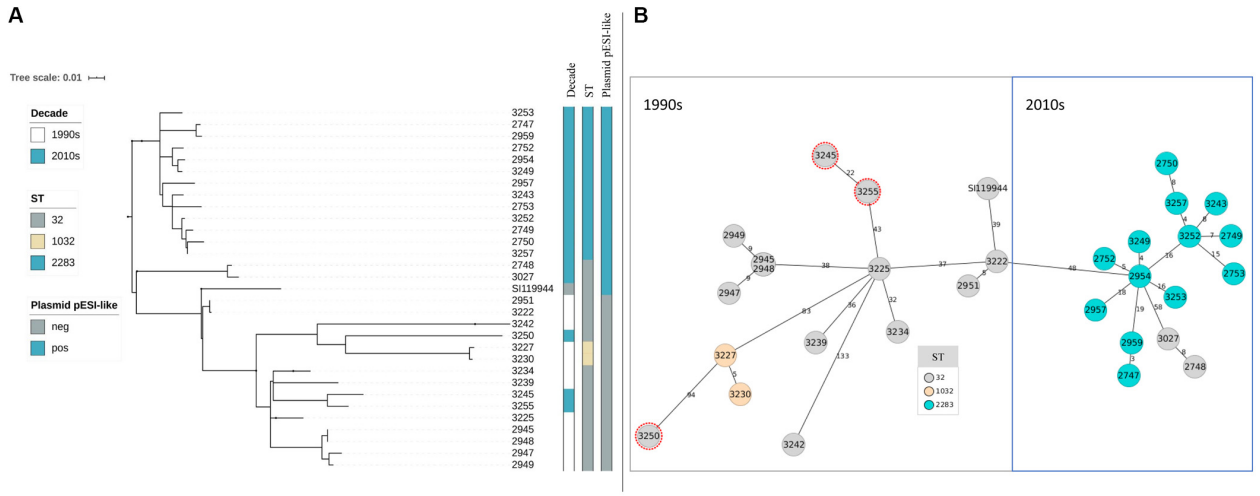


FIGURE 1 | (A,B) Phylogeny of broiler-derived *S.* Infantis strains in Germany and *S.* Infantis SI119944 from Israel. **(A)** Phylogenetic tree based on SNPs representing two main decade- and ST-related groups as the minimum spanning tree based on cgMLST. **(B)** Plasmid pESI-like is presented in the majority of samples from the 2010s (n = 15) and absent for the data from the 1990s. Samples 3227 and 3230 that belong to ST1032 are grouped within the 1990s group as they do not present the plasmid pESI-like. Note that samples 2748 and 3027 belong to ST32 but they are grouped with the samples collected in the 2010s **(B)** and that three strains from the 2010s are included within the group of the 1990s as they have ST32 (marked in red in **B**).

2016, 2017) and Israel (Aviv et al., 2014) as described above and analyzed them using our pipeline (**Supplementary Table S1**). Genotyping revealed that ST32 was the only sequence type of *S.* Infantis strains presented within the data from Israel, Hungary and Italy while ST2283 was not detected. PlasmidFinder found the replicon IncFIB(pN55391) in 13 out of 17 of the strains and the majority of them (n = 12) presented the IncI profile: *ardA*2, *pilL*3, *sogS*9, *trbA*21, and *repI*1 absent (**Supplementary Table S6**). Additionally, we detected two replicons (IncX1 and IncX3) in two Hungarian strains (SI240/16 and SI3337/12) and IncI1-I(Gamma) in one Italian strain (ERS846145) with IncI profile including the *repI*1. As expected, samples from Hungary SI69/04, Israel 335-3 and United Kingdom 1326/28 collected before the 2000s did contain neither plasmid replicons, nor IncI genes. Abricate using ResFinder, revealed a variety of 15 different resistance genes among the *S.* Infantis harboring the pESI-like plasmid (**Supplementary Table S4**). The ESIr [*ant(3")-la*, *sul1*, *tet(A)* and *qacEdelta1*] was found among 10 out of 17 strains from the three countries. The remaining strains had a variation of this pattern and presented other additional genes. Especially, the Italian strains present a wide variety of resistance genes including resistance genes related to extended-spectrum β-lactamase (ESBL), like *blaCTX-M-1* as reported previously (Franco et al., 2015). Three out of seven Hungarian strains presented ESBL genes as well like the *blaCTX-M-14* or *blaTEM-104*.

Chromosomal mutations of gene *gyr*A were found as well: *gyrA*-S83Y (8 out of 17) and *gyrA*-D87G (4 out of 17), *gyrA*_D87Y was found only in SI119944. The ESIv pattern described above within the German *S.* Infantis was found in 12 out of 16 strains (**Supplementary Table S5**). Therefore, the results from this study are in line with the findings in other European countries regarding the emergence of multidrug and virulent *S.* Infantis clones.

To see if there is a clonal transmission of the strains, a minimum spanning tree based on cgMLST was constructed (**Supplementary Figure S3**). German *S.* Infantis strains of ST2283 form 2010s are close to two Hungarian strains that were reported as new *S.* Infantis clones (Nógrády et al., 2012). The smallest difference between external strains and German pESI-like strains is 37 alleles between the Hungarian SI54/04 and the German 2954, therefore we could not detect any clonal relatedness. The two ST32 strains from the 2010s are most closely related to two emergent *S.* Infantis strains from Italy. We could detect clonal transmission between two Italian strains where the smallest number of different alleles was 1.

DISCUSSION

Studies performed within and outside Europe revealed an emergent dissemination of *S.* Infantis clones in humans and several animal species (Nógrády et al., 2007, 2008, 2012; Aviv et al., 2014; Franco et al., 2015; Olasz et al., 2015; Yokoyama et al., 2015; Wilk et al., 2016; Hindermann et al., 2017; Szmolka et al., 2018; Acar et al., 2019; Bogomazova et al., 2019). There is evidence that the acquisition of a conjugative megaplasmid provides the bacteria with new resistance properties which might have contributed to the increased occurrence of this serovar. An in-depth analysis of the genetic characteristics of the plasmid called pESI (Aviv et al., 2014) or similar pESI-like plasmids (Franco et al., 2015; Szmolka et al., 2018) revealed that they also encode for virulence-associated characteristics, resistance to heavy metals or disinfectants and fitness characteristics. The

TABLE 3 | Virulence and fitness genes detected in broiler-derived *S*. Infantis strains from Germany (ESIv pattern) and comparison with the complete genome sequence of plasmid pESI detected in *S.* Infantis SI119944 strain from Israel.

Sample	Year of isolation	K88-like fimbria (Klf)	Infantis plasmid encoded fimbria (Ipf)	*mer* operon	Yersiniabactin system	Toxin/antitoxin system (T/AT)
2945	1992	–	–	–	–	–
2947	1992	–	–	–	–	–
2948	1992	–	–	–	–	–
2949	1992	–	–	–	–	–
2951	1994	–	–	–	–	–
3222	1994	–	–	–	–	–
3225	1995	–	–	–	–	–
3227	1995	–	–	–	–	–
3230	1996	–	–	–	–	–
3234	1997	–	–	–	–	–
3239	1998	–	–	–	–	–
3242	1998	–	–	–	–	–
3243	2014	+	+	+	+	+
3245	2014	–	–	–	–	–
3249	2015	+	+	+	+	+
3250	2015	–	–	–	–	–
3252	2015	+	+	+	+	+
3253	2016	+	+	+	+	+
3255	2016	–	–	+	–	–
3257	2016	+	+	+	+	+
2954	2017	+	+	+	+	+
2957	2018	+	+	+	+	+
2959	2019	+	+	+	+	+
2747	2019	+	+	+	+	+
2748	2019	+	+	+	+	+
2749	2019	+	+	+	+	+
2750	2019	+	+	+	+	+
2752	2019	+	+	+	+	+
2753	2019	+	+	+	+	+
3027	2020	+	+	+	+	+
SI119944	**2008**	**+**	**+**	**+**	**+**	**+**

analysis and genomic comparison of the complete genome sequence of SI119944 from Israel (Cohen et al., 2020) and strain 2747 from Germany demonstrate and confirm the characteristic resistance, virulent as well as fitness traits encoded on a pESI-like plasmid. Furthermore, results from this study confirm the observed switch in the occurrence of *S.* Infantis organisms in broilers from non-MDR strains screened until the 2000s (Asai et al., 2007; Shahada et al., 2008; Wilk et al., 2017) to the emergence of MDR clones collected during and after the end of the 2000s (Nógrády et al., 2007, 2008, 2012). ST32 is a highly conserved sequence type of *S.* Infantis (Monte et al., 2019) and was the dominating MLST type isolated from various and numerous sources (broilers, pigs, cattle, food, human) in different European countries (Hauser et al., 2012; Hindermann et al., 2017; Gymoese et al., 2019). In this study, we describe the occurrence of two single locus variants of ST32 within the *S.* Infantis from German broiler production: ST2283 and ST1032. The emergence of ST2283 in *S.* Infantis organisms from the 2010s is linked with the presence of a pESI-like plasmid that confers a MDR pattern that has not been found among *S.* Infantis ST32 strains originating from the 1990s. However, we also identified two ST32 strains of *S.* Infantis from the 2010s harboring the resistant coding plasmid. We observed general concordance in cluster separation between the chromosome-based tree and plasmid-based tree in all strains harboring pESI-like megaplasmid (ST32 and ST2283) as shown also by Alba et al. (2020). Therefore, we hypothesize that the plasmid has co-evolved with the chromosome and both STs gained the plasmid in two (or more) evolutionary independent events. However, a rather rare occurrence of MDR clone ST32 compared with the higher prevalence of MDR clone ST2283 in recent years indicates an obvious greater tendency of dissemination of this clone in Germany and perhaps European broiler production, and therefore, another until now not detected property of ST2283. We also detected two non-resistant ST1032 clones from the 1990s. This variant of ST32 had been described before in a non-resistant isolate from food (Alba et al., 2020). In this study, bioinformatics analysis revealed a correlation between

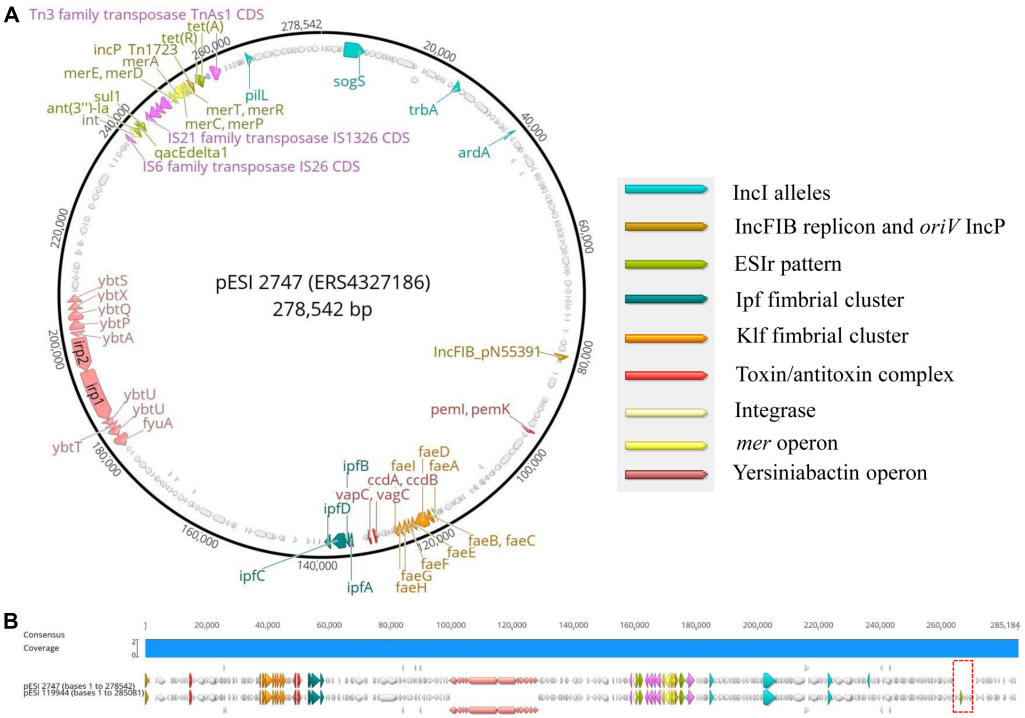


FIGURE 2 | (A,B) Main genomic traits and alignment of the complete plasmid sequence of pESI119944 from the Israelian strain SI119944 and pESI2747 from the German strain 19PM0148. The red frame indicates the non-common fragment in the alignment.

the resistance gene pattern named as ESIr [*ant(3")-la, sul1, tet(A)*] and the phenotypic resistance profile to aminoglycosides, sulfonamides, and TETs. Furthermore, gen *aac(6')-Iaa* was not located on pESI119944 but on the chromosome. In line with the literature, we found the gene *aac(6')-Iaa* to be chromosomal-encoded gene (Salipante and Hall, 2003). On the other hand, German broiler derived *S.* Infantis strains showed phenotypic resistance to quinolones like CIP and NAL. Genotype findings do correlate with the phenotypic results as we detected the well-studied mutation in *gyrA* gene that codifies for resistance to those antibiotics (Chen et al., 2019). The predominant MDR pattern found among the emergent *S.* Infantis clones from Europe consists mainly of antimicrobials that belong to the major classes of antibiotics FOT, CIP, cephalosporin, TET, sulfonamide, fluoroquinolone, and TMP (Cloeckaert et al., 2007; Franco et al., 2015; Acar et al., 2019). Phenotypic and genotypic variants of this pattern have been observed in Hungary related to two different pulsotype clusters (Nógrády et al., 2012). Variants of this MDR pattern including ESBL resistant isolates of *S.* Infantis were also found in Italy and Germany (Franco et al., 2015; Fischer et al., 2017). Resistance and virulence traits coevolved and interfere in the ecology of a strain (Beceiro et al., 2013). Thus, the increasing emergence of a strain is not only dependent on antimicrobial resistance, but also on virulence, bacteriocin secretion, biocide resistance and, biofilm formation (Acar et al., 2019). Consequently, in this study, the bioinformatics analysis for virulence determinants showed a common pattern in virulence and fitness genes within the MDR isolates. *Salmonella* pathogenicity islands and other different gene complexes that encode for fimbriae production, adherent and non-adherent products, as well as curli structures, are of special interest because of their involvement in host colonization, persistence, motility, and invasion (Barnhart and Chapman, 2006; Rychlik et al., 2009; Aviv et al., 2017). The strong dissemination of *S.* Infantis not only in broilers during the last two decades wonders whether the increased antimicrobial resistance, the swift in MLST type, the virulence properties, the capability of biofilm formation or other unknown factors are responsible for this emergence. Different hypotheses try to explain this phenomenon. For example, it is suspected that the increased prevalence of *S.* Infantis could be due to the general decreased prevalence of *S.* Enteritidis in poultry farms (Szmolka et al., 2018). It is also suggested that the EU trade of broiler chicken and the pyramidal structure of the poultry industry may be factors of the rapid spread of emergent clones of *S.* Infantis carrying the pESI-like plasmid beyond national borders (Alba et al., 2020; Nagy et al., 2020). The long term use of special groups of antimicrobial substances might have resulted in selection pressure and increased emergence of particular bacterial organisms (Nógrády et al., 2007, 2012). However, it is also stated that antimicrobial usage is not always linked to a higher *Salmonella* prevalence (Asai et al., 2007). The acquisition of the megaplasmid pESI does not result in a significant burden to its hosts as it is presented only as a single copy in the bacteria genome, therefore, it does not seem to limit the dissemination of the organisms (Aviv et al., 2016).

Production of fimbriae and the ability to form biofilms are discussed as factors enabling a long term persistence of *Salmonella* organisms at poultry farms. The gene *fyuA* (Schubert et al., 1998) together with the genes *irp1* and *irp2* are involved in biofilm formation (Hancock et al., 2008). In this study, gene *irp1* was found together with *irp2* in all strains that carry the pESI-like plasmid suggesting a possible role in persistence of *S.* Infantis organisms. However, the association of yersiniabactin and biofilm in serovar Infantis has not been yet wide studied in contrast with other microorganisms as in Uropathogenic *Escherichia coli* (UPEC) (Zamani and Salehzadeh, 2018). It has been demonstrated before a higher biofilm formation for pESI positive strains (Aviv et al., 2014). Besides, very recently, it has been demonstrated a higher cell adhesion of *S.* Infantis compared with other serovars. Resistance to heavy metals or biocides like QACs might also play a role in the emergence of *S.* Infantis. The gene *qacEdelta1* (Chuanchuen et al., 2007), located on pESI-like megaplasmids codes for resistance against QACs and was found in this study exclusively in strains from the 2010s. On the other hand, detailed analysis of the sequence of pESI-like pESI2747 has revealed not only resistance genes but also virulence genes, toxin/antitoxin systems that as previously described (Aviv et al., 2014; Acar et al., 2017) play a clue role in the emergence of *S.* Infantis in poultry. However, it is open whether the encoded resistance against disinfectants might contribute to a higher *S.* Infantis persistence at broiler farms.

In conclusion, broiler derived *S.* Infantis strains of ST2283 in Germany show similarities to emergent *S.* Infantis strains from Europe including the possession of the pESI-like megaplasmid which encodes for antimicrobial resistance, virulence genes (fimbrial clusters) and fitness determinants (toxin/antitoxin system) that enhance bacterial adaptability. Therefore, the involvement of the megaplasmid might explain the current spread of these emergent *S.* Infantis organisms. However, specific reasons for a suspected higher persistence of MLST2283 could not be identified in this study. It seems that *S.* Infantis persistence in broiler farms is caused by its occurrence in the primary broiler production and ineffective cleaning and disinfection protocols at least for these special clones. Epidemiological studies on the occurrence of *S.* Infantis ST2283 in the whole broiler production chain and the establishment of effective control measures are essential to prevent these organisms from entering the food chain.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. Sequencing data of this study can be found at https://www.ebi.ac.uk/ena/data/view/PRJEB36784. The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

UM conceived and coordinated the study, performed serotyping and MIC determination, and wrote the manuscript. JL coordinated the study, developed and implemented the bioinformatics pipeline, and wrote the manuscript. SG-S performed the bioinformatics analysis, performed in-depth analysis of the plasmid, and wrote the manuscript. MA-G performed MinION sequencing and data analysis. HT critically read the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We sincerely thank A. Hackbart and S. Keiling for their excellent technical assistance. We thank Dr. I. Szabo (Head of National Reference Laboratory for *Salmonella* at the German Federal Institute for Risk Assessment, BfR) for the kind provision of *S.* Infantis strains. We also thank Dr. Belén Gonzalez-Santamarina for her suggestions regarding the analysis of SPIs and resistance genes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01741/full#supplementary-material

FIGURE S1 | Mapping coverage of all pESI-like positive strains across the complete genome sequence of plasmid pESI119944.

FIGURE S2 | Tanglegram plot representing the comparison between pESI-like positive strains using the chromosome sequence of *S.* Infantis ESI119944 (CP047881) (left) as reference and the plasmid pESI119944 sequence (CP047882) as reference (right).

FIGURE S3 | Minimum spanning tree for previously reported emergent *S.* Infantis clones from Israel (Aviv et al., 2014, 2016), Hungary (Olasz et al., 2015; Wilk et al., 2016, 2017), and Italy (Franco et al., 2015) used in this study based on cgMLST.

TABLE S1 | *S.* Infantis genomes for previously reported emergent *S.* Infantis clones from Israel (Aviv et al., 2014, 2016), Hungary (Olasz et al., 2015; Wilk et al., 2016, 2017), and Italy (Franco et al., 2015) used in this study and their general genome characteristics and genoserotyping.

TABLE S2 | Specific gene sequences downloaded for the creation of customized databases for Abricate and available databases downloaded. These databases include sequences of genes encoding for Salmonella pathogenicity islands (SPIs), fimbrial clusters lpf and K88-like and pESI encoded fitness determinants as the toxin-antitoxin (T/AT) system (CcdAB and PemKI/MazF), mercury operon and allele sequences for incompatibility groups plasmids IncI1 and IncF for plasmid typing.

TABLE S3 | Whole Genome Sequencing general characteristics and genoserotyping of the *S.* Infantis strains from Germany used in this study.

TABLE S4 | Resistance genes and chromosomal point mutations found among the *S.* Infantis strains used in this study.

TABLE S5 | Virulence genes, SPIs, fimbrial cluster genes and fitness genes found among the *S.* Infantis strains used in this study.

TABLE S6 | Plasmid replicons, plasmid pMLST typing and genes encoding for plasmid origin of replication found among the *S.* Infantis strains used in this study.

REFERENCES

Acar, S., Bulut, E., Durul, B., Uner, I., Kur, M., Avsaroglu, M. D., et al. (2017). Phenotyping and genetic characterization of *Salmonella enterica* isolates from Turkey revealing arise of different features specific to geography. *Int. J. Food Microbiol.* 241, 98–107. doi: 10.1016/j.ijfoodmicro.2016.09.031

Acar, S., Bulut, E., Stasiewicz, M. J., and Soyer, Y. (2019). Genome analysis of antimicrobial resistance, virulence, and plasmid presence in Turkish *Salmonella* serovar Infantis isolates. *Int. J. Food Microbiol.* 307:108275. doi: 10.1016/j.ijfoodmicro.2019.108275

Alba, P., Leekitcharoenphon, P., Carfora, V., Amoruso, R., Cordaro, G., Di Matteo, P., et al. (2020). Molecular epidemiology of *Salmonella* Infantis in Europe: insights into the success of the bacterial host and its parasitic pESI-like megaplasmid. *Microb. Genom* 6. doi: 10.1099/mgen.0.000365 [Epub ahead of print].

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

Andrews, S. (2018). *FastQC: A Quality Control Tool for High Throughput Sequence Data". v. 0.11*, 5 Edn. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed February, 2020).

Asai, T., Ishihara, K., Harada, K., Kojima, A., Tamura, Y., Sato, S., et al. (2007). Long-term prevalence of antimicrobial-resistant *Salmonella enterica* subspecies enterica Serovar Infantis in the broiler chicken industry in Japan. *Microbiol. Immunol.* 51, 111–115. doi: 10.1111/j.1348-0421.2007.tb03881.x

Aviv, G., Elpers, L., Mikhlin, S., Cohen, H., Vitman Zilber, S., Grassl, G. A., et al. (2017). The plasmid-encoded Ipf and Klf fimbriae display different expression and varying roles in the virulence of *Salmonella enterica* serovar Infantis in mouse vs. avian hosts. *PLoS Pathog.* 13:e1006559. doi: 10.1371/journal.ppat.1006559

Aviv, G., Tsyba, K., Steck, N., Salmon-Divon, M., Cornelius, A., Rahav, G., et al. (2014). A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. *Environ. Microbiol.* 16, 977–994. doi: 10.1111/1462-2920.12351

Aviv, G., Rahav, G., and Gal-Mor, O. (2016). Horizontal transfer of the *Salmonella enterica* serovar infantis resistance and virulence plasmid pESI to the gut microbiota of warm-blooded hosts. *mBio* 7:e01395-16. doi: 10.1128/mBio.01395-16

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barnhart, M. M., and Chapman, M. R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.* 60, 131–147. doi: 10.1146/annurev.micro.60.080805.142106

Beceiro, A., Tomas, M., and Bou, G. (2013). Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* 26, 185–230. doi: 10.1128/CMR.00059-12

Bogomazova, A. N., Gordeeva, V. D., Krylova, E. V., Soltynskaya, I. V., Davydova, E. E., Ivanova, O. E., et al. (2019). Mega-plasmid found worldwide confers multiple antimicrobial resistance in *Salmonella* Infantis of broiler origin in Russia. *Int. J. Food Microbiol.* 319:108497. doi: 10.1016/j.ijfoodmicro.2019.108497

Carattoli, A., and Hasman, H. (2020). PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). *Methods Mol. Biol.* 2075, 285–294. doi: 10.1007/978-1-4939-9877-7_20

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Chen, K., Dong, N., Chan, E. W., and Chen, S. (2019). Transmission of ciprofloxacin resistance in *Salmonella* mediated by a novel type of conjugative helper plasmids. *Emerg. Microbes Infect.* 8, 857–865. doi: 10.1080/22221751.2019.1626197

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi: 10.1093/nar/gki008

Chuanchuen, R., Khemtong, S., and Padungtod, P. (2007). Occurrence of qacE/qacEDelta1 genes and their correlation with class 1 integrons in *Salmonella enterica* isolates from poultry and swine. *Southeast Asian J. Trop. Med. Public Health* 38, 855–862.

Cloeckaert, A., Praud, K., Doublet, B., Bertini, A., Carattoli, A., Butaye, P., et al. (2007). Dissemination of an extended-spectrum-beta-lactamase blaTEM-52 gene-carrying IncI1 plasmid in various *Salmonella enterica* serovars isolated from poultry and humans in Belgium and France between 2001 and 2005. *Antimicrob. Agents Chemother.* 51, 1872–1875. doi: 10.1128/AAC.01514-06

Cohen, E., Rahav, G., and Gal-Mor, O. (2020). Genome Sequence of an Emerging *Salmonella enterica* Serovar Infantis and Genomic Comparison with Other S. Infantis Strains. *Genome Biol. Evol.* 12, 151–159. doi: 10.1093/gbe/evaa048

EFSA, (2019). The European Union One Health 2018 zoonoses report. *EFSA J.* 17:e05926. doi: 10.2903/j.efsa.2019.5926

EFSA, (2020). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2017/2018. *EFSA J.* 18:e06007. doi: 10.2903/j.efsa.2020.6007

EUCAST, (2018). *The European Committee on Antimicrobial Susceptibility Testing. Breakpoint Tables for Interpretation of MICs and Zone Diameters, Version 8.0.*

Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019a). Using the NCBI AMRFinder tool to determine antimicrobial resistance genotype-phenotype correlations within a collection of NARMS Isolates. *bioRxiv* [Preprint]. doi: 10.1101/550707

Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019b). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of Isolates. *Antimicrob. Agents Chemother.* 63:e00483-419. doi: 10.1128/AAC.00483-419

Fischer, J., Borowiak, M., Baumann, B., Szabo, I., and Malorny, B. (2017). "Whole-genome sequencing analysis of multidrug-resistant *Salmonella* Infantis isolatescirculating in the German food-production chain," in *27th ECCMID*, (Vienna).

FLI_Bioinfo (2020). *WGSBAC. Modules for Genotyping and Characterization of Bacterial Isolates Using Whole-Genome-Sequencing Data". v2.0*, 0 Edn. Available online at: https://gitlab.com/FLI_Bioinfo/WGSBAC (accessed February, 2020).

Franco, A., Leekitcharoenphon, P., Feltrin, F., Alba, P., Cordaro, G., Iurescia, M., et al. (2015). Emergence of a clonal lineage of multidrug-resistant ESBL-producing *Salmonella* Infantis transmitted from broilers and broiler meat to humans in Italy between 2011 and 2014. *PLoS One* 10:e0144802. doi: 10.1371/journal.pone.0144802

Grimont, P., and Weill, F.-X. (2007). *Antigenic Formulae-Grimont-Weill.pdf. WHO Collaborating Center for Reference and Research on Salmonella*, 9th Edn. Paris: Institut Pasteur.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Gymoese, P., Kiil, K., Torpdahl, M., Osterlund, M. T., Sorensen, G., Olsen, J. E., et al. (2019). WGS based study of the population structure of *Salmonella enterica* serovar Infantis. *BMC Genomics* 20:870. doi: 10.1186/s12864-019-6260-6

Hancock, V., Ferrieres, L., and Klemm, P. (2008). The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology* 154(Pt 1), 167–175. doi: 10.1099/mic.0.2007/011981-0

Hauser, E., Tietze, E., Helmuth, R., Junker, E., Prager, R., Schroeter, A., et al. (2012). Clonal dissemination of *Salmonella enterica* serovar Infantis in Germany. *Foodborne Pathog. Dis.* 9, 352–360. doi: 10.1089/fpd.2011.1038

Hindermann, D., Gopinath, G., Chase, H., Negrete, F., Althaus, D., Zurfluh, K., et al. (2017). *Salmonella enterica* serovar infantis from food and human infections, Switzerland, 2010–2015: poultry-related multidrug resistant clones and an emerging ESBL producing clonal lineage. *Front. Microbiol.* 8:1322. doi: 10.3389/fmicb.2017.01322

Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their

applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Lindqvist, N., and Pelkonen, S. (2007). Genetic surveillance of endemic bovine *Salmonella* Infantis infection. *Acta Vet. Scand.* 49:15. doi: 10.1186/1751-0147-49-15

Monte, D. F., Lincopan, N., Berman, H., Cerdeira, L., Keelara, S., Thakur, S., et al. (2019). Genomic features of high-priority *Salmonella enterica* serovars circulating in the food production Chain, Brazil, 2000-2016. *Sci. Rep.* 9:11058. doi: 10.1038/s41598-019-45838-0

Nagy, T., Szmolka, A., Wilk, T., Kiss, J., Szabo, M., Paszti, J., et al. (2020). Comparative genome analysis of hungarian and global Strains of *Salmonella* Infantis. *Front. Microbiol.* 11:539. doi: 10.3389/fmicb.2020.00539

Nógrády, N., Kardos, G., Bistyak, A., Turcsanyi, I., Meszaros, J., Galantai, Z., et al. (2008). Prevalence and characterization of *Salmonella* infantis isolates originating from different points of the broiler chicken-human food chain in Hungary. *Int. J. Food Microbiol.* 127, 162–167. doi: 10.1016/j.ijfoodmicro.2008.07.005

Nógrády, N., Kiraly, M., Davies, R., and Nagy, B. (2012). Multidrug resistant clones of *Salmonella* Infantis of broiler origin in Europe. *Int. J. Food Microbiol.* 157, 108–112. doi: 10.1016/j.ijfoodmicro.2012.04.007

Nógrády, N., Toth, A., Kostyak, A., Paszti, J., and Nagy, B. (2007). Emergence of multidrug-resistant clones of *Salmonella* Infantis in broiler chickens and humans in Hungary. *J. Antimicrob. Chemother.* 60, 645–648. doi: 10.1093/jac/dkm249

Olasz, F., Nagy, T., Szabo, M., Kiss, J., Szmolka, A., Barta, E., et al. (2015). Genome sequences of three *Salmonella enterica* subsp. enterica serovar infantis strains from healthy broiler chicks in hungary and in the United Kingdom. *Genome Announc.* 3:e01468-14. doi: 10.1128/genomeA.01468-14

Pate, M., Micunovic, J., Golob, M., Vestby, L. K., and Ocepek, M. (2019). *Salmonella* infantis in broiler flocks in slovenia: the prevalence of multidrug resistant strains with high genetic homogeneity and low biofilm-forming ability. *Biomed. Res. Int.* 2019:4981463. doi: 10.1155/2019/4981463

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Rajic, A., Keenliside, J., McFall, M. E., Deckert, A. E., Muckle, A. C., O'Connor, B. P., et al. (2005). Longitudinal study of *Salmonella* species in 90 Alberta swine finishing farms. *Vet. Microbiol.* 105, 47–56. doi: 10.1016/j.vetmic.2004.10.005

Ross, I. L., and Heuzenroeder, M. W. (2008). A comparison of three molecular typing methods for the discrimination of *Salmonella enterica* serovar Infantis. *FEMS Immunol. Med. Microbiol.* 53, 375–384. doi: 10.1111/j.1574-695X.2008.00435.x

Rychlik, I., Karasova, D., Sebkova, A., Volf, J., Sisak, F., Havlickova, H., et al. (2009). Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. *BMC Microbiol.* 9:268. doi: 10.1186/1471-2180-9-268

Salipante, S. J., and Hall, B. G. (2003). Determining the limits of the evolutionary potential of an antibiotic resistance gene. *Mol. Biol. Evol.* 20, 653–659. doi: 10.1093/molbev/msg074

Schubert, S., Rakin, A., Karch, H., Carniel, E., and Heesemann, J. (1998). Prevalence of the "High-Pathogenicity Island" of Yersinia Species among *Escherichia coli* Strains That Are Pathogenic to Humans. *Infect. Immun.* 66, 480–485. doi: 10.1128/iai.66.2.480-485.1998

Seemann, T. (2014a). *mlst GitHub*. Available online at: https://github.com/tseemann/mlst (accessed February, 2020).

Seemann, T. (2014b). *Snippy GitHub*. Available online at: https://github.com/tseemann/snippy (accessed February, 2020).

Seemann, T. (2015). *Abricate Github*. Available online at: https://github.com/tseemann/abricate (accessed February, 2020).

Seemann, T. (2018). *Shovill GitHub. Assemble Bacterial Isolate Genomes From Illumina Paired-End Reads*. Available online at: https://github.com/tseemann/shovill (accessed February, 2020).

Shahada, F., Chuma, T., Okamoto, K., and Sueyoshi, M. (2008). Temporal distribution and genetic fingerprinting of *Salmonella* in broiler flocks from southern Japan. *Poul. Sci.* 87, 968–972. doi: 10.3382/ps.2007-00455

Szmolka, A., Szabo, M., Kiss, J., Paszti, J., Adrian, E., Olasz, F., et al. (2018). Molecular epidemiology of the endemic multiresistance plasmid pSI54/04 of *Salmonella* Infantis in broiler and human population in Hungary. *Food Microbiol.* 71, 25–31. doi: 10.1016/j.fm.2017.03.011

Vaser, R., Sovic, I., Nagarajan, N., and Sikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wilk, T., Szabo, M., Szmolka, A., Kiss, J., Barta, E., Nagy, T., et al. (2016). Genome sequences of multidrug-resistant *Salmonella enterica* subsp. enterica serovar infantis strains from broiler chicks in hungary. *Genome Announc.* 4:e01400-16. doi: 10.1128/genomeA.01400-16

Wilk, T., Szabo, M., Szmolka, A., Kiss, J., Olasz, F., and Nagy, B. (2017). Genome sequences of *Salmonella enterica* subsp. enterica serovar infantis strains from hungary representing two peak incidence periods in three decades. *Genome Announc.* 5:e01735-16. doi: 10.1128/genomeA.01735-16

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0

Yokoyama, E., Ando, N., Ohta, T., Kanada, A., Shiwa, Y., Ishige, T., et al. (2015). A novel subpopulation of *Salmonella enterica* serovar Infantis strains isolated from broiler chicken organs other than the gastrointestinal tract. *Vet. Microbiol.* 175, 312–318. doi: 10.1016/j.vetmic.2014.11.024

Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P., Nash, J. H., et al. (2016). The *Salmonella* in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping Draft *Salmonella* genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101

Zamani, H., and Salehzadeh, A. (2018). Biofilm formation in uropathogenic *Escherichia coli*: association with adhesion factor genes. *Turk J. Med. Sci.* 48, 162–167. doi: 10.3906/sag-1707-03

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

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.

3.1. Supplementary figures from the Publication

Figure S1. Mapping coverage of all pESI-like positive strains across the complete genome sequence of plasmid pESI119944.

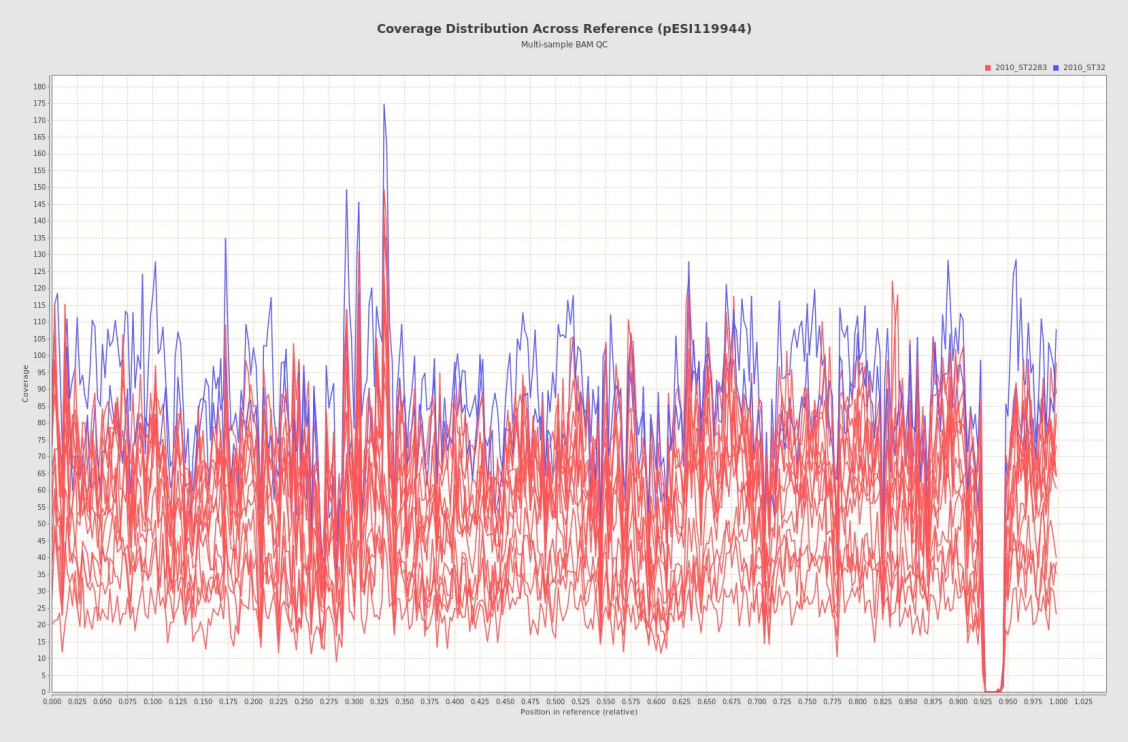


Figure S2. Tanglegram plot representing the comparison between pESI-like positive strains using the chromosome sequence of *S*. Infantis ESI119944 (CP047881) (left) as reference and the plasmid pESI119944 sequence (CP047882) as reference (right).

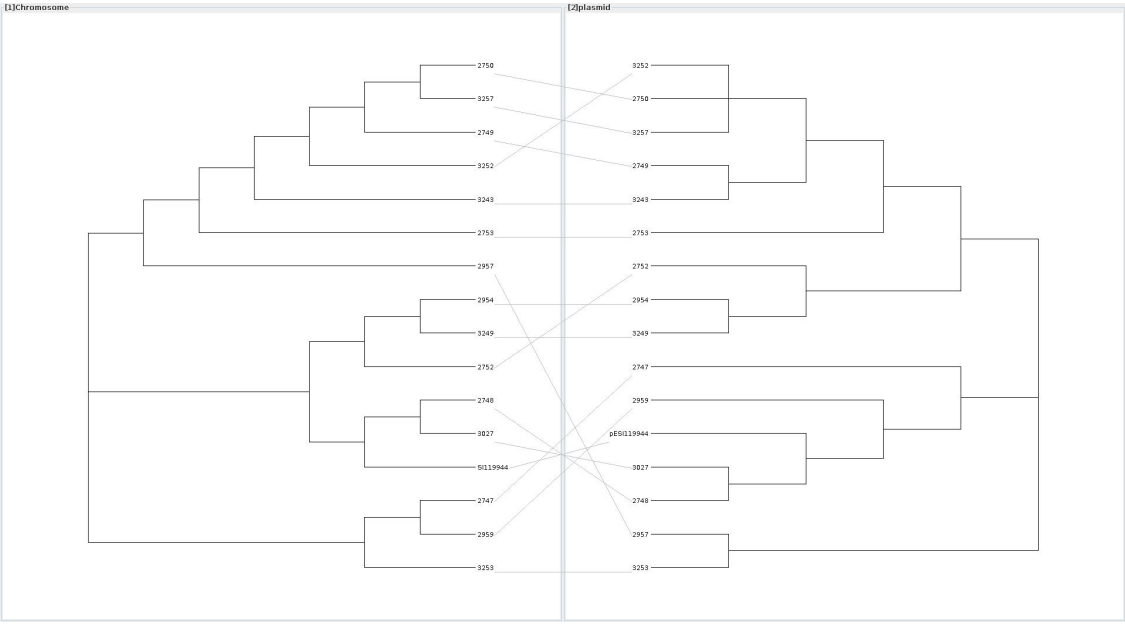
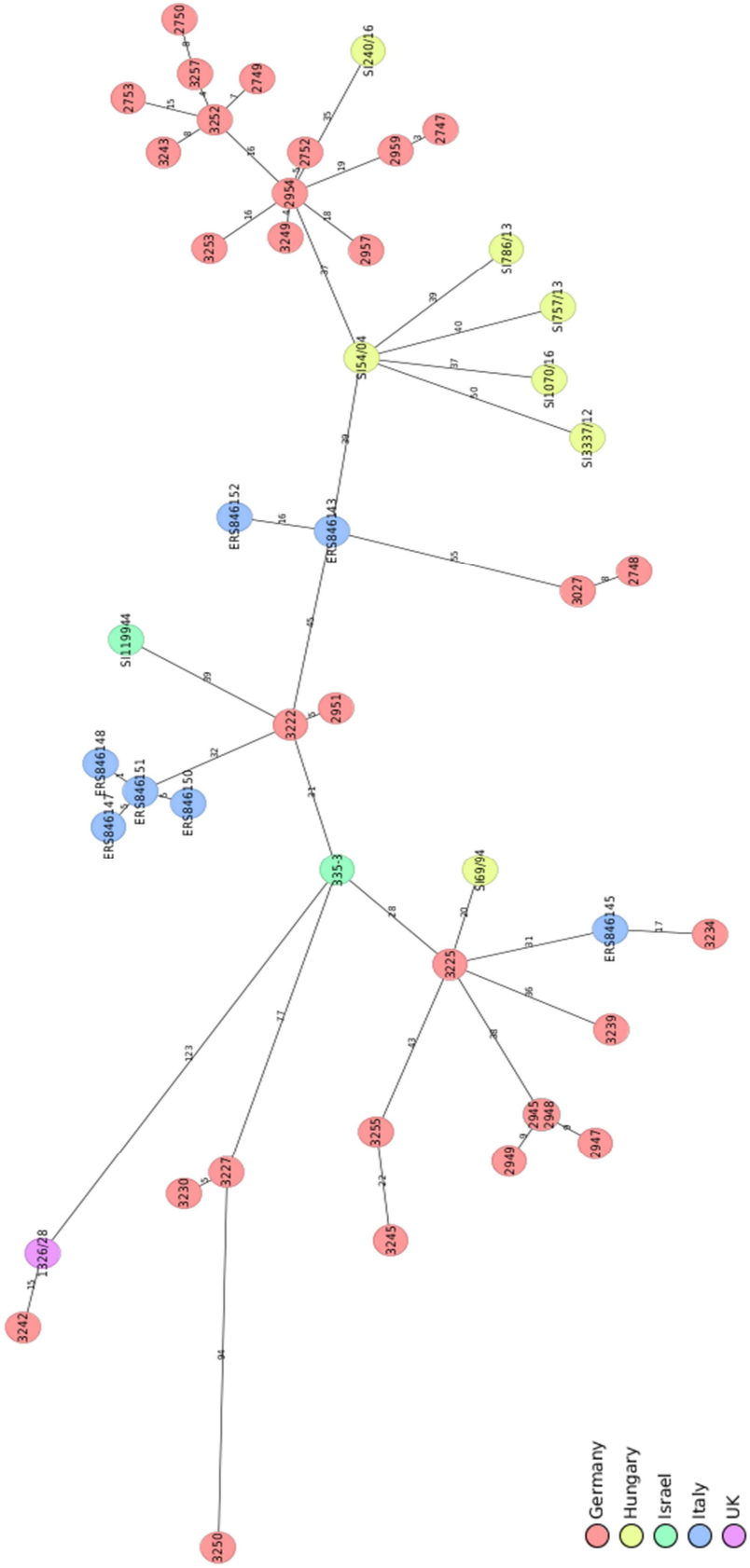


Figure S3. Minimum spanning tree for previously reported emergent *S.* Infantis clones from Israel, Hungary, and Italy used in this study based on cgMLST.



4. DISCUSSION

The work presented in this thesis consisted of two parts: i) implementation and evaluation of an in–house bioinformatics pipeline (WGSBAC) for *Salmonella in silico* serotyping and ii) the application of the WGSBAC pipeline to study the potential occurrence and high tendency of dissemination of a MDR emergent *S*. Infantis population in German broiler farms.

4.1. Implementation and evaluation of an in–house bioinformatics pipeline (WGSBAC) for *Salmonella in silico* serotyping

Serotyping is an essential step in the investigation of outbreaks concerning *Salmonella*. It was proposed as a classification method for *Salmonella* genus in 1975 (KAUFFMANN 1975) and still today, it is considered the gold-standard for *Salmonella* typing (GRIMONT et al. 2007). However, the successful performance of WGS-based serotyping tools raises the question of whether the use of conventional *Salmonella* serotyping based on agglutination assays is coming to an end (ZHANG et al. 2015, ASHTON et al. 2016, YOSHIDA et al. 2016, YACHISON et al. 2017, RENE et al. 2018, ROBERTSON et al. 2018, ZHANG et al. 2019, UELZE et al. 2020). The list of disadvantages of conventional *Salmonella* serotyping includes among others low throughput, long turnaround times, demand of expertise (BOXRUD 2010), and production and quality control of a large number (more than 250) of antisera from immunized rabbits (MCQUISTON et al. 2004). Several molecular-based methods that aimed to replace antisera agglutination-based serotyping have been proposed in the last years (TANG et al. 2019). Among the molecular-based methods, PFGE and MLVA have shown useful results for subtyping in outbreak investigation, however, their use for serovar prediction supposes often a challenge in the case of polyphyletic serovars (e.g. Newport, Paratyphi B, or Kentucky serovars) and the standardization of profiles and protocols for inter and external-laboratory comparisons are required (TANG et al. 2019). Other attempts of molecular-based serotyping methods are rep-PCR, ribotyping, or MLST however, they have difficulties reflecting genetic relatedness between different serovars (XU et al. 2020). Recently, a genoserotyping system based on Multiplex Oligonucleotide Ligation – PCR (MOL-PCR) has been developed for the determination of specific invasive serovars in the poultry and pork sector (GAND et al. 2020).

To evaluate the performance of three different *Salmonella* genoserotyping tools included in the previous version of WGSBAC (v. 2.0.0), we sequenced the genomes of 43 *Salmonella* strains of 26 different serovars. They were selected based on specific characteristics such as

their dificulty to be resolved by conventional serotyping (e.g. “rough” form isolates). Serovar prediction was carried out by three bioinformatics tools (SISTR, SeqSero, and SeqSero2) that were integrated within the previous version (v. 2.0.0) of the WGSBAC pipeline. The analysis concluded with the examination of the correlation between the serovar determination by slide agglutination test and the serovar prediction from *in silico* serotyping. The three tools performed the alignment of the sequenced query data to a curated database of O and H antigen allele reference sequences and the consequent assignment of the antigenic formula and/or the serovar name. SISTR works on assembled sequences by determination of the antigenic genes and the examination of cgMLST alleles and combines both approaches to report accurately the serovar (YOSHIDA et al. 2016, ROBERTSON et al. 2018). On the other hand, SeqSero and its updated version, SeqSero2, are mapping-based tools and work on raw reads as well as contigs (ZHANG et al. 2015, ZHANG et al. 2019). Besides, SeqSero2 has appeared as optimization of SeqSero including a k-mer based algorithm (“k-mer mode”) for rapid serovar prediction and a targeted micro-assembly approach (“allele mode”) (ZHANG et al. 2019). The results obtained were categorized into three main groups: “correct result”, “incorrect result” and “no result”. Furthermore, we subclassified the correct results into “full match”, “inconclusive” and “incongruent” according to the classification proposed by YACHISON et al. (2017) and also applied by UELZE et al. (2020)..Table S1, S2 and Figure S1 in the Appendix (section 8.2) present a detailed resolution of the results.

We observed the highest correlation for SISTR (SISTR_consensus), reporting 34/43 (79.1%) of correct matches followed by SeqSero2 (reads in “k-mer mode”) (72.1%), SeqSero2 (reads in “allele mode”) (67.4%), SeqSero2 (assemblies in “k-mer mode”) (67.4%), SeqSero (with reads) (60.5%) and SeqSero (with assemblies) (58.1%). These results are in line with the study performed by UELZE et al. (2020) that employed a dataset consisting of 1624 *Salmonella* isolates of 72 different serovars and concluded that SISTR with 94% of correctly typed isolates is the most suitable prediction tool for routine serotyping. Another study performed in Canada (YACHISON et al. 2017) reported as well the highest record of correctly serotyped serovar for SISTR. Furthermore, our data is largely consistent with the benchmarking study conducted by the international ENGAGE project (RENE et al. 2018). The ENGAGE project was a European collaborative study that performed benchmarking exercises to promote the implementation of WGS for replacing conventional typing for outbreak investigations. Within the benchmarking exercises concerning *Salmonella* subtyping, they analyzed a total of 786

serotyped isolates with the tools SISTR, SeqSero, Metric Oriented Sequence Typer (MOST) (TEWOLDE et al. 2016), and SalmonellaTypeFinder (based on SRST2, MLST, and SeqSero) (INOUYE et al. 2014, ZHANG et al. 2015, ASHTON et al. 2016) and concluded that SISTR with a 88% of correct results, shows the best correlation.

Additionally to the correct matches, in our study, SISTR reported the highest percentage of incongruent results (7%). We considered as incongruent results those for isolates which serovar could not be resolved by conventional serotyping (e.g. “rough” forms isolates). Among them, we found that all the tools were able to predict serovar Dublin for two “rough” form isolates within the database, and only SISTR detected the variant Java for serovar Paratyphi B. The percentage of incongruent results was equivalent to the rest of the tools. UELZE et al. (2020) and the study performed by YACHISON et al. (2017) reported as well agreement for all the tools including SISTR (4.1% and 3.14% respectively).

Regarding inconclusive predictions, those cases for which the tool listed at least two possible serovars, our analysis reported a percentage ranging from 0% (by SISTR) to 18.6% (by SeqSero using raw reads). However, the implementation of SeqSero2 (using reads in “allele mode”) could resolve half of these inconclusive predictions (9.3%). This occurred for serovars Bovismorbificans (typed by SeqSero as Bovismorbificans or Hindmarsh), serovar Hadar (typed as Hadar or Istanbul), serovar Indiana (typed as Indiana or II 4,12:z:1,7), and serovar Kottbus (typed as Kottbus or Ferruch). However, in three cases, both SeqSero and SeqSero2 could not resolve the correct serovar. For example for serovar Gallinarum, (typed as Gallinarum or Enteritidis), Goldcoast (typed as Goldcoast or Brikama), and Cholerasuis (typed as Cholerasuis or Paratyphi C or Typhisuis).

Both YACHISON et al. (2017) and UELZE et al. (2020) showed as well a low percentage of inconclusive matches for SISTR (1.1% and 1.23% respectively) in contrast to high percentages for SeqSero (30% and 12.56% respectively). We agree with UELZE et al. (2020) that the ability of SISTR to resolve ambiguous results may be due to the high discrimination power offered by its cgMLST approach. On the other hand, it should be taken into consideration that the cgMLST approach does not allow SISTR to distinguish monophasic isolates in the cases in which the cgMLST distance is very small (e.g monophasic variant of serovar Typhimurium). In those cases, further analysis using the conventional serotyping technique would be required (YACHISON et al. 2017). In the case of our analysis, the prediction of the monophasic variants

of serovar Typhimurium did not suppose an issue. In all the cases, SISTR could predict correctly the serovar with the antigenic formula I 4,[5],12:i:- and SeqSero and SeqSero2 could resolve it as "potential monophasic variant of Typhimurium". However, we should take into consideration the small size of our dataset and the number of isolates of this serovar is not representative (7% (3/43) strains typed as "monophasic Typhimurium" and 5% (2/43) typed as Typhimurium).

Surprisingly, in our study SISTR reported as well the highest number of incorrectly predicted serovars (14%) in an equal percentage as SeqSero using assemblies (14%). They were followed by SeqSero (using reads) (11.6%) and SeqSero2 (using reads in "allele mode") (11.6%). However, SeqSero2 in "k-mer mode" using indistinctly raw reads or contigs reported the fewest percentage of incorrect results (9.3%). Interestingly for six isolates reported by traditional serotyping as serovar Panama, Abony, Tennessee, Enteritidis, Mbandaka, and Panama all the tools failed in the serovar prediction or were not able to determine any serovar. The mismatch was evident, for serovar Panama (isolate 16PM0121), for which all the tools agreed on the prediction of serovar Goettingen and for serovar Tenessee (isolate 16PM0161), for which all the tools agreed on the prediction of serovar Mbandaka (Supplementary Table 1). We observe that the major difference between the mismatched serovar formulas resides in the second phase for the flagellar antigen. As reported before by YACHISON et al. (2017), one possible reason for the incorrect results is the inhability of the tool to call the correct flagellar antigenic determinant for close related serovars. Another reason could be the mistakenly reported serovar by conventional serotyping, however, this can not be supported in our study as the isolates were not retested either by conventional serotyping or bioinformatics analysis.

SeqSero using contigs was the tool with a high percentage of non-determined serovars (14.0%), including one *Salmonella enterica* subsp. *diarizonae* (IIIb) strain that was correctly typed by the rest of the tools as serovar 61:k:1,5,7. On the contrary, when using raw reads, SeqSero reported only 4.7% of no results. Besides, SeqSero2 reported 11.6% of no results when used with assemblies in "k-mer mode". For the majority of these cases, the tool reported possible inter serotype contamination and 7% when used with reads regardless of the mode.

Our results demonstrate that WGSBAC is feasible to determine the antigenic profile in most of the *Salmonella* strains while specific serovars need improved tools. All in all, we concluded

that SeqSero 2 for reads in k-mer mode appeared as accurate as SISTR. This largely correlates with recent results (UELZE et al. 2020). We are aware that the small size of our dataset suppose a handicap to observe the performance of the three bioinformatics tools tested. This fact affected, for example the recognition of the GC bias effect of the GC for some serovars or the differences due to the use of different library preparation kits (Nextera XT DNA library preparation kit/Nextera DNA Flex library preparation kit).

4.2. Application of the WGSBAC pipeline to study the potential occurrence and high tendency of dissemination of a MDR emergent *S*. Infantis population in German broiler farms.

S. Infantis has doubled its prevalence in broiler flocks and the number of humans affected by this NTS serovar has reached a high level so that it is now considered a potential hazard for public health (EFSA 2019a). Only during the last year, several studies including the present one, have reported the emergent situation and rapid worldwide dissemination of *S*. Infantis isolates with recurrent resistance and virulence profiles (ALBA et al. 2020, GARCÍA-SOTO et al. 2020, JOVCIC et al. 2020, LAPIERRE et al. 2020, MCMILLAN et al. 2020, NAGY et al. 2020, PROIETTI et al. 2020, TYSON et al. 2020, KUREKCI et al. 2021). Studies agree that the possession of a megaplasmid named pESI (plasmid of Emergent Salmonella Infantis) or similar ones (pESI-like plasmids) characterize the emergent S. Infantis (ESI) population. In this study, the presence of a conjugative pESI-like plasmid has been detected within German broiler-derived *S*. Infantis strains. The genome comparison of the German plasmid pESI-like (pESI2747) and the Israeli (pESI119944) revealed a homology of 97.37%. Although in different locations, both contained the replicon IncFIB, the *repB* gene coding for RepFIB replication protein A, the complete IncI gene profile (*ardA*, *pilL*, *sogS*, *trbA,* and *repI1* absent), and the origin of replication of an IncP plasmid. The homology in terms of genetic composition between the Israeli pESI and the German pESI-like has been observed within the strains collected from the 2010s and not within the ones from the 1990s. Besides, the in-detailed comparative genome analysis of the complete genome sequence of one pESI-like positive strain within the German dataset (2747) with the complete genome sequence of *S*. Infantis from Israel (ESI119944) and other ESI clones from Europe (FRANCO et al. 2015, OLASZ et al. 2015, WILK et al. 2016, WILK et al. 2017) confirms the homology between the resistance, virulence, and fitness genetic traits encoded on the pESI-like plasmid. These findings are in agreement with other studies that indicate that

the possession of the megaplasmid might be a potential reason for the switch in the occurrence of *S*. Infantis in broilers from non-MDR strains (until the 2000s) (ASAI et al. 2007, SHAHADA et al. 2008, WILK et al. 2017) to the emergence of MDR clones (during and after the end of the 2000s) (NÓGRÁDY et al. 2007, NÓGRÁDY et al. 2008, NÓGRÁDY et al. 2012).

The predominant MDR phenotypic profile of the ESI clones from Europe consists mostly of resistance to the antimicrobial families (aminoglycosides, sulphonamides (trimethoprim), and tetracyclines) (CLOECKAERT et al. 2007, FRANCO et al. 2015, ACAR et al. 2019, PROIETTI et al. 2020). The application of these antimicrobials for the prevention and therapy of diseases in poultry may facilitate the occurrence of this AMR profile among *S*. Infantis strains (WASSENAAR. 2005). The variable nature of the pESI plasmid as recently indicated by COHEN et al. (2020) enables genetic and phenotypic diversity, especially to the AMR profile. Among the AMR pattern variations found in other ESI clones from Europe, the presence of ESBL-positive clones found mostly in Italy (FRANCO et al. 2015, PROIETTI et al. 2020), Hungary (SZMOLKA et al. 2018), and Switzerland (HINDERMANN et al. 2017) needs to be mentioned. The Italian ESI clones carry the gene $bla_{\text{CTX-M-1}}$ conferring resistance to cephalosporin as well as the gene *mcr-1* gene mediating colistin resistance (CARFORA et al. 2018). However, another recent study has reported reduced susceptibility to colistin among ESI isolates from poultry farms in Serbia (JOVCIC et al. 2020). They hypothesized that this reduction is due to the more prudent use of enrofloxacin in the past years. Furthermore, in Hungary, ESI was identified to carry the gene $bla_{\text{TEM-1}}$ and the *qnrS* for beta-lactam and fluoroquinolone resistance on the pESI-like plasmid (pESI54/04) (SZMOLKA et al. 2018) while in Switzerland, only one strain contained the gene $bla_{\text{CTX-M-65}}$ (HINDERMANN et al. 2017) reflecting the uncommon presence of the gene $bla_{\text{CTX-M-65}}$ in the EU food-production chain. Contrarily, in North and South America due to the recurrent presence of the gene $bla_{\text{CTX-M-65}}$, it has been used as a predictor for the detection of pESI plasmids (TATE et al. 2017, FUENTES-CASTILLO et al. 2019, MCMILLAN et al. 2020).

In the present study, *in silico* phenotyping has confirmed the recurrent MDR phenotype in the majority of German strains from the 2010s. We named the pattern "ESIr" (Emergent Salmonella Infantis resistance) and it consisted of the genes *ant(3")-Ia*, *sul1,* and *tet(A)* that correlate with the phenotypic resistance profile of the strains to aminoglycosides, sulfonamides, and tetracyclines, respectively. Some authors observed a phenotype-genotype

incongruence for the genes *ant(3")-Ia* and *aac(6')-Iaa* mediating resistance to aminoglycosides (ACAR et al. 2019, KUREKCI et al. 2021). We did not find the gene *aac(6')-Iaa* within the pESI-like plasmid, but on the chromosome, suggesting its nature as a chromosomal-encoded gene as indicated before (SALIPANTE et al. 2003). Additionally, within the chromosome of German ESI strains, a mutation in the *gyrA* gene was detected, known to deal with resistance to fluoroquinolones (ciprofloxacin and nalidixic acid)(CHENG et al. 2019). Recently, four different chromosomal point mutations of the gene *gyrA* were found as well in different proportions among European strains (ALBA et al. 2020). Within the German strains, we did not find any ESBL-positive clone, however, a study performed in Germany showed the presence of genes for beta-lactamases circulating in the food chain (FISCHER 2017).

The backbone of pESI and pESI-like plasmids are not only defined to confer MDR profile to the ESI clones but also a specific set of virulence markers (AVIV et al. 2014, COHEN et al. 2020). In general, the virulence pattern consists of several virulence genes coding for structures involved in increasing host invasion and colonization, motility, and persistence of the ESI in the farms (AVIV et al. 2014, AVIV et al. 2016, AVIV et al. 2017). Among them, the yersiniabactin-iron acquisition system (Ybt) and the novel fimbriae complexes Klf (K88-like fimbria), and the Ipf (Infantis plasmid-encoded fimbria) have been detected. Furthermore, the toxin/antitoxin complexes such as CcdA/B, PemK/I, MazE/F, and VagCD have been reported as systems activated under stress conditions and play a role in the adaptability of plasmids and persistence of certain *Salmonella* serovars (GOEDERS et al. 2014, DI CESARE et al. 2016). Additionally, the mercury (*mer)* operon conferring bacterial tolerance to mercury (COHEN et al. 2020) and the gene *qacEdelta1* known to provide resistance to heavy metals or biocides like quaternary ammonium compounds were found (QACs) (CHUANCHUEN et al. 2007).

Regarding the virulence profile of the German ESI clones, the bioinformatics analysis found this specific virulence and fitness-associated genes pattern with some variations. We termed this pattern "ESIv" (Emergent Salmonella Infantis virulence) and it was only present among the strains collected from the 2010s. In contrast to the variable nature of the pESI-like plasmids, the chromosome of ESI strains provides a conserved distribution of *Salmonella* pathogenicity islands (SPIs) across *S*. Infantis and other ESI clones genomes (COHEN et al. 2020). In this study, within the chromosome of the ESI German clones, we found 10 intact SPIs (SPIs-1-6, SPI-9, SPIs-11-12, and CS-54) also detected recently by others (JOVCIC et al. 2020).

Other gene complexes related to fimbrial production and known to be part of the core genome of NTS serovars (DHANANI et al. 2015) such us *fim* (type 1 fimbria) and *lpf* (long polar fimbria), were found among the chromosome of the *S.* Infantis from this study.

S. enterica and other Enterobacteriaceae are known to carry plasmids involved in resistance and virulence (CARATTOLI 2009). Previous studies reported that certain *Salmonella* serovars commonly carry large (>40 kb), virulence-plasmids in low-copy numbers (1-2 copies) (SANCHEZ-ROMERO et al. 2020). The acquisition of virulence-plasmids may benefit specially host-adapted serovars as they may expand their host range (ROTGER et al. 1999). Some plasmids carry as well genes that control the expression of essential bacterial structures and others are also reservoirs of AMR (EMOND-RHEAULT et al. 2020, SANCHEZ-ROMERO et al. 2020). Furthermore, there are so-called hybrid virulence-resistance plasmids in NTS such as *S*. Typhimurium and *S*. Cholerasuis (MENDOZA MDEL et al. 2009).

As occurred in other bacterial pathogens, the presence of pESI or pESI-like plasmid-borne genes may ensure the evolutive success of the emergent *S*. Infantis population. It is accepted that the AMR and virulence-associated determinants encoded in pESI plasmid may play a beneficial role in the pathogenesis and rapid dissemination of the emergent *S*. Infantis population (AVIV et al. 2014, COHEN et al. 2020). However, the carriage of such large and complex plasmids may impose as well fitness stress and a metabolic cost on the host. The megaplasmid pESI is presented in a single copy and despite its large size (258081 bp), its presence does not result in a significant burden to its host and does not limit its transmission and dissemination potential (AVIV et al. 2016). Evidence of this is that fixing in the population has been demonstrated by AVIV et al. (2016) as well as an interspecies transfer into the mouse microbiota (from *S*. Infantis to *Escherichia coli*). The authors hypothesize about the role of microbiota in the transfer of genes between pathogenic and commensal bacteria. Furthermore, due to the fitness cost that conjugative transfer supposes, the plasmids encode mechanisms associated with their maintenance (PILLA et al. 2018). The comparative genome analysis performed by BOGOMAZOVA et al. (2019) showed that pESI-like plasmids contain a conservative 173kB sequence encoding genes responsible for plasmid maintenance and conjugative transfer. Among them transfer (*tra*) genes and pilus *(pil)* genes present also within the German pESI-like plasmid may play a role in the regulation of the plasmid metabolism. They have been as well observed among the Turkey strains (ACAR et al. 2019). As occurred for

the virulence-plasmid of *S*. Typhimurium in strain LT2 (pSLT), toxin-antitoxin systems may ensure successful stability of megaplasmids (SANCHEZ-ROMERO et al. 2020). In particular, the CcdA-CcdB complex may be implicated as well in the adaptability and stability of pESI and pESI-like megaplasmids (HAYES 2003, GOEDERS et al. 2014)

Besides, genotyping and phylogenetic studies of the ESI clones revealed that the major ST among the ESI clones from Europe is ST32 (FRANCO et al. 2015, ALBA et al. 2020, KUREKCI et al. 2021). However, novels STs have been found recently in Serbia (ST413 and ST11) in strains corresponding to different geographical regions. Moreover, in Turkey, a single novel ST7091 was found in a minor proportion compared to the dominant ST32 (KUREKCI et al. 2021). In this study, we found the ST32 within half of the German dataset (15 out of 30) as well as a novel ST2283 within the strains from the 2010s and the previously reported ST1032 (ALBA et al. 2020) within non-emergent isolates from the 1990s. Interestingly, and in contrast to previous studies, we found ST32 among the majority of non-emergent strains from the 1990s and only two ESI strains from the 2010s. Meanwhile, the novel ST2283 was predominant among the ESI strains from the 2010s. This might indicate a greater tendency of dissemination of ESI clone ST2283 in Germany. A SNP-based and chromosome-based phylogenetic analysis showed general concordance in cluster separation between the chromosome- and the plasmid-based tree in all the strains (ST32 and ST2283) harboring the pESI-like plasmid as also shown by ALBA et al. (2020). Therefore, we hypothesized that the plasmid has co-evolved with the chromosome and both STs (ST32 and ST2283) gained the plasmid in two (or more) evolutionary independent events. GYMOESE et al. (2019) observed as well separation of clusters between *S*. Infantis strains carry pESI-like plasmids and plasmidless *S*. Infantis population. They suggested that the use of antimicrobials at the beginning of the poultry production industry (around 60 years ago) may select for the pESI-like positive *S*. Infantis population.

The prompt dissemination of an ESI population observed during the last decade in Germany supposes an urgent public health concern. The study presented here gives evidence of the occurrence of an emergent MDR *S*. Infantis population in Germany able to strong dissemination within two decades. The acquisition of AMR, virulence-associated determinants, as well as the pESI-like encoded mechanisms such as toxin-antitoxin systems or transfer and pilus genes in favor of plasmid maintenance and conjugation transfer, may

promote this worrying epidemiological event. The use of WGS and our own implemented bioinformatics pipeline has successfully enabled i) the reliable serovar prediction of the strains driven by the previous implementation of an in-house pipeline for serovar prediction and ii) the power for the accurate and detailed characterization of German strains and comparison with other *S*. Infantis clones circulating in Europe and outside Europe. Nevertheless, we are aware of the limitations of this study and that further analysis is needed to provide a more robust explanation for this epidemic hazard in Germany.

5. ZUSAMMENFASSUNG

Silvia García Soto

Eine auf Gesamtgenomsequenzierung basierende Studie zur Verbreitung multiresistenter *Salmonella enterica* subspecies *enterica* serovar Infantis-Klone in deutschen Broilerbetrieben

Institut für Tierhygiene und Öffentliches Veterinärwesen, Veterinärmedizinischen Fakultät, Universität Leipzig

Eingereicht im März 2021

(48 Seiten, 1 Figuren, 1 Tabelle, 148 Literaturangaben, 1 Publikationen, Anhänge)

Schlüsselwörter: *S. Infantis*, Broiler, Multidrug-Resistenz, pESI-Plasmid, Gesamtgenomsequenzierung

Einleitung: *Salmonella enterica* subspecies *enterica* serovar Infantis (*S*. Infantis) nimmt die vierte Position in der Rangliste der am häufigsten gemeldeten Salmonella-Serovare in Europa ein. Während des letzten Jahrzehnts hat das Auftreten einer multiresistenten *Salmonella enterica* subspecies *enterica* serovar Infantis (*S*. Infantis)-Population rapide zugenommen und ist in der europäischen und außereuropäischen Broilerproduktion weit verbreitet.

Ziel der Untersuchungen: Die Studie zielte darauf ab, i) eine bioinformatische Pipeline für die in silico-Serotypisierung von Salmonellen zu implementieren und deren Leistungsfähigkeit zu bewerten, und ii) die genetischen Determinanten für das in der deutschen Broilerproduktion während des letzten Jahrzehnts beobachtete vermehrte Auftreten von *S*. Infantis zu identifizieren.

Tiere, Material und Methoden: Zunächst wurde eine Bioinformatik-Pipeline (WGSBAC) und drei bioinformatische Tools (SISTR, SeqSero und SeqSero2) zur Genoserotypisierung von 43 *Salmonella spp*.-Stämmen von 26 verschiedenen Serovaren durchgeführt. Zweitens führten wir die Genomsequenzierung von 30 *S*. Infantis Broiler-Isolaten durch, die in zwei unterschiedlichen Jahrzehnten (den 1990er und den 2010er Jahren) gesammelt wurden. Wir setzen die WGSBAC-Pipeline und externe Bioinformatik-Software ein, um i) die Qualität der sequenzierten Reads zu bewerten und zu kontrollieren, ii) Assemblierungen und Qualitätskontrollen von und iii) Annotation, Typisierung durch klassischen MLST, cgMLST, Genoserotypisierung, phylogenetische Rekonstruktion mittels Einzelnukleotidänderungen

(SNPs) und *in-silico*-Phänotyp-Vorhersage, einschließlich antimikrobieller Resistenzgene (AMR), Virulenzgene und Plasmid-Replikons zu erkennen.

Ergebnisse: Die Bioinformatik-Pipeline WGSBAC ist geeignet, das antigene Profil der meisten der in der Studie verwendeten Salmonella-Stämme zu bestimmen. Das Tool SISTR zeigte dabei die höchste Übereinstimmung (79,1 %), gefolgt von SeqSero2 (72,1 %) und SeqSero (60,5 %). Die Untersuchung der *S*. Infantis-Stämme ergab, dass im Gegensatz zu den Isolaten aus den 1990er Jahren, die Mehrheit der Stämme aus den 2010er Jahren das Vorhandensein eines Megaplasmids zeigte, das homolog zu dem pESI-Plasmid aus Israel und anderen pESI-ähnlichen Plasmiden aus europäischen Isolaten ist. Das deutsche pESI-ähnliche Plasmid kodierte für ein Muster von multiresistenten Genen (MDR), ein Muster von Virulenzgenen und mehrere Fitness-assoziierte Determinanten, welches wir "ESIr" nannten. Dieses korrelierte mit mindestens drei antimikrobiellen Familien: ant(3")-Ia (Aminoglykoside), sul1 (Sulfonamide) und tet(A) (Tetracycline). Der Genotyp korreliert hier vollständig mit dem antimikrobiellen Phänotyp. Außerdem bezeichneten wir das Virulenzmuster als "ESIv", welches Gene für Fimbrien-Cluster, Yersiniabactin-Siderophore, Quecksilberresistenz und Antitoxin/Antitoxin-Systeme mit einschließt. Die Genotypisierungsanalyse ergab das Vorhandensein eines neuen Sequenztyps (ST2283) bei der Mehrzahl der Stämme aus den 2010er Jahren und ST32 und ST1032 bei den Stämmen aus den 1990er Jahren.

Schlussfolgerungen: Anhand eines auf Gesamtgenomsequenzierung-basierten Ansatzes zeigte diese Studie, dass MDR *S*. Infantis ST2283 Stämme, die ein pESI-ähnliches Plasmid tragen, während des letzten Jahrzehnts entstanden sind und derzeit in der deutschen Geflügelproduktionskette zirkulieren. Der Erwerb eines Megaplasmids, das für Resistenzen, Virulenz-assoziierte Determinanten und Fitnessmechanismen kodiert, könnte dieses schnelle und besorgniserregende epidemiologische Ereignis erklären. Dieses Ereignis stellt eine Gefahr für die öffentliche Gesundheit dar. Daher sind Kontrollmaßnahmen und die Unterstützung epidemiologischer Studien erforderlich, um den Eintritt, die Übertragung und die weitere Verbreitung dieser klonalen Population in der Lebensmittelkette zu verhindern.

6. SUMMARY

Silvia García Soto

A Whole-Genome sequencing-based study of the emergent multidrug-resistant *Salmonella enterica* subspecies *enterica* serovar Infantis clones in German broiler farms

Institute of Animal Hygiene and Veterinary Public Health of the Faculty of Veterinary Medicine, Leipzig University

Submitted in March 2021

(48 pages, 1 figure, 1 table, 148 references, 1 publication, appendices)

Keywords: *Salmonella Infantis*, broiler, emergence, multidrug-resistance, pESI-like plasmid, whole-genome sequencing

Background: *Salmonella enterica* subspecies *enterica* serovar Infantis (*S*. Infantis) places the fourth position in the ranking of most reported *Salmonella* serovars in Europe. During the last decade, a multi-drug resistant (MDR) *S.* Infantis population has rapidly increased and widespread in European and non-European broiler production.

Goal: The study proposed here aimed to i) implement and evaluate the performance of a bioinformatics pipeline named WGSBAC for *Salmonella in silico* serotyping, and ii) identify the genetic determinants for the increased emergence of broiler-derived *S*. Infantis observed in Germany.

Animals, material, and methods: First, we conducted an evaluation of WGSBAC and three bioinformatic tools (SISTR, SeqSero, and SeqSero2) for the characterization and genoserotyping of 43 *Salmonella* strains of 26 different serovars. Second, we performed sequencing of 30 broiler-derived *S*. Infantis isolates collected from two distant decades (the 1990s and the 2010s). We applied the WGSBAC pipeline and external bioinformatics software to i) assess and control the quality of the sequenced reads ii) assembly and quality control of assemblies, and iii) annotation, typing by classical MLST, cgMLST, genoserotyping, SNPs-based phylogenetic reconstruction and *in silico* phenotype prediction including antimicrobial resistance genes (AMR), virulence genes and plasmid replicons detection. To detect possible clonal relatedness with other *S*. Infantis clones from Europe, we performed a further

comparative genome analysis using 17 public genomes of other *S*. Infantis clones circulating in Europe.

Results: WGSBAC was feasible for the serovar prediction of most of the 43 *Salmonella* strains. The tool SISTR reported the highest correlation (79.1%) followed by SeqSero2 (72.1%) and SeqSero (60.5%). The study of the *S*. Infantis strains revealed that in contrast to the isolates from the 1990s, the majority of the strains from the 2010s revealed the presence of a megaplasmid that carried a multidrug-resistant genes (MDR) pattern, a virulence genes pattern, and several fitness-associated determinants. We termed the MDR gene pattern "ESIr" and it coded for at least three antimicrobial families: *ant(3")-Ia* (aminoglycosides), *sul1* (sulfonamides), and *tet(A)* (tetracyclines). Besides, we termed the virulence pattern as "ESIv" which includes genes for fimbriae cluster, yersiniabactin siderophore, mercury resistance, and antitoxin/antitoxin systems. Furthermore, the genotyping analysis revealed the presence of a novel sequence type (ST2283) among the majority of the strains from the 2010s and ST32 and ST1032 within the strains from the 1990s. This genetic traits may promote the rapid incidence and dissemination of a novel MDR *S*. Infantis population.

Conclusion: Following a WGS-based approach, this study evidences that MDR *S*. Infantis ST2283 strains carrying a pESI-like plasmid have emerged during the last decade and are currently circulating in the German poultry production chain. This event results in an urgent public health hazard, thus, control measures and the support of epidemiological studies are needed to prevent the entrance, transmission, and further dissemination of this clonal population in the food chain.

7. REFERENCES

Acar S, Bulut E, Stasiewicz MJ, Soyer Y. Genome analysis of antimicrobial resistance, virulence, and plasmid presence in Turkish Salmonella serovar Infantis isolates. Int J Food Microbiol. 2019;307:108275.

Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog. 2012;8(6):e1002776.

Achtman M, Zhou Z, Alikhan N, Tyne W, Parkhill J, Cormican M, et al. Genomic diversity of Salmonella enterica -The UoWUCC 10K genomes project [version 1; peer review: 2 approved]. 2020;5(223).

Alba P, Leekitcharoenphon P, Carfora V, Amoruso R, Cordaro G, Di Matteo P, et al. Molecular epidemiology of Salmonella Infantis in Europe: insights into the success of the bacterial host and its parasitic pESI-like megaplasmid. Microb Genom. 2020.

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517-D525.

Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of Salmonella. PLoS Genet. 2018;14(4):e1007261.

Andrews S. FastQC: A quality control tool for high throughput sequence data. Version 0.11.5 (software) 2016 Mar 08 (cited 2020 Dec 10) <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/2018>

Asai T, Ishihara K, Harada K, Kojima A, Tamura Y, Sato S, et al. Long-term prevalence of antimicrobial-resistant Salmonella enterica subspecies enterica Serovar Infantis in the broiler chicken industry in Japan. Microbiol Immunol. 2007;51(1):111-5.

Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, et al. Identification of Salmonella for public health surveillance using whole genome sequencing. PeerJ. 2016;4:e1752.

Aviv G, Elpers L, Mikhlin S, Cohen H, Vitman Zilber S, Grassl GA, et al. The plasmid-encoded Ipf and Klf fimbriae display different expression and varying roles in the virulence of Salmonella enterica serovar Infantis in mouse vs. avian hosts. PLoS Pathog. 2017;13(8):e1006559.

Aviv G, Rahav G, Gal-Mor O. Horizontal Transfer of the Salmonella enterica Serovar Infantis Resistance and Virulence Plasmid pESI to the Gut Microbiota of Warm-Blooded Hosts. MBio. 2016;7(5).

Aviv G, Tsyba K, Steck N, Salmon-Divon M, Cornelius A, Rahav G, et al. A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent Salmonella enterica serovar Infantis strain. Environ Microbiol. 2014;16(4):977-94.

Banerji S, Simon S, Tille A, Fruth A, Flieger AA-O. Genome-based Salmonella serotyping as the new gold standard. 2020(2045-2322 (Electronic)).

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.

Bundesinstitut für Risikobewertung (BfR) 2020. Salmonellen-Bekämpfungsprogramm – Ergebnisse für das Jahr 2019: Salmonella Enteritidis und Salmonella Typhimurium bei Legehennen rückläufig 2020. 2020 Jul 28 (cited 2021 Dec 10) <https://www.bfr.bund.de/cm/343/salmonellen-bekaempfungsprogramm-ergebnisse-fuer-2019.pdf>

Bogomazova AN, Gordeeva VD, Krylova EV, Soltynskaya IV, Davydova EE, Ivanova OE, et al. Mega-plasmid found worldwide confers multiple antimicrobial resistance in Salmonella Infantis of broiler origin in Russia. Int J Food Microbiol. 2019;319:108497.

Boxrud D. Advances in subtyping methods of foodborne disease pathogens. Curr Opin Biotechnol. 2010;21(2):137-41.

Bundesministerium. Verordnung zum Schutz gegen bestimmte Salmonelleninfektionen beim Haushuhn und bei Puten (Geflügel-Salmonellen-Verordnung - GflSalmoV). Bundesgesetzblatt. 2014;I.

Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) 2019. Berichte zur Lebensmittelsicherheit - Zoonosen-Monitoring 2018. Berlin: Bundesamt für Verbraucherschutz und Lebensmittelsicherheit; 2019 (cited 2021 Feb 20) <https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2018.pdf?__blob=publicationFile&v=7>

Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) 2018. Zoonoses Monitoring 2017 - Summary of Findings and Conclusions. 2018 (cited 2021 Feb 20). <https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2018_ensummary.pdf;jsessionid=334DF874445DCCD03907D92AD593C485.2_cid341?__blob=publicationFile&v=3>

Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009;53(6):2227-38.

Carattoli A, Hasman H. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). Methods Mol Biol. 2020;2075:285-294.

Carfora V, Alba P, Leekitcharoenphon P, Ballaro D, Cordaro G, Di Matteo P, et al. Colistin Resistance Mediated by mcr-1 in ESBL-Producing, Multidrug Resistant Salmonella Infantis in Broiler Chicken Industry, Italy (2016-2017). Front Microbiol. 2018;9:1880.

Carrico JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. Clin Microbiol Infect. 2018;24(4):342-349.

Cartelle Gestal M, Zurita J, Paz YMA, Ortega-Paredes D, Alcocer I. Characterization of a small outbreak of Salmonella enterica serovar Infantis that harbour CTX-M-65 in Ecuador. Braz J Infect Dis. 2016;20(4):406-7.

Cheng RA, Eade CR, Wiedmann M. Embracing Diversity: Differences in Virulence Mechanisms, Disease Severity, and Host Adaptations Contribute to the Success of Nontyphoidal Salmonella as a Foodborne Pathogen. Front Microbiol. 2019;10:1368.

Chuanchuen R, Khemtong S, Padungtod P. Occurrence of qacE/qacEDelta1 genes and their correlation with class 1 integrons in Salmonella enterica isolates from poultry and swine. Southeast Asian J Trop Med Public Health. 2007;38(5):855-62.

Cloeckaert A, Praud K, Doublet B, Bertini A, Carattoli A, Butaye P, et al. Dissemination of an extended-spectrum-beta-lactamase blaTEM-52 gene-carrying IncI1 plasmid in various Salmonella enterica serovars isolated from poultry and humans in Belgium and France between 2001 and 2005. Antimicrob Agents Chemother. 2007;51(5):1872-5.

Cohen E, Rahav G, Gal-Mor O. Genome Sequence of an Emerging Salmonella enterica Serovar Infantis and Genomic Comparison with Other S. Infantis Strains. Genome Biol Evol. 2020;12(3):151-159.

Dhanani AS, Block G, Dewar K, Forgetta V, Topp E, Beiko RG, et al. Genomic Comparison of Non-Typhoidal Salmonella enterica Serovars Typhimurium, Enteritidis, Heidelberg, Hadar and Kentucky Isolates from Broiler Chickens. PLoS One. 2015;10(6):e0128773.

Di Cesare A, Losasso C, Barco L, Eckert EM, Conficoni D, Sarasini G, et al. Diverse distribution of Toxin-Antitoxin II systems in Salmonella enterica serovars. Sci Rep. 2016;6:28759.

European Centre for Disease Prevention and Control (ECDC). Expert opinion on whole genome sequencing for public health surveillance. Stockholm: European Centre for Disease Prevention and Control, ECDC; 2016 (cited 2021 Feb 20). <https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/whole-genome-sequencing-for-public-health-surveillance.pdf>.

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2017. Multi-country outbreak of Salmonella Enteritidis infections linked to Polish eggs. EFSA supporting publication 2017:EN-1353. 2017 (cited 2021 Feb 20). <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2017.EN-1353>

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2018. Multi-country outbreak of Salmonella Agona infections linked to infant formula. EFSA supporting publication 2018 (cited 2021 Feb 20): EN-1365. 2018a <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2018.EN-1365>

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2018. Multi-country outbreak of Salmonella Agona infections possibly linked to ready-to-eat food. EFSA supporting publication 2018 (cited 2021 Feb 20):EN-1465. 2018b <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2018.EN-1465>

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2019. Multi-country outbreak Salmonella Poona infections linked to consumption of infant formula. EFSA supporting publication 2019 (cited 2021 Feb 20):EN-1594. 2019. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2019.EN-1594>

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2020. Multi-country outbreak of Salmonella Enteritidis infections linked to eggs, third update. 2020 Feb 06 (cited 2021 Feb 20). 2020a <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2020.EN-1799>

European Centre for Disease Prevention and Control, European Food Safety Authority (ECDC-EFSA) 2020. Multi-country outbreak of Salmonella Typhimurium and S. Anatum infections linked to Brazil nuts. 2020 October 21 (cited 2020 Dec 10) 2020b. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2020.EN-1944>.

European Food Safety Authority (EFSA) 2019. The European Union One Health 2018 Zoonoses Report. EFSA Journal. 2019a;17(12). 2019 Nov 19 (cited 2020 Dec 10) <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2019.5926>

European Food Safety Authority (EFSA) 2019. Salmonella control in poultry flocks and its public health impact. EFSA Journal. 2019b;17(2). 2019 Jan 16 (cited 2020 Dec 10) <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2019.5596>

European Food Safety Authority (EFSA) 2020. The European Union Summary Report on Antimicrobial Resistance in zoonotic and indicator bacteria from humans, animals and food in 2017/2018. EFSA Journal. 2020;18(3). 2020 Jan 31 (cited 2021 Feb 20) <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2020.6007>

Emond-Rheault JG, Hamel J, Jeukens J, Freschi L, Kukavica-Ibrulj I, Boyle BA-O, et al. The Salmonella enterica Plasmidome as a Reservoir of Antibiotic Resistance. LID - 10.3390/microorganisms8071016 [doi] LID - 1016. 2020(2076-2607 (Print)).

Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, 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(11).

Fischer J, Borowiak, M., Baumann, B., Szabo, I., Malorny, B. Whole-genome sequencing analysis of multidrug-resistant Salmonella Infantis isolates circulating in the German food-production chain. 27th ECCMID; 24 April 2017; Viena, Austria 2017. p. P1080.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science. 1995;269(5223):496-512.

Franco A, Leekitcharoenphon P, Feltrin F, Alba P, Cordaro G, Iurescia M, et al. Emergence of a Clonal Lineage of Multidrug-Resistant ESBL-Producing Salmonella Infantis Transmitted from Broilers and Broiler Meat to Humans in Italy between 2011 and 2014. PLoS One. 2015;10(12):e0144802.

Fuentes-Castillo D, Farfan-Lopez M, Esposito F, Moura Q, Fernandes MR, Lopes R, et al. Wild owls colonized by international clones of extended-spectrum beta-lactamase (CTX-M)-producing Escherichia coli and Salmonella Infantis in the Southern Cone of America. Sci Total Environ. 2019;674:554-562.

Gal-Mor O, Valinsky L, Weinberger M, Guy S, Jaffe J, Schorr YI, et al. Multidrug-resistant Salmonella enterica serovar Infantis, Israel. Emerg Infect Dis. 2010;16(11):1754-7.

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.

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.

Gill RK, Hecht GA. Chapter 64 - Host-Pathogen Interactions in Pathophysiology of Diarrheal Disorders. In: Said HM, editor. Physiology of the Gastrointestinal Tract (Sixth Edition): Academic Press; 2018. p. 1547-1577.

Goeders N, Van Melderen L. Toxin-antitoxin systems as multilevel interaction systems. Toxins (Basel). 2014;6(1):304-24.

Grimont P, Weill F-X. Antigenic formulae-Grimont-Weill.pdf. WHO Collaborating Center for Reference and Research on Salmonella, Institut Pasteur 9th Edn. 2007 (cited 2021 Feb 20) <https://www.pasteur.fr/sites/default/files/veng_0.pdf>

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.

Gymoese P, Kiil K, Torpdahl M, Osterlund MT, Sorensen G, Olsen JE, et al. WGS based study of the population structure of Salmonella enterica serovar Infantis. BMC Genomics. 2019;20(1):870.

Hayes F. Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. Science. 2003;301(5639):1496.

Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using Genomics to Track Global Antimicrobial Resistance. Front Public Health. 2019;7:242.

Henzler DJ, Opitz HM. The role of mice in the epizootiology of Salmonella enteritidis infection on chicken layer farms. 1992(0005-2086 (Print)).

Hernández M, Quijada NM, Rodríguez-Lázaro D, Eiros JM. Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. Rev Argent Microbiol. 2020;52(2):150-161.

Hindermann D, Gopinath G, Chase H, Negrete F, Althaus D, Zurfluh K, et al. Salmonella enterica serovar Infantis from Food and Human Infections, Switzerland, 2010-2015: Poultry-Related Multidrug Resistant Clones and an Emerging ESBL Producing Clonal Lineage. Front Microbiol. 2017;8:1322.

Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90.

International Organization for Standardization (ISO). ISO/DIS 23418 Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance. 1 ed: ISO; 2020. 42 p.

Jagadeesan B, Gerner-Smidt P, Allard MW, Leuillet S, Winkler A, Xiao Y, et al. The use of next generation sequencing for improving food safety: Translation into practice. Food Microbiol. 2019;79:96-115.

Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17(1):239.

Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.

Jovcic B, Novovic K, Filipic B, Velhner M, Todorovic D, Matovic K, et al. Genomic Characteristics of Colistin-Resistant Salmonella enterica subsp. enterica Serovar Infantis from Poultry Farms in the Republic of Serbia. Antibiotics (Basel). 2020;9(12).

Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4):294-6.

Kauffmann F. Classification of bacteria : a realistic scheme with special reference to the classification of Salmonella- and Escherichia-species: Copenhagen : Munksgaard; 1975.

Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. Semin Thromb Hemost. 2019;45(7):661-673.

Kurekci C, Sahin S, Iwan E, Kwit R, Bomba A, Wasyl D. Whole-genome sequence analysis of Salmonella Infantis isolated from raw chicken meat samples and insights into pESI-like megaplasmid. Int J Food Microbiol. 2021;337:108956.

Lapierre L, Cornejo J, Zavala S, Galarce N, Sanchez F, Benavides MB, et al. Phenotypic and Genotypic Characterization of Virulence Factors and Susceptibility to Antibiotics in Salmonella Infantis Strains Isolated from Chicken Meat: First Findings in Chile. Animals (Basel). 2020;10(6).

Lärkeryd A, Myrtennäs K, Karlsson E, Dwibedi CK, Forsman M, Larsson P, et al. CanSNPer: a hierarchical genotype classifier of clonal pathogens. Bioinformatics. 2014;30(12):1762-1764.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W259.

Lindqvist N, Pelkonen S. Genetic surveillance of endemic bovine Salmonella Infantis infection. Acta Vet Scand. 2007;49:15.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376-80.

McMillan EA, Wasilenko JL, Tagg KA, Chen JC, Simmons M, Gupta SK, et al. Carriage and Gene Content Variability of the pESI-Like Plasmid Associated with Salmonella Infantis Recently Established in United States Poultry Production. Genes (Basel). 2020;11(12).

McQuiston JR, Parrenas R, Ortiz-Rivera M, Gheesling L, Brenner F, Fields PI. Sequencing and comparative analysis of flagellin genes fliC, fljB, and flpA from Salmonella. J Clin Microbiol. 2004;42(5):1923-32.

Mendoza Mdel C, Herrero A, Rodicio MR. Evolutionary engineering in Salmonella: emergence of hybrid virulence-resistance plasmids in non-typhoid serotypes. Enferm Infecc Microbiol Clin. 2009;27(1):37-43.

Methner U. Salmonella Enteritidis. Das Geflügel und der Mensch-ist die Pandemie vorüber? Im Fokus Friedrich-Loeffler-Institut. 2013. 2013 Juli 15 (cited 2020 Dec 15) <https://www.openagrar.de/servlets/MCRFileNodeServlet/Document_derivate_00011848/Im_Fokus_01-2013.pdf>

Methner U. Salmonellose der Rinder - Salmonellosis in cattle. Tiergesundheitsjahresberichte 2010-2019. Friedrich-Loeffler-Institut; 2019. 2020 Dec 11 (cited 2020 Dec 15) <https://www.openagrar.de/servlets/MCRFileNodeServlet/openagrar_derivate_00034561/TGJB_2019.pdf>

Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.

Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Euro Surveill. 2017;22(23).

Nagy T, Szmolka A, Wilk T, Kiss J, Szabo M, Paszti J, et al. Comparative Genome Analysis of Hungarian and Global Strains of Salmonella Infantis. Front Microbiol. 2020;11:539.

Nógrády N, Kardos G, Bistyak A, Turcsanyi I, Meszaros J, Galantai Z, et al. Prevalence and characterization of Salmonella infantis isolates originating from different points of the broiler chicken-human food chain in Hungary. Int J Food Microbiol. 2008;127(1-2):162-7.

Nógrády N, Kiraly M, Davies R, Nagy B. Multidrug resistant clones of Salmonella Infantis of broiler origin in Europe. Int J Food Microbiol. 2012;157(1):108-12.

Nógrády N, Toth A, Kostyak A, Paszti J, Nagy B. Emergence of multidrug-resistant clones of Salmonella Infantis in broiler chickens and humans in Hungary. J Antimicrob Chemother. 2007;60(3):645-8.

Olasz F, Nagy T, Szabo M, Kiss J, Szmolka A, Barta E, et al. Genome Sequences of Three Salmonella enterica subsp. enterica Serovar Infantis Strains from Healthy Broiler Chicks in Hungary and in the United Kingdom. Genome Announc. 2015;3(1).

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3.

Parliament E, Council. Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. Official Journal L 325 2003a. p. 0031 - 0040.

Parliament E, Council. Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of salmonella and other specified food-borne zoonotic agents. Official Journal L 325 2003b. p. 0001 - 0015.

Pate M, Micunovic J, Golob M, Vestby LK, Ocepek M. Salmonella Infantis in Broiler Flocks in Slovenia: The Prevalence of Multidrug Resistant Strains with High Genetic Homogeneity and Low Biofilm-Forming Ability. Biomed Res Int. 2019;2019:4981463.

Perez-Sepulveda BM, Heavens D, Pulford CV, Predeus AV, Low R, Webster H, et al. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv. 2020:2020.07.22.200840.

Petzold M, Prior K, Moran-Gilad J, Harmsen D, Lück C. Epidemiological information is key when interpreting whole genome sequence data - lessons learned from a large Legionella pneumophila outbreak in Warstein, Germany, 2013. LID - 10.2807/1560-7917.ES.2017.22.45.17-00137 [doi] LID - 17-00137. 2017(1560-7917 (Electronic)).

Pilla G, Tang CM. Going around in circles: virulence plasmids in enteric pathogens. Nat Rev Microbiol. 2018;16(8):484-495.

Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641-50.

Proietti PC, Stefanetti V, Musa L, Zicavo A, Dionisi AM, Bellucci S, et al. Genetic Profiles and Antimicrobial Resistance Patterns of Salmonella Infantis Strains Isolated in Italy in the Food Chain of Broiler Meat Production. Antibiotics (Basel). 2020;9(11).

Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev. 2017;30(4):1015-1063.

Rabsch W, Simon S, Humphrey T. Public health Aspects of Salmonella Infections. In: Barrow PA, Methner U, editors. Salmonella in Domestic Animals. 2nd ed: CAB International; 2013. p. 547.

Rajic A, Keenliside J, McFall ME, Deckert AE, Muckle AC, O'Connor BP, et al. Longitudinal study of Salmonella species in 90 Alberta swine finishing farms. Vet Microbiol. 2005;105(1):47-56.

Rene SH, Susann KP, Pimlapas L, Burkhard M, Maria B, Antonio B, et al. Final report of ENGAGE - Establishing Next Generation sequencing Ability for Genomic analysis in Europe. EFSA supporting publication; 2018 (cited 2020 Dec 15):EN-1431. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2018.EN-1431>

Ribot EM, Hise KB. Future challenges for tracking foodborne diseases: PulseNet, a 20-year-old US surveillance system for foodborne diseases, is expanding both globally and technologically. EMBO Rep. 2016;17(11):1499-1505.

Robert Koch Institut (RKI) 2019. Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2018. Berlin 2019 Mar 01 (cited 2021 Feb 20). <https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch_2018.pdf?__blob=publicationFile>

Robert Koch Institut (RKI) 2020. Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2019. Berlin 2020 Mar 01 (cited 2021 Feb 20). <https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch_2019.pdf?__blob=publicationFile>

Robertson J, Yoshida C, Kruczkiewicz P, Nadon C, Nichani A, Taboada EN, et al. Comprehensive assessment of the quality of Salmonella whole genome sequence data available in public sequence databases using the Salmonella in silico Typing Resource (SISTR). Microb Genom. 2018;4(2).

Rotger R, Casadesús J. The virulence plasmids of Salmonella. 1999(1139-6709 (Print)).

Salipante SJ, Hall BG. Determining the limits of the evolutionary potential of an antibiotic resistance gene. Mol Biol Evol. 2003;20(4):653-9.

Salmon D, Smith T. The bacterium of swine-plague. Am Month Micr. 1886;7:204-205.

Sanchez-Romero MA, Merida-Floriano A, Casadesus J. Copy Number Heterogeneity in the Virulence Plasmid of Salmonella enterica. Front Microbiol. 2020;11:599931.

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94(3):441-8.

Seemann T. mlst GitHub. <https://github.com/tseemann/mlst>. 2014 (cited 2020 Dec 15).

Seemann T. Snippy GitHub. <https://github.com/tseemann/snippy>. 2014 (cited 2020 Dec 15).

Seemann T. Abricate Github. <https://github.com/tseemann/abricate>. 2015 (cited 2020 Dec 15).

Seemann T. Shovill GitHub. Assemble bacterial isolate genomes from Illumina paired-end reads. 2018 <https://github.com/tseemann/shovill> (cited 2020 Dec 15).

Shahada F, Chuma T, Okamoto K, Sueyoshi M. Temporal distribution and genetic fingerprinting of Salmonella in broiler flocks from southern Japan. Poultry Sci. 2008;87(5):968-972.

Simon S, Trost E, Bender J, Fuchs S, Malorny B, Rabsch W, et al. Evaluation of WGS based approaches for investigating a food-borne outbreak caused by Salmonella enterica serovar Derby in Germany. Food Microbiol. 2018;71:46-54.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121-132.

Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol. 2018;122(1):e59.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

Szmolka A, Szabo M, Kiss J, Paszti J, Adrian E, Olasz F, et al. Molecular epidemiology of the endemic multiresistance plasmid pSI54/04 of Salmonella Infantis in broiler and human population in Hungary. Food Microbiol. 2018;71:25-31.

Tang S, Orsi RH, Luo H, Ge C, Zhang G, Baker RC, et al. Assessment and Comparison of Molecular Subtyping and Characterization Methods for Salmonella. Front Microbiol. 2019;10:1591.

Tate H, Folster JP, Hsu CH, Chen J, Hoffmann M, Li C, et al. Comparative Analysis of Extended-Spectrum-beta-Lactamase CTX-M-65-Producing Salmonella enterica Serovar Infantis Isolates from Humans, Food Animals, and Retail Chickens in the United States. Antimicrob Agents Chemother. 2017;61(7).

Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, et al. MOST: a modified MLST typing tool based on short read sequencing. PeerJ. 2016;4:e2308.

Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res. 2018;3:93.

Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biology. 2014;15(11):524.

Tyson GH, Li C, Harrison LB, Martin G, Hsu CH, Tate H, et al. A Multidrug-Resistant Salmonella Infantis Clone Is Spreading and Recombining in the United States. Microbial drug resistance (Larchmont, NY). 2020.

Uelze L, Becker N, Borowiak M, Busch U, Dangel A, Deneke C, et al. Toward an Integrated Genome-Based Surveillance of Salmonella enterica in Germany. Frontiers in Microbiology. 2021;12:200.

Uelze L, Borowiak M, Deneke C, Szabo 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(5).

Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G, Bernard S, et al. Host adapted serotypes of Salmonella enterica. Epidemiol Infect. 2000;125(2):229-55.

Vallejos-Sanchez K, Tataje-Lavanda L, Villanueva-Perez D, Bendezu J, Montalvan A, Zimic-Peralta M, et al. Whole-Genome Sequencing of a Salmonella enterica subsp. enterica Serovar Infantis Strain Isolated from Broiler Chicken in Peru. Microbiol Resour Announc. 2019;8(43).

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-26.

Vinueza-Burgos C, Baquero M, Medina J, De Zutter L. Occurrence, genotypes and antimicrobial susceptibility of Salmonella collected from the broiler production chain within an integrated poultry company. Int J Food Microbiol. 2019;299:1-7.

Waldram A, Dolan G, Ashton PM, Jenkins C, Dallman TJ. Epidemiological analysis of Salmonella clusters identified by whole genome sequencing, England and Wales 2014. Food Microbiol. 2018;71:39-45.

Ward LR, Threfall J, Smith HR, O'Briend S, Kass P, Cliver D, et al. Salmonella enteritidis epidemic. Science. 2000;287(5459).

Wassenaar TM. Use of Antimicrobial Agents in Veterinary Medicine and Implications for Human Health. Crit. Rev. Microbiol. 2005;31(3):155-169.

Wilk T, Szabo M, Szmolka A, Kiss J, Barta E, Nagy T, et al. Genome Sequences of Multidrug-Resistant Salmonella enterica subsp. enterica Serovar Infantis Strains from Broiler Chicks in Hungary. Genome Announc. 2016;4(6).

Wilk T, Szabo M, Szmolka A, Kiss J, Olasz F, Nagy B. Genome Sequences of Salmonella enterica subsp. enterica Serovar Infantis Strains from Hungary Representing Two Peak Incidence Periods in Three Decades. Genome Announc. 2017;5(9).

Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. Nat Commun. 2016;7:12827.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.

Xu F, Ge C, Luo H, Li S, Wiedmann M, Deng X, et al. Evaluation of real-time nanopore sequencing for Salmonella serotype prediction. Food Microbiol. 2020;89:103452.

Yachison CA, Yoshida C, Robertson J, Nash JHE, Kruczkiewicz P, Taboada EN, et al. The Validation and Implications of Using Whole Genome Sequencing as a Replacement for Traditional Serotyping for a National Salmonella Reference Laboratory. Front Microbiol. 2017(1664-302X (Print)).

Yokoyama E, Ando N, Ohta T, Kanada A, Shiwa Y, Ishige T, et al. A novel subpopulation of Salmonella enterica serovar Infantis strains isolated from broiler chicken organs other than the gastrointestinal tract. Vet Microbiol. 2015;175(2-4):312-8.

Yokoyama E, Murakami K, Shiwa Y, Ishige T, Ando N, Kikuchi T, et al. Phylogenetic and population genetic analysis of Salmonella enterica subsp. enterica serovar Infantis strains isolated in Japan using whole genome sequence data. Infect Genet Evol. 2014;27:62-8.

Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, et al. 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(1):e0147101.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640-4.

Zascavage RR, Hall CL, Thorson K, Mahmoud M, Sedlazeck FJ, Planz JV. Approaches to Whole Mitochondrial Genome Sequencing on the Oxford Nanopore MinION. Curr Protoc Hum Genet. 2019;104(1):e94.

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(23).

Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, et al. Salmonella serotype determination utilizing high-throughput genome sequencing data. J Clin Microbiol. 2015;53(5):1685-92.

Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 2020;30(1):138-152.

8. APPENDIX

8.1. Supplementary Tables from the publication

Table S1. *S*. Infantis genomes for previously reported emergent *S*. infantis clones from Israel (AVIV et al. 2014, AVIV et al. 2016), Hungary (OLASZ et al. 2015, WILK et al. 2016, WILK et al. 2017), and Italy (FRANCO et al. 2015) used in this study and their general genome characteristics and genoserotyping (*Not applicable: N/A: CoG: contig; Cov: coverage).

ID isolate	Accesion Number	Reference	Source	Year of isolation	Region of isolation
335-3	ATHK00000000	Aviv et al., 2014	Human	1970	Israel
1326/28	LN649235	Olasz et al., 2015	Broiler	1973	United Kingdom
SI69/94	JRXB00000000	Olasz et al., 2015	Broiler	1994	Hungary
SI54/04	JRXC00000000	Olasz et al., 2015	Broiler	2004	Hungary
ERS846145	ERR1014111	Franco et al., 2015	Broiler	2006	Italy
ERS846143	ERR1014109	Franco et al., 2015	Broiler	2007	Italy
SI119944	GCA010919335	Aviv et al., 2014	Broiler	2008	Israel
ERS846152	ERR1014118	Franco et al., 2015	Broiler	2009	Italy
ERS846151	ERR1014117	Franco et al., 2015	Broiler	2012	Italy
SI3337/12	MIJS00000000	Wilk et al., 2016	Broiler	2012	Hungary
ERS846147	ERR1014113	Franco et al., 2015	Broiler	2013	Italy
ERS846148	ERR1014114	Franco et al., 2015	Broiler	2013	Italy
SI757/13	MIJT00000000	Wilk et al., 2016	Broiler	2013	Hungary
SI786/13	MIJR00000000	Wilk et al., 2016	Broiler	2013	Hungary
ERS846150	ERR1014116	Franco et al., 2015	Broiler	2014	Italy
SI1070/16	MRUX00000000	Wilk et al., 2017	Broiler	2016	Hungary
SI240/16	MRUW00000000	Wilk et al., 2017	Broiler	2016	Hungary

Continuation **Table S1**

ID isolate	Serovar	CoG	Cov	N50 (bp)	GC (%)	Genome size (Mbp)	ST	SISTR serovar
335-3	Infantis	79	N/A	152,157	52.3	4.60	32	Infantis
1326/28	Infantis	1	N/A	4,710,675	52.3	4.70	32	Infantis
SI69/94	Infantis	175	N/A	479,709	52.1	4.90	32	Infantis
SI54/04	Infantis	78	N/A	290,766	52.1	5.00	32	Infantis
ERS846145	Infantis	44	151	202,383	52.2	4.80	32	Infantis
ERS846143	Infantis	49	85	416,324	52.1	4.90	32	Infantis
SI119944	Infantis	2	N/A	4,725,957	52.2	5.00	32	Infantis
ERS846152	Infantis	47	99	416,371	52.2	4.90	32	Infantis
ERS846151	Infantis	47	127	386,563	52.2	4.90	32	Infantis
SI3337/12	Infantis	47	N/A	264,211	52.1	5.00	32	Infantis
ERS846147	Infantis	50	186	264,675	52.2	4.90	32	Infantis

Continuation **Table S1**

ID isolate	Serovar	CoG	Cov	N50 (bp)	GC (%)	Genome size (Mbp)	ST	SISTR serovar
ERS846148	Infantis	50	174	386,579	52.2	4.90	32	Infantis
SI757/13	Infantis	53	N/A	343,219	52.2	4.90	32	Infantis
SI786/13	Infantis	48	N/A	285,944	52.2	4.90	32	Infantis
ERS846150	Infantis	47	120	386,585	52.2	4.90	32	Infantis
SI1070/16	Infantis	58	N/A	445,038	52.1	4.90	32	Infantis
SI240/16	Infantis	53	N/A	444,648	52.1	5.00	32	Infantis

Table S2. Specific gene sequences downloaded for the creation of customized databases for ABRicate and available databases downloaded. These databases included sequences of genes encoding for SPIs, fimbrial clusters Ipf and K88-like and pESI encoded fitness determinants as the toxin-antitoxin system (CcdAB and PemKI/MazF), mercury operon, and allele sequences for incompatibility groupsplasmids IncI1 and IncF for plasmids typing (DB: database)

NAME GEN/LOCUS	STRAIN	COORDINATES	SIZE (bp)	ACCESION NUMBER	REFERENCE	PROTEIN DESCRIPTION	NAME DB
ccdA	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	49897..50115	219	CP047882	Cohen et al., 2020 (PMID: 32145019)	type II toxin-antitoxin system antitoxin CcdA	fitness_IS
ccdB	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	50117..50422	306	CP047882	Cohen et al., 2020 (PMID: 32145019)	type II toxin-antitoxin system toxin CcdB	fitness_IS
PemK/MazF	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	14789..15121	333	CP047882	Cohen et al., 2020 (PMID: 32145019)	type II toxin-antitoxin system PemK/MazF family toxin	fitness_IS
PemI-like	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	14531..14788	258	CP047882	Cohen et al., 2020 (PMID: 32145019)	antitoxin PemI-like	fitness_IS
merE	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(169042..169278)	237	CP047882	Cohen et al., 2020 (PMID: 32145019)	broad-spectrum mercury transporter MerE	fitness_IS
merD	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(169275..169637)	363	CP047882	Cohen et al., 2020 (PMID: 32145019)	mercury resistance co-regulator MerD	fitness_IS

Continuation **Table S2**

NAME GEN/LOCUS	STRAIN	COORDINATES	SIZE (bp)	ACCESION NUMBER	REFERENCE	PROTEIN DESCRIPTION	NAME DB
GVI52_23630_mercury(II)reductase	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(169655..171349)	1695	CP047882	Cohen et al., 2020 (PMID: 32145019)	merA mercury(II) reductase	fitness_IS
merC	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(171401..171823)	423	CP047882	Cohen et al., 2020 (PMID: 32145019)	organomercurial transporter MerC	fitness_IS
merP	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(171859..172134)	276	CP047882	Cohen et al., 2020 (PMID: 32145019)	mercury resistance system periplasmic binding protein MerP	fitness_IS
merT	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	complement(172148..172498)	351	CP047882	Cohen et al., 2020 (PMID: 32145019)	mercuric transport protein MerT	fitness_IS
merR	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	172570..173004	435	CP047882	Cohen et al., 2020 (PMID: 32145019)	Hg(II)-responsive transcriptional regulator	fitness_IS
faeC_GVI52_23065	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	38569..39108	540	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbrial protein	fimbrial_clusters

Continuation **Table S2**

NAME GEN/LOCUS	STRAIN	COORDINATES	SIZE (bp)	ACCESION NUMBER	REFERENCE	PROTEIN DESCRIPTION	NAME DB
faeD	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	39133..41565	2433	CP047882	Cohen et al., 2020 (PMID: 32145019)	F4 (K88) fimbrial usher FaeD	fimbrial_clusters
faeE_GVI52_23075	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	41588..42373	786	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbria/pilus periplasmic chaperone	fimbrial_clusters
faeF_GVI52_23080	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	42409..42900	492	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbrial protein	fimbrial_clusters
faeH	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	44110..44901	792	CP047882	Cohen et al., 2020 (PMID: 32145019)	F4 (K88) fimbria minor subunit FaeH	fimbrial_clusters
faeI_GVI52_23095	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	44929..45693	765	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbrial protein	fimbrial_clusters
ipfA_GVI52_23160	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	53160..53696	537	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbrial protein	fimbrial_clusters

Continuation **Table S2**

NAME GEN/LOCUS	STRAIN	COORDINATES	SIZE (bp)	ACCESION NUMBER	REFERENCE	PROTEIN DESCRIPTION	NAME DB
ipfB_GVI52_23165	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	53818..54480	663	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbria/pilus periplasmic chaperone	fimbrial_clusters
ipfC_GVI52_23170	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	54521..57073	2553	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbria/pilus outer membrane usher protein	fimbrial_clusters
ipfD_GVI52_23175	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	57230..58279	1050	CP047882	Cohen et al., 2020 (PMID: 32145019)	fimbrial protein	fimbrial_clusters
repB_GVI52_22855	Salmonella enterica subsp. enterica serovar Infantis strain 119944 plasmid pESI, complete sequence	1..1011	1011	CP047882	Cohen et al., 2020 (PMID: 32145019)	RepB family plasmid replication initiator protein	repB_pESI119944
incP_oriV	IncP-1alpha plasmid pBS228 complete sequence	1..89147	89147	AM261760	Haines AS et al., 2007 (PMID: 17320955)	-	incP_plasmids
incP_pTB11	uncultured bacterium pTB11 plasmid complete genome	1..68869	68869	AJ744860	Tennstedt T et al., 2005 (PMID: 15848226)	-	incP_plasmids
incP_Tn1723	Plasmid RK2 (from E.coli) DNA with transposon (Tn1723) insertion sites	1..811	811	M20134	Cross MA et al., 1986 (PMID: 3010353)	-	incP_plasmids

Continuation **Table S2**

NAME	ACCESION NUMBER	SIZE (bp)	START	END	STRAIN	DESCRIPTION
SPI-1	NC_006905_P5	43488	2960260	3003748	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome.	NC_006905.1:2960260-3003748 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-1	NC_003198_P5	41851	2858736	2900586	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:2858736-2900586 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-1	NC_004631_P2	41851	2844593	2886443	Salmonella enterica subsp. enterica serovar Typhi Ty2	NC_004631.1:2844593-2886443 Salmonella enterica subsp. enterica serovar Typhi str. Ty2, complete sequence
SPI-2	NC_006905_P3	41829	1497670	1539498	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome.	NC_006905.1:1497670-1539498 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-2	NC_003198_P3	41605	1624920	1666524	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:1624920-1666524 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-2	NC_004631_P1	41610	1314607	1356216	Salmonella enterica subsp. enterica serovar Typhi Ty2	NC_004631.1:1314607-1356216 Salmonella enterica subsp. enterica serovar Typhi str. Ty2, complete sequence
SPI-3	AF106566	17039	1	17039	Salmonella typhimurium pathogenicity island SPI-3, complete sequence.	AF106566.1 Salmonella typhimurium pathogenicity island SPI-3, complete sequence

Continuation **Table S2**

NAME	ACCESION NUMBER	SIZE (bp)	START	END	STRAIN	DESCRIPTION
SPI-3	NC_006905_P6	12819	3890879	3903697	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome.	NC_006905.1:3890879-3903697 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-3	NC_003198_P7	16941	3883613	3900553	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:3883613-3900553 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-4	AF060869	27290	1	27290	Salmonella enterica LT2	AF060869.1 Salmonella typhimurium excision nuclease UvrA (uvrA) gene, partial cds; single-strand binding protein (ssb) gene, complete cds; tRNA-Thr gene, complete sequence; pathogenicity island SPI-4 operon, complete sequence; yjcB gene, complete cds; and yjcC gene, partial cds
SPI-4	AJ576316	24660	1	24660	Salmonella enterica ST4/74	AJ576316.1 Salmonella typhimurium Salmonella Pathogenicity Island 4 siiABCDEF genes
SPI-4	NC_006905_P7	26698	4411902	4438599	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	NC_006905.1:4411902-4438599 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-4	NC_003198_P8	23391	4322993	4346383	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:4322993-4346383 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome

Continuation **Table S2**

NAME	ACCESION NUMBER	SIZE (bp)	START	END	STRAIN	DESCRIPTION
SPI-5	AF060858	9739	1	9739	Salmonella enterica 2229	AF060858.1 Salmonella dublin regulatory protein CopR (copR), histidine kinase (copS), SPI-4 pathogenicity island containing dipeptidase homolog (pipD), SopB (sopB), PipC (pipC), PipB (pipB), and PipA (pipA) genes, complete cds; and tRNA-Ser gene, complete sequence; and unknown genes
SPI-5	NC_006905_P1	5689	1155936	1161624	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	NC_006905.1:1155936-1161624 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-5	NC_003198_P2	7496 bp	1085068	1092563	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:1085068-1092563 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-6	NC_003198_P1	58666 bp	302092	360757	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:302092-360757 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-7	NC_003198_P9	133638 bp	4409511	4543148	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:4409511-4543148 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-7	NC_004631_P3	131749 bp	4394302	4526050	Salmonella enterica subsp. enterica serovar Typhi Ty2	NC_004631.1:4394302-4526050 Salmonella enterica subsp. enterica serovar Typhi str. Ty2, complete sequence
SPI-8	NC_003198_P6	6885	3132530	3139414	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:3132530-3139414 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome

Continuation **Table S2**

NAME	ACCESION NUMBER	SIZE (bp)	START	END	STRAIN	DESCRIPTION
SPI-9	NC_003198_P4	15696	2743495	2759190	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:2743495-2759190 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-10	NC_003198_P10	32934	4683605	4716538	Salmonella enterica subsp. enterica serovar Typhi CT18	NC_003198.1:4683605-4716538 Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome
SPI-11	NC_006905_P2	15686	1350481	1366166	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome.	NC_006905.1:1350481-1366166 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
SPI-12	NC_006905_P4	11075	2354604	2365678	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome.	NC_006905.1:2354604-2365678 Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67, complete genome
CS54	AF140550	25252	1	25252	Salmonella enterica ATCC14028	AF140550 AF140550.2 Salmonella typhimurium exonuclease VII (xseA), ShdA (shdA), RatC (ratC), RatB (ratB), RatA (ratA), SinI (sinI), and SinH (sinH) genes, complete cds; and YfgK (yfgK) gene, partial cds

Continuation **Table S2**

DATABASE	**SOURCE**
incI1_plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
incF_plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/incf.fsa
PlasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ResFinder	https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/
VFDB	http://www.mgc.ac.cn/VFs/
NCBI_database	https://www.ncbi.nlm.nih.gov/bioproject/313047
paidb*	http://www.paidb.re.kr/browse_pais.php?m=p

Table S3. Whole-Genome Sequencing general characteristics and genoserotyping of the *S.* Infantis strains from Germany used in this study.

		General genome characteristics				Kraken classification			
ID isolate	**ID strain**	**Total reads (bp)**	**Total contigs**	**Coverage**	**N50 (bp)**	**Identified species**	**% 1st match**	**Identified genera**	**% 1st match**
2945	19PM0346	1,811,804	42	93	264.378	*Salmonella enterica*	97.1	Salmonella	97.79
2947	19PM0348	1,771,342	46	90	332.767	*Salmonella enterica*	91.87	Salmonella	92.5
2948	19PM0349	2,124,116	51	107	201.973	*Salmonella enterica*	97.12	Salmonella	97.83
2949	19PM0350	1,945,002	54	102	240.386	*Salmonella enterica*	87.27	Salmonella	87.83
2951	19PM0351	2,509,738	37	128	353.754	*Salmonella enterica*	97.15	Salmonella	97.81
3222	20PM0240	991.352	49	47	204.015	*Salmonella enterica*	96.41	Salmonella	97.21
3225	20PM0243	1,470,222	46	67	304.608	*Salmonella enterica*	96.04	Salmonella	96.96
3227	20PM0245	2,191,574	38	96	397.341	*Salmonella enterica*	95.96	Salmonella	97.03
3230	20PM0248	524.688	85	24	90.139	*Salmonella enterica*	96.41	Salmonella	97.27
3234	20PM0252	832.998	91	37	90.406	*Salmonella enterica*	95.5	Salmonella	96.64
3239	20PM0257	90.067	52	42	184.048	*Salmonella enterica*	95.86	Salmonella	96.71
3242	20PM0260	1,181,000	60	57	138.816	*Salmonella enterica*	96.64	Salmonella	97.51
3243	20PM0261	1,270,948	53	59	204.015	*Salmonella enterica*	94.14	Salmonella	94.99
3245	20PM0263	1,049,692	50	50	202.006	*Salmonella enterica*	95.21	Salmonella	96.03

Continuation **Table S3**

		General genome characteristics				Kraken classification			
3249	20PM0267	1,522,276	52	73	213.614	*Salmonella enterica*	94.59	Salmonella	95.34
3250	20PM0268	1,271,966	45	59	229.154	*Salmonella enterica*	96.4	Salmonella	97.42
3252	20PM0270	793.374	89	35	123.937	*Salmonella enterica*	93.52	Salmonella	94.55
3253	20PM0271	560.54	86	26	118.236	*Salmonella enterica*	94.31	Salmonella	95.15
3255	20PM0273	1,163,972	48	55	202.039	*Salmonella enterica*	93.11	Salmonella	94.05
3257	20PM0275	660.432	56	32	212.700	*Salmonella enterica*	94.83	Salmonella	95.52
2954	19PM0355	1,159,090	50	52	222.196	*Salmonella enterica*	93.72	Salmonella	94.79
2957	19PM0358	1,529,268	52	76	353.112	*Salmonella enterica*	94.76	Salmonella	95.39
2959	19PM0360	1,357,154	49	64	255025	*Salmonella enterica*	93.95	Salmonella	94.92
2747	19PM0148	2,071,834	49	92	416.366	*Salmonella enterica*	93.51	Salmonella	94.56
2748	19PM0149	2,031,940	48	90	416.095	Salmonella *enterica*	94.39	*Salmonella*	93.32
2749	19PM0150	2,022,404	46	91	416.184	Salmonella	94.59	*Salmonella enterica*	93.58
2750	19PM0151	1,301,816	51	56	333.573	Salmonella	89.8	*Salmonella enterica*	88.75
2752	19PM0153	1,629,922	48	72	333.259	Salmonella	94.22	*Salmonella enterica*	93.15
2753	19PM0154	1,676,256	51	74	232.037	Salmonella	94.61	*Salmonella enterica*	93.51
3027	20PM0045	1,861,120	51	99	445.475	Salmonella	96.18	*Salmonella enterica*	95.71

Continuation **Table S3**

		Kraken classification		SISTR *in silico* serovar prediction						Typing
ID	**ID strain**	**Identified species**	**% 1st match**	**O antigen**	**h1 antigen**	**h2 antigen**	**Serovar**	**Serovar_antigen**	**Serovar_cg MLST**	**MLST (ST)**
2945	19PM0346	*Salmonella enterica*	97.1	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2947	19PM0348	*Salmonella enterica*	91.87	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2948	19PM0349	*Salmonella enterica*	97.12	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2949	19PM0350	*Salmonella enterica*	87.27	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2951	19PM0351	*Salmonella enterica*	97.15	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3222	20PM0240	*Salmonella enterica*	96.41	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3225	20PM0243	*Salmonella enterica*	96.04	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3227	20PM0245	*Salmonella enterica*	95.96	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3230	20PM0248	*Salmonella enterica*	96.41	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3234	20PM0252	*Salmonella enterica*	95.5	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32

Continuation **Table S3**

		Kraken classification		SISTR *in silico* serovar prediction						Typing
3239	20PM0257	*Salmonella enterica*	95.86	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3242	20PM0260	*Salmonella enterica*	96.64	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3243	20PM0261	*Salmonella enterica*	94.14	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3245	20PM0263	*Salmonella enterica*	95.21	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3249	20PM0267	*Salmonella enterica*	94.59	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3250	20PM0268	*Salmonella enterica*	96.4	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3252	20PM0270	*Salmonella enterica*	93.52	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3253	20PM0271	*Salmonella enterica*	94.31	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3255	20PM0273	*Salmonella enterica*	93.11	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3257	20PM0275	*Salmonella enterica*	94.83	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2954	19PM0355	*Salmonella enterica*	93.72	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32

Continuation **Table S3**

		Kraken classification		SISTR *in silico* serovar prediction						Typing
2957	19PM0358	*Salmonella enterica*	94.76	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2959	19PM0360	*Salmonella enterica*	93.95	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2747	19PM0148	*Salmonella enterica*	93.51	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2748	19PM0149	*Salmonella enterica*	93.32	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2749	19PM0150	*Salmonella enterica*	93.58	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2750	19PM0151	*Salmonella enterica*	88.75	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2752	19PM0153	*Salmonella enterica*	93.15	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
2753	19PM0154	*Salmonella enterica*	93.51	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32
3027	20PM0045	*Salmonella enterica*	95.71	6,7,14	r	1,5	Infantis	Infantis|Senegal	Infantis	32

Table S4. Resistance genes and chromosomal point mutations found among the *S. infantis* strains used in this study.

SEQUENCE	ISOLATE	COUNTRY	YEAR	GENES FOUND	*aac(3)-VIa*	*aac(6')-Iaa*	*aadA13*	*aadA2*	*ant(3'')-Ia*	*aph(3')-Ia*	*blaCTX-M-14*	*blaCTX-M-1*	*blaTEM-104*	*cmlA1*	*dfrA14*	*dfrA1*	*anrS1*	*sul1*	*sul3*	*tet(A)*	*qacEdelta1*	*gyrA-S83Y*	*gyrA-D87G*	*gyrA-D87Y*
ATHK000000 00	335-3	Israel	1970	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
LN649235	1326/28	UK	1973	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
19PM0346	2945	Germany	1992	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
19PM0348	2947	Germany	1992	4	100	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	-	100		-	-
19PM0349	2948	Germany	1992	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
19PM0350	2949	Germany	1992	4	100	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	-	100		-	-
19PM0351	2951	Germany	1994	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0240	3222	Germany	1994	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
JRXB000000 00	SI69/94	Hungary	1994	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0243	3225	Germany	1995	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0245	3227	Germany	1995	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0248	3230	Germany	1996	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0252	3234	Germany	1997	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0257	3239	Germany	1998	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0260	3242	Germany	1998	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
JRXC000000 00	SI54/04	Hungary	2004	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	92.24	100	-	-
ERR1014111	ERS846145	Italy	2006	6	-	100	100	97.92	82.51	-	-	-	-	100	-	-	-	.	100	-	-	-	-	-
ERR1014109	ERS846143	Italy	2007	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	99.89	100	-	-
GCA010919335	SI119944	Israel	2008	5	-	100	-	-	99.59	-	-	-	-	-	100	-	-	100	-	97.8	98.17	-	-	100
ERR1014118	ERS846152	Italy	2009	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.61	-	-
ERR1014117	ERS846151	Italy	2012	7	-	100	-	-	.	100	-	100	-	-	100	100	-	100	-	97.8	98.17	-	100	-

Continuation **Table S4**

SEQUENCE	ISOLATE	COUNTRY	YEAR	GENES FOUND	*aac(3)-VIa*	*aac(6')-Iaa*	*aadA13*	*aadA2*	*ant(3'')-Ia*	*aph(3')-Ia*	*blaCTX-M-14*	*blaCTX-M-1*	*blaTEM-104*	*cmlA1*	*dfrA14*	*dfrA1*	*qnrS1*	*sul1*	*sul3*	*tet(A)*	*qacEdelta1*	*gyrA-S83Y*	*gyrA-D87G*	*gyrA-D87Y*
MIJS000000 00	SI3337/12	Hungary	2012	5	-	100	-	-	99.59	-	-	-	82.46	-	-	-	-	100	-	97.8	100	97.17	-	-
ERR1014113	ERS846147	Italy	2013	7	-	100	-	-	82.51	-	-	100	-	-	100	100	-	100	-	97.8	100	-	92.24	-
ERR1014114	ERS846148	Italy	2013	7	-	100	-	-	.	100	-	100	-	-	100	100	-	100	-	97.8	98.17	-	99.89	-
ERR1014109	ERS846143	Italy	2007	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	99.89	100	-	-
GCA0109193 35	SI119944	Israel	2008	5	-	100	-	-	99.59	-	-	-	-	-	100	-	-	100	-	97.8	98.17	-	-	100
ERR1014118	ERS846152	Italy	2009	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.61	-	-
ERR1014117	ERS846151	Italy	2012	7	-	100	-	-	.	100	-	100	-	-	100	100	-	100	-	97.8	98.17	-	100	-
MIJS000000 00	SI3337/12	Hungary	2012	5	-	100	-	-	99.59	-	-	-	82.46	-	-	-	-	100	-	97.8	100	97.17	-	-
ERR1014113	ERS846147	Italy	2013	7	-	100	-	-	82.51	-	-	100	-	-	100	100	-	100	-	97.8	100	-	92.24	-
ERR1014114	ERS846148	Italy	2013	7	-	100	-	-	.	100	-	100	-	-	100	100	-	100	-	97.8	98.17	-	99.89	-
MIJT000000 00	SI757/13	Hungary	2013	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	97.17	-	-
MIJR000000 00	SI786/13	Hungary	2013	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	98.17	-	-
20PM0261	3243	Germany	2014	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.61	-	-
20PM0263	3245	Germany	2014	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
ERR1014116	ERS846150	Italy	2014	6	-	100	-	-	82.51	-	-	100	-	-	-	100	-	100	-	97.8	100	-	100	-
20PM0267	3249	Germany	2015	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	98.28	-	-
20PM0268	3250	Germany	2015	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0270	3252	Germany	2015	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.89	-	-
20PM0271	3253	Germany	2016	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	97.17	-	-
20PM0273	3255	Germany	2016	1	-	100	-	-	.	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-
20PM0275	3257	Germany	2016	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	98.28	-	-

Continuation **Table S4**

SEQUENCE	ISOLATE	COUNTRY	YEAR	GENES FOUND	*aac(3)-VIa*	*aac(6')-Iaa*	*aadA13*	*aadA2*	*ant(3'')-Ia*	*aph(3')-Ia*	*blaCTX-M-14*	*blaCTX-M-1*	*blaTEM-104*	*cmlA1*	*dfrA14*	*dfrA1*	*qnrS1*	*sul1*	*sul3*	*tet(A)*	*qacEdelta1*	*gyrA-S83Y*	*gyrA-D87G*	*gyrA-D87Y*
MRUX00000000	SI1070/16	Hungary	2016	2	-	100	-	-	.	-	100	-	-	-	-	-	-	.	-	-	-	98	-	-
MRUW00000000	SI240/16	Hungary	2016	6	-	100	-	-	99.59	-	-	-	82.46	-	-	-	100	100	-	97.8	100	97.17	-	-
19PM0355	2954	Germany	2017	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	100	-	-
19PM0358	2957	Germany	2018	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	93.75	98.28	-	-
19PM0148	2747	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	90.11	98.28	-	-
19PM0149	2748	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.61	-	-
19PM0150	2749	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	90.11	98.28	-	-
19PM0151	2750	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	90.11	99.91	-	-
19PM0153	2752	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.91	-	-
19PM0154	2753	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	98.28	-	-
19PM0360	2959	Germany	2019	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	99.61	-	-
20PM0045	3027	Germany	2020	4	-	100	-	-	99.59	-	-	-	-	-	-	-	-	100	-	97.8	100	98.28	-	-

Table S5. Virulence genes, SPIs, fimbrial cluster genes, and fitness genes found among the *S.* Infantis strains used in this study (CoG: contig; Cov: coverage; ID: identity; DB: database).

SEQUENCE	ISOLATE	COUNTRY	YEAR	GENES FOUND	*avrA*	*csgA*	*csgB*	*csgC*	*csgD*	*csgE*	*csgF*	*csgG*	*entB*	*faeD*	*faeE*
ATHK00000000	335-3	Israel	1970	104	100	100	100	100	100	100	100	100	99.18	-	-
LN649235	1326/28	UK	1973	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0346	2945	Germany	1992	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0348	2947	Germany	1992	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0349	2948	Germany	1992	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0350	2949	Germany	1992	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0351	2951	Germany	1994	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0240	3222	Germany	1994	104	100	100	100	100	100	100	100	100	99.18	-	-
JRXB00000000	SI69/94	Hungary	1994	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0243	3225	Germany	1995	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0245	3227	Germany	1995	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0248	3230	Germany	1996	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0252	3234	Germany	1997	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0257	3239	Germany	1998	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0260	3242	Germany	1998	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
JRXC00000000	SI54/04	Hungary	2004	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
ERR1014111	ERS846145	Italy	2006	104	100	100	100	100	100	100	100	100	99.18	-	-
ERR1014109	ERS846143	Italy	2007	104	100	100	100	100	100	100	100	100	99.18	-	-
GCA010919335	119944	Israel	2008	104	100	100	100	100	100	100	100	100	99.18	-	-
ERR1014118	ERS846152	Italy	2009	104	100	100	100	100	100	100	100	100	99.18	-	-
ERR1014117	ERS846151	Italy	2012	103	100	100	100	100	100	100	100	100	99.18	-	-
MIJS00000000	SI3337/12	Hungary	2012	104	100	100	100	100	100	100	100	100	99.18	-	-
ERR1014113	ERS846147	Italy	2013	104	100	100	100	100	100	100	100	100	99.18	-	-
ERR1014114	ERS846148	Italy	2013	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
MIJT00000000	SI757/13	Hungary	2013	104	100	100	100	100	100	100	100	100	99.18	-	-
MIJR00000000	SI786/13	Hungary	2013	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25

Continuation **Table S5**

SEQUENCE	ISOLATE	COUNTRY	YEAR	GENES FOUND	*avrA*	*csgA*	*csgB*	*csgC*	*csgD*	*csgE*	*csgF*	*csgG*	*entB*	*faeD*	*faeE*
20PM0261	3243	Germany	2014	103	100	100	100	100	100	100	100	100	99.18	-	-
20PM0263	3245	Germany	2014	116	100	100	100	100	100	100	100	100	99.18	96.60	91.25
ERR1014116	ERS846150	Italy	2014	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0267	3249	Germany	2015	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0268	3250	Germany	2015	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0270	3252	Germany	2015	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0271	3253	Germany	2016	104	100	100	100	100	100	100	100	100	99.18	-	-
20PM0273	3255	Germany	2016	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0275	3257	Germany	2016	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
MRUX00000000	SI1070/16	Hungary	2016	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
MRUW00000000	SI240/16	Hungary	2016	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0355	2954	Germany	2017	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0358	2957	Germany	2018	103	100	100	100	100	100	100	100	100	99.18	-	-
19PM0360	2959	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0148	2747	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0149	2748	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0150	2749	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0151	2750	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0153	2752	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
19PM0154	2753	Germany	2019	117	100	100	100	100	100	100	100	100	99.18	96.60	91.25
20PM0045	3027	Germany	2020	104	100	100	100	100	100	100	100	100	99.18	-	-

Continuation **Table S5**

SEQUENCE	ISOLATE	*fepC*	*fepG*	*fimC*	*fimD*	*fimF*	*fimH*	*fimI*	*fyuA*	*invA*	*invB*	*invC*	*invE*	*invF*	*invG*	*invH*	*invI*	*invJ*
ATHK00000000	335-3	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
LN649235	1326/28	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0349	2948	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0350	2949	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
JRXB00000000	SI69/94	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0243	3225	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0245	3227	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0248	3230	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0252	3234	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014111	ERS846145	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014109	ERS846143	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
GCA010919335	119944	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014118	ERS846152	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014117	ERS846151	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
MIJS00000000	SI3337/12	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014113	ERS846147	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014114	ERS846148	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
MIJR00000000	SI786/13	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE	ISOLATE	*fepC*	*fepG*	*fimC*	*fimD*	*fimF*	*fimH*	*fimI*	*fyuA*	*invA*	*invB*	*invC*	*invE*	*invF*	*invG*	*invH*	*invI*	*invJ*
20PM0261	3243	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0263	3245	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0268	3250	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0270	3252	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
20PM0273	3255	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
19PM0360	2959	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0150	2749	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	96.45	91.54	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	96.45	91.54	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*irp1*	*irp2*	*lpfA*	*lpfB*	*lpfC*	*lpfD*	*lpfE*	*mgtB*	*mgtC*	*mig.14*	*misL*	*ompA*	*orgA*	*orgB*	*orgC*	*pipB*
ATHK00000000	335-3	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LN649235	1326/28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0349	2948	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0350	2949	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXB00000000	SI69/94	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0243	3225	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0245	3227	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0248	3230	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0252	3234	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014111	ERS846145	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014109	ERS846143	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GCA010919335	119944	-	-	100	100	100	100	100	100	100	100	100	100	100	99.85	100	100
ERR1014118	ERS846152	-	-	100	100	100	100	100	100	100	100	100	100	100	99.85	100	100
ERR1014117	ERS846151	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJS00000000	SI3337/12	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014113	ERS846147	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014114	ERS846148	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJR00000000	SI786/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0261	3243	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0263	3245	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0268	3250	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*irp1*	*irp2*	*lpfA*	*lpfB*	*lpfC*	*lpfD*	*lpfE*	*mgtB*	*mgtC*	*mig.14*	*misL*	*ompA*	*orgA*	*orgB*	*orgC*	*pipB*
20PM0270	3252	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0273	3255	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	-	-	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0360	2959	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0150	2749	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	-	-	100; 100	100	100	100	100	100	100	100	100	100	100	100	100	100

SEQUENCE ID	ISOLATE ID	*pipB2*	*prgH*	*prgI*	*prgJ*	*prgK*	*ratB*	*shdA*	*sicA*	*sicP*	*sifA*	*sifB*	*sinH*	*sipA.sspA*	*sipB.sspB*	*sipC.sspC*	*sipD*
ATHK00000000	335-3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LN649235	1326/28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*pipB2*	*prgH*	*prgI*	*prgJ*	*prgK*	*ratB*	*shdA*	*sicA*	*sicP*	*sifA*	*sifB*	*sinH*	*sipA.sspA*	*sipB.sspB*	*sipC.sspC*	*sipD*
19PM0349	2948	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0350	2949	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXB00000000	SI69/94	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0243	3225	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0245	3227	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0248	3230	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0252	3234	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014111	ERS846145	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014109	ERS846143	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GCA010919335	119944	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014118	ERS846152	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014117	ERS846151	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJS00000000	SI3337/12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014113	ERS846147	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014114	ERS846148	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJR00000000	SI786/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0261	3243	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0263	3245	100	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*pipB2*	*prgH*	*prgI*	*prgJ*	*prgK*	*ratB*	*shdA*	*sicA*	*sicP*	*sifA*	*sifB*	*sinH*	*sipA.sspA*	*sipB.sspB*	*sipC.sspC*	*sipD*
20PM0268	3250	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0270	3252	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0273	3255	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	100	100	100	100	100	100	-	100	100	100	100	100	100	100	100	100
19PM0360	2959	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0150	2749	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	100	100	100	100	100	100	83.32	100	100;100	100	100	100	100	100	100	100

SEQUENCE ID	ISOLATE ID	*slrP*	*sopA*	*sopB.sigD*	*sopD*	*sopD2*	*sopE2*	*spaO*	*spaP*	*spaQ*	*spaR*	*spaS*	*spiC.ssaB*	*sptP*	*ssaC*	*ssaD*	*ssaE*
ATHK00000000	335-3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LN649235	1326/28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0349	2948	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*slrP*	*sopA*	*sopB.sigD*	*sopD*	*sopD2*	*sopE2*	*spaO*	*spaP*	*spaQ*	*spaR*	*spaS*	*spiC.ssaB*	*sptP*	*ssaC*	*ssaD*	*ssaE*
19PM0350	2949	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXB00000000	SI69/94	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0243	3225	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0245	3227	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0248	3230	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0252	3234	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014111	ERS846145	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014109	ERS846143	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GCA010919335	119944	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014118	ERS846152	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014117	ERS846151	100	89.70	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJS00000000	SI3337/12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014113	ERS846147	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014114	ERS846148	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJR00000000	SI786/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0261	3243	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0263	3245	100	100	100	100	88.54	100	100	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*slrP*	*sopA*	*sopB.sigD*	*sopD*	*sopD2*	*sopE2*	*spaO*	*spaP*	*spaQ*	*spaR*	*spaS*	*spiC.ssaB*	*sptP*	*ssaC*	*ssaD*	*ssaE*
20PM0268	3250	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0270	3252	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0273	3255	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0360	2959	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0150	2749	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	100	100	100	100	100;100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	100	100	100	100	100	100	100	100	100	100	100	100;100	100	100	100	100

SEQUENCE ID	ISOLATE ID	*ssaG*	*ssaH*	*ssaI*	*ssaJ*	*ssaK*	*ssaL*	*ssaM*	*ssaN*	*ssaO*	*ssaP*	*ssaQ*	*ssaR*	*ssaS*	*ssaT*	*ssaU*	*ssaV*	*sscA*
ATHK00000000	335-3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LN649235	1326/28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0349	2948	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0350	2949	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*ssaG*	*ssaH*	*ssaI*	*ssaJ*	*ssaK*	*ssaL*	*ssaM*	*ssaN*	*ssaO*	*ssaP*	*ssaQ*	*ssaR*	*ssaS*	*ssaT*	*ssaU*	*ssaV*	*sscA*
JRXB00000000	SI69/94	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0243	3225	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0245	3227	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0248	3230	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0252	3234	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014111	ERS846145	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014109	ERS846143	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GCA010919335	119944	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014118	ERS846152	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014117	ERS846151	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJS00000000	SI3337/12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014113	ERS846147	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014114	ERS846148	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJR00000000	SI786/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0261	3243	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0263	3245	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0268	3250	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0270	3252	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0273	3255	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	100	100	100	100	100	99.80	100	100	100	100	99.79	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*ssaG*	*ssaH*	*ssaI*	*ssaJ*	*ssaK*	*ssaL*	*ssaM*	*ssaN*	*ssaO*	*ssaP*	*ssaQ*	*ssaR*	*ssaS*	*ssaT*	*ssaU*	*ssaV*	*sscA*
19PM0360	2959	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0150	2749	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

SEQUENCE ID	ISOLATE ID	*sscB*	*sseA*	*sseB*	*sseC*	*sseD*	*sseE*	*sseF*	*sseG*	*sseJ*	*sseK1*	*sseK2*	*sseL*	*sspH2*	*steA*	*steB*	*steC*	*ybtA*
ATHK00000000	335-3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
LN649235	1326/28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0346	2945	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0348	2947	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0349	2948	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0350	2949	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0351	2951	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0240	3222	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
JRXB00000000	SI69/94	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0243	3225	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0245	3227	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0248	3230	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0252	3234	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0257	3239	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0260	3242	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
JRXC00000000	SI54/04	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*sscB*	*sseA*	*sseB*	*sseC*	*sseD*	*sseE*	*sseF*	*sseG*	*sseJ*	*sseK1*	*sseK2*	*sseL*	*sspH2*	*steA*	*steB*	*steC*	*ybtA*
ERR1014111	ERS846145	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
ERR1014109	ERS846143	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
GCA010919335	119944	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
ERR1014118	ERS846152	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
ERR1014117	ERS846151	100	100	100	100	100	100	100	100	100	100	100	-	100	100	100	100	-
MIJS00000000	SI3337/12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
ERR1014113	ERS846147	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
ERR1014114	ERS846148	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MIJT00000000	SI757/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
MIJR00000000	SI786/13	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0261	3243	100	100	100	100	100	100	100	100	100	100	100	100	-	100	100	100	-
20PM0263	3245	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ERR1014116	ERS846150	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0267	3249	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0268	3250	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0270	3252	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0271	3253	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
20PM0273	3255	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0275	3257	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUX00000000	SI1070/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
MRUW00000000	SI240/16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0355	2954	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0358	2957	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-
19PM0360	2959	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0148	2747	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0149	2748	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*sscB*	*sseA*	*sseB*	*sseC*	*sseD*	*sseE*	*sseF*	*sseG*	*sseJ*	*sseK1*	*sseK2*	*sseL*	*sspH2*	*steA*	*steB*	*steC*	*ybtA*
19PM0150	2749	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0151	2750	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0153	2752	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
19PM0154	2753	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
20PM0045	3027	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	-

SEQUENCE ID	ISOLATE ID	*ybtE*	*ybtP*	*ybtQ*	*ybtS*	*ybtT*	*ybtU*	*ybtX*
ATHK00000000	335-3	-	-	-	-	-	-	-
LN649235	1326/28	100	100	100	99.62	100	100	100
19PM0346	2945	100	100	100	99.62	100	100	100
19PM0348	2947	100	100	100	99.62	100	100	100
19PM0349	2948	100	100	100	99.62	100	100	100
19PM0350	2949	100	100	100	99.62	100	100	100
19PM0351	2951	100	100	100	99.62	100	100	100
20PM0240	3222	-	-	-	-	-	-	-
JRXB00000000	SI69/94	-	-	-	-	-	-	-
20PM0243	3225	-	-	-	-	-	-	-
20PM0245	3227	-	-	-	-	-	-	-
20PM0248	3230	-	-	-	-	-	-	-
20PM0252	3234	100	100	100	99.62	100	100	100
20PM0257	3239	100	100	100	99.62	100	100	100
20PM0260	3242	100	100	100	99.62	100	100	100
JRXC00000000	SI54/04	100	100	100	99.62	100	100	100
ERR1014111	ERS846145	-	-	-	-	-	-	-

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*ybtE*	*ybtP*	*ybtQ*	*ybtS*	*ybtT*	*ybtU*	*ybtX*
ERR1014109	ERS846143	-	-	-	-	-	-	-
GCA010919335	119944	-	-	-	-	-	-	-
ERR1014118	ERS846152	-	-	-	-	-	-	-
ERR1014117	ERS846151	-	-	-	-	-	-	-
MIJS00000000	SI3337/12	-	-	-	-	-	-	-
ERR1014113	ERS846147	-	-	-	-	-	-	-
ERR1014114	ERS846148	100	100	100	99.62	100	100	100
MIJT00000000	SI757/13	-	-	-	-	-	-	-
MIJR00000000	SI786/13	100	100	100	99.62	100	100	100
20PM0261	3243	-	-	-	-	-	-	-
20PM0263	3245	100	100	100	99.62	100	100	100
ERR1014116	ERS846150	100	100	100	99.62	100	100	100
20PM0267	3249	-	-	-	-	-	-	-
20PM0268	3250	100	100	100	99.62	100	100	100
20PM0270	3252	100	100	100	99.62	100	100	100
20PM0271	3253	-	-	-	-	-	-	-
20PM0273	3255	100	100	100	99.62	100	100	100
20PM0275	3257	100	100	100	99.62	100	100	100
MRUX00000000	SI1070/16	100	100	100	99.62	100	100	100
MRUW00000000	SI240/16	100	100	100	99.62	100	100	100
19PM0355	2954	100	100	100	99.62	100	100	100
19PM0358	2957	-	-	-	-	-	-	-
19PM0360	2959	100	100	100	99.62	100	100	100
19PM0148	2747	100	100	100	99.62	100	100	100
19PM0149	2748	100	100	100	99.62	100	100	100

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	*ybtE*	*ybtP*	*ybtQ*	*ybtS*	*ybtT*	*ybtU*	*ybtX*
19PM0149	2748	100	100	100	99.62	100	100	100
19PM0150	2749	100	100	100	99.62	100	100	100
19PM0151	2750	100	100	100	99.62	100	100	100
19PM0153	2752	100	100	100	99.62	100	100	100
19PM0154	2753	100	100	100	99.62	100	100	100
20PM0045	3027	-	-	-	-	-	-	-

SEQUENCE ID	ISOLATE ID	COUNTRY	YEAR	SPI-1	SPI-2	SPI-3	SPI-4	SPI-5	SPI-6	SPI-7	SPI-8	SPI-9	SPI-10	SPI-11	SPI-12	CS54
ATHK00000000	335-3	Israel	1970	+	+	+	+	+	+	-	-	+	-	+	+	+
132628	1326/28	UK	1973	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0346	2945	Germany	1992	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0348	2947	Germany	1992	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0349	2948	Germany	1992	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0350	2949	Germany	1992	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0351	2951	Germany	1994	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0240	3222	Germany	1994	+	+	+	+	+	+	-	-	+	-	+	+	+
JRXB00000000	SI69/94	Hungary	1994	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0243	3225	Germany	1995	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0245	3227	Germany	1995	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0248	3230	Germany	1996	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0252	3234	Germany	1997	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0257	3239	Germany	1998	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0260	3242	Germany	1998	+	+	+	+	+	+	-	-	+	-	+	+	+
JRXC00000000	SI54/04	Hungary	2004	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014111	ERS846145	Italy	2006	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014109	ERS846143	Italy	2007	+	+	+	+	+	+	-	-	+	-	+	+	+
GCA010919335	SI119944	Israel	2008	+	+	+	+	+	+	-	-	+	-	+	+	+

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	COUNTRY	YEAR	SPI-1	SPI-2	SPI-3	SPI-4	SPI-5	SPI-6	SPI-7	SPI-8	SPI-9	SPI-10	SPI-11	SPI-12	CS54
ERR1014118	ERS846152	Italy	2009	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014117	ERS846151	Italy	2012	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJS00000000	SI3337/12	Hungary	2012	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014113	ERS846147	Italy	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014114	ERS846148	Italy	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJT00000000	SI757/13	Hungary	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJR00000000	SI786/13	Hungary	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0261	3243	Germany	2014	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0263	3245	Germany	2014	+	+	+	+	+	+	-	-	+	-	+	+	+
ERR1014116	ERS846150	Italy	2014	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0267	3249	Germany	2015	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0268	3250	Germany	2015	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0270	3252	Germany	2015	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0271	3253	Germany	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0273	3255	Germany	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0275	3257	Germany	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
MRUX00000000	SI1070/16	Hungary	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
MRUW00000000	SI240/16	Hungary	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0355	2954	Germany	2017	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0358	2957	Germany	2018	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0148	2747	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0149	2748	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0150	2749	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0151	2750	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0153	2752	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0154	2753	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
19PM0360	2959	Germany	2019	+	+	+	+	+	+	-	-	+	-	+	+	+
20PM0045	3027	Germany	2020	+	+	+	+	+	+	-	-	+	-	+	+	+

Continuation **Table S5**

SEQUENCE ID	ISOLATE ID	COUNTRY	YEAR	SPI-1	SPI-2	SPI-3	SPI-4	SPI-5	SPI-6	SPI-7	SPI-8	SPI-9	SPI-10	SPI-11	SPI-12	CS54
ERR1014119	ERS846153	Italy	2014	+	+	+	+	+	+	-	-	+	-	+	+	+
MRUX00000000	SI1070/16 (GCA001906515)	Hungary	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
MRUW00000000	SI240/16 (GCA001906535)	Hungary	2016	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJS00000000	SI3337/12 (GCA001766495)	Hungary	2012	+	+	+	+	+	+	-	-	+	-	+	+	+
JRXC00000000	SI54/04	Hungary	2004	+	+	+	+	+	+	-	-	+	-	+	+	+
JRXB00000000	SI69/94	Hungary	1994	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJT00000000	SI757/13 (GCA001766505)	Hungary	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
MIJR00000000	SI786/13 (GCA001766515)	Hungary	2013	+	+	+	+	+	+	-	-	+	-	+	+	+
CP047882	pESI119944	Israel	2008	+	+	+	+	+	+	-	-	+	-	+	+	+
GCA010919335	SI119944	Israel	2008	+	+	+	+	+	+	-	-	+	-	+	+	+

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0148	2747	12	83401	84450	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	84607	87159	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	87200	87862	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	87984	88520	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	95987	96751	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	96779	97570	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0148	2747	12	98780	99271	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	99307	100092	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	100115	102547	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0148	2747	12	102572	103111	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	20554	21603	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	21760	24312	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	24353	25015	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	25137	25673	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	33140	33904	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	33932	34723	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	35933	36424	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	36460	37245	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	37268	39700	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0149	2748	19	39725	40264	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	83402	84451	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0150	2749	13	84608	87160	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	87201	87863	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	87985	88521	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	95988	96752	faeI_GVI52_23095	1-765/765	========	0/0	100.00	99.87	Fimbrial clusters	This study
19PM0150	2749	13	96780	97571	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	98781	99272	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	99308	100093	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	100116	102548	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0150	2749	13	102573	103112	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0151	2750	11	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0151	2750	11	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0153	2752	17	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	83402	84451	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0154	2753	13	84608	87160	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	87201	87863	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	87985	88521	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	95988	96752	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	96780	97571	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	98781	99272	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	99308	100093	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	100116	102548	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0154	2753	13	102573	103112	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0355	2954	22	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0355	2954	22	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0358	2957	16	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0360	2959	23	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
19PM0360	2959	23	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	14475	15524	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	15681	18233	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	18274	18936	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	19058	19594	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	27061	27825	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	27853	28644	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0045	3027	17	29854	30345	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	30381	31166	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	31189	33621	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0045	3027	17	33646	34185	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0261	3243	14	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0267	3249	12	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0267	3249	12	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	27605	28144	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	28169	30601	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	30624	31409	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	31445	31936	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	33146	33937	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	33965	34729	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0270	3252	10	42196	42732	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	42854	43516	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	43557	46109	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0270	3252	10	46266	47315	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	6240	7289	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	7446	9998	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	99.96	Fimbrial clusters	This study
20PM0271	3253	32	10039	10701	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	10823	11359	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	18826	19590	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	19618	20409	faeH	1-792/792	========	0/0	100.00	99.87	Fimbrial clusters	This study
20PM0271	3253	32	21619	22110	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	22146	22931	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	22954	25386	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0271	3253	32	25411	25950	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	656	1195	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0275	3257	28	1220	3652	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	3675	4460	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	4496	4987	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	6197	6988	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	7016	7780	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	15247	15783	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	15905	16567	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	16608	19160	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
20PM0275	3257	28	19317	20366	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	27602	28141	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	28166	30598	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	30621	31406	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	31442	31933	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	33143	33934	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	33962	34726	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERR1014109	ERS846143	19	42194	42730	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	42852	43514	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	43555	46107	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014109	ERS846143	19	46264	47313	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	27626	28165	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	28190	30622	faeD	1-2433/2433	========	0/0	100.00	99.96	Fimbrial clusters	This study
ERR1014113	ERS846147	12	30645	31430	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	31466	31957	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	33167	33958	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	33986	34750	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	42217	42753	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	42875	43537	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	43578	46130	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014113	ERS846147	12	46287	47336	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	27626	28165	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERR1014114	ERS846148	18	28190	30622	faeD	1-2433/2433	========	0/0	100.00	99.96	Fimbrial clusters	This study
ERR1014114	ERS846148	18	30645	31430	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	31466	31957	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	33167	33958	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	33986	34750	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	42217	42753	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	42875	43537	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	43578	46130	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014114	ERS846148	18	46287	47336	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	27626	28165	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	28190	30622	faeD	1-2433/2433	========	0/0	100.00	99.96	Fimbrial clusters	This study
ERR1014116	ERS846150	18	30645	31430	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	31466	31957	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	33167	33958	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	33986	34750	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERR1014116	ERS846150	18	42217	42753	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	42875	43537	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	43578	46130	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014116	ERS846150	18	46287	47336	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	27618	28157	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	28182	30614	faeD	1-2433/2433	========	0/0	100.00	99.96	Fimbrial clusters	This study
ERR1014117	ERS846151	18	30637	31422	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	31458	31949	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	33159	33950	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	33978	34742	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	42209	42745	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	42867	43529	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	43570	46122	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014117	ERS846151	18	46279	47328	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	20540	21589	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERR1014118	ERS846152	17	21746	24298	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	24339	25001	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	25123	25659	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	33127	33891	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	33919	34710	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	35920	36411	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	36447	37232	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	37255	39687	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
ERR1014118	ERS846152	17	39712	40251	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	24232	25281	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	25438	27990	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	28031	28693	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	28815	29351	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	36818	37582	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	37610	38401	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
JRXC01000000	SI54/04	1	39611	40102	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	40138	40923	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	40946	43378	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
JRXC01000000	SI54/04	1	43403	43942	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	20423	21472	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	21629	24181	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	24222	24884	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	25006	25542	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	33009	33773	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	33801	34592	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	35802	36293	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	36329	37114	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	37137	39569	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJR01000000	SI786/13	1	39594	40133	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	20423	21472	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
MIJS01000000	SI3337/12	1	21629	24181	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	24222	24884	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	25006	25542	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	33009	33773	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	33801	34592	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	35802	36293	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	36329	37114	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	37137	39569	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJS01000000	SI3337/12	1	39594	40133	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	20423	21472	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	21629	24181	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	24222	24884	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	25006	25542	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	33009	33773	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	33801	34592	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
MIJT01000000	SI757/13	1	35802	36293	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	36329	37114	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	37137	39569	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
MIJT01000000	SI757/13	1	39594	40133	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	20423	21472	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	21629	24181	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	24222	24884	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	25006	25542	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	33009	33773	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	33801	34592	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	35802	36293	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	36329	37114	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	37137	39569	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUW01000000	SI240/16	1	39594	40133	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	20405	21454	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
MRUX01000000	SI1070/16	1	21611	24163	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	24204	24866	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	24988	25524	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	32991	33755	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	33783	34574	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	35784	36275	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	36311	37096	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	37119	39551	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
MRUX01000000	SI1070/16	1	39576	40115	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	38569	39108	faeC_GVI52_23065	1-540/540	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	39133	41565	faeD	1-2433/2433	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	41588	42373	faeE_GVI52_23075	1-786/786	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	42409	42900	faeF_GVI52_23080	1-492/492	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	44110	44901	faeH	1-792/792	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	44929	45693	faeI_GVI52_23095	1-765/765	========	0/0	100.00	100.00	Fimbrial clusters	This study

Continuation **Table S5**

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
GCA010919335	SI119944	1	53160	53696	ipfA_GVI52_23160	1-537/537	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	53818	54480	ipfB_GVI52_23165	1-663/663	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	54521	57073	ipfC_GVI52_23170	1-2553/2553	========	0/0	100.00	100.00	Fimbrial clusters	This study
GCA010919335	SI119944	1	57230	58279	ipfD_GVI52_23175	1-1050/1050	========	0/0	100.00	100.00	Fimbrial clusters	This study

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0148	2747	12	91258	91563	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	12	91565	91783	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	12	126538	126870	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	12	126871	127128	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0148	2747	15	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0149	2748	14	90537	90971	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	91043	91393	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	91407	91682	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	91718	92140	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	92192	93886	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	93904	94266	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	14	94263	94499	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	19	28411	28716	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	19	28718	28936	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	19	63700	64032	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0149	2748	19	64033	64290	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	13	91259	91564	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	13	91566	91784	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	13	126539	126871	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	13	126872	127129	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0150	2749	16	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0150	2749	16	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	11	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	11	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	11	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	11	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0151	2750	14	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	90537	90971	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0153	2752	14	91043	91393	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	91407	91682	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	91718	92140	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	92192	93886	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	93904	94266	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	14	94263	94499	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	17	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	17	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	17	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0153	2752	17	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	13	91259	91564	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	13	91566	91784	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	13	126539	126871	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	13	126872	127129	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	90537	90971	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	91043	91393	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	91407	91682	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0154	2753	17	91718	92140	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	92192	93886	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	93904	94266	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0154	2753	17	94263	94499	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	16	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	22	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	22	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	22	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0355	2954	22	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	12	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	12	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0358	2957	12	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	12	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	12	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	12	13287	13637	merT	1-351/351	=========	0/0	100.00	99.72	fitness_SI	This study
19PM0358	2957	12	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	16	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	16	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	16	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0358	2957	16	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	90537	90971	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	91043	91393	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	91407	91682	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	91718	92140	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	92192	93886	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	93904	94266	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	16	94263	94499	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	23	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0360	2959	23	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	23	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
19PM0360	2959	23	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	90537	90971	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	91043	91393	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	91407	91682	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	91718	92140	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	92192	93886	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	93904	94266	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	11	94263	94499	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	17	22332	22637	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	17	22639	22857	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	17	57621	57953	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0045	3027	17	57954	58211	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	14	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	14	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	14	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0261	3243	14	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0261	3243	17	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	12	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	12	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	12	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	12	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0267	3249	18	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0267	3249	18	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	10	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	10	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	10	38933	39151	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	10	39153	39458	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0270	3252	23	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	65058	65492	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	65564	65914	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	65928	66203	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	66239	66661	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0271	3253	24	66713	68407	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	68425	68787	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	24	68784	69020	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	32	14097	14402	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	32	14404	14622	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	32	49377	49709	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0271	3253	32	49710	49967	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0273	3255	17	33319	33716	merR	1-398/435	=========	0/0	91.49	86.94	fitness_SI	This study
20PM0273	3255	17	33805	34153	merT	1-349/351	=========	0/0	99.43	87.39	fitness_SI	This study
20PM0273	3255	17	34169	34437	merP	8-276/276	=========	0/0	97.46	84.39	fitness_SI	This study
20PM0273	3255	17	34472	34876	merC	22-423/423	=========	#####	94.80	78.08	fitness_SI	This study
20PM0273	3255	17	34936	36621	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	Oct-31	98.82	81.54	fitness_SI	This study
20PM0273	3255	17	36639	36992	merD	13-363/363	=========	#####	96.69	81.07	fitness_SI	This study
20PM0273	3255	17	37002	37237	merE	1-236/237	=========	0/0	99.58	80.08	fitness_SI	This study
20PM0275	3257	15	9925	10161	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	15	10158	10520	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	15	10538	12232	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0275	3257	15	12284	12706	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	15	12742	13017	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	15	13031	13381	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	15	13453	13887	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	28	11984	12202	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	28	12204	12509	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	31	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
20PM0275	3257	31	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	14531	14788	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	14789	15121	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	49897	50115	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	50117	50422	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	169042	169278	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	169275	169637	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	169655	171349	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	171401	171823	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	171859	172134	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
CP047882	SI119944	1	172148	172498	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
CP047882	SI119944	1	172570	173004	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846143	ERR1014109	19	3564	3821	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846143	ERR1014109	19	3822	4154	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846143	ERR1014109	19	38930	39148	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846143	ERR1014109	19	39150	39455	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	12	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	12	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	12	38954	39172	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	12	39174	39479	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	90535	90969	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	91041	91391	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	91405	91680	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	91716	92138	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	92190	93884	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	93902	94264	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846147	ERR1014113	15	94261	94497	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERS846148	ERR1014114	13	7718	7954	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	7951	8313	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	8331	10025	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	10077	10499	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	10535	10810	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	10824	11174	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	13	11246	11680	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	18	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	18	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	18	38954	39172	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846148	ERR1014114	18	39174	39479	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	90535	90969	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	91041	91391	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	91405	91680	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	91716	92138	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	92190	93884	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	13	93902	94264	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERS846150	ERR1014116	13	94261	94497	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	18	3588	3845	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	18	3846	4178	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	18	38954	39172	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846150	ERR1014116	18	39174	39479	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014116	13	7710	7946	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	7943	8305	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	8323	10017	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	10069	10491	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	10527	10802	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	10816	11166	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	13	11238	11672	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	18	3580	3837	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	18	3838	4170	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	18	38946	39164	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846151	ERR1014117	18	39166	39471	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	10181	10417	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
ERS846152	ERR1014118	13	10414	10776	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	10794	12488	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	12540	12962	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	12998	13273	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	13287	13637	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	13	13709	14143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	17	28398	28703	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	17	28705	28923	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	17	63699	64031	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
ERS846152	ERR1014118	17	64032	64289	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	10075	10311	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	10308	10670	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	10688	12382	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	12434	12856	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	12892	13167	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	13181	13531	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	13603	14037	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
JRXC01000000	SI54/04	1	32089	32394	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	32396	32614	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	67390	67722	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
JRXC01000000	SI54/04	1	67723	67980	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	90708	91142	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	91214	91564	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	91578	91853	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	91889	92311	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	92363	94057	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	94075	94437	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	94434	94670	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	28280	28585	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	28587	28805	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	63581	63913	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
MIJS01000000	SI3337/12	1	63914	64171	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	90709	91143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	91215	91565	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
MIJT01000000	SI757/13	1	91579	91854	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	91890	92312	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	92364	94058	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	94076	94438	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	94435	94671	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	28280	28585	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	28587	28805	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	63581	63913	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
MIJT01000000	SI757/13	1	63914	64171	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001766515	SI786/13	1	28280	28585	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001766515	SI786/13	1	28587	28805	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001766515	SI786/13	1	63581	63913	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001766515	SI786/13	1	63914	64171	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906515	SI1070/16	1	28280	28585	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906515	SI1070/16	1	28587	28805	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906515	SI1070/16	1	63581	63913	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906515	SI1070/16	1	63914	64171	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
GCA001906535	SI240/16	1	90709	91143	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	91215	91565	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	91579	91854	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	91890	92312	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	92364	94058	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	94076	94438	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	94435	94671	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	28262	28567	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	28569	28787	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	63563	63895	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
GCA001906535	SI240/16	1	63896	64153	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	14531	14788	PemI-like	1-258/258	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	14789	15121	PemK/MazF	1-333/333	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	49897	50115	ccdA	1-219/219	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	50117	50422	ccdB	1-306/306	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	169042	169278	merE	1-237/237	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	169275	169637	merD	1-363/363	=========	0/0	100.00	100.00	fitness_SI	This study

Continuation **Table S5**

SEQUENCE	SAMPLE	CoG	START	END	GENE	COVERAGE	Cov MAP	GAPS	%Cov	%ID	DB	SOURCE
GCA010919335	SI119944	1	169655	171349	GVI52_23630_mercury(II)reductase	1-1695/1695	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	171401	171823	merC	1-423/423	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	171859	172134	merP	1-276/276	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	172148	172498	merT	1-351/351	=========	0/0	100.00	100.00	fitness_SI	This study
GCA010919335	SI119944	1	172570	173004	merR	1-435/435	=========	0/0	100.00	100.00	fitness_SI	This study

Table S6. Plasmid replicons, plasmid pMLST typing and genes encoding for plasmid origin of replication found among the S. Infantis strains used in this study (CoG: contig; Cov: coverage; ID: identity; DB: database).

SEQUENCE	SAMPLE	COUNTRY	YEAR	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	ACCESSION
19PM0348	2947	Germany	1992	15	16456	16597	IncI1-I(Gamma)_1__AP005147	1-142/142	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0350	2949	Germany	1992	29	14321	14618	IncFIC(FII)_1__AP001918	1-297/499	========/=.....	3/3	59.32	91.64	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0350	2949	Germany	1992	29	14321	14582	IncFII(pSFO)_1__AF401292	1-258/258	========/======	3/4	100	90.46	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0350	2949	Germany	1992	30	16456	16597	IncI1-I(Gamma)_1__AP005147	1-142/142	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
JRXC000000000	SI54/04	Hungary	2004	1	25826	26385	IncFIB(pN5539 1)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014111	ERS846145	Italy	2006	18	89114	89255	IncI1-I(Gamma)_1__AP005147	1-142/142	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014109	ERS846143	Italy	2007	27	9213	9772	IncFIB(pN5539 1)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
GCA010919335	SI119944	Israel	2008	1	300	859	IncFIB(pN5539 1)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014118	ERS846152	Italy	2009	25	14766	15325	IncFIB(pN5539 1)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014117	ERS846151	Italy	2012	26	9227	9786	IncFIB(pN5539 1)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MIJS000000000	SI3337/12	Hungary	2012	1	6216	6589	IncX1_1__EU370913	1-374/374	==========	0/0	100	98.66	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	COUNTRY	YEAR	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	ACCESSION
MIJS000000000	SI3337/12	Hungary	2012	1	7236	7578	IncX3_1__JN247852	37-374/374	.=======/======	5/4	90.37	80.17	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MIJS000000000	SI3337/12	Hungary	2012	1	14938	15497	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014113	ERS846147	Italy	2013	27	14787	15346	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014114	ERS846148	Italy	2013	25	14787	15346	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MIJT000000000	SI757/13	Hungary	2013	1	14938	15497	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MIJR000000000	SI786/13	Hungary	2013	1	14938	15497	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
ERR1014116	ERS846150	Italy	2014	25	9235	9794	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0267	3249	Germany	2015	27	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0270	3252	Germany	2015	50	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0271	3253	Germany	2016	54	2156	2715	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0273	3255	Germany	2016	17	30811	31110	IncFIC(FII)_1__AP001918	1-297/499	========/=.....	9/20	58.52	82.30	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	COUNTRY	YEAR	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	ACCESSION
20PM0273	3255	Germany	2016	17	30856	31103	IncFII(S)_1__CP000858	2-251/262	========/======	6/3	93.89	93.25	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0273	3255	Germany	2016	17	31363	31807	IncFII(SARC14)_1__JQ418540	1-445/445	===============	0/0	100	94.83	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0275	3257	Germany	2016	33	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MRUX00000000	SI1070/16	Hungary	2016	1	14938	15497	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MRUW00000000	SI240/16	Hungary	2016	1	7001	7374	IncX1_1__EU370913	1-374/374	=========	0/0	100	98.93	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MRUW00000000	SI240/16	Hungary	2016	1	8021	8363	IncX3_1__JN247852	37-374/374	.=======/======	5/4	90.37	80.17	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
MRUW00000000	SI240/16	Hungary	2016	1	14938	15497	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0355	2954	Germany	2017	31	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0358	2957	Germany	2018	27	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0148	2747	Germany	2019	26	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0149	2748	Germany	2019	27	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	COUNTRY	YEAR	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	ACCESSION
19PM0150	2749	Germany	2019	25	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0151	2750	Germany	2019	27	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0153	2752	Germany	2019	25	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0154	2753	Germany	2019	27	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
19PM0360	2959	Germany	2019	29	14766	15325	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa
20PM0045	3027	Germany	2020	27	9237	9796	IncFIB(pN55391)_1__CP016411	1-560/560	==========	0/0	100	100	plasmidFinder	https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/enterobacteriaceae.fsa

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0148	2747	15	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2747	15	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2747	15	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2747	15	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0149	2748	14	27168	27510	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0149	2748	14	39886	40350	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0149	2748	14	59317	59551	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0149	2748	14	78302	78555	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0150	2749	16	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0150	2749	16	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0150	2749	16	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0150	2749	16	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0151	2750	14	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0151	2750	14	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0151	2750	14	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0151	2750	14	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0153	2752	14	27168	27510	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0153	2752	14	39886	40350	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0153	2752	14	59317	59551	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0153	2752	14	78302	78555	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0154	2753	17	27168	27510	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0154	2753	17	39886	40350	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0154	2753	17	59317	59551	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0154	2753	17	78302	78555	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2947	15	10235	10488	pilL_1	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2947	15	16494	16576	repI1_1	1-83/83	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2947	15	64578	64920	ardA_4	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2947	15	77318	77782	trbA_13	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0148	2947	15	96212	96446	sogS_2	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0350	2949	20	10374	10608	sogS_2	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	20	29038	29502	trbA_13	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	20	41900	42242	ardA_4	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	29	14339	14496	FII_91	1-158/158	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	29	14454	14618	FIC_3	36-200/200	========	2/2	82	92,77	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	30	10235	10488	pilL_1	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0350	2949	30	16494	16576	repI1_1	1-83/83	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0355	2954	16	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0355	2954	16	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0355	2954	16	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0355	2954	16	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0358	2957	12	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0358	2957	12	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0358	2957	12	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0358	2957	12	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0360	2959	16	27168	27510	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0360	2959	16	39886	40350	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0360	2959	16	59317	59551	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
19PM0360	2959	16	78302	78555	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0045	3027	11	27168	27510	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0045	3027	11	39886	40350	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0045	3027	11	59317	59551	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0045	3027	11	78302	78555	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0261	3243	17	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0261	3243	17	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0261	3243	17	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0261	3243	17	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0267	3249	18	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0267	3249	18	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0267	3249	18	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0267	3249	18	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0270	3252	23	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0270	3252	23	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0270	3252	23	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0270	3252	23	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0271	3253	24	1689	2031	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0271	3253	24	14407	14871	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0271	3253	24	33838	34072	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0271	3253	24	52823	53076	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0273	3255	17	30811	30974	FIC_3	36-197/200	========	4/4	80.50	84.24	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0273	3255	17	30902	31111	FIIS_5	1-205/208	========	2/7	98.08	95.73	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0275	3257	15	25869	26122	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0275	3257	15	44873	45107	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0275	3257	15	64074	64538	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
20PM0275	3257	15	76914	77256	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047882	184986	185239	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047882	203990	204224	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047882	223191	223655	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047882	236031	236373	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846143	ERR1014109	18	27144	27486	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846143	ERR1014109	18	39862	40326	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846143	ERR1014109	18	59293	59527	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
ERS846145	ERR1014111	18	9862	10096	sogS_2	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846145	ERR1014111	18	28256	28990	trbA_13	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846145	ERR1014111	18	43974	44316	ardA_4	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846145	ERR1014111	18	89135	89217	repI1_1	1-83/83	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846145	ERR1014111	18	96559	96812	pilL_1	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846147	ERR1014113	15	27166	27508	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846147	ERR1014113	15	39884	40348	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846147	ERR1014113	15	59315	59549	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846147	ERR1014113	15	78300	78553	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846148	ERR1014114	13	23662	23915	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846148	ERR1014114	13	42666	42900	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846148	ERR1014114	13	61867	62331	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846148	ERR1014114	13	74707	75049	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
ERS846150	ERR1014116	13	27166	27508	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846150	ERR1014116	13	39884	40348	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846150	ERR1014116	13	59315	59549	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846150	ERR1014116	13	78300	78553	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846151	ERR1014117	13	23654	23907	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846151	ERR1014117	13	42658	42892	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846151	ERR1014117	13	61859	62323	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846151	ERR1014117	13	74699	75041	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846152	ERR1014118	13	26125	26378	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846152	ERR1014118	13	45129	45363	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846152	ERR1014118	13	64330	64794	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
ERS846152	ERR1014118	13	77170	77512	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJS00000000	SI3337/12	1	27340	27682	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
MIJS00000000	SI3337/12	1	40058	40522	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJS00000000	SI3337/12	1	59489	59723	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJS00000000	SI3337/12	1	78474	78727	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJR00000000	SI786/13	1	27340	27682	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJR00000000	SI786/13	1	40058	40522	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJR00000000	SI786/13	1	59489	59723	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MIJR00000000	SI786/13	1	78474	78727	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUX000000000	SI1070/16	1	4816	5069	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUX000000000	SI1070/16	1	23820	24054	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUX000000000	SI1070/16	1	43980	44444	trbA_5	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUX000000000	SI1070/16	1	56831	57173	ardA_4	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUW00000000	SI240/16	1	27340	27682	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
MRUW00000000	SI240/16	1	40058	40522	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUW00000000	SI240/16	1	59489	59723	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
MRUW00000000	SI240/16	1	78474	78727	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047881	184986	185239	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047881	203990	204224	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047881	223191	223655	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
GCA010919335	SI119944	CP047881	236031	236373	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
JRXC00000000	SI54/04	1	9661	10003	ardA_2	1-343/343	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
JRXC00000000	SI54/04	1	22379	22843	trbA_21	1-465/465	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
JRXC00000000	SI54/04	1	41810	42044	sogS_9	1-235/235	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa
JRXC00000000	SI54/04	1	60795	61048	pilL_3	1-254/254	========	0/0	100	100	incI1 plasmids	https://bitbucket.org/genomicepidemiology/pmlst_db/src/master/inci1.fsa

SEQUENCE	ISOLATE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0148	2747	15	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0148	2747	26	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0149	2748	14	89558	90369	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0149	2748	27	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0150	2749	16	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0150	2749	25	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0151	2750	14	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
19PM0151	2750	27	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0153	2752	14	89558	90369	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0153	2752	25	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0154	2753	17	89558	90369	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0154	2753	27	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0355	2954	16	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0355	2954	31	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0358	2957	12	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0358	2957	27	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
19PM0360	2959	16	89558	90369	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
19PM0360	2959	29	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0045	3027	11	89558	90369	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
20PM0045	3027	27	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0261	3243	17	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
20PM0261	3243	30	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0267	3249	18	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
20PM0267	3249	27	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0270	3252	23	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
20PM0270	3252	50	9085	10095	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0271	3253	24	64079	64890	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
20PM0271	3253	54	2004	3014	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
20PM0275	3257	15	14055	14866	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
20PM0275	3257	33	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846143	ERR1014109	27	9061	10071	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846147	ERR1014113	15	89556	90367	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
ERS846147	ERR1014113	27	14488	15498	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846148	ERR1014114	13	11848	12659	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
ERS846148	ERR1014114	25	14488	15498	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846150	ERR1014116	13	89556	90367	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
ERS846150	ERR1014116	25	9083	10093	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846151	ERR1014117	13	11840	12651	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
ERS846151	ERR1014117	26	9075	10085	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
ERS846152	ERR1014118	13	14311	15122	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
ERS846152	ERR1014118	25	14467	15477	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
MIJS01000000	SI3337/12	1	89729	90540	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
MIJS01000000	SI3337/12	1	14639	15649	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
MIJT01000000	SI757/13	1	89730	90541	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
MIJT01000000	SI757/13	1	14639	15649	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
MIJR01000000	SI786/13	1	14639	15649	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
MRUX01000000	SI1070/16	1	14639	15649	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
MRUW01000000	SI1070/16	1	89730	90541	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study

Continuation **Table S6**

SEQUENCE	SAMPLE	CoG	START	END	GENE	Cov	Cov_MAP	GAPS	%Cov	%ID	DB	SOURCE
MRUW01000000	SI1070/16	1	14639	15649	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
GCA010919335	SI119944	CP047882	173172	173983	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
GCA010919335	SI119944	CP047882	1	1011	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study
JRXC01000000	SI54/04	1	72051	72862	incP_Tn1723	1-811/811	========/======	1/1	100	99,63	incP_plasmids	This study
JRXC01000000	SI54/04	1	25527	26537	repB_GVI52_22855	1-1011/1011	==========	0/0	100	100	repB_pESI119944	This study

8.2. Supplementary Table from the Dissertation

Table S1. Results of the analysis of Salmonella genoserotyping study performed by the bioinformatics pipeline WGSBAC and the tools SISTR, SeqSero, and SeqSero2.

	Traditional analysis (KWS)		Analysis by SISTR (WGS)			
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar antigen**	**Predicted serovar cgMLST**	**Predicted serovar consensus**	**Predicted formula**
16PM0056	Bovismorbificans	6,8,20:r:1,5	Bovismorbificans | Hindmarsh	Bovismorbificans	Bovismorbificans	6,8,20:r:1,5
16PM0089	Hadar	6_1,8,z_{10},e,n,x	Hadar | Istanbul	Hadar	Hadar	6,8:z_{10}:e,n,x
16PM0105	Gallinarum	1,9,12	Gallinarum | Pullorum	Gallinarum	Gallinarum	1,9,12:-:-
16PM0121	Panama	1,9,12,l,v,1,5	Goettingen	Goettingen	Goettingen	9,12:l,v:e,n,z_{15}
16PM0122	Abony	1,4,(5),12,27,b,e,n,x	I 4,[5],12:b:- | Schleissheim	Mississippi	I 4,[5],12:b:- | Schleissheim	4,[5],12:b:-
16PM0124	III_diarizonae	61:k:1,5,7	IIIb 61:k:1,5,7	O61:k:1,5,7	O61:k:1,5,7	61:k:1,5,7
16PM0161	Tennessee	6,7,14:z29:(1,2,7)	Mbandaka	Mbandaka	Mbandaka	6,7,14:z_{10}:e,n,z,15
16PM0171	Livingstone	6,7,14:d:l,w	Livingstone	Livingstone	Livingstone	6,7,14:d:l,w
16PM0176	Stourbridge	6,8:b:1,6	Stourbridge	Stourbridge	Stourbridge	6,8:b:1,6
16PM0256	Goldcoast	6,8:r:l,w	Brikama | Goldcoast	Goldcoast	Goldcoast	6,8:r:l,w
16PM0296	Anatum	3,(10,15,34):e,h:1,6	Anatum | Hayindogo	Anatum	Anatum	3,{10}{15}{15,34}:e,h:1,6
17PM0009	Enteritidis	1,9,12,g,m	I 9,46:g,m:1,2	Coeln	I 9,46:g,m:1,2	9,12:g,m:1,2
17PM0024	Mbandaka	6,7,14:z10:e,n,z15	Tennessee	Tennessee	Tennessee	6,7,14:z_{29}:-
17PM0051	Kentucky	8,20:i:z6	Kentucky	Kentucky	Kentucky	8,20:i:z_6
17PM0053	Agona	4,(5),f,g,s,(1,2)	Agona | Budapest | Derby	Agona	Agona	1,4,[5];12:f,g,s:-,

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SISTR (WGS)			
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar antigen**	**Predicted serovar cgMLST**	**Predicted serovar consensus**	**Predicted formula**
17PM0054	Muenster	3,(10,15,34),e,h,1,5	Muenster|Vilvoorde	Muenster	Muenster	3,{10}{15}{15,34}:e,h:1,5
17PM0072	Rauform	-	Blegdam|Dublin|Enteritidis|Gueulletapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,b:-
17PM0167	Meleagridis	3,(10,15,34):e,h:l,w	Calabar|Meleagridis	Meleagridis	Meleagridis	3,{10}{15}{15,34}:e,h:l,w
17PM0271	Indiana	1,4,12:z:1,7	Indiana	Indiana	Indiana	1,4,12:z:1,7
17PM0282	Typhimurium	4,5:i:1,2	Typhimurium	Typhimurium	Typhimurium	1,4,[5],12:i:1,2
17PM0296	Typhimurium mono	4,5:i:1-,2-,	I 4,[5],12:i:-	I 4,[5],12:i:-	I 4,[5],12:i:-	4,[5],12:i:-
17PM0299	Typhimurium mono	4,5-:i:1-,2-,	I 4,[5],12:i:-	I 4,[5],12:i:-	I 4,[5],12:i:-	4,[5],12:i:-
18PM0004	Dublin	1,9,12:g,p	Blegdam|Dublin|Enteritidis|Gueulletapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0007	Coeln	1,4,(5),y,1,2	Coeln	Coeln	Coeln	1,4,[5],12:y:1,2
18PM0020	Infantis	6_1,7,r,1,5	Infantis|Senegal	Infantis	Infantis	6,7,14:r:1,5
18PM0021	Typhimurium mono	4, 5, i, 1-, 2-,	I 4,[5],12:i:-	I 4,[5],12:i:-	I 4,[5],12:i:-	| 4, [5], 12:i:-
18PM0026	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulletapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0031	Choleraesuis	6,7:c:1,5	Chiredzi|Choleraesuis|Paratyphi C|Typhisuis	Choleraesuis	Choleraesuis	6,7:c:1,5

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SISTR (WGS)			
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar antigen**	**Predicted serovar cgMLST**	**Predicted serovar consensus**	**Predicted formula**
18PM0031 B	Choleraesuis	6,7:c:1,5	Chiredzi|Choleraesuis|Paratyphi C|Typhisuis	Choleraesuis	Choleraesuis	6,7:c:1,5
18PM0045	Derby	4,(5),f,g,(1,2)	Agona|Budapest|Derby	Derby	Derby	1,4,[5],12:f,g:-
18PM0045 B	Derby	4,(5),f,g,(1,2)	Agona|Budapest|Derby	Derby	Derby	1,4,[5],12:f,g:-
18PM0089	Rauform	-	Blegdam|Dublin|Enteritidis|Gueulettapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0092	Paratyphi B	4,5-,b,1,2	Paratyphi B|Paratyphi B var. Java	Paratyphi B var. Java	Paratyphi B var. Java	1,4,[5],12:b:1,2
18PM0109	Kottbus	6_1, 8, e, h, 1, 5	Ferruch|Kottbus	Kottbus	Kottbus	6,8:e,h:1,5
18PM0111	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulettapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0112	Typhimurium	4,5-,i,1,2	Typhimurium	Typhimurium	Typhimurium	1,4,[5],12:i:1,2
18PM0128	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulettapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0135	Kottbus	6_1,8,e,h,1,5	Ferruch|Kottbus	Kottbus	Kottbus	6,8:e,h:1,5
18PM0158	Anatum		Anatum|Hayindogo	Anatum	Anatum	3,{10}{15}{15,34}:e,h:1,6
18PM0174	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulettapee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SISTR (WGS)			
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar antigen**	**Predicted serovar cgMLST**	**Predicted serovar consensus**	**Predicted formula**
18PM0195	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulette apee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0246	Stourbridge	6, 8, b, 1, 6	Stourbridge	Stourbridge	Stourbridge	6,8:b:1,6
18PM0262	Dublin	1, 9, 12, g, p	Blegdam|Dublin|Enteritidis|Gueulette apee|Hillingdon|Kiel|Moscow|Naestved|Nitra|Rostock	Dublin	Dublin	1,9,12[Vi]:g,p:-
18PM0291	Panama	1, 9, 12, l, v, 1, 5	Mathura	I 4,[5],12:i:-	Mathura	9,46: i: e,n, z_{15}
19PM0011	Newport	-	Bardo|Newport	Newport	Newport	6,8,20:e,h:1,2
	Traditional analysis (KWS)		**Analysis by SeqSero using reads**		**Analysis by SeqSero using contigs**	
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar reads**	**Predicted formula**	**Predicted serovar contigs**	**Predicted antigen profile**
16PM0056	Bovismorbificans	6,8,20:r:1,5	Hindmarsh or Bovismorbificans	8:r:1,5	N/A The predicted antigenic profile does not exist in the KWS	8:r:-
16PM0089	Hadar	6_1,8,z_{10},e,n,x	Hadar or Istanbul*	8:z10:e,n,x	Hadar or Isanbul*	8:z10:e,n,x
16PM0105	Gallinarum	1,9,12	*	9:g,m:-	*	9:g,m:-
16PM0121	Panama	1,9,12,l,v,1,5	Goettingen	9:l,v:e,n,z15	Goettingen	9:l,v:e,n,z15
16PM0122	Abony	1,4,(5),12,27,b,e,n,x	Potential monophasic variant of Paratyphi B*	4:b:-	Potential monophasic variant of Paratyphi B*	4:b:-
16PM0124	III_diarizonae	61:k:1,5,7	N/A The predicted antigen profile does not exist in the KWS	61:k:1,5	IIIb 61:k:1,5,7	61:k:1,5,(7)

Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero using reads		Analysis by SeqSero using contigs	
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar reads**	**Predicted formula**	**Predicted serovar contigs**	**Predicted antigen profile**
16PM0161	Tennessee	6,7,14:z29:(1,2,7)	Mbandaka	7:z10:e,n,z15	Mbandaka	7:z10:e,n,z15
16PM0171	Livingstone	6,7,14:d:l,w	Livingstone	7:d:l,w	Livingstone	7:d:l,w
16PM0176	Stourbridge	6,8:b:1,6	Stourbridge	8:b:1,6	Stourbridge	8:b:1,6
16PM0256	Goldcoast	6,8:r:l,w	Goldcoast or Brikama	8:r:l,w	IIIb 8:r:z	8:r:z
16PM0296	Anatum	3,(10,15,34):e,h:1,6	Anatum	3,10:e,h:1,6	Anatum	3,10:e,h:1,6
17PM0009	Enteritidis	1,9,12,g,m	N/A The predicted antigen profile does not exist in the KWS	9:y:1,2	N/A The predicted antigen profile does not exist in the KWS	9:y:1,2
17PM0024	Mbandaka	6,7,14:z10:e,n,z15	II 6,7:z29:[z42] or Tenesse*	7:z29:-	II 6,7:z29:[z42] or Tenesse*	7:z29:-
17PM0051	Kentucky	8,20:i:z6	Kentucky	8:i:z6	Kentucky	8:i:z6
17PM0053	Agona	4,(5),f,g,s,(1,2)	Agona	4:f,g,s:-	Agona	4:f,g,s:-
17PM0054	Muenster	3,(10,15,34),e,h,1,5	Muenster	3,10:e,h:1,5	Muenster	3,10:e,h:1,5
17PM0072	Rauform	-	Dublin	9:g,p:-	Dublin	9:g,p:-
17PM0167	Meleagridis	3,(10,15,34):e,h:l,w	Meleagridis	3,10:e,h:l,w	Meleagridis	3,10:e,h:l,w
17PM0271	Indiana	1,4,12:z:1,7	Indiana or II 4,12:z:1,7*	4:z:1,7	Indiana or II 4,12:z:1,7*	4:z:1,7
17PM0282	Typhimurium	4,5:i:1,2	Typhimurium	4:i:1,2	Typhimurium	4:i:1,2
17PM0296	Typhimurium mono	4,5:i:1-,2-,	Potential monophasic variant of Typhimurium	4:i:-	Potential monophasic variant of Typhimurium	4:i:-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero using reads		Analysis by SeqSero using contigs	
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar reads**	**Predicted formula**	**Predicted serovar contigs**	**Predicted antigen profile**
17PM0299	Typhimurium mono	4,5-:i:1-,2-,	Potential monophasic variant of Typhimurium(O5-)*	4:i:-	Potential monophasic variant of Typhimurium(O5-)*	4:i:-
18PM0004	Dublin	1,9,12:g,p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0007	Coeln	1,4,(5),y,1,2	Coeln	4:y:1,2	Coeln	4:y:1,2
18PM0020	Infantis	6_1,7,r,1,5	Infantis	7:r:1,5	N/A The predicted antigen profile does not exist in the KWS	7:r:-
18PM0021	Typhimurium mono	4, 5, i, 1-, 2-,	Potential monophasic variant of Typhimurium	4:i:-	Potential monophasic variant of Typhimurium	4:i:-
18PM0026	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0031	Choleraesuis	6,7:c:1,5	Paratyphi C or Cholerasuis or Typhisuis*	7:c:1,5	Hissar	7:c:1,2
18PM0031B	Choleraesuis	6,7:c:1,5	Paratyphi C or Cholerasuis or Typhisuis*	7:c:1,5	Hissar	7:c:1,2
18PM0045	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	Derby	4:f,g:-
18PM0045B	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	Derby	4:f,g:-
18PM0089	Rauform	-	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0092	Paratyphi B	4,5-,b,1,2	Paratyphi B	4:b:1,2	Paratyphi B	4:b:1,2
18PM0109	Kottbus	6_1, 8, e, h, 1, 5	Kottbus or Ferruch	8:e,h:1,5	N/A The predicted antigen profile does not exist in the KWS	8:e,h:-
18PM0111	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero using reads		Analysis by SeqSero using contigs	
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar reads**	**Predicted formula**	**Predicted serovar contigs**	**Predicted antigen profile**
18PM0112	Typhimurium	4,5-,i,1,2	Typhimurium(O5-)*	4:i:1,2	Typhimurium(O5-)*	4:i:1,2
18PM0128	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0135	Kottbus	6_1,8,e,h,1,5	Kottbus or Ferruch*	8:e,h:1,5	Kottbus or Ferruch*	8:e,h:1,5
18PM0158	Anatum	-	Anatum	3,10:e,h:1,6	Anatum	3,10:e,h:1,6
18PM0174	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0195	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0246	Stourbridge	6, 8, b, 1, 6	Stourbridge	8:b:1,6	Stourbridge	8:b:1,6
18PM0262	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	Dublin	9:g,p:-
18PM0291	Panama	1, 9, 12, l, v, 1, 5	Goettingen	9:l,v:e,n,z15	N/A The predicted antigen profile does not exist in the KWS	9:-:e,n,z15
19PM0011	Newport	-	Newport	8:e,h:1,2	N/A The predicted antigen profile does not exist in the KWS	8:e,h:-

	Traditional analysis (KWS)		Analysis by SeqSero2 using reads and allele microassembly workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
16PM0056	Bovismorbificans	6,8,20:r:1,5	Bovismorbificans	8:r:1,5	-
16PM0089	Hadar	6_1,8,z_{10},e,n,x	Hadar	8:z10:e,n,x	-
16PM0105	Gallinarum	1,9,12	Gallinarum or Enteritidis	9:g,m:-	sdf gene not detected. The predicted serotypes share the same general formula: 9:g,m:-.

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero2 using reads and allele microassembly workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
16PM0121	Panama	1,9,12,l,v,1,5	Goettingen	9:l,v:e,n,z15	-
16PM0122	Abony	1,4,(5),12,27,b,e,n,x	I 4:b:-	4:b:-	This predicted serotype is not in the Kauffman-White scheme.
16PM0124	III_diarizonae	61:k:1,5,7	IIIb 61:k:1,5,(7)	61:k:1,5,7	-
16PM0161	Tennessee	6,7,14:z29:(1,2,7)	Mbandaka	7:z10:e,n,z15	-
16PM0171	Livingstone	6,7,14:d:l,w	Livingstone	7:d:l,w	-
16PM0176	Stourbridge	6,8:b:1,6	Stourbridge	8:b:1,6	-
16PM0256	Goldcoast	6,8:r:l,w	Goldcoast or Brikama	8:r:l,w	The predicted serotypes share the same general formula: 8:r:l,w.
16PM0296	Anatum	3,(10,15,34):e,h:1,6	Anatum	3,10:e,h:1,6	-
17PM0009	Enteritidis	1,9,12,g,m	I 4:g,m:1,2	4:g,m:1,2	This predicted serotype is not in the Kauffman-White scheme. Co-existence of multiple serotypes detected, indicating potential inter-serotype contamination. See 'Extracted_antigen_alleles.fasta' for detected serotype determinant alleles.
17PM0024	Mbandaka	6,7,14:z10:e,n,z15	Tennessee	7:z29:-	-
17PM0051	Kentucky	8,20:i:z6	Kentucky	8:i:z6	-
17PM0053	Agona	4,(5),f,g,s,(1,2)	Agona	4:f,g,s:-	-
17PM0054	Muenster	3,(10,15,34),e,h,1,5	Muenster	3,10:e,h:1,5	Co-existence of multiple serotypes detected, indicating potential inter-serotype contamination. See 'Extracted_antigen_alleles.fasta' for detected serotype determinant alleles.

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero2 using reads and allele microassembly workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
17PM0072	Rauform	-	Dublin	9:g,p:-	-
17PM0167	Meleagridis	3,(10,15,34):e,h:l,w	Meleagridis	3,10:e,h:l,w	-
17PM0271	Indiana	1,4,12:z:1,7	Indiana	4:z:1,7	-
17PM0282	Typhimurium	4,5:i:1,2	Typhimurium	4:i:1,2	-
17PM0296	Typhimurium mono	4,5:i:1-,2-,	I 4:-:-	4:-:-	H antigens were not detected. This is an atypical result that should be further investigated. Most Salmonella strains have at least fliC, encoding the Phase 1 H antigen, even if it is not expressed.
17PM0299	Typhimurium mono	4,5-:i:1-,2-,	I 4,[5],12:i:-	4:i:-	Detected a deletion that causes O5- variant of Typhimurium.
18PM0004	Dublin	1,9,12:g,p	Dublin	9:g,p:-	-
18PM0007	Coeln	1,4,(5),y,1,2	Coeln	4:y:1,2	-
18PM0020	Infantis	6_1,7,r,1,5	Infantis	7:r:1,5	-
18PM0021	Typhimurium mono	4, 5, i, 1-, 2-,	I 4,[5],12:i:-	4:i:-	-
18PM0026	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0031	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	The predicted serotypes share the same general formula: 7:c:1,5.
18PM0031 B	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	-
18PM0045	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-
18PM0045 B	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero2 using reads and allele microassembly workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
18PM0089	Rauform	-	Dublin	9:g,p:-	-
18PM0092	Paratyphi B	4,5-,b,1,2	- -:-:-	-:-:-	The input genome cannot be identified as Salmonella. Check the input for taxonomic ID, contamination, or sequencing quality.
18PM0109	Kottbus	6_1, 8, e, h, 1, 5	Kottbus	8:e,h:1,5	-
18PM0111	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0112	Typhimurium	4,5-,i,1,2	Typhimurium	4:i:1,2	Detected a deletion that causes O5- variant of Typhimurium.
18PM0128	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0135	Kottbus	6_1,8,e,h,1,5	Kottbus	8:e,h:1,5	-
18PM0158	Anatum		Anatum	3,10:e,h:1,6	-
18PM0174	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0195	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0246	Stourbridge	6, 8, b, 1, 6	Stourbridge	8:b:1,6	-
18PM0262	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0291	Panama	1, 9, 12, l, v, 1, 5	Brandenburg	4:l,v:e,n,z15	Co-existence of multiple serotypes detected, indicating potential inter-serotype contamination. See 'Extracted_antigen_alleles.fasta' for detected serotype determinant alleles.
19PM0011	Newport		Newport	8:e,h:1,2	-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero2 using reads and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
16PM0056	Bovismorbificans	6,8,20:r:1,5	Bovismorbificans	8:r:1,5	-
16PM0089	Hadar	6_1,8,z_{10},e,n,x	Hadar	8:z10:e,n,x	-
16PM0105	Gallinarum	1,9,12	Gallinarum or Enteritidis	9:g,m:-	-
16PM0121	Panama	1,9,12,l,v,1,5	Goettingen	9:l,v:e,n,z15	-
16PM0122	Abony	1,4,(5),12,27,b,e,n,x	I 4:b:-	4:b:-	This predicted serotype is not in the Kauffman-White scheme.
16PM0124	III_diarizonae	61:k:1,5,7	IIIb 61:k:1,5,(7)	61:k:1,5,7	-
16PM0161	Tennessee	6,7,14:z29:(1,2,7)	Mbandaka	7:z10:e,n,z15	-
16PM0171	Livingstone	6,7,14:d:l,w	Livingstone	7:d:l,w	-
16PM0176	Stourbridge	6,8:b:1,6	Stourbridge	8:b:1,6	-
16PM0256	Goldcoast	6,8:r:l,w	Goldcoast or Brikama	8:r:l,w	The predicted serotypes share the same general formula: 8:r:l,w.
16PM0296	Anatum	3,(10,15,34):e,h:1,6	Anatum	3,10:e,h:1,6	-
17PM0009	Enteritidis	1,9,12,g,m	I 9:g,m:1,2	9:g,m:1,2	This predicted serotype is not in the Kauffman-White scheme.
17PM0024	Mbandaka	6,7,14:z10:e,n,z15	Tennessee	7:z29:-	-
17PM0051	Kentucky	8,20:i:z6	Kentucky	8:i:z6	-
17PM0053	Agona	4,(5),f,g,s,(1,2)	Agona	4:f,g,s:-	-
17PM0054	Muenster	3,(10,15,34),e,h,1,5	Muenster	3,10:e,h:1,5	-

Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero2 using reads and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
17PM0072	Rauform	-	Dublin	9:g,p:-	-
17PM0167	Meleagridis	3,(10,15,34):e,h:l,w	Meleagridis	3,10:e,h:l,w	-
17PM0271	Indiana	1,4,12:z:1,7	Indiana	4:z:1,7	
17PM0282	Typhimurium	4,5:i:1,2	Typhimurium	4:i:1,2	-
17PM0296	Typhimurium mono	4,5:i:1-,2-,	I 4:-:-	4:-:-	H antigens were not detected. This is an atypical result that should be further investigated. Most Salmonella strains have at least fliC, encoding the Phase 1 H antigen, even if it is not expressed.
17PM0299	Typhimurium mono	4,5-:i:1-,2-,	I 4,[5],12:i:-	4:i:-	Detected a deletion that causes O5- variant of Typhimurium.
18PM0004	Dublin	1,9,12:g,p	Dublin	9:g,p:-	-
18PM0007	Coeln	1,4,(5),y,1,2	Coeln	4:y:1,2	-
18PM0020	Infantis	6_1,7,r,1,5	Infantis	7:r:1,5	-
18PM0021	Typhimurium mono	4, 5, i, 1-, 2-,	I 4,[5],12:i:-	4:i:-	-
18PM0026	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0031	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	The predicted serotypes share the same general formula: 7:c:1,5.
18PM0031B	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	The predicted serotypes share the same general formula: 7:c:1,5.

Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero2 using reads and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
18PM0045	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-
18PM0045B	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-
18PM0089	Rauform	-	Dublin	9:g,p:-	-
18PM0092	Paratyphi B	4,5-,b,1,2	Paratyphi B	4:b:1,2	-
18PM0109	Kottbus	6_1, 8, e, h, 1, 5	Kottbus	8:e,h:1,5	-
18PM0111	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0112	Typhimurium	4,5-,i,1,2	Typhimurium	4:i:1,2	-
18PM0128	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0135	Kottbus	6_1,8,e,h,1,5	Kottbus	8:e,h:1,5	-
18PM0158	Anatum		Anatum	3,10:e,h:1,6	-
18PM0174	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0195	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0246	Stourbridge	6, 8, b, 1, 6	Stourbridge	8:b:1,6	-
18PM0262	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0291	Panama	1, 9, 12, l, v, 1, 5	I 9:i:e,n,z15	9:i:e,n,z15	This predicted serotype is not in the Kauffman-White scheme.
19PM0011	Newport		Newport	8:e,h:1,2	-

Continuation **Table S1**

	Traditional analysis (KWS)		Analysis by SeqSero2 using assemblies and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
16PM0056	Bovismorbificans	6,8,20:r:1,5	Bovismorbificans	8:r:1,5	-
16PM0089	Hadar	6_1,8,z_{10},e,n,x	Hadar	8:z10:e,n,x	-
16PM0105	Gallinarum	1,9,12	Gallinarum or Enteritidis	9:g,m:-	sdf gene not detected. The predicted serotypes share the same general formula: 9:g,m:-.
16PM0121	Panama	1,9,12,l,v,1,5	Goettingen	9:l,v:e,n,z15	-
16PM0122	Abony	1,4,(5),12,27,b,e,n,x	- 4:-:-	4:-:-	The input genome cannot be identified as Salmonella. Check the input for taxonomic ID, contamination, or sequencing quality.
16PM0124	III_diarizonae	61:k:1,5,7	IIIb 61:k:1,5,(7)	61:k:1,5,7	-
16PM0161	Tennessee	6,7,14:z29:(1,2,7)	Mbandaka	7:z10:e,n,z15	-
16PM0171	Livingstone	6,7,14:d:l,w	Livingstone	7:d:l,w	-
16PM0176	Stourbridge	6,8:b:1,6	Stourbridge	8:b:1,6	-
16PM0256	Goldcoast	6,8:r:l,w	Goldcoast or Brikama	8:r:l,w	The predicted serotypes share the same general formula: 8:r:l,w.
16PM0296	Anatum	3,(10,15,34):e,h:1,6	Anatum	3,10:e,h:1,6	-
17PM0009	Enteritidis	1,9,12,g,m	I 4:g,m:1,2	4:g,m:1,2	This predicted serotype is not in the Kauffman-White scheme.
17PM0024	Mbandaka	6,7,14:z10:e,n,z15	Tennessee	7:z29:-	-

Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero2 using assemblies and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
17PM0051	Kentucky	8,20:i:z6	Kentucky	8:i:z6	-
17PM0053	Agona	4,(5),f,g,s,(1,2)	Agona	4:f,g,s:-	-
17PM0054	Muenster	3,(10,15,34),e,h,1,5	Muenster	3,10:e,h:1,5	-
17PM0072	Rauform	-	Dublin	9:g,p:-	-
17PM0167	Meleagridis	3,(10,15,34):e,h:l,w	Meleagridis	3,10:e,h:l,w	-
17PM0271	Indiana	1,4,12:z:1,7	Indiana	4:z:1,7	-
17PM0282	Typhimurium	4,5:i:1,2	I -:i:1,2	-:i:1,2	O antigen was not detected. This result may be due to a rough strain that has deleted the rfb region. For raw reads input, the k-mer workflow is sometimes more sensitive than the microassembly workflow in detecting O antigen. Caution should be used with this approach because the k-mer result may be due to low levels of contamination.
17PM0296	Typhimurium mono	4,5:i:1-,2-,	- -:-:-	-:-:-	The input genome cannot be identified as Salmonella. Check the input for taxonomic ID, contamination, or sequencing quality.
17PM0299	Typhimurium mono	4,5-:i:1-,2-,	I 4,[5],12:i:-	4:i:-	Detected a deletion that causes O5- variant of Typhimurium.

Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero2 using assemblies and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
18PM0004	Dublin	1,9,12:g,p	Dublin	9:g,p:-	-
18PM0007	Coeln	1,4,(5),y,1,2	Coeln	4:y:1,2	-
18PM0020	Infantis	6_1,7,r,1,5	Infantis	7:r:1,5	-
18PM0021	Typhimurium mono	4, 5, i, 1-, 2-,	I 4,[5],12:i:-	4:i:-	-
18PM0026	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0031	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	The predicted serotypes share the same general formula: 7:c:1,5.
18PM0031B	Choleraesuis	6,7:c:1,5	Paratyphi C or Choleraesuis or Typhisuis	7:c:1,5	The predicted serotypes share the same general formula: 7:c:1,5.
18PM0045	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-
18PM0045B	Derby	4,(5),f,g,(1,2)	Derby	4:f,g:-	-
18PM0089	Rauform	-	Dublin	9:g,p:-	-
18PM0092	Paratyphi B	4,5-,b,1,2	- -:-:-	-:-:-	The input genome cannot be identified as Salmonella. Check the input for taxonomic ID, contamination, or sequencing quality.
18PM0109	Kottbus	6_1, 8, e, h, 1, 5	Kottbus	8:e,h:1,5	-
18PM0111	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0112	Typhimurium	4,5-,i,1,2	Typhimurium	4:i:1,2	Detected a deletion that causes O5- variant of Typhimurium.

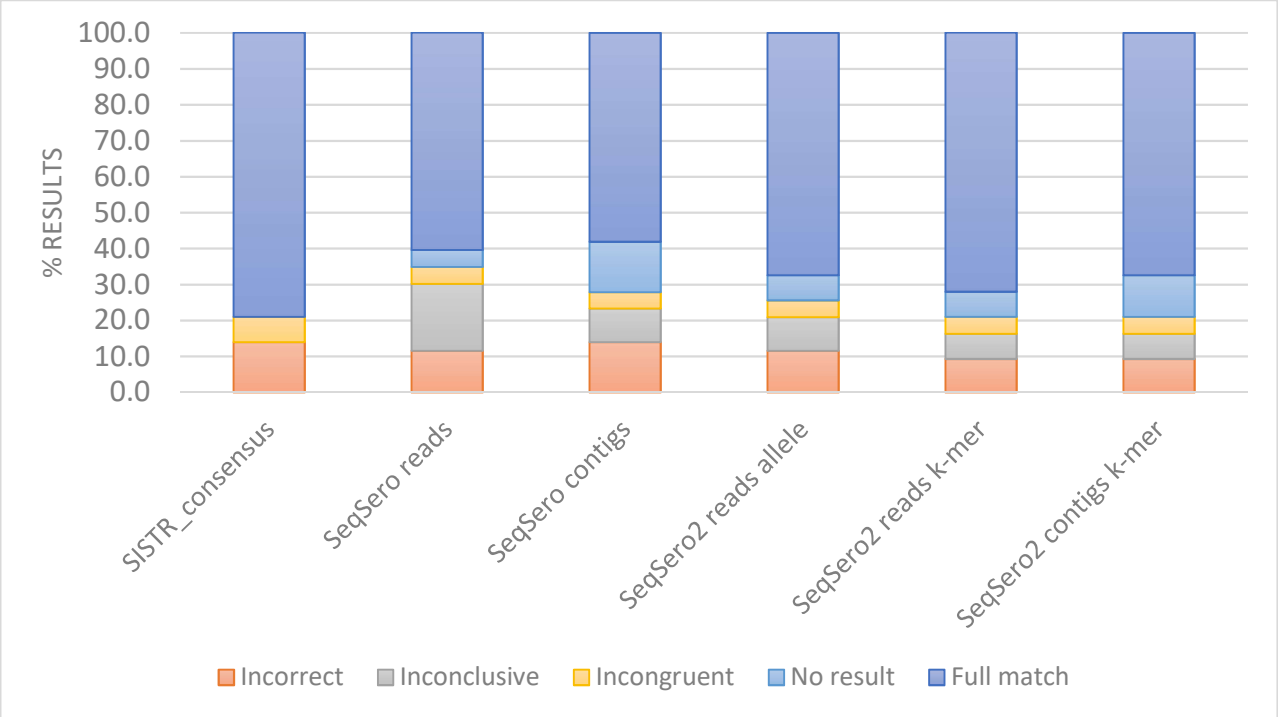
Continuation **Table S1**

	Traditional analyisis (KWS)		Analysis by SeqSero2 using assemblies and k-mer-based workflow		
SAMPLE	**Reported serovar**	**Reported formula**	**Predicted serovar(s)**	**Predicted antigenic profile**	**NOTES**
18PM0128	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0135	Kottbus	6_1,8,e,h,1,5	Kottbus	8:e,h:1,5	-
18PM0158	Anatum		Anatum	3,10:e,h:1,6	-
18PM0174	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0195	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0246	Stourbridge	6, 8, b, 1, 6	Stourbridge	8:b:1,6	-
18PM0262	Dublin	1, 9, 12, g, p	Dublin	9:g,p:-	-
18PM0291	Panama	1, 9, 12, l, v, 1, 5	I 9:i:e,n,z15	9:i:e,n,z15	This predicted serotype is not in the Kauffman-White scheme.
19PM0011	Newport		Newport	8:e,h:1,2	-

Table S2. Performance (% of strains) of the genoserotyping tools used within WGSBAC

	SISTR			SeqSero		SeqSero2		
	antigen	cgMLST	consensus	reads	Contigs	reads allele	reads k-mer	contigs k-mer
Full match	27.9	79.1	79.1	60.5	58.1	67.4	72.1	67.4
Incorrect	14.0	14.0	14.0	11.6	14.0	11.6	9.3	9.3
Inconclusive	58.1	0.0	0.0	18.6	9.3	9.3	7.0	7.0
Incongruent	0.0	7.0	7.0	4.7	4.6	4.7	4.7	4.7
No result	0.0	0.0	0.0	4.7	14.0	7.0	7.0	11.6

Figure S1. Performance (% of results) of the genoserotyping tools used within WGSBAC



9. ACKNOWLEDGMENTS

It is time to sincerely thank all the valuable people who have accompanied me on the journey of the creation of this Doctoral Thesis. A journey of professional and personal development in which I have learned from the following persons.

First, I would like to express my sincere gratitude to Dr. Tomaso, Dr. Linde, and Prof. Neubauer for allowing me to start the first stage of this journey that took me from Madrid to Jena. A special thank you to Dr. Methner for his willingness to support, sharing knowledge and supply samples when the resources were scarce and to Prof. Pfeffer for his always quick and positive answers that supported me when my legs were weak. I dedicate as well my gratitude to the laboratory technicians from the AGr110 group at IBIZ for their technical assistance as well as for the helpful german talks during the coffee breaks. I would like to thank especially my colleagues Fiona Balzer, Dr. Belén González-Santamarina, and Dr. Mostafa Y Abdel-Glil.

On the way, it is necessary to stop from time to time to rest. Sometimes in makeshift shelters and sometimes in others accommodations that could be considered a real home. I especially thank Carla and Xavi to create my improvised family during my first month in Jena. My warmest thanks go to Farzaneh for giving me unique doses of endorphins when the way was hard to continue. Danke, gracias, merci, grazie, shukraan, kam sa hae yo, faleminderit ... to you, my international family (do not blame me for not mentioning your names, you know who you are) thank you for your friendship, patience, and understanding. You taught me the importance to stop from time to time during the way.

And finally, my gratitude goes to the most important people in my life, whom I always carry in my deepest inner place. Thanks to my parents, Juan Miguel and Isabel, they have paved the way and removed many stones from it that I was often not aware of. They have made it possible to pursue my dream of becoming a veterinarian and have allowed me to start many and varied paths giving me the freedom to set my limits. To my grandparents thank you, Luis and Isabel, you never lacked pleasant words of love, warmth, and affection. And to my sister, Isa, her footprints have always been the mold to leave mine and her way is that of a strong fearless woman. The difficulties of this year have affected everyone's paths in some way, but we always have to keep moving forward, leaning on and trusting our family (whether or not you share genes with) and walking without thinking of the goal as the end but as a part of the way.