First draft genome sequence of *Salmonella enterica* subsp. *enterica* serovar Enteritidis isolated from the chicken meat in Russia

Alexey V. Rakov

Somov Institute of Epidemiology and Microbiology, Vladivostok, Russia

BACKGROUND

Salmonella spp. stand out as some of the most common causes of human bacterial food poisoning (EFSA and ECDC, 2018). Specifically, *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*S.* Enteritidis) represented 61.2% of all reported serovars of confirmed human cases in 2017 in the European Union (EFSA and ECDC, 2018) and about 70% of all reported serovars of confirmed human cases in 1988-2018 in Russia (Rakov et al., 2020). Chickens are known to be the main reservoir for *S.* Enteritidis (EFSA and ECDC, 2018). However, there is a lack of information on full-genome sequences of *S.* Enteritidis, isolated in Russia.

PURPOSE

The purpose of the study was to perform the whole genome sequencing (WGS) of the typical *S.* Enteritidis strain isolated from chicken in Russia known to be a major cause of *Salmonella* infection in humans and to compare their genomic characteristics to ones from the strains of other countries.

MATERIALS & METHODS

Bacterial strain. *S.* Enteritidis strain S-25048 was isolated from a local poultry farm in Artyom, Primorsky Krai of Russia from the chicken (*Gallus gallus domesticus*) meat in 2016. The strain was serotyped as *S.* Enteritidis (the seroformula is (1),9,12:g,m:-). Plasmid profile analysis showed that it contains two plasmids of size 59 kb and 2.1 kb.

Whole genome sequencing. Genomic DNA was isolated from overnight culture using the AmpliSens DNA-sorb-B DNA extraction kit (AmpliSens Biotechnologies, Moscow, Russia), and libraries were constructed using the Nextera XT DNA library preparation kit (Illumina, CA, USA), both as per the manufacturers' directions. Whole-genome sequencing was performed on NextSeq 550 platform (Illumina, CA, USA), using High Output Kit v2.5 (300 cycles).

Assembling and annotation. Trimmed reads were assembled *de novo* by using SPAdes version 3.14.1 with default settings (Bankevich et al., 2012), The draft genome assembly quality was assessed with QUAST version 5.0.2 (Gurevich et al., 2013). All contigs with <500 bp were manually excluded. Contigs were manually oriented and ordered according to the reference genome of first sequenced *S.* Enteritidis P125109 (RefSeq Acc. No. NC_011294.1) by using Mauve version 2015-02-26 (Darling et al., 2004). The draft genome sequence was annotated using the automated NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.11 and subsequently deposited at GenBank (Tatusova et al., 2016).

Genomic characterization. The assembled whole genome sequence of S-25048 isolate was uploaded to the next available online tools from the Center of Genomic Epidemiology (CGE) (https://cge.cbs.dtu.dk/): SeqSero 1.2 (Zhang et al., 2015), MLST 2.0 (Larsen et al., 2012), PlasmidFinder 2.1 (Caratolli et al., 2014), pMLST 2.0 (Caratolli et al., 2014), cgMLSTFinder 1.1 (Clausen et al., 2018), ResFinder 4.1 (Bortolaia et al., 2020), SPIFinder 1.0 (Roer et al., 2016), and CSIPhylogeny 1.4 (Kaas et al., 2014).

RESULTS & DISCUSSION

General information. The assembly size was 4,695,145 bp, with an N50 value of 371,562 bp, an average read depth of around 69×, and 52.14% GC content. The assembly size is comparable to that of recently announced *S.* Enteritidis genome sequences, and contains 4,565 coding sequences (CDSs), 4 rRNAs, 62 tRNAs, and 14 noncoding RNAs. Two plasmids were identified and placed as two last contigs of size 58,347 bp and 2,173 bp.

Identification. SeqSero 1.2 and SISTR Web service version 1.0 (Yoshida et al., 2016) confirmed the serotype: Enteritidis with antigenic profile 9:g,m:-.

Typing. Unsurprisingly, *S.* Enteritidis S-20504 isolate belongs to the sequence type ST11 and eBurst group eBG4, representing the most common sequence type of the *S.* Enteritidis isolated from different source worldwide (Achtman et al., 2012), composing about 92.4% of all *S.* Enteritidis STs submitted to EnteroBase MLST database.

Core genome multilocus sequence typing (cgMLST) analysis revealed 139916 profile, exactly the same as in *S.* Enteritidis strain SL1_8 isolated from Russia of unknown origin.

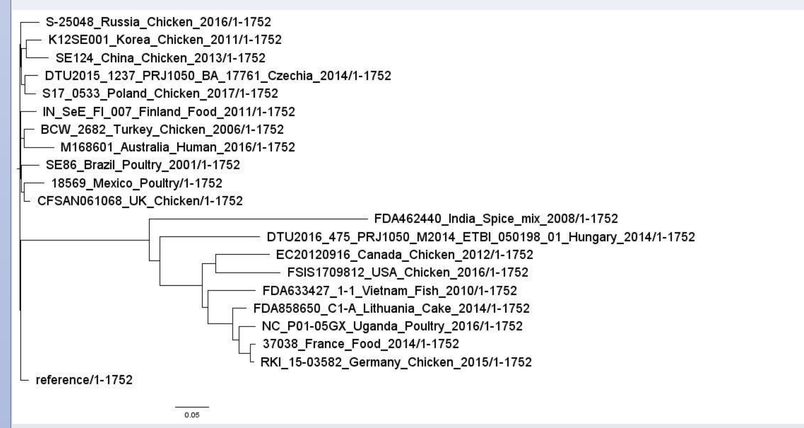
The replicon typing of the plasmids (IncFII(S)/IncFIB(S) and ColpVC) was performed using PlasmidFinder version 2.1. Plasmid MLST typing of 59 kb plasmid identified FAB formula S1:A-:B22 inherent for *S.* Enteritidis virulence plasmid (Villa et al., 2010).

Homologues of *Salmonella* pathogenicity islands (SPIs) 1, 3, 5, 13, and 14 as well as the centisome 63 pathogenicity island (C63PI) were detected in genome using SPIFinder 1.0 with default parameters.

Antibiotic resistance prediction. Using ResFinder version 4.1, only chromosomal-encoded *aac(6')-Iaa* gene, which confers aminoglycoside resistance, and *gyrA* S83Y point mutation, which confers resistance to nalidixic acid and ciprofloxacin, were found.

Phylogeny. We have compared the sequence of S-25048 with publically available sequences of *S.* Enteritidis isolates from 20 countries that were deposited at EnteroBase. We have chosen one chicken *S.* Enteritidis isolate from each of the countries where it available, based on assembly quality (the least number of scaffolds). CSIPhylogeny 1.4 revealed 1752 SNPs and divided 20 genomes from different countries into two clusters, A and B. These obtained data are in agreement with previously shown data based on high discriminatory methods allowing to distinguish this highly homogeneous serovar. It was shown that *S.* Enteritidis may be divided into at least two distinct lineages A and B based on MLVA, cgMLST, and SNP typing methods (Toro et al., 2016; Guard et al., 2020; Ksibi et al., 2020). On the tree, Russian isolate nested with isolates from East Asia (Korea and China) and Eastern Europe (Poland and Czechia) clades of cluster A. All these countries are located close to Russia.

Figure 1. SNP-based maximum likelihood phylogenetic tree of 20 *S.* Enteritidis genomes with reference strain P125109 (NC_011294.1) (UK, 1991). Designations: isolate number, country, host, year of isolation.



Traditional molecular typing techniques, such as pulse-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST), which still used widely and considered as "gold standard" for molecular epidemiology, showed insufficient discriminatory power to distinguish *S.* Enteritidis due to genetic homogeneity of this serovar, where the major genotype contained up to 90% of all typeable isolates (Deng et al., 2015; Tang et al., 2019). Using WGS based typing methods we have a powerful tool to reveal lineages, clades, and subclades among *S.* enteritidis.

Data availability. This Whole Genome Shotgun project of *S.* Enteritidis isolate S-25048 has been deposited at DDBJ/ENA/GenBank under the accession JACEGM000000000 (BioProject, PRJNA638532; BioSample, SAMN15196029). The version described here is version JACEGM010000000.

CONCLUSIONS

This sequence provides the first draft-quality reference for genome assemblies of *S.* Enteritidis, isolated in Russia, and for future virulence analyses or population structure studies of this epidemiologically and clinically relevant *S. enterica* serovar. To our knowledge, this is the first publically deposited annotated genome of this serovar isolated in Russia.

CONTACT

Corresponding Author: Alexey V. Rakov, E-mail: rakovalexey@gmail.com