



Data Article

Draft genome sequences data of four *Salmonella enterica* subsp. *enterica* serovar Dublin archival strains originating from animals in Poland, 1956 – 1957



Milena Skóra^{a,*}, Renata Kwit^a, Magdalena Zając^a, Marta Pietruk^b,
Magdalena Skarżyńska^a, Ewelina Skrzypiec^a, Katarzyna Tłuścik^b,
Anna Lalak^a, Dariusz Wasyl^{a,b}

^aDepartment of Microbiology, National Veterinary Research Institute, Pulawy, Poland

^bDepartment of Omic Analyses, National Veterinary Research Institute, Pulawy, Poland

ARTICLE INFO

Article history:

Received 13 October 2022

Accepted 28 October 2022

Available online 3 November 2022

Dataset link: [Salmonella Dublin isolated from animal source Genome sequencing and assembly \(Original data\)](#)

Dataset link: [Salmonella enterica subsp. enterica serovar Dublin str. ATCC 39184 chromosome, complete sequence \(Reference data\)](#)

Dataset link: [GenomeTrakr Project: US Food and Drug Administration \(Reference data\)](#)

Keywords:

Salmonella Dublin

Cattle

Fox

Whole-genome sequencing

Pathogen

ABSTRACT

Salmonella enterica subsp. *enterica* serovar Dublin (*S. Dublin*) is a zoonotic pathogen causing infections in animals, especially in cattle. In this study, we report draft genome sequences of four *S. Dublin* isolated between 1956 and 1957 from cattle and fox in Poland.

Whole genome sequencing was performed on the Illumina platform and the data is available at National Center for Biotechnology Information under the BioProject accession number PRJNA865912. In order to better understand the genetic basis of epidemiology of *S. Dublin* infection, the obtained sequences were analyzed using the tools which are available at Center of Genomic Epidemiology (<https://www.genomicepidemiology.org/>) including core genome multilocus sequence typing (cgMLST) and core genome single nucleotide polymorphisms (cgSNPs).

© 2022 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: milena.skora@piwet.pulawy.pl (M. Skóra).

Specifications Table

Subject	Biological sciences
Specific subject area	Microbiology: Bacteriology
Type of data	Genome sequence data, table, figure
How the data were acquired	Whole genome sequencing: Illumina MiSeq, Quality control: FastQC v0.11.5, Trimming sequences: Trimmomatic v0.36, Merge trimmed reads: BBMerge from bbtools software suite, Assembly: SPAdes v3.9.0, Assembly statistics: QUAST, Genome annotation: PGAP, Genome analysis: tools from Center of Genomic Epidemiology (CGE), Phylogenetic analysis: MEGA 6, Sequences visualization: Proksee server
Data format	Raw, filtered and assembled genome sequences
Description of data collection	Four lyophilized strains of <i>Salmonella</i> Dublin were revived, streaked onto Xylose Lysine Deoxycholate (XLD) medium and passaged onto nutrient agar. Genomic DNA was isolated from pure nutrient agar culture with the Maxwell RSC cultured cells DNA kit (Promega) sequenced with the Illumina Miseq platform.
Data source location	<ul style="list-style-type: none"> • Institution: National Veterinary Research Institute (PIWet) • City/Town/Region: Pulawy • Country: Poland • Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: PIW15: 52.73371238 N 15.23034418 E PIW16, PIW19, PIW21: 54.46904475 N 17.04150334 E
Data accessibility	Assembled sequences and raw reads have been deposited in GenBank under the BioProject accession number PRJNA865912 and BioSample accession number: PIW 15 (SAMN30076191), PIW 16 (SAMN30076192), PIW 19 (SAMN30076193), PIW 21 (SAMN30076194).

Value of the Data

- The draft genome data may be useful for estimating the degree of genetic diversity of *Salmonella* Dublin strains.
- The scientists could use these genome data for comparative genome analysis and also identify evolutionary changes among *Salmonella* Dublin.
- The draft genome data allow better understanding of *Salmonella* epidemiology.

1. Objective

The incidence of *S. Dublin* infection in cattle is a major problem from both animal and human health perspective. Whole genome sequencing allows accurate genome analysis of pathogens, invasiveness and pathogenicity mechanisms. This study describes the *S. Dublin* sequences genome to better understand the evolutionary changes that have occurred over dozens of years.

2. Data Description

Salmonella enterica serovar Dublin (*S. Dublin*) is one of the host-specific serovars adapted to cattle. Infections are detected in both calves and adult animals and can cause substantial losses in livestock production. Symptoms of salmonellosis include diarrhea, fever, loss of appetite, in pregnant cows abortion may occur [1]. *S. Dublin* was confirmed also in foxes and other fur-bearing animals [2] and can spread to different species as a result of interspecies transmission

Table 1Genome characteristics of *S. Dublin* sequences.

Isolate	PIW 15	PIW 16	PIW 19	PIW 21
Species	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>
Serovar	Dublin	Dublin	Dublin	Dublin
Isolation date	1956	1956	1957	1956
Lyophilization date	1958	1960	1958	1958
Host	cattle	cattle (calf)	cattle	fox
Sample type	feces	internal organs	internal organs	internal organs
Genome size (bp)	4,875,403	4,876,577	4,881,137	4,878,038
No. of contigs	31	34	36	37
Total no. of reads	1,641,274	1,545,018	1,570,976	1,062,786
Overall read coverage (x fold)	101	95	97	66
%GC	52.1	52.1	52.1	52.1
N ₅₀ (bp)	679 274	679 108	560 267	480 263
MLST type	ST 10	ST 10	ST 10	ST 10
cgMLST	219,058	279,553	279,536	12,682
Plasmid replicon (identity%)	IncFII(S) 97.71, IncX1 98.66	IncFII(S) 97.71, IncX1 98.66	IncFII(S) 97.71, IncX1 98.66	Col(pHAD28) 91.15, IncFII(S) 97.71, IncX1 98.66
BioSample no.	SAMN30076191	SAMN30076192	SAMN30076193	SAMN30076194
Genome Accession no.	JANKYP000000000	JANKYQ000000000	JANKYR000000000	JANKYS000000000

[3]. Human infections are caused by the consumption of food contaminated e.g. milk, dairy products [4], and after contact with infected animals [5]. Here, we present draft genome sequences of four isolates *S. Dublin* from cattle and fox isolated in the National Veterinary Research Institute (PIWet) between 1956 and 1957 and lyophilized between 1958 and 1960 and stored under refrigerated conditions until current use. *S. Dublin* genomes varied from 4,875,403 to 4,881,137 bp, and had between 31 and 37 contigs with an average GC content of 52.1%. More information about the described sequences was included in Table 1.

Core genome SNPs were used to create a phylogenetic tree (Fig. 1) of described isolates, the reference *S. Dublin* (ATCC 39184) and the another 20 genome sequences of cattle *S. Dublin* [6] available from NCBI database. The comparison of PIW15, PIW16, PIW19 and PIW21 sequences was also presented in the form of a circular map (Fig. 2).

3. Experimental Design, Materials and Methods

The lyophilized strains were rehydrated by adding 0,5 ml NaCl 0.85% Medium (bioMérieux) to each ampoule, the contents were gently mixed, then 10 μ l of each suspensions were streaked onto Xylose Lysine Deoxycholate (XLD) medium and incubated overnight at 37 \pm 1 $^{\circ}$ C. Single colonies were passaged overnight at 37 \pm 1 $^{\circ}$ C onto nutrition agar and used for further analysis [7]. Strains identity was validated on the basis of biochemical reaction (VITEK System, bioMérieux), matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) using the extraction method following the producer guidelines (Bruker Daltonik GmbH) and serologically according to the White-Kaufmann-Le Minor scheme. Isolation of genomic DNA was executed using Maxwell Rapid Sample Concentrator (RSC) cultured cells DNA Kit (Promega). Sequencing libraries were constructed using the Nextera XT sample preparation kit following the manufacturer's recommendations and evaluated by capillary electrophoresis (Fragment Analyzer). Whole genome sequencing was performed on the MiSeq platform (Illumina) with the MiSeq reagent kit (2 \times 300-bp paired-end protocol, to 100 \times depth of sequencing). Default parameters were used for all software unless otherwise specified. Raw paired-end reads were quality controlled using FastQC v0.11.5 [8] (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Trimmomatic 0.36 [9] was used to trimmed and removed adapter sequence with the following pa-

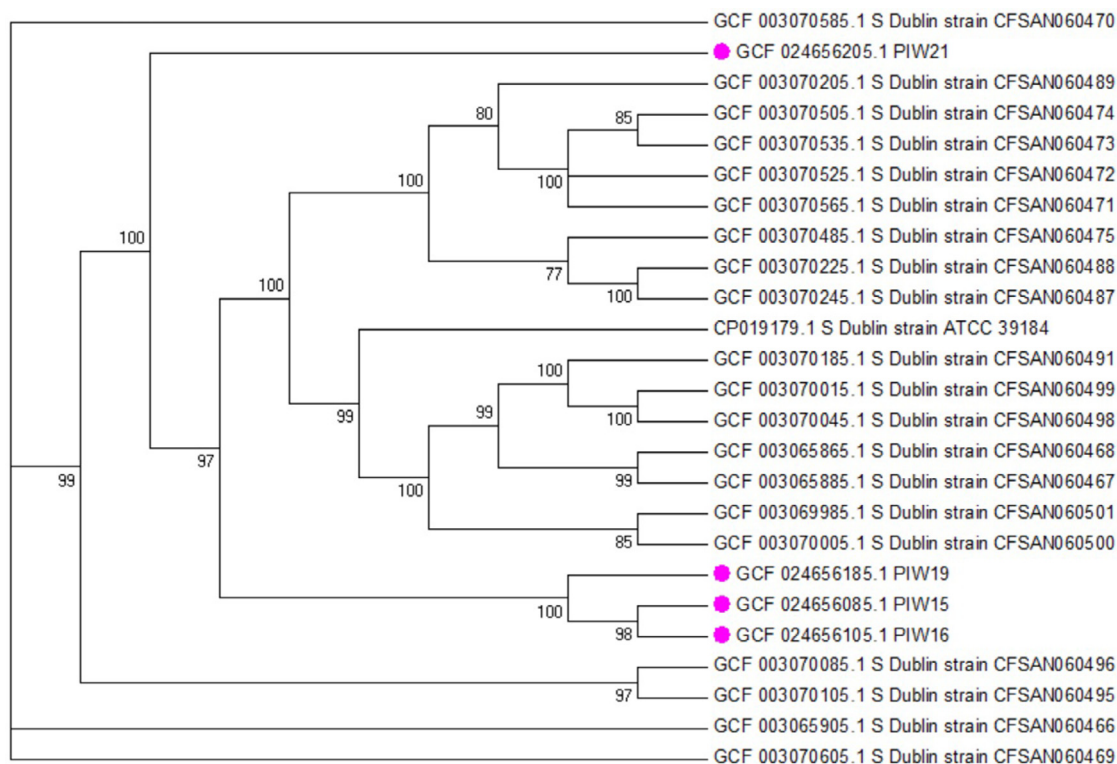


Fig. 1. Phylogenetic tree of *S. Dublin* isolates (reported strains are marked with pink circles) based on cgSNP and reference strains downloaded from NCBI database. The bootstrap values are shown on branches. Tree visualization was made in MEGA 6.

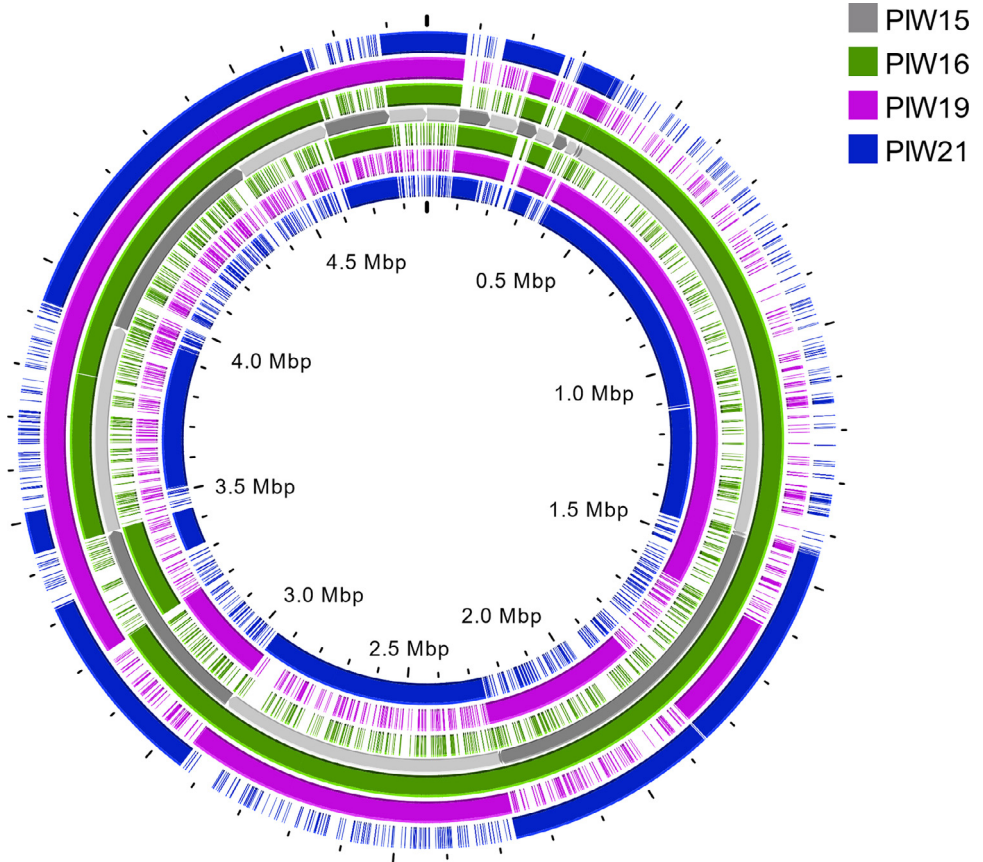


Fig. 2. Circular map of the comparison of PIW15 (used as the reference), PIW16, PIW19, and PIW21. Outside from the PIW15 are CDSs on the forward strands, inside from PIW15 are CDSs on the reverse strands.

rameters: ILLUMINACLIP: 2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:36. To merge the trimmed reads the BBMerge from bbtools software suite was used (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmerge-guide/>) and assembled using SPAdes v3.9.0 [10] with the “-careful” flag. The genome statistics and annotation of the *Salmonella* strains were determined using the Quality Assessment Tool for Genome Assemblies (QUAST) [11] and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [12]. Bioinformatics tools from Center of Genomic Epidemiology (CGE) have been used to determine MLST type (MLST 2.0) and cgMLST (cgMLSTFinder 1.2) [13,14]. A phylogenetic tree was created in CSIPhylogeny and visualized in MEGA 6 [15]. The similarity of the sequences is shown on the circular map and generated in Proksee server (<https://proksee.ca/>).

Ethics Statements

Not required.

CRedit Author Statement

Milena Skóra: Investigation, Writing – Original Draft; **Renata Kwit:** Investigation, Writing – review & editing; **Magdalena Zajac:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing; **Marta Pietruk:** Writing – review & editing; **Magdalena Skarzyńska:** Writing – review & editing; **Ewelina Skrzypiec:** Writing – review & editing; **Katarzyna Tłuścik:** Writing – review & editing; **Anna Lalak:** Writing – review & editing; **Dariusz Wasyl:** Data curation, Writing – review & editing.

Declaration of Competing Interest

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

Data Availability

[Salmonella Dublin isolated from animal source Genome sequencing and assembly \(Original data\)](#) (National Center for Biotechnology Information)

[Salmonella enterica subsp. enterica serovar Dublin str. ATCC 39184 chromosome, complete sequence \(Reference data\)](#) (National Center for Biotechnology Information)

[GenomeTrakr Project: US Food and Drug Administration \(Reference data\)](#) (National Center for Biotechnology Information)

Acknowledgments

This work was supported by funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement No 773830: One Health European Joint Programme and from Ministry of Science funds for science in 2018–2022 allocated to the implementation of an international co-financed project, agreement No. 3932/H2020/2018/2.

References

- [1] L.F. Costa, T.A. Paixão, R.M. Tsolis, A.J. Bäumler, R.L. Santos, Salmonellosis in cattle: advantages of being an experimental model, *Res. Vet. Sci.* 93 (2012) 1–6, doi:[10.1016/j.rvsc.2012.03.002](https://doi.org/10.1016/j.rvsc.2012.03.002).
- [2] H.H. Dietz, M. Chriél, T.H. Andersen, J.C. Jørgensen, M. Torpdahl, H. Pedersen, K. Pedersen, *Article Outbreak of Salmonella Dublin-associated abortion in Danish fur farms, 2006*.
- [3] W. Glawischnig, J. Lazar, A. Wallner, C. Kornschober, Cattle-derived *Salmonella enterica* serovar Dublin Infections in Red Foxes (*Vulpes vulpes*) in Tyrol, Austria, *J. Wildl. Dis.* 53 (2017) 361–363, doi:[10.7589/2016-04-087](https://doi.org/10.7589/2016-04-087).
- [4] M. Mohammed, S. Hello, P. Leekitcharoenphon, R. Hendriksen, The invasive of *Salmonella* Dublin as revealed by whole genome sequencing, *BMC Infect. Dis.* (2017) 17, doi:[10.1186/s12879-017-2628-X](https://doi.org/10.1186/s12879-017-2628-X).
- [5] A. Andino, I. Hanning, *Salmonella enterica*: survival, colonization, and virulence differences among serovars, *ScientificWorldJ.* (2015) 2015, doi:[10.1155/2015/520179](https://doi.org/10.1155/2015/520179).
- [6] F. Campioni, F.P. Vilela, G. Cao, G. Kastanis, D. Miller, M.S. Leon, M.R. Tiba-Casas, S.A. Fernandes, D. dos P. Rodrigues, R.G. Costa, M.W. Allard, J.P. Falcão, Draft genome sequences of 112 salmonella enterica serovar dublin strains isolated from humans and animals in Brazil, *Genome Announc* 6 (2018), doi:[10.1128/GENOMEA.00405-18](https://doi.org/10.1128/GENOMEA.00405-18).
- [7] ISO - ISO/TR 6579-3:2014 - Microbiology of the food chain – Horizontal method for the detection, enumeration and serotyping of *Salmonella* – Part 3: guidelines for serotyping of *Salmonella* spp., (n.d.). <https://www.iso.org/standard/56714.html> (accessed September 8, 2022).
- [8] S. Andrews, *FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010*.
- [9] A.M. Bolger, M. Lohse, B. Usadel, *Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data*, 30 (2014) 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- [10] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477, doi:[10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).

- [11] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, Genome analysis QUASt: quality assessment tool for genome assemblies, 29 (2013) 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- [12] M.V. Larsen, S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R.L. Marvig, L. Jelsbak, T. Sicheritz-Pontén, D.W. Ussery, F.M. Aarestrup, O. Lund, Multilocus sequence typing of total-genome-sequenced bacteria, *J. Clin. Microbiol.* 50 (2012) 1355–1361, doi:[10.1128/JCM.06094-11/FORMAT/EPUB](https://doi.org/10.1128/JCM.06094-11/FORMAT/EPUB).
- [13] P.T.L.C. Clausen, F.M. Aarestrup, O. Lund, Rapid and precise alignment of raw reads against redundant databases with KMA, *BMC Bioinformatics* 19 (2018), doi:[10.1186/S12859-018-2336-6](https://doi.org/10.1186/S12859-018-2336-6).
- [14] N.F. Alikhan, Z. Zhou, M.J. Sergeant, M. Achtman, A genomic overview of the population structure of *Salmonella*, *PLoS Genet.* 14 (2018), doi:[10.1371/JOURNAL.PGEN.1007261](https://doi.org/10.1371/JOURNAL.PGEN.1007261).
- [15] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.* 30 (2013) 2725, doi:[10.1093/MOLBEV/MST197](https://doi.org/10.1093/MOLBEV/MST197).