



HAL
open science

Complete genome sequence of *Yersinia pseudotuberculosis* strain SP-1303 from lineage 8, associated with Far East scarlet-like fever

Marion Lemarignier, Cyril Savin, Pierre Lê-Bury, Olivier Dussurget, Javier Pizarro-Cerdá

► To cite this version:

Marion Lemarignier, Cyril Savin, Pierre Lê-Bury, Olivier Dussurget, Javier Pizarro-Cerdá. Complete genome sequence of *Yersinia pseudotuberculosis* strain SP-1303 from lineage 8, associated with Far East scarlet-like fever. *Microbiology Resource Announcements*, 2023, pp.e00838-23. 10.1128/MRA.00838-23 . pasteur-04266523

HAL Id: pasteur-04266523

<https://pasteur.hal.science/pasteur-04266523>

Submitted on 31 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Complete genome sequence of *Yersinia pseudotuberculosis* strain SP-1303 from lineage 8, associated with Far East scarlet-like fever

Marion Lemarignier,¹ Cyril Savin,^{1,2} Pierre Lê-Bury,¹ Olivier Dussurget,¹ Javier Pizarro-Cerdá^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genome sequence of *Yersinia pseudotuberculosis* strain SP-1303, identified as part of lineage 8 and associated with Far East scarlet-like fever. The genome includes the chromosome, the *Yersinia*-virulence plasmid (pYV) encoding a type III secretion system essential for virulence, the pVM82 plasmid, and two cryptic plasmids.

KEYWORDS *Yersinia pseudotuberculosis*, Far East scarlet-like fever, genome

Yersinia pseudotuberculosis is one of the 26 species of the genus *Yersinia*, a member of the *Yersiniaceae* family from the order *Enterobacterales* (1, 2). *Y. pseudotuberculosis* is a foodborne pathogenic bacterium, known to cause self-limiting acute enteropathy in humans, taking the form of mesenteric lymphadenitis and gastroenteritis with fever, abdominal pain, and occasionally diarrhea, which in rare cases takes a septicemic form (3–5). The Far East scarlet-like fever (FESLF) is another form of yersiniosis, first reported in 1959 during an epidemic in Vladivostok and mainly present in Far Eastern countries, with symptoms including fever, hyperemic tongue, skin rash, jaundice, and desquamation (6, 7). A single complete genome sequence of *Y. pseudotuberculosis* strain IP31758 causing FESLF was published in 2007 by M. Eppinger and colleagues (8). However, the strain IP31758 does not harbor the *Yersinia*-virulence plasmid (pYV) encoding a type III secretion system essential for virulence, which is often lost under laboratory culture conditions (9), making the IP31758 of limited use. The strain SP-1303 was isolated from mouse stools in 1988 in Moscow. It was recovered by the Shimane Prefectural Institute of Public Health and Environmental Science (Matsue, Japan) and given at –80°C in LB-glycerol to the *Yersinia* Research Unit by Professor H. Fukushima in 2011. The strain SP-1303 belongs to the *Y. pseudotuberculosis* lineage 8 together with strains associated with FESLF (including strain IP31758) and harbors the pYV, allowing for the first time to fully understand the complete repertoire of virulence factors displayed by these strains, which is of major interest to study this poorly characterized disease (10).

Genomic DNA from *Y. pseudotuberculosis* strain SP-1303, grown in LB medium at 28°C for 16 h, was extracted using the PureLink Genomic DNA kit (Invitrogen) according to the manufacturer's instruction, except the final elution step carried out in H₂O (Ambion). Two complementary sequencing technologies were applied using the same purified DNA: Oxford Nanopore Libraries (ONT) long reads and Illumina short reads (paired-end reads of 150 nucleotides). Libraries were prepared using the Rapid Barcoding kit (SQK-RBK004) for R9.4.2 ONT MinION flow cell sequencing, and the Nextera kit (Illumina) for Illumina NextSeq 550 sequencing. ONT Guppy v6.4.2 was used for long reads base-calling in high-accuracy mode (configuration dna_r9.4.1_450bps_hac.cfg). All reads were provided as FASTQ files and their information is detailed in Table 1. Hybrid genome assembly with both Illumina and ONT sequencing data was carried out as previously described by R. Wick and colleagues (11). Briefly, reads were filtered with FilTlong (<https://github.com/rwick/FilTlong>) (ONT) and fqCleanER ([**Editor** André O. Hudson, Rochester Institute of Technology, Rochester, New York, USA](https://</p></div><div data-bbox=)

Address correspondence to Javier Pizarro-Cerdá, pizarroj@pasteur.fr.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 6 September 2023

Accepted 26 September 2023

Published 31 October 2023

Copyright © 2023 Lemarignier et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Assembly and general features of *Y. pseudotuberculosis* strain SP-1303

Parameter	Whole genome	Chromosome	pVM82	pYV	pYpstbSP1303.1	pYpstbSP1303.2
GenBank accession no.		CP130901	CP130902	CP130903	CP130904	CP130905
No. of circular contigs	5	1	1	1	1	1
Size (bp)	4,979,983	4,741,829	153,152	74,956	5,546	4,500
G + C content (%)	45.6	47.5	40	44.5	48	48
No. of annotated protein-coding genes	4,186	3,939	161	73	6	7
No. of pseudogenes	155	126	3	24	1	1
No. of rRNAs	22	22	0	0	0	0
No. of tRNAs	84	84	0	0	0	0
No. of ncRNAs	10	10	0	0	0	0
No. of CRISPR arrays	2	2	0	0	0	0
Total no. of reads	17,811 (ONT); 5,355,151 (Illumina)					
ONT reads information (20)	N50 length: 18,828; mean length: 10,193; median length: 6,400; max length: 109,206					
Total no. of bases	181,551,528 (ONT); 1,572,500,161 (Illumina)					
Sequencing depth (x)	36.5 (ONT); 284 (Illumina)					

gitlab.pasteur.fr/GIPhy/fqCleanER (Illumina) for assembly using Tricycler/0.5.0 (12), Flye v2.9 (13), Raven/1.6.0 (14), Miniasm/0.3, and Minipolish/0.1.3 (15). Assembled reads were polished with Medaka/1.4.4 (<https://github.com/nanoporetech/medaka>) (ONT) and Polypolish/0.5.0 (16) (Illumina). Contigs were circularized and rotated by Tricycler/0.5.0 (12). Genome annotation was performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (17–19) (Table 1).

Our work reveals a circular chromosome of 4.7 Mb encoding the superantigen Ypma (21), the pYV (74.9 kb) plasmid found in pathogenic *Yersinia* species encoding a type III secretion system (22), and the pVM82 (153.1 kb) plasmid also present in the IP31758 FESLF strain, which harbors a type IVB secretion system that could be involved in the pathogenesis of FESLF (8, 23–26). Two additional cryptic plasmids were identified with MOB-suite (27, 28): pYpstbSP1303.1 (5.5 kb) and pYpstbSP1303.2 and compared using BLAST. The plasmid pYpstbSP1303.1 is close to plasmids present in *Klebsiella pneumonia* (identity: 98.15%, query cover: 54%), *Escherichia coli* (identity: 97.96%, query cover: 49%), and *Salmonella enterica* (identity: 97.84%, query cover: 49%). The second plasmid pYpstbSP1303.2 (4.5 kb) resembles the pHW66 plasmid of *Rhanella sp.* (identity: 95.24%, query cover: 63%), a ColE-2-like plasmid (29). No antimicrobial resistance genes were detected using ResFinder 4.3.3 (30).

ACKNOWLEDGMENTS

We thank Professor H. Fukushima for generously providing the *Y. pseudotuberculosis* SP-1303 strain. We thank the Plateforme de microbiologie mutualisée (P2M) of Institut Pasteur for Illumina sequencing.

The project received funding from Institut Pasteur and LabEX Integrative Biology of Emerging Infectious Diseases (ANR-10-LBX-62-IBEID). M.L. was supported by a MENESR (Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche) fellowship. P.L.-B. was supported by Agence de l'Innovation de Défense (AID-DGA), Fondation pour la Recherche Médicale (FDT202204015222), and the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005).

AUTHOR AFFILIATIONS

¹Institut Pasteur, Université Paris Cité, CNRS UMR6047, *Yersinia* Research Unit, Paris, Ile de France, France

²Institut Pasteur, Université Paris Cité, *Yersinia* National Reference Laboratory, WHO Collaborating Research & Reference Centre for Plague FRA-140, Paris, Île-de-France, France

AUTHOR ORCID*s*

Marion Lemarignier  <http://orcid.org/0000-0003-4025-610X>

Cyril Savin  <http://orcid.org/0000-0003-2659-3433>

Pierre Lê-Bury  <http://orcid.org/0000-0002-6010-2952>

Olivier Dussurget  <http://orcid.org/0000-0001-9924-0837>

Javier Pizarro-Cerdá  <http://orcid.org/0000-0002-4343-0508>

FUNDING

Funder	Grant(s)	Author(s)
Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche		Marion Lemarignier
French Ministry of Armed Forces Agence de l'innovation de Défense (AID)	AID-DGA	Pierre Lê-Bury
Fondation pour la Recherche Médicale (FRM)	FDT202204015222	Pierre Lê-Bury
Inception program	ANR-16-CONV-0005	Pierre Lê-Bury

AUTHOR CONTRIBUTIONS

Marion Lemarignier, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Cyril Savin, Data curation, Formal analysis, Writing – review and editing | Pierre Lê-Bury, Conceptualization, Data curation, Methodology, Writing – review and editing | Olivier Dussurget, Funding acquisition, Supervision, Writing – review and editing | Javier Pizarro-Cerdá, Funding acquisition, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome sequence of *Y. pseudotuberculosis* strain SP-1303 has been submitted to NCBI under the BioProject ID [PRJNA998338](https://ncbi.nlm.nih.gov/bioproject/PRJNA998338) and the Biosample accession number [SAMN36701287](https://ncbi.nlm.nih.gov/sra/SAMN36701287). The assembled genome is available in GenBank under assembly accession number [GCA_030644845](https://ncbi.nlm.nih.gov/genbank/GCA_030644845). Sequencing raw reads are available on Sequencing Read Archive (SRA) accession numbers [SRR25411577](https://ncbi.nlm.nih.gov/sra/SRR25411577) (Illumina) and [SRR25411576](https://ncbi.nlm.nih.gov/sra/SRR25411576) (ONT). In addition, the genome sequence is also available on Yersiniomics, a multi-omics interactive web-based platform for easy visualization of genome organization, rapid access to structural and functional protein properties, and search for protein homologs within the *Yersinia* genus (31).

REFERENCES

- Adeolu M, Alnajjar S, Naushad S, Gupta R. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacteriales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* 66:5575–5599. <https://doi.org/10.1099/ijsem.0.001485>
- Le Guern A-S, Savin C, Angermeier H, Brémont S, Clermont D, Mühle E, Orozova P, Najdenski H, Pizarro-Cerdá J. 2020. *Yersinia artesisiana* sp. nov., *Yersinia proxima* sp. nov., *Yersinia alsatica* sp. nov., *Yersinia vastinensis* sp. nov., *Yersinia thracica* sp. nov. and *Yersinia occitanica* sp. nov., isolated from humans and animals. *Int J Syst Evol Microbiol* 70:5363–5372. <https://doi.org/10.1099/ijsem.0.004417>
- Carniel E, Mollaret HH. 1990. Yersiniosis. *Comp Immunol Microbiol Infect Dis* 13:51–58. [https://doi.org/10.1016/0147-9571\(90\)90516-v](https://doi.org/10.1016/0147-9571(90)90516-v)
- Attwood SE, Mealy K, Cafferkey MT, Buckley TF, West AB, Boyle N, Healy E, Keane FB. 1987. *Yersinia* infection and acute abdominal pain. *Lancet* 329:529–533. [https://doi.org/10.1016/s0140-6736\(87\)90175-9](https://doi.org/10.1016/s0140-6736(87)90175-9)
- Ljungberg P, Valttonen M, Harjola VP, Kaukoranta-Tolvanen SS, Vaara M. 1995. Report of four cases of *Yersinia pseudotuberculosis* septicemia and a literature review. *Eur J Clin Microbiol Infect Dis* 14:804–810. <https://doi.org/10.1007/BF01690998>
- Amphlett A. 2016. Far East scarlet-like fever: a review of the epidemiology, symptomatology, and role of superantigenic toxin: *Yersinia pseudotuberculosis*-derived mitogen A. *Open Forum Infect Dis* 3:fv202. <https://doi.org/10.1093/ofid/ofv202>
- Timchenko NF, Adgamov RR, Popov AF, Psareva EK, Sobyanyan KA, Gintsburg AL, Ermolaeva SA. 2016. Far East scarlet-like fever caused by a few related genotypes of *Yersinia pseudotuberculosis*, Russia. *Emerg Infect Dis* 22:503–506. <https://doi.org/10.3201/eid2203.150552>

8. Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, Fayolle C, Lindler LE, Carniel E, Ravel J. 2007. The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. *PLoS Genet* 3:e142. <https://doi.org/10.1371/journal.pgen.0030142>
9. Cornelis GR, Boland A, Boyd AP, Geuijen C, Iriarte M, Neyt C, Sory MP, Stainier I. 1998. The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol Mol Biol Rev* 62:1315–1352. <https://doi.org/10.1128/MMBR.62.4.1315-1352.1998>
10. Savin C, Criscuolo A, Guglielmini J, Le Guern A-S, Carniel E, Pizarro-Cerdá J, Brisse S. 2019. Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. *Microb Genom* 5:e000301. <https://doi.org/10.1099/mgen.0.000301>
11. Wick RR, Judd LM, Holt KE. 2023. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* 19:e1010905. <https://doi.org/10.1371/journal.pcbi.1010905>
12. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Tricycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 22:266. <https://doi.org/10.1186/s13059-021-02483-z>
13. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
14. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 1:332–336. <https://doi.org/10.1038/s43588-021-00073-4>
15. Wick RR, Holt KE. 2019. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 8:2138. <https://doi.org/10.12688/f1000research.21782.4>
16. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>
17. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
18. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>
19. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
20. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* 35:523–525. <https://doi.org/10.1093/bioinformatics/bty654>
21. Carnoy C, Mullet C, Müller-Alouf H, Leteurtre E, Simonet M. 2000. Superantigen YPMa exacerbates the virulence of *Yersinia pseudotuberculosis* in mice. *Infect Immun* 68:2553–2559. <https://doi.org/10.1128/IAI.68.5.2553-2559.2000>
22. Cornelis GR. 2002. The *Yersinia* Ysc-Yop 'type III' weaponry. *Nat Rev Mol Cell Biol* 3:742–752. <https://doi.org/10.1038/nrm932>
23. Shubin FN, Gintsburg AL, Kitaev VM, Ianishevskii NV, Zenkova ZG. 1989. Analysis of the plasmid composition of *Yersinia pseudotuberculosis* strains and its use for typing *pseudotuberculosis* pathogens. *Mol Gen Mikrobiol Virusol*:20–25.
24. Gintsburg AL, Shubin FN, Shovadaeva GA, Kulichenko AN, Ianishevskii NV. 1988. A new pathogenic trait encoded by *Yersinia pseudotuberculosis* pVM82 plasmid. *Genetika* 24:1562–1571.
25. Dubrovina VI, Golubinskiĭ EP, Borsuk GI, Balakhonov SV, Konvalova ZA. 1999. Characteristics of phagocytosis of *Yersinia pseudotuberculosis* with various sets of plasmids. *Med Parazitol (Mosk)*:50–53.
26. Shovadaeva GA, Markov AP, Pokrovskaja MS, Chetina EV, Ianishevskii NV, Shubin FN, Gintsburg AL, Smirnov GB. 1990. Structural-functional characteristics of *Yersinia pseudotuberculosis* plasmid pVM82. *Genetika* 26:621–629.
27. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 4:e000206. <https://doi.org/10.1099/mgen.0.000206>
28. Robertson J, Bessonov K, Schonfeld J, Nash JHE. 2020. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. *Microbial Genomics* 6:mgen000435. <https://doi.org/10.1099/mgen.0.000435>
29. Rozhon W, Petutschnig E, Khan M, Summers DK, Poppenberger B. 2010. Frequency and diversity of small cryptic plasmids in the genus *Rahnella*. *BMC Microbiol* 10:56. <https://doi.org/10.1186/1471-2180-10-56>
30. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. 2022. ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* 8:000748. <https://doi.org/10.1099/mgen.0.000748>
31. Lê-Bury P, Druart K, Savin C, Lechat P, Mas Fiol G, Matondo M, Bécavin C, Dussurget O, Pizarro-Cerdá J. 2023. Yersiniomics, a multi-omics interactive database for *Yersinia*. *Microbiol Spectr* 11:e03826–22. <https://doi.org/10.1128/spectrum.03826-22>