## The Brazilian Journal of
# INFECTIOUS DISEASES

## Brief communication

# Next-generation sequencing virulome analysis of a *Yersinia enterocolitica* subsp. *palearctica* bioserotype 4/O:3 ST18 isolated from human blood in Brazil

*Leonardo Alves Rusak*[a,*], *Ricardo Magrani Junqueira*[b], *Ernesto Hofer*[c], *Deyse Christina Vallim*[c], *Marise Dutra Asensi*[a]

[a] *Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Laboratório de Pesquisa em Infecção Hospitalar, Rio de Janeiro, RJ, Brazil*
[b] *Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Laboratório de Biologia Computacional e Sistemas, Rio de Janeiro, RJ, Brazil*
[c] *Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Laboratório de Zoonoses Bacterianas/Setor Listeria, Rio de Janeiro, RJ, Brazil*

## ARTICLE INFO

## ABSTRACT

*Yersinia enterocolitica* is a widespread Gram-negative bacterium that causes gastrointestinal disease and other clinical manifestations in humans. Potentially pathogenic *Y. enterocolitica* has been isolated in Brazil, from human, environmental, food, and animal sources. Herein we report a genome sequence of *Y. enterocolitica* subsp. *palearctica* strain YE 19, serotype O:3, biotype 4, sequence type 18, with virulence determinants isolated from human blood in Rio de Janeiro in 2005. The results corroborate other findings that this strain harbors a set of virulence determinants that could play a role in host pathoadaptation and may also justify the successful dissemination of bioserotype 4/O:3 in Brazil. The presence of strains harboring all of these virulence genes in Brazil is a potential threat to young children and immunocompromised individuals, for whom yersiniosis are a significant source of morbidity and mortality. The results of a genomic data analysis will help understand the virulence of Brazilian strains and provide data for *Y. enterocolitica* studies worldwide.

*Yersinia enterocolitica* is an enteric pathogen with a wide range of clinical and immunological manifestations, which depend on age and physical condition of the patient. Responsible for intestinal diseases, the most frequent occurrence in infants and children, including enterocolitis with an inflammatory diarrhea; however, in older children and young adults the symptoms include acute terminal ileitis and mesenteric lymphadenitis (mimicking appendicitis). On the other hand, in some cases, extra-intestinal manifestations have also been reported, including urinary tract and respiratory tract infection (empyema), osteoarticular infection (reactive arthritis), erythema nodosum, infected mycotic aneurysm, axillary abscesses, and endocarditis.[1,2] *Y. enterocolitica* is the major

causative agent of yersiniosis worldwide and is typically transmitted via the fecal–oral route to humans and animals.[3] *Y. enterocolitica* can be classified by biochemical characteristics and serotyped into six biotypes (BT) (1A, 1B, 2, 3, 4, and 5).[1] Strains that are non-pathogenic to humans and animals belong to BT1A, whereas highly pathogenic strains belong to BT1B; BT2–5 are low to moderate-pathogenic strains. The serotype O:3 biotype 4 comprises about 80–90% of human isolates in Germany and Europe, with rising global relevance.[4]

Little attention has been given to this microorganism in Brazil, where it is still poorly studied. Although Brazil has no official data about the incidence of this pathogen, sporadic cases have been published and some studies demonstrate the pathogenic potential of strains isolated from many sources, including humans, the environment, food, and animals. The bioserotype 4/O:3 is the main bioserotype isolated from human clinical cases.[5,6]

The aim of this brief communication was to analyze the virulence profile and announces the genome sequence of Y. *enterocolitica* subsp. *palearctica*, strain YE 19, serovar O:3, biotype 4 isolated in 2005 from blood of a hospitalized patient in Rio de Janeiro. It was deposited in the Collection of Listeria (CLIST) located at the Bacterial Zoonoses Laboratory (LABZOO/Oswaldo Cruz Institute/Oswaldo Cruz Foundation) under the accession number CLIST 4117. Sequence type 18 (ST18) was determined by multilocus sequence typing (MLST) and by use of the *Yersinia* MLST website. The ST18 comprises the main Y. *enterocolitica* serotype O:3 sequence type based on MLST database.

A next-generation sequencing (NGS) technique was used to analyze the virulome of YE 19 strain. Genomic DNA was extracted using DNeasy® Blood & Tissue Kit (Qiagen, Germany) and the DNA library was prepared with Nextera XT (Illumina, San Diego, CA). Sequencing was performed on an Illumina HiSeq 2500 platform (Illumina Inc., USA) generating 100 bp paired-end reads. The sequence data were processed using Trimmomatic (v0.35) and paired-end reads were merged using FLASh (v1.2.11).[7] De novo assembly was performed using SPAdes (v3.7.1).[8] Ragout (v1.2)[9] was utilized for reference assisted scaffolding of contigs against Y. *enterocolitica* subsp. *palearctica* Y11 genome (NC_017564.1). Sealer (v1.9.0)[10] was used for gap filling and Pilon (v1.16)[11] performed polishing. The resulting assembly was evaluated using REAPR (v1.0.18).[12] The NCBI Prokaryotic Genome Annotation Pipeline (released 2013) annotated the final genome model comprising 19 scaffolds with a total length of 4,550,059 bp and 47% GC. The summary of genomic statistics is reported in Table 1. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LYBL00000000. The version described in this paper is version LYBL01000000.

The presence of virulence factors was investigated using SRST2 program (v0.1.7) against the Virulence Factor Database (VFDB). The details of all virulence genes found are reported in supplementary file 1. Fig. 1 is built to show where each virulence gene is located on YE 19 genome.

A total of 103 virulence genes have been identified from the VFDB. These virulence genes are related to motility, the secretion system, toxins, the O antigen (lipopolysaccharide – LPS), hemin uptake, invasion, and adherence.

**Table 1 – Genome features.**

| Property | YE 19 |
| --- | --- |
| Total length (bp) | 4550059 |
| Scaffolds | 19 |
| Contigs | 135 |
| N50 | 620817 |
| GC (%) | 47 |
| Genes (total) | 4138 |
| CDS (total) | 4034 |
| Genes (coding) | 3857 |
| CDS (coding) | 3857 |
| Genes (RNA) | 104 |
| rRNAs | 8, 7, 7 (5S, 16S, 23S) |
| tRNAs | 73 |
| ncRNAs | 9 |
| Pseudo Genes (total) | 177 |

We detected virulence determinants of motility controlled by *flhE*, *fliN fliL*, *flgA*, *fliD*, *fliC*, *fliP*, *flgD*, *flhA*, *fliS*, *flhB*, *flgK*, *fliF*, *flgB*, *flgJ*, *flgC*, *flgL*, *flgN*, *flgM*, *flhC*, *fliE*, *fliO*, *fliH*, *flgE*, *fliB*, *fliA*, *fliK*, *fliI*, *flgF*, *flgI*, *fliG*, *fliR*, *flgH*, *flhD*, *flgG*, *fliT*, *fliM*, *fliZ*, and *fliQ*. These genes act in other processes, such as cellular invasion, biofilm formation, and the secretion system (the type III secretion/flagella biosynthesis apparatus).[13,14] We found the *cheA*, *cheB*, *cheD*, *cheR*, *cheW*, *cheY*, *cheZ* genes, which act in the chemotaxis mechanism, by which bacteria efficiently and rapidly respond to changes in the chemical composition of their environment. This behavior modulates the direction of flagellar rotation.[15] Finally, we found the *motA* and *motB* genes, which encode flagellar motor protein MotA and MotB, respectively, and build the flagellar-motor supramolecular complex stator.[15]

The *yst1L*, *yst1J*, *yst1K*, *yst1F*, *yst1E*, *yst1G*, *yst1C* and YE105_C0912 genes encode proteins that build the type II secretion system (T2SS). All *Yersinia* species, non-pathogenic as well as pathogenic, possess at least one of the two T2SS. The one T2SS that has so far been studied is termed Yts1, which is involved in dissemination and colonization of deeper tissues like liver and spleen in mouse infection experiments. Furthermore, the *Yts1O* gene encodes the Yts1O protein, which is a putative prepilin-like peptidase.[16]

YE105_C0309, YE105_C0310, YE105_C0312, YE105_C0313, YE105_C0314, YE105_C0323, YE105_C0324, YE105_C0325, YE105_C0326, YE105_C0327, YE105_C0328, YE105_C0332, YE105_C0333, YE105_C0334, YE105_C0335, YE105_C0336, YE105_C0338, YE105_C0339, YE105_C0340, YE105_C0341 are a set of genes that received the name of the strain from where they were recognized for the first time. These genes encode proteins that build the Ysc type III secretion system (pYV-encoded T3SS/TTSS). The T3SS is the mechanism by which pathogenic *Yersinia* injects proteins directly into the eukaryotic cell cytosol.[3] Y. *enterocolitica* has three different type III secretion systems, Ysc (pYV-encoded), Ysa (exclusive to biovar 1B), and the flagellar.[2,3,13] Surprisingly, the genes found on YE 19 genome were related to the proteins (YscC, R, and U) that built the Ysc T3SS[17] (Supplementary file 1). These findings reinforce the hypothesis that this strain already had the pYV and, somehow, these genes could be integrated into the chromosome. Meanwhile, further studies are needed to confirm this possibility.
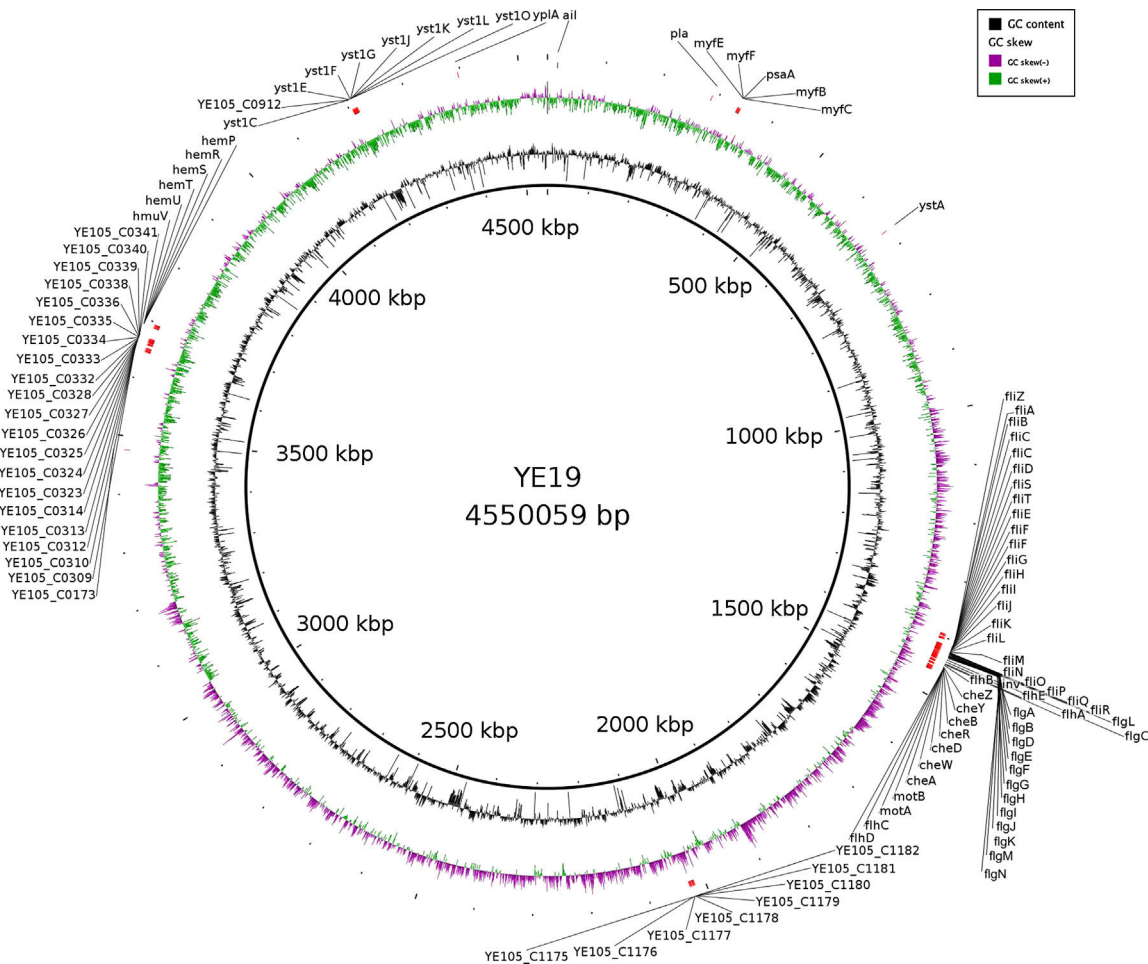
**Fig. 1 – YE 19 strain virulome. Circular map of the YE 19 strain chromosome is playing the distribution of the identified virulence genes.**

We found the *Yersinia* stable toxin *ystA* gene, which encodes the production of a heat-stable enterotoxin that causes diarrhea.[2]

The presence of YE105_C0173, YE105_C1175, YE105_C1176, YE105_C1177, YE105_C1178, YE105_C1179, YE105_C1180, YE105_C1181, YE105_C1182 genes related to the O antigen were found. The O antigen or O polysaccharide chain is associated with antigenic properties and is involved in invasion and colonization processes. In addition, the O antigen is required for proper functionality or expression of other outer membrane virulence factors.[2]

We found the *hemP*, *hemR*, *hemS*, *hemT*, *hemU* and *hmuV* genes associated with hemin uptake. This set of genes permits the strain to use the heme-containing compounds as a source of iron.[18] However, these genes are not the same genes found on yersiniabactin.[2]

The common virulence genes *inv* (invasive gene) and *ail* (attachment invasion locus) were present. They encode an invasin, which directly initiates cell penetration, and an outer membrane protein called Ail, which encodes a factor that also promotes the invasion of epithelial cells, respectively.[1] The *psaA* (*myfA*) gene is related to adhesion to the host cell

and its pH 6 antigen homolog in Y. *pseudotuberculosis* and Y. *pestis*.[19] The *myfB*, *myfC*, *myfE* and *myfF* genes are encoded by the *myf* operon, known as the mucoid *Yersinia* factor. These genes encode proteins that constitute a fibrillar structure that acts during adhesion.[2] The Y*pla* gene encodes phospholipase A that stimulates the acute inflammatory response of the host and is required for survival of Y. *enterocolitica* in Peyer's patches.[3] Finally, the *pla* gene, which has been described in the Y. *enterocolitica* W22703 strain,[20] encodes a peptidase A26 omptin protein that has proteolytic and adhesion functions.[19] A Blast search on Genbank for this *pla* gene showed only 30% nucleotide identity with the *pla* gene from Y. *pestis* (strain I-3190 plasmid pPst); thus, further studies are needed to determine the relationship between these two genes.

These results corroborate the findings of Batzilla et al. and suggest that this strain harbors a set of virulence determinants that may play role in host pathoadaptation. This could also justify the successful dissemination of bioserotype 4/O:3 in Brazil and worldwide. As the pYV virulence plasmid[1] was not found, we cannot rule out the possibility that it might have been lost during storage. The presence of strains harboring all of these virulence genes in Brazil is a potential

threat to young children and immunocompromised individuals, for whom yersiniosis can represent a significant source of morbidity and mortality.[3] We hope this genomic data analysis helps understanding the virulence issue of Brazilian strains and contributes for *Y. enterocolitica* studies worldwide.

## Conflicts of interest

The authors declare no conflicts of interest.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bjid.2017.04.005.

## REFERENCES

1. Sabina Y, Rahman A, Ray RC, Montet D. *Yersinia enterocolitica*: mode of transmission, molecular insights of virulence, and pathogenesis of infection. J Pathog. 2011;2011:1–10.
2. Fàbrega A, Vila J. *Yersinia enterocolitica*: pathogenesis, virulence and antimicrobial resistance. Enferm Infecc Microbiol Clin. 2012;30:24–32.
3. Galindo CL, Rosenzweig JA, Kirtley ML, Chopra AK. Pathogenesis of Y. *enterocolitica* and Y. *pseudotuberculosis* in Human Yersiniosis. J Pathog. 2011;2011:1–16 [Internet]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3335670&tool=pmcentrez&rendertype=Abstract
4. Batzilla J, Antonenka U, Höper D, Heesemann J, Rakin A. *Yersinia enterocolitica palearctica* serobiotype O:3/4 – a successful group of emerging zoonotic pathogens. BMC Genomics. 2011;12:348 [Internet]. Available from: http://www.biomedcentral.com/1471-2164/12/348
5. Falcão JP, Falcão DP. Importance of *Yersinia enterocolitica* in medical microbiology. Rev Cienc Farm Basica Apl. 2006;27:9–19 [in Portuguese].
6. Rusak LA, Moura Falavina Dos Reis C, Barbosa AV, et al. Phenotypic and genotypic analysis of bio-serotypes of *Yersinia enterocolitica* from various sources in Brazil. J Infect Dev Ctries. 2014;8:1533–40.
7. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957–63.
8. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
9. Kolmogorov M, Raney B, Paten B, Pham S. Ragout – a reference-assisted assembly tool for bacterial genomes. Bioinformatics. 2014;30:302–9.
10. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinform. 2015;16:230 [Internet]. Available from: http://www.biomedcentral.com/1471-2105/16/230
11. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE. 2014;9.
12. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. REAPR: a universal tool for genome assembly evaluation. Genome Biol. 2013;14:R47 [Internet]. Available from: http://genomebiology.com/2013/14/5/R47
13. Kim TJ, Young BM, Young GM. Effect of flagellar mutations on *Yersinia enterocolitica* biofilm formation. Appl Environ Microbiol. 2008;74:5466–74.
14. Kapatral V, Campbell JW, Minnich SA, Thomson NR, Matsumura P, Prüß BM. Gene arrays analysis of *Yersinia enterocolitica* FlhD and FlhC: regulation of enzymes affecting synthesis and degradation of carbamoylphosphate. Microbiology. 2004;150:2289–300.
15. Bren A, Eisenbach M. How signals are heard during bacterial chemotaxis: protein–protein interactions in sensory signal propagation. J Bacteriol. 2000;182:6865–73 [minireview].
16. von Tils D, Blädel I, Schmidt MA, Heusipp G. Type II secretion in *Yersinia* – a secretion system for pathogenicity and environmental fitness. Front Cell Infect Microbiol. 2012;2:160 [Internet]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3521999&tool=pmcentrez&rendertype=Abstract
17. Cornelis GR. The *Yersinia* Ysc–Yop "Type III" weaponry. Nat Rev Mol Cell Biol. 2002;3:742–54 [Internet]. Available from: http://www.nature.com/doifinder/10.1038/nrm932
18. Stojiljkovic I, Hantke K. Hemin uptake system of *Yersinia enterocolitica*: similarities with other TonB-dependent systems in gram-negative bacteria. EMBO J. 1992;11:4359–67.
19. Mikula KM, Kolodziejczyk R, Goldman A. Yersinia infection tools – characterization of structure and function of adhesins. Front Cell Infect Microbiol. 2012;2:169 [Internet]. Available from: /pmc/articles/PMC3539135/?report=Abstract
20. Fuchs TM, Brandt K, Starke M, Rattei T. Shotgun sequencing of *Yersinia enterocolitica* strain W22703 (biotype 2, serotype O:9): genomic evidence for oscillation between invertebrates and mammals. BMC Genomics. 2011;12:168 [Internet]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3079665&tool=pmcentrez&rendertype=Abstract

## WEB REFERENCES

*Yersinia* MLST website (available from: http://pubmlst.org/yersinia/) developed by Keith Jolley and sited at the University of Oxford (Jolley & Maiden 2010, BMC Bioinformatics, 11:595) [accessed in 08.29.16].
The Virulence Factor database (VFDB) (available from: http://www.mgc.ac.cn/VFs/) [accessed in 08.29.16].