AMERICAN SOCIETY FOR MICROBIOLOGY | **Microbiology**® Resource Announcements

Draft Genome Sequence of *Escherichia marmotae* E690, Isolated from Beef Cattle

Medelin Ocejo,ª Maitane Tello,ª Beatriz Oporto,ª José Luis Lavín,ᵇ Ana Hurtadoª

ªAnimal Health Department, NEIKER–Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Derio, Bizkaia, Spain
ᵇApplied Mathematics Department, Bioinformatics Unit, NEIKER–Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Derio, Bizkaia, Spain

ABSTRACT   We report here the draft genome sequence of an extended-spectrum β-lactamase (ESBL)-producing *Escherichia* species isolated from rectal feces collected from beef cattle in northern Spain. Analysis of the draft genome identified the strain as a member of the newly described species *Escherichia marmotae*.

The genus *Escherichia* (family *Enterobacteriaceae*) includes four species, *Escherichia coli* (1), *Escherichia fergusonii* (2), *Escherichia albertii* (3), and the recently described *Escherichia marmotae* (4, 5), along with several cryptic clades (6).

We report here the draft genome sequence of *E. marmotae* strain E690, recovered in 2016 from rectal feces collected from beef cattle in the Basque Country (northern Spain) as part of a cross-sectional survey conducted to estimate the herd prevalence of extended-spectrum β-lactamase/AmpC-producing *E. coli* in ruminants (7). Feces (25 g) diluted 1:10 in modified tryptic soy broth (bioMérieux) supplemented with novobiocin (Biolife) was incubated at $41 \pm 1°C$ (6 to 7 h), preenriched in MacConkey broth with 1 mg/liter cefotaxime ($37 \pm 1°C$, 24 h), and subcultured onto MacConkey agar with cefotaxime (1 mg/liter).

DNA extracted from pure culture (Wizard genomic DNA purification kit, Promega) was submitted to Eurofins Genomics, where libraries were prepared using the NEBNext Ultra II FS DNA library prep kit (Illumina). The genome was sequenced using an Illumina NovaSeq 6000 instrument (150-bp paired-end reads), resulting in 23,934,128 reads ($718\times$ coverage). Quality control was assessed using FastQC v.0.11.9 (8) and then analyzed via TORMES v.1.0 (9). Briefly, the reads were quality filtered using Trimmomatic (10) and PRINTSEQ (11) and *de novo* assembled using SPAdes (12). The quality of the assemblies was assessed with QUAST (13), discarding contigs below 200 bp with PRINSEQ. BLASTn v.2.9.1+ (14) and ABRicate were used to screen for acquired antimicrobial resistance genes in ResFinder (15), chromosomal point mutations associated with antimicrobial resistance in PointFinder (16), and virulence genes in the Virulence Factors Database (VFDB) (17). Plasmid replicons were identified using PlasmidFinder (18). PlasFlow (19) predicted plasmid- and chromosome-derived contigs. *In silico* FimH typing was achieved using FimTyper (20). Multilocus sequence types (MLSTs) (Achtman scheme) were queried against the *E. coli* MLST database PubMLST (21) using mlst, and the core-genome MLST (cgST) was assigned using cgMLSTFinder, following the EnteroBase *E. coli* cgMLST scheme (22). Gene identification and annotation were retrieved from the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (23). See Table 1 for versions/commands used for specific tools.

The chromosome sequence length was 4,303,797 bp (63 contigs; N_{50}, 175,886 bases), with a 50.4% G+C content. The chromosome contains 4,094 genes (3,951 protein-coding, 74 pseudogenes, and 69 RNA genes). Moreover, two plasmid incompatibility group-determining sequences were identified, IncFI and IncI1. Pairwise com-

TABLE 1 Bioinformatics tools used in the study along with the versions, scripts, and references

Tool	Version	Script(s) run on command line	Reference or source
Trimmomatic	0.38	trimmomatic PE -phred33 Raw_reads/Sample_R1.fastq.gz Raw_reads/Sample_R2.fastq.gz /cleaned_reads/Sample_noadapt.R1.fastq.gz /dev/null cleaned_reads/Sample_noadapt.R2.fastq.gz /dev/null ILLUMINACLIP: adapters.fa:1:30:1	10
PRINSEQ	0.20.4	raw data quality filtering: perl prinseq-lite.pl -verbose -fastq cleaned_reads/Sample_noadapt.R1.fastq -fastq2 cleaned_reads/Sample_noadapt.R2.fastq -out_good cleaned_reads/Sample_ok -out_format 3 -out_bad null -min_len 125 -min_qual_mean 25 -trim_qual_right 25 -trim_qual_window 15 -trim_qual_type mean contigs filtering: perl prinseq-lite.pl -fasta assembly/Sample_assembly/contigs.fasta -min_len 200 -out_good genomes/Sample -out_bad null	11
SPAdes	3.13.0	python spades.py --careful -1 cleaned_reads/Sample_ok_1.fastq.gz -2 cleaned_reads/Sample_ok_2.fastq.gz -o assembly/Sample_assembly -t 8	12
QUAST	5.0.2	quast genomes/Sample.fasta -o /genome_stats/Sample_genome_stats -t 16 --min-contig 200 --no-icarus --silent --no-sv	13
ABRicate	0.8.10	abricate genomes/Sample.fasta --db db* –nopath (in the curation of the results table, all hits with coverage below 60% and identity below 90% were removed) *db = Resfinder, PlasmidFinder, VFDB	https://github.com/tseemann/abricate
PointFinder	3.1.0	python PointFinder.py -i genomes/Sample.fasta -p pointfinder_db -m blastn -m_p blastn -s ecoli -o point_mutations/Sample	16
PlasFlow	1.1	PlasFlow.py --input genomes/Sample.fasta --output PlasFlow/Sample.predictions --threshold 0.7	19
FimTyper	1.1	perl fimtyper.pl -d fimtyper_db -b ncbi-blast-2.9.0+/ -i genomes/Sample.fasta -o /fimH_typing/Sample -k 95.00 -l 0.80	20
mlst	2.16.1	mlst genomes/Sample.fasta --nopath --quiet > mlst/mlst.tab	https://github.com/tseemann/mlst
cgMLSTFinder	1.1	docker run --rm -it -v $(workdir):/workdir cgmlstfinder -o output -s ecoli -db cgmlstfinder_db -t temp cleaned_reads/Sample.fastq.gz	https://bitbucket.org/genomicepidemiology/cgmlstfinder/src/master/
PGAP	4.11 2020-03-30.build4489	./pgap.py -r -o Sample_results genomes/Sample/input.yaml	23

parisons of the E690 genome versus closely related strain genomes performed at the Type (Strain) Genome Server (TYGS) (24) identified the genome of *E. marmotae* HT073016^T as the closest match. Intergenomic comparison of E690 and *E. marmotae* HT073016^T by digital DNA-DNA hybridization (d_4 = 94.6 [93.0 to 95.9]) and G+C content (0.18 difference) identified strain E690 as *E. marmotae*.

MICs were determined by broth microdilution following the recommendations in Commission Decision 2013/652/EU using two Sensititre MIC susceptibility plates (EUVSEC1 and EUVSEC2, Thermo Fisher Scientific). *E. marmotae* E690 exhibits microbiological resistance to tetracycline (MIC > 64 mg/liter) and the β-lactams ampicillin (MIC > 64 mg/liter), cefotaxime (MIC = 8 mg/liter), ceftazidime (MIC = 16 mg/liter), and cefepime (MIC = 1 mg/liter) and carries the *tet*(A) and *bla*$_{SHV-12}$ genes in an IncI1 plasmid. It carries mutations in the topoisomerase genes, *parC* (p.S57T) and *parE* (p.I355T) but is susceptible to nalidixic acid (MIC < 4 mg/liter) and ciprofloxacin (MIC = 0.03 mg/liter). It belongs to MLST ST-6495 and cgST-141216 and carries a *fim*H160 allele. *E. marmotae* E690 harbors virulence factors related to extraintestinal pathogenic *E. coli* (ExPEC), like F1 fimbriae, the K1 capsule, and the protein OmpA (chromosomally encoded), and to animal enterotoxigenic *E. coli* (ETEC), like F4 fimbriae and a heat-stable enterotoxin EAST1 (plasmid encoded).

Data availability. The genome sequence was deposited under GenBank accession number JABXGM000000000, BioProject accession number PRJNA632731, and BioSample accession number SAMN14918579.

Data curation, formal analysis, writing – review & editing; A.H.: conceptualization, supervision, writing – original draft.

REFERENCES

1. Castellani A, Chalmers AJ. 1919. Manual of tropical medicine, 3rd ed. Baillière, Tindall and Cox, London, United Kingdom.
2. Farmer JJ, III, Fanning GR, Davis BR, O'Hara CM, Riddle C, Hickman-Brenner FW, Asbury MA, Lowery VA, III, Brenner DJ. 1985. *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of Enterobacteriaceae isolated from clinical specimens. J Clin Microbiol 21:77–81. https://doi.org/10.1128/JCM.21.1.77-81.1985.
3. Huys G, Cnockaert M, Janda JM, Swings J. 2003. *Escherichia albertii* sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. Int J Syst Evol Microbiol 53:807–810. https://doi.org/10.1099/ijs.0.02475-0.
4. Liu S, Jin D, Lan R, Wang Y, Meng Q, Dai H, Lu S, Hu S, Xu J. 2015. *Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. Int J Syst Evol Microbiol 65:2130–2134. https://doi.org/10.1099/ijs.0.000228.
5. Liu S, Feng J, Pu J, Xu X, Lu S, Yang J, Wang Y, Jin D, Du X, Meng X, Luo X, Sun H, Xiong Y, Ye C, Lan R, Xu J. 2019. Genomic and molecular characterisation of *Escherichia marmotae* from wild rodents in Qinghai-Tibet plateau as a potential pathogen. Sci Rep 9:10619. https://doi.org/10.1038/s41598-019-46831-3.
6. Clermont O, Gordon DM, Brisse S, Walk ST, Denamur E. 2011. Characterization of the cryptic *Escherichia* lineages: rapid identification and prevalence. Environ Microbiol 13:2468–2477. https://doi.org/10.1111/j.1462-2920.2011.02519.x.
7. Tello M, Ocejo M, Oporto B, Hurtado A. 2020. Prevalence of cefotaxime-resistant Escherichia coli isolated from healthy cattle and sheep in Northern Spain: phenotypic and genome-based characterization of antimicrobial susceptibility. Appl Environ Microbiol 86:e00742-20. https://doi.org/10.1128/AEM.00742-20.
8. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
9. Quijada NM, Rodríguez-Lázaro D, Eiros JM, Hernández M. 2019. TORMES: an automated pipeline for whole bacterial genome analysis. Bioinformatics 35:4207–4212. https://doi.org/10.1093/bioinformatics/btz220.
10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
11. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
14. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214. https://doi.org/10.1089/10665270050081478.
15. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
16. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. 2017. PointFinder: a novel Web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother 72:2764–2768. https://doi.org/10.1093/jac/dkx217.
17. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328. https://doi.org/10.1093/nar/gki008.
18. Carattoli A, 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. https://doi.org/10.1007/978-1-4939-9877-7_20.
19. Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res 46:e35. https://doi.org/10.1093/nar/gkx1321.
20. Roer L, Tchesnokova V, Allesøe R, Muradova M, Chattopadhyay S, Ahrenfeldt J, Thomsen MCF, Lund O, Hansen F, Hammerum AM, Sokurenko E, Hasman H. 2017. Development of a Web tool for *Escherichia coli* subtyping based on *fimH* alleles. J Clin Microbiol 55:2538–2543. https://doi.org/10.1128/JCM.00737-17.
21. Jolley KA, Maiden MCJ. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
22. Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Agama Study Group, Achtman M. 2020. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. Genome Res 30:138–152. https://doi.org/10.1101/gr.251678.119.
23. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics 28:416–418. https://doi.org/10.1093/bioinformatics/btr655.
24. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3.