DATA NOTE

Draft genome sequence of *Enterococcus faecium* SP15, a potential probiotic strain isolated from spring water

Fauzia Aziz[1,2,3]* , Muhammad Nasim Khan[1], Safia Ahmed[4] and Simon Colin Andrews[2]

Abstract

Objectives: Enterococci are Gram-positive lactic acid bacteria and common inhabitants of the gastrointestinal tract of mammals, including humans. They are also widely distributed in diverse environments such as soil, water, vegetables and food. *Enterococcus faecium* is able to produce antimicrobial compounds (enterocins) and thus can act as a probiotic. *E. faecium* SP15 is a newly identified enterocin-producing strain from spring water that has been subjected to genome sequence analysis to provide understanding of its antimicrobial and probiotic properties.

Data description: The draft genome of *E. faecium* SP15 comprises of 2,783,033 bp with a G+C content of 38.08%. Five genetic loci predicted to specify enterocin production were identified, but no virulence factors could be detected and only two potential antibiotic resistance genes were noted.

Keywords: *Enterococcus faecium*, Probiotic, Draft genome assembly, Enterocin, Spring water

Objective

Enterococci are Gram-positive lactic acid bacteria with a wide environmental distribution, encompassing many species from a wide variety of ecological niches [1]. *Enterococcus faecium* is a major nosocomial pathogen often causing neonatal meningitis or endocarditis [2]. However, certain strains of *E. faecium* have beneficial effects on human health due to their probiotic activity [3]. For example, *E. faecium* T-110 is a syndicate member in several probiotic products including BIO-THREE[R] which is widely prescribed for human, animal and aquacultural use [3] and *E. faecium* strain L-3 is the principle organism in the probiotic Laminolact [4]. *E. faecium* is well known for its ability to produce bacteriocins, but there are relatively few reports in the literature on the genome sequence of *E. faecium* from non-clinical sources [5, 6] and there are no current NCBI database genome entries for *E. faecium* isolated from natural water sources.

Data description

Enterococcus faecium SP15 was isolated from spring water in Rawalakot (Azad Kashmir, Pakistan), a site of relative isolation with little waste water contamination. The strain exhibited strong antimicrobial activity against a panel of seven indicator strains, including *Listeria monocytogenesis*, indicative of enterocin production. *E. faecium* SP15 genomic DNA was extracted using a GeneJET genomic DNA purification kit (Thermofisher Scientific) as recommended by the vendor and assessed using a NanoDrop ND-1000 spectrophotometer and 0.7% agarose gel electrophoresis. Genome sequencing was performed by MicrobesNG (University of Birmingham, UK) using Illumina MiSeq and HiSeq 2500 platforms (Illumina, UK) with 2×250 bp paired-end reads. The reads were trimmed using Trimmomatic version 0.30 [7] and the quality was assessed using in-house scripts combined with BWA-MEM software 0.7.16 [8]. De novo assembly was performed with SPAdes software version 3.9.0 [9] and assembly metrics were calculated using QUAST version 2.0 [10]. Gene prediction and annotation were carried out using the Pathosystems Resources Integration Center (PATRIC) web server [11], RAST version 2.0 [12] and the NCBI, PGAP version 4.6 [13]. The

*Correspondence: f.aziz@reading.ac.uk
[2] School of Biological Sciences, University of Reading, Reading RG6 6AJ, UK
Full list of author information is available at the end of the article

Table 1 Overview of data files

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file	Whole genome shotgun project	FASTA	DDBJ/ENA/GenBank (accession RDQA00000000)

tRNA genes were predicted by tRNA scan-SE 2.0 [14]. Antimicrobial mechanisms (e.g. enterocin production) were explored with BAGEL 3 [15] and anti-SMASH V4 [16]. Virulence factors were identified using the virulence factor database, VFDB [17]. Antibiotic resistance gene were identified using the Comprehensive Antibiotic Resistance Database, CARD [18], and acquired resistance genes were predicted by Resistance Finder 3.0 [19]. Contigs were ordered by alignment against the most closely related sequence in GenBank [20] (*E. faecium* T110, CP006030; 99% identity) using progressive Mauve version 2.4.0 [21]. Intact and incomplete prophage regions were identified through the integrated search and annotation tool, PHAST [22]. Clustered regularly interspaced short palindromic repeat (CRISPR) arrays were identified using CRISPR finder [23].

The draft genome assembly consisted of 121 contigs with a total size of 2,783,033 bp (Table 1). The genome sequence data was at 30× coverage with an N50 of 102,590 bp and mean GC content 38.08%. A total of 2900 protein-encoding genes were predicted of which 2063 were assigned putative functions while 837 remain hypothetical. A total of 63 tRNA structural genes were identified. BAGLE 3 predicted five bacteriocin biosynthetic gene clusters (enterocin-HF, enterocin-P, enterocin SE-K4, enterocin L50A/L50B and enterolysin). No virulence factors (AS, Ace, Acm, Scm, EfaA, EcbA, Esp, Cyl, GelE and SprE) were detected. Two antibiotic resistance genes *aac* (6′)-li and *msrC* (98 and 97% identity, respectively) were identified conferring resistance to aminoglycosides, and macrolides and streptogramin B antibiotics, but no acquired resistance determinants were found. Four prophage loci were predicted of which three were intact (Strept_9871, Lactob_phig1e and Staphy_SPbeta (40, 36.9 and 32.7 kb, respectively) and one incomplete (Salmon_SJ46, 17.4 kb). One CRISPR array was identified of 190 bp, containing three spacers with a highly conserved 24 bp DR region, and although two *cas* gene clusters (*cas*3_typeI, *cas*4_typeI-II) were found, these were not associated with the CRISPR array.

Two major replicons are apparent: a chromosome of ~ 2,545,000 bp and a plasmid of ~ 149,300 bp (related to plasmid pNB2354 from *E. faecium* NRRL-B-2354, CP004064).

Limitations

Current data is based on the draft level genome such that the exact length of the genome, and the number of rRNA genes and repetitive elements, cannot be absolutely determined. Furthermore, the genome includes extrachromosomal elements that cannot be predicted precisely.

Abbreviations
PGAP: prokaryotic genome annotation pipeline; BAGEL 3: BActeriocin GEnome mining tooL version 3; antiSMASH: antibiotic and secondary Metabolite Analysis SHell.

Authors' contributions
FA obtained the samples and completed the sequence analysis. MNK and SA advised on the study design and SCA advised and assisted the experimental work. All authors read and approved the final manuscript.

Author details
[1] Department of Zoology, The University of Azad Jammu and Kashmir, Muzaffarabad, Azad Kashmir, Pakistan. [2] School of Biological Sciences, University of Reading, Reading RG6 6AJ, UK. [3] Department of Zoology, The University of Poonch, Rawalakot, Azad Kashmir, Pakistan. [4] Department of Microbiology, Quaid-I-Azam University, Islamabad 45230, Pakistan.

Competing interests
The authors declare that they have no competing interests.

Availability of data materials
This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number RDQA00000000. The version described in this paper is version RDQA01000000. The data described in this Data Note can be freely and openly accessed at DDBJ/ENA/GenBank. Please see Table 1 for details and links to the data.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Aziz *et al. BMC Res Notes*   (2019) 12:99

Page 3 of 3

References

1. Franz CM, Holzapfel WH, Stiles ME. Enterococci at the crossroads of food safety. Int J Food Microbiol. 1999;47:1–24.
2. Linden SM, Miller CB. Vancomycin-resistant enterococci: the clinical effect of a common nosocomial pathogen. Diagn Microbiol Infect Dis. 1999;33:113–20.
3. Purushothaman M, Parani M. First complete genome sequence of a probiotic *Enterococcus faecium* strain T-110 and its comparative genome analysis with pathogenic and non-pathogenic *Enterococcus faecium* genome. J Genet Genomics. 2015;42:43–6. https://doi.org/10.1016/j.jgg.2014.07.002.
4. Karaseva A, Tsapieva A, Pachebat J, Suvorov A. Draft genome sequence of probiotic *Enterococcus faecium* strain L-3. Genome Announ. 2016;4:e01622.
5. Qin X, Galloway-Peña JR, Sillanpaa J, Roh JH, Nallapareddy SR, Chowdhury S, Bourgogne A, Choudhury T, Muzny DM, Buhay CJ, Ding Y, Dugan-Rocha S, Liu W, Kovar C, Sodergren E, Highlander S, Petrosino JF, Worley KC, Gibbs RA, Weinstock GM, Murray BE. Complete genome sequence of *Enterococcus faecium* strain TX16 and comparative genomic analysis of *Enterococcus faecium* genomes. BMC Microbiol. 2012;12:135.
6. Van Hal SJ, Ip CLC, Ansari MA, Wilson DJ, Espedido BA, Jensen SO, Bowden R. Evolutionary dynamics of *Enterococcus faecium* reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. Microbial Genomics. 2016. https://doi.org/10.1099/mgen.0.000048.
7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–200.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler Transform. Bioinformatics. 2009;25:1754–60.
9. 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. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
10. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
11. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. 2017;45:D535–D542542.
12. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614–24. https://doi.org/10.1093/nar/gkw569.
14. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
15. van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. BAGEL3: automated identification of genes encoding bacteriocins and (non-bactericidal post translationally modified peptides. Nucleic Acids Res. 2013;41:W448–W45353. https://doi.org/10.1093/nar/gkt391.
16. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite bio-synthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39:W339–W34646. https://doi.org/10.1093/nar/gkr466.
17. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33:D325–D328328.
18. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57(7):3348–57. https://doi.org/10.1128/AAC.00419-13.
19. Zankari E, Hasman H, Consentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012. https://doi.org/10.1093/jac/dks261.
20. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018;46(D1):D8–D13. https://doi.org/10.1093/nar/gkx1095.
21. Darling AE, Mau B, Perna NT. Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010;5(6):e11147. https://doi.org/10.1371/journal.pone.0011147.
22. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:W347–W352352.
23. Grissa I, Vergnaud G, Pource LC. CRISPR Finder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52–W5757. https://doi.org/10.1093/nar/gkm360.