Data in Brief 48 (2023) 109151

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome sequence data of *Enterococcus faecium* R9, a multiple enterocins-producing strain

Abrar Akbar, Sabah Al-Momin*, Mohamed Kishk, Abdulaziz Al-Ateeqi, Anisha Shajan, Rita Rahmeh

Environment and Life Sciences Research Centre, Kuwait Institute for Scientific Research, Kuwait

ARTICLE INFO

Article history:
Received 28 December 2022
Revised 2 April 2023
Accepted 6 April 2023
Available online 14 April 2023

Dataset link: Enterococcus faecium strain R9, whole genome shotgun sequencing project (Original data)

Keywords:
Enterocins
Enterococcus faecium
Illumina HiSeq
Food contamination

ABSTRACT

Food contamination by pathogens results in serious health problems and economic losses. Chemical food preservatives pose a risk to human health when used in food preservation. To increase the shelf life of the products and prevent spoilage, the dairy sector is considering natural preservatives such the ribosomally synthesized peptides, bacteriocins. Here we present the draft genome sequence of *Enterococcus faecium* strain R9 producing three bacteriocins isolated from raw camel milk. These bacteriocins showed valuable technological properties, such as sensitivity to proteolytic enzymes, heat stability, and wide range of pH tolerance. The 2 × 250 bp paired end reads sequencing was performed on Illumina HiSeq 2500 sequencing. The genome sequence consisted of 3,598,862 bases, with a GC content of 37.94% bases. The number of raw reads was 4,670,510, and the assembly N50 score was 65,355 bp with a 310.28 average coverage. A total of 3,086 coding sequences (CDSs) was predicted with 2,126 CDSs with a known function and 127 with a signal peptide. Annotation of the genome sequence revealed bacteriocins encoding genes, namely, enterocin B, enterocin P, and two-component enterocin X (X-alfa and X-beta subunits). These enterocins are beneficial for controlling *Listeria monocytogenes* in the food industry. Genome sequence of *Ente-*

* Corresponding author.
E-mail address: smomen@kisr.edu.kw (S. Al-Momin).
Social media: @meyadw (M. Kishk)

https://doi.org/10.1016/j.dib.2023.109151

rococcus faecium R9 has been deposited at the gene bank under BioSample accession number JALJED000000000 and are available in Mendeley Data [1].

Specifications Table

Subject:	*Biology*
Specific subject area:	*Genomics, applied microbiology*
Type of data:	*Raw sequencing reads, tables, figures, Assembled Genome, Annotated Genome*
How the data were acquired:	*Whole genome sequencing using Illumina HiSeq 2500 platform*
Data format:	*Raw data and analysed reads*
Description of data collection:	*Enterococcus faecium strain R9 was isolated from raw camel milk collected from Kuwait. DNA was isolated using GenElute Bacterial Genomic DNA Kit and sequenced on Illumina HiSeq 2500 platform. De novo assembly of the reads, including scaffolding and gap-filling were performed. genome annotation of predicted genes was achieved.*
Data source location:	• *Institution: Kuwait Institute for Scientific Research* • *City/Town/Region: Kuwait* • *Country: Kuwait* • *Latitude and longitude (29.25 N 47.91 E), Kuwait.*
Data accessibility:	*Repository name: Mendeley Data [1].* *Data identification number: DOI: 10.17632/yvnkzzb7yk.1* *Direct URL to data:* https://data.mendeley.com/datasets/yvnkzzb7yk *Repository name: Figshare* *Data identification number: 10.6084/m9.figshare.21786662* *Direct URL to data:* https://figshare.com/articles/dataset/R9_contig-sequences-gapclosed_fa/21786662 *Repository name: National center for Biotechnology Information (NCBI)* *Data identification number: Accession: JALJED000000000* *Direct URL to data:* https://www.ncbi.nlm.nih.gov/nuccore/JALJED000000000

Value of the Data

- The data provides the genome sequence of *Enterococcus faecium* R9, a strain producing three enterocins with antilisterial activity.
- The data produced can have a significant impact in the food industry's ability to control *Listeria monocytogenes*.
- The sequences can be utilized in future studies aiming for enterocins overexpression.
- The data is publically available for the scientist for comparison with the sequence of other *Enterococcus faecium* strains isolated from food matrices.

1. Objective

This data article describes the draft genome sequence of *Enterococcus faecium* strain R9 producing three bacteriocins isolated from raw camel milk. These bacteriocins display antilisterial activity [2]. The data reveal the number of raw reads, the coding sequences and the bacteriocins encoding genes.

2. Data Description

Genome sequence of *Enterococcus faecium* R9 is available at https://www.ncbi.nlm.nih.gov/nuccore/JALJED000000000.

Enterococcus faecium strain R9 was isolated from raw camel milk in Kabad (northwest region, Kuwait). The strain exhibited strong antimicrobial activity against *Listeria monocytogenes* [2]. This paper reports the genome sequence of *Enterococcus faecium* strain R9. The sequence was annotated and analysed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The sequences consisted of 3597,440 bases, with a GC content of 37.94% bases. The number of raw reads was 4670,510, and the assembly N50 score was 65,355 bp with a 310.28 average coverage. A total of 4.412 coding sequences (CDSs) was predicted with 2126 CDSs with a known function and 127 with a signal peptide (Table 1). A circular map of the genome sequence of *Enterococcus faecium* strain R9 was constructed using GCView software (Fig. 1).

During the metabolic process, *Enterococcus* secretes extracellular proteins. Table 2 summarizes the identified extracellular proteins with known function. The annotated sequences showed 16 extracellular proteins that were annotated using Prokka Prokaryotic Genome Annotation System.

Several CDSs for the production of bacteriocins, namely, enterocin B, enterocin P, and two-component enterocin X (X-alfa and X-beta subunits) were found using BAGEL4 analysis software.

Table 1
Basic statistics of *Enterococcus faecium* strain R9.

Species Attribute	Value
Total sequence length	3597,440 bp
GC content	37.94%
The number of raw reads	4670,510
Number of scaffolds	1655
Scaffold N50	65,355
Scaffold L50	13
Number of contigs	1658
CDSs (total)	4412
Total Genes	4489
Coding Genes	4304

Table 2
Known Extracellular proteins with function, identified in the sequence of *Enterococcus faecium* strain R9.

locus_tag	gene	length_aa_seq	Protein name
Prokka_01359	lytG_1	206	Exo-glucosaminidase LytG
Prokka_01538	usp45	513	Secreted 45 kDa protein
Prokka_01570	Araf43A	320	Extracellular exo-alpha-(1->5)-L-arabinofuranosidase
Prokka_01726	oppA	594	Oligopeptide-binding protein OppA
Prokka_02011	yddH	334	Probable endopeptidase YddH
Prokka_00238	Ent B	72	Bacteriocins enterocin-B
Prokka_00241	EntX alpha	54	Bacteriocin enterocin X-alpha
Prokka_00429	sodA	203	Superoxide dismutase [Fe]
Prokka_02739	lytG_2	215	Exo-glucosaminidase LytG
Prokka_00677	SAOUHSC_02979	715	N-acetylmuramoyl-L-alanine amidase domain-containing protein SAOUHSC_02979
Prokka_02854	yycI_2	204	Two-component system WalR/WalK regulatory protein YycI
Prokka_02514	xynB_2	134	Endo-1
Prokka_02520	entP	72	Bacteriocins enterocin-P
Prokka_01185	EF_11	296	Autolysin
Prokka_01195	ponA	794	Penicillin-sensitive transpeptidase
Prokka_01316	EHR_05900	673	Muramidase-2

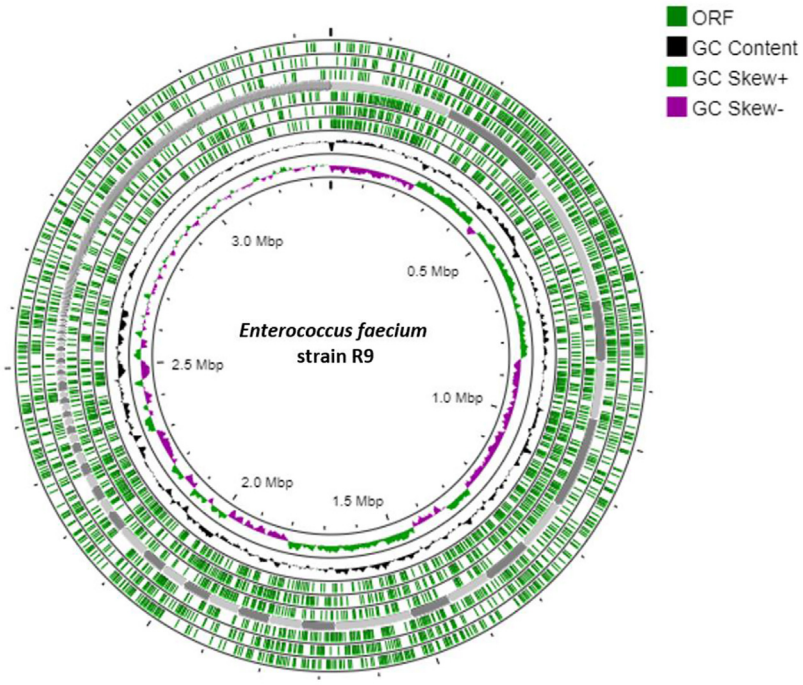

Fig. 1. Circular map of the genome sequence of *Enterococcus faecium* strain R9.

Enterocins are extracellular proteins encoded by *Enterococcus* genes and synthesized by ribosomes. They possess bacteriostatic abilities as a defense mechanism of bacteria that can stop the growth of other bacteria and give bacteriocin-producing bacteria a competitive advantage [3]. Most enterocins are class II bacteriocins known for their antilisterial activity but they can also inhibit other foodborne pathogens [4,5]. The enterocins secreted by *Enterococcus faecium* strain R9 (Enterocin B, enterocin P, and two-component enterocin X) belong to Class II bacteriocins that consists of unmodified small-molecule (<10 kDa) thermostable bacteriocins after translation. The maturation and activation of this class is completed without requirement of special enzymes other than signal peptides or transporters [3]. Enterocin B belongs to class IId, a non-pediocin-like, single-peptide bacteriocin. It shows strong homology to carnobacteriocin A [5,6]. Enterocin P belongs to class IIa, unmodified, pediocin-like bacteriocins [7]. Enterocin X is a class IIb bacteriocin composed of two antibacterial peptides X-alpha and X-beta with distinct properties. These two peptides, when combined in equal amounts, can produce antibacterial activity [8].

The structural genes for enterocin B and enterocin X-alpha and X-beta are located in close proximity to each other. Enterocin B gene cluster includes Lactococcin G processing and transport ATP-binding protein IagD (Fig. 2). ABC transporter gene (AbcA) in addition to Sensor histidine kinase LiaS and three bacteriocin immunity proteins were also identified in the sequence.

Furthermore, ResFinder software 3.0 was used to identify the acquired resistance genes. The resistance genes were Tobramycin, Virginiamycin, quinupristin, gentamicin, erythromycin, telithromycin, sisomicin, doxycycline, pristinamycin, azithromycin, chloramphenicol, dibekacin, tetracycline, netilmicin, lincomycin, minocycline, streptomycin (Table 3).

This work highlights the potential application of the enterocins produced by *Enterococcus faecium* strain R9 in the food industry for food preservation.

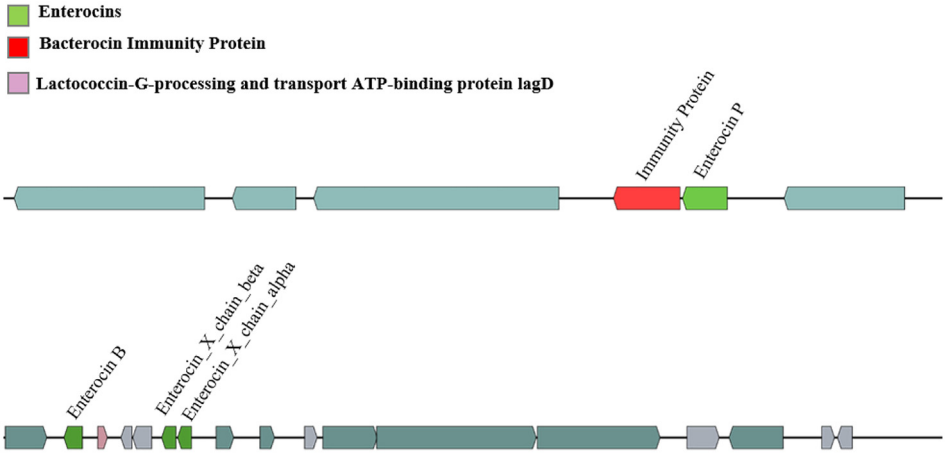

Fig. 2. Representation of the location of different genes encoding for enterocins.

Table 3
Summary of resistance genes in the genome of *Enterococcus faecium* strain R9 using ResFinder software 3.

Gene	Identity (%)	Alignment length	Contig	Phenotype
msr(C)	98.9	1479/1479	contig_00011	Erythromycin, telithromycin, quinupristin, pristinamycinia virginiamycin
msr(D)	95.4	1464/1464	contig_00019	Erythromycin, azithromycin, telithromycin, quinupristin, pristinamycinia virginiamycin
aac(6′)-Ii	99.8	549/549	contig_00030	Tobramycin, dibekacin, gentamicin, sisomicin, netilmicin
tet(L)	100	1377/1377	contig_00019	Doxycycline, tetracycline
tet(M)	96.4	1920/1920	contig_00019	Doxycycline, tetracycline, Minocycline
lnu(G)	100	804/804	contig_00002	Lincomycin
cat(pC194)	99.6	651/651	contig_00049	Chloramphenicol
mef(A)	98.1	1218/1218	contig_00019	Erythromycin, azithromycin
ant(6)-Ia	100	864/864	contig_00019	Streptomycin

3. Experimental Design, Materials and Methods

3.1. Genomic DNA Isolation

Enterococcus faecium strain R9 was cultured in Man, Rogosa, and Sharpe (MRS) broth (Thermo Fisher Scientific, Waltham, USA) overnight under aerobic conditions at 37 °C. Genomic DNA extraction from the overnight cultured bacteria was carried out using the GeneElute bacterial genomic DNA kit (Sigma-Aldrich, USA) according to the manufacturer's instructions. The concentration of the extracted DNA was determined using a Qubit 3.0 Fluorometer (Invitrogen, USA). The extracted DNA was analyzed by electrophoresis in a 0.8% agarose gel.

3.2. Genome Sequencing, Assembly, and Annotation

The genome of *Enterococcus faecium R9* was sequenced using Illumina HiSeq 2500 technology platforms. FASTQ sequence files were generated using bcl2fastq2 version 2.18. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads con-

taining PhiX control signal were removed. In addition, reads containing (partial) adapters were clipped (up to minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.11.5.

De novo assembly of the reads, including scaffolding and gap-filling were performed. The read error correction module BayesHammer (SPAdes version 3.10 genome assembly toolkit) was used to improve the quality of the FASTQ sequences [9]. Using SPAdes software, the high-quality reads were put together into contigs. Pilon version 1.11 [10] was used to correct misassemblies and nucleotide disagreement between the Illumina data and the contig sequences. The contigs were joined and put into scaffolds, where the orientation, order, and distance between them were determined using the insert size between the paired-end and/or matepair reads. The SSPACE Premium Scaffolder version 2.3 has been used to conduct the analysis [11]. The gapped regions within the scaffolds were (partially) closed in an automated manner using GapFiller version 1.10) [12].

Genome annotation has been carried out on the assembled contig or scaffold sequences using the Prokka Prokaryotic Genome Annotation System (http://vicbioinformatics.com/). The genome was analysed using ResFinder software 3.0 (https://cge.food.dtu.dk/services/ResFinder/) to identify the acquired resistance genes. BAGEL4 analysis software (http://bagel4.molgenrug.nl/) was used to screen the sequence and identify genes encoding bacteriocins. Unless stated otherwise, default parameters were used for all software tools.

Ethics Statements

All experimental protocols were approved by the Center Proposal Evaluation Committee (PEC) of Kuwait Institute for scientific research. All methods were performed in accordance with relevant institutional guideline and regulations with Reference No. PMO/PV/GM/073/2015, in compliance with the standards of animal rights.

CRediT Author Statement

Abrar Akbar: Software, Data curation; **Sabah Al-Momin**: Supervision, Investigation; **Mohamed Kishk**: Methodology; **Abdulaziz Al-Ateeqi**: Methodology; **Anisha Shajan**: Methodology; **Rita Rahmeh**: Conceptualization, Supervision, Validation. All co-authors co-wrote the paper.

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

Enterococcus faecium strain R9, whole genome shotgun sequencing project (Original data) (National center for Biotechnology Information (NCBI)).

Acknowledgments

This project was funded by Kuwait Foundation for the Advancement of Sciences (KFAS) under the project code PR20-12MI-01. The authors thank KFAS (Kuwait City, Kuwait), and Kuwait Institute for Scientific Research (Kuwait City, Kuwait) for their financial support.

References

[1] Akbar, Abrar; Rahmeh, Rita; Kishk, Mohamed; Al-Momen, Sabah. R9 a bacteriocin producing lactic acid bacteria isolated from camel milk. Mendeley Data, V1. 2022, doi:10.17632/yvnkzzb7yk.1.

[2] R. Rahmeh, A. Akbar, M. Kishk, T. Al Onaizi, A. Al-Shatti, A. Shajan, B. Akbar, S. Al-Mutairi, A. Yateem, Characterization of semipurified enterocins produced by Enterococcus faecium strains isolated from raw camel milk, J. Dairy Sci. 101 (2018) 4944–4952 S0022-0302(18)30218-2 [pii].

[3] Y. Wu, X. Pang, Y. Wu, X. Liu, X. Zhang, Enterocins: classification, synthesis, antibacterial mechanisms and food applications, Molecules 27 (2022) 2258.

[4] O. Ben Braiek, S. Smaoui, Enterococci: between emerging pathogens and potential probiotics, Biomed. Res. Int. 2019 (2019).

[5] M.E. Kasimin, S. Shamsuddin, A.M. Molujin, M.K. Sabullah, J.A. Gansau, R Jawan, Enterocin: promising biopreservative produced by enterococcus sp, Microorganisms 10 (2022) 684.

[6] P. Casaus, T. Nilsen, L.M. Cintas, I.F. Nes, P.E. Hernández, H. Holo, Enterocin B, a new bacteriocin from Enterococcus faecium T136 which can act synergistically with enterocin A, Microbiology 143 (1997) 2287–2294.

[7] L.M. Cintas, P. Casaus, L.S. Håvarstein, P.E. Hernandez, I.F. Nes, Biochemical and genetic characterization of enterocin P, a novel sec-dependent bacteriocin from Enterococcus faecium P13 with a broad antimicrobial spectrum, Appl. Environ. Microbiol. 63 (1997) 4321–4330.

[8] C. Hu, W. Malaphan, T. Zendo, J. Nakayama, K. Sonomoto, Enterocin X, a novel two-peptide bacteriocin from Enterococcus faecium KU-B5, has an antibacterial spectrum entirely different from those of its component peptides, Appl. Environ. Microbiol. 76 (2010) 4542–4545.

[9] 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, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.

[10] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS ONE 9 (2014) e112963 (Accessed 16 March 2022), doi:10.1371/journal.pone.0112963.

[11] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinformatics 27 (2011) 578–579 (Accessed 16 March 2022), doi:10.1093/bioinformatics/btq683.

[12] M. Boetzer, W. Pirovano, Toward almost closed genomes with GapFiller, Genome Biol. 13 (2012) R56, doi:10.1186/gb-2012-13-6-r56.