

Genome sequences and phenotypic profiles of three calf fecal Parabacteroides isolates

Hae-In Joe

Kyung Hee University

Jae-Yun Lee

Kyung Hee University

Hyun Sik Kim

Kyung Hee University Hospital

Jin-Woo Bae

baejw@khu.ac.kr

Kyung Hee University

Data Note

Keywords: Parabacteroides, calf-associated isolates, genome-based taxonomy, phylogenetic characterization, comparative genomics, average nucleotide identity, antimicrobial susceptibility

Posted Date: August 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9787371/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomic Data on October 1st, 2026. See the published version at <https://doi.org/10.1186/s12863-026-01491-x>.

Abstract

Objective

Parabacteroides is increasingly recognized as an important gut-associated bacterial genus involved in mammalian health and disease. Despite this, the taxonomy of mammal-derived *Parabacteroides* spp., particularly those from gastrointestinal tract-associated samples, remains understudied and is frequently misclassified. The data were collected to provide genome-resolved resources for three *Parabacteroides* strains isolated from feces of clinically healthy calves.

Data description:

The data include high-quality genome sequences for KHUB-01, KHUB-02, and KHUB-03 generated using PacBio single-molecule real-time sequencing with additional Illumina short reads for polishing. De novo assembly yielded a single circular contig for KHUB-01 and KHUB-03, whereas KHUB-02 was assembled into two contigs. Additional data files summarize sequencing output, assembly features, genome quality, annotation statistics, comparative genome metrics, reference genomes, and ResFinder-based antimicrobial resistance gene hits. Antimicrobial susceptibility profiles are also provided.

Objective

The mammalian gastrointestinal tract harbors a complex and dynamic community of microorganisms that play a pivotal role in maintaining host homeostasis, modulating the immune system, and protecting against pathogen colonization (1). Members of the genus *Parabacteroides* are Gram-negative, obligate anaerobes, non-spore-forming and non-motile bacilli. Since its reclassification from the genus *Bacteroides* in 2006, several species, including *Parabacteroides distasonis*, *Parabacteroides goldsteinii*, and *Parabacteroides merdae*, have been frequently identified as prevalent members of the gut microbiota of healthy mammals (1).

Increasing evidence indicates substantial strain-level genomic heterogeneity, which may contribute to functional differences among strains (2–5). Although these genomic insights have been characterized most extensively in *P. distasonis*, they underscore the importance of comparative genome analysis for newly isolated *Parabacteroides* strains. In this study, three *Parabacteroides* isolates were isolated from fecal samples of healthy calves. These microbial strains were identified in previous research as being associated with the alleviation of calf diarrhea following fecal microbiota transplantation (6). Using whole-genome sequencing, high-quality genome assemblies were obtained from the isolates and comparative genomic analysis was conducted to provide data on functional potential, metabolic pathways, virulence factors, and antimicrobial resistance profiles.

Data description

Among 20 healthy calves aged 21–50 days selected on the basis of physical condition and nutritional status, six met the predefined inclusion and exclusion criteria across the four study stages (6). Fecal samples were collected by rectal enema using clean disposable latex gloves. The collected samples were transported to the laboratory on dry ice and stored at -80°C until use.

A total of nine fecal samples from healthy calves were used for bacterial isolation, and all procedures were performed under strict anaerobic conditions in an anaerobic chamber containing 5% H₂, 5% CO₂, and 90% N₂. Each sample was suspended in sterile phosphate-buffered saline and subjected to tenfold serial dilution. Aliquots were spread onto GAM agar plates, and 382 colonies were randomly picked after 24 or 48 h of incubation. Selected colonies were subjected to at least three rounds of subculturing to obtain pure isolates. Colonies were subjected to species-level identification by 16S rRNA gene sequencing using the universal bacterial primers 27F and 1492R (7).

Genomic DNA was extracted using the Maxwell RSC Tissue DNA Kit following the manufacturer's instructions, with enzymatic pretreatment for cell lysis. Whole-genome sequencing was performed using single-molecule real-time sequencing technology (8). High-molecular-weight genomic DNA was used to generate SMRTbell libraries with a target insert size of 7–12 kb. Short-read data generated on an Illumina platform were quality filtered and used for Pilon-based error correction of the PacBio de novo assemblies.

Illumina sequencing generated 21,223,242 reads for KHUB-01, 26,802,174 reads for KHUB-02, and 23,141,090 reads for KHUB-03. PacBio sequencing generated 312,510,083 bases from 41,891 reads for KHUB-01, 512,045,320 bases from 62,184 reads for KHUB-02, and 494,786,598 bases from 62,454 reads for KHUB-03. De novo assembly resulted in a single circular contig of 3,872,635 bp for KHUB-01 and 3,615,838 bp for KHUB-03, with GC contents of 39.30% and 40.20%, respectively. KHUB-02 was assembled into two contigs with a total genome size of 5,408,123 bp, an N50 of 4,849,314 bp, and a GC content of 45.20%. Genome quality assessment using CheckM (9) indicated high completeness and low contamination for all assemblies.

Taxonomic classification of the assembled genomes was performed using GTDB-Tk (10). Whole-genome average nucleotide identity was calculated using fastANI (11), and phylogenomic analysis was conducted using the UBCG pipeline (12). Virulence-associated genes and antimicrobial resistance genes were identified using ABRicate by querying the VFDB (13) and ResFinder databases (14). Antimicrobial susceptibility of the isolates was evaluated using the disk diffusion method (15).

Data file 1–3 contain the genome sequence files for KHUB-01, KHUB-02, and KHUB-03, respectively. Data file 4 summarizes sequencing output, assembly statistics, genome quality, and annotation metrics. Data files 5–7 provide comparative assembly metrics, reference genome information, and ResFinder-based antimicrobial resistance gene hits. Data file 8 contains the phylogenomic tree and average nucleotide identity heatmap, and Data file 9 provides the full methodology description. The data files and data sets are summarized in Table 1.

Table 1
Overview of data files/data sets.

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Genome sequence of Parabacteroides sp. KHUB-01	FASTA (.fna)	GenBank https://identifiers.org/ncbi/insdc:JBWQEU0000000000(16) ; Data-use terms: NCBI/INSDC public data policy
Data file 2	Genome sequence of Parabacteroides sp. KHUB-02	FASTA (.fna)	GenBank https://identifiers.org/ncbi/insdc:JBWQEV0000000000(17) ; Data-use terms: NCBI/INSDC public data policy
Data file 3	Genome sequence of Parabacteroides sp. KHUB-03	FASTA (.fna)	GenBank https://identifiers.org/ncbi/insdc:JBWQEW0000000000(18) ; Data-use terms: NCBI/INSDC public data policy
Data file 4	Summary of sequencing output, assembly features, genome quality, and annotation statistics for KHUB-01, KHUB-02, and KHUB-03	MS Excel file (.xlsx)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.
Data file 5	Comparative genome assembly quality metrics of KHUB-01, KHUB-02, KHUB-03, and publicly available Parabacteroides reference genomes	MS Excel file (.xlsx)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.
Data file 6	Assembly names and organism names of reference genomes included in the comparative genome analysis	MS Excel file (.xlsx)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 7	ResFinder-based antimicrobial resistance gene hits identified in KHUB-01, KHUB-02, KHUB-03, and reference <i>Parabacteroides</i> genomes	MS Excel file (.xlsx)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.
Data file 8	Phylogenomic tree and average nucleotide identity heatmap for KHUB-01, KHUB-02, KHUB-03, and reference <i>Parabacteroides</i> genomes	TIFF image (.tif)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.
Data file 9	Full methodology description for isolate recovery, sequencing, genome assembly, polishing, quality assessment, annotation, comparative genomics, antimicrobial resistance screening, and antimicrobial susceptibility testing	PDF (.pdf)	figshare. Reserved DOI: https://doi.org/10.6084/m9.figshare.32315463(19) ; Private peer-review link: https://figshare.com/s/ff29bb1093a15a986f30 . License: CC BY 4.0.

Limitations

- The dataset comprises three *Parabacteroides* isolates recovered from healthy calf feces and is intended to provide genome-resolved resources for this defined isolate set rather than to represent the full diversity of calf gut *Parabacteroides* populations.

- Partial discordance was observed between genome-based AMR gene predictions and phenotypic susceptibility profiles, suggesting that additional mechanisms or uncharacterized resistance determinants may exist beyond the currently annotated AMR genes in the ResFinder database.
- Virulence-associated gene detection was based solely on VFDB annotation, which may not capture all relevant virulence determinants, particularly novel or uncharacterized ones.
- Antimicrobial susceptibility was evaluated using the disk diffusion method, and the results were interpreted using study-specific criteria; therefore, direct comparisons with studies using standardized interpretive criteria should be made with caution.

Abbreviations

AMR	
Antimicrobial resistance	
GAM	
Gifu anaerobic medium	
GC	
Guanine-cytosine	
GTDB-Tk	
Genome Taxonomy Database Toolkit	
N50	
N50 statistic (contig length at which 50% of the total assembly is covered)	
NCBI	
National Center for Biotechnology Information	
UBCG	
Up-to-date bacterial core gene	
VFDB	
Virulence factor database	

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board of Kyung Hee University [KHUASP(SE)-17-145] and complied with the Animal Protection Act and the Animal Welfare Guidelines of the World Organisation for Animal Health (WOAH; formerly OIE).

Consent for publication

Not applicable.

Author Contribution

H-I.J., J-Y.L., and H.S.K. contributed to bacterial isolation, genome analysis, data collection, manuscript preparation, and manuscript editing. H.S.K. and J-W.B. contributed to the conception and design of the study and manuscript editing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The genome sequences of *Parabacteroides* sp. KHUB-01 (16), KHUB-02 (17), and KHUB-03 (18) have been deposited in GenBank. For KHUB-01, the BioProject, BioSample, and genome accession records are available at <https://identifiers.org/ncbi/bioproject:PRJNA1406019>, <https://identifiers.org/ncbi/biosample:SAMN54759956>, and <https://identifiers.org/ncbi/insdc:JBWQEU000000000>, respectively. For KHUB-02, the corresponding records are available at <https://identifiers.org/ncbi/bioproject:PRJNA1406034>, <https://identifiers.org/ncbi/biosample:SAMN54760218>, and <https://identifiers.org/ncbi/insdc:JBWQEV000000000>, respectively. For KHUB-03, the corresponding records are available at <https://identifiers.org/ncbi/bioproject:PRJNA1406118>, <https://identifiers.org/ncbi/biosample:SAMN54760312>, and <https://identifiers.org/ncbi/insdc:JBWQEW000000000>, respectively. Data files 4–9 have been deposited in figshare (19). The dataset has been assigned the reserved DOI <https://doi.org/10.6084/m9.figshare.32315463>, which will resolve upon public release. During peer review, the dataset is accessible through the following private peer-review link: <https://figshare.com/s/ff29bb1093a15a986f30>.

References

1. Sakamoto M, Benno Y. Reclassification of *Bacteroides distasonis*, *Bacteroides goldsteinii* and *Bacteroides merdae* as *Parabacteroides distasonis* gen. nov., comb. nov., *Parabacteroides goldsteinii* comb. nov. and *Parabacteroides merdae* comb. nov. *Int J Syst Evol Microbiol.* 2006;56(Pt 7):1599–605.
2. Ezeji JC, Sarikonda DK, Hopperton A, Erkkila HL, Cohen DE, Martinez SP, et al. *Parabacteroides distasonis*: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes.* 2021;13(1):1922241.
3. Feng X, Liu Y, Xu S, Ma J, Yuan H, Wang H, et al. Functional analysis of *Parabacteroides distasonis* F4: a novel probiotic strain linked to calf growth and rumen fermentation. *J Anim Sci Biotechnol.* 2025;16(1):50.

4. Bank NC, Singh V, Rodriguez-Palacios A. Classification of *Parabacteroides distasonis* and other Bacteroidetes using O- antigen virulence gene: RfbA-Typing and hypothesis for pathogenic vs. probiotic strain differentiation. *Gut Microbes*. 2022;14(1):1997293.
5. Chamarande J, Cunat L, Alauzet C, Cailliez-Grimal C. In Silico Study of Cell Surface Structures of *Parabacteroides distasonis* Involved in Its Maintenance within the Gut Microbiota. *Int J Mol Sci*. 2022;23(16).
6. Kim HS, Whon TW, Sung H, Jeong YS, Jung ES, Shin NR, et al. Longitudinal evaluation of fecal microbiota transplantation for ameliorating calf diarrhea and improving growth performance. *Nat Commun*. 2021;12(1):161.
7. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64(Pt 2):346–51.
8. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323(5910):133–8.
9. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043–55.
10. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;36(6):1925–7.
11. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114.
12. Na SI, Kim YO, Yoon SH, Ha SM, Baek I, Chun J. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol*. 2018;56(4):280–5.
13. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019;47(D1):D687–92.
14. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75(12):3491–500.
15. Joe HI, Choi JW, Lee JY, Sung H, Jeong SW, Jeong YS, et al. Description of *Luteibacter aegosomatis* sp. nov., *Luteibacter aegosomaticola* sp. nov., and *Luteibacter aegosomatissinici* sp. nov. isolated from the Intestines of *Aegosoma sinicum* Larvae. *J Microbiol*. 2023;61(6):603–13.
16. Bae J-W, Joe H-I, Lee J-Y, Kim HS. *Parabacteroides* sp. KHUB-01, whole genome shotgun sequencing project. National Center for Biotechnology Information. 2026 [cited 21 May 2026]. Available from: <https://identifiers.org/ncbi/insdc:JBWQEU000000000>
17. Bae J-W, Joe H-I, Lee J-Y, Kim HS. *Parabacteroides* sp. KHUB-02, whole genome shotgun sequencing project. National Center for Biotechnology Information. 2026 [cited 21 May 2026]. Available from: <https://identifiers.org/ncbi/insdc:JBWQEV000000000>
18. Bae J-W, Joe H-I, Lee J-Y, Kim HS. *Parabacteroides* sp. KHUB-03, whole genome shotgun sequencing project. National Center for Biotechnology Information. 2026 [cited 21 May 2026].

Available from: <https://identifiers.org/ncbi/insdc:JBWQEW000000000>

19. Bae J-W, Joe H-I, Lee J-Y, Kim HS. Supporting data for genome sequence and phenotypic datasets for three calf-associated Parabacteroides isolates. figshare. 2026. Available from: <https://doi.org/10.6084/m9.figshare.32315463>