Genomic characterization of two *Faecalibacterium* strains isolated from a healthy Japanese individual

Hidehiro Toh[a#], Yuta Fujimura[b], Soma Ryuto[b], Tien Thi Thuy Nguyen[b,c], Yusuke Fujii[b], Takehiro Matsubara[d], Hidetoshi Morita[b], Kensuke Arakawa[b#]

a, Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka, Japan.
b, Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, Okayama, Japan.
c, College of Agriculture and Forestry, Hue University, Hue, Vietnam.
d, Division of Biobank, Center for Comprehensive Genomic Medicine, Okayama University Hospital, Okayama, Japan.

#Address correspondence to Hidehiro Toh (toh@nig.ac.jp) and Kensuke Arakawa (karakawa@okayama-u.ac.jp)

Abstract

The genome sequences of two *Faecalibacterium* strains isolated from a healthy Japanese individual were analyzed. Comparative genomics revealed the smallest genome within the genus to date and identified core gene clusters, providing new insights into the minimal genomic requirements and evolutionary adaptation of *Faecalibacterium* species.

Running title

Faecalibacterium genomes from a Japanese individual

Announcement

Faecalibacterium species are among the most abundant butyrate-producing bacteria in the human gut and play a key role in maintaining gut barrier integrity and modulating inflammation (1). Increasing evidence supports their health-promoting properties and highlights them as promising next-generation probiotics (2). However, genomic information on *Faecalibacterium* species other than *F. prausnitzii* remains limited, especially in Asian populations (3, 4). To address this gap and better understand the diversity and functional potential of *Faecalibacterium* in the Japanese gut, we isolated two novel strains, F15 and a30, from a healthy individual and determined their draft genome sequences.

A fecal sample from a healthy 22-year-old Japanese man was serially diluted under anaerobic conditions and plated on modified GAM agar containing 1% pectin. After 48 hours of anaerobic incubation at 37°C, colonies with typical *Faecalibacterium* morphology were selected and identified by 16S rRNA gene sequencing. The isolates were cultured in modified GAM broth with 0.5% sodium acetate at 37°C for 24 hours anaerobically. Cells were harvested, washed, and treated sequentially with lysozyme, achromopeptidase, proteinase K, and SDS. Genomic DNA was extracted using phenol:chloroform:isoamyl alcohol, purified by ethanol precipitation and centrifugation, treated with RNase, further purified using NaCl and PEG precipitation, washed with ethanol, and resuspended in TE buffer.

Genomic libraries were prepared using the Illumina Nextera XT DNA Library Prep Kit and sequenced with paired-end reads on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 (600 cycles). Sequencing reads were assembled using CLC Genomics Workbench v10.1.1 with default parameters, and protein-coding genes were annotated using DFAST v1.6.0 (5). A summary of the sequencing and genome assembly statistics is presented in Table 1. For species-level classification of *Faecalibacterium*, the *recA* gene proved more reliable than the 16S rRNA gene as a phylogenetic marker (6). The *recA* sequence of strain F15 showed 99.1% identity to *Faecalibacterium duncaniae* JCM 31915^{T} (7) and an average nucleotide identity (ANI) of 96.7%, supporting its classification as *F. duncaniae* (Fig. 1). Strain a30 showed 99.1% *recA* sequence identity to *Faecalibacterium taiwanense* HLW78^{T} (8) and an ANI of 97.8%, consistent with classification as *F. taiwanense* (Fig. 1).

Remarkably, strain a30 has the smallest genome (2.55 megabases) among all *Faecalibacterium* species reported to date. To estimate the core genome of the genus, we compared the gene clusters of strain a30 with those of 11 other reference strains representing *Faecalibacterium*. A total of 1,203 genes were shared by all 12 genomes and were inferred to constitute the core gene cluster of *Faecalibacterium* species.

These findings provide new insights into the minimal genomic requirements for the genus and establish a foundation for understanding its essential functions and evolutionary adaptations.

Data availability. This study is associated with the BioProject registered under accession number PRJDB15359. Draft genome sequences for the two strains (F15 and a30) have been submitted to the DDBJ/GenBank/EMBL databases under the accession numbers BSSS01000001–BSSS01000141 and BSSR01000001–BSSR01000089, respectively. The raw sequencing data are available in the Sequence Read Archive under accession number DRA020529.

ACKNOWLEDGMENTS

We thank the Okayama University Hospital Biobank for providing Illumina sequencing services. The human fecal sample was obtained from a healthy Japanese man who provided informed consent. This study was approved by the Okayama University Ethics Committee (#1610-026). This work was supported by a JSPS KAKENHI Grant-in-Aid for Scientific Research (grant number JP23K18086). The authors note with sadness that Hidetoshi Morita, a co-author of this manuscript, passed away prior to its submission. His contributions are gratefully acknowledged.

REFERENCES

1. Singh V, Lee G, Son H, Koh H, Kim ES, Unno T, Shin JH. 2023. Butyrate producers, “The Sentinel of Gut”: Their intestinal significance with and beyond butyrate, and prospective use as microbial therapeutics. Front Microbiol 13:1103836.
2. Martín R, Rios-Covian D, Huillet E, Auger S, Khazaal S, Bermúdez-Humarán, LG, Sokol H, Chatel JM, Langella P. 2023. *Faecalibacterium*: a bacterial genus with promising human health applications. FEMS Microbiol Rev 47:fuad039.
3. Lindstad LJ, Lo G, Leivers S, Lu Z, Michalak L, Pereira GV, Røhr AK, Martens EC, McKee LS, Louis P, Duncan SH, Westereng B, Pope PB, La Rosa SL. 2021. Human gut *Faecalibacterium prausnitzii* deploys a highly efficient conserved system to cross-feed on β-mannan-derived oligosaccharides. mBio 12:e0362820.
4. Ueda A, Shinkai S, Shiroma H, Taniguchi Y, Tsuchida S, Kariya T, Kawahara T, Kobayashi Y, Kohda N, Ushida K, Kitamura A, Yamada T. 2021. Identification of *Faecalibacterium prausnitzii* strains for gut microbiome-based intervention in Alzheimer’s-type dementia. Cell Rep Med 2:100398.

5. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039.
6. Tanno H, Chatel JM, Martin R, Mariat D, Sakamoto M, Yamazaki M, Salminen S, Gueimonde M, Endo A. 2023. New gene markers for classification and quantification of *Faecalibacterium* spp. in the human gut. FEMS Microbiol Ecol 99:fiad035.
7. Sakamoto M, Sakurai N, Tanno H, Iino T, Ohkuma M, Endo A. 2022. Genome-based, phenotypic and chemotaxonomic classification of *Faecalibacterium* strains: proposal of three novel species *Faecalibacterium duncaniae* sp. nov., *Faecalibacterium hattorii* sp. nov. and *Faecalibacterium gallinarum* sp. nov. Int J Syst Evol Microbiol 72:005379.
8. Liou JS, Zhang WL, Hsu LW, Chen CC, Wang YT, Mori K, Hidaka K, Hamada M, Huang L, Watanabe K, Huang CH. 2024. *Faecalibacterium taiwanense* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 74:006413.
9. Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027.

TABLE 1 Genomic features of two *Faecalibacterium* strains analyzed in this study

Species	Strain	No. of paired reads	No. of contigs	Genome size (bp)	N_{50} contig size (bp)	GC content (%)	Genome coverage (×)	No. of protein-coding genes
F. duncaniae	F15	550,916	141	2,988,843	44,918	56.2	55.3	2,791
F. taiwanense	a30	481,940	89	2,554,504	55,053	56.6	56.6	2,308

Figure legends

Figure 1. Neighbor-joining phylogenetic tree based on the *recA* gene sequences showing the relationships between the two analyzed strains and related *Faecalibacterium* species. The *recA* gene sequences for each species were obtained

from reference genomes available at NCBI. Multiple sequence alignment was performed using ClustalW2, and the phylogenetic tree was constructed using MEGA11 (9). Values in parentheses indicate genome size in megabases. The scale bar indicates the number of nucleotide substitutions per site.

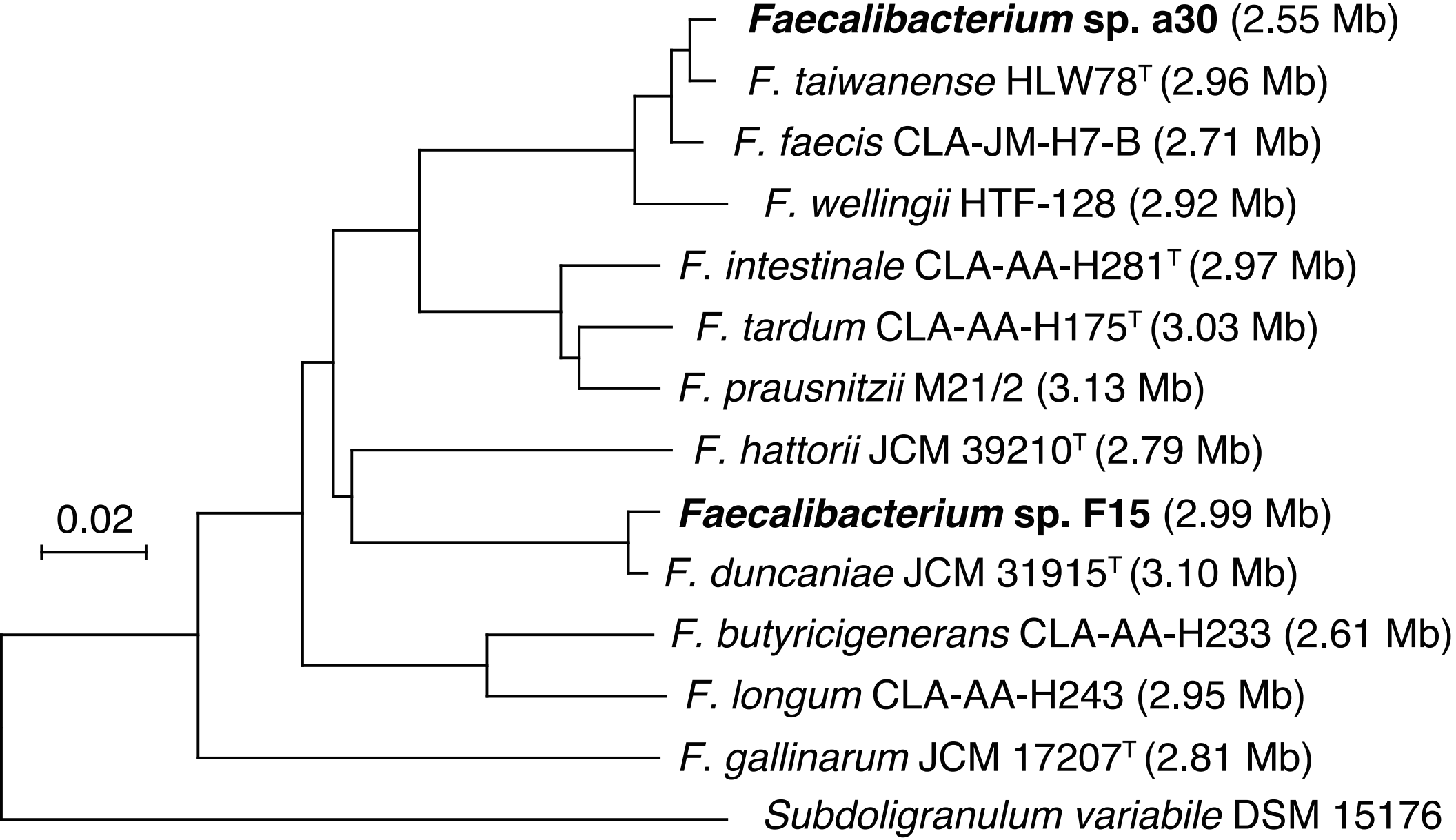
***Faecalibacterium* sp. a30** (2.55 Mb)
F. taiwanense HLW78^{T} (2.96 Mb)
F. faecis CLA-JM-H7-B (2.71 Mb)
F. wellingii HTF-128 (2.92 Mb)
F. intestinale CLA-AA-H281^{T} (2.97 Mb)
F. tardum CLA-AA-H175^{T} (3.03 Mb)
F. prausnitzii M21/2 (3.13 Mb)
F. hattorii JCM 39210^{T} (2.79 Mb)
***Faecalibacterium* sp. F15** (2.99 Mb)
F. duncaniae JCM 31915^{T} (3.10 Mb)
F. butyricigenerans CLA-AA-H233 (2.61 Mb)
F. longum CLA-AA-H243 (2.95 Mb)
F. gallinarum JCM 17207^{T} (2.81 Mb)
Subdoligranulum variabile DSM 15176
0.02

Figure 1