Complete genome sequences of *Fusobacterium watanabei* sp. isolates

Martha A. Zepeda-Rivera[1], Falk Ponath[2], Kaitlyn N. Lewis[1], Rutika P. Gavate[1], Floyd E. Dewhirst[3,4], Junko Tomida[5], Yoshiaki Kawamura[5], Kaori Tanaka[6], Susan Bullman[2], Christopher D. Johnston[1#]

[1]Genomic Medicine, MD Anderson Cancer Center, Houston, Texas, USA

[2]Immunology, James P. Allison Institute, MD Anderson Cancer Center, Houston, Texas, USA

[3]ADA Forsyth Institute, Cambridge, MA, USA

[4]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts, USA

[5]Department of Microbiology, School of Pharmacy, Aichi Gakuin University, Nagoya, Aichi, Japan

[6]Division of Anaerobe Research, Life Science Research Center, Gifu University, Yanagido, Gifu, Japan

[#]**Address Correspondence to**:

Christopher D. Johnston, PhD, cdjohnston@mdanderson.org

Keywords: Fusobacterium, nomenclature, Illumina sequencing

Abstract

We report the complete genome sequences of eight *Fusobacterium watanabei* clinical isolates, ranging from 1.95 to 2.09 Mbp. Analysis against the Genome Taxonomy Database (GTDB) indicates that *Fusobacterium watanabei* genomes are part of the "*Fusobacterium nucleatum_J*" group, which also encompasses the previously published FNU strain and *Fna* C1 isolates.

Announcement

Fusobacterium watanabei was initially described as a novel species closely related to the *Fusobacterium nucleatum* sensu lato subspecies: *F. nucleatum* subsp. *animalis, F. nucleatum* subsp. *nucleatum, F. nucleatum* subsp. *polymorphum, F. nucleatum* subsp. *vincentii*[1]. Species designation was based on whole genome comparisons of eight *Fusobacterium watanabei* clinical isolates against *Fusobacterium nucleatum* sensu lato subspecies members. Analysis indicated average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) metrics consistent with species-level assignment[1]. Here, we report and release the genome sequences for these eight clinical isolates of *Fusobacterium watanabei* (**Table 1**). Genome assembly indicates a genome size range from 1.95 to 2.09 Mbp (average of 2.01 Mbp) with GC content ranging from 26.82% to 26.99% (average of 26.92%).

Although the PAGU 1796 (type strain) 16S rRNA gene and partial sequences of three marker genes were available at National Center for Biotechnology Information (NCBI), the draft genome was only available in the Short Read Archive. Here, analysis of these eight *Fusobacterium watanabei* assembled genomes, including PAGU 1796 (type strain), against the Genome Taxonomy Database using GTDB-tk[2] assigned them all to the currently designated "*Fusobacterium nucleatum_J*" group, a member of the *Fusobacterium nucleatum* sensu lato group

of taxa. "*Fusobacterium nucleatum_J*" also encapsulates the strain 13-08-02, strain "*Fusobacterium nucleatum* subsp. *unique* (FNU)" previously described in Ma et al.[3], and members of the group we recently described as "*Fusobacterium nucleatum* subsp. *animalis* clade 1 (*Fna* C1)"[4]. ANI values between "*Fusobacterium watanabei*", "*FNU*", and "*Fna* C1" genomes range from 98.1228% to 99.9995%, supporting the observation that all three of these labels represent the same *Fusobacterium* lineage (**Figure 1**). The lack of an annotated genome for *Fusobacterium watanabei* in NCBI limited the ability of investigators to recognize and apply the name "*Fusobacterium watanabei*" to describe FNU[3] and *Fna* C1[4] genomes. Nevertheless, despite the use of different names, observations regarding this lineage are in agreement across publications[1,3,4].

The taxonomy of *Fusobacterium* species has recently undergone significant revision[5]. Based on the proposal by Kook et al.[6] to elevate *Fusobacterium nucleatum* sensu lato subspecies to the species level assignments, the four historic subspecies have been revised to individual species: *Fusobacterium animalis, Fusobacterium nucleatum, Fusobacterium polymorphum* and *Fusobacterium vincentii.* Within this context of changing nomenclature, it is fortuitous that *Fusobacterium watanabei* is a recognized species name under the List of Prokaryotic names with Standing in Nomenclature (LPSN)[7] and can now be used for genomes in databases designated as "*FNU*", "*Fna* C1" and "*F. nucleatum_J".* Here, we have publicly released genomes for strains PAGU 1795, PAGU 1796, PAGU 1797, PAGU 1798, PAGU 1799, PAGU 1800, PAGU 1801, and PAGU 1802. This increases the number of representative genomes to 34 for the *Fusobacterium watanabei* ("*FNU*"/"*Fna* C1"/"*F. nucleatum_J"*) group, which remains the closest phylogenetic lineage to the colorectal cancer-associated *Fusobacterium animalis*[4]. Finally, it is noteworthy that the PAGU 1796 *Fusobacterium watanabei* type strain, previously deposited to DSMZ under CCUG 74246, now has an available assembled genome deposited under PRJEB87340.

Figure

Figure 1: Average nucleotide identity between "*Fusobacterium watanabei*", "*FNU*", "*Fna* C1" genomes

Clustered dendrogram depicts the FastANI values between *Fusobacterium watanabei* genomes, including the previously published "*FNU*"[3] (‡) , and "*Fna* C1"[4] (*) strains. Color of each individual box represents the fastANI value, ranging from 98.0% (white) to 100% (dark green).

Table

Genome Name	Size (bp)	GC (%)	No. Contigs	No. Reads	N50
PAGU 1795	2,015,742	26.89	48	1,310,315	177,963
PAGU 1796	1,951,877	26.86	44	1,292,750	144,056
PAGU 1797	2,005,769	26.98	34	1,235,122	278,257
PAGU 1798	2,012,989	26.85	44	1,422,737	317,217
PAGU 1799	2,087,850	26.82	53	1,421,568	217,125
PAGU 1800	2,009,578	26.90	92	1,342,333	120,039
PAGU 1801	2,004,759	26.99	41	1,461,651	278,257
PAGU 1802	2,005,282	26.98	42	1,595,261	159,188

Table 1: *Fusobacterium watanabei* genomes
Table shows each *Fusobacterium watanabei* isolate name, with the size and the percentage of GC content of each resulting assembly.

Data Availability

Previously deposited draft genome data is available under DRA009376 and DRA009930. Previously deposited partial gene sequences for PAGU 1796 are available under accession numbers NR_179326.1, LC514068.1, LC514067.1, LC514066.1, and LC514065.1. Whole-genome sequences for PAGU 1795, PAGU 1796, PAGU 1797, PAGU 1798, PAGU 1799, PAGU 1800, PAGU 1801, and PAGU 1802 are available in ENA under the study accession number PRJEB87340.

Acknowledgements

Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award number [R01 DE027850] (to C.D.J.), and National Cancer Institute [R01CA288827] (to C.D.J and S.B], as well as start-up funds provided by UT MD Anderson Cancer Center (to S.B and C.D.J).

96 **Methods**

97 ***Bacterial culturing***

98 Strains PAGU 1795, PAGU 1796, PAGU 1797, PAGU 1798, PAGU 1799, PAGU 1800, PAGU
99 1801, and PAGU 1802 were grown on anaerobic blood agar plates (Remel™, CDC
100 formulation, R01036) at 37 °C degree under anaerobic conditions (Baker Concept
101 1000; N_2:H_2:CO_2 90:5:5).White-greyish colonies appeared between 24h to 48h which were used
102 for the downstream isolation of genomic DNA.

103 ***Genomic DNA isolation***

104 Genomic DNA was extracted using a modified protocol with the Monarch® Spin gDNA Extraction
105 Kit (New England Biolabs, Ipswich, MA, USA). Briefly, the resuspended bacterial pellet and lysis
106 buffer were combined and transferred to an MP Lysing Matrix B tube (MP Biomedicals, Santa
107 Ana, CA, USA), followed by mechanical disruption using the MP Biomedicals Fastprep-24 Bead
108 Beater (MP Biomedicals, Santa Ana, CA, USA) at 4.0 m/s for 20 seconds. Subsequent steps
109 followed the standard procedure outlined in the Monarch® Spin gDNA Extraction Kit protocol.

110 ***Illumina DNA Sequencing***

111 Illumina sequencing libraries were prepared using the tagmentation-based and PCR-based
112 Illumina DNA Prep kit (Illumina, San Diego, CA, USA) and custom IDT 10bp unique dual indices
113 (UDI) (Integrated DNA Technologies, Inc., Coralville, Iowa, USA) with a target insert size of
114 280bp. No additional DNA fragmentation or size selection steps were performed. Illumina
115 sequencing was performed on an Illumina Novaseq X Plus sequencer in one or more multiplexed
116 shared-flow-cell runs, producing 2X151bp paired-end reads. Demultiplexing, quality control and
117 adapter trimming was performed with bcl-convert (v.4.2.4). Short read assembly was performed
118 with Unicycler[8].

Phylogenetic classification in Genome Database Taxonomy (GTDB)

Phylogenetic classifications were assessed using GTDB-tk [2] as listed in Table 2 (https://github.com/Ecogenomics/GTDBTk).

Average Nucleotide Identity

Pairwise average nucleotide identity values were calculated using FastANI [9] via the DOE Systems Biology Knowledgebase (KBase) [10]. Clustered dendogram generated with pheatmap R package, RStudio version 2024.12.0+467.

References

1 Tomida, J. *et al. Fusobacterium watanabei* sp. nov. As additional species within the genus Fusobacerium, isolated from human clinical specimens. *Anaerobe* **69**, 102323 (2021). https://doi.org/10.1016/j.anaerobe.2021.102323

2 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927 (2019). https://doi.org/10.1093/bioinformatics/btz848

3 Ma, X. *et al.* Pangenomic Study of *Fusobacterium nucleatum* Reveals the Distribution of Pathogenic Genes and Functional Clusters at the Subspecies and Strain Levels. *Microbiol Spectr* **11**, e0518422 (2023). https://doi.org/10.1128/spectrum.05184-22

4 Zepeda-Rivera, M. *et al.* A distinct *Fusobacterium nucleatum* clade dominates the colorectal cancer niche. *Nature* **628**, 424-432 (2024). https://doi.org/10.1038/s41586-024-07182-w

5 Krieger, M., Guo, M. & Merritt, J. Reexamining the role of *Fusobacterium nucleatum* subspecies in clinical and experimental studies. *Gut Microbes* **16**, 2415490 (2024). https://doi.org/10.1080/19490976.2024.2415490

6 Kook, J. K. *et al.* Genome-Based Reclassification of *Fusobacterium nucleatum* Subspecies at the Species Level. *Curr Microbiol* **74**, 1137-1147 (2017). https://doi.org/10.1007/s00284-017-1296-9

7 Oren, A. & Garrity, G. M. Valid publication of new names and new combinations effectively published outside the IJSEM. *Int J Syst Evol Microbiol* **71** (2021). https://doi.org/10.1099/ijsem.0.004846

8 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595 (2017). https://doi.org/10.1371/journal.pcbi.1005595

9 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114 (2018). https://doi.org/10.1038/s41467-018-07641-9

10 Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* **36**, 566-569 (2018). https://doi.org/10.1038/nbt.4163