AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Neopoerus faecalis* gen. nov., sp. nov., an *Oscillospiraceae* Strain Isolated from Human Feces

Marta Selma-Royo,[a] Liviana Ricci,[a] Davide Golzato,[a] Charlotte Servais,[a] Federica Armanini,[a] Francesco Asnicar,[a] Federica Pinto,[a] Nicola Segata[a]

[a]CIBIO Department, University of Trento, Povo, Trento, Italy

Marta Selma-Royo and Liviana Ricci contributed equally to this article. Author order was determined by project responsability assignment.

ABSTRACT Here, we report the isolation and genome assembly of a strictly anaerobic bacterium from a previously uncharacterized species in the *Oscillospiraceae* family, isolated from a fecal sample from a healthy adult human. The name *Neopoerus faecalis* gen. nov., sp. nov. is proposed.

The family *Oscillospiraceae* includes genera such as *Faecalibacterium* and *Ruminococcus* that are linked to host health (1) and are predicted to produce beneficial metabolites (2), while several representatives of this family remain to be cultivated.

We describe the isolation of an *Oscillospiraceae* strain in an anaerobic chamber (Baker Ruskinn Concept 400) (95% N_2/5% H_2). The fecal sample was collected from a healthy donor in November 2022 (Trento, Italy), and it was processed within the hour. Two fecal dilutions (10^{-5} and 10^{-7}) were spread onto modified brain-heart infusion medium (BHIm; Fluka AG) agar plates supplemented with 2.5 g/L yeast extract (Fluka AG), 2 g/L cellobiose (Alfa Aesar), 2 g/L maltose (Sigma-Aldrich), 2 g/L soluble starch (Rozzano), 1 g/L L-cysteine (Sigma-Aldrich), 1 mg/L resazurin sodium salt (Sigma-Aldrich), 0.005% vitamin K1 (Alfa-Aesar), 5 mg/L hemin (Thermo Fisher), and 4% defibrinated sheep blood (SB, Microbiol Diagnostics). After 3 days of incubation (37°C), one single colony was inoculated in 5 ml of BHIm broth plus SB for 48h. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) and used for sequencing library preparation with the Illumina DNA prep and tagmentation kit. The libraries were sequenced (150-bp paired-end reads) on a NovaSeq 6000 instrument with S4 flow cell reagents (Illumina) at the University of Trento (Italy), after a cleaning step using 0.6× Agencourt AMPure XP beads.

The raw reads were preprocessed using Trim Galore (parameters: "–stringency 5 –length 75 –quality 20 –max_n 2 –trim-n") (https://github.com/FelixKrueger/TrimGalore). A total of 7,772,460 high-quality paired-end reads (mean quality [Q] value, 35.47) were retained, with a mean read length of 148.21 bp. Genome assembly was performed using SPAdes v3.15.2 (3) (parameters: –careful -k 21,33,55,77,99,127) using 30% of the high-quality reads (https://github.com/lh3/seqtk; parameters: sample module), as they were already above 100× coverage for a putative genome of 3.5 Mb. After standard filtering using the NCBI pipeline and removal of contigs of <1,000 bp, we obtained 36 contigs spanning a total length of 2.54 Mb. The assembly and annotation statistics are reported in Table 1. The assembly statistics were computed using QUAST v5.1.0rc1 (4), and measures of completeness and contamination were obtained using the lineage_wf workflow from CheckM v1.1.2 (5). Unless specified, all the computational tools used in this work were used with default parameters.

Annotation using Prokka v1.14 (6) revealed that the strain harbors a total of 2,475 genes. Screening for antibiotic resistance genes using the Resistance Gene Identifier v5.1.1 (7) identified resistance to tetracycline, diaminopyrimidine, and aminoglycoside. Phylogenetic

Editor Simon Roux, DOE Joint Genome Institute

Address correspondence to Nicola Segata, nicola.segata@unitn.it.

The authors declare no conflict of interest.

Received 19 April 2023
Accepted 6 June 2023
Published 21 June 2023

TABLE 1 Summary of the statistics of the *Neopoerus faecalis* sp. nov. genome assembly[a]

Parameter	Value
Total length (bp)	2,542,847
No. of scaffolds	36
GC content (%)	56.22
Mean coverage (×)	116
Size of longest scaffold (bp)	496,649
N_{50} (bp)	274,696
L_{50}	4
No. of coding sequences	2,408
No. of rRNAs	7
No. of tRNAs	57
Estimated completeness (%)	98.66
Estimated contamination (%)	0

[a] The genome assembly was annotated using Prokka v1.14. The assembly statistics were computed using QUAST v5.1.0rc1, and the completeness and contamination were obtained using the lineage_wf workflow from CheckM v1.1.2. The mean coverage was computed using CMSeq v1.0.4 (https://github.com/SegataLab/cmseq).

analysis using the PhyloPhlAn v3.0 pipeline (January 21 version; parameters: -d phylophlan –diversity low –fast –force_nucleotides) (8) showed that the closest taxonomically defined species was *Dysosmobacter welbionis* (GenBank accession number GCF_005121165.3) at <80% average nucleotide identity (ANI; computed using FastANI v1.33 [9]) (Fig. 1). This genetic distance to the closest known taxa warrants the definition of a new genus and species (10). The closest reference genome available in the NCBI database was that of *Dysosmobacter* sp. strain BX15 (ANI, 85.52%; GCA_014297285), a genus-level mislabeled strain, since it shows an ANI of <85% with the *Dysosmobacter* sp. reference genome (Fig. 1), thus warranting the classification of the new strain as *Neopoerus faecalis* sp. nov. Pipeline analysis using METABOLIC (11) revealed genes related to complex carbohydrate degradation, such as cellulose and chitin. Previous surveys estimated the prevalence of *Neopoerus faecalis* sp. nov. at 40.6% after the mapping of its specific marker genes across more than 24,500 available metagenomes from five continents (12).

Data availability. This study project is available under NCBI BioProject accession number PRJNA939950 and BioSample accession number SAMN33549472. The assembly

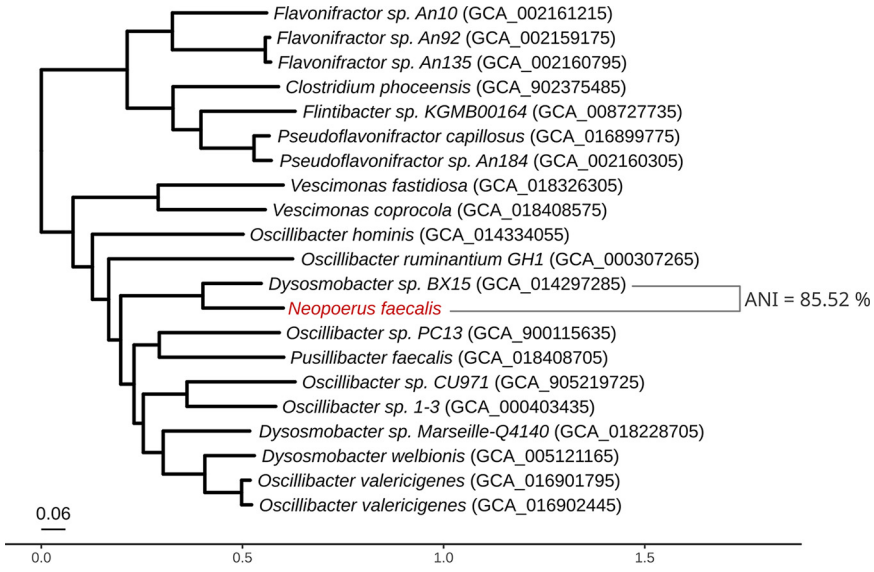

FIG 1 Phylogenetic tree of the *Neopoerus faecalis* gen. nov, sp. nov. isolate and related taxa with available reference genomes based on whole-genome sequencing (WGS). Sequence accession numbers are given in parentheses. The closest taxonomically defined species is *Dysosmobacter welbionis* (GenBank accession number GCA_005121165), with an ANI of <80%. The closest available reference genome, *Dysosmobacter* sp. strain BX15 (GCA_014297285), is mislabeled at the genus level, since it has an ANI of <85% to the *Dysosmobacter* reference strain (*Dysosmobacter welbionis*).

and the reads used to assemble the genome are available under GenBank accession number JARFXX000000000 and SRA accession number SRX20179702, respectively.

ACKNOWLEDGMENT

This project was funded by the European Union (ERC, microTOUCH, 101045015; ERC META-PG 716575; "ONCOBIOME" 825410) and from the Italian Ministry of Health (RF-2016-02364814).

REFERENCES

1. Leylabadlo HE, Ghotaslou R, Feizabadi MM, Farajnia S, Moaddab SY, Ganbarov K, Khodadadi E, Tanomand A, Sheykhsaran E, Yousefi B, Kafil HS. 2020. The critical role of Faecalibacterium prausnitzii in human health: an overview. Microb Pathog 149:104344. https://doi.org/10.1016/j.micpath.2020.104344.
2. Yang J, Li Y, Wen Z, Liu W, Meng L, Huang H. 2021. Oscillospira—a candidate for the next-generation probiotics. Gut Microbes 13:1987783. https://doi.org/10.1080/19490976.2021.1987783.
3. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.
4. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150. https://doi.org/10.1093/bioinformatics/bty266.
5. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
6. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
7. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935.
8. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500. https://doi.org/10.1038/s41467-020-16366-7.
9. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.
10. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N. 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 176:649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001.
11. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, Karaoz U, Anantharaman K. 2020. METABOLIC: high-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. Microbiome 10:33. https://doi.org/10.1186/s40168-021-01213-8.
12. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM, Nickols WA, Piccinno G, Piperni E, Punčochář M, Valles-Colomer M, Tett A, Giordano F, Davies R, Wolf J, Berry SE, Spector TD, Franzosa EA, Pasolli E, Asnicar F, Huttenhower C, Segata N. 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. Nat Biotechnol https://doi.org/10.1101/2022.08.22.504593.