AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

Article title

Draft Genome Sequence of *Intestinibacter bartletti DSM 16795*, a Human Gut Isolate Associated with Autism Microbiome Studies

Authors

Christina Johne[1,#],Mason Kramer[1,#], Raneem Makrai[1,#], Holly Moffett[1,#], Fiona Shavilof [1,#], Tricia A. Van Laar [1,]✉

All authors contributed equally to this work. Author order was determined alphabetically by surname.

Affiliations

[a]Department of Biological Sciences, California State University, Stanislaus, Turlock, California, USA

Running title

Genome of *I. bartletti* DSM 16795 from Human Gut

Corresponding author's email address

#Address correspondence to Tricia Van Laar, tvanlaar@csustan.edu.

Abstract

Here we present the draft genome sequence of *Intestinibacter bartletti* DSM 16795, a human gut isolate of interest in autism-associated microbiome studies. The genome spans 2.97 Mb across 22 contigs, with a GC content of 28.84% and 100% estimated completeness. Key genomic features include virulence factors, antibiotic resistance genes, biosynthetic clusters, and CRISRP-Cas loci.

Announcement

Gut microbiota dysbiosis has been increasingly implicated in late-onset autism, particularly in individuals with elevated levels of *Clostridial species* compared to controls (1). One such bacterium of interest is *Intestinibacter bartlettii* DSM 16795, a Gram-positive, spore-forming, rod-shaped organism initially known as *Clostridium bartlettii*. It was first isolated in 2000 from human stool and gastric/duodenal samples at Rush Children's Hospital, Chicago (1, 2). The strain (WAL 16138, DSM 16795) was cultured anaerobically, transported on dry ice, and grown on Brucella and CDC blood agar at 37°C for 120 hours (1). It is stored at DSMZ and ATCC under accession number BAA-827 (1). The taxonomic identification was verified through 16S rRNA gene sequencing with accession number AY438672 (3). This strain is part of the JGI initiative to study gut microbiota in children with late-onset autism and its potential involvement in the condition (1, 4).

Details on organism growth and DNA isolation to be provided by DSMZ*.* Genomic DNA from *I. bartletti* DSM 16795 DNA was sequenced by the DOE Joint Genome Institute using Illumina HiSeq2000. A shotgun library generated 11,188,166 total reads (~1,678.2 Mb). Raw reads were quality-filtered using DUK software (5). Filtered reads were then assembled using Velvet v 1.2.07 with modified parameters to generate initial contigs. Wgsim generated 1-3 kb paired-end reads for improved assembly (6, 7). The final genome was assembled with Allpaths-LG v r46652 (8). The assembled genome contained 22 contigs across 19 scaffolds and totaled 2.9 Mb. The final assembly utilized 1,500 Mb of Illumina data and incorporated 300 bp input read coverage to enhance accuracy. The JGI Microbial Genome Annotation Pipeline was used to annotate the genome (9). Prodigal generated gene prediction, which were manually curated using GenePRIMP. For functional annotation, NCBI, UniProt, TIGFAM, pfam, KEGG, COG, and InterPro were utilized. The rRNA and tRNA genes present were identified with tRNAscan-SE and SILVA. Non-coding RNA was identified with INFERNAL. CheckM2 confirmed 100% completeness and 0% contamination (10). Potential contaminants were detected but not removed. This resulted in 89.4% of the initial dataset being retained (10).

The draft genome of *Intestinibacter bartlettii* DSM 16795 is 2,971,856 bp, 22 contigs, 28.84% GC, 2,878 genes (2,787 coding), 28 rRNAs, and 59 tRNAs (Table 1). The phylogenetic tree is midpoint rooted, 79.1% average branch support, and shows close relation to *Terrisporobacter (Figure1).* Genomic analysis of *I. bartletti* identified pathogenic features. Pathogen Finder reported an 87.5% probability of human pathogenicity (11). The virulence Factor Database (VFDB) reported 19 genes related to pathogenic genes (CarB) (Table 2) (12). Comprehensive

Antibiotic Resistance Database (CARD) v 3.2.4 identified antibiotic-resistant genes ErmQ (MLS resistance) and vanW, vanT, and vanY (glycopeptide resistance) (13). CRISPRCasFinder v 4.3.2 confirmed two CRISPR/Cas regions with strong evidence (two CRISPR, seven cas genes) and one weaker evidence CRISPR region (14). AntiSMASH (v8.0, relaxed settings) predicted three biosynthetic gene clusters: terpene precursor (region 9.1) and two ranthipeptides (regions 12.1 and 13.1) (15).

TABLE 1 Genomic Features of *Intestinibacter bartletti* DSM 16795

Feature	Finding
Length (bp)	2,971,856
Status	Draft genome
No. of contigs	22
GC content (%)	28.84%
No. of genes	2,875
No. of coding sequences	2,787
No. of rRNAs	28
No. of tRNAs	59

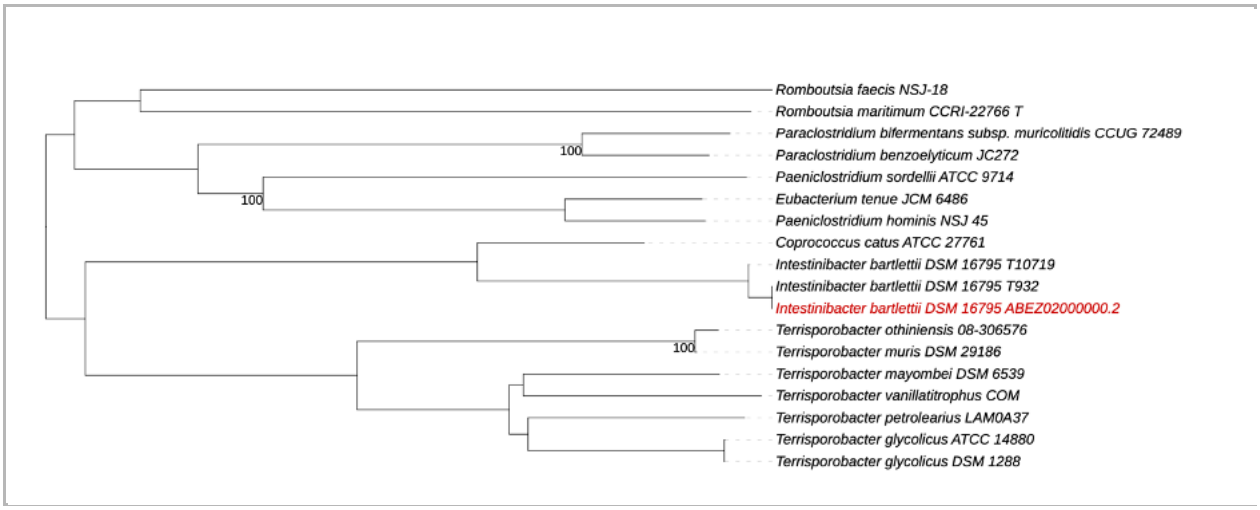

Figure1. Phylogenetic analysis of *Intestinibacter bartletti* DSM 16795 based on whole-genome comparisons. The genome BLAST distance phylogeny (GBDP) tree was constructed using the Type Strain Genome Server (TYGS) and inferred with FastME 2.1.6.. Branch lengths are scareled according to the GBDP distance formula d_5. Bootstraps values (>60%) from 100 replicates are indicated at the nodes, The tree was midpoint-rooted, with an average branch support of was 79.1%.

TABLE 2 Virulence factors found in the genome.

Gene	Description
espX6	Effector delivery system
tufA	Adherence
lap	Adherence
fbpA/fbp68	Adherence
clpC	Stress survival
bont	Exotoxin
capB	Immune modulation
cpsJ	Immune modulation

galE	Immune modulation
capA	Immune modulation
FTT_RS04140	Immune modulation
Cj1138	Immune modulation
Cj1136	Immune modulation
groEL	Adherence
htpB	Adherence
clpC	Stress survival
carB	Nutritional/Metabolic factor
toxB	Exotoxin
msbA	Immune modulation

Data availability statement

The genome sequence for *Intestinibacter bartletti* DSM 16795 is available in GenBank under accession number DS499552.1. Raw sequencing reads have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRX2066076. Further genomic data can be accessed through the Integrated Microbial Genomes with Microbiomes (IMG/M) system hosted by JGI using taxon ID 641736113.

Acknowledgments

We thank JGI for supporting student participation in the "Adopt-a-Genome" project and for including institutions within the California State University (CSU) system, which are not classified as R1 research universities. This opportunity allowed undergraduates to engage in genome analysis and scientific writing.

The work (proposal DOI[]:xxx) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under contract no. DE-AC02-05CH11231.

References

1. Finegold, S. M., Molitoris, D., Song, Y., Liu, C., Vaisanen, M. L., Bolte, E., McTeague, M., Sander, R., Wexler, H., Marlowe, E. M., Collins, M. D., Lawson, P. A., Summanen, P., Baysallar, M., Tomzyski, T. J., Read, E., Johnson, E., Rolfe, R., Nasir, P., Shah, H., Haake, D. A., Manning, P., & Kaul, A. (2002). *Clinical Infectious Diseases, 35*(Supplement_1), S6–S16. https://doi.org/10.1086/341914

2. Song, Y. L., Liu, C. X., McTeague, M., Summanen, P., & Finegold, S. M. (2004). *Anaerobe, 10*(3), 179–184. https://doi.org/10.1016/j.anaerobe.2004.06.004

3. Gerritsen, J., Fuentes, S., Grievink, W., van Niftrik, L., Tindall, B. J., Timmerman, H. M., Rijkers, G. T., & Smidt, H. (2014). *International Journal of Systematic and Evolutionary Microbiology, 64*(Pt_5), 1600–1616. https://doi.org/10.1099/ijs.0.059447-0

4. Kyrpides, N. C., Woyke, T., Eisen, J. A., Garrity, G., Lilburn, T. G., Beck, B. J., Whitman, W. B., Hugenholtz, P., & Klenk, H. P. (2014). *Standards in Genomic Sciences, 9*, 1278–1284. https://doi.org/10.1186/1944-3277-9-1278

5. Mingkun, L., Copeland, A., & Han, J. DUK. (2011). Unpublished.

6. Zerbino, D., & Birney, E. (2008). *Genome Research, 18*(5), 821–829. https://doi.org/10.1101/gr.074492.107

7. Li, H. wgsim. GitHub. https://github.com/lh3/wgsim

8. Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., & Jaffe, D. B. (2011). *Proceedings of the National Academy of Sciences, 108*(4), 1513–1518. https://doi.org/10.1073/pnas.1017351108

9. Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I. A., Pati, A., Nielsen, T., Markowitz, V. M., & Kyrpides, N. C. (2015). *Standards in Genomic Sciences, 10*, 86. https://doi.org/10.1186/s40793-015-0077-y

10. Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). *Nature Methods, 20*(8), 1203–1212. https://doi.org/10.1038/s41592-023-01987-0

11. Cosentino, S., Voldby Larsen, M., Møller Aarestrup, F., & Lund, O. (2013). *PLOS ONE, 8*(12). https://doi.org/10.1371/annotation/b84e1af7-c127-45c3-be22-76abd977600f

12. Liu, B., Zheng, D., Zhou, S., Chen, L., & Yang, J. (2022). *Nucleic Acids Research, 50*(D1), D912–D917. https://doi.org/10.1093/nar/gkab1107

13. Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z. C., Niu, K. C., Shan, P., Yang, J. Y. J., Gray, K. L., Hoad, G. R., Jia, B., Bhando, T., Carfrae, L. A., Farha, M. A., French, S., Gordzevich, R., Rachwalski, K., Tu, M. M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H. L., Brown, E. D., Maguire, F., Beiko, R. G., Hsiao, W. W. L., Brinkman, F. S. L., Van Domselaar, G., & McArthur, A. G. (2023). *Nucleic Acids Research, 51*(D1), D690–D699. https://doi.org/10.1093/nar/gkac920

14. Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). *Nucleic Acids Research, 46*(W1), W246–W251. https://doi.org/10.1093/nar/gky425

15. Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B. R., Metcalf, W. W., Helfrich, E. J. N., van Wezel, G. P., Medema, M. H., & Weber, T. (2023). *Nucleic Acids Research, 51*(W1), W46–W50. https://doi.org/10.1093/nar/gkad344