Faecalibacterium prausnitzii M21/2 Scfld_02_24, whole genome shotgun sequence

NCBI Reference Sequence: NZ_DS483503.1

FASTA Graphics

Go to:

LOCUS NZ_DS483503 3864 bp DNA linear CON 16-APR-2025
DEFINITION Faecalibacterium prausnitzii M21/2 Scfld_02_24, whole genome shotgun sequence.
ACCESSION NZ_DS483503 REGION: complement(295630..299493)
VERSION NZ_DS483503.1
DBLINK BioProject: PRJNA224116
BioSample: SAMN02299416
Assembly: GCF_000154385.1
KEYWORDS WGS; RefSeq.
SOURCE Faecalibacterium prausnitzii M21/2
ORGANISM Faecalibacterium prausnitzii M21/2
Bacteria; Bacillati; Bacillota; Clostridia; Eubacteriales; Oscillospiraceae; Faecalibacterium.
REFERENCE 1 (bases 1 to 3864)
AUTHORS Sudarsanam,P., Ley,R., Guruge,J., Turnbaugh,P.J., Mahowald,M., Liep,D. and Gordon,J.
TITLE Draft genome sequence of Faecalibacterium prausnitzii M21/2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3864)
AUTHORS Sudarsanam,P., Ley,R., Guruge,J., Turnbaugh,P.J., Mahowald,M., Liep,D. and Gordon,J.
TITLE Draft genome sequence of Faecalibacterium prausnitzii M21/2
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3864)
AUTHORS Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H., Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R. and Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 3864)
AUTHORS Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H., Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R. and Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 3864)

AUTHORS Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Mardis,E.R. and Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (12-JUL-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA

COMMENT REFSEQ INFORMATION: The reference sequence is identical to DS483503.1.
Faecalibacterium prausnitzii M21/2 (GenBank Accession Number for 16S rDNA gene: AY305307) is a member of the division Firmicutes. In one comprehensive 16S rDNA sequence-based enumeration of the colonic microbiota of three healthy adult humans, it represents, on average, 3.804% of all 16S rDNA sequences and 7.490% of the sequences in its division (Eckburg et. al. (2005)). The sequenced strain was obtained from the laboratory of Prof. Harry Flint, Rowett Research Institute, Aberdeen, Scotland, UK (H.Flint@rowett.ac.uk).

We have collected 8.4X coverage in plasmid end reads and 454 reads. We will be performing one round of automated sequence improvement (pre-finishing).

Sequencing/Assembly: The genomic DNA was purified from liquid culture derived from a single bacterial colony. A hybrid sequencing strategy that utilized reads from both 454 GS-20 and ABI 3730xl sequencers was devised and implemented to generate the draft genome sequences. 454 reads were assembled using Newbler (454 Life Sciences) into 454 de novo contigs. These de novo contigs were converted in silico to 800 base paired reads ('superreads') with 400 base overlaps with neighboring superreads. Finally, PCAP (Huang, et al, Genome Research, 13:2164, (2003)) was used to assemble the super-reads and the conventional 3730xl capillary reads.

This sequenced strain is part of a comprehensive, sequence-based survey of members of the normal human gut microbiota. A joint effort of the WU-GSC and the Center for Genome Sciences at Washington University School of Medicine, the purpose of this survey is to provide the general scientific community with a broad view of the gene content of 100 representatives of the major divisions represented in the intestine's microbial community. This information should provide a frame of reference for analyzing metagenomic studies of the human gut microbiome. Further details of this effort are described in a white paper entitled 'ExtendingOur

View of Self: the Human Gut Microbiome Initiative (HGMI)'

(http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/HGMISeq.pdf). These studies are supported by National Human Genome Research Institute.

For answers to your questions regarding this assembly or project, or any other GSC genome project, please visit our Genome Groups web page (http://genome.wustl.edu/genome_group_index.cgi) and email the designated contact person.
Faecalibacterium prausnitzii M21/2 (GenBank Accession Number for 16S rDNA gene: AY305307) is a member of the division Firmicutes. In one comprehensive 16S rDNA sequence-based enumeration of the colonic microbiota of three healthy adult humans, it represents, on average, 3.804% of all 16S rDNA sequences and 7.490% of the sequences in its division (Eckburg et. al. (2005)). The sequenced strain was obtained from the laboratory of Prof. Harry Flint, Rowett Research Institute, Aberdeen, Scotland, UK (H.Flint@rowett.ac.uk).

We have collected 8.4X coverage in plasmid end reads and 454 reads. We will be performing one round of automated sequence improvement (pre-finishing).

Sequencing/Assembly: The genomic DNA was purified from liquid culture derived from a single bacterial colony. A hybrid sequencing strategy that utilized reads from both 454 GS-20 and ABI 3730xl sequencers was devised and implemented to generate the draft genome sequences. 454 reads were assembled using Newbler (454 Life Sciences) into 454 de novo contigs. These de novo contigs were converted in silico to 800 base paired reads ('superreads') with 400 base overlaps with neighboring superreads. Finally, PCAP (Huang, et al, Genome Research, 13:2164, (2003)) was used to assemble the super-reads and the conventional 3730xl capillary reads.

This sequenced strain is part of a comprehensive, sequence-based survey of members of the normal human gut microbiota. A joint effort of the WU-GSC and the Center for Genome Sciences at Washington University School of Medicine, the purpose of this survey is to provide the general scientific community with a broad view of the gene content of 100 representatives of the major divisions represented in the intestine's microbial community. This information should provide a frame of reference for analyzing

metagenomic studies of the human gut microbiome. Further details of this effort are described in a white paper entitled 'ExtendingOur View of Self: the Human Gut Microbiome Initiative (HGMI)'

(http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/HGMISeq.pdf). These studies are supported by National Human Genome Research Institute.

Coding sequences were predicted using GeneMark v3.3 and Glimmer2 v2.13. Intergenic regions not spanned by GeneMark and Glimmer2 were blasted against NCBI's non-redundant (NR) database and predictions generated based on protein alignments. RNA genes were determined using tRNAscan-SE 1.23 or Rfam v8.0. Gene names are generated at the contig level and may not necessarily reflect any known order or orientation between contigs.These studies are supported by National Human Genome Research Institute.

For answers to your questions regarding this assembly or project, or any other GSC genome project, please visit our Genome Groups web page (http://genome.wustl.edu/genome_group_index.cgi) and email the designated contact person.

Annotation was added to the contigs in October 2007, and the CDS comments were updated in January 2008.

This is a reference genome for the Human Microbiome Project. This project is co-owned with the Human Microbiome Project DACC.
Product names were updated in August 2012.
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Annotation-Data-START##

Annotation Provider	:: NCBI RefSeq
Annotation Name	:: GCF_000154385.1-RS_2025_04_10
Annotation Date	:: 04/10/2025 08:38:18
Annotation Pipeline	:: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method	:: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision	:: 6.10
Features Annotated	:: Gene; CDS; rRNA; tRNA; ncRNA
Genes (total)	:: 3,029
CDSs (total)	:: 2,942

```
            Genes (coding)                 :: 2,900
            CDSs (with protein)             :: 2,900
            Genes (RNA)                    :: 87
            rRNAs                          :: 5, 6, 6 (5S, 16S, 23S)
            complete rRNAs                  :: 5, 6, 6 (5S, 16S, 23S)
            tRNAs                          :: 66
            ncRNAs                         :: 4
            Pseudo Genes (total)            :: 42
            CDSs (without protein)          :: 42
            Pseudo Genes (ambiguous residues) :: 0 of 42
            Pseudo Genes (frameshifted)     :: 10 of 42
            Pseudo Genes (incomplete)       :: 32 of 42
            Pseudo Genes (internal stop)    :: 3 of 42
            Pseudo Genes (multiple problems)  :: 2 of 42
            ##Genome-Annotation-Data-END##
FEATURES           Location/Qualifiers
     source         1..3864
                    /organism="Faecalibacterium prausnitzii M21/2"
                    /mol_type="genomic DNA"
                    /strain="M21/2"
                    /isolation_source="biological product [ENVO:02000043]"
                    /host="Homo sapiens"
                    /db_xref="taxon:411485"
     gene           1..3864
                    /locus_tag="FAEPRAM212_RS11485"
                    /old_locus_tag="FAEPRAM212_02796"
                    /db_xref="GeneID:75069035"
     CDS            1..3864
                    /locus_tag="FAEPRAM212_RS11485"
                    /old_locus_tag="FAEPRAM212_02796"
                    /EC_number="2.7.7.6"
                    /inference="COORDINATES: similar to AA
                    sequence:RefSeq:WP_005926187.1"
                    /GO_function="GO:0003677 - DNA binding [Evidence IEA];
                    GO:0003899 - DNA-directed 5'-3' RNA polymerase activity
                    [Evidence IEA]; GO:0032549 - ribonucleoside binding
                    [Evidence IEA]"
                    /GO_process="GO:0006351 - DNA-templated transcription
                    [Evidence IEA]"
                    /note="Derived by automated computational analysis using
                    gene prediction method: Protein Homology."
                    /codon_start=1
                    /transl_table=11
                    /product="DNA-directed RNA polymerase subunit beta"
```

/protein_id="WP_005926187.1"
/db_xref="GeneID:75069035"
/translation="MMKVKPVKLGKTERMSFSHIDEVISMPNLIEVQKNSYQWFLDEG
LKEVFHDIGTIEDYTGNLALSFVDFRLDKEPKYSIKECKERDVTFAAPLRVTARLLNK
ETGEVKDQEIFMGDFPLMTDAGTFVINGAERAIVSQLVRSPGVFYGHAKDKVGNDLYS
ATMNPNRGAWLEYETDAANVFYVRIDKNRKLPVTVLCRALGLSSNEDILNYFGEDERI
LATLEKDTTKNREEGLLEVYRKLRPGEPPTVESATSQIDMLFFDPRRYDLSRFGRYKM
NKKLSLARRIMDHVAAENVVDPFTGEILVEADKKIDRKLAEQIDAAGVNLVVLKIDDP
MKDQPHKVKVITNGCVDAQAIIDSYYPAFKGVDVKECGINERCCLKELRKILDNASSA
EEVMESLKKDHDLLIGRTVTIDDILSSINYLNGLGYGIGTTDDIDHLGNRRIRSVGEL
LQNQFRIGFSRMERVIRERMTLQSQDQSVITPQALINIRPVVAAIKEFFGSSPLSQFM
DQNNPLAELTHKRRLSALGPGGLSRDRAGFEVRDVHYSHYGRMCPIETPEGPNIGLIS
YLASYAKINEYGFVEAPYRKVKKIYDENNNLIEQVVTDEVEYMTADVEDEYVVAQANE
PLDEGKHFIRPRVSARRRDEILEIDAEKVDYMDVSPRMMVSVATACIPFLENDDCNRA
LMGSNMQRQAVPLMVTQQPIVATGMEYKAATDSGTAVLAKSNGVVEKVDADHVVVRNE
QGALEDYSLIKFARSNAGTCINQRPIVEVGETVTAGQVLADGPAMRNGEISLGKNALI
GFMTWEGYNYEDAVLLNEKIVREDVYTSIHIEEYETESRDTKLGPEEITRDIPNVSED
ALKDLDERGIIRIGAEVKSGDILVGKVTPKGETELTAEERLLRAIFGEKAREVRDTSL
RVPHGAYGIIVDVKVFTPENSDELQPGVREVVRCYIAQKRKISVGDKMAGRHGNKGVV
SRILPQEDMPYLPDGTPLDIVLNPLGVPSRMNIGQVLEVNLGYAAKACGIKVMTPVFD
SARENDIGDTFDTAREMWHGENAPAYPTKLPKIMGEKGHIIDFSKIELDRDGKTTVYD
GRTGEKFDNRVTVGYMYYLKLHHLVDDKIHARSTGPYSLVTQQPLGGKAQFGGQRFGE
MEVWALEAYGAAYTLQEILTVKSDDVEGRVKTYEAIVKGEPIPQPGIPESFRVMLKEL
QSLGMDVIVQDKDGNEIDMRQNFDDEETGFDMRDVAGTETVVQESELTDYNIKDADAG
FDDPSVLEDDNSGAEAPAASDEVDF"

ORIGIN

```
       1 atgatgaaag tcaagcccgt aaagctcggc aaaaccgagc gcatgagctt ctcccacatc
      61 gacgaagtca tcagtatgcc gaacctgatc gaggtgcaga agaactcgta tcagtggttc
     121 ctggatgaag gcctgaagga ggtcttccac gatatcggca ccattgagga ctacaccggc
     181 aatctggcac tgagctttgt ggacttccgg ctggataagg agccgaagta cagcatcaag
     241 gagtgcaagg agcgcgacgt tacctttgca gccccccctgc gcgtcaccgc ccgtctgctg
     301 aacaaggaga ccggcgaggt aaaggatcag gagatcttca tgggcgactt ccccctgatg
     361 accgacgcag gcacctttgt catcaacggt gctgagcgtg caatcgtcag ccagttggtc
     421 cgctctcccg gtgtgttcta cggccatgcg aaggataagg tcggcaacga cctgtacagc
     481 gccaccatga accccaaccg cggtgcatgg ctggagtacg agaccgatgc cgccaacgtg
     541 ttctatgtcc gtatcgataa gaaccgtaag ctgcccgtca ccgtgctgtg ccgtgcactg
     601 ggcctgtcct ctaacgagga tatcctgaac tacttcggcg aggacgagcg cattctggcc
     661 acgctggaaa aggataccac caagaaccgc gaggagggcc tgctggaggt ctaccgcaag
     721 ctgcgccccg gcgagcccc cacggtggag tccgctacca gccagatcga tatgctgttc
     781 ttcgatccgc gccggtacga cctgtcccgc ttcggccgct ataagatgaa taagaagctg
     841 tctctggcac gccgcatcat ggaccatgtg gccgcagaga acgtggtgga tcccttcacc
     901 ggcgagatcc ttgtcgaggc cgacaagaag atcgaccgca agctggctga gcagatcgat
     961 gctgccggcg tgaatctggt cgtgctgaag atcgatgatc ccatgaagga tcagccccac
    1021 aaggtcaagg tcatcaccaa cggctgcgtg gatgcacagg ccatcatcga cagctactac
```

1081 cccgcattca agggtgtgga tgtcaaggag tgcggcatca acgagcgctg ctgcctcaag
1141 gagctgcgca agatccttga taacgcctcc agcgccgagg aggtcatgga gagcctgaag
1201 aaggatcacg acctgctcat cggccgtacc gtcaccatcg acgatatcct gtcctccatc
1261 aactacctga acggtctggg ctacggcatc ggcaccaccg atgatatcga ccatctgggc
1321 aaccgccgta tccgcagcgt gggcgagctg ctgcagaacc agttccgcat cggtttctcc
1381 cgtatggagc gcgtcatccg tgagcgtatg accctgcaga gtcaggatca gagcgtcatc
1441 actccgcagg cactgatcaa catccgtcct gtggtggcag ccatcaagga gttcttcggc
1501 tcctctccgc tgtctcagtt catggatcag aataaccctc tggctgagct gacccacaag
1561 cgccgtctgt ctgctctggg ccccggcggt ctgagccgtg accgcgcagg cttcgaagtc
1621 cgcgacgttc attacagcca ctatggccgt atgtgcccca ttgagacccc tgaaggcccc
1681 aacatcggtc tgatctccta tctggcatcc tacgccaaga tcaacgagta cggcttcgtg
1741 gaggctccct accgcaaggt caagaagatc tacgacgaga acaacaatct catcgagcag
1801 gtcgtcaccg acgaggtgga gtatatgacc gccgatgtgg aggacgagta cgttgtggca
1861 caggccaacg agccgctgga tgaaggcaag cacttcatcc gtccccgtgt gtctgcccgc
1921 cgccgcgacg agatcttgga aatcgacgca gagaaggtcg attacatgga cgtttctccg
1981 cgaatgatgg tctctgttgc taccgcttgc atccccttcc tggagaacga tgactgtaac
2041 cgtgcactga tgggctctaa catgcagcgt caggcagtgc ctctgatggt cacccagcag
2101 cccattgttg ctaccggtat ggagtacaag gctgctaccg attccggcac tgctgtgctg
2161 gccaagagca acggcgttgt ggaaaaggtc gatgccgacc atgttgtggt gcgcaacgag
2221 cagggcgctc tggaggatta ttctctgatc aagtttgccc gctccaacgc aggcacctgc
2281 atcaaccagc gccccatcgt agaagtgggc gagaccgtca ccgccggtca ggtgctggcc
2341 gatggccccg caatgcgcaa cggcgagatc tctctgggta agaacgccct gatcggcttc
2401 atgacctggg agggctacaa ctacgaggac gccgtcctgc tgaacgagaa gatcgtgcgt
2461 gaggacgtgt acacctccat ccacatcgag gagtacgaga ccgagagccg cgataccaag
2521 ctgggacccg aagaaatcac ccgcgacatc cccaacgttt ctgaggatgc cctgaaggat
2581 ctggacgagc gcggcatcat ccgcatcggt gccgaggtca agagcggcga tattctggtc
2641 ggtaaggtca ccccgaaggg cgagactgag ctgactgctg aggagcgcct gctgcgcgcc
2701 atcttcggtg agaaggcccg cgaggtgcgc gacacctctc tgcgtgtgcc ccacggcgca
2761 tacggcatca ttgtggatgt caaggtcttt accccggaga acagcgacga gctgcagccc
2821 ggtgtgcgcg aggtcgtccg ttgctatatc gcccagaagc gtaagatcag cgtcggcgat
2881 aagatggcag gccgtcacgg taacaagggt gtcgtttccc gcatcctgcc gcaggaggat
2941 atgccctatc tgcccgacgg cactccgctg gacatcgtgc tgaacccgct gggcgtgcct
3001 tcccgtatga atatcggtca ggtgctggaa gtcaacctcg gctacgctgc caaggcctgc
3061 ggcatcaagg tcatgacccc tgtctttgac tctgcccgtg agaacgacat cggcgatacc
3121 tttgataccg cccgcgagat gtggcacggc gagaatgctc ccgcatatcc caccaagctg
3181 cccaagatca tgggcgagaa gggccacatc atcgacttct ccaagatcga gctggatcgt
3241 gacggcaaga ccaccgttta cgatggccgt accggtgaaa agtttgataa ccgcgttacc
3301 gttggttata tgtactacct caagctgcat cacctggtcg acgataaaat ccacgcccgt
3361 tccaccggcc cctactctct ggtcactcag cagcctctgg gcggcaaggc tcagttcggc
3421 ggtcagcgct tcggcgagat ggaggtctgg gcactggaag cttacggcgc cgcttacacc
3481 ctgcaggaga tcctgactgt caagtccgat gatgtggagg gtcgtgtcaa gacctacgaa
3541 gccatcgtca agggcgagcc catcccgcag cccggcatcc ccgagtcctt ccgcgttatg
3601 ctgaaggaac tgcagtctct gggtatggac gtgatcgtgc aggataagga cggcaacgag
3661 atcgatatgc gccagaactt cgacgatgaa gagaccggct ttgatatgcg tgacgttgcc

```
    3721 ggtaccgaga ctgtggtgca ggagagcgaa ctgactgatt acaacatcaa ggacgccgat
    3781 gcaggtttcg atgatccttc tgtgctggag gacgataact ccggtgccga ggctccggct
    3841 gcttctgacg aagtagattt ctga
//
```