**Supplementary Data**

**Data D1.** The GenBank format of AcetoBase accession for *Tepidanaerobacter acetatoxydans* Re1. ACCESSION is the flag for NCBI LinkOut, where as, VERSION is the AcetoBase LinkIn flag.

LOCUS NN\_0000000445 1686 bp DNA BCT

DEFINITION Formate--tetrahydrofolate ligase (Formyltetrahydrofolate synthe

 tase) (FTHFS) [Tepidanaerobacter acetatoxydans]

ACCESSION   [CP002728.1](https://www.ncbi.nlm.nih.gov/nuccore/CP002728.1)

VERSION     [NN\_0000000445](https://acetobase.molbio.slu.se/seq/NN_0000000445)

KEYWORDS AcetoBase

SOURCE Tepidanaerobacter acetatoxydans

 ORGANISM Bacteria;Terrabacteria group;Firmicutes;Clostridia;Thermoanaero

 bacterales;Thermoanaerobacteraceae;Tepidanaerobacter;Tepidanaer

 obacter acetatoxydans;Re1

FEATURES Location/Qualifiers

 source 1..1686

 gene 329196..330881

 /gene = "FTHFS\_1"

 /orientation = "Sense Strand"

                     /db\_xref = taxon:[1209989](https://acetobase.molbio.slu.se/db_xref/1209989)
                     /WLP\_genes = [1209989](https://acetobase.molbio.slu.se/wlp_genes/1209989)
                     /16S ribosomal RNA gene = [CP002728.1](https://acetobase.molbio.slu.se/16S/CP002728.1)

 /protein\_id = "AEE90506.1"

 /translation = "MTFKSDIEIAQSVKLQDIREIAAKLGLTEDDIDLYGKY

 KAKVDYNLLNNCNGKKAKLILTTAITPTPAGEGKTTTTIGAADALTRLGKKTIV

 ALREPSLGPVFGIKGGAAGGGYAQVVPMEDINLHFTGDIHAITAANNLLAAMVD

 NHIFQGNQLNIDTRRVVWRRAVDINDRQLRFVIDGLGGKANGVPREDGFDITVA

 SEVMAIFCLANDIMDLKERLAKIVVAYDREGKPVTAGDLKAQGAMAALLKDALK

 PNLVQTLEGTPAFVHGGPFANIAHGCNSVIATKMAMHLADYVVTEAGFGADLGA

 EKFIDIKCRLAGLKPDAVIIVATVRALKYNGGLAKADLEKEDLNALAKGIPNLL

 KHVENITQVFKLPAVVAINRFPQDTEAELKLVEDRCNELGVNVALSEVWAKGGE

 GGIALAEELIRLTEDNKASNKGLAYVYELDMPIEEKIRAISQKIYGADDVMFTD

 KALKEIANLEKLGFGKMPVCIAKTQYSLTDDPKKLGRPSGFNITVRDVSVSAGA

 GFIVAVTGDIMKMPGLPKVPAAEKIDVDEKGVISGLF"

ORIGIN

 1 ATGACTTTTA AATCCGACAT TGAGATTGCT CAATCAGTGA AACTCCAAGA TATCAGAGAG

 61 ATTGCGGCAA AGCTAGGTCT TACCGAAGAC GATATTGATT TGTACGGCAA ATATAAAGCA

 121 AAGGTCGACT ATAATTTACT AAATAATTGT AATGGTAAAA AAGCAAAGTT AATTTTAACA

 181 ACTGCCATTA CTCCTACACC TGCAGGGGAA GGCAAAACTA CAACCACAAT CGGTGCTGCG

 241 GATGCACTTA CAAGACTTGG GAAAAAGACT ATTGTTGCAT TACGTGAGCC GTCATTGGGT

 301 CCTGTTTTTG GAATTAAAGG TGGTGCGGCC GGGGGCGGAT ATGCACAGGT TGTTCCTATG

 361 GAAGATATAA ACCTGCACTT TACCGGGGAT ATTCACGCTA TTACGGCTGC CAACAATCTA

 421 TTAGCCGCCA TGGTGGATAA TCATATTTTC CAGGGGAATC AATTAAATAT CGACACTAGG

 481 CGTGTTGTTT GGAGACGTGC TGTTGATATA AATGATAGGC AATTAAGATT TGTCATTGAT

 541 GGTCTTGGAG GAAAAGCCAA TGGGGTGCCA AGAGAGGACG GCTTTGATAT TACTGTTGCT

 601 TCGGAAGTAA TGGCTATTTT CTGTCTTGCT AACGATATTA TGGATTTGAA AGAACGGCTG

 661 GCGAAGATTG TTGTAGCATA TGATCGGGAG GGTAAGCCCG TTACTGCGGG AGATTTAAAA

 721 GCACAAGGTG CCATGGCGGC TCTACTCAAA GATGCTTTAA AACCCAATCT GGTACAAACA

 781 TTAGAAGGTA CTCCTGCTTT CGTGCATGGT GGTCCCTTTG CAAATATAGC ACACGGATGT

 841 AACTCGGTAA TTGCAACTAA AATGGCAATG CATCTTGCTG ATTATGTGGT AACGGAGGCC

 901 GGCTTCGGCG CGGATTTAGG TGCTGAGAAA TTTATAGATA TAAAATGCCG ATTGGCTGGG

 961 CTAAAACCTG ATGCGGTGAT TATAGTTGCA ACGGTTAGAG CCCTCAAATA CAATGGAGGT

 1021 TTGGCAAAGG CTGACTTGGA GAAGGAAGAC TTGAATGCAC TAGCCAAGGG TATTCCGAAT

 1081 CTATTAAAAC ATGTAGAAAA CATCACACAA GTATTCAAAC TTCCTGCTGT TGTAGCAATA

 1141 AACAGATTTC CTCAGGACAC CGAAGCCGAA TTGAAACTGG TTGAAGACAG ATGTAATGAG

 1201 CTTGGGGTGA ATGTAGCATT ATCTGAAGTA TGGGCAAAAG GCGGGGAAGG CGGAATAGCC

 1261 CTTGCGGAAG AATTAATAAG GCTAACAGAA GACAATAAGG CAAGTAATAA GGGGTTGGCG

 1321 TATGTTTATG AGCTAGATAT GCCAATAGAA GAAAAAATAC GTGCAATCTC GCAAAAAATA

 1381 TATGGTGCAG ATGATGTGAT GTTTACAGAT AAGGCATTAA AGGAAATTGC GAATCTTGAA

 1441 AAACTCGGTT TTGGTAAAAT GCCGGTTTGT ATAGCAAAAA CCCAGTACTC ACTGACTGAC

 1501 GATCCTAAAA AATTAGGTAG GCCGTCAGGA TTCAATATTA CCGTAAGGGA TGTTTCAGTA

 1561 TCGGCTGGAG CCGGATTTAT AGTGGCTGTT ACAGGCGATA TTATGAAGAT GCCGGGCTTG

 1621 CCTAAGGTAC CTGCAGCAGA AAAAATTGAT GTGGATGAAA AGGGTGTTAT TTCAGGGCTA

 1681 TTCTAA

//

**Data D2.** Downloading the fasta format of the AcetoBase accessions

# Downloading the AcetoBase accession in fasta format on command line interface

$ curl -o outputfile.fasta <https://acetobase.molbio.slu.se/seq/download/NN_0000000445>

Or

$ wget -O outputfile.fasta <https://acetobase.molbio.slu.se/seq/download/NN_0000000445>

# Visualizing first 10 lines of Downloaded file

$ head outputfile.fasta

>NN\_0000000445 Formate--tetrahydrofolate ligase (Formyltetrahydrofolate synthetase) (FTHFS) [Tepidanaerobacter acetatoxydans]

ATGACTTTTAAATCCGACATTGAGATTGCTCAATCAGTGAAACTCCAAGATATCAGAGAGATTGCGGCAAAGCTAGGTCT

TACCGAAGACGATATTGATTTGTACGGCAAATATAAAGCAAAGGTCGACTATAATTTACTAAATAATTGTAATGGTAAAA

AAGCAAAGTTAATTTTAACAACTGCCATTACTCCTACACCTGCAGGGGAAGGCAAAACTACAACCACAATCGGTGCTGCG

GATGCACTTACAAGACTTGGGAAAAAGACTATTGTTGCATTACGTGAGCCGTCATTGGGTCCTGTTTTTGGAATTAAAGG

TGGTGCGGCCGGGGGCGGATATGCACAGGTTGTTCCTATGGAAGATATAAACCTGCACTTTACCGGGGATATTCACGCTA

TTACGGCTGCCAACAATCTATTAGCCGCCATGGTGGATAATCATATTTTCCAGGGGAATCAATTAAATATCGACACTAGG

CGTGTTGTTTGGAGACGTGCTGTTGATATAAATGATAGGCAATTAAGATTTGTCATTGATGGTCTTGGAGGAAAAGCCAA

TGGGGTGCCAAGAGAGGACGGCTTTGATATTACTGTTGCTTCGGAAGTAATGGCTATTTTCTGTCTTGCTAACGATATTA

TGGATTTGAAAGAACGGCTGGCGAAGATTGTTGTAGCATATGATCGGGAGGGTAAGCCCGTTACTGCGGGAGATTTAAAA

**Data D3.** The whole genomic metadata information can be accessed on command line interface as follows:

# Downloading the data and directing it to outputfile in text format

$ curl <https://acetobase.molbio.slu.se/db_xref/1209989> > outputfile.txt

Or

$ wget <https://acetobase.molbio.slu.se/db_xref/1209989> > outputfile.txt

# Visualizing the first 20 lines of outputfile.txt

$ head -20 outputfile.txt

===>taxon: 1209989 ===> taxon\_oid 650716094

Domain:Bacteria

Sequencing Status:Finished

Study Name:DOEM\_782014

Genome Name / Sample Name:Tepidanaerobacter sp. Re1

Sequencing Center:DOE Joint Genome Institute (JGI)

IMG Genome ID :650716094

Phylum:Firmicutes

Class:Clostridia

Order:Thermoanaerobacterales

Family:Thermoanaerobacteraceae

Genus:Tepidanaerobacter

Species:Tepidanaerobacter acetatoxydans

IMG Product Assignment:NA

IMG Release:IMG/W 3.5

IMG Submission ID:0

JGI Project ID / ITS SP ID:402559

JGI Analysis Product Name:NA

JGI Analysis Project Type:Genome Analysis

**Data D4.** From the command line interface information regarding Wood-Ljungdahl pathway can be accessed as follows:

# Downloading the data and directing it to outputfile in text format

$ curl <https://acetobase.molbio.slu.se/wlp_genes/1209989> > outputfile.txt

# Visualizing the first 10 lines of outputfile.txt

$ head outputfile.txt

enzyme: acetate kinase

taxid: 1209989

EC=2.7.2.1\_ack

seqname: CP002728.1

source: Genbank

feature: CDS

start: 1484110

stop: 1485312

score: .

strand: -

**Data D5.** Downloading 16S rRNA gene associated with the AcetoBase accession

# Downloading the 16s rRNA gene fasta sequence

$ curl -o 16s.fasta <https://acetobase.molbio.slu.se/16S/CP002728.1>

Or

$ wget -O 16s.fasta <https://acetobase.molbio.slu.se/16S/CP002728.1>

# Visualizing the fasta header of all 16s rRNA gene sequences

$ grep -e ">" 16s.fasta

>accession|CP002728 start|250097|stop|251608

>accession|CP002728 start|1216885|stop|1218396