Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

16s rRNA sequence analysis and identification of nif genes in a new strain of *Rhodopseudomonas* sps, as a potential candidate for increasing efficiency of biogas production

Aysha Sherieff[1], M.Mohibbe Azam[2], K.Sesha Maheswaramma[3]

[1]*Research Scholar, JNTUA, Anantapur, Andhra Pradesh, India*
[2]*Indian Institute of Rice Research (ICAR-IIRR), Rajendra Nagar, Hyderabad, Telangana State, India*
[3]*Department of Chemistry, JNTUACEP, Pulivendula, Kadapa, Andhra Pradesh, India*

A new strain of *Rhodopseudomonas* sps with partial 16s rRNA sequence was performed previously using sequencing methods and further analysed using Bioinformatics tools which is being reported in the present paper. GC profile content of the sequence, phylogenetic tree analysis and functional analysis of 16s rRNA sequence was performed using various tools like RDP classifier, BLAST, FASTA, Expasy-Translate, ORF Finder, EMBOSS Seqret, CLUSTALW, Addgene, InterPro scan and CLUSTAL Omega. The insilico studies have revealed its close relation to *Rhodopseudomonas faecalis* and also supporting evidences that the strain has genes for photosynthetic activity and consists of nif genes for Nitrogenase activity. Several researchers reported the presence of nitrogen rich biomass as raw material and have yielded high methane production. This shows the presence of nif genes in *Rhodopseudomonas faecalis* increases the efficiency of biogas production and the insilico research work proposes to utilise the bacterium as a potential candidate.

Keywords: Insilico, rRNA, *Rhodopseudomonas* sps, Biogas, Nitroginase, nif genes, BLAST, FASTA, CLUSTALW.

1. Introduction

The microbiome structure in the Biogas production through Anaerobic Digestion (AD) process is having more scope to analyse and understand. The increase in organic waste have opened avenues for this process to explore and deploy technologies for efficient Biogas production. A complex consortium of bacteria and archaea performs 4 reactions viz., hydrolysis, acidogenesis, acetogenesis, and methanogenesis in an AD process (Angenent LT et al., 2004

& Hassa J et al., 2018). The microbial communities in AD are extremely complex and the microbial compositions, their interactions among microbes remain largely unclear (Narihiro T et al., 2015).

Purple non-sulfur phototrophic bacteria grow in wide range of environmental conditions and most of the genus belong to *Rhodopseudomonas*. They are able to grow anaerobically in the light or aerobically in the dark with different carbon sources and electron donors. In Bergey's Manual of Systematic Bacteriology, the genus *Rhodopseudomonas* included seven species: *Rhodopseudomonas palustris, Rhodopseudomonas blastica, Rhodopseudomonas viridis, Rhodopseudomonas sulfoviridis, Rhodopseudomonas acidophila, Rhodopseudomonas rutila* and *Rhodopseudomonas marina* (Imhoff & Trusper, 1989).Later *Rhodopseudomonas faecalis* was isolated and characterised from anaerobic reactor that digests chicken faeces (Demin Zhang et al., 2002).

Based on high-throughput sequencing of culture-independent technologies have enabled the deep investigation of microbial compositions and functions. The taxonomic profile of AD microbial communities was analysed frequently using high-throughput 16S ribosomal RNA gene sequencing (De Vrieze J et al.,2018 & Mei R et al.,2017). Metagenomic approaches alone or besides with meta-proteomics, metabolomics and meta-transcriptomics are applied to decipher the gene functions, enzyme profiles, and metabolic processes of microbial communities in AD (Jia Y et al.,2018 and Treu L et al., 2016).

Previously, a new strain of *Rhodopseudomonas* sps was isolated that is responsible for producing maximum biogas in a digester. In this study, the sequence analysis using bioinformatics tools of the 16s rRNA of the bacterial organism was found to reveal that the it is is closely related to *Rhodopseudomonas faecalis* and have characteristics compatible with respect to parameters of the biogas digester thereby proving it as a potential candidate for increased biogas production.

2. Methodology

Source of the sequence

The organism was isolated from the samples of the Sewage Treatment Plant (STP) of Association of Lady Entrepreneurs of India (ALEAP) Industrial Park in Hyderabad of Telangana State, India and the pure cultures were screened and identified as *Rhodopseuodmonas* sps based upon the morphological and biochemical methods at Centre for Environment, Institute of Science and Technology (IST), Jawaharlal Nehru Technological University(JNTU), Hyderabad, Telangana State, India. The organism was sequenced by 16s rRNA methodology , partial sequence was obtained in FASTA format.

GC profile and Classification of the Strain

The GC percentage of the 16s rRNA sequence was known using the tool GC profile (Feng Gao and Chun-Ting Zhang, 2006), a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. The classification of the sequence was obtained by Ribosomal Database Project (RDP) Classifier (Cole JR et al., 2014). It provides quality-

controlled, aligned and annotated bacterial and archaeal 16S rRNA sequences. A nearest neighbor sequence matching based upon the most number of shared k-mers between a query and reference sequence was analysed using Sequence Match RDP tool (cole JR et al., 2005).

Phylogenetic Tree Analysis

BLAST was run (McGinnis S and Madden TL, 2004) for both nucleotide and protein sequences of *Rhodopseudomonas* sps. The same tool was also used for generating phylogenetic tree. The protein sequence was obtained from the Expasy-Translate tool (Gasteiger E et al.,2003) that performs translation of a nucleotide sequence in to a protein sequence. The protein sequence was converted in to FASTA format using the EMBOSS Seqret (Hamish McWilliam et al.,2013). Using ORF Viewer (Wheeler DL et al.,2003), ORF's will be determined, it finds Open Reading Frames (ORFs) from the query nucleotide sequence. The program returns the range of each ORF along with its protein translation. The ORF finder searches newly sequenced nucleotide for potential protein encoding segments. A CLUSTALW (Thompson et al.,1994) program was run to know the multiple sequence alignment and analysis of the similarity of the 16S rRNA gene.

Functional analysis of the 16s rRNA sequence

The Circular and Linear structures of the nucleotide along with features and restriction sites containing restriction enzyme sites were obtained using Addgene (Kamens J , 2015). It analyses the sequence with maps, features and translate the nucleotide sequence in to an aminoacid sequence. The InterPro Scan (Quevillon E et al.,2005) provides functional analysis of proteins by classifying them into families and predicting domains and important sites. CLUSTAL Omega was run (Sievers F et al.,2007) to generate fast, scalable and high quality protein multiple sequence alignments.

3. Discussion and Results

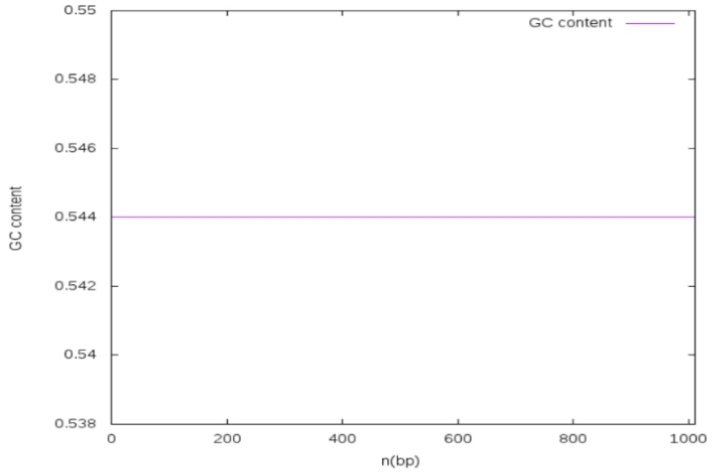
The FASTA sequence of the Anaerobic bacteria, a Purple Non Sulphur Bacteria (PNSB) was named with genus *Rhodopseudomonas* sps based upon morphological and biochemical studies. The sequence was published for first time in this Journal (Fig 1).

```
>PNSB
TAACGCGTGGGAACGTACCTTTTGGTTCGGAACAACTGAGGGAAACTTCAGCTAATACCGGATAAGCCCTTACGGGGAAAGATTTATCGC
CGAAAGATCGGCCCGCGTCTGATTAGCTAGTTGGTGGGGTAATGGCCCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGC
CACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCC
GCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCGT
GCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGTGCGTAGGCGGGTTTCTAAGTCAGAGGTG
AAAGCCTGGAGCTCAACTCCAGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGCAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAA
TTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGGCCATTACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATGCCAGCCGTTAGTGGGTTTACTCACTAGTGGCGCAGCTAACGCTTTAAGCAT
TCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGC
AACGCGCAGAACCTTACCAGCCCTTGACATGTCCAGGACCGGTCGCAGAGATGTGACCTTCTCTTCGGAGCCTGGAGCACAGGTGCTGCA
TGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCCGTCCTTAGTTGCTACCATTTAGTTGAGCACTCTAAGGAGACTGCCGGTGA
```

Fig 1 : 16s rRNA sequence of the Purple Non Sulphur Bacteria (PNSB),*Rhodopseudomonas* sps

Halting parameter = 100.00 Filtered gap size = 10 bp Minimum length = 3000 bp

Start (bp)	Stop (bp)	Length (bp)	GC content (%)
1	1011	1011	54.40



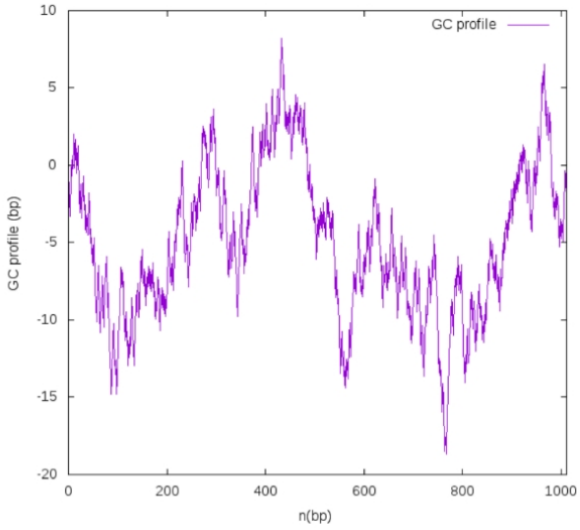


Fig 2 : GC profile of the sequence of *Rhodopseudomonas* sps

Fig2 specifies the GC content in the sequence with 1011 bp and with 54.40% in total. The GC content of a gene region can impact its coverage, with regions having 50–60% GC content receiving the highest coverage while regions with high (70–80%) or low (30–40%) GC content having significantly decreased coverage. Yet, adequate coverage of GC-rich regions (which are commonly present in the promoter and first exon of many genes) is necessary for a high analytical sensitivity of a targeted gene panel (Sami S. Amr and Birgit Funke, 2015). Evidence of GC ratio with that of length of the coding region of a gene has shown that the length of the coding sequence is directly proportional to higher G+C content (Pozzoli U et al.,2008).

The RDP Classifier shows that the sequence belongs to the following classification (Fig 3)

Domain : Bacteria

Phylum : Proteobacteria

Class : Alphaproteobacteria

Order: Rhizobiales

Family : Bradyrhizobiaceae

Genus : *Rhodopseudomonas*

It reveals that the organism belongs to *Rhodopseudomonas* sps.

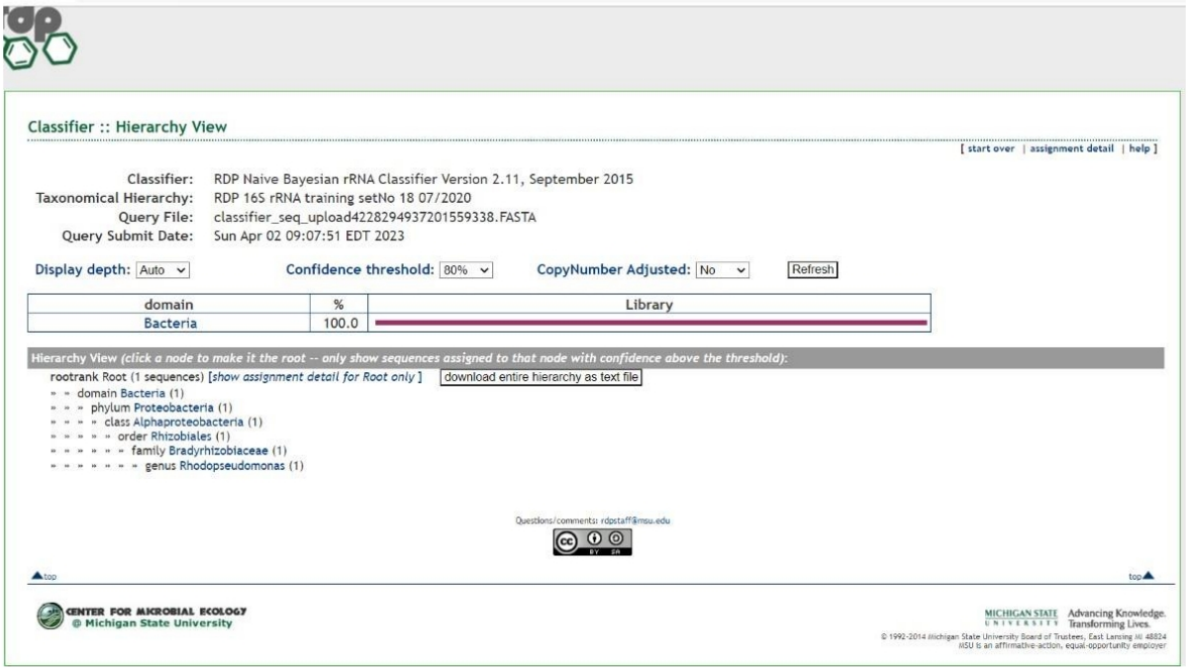


Fig 3 : RDP Classifier showing the classification (Hierarchy view) of the query sequence

Further RDP is performed with Seq Match to find out the nearest neighbors of the query sequence by picking up the database sequences that have the highest numbers of shared 7-mers ("word") with the specific query sequence. A S_ab score that is the percentage of shared words between compared sequences were obtained in the results (Fig 4).

Out of total 20 sequences ,8 sequences were belongs to *Rhodopseudomonas* sps and 12 are

Non Purple Sulphur Bacteria. A maximum of 1.000 S_ab score was obtained for 6 *Rhodopseudomonas faecalis* and minimum of 0.976 S_ab score was obtained for *Rhodopseudomonas palustris*, where in it can be conferred that the query sequence is closet to *Rhodopseudomonas faecalis*.

S.NO	Name of the Organism	Short ID of RDP	S_ab Score
1	*Rhodopseudomonas faecalis*; HR; HQ154127	S002233990	1.000
2	*Rhodopseudomonas faecalis*; DBNR4-1; KF668620	S004056606	1.000
3	*Rhodopseudomonas faecalis*; DBNR4-2; KF668621	S004056607	1.000
4	*Rhodopseudomonas faecalis*; DBNR4-3; KF668622	S004056608	1.000
5	*Rhodopseudomonas faecalis*; DBNR4-4; KF668623	S004056609	1.000
6	*Rhodopseudomonas faecalis*; DBNRh16; KJ776421	S004221214	1.000

Table 1: Seq Match of the query sequence with 16srRNA database RDP showing ID and S_ab Scores

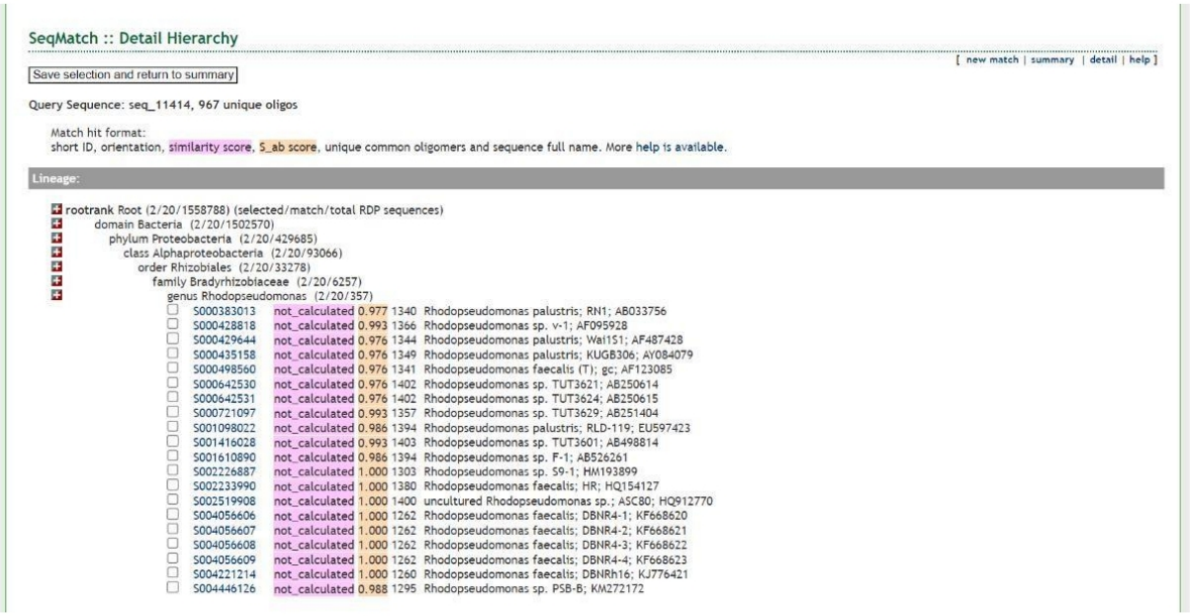


Fig 4 : RDP Seq Match results

Table 2 shows that there are 100 sequences primarily were found to have percent similarity from a maximum of 100% to a minimum of 99.21 % . The query sequence have shown 100% Percent Identity with *Rhodopseudomonas faecalis* and besides secondly, Percent Identity was shown close to *Rhodopseudomonas palustris* (Table 3).

Mostly the query strain's sequence shows Percent Identity in the ascending order with *Rhodopseudomonas faecalis* (100%, 99.90%, 99.70%, 99.41%,99.60%, 99.51%), *Rhodopseudomonas palustris* (99.60%, 99.51%, 99.41%, 99.31%, 99.21%), *Rhodopseudomonas* sp (100%, 99.90%, 99.80%, 99.60%, 99.51%, 99.41%, 99.21%) , *Rhodobacter sphaeroides* (99.60%), uncultured alpha proteobacterium(99.41%),

Rhodopseudomonas Pentothenatexigens (99.31%), *Rhodopseudomonas oryzae* (99.21%) *Rhodopseudomonas thermotolerans* (99.21%).

S.NO	Percent Identity	Number of Sequences
1	99.21	53
2	99.31	2
3	99.41	4
4	99.51	7
5	99.60	19
6	99.70	1
7	99.80	2
8	99.90	4
9	100	8

Table 2: Percent Identity of the query sequence of *Rhodoseudomonas* sps against BLASTN

S.NO	Name of the Organism	Accession Number	Percent Identity(%)
1	*Rhodopseudomonas faecalis* strain DBNRh33 16S ribosomal RNA gene, partial Sequence	KT180200.1	100
2	*Rhodopseudomonas palustris* gene for 16S rRNA, partial sequence, strain:KP0014	AB167545.1	99.60

Table 3: Showing the sequences having nearest Percent Identity value for the query sequence of *Rhosopseudomonas* sps

From Fig 5 it can be concluded that the organism is near to *Rhodopseudomonas faecalis* and also the same was conferred from the Table 1, 2 & 3 that the new strain of the organism belongs to the same genus and species.

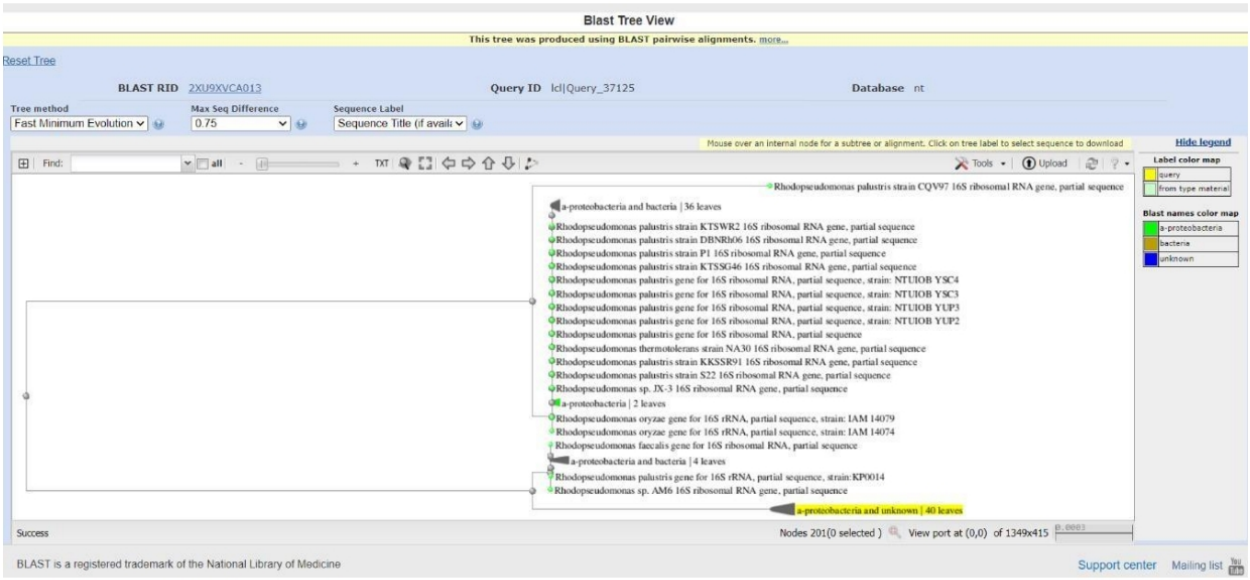


Fig 5 : BLASTN Tree view of the query sequence of *Rhodoseudomonas* sps

The protein sequence of the *Rhodopseudomonas* sps was converted in to the FASTA format using EMBOSS Seqret showing 1754bp in length of aminoacids.

>seq_3168 1754 bp

RKIGPRLISLVGWPTKATISSWSERMISHIGTETRPKLLREAAVGNIGQWAQASSHAA
VMKALGLSSFVREDNDGTARISPGLRASSRGN
TKGASVARNHWARVRRRVSKSEVKAWSSTPELPLILEVLSMAEVSGTASVEVKFVD
IRKNTSGEGGSLGHYRGTKAWGANRIRYPGSPRR

KRMPAVSGFTHWRSRFKHSAWGVRSQDNSKELTGARTSGGACGLIRRNAQNLTSP
HVQDRSQRCDLLFGAWSTGAAWLSSARVVRCWVKS

RNERNPRPLLPFSALGDCRERSARVLASWWGNGPPRRRSVAGLRGSATLGLRHGPN
SYGRQQWGILDNGRKPDPAMPRERPGCKALLCGK

IMTVPQEAPANFVPAAAVIRRGLALLGITGRKGCVGGFLSQRKPGAQLQNCLYWKS
VWQRVELRVRNSIFARTPVAKAAHWAITDAEARK

RGEQTGLDTLVVHAVNDECQPLVGLLTSGAANALSIPPGEYGRKIKTQRNRGPAQA
VEHVVFDATRRTLPALDMSRTGRRDVTFSSEPGA

QVLHGCRQLVSDVGLSPATSATPVLSCYHLVEHSKETAGKDRPASDLVGGVMAQG
DDQLVEDDQPHWDDTAQTPTGGSSGEYWTMGASLI

QPCRVSDEGPRVVKLFCAGRRYRKNKPRLTSCQQPRYEGGRCSESLGVKGAAGFVR
GESLELNSRTAFDTGSLEYGRGEWNCECRGEIRR

YSQEHQWRRRLTGPLLTLRHESVGSKQDIPWSTPTMNASRWVYSLVAQLTLAFRLG
STVARLKLKGIDGGPHKRWSMWFNSTQRAEPYQP

LTCPGPVAEMPSLRSLEHRCCMAVVSSCREMLGVPQRAQPPSLVATILSTLRRLPVS
PAVSLECSTKWQLRTGVALVAGLNPTSHDTSRQ

PCSTCAPGSEEKVTSLRPVLDMSRAGKVLRVASNTTCSTACAGPRQFLVLILRPYSP
GGMLKALAAPLVSKPTNGWHSSFTAWTTRVSNP

VCSPRFRASASVMAQAAFATGVLANIYEFHLYTRSSTHLCHTQDFQYQRQFWSAPG
FHLLRNPPTHPLRPVIPSNASPLRITAAAGTKLA

GAYSCGTVIIFPHKRALQPGLHHSRGMAGSGLRPLSNIPHCCLPEFGPCLSPNVADHP
LRPATDRRLGGPLPHQLANQTRADLSHRQSPS

AQLNGSNGRGLRSLRDLTQHLTTRADDSHAAPVLQAPKRRSHLCDRSWTCQGLVR
FCALRRIKPHAPPLVRAPVNSFEFSCDRTPQAECL

KRLRHVNPLTAGIHRLRRGLPGYLILFAPHAFVPQRQWPSEPPSPLVFLRISTNFTSTL
AVPLTSAILKTSSIKGSSGVELQAFTSDLET

RLRTLYAQFRATLAPFVLPRLLARSPGLILAVPSLSSRTKELYNPRAFITHAAWLDQA
CAHCPIFPTAASRRSLGRVSVPMWLIILSDQL

LIVALVGHYPTNLIRRGPIFTGSLLRVLNMVATKDGGCARCGTPNISRHELTTAMQH
LCSRLRREGHISATGPGHVKGWGSARCVELNHM

LHRLCGPPSIPLSFNLATVLPRRNASVSCATSETHRLAFIVYGVDYQGISCLLPTLSCL
SVSNGPVSRLRHWCSCEYLRISPLHSQFHSP

LPYSRLPVSKAVLELSSRLSPLTKPAYAPFTPSDSEQRPPSYYRGCWHEVSRGLFLRY
RHYLPAQKSFTTLGPSSLTRHGWIRLAPIVQYSPLLPPVGVWAVSQSQCGSSSQTSYSS
PWWAITPPTSSDAGRSF

Fig 6: FASTA format of the protein sequence of the *Rhodopseudomonas* sps

In Fig 7, the nucleotide query sequence was run to obtain a protein sequence with Three frames in 5` 3` and 3` 5` directions and the ORF's were shown in red colour.

Fig 7 : Protein sequence of the *Rhodopseudomonas* sps

From Fig 8 it can be concluded that the organism is related to the protein sequences of the genus like g-proteobacteria, Glucanobacter , Lupinus, Lactobacillus, Streptomyces etc.

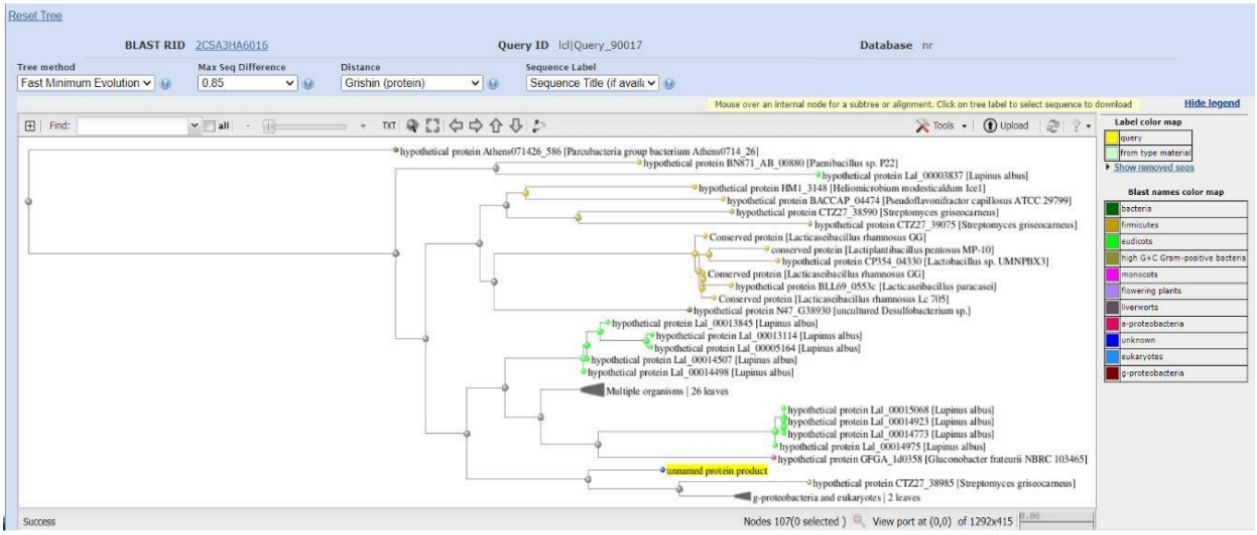


Fig 8 : BLASTP Tree view of the protein sequence of *Rhodoseudomonas* sps

The National Laboratory of Medicine ORF Viewer shows 7ORFs for the Nucleotide sequence

of *Rhodopseudomonas* sps having 1101bp length with 210 aminoacids and 651 nucleotide length (Fig 9). The complete ORF protein sequence reads were also shown in Table 4.

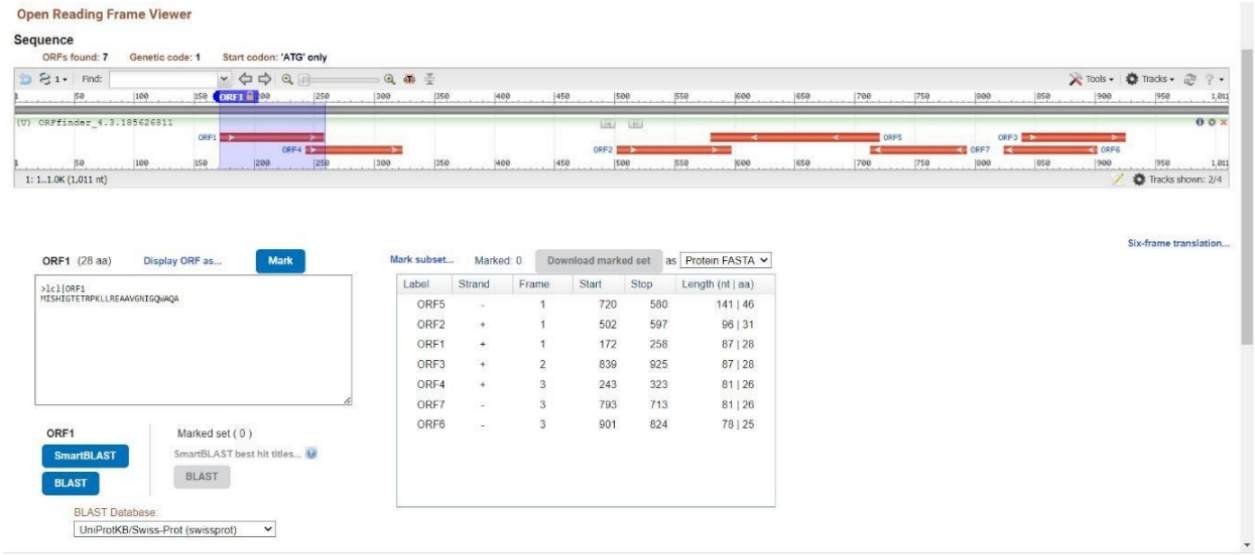


Fig 9 : ORF viewer of NLM database search showing 7 ORF of the *Rhodopseudomonas* sps

S.NO	ORF Number	ORF Sequence read
1	ORF 1	>lcl|ORF1 MISHIGTETRPKLLREAAVGNIGQWAQA
2	ORF2	>lcl|ORF2 MAEVSGTASVEVKFVDIRKNTSGEGGSLGHY
3	ORF3	>lcl|ORF3 MSRTGRRDVTFSSEPGAQVLHGCRQLVS
4	ORF4	>lcl|ORF4 MGASLIQPCRVSDEGPRVVKLFCAGR
5	ORF5	>lcl|ORF5 MLKALAAPLVSKPTNGWHSSFTAWTTRVSNPVCSPRFRASASVMAQ
6	ORF6	>lcl|ORF6 MQHLCSRLRREGHISATGPGHVKGW
7	ORF7	>lcl|ORF7 MLHRLCGPPSIPLSFNLATVLPRRNA

Table 4 : ORF Protein Sequences of the *Rhodopseudomonas* sps

A CLUSTAL W program was run against the sequences generated from RDP and BLASTN along with an out group genus. The results shows that the query sequence is having maximum similarity with *Rhodopseudomonas faecalis* (Fig 10 & 11).

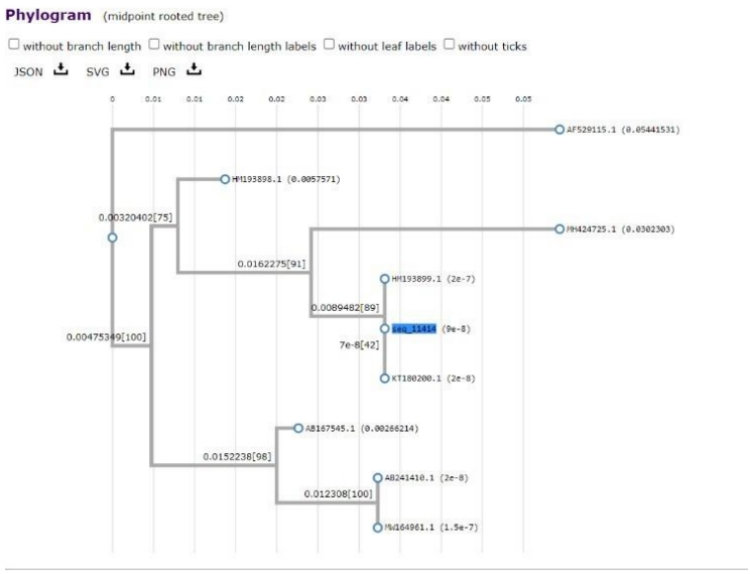


Fig 10 : Phylogram using CLUSTALW for BLASTN generated sequences

Legend:

>AF529115.1 Uncultured alpha proteobacterium clone FTLM116 16S ribosomal RNA gene,partial sequence

>HM193898.1 *Rhodobacter sphaeroides* strain S10-1 16S ribosomal RNA gene, partialsequence

>MH424725.1 *Rhodopseudomonas pentothenatexigens* strain Y7 16S ribosomal RNA gene,partial sequence

>HM193899.1 *Rhodopseudomonas* sp. S9-1 16S ribosomal RNA gene, partial sequence

>seq_11414 1011 bp (Blue color) Query sequence

>KT180200.1 *Rhodopseudomonas faecalis* strain DBNRh33 16S ribosomal RNA gene, partialsequence

>AB167545.1 *Rhodopseudomonas palustris* gene for 16S rRNA, partial sequence,strain:KP0014

>AB241410.1 *Rhodopseudomonas oryzae* gene for 16S rRNA, partial sequence, strain: IAM14079

>MW164961.1 *Rhodopseudomonas thermotolerans* strain NA30 16S ribosomal RNA gene,partial sequence

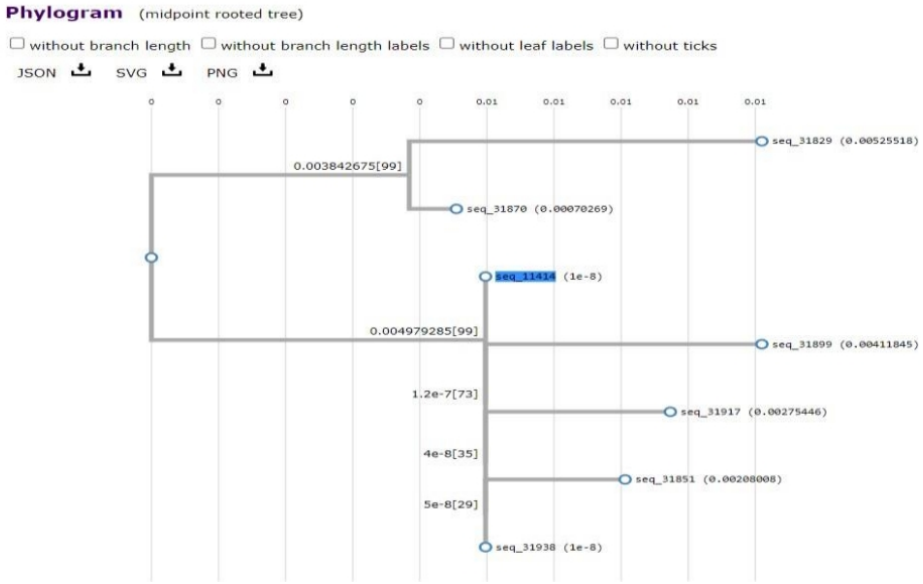


Fig 11: Phylogram using CLUSTALW for RDP generated sequences

Legend:

>seq_31829 1420 bp S000383013 *Rhodopseudomonas palustris*; RN1; AB033756

>seq_31870 1425 bp S000429644 *Rhodopseudomonas palustris*; Wai1S1; AF487428

>seq_11414 1011 bp (Blue color) Query sequence

>seq_31899 1482 bp S001098022 *Rhodopseudomonas palustris*; RLD-119; EU597423

>seq_31917 1462 bp S002233990 *Rhodopseudomonas faecalis*; HR; HQ154127

>seq_31851 1446 bp S000428818 *Rhodopseudomonas* sp. v-1; AF095928

>seq_31938 1331 bp S004056606 *Rhodopseudomonas faecalis*; DBNR4-1; KF668620

The Circular and Linear maps of the sequence of *Rhodopseudomonas* sps was obtained (Fig 12) where in, the restriction enzymes along with cuts , source of the organism and recognition sequences was also shown in Table 5. The similar sequence cuts and restriction enzymes were found in pCMV 104 plasmid of *Rhodopseudomonas palustris* (Kaberniuk AA et al., 2016).

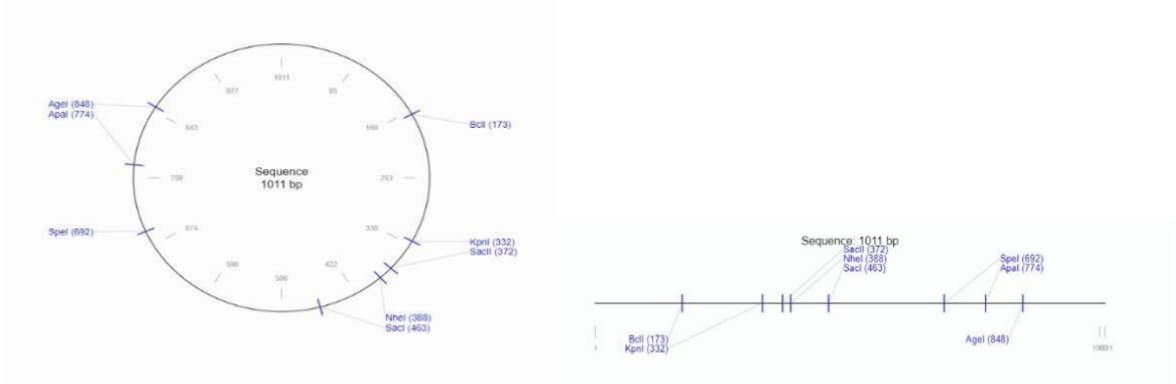


Fig 12: The Circular and Linear map of the sequence of the *Rhodopseudomonas* sps

S.NO	Enzyme Name	Cut	Source of the Organism	Recognition sequence
1	Bcl I	173	*Bacillus caldolyticus*	T/GATCA
2	Kpn I	332	*Klebsiella pneumoniae*	5′-GGTAC/C-3′
3	Sac II	372	*Streptomyces achromogenes*	CCGC/GG
4	Nhe I	388	*Neisseria mucosa heidelbergensis*	G/CTAGC
5	Sac I	463	*Streptomyces achromogenes*	GAGCT/C
6	Spe I	692	*Sphaerotilus species*	A/CTAGT
7	Apa I	774	*Acetobacter pasteurianus*	5′-GGGCC/C-3′
8	Age I	848	*Agrobacterium gelatinovorum*	A/CCGGT

Table 5 : The total number of enzymes along with cuts were identified from the Sequence map

Fig 13 reveals that the protein sequence is having chloroplastic 30s Ribosomal protein S12 which is a confirmation that the organism is a photosynthetic one. The prokaryotic small ribosomal subunit or 30S subunit is the smaller subunit of the 70S ribosome. It is a complex of 16S rRNA and 19 proteins.

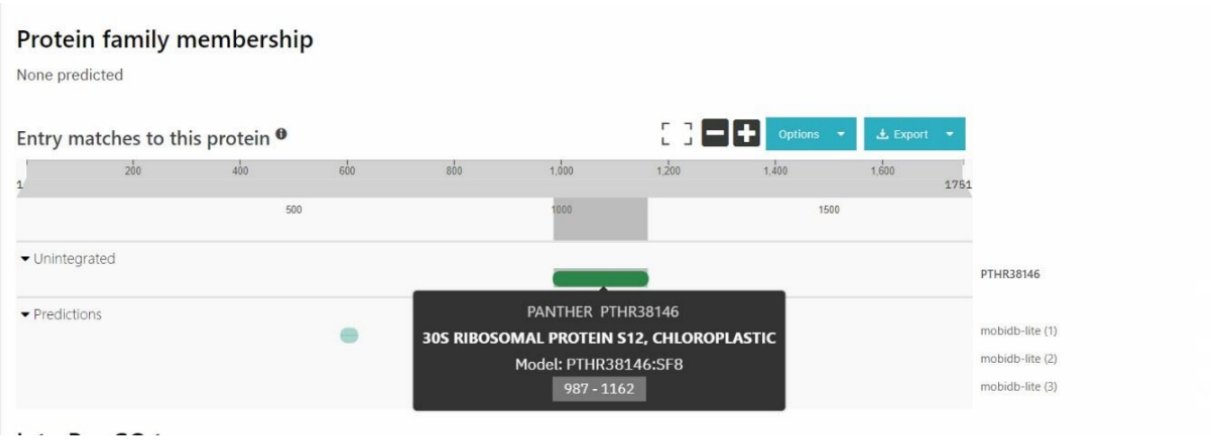


Fig 13 : Conserved Domain Identification using InterPro Scan of protein sequence of *Rhodoseudomonas* sps

Fig 14 shows results of CLUSTAL Omega, it was run for the query sequence against nucleotide sequences of nif genes having different species of *Rhodopseudomonas* along with an outgroup member. The similarity was found to be more near to *Rhodopseudomonas faecalis* , nif H gene for reductase of nitrogenase. Whereas Fig 15 shows that CLUSTAL Omega of protein query sequence against the proteins of nif genes of *Rhodopseudomonas* sps. It shows maximum similarity with nif-specific transcriptional activator nif A of *Rhodopseudomonas palustris* and with Nif-specific ferredoxin III of *Rhodopseudomonas faecalis*. The strains of *Rhodopseudomonas palustris*, PS3 and YSC3 have been found with nif nitrogenase-related genes (Lo, KJ et al., 2018).

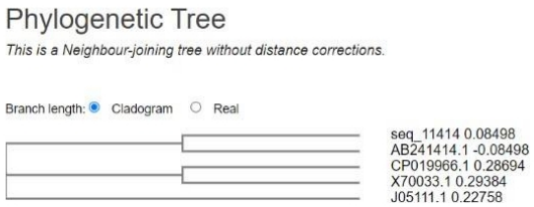


Fig 14: CLUSTAL Omega of nucleotide sequences of nif genes of different genus and species

Legend:

>seq_11414 1011 bp (Query sequence)

>AB241414.1 *Rhodopseudomonas faecalis* nifH gene for reductase of nitrogenase, partial cds,strain: JCM 11668

>CP019966.1 *Rhodopseudomonas palustris* strain PS3 chromosome, complete genome

>X70033.1 R.capsulatus anfH, anfD, anfG and anfK genes for alternative nitrogenase

>J05111.1 Anabaena PCC7120 nitrogenase (nifB), ferrodoxin-like protein (fdxN), nifS (nifS),nifU (nifU), and nitrogenase reductase (nifH) genes, complete cds

Fig 15 : CLUSTAL Omega of protein sequences of nif genes of different genus and species

Legend:

>seq_3168 1754 bp (Query Sequence)

>TAH65830.1 MAG: nif-specific transcriptional activator NifA [*Rhodopseudomonas* palustris]

>PYF04189.1 Nif-specific ferredoxin III [*Rhodopseudomonas faecalis*]

>WP_027278297.1 ferredoxin III, nif-specific [*Rhodopseudomonas faecalis*]

>TAH65812.1 MAG: ferredoxin III, nif-specific [*Rhodopseudomonas palustris*]

In a study done on obtaining 1,817 Gb metagenomic data, derived from digestate samples of 56 full-scale biogas plants fed with diverse feedstocks, the construction of microbial gene catalog of AD (22,840,185 genes) have revealed that the anaerobic microbes are having only species that belongs to methanobacter and others (Shichun Ma et al., 2021).

In another study it was demonstrated that from the complex nitrogen sources such as yeast extract and casamino acids, it was showed that the highest methane production of 600 ml methane per mole of nitrogen was reported, whereas by the use of skim milk as a source there

was no methane production observed (Andreas Otto Wagner et al.,2012).

In one research, a comparative study has been developed between pig manure and leguminous plant biomass at different Organic Load Rate values. It was found that most elevated methane yield was reported with nitrogen-rich biomass. Methane-yield results of nitrogen-rich biomass are higher than pig manure results for all OLR studied values (Isabel et al., 2022).

4. Conclusion

The partial 16s rRNA sequence was analysed with phylogenetic analysis and other tools proving that the strain was closet to genus and species of *Rhodopseudomonas faecalis*. The GC content of the sequence has 1011 bp and it contributes 54.40% of the total sequence base pairs. The RDP database shows that the query sequence has 100% Percent Identity with *Rhodopseudomonas faecalis* strains. BLAST N results shows that the organism is having high similarity of 100% with *Rhodopseudomonas faecalis*, EMBOSS Seqret results shows the organism has 1754bp of aminoacids. ORF Viewer shows 7ORFs for the Nucleotide sequence of *Rhodopseudomonas* sps and consists of 210 aminoacids and a total of 651 nucleotide length.

When CLUSTALW was run for RDP and BLASTN results, it reveals that the query sequence was 100% similar to *Rhodopseudomonas faecalis*. The Circular and Linear maps of the sequence of *Rhodopseudomonas* sps was also obtained. Conserved Domain Identification using InterPro Scan of protein sequence of Rhodoseudomonas sps. It shows that the protein sequence is having chloroplastic 30s Ribosomal protein S12, a confirmation that the organism is a photosynthetic one. When CLUSTAL Omega was run for *Rhodopseudomonas* sps sequence against nucleotide and protein sequences of nif genes having different species of *Rhodopseudomonas* genus. The results show the maximum similarity to nif H gene for reductase of nitrogenase and nif-specific ferredoxin III of *Rhodopseudomonas faecalis* respectively.

The organism is the novel strain of *Rhodopseudomonas faecalis* with nif genes, protein sequences for Nitrogenease activity and exhibits photosynthetic activity. This demonstrates that this novel strain of *Rhodopseudomonas faecalis* have potential to fix the nitrogen through nitrogenase activity and helps in increasing the methane production.

References

1. Andreas Otto Wagner, Peter Hohlbrugger, Philipp Lins, Paul Illmer, Effects of different nitrogen sources on the biogas production – a lab-scale investigation, Microbiological Research, Volume 167, Issue 10, 2012, Pages 630-636, ISSN 0944-5013, https://doi.org/10.1016/j.micres.2011.11.007.
2. Angenent LT, Karim K, Al-Dahhan MH, et al. Production of bioenergy and biochemicals from industrial and agricultural wastewater. Trends Biotechnol. 2004;22:477–85.
3. Bingham, A. H. A. et al. (1978) Nucleic Acids Res. 5, 3457
4. Demin Zhang, Huifang Yang, Zhiyong Huang, Wei Zhang and Shuang-Jiang Liu, "*Rhodopseudomonas faecalis* sp. nov., a phototrophic bacterium isolated from an anaerobic reactor that digests chicken faeces", International Journal of Systematic and Evolutionary Microbiology (2002), 52, 2055–2060.

5. De Vrieze J, Saunders AM, He Y, et al. Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. Water Res. 2015;75:312–23.
6. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D294-6. doi: 10.1093/nar/gki038. PMID: 15608200; PMCID: PMC539992.
7. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014 Jan;42(Database issue):D633-42. doi: 10.1093/nar/gkt1244. Epub 2013 Nov 27. PMID: 24288368; PMCID: PMC3965039.
8. Feng Gao, Chun-Ting Zhang, GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences, Nucleic Acids Research, Volume 34, Issue suppl_2, 1 July 2006, Pages W686–W691, https://doi.org/10.1093/nar/gkl040
9. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003 Jul 1;31(13):3784-8. doi: 10.1093/nar/gkg563. PMID: 12824418; PMCID: PMC168970.
10. Hamish McWilliam, Weizhong Li, Mahmut Uludag, Silvano Squizzato, Young Mi Park, Nicola Buso, Andrew Peter Cowley, Rodrigo Lopez, Analysis Tool Web Services from the EMBL-EBI, Nucleic Acids Research, Volume 41, Issue W1, 1 July 2013, Pages W597–W600, https://doi.org/10.1093/nar/gkt376.
11. Hassa J, Maus I, Off S, et al. Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions and functional relationships of microbial communities residing in biogas plants. Appl Microbiol Biotechnol. 2018;102:5045–63.
12. Imhoff, J. F. & Tru$per, H. G. (1989). Purple nonsulfur bacteria. In Bergey's Manual of Systematic Bacteriology, vol. 3, pp. 1658–1682. Edited by J. T. Staley, M. P. Bryant, N. Pfennig & J. G. Holt. Baltimore: Williams & Wilkins.
13. Isabel, P.A.A.; Luis, R.B.; Juan, C.P.; Jerónimo, G.C. Biogas from Nitrogen-Rich Biomass as an Alternative to Animal Manure Co-Substrate in Anaerobic Co-Digestion Processes. Energies 2022, 15, 5978. https://doi.org/10.3390/en15165978.
14. Jia Y, Ng SK, Lu H, et al. Genome-centric metatranscriptomes and ecological roles of the active microbial populations during cellulosic biomass anaerobic digestion. Biotechnol Biofuels. 2018;11:117.
15. Kaberniuk AA, Shemetov AA, Verkhusha VV. Nat Methods.,A bacterial phytochrome-based optogenetic system controllable with near-infrared light. 2016 May 9. doi: 10.1038/nmeth.3864. 10.1038/nmeth.3864 PubMed 27159085.
16. Kamens J. The Addgene repository: an international nonprofit plasmid and data resource. Nucleic Acids Res. 2015 Jan;43(Database issue):D1152-7. doi: 10.1093/nar/gku893. Epub 2014 Nov 11. PMID: 25392412; PMCID: PMC4384007.
17. Lo, KJ., Lin, SS., Lu, CW. et al. Whole-genome sequencing and comparative analysis of two plant-associated strains of *Rhodopseudomonas palustris* (PS3 and YSC3). Sci Rep 8, 12769 (2018). https://doi.org/10.1038/s41598-018-31128-8
18. Luo G, Fotidis IA, Angelidaki I. Comparative analysis of taxonomic, functional, and metabolic patterns of microbiomes from 14 full-scale biogas reactors by metagenomic sequencing and radioisotopic analysis. Biotechnol Biofuels. 2016;9:51.
19. Mei R, Nobu MK, Narihiro T, et al. Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. Water Res. 2017;124:77–84.
20. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W20-5. doi: 10.1093/nar/gkh435. PMID: 15215342; PMCID: PMC441573.
21. Narihiro T, Nobu MK, Kim NK, et al. The nexus of syntrophy-associated microbiota in

anaerobic digestion revealed by long-term enrichment and community survey. Environ Microbiol. 2015;17:1707–20.

22. Pozzoli U, Menozzi G, Fumagalli M, et al. (2008). "Both selective and neutral processes drive GC content evolution in the human genome". BMC Evol. Biol. 8: 99.
23. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W116-20. doi: 10.1093/nar/gki442. PMID: 15980438; PMCID: PMC1160203.
24. Sami S. Amr, Birgit Funke, Targeted Hybrid Capture for Inherited Disease Panels, Clinical Genomics, 2015.
25. Schnürer A. Biogas production: Microbiology and technology. Adv Biochem Eng Biotechnol. 2016;156:195–234.
26. Shichun Ma, Fan Jiang, Yan Huang, Yan Zhang, Sen Wang, Hui Fan, Bo Liu, Qiang Li, Lijuan Yin, Hengchao Wang, Hangwei Liu, Yuwei Ren, Shuqu Li, Lei Cheng, Wei Fan, Yu Deng, A microbial gene catalog of anaerobic digestion from full-scale biogas plants, GigaScience, Volume 10, Issue 1, January 2021, giaa164, https://doi.org/10.1093/gigascience/giaa164.
27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct 11;7:539. doi: 10.1038/msb.2011.75. PMID: 21988835; PMCID: PMC3261699.
28. Stolze Y, Bremges A, Rumming M, et al. Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. Biotechnol Biofuels. 2016;9:156.
29. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680.
30. Treu L, Kougias PG, Campanaro S, et al. Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. Bioresour Technol. 2016;216:260–6.
31. Tyagi VK, Lo SL. Sludge: A waste or renewable source for energy and resources recovery?. Renew Sust Energ Rev. 2013;25:708–28.
32. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 2003 Jan 1;31(1):28-33. doi: 10.1093/nar/gkg033. PMID: 12519941; PMCID: PMC165480.