

# Data Sequence untuk BLAST Search

**MK Bioinformatika ATA 2023/2024**

**KELAS C**

# Identitas Mahasiswa

**Nama: Bimo Hadi Permana**

**NPM: 2306157103**

**Kelas: C Bioinformatika**

**Absensi: 20**

# Tugas Individu

1. Pilih data unknown sequence.
2. Kirim data sequence ke NCBI BLAST Search Tool.
3. **Buat beberapa opsi** dengan click: “**Exclude**” dan “**Limit to**”
4. Buatlah Hasil Identifikasi berdasarkan hasil BLAST Search sesuai dengan beberapa opsi yang dipilih (Butir no.3). Buatlah Kesimpulan apakah terdapat perbedaan hasil identifikasi?
5. Buatlah Laporan Hasil identifikasi *Unknown sequence* tersebut dalam bentuk PPT, tuliskan Nama mahasiswa, NPM, dan beri keterangan kelas Bioinformatika. Beri nama file dengan: **No. Absen\_Kelas\_Nama Lengkap\_Tugas 1.**
6. Upload file di tempat yang telah disediakan di EMAS.

# FORMAT HASIL IDENTIFIKASI QUERY SEQUENCE

1. QUERY sequence name: \_\_\_\_\_

2. Search Set: \_\_\_\_\_

3. BLAST Search Result:

- Query length: \_\_\_\_\_ bp
- Gen: \_\_\_\_\_
- Top Hit: \_\_\_\_\_ , dengan persentase identitas \_\_\_\_\_%.
- Max Score: \_\_\_\_; Total Score: \_\_\_\_: Query cover: \_\_\_\_%; E value: \_\_\_\_\_
- Organisme: \_\_\_\_\_.

*Nama organisme dan persentase identitas (%)*

# SEQUENCE 1

## Pencarian sequence dengan BLASTN

>Sequence-1

```
AGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGGAAAGGCCCTTCGGGGGTACTCGAGCGGCGAACGGGTGAGTAACA
CGTGAGCAACCTGCCCCTGACTCTGGGATAAGCCTGGGAAACCGGGTCTAATACCGGATATGACCCTTCATCGCATGGTGTGAGGGTGGAAA
GTTTTTCGGTTGGGGATGGGCTCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACGGGTAAACGGGCCTGAGAGGG
CGACCGGTCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGCGGAAGCCTGACGC
AGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGACGAAGCTAACGTGACGGTACCTGCAGAAGAAGCG
CCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGT
GCGTCCGTCGTGAAAGCCACGGCTTAACCTGTGGGTCTGCGGTGGATACGGGCAGACTAGAGGCAGGTAGGGGAGCATGGAATTCCCGGT
GTAGCGGTGAAATGCGCAGATATCGGGAGGAACACCGGTGGCGAAGGCGGTGCTCTGGGCCTGTACTGACGCTGAGGAGCGAAAGCGTGG
GGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGCGCTAGGTGTGGGGTCCTTCCACGGATTCCGCGCCCGTAGCTA
ACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAAGCGGCGGAGCATGTT
GCTTAATTCGACGCAACGCGAAGAACCTTACCAAGGCTTGACATCGCCGGAAAACCTCGCAGAGATGCGGGGGTCCTTTTGGGCGCGGTGACAG
GTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTCGTTCCATGTTGCCAGCACGTAGTG
GTGGGGACTCATGGGAGACCGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGTCTTGGGCTGCAAAC
ATGCTACAATGGCCGGTACAGAGGGCTGCGATACCGTGAGGTGGAGCGAATCCCTTAAAGCCGGTCTCAGTTCGGATCGAAGTCTGCAACTC
GACTTCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCCGTCACGTCA
CGAAAGTCGGCAACACCCGAAGCCCGTGGCCCAACCCTTGTGGGGGGAGCGGTCTGAAGGTGGGGCCGGCGATTGGGACGAAGTCGTAA
CAAGGTAGCCGTACCGGAAG GTGC
```

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. QUERY sequence name: Sequence-1
2. Search Set: 16S ribosomal RNA seq; sequences from type material ; and exclude: models (XM/XP), uncultured/environmental sample sequences
3. BLAST Search Result:
  - Query length: 1474 bp
  - Gen: 16S ribosomal RNA
  - Top Hit: Actinomadura vinacea strain NBRC 14688 16S ribosomal RNA, partial sequence, dengan persentase identitas 98,85 %.
  - Max Score: 2628; Total Score: 2628; Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Actinomycetota; Actinomycetes; Streptosporangiales; Thermomonosporaceae; Actinomadura; Actinomadura vinacea strain NBRC 14688.

*Actinomadura vinacea strain NBRC 14688  
dengan persentase identitas 98,85 %.*

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. QUERY sequence name: Sequence-1
2. Search Set: 16S ribosomal RNA seq; sequences from type material ; and exclude: -
3. BLAST Search Result:
  - Query length: 1474 bp
  - Gen: 16S ribosomal RNA
  - Top Hit: *Actinomadura vinacea* strain NBRC 14688 16S ribosomal RNA, partial sequence, dengan persentase identitas 98,85 %.
  - Max Score: 2628; Total Score: 2628: Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Actinomycetota; Actinomycetes; Streptosporangiales; Thermomonosporaceae; *Actinomadura*; *Actinomadura vinacea* strain NBRC 14688 .

*Actinomadura vinacea* strain NBRC 14688  
dengan persentase identitas 98,85 %.

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. QUERY sequence name: Sequence-1
2. Search Set: 16S ribosomal RNA seq; tanpa type material, exclude: -
3. BLAST Search Result:
  - Query length: 1474 bp
  - Gen: 16S ribosomal RNA
  - Top Hit: *Actinomadura* sp. HAT23 16S ribosomal RNA gene, partial sequence, dengan persentase identitas 100 %.
  - Max Score: 2723; Total Score: 2723; Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Actinomycetota; Actinomycetes; Streptosporangiales; Thermomonosporaceae; *Actinomadura*; *Actinomadura* sp. HAT23.

*Actinomadura* sp. HAT23  
dengan persentase identitas 100%.



# HASIL IDENTIFIKASI QUERY SEQUENCE

BLAST® » blastn suite » results for RID-VXGYAVNT013

Home Recent Results Saved Strategies Help

[← Edit Search](#) [Save Search](#) [Search Summary ▾](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title **Sequence-1**

RID [VXGYAVNT013](#) Search expires on 02-27 14:33 pm [Download All ▾](#)

Program BLASTN [Citation ▾](#)

Database rRNA\_typestrains/16S\_ribosomal\_RNA [See details ▾](#)

Query ID lc|Query\_1614411

Description None

Molecule type dna

Query Length 1474

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
✓	<a href="#">Actinomadura vinacea strain NBRC 14688 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura vinacea</a>	2628	2628	100%	0.0	98.85%	1471	<a href="#">NR_112737.1</a>
✓	<a href="#">Actinomadura montaniterrae strain CYP1-1B 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura montaniterrae</a>	2590	2590	100%	0.0	98.37%	1488	<a href="#">NR_151939.1</a>
✓	<a href="#">Actinomadura barringtoniae strain GKU 128 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura barringtoniae</a>	2586	2586	100%	0.0	98.31%	1497	<a href="#">NR_178624.1</a>
✓	<a href="#">Actinomadura jiaoheensis strain NEAU-Jh1-3 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura jiaoheensis</a>	2584	2584	100%	0.0	98.31%	1511	<a href="#">NR_146004.1</a>
✓	<a href="#">Actinomadura sporangiiformans strain NEAU-Jh2-5 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura sporangiiformans</a>	2584	2584	100%	0.0	98.31%	1511	<a href="#">NR_146003.1</a>
✓	<a href="#">Actinomadura geliboluensis strain A8036 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura geliboluensis</a>	2575	2575	100%	0.0	98.17%	1490	<a href="#">NR_109059.1</a>

**i** Your search is limited to records that include: sequences from type material ; and exclude: models (XM/XP), uncultured/environmental sample sequences

Job Title **Sequence-1**

RID [VXGXR6SJ016](#) Search expires on 02-27 14:33 pm [Download All ▾](#)

Program BLASTN [Citation ▾](#)

Database rRNA\_typestrains/16S\_ribosomal\_RNA [See details ▾](#)

Query ID lc|Query\_7073719

Description None

Molecule type dna

Query Length 1474

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
✓	<a href="#">Actinomadura vinacea strain NBRC 14688 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2628	2628	100%	0.0	98.85%	1471	<a href="#">NR_112737.1</a>
✓	<a href="#">Actinomadura montaniterrae strain CYP1-1B 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2590	2590	100%	0.0	98.37%	1488	<a href="#">NR_151939.1</a>
✓	<a href="#">Actinomadura barringtoniae strain GKU 128 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2586	2586	100%	0.0	98.31%	1497	<a href="#">NR_178624.1</a>
✓	<a href="#">Actinomadura jiaoheensis strain NEAU-Jh1-3 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2584	2584	100%	0.0	98.31%	1511	<a href="#">NR_146004.1</a>
✓	<a href="#">Actinomadura sporangiiformans strain NEAU-Jh2-5 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2584	2584	100%	0.0	98.31%	1511	<a href="#">NR_146003.1</a>
✓	<a href="#">Actinomadura geliboluensis strain A8036 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadur...</a>	2575	2575	100%	0.0	98.17%	1490	<a href="#">NR_109059.1</a>

Kiri: Tanpa filter

Kanan: Filter exclude (seluruh) dan sequence from type material

# HASIL IDENTIFIKASI QUERY SEQUENCE

BLAST® » blastn suite » results for RID-W63PPFPF016

Home Recent Results Saved Strategies Help

[← Edit Search](#) [Save Search](#) [Search Summary ▼](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title **Nucleotide Sequence**

RID [W63PPFPF016](#) Search expires on 03-02 20:42 pm [Download All ▼](#)

Program BLASTN [Citation ▼](#)

Database core\_nt [See details ▼](#)

Query ID lclQuery\_7069109

Description None

Molecule type dna

Query Length 1474

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#) [back](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	<a href="#">Actinomadura sp. HAT23 16S ribosomal RNA gene, partial sequence</a>	<a href="#">Actinomadura sp. HAT23</a>	2723	2723	100%	0.0	100.00%	1474	<a href="#">KT210129.1</a>
✓	<a href="#">Actinomadura sp. 21ATH 16S ribosomal RNA gene, partial sequence</a>	<a href="#">Actinomadura sp. 21ATH</a>	2717	2717	100%	0.0	99.93%	1485	<a href="#">KT876994.1</a>
✓	<a href="#">Actinomadura rugatobispora JCM 3366 DNA, complete genome</a>	<a href="#">Actinomadura rugatobispora</a>	2660	13290	100%	0.0	99.25%	12447268	<a href="#">AP029191.1</a>
✓	<a href="#">Actinomadura sp. strain 7C222 16S ribosomal RNA gene, partial sequence</a>	<a href="#">Actinomadura sp.</a>	2628	2628	100%	0.0	98.85%	1480	<a href="#">KX928701.1</a>
✓	<a href="#">Actinomadura vinacea strain NBRC 14688 16S ribosomal RNA, partial sequence</a>	<a href="#">Actinomadura vinacea</a>	2628	2628	100%	0.0	98.85%	1471	<a href="#">NR_112737.1</a>

Hasil tanpa filter (*no type material dan exclude*)

# KESIMPULAN QUERY SEQUENCE-1

Data exclude model menggunakan data cloning hasil lab dan untuk data *uncultured/environmental* biasanya menggunakan data hasil lingkungan seperti *barcoding* dan metagenomik. Type material memiliki peran *intrinsic* dalam penamaan spesies.

Pada exclude (pengecualian), data model menggunakan data yang panjang sedangkan untuk *uncultured/environmental* 300 bp saja sehingga banyak nukleotida yang dibuang.

Hasil yang ditunjukkan pada sequence-1 dengan filter (*sequences from type material ; and exclude: models (XM/XP), uncultured/environmental sample sequences*) menunjukkan hasil yang sama dengan tanpa filter tapi dengan type material, yaitu dengan persentase identitas 98,85%. Arti dari hal tersebut yaitu adanya indikasi sequence cocok pada *Actinomadura vinacea* strain NBRC 14688, namun terdapat 1% perbedaan pada hasil *query* dengan *subject* database.

Hasil lain pada tanpa type material dan exclude, yaitu 100% pada *Actinomadura* sp. HAT23. Hasil tersebut mengindikasikan bahwa *sequence* berasal dari organisme yang belum dijelaskan penamaan spesiesnya, sehingga dapat muncul apabila tidak menunjukkan *type material*.

Terdapat perbedaan pada organisme hasil exclude dan type material yang tidak dipilih dan dipilih, sedangkan untuk hasil tanpa exclude dengan type material, sama dengan opsi dipilih exclude dengan type material

# SEQUENCE 2

## Pencarian sequence dengan BLASTX

>Sequence-2

```
ATGTCCCGCCTTATATATCGAACAGAATTGGGTATATTTGCTGCCTGTCTGGCCCTTGGTACCCTGCTCTTCTTTACGGTGTATCAGACCCGACCT
GTCCAGGCAGGTAAAGAAGAGGACAAGGTGCAGGCCTGGCAGACGACGGCTGATCTGAGCACCCATCTGGCTCCACAGGCCAATAAAACG
TTTACGACTGGCAGTGGGTCAGACGCGACAAAGATCACGGTTGATGAGAATCAGCAGTATCAACAGATGGACGGCTTTGGTGCCTCGTTTACTG
ATTCTTCGGCCTGGCTGGTCTCCAGTAAAATGGATACTAAGCAACGCGATACGTTAATGAAAAATCTTTTTGATGCTCATCAAGGTATTGGCCTGAA
CTTTTTACGCCAGCCAATGGGTGCCTCAGATCTGACACGGCCAGCACCCGAGGTCGGTGAGTATTCGTATGACGATATGCCACAGGGTCAGA
CTGATCCTAATCTGGATCATTCTCCATTGACCACGATACGAGTTATATCATTCCGGTACTGCAGCAGGCGCTGAAAGAGAATAAAGATATCAAGAT
TATGGCCTCCCCATGGAGCGCACCTGGTTGGATGAAGTCCAGCGGCTCGATGGAAGGGGGAACCCTCAATAGTAGCGCCTATACGGCCTAT
GCTAACTACTTTGTGAAATTCATTCAAGCCTATCAGGCTCAGGGTCTCCCCATCTATGCGGTCACACCGCAGAATGAACCACTCTATGTGCCTTC
CGGCTATCCAGGCATGTCTTTTCCTGCTAAAGATGAAACTACCTTTATCAGGGACTACCTGGGACCAGCAATGGTGAGCAATGGCCTGAGCACT
AAAATTCTGGGTTATGATCATAACTGGGATCAGCCAGGGTATCCAACCACCATTCTGAGTGATCCTGGCACCACACCATATGCCACAGGTACCG
CCTGGCATTGCTATGGCGGGTCCGTCAGTGCTCAGACGCCGATCCATAATGCGTATCCAGACAAAGACACCTATGAGACCGATTGTTCTGGTG
GTCAGTGGGAAGGCAGCAATGGCTTTGCCAATACGATGGATCTCCTGATCGGCACCACGCGCAACTGGAGCAAGTCAGTAGTTCGCTGGGGA
ATGGCCCTGGATCCTGATGGGAAACCAAATCTGGGTACTGGCGCCGCCTGCACACAGTGCCGTGGTATTGTGACGGTTGATCAGACCAATGG
CACTGTCACCTACAATGGCGATTACTATGGTCTGGGACAGGCCAGCAAGTTTGTCCAACCCGGTGCCTATCGTGTGCGCATCCAGTTCTGGCCT
CAATGGCATTAAAGATGTCGCCTTTAAGAATGCGGATGGTTCTGAAGGTGCTGGTGACGTATAATTCGTCTTCCGACCCACAAACGTTTGATGTGC
AGTGGGGCGATAAATGGTTCGCTGACACCCTGCCCGCAGGCGCAGCCACAACCTTCAAATGGACTGGCAAACAATCTACCAATAACCACAGT
GCGCTGAATCGCACTGGTTGGACCACTTCCACCAGCGCTACACCCGCCAGCACCGACAATAGCGCGGGCCAATGTGCTGGATGATGATTATAC
GACGCACTGGTCAAGTACAGAGAACCAGACCAATGGACAGTGGTTCCAGATAGATATGGGCTCAACCCAGACCTTCAGTCAGATGACACTGGA
TGCAGGGAGTGACGGCGACTATCCACACGGTTATCAGGTATCTGTCTCGGATGATGGGGCCAACCTGGAGCAATCCTGTTGCCAGTGGCAATG
GGAGTGCGCAACAACCCCTGAGCATCTCATTGCGGCCCCAGACCGCTCGCTATATTCGGATTGTTTCAGACAGGGAGCGCGGACAGCAAATG
GTCCATTATGAAATGAATGTGTATCCGTAA
```

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. **QUERY** sequence name: Sequence-2
2. **Search Set:** Standard database (nr etc.), exclude: non-redundant RefSeq proteins (WP)
3. **BLAST Search Result:**
  - Query length: 1896 bp (631 asam amino)
  - Protein: Glucosylceramidase
  - Top Hit: glucosylceramidase [*Dictyobacter vulcani*] , dengan persentase identitas 100%.
  - Max Score: 1269; Total Score: 1269: Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Chloroflexota; Ktedonobacterales; *Dictyobacter*; *Dictyobacter vulcani*.

*Dictyobacter vulcani* dan seratus persentase identitas (100%)

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. **QUERY** sequence name: Sequence-2
2. **Search Set:** Standard database (nr etc.), exclude: models (XM/XP) dan non-redundant RefSeq proteins (WP)
3. **BLAST Search Result:**
  - Query length: 1896 bp (631 asam amino)
  - Protein: Glucosylceramidase
  - Top Hit: glucosylceramidase [*Dictyobacter vulcani*] , dengan persentase identitas 100%.
  - Max Score: 1269; Total Score: 1269; Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Chloroflexota; Ktedonobacterales; *Dictyobacter*; *Dictyobacter vulcani*.

*Dictyobacter vulcani* dan seratus persentase identitas (100%)

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. **QUERY sequence name:** Sequence-2
2. **Search Set:** Standard database (nr etc.), exclude: -
3. **BLAST Search Result:**
  - Query length: 1896 bp (631 asam amino)
  - Protein: discoidin domain-containing protein
  - Top Hit: discoidin domain-containing protein [*Dictyobacter vulcani*], dengan persentase identitas 100%.
  - Max Score: 1269; Total Score: 1269; Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Chloroflexota; Ktedonobacterales; *Dictyobacter*; *Dictyobacter vulcani*

*Dictyobacter vulcani* dan seratus persentase identitas (100%)

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. QUERY sequence name: Sequence-2
2. Search Set: Standard database (nr etc.), exclude: models (XM/XP), non-redundant RefSeq proteins (WP) dan Uncultured/environmental
3. BLAST Search Result:  
tidak terbaca

Tidak terbaca



# HASIL IDENTIFIKASI QUERY SEQUENCE

BLAST® » blastx » results for RID-VXHU9WEE013

HomeRecent ResultsSaved StrategiesHelp

Edit Search

Save Search

Search Summary ▾

How to read this report?

BLAST Help Videos

Back to Traditional Results Page

Job Title

Nucleotide Sequence

RID

VXHU9WEE013

Search expires on 02-27 14:48 pm

Download All ▾

Program

BLASTX

Citation ▾

Database

nr

See details ▾

Query ID

lcl|Query\_4152290

Description

None

Molecule type

dna

Query Length

1896

Other reports

?

Filter Results

Organism

only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity

E value

Query Coverage

to

to

to

Filter

Reset

There was a problem with the search. Please, contact [Help Desk](#) and include RID VXHTSASY013.

CPU usage limit was exceeded. You may need to change your search strategy. Helpful changes include reducing the number of queries, choosing a smaller database and setting an organism limit. You may also need to adjust the Algorithm parameters in the bottom section of the form. Choose a smaller number of target sequences, set a smaller Expect cut-off, use a larger word-size and turn on species specific repeats.

Please consider running your BLAST searches with [ElasticBLAST](#) or [BLAST+](#).

Job Title

Nucleotide Sequence

RID

VXHTSASY013

Search expires on 02-27 14:48 pm

Download All ▾

Program

Citation ▾

Database

nr

See details ▾

Query ID

lcl|Query\_4124270

Description

None

Molecule type

dna

Query Length

1896

Other reports

?

Filter Results

Percent Identity

E value

Query Coverage

to

to

to

Filter

Reset

No significant similarity found. For reasons why, [click here](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	<a href="#">discoidin domain-containing protein [Dictyobacter vulcani]</a>	<a href="#">Dictyobacter vu...</a>	1269	1269	100%	0.0	100.00%	631	<a href="#">WP_233097673.1</a>
✓	<a href="#">discoidin domain-containing protein [Ktedonobacteraceae bacterium]</a>	<a href="#">Ktedonobactera...</a>	796	892	100%	0.0	61.43%	777	<a href="#">HEX6555940.1</a>
✓	<a href="#">discoidin domain-containing protein [Ktedonobacteraceae bacterium]</a>	<a href="#">Ktedonobactera...</a>	784	883	100%	0.0	61.02%	777	<a href="#">MBA2395303.1</a>
✓	<a href="#">discoidin domain-containing protein [Ktedonobacteraceae bacterium]</a>	<a href="#">Ktedonobactera...</a>	776	881	100%	0.0	59.22%	776	<a href="#">MBO0779299.1</a>
✓	<a href="#">discoidin domain-containing protein [Chloroflexota bacterium]</a>	<a href="#">Chloroflexota b...</a>	768	877	100%	0.0	59.32%	779	<a href="#">MDQ2903005.1</a>
✓	<a href="#">discoidin domain-containing protein [Ktedonobacteraceae bacterium]</a>	<a href="#">Ktedonobactera...</a>	764	873	100%	0.0	59.63%	777	<a href="#">HEX6477783.1</a>

Kiri: Tanpa filter exclude;  
Kanan: Filter exclude (seluruh), tidak dapat ditemukan organismse karena CPU exceeded

# KESIMPULAN QUERY SEQUENCE-2

Untuk pembuktian data mana yang membuat hasil tidak sesuai kapasitas (parameter seharusnya exclude seluruh opsi dan non-exclude). Saya menemukan masalah pada saat menggunakan filter uncultured/environmental yang menyiratkan bahwa antara sequence terlalu panjang (631 asam amino), dalam jumlah besar memiliki query sequence pada satu pencarian dan query sequence memiliki kompleksitas rendah atau berulang tanpa filtering dan masking.

Sequence-2 dapat ditemukan dengan menggunakan keywords Refseq untuk prokariotik bakteri non-redundant protein record (WP\_). Pada data hasil yang sudah difilter dapat ditemukan pada exclude XM/XP dan WP, sedangkan pada hasil dengan uncultured tidak dapat ditemukan karena CPU yang tidak memadai dan permasalahan lain seperti kompleksitas rendah apabila uncultured/environmental dikecualikan.

Solusi dari CPU exceeded yaitu spesifikasi parameter, cut-off sequence, menggunakan spesies repeat sequence dan menggunakan ElasticBLAST atau BLAST+.

Terdapat hasil perbedaan saat mencari dengan exclude (all filter) tidak dapat dicari (CPU exceeded) dan non-exclude ditemukan discoidin domain-containing protein [*Dictyobacter vulcani*] dengan persentase identitas 100%.

# KESIMPULAN QUERY SEQUENCE-2

Pada hasil menggunakan exclude RefSeq mendapatkan hasil protein glucosylceramidase dengan organisme sama dengan discoidin domain-containing protein (ddc), yaitu *Dictyobacter vulcani*. Hasil lebih merujuk ke protein ddc dibandingkan dengan glucosylceramidase karena merupakan sequence referensi. Lalu hasil seperti ini dapat terjadi pada struktur protein memiliki kesamaan urutan asam aminonya (*identical protein group*).

Identical Protein Groups

Identical Protein

Search

Advanced

Help

Identical Protein Groups

Send to:

discoidin domain-containing protein

GenPept FASTA Graphics BLAST

Name: discoidin domain-containing protein

RefSeq Selected Product: WP\_233097673.1, 631 amino acids

Taxonomic Group: GNS bacteria

Assembly Accessions: 2

Protein Accessions: 2

CDS Regions: 2

Total Rows: 2

Protein Classes

☒ WP\_233097673.1 - 1 row

☒ INSDC - 1 row

Update Report

Source	CDS Region in Nucleotide	Protein	Name	Organism	Strain	Assembly
RefSeq	NZ_BKZW01000001.1 2668138-2670033 (-)	WP_233097673.1	discoidin domain-containing protein	Dictyobacter vulcani	W12	GCF_008974265.1
INSDC	BKZW01000001.1 2668138-2670033 (-)	GER68093.1	glucosylceramidase	Dictyobacter vulcani	W12	GCA_008974265.1

Taxonomic Groups

GNS bacteria [1]

Top Organisms

Recent activity

Turn Off Clear

glucosylceramidase [Dictyobacter vulcani]

Protein

discoidin domain-containing protein [Dictyobacter vulcani]

Protein

MAG: discoidin domain-containing protein [Ktedonobacteraceae bacterium]

Protein

Actinomadura vinacea strain NBRC 14688  
16S ribosomal RNA, partial sequenc

Nucleotide

The NCBI Handbook

See more...

Terdapat hasil perbedaan saat mencari dengan exclude (all filter) tidak dapat dicari (CPU exceeded) dan non-exclude ditemukan discoidin domain-containing protein [*Dictyobacter vulcani*] dengan persentase identitas 100%.

# SEQUENCE 3

## Pencarian sequence dengan BLASTP

>Sequence-3

```
MLLAGDIGGKTNLAVFSSKDELRKPLYAKKFPSAQYTTLAALVTDFLSEVDLPIDRAVFGVAGPVLEGKAKITNLPWVMEENQLQSALKIPTIRLINDL  
AATAQSIPALEPADLHTLNAGEPMKNGTMSVVAPGTGLGEAFLVWDGSKYAIYPSEGGHADFAPTNAFEVGLLVYMLERLPHVSYEHVCSGIGLPNI  
YAYIKESGMFVEPEWLSEKLANAADRTTPVIADGAMAAEPAPICMAALKSF AAILGAEAGNMAIKVLSTGGVYLGGGIPPRILPFLESDDFMRAFRNKG  
RFSNMLGNVPVHVILRPDAGLIGAAAYGFNML
```

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. **QUERY** sequence name: Sequence-3
2. **Search Set:** Database non-redundant protein sequences (nr);  
exclude: -
3. **BLAST Search Result:**
  - Query length: 327 asam amino
  - Protein: glucokinase
  - Top Hit: glucokinase [*Tengunoibacter tsumagoiensis*], dengan persentase identitas 100%.
  - Max Score: 663; Total Score: 663: Query cover: 100%; E value: 0.0
  - Organisme: Bacteria; Bacillati; Chloroflexota; Ktedonobacteria; Ktedonobacterales; Dictyobacteraceae; *Tengunoibacter*; *Tengunoibacter tsumagoiensis*

*Tengunoibacter tsumagoiensis dan seratus persentase identitas (100%)*

# HASIL IDENTIFIKASI QUERY SEQUENCE

1. **QUERY sequence name:** Sequence-3
2. **Search Set:** Database non-redundant protein sequences (nr), exclude models (XM/XP), uncultured/environmental sample sequences, non-redundant RefSeq proteins (WP)
3. **BLAST Search Result:**  
tidak terbaca

Tidak terbaca

# HASIL IDENTIFIKASI QUERY SEQUENCE

[← Edit Search](#) [Save Search](#) [Search Summary ▾](#) [? How to read this report?](#) [▶ BLAST Help Videos](#) [↶ Back to Traditional Results Page](#)

Job Title

Protein Sequence

RID

[VXHEX3CG016](#) Search expires on 02-27 14:42 pm [Download All ▾](#)

Program

BLASTP [? Citation ▾](#)

Database

nr [See details ▾](#)

Query ID

lcl|Query\_3475679

Description

unnamed protein product

Molecule type

amino acid

Query Length

327

Other reports

[Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear ☐ exclude

[+ Add organism](#)

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
✓	<a href="#">glucokinase [Tengunoibacter tsumagoiensis]</a>	<a href="#">Tengunoibacter tsumagoiensis</a>	663	663	100%	0.0	100.00%	327	<a href="#">WP_126583343.1</a>
✓	<a href="#">glucokinase [Dictyobacter formicarum]</a>	<a href="#">Dictyobacter formicarum</a>	472	472	100%	6e-164	67.94%	354	<a href="#">GHO83574.1</a>
✓	<a href="#">glucokinase [Dictyobacter formicarum]</a>	<a href="#">Dictyobacter formicarum</a>	471	471	100%	1e-163	67.94%	340	<a href="#">WP_201361243.1</a>
✓	<a href="#">glucokinase [Dictyobacter kobayashii]</a>	<a href="#">Dictyobacter kobayashii</a>	470	470	99%	1e-163	67.55%	340	<a href="#">WP_126556085.1</a>
✓	<a href="#">glucokinase [Dictyobacter alpinus]</a>	<a href="#">Dictyobacter alpinus</a>	467	467	100%	3e-162	65.09%	338	<a href="#">WP_126630717.1</a>
✓	<a href="#">glucokinase [Dictyobacter aurantiacus]</a>	<a href="#">Dictyobacter aurantiacus</a>	461	461	99%	5e-160	65.68%	340	<a href="#">WP_126601200.1</a>

! Your search is limited to records that exclude: models (XM/XP), uncultured/environmental sample sequences, non-redundant RefSeq proteins (WP)

! There was a problem with the search. Please, contact [Help Desk](#) and include RID VXHE6HX4016. CPU usage limit was exceeded. You may need to change your search strategy. Helpful changes include reducing the number of queries, choosing a smaller database and setting an organism limit. You may also need to adjust the Algorithm parameters in the bottom section of the form. Choose a smaller number of target sequences, set a smaller Expect cut-off, use a larger word-size and turn on species specific repeats. Please consider running your BLAST searches with [ElasticBLAST](#) or [BLAST+](#).

Job Title

Protein Sequence

RID

[VXHE6HX4016](#) Search expires on 02-27 14:41 pm [Download All ▾](#)

Program

[? Citation ▾](#)

Database

nr [See details ▾](#)

Query ID

lcl|Query\_3427628

Description

unnamed protein product

Molecule type

amino acid

Query Length

327

Other reports

[?](#)

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Feedback

Kiri: tanpa filter exclude;  
Kanan: filter exclude (seluruh), tidak dapat ditemukan organisme karena CPU exceeded

# KESIMPULAN QUERY SEQUENCE-3

Pada data hasil sama dengan sequence-3, dengan hasil yang tidak dapat ditemukan pada data dengan exclude seluruh opsi (models (XM/XP), uncultured/environmental sample sequences, non-redundant RefSeq proteins (WP)). Lalu apabila filter exclude tidak dipilih, maka hasil mengarah pada *Tengunoibacter tsumagoiensis* dan seratus persentase identitas (100%) dengan protein glucokinase.

Sudah dicoba oleh pelaku (penulis) pada beberapa percobaan, hasil yang sulit untuk dicari apabila terdapat exclude uncultured/environmental. Hal tersebut dikarenakan oleh hasil data metagenomik yang banyak dan dapat berkorelasi dengan sequence-3. Masalah lain seperti sequence spesifik yang hanya ada di *Tengunoibacter tsumagoiensis* dan seratus persentase identitas (100%) , sedangkan organisme lain pada produk protein tidak ada yang highly identic.

*Terdapat hasil perbedaan saat mencari dengan exclude (all filter) tidak dapat dicari (CPU exceeded) dan non-exclude ditemukan glucokinase [Tengunoibacter tsumagoiensis], dengan persentase identitas 100%.*



# ACCESSION NUMBER SEARCH

1. OR169734

2. AB008413

- **Screenshot hasil pencarian identitas**
- **Cantumkan FASTA filenya**

# ACCESSION NUMBER SEARCH

## 1. OR169734

- Screenshot hasil pencarian identitas

**Euconnus simplex isolate MSC\_11760 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial**

GenBank: OR169734.1

[FASTA](#) [Graphics](#)

[Go to:](#) 

LOCUS	OR169734	421 bp	DNA	linear	INV 27-JUN-2023
DEFINITION	Euconnus simplex isolate MSC_11760 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial.				
ACCESSION	OR169734				
VERSION	OR169734.1				
KEYWORDS	.				
SOURCE	mitochondrion Euconnus simplex				
ORGANISM	<a href="#">Euconnus simplex</a> Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Staphyliniformia; Staphylinidae; Scydmaeninae; Euconnus.				
REFERENCE	1 (bases 1 to 421)				
AUTHORS	Caterino,M.S. and Recuero,E.				
TITLE	Caterino, M.S. and Recuero E. (in review) Shedding light on dark taxa in high Appalachian leaf litter - assessing patterns of endemcity using large scale, voucher-based barcoding. INSECT CONSERVATION & DIVERSITY				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 421)				
AUTHORS	Caterino,M.S. and Recuero,E.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JUN-2023) Plant & Environmental Sciences, Clemson University, 277 Poole Ag. Center, Clemson, SC 29634, USA				
COMMENT	##Assembly-Data-START##				

- Fasta File

```
>OR169734.1 Euconnus simplex isolate MSC_11760
cytochrome c oxidase subunit I (COX1) gene, partial cds;
mitochondrial
ACGAATAAATAATTTAAGTTTTTGATTACTACCTCCTGCCATTTCT
TTATTACTAATAAGTTCAATAATT
GAAAGAGGGAAGTGGAACTGGCTGAACGGTATATCCCCCCTT
TCATCAAATTTAGCTCATAGAGGATCAT
CAGTTGATTTAACAATCTTTTCTTTACATTAGCAGGTATTTTCATC
AATTTTAGGAGCTATTAATTTTAT
TTCCACAATTTTAAATATAAAACCTTTAGAAATAAAATTTGATATA
ATACCTTTATTTATTTGAGCAGTT
TTAATTACCGCCATTTTACTTTTATTATCCTTACCAGTACTAGCA
GGAGCTATTACAATATTACTAACAG
ATCGTAATATTAATACTTCATTCTTTGACCCTAGAGGAGGAGGA
GACCCAATTTTGTATCAACATTTATT
T
```

# ACCESSION NUMBER SEARCH

## 2. AB008413

- Screenshot hasil pencarian identitas

**Aspergillus zonatus** gene for 18S rRNA, partial sequence

GenBank: AB008413.1

[FASTA](#) [Graphics](#)

---

Go to: ☐

LOCUS	AB008413	1733 bp	DNA	linear	PLN 13-FEB-1999
DEFINITION	Aspergillus zonatus gene for 18S rRNA, partial sequence.				
ACCESSION	AB008413				
VERSION	AB008413.1				
KEYWORDS	18S ribosomal RNA.				
SOURCE	Penicillliopsis zonata				
ORGANISM	<a href="#">Penicillliopsis zonata</a> Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Penicillliopsis.				
REFERENCE	1				
AUTHORS	Nikkuni,S., Nakajima,H., Hoshina,S.I., Ohno,M., Suzuki,C., Kashiwagi,Y. and Mori,K.				
TITLE	Evolutionary relationships among Aspergillus oryzae and related species based on the sequences of 18S rRNA genes and internal transcribed spacers				
JOURNAL	J. Gen. Appl. Microbiol. 44 (3), 225-230 (1998)				
PUBMED	<a href="#">12501432</a>				
REFERENCE	2 (bases 1 to 1733)				
AUTHORS	Nikkuni,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-OCT-1997) Sayuki Nikkuni, Japan International Research Center for Agricultural Sci.; Ohwashi 1-2, Tsukuba, Ibarakki 305, Japan (E-mail:nikkuni@jircas.affrc.go.jp, Tel:0298-38-6362, Fax:0298-38-6652)				
FEATURES	Location/Qualifiers				
SOURCE	1 1733				

FEATURES

source

[rRNA](#)

1..1733

/organism="Penicillliopsis zonata"

/mol\_type="genomic DNA"

/isolate="IF08817"

/db\_xref="taxon:[41063](#)"

<1..1733

/product="18S ribosomal RNA"

# ACCESSION NUMBER SEARCH

## 2. AB008413

- **Fasta file**

```
>AB008413.1 Aspergillus zonatus gene for 18S rRNA, partial sequence
AAAGATTAAGCCATGCATGTCTAAGTATAAGCAATTTATACGGTGAAACTGCGAATGGCTCATTAAATCA
GTTATCGTTTATTTGATAGTACCTTGCTACATGGATACCTGTGGTAATTCTAGAGCTAATACATGCTGAA
AACCTCGACTTCGGAAGGGGTGTATTTATTAGATAAAAAACCAATGCCCCCTCGGGGCTCCTTGGTGATTG
ATAATAACTAAACGAATCGCATGGCCTTGCGCCGGCGATGGTTCATTCAAATTTCTGCCCTATCAACTTT
CGATGGTAGGATAGTGGCCTACCATGGTGGCAACGGGTAAACGGGGAATTAGGGTTCGATTCCGGAGAGGG
AGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCCGATTGCGGGA
GGTAGTGACAATAAATACTGATACGGGGCTCTTTTGGGTCTCGTAATTGGAATGAGTACAATTTAAATCC
CTTAACGAGGAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTAT
ATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACCTTGGGTCTGGCTGGCCGGTCCGCCTCACCGCG
AGTACTGGTCCGGCTGGACCTTTCCTTCTGGGGAACTCCATGGCCTTCACTGGCCGTGGGGGGAACCAGG
ACTTTTACTGTGAAAAAATTAGAGTGTTCAAAGCAGGCCTTTGCTCGAATACATTAGCATGGAATAATAG
AATAGGACGTGCGGTTCTATTTTGTGGTTTCTAGGACCGCCGTAATGATTAATAGGGATAGTCGGGGGC
GTCAGTATTGAGCTGTCAGAGGTGAAATCTTGGATTGCTGAAGACTAACTACTGCGAAAGCATTGCCC
AAGGATGTTTTCATTAATCAGGGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTA
ACCATAAACTATGCCGACTAGGGATCGGACGGTGTTCTTTTATGACCCGTTCCGGCACCTTACGAGAAAT
CAAAGTTTTTGGGTTCTGGGGGGAGTATGGTCGCAAGGCTGAACTTAAAGAAATTGACGGAAGGGCACC
ACAAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGGAACTCACCAGGTCCAGACAAAATAAGG
ATTGACAGATTGAGAGCTCTTCTTGATCTTTGGATGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTG
ATTTGTCTGCTTAATTGCGATAACGAACGAGACCTCGGCCCTTAAATAGCCCGGTCCGCATTTGCGGGCC
GCTGGCTTCTTAGGGGGACTATCGGCTCAAGCCGATGGAAGTGCGCGGCAATAACAGGTCTGTGATGCCC
TTAGATGTTCTGGGCCGCACGCGCGCTACACTGACAGGGCCAGCGAGTACATCACCTTGGCCGAAAGGTC
TGGGTAATCTTGTTAAACCCTGTCGTGCTGGGGATAGAGCATTGCAATTATTGCTCTTCAACGAGGAATG
CCTAGTAGGCACGAGTCATCAGCTCGTGCCGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTAC
TACCGATTGAATGGCTCAGTGAGGCCTTCGGACTGGCTCAGGAGGGTTGGCAACGACCCCCCAGAGCCGG
AAAGTTGGTCAAACCTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGTTTC
```

# References

1. NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information, Nucleic Acids Research, Volume 43, Issue D1, 28 January 2015, Pages D6–D17, <https://doi.org/10.1093/nar/gku1130>
2. Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.
3. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
4. Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.