



UNIVERSITAS
GADJAH MADA

Final Project BioInformatika

Athaya Naura Khalilah	(23/512716/PA/21899)
Bagus Cipta Pratama	(23/516539/PA/22097)
Kosmas Rio Legowo	(23/512012/PA/21863)
Muhammad Akmal Fauzan	(23/519741/PA/22303)
Muhammad Hanif Zuhair	(23/516550/PA/22099)

Kontribusi

Nama	Kontribusi
Athaya Naura Khalilah	Implementasi, Analisis
Bagus Cipta Pratama	Implementasi, Analisis
Kosmas Rio Legowo	Implementasi, Analisis
Muhammad Akmal Fauzan	Implementasi, Analisis
Muhammad Hanif Zuhair	Implementasi, Analisis

Data

Sequence 1: Nucleotide, *Mycobacterium tuberculosis* strain yt180643 RNA polymerase beta subunit (*rpoB*) mRNA, complete cds.

```
>MN221386.1 Mycobacterium tuberculosis strain yt180643 RNA polymerase beta subunit (rpoB)
mRNA, complete cds
TTGGCAGATTCGGCCAGAGCAAACAGCCGCTAGTCTAGTCCGAGTCGCCGC
AAAGTCCTCGAATA
ACTCGTACCCGGAGCCTAACCGGTCTCTTCGCTAAAGCTGC
CGAACCAC
TTGAGGTTCCGGACT
CCTTGACGTCGACGAGCATTCTCGAGTGGCTGATCGGTTGCCGCGTGGCG
GAATCGCCGCGAG
CGGGGTGATGTCACCCAGTGGGCTGGAAAGAGGTGCTACGAGCTGTC
CCGATCGAGGACTTCT
CCGGGTGATGTCGTTGTCGTTCTGACCCCTCGTTCGACGATGTC
AAGGCACCCGTCGACGAGTGCAA
AGACAAGGACATGACGTACGGGGCTTCACTGTTGTCACCCGGAGTTCA
ACAACAACACCCGGTGAG
ATCAAGAGTCAGACGGTTCATGGGTGACTTCCGATGATGACCGAGAAGGGC
ACGTTCATCATCAACG
GGACCGAGCGTGGTGGTCAGCCAGCTGGTGC
GGTCGCCGGGTGACTTCGACGAGACCAATTGACAA
GTCACCGACAAGACGCTGCACAGCGTAAGGTGATCCCAGGCCGCGCGT
GGCTCGAGTTGACGTC
```



Sequence 2: Protein, thymidine kinase, TK [varicella zoster virus VZV, Peptide, 341 aa]

```
>AAB24914.1 thymidine kinase, TK [varicella zoster virus VZV, Peptide, 341 aa]
MSTDKTDVKMVLRIYLDGAYGIGKTTAAEELHHFAITPNRILLIGEPLSYWRNLAGE
DAICGIYGTQT
RRLNGDVSPEDAQRLLTAHFQSLFCSPHAIMHKISALMDTSTS
DLVQVNKEPYKIMLSDRPIASTICFP
LSRYLVGDMSPAALPGLLFTLPAEPPGTNLVVCTVSLP
SHLSRVSKRARPGETVNL
PFVMVLNVYIMLI
NTIIIFLKTNNWHAGWNTLSFCNDVFQKLQKSEC
IKLREVPGIEDTLFAVLKLP
ELCGEFGN
ILPLWA
G
METLSNCRSMSPFVLSLEQTPHAAQELKTL
PQMTPANMSGAWN
ILKELVNAVQDNTS
```

Keterangan

HIT: Hasil temuan atau kandidat sekuens

ID: Kode akses unik (*Accession Number*)

Deskripsi: Nama organisme dan nama gen/protein yang terindikasi

Panjang: Jumlah nukleotida atau asam amino

E-value: Nilai probabilitas kecocokan terjadi karena kebetulan

Bit score: Skor kualitas alignment

Identity: Persentase karakter yang sama persis antara input dan database



OUTPUT SEQ 1

HIT	ID	Deskripsi	Panjang	E-Value	Bit Score	Identity
1	gi 1917459103 gb MN221386.1	Mycobacterium tuberculosis strain yt180643 RNA polymerase beta subunit (rpoB) mRNA, complete cds	3519	0.0	6347.34	100.00% (3519/3519)
2	gi 2556671655 gb CP130778.1	Mycobacterium tuberculosis strain MTb-Oman-3213831 chromosome, complete genome	4321982	0.0	6338.32	99.94% (3517/3519)
3	gi 3020066102 gb CP194235.1	Mycobacterium tuberculosis strain LP-0504069-RM2 chromosome, complete genome	4411637	0.0	6338.32	99.94% (3517/3519)
4	gi 3020066110 gb CP194243.1	Mycobacterium tuberculosis strain LP-0261867-RM7 chromosome, complete genome	4410169	0.0	6338.32	99.94% (3517/3519)
5	gi 2221354464 gb CP069070.1	Mycobacterium tuberculosis strain N1283 chromosome, complete genome	4419495	0.0	6338.32	99.94% (3517/3519)

OUTPUT SEQ 2

HIT	ID	Deskripsi	Panjang	E-Value	Bit Score	Identity
1	ref NP_040159.1	thymidine kinase [Human alphaherpesvirus 3] >sp P09250.1 RecName: Full=Thymidine kinase [Human herpesvirus 3 strain Dumas] >gb AAB24914.1 thymidine kinase, TK [varicella zoster virus VZV, Peptide, 341 aa] [Human alphaherpesvirus 3] >gb AEW88445.1 thymidine kinase [Human alphaherpesvirus 3] >gb AEW88733.1 thymidine kinase [Human alphaherpesvirus 3] >emb CAA27919.1 deoxypyrimidine kinase [Human herpesvirus 3 strain Dumas]	341	0.0	709.909	100.00% (341/341)

OUTPUT SEQ 2

HIT	ID	Deskripsi	Panjang	E-Value	Bit Score	Identity
2	sp P0C0E6.1	<p>RecName: Full=Thymidine kinase [Human herpesvirus 3 strain Oka vaccine] >pdb 1OSN A Chain A, Thymidine kinase [Human alphaherpesvirus 3] >pdb 1OSN B Chain B, Thymidine kinase [Human alphaherpesvirus 3] >pdb 1OSN C Chain C, Thymidine kinase [Human alphaherpesvirus 3] >pdb 1OSN D Chain D, Thymidine kinase [Human alphaherpesvirus 3] >gb WRP98722.1 ORF36 [synthetic construct] >gb XXP31349.1 thymidine kinase [Varicellovirus humanalpha3] >gb AAA45865.1 thymidine kinase [Human alphaherpesvirus 3] >gb AAT07718.1 thymidine kinase [Human alphaherpesvirus 3] >gb AAT07794.1 thymidine kinase [Human alphaherpesvirus 3]</p>	341	0.0	706.827	99.71% (340/341)



OUTPUT SEQ 2

HIT	ID	Deskripsi	Panjang	E-Value	Bit Score	Identity
3	gb XXP331 74.1	thymidine kinase [Varicellovirus humanalpha3] >gb XXP33247.1 thymidine kinase [Varicellovirus humanalpha3] >gb XXP33320.1 thymidine kinase [Varicellovirus humanalpha3]	341	0.0	706.442	99.41% (339/341)
4	gb AAA935 09.1	thymidine kinase:phleomycin/zeocin resistance fusion protein [Cloning vector pZEO-SG2]	475	0.0	706.057	99.71% (339/340)
5	gb ABF220 48.1	deoxypyrimidine kinase [Human alphaherpesvirus 3] >gb QCA46240.1 ORF36 [Human alphaherpesvirus 3] >gb QCA46751.1 ORF36 [Human alphaherpesvirus 3] >gb QCA47189.1 ORF36 [Human alphaherpesvirus 3] >gb QCA47262.1 ORF36 [Human alphaherpesvirus 3]	341	0.0	706.057	99.41% (339/341)

ANALISIS SEQ 1

Identifikasi Organisme (Top Hit):

- **Hit #1 (MN221386.1):** Teridentifikasi sebagai *Mycobacterium tuberculosis strain yt180643* (Gen rpoB).
- **Perfect Match:** Skor Identity mencapai **100.00% (3519/3519)**. Sequence input sama persis dengan database tanpa ada perbedaan satu huruf pun.

Perbandingan dengan Hit Lain (Hit #2 – #5):

- Hit #2 s.d #5 merupakan **Complete Genome** (Genom Utuh Bakteri), terlihat dari kolom *Panjang* yang mencapai >4 juta bp.
- Meski genomnya panjang, BLAST hanya mencocokkan pada area spesifik (rpoB), dengan kemiripan **99.94%**.
- Selisih 0.06% (2 huruf berbeda) menyebabkan skor *Bit Score* Hit #2 (6338.32) lebih rendah dari Hit #1 (6347.34).

Validitas Statistik:

- **E-Value 0.0:** Konsisten di semua top hits. Menunjukkan probabilitas error nol; identifikasi organisme ini **sangat signifikan** dan valid secara biologis.

ANALISIS SEQ 2

Identifikasi Organisme (Top Hit):

- Hit #1 (NP_040159.1) teridentifikasi sebagai thymidine kinase (TK) dari Human alphaherpesvirus 3 / Varicella zoster virus (VZV).
- **Perfect Match:** Skor Identity mencapai **100.00% (341/341)**. Sequence input sama persis dengan database tanpa ada satu pun perbedaan residu..

Perbandingan dengan Hit Lain (Hit #2 – #5):

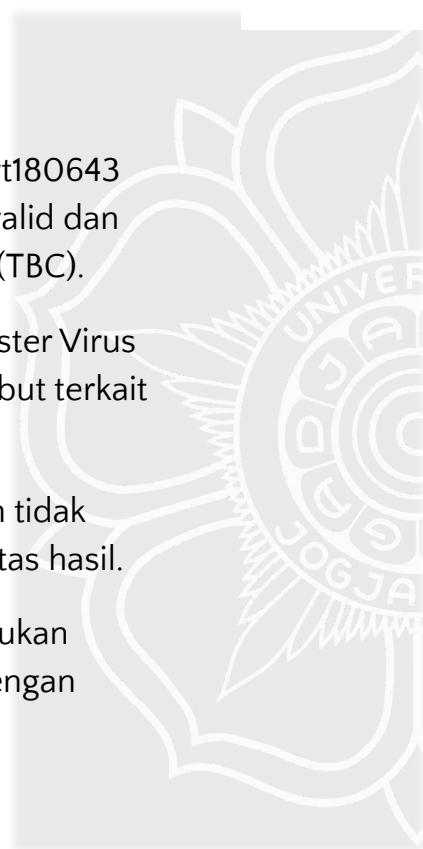
- Hit #2 s.d #5 merupakan varian thymidine kinase (TK) Varicella zoster virus/Human alphaherpesvirus 3 atau turunannya), terlihat dari panjang domain yang sejajar -340–341 aa dan deskripsi fungsinya yang sama.
- Meski ada perbedaan strain/konstruksi, BLAST hanya mencocokkan pada domain TK yang konservatif, dengan tingkat kemiripan sangat tinggi (99.41–99.71%, selisih hanya 1–2 asam amino).
- Perbedaan kecil ini menyebabkan skor Bit Score Hit #2–#5 sedikit lebih rendah dibanding Hit #1

Validitas Statistik:

- **E-Value 0.0:** Konsisten di semua top hits. Menunjukkan probabilitas error nol; identifikasi organisme ini **sangat signifikan** dan valid secara biologis.

Kesimpulan

- Sequence 1 (nukleotida) teridentifikasi sebagai *Mycobacterium tuberculosis* strain yt180643 pada gen *rpoB* dengan tingkat kecocokan 100% dan E-value 0.0, sehingga sangat valid dan menunjukkan bahwa sekuens tersebut berasal dari bakteri penyebab Tuberkulosis (TBC).
- Sequence 2 (protein) teridentifikasi sebagai Thymidine Kinase (TK) dari Varicella Zoster Virus (VZV) dengan identity 100% dan E-value 0.0, mengonfirmasi bahwa sekuens tersebut terkait virus penyebab Cacar Air.
- Kedua sekuens menunjukkan perfect match, sehingga identifikasi sangat akurat dan tidak menyisakan ambiguitas. Perbedaan kecil pada hit lainnya tidak memengaruhi validitas hasil.
- Secara keseluruhan, analisis ini menunjukkan bahwa metode BLAST mampu melakukan identifikasi organisme dan protein secara tepat berdasarkan kesesuaian sekuens dengan database referensi.





UNIVERSITAS
GADJAH MADA

Terima Kasih

<https://colab.research.google.com/drive/1ogHf5xCnpGEAWhM-sNCbq4Kk8FTqhFMy?usp=sharing>

