

## HIV i retslokalet -sådan finder man data i Genbank

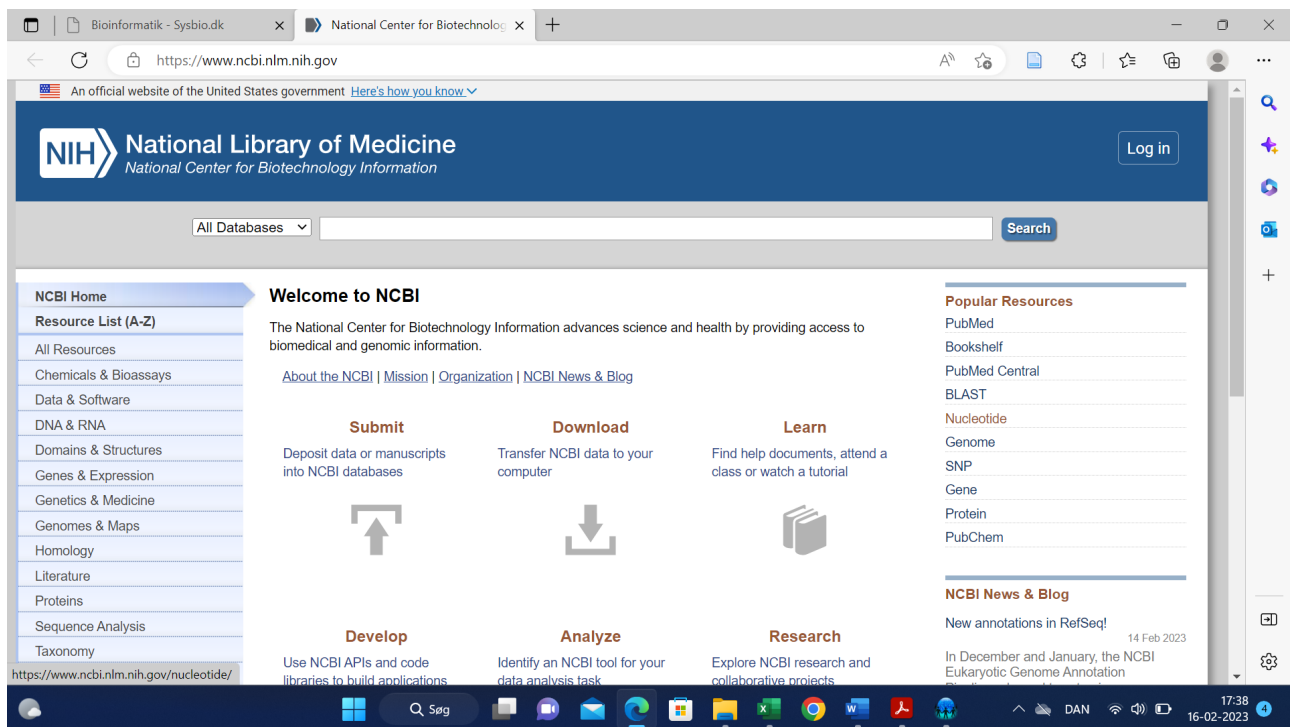
Af: *Kresten Cæsar Torp, Aalborghus Gymnasium*

*Materialet inddrager dele af: Regassa, Laura; Naowarra Cheeptham og Michèle Shuster: Murder by HIV?, NSTA, Link: [Murder by HIV? Grades 9-12 Edition | NSTA](#)*

### Oprettelse af bibliotek

Første skridt er at finde de sekvenser som skal sammenlignes. Dem samler man i en fil, en form for bibliotek. De skal have et format, som kan genkendes af programmet, når de skal sammenlignes. Selve sammenligningen kaldes et alignment, og man kan ud fra det konstruere et fylogentisk træ. Det kan du arbejde videre med i filen: *HIV i retsgenetik -alignment og fylogentisk træ.*

1. Gå ind på NCBI's database Genbank. Link: [National Center for Biotechnology Information \(nih.gov\)](https://www.ncbi.nlm.nih.gov)
2. Vælg "Nucleotide" i menuen til højre.



3. Find den sekvens, vi vil sammenligne med de øvrige. Det gør man vha. en accession number, dvs. et ID, som sekvensen har i databasen. Kopier Accession Number: **AY156807**.
4. Indsæt nummeret i søgefeltet.



FASTA -

HIV-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds

GenBank: AY156807.1

```
>AY156807.1 HIV-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds
GTAGGCTACACCTGTCAACATATTGGAAGAAATCTGTTGACTCAGATTGGTTCACCTTTAAATTTTC
CCATAAGTCCATTGGAACCTGTACCAAGTAAATTAAGCCAGGAATGGCCCAAAAGTTAAACAATG
GCCACTGACAGAGAAAAAATAAAGCATTAGTAGAATTTGTACAGAAATAGAAAAAGGAGAAAAATT
TCAAAAATTGGGCTGAATAATCCATACAATCTCCAGTATTGCCATAAAGAAAAAACAAGTACTAGAT
GGAAAAATTAGTAGATTTCAGAGAACTTAATAGAGAATCAAGACTCTGGGAAGTCAATTAGGAAT
ACCACATCTGCAGGGTTAAAAAAGAAAAATCAGTAACAGTCTGGATGGGTGATGCATATTTTCA
GTTCCCTTAGATAAAGAGTTCAGGAAATATACTGCATTTACCATACCTAGTATAACAATGAGACACCAG
AGATTAGATACAGTACAATGTGCTCCACAGGGATGGAAAGGATCACCAGCAATATCCAAAGTAGCAT
GACAAAAATCTTAGAGCCTTTAGAAAAAATAATCCAGACATAGTTATCTATCAATACATGGATGATCTG
TATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAAATAAACAGAGGAATCAAGACACATCTGTTGA
AGTGGGATTTTTCACACCAGAAAAACACAGAAAGAACCTCCATCCGTTGGATGGGTATGAACT
CCATCTGATAAATGGACAGTACAGCTATAGTGC
```

Første linje indeholder en række oplysninger om sekvensen. Dem vender vi lige med...

7. Marker kun selve DNA-sekvensen (GTAG.....GTGC) og kopier den.
8. Gå tilbage til Genbanks hovedside.
9. Vælg BLAST (lige under "Nucleotide")

BLAST®

Home Recent Results Saved Strategies Help

**Basic Local Alignment Search Tool**

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**NEWS**

ElasticBLAST 1.0.0 is Now available!

ElasticBLAST version 1.0.0 has support for faster cheaper disks at AWS and better supports Kubernetes on GCP!

Mon, 09 Jan 2023 [More BLAST news...](#)

**Web BLAST**

**Nucleotide BLAST**  
nucleotide → nucleotide

**blastx**  
translated nucleotide → protein

**tblastn**  
protein → translated nucleotide

**Protein BLAST**  
protein → protein

BLAST er et program som anvendes til at søge sekvenser i databasen, som ligner den sekvens, man indtaster (Query).

10. Tryk på Nucleotide BLAST.
11. Indsæt sekvensen fra før.

The screenshot shows the NCBI BLAST search page. The 'Enter Query Sequence' section is highlighted with a red box, containing a FASTA sequence. Below it, the 'Choose Search Set' section is visible, with 'Standard databases (nr etc.)' selected and 'Nucleotide collection (nr/nt)' chosen in the dropdown menu.

12. Under "Choose Search Set" vælg: "Standard database" og "Nucleotide collection (nr/nt)".

13. Under "Program Selection": Vælg "Somewhat similar"

This is a close-up of the 'Choose Search Set' section. The 'Database' section has 'Standard databases (nr etc.)' selected. The 'Nucleotide collection (nr/nt)' is selected in the dropdown menu. The 'Organism' section is optional, and the 'Exclude' section has 'Sequences from type material' selected.

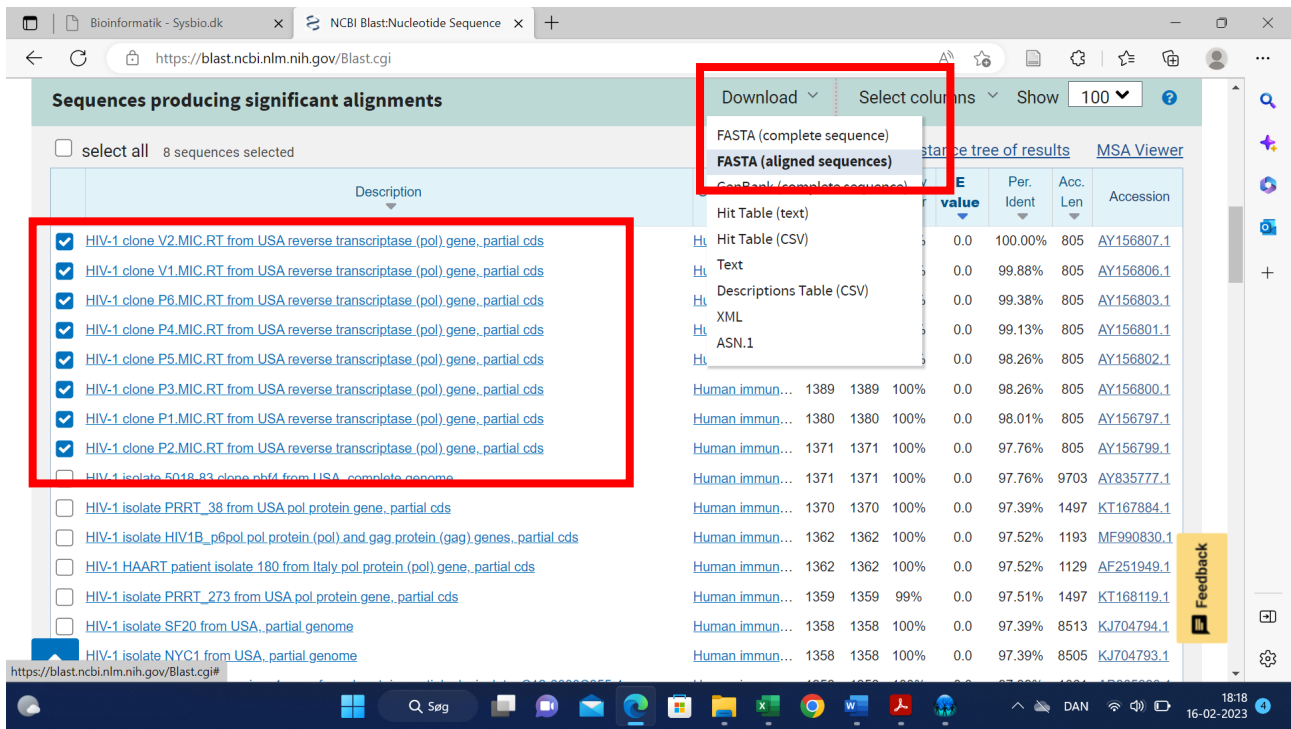
14. Tryk: "BLAST". Det tager lidt tid af lave matchet i databasen. Husk at den indeholder millioner af sekvenser.

15. Scroll lidt ned. Her finder du en liste over de hits i databasen, som minder mest om den indtastede. Marker de første 8. De stammer fra retssagen:

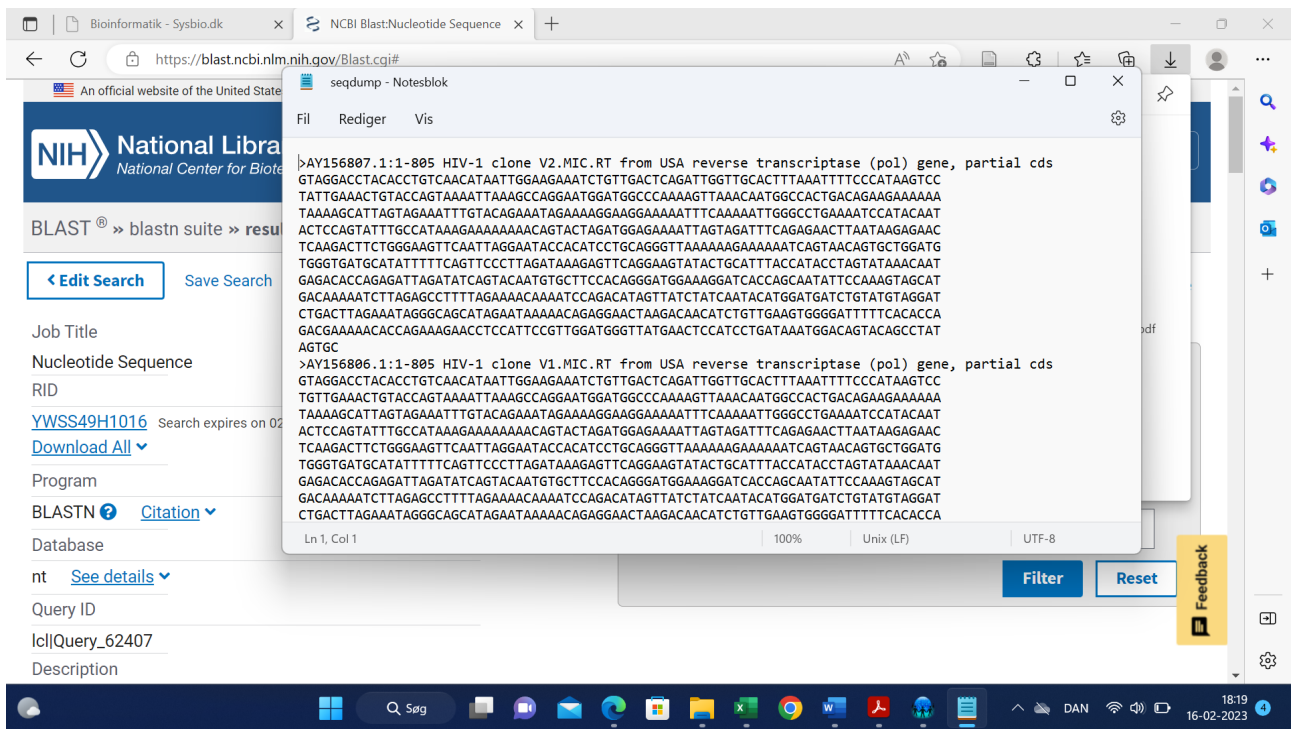
- Klon V1 og Klon V2 er virusvarianter fra offeret (victim)
- Klonerne P1-P6 stammer fra patienten.
- "MIC.RT." angiver metoden: Revers Transkriptase
- "(pol) gene" skyldes, at sekvensen kun er et udsnit af HIV-genomet, nemlig pol-genet.

16. Download filerne ved at trykke på "Download".

17. Vælg "FASTA (aligned sequence)".



18. Gem filerne som en tekstfil (.txt) ved copy-paste. Word ol. Du'r ikke.



19. For at undersøge, om virus fra offer og patient er tæt beslægtede, må man sammenligne med virus fra andre personer, der ikke har været involverede i sagen. Suppler filen med sekvenser fra tre andre personer fra USA og tre fra samme stat (Louisiana). Sekvensernes aquisition number er:

- AY835777 >USA\_1
- AY835778 >USA\_2

- AY835769 >USA\_3
- AY156793 >LA\_1
- AY156789 >LA\_2
- AY156788 >LA\_3

20. Nu skal filen rettes til, så sekvenserne får håndterbare navne (første linje inden sekvensen starter). De vigtigste regler er:

Sekvensen skal altid starte med ">"

Der må ikke være mellemrum. Brug "\_"

Du kan fx omdøbe sekvenserne til (men kan også vælge andre navne, som giver mening):

Acquisition number	Nyt navn
AY156807	>victim_clone1
AY156806	>victim_clone2
AY156803	>patient_clone1
AY156801	>patient_clone2
AY156802	>patient_clone3
AY156800	>patient_clone4
AY156797	>patient_clone5
AY156799	>patient_clone6
AY835777	>USA_1
AY835778	>USA_2
AY835769	>USA_3
AY156793	>LA_1
AY156789	>LA_2
AY156788	>LA_3

Du har nu en FASTA-fil du kan anvende til at lave det fylogenetiske træ. Gem din fil med et navn, som du kan genkende.