SESSION 8. PRACTICE

Evolution Resolving a criminal case

Phylogenetic analysis of HIVs

DNA samples from Victim, patients, and controls (from Lafayette regions)

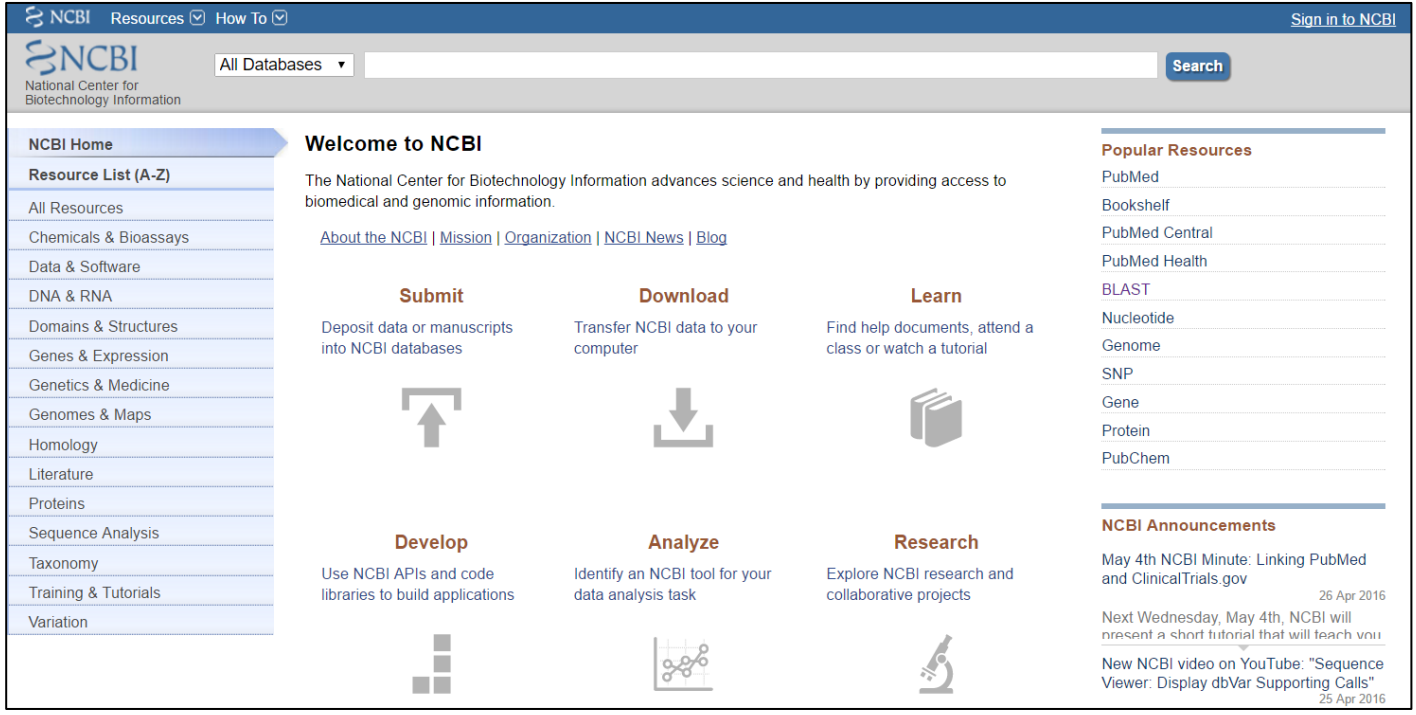
PCR DNAs with primers of Env and RT and sequenced them.

Computational analysis of phylogenetic trees with the sequences using ClustalW

Data is publicly available in NCBI Entrez (AY156734-AY156907)

- 132 env sequences
- 42 RT sequences
- clustalw2 rt.fa

How to download Data from NCBI database



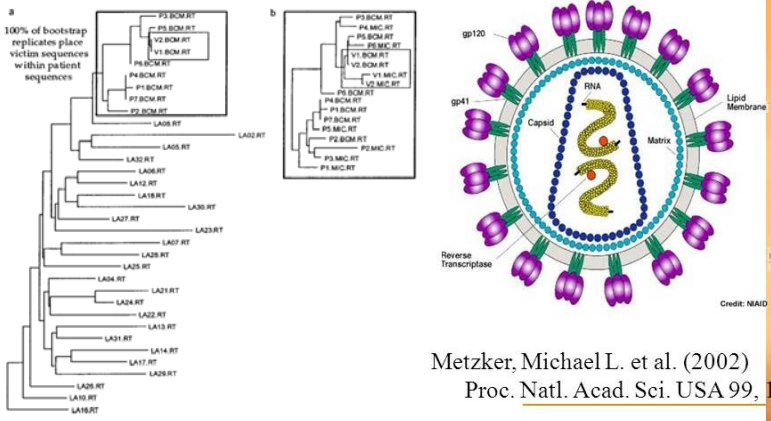
Phylogenetic tree as both scientific and legal evidence in crime scene

Molecular evidence of HIV-1 transmission in a criminal case

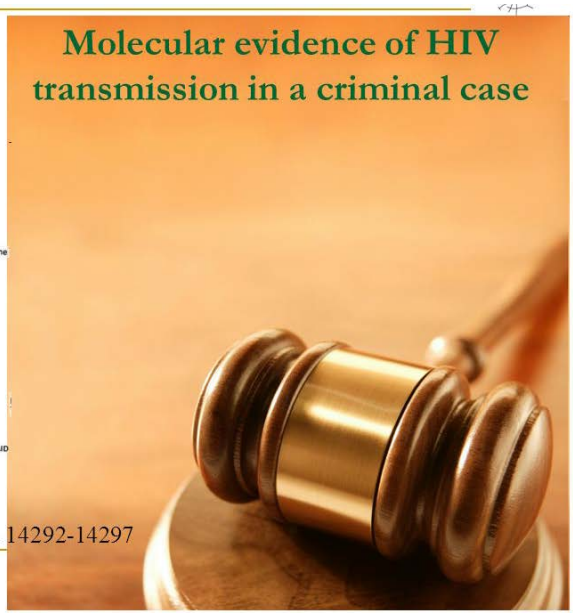
Michael L. Metzker*†, David P. Mindell‡, Xiao-Mei Liu*§, Roger G. Ptak¶‖, Richard A. Gibbs*, and David M. Hillis**

*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; ‡Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079; §School of Dentistry, Biologic and Materials Sciences, University of Michigan, Ann Arbor, MI 48109; and **Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, TX 78712

Edited by Walter M. Fitch, University of California, Irvine, CA, and approved September 4, 2002 (received for review May 2, 2002)

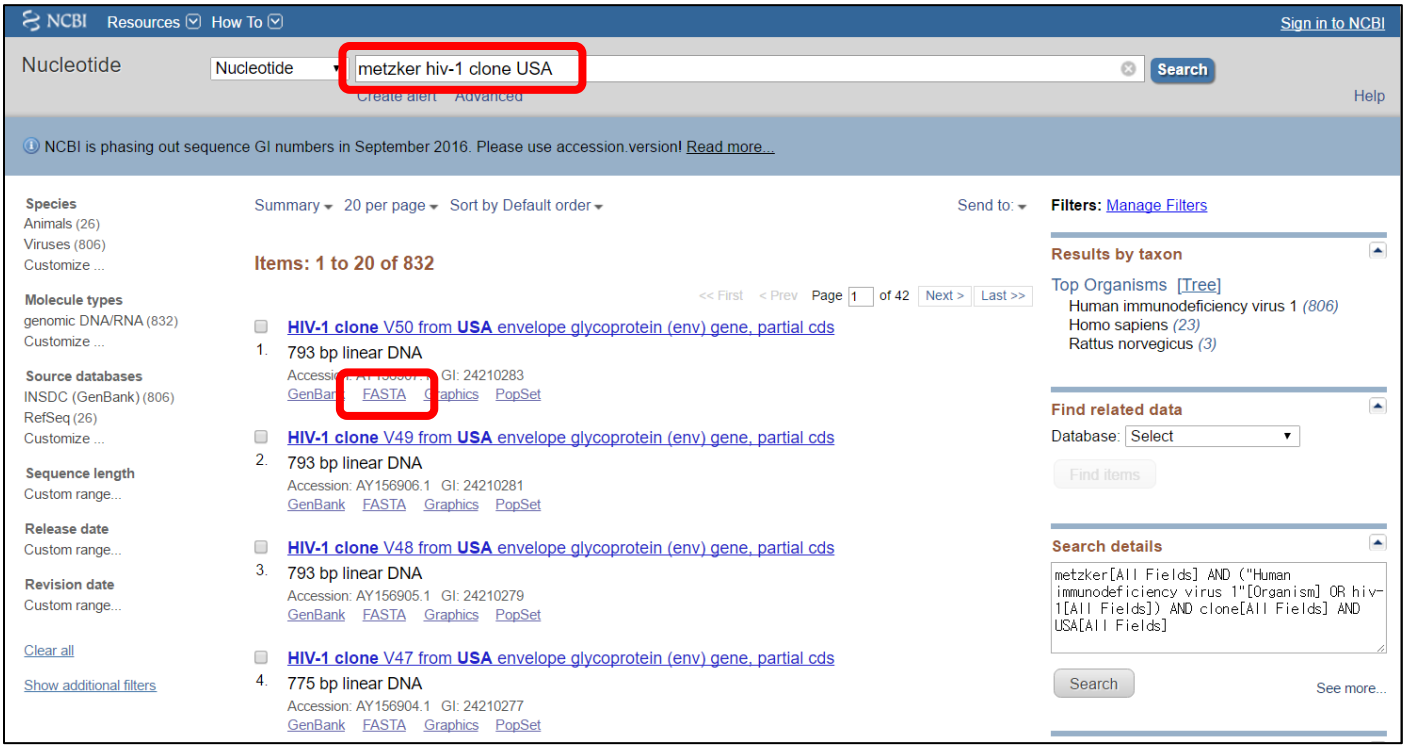


Metzker, Michael L. et al. (2002)
Proc. Natl. Acad. Sci. USA 99, 14292-14297



https://slideplayer.com/slide/8825829/

How to download Data from NCBI database (data from the paper in previous slide)



How to download Data from NCBI database (data from the paper in previous slide)

Basic Shell Commands

```
$ cd [User_Folder]

$ mkdir Session8

$ cd Session8
```

Using clustalw2 for phylogenetic analysis

```
$ ln -s /home/biguser/tutor/Session8/rt.fa .
$ ll
```

```
[biguser@biglab-master session8]$ ll
total 36
-r--r--r-- 1 biguser biguser 34710 Apr 27 10:37 rt.fa
```

Using clustalw2 for phylogenetic analysis

```
$ clustalw2 rt.fa
$ clustalw2 rt.aln –tree
$ clustalw2 rt.aln –bootstrap=1000
```

```
[biguser@biglab-master session8]$ ll
total 108
-rw-rw-r-- 1 biguser biguser 58796 Apr 27 15:10 rt.aln
-rw-rw-r-- 1 biguser biguser  1983 Apr 27 15:10 rt.dnd
-r--r--r-- 1 biguser biguser 34710 Apr 27 10:37 rt.fa
-rw-rw-r-- 1 biguser biguser  1984 Apr 27 15:10 rt.ph
-rw-rw-r-- 1 biguser biguser  2172 Apr 27 15:10 rt.phb
```

Using clustalw2 for phylogenetic analysis

```
$ less rt.phb
```

```
(
(
gi|24209945|gb|AY156737.1|:0.00000,
gi|24209939|gb|AY156734.1|:0.00146)
:0.00000[590],
gi|24209951|gb|AY156740.1|:0.00000)
:0.00162[976],
gi|24210015|gb|AY156802.1|:0.00128)
:0.00077[431],
(
gi|24210009|gb|AY156799.1|:0.00845,
gi|24210011|gb|AY156800.1|:0.00149)
:0.00115[318])
:0.00080[295],
gi|24209941|gb|AY156735.1|:0.00338)
:0.00142[389],
gi|24210007|gb|AY156797.1|:0.00327)
:0.00151[590],
(
(
(
gi|24209943|gb|AY156736.1|:0.00267,
gi|24210013|gb|AY156801.1|:0.00313)
:0.00070[417],
(
(
(
(
gi|24209953|gb|AY156741.1|:0.00000,
gi|24209955|gb|AY156742.1|:0.00000)
:0.00018[961],
gi|24209947|gb|AY156738.1|:0.00127)
:0.00017[697],
gi|24210017|gb|AY156803.1|:0.00273)
:0.00071[400],
(
gi|24210019|gb|AY156806.1|:0.00134,
gi|24210021|gb|AY156807.1|:0.00000)
:0.00318[959])
:0.00158[549])
:0.00245[788],
```

Code 9.1 reformat_giline.ipynb

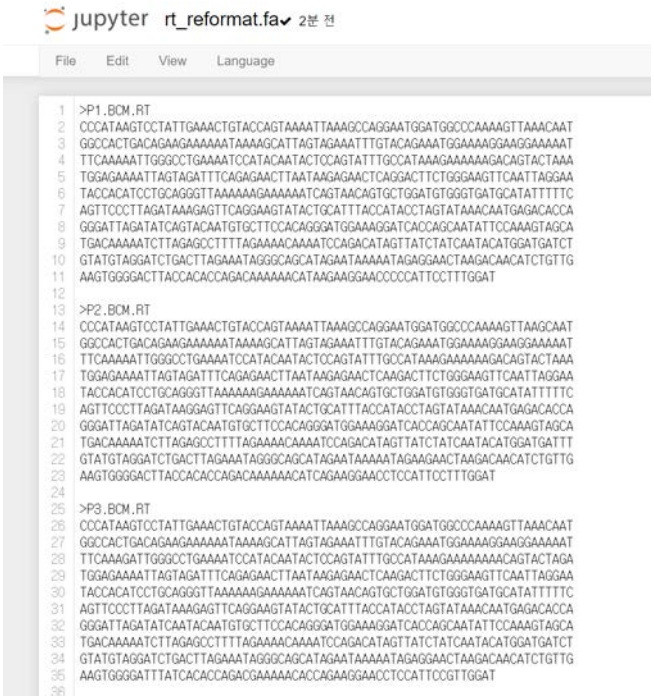
In [4]:

```
import sys
import re

output_open= open("rt_reformat.fa","w")
for line in open("rt.fa", "r"):
    line= line.strip()
    match= re.search(">.*clone (₩S+) ", line)
    if match:
        id= ">"+ match.group(1)+ "₩n"
        output_open.write(id)
    else:
        output_open.write(line.rstrip()+ "₩n")
output_open.close()
```

*Regular expression "\S" - Matches any character which is not a Unicode whitespace character

Code 9.1 reformat_giline.ipynb



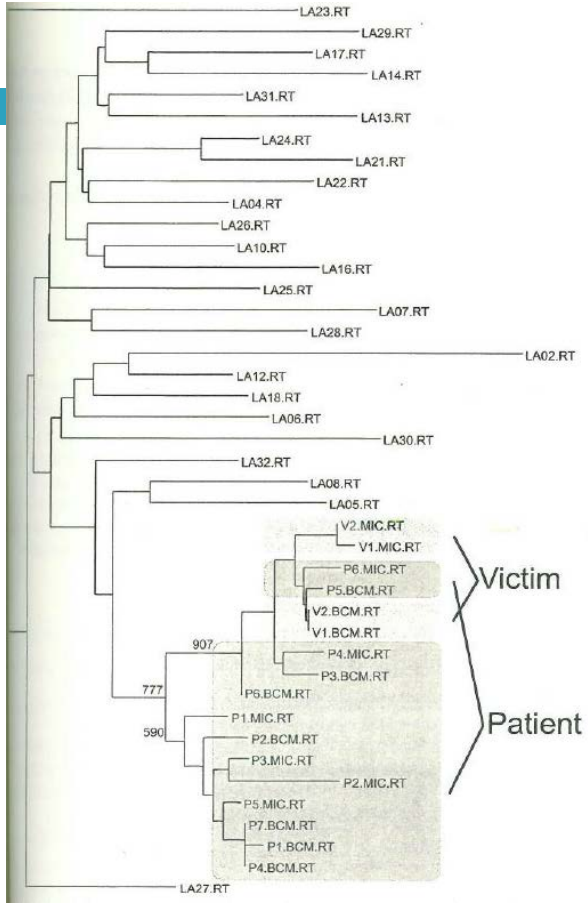
Using clustalw2 for phylogenetic analysis

```
$ clustalw2 rt_reformat.fa
$ clustalw2 rt_reformat.aln –tree
$ clustalw2 rt_reformat.aln –bootstrap=1000
```

```
[biguser@biglab-master session8]$ ll
total 236
-r--r--r-- 1 biguser biguser 117608 Apr 27 14:00 env.fa
-r--r--r-- 1 biguser biguser  34710 Apr 27 10:37 rt.fa
-rw-rw-r-- 1 biguser biguser  44904 Apr 27 15:04 rt_reformat.aln
-rw-rw-r-- 1 biguser biguser   1219 Apr 27 15:04 rt_reformat.dnd
-rw-rw-r-- 1 biguser biguser  30762 Apr 27 15:04 rt_reformat.fa
-rw-rw-r-- 1 biguser biguser   1408 Apr 27 15:05 rt_reformat.phb
```

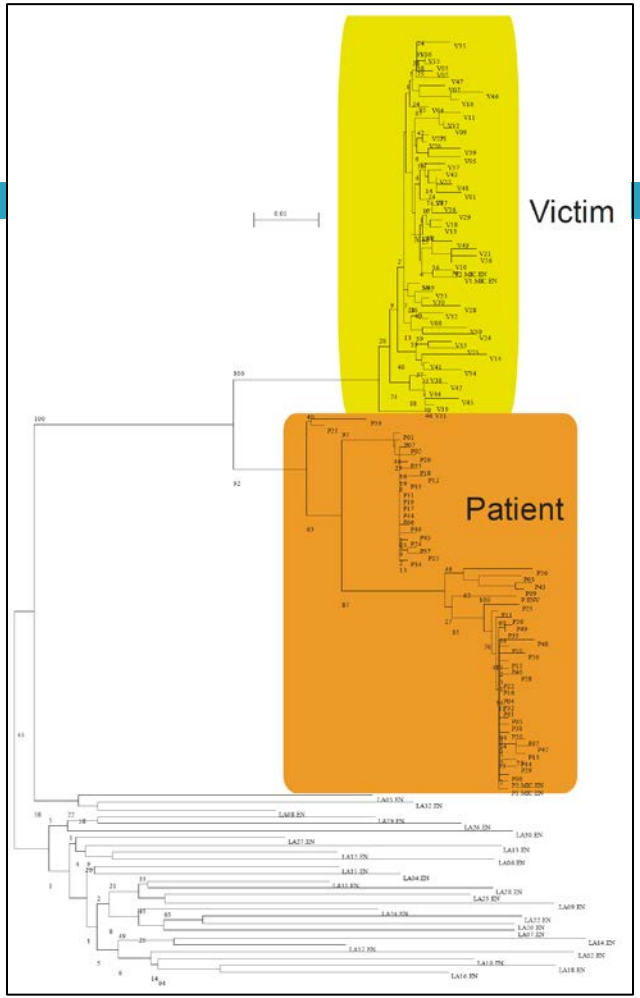
Using clustalw2 for phylogenetic analysis ("rt.fa")

```
(
(
P4.BCM.RT:0.00000,
P1.BCM.RT:0.00146)
:0.00000[590],
P7.BCM.RT:0.00000)
:0.00162[976],
P5.MIC.RT:0.00128)
:0.00077[431],
(
P2.MIC.RT:0.00845,
P3.MIC.RT:0.00149)
:0.00115[318])
:0.00080[295],
P2.BCM.RT:0.00338)
:0.00142[389],
P1.MIC.RT:0.00327)
:0.00151[590],
(
(
(
P3.BCM.RT:0.00267,
P4.MIC.RT:0.00313)
:0.00070[417],
(
(
(
(
V1.BCM.RT:0.00000,
V2.BCM.RT:0.00000)
:0.00018[961],
P5.BCM.RT:0.00127)
:0.00017[697],
P6.MIC.RT:0.00273)
:0.00071[400],
```



Using clustalw2 for phylogenetic analysis (“env.fa”)

```
(
(
(
LA16.EN:0.03515,
LA18.EN:0.05200)
:0.00998[68],
LA02.EN:0.06466)
:0.00325[9],
(
LA10.EN:0.05145,
LA26.EN:0.04984)
:0.00366[33])
:0.00593[2],
(
LA17.EN:0.02509,
LA14.EN:0.06339)
:0.01176[55])
:0.00212[1],
(
(
LA07.EN:0.05054,
LA24.EN:0.03889)
:0.00531[25],
(
LA20.EN:0.04773,
LA22.EN:0.04938)
:0.00634[49])
:0.00281[4])
:0.00036,
(
(
LA06.EN:0.06420,
LA05.EN:0.04289)
:0.00278[15],
(
(
LA09.EN:0.06125,
LA25.EN:0.04618)
:0.00343[25],
(
LA28.EN:0.05127,
```



Exercise

- Make a Python script to derive one of the distance matrices shown in Fig. 9.1. The starting point for the script is the multiple alignment in the same figure.

```
>A
GGACCACTACGAGCGCCTACGACGTA
>B
GGACCCCTACGAGCCCCTACGACGTA
>C
GGACCGCTGCGAGCTTCTACGACGTA
>D
GGACCTCTCCGGGCAGCTAGGACGTA
```

	A	B	C	D
A	0	2	4	6
B	2	0	4	6
C	4	4	0	6
D	6	6	6	0

Answer for Exercise

i : row
j : column

for j in xrange

for i in xrange

	A	B	C	D
A	0	2	4	6
B	2	0	4	6
C	4	4	0	6
D	6	6	6	0

```
seqs= ["GGACCACTACGAGCGCCTACGACGTA",
       "GGACCCCTACGAGCCCCTACGACGTA",
       "GGACCGCTGCGAGCTTCTACGACGTA",
       "GGACCTCTCCGGGCAGCTAGGACGTA"]
for i in xrange(len(seqs)):
    rowseq= seqs[i]
    for j in xrange(len(seqs)):
        colseq= seqs[j]
        diff=0
        for k in range(len(colseq)):
            if rowseq[k] != colseq[k]:
                diff +=1
        print diff,
    print "Wn"
```

```
0 2 4 6

2 0 4 6

4 4 0 6

6 6 6 0
```

이번주 과제는 없습니다. 중간고사 시험 준비 잘 하시기 바랍니다.