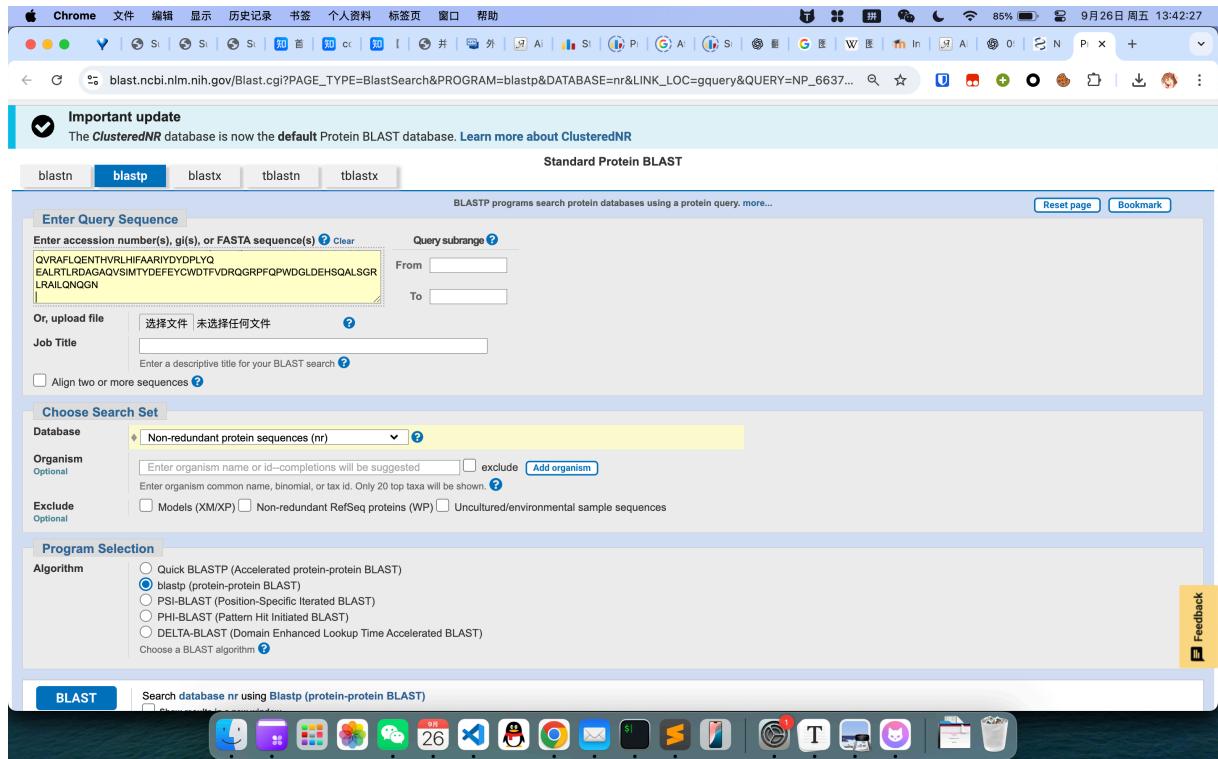


Bioinfo Lab 1

Bohan YANG

September 2025

1 Task 1



The screenshot shows the NCBI BLAST search interface. The 'blast' tab is selected. The query sequence is a protein sequence: QVRAFLQENTHVRHLIFAAIRIVYDPLYQ EALRTLRLDAGAQVSIMTYDEFYEYCWDTFVRQGRPFQPWDGLDEHSQALSGR LRALQNQGN. The search parameters include the 'Non-redundant protein sequences (nr)' database, the 'blastp' algorithm, and the 'protein-protein BLAST' program. A yellow box highlights the 'blastp' button.

2 Task 2

Results for Job ID clustalo-l20250926-064814-0742-1069241-p1m

Tool Output

CLUSTAL O(1.2.4) multiple sequence alignment

Download

```

APOBEC3A_Homo_sapiens ..... 0
APOBEC3A_Sympalangus_syndactylus ..... 0
APOBEC3A_Nomascus_leucogenys ..... 0
APOBEC3A_Hylobates_lar ..... 0
APOBEC3A_Pongo_pygmaeus ..... 0
APOBEC3A_Pongo_abelii ..... 0
APOBEC3A_Macaca_fasciolaris ..... 0
APOBEC3A_Macaca_leonina ..... 0
APOBEC3A_Theropithecus_gelada ..... 0
APOBEC3A_Macaca_mulatta MNPQIRNPMERMYQRTFYHHENKPLYLGRSYTHLCLYEVKIRKOPSKLPWOTGVFRGQVY 69
APOBEC3A_Trachypithecus_francoisi ..... 0
APOBEC3A_Rhinopithecus_xoeliana ..... 0
APOBEC3A_Piliocolobus_tephrosceles ..... 0
APOBEC3A_Colobus_gureza ..... 0
APOBEC3A_Colobus_angolensis_palliatus ..... 0

APOBEC3A_Homo_sapiens ..... 0
APOBEC3A_Sympalangus_syndactylus ..... 0
APOBEC3A_Nomascus_leucogenys ..... 0
APOBEC3A_Hylobates_lar ..... 0
APOBEC3A_Pongo_pygmaeus ..... 0
APOBEC3A_Pongo_abelii ..... 0
APOBEC3A_Macaca_fasciolaris ..... 0
APOBEC3A_Macaca_leonina ..... 0
APOBEC3A_Theropithecus_gelada ..... 0
APOBEC3A_Macaca_mulatta SKPEHIIAECMFLSRFCGNOLPAYKRFQITWFVSNPCDCAVAKVIEFLAEHPNVTLTIST 120
APOBEC3A_Trachypithecus_francoisi ..... 0
APOBEC3A_Rhinopithecus_xoeliana ..... 0
APOBEC3A_Piliocolobus_tephrosceles ..... 0
APOBEC3A_Colobus_gureza ..... 0
APOBEC3A_Colobus_angolensis_palliatus ..... 0

```

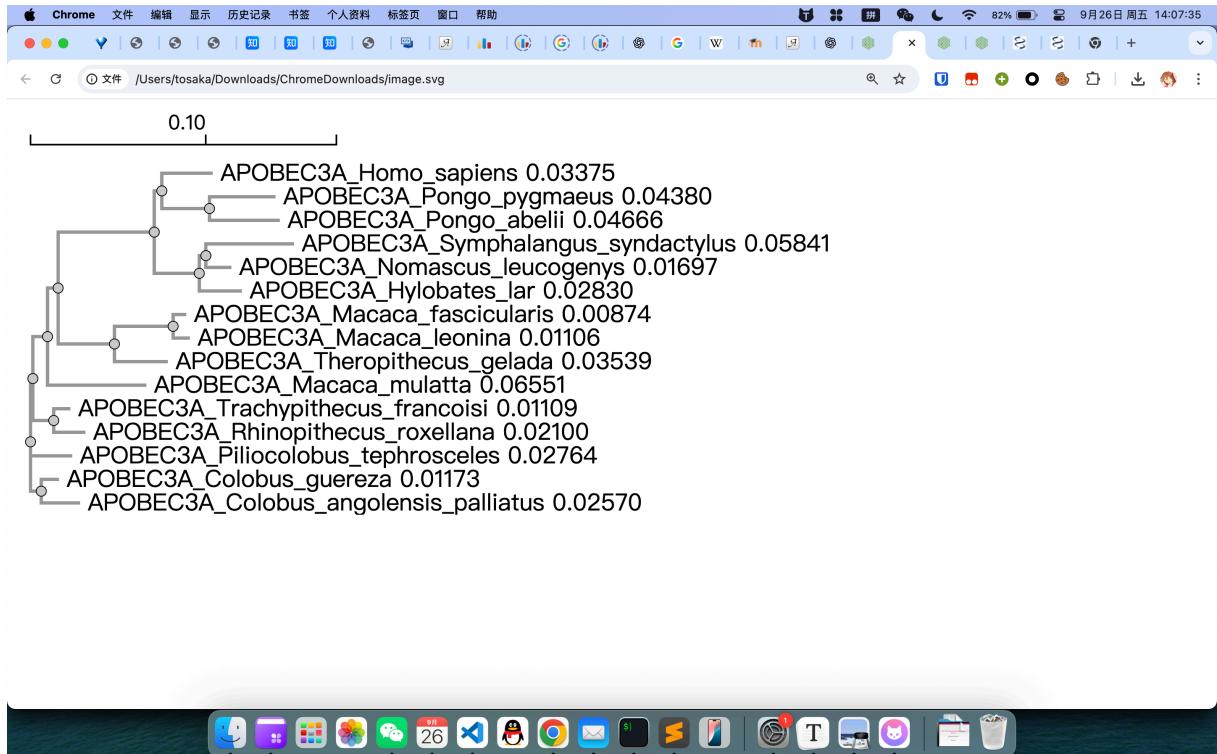
Selected 0 branches with current label

Phylogram

```

root-> APOBEC3A_Pongo_pygmaeus:0.03375
      | APOBEC3A_Pongo_abelii:0.04380
      | APOBEC3A_Pongo_abelii:0.044665
      | APOBEC3A_Nomascus_leucogenys:0.05841
      | APOBEC3A_Hylobates_lar:0.05937
      | APOBEC3A_Macaca_fasciolaris:0.03894
      | APOBEC3A_Macaca_leonina:0.0106
      | APOBEC3A_Trachypithecus_francoisi:0.02830
      | APOBEC3A_Theropithecus_gelada:0.03539
      | APOBEC3A_Macaca_mulatta:0.06551
      | APOBEC3A_Trachypithecus_francoisi:0.01109
      | APOBEC3A_Rhinopithecus_xoeliana:0.02100
      | APOBEC3A_Piliocolobus_tephrosceles:0.02764
      | APOBEC3A_Colobus_gureza:0.01173
      | APOBEC3A_Colobus_angolensis_palliatus:0.02570
      |
      | APOBEC3A_Colobus_angolensis_palliatus:0.01712;
  
```

3 Task 3



The tree demonstrates that *Homo sapiens* clusters most closely with *Pongo pygmaeus* (Bornean orangutan) and *Pongo abelii* (Sumatran orangutan), indicating a very close evolutionary relationship.

The phylogenetic tree aligns perfectly with the established evolutionary relationships among primates. The close grouping of humans (*Homo sapiens*) with orangutans (*Pongo*) reflects their shared ancestry as hominids. The more distant relationships with gibbons, macaques, and colobus monkeys are also consistent with known primate phylogeny.

This indicates that the APOBEC3A gene has evolved in a manner that mirrors the speciation events of its host organisms.

4 Task 4

The screenshot shows two instances of the Chrome browser on a Mac OS X desktop. Both instances are running the EBI's emboss_needle tool at ebi.ac.uk/jdispatcher/psa/emboss_needle.

Top Window (Job ID: emboss_needle-120250926-071721-0536-46945328-p1m):

- Input sequence:** Sequence type is set to Protein.
- Paste your first sequence here - or use the example sequence:**

```
>APOBEC3A_Homo_sapiens
MEASPASGPRHLMDPHIFTSNFNNGIGRKHTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAERLFLLVP
SLOLDPAQIYRVTWIFSWSPCFSGCAGEVRAFLQENTHVRRLRIFAARIYDPLYKEALQMLRDAGAQVSIMTYDEFKH
CWDTFVDHQGCPFQPWDGLDEHSQALSGKLQAILLNQGN
```
- Paste your second sequence here - or use the example sequence:**

```
>APOBEC3A_Pongo_pygmaeus
MEASPASGPRHLMDPVCFTSNFNNGIGRKHTYLCYEVERLDNGTWVKMDQHRGFLHNQARNPLYGLDRHAERLFGLLP
YWQLDPAQIYRVTWIFSWPCFSWGCAQVRAFLQENTHVRRLRIFAARIYDPLYKEALQMLRDAGAQVSIMTYDEFY
CWNTFVDHQGCPFQPWDGLEEHSQALSGKLQAILLNQGN
```
- Parameters:** OUTPUT FORMAT is set to pair.

Bottom Window (Job ID: emboss_needle-120250926-071721-0536-46945328-p1m):

- Results for Job ID:** emboss_needle-120250926-071721-0536-46945328-p1m
- Tool output:** The output shows the command-line parameters used for the alignment, including the sequences, matrix, gap penalties, and alignment format.

The pairwise alignment of APOBEC3A from *Homo sapiens* and *Pongo pygmaeus* yielded the following summary statistics:

- **Length:** 199 amino acids
- **Identity:** 176/199 (88.4%)
- **Similarity:** 186/199 (93.5%)

- **Gaps:** 0/199 (0.0%)
 - **Score:** 962.0

The detailed alignment output is presented in Listing 1.

```
1 #=====
2 # Aligned_sequences: 2
3 # 1: APOBEC3A_Homo_sapiens
4 # 2: APOBEC3A_Pongo_pygmaeus
5 # Matrix: EBLOSUM62
6 # Length: 199
7 # Identity: 176/199 (88.4%)
8 # Similarity: 186/199 (93.5%)
9 # Gaps: 0/199 ( 0.0%)
10 # Score: 962.0
11 #=====

12
13 APOBEC3A_Homo      1 MEASPASGPRHLMDFHITSNFNNNGIRHKTYLCYEVERLDNGTSVKMDQ 50
14          |||||||..|||||||.|||..||||||||||||||||.|||||..|||||
15 APOBEC3A_Pong       1 MEASPASGPRHLMDFPCVFTSNFNNNGIRWHKTYLCYEVERLDNGTWVKMDQ 50
16
17 APOBEC3A_Homo      51 HRGFLHNQAKNLLCFGYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSP 100
18          |||||||.:..|.|||..|||||||||.:.||..|||||||||||||||||
19 APOBEC3A_Pong       51 HRGFLHNQARNPLYGLDGRHAELRFLGLLPYWLDPDAQIYRVTFISWSP 100
20
21 APOBEC3A_Homo      101 CFSWGCAGEVR AFLQENTH VRLRIFAARIYDYDPLYKEALQMLRDAGAQV 150
22          |||||||.:|||||||..|||||||||||||||||||||||||||||
23 APOBEC3A_Pong       101 CFSWGCARQVRAFLQENTH VRLRIFAARIYDYDPLYKEALQMLRDAGAQV 150
24
25 APOBEC3A_Homo      151 SIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN 199
26          |||||||.:|||.||||||||||||.||||||||.|||..|||||
27 APOBEC3A_Pong       151 SIMTYDEFEYCWNTFVDHQGCPFQPWDGLEEHQSQALSGKLQAILLNQGN 199
```

Listing 1: EMBOSS Needle Alignment Output

Specific Amino Acid Differences: Despite the overall similarity, 23 amino acid positions differ between the two sequences. Some notable substitutions include:

- **Position 15:** Histidine (H) in humans is replaced by Cysteine (C) in orangutans.
 - **Position 61-64:** The ‘KNLL’ motif in humans corresponds to ‘RNPL’ in orangutans, showing a cluster of substitutions in this region.
 - **Position 158-159:** The ‘HC’ (Histidine-Cysteine) motif in humans is substituted by ‘EY’ (Glutamic acid-Tyrosine) in orangutans, representing two consecutive changes to residues with different biochemical properties.

It is important to note that many substitutions are conservative (e.g., Arginine ‘R’ to Lysine ‘K’ at position 179), where the replacing amino acid shares similar properties (in this case, both are basic and positively charged). Such changes are less likely to disrupt protein structure and function.