

# BIOINFORMATICS SESSION 10. PRACTICE

2023-11-06

Royal blood and order In the sequence universe

# Basic Shell Commands

```
$ cd 2023123456_HyunWoo  
$ mkdir session10  
$ cd session10
```

# Unix commands

\*\* 반드시 여러분의 working directory에서 아래 명령어를 수행하세요!

```
$ ln -s /home/biguser/tutor/session10/tax.txt .
```

```
$ ln -s /home/biguser/tutor/session10/mito.fa .
```

```
$ less tax.txt
```

```
1 root other sequences
2 Bacteria eubacteria Bacteria
6 Azorhizobium Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Xanthobacteraceae
7 Azorhizobium caulinodans Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Xanthobacteraceae; Azorhizobium
9 Buchnera aphidicola Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera
10 Cellvibrio Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae
11 Cellvibrio gilvus Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Cellvibrio
13 Dictyoglomus Bacteria; Dictyoglomi; Dictyoglomales; Dictyoglomaceae
14 Dictyoglomus thermophilum Bacteria; Dictyoglomi; Dictyoglomales; Dictyoglomaceae; Dictyoglomus
16 Methylophilus Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae
17 Methylophilus methylotrophus Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae; Methylophilus
18 Pelobacter Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; Pelobacteraceae
19 Pelobacter carbinolicus Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; Pelobacteraceae; Pelobacter
20 Phenylbacterium Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae
21 Phenylbacterium immobile Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; Phenylbacterium
22 Shewanella Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Shewanellaceae
23 Shewanella colwelliana Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Shewanellaceae; Shewanella
24 Shewanella putrefaciens Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Shewanellaceae; Shewanella
25 Shewanella hanedai Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Shewanellaceae; Shewanella
27 halophilic eubacterium NRCC 41227 Bacteria
```

# Unix commands

```
$ less mito.fa
```

```
>Dasyurus_hallucatus
GCCTTA-CTGTTAATTTTTATTAAACCTACACATGCAAGT
TTCCGCAATCCAGTGAGT-ACGCCCTTTTAACT-TGICTA
-----GAGCATAAAGGAGCTGGTATCAGGCACACT-CGAT
TGAGCAGCCCATGACACCTTGTCCAACCACA-CCCCCAGG
GGTTACAGCAGTGACTAACATTGAGCTATAAACGAAAGTT
TGA-CTAAATTATAATAAAC--AAGGGTTGGTAAATCTCG
TGCCAGCCACCAGCGGTTCATACGATTAACCCAAATTAACAG
AAAAACGGCGTAAAGGGTGTTTAAGCAT---AAACCTTGC
AA-TAAAGTTAAAGCTCAACTACGCTGTAATACGCCACAG
TTGATATTTAAATAAGCAACTTACGTGACTTTTATTAAC--
GCTGAAGACACTAAAACCTAAGGTACAACTGGGATTAGAG
ACCCCACTATGCTTAGTTCGTAACCTAGGTAATTTAAA-TA
ACAAAATTACTCGCCAGAGAACTACTAGTACTGCTTAAA
ACTCAAAGGACTTGGCGGTGCCCTAGACCCTCCTAGAGGA
GCCTGTTCCTATAATCGATAAAACCCCGATAGACCTCACCCC
TCCTCGCTC-AACAGTCTATATACCGCCATCGTCAGCTCA
CCCCAATAGGGCTTAAAAGTGAGCAAAA-TCATCAAAC-C
ATAAAAACGTTAGGTCAAGGTGTAGCATATGGAGGGGAA
GTGATGGGCTACATTTTCTATACTAGAACAT--AACGGAT
TGCTACTGAAAT----AAAGACATGAAGGAGGATTTAGT
AGTAAATAAAGATAGAGAGCTTAATTGAAATAGGCAATA
```



```
>Phascogale_tapoatafa
GCCTTA-CTGTTAATTTTTATTAGACCTACACATGCAAGT
TTCCGCTACCCAGTGAGT-ATGCCCTTTTAGCT-TTCTTA
-----GAGTATAAAGGAGTAGGTATCAGGCACACTTCTGT
GAAGTAGCCCATGACACCTAGTTTGACCACA-CCCCCAGG
GGCTACAGCAGTGACTAACATTGAGCTATAAACGAAAGTT
TGA-CTAAATCATAATAAAA--AAGGGTTGGTAAATTTTCG
TGCCAGCCACCAGCGGTTCATACGATTAACCCGAATTAACAG
AAGAACGGCGTAAAGTGTGTTTAAAGCAATAAGAATTTTCC
AAATAAGGTTAAAGATCAACTAAGCTGTAATACGCTCAGG
TTGATGTTAAAATACGCAACTTACGTGACTTTTACCCCT-
GCTGAAGACATTAAGCTAAGGTACAAACTGGGATTAGAG
ACCCCACTATGCTTAGCCGTAACCCGAGGTAGTTATA-TA
ACAAGACTATCCGCCAGAGAACTACGAGCCACTGCTTAAA
ACTCAAAGGACTTGGCGGTGCCCTAGACCCTCCTAGAGGA
GCCTGTTCCTGTAATCGATAAAACCCCGATACCTCACCTC
TCCTGGCT--GTCAGTCTATATACCGCCATCGTCAGCTCA
CCCCAATAGGGTACAAAAGTGAGCAAGA-TCATGAAAC-C
ATAAAAACGTTAGGTCAAGGTGTAGCATATGGAAAGGGAA
GTAATGGGCTACATTTTCTATATTAGAACAT--AACGGAT
```

# Unix commands

```
$ cat tax.txt
```

```
$ cat mito.fa
```

## 아래 명령어는 실제로 실행하지는 마세요. 단순 예시입니다!

```
$ cat tax.txt mito.fa > taxmito.txt
```

\*cat = concatenate

```
*cat file1 file2 file3 ... > merged_file
```

# Unix commands

```
$ grep --help
```

```
[biguser@biglab-master session10]$ grep --help
Usage: grep [OPTION]... PATTERN [FILE]...
Search for PATTERN in each FILE or standard input.
PATTERN is, by default, a basic regular expression (BRE).
Example: grep -i 'hello world' menu.h main.c

Regexp selection and interpretation:
  -E, --extended-regexp    PATTERN is an extended regular expression (ERE)
  -F, --fixed-strings      PATTERN is a set of newline-separated fixed strings
  -G, --basic-regexp       PATTERN is a basic regular expression (BRE)
  -P, --perl-regexp        PATTERN is a Perl regular expression
  -e, --regexp=PATTERN     use PATTERN for matching
  -f, --file=FILE          obtain PATTERN from FILE
  -i, --ignore-case        ignore case distinctions
  -w, --word-regexp        force PATTERN to match only whole words
  -x, --line-regexp        force PATTERN to match only whole lines
  -z, --null-data          a data line ends in 0 byte, not newline
```

# Unix commands

```
$ grep --help
```

- `-c` : 패턴이 일치하는 행의 수를 출력한다.
- `-i` : 비교시 대소문자를 구별하지 않는다.
- `-v` : 지정한 패턴과 일치하지 않는 행만 출력한다.
- `-n` : 행의 번호를 함께 출력한다.
- `-l` : 패턴이 포함된 파일의 이름을 출력한다.
- `-w` : 패턴이 전체 단어와 일치하는 행만 출력한다.

# Unix commands

```
$ grep -e ">" mito.fa
## -e, --regexp=PATTERN, use PATTERN for matching
```

```
[biguser@biglab-master session10]$ grep -e ">" mito.fa
>Dasyurus_hallucatus
>Phascogale_tapoatafa
>Sminthopsis_crassicaudata
>Myrmecobius_fasciatus
>Thylacinus_cynocephalus
>Isoodon_macrourus
>Echymipera_rufescens_australis
>Monodelphis_domestica
>Trichosurus_vulpecula
>Phalanger_interpositus
>Vombatus_ursinus
>Macropus_robustus
```

```
$ grep -c ">" mito.fa
## -c, --count, print only a count of matching lines per
FILE
```

```
[biguser@biglab-master session10]$ grep -c ">" mito.fa
16
```



# Unix commands

```
$ cut --help
```

```
[biguser@biglab-master session10]$ cut --help
Usage: cut OPTION... [FILE]...
Print selected parts of lines from each FILE to standard output.

Mandatory arguments to long options are mandatory for short options too.
  -b, --bytes=LIST      select only these bytes
  -c, --characters=LIST select only these characters
  -d, --delimiter=DELIM use DELIM instead of TAB for field delimiter
  -f, --fields=LIST     select only these fields; also print any line
                       that contains no delimiter character, unless
                       the -s option is specified
  -n                   with -b: don't split multibyte characters
  --complement         complement the set of selected bytes, characters
                       or fields
  -s, --only-delimited do not print lines not containing delimiters
  --output-delimiter=STRING use STRING as the output delimiter
                       the default is to use the input delimiter
  --help              display this help and exit
  --version            output version information and exit

Use one, and only one of -b, -c or -f. Each LIST is made up of one
range, or many ranges separated by commas. Selected input is written
in the same order that it is read, and is written exactly once.
Each range is one of:

N      N'th byte, character or field, counted from 1
N-     from N'th byte, character or field, to end of line
N-M    from N'th to M'th (included) byte, character or field
-M     from first to M'th (included) byte, character or field

With no FILE, or when FILE is -, read standard input.

Report cut bugs to bug-coreutils@gnu.org
GNU coreutils home page: <http://www.gnu.org/software/coreutils/>
General help using GNU software: <http://www.gnu.org/gethelp/>
For complete documentation, run: info coreutils 'cut invocation'
```

# Unix commands

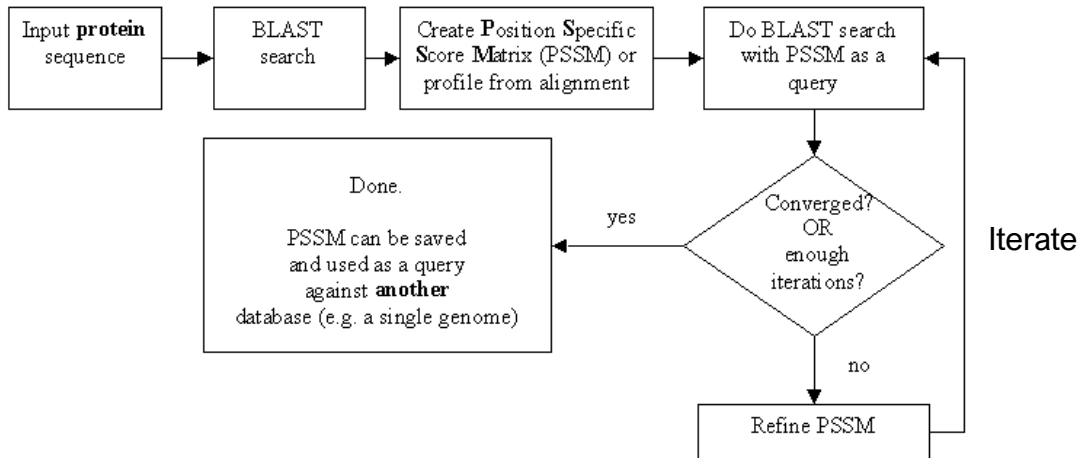
```
$ cut -f 1-4 tax.txt
```

```
##-f, --fields=LIST
```

select only these fields

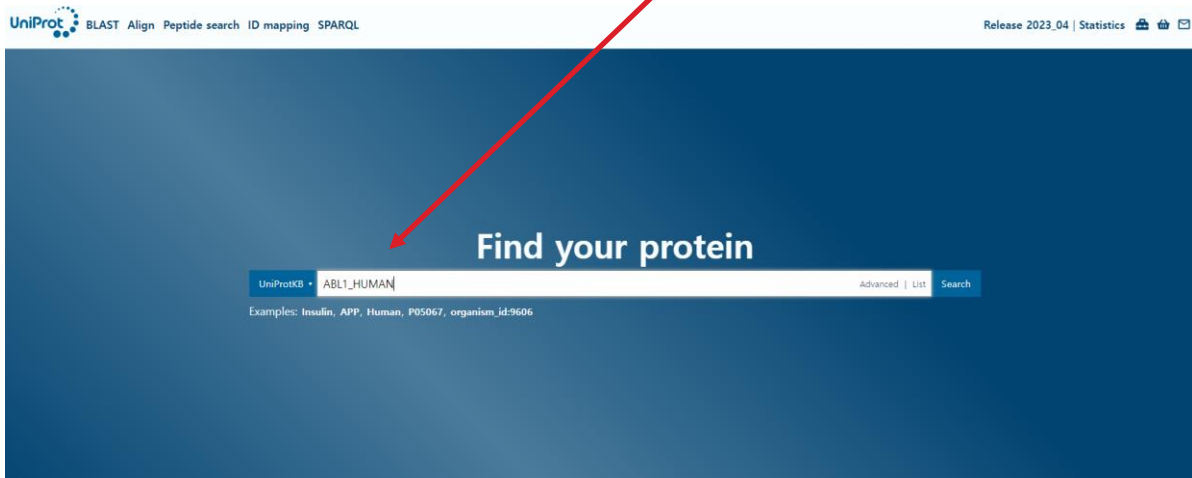
|    |    |    |                              |  |
|----|----|----|------------------------------|--|
| 1  | ro | 1  | root                         | other sequences  |
| 2  | Ba | 2  | Bacteria                     | eubacteria Bacteria  |
| 6  | Az | 6  | Azorhizobium                 | Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Xanthobacteraceae                  |
| 7  | Az | 7  | Azorhizobium caulinodans     | Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Xanthobacteraceae; Azorhizobium    |
| 9  | Bu | 9  | Buchnera aphidicola          | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera |
| 10 | C  | 10 | Cellvibrio                   | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae               |
| 11 | C  | 11 | Cellvibrio gilvus            | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Cellvibrio   |
| 13 | D  | 13 | Dictyoglomus                 | Bacteria; Dictyoglomi; Dictyoglomales; Dictyoglomaceae   |
| 14 | D  | 14 | Dictyoglomus thermophilum    | Bacteria; Dictyoglomi; Dictyoglomales; Dictyoglomaceae; Dictyoglomus                           |
| 16 | M  | 16 | Methylophilus                | Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae                |
| 17 | M  | 17 | Methylophilus methylotrophus | Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae; Methylophilus |
| 18 | P  | 18 | Pelobacter                   | Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; Pelobacteraceae             |
| 19 | P  | 19 | Pelobacter carbinolicus      | Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; Pelobacteraceae; Pelobacter |
| 20 | P  | 20 | Phenylobacterium             | Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae               |

# PSI-BLAST workflow



# Get sequence from UniProt

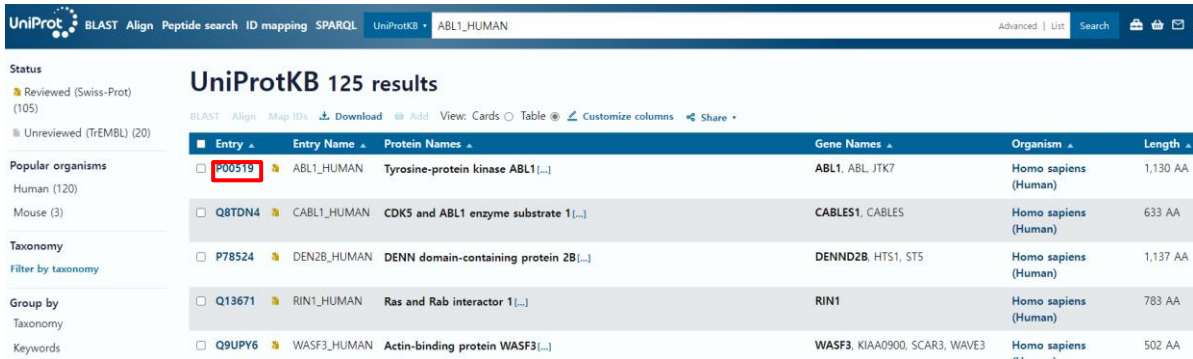
- Input: ABL1 (ABL1\_HUMAN) – SH domain “ABL1\_HUMAN” and press search



The screenshot shows the UniProt search interface. At the top left, the UniProt logo is followed by navigation links: BLAST, Align, Peptide search, ID mapping, and SPARQL. At the top right, it says "Release 2023\_04 | Statistics" with icons for a printer, a folder, and a mail icon. The main content area has a dark blue background with the text "Find your protein" in white. Below this is a search bar with "UniProtKB" on the left and "ABL1\_HUMAN" in the center. To the right of the search bar are the options "Advanced | List" and a blue "Search" button. Below the search bar, there are examples: "Examples: Insulin, APP, Human, P05067, organism\_id:9606". A red arrow points from the text "“ABL1\_HUMAN” and press search" to the "Search" button.

# Get sequence from UniProt

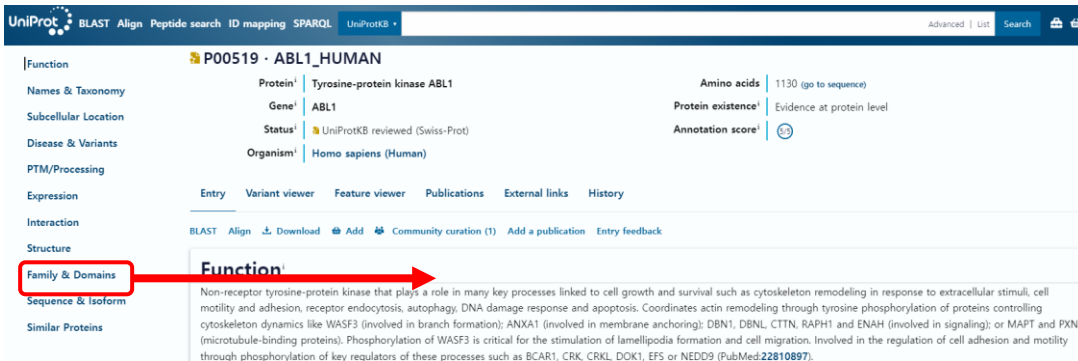
- Input: ABL1 (ABL1\_HUMAN) – SH domain



UniProtKB 125 results

| Entry                           | Entry Name  | Protein Names                           | Gene Names                    | Organism             | Length   |
|---------------------------------|-------------|---|-------------------------------|----------------------|----------|
| <input type="checkbox"/> P00519 | ABL1_HUMAN  | Tyrosine-protein kinase ABL1 [...]      | ABL1, ABL, JTK7               | Homo sapiens (Human) | 1,130 AA |
| <input type="checkbox"/> Q8TDN4 | CABL1_HUMAN | CDK5 and ABL1 enzyme substrate 1 [...]  | CABLES1, CABLES               | Homo sapiens (Human) | 633 AA   |
| <input type="checkbox"/> P78524 | DEN2B_HUMAN | DENN domain-containing protein 2B [...] | DENND2B, HTS1, ST5            | Homo sapiens (Human) | 1,137 AA |
| <input type="checkbox"/> Q13671 | RIN1_HUMAN  | Ras and Rab interactor 1 [...]          | RIN1                          | Homo sapiens (Human) | 783 AA   |
| <input type="checkbox"/> Q9UPY6 | WASF3_HUMAN | Actin-binding protein WASF3 [...]       | WASF3, KIAA0900, SCAR3, WAVE3 | Homo sapiens (Human) | 502 AA   |

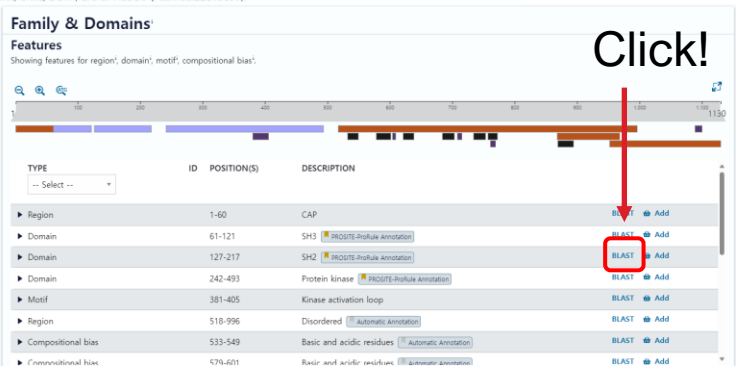
# Get sequence from UniProt



The screenshot shows the UniProt entry for P00519 - ABL1\_HUMAN. The 'Function' section is highlighted with a red box and a red arrow pointing to the 'Function' header. The 'Function' text describes the protein as a non-receptor tyrosine-protein kinase involved in various cellular processes like cell growth, survival, and cytoskeleton remodeling.

**Function**

Non-receptor tyrosine-protein kinase that plays a role in many key processes linked to cell growth and survival such as cytoskeleton remodeling in response to extracellular stimuli, cell motility and adhesion, receptor endocytosis, autophagy, DNA damage response and apoptosis. Coordinates actin remodeling through tyrosine phosphorylation of proteins controlling cytoskeleton dynamics like WASF3 (involved in branch formation); ANXA1 (involved in membrane anchoring); DBN1, DBNL, CTTN, RAPH1 and ENAH (involved in signaling); or MAPT and PXN (microtubule-binding proteins). Phosphorylation of WASF3 is critical for the stimulation of lamellipodia formation and cell migration. Involved in the regulation of cell adhesion and motility through phosphorylation of key regulators of these processes such as BCAR1, CRK, CRKL, DOK1, EFS or NEDD9 (PubMed:22810897).



The screenshot shows the 'Family & Domains' section of the UniProt entry. A red box highlights the 'BLAST' button next to the SH2 domain (residues 127-217). A red arrow points to this button with the text 'Click!'.

**Family & Domains**

Features

Showing features for region, domain, motif, compositional bias.

| TYPE               | ID      | POSITION(S) | DESCRIPTION   | BLAST | Add |
|--------------------|---------|-------------|---|-------|-----|
| Region             | 1-60    |             | CAP   | BLAST | Add |
| Domain             | 61-121  |             | SH3 <small>PROSITE-ProRule Annotation</small>                 | BLAST | Add |
| Domain             | 127-217 |             | SH2 <small>PROSITE-ProRule Annotation</small>                 | BLAST | Add |
| Domain             | 242-493 |             | Protein kinase <small>PROSITE-ProRule Annotation</small>      | BLAST | Add |
| Motif              | 381-405 |             | Kinase activation loop  | BLAST | Add |
| Region             | 518-996 |             | Disordered <small>Automatic Annotation</small>                | BLAST | Add |
| Compositional bias | 533-549 |             | Basic and acidic residues <small>Automatic Annotation</small> | BLAST | Add |
| Compositional bias | 579-601 |             | Basic and acidic residues <small>Automatic Annotation</small> | BLAST | Add |

# Get sequence from UniProt

Click!

## BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4\_HUMAN or UPI0000000001).

UniProt IDs

OR

Enter one or more sequences (5 max). You may also load from a text file.

```
esp|P00519|ABL1_HUMAN|127-217 Tyrosine-protein kinase ABL1 OS:Homo sapiens OX:9606 GN:ABL1 PE:1 SV:4  
MYHGPSRMA AEYLLSSGIN GSFLVRESLS SPGQRSISLR YEGRVVHYRL NIASDGKLYV  
KSESFRNTLA ELWRRSTVA DGLITTLHP A
```

Copy the text



BLAST

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

**BLAST: Basic Local Alignment Search Tool**

2023년 8월 24일 - Establish taxonomy for uncultured or environmental sequences. The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. ...

## Protein BLAST

QuickBLASTP is an accelerated version of BLASTP that is ver...

## Nucleotide BLAST

Nucleotide BLAST - BLAST: Basic Local Alignment Search Tool

## BLAST Algorithm Help

BLAST Algorithm Help - BLAST: Basic Local Alignment Search Tool

## Blastx

Blastx - BLAST: Basic Local Alignment Search Tool

## Primer-BLAST Help

Primer-BLAST was developed at NCBI to help users make pri...

## Global Alignment

Local alignments algorithms (such as BLAST) are most often ...

[blast.ncbi.nlm.nih.gov](https://blast.ncbi.nlm.nih.gov)의 검색 결과만 보기

Your input contains 1 sequence

# PSI-BLAST

NIH National Library of Medicine  
National Center for Biotechnology Information

BLAST®

Check out the ClusteredNR database on BLAST+ [Learn more](#) [Give us feedback](#)

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](#)

BLAST Quick Start guides!  
Need some help getting started with BLAST?  
Thu, 22 Jun 2023 [More BLAST news...](#)

**Web BLAST**

**Nucleotide BLAST**  
nucleotide > nucleotide

**blastx**  
translated nucleotide > protein

**tblastn**  
protein > translated nucleotide

**Protein BLAST**  
protein > protein

**BLAST Genomes**

Enter organism common name, scientific name, or tax id [Search](#)

Human Mouse Rat Microbes

Click!

## 1. Paste the copied sequence of SH2

BLASTP programs search protein databases using a protein query

Enter Query Sequence

Enter accession number(s), GI(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

From

To

Or, upload file [파일 선택](#) [선택된 파일 전송](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Databases  Standard databases (nr etc.)  Experimental databases [Try experimental clustered nr database](#) [For more info see What is clustered nr?](#)

Compare  Select to compare standard and experimental database [?](#)

**Standard**

Database  [?](#)

Organism   exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude  Models (XM/XP)  Non-redundant RefSeq proteins (WP)  Uncultured/environmental sample sequences

**Program Selection**

Algorithm

Quick BLAST (Accelerated protein protein BLAST)

**PSI-BLAST (Position-Specific Iterated BLAST)**

PHI-BLAST (Pattern Hit Initiated BLAST)

DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

**BLAST** Search database nr using PSI-BLAST (Position-Specific Iterated BLAST)

Show results in a new window

2. Check PSI-BLAST

3. Run BLAST



# PSI-BLAST

Job Title **sp|P00519|ABL1\_HUMAN|127-217 Tyrosine-protein...**  
 RID **KY782.DS016** Search expires on 10-31 16:10 pm [Download All](#) ▾  
 Program **PSI-BLAST Iteration 1** Citation ▾  
 Database **nr** [See details](#) ▾  
 Query ID **klQuery\_85319**  
 Description **sp|P00519|ABL1\_HUMAN|127-217 Tyrosine-protein kinase...**  
 Molecule type **amino acid**  
 Query Length **91**  
 Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism only top 20 will appear  exclude  
 Type common name, binomial, taxid or group name  
 + [Add organisms](#)

Percent Identity  to  E value  to  Query Coverage  to

PSI-BLAST incl. threshold      
 0.005

Run PSI-BLAST Iteration 2  
 Number of sequences  500

Compare these results against the new Clustered nr database [?](#)

**Descriptions** | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

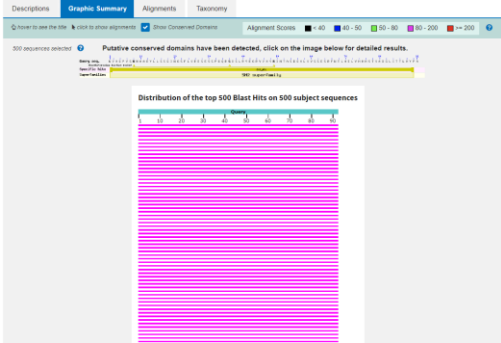
Sequences producing significant alignments Download ▾ Select columns ▾ Show 500 ▾

500 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

**Sequences with E value BETTER than threshold**

select all 500 sequences selected

| PSI-BLAST Iteration 1   |                               |           |             |             |         |           |         |                |                                     |                                     |             |
|---|-------------------------------|-----------|-------------|-------------|---------|-----------|---------|----------------|-------------------------------------|-------------------------------------|-------------|
| Description   | Scientific Name               | Max Score | Total Score | Query Cover | E value | Per Ident | Acc Len | Accession      | Select for PSI blast                | Used for PSSM                       | newly added |
| <input checked="" type="checkbox"/> Chain A, Proto-oncogene tyrosine-protein kinase ABL1 (Homo sapiens)   | <i>Homo sapiens</i>           | 185       | 185         | 100%        | 2e-58   | 100.00%   | 112     | JGCM_A         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> Chain B, Tyrosine-protein kinase ABL1 (Homo sapiens)                  | <i>Homo sapiens</i>           | 185       | 185         | 100%        | 3e-58   | 100.00%   | 121     | SDCC_E         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> Chain A, C-ABL TYROSINE KINASE SH2 DOMAIN (Homo sapiens)              | <i>Homo sapiens</i>           | 184       | 184         | 100%        | 4e-58   | 100.00%   | 109     | IAE2_A         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> Chain A, Tyrosine-protein kinase ABL1 (Homo sapiens)                  | <i>Homo sapiens</i>           | 184       | 184         | 100%        | 4e-58   | 100.00%   | 123     | JTHL_A         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> Chain A, ABL TYROSINE KINASE (Homo sapiens)                           | <i>Homo sapiens</i>           | 185       | 185         | 100%        | 1e-57   | 100.00%   | 163     | ZMHL_A         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> ABL, erythrocytosis 1 non-receptor tyrosine kinase (Larus michadalis) | <i>Larus michadalis</i>       | 185       | 185         | 100%        | 2e-57   | 100.00%   | 175     | AUSJ3946.1     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> klQuery_85319_1343 (Homo sapiens)                                     | <i>Homo sapiens</i>           | 187       | 187         | 100%        | 2e-57   | 100.00%   | 235     | GAM33989.1     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)                           | <i>Caenorhabditis elegans</i> | 186       | 186         | 100%        | 3e-57   | 100.00%   | 218     | XP_013204503.2 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> klQuery_85319_1343 (Homo sapiens)                                     | <i>Homo sapiens</i>           | 186       | 186         | 100%        | 5e-57   | 100.00%   | 260     | GAM33929.1     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |
| <input checked="" type="checkbox"/> klQuery_85319_1343 (Homo sapiens)                                     | <i>Homo sapiens</i>           | 188       | 188         | 100%        | 5e-57   | 100.00%   | 307     | GAM31077.1     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |             |



|   |                               |     |     |      |       |                |     |              |                                     |                                     |  |
|---|-------------------------------|-----|-----|------|-------|----------------|-----|--------------|-------------------------------------|-------------------------------------|--|
| <input checked="" type="checkbox"/> tyrotheoretical protein K5549_D11052/Canara hircu1    | <i>Canara hircu1</i>          | 191 | 191 | 100% | 8e-54 | 100.00%        | 162 | KAJ1077275.1 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Homo sapiens</i>           | 191 | 191 | 100% | 1195  | XP_030775418.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> PHECCTED, tyrosine-protein kinase ABL1 (Homo sapiens) | <i>Bufo marinus</i>           | 191 | 191 | 100% | 1123  | XP_019133910.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Quus americana</i>         | 191 | 191 | 100% | 1109  | XP_054704998.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Quus flavescens</i>        | 191 | 191 | 100% | 1143  | XP_057820264.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Acetabularia clausenii</i> | 191 | 191 | 100% | 1144  | XP_051822920.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> PHECCTED, tyrosine-protein kinase ABL1 (Homo sapiens) | <i>Caenorhabditis elegans</i> | 191 | 191 | 100% | 1126  | XP_009684831.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrotheoretical protein HGM15179_309445/Canara hircu1 | <i>Canara hircu1</i>          | 191 | 191 | 100% | 1127  | IR217646.1     |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Lonicera xylosteum</i>     | 191 | 191 | 100% | 1127  | XP_021399177.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Therapsid</i>              | 191 | 191 | 100% | 1127  | XP_021399177.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Siamese cat</i>            | 191 | 191 | 100% | 1130  | XP_023020918.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> tyrosine-protein kinase ABL1 (Homo sapiens)           | <i>Hirundo</i>                | 191 | 191 | 100% | 1130  | XP_033021326.1 |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |
| <input checked="" type="checkbox"/> ABL1 (Homo sapiens)                                   | <i>Felis tigris</i>           | 191 | 191 | 100% | 1130  | XP098878.1     |     |              | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |  |

Run PSI-BLAST Iteration 2 with max number of sequences  500

# PSI-BLAST

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 500

500 sequences selected  sequences newly added this iteration [GenPlot](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSAViewer](#)

Sequences with E value BETTER than threshold

select all 500 sequences selected [Skip to the first new sequence](#) PSI-BLAST iteration 2

| Description   | Scientific Name          | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession                      | Select for PSI blast                | Used to build PSSM                  | Newly added                         |
|---|--------------------------|-----------|-------------|-------------|---------|------------|----------|--------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 (Parthenococcus)</a>                                      | <i>Parthenococcus</i>    | 187       | 187         | 100%        | 5e-55   | 100.00%    | 308      | <a href="#">XP_04980264.1</a>  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">AB1.1 kinase (Cochlosoma macleodensis)</a>  | <i>Cochlosoma</i>        | 187       | 187         | 100%        | 9e-56   | 100.00%    | 323      | <a href="#">NC026983.1</a>     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 system X2 (Cyclotus spiro)</a>                            | <i>Cyclotus spiro</i>    | 191       | 191         | 100%        | 9e-56   | 100.00%    | 545      | <a href="#">KAF599842.1</a>    | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">PREDICTED Xtreme-ortem kinase AB_1 like (Hemite carolinensis)</a>                   | <i>Acetabularia</i>      | 187       | 187         | 100%        | 1e-55   | 100.00%    | 341      | <a href="#">XP_009322772.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 (Puffin lactiflora)</a>                                   | <i>Puffin lactiflora</i> | 189       | 189         | 100%        | 1e-55   | 98.96%     | 476      | <a href="#">XP_026028108.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">PREDICTED Xtreme-ortem kinase AB_1 system X2 (Lecosphaera scutellata)</a>           | <i>Lecosphaera</i>       | 191       | 191         | 100%        | 1e-55   | 100.00%    | 573      | <a href="#">XP_015225238.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 (Laminothus socialis)</a>                                 | <i>Laminothus</i>        | 188       | 188         | 100%        | 2e-55   | 100.00%    | 442      | <a href="#">KAE8280546.1</a>   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Chain A Xtreme-ortem kinase AB1.1 (Hemite saevata)</a>                              | <i>Hemite saevata</i>    | 188       | 188         | 100%        | 2e-55   | 100.00%    | 448      | <a href="#">H5291.6</a>        | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Chain A Photo-encapsome Xtreme-ortem kinase AB1.1 (S. S/OFCRM) (Hemite saevata)</a> | <i>Hemite saevata</i>    | 189       | 189         | 100%        | 2e-55   | 100.00%    | 495      | <a href="#">2E03.6</a>         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Chain A Xtreme-ortem kinase AB1.1 (Hemite saevata)</a>                              | <i>Hemite saevata</i>    | 189       | 189         | 100%        | 2e-55   | 100.00%    | 495      | <a href="#">5M24.6</a>         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 system X1 (Mux muscicola)</a>                             | <i>Mux muscicola</i>     | 191       | 191         | 100%        | 2e-55   | 100.00%    | 594      | <a href="#">XP_006482684.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtremofactin-ortem LIGNO_02153060 (Lacifera dentata)</a>                            | <i>Lacifera dent</i>     | 187       | 187         | 100%        | 2e-55   | 100.00%    | 451      | <a href="#">KAF7465166.1</a>   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">PREDICTED Xtreme-ortem kinase AB1.1 system X1 (Lecosphaera scutellata)</a>          | <i>Lecosphaera</i>       | 190       | 190         | 100%        | 3e-55   | 100.00%    | 609      | <a href="#">XP_015225238.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 system X3 (Lapocrothochus obtusoides)</a>                 | <i>Lapocrothochus</i>    | 189       | 189         | 100%        | 3e-55   | 100.00%    | 639      | <a href="#">XP_020896322.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtremofactin-ortem OHT09_005314 (Mammota monax)</a>                                 | <i>Mammota monax</i>     | 184       | 184         | 100%        | 3e-55   | 100.00%    | 344      | <a href="#">KAF7465052.1</a>   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Chain A Photo-encapsome Xtreme-ortem kinase AB1.1 (Mux muscicola)</a>               | <i>Mux muscicola</i>     | 188       | 188         | 100%        | 3e-55   | 100.00%    | 495      | <a href="#">10FK.6</a>         | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 system X2 (Corythochthys altilateralis)</a>               | <i>Corythochthys</i>     | 195       | 195         | 100%        | 3e-55   | 100.00%    | 1071     | <a href="#">XP_029175548.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> <a href="#">PREDICTED Xtreme-ortem kinase AB1.1 (Cochlosoma rotunda)</a>                        | <i>Cochlosoma rot</i>    | 192       | 192         | 100%        | 3e-55   | 100.00%    | 606      | <a href="#">XP_026817336.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 like system X3 (Mastacembelus armatus)</a>                | <i>Mastacembelus</i>     | 189       | 189         | 100%        | 4e-55   | 100.00%    | 961      | <a href="#">XP_026153074.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |
| <input checked="" type="checkbox"/> <a href="#">PREDICTED Xtreme-ortem kinase AB_1 like system X2 (Xenocarpus cometi)</a>           | <i>Xenocarpus</i>        | 194       | 194         | 100%        | 4e-55   | 100.00%    | 1065     | <a href="#">XP_019723551.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> <a href="#">Xtreme-ortem kinase AB1.1 system X4 (Diverothus angul)</a>                          | <i>Diverothus a.</i>     | 189       | 189         | 100%        | 4e-55   | 100.00%    | 531      | <a href="#">XP_031120933.1</a> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |                                     |

Job title: sp|P00519|127-217

RID [GXGD7IG8015](#) (Expires on 05-04 10:13 am)

Query ID [kcl|Query\\_339575](#)  
 Description [sp|P00519|127-217](#)  
 Molecule type amino acid  
 Query Length 91

Database Name [nr](#)  
 Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
 Program [BLASTP 2.6.1+ p-Citation](#)

**No new sequences were found above the 0.005 threshold**

# Data download

[HOME](#) | [SEARCH](#) | [BROWSE](#) | [FTP](#) | [HELP](#) | [ABOUT](#)

## Pfam 35.0 (November 2021, 19632 entries)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

### [QUICK LINKS](#)

- [SEQUENCE SEARCH](#)
- [VIEW A PFAM ENTRY](#)
- [VIEW A CLAN](#)
- [VIEW A SEQUENCE](#)
- [VIEW A STRUCTURE](#)
- [KEYWORD SEARCH](#)
- [JUMP TO](#)

### YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...

Analyze your protein sequence for Pfam matches

View Pfam annotation and alignments

See groups of related entries

Look at the domain organisation of a protein sequence

Find the domains on a PDB structure

Query Pfam by keywords

[Go](#) [Example](#)

Enter any type of accession or ID to jump to the page for a Pfam entry or clan, UniProt sequence, PDB structure, etc.

Or view the [help](#) pages for more information

# Data download

## Index of /pub/databases/Pfam

| Name                             | Last modified    | Size | Description |
|----------------------------------|------------------|------|-------------|
| <a href="#">Parent Directory</a> | -                | -    | -           |
| <a href="#">AntiFam/</a>         | 2021-09-01 14:43 | -    | -           |
| <a href="#">RoseTTAfold_aln/</a> | 2021-11-11 15:16 | -    | -           |
| <a href="#">Tools/</a>           | 2017-03-08 09:16 | -    | -           |
| <a href="#">current_release/</a> | 2023-09-12 14:14 | -    | -           |
| <a href="#">mappings/</a>        | 2023-10-25 03:00 | -    | -           |
| <a href="#">papers/</a>          | 2021-04-01 13:25 | -    | -           |
| <a href="#">releases/</a>        | 2023-09-14 11:22 | -    | -           |
| <a href="#">vm/</a>              | 2013-05-22 06:25 | -    | -           |

## Index of /pub/databases/Pfam/current\_release

| Name   | Last modified    | Size | Description         |
|--|------------------|------|---------------------|
| <a href="#">Parent Directory</a>                   | -                | -    | -                   |
| <a href="#">Pfam-A.clans.tsv.gz</a>                | 2023-09-12 13:30 | 377K |                     |
| <a href="#">Pfam-A.dead.gz</a>                     | 2023-09-12 13:30 | 24K  |                     |
| <a href="#">Pfam-A.fasta.gz</a>                    | 2023-09-12 13:31 | 4.8G |                     |
| <a href="#">Pfam-A.full.gz</a>                     | 2023-09-12 13:32 | 18G  |                     |
| <a href="#">Pfam-A.full.uniprot.gz</a>             | 2023-09-12 13:36 | 33G  |                     |
| <a href="#">Pfam-A.hmm.dat.gz</a>                  | 2023-09-12 13:36 | 540K |                     |
| <a href="#">Pfam-A.hmm...</a>                      | 2023-09-12 13:36 | 286M |                     |
| <a href="#">Pfam-A.regio...</a>                    | 2023-09-12 13:36 | 3.5G | 새 탭에서 링크 열기         |
| <a href="#">Pfam-A.regio...</a>                    | 2023-09-12 13:36 | 9.6G | 새 창에서 링크 열기         |
| <a href="#">Pfam-A.rp15...</a>                     | 2023-09-12 13:38 | 1.2G | InPrivate 창에서 링크 열기 |
| <a href="#">Pfam-A.rp35...</a>                     | 2023-09-12 13:38 | 4.0G | 분할 화면 창에서 링크 열기     |
| <a href="#">Pfam-A.rp55...</a>                     | 2023-09-12 13:39 | 8.2G | (으)로 링크 저장          |
| <a href="#">Pfam-A.rp75...</a>                     | 2023-09-12 13:40 | 13G  |                     |
| <a href="#">Pfam-A.seed</a>                        | 2023-09-12 13:40 | 144M | 링크 복사               |
| <a href="#">Pfam-B.tgz</a>                         | 2023-09-12 13:41 | 2.5G | 질액선에 추가             |
| <a href="#">Pfam-C.gz</a>                          | 2023-09-12 13:41 | 178K |                     |
| <a href="#">Pfam.version</a>                       | 2023-09-12 13:41 | 115  | 공유                  |
| <a href="#">active_site.d...</a>                   | 2023-09-12 13:41 | 12K  | 경사                  |
| <a href="#">database.file...</a>                   | 2023-09-12 13:41 | 658  |                     |
| <a href="#">diff.gz</a>                            | 2023-09-12 13:41 | 250K |                     |
| <a href="#">md5_checksums</a>                      | 2023-09-12 13:41 | 1.4K |                     |
| <a href="#">pdbmap.gz</a>                          | 2023-09-12 13:41 | 4.9M |                     |
| <a href="#">pfamseq.gz</a>                         | 2023-09-12 13:43 | 19G  |                     |
| <a href="#">proteomes/</a>                         | 2023-09-12 14:54 | -    |                     |
| <a href="#">relnotes.txt</a>                       | 2023-09-12 13:43 | 25K  |                     |
| <a href="#">swisspfam.gz</a>                       | 2023-09-12 13:43 | 3.3G |                     |
| <a href="#">trees.tgz</a>                          | 2023-09-12 13:43 | 24M  |                     |
| <a href="#">uniprot.gz</a>                         | 2023-09-12 13:52 | 51G  |                     |
| <a href="#">uniprot_reference_proteomes.dat.gz</a> | 2023-09-12 13:58 | 51G  |                     |
| <a href="#">uniprot_sprot.dat.gz</a>               | 2023-09-12 13:58 | 615M |                     |
| <a href="#">uniprot_trembl.dat.gz</a>              | 2023-09-12 14:14 | 150G |                     |
| <a href="#">userman.txt</a>                        | 2023-09-12 14:14 | 17K  |                     |

# Data download

```
$ wget [copy the URL link]
```

→ 실제로 실행하지는 말고 아래 link 명령어를 사용하세요

```
biguser@R440 session10]$ wget http://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/Pfam-A.hmm.gz
--2023-10-30 17:25:13-- http://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/Pfam-A.hmm.gz
Resolving ftp.ebi.ac.uk (ftp.ebi.ac.uk)... 193.62.193.165
Connecting to ftp.ebi.ac.uk (ftp.ebi.ac.uk)|193.62.193.165|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 299797995 (286M) [application/x-gzip]
Saving to: 'Pfam-A.hmm.gz'
```

```
1% [>] 5,829,145 677KB/s eta 8m 37s
```



```
$ ln -s /home/biguser/tutor/session10/Pfam-A.hmm .
```

# Build Pfam-A

\$ less Pfam-A.hmm

```

HMMER3/T [3.1b2 | February 2015]
NAME      1-cysPrx_C
ACC       PF19417.11
DESC      C-terminal domain of 1-Cys peroxiredoxin
LENG      40
ALPH      amino
RF        no
RM        no
CONS      yes
CS        yes
MAP       yes
DATE      Wed Feb 24 18:37:46 2021
NSEQ      40
EFFN      17.426758
CKSUM     4086680297
GA        21.10 21.10;
TC        21.10 21.10;
NC        21.00 21.00;
BM        hmmbuild HMM.ann SEED.ann
SM        hmmssearch -Z 57096847 -E 1000 --cpu 4 HMM pfamseq
STATS     LOCAL MSV      -7.5463  0.71948
          LOCAL VITERBI  -7.8624  0.71948
          LOCAL FORWARD  -4.3303  0.71948
HMM
          A          C          D          E          F          G          H          I          K          L          M          N          P          Q          R          S          T          V          W          Y
COMPO    m->m    m->l    m->d    l->m    l->l    d->m    d->d
5         2.28846  4.31288  2.83393  2.63913  3.90855  2.69988  3.89812  3.33401  2.56310  2.85023  3.99954  3.22924  2.52123  2.90328  3.31238  2.94055  2.70512  2.59551  3.49266  3.8271
3         2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.29354  2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.77519  2.98518  4.58477  3.6150
1         0.00226  6.48754  7.20989  0.61958  0.77255  0.00000  *
7         0.29666  6.14436  6.78514  6.79783  7.06332  2.55785  7.22049  6.57837  6.66651  6.27638  3.20757  5.91223  5.83978  6.69238  6.58162  2.20136  4.83343  5.59959  8.41086  7.4310
3         2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.29354  2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.77519  2.98518  4.58477  3.6150
2         0.00226  6.48754  7.20989  0.61958  0.77255  0.48576  0.95510
2         4.59591  5.92009  6.57211  5.96147  1.92899  5.81035  6.10135  2.33093  5.75927  0.69439  2.86149  5.97020  6.07717  5.70793  5.72916  5.13924  4.81708  2.59612  3.18569  3.3584
3         2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.29354  2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.77519  2.98518  4.58477  3.6150
3         0.00226  6.48754  7.20989  0.61958  0.77255  0.48576  0.95510
5         4.81290  7.05274  3.71096  4.47757  6.00126  5.41623  3.72993  5.92180  2.06538  3.59487  6.10993  4.89014  5.75663  0.42291  2.54802  4.76779  4.95656  5.56452  7.24472  6.0861
3         2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.29354  2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.77519  2.98518  4.58477  3.6150
3         0.00226  6.48754  7.20989  0.61958  0.77255  0.48576  0.95510
8         2.64100  5.20125  5.84007  3.33362  1.00025  5.06096  2.72027  3.71332  5.01717  1.75203  2.65498  5.22200  5.43290  5.15329  5.03455  4.37913  1.00041  2.31249  5.90246  2.6329
4         2.68618  4.42225  2.77519  2.73123  3.46354  2.40513  3.72494  3.29354  2.67741  2.69355  4.24690  2.90347  2.73739  3.18146  2.89801  2.37887  2.77519  2.98518  4.58477  3.6150

```

# Get sequence from swissprot

```
$ cp /home/biguser/tutor/session10/bc_seqid.txt .
```

```
$ less bc_seqid.txt
```

```
FA7_HUMAN  
FA8_HUMAN  
FA9_HUMAN  
FA10_HUMAN  
FA11_HUMAN  
FA12_HUMAN  
TF_HUMAN  
PLMN_HUMAN  
TPA_HUMAN  
UROK_HUMAN  
THRB_HUMAN  
KLKB1_HUMAN  
HGF_HUMAN  
HGFA_HUMAN  
bc_seqid.txt (END)
```

```
$ ln -s /home/biguser/your_directory/session6/swissprot* .
```

```
$ blastdbcmd -entry_batch bc_seqid.txt -db swissprot -long_seqs > clotting.fa
```

# hmmsearch from Pfam-A.hmm

## HMM profile indexing

```
$ hmmpress Pfam-A.hmm
```

```
[biguser@R440 session10]$ hmmpress Pfam-A.hmm
Working... done.
Pressed and indexed 19179 HMMs (19179 names and 19179 accessions).
Models pressed into binary file: Pfam-A.hmm.h3m
SSI index for binary model file: Pfam-A.hmm.h3i
Profiles (MSV part) pressed into: Pfam-A.hmm.h3f
Profiles (remainder) pressed into: Pfam-A.hmm.h3p
```



# hmmsearch from Pfam-A.hmm

```
[biguser@r440 session10]$ hmmsearch -h
# hmmsearch :: search sequence(s) against a profile database
# HMMER 3.3.2 (Nov 2020); http://hmmer.org/
# Copyright (C) 2020 Howard Hughes Medical Institute.
# Freely distributed under the BSD open source license.
# -----
Usage: hmmsearch [-options] <hmmdb> <seqfile>

Basic options:
  -h : show brief help on version and usage

Options controlling output:
  -o <f>          : direct output to file <f>, not stdout
  --tblout <f>   : save parseable table of per-sequence hits to file <f>
  --domtblout <f> : save parseable table of per-domain hits to file <f>
  --pfamtblout <f> : save table of hits and domains to file, in Pfam format <f>
  --acc          : prefer accessions over names in output
  --noall        : don't output alignments, so output is smaller
  --notextv      : unlimit ASCII text output line width
  --textw <n>    : set max width of ASCII text output lines [120] (n>=120)
```

## Searching HMM profile with a query sequence

```
$ hmmsearch --domtblout clotting.tab Pfam-A.hmm clotting.fa
```

```
[biguser@r440 session10]$ hmmsearch --domtblout clotting.tab Pfam-A.hmm clotting.fa
# hmmsearch :: search sequence(s) against a profile database
# HMMER 3.3.2 (Nov 2020); http://hmmer.org/
# Copyright (C) 2020 Howard Hughes Medical Institute.
# Freely distributed under the BSD open source license.
# -----
# query sequence file:      clotting.fa
# target HMM database:      Pfam-A.hmm
# per-dom hits tabular output: clotting.tab
# -----
Query:      sp|P08709|FAJ_HUMAN [L=466]
Description: Coagulation factor VII 05-Homo sapiens OX-9606 (h-F7 PE-1 SV-1)
Scores for complete sequence (score includes all domains):
-----
E-value score bias  E-value score bias  exp N Model      Description
-----
3.2e-59 200.5  0.1  4.7e-59 200.0  0.1  1.3  1  Trypsin
8.3e-19  67.6  6.3  8.8e-19  67.5  4.9  1.7  2  Gla
0.4e-09  35.8  8.0  1.7e-08  35.0  0.0  1.4  1  DUF1980
5.6e-08  33.9  11.4  5.6e-08  33.0  11.4  2.9  2  Pxx inhibition
6e-05  23.4  4.9  6e-05  23.4  4.9  2.9  2  EGF
0.00012 22.8  1.2  0.00012 22.0  1.2  3.7  4  EGF
0.0012  19.5  2.1  0.0012  19.5  2.1  3.5  2  NEGf
----- Inclusion threshold -----
0.0  10.2  24.5  0.21  12.1  8.8  3.2  2  EGF_2
8.8  6.7  25.7  1.4  9.2  11.1  2.7  2  EGF_3
EGF domain

Domain annotation for each model (and alignments):
>> Trypsin Trypsin
# score bias c-Evalue l-Evalue hmmfrom hmm to alifrom all to envfrom env to acc
-----
1 1 200.0 0.1 4.7e-62 4.7e-59 1 221 11 213 447 213 447 0.94
```

# Result of hmmscan

\$ less clotting.tab

| # | target name    | accession  | tlen | query name          | accession | qlen | --- full sequence --- |       |      |   | --- this domain --- |          |          |       | hmm coord |      | all coord |      | env coord |      |      |      |
|---|----------------|------------|------|---------------------|-----------|------|-----------------------|-------|------|---|---------------------|----------|----------|-------|-----------|------|-----------|------|-----------|------|------|------|
| # |                |            |      |                     |           |      | E-value               | score | bias | # | of                  | c-Evalue | i-Evalue | score | bias      | from | to        | from | to        | from | to   |      |
|   | Trypsin        | PF00809.28 | 221  | sp P08709 FA7_HUMAN | -         | 466  | 3.2e-59               | 200.5 | 0.1  | 1 | 1                   | 2.2e-62  | 4.7e-59  | 200.0 | 0.1       | 1    | 221       | 213  | 447       | 213  | 447  | 0.94 |
|   | Gla            | PF00594.22 | 41   | sp P08709 FA7_HUMAN | -         | 466  | 8.3e-19               | 67.6  | 6.3  | 1 | 2                   | 2.4      | 5.2e+03  | -2.2  | 0.0       | 20   | 36        | 48   | 64        | 46   | 64   | 0.87 |
|   | Gla            | PF00594.22 | 41   | sp P08709 FA7_HUMAN | -         | 466  | 8.3e-19               | 67.6  | 6.3  | 2 | 2                   | 4.1e-22  | 8.8e-19  | 67.5  | 4.9       | 1    | 40        | 65   | 104       | 65   | 105  | 0.98 |
|   | DUF1986        | PF09342.13 | 116  | sp P08709 FA7_HUMAN | -         | 466  | 9.4e-09               | 35.8  | 0.0  | 1 | 1                   | 7.8e-12  | 1.7e-08  | 35.8  | 0.0       | 1    | 100       | 224  | 327       | 224  | 335  | 0.78 |
|   | FXa_inhibition | PF14670.8  | 36   | sp P08709 FA7_HUMAN | -         | 466  | 5.6e-08               | 33.0  | 11.4 | 1 | 2                   | 1.2      | 2.5e+03  | -1.0  | 6.0       | 8    | 29        | 115  | 138       | 108  | 147  | 0.74 |
|   | FXa_inhibition | PF14670.8  | 36   | sp P08709 FA7_HUMAN | -         | 466  | 5.6e-08               | 33.0  | 11.4 | 2 | 2                   | 2.6e-11  | 5.6e-08  | 33.0  | 11.4      | 1    | 36        | 151  | 187       | 151  | 187  | 0.94 |
|   | EGF            | PF00808.29 | 32   | sp P08709 FA7_HUMAN | -         | 466  | 6e-05                 | 23.4  | 4.9  | 1 | 2                   | 2.8e-08  | 6e-05    | 23.4  | 4.9       | 1    | 31        | 110  | 139       | 110  | 140  | 0.90 |
|   | EGF            | PF00808.29 | 32   | sp P08709 FA7_HUMAN | -         | 466  | 6e-05                 | 23.4  | 4.9  | 2 | 2                   | 0.1      | 2.1e+02  | 2.4   | 7.2       | 8    | 29        | 155  | 179       | 151  | 180  | 0.82 |
|   | cEGF           | PF12662.9  | 24   | sp P08709 FA7_HUMAN | -         | 466  | 0.00012               | 22.0  | 1.2  | 1 | 4                   | 0.99     | 2.1e+03  | -1.2  | 0.3       | 15   | 21        | 117  | 123       | 111  | 123  | 0.83 |
|   | cEGF           | PF12662.9  | 24   | sp P08709 FA7_HUMAN | -         | 466  | 0.00012               | 22.0  | 1.2  | 2 | 4                   | 0.14     | 3e+02    | 1.5   | 1.0       | 1    | 8         | 127  | 134       | 127  | 142  | 0.77 |
|   | cEGF           | PF12662.9  | 24   | sp P08709 FA7_HUMAN | -         | 466  | 0.00012               | 22.0  | 1.2  | 3 | 4                   | 5.8e-08  | 0.00012  | 22.0  | 1.2       | 3    | 20        | 171  | 188       | 170  | 189  | 0.95 |
|   | cEGF           | PF12662.9  | 24   | sp P08709 FA7_HUMAN | -         | 466  | 0.00012               | 22.0  | 1.2  | 4 | 4                   | 3.5      | 7.5e+03  | -2.9  | 0.2       | 8    | 19        | 390  | 400       | 390  | 401  | 0.54 |
|   | NEGF           | PF12661.9  | 22   | sp P08709 FA7_HUMAN | -         | 466  | 0.0012                | 19.5  | 2.1  | 1 | 2                   | 5.5e-07  | 0.0012   | 19.5  | 2.1       | 1    | 20        | 115  | 134       | 115  | 136  | 0.96 |
|   | NEGF           | PF12661.9  | 22   | sp P08709 FA7_HUMAN | -         | 466  | 0.0012                | 19.5  | 2.1  | 2 | 2                   | 0.12     | 2.6e+02  | 2.6   | 4.8       | 7    | 22        | 162  | 178       | 150  | 178  | 0.90 |
|   | EGF_2          | PF07974.15 | 32   | sp P08709 FA7_HUMAN | -         | 466  | 0.8                   | 10.2  | 24.5 | 1 | 2                   | 0.0001   | 0.21     | 12.1  | 8.8       | 1    | 32        | 110  | 141       | 110  | 141  | 0.85 |
|   | EGF_2          | PF07974.15 | 32   | sp P08709 FA7_HUMAN | -         | 466  | 0.8                   | 10.2  | 24.5 | 2 | 2                   | 0.092    | 2e+02    | 2.6   | 7.7       | 5    | 28        | 148  | 179       | 144  | 187  | 0.70 |
|   | EGF_3          | PF12947.9  | 36   | sp P08709 FA7_HUMAN | -         | 466  | 8.8                   | 6.7   | 25.7 | 1 | 2                   | 0.069    | 1.5e+02  | 2.8   | 6.0       | 6    | 31        | 113  | 138       | 110  | 141  | 0.86 |
|   | EGF_3          | PF12947.9  | 36   | sp P08709 FA7_HUMAN | -         | 466  | 8.8                   | 6.7   | 25.7 | 2 | 2                   | 0.00068  | 1.4      | 9.2   | 11.1      | 1    | 36        | 151  | 187       | 151  | 187  | 0.89 |
|   | F5_F8_type_C   | PF00754.27 | 127  | sp P00451 FA8_HUMAN | -         | 2351 | 8.5e-55               | 184.4 | 0.3  | 1 | 2                   | 6.5e-28  | 1.8e-24  | 86.4  | 0.0       | 1    | 127       | 2055 | 2185      | 2055 | 2185 | 0.88 |
|   | F5_F8_type_C   | PF00754.27 | 127  | sp P00451 FA8_HUMAN | -         | 2351 | 8.5e-55               | 184.4 | 0.3  | 2 | 2                   | 1.2e-30  | 3.3e-27  | 25.2  | 0.2       | 1    | 127       | 2208 | 2342      | 2208 | 2342 | 0.94 |
|   | Cu-oxidase_3   | PF07732.17 | 119  | sp P00451 FA8_HUMAN | -         | 2351 | 3.5e-16               | 59.4  | 0.0  | 1 | 3                   | 8.5e-08  | 0.00023  | 21.2  | 0.0       | 24   | 114       | 90   | 197       | 88   | 202  | 0.78 |
|   | Cu-oxidase_3   | PF07732.17 | 119  | sp P00451 FA8_HUMAN | -         | 2351 | 3.5e-16               | 59.4  | 0.0  | 2 | 3                   | 3.3e-08  | 9.2e-05  | 22.6  | 0.0       | 11   | 114       | 454  | 572       | 444  | 577  | 0.76 |
|   | Cu-oxidase_3   | PF07732.17 | 119  | sp P00451 FA8_HUMAN | -         | 2351 | 3.5e-16               | 59.4  | 0.0  | 3 | 3                   | 0.00015  | 0.4      | 10.8  | 0.0       | 24   | 90        | 1777 | 1841      | 1772 | 1880 | 0.77 |
|   | Cu-oxidase_2   | PF07731.16 | 137  | sp P00451 FA8_HUMAN | -         | 2351 | 3.2e-15               | 56.1  | 7.5  | 1 | 4                   | 0.3      | 8.1e+02  | -0.2  | 0.0       | 35   | 78        | 95   | 136       | 82   | 149  | 0.65 |
|   | Cu-oxidase_2   | PF07731.16 | 137  | sp P00451 FA8_HUMAN | -         | 2351 | 3.2e-15               | 56.1  | 7.5  | 2 | 4                   | 0.0021   | 5.7      | 6.8   | 3.3       | 39   | 133       | 268  | 345       | 230  | 347  | 0.80 |
|   | Cu-oxidase_2   | PF07731.16 | 137  | sp P00451 FA8_HUMAN | -         | 2351 | 3.2e-15               | 56.1  | 7.5  | 3 | 4                   | 1.3e-05  | 0.035    | 14.0  | 0.0       | 86   | 133       | 680  | 727       | 660  | 730  | 0.88 |
|   | Cu-oxidase_2   | PF07731.16 | 137  | sp P00451 FA8_HUMAN | -         | 2351 | 3.2e-15               | 56.1  | 7.5  | 4 | 4                   | 8.7e-11  | 2.6e-07  | 30.7  | 0.0       | 6    | 134       | 1920 | 2036      | 1915 | 2039 | 0.93 |
|   | Cu-oxidase     | PF00394.24 | 159  | sp P00451 FA8_HUMAN | -         | 2351 | 2.7e-10               | 40.7  | 0.0  | 1 | 2                   | 8.3e-10  | 2.3e-06  | 27.0  | 0.0       | 7    | 158       | 224  | 348       | 218  | 349  | 0.91 |
|   | Cu-oxidase     | PF00394.24 | 159  | sp P00451 FA8_HUMAN | -         | 2351 | 2.7e-10               | 40.7  | 0.0  | 2 | 2                   | 0.00031  | 0.84     | 9.8   | 0.0       | 7    | 94        | 1905 | 1984      | 1900 | 2037 | 0.95 |
|   | CytadhesinP1   | PF12378.10 | 260  | sp P00451 FA8_HUMAN | -         | 2351 | 0.0019                | 18.1  | 0.1  | 1 | 1                   | 1.8e-06  | 0.0049   | 16.7  | 0.1       | 81   | 175       | 954  | 1050      | 948  | 1068 | 0.83 |

# Searching with custome HMM profiles

## Building HMM profile

```
$ clustalw2 clotting.fa  
$ hmmbuild clotting.hmm clotting.aln
```

```
$ less clotting.hmm
```

```
HMMER3/F [3.3.2 | Nov 2020]  
NAME clotting  
LENG 2956  
ALPH amino  
1F no  
1M no  
CONS yes  
CS no  
MAP yes  
DATE Mon Oct 30 17:43:27 2023  
MSD 14  
EFIN 1_261230  
CKSUM 3813876788  
STATS LOCAL REV -13.7261 0.69433  
STATS LOCAL VITERBI -15.0680 0.69433  
STATS LOCAL FORWARD -7.6657 0.69433  
HMM  
A C D E F G H I K L M N P Q R S T V W Y  
m->m m->l m->d l->m l->l d->m d->d  
3 COMPO 2.50789 4.13958 2.92648 2.69851 3.26267 2.88968 3.60744 3.29354 2.67246 2.48341 3.65953 3.86617 3.25964 3.86259 2.93713 2.59136 2.81196 2.69529 4.48089 3.4832  
2 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
1 0.83139 3.87272 4.59587 0.61958 0.77255 0.00000 +  
2 2.96932 4.51802 4.84964 3.64387 3.20771 3.83424 4.37815 2.38333 3.41115 1.80230 1.37745 3.92655 4.29481 3.79299 3.61645 3.31848 3.27460 2.38949 5.04860 3.8104  
1 m 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
2 0.83139 3.87272 4.59587 0.61958 0.77255 0.48576 0.95510  
2 2.88081 4.86667 2.90295 2.65761 4.03528 3.36611 3.81594 3.66681 2.47836 3.16834 4.19612 3.11737 3.91952 1.23359 2.76256 2.92851 3.17774 3.39564 5.30721 4.0089  
2 q 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
2 0.83139 3.87272 4.59587 0.61958 0.77255 0.48576 0.95510  
3 3.83128 4.44988 4.33616 3.95191 3.37428 4.86698 4.67994 1.07759 3.78886 1.96019 3.38953 4.19721 4.49883 4.12828 3.97659 3.56217 3.33842 1.84576 5.26521 4.0835  
3 i . . . . 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
3 0.83139 3.87272 4.59587 0.61958 0.77255 0.48576 0.95510  
4 2.95238 5.06366 2.48751 1.02501 4.40988 3.27413 3.86886 3.86383 2.74141 3.48087 4.45482 2.93286 3.07871 3.07647 3.15034 2.93470 3.25659 3.54762 5.50140 4.3035  
4 e . . . . 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
3 0.83139 3.87272 4.59587 0.61958 0.77255 0.48576 0.95510  
5 3.15548 4.61764 4.18992 3.82971 3.13669 3.98251 4.48536 2.33967 3.59746 0.87488 3.15844 4.10516 4.41778 3.96572 3.77351 3.55331 3.44970 2.37709 5.01514 3.7385  
5 l . . . . 2.68618 4.42225 2.77519 2.73123 3.46354 2.49513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.6150  
3 0.83139 3.87272 4.59587 0.61958 0.77255 0.48576 0.95510  
6 2.31951 4.22945 3.26052 3.82223 4.07648 2.96532 4.12911 3.59910 3.89681 3.31928 4.24529 3.24702 3.78413 3.44529 3.38289 1.89868 2.79515 3.13415 5.44894 4.1599
```

# Searching with custom HMM profiles

## HMM profile indexing

```
hmmcompress clotting.hmm
```

```
[biguser@R440 session10]$ hmmcompress clotting.hmm
Working... done.
Pressed and indexed 1 HMMs (1 names).
Models pressed into binary file: clotting.hmm.h3m
SSI index for binary model file: clotting.hmm.h3i
Profiles (MSV part) pressed into: clotting.hmm.h3f
Profiles (remainder) pressed into: clotting.hmm.h3p
```

## Searching HMM profile with a query sequence

```
hmmsearch --domtblout query.tab clotting.hmm query.fa
```

# Code 1 parse\_hmmscan.ipynb

```
import re
import sys

print('protein_name\tlen\tdomain\tbegin\tend\tValue')

for line in open("clotting.tab", 'r'):
    if not re.search('^##', line): # avoid all lines beginning
                                    # with the '#' character
        col = re.split(' +', line)
        domname = col[0]
        protname = col[3]
        protname = re.sub('.*##|', '', protname)
        length = col[5]
        value = float(col[12])
        begin = col[17]
        end = col[18]
        if value < 1e-5:
            print(protname, '\t', length, '\t', domname, end = '')
            print('\t', begin, '\t', end, '\t', str(value))
            pass
        else:
            pass
```

# Code 1 parse\_hmmscan.ipynb

| protein_name | len  | domain         | begin | end  | Value        |
|--------------|------|----------------|-------|------|--------------|
| FA7_HUMAN    | 466  | Trypsin        |       | 213  | 447 4.7e-59  |
| FA7_HUMAN    | 466  | Gla            | 65    | 104  | 8.8e-19      |
| FA7_HUMAN    | 466  | DUF1986        |       | 224  | 327 1.7e-08  |
| FA7_HUMAN    | 466  | FXa_inhibition |       | 151  | 187 5.6e-08  |
| FA8_HUMAN    | 2351 | F5_F8_type_C   |       | 2055 | 2185 1.8e-24 |
| FA8_HUMAN    | 2351 | F5_F8_type_C   |       | 2208 | 2342 3.3e-27 |
| FA8_HUMAN    | 2351 | Cu-oxidase_2   |       | 1920 | 2036 2.4e-07 |
| FA8_HUMAN    | 2351 | Cu-oxidase     |       | 224  | 348 2.3e-06  |
| FA9_HUMAN    | 461  | Trypsin        |       | 227  | 454 1.3e-69  |
| FA9_HUMAN    | 461  | Gla            | 52    | 92   | 1.1e-20      |
| FA9_HUMAN    | 461  | FXa_inhibition |       | 134  | 170 2.2e-09  |
| FA9_HUMAN    | 461  | EGF            | 97    | 127  | 7.5e-06      |
| FA10_HUMAN   | 488  | Trypsin        |       | 235  | 462 7e-70    |
| FA10_HUMAN   | 488  | Gla            | 45    | 85   | 2e-21        |
| FA10_HUMAN   | 488  | FXa_inhibition |       | 129  | 164 6.9e-09  |
| FA10_HUMAN   | 488  | EGF            | 90    | 120  | 3.1e-07      |
| FA11_HUMAN   | 625  | Trypsin        |       | 388  | 618 4e-73    |
| FA11_HUMAN   | 625  | PAN_1          | 21    | 103  | 2e-07        |
| FA11_HUMAN   | 625  | PAN_1          | 119   | 193  | 3.4e-08      |
| FA11_HUMAN   | 625  | PAN_1          | 203   | 283  | 2.7e-08      |
| FA11_HUMAN   | 625  | PAN_1          | 296   | 371  | 3.5e-08      |
| FA11_HUMAN   | 625  | PAN_4          | 299   | 349  | 4.3e-07      |
| FA12_HUMAN   | 615  | Trypsin        |       | 373  | 609 3.6e-63  |
| FA12_HUMAN   | 615  | Kringle        |       | 217  | 295 5.7e-24  |
| FA12_HUMAN   | 615  | fn2            | 47    | 88   | 6.8e-17      |
| FA12_HUMAN   | 615  | EGF            | 98    | 129  | 6.2e-06      |
| FA12_HUMAN   | 615  | EGF            | 178   | 207  | 5.5e-08      |
| FA12_HUMAN   | 615  | fn1            | 135   | 170  | 1.7e-10      |

# Exercise

- Merge the BCR\_HUMAN.fa, ABL1\_HUMAN.fa, BCR\_ABL1\_fusion\_HUMAN.fa into single file using “cat” command. Then run “hmmsearch” to search for the similar proteins (domains) of query proteins. After obtaining the output file named as “bcr\_abl1.tab”, run the “parse\_hmmsearch.py” to get summary of the hmmsearch search.

아래 명령어를 실행해서 BCR\_HUMAN.fa, ABL1\_HUMAN.fa, BCR\_ABL1\_fusion\_HUMAN.fa 파일들을 여러분들의 directory에 복사해 주세요

```
$ cp /home/biguser/tutor/session10/bcr_abl_sequences/* .
```