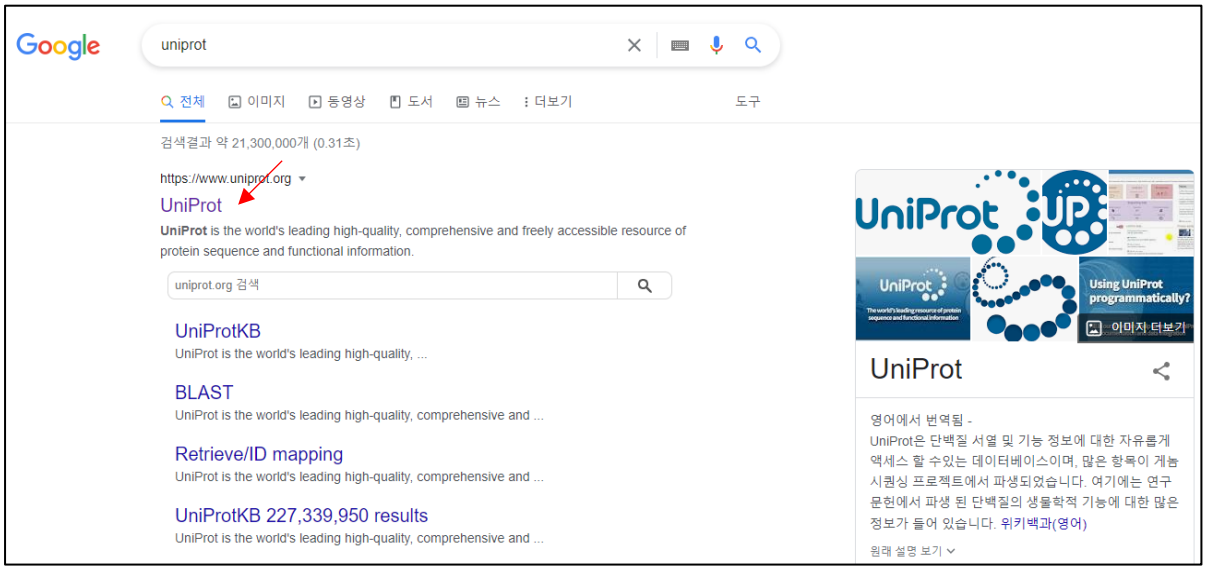
Molecular Biology Laboratory

Bioinformatics and Genomics Lab.

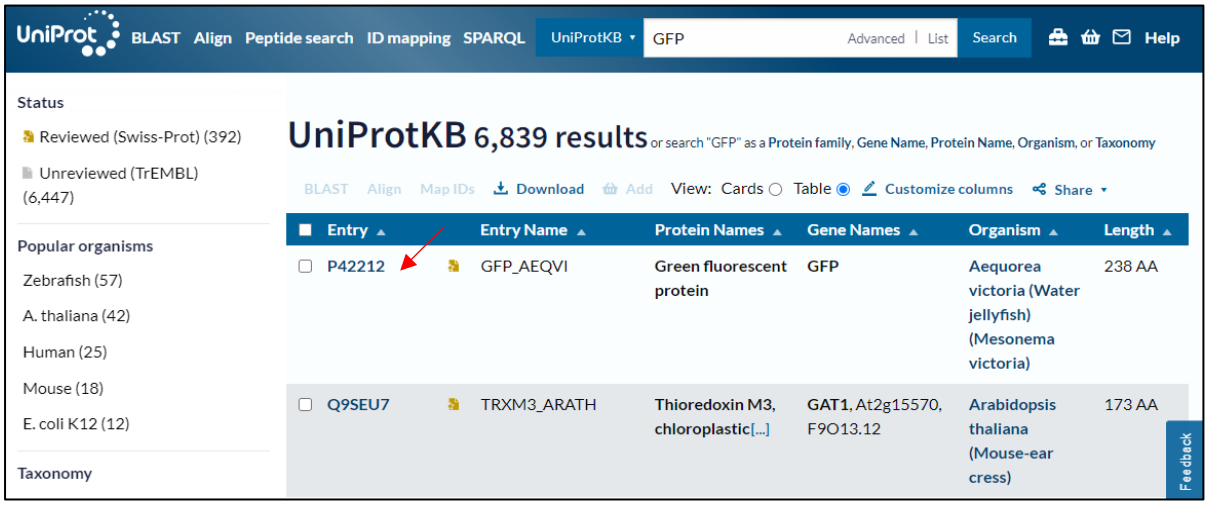
Week2. Protein Sequence Alignment (BLAST, Clustal Omega) & Domain Search

1. Practice how to use UniProt database

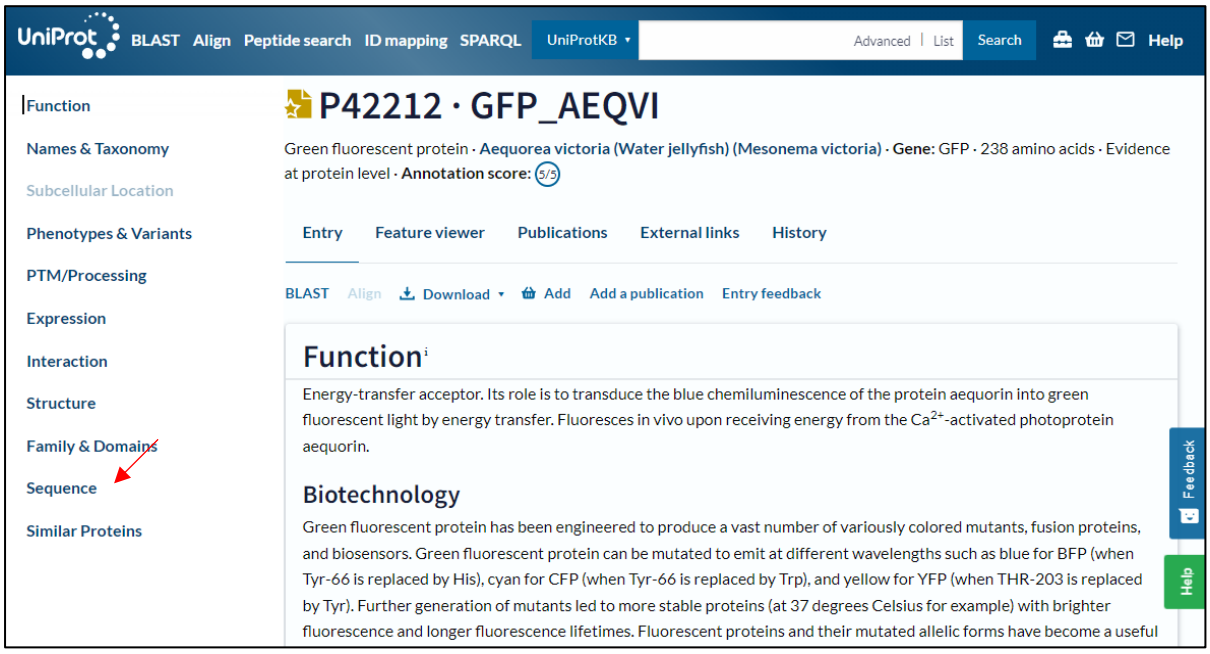
- Find information on GFP and get the amino acid sequence
  - ➢ Search "uniprot" in google and access UniProt.

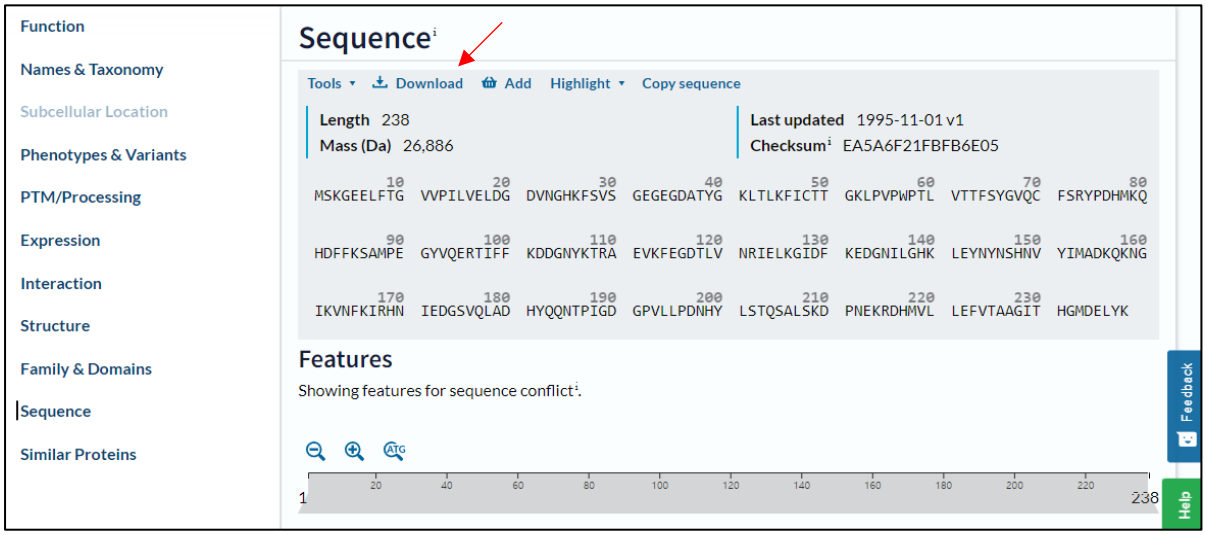


  - ➢ Search "GFP" and select 1st one (P42212).



- We can find information on GFP protein. To get the amino acid sequence, click "Sequence"- "Download".

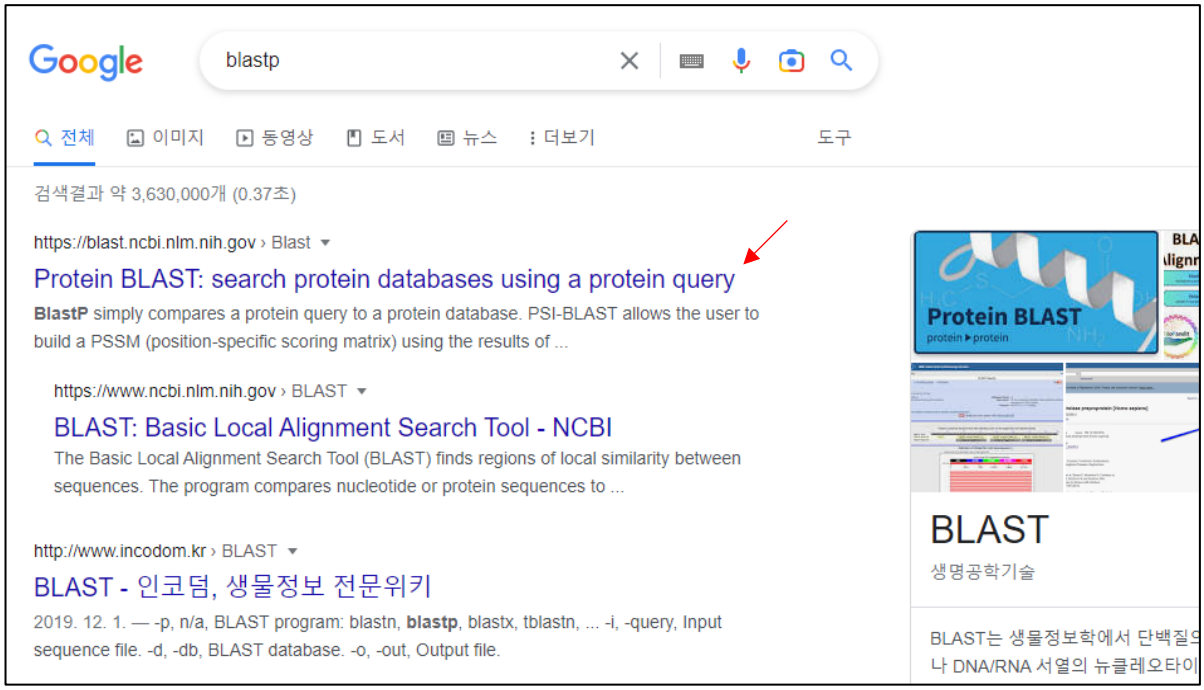




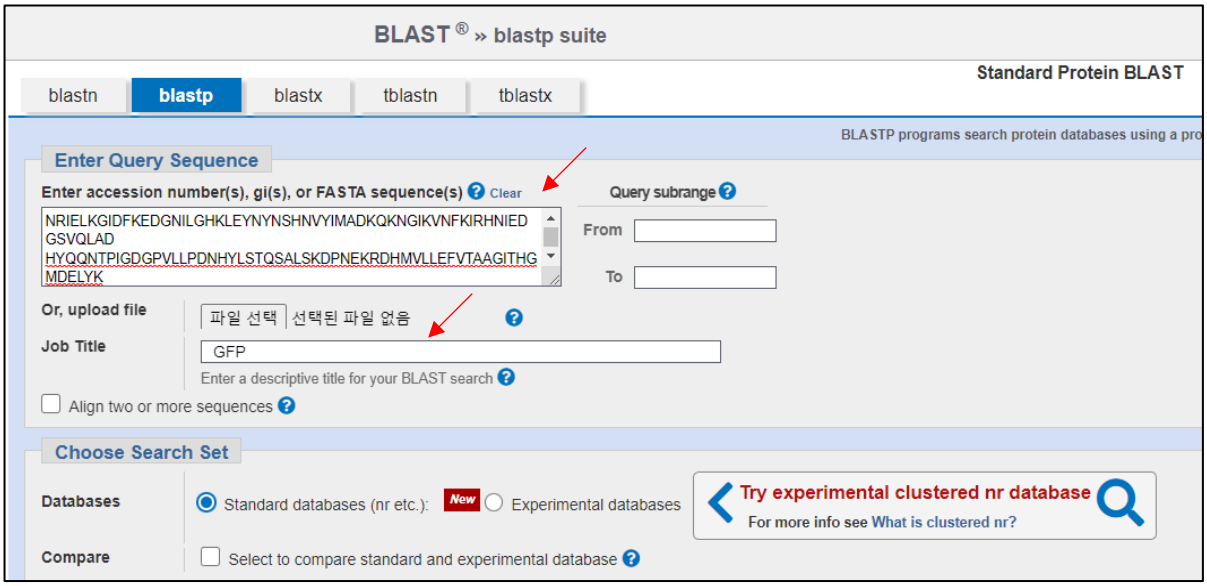

```
>sp|P42212|GFP_AEQVI Green fluorescent protein OS=Aequorea victoria OX=6100 GN=GFP PE=1 SV=1
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK
```

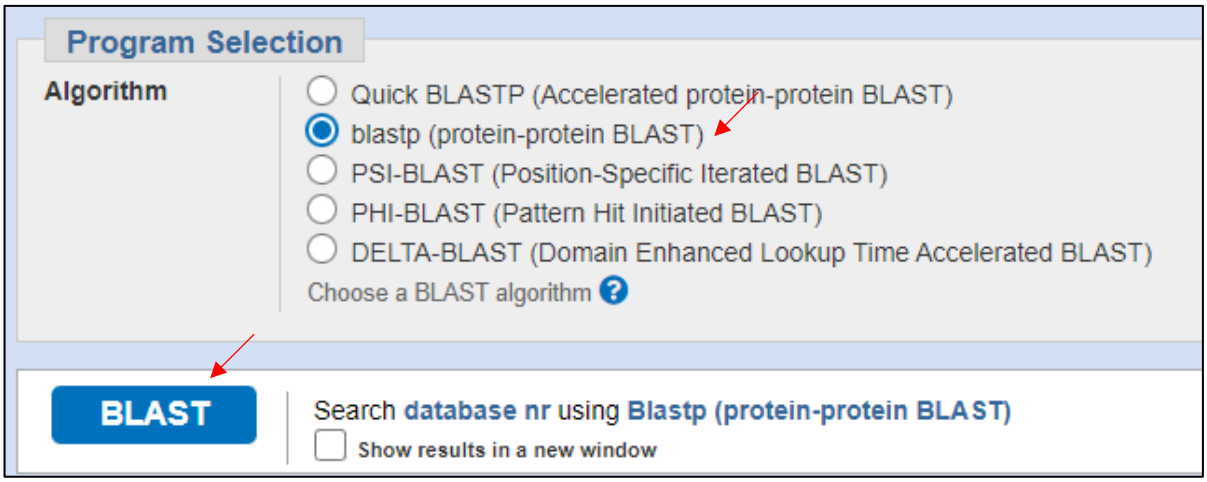
2. Practice how to use BLASTP

- Try sequence alignment of GFP
  - Search "blastp" in google and access BLASTP

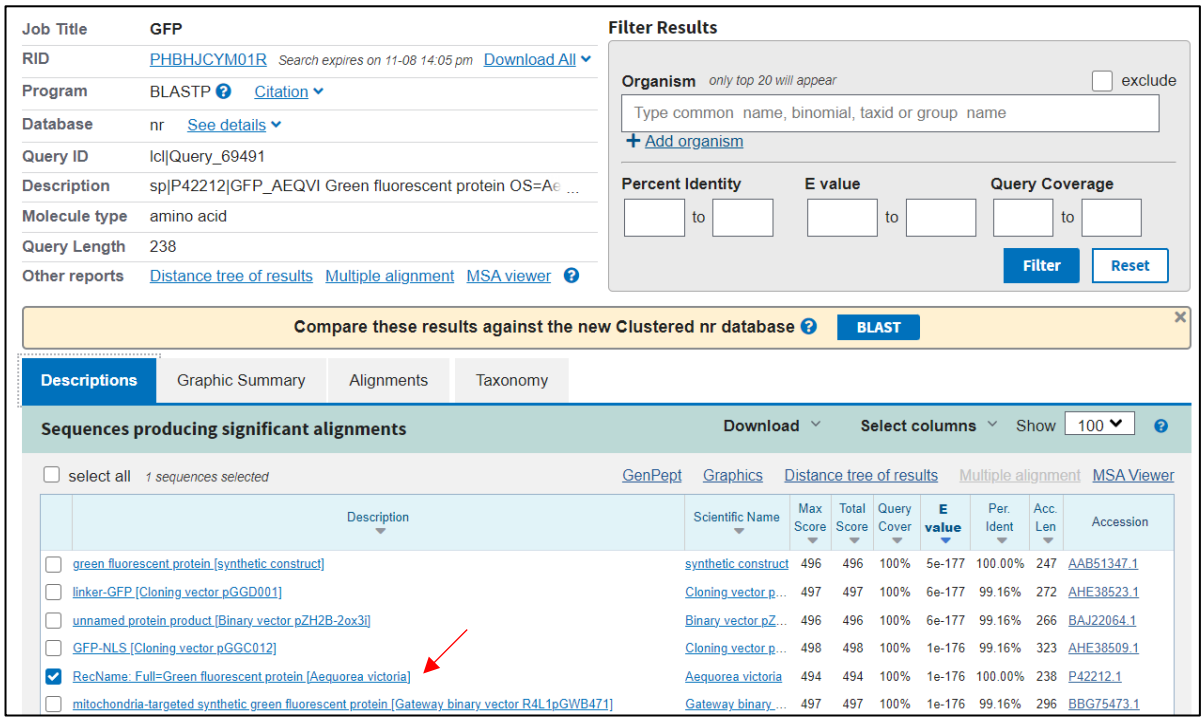


  - Copy and paste the GFP sequence to the query sequence box and set job title. Select program algorithm as "blastp" and click "BLAST" button to run the tool



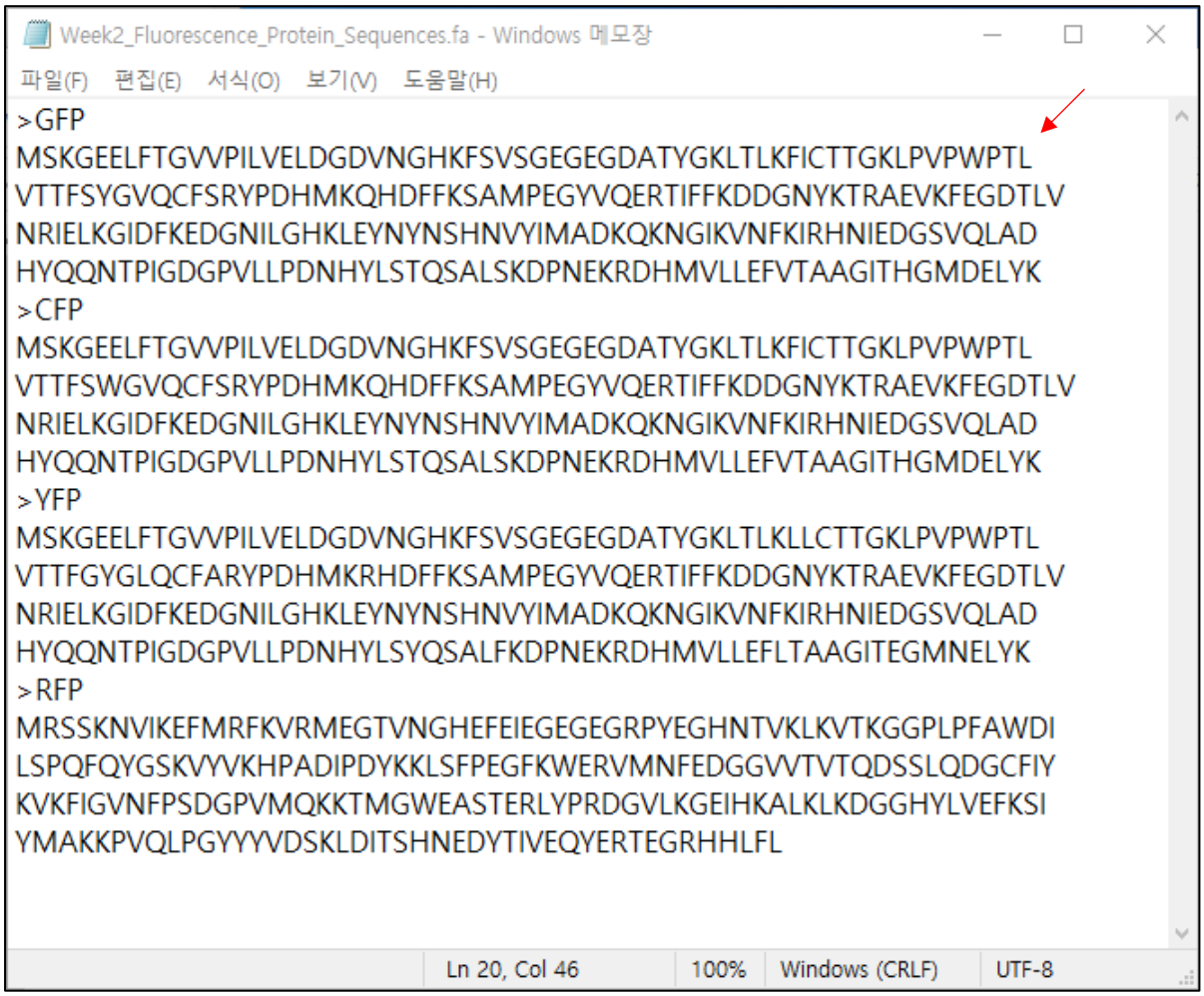


➢ Select GFP of "*Aequorea victoria*" and check information (Query cover, Per. Ident, Graphic Summary etc.)

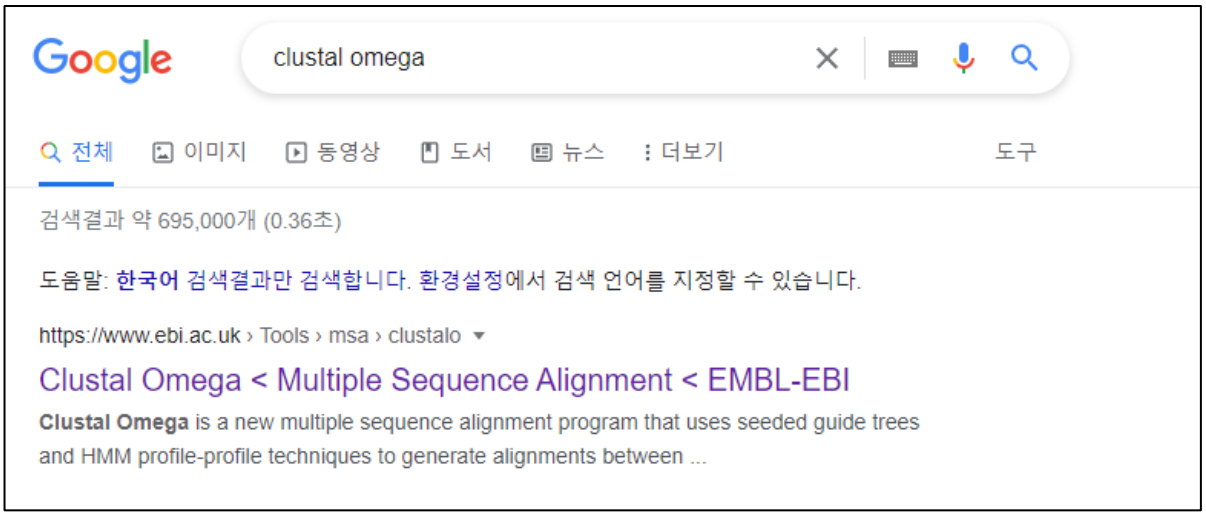


3. Practice how to use Clustal Omega

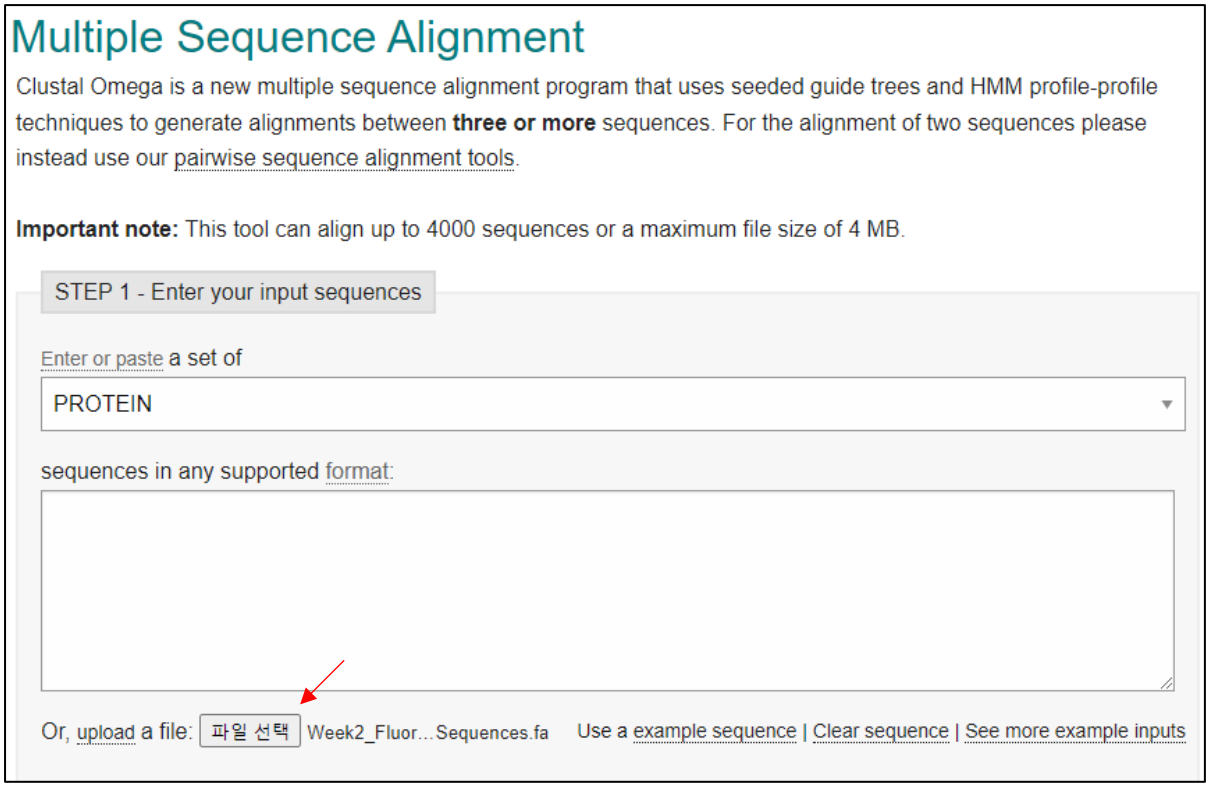
- Try multiple sequence alignment (MSA) of GFP, CFP, YFP, and RFP
  - Copy and paste sequence to the given "FASTA" file (Week2_Fluorescence_Protein_Sequences_for_Upload.fa). Edit header part of GFP as ">GFP" for MSA.



  - Search "clustal omega" in google and access Clustal Omega.



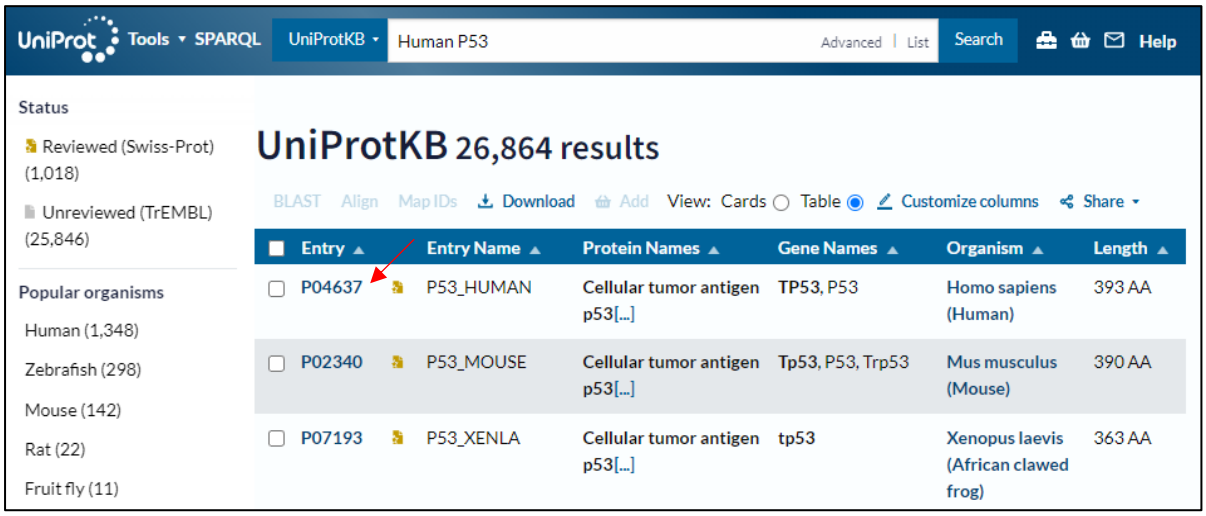
- Upload "FASTA" file, go to below and click "Submit" for MSA



- In the result, we can see how similar each fluorescence proteins.

4. Practice how to find protein domain

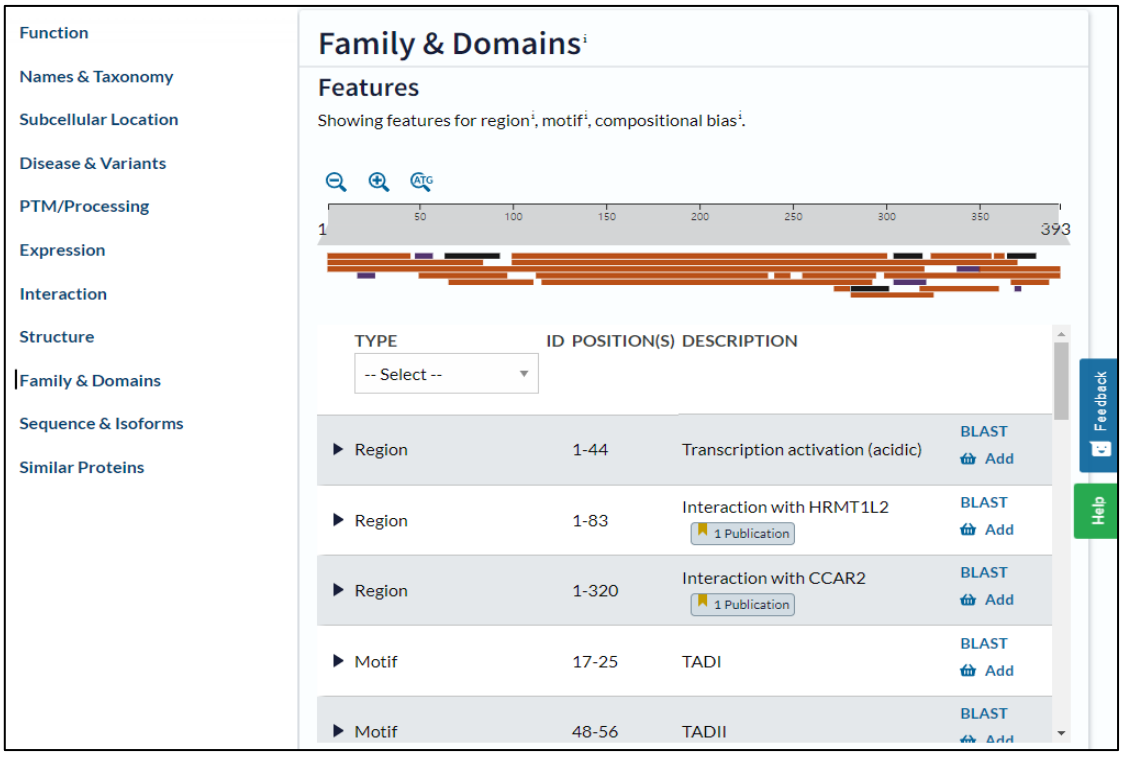
- Find sequence and domains of TP53 using "UniProt"
  - Search "Human P53" and select 1st one (P04637).



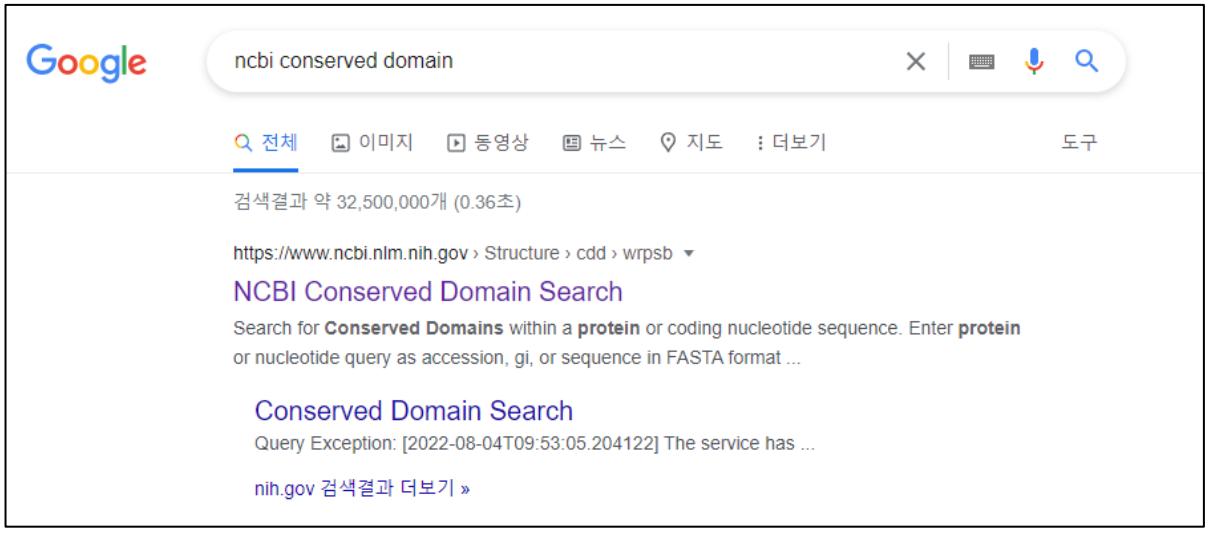
  - Get the protein sequence of TP53.

```
>sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens OX=9606 GN=TP53 PE=1 SV=4
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
```

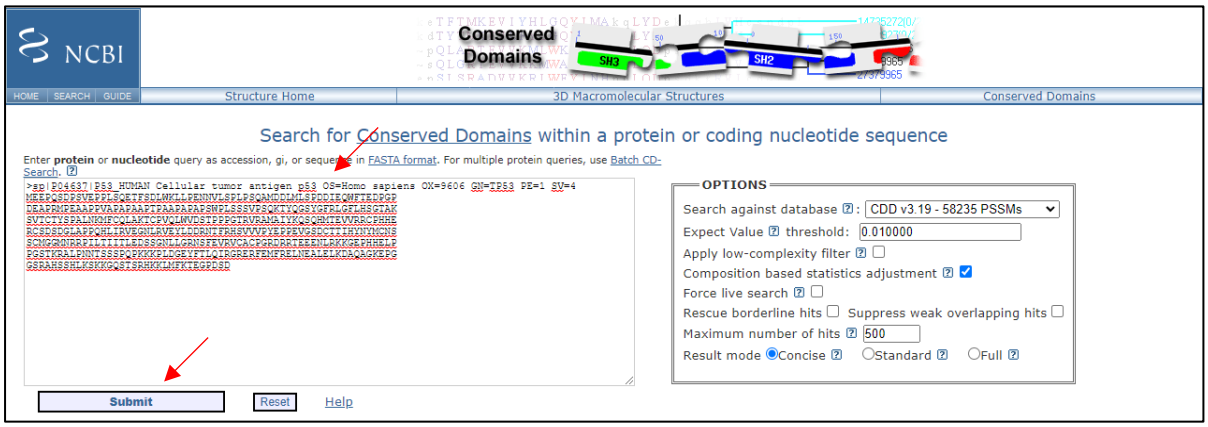
  - Click "Family & Domains" to see the domain information of TP53.



- Find domains of TP53 using "NCBI Conserved Domains"
  - Search "ncbi conserved domain" in google and access NCBI Conserved Domain Search.



  - Copy and paste the sequence of TP53 in the box and click "Submit".



  - In the result, we can see domains of TP53.