Лабораторна робота № 6

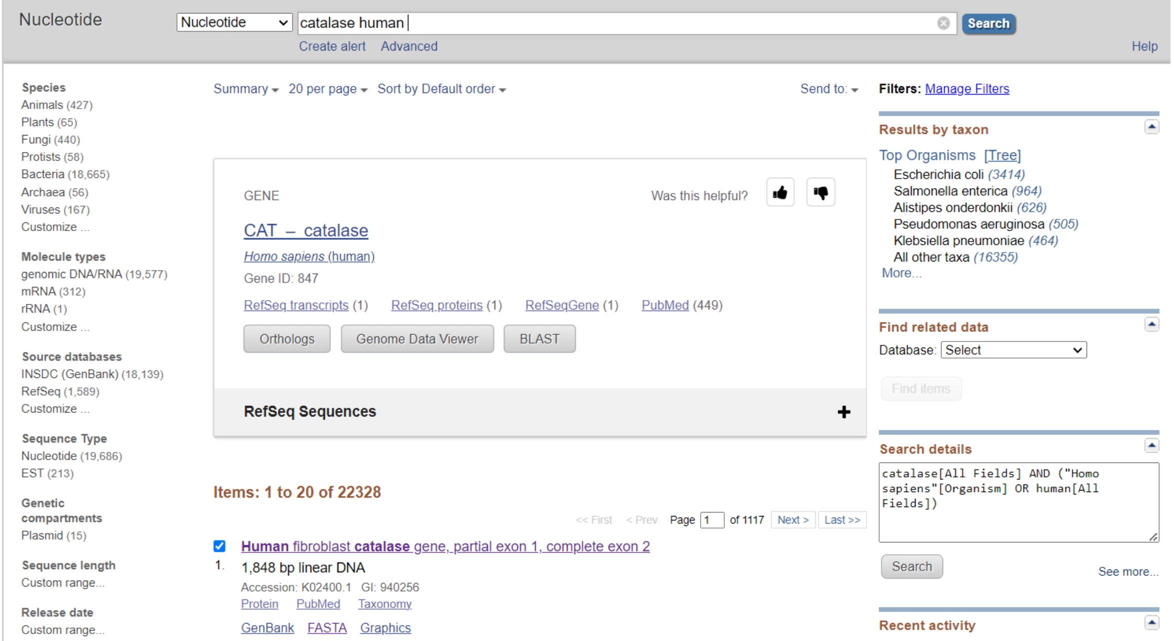
Тема: вирівнювання нуклеотидних послідовностей

Мета: Навчитися здійснювати множинне вирівнювання для встановлення структурних, функціональних і еволюційних відносин між нуклеотидними послідовностями ДНК.

Завдання: За допомогою програми ClustalW виконайте множинне вирівнювання нуклеотидних послідовностей каталази людини (human), гелікобактера пилорі (Helicobacter pylori), грибка ризостонії (Rhizoctonia solani).

Хід дослідження:

1. Заходимо на веб-ресурс GenBank – http://www.ncbi.nlm.nih.gov/Genbank. Знаходимо нуклеотидні послідовності у FASTA-форматі для потрібних організмів. Знаходимо **каталазу людини (human).**



GENE

CAT – catalase

Homo sapiens (human)

Gene ID: 847

RefSeq transcripts (1) RefSeq proteins (1) RefSeqGene (1) PubMed (449)

Items: 1 to 20 of 22328

☑ 1. **Human** fibroblast **catalase** gene, partial exon 1, complete exon 2

1,848 bp linear DNA

Accession: K02400.1 GI: 940256

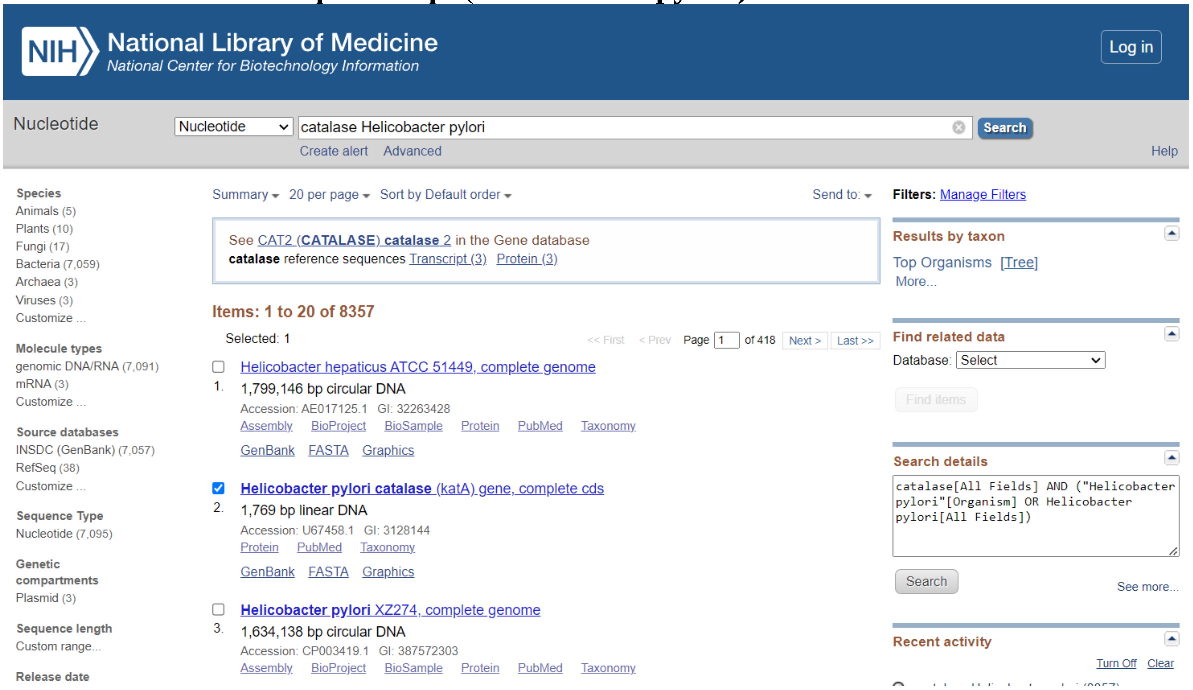
Protein PubMed Taxonomy

GenBank FASTA Graphics

Escherichia coli (3414)
Salmonella enterica (964)
Alistipes onderdonkii (626)
Pseudomonas aeruginosa (505)
Klebsiella pneumoniae (464)
All other taxa (16355)

```
catalase[All Fields] AND ("Homo
sapiens"[Organism] OR human[All
Fields])
```

Каталаза гелікобактера пилорі (Helicobacter pylori):



See CAT2 (**CATALASE**) **catalase** 2 in the Gene database
catalase reference sequences Transcript (3) Protein (3)

Items: 1 to 20 of 8357

Selected: 1

☐ 1. Helicobacter hepaticus ATCC 51449, complete genome

1,799,146 bp circular DNA

Accession: AE017125.1 GI: 32263428

Assembly BioProject BioSample Protein PubMed Taxonomy

GenBank FASTA Graphics

☑ 2. **Helicobacter pylori catalase** (katA) gene, complete cds

1,769 bp linear DNA

Accession: U67458.1 GI: 3128144

Protein PubMed Taxonomy

GenBank FASTA Graphics

☐ 3. **Helicobacter pylori** XZ274, complete genome

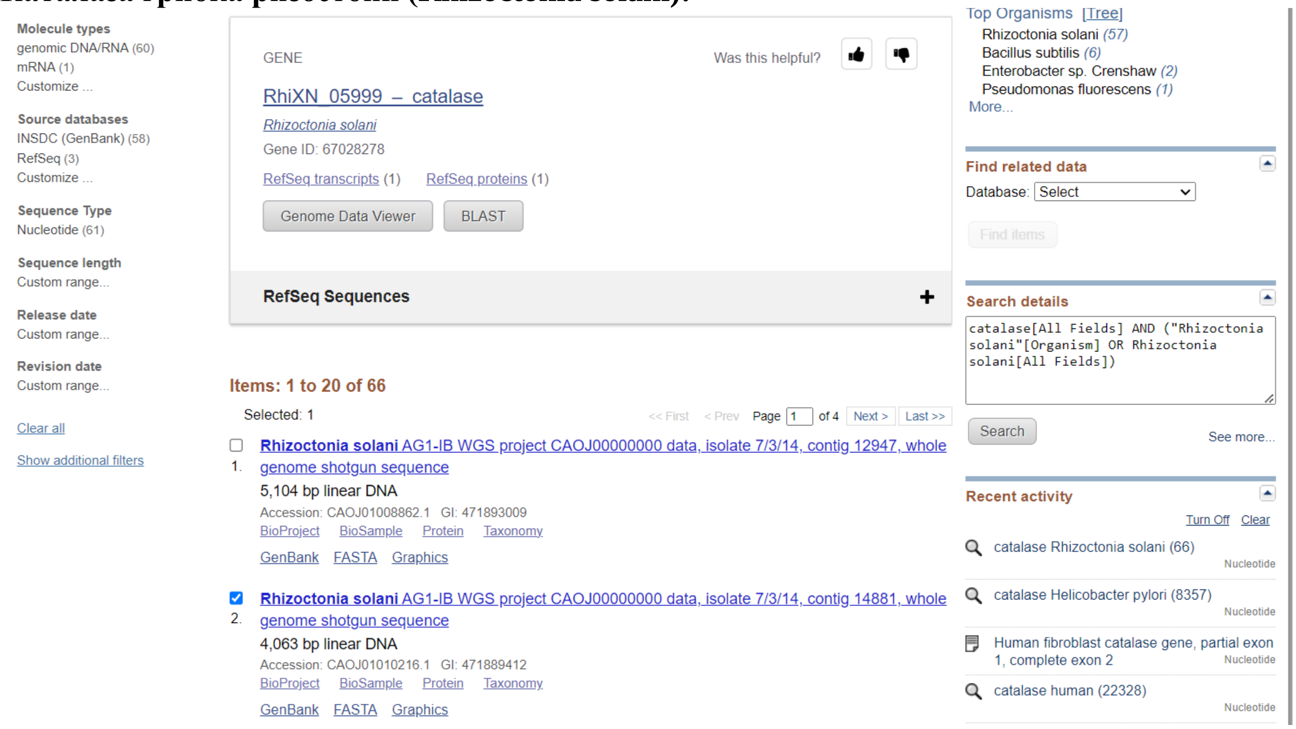
1,634,138 bp circular DNA

Accession: CP003419.1 GI: 387572303

Assembly BioProject BioSample Protein PubMed Taxonomy

```
catalase[All Fields] AND ("Helicobacter
pylori"[Organism] OR Helicobacter
pylori[All Fields])
```

Каталаза грибка ризостонії (Rhizoctonia solani):



Molecule types
genomic DNA/RNA (60)
mRNA (1)
Customize ...

Source databases
INSDC (GenBank) (58)
RefSeq (3)
Customize ...

Sequence Type
Nucleotide (61)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

GENE

Was this helpful?

RhiXN_05999 – catalase
Rhizoctonia solani
Gene ID: 67028278
RefSeq transcripts (1) RefSeq proteins (1)

Genome Data Viewer BLAST

RefSeq Sequences +

Items: 1 to 20 of 66

Selected: 1

<< First < Prev Page 1 of 4 Next > Last >>

☐ 1. Rhizoctonia solani AG1-IB WGS project CAOJ00000000 data, isolate 7/3/14, contig 12947, whole genome shotgun sequence
5,104 bp linear DNA
Accession: CAOJ01008862.1 GI: 471893009
BioProject BioSample Protein Taxonomy
GenBank FASTA Graphics

☑ 2. Rhizoctonia solani AG1-IB WGS project CAOJ00000000 data, isolate 7/3/14, contig 14881, whole genome shotgun sequence
4,063 bp linear DNA
Accession: CAOJ01010216.1 GI: 471889412
BioProject BioSample Protein Taxonomy
GenBank FASTA Graphics

Top Organisms [Tree]
Rhizoctonia solani (57)
Bacillus subtilis (6)
Enterobacter sp. Crenshaw (2)
Pseudomonas fluorescens (1)
More...

Find related data
Database: Select
Find items

Search details
catalase[All Fields] AND ("Rhizoctonia solani"[Organism] OR Rhizoctonia solani[All Fields])
Search
See more...

Recent activity
Turn Off Clear
catalase Rhizoctonia solani (66) Nucleotide
catalase Helicobacter pylori (8357) Nucleotide
Human fibroblast catalase gene, partial exon 1, complete exon 2 Nucleotide
catalase human (22328) Nucleotide

2. Створюємо файл із нуклеотидними послідовностями, які хочемо проаналізувати, та зберігаємо їх у форматі FASTA.

Ген каталази людини:

>K02400.1 Human fibroblast catalase gene, partial exon 1, complete exon 2
AAAGGAGCAGGGGCCTTTGGCTACTTTGAGGTCACACATGACATTACCAAATACTCCAAGGCA
AAGGTATTTGAGCATATTGGAAAGAAGACTCCCATCGCAGTTCGGTTCTCCACTGTTGCTGGAG
AATCGGGTTCAGCTGACACAGTTCGGGACCCTCGTGGGTTTGCAGTGAAATTTTACACAGAAGA
TGGTAACTGGGATCTCGTTGGAAATAACACCCCCATTTTCTTCATCAGGGATCCCATATTGTTTC
CATCTTTTATCCACAGCCAAAAGAGAAATCCTCAGACACATCTGAAGGATCCGGACATGGTCTG
GGACTTCTGGAGCCTACGTCCTGAGTCTCTGCATCAGGTTTCTTTCTTGTTCAGTGATCGGGGGA
TTCCAGATGGACATCGCCACATGAATGGATATGGATCACATACTTTCAAGCTGGTTAATGCAAA
TGGGGAGGCAGTTTATTGCAAATTCCATTATAAGACTGACCAGGGCATCAAAAACCTTTCTGTT
GAAGATGCGGCGAGACTTTCCCAGGAAGATCCTGACTATGGCATCCGGGATCTTTTTAACGCCA
TTGCCACAGGAAAGTACCCCTCCTGGACTTTTTACATCCAGGTCATGACATTTAATCAGGCAGA
AACTTTTCCATTTAATCCATTCGATCTCACCAAGGTGAGTCAGTAAACAACTATATTGTTTTCTT
TTTTAAGTCTCTTCTTACCTAATTAGAAAAAAAATCTAGTCAAACAATTATAATAATGGGGAAG
TCATATACAAAATACAGAGGGTACCACTTCAGAGTGTCCTAAGCTGTGAATGAGTGCTTACCAG
CATCTTACTTCCACGTTCCTGTTTGTCATTTCATTGAGTATGTGTATGTGGCTTCATATATTGTTA
TTAACAGGGAACAGATTATGAAAAGCTGATGTACTTTTTCCTGGGGAAACTGTCAGTATTTACC
ACTTACTATTGTGAAAGATTTAACTAAGGCACTCATCTTAAATTCTTATGTTTTATTGGATTTAA
AAATTATTTTCATTGGCTTGATTGTATTTGAAATCTGGTATTTTTGTGGGTAGCTTTGATTTCCTT
CAGTTGATTGCCTGGTAATTGTGAATATGACATCATTTTCAGGTTTGGCCTCACAAGGACTACCC
TCTCATCCCAGTTGGTAAACTGGTCTTAAACCGGAATCCAGTTAATTACTTTGCTGAGGTTGAAC
AGATAGCCTTCGACCCAAGCAACATGCCACCTGGCATTGAGGCCAGTCCTGACAAAATGCTTCA
GGGCCGCCTTTTTGCCTATCCTGACACTCACCGCCATCGCCTGGGACCCAATTATCTTCATATAC
CTGTGAACTGTCCCTACCGTGCTCGAGTGGCCAACTACCAACGTGACGGCCCGATGTGCATGCA
GGACAATCAGGGTGGTGCTCCAAATTACTACCCCAACAGCTTTGGTGCTCCGGAACAACAGCCT
TCTGCCTTGGAGCACAGCATCCAATATTCTGGAGAAGTGCGGAGATTCAACACTGCCAATGATG

ATAACGTTACTCAGGTGCGGGCATTCTATGTGAACGTGCTGAATGAGGAACAGAGGAAACGTC
TGTGTGAGAACATTGCCGGCCACCTGAAGGATGCACAAATTTTCATCCAGAAGAAAGCGGTCA
AGAACTTCACTGAGGTCCACCCTGACTACGGGAGCCACATCCAGGCTCTTCTGGACAAGTACAA
TGCTGAGAAGCCTAAGAATGCGATTCACACCTTTGTGCAGTCCGGATCTCACTTGGCGGCAAGG
GAGAAGGCAAATCTGTGAGGCCGGGGCCCTGCACCTGTGCATGAAGCT

Ген гелікобактера пилорі (Helicobacter pylori):

>U67458.1 Helicobacter pylori catalase (katA) gene, complete cds
AAGATTGAGATAGAATAGTATGATAATTATTATTAAAACCAGATTAAAAATAAAATTTTGTTTT
TAATCTTTCTTATTTTCATTAATTGTTACGAATAGAAATACTTAAGGGGGTTTTTCATTCTTAAA
AAAAGGATTTTTTAAGGAAATTGAATCTTGTTAGTCTTTATATAACAAATTATGTAATAATCACC
ACAAGTAATCGGCTTGGTGTCAGATTACGAAAATCTAAATCAAATAAAGGAAAGAAGATGGTT
AATAAAGATGTGAAACAAACCACTGCTTTTGGCGCTCCCGTTTGGGATGATAACAATGTGATTA
CGGCTGGTCCTAGAGGTCCTGTTTTATTACAAAGCACTTGGTTTTTGGAAAAGTTAGCAGCGTTT
GACAGAGAAAGGATCCCTGAAAGGGTAGTGCATGCTAAAGGAAGCGGGGCTTATGGCACTTTC
ACCGTGACTAAAGACATCACTAAATACACTAAAGCGAAGATTTTTTTTAAAGTGGGCAAAAAA
ACCGAATGCTTTTTCAGGTTTTTTACTGTGGCTGGTGAAAGAGGCAGTGCGGATGCAGTGAGAG
ACCCTAGAGGTTTTGCGATGAAGTATTACACTGAAGAAGGTAACTGGGATTTAGTAGGGAACA
ACACGCCTGTTTTCTTTATCCGTGATGCGATCAAATTCCCTGATTTCATCCACACCCAAAAAAGA
GACCCTCAAACCAATTTGCCTAACCACGACATGGTATGGGATTTTTGGAGTAATGTTCCTGAAA
GCTTGTATCAAGTAACATGGGTTATGAGCGATAGAGGGATCCCTAAATCTTTCCGCCACATGGA
TGGTTTTGGCAGCCACACTTTCAGTCTTATCAACGCTAAGGGCGAACGCTTTTGGGTGAAATTCC
ACTTTCACACCATGCAAGGCGTTAAGCACTTGACTAACGAAGAAGCCGCAGAAGTCAGAAAAT
ATGATCCTGATTCCAATCAAAGGGATTTATTCAATGCGATCGCTAGAGGGGATTTCCCAAAATG
GAAATTAAGCATTCAAGTGATGCCAGAAGAAGATGCTAAGAAGTATCGATTCCATCCGTTTGAT
GTTACTAAAATTTGGTATCTCCAAGATTATCCGTTGATGGAAGTGGGCATTGTAGAGTTGAATA
AAAATCCTGAAAACTATTTTGCAGAAGTGGAGCAAGTGGCATTCACTCCGGCTAATGTCGTTCC
TGGAATTGGCTATAGCCCTGATAGGATGTTACAAGGACGCTTGTTCTCTTATGGGGACACACAC
CGCTACCGCTTAGGGGTTAATTATCCTCAAATACCGGTTAATAAACCAAGATGCCCGTTCCACT
CTTCTAGCAGAGATGGTTACATGCAAAACGGATACTACGGCTCTTTACAAAACTATACGCCTAG
CTCATTGCCAGGTTATAAAGAAGATAAGAGCGCGAGAGATCCTAAGTTCAACTTAGCTCATATT
GAGAAAGAGTTTGAAGTGTGGAATTGGGATTACAGGGCTGATGATAGCGATTACTACACCCAA
CCAGGTGATTACTACCGCTCATTGCCAGCTGATGAAAAAGAAAGGTTGCATGACACTATTGGAG
AGTCTTTGGCTCATGTTACCCATAAGGAAATTGTGGATAAACAATTGGAGCATTTCAAGAAAGC
TGATCCCAAATACGCTGAGGGAGTTAAAAAAGCTCTTGAAAAACACCAAAAGATGATGAAAGA
CATGCATGGAAAAGACATGCACCACACAAAAAAGAAAAAGTAA

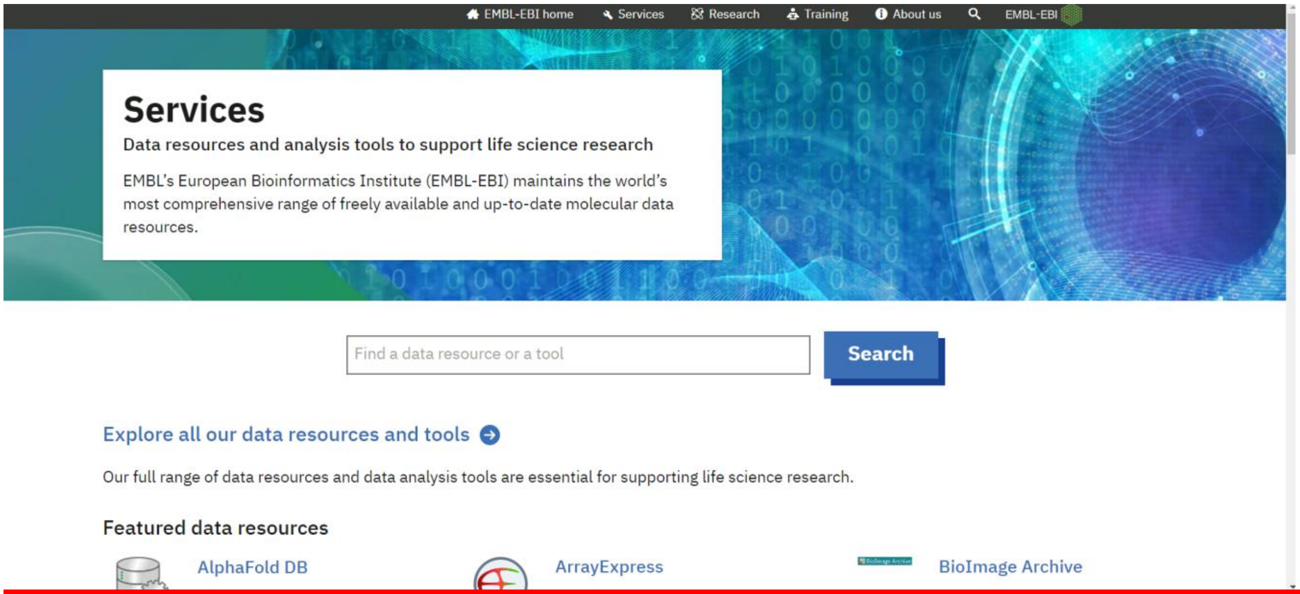
Ген каталази грибка ризостонії (Rhizoctonia solani):

>CAOJ01010216.1 Rhizoctonia solani AG1-IB WGS project CAOJ00000000 data, isolate 7/3/14, contig 14881, whole genome shotgun sequence
CTATAAGCACTTACCGAGCCATGAATGGACAATTAACAGGGATCTGGTGAATGTTGGCACTTCC
AAGACGGTGGTACTGAGCATCCCGATAGAAGAATGCGCGCCATTGAAGCAAAGTGTCCGGAGA
CAGTTCAATTCCCGGAACCAGGGAACCAGGAGAGAACGCAGCTTGTTCAACATCACGGTGATA
ATCCTTATTCTCACATCAGCGAGAGCTACTCGAATAAAGGCAAGAGACCTTACTTCGGGATTAC
GATTCAAAATTAGTTCTCCGACTTCGTGCATCGGGAACTTGTTGCGCGGCCAGACCTTGGTGAC
GTCGAATGCATCGAAATCGATTTTGCCCGAACTGAGGTCTTCGGGTGTGATAACCTGGACAAAC
AACGTCCATCGAGCGTCGCCTCCCTTTTCGATATGCTCCCACAAGTCGCGTTTGGCAAAGTCAG

GGTCTTCACCACACATGGCGACAGATTCTTCATGAGTGAACTCTTTGGTGCCTTGTTGGGATCGC
CAATGATACTAAAAACGACCATCAATCAGGTGAATACCAAAGAGATCAAAGGGCTTTTTCTATA
GCTCACCTTGACAAAGATTCCCTTCCCTTGGCTGTTCGTCCAACGGAAAGTATGGCAGCCATAG
CCACTCATGAAGCGCCATCCGACGGGAGTAGCGTGATCCGAAAGCAGCATGAGTCCAGCGTGC
TGAGACTCGGGTACGTTTGCAAGGAAATCAAACCAGGCGTTGTAATCGATGAGTCCGCTCTTTG
GGTTACGTTGCTGCGACCTGATGTTATCGGGTCCCATAAAGGGGTCGCGGACGAAAAAGACAG
GCCAGTTCAAGCCAACCAAGTCATAGTTTCCTAGGGCAGATTTTAATAAATGATCGCAGTAATG
TAATGATCAAGGTATACCTTCCTCGGTATAAAACTTAGTGGCGAACCCACGGGGATTGCGCGCA
CTATCAGGGAATTCCCGCTGATGGCTCCCAGTCAGCAAGCAGAACGAGTGTACAATCAATCCAT
ACTCACGCCGTACGTGACAGTCGAGAAACGGGTATAAGCAGGTGTCTTCTTCCCCGGTTGGAAC
AGAGCAGCCTTGGTCAGATGAGAAACATCCTTGGTAACTTGGAAGTATCCAAACCTACAAGTCA
AAATGTTACATATTTCCCAGAAGAGCACCTTTAGACGCAAAAACTCACGCGGATGAGCCACAC
GGATGAACAATACGTTCCTGGATTTTACCCCGATCGAAAGCCTGCTGCTTCTCGAGCAACAAAG
TATCCCCCTGATACGGAATACCACCAATGTTCAAAGCGTGGCTAAAACAACGACATCAGTACAT
ACCAGACATTAATACATTTCACAATCCCACTTACCCAGAGTCAGGAATAGGCTGTCCAAAATTA
CTAGTATAAATCGGAGAATCGTCCTTGGCATGAGACTCCTGAACAAGTCCAGCCAGACGAGCG
GCGGTAGCCGCCGGCCCATCATTCAATCCGCTAAACGCTTGGTTCCGCTGAATCACCCTCGGGT
CGACTCCGGGTATTTTAGAAAGCGCATCTTGAATAAACGACGTCATGGGATTTGCAGTTTGGTT
GTAACCCCCAGCGAGGAGTAAACAGGGAGCTTTGGGGAGTGGGGAGGGATGAAATGTGCCCCA
AAGCGAGTGGTCTGACTTTTATCTCTTTGCTGATTGGGCTCGAAGACGATTCATACGCGCACAT
ACTGGCTCATAATGACCTTTTGCTACGTGGTATGATTCAAATGACGAACGGGTTTGTCGAGATC
GTGCGTCCCAAGTTGGCGTTGGTATGTACGGAATTCAAGCGGAGAGACTTGGCTTATTAGAATT
ACGTTGGTCGAAGATCAAGAACCATGAACCCTGAGTCTTTGGCTGCGTTTGGACAAAATCGTCA
AGTTGCGCCATGGATCAACGCACGTTCCCAAAGCCGACTTTTGCGCTCTGGCTGACCGGGCTCG
ACTTCAAGTGGAATGAACCACCACCACCACCAGTCAATCGGAATGCACGACCCCCCCAGATCC
AGCTCAACTTTTTAAGCACACATTAATTAACCCTATATAAAAAGAAAAACAGCGCAACGAGCG
ACTAGCAAACAGCAGAGATCGACAAAATAGGTGAGATCGAAACAGATGTCCATGGGGGAAAC
AAACAAGCAAACAAAACAACAAAAAATTCTAAAAAACAAAAAAGTCCGTACACACACACACA
TGTATTTTCTACGCGGGCGTGCCAGTCTCACGAGGCATGGCGTGACGGCCTCGGAATACCCTCT
GGGCTGTACCGCTGCATGCAGTCTGGGTAGCGCTGACACCGACAGTACCCCAACCGATGCTAGG
ATACGCACATCCAGCGACGGGCGTGTAGAGGCCCTGTCGGTCGGCAGCGTTAGCCCATCCAGA
CCCAGCATTGACACCTCCGGTAAACGAAGCGGGGGCAAGAGTGATGAGAGCGCCGGTGGCGGT
GTAGGTAGGGATCACAGCAGCGTTGGGGACGTTGTTGAGCGAGGTGGGCGGCCAGACGCGGGT
GGCGGAGTAGGAAGCTGGGATTGTACCGGCTCCGACACCACCGGTCTTGGATGCTGGGAGAGG
ACCATTGAAAGGAGAGCTGTTGCCGCAGGTTCCGACGGCTTCGCGAGGATCGGTGGGCATCCAT
CCGCCTTCAAGACCGTGCTGGTAAGACCAGAGCGGGGCAGCGATGGTTCCCCACGTCGAGGAC
CTTCCGATCTTCCATGTCCAGAAGAAGAAGTTCTATAACAAATAGGTGAGAGATGGTTCAAGGG
GGGGAAATGTGGAGAACGTACCTGCATGGCATCCATGCTGGACAACGCAAACTGCTTCAGTGC
GGCCTTTTCGGCGGCGCTCCACCTAGCAGAATCCATCCACTTGACACAAGAGCCAGCACCACCG
TTGGGACCTCCGTAACCGCCAAAAGTACCGTCATAACGAGCACCACGGTTGACACCGTTGAGG
AAAAGTCCGCAATCGTTGAACGCGTTGGAGAACTCTCCGGCAGTGGTCATACCGAACGAAGAC
ATGGACTGGTTGAAGGCTGCAGCACGCGAGGCACATGGCTTGTTGGTCTGGGTGGTGATGGCAC
CAGTGTCTTGGTCTCCGAAGCAGAAGTAAGGGTGAGCATCGCTAGAATAAAAGGTTAGGCTGT
GGGACACGCGTAGGGGACAAAAATACTAACAGAGCGACGCGGTCTCCACCTGGCAAGAACTCG
GTCCAGCGGTCGAGGGGGTCGAAGCCGTCATGGACGCTAATCCAGGGGCCCTTGCCCTCACCA
ATACCAGTAACGCTCCTCATAATATCGTGGACCTCGAGGTAGCTGGACAAGCAAAATTAATACA
GGAAAGCCACGACCAAGGGCTTATACGCACAAGCTCTCGAGAACGTCGCGACCGATGGTGGCA
ATGAGTGCCTCGTTGATGACACCGAACATGGGCACGACATCGCGGTATTCGGGTTGCGAGATGA
ACTGAGTAATGATGCGGATGTAGTTCAAAGTACGCTGAGCGTTGGCGATACCCATGTTTCCGCT

GAGCCTGAAAGAGGATCATGTTTAGCTACCGGTACAATAGGCAAGAAAATGAACCCTACCATC
CGACAGGGCCCAACATGCCAGAGTGGTTGTATCCATTTTGCGATCCAGGAACAGTGTGGAGATC
AAGGTTGATACGAAGACCATACTTGCGCGCCCACTCGATAGCCTTGAGGAAGTACTTCCATGCA
ACGCCCTCTAGGAACGGTTCACCAGGATAGACCTCGATAGCCCAGTACGGAATGGGGATACGA
ATCCAGTTCAGGCCCGCCGCGGCGATCTGAGCAAAGTCTTCCTCGGTCTAGTCATAACGATCAG
CGGACAAACCCTAAAAAAGGTAAATTCCAACCACGTACAATAAAGGTCTTGTAGTGATCCTCG
ATGGCCTTGGCAACGCCTCCAGACGAAGCATCGCCTGCAAGATTCTCGCAGAGAGTCCACTCAT
CGATGGCAGGGACGGTCGCGTTCATGTAGGGTTCGTACATCGCCGGGCTGATAAAGGG

3. Здійснюємо вхід на сторінку браузера веб-сервера EMBL-EBI – http://www.ebi.ac.uk/services.



Services

Data resources and analysis tools to support life science research

EMBL's European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Explore all our data resources and tools

Our full range of data resources and data analysis tools are essential for supporting life science research.

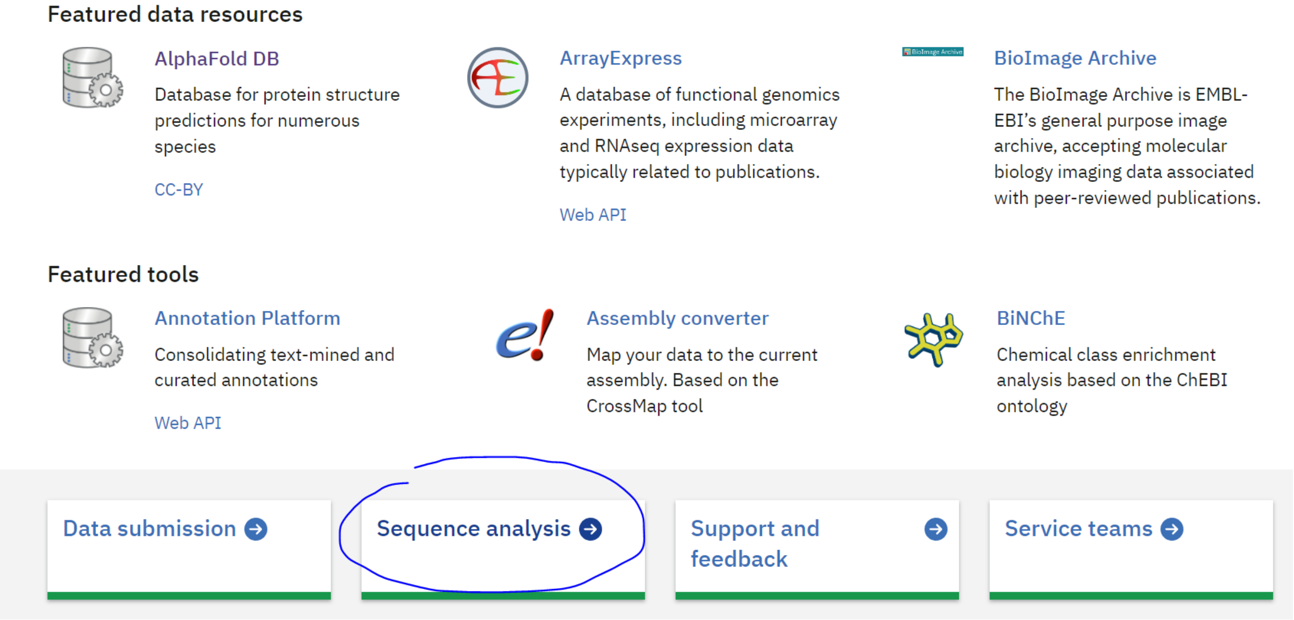
Featured data resources

AlphaFold DB

ArrayExpress

BioImage Archive

Вибираємо опцію «Sequence analysis – аналіз послідовностей» і переходимо на сторінку з вікном програми.



Featured data resources

AlphaFold DB
Database for protein structure predictions for numerous species
CC-BY

ArrayExpress
A database of functional genomics experiments, including microarray and RNAseq expression data typically related to publications.
Web API

BioImage Archive
The BioImage Archive is EMBL-EBI's general purpose image archive, accepting molecular biology imaging data associated with peer-reviewed publications.

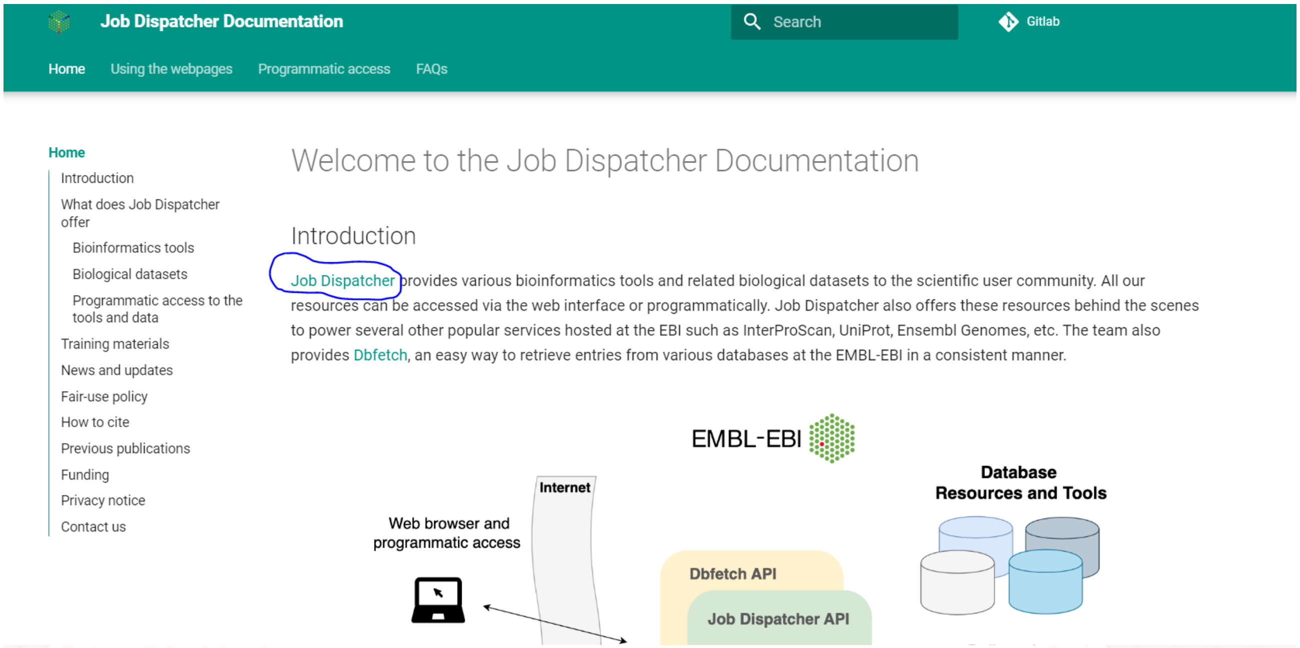
Featured tools

Annotation Platform
Consolidating text-mined and curated annotations
Web API

Assembly converter
Map your data to the current assembly. Based on the CrossMap tool

BiNChE
Chemical class enrichment analysis based on the ChEBI ontology

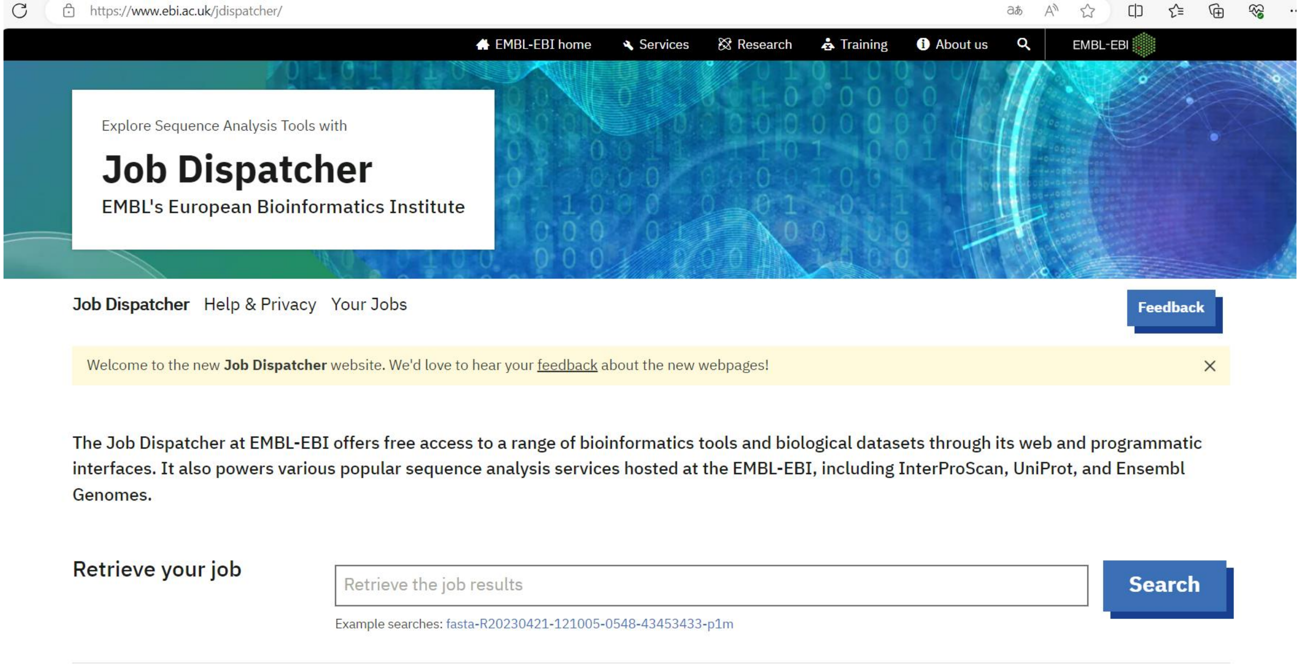
Переходимо на наступну сторінку та обираємо диспетчера робіт (Job Dispatcher)^



Welcome to the Job Dispatcher Documentation

Introduction

Job Dispatcher provides various bioinformatics tools and related biological datasets to the scientific user community. All our resources can be accessed via the web interface or programmatically. Job Dispatcher also offers these resources behind the scenes to power several other popular services hosted at the EBI such as InterProScan, UniProt, Ensembl Genomes, etc. The team also provides Dbfetch, an easy way to retrieve entries from various databases at the EMBL-EBI in a consistent manner.



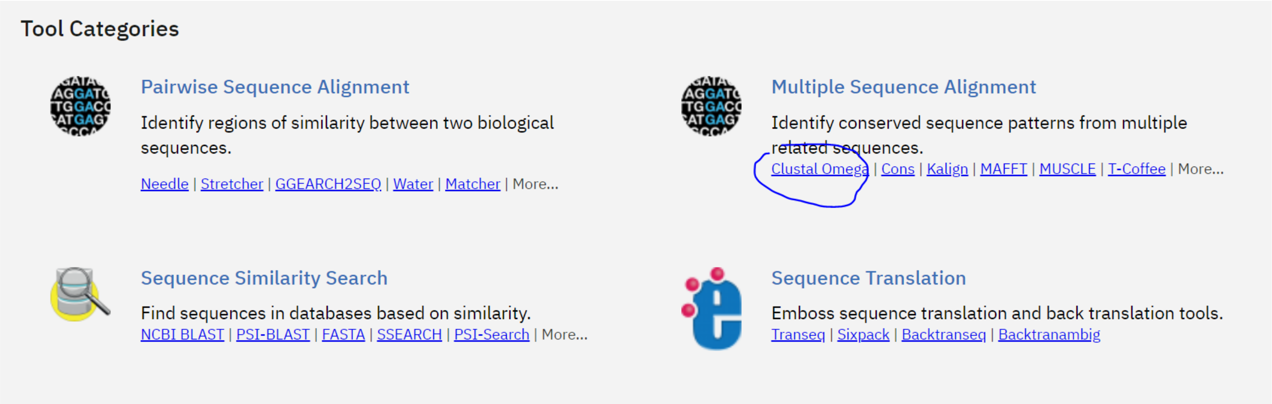
Welcome to the new **Job Dispatcher** website. We'd love to hear your feedback about the new webpages!

The Job Dispatcher at EMBL-EBI offers free access to a range of bioinformatics tools and biological datasets through its web and programmatic interfaces. It also powers various popular sequence analysis services hosted at the EMBL-EBI, including InterProScan, UniProt, and Ensembl Genomes.

Retrieve your job

Example searches: fasta-R20230421-121005-0548-43453433-p1m

Знизу натискаємо в опції Multiple Sequence Alignment (вирівнювання кількох послідовностей) «Clustal Omega»



Tool Categories

Pairwise Sequence Alignment

Identify regions of similarity between two biological sequences.

Needle | Stretcher | GGEARCH2SEQ | Water | Matcher | More...

Multiple Sequence Alignment

Identify conserved sequence patterns from multiple related sequences.

Clustal Omega | Cons | Kalign | MAFFT | MUSCLE | T-Coffee | More...

Sequence Similarity Search

Find sequences in databases based on similarity.

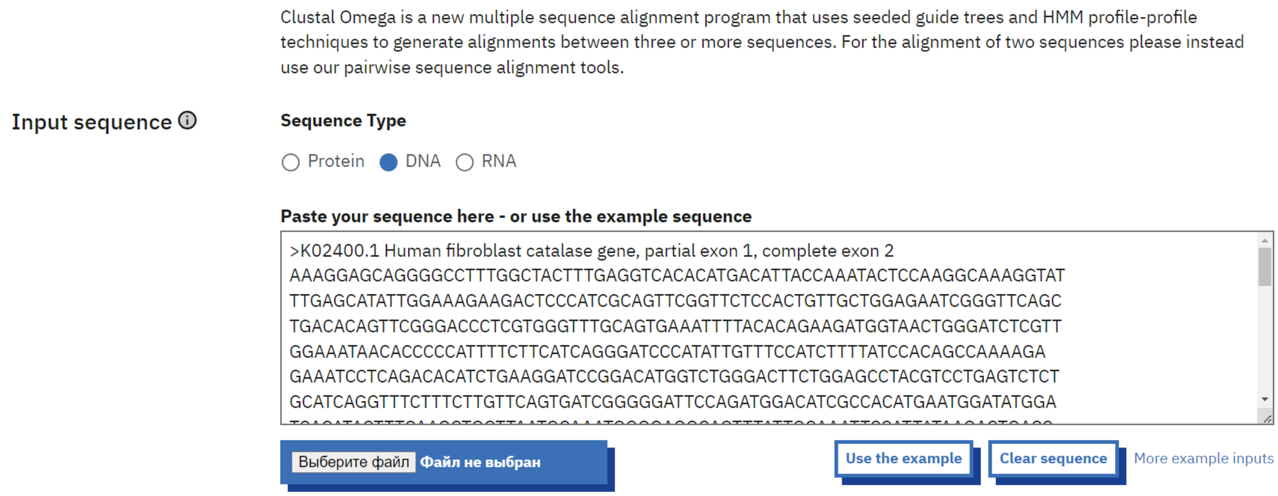
NCBI BLAST | PSI-BLAST | FASTA | SSEARCH | PSI-Search | More...

Sequence Translation

Emboss sequence translation and back translation tools.

Transeq | Sixpack | Backtranseq | Backtranambig

4. ***Вибираємо*** опцію «DNA», в поле вставляємо сукупність нуклеотидних послідовностей, які раніше зберегли (без пропущених строк між ними) і натискаємо «Use the example»



Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

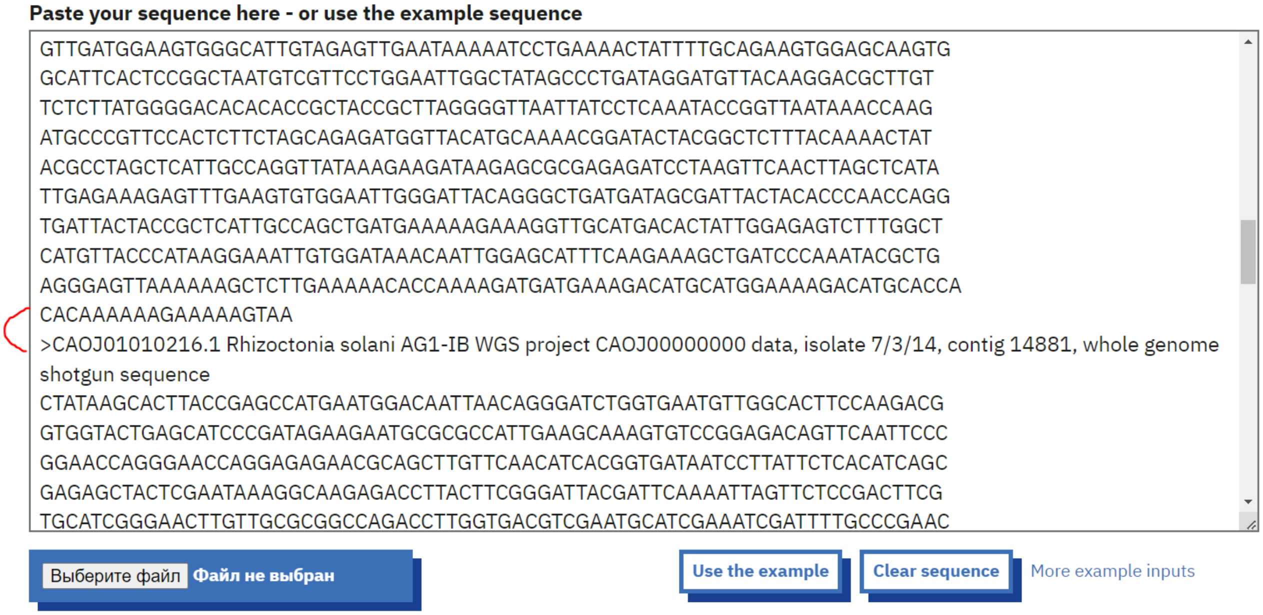
Input sequence

Sequence Type

○ Protein ● DNA ○ RNA

Paste your sequence here - or use the example sequence

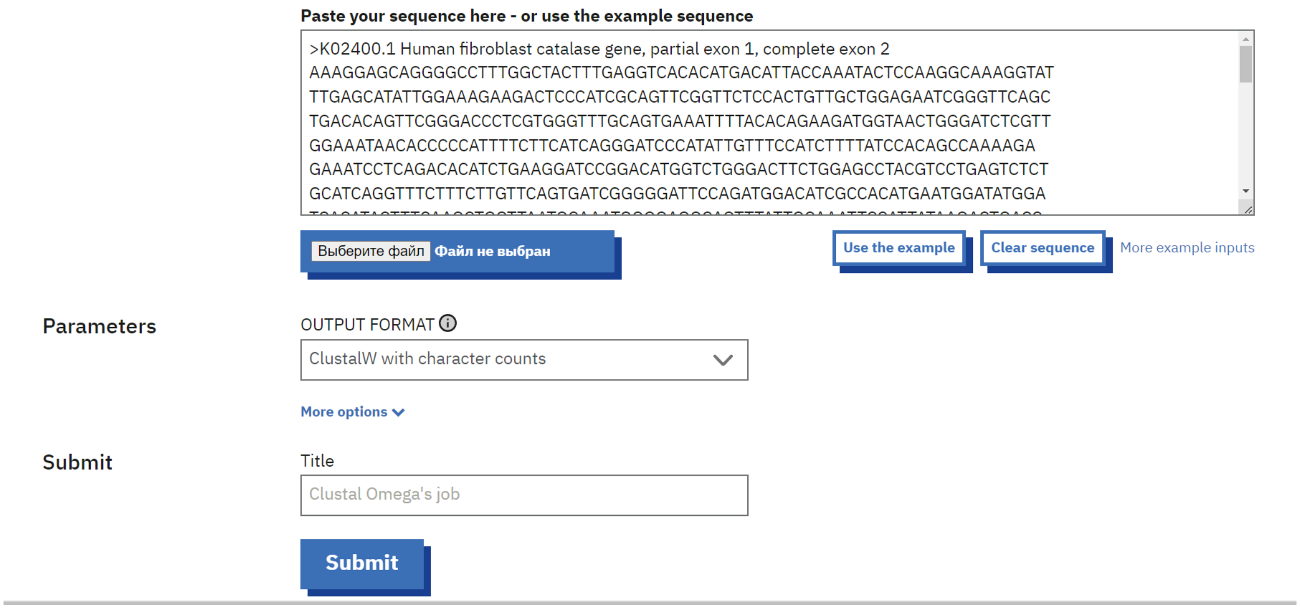
>K02400.1 Human fibroblast catalase gene, partial exon 1, complete exon 2
AAAGGAGCAGGGGCCTTTGGCTACTTTGAGGTCACACATGACATTACCAAATACTCCAAGGCAAAGGTAT
TTGAGCATATTGGAAAGAAGACTCCCATCGCAGTTCGGTTCTCCACTGTTGCTGGAGAATCGGGTTCAGC
TGACACAGTTCGGGACCCTCGTGGGTTTGCAGTGAAATTTTACACAGAAGATGGTAACTGGGATCTCGTT
GGAAATAACACCCCCATTTTCTTCATCAGGGATCCCATATTGTTTCCATCTTTTATCCACAGCCAAAAGA
GAAATCCTCAGACACATCTGAAGGATCCGGACATGGTCTGGGACTTCTGGAGCCTACGTCCTGAGTCTCT
GCATCAGGTTTCTTTCTTGTTCAGTGATCGGGGGATTCCAGATGGACATCGCCACATGAATGGATATGGA



Paste your sequence here - or use the example sequence

GTTGATGGAAGTGGGCATTGTAGAGTTGAATAAAAATCCTGAAAACTATTTTGCAGAAGTGGAGCAAGTG
GCATTCACTCCGGCTAATGTCGTTCCTGGAATTGGCTATAGCCCTGATAGGATGTTACAAGGACGCTTGT
TCTCTTATGGGGACACACACCGCTACCGCTTAGGGGTTAATTATCCTCAAATACCGGTTAATAAACCAAG
ATGCCCGTTCCACTCTTCTAGCAGAGATGGTTACATGCAAAACGGATACTACGGCTCTTTACAAAACTAT
ACGCCTAGCTCATTGCCAGGTTATAAAGAAGATAAGAGCGCGAGAGATCCTAAGTTCAACTTAGCTCATA
TTGAGAAAGAGTTTGAAGTGTGGAATTGGGATTACAGGGCTGATGATAGCGATTACTACACCCAACCAGG
TGATTACTACCGCTCATTGCCAGCTGATGAAAAAGAAAGGTTGCATGACACTATTGGAGAGTCTTTGGCT
CATGTTACCCATAAGGAAATTGTGGATAAACAATTGGAGCATTTCAAGAAAGCTGATCCCAAATACGCTG
AGGGAGTTAAAAAAGCTCTTGAAAAACACCAAAAGATGATGAAAGACATGCATGGAAAAGACATGCACCA
CACAAAAAAGAAAAAGTAA
>CAOJ01010216.1 Rhizoctonia solani AG1-IB WGS project CAOJ00000000 data, isolate 7/3/14, contig 14881, whole genome shotgun sequence
CTATAAGCACTTACCGAGCCATGAATGGACAATTAACAGGGATCTGGTGAATGTTGGCACTTCCAAGACG
GTGGTACTGAGCATCCCGATAGAAGAATGCGCGCCATTGAAGCAAAGTGTCCGGAGACAGTTCAATTCCC
GGAACCAGGGAACCAGGAGAGAACGCAGCTTGTTCAACATCACGGTGATAATCCTTATTCTCACATCAGC
GAGAGCTACTCGAATAAAGGCAAGAGACCTTACTTCGGGATTACGATTCAAAATTAGTTCTCCGACTTCG
TGCATCGGGAACTTGTTGCGCGGCCAGACCTTGGTGACGTCGAATGCATCGAAATCGATTTTGCCCGAAC

Натискаємо «Submit» - надіслати

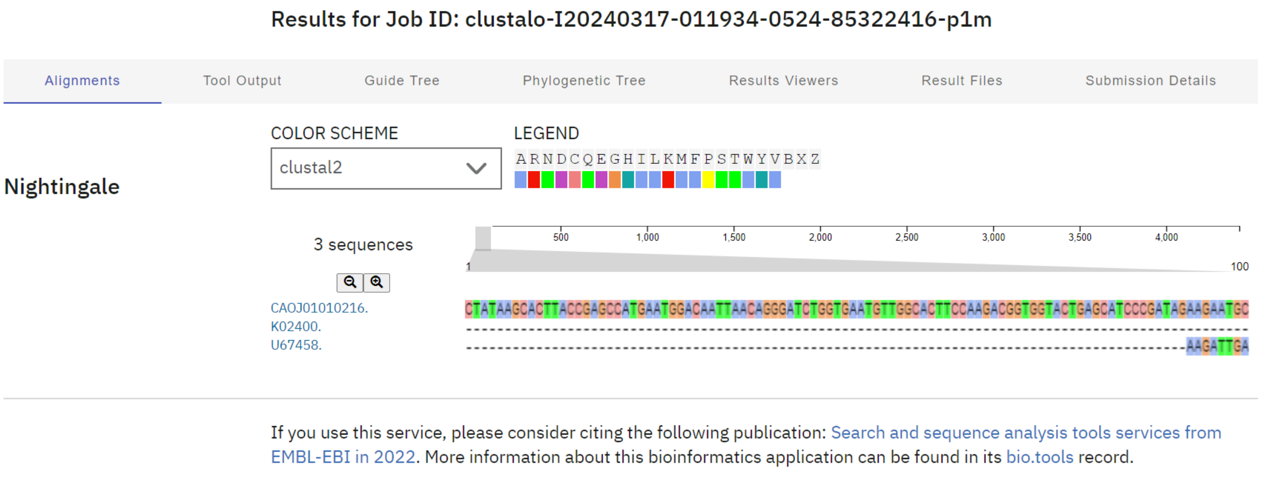


```
>K02400.1 Human fibroblast catalase gene, partial exon 1, complete exon 2
AAAGGAGCAGGGGCCTTTGGCTACTTTGAGGTCACACATGACATTACCAAATACTCCAAGGCAAAGGTAT
TTGAGCATATTGGAAAGAAGACTCCCATCGCAGTTCGGTTCTCCACTGTTGCTGGAGAATCGGGTTCAGC
TGACACAGTTCGGGACCCTCGTGGGTTTGCAGTGAAATTTTACACAGAAGATGGTAACTGGGATCTCGTT
GGAAATAACACCCCCATTTTCTTCATCAGGGATCCCATATTGTTTCCATCTTTTATCCACAGCCAAAAGA
GAAATCCTCAGACACATCTGAAGGATCCGGACATGGTCTGGGACTTCTGGAGCCTACGTCCTGAGTCTCT
GCATCAGGTTTCTTTCTTGTTCAGTGATCGGGGGATTCCAGATGGACATCGCCACATGAATGGATATGGA
```

OUTPUT FORMAT: ClustalW with character counts

Title: Clustal Omega's job

5. ***Отримуємо*** результати вирівнювання:



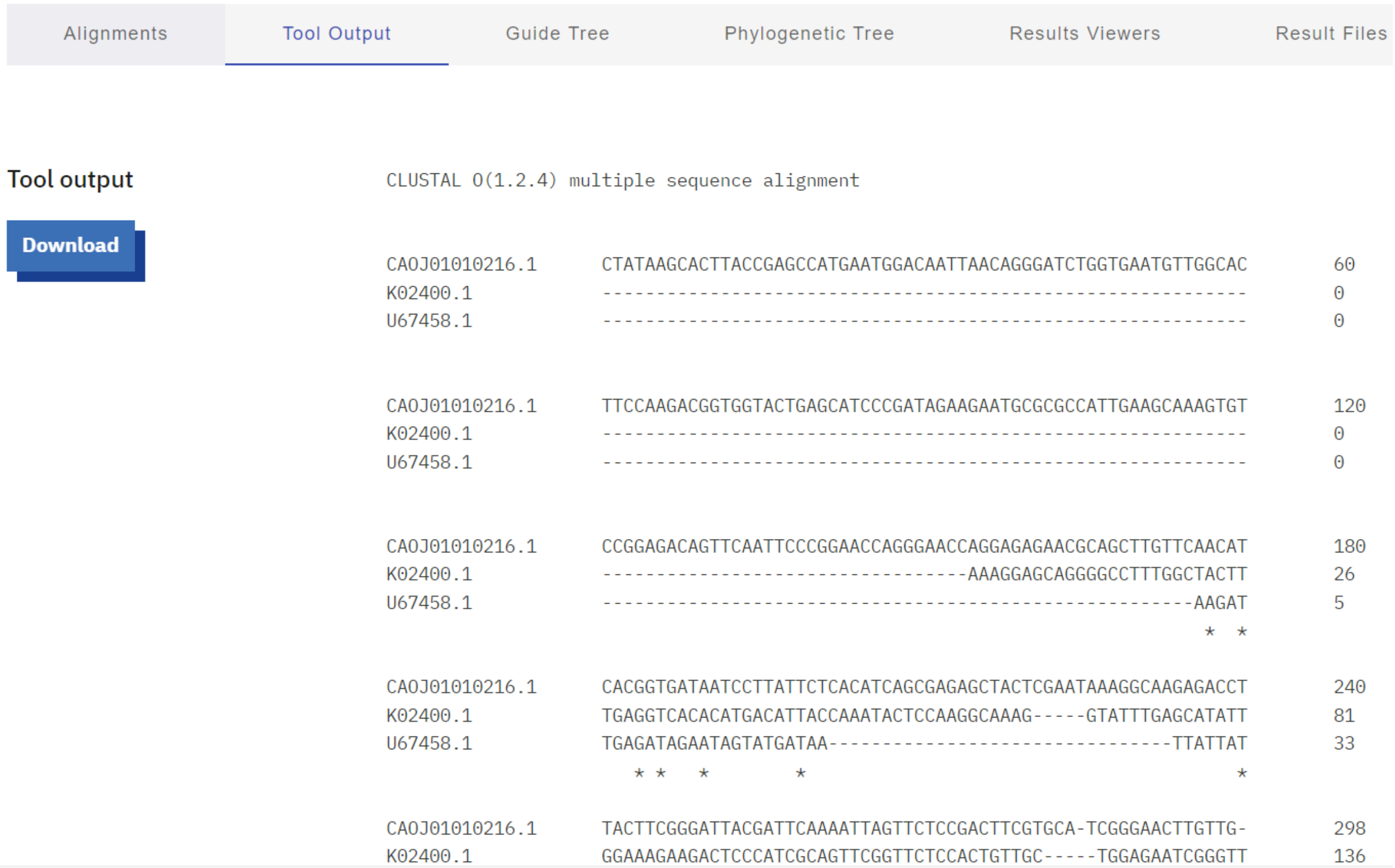
Results for Job ID: clustalo-I20240317-011934-0524-85322416-p1m

COLOR SCHEME: clustal2

```
CAOJ01010216.  CTATAAGCACTTACCGAGCCATGAATGGACAATTAACAGGGATCTGGTGAATGTTGGCACTTCCAAGACGGTGGTACTGAGCATCCCGATAGAAGAATGC
K02400.        ----------------------------------------------------------------------------------------------------
U67458.        --------------------------------------------------------------------------------------------AAGATTGA
```

If you use this service, please consider citing the following publication: Search and sequence analysis tools services from EMBL-EBI in 2022. More information about this bioinformatics application can be found in its bio.tools record.

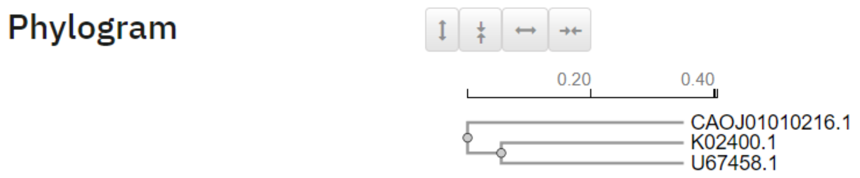
При натисканні **Tool Output** отримаємо такий вид результату вимірювання:



```
CLUSTAL O(1.2.4) multiple sequence alignment


CAOJ01010216.1      CTATAAGCACTTACCGAGCCATGAATGGACAATTAACAGGGATCTGGTGAATGTTGGCAC        60
K02400.1            ------------------------------------------------------------        0
U67458.1            ------------------------------------------------------------        0


CAOJ01010216.1      TTCCAAGACGGTGGTACTGAGCATCCCGATAGAAGAATGCGCGCCATTGAAGCAAAGTGT        120
K02400.1            ------------------------------------------------------------        0
U67458.1            ------------------------------------------------------------        0


CAOJ01010216.1      CCGGAGACAGTTCAATTCCCGGAACCAGGGAACCAGGAGAGAACGCAGCTTGTTCAACAT        180
K02400.1            ----------------------------------AAAGGAGCAGGGGCCTTTGGCTACTT        26
U67458.1            -------------------------------------------------------AAGAT        5
                                                                        *  *

CAOJ01010216.1      CACGGTGATAATCCTTATTCTCACATCAGCGAGAGCTACTCGAATAAAGGCAAGAGACCT        240
K02400.1            TGAGGTCACACATGACATTACCAAATACTCCAAGGCAAAG-----GTATTTGAGCATATT        81
U67458.1            TGAGATAGAATAGTATGATAA--------------------------------TTATTAT        33
                       * *   *        *                                        *

CAOJ01010216.1      TACTTCGGGATTACGATTCAAAATTAGTTCTCCGACTTCGTGCA-TCGGGAACTTGTTG-        298
K02400.1            GGAAAGAAGACTCCCATCGCAGTTCGGTTCTCCACTGTTGC-----TGGAGAATCGGGTT        136
```

Де CAOJ01010216 - ген каталази грибка ризостонії (Rhizoctonia solani)
K02400 - ген каталази людини
U67458 - ген гелікобактера пилорі (Helicobacter pylori)

Філограма виглядає таким чином:



Зробити висновок щодо результатів вимірювання послідовностей вказаних організмів:

Завдання для самостійного виконання:

1) **Виконайте** вирівнювання нуклеотидних послідовностей мРНК, що кодують молекули актину у **людини** (*Homo sapiens actin alpha 2, smooth muscle (ACTA2), transcript variant 4, mRNA*, актин альфа 2 гладкої мускулатури), **мухи дрозофіли** (*Drosophila melanogaster actin 5C, transcript variant E*), **дощового черв'яка** (*Lumbricus rubellus mRNA for actin, partial*), **червоної джунглевої курки** (*Gallus gallus actin gamma 1 like (ACTG1L), mRNA*), **бика** (*Bos taurus actin gamma 1 (ACTG1), mRNA*) та **горили** (*Gorilla gorilla gorilla actin alpha 2, smooth muscle (ACTA2), transcript variant X14, mRNA*).

Надати у вигляді скріншотів результати порівняльного вирівнювання за п.5.

2) **Надати у вигляді скріншоту філограму та її аналіз**

3) В опції **Tool Output** вибрати будь який фрагмент РНК з шістьома послідовностями нуклеотидів усіх організмів і за таблицею генетичного коду побудувати шість ланцюгів білка. Виділити амінокислотні залишки, що відрізняються.

Приклад фрагментів:

```
NM_001406462.1 TGCCTTGGTGTGTGACAATGGCTCTGGGCTCTGTAAGGCCGGCTTTGCTGGGGACGATGC 580
XM_004049748.4 TGCCTTGGTGTGTGACAATGGCTCTGGGCTCTGTAAGGCCGGCTTTGCTGGGGACGATGC 340
NM_001007824.3 TGCCCTGGTGGTGGACAATGGCTCTGGTATGTGCAAGGCAGGCTTTGCTGGGGATGATGC 129
NM_001033618.1 CGCCCTCGTCATCGACAATGGTTCTGGCATGTGCAAAGCTGGCTTTGCTGGGGATGACGC 174
NM_001297986.1 TGCTCTGGTTGTCGACAACGGCTCTGGCATGTGCAAGGCCGGATTTGCCGGAGACGATGC 241
Y09623.1 CGCCCTGGTCGTCGACAATGGATCTGGAATGTGCAAAGCCGGTTTCGCCGGTGATGACGC 68
                    ** * ** ***** ** ***** * ** ** ** ** ** ** ** ** ** **
```

Зробити висновки.