před přednáškou do pdf skrýt výsledky procvičování

Základy praktické bioinformatiky

PETRA MATOUŠKOVÁ

2025/2026

Základy praktické bioinformatiky

Téma 4/10

Proteinová bioinformatika III

Cíle:

Student bude schopen porovnat sekvence dvou či více proteinů, určit míru identity, zobrazit jednoduchý fylogenetický strom vybraných proteinů a najít 3D strukturu a interakční partnery proteinu.

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod…)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce funkce → Hledání známých motivů

3D-struktura, vizualizace

→ Vyhledání **a porovnání** podobných sekvencí

Evoluční příbuznost sekvencí

Vyhledání interakcí mezi proteiny

…

Párové porovnání (pairwise alignment)

Globální porovnání – porovná sekvence po celé délce

Lokální porovnání – porovná sekvence částečně v „nejpodobnější“ oblasti (např pro vyhledání podobných domén, odhalení repetitivních sekvencí)

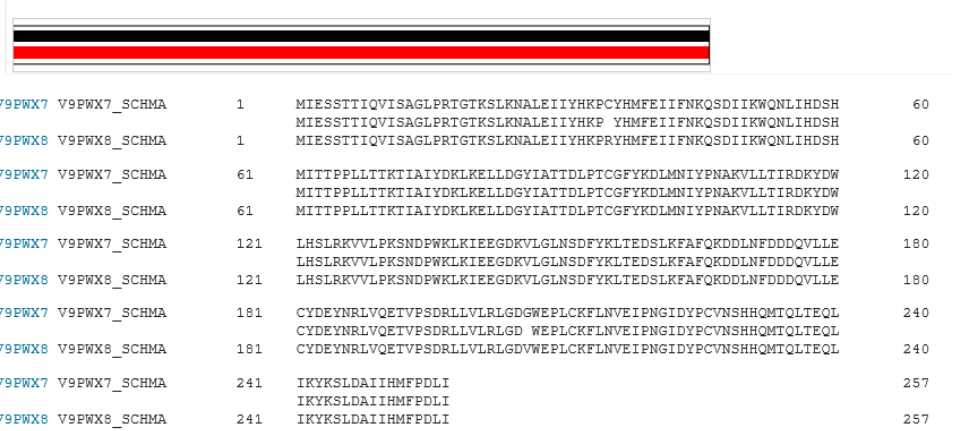
```
seq1   EARDF-NQYYSSIKRSGSIQ
         . : .:::::::::. . .
seq2   LPKLFIDQYYSSIKRTMG-H
```

global sequence alignment

```
seq1   NQYYSSIKRS
         .:::::::::.
seq2   DQYYSSIKRT
```

local sequence alignment

BLAST-součástí výstupu je párové porovnání (optimalizován na rychlost, ne přesnost)

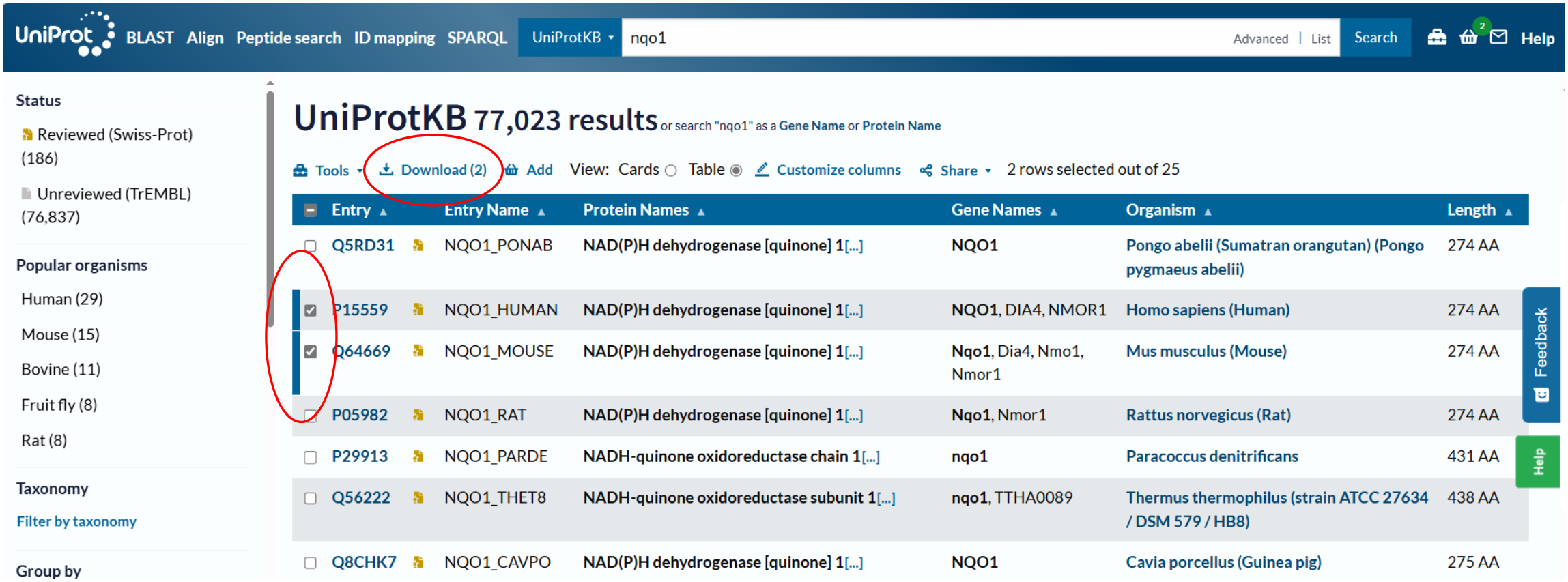


```
V9PWX7 V9PWX7_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPCYHMFEIIFNKQSDIIKWQNLIHDSH      60
                               MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKP YHMFEIIFNKQSDIIKWQNLIHDSH
V9PWX8 V9PWX8_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPRYHMFEIIFNKQSDIIKWQNLIHDSH      60

V9PWX7 V9PWX7_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW     120
                               MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW
V9PWX8 V9PWX8_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW     120

V9PWX7 V9PWX7_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE     180
                               LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE
V9PWX8 V9PWX8_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE     180

V9PWX7 V9PWX7_SCHMA     181    CYDEYNRLVQETVPSDRLLVLRLGDGWEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL     240
                               CYDEYNRLVQETVPSDRLLVLRLGD WEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL
V9PWX8 V9PWX8_SCHMA     181    CYDEYNRLVQETVPSDRLLVLRLGDVWEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL     240

V9PWX7 V9PWX7_SCHMA     241    IKYKSLDAIIHMFPDLI                                          257
                               IKYKSLDAIIHMFPDLI
V9PWX8 V9PWX8_SCHMA     241    IKYKSLDAIIHMFPDLI                                          257
```

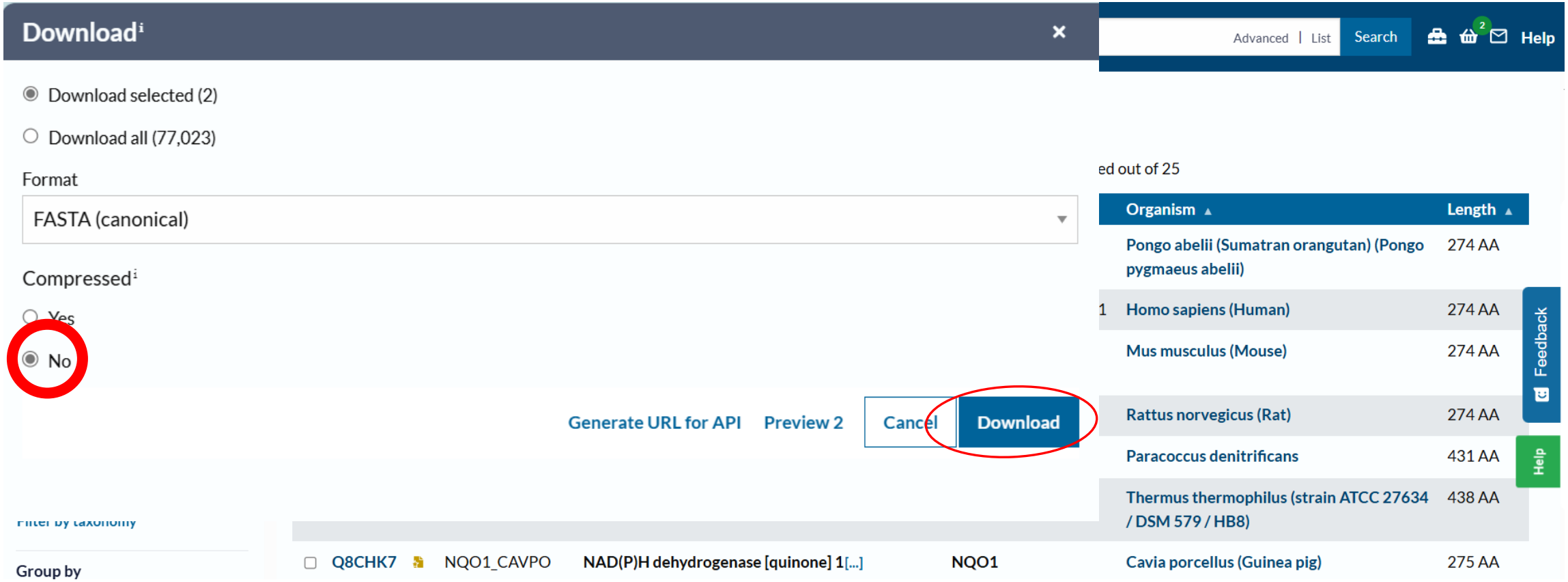
```
V9PWX7 V9PWX7_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPCYHMFEIIFNKQSDIIKWQNLIHDSH      60
                               M ESS  + VI AGLPRTGTKSLKNALEIIYHKPCYHM EII  + +DI KWQ L  ++
A0A183QDM9 A0A183QDM9_9TREM1   MSESSNDLLVIGAGLPRTGTKSLKNALEIIYHKPCYHMTEIIIKQHNDIDKWQKLFDEAL      60

V9PWX7 V9PWX7_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW     120
                                +       T  + I D LKE+L  Y A TD+P CGFYK+LMNIYPNAKVLLTIRDKYDW
A0A183QDM9 A0A183QDM9_9TREM61  KMEP-----TNELMINDGLKEILMNYGAVTDVPACGFYKELMNIYPNAKVLLTIRDKYDW     115

V9PWX7 V9PWX7_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKV----------------------------------     146
                               LHSLRKVVLPKSNDPWKLKIEEGDKV
A0A183QDM9 A0A183QDM9_9TREM116 LHSLRKVVLPKSNDPWKLKIEEGDKVILTIRNKYDWLSSFRQTLMPKSNDSSNRTIDEAD     175

V9PWX7 V9PWX7_SCHMA     147    --LGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLECYDEYNRLVQETVPSDRLLVLRLG     204
                                 L L   F K+  DS+K AF+K D + D+D  +L+C+DEYNR V ETVPS+RLL+ +LG
A0A183QDM9 A0A183QDM9_9TREM176 EILKLGPKFIKMAIDSMKLAFRKIDFDIDNDNEMLQCFDEYNRTVIETVPSERLLIRKLG     235

V9PWX7 V9PWX7_SCHMA     205    DGWEPLCKFLNVEIPNGIDYPCVN                                   228
                               DGWEPLC+FLNV++P G+ YP +N
A0A183QDM9 A0A183QDM9_9TREM236 DGWEPLCRFLNVDVPEGVSYPYIN                                   259
```

Nevýhoda-ukáže jen jeden „nejlepší“ alignment

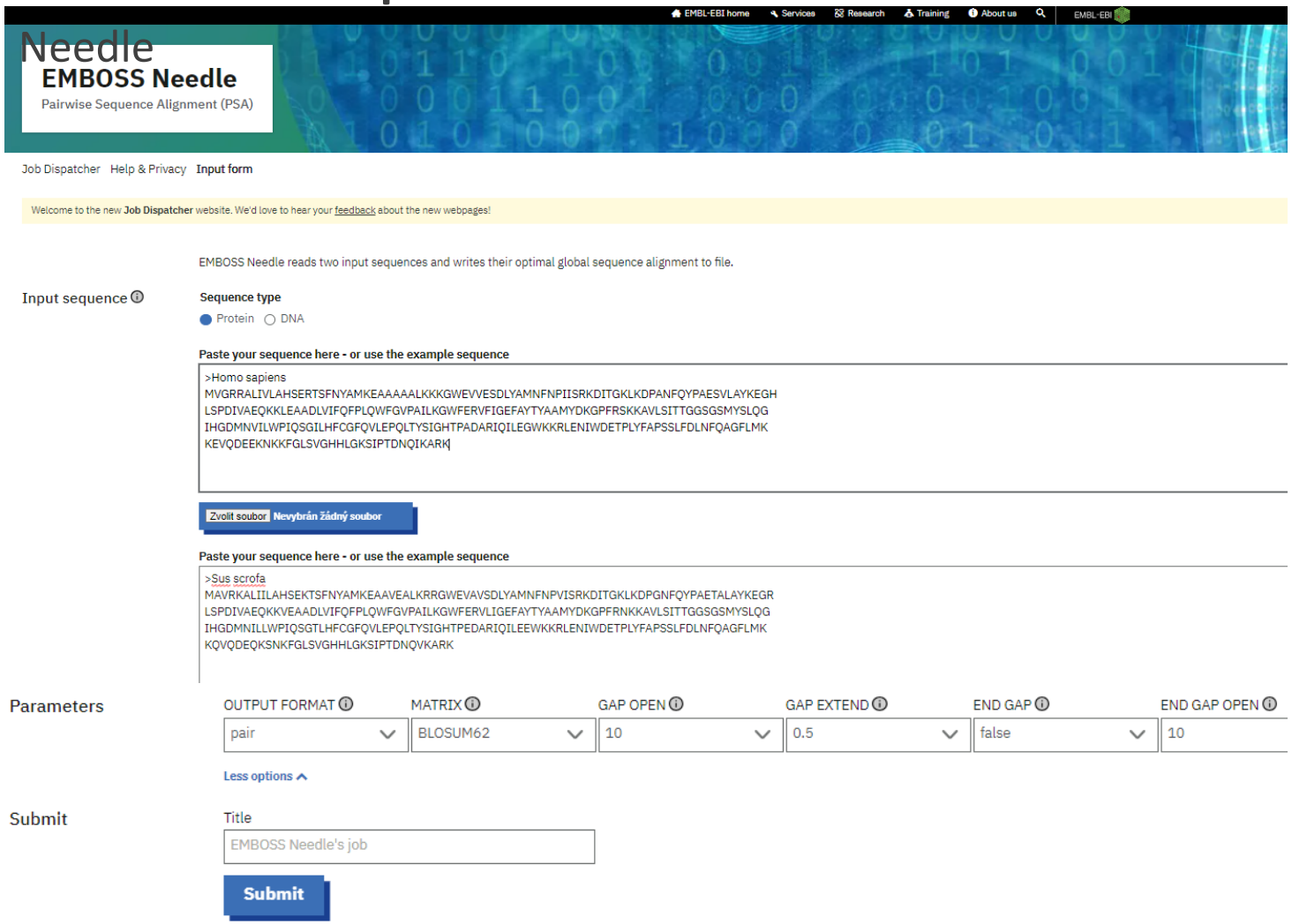
Stažení víc sekvencí najednou:



Stažení víc sekvencí najednou:

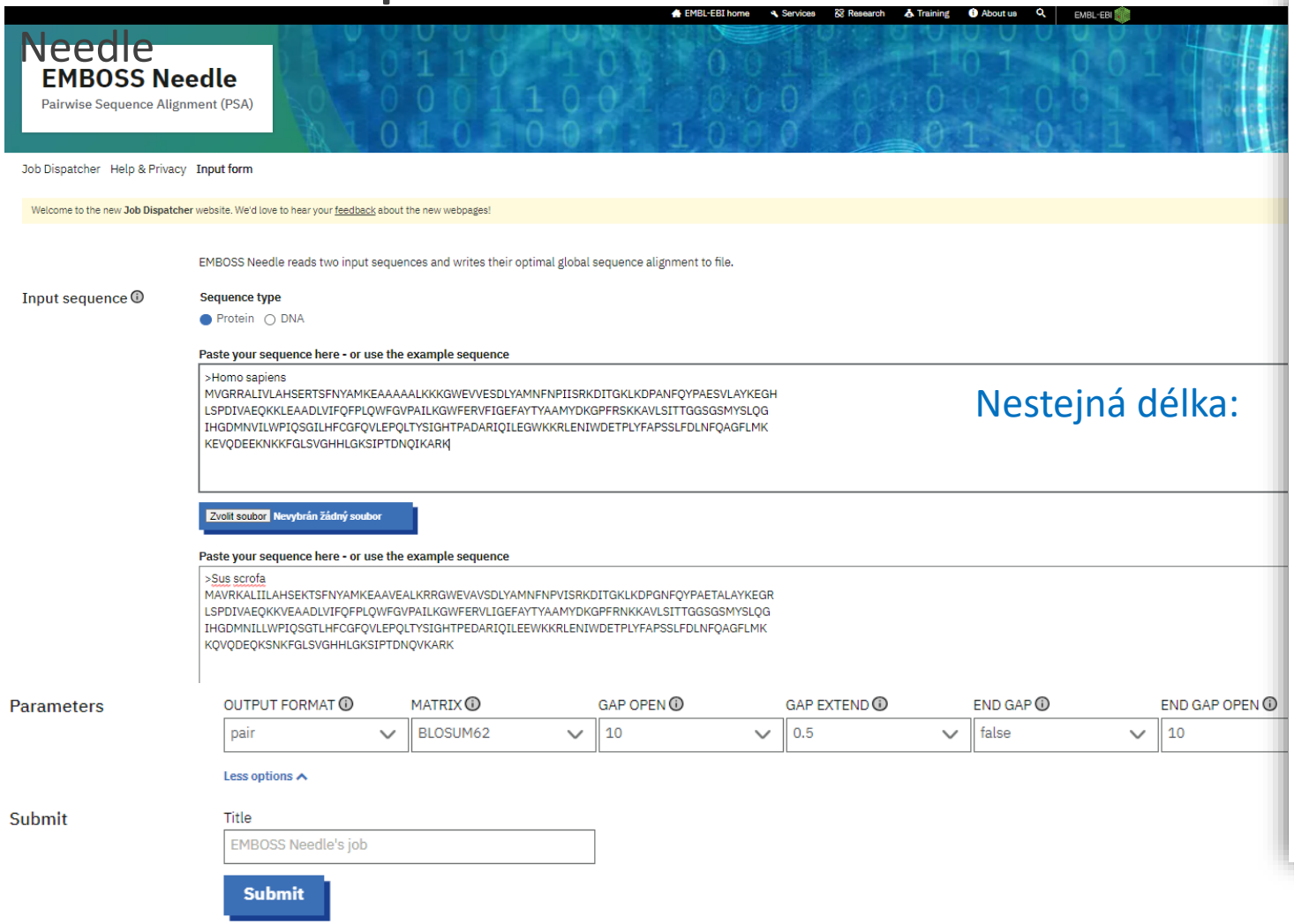


Párové porovnání - Globální




```
#=======================================
#
# Aligned_sequences: 2
# 1: Homo
# 2: Sus
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 274
# Identity:     245/274 (89.4%)
# Similarity:   260/274 (94.9%)
# Gaps:           0/274 ( 0.0%)
# Score: 1320.0
#
#
#=======================================

Homo              1 MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPI     50
                    |..|:|||:|||||:||||||||||..|||::|||||..||||||||||:
Sus               1 MAVRKALIILAHSEKTSFNYAMKEAAVEALKRRGWEVAVSDLYAMNFNPV     50

Homo             51 ISRKDITGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIF    100
                    |||||||||||||.|||||||:.||||||.|||||||||||:||||||||
Sus              51 ISRKDITGKLKDPGNFQYPAETALAYKEGRLSPDIVAEQKKVEAADLVIF    100

Homo            101 QFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTG    150
                    ||||||||||||||||||||.||||||||||||||||||:||||||||||
Sus             101 QFPLQWFGVPAILKGWFERVLIGEFAYTYAAMYDKGPFRNKKAVLSITTG    150

Homo            151 GSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADA    200
                    |||||||||||||::||||||||.|||||||||||||||||||||.||
Sus             151 GSGSMYSLQGIHGDMNILLWPIQSGTLHFCGFQVLEPQLTYSIGHTPEDA    200

Homo            201 RIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNK    250
                    ||||||.||||||||||||||||||||||||||||||||:||||:|:.
Sus             201 RIQILEEWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKKQVQDEQKSN    250

Homo            251 KFGLSVGHHLGKSIPTDNQIKARK    274
                    ||||||||||||||||||:||||
Sus             251 KFGLSVGHHLGKSIPTDNQVKARK    274
```

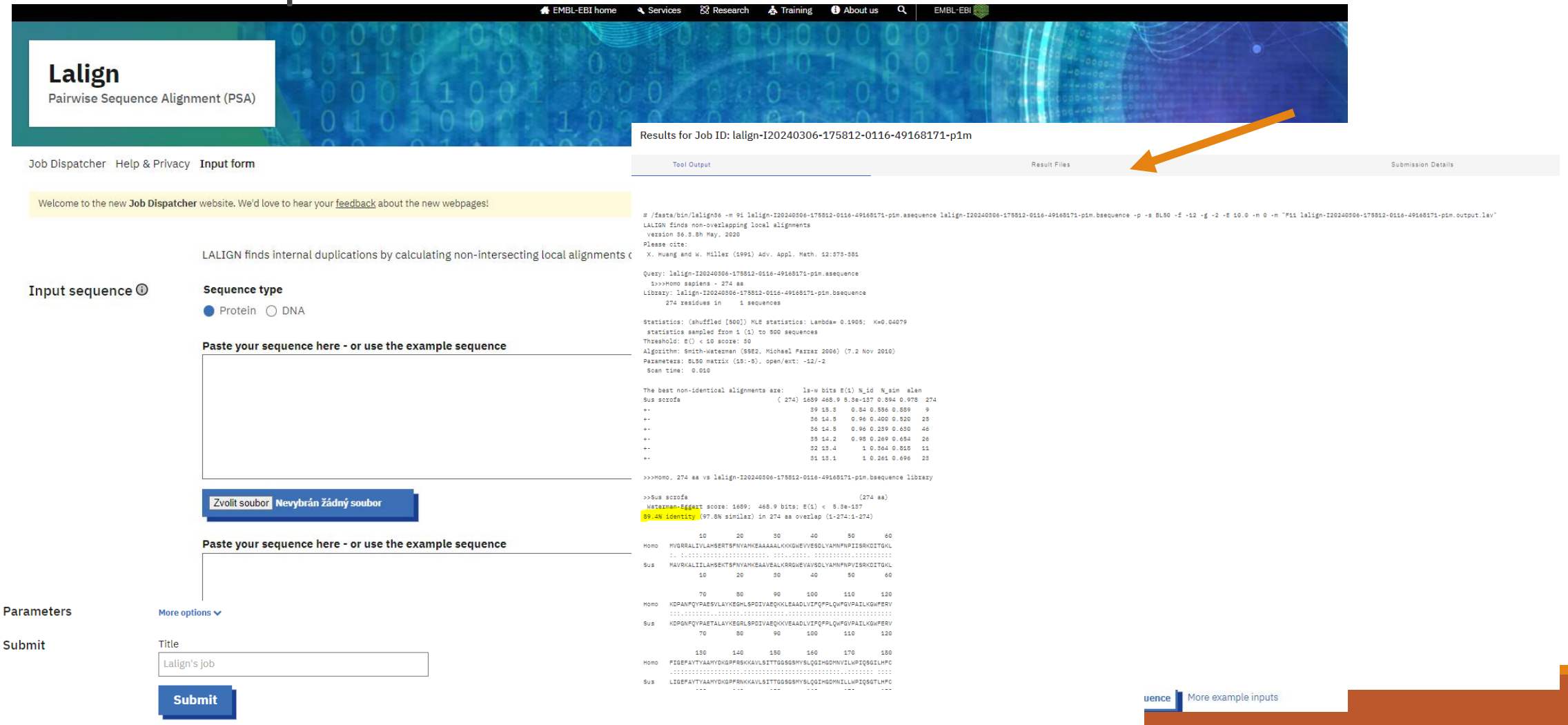
Párové porovnání - Globální



Nestejná délka:

```
#=======================================
#
# Aligned_sequences: 2
# 1: AAH09679.1
# 2: AAI47025.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 342
# Identity:     138/342 (40.4%)
# Similarity:   196/342 (57.3%)
# Gaps:          51/342 (14.9%)
# Score: 646.5
#
#
#=======================================

AAH09679.1         1 MVSPATRKSLPKVKAMDFITSTAILPLLFGCLGVFGLFRLLQWVR---GK      47
                                          .:..:||||.  ||:.||..:.|.|.   .|
AAI47025.1         1 ------------------MAMLMLPLLL--LGISGLLFIYQEVSRLWSK      29

AAH09679.1        48 AYLRNAVVVITGATSGLGKECAKVFYAAGAKLVLCGRNGGALEELIRELT      97
                     :.::|.|||||.|.||||||||:||:..||:|||||:|...||.|...|
AAI47025.1        30 SAVQNKVVVITDAISGLGKECARVFHTGGARLVLCGKNWERLENLYDAL-      78

AAH09679.1        98 ASHATKVQTHKPYLVTFDLTDSGAIVAAAAEILQCFGYVDILVNNAGISY     147
                     .|.|...:|..|.||..||:|...:...|.|:|.|:|.||||:|||.:..
AAI47025.1        79 ISVADPSKTFTPKLVLLDLSDISCVPDVAKEVLDCYGCVDILINNASVKV     128

AAH09679.1       148 RGTIMDTTVDVDKRVMETNYFGPVALTKALLPSMIKRRQGHIVAISSIQG     197
                     :|.....:::::||::|:.|||||:.|||||||:||.||.|.||.:::|||
AAI47025.1       129 KGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRRTGQIVLVNNIQG     178

AAH09679.1       198 KMSIPFRSAYAASKHATQAFFDCLRAEMEQYEIEVTVISPGYIHTNLSVN     247
                     |..||||:.|||||||...||||||||:|:|::.::.:||.:|.   |.:
AAI47025.1       179 KFGIPFRTTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIR---SYH     225

AAH09679.1       248 AITADGS-------------RYGVMDTTTAQGRSPVEVAQDVLAAVGKKK     284
                     .....|:             .|||         .|||||::|:..|.:||
AAI47025.1       226 VYPEQGNWEASIWKFFFRKLTYGV---------HPVEVAEEVMRTVRRKK     266

AAH09679.1       285 KDVILADLLPSLAVYLRTLAPGLFFSLMASRARKERKSKNS-     325
                     ::|.:|:.|..|||:||..|..||:::|...:::.....
AAI47025.1       267 QEVFMANPIPKAAVYVRTFFPEFFFAVVACGVKEKLNVPEEG     308
```

Párové porovnání - Lokální



Párové porovnání – Lokální (vizualizace)

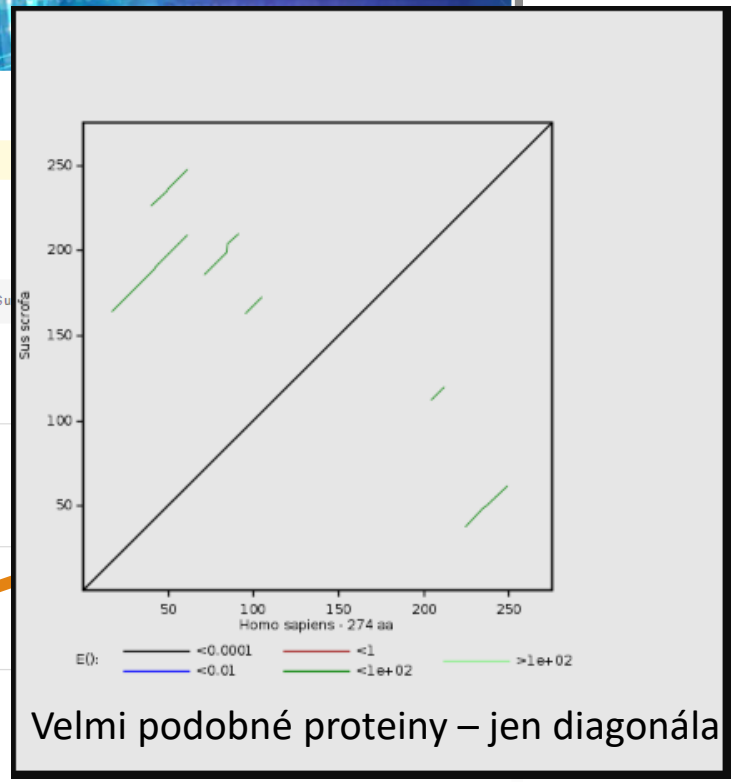
Lalign
Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy **Input form**

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

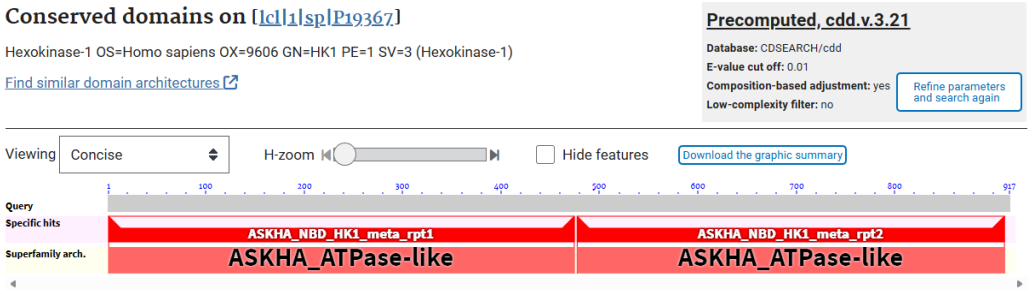
Results for Job ID: lalign-I20240306-175812-0116-49168171-p1m

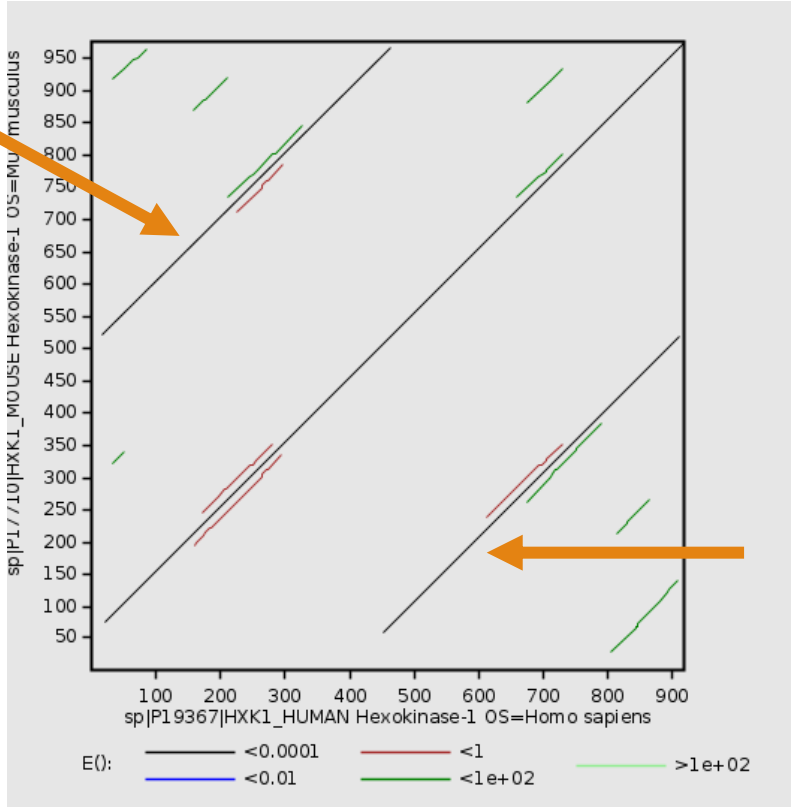
Tool Output	Result Files	
Tool Output	lalign-I20240306-175812-0116-49168171-p1m.out	Download
Visual Result (SVG)	lalign-I20240306-175812-0116-49168171-p1m.visual-svg	Download
Visual Result (PNG)	lalign-I20240306-175812-0116-49168171-p1m.visual-png	Download
Visual Result (JPEG)	lalign-I20240306-175812-0116-49168171-p1m.visual-jpg	Download
First Input Sequence	lalign-I20240306-175812-0116-49168171-p1m.asequence	Download



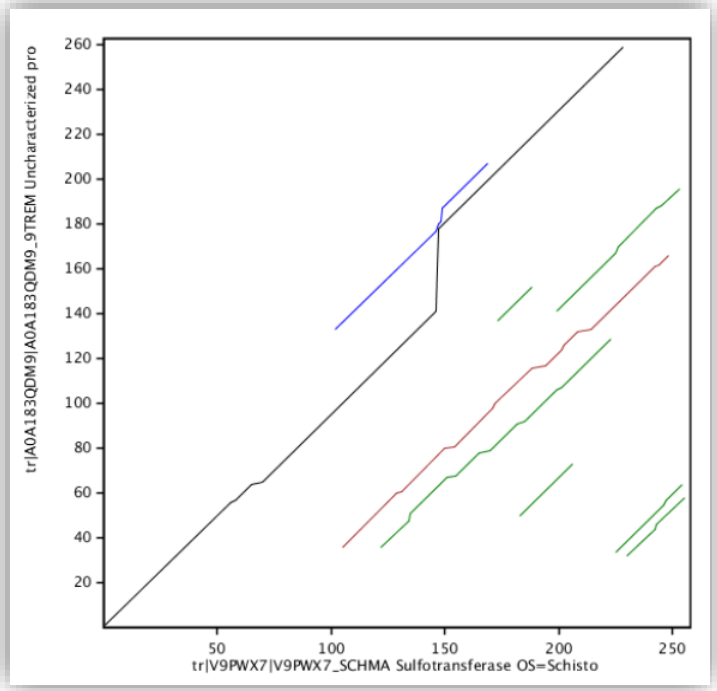
Velmi podobné proteiny – jen diagonála

– Lokální (vizualizace)





Dvě „opakující“ se domény



Podobnější proteiny (kratší je na vodorovné ose – svislá část na diagonále= chybějící úsek)

Vyzkoušejte si porovnat 2 sekvence

Porovnejte váš (lidský) protein s homologním proteinem z myši (Mus musculus). (DÚ)

Vyzkoušejte oba programy:

- Jaká je identita/similarita těchto dvou proteinů?
  → globální porovnání
- Jak dlouhé je nejlepší lokální porovnání těchto dvou sekvencí?
  → lokální porovnání

Vyzkoušejte si BLAST/porovnání

BLASTp:

Existuje homologní protein také u rostlin? (např Arabidopsis thaliana, huseníček rolní)?

- **jak je ten na tom s podobností?**

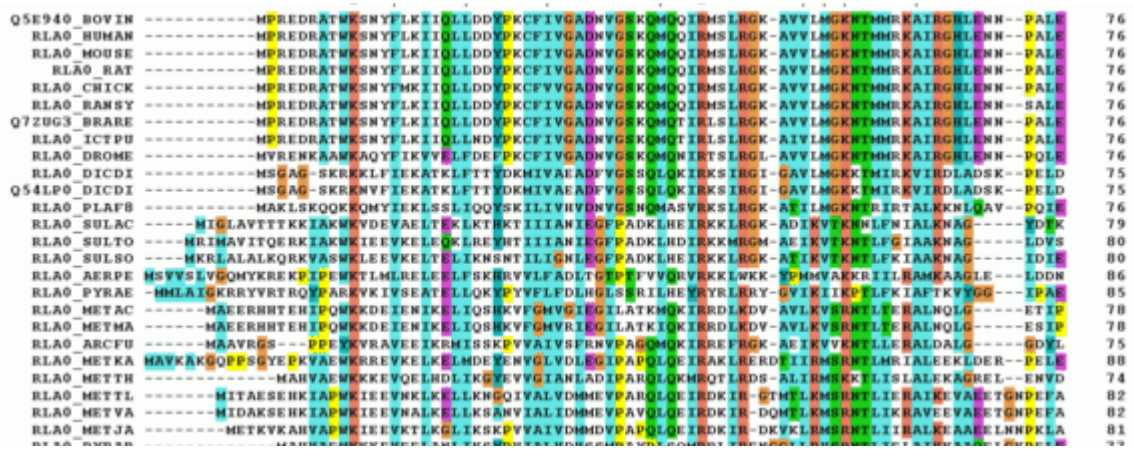
Mnohonásobné porovnání (multiple sequence alignment)

= přepsání sekvencí, tak aby stejné aminokyseliny či oblasti byly srovnány pod sebou ve sloupcích

Získání sekvencí (např: porovnání různých isoforem, porovnání vícero výstupů z BLASTu..)

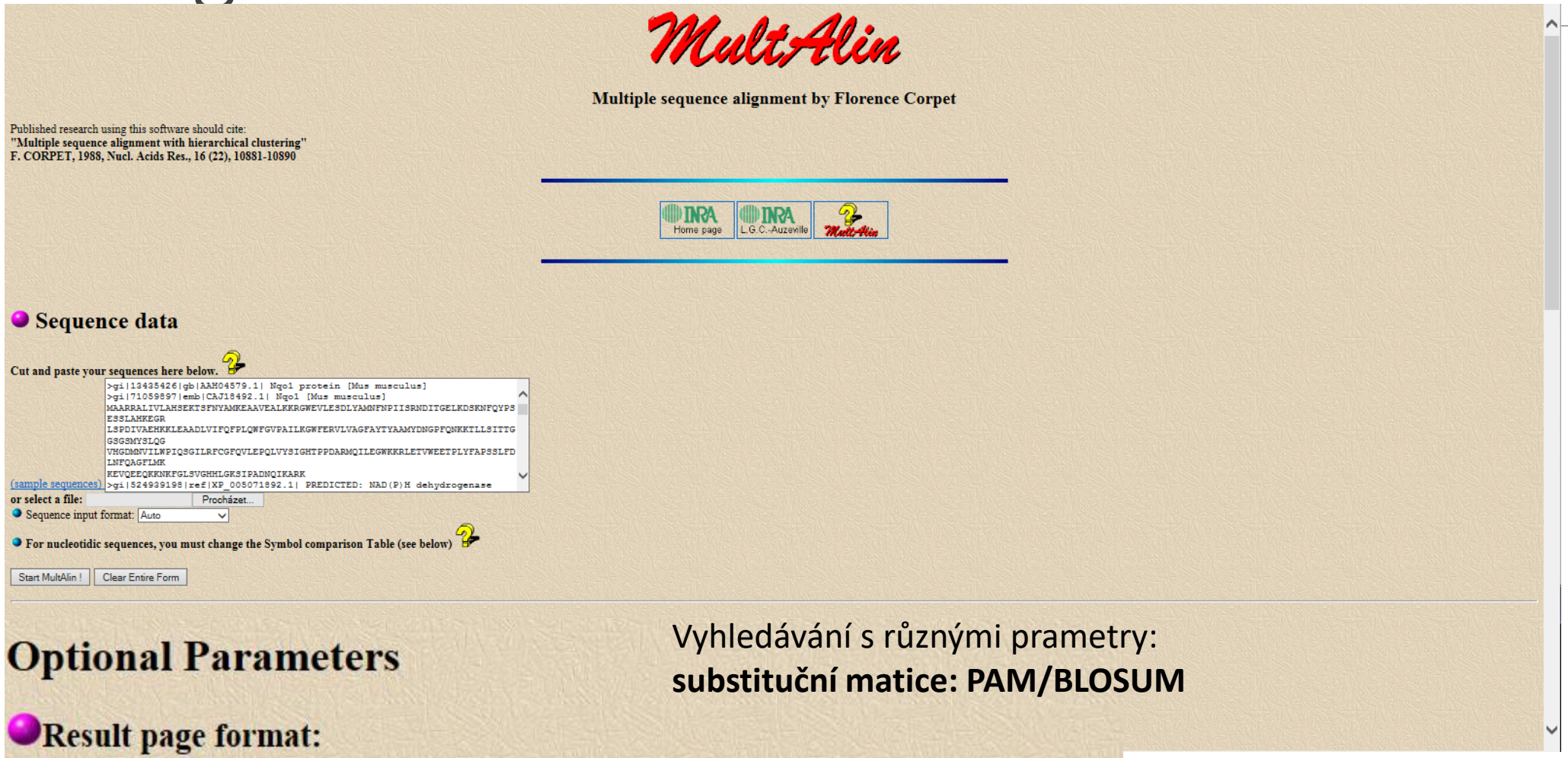
Vlastní porovnání (různé programy, různé matice, různé „výstupy“…)

Umožňuje fylogenetickou analýzu (fylogenetické stromy - evoluční příbuznost)



Alignment: MultAlin

http://multalin.toulouse.inra.fr/multalin/



Vyhledávání s různými prametry:
substituční matice: PAM/BLOSUM

MultAlin

Multalin result page

Back | Home page | L.G.C.-Auzeville

Go directly to Alignment

```
Multalin version 5.4.1
Copyright I.N.R.A. France 1989, 1991, 1994, 1996
Published research using this software should cite
Multiple sequence alignment with hierarchical clustering
F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890
Symbol comparison table: blosum62
Gap weight: 12
Gap length weight: 2
Consensus levels: high=90% low=50%
Consensus symbols:
 ! is anyone of IV
 $ is anyone of LM
 % is anyone of FY
 # is anyone of NDQEBZ

 MSF:   274    Check:   0        .
 Name: gi|13435426|gb Len:   274  Check: 6705  Weight:  1.23
 Name: gi|524939198|r Len:   274  Check: 6867  Weight:  1.23
 Name: gi|227430403|r Len:   274  Check: 6661  Weight:  0.89
 Name: gi|426242583|r Len:   274  Check: 5108  Weight:  0.89
 Name: gi|386781783|r Len:   274  Check: 4019  Weight:  0.89
 Name: gi|30230685|gb Len:   274  Check: 4190  Weight:  0.89
 Name: Consensus      Len:   274  Check: 4506  Weight:  0.00

//
```

Available files:
-Sequence Input file
-Cluster file
-Results as a fasta file
-Results as a text page (msf)

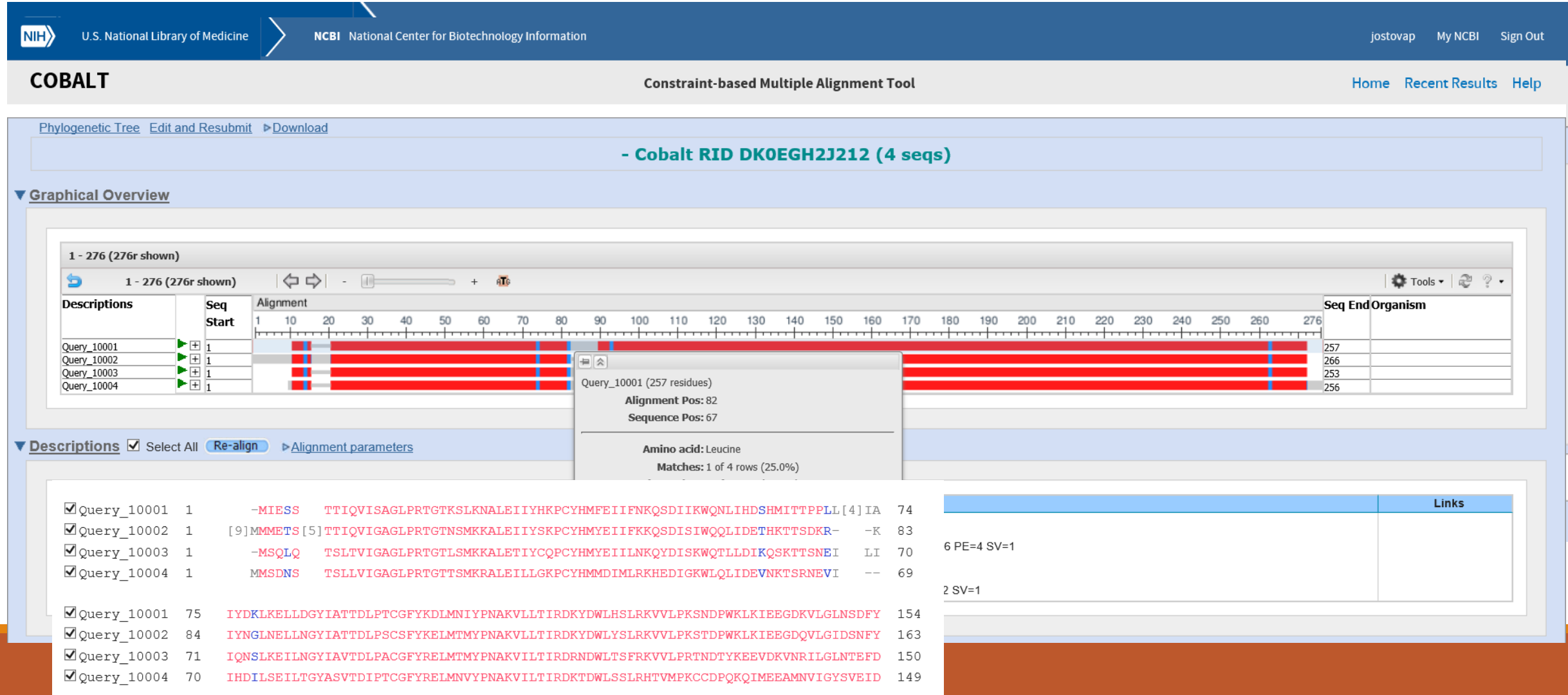
-Results as postscript page(s) with ESPript (protein only)
-Alignment and tree description (rfd) Get a better view of your protein family : phylogenetic tree, pruned tree and subtrees, summarised coloured alignment and subalignments.
- Results as an html page (needs to enable style sheets)

-Results as a text page with colour indications (need a text editor)
-Results as a gif image

Table 9-7 Patterns of Conservation in Multiple Sequence Alignments

Amino Acid	*Characteristic*
W,Y, F	It is common to find conserved tryptophans. Tryptophan is a large hydrophobic residue that sits deep in the core of proteins. It plays an important role in their stability and is therefore difficult to mutate. When tryptophan mutates, it is usually replaced by another aromatic amino acid, such as phenylalanine or tyrosine. Patterns of conserved aromatic amino acids constitute the most common signatures for recognizing protein domains.
G, P	It is common to find conserved columns with a glycine or a proline in a multiple alignment. These two amino acids often coincide with the extremities of well-structured beta strands or alpha helices. (For more on these structures, see Chapter 11.)
C	Cysteines are famous for making C-C (disulphide) bridges. Conserved columns of cysteines are rather common and usually indicate such bridges. Columns of conserved cysteines with a specific distance provide a useful signature for recognizing protein domains and folds.
H, S	Histidine and serine are often involved in catalytic sites, especially those of proteases. Conserved histidine or a conserved serine are good candidates for being part of an active site.
K, R, D, E	These charged amino acids are often involved in ligand binding. Highly conserved columns can also indicate a salt bridge inside the core of the protein.
L	Leucines are rarely very conserved unless they're involved in protein-protein interactions such as a leucine zipper.

Alignment: NCBI/COBALT

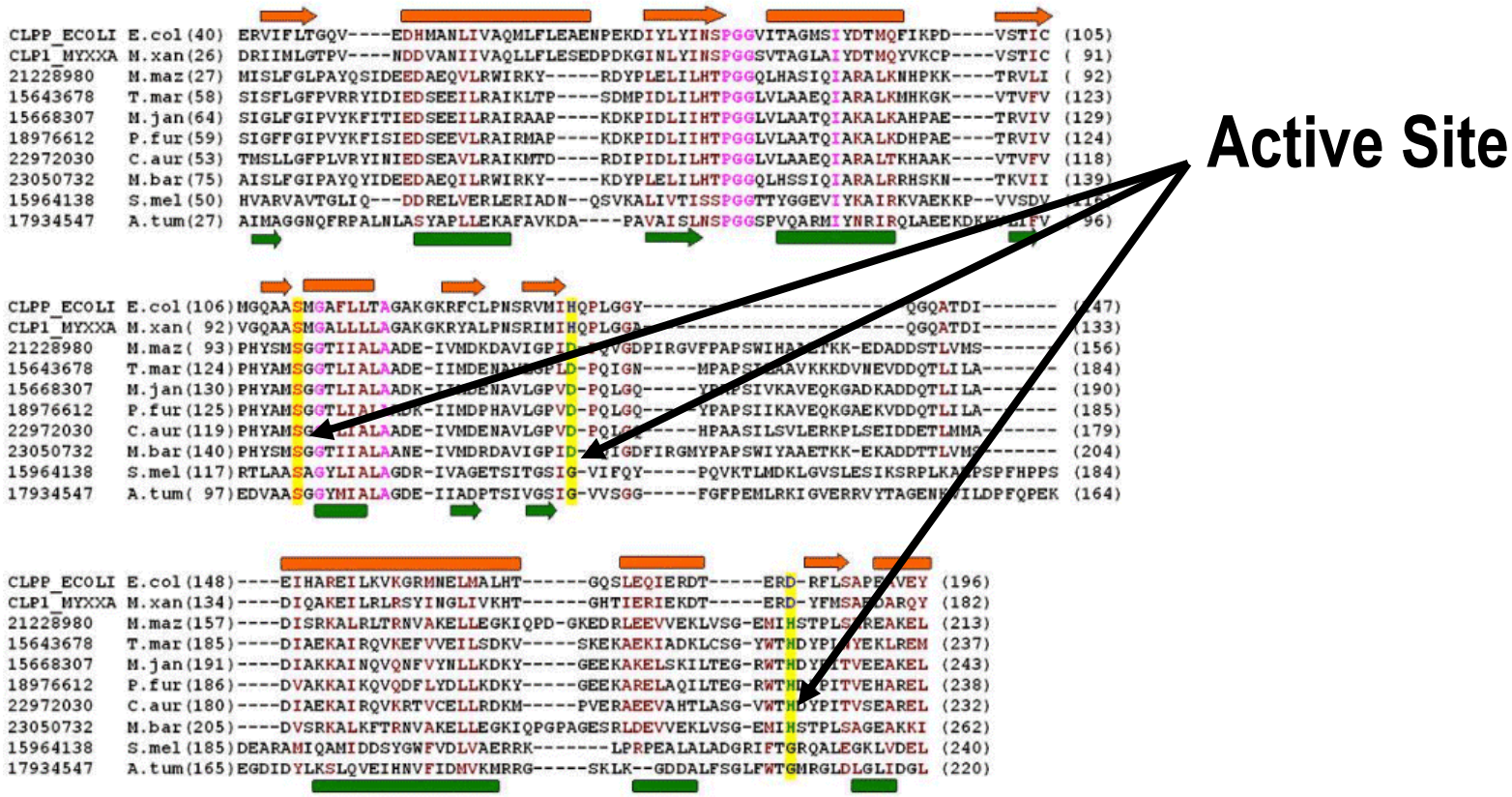


„Evoluce v kostce“

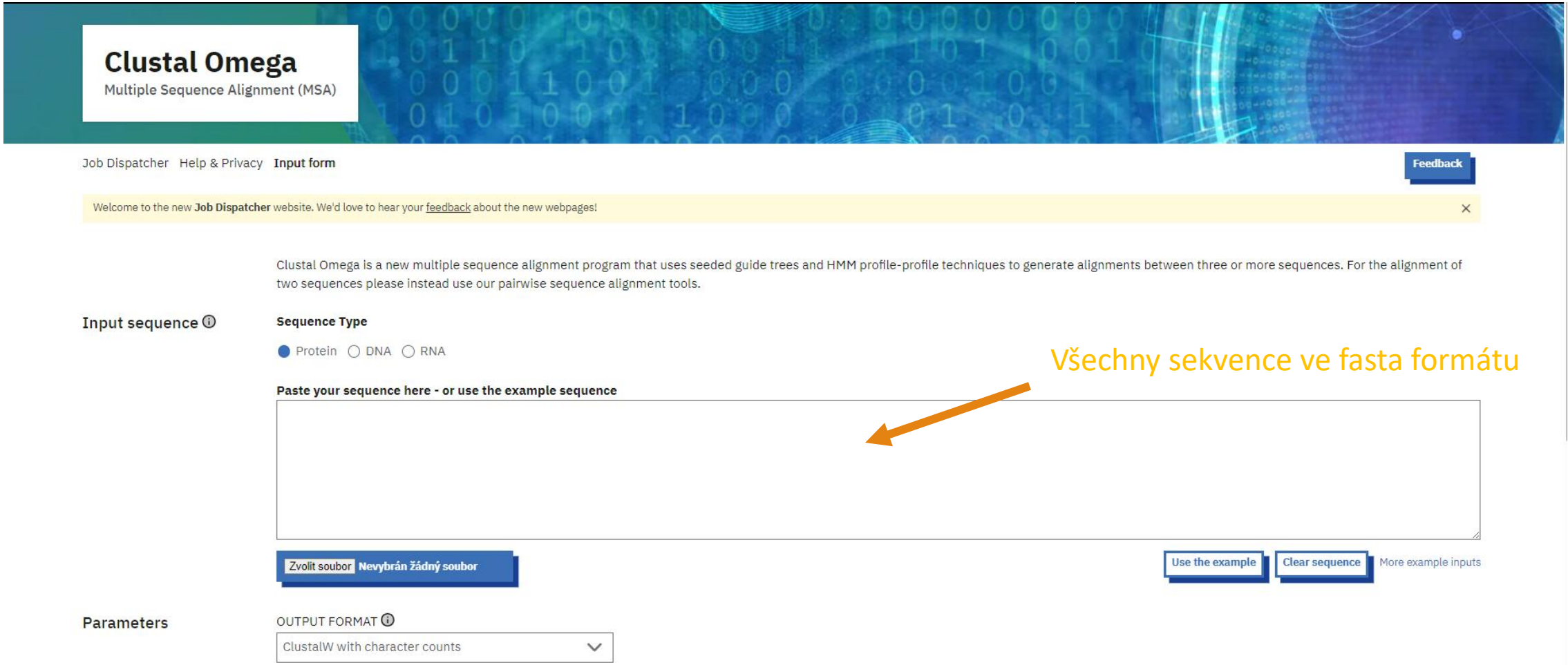
- Aminokyseliny mutují náhodně
- Mutace jsou následně organismem akceptovány nebo odmítnuty
- Pokud je to mutace škodlivá, je odmítnuta
  - Vymizí z genomu
- Mutace důležitých míst / pozic v proteinu (např. podílejících se na aktivním místě) jsou vždycky škodlivé
- **Významné pozice se snadno identifikují, protože nemutují!**

- Mnohonásobné porovnání umožňuje identifikovat tyto ***konzervované*** pozice

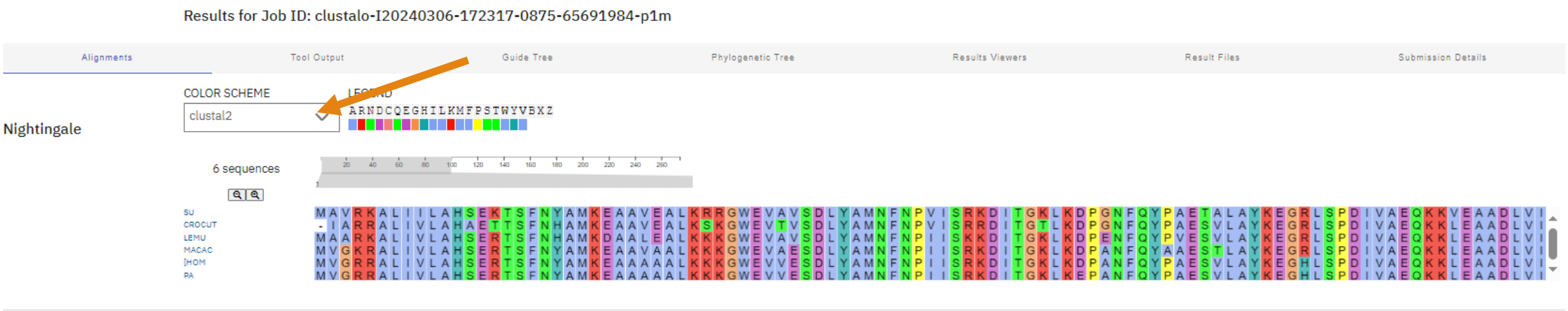
Ukázka aktivních míst:
(The Serine Proteases Active Site)



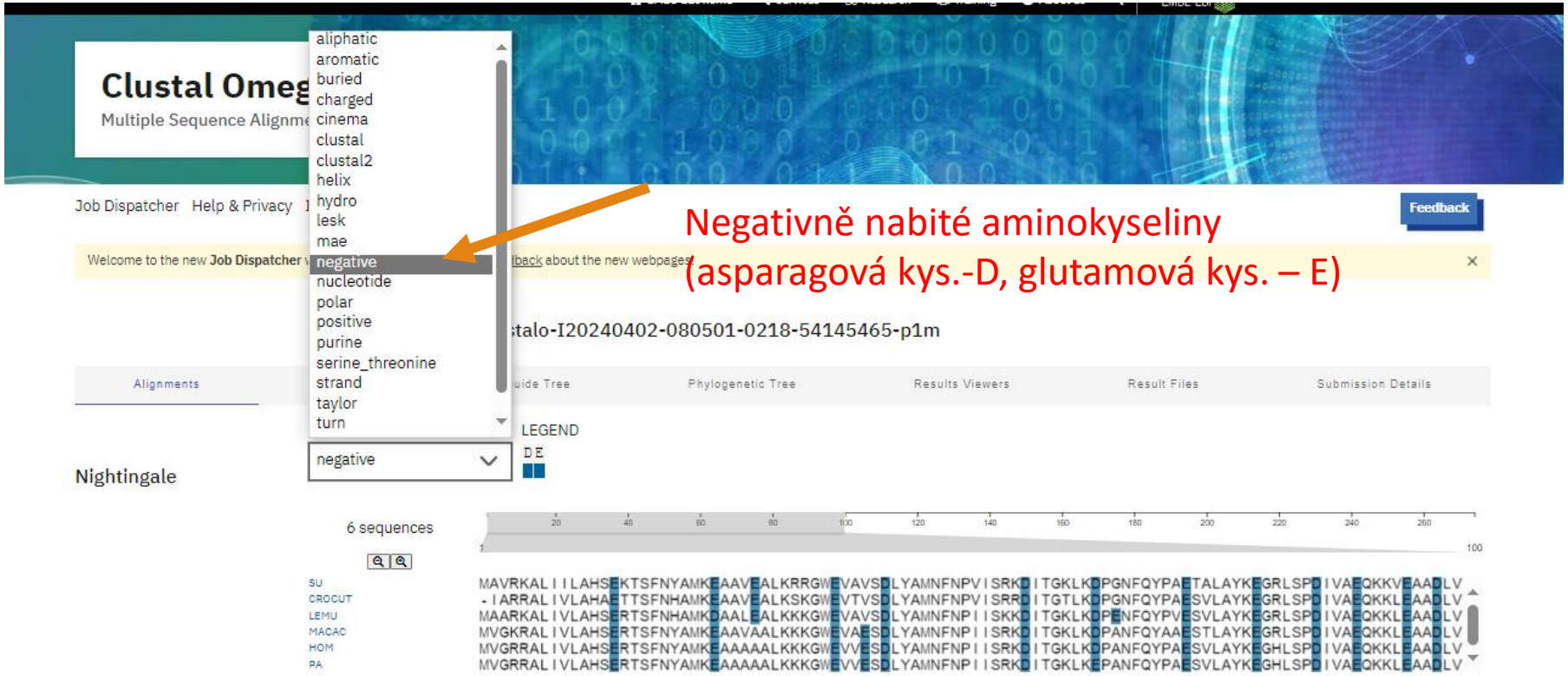
Alignment: Clustal Omega



Alignment: Clustal Omega



Alignment: Clustal Omega



Alignment: Clustal Omega

Results for Job ID: clustalo-I20240306-17[illegible]17-0875-65691984-p1m

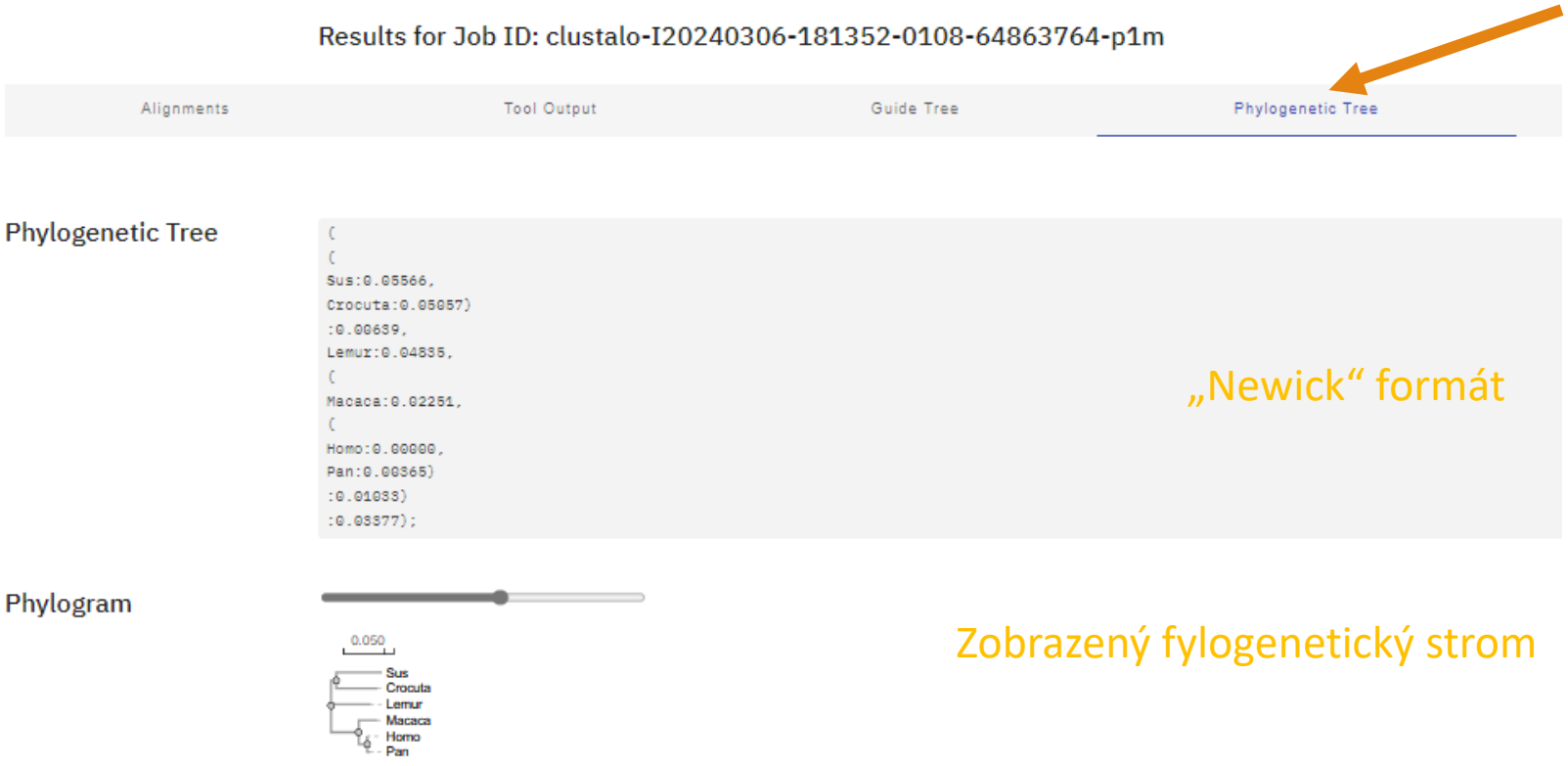
Alignments | Tool Output | Guide Tree | Phylogenetic Tree | Results Viewers | Result Files | Submission Details

Nightingale

```
CLUSTAL O(1.2.4) multiple sequence alignment

Sus         MAVRKALIILAHSEKTSFNYAMKEAAVEALKRRGWEVAVSDLYAMNFNPVISRKDITGKL      60
Crocuta     -IARRALIVLAHAETTSFNHAMKEAAVEALKSKGWEVTVSDLYAMNFNPVISRRDITGTL      59
Lemur       MAARKALIVLAHSERTSFNHAMKDAALEALKKKGWEVAVSDLYAMNFNPIISKKDITGKL      60
Macaca      MVGKRALIVLAHSERTSFNYAMKEAAVAALKKKGWEVAESDLYAMNFNPIISRKDITGKL      60
[Homo       MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL      60
Pan         MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL      60
                ::***:***:* ****:***:**  *** :****. **********:**::****.*

Sus         KDPGNFQYPAETALAYKEGRLSPDIVAEQKKVEAADLVIFQFPLQWFGVPAILKGWFERV      120
Crocuta     KDPGNFQYPAESVLAYKEGRLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFDRV      119
Lemur       KDPENFQYPVESVLAYKEGRLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV      120
Macaca      KDPANFQYAAESTLAYKEGRLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV      120
[Homo       KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV      120
Pan         KEPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV      120
            *:* **** .*:.******:**********:**************************:**

Sus         LIGEFAYTYAAMYDKGPFRNKKAVLSITTGGSGSMYSLQGIHGDMNILLWPIQSGTLHFC      180
Crocuta     LIGGFAYTYAAMYDNGPFRNKKTVLSITTGGSGSMYSLQGIHGDMNIILWPIQSGTLHFC      179
Lemur       LIGEFAYSYAAMYDKGPFRNKKTLLSITTGGSGSMYSLQGVHGDMNVILWPLQSGTLHFC      180
Macaca      FVGEFAYTLAAMYDKGPFQSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC      180
[Homo       FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC      180
Pan         FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC      180
            ::* ***: *****:***:.**::****************:*****::***:*** ****

Sus         GFQVLEPQLTYSIGHTPEDARIQILEEWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK      240
Crocuta     GFQVLEPQLTYSIGHTPVDARIQILDGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK      239
Lemur       GFQVLEPQLTYSIGHTPADARIQILAGWKKRLENIWNETPLYFAPSSLFDLNFQAGFVMK      240
Macaca      GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK      240
[Homo       GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK      240
Pan         GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK      240
            ********************* *******  *********:**********************:**

Sus         KQVQDEQKSNKFGLSVGHHLGKSIPTDNQVKARK 274
Crocuta     KEVQDEQKNKKFGLSVGHHLGKSIPADNQIRARK 273
Lemur       KEVQDEQKNKKFGLSVGHHLGKSIPTDNQIKARK 274
Macaca      KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK 274
[Homo       KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK 274
Pan         KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK 274
            *:****:*.:*****************:***::***
```

(*) konzervované aminokyseliny
(:) aminokyseliny s podobnou velikostí a hydrofobicitou
(.) aminokyseliny s podobnou velikosti NEBO hydrofobicitou

Alignment: Clustal Omega

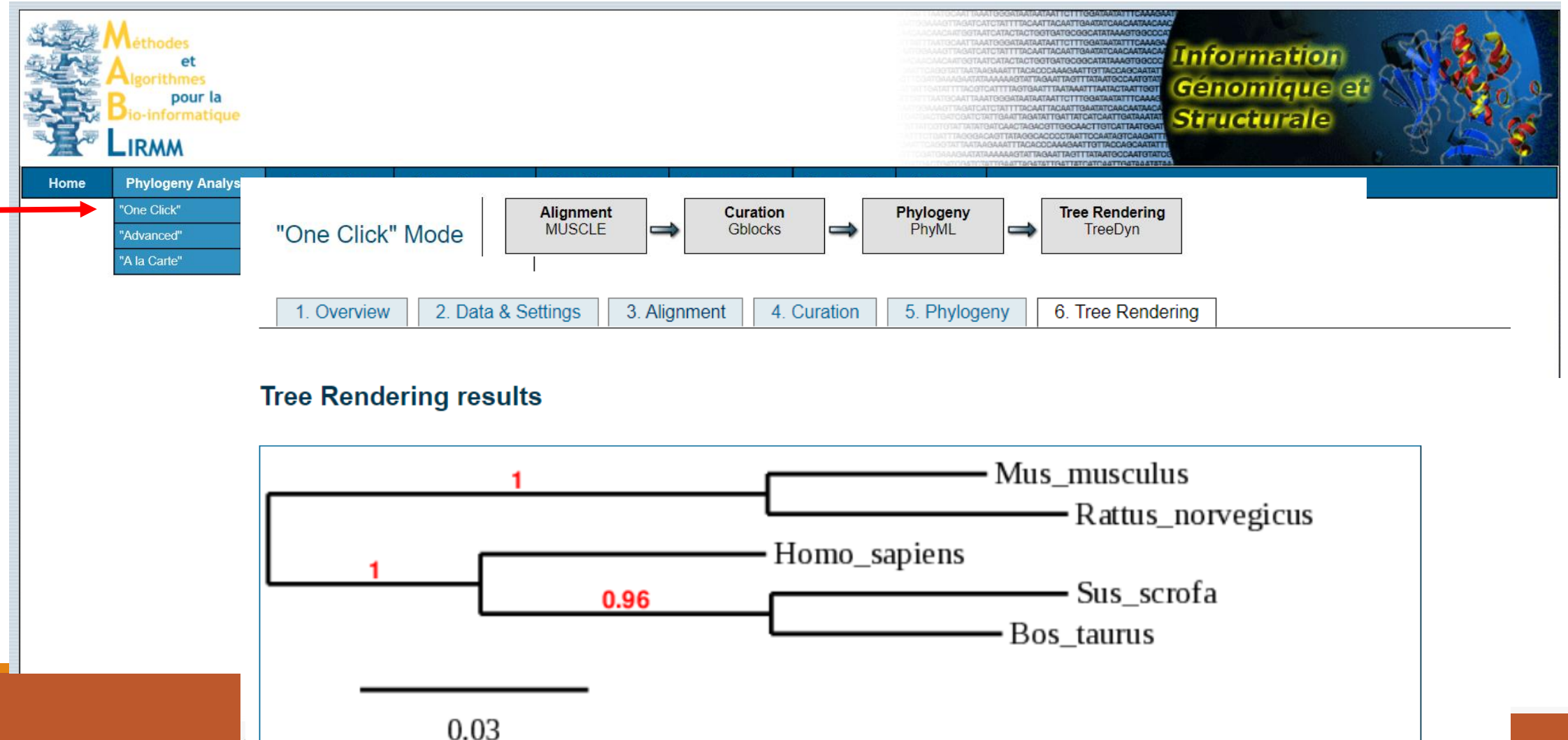


Vyzkoušejte si porovnat sekvence uložené v minulém úkolu

Vyzkoušejte víc programů!

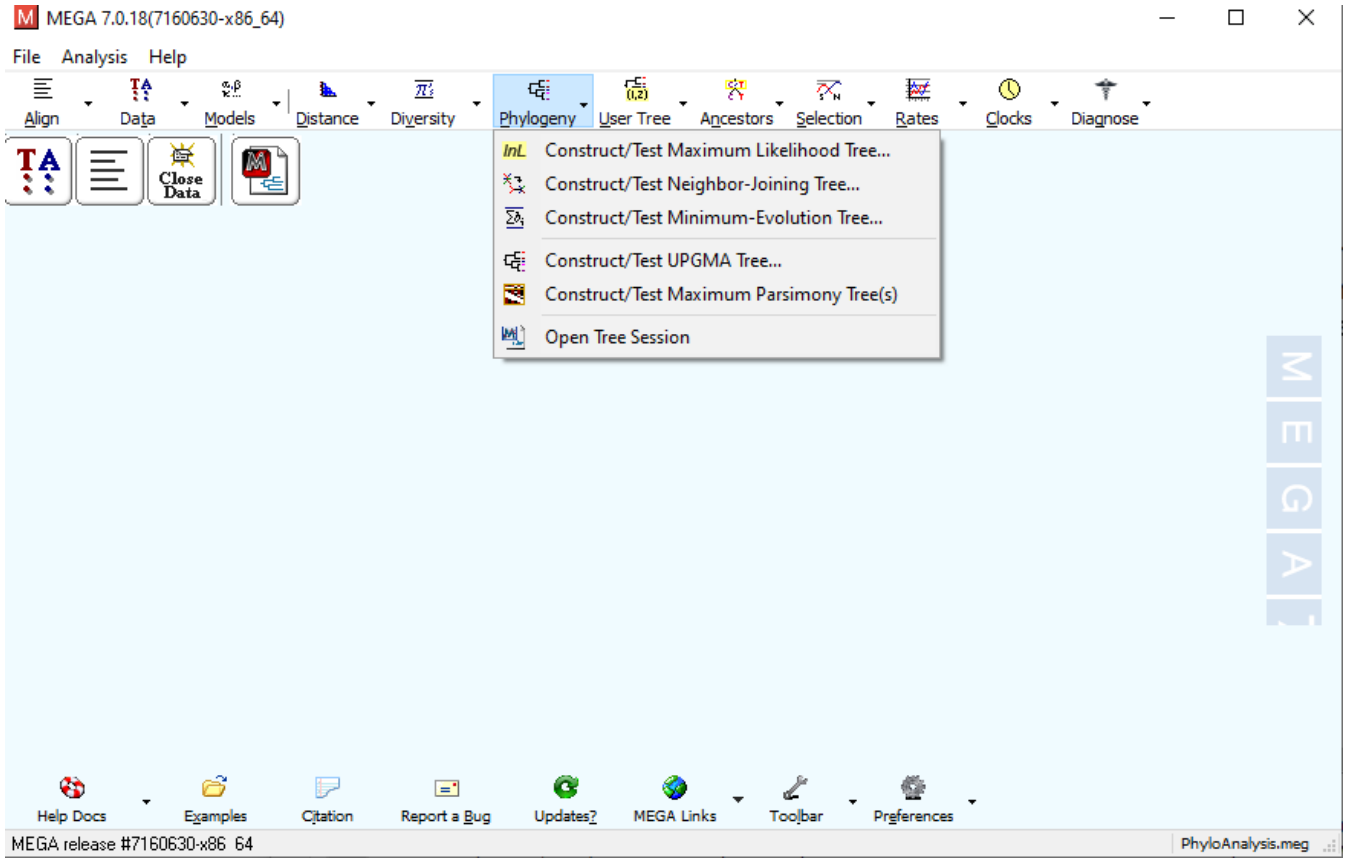
„pokročilejší“ fylogenetické porovnávání

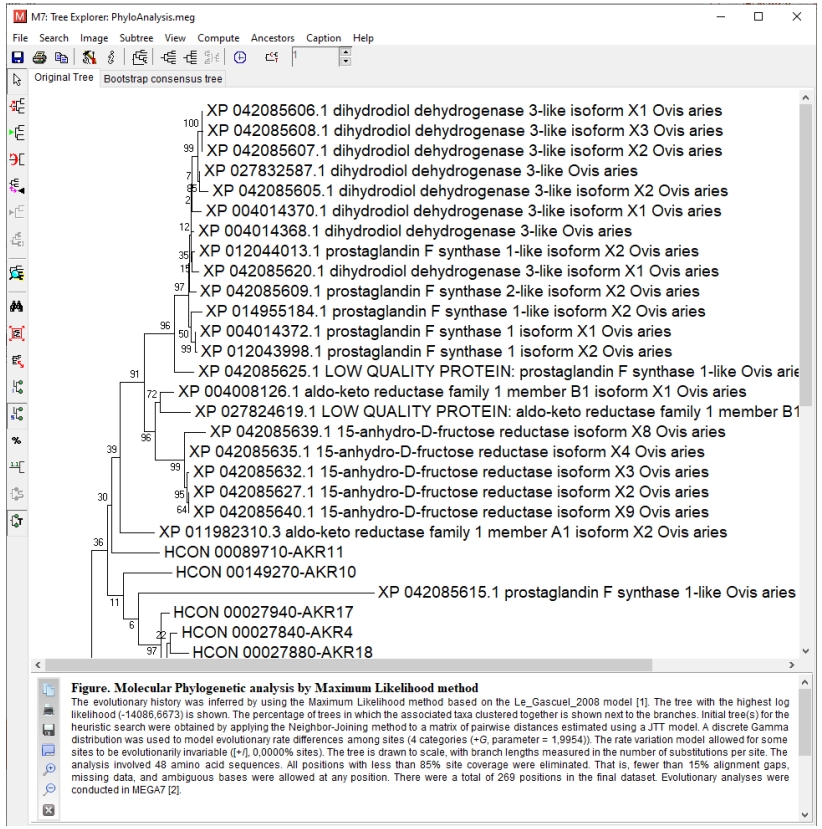
„pokročilejší“ fylogenetické porovnávání



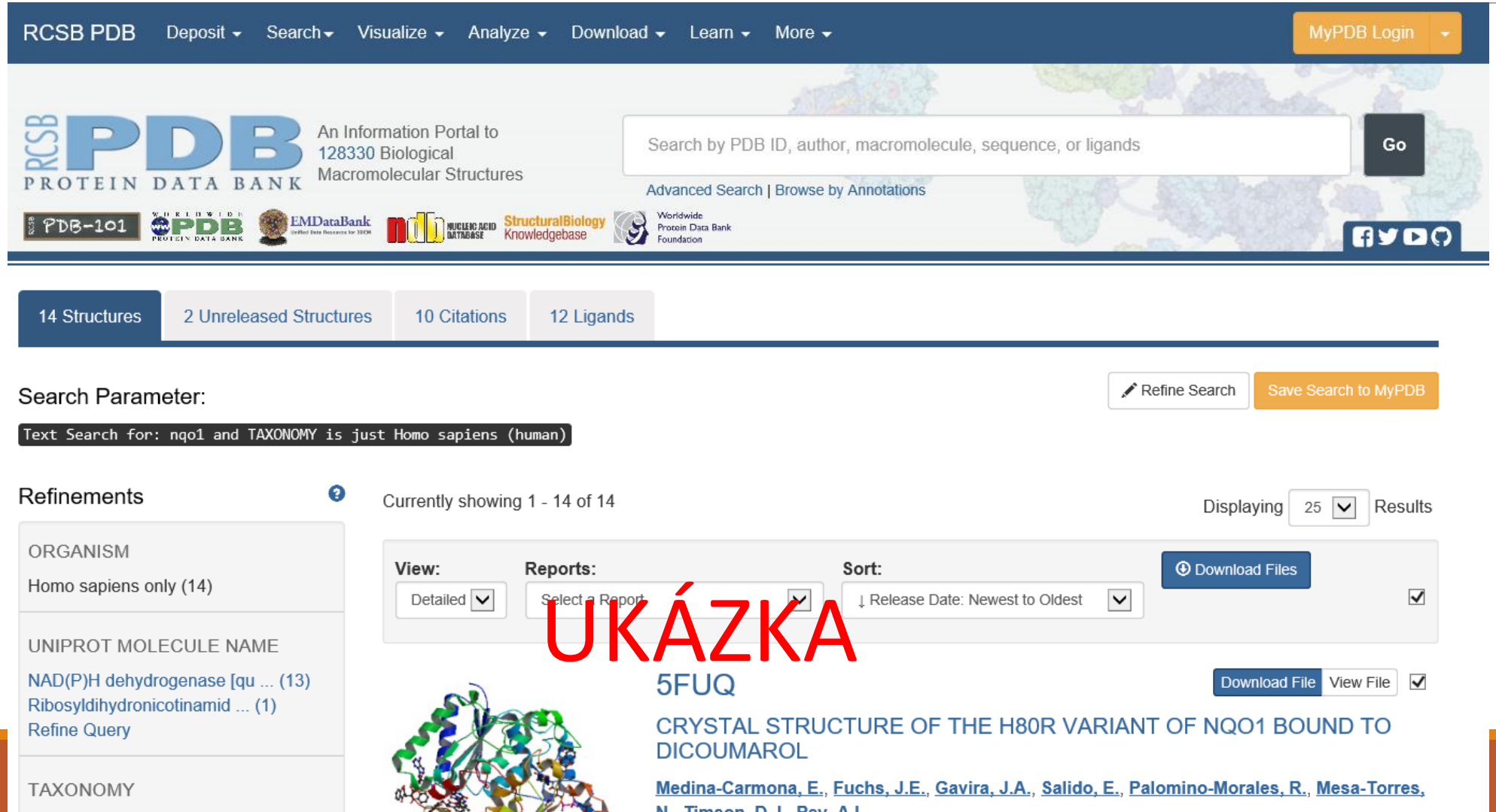
„ještě pokročilejší“ fylogenetické porovnávání

program MEGA (nutná instalace)

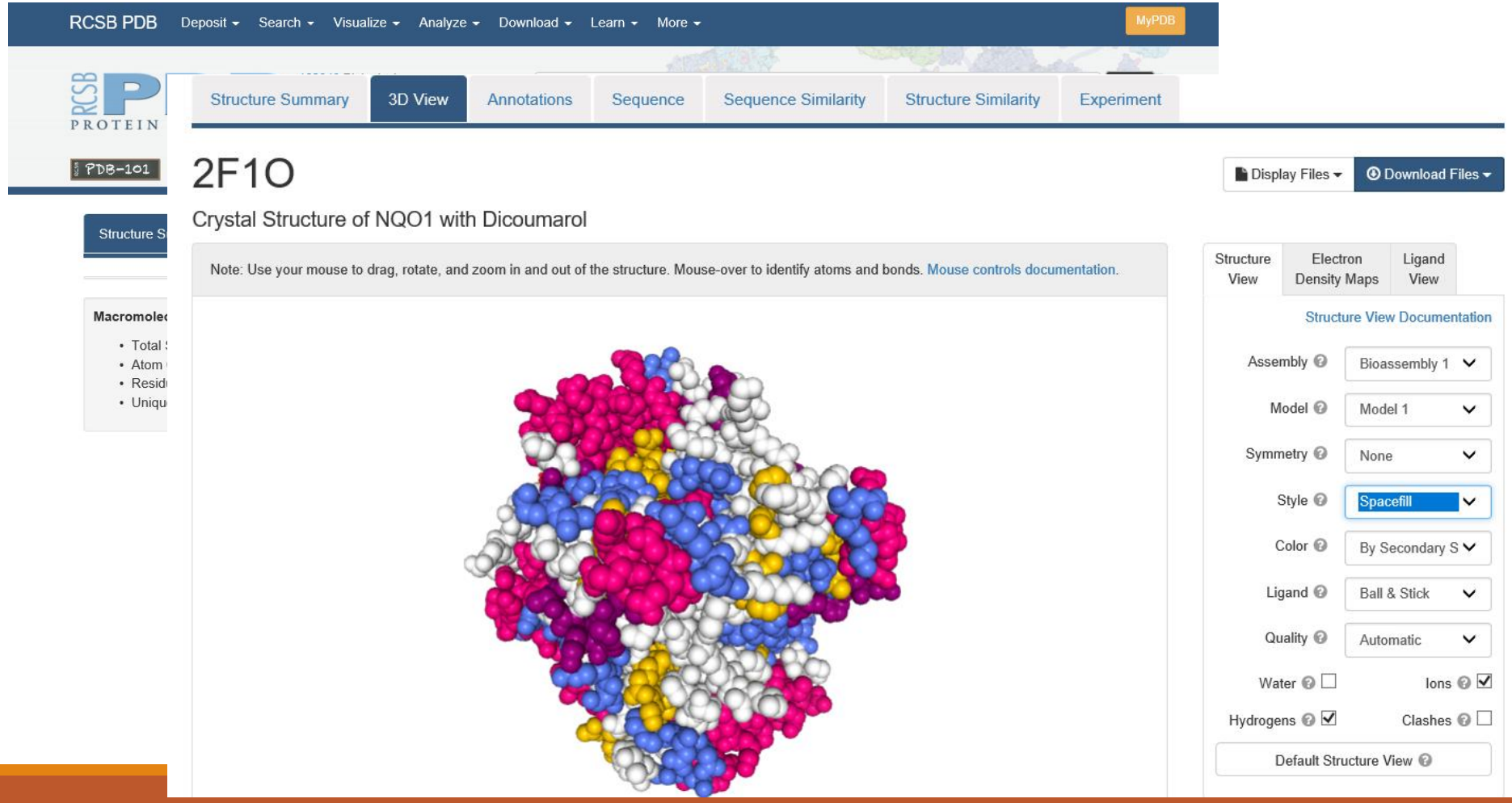




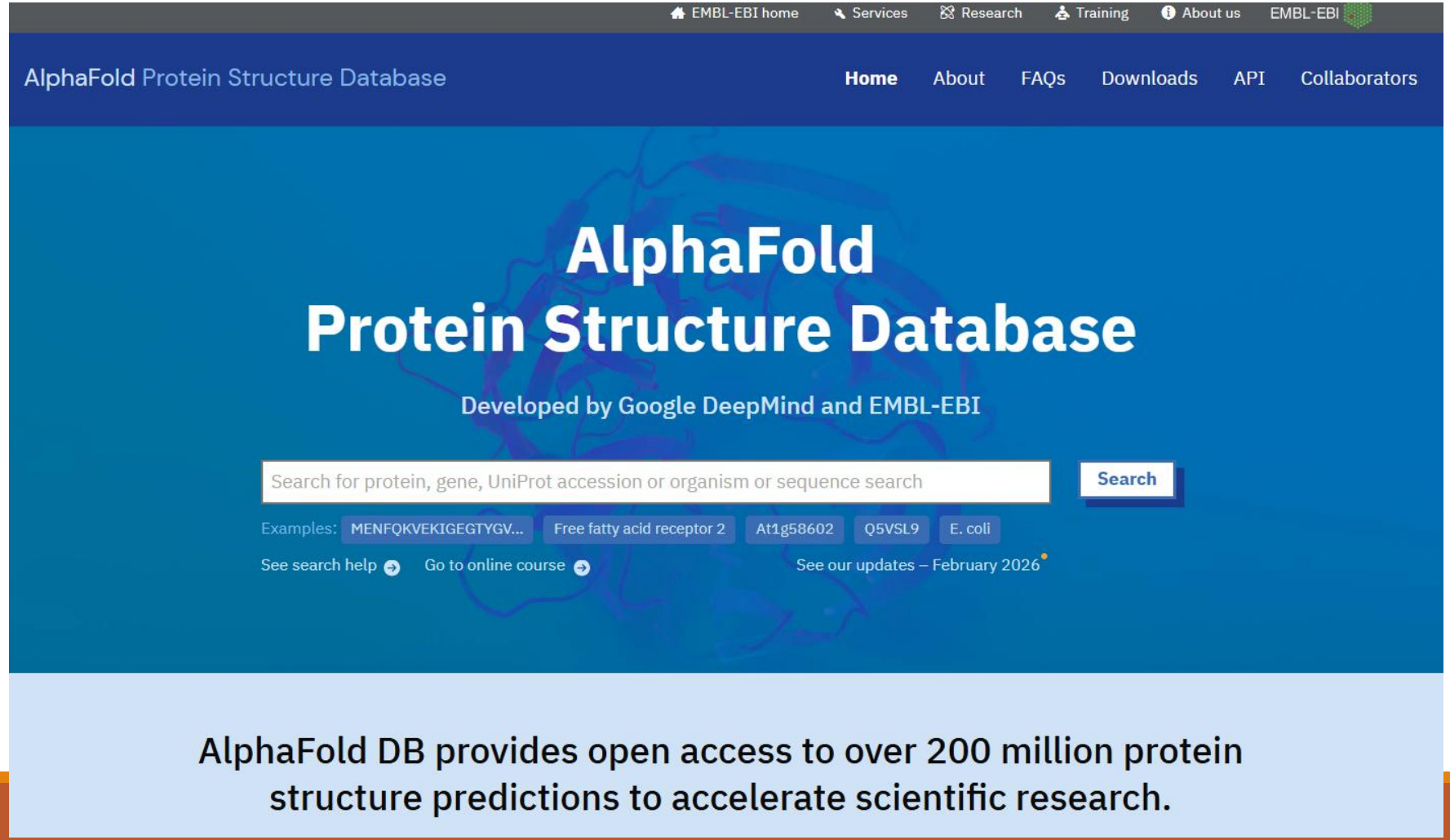
3-D struktura proteinů: PDB



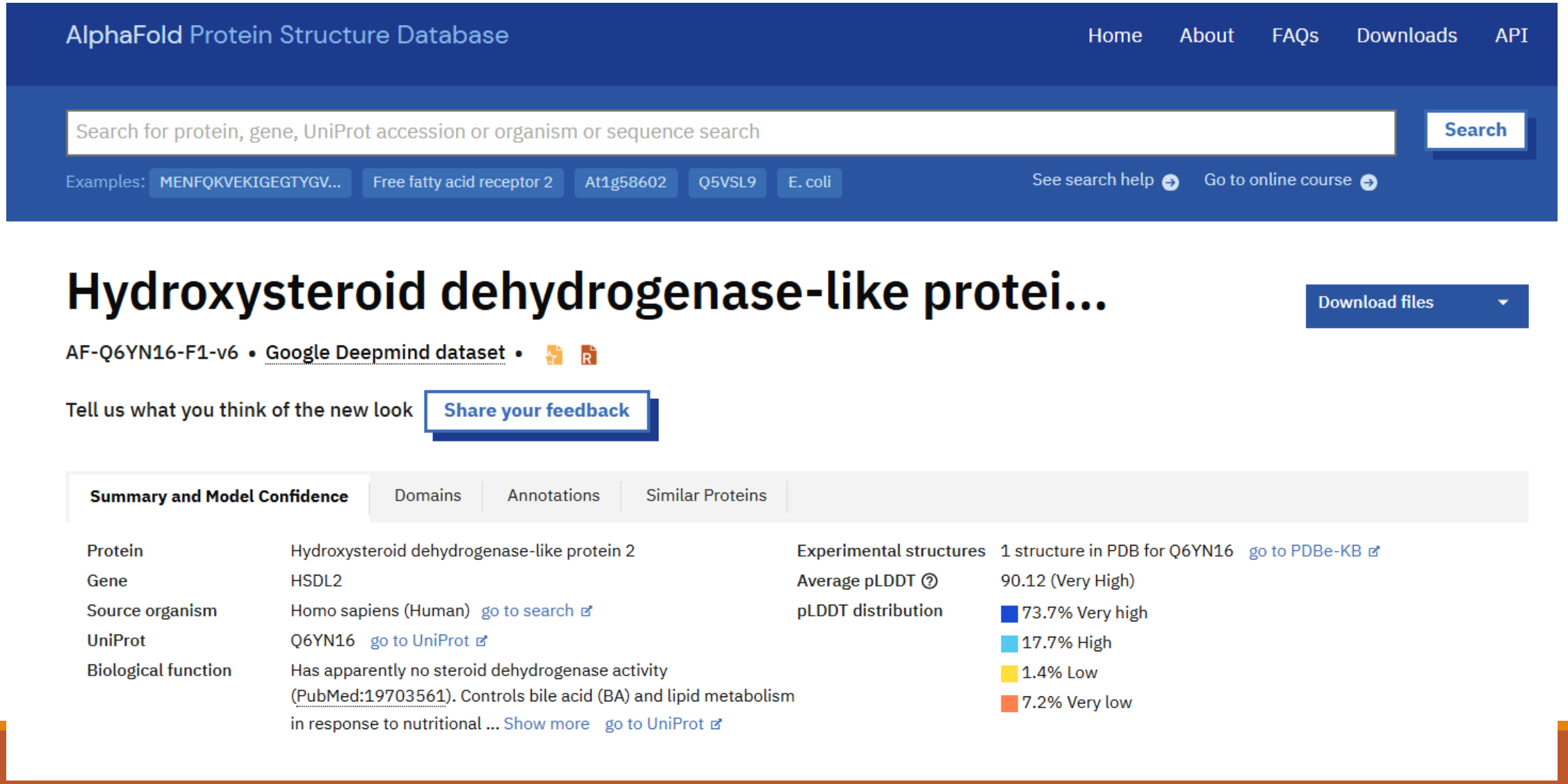
3-D struktura proteinů: PDB



3D struktura „vypočítaná“ **Alphafold**



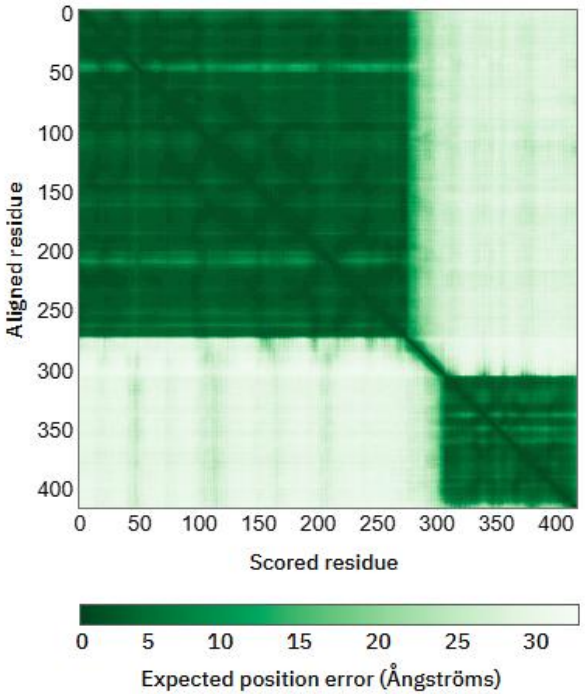
Alphafold

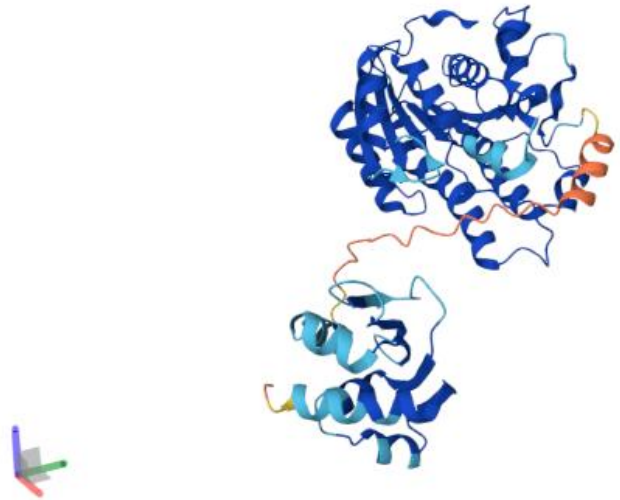

AlphaFold Protein Structure Database
Home
About
FAQs
Downloads
API
Search for protein, gene, UniProt accession or organism or sequence search
Search
Examples:
MENFQKVEKIGEGTYGV...
Free fatty acid receptor 2
At1g58602
Q5VSL9
E. coli
See search help
Go to online course
Hydroxysteroid dehydrogenase-like protei...
Download files
AF-Q6YN16-F1-v6 • Google Deepmind dataset •
Tell us what you think of the new look
Share your feedback
Summary and Model Confidence
Domains
Annotations
Similar Proteins
Protein
Hydroxysteroid dehydrogenase-like protein 2
Gene
HSDL2
Source organism
Homo sapiens (Human) go to search
UniProt
Q6YN16 go to UniProt
Biological function
Has apparently no steroid dehydrogenase activity (PubMed:19703561). Controls bile acid (BA) and lipid metabolism in response to nutritional ... Show more go to UniProt
Experimental structures
1 structure in PDB for Q6YN16 go to PDBe-KB
Average pLDDT
90.12 (Very High)
pLDDT distribution
73.7% Very high
17.7% High
1.4% Low
7.2% Very low

Alphafold

Hydroxysteroid dehydrogenase-like protein 2, Sequence length 418 Copy sequence

```
MLPNTGRLAGCTVFITGASRGIGKAIALKAAKDGANIVIAAKTAQPHPKLLGTIYTAAEEIEAVGGKALPCIVDVRDEQQISAAVEKAIKKFGGIDILVNNASAISLTNTLDTPTKRLDLMMNVNTRGTYLASKACIPYLKKSKVAHILNISPPL
NLNPVWFKQHCAYTIAKYGMSMYVLGMAEEFKGEIAVNALWPKTAIHTAAMDMLGGPGIESQCRKVDIIADAAYSIFQKPKSFTGNFVIDENILKEEGIENFDVYAIKPGHPLQPDFFLDEYPEAVSKKVESTGAVPEFKEEKLQLQPKPRSGAV
EETFRIVKDSLSDDVVKATQAIYLFELSGEDGGTWFLDLKSKGGNVGYGEPSDQADVVMSMTTDDFVKMFSGKLKPTMAFMSGKLKIKGNMALAIKLEKLMNQMNARL
```





Model Confidence

- Very high (pLDDT > 90)
- High (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

pLDDT is a per-residue measure of local confidence. Learn more...

Domains (2)

Annotations

Barva ukazuje spolehlivost: pLDDT

Predicted Aligned Error (PAE)

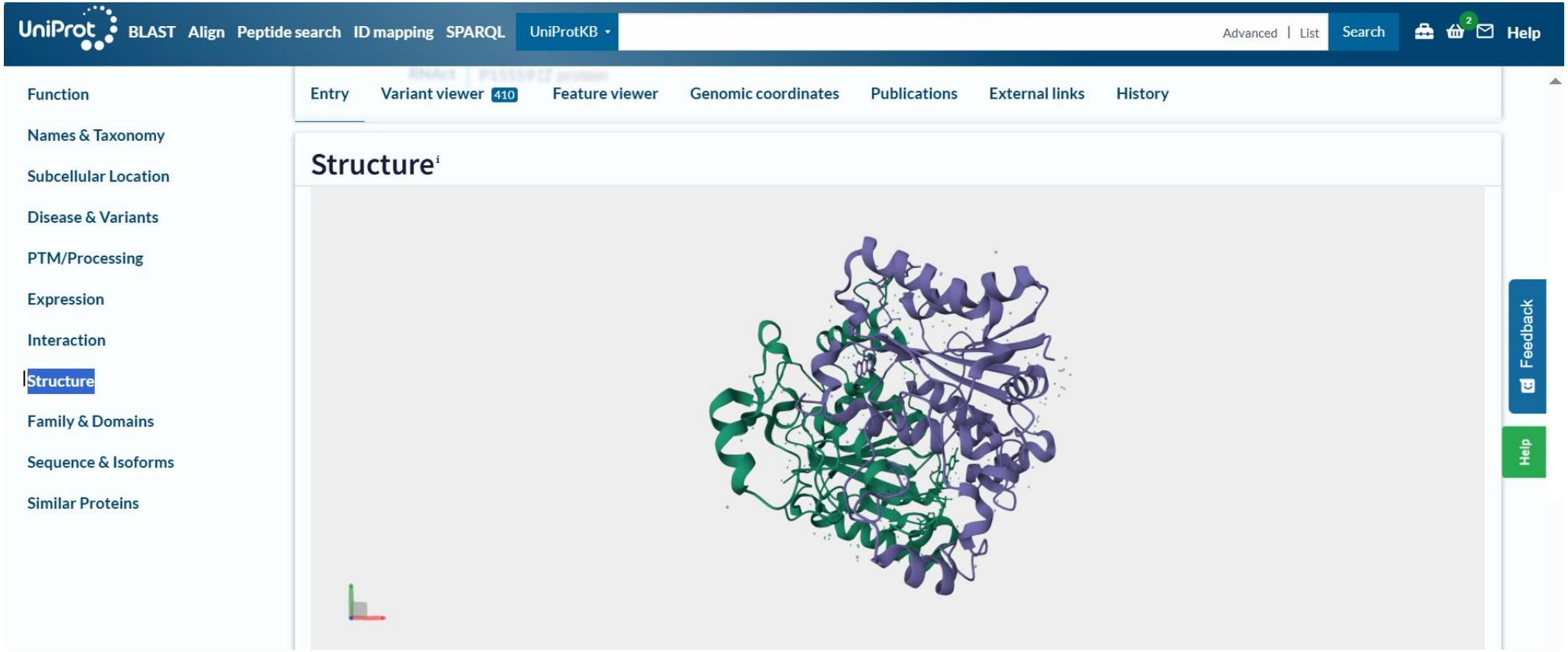
PAE measures the confidence in the relative position of two residues - **see Help section below** for more information.

Audit history

Model creation date: 1 Aug 2025

Sequence version date: 5 Jul 2004

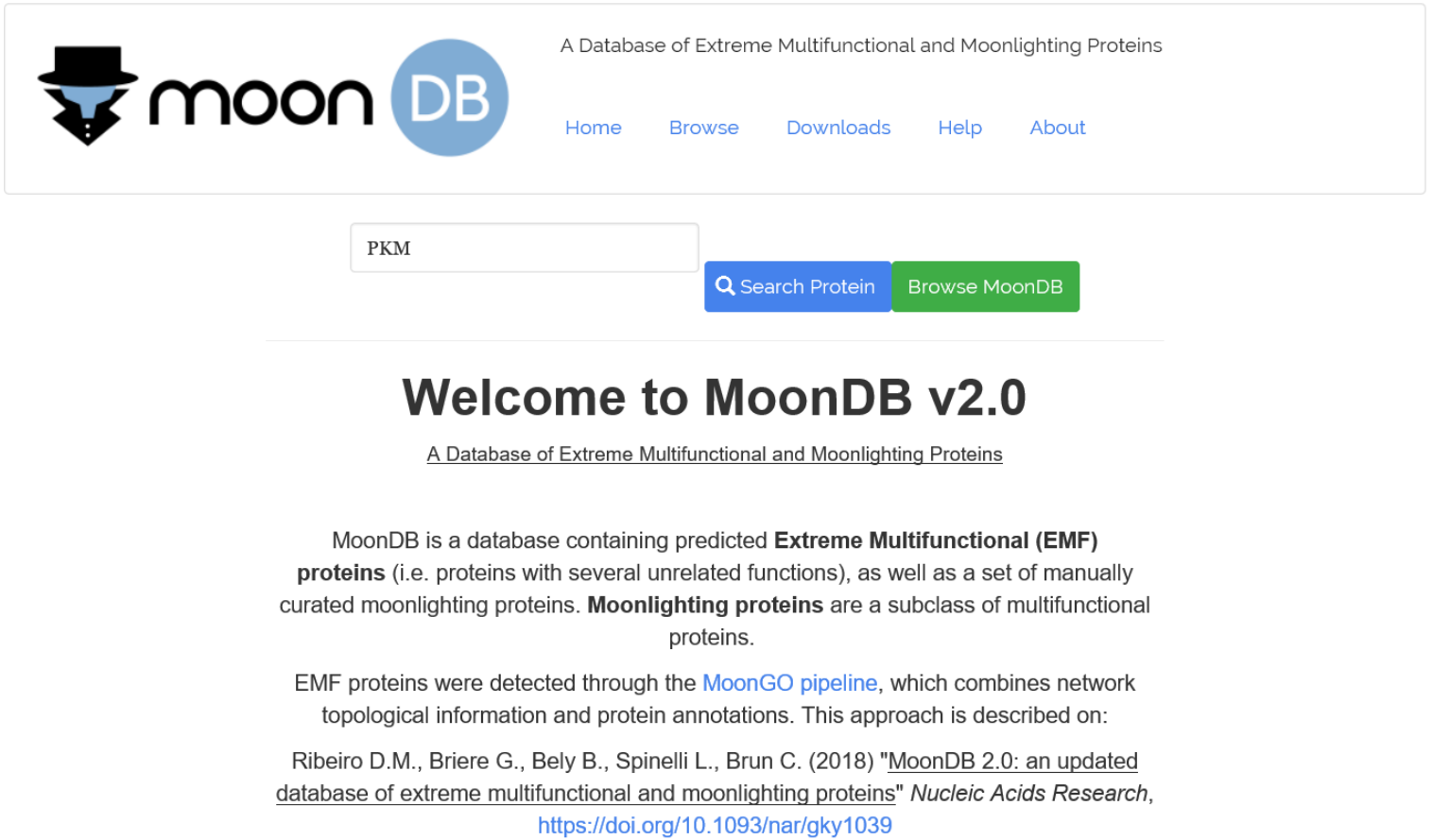
Cesta ke struktuře přes Uniprot:



Podívejte se po 3D struktuře:

- Vyzkoušejte si PDB databázi
- Vyzkoušejte Alphafold
- Podívejte se do Uniprotu

Multifunkční proteiny: moondb



Fosforylace proteinů:PhosphoSite

PhosphoSitePlus®

Powered by Cell Signaling Technology

Feedback | Contact | Login

Home Search Browse Tools Downloads About

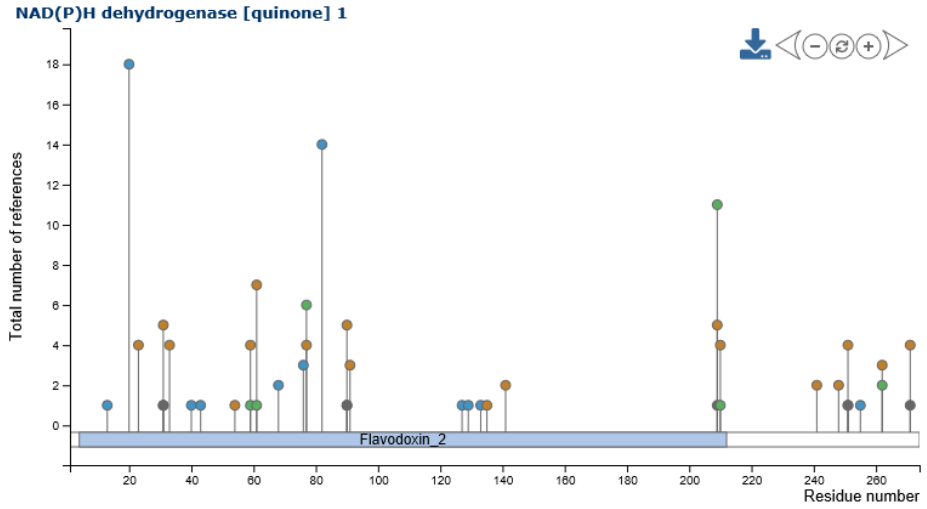
PhosphoSitePlus® provides comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, and more. The web use is free for everyone including commercial.

Protein or Substrate Search

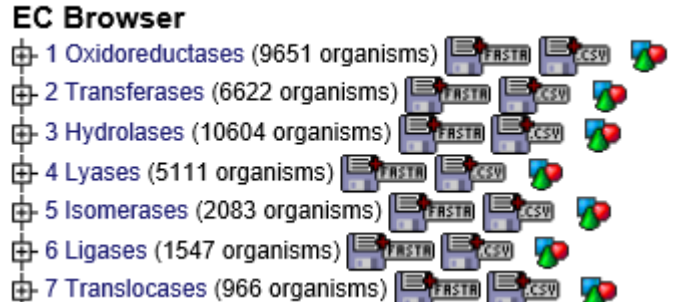
Protein Name | Search

Alternative Search Options

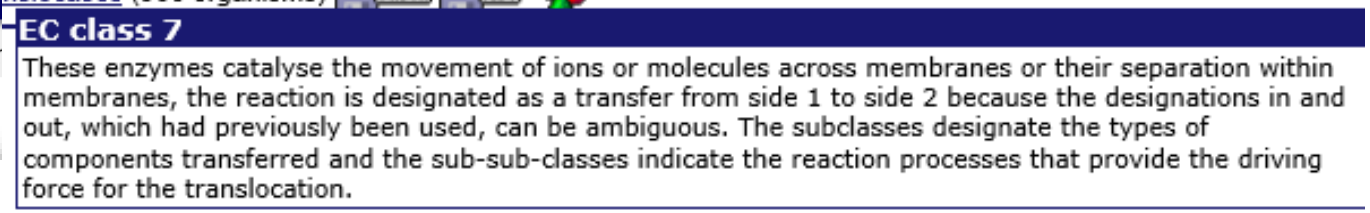
- Protein, Sequence, or Reference Search
- Site Search
- Comparative Site Search
- Browse MS2 Data by Disease
- Browse MS2 Data by Cell Line
- Browse MS2 Data by Tissue



Enzymová databáze: Brenda



A new class EC 7, Translocases, is



Rozdělení do tříd

- 1. OXIDOREDUKTASY – oxidačně-redukční děje
  - donor + akceptor → oxidovaný donor + redukovaný akceptor
  - **Systematický název**: donor : akceptor-oxidoreduktasa
- 2. TRANSFERASY – přenos funkčních skupin
  - donor_SK + akceptor → donor + akceptor_SK
  - **Systematický název** : donor : akceptor_skupinatransferasa
- 3. HYDROLASY – hydrolýza
  - A - B + H_2O → AOH + HB
  - **Systematický název**: substrát (skupina) hydrolasa
- 4. LYASY – eliminace skupin za vzniku dvojných vazeb
  - substrát 1 (+ substrát2) → produkt1 + produkt2 (malý)
  - **Systematický název**: substrát1 (:substrát 2)_ produkt2lyasa
- 5. ISOMERASY – izomerace
  - **Systematický název**: substráttyp
- 6. LIGASY – tvorba vazeb spojená s hydrolýzou ATP
  - substrát1 + substrát2 + A(G) TP → substrát1_substrát2 + ADP + Pi
  - substrát1 + substrát2 + ATP → substrát1_substrát2 + AMP + PPi
  - **Systematický název**: substrát1 :substrát2_ligasa (tvořící ADP/AMP)

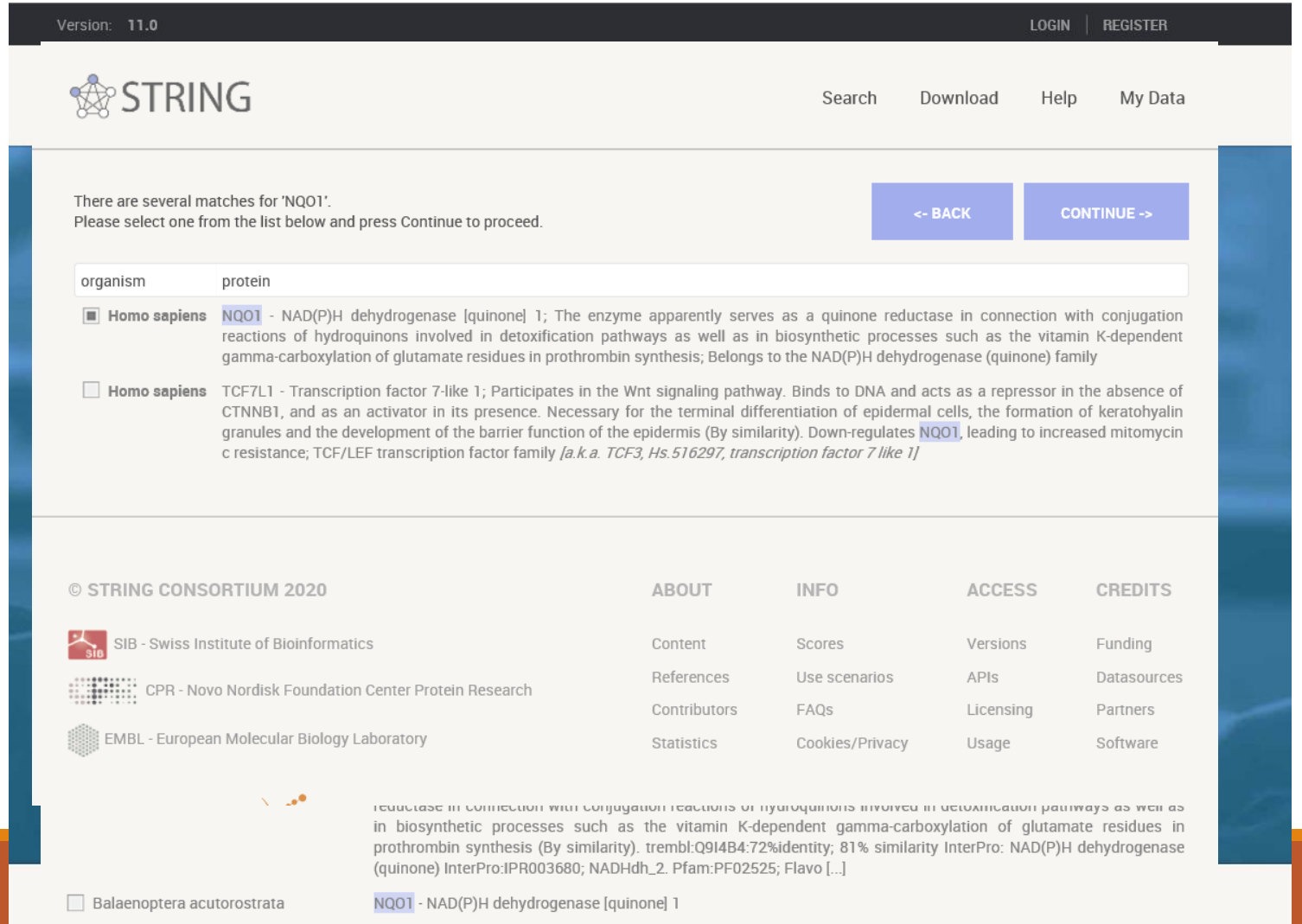
Příklad systematického názvosloví

- E.C. Enzyme Classification
- 1 Oxidoreduktasa
  - 1.1 působící na CH-OH skupiny donoru
    - 1.1.1 s NAD nebo NADP jako akceptorem
      - 1.1.1.1 alkohol dehydrogenasa ☺

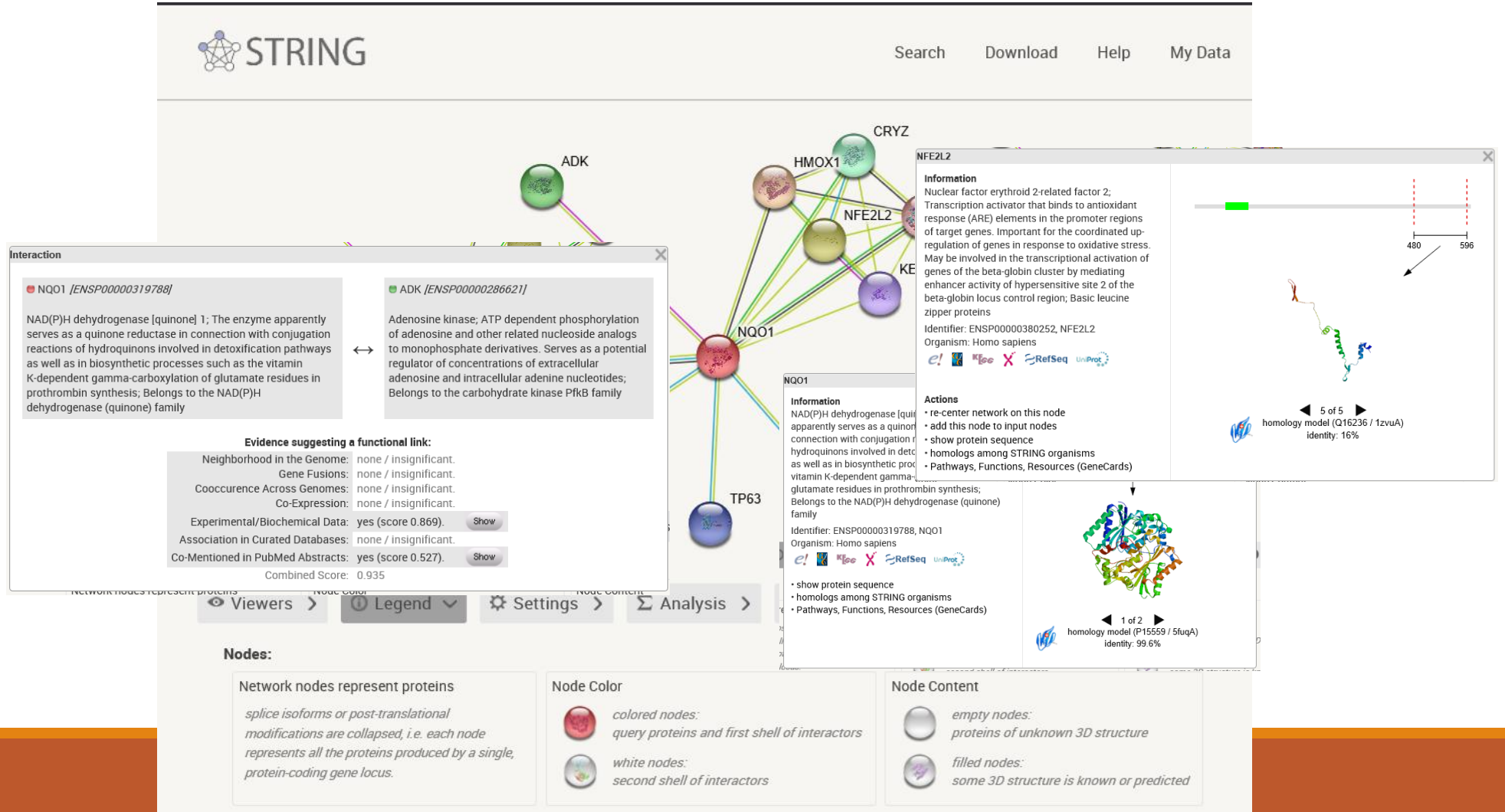
EC 7 – translokázy: Transport látek, nejčastěji přes biologické membrány;

některé translokázy vyžadují ATP; tato třída byla nově zavedena v r. 2018.

Interakce proteinů



Interakce proteinů



Podívejte se do vybraných databází

Je váš protein multifunkční?

Má typická fosforylační místa?

Má nějaké interakční partnery?

Je váš protein enzym? Jaké má enzymové číslo (E.C.x.x.x.x)? (DÚ)

DÚ4: porovnávání proteinů, 3D struktury

Pracujte s „vaším“ genem/proteinem/enzymem (př. NQO1)

1) Porovnejte „váš“ protein se „stejným“ proteinem z myši – párové porovnání. Jaká je identita těchto dvou sekvencí? Proběhlo porovnání celé délky sekvence?
2) Vytvořte mnohonásobné porovnání všech sekvencí z minulého úkolu (DÚ3) + vystřihněte „evoluční strom“
3) Má váš protein nějaké isoformy? Porovnejte je…
4) Byla určena 3D struktura vašeho proteinu? Vystřihněte jednu na ukázku.
5) Je váš protein enzym? Jaké má enzymové číslo (EC)?

Příklad řešení

DÚ4

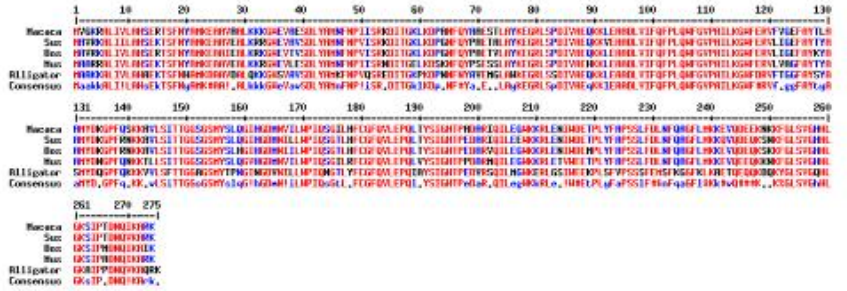
1)

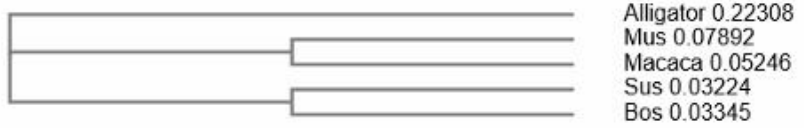
```
>>sp|Q64669|NQO1_MOUSE NAD(P)H dehydrogenase [quinone] 1  (274 aa)
 Waterman-Eggert score: 1626;  421.3 bits; E(1) <  1.1e-122
 86.5% identity (97.8% similar) in 274 aa overlap (1-274:1-274)

                 10        20        30        40        50        60
sp|P15 MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
       :.:::::::::::.:::::::::::. :::::.::::.::::::::::::::::.:::::.:
sp|Q64 MAARRALIVLAHSERTSFNYAMKEAAVEALKKRGWEVLESDLYAMNFNPIISRNDITGEL
                 10        20        30        40        50        60

                 70        80        90       100       110       120
sp|P15 KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
       ::  :::::.:: :::::::.::::::::.:::::::::::::::::::::::::::::::
sp|Q64 KDSKNFQYPSESSLAYKEGRLSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKGWFERV
                 70        80        90       100       110       120

                130       140       150       160       170       180
sp|P15 FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC
       ... ::::::::::.:::..::.:::::::::::::::::.:::::::::::::::::.::
sp|Q64 LVAGFAYTYAAMYDNGPFQNKRTLLSITTGGSGSMYSLQGVHGDMNVILWPIQSGILRFC
                130       140       150       160       170       180

                190       200       210       220       230       240
sp|P15 GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
       :::::::::.::::::: :::.:::::::::::..:.:::::::::::::::::::::::
sp|Q64 GFQVLEPQLVYSIGHTPPDARMQILEGWKKRLETVWEETPLYFAPSSLFDLNFQAGFLMK
                190       200       210       220       230       240

                250       260       270
sp|P15 KEVQDEEKNKKPGLSVGHHLGKSIPTDNQIKARK
       ::::.:.:..:::::::::::::::.:::::::
sp|Q64 KEVQEEQKNKKPGLSVGHHLGKSIPADNQIKARK
                250       260       270
```

i:86,5%

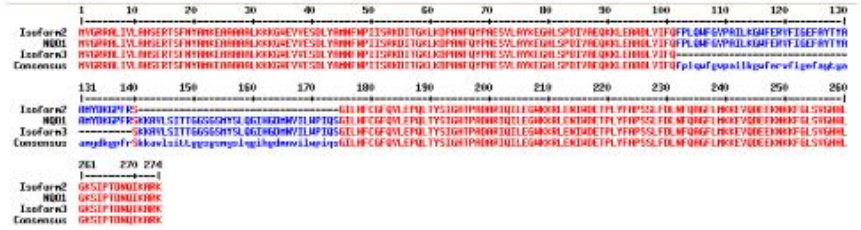
Porovní proběhlo v celém rozsahu obou proteinů.

2)

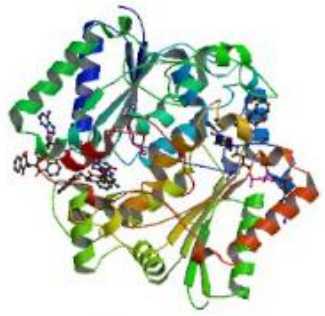




3)



4)



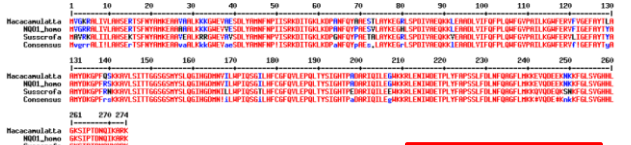
shrnutí I - Rešeršní projekt: NQO1

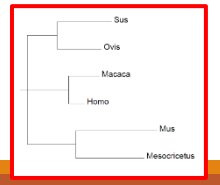
Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

Nalezněte podobné sekvence a porovnejte je na proteinové úrovni.

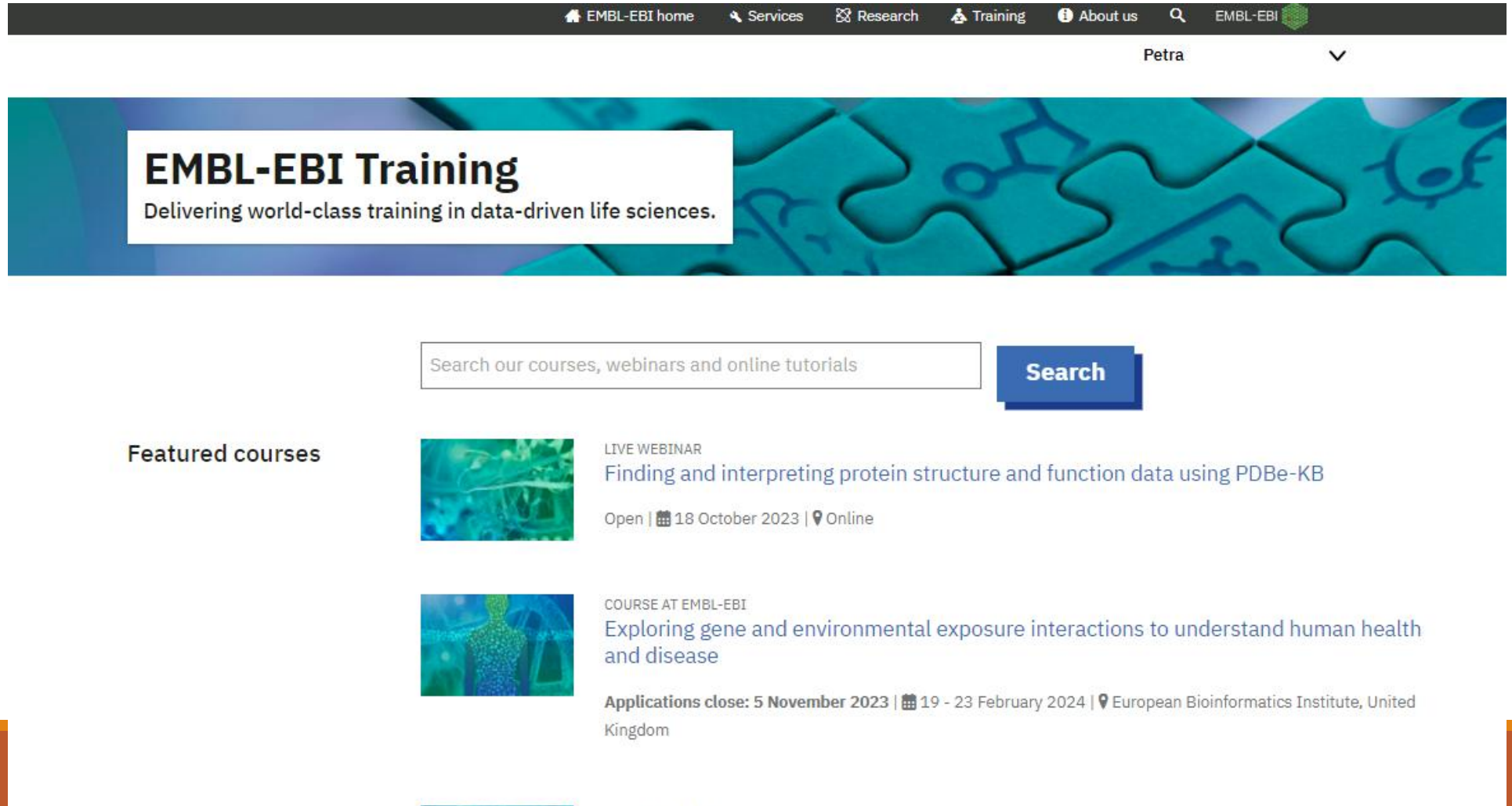
- NAD(P)H:chinonoxidoreduktasa / reference
- Protein: NP_000894, P15559 / 274 AMK; 30,8 kDa; trypsin štěpí 33x (59AMK nejdelší peptid)…
- **Flavodoxinová doména**

, N-terminální signální peptid
- **Rozpustný protein**
- **Podobné proteiny vyhledány, FASTA soubory uloženy**
- **Sekvence porovnány (evoluční strom)**
- **3D struktura**





EMBL-EBI Training (interaktivní kurzy)



WEBINAR

Exploring protein sequences and functional annotations with UniProt

Overview | How to attend | Competencies

UniProt provides the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

This webinar will guide you through ways in which students and researchers at all career stages may take a structured approach to access the data in the UniProt KnowledgeBase (UniProtKB) section of the UniProt protein function database.

We will discuss the types of protein data that can be accessed, and using sequence analysis tools, how this data can be analysed to answer biological questions. We will demonstrate how to find the function of a protein, how the protein components of a pathway may be identified, features of proteins of interest can be compared, similar sections of sequence can be sought in other species and how whole data sets containing all the proteins expressed in an organism can be obtained. By accessing sequence, structural and functional data through UniProt, we want researchers to be part of a journey through biology and support further work and adventures along the way!

Who is this course for?

This webinar is suitable for students or early career researchers beginning to use bioinformatics resources in their studies/research who wish to learn more about UniProt. No prior knowledge of bioinformatics is required, but undergraduate level knowledge of biology would be useful.

This event is part of a webinar series - Molecular building blocks of life - showcasing the EMBL-EBI resources that provide data and tools for studying molecular building blocks of life at molecular level such as metabolites, DNA/RNA, proteins, and at systems level such as cells. Check out other webinars in this

Register

Mark as favourite

Mark as complete

01 November 2023

15:30 - 16:30 (GMT)

Online

Free

First come, first served
1000 places

Contact
Ajay Mishra

Organisers

Ajay Mishra
EMBL-EBI

Speakers

Paul Denny
EMBL-EBI

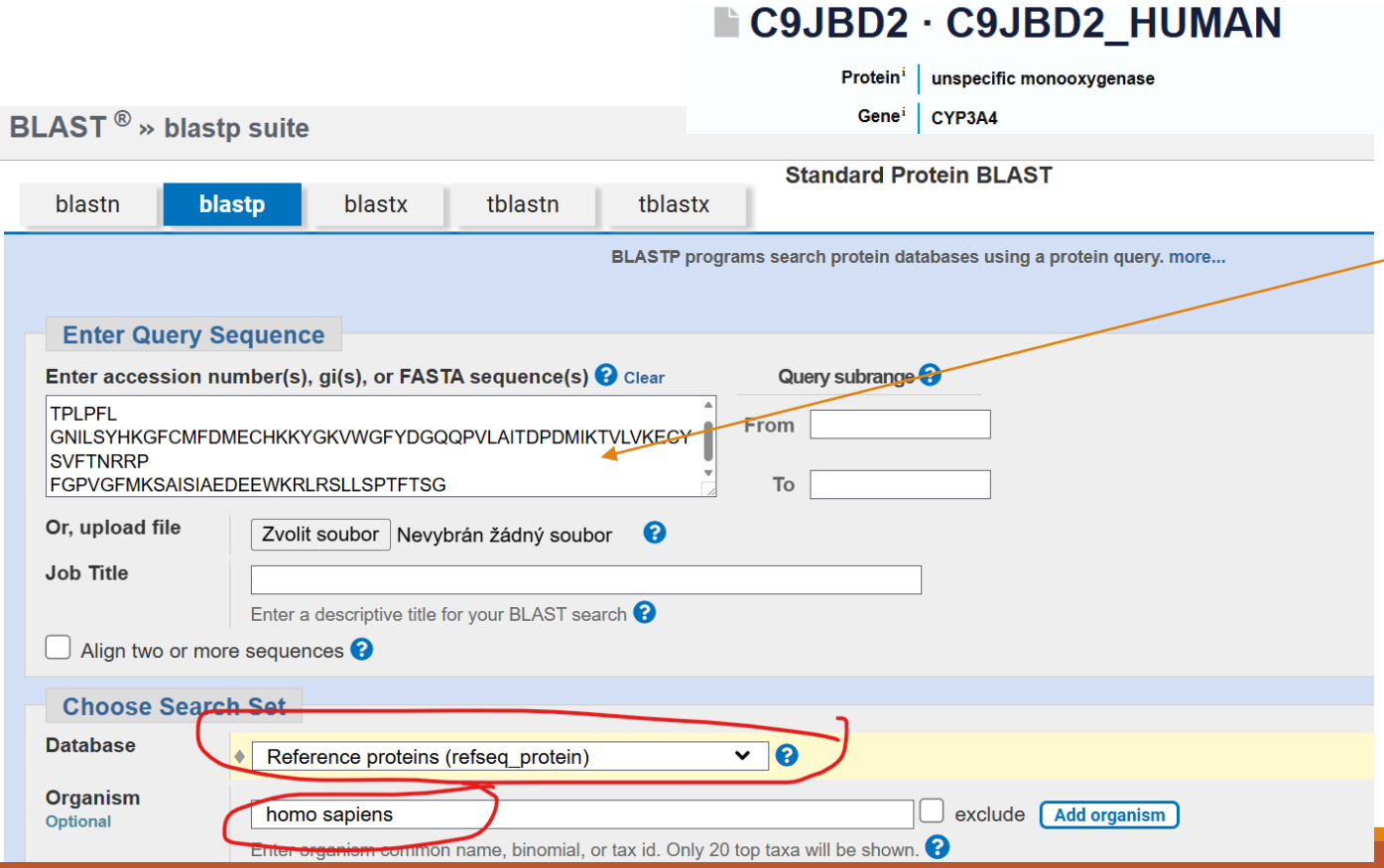
Procvičování (proteinová část)

Zjistěte kolik existuje lidských homologů proteinu CYP3A4.

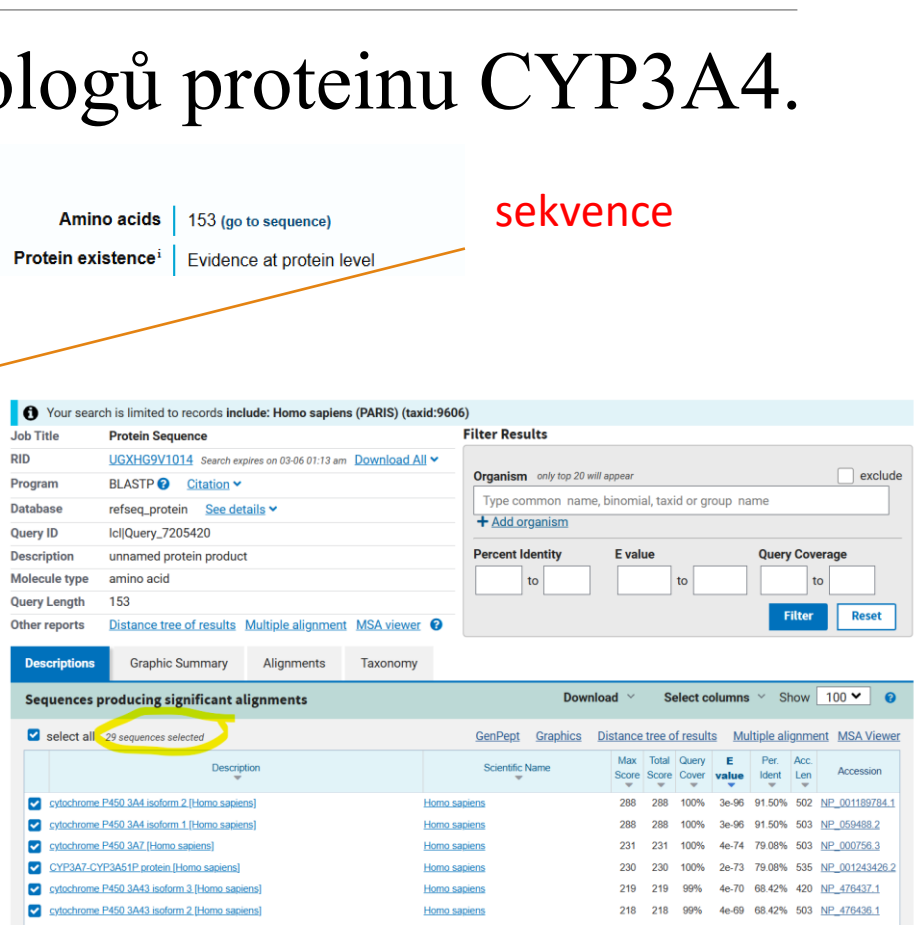
Řešení:

Procvičování (proteinová část)

Zjistěte kolik existuje lidských homologů proteinu CYP3A4.

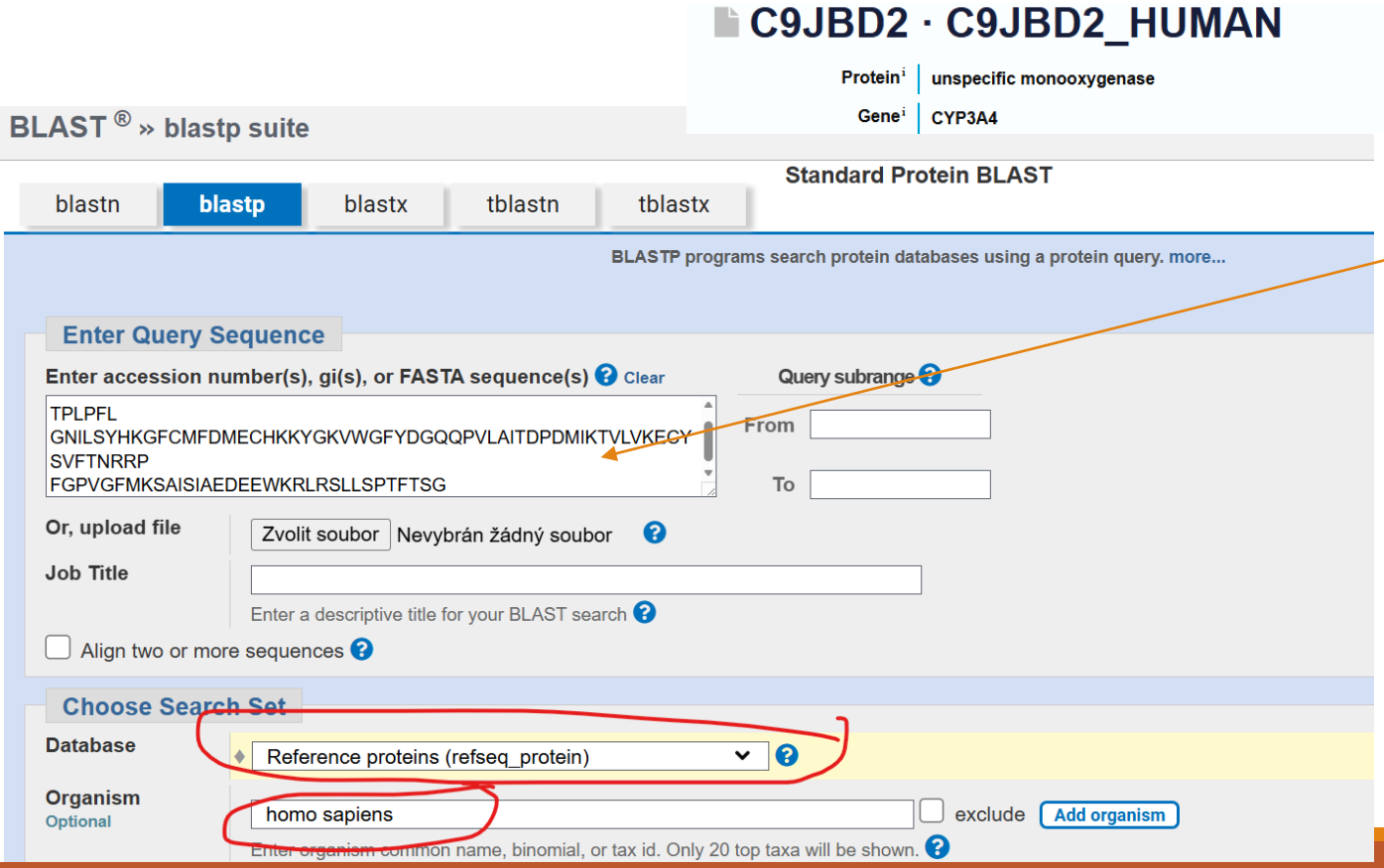


sekvence

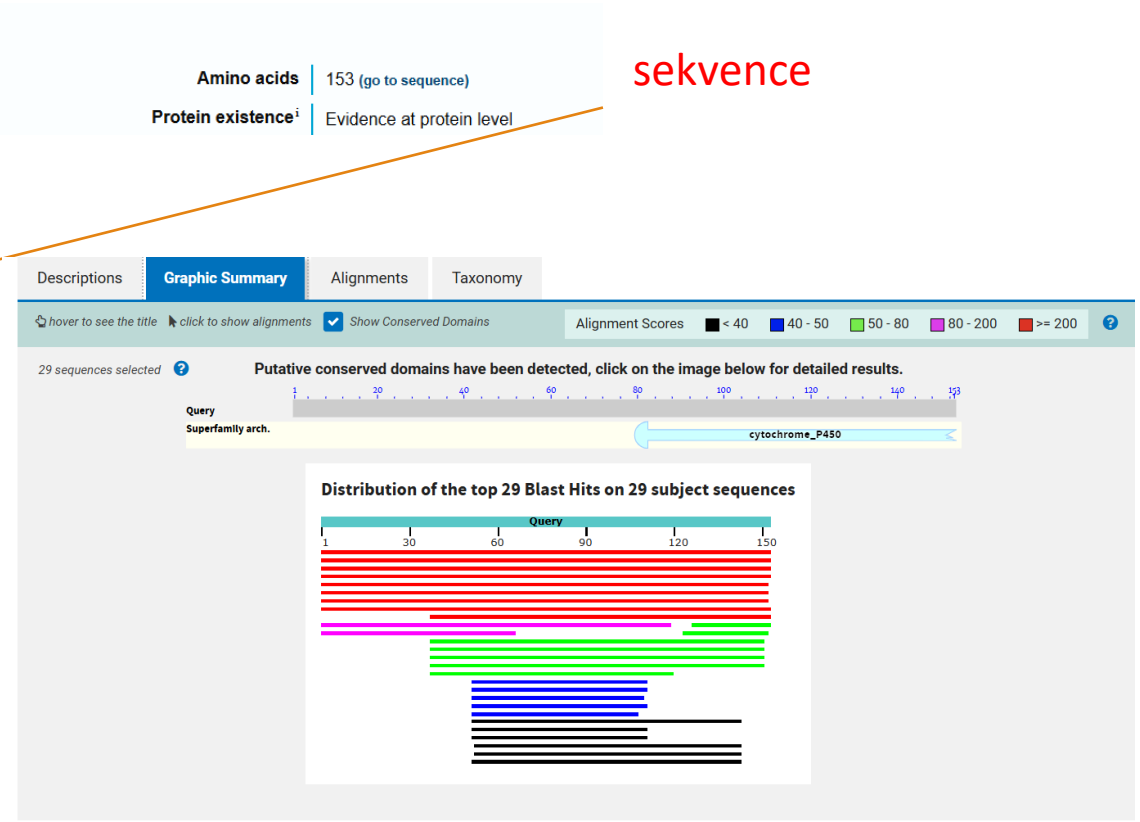


Procvičování (proteinová část)

Zjistěte kolik existuje lidských homologů proteinu CYP3A4.



sekvence



Procvičování (proteinová část)

Pracujte s následující sekvencí peptidu (Moodle-procvičování)

```
VTNLFILNLAISDLLVGIFCMPITLLDNIIAGWPFGNTMCKISGLVQGISVAASVFTLVA
IAVDRFQCVVYPFKPKLTIKTAFVIIMIIWVLAITIMSPSAVMLHVQEEKYYRVRLNSQN
KTSPVYWCREDWPNQEMRKIYTTVLFANIYLAPLSLIVIMYGRIGISLFRAAVPHTGRKN
QEQWHVVSRKKQKIIKMLLIVALLFILSWLPLWTLMMLSDYADLSPNELQIINIYIYPFA
HWLAFGNSSVNPIIYGFFNENFRRGFQEAFQLQLCQKRAKPMEAYALKAKSHVLINTSNQ
```

1) Identifikujte příslušný protein, zapište přístupový kód referenční sekvence.

2) Jaká je molekulová hmotnost tohoto peptidu? Jaká je molekulová hmotnost celého identifikovaného proteinu?

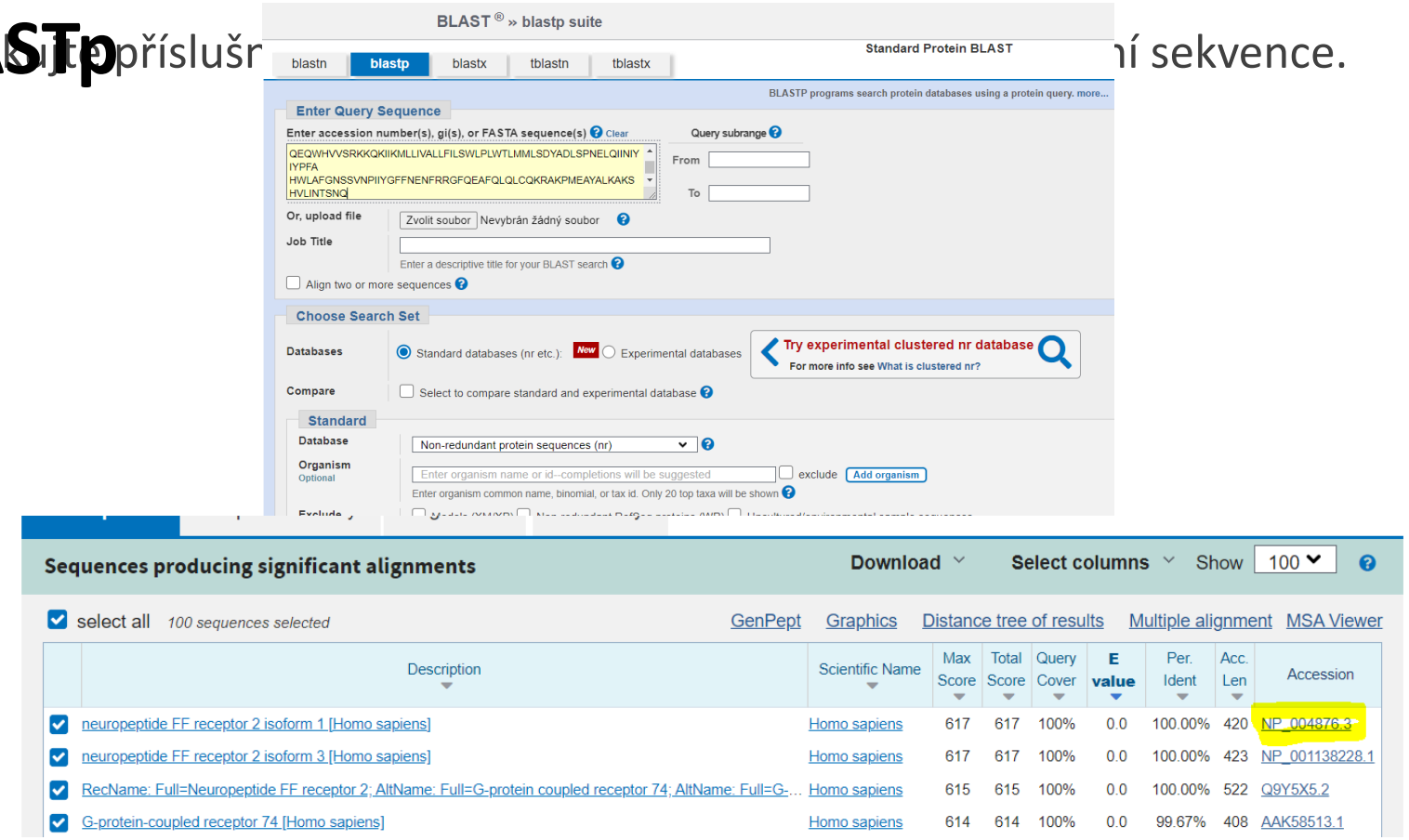
3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky?

4) Kde je tento protein v buňce lokalizován?

řešení

→BLASTp

1) identifikujte příslušn ní sekvence.

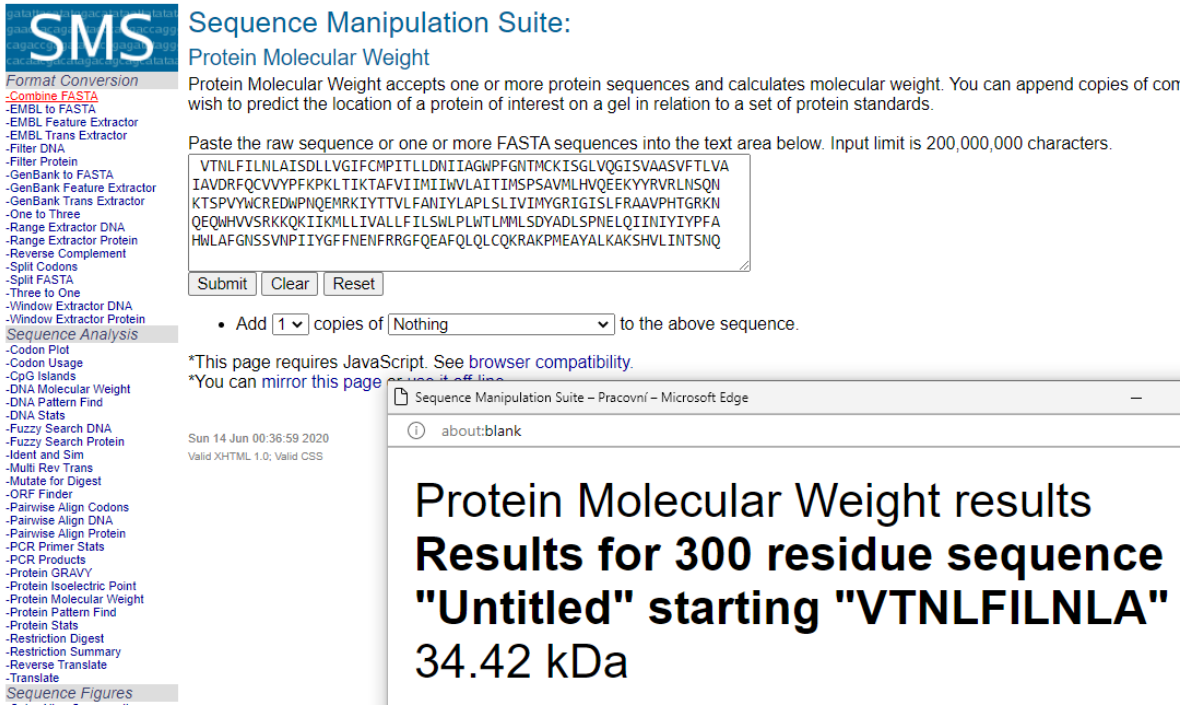


řešení

2) Jaká je molekulová hmotnost tohoto peptidu?

→ SMS Suite: Protein Molecular Weight

2) 34,42kDa

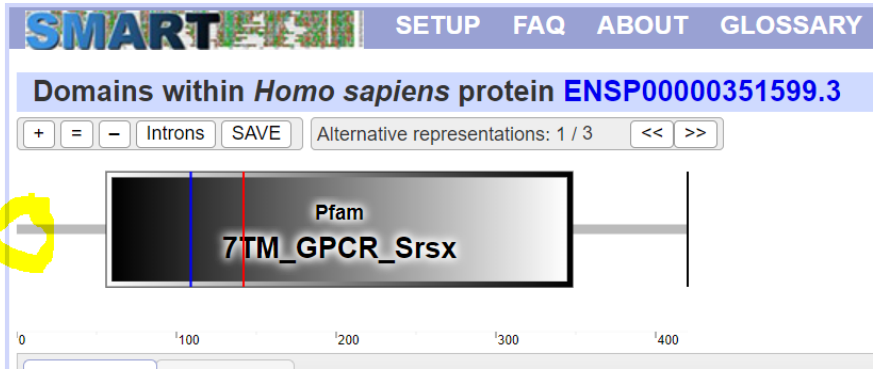


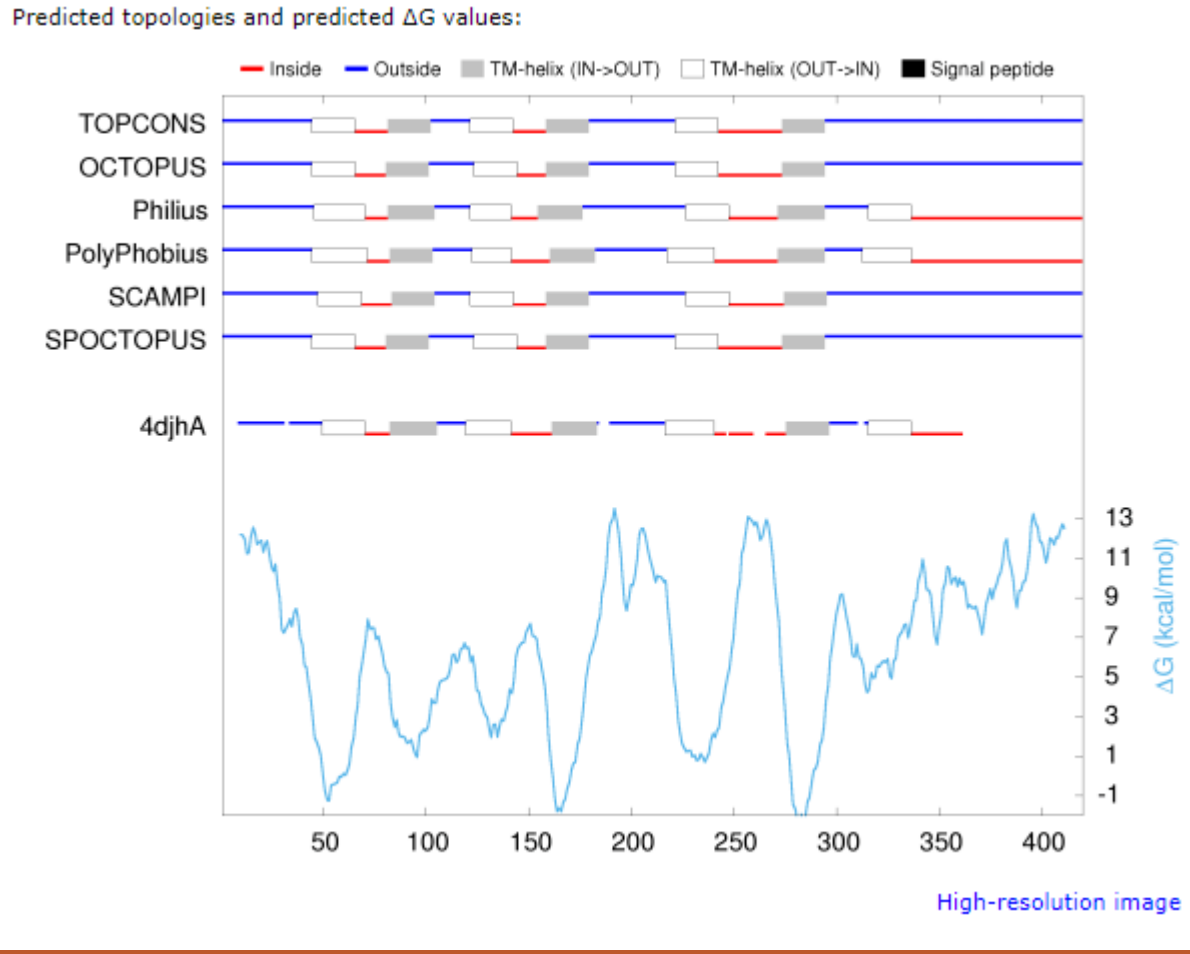
řešení

3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky?

→ různé programy (SignalP, TMHMM, topcons…)

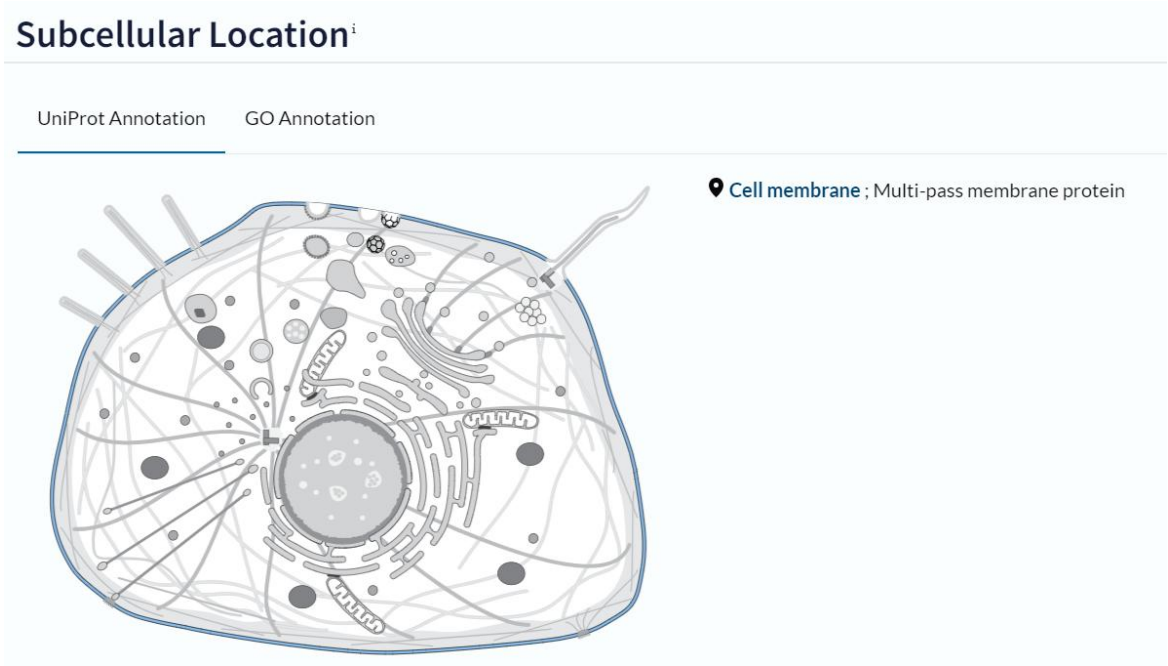
3) ne, ano (6xTM)





řešení

4) Kde je tento protein v buňce lokalizován?

→ UNIPROT (Q9Y5X5 · NPFF2_HUMAN)

4) Na membráně buňky