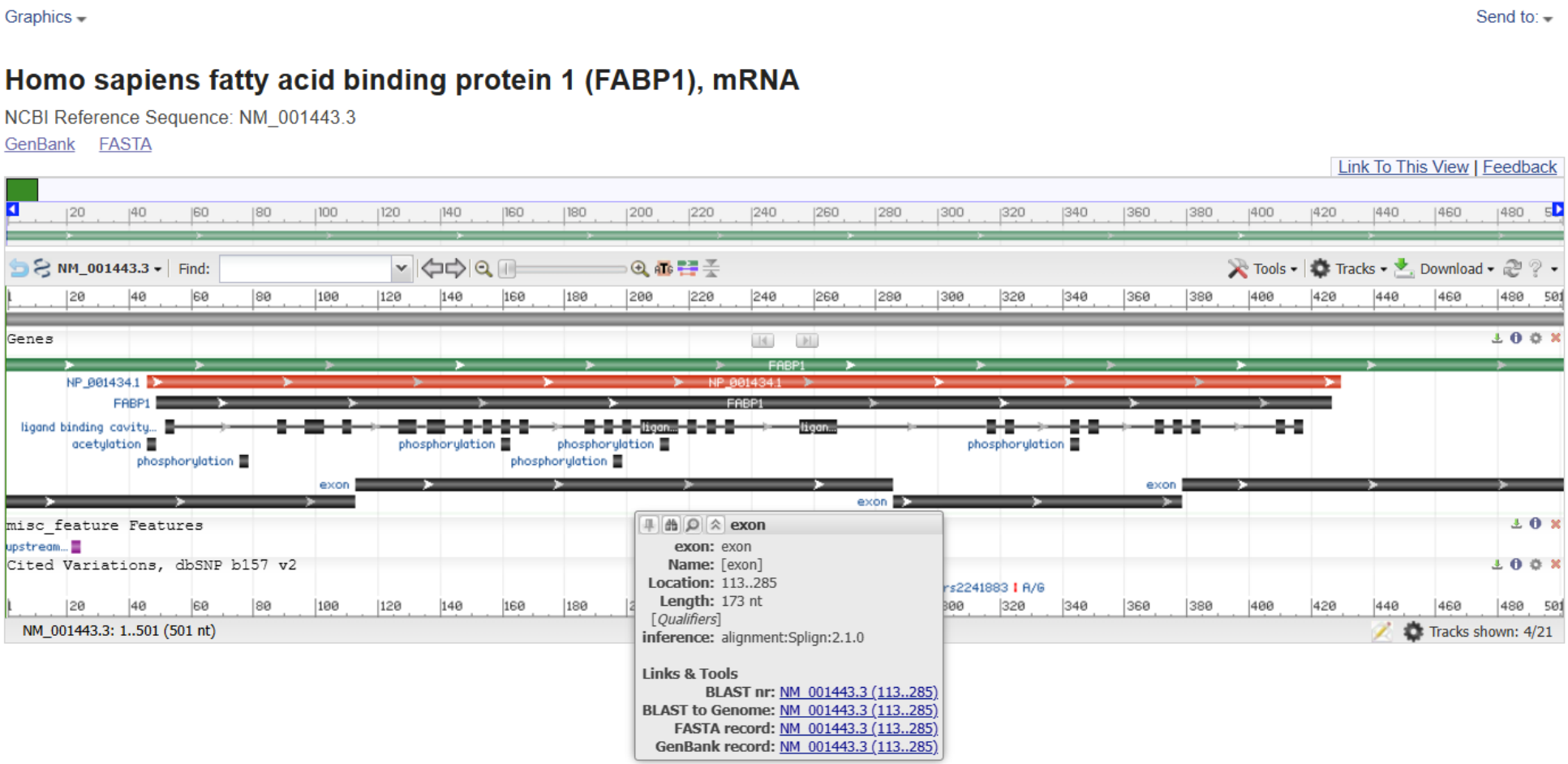
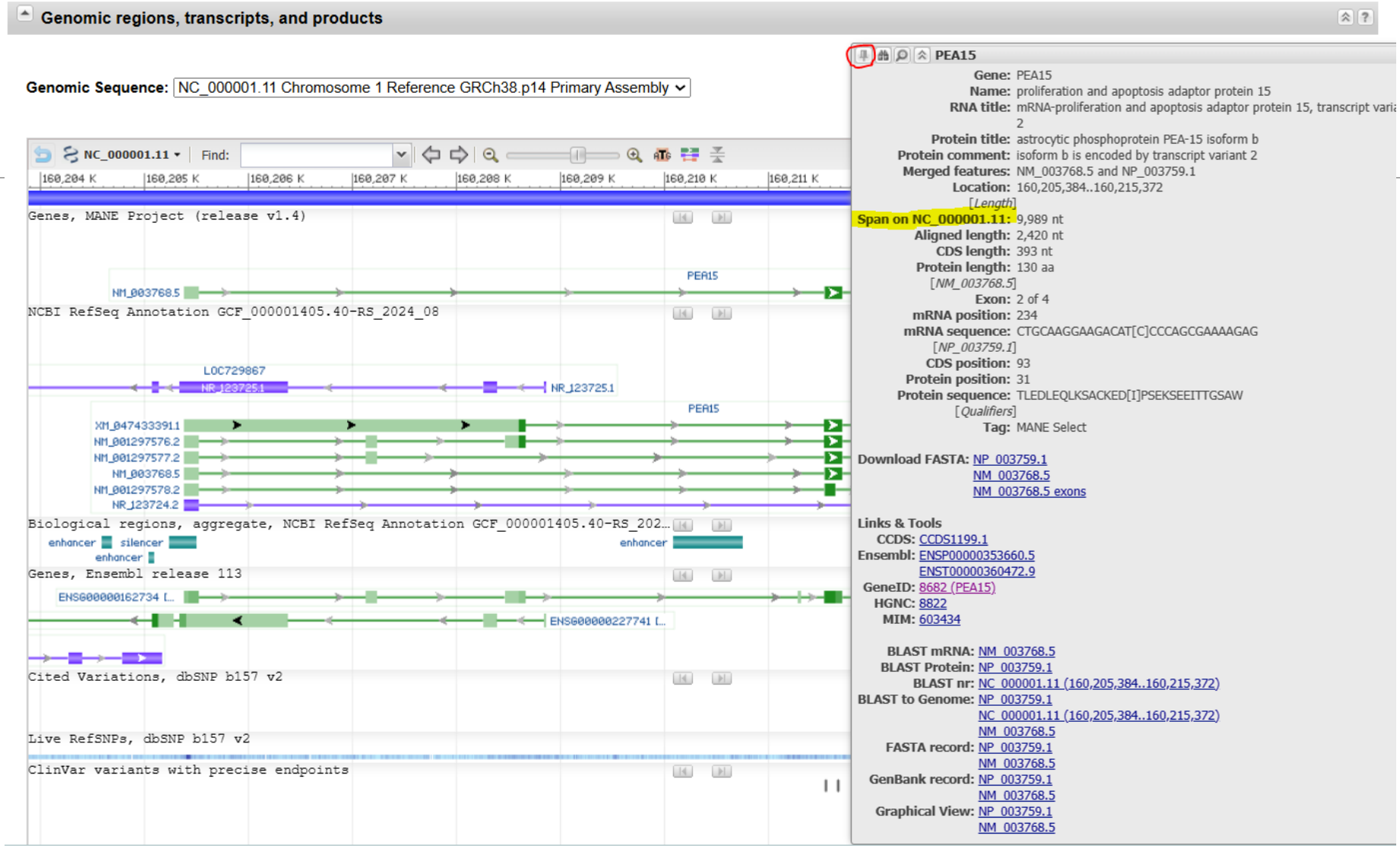
délka exonu:


Graphics
Send to:
Homo sapiens fatty acid binding protein 1 (FABP1), mRNA
NCBI Reference Sequence: NM_001443.3
GenBank FASTA
Link To This View | Feedback
NM_001443.3
Find:
Tools
Tracks
Download
Genes
FABP1
NP_001434.1
ligand binding cavity...
acetylation
phosphorylation
exon
misc_feature Features
Cited Variations, dbSNP b157 v2
NM_001443.3: 1..501 (501 nt)
Tracks shown: 4/21
exon
exon: exon
Name: [exon]
Location: 113..285
Length: 173 nt
[Qualifiers]
inference: alignment:Splign:2.1.0
Links & Tools
BLAST nr: NM_001443.3 (113..285)
BLAST to Genome: NM_001443.3 (113..285)
FASTA record: NM_001443.3 (113..285)
GenBank record: NM_001443.3 (113..285)


Genomic regions, transcripts, and products
Genomic Sequence: NC_000001.11 Chromosome 1 Reference GRCh38.p14 Primary Assembly
NC_000001.11
Find:
160,204 K
160,205 K
160,206 K
160,207 K
160,208 K
160,209 K
160,210 K
160,211 K
Genes, MANE Project (release v1.4)
PEA15
NM_003768.5
NCBI RefSeq Annotation GCF_000001405.40-RS_2024_08
LOC729867
NR_123725.1
NR_123725.1
PEA15
XM_047433391.1
NM_001297576.2
NM_001297577.2
NM_003768.5
NM_001297578.2
NR_123724.2
Biological regions, aggregate, NCBI RefSeq Annotation GCF_000001405.40-RS_202...
enhancer
silencer
enhancer
enhancer
Genes, Ensembl release 113
ENSG00000162734 [...
ENSG00000227741 [...
Cited Variations, dbSNP b157 v2
Live RefSNPs, dbSNP b157 v2
ClinVar variants with precise endpoints
PEA15
Gene: PEA15
Name: proliferation and apoptosis adaptor protein 15
RNA title: mRNA-proliferation and apoptosis adaptor protein 15, transcript vari
2
Protein title: astrocytic phosphoprotein PEA-15 isoform b
Protein comment: isoform b is encoded by transcript variant 2
Merged features: NM_003768.5 and NP_003759.1
Location: 160,205,384..160,215,372
[Length]
Span on NC_000001.11: 9,989 nt
Aligned length: 2,420 nt
CDS length: 393 nt
Protein length: 130 aa
[NM_003768.5]
Exon: 2 of 4
mRNA position: 234
mRNA sequence: CTGCAAGGAAGACAT[C]CCCAGCGAAAAGAG
[NP_003759.1]
CDS position: 93
Protein position: 31
Protein sequence: TLEDLEQLKSACKED[I]PSEKSEEITTGSAW
[Qualifiers]
Tag: MANE Select
Download FASTA: NP_003759.1
NM_003768.5
NM_003768.5 exons
Links & Tools
CCDS: CCDS1199.1
Ensembl: ENSP00000353660.5
ENST00000360472.9
GeneID: 8682 (PEA15)
HGNC: 8822
MIM: 603434
BLAST mRNA: NM_003768.5
BLAST Protein: NP_003759.1
BLAST nr: NC_000001.11 (160,205,384..160,215,372)
BLAST to Genome: NP_003759.1
NC_000001.11 (160,205,384..160,215,372)
NM_003768.5
FASTA record: NP_003759.1
NM_003768.5
GenBank record: NP_003759.1
NM_003768.5
Graphical View: NP_003759.1
NM_003768.5

Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2025/2026

Základy praktické bioinformatiky

Téma 6/10

Nukleotidová bioinformatika II

Cíle:

Student bude schopen přeložit sekvenci nukleotidovou do proteinové a nalézt smysluplný čtecí rámec. Bude schopen analyzovat sekvenci získanou sekvenováním, určit a porovnat neznámou sekvenci.

„Bioinformatika nukleových kyselin II“

Vyhledávání NK sekvencí

Analýza vlastností sekvencí-složení, reverse complement, identifikace restrikčních míst (Palindromy)

Práce s kódující DNA=práce s proteiny / překlad DNA sekvence-otvírání čtecího rámce

Návrh primerů pro PCR, rt-PCR

Předpověď sekundárních struktur

Porovnávání sekvencí, identifikace neznámé sekvence

(Vyhledání SNPs)

„čtení“ sekvenačních dat a spojování fragmentů

Vyhledávání hladin expresí jednotlivých genů

mikroRNA

Celé genomy

….

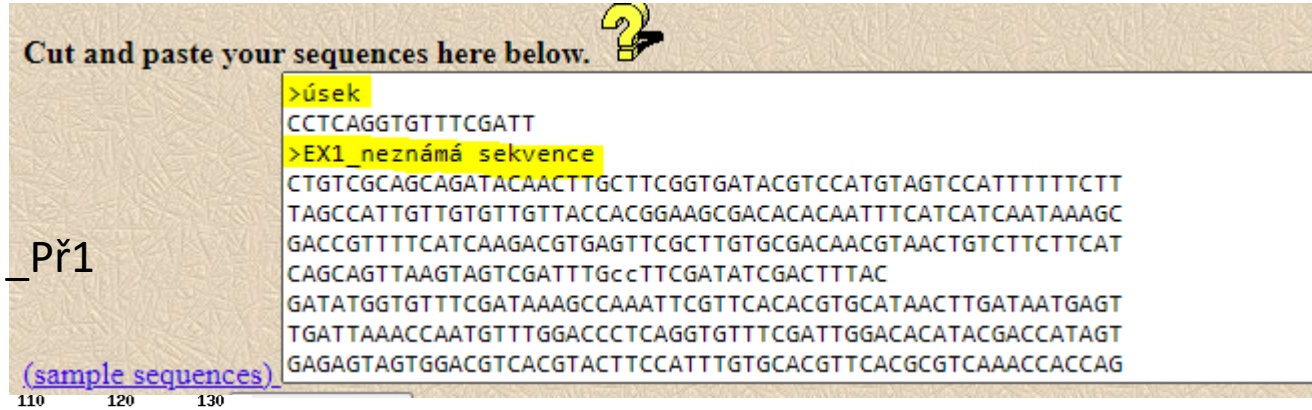
Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Např: multalin



- K vyhledání krátkých úseků (primerů, exonů apod.) _Př1

Default:BLOSUM62

```
                1       10        20        30        40        50        60        70        80        90       100       110       120       130
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
  nezn n        GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGTTTAGCAATGATAAAGAAAATAACCTTCTGAAAATTTGTATAGATCAGAAATAAAGTATTTTTTGTGGAAG
      f
Consensus       ..................................................................................................................................

                131     140       150       160       170       180       190       200       210       220       230       240       250       260
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
  nezn n        ACTATTTTTAAGTATTGAAGGTACTATTTCCTTTCTTGAATTCATATTGCAGATGTACGGTGTGGATTTATTGGTTTATCTCTGCAAACCTTAAAGTAGAAGATTGCAAGGGCCAGGTGTGGTGGATCAC
      f
Consensus       ..................................................................................................................................

                261     270       280       290       300       310       320       330       340       350       360       370       380       390
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
  nezn n        GCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTAGGCGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACGCCGTCTCTACTAAAAATACAAAAATTAGTCAGGCATGATGG
      f          CCTGTAATCCCAGCAC
Consensus       .CCTGTAATCCCAGCAC................................................................................................................

                391     400       410       420       430       440       450       460       470       480       490       500       510       520
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
  nezn n        CAGGCGCCTGTCATCCCAGCTACTAGGGAGGCTGAGGCAGGAGAATTGCTGGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGTAACAGAGCAAGACTGCTTC
      f
Consensus       ..................................................................................................................................

                521     530       540       550       560       570       580       590       600
                |--------+---------+---------+---------+---------+---------+---------+---------|
  nezn n        AAAAAAAAATTTTTTTTAATTAAAAAAAAAGTAGATGACTGCAGCAAAGAAGAGATTTATTAGCTCTTCCTTTCAAGGTA
      f
Consensus       ........................................................................
```

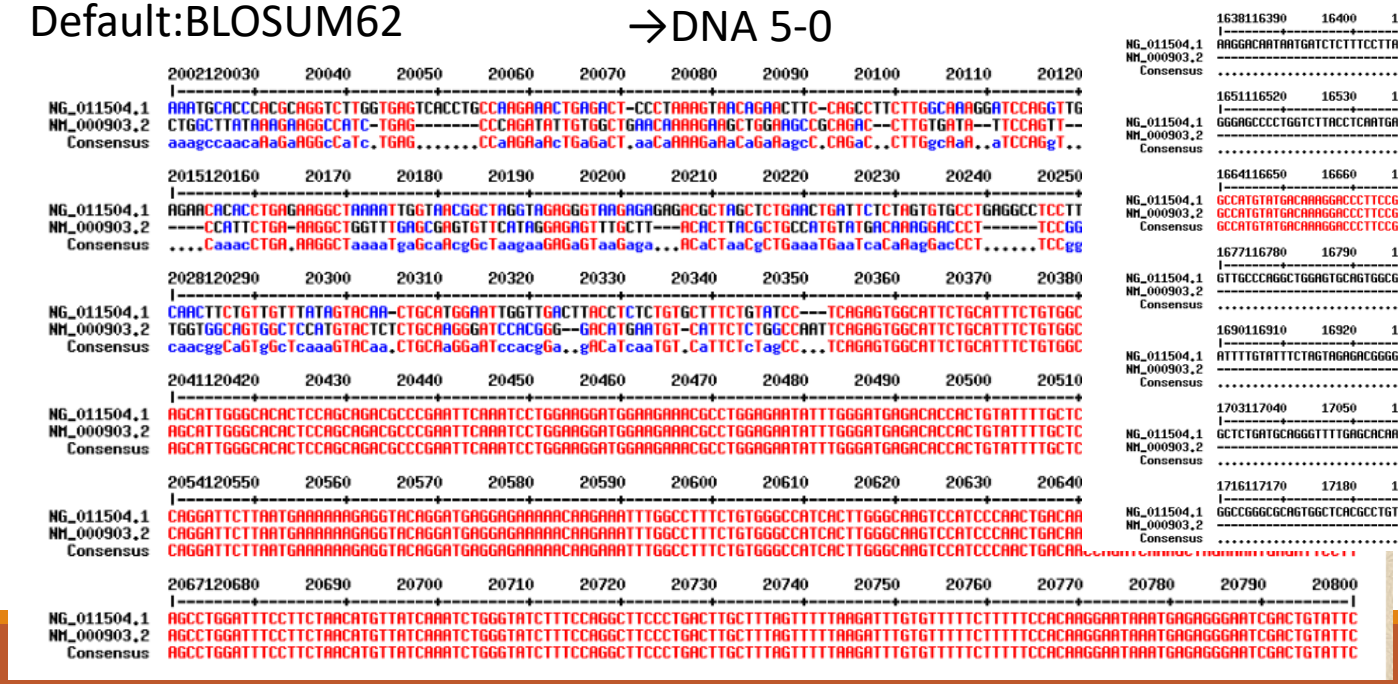
Porovnání nukleotidových sekvencí

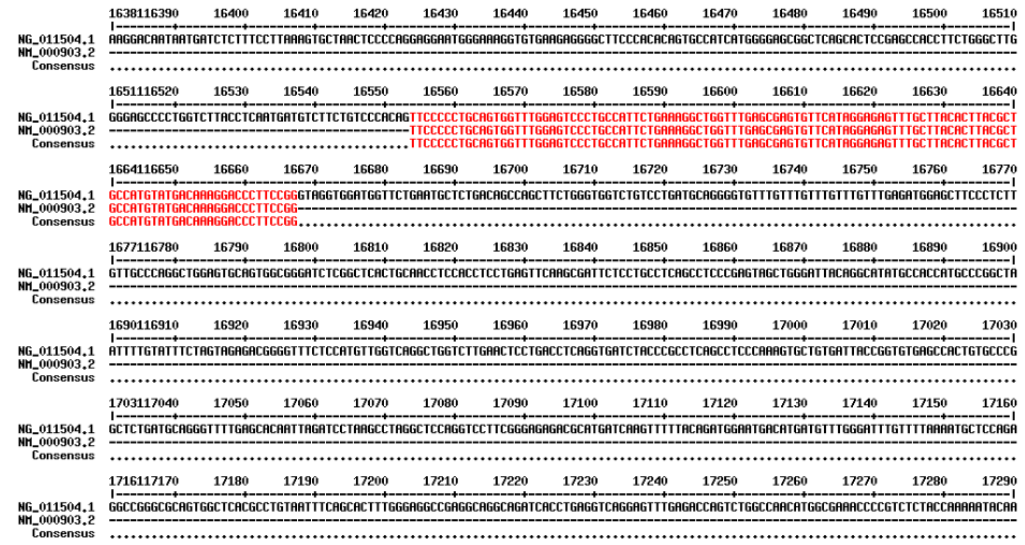
Velmi podobné proteinovému porovnávání

Např: multalin

- K porovnání sekvencí genomových a cDNA (mRNA)_Př2

Default:BLOSUM62 →DNA 5-0





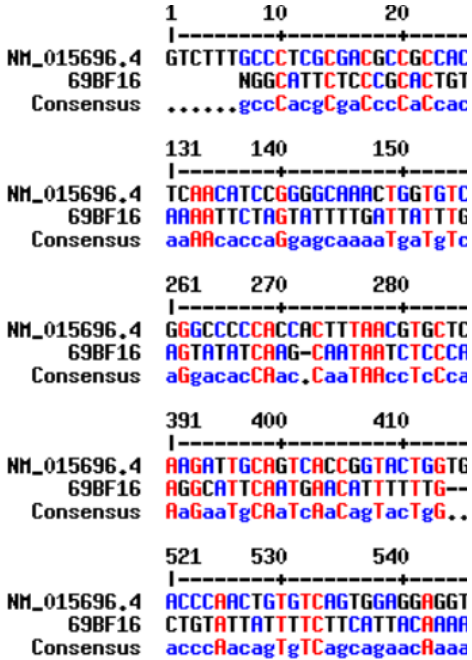
Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

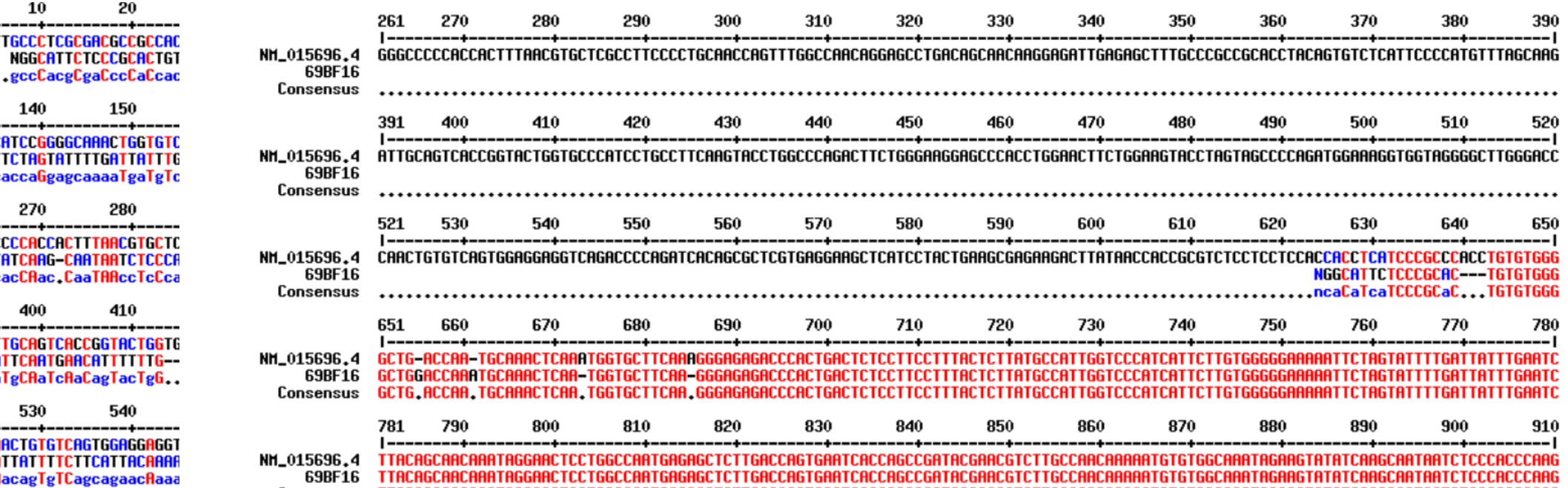
Např: multalin

- K „porovnání“ sekvencí identifikovaných a neznámých

Default:BLOSUM62



→DNA 5-0



Vyzkoušejte si porovnání

1) vyzkoušejte si porovnat sekvence v Př1 a Př2.

- porovnejte sekvence v multalinu s původním nastavením (blosum62 a s maticí DNA-5-0)

2)Porovnejte CDS a celou mRNA (DÚ)

3) Zkuste najít genomovou sekvenci odpovídající vašemu genu (referenční: **NG_**...)

(Nebo přes gene databázi výsek genu z NC – kratší než 30k nt)

Porovnejte ji s „mRNA“ v multalinu s původním nastavením (blosum62 a s maticí DNA-5-0)

4) Zkuste: pro gen **FABP1** získat mRNA (NM_...) a sekvenci celého genu (v databázi Gene, přes NC_... A konkrétní pozici)

Vždy je nutné „posoudit“ zda je to „očekávatelný“ výstup nebo je potřeba „něco“ pozměnit!

„čtení“ DNA = translace/překlad

Genetický kód založený na tripletech: tři „čtecí rámce“ = ORF (open reading frame)

1. DNA sekvence:

5´-**ATGGA**AGTATTTAAAGCGCCACCTATTTAA-3´

2. Rozdělená na triplety:

ATG GAA GTA TTT AAA GCG CCA CCT ATT TAA

A **TGG** AAG **TAT** TTA **AAG** CGC **CAC** CTA **TTT** AA

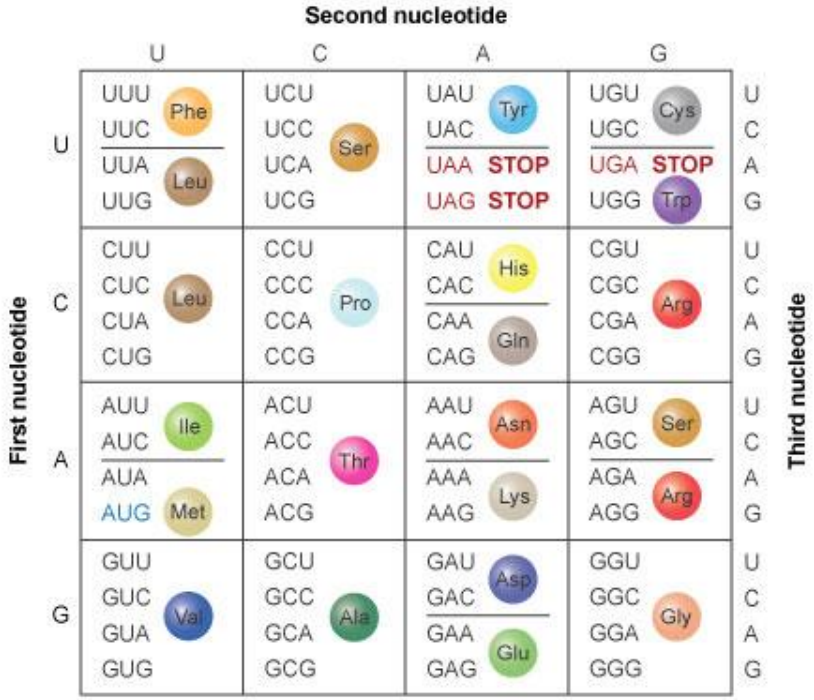
AT **GGA** AGT ATT TAA AGC GCC ACC TAT TTA A

3. Každý triplet přeložený do příslušné aminokyseliny:

M E V F K A P P I STOP(*)

W K Y L K R H L F

G R I * S A T Y L



„čtení“ DNA = translace/překlad

DNA sekvence psaná v orientaci 5' - 3', vznikne aminokyselinová sekvence v orientaci od **N-** k **C-** konci.

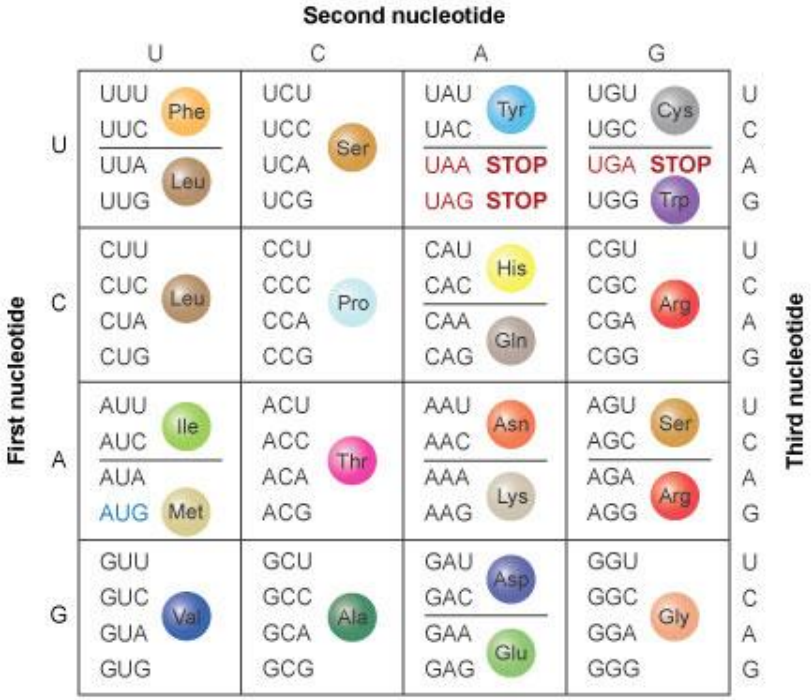
X nevíme který z dsDNA je kódující řetězec:

5´-ATGGAAGTATTTAAAGCGCCACCTATTTAA-3´
3´-TACCTTCATAAATTTCGCGGTGGATAAATT-5´

5´-TTAAATAGGTGGCGCTTTAAATACTTCCAT-3´

TTA AAT AGG TGG CGC TTT AAA TAC TTC CAT
T TAA ATA GGT GGC GCT TTA AAT ACT TCC AT
TT AAA TAG GTG GCG CTT TAA ATA CTT CCA T

L N R W R F K Y F H
* I G G A L N T S
K * V A L * I L P



„čtení“ DNA = translace/překlad

→ je třeba „číst v 6(!) čtecích rámcích = **ORF** (open reading frame)

5´- ATGGAAGTATTTAAAGCGCCACCTATTTAA-3´

ATG GAA GTA TTT AAA GCG CCA CCT ATT TAA

A **TGG** AAG **TAT** TTA **AAG** CGC **CAC** CTA **TT**T AA

AT **GGA** AGT ATT TAA AGC GCC ACC TAT TTA A

M E V F K A P P I STOP(*)

W K Y L K R H L F

G R I * S A T Y L

5´-TTAAATAGGTGGCGCTTTAAATACTTCCAT-3´

TTA AAT AGG TGG CGC TTT AAA TAC TTC CAT

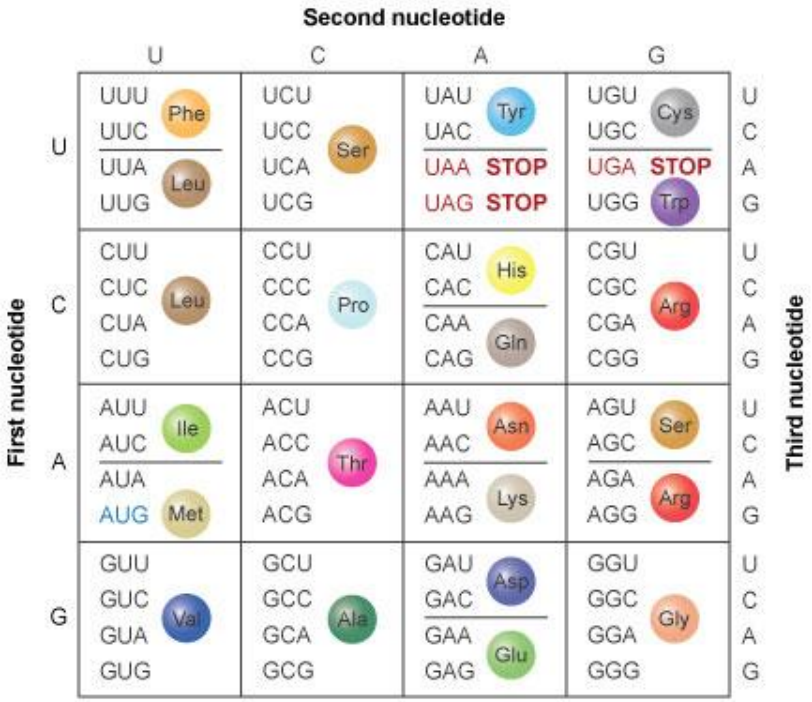
T TAA ATA GGT GGC GCT TTA AAT ACT TCC AT

TT AAA TAG GTG GCG CTT TAA ATA CTT CCA T

L N R W R F K Y F H

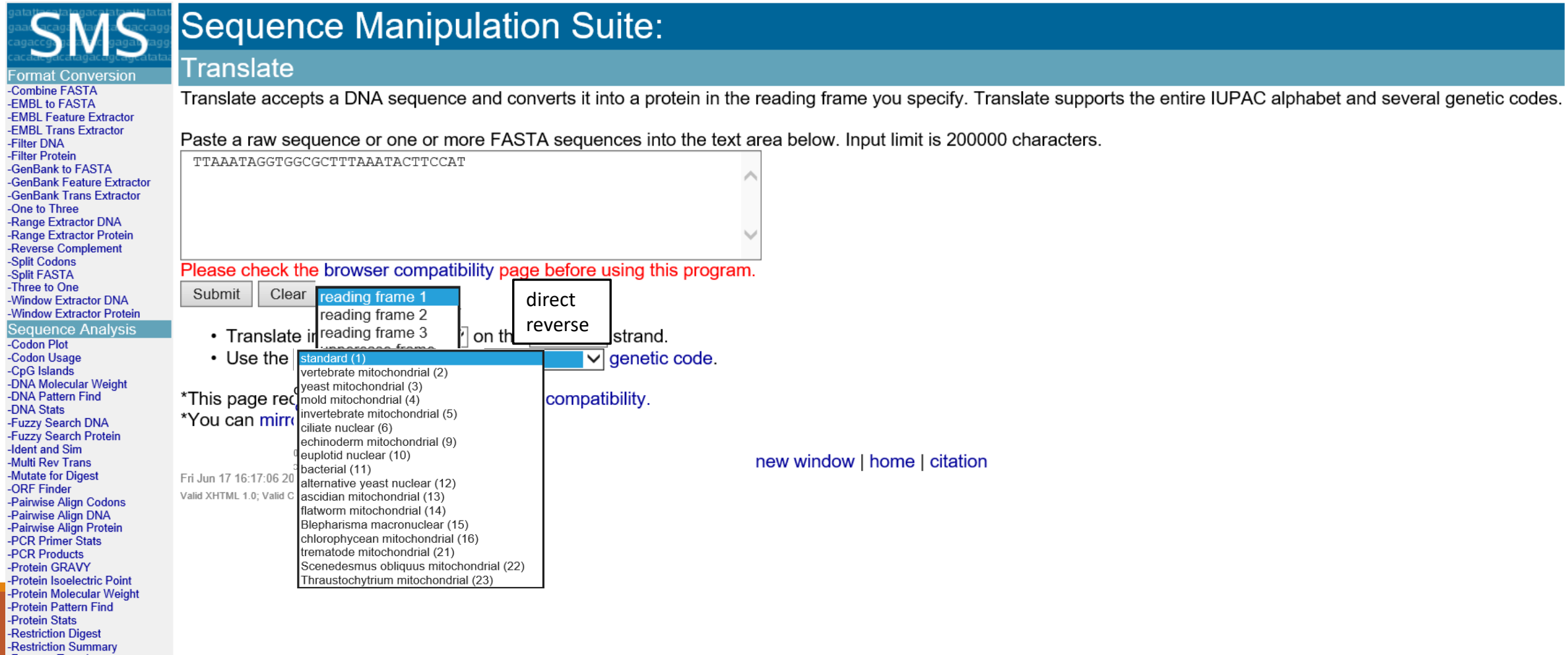
* I G G A L N T S

K * V A L * I L P



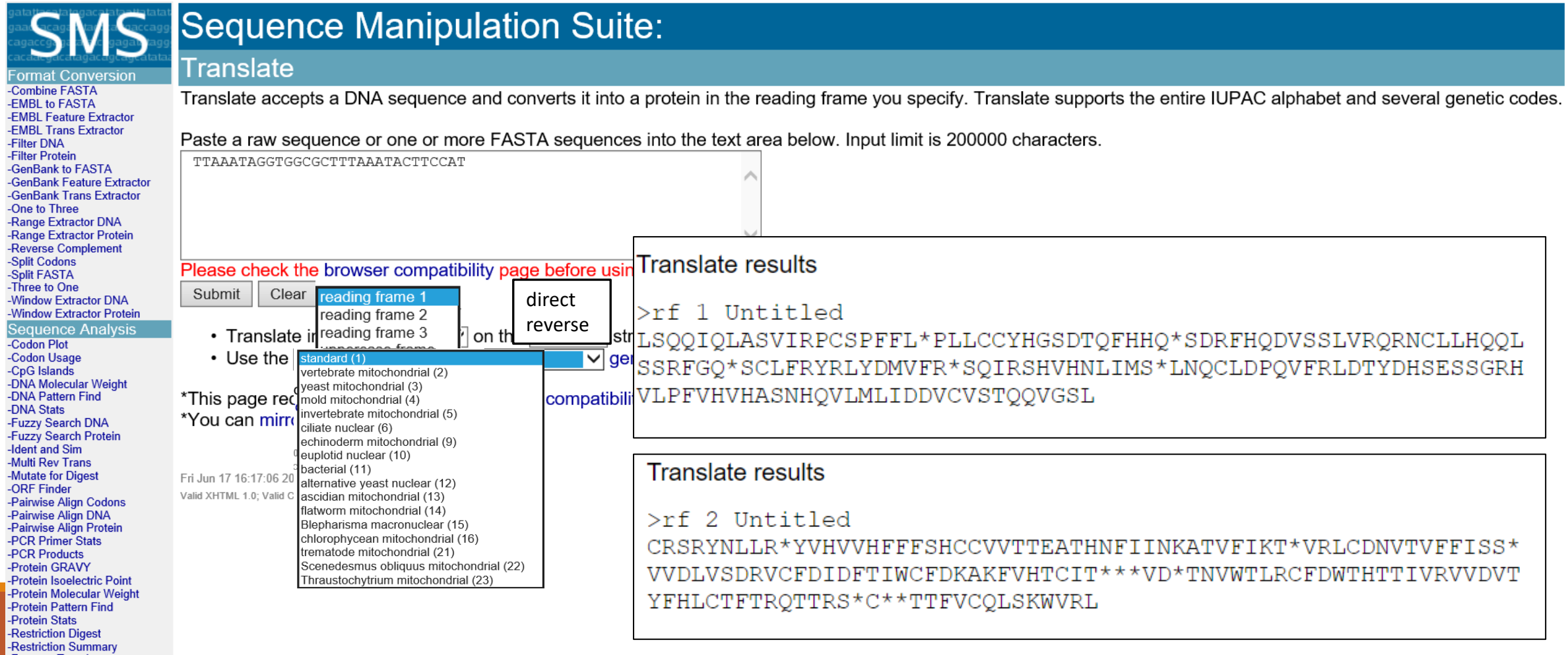
„čtení“ DNA = translace/překlad

SMS/Translate → vhodné známe-li kde je čtecí rámec



„čtení“ DNA = translace/překlad

SMS/Translate → vhodné známe-li kde je čtecí rámec



vyzkoušejte si

zkuste si v programu SMS/translate přeložit

- „celou“ nukleotidovou sekvenci (mRNA)

ve frame „1“ direct

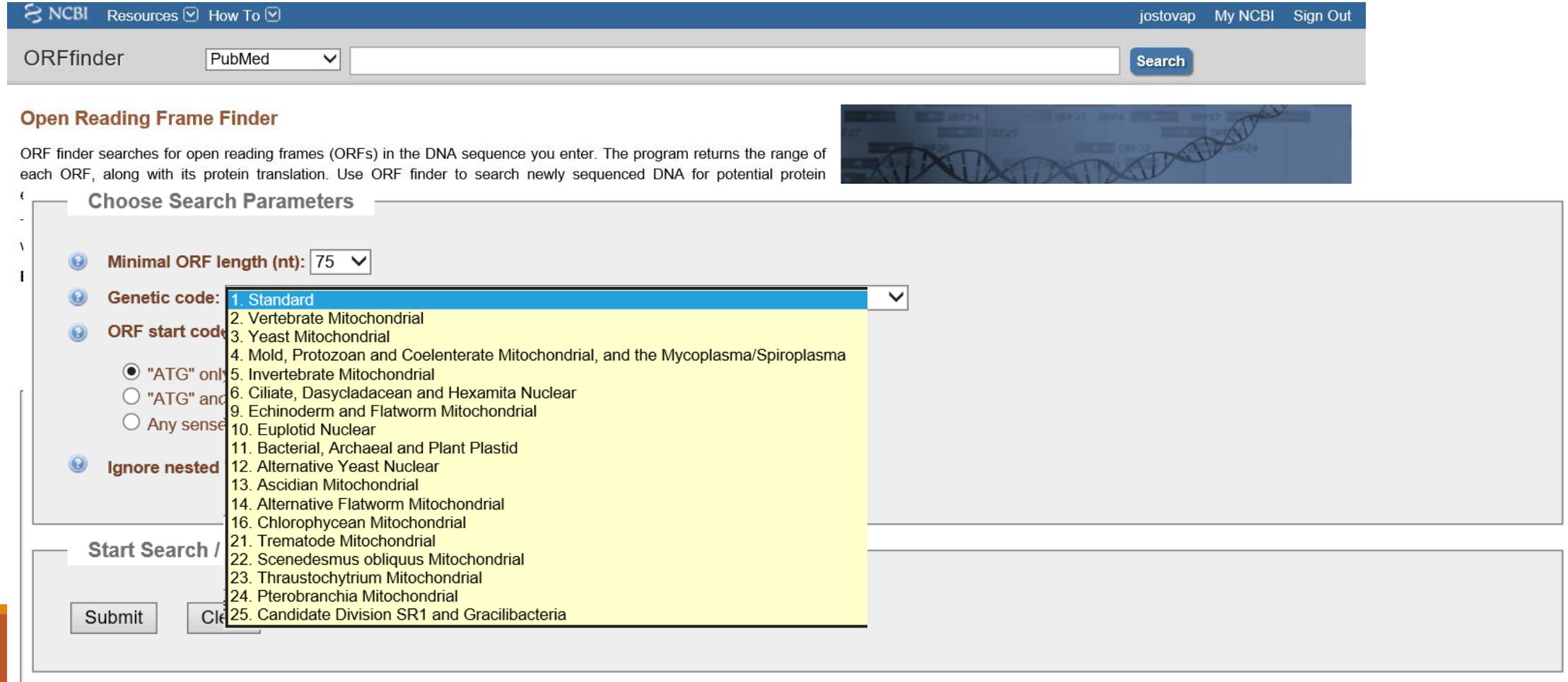
– dokážete určit zda je to „správný čtecí rámec“ pro vaši sekvenci?

zkuste si v programu SMS/translate přeložit

- jen CDS

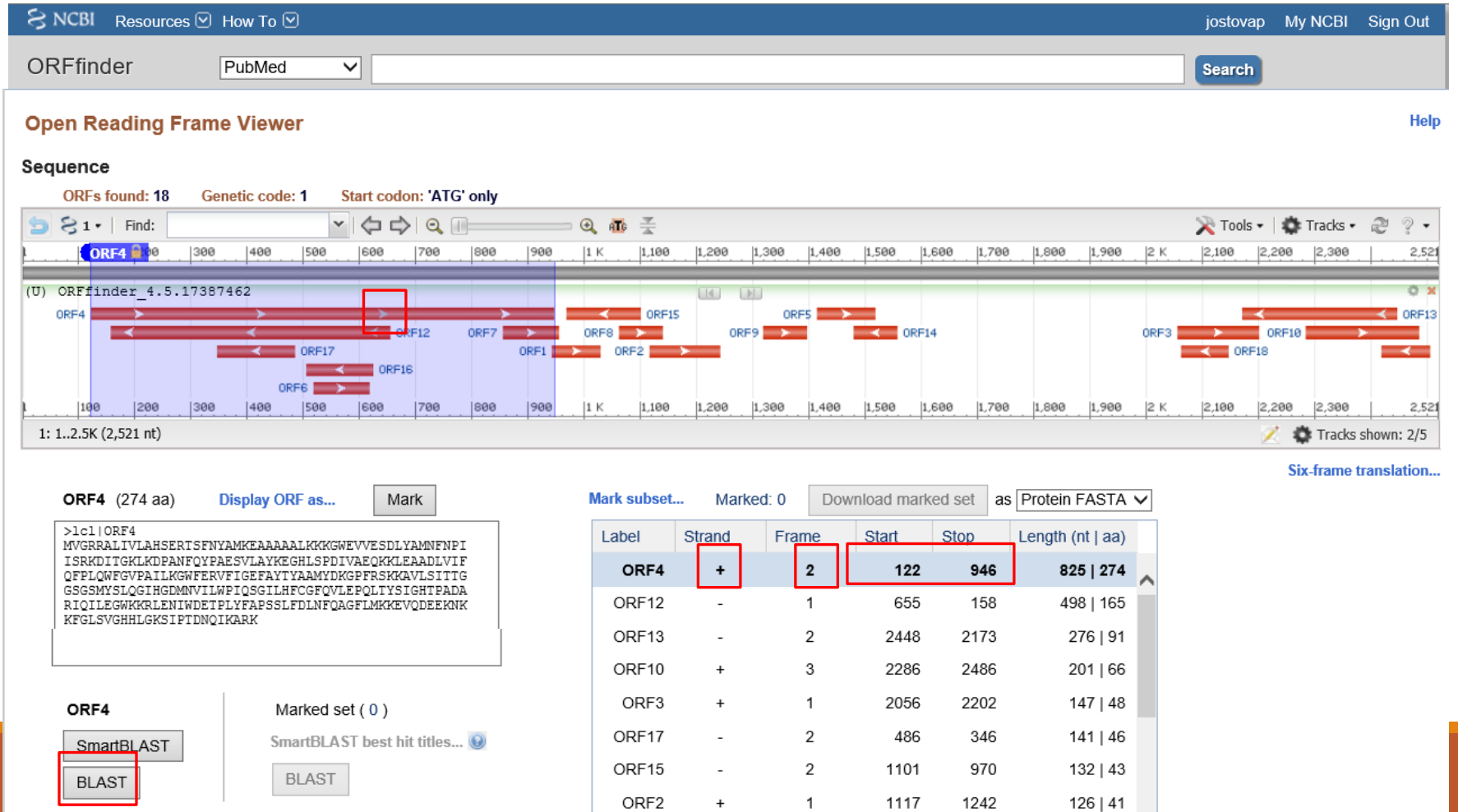
„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“



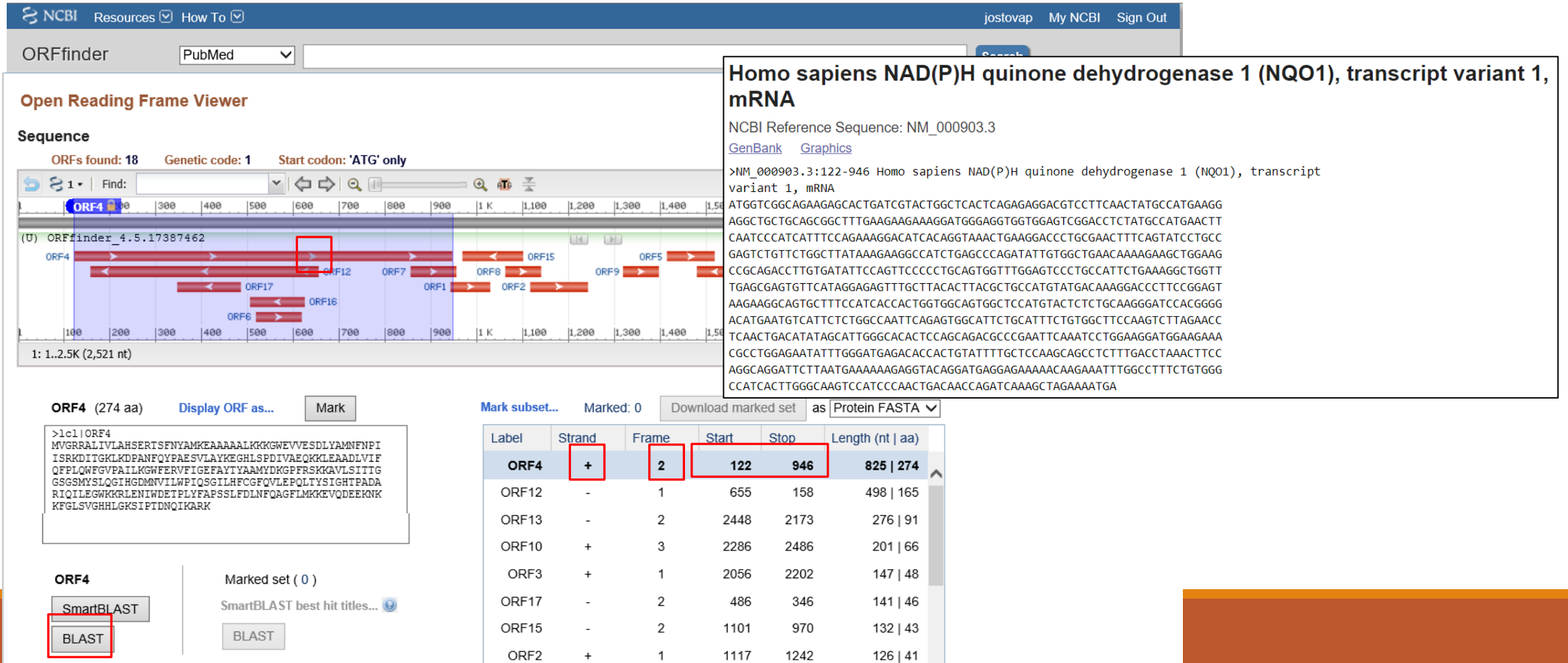
„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

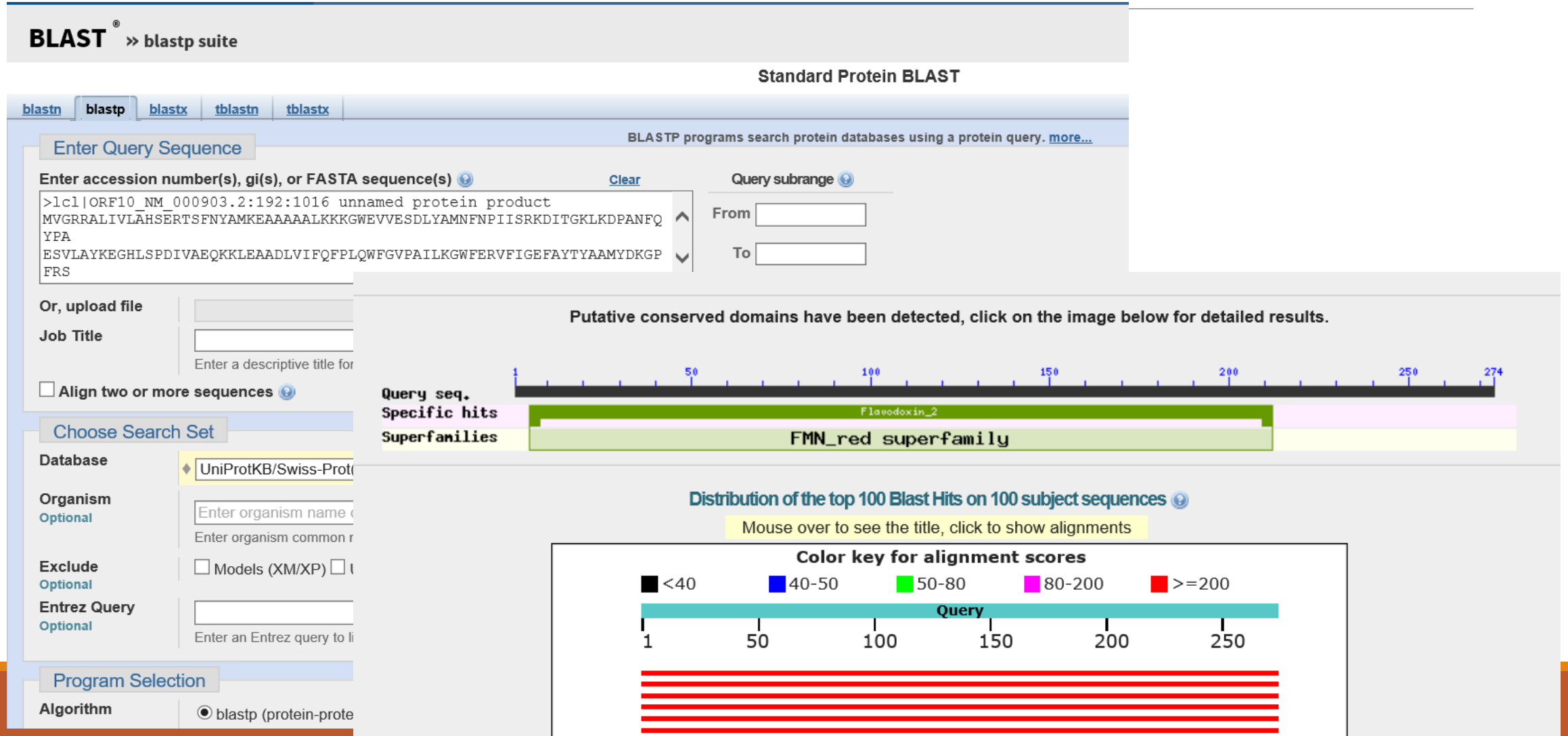


„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

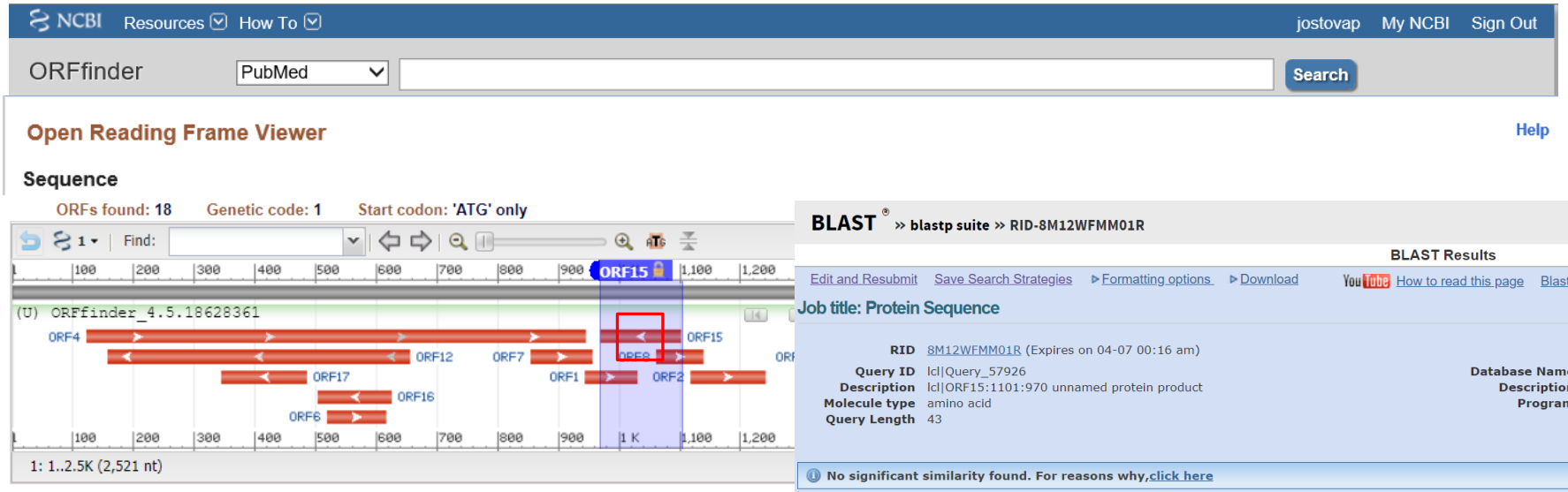


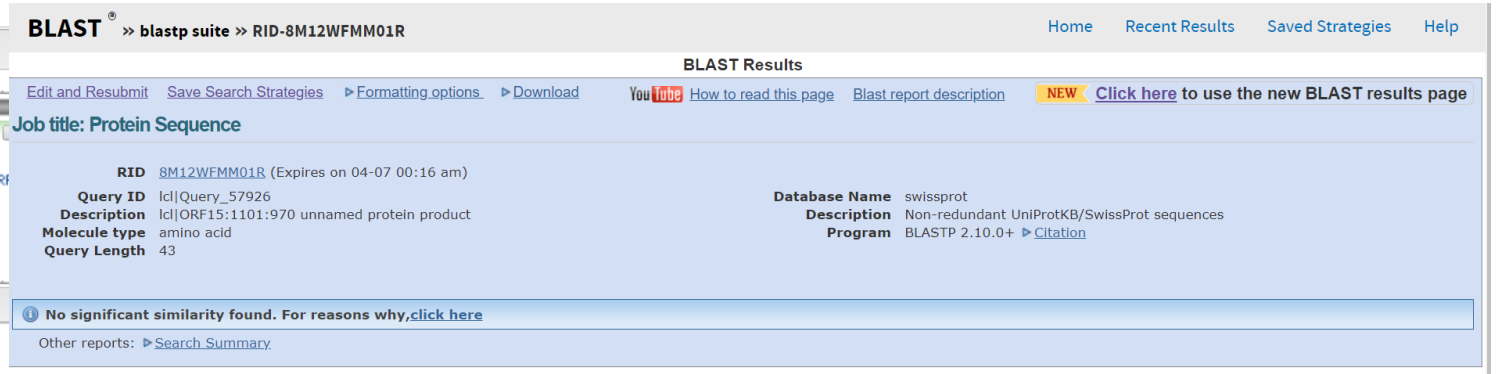
„čtení“ DNA = translace/překlad

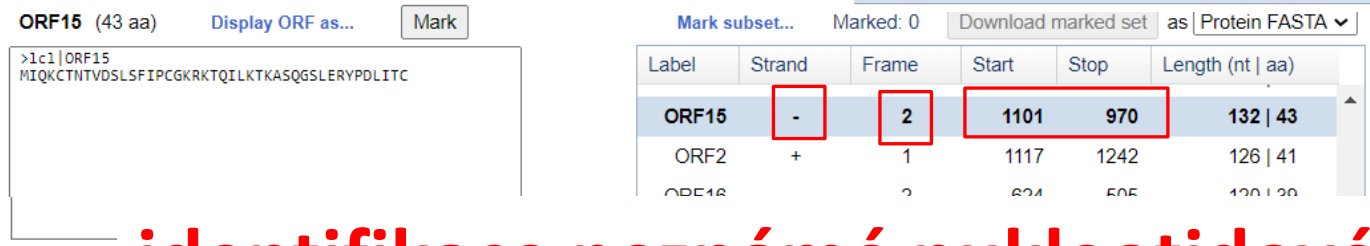


„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“







identifikace neznámé nukleotidové sekvence „přes protein“

Vyzkoušejte si „překládání“ nt sekvence

1) „přeložte“ vaši nukleotidovou sekvenci pomocí ORFfinderu (kompletní mRNA, včetně UTR) →porovnejte váš překlad se záznamem v NCBI

2) vyzkoušejte BLASTp i s nějakým jiným čtecím rámcem (ORFx)

Vyzkoušejte si „překládání“ nt sekvence

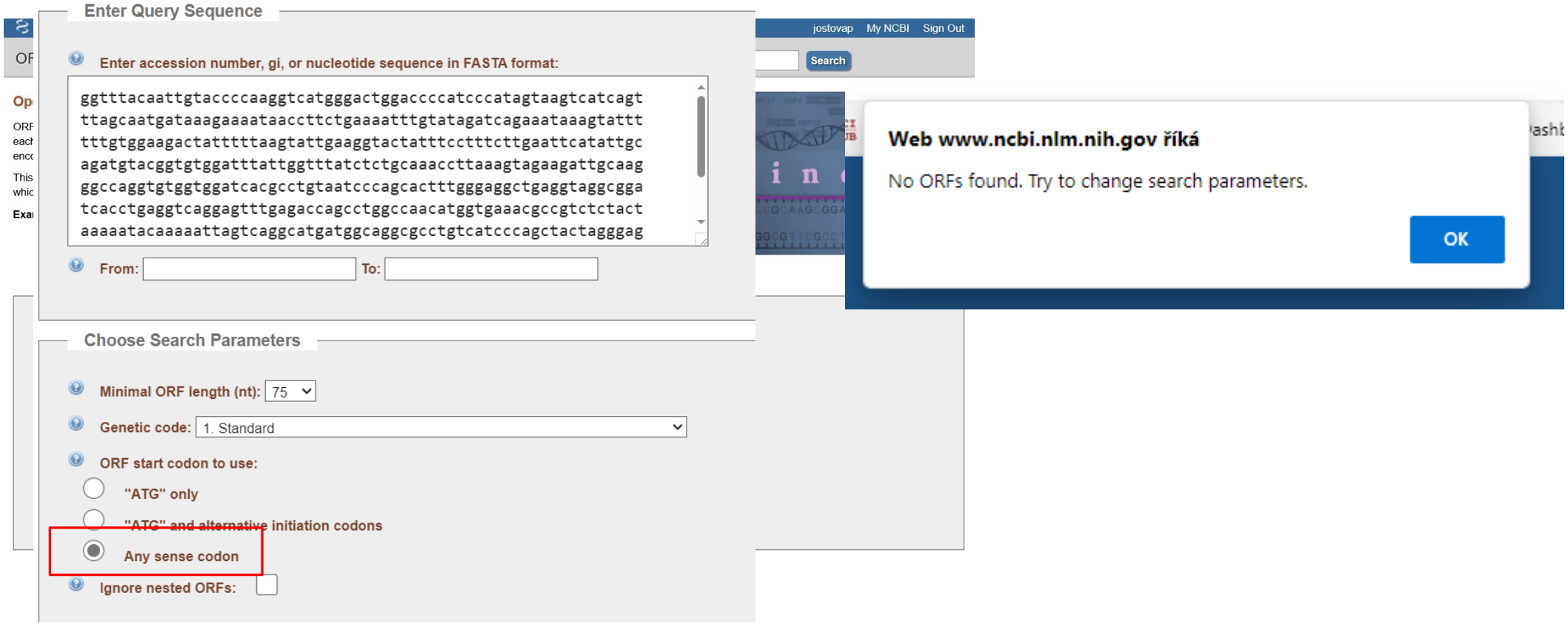
3) vyzkoušejte překlad s **neznámou sekvencí 1** z příkladu 2

-najděte ve kterém čtecím se otevírá „ Př1 neznámá sekvence“

-pomocí programu **BLASTp** (nebo SMART blast přímo z ORFfinderu) identifikujte o jakou sekvenci se jedná

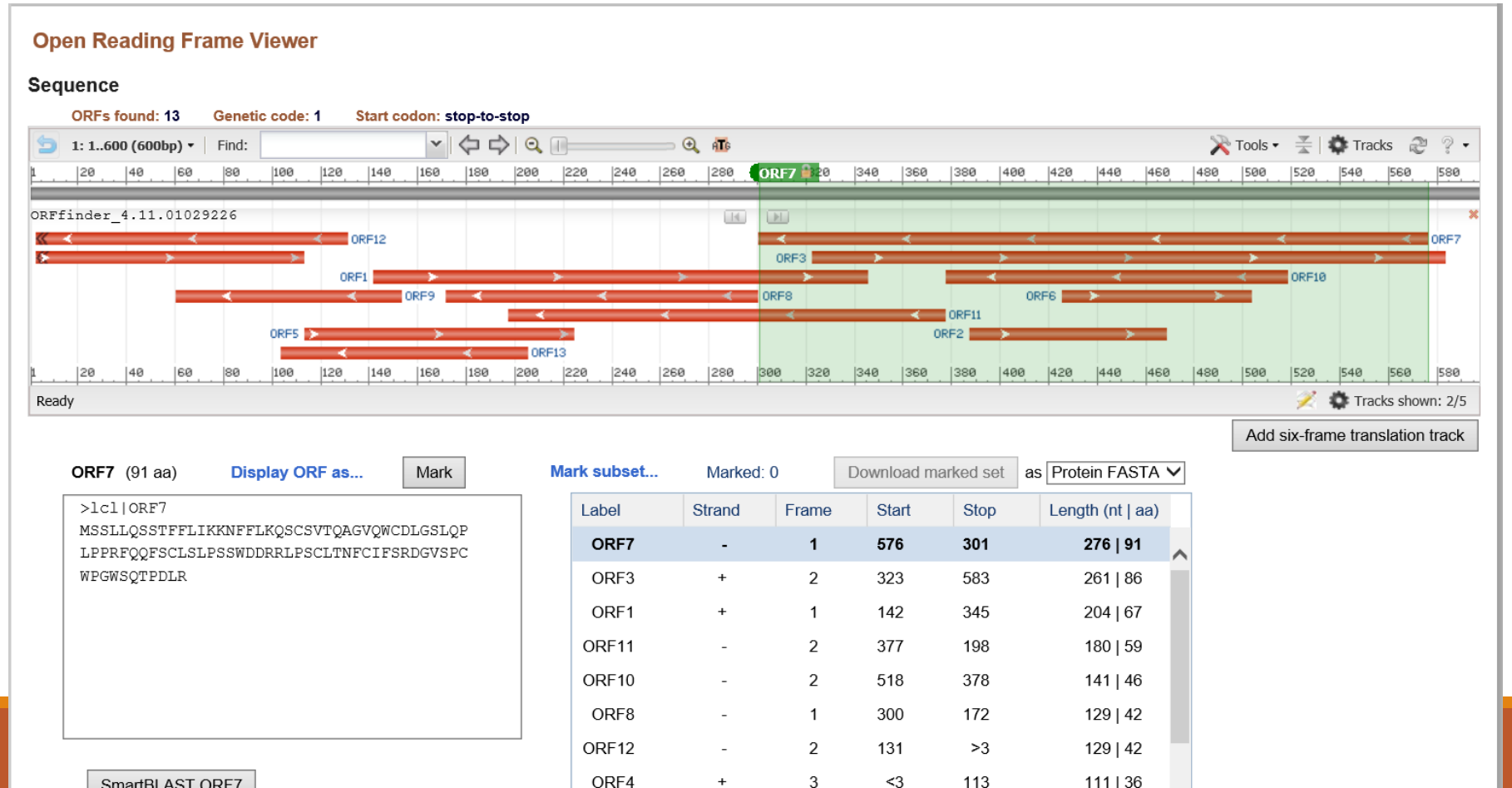
4) vyzkoušejte totéž s **neznámou sekvencí 2** z příkladu 2

„čtení“ DNA = translace/překlad



Identifikace **neznámé** sekvence

-nemá-li „smysluplný“ nebo „jednoznačný“ ORF



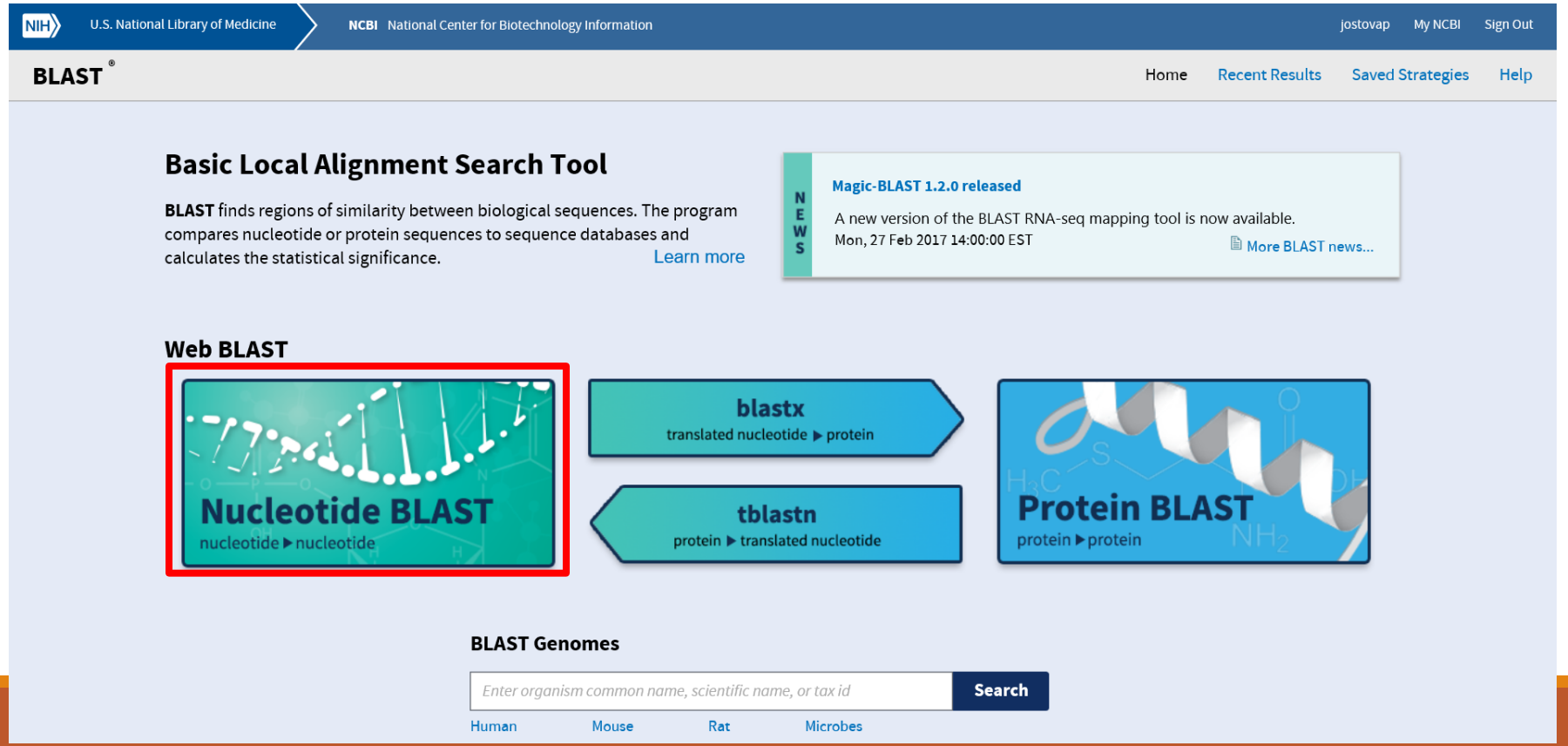
Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF7	**-**	**1**	**576**	**301**	**276 | 91**
ORF3	+	2	323	583	261 | 86
ORF1	+	1	142	345	204 | 67
ORF11	-	2	377	198	180 | 59
ORF10	-	2	518	378	141 | 46
ORF8	-	1	300	172	129 | 42
ORF12	-	2	131	>3	129 | 42
ORF4	+	3	<3	113	111 | 36

Identifikace **neznámé** sekvence

-nemá-li „smysluplný“ nebo „jednoznačný“ ORF

→ vyhledání příbuzných sekvencí

BLAST!

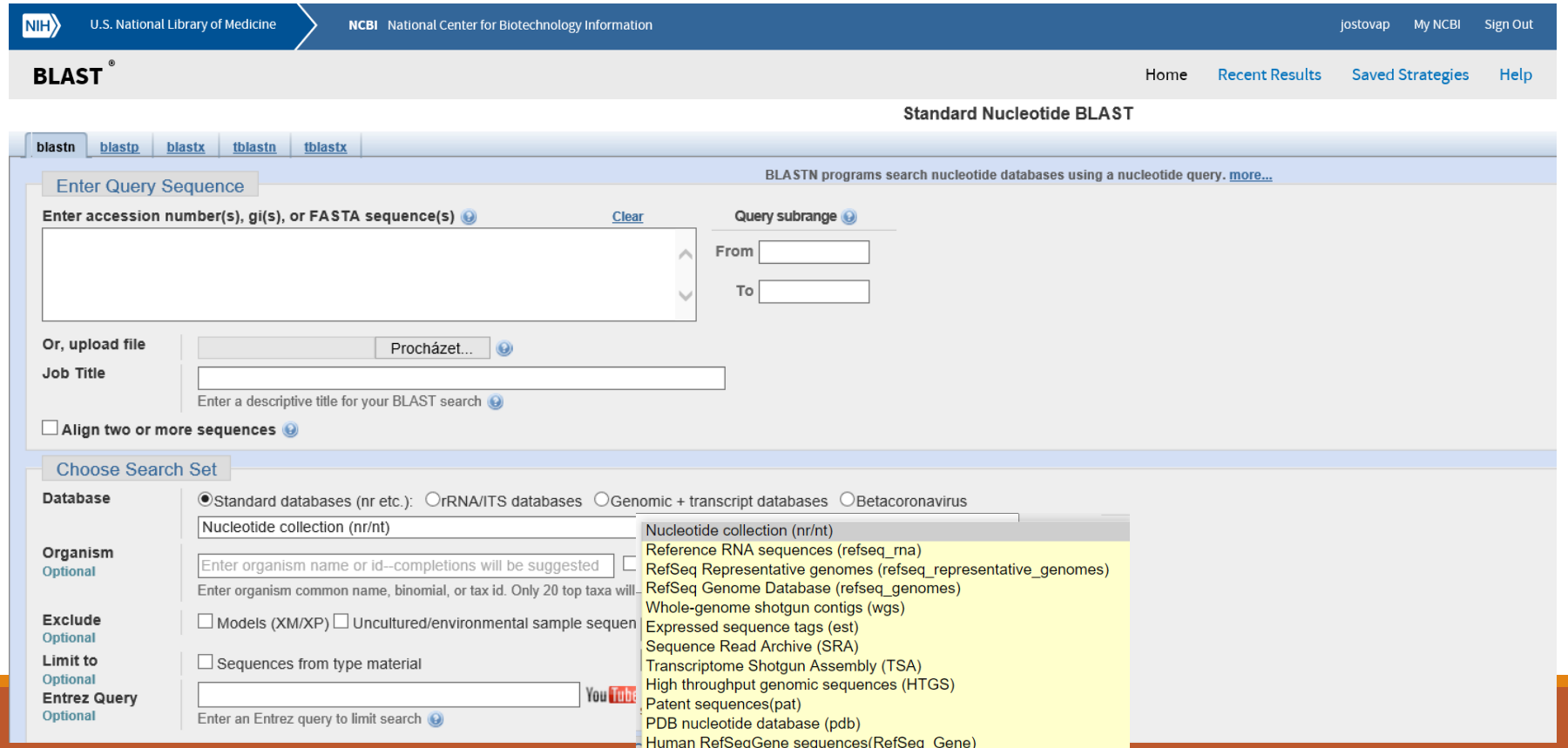


Identifikace **neznámé** sekvence

-nemá-li „smysluplný“ nebo „jednoznačný“ ORF

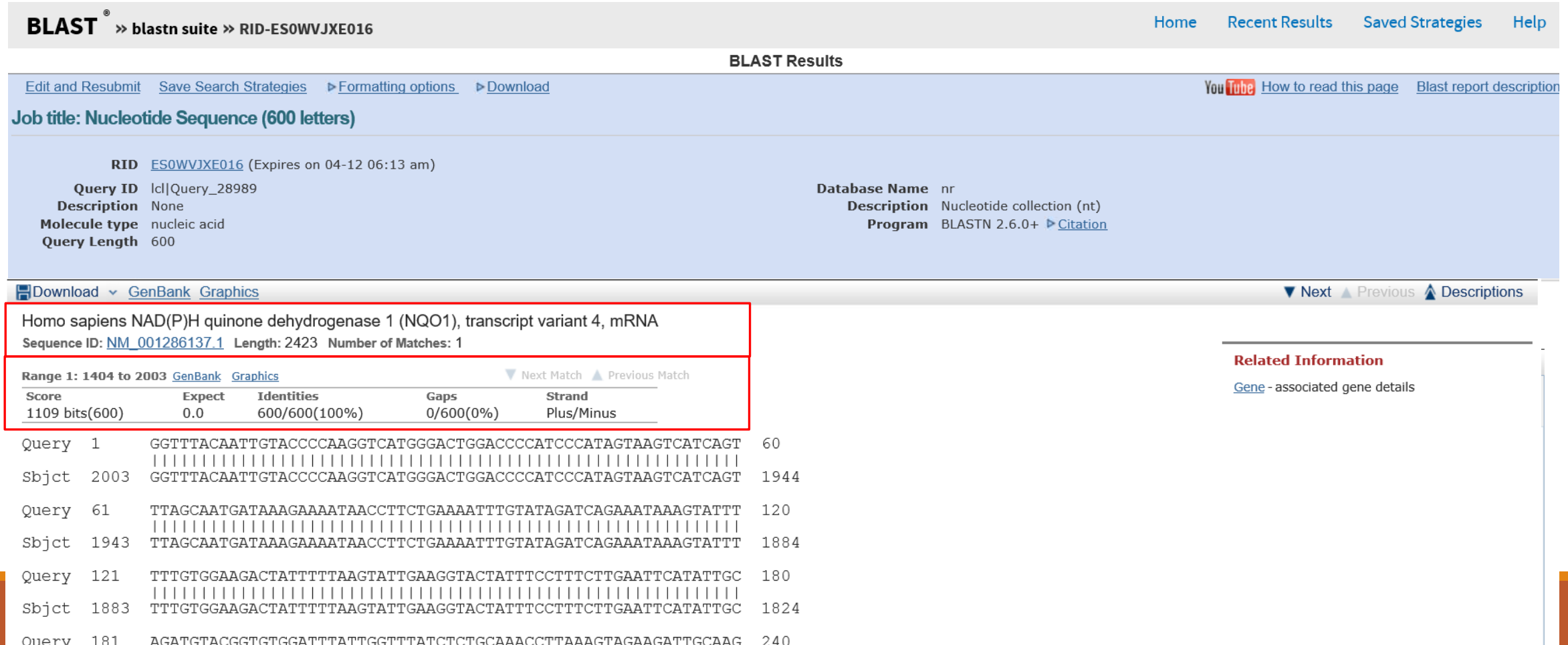
→ vyhledání příbuzných sekvencí

BLAST!



Identifikace **neznámé** sekvence

-nemá-li „smysluplný“ nebo „jednoznačný“ ORF → BLASTn



Vyzkoušejte si identifikaci neznámé sekvence

Zkopírujte si „Př2.neznámou sekvenci“ a BLASTujte (Blastn)

„Bioinformatika nukleových kyselin II“

Vyhledávání NK sekvencí

Analýza vlastností sekvencí-složení, reverse complement, identifikace restrikčních míst (Palindromy)

Práce s kódující DNA=práce s proteiny / překlad DNA sekvence-otvírání čtecího rámce

Návrh primerů pro PCR, rt-PCR

Předpověď sekundárních struktur

Porovnávání sekvencí, identifikace neznámé sekvence

(Vyhledání SNPs)

„čtení“ sekvenačních dat a spojování fragmentů

Vyhledávání hladin expresí jednotlivých genů

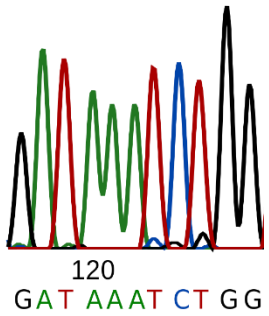
mikroRNA

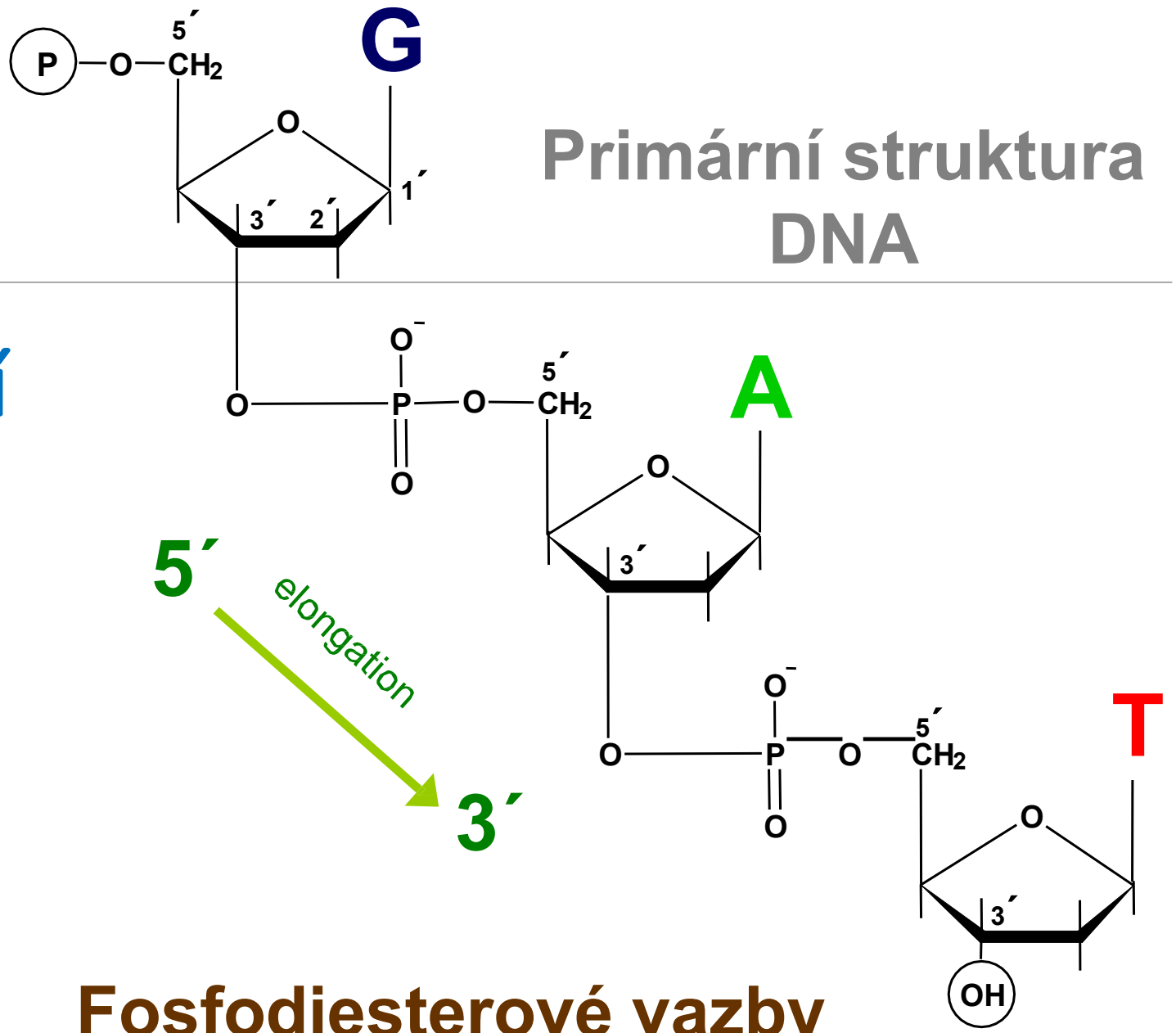
Celé genomy

….

Primární struktura DNA

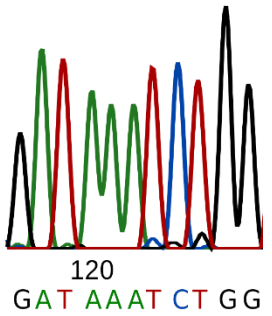
DNA sekvenování





Fosfodiesterové vazby

DNA sekvenování

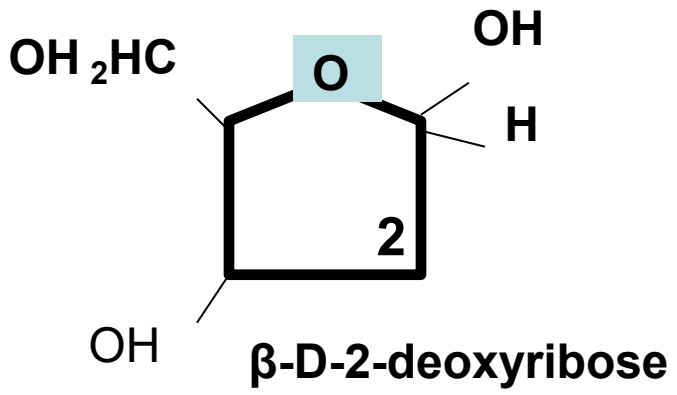


„Klasické“ Sangerovo sekvenování (1977)

→ Délka čtení: 200-1000nt

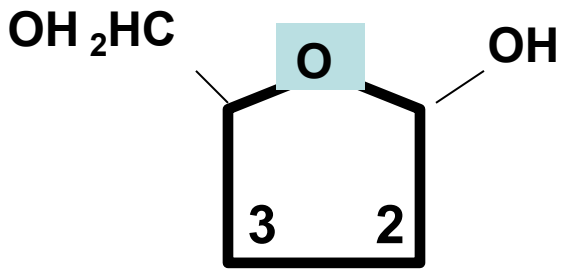
→ Výsledek: „text“ 4 písmena (ACTG)

Deoxynukleotidy x **dideoxynukleotidy**



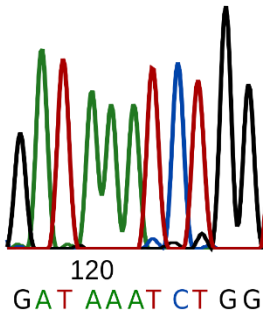
β-D-2-deoxyribose

~1:100



β-D-2,3-dideoxyribose

DNA sekvenování

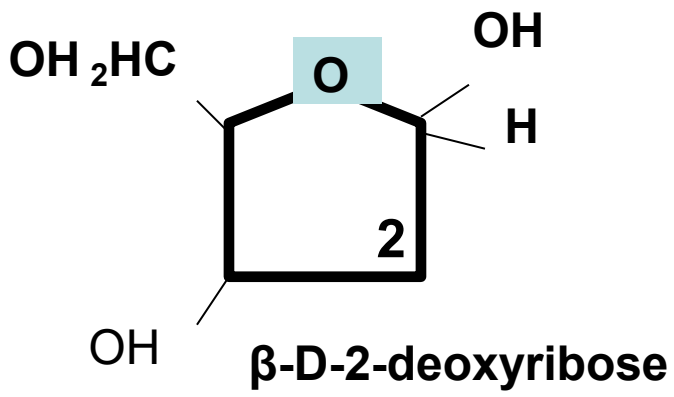


„Klasické“ Sangerovo sekvenování (1977)

- Asymetrická DNA amplifikace (DNA polymeráza I, 1 primer)

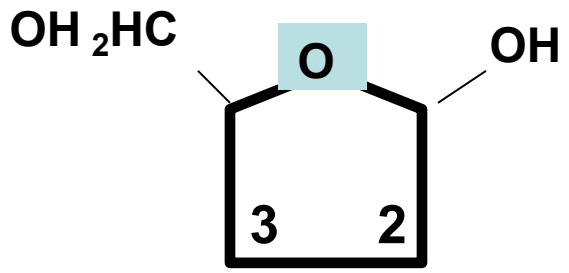
-směr syntézy 5´→ 3´

Deoxynukleotidy x **dideoxynukleotidy**



β-D-2-deoxyribose

~1:100



β-D-2,3-dideoxyribose

DNA sekvenování

Enzymatické „Sangerovo“ sekvenování (1977)

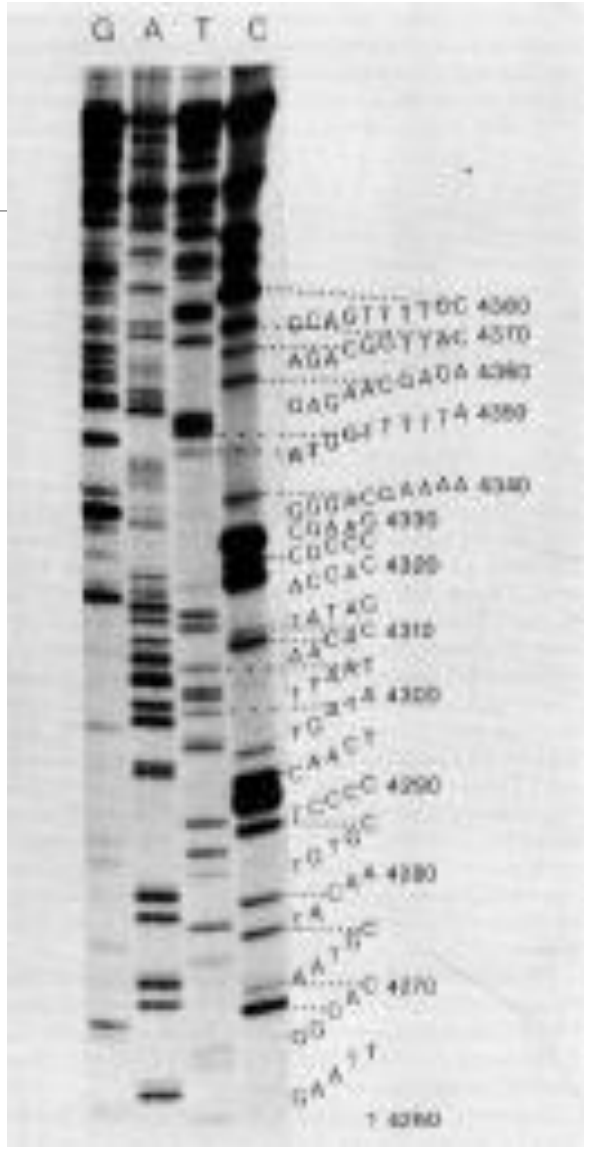
Asymetrická DNA amplifikace, značení ^{32}P-dATP

Specifické zakončení polymerace

G	A	T	C
Reakční mix 1 *3 dNTPs* *1:100 dGTP:ddGTP*	*Reakční mix 2* *3 dNTPs* *1:1 dATP:ddATP*	*Reakční mix 3* *3 dNTPs* *1:100 dTTP:ddTTP*	*Reakční mix 4* *3 dNTPs* *1:50 dCTP:ddCTP*

PAGE/UREA > Southern blotting

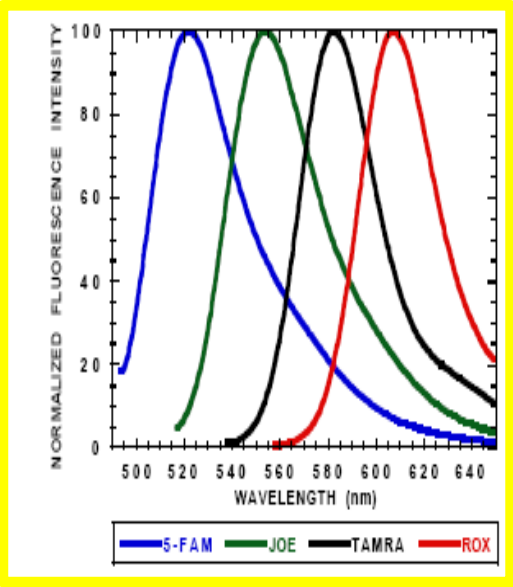
RA detekce



DNA sekvenování

1980s nové fluorofory umožnily automatizaci DNA sekvenování

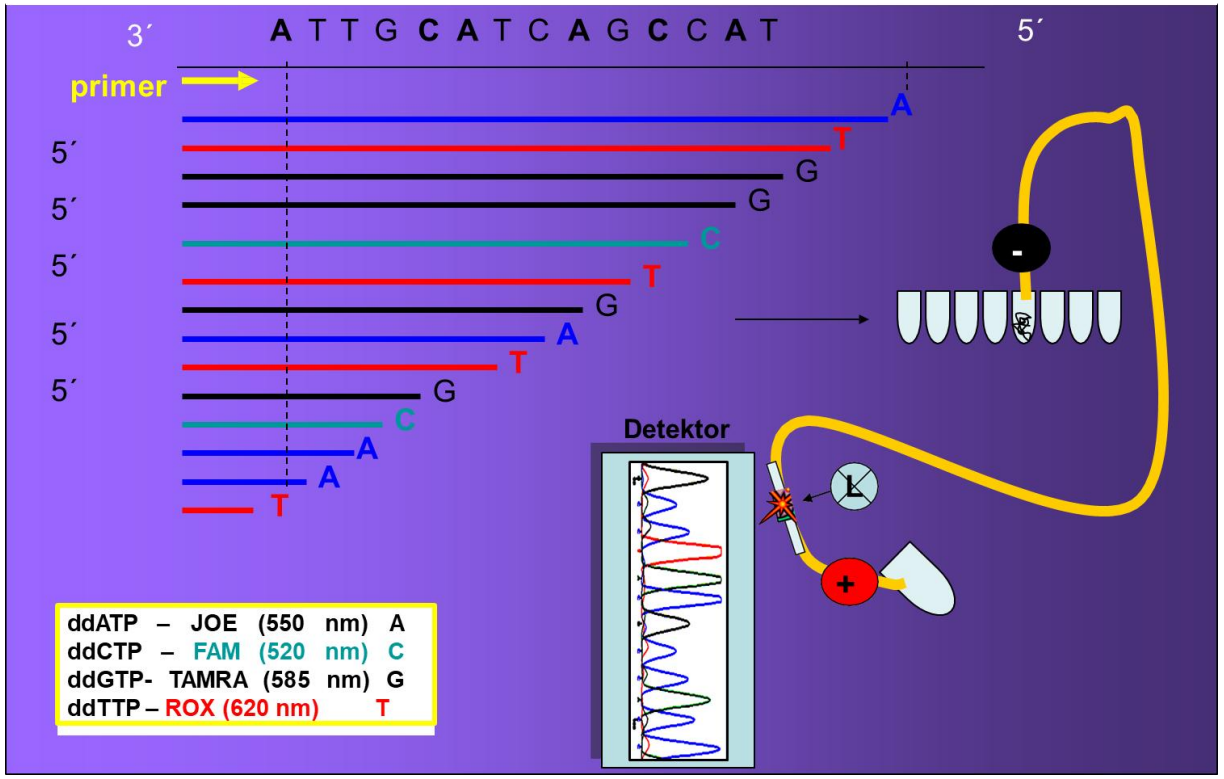
ddATP – JOE (550 nm) A
ddCTP – FAM (520 nm) C
ddGTP- TAMRA (585 nm) G
ddTTP – ROX (620 nm) T



DNA sekvenování

Automatické analyzátory (od 1990s)

Princip: Sangerovo sekvenování a kapilární elektroforéza



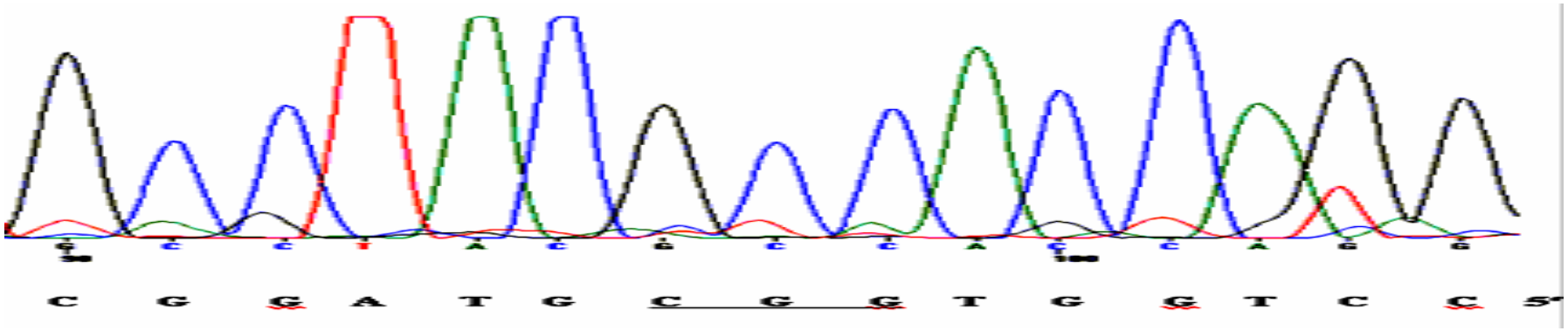
Analýza sekvenčních dat

Chromas → umožní „čtení sekvenčních dat (.scf, .abi, .ab1)

„plain text“: CGGATGCGGTGGTCG

„fasta“: >identifikace
CGGATGCGGTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



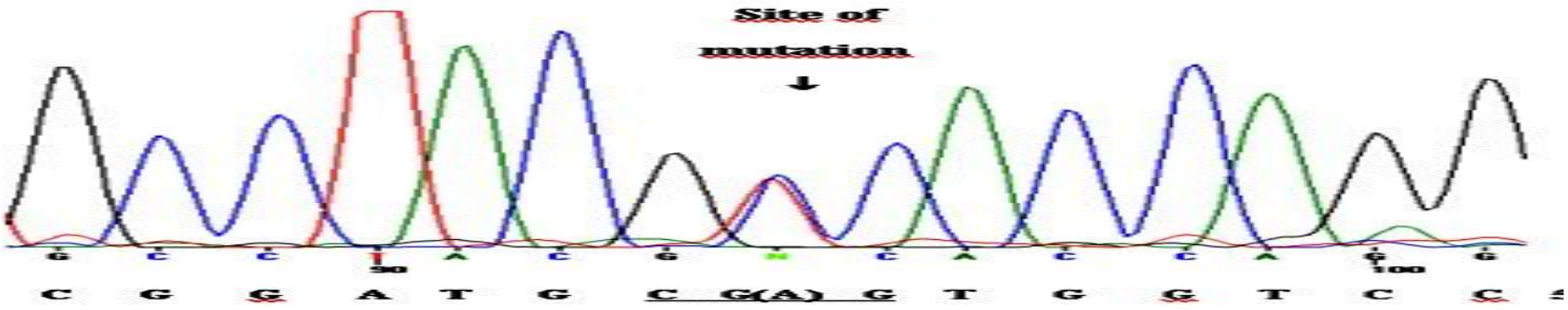
Analýza sekvenčních dat

Chromas → umožní „čtení sekvenčních dat (.scf, .abi, .ab1)

„plain text“: CGGATGCNGTGGTCG

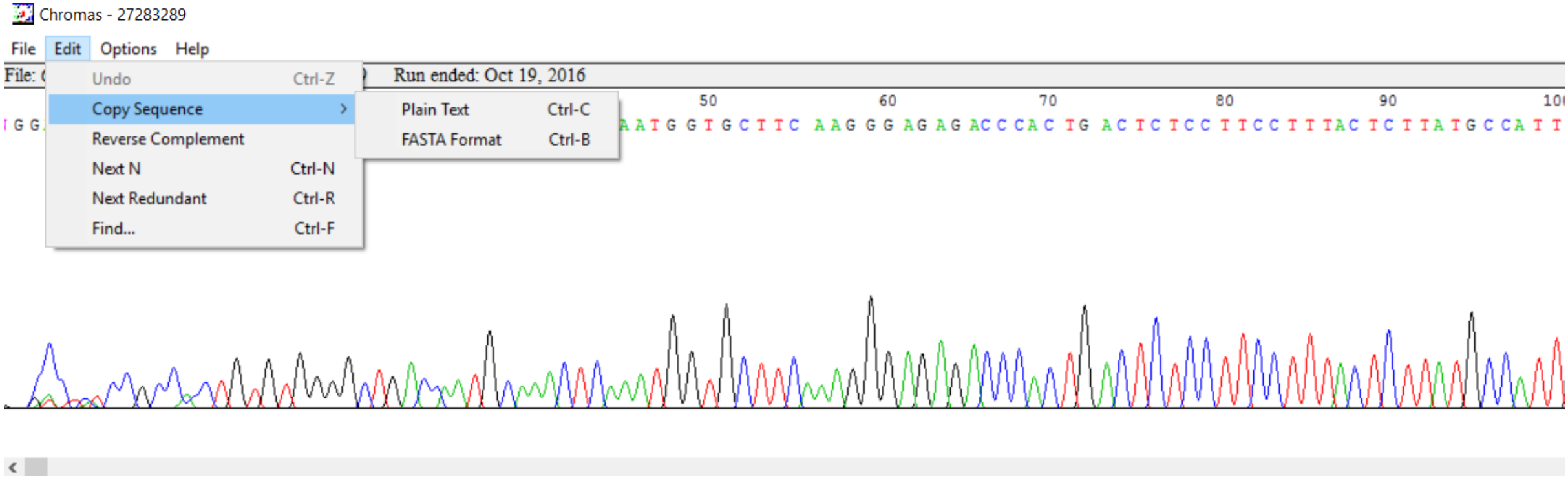
„fasta“: >identifikace
CGGATGC**N**GTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



Analýza sekvenčních dat

Chromas → umožní „čtení sekvenčních dat (.scf, .abi, .ab1)



Vyzkoušejte si čtení sekvenačních dat

Spusťte si „chromas“

Stáhněte a uložte sekvenci př3

Otevřete v chromasu

- Identifikujte sekvenci

  (stáhněte text/fasta, vložte do BLASTu/ncbi BLASTn)

Analýza sekvenčních dat – příklad (př3)

BLAST® » blastn suite » RID-ES1YNVDF016 Home Recent Results Saved Strategies Help

BLAST Results

Edit and Resubmit | Save Search Strategies | Formatting options | Download | How to read this page | Blast report description

Job title: Nucleotide Sequence (1167 letters)

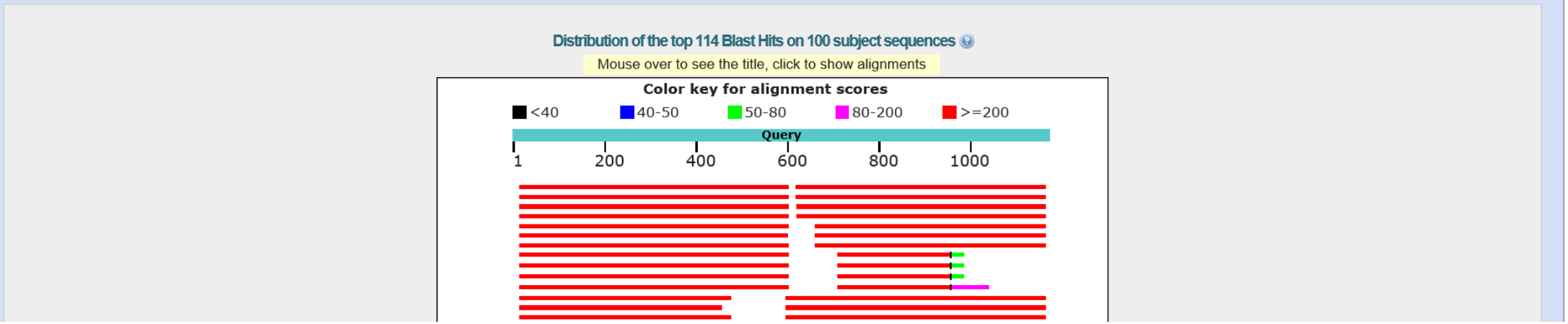
RID ES1YNVDF016 (Expires on 04-12 06:31 am)

Query ID lcl|Query_42683
Description None
Molecule type nucleic acid
Query Length 1167

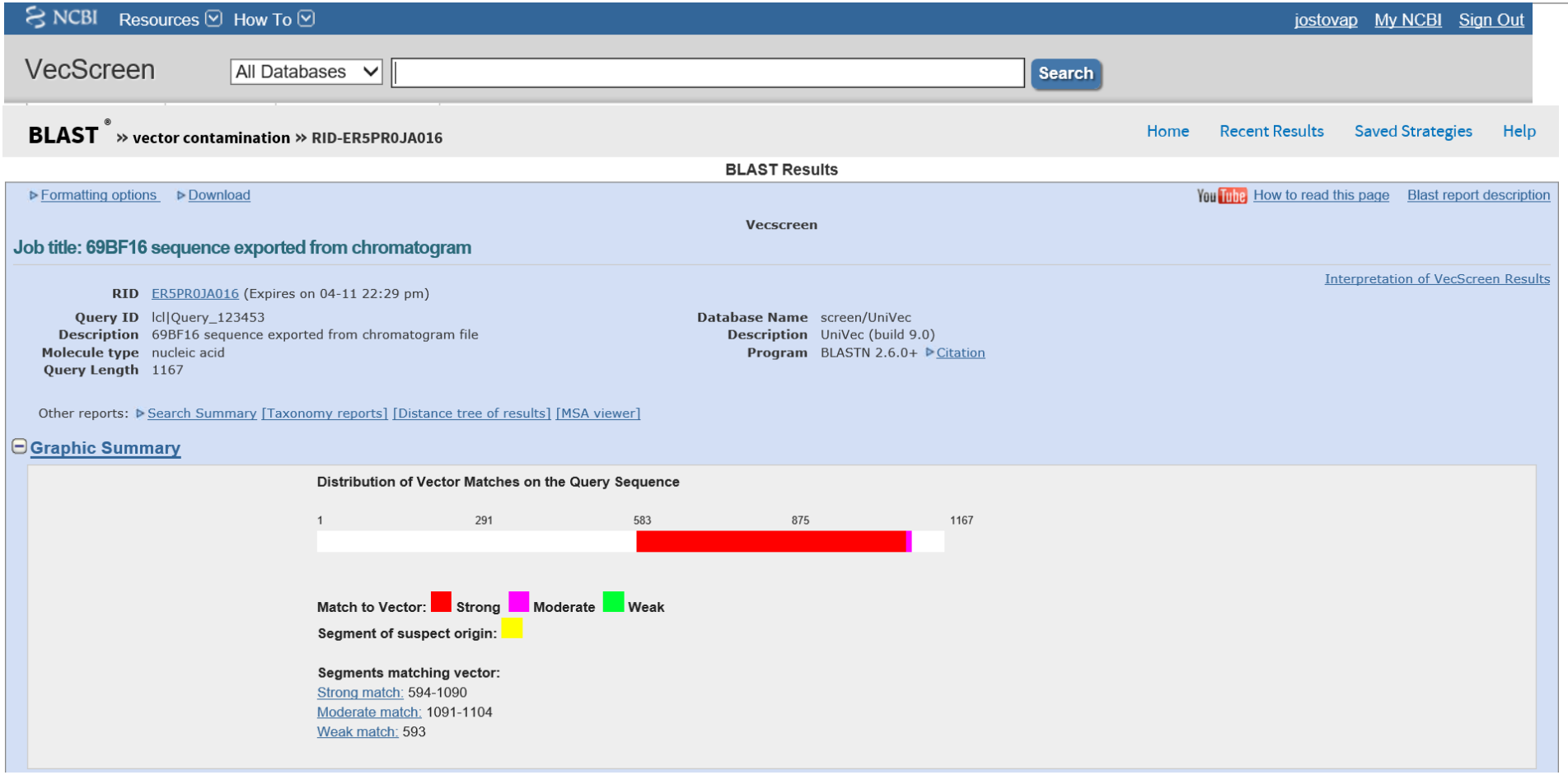
Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.6.0+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [MSA viewer]

Graphic Summary



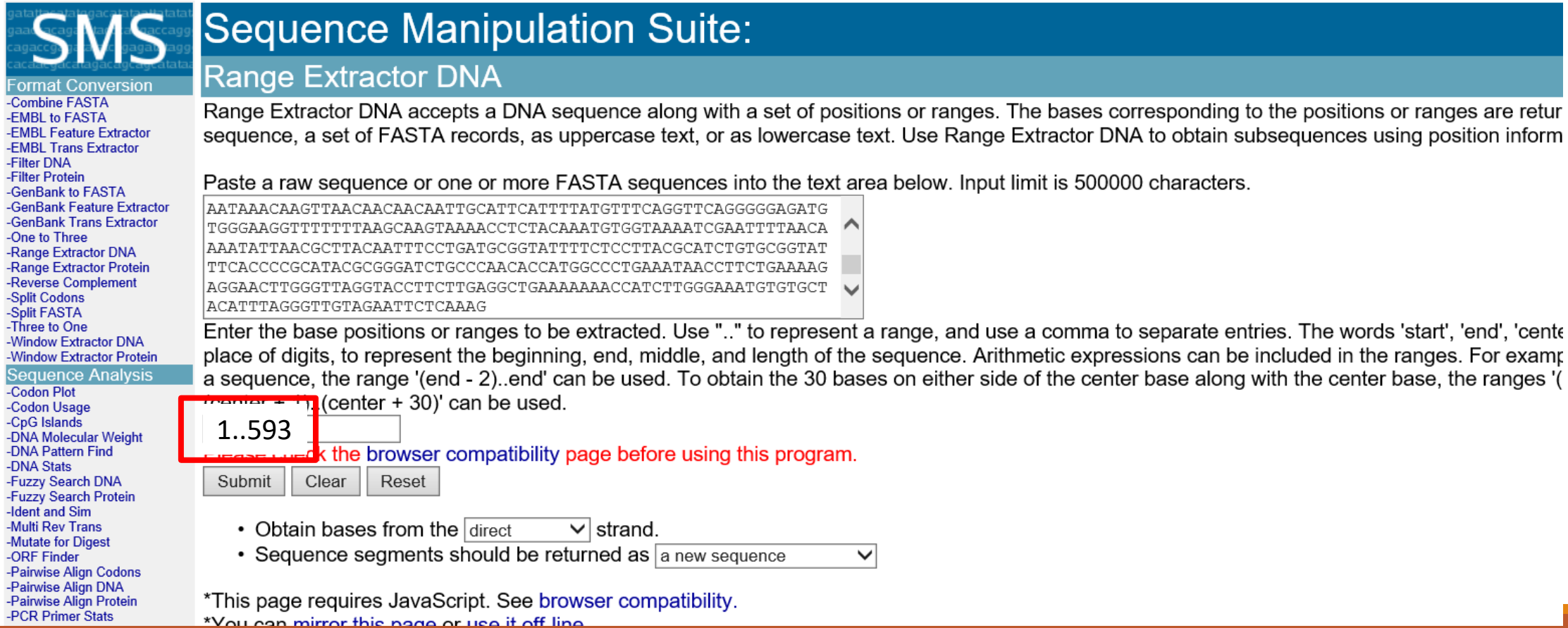
Analýza sekvenčních dat – příklad (př3)



Analýza sekvenčních dat – příklad (př3)

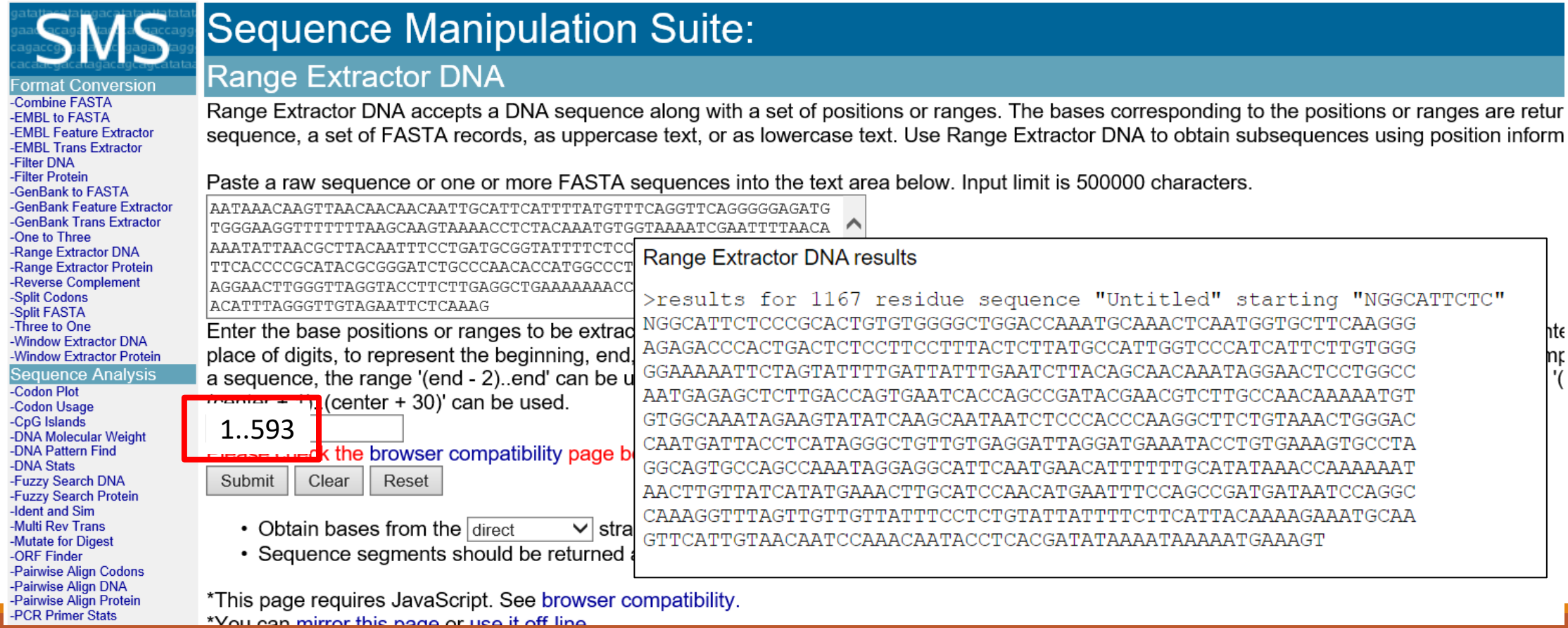
Analýza sekvenčních dat: odstranění kontaminace

SMS „Range Extraktor“



Analýza sekvenčních dat: odstranění kontaminace

SMS „Range Extraktor“



DÚ6

1) Porovnejte celou mRNA a CDS „vaší“ nukleotidové sekvence.

2) Přeložte „vaši“ nukleotidovou sekvenci (celou mRNA), ve kterém čtecím rámci se otevírá?

3) Stáhněte si DU6-neznámou sekvenci v úkolu a otevřete ji programem chromas.

4) Zkontrolujte přítomnost „vektorů“ a „Očištěnou“ sekvenci identifikujte

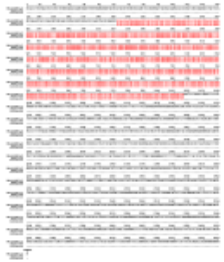
5) Shlédněte toto video PCR Primer Design: https://www.youtube.com/watch?v=c-f1H07D_70

DÚ6-řešení

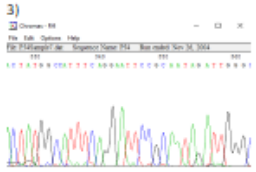
1) NQO1 (mRNA): Čtecí rámec +3

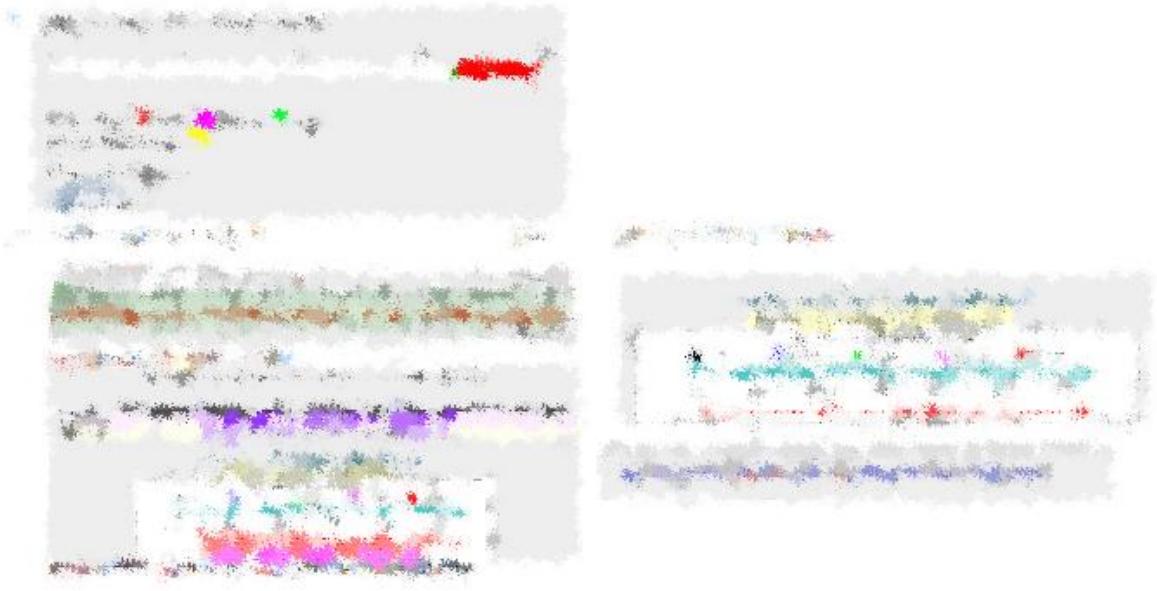
Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF10	+	3	192	1016	825 | 274

2) Porovnání NQO1 mRNA a CDS



3)



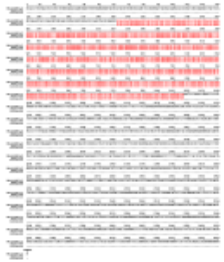
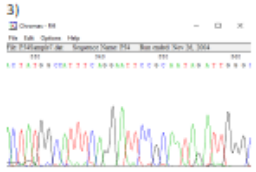
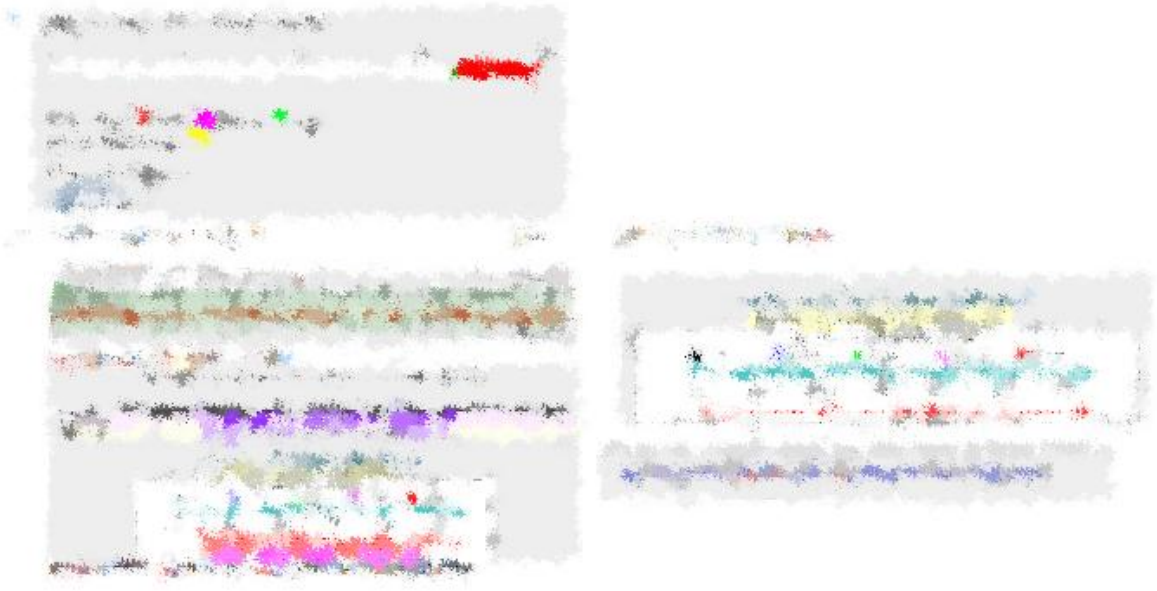


DÚ6-řešení

1) NQO1 (mRNA): Čtecí rámec +3

Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF10	+	3	192	1016	825 | 274

2) Porovnání NQO1 mRNA a CDS



3)