Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2023/2024

6/10

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

Local Alignment

Target Sequence

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
             |||| |||||| |||||||||||||||
          5' TACTCACGGATGAGGTACTTTAGAGGC 3'
```

Query Sequence

Global Alignment

Target Sequence

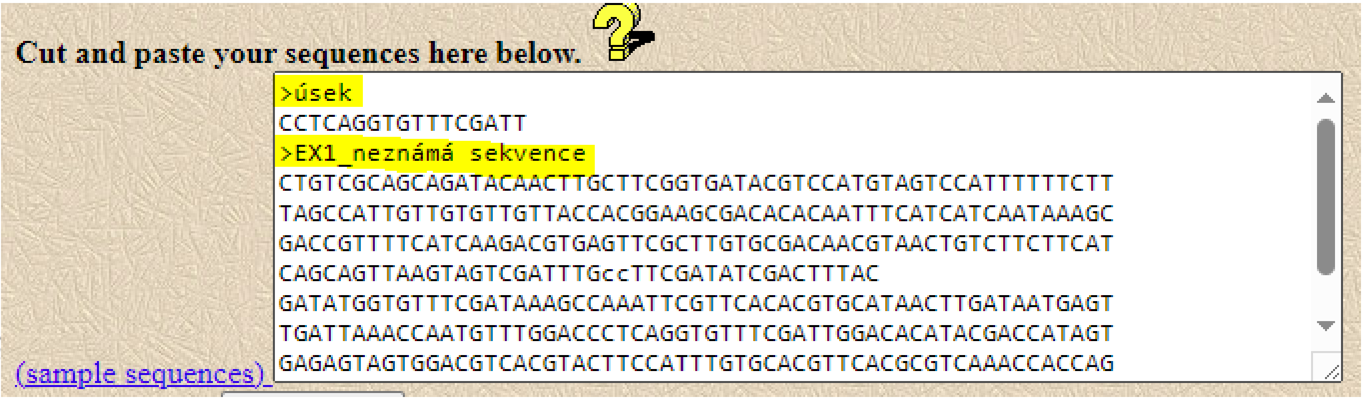
```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
   |||||||||||    |||||||  |||||||||||||| |||||||
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'
```

Query Sequence

Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Např: multalin

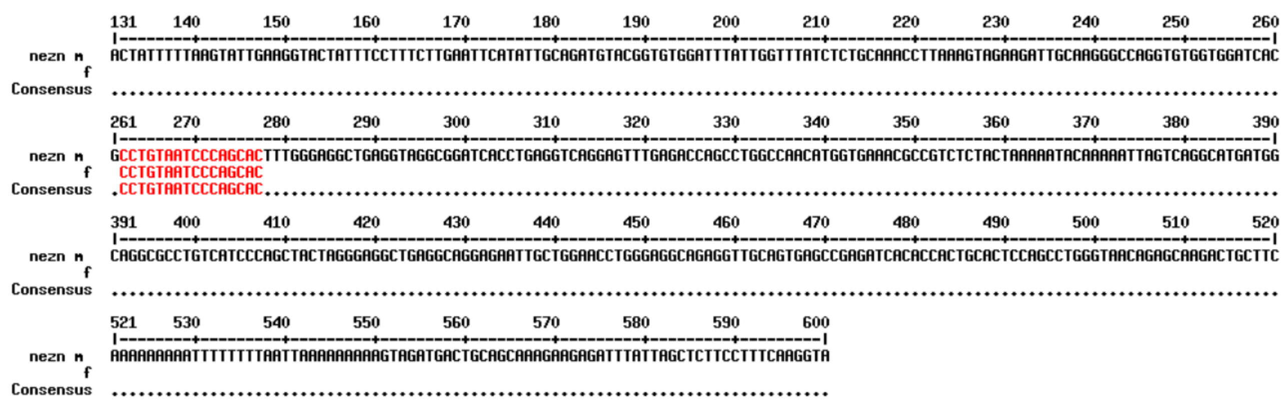


Cut and paste your sequences here below.

```
>úsek
CCTCAGGTGTTTCGATT
>EX1_neznámá sekvence
CTGTCGCAGCAGATACAACTTGCTTCGGTGATACGTCCATGTAGTCCATTTTTTCTT
TAGCCATTGTTGTGTTGTTACCACGGAAGCGACACACAATTTCATCATCAATAAAGC
GACCGTTTTCATCAAGACGTGAGTTCGCTTGTGCGACAACGTAACTGTCTTCTTCAT
CAGCAGTTAAGTAGTCGATTTGccTTCGATATCGACTTTAC
GATATGGTGTTTCGATAAAGCCAAATTCGTTCACACGTGCATAACTTGATAATGAGT
TGATTAAACCAATGTTTGGACCCTCAGGTGTTTCGATTGGACACATACGACCATAGT
GAGAGTAGTGGACGTCACGTACTTCCATTTGTGCACGTTCACGCGTCAAACCACCAG
```

- ➢ K vyhledání krátkých úseků (primerů)

Default:BLOSUM62



```
           1       10        20        30        40        50        60        70        80
           |--------+---------+---------+---------+---------+---------+---------+---------+----
   nezn м  GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGTTTAGCAATGATAAAGAAAATAACC
        f
Consensus  ....................................................................................

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

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

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

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

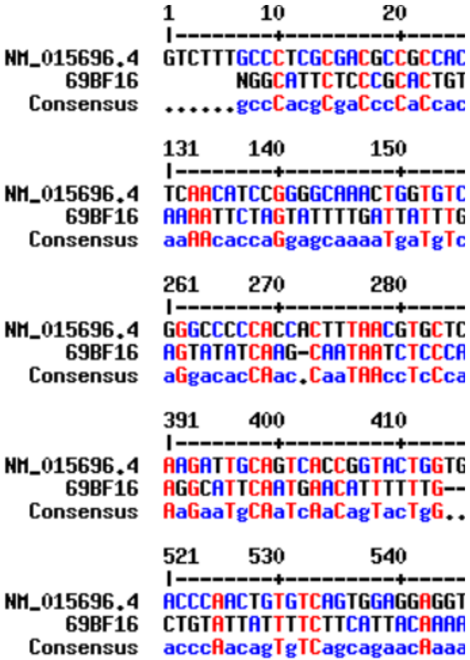
Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Např: multalin

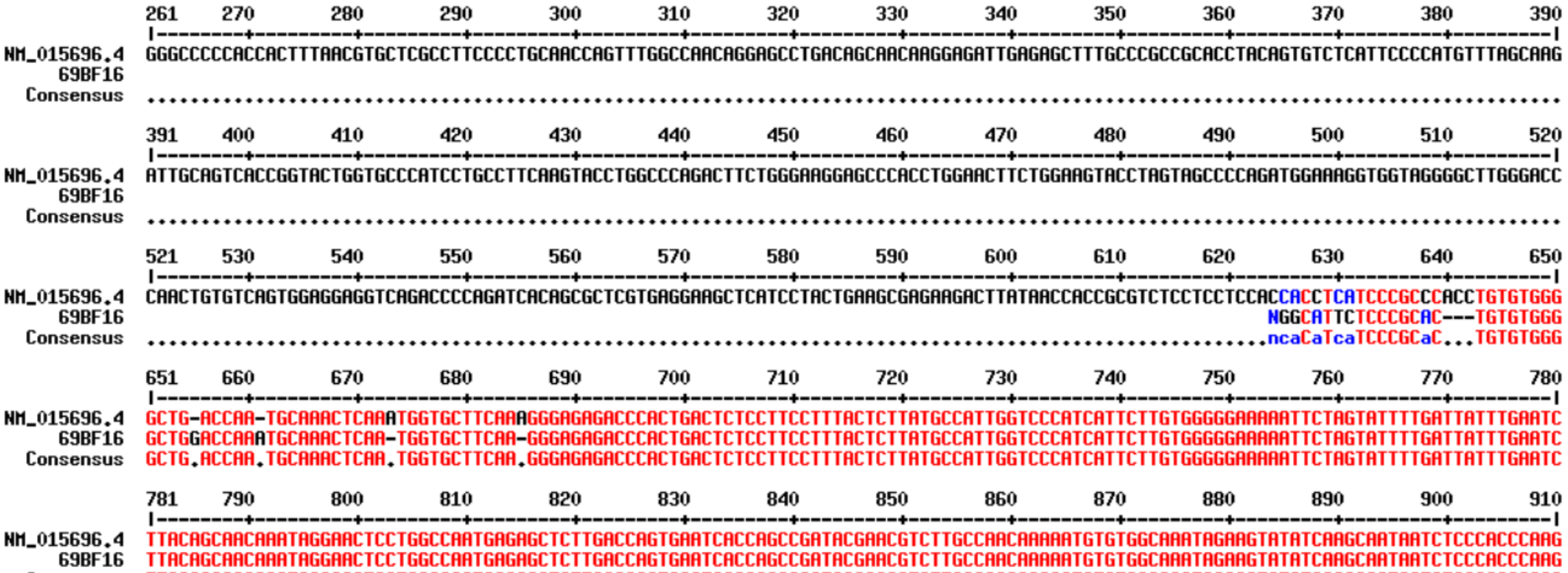
- K „porovnání“ sekvencí identifikovaných a neznámých

Default:BLOSUM62



```
            1       10        20
            |--------+---------+-----
NM_015696.4 GTCTTTGCCCTCGCGACGCCGCCAC
     69BF16       NGGCATTCTCCCGCACTGT
  Consensus ......gccCacgCgaCccCaCcac

            131     140       150
            |--------+---------+-----
NM_015696.4 TCAACATCCGGGGCAAACTGGTGTC
     69BF16 AAAATTCTAGTATTTTGATTATTTG
  Consensus aaAAcaccaGgagcaaaaTgaTgTc

            261     270       280
            |--------+---------+-----
NM_015696.4 GGGCCCCCACCACTTTAACGTGCTC
     69BF16 AGTATATCAAG-CAATAATCTCCCA
  Consensus aGgacacCAac.CaaTAAccTcCca

            391     400       410
            |--------+---------+-----
NM_015696.4 AAGATTGCAGTCACCGGTACTGGTG
     69BF16 AGGCATTCAATGAACATTTTTTG--
  Consensus AaGaaTgCAaTcAaCagTacTgG..

            521     530       540
            |--------+---------+-----
NM_015696.4 ACCCAACTGTGTCAGTGGAGGAGGT
     69BF16 CTGTATTATTTTCTTCATTACAAAA
  Consensus acccAacagTgTCagcagaacAaaa
```

→DNA 5-0



```
            261     270       280       290       300       310       320       330       340       350       360       370       380       390
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_015696.4 GGGCCCCCACCACTTTAACGTGCTCGCCTTCCCCTGCAACCAGTTTGGCCAACAGGAGCCTGACAGCAACAAGGAGATTGAGAGCTTTGCCCGCCGCACCTACAGTGTCTCATTCCCCATGTTTAGCAAG
     69BF16
  Consensus ..................................................................................................................................

            391     400       410       420       430       440       450       460       470       480       490       500       510       520
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_015696.4 ATTGCAGTCACCGGTACTGGTGCCCATCCTGCCTTCAAGTACCTGGCCCAGACTTCTGGGAAGGAGCCCACCTGGAACTTCTGGAAGTACCTAGTAGCCCCAGATGGAAAGGTGGTAGGGGCTTGGGACC
     69BF16
  Consensus ..................................................................................................................................

            521     530       540       550       560       570       580       590       600       610       620       630       640       650
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_015696.4 CAACTGTGTCAGTGGAGGAGGTCAGACCCCAGATCACAGCGCTCGTGAGGAAGCTCATCCTACTGAAGCGAGAAGACTTATAACCACCGCGTCTCCTCCTCCACCACCTCATCCCGCCCACCTGTGTGGG
     69BF16                                                                                                        NGGCATTCTCCCGCAC---TGTGTGGG
  Consensus .......................................................................................................ncaCaTcaTCCCGCaC...TGTGTGGG

            651     660       670       680       690       700       710       720       730       740       750       760       770       780
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_015696.4 GCTG-ACCAA-TGCAAACTCAAATGGTGCTTCAAAGGGAGAGACCCACTGACTCTCCTTCCTTTACTCTTATGCCATTGGTCCCATCATTCTTGTGGGGGAAAAATTCTAGTATTTTGATTATTTGAATC
     69BF16 GCTGGACCAAATGCAAACTCAA-TGGTGCTTCAA-GGGAGAGACCCACTGACTCTCCTTCCTTTACTCTTATGCCATTGGTCCCATCATTCTTGTGGGGGAAAAATTCTAGTATTTTGATTATTTGAATC
  Consensus GCTG.ACCAA.TGCAAACTCAA.TGGTGCTTCAA.GGGAGAGACCCACTGACTCTCCTTCCTTTACTCTTATGCCATTGGTCCCATCATTCTTGTGGGGGAAAAATTCTAGTATTTTGATTATTTGAATC

            781     790       800       810       820       830       840       850       860       870       880       890       900       910
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_015696.4 TTACAGCAACAAATAGGAACTCCTGGCCAATGAGAGCTCTTGACCAGTGAATCACCAGCCGATACGAACGTCTTGCCAACAAAAATGTGTGGCAAATAGAAGTATATCAAGCAATAATCTCCCACCCAAG
     69BF16 TTACAGCAACAAATAGGAACTCCTGGCCAATGAGAGCTCTTGACCAGTGAATCACCAGCCGATACGAACGTCTTGCCAACAAAAATGTGTGGCAAATAGAAGTATATCAAGCAATAATCTCCCACCCAAG
```

Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

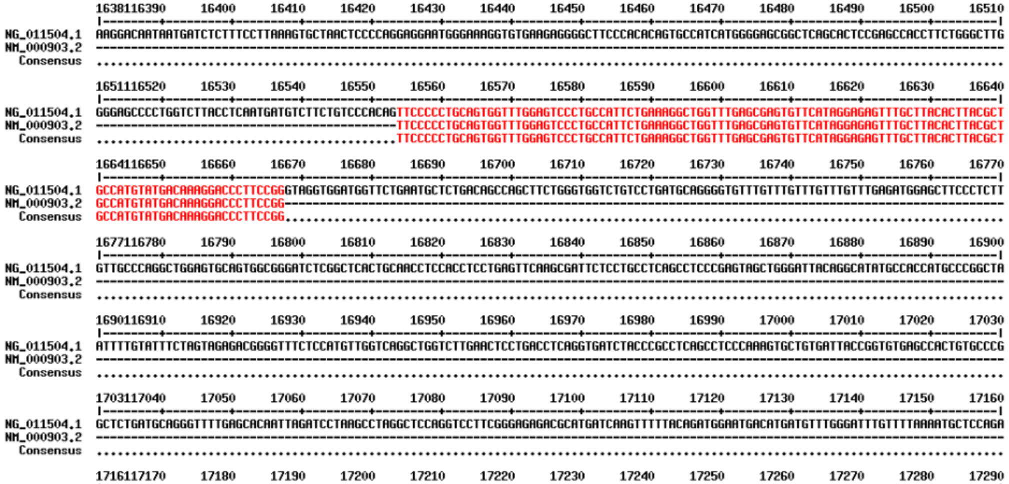
Např: multalin

- K porovnání sekvencí genomových a cDNA (mRNA)

Default:BLOSUM62

→DNA 5-0

```
            2002120030     20040     20050     20060     20070     20080     20090     20100     20110     20120
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
NG_011504.1 AAATGCACCCACGCAGGTCTTGGTGAGTCACCTGCCAAGAAACTGAGACT-CCCTAAAGTAACAGAACTTC-CAGCCTTCTTGGCAAAGGATCCAGGTTG
NM_000903.2 CTGGCTTATAAAGAAGGCCATC-TGAG-------CCCAGATATTGTGGCTGAACAAAAGAAGCTGGAAGCCGCAGAC--CTTGTGATA--TTCCAGTT--
  Consensus aaagccaacaAaGaAGGcCaTc.TGAG.......CCaAGAaAcTGaGaCT.aaCaAAAGaAaCaGaAagcC.CAGaC..CTTGgcAaA..aTCCAGgT..

            2015120160     20170     20180     20190     20200     20210     20220     20230     20240     20250
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
NG_011504.1 AGAACACACCTGAGAAGGCTAAAATTGGTAACGGCTAGGTAGAGGGTAAGAGAGAGACGCTAGCTCTGAACTGATTCTCTAGTGTGCCTGAGGCCTCCTT
NM_000903.2 ----CCATTCTGA-AAGGCTGGTTTGAGCGAGTGTTCATAGGAGAGTTTGCTT---ACACTTACGCTGCCATGTATGACAAAGGACCCT------TCCGG
  Consensus ....CaaacCTGA.AAGGCTaaaaTgaGcaAcgGcTaagaaGAGaGTaaGaga...ACaCTaaCgCTGaaaTGaaTcaCaAagGacCCT......TCCgg

            2028120290     20300     20310     20320     20330     20340     20350     20360     20370     20380
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
NG_011504.1 CAACTTCTGTTGTTTATAGTACAA-CTGCATGGAATTGGTTGACTTACCTCTCTGTGCTTTCTGTATCC---TCAGAGTGGCATTCTGCATTTCTGTGGC
NM_000903.2 TGGTGGCAGTGGCTCCATGTACTCTCTGCAAGGGATCCACGGG--GACATGAATGT-CATTCTCTGGCCAATTCAGAGTGGCATTCTGCATTTCTGTGGC
  Consensus caacggCaGTgGcTcaaaGTACaa.CTGCAaGGaATccacgGa..gACaTcaaTGT.CaTTCTcTagCC...TCAGAGTGGCATTCTGCATTTCTGTGGC

            2041120420     20430     20440     20450     20460     20470     20480     20490     20500     20510
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
NG_011504.1 AGCATTGGGCACACTCCAGCAGACGCCCGAATTCAAATCCTGGAAGGATGGAAGAAACGCCTGGAGAATATTTGGGATGAGACACCACTGTATTTTGCTC
NM_000903.2 AGCATTGGGCACACTCCAGCAGACGCCCGAATTCAAATCCTGGAAGGATGGAAGAAACGCCTGGAGAATATTTGGGATGAGACACCACTGTATTTTGCTC
  Consensus AGCATTGGGCACACTCCAGCAGACGCCCGAATTCAAATCCTGGAAGGATGGAAGAAACGCCTGGAGAATATTTGGGATGAGACACCACTGTATTTTGCTC

            2054120550     20560     20570     20580     20590     20600     20610     20620     20630     20640
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
NG_011504.1 CAGGATTCTTAATGAAAAAAGAGGTACAGGATGAGGAGAAAAACAAGAAATTTGGCCTTTCTGTGGGCCATCACTTGGGCAAGTCCATCCCAACTGACAA
NM_000903.2 CAGGATTCTTAATGAAAAAAGAGGTACAGGATGAGGAGAAAAACAAGAAATTTGGCCTTTCTGTGGGCCATCACTTGGGCAAGTCCATCCCAACTGACAA
  Consensus CAGGATTCTTAATGAAAAAAGAGGTACAGGATGAGGAGAAAAACAAGAAATTTGGCCTTTCTGTGGGCCATCACTTGGGCAAGTCCATCCCAACTGACAA

            2067120680     20690     20700     20710     20720     20730     20740     20750     20760     20770     20780     20790     20800
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 AGCCTGGATTTCCTTCTAACATGTTATCAAATCTGGGTATCTTTCCAGGCTTCCCTGACTTGCTTTAGTTTTTAAGATTTGTGTTTTTCTTTTTCCACAAGGAATAAATGAGAGGGAATCGACTGTATTC
NM_000903.2 AGCCTGGATTTCCTTCTAACATGTTATCAAATCTGGGTATCTTTCCAGGCTTCCCTGACTTGCTTTAGTTTTTAAGATTTGTGTTTTTCTTTTTCCACAAGGAATAAATGAGAGGGAATCGACTGTATTC
  Consensus AGCCTGGATTTCCTTCTAACATGTTATCAAATCTGGGTATCTTTCCAGGCTTCCCTGACTTGCTTTAGTTTTTAAGATTTGTGTTTTTCTTTTTCCACAAGGAATAAATGAGAGGGAATCGACTGTATTC
```



```
            1638116390     16400     16410     16420     16430     16440     16450     16460     16470     16480     16490     16500     16510
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 AAGGACAATAATGATCTCTTTCCTTAAAGTGCTAACTCCCCAGGAGGAATGGGAAAGGTGTGAAGAGGGGCTTCCCACACAGTGCCATCATGGGGAGCGGCTCAGCACTCCGAGCCACCTTCTGGGCTTG
NM_000903.2 ----------------------------------------------------------------------------------------------------------------------------------
  Consensus ..................................................................................................................................

            1651116520     16530     16540     16550     16560     16570     16580     16590     16600     16610     16620     16630     16640
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 GGGAGCCCCTGGTCTTACCTCAATGATGTCTTCTGTCCCACAGTTCCCCCTGCAGTGGTTTGGAGTCCCTGCCATTCTGAAAGGCTGGTTTGAGCGAGTGTTCATAGGAGAGTTTGCTTACACTTACGCT
NM_000903.2 -------------------------------------------TTCCCCCTGCAGTGGTTTGGAGTCCCTGCCATTCTGAAAGGCTGGTTTGAGCGAGTGTTCATAGGAGAGTTTGCTTACACTTACGCT
  Consensus ...........................................TTCCCCCTGCAGTGGTTTGGAGTCCCTGCCATTCTGAAAGGCTGGTTTGAGCGAGTGTTCATAGGAGAGTTTGCTTACACTTACGCT

            1664116650     16660     16670     16680     16690     16700     16710     16720     16730     16740     16750     16760     16770
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 GCCATGTATGACAAAGGACCCTTCCGGGTAGGTGGATGGTTCTGAATGCTCTGACAGCCAGCTTCTGGGTGGTCTGTCCTGATGCAGGGGTGTTTGTTTGTTTGTTTGTTTGAGATGGAGCTTCCCTCTT
NM_000903.2 GCCATGTATGACAAAGGACCCTTCCGG-------------------------------------------------------------------------------------------------------
  Consensus GCCATGTATGACAAAGGACCCTTCCGG.......................................................................................................

            1677116780     16790     16800     16810     16820     16830     16840     16850     16860     16870     16880     16890     16900
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 GTTGCCCAGGCTGGAGTGCAGTGGCGGGATCTCGGCTCACTGCAACCTCCACCTCCTGAGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATATGCCACCATGCCCGGCTA
NM_000903.2 ----------------------------------------------------------------------------------------------------------------------------------
  Consensus ..................................................................................................................................

            1690116910     16920     16930     16940     16950     16960     16970     16980     16990     17000     17010     17020     17030
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 ATTTTGTATTTCTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCTACCCGCCTCAGCCTCCCAAAGTGCTGTGATTACCGGTGTGAGCCACTGTGCCCG
NM_000903.2 ----------------------------------------------------------------------------------------------------------------------------------
  Consensus ..................................................................................................................................

            1703117040     17050     17060     17070     17080     17090     17100     17110     17120     17130     17140     17150     17160
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 GCTCTGATGCAGGGTTTTGAGCACAATTAGATCCTAAGCCTAGGCTCCAGGTCCTTCGGGAGAGACGCATGATCAAGTTTTTACAGATGGAATGACATGATGTTTGGGATTTGTTTTAAAATGCTCCAGA
NM_000903.2 ----------------------------------------------------------------------------------------------------------------------------------
  Consensus ..................................................................................................................................

            1716117170     17180     17190     17200     17210     17220     17230     17240     17250     17260     17270     17280     17290
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NG_011504.1 GGCCGGGCGCAGTGGCTCACGCCTGTAATTTCAGCACTTTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAGTCTGGCCAACATGGCGAAACCCCGTCTCTACCAAAAATACAA
NM_000903.2 ----------------------------------------------------------------------------------------------------------------------------------
  Consensus ..................................................................................................................................
```

Vyzkoušejte si porovnání

1) vyzkoušejte si porovnat sekvence v Př0.

- 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)

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

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 **TT**T AA

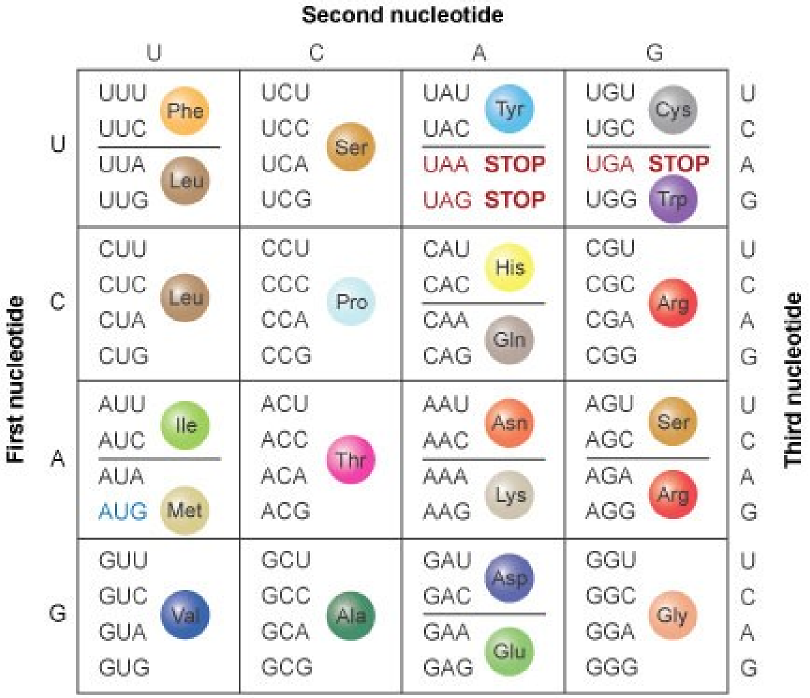
AT **GGA** AGT ATT TAA AGC GCC ACC TAT TTA A

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



	Second nucleotide				
First nucleotide	U	C	A	G	Third nucleotide
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA STOP	UGA STOP	A
	UUG	UCG	UAG STOP	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met	ACG	AAG	AGG	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu	GGA	A
	GUG	GCG	GAG	GGG	G

„čtení“ DNA = translace/překlad

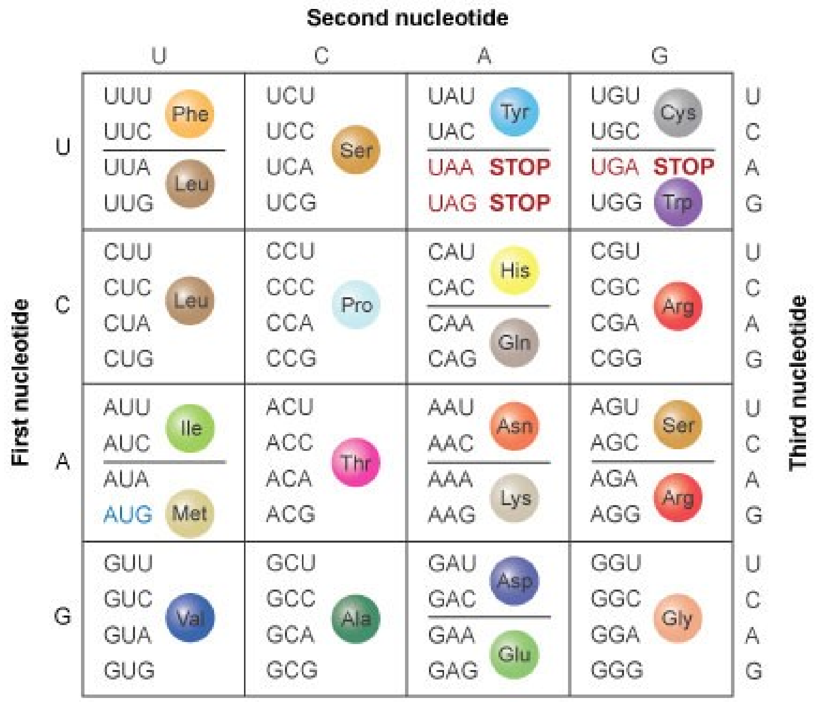
DNA sekvence psaná v orientaci 5' - 3', vznikne aminokyselinová sekvence v orientaci od **N-** k **C-** konci.

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
```



Second nucleotide

First nucleotide	U	C	A	G	Third nucleotide
U	UUU UUC Phe UUA UUG Leu	UCU UCC UCA UCG Ser	UAU UAC Tyr UAA STOP UAG STOP	UGU UGC Cys UGA STOP UGG Trp	U C A G
C	CUU CUC CUA CUG Leu	CCU CCC CCA CCG Pro	CAU CAC His CAA CAG Gln	CGU CGC CGA CGG Arg	U C A G
A	AUU AUC AUA Ile AUG Met	ACU ACC ACA ACG Thr	AAU AAC Asn AAA AAG Lys	AGU AGC Ser AGA AGG Arg	U C A G
G	GUU GUC GUA GUG Val	GCU GCC GCA GCG Ala	GAU GAC Asp GAA GAG Glu	GGU GGC GGA GGG Gly	U C A G

„čtení“ DNA = translace/překlad

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

5´- ATGGAAGTATTTAAAGCGCCACCTATTTAA-3´

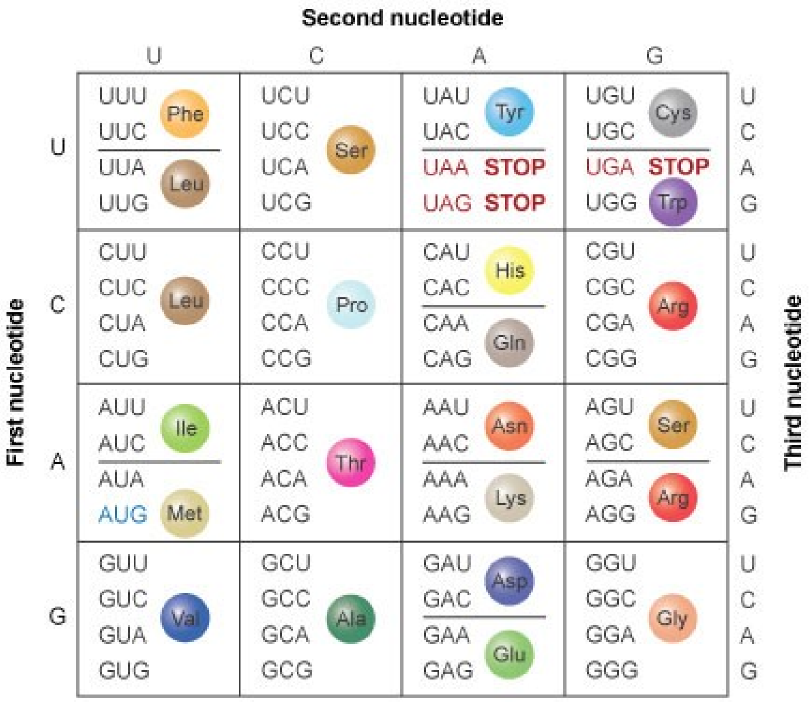
ATG GAA GTA TTT AAA GCG CCA CCT ATT TAA
A **TGG** AAG **TAT** TTA **AAG** CGC **CAC** CTA **TTT** AA
AT **GGA** AGT ATT TAA AGC GCC ACC TAT TTA A

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

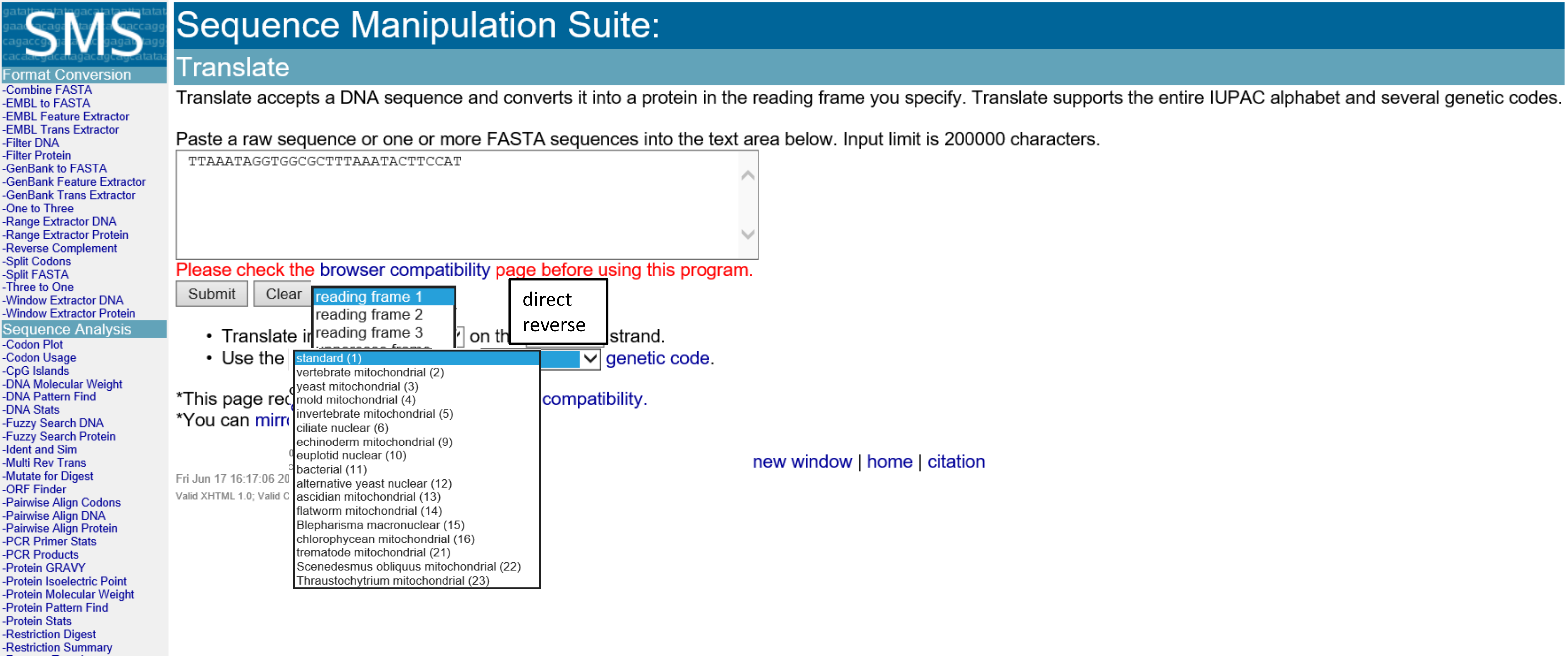


Second nucleotide

First nucleotide	U	C	A	G	Third nucleotide
U	UUU UUC Phe UUA UUG Leu	UCU UCC UCA UCG Ser	UAU UAC Tyr UAA STOP UAG STOP	UGU UGC Cys UGA STOP UGG Trp	U C A G
C	CUU CUC CUA CUG Leu	CCU CCC CCA CCG Pro	CAU CAC His CAA CAG Gln	CGU CGC CGA CGG Arg	U C A G
A	AUU AUC AUA Ile AUG Met	ACU ACC ACA ACG Thr	AAU AAC Asn AAA AAG Lys	AGU AGC Ser AGA AGG Arg	U C A G
G	GUU GUC GUA GUG Val	GCU GCC GCA GCG Ala	GAU GAC Asp GAA GAG Glu	GGU GGC GGA GGG Gly	U C A G

„čtení“ DNA = translace/překlad

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



Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

```
TTAAATAGGTGGCGCTTTAAATACTTCCAT
```

Please check the browser compatibility page before using this program.

- Translate i on th strand.
- Use the genetic code.

*This page req compatibility.
*You can mirr

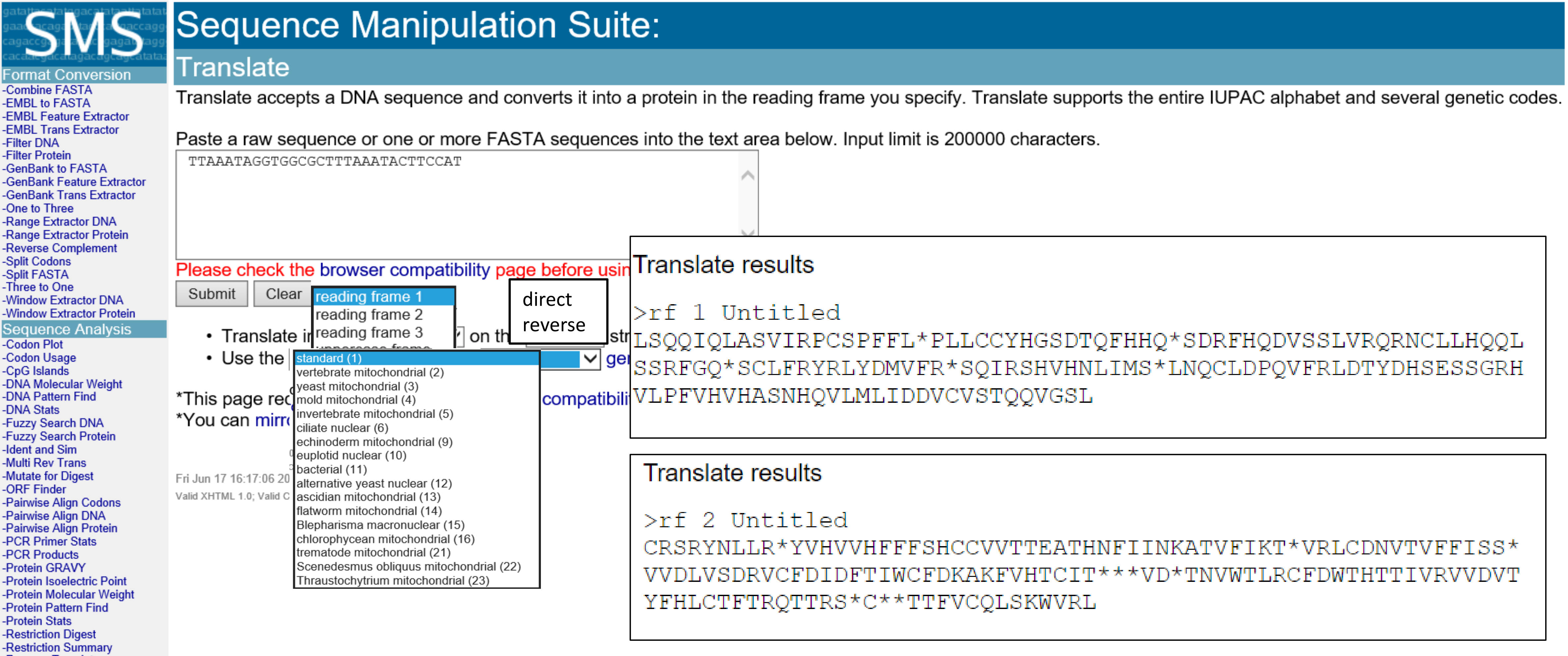
new window | home | citation

Fri Jun 17 16:17:06 20

Valid XHTML 1.0; Valid C

„čtení“ DNA = translace/překlad

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



Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

```
TTAAATAGGTGGCGCTTTAAATACTTCCAT
```

Please check the browser compatibility page before usin

- Translate i on th str
- Use the ge

*This page re compatibili

*You can mirr

Fri Jun 17 16:17:06 20

Valid XHTML 1.0; Valid C

Translate results

```
>rf 1 Untitled
LSQQIQLASVIRPCSPFFL*PLLCCYHGSDTQFHHQ*SDRFHQDVSSLVRQRNCLLHQQL
SSRFGQ*SCLFRYRLYDMVFR*SQIRSHVHNLIMS*LNQCLDPQVFRLDTYDHSESSGRH
VLPFVHVHASNHQVLMLIDDVCVSTQQVGSL
```

Translate results

```
>rf 2 Untitled
CRSRYNLLR*YVHVVHFFFSHCCVVTTEATHNFIINKATVFIKT*VRLCDNVTVFFISS*
VVDLVSDRVCFDIDFTIWCFDKAKFVHTCIT***VD*TNVWTLRCFDWTHTTIVRVVDVT
YFHLCTFTRQTTRS*C**TTFVCQLSKWVRL
```

vyzkoušejte si

zkuste si v programu SMS/translate přeložit

- „celou“ nukleotidovou sekvenci (mRNA)

ve frame „1“ direct

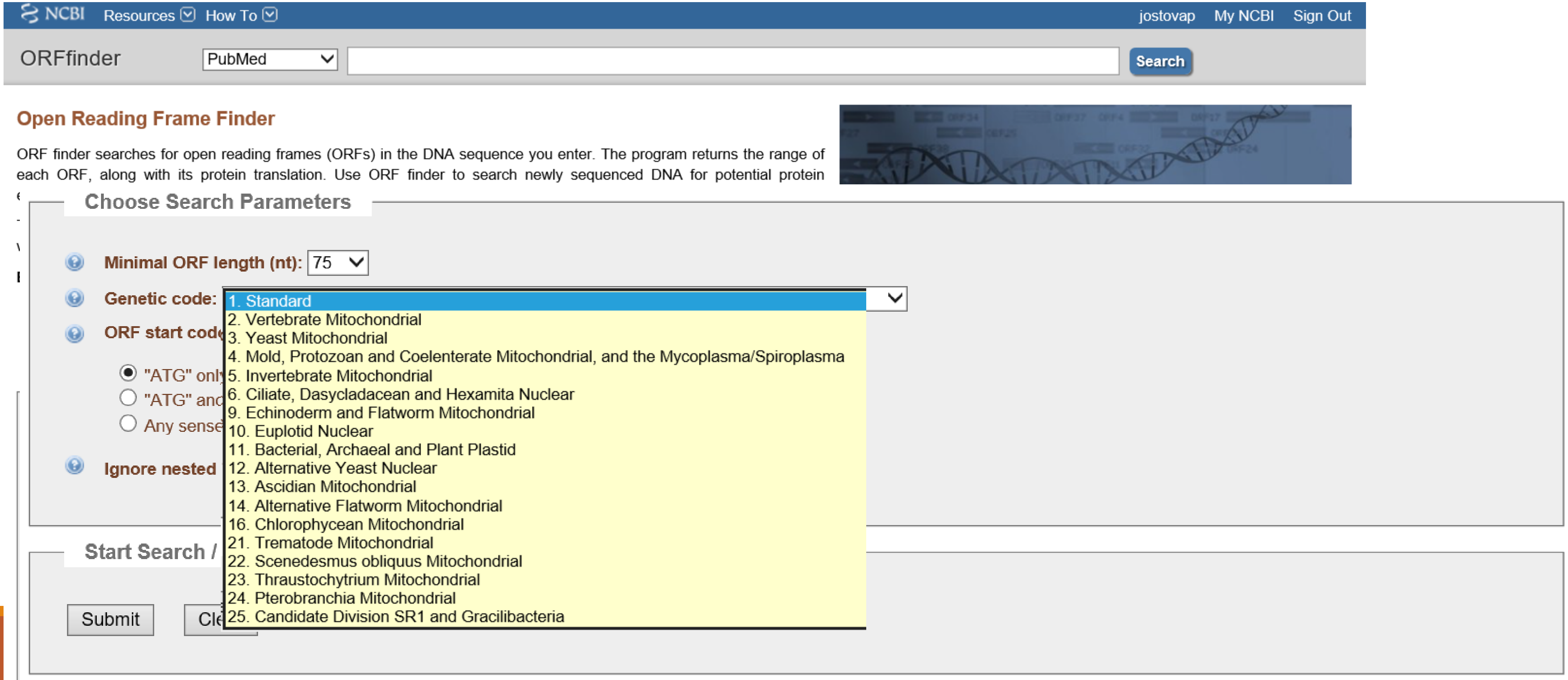
– dokážete určit zda je to „správný čtecí rámec“ pro vaši sekvenci?

zkuste si v programu SMS/translate přeložit

- jen CDS

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“



Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code:

1. Standard
2. Vertebrate Mitochondrial
3. Yeast Mitochondrial
4. Mold, Protozoan and Coelenterate Mitochondrial, and the Mycoplasma/Spiroplasma
5. Invertebrate Mitochondrial
6. Ciliate, Dasycladacean and Hexamita Nuclear
9. Echinoderm and Flatworm Mitochondrial
10. Euplotid Nuclear
11. Bacterial, Archaeal and Plant Plastid
12. Alternative Yeast Nuclear
13. Ascidian Mitochondrial
14. Alternative Flatworm Mitochondrial
16. Chlorophycean Mitochondrial
21. Trematode Mitochondrial
22. Scenedesmus obliquus Mitochondrial
23. Thraustochytrium Mitochondrial
24. Pterobranchia Mitochondrial
25. Candidate Division SR1 and Gracilibacteria

ORF start cod

● "ATG" onl

○ "ATG" and

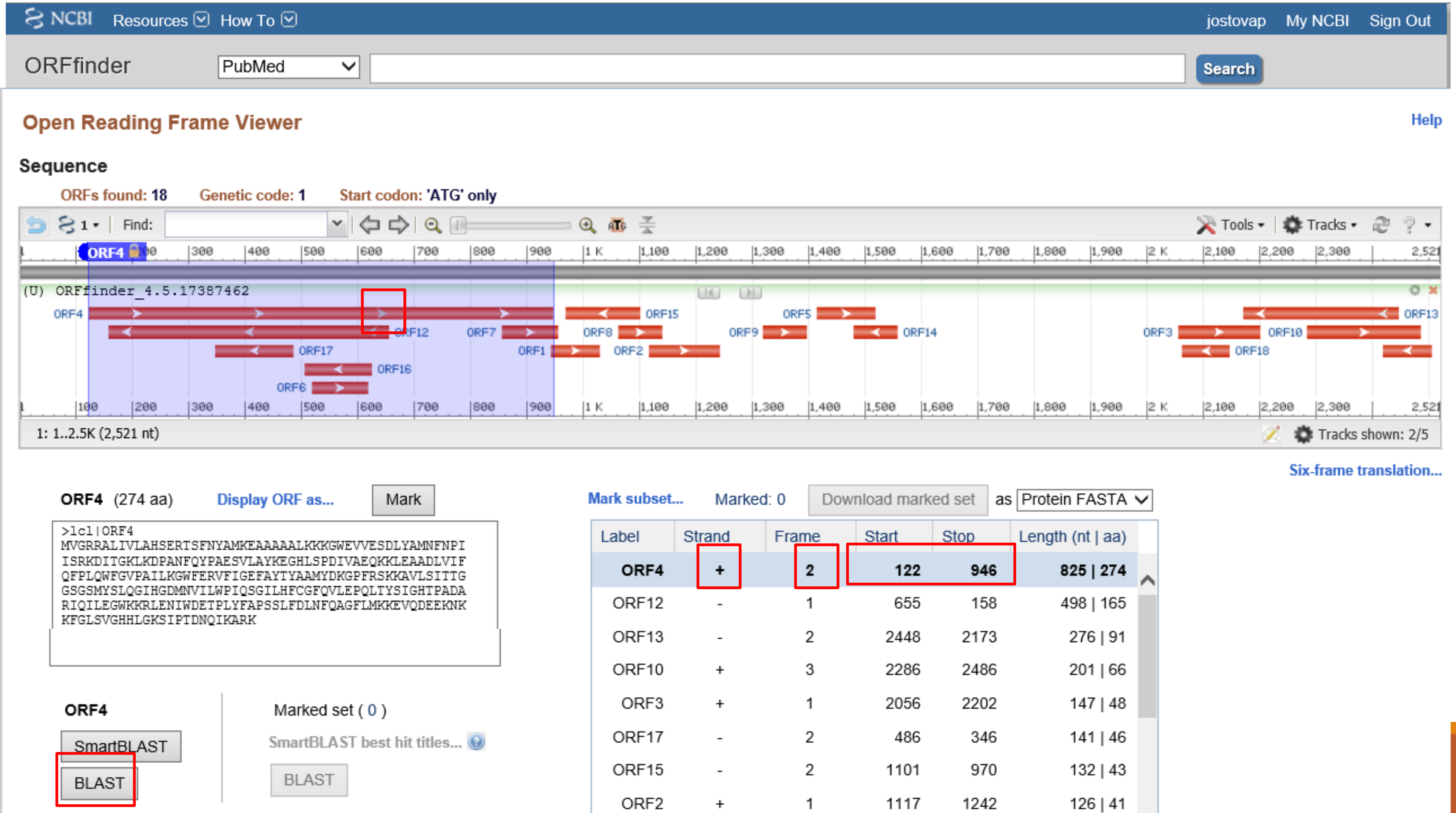
○ Any sense

Ignore nested

Start Search /

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“



Open Reading Frame Viewer

Sequence

ORFs found: 18 Genetic code: 1 Start codon: 'ATG' only

1: 1..2.5K (2,521 nt)

ORF4 (274 aa)

```
>lcl|ORF4
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPI
ISRKDITGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIF
QFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTG
GSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADA
RIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNK
KFGLSVGHHLGKSIPTDNQIKARK
```

Marked: 0 as Protein FASTA

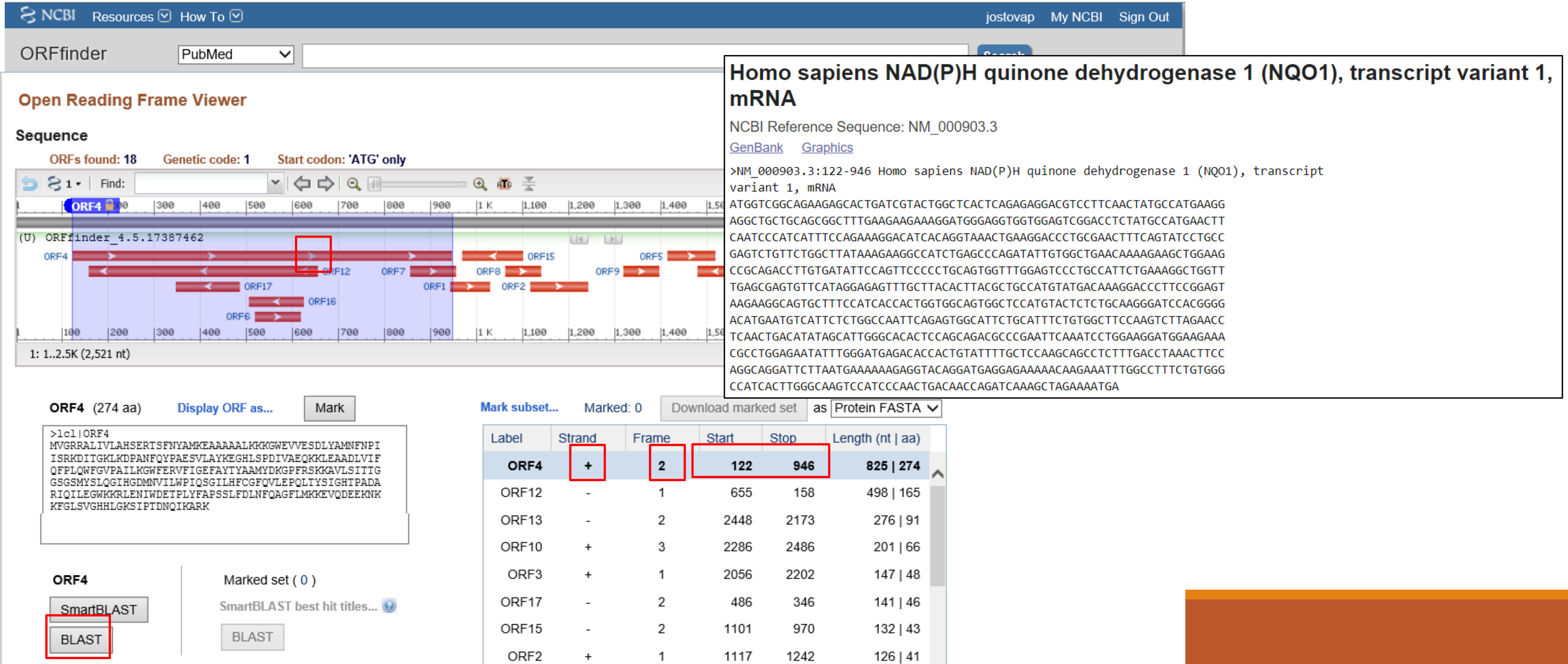
Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF4	**+**	**2**	**122**	**946**	**825 | 274**
ORF12	-	1	655	158	498 | 165
ORF13	-	2	2448	2173	276 | 91
ORF10	+	3	2286	2486	201 | 66
ORF3	+	1	2056	2202	147 | 48
ORF17	-	2	486	346	141 | 46
ORF15	-	2	1101	970	132 | 43
ORF2	+	1	1117	1242	126 | 41

ORF4

Marked set (0)

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“



ORFs found: 18 Genetic code: 1 Start codon: 'ATG' only

ORF4 (274 aa)

```
>lcl|ORF4
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPI
ISRKDITGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIF
QFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTG
GSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADA
RIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNK
KFGLSVGHHLGKSIPTDNQIKARK
```

Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF4	+	2	122	946	825 | 274
ORF12	-	1	655	158	498 | 165
ORF13	-	2	2448	2173	276 | 91
ORF10	+	3	2286	2486	201 | 66
ORF3	+	1	2056	2202	147 | 48
ORF17	-	2	486	346	141 | 46
ORF15	-	2	1101	970	132 | 43
ORF2	+	1	1117	1242	126 | 41

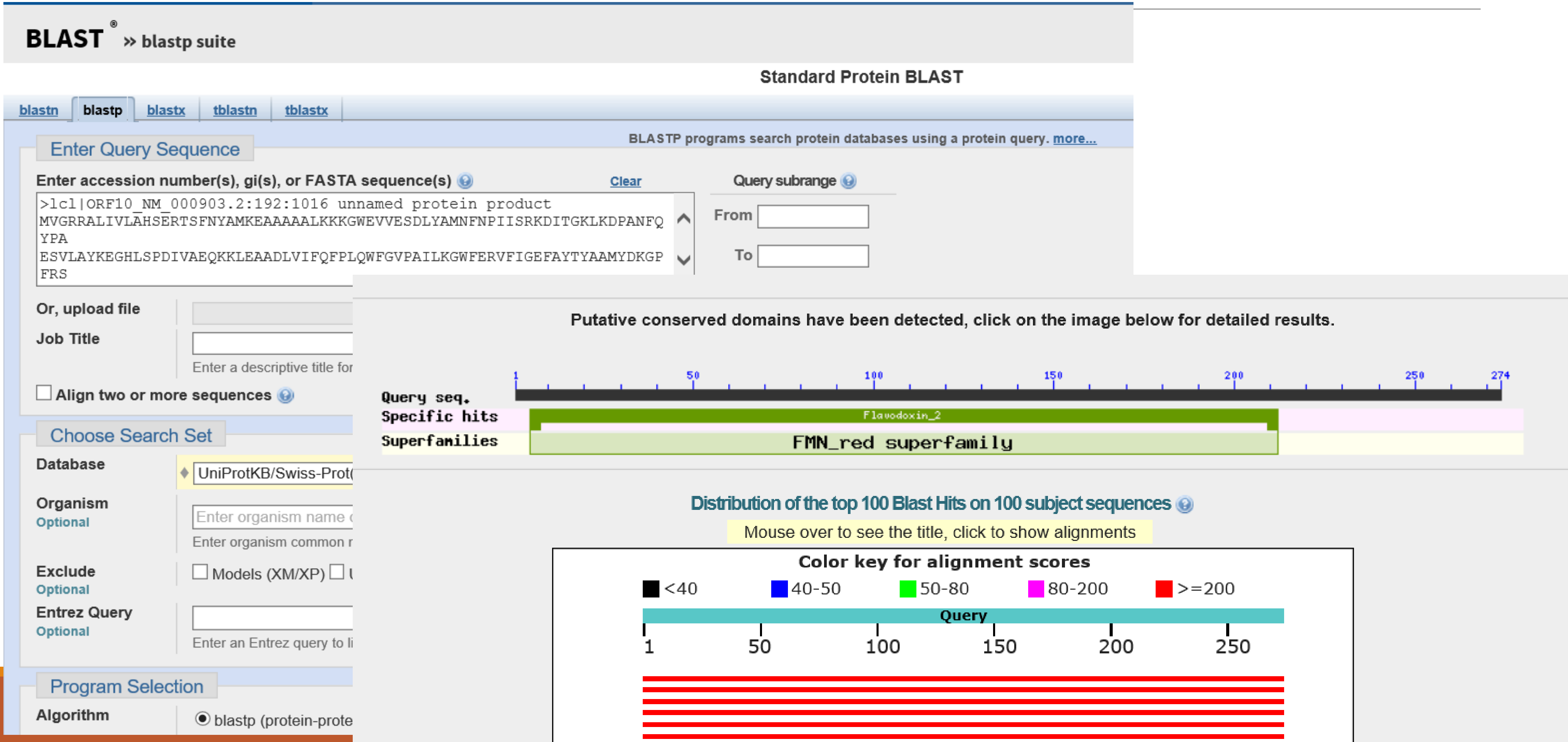
Homo sapiens NAD(P)H quinone dehydrogenase 1 (NQO1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000903.3

GenBank Graphics

```
>NM_000903.3:122-946 Homo sapiens NAD(P)H quinone dehydrogenase 1 (NQO1), transcript
variant 1, mRNA
ATGGTCGGCAGAAGAGCACTGATCGTACTGGCTCACTCAGAGAGGACGTCCTTCAACTATGCCATGAAGG
AGGCTGCTGCAGCGGCTTTGAAGAAGAAAGGATGGGAGGTGGTGGAGTCGGACCTCTATGCCATGAACTT
CAATCCCATCATTTCCAGAAAGGACATCACAGGTAAACTGAAGGACCCTGCGAACTTTCAGTATCCTGCC
GAGTCTGTTCTGGCTTATAAAGAAGGCCATCTGAGCCCAGATATTGTGGCTGAACAAAAGAAGCTGGAAG
CCGCAGACCTTGTGATATTCCAGTTCCCCCTGCAGTGGTTTGGAGTCCCTGCCATTCTGAAAGGCTGGTT
TGAGCGAGTGTTCATAGGAGAGTTTGCTTACACTTACGCTGCCATGTATGACAAAGGACCCTTCCGGAGT
AAGAAGGCAGTGCTTTCCATCACCACTGGTGGCAGTGGCTCCATGTACTCTCTGCAAGGGATCCACGGGG
ACATGAATGTCATTCTCTGGCCAATTCAGAGTGGCATTCTGCATTTCTGTGGCTTCCAAGTCTTAGAACC
TCAACTGACATATAGCATTGGGCACACTCCAGCAGACGCCCGAATTCAAATCCTGGAAGGATGGAAGAAA
CGCCTGGAGAATATTTGGGATGAGACACCACTGTATTTTGCTCCAAGCAGCCTCTTTGACCTAAACTTCC
AGGCAGGATTCTTAATGAAAAAAGAGGTACAGGATGAGGAGAAAAACAAGAAATTTGGCCTTTCTGTGGG
CCATCACTTGGGCAAGTCCATCCCAACTGACAACCAGATCAAAGCTAGAAAATGA
```

„čtení“ DNA = translace/překlad



BLAST® » blastp suite

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>lcl|ORF10_NM_000903.2:192:1016 unnamed protein product
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQ
YPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGP
FRS
```

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for

☐ Align two or more sequences

Choose Search Set

Database UniProtKB/Swiss-Prot(

Organism Optional

Enter organism name o

Enter organism common r

Exclude Optional ☐ Models (XM/XP) ☐ U

Entrez Query Optional

Enter an Entrez query to li

Program Selection

Algorithm ● blastp (protein-prote

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

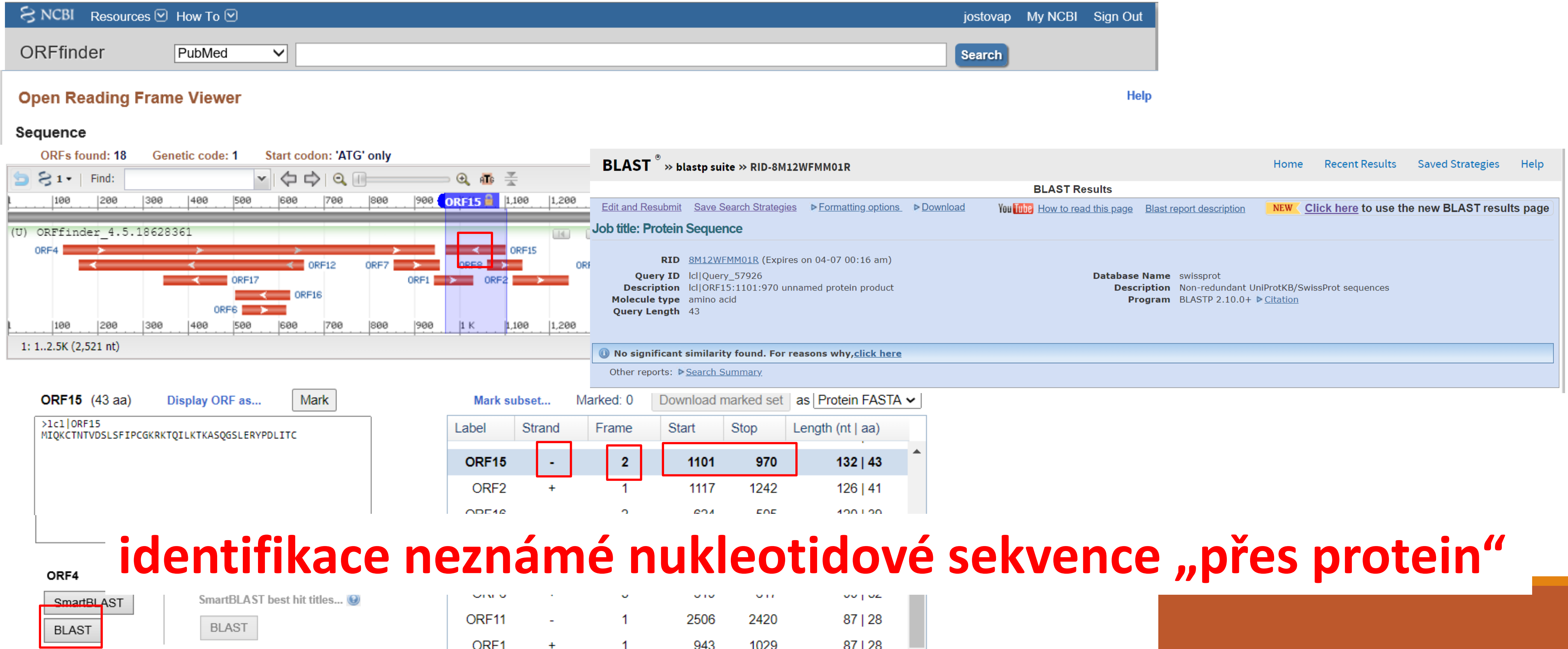
Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40 🟦 40-50 🟩 50-80 🟪 80-200 🟥 >=200

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“



Open Reading Frame Viewer

Sequence

ORFs found: 18 Genetic code: 1 Start codon: 'ATG' only

(U) ORFfinder_4.5.18628361

ORF15 (43 aa) Display ORF as...

```
>lcl|ORF15
MIQKCTNTVDSLSFIPCGKRKTQILKTKASQGSLERYPDLITC
```

ORF4

BLAST » blastp suite » RID-8M12WFMM01R

BLAST Results

Click here to use the new BLAST results page

Job title: Protein Sequence

RID	8M12WFMM01R (Expires on 04-07 00:16 am)		
Query ID	lcl|Query_57926	**Database Name**	swissprot
Description	lcl|ORF15:1101:970 unnamed protein product	**Description**	Non-redundant UniProtKB/SwissProt sequences
Molecule type	amino acid	**Program**	BLASTP 2.10.0+
Query Length	43		

No significant similarity found. For reasons why, click here

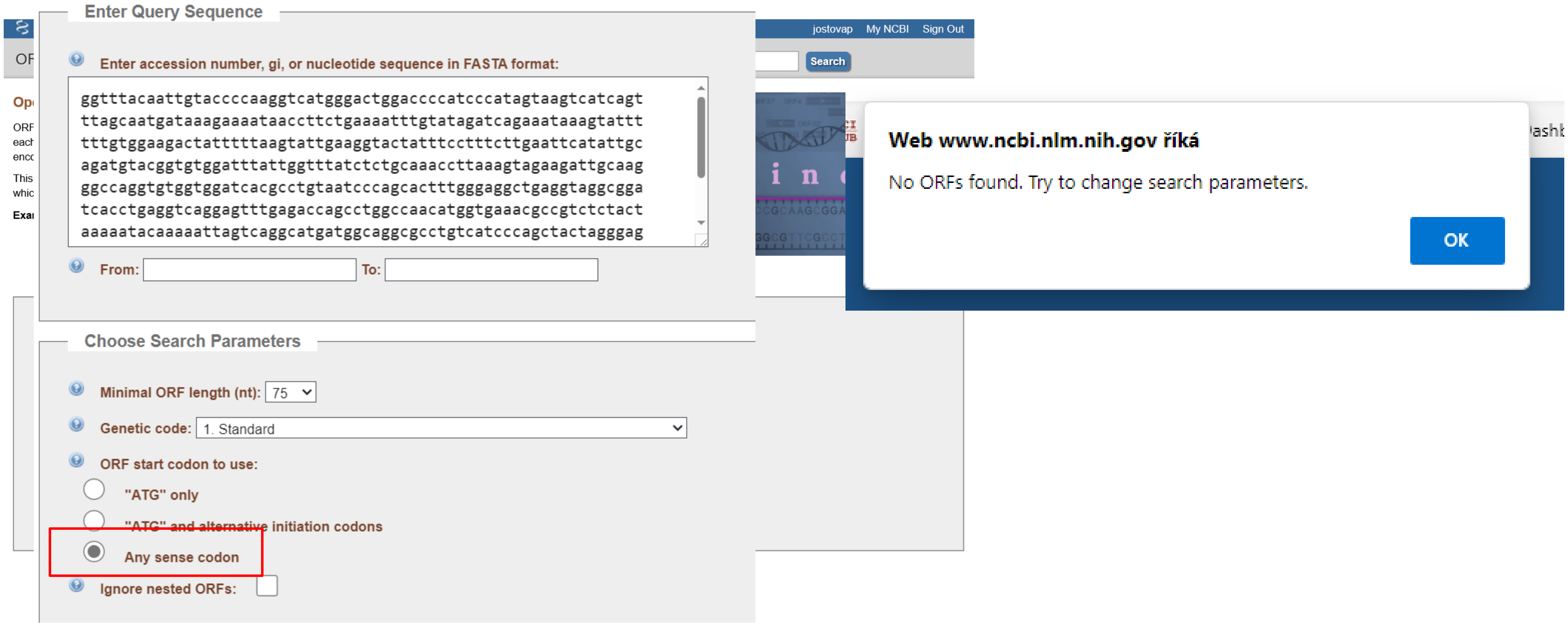
Other reports: Search Summary

Marked: 0

Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF15	**-**	**2**	**1101**	**970**	**132 | 43**
ORF2	+	1	1117	1242	126 | 41
ORF11	-	1	2506	2420	87 | 28
ORF1	+	1	943	1029	87 | 28

identifikace neznámé nukleotidové sekvence „přes protein“

„čtení“ DNA = translace/překlad



Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
ggtttacaattgtaccccaaggtcatgggactggaccccatcccatagtaagtcatcagt
ttagcaatgataaagaaaataaccttctgaaaatttgtatagatcagaaataaagtattt
tttgtggaagactatttttaagtattgaaggtactatttcctttcttgaattcatattgc
agatgtacggtgtggatttattggtttatctctgcaaaccttaaagtagaagattgcaag
ggccaggtgtggtggatcacgcctgtaatcccagcactttgggaggctgaggtaggcgga
tcacctgaggtcaggagtttgagaccagcctggccaacatggtgaaacgccgtctctact
aaaaatacaaaaattagtcaggcatgatggcaggcgcctgtcatcccagctactagggag
```

From: To:

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code: 1. Standard

ORF start codon to use:

○ "ATG" only

○ "ATG" and alternative initiation codons

● Any sense codon

Ignore nested ORFs: ☐

Web www.ncbi.nlm.nih.gov říká

No ORFs found. Try to change search parameters.

Vyzkoušejte si „překládání“ nt sekvence

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

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

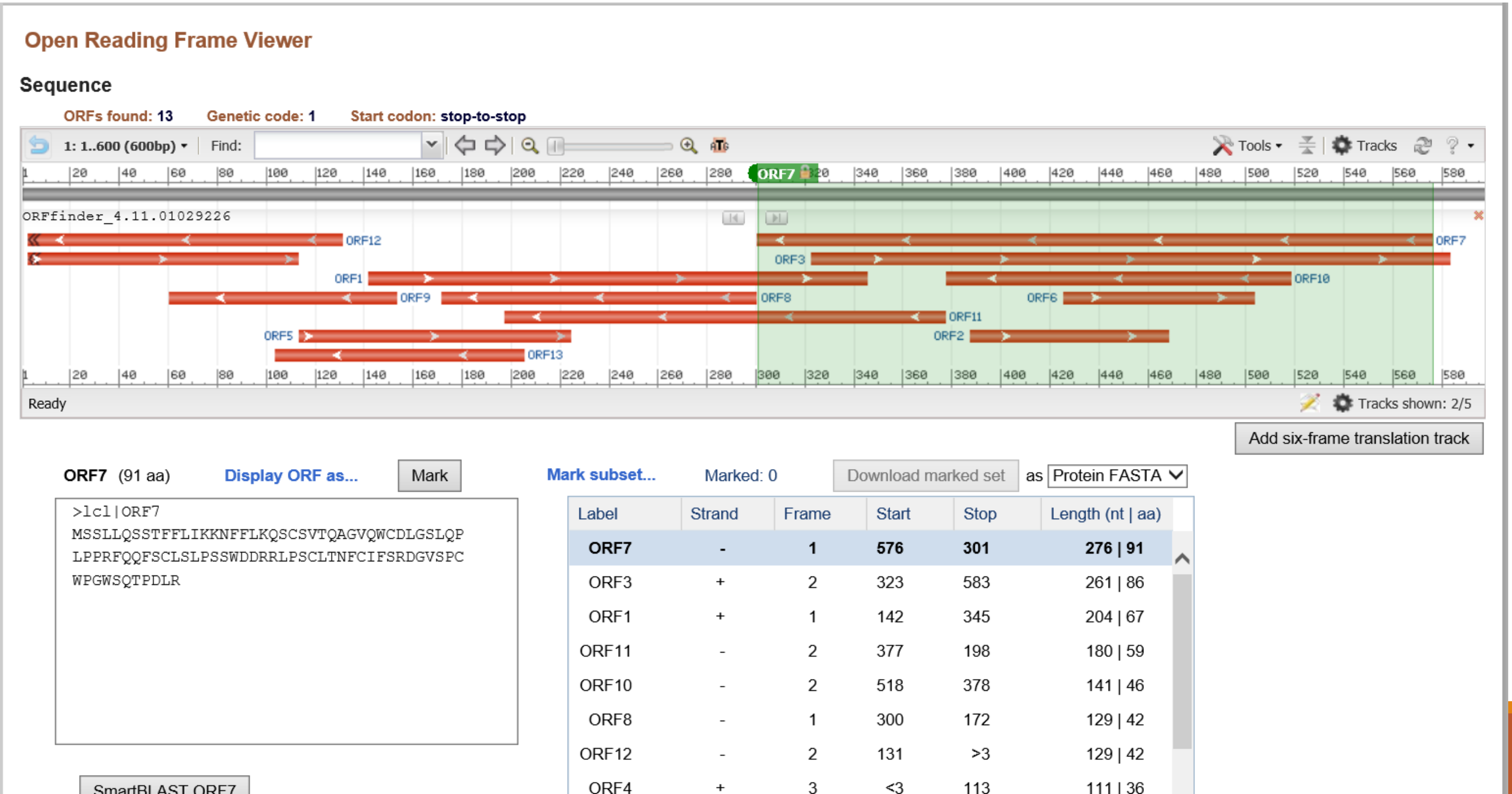
-najděte ve kterém čtecím se otevírá „ Př1 neznámá sekvence“

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

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

Identifikace **neznámé** sekvence

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



ORFs found: 13 Genetic code: 1 Start codon: stop-to-stop

ORF7 (91 aa)

```
>lcl|ORF7
MSSLLQSSTFFLIKKNFFLKQSCSVTQAGVQWCDLGSLQP
LPPRFQQFSCLSLPSSWDDRRLPSCLTNFCIFSRDGVSPC
WPGWSQTPDLR
```

Marked: 0

as Protein FASTA

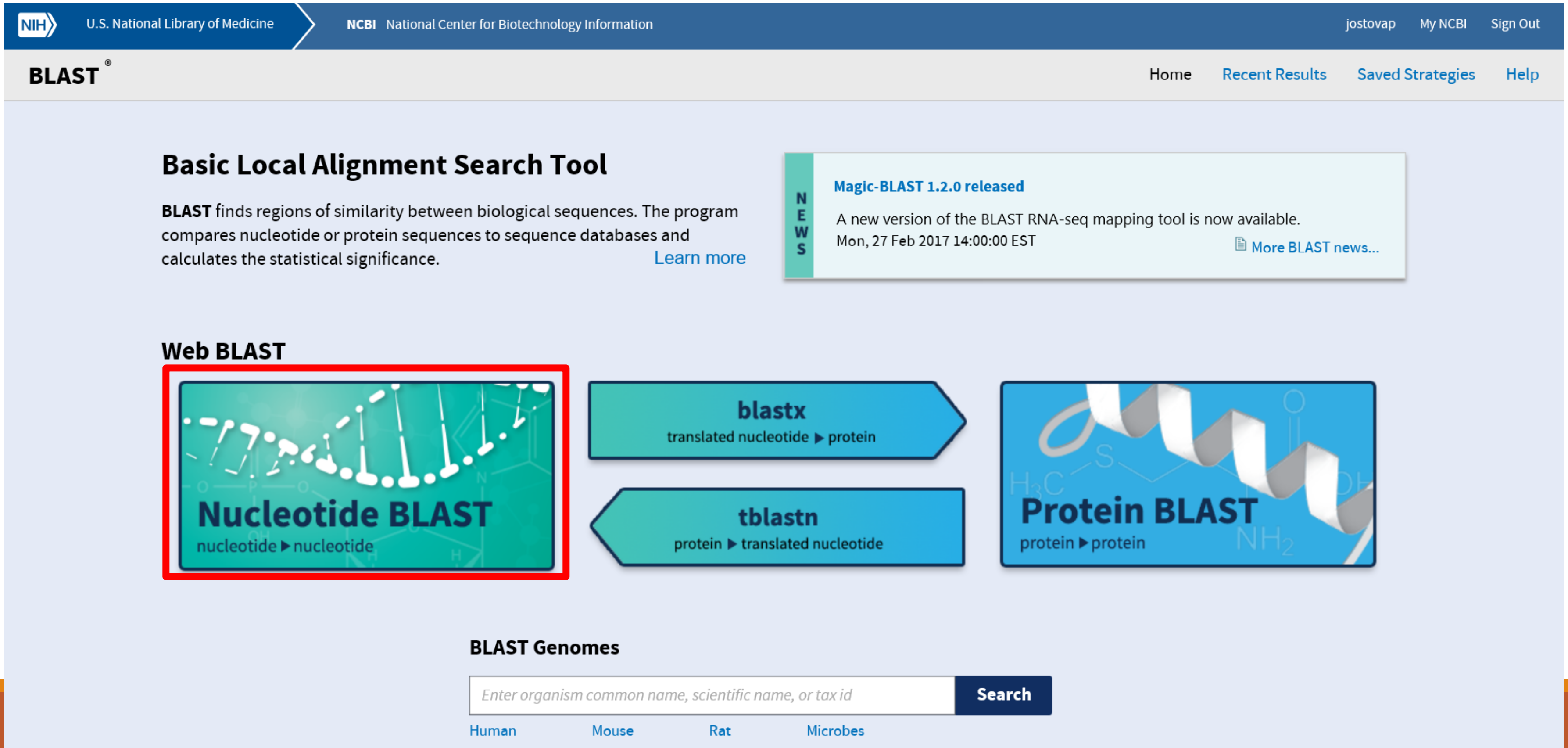
Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF7	**-**	**1**	**576**	**301**	**276 | 91**
ORF3	+	2	323	583	261 | 86
ORF1	+	1	142	345	204 | 67
ORF11	-	2	377	198	180 | 59
ORF10	-	2	518	378	141 | 46
ORF8	-	1	300	172	129 | 42
ORF12	-	2	131	>3	129 | 42
ORF4	+	3	<3	113	111 | 36

Identifikace **neznámé** sekvence

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

→ vyhledání příbuzných sekvencí

BLAST!



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS

Magic-BLAST 1.2.0 released

A new version of the BLAST RNA-seq mapping tool is now available.
Mon, 27 Feb 2017 14:00:00 EST

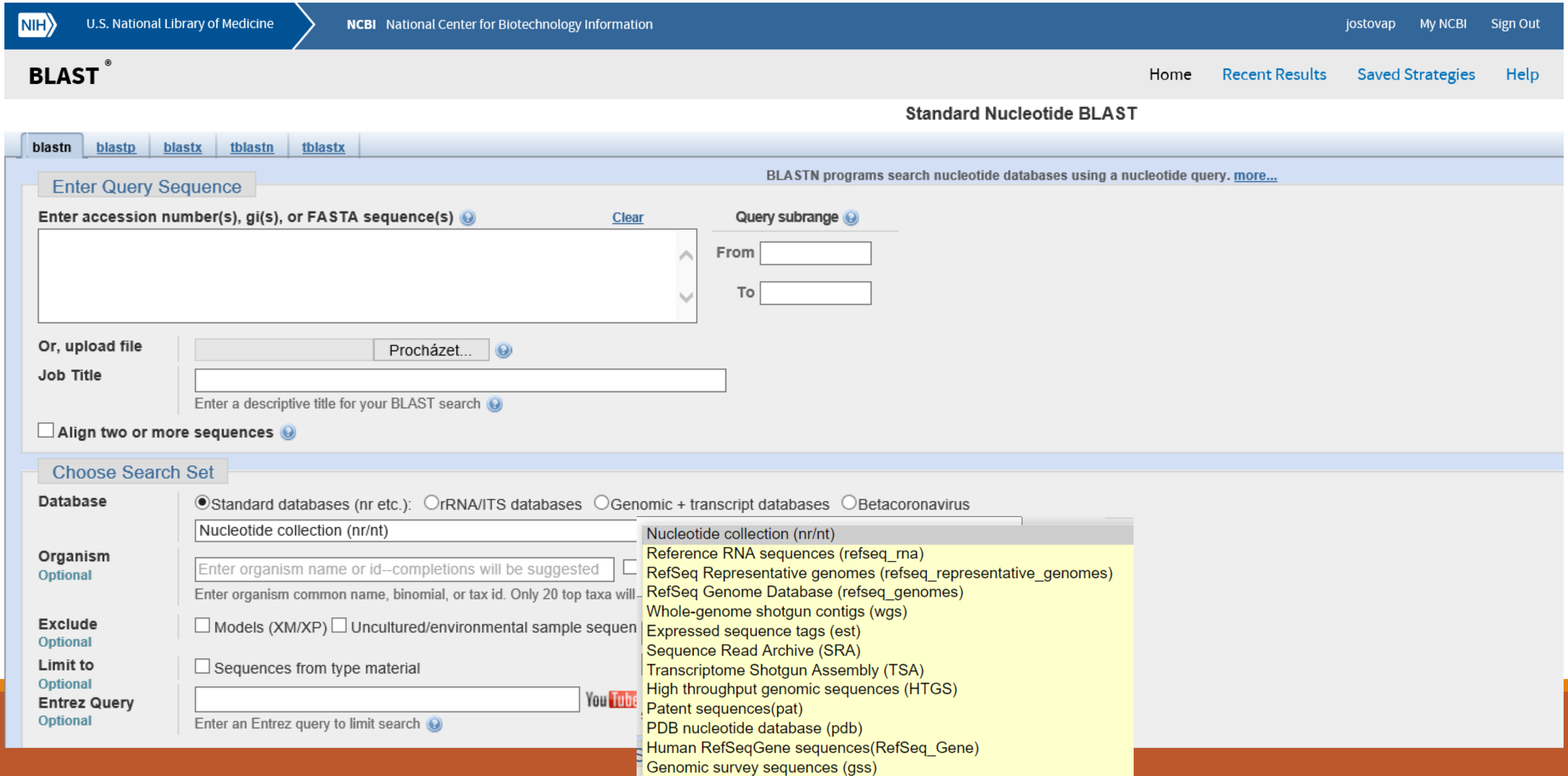
More BLAST news...

Identifikace **neznámé** sekvence

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

→ vyhledání příbuzných sekvencí

BLAST!



Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

Nucleotide collection (nr/nt)

- Nucleotide collection (nr/nt)
- Reference RNA sequences (refseq_rna)
- RefSeq Representative genomes (refseq_representative_genomes)
- RefSeq Genome Database (refseq_genomes)
- Whole-genome shotgun contigs (wgs)
- Expressed sequence tags (est)
- Sequence Read Archive (SRA)
- Transcriptome Shotgun Assembly (TSA)
- High throughput genomic sequences (HTGS)
- Patent sequences(pat)
- PDB nucleotide database (pdb)
- Human RefSeqGene sequences(RefSeq_Gene)
- Genomic survey sequences (gss)

Organism
Optional

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will

Exclude
Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequen

Limit to
Optional

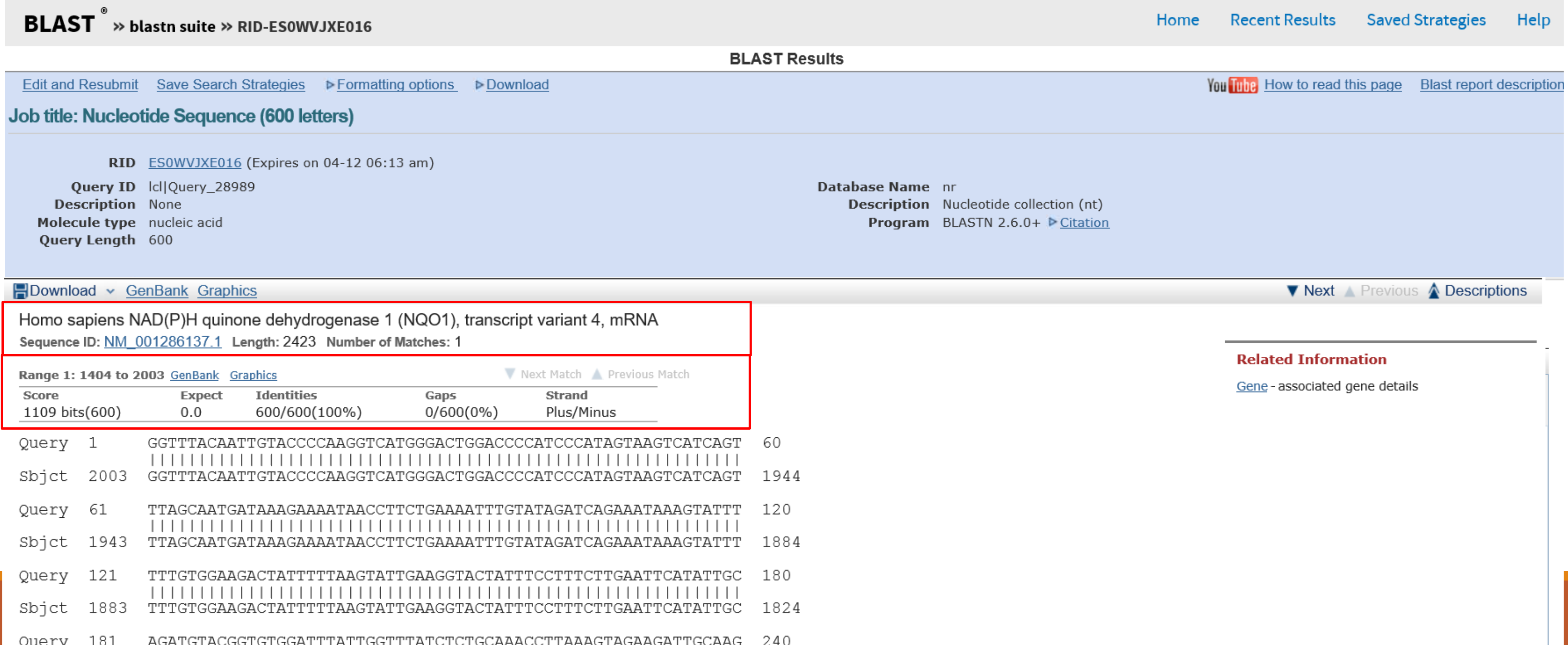
☐ Sequences from type material

Entrez Query
Optional

Enter an Entrez query to limit search

Identifikace **neznámé** sekvence

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



Job title: Nucleotide Sequence (600 letters)

RID	ES0WVJXE016 (Expires on 04-12 06:13 am)		
Query ID	lcl|Query_28989	**Database Name**	nr
Description	None	**Description**	Nucleotide collection (nt)
Molecule type	nucleic acid	**Program**	BLASTN 2.6.0+ Citation
Query Length	600		

Homo sapiens NAD(P)H quinone dehydrogenase 1 (NQO1), transcript variant 4, mRNA

Sequence ID: NM_001286137.1 **Length:** 2423 **Number of Matches:** 1

Range 1: 1404 to 2003

Score	Expect	Identities	Gaps	Strand
1109 bits(600)	0.0	600/600(100%)	0/600(0%)	Plus/Minus

```
Query  1     GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2003  GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGT  1944

Query  61    TTAGCAATGATAAAGAAAATAACCTTCTGAAAATTTGTATAGATCAGAAATAAAGTATTT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1943  TTAGCAATGATAAAGAAAATAACCTTCTGAAAATTTGTATAGATCAGAAATAAAGTATTT  1884

Query  121   TTTGTGGAAGACTATTTTTAAGTATTGAAGGTACTATTTCCTTTCTTGAATTCATATTGC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1883  TTTGTGGAAGACTATTTTTAAGTATTGAAGGTACTATTTCCTTTCTTGAATTCATATTGC  1824

Query  181   AGATGTACGGTGTGGATTTATTGGTTTATCTCTGCAAACCTTAAAGTAGAAGATTGCAAG  240
```

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

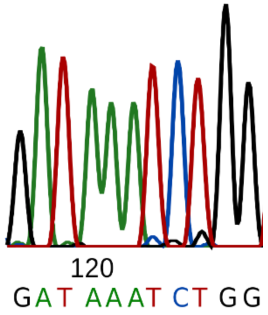
mikroRNA

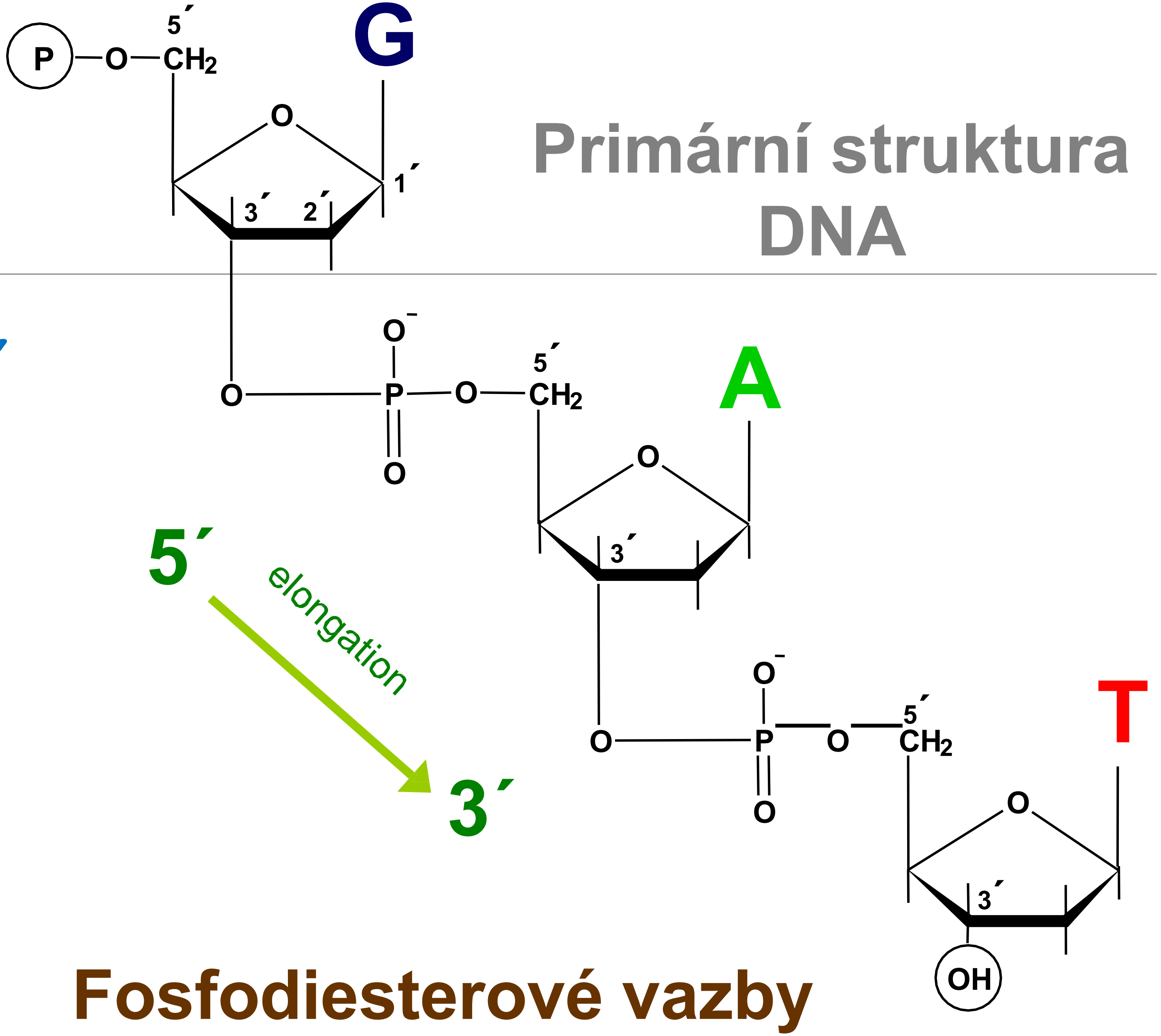
Celé genomy

….

DNA sekvenování

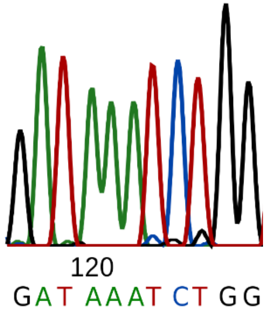
Primární struktura DNA





Fosfodiesterové vazby

DNA sekvenování

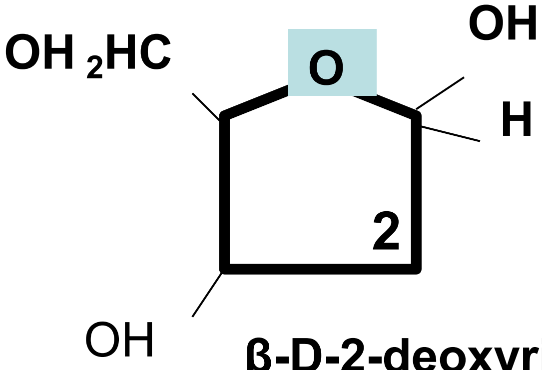


„Klasické“ Sangerovo sekvenování (1977)

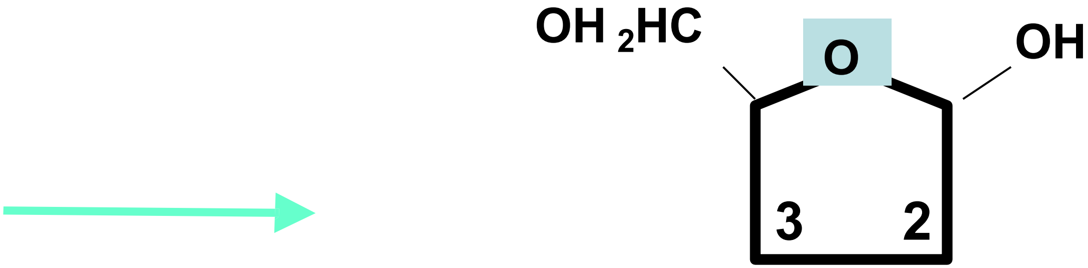
→ Délka čtení: 200-1000nt

→ Výsledek: „text“ 4 písmena (ACTG)

Deoxynukleotidy x **dideoxynukleotidy**



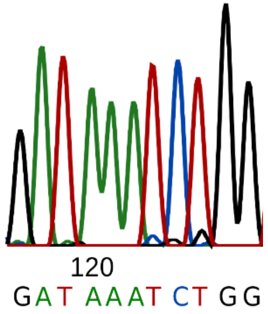
β-D-2-deoxyribose



~1:100

β-D-2,3-dideoxyribose

DNA sekvenování

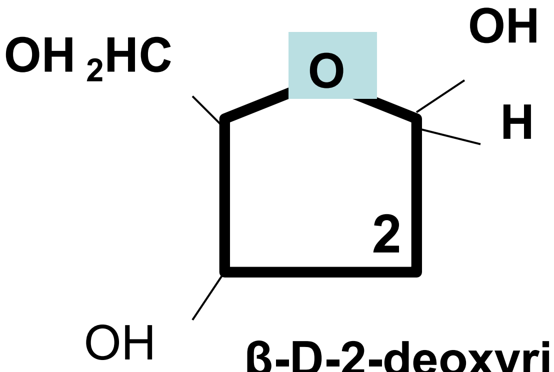


„Klasické“ Sangerovo sekvenování (1977)

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

-směr syntézy 5´→ 3´

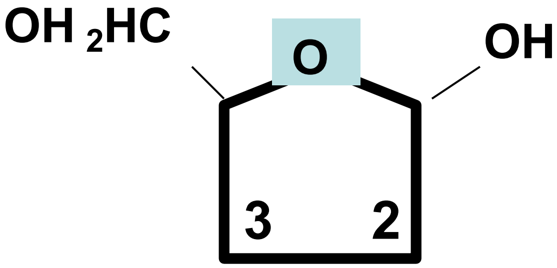
Deoxynukleotidy x **dideoxynukleotidy**



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***

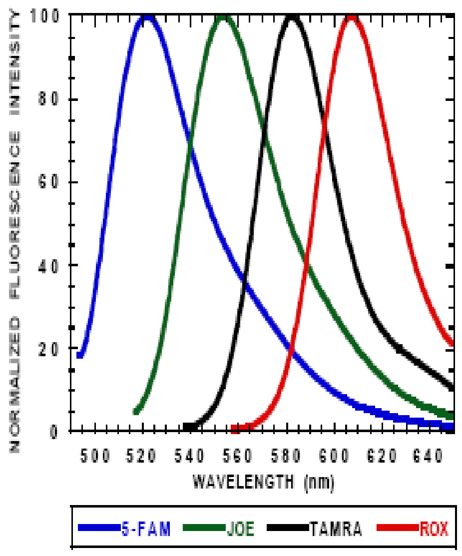
PAGE/UREA > Southern blotting

RA detekce

DNA sekvenování

1980s nové fluorofory umožnily automatizaci DNA sekvenování

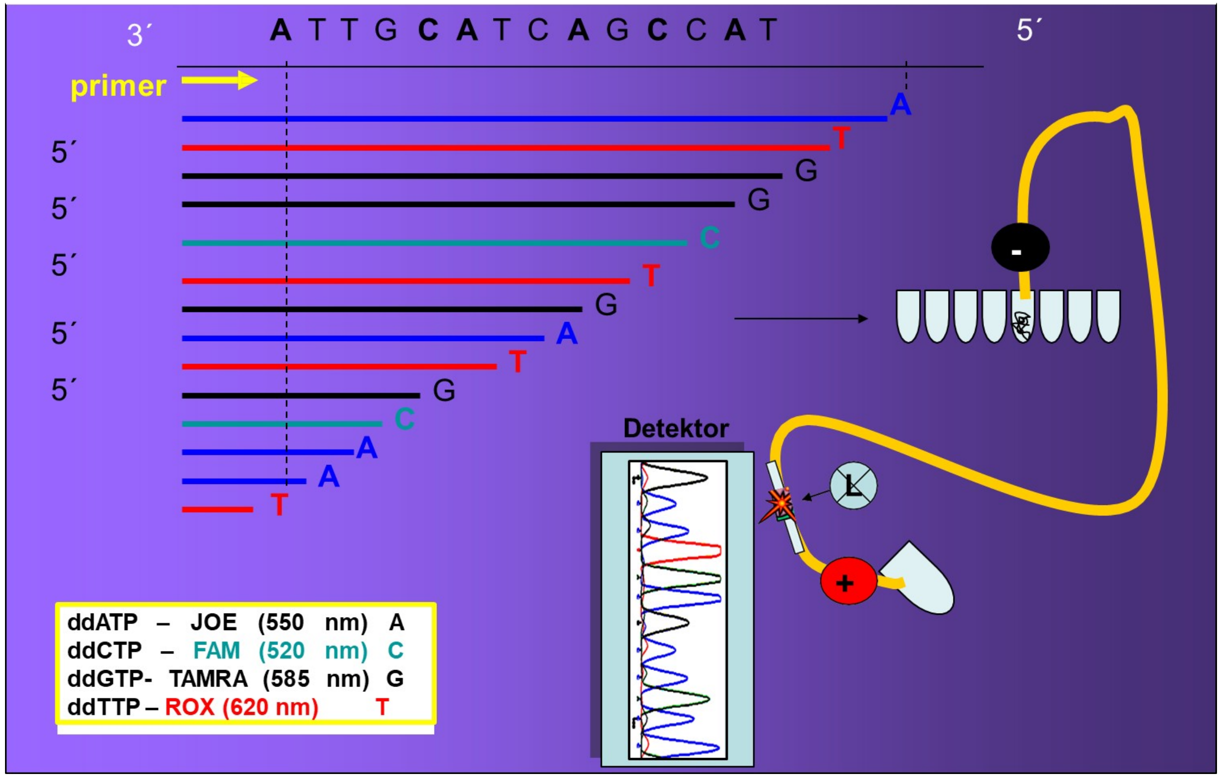
ddATP – JOE (550 nm) A
ddCTP – FAM (520 nm) C
ddGTP- TAMRA (585 nm) G
ddTTP – ROX (620 nm) T



DNA sekvenování

Automatické analyzátory (od 1990s)

Princip: Sangerovo sekvenování a kapilární elektroforéza



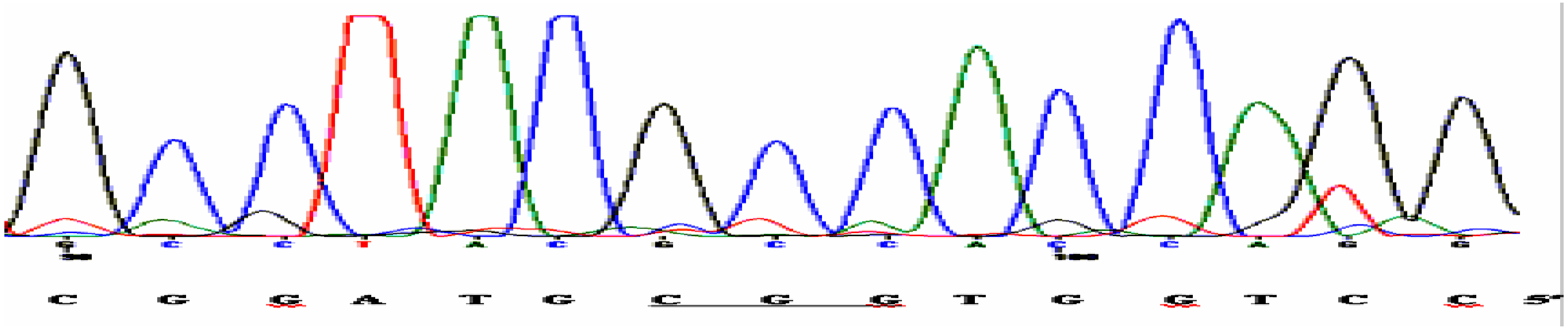
Analýza sekvenčních dat

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

„plain text“: CGGATGCGGTGGTCG

„fasta“: >identifikace
CGGATGCGGTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



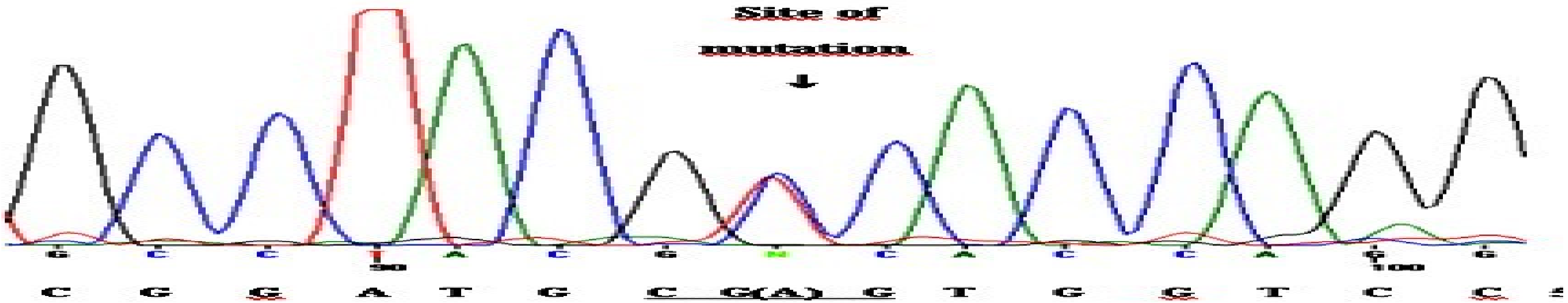
Analýza sekvenčních dat

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

„plain text“: CGGATGCNGTGGTCG

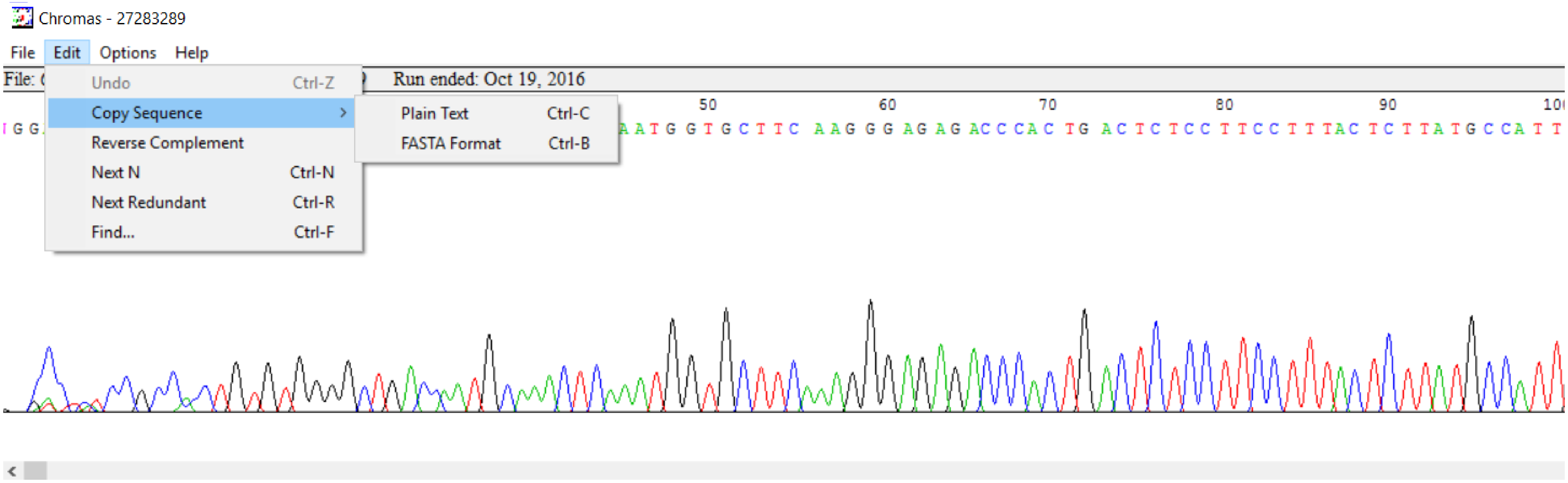
„fasta“: >identifikace
CGGATGC**N**GTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



Analýza sekvenčních dat

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



Vyzkoušejte si čtení sekvenačních dat

Spusťte si „chromas“

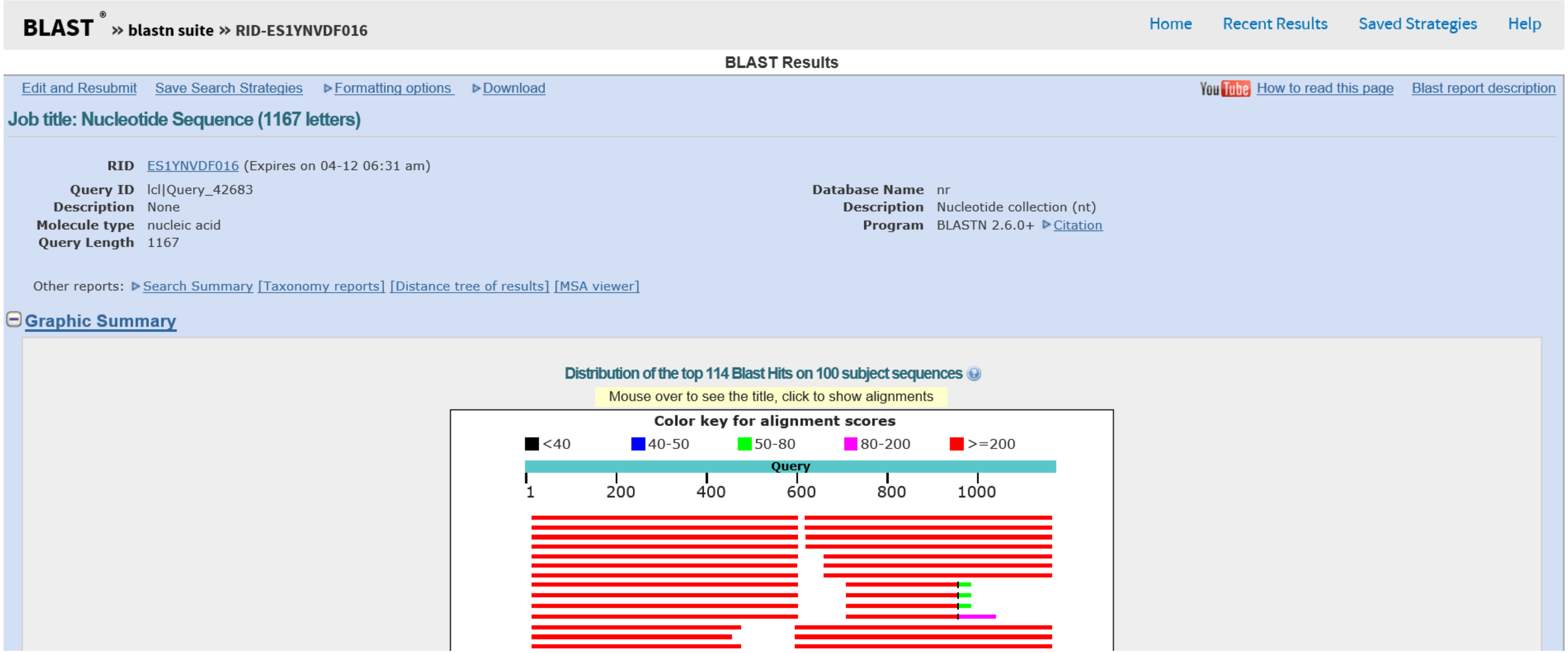
Stáhněte a uložte sekvenci př3

Otevřete v chromasu

- Identifikujte sekvence

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

Analýza sekvenčních dat – příklad (př3)



BLAST® » blastn suite » RID-ES1YNVDF016

BLAST Results

Job title: Nucleotide Sequence (1167 letters)

RID	ES1YNVDF016 (Expires on 04-12 06:31 am)		
Query ID	lcl|Query_42683	**Database Name**	nr
Description	None	**Description**	Nucleotide collection (nt)
Molecule type	nucleic acid	**Program**	BLASTN 2.6.0+ ▷Citation
Query Length	1167		

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

Graphic Summary

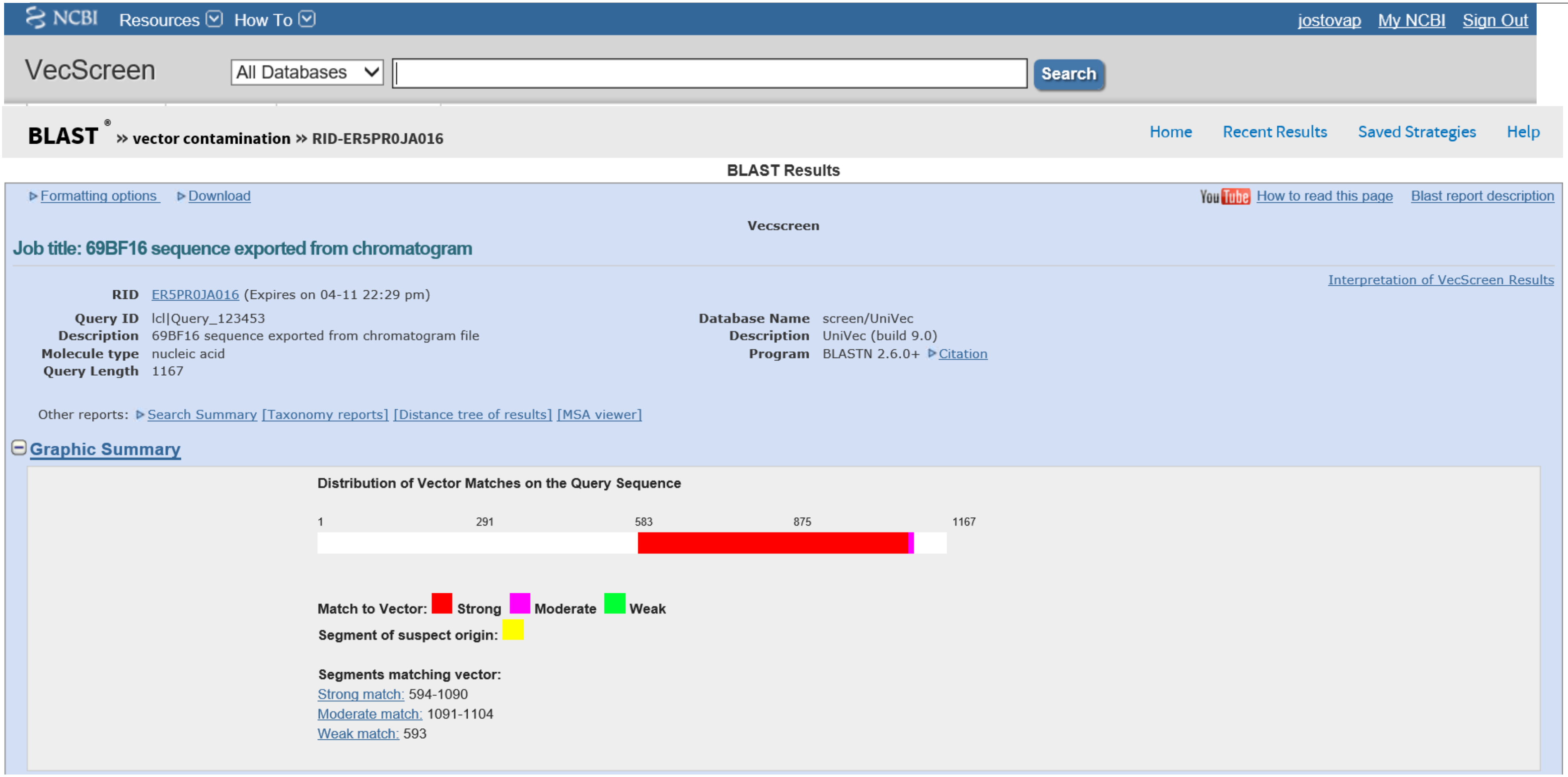
Distribution of the top 114 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

⬛ <40 🟦 40-50 🟩 50-80 🟪 80-200 🟥 >=200

Analýza sekvenčních dat – příklad (př3)



BLAST Results

Vecscreen

Job title: 69BF16 sequence exported from chromatogram

RID ER5PR0JA016 (Expires on 04-11 22:29 pm)

Query ID lcl|Query_123453
Description 69BF16 sequence exported from chromatogram file
Molecule type nucleic acid
Query Length 1167

Database Name screen/UniVec
Description UniVec (build 9.0)
Program BLASTN 2.6.0+ Citation

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

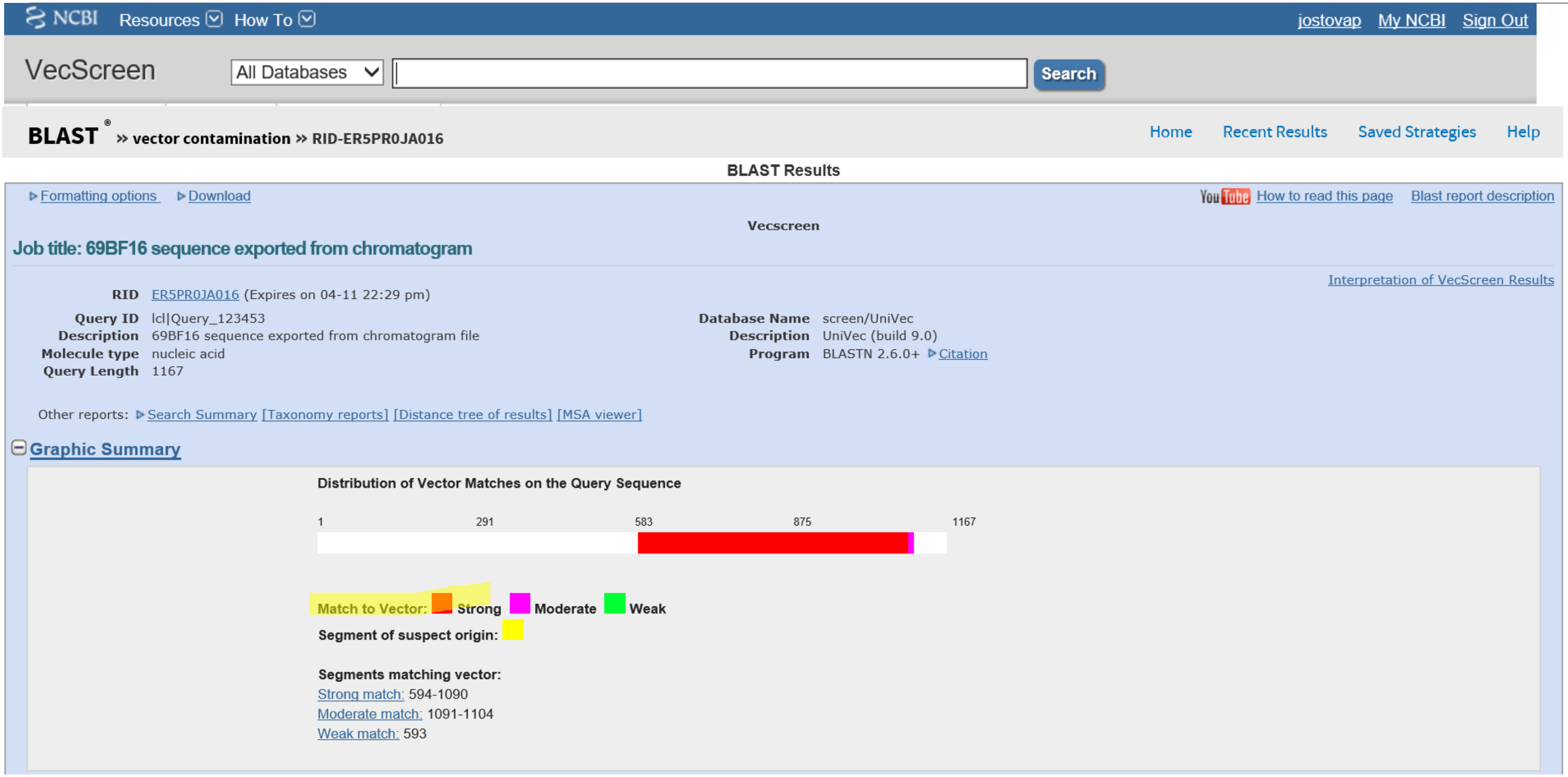
Graphic Summary

Distribution of Vector Matches on the Query Sequence

Match to Vector: 🟥 **Strong** 🟪 **Moderate** 🟩 **Weak**
Segment of suspect origin: 🟨

Segments matching vector:
Strong match: 594-1090
Moderate match: 1091-1104
Weak match: 593

Analýza sekvenčních dat – příklad (př3)



BLAST Results

Vecscreen

Job title: 69BF16 sequence exported from chromatogram

RID	ER5PR0JA016 (Expires on 04-11 22:29 pm)		
Query ID	lcl|Query_123453	**Database Name**	screen/UniVec
Description	69BF16 sequence exported from chromatogram file	**Description**	UniVec (build 9.0)
Molecule type	nucleic acid	**Program**	BLASTN 2.6.0+ Citation
Query Length	1167		

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

Graphic Summary

Distribution of Vector Matches on the Query Sequence

Match to Vector: 🟥 **Strong** 🟪 **Moderate** 🟩 **Weak**

Segment of suspect origin: 🟨

Segments matching vector:

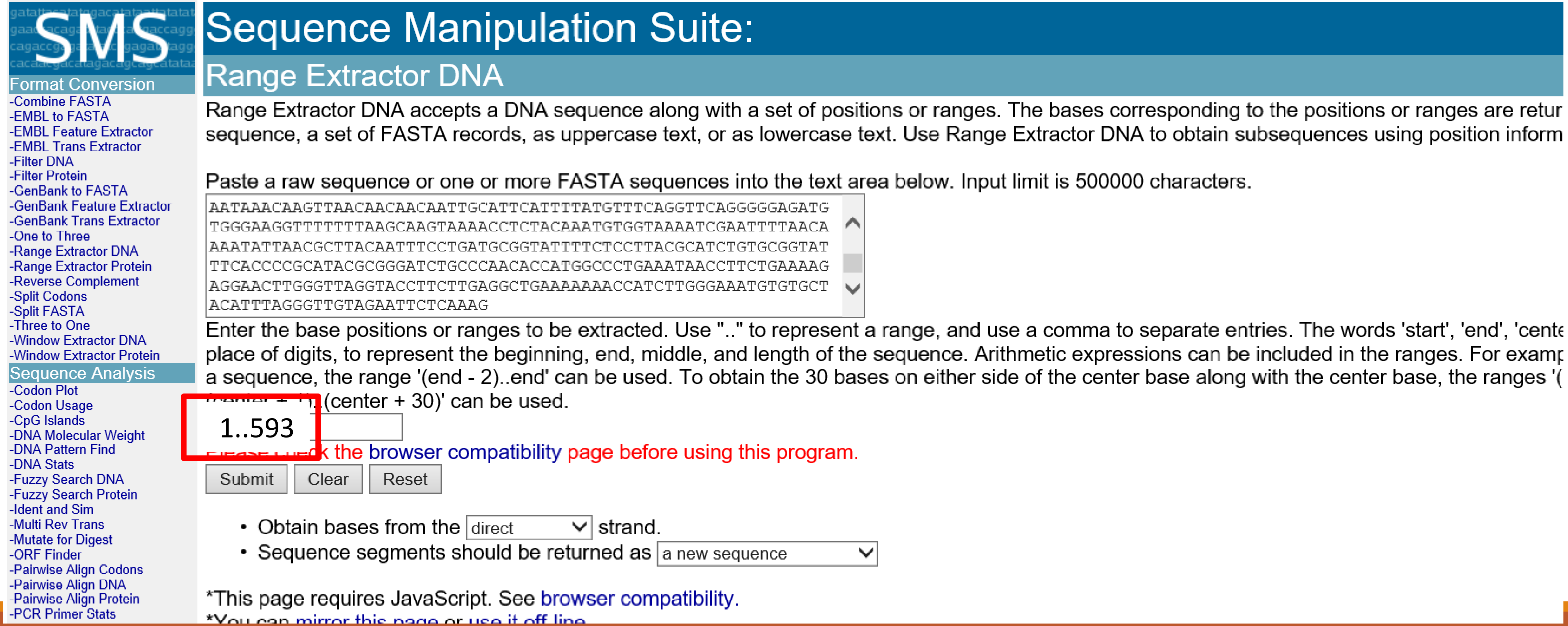
Strong match: 594-1090

Moderate match: 1091-1104

Weak match: 593

Analýza sekvenčních dat: odstranění kontaminace

SMS „Range Extraktor“



SMS

Sequence Manipulation Suite:

Range Extractor DNA

Format Conversion

- -Combine FASTA
- -EMBL to FASTA
- -EMBL Feature Extractor
- -EMBL Trans Extractor
- -Filter DNA
- -Filter Protein
- -GenBank to FASTA
- -GenBank Feature Extractor
- -GenBank Trans Extractor
- -One to Three
- -Range Extractor DNA
- -Range Extractor Protein
- -Reverse Complement
- -Split Codons
- -Split FASTA
- -Three to One
- -Window Extractor DNA
- -Window Extractor Protein

Sequence Analysis

- -Codon Plot
- -Codon Usage
- -CpG Islands
- -DNA Molecular Weight
- -DNA Pattern Find
- -DNA Stats
- -Fuzzy Search DNA
- -Fuzzy Search Protein
- -Ident and Sim
- -Multi Rev Trans
- -Mutate for Digest
- -ORF Finder
- -Pairwise Align Codons
- -Pairwise Align DNA
- -Pairwise Align Protein
- -PCR Primer Stats

Range Extractor DNA accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are retur
sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position inform

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 500000 characters.

```
AATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATG
TGGGAAGGTTTTTTTAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGAATTTTAACA
AAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTAT
TTCACCCCGCATACGCGGGATCTGCCCAACACCATGGCCCTGAAATAACCTTCTGAAAAG
AGGAACTTGGGTTAGGTACCTTCTTGAGGCTGAAAAAAACCATCTTGGGAAATGTGTGCT
ACATTTAGGGTTGTAGAATTCTCAAAG
```

Enter the base positions or ranges to be extracted. Use ".." to represent a range, and use a comma to separate entries. The words 'start', 'end', 'cente
place of digits, to represent the beginning, end, middle, and length of the sequence. Arithmetic expressions can be included in the ranges. For examp
a sequence, the range '(end - 2)..end' can be used. To obtain the 30 bases on either side of the center base along with the center base, the ranges '(
..(center + 30)' can be used.

1..593

k the browser compatibility page before using this program.

Submit | Clear | Reset

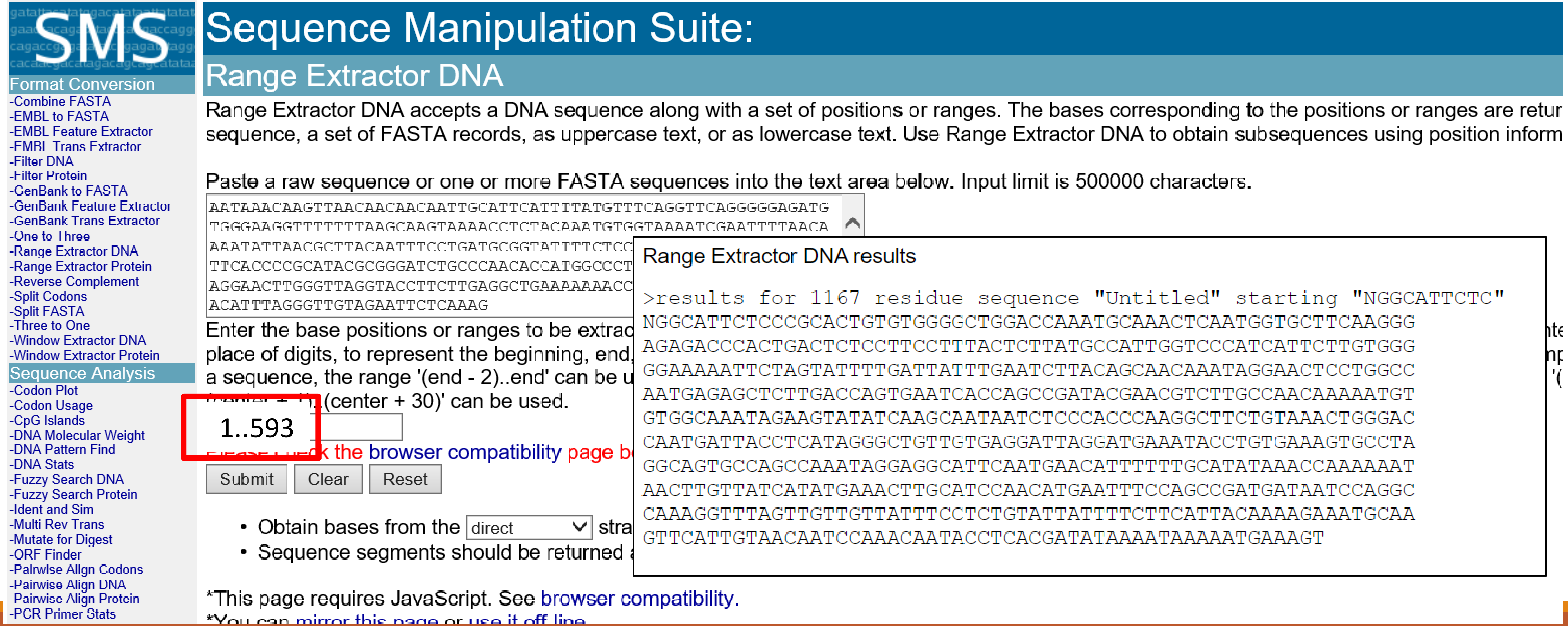
- Obtain bases from the direct strand.
- Sequence segments should be returned as a new sequence

*This page requires JavaScript. See browser compatibility.

*You can mirror this page or use it off line.

Analýza sekvenčních dat: odstranění kontaminace

SMS „Range Extraktor“



Sequence Manipulation Suite:

Range Extractor DNA

Range Extractor DNA accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are retur sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position inform

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 500000 characters.

```
AATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATG
TGGGAAGGTTTTTTTAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGAATTTTAACA
AAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCC
TTCACCCCGCATACGCGGGATCTGCCCAACACCATGGCCCT
AGGAACTTGGGTTAGGTACCTTCTTGAGGCTGAAAAAAACC
ACATTTAGGGTTGTAGAATTCTCAAAG
```

Enter the base positions or ranges to be extrac place of digits, to represent the beginning, end, a sequence, the range '(end - 2)..end' can be u (center + 30)' can be used.

1..593

k the browser compatibility page b

- Obtain bases from the direct stra
- Sequence segments should be returned a

*This page requires JavaScript. See browser compatibility.

*You can mirror this page or use it off line.

Range Extractor DNA results

```
>results for 1167 residue sequence "Untitled" starting "NGGCATTCTC"
NGGCATTCTCCCGCACTGTGTGGGGCTGGACCAAATGCAAACTCAATGGTGCTTCAAGGG
AGAGACCCACTGACTCTCCTTCCTTTACTCTTATGCCATTGGTCCCATCATTCTTGTGGG
GGAAAAATTCTAGTATTTTGATTATTTGAATCTTACAGCAACAAATAGGAACTCCTGGCC
AATGAGAGCTCTTGACCAGTGAATCACCAGCCGATACGAACGTCTTGCCAACAAAAATGT
GTGGCAAATAGAAGTATATCAAGCAATAATCTCCCACCCAAGGCTTCTGTAAACTGGGAC
CAATGATTACCTCATAGGGCTGTTGTGAGGATTAGGATGAAATACCTGTGAAAGTGCCTA
GGCAGTGCCAGCCAAATAGGAGGCATTCAATGAACATTTTTTGCATATAAACCAAAAAAT
AACTTGTTATCATATGAAACTTGCATCCAACATGAATTTCCAGCCGATGATAATCCAGGC
CAAAGGTTTAGTTGTTGTTATTTCCTCTGTATTATTTTCTTCATTACAAAAGAAATGCAA
GTTCATTGTAACAATCCAAACAATACCTCACGATATAAAATAAAAATGAAAGT
```

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

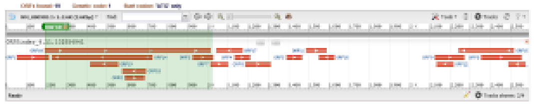
5) Shlédněte toto video PCR Primer Design: https://www.youtube.com/watch?v=c-f1H07D_70

DÚ6-řešení

DÚ6

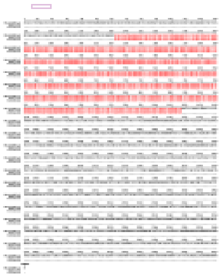
1) NQO1 (mRNA): Čtecí rámec +3

NM_000903.2 Homo sapiens NAD(P)H quinone dehydrogenase 1 (NQO1), transcript variant 1, mRNA

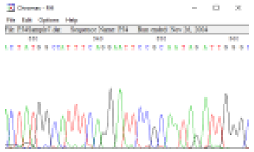


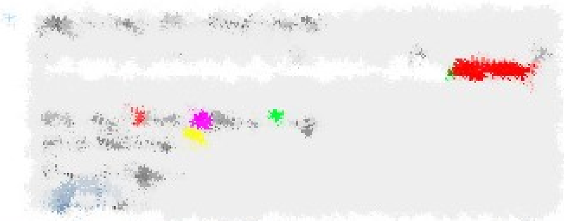
Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF10	+	3	192	1016	825 | 274

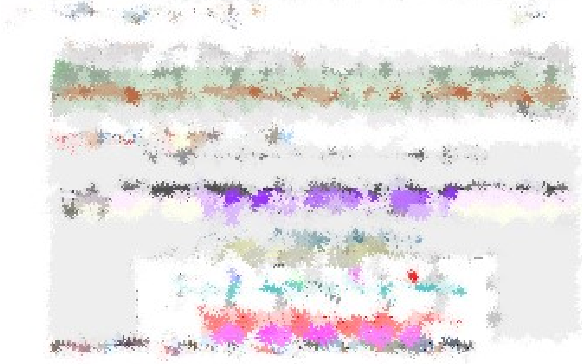
2) Porovnání NQO1 mRNA a CDS

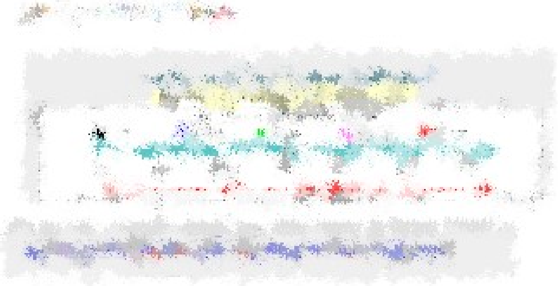


3)





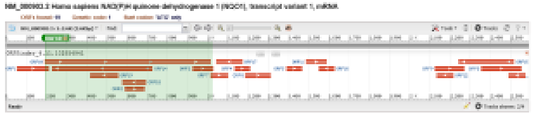




DÚ6-řešení

DÚ6

1) NQO1 (mRNA): Čteci rámec +3



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

2) Porovnání NQO1 mRNA a CDS

3)