

Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2024/2025

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'

||||| ||||| ||||| ||||| |||||

Query Sequence

5' TACTCACGGATGAGGTACTTTAGAGGC 3'

Global Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||| |||||

5' ACTACTAGATT---ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'

Query Sequence

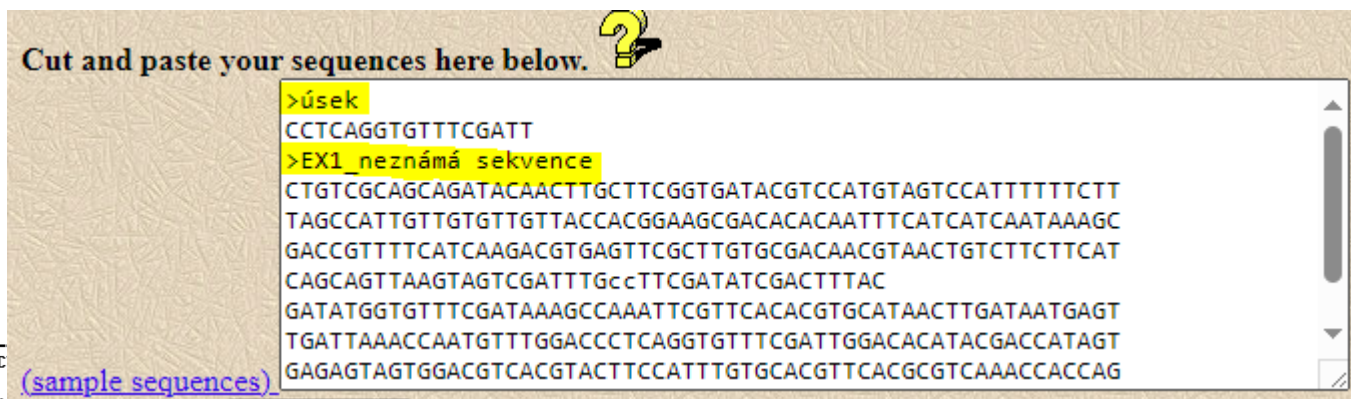
Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Např: multalin

- K vyhledání krátkých úseků (primerů)

Default:BLOSUM62



Porovnání nukleotidových sekvencí

Velmi podobné proteinovému porovnávání

Např: multalin

- K „porovnání“ sekvencí identifikovaných a neznámých

Default: BLOSUM62

→ DNA 5-0

```
      1      10      20
      |-----|-----|
NM_015696.4 GTCTTTGCCTCCGCACGCCGCCAC
69BF16      NGGCATTCTCCGCACTGT
Consensus  .....gccCacgCgaCccCaCcac

      131     140     150
      |-----|-----|
NM_015696.4 TCAACATCCCGGGGCAACTGGTGTC
69BF16      AAAATTCTAGTATTTTGATTTTTG
Consensus  aaAAcaccaGgagcaaaaTgaTGTc

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

      391     400     410
      |-----|-----|
NM_015696.4 AAGATTGCAGTCACCGTACTGGTG
69BF16      AGGCATTCATGAACATTTTTG--
Consensus  AaGaaTgCAaTcAaCagTacTG..

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

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

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

      521     530     540     550     560     570     580     590     600     610     620     630     640     650
      |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
NM_015696.4 CAACTGTGTCAGTGGAGGAGGTCAGACCCAGATCAGCGCTCGTGAGGAGCTCATCCTACTGAGCGAGAGACTTATAACCACCGCGTCTCCTCCTCCACCACTCATCCCGCCACCTGTGTGGG
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-TGCAACTCAATGGTGCTTCAAGGGAGAGACCCACTGACTCTCCTTCTTACTCTTATGCCATTGGTCCCATATTCTTGTGGGGGAAAAATCTAGTATTTTGATTATTTGAATC
69BF16      GCTGGACCAAATGCAACTCAA-TGGTGCTTCAA-GGGAGAGACCCACTGACTCTCCTTCTTACTCTTATGCCATTGGTCCCATATTCTTGTGGGGGAAAAATCTAGTATTTTGATTATTTGAATC
Consensus  GCTG.ACCAA.TGCAACTCAA.TGGTGCTTCAA.GGGAGAGACCCACTGACTCTCCTTCTTACTCTTATGCCATTGGTCCCATATTCTTGTGGGGGAAAAATCTAGTATTTTGATTATTTGAATC

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

Porovnání nukleotidových sekvencí

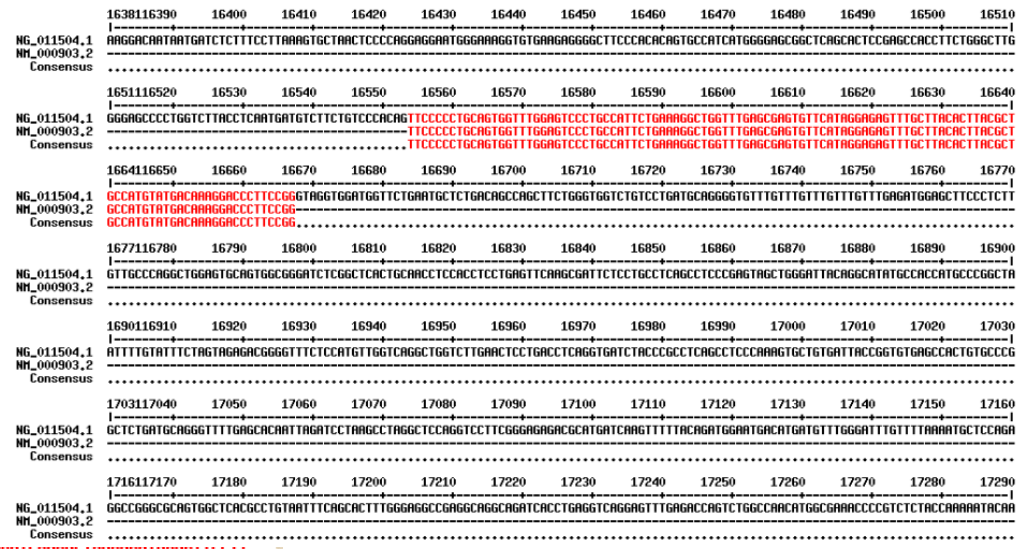
Velmi podobné proteinovému porovnávání

Např: multalin

- K porovnání sekvencí genomových a cDNA (mRNA)

Default: BLOSUM62

→ DNA 5-3



Vyzkoušejte si porovnání

1) vyzkoušejte si porovnat sekvence v PŘO.

- 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** AGTATT TAAAGCGCCACCTATT TAA – 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

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

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

„čtení“ DNA = translace/překlad

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

SMS

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
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find
- Protein Stats
- Restriction Digest
- Restriction Summary

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.

TTAAATAGGTGGCGCTTAAATACTTCCAT

Please check the [browser compatibility page](#) before using this program.

Submit Clear

reading frame 1
reading frame 2
reading frame 3
unspecified frame

direct
reverse

• Translate in

reading frame 1

 on the

direct

 strand.

• Use the

standard (1)

 genetic code.

*This page records your session. [You can mirror it](#) for [compatibility](#).

Fri Jun 17 16:17:06 2005
Valid XHTML 1.0; Valid CSS

[new window](#) | [home](#) | [citation](#)

„čtení“ DNA = translace/překlad

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

SMS

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
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find
- Protein Stats
- Restriction Digest
- Restriction Summary

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.

Submit

Clear

reading frame 1

reading frame 2

reading frame 3

unspecified frame

direct

reverse

• Translate in on the strand
• Use the genetic code

*This page records your session
*You can [mirror](#) your session

Fri Jun 17 16:17:06 20
Valid XHTML 1.0; Valid CSS

Translate results

```
>rf 1 Untitled
LSQQIQLASVIRPCSPFFL*PLLCCYHGSDTQFHHQ*SDRFHQDVSSLVRQRNCLLHQQ
SSRFGQ*SCLFRYRLYDMVFR*SQIRSHVHNLIMS*LNQCLDPQVFRLDTYDHSESSGRH
VLPFVHVHASNHQVLMLIDDVCVSTQQVGS
```

Translate results

```
>rf 2 Untitled
CRSRYNLLR*YVHVHFFFSHCCVVTTEATHNFIINKATVFIKT*VRLCDNVTVFFISS*
VVDLVSDRVCFDIDETIWCDFDAKFVHTCIT***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

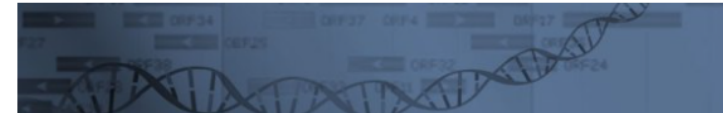
ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

NCBI Resources ▾ How To ▾ jostovap My NCBI Sign Out

ORFfinder PubMed ▾ Search

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 ▾

ORF start code: ☒ "ATG" only ☐ "ATG" and ☐ Any sense

Ignore nested: ☐

Start Search /

Submit Clear

- 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

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

NCBI Resources How To jostovap My NCBI Sign Out

ORFfinder PubMed Search

Open Reading Frame Viewer Help

Sequence

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

Find: Tools Tracks ?

ORF4 (274 aa) Display ORF as... Mark

Mark subset... Marked: 0 Download marked set as Protein FASTA

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

Six-frame translation...

ORF4 (274 aa)

SmartBLAST BLAST

Marked set (0)

SmartBLAST best hit titles... BLAST

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

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

NCBI Resources How To

ORFfinder PubMed

Open Reading Frame Viewer

Sequence

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

Find: 1 300 400 500 600 700 800 900 1 K 1,100 1,200 1,300 1,400 1,500

ORF4 ORF12 ORF7 ORF8 ORF15 ORF9 ORF5 ORF17 ORF16 ORF6

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

ORF4 (274 aa) Display ORF as... Mark

>lc1|ORF4
MVGRRALIVLAHSERISFNYAMKEAAAAALKKKGWEVVEDLYAMNENPI
ISRKIDITGKLDKDPANFQYPAESVLAKEGHLSPDIVAEQKLEADLVIF
QFPLQWFGVPAILKGFERFVFIGEFAYTYAAMYDKGPFPRSKKAVLSITIG
GSGSMYSLQGIHGMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADA
RIQILEGKKRLENIWDETPLYFAPSSFLDLNFQAGFLMKKEVQDEEKNK
KFLSVGHHLGKSIPTDNQIKARK

ORF4 SmartBLAST BLAST

Marked set (0)

SmartBLAST best hit titles... BLAST

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

ATGGTCGGCAGAAGAGCACTGATCGTACTGGCTCACTCAGAGAGGACGTCTTCAACTATGCCATGAAGG
AGGCTGCTGCAGCGGCTTTGAAGAAGAAAGGATGGGAGGTGGTGGAGTCGGACCTCTATGCCATGAACCT
CAATCCCATCATTTCCAGAAAGGACATCACAGGTAAGTGAAGGACCTGCGAAGCTTCAGTATCCTGCC
GAGTCTGTTCTGGCTTATAAAGAAGGCCATCTGAGCCAGATATTGGGCTGAACAAAAGAAGCTGGAAG
CCGCAGACCTTGTGATATTCCAGTTCCTCCCTGCAAGTGGTTGGAGTCCCTGCCATTCTGAAGGCTGGTT
TGAGCGAGTGTTTCATAGGAGAGTTTGTCTTACACTTACGCTGCCATGTATGACAAAGGACCTTCCGGAGT
AAGAAGGCAGTGCTTTCCATCACCACCTGGTGGCAGTGGCTCCATGTACTCTCTGCAAGGGATCCACGGGG
ACATGAATGTCATTCTCTGGCCAATTCAGAGTGGCATTCTGCATTCTGTGGCTTCCAAGTCTTAGAACC
TCAACTGACATATAGCATTGGGCACACTCCAGCAGACGCCCAATTCAAATCTGGAAGGATGGAAGAAA
CGCTGGAGAATATTTGGGATGAGACCACTGTATTTTGTCTCAAGCAGCTCTTTGACCTAACTTCC
AGGCAGGATTCTTAATGAAAAAGAGGTACAGGATGAGGAGAAAAACAAGAAATTTGGCCTTTCTGTGGG
CCATCACTTGGGCAAGTCCATCCCACTGACAACCATGCAAGCTAGAAATGA

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

„čtení“ DNA = translace/překlad

ORFfinder → vhodné pro vyhledávání nejdelších „ORFů“

NCBI Resources How To jostovap My NCBI Sign Out

ORFfinder PubMed Search

Open Reading Frame Viewer Help

Sequence

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

Find: 100 200 300 400 500 600 700 800 900 1,000 1,200

ORFfinder_4.5.18628361

ORF4 ORF17 ORF12 ORF7 ORF15 ORF1 ORF2 ORF6 ORF16

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

BLAST » blastp suite » RID-8M12WFM01R

BLAST Results

Edit and Resubmit Save Search Strategies Formatting options Download YouTube How to read this page Blast report description NEW Click here to use the new BLAST results page

Job title: Protein Sequence

RID 8M12WFM01R (Expires on 04-07 00:16 am)

Query ID Icl|Query_57926

Description Icl|ORF15:1101:970 unnamed protein product

Molecule type amino acid

Query Length 43

Database Name swissprot

Description Non-redundant UniProtKB/SwissProt sequences

Program BLASTP 2.10.0+ Citation

No significant similarity found. For reasons why, click here

Other reports: Search Summary

ORF15 (43 aa) Display ORF as... Mark

>Icl|ORF15
MIQKCTNTVDLSFIPCGKRKTQILKTKASQGSLEYPDLITC

Mark subset... Marked: 0 Download marked set as Protein FASTA

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF15	-	2	1101	970	132 43
ORF2	+	1	1117	1242	126 41
ORF16	-	2	624	505	120 40
ORF11	-	1	2506	2420	87 28
ORF1	+	1	943	1029	87 28

ORF4

SmartBLAST

BLAST

SmartBLAST best hit titles...

BLAST

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

Enter Query Sequence

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

```
ggtttacaattgtacccaagggtcatgggactggacccatcccatagtaagtcacagt
ttagcaatgataaagaaaataaccttctgaaaattgtatagatcagaaataaagtattt
tttgtggaagactatttttaagtattgaaggtaactatttccttcttgaattcatattgc
agatgtacggtgtggatttattggttatctctgcaaaccttaaagtagaagattgcaag
ggccagggtgtggatcacgcctgtaatcccagcactttgggaggctgaggtaggcgga
tcacctgaggtcaggagtttgagaccagcctggccaacatggtgaaacgccgtctctact
aaaaatacaaaaattagtcaggcatgatggcaggcgctgtcatcccagctactagggag
```

From: To:

Choose Search Parameters

Minimal ORF length (nt):

Genetic code:

ORF start codon to use:

☐ "ATG" only

☐ "ATG" and alternative initiation codons

☒ Any sense codon

Ignore nested ORFs: ☐

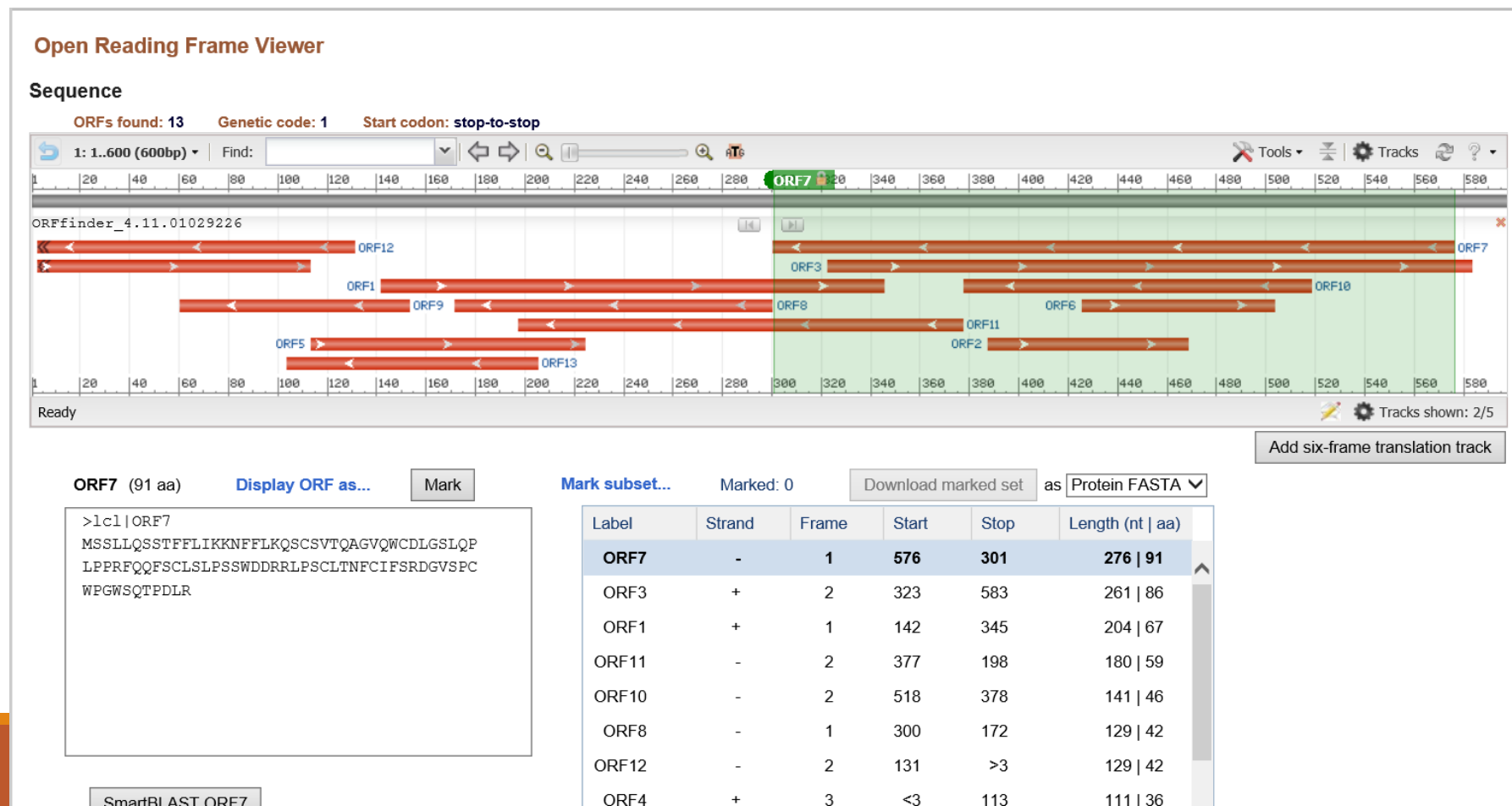
Web www.ncbi.nlm.nih.gov říká

No ORFs found. Try to change search parameters.

OK

Identifikace neznámé sekvence

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



Identifikace neznámé sekvence

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

→ vyhledání příbuzných sekvencí

BLAST!

The screenshot shows the NCBI BLAST website. At the top is a navigation bar with the NIH logo, "U.S. National Library of Medicine", "NCBI National Center for Biotechnology Information", and user links for "jostovap", "My NCBI", and "Sign Out". Below this is a sub-header with the "BLAST" logo and navigation links: "Home", "Recent Results", "Saved Strategies", and "Help".

The main content area is divided into several sections:

- Basic Local Alignment Search Tool**: A text block explaining that BLAST finds regions of similarity between biological sequences and compares nucleotide or protein sequences to sequence databases. It includes a "Learn more" link.
- News**: A sidebar on the right with a "NEWS" label and a "Magic-BLAST 1.2.0 released" announcement, dated Mon, 27 Feb 2017 14:00:00 EST, with a "More BLAST news..." link.
- Web BLAST**: A section with three main options, each in a blue box with a DNA/RNA/protein graphic:
 - Nucleotide BLAST**: "nucleotide ► nucleotide". This box is highlighted with a red border.
 - blastx**: "translated nucleotide ► protein".
 - tblastn**: "protein ► translated nucleotide".
- Protein BLAST**: "protein ► protein".

At the bottom, there is a **BLAST Genomes** section with a search input field containing the placeholder text "Enter organism common name, scientific name, or tax id" and a "Search" button. Below the input field are links for "Human", "Mouse", "Rat", and "Microbes".

Identifikace neznámé sekvence

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

→ vyhledání příbuzných sekvencí

BLAST!

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information jostovap My NCBI Sign Out

BLAST® Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [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

Organism [Optional](#) [Nucleotide collection \(nr/nt\)](#)

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [Optional ☐ Sequences from type material](#)

Entrez Query [Optional](#) [YouTube](#)

Enter an Entrez query to limit search [?](#)

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)

Identifikace neznámé sekvence

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

BLAST® » blastn suite » RID-ES0WVJXE016

HomeRecent ResultsSaved StrategiesHelp

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: Nucleotide Sequence (600 letters)

RID [ES0WVJXE016](#) (Expires on 04-12 06:13 am)

Query ID lcl|Query_28989
Description None
Molecule type nucleic acid
Query Length 600

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

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

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 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

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

Query 1GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGT60

Sbjct 2003GGTTTACAATTGTACCCCAAGGTCATGGGACTGGACCCCATCCCATAGTAAGTCATCAGT1944

Query 61TTAGCAATGATAAAAGAAAATAACCTTCTGAAAATTTGTATAGATCAGAAATAAAGTATTT120

Sbjct 1943TTAGCAATGATAAAAGAAAATAACCTTCTGAAAATTTGTATAGATCAGAAATAAAGTATTT1884

Query 121TTTGTGGAAGACTATTTTAAAGTATTGAAGGTACTATTCCTTTCTTGAATTCATATTGC180

Sbjct 1883TTTGTGGAAGACTATTTTAAAGTATTGAAGGTACTATTCCTTTCTTGAATTCATATTGC1824

Query 181AGATGTACGGTGTGGATTTATTGGTTTATCTCTGCAAACCTTAAAGTAGAAGATTGCAAG240

Related Information

[Gene](#) - associated gene details

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

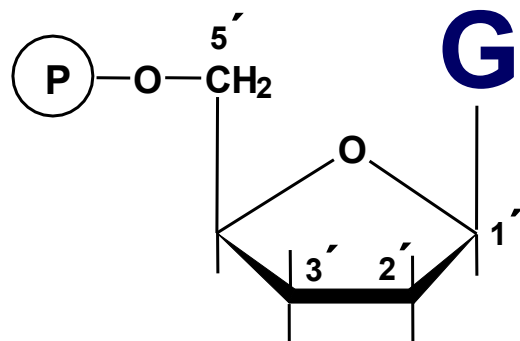
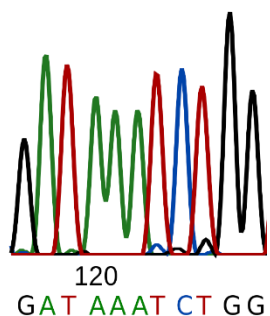
Vyhledávání hladin expresí jednotlivých genů

mikroRNA

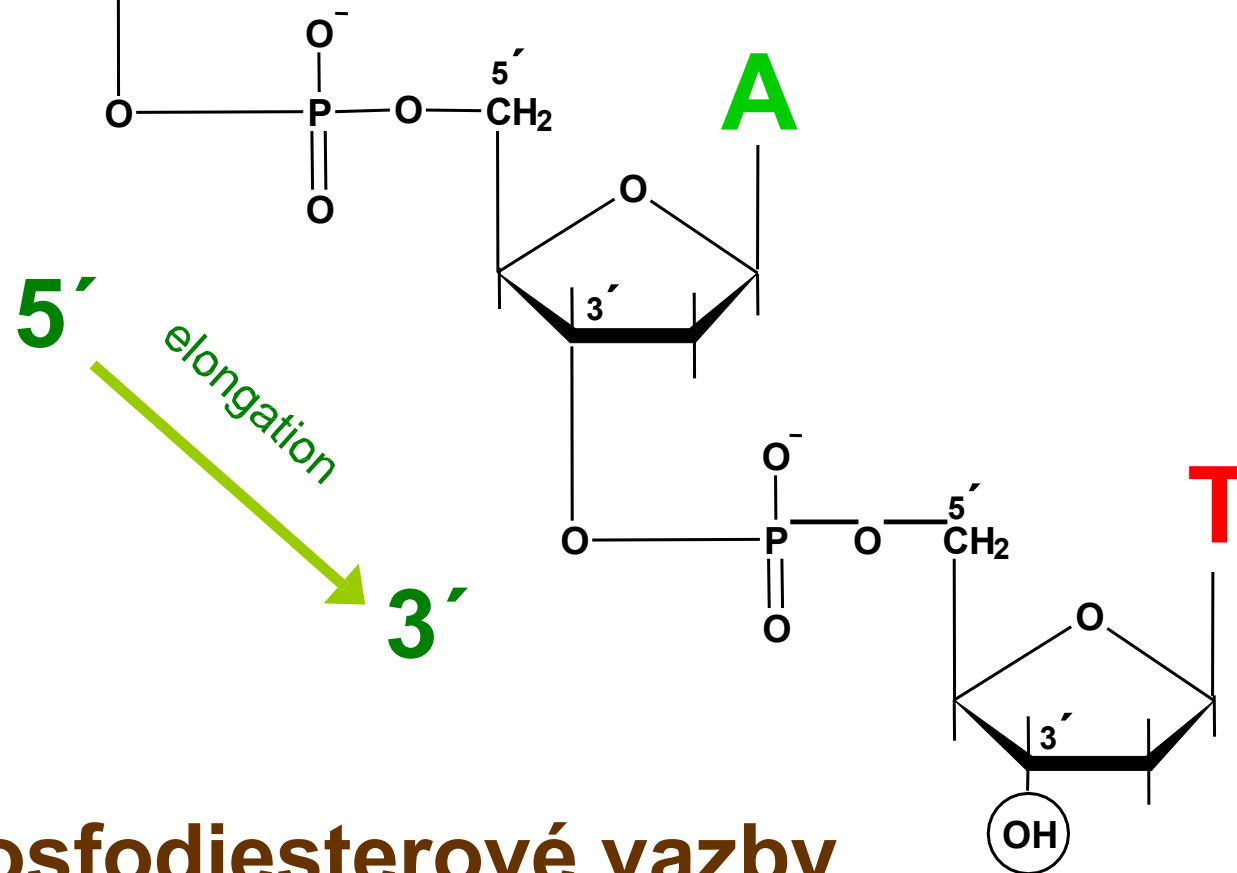
Celé genomy

....

DNA sekvenování

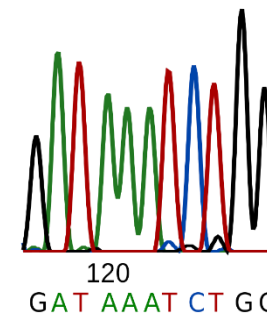


Primární struktura DNA



Fosfodiesterové vazby

DNA sekvenování

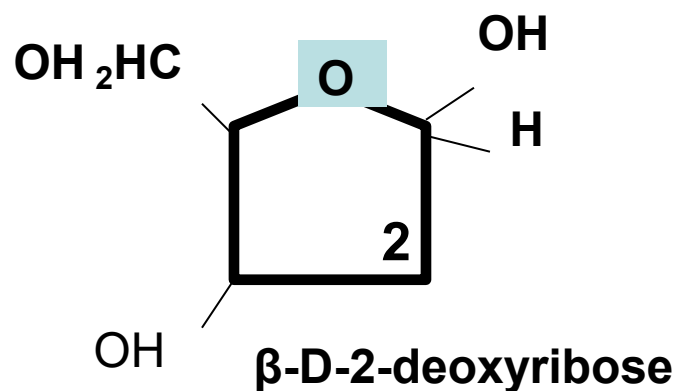


„Klasické“ Sangerovo sekvenování (1977)

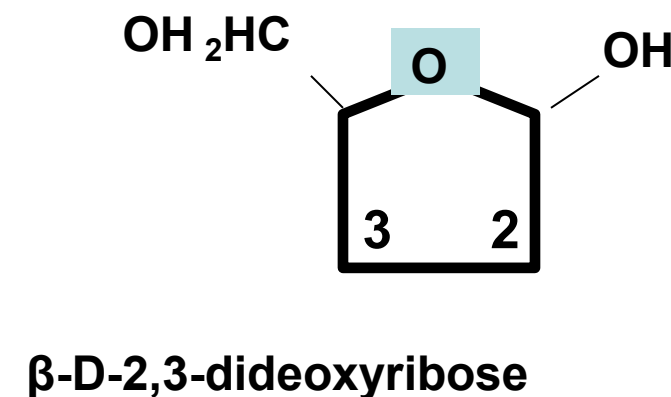
→ Délka čtení: 200-1000nt

→ Výsledek: „text“ 4 písmena (ACTG)

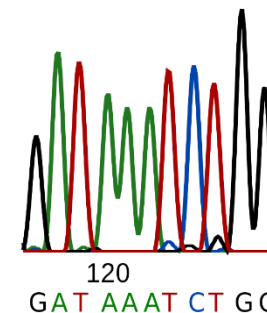
Deoxynukleotidy x **dideoxynukleotidy**



→
~1:100



DNA sekvenování

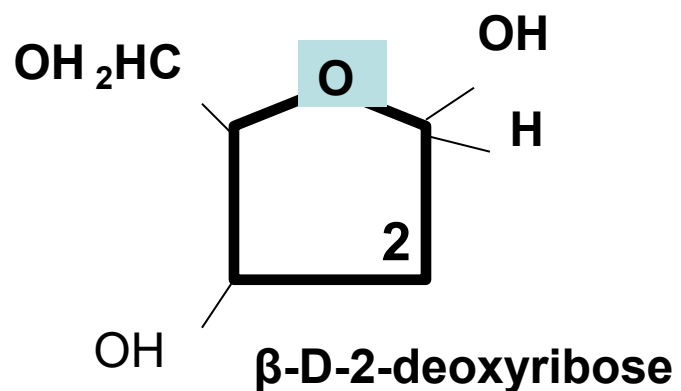


„Klasické“ Sangerovo sekvenování (1977)

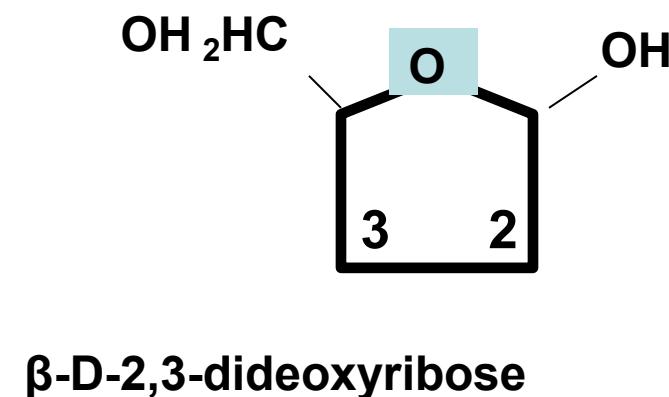
- Asymetrická DNA amplifikace (DNA polymeráza I, 1 primer)

-směr syntézy $5' \rightarrow 3'$

Deoxynukleotidy x **dideoxynukleotidy**



$\sim 1:100$



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

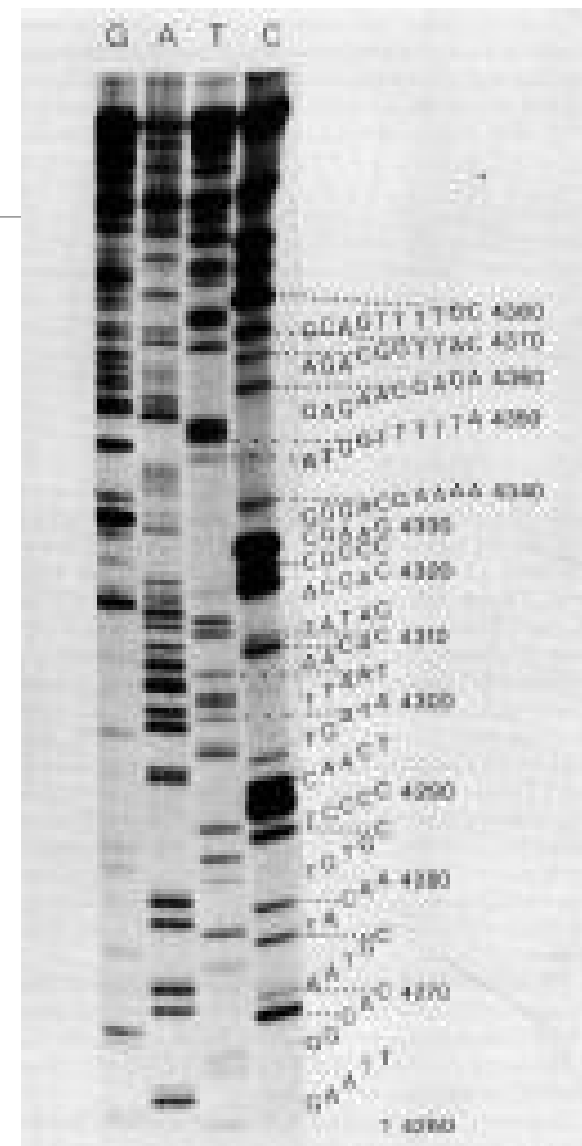
Reakční mix 2
3 dNTPs
1:1 dATP:ddATP

Reakční mix 3
3 dNTPs
1:100 dTTP:ddTTP

Reakční mix 4
3 dNTPs
1:50 dCTP:ddCTP

PAGE/UREA > Southern blotting

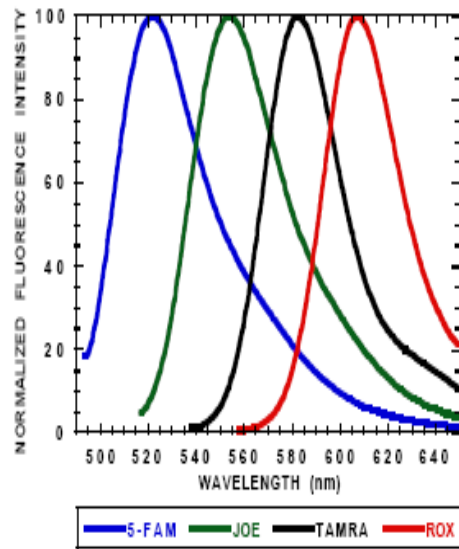
RA detekce



DNA sekvenování

1980s nové fluorofory umožnily automatizaci DNA sekvenování

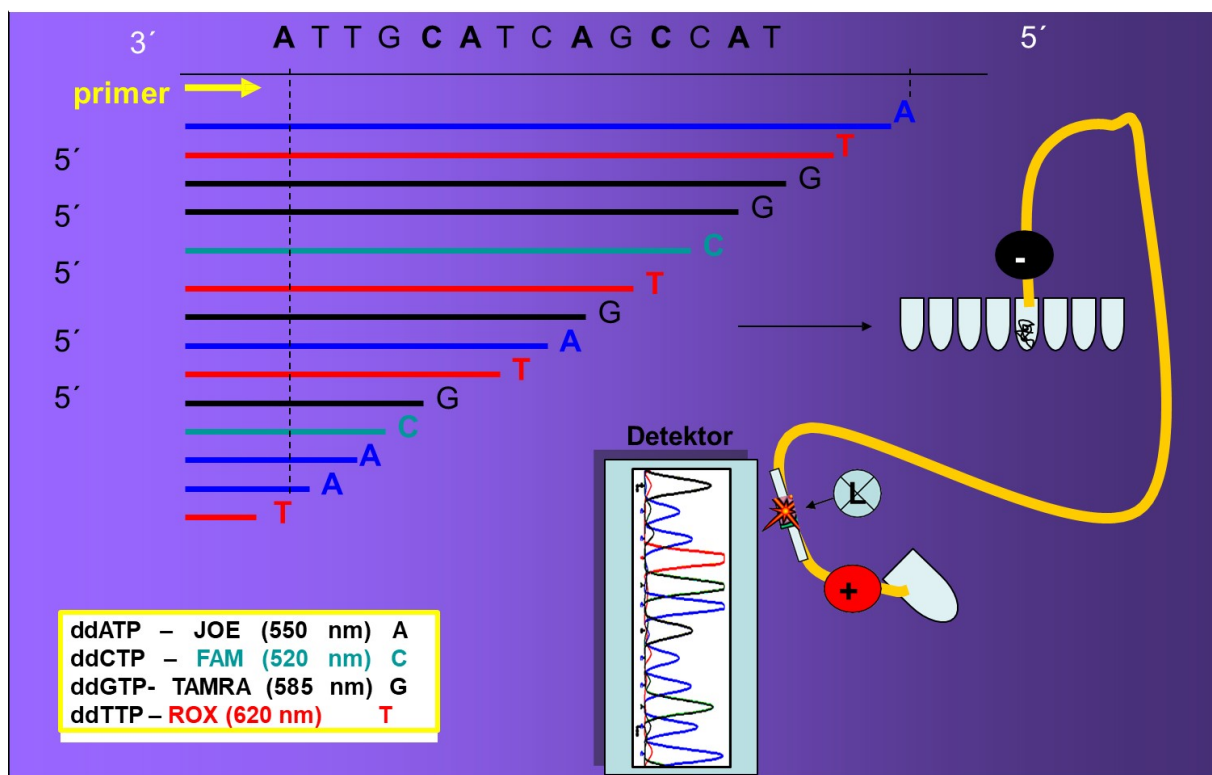
ddATP – JOE (550 nm)	A
ddCTP – FAM (520 nm)	C
ddGTP- TAMRA (585 nm)	G
ddTTP – ROX (620 nm)	T



DNA sekvenování

Automatické analyzátoři (od 1990s)

Princip: Sangerovo sekvenování a kapilární elektroforéza



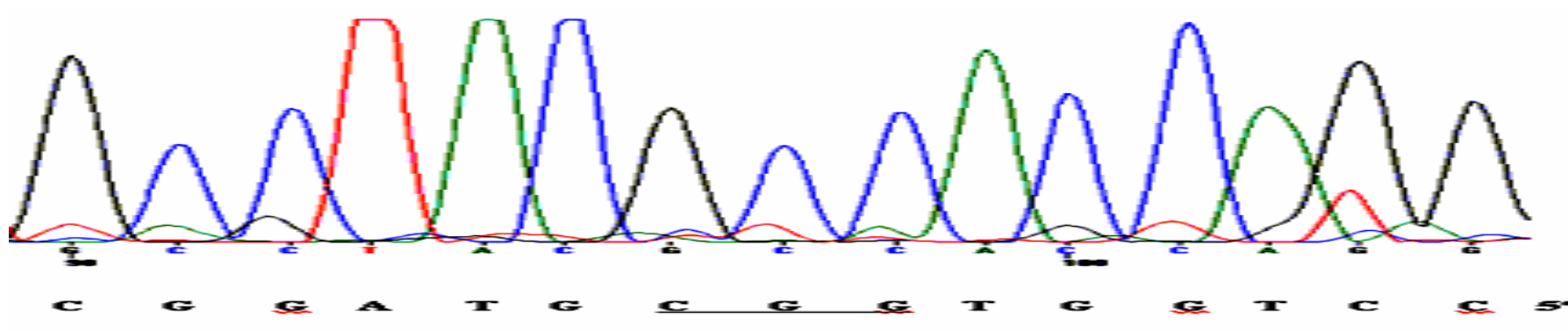
Analýza sekvenčních dat

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

„plain text“: CGGATGCGGTGGTCG

„fasta“: >identifikace
CGGATGCGGTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



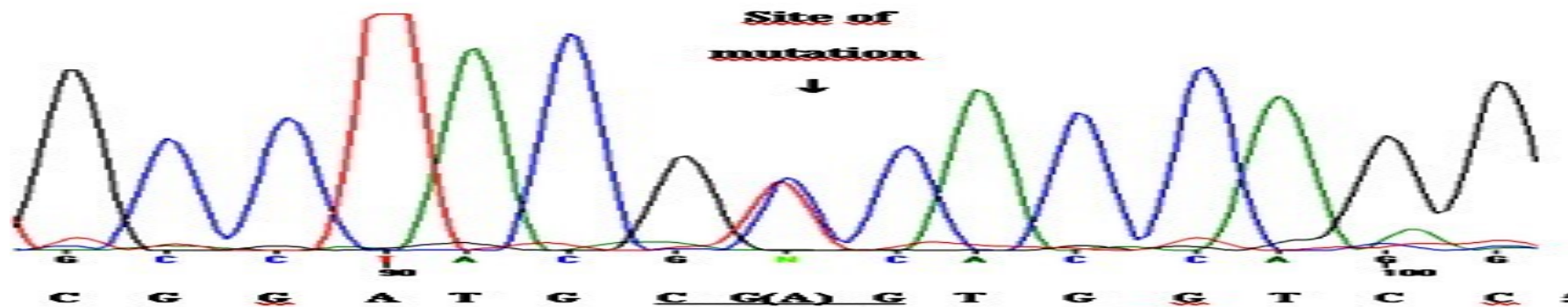
Analýza sekvenčních dat

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

„plain text“: CGGATGCNGTGGTCG

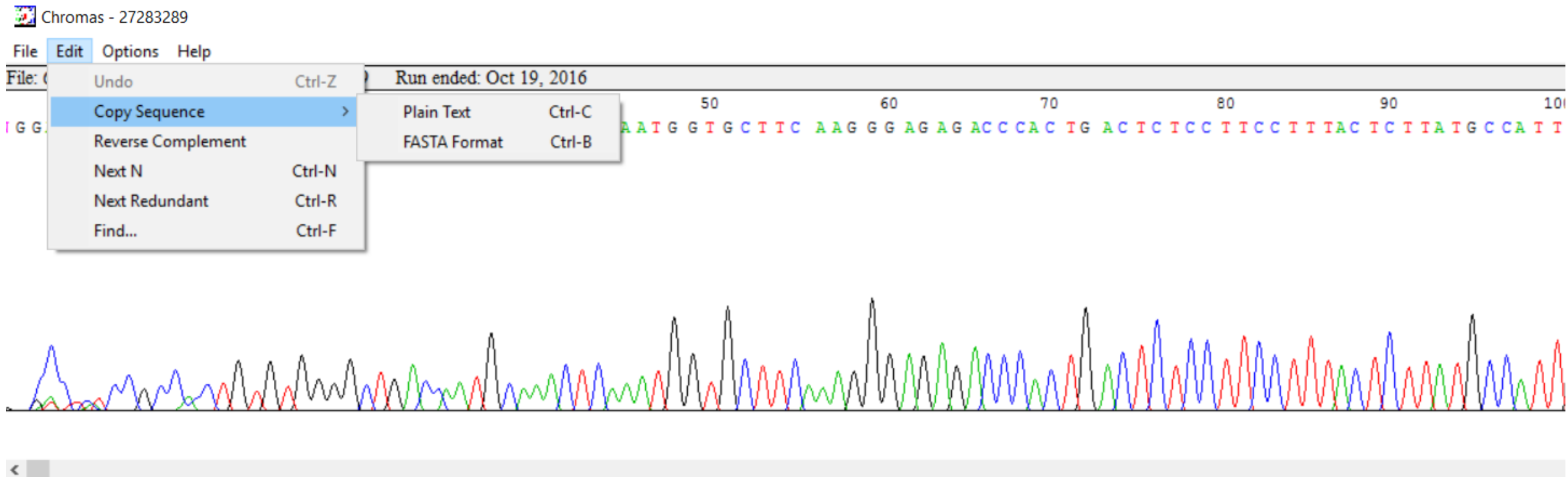
„fasta“: >identifikace
CGGATGCNGTGGTCG

„sekvenační formát“(.scf, .abi, .ab1)



Analýza sekvenčních dat

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



Vyzkoušejte si čtení sekvenačních dat

Spusťte si „chromas“

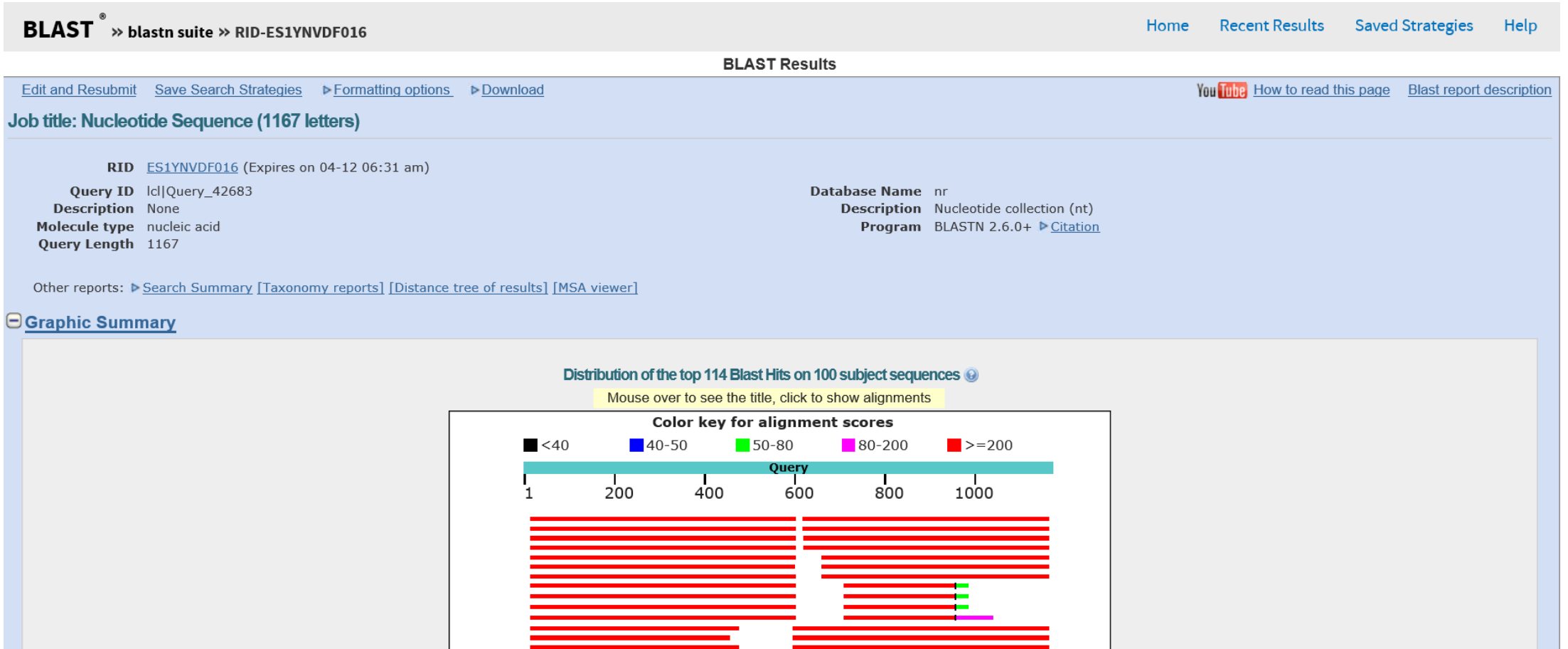
Stáhněte a uložte sekvenci př3

Otevřete v chromasu

➤ Identifikujte sekvenci

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

Analýza sekvenčních dat – příklad (př3)



Analýza sekvenčních dat – příklad (př3)

NCBI Resources ▾ How To ▾ jostovap My NCBI Sign Out

VecScreen All Databases ▾ Search

BLAST® » vector contamination » RID-ER5PR0JA016 Home Recent Results Saved Strategies Help

BLAST Results

► Formatting options ► Download YouTube How to read this page Blast report description

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



1 291 583 875 1167

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)

NCBI Resources ▾ How To ▾ jostovap My NCBI Sign Out

VecScreen All Databases ▾ Search

BLAST® » vector contamination » RID-ER5PR0JA016 Home Recent Results Saved Strategies Help

BLAST Results

► Formatting options ► Download YouTube How to read this page Blast report description

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

1 291 583 875 1167

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

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

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 returned as a sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position information.

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

AATAAACAAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGATG
TGGGAAGGTTTTTTTAAGCAAGTAAACCTCTACAAATGTGGTAAATCGAATTTTAACA
AAATATTAACGCTTACAATTTCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTAT
TTCACCCCGCATACGCGGGATCTGCCAACACCATGGCCCTGAAATAACCTTCTGAAAAG
AGGAACTTGGGTTAGGTACCTTCTTGAGGCTGAAAAAACCATCTTGGGAAATGTGTGCT
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', 'center', 'begin', 'end', 'middle', and length of the sequence. Arithmetic expressions can be included in the ranges. For example, to obtain the 30 bases on either side of the center base along with the center base, the ranges '(center - 30)..(center + 30)' can be used.

1..593

[Please check 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“

SMS

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

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 returned as a sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position information.

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

AATAAACAAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGATG
TGGGAAGGTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAATCGAATTTTAACA

AAATATTAACGCTTACAATTTCTGATGCGGTATTTCTCT
TTCACCCCGCATACGCGGGATCTGCCAACACCATGGCCCT
AGGAACTTGGGTTAGGTACCTTCTTGAGGCTGAAAAAACC
ACATTTAGGGTTGTAGAATTCTCAAAG

Enter the base positions or ranges to be extracted. To represent the beginning, end, or a sequence, the range '(end - 2)..end' can be used. To represent the center, the range '(center + 30)' can be used.

1..593

Submit

Clear

Reset

- Obtain bases from the strategy
- Sequence segments should be returned as a set of FASTA records, as uppercase text, or as lowercase text.

*This page requires JavaScript. See [browser compatibility](#).
*You can [mirror this page](#) or [use it off line](#).

DÚ6

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

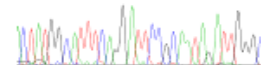
1) NQO1 (mRNA): Čtecí rámec +3

HM_000003.2 Homo sapiens NAD(P) quinone dehydrogenase 1 (NQD1), transcript variant 1, mRNA

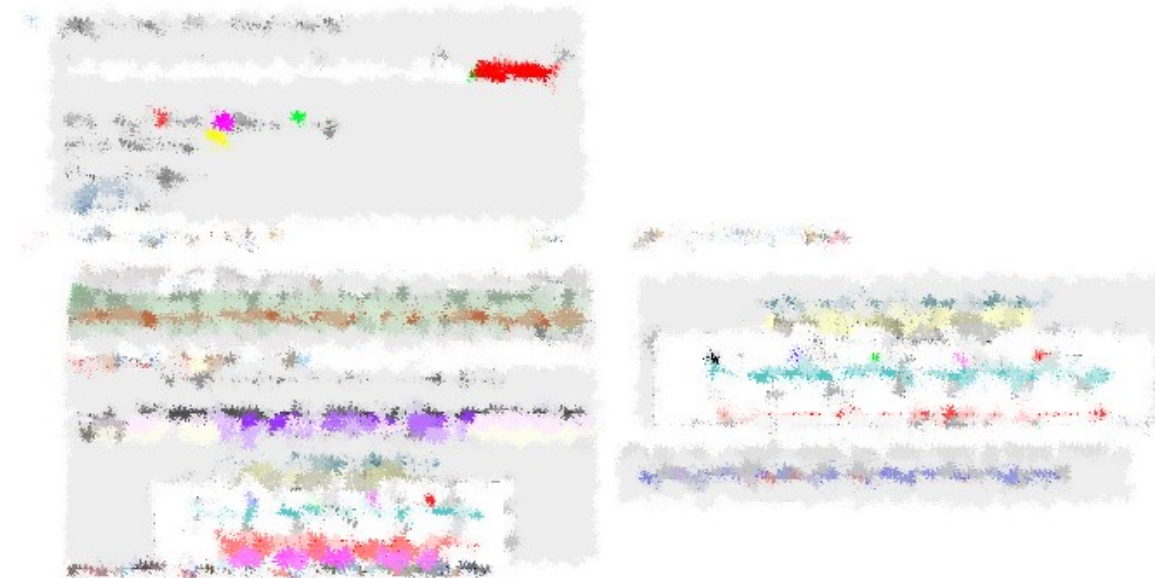


Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF10	+	3	192	1016	825 274

3)

[illegible]

2) Porovnání NQO1 mRNA a CDS



DÚ6-řešení

DÚ6

1) NQO1 (mRNA): Čtecí rámec +3

HM_000003.2 Homo sapiens NAD(P)H quinone dehydrogenase 1 (NQD1), transcript variant 1, mRNA



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

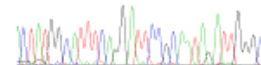
3)

Chromas-BE

File Edit Options Help

File	Sequence Name	Run
File	File	File

A T T A T T G C C T T C T C A G G A A C C C C C G G T A G A T T G G G



2) Porovnání NQO1 mRNA a CDS

