

V. Cvičení sekvence

RNDr. Karel Berka, Ph.D.

Univerzita Palackého v Olomouci

Vyhledání na NCBI-Entrez

- Check [NCBI](#) site and how to ask [ENTREZ](#), especially how to use [fields](#).
- What is the latest version of dna with ac number [NM_000770.3](#) ?
- Rad32 is the name of the Schizosaccharomyces Pombe gene. What is the name of it's mouse ortholog?
- What is the ac number of Saccharomyces cerevisiae XRS2 gene (complete cds)?
- What is the name of the first author, who sequenced dna with locus name [NM_000770.3](#)?
- What is the origin of dna sequence with gi 54290088?
- What is the size of the protein, which is encoded by dna with id BC078973.1?
- What is the name of biochemical pathway where enzyme with ac NP_059144 appears?
- What is the exact size of the Humo Sapiens genome?

Sekvenování - úkol

You have just sequenced an Arabidopsis cDNA fragment cloned in the EcoRI site of pBluescriptII SK-vector and obtained the following raw sequence:

```
CCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAAT
TCGGTGACCCCGGCAAAGCTTGCTTAATCCGAAGACGTTTCTGTTTCATC
TTCTTAAATCCGGGCCAACNGCGTTTACGAGACTAAACGCGTTTCTCTTT
AGGGCTTAATTATTATCCAGAGATGGCTCATCATAGCAAATGTTTACAAA
CGTTGGATTTAGCTTGTAAGAGCTGAGATCTCGTGGCTTGTTTGTGAAG
CTTTTGGAGGCAATACTTAAAGCTGGAAACAGAATGAACGCGGGTACCGC
GAGAGGAAACGCTCAAGCGTTTAATCTAACCGCGCTTTTGAAGCTTTCGG
ATGTTAAAAGCGTTGATGGGAAGACTTCTTTGCTTAACTTTGTAGTGGAG
GAAGTTGTTAGATCGGAAGGAAAACGTTGTGTTATGAATAGAAGAAGCCA
TAGCTTAACACGAAGCGGTAGTAGTAACAATGGTGGTAATAGTAGTC
TTCAGGTTATGTGCGAAAGAAGAGCAAGAGAAAGAGTACTTGAAGCTTGGT
TTACCAGTTGTTGGTGGATTGAGCTCTGAGTTTTTCAAACGTGAAGAAAGC
TGCTTGTGTGGACTATGAAACGGTTGTTGCAACTTGTTCTGCTCTTGCGG
TTAGAGCGAAAGATGCGAAAACGGTGATTGGAGAATGTGAAGATGGAGAA
GGAGGGAGGTTTGTGAAAACGATGATGACGTTTCNTGATTTCGGTAGAGGA
AGAGGTGAAAATAGCGAAAGGTGAAGAGAGGAAAGTGATCCCTGA
```

- Copy the sequence into a text file and convert it to the FASTA format.
- Perform **restriction analysis** to identify sites from the vector polylinker and remove vector sequences so that you have only clean cDNA left. Check the vector-insert boundary using the map provided below. Save the clean cDNA sequence (in FASTA format) as a text file.
- What is the first letter of your insert?

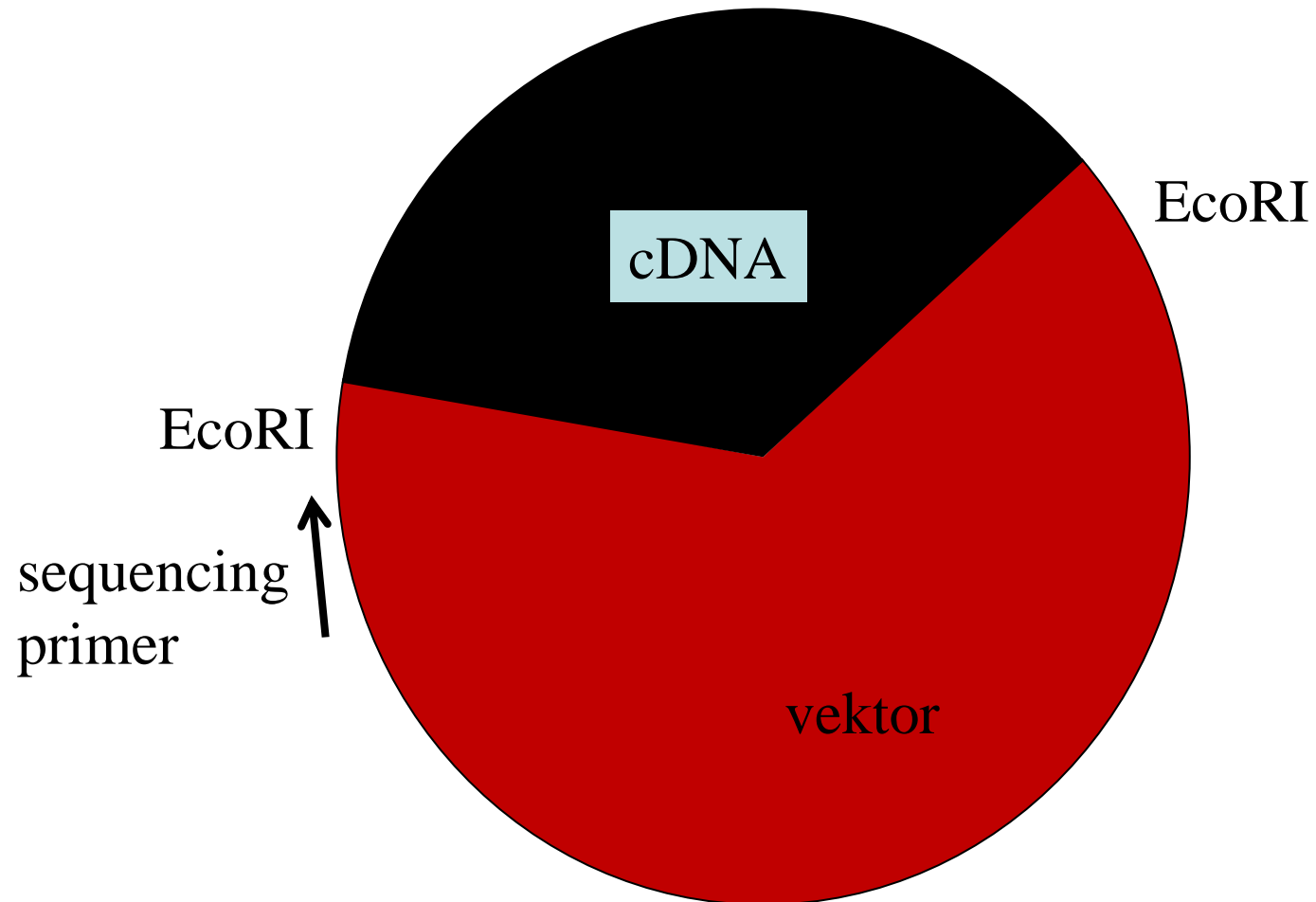
Pojmy

- Arabidopsis
 - (huseníček), často používaný rostlinný model
- Vektor
 - vnášená DNA
- cDNA – komplementární DNA
 - vznikne přepisem mRNA pomocí reverzní transkriptázy
- EcoRI
 - restriční endonukleáza, která štěpí dsDNA v sekvenci:



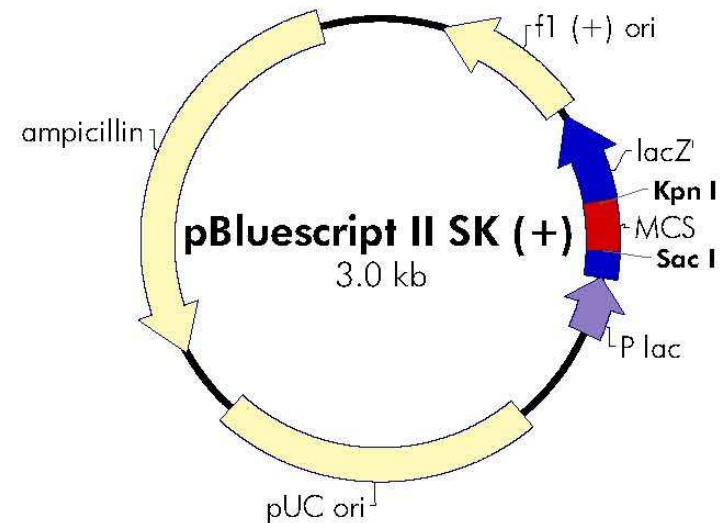
G A A T T C
C T T A A G

Klonování - příklad

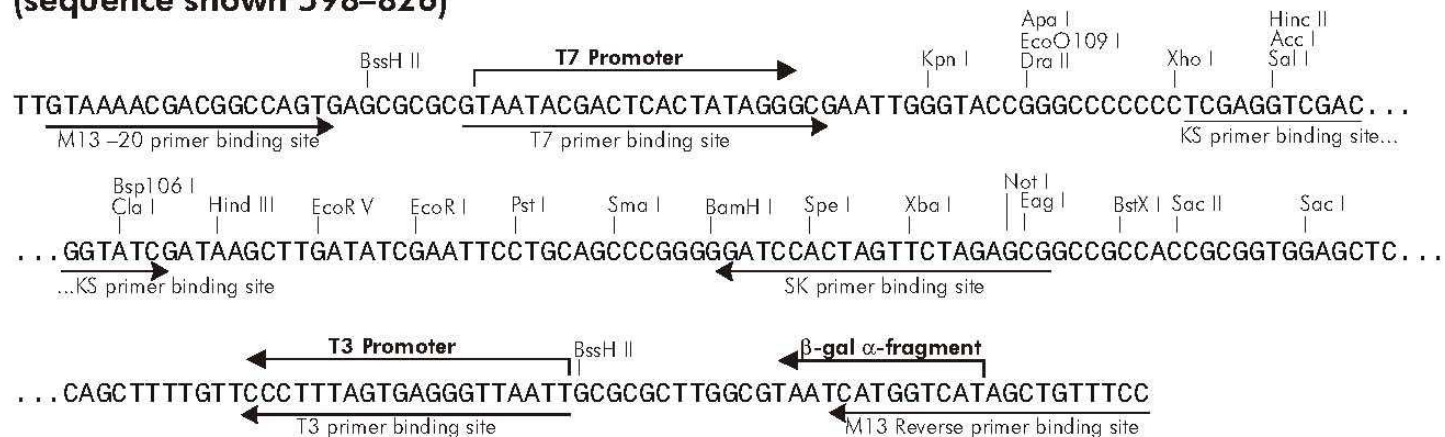


vektor pBlueskript II SK-

f1 (+) origin 135–441
 β -galactosidase α -fragment 460–816
multiple cloning site 653–760
lac promoter 817–938
pUC origin 1158–1825
ampicillin resistance (*bla*) ORF 1976–2833



pBluescript II SK (+/-) Multiple Cloning Site Region (sequence shown 598–826)



Hledání restriční místa

vector	EcoRI	insert
..CCCCGGGCTGCAGG	AATTCGGTGACCCC..	
..GGGGCCCGACGTC	CTTAA	GCCACTGGGG..

Výsledek

>gi cDNA	sekvence vektoru	restrikční místo
CCGCGGTGGCGGGCCGCTCTAGAAGTAGTGGATCCCCCGGGCTGCAG GGAAT		
TC GGTGACCCCGGCAAAGCTTGCTTAATCCGAAGACGTTTCTGTTTCATC		
TTCTTAAATCCGGGCCAACNGCGTTTACGAGACTAAACGCGTTTCTCTTT		
AGGGCT' vložená cDNA CAGAGATGGCTCATCATAGCAAATGTTTACAAA		
CGTTGGATTTTAGCTTTGTAAGAGCTGAGATCTCGTGGCTTGTTTGTGAAG		
CTTTTGGAGGCAATACTTAAAGCTGGAAACAGAATGAACGCGGGTACCGC		
GAGAGGAAACGCTCAAGCGTTTAATCTAACCGCGCTTTTGAAGCTTTCGG		
ATGTTAAAAGCGTTGATGGGAAGACTTCTTTGCTTAACTTTGTAGTGGAG		
GAAGTTGTTAGATCGGAAGGAAAACGTTGTGTTATGAATAGAAGAAGCCA		
TAGCTTAACACGAAGCGGTAGTAGTAACAATGGTGGTAATAGTAGTC		
TTCAGGTTATGTCGAAAGAAGAGCAAGAGAAAGAGTACTTGAAGCTTGGT		
TTACCAGTTGTTGGTGGATTGAGCTCTGAGTTTTCAAACGTGAAGAAAGC		
TGCTTGTGTGGACTATGAAACGGTTGTTGCAACTTGTCTGCTCTTGCGG		
TTAGAGCGAAAGATGCGAAAACGGTGATTGGAGAATGTGAAGATGGAGAA		
GGAGGGAGGTTTGTGAAAACGATGATGACGTTTCNTGATTCGGTAGAGGA		
AGAGGTGAAAATAGCGAAAGGTGAAGAGAGGAAAGTGATCCCTGA		

Sekvence – cvičení 2

- You have just sequenced the following *Arabidopsis* cDNA fragment.

>cDNA_2_At

```
CAAGCAACCTTTACACATATAGAAGAAGAAAAACACTTCTTTGTTTCTGTCATTAATTCCCTCCCTCTAT
ATATATATATTTAAATCTATTATGACAAAACAATCCAATACTGGATACTTTTTACAACAACATGCACAGA
CAAAGCTGAGATCTCACCTTAAGAAACTAATGAGATTGAGTTATGTTTCGTCTTCATCAGACGAAGAGGTT
GAAGACGACGACGAAGAAGAAGAAGATTGTCTTCGTCCAACGAGTCCAGGAAGAGGTTGTGGCATCATTG
GATTCACTGGAACAGGAACTTATGAGCAGAACTAACCATTGTTCTTTTCGTTTATCATCCCTACTTCTTT
GCAAACCTCTGTCTACTACTCCAAGGAAGTCTCTAACCACCAAGAATATTCTAAACGGATGCGCTTCTTCT
TTAGCCGAGTTTCCATGGAAATACTCTGTGATTTCTTTTACAAGTGATAACGCTACGCTCTCTTGAGCTT
GTACTCTGATGATCTCTTCCTCAGCTCTTTTCAGAAACNTTTTCATCGATTCCGNNAACCTCTGACTGTT
GCTTTCTTCTGTGATTGTTGATTGGACTTGGATTGCTTCGTTGATCTTGGCAATGNNTTGAGGAAAGCTT
GGAGANGTAGCTGCTTAGTACTTCTGAGTCCATCANNAGC
```

- Convert the sequence to antisense (reverse complement) and keep it as a FASTA file.
- Look at the coding capacity of your cDNA in both directions. Select the longest reading frame (now in the sense orientation) and keep it as a FASTA file. (NOTE: you might not necessarily get a reasonable ORF - assume that this was a raw first run sequence, containing possible frameshifts).

Credits

- Při přípravě této přednášky byly použity přednášky:
 - Jan Pačes a Jiří Vondrášek – Bioinformatika (UK Praha)
 - Aplikovaná proteomika (UO Hradec Králové)