

CG020 Genomika Bi7201 Základy genomiky

Přednáška 2

Identifikace genů

Jan Hejátko

Funkční genomika a proteomika rostlin,
Mendelovo centrum genomiky a proteomiky rostlin,
Středoevropský technologický institut (CEITEC), Masarykova univerzita, Brno
hejatko@sci.muni.cz, www.ceitec.muni.cz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Literatura

▪ Zdrojová literatura ke kapitole 2

- Plant Functional Genomics, ed. Erich Grotewold, 2003, Humana Press, Totowa, New Jersey
- Majoros, W.H., Pertea, M., Antonescu, C. and Salzberg, S.L. (2003) GlimmerM, Exonomy, and Unveil: three ab initio eukaryotic gene finders. *Nucleic Acids Research*, **31**(13).
- Singh, G. and Lykke-Andersen, J. (2003) New insights into the formation of active nonsense-mediated decay complexes. *TRENDS in Biochemical Sciences*, **28** (464).
- Wang, L. and Wessler, S.R. (1998) Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize R gene. *Plant Cell*, **10**, (1733)
- de Souza et al. (1998) Toward a resolution of the introns early/late debate: Only phase zero introns are correlated with the structure of ancient proteins *PNAS*, **95**, (5094)
- Feuillet and Keller (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution *Ann Bot*, **89** (3-10)
- Frobius, A.C., Matus, D.Q., and Seaver, E.C. (2008). Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS One* **3**, e4004



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven pomocí technologie metylačního filtrování
 - EST knihovny
 - přímá a reverzní genetiky



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*

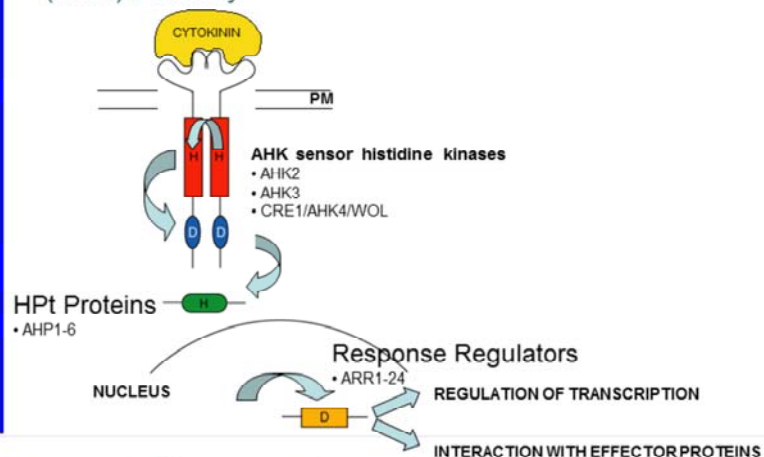


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21*

Recent Model of the CK Signaling via Multistep Phosphorelay (MSP) Pathway



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*
- Mutant identifikován vyhledáváním v databázi inzerčních mutantů (SINS-sequenced insertion site) pomocí programu BLAST



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

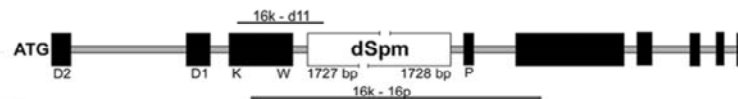
Identifikace role genu *ARR21* – izolace inz. mutanta

- vyhledávání v databázi inzerčních mutantů (SINS)

```
Insert_SINS: 01_09_64
Query: 80 tcttagcggttcgatgagcgtaaccatacttgacaanagagaacgtagccagccattacagg 139
          |||
Sbjct: 58319 tcttagcggttcgatgagcgtaaccatacttgacaagagagaacgtagccagccattacagg 58378
Arr21: 1830
```

```
Insert_SINS: 01_09_64
Query: 140 ttgtgatetcttgcaaaaatgttttggatttactgt 179
          |||
Sbjct: 58379 ttgtgatetcttgcaaaaatgttttggatttactgt 58418
Arr21: 1890
```

- lokalizace inzerce *dSpm* v genomové sekvenci *ARR21* pomocí sekvenace PCR produktů



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Operativní
SP Vzdělávání
pro konkurenceschopnost



EVROPSKÉ VZDĚLÁVÁNÍ

pro konkurenceschopnost

Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*
- Mutant identifikován vyhledáváním v databázi inzerčních mutantů (SINS-sequenced insertion site) pomocí programu BLAST
- Exprese *ARR21* u standardního typu a Inhibice exprese u inzerčního mutantu potvrzena na úrovni RNA



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

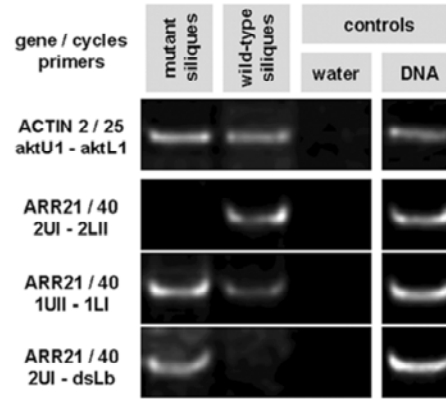
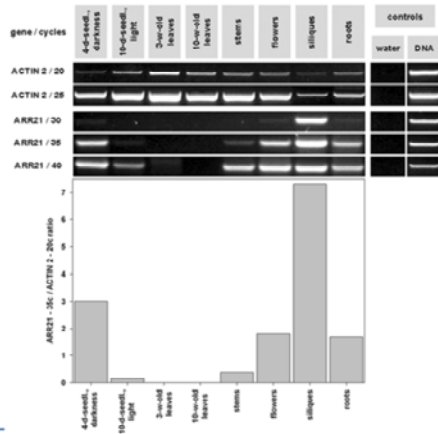
Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu

ARR21 – analýza exprese

Standardní typ

Inzerční mutant



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*
- Mutant identifikován vyhledáváním v databázi inzerčních mutantů (SINS-sequenced insertion site) pomocí programu BLAST
- Exprese *ARR21* u standardního typu a Inhibice exprese u inzerčního mutantu potvrzena na úrovni RNA
- Analýza fenotypu inzerčního mutantu



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

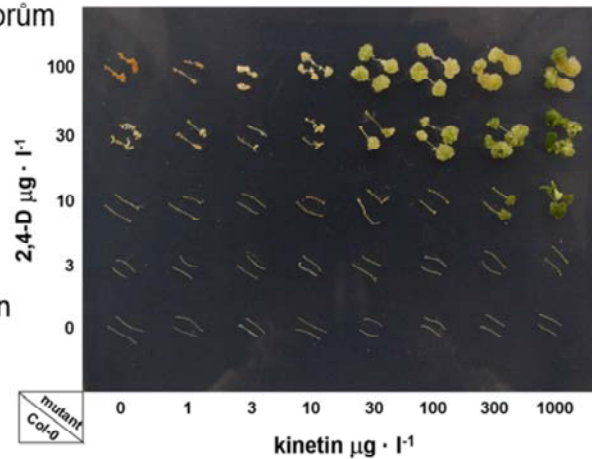
Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21* – analýza fenotypu mutantu

- Analýza citlivosti k regulátorům růstu rostlin

- 2,4-D a kinetin
- etylén
- světlo různých vlnových délek

- Doba kvetení i počet semen nezměněn



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21* – příčiny absence fenotypu

- Funkční redundance v rámci genové rodiny?



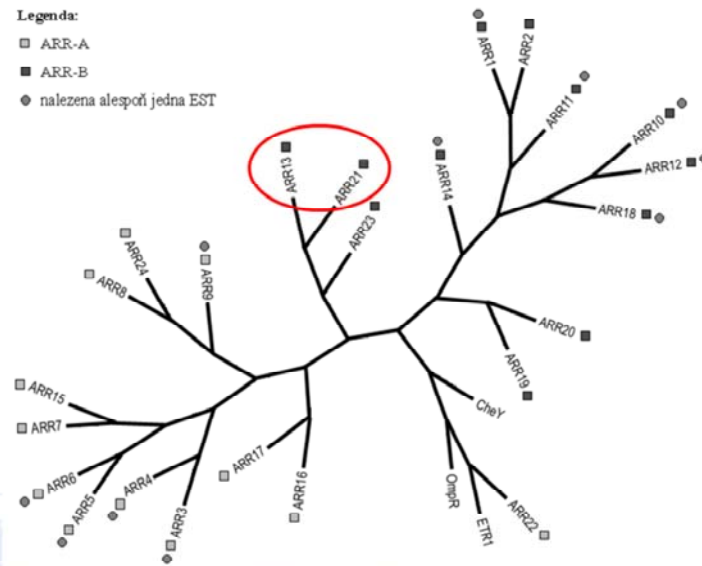
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21* – příbuznost ARR genů

Legenda:

- ARR-A
- ARR-B
- nalezena alespoň jedna EST



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost



ZVOJE VZDĚLÁVÁNÍ

zvyšuje je kvalifikovanost
zlepšuje sociální fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21* – příčiny absence fenotypu

- Funkční redundance v rámci genové rodiny?
- Fenotypový projev pouze za velmi specifických podmínek (?)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace role genu *ARR21* – shrnutí

- Gen *ARR21* identifikován pomocí srovnávací analýzy genomu *Arabidopsis*
- Na základě analýzy sekvence byla předpovězena jeho funkce
- Byla prokázána místně specifická exprese genu *ARR21* na úrovni RNA
- Identifikace funkce genu pomocí inzerční mutagenese v případě *ARR21* ve vývoji *Arabidopsis* byla neúspěšná, pravděpodobně v důsledku funkční redundance v rámci genové rodiny



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání

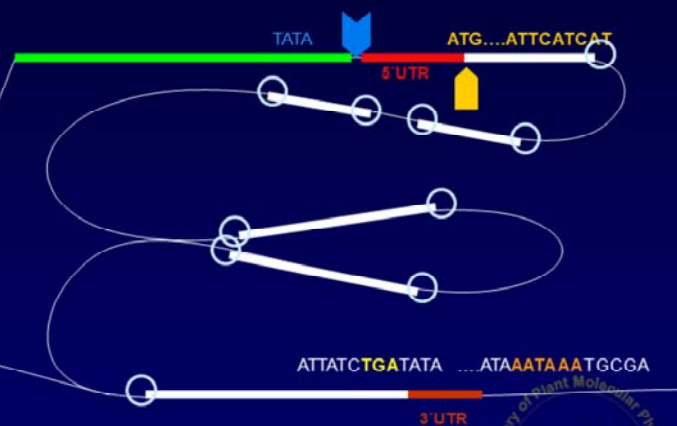


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

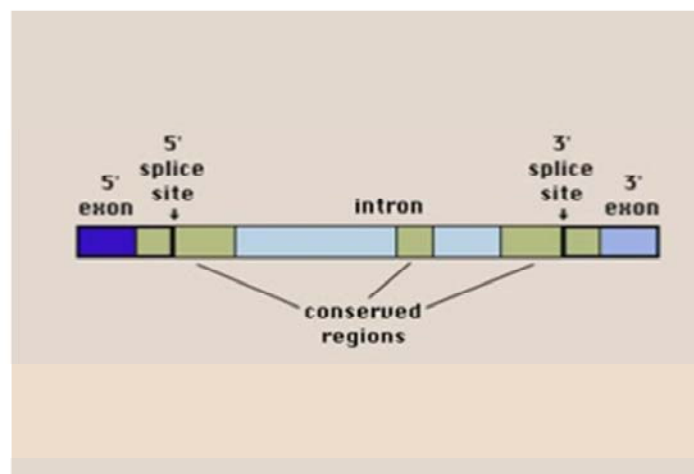
Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Struktura genů

- promotor
- počátek transkripce
- 5'UTR
- počátek translace
- místa sestřihu
- stop kodon
- 3'UTR
- polyadenylační signál



Sestřih RNA



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace genů *ab initio*

- zanedbání 5' a 3' UTR
- identifikace počátku translace (ATG) a stop kodonu (TAG, TAA, TGA)
- nalezení donorových (většinou GT) a akceptorových (AG) míst sestřihu
- většina ORF není skutečně kódujícími sekvencemi – u *Arabidopsis* je asi 350 mil. ORF na každých 900 bp (!)
- využití různých statistických modelů (např. Hidden Markov Model, HMM, viz doporučená studijní literatura, Majoros et al., 2003) k posouzení a ohodnocení váhy identifikovaných donorových a akceptorových míst



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Predikce míst sestřihu

- programy pro predikci míst sestřihu (specifita přibližně 35%)
 - GeneSplicer (http://www.tigr.org/tdb/GeneSplicer/gene_spl.html)
 - SplicePredictor (<http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi>)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Predikce míst sestřihu

BCB @ ISU Bioinformatics 2 Download Help Tutorial References Contact
Go

SplicePredictor

- a method to identify potential splice sites in (plant) pre-mRNA by sequence inspection using Bayesian statistical models
(click [here](#) to access the older method using logitlinear models)

Sequences should be in the one-letter-code ({a,b,c,g,h,k,m,n,r,s,t,u,w,y}), upper or lower case; all other characters are ignored during input. Multiple sequence input is accepted in **FASTA** format (sequences separated by identifier lines of the form ">SQ:name_of_sequence comments") or in **GenBank** format.

Paste your genomic DNA sequence here:

```
GAGGAGGCACAAAATGACGAATATACAAAATGATCTTAAACAGCTAAACTATATTGGACATTTTTTCGATCTCAGATATA  
AAAGATTTTCATTCAATATAAATACTGGATAAATACTTTATTTTCTTTAGTTTATTAACAAAAAACCTCTAATAAAT  
ACGAGTTTAAAGTCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAAACGATAAAGTTTACAAAAGTAATATCC  
AAGTATCTCATAGTCAACATATATATAGTAATAATTAGTTGACGTATAAGAAAAATAAAATAAATAAATTAGTATCTTAT  
TTTGGGTGGTGTGACTGGTGAATGCTGCAGAAATGCTCGGCAAAATGGAACCATATCCCAAGACATGGGTTTTAGAT
```

... or upload your sequence file (specify file name):

... or type in the GenBank accession number of your sequence:



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Predikce míst sestřihu

What do the output columns mean?

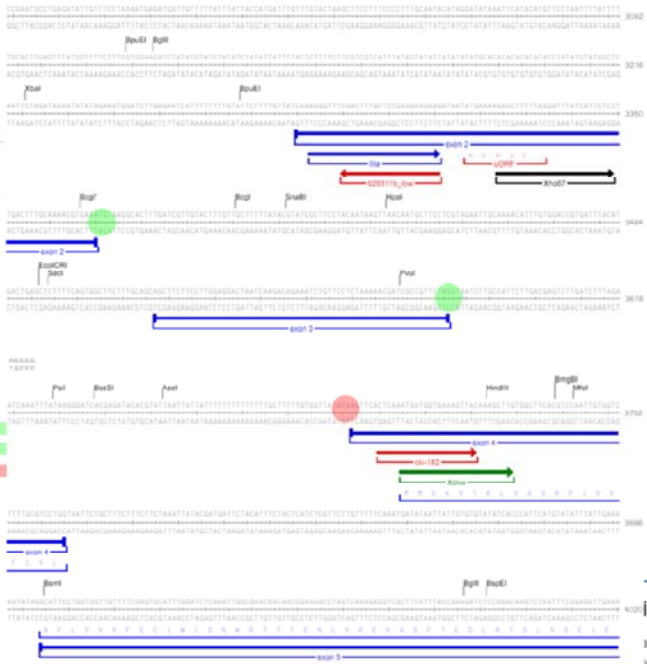
SplicePredictor, Version of February 13, 2005.
Date run: Wed Nov 9 11:30:16 2005

Species: Homo sapiens
Model: 2-class Bayesian
Prediction cutoff (2 ln[P]): 3.00
Local pruning: on
Non-canonical sites: not scored

Sequence: 1: your+sequence, from 1 to 9430.

Potential splice sites

t	q	loc	sequence	P	o	etc	gamma	+PS+D*
A	---	75	ttttttgagatctAGat	0.973	7.14	0.000	0.000	7 15 3 11
A	---	134	attattttctcttAGAt	0.999	14.80	0.000	0.000	7 15 3 11
A	---	300	gattttgttttAGGc	0.977	7.48	0.000	0.000	7 15 3 11
A	---	788	ttgtttctgtctAGGc	0.984	8.26	0.000	0.000	7 15 3 11
A	---	848	tattttttgaaatAGAt	0.968	6.80	0.000	0.000	7 15 3 11
A	---	1051	caatttttttttAGGc	0.930	5.19	0.000	0.000	7 15 3 11
A	---	1211	ttatttttttttAGAt	0.998	12.14	0.000	0.000	7 15 3 11
A	---	1373	ttctctctctctAGGc	0.999	13.17	0.000	0.000	7 15 3 11
A	---	1487	ttttatattttAGGc	0.983	4.04	0.000	0.000	7 15 3 11
A	---	1581	atgtttttgtttAGGc	0.982	8.03	0.388	0.000	7 15 3 11
A	---	1791	ggttttggaaatAGGc	0.958	4.10	0.000	0.000	7 15 3 11
A	---	2440	tatttttttttAGAt	0.939	5.46	0.000	0.000	7 15 3 11
A	---	2478	caottttttttAGAt	0.942	5.59	0.000	0.000	7 15 3 11
D	---->	2544	aaGTtaGta	0.909	4.41	0.885	1.903	15 15 5 5
A	---	2572	tttttttttttAGGc	0.930	5.16	0.000	0.000	7 15 3 11
A	---	2763	ctcaatttttttAGGc	0.873	3.86	0.185	0.000	11 15 5 11
A	---	2782	tttttttttttAGGc	0.952	5.98	0.220	0.000	11 15 5 11
A	---	3022	tttttttttttAGGc	0.958	6.16	0.221	0.000	11 15 5 11
A	---	3048	cttttttttttAGGc	0.973	7.10	0.229	0.000	11 15 5 11
A	---	3177	agtttttttttAGGc	0.968	8.74	0.000	0.000	7 15 3 11
A	---	3281	tttttttttttAGGc	0.933	10.63	0.000	0.000	8 15 3 11
D	---->	3282	aaTTtaGta	0.933	5.28	0.955	1.849	15 15 5 5
A	---	3414	gatttttttttAGGc	0.916	4.71	0.193	0.000	12 15 5 11
A	---	3481	gatttttttttAGGc	0.910	4.41	0.194	0.000	7 15 3 11
D	---->	3482	aaTTtaGta	0.910	5.25	0.000	1.849	11 15 3 11
A	---	3483	tttttttttttAGGc	0.987	4.08	0.000	0.000	7 15 3 11
A	---	4234	attattttctcttAGAt	0.958	17.42	0.000	0.000	4 15 3 11
A	---	4351	tttttttttttAGGc	0.991	9.42	0.000	0.000	7 15 3 11
A	---	4633	gtttttttttAGGc	0.979	1.87	0.000	0.000	7 15 3 11
A	---	4974	cttttttttttAGGc	0.952	5.98	0.000	0.000	7 15 3 11
A	---	5004	tttttttttttAGGc	0.984	11.17	0.000	0.000	7 15 3 11
D	---->	5034	aaTTtaGta	0.921	3.04	0.387	0.000	11 15 3 11
D	---->	5384	tttttttttttAGGc	0.941	3.54	0.478	0.090	13 15 3 11
A	---	5405	aaatttttttAGGc	0.894	4.26	0.000	0.000	7 15 3 11
A	---	5441	cttttttttttAGGc	0.995	10.43	0.387	0.000	11 15 3 11
A	---	5472	tttttttttttAGGc	0.965	6.42	0.478	0.090	13 15 3 11
D	---->	5745	aaTTtaGta	0.991	9.48	0.990	1.806	13 15 3 11
A	---	5808	caatatttttAGGc	0.948	5.83	0.458	0.000	11 15 3 11
A	---	6125	ggttttttttAGGc	0.999	13.29	0.000	0.000	12 15 3 11
A	---	6552	ggttttttttAGGc	0.938	5.42	0.000	0.000	7 15 3 11



Identifikace genů *ab initio*

- programy pro predikci míst sestřihu (specifita přibližně 35%)
 - GeneSplicer (http://www.tigr.org/tdb/GeneSplicer/gene_spl.html)
 - SplicePredictor (<http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi>)
 - NetGene2 (<http://www.cbs.dtu.dk/services/NetGene2/>)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Predikce míst sestřihu

Prediction done

***** NetGene2 v. 2.4 *****

The sequence: Sequence has the following composition:

Length: 9490 nucleotides.
31.8% A, 17.0% C, 19.6% G, 31.7% T, 0.0% X, 36.5% G+C

Donor splice sites, direct strand

pos 5'>3'	phase	strand	confidence	5'	exon	intron	3'
1704	0	+	0.87	TTGGAAAGC	AGTAAATT		
1904	0	+	0.93	CGGTGACGG	GTAGACAT		
3182	1	+	1.00	GGGCTTATG	GAATCTGG		
3780	1	+	1.00	TGGGAGGAG	GAATCTGG		
4134	0	+	0.74	TCAACACAG	GTCTTAAA		
4619	1	+	0.74	AGCAAGAA	GTCTTCTTC		
4915	0	+	0.94	CGTCTCTTC	GAATCTGG		
5394	0	+	0.87	TCTCAGCAA	GGATATTT		
5384	1	+	1.00	GATTTGGTG	GAAGACTCT		
5809	1	+	1.00	TATCCTAAG	GTCTGCCAA		
4857	0	+	1.00	GCAGCTCTT	GAAGACTCT		
4094	1	+	0.74	CTCTTCACA	GAAGACTCT		
7369	0	+	1.00	GGACTGCCA	GTAAAGTTAA		
7884	0	+	0.74	GAACAAATG	GTAGATGAA		
9323	0	+	0.74	GAAGATTAG	GTCTTCTTC		

Donor splice sites, complement strand

pos 3'>5'	pos 5'>3'	phase	strand	confidence	5'	exon	intron	3'
1213	0	+	0.59	TATTTTTTA	TTATGGAG			
1221	2	+	0.87	AGTTATGAG	ACAAGAATCG			
1373	0	+	0.71	TCTTACAG	GTACAGAG			
1487	1	+	0.81	ATATTGATG	TGGACATTA			
3284	0	+	0.87	GTATGCAAG	GGTCTGCAC			
4254	0	+	1.00	TCTTCTTC	ATCCGACAT			
4832	2	+	0.54	AAANTGGG	TCCATGGG			
5004	0	+	0.94	TTTTGGCC	AGATACAC			
5472	1	+	0.96	AAAATTAC	GTCTGTCAA			
6135	0	+	1.00	ATTATTATG	GTAAAGTTAA			
6490	1	+	0.90	AAATTACAG	TGGTGGGA			
6744	0	+	0.59	TGTCAACAG	TTCCGAGAG			
7447	0	+	0.96	TTCCGACAG	ATCCGAGAA			
7780	2	+	0.74	TCCATTTC	ATACAGACA			
7786	2	+	0.92	TCGATACAG	AAACATGCA			

Acceptor splice sites, direct strand

pos 5'>3'	phase	strand	confidence	5'	intron	exon	3'
1213	0	+	0.59	TATTTTTTA	TTATGGAG		
1221	2	+	0.87	AGTTATGAG	ACAAGAATCG		
1373	0	+	0.71	TCTTACAG	GTACAGAG		
1487	1	+	0.81	ATATTGATG	TGGACATTA		
3284	0	+	0.87	GTATGCAAG	GGTCTGCAC		
4254	0	+	1.00	TCTTCTTC	ATCCGACAT		
4832	2	+	0.54	AAANTGGG	TCCATGGG		
5004	0	+	0.94	TTTTGGCC	AGATACAC		
5472	1	+	0.96	AAAATTAC	GTCTGTCAA		
6135	0	+	1.00	ATTATTATG	GTAAAGTTAA		
6490	1	+	0.90	AAATTACAG	TGGTGGGA		
6744	0	+	0.59	TGTCAACAG	TTCCGAGAG		
7447	0	+	0.96	TTCCGACAG	ATCCGAGAA		
7780	2	+	0.74	TCCATTTC	ATACAGACA		
7786	2	+	0.92	TCGATACAG	AAACATGCA		



MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost

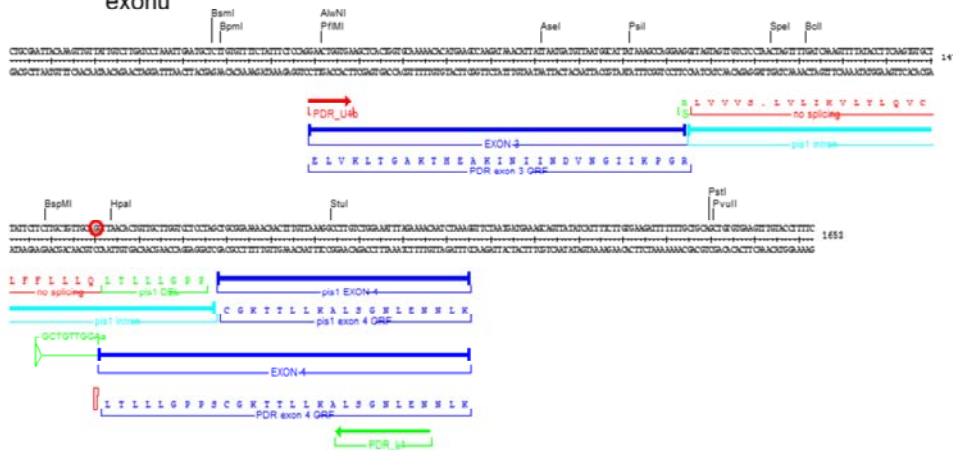


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Sestřih RNA a adaptace

- odchylky rozpoznávání míst sestřihu u rostlin v praxi - příklad vývojové plasticity (nejen) rostlin
 - identifikace mutanta s bodovou mutací (tranzice G→A) přesně v místě sestřihu na 5' konci 4. exonu



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

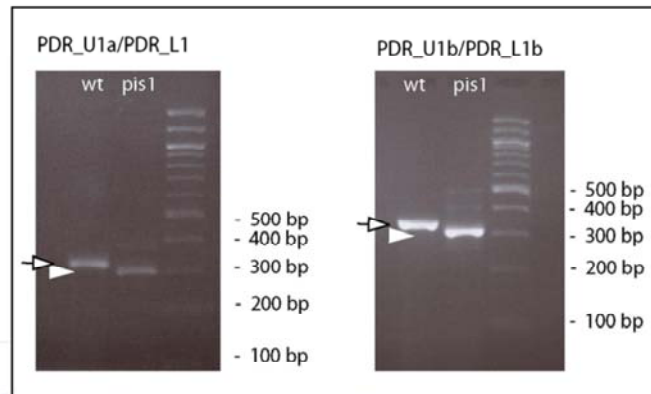
OP Vzdělávání
pro konkurenceschopnost



a státním rozpočtem České republiky

Sestřih RNA a adaptace

- identifikace mutanta s bodovou mutací (tranzice G→A) přesně v místě sestřihu na 5' konci 4. exonu
- analýza pomocí RT PCR prokázala přítomnost fragmentu kratšího než by odpovídalo cDNA po normálním sestřihu



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost

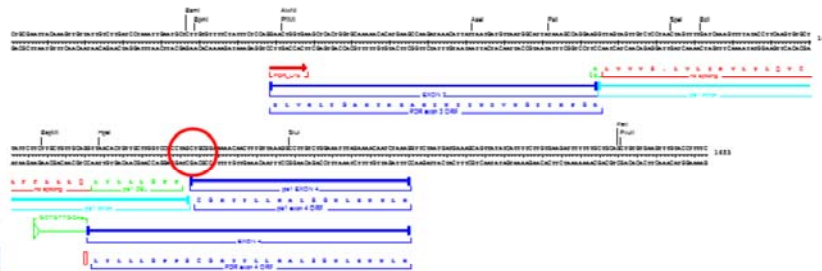


ROZVOJE VZDĚLÁVÁNÍ

pro prezentaci je poskytnuta
Evropským sociálním fondem
a státním rozpočtem České republiky

Sestřih RNA a adaptace

- odchylky rozpoznávání míst sestřihu u rostlin v praxi - příklad vývojové plasticity (nejen) rostlin
 - identifikace mutantu s bodovou mutací (tranzice G→A) přesně v místě sestřihu na 5' konci 4. exonu
 - analýza pomocí RT PCR prokázala přítomnost fragmentu kratšího než by odpovídalo cDNA po normálním sestřihu
 - sekvenace tohoto fragmentu pak ukázala na alternativní sestřih s využitím nejbližšího možného místa sestřihu v exonu 4



MINISTERSTVO VYŠŠÍ ŠKOLY, Mládeží a tělovýchovy

OP Vzdělávání pro konkurenceschopnost



DĚLAVÁNÍ

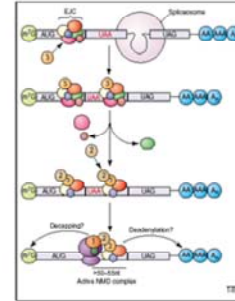
Autoremováno

Evropským sociálním fondem

a státním rozpočtem České republiky

Sestřih RNA a adaptace

- odchylky rozpoznávání míst sestřihu u rostlin v praxi - příklad vývojové plasticity (nejen) rostlin
 - identifikace mutanta s bodovou mutací (tranzice G→A) přesně v místě sestřihu na 5' konci 4. exonu
 - analýza pomocí RT PCR prokázala přítomnost fragmentu kratšího než by odpovídalo cDNA po normálním sestřihu
 - sekvenace tohoto fragmentu pak ukázala na alternativní sestřih s využitím nejbližšího možného místa sestřihu v exonu 4
 - existence podobných obranných mechanismů prokázána i u jiných organismů (např. nestabilita mutantní mRNA se vznikem předčasného stopkodonu (> 50-55 bp před normálním stop kodonem) u eukaryot, viz doporučená studijní literatura, Singh and Lykke-Andersen, 2003)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace genů *ab initio*

- programy pro predikci exonů
 - 4 typy exonů (podle polohy):
 - iniciační
 - vnitřní
 - terminální
 - jednoduché
 - programy kromě rozpoznávání míst sestřihu zohledňují i strukturu jednotlivých typů exonů
- iniciační:
 - Genescan (<http://genes.mit.edu/GENSCAN.html>)
 - GeneMark.hmm (<http://opal.biology.gatech.edu/GeneMark/>)
- interní:
 - MZEF (<http://rulai.cshl.org/tools/genefinder/>)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace genů *ab initio*

GENSCANW output for sequence CKII

```

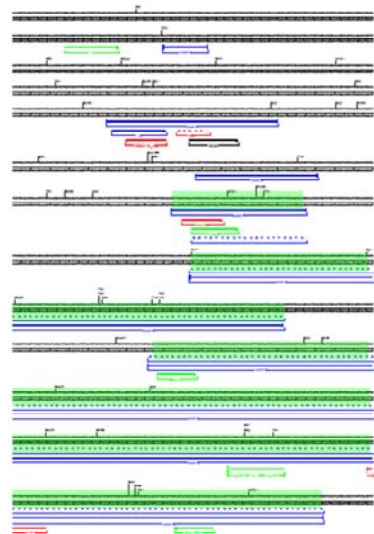
GENSCAN 1.0  Date run: 10-Nov-105  Time: 02:24:26
Sequence CKII : 9490 bp : 36,53% C+G : Isochores 1 ( 0 - 43 C+G)
Parameter matrix: Arabidopsis.mat
Predicted genes/exons:

Gn.Ex Type S .Begin .End .Len Fr Ph I/Ac Do/T CodRg P... Tscr..
-----
1.00 Prom + 1497 1536 40 - - - - - - - - - - -3.85
1.01 Init + 3708 3764 57 2 0 43 51 37 0.489 -4.038
1.02 Intr + 3994 4133 240 2 0 -3 7 327 0.713 17.32
1.03 Intr + 4255 4914 660 0 0 86 59 296 0.771 22.57
1.04 Intr + 5605 5303 379 0 1 70 91 343 0.772 31.41
1.05 Intr + 5473 6056 584 2 2 38 99 582 0.722 50.76
1.06 Intr + 6136 7368 1233 0 0 68 108 655 0.977 56.86
1.07 Term + 7448 7660 213 1 0 43 35 212 0.999 12.65
1.08 PlyA + 7910 7915 6 - - - - - - - - - -0.45

2.03 PlyA - 7976 7971 6 - - - - - - - - - -4.83
2.02 Term - 8793 8050 744 0 0 107 37 542 0.997 48.46
2.01 Init - 9253 8936 318 1 0 105 73 386 0.999 41.18

Suboptimal exons with probability > 0.100

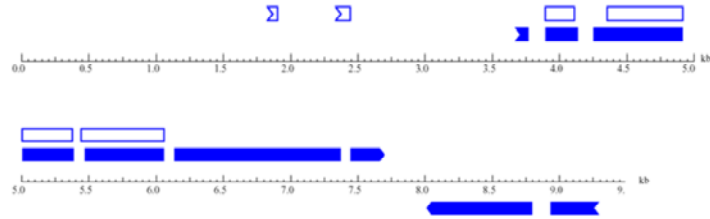
Exon Type S .Begin .End .Len Fr Ph B/Ac Do/T CodRg P... Tscr..
-----
8.001 Init + 1867 1905 39 0 0 64 40 57 0.298 3.74
8.002 Init + 2374 2442 69 0 0 55 95 -11 0.132 2.40
8.003 Intr + 3894 4110 217 2 1 -3 -34 307 0.177 11.55
8.004 Intr + 4352 4914 563 0 2 75 59 338 0.187 26.20
8.005 Intr + 8056 8279 276 0 0 70 8 238 0.212 22.99
8.006 Intr + 5442 6056 615 2 0 95 99 589 0.208 57.32
    
```



EVROPSKÁ UNIE
 ESF
 INSTITUT MIKROBIOLOGIE
 ČESKÉ AKADEMIE VĚD
 UNIVERZITA JYVÄSKYLÄ
 FENOLINEN ET AL. 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 2681, 2682, 2683, 2684, 2685, 2686, 2687, 2688, 2689, 2690, 2691, 2692, 2693, 2694, 2695, 2696, 2697, 2698, 2699, 2700, 2701, 2702, 2703, 2704, 2705, 2706, 2707, 2708, 2709, 2710, 2711, 2712, 2713, 2714, 2715, 2716, 2717, 2718, 2719, 2720, 2721, 2722, 2723, 2724, 2725, 2726, 2727, 2728, 2729, 2730, 2731, 2732, 2733, 2734, 2735, 2736, 2737, 2738, 2739, 2740, 2741, 2742, 2743, 2744, 2745, 2746, 2747, 2748, 2749, 2750, 2751, 2752, 2753, 2754, 2755, 2756, 2757, 2758, 2759, 2760, 2761, 2762, 2763, 2764, 2765, 2766, 2767, 2768, 2769, 2770, 2771, 2772, 2773, 2774, 2775, 2776, 2777, 2778, 2779, 2780, 2781, 2782, 2783, 2784, 2785, 2786, 2787, 2788, 2789, 2790, 2791, 2792, 2793, 2794, 2795, 2796, 2797, 2798, 2799, 2800, 2801, 2802, 2803, 2804, 2805, 2806, 2807, 2808, 2809, 2810, 2811, 2812, 2813, 2814, 2815, 2816, 2817, 2818, 2819, 2820, 2821, 2822, 2823, 2824, 2825, 2826, 2827, 2828, 2829, 2830, 2831, 2832, 2833, 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843, 2844, 2845, 2846, 2847, 2848, 2849, 2850, 2851, 2852, 2853, 2854, 2855, 2856, 2857, 2858, 2859, 2860, 2861, 2862, 2863, 2864, 2865, 2866, 2867, 2868, 2869, 2870, 2871, 2872, 2873, 2874, 2875, 2876, 2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885, 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895, 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905, 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914, 2915, 2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924, 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934, 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944, 2945, 2946, 2947, 2948, 2949, 2950, 2951, 2952, 2953, 2954, 2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963, 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973, 2974, 2975, 2976, 2977, 2978, 2979, 2980, 2981, 2982, 2983, 2984, 2985, 2986, 2987, 2988, 2989, 2990, 2991, 2992, 2993, 2994, 2995, 2996, 2997, 2998, 2999, 3000, 3001, 3002, 3003, 3004, 3005, 3006, 3007, 3008, 3009, 3010, 3011, 3012, 3013, 3014, 3015, 3016, 3017, 3018, 3019, 3020, 3021, 3022, 3023, 3024, 3025, 3026, 3027, 3028, 3029, 3030, 3031, 3032, 3033, 3034, 3035, 3036, 3037, 3038, 3039, 3040, 3041, 3042, 3043, 3044, 3045, 3046, 3047, 3048, 3049, 3050, 3051, 3052, 3053, 3054, 3055, 3056, 3057, 3058, 3059, 3060, 3061, 3062, 3063, 3064, 3065, 3066, 3067, 3068, 3069, 3070, 3071, 3072, 3073, 3074, 3075, 3076, 3077, 3078, 3079, 3080, 3081, 3082, 3083, 3084, 3085, 3086, 3087, 3088, 3089, 3090, 3091, 3092, 3093, 3094, 3095, 3096, 3097, 3098, 3099, 3100, 3101, 3102, 3103, 3104, 3105, 3106, 3107, 3108, 3109, 3110, 3111, 3112, 3113, 3114, 3115, 3116, 3117, 3118, 3119, 3120, 3121, 3122, 3123, 3124, 3125, 3126, 3127, 3128, 3129, 3130, 3131, 3132, 3133, 3134, 3135, 3136, 3137, 3138, 3139, 3140, 3141, 3142, 3143, 3144, 3145, 3146, 3147, 3148, 3149, 3150, 3151, 3152, 3153, 3154, 3155, 3156, 3157, 3158, 3159, 3160, 3161, 3162, 3163, 3164, 3165, 3166, 3167, 3168, 3169, 3170, 3171, 3172, 3173, 3174, 3175, 3176, 3177, 3178, 3179, 3180, 3181, 3182, 3183, 3184, 3185, 3186, 3187, 3188, 3189, 3190, 3191, 3192, 3193, 3194, 3195, 3196, 3197, 3198, 3199, 3200, 3201, 3202, 3203, 3204, 3205, 3206, 3207, 3208, 3209, 3210, 3211, 3212, 3213, 3214, 3215, 3216, 3217, 3218, 3219, 3220, 3221, 3222, 3223, 3224, 3225, 3226, 3227, 3228, 3229, 3230, 3231, 3232, 3233, 3234, 3235, 3236, 3237, 3238, 3239, 3240, 3241, 3242, 3243, 3244, 3245, 3246, 3247, 3248, 3249, 3250, 3251, 3252, 3253, 3254, 3255, 3256, 3257, 3258, 3259, 3260, 3261, 3262, 3263, 3264, 3265, 3266, 3267, 3268, 3269, 3270, 3271, 3272, 3273, 3274, 3275, 3276, 3277, 3278, 3279, 3280, 3281, 3282, 3283, 3284, 3285, 3286, 3287, 3288, 3289, 3290, 3291, 3292, 3293, 3294, 3295, 3296, 3297, 3298, 3299, 3300, 3301, 3302, 3303, 3304, 3305, 3306, 3307, 3308, 3309, 3310, 3311, 3312, 3313, 3314, 3315, 3316, 3317, 3318, 3319, 3320, 3321, 3322, 3323, 3324, 3325, 3326, 3327, 3328, 3329, 3330, 3331, 3332, 3333, 3334, 3335, 3336, 3337, 3338, 3339, 3340, 3341, 3342, 3343, 3344, 3345, 3346, 3347, 3348, 3349, 3350, 3351, 3352, 3353, 3354, 3355, 3356, 3357, 3358, 3359, 3360, 3361, 3362, 3363, 3364, 3365, 3366, 3367, 3368, 3369, 3370, 3371, 3372, 3373, 3374, 3375, 3376, 3377, 3378, 3379, 3380, 3381, 3382, 3383, 3384, 3385, 3386, 3387, 3388, 3389, 3390, 3391, 3392, 3393, 3394, 3395, 3396, 3397, 3398, 3399, 3400, 3401, 3402, 3403, 3404, 3405, 3406, 3407, 3408, 3409, 3410, 3411, 3412, 3413, 3414, 3415, 3416, 3417, 3418, 3419, 3420, 3421, 3422, 3423, 3424, 3425, 3426, 3427, 3428, 3429, 3430, 3431, 3432, 3433, 3434, 3435, 3436, 3437, 3438, 3439, 3440, 3441, 3442, 3443, 3444, 3445, 3446, 3447, 3448, 3449, 3450, 3451, 3452, 3453, 3454, 3455, 3456, 3457, 3458, 3459, 3460, 3461, 3462, 3463, 3464, 3465, 3466, 3467, 3468, 3469, 3470, 3471, 3472, 3473, 3474, 3475, 3476, 3477, 3478, 3479, 3480, 3481, 3482, 3483, 3484, 3485, 3486, 3487, 3488, 3489, 3490, 3491, 3492, 3493, 3494, 3495, 3496, 3497, 3498, 3499, 3500, 3501, 3502, 3503, 3504, 3505, 3506, 3507, 3508, 3509, 3510, 3511, 3512, 3513, 3514, 3515, 3516, 3517, 3518, 3519, 3520, 3521, 3522, 3523, 3524, 3525, 3526, 3527, 3528, 3529, 3530, 3531, 3532, 3533, 3534, 3535, 3536, 3537, 3538, 3539, 3540, 3541, 3542, 3543, 3544, 3545, 3546, 3547, 3548, 3549, 3550, 3551, 3552, 3553, 3554, 3555, 3556, 3557, 3558, 3559, 3560, 3561, 3562, 3563, 3564, 3565, 3566, 3567, 3568, 3569, 3570, 3571, 3572, 3573, 3574, 3575, 3576, 3577, 3578, 3579, 3580, 3581, 3582, 3583, 3584, 3585, 3586, 3587, 3588, 3589, 3590, 3591, 3592, 3593, 3594, 3595, 3596, 3597, 3598, 3599, 3600, 3601, 3602, 3603, 3604, 3605, 3606, 3607, 3608, 3609, 3610, 3611, 3612, 3613, 3614, 3615, 3616, 3617, 3618, 3619, 3620, 3621, 3622, 3623, 3624, 3625, 3626, 3627, 3628, 3629, 3630, 3631, 3632, 3633, 3634, 3635, 3636, 3637, 3638, 3639, 3640, 3641, 3642, 3643, 3644, 3645, 3646, 3647, 3648, 3649, 3650, 3651, 3652, 3653, 3654, 3655, 3656, 3657, 3658, 3659, 3660, 3661, 3662, 3663, 3664, 3665, 3666, 3667, 3668, 3669, 3670, 3671, 3672, 3673, 3674, 3675, 3676, 3677, 3678, 3679, 3680, 3681, 3682, 3683, 3684, 3685, 3686, 3687, 3688, 3689, 3690, 3691, 3692, 3693, 3694, 3695, 3696, 3697, 3698, 3699, 3700, 3701, 3702, 3703, 3704, 3705, 3706, 3707, 3708, 3709, 3710, 3711, 3712, 3713, 3714, 3715, 3716, 3717, 3718, 3719, 3720, 3721, 3722, 3723, 3724, 3725, 3726, 3727, 3728, 3729, 3730, 3731, 3732, 3733, 3734, 3735, 3736, 3737, 3738, 3739, 3740, 3741, 3742, 3743, 3744, 3745, 3746, 3747, 3748, 3749, 3750, 3751, 3752, 3753, 3754, 3755, 3756, 3757, 3758, 3759, 3760, 3761, 3762, 3763, 3764, 3765, 3766, 3767, 3768, 3769, 3770, 3771, 3772, 3773, 3774, 3775, 3776, 3777, 3778, 3779, 3780, 3781, 3782, 3783, 3784, 3785, 3786, 3787, 3788, 3789, 3790, 3791, 3792, 3793, 3794, 3795, 3796, 3797, 3798, 3799, 3800, 3801, 3802, 3803, 3804, 3805, 3806, 3807, 3808, 3809, 3810, 3811, 3812, 3813, 3814, 3815, 3816, 3817, 3818, 3819, 3820, 3821, 3822, 3823, 3824, 3825, 3826, 3827, 3828, 3829, 3830, 3831, 383

Identifikace genů *ab initio*

GENSCAN predicted genes in sequence 02:56:23



Key: Initial exon Internal exon Terminal exon Single-exon gene Optimal exon Suboptimal exon

MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost



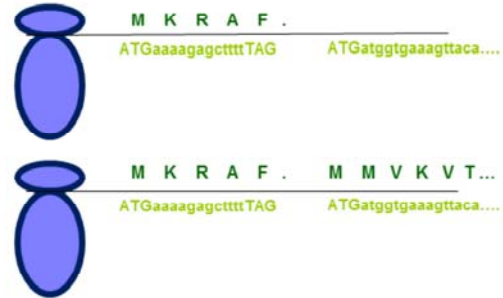
ROZVOJE VZDĚLÁVÁNÍ

Investice je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Regulace translace

• Funkční význam sestřihu v nepřekládaných oblastech - důležitá regulační součást genů

- Translační represe prostřednictvím krátkých ORF v 5'UTR
- Identifikováno např. u kukuřice (Wang and Wessler, 1998, viz doporučená lit.)
- V případě CKI1 pokus prokázat tento způsob regulace genové exprese pomocí transgenních linií nesoucích *uidA* pod kontrolou dvou verzí promotoru, zatím nepotvrzeno

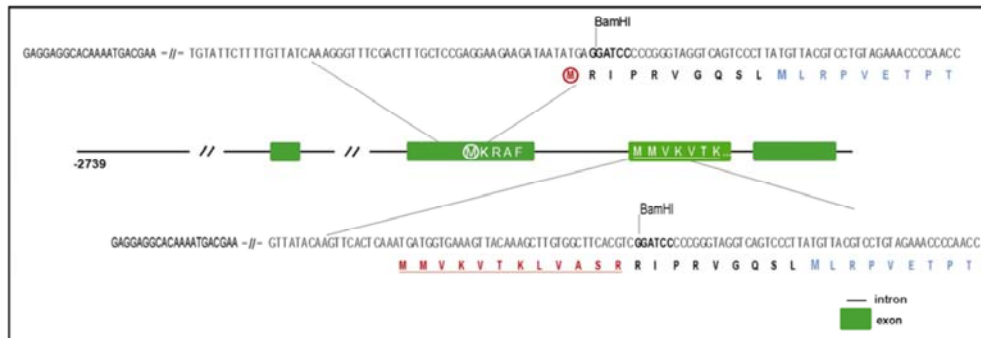


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Regulace translace

- Funkční význam sestřihu v nepřekládaných oblastech - důležitá regulační součást genů
- V případě CK11 pokus prokázat tento způsob regulace genové exprese pomocí transgenních linií nesoucích *uidA* pod kontrolou dvou verzí promotoru, zatím nepotvrzeno



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genové modelování

- programy pro genové modelování
 - zohledňují také další parametry, např. návaznost ORF
 - **Genescan** (<http://genes.mit.edu/GENSCAN.html>)
velice dobrý pro predikci exonů v kódujících oblastech
(testováno na genu *PDR9*, identifikoval všech 23 (!) exonů)
 - **GeneMark.hmm** (<http://opal.biology.gatech.edu/GeneMark/>)
 - **GlimmerHMM** (<http://http://ccb.jhu.edu/software/glimmerhmm/>)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Identifikace genů *ab initio*

Result of last submission:

[View PDF Graphical Output](#)

[GeneMark.hmm Listing](#)

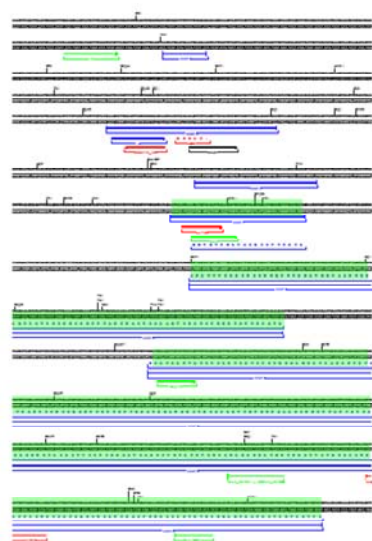
Go to: [GeneMark.hmm Protein Translations](#)

Go to: [Job Submission](#)

Dukariotyc GeneMark.hmm version bp 3.9 April 25, 2008
 Sequence name: CK11
 Sequence length: 5049 bp
 GC content: 38.724
 Matrices file: /home/geneMark/euk_gmm/matrices/ambalima_hmm0.0mod
 Thu Oct 1 11:09:24 2009

Predicted genes/exons

Gene #	Exon #	Strand	Exon Type	Exon Range	Exon Length	Start/End Frame
1	1	+	Initial	969 1025	57	1 0 -
1	2	+	Internal	1155 1394	240	1 0 -
1	3	+	Internal	1516 2175	660	1 0 -
1	4	+	Internal	2266 2644	379	1 1 -
1	5	+	Internal	2794 3217	424	2 0 -
1	6	+	Internal	3397 4659	1262	1 0 -
1	7	+	Terminal	4709 4921	213	1 0 -



/ZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
 Evropským sociálním fondem
 a státním rozpočtem České republiky

Identifikace genů *ab initio*

Result of last submission:

[View PDF Graphical Output](#)

GeneMark.hmm Listing

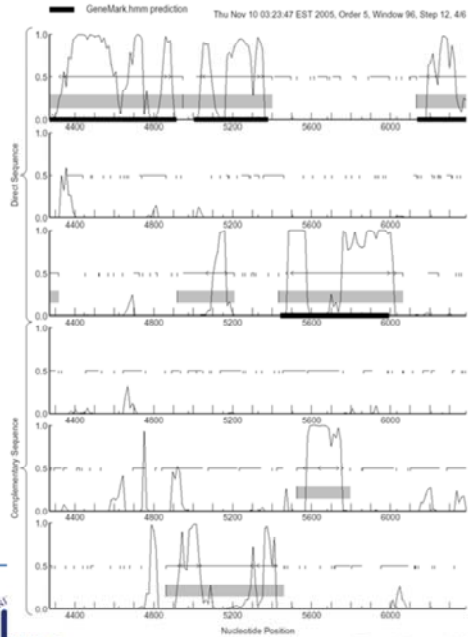
Go to: [GeneMark.hmm Protein Translations](#)

Go to: [Job Submission](#)

DNA: eukaryotic GeneMark.hmm version bp 0.9 April 25, 2008
 Sequence name: CK11
 Sequence length: 5049 bp
 GC content: 28.72%
 Matrices file: /home/genemark/euk_gmm/matrices/ambalarna_hmm0.0mod
 Thu Oct 1 11:09:24 2009

Predicted genes/exons

Gene #	Exon #	Strand	Exon Type	Exon Range	Exon Length	Start/End Frame
1	1	+	Initial	969 1025	57	1 0 --
1	2	+	Internal	1155 1394	240	1 0 --
1	3	+	Internal	1516 2175	660	1 0 --
1	4	+	Internal	2266 2644	379	1 1 --
1	5	+	Internal	2794 3217	424	2 0 --
1	6	+	Internal	3397 4659	1262	1 0 --
1	7	+	Terminal	4709 4921	213	1 0 --



AVÁNI

ancována

Evropským sociálním fondem
a státním rozpočtem České republiky

Genové homologie

- vyhledávání genů podle homologií
 - porovnávání s EST databázemi
 - BLASTN (<http://www.ncbi.nlm.nih.gov/BLAST/>, <http://workbench.sdsc.edu/>)
 - porovnávání s proteinovými databázemi
 - BLASTX (<http://www.ncbi.nlm.nih.gov/BLAST/>, <http://workbench.sdsc.edu/>)
 - Genewise (<http://www.ebi.ac.uk/Wise2/>)

porovnávají proteinovou sekvenci s genomovou DNA (po zpětném překladu), je nutná znalost aminokyselinové sekvence
 - porovnávání s homologními genomovými sekvencemi z příbuzných druhů
 - VISTA/AVID (<http://www.lbl.gov/Tech-Transfer/techs/lbnl1690.html>)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genomová kolinearita

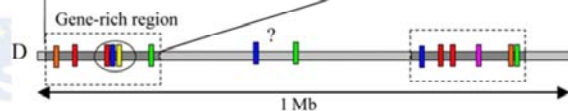
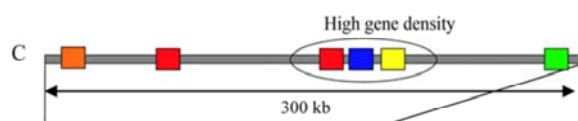
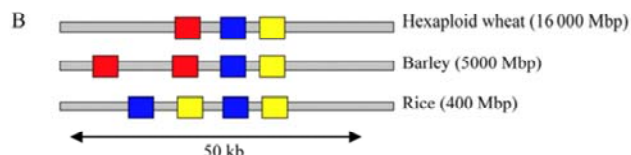
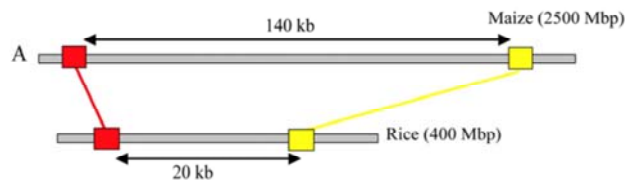
- genomy příbuzných druhů se přes značné odlišnosti vyznačují podobnostmi v uspořádání i sekvencích, možnost využití při identifikaci genů u příbuzných organismů pomocí vyhledávání v databázích
- obecné schéma postupu při využívání genomové kolinearity (také „komparativní genomika“) při experimentální identifikaci genů příbuzných organismů:
 - mapování malých genomů s využitím nízkokopiových DNA markerů (např. RFLP)
 - využití těchto markerů k identifikaci orthologních genů (genů se stejnou nebo podobnou funkcí) příbuzného organismu
 - malý genom (např. rýže, 466 Mbp) může sloužit jako vodítko, kdy jsou identifikovány molekulární nízkokopiové markery (např. RFLP) ve vazbě s genem zájmu a tyto oblasti jsou pak použity jako sonda při vyhledávání v BAC knihovnách při identifikaci orthologních oblastí velkých genomů (např. ječmene nebo pšenice, 5000, resp. 16000 Mbp)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genomová kolinearita



Feuillet and Keller, 2002

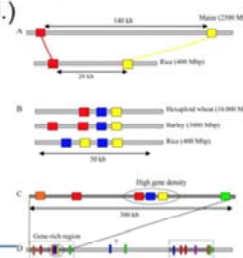
ŠTÍČKA DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky



Genomová kolinearita

- zejména využitelné u trav (např. využití příbuznosti u ječmene, pšenice, rýže a kukuřice)
- malé genomové přestavby (dalece, duplikace, inverze a translokace menší než několik cM) jsou pak detekovány podrobnou sekvenční komparativní analýzou
- během evoluce dochází u příbuzných druhů k odchylkám především v nekódujících oblastech (invaze retrotranspozonů atd.)

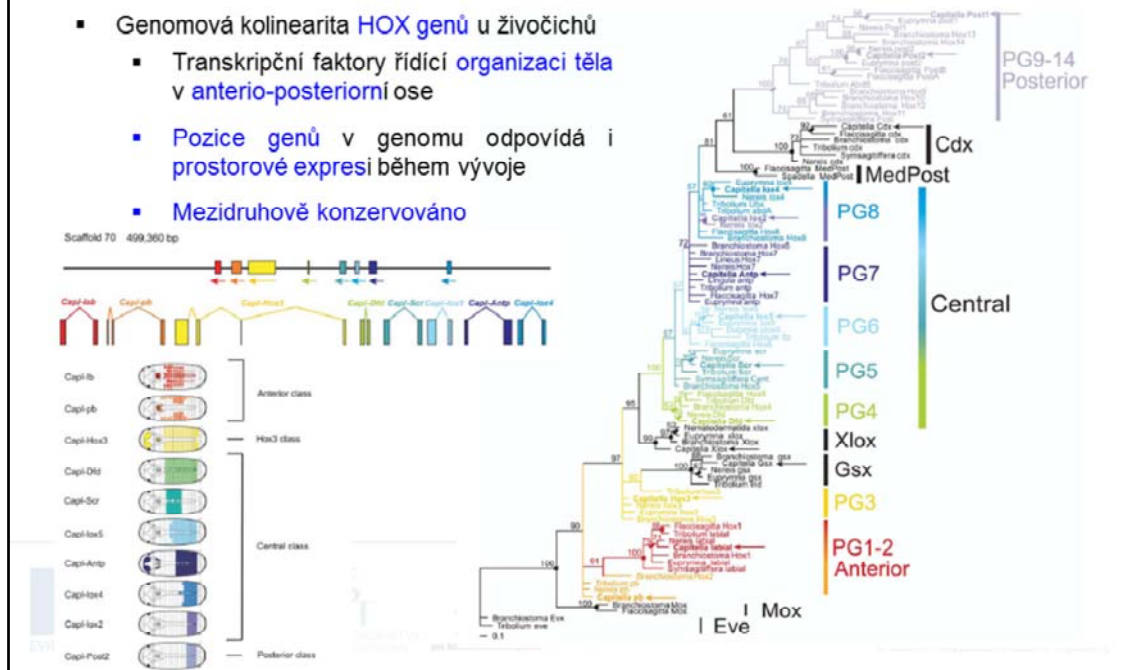


INVESTICE DO ROZVOJE VZDELAVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genomová kolinearita

- Genomová kolinearita **HOX genů** u živočichů
 - Transkripční faktory řídící **organizaci těla** v **anterio-posteriorní ose**
 - **Pozice genů** v genomu odpovídá i **prostorové expresi** během vývoje
 - **Mezidruhově konzervováno**



Genomic organization of the *Capitella* sp. I Hox cluster. A total of 11 *Capitella* sp. I Hox genes are distributed among three scaffolds. Black lines depict two scaffolds, which contain 10 of the *Capitella* sp. I Hox genes. The eleventh gene, *Cap1-Post1*, is located on a separate scaffold surrounded by ORFs of non-Hox genes (unpublished data). No predicted ORFs were identified between adjacent linked Hox genes. Transcription units are shown as boxes denoting exons, connected by lines that denote introns. Transcription orientation is denoted by arrows beneath each box. Color coding is the same as that used in on the right-hand side for each ortholog.

The phylogenetic tree on the right-hand side shows that the order of the genes on the chromosome is retained in several species (genome colinearity).

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven pomocí technologie metylačního filtrování



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Metylační filtrování

- příprava genově obohacených knihoven pomocí technologie metylačního filtrování
- **geny** jsou (většinou!) **hypometylované**, kdežto **nekódující oblasti** jsou **metylované**
- využití bakteriálního RM systému, který rozpoznává metylovanou DNA pomocí rest. enzymů McrA a McrBC
 - McrBC rozpoznává v DNA metylovaný cytozin, který předchází purin (G nebo A)
 - pro štěpení je nutná vzdálenost těchto míst z 40-2000 bp



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Metylační filtrování

- příprava genově obohacených knihoven pomocí technologie metylačního filtrování
- schéma postupu při přípravě BAC genomových knihoven pomocí metylačního filtrování:
 - příprava genomové DNA bez příměsí organelární DNA (chloroplasty a mitochondrie)
 - fragmentace DNA (1-4 kbp) a ligace adaptorů
 - příprava BAC knihovny v *mcrBC+* kmeni *E. coli*
 - selekce pozitivních klonů
- omezené využití: obohacení o kódující DNA o pouze cca 5-10 %



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Osnova

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven pomocí technologie metylačního filtrování
 - EST knihovny

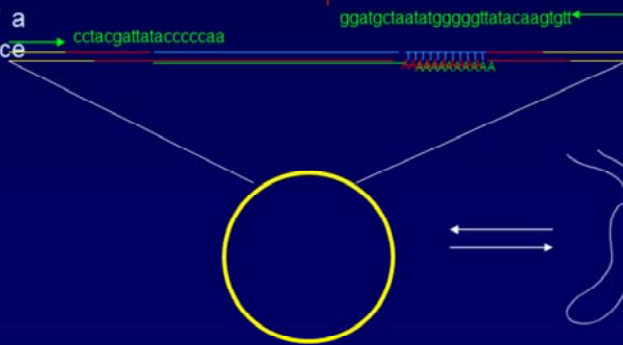


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

EST knihovny

- příprava EST knihoven
 - izolace mRNA
 - RT
 - ligace linkerů a syntéza druhého řetězce cDNA
 - klonování do vhodného bakteriálního vektoru
 - transformace do bakterií a izolace DNA (amplifikace DNA)
 - sekvenace s použitím primerů specifických pro použitý plasmid
 - uložení výsledků sekvenace do veřejné databáze



Shrnutí

- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven pomocí technologie metylačního filtrování
 - EST knihovny
 - přímá a reverzní genetika (přednáška 03)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Diskuse



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky