

# 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](mailto:hejatko@sci.muni.cz), [www.ceitec.muni.cz](http://www.ceitec.muni.cz)



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

# Literatura

## ▪ Zdrojová literatura ke kapitole 2

- Plant Functional Genomics, ed. Erich Grotewold, 2003, Humana Press, Totowa, New Jersey
- Majoros, W.H., Perlea, M., Antonescu, C. and Salzberg, S.L. (2003) GlimmerM, Exonomy, and Unveil: three ab initio eukaryotic genefinders. *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 genetika



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 funkci



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Přímá vs. reverzní genetika

## Revoluce v chápání pojmu genu

Přístupy „klasické“ genetiky



3

:

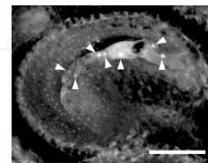
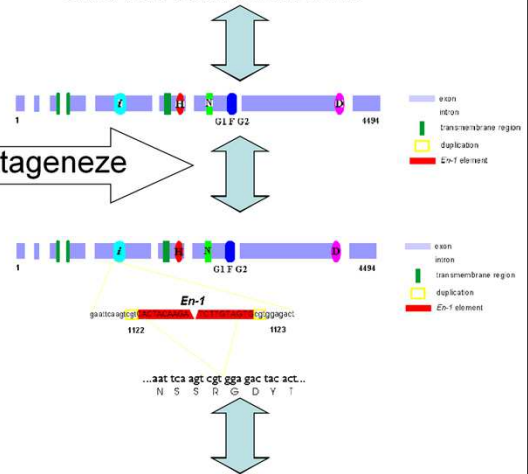
1



„Reverzně genetický“ přístup

5'TTATATATATATATATAAAAAATAAAATAAAA  
GAACAAAAAGAAAAATAAAATA...3'

inzerční mutageneze



POJE VZDĚLÁVÁNÍ  
stace je spolufinancována  
státním sociálním fondem  
z rozpočtu České republiky

# Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*

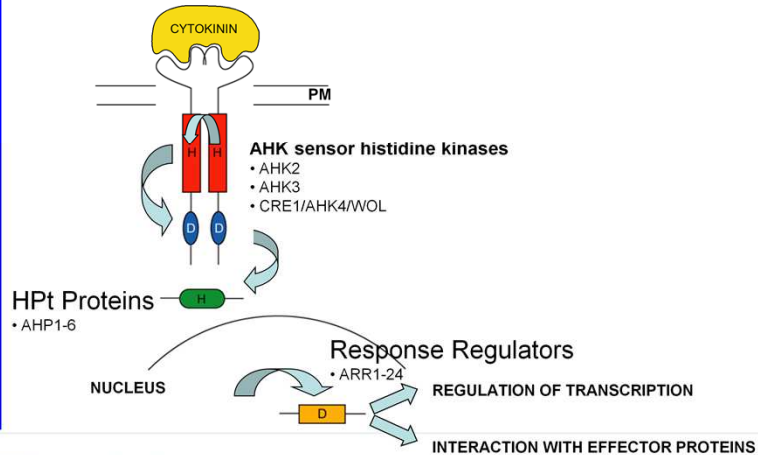


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



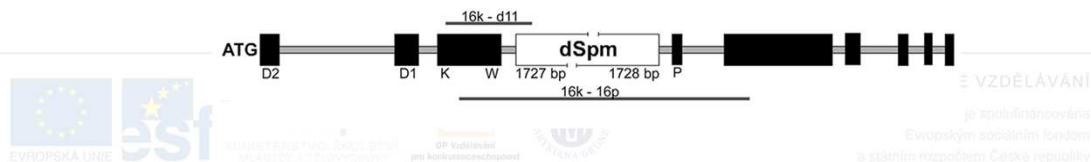
# Identifikace role genu *ARR21* – izolace inz. mutanta

- vyhledávání v databázi inzerčních mutantů (SINS)

```
Insert_SINS: 01_09_64
Query: 80      tcctageggtcatgagcggtaccatacttgacaanagagaacgtagccagccatttacagg 139
              |||
Sbjct: 58319  tcctageggtcatgagcggtaccatacttgacaagagagaacgtagccagccatttacagg 58378
Arr21: 1830
```

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

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



# 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



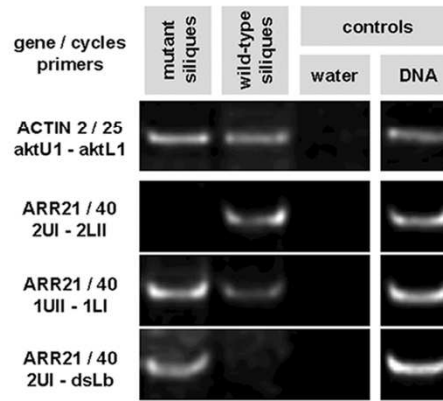
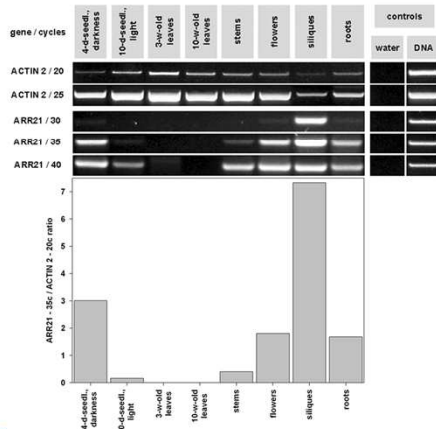
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

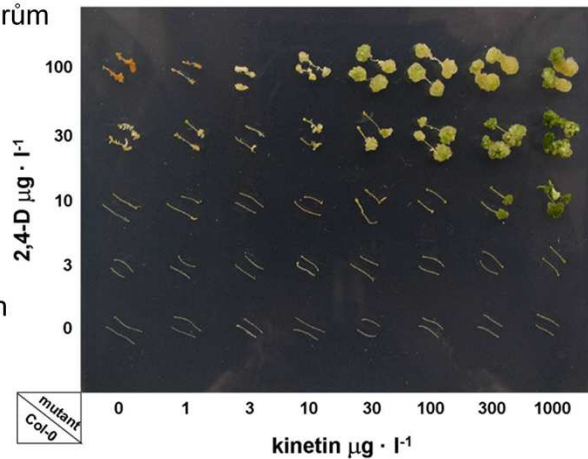
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# 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?

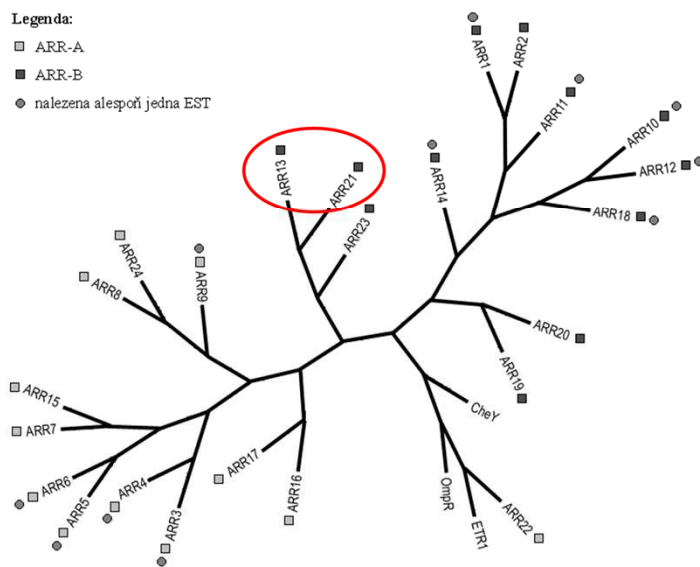


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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 AGRÁRNÍ  
VÝVOJE A RYBNÁŘSTVÍ

OP Vzdělávání  
pro konkurenceschopnost



ZVOJE VZDĚLÁVÁNÍ

záměra je spolufinancována  
evropským sociálním 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
- Inzerční mutagenese v případě identifikace funkce genu *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

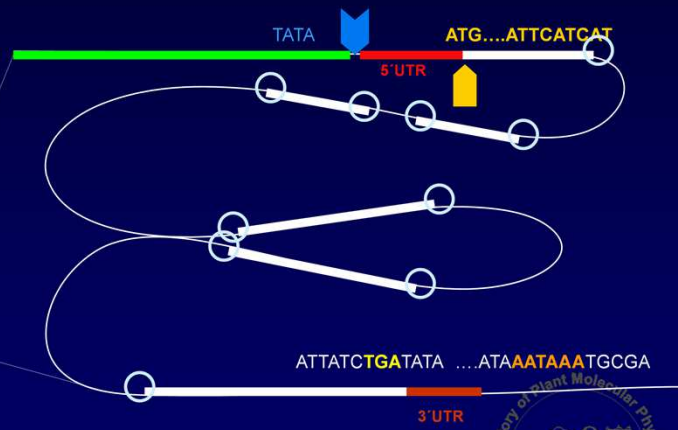


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

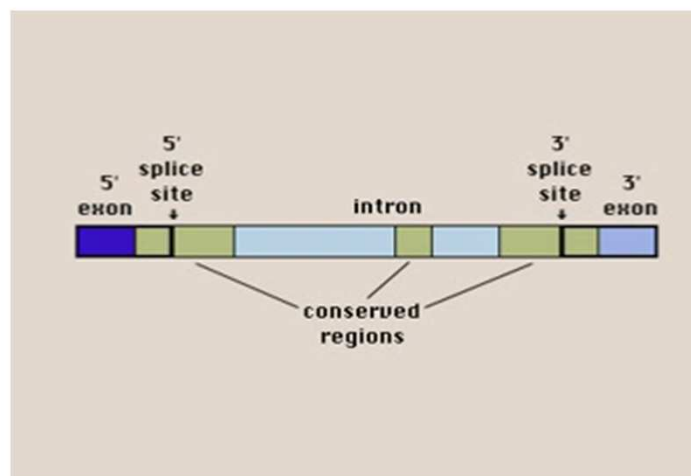
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# 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 (specifická přibližně 35%)
  - GeneSplicer ([http://www.tigr.org/tdb/GeneSplicer/gene\\_spl.html](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  
AAAGATTTTCATTCAATATAATACTTGGATAAATACTCTTATTATTTTTCTTTAGTTTATTAACAAAAACCTCAATAAAT  
ACGAGTTTAAAGTCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAAACGATAAAGTTACAAAAGTAATAATCC  
AAGTATCTCATAGTCAACATATATATAGTAATAATTAGTTGACGTATAAGAAAATAAAAAATAAATAAATTAGTATCTTAT  
TTTGGGTGGTGCTGACTGGTGACTGGTGACTGCAGAATGCTCGGCAAAATGGAACCATATCCCAAGACATGGGTTTTAGAT
```

... 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?

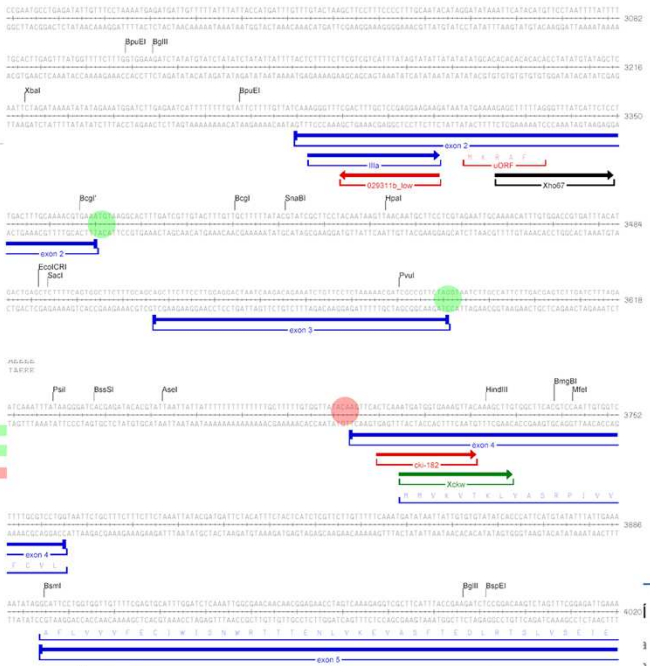
SplicePredictor. Version of February 13, 2005.  
Date run: Wed Nov 9 11:30:14 2005

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

Sequence: 1: your=sequence, from 1 to 8490.

### Potential splice sites

t	q	loc	sequence	P	c	rho	gamma	* P*Q*
A	<<<	75	tttttccgattctcagat	0.973	7.14	0.000	0.000	7 (5 1 1)
A	<<<	134	attattctctcattAGct	0.999	14.86	0.000	0.000	7 (5 1 1)
A	<<<	500	gattttgttgcttAGtc	0.977	7.48	0.000	0.000	7 (5 1 1)
A	<<<	780	tcgtctactctgattAGct	0.966	6.56	0.000	0.000	7 (5 1 1)
A	<<<	848	tatttttgaattAGat	0.968	6.80	0.000	0.000	7 (5 1 1)
A	<<<	1051	caatttttttattAGaa	0.930	5.19	0.000	0.000	7 (5 1 1)
A	<<<	1211	ttattttttttttAGct	0.996	12.14	0.000	0.000	7 (5 1 1)
A	<<<	1373	tttctctctctcaccAGaa	0.999	13.17	0.000	0.000	7 (5 1 1)
A	<<<	1481	ttctacaccttattAGct	0.983	8.04	0.000	0.000	7 (5 1 1)
A	<<<	1581	atgtttgttgcttAGaa	0.982	8.03	0.000	0.000	7 (5 1 1)
A	<<<	1781	gatttggggaaatAGgg	0.886	4.10	0.000	0.000	7 (5 1 1)
A	<<<	2440	tactcaaaaattAGat	0.959	5.46	0.000	0.000	7 (5 1 1)
A	<<<	2479	caatcaaaaattAGat	0.942	5.59	0.000	0.000	7 (5 1 1)
D	>>>>	2546	aagTtagta	0.909	4.61	0.885	1.903	15 (5 5 5)
A	<<<	2572	ttttttttttttttAGca	0.930	5.16	0.000	0.000	7 (5 1 1)
A	<<<	2763	ctcaaatccaaaAGgt	0.873	3.86	0.185	0.000	11 (5 5 1)
A	<<<	2782	tttcttttccattAGgg	0.952	5.98	0.220	0.000	11 (5 5 1)
A	<<<	3022	ttgttttgcactAGct	0.836	6.16	0.221	0.000	11 (5 5 1)
A	<<<	3048	ctttgcaatacaAGaa	0.973	7.13	0.229	0.000	11 (5 5 1)
A	<<<	3171	agttgcaatttAGct	0.988	8.74	0.000	0.000	7 (5 1 1)
A	<<<	3285	cttttggctcaccAGgg	0.993	10.03	0.000	0.006	6 (5 1 2)
D	>>>>	3372	akGTaaag	0.913	5.25	0.855	1.849	15 (5 5 5)
A	<<<	3451	gatattttctcttAGaa	0.916	4.73	0.293	0.000	12 (5 5 2)
D	>>>>	3581	cgatggcttctAGgt	0.850	3.47	0.000	0.000	7 (5 1 1)
D	>>>>	3649	caacttatta	0.933	5.25	0.000	1.848	11 (5 1 5)
A	<<<	3688	ttttgtttattcaccAGct	0.907	6.26	0.000	0.000	7 (5 1 1)
A	<<<	4254	attattctctctcAGat	0.998	12.82	0.000	0.002	8 (5 1 2)
A	<<<	4331	ttctctacattgcaAGaa	0.991	9.62	0.000	0.000	7 (5 1 1)
A	<<<	4633	gtctttgtttcttAGgg	0.879	6.97	0.000	0.000	7 (5 1 1)
A	<<<	4976	ctttgtttctctcAGct	0.952	5.98	0.000	0.000	7 (5 1 1)
A	<<<	5005	ttttttttttttttAGgg	0.990	11.17	0.000	0.000	7 (5 1 1)
D	>>>>	5356	caagTgaa	0.821	3.04	0.387	0.000	11 (5 5 1)
D	>>>>	5384	ttgtTaaag	0.941	5.54	0.478	0.090	13 (5 5 3)
A	<<<	5403	attttttttttttAGct	0.984	4.26	0.000	0.000	7 (5 1 1)
A	<<<	5441	ctttctctctcaccAGaa	0.995	10.43	0.387	0.000	11 (5 5 1)
A	<<<	5472	ttgttaaaattcaccAGct	0.965	6.62	0.478	0.090	13 (5 5 3)
D	>>>>	5745	agcGTaaag	0.991	9.48	0.990	1.856	15 (5 5 5)
A	<<<	5808	caatcaatccaaaAGgt	0.948	5.83	0.458	0.000	11 (5 5 1)
A	<<<	6133	ggttctcactcaccAGgt	0.999	13.59	0.308	0.058	12 (5 5 2)
A	<<<	6552	gpattttccactcAGgg	0.938	5.42	0.000	0.000	7 (5 1 1)



EVROPSKÁ UNIE



EVROPSKÝ SOCIÁLNÍ FOND  
MLÁDEŽE A TĚLOVÝCHOVY

pro konkurenceschopnost



EVROPSKÝ SOCIÁLNÍ FOND  
MLÁDEŽE A TĚLOVÝCHOVY

a státním rozpočtem České republiky



# Identifikace genů *ab initio*

- programy pro predikci míst sestřihu (specificita přibližně 35%)
  - GeneSplicer ([http://www.tigr.org/tdb/GeneSplicer/gene\\_spl.html](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



CBS >> Prediction Servers >> NetGene2

## NetGene2 Server

The NetGene2 server is a service producing neural network predictions of splice sites in human, *C. elegans* and *A. thaliana*

[Instructions](#)   [Output format](#)   [Abstract](#)   [Performanc](#)

### SUBMISSION

Submission of a local file with a single sequence:

File in **FASTA** format

- Human
- C. elegans*
- A. thaliana*

Submission by pasting a single sequence:

#### Sequence name

- Human
- C. elegans*
- A. thaliana*

#### Sequence

```
GAGGAGCCACAAAATGACGAATATACAAAATGATCTTAAACAGCTAAACTATATTGGACATTTTTTCGATC  
TCAGATATA  
AAAGATTTTCATTCATATAATACTTGGATAAATACTCTTATTATTTTTCTTTAGTTATTATFAAAAAAACCT  
CTAATAAAT  
ACGAGTTTAAAGTCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAACGATAAAGTTTACAAAA
```

**NOTE:** The submitted sequences are kept confidential and will be erased immediately after processing.



### 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	TCCCAACAC	GTGAAATTT		
1906	0	+	0.99	CDGTGAACG	GTGAGACAT		
3582	1	+	1.00	GGCCGCTCAG	GTAAATCTGC	H	
3762	2	+	1.00	TTCGCTCCTG	GTAAATCTGC	H	
4134	0	+	0.74	TCAACACAG	GTGTTAAAA		
4619	1	+	0.74	AGCAAGAAAG	GTCTTGTTC		
4915	0	+	0.94	CGTTCTCTG	GTAAATCTG		
5356	0	+	0.87	TCTCAACCA	GTGAAATTT		
5384	1	+	1.00	GATTGCTTG	GTAGACTCT	H	
5809	1	+	1.00	TATCCTAAG	GTGTGCCAA		
6053	0	+	1.00	GCAGCTTTC	GTAACTACT	H	
6096	1	+	0.74	CTCTTCACA	GTAAATCTG		
7369	0	+	1.00	GGACTGCCA	GTAAATCTG	H	
7886	0	+	0.74	GAACAAATG	GTAAATCTG		
9323	0	+	0.74	GGAGATTAG	GTGTTCTCT		

Donor splice sites, complement strand

pos 3'>5'	pos 5'>3'	phase	strand	confidence	5'	intron	exon	3'
	1704	0	+	0.87	ATATTTTAA	TTATGGAGAC		
	1906	0	+	0.99	ACATGTCAC	GTGAGACATCG		
	3582	1	+	0.71	CTCTTCACG	GTAACTACT		
	4134	0	+	0.81	ATATGATAG	GTGGACATTA		
	4619	0	+	0.87	GTATCAAAG	GGTTCGACT		
	4915	0	+	1.00	TGTCTTCA	ATCCGACAT	H	
	5356	2	+	0.54	AAATTCAG	TTCGAGTGC		
	5004	0	+	0.94	TTTTGCCAG	AGATACACAC		
	5472	1	+	0.96	AAAATTACAG	CTCTGTCAA		
	6135	0	+	1.00	ATTATTATAG	GTAAAGTTAA	H	
	6490	1	+	0.90	AAATTCAG	TGTGGAGA		
	6744	0	+	0.59	TGCAACAG	TTCTGTAGAG		
	7447	0	+	0.96	TTCGCACAG	ATSCCAGAAA		
	7780	2	+	0.76	TCCATTCAG	ATACAGACA		
	7786	2	+	0.92	TCAGATCAG	AACACATCA		

Acceptor splice sites, direct strand

pos 5'>3'	phase	strand	confidence	5'	intron	exon	3'
1213	0	+	0.59	TATTTTTAG	TTATGGAGAC		
1221	2	+	0.87	AGTATGGAG	ACAAGAACTG		
1373	0	+	0.71	CTCTTCACG	GTAACTACT		
1487	1	+	0.81	ATATGATAG	GTGGACATTA		
3264	0	+	0.87	GTATCAAAG	GGTTCGACT		
4254	0	+	1.00	TGTCTTCA	ATCCGACAT	H	
4832	2	+	0.54	AAATTCAG	TTCGAGTGC		
5004	0	+	0.94	TTTTGCCAG	AGATACACAC		
5472	1	+	0.96	AAAATTACAG	CTCTGTCAA		
6135	0	+	1.00	ATTATTATAG	GTAAAGTTAA	H	
6490	1	+	0.90	AAATTCAG	TGTGGAGA		
6744	0	+	0.59	TGCAACAG	TTCTGTAGAG		
7447	0	+	0.96	TTCGCACAG	ATSCCAGAAA		
7780	2	+	0.76	TCCATTCAG	ATACAGACA		
7786	2	+	0.92	TCAGATCAG	AACACATCA		



MINISTERSTVO ŠKOLSTVÍ  
Mládeže a tělovýchovy

OP Vzdělávání  
pro konkurenceschopnost



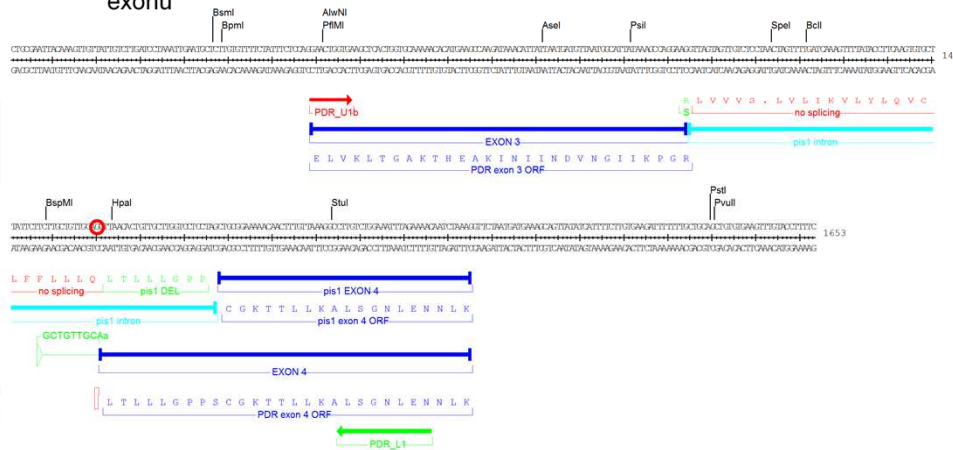
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

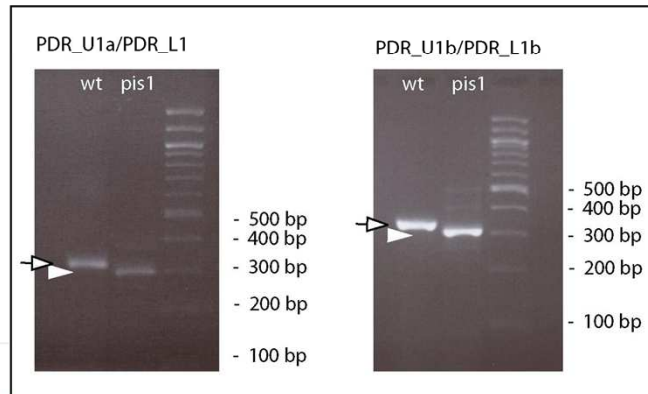
OP Vzdělávání  
pro konkurenceschopnost



a státním rozpočtem České republiky

# Sestřih RNA a adaptace

- 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



MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY  
OP Vzdělávání pro konkurenceschopnost

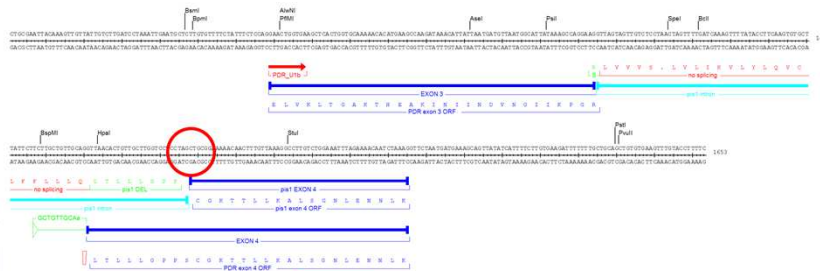


ROZVOJE VZDĚLÁVÁNÍ

o 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
  - 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



ESF

OP Vzdělávání a odborná příprava

OP Výzkum a inovace pro konkurenceschopnost



DĚLÁVÁNÍ

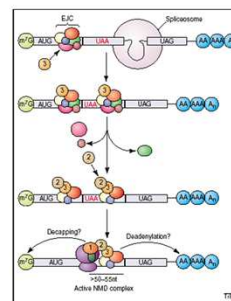
financováno

státním 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
  - 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*

**The New GENSCAN Web Server at MIT**  
**Identification of complete gene structures in genomic DNA**



[For information about Genscan, click here](#)

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see instructions at the bottom of this page) or use the GENSCAN email server. If your browser (e.g., Lynx) does not support file upload or multipart forms, use the older version.

Organism:  Suboptimal exon cutoff (optional):

Sequence name (optional):

Print options:  Predicted peptides only

Upload your DNA sequence file (one-letter code, upper or lower case, spaces/numbers ignored):

Or paste your DNA sequence here (one-letter code, upper or lower case, spaces/numbers ignored):

```
GGGGGGGCAAAATGAGGATATACAAAATGATCTTAAGAGCTAAAGTATATGGACATTTTTCGATC
TCAGATATA
AAAGATTTCTTCAATATATAGCTTGGATAAATCTCTATATATTTTCTTAGTTTATAAAAAAACCT
CTATATAAT
ACGAGTTTACGCCACAAAATCGCTTAGACTAAAATACACCATATATTTCAAACGATAAAATTTACAAAA
GGAAATGC
AAGTATCTCATAGTCAAGATATATATAGTAAATAATTAAGTGGATATAAGAAAATAAAATAAATTA
GGATCTTAT
TTTGGGTTGCTGACTGGTGAAGTGGTGAAGTGGTGGAAATGGAAATGGAAATATCCAAAGCATGG
GTTTGAAGT
AGAACAAATAGTGTCCGAGGAATGATTTAAAGTCAATAGAAATATATAATATGTAATTAGCA
ATCAAAAGC
```

To have the results mailed to you, enter your email address here (optional):

[back to the top](#)



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Identifikace genů *ab initio*

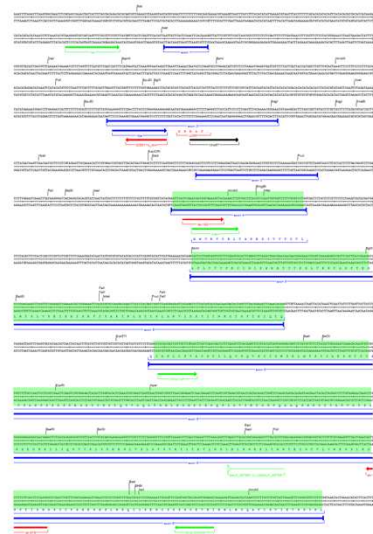
GENSCANW output for sequence CK11

GENSCAN 1.0 Date run: 10-Nov-105 Time: 02:24:26  
 Sequence CK11 : 9490 bp : 36.53% C+G : Isochore 1 ( 0 - 43 C+G%)  
 Parameter matrix: Arabidopsis1.smat  
 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	Intr	+	3708	3784	77	2	0	63	51	37	0.499	-4.08
1.02	Intr	+	3894	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	+	5005	5383	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	0.877	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

Exnum	Type	S	.Begin	...End	.Len	Fr	Ph	B/Ac	Do/T	CodRg	P....	Tscr..
S.001	Init	+	1867	1905	39	0	0	64	40	57	0.298	3.74
S.002	Init	+	2374	2442	69	0	0	55	95	-11	0.132	2.40
S.003	Intr	+	3894	4110	217	2	1	-3	-34	307	0.177	11.55
S.004	Intr	+	4352	4914	563	0	2	75	59	338	0.187	26.20
S.005	Intr	+	5005	5379	375	0	0	70	8	335	0.212	22.99
S.006	Intr	+	5442	6056	615	2	0	95	99	589	0.208	57.32



OP Vzdělávání pro konkurenceschopnost

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

**Explanation Gn.Ex** : gene number, exon number (for reference) **Type** : Init = Initial exon (ATG to 5' splice site) Intr = Internal exon (3' splice site to 5' splice site) Term = Terminal exon (3' splice site to stop codon) Sngl = Single-exon gene (ATG to stop) Prom = Promoter (TATA box / initiation site) PlyA = poly-A signal (consensus: AATAAA) **S** : DNA ...Len strand (+ = input strand; - = opposite strand) **Begin** : beginning of exon or signal (numbered on input strand) **End** : end point of exon or signal (numbered on input strand) **Len** : length of exon or signal (bp) **Fr** : reading frame (a forward strand codon ending at x has frame  $x \bmod 3$ ). For example, if nucleotides 1,2,3 of the sequence are read as a codon, that's called reading frame 0. If 2,3,4 are read as a codon, that's reading frame 1. If 3,4,5 are read as a codon, that's reading frame 2, and so on. This information, together with the starting and ending positions of the exon, is sufficient to give the amino acid sequence encoded by the exon. Another use of the reading frame is that if you see two adjacent predicted exons separated by a relatively short intron which share the same reading frame, it may be worth looking at the possibility that the intervening intron is not correct, i.e. that the two exons plus the intervening intron might form one long exon (assuming there are no inframe stops in the intron, of course). **Ph** : net phase of exon (exon length modulo 3). For example, an exon of length 15 bp has net phase 0 since 15 is divisible by 3, an exon of length 16 bp has net phase 1 because 16 divided by 3 leaves a remainder of 1, an exon of length 17 bp has net phase 2, and an exon of length 18 bp has net phase 0 again. The point of this is that exons whose net phase is 0 can be omitted from

the gene without disrupting the reading frame: such exons are candidates for being either 1) incorrect, or 2) alternatively spliced. **I/Ac** : initiation signal or 3' splice site score (tenth bit units; x 10). If below zero, probably not a real acceptor site. **Do/T** : 5' splice site or termination signal score (tenth bit units; x 10) If below zero, probably not a real donor site. **CodRg** : coding region score (tenth bit units) **P** : probability of exon (sum over all parses containing exon). This quantity is close to the actual probability that the predicted exon is correct. **Tscr** : exon score (depends on length, I/Ac, Do/T and CodRg scores).

**Comments** The SCORE of a predicted feature (e.g., exon or splice site) is a log-odds measure of the quality of the feature based on local sequence properties. For example, a predicted 5' splice site with score > 100 is strong; 50-100 is moderate; 0-50 is weak; and below 0 is poor (more than likely not a real donor site). The PROBABILITY of a predicted exon is the estimated probability under GENSCAN's model of genomic sequence structure that the exon is correct. This probability depends in general on global as well as local sequence properties, e.g., it depends on how well the exon fits with neighboring exons. It has been shown that predicted exons with higher probabilities are more likely to be correct than those with lower probabilities.

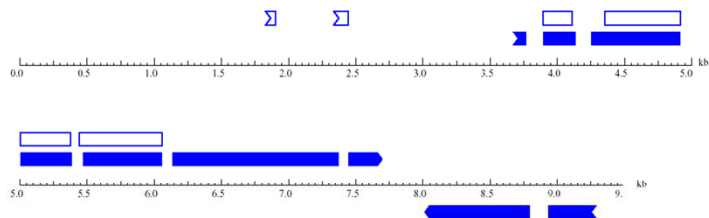
What are the suboptimal exons?

Under the probabilistic model of gene structural and compositional properties used by GENSCAN, each possible "parse" (gene structure description) which is compatible with the sequence is assigned a probability. The default output of the program is simply the "optimal" (highest probability) parse of the sequence. The exons in this optimal parse are referred to as "optimal exons" and the translation products of the corresponding "optimal genes" are printed as GENSCAN predicted peptides. (All the data in our J Mol Biol paper and on the other GENSCAN web pages refer exclusively to the optimal parse/optimal exons.) Of course, the optimal parse does not always correspond to the actual (biological) parse of the sequence, that is, the actual set of exons/genes present. In addition, there may be more than one parse which can be considered "correct", for example, in the case of a gene which is alternatively transcribed, translated or spliced. For both of these reasons, it may be of interest to consider "suboptimal" ("near-optimal") exons as well, i.e. exons which have reasonably high probability but are not present in the optimal parse. Specifically, for every potential exon E in the sequence, the probability P(E) is defined as the sum of the probabilities under the model of all possible "parses" (gene structures) which contain the exact exon E in the correct reading frame. (This quantity is calculated as described on the [GENSCAN exon probability page](#).) Given a probability cutoff C, suboptimal exons are those potential exons with P(E) > C which are not present in the optimal parse.

Suboptimal exons have a variety of potential uses. First, suboptimal exons sometimes correspond to real exons which were missed for whatever reason by the optimal parse of the sequence. Second, regions of a prediction which contain multiple overlapping and/or incompatible optimal and suboptimal exons may in some cases indicate alternatively spliced regions of a gene (Burge & Karlin, in preparation). The probability cutoff  $C$  used to determine which potential exons qualify as suboptimal exons can be set to any of a range of values between 0.01 and 1.00. The default value on the web page is 1.00, meaning that no suboptimal exons are printed. For most applications, a cutoff value of about 0.10 is recommended. Setting the value much lower than 0.10 will often lead to an explosion in the number of suboptimal exons, most of which will probably not be useful. On the other hand, if the value is set much higher than 0.10, then potentially interesting suboptimal exons may be missed.

# Identifikace genů *ab initio*

GENSCAN predicted genes in sequence 02:56:23



Key:



MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY  
OP Vzdělávání pro konkurenceschopnost



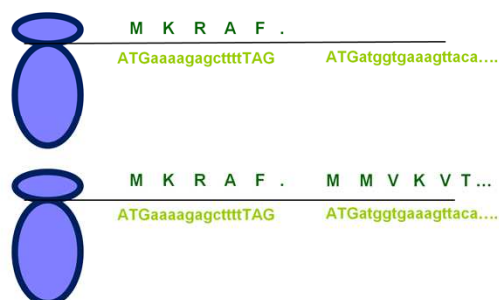
OZVOJE VZDĚLÁVÁNÍ

Inovace 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

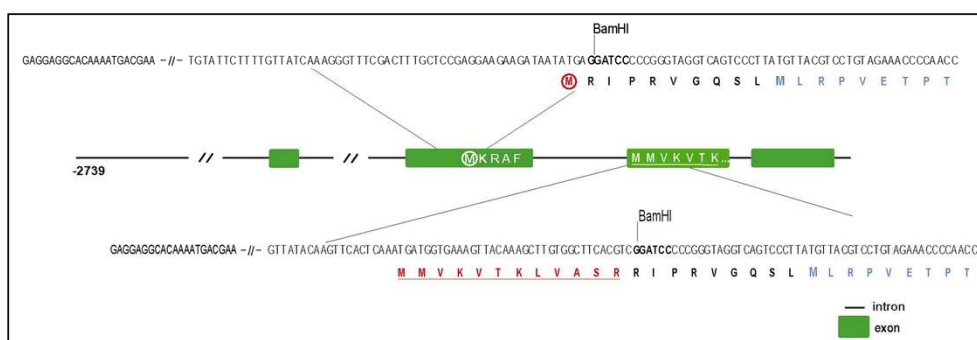


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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 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

# 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

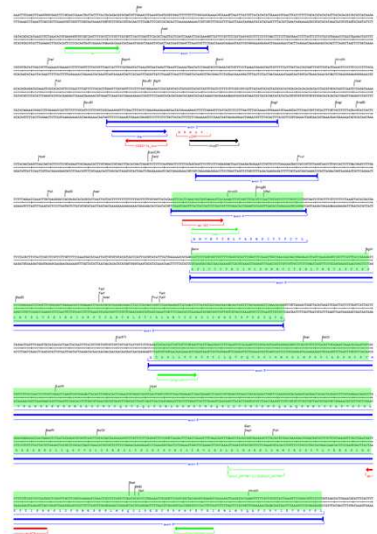
Go to: [GeneMark.hmm Protein Translations](#)

Go to: [Job Submission](#)

Eukaryotic GeneMark.hmm version by 3.9 April 25, 2008  
 Sequence name: CK11  
 Sequence length: 5040 bp  
 GC content: 38.794  
 Matrices file: /home/genemark/euk\_gmm/matrices/abhalima\_hmm3.0mod  
 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 3 - -		
1	2	+	Internal	1155 1394 240		1 3 - -
1	3	+	Internal	1516 2175 660		1 3 - -
1	4	+	Internal	2266 2644 379		1 1 - -
1	5	+	Internal	2794 3317 524		2 3 - -
1	6	+	Internal	3297 4629 1333		1 3 - -
1	7	+	Terminal	4709 4921 213		1 3 - -



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](#)

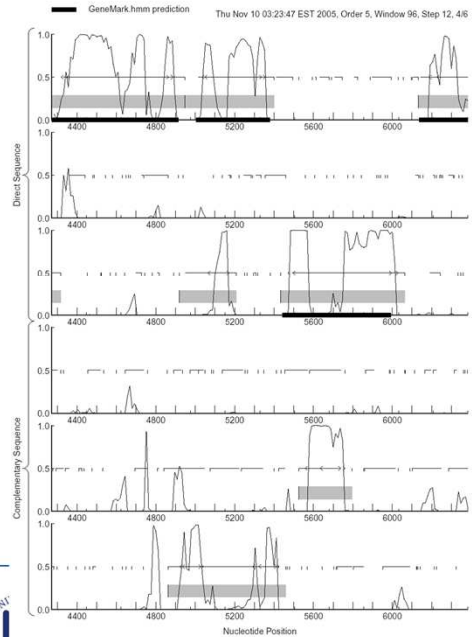
[Go to: GeneMark.hmm Protein Translations](#)

[Go to: Job Submission](#)

Eukaryotic GeneMark.hmm version by 3.9 April 25, 2008  
 Sequence name: CK11  
 Sequence length: 5040 bp  
 GC content: 38.794  
 Matrices file: /home/genemark/uk\_gmm/matrices/abhalima\_hmm3.0mod  
 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 3 --
1	2	+	Internal	1155 1394	240	1 3 --
1	3	+	Internal	1516 2175	660	1 3 --
1	4	+	Internal	2266 2644	379	1 1 --
1	5	+	Internal	2794 3317	524	2 3 --
1	6	+	Internal	3297 4629	1333	1 3 --
1	7	+	Terminal	4709 4921	213	1 3 --



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/lbn1690.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

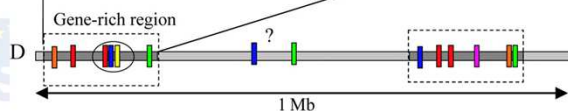
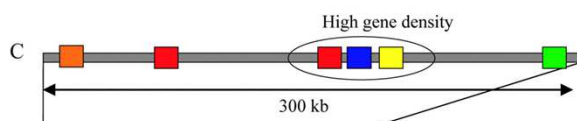
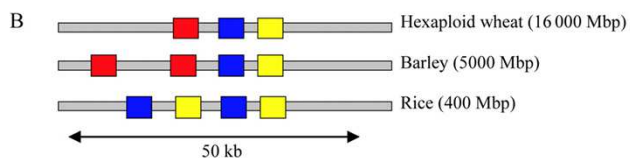
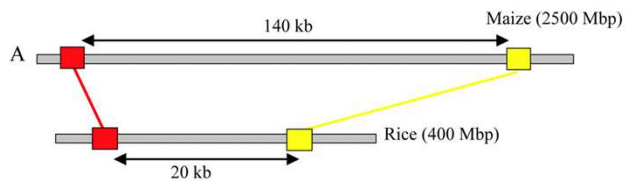
# Genomová kolinearita

- 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

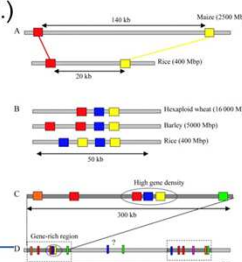
ŠTÍČKA DO ROZVOJE VZDĚLÁVÁNÍ

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.)



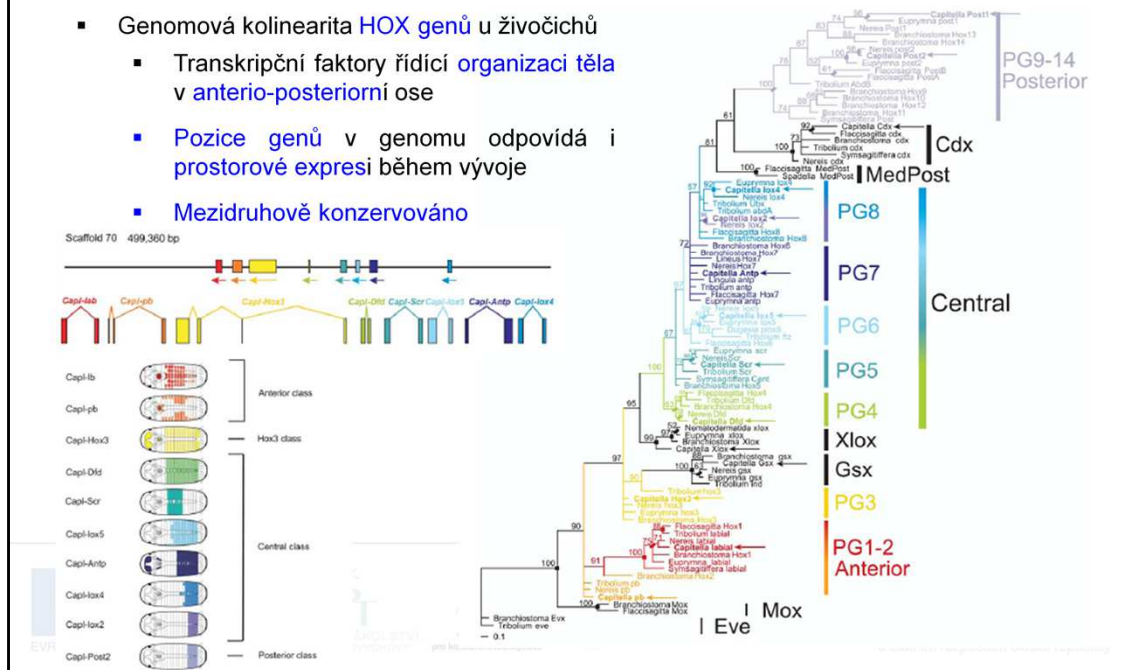
INVESTICE DO ROZVOJE VZDELAVÁNÍ

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 antero-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, *CapI-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

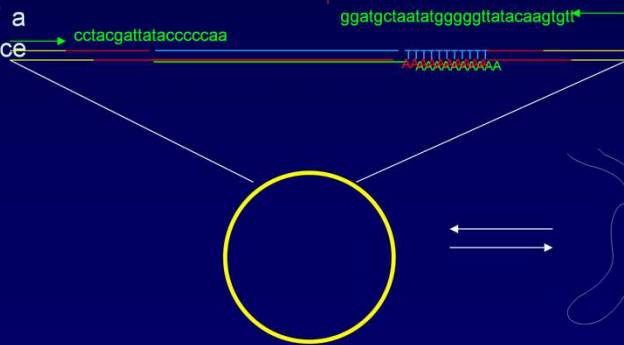


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# EST knihovny

- příprava EST knihoven
  - izolace mRNA
  - RT PCR
  - ligace linkerů a syntéza druhého řetězce cDNA
  - klonování do vhodného bakteriálního vektoru
  - transformace do bakterií a izolace DNA (amplifikace DNA)
  - sekvenace s použitím specifických primerů 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