

# CG020 Genomika

## Přednáška 2

Identifikace genů

Jan Hejátko

Funkční genomika a proteomika rostlin,  
Mendelovo centrum genomiky a proteomiky rostlin,  
Středoevropský technologický institut (CEITEC), Masarykova univerzita, Brno  
[hejatk@sci.muni.cz](mailto:hejatk@sci.muni.cz), [www.ceitec.muni.cz](http://www.ceitec.muni.cz)



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERZITA  
BRNO

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 genefinders. *Nucleic Acids Research*, **31**(13).
- Singh, G. and Lykke-Andersen, J. (2003) New insights into the formation of active nonsensemediated 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



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

# 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
  - rozdíl v myšlenkových přístupech k identifikaci genů a jejich funkcí



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MUNI  
MASARYKŮVA UNIVERZITA  
BRNO

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

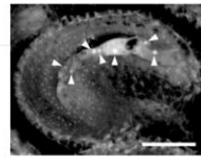
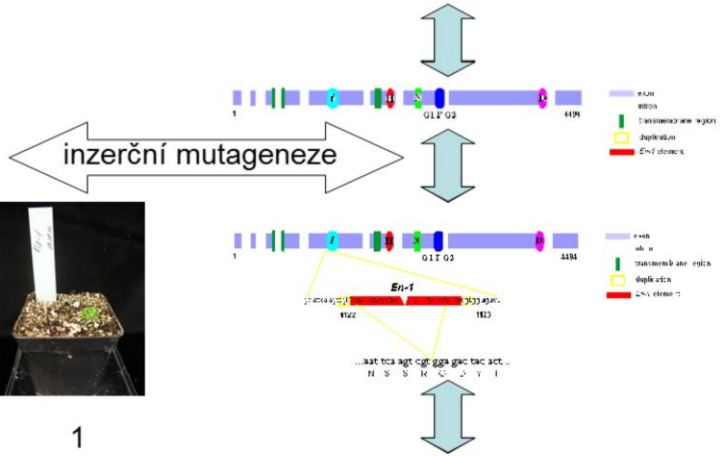
:



1

„Reverzně genetický“ přístup

5'TTATATATATATATATAAAAATAAAAAATAAAA  
GAACAAAAAGAAAAATAAAATA....3'



PROJEKT VZDĚLÁVÁNÍ

... je spolufinancována  
... a dalšími partnery  
... a dalšími partnery  
... a dalšími partnery

# Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



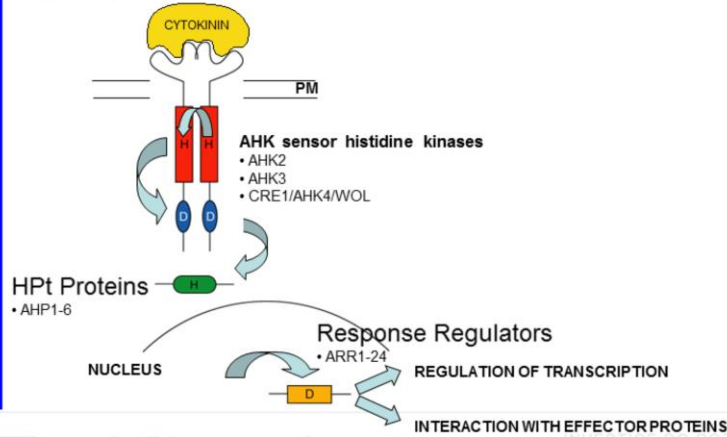
MASARYKŮVA UNIVERZITA  
BRNO

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



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERSITA BRNO

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



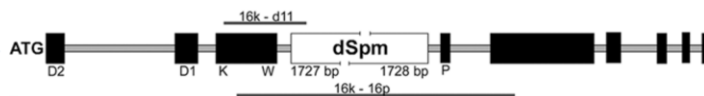
# Identifikace role genu *ARR21* – izolace inz. mutanta

- vyhledávání v databázi inzerčních mutantů (SINS)

```
Insert_SINS: 01_09_64  
Query: 80 tcttagcggtcatgagcgtaccatacttgacaanagagaaactagccagccatttacagg 139  
|||||  
Sbjct: 58319 tcttagcggtcatgagcgtaccatacttgacaagagaaactagccagccatttacagg 58378  
Arr21: 1830
```

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

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



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

Operační program  
Vzdělávání pro konkurenceschopnost



VZDĚLÁVÁNÍ

je spolufinancováno  
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



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



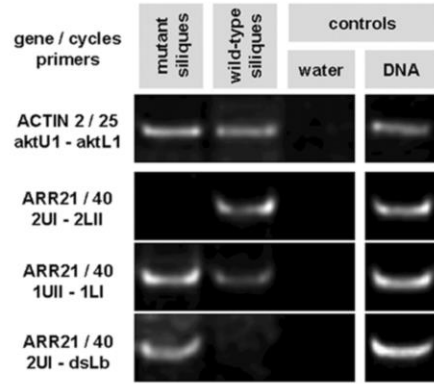
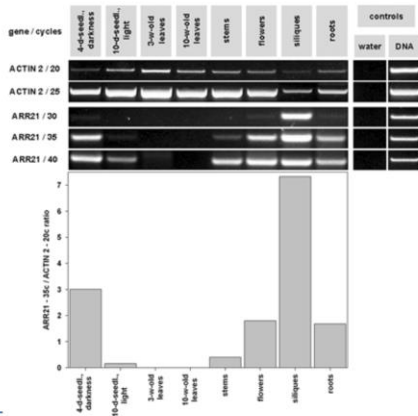
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Identifikace role genu *ARR21* – analýza exprese

Standardní typ

Inzerční mutant



MINISTERSTVO ŠKOLSTVÍ, Mládeže a Tělovýchovy



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



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

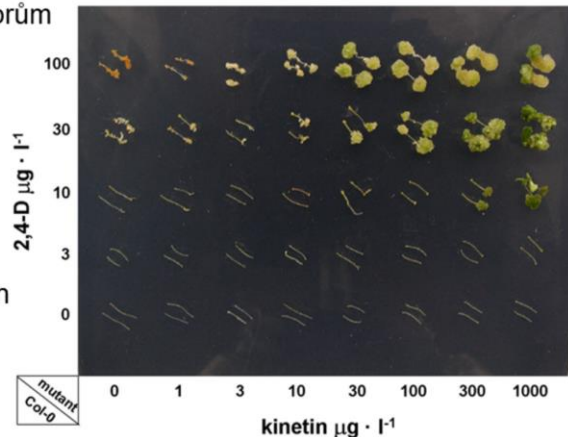
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# 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?



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



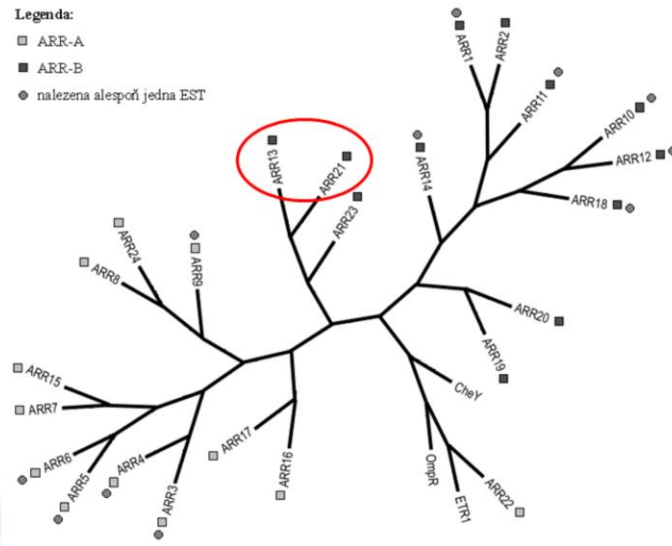
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Identifikace role genu *ARR21* – příbuznost ARR genů

Legenda:

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



MINISTERSTVO ŠKOLSTVÍ  
Mládež a tělovýchovy

OP Vzdělávání  
pro konkurenceschopnost



ZVOJE VZDĚLÁVÁNÍ

Centra 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 (?)



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERSITA  
BRNO

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



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



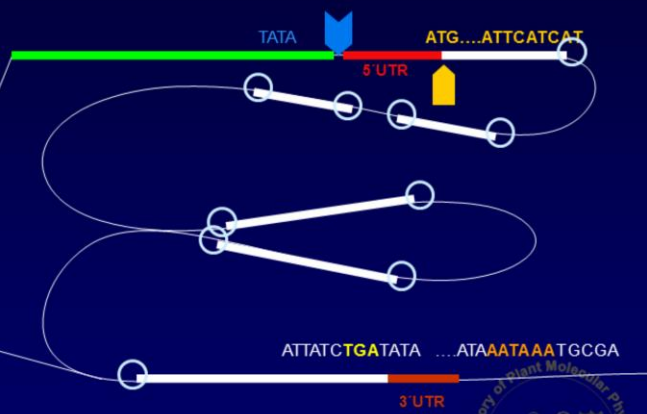
MASARYKŮVA UNIVERZITA  
BRNO

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

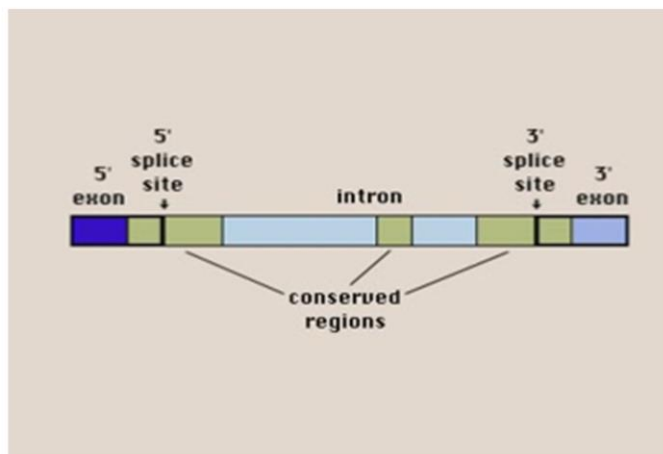
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# 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>)



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MUNI  
MASARYKŮVA UNIVERZITA  
BRNO

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  
AAAGATTTCAATCAATATAAATACTTGGATAAATACTTATATTTTCCTTAGTTTATTAACAAAAACCTCTAATAAAT  
ACGAGTTTAAAGCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAAACGATAAAGTTACAAAAGTAATATCC  
AAGTATCTCATAGTCAACATATATATAGTAATAATTAGTTGACGTATAAGAAAAATAAAAAATAAATAGTATCTTAT  
TTTGGTGGTGCTGACTGGTGACTGGTGACTGCAGAATGCTCGGCARAATGGAAACCATATCCCAAGACATGGGTTTAGAT
```

... 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?

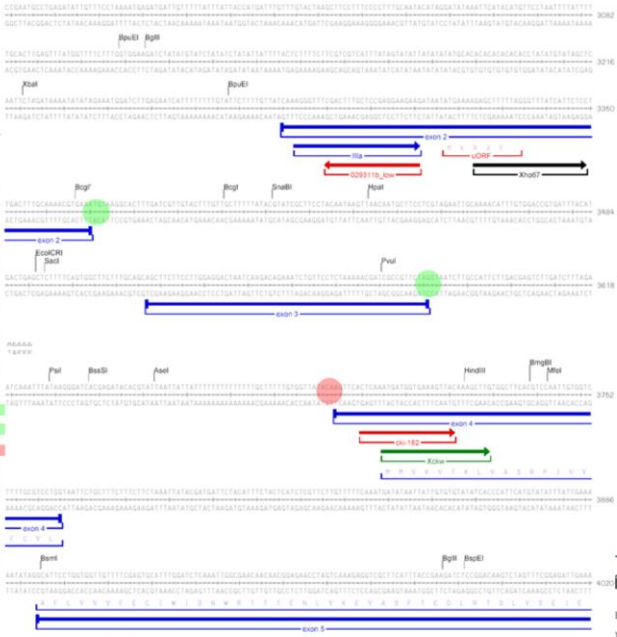
splicePredictor, Version of February 13, 2005.  
Date run Wed Nov 9 11:30:14 2005.

Species: Homo sapiens  
Model: 2-class Bayesian  
Prediction cutoff (2 in [BF]): 3.00  
Local pruning: on  
Non-ambigial sites: not scored

Sequence: | your=sequence, from 1 to 9490.

### Potential splice sites

L	g	loc	sequence	F	c	rho	gamma	* P*RG*
A	<<<	75	tttttttgcagctgagat	0.973	7.16	0.000	0.000	7 (5 1 1)
A	<<<	124	atattttttttttttagctt	0.999	14.86	0.000	0.000	7 (5 1 1)
A	<<<	500	gattttttttttttttttagctt	0.975	7.48	0.000	0.000	7 (5 1 1)
A	<<<	780	tccttttttttttttttttttagctt	0.986	8.56	0.000	0.000	7 (5 1 1)
A	<<<	848	tatttttttttttttttttttagctt	0.968	6.80	0.000	0.000	7 (5 1 1)
A	<<<	1051	caatttttttttttttttttttagctt	0.950	5.59	0.000	0.000	7 (5 1 1)
A	<<<	1213	ttatttttttttttttttttttagctt	0.999	12.14	0.000	0.000	7 (5 1 1)
A	<<<	1373	ttcttttttttttttttttttagctt	0.999	13.27	0.000	0.000	7 (5 1 1)
A	<<<	1487	ttttatatatttttttttttttagctt	0.983	4.04	0.000	0.000	7 (5 1 1)
A	<<<	1581	atttttttttttttttttttagctt	0.962	8.03	0.000	0.000	7 (5 1 1)
A	<<<	1781	gpttttttttttttttttttagctt	0.986	4.10	0.000	0.000	7 (5 1 1)
A	<<<	2440	takttttttttttttttttttagctt	0.939	5.46	0.000	0.000	7 (5 1 1)
A	<<<	2479	caatttttttttttttttttttagctt	0.942	5.59	0.000	0.000	7 (5 1 1)
D	<<<<<	2544	agpttttttttttttttttttagctt	0.909	4.61	0.985	1.903	15 (5 5 5)
A	<<<	2572	tttttttttttttttttttagctt	0.930	5.14	0.000	0.000	7 (5 1 1)
A	<<<<	2763	ctcttttttttttttttttttagctt	0.973	3.86	0.385	0.000	11 (5 5 1)
A	<<<<<	2782	tttttttttttttttttttagctt	0.952	5.99	0.220	0.000	11 (5 5 1)
A	<<<<<<	3022	tttttttttttttttttttagctt	0.956	6.16	0.221	0.000	11 (5 5 1)
A	<<<<	3048	tttttttttttttttttttagctt	0.973	7.13	0.229	0.000	11 (5 5 1)
A	<<<<	3171	gpttttttttttttttttttagctt	0.968	8.74	0.000	0.000	7 (5 1 1)
A	<<<<	3234	tttttttttttttttttttagctt	0.930	10.24	0.000	0.006	8 (5 5 1)
D	<<<<<<	3372	tttttttttttttttttttagctt	0.933	5.28	0.855	1.849	15 (5 5 1)
A	<<<<<<<	3511	tttttttttttttttttttagctt	0.916	4.77	0.233	0.000	11 (5 5 1)
A	<<<<<<<<	3591	tttttttttttttttttttagctt	0.910	4.87	0.000	0.000	11 (5 1 1)
D	<<<<<<<<<	3649	tttttttttttttttttttagctt	0.933	5.21	0.000	1.848	11 (5 1 1)
D	<<<<<<<<<<<	3698	tttttttttttttttttttagctt	0.907	4.08	0.000	0.000	15 (5 1 1)
A	<<<<<	4234	atttttttttttttttttttagctt	0.998	12.82	0.000	0.002	8 (5 1 1)
A	<<<<	4302	tttttttttttttttttttagctt	0.991	9.42	0.000	0.000	7 (5 1 1)
A	<<<<	4633	gttttttttttttttttttagctt	0.979	3.97	0.000	0.000	7 (5 1 1)
A	<<<<<	4976	tttttttttttttttttttagctt	0.952	5.89	0.000	0.000	7 (5 1 1)
A	<<<<	5004	tttttttttttttttttttagctt	0.996	11.17	0.000	0.000	7 (5 1 1)
D	<<<<<<	5334	tttttttttttttttttttagctt	0.821	3.04	0.387	0.000	11 (5 5 1)
D	<<<<<<<	5384	tttttttttttttttttttagctt	0.941	5.28	0.478	0.990	13 (5 5 1)
A	<<<<	5403	tttttttttttttttttttagctt	0.994	4.24	0.000	0.000	7 (5 1 1)
A	<<<<<<	5441	tttttttttttttttttttagctt	0.995	10.43	0.387	0.000	11 (5 5 1)
A	<<<<<<<	5472	tttttttttttttttttttagctt	0.965	6.62	0.478	0.990	13 (5 5 1)
D	<<<<<<<<	5745	tttttttttttttttttttagctt	0.991	9.48	0.990	1.934	15 (5 5 1)
A	<<<<<<<<	5808	tttttttttttttttttttagctt	0.948	5.83	0.468	0.000	11 (5 5 1)
A	<<<<<<	6135	gttttttttttttttttttagctt	0.999	13.59	0.308	0.000	12 (5 5 1)
A	<<<<<<<<	6302	gttttttttttttttttttagctt	0.938	5.42	0.000	0.000	7 (5 1 1)



EVROPSKÁ UNIE

MLÁDEŽE A TĚLOVÝCHOVY

pro konkurenceschopnost

FRANCA BRU

a státním rozpočtem České republiky



# Identifikace genů *ab initio*

- programy pro predikci míst sestřihu (specifita 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/>)



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERSITA BRNO

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Predikce míst sestřihu



## 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

GAGGAGGCACAAAATGACGAATATACAAAATGATCTTAAACAGCTAAACTATATGGACATTTTTTCGATCTCAGATATAAAGATTTCAATCAATATAACTTTGGATAAATACTCTTATTATTTTTCTTTAGTTTTAAAAAAAACCTCTAATAAATACGAGTTTAGTCCACAAAATCGCTTAGACTAAATACACCATATAATTCAAACGATAAAGTTTACAAA

**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 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

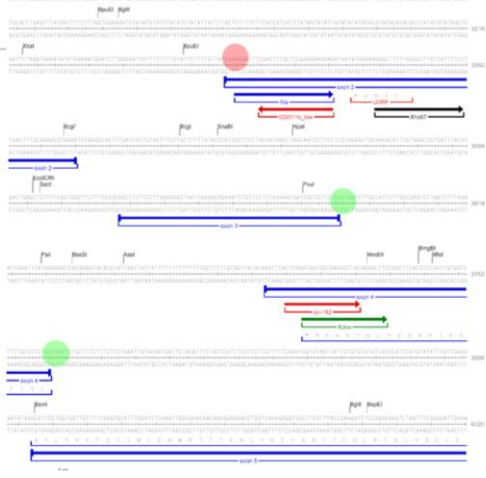
pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
1704	0	+	+	0.87	TTCGAAACAC*	TGATATTT		
1906	0	+	+	0.99	CGGTGAACGG*	GTGAGACAT		
3782	1	+	+	1.00	GGCTCTCTGG*	GTATCCGCG		
3745	1	+	+	1.00	TTTCTCTCTG*	GTATCCGCG		
4134	0	+	+	0.74	TCAAACACAG*	GTGTGTAAA		
4619	1	+	+	0.74	AGCAAGAAG*	GTCTTGTTC		
4915	0	+	+	0.94	CGTTCCTCTG*	GTAAATACG		
5356	0	+	+	0.87	TCTCAACAA*	GTGAATGTT		
5384	1	+	+	1.00	GATTTGTGTT*	GTAGACTCT		
5809	1	+	+	1.00	TATCCCTAAG*	GTGTGTGAA		
6057	0	+	+	1.00	GCAGTCTTGT*	GTAACTACT		
4096	1	+	+	0.74	CTCTTCACAA*	GTAAACTAG		
7469	0	+	+	1.00	GGACTCCCAA*	GTAACTTAA		
7886	0	+	+	0.74	GAACAAAATG*	GTAGATGAA		
9323	0	+	+	0.74	GAAGATTAGG*	GTTTTCTCT		

### Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
-----	-------	-----	-------	-------	--------	------------	----	------	--------	----

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
1213	0	+	+	0.59	TATTTTTTAA*	TTATGGAGAC		
1221	2	+	+	0.87	AGTTATGAG*	ACAGAGATCG		
1373	0	+	+	0.71	TCTTTCACAG*	GACACAGAT		
1487	1	+	+	0.81	ATATTGATAG*	TGGACATTA		
2284	0	+	+	0.87	GTATCCAAAG*	GGTTCGACT		
4254	0	+	+	1.00	TGTTTCTCTG*	ATCCGACAT		
4832	2	+	+	0.54	AAAATTGCAG*	TCCAGTGGC		
5004	0	+	+	0.94	TTTTTGCCAG*	AGATACACAC		
5472	1	+	+	0.96	AAATTACAG*	CTCTCTCAA		
6135	0	+	+	1.00	ATATTATAG*	GTAAAGTAA		
6490	1	+	+	0.90	AAAGTTACAG*	TGGTGGAGAA		
6744	0	+	+	0.59	TGTCAAACAG*	TTTCGTAGAG		
7447	0	+	+	0.96	TCTCCACAG*	ATGCCAGAA		
7780	2	+	+	0.76	TCCATTTCAG*	ATACAGAACA		
7786	2	+	+	0.92	TCGATACAG*	AACACATCCA		



MINISTERSTVO ŠKOLSTVÍ, Mládeží a tělovýchovy  
ČP Vzdělávání pro konkurenceschopnost



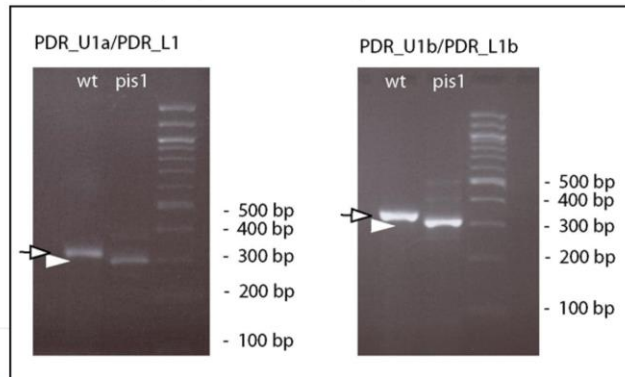
## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
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

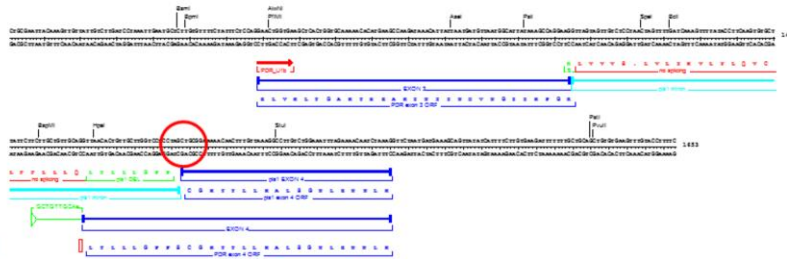
OP Vzdělávání  
pro konkurenceschopnost



ROZVOJE VZDĚLÁVÁNÍ  
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
  - 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 ŠKOLSTVÍ  
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání  
pro konkurenceschopnost



DĚLÁVÁNÍ

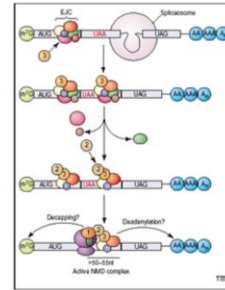
Řízené činnosti

záměrným 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/>)

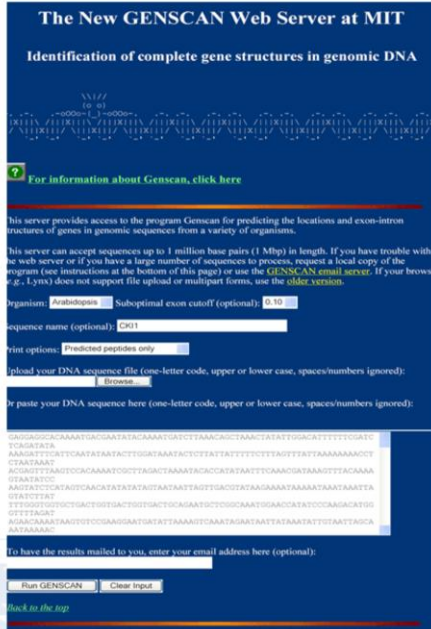


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Identifikace genů *ab initio*



The screenshot shows the Genscan web server interface. At the top, it reads "The New GENSCAN Web Server at MIT" and "Identification of complete gene structures in genomic DNA". Below this is a diagram of a gene structure with exons and introns. A green question mark icon is followed by the text "For information about Genscan, click here". The main text describes the server's capabilities: "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." It also mentions that the server can accept sequences up to 1 million base pairs (1 Mbp) in length. The interface includes several input fields: "Organism" with a dropdown menu set to "Arabidopsis", "Suboptimal exon cutoff (optional)" with a value of "0.10", and "Sequence name (optional)" with the value "001". There are also checkboxes for "Predicted peptides only" and "Reload your DNA sequence file (one-letter code, upper or lower case, spaces/numbers ignored)". A "Browse" button is next to the file upload field. Below these fields is a text area for pasting DNA sequences. At the bottom, there is an email field with the prompt "To have the results mailed to you, enter your email address here (optional):" and two buttons: "Run GENSCAN" and "Clear input". A "Back to the top" link is also present.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Identifikace genu *ab initio*

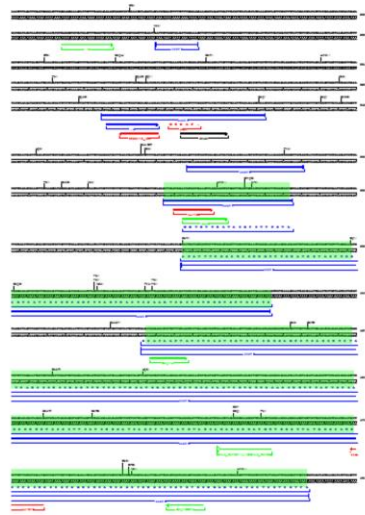
GENSCANW output for sequence CKII

```

GENSCAN 1.0   Date run: 10-Nov-105   Time: 02:24:26
Sequence CKII : 9490 bp : 36,53% C+G : Isochore 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 63 51 37 0.499 4.03
1.02 Intr + 3894 4133 240 2 0 -3 7 327 0.713 17.32
1.03 Intr + 4255 4914 660 0 0 86 59 286 0.771 22.59
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 0.977 56.86
1.07 Term + 7448 7660 213 1 0 43 35 212 0.999 12.65
1.08 PolyA + 7910 7915 6 -0.45

2.03 PolyA - 7976 7971 6 -4.83
2.02 Term - 8793 8050 257 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..
-----
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 + 5005 5379 375 0 0 70 8 335 0.212 22.59
8.006 Intr + 5442 6056 615 2 0 95 99 589 0.208 57.32
    
```



INSTITUT PRO BIOMOLEKULÁRNÍ PĚVĚLAVÁNÍ  
 Ústav pro biotecnologii  
 a šlechtění rostlinných kultivarů

**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) PolyA = poly-A signal (consensus: AATAAAA) S : DNA 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 mod 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.

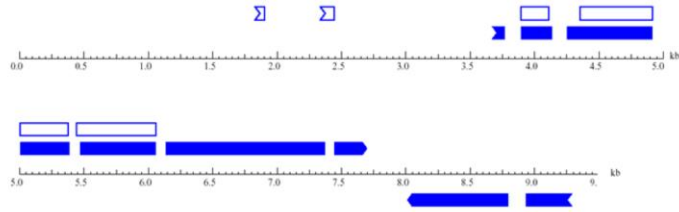
What are the suboptimal exons?

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: ■ Initial exon ■ Internal exon ■ Terminal exon ■ Single-exon gene ■ Optimal exon □ Suboptimal exon

MINISTERSTVO ŠKOLSTVÍ, Mládeží a tělovýchovy

OP Vzdělávání pro konkurenceschopnost



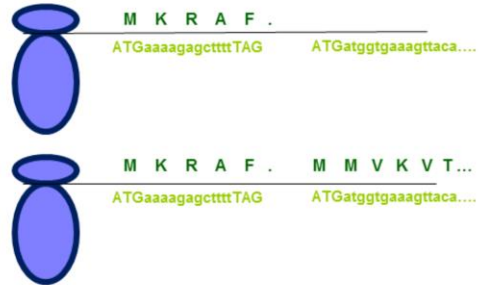
OZVOJE VZDĚLÁVÁNÍ

prezentace 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 CK11 pokus prokázat tento způsob regulace genové exprese pomocí transgenních linií nesoucích *uidA* pod kontrolou dvou verzí promotoru, zatím nepotvrzeno

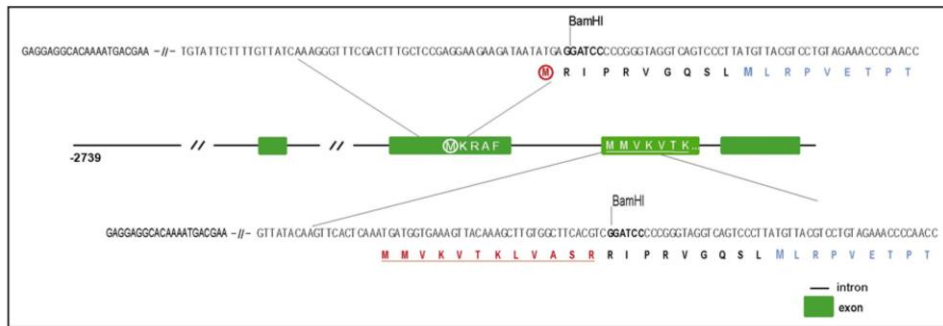


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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/>)



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERSITA  
BRNO

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](#)

GeneMarkhm Listing

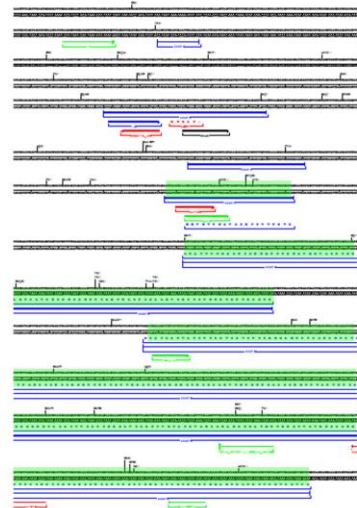
Go to: [GeneMarkhm Protein Translations](#)

Go to: [Job Submission](#)

Eukaryotic GeneMarkhm version by 3.9 April 25, 2008  
Sequence name: CK11  
Sequence length: 5042 bp  
GC content: 38.73%  
Matrix file: /home/genemark/euk\_gbm.matrices/zhhalima\_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 1294	140	1 3 - -
1	3	+	Internal	1516 2175	660	1 3 - -
1	4	+	Internal	2266 2644	379	1 1 - -
1	5	+	Internal	2794 3217	424	2 3 - -
1	6	+	Internal	3397 4629	1232	1 3 - -
1	7	+	Terminal	4709 4921	213	1 3 - -



/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](#)

GeneMarkhm Listing

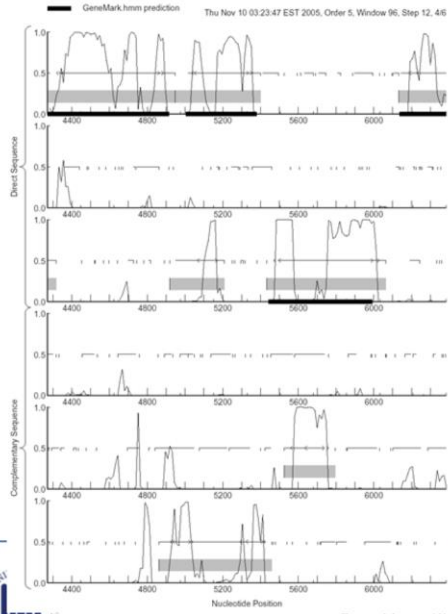
Go to: [GeneMarkhm Protein Translations](#)

Go to: [Job Submission](#)

Eukaryotic GeneMarkhm version by 3.9 April 25, 2008  
 Sequence name: CK11  
 Sequence length: 5042 bp  
 GC content: 38.73%  
 Matrices file: /home/genemark/euk\_gbm.matrices/zhhalama\_hmm2.0mod  
 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 3 --
1	3	+	Internal	1516 2175	660	1 3 --
1	4	+	Internal	2266 2644	379	1 1 --
1	5	+	Internal	2794 3217	424	2 3 --
1	6	+	Internal	3397 4659	1262	1 3 --
1	7	+	Terminal	4709 4921	213	1 3 --



AVÁNÍ  
 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 genomová homologie



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA  
UNIVERSITA  
BRNO

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomová kolinearita

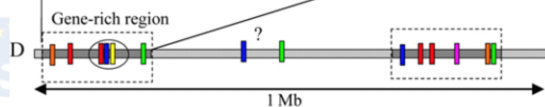
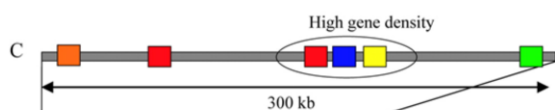
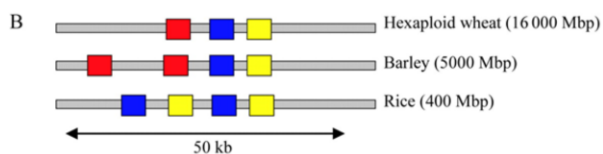
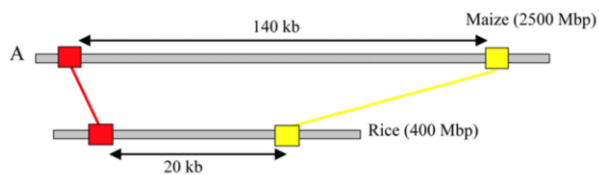
- genomy příbuzných druhů se přes značné odlišnosti vyznačují podobnostmi v uspořádání i sekvencích, možnost využití při identifikaci genů u příbuzných organismů pomocí vyhledávání v databázích
- obecné schéma postupu při využívání genomové kolinearity (také „komparativní genomika“) při experimentální 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

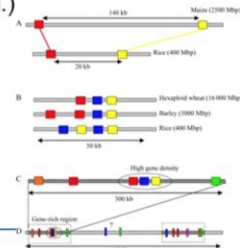
ŠTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MUNI  
MASARYKIANA BRNO

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



EVROPSKÁ UNIE

esf



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



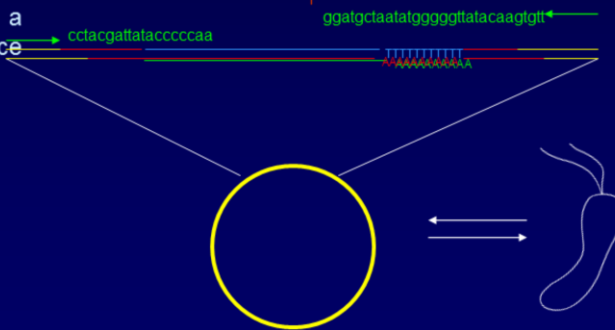
MASARYKŮVA UNIVERZITA  
BRNO

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 genomová homologie
- Experimentální identifikace genů
  - příprava genomově 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