

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
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 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

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

Osnova

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

Přímá vs. reverzní genetika

Revoluce v chápání pojmu genu

Přístupy „klasické“ genetiky

„Reverzně genetický“ přístup

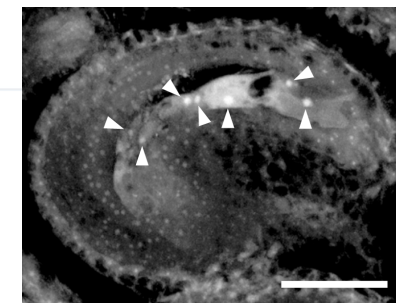
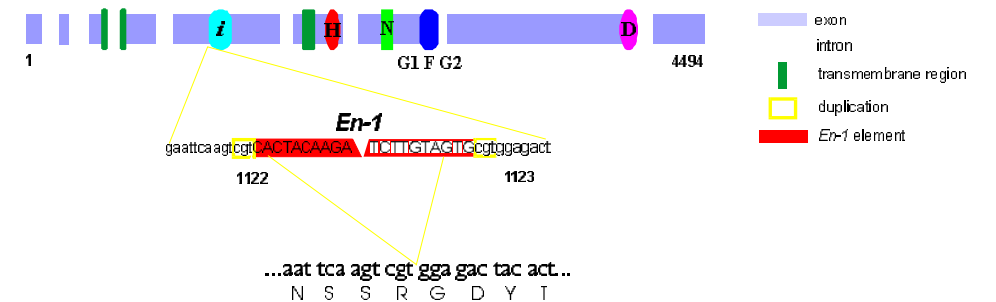
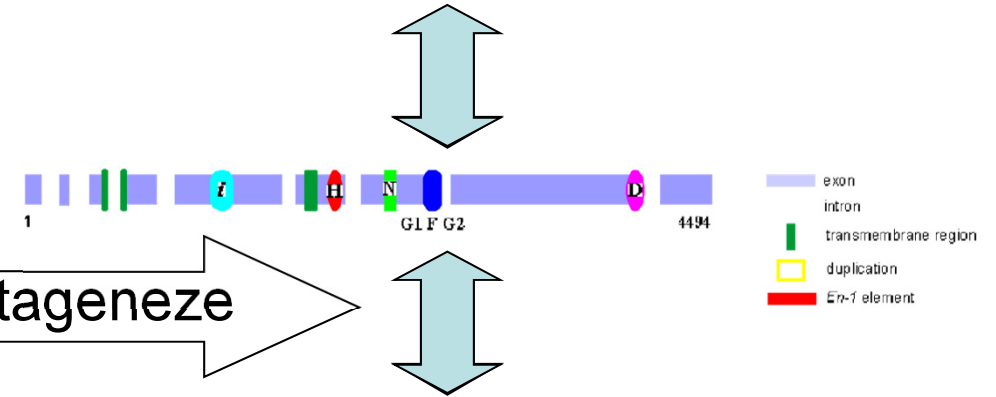
5'TTATATATATATATATAAAAAATAAAATAAAA
GAACAAAAAAGAAAATAAAATA....3'



3

:

1



Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*

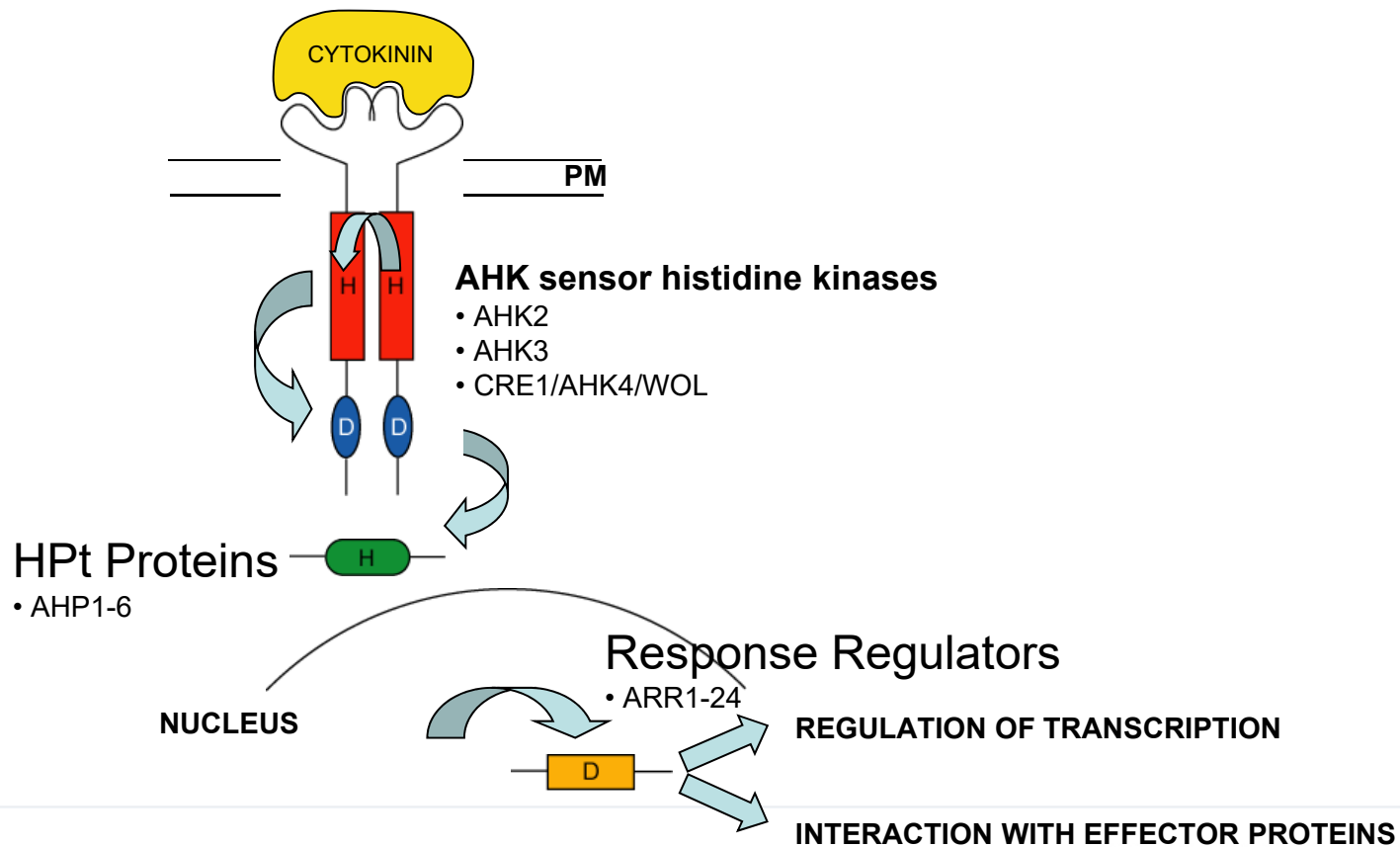


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



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

Identifikace role genu *ARR21* – izolace inz. mutanta

- vyhledávání v databázi inzerčních mutantů (SINS)

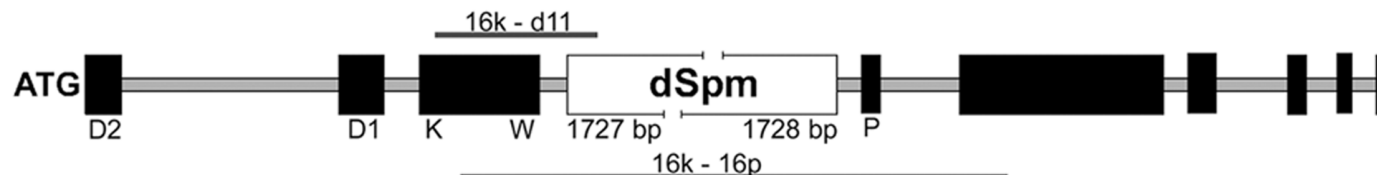
```

Insert_SINS: 01_09_64
Query: 80      tcctagcgttcatgagcgtaccatacttgacaanagagaacgtagccagccatttacagg 139
              |||
Sbjct: 58319  tcctagcgttcatgagcgtaccatacttgacaagagagaacgtagccagccatttacagg 58378
Arr21: 1830
    
```

```

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

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



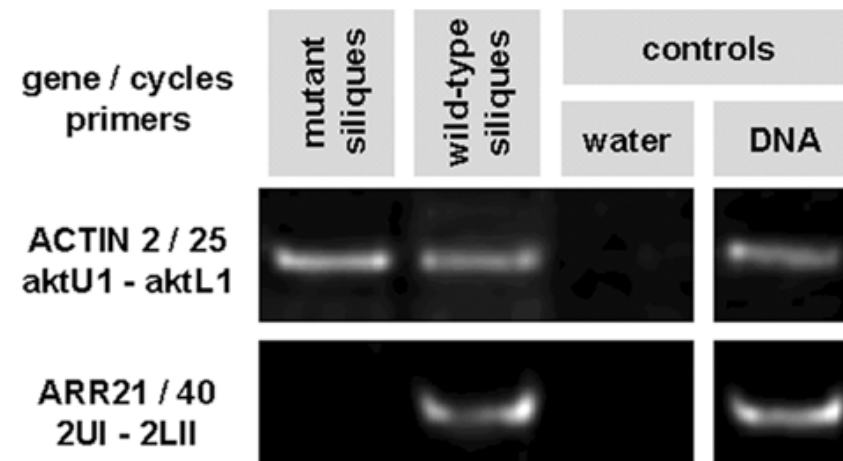
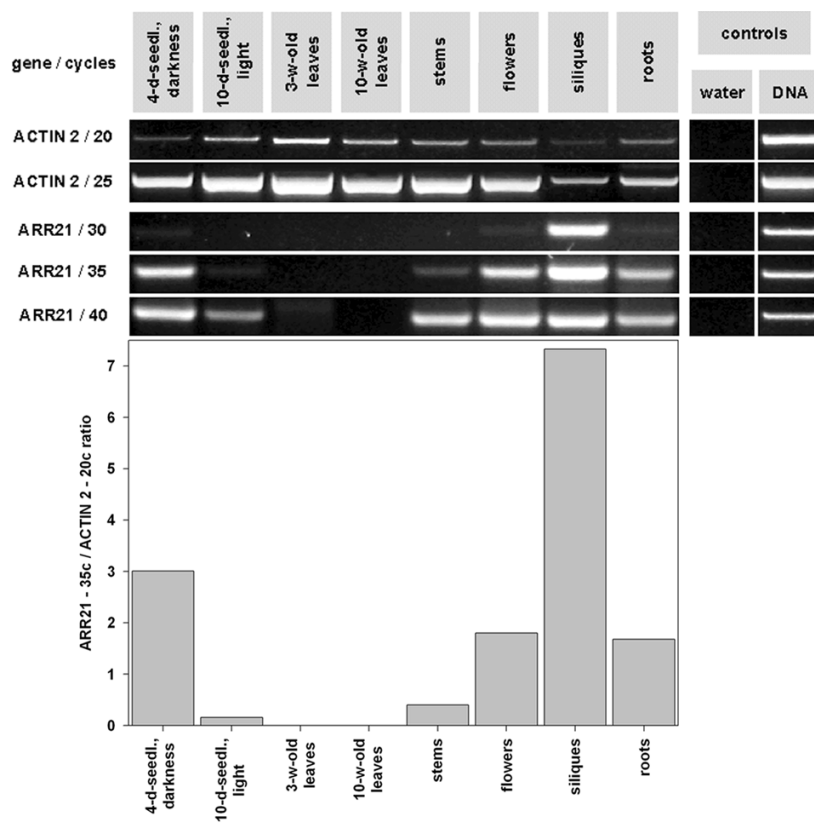
Identifikace role genu *ARR21*

- 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 mutanta potvrzena na úrovni RNA

Identifikace role genu *ARR21* – analýza exprese

Standardní typ

Inzerční mutant

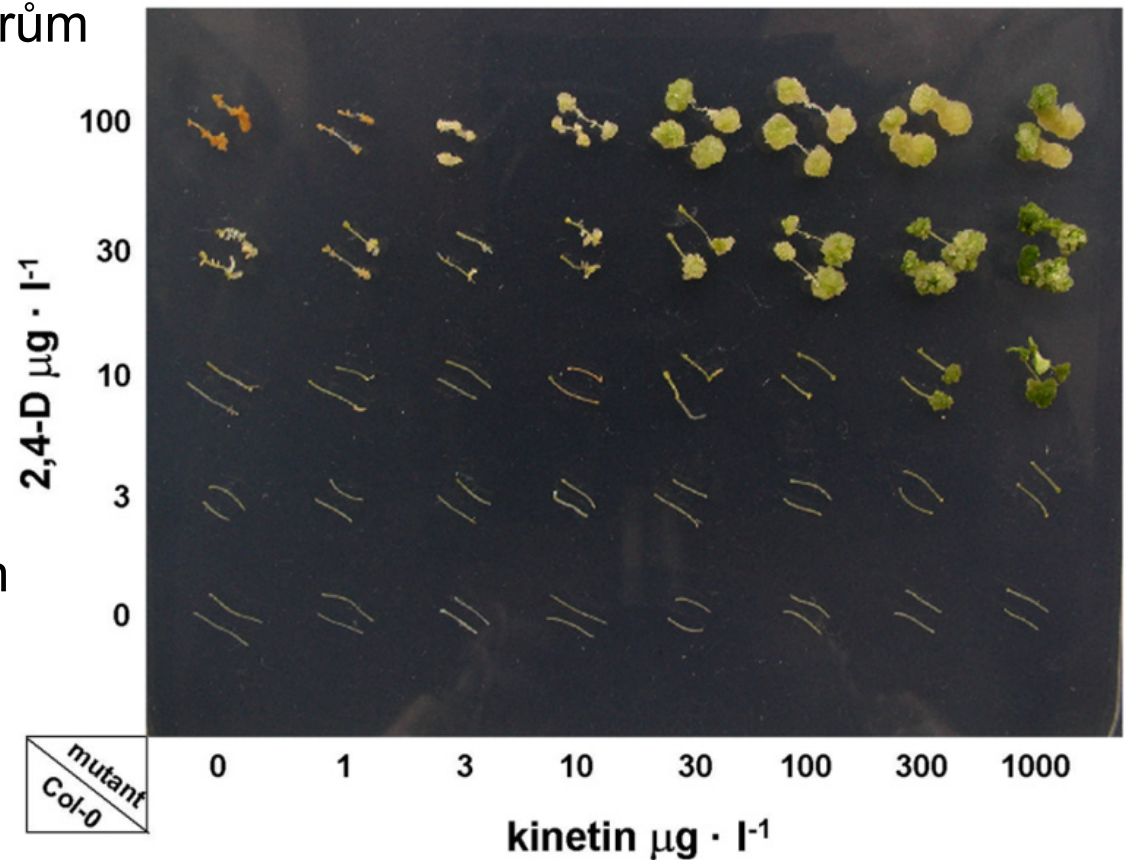


Identifikace role genu *ARR21*

- 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 mutanta potvrzena na úrovni RNA
- Analýza fenotypu inzerčního mutanta

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



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

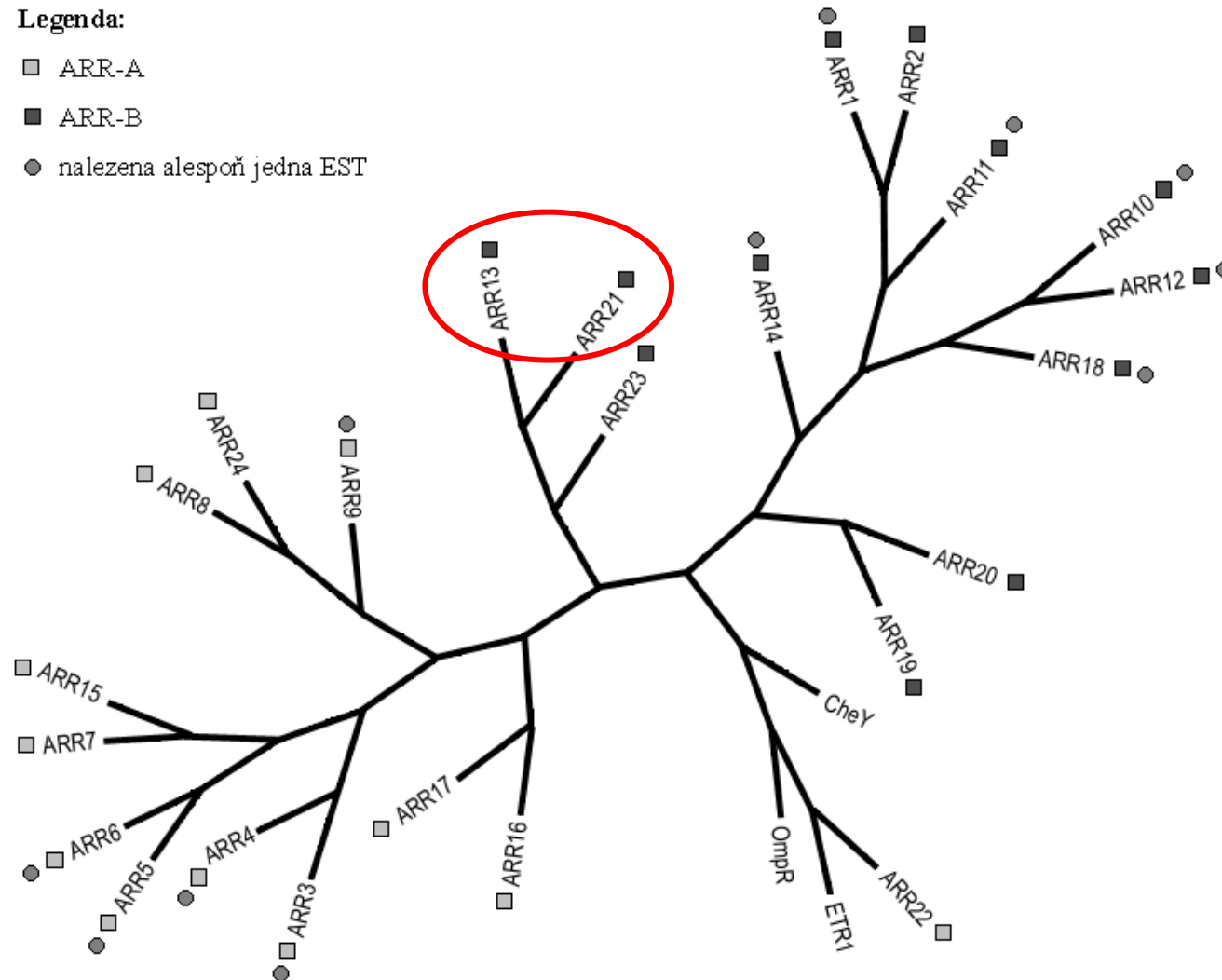
ARR21 – příbuznost ARR genů

Legenda:

□ ARR-A

■ ARR-B

● nalezena alespoň jedna EST



Identifikace role genu *ARR21* – příčiny absence fenotypu

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

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

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

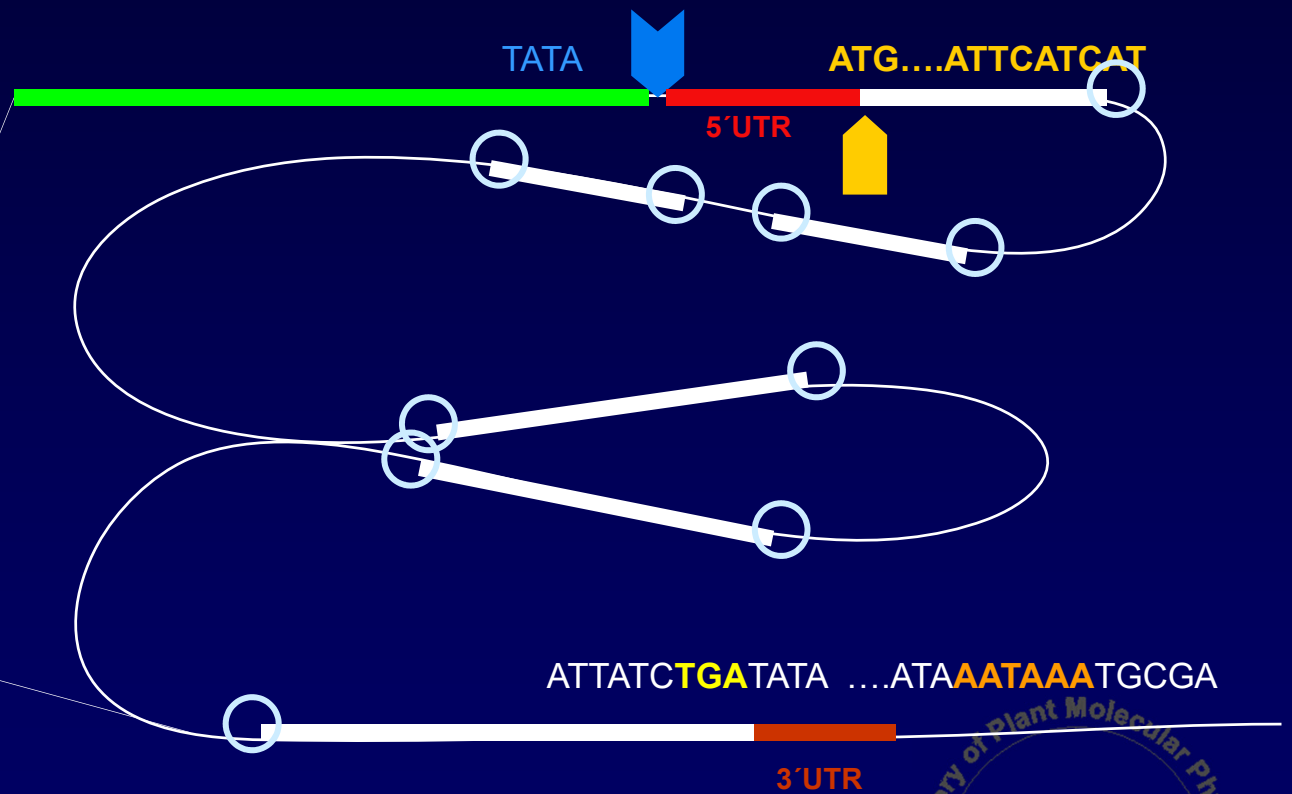


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

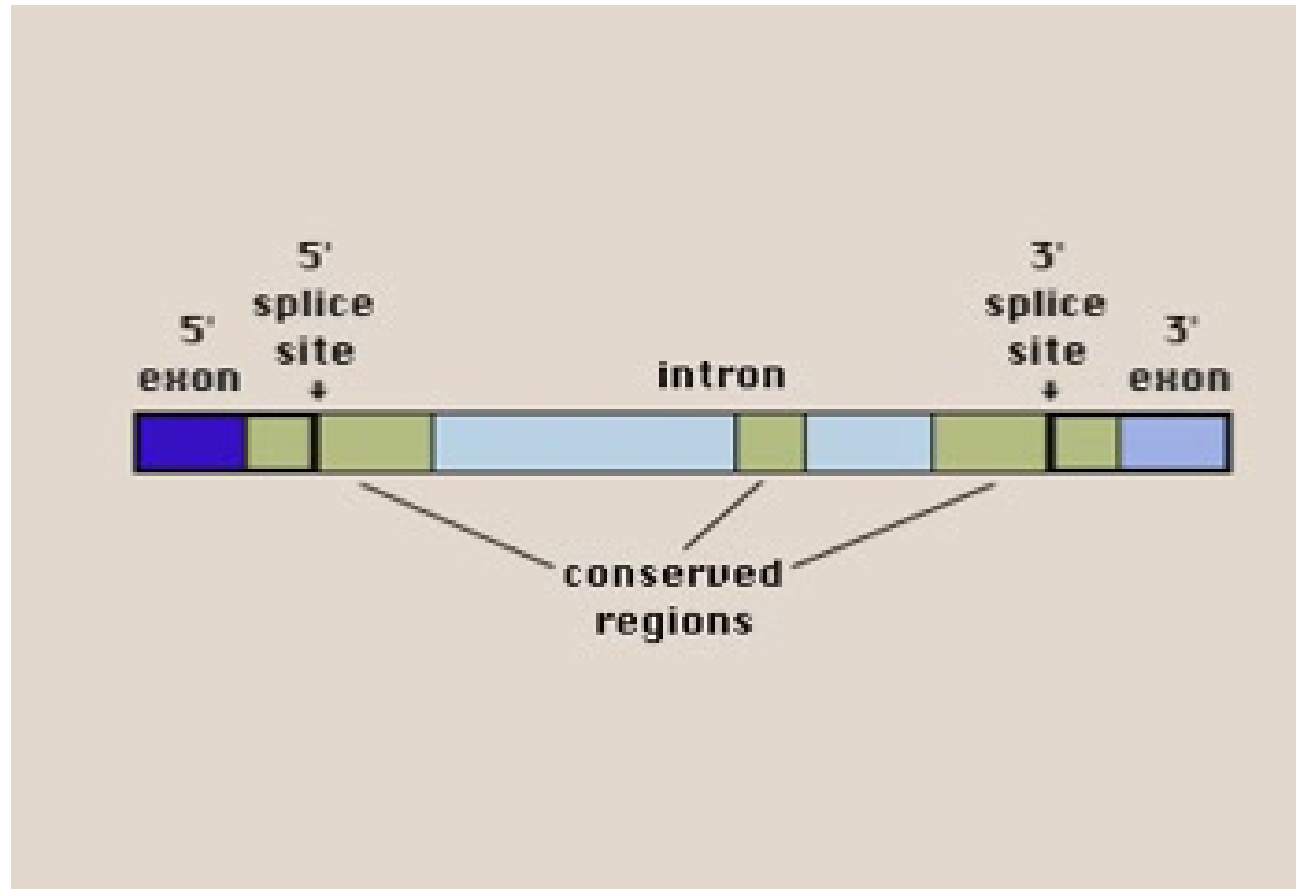
Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Struktura genů

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



Sestřih RNA



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
- 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

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

SplicePredictor

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
AAAGATTTTCATTCAATATAAATACTTGGATAAATACTCTTATTATTTTTCTTTAGTTTATTAAAAAAAACCTCTAATAAAT
ACGAGTTTAAAGTCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAAACGATAAAGTTTACAAAAGTAATATCC
AAGTATCTCATAGTCAACATATATATAGTAATAATTAGTTGACGTATAAGAAAATAAAAAATAAATAAATTAGTATCTTAT
TTTGGGTGGTGCTGACTGGTGACTGGTGACTGCAGAATGCTCGGCAAATGGAACCATATCCCAAGACATGGGTTTTAGAT
```

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

Browse...

... 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

SplicePredictor

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[BF]): 3.00
Local pruning: on
Non-canonical sites: not scored

Sequence 1: your-sequence, from 1 to 9490.

Potential splice sites

t	q	loc	sequence	P	c	rho	gamma	*	P*R*G*
A	<--	75	ttttttcgatctcAGat	0.973	7.16	0.000	0.000	7	(5 1 1)
A	<--	134	attatttttctttAGtt	0.999	14.86	0.000	0.000	7	(5 1 1)
A	<--	500	gatttttggtttAGtc	0.977	7.48	0.000	0.000	7	(5 1 1)
A	<--	780	tctgttattgtatAGct	0.986	8.56	0.000	0.000	7	(5 1 1)
A	<--	848	tattttttgaaatAGat	0.968	6.80	0.000	0.000	7	(5 1 1)
A	<--	1051	caatttatattttAGaa	0.930	5.19	0.000	0.000	7	(5 1 1)
A	<--	1213	ttatttatattttAGtt	0.998	12.14	0.000	0.000	7	(5 1 1)
A	<--	1373	tttctctctcacAGga	0.999	13.17	0.000	0.000	7	(5 1 1)
A	<--	1487	tttatatttattAGtg	0.883	4.04	0.000	0.000	7	(5 1 1)
A	<--	1581	atgtgttcttggAGga	0.982	8.03	0.000	0.000	7	(5 1 1)
A	<--	1781	ggttgtgcaaatAGgg	0.886	4.10	0.000	0.000	7	(5 1 1)
A	<--	2440	taattaaaaattAGat	0.939	5.46	0.000	0.000	7	(5 1 1)
A	<--	2479	catctaaaaattAGat	0.942	5.59	0.000	0.000	7	(5 1 1)
D	---->	2546	aagGTtagta	0.909	4.61	0.885	1.903	15	(5 5 5)
A	<--	2572	tttttttttggcAGca	0.930	5.16	0.000	0.000	7	(5 1 1)
A	<--	2763	ctcaaatccacaaAGgt	0.873	3.86	0.185	0.000	11	(5 5 1)
A	<--	2782	tttcgttttcattAGcg	0.952	5.98	0.220	0.000	11	(5 5 1)
A	<--	3022	ttgttttgaactaAGct	0.956	6.16	0.221	0.000	11	(5 5 1)
A	<--	3048	ctttgcaaatcatAGga	0.973	7.15	0.229	0.000	11	(5 5 1)
A	<--	3171	cgctgtcatttatAGta	0.988	8.74	0.000	0.000	7	(5 1 1)
A	<--	3284	cttttggatcaaaAGgg	0.993	10.03	0.000	0.006	8	(5 1 2)
D	---->	3372	aatGTaagg	0.933	5.28	0.855	1.849	15	(5 5 5)
A	<--	3451	aatgcttcctcgtAGaa	0.916	4.77	0.293	0.065	12	(5 5 2)
A	<--	3581	cgatccgcttctAGgt	0.850	3.47	0.000	0.000	7	(5 1 1)
D	---->	3649	cacGTatta	0.933	5.25	0.000	1.848	11	(5 1 5)
A	<--	3695	ttgtggttatacaAGtt	0.907	4.56	0.000	0.000	7	(5 1 1)
A	<--	4254	attattgtctctcAGat	0.998	12.82	0.000	0.002	8	(5 1 2)
A	<--	4351	tttcttacattgcAGaa	0.991	9.42	0.000	0.000	7	(5 1 1)
A	<--	4633	gtcttgtttctttAGgg	0.879	3.97	0.000	0.000	7	(5 1 1)
A	<--	4976	cttgtgtttctcAGct	0.952	5.98	0.000	0.000	7	(5 1 1)
A	<--	5004	tttttttttggcAGag	0.996	11.17	0.000	0.000	7	(5 1 1)
D	---->	5356	caaGTgaat	0.821	3.04	0.387	0.000	11	(5 5 1)
D	---->	5384	ttgGTaaga	0.941	5.54	0.478	0.090	13	(5 5 3)
A	<--	5403	actctgtttctttAGct	0.894	4.26	0.000	0.000	7	(5 1 1)
A	<--	5441	ctttctcttaacAGaa	0.995	10.43	0.387	0.000	11	(5 5 1)
A	<--	5472	ttgttaaaattacAGct	0.965	6.62	0.478	0.090	13	(5 5 3)
D	---->	5745	gcgGTaaga	0.991	9.48	0.990	1.956	15	(5 5 5)
A	<--	5808	catcatatcctaaAGgt	0.948	5.83	0.458	0.000	11	(5 5 1)
A	<--	6135	ggctattattatAGgt	0.999	13.59	0.508	0.050	12	(5 5 2)
A	<--	6552	ggattttcacctcAGag	0.938	5.42	0.000	0.000	7	(5 1 1)



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/>)

NetGene 2

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	TTCCAACAC	^	GTTAATATTT	
1906	0	+	0.99	CGGTGAACGG	^	GTGAGAACAT	
3582	1	+	1.00	GCCGTTCTAG	^	GTAATCTGCG	H
3765	1	+	1.00	TTGCGTCTCG	^	GTAATCTGCG	H
4134	0	+	0.74	TCAAACACAG	^	GTTGTTAAAA	
4619	1	+	0.74	AGCAAGAAAG	^	GTCTTGTTTC	
4915	0	+	0.94	CGTTCCTCTG	^	GTAATACTG	
5356	0	+	0.87	TCTCAACCA	^	GTGAATGTTT	
5384	1	+	1.00	GATTTGTTG	^	GTAAGACTCT	H
5809	1	+	1.00	TATCCTAAG	^	GTTGTGCCAA	
6057	0	+	1.00	GCAGTCTTTG	^	GTAAGCTACT	H
6096	1	+	0.74	CTCTTCACAA	^	GTAATCTAG	
7369	0	+	1.00	GGACTGCCAA	^	GTAAGTTTAA	H
7886	0	+	0.74	GAACAAAATG	^	GTTTAGATGAA	
9323	0	+	0.74	GAAGATTAGG	^	GTTTTTCTCT	

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	TATTTTTT	AG	^TTATGGAGAC	
1221	2	+	0.87	AGTTATGGAG	^	ACAAGAATCG	
1373	0	+	0.71	TCTCTCACAG	^	GACACAGAAT	
1487	1	+	0.81	ATATTGATAG	^	TGGGACATTA	
3284	0	+	0.87	GTTATCAAAG	^	GGTTTCGACT	
4254	0	+	1.00	TGTTCTTCAG	^	ATCGCACCAT	H
4832	2	+	0.54	AAAATTGCAG	^	TCCAGTGGC	
5004	0	+	0.94	TTTTTGCCAG	^	AGATACACAC	
5472	1	+	0.96	AAAATTACAG	^	CTCTGCTCAA	
6135	0	+	1.00	ATTATTATAG	^	GTAAGATTAA	H
6490	1	+	0.90	AAAGTTACAG	^	TGGTGGAGAA	
6744	0	+	0.59	TGTCAAACAG	^	TTCGTAGAG	
7447	0	+	0.96	TTCTGCACAG	^	ATGCCAGAAA	
7780	2	+	0.76	TCCATTTT	TCAG	^ATACAGAACA	
7786	2	+	0.92	TCAGATACAG	^	AACACATGCA	



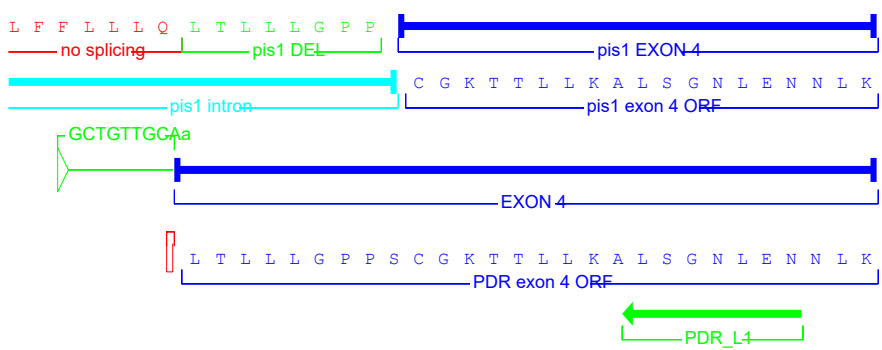
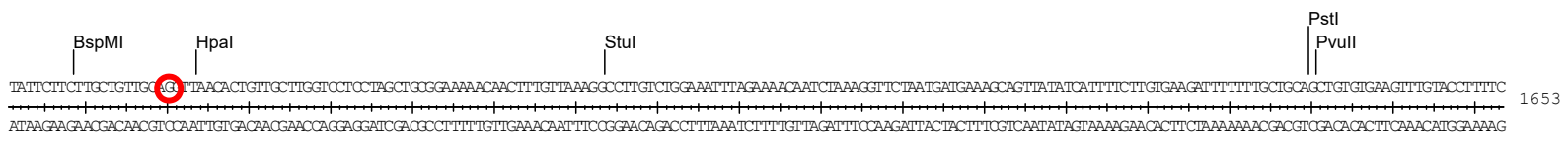
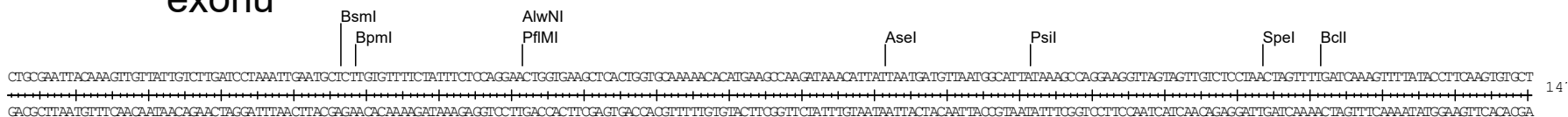
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

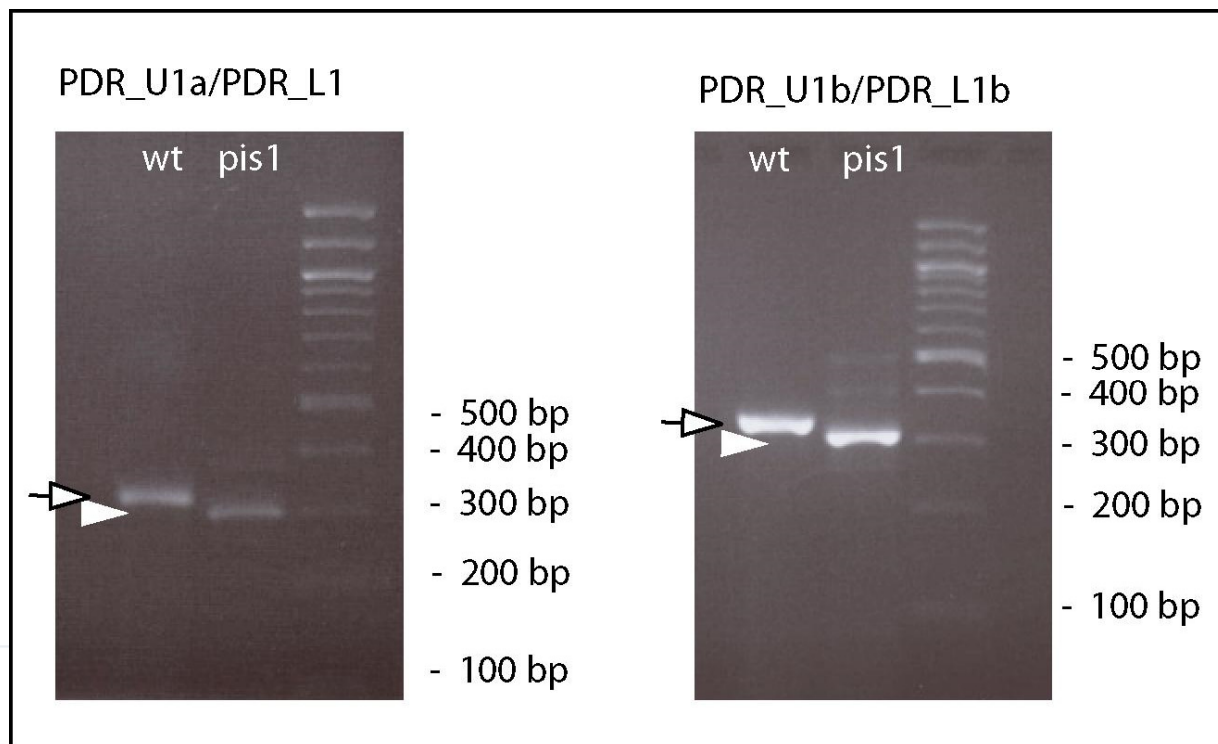
OP Vzdělávání
pro konkurenceschopnost



Evropským sociálním fondem
a státním rozpočtem České republiky

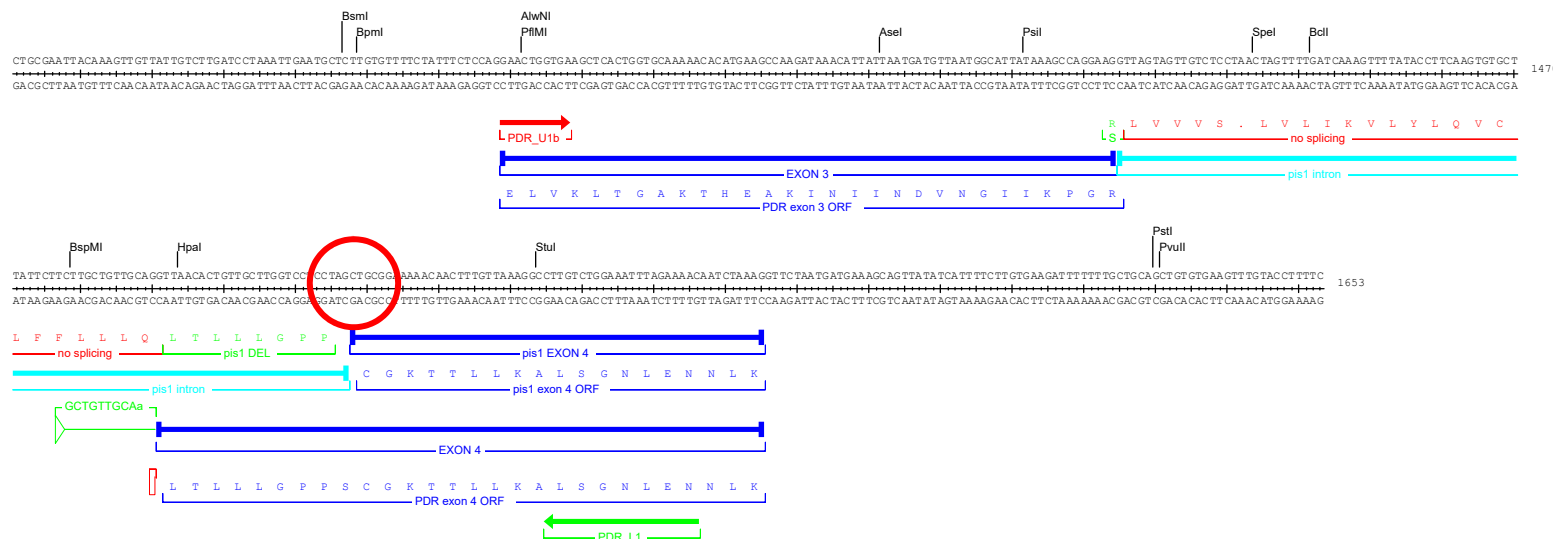
Sestřih RNA a adaptace

- 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



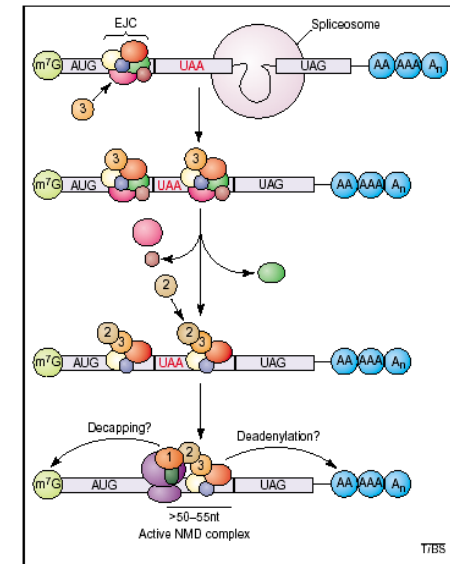
Sestřih RNA a adaptace

- 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



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)



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://hollywood.mit.edu/GENSCAN.html>)
 - GeneMark.hmm (<http://opal.biology.gatech.edu/GeneMark/>)
- interní:
 - MZEF (<http://rulai.cshl.org/tools/genefinder/>)

GENESCAN

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):

```
GAGGAGGCACAAAATGACGAATATACAAAATGATCTTAAACAGCTAAACTATATTGGACATTTTTCGATC
TCAGATATA
AAAGATTTTCATTCAATATAACTTGGATAAATACTCTTATTATTTTCTTTAGTTTATAAAAAAACCT
CTAATAAAT
ACGAGTTTAAAGTCCACAAAATCGCTTAGACTAAAATACACCATATAATTTCAAACGATAAAGTTTACAAA
GTAATATCC
AAGTATCTCATAGTCAACATATATATAGTAATAATTAGTTGACGTATAAGAAAAATAAATAAATAAATTA
GTATCTTAT
TTTGGGTGGTCTGACTGGTGACTGGTGACTGCAGAATGCTCGGCAAAATGGAACCATATCCCAAGACATGG
GTTTTAGAT
AGAACAAAATAAGTGTCCGAAGGAATGATATTTAAAAGTCAAATAGAATAATTATAAATATTGTAATTAGCA
AATAAAAAC
```

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

[Back to the top](#)



EVROPSKÁ UNIE



MLÁDEŽE A TĚLOVÝCHOVY

pro konkurenceschopnost

"ČINA B"

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

GENESCAN

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%)

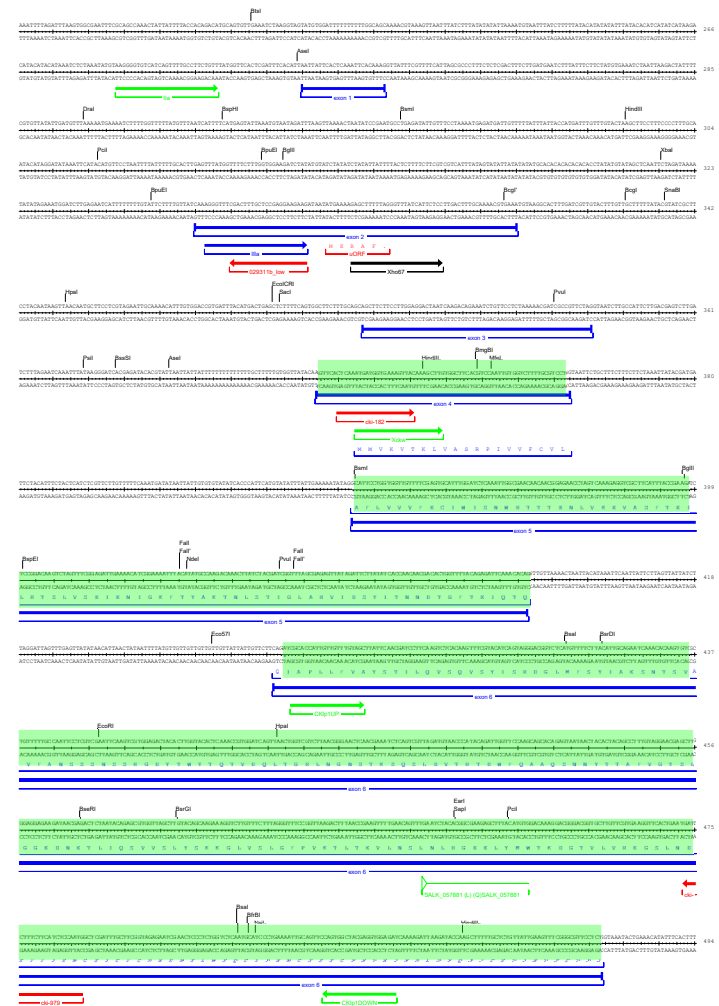
Parameter matrix: Arabidopsis.smat

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	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	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	0.977	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

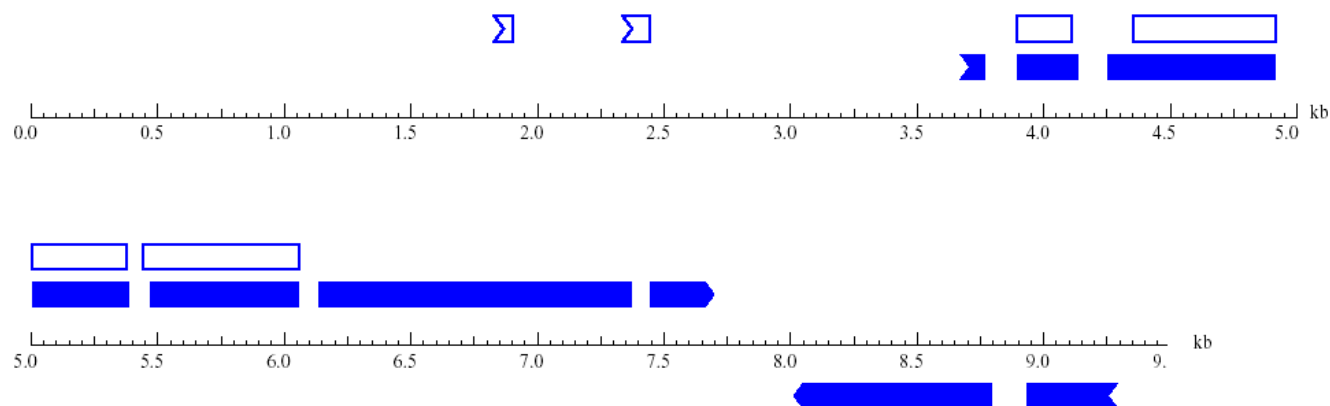
Suboptimal exons with probability > 0.100

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



GENESCAN

GENSCAN predicted genes in sequence 02:56:23



Key:



Initial exon



Internal exon



Terminal exon



Single-exon gene

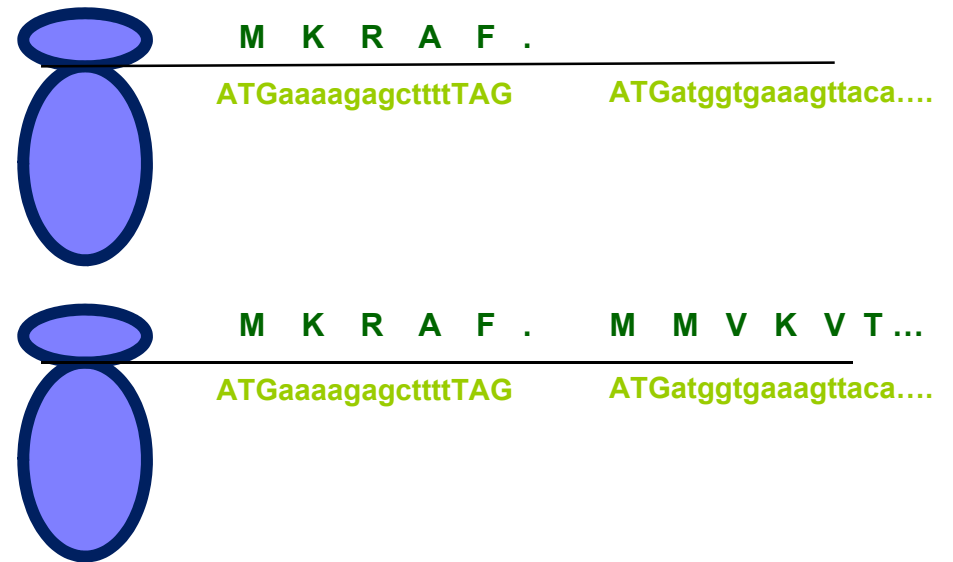
Optimal exon

Suboptimal exon

Regulace translace

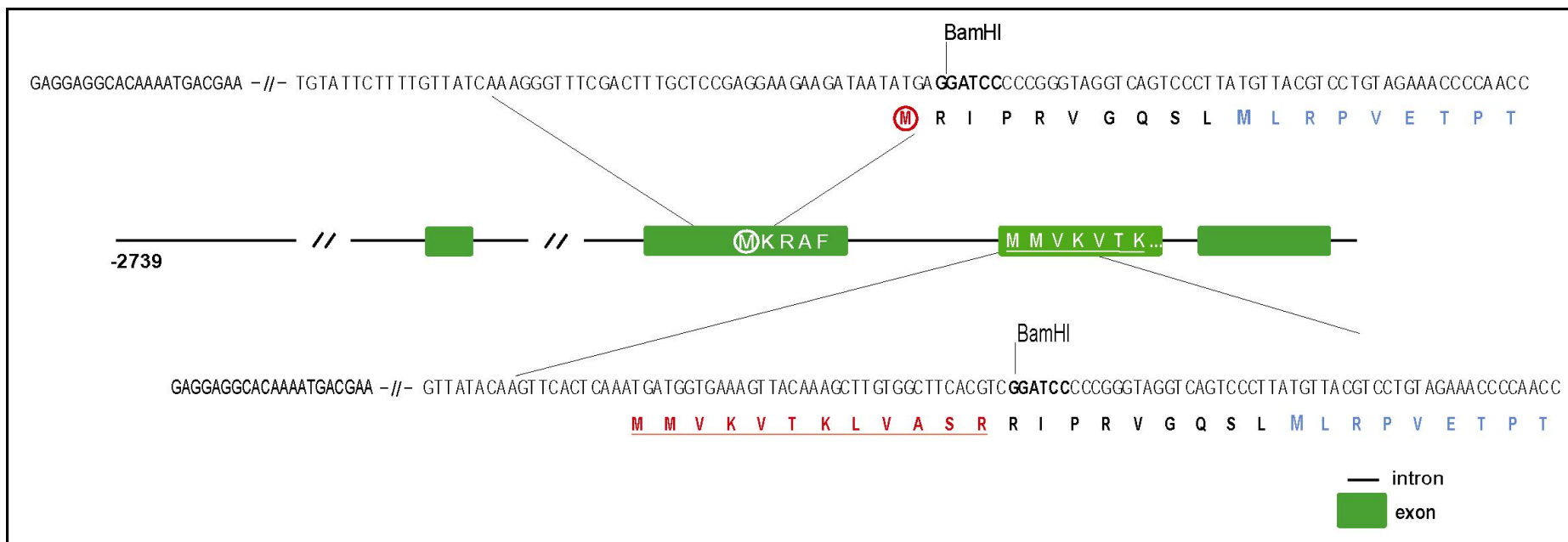
- Funkční význam sestřihu v nepřekládaných oblastech - důležitá regulační součást genů

- 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



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



Genové modelování

- programy pro genové modelování
 - zohledňují také další parametry, např. návaznost ORF
 - **Genescan** (<http://hollywood.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** (<https://ccb.jhu.edu/software/glimmerhmm/>)

GeneMark

Result of last submission:

[View PDF Graphical Output](#)

GeneMark.hmm Listing

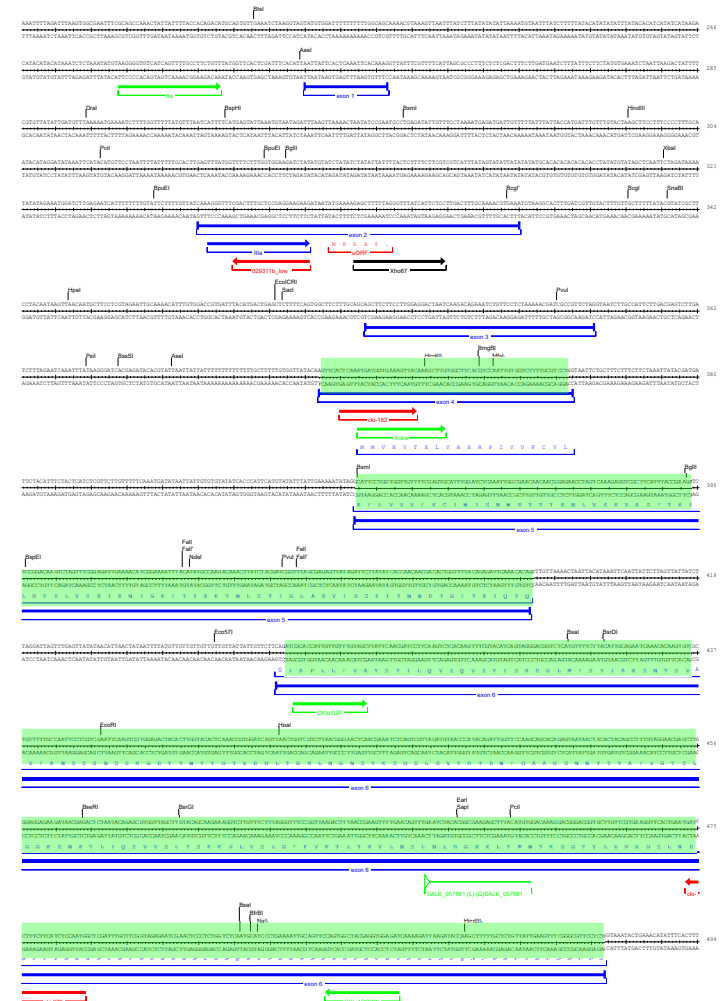
Go to: [GeneMark.hmm Protein Translations](#)

Go to: [Job Submission](#)

Eukariotic GeneMark.hmm version bp 3.9 April 25, 2008
 Sequence name: CK11
 Sequence length: 5043 bp
 G+C content: 38.79%
 Matrices file: /home/genemark/euk_gbm.matrices/athaliazna_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	2734 3317 584		2 3 - -
1	6	+	Internal	3397 4629 1233		1 3 - -
1	7	+	Terminal	4709 4921 213		1 3 - -



/ZDĚLÁVÁNÍ

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

GeneMark

Result of last submission:

[View PDF Graphical Output](#)

GeneMark.hmm Listing

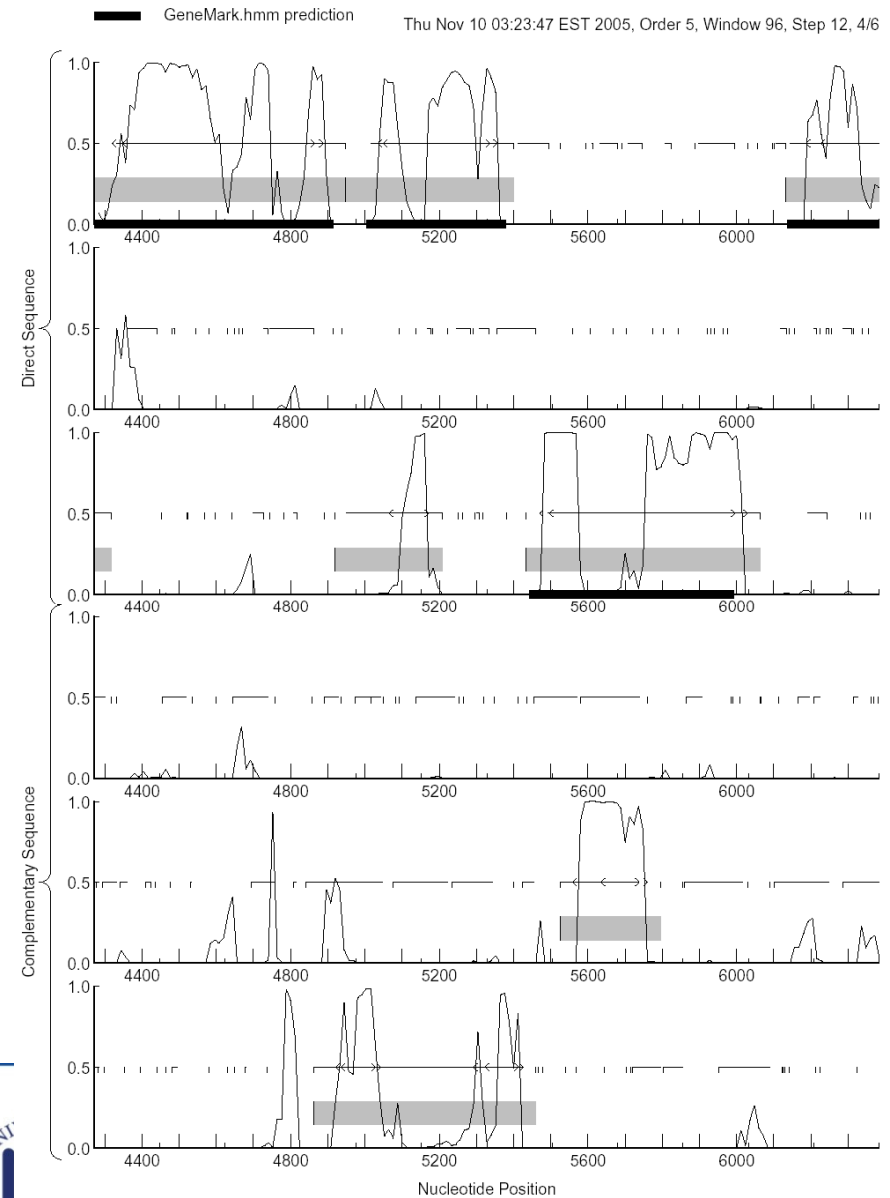
Go to: [GeneMark.hmm Protein Translations](#)

Go to: [Job Submission](#)

Eukaryotic GeneMark.hmm version bp 3.9 April 25, 2008
 Sequence name: CK11
 Sequence length: 5043 bp
 G+C content: 38.79%
 Matrices file: /home/genemark/euk_gbm.matrices/athaliana_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	2734 3317	584	2 3 - -
1	6	+	Internal	3397 4629	1233	1 3 - -
1	7	+	Terminal	4709 4921	213	1 3 - -



ÁVÁNÍ

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/>)
 - porovnávání s proteinovými databázemi
 - **BLASTX** (<http://www.ncbi.nlm.nih.gov/BLAST/>)
 - **Genewise** (<https://www.ebi.ac.uk/Tools/psa/genewise/>)

porovnávají proteinovou sekvenci s genomovou DNA (po zpětném překladu), je nutná znalost aminokyselinové sekvence
 - porovnávání s homologními genomovými sekvencemi z příbuzných druhů
 - **VISTA** (<http://genome.lbl.gov/vista/index.shtml>)

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



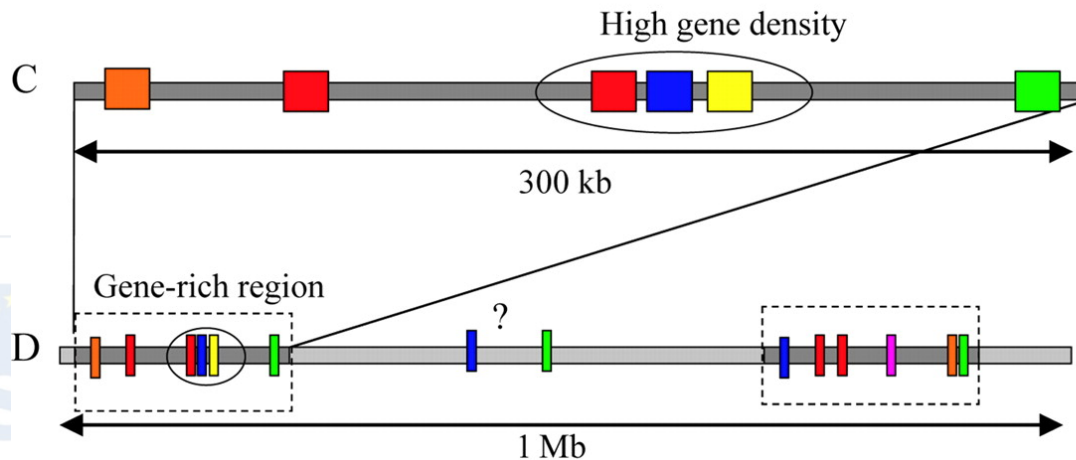
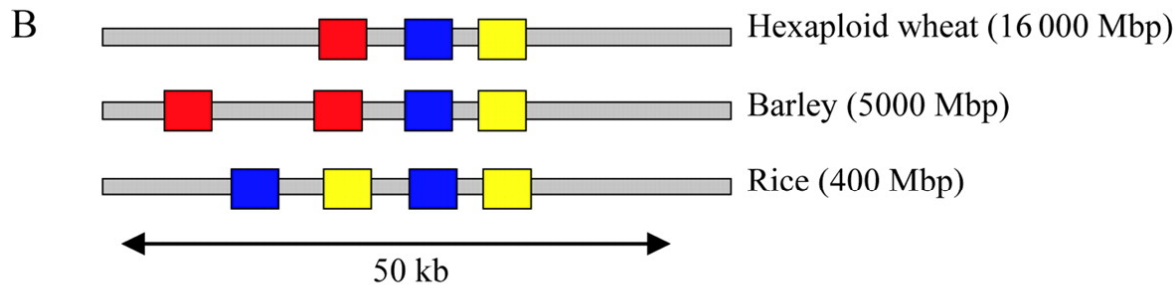
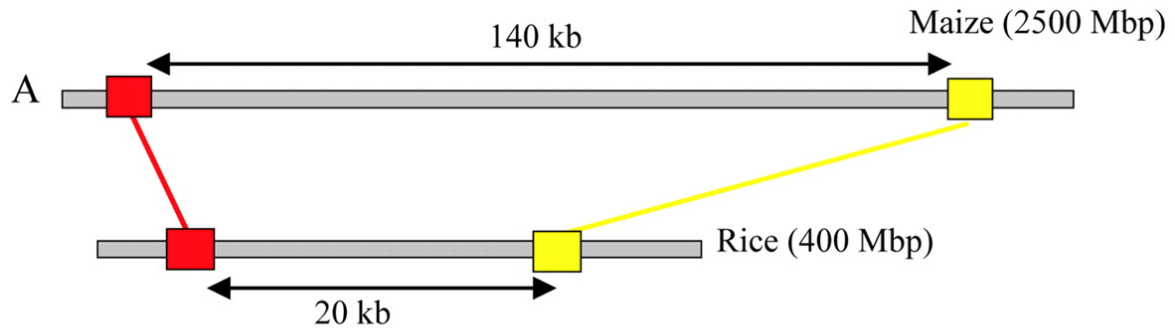
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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)

Genomová kolinearita



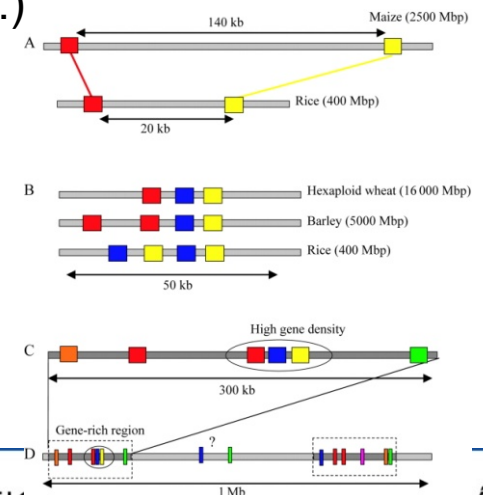
Feuillet and Keller, 2002

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.)

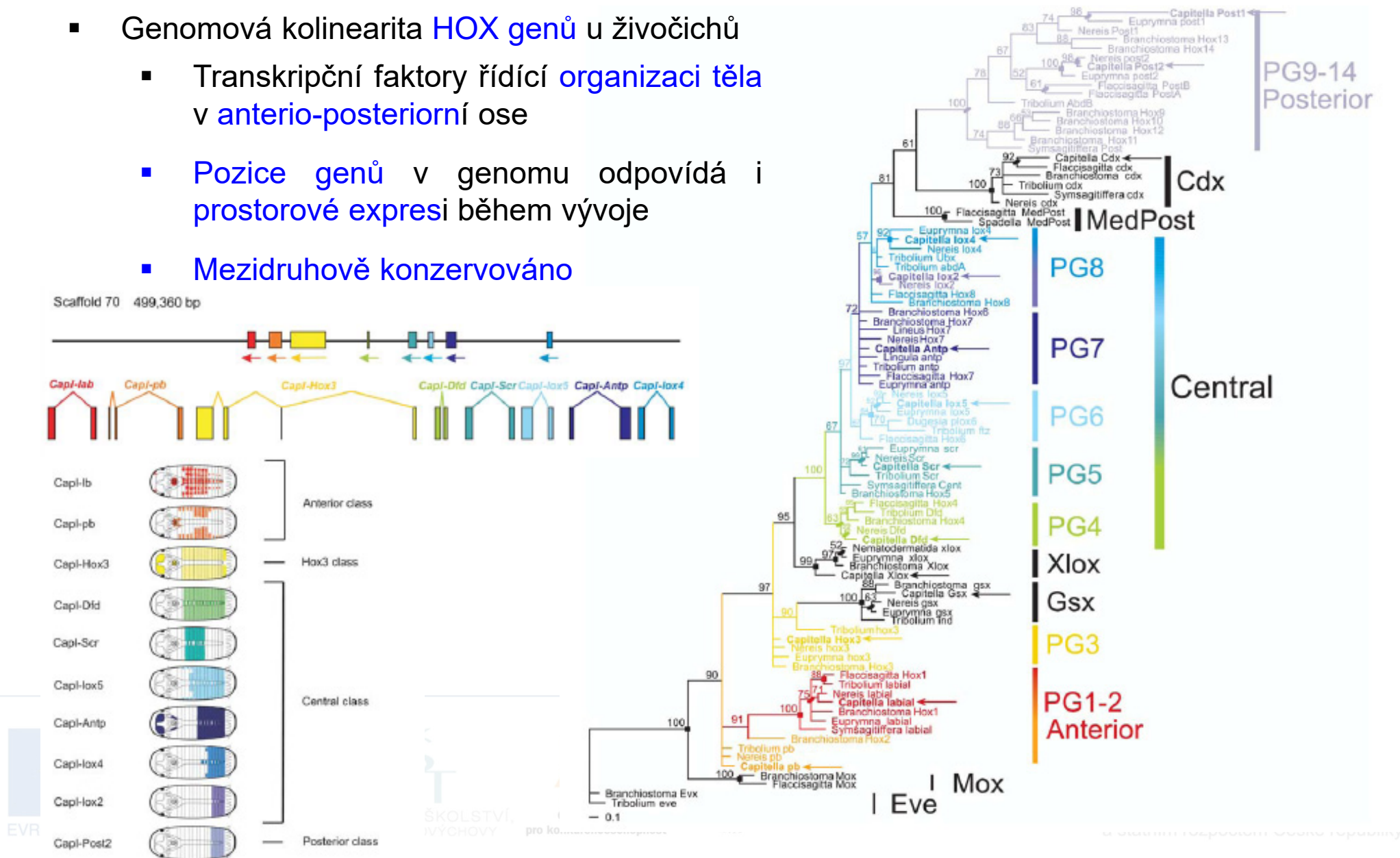


INVESTICE DO ROZVOJE VZDELAVÁNÍ

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



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

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

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 %

Osnova

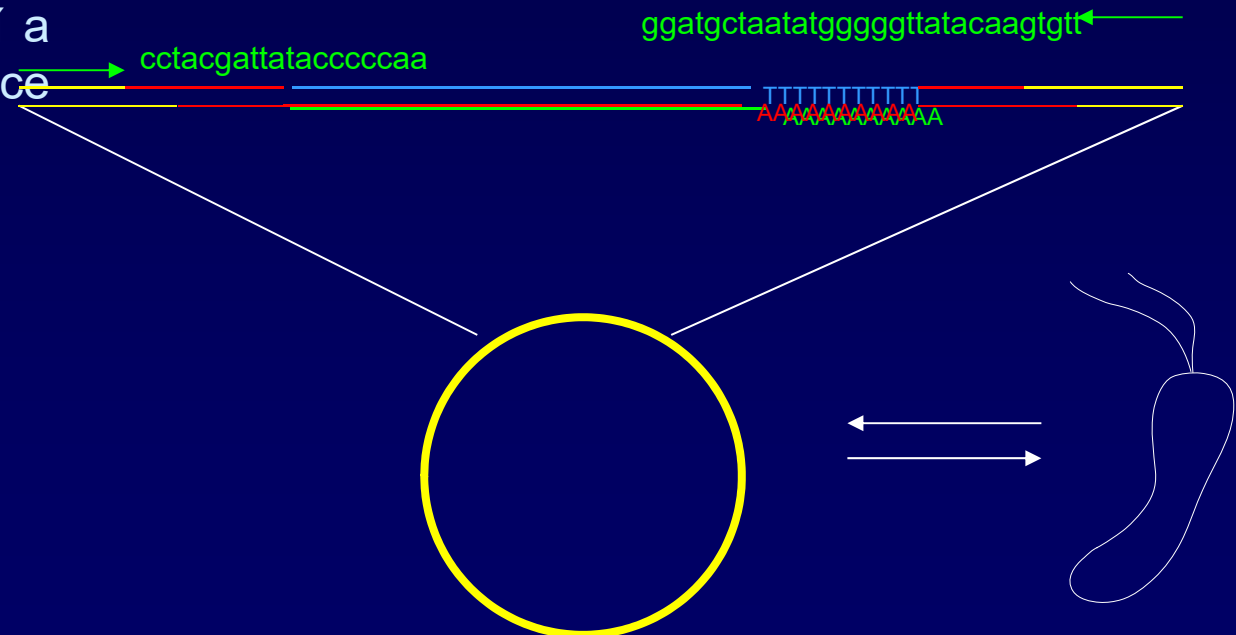
- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Identifikace genů *ab initio*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven pomocí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

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

NCBI Nucleotide database entry for Agrobacterium tumefaciens plasmid Ti. The entry includes details such as LOCUS, DEFINITION, ACCESSION, VERSION, FEATURES, and COMMENT. The COMMENT section includes a link to the GenBank entry for the plasmid.



Klíčové koncepty

- **Přímá vs. reverzní genetika**
 - Gen jako faktor určující frekvenci fenotypu vs. fyzická entita, která existuje nezávisle na fenotypu
- **Identifikace genů *ab initio***
 - struktura genů a často i jejich poloha v genomu je konzervovaná
- **Experimentální identifikace genů**
 - příprava genově obohacených knihoven
 - EST knihovny
 - přímá a reverzní genetika (přednáška 03)

Diskuse



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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