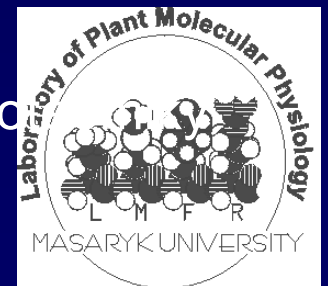


Základy genomiky



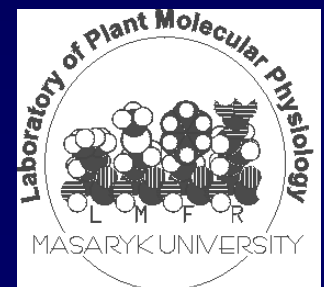
Jan Hejato

Masarykova univerzita, Laboratoř funkční genomiky a proteomiky
Laboratoř molekulární fyziologie rostlin



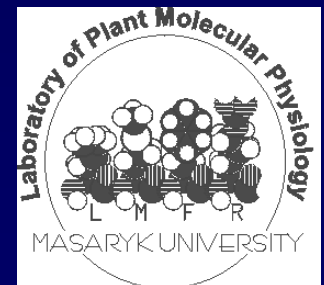
Základy genomiky II.

- Zdrojová literatura ke kapitole II:
 - 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)



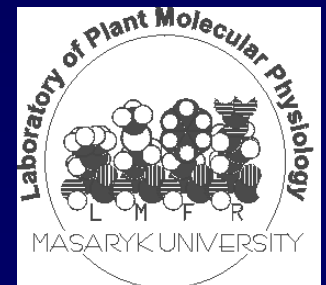
Základy genomiky II.

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



Základy genomiky II.

- 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



3

:

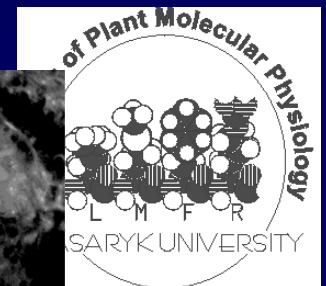
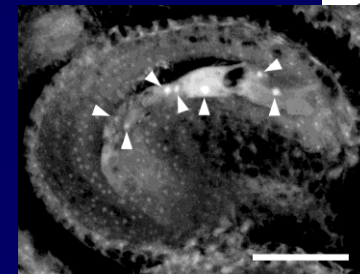
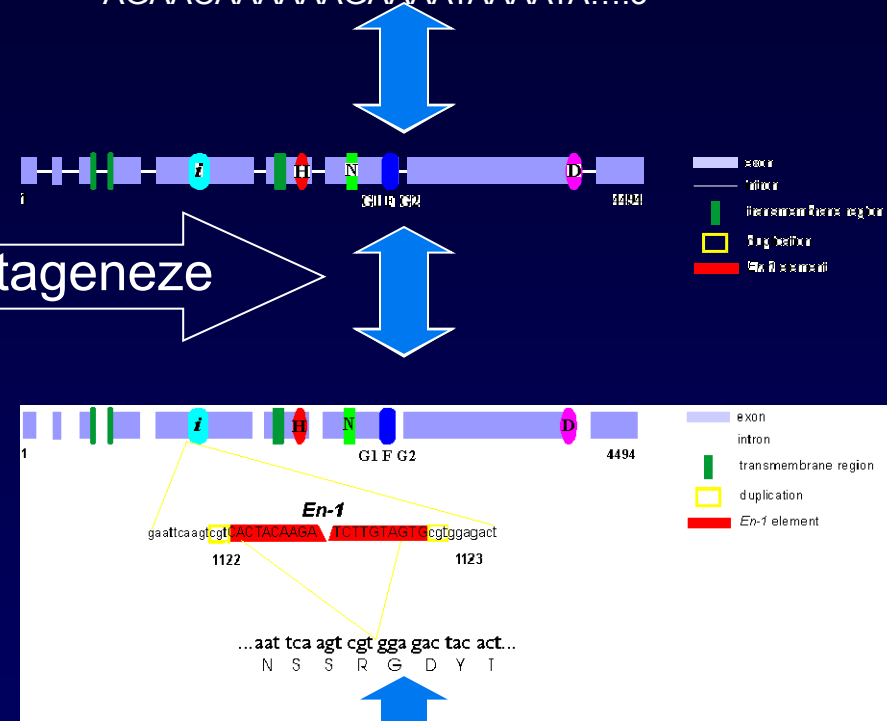
1



?

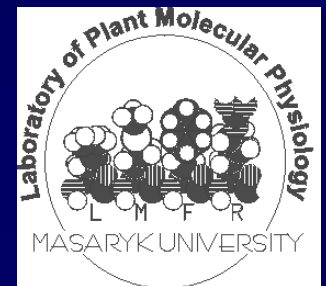
„Reverzně genetický“ přístup

5'TTATATATATATATATTAATAATAATAATA
AGAACAAAAAGAAAATAATA...3'

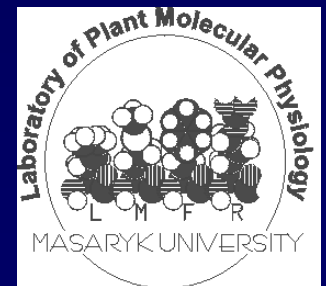
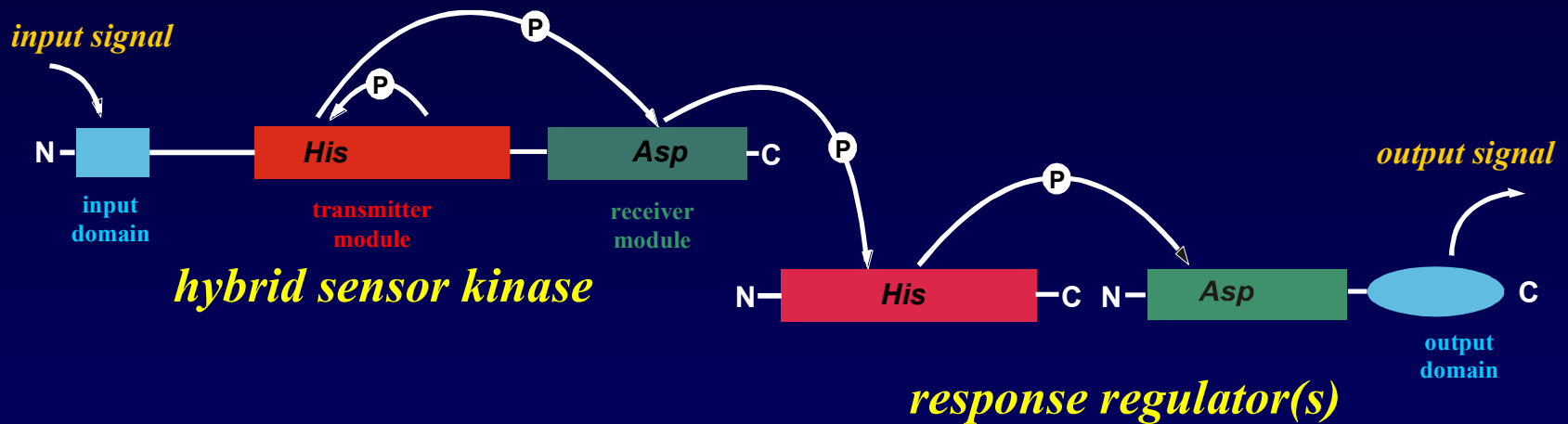


Identifikace role genu *ARR21*

- Předpokládaný přenašeč signálu u dvoukomponentního signálního systému *Arabidopsis*

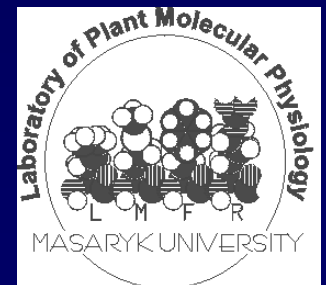


Identifikace role genu *ARR21* regulátor odezvy v dvoukomponentním signálním systému



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



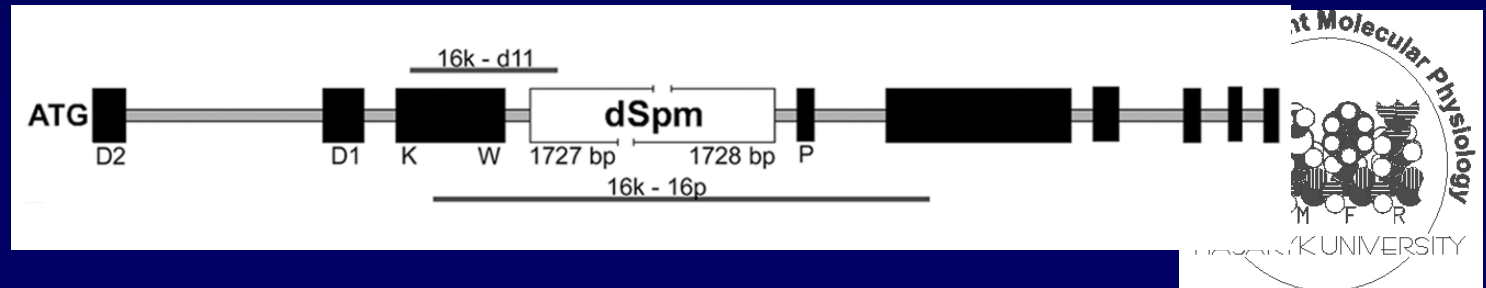
Identifikace role genu *ARR21* identifikace inzerčního mutanta

- vyhledávání v databázi inzerčních mutantů (SINS)

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

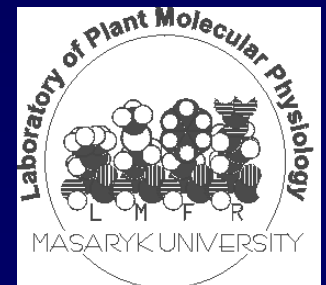
Insert_SINS: 01_09_64
Query: 140     tttgatattctctgtgcaaaaatgttttggattttactgt 179
              |||
Sbjct: 58379  tttgatattctctgtgcaaaaatgttttggattttactgt 58418
Arr21: 1890
```

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



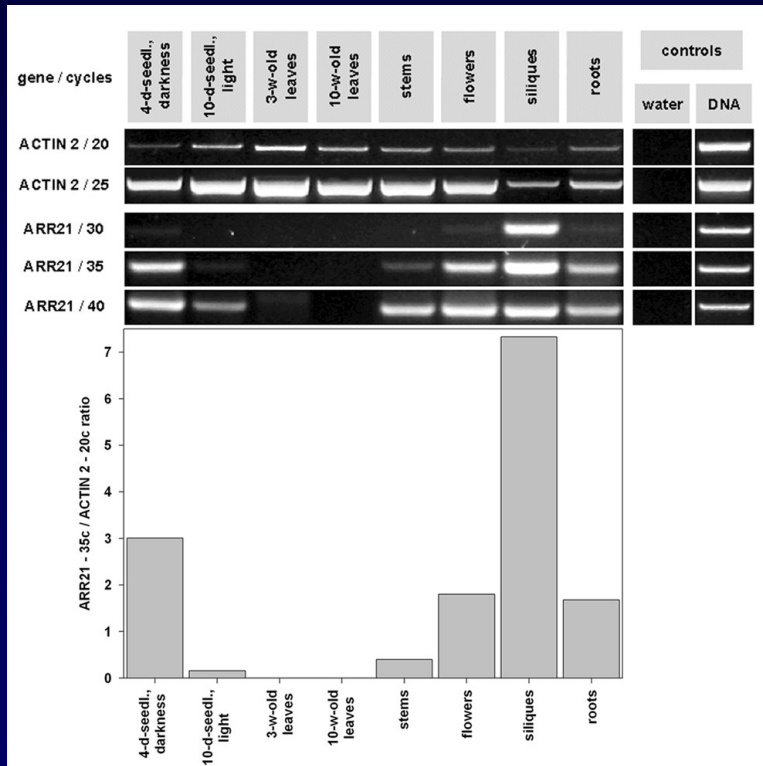
Identifikace role genu *ARR21*

- 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

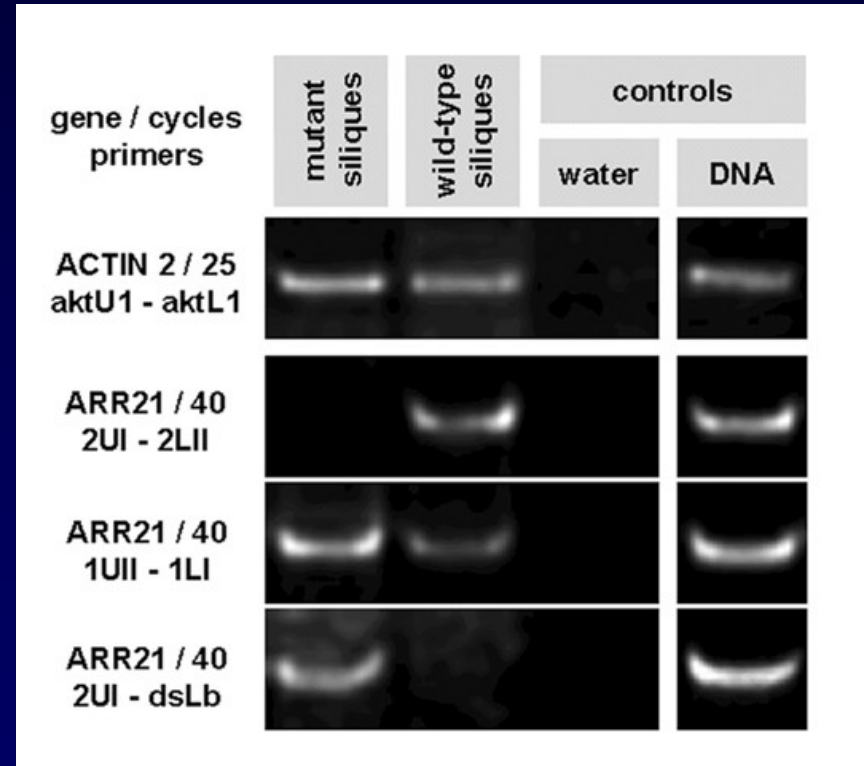


Identifikace role genu *ARR21* analýza expresního profilu

Standardní typ

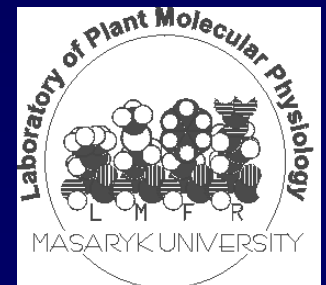


Inzerční mutant



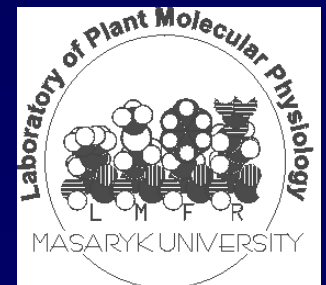
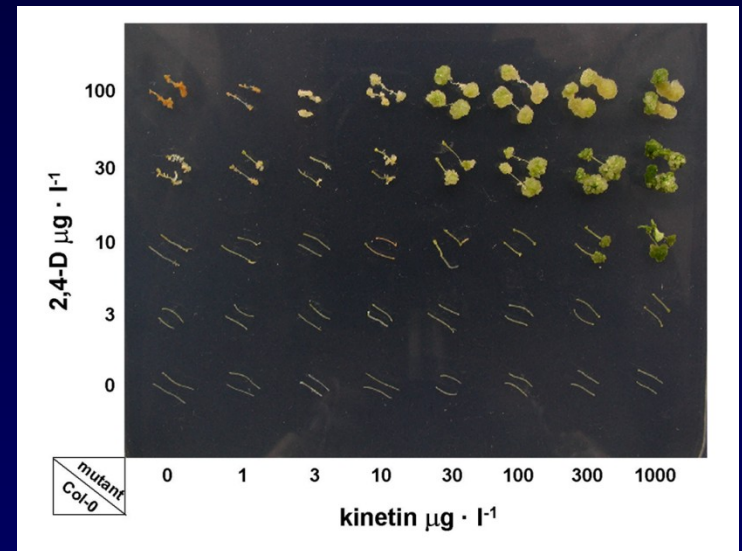
Identifikace role genu *ARR21*

- 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 mutantu potvrzena na úrovni RNA
- Analýza fenotypu inzerčního mutantu



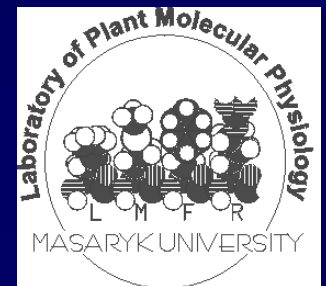
Identifikace role genu *ARR21* analýza fenotypu inzerčního mutanta

- Analýza citlivosti k regulátorům růstu rostlin
 - 2,4-D a kinetin
 - etylén
 - světlo různých vlnových délek
- Doba kvetení i počet semen nezměněn



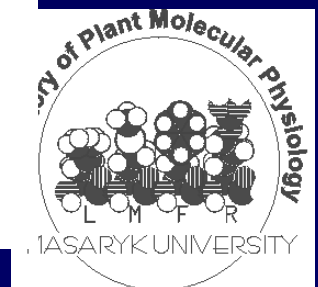
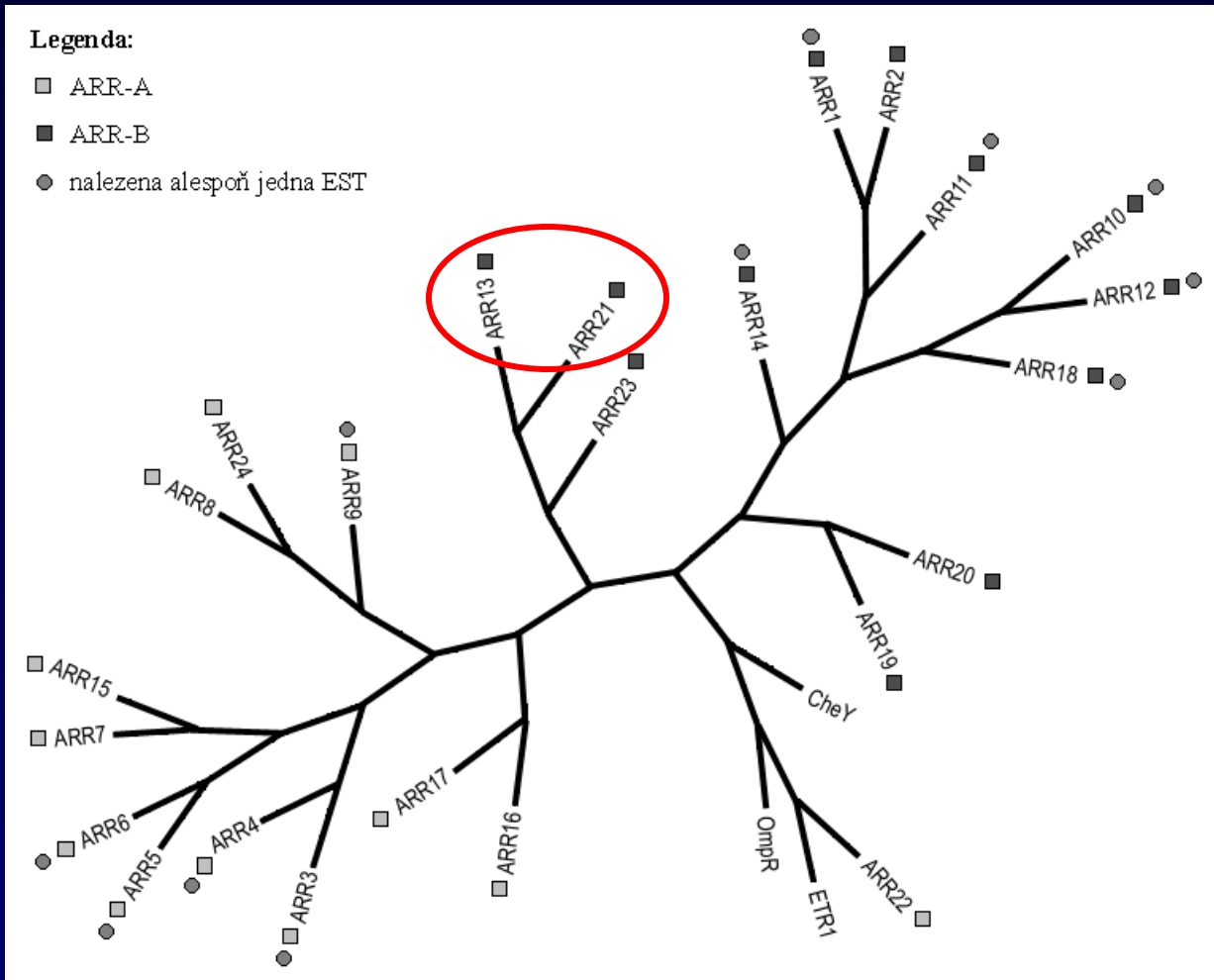
Identifikace role genu *ARR21* možné příčiny absence odchylek fenotypu u inzerčního mutanta

- Funkční redundance v rámci genové rodiny?



Identifikace role genu *ARR21*

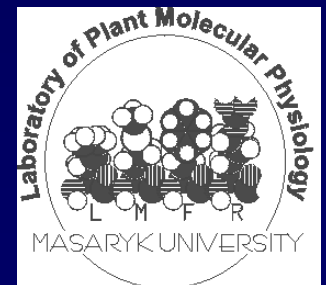
příbuznost jednotlivých ARR genů u *Arabidopsis*



Identifikace role genu *ARR21*

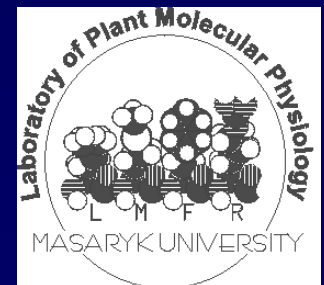
možné příčiny absence odchylek fenotypu u
inzerčního mutanta

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



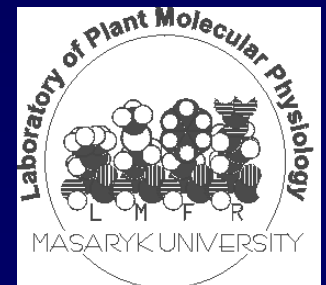
Inzerční mutageneze ve funkční genomice *Arabidopsis thaliana*

- 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 mutageneze v případě identifikace funkce genu *ARR21* ve vývoji *Arabidopsis* byla neúspěšná, pravděpodobně v důsledku funkční redundance v rámci genové rodiny



Základy genomiky II.

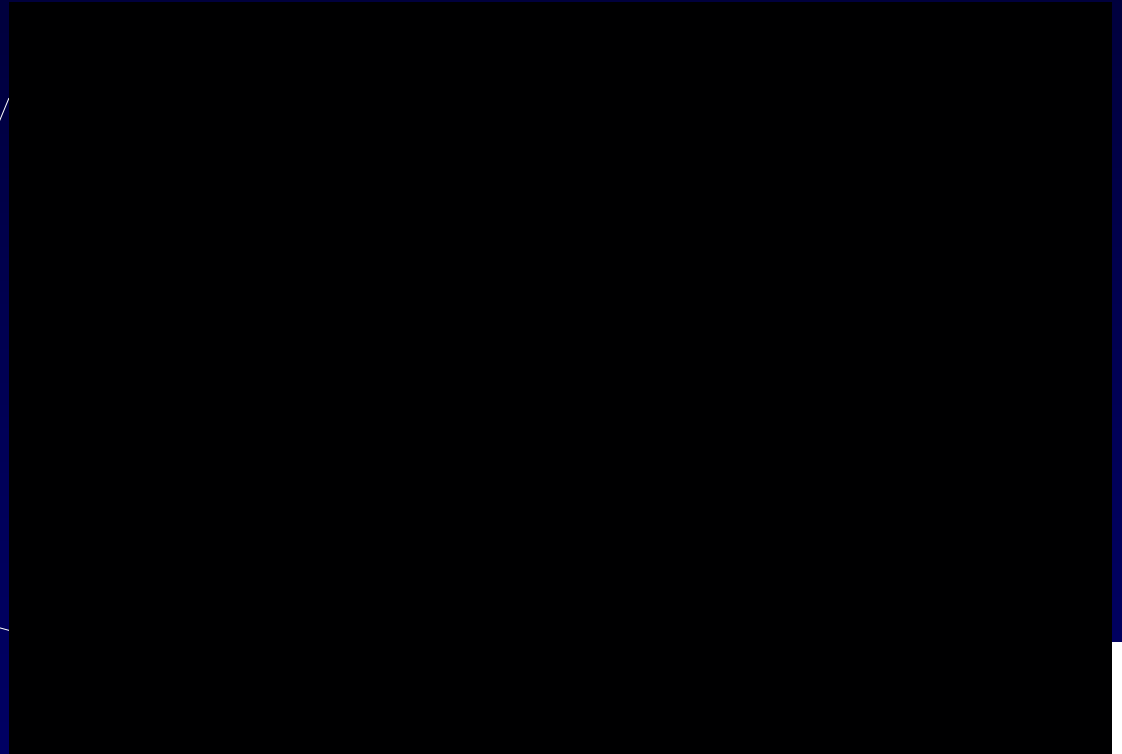
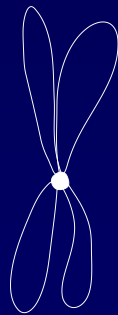
- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Predikce funkce genů *in silico*
 - struktura genů a jejich vyhledávání



Predikce funkce genů *in silico*

struktura genů

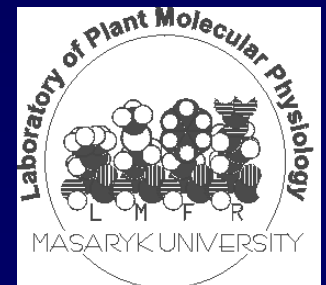
- struktura genů
 - promotor
 - počátek transkripce
 - 5 UTR
 - počátek translace
 - místa sestřihu
 - stop kodon
 - 3 UTR
 - polyadenylační signál



Predikce funkce genů *in silico*

vyhledávání genů

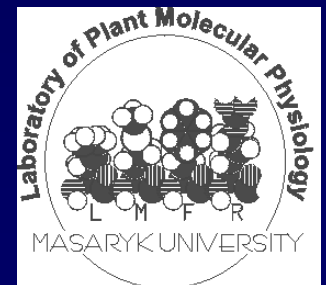
- vyhledávání genů *ab initio*
 - zanedbání 5' a 3' UTR
 - identifikace počátku translace (ATG) a stop kodonu (TAG, TAA, TGA)
 - nalezení donorových (většinou GT) a akceptorových (AG) míst sestřihu
 - většina ORF není skutečně kódujícími sekvencemi – u *Arabidopsis* je asi 350 mil. ORF na každých 900 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



Predikce funkce genů *in silico*

vyhledávání genů

- vyhledávání 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>)



Predikce funkce genů *in silico* vyhledávání genů

What do the output columns mean?

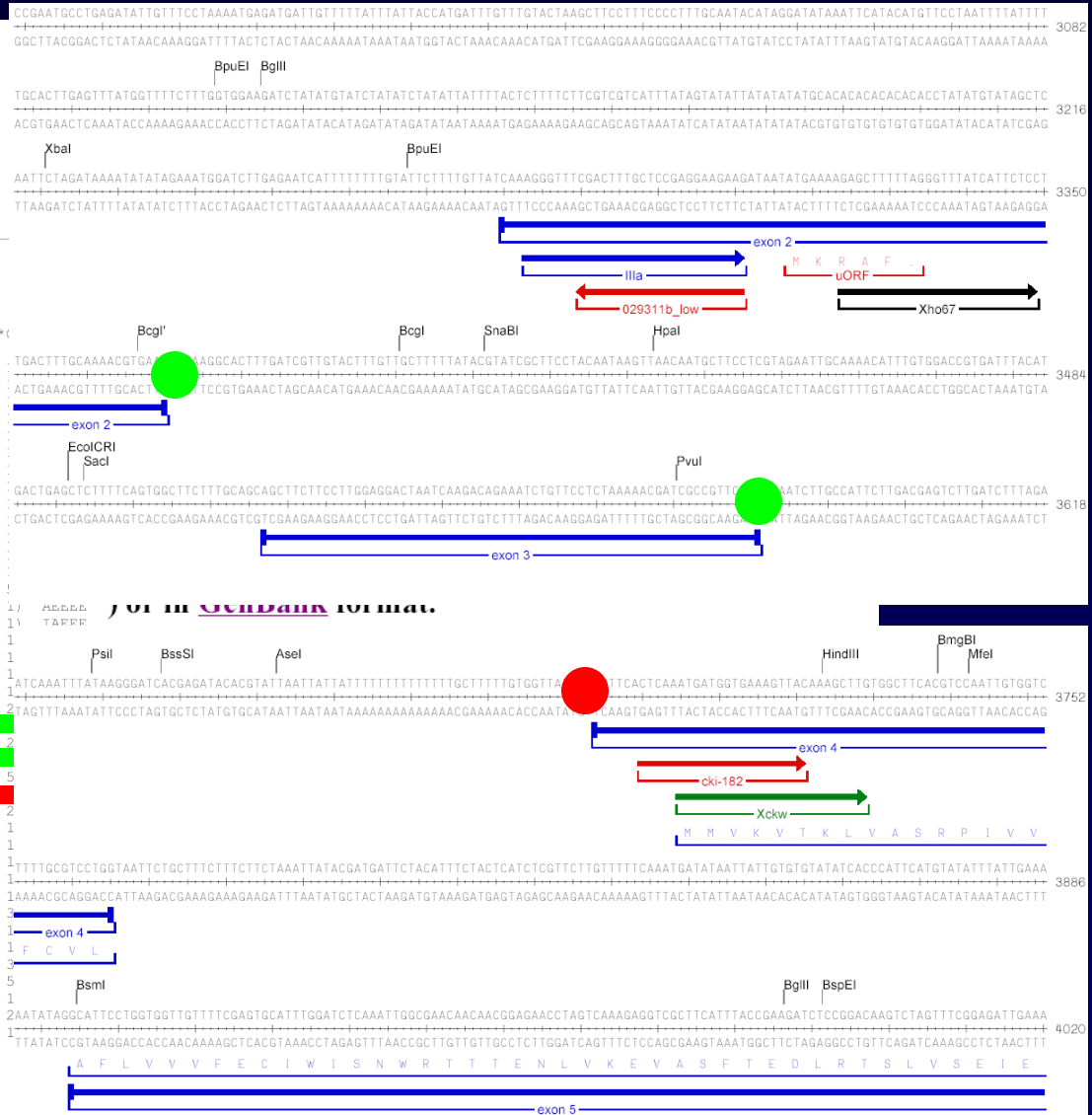
SplicePredictor. Version of February 13, 2005.
Date run: Wed Nov 9 11:30:14 2005

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

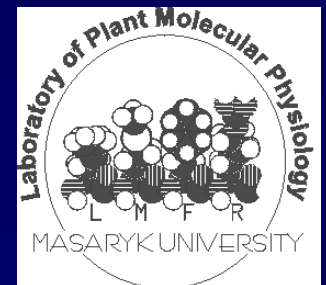
t	q	loc	sequence	P	c	rho	gamma	*	P*R*(
A	<--	75	ttttttgatctcAGat	0.973	7.16	0.000	0.000	7	(5 1
A	<--	134	attattttttcAGt	0.999	14.86	0.000	0.000	7	(5 1
A	<--	500	gattttgtgttAGtc	0.977	7.48	0.000	0.000	7	(5 1
A	<--	780	tctgttattgtatAGct	0.986	8.56	0.000	0.000	7	(5 1
A	<--	848	tatttttgaagtAGat	0.968	6.80	0.000	0.000	7	(5 1
A	<--	1051	caatttttttaAGaa	0.930	5.19	0.000	0.000	7	(5 1
A	<--	1213	ttattttttttAGtt	0.998	12.14	0.000	0.000	7	(5 1
A	<--	1373	ttctctctcacAGga	0.999	13.17	0.000	0.000	7	(5 1
A	<--	1487	ttatataattgatAGtg	0.883	4.04	0.000	0.000	7	(5 1
A	<--	1581	atgtgttcctgtAGga	0.982	8.03	0.000	0.000	7	(5 1
A	<--	1781	gggtgtgcgaagtAGgg	0.886	4.10	0.000	0.000	7	(5 1
A	<--	2440	taataaaaaattAGat	0.939	5.46	0.000	0.000	7	(5 1
A	<--	2479	catctaaaaattAGat	0.942	5.59	0.000	0.000	7	(5 1
D	---->	2546	aagTtagta	0.909	4.61	0.885	1.903	15	(5 5
A	<--	2572	ttttttttggcAGca	0.930	5.16	0.000	0.000	7	(5 1
A	<--	2763	ctcaaatccaaaAGgt	0.873	3.86	0.185	0.000	11	(5 5 1
A	<--	2782	tttcttttcattAGcg	0.952	5.98	0.220	0.000	11	(5 5 1
A	<--	3022	ttgtttgtactaAGct	0.956	6.16	0.221	0.000	11	(5 5 1
A	<--	3048	ctttgcaatcatAGga	0.973	7.15	0.229	0.000	11	(5 5 1
A	<--	3171	cgctgcattttatAGta	0.988	8.74	0.000	0.000	7	(5 1
A	<--	3284	ctttgttatcaaaAGgg	0.993	10.03	0.000	0.006	8	(5 1
A	<--	3451	aatgcttctcgtAGaa	0.916	4.77	0.293	0.065	12	(5 5 2
D	---->	3649	cacGTatta	0.933	5.25	0.000	1.848	11	(5 1 5
A	<--	4254	attattgtttcAGat	0.998	12.82	0.000	0.002	8	(5 1 2
A	<--	4351	tttcttacatgcAGaa	0.991	9.42	0.000	0.000	7	(5 1 1
A	<--	4633	gtttgtttttttAGgg	0.879	3.97	0.000	0.000	7	(5 1 1
A	<--	4976	cttgtgtttttcAGct	0.952	5.98	0.000	0.000	7	(5 1 1
A	<--	5004	ttttttttttgcaAGag	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
D	---->	5384	ttgTtaaga	0.941	5.54	0.478	0.090	13	(5 5 3
A	<--	5403	actctgtttcttAGct	0.894	4.26	0.000	0.000	7	(5 1 1
A	<--	5441	ctttctctctaacAGaa	0.995	10.43	0.387	0.000	11	(5 5 1 F C V L
A	<--	5472	ttgttaaaattacAGct	0.965	6.62	0.478	0.090	13	(5 5 3
D	---->	5745	gcgTtaaga	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
A	<--	6552	ggattttcacctcAGag	0.938	5.42	0.000	0.000	7	(5 1



Predikce funkce genů *in silico*

vyhledávání genů

- vyhledávání genů *ab initio*
 - programy pro predikci míst sestřihu (specifická 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/>)



Predikce funkce genů *in silico* vyhledávání genů

Prediction done

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

The sequence: Sequence has the following
Length: 9490 nucleotides.
31.8% A, 17.0% C, 19.6% G, 31.7% T, 0.0% N

Donor splice sites, direct strand

pos 5'→3'	phase	strand	con
1704	0	+	
1906	0	+	
4134	0	+	
4619	1	+	
4915	0	+	
5356	0	+	
5384	1	+	
5809	1	+	
6057	0	+	
6096	1	+	
7369	0	+	
7886	0	+	
9323	0	+	

Donor splice sites, complement strand

pos 3'→5' pos 5'→3' phase strand con

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	con
1213	0	+	
1221	2	+	
1373	0	+	
1487	1	+	
4254	0	+	
4832	2	+	
5004	0	+	
5472	1	+	
6135	0	+	
6490	1	+	
6744	0	+	
7447	0	+	
7780	2	+	
7786	2	+	

0.92 TCAGATAC AACACATGCA



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 Performance

SUBMISSION

Submission of a local file with a single sequence:

File in FASTA format

Browse...

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

Clear fields Send file

Submission by pasting a single sequence:

Sequence name

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

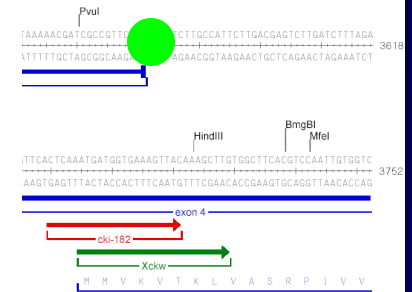
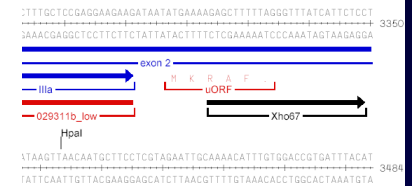
Sequence

GAGGAGGCACAAAACCGAATATACAAAATGATCTTAAACAGCTAAACTATATTGGACATTTTTCGATC
TCAGATATA
AAAGATTTTCATCAATAAATACTGGATAAATACTCTTATATATTTTCTTTAGTTTATTAATAAAAAACCT
CTAATAAAT
ACGAGTTTAAGTCCAAATCGCTTAGACTAAAATACACCATATAAATTTCAAACGATAAAGTTTACAAAA

Clear fields Send file

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

CCGAATGCCTGAGATATGTTTCCTAAATGAGATGATGTTTTATTATACCATGATTTGTTGTACTAAGCTTCCTTTCCCTTTGCAATACATAGGATAAATTCATACATGTTCTCAATTTATTTT
GGCTTAGCGAGCTATAACACAAGGATTTTACTCTACTAACARAAATAAATGAGTACTAACACAACATGATTCGAGGAAGGGGAAACGTTAGTCTATCTATATTAAAGTATGTACAGGATTAATAAAAA
BpuEI BglII
TGCACTTGAGTTATGGTTTTCTTTGGTGGAGATGATATGTATCTATCTATATATTTACTCTTTCTTGGTCGTCATTATAGTATATATATATGCACACACACACACCTATGTATAGCTC
AGCATAAATCATATAATAATATATAGTGTGTGTGTGGATATACATATCGAG

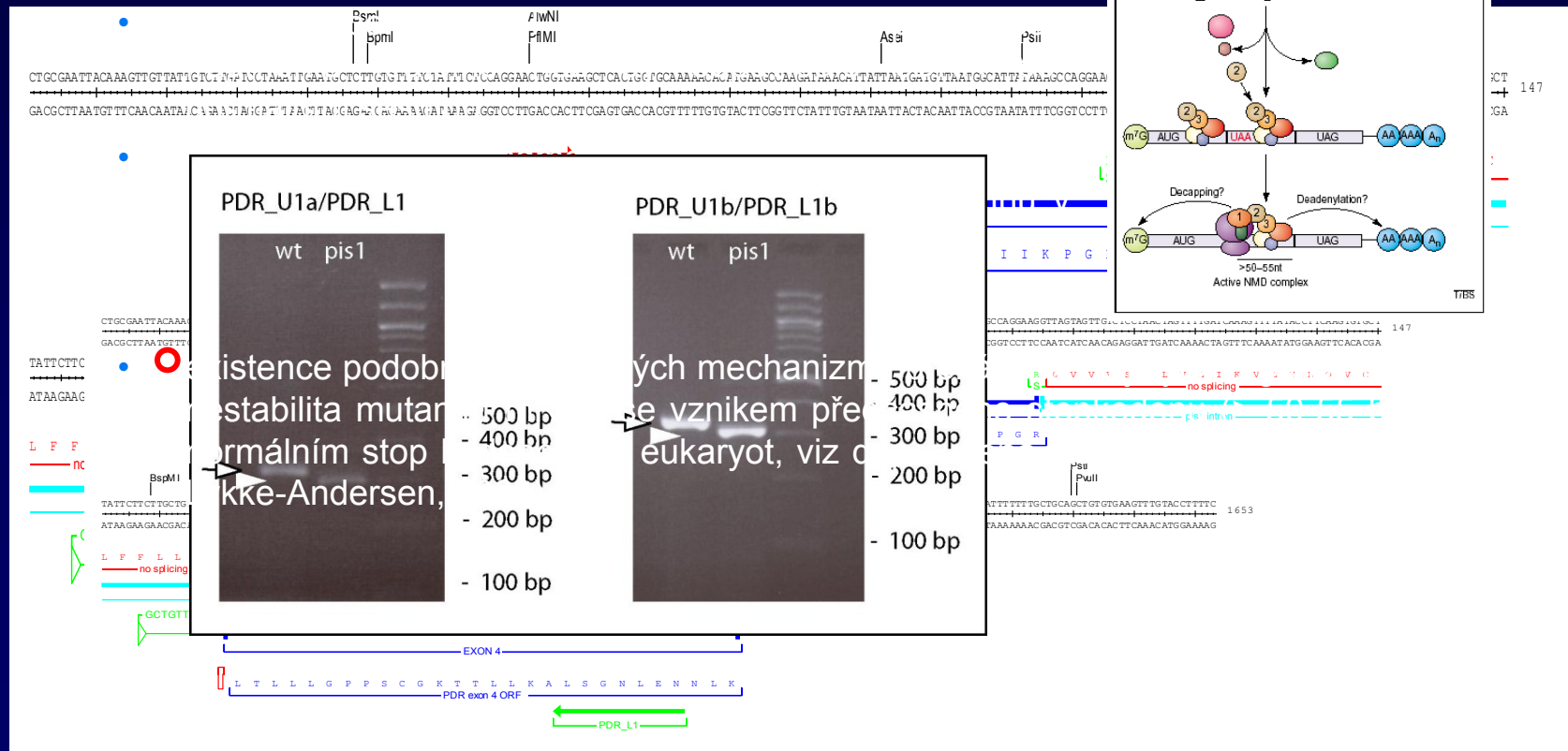


AAGAGTCCCTTCATTTACCGAAGATCTCCGACAAAGCTAGTTCCGGAGATTGAAA
TTCTCCAGCGAAGTAATGGCTTCTAGAGGCTGTTCAGATCAAGGCTCTAACCTTT
K E V A S P T E D L R I S L V S E I E



Predikce funkce genů *in silico* vyhledávání genů

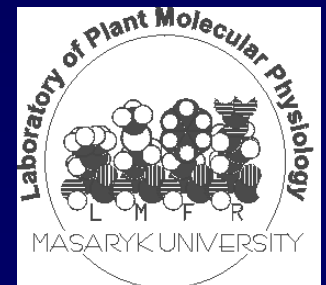
- odchyly rozpoznávání míst sestřihu u rostlin v praxi - příklad vývojové plasticity (nejen) rostlin
 - identifikace mutanta s bodovou mutací (tranzice G→A) přesně v místě sestřihu na 5' konci 4. exonu



Predikce funkce genů *in silico*

vyhledávání genů

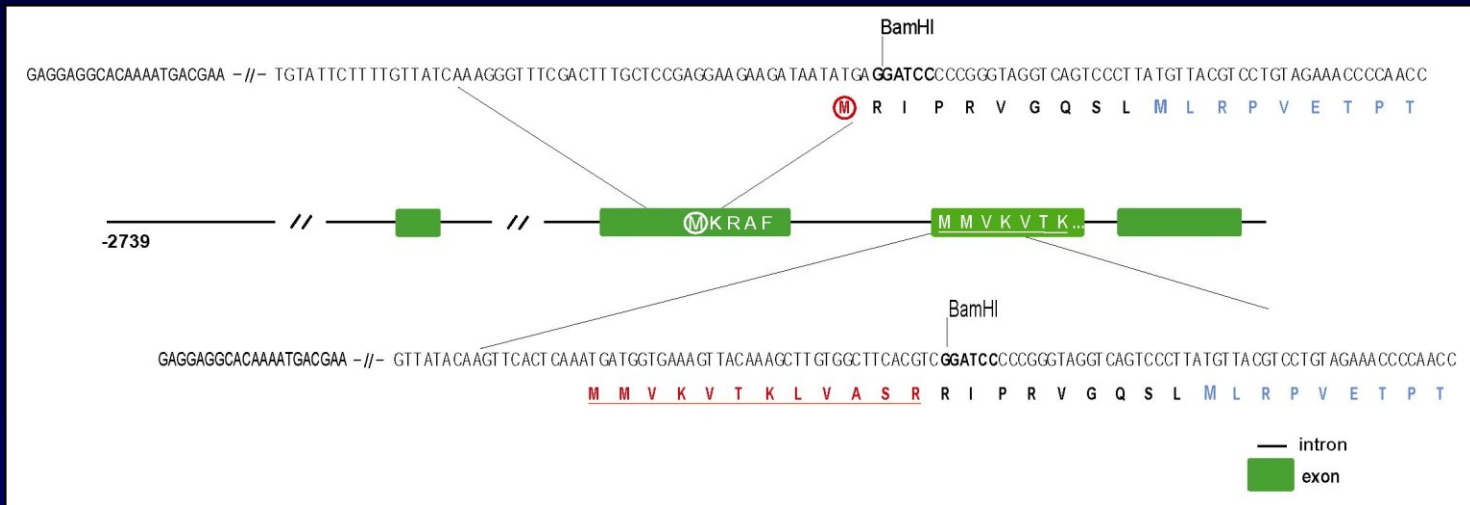
- vyhledávání genů *ab initio*
 - programy pro predikci míst sestřihu (specifita přibližně 35%)
 - GeneSplicer (http://www.tigr.org/tdb/GeneSplicer/gene_spl.html)
 - SplicePredictor (<http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi>)
 - NetGene2 (<http://www.cbs.dtu.dk/services/NetGene2/>)
 - programy pro predikci exonů
 - 4 typy exonů (podle polohy):
 - iniciační
 - vnitřní
 - terminální a
 - 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/>)



Predikce funkce genů *in silico*

vyhledávání genů

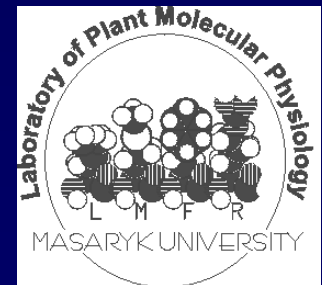
- Funkční význam sestřihu v nepřekládaných oblastech - důležitá regulační součást genů



aaagtaca...

K V T ...
aaagtaca...

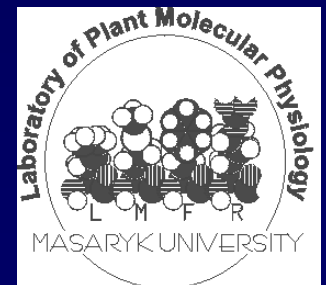
- V případě CK11 pokus prokázat tento způsob regulace genové exprese pomocí transgenních linií nesoucích *uidA* pod kontrolou dvou verzí promotoru, zatím nepotvrzeno



Predikce funkce genů *in silico*

vyhledávání genů

- vyhledávání genů *ab initio*
 - 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/>)



Predikce funkce genů *in silico* vyhledávání genů

Eukarvotic GeneMark.hmm^(1,2) [\(Reload this page\)](#)

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/genmark/euk_ghm.matrices/athaliana

Thu Oct 1 11:09:24 2009

Predicted genes/exons

Gene #	Exon #	Strand	Exon Type	Exon Range	Exon Length
1	1				
1	1				
1	1				
1	4	+	Internal	2266 - 2544	3
1	5	+	Internal	2734 - 3317	5
1	6	+	Internal	3397 - 4529	12
1	7	+	Terminal	4709 - 4921	2

[Procházet...](#)

Species: Athaliana ES-30 [Model description](#)

Output Options

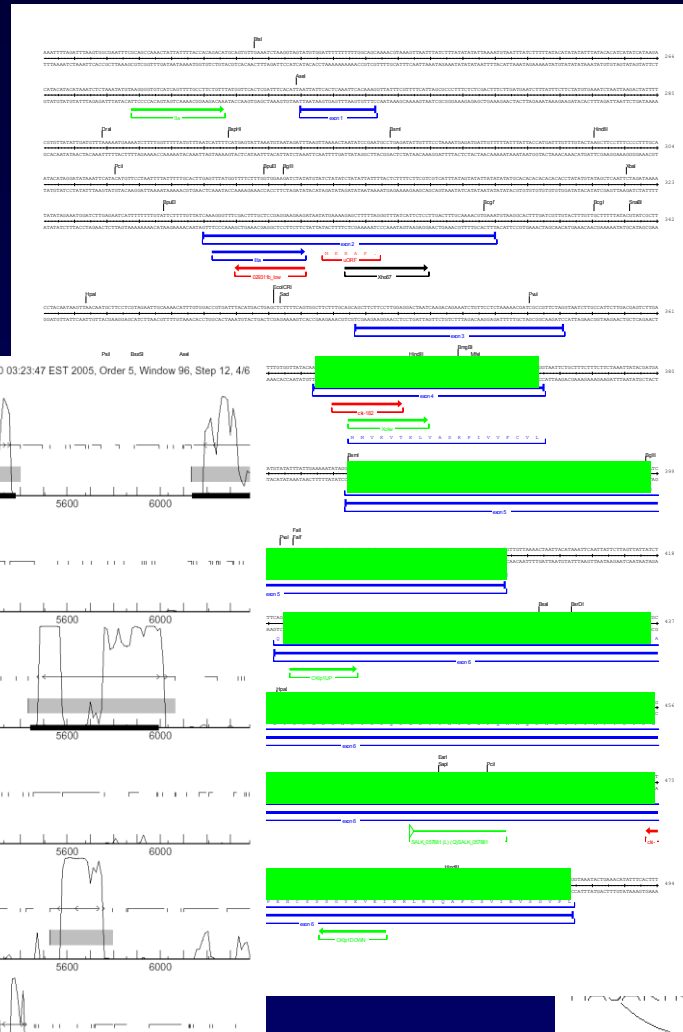
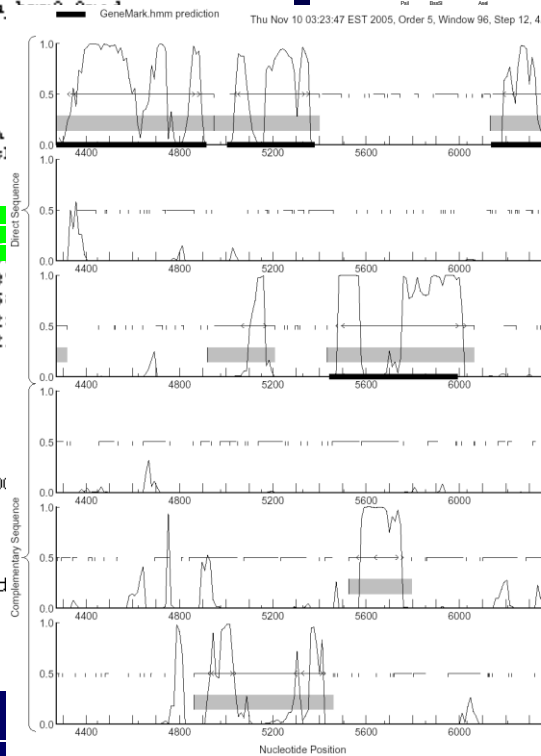
Email Address: (required for graphical output or sequences longer than 4000)

- Generate PDF graphics (screen)
- Generate PostScript graphics (email)
- Print GeneMark 2.4 predictions in addition to GeneMark.hmm pred
- Translate predicted genes into protein

Run

[Default](#)

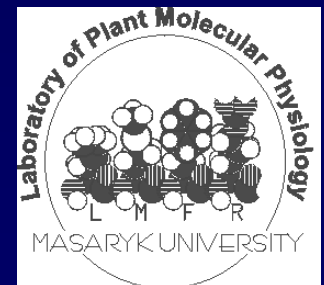
[Start GeneMark.hmm](#)



Predikce funkce genů *in silico*

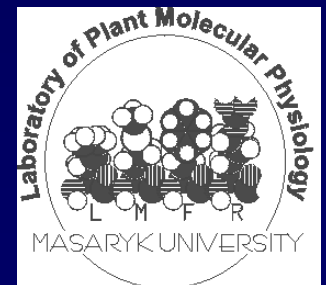
vyhledávání genů

- 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/lbn11690.html>)



Základy genomiky II.

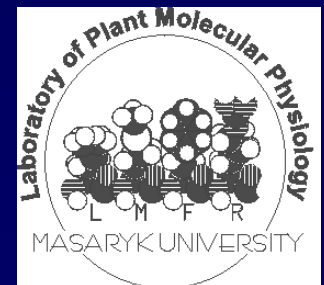
- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Predikce funkce genů *in silico*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie



Predikce funkce genů *in silico*

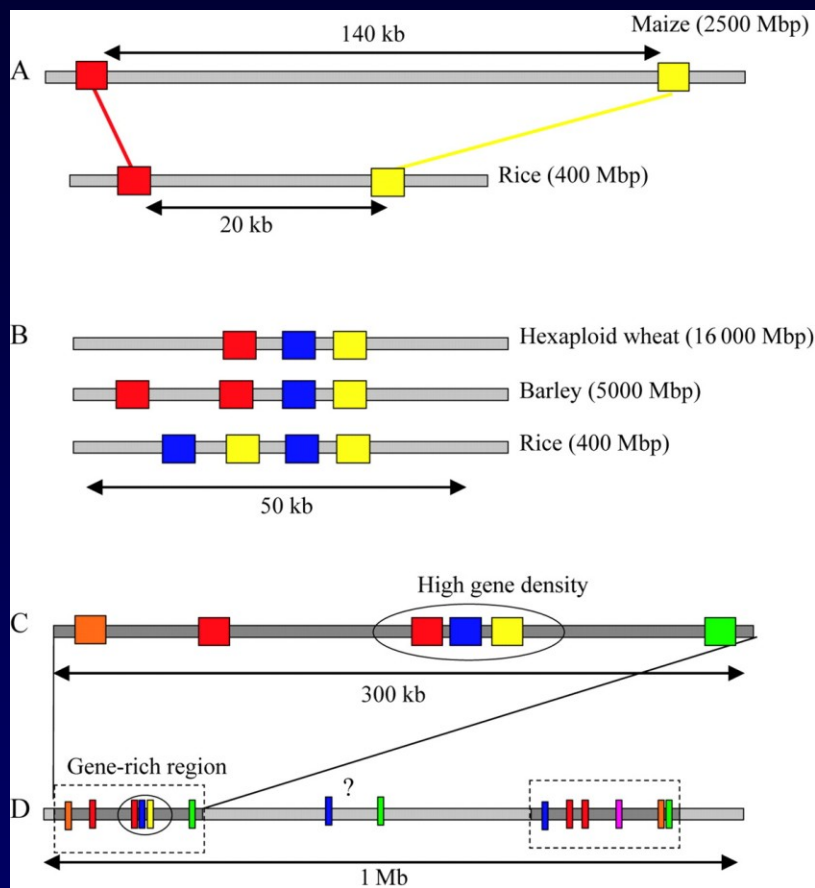
vyhledávání genů

- genomová kolinearita a genová homologie
 - 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, 46-55 tis. genů) 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)

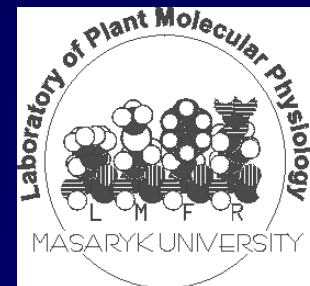


Predikce funkce genů *in silico*

vyhledávání genů-genomová kolinearita



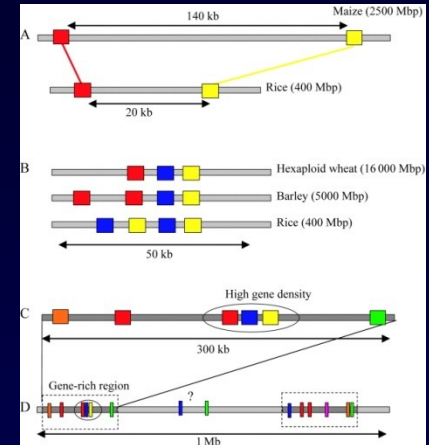
Feuillet and Keller, 2002



Predikce funkce genů *in silico* vyhledávání genů

- genomová kolinearita a genová homologie

- 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 kolinearit (také „komparativní genomika“) při experimentální identifikaci genů příbuzných organismů:

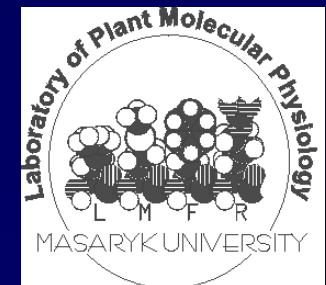


- mapování malých genomů s využitím nízkokopiových DNA markerů (např. RFLP)

- využití těchto markerů k identifikaci orthologních genů (genů se stejnou nebo podobnou funkcí) příbuzného organismu

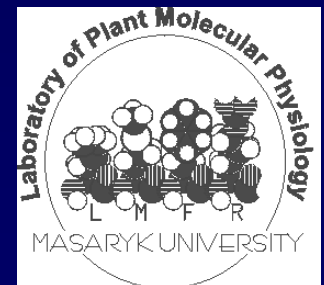
- malý genom (např. rýže, 466 Mbp, 46-55 tis. genů) může sloužit jako vodítka, kdy jsou identifikovány molekulární nízkokopiové markery (např. RFLP) ve vazbě s genem zájmu a tyto oblasti jsou pak použity jako sonda při vyhledávání v BAC knihovnách při identifikaci orthologních oblastí velkých genomů (např. ječmene nebo pšenice, 5000, resp. 16000 Mbp)

- 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.)



Základy genomiky II.

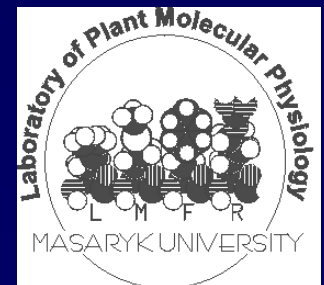
- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Predikce funkce genů *in silico*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven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



Predikce funkce genů *in silico*

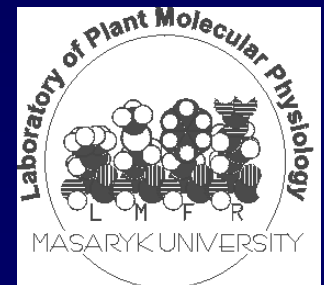
vyhledávání genů

- 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
 - 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 %



Základy genomiky II.

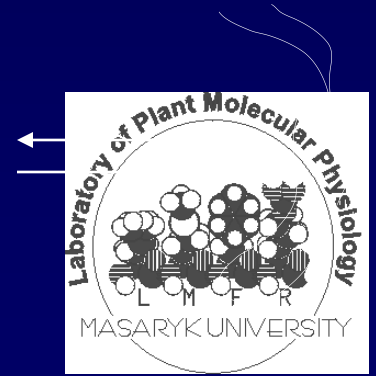
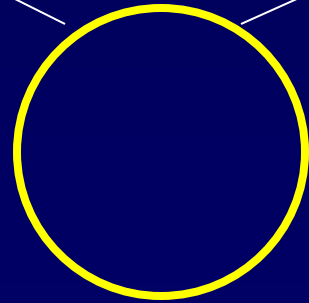
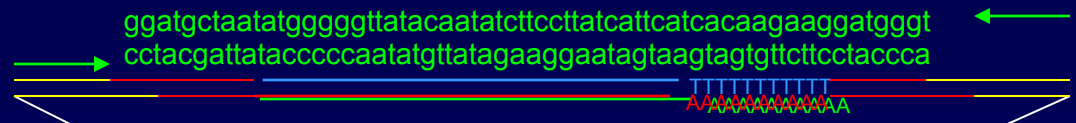
- Postupy „přímé“ a reverzní genetiky
 - rozdíly v myšlenkových přístupech k identifikaci genů a jejich funkcí
- Predikce funkce genů *in silico*
 - struktura genů a jejich vyhledávání
 - genomová kolinearita a genová homologie
- Experimentální identifikace genů
 - příprava genově obohacených knihoven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



Predikce funkce genů *in silico* vyhledávání genů-EST knihovny

- příprava EST knihoven

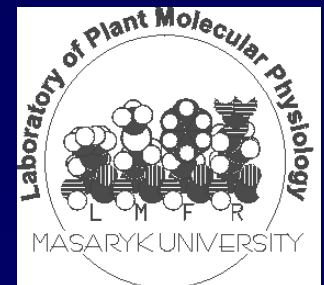
- izolace mRNA
- RT PCR
- ligace linkerů a syntéza druhého řetězce cDNA
- klonování do vhodného bakteriálního vektoru
- transformace do bakterií a izolace DNA (amplifikace DNA)
- sekvenace s použitím primerů specifických pro použitý plasmid
- uložení výsledků sekvenace do veřejné databáze



Základy genomiky II.

shrnutí

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



Základy genomiky II.

diskuse

