

# PB051 Výpočetní metody v bioinformatice a systémové biologii

## Týden 1

Katedra informačních technologií  
Masarykova Univerzita Brno

Jaro 2016

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

## Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

- ▶ Dr. Matej Lexa, C506 (lexa@fi.muni.cz)
- ▶ Kurz: Út 8:00-9:50 (A219)
- ▶ Konzultace: Út 13:00-15:00 (C506)
- ▶ <http://www.fi.muni.cz/~lexa/teaching.html>

## Informace o kurzu

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ Dr. David Šafránek, A408 ([xsafran1@fi.muni.cz](mailto:xsafran1@fi.muni.cz))

## Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

- ▶ Hodnotí se
  - ▶ Úkoly 4 x 5 bodů
  - ▶ Semestrální úkol 30 bodů
  - ▶ Zkouška 50 bodů
- ▶ Klasifikační stupnice
  - ▶ A 91 - 100
  - ▶ B 81 - 90
  - ▶ C 71 - 80
  - ▶ D 61 - 70
  - ▶ E 51 - 60
  - ▶ F méně než 51

## Informace o kurzu

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ Genomové anotace
- ▶ Výpočty nad sekvencemi, konsenzus, repetice, mapování a skládání sekvencí
- ▶ Aplikace Markovovských modelů v bioinformatice
- ▶ Statická analýza sítí
- ▶ Dynamická analýza sítí

## Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

- ▶ Prohlížeče genomů (UCSC, Ensembl, GBrowse, IGV)
- ▶ Pokročilé funkce UCSC Genome Browser (Table Browser)
- ▶ Programovatelný přístup ke genomu přes rozhraní Biomart, GenomicRanges (R/Bioconductor)
- ▶ Bioinformatika genové regulace (JASPAR, TFBSTools)

# Výpočty nad sekvencemi, konsenzus, repetice, mapování

## Informace o kurzu

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ Mnohočetné zarovnání a konsenzuální sekvence
- ▶ Diagram "dot plot" a tandemové repetice
- ▶ Mapování sekvencí (BLAST, Bowtie-2)



# Aplikace Markovovských modelů v bioinformatice

## Informace o kurzu

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ Markovovy řetězce
- ▶ Markovovy řetězce proměnného řádu
- ▶ Skryté Markovovské modely (HMM)
- ▶ HMM profily
- ▶ HMM pro identifikaci genů

## Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

- ▶ metody a nástroje statické analýzy a integrace dat
  - ▶ integrace dat
  - ▶ rekonstrukce sítě genových interakcí z experimentálních dat
  - ▶ analýza interakční sítě jako obecného grafu

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

- ▶ statická analýza sítí a integrace dat
  - ▶ nástroje: Cytoscape s několika moduly
- ▶ rekonstrukce genových regulačních sítí z microarray dat
  - ▶ nástroje: GeneNetworks, GinSim
- ▶ dynamická analýza pravděpodobnostních modelů genových sítí
  - ▶ nástroje: Dizzy
- ▶ metabolické sítě a jejich analýza
  - ▶ nástroje: KEGG, metacyc, COPASI

Informace o kurzu

**Struktura genomu**

Dynamika genomu

Genomové data prohlížeče

Informace o kurzu

**Struktura genomu**

Dynamika genomu

Genomové data prohlížeče



Informace o kurzu

**Struktura genomu**

Dynamika genomu

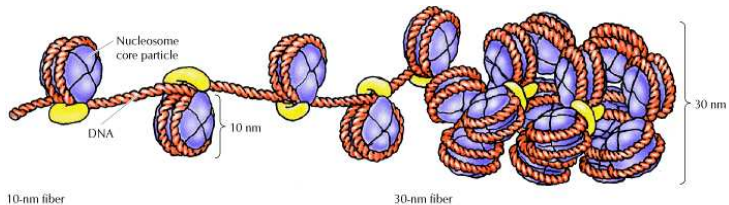
Genomové data prohlížeče

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

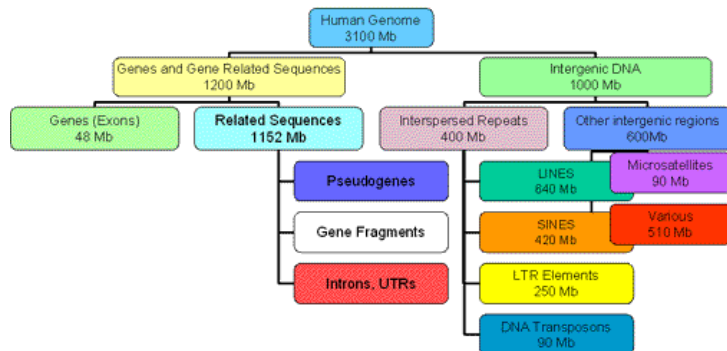


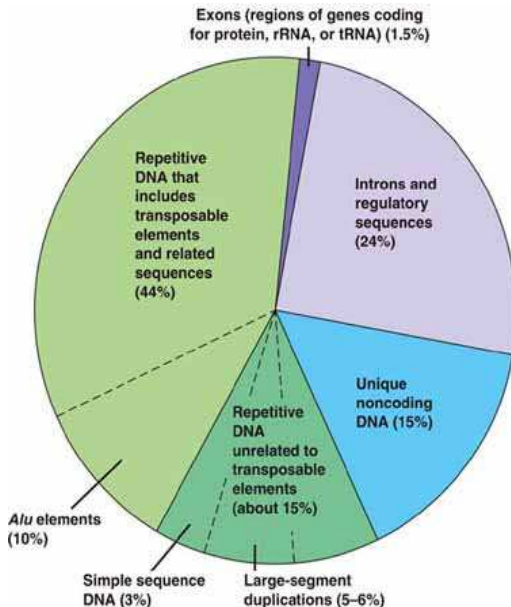
Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče





Informace o kurzu

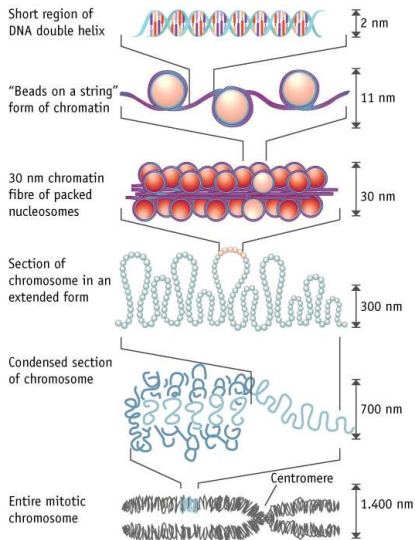
Struktura genomu

Dynamika genomu

Genomové data prohlížeče



# Genome compacting



images/chromosomes

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

# Eukaryotic transcription

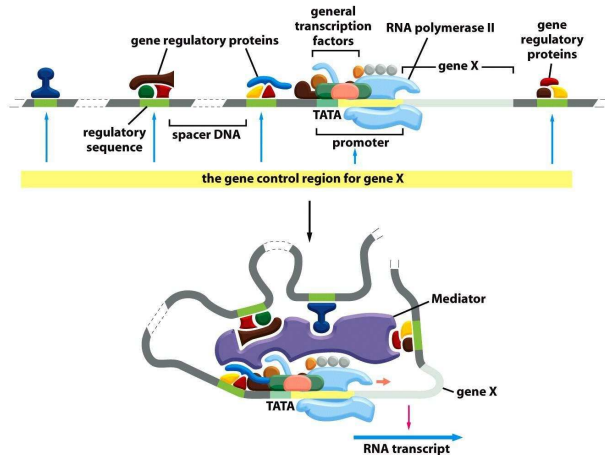


Figure 7-44 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

# Selected organizational characteristics of genomes

- ▶ by topology
  - ▶ modular
  - ▶ hierarchical
  - ▶ discrete, but elements sometimes fuzzy
- ▶ by end-product
  - ▶ DNA (structure)
  - ▶ RNA (tRNA, ncRNA, rRNA, siRNA...)
  - ▶ protein
  - ▶ complexes and particles (membrane sensor, flagellum, viral particle)

- ▶ Geny
  - ▶ proteiny (kódující, exon, intron)
  - ▶ RNA
- ▶ Regulační sekvence
  - ▶ promotory
  - ▶ enhancery
  - ▶ jiné
- ▶ Repetitivní sekvence
  - ▶ mikrosatelity (STR)
  - ▶ minisatelity (VNTR)
  - ▶ satelity
    - ▶ DNA transpozony, helitrony
    - ▶ retrotranspozony (LINE, SINE, LTR)
- ▶ Cizí sekvence
  - ▶ viry
  - ▶ endo(retro)viry
- ▶ Oblasti (ne)podobnosti (homology)
  - ▶ SNP
  - ▶ delší strukturní variace
  - ▶ Genomické ostrovy

[Informace o kurzu](#)

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

Informace o kurzu

Struktura genomu

**Dynamika genomu**

Genomové data prohlížeče

Informace o kurzu

Struktura genomu

**Dynamika genomu**

Genomové data prohlížeče

# Genome changes (in sequence or number)

- ▶ topology unchanged
  - ▶ SNPs (point mutations)
  - ▶ tandem repeat expansion/contraction
- ▶ 1-point translocation
  - ▶ chromosome breakage
  - ▶ chromosome fusion
- ▶ 2-point translocation
  - ▶ deletions, conversions and exchanges (recombination effects)
  - ▶ cut-paste (DNA transposon)
  - ▶ copy-paste (retrotransposons)
  - ▶ rolling-circle (helitrons)
- ▶ 2-point translocation
  - ▶ DNA methylation
  - ▶ Histone methylation/acetylation

[Informace o kurzu](#)

[Struktura genomu](#)

**[Dynamika genomu](#)**

[Genomové data prohlížeče](#)

- ▶ at DNA level
  - ▶ exonization
  - ▶ exon shuffling
  - ▶ gene migration
  - ▶ genome duplication
- ▶ at RNA level
  - ▶ alternative splicing
  - ▶ transcriptional fusion
- ▶ at protein level
  - ▶ translational fusion

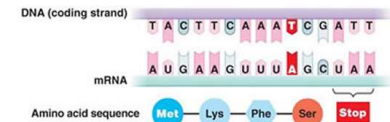
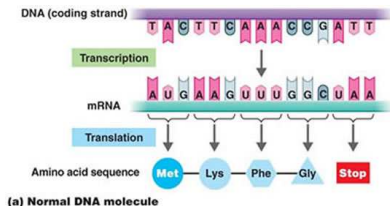
# Mutations

Informace o kurzu

Struktura genomu

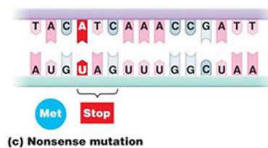
Dynamika genomu

Genomové data prohlížeče

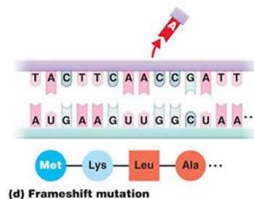


**(b) Missense mutation**

Copyright © 2010 Pearson Education, Inc.



**(c) Nonsense mutation**



**(d) Frameshift mutation**



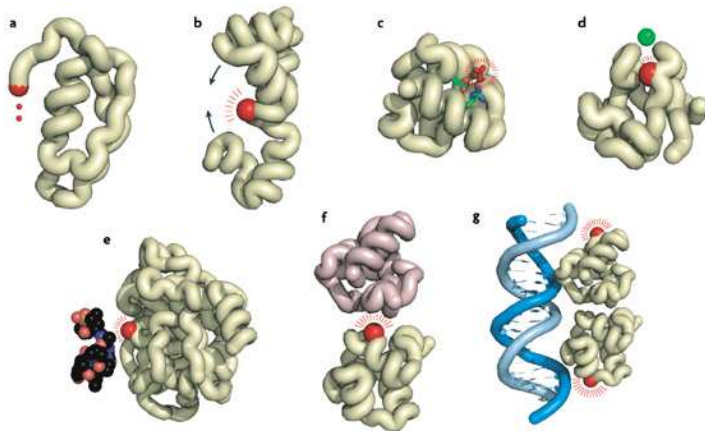
# Mutations

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



# Tandem repeat expansion or contraction

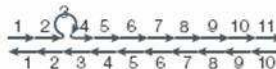
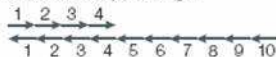
Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

Increase in repeat length



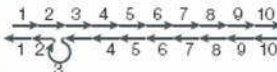
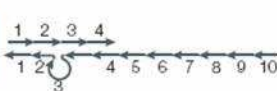
Initiation

Dissociation

Rehybridization  
and  
misalignment

The new strand  
is a different  
length to the  
template

Decrease in repeat length



# Chromosome breakage and repair

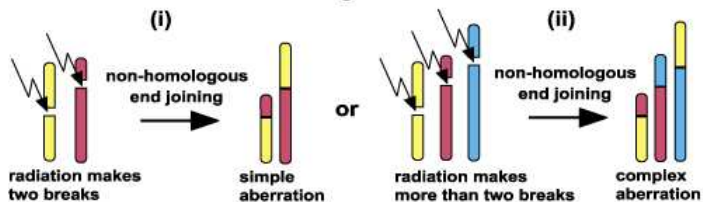
Informace o kurzu

Struktura genomu

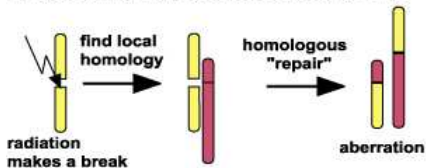
Dynamika genomu

Genomové data prohlížeče

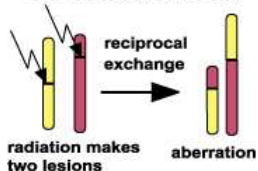
## A. Breakage-and-reunion



## B. Recombinational misrepair (1-hit)



## C. Exchange theory



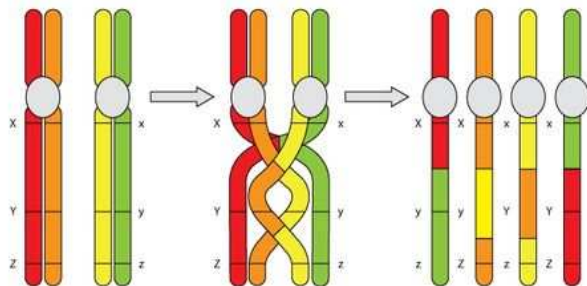
# Chiasma and crossing-over during meiosis

Informace o kurzu

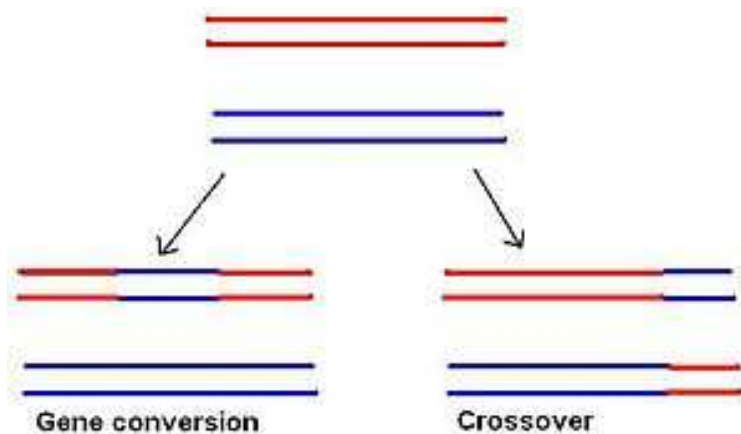
Struktura genomu

Dynamika genomu

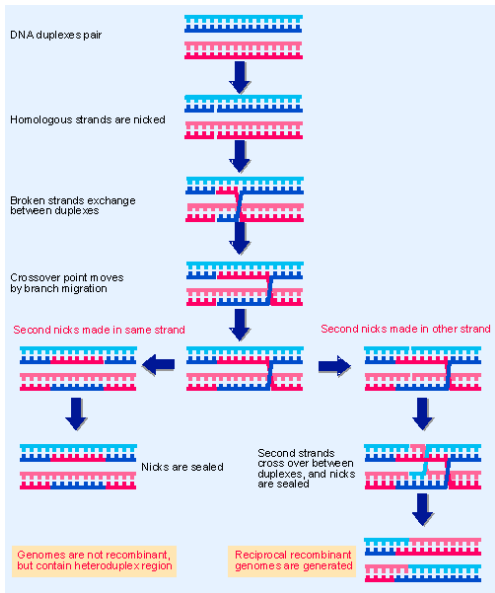
Genomové data prohlížeče



# Gene cross-over or conversion



# Gene cross-over or conversion



Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

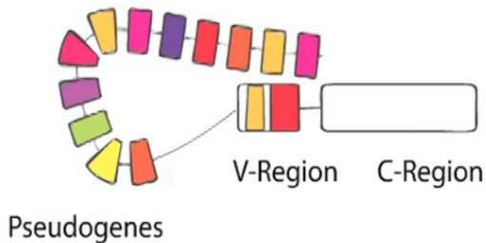
# Gene conversion in immunoglobulin

[Informace o kurzu](#)

[Struktura genomu](#)

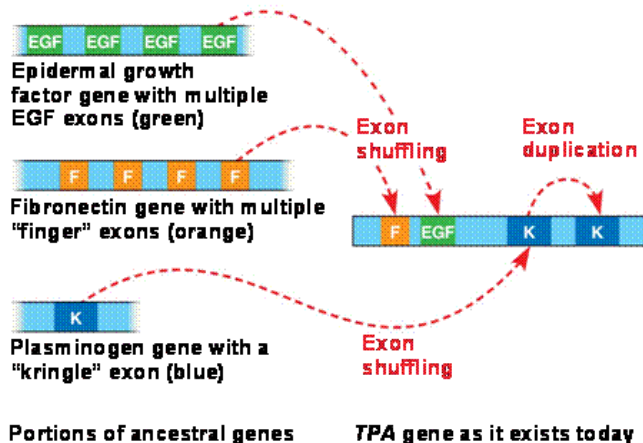
[Dynamika genomu](#)

[Genomové data prohlížeče](#)



# Exon shuffling

Fig. 21-14



Copyright © 2008 Pearson Education, Inc., publishing as Benjamin Cummings

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



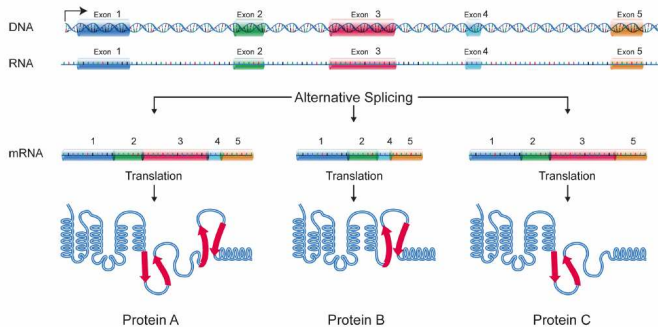
# Alternative splicing

Informace o kurzu

Struktura genomu

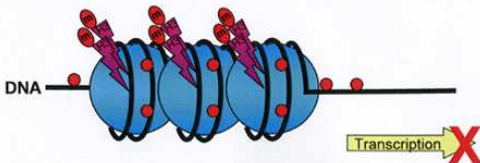
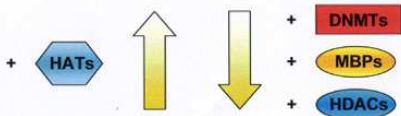
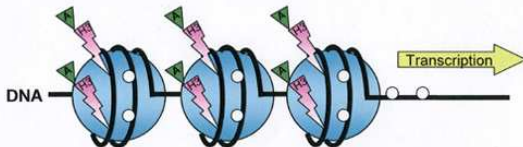
Dynamika genomu

Genomové data prohlížeče



# Epigenetic regulation of gene activity

## A. Transcriptionally active chromatin



## B. Transcriptionally inactive chromatin

# RNA interference

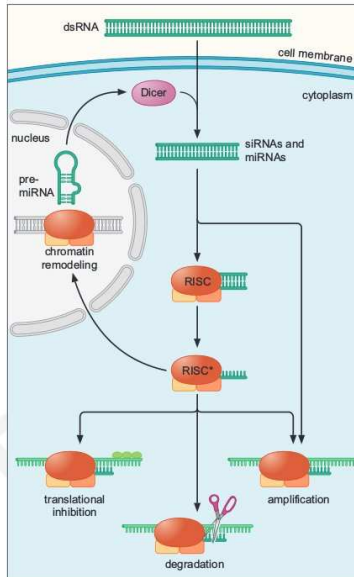
Informace o kurzu

Struktura genomu

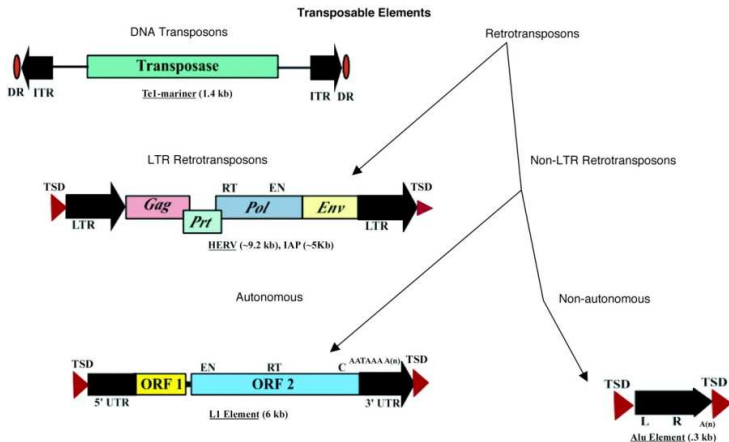
Dynamika genomu

Genomové data prohlížeče

**FIGURE 18-6 RNAi silencing.** RNAi switches off the expression of a gene when dsRNA molecules that have homology to that gene are introduced, or made, in the cell. This effect involves processing of the dsRNA to make siRNAs and miRNAs by the enzyme Dicer. Another enzyme involved only in the case of miRNAs—Drosha—is not shown here, but is described later. The siRNAs and miRNAs direct a complex called RISC (RNA-induced silencing complex) to repress genes in three ways. It attacks and digests mRNA that has homology with the siRNA; it interferes with translation of those mRNAs; or it directs chromatin-modifying enzymes to the promoters that direct expression of those mRNAs. (Adapted, with permission, from Hannon G.J. 2002. *Nature* 418: 244–251, Fig. 5. © Macmillan.)



# Transposons



Informace o kurzu

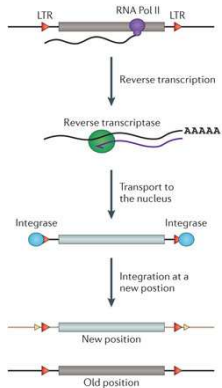
Struktura genomu

Dynamika genomu

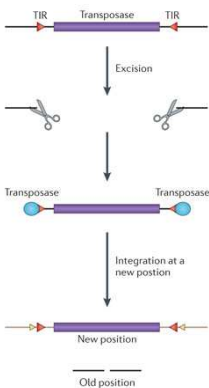
Genomové data prohlížeče

# Transposons

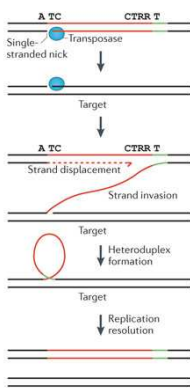
Class I element



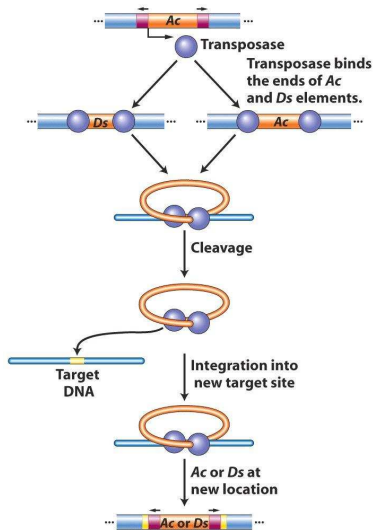
Class II element



Helitron



# Transposase is the key protein in DNA transposons



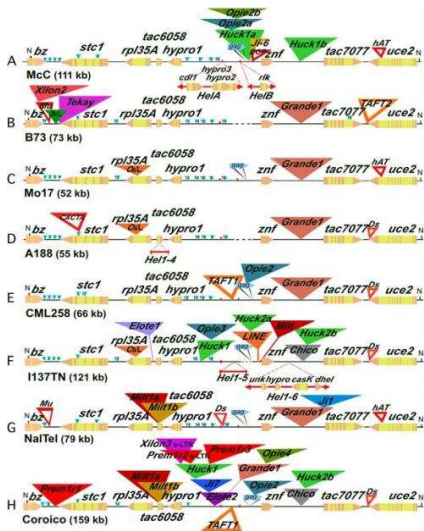
# Transposons shape genomes of maize varieties

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



# Transposons shape phenotypes of maize kernels

Informace o kurzu

Struktura genomu

**Dynamika genomu**

Genomové data prohlížeče





# Transposons shape phenotypes of grape varieties

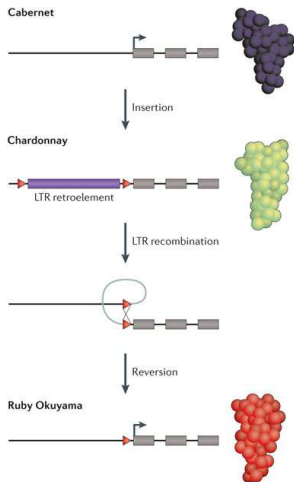
PB051 Výpočetní metody v  
bioinformatice a systémové  
biologii - Týden 1

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



[Informace o kurzu](#)

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ organizace (konfokální a elektronová mikroskopie)
- ▶ sekvenace
- ▶ mapování metylace, nukleozomů
- ▶ měření transkripce (RNA-Seq, DNA čipy)
- ▶ identifikace regulačních sekvencí (Chip-Seq)
- ▶ funkce genů - podrobný výzkum

[Informace o kurzu](#)

[Struktura genomu](#)

[Dynamika genomu](#)

[Genomové data prohlížeče](#)

- ▶ predikce genů (např. GeneMark)
- ▶ homologie (zjišťování podobnosti sekvencí) (BLAT, MUMMER, BLAST)
- ▶ identifikace opakování (např. RepeatMasker, LTR Finder)

Informace o kurzu

Struktura genomu

Dynamika genomu

**Genomové data prohlížeče**

Informace o kurzu

Struktura genomu

Dynamika genomu

**Genomové data prohlížeče**

# UCSC Genome Browser

PB051 Výpočetní metody v  
bioinformatice a systémové  
biologii - Týden 1

The screenshot displays the UCSC Genome Browser interface for Human chromosome 5 (chr5:70,256,524-70,284,592). The browser window title is "Human chr5:70,256,524-70,284,592 - UCSC Genome Browser v134 - Konqueror". The address bar shows the URL "id=73350821&knownGene=full". The main navigation bar includes links for Home, Genomes, Blat, Tables, Gene Sorter, PCR, DNA, Convert, PDF/PS, and Help. The title of the page is "UCSC Genome Browser on Human Mar. 2006 Assembly".

The interface features a navigation area with "move" buttons (left and right arrows) and "zoom in" (1.5x, 3x, 10x, base) and "zoom out" (1.5x, 3x) buttons. A "position/search" field contains "chr5:70,256,524-70,284,592" with "jump", "clear", "size 28,069 bp.", and "configure" buttons. A chromosome map shows the location of the region on chromosome 5 (q13.2).

The main data track displays various genomic features:

- STS Markers:** STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps.
- UCSC Known Genes:** UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA.
- RefSeq Genes:** RefSeq Genes.
- Human mRNAs:** Human mRNAs from GenBank.
- Spliced ESTs:** Human ESTs That Have Been Spliced.
- Conservation:** Vertebrate Multiz Alignment & Conservation (17 Species) showing conservation across mouse, rat, rabbit, dog, armadillo, elephant, opossum, and chicken.

The URL at the bottom of the browser window is <http://genome.ucsc.edu/cgi-bin/hgc?hgsid=733...523&r=70284592&db=hg18&pix=620>.

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

# Ensembl Genome Browser

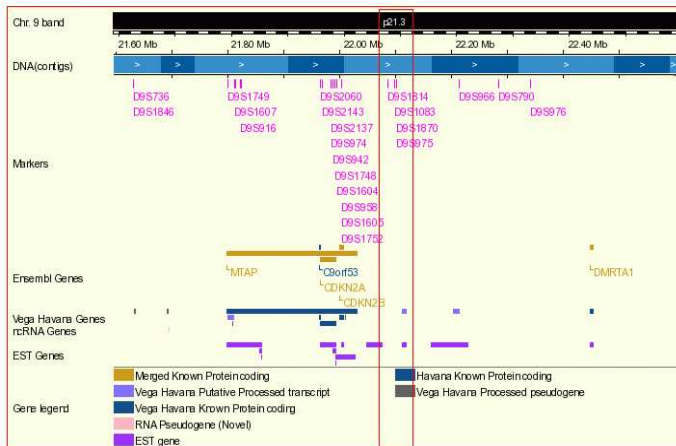
PB051 Výpočetní metody v  
bioinformatice a systémové  
biologii - Týden 1

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

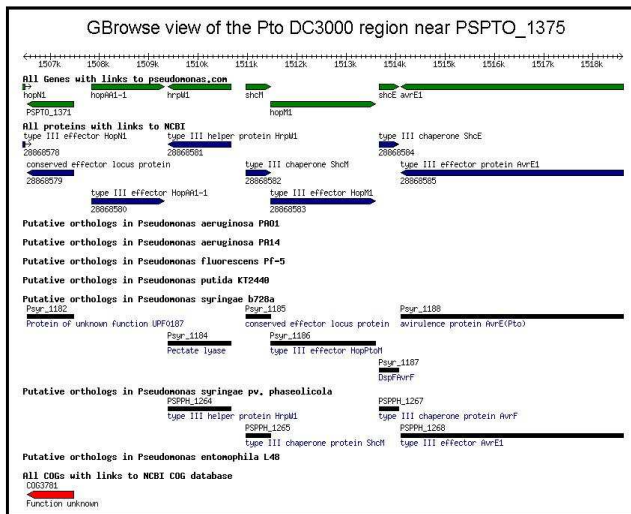


Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

The screenshot displays the Argo genome browser interface. At the top, there is a menu bar with options: Argoc, File, Track, Edit, Select, View, Zoom, Rulers, Analyze, User, Bookmarks, Window, Help. Below the menu bar, the main window title is 'Argo'. The main track area shows a genomic map for 'Human Chr15 contig 1.1 (1 - 1399746:1-100000)'. The track contains various annotations, including gene models, exons, and introns. A 'PolyA Signal View: Novel Transcrip...' popup is visible, showing the following data:

AATACA	(1.2%)	415-420	82808-82813
AATACA	(1.2%)	470-475	86534-86539
ATYAAA	(14.8%)	804-808	89548-89553

Below the main track area, there is an 'Inspector' panel with tabs for 'Properties', 'DNA', 'mRNA', and 'Protein'. The 'DNA' tab is selected, showing a sequence view with a highlighted region. To the right of the Inspector is a 'Finder' panel with the following settings:

Select Features whose:

- Label: contains repeats
- Protein length: > 50
- mRNA Sequence: contains gataca

A 'Search' button is located at the bottom of the Finder panel.



# DecodeMe Browser

PB051 Výpočetní metody v  
bioinformatice a systémové  
biologii - Týden 1

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



# Golden Helix Genome Browser

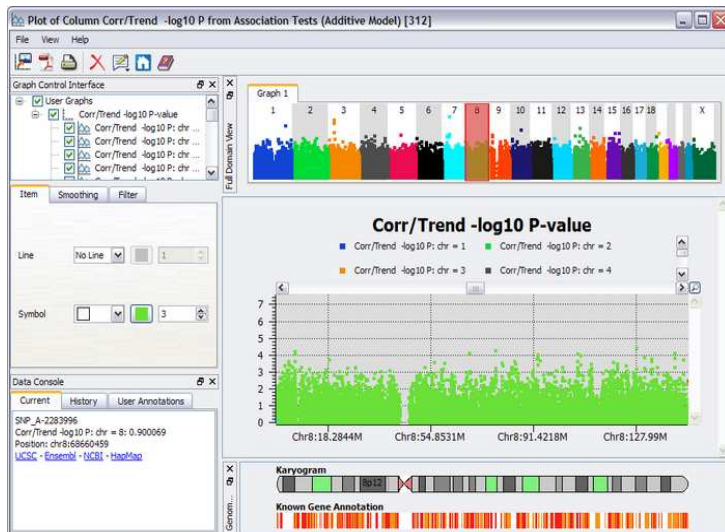
PB051 Výpočetní metody v  
bioinformatice a systémové  
biologii - Týden 1

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

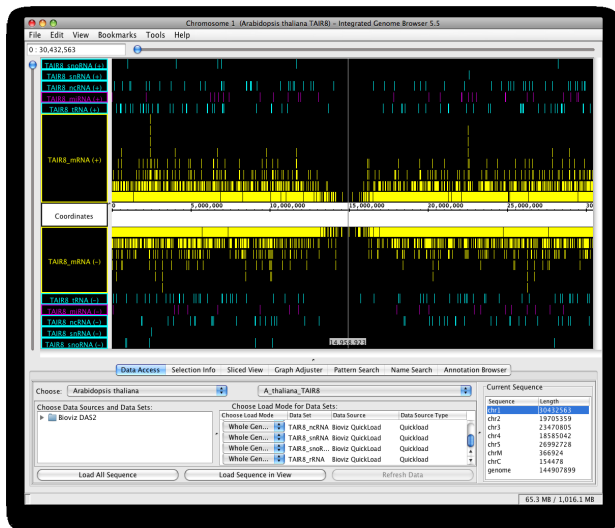


Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

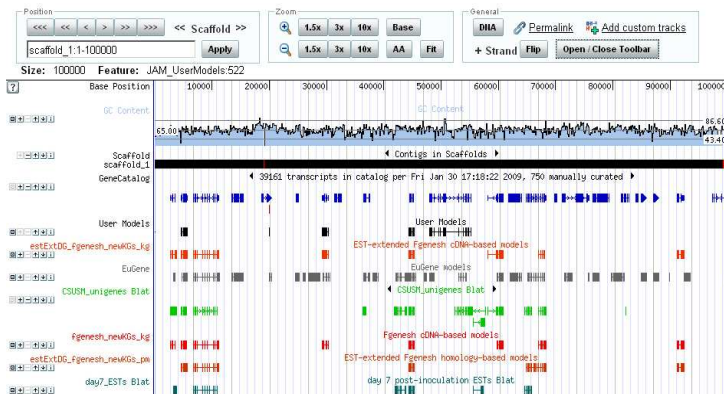


Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče



# RIKEN Genome Browser

PB051 Výpočetní metody v  
bioinformatické a systémové  
biologii - Týden 1

The screenshot displays the RIKEN Genome Browser interface for the *Mus musculus* gene *Irf1*. The interface is divided into several numbered sections:

- 1**: Database list on the left, showing various annotations like Mouse mRBase version, Mouse FANTOM3, Mouse Fantom3 Mega Genes, Mouse CAGE, Mouse Mutant Resources, Mouse Ensembl Transcripts, Mouse Ensembl Vega Genes, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, Mouse UniProtKB SwissProt, Mouse UniProtKB TrEMBL, Mouse Riken Transcription, Mouse dbSNP NCBI/Ensembl, Mouse Ensembl Gene 40, Mouse Ensembl Vega Gene, Mouse Entrez Gene NCBI, Mouse MGI Gene, and Mouse RefSeq Protein.
- 2**: Expert's set / User's set on the left, showing Genome View (Mouse, Human Homology, C elegans Homology, Other Homology, All, dbSNP) and Medline (PosMed (Positional Medline)), Transcriptome (FANTOM, CAGE).
- 3**: Search and navigation controls at the top left, including "Go to Search page", "Register current interval", and "Filter by keyword".
- 4**: Gene structure diagram at the top right, showing the *Irf1* gene structure with exons and introns, and a scale bar from 82,100,466 bp to 82,182,103 bp.
- 5**: Gene information and coordinates at the top right, including the gene symbol *Irf1*, marker symbols, and coordinates: 4700 2 | RefSeq\_dnaNM\_010570.2 | UniProt / SPTREMBL\_Q543V3\_MOUSE | Entrez.
- 6**: Genomic tracks on the right, including Mouse FANTOM3 (cDNA and cDNA-), Mouse CAGE (all tissues) Expression (TAG=249, TPM=6.62), Mouse Ensembl Transcripts (43-36a), Mouse Ensembl Vega Gene Transcripts (43-36a), Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, and Mouse UniProtKB SwissProt.

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

Informace o kurzu

Struktura genomu

Dynamika genomu

Genomové data prohlížeče

