

PB051 Výpočetní metody v bioinformatice a systémové biologii

Týden 1

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2012

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



Informace o kurzu

Genomové data prohlížeče

Informace o kurzu

Genomové data prohlížeče

- ▶ Dr. Matej Lexa, C506 (lexa@fi.muni.cz)
- ▶ Kurz: Pá 8:00-9:50 (B116)
- ▶ Konzultace: Čt 13:00-15:00 (C506)
- ▶ <http://www.fi.muni.cz/~lexa/teaching.html>

- ▶ Hodnotí se
 - ▶ Úkoly 4 x 5 bodů
 - ▶ Cvičení 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

- ▶ Analýza dat v genomovém kontextu
- ▶ Aplikace Markovovských modelů v bioinformatice
- ▶ Statická analýza sítí
- ▶ Dynamická analýza sítí

- ▶ Prohlížeče genomů (UCSC, Ensembl, Argo, GBrowse)
- ▶ Pokročilé funkce UCSC Genome Browser
- ▶ Programovatelný přístup ke genomu přes rozhraní Biomart
- ▶ Bioinformatika genové regulace (TRANSFAC, MatInspector))

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

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

- ▶ integrace dat
 - ▶ nástroje: BioMet toolbox, CellDesigner
- ▶ rekonstrukce sítí
 - ▶ nástroje: Genomica, GeneNetworks
- ▶ analýza sítí a toků
 - ▶ nástroje: Cytoscape, COPASI

Informace o kurzu

Genomové data prohlížeče

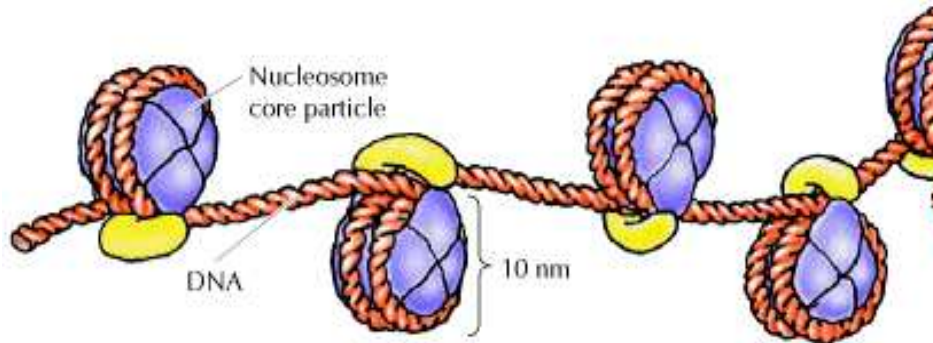
Informace o kurzu

Genomové data prohlížeče



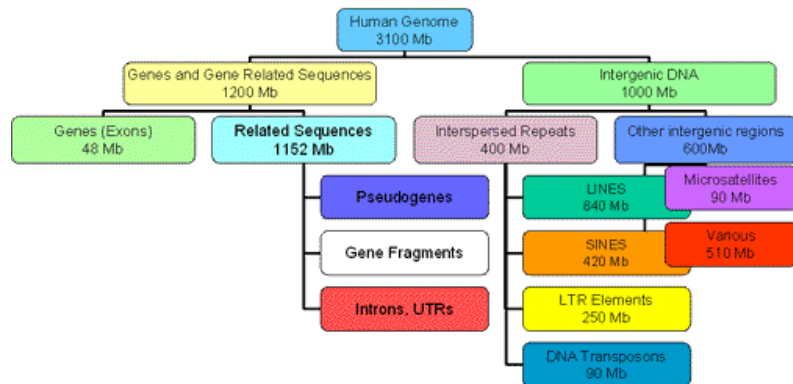
Informace o kurzu

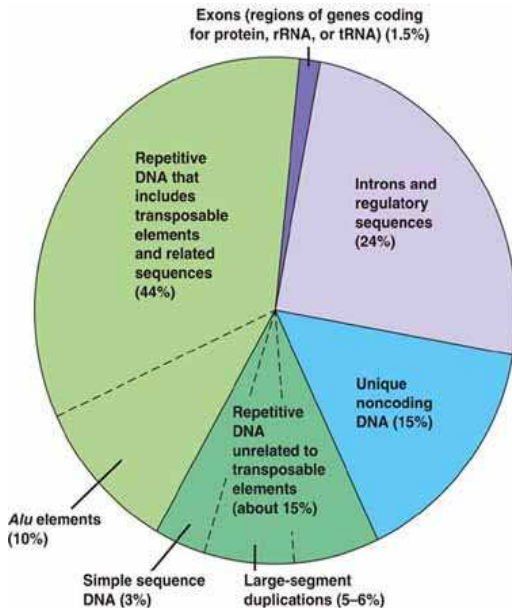
Genomové data prohlížeče



10-nm fiber

30-nm





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

- ▶ predikce genů
- ▶ homologie (zjišťování podobnosti sekvencí)

- ▶ Geny kódující
 - ▶ proteiny (exon, intron)
 - ▶ RNA
- ▶ Regulační sekvence
 - ▶ promotory
 - ▶ enhancery
 - ▶ jiné
- ▶ Repetitivní sekvence
- ▶ Cizí sekvence
 - ▶ viry
 - ▶ transpozony
- ▶ Oblasti podobnosti (homology)
 - ▶ SNP
 - ▶ delší strukturní variace

UCSC Genome Browser

Informace o kurzu

Genomové data prohlížeče

Human chr5: 70,256,524-70,284,592 - UCSC Genome Browser v134 - Konqueror

Location Edit View Bookmarks Tools Settings Help

id=73350821&knownGene=full

Human chr5:70,256,524-70,28...

Home Genomes Blat Tables Gene Sorter PCR DNA Convert PDF/PS Help

UCSC Genome Browser on Human Mar. 2006 Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x

10x

position/search chr5:70,256,524-70,284,592 jump clear size 28,069 bp. configure

chr5 (q13.2)

chr5: 70260000 70265000 70270000 70275000 70280000

STS Markers STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps

UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

Human mRNAs Human mRNAs From GenBank

Spliced ESTs Human ESTs That Have Been Spliced

Conservation Vertebrate Multiz Alignment & Conservation (17 Species)

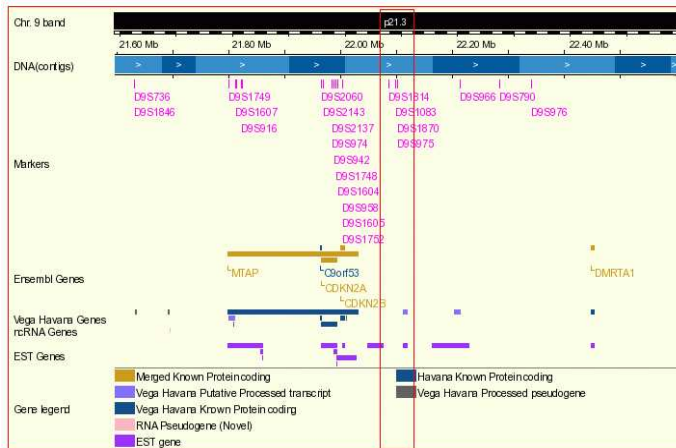
mouse rat rabbit dog armadillo elephant opossum chicken

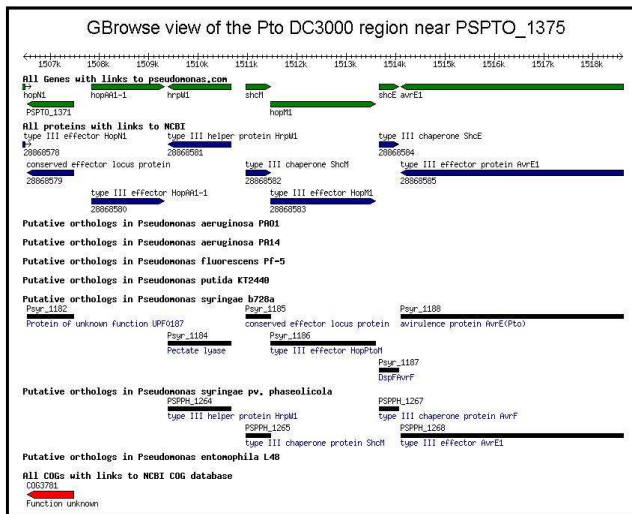
<http://genome.ucsc.edu/cgi-bin/hgc?hgid=733...523&r=70284592&db=hg18&pix=620>

Ensembl Genome Browser

Informace o kurzu

Genomové data prohlížeče





Informace o kurzu

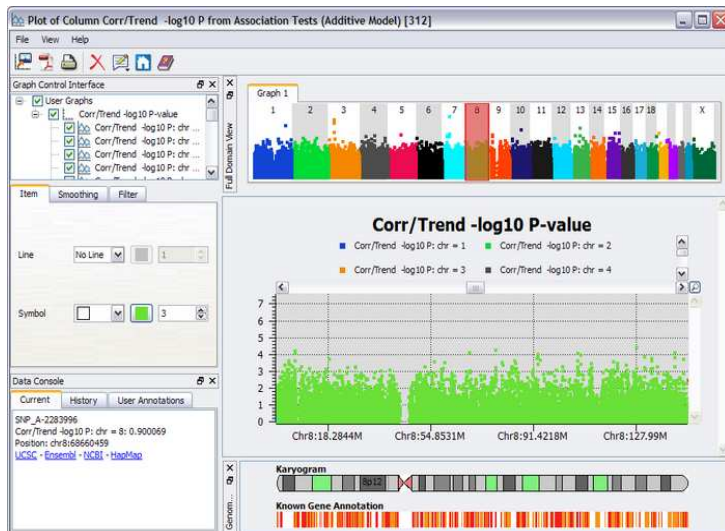
Genomové data prohlížeče



Golden Helix Genome Browser

Informace o kurzu

Genomové data prohlížeče



Informace o kurzu

Genomové data prohlížeče

Chromosome 1 (Arabidopsis thaliana TAIR8) - Integrated Genome Browser 5.5

File Edit View Bookmarks Tools Help

0 - 30,432,563

TAIR8_snoRNA (+)
TAIR8_ncRNA (+)
TAIR8_rRNA (+)
TAIR8_tRNA (+)
TAIR8_mRNA (+)
Coordinates
TAIR8_mRNA (-)
TAIR8_rRNA (-)
TAIR8_ncRNA (-)
TAIR8_snoRNA (-)

0 5,000,000 10,000,000 15,000,000 20,000,000 25,000,000 30,000,000

NEWSP4378

Data Access Selection Info Sliced View Graph Adjuster Pattern Search Name Search Annotation Browser

Choose: Arabidopsis thaliana A. thaliana_TAIR8

Choose Data Sources and Data Sets: Bovivz DAS2

Choose Load Mode for Data Sets:

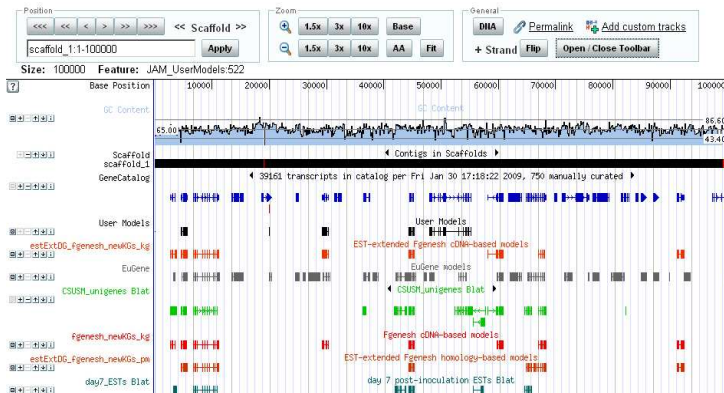
Choose Load Mode	Data Set	Data Source	Data Source Type
Whole Gen...	TAR8_ncRNA	Bovivz QuickLoad	Quickload
Whole Gen...	TAR8_snoRNA	Bovivz QuickLoad	Quickload
Whole Gen...	TAR8_snoR...	Bovivz QuickLoad	Quickload
Whole Gen...	TAR8_rRNA	Bovivz QuickLoad	Quickload

Load All Sequence Load Sequence in View Refresh Data

Current Sequence

Sequence	Length
chr1	104,925,63
chr2	19,705,559
chr3	21,470,805
chr4	18,585,042
chr5	26,992,728
chrM	36,692,4
chrC	15,447,8
genome	144,907,899

65.3 MB / 1,016.1 MB



RIKEN Genome Browser

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

- 1**: Navigation and search options, including a search bar and a "Register current interval" button.
- 2**: A sidebar menu for database selection, listing various databases such as Mouse miRBase, FANTOM3, CAGE, and RefSeq.
- 3**: A genomic map showing the current interval (82,100,466 bp to 82,182,103 bp) with a specific marker (821,15634-821,70440) highlighted.
- 4**: A detailed view of the genomic region, showing the marker symbol (IRS1_MOUSE) and the RefSeq peptide (NP_034700.2).
- 5**: A list of tracks for the selected marker, including FANTOM3, CAGE, Ensembl Transcripts, and RefSeq Peptide.
- 6**: A detailed view of the selected marker, showing the gene structure and the RefSeq peptide sequence.

