

PB051 Výpočetní metody v bioinformatice a systémové biologii

Týden 1

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2013

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: Čt 10:00-11:50 (B117)
- ▶ Konzultace: Čt 13:00-15:00 (C506)
- ▶ <http://www.fi.muni.cz/~lexa/teaching.html>

- ▶ 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

- ▶ 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 a Table Browser
- ▶ Programovatelný přístup ke genomu přes rozhraní Biomart z prostředí R/Bioconductor
- ▶ Bioinformatika genové regulace (TRANSFAC, MatInspector))
- ▶ Datové zdroje týkající se interakčních sítí

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

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

- ▶ 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

Genomové data prohlížeče

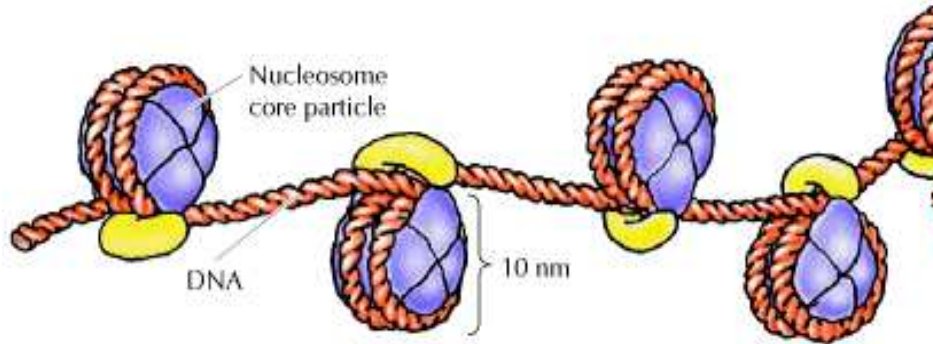
Informace o kurzu

Genomové data prohlížeče



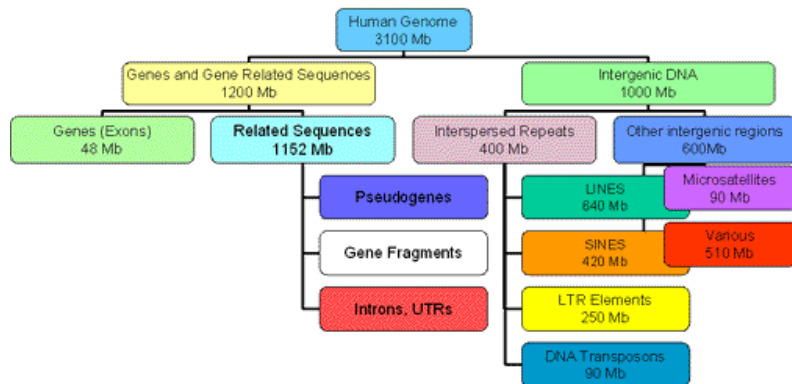
Informace o kurzu

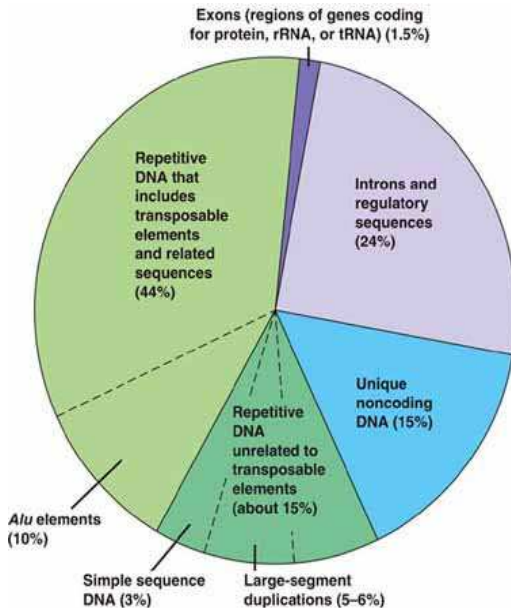
Genomové data prohlížeče



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

- ▶ 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

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)

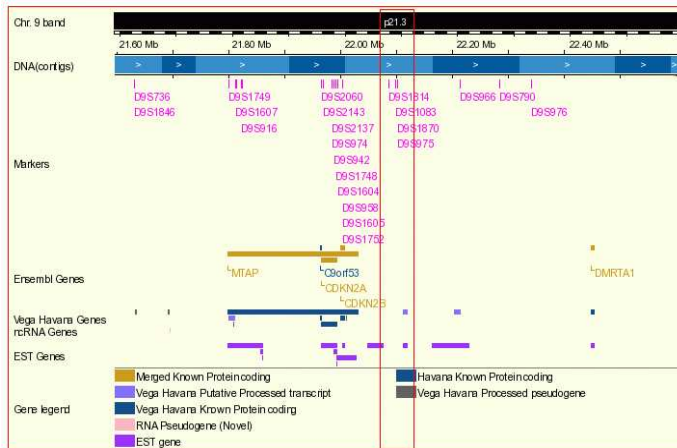
mouse rat rabbit dog armadillo elephant opossum chicken

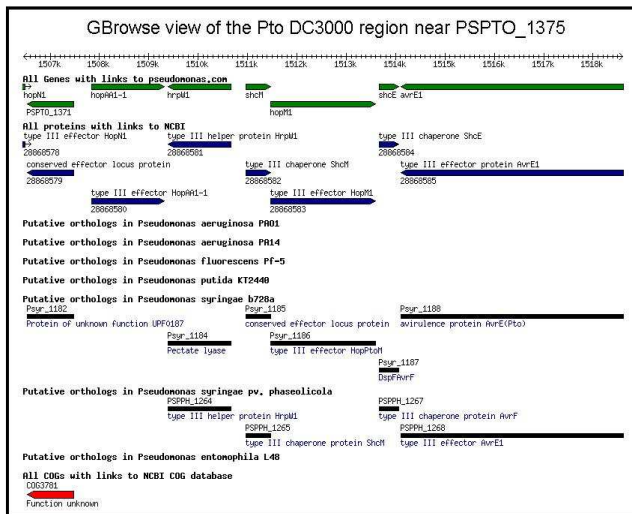
<http://genome.ucsc.edu/cgi-bin/hgc?hgid=733...523&r=70284592&db=hg18&pix=620>

Ensembl Genome Browser

Informace o kurzu

Genomové data prohlížeče





Informace o kurzu

Genomové data prohlížeče

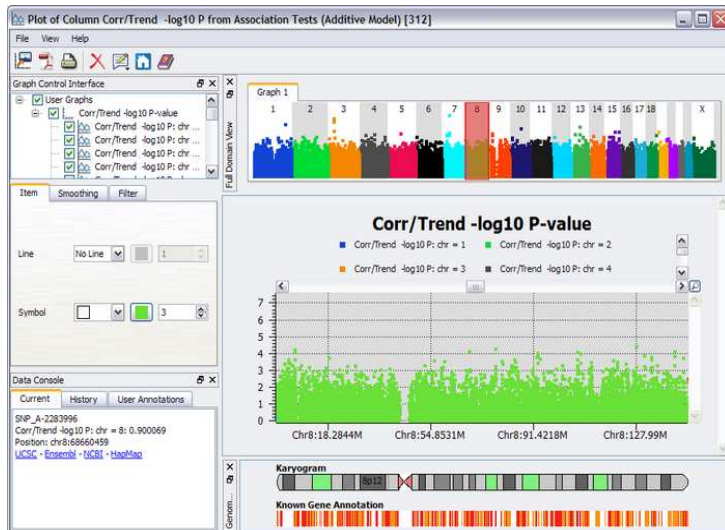
The screenshot displays the Argo genome browser interface. At the top, there is a menu bar with options: Argoc, File, Track, Edit, Select, View, Zoom, Rulers, Analyze, User, Bookmarks, Window, Help. The main window shows a genomic track for Human Chr15, contig 1.1 (1,399,746:1-100,000). The track includes various features such as genes, transcripts, and repeats, represented by colored arrows and boxes. A 'PolyA Signal View: Novel Transcrip...' window is open, showing a list of polyA signals with their coordinates and scores:

Signal	Score	Coordinates
AATACA	(1.28)	415-420 82808-82813
AATACA	(1.28)	470-475 86534-86539
AATAAA	(14.91)	804-809 89548-89553

Below the main track, there is an 'Inspector' panel with tabs for Properties, DNA, mRNA, and Protein. The DNA tab is selected, showing the sequence: `AGTAATATAGAA...TGGACAGATGAA...AAAATAGCAATGTCATAAGGACGTCCTACACCTGGGCTGGGCGCAATGGCCGACGACAGAGTGTATACGCTCTGGTATGATGAGAAAGTGTGACGCTGTACACCTTTCGATGGTAAAGACAGGCAACCTCTG...AAGCGTGGCGCAATGCCAGAGGGAGGTTGTGCAAAATATCTCATAGATCTCTGGTCTGATCCAAATATTTAGATGATGTTATGACCAACAGCTGTCCATTATGCTCTTTATGGTGG...Nucleotide C 127/80162...AATTTGCTGTCTCTGGTGCACACATGCAAGTGTGAAGAACAGCGCTGGCCACACACCC...GAGAGGTTGAGAAATGTTGGAAATTTTACGACAGAAATATGCAATGCAATGCTGATATATTTAAATTCATTCACTACACACTTTTTC...AATACAACCAAAAGATATCTAAAATTCCTAAAATGACAACTCCAGAAAGGAGATCTGAGGAGACCTGACGAGAGCTGCACCTTGGTGGAAAGAACACTGAC...GCTGAAAGCTTGGTGGAAAGAACACTGACGA...GATACAGTGAATTCCTTCAAAAGATTTAGCCTGTAAAATTCCTTAAAATTCGAAAGAGGGGTTGATGATACAGTGAATTCCT`. The 'Finder' panel on the right allows searching for features, with filters for Label, Protein length, and mRNA Sequence.



Golden Helix Genome Browser



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_mRNA (+)
TAIR8_rRNA (+)
TAIR8_tRNA (+)

TAIR8_mRNA (+)

Coordinates

TAIR8_mRNA (-)
TAIR8_rRNA (-)
TAIR8_miRNA (-)
TAIR8_ncRNA (-)
TAIR8_snoRNA (-)
TAIR8_tRNA (-)

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:

Choose Load Mode for Data Sets:

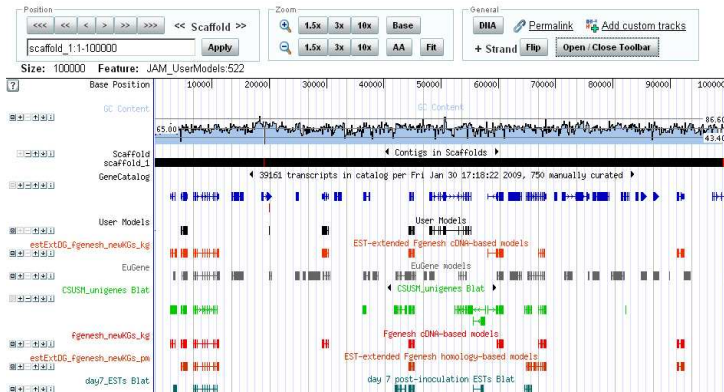
Choose Load Mode	Data Set	Data Source	Data Source Type
Whole Gen...	TAIR8_mRNA	Boviz QuickLoad	Quickload
Whole Gen...	TAIR8_ncRNA	Boviz QuickLoad	Quickload
Whole Gen...	TAIR8_snoR...	Boviz QuickLoad	Quickload
Whole Gen...	TAIR8_rRNA	Boviz QuickLoad	Quickload

Current Sequence

Sequence	Length
chr1	104,253,63
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

Load All Sequence Load Sequence in View Refresh Data

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**: Browser navigation and database selection. The top left shows a search bar with "Go to Search page" and "Register current interval". Below it, a "Database list" is visible, including Mouse miRBase version 5, Mouse FANTOM3, Mouse Fantom3 Mega Genes, Mouse CAGE, Mouse Mutant Resources, Mouse Ensembl Transcripts, Mouse Ensembl VegaGene, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, Mouse UniProtKB SwissProt, Mouse UniProtKB TrEMBL, Mouse Riken Transcription, Mouse dbSNP NCBI, Mouse Ensembl Gene 40, Mouse Ensembl VegaGene, Mouse Ensembl Gene NCBI, Mouse MGI Gene, and Mouse RefSeq Oribacterium.
- 2**: Expert's set and User's set. The bottom left shows a sidebar with "Expert's set" and "User's set" options. Under "Genome View", there are links for Mouse, Human Homology, Catagans Homology, Other Homology, All, dbSNP, Medline (Positional Medline), Transcriptome (FANTOM, CAGE).
- 3**: Search and filter options. The top center shows a search bar with "Filter by: keyword" and a "Go" button.
- 4**: Genomic map. The top right shows a genomic map of a region on chromosome 1 of *Mus musculus*, with coordinates 82,100,466 bp, 81,637 bp, and 82,182,103 bp. A red arrow points to a specific location on the map.
- 5**: Gene and transcript information. The middle right shows a list of genes and transcripts, including:
 - ENSMUSG0000055880: MarkerSymbol:irs1|Uniprot/SWISSPROT:IRS1_MOUSE|RefSeq_peptide:NP_034700.2|RefSeq_dnaNM:010570.2|Uniprot/SPTREMBL:Q543V3_MOUSE|Entrez
 - Mouse FANTOM3: (-DNA(+), -DNA(-))
 - Mouse CAGE (all tissues) Expression(TAG=249, TPM=34.62): Histogram(+), Histogram(-)
 - Mouse Ensembl Transcript 45.36a: transcript(+), transcript(-)
 - Mouse Ensembl VegaGene Transcript 45.36a: transcript(+), transcript(-)
 - Mouse GTOP: gene(+), gene(-)
 - Mouse RefSeq Peptide: protein(+), protein(-)
 - Mouse RefSeq DNA: dna(+), dna(-)
 - Mouse UniProtKB SwissProt: protein(+), protein(-)
- 6**: Navigation and zoom controls. The bottom right shows navigation buttons for "Select strand", "Grid", and zoom controls (+, -, up, down, left, right).

Informace o kurzu

Genomové data prohlížeče

