

IV107 Bioinformatika I

Přednáška 5

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2010

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

- ▶ Struktura genu
 - ▶ prokaryotického
 - ▶ eukaryotického
- ▶ Porovnání sekvencí
 - ▶ globální (Needleman–Wunsch)
 - ▶ semi-globální
 - ▶ lokální (Smith–Waterman)

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Analýza proteinových sekvencí, strukturních a funkčních dat

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

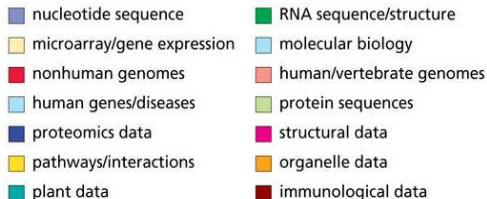
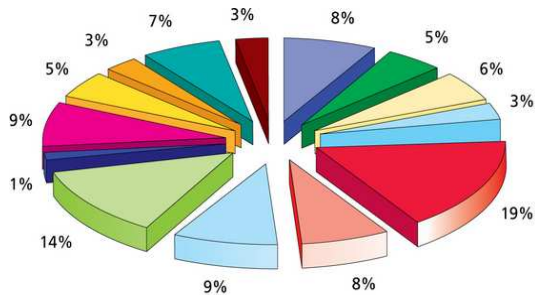
Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat



Bioinformatické databáze

Databáze GenBank
Databáze UniProt
Protein Data Bank
Gene Ontology
KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

<http://www.agr.kuleuven.ac.be/vak>

Nárůst databáze GenBank

IV107 Bioinformatika I -
Přednáška 4

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

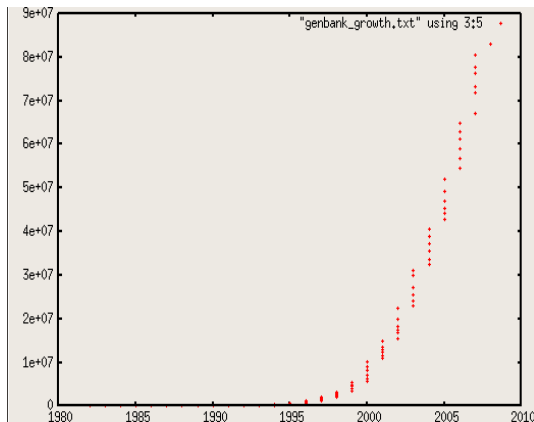
Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturálních a funkčních dat



Genetic Sequence Data Bank
February 15 2008
NCBI-GenBank Flat File Release 164.0
National Center for Biotechnology Information

- ▶ 85759586764 bases
- ▶ 82853685 sequences

<ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt>

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

- ▶ INV, VRT, MAM, PLN, PRI, ROD, BCT, VRL
- ▶ PAT (Patents)
- ▶ HTGS (High Throughput Genomic Sequences)
- ▶ GSS (Genome Survey Sequences)
- ▶ ETS (Expressed Sequence Tags)
- ▶ STS (Sequence Tagged Sites)

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Příklad záznamu v databázi GenBank

LOCUS SCU49845 5028 bp DNA
DEFINITION Saccharomyces cerevisiae TCP1-beta gene,
Axl2p
(AXL2) and Rev7p (REV7) genes, complete
ACCESSION U49845
VERSION U49845.1 GI:1293613
KEYWORDS .
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomy
Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; S

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

- ▶ textové (klíčová slova)
- ▶ sekvenční (BLAST)

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Příklad záznamu v databázi UniProt

Bioinformatické databáze

Databáze GenBank

Databáze UniProt


Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Entry information	
Entry name	LMO7_HUMAN
Primary accession number	Q8WW11
Secondary accession numbers	O15462 Q95346 Q9UKC1 Q9UQM5 Q9Y6A7
Integrated into Swiss-Prot on	March 15, 2004
Sequence was last modified on	March 15, 2004 (Sequence version 2)
Annotations were last modified on	July 25, 2006 (Entry version 39)
Name and origin of the protein	
Protein name	LIM domain only protein 7
Synonyms	LOMP F-box only protein 20
Gene name	Name: LMO7 Synonyms: FBX20, FBXO20, KIAA0858
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eulacostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhina; Catarrhini; Hominidae; Homo.
References	
[1]	NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 3), AND TISSUE SPECIFICITY. TISSUE=Brain, and Peripheral blood leukocyte; DOI=10.1007/s00439-001-0646-6; PubMed=11935316 [NCBI, ExPASy, EBI, Israel, Japan] Rozenblum E., Vahteristo P., Sandberg T., Bergthorsson J.T., Syrjakoski K., Weaver D., Haraldsson K., Johannsdottir H.K., Vehmanen P., Nigam S., Golberger N., Robbins C., Pak E., Dutra A., Gillander E., Stephan D.A., Bailey-Nilson J., Juo S.-H.H., Kainu T.,  , Kallioniemi O.-P.; "A genomic map of a 6-Mb region at 13q21-q22 implicated in cancer development: identification and characterization of candidate genes."; Hum. Genet. 110:111-121(2002).

<http://www.uniprot.org/>

Příklad záznamu v databázi UniProt

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Key	From	To	Length	Description	FTId
CHAIN	1	1683	1683	LIM domain only protein 7.	PRO_0000075824
DOMAIN	54	168	115	CH.	
DOMAIN	1042	1128	87	PDZ.	
DOMAIN	1612	1678	67	LIM zinc-binding.	

```
      10      20      30      40      50      60
MKKIRICHIF TFYSWMSYDV LFQRTTELGA EIWRQLICAH VCICVGWLYL RDRVCSKKDI

      70      80      90     100     110     120
ILRTEQNSGR TILIKAVTEK NFETKDFRAS LENGVLLCDL INKLKPGVIK KINRLSTPIA

     130     140     150     160     170     180
GLDNINVFLK ACEQIGLKEA QLFHPGDLQD LSNRVTVKQE ETDRRVKNVL ITLYWLGRKA
```

RCSB **PDB**
PROTEIN DATA BANK

A MEMBER OF THE **CPDB**

An Information Portal to Biological Macromolecular Structures


Contact Us | Help | Print Page

PDB ID or keyword | Author | | |

91 Structure Hits | 127 Web Page Hits | 1 Unreleased Structure

1 2 3 4 5 .. 10 ↩

- Results (1-10 of 91)
- Results ID List
- Refine this Search
- 1 Structures Awaiting Release
- Select All
- Deselect All
- Download Selected
- Tabulate
- Narrow Query
- Sort Results
- Results per Page
- Show Query Details
- Results Help

<input checked="" type="checkbox"/>	1X6Z		Characteristics Classification Compound Authors	Solution structure of the LIM domain of carboxyl terminal LIM domain protein 1 Release Date: 17-Nov-2005 Exp. Method: NMR 20 Structures Structural Protein Mol. Id: 1 Molecule: C Terminal Lim Domain Protein 1 Fragment: Lim Domain Qin, X.R., Nagashima, T., Hayashi, F., Yokoyama, S.
<input checked="" type="checkbox"/>	1X4K		Characteristics Classification Compound Authors	Solution structure of LIM domain in LIM-protein 3 Release Date: 14-Nov-2005 Exp. Method: NMR 20 Structures Metal Binding Protein Mol. Id: 1 Molecule: Skeletal Muscle Lim Protein 3 Fragment: Lim Domain He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,
<input checked="" type="checkbox"/>	1X4L		Characteristics Classification Compound Authors	Solution structure of LIM domain in Four and a half LIM domains protein 2 Release Date: 14-Nov-2005 Exp. Method: NMR 20 Structures Metal Binding Protein Mol. Id: 1 Molecule: Skeletal Muscle Lim Protein 3 Fragment: Lim Domain He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Přístě

 Analýza proteinových sekvencí,
 strukturních a funkčních dat

```

HEADER      HYDROLASE (O-GLYCOSYL)                20-JAN-92  1HEW      1HEW  2
COMPND      LYSOZYME (E.C.3.2.1.17) COMPLEXED WITH THE INHIBITOR  1HEW  3
COMPND      2 TRI-N-ACETYLCHITOTRIOSE              1HEW  4
SOURCE      HEN (GALLUS GALLUS) EGG WHITE          1HEW  5
AUTHOR      J.C.CHEETHAM,P.J.ARTYMIUK,D.C.PHILLIPS  1HEW  6
REVDAT      1 31-JAN-94 1HEW 0                    1HEW  7
JRNL        AUTH  J.C.CHEETHAM,P.J.ARTYMIUK,D.C.PHILLIPS  1HEW  8
JRNL        TITL  REFINEMENT OF AN ENZYME COMPLEX WITH INHIBITOR  1HEW  9
JRNL        TITL  2 BOUND AT PARTIAL OCCUPANCY. HEN EGG-WHITE  1HEW 10
JRNL        TITL  3 LYSOZYME AND TRI-N-ACETYLCHITOTRIOSE AT 1.75  1HEW 11
JRNL        TITL  4 ANGSTROMS RESOLUTION              1HEW 12
JRNL        REF   J.MOL.BIOL.                          V. 224  613 1992  1HEW 13
JRNL        REFN  ASTM JMOBAK UK ISSN 0022-2836        070  1HEW 14
REMARK      1                                          1HEW 15
REMARK      1 REFERENCE 1                              1HEW 16
REMARK      1 AUTH  L.N.JOHNSON,J.C.CHEETHAM,P.J.MC*LAUGHLIN,  1HEW 17
REMARK      1 AUTH  2 K.R.ACHARYA,D.BARFORD,D.C.PHILLIPS      1HEW 18
REMARK      1 TITL  PROTEIN-OLIGOSACCHARIDE INTERACTIONS: LYSOZYME,  1HEW 19
REMARK      1 TITL  2 PHOSPHORYLASE, AMYLASES            1HEW 20
REMARK      1 REF   CURR.TOP.MICROBIOL.IMMUNOL.        V. 139   81 1988  1HEW 21
REMARK      1 REFN  ASTM CTMIA3 GW ISSN 0070-217X        761  1HEW 22
    
```

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

```
REMARK      5 THE THREE SUGAR UNITS OF THE INHIBITOR MOLECULE ARE BOUND      1HEW      56
REMARK      5 IN THE UPPER THREE SITES (A TO C) OF THE LYSOZYME ACTIVE      1HEW      57
REMARK      5 SITE CLEFT.  NAG MOLECULES, NUMBERED 203, 202, AND 201, ARE  1HEW      58
REMARK      5 BOUND IN SITES A, B, AND C, RESPECTIVELY.                    1HEW      59
SEQRES      1   129  LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS      1HEW      60
SEQRES      2   129  ARG HIS GLY LEU ASP ASN TYR ARG GLY TYR SER LEU GLY      1HEW      61
SEQRES      3   129  ASN TRP VAL CYS ALA ALA LYS PHE GLU SER ASN PHE ASN      1HEW      62
SEQRES      4   129  THR GLN ALA THR ASN ARG ASN THR ASP GLY SER THR ASP      1HEW      63
SEQRES      5   129  TYR GLY ILE LEU GLN ILE ASN SER ARG TRP TRP CYS ASN      1HEW      64
SEQRES      6   129  ASP GLY ARG THR PRO GLY SER ARG ASN LEU CYS ASN ILE      1HEW      65
SEQRES      7   129  PRO CYS SER ALA LEU LEU SER SER ASP ILE THR ALA SER      1HEW      66
SEQRES      8   129  VAL ASN CYS ALA LYS LYS ILE VAL SER ASP GLY ASN GLY      1HEW      67
SEQRES      9   129  MET ASN ALA TRP VAL ALA TRP ARG ASN ARG CYS LYS GLY      1HEW      68
SEQRES     10   129  THR ASP VAL GLN ALA TRP ILE ARG GLY CYS ARG LEU      1HEW      69
HET      NAG      201      15      N-ACETYL-D-GLUCOSAMINE      1HEW      70
HET      NAG      202      14      N-ACETYL-D-GLUCOSAMINE      1HEW      71
HET      NAG      203      14      N-ACETYL-D-GLUCOSAMINE      1HEW      72
FORMUL      2  NAG      3(C8 H15 N1 O6)      1HEW      73
```

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

```

HELIX  1  A  ARG      5  HIS      15  1                                LHEW  75
HELIX  2  B  LEU     25  GLU     35  1                                LHEW  76
HELIX  3  C  CYS     80  LEU     84  5                                LHEW  77
HELIX  4  D  THR     89  ILE     98  1                                LHEW  78
HELIX  5  E  VAL    109  ASN    113  1                                LHEW  79
SHEET  1  S1 2  LYS      1  PHE      3  0                                LHEW  80
SHEET  2  S1 2  PHE     38  THR     40 -1  N   THR     40  O   LYS     1  LHEW  81
SHEET  1  S2 3  ALA     42  ASN     46  0                                LHEW  82
SHEET  2  S2 3  SER     50  GLY     54 -1  O   SER     50  N   ASN     46  LHEW  83
SHEET  3  S2 3  GLN     57  SER     60 -1  O   ILE     58  N   TYR     53  LHEW  84
TURN   1  T1  MET     12  HIS     15          TYPE III                                LHEW  85
TURN   2  T2  LYS     13  GLY     16          TYPE I                                 LHEW  86
TURN   3  T3  LEU     17  TYR     20          TYPE II                                LHEW  87
TURN   4  T4  ASN     19  GLY     22          DISTORTED TYPE II                       LHEW  88
TURN   5  T5  TYR     20  TYR     23          TYPE I '                                LHEW  89
TURN   6  T6  SER     24  ASN     27          TYPE III                                LHEW  90
TURN   7  T7  LEU     25  TRP     28          TYPE III                                LHEW  91
TURN   8  T8  SER     36  ASN     39          TYPE III '                               LHEW  92
    
```

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

```

CRYST1      78.860   78.860   38.250   90.00   90.00   90.00 P  43 21 2      8  1HEW 113
ORIGX1          1.000000   0.000000   0.000000          0.000000          1HEW 114
ORIGX2          0.000000   1.000000   0.000000          0.000000          1HEW 115
ORIGX3          0.000000   0.000000   1.000000          0.000000          1HEW 116
SCALE1         0.012681   0.000000   0.000000          0.000000          1HEW 117
SCALE2         0.000000   0.012681   0.000000          0.000000          1HEW 118
SCALE3         0.000000   0.000000   0.026144          0.000000          1HEW 119
ATOM          1  N   LYS      1          3.398   9.981  10.408   1.00  30.48          1HEW 120
ATOM          2  CA  LYS      1          2.459  10.365   9.364   1.00  28.03          1HEW 121
ATOM          3  C   LYS      1          2.458  11.880   9.149   1.00  21.93          1HEW 122
ATOM          4  O   LYS      1          2.481  12.672  10.100   1.00  14.10          1HEW 123
ATOM          5  CB  LYS      1          1.026   9.935   9.695   1.00  30.54          1HEW 124
ATOM          6  CG  LYS      1          0.028  10.169   8.558   1.00  37.93          1HEW 125
ATOM          7  CD  LYS      1          -1.415  10.089   9.048   1.00  33.23          1HEW 126
ATOM          8  CE  LYS      1          -2.357  10.822   8.082   1.00  32.17          1HEW 127
ATOM          9  NZ  LYS      1          -3.661  10.090   8.025   1.00  31.92          1HEW 128
ATOM         10  N   VAL      2          2.429  12.232   7.880   1.00  17.30          1HEW 129
ATOM         11  CA  VAL      2          2.395  13.653   7.465   1.00  14.47          1HEW 130
ATOM         12  C   VAL      2          0.977  13.868   6.903   1.00  17.58          1HEW 131
ATOM         13  O   VAL      2          0.642  13.368   5.826   1.00  32.65          1HEW 132
ATOM         14  CB  VAL      2          3.533  14.012   6.536   1.00  22.88          1HEW 133
    
```


Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

- ▶ Funkce genů a proteinů zjišťujeme experimentálně
- ▶ Slovní popis není jednoznačný
 - ▶ syntéza proteinů
 - ▶ syntéza polypeptidů
 - ▶ translace
 - ▶ aktivita ribozomů
- ▶ Ontologie je způsob jak do používaných termínů vnést systém

- ▶ Molekulární proces
 - ▶ katalytická aktivita
 - ▶ transport
 - ▶ intermolekulární vazba
- ▶ Biologický proces
 - ▶ přenos signálu
 - ▶ aktivace imunitního systému
 - ▶ regulace genů
- ▶ Buněčná složka
 - ▶ buněčné jádro
 - ▶ plazmatická membrána

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturálních a funkčních dat

Curator-assigned Evidence Codes

- ▶ Experimental Evidence Codes
 - ▶ IDA: Inferred from Direct Assay
 - ▶ IPI: Inferred from Physical Interaction
 - ▶ IMP: Inferred from Mutant Phenotype
 - ▶ IGI: Inferred from Genetic Interaction
 - ▶ IEP: Inferred from Expression Pattern
- ▶ Computational Analysis Evidence Codes
 - ▶ ISS: Inferred from Sequence or Structural Similarity
 - ▶ IGC: Inferred from Genomic Context
 - ▶ RCA: inferred from Reviewed Computational Analysis
- ▶ Author Statement Evidence Codes
 - ▶ TAS: Traceable Author Statement
 - ▶ NAS: Non-traceable Author Statement
- ▶ Curator Statement Evidence Codes
 - ▶ IC: Inferred by Curator
 - ▶ ND: No biological Data available
- ▶ Automatically-assigned Evidence Codes
 - ▶ IEA: Inferred from Electronic Annotation
- ▶ Obsolete Evidence Codes

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

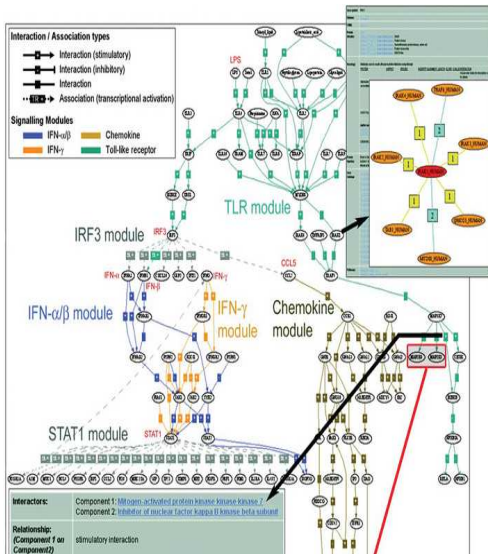
Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturálních a funkčních dat

Metabolické dráhy



Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

<http://www.genome.jp/kegg/>

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?hgsid=733...523&r=70284592&db=hg18&pix=620>

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

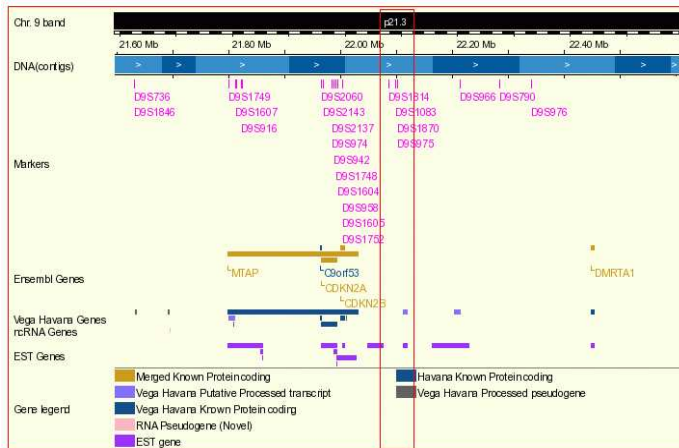
Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat



Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturálních a funkčních dat

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

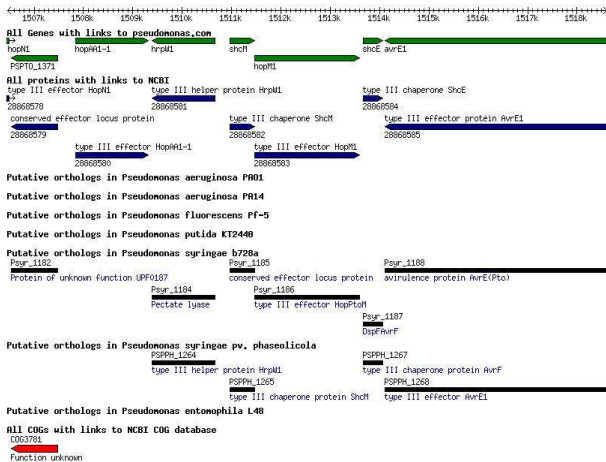
Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturálních a funkčních dat

GBrowse view of the Pto DC3000 region near PSPTO_1375



The screenshot displays the Argo genome browser interface. At the top, the menu bar includes: Argo, File, Track, Edit, Select, View, Zoom, Rulers, Analyze, User, Bookmarks, Window, Help. The status bar shows '(9990) Sat 5:24 AM'.

The main window shows a 'Feature Map: Human Chr15 contig 1.1 (L = 1399746):1-100000'. The track displays various genomic features, including genes and transcripts, represented by colored arrows and lines. A 'PolyA Signal View: Novel Transcrip...' window is open, showing a list of polyA signals:

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

The 'Inspector' panel at the bottom left shows the 'Properties' tab with 'DNA', 'mRNA', and 'Protein' options. The 'mRNA' tab is selected, displaying the following sequence:

```

AGTAATATAGAA...TGGACAGATGAA...AAAATAGCAATGTCATAAGGACGTCCTACAGCTGGGCTGGG
CAAATGGCCACGCAAGGTTGATACGCTCTGTTGATGATGAAAGTGTGACGCTGTGACATCTTGAAGTGGTAAACAGGA
CAAGCTCG...AAGCGTGGCAATGCCAAGGAGGGGTTGTGCAAAATATCTCATGATCTCTGGCTGATCCAAAT
ATTGTAGATGTTATGACCAACACAGCTGTCCATTATGCTCTTTATGGT...Nucleotide C 127/80162...AATGTCT
GTCTCTGTGGTCAACATGCAAGTGAAGAAACAGGCTGGCCACAGCCCA...GGAAGGTG
MCKAAATGTGGAAATTTTACCTACGAAABATACAAATGCAATGCAATGCAAGTGTATATCTTAAATCATTTCATCAACAA
CTTTTGG...AATACA...AACAAAGATATCTAAAAATCTCAAAATAGCAATCCACAAAGGAGATCTGAAAGAACCTGA
CGAGCGTCACCCCTTGGTGGAAAGAACACTGAC...GCTGAAAGCTTGGTGGAAAGAACACTGACGA...GATA
CAGTGAATGCTCTCAAGAAATTTAGCCTGTAAAAATCTTAAAAATTCAAAGAGGGGTTAGTACAGTGAATCT

```

The 'Finder' panel at the bottom right allows for searching features. It includes a 'Select Features whose:' section with the following filters:

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

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

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat



Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

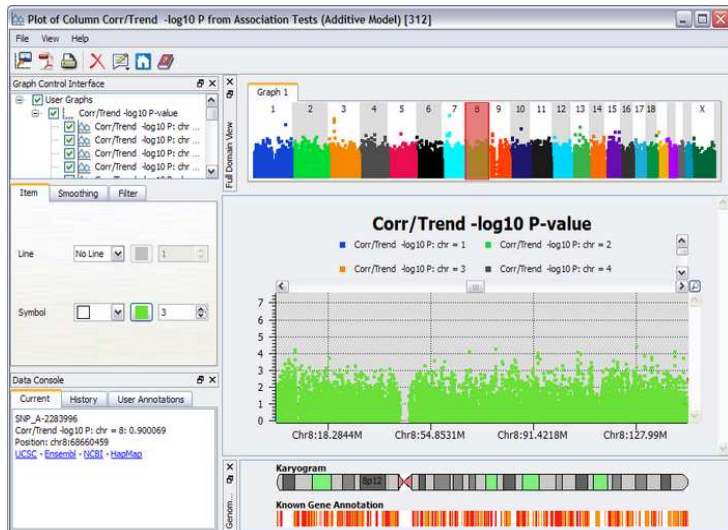
Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Golden Helix Genome Browser



Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Bioinformatické databáze

Databáze GenBank

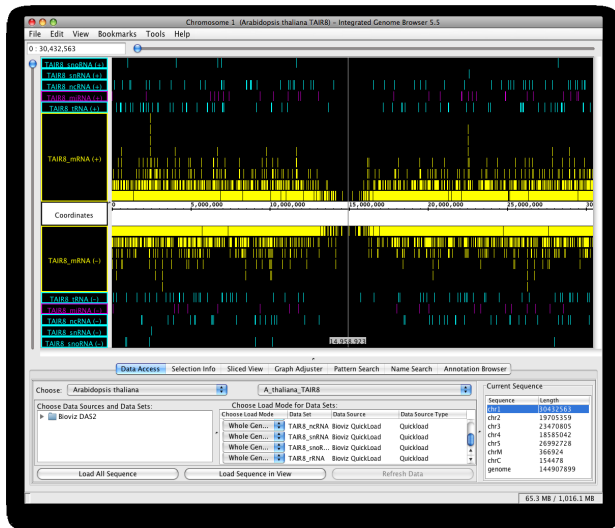
Databáze UniProt

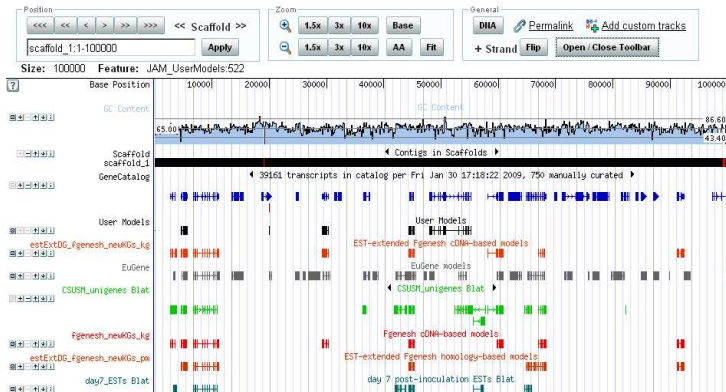
Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat



Bioinformatické databáze

Databáze GenBank
Databáze UniProt
Protein Data Bank
Gene Ontology
KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

1 Database list: Mouse miRBase version 5, Mouse FANTOM3, Mouse FANTOM3 Mega Genes, Mouse CAGE, Mouse Mutant Resources, Mouse Ensembl Transcription, Mouse Ensembl VegaGene, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, Mouse UniProtKB SwissProt, Mouse UniProtKB TrEMBL, Mouse Riken Transcription, Mouse dbSNP NCBIn38, Mouse Ensembl Gene 40, Mouse Ensembl VegaGene, Mouse Ensembl Gene NCBIn38, Mouse MGI Gene, Mouse RefSeq Peptide.

2 Expert's set: Mouse, Human Homology, Catagens Homology, Other Homology, All, dbSNP, Medline (Positional Medline), Transcriptome (FANTOM, CAGE).

3 Go to Search page, Register current interval, Filter by: keyword.

4 Mus musculus: Irf1, 82,100,466 bp, 81,637 bp, 82,182,103 bp.

5 ENSMUSG0000055880, MarkerSymbol:Irf1|UniProt/SWISSPROT:IRF1_MOUSE|RefSeq_peptide:NP_034700.2|RefSeq_dnaNM:010570.2|UniProt/SPTREMBL:Q543V3_MOUSE|Entrez

6 Mouse FANTOM3, Mouse CAGE (all tissues) Expression(TAG=249, TPM=34.62), Mouse Ensembl Transcript 43.36a, Mouse Ensembl VegaGene Transcript 43.36a, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, Mouse UniProtKB SwissProt.

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Bioinformatické databáze

Databáze GenBank

Databáze UniProt

Protein Data Bank

Gene Ontology

KEGG

Příště

Analýza proteinových sekvencí,
strukturních a funkčních dat

Analýza proteinových sekvencí, strukturních a funkčních dat

Dodatek

Dodatek

For Further Reading

For Further Reading

Dodatek

For Further Reading

X