

Bi7201 Základy genomiky

Přednáška 1

Úvod do bioinformatiky

Jan Hejátko

**Laboratoř molekulární fyziologie rostlin,
Oddělení Funkční Genomiky a Proteomiky,
Masarykova univerzita, Brno**

hejatko@sci.muni.cz, www.sci.muni.cz/FGP/



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky



Osnova

- Schéma předmětu
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje
- Analytické nástroje
 - Vyhledávání homologií
 - Vyhledávání sekvenčních motivů, otevřených čtecích rámců, restričních míst....
 - Další www genomové nástroje

Schéma přednášky

- **Přednáška 1**
 - Úvod do bioinformatiky
 - Teorie základních bioinformatických nástrojů

- **Přednáška 2**
 - Identifikace genů
 - *In silico* i experimentální

- **Přednáška 3**
 - Přístupy reverzní genetiky

- **Přednáška 4**
 - Přístupy genetiky přímé

Schéma přednášky

- **Přednáška 5**
 - Přístupy funkční genomiky

- **Přednáška 6**
 - Protein-protein interakce a jejich analýza
 - Praktické aplikace funkční genomiky
 - Využití genomiky ve šlechtění rostlin
 - Individualizovaná medicína
 - Pokročilé biotechnologie
 - Význam a bezpečnost GMO



Literatura

- Zdrojová literatura ke kapitole I:
 - Plant Functional Genomics, ed. Erich Grotewold, 2003, Humana Press, Totowa, New Jersey
 - Úvod do praktické bioinformatiky, Fatima Cvrčková, 2006, Academia, Praha

GENOMIKA-co to je?

- V širším pojetí-zkoumá **STRUKTURU** a **FUNKCI** genomů
 - Předpokladem je znalost genomu (sekvencí)-práce s databázemi
- V užším pojetí zkoumá **FUNKCI** jednotlivých genů - **FUNKČNÍ GENOMIKA**
 - používá zejména přístupy REVERZNÍ GENETIKY

GENOMIKA-co to je?

role BIOINFORMATIKY ve FUNKČNÍ GENOMICE

Přístupy „klasické“ genetiky



3

:

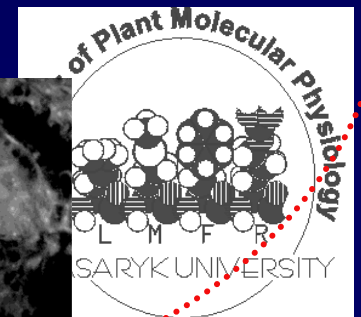
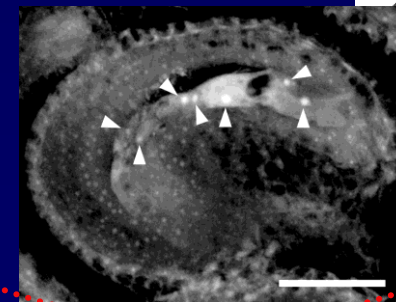
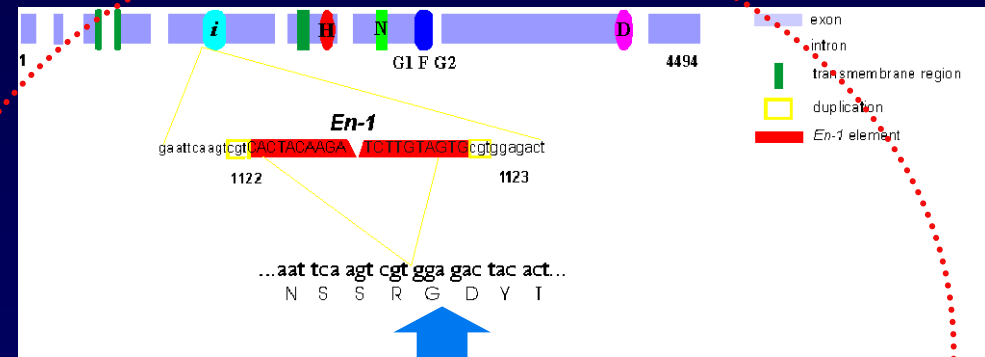
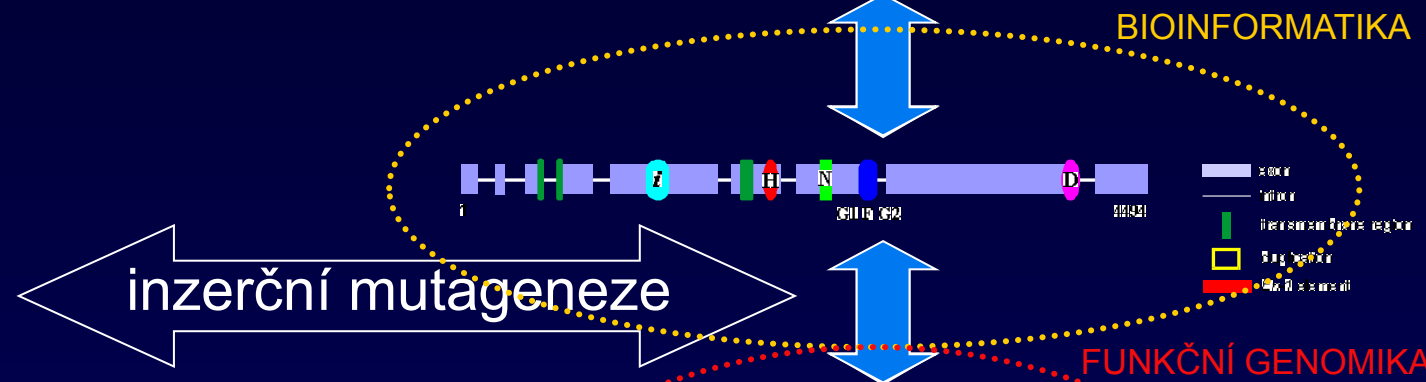
1



?

„Reverzně genetický“ přístup

5'TTATATATATATATTAATAATAATAATA
GAACAAAAAGAAAATAAATA...3'





Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů

Spektrum on-line zdrojů

EMBNet National Nodes

| | | |
|------------------|-------------|---|
| Vienna Biocenter | Austria | http://www.at.embnet.org/ |
| BEN | Belgium | http://www.be.embnet.org/ |
| BioBase | Denmark | http://biobase.dk/ |
| CSC | Finland | http://www.fi.embnet.org/ |
| INFOBIOGEN | France | http://www.infobiogen.fr/ |
| GENIUSnet | Germany | http://genome.dkfz-heidelberg.de/biounit/ |
| IMBB | Greece | http://www.imbb.forth.gr/ |
| HEN | Hungary | http://www.hu.embnet.org/ |
| INCBI | Ireland | http://acer.gen.tcd.ie/ |
| INN | Israel | http://dapsas.weizmann.ac.il/bcd/inn.html |
| IEN-ADR | Italy | http://bio-www.ba.cnr.it:8000/BioWWW/Bio-WWW.htm |
| CAOS/CAMM | Netherlands | http://www.caos.kun.nl/ |
| Bio | Norway | http://www.no.embnet.org/ |
| IBB | Poland | http://www.ibb.waw.pl/ |
| IGC | Portugal | http://www.igc.gulbenkian.pt/ |
| GeneBee | Russia | http://www.genebee.msu.su/ |
| CNB-CSIC | Spain | http://www.es.embnet.org/ |
| BMC | Sweden | http://www.embnet.se/ |
| SIB | Switzerland | http://www.ch.embnet.org/ |
| SEQNET | UK | http://www.seqnet.dl.ac.uk/ |

EMBNet Specialist Nodes

| | | |
|---------------------|-------------|---|
| MIPS | Germany | http://www.mips.biochem.mpg.de/ |
| ICGEB | Italy | http://www.icgeb.trieste.it/ |
| Pharmacia Upjohn | Sweden | http://www.pnu.com/ |
| F.Hoffmann-La Roche | Switzerland | http://www.roche.com/ |
| EBI | UK | http://www.ebi.ac.uk/ |
| HGMP-RC | UK | http://www.hgmp.mrc.ac.uk/ |
| Sanger | UK | http://www.sanger.ac.uk/ |
| UMBER | UK | http://www.bioinf.man.ac.uk/dbbrowser |

EMBNet Associate Nodes

| | | |
|-------|--------------|---|
| IBBM | Argentina | http://sol.biol.unlp.edu.ar/embnet |
| ANGIS | Australia | http://www.angis.su.oz.au/ |
| CBI | China | http://www.cbi.pku.edu.cn/ |
| CIGB | Cuba | http://bio.cigb.edu.cu/ |
| CDFO | India | http://salarjung.embnet.org.in/ |
| SANBI | South Africa | http://www.sanbi.ac.za |

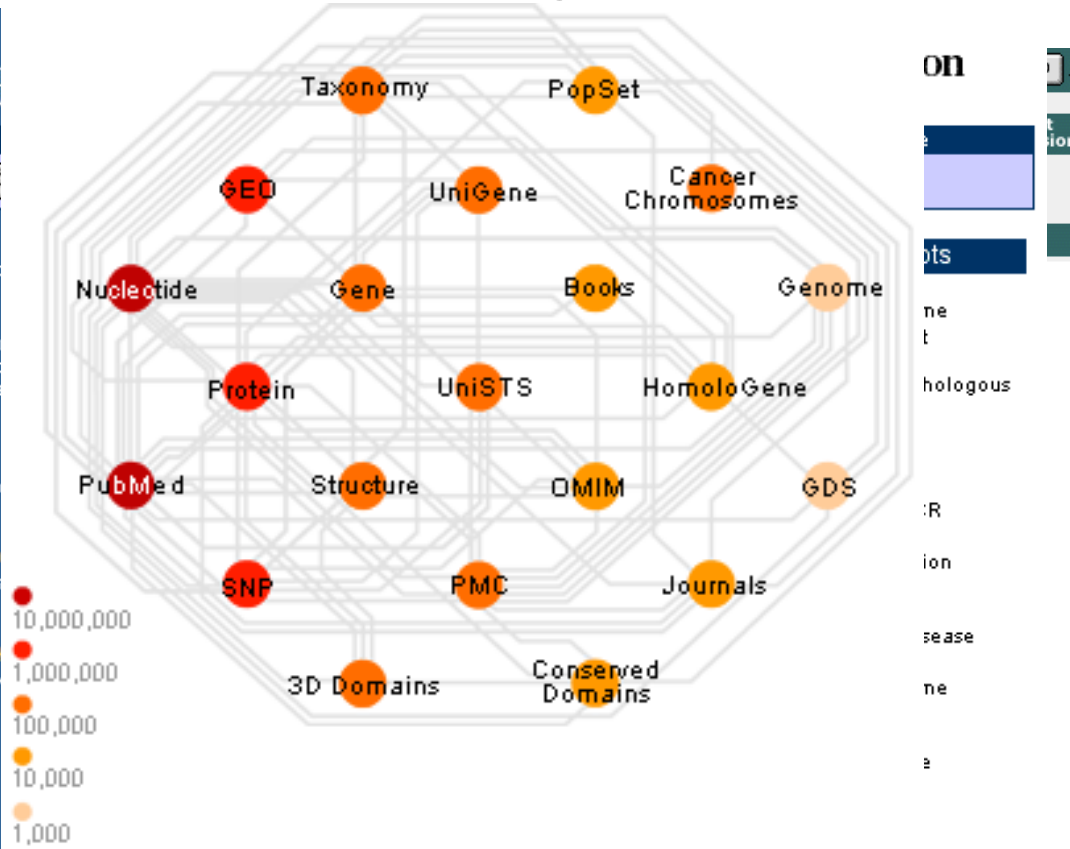
USA Information Providers

| | | |
|------|-----|---|
| NCBI | USA | http://www.ncbi.nlm.nih.gov/ |
| NLM | USA | http://www.nlm.nih.gov/ |
| NIH | USA | http://www.nih.gov/ |

Spektrum on-line zdrojů

EBI <http://www.ebi.ac.uk/services>

NCBI
PubMed
Search Nucleotide
SITE MAP
Guide to NCBI resources
About NCBI
The science behind our resources. An introduction for researchers, educators and the public.
GenBank
Sequence submission, support and software
Molecular databases
Sequences, structures, taxonomy
Literature databases
PubMed, OMIM, Books and PubMed Central
Genomic biology
The human genome, whole genomes and related resources





Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze

Primární databáze

- zahrnují soubory primárních dat – sekvencí DNA a proteinů
 - Sekvence v databázích tzv. „Velké trojky“:
 - EMBL, <http://www.ebi.ac.uk/embl/>
 - GenBank, <http://www.ncbi.nih.gov/Genbank/GenbankSearch.html>
 - DDBJ, <http://www.ddbj.nig.ac.jp>
 - denně vzájemná výměna a zálohování dat
 - velká datová náročnost (kapacita i software)
 - září 2003 27,2 x 10⁶ záznamů o zhruba 33 x 10⁹ bp
 - srpen 2005 100 x 10⁹ bp ze 165.000 organizmů

Primární databáze

- zahrnují soubory primárních dat – sekvencí DNA a proteinů
 - Proteinové sekvence:
 - PIR, <http://pir.georgetown.edu/>
 - MIPS, <http://www.mips.biochem.mpg.de>
 - SWISS-PROT, <http://www.expasy.org/sprot/>

Primární databáze

- Typy sekvencí v primárních databázích
 - standardní nukleotidové sekvence získané kvalitním sekvencováním
 - **ESTs (Expressed Sequence Tags)**
 - **HGTS (High Throughput Genome Sequencing)**
 - neanotované „surové“ výsledky sekvenačních projektů
 - referenční sekvence anotovaných genomů
 - **TPAs (Third Party Annotation)**
 - sekvence anotované jinými než původními autory

Primární databáze

GenBank (NCBI) <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI website interface. At the top, the NCBI logo is on the left, and the text "National Center for Biotechnology Information" is in the center, with "National Library of Medicine" and "National Institutes of Health" below it. A navigation bar contains links for "PubMed", "Entrez", "BLAST", "OMIM", "Books", "TaxBrowser", and "Structure". Below this is a search bar with a dropdown menu set to "Nucleotide" and a "Go" button. The main content area is divided into several sections: "SITE MAP" (Guide to NCBI resources), "About NCBI" (The science behind our resources), "GenBank" (Sequence submission support and software), "Molecular databases" (Sequences, structures and taxonomy), "Literature databases" (PubMed, OMIM, Books and PubMed Central), and "Genomic biology" (The human genome, whole genomes and related resources). A central section titled "What does NCBI do?" describes the center's mission. To the right, a "Hot Spots" section lists various projects like "Cancer genome anatomy project" and "Clusters of orthologous groups". A "Mouse Genome" section features a mouse illustration and links to "Map Viewer", "Sequencing Progress", and "Human-Mouse Homology". At the bottom, a "BLink" section promotes fast results.

Sekundární databáze

- databáze funkčních nebo strukturálních *motivů* získaných srovnáním primárních dat (sekvencí)

PROSITE, <http://www.expasy.org/prosite/> browser/PRINTS/

>PDOC50109 PS50109 HIS_KIN Histidine kinase domain [profile].

```
402 - 671 NASHDIRGALAGMKGLIDICRDGVKPGSDVDTTINQVMVCAKDLVALLNSVLEMSKIRSG
KMQLVREDFNLSKLLLEDVIDFYHPVAMKKGVVLDPHDgsvEKFSNVRGDSGRLKQILN
NLVSNARVFTVD - GHIIVRAWAQrpgnssavvlasyppkgsakfvksmfcknkeesatye
teisnsirnnanTMEFVFEVDDTGKGIHMEHREKSVPRNYVQVREtAQGHQGTGLGLGIIVQ
SLVRLMGGEIRITDKAMGeKGTCPQPNVLLTT
```

>PDOC50110 PS50110 RESPONSE_REGULATORY Response regulatory domain [profile].

```
987 - 1085 RVLVDDNFIQRKVTGKLNKMGVSeVEQCDSGKEALRLVTEGLtqreeggsvdKlpFDY
IFMDQMPEMDGYRATREIRkvekSYGVRTPIIAVSGHD-----
```

Graphical summary of hits (java applet)



98 hits with 12 PROSITE entries

[Expasy Home page](#) | [Site Map](#) | [Search ExPASy](#) | [Contact us](#) | [Swiss-Prot](#) | [PROSITE](#) | [Proteomics tools](#)

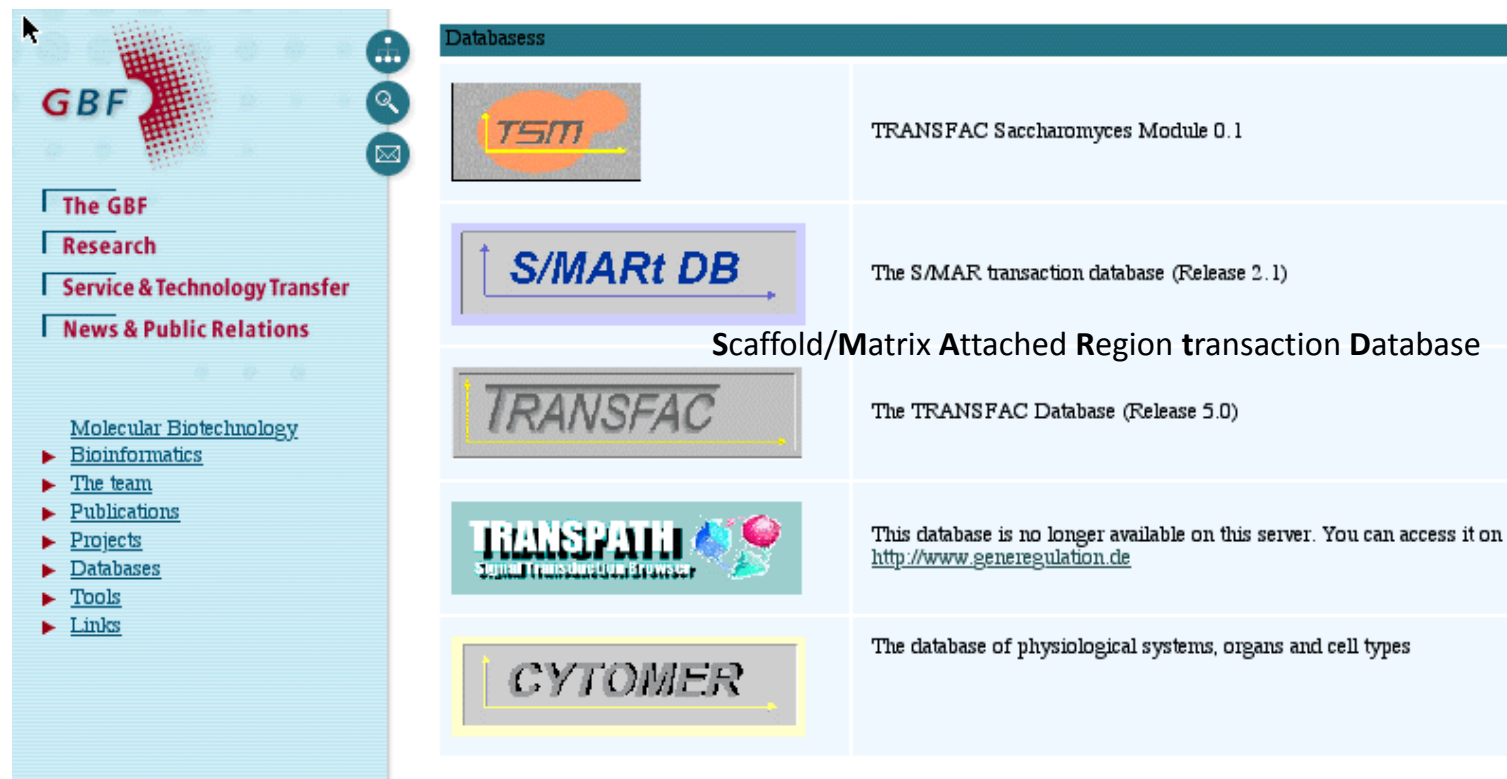
plain text output

START THE SCAN | RESET





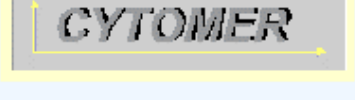
match mode: (for patterns, see [help](#))
randomize databases: (to test a pattern, see [help](#))

Sekundární databáze

- TRANSFAC <http://www.gene-regulation.com/>

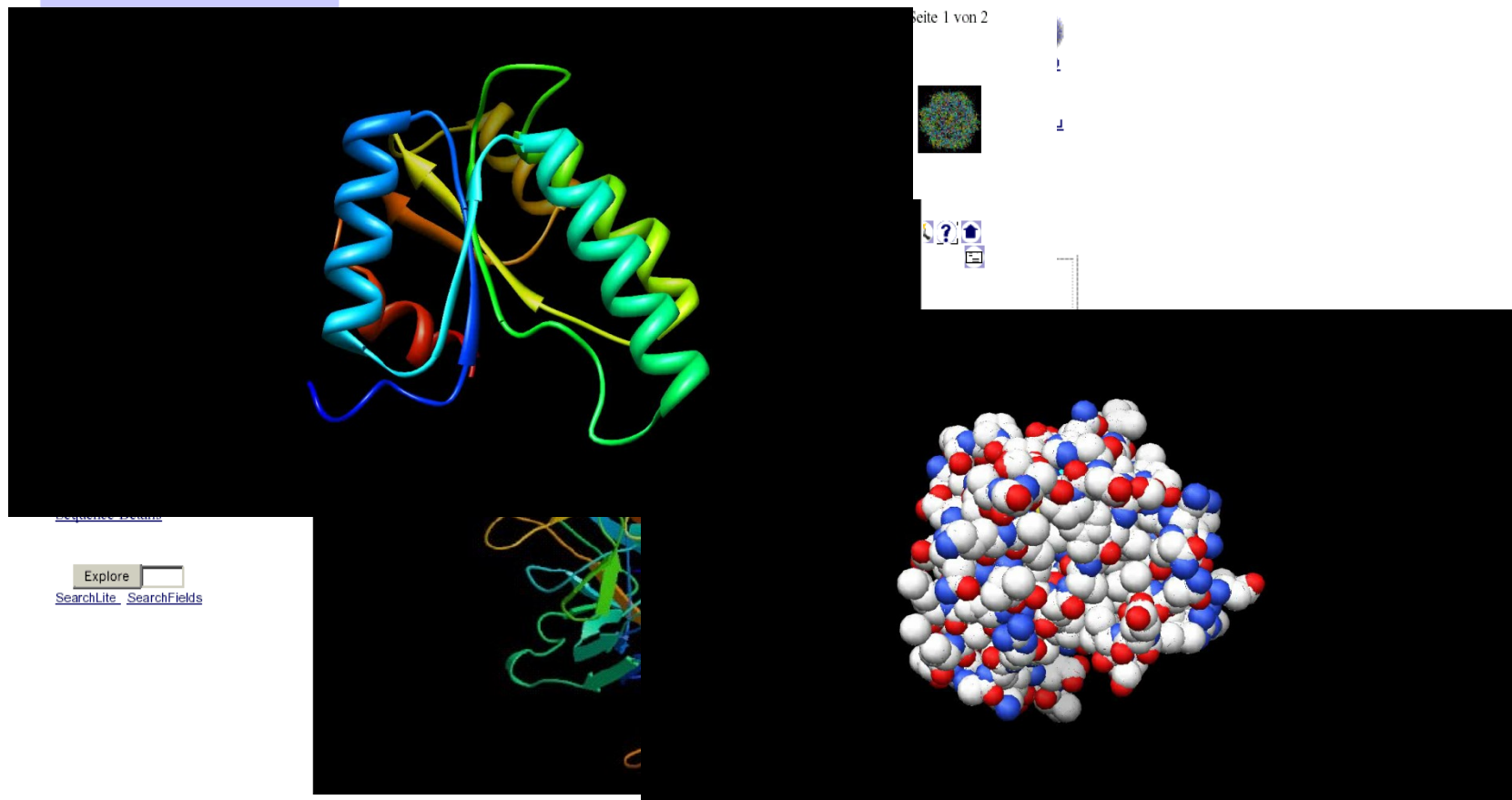


The screenshot shows the GBF website interface. On the left is a navigation menu with the GBF logo and links for 'The GBF', 'Research', 'Service & Technology Transfer', and 'News & Public Relations'. Below these are links for 'Molecular Biotechnology', 'Bioinformatics', 'The team', 'Publications', 'Projects', 'Databases', 'Tools', and 'Links'. The main content area is titled 'Databases' and contains a table of database entries.

| Databases | |
|--|---|
|  | TRANSFAC Saccharomyces Module 0.1 |
|  | The S/MAR transaction database (Release 2.1) Scaffold/Matrix Attached Region transaction Database |
|  | The TRANSFAC Database (Release 5.0) |
|  | This database is no longer available on this server. You can access it on http://www.generegulation.de |
|  | The database of physiological systems, organs and cell types |

Strukturální databáze

- PDB <http://www.rcsb.org/pdb/>





Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje

Genomové zdroje

NCBI
National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search for

SITE MAP
Guide to NCBI resources

About NCBI
The science behind our resources. An introduction for researchers, educators and the public.

GenBank
Sequence submission support and software

Molecular databases
Sequences, structures and taxonomy

Literature databases
PubMed, OMIM, Books and PubMed Central

Genomic biology
The human genome, whole genomes and related resources

What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots

- ▶ Cancer genome anatomy project
- ▶ Clusters of orthologous groups
- ▶ Coffee Break
- ▶ Electronic PCR
- ▶ Gene expression omnibus
- ▶ Genes and disease
- ▶ Human genome resources
- ▶ Human/mouse homology maps
- ▶ LocusLink

Mouse Genome
Resources: explore tools for manipulating the mouse genome.

Try these: Map Viewer Sequencing Progress Human-Mouse Homology

BLink and get results fast!

Genomové zdroje

NCBI Entrez Genomes

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Help

Search for on chromosome(s)

Show linked entries [Help](#) [FTP](#)

Entrez Genomes
MapViewer Home

Prominent organisms

FTP SITE

Related Databases:
TAIR
TIGR
MIPS
KAOS

Sequencing Projects:
SPP Consortium
CSH / WashU
TIGR
Kazusa
ESSA
Genoscope

Arabidopsis thaliana genome view [BLAST search Arabidopsis genome](#)

I **II** **III** **IV** **V** **MT** **CHL**

Lineage: [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Embryophyta](#); [Tracheophyta](#); [Spermatophyta](#); [Magnoliophyta](#); [eudicotyledons](#); [core eudicots](#); [Fosidae](#); [eurosids II](#); [Brassicales](#); [Brassicaceae](#); [Arabidopsis](#)

Arabidopsis thaliana is a small flowering plant that is widely used as a model organism in plant biology. Arabidopsis is a member of the mustard (Brassicaceae) family, which includes cultivated species such as cabbage and radish. Arabidopsis is not of major agronomic significance, but it offers important advantages for basic research in genetics and molecular biology. Its genome has been sequenced by an international collaboration collectively termed the [Arabidopsis Genome Initiative \(AGI\)](#) ([The Arabidopsis Genome Initiative, 2000, Nature, 408:796-815](#)).

This sequence, map, and annotations are the result of a collaboration between [TIGR](#), [MIPS](#), and [TAIR](#). The non-redundant sequence of the chromosomes (pseudomolecules) and their annotations were provided to NCBI by TIGR on behalf of the collaborators.

Genomové zdroje

NCBI Arabidopsis thaliana Map View

Chromosome: **I** | II | III | IV | V

Master Map: Gene **Maps & Options**

Total Genes On Chromosome: **7173**

Region Displayed: **0-30M bp** [Download/View Sequence/Evidence](#)

Genes Labeled: **20** Total Genes in Region: **7173**

| Gene | Accession | Function |
|-----------|------------|--|
| Atlg05180 | YUP8H12.21 | <input type="checkbox"/> TIGR MIPS TAIR NM_100396 NP_172010 auxin-resistance protein AXR1 |
| Atlg08210 | T23G18.7 | <input type="checkbox"/> TIGR MIPS TAIR NM_100695 NP_563808 expressed protein |
| Atlg11250 | T28P6.10 | <input type="checkbox"/> TIGR MIPS TAIR NM_100997 NP_172591 syntaxin-related protein At-SYR1, putative |
| Atlg14670 | T5E21.15 | <input type="checkbox"/> TIGR MIPS TAIR NM_101334 NP_172919 endomembrane protein, putative |
| Atlg17790 | F2H15.2 | <input type="checkbox"/> TIGR MIPS TAIR NM_101642 NP_564037 expressed protein |
| Atlg21050 | T22I11.13 | <input type="checkbox"/> TIGR MIPS TAIR NM_101958 NP_564130 expressed protein |
| Atlg24210 | F3I6.14 | <input type="checkbox"/> TIGR MIPS TAIR NM_102267 NP_564212 expressed protein |
| Atlg28370 | F3M18.20 | <input type="checkbox"/> TIGR MIPS TAIR NM_102603 NP_174159 ethylene-responsive element binding factor, putative |
| Atlg31885 | F5M6.28 | <input type="checkbox"/> TIGR MIPS TAIR NM_102926 NP_174472 major intrinsic protein, putative |
| Atlg35670 | F15O4.8 | <input type="checkbox"/> TIGR MIPS TAIR NM_103271 NP_174807 calcium-dependent protein kinase |
| Atlg48160 | F21D18.11 | <input type="checkbox"/> TIGR MIPS TAIR NM_103712 NP_175250 signal recognition particle 19 kDa protein subunit, putative |
| Atlg51980 | F5F19.4 | <input type="checkbox"/> TIGR MIPS TAIR NM_104079 NP_175610 mitochondrial processing peptidase alpha subunit, putative |
| Atlg55150 | T7N22.9 | <input type="checkbox"/> TIGR MIPS TAIR NM_104388 NP_175911 ethylene-responsive RNA helicase, putative |
| Atlg60140 | T13D8.4 | <input type="checkbox"/> TIGR MIPS TAIR NM_104705 NP_176221 trehalose-6-phosphate synthase, putative |
| Atlg63750 | F24D7.6 | <input type="checkbox"/> TIGR MIPS TAIR NM_105052 NP_176562 putative disease resistance protein |
| Atlg67090 | F5A8.1 | <input type="checkbox"/> TIGR MIPS TAIR NM_105379 NP_176880 ribulose-bisphosphate carboxylase small unit, putative |
| Atlg69870 | T17F3.10 | <input type="checkbox"/> TIGR MIPS TAIR NM_105655 NP_564979 putative peptide transporter |
| Atlg72970 | F3N23.17 | <input type="checkbox"/> TIGR MIPS TAIR NM_105955 NP_565050 expressed protein |
| Atlg75780 | T4O12.1 | <input type="checkbox"/> TIGR MIPS TAIR NM_106228 NP_177706 tubulin beta-1 chain |
| Atlg78700 | F9K20.26 | <input type="checkbox"/> TIGR MIPS TAIR NM_106517 NP_565187 expressed protein |

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Genomové zdroje

NCBI Nucleotide search interface. Search for Nucleotide. History has expired due to inactivity. Display: 1. Send to: File. Get Subsequence.

1: NC_003070. Arabidopsis thali...[gi:22330780]

Reverse Complement Strand. View on plus strand. Protein coding genes. Hide Toolbar.

Search for gene: Find. CDS with gene and mRNA. Hide sequence. Refresh.

30028659M 20M 10M 5' 3'.

At1g05200 At1g05190 At1g05180 At1g05170.

Legend: - CDS - RNA - gene - sequence fragment shown.

Sequence:

```
1501830 CTTTTTGTTC ATCAGTTCAC CGGAGCCAAA ATCGCTCTCT CGCTTGAGCT GCGAGATGC At1g05180 mRNA-protein id: CDS M
1501778 AAGCAGTAAA AAGATCCAGG AGCATGTTG AAGAGAGCC AACAATGCTA GACCTAARA At1g05180 mRNA-protein id: CDS Q A V K R S R R H V E E E P T M V E P K
1501710 CCRACTACGA TCGTCAGCTC AGGTATACAT ACTCTTTTC CTTAACTCT ACTTCGACT At1g05180 mRNA-protein id: CDS T K Y D R Q L R
1501650 ACTCTGTCC GAGCGAARAT TGATGAGTA GTAATCCTG GGTGACTGA TTAGATTTC At1g05180 mRNA-protein id: CDS I W
1501590 GGGGAGCTA GGTCAAGCG CTTGAGAGA ACCGAGTATC TGTTTACTA ATTGTGCCCC At1g05180 mRNA-protein id: CDS G E V G Q A A L E E A S I C L L N C G P
```


Genomové zdroje

- Human Genome Browser <http://genome.ucsc.edu/cgi-bin/hgGateway>

UCSC Genome Bioinformatics

[Genome Browser](#) - [Family Browser](#) - [Blat](#) - [Table Browser](#) - [FAQ](#) - [Help](#)

[Genome Browser](#)

About the UCSC Genome Bioinformatics Site

This site contains the reference sequence for the human and *C. elegans* genomes and working drafts for the mouse, rat, Fugu, *Drosophila*, *C. briggsae*, and SARS genomes. It also contains the CFTR (cystic fibrosis) region in 13 species.

[Family Browser](#)

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Family Browser shows expression, homology and other information on groups of genes that can be related in many ways. The Table Browser provides convenient access to the underlying database. Blat quickly maps your sequence to the genome.

[Blat](#)

[Tables](#)

[Downloads](#)

News

[News Archives](#) ▶

[Release Log](#)

22 Dec. 2003 - [Advance Warning of Browser Outage 30 Dec. 2003](#)

[Custom Tracks](#)

Electrical power to the UCSC School of Engineering building will be shut down on 30 December from 7 a.m. to 3 p.m. PST. The Genome Browser and Blat servers will not be available during this time period.

[Mirrors](#)

[Archives](#)

10 Dec. 2003 - [Human/Chimp Alignment Tracks Released](#)

[Credits](#)

UCSC has released alignments of the Nov. 2003 chimpanzee draft assembly to the July 2003 human assembly in the Genome Browser. These alignments may be viewed on the [Human July 2003](#) assembly. This release coincides with today's [announcement](#) by the National Human Genome Research Institute (NHGRI) of the first draft assembly of the chimpanzee genome.

[Pubs](#)

[Cite Us](#)

The set of human/chimpanzee alignments consists of a reciprocal best-in-genome net track and a chimp chain track. These alignments were generated using the blastz program developed at Pennsylvania State University and the programs blat, actChain, chainNet, and netSyntenic developed at UCSC by Jim Kent. Research scientists should find these tracks useful for locating orthologous regions and studying genome rearrangement in the two species.

[Licenses](#)

[Jobs](#)

For more information about the alignment tracks, refer to the track description pages. The tables may be downloaded from the Genome Browser FTP server's [hg16 database directory](#). The chimp sequence and alignment data are downloadable from the [hg16 human/chimp alignments directory](#).

[Contact Us](#)

The chimp sequence used in these alignments was obtained from the 13 Nov. 2003 Arachne assembly. We'd like to thank NHGRI, the Eli & Edythe L. Broad Institute at MIT/Harvard, and Washington University School of Medicine for providing this sequence, and LaDeana Hillier, Washington University School of Medicine, and the Whitehead Institute for their work on the alignments. We'd also like to acknowledge the members of the UCSC team who contributed to the release of these alignments in the Genome Browser: Jim Kent, Kate Rosenbloom, Heather Trumbower, and Donna Karolchik.

24 Nov. 2003 - [Build 32 Mouse Assembly \(Oct. 2003\) Released](#): We have released a Genome Browser and Blat server for the latest mouse genome assembly, NCBI Build 32 (UCSC v. mm4). [Read more.](#)

31 Oct. 2003 - [V. 3.1 D. melanogaster Assembly \(Jan. 2003\) Released](#): We have added the *Drosophila melanogaster* (fruitfly) assembly to



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genomové zdroje

- TAIR, The Arabidopsis Information Resource, <http://www.arabidopsis.org>



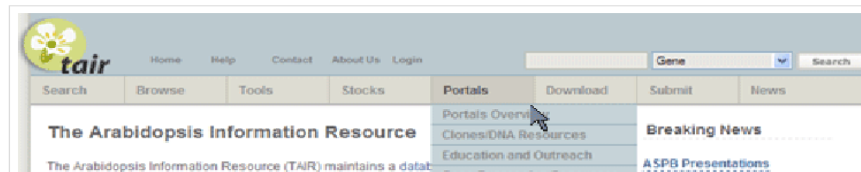
The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a [database](#) of genetic and [molecular biology data](#) for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new and updated genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The [Arabidopsis Biological Resource Center](#) at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of *Arabidopsis thaliana* and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

The NEW arabidopsis.org

We've added new dropdown headers and left navigation bars and reorganized our web pages to make it easier to locate information and resources in TAIR. Please contact us if you experience any problems with our new site.



Breaking News

Data Updates Suspended

[October 19, 2006]
Some TAIR data updates, including loading of new ABRC stocks, will be suspended from Oct 20-Nov 17 while we move our servers.

New Phenotype Search Option

[October 15, 2006]
Search for [genes](#), [germplasms](#), and [polymorphisms](#) using associated phenotype, and see improved phenotype data display in results and detail pages.

ASPB Presentations

[August 15, 2006]
Following heavy demand, the TAIR workshop presentations given at the ASPB meeting in Boston have been made available from the TAIR website for download.

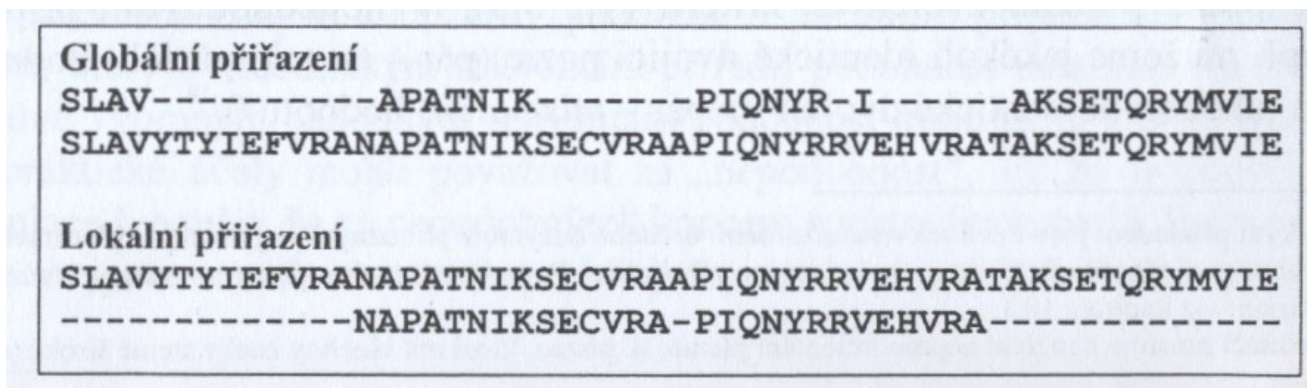


Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje
- Analytické nástroje
 - Vyhledávání homologií

Analytické nástroje

□ Globální vs. lokální přiřazení

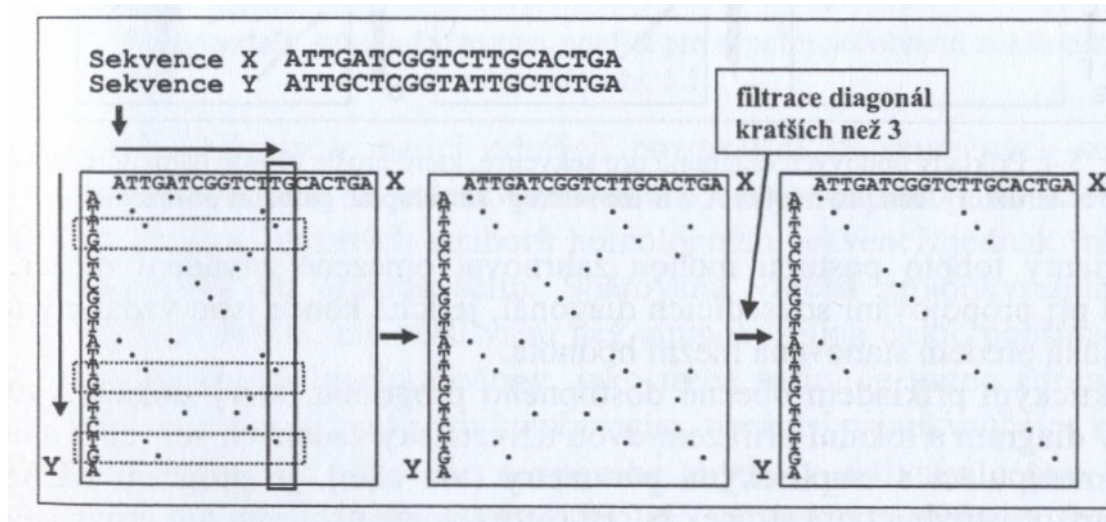


Cvrčková, Úvod do praktické bioinformatiky

- globální přiřazení pouze u sekvencí, které jsou si podobné (za cenu vnášení mezer do jedné nebo obou sekvencí)
- globální přiřazení se používá především v případě mnohačetného přiřazování (CLUSTALW, viz dále)
- lokální přiřazení umožní identifikaci a srovnání i v případě porovnávání pouze **úseků sekvencí** s významnou mírou podobnosti, např. i při záměně pořadí proteinových domén během evoluce

Analytické nástroje

- Volba správného typu přiřazení pomocí bodového diagramu (dotplot)

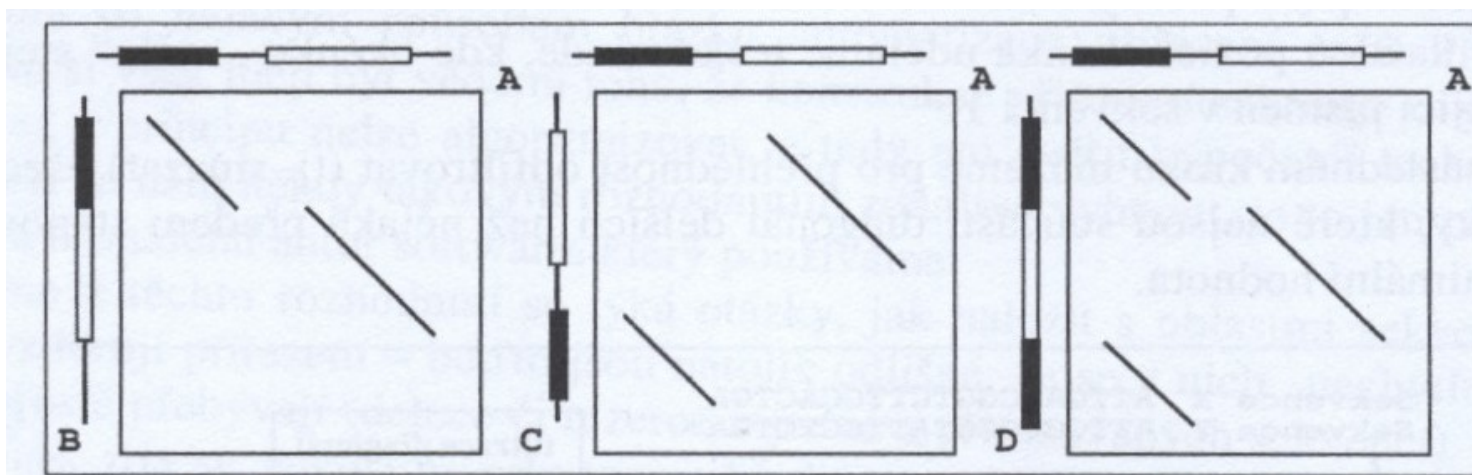


Cvrčková, Úvod do praktické bioinformatiky

- vynesení sekvencí proti sobě
- identifikace shody v okně o dané velikosti (např. 2 bp)
- „odfiltrování“ diagonál o délce menší než je mezní hodnota (threshold)

Analytické nástroje

- příklady srovnání sekvencí pomocí bodového diagramu



Cvrčková, Úvod do praktické bioinformatiky

- globálně lze srovnávat pouze sekvence A, B
- ostatní sekvence prošly během evoluce záměnou domén a je nutné je porovnávat lokálně
- bodový diagram lze získat pomocí srovnávacího programu BLAST2 (viz dále)

Analytické nástroje

- BLAST <http://ncbi.nlm.nih.gov/BLAST/>

NCBI *nucleotide-nucleotide* **BLAST**
Nucleotide Protein Translations Retrieve results for an RID

[Search](#)

```
aacccacccgc  
acaccatcat cattatcacc atcgtttttg ggcgatggtg tgtgggtcca  
gogtattaat  
ataattaatt tattccacat gagatatgat atgatatact atgtattttt  
tgtttttttt  
ttatttgtaa acctttaata taacaagaac tacaaaaaat gaaaa
```

[Set subsequence](#) From: To:

[Choose database](#)

Now: **BLAST!** or

BLAST

Basic Local Alignment Search Tool

- Velikost vyhledávacího slova (word size): 10-11 bp, resp. 2-3 aa
 - Primární podobnosti (seed matches)
 - Rozšiřování oblasti homologie doprava i doleva
- Hodnocení homologie pomocí matice PAM (Point Accepted Mutation) nebo BLOSUM (BLOcks Substitution Matrix)
- Zobrazení výsledků

| | A | T | G | C |
|---|---|---|---|---|
| A | 1 | 0 | 0 | 0 |
| T | 0 | 1 | 0 | 0 |
| G | 0 | 0 | 1 | 0 |
| C | 0 | 0 | 0 | 1 |

hodnota nepáru G-A

hodnota páru G-G

Cvrčková, Úvod do praktické bioinformatiky

Matrice PAM 250

| C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|----|
| 12 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -2 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -3 | 1 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -2 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -3 | 1 | 0 | -1 | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -4 | 1 | 0 | -1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5 | 0 | 0 | -1 | 0 | 1 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5 | 0 | 0 | -1 | 0 | 0 | 1 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5 | -1 | -1 | 0 | 0 | -1 | 1 | 2 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -3 | -1 | -1 | 0 | -1 | -2 | 2 | 1 | 1 | 3 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -4 | 0 | -1 | 0 | -2 | -3 | 0 | -1 | -1 | 1 | 2 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5 | 0 | 0 | -1 | -1 | -2 | 1 | 0 | 0 | 1 | 0 | 3 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5 | -2 | -1 | -2 | -1 | -3 | -2 | -3 | -2 | -1 | -2 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| -2 | -1 | 0 | -2 | -1 | -3 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2 | 5 | 0 | 0 | 0 | 0 | 0 |
| -6 | -3 | -2 | -3 | -2 | -4 | -3 | -4 | -3 | -2 | -2 | -3 | -3 | 4 | 2 | 6 | 0 | 0 | 0 | 0 |
| -2 | -1 | 0 | -1 | 0 | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2 | 4 | 2 | 4 | 0 | 0 | 0 |
| -4 | -3 | -3 | -5 | -4 | -5 | -4 | -6 | -5 | -5 | -2 | -4 | -5 | 0 | 1 | 2 | -1 | 9 | 0 | 0 |
| 0 | -3 | -3 | -5 | -3 | -5 | -2 | -4 | -4 | -4 | 0 | -4 | -4 | -2 | -1 | -1 | -2 | 7 | 10 | 0 |
| -8 | -2 | -5 | -6 | -6 | -7 | -4 | -7 | -7 | -5 | -3 | 2 | -3 | -4 | -5 | -2 | -6 | 0 | 0 | 17 |
| C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |

BLAST

Basic Local Alignment Search Tool



- „expectancy value“ udává předpokládaný počet sekvencí se stejnou nebo lepší podobností při vyhledávání ve stejně velké databázi složené z náhodných sekvencí
- výsledek udává frakci totožných a u proteinů i podobných pozic, příp. počet vložených mezer

BLAST

Specializované verze

- V současnosti existuje celá řada specializovaných verzí programu BLAST
 - vyhledávání podle zdroje (organismu) sekvencí, např. známých genomů mikroorganismů
 - **BLASTP**
 - vyhledávání podobnosti k proteinu v databázi proteinových sekvencí
 - **BLASTN**
 - vyhledávání podobnosti k nukleotidové sekvenci v databázi nukleotidových sekvencí
 - další varianty jako např. MEGABLAST pro identifikaci totožných nebo velice podobných sekvencí (vyhledává dlouhé podobné úseky nukl. sekvencí)
 - **BLASTX**
 - vyhledávání podobnosti k proteinu v databázi nukleotidových sekvencí přeložených do sekvence aa





BLAST

Specializované verze

- V současnosti existuje celá řada specializovaných verzí programu BLAST
 - **TBLASTN**
 - vyhledávání k sekvenci nukleotidů přeložené do sekvence aa v databázi proteinů
 - **TBLASTX**
 - vyhledávání k sekvenci nukleotidů přeložené do sekvence aa v databázi nukleotidových sekvencí přeložených do sekvence aa



BLAST

Specializované verze

- V současnosti existuje celá řada specializovaných verzí programu BLAST
 - **PSI-BLAST (Position-Specific Iterated Blast)**
 - Prvním krokem je standardní BLAST, při kterém PSI-BLAST identifikuje skupinu podobných sekvencí s E hodnotou lepší než minimální hodnota (standardně 0,005)
 - PSI-BLAST vytváří pro každé přiřazení tzv. PSSM (position specific substitution matrix)
 - PSSM matice zohledňuje výskyt jedné aminokyseliny ve stejné pozici se zvýšenou frekvencí u sekvencí identifikovaných jako podobné v prvním kole pomocí BLAST, což může znamenat funkční konzervovanost





BLAST

Specializované verze

- V současnosti existuje celá řada specializovaných verzí programu BLAST
 - **PHI-BLAST (Pattern-Hit InitiatedBlast)**
 - Určen k identifikaci specifické sekvence, např. motivu (pattern) v sekvenci podobných proteinových sekvencí
 - Sekvenci motivu je třeba vložit pomocí speciálního syntaxu
 - [LVIMF] znamená buď Leu, Val, Ile, Met nebo Phe
 - - je oddělovník (neznamená nic)
 - x(5) znamená 5 jakýchkoliv aminokyselin
 - x(3, 5) znamená 3 až 5 jakýchkoliv aminokyselin

BLAST

Specializované verze

□ Příklad vyhledávání pomocí PHI-BLAST

```
>gi|4758958|ref|NP_004148.1| Human cAMP-dependent protein kinase  
MSHIQIPPGLTELLQGYTVEVLRQQPPDLVEFAVEYFTRLREARAPASVLPAAATPRQSLGHPPPEPGPDR  
VADAKGDSESEEDLLEVPVPSRFNRRVSVCAETYNPDEEEEDTDPRVIHPKTDEQRCRLQEACKDILLF  
KNLDQEQLSQVLDAMFERIVKADEHVIDQGDDGDNFYVIERGTYDILVTKDNQTRSVGQYDNRGS  
TSEGSLWGLDRVTFRRIIVKNNAKKRKMFESFIESVPLLKSLEVSERMKIVDVIgek  
IYKDGERIITQGEKADSFYIESGEVSIILRSRTKSNKDGGNQEVEIARCHKGQYFGELALVTNKPRAAS  
AYAVGDVKCLVMDVQAFERLLGPCMDIMKRNI SHYEEQLVKMFGSSVDLGNLQ
```

```
[LIVMF] -G-E-x- [GAS] - [LIVM] -x(5,11) -R- [STAQ] -A-x- [LIVMA] -x- [STACV] .
```



Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje
- Analytické nástroje
 - Vyhledávání homologií
 - Vyhledávání sekvenčních motivů, otevřených čtecích rámců, restričních míst....

Analytické nástroje

- <http://workbench.sdsc.edu/>

Biology WorkBench
click here to toggle between menus and buttons
WE Moved! <http://workbench.sdsc.edu/>
Version 3.2

Session Tools Protein Tools **Nucleic Tools** Alignment Tools Structure Tools (Alpha)

beta-glucosidase

GBPLN:804655 **Hordeum vulgare L. beta-glucosidase (BGQ60) gene, complete cds.**
 GBPLN:170248 **Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds.**

Select All Deselect All Ndjinn BATCH Add Edit Delete Copy View Download ViewRecords
BL2SEQ BL2SEQX BLASTN BLASTX TBLASTX FASTA FASTX FASTY SSEARCH CLUSTALW
CLUSTALWPROF ALIGN LALIGN LFASTA PATTERNMATCHDB PATTERNMATCH TACG PRIMER3
NASTATS BESTSCOR PFSCAN PRIMERCHECK PRIMERTM SIXFRAME REVCOMP RANDSEQ

Copyright (C) 1999, Board of Trustees of the University of Illinois.
SDSC

Analytické nástroje

- <http://workbench.sdsc.edu/>

View
View Nucleic Sequence(s)

Format Case

[Download/view all sequences in text format](#)

[\[NEXT\]](#) [\[BOTTOM\]](#)

Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds.
GBPLN:170248, 4699 bp

> 170248
GAGCTCCCTTGGGGGGCAAGGGCAAAACTTTTTGCTAAATGGAAAAATATTATACCAAGTGTGTAATA
GTTACTCAATTTGAATTAACAAAGGGGCAAAATTTGACTATTTTGCCCTTATATCTTTTGGTCACAAAAAC
ATAAAATATCCCATCCGAAATTCCAAATGGTCCATTATCGGCCAAGTAGCTTTCTTTAATTATAGTTAGTT
GACAAAACACTATCAAGATATCATTATTATAAATAAATTCAAAGTCCATCATCTTAGCTGCCTCCTCA
GTAGAGCCGCCAGTAAATAAAGACCGATCAAAATAAAGCCGCCATTAAAAAATGAATTTTAGGACTCTC
GATTGGCACGTAAGTGCCAAAACCTTTCCAATACTTTGCTGCAACTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATTTCTAAGTTTTATCTCTAATTTACATCTCAACTAATATTAAGAAATTAACAGGTA
CAGCAAAATCATAAAATTTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCTTTTTCAGAG
TCTGCATGCCATATTCCTAAGGGGTCGTTTGGTACAAGAAATAAATAAATAAATTTTCGGGATAGAATTT
GAGATTGCATTTATCTTGTGTTTTAATTATAAGTATTAGCTAATTTTCAGAAATAAATTTTACTAAAATAG
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC
TTATTTATCTCACTATTTTACCAAATGATCGGTTAGTCTTTCATGAGAATCCAGTATCCTCAATAAATGCA
GTAAGAAGTTAGAAAAATTTTCAITTAATCAATTCATATAAATTTAAAAAATATTAGATATGGAGCACTTAAG
ATACAATAAAAGATGTACCGTTAATAAATAAAGATAAGATAGAGTTTTTAAATAGGAAAAAAAAAACGGTT
CGAGACTCTTTATGGAAGGCGTTTCTTCAAAATGAGATTCTCATTCATTGCTCTGGTGC AATAGCAAAA
TGACATCTTACTCTTAAGATACAGCGAGCCACTCTACAATCTTCTATTGTATACTCAAATGAAAGTTTTA
GAGAATTTCAAATCTCTCAACTACTTTTTAAGGGAATTCAAAATACGACC AATATTTATTACTTACTTAC
TTATAGTTAAATGATATGAATTTTTATTTTAAATTTGAATTGAAAAATTTAAATTTACTTGAATTTAATATAA

Analytické nástroje

- <http://workbench.sdsc.edu/>

Regex pattern:

ett. {1,32}ett

0 sequences were searched

1 match was found

Matches are indicated in blue

>170248

```
GAGCTCCCTTTGGGGGGCAAGGGCAAAACTTTTGTCTAAATGGAAAAATATTATACCAAGTGTGTTGTAATA
GTTACTCAATTTGAATTAACAAAGGGGCAAAATTTGACTATTTTGGCCCTTATATCTTTTGGTCCAAAAAC
ATAAAATATCCCATCCGAAATTCCAAATGGTCCATTATCGGCAAGTAGCTTTCTTTAAATATAGTTAGTT
GACAAAACACTATCAAGATATCATTATTATAATAATAAACTTCAAGTCCATCATCTTTAGCTGCCTCCTCA
GTAGAGCCGCCAGTAAAAAAGACCAGATCAAAATAAAAGCCGCCATTAAAAATAATGAATTTTAGGACTCTC
GATTTGGCAGGTAAGTGGCAAAACTCTTCCAATACTTTTGTGCAACTTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATTTCTAAGTTTATCTCCTAATTTACATCTCAACTAATATTAAGAAATTAACAGGTA
CAGCAATCATAAAATTTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCCTTTTTCAGAG
TCTGCATGCCATATTTCACTAAGGGGTCGTTTGGTACAAGAAATAATAATAAATTTTCGGGATAGAATTT
GAGATTGCATTTATCTTGTGTTTAAATATAAGTATTAGCTAATTTTCAAGAATAAATTTTACTAAAATAG
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC
TTATTTATCTCACTATTTTACCAATGATCGGTTAGTCTTCATGAGAATCCAGTATCCTCAATAAATGCA
GTAAGAAGTTAGAAAAATTTTCAATTAATCAATTCATATAATTTAAAAATATTAGATATGGAGCACTTAAG
ATACAATAAAGATGTACCGTTAATAATAAAGATAAGATAGAGTTTAAATAGGAAAAAACAACGGTT
CGAGACACTCTTATGGAAGGCGTTGTCTTCAAAGTAGATTTCTCATTCATTGCTCTGGTGCATAGCAAAA
TGACATCTTACTCTTAAGATACAGGAGCCACTCTACAATCTTCTATTGTATACTCAAAATGAAAGTTTTA
GAGAACTTTTCAAACTCTCAACTACTTTTAAAGGGAATTCAAAAATACGACCAATATTTATTACTTAC
TTATAGTTAAATGATATGAATTTTAAATTTGAAATTTGAAATATTAAATTTACTTGAATTAATATAA
ACAAATAGATATCGCTAAGTATTTACCACAAACATGGAGATACTACAGAAGATTTTATTATTTGTAACGAT
GATTAAGCAGCTATTCATCTGGTTTGTGCAGGATGAAAGAAAGTAACTAGCTATAATTTCTTTTGTAAAGT
```

Analytické nástroje

- <http://workbench.sdsc.edu/>

Frame 1, 1 stop codon

Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds. Tran

```
>170248 Translated - Frame 1  
ELPWGARAKLFAKWKNIIIPSVCSYSI*INKGANLTILPL
```

```
      E L P W G A R A K L F A K W K N I I P S  
1    gagctcccttggggggcaagggcaaaactttttgctaaatggaaaaatattataccaagt 60  
      V C N S Y S I * I N K G A N L T I L P L  
61   gtttgaatagttactcaatttgaattaacaaaggggcaatttgactattttgcctta 120
```

Frame 2, 1 stop codon

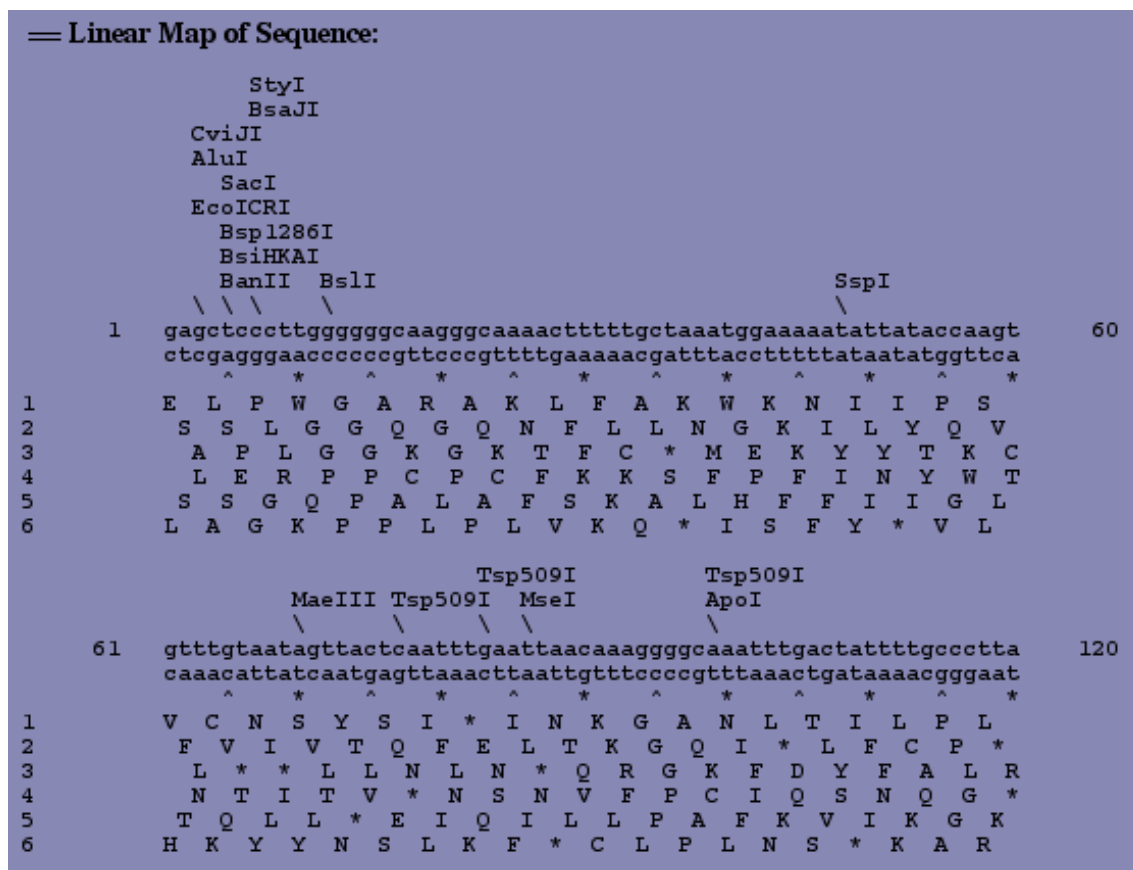
Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds. Tran

```
>170248 Translated - Frame 2  
SSLGGQGQNFLLNGKILYQVFVIVTQFELTKGQI*LFCP
```

```
      S S L G G Q G Q N F L L N G K I L Y Q V  
2    agctcccttggggggcaagggcaaaactttttgctaaatggaaaaatattataccaagtg 61  
      F V I V T Q F E L T K G Q I * L F C P  
62   tttgtaatagttactcaatttgaattaacaaaggggcaatttgactattttgcctta 120
```

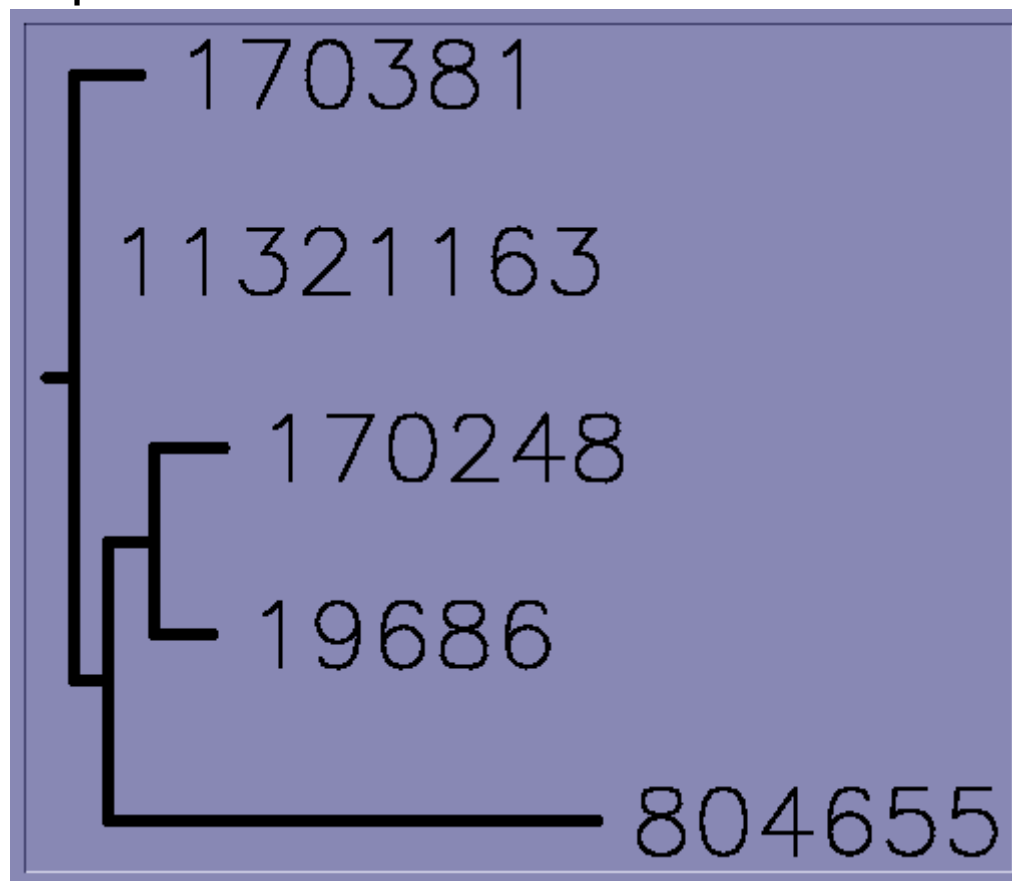
Analytické nástroje

- <http://workbench.sdsc.edu/>




Analytické nástroje

- <http://workbench.sdsc.edu/>



Analytické nástroje

- VPCR <http://grup.cribi.unipd.it/cgi-bin/mateo/vpcr2.cgi>

SEARCH  [ABOUT](#) [DOWNLOAD](#) [LINKS](#)

VPCR 2.0 (WWW interface) - Please, enter nucleotide primer sequences ([UB codes](#) allowed for degenerate primers). VPCR 2.0 searches the specified database for matches to the primers. If matches are found within 10000 bases, a PCR simulation model predicts amplification. Calculated PCR products are displayed within a minute.

NOTE: Abilities of VPCR 2.0 are still limited by BLAST capabilities and settings, as well as inability of our current software to deal with more than a couple thousand matches per primer. For example, using primers shorter or roughly equal to our 11-base word size misses most matches. Primers with overrepresented sequences cause problems as well. We are now busy solving most of these problems, please, be patient. If you have a minute, please, let us know what kind of expectations you have for VPCR 2.0 etc. Currently, this address is for testing VPCR 2.0, stable features will be installed on [VPCR 2.0 Homepage](#).

Search using in the database for

Primer 1

Primer 2

Primer 3

Primer 4


Primer 5

Primer 6

Primer 7

Primer 8

Annealing temperature



Analytické nástroje

- VPCR <http://grup.cribi.unipd.it/cgi-bin/mateo/vpccr2.cgi>





Osnova

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje
- Analytické nástroje
 - Vyhledávání homologií
 - Vyhledávání sekvenčních motivů, otevřených čtecích rámců, restričních míst....
 - Další [www genomové nástroje](#)

Další WWW zdroje

- TIGR (The Institute for Genomic Research, <http://www.tigr.org/software/>)
 - Recently part of the J. Craig Venter Institute

The screenshot displays the NCBI Gene database entry for PHACTR4 phosphatase and actin regulator 4 [Homo sapiens]. The page is viewed in a Mozilla Firefox browser. The main content includes:

- Gene ID:** 65979, updated on 27-Aug-2011
- Summary:**
 - Official Symbol:** PHACTR4 provided by HGNC
 - Official Full Name:** phosphatase and actin regulator 4 provided by HGNC
 - Primary source:** HGNC:25793
 - Locus tag:** RP11-442N24_A.1
 - See related:** [Ensembl:ENSG00000204138](#); [HPRD:07816](#); [MIM:608726](#)
 - Gene type:** protein coding
 - RefSeq status:** REVIEWED
 - Organism:** [Homo sapiens](#)
 - Lineage:** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo
 - Also known as:** FLJ13171; MGC20618; MGC34186; DKFZp686L07205; RP11-442N24_A.1
 - Summary:** This gene encodes a member of the phosphatase and actin regulator (PHACTR) family. Other PHACTR family members have been shown to inhibit protein phosphatase 1 (PP1) activity, and the homolog of this gene in the mouse has been shown to interact with actin and PP1. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
- Genomic context:**
 - Location:** 1p35.3
 - Sequence:** Chromosome 1; NC_000001.10 (28896093..28826881)
 - Diagram:** A genomic map of Chromosome 1 - NC_000001.10 showing the location of PHACTR4 and other genes (SESN2, NEDL6, SNORA73A, SNORA73B, SNORA73C, SNORA73D, SNORA73E, SNORA73F, SNORA73G, SNORA73H, SNORA73I, SNORA73J, SNORA73K, SNORA73L, SNORA73M, SNORA73N, SNORA73O, SNORA73P, SNORA73Q, SNORA73R, SNORA73S, SNORA73T, SNORA73U, SNORA73V, SNORA73W, SNORA73X, SNORA73Y, SNORA73Z, SNORA73AA, SNORA73AB, SNORA73AC, SNORA73AD, SNORA73AE, SNORA73AF, SNORA73AG, SNORA73AH, SNORA73AI, SNORA73AJ, SNORA73AK, SNORA73AL, SNORA73AM, SNORA73AN, SNORA73AO, SNORA73AP, SNORA73AQ, SNORA73AR, SNORA73AS, SNORA73AT, SNORA73AU, SNORA73AV, SNORA73AW, SNORA73AX, SNORA73AY, SNORA73AZ, SNORA73BA, SNORA73BB, SNORA73BC, SNORA73BD, SNORA73BE, SNORA73BF, SNORA73BG, SNORA73BH, SNORA73BI, SNORA73BJ, SNORA73BK, SNORA73BL, SNORA73BM, SNORA73BN, SNORA73BO, SNORA73BP, SNORA73BQ, SNORA73BR, SNORA73BS, SNORA73BT, SNORA73BU, SNORA73BV, SNORA73BW, SNORA73BX, SNORA73BY, SNORA73BZ, SNORA73CA, SNORA73CB, SNORA73CC, SNORA73CD, SNORA73CE, SNORA73CF, SNORA73CG, SNORA73CH, SNORA73CI, SNORA73CJ, SNORA73CK, SNORA73CL, SNORA73CM, SNORA73CN, SNORA73CO, SNORA73CP, SNORA73CQ, SNORA73CR, SNORA73CS, SNORA73CT, SNORA73CU, SNORA73CV, SNORA73CW, SNORA73CX, SNORA73CY, SNORA73CZ, SNORA73DA, SNORA73DB, SNORA73DC, SNORA73DD, SNORA73DE, SNORA73DF, SNORA73DG, SNORA73DH, SNORA73DI, SNORA73DJ, SNORA73DK, SNORA73DL, SNORA73DM, SNORA73DN, SNORA73DO, SNORA73DP, SNORA73DQ, SNORA73DR, SNORA73DS, SNORA73DT, SNORA73DU, SNORA73DV, SNORA73DW, SNORA73DX, SNORA73DY, SNORA73DZ, SNORA73EA, SNORA73EB, SNORA73EC, SNORA73ED, SNORA73EE, SNORA73EF, SNORA73EG, SNORA73EH, SNORA73EI, SNORA73EJ, SNORA73EK, SNORA73EL, SNORA73EM, SNORA73EN, SNORA73EO, SNORA73EP, SNORA73EQ, SNORA73ER, SNORA73ES, SNORA73ET, SNORA73EU, SNORA73EV, SNORA73EW, SNORA73EX, SNORA73EY, SNORA73EZ, SNORA73FA, SNORA73FB, SNORA73FC, SNORA73FD, SNORA73FE, SNORA73FF, SNORA73FG, SNORA73FH, SNORA73FI, SNORA73FJ, SNORA73FK, SNORA73FL, SNORA73FM, SNORA73FN, SNORA73FO, SNORA73FP, SNORA73FQ, SNORA73FR, SNORA73FS, SNORA73FT, SNORA73FU, SNORA73FV, SNORA73FW, SNORA73FX, SNORA73FY, SNORA73FZ, SNORA73GA, SNORA73GB, SNORA73GC, SNORA73GD, SNORA73GE, SNORA73GF, SNORA73GG, SNORA73GH, SNORA73GI, SNORA73GJ, SNORA73GK, SNORA73GL, SNORA73GM, SNORA73GN, SNORA73GO, SNORA73GP, SNORA73GQ, SNORA73GR, SNORA73GS, SNORA73GT, SNORA73GU, SNORA73GV, SNORA73GW, SNORA73GX, SNORA73GY, SNORA73GZ, SNORA73HA, SNORA73HB, SNORA73HC, SNORA73HD, SNORA73HE, SNORA73HF, SNORA73HG, SNORA73HH, SNORA73HI, SNORA73HJ, SNORA73HK, SNORA73HL, SNORA73HM, SNORA73HN, SNORA73HO, SNORA73HP, SNORA73HQ, SNORA73HR, SNORA73HS, SNORA73HT, SNORA73HU, SNORA73HV, SNORA73HW, SNORA73HX, SNORA73HY, SNORA73HZ, SNORA73IA, SNORA73IB, SNORA73IC, SNORA73ID, SNORA73IE, SNORA73IF, SNORA73IG, SNORA73IH, SNORA73II, SNORA73IJ, SNORA73IK, SNORA73IL, SNORA73IM, SNORA73IN, SNORA73IO, SNORA73IP, SNORA73IQ, SNORA73IR, SNORA73IS, SNORA73IT, SNORA73IU, SNORA73IV, SNORA73IW, SNORA73IX, SNORA73IY, SNORA73IZ, SNORA73JA, SNORA73JB, SNORA73JC, SNORA73JD, SNORA73JE, SNORA73JF, SNORA73JG, SNORA73JH, SNORA73JI, SNORA73JJ, SNORA73JK, SNORA73JL, SNORA73JM, SNORA73JN, SNORA73JO, SNORA73JP, SNORA73JQ, SNORA73JR, SNORA73JS, SNORA73JT, SNORA73JU, SNORA73JV, SNORA73JW, SNORA73JX, SNORA73JY, SNORA73JZ, SNORA73KA, SNORA73KB, SNORA73KC, SNORA73KD, SNORA73KE, SNORA73KF, SNORA73KG, SNORA73KH, SNORA73KI, SNORA73KJ, SNORA73KL, SNORA73KM, SNORA73KN, SNORA73KO, SNORA73KP, SNORA73KQ, SNORA73KR, SNORA73KS, SNORA73KT, SNORA73KU, SNORA73KV, SNORA73KW, SNORA73KX, SNORA73KY, SNORA73KZ, SNORA73LA, SNORA73LB, SNORA73LC, SNORA73LD, SNORA73LE, SNORA73LF, SNORA73LG, SNORA73LH, SNORA73LI, SNORA73LJ, SNORA73LK, SNORA73LL, SNORA73LM, SNORA73LN, SNORA73LO, SNORA73LP, SNORA73LQ, SNORA73LR, SNORA73LS, SNORA73LT, SNORA73LU, SNORA73LV, SNORA73LW, SNORA73LX, SNORA73LY, SNORA73LZ, SNORA73MA, SNORA73MB, SNORA73MC, SNORA73MD, SNORA73ME, SNORA73MF, SNORA73MG, SNORA73MH, SNORA73MI, SNORA73MJ, SNORA73MK, SNORA73ML, SNORA73MN, SNORA73MO, SNORA73MP, SNORA73MQ, SNORA73MR, SNORA73MS, SNORA73MT, SNORA73MU, SNORA73MV, SNORA73MW, SNORA73MX, SNORA73MY, SNORA73MZ, SNORA73NA, SNORA73NB, SNORA73NC, SNORA73ND, SNORA73NE, SNORA73NF, SNORA73NG, SNORA73NH, SNORA73NI, SNORA73NJ, SNORA73NK, SNORA73NL, SNORA73NM, SNORA73NN, SNORA73NO, SNORA73NP, SNORA73NQ, SNORA73NR, SNORA73NS, SNORA73NT, SNORA73NU, SNORA73NV, SNORA73NW, SNORA73NX, SNORA73NY, SNORA73NZ, SNORA73OA, SNORA73OB, SNORA73OC, SNORA73OD, SNORA73OE, SNORA73OF, SNORA73OG, SNORA73OH, SNORA73OI, SNORA73OJ, SNORA73OK, SNORA73OL, SNORA73OM, SNORA73ON, SNORA73OO, SNORA73OP, SNORA73OQ, SNORA73OR, SNORA73OS, SNORA73OT, SNORA73OU, SNORA73OV, SNORA73OW, SNORA73OX, SNORA73OY, SNORA73OZ, SNORA73PA, SNORA73PB, SNORA73PC, SNORA73PD, SNORA73PE, SNORA73PF, SNORA73PG, SNORA73PH, SNORA73PI, SNORA73PJ, SNORA73PK, SNORA73PL, SNORA73PM, SNORA73PN, SNORA73PO, SNORA73PP, SNORA73PQ, SNORA73PR, SNORA73PS, SNORA73PT, SNORA73PU, SNORA73PV, SNORA73PW, SNORA73PX, SNORA73PY, SNORA73PZ, SNORA73QA, SNORA73QB, SNORA73QC, SNORA73QD, SNORA73QE, SNORA73QF, SNORA73QG, SNORA73QH, SNORA73QI, SNORA73QJ, SNORA73QK, SNORA73QL, SNORA73QM, SNORA73QN, SNORA73QO, SNORA73QP, SNORA73QQ, SNORA73QR, SNORA73QS, SNORA73QT, SNORA73QU, SNORA73QV, SNORA73QW, SNORA73QX, SNORA73QY, SNORA73QZ, SNORA73RA, SNORA73RB, SNORA73RC, SNORA73RD, SNORA73RE, SNORA73RF, SNORA73RG, SNORA73RH, SNORA73RI, SNORA73RJ, SNORA73RK, SNORA73RL, SNORA73RM, SNORA73RN, SNORA73RO, SNORA73RP, SNORA73RQ, SNORA73RR, SNORA73RS, SNORA73RT, SNORA73RU, SNORA73RV, SNORA73RW, SNORA73RX, SNORA73RY, SNORA73RZ, SNORA73SA, SNORA73SB, SNORA73SC, SNORA73SD, SNORA73SE, SNORA73SF, SNORA73SG, SNORA73SH, SNORA73SI, SNORA73SJ, SNORA73SK, SNORA73SL, SNORA73SM, SNORA73SN, SNORA73SO, SNORA73SP, SNORA73SQ, SNORA73SR, SNORA73SS, SNORA73ST, SNORA73SU, SNORA73SV, SNORA73SW, SNORA73SX, SNORA73SY, SNORA73SZ, SNORA73TA, SNORA73TB, SNORA73TC, SNORA73TD, SNORA73TE, SNORA73TF, SNORA73TG, SNORA73TH, SNORA73TI, SNORA73TJ, SNORA73TK, SNORA73TL, SNORA73TM, SNORA73TN, SNORA73TO, SNORA73TP, SNORA73TQ, SNORA73TR, SNORA73TS, SNORA73TT, SNORA73TU, SNORA73TV, SNORA73TW, SNORA73TX, SNORA73TY, SNORA73TZ, SNORA73UA, SNORA73UB, SNORA73UC, SNORA73UD, SNORA73UE, SNORA73UF, SNORA73UG, SNORA73UH, SNORA73UI, SNORA73UJ, SNORA73UK, SNORA73UL, SNORA73UM, SNORA73UN, SNORA73UO, SNORA73UP, SNORA73UQ, SNORA73UR, SNORA73US, SNORA73UT, SNORA73UU, SNORA73UV, SNORA73UW, SNORA73UX, SNORA73UY, SNORA73UZ, SNORA73VA, SNORA73VB, SNORA73VC, SNORA73VD, SNORA73VE, SNORA73VF, SNORA73VG, SNORA73VH, SNORA73VI, SNORA73VJ, SNORA73VK, SNORA73VL, SNORA73VM, SNORA73VN, SNORA73VO, SNORA73VP, SNORA73VQ, SNORA73VR, SNORA73VS, SNORA73VT, SNORA73VU, SNORA73VV, SNORA73VW, SNORA73VX, SNORA73VY, SNORA73VZ, SNORA73WA, SNORA73WB, SNORA73WC, SNORA73WD, SNORA73WE, SNORA73WF, SNORA73WG, SNORA73WH, SNORA73WI, SNORA73WJ, SNORA73WK, SNORA73WL, SNORA73WM, SNORA73WN, SNORA73WO, SNORA73WP, SNORA73WQ, SNORA73WR, SNORA73WS, SNORA73WT, SNORA73WU, SNORA73WV, SNORA73WW, SNORA73WX, SNORA73WY, SNORA73WZ, SNORA73XA, SNORA73XB, SNORA73XC, SNORA73XD, SNORA73XE, SNORA73XF, SNORA73XG, SNORA73XH, SNORA73XI, SNORA73XJ, SNORA73XK, SNORA73XL, SNORA73XM, SNORA73XN, SNORA73XO, SNORA73XP, SNORA73XQ, SNORA73XR, SNORA73XS, SNORA73XT, SNORA73XU, SNORA73XV, SNORA73XW, SNORA73XX, SNORA73XY, SNORA73XZ, SNORA73YA, SNORA73YB, SNORA73YC, SNORA73YD, SNORA73YE, SNORA73YF, SNORA73YG, SNORA73YH, SNORA73YI, SNORA73YJ, SNORA73YK, SNORA73YL, SNORA73YM, SNORA73YN, SNORA73YO, SNORA73YP, SNORA73YQ, SNORA73YR, SNORA73YS, SNORA73YT, SNORA73YU, SNORA73YV, SNORA73YW, SNORA73YX, SNORA73YY, SNORA73YZ, SNORA73ZA, SNORA73ZB, SNORA73ZC, SNORA73ZD, SNORA73ZE, SNORA73ZF, SNORA73ZG, SNORA73ZH, SNORA73ZI, SNORA73ZJ, SNORA73ZK, SNORA73ZL, SNORA73ZM, SNORA73ZN, SNORA73ZO, SNORA73ZP, SNORA73ZQ, SNORA73ZR, SNORA73ZS, SNORA73ZT, SNORA73ZU, SNORA73ZV, SNORA73ZW, SNORA73ZX, SNORA73ZY, SNORA73ZZ
- Genomic regions, transcripts, and products:**
 - Genomic Sequence:** NC_000001 chromosome 1 reference GRCh37.p5 Primary Assembly
 - Links:** Order cDNA clone, BioAssay, by Gene target, BioProjects, CCDS, Conserved Domains, dbVar, EST, Full text in PMC, Genome, GEO Profiles, HomoloGene, Map Viewer, Nucleotide, OMIM, Probe, Protein, PubChem Compound, PubChem Substance, PubMed, PubMed (GeneRIF), PubMed (OMIM), RefSeq Proteins



Shrnutí

- Schéma přednášky
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY
- Databáze
 - Spektrum „on-line“ zdrojů
 - PRIMÁRNÍ, SEKUNDÁRNÍ a STRUKTURÁLNÍ databáze
 - GENOMOVÉ zdroje
- Analytické nástroje
 - Vyhledávání homologií
 - Vyhledávání sekvenčních motivů, otevřených čtecích rámců, restričních míst....
 - Další www genomové nástroje



Diskuse



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

● ● ● Přístupy „klasické“ genetiky



3

:
?

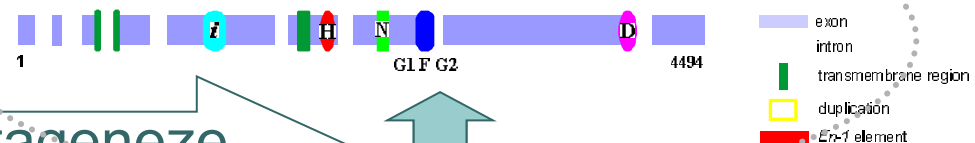


1

„Reverzně genetický“ přístup

5'TTATATATATATATTAATAAAATAAAATAAAA
GAACAAAAAGAAAATAAAATA...3'

BIOINFORMATIKA

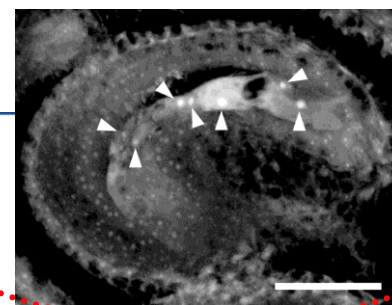


FUNKČNÍ GENOMIKA



En-1
gaattca agt cgt CAC TACAAGA TCT TGTASTG cgtggagact
1122 1123

...aat tca agt cgt gga gac tac act...
N S S R G D Y I



VOJE VZDĚLÁVÁNÍ

entace je spolufinancována
ropským sociálním fondem
a státním rozpočtem České republiky