

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 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

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**
 - 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

:

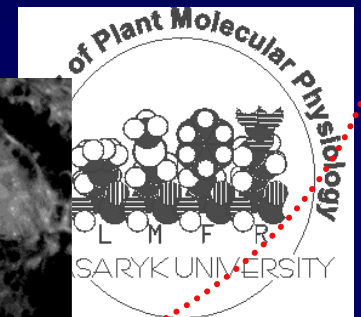
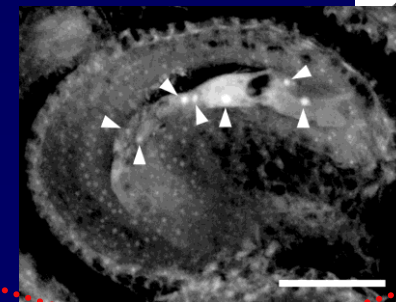
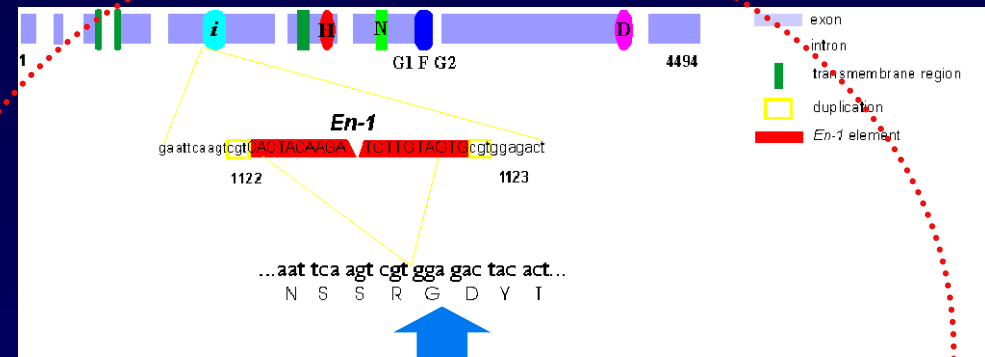
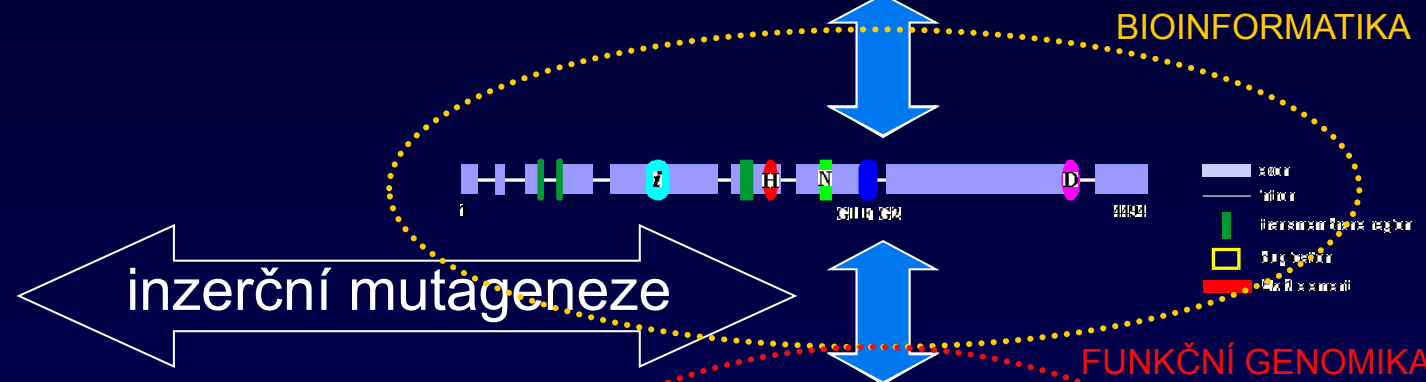
1



?

„Reverzně genetický“ přístup

5'TTATATATATATATTA AAAAATAAAATAAAA
GAACAAAAAGAAAATAAAATA...3'





Osnova

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

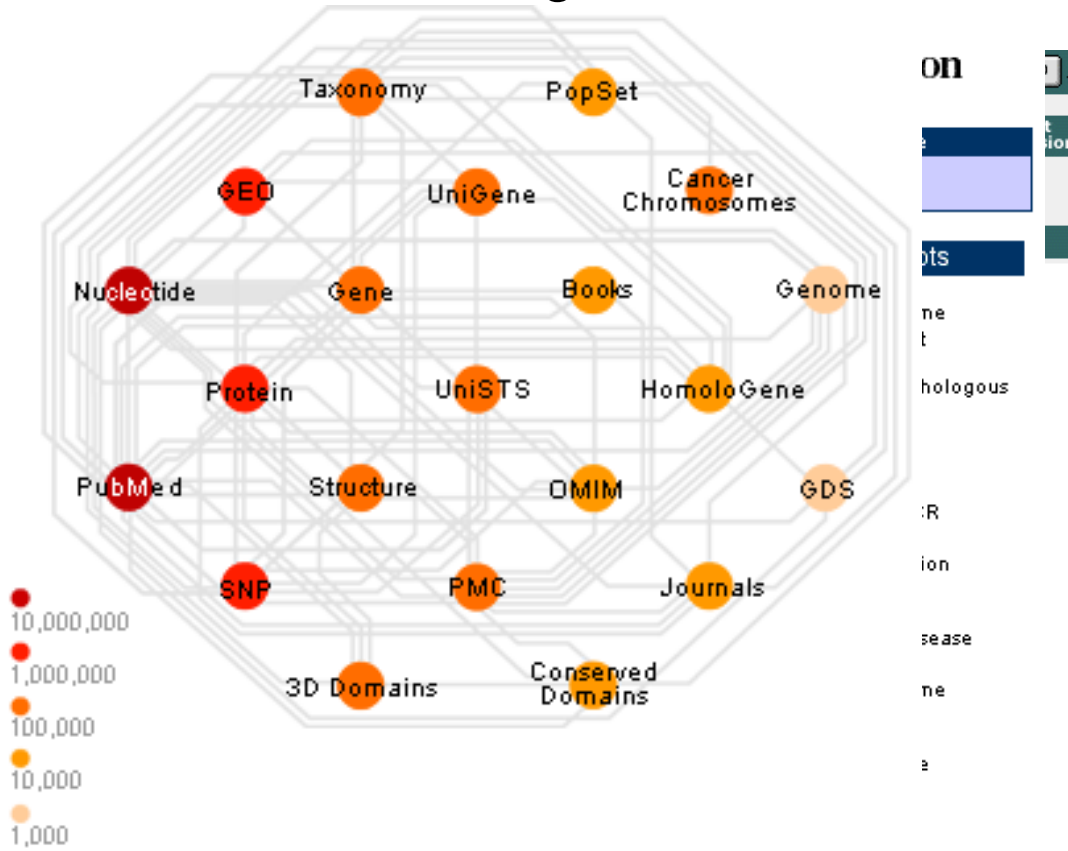
Spektrum on-line zdrojů

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

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 displays the NCBI website interface. At the top, the NCBI logo is on the left, and the text "National Center for Biotechnology Information" is centered, 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: "What does NCBI do?" (Established in 1988 as a national resource for molecular biology information...), "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), and "BLink and get results fast!". On the left side, there is a "SITE MAP" section with links to "About NCBI", "GenBank", "Molecular databases", "Literature databases", and "Genomic biology".

Primární databáze

Přístupový
kód

```
/translation="MNGRYSPTRODPFTGAKPWSILALIVAMIFAFNAVASWQDNAT
TQAILSQLRSINADSASLORDVLRRAHTGTVANYRPIISRLGALKRNKLEDLKLQJLQROSH
IVSEBNAQQLLRQLEYSLMSADAAVAAPGAQNVRLQDSLASPTRALSSLGKASTDQT
LEKPTLACMMLQFLOPSPAISEFISLELERLQKQGLDRAPVRILAREGPIILELL
POKEDLVNMIQCTDARIEMLQRECLVYSLKNVRESRIFLGSASVGLCLYIITL
LQKQFEGHAATTSQAALRIIQRPDADTICALAL
KQKRTKADERATVFRIISEKKIVHLPLEIP
FIQLLELATACLCHYIDVREKQTECD
SSILGHAELAQNSVRSVTRYIDYII
KPFVSELVTEIAPLLEKALPPIELSPFDQM
PFLRLQVLINICKNASQAMTANGQIDIIISQAPLQVKKILAHGVMPFDYVL
LSEKNGGIPRAVLPHIFPPFTTRRNGGTGLGLASVHGHSAPAGYIDVSVTVGH
GTRFDIYLFPSKPEFVMPDSFFGRNKAPRGMGEIVALVEPDLLEAYEDKIALGLEY
FVGFPTPNRIRDWISKGNADLVMDQASLPHDQSPNSVDLVKTSASIIIGNDLKMT
LSREDVTRDPLPKPTSSRMAHALTKIK"
```

GeneBank Identifier

```
ORIGIN
1 atgaacggaa gatattcaac gacgggag gattttaaga caggcgcaa gccctgtct
61 atattggcc ttatggtt gatt ttcggttca tggcggtgc gtcctggcag
gatttaacgc cgacagcgcc
cggaaacta ccgcccatt
agcaatt atttagacaa
gcccagtaga agtgtctcta
atgacgccc ccgcaaaa atgacgccc gcaagattcg
421 ctggccagtt tcaactgtgc tttagcagtt cttccaggaa aagcctcaac cgcacagact
481 ttgaaaaaac caacagaatt ggttagcagc atgctccaat ttcttggca accaagcccg
541 gctatttcat togagatcag cttgaaacta gagaggtccc aaaaaaacg cggctctgat
601 gaagctcccg tgcgcactct tgcacgtgaa ggtcccatta tcttctcgc ttgcccacag
661 gtgaaaagtc tggtagaacat gattcagacg tctgacacgc cagaatttgc gggagatgct
721 cagcgcaggt gtttggaggt ctatagcttg aaaaatgtag aggagccggag cgcacgtatc
781 ttctctgggt ccgcttcagt gggctcttgc ctctacatca tcaacttagt ctataggcta
841 cggaaaaaaa ccgattgggt agcggcgctg ttagattacg aagagctaat caaagagatc
901 ggagtatggt ttgaagtgga ggcggccacc acgtctcgc ccgaagctgc acctctatt
961 atccagcgtc tctttgatgc cgtactgtgc gcttagctc tagtgacca tgaccagtaga
1021 tgggtgtgct aaacattcgg tgcgaaacac caaaaacctg tgtgggacga cagcgtgcta
1081 cgcgaaatag tctctgtac caaaggcgac gaacggggca cggattctcg cactaatatg
1141 tcgaaaaaaa tctacattt gccctctgaa attccaggtc tctcgaact actggtctac
1201 aaatccacag ataaactaat tgcggtttgt tcaactgggt accaaaagcta tgcgctcga
1261 ccttgccaag gogaattca gctcttgaa ctgcgcaacg cctgctctg tcactatate
1321 gatgttcggc gtaagcagac cgaatggac gttttggcca gacgattgga geatgcgcaa
1381 cgccttgagg cagttggtac acttgcggc ggaatagcac atgaatttaa taacatttg
1441 ggtccaatcc tgggcaacgc agaattagca caaaactcgg tctctogaac atctgcacc
1501 cgaagatata ttgactatc cttctgca ggagcagag ccatgctcat tctcagcag
1561 atcttgacgc tgagcogaaa acagggagc atgactcaag catttagtgt ctacagcctt
1621 gtgacogaaa toctccctt gctacgtatg gctctccgc caaacatcga gcttagttc
1681 agatttgatc aaatcagag cgtgatcgaa ggaagccgc ttgaacttca acaggtacta
1741 attaacatc gcaagaatgc ttcccaagc atgactgcaa atggtcaaat cgcacatcct
1801 atcagcgaag ctttttacc agttaagaaa atctgtgcgc atggtgttat gccacctggc
1861 gactatgttc tctatctat tagcgacaaat ggtggaggca tcccgaggc tgtgttacc
1921 cacatttttg aacctctct tgcgacaga gctgcgaacg gtggaaaggg tctcggcctt
1981 gctctgtgct atggtcaat cagcggcttt ggggttaca tgcaggttag tccaactgtt
2041 gggcagggga cggcttga catttatctc cctcgtctt ctaaggaacc cgtaaatcca
2101 gacagttttt tggcgcgcaa taaggcaccg cgtggaacg gggagattgt ggcacttgtt
2161 gacccgatg acctctgag gggagctat gaagacaaga tgcgcgtct aggatagag
2221 ccggtcgggt tctgacct taatgaaat ccgcatgga tttcaaaag caatgaagcc
2281 gatctggtca tggtagacca agcgtctctt cctgaagatc aaagtctaa tctcgtgat
2341 ttatgtctca agaaccctc cactatcatt ggggaaatg atctcaaat gccctttda
```



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

Sekundární databáze

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

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

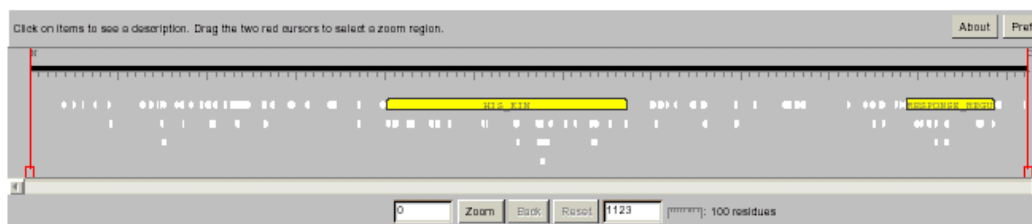
>PDOC50109 PS50109 HIS_KIN Histidine kinase domain [profile].

```
402 - 671 NASHDIRGALAGMKGLIDICRDGVKPGSDVDTTINQVMVCAKDLVALLNSVLEMSKIRSG
KMQLVREDFNLSKLLLEDVIDFYHPVAMKKGVDVLDPHDgsvEKFSNVRGDSGRLKQILN
NLVSNARVFTVD - GHIIVRAWAQrpggnasvvlasyppkgsakfvkkmfcknkeesatye
teiansirnnanTMEFVFEVDDTGGIIFMHRKESVPRNYVQVREtAQGHQGTGLGLGIIVQ
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

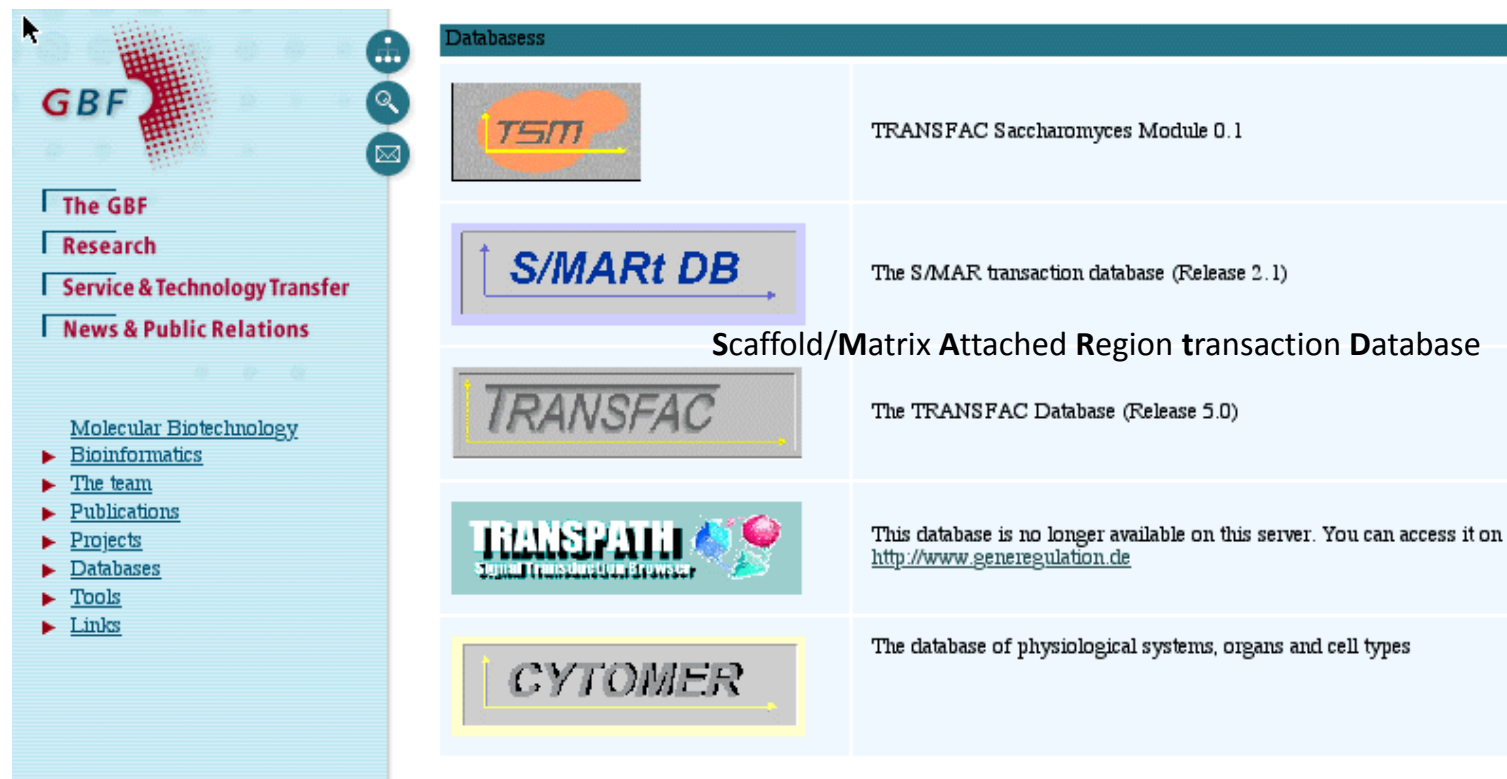
START THE SCAN | RESET

match mode: (for patterns, see [help](#))





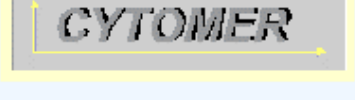
randomize databases: (to test a pattern, see [help](#))

Sekundární databáze

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

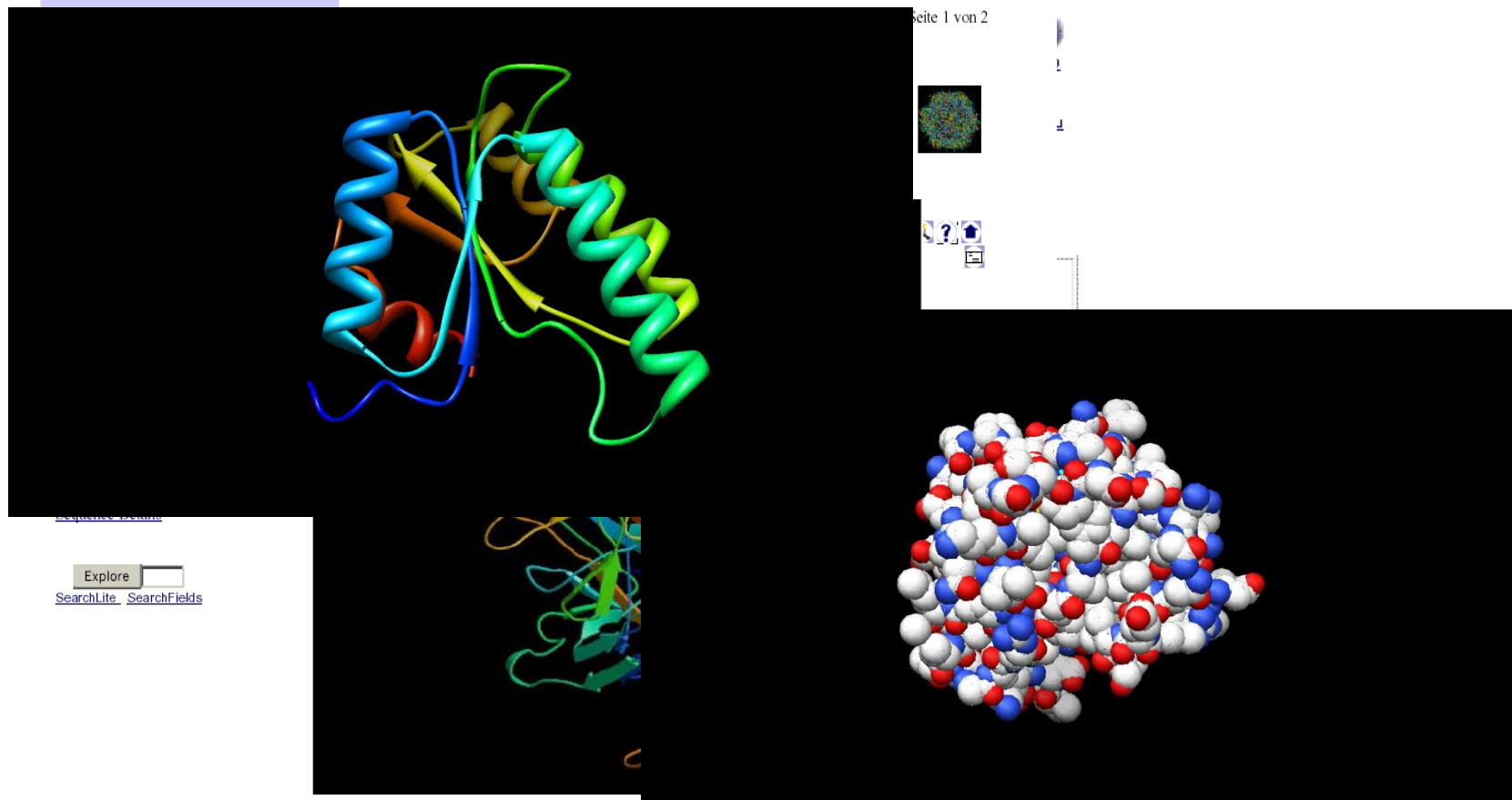


The screenshot shows the GBF website interface. On the left is a navigation menu with the GBF logo and links for 'The GBF', 'Research', 'Service & Technology Transfer', and 'News & Public Relations'. Below these are links for 'Molecular Biotechnology', 'Bioinformatics', 'The team', 'Publications', 'Projects', '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** **VI** **VII**

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	TIGR MIPS TAIR NM_100396 NP_172010 auxin-resistance protein AXR1
Atlg08210	T2G3I8.7	TIGR MIPS TAIR NM_100695 NP_563808 expressed protein
Atlg11250	T28P6.10	TIGR MIPS TAIR NM_100997 NP_172591 syntaxin-related protein At-SYR1, putative
Atlg14670	T5E2I.15	TIGR MIPS TAIR NM_101334 NP_172919 endomembrane protein, putative
Atlg17790	F2H15.2	TIGR MIPS TAIR NM_101642 NP_564037 expressed protein
Atlg21050	T22I11.13	TIGR MIPS TAIR NM_101958 NP_564130 expressed protein
Atlg24210	F3I6.14	TIGR MIPS TAIR NM_102267 NP_564212 expressed protein
Atlg28370	F3M18.20	TIGR MIPS TAIR NM_102603 NP_174159 ethylene-responsive element binding factor, putative
Atlg31885	F5M6.28	TIGR MIPS TAIR NM_102926 NP_174472 major intrinsic protein, putative
Atlg35670	F15O4.8	TIGR MIPS TAIR NM_103271 NP_174807 calcium-dependent protein kinase
Atlg48160	F21D18.11	TIGR MIPS TAIR NM_103712 NP_175250 signal recognition particle 19 kDa protein subunit, putative
Atlg51980	F5F19.4	TIGR MIPS TAIR NM_104079 NP_175610 mitochondrial processing peptidase alpha subunit, putative
Atlg55150	T7N22.9	TIGR MIPS TAIR NM_104388 NP_175911 ethylene-responsive RNA helicase, putative
Atlg60140	T13D8.4	TIGR MIPS TAIR NM_104705 NP_176221 trehalose-6-phosphate synthase, putative
Atlg63750	F24D7.6	TIGR MIPS TAIR NM_105052 NP_176562 putative disease resistance protein
Atlg67090	F5A8.1	TIGR MIPS TAIR NM_105379 NP_176880 ribulose-bisphosphate carboxylase small unit, putative
Atlg69870	T17F3.10	TIGR MIPS TAIR NM_105655 NP_564979 putative peptide transporter
Atlg72970	F3N23.17	TIGR MIPS TAIR NM_105955 NP_565050 expressed protein
Atlg75780	T4O12.1	TIGR MIPS TAIR NM_106228 NP_177706 tubulin beta-1 chain
Atlg78700	F9K20.26	TIGR MIPS TAIR NM_106517 NP_565187 expressed protein

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

Genomové zdroje

NCBI Nucleotide search interface showing a search for Arabidopsis thaliana sequence. The search results show a single entry: 1: NC_003070. Arabidopsis thali...[gi:22330780].

The interface includes options for Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, OMIM, and Books. The search results show a single entry: 1: NC_003070. Arabidopsis thali...[gi:22330780].

The detailed view shows the reverse complement strand and protein coding genes. The search for gene is set to "Find" and "CDS with gene and mRNA" is checked. The sequence is displayed with a scale from 300286 to 1. The sequence is shown in both DNA and protein format. The protein sequence is: M Q A V K R S R R H V E E E P T M V E P K T K Y D R Q L R I W G E V G Q A A L E E A S I C L L N C G P.

Legend:

- CDS
- RNA
- gene
- sequence fragment shown

Sequence:

```

1501830 CTTTTTGTTC ATCAGTTCAC CGGAGGCCAAA ATCGCTCTCT CGCTTGAGCT GCGAGAGTGC At1g05180 mRNA-protein id:
                                     M
1501778 AAGCAGTAAA AAGATCCAGG AGCATGTTG AAGAGAGGCC AACAATGCTA GACCTAATAA At1g05180 mRNA-protein id:
Q A V K R S R R H V E E E P T M V E P K
1501710 CCRAGTACGA TCGTCAGCTC AGGTATACAT ACTCTTTTC CTTAACTCT ACTTCGAACT At1g05180 mRNA-protein id:
T K Y D R Q L R
1501650 ACTCTGTGCG GGAGCAAAAT TGATGAGTA GTAATCGTTG GGTGACTGA TTAGGATTTG At1g05180 mRNA-protein id:
I W
1501590 GGGGGAGCTA GGTCAAGCGG CCTTGAGAGA ACCGAGTATC TGTTTACTCA ATTGTGGCCC At1g05180 mRNA-protein id:
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](#)

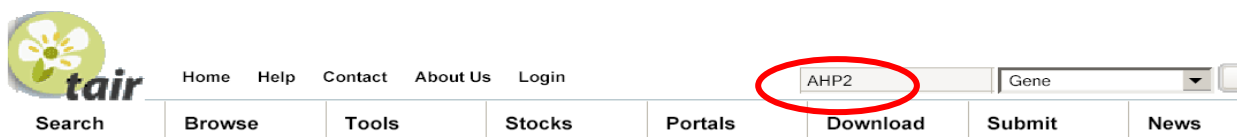
The chimp sequence used in these alignments was obtained from the 13 Nov. 2003 Arachne assembly. We'd like to thank NHGRI, the Eli & Edythe L. Broad Institute at MIT/Harvard, and Washington University School of Medicine for providing this sequence, and LaDeana Hillier, Washington University School of 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

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.

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

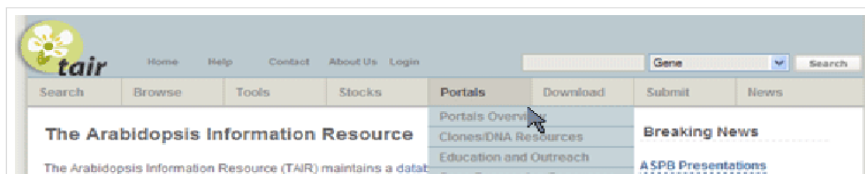
[October 15, 2006]
Search for [genes](#), [germplasms](#), and [polymorphisms](#) using associated phenotype, and see improved phenotype data display in results and detail pages.

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.

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.



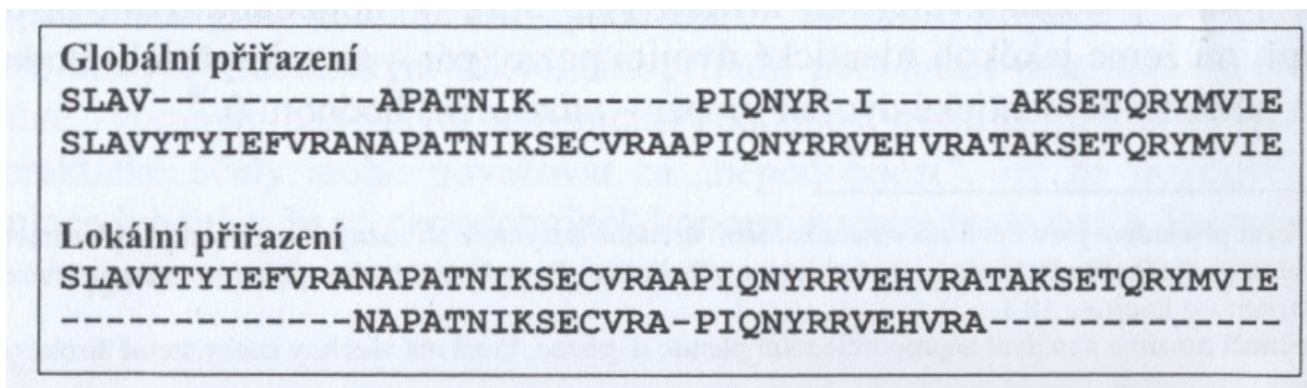


Osnova

- 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

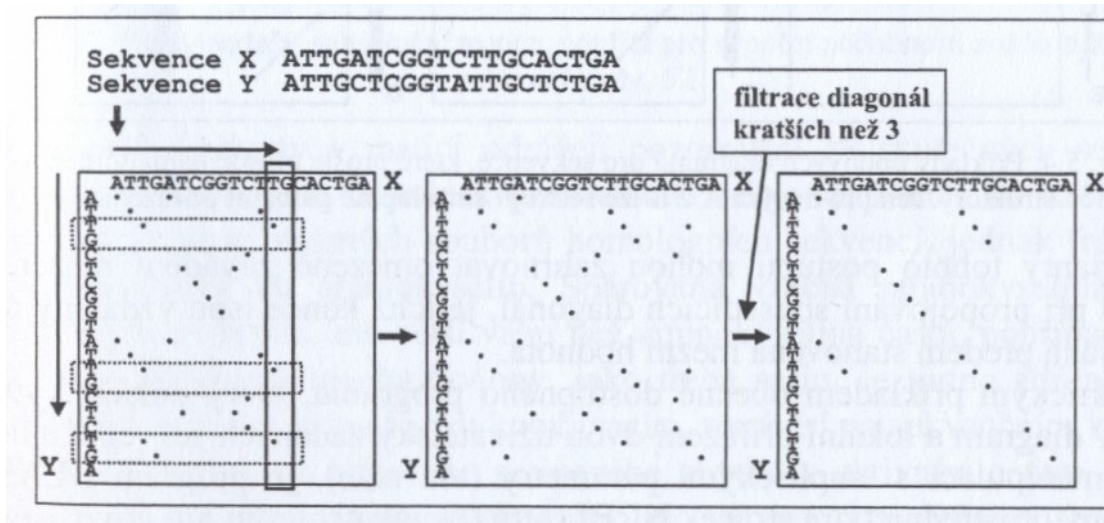


Cvrčková, Úvod do praktické bioinformatiky

- 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)

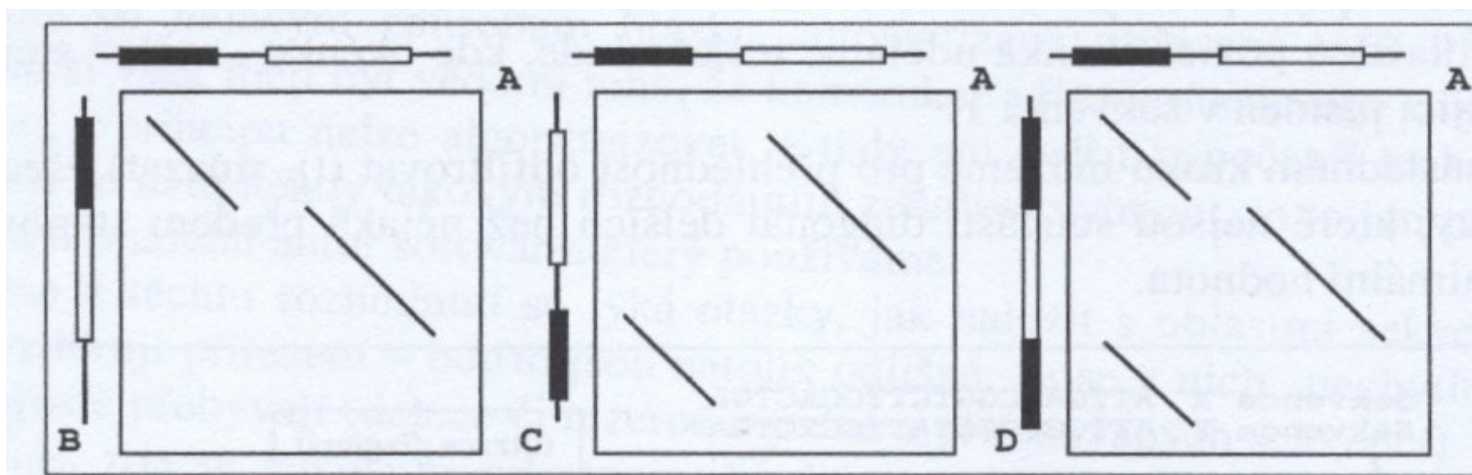


Cvrčková, Úvod do praktické bioinformatiky

- 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

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

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

[Search](#)

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

[Set subsequence](#) From: To:

[Choose database](#) nr

Now: **BLAST!** or **Reset query** **Reset all**

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	2	-2	-3	-3	-4	-5	-5	-5	-4	-4	-5	-2	0	-3	-3	-4	0	-8
2	0	1	1	1	1	1	0	0	0	1	1	1	-2	-1	-1	-1	-3	-3	-2
3	1	0	6	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
6	1	0	6	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
5	1	0	-1	1	5	-4	-5	-5	-5	2	2	2	-2	-3	-3	-3	-4	-4	-4
2	0	-1	0	0	2	-4	-5	-5	-5	2	2	2	-2	-3	-3	-3	-4	-4	-4
4	0	-1	0	0	2	4	-5	-5	-5	4	4	4	-2	-3	-3	-3	-4	-4	-4
4	0	-1	0	0	1	3	4	-5	-5	4	4	4	-2	-3	-3	-3	-4	-4	-4
4	-1	-1	0	0	-1	1	2	2	4	2	2	2	-2	-2	-2	-2	-2	-2	-2
6	-1	-1	0	-1	-2	2	1	1	3	6	6	6	-2	-3	-3	-3	-4	-4	-4
6	0	-1	0	-2	-3	0	-1	-1	1	2	6	6	-2	-3	-3	-3	-4	-4	-4
5	0	0	-1	-1	-2	1	0	0	1	0	3	5	-2	-2	-2	-2	-2	-2	-2
6	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6	6	6	6	6	6	6
5	-2	1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	2	5	5	5	5	5	5
6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6	6	6	6	6
4	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	2	4	2	4	4	4	4
9	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9	9
10	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10
17	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	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 Initiated Blast)**
 - 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  
IYKDGERIITQGEKADSFYIESGEVSIILRSRTKSNKDGGNQEEVEIARCHKGQYFGELALVTNKPRAAS  
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
GAGCTCCCTTGGGGGGCAAGGGCAAAAACTTTTGCTAAATGGAAAAATATTATACCAAGTGTGTAATA
GTTACTCAATTTGAATTAACAAAGGGGCAAAATTTGACTATTTTGCCCTTATATCTTTTGGTCACAAAAAC
ATAAAATATCCCATCCGAAATTCCAAATGGTCCATTATCGGCCAAGTAGCTTTCTTTAATTATAGTTAGTT
GACAAAACACTATCAAGATATCATTATTATAAATAAATTCAAAGTCCATCATCTTAGCTGCCTCCTCA
GTAGAGCCGCCAGTAAATAAAGACCGATCAAAATAAAGCCGCCATTAAAAAATGAATTTTAGGACTCTC
GATTGGCAGGTAAGTGCCAAAACCTTTCCAATACCTTTGCTGCAACTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATTTCTAAGTTTTATCTCTAATTTACATCTCAACTAATATTAAGAAAATTAACAGGTA
CAGCAAAATCATAAAAATTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCCTTTTAGAG
TCTGCATGCCATATTCCTAAGGGGTCGTTTGGTACAAGAAAATAAATAAATAAATTTCCGGATAGAATTT
GAGATTGCATTTATCTTGTGTTTTAATTATAAGTATTAGCTAATTTTCAGAAATAAATTTTACTAAAATAG
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC
TTATTTATCTCACTATTTTACCAAATGATCGGTTAGTCTTTCATGAGAATCCAGTATCCTCAATAAATGCA
GTAAGAAGTTAGAAAATTTTCAITTAATCAATTCATATAAATTTAAAAATATTAGATATGGAGCACTTAAG
ATACAATAAAAGATGTACCGTTAATAAATAAAGATAAGATAGAGTTTTAAATAGGAAAAAAAACCGTT
CGAGACTCTTTATGGAAGGCGTTTCTTCAAAATGAGATTCTCATTTCATTGCTCTGGTGC AATAGCAAAA
TGACATCTTACTCTTAAGATACAGCGAGCCACTCTACAATCTTCTATTGTATACTCAAATGAAAGTTTTA
GAGAATTTCAAATCTCTCAACTACTTTTTAAGGGAATTCAAAATACGACC AATATTTATTACTTACTTAC
TTATAGTTAAATGATATGAATTTTTATTTAAATTTGAATTGAAAATATTAATTTACTTGAATTTAATATAA

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
GACAAAACACTATCAAGATATCATTATTATAAATAAACTTCAAGTCCATCATCTTTAGCTGCCTCCTCA
GTAGAGCCGCCAGTAAAAAAGACCAGATCAAAATAAAAGCCGCCATTAAAAATAATGAATTTTAGGACTCTC
GATTTGGCAGGTAAGTGGCAAAACTCTTCCAATACTTTTGTGCAACTTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATTTCTAAGTTTATCTCCTAATTTACATCTCAACTAATATTAAGAAATTAACAGGTA
CAGCAAATCATAAAATTTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCCTTTTCAGAG
TCTGCATGCCATATTTCACTAAGGGGTCGTTTGGTACAAGAAATAAATAAATAAATTTTCGGGATAGAATTT
GAGATTGCATTTATCTTGTGTTTAAATTAAGTATTAGCTAATTTTCAAGAATAAATTTTACTAAAATAG
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC
TTATTTATCTCACTATTTTACCAATGATCGGTTAGTCTTCATGAGAATCCAGTATCCTCAATAAATGCA
GTAAGAAGTTAGAAAAATTTTCAATTAATCAATTCATATAATTTAAAAATATTAGATATGGAGCACTTAAG
ATACAATAAAGATGTACCGTTAATAAATAAAGATAAGATAGAGTTTAAATAGGAAAAAAAACGGTT
CGAGACACTCTTATGGAAGGCGTTGTCTTCAAAGTAGATTCTCATTCATTGCTCTGGTGCATAGCAAAA
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 gagtcccttggggggcaagggcaaaactttttgctaaatggaaaaatattataccaagt 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

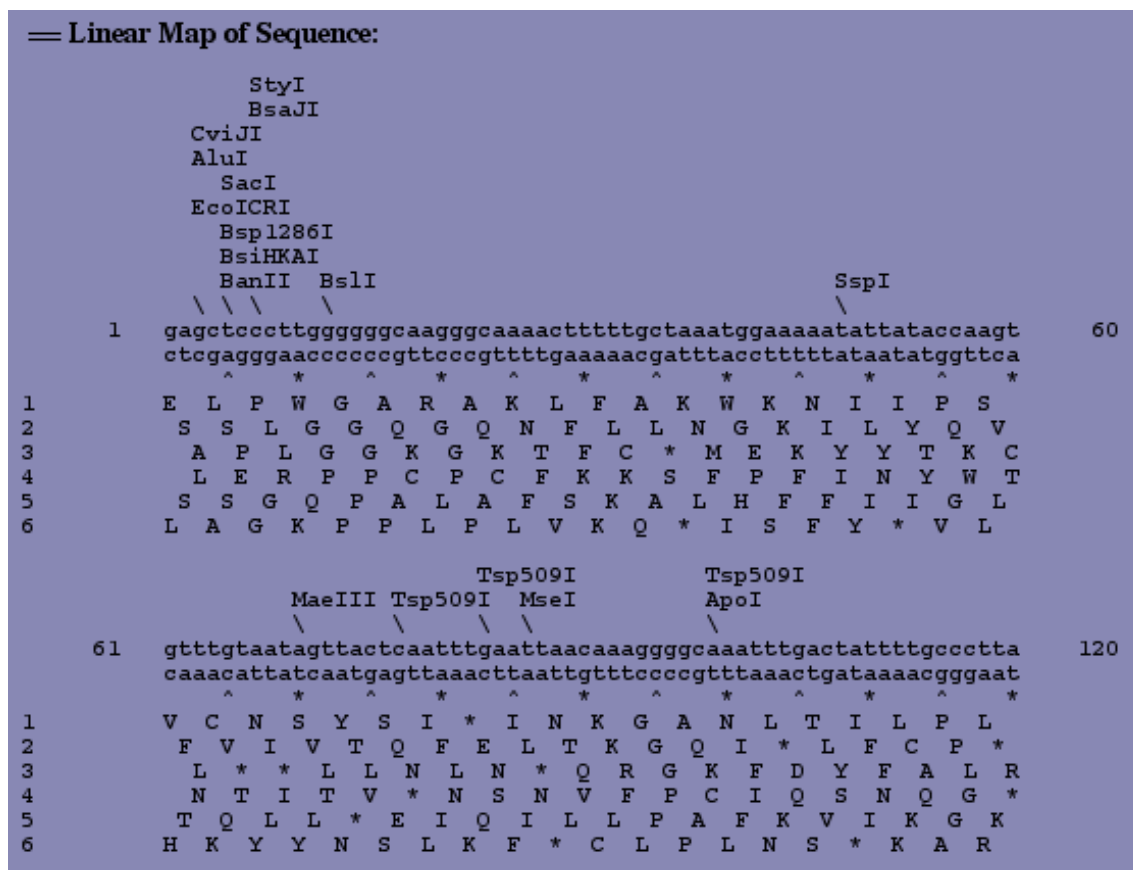
```
>170248 Translated - Frame 2
SSLGGQGQNFLLNGKILYQVFVIVTQFELTKGQI*LFCP
```

```

S S L G G Q G Q N F L L N G K I L Y Q V
2 agtcccttggggggcaagggcaaaactttttgctaaatggaaaaatattataccaagt 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/>

Selected Sequence(s)

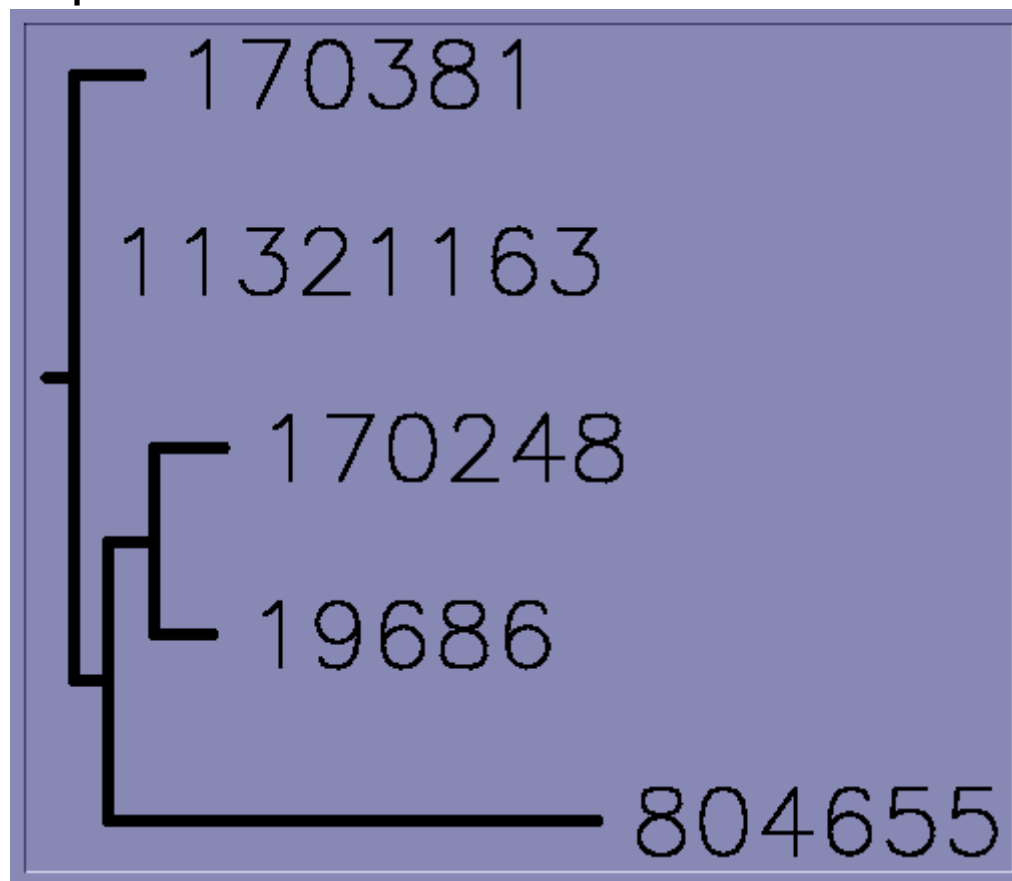
- Lycopersicon esculentum beta-1,3-glucanase mRNA, complete cds.
- Capsicum annuum clone GC170 beta-1,3-glucanase-like protein gene.
- Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds.
- Nicotiana plumbaginifolia beta-(1,3)-glucanase gene for a vacuolar.
- Hordeum vulgare L. beta-glucosidase (BGQ60) gene, complete cds.

[Download a PostScript version of the output](#)

```
-----
2560 GTTTGGTTGGTGTGTGCTTGCAGAACTTGGAGTGGAGACTGGGGTAGAGTGGGGTTTGGG 804655
      2850      2900      2950      3000      3050      3100
24  ..... A A A T G G C T . 170381
1  ..... 11321163
2430 ..... C A A G A A T T . 170248
1743 CAGTGAAATGATTGACAGAACTGCAAAAAACAAGCCAAAATGGTAAAAAATAAATTC 19686
2620 GATCGTCTATGTGGACTTCAATACTGTGAAGAGGTACGCCAAGGAGTCAAGGCTTTGGTT 804655
      2710      2720      2730      2740      2750      2760
32  ..... A T T A T G T G G T T C T A C G A T T C T T G T G G C C A . C C A A C A T T E A G A T A G 170381
1  ..... 11321163
2438 ..... A G A T A A T G A T T T A C T T T C T A A G A C T A A T T . C T A A T T C T T A T T G A G G 170248
1803 ACGATGTTTACATTCCTATGTCCAAACGGCCACTGACTATTTTC A A T T C A T T T G A G G 19686
2680 CAAGAAATGCTGTCCGAAAGAAAGACAGCTAGGATCGCAACAGGATCGGGAGGATC 804655
      2770      2780      2790      2800      2810      2820
79  A G A G G G T F A A . . . . A T A G G T G T . . . . T T G T T A T G C A A T C A T G G C C A C A C A C T T G C C A T G A C 170381
1  T . . . . A T G G G T C T . . . . T T G G T A T G C A A T C A T G G C C A C A C A C T T G C C A T T G A C 11321163
2484 A C C G G G T F A A T C A A T A G G T G T . . . . T T G G T A T G C A A T C T A G C C A C A C A C T T G C C A A A T C 170248
1863 A C C G G G T C A A T G G A T A G G T G T . . . . T T G G T A T G C A A T C T A G C C A C A C A C T T G C C A A A T C 19686
2740 A C T G G T T C A G G T T C A G A A A A A A A G A T A T G T A A T C T T T T A T C A T A G A A A C T G A G 804655
      2830      2840      2850      2860      2870      2880
132 A T T G T A A A C T T A T A C A G C . . . . T C T A C A A G T G G A A A A C A T T C A A A C T G A G C C T T T A T G A 170381
45 A T T C G G A A C T T A T A C A G C . . . . T C T A C A A G T C A A A A A C A T T C C A A A T G A A G C C T T T A T G A 11321163
2540 A T T C G G A A C T T A T A C A G C . . . . T C T A C A A G T C A A A A A C A T T C C A A A C T G A G C C T T T A T G A 170248
1919 A T T C G G A A C T T A T A C A G C . . . . T C T A C A A G T C A A A A A C A T T C C A A A C T G A G C C T T T A T G A 19686
2800 A C T T A G G G C C T C T T G C C T A A A G G C C T C G C C A A T A T G C C A G . . . . C C G G A A A T T G G A G 804655
```


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>



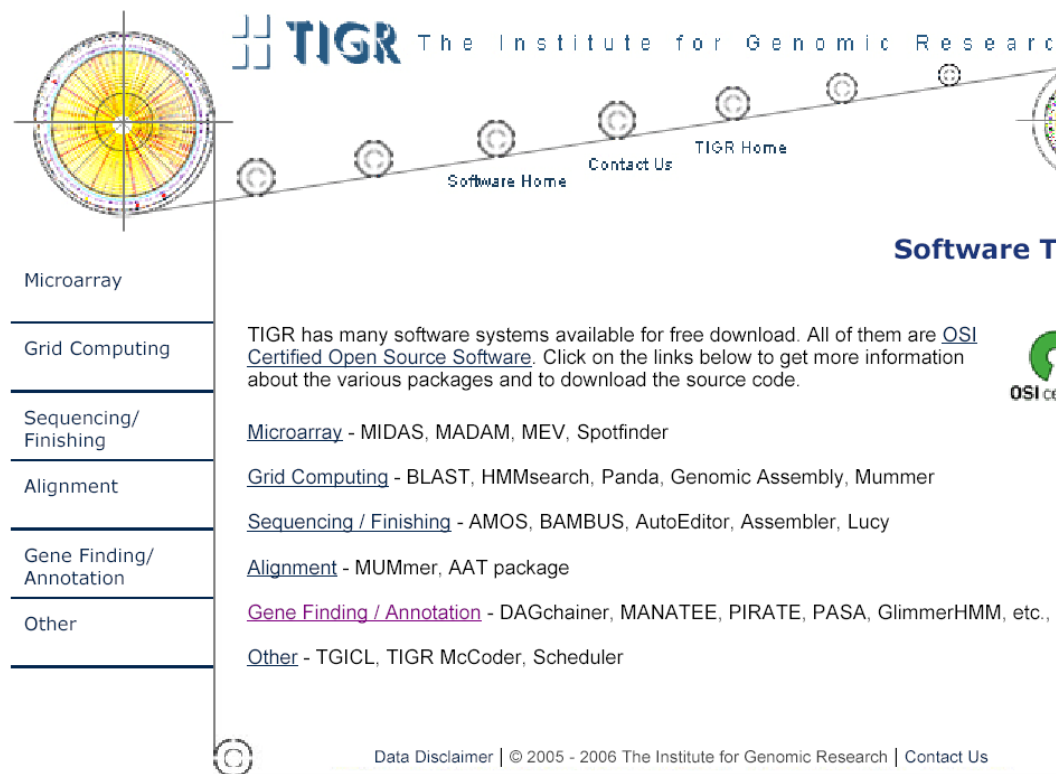


Osnova

- 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/>)



The screenshot shows the TIGR website interface. At the top left is a circular genomic map. To its right is the TIGR logo and the text "The Institute for Genomic Research". Below the logo is a navigation bar with links: "Software Home", "Contact Us", and "TIGR Home". On the right side, there is a "Software T" section with an OSI logo. A vertical menu on the left lists categories: "Microarray", "Grid Computing", "Sequencing/Finishing", "Alignment", "Gene Finding/Annotation", and "Other". The main content area provides descriptions and links for each category.

TIGR The Institute for Genomic Research

Microarray

Grid Computing

Sequencing/Finishing

Alignment

Gene Finding/Annotation

Other

TIGR has many software systems available for free download. All of them are [OSI Certified Open Source Software](#). Click on the links below to get more information about the various packages and to download the source code.

[Microarray](#) - MIDAS, MADAM, MEV, Spotfinder

[Grid Computing](#) - BLAST, HMMsearch, Panda, Genomic Assembly, Mummer

[Sequencing / Finishing](#) - AMOS, BAMBUS, AutoEditor, Assembler, Lucy

[Alignment](#) - MUMmer, AAT package

[Gene Finding / Annotation](#) - DAGchainer, MANATEE, PIRATE, PASA, GlimmerHMM, etc.,

[Other](#) - TGICL, TIGR McCoder, Scheduler

Data Disclaimer | © 2005 - 2006 The Institute for Genomic Research | Contact Us

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

:
?

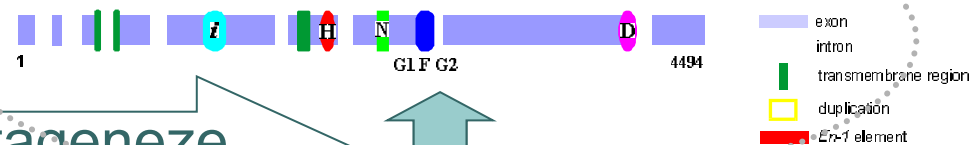


1

„Reverzně genetický“ přístup

5'TTATATATATATATTAATAATAATAATAA
GAACAAAAAGAAAATAAATA...3'

BIOINFORMATIKA

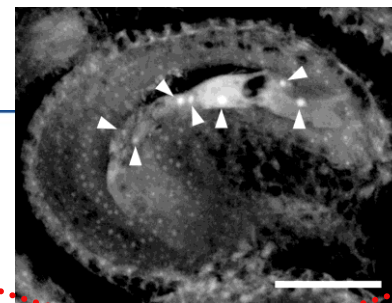


FUNKČNÍ GENOMIKA



En-1
ga att ca agt cgt CAC TACAAGA TCT TGTASTG cgt gag act
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