

# CG020 Genomika

## Bi7201 Základy genomiky

### Přednáška 1

#### Úvod do bioinformatiky

Jan Hejátko

**Funkční genomika a proteomika rostlin,**  
Mendelovo centrum genomiky a proteomiky rostlin,  
Středoevropský technologický institut (CEITEC), Masarykova univerzita, Brno  
[hejatko@sci.muni.cz](mailto:hejatko@sci.muni.cz), [www.ceitec.muni.cz](http://www.ceitec.muni.cz)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Osnova

- Schéma předmětu
- Definice
- 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 předmětu

- **Kapitola 01** (CG020 , Bi7201)
  - Úvod do bioinformatiky
  
- **Kapitola 02** (CG020 , Bi7201)
  - Identifikace genů
  
- **Kapitola 03** (CG020 , Bi7201)
  - Přístupy reverzní genetiky
  
- **Kapitola 04** (CG020 , Bi7201)
  - Přístupy genetiky přímé

# Schéma předmětu

- **Kapitola 05** (CG020 , Bi7201)
  - Přístupy funkční genomiky
  
- **Kapitola 06** (CG020 , Bi7201)
  - Protein-protein interakce a jejich analýza
  
- **Kapitola 07** (CG020)
  - Moderní postupy funkční genomiky
  
- **Kapitola 08** (CG020)
  - Strukturní genomika

# Schéma předmětu

- **Kapitola 09** (CG020)
  - Lokalizace genů a genových produktů v buňce
  
- **Kapitola 10** (CG020)
  - Genomika a systémová biologie
  
- **Kapitola 11** (CG020)
  - Praktické aspekty funkční genomiky
  
- **Kapitola 12** (CG020)
  - Lokalizace genů a genových produktů v buňce

# Literatura

- Zdrojová literatura ke kapitole I:
  - **Bioinformatics and Functional Genomics**, 2009, Jonathan Pevsner, Willey-Blackwell, Hoboken, New Jersey  
<http://www.bioinfbook.org/index.php>
  - **Úvod do praktické bioinformatiky**, Fatima Cvrčková, 2006, Academia, Praha
  - **Plant Functional Genomics**, ed. Erich Grotewold, 2003, Humana Press, Totowa, New Jersey

# Osnova

- Schéma předmětu
- Definice

# 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

„Reverzně genetický“ přístup

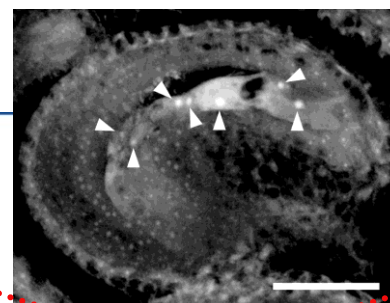
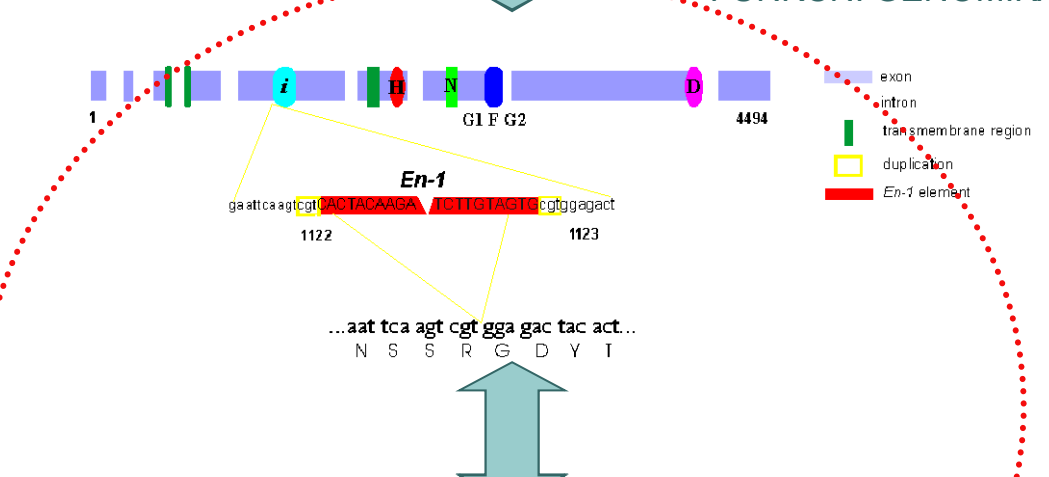
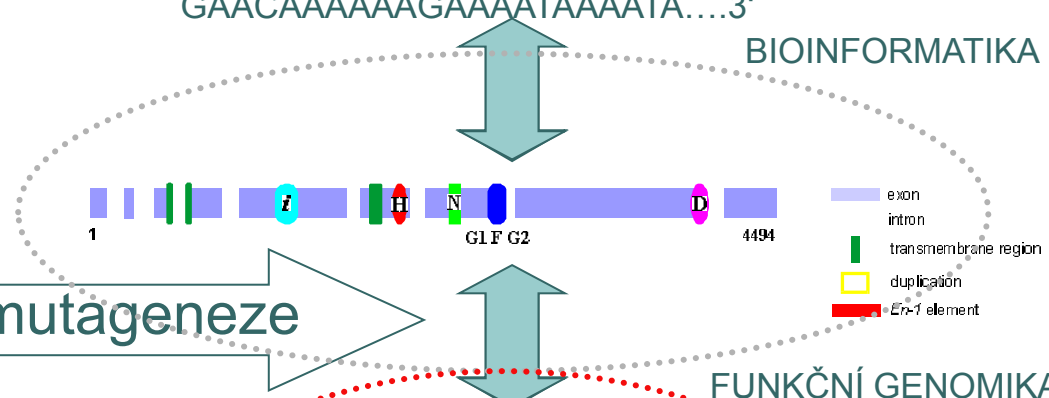
5'TTATATATATATATTAATAATAATAATAA  
GAACAAAAAGAAAATAAATA...3'



3



1



VOJE VZDĚLÁVÁNÍ

entace je spolufinancována  
ropským sociálním fondem  
a státním rozpočtem České republiky



EVROPSKÁ UNIE



MLÁDEŽE A TĚLOVÝCHOVY

pro konkurenceschopnost

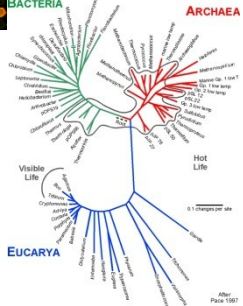
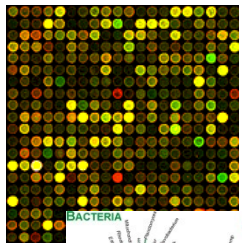


# Osnova

- Schéma předmětu
- Definice
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY

# Bioinformatika

- **Definice bioinformatiky** (podle NIH vědeckého a technologického konsorcia pro biomedicínské informace)



**Výzkum, vývoj nebo aplikace výpočetních nástrojů a přístupů za účelem zvyšování rozvoje využití biologických, lékařských, dat o chování nebo zdraví, včetně těch, které umožňují taková data získávat, ukládat, organizovat, archivovat, analyzovat nebo vizualizovat.**

# What is bioinformatics?

- Interface of biology and computers
- Analysis of proteins, genes and genomes using computer algorithms and computer databases
- Genomics is the analysis of genomes. The tools of bioinformatics are used to make sense of the billions of base pairs of DNA that are sequenced by genomics projects.

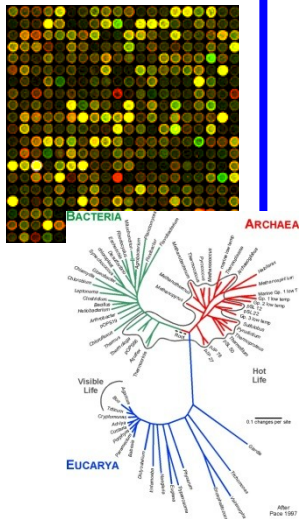
J. Pevsner,  
<http://www.bioinfbook.org/index.php>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Bioinformatika



- **Bioinformatika ve funkční genomice**
  - **Zpracování a analýza sekvenačních dat**
    - Identifikace referenčních sekvencí
    - Identifikace genů
    - Identifikace homologů, ortologů a paralogů
    - Korelační analýzy mezi sekvencemi a fenotypy (včetně člověka)
  - **Zpracování a analýza transkripčních dat**
    - Transkripční profilování pomocí DNA čipů nebo next-gen sekvenování
  - **Vyhodnocování experimentálních dat a predikce nových regulací v přístupech systémové biologie**
    - Matematické modelování genových regulačních sítí

# Osnova

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

# Spektrum on-line zdrojů

<b>EMBnet National Nodes</b>		
Vienna Biocenter	Austria	<a href="http://www.at.embnet.org/">http://www.at.embnet.org/</a>
BEN	Belgium	<a href="http://www.be.embnet.org/">http://www.be.embnet.org/</a>
BioBase	Denmark	<a href="http://biobase.dk/">http://biobase.dk/</a>
CSC	Finland	<a href="http://www.fi.embnet.org/">http://www.fi.embnet.org/</a>
INFOBIOGEN	France	<a href="http://www.infobiogen.fr/">http://www.infobiogen.fr/</a>
GENIUSnet	Germany	<a href="http://genome.dkfz-heidelberg.de/biounit/">http://genome.dkfz-heidelberg.de/biounit/</a>
IMBB	Greece	<a href="http://www.imbb.forth.gr/">http://www.imbb.forth.gr/</a>
HEN	Hungary	<a href="http://www.hu.embnet.org/">http://www.hu.embnet.org/</a>
INCBI	Ireland	<a href="http://acer.gen.tcd.ie/">http://acer.gen.tcd.ie/</a>
INN	Israel	<a href="http://dapsas.weizmann.ac.il/bcd/inn.html">http://dapsas.weizmann.ac.il/bcd/inn.html</a>
IEN-ADR	Italy	<a href="http://bio-www.ba.cnr.it:8000/BioWWW/Bio-WWW.htm">http://bio-www.ba.cnr.it:8000/BioWWW/Bio-WWW.htm</a>
CAOS/CAMM	Netherlands	<a href="http://www.caos.kun.nl/">http://www.caos.kun.nl/</a>
Bio	Norway	<a href="http://www.no.embnet.org/">http://www.no.embnet.org/</a>
IBB	Poland	<a href="http://www.ibb.waw.pl/">http://www.ibb.waw.pl/</a>
IGC	Portugal	<a href="http://www.igc.gulbenkian.pt/">http://www.igc.gulbenkian.pt/</a>
GeneBee	Russia	<a href="http://www.genebee.msu.su/">http://www.genebee.msu.su/</a>
CNB-CSIC	Spain	<a href="http://www.es.embnet.org/">http://www.es.embnet.org/</a>
BMC	Sweden	<a href="http://www.embnet.se/">http://www.embnet.se/</a>
SIB	Switzerland	<a href="http://www.ch.embnet.org/">http://www.ch.embnet.org/</a>
SEQNET	UK	<a href="http://www.seqnet.dl.ac.uk/">http://www.seqnet.dl.ac.uk/</a>
<b>EMBnet Specialist Nodes</b>		
MIPS	Germany	<a href="http://www.mips.biochem.mpg.de/">http://www.mips.biochem.mpg.de/</a>
ICGEB	Italy	<a href="http://www.icgeb.trieste.it/">http://www.icgeb.trieste.it/</a>
Pharmacia Upjohn	Sweden	<a href="http://www.pnu.com/">http://www.pnu.com/</a>
F.Hoffmann-La Roche	Switzerland	<a href="http://www.roche.com/">http://www.roche.com/</a>
EBI	UK	<a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a>
HGMP-RC	UK	<a href="http://www.hgmp.mrc.ac.uk/">http://www.hgmp.mrc.ac.uk/</a>
Sanger	UK	<a href="http://www.sanger.ac.uk/">http://www.sanger.ac.uk/</a>
UMBER	UK	<a href="http://www.bioinf.man.ac.uk/dbbrowser">http://www.bioinf.man.ac.uk/dbbrowser</a>
<b>EMBnet Associate Nodes</b>		
IBBM	Argentina	<a href="http://sol.biol.unlp.edu.ar/embnet">http://sol.biol.unlp.edu.ar/embnet</a>
ANGIS	Australia	<a href="http://www.angis.su.oz.au/">http://www.angis.su.oz.au/</a>
CBI	China	<a href="http://www.cbi.pku.edu.cn/">http://www.cbi.pku.edu.cn/</a>
CIGB	Cuba	<a href="http://bio.cigb.edu.cu/">http://bio.cigb.edu.cu/</a>
CDFO	India	<a href="http://salarjung.embnet.org.in/">http://salarjung.embnet.org.in/</a>
SANBI	South Africa	<a href="http://www.sanbi.ac.za">http://www.sanbi.ac.za</a>
<b>USA Information Providers</b>		
NCBI	USA	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
NLM	USA	<a href="http://www.nlm.nih.gov/">http://www.nlm.nih.gov/</a>
NIH	USA	<a href="http://www.nih.gov/">http://www.nih.gov/</a>



# Spektrum on-line zdrojů

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

The image shows a screenshot of the EMBL-EBI website interface. At the top, there is a search bar with the text "Nucleotide sequences" and a "Go" button. Below the search bar, the EMBL-EBI logo is displayed, along with the text "European Bioinformatics Institute". A navigation menu includes links for "EBI Home", "About EBI", "Research", "Services", "Toolbox", "Databases", "Downloads", and "Submissions".

Below the navigation menu, a tree diagram illustrates the structure of the "Services Overview". The tree is rooted at "Services Overview" and branches into several categories:

- Databases**
  - Database Browsing
    - SRS
  - Nucleotide Databases
    - EMBL Nucleo. Sequence
    - Ensembl
    - Genomes Server
    - Genome MOT
    - EMBL-Align
    - Simple Queries
    - dbSTS Queries
    - Parasites
    - Mutations
    - IMG
  - Protein Databases
    - SWISS-PROT
    - TrEMBL
    - InterPro
    - CluSTR
    - IPI
    - GQA
    - Proteome Analysis
    - HPI
    - IntEnz
- Toolbox**
  - Homology & Similarity
    - Fasta
    - WU-Blast2
    - NCBI-Blast2
    - Blast2\_EVEC
    - Genome/Proteome Fasta
    - MPsrch
    - Scanps2.3
    - Parasite-Blast
    - EGI-Blast
    - SNP-Fasta3 Server
  - Prot. Function. Analysis
    - CluSTR Search
    - InterPro Scan
    - FingerPRINTScan
    - ppsearch
    - Gene Quiz
    - Pratt
    - Radar
- Submissions**
  - EMBL via WEBIN
  - EMBL-Info. Submitters
  - SWISS-PROT
  - Webin-Align
  - PDB-AutoDep.
  - M/AMExpress
  - IMG/HLA
  - Sequin Software
- Downloads**
  - FTP Server
  - Database Repository
  - Software Repository
  - Downloads Help Files
- Services Help**

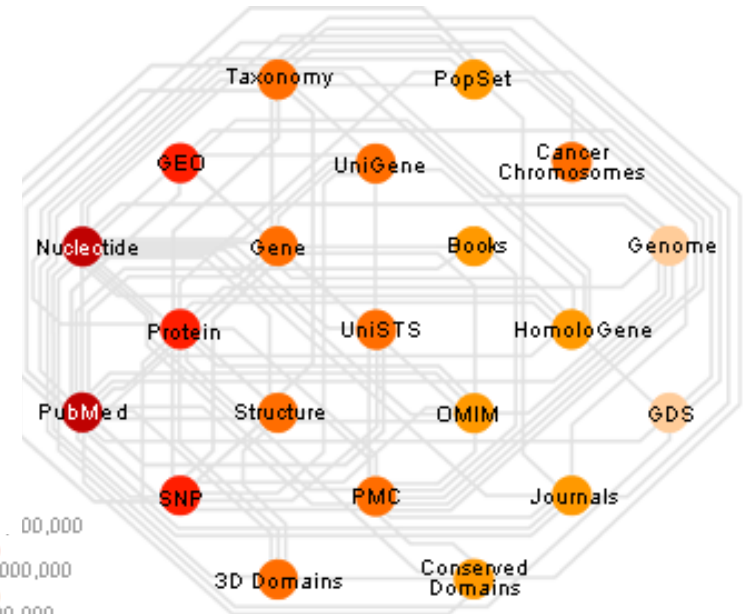


# Spektrum on-line zdrojů

□ NCBI <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI website interface. At the top, there's a navigation bar with 'NCBI', 'Resources', and 'How To'. Below it is a search bar labeled 'All Databases'. The main content area is divided into several sections:
 

- Welcome to NCBI:** A brief introduction to the center's mission.
- Get Started:** A list of links for 'Tools', 'Downloads', 'How-To's', and 'Submissions'.
- Popular Resources:** A list of key databases and services like PubMed, Bookshelf, and UniGene.
- NCBI YouTube channel:** A section promoting video tutorials.
- NCBI Announcements:** A section for news and updates.



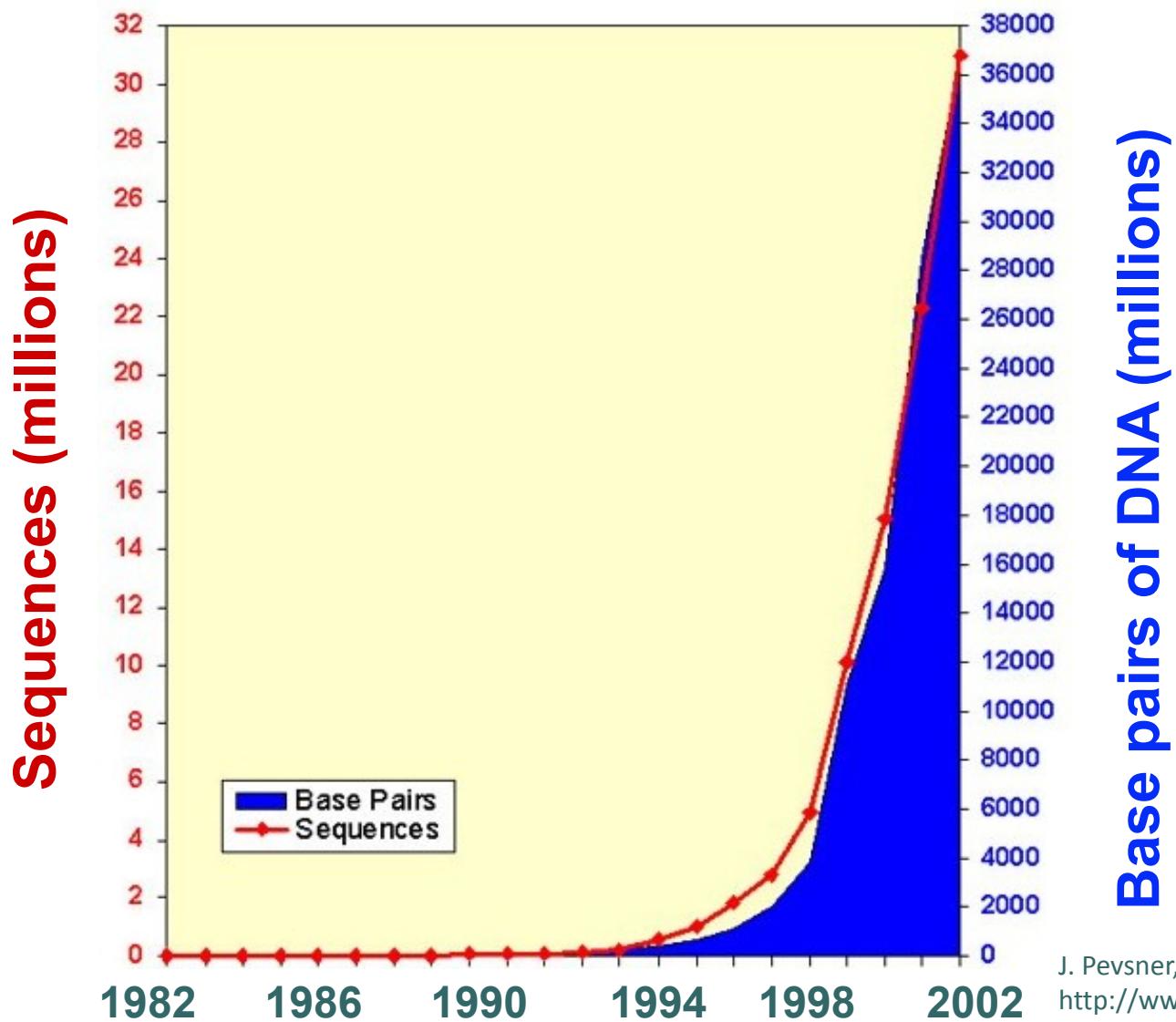
# Osnova

- Schéma předmětu
- Definice
- 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 10<sup>6</sup> záznamů o zhruba 33 x 10<sup>9</sup> bp
  - srpen 2005 100 x 10<sup>9</sup> bp ze 165.000 organizmů

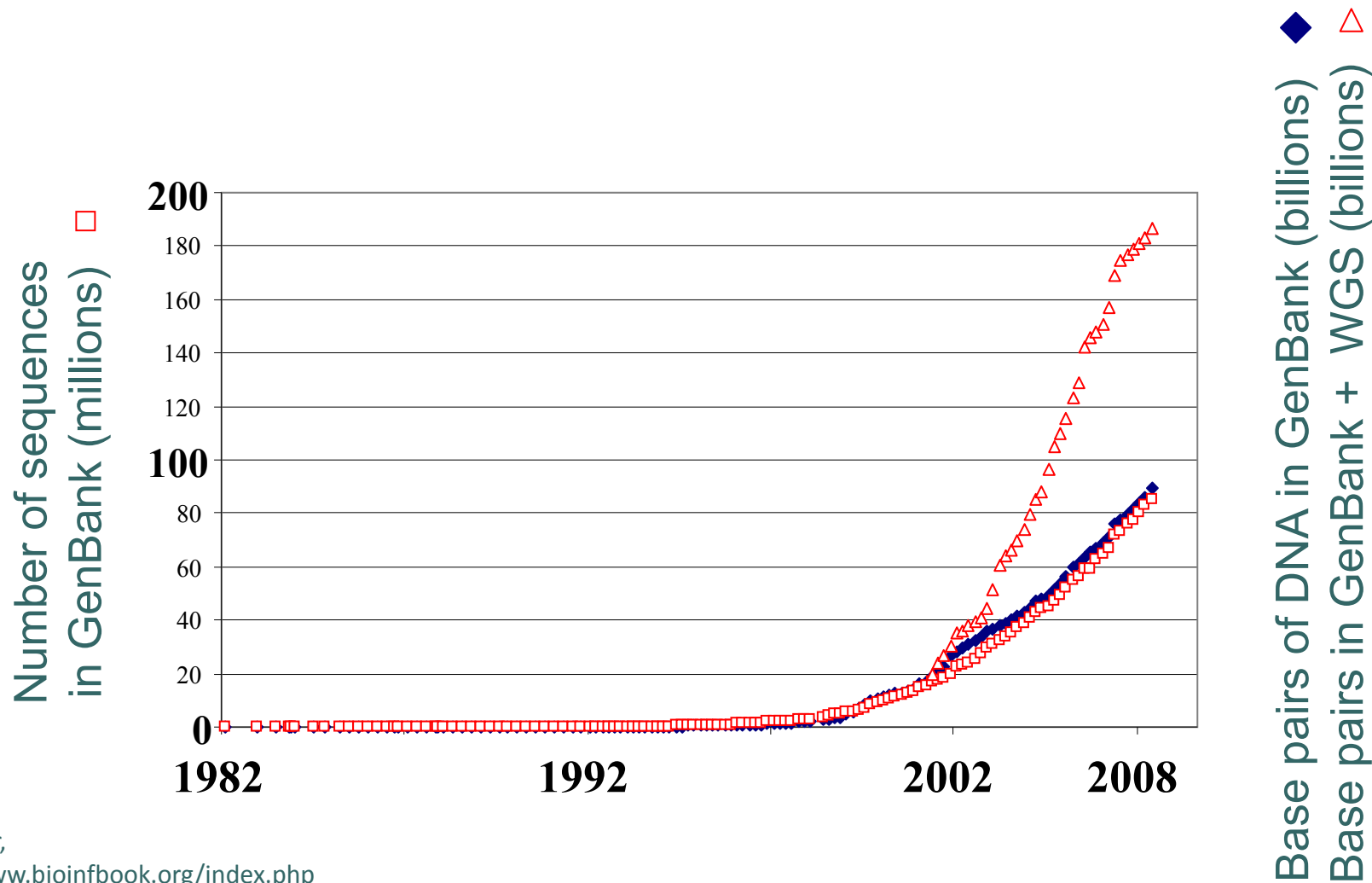
# Growth of GenBank



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Growth of GenBank + Whole Genome Shotgun (1982-November 2008): we reached 0.2 terabases



J. Pevsner,  
<http://www.bioinfbook.org/index.php>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# WGS

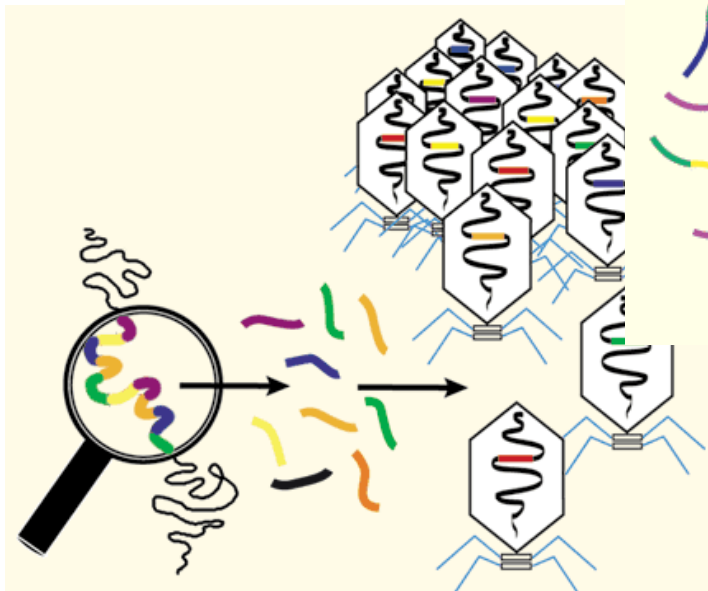
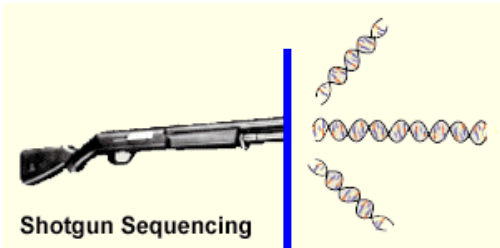


Fig 1: Genomic DNA is fragmented, ligated into viral DNA and packaged into viral particles to create a library

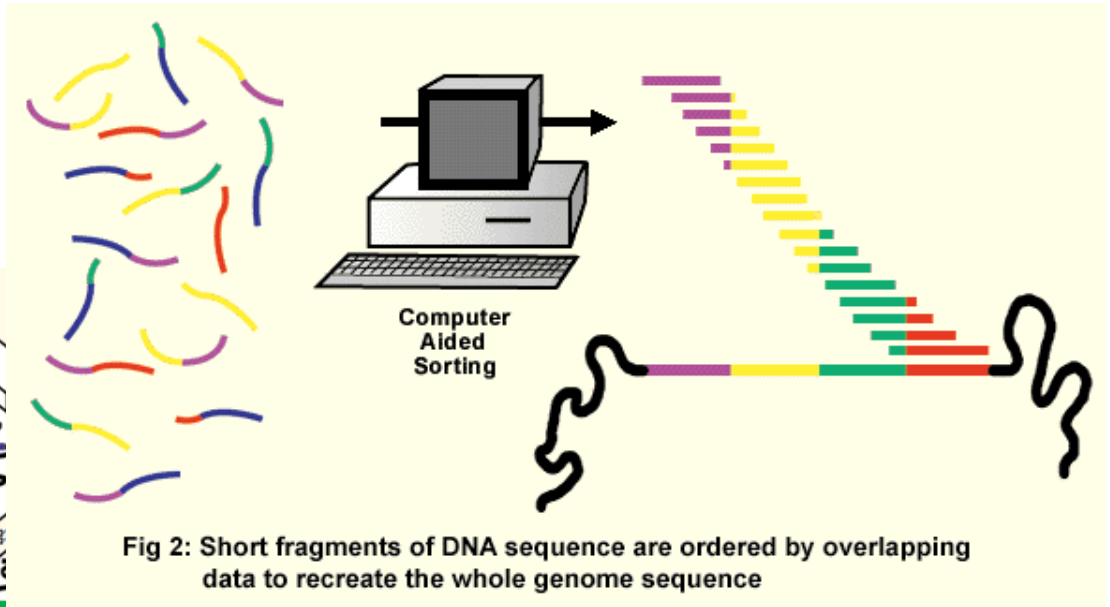
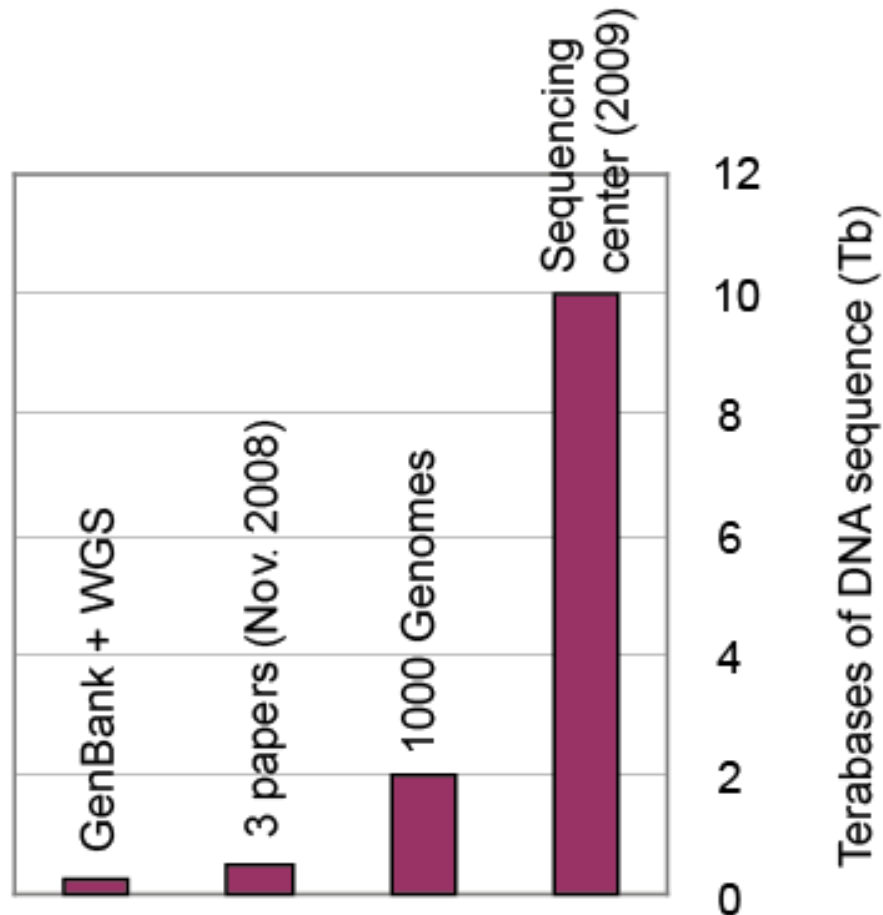


Fig 2: Short fragments of DNA sequence are ordered by overlapping data to recreate the whole genome sequence

Interactive concepts in biochemistry, Rodney Boyer, Wiley, 2002, <http://www.wiley.com//college/boyer/0470003790/>



# Arrival of next-generation sequencing: In two years we have gone from 0.2 terabases to 71 terabases (71,000 gigabases) (November 2010)



J. Pevsner,  
<http://www.bioinfbook.org/index.php>

# 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 homepage with a navigation menu on the left, a search bar at the top, and a main content area with sections for 'Welcome to NCBI', 'Get Started', and 'Popular Resources'. The 'Get Started' section lists links for Tools, Downloads, How-To's, and Submissions. The 'Popular Resources' section lists various databases and tools like PubMed, Bookshelf, and BLAST.

**NCBI Home**

- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

**Welcome to NCBI**

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

**Get Started**

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

**Popular Resources**

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

**NCBI YouTube channel**

Learn how to get the most out of NCBI tools and databases with video tutorials on the NCBI YouTube Channel. [GO](#)

**NCBI Announcements**

- New version of GenBank available
- An integrated, downloadable tool for viewing and analyzing sequence data
- NCBI's July Newsletter

# Primární databáze

**Gene symbol** virA  
**Gene description** two-component VirA-like sensor kinase  
**Locus tag** pTI\_125  
**Gene type** protein coding  
**RefSeq status** PROVISIONAL  
**Organism** *Agrobacterium tumefaciens* (old-name: *Agrobacterium tumefaciens*, gb-synonym: *Rhizobium radiobacter*)  
**Lineage** Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium; Agrobacterium tumefaciens complex

**Genomic context**  
**Location:** plasmid: Ti  
**Sequence:** NC\_002377.1 (145694..148183)

**Genomic regions, transcripts, and products**  
Genomic Sequence: NC\_002377

**Related articles**

1. [Sequence analysis of the virA locus of Agrobacterium tumefaciens octopine Ti plasmid pTI15955](#), Schrammeyer B, et al. J Exp Bot. 2000 Jun. PMID 10948245.
2. [The virA promoter is a host-range determinant in Agrobacterium tumefaciens](#), Turk SC, et al. Mol Microbiol. 1993 Mar. PMID 8469115.
3. [Characterization of the virA locus of Agrobacterium tumefaciens: a transcriptional regulator and host range determinant](#), Leroux B, et al. EMBO J. 1987 Apr. PMID 3595559.
4. [Analysis of the complete nucleotide sequence of the Agrobacterium tumefaciens virB operon](#), Thompson DV, et al. Nucleic Acids Res. 1988 May 25. PMID 2837739.

**GeneRIFs: Gene References Into Functions** [What's a GeneRIF?](#)  
Submit: [New GeneRIF](#) [Correction](#)

# Primární databáze

The screenshot displays a web browser window with the URL [www.ncbi.nlm.nih.gov/geo/1224316](http://www.ncbi.nlm.nih.gov/geo/1224316). The main content shows a genomic map for NC\_002377.1 (145K..148K, 2.9Kbp). A red bar represents the gene NP\_059797.1. A tooltip window is open over this gene, providing the following information:

- NP\_059797.1**
- NP\_059797.1: two-component VirA-like sensor kinase
- total range: NC\_002377.1 (145,694..148,183)
- total length: 2,490
- strand: plus
- protein product length: 829
- Links & Tools**
- GenBank View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)
- FASTA View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)
- BLAST Genomic: [NC\\_002377.1 \(145,694..148,183\)](#)
- Graphical View: [NP\\_059797.1](#)
- BLAST Protein: [NP\\_059797.1](#)
- BLINK Results: [NP\\_059797.1](#)

Below the gene information, there are sections for **Bibliography** and **Related articles in PubMed**. The browser's taskbar at the bottom shows several open applications, including Firefox, Windows Media Center, and various document editors.



# Primární databáze

```
/translation="MNGRYSPTRODFKTKGAKPWSILALIYAAMIAPFMAVASWQDNMT  
TQAILSQLRSINADSASLQRDVLRAHTCTVANYRPIISRLGALRNKLEDLKQLFRQSH  
IVSEBNAQLLRQLEVSLSADAAVAAPGAQNVRLQDSIASPTRALSSSLFGKASTDQT  
LEKPTELASMMQLQFLRQSPAISFPISELELELQKQRLDEAFVILAREGPIILSLL  
PQVKDLVNMQISTDIAEMLQRBCLEVYSLKNVEERSARIPLSSASVGLCLYIITL  
VYLRKKTDWLARRLDYELIKEIIVCFBGEAATTSSAQALRIIQRPFDADTCALAL  
VDHRRWAVETFOAKHFKPVWDSVLRRIVSRTEKADERATVFRILSSKKIVHLFLHIP  
GLSILLAHKSTDKLIAVCSLGYQSYRFPFCQGETQLLELATACLCHYIDVRRKQTECD  
VLARRLEHAQRLAVGTLAGGIAHFNNILGSLGHAELAQNSVSRTEVTRRYIDYII  
SSGDRAMLIIDQILTLRKRQEMIKPPSVSELVTEIAFLRMLPFPNIELEFRPDQMC  
SVI EGSPLRLQQLINICKNASQAMTANQIDIIIGQAPLPVKKILAHGVMPFGDYVL  
LSISDNQGGIPRAVLPHIFEPFFTTARNGGTGLGLASVHGHSAPAGYIDVSTVGH  
GTRFDIYLPFSSKPFVNPDSFFGRNKAPRGMGHI VALVFPDDLREAYRDKIAALGYE  
PVGFRTFNKIRDWISKGNADLVMDQASLPEDQSPNSVDLVKTAIIIGNDLKM  
LSREDVTADLYLFPKPISSRTMAHAHLTKIKT"
```

```
ORIGIN  
1 atgaacggaa gatattcacc gaecggcgag gattttaaga caggcgcgaa gcccttggct  
61 atattggccc ttatcgttgc tgaatgatt ttocogttca tggcggttgc gtccttggcag  
121 gacaatgcca ctaccacgga aatcctcage caactacgat cgat taacgc cgcacagcgc  
181 tcaactgcaag cagatgact cccgcctcac accggcaccgc tggcgaacta cgcgcccat  
241 atctccaggg tgggagctct cgggaagaat ctggaagatt tgaagcaatt atttagacaa  
301 tctcatattg taagtgagag caatgctgct caactgctac gccagctaga agtctctcta  
361 aatcggctgt acgcgcggtg cgcgccttt ggtgcgcaaa atgtacgct gcaagattcg  
421 ctggcagctt tcaactcgtg ttgagcagc ctccagaa aagcctcaac cgtacagact  
481 ttgaaaaaac caacagaatt ggttagcagc atgctccaat tctctggca accaaagccg  
541 gctatttcat toagatcag ccttgaacta gagaggtccc aaaaacaag cggcttggat  
601 gaagctcccg tgcgcaact tgcacgtgaa ggtcccatta tcttatcgt tttgccacag  
661 gtgaaagatc tgggtaacat gattcagcgc tctgacaccg cagaatggc gtagatgctg  
721 cagcgcagat gtttggaggt ctatagctt aaaaatgtag aggagcggag cgcacgtatc  
781 ttcttgggt cccctcagc ggttcttgc ctctacatca tcccttagt ctataggcta  
841 cgcacacaaa cagatgggt agcgcggcgt ttgatatac aagagctaat caaagagatc  
901 ggagtagttt ttgaaagtga ggcggccacc acgtcgtccg cgcacagctc actctgtatt  
961 atcagcgcct tcttggatgc cgtacgtgce gcttagctc tagtggacca tgacgtaga  
1021 tgggctgtgc aaacattcgg tgcgaaacac ccaaacctgt tctgggacga cagcgtgcta  
1081 cgcgaaatag tctctcgtac caaagcggac gaacgggcca cggatctccg catcatatcg  
1141 tgcacacaaa togtacattt gccctcogaa atccagctc tctcagatc actggctcac  
1201 aaatccacag ataaactaat tgcggttgt tcaactgggt accaaagcta tgcctctoga  
1261 ccttgcacag gcaaaatca gctcttggaa ctgcacaccg cctgcctctg tcatatatac  
1321 gatgtcggc gtaagcagac cgaatgagac gtttggcca gacgatgga gcatgcccga  
1381 cgccttggag cagttggtac acttgcggc ggaatgacac atgaatttaa taacatttg  
1441 ggtcacaatc cgggcaagc agaattagca caaacctcgg tctctgcaac atctgtcac  
1501 cgaagatata ttgactatc cattctgca ggcgacagag ccatgctcat tctgatcag  
1561 atctgacgc tgagccgaaa acaggagcgc atgatcaagc catttagtgt ctcagagctt  
1621 gtgaccgaaa toctcctct gctacgtatg gctctccgc caaacatgca gcttagttc  
1681 agatgtgac aaatcgagag cgtgatcgaa ggaagccgc ttgaactca acaggtacta  
1741 ataacatct gcaagaatgc tcccaagcc atgactgcaa atgctcaaat cgcacatc  
1801 atcagccaaag ctttttacc agttaagaaa attctgggc atggtttat gccocctggc  
1861 gactatgttc tctatctat tagcgaact ggtggaggca tcccgagc tgtttacc  
1921 cacattttg aacctctct tacgacacga gctgcacag gtggaaacgg tctcggcctt  
1981 gcttctgtgc atggtcatc cagcgcgttt cggggttaca tgcagcttag ttoactgtt  
2041 gggcatggga cgcgcttga catttatctc cctcctgtt ctaaagaaec cgtaaatca  
2101 gacagttttt cggcccgcaa taaggccacc cgtgaaacgc gggagattgt ggcactgtt  
2161 gagccgatg acctcctgag gtagcgtat gaagacaaga tgcgcctct aggatagag  
2221 cggctgggt tctgactct taatgaactt cgcgatgga tttcaaaag caatgaagc  
2281 gatctggtca tggtcgaca agcgtctctt cctgaagatc aaagtctaa tctcgtggt  
2341 ttagtgtca agacgcctc catcatcatt ggcgaaatg atctcaaat gacccttca
```



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# What is an accession number?

An accession number is label that used to identify a sequence. It is a string of letters and/or numbers that corresponds to a molecular sequence.

Examples (all for retinol-binding protein, RBP4):

X02775	GenBank genomic DNA sequence
NT_030059	Genomic contig
Rs7079946	dbSNP (single nucleotide polymorphism)

**DNA**

N91759.1	An expressed sequence tag (1 of 170)
NM_006744	RefSeq DNA sequence (from a transcript)

**RNA**

NP_007635	RefSeq protein
AAC02945	GenBank protein
Q28369	SwissProt protein
1KT7	Protein Data Bank structure record

**protein**

J. Pevsner,  
<http://www.bioinfbook.org/index.php>

# NCBI's important RefSeq project: best representative sequences

RefSeq (accessible via the main page of NCBI) provides an expertly curated accession number that corresponds to the most stable, agreed-upon “reference” version of a sequence.

RefSeq identifiers include the following formats:

Complete genome	NC_#####
Complete chromosome	NC_#####
Genomic contig	NT_#####
mRNA (DNA format)	NM_##### e.g. NM_006744
Protein	NP_##### e.g. NP_006735

J. Pevsner,  
<http://www.bioinfbook.org/index.php>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# RefSeq

The screenshot shows the NCBI RefSeq page for the gene 'two-component VirA-like sensor kinase'. The page is viewed in a Firefox browser window. The main content area is titled 'two-component VirA-like sensor kinase' and includes a 'Genome Annotation' section. Below this, there is a 'Reference assembly' section with a 'Genomic' subsection. The first entry is 'NC\_003065.3' with a range of 180831..183332 and download links for GenBank, FASTA, and Sequence Viewer. The 'mRNA and Protein(s)' section lists 'NP\_396486.1 two component sensor kinase [Agrobacterium tumefaciens str. C58]'. This entry includes UniProtKB/Swiss-Prot ID P18540 and three conserved domains: cd00075 (HATPase\_c), cd00082 (HisKA), and PRK13837 (two-component VirA-like sensor kinase). The PRK13837 domain has a location of 14-833 and a blast score of 2944. The 'Related Sequences' section is partially visible at the bottom.

# NCBI's RefSeq project: many accession number formats for genomic, mRNA, protein sequences

<u>Accession</u>	<u>Molecule</u>	<u>Method</u>	<u>Note</u>
AC_123456	Genomic	Mixed	Alternate complete genomic
AP_123456	Protein	Mixed	Protein products; alternate
NC_123456	Genomic	Mixed	Complete genomic molecules
NG_123456	Genomic	Mixed	Incomplete genomic regions
NM_123456	mRNA	Mixed	Transcript products; mRNA
NM_123456789	mRNA	Mixed	Transcript products; 9-digit
NP_123456	Protein	Mixed	Protein products;
NP_123456789	Protein	Curation	Protein products; 9-digit
NR_123456	RNA	Mixed	Non-coding transcripts
NT_123456	Genomic	Automated	Genomic assemblies
NW_123456	Genomic	Automated	Genomic assemblies
NZ_ABCD12345678	Genomic	Automated	Whole genome shotgun data
XM_123456	mRNA	Automated	Transcript products
XP_123456	Protein	Automated	Protein products
XR_123456	RNA	Automated	Transcript products
YP_123456	Protein	Auto. & Curated	Protein products
ZP_12345678	Protein	Automated	Protein products

J. Pevsner,  
<http://www.bioinfbook.org/index.php>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

NC\_002377.1: 145K..148K (2.9Kbp)

Genes

**NP\_059797.1**

NP\_059797.1: two-component VirA-like sensor kinase  
total range: NC\_002377.1 (145,694..148,183)  
total length: 2,490  
strand: plus  
protein product length: 829

**Links & Tools**

GenBank View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)  
FASTA View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)  
BLAST Genomic: [NC\\_002377.1 \(145,694..148,183\)](#)  
Graphical View: [NP\\_059797.1](#)  
BLAST Protein: [NP\\_059797.1](#)  
BLINK Results: [NP\\_059797.1](#)

**Bibliography**

**Related articles in PubMed**

# Primární databáze

Display Settings: FASTA

Showing 2.49kb region from base 145694 to 148183.

### Agrobacterium tumefaciens plasmid Ti, complete sequence

NCBI Reference Sequence: NC\_002377.1

[GenBank](#) [Graphics](#)

```
>gi|10955016:145694-148183 Agrobacterium tumefaciens plasmid Ti, complete sequence
ATGAACGGGAAGATATTCACCGACGCGGCAGGATTTAAGACAGCGCGGAAGCCTTGGCTATATTTGGCCC
TTATCGTTGCTGCAATGATTTTCGCGTTTCATGGCGTTGCGTCTGGCAGGACAAATGCGACTACCCAGGC
AATCTCAGCCAACATCAGATCGATTAACGCGCACAGCCCTCACTGCAGCGCATGTACTCCGCGCTCAC
ACGGCACCGTGGCGAATACCGCCCATTTATCTCCAGGCTGGGAGCTCTGCGGAAGAAATCTGAAAGATT
TGAAGCAATTTAGCAATCTCATATTTGAAGTGAAGCAATGCTGCTCAACTGCTACGCGCAGCTAGA
AGTGTCTAAATTCGGCTGACGCGCGCTCGCCGCTTTGGTGGCAAAATGTACGCTCAAAATTCG
CTGGCCAGTTTCACTCGTGTCTTGAGCAGTCTCCAGGAAAAGCCTCAACCGATCAGACTTTAGAAAAAC
CAACAGAATTTGGTAGCATGATGCTCCAATTTCTTCGGCAACCAAGCCCGCTATTTCAATTCGAGATCAG
CCTTGAATAGAGAGGCTCCAAAAACAACGCGCTTGATGAAGCTCCCGTGGCAGTACTTGCACGTGAA
GGTCCCATTTATCGCTTTTGGCCAGGTTGAAAGATCTGGTGAACATGATTCAGAGCTCTGACACCG
CAGAAATTCGGAGATGCTGCAGCGGAGTGTGGAGGCTATAGCTTGAATAATGTAGAGGACGGAG
CGCAGCTATCTTTCTGGTCCGCTTCAGTGGTCTTTGCCTCTACATCATCACTTAGTCTATAGGCTA
CGCAAAAAACCGATTTGGTTCAGCGCGGCTTTAGATTACGAAGAGCTAATCAAGAGATCGGAGTATGTT
TTGAAGTGAAGCGGCCACCACTGCTCGCGCAAGCTGCATTTGATTTTTCAGCGCTTTTGGATGC
CGATACGTCGCGCTTAGCTTAGTGGACCATGACCGTAGTGGGCTGTGAAAATTTGCTGCGAAACAC
CCAAAACCTGTGGGACGACAGCGTGTACGCGAAATAGTCTCTGTAACAAAGCGGACGAAACGGCGCA
CGGATTCGCGATCATGCTGCAAAAAATCGTACATTTGCTCTCGAAATTCAGGCTCTCTGATACT
ACTGGCTCAAAATCCACAGATAAATTAATGCGCTTTGCTCACTGGGTACCAAGCTTTCGCGCTCGA
CCTTGCCAGGGGAAATTCAGCTTTTGAATCGCCACCGCTGCTCTGACTATATCGATGTTGGCG
GTAAGCAGACCGAATGCGACGTTTGGCCAGACGATTGGAGCATGCGCAACGCTTGAGGCAATTTGATC
ACTTCCGCGCGAATAGCACATGAATTAATAACATTTTGGCTCAATCTCGGGCAGCAGAAATAGCA
CAAACTCGGTCTCGAACATCTGTACCCGAAAGATATATGACTATATCATTTGCTCAGGCGACAGAG
CCATGCTCATATCGATCAGATCTTGACGCTGAGCGGAAACAGGAGCGCATGATCAAGCCATTTAGTGT
CTCAGACTGTGACCGAAATCGCTCCCTTGTACGATGGCTTTCGCGCAACATCGAGCTTAGTTC
AGATTTGATCAAAATCGAGCGGTGATCGAAGGAAGCCCGCTTGAACCTCAACAGGTACTAATTAACAT
GCAAGATGCTTCCAAAGCCATGACTGCAAAATGTTCAATCGACATCATCAGCCAAAGCTTTTTTACC
AGTTAAGAAAATTCGGCGCATGGTGTATGCCACCTGGCGACTATGTTCTCCTATCTATAGCGACAAT
GGTGGAGGATTCGGAGGCTGTGTACCCACATTTTGAACCTTTTACGACACGAGCTCGCAACG
GTGAAACGGGCTCGGCTGCTGTGTCATGTTGATCAGCGGCTTTCGCGGTTACATCGAGCTTAG
TTCAACTGTGGGATGGGACCGCTTTGACATTTATCTCCCTCCGCTTCTAAGGAACCGGTAATCCA
GACAGTTTTTTCGGCCGCAATAAGGCAACCGCGTGGAAACGGGAGATTGCGCACTTTGAGCCCGATG
ACCTCCTCGGGGAGGCGTATGAAACAAGATCGCCGCTTAGGATATGAGCGGTCGGTTTTTCTGATCCTT
TAATGAAATTCGCGATTGGATTTCAAAAGGCAATGAAGCCGATCTGGTCATGTTGACCAAGCGCTCTCT
CCTGAAGATCAAACTCCTAATTCGGTGGATTTAGTGTCAAGACCGCCATCATCATTTGGCGAAATG
ATCTCAAAATGACCCCTTCAAGGGAGGATGTGACCGGACCTTTATCTCCGAGCGGATATCGTCCAG
AACTATGGCGCATGCAATCTAACAAAATCAAGACGATG
```



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/>

<a href="#">Expasy Home page</a>	<a href="#">Site Map</a>	<a href="#">Search Expasy</a>	<a href="#">Contact us</a>	<a href="#">Swiss-Prot</a>	<a href="#">PROSITE</a>	<a href="#">Proteomics tools</a>
----------------------------------	--------------------------	-------------------------------	----------------------------	----------------------------	-------------------------	----------------------------------

Hosted by [SIB Switzerland](#) | Mirror sites: [Australia](#) | [Belgium](#) | [Canada](#) | [China](#) | [Korea](#) | [Taiwan](#) | [USA](#)

Search  for

### prosite ScanProsite

This program allows to scan a protein sequence (either from [Swiss-Prot](#) or [TrEMBL](#), or provided by the user) for the occurrence of patterns and profiles stored in the [PROSITE](#) database, or to search protein databases with a user-entered pattern [[Reference](#) / [Download ps\\_scan, the standalone version](#)]. The program [PRATI](#) can be used to generate your own patterns. You may either:

- enter a PROSITE accession number or pattern to search the Swiss-Prot/TrEMBL and/or PDB databases with a pattern, **OR**
- enter a sequence or a Swiss-Prot/TrEMBL accession number to scan the sequence with all patterns, profiles and rules in PROSITE, **OR**
- fill in both fields to find all occurrences of a pattern or profile in a sequence.

Scan a protein for PROSITE matches	Search Swiss-Prot with a PROSITE entry
<p>Enter a Swiss-Prot/TrEMBL accession number (AC) (for example <b>P05130</b>) or a sequence identifier (ID) (for example <b>NOTC_DROME</b>), or a PDB identifier, or paste your own protein sequence in the box below:</p> <pre>SMVKVTKLYASEPTVFPVCLAPLVVPECTWISNMTTTE NLVKEVASFTDLRLSLVGERIENIKGPTVAKTHLSTGLA RVIDSVITNNDTQPTFRIQTQIAFLPLFVAVSTILQVQSVY ISRDGIMPFSYIARSNTSVAVFANSSNSRQDWTYTQTV DQLTGRLRHNGHSTRSQSLDVHTDWFQQAQSHNYTTAPVGT SLGSEMETLIQSVLSYERGLVSLGFPFKTLTEVNL NLHRELIMWTEDVLYRELSLNDFFISGSIQFEE NSLMSQCIPENCSSGVEVEIKRLRYQAQPCSVIRVSGVPL</pre> <p><input type="button" value="Clear"/></p> <p>and specify which motifs to use:</p> <p>Scan <input checked="" type="checkbox"/> patterns <input checked="" type="checkbox"/> profiles <input checked="" type="checkbox"/> rules [<a href="#">User Manual</a>] (You may also specify a PROSITE entry in the box to the right)</p> <p><input type="checkbox"/> Exclude <a href="#">patterns with a high probability of occurrence</a></p> <p>Your e-mail (optional): <input type="text"/> (will send results by e-mail)</p> <p><input type="checkbox"/> plain text output</p> <p><input type="button" value="START THE SCAN"/> <input type="button" value="RESET"/></p>	<p>Enter a PROSITE accession number (for example <b>PS01253</b>), or type your pattern in <a href="#">PROSITE format</a>:</p> <p>(leave this box blank to scan a sequence with the entire PROSITE database)</p> <p><input type="text"/></p> <p>and specify your search limits:</p> <ul style="list-style-type: none"><li>• The <input checked="" type="checkbox"/> Swiss-Prot <input type="checkbox"/> TrEMBL <input type="checkbox"/> TrEMBLnew <input type="checkbox"/> PDB databases (You may also specify a protein in the box to the left) <input checked="" type="checkbox"/> including splice variants</li><li>• The following taxa: <input type="text"/> (see <a href="#">NCBI Taxonomy</a>; separate multiple taxa with a semicolon, e.g. <i>Homo sapiens; Drosophila</i>. Not available for PDB.)</li><li>• Sequences with at least <input type="text"/> hits</li><li>• At most <input type="text"/> matches</li></ul> <p>Advanced options: <input type="checkbox"/> FASTA output <input type="checkbox"/> retrieve complete sequences</p> <p>allow at most <input type="text"/> X sequence characters to match a conserved position in the pattern</p> <p><a href="#">match mode</a>: <input type="text" value="greedy, overlaps, no includes"/> (for patterns, see <a href="#">help</a>)</p> <p><a href="#">randomize databases</a>: <input type="text" value="no"/> (to test a pattern, see <a href="#">help</a>)</p>



# Sekundární databáze

- databáze funkčních nebo strukturálních *motivů* získaných srovnáním primárních dat (sekvencí)
- PROSITE, <http://www.expasy.org/prosite/>

>[PDOC0003 PS00003](#) SULFATION Tyrosine sulfation site [rule] [Warning: rule with a high probability of occurrence].

```
571 - 585 nkeesstYeteisns
```

>[PDOC0004 PS00004](#) CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
744 - 747 RRvT
814 - 817 KRrS
```

>[PDOC0005 PS00005](#) PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
148 - 150 S#R
164 - 166 TgR
171 - 173 StK
219 - 221 SkK
369 - 371 TrR
460 - 462 SgK
513 - 515 SgR
585 - 587 SiR
602 - 604 TgK
652 - 654 TdK
716 - 718 SpR
726 - 728 SpK
747 - 749 TeK
794 - 796 S#R
854 - 856 ScK
864 - 866 StR
868 - 870 SeR
921 - 923 SpK
957 - 959 SvR
960 - 962 TgR
974 - 976 TeK
997 - 999 SrK
1002 - 1004 TgK
1018 - 1020 SgK
1031 - 1033 TgR
1119 - 1121 SkR
```

# Sekundární databáze

- databáze funkčních nebo strukturálních *motivů* získaných srovnáním primárních dat (sekvencí)
- PROSITE, <http://www.expasy.org/prosite/>

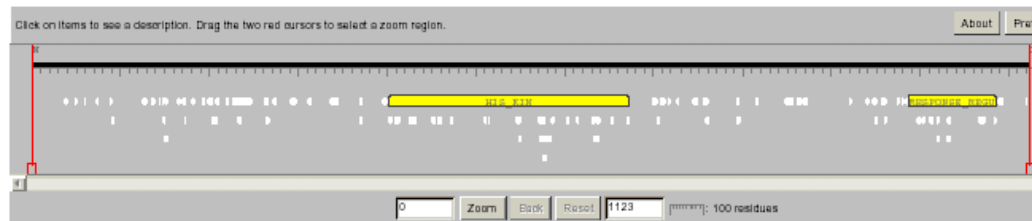
>[PDOC50109 PS50109 HIS\\_KIN](#) Histidine kinase domain [profile].

```
402 - 671 NASHDIRGALAGMKGLIDICRDGVKPGSDVDTTLNQVMVCAKDLVALLNSVLEMSKIESG
KMQLVREDFNLSKLLLEDVIDFYHFVAMKKGVDFVLDPHDgveEKPSNVRGDSGRLLKQILN
NLVSNVAVRFTVD--GHIAVRAWAQrpgensavvlasyppgvkfvkkmfckkkesatye
teienairnnaTMEFVFEVDDTGKGIHMEMRKSVPBNYVQVREtAQGHQGTGLGLGIWQ
SLVRLMGGEIRITDKAMGeKGTCPQPNVLLTT
```

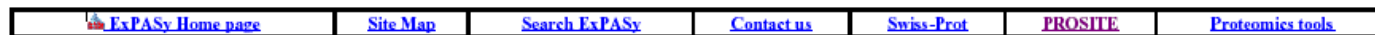
>[PDOC50110 PS50110 RESPONSE\\_REGULATORY](#) Response regulatory domain [profile].

```
987 - 1085 RVLVDDNPISRKVTGKLNKMGVSeVEQCDSGKEALRLVTEGLtqreeggsvdklpPDY
IFMDQMPEMDGYRATREIRkvekSYGVRTPIIAVSGHD-----
```

Graphical summary of hits (*java applet*)



98 hits with 12 PROSITE entries



# Sekundární databáze

- databáze funkčních nebo strukturálních *motivů* získaných srovnáním primárních dat (sekvencí)
- PRINTS, <http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/>



*PRINTS* is a compendium of protein **fingerprints**. A fingerprint is a group of conserved motifs used to characterise a protein family; its diagnostic power is refined by iterative scanning of a *SWISS-PROT/EMBL* composite. Usually the motifs do not overlap, but are separated along a sequence, though they may be contiguous in 3D-space. Fingerprints can encode protein folds and functionalities more flexibly and powerfully than can single motifs, full diagnostic potency deriving from the mutual context provided by motif neighbours. [References](#)

#### New:

- [SPRINT](#) - Search *PRINTS-S* (relational *PRINTS*)
- [prePRINTS](#) - Search *PRINTS*' automatic supplement
- [InterPro](#) - Search the integrated *InterPro* family database

#### Direct PRINTS access:

- [By accession number](#)
- [By PRINTS code](#)
- [By database code](#)
- [By text](#)
- [By sequence](#)
- [By title](#)
- [By number of motifs](#)
- [By author](#)
- [By query language](#)

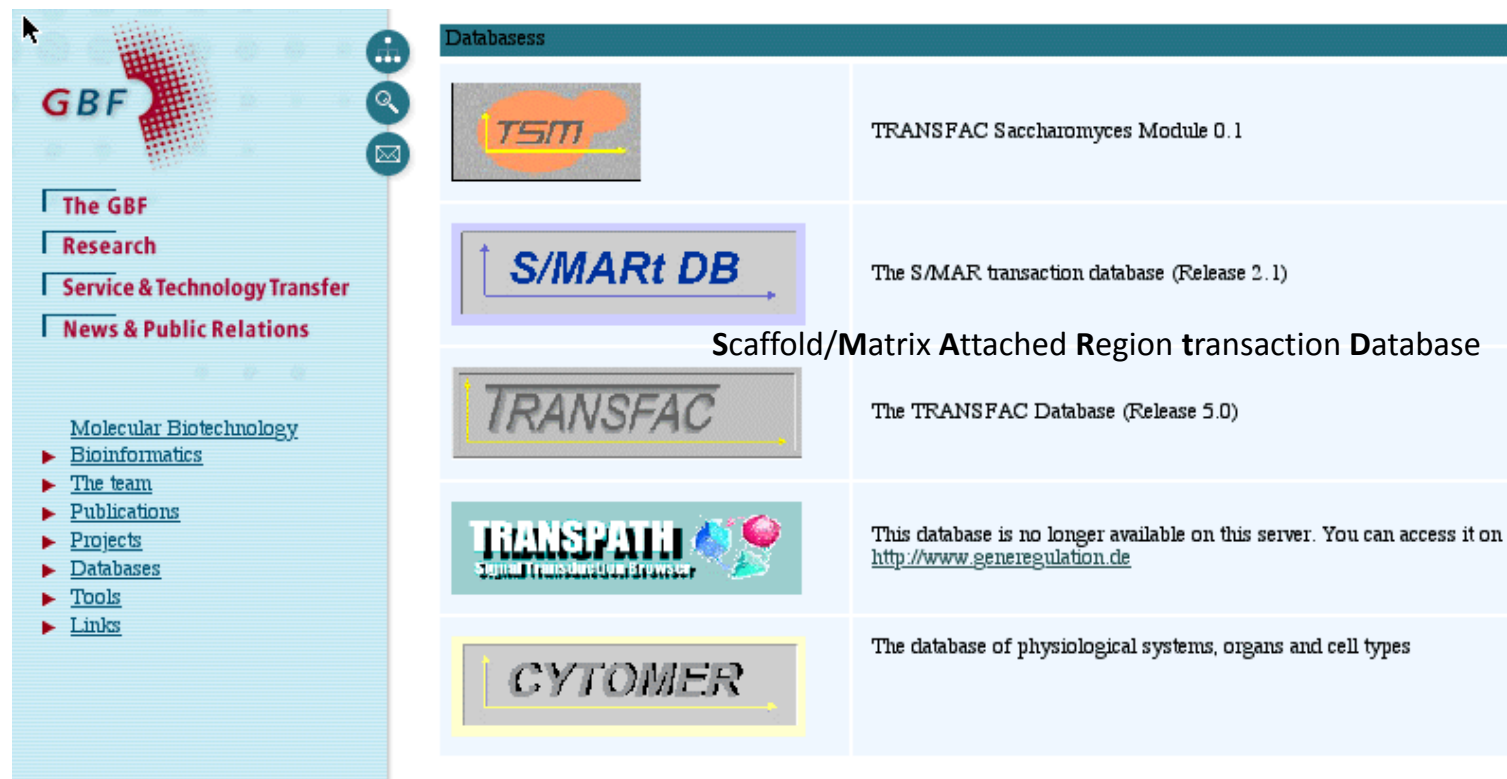
#### PRINTS search:

- [Search PRINTS with NEW FingerPRINTScan](#)
- [FPScan](#)
- [GRAPHScan](#)
- [MULScan](#)
- FingerPRINTScan binaries and source are available: [contact.scordis@bioinf.man.ac.uk](mailto:contact.scordis@bioinf.man.ac.uk)





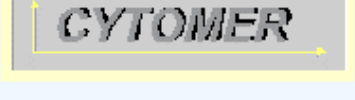


# Sekundární databáze

- 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) <b>Scaffold/Matrix Attached Region transaction Database</b>
	The TRANSFAC Database (Release 5.0)
	This database is no longer available on this server. You can access it on <a href="http://www.generegulation.de">http://www.generegulation.de</a>
	The database of physiological systems, organs and cell types

# Strukturální databáze

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

The screenshot shows the PDB website with the following elements:

- Navigation Links:** [DEPOSIT data](#), [DOWNLOAD files](#), [browse LINKS](#), [BETA TEST new features](#), [BETA mmCIF files](#)
- Current Holdings:** 19623 Structures, Last Update: 30-Dec-2002, PDB Statistics
- Molecule of the Month:** Cytochrome c (with a 3D ribbon diagram of the protein structure)
- Search the Archive:** Includes a search box for PDB ID or keyword, a "Find a structure" button, and options for "query by PDB id only", "match exact word", and "remove sequence homologues".
- News:** A news item dated 23-Dec-2002 titled "Happy Holidays from the PDB!" with a small image of lit candles.
- PDB Mirrors:** A list of mirror sites including San Diego Supercomputer Center, Rutgers University, National Institute of Standards and Technology, Cambridge Crystallographic Data Centre, National University of Singapore, Osaka University, Universidade Federal de Minas Gerais, and Max Delbrück Center for Molecular Medicine.
- Footer:** "OTHER SITES" with a link to a list of additional mirrors.

# Strukturální databáze

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

Structure Explorer - 1P5Y



## Structure Explorer - 1P5Y

**Title** The Structures Of Host Range Controlling Regions Of The Capsids Of Canine and Feline Parvoviruses and Mutants  
**Classification** Virus/Viral Protein  
**Compound** Mol\_Id: 1; Molecule: Coat Protein Vp2; Chain: A; Fragment: Sequence Database Residues 190-737; Engineered: Yes; Mutation: Yes  
**Exp. Method** X-ray Diffraction



[View Structure](#)

[Summary Information](#)

[View Structure](#)

[Download/Display File](#)

[Structural Neighbors](#)

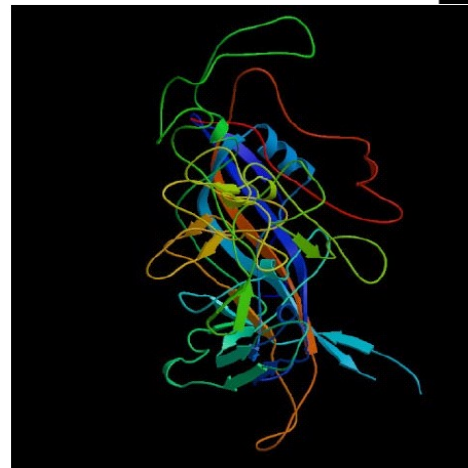
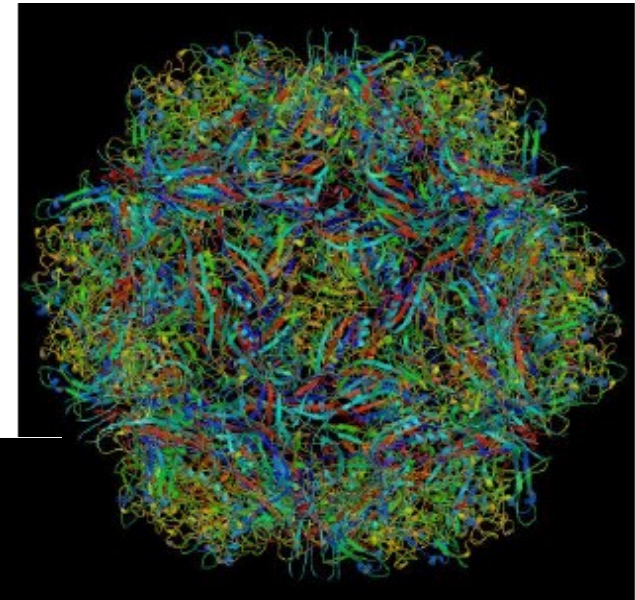
[Geometry](#)

[Other Sources](#)

[Sequence Details](#)

Explore

[SearchLite](#) [SearchFields](#)

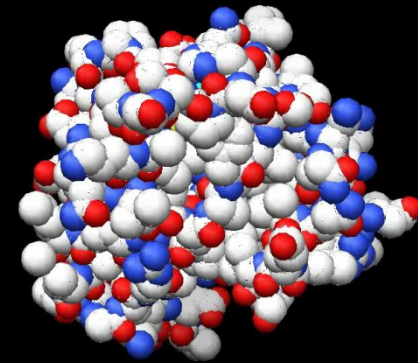
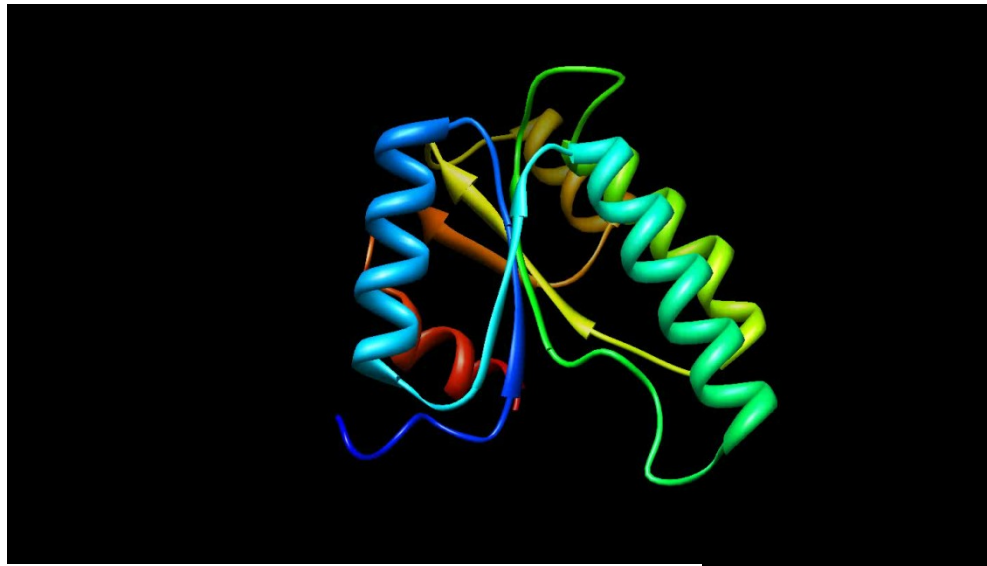


<http://www.rcsb.org/pdb/cgi/explore.cgi?job=graphics;pdbId=1P5Y;page=:pid=173561064349344&bio=1&opt=show&size=500>

12/29/2003

# Strukturální databáze

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



Pekárová et al., *Plant Journal* (2011)

# Osnova

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



# Genomové zdroje

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

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr21:33,031,597-33,041,570	enter position, gene symbol or search terms

[Click here to reset](#) the browser user interface settings to their defaults.

[track search](#) [add custom tracks](#) [track hubs](#) [configure tracks and display](#)

[Add your own custom tracks](#)

### Human Genome Browser - hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

#### Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gI000212	Displays all of the unplaced contig gi000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061,RH80175 15q11,15q13 rs1042522;rs1800370	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11 to 15q13, or SNPs rs1042522 and rs1800370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans, J.E.	Lists mRNAs deposited by co-author J.E. Evans

**U C S C**  
Homo sapiens  
(Graphic courtesy of CSHL)

# Genomové zdroje

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

The screenshot displays the UCSC Genome Browser interface. The main content area is divided into three sections:

- UCSC Genes:** Lists genes such as HBB, HBD, HBE1, HBE2, SATB1, SATB2, KLF1, KLF11, GATA1, MAFG, and HBG1, along with their genomic coordinates and descriptions.
- Non-Human RefSeq Genes:** Lists orthologous genes from other species, including LOC100174873, LOC100190885, LOC100136584, LOC100135791, and LOC100135791.
- Human Aligned mRNA Search Results:** Lists mRNA sequences like A18171, AF349114, M25079, M12050, Y00497, Y00500, M14574, M11428, and AY195861.
- Human Unaligned mRNA Search Results:** Lists unaligned mRNA sequences like S77349.
- Non-Human Aligned mRNA Search Results:** Lists orthologous mRNA sequences from other species like M19548, HG843793, HG843792, AY383035, AY775302, and AY329629.

The browser interface includes a search bar at the top, navigation buttons, and a taskbar at the bottom showing various open applications and system information.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Genomové zdroje

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

The screenshot displays the UCSC Genome Browser interface for Human chr11:5,246,696-5,248,301. The main track shows the RefSeq gene structure for HES1. Below it are tracks for Human ESTs, H3K27ac marks, and Digital DNase hypersensitivity clusters. The interface includes navigation controls (move, zoom in/out, track search) and a list of available tracks categorized into Mapping and Sequencing Tracks and Phenotype and Disease Associations. The browser window title is "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly".



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

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

**Human Gene HBB (uc001mae.1) Description and Page Index**

**Description:** Homo sapiens hemoglobin, beta (HBB), mRNA.  
**RefSeq Summary (NM\_000518):** The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon - gamma-G - gamma-A - delta - beta-3' [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##RefSeq-Attributes-END##  
Transcript\_exon\_combination\_evidence :: V00497.1, BU659180.1 [ECO:0000332] ##RefSeq-Attributes-END##  
Transcription Chromosome: chr11 Strand: - Size: 1,606 Start: 5,246,695 End: 5,248,301 Exon Count: 3  
Coding Size: 1,424 Start: 5,246,827 End: 5,248,251 Exon Count: 3

Page Index	Sequence and Links	UniProtKB Comments	Genetic Associations	CTD	Microarray
RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Pathways
Other Names	GeneReviews	Model Information	Methods		

Data last updated: 2011-12-21

**Sequence and Links to Tools and Databases**

Genomic Sequence (chr11:5,246,696-5,248,301)	mRNA (may differ from genome)	Protein (147 aa)			
Gene Sorter	Genome Browser	Protein FASTA	VisiGene	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneNetwork
Gepis Tissue	H-INV	HGNC	HPRD	Jackson Lab	MOPED
OMIM	PubMed	Reactome	Stanford SOURCE	Treefam	UniProtKB
Wikipedia					

**Comments and Description Text from UniProtKB**

**ID:** HBB\_HUMAN  
**DESCRIPTION:** RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName: Full=LVV-hemorphin-7;  
**FUNCTION:** Involved in oxygen transport from the lung to the various peripheral tissues.  
**FUNCTION:** LVV-hemorphin-7 potentiates the activity of bradykinin, causing a decrease in blood pressure.  
**SUBUNIT:** Heterotetramer of two alpha chains and two beta chains in adult hemoglobin A (HbA).  
**INTERACTION:** P69905:HBA2; NbExp=19, IntAct=EBI-715554, EBI-714680,  
**TISSUE SPECIFICITY:** Red blood cells.  
**PTM:** Glucose reacts non-enzymatically with the N-terminus of the beta chain to form a stable ketoamine linkage. This takes place slowly and continuously throughout the 120-day life span of the red blood cell. The rate of glycation is increased in patients with diabetes mellitus.  
**PTM:** S-nitrosylated; a nitric oxide group is first bound to Fe(2+) and then transferred to Cys-94 to allow capture of O(2).  
**PTM:** Acetylated on Lys-60, Lys-83 and Lys-145 upon aspirin exposure. PubMed 16916647 reports the identification of HBB acetylated on Lys-145 in the cytosolic fraction of HeLa cells. This may have resulted from contamination of the sample.  
**MASS SPECTROMETRY:** Mass=1310, Method=FAB, Range=33-42, Source=PubMed 1575724.  
**DISEASE:** Defects in HBB may be a cause of Heinz body anemias (HEIBAN) [MIM:140700]. This is a form of non-spherocytic hemolytic anemia of Dacie type 1. After splenectomy, which has little benefit, basophilic inclusions called Heinz bodies are demonstrable in the erythrocytes. Before splenectomy, diffuse or punctate basophilia may be evident. Most of these cases are probably instances of hemoglobinopathy. The hemoglobin demonstrates heat lability. Heinz bodies are observed also with the Ivemark syndrome (asplenia with cardiovascular anomalies) and with glutathione peroxidase deficiency.  
**DISEASE:** Defects in HBB are the cause of beta-thalassemia (B-THAL) [MIM:604131]. A form of thalassemia. Thalassemias are common monogenic diseases occurring mostly in Mediterranean and Southeast Asian populations. The hallmark of beta-thalassemia is an imbalance in globin-chain production in the adult HbA molecule. Absence of beta chain causes beta(0)-thalassemia, while reduced amounts of detectable beta globin causes beta(+)-thalassemia. In the severe forms of beta-thalassemia, the excess alpha globin chains accumulate in the developing erythroid precursors in the marrow. Their deposition leads to a vast increase in erythroid apoptosis that in turn causes ineffective erythropoiesis and severe microcytic hypochromic anemia. Clinically, beta-thalassemia is divided into thalassemia major which is transfusion dependent, thalassemia intermedia (of intermediate severity), and thalassemia minor that is asymptomatic.  
**DISEASE:** Defects in HBB are the cause of sickle cell anemia (SKCA) [MIM:603903]; also known as sickle cell disease. Sickle cell anemia is characterized by abnormally shaped red cells resulting in chronic anemia and periodic episodes of pain, serious infections and damage to vital organs. Normal red blood cells are round and flexible and flow easily through blood vessels, but in sickle cell anemia, the abnormal hemoglobin (called Hb S) causes red blood cells to become stiff. They are C-shaped and resembles a sickle. These stiffer red blood cells can lead to microvascular occlusion thus cutting off the blood supply to nearby tissues.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

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

Genomic Sequence Near Gene

### Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the [Table Browser](#) using the output format sequence.

**Sequence Retrieval Region Options:**

- Promoter/Upstream by 1000 bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by 1000 bases
- One FASTA record per gene.
- One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 0 extra downstream (3')
- Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

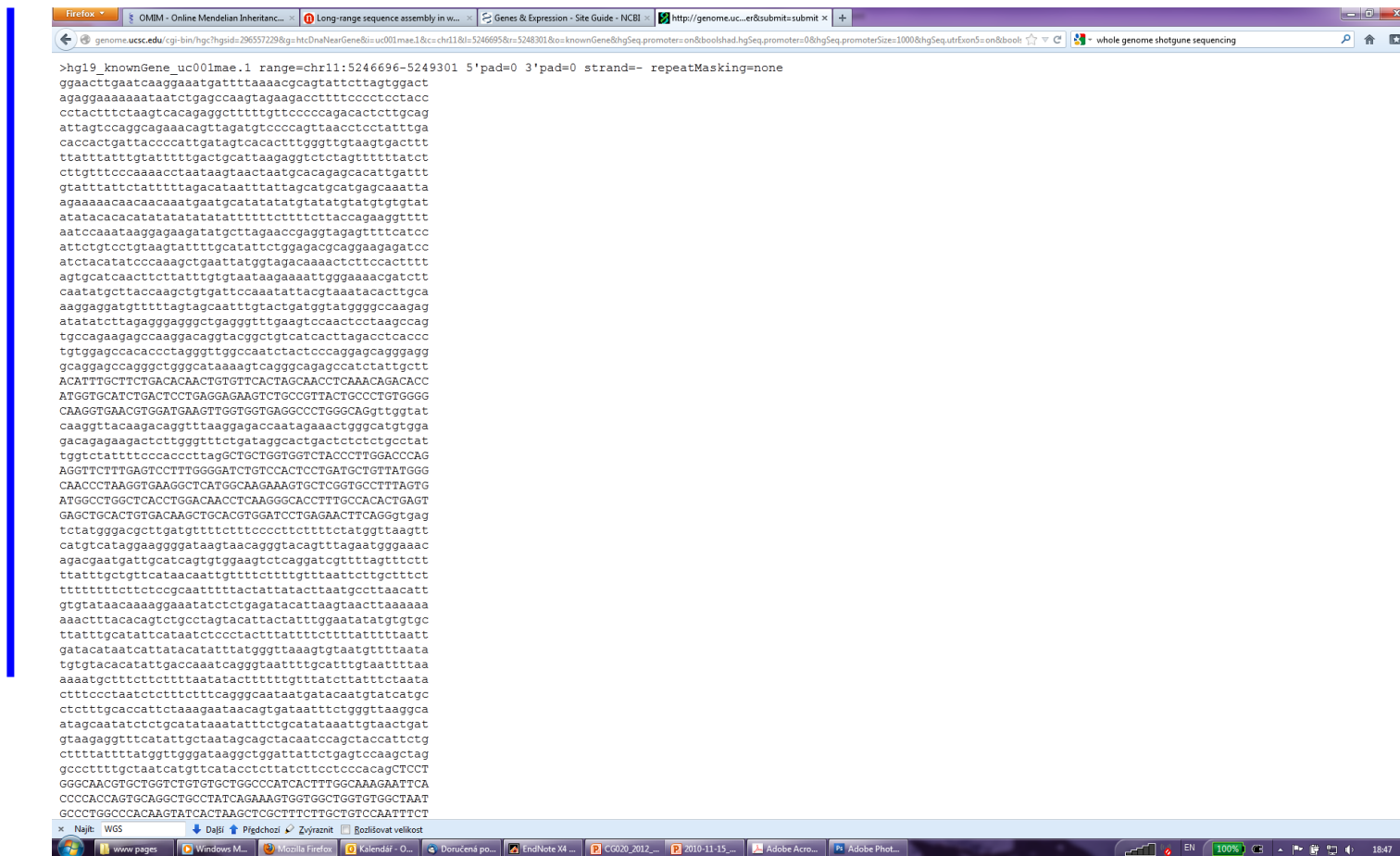
**Sequence Formatting Options:**

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats:  to lower case  to N

submit

# Genomové zdroje

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

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

The screenshot shows the TAIR website homepage. The browser window title is "TAIR - Home Page". The website features a search bar at the top right and a navigation menu with options like Home, Help, Contact, About Us, and Login/Register. Below the navigation, there are tabs for Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area is titled "The Arabidopsis Information Resource" and includes a detailed description of the database, which maintains genetic and molecular biology data for *Arabidopsis thaliana*. It also mentions the 2012 MASC Report Now Available (July 11, 2012) and a new protein chip and cell cultures at ABRC (May 9, 2012). A prominent banner in the center encourages users to "Click here to try our new online submission form" and provides instructions on what to submit, such as molecular function, biological process, localization, or interacting partner of a gene. The footer of the browser window shows the Windows taskbar with various open applications and the system tray.



# 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

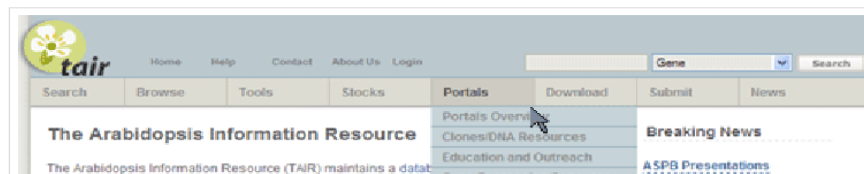
[October 15, 2006]  
Search for [genes](#), [germplasms](#), and [polymorphisms](#) using associated phenotype, and see improved phenotype data display in results and detail pages.

### 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.



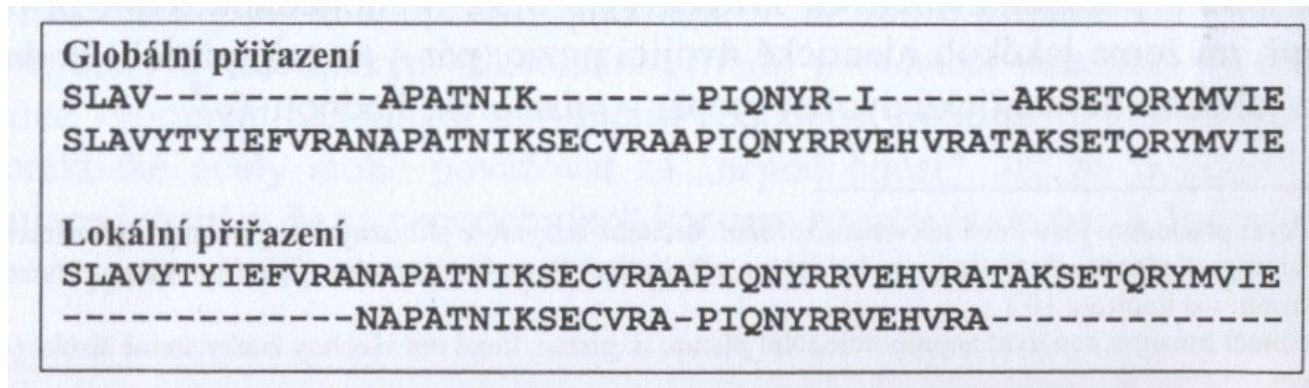
# Osnova

- 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

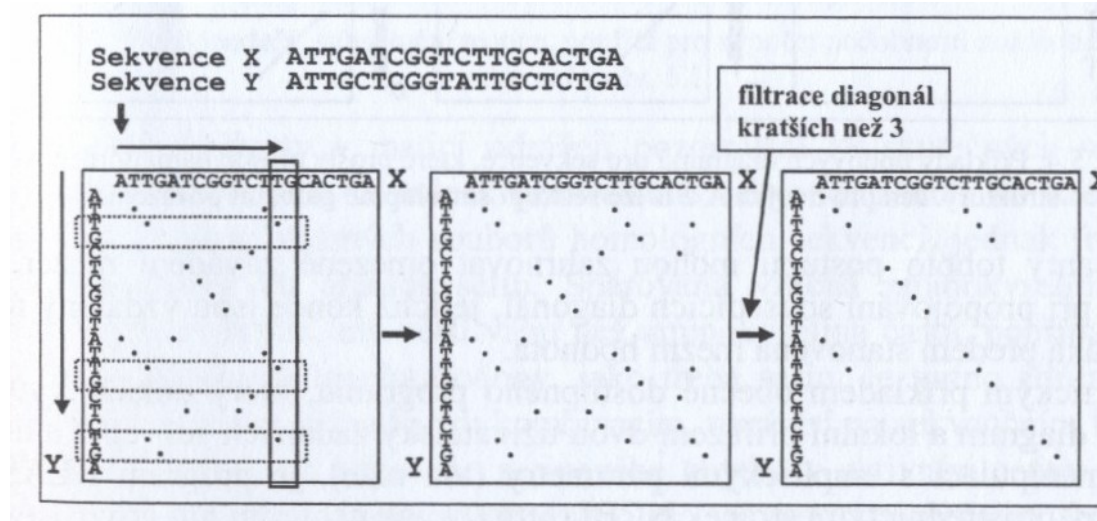


Cvrčková, Úvod do praktické bioinformatiky

- 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)

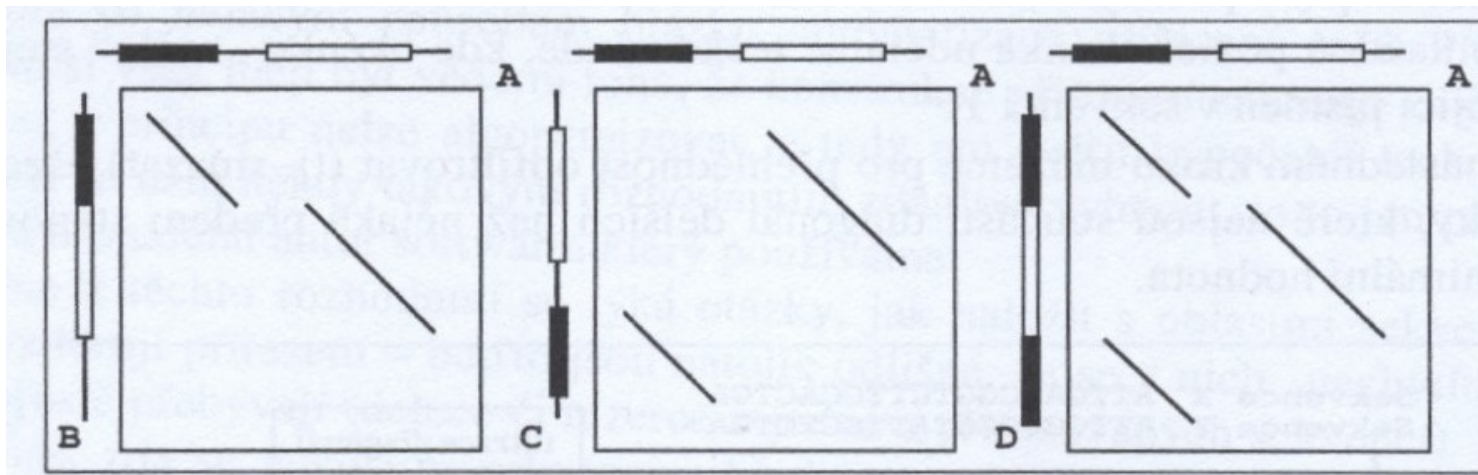


Cvrčková, Úvod do praktické bioinformatiky

- 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

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

NCBI *nucleotide-nucleotide* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

[Search](#)

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

[Set subsequence](#) From:  To:

[Choose database](#)

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	-3	-2	-3	-4	-5	-5	-2	-4	-4	-2	-5	-3	-3	-2	-4	-3	-8
2	1	3	1	1	1	1	0	0	-1	-1	0	0	-2	-1	-1	-1	-3	-3	-2
3	0	6	1	1	1	1	1	1	0	0	1	0	-2	-1	-1	-1	-4	-4	-2
6	0	-1	1	2	1	1	1	1	1	1	1	0	-2	-1	-1	-1	-4	-4	-2
5	0	-1	1	1	5	1	1	1	1	1	1	0	-2	-1	-1	-1	-4	-4	-2
2	0	0	0	0	0	2	2	2	2	2	2	0	-2	-1	-1	-1	-4	-4	-2
4	0	0	0	0	0	1	4	1	3	4	1	0	-2	-1	-1	-1	-4	-4	-2
4	0	0	0	0	0	1	2	2	4	1	2	0	-2	-1	-1	-1	-4	-4	-2
6	0	0	0	0	0	1	1	3	3	6	1	0	-2	-1	-1	-1	-4	-4	-2
6	0	0	0	0	0	1	0	0	1	2	6	0	-2	-1	-1	-1	-4	-4	-2
5	0	0	0	0	0	1	0	0	1	0	3	5	-2	-1	-1	-1	-4	-4	-2
6	0	0	0	0	0	1	0	0	0	0	0	0	6	2	2	2	5	2	10
5	0	0	0	0	0	1	0	0	0	0	0	0	2	5	4	2	6	7	10
4	0	0	0	0	0	1	0	0	0	0	0	0	2	2	4	2	4	7	10
4	0	0	0	0	0	1	0	0	0	0	0	0	2	2	4	2	4	7	10
9	0	0	0	0	0	1	0	0	0	0	0	0	2	2	4	2	4	7	10
10	0	0	0	0	0	1	0	0	0	0	0	0	2	2	4	2	4	7	10
17	0	0	0	0	0	1	0	0	0	0	0	0	2	2	4	2	4	7	10

# BLAST

## Basic Local Alignment Search Tool



E= expectancy value

Expect = 0.0

- „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

# Primární databáze

The screenshot displays a web browser window with the URL [www.ncbi.nlm.nih.gov/geo/1224316](http://www.ncbi.nlm.nih.gov/geo/1224316). The main content shows a genomic map for NC\_002377.1 (145K..148K, 2.9Kbp). A red bar represents the gene NP\_059797.1. A tooltip provides the following details:

- NP\_059797.1**
- NP\_059797.1: two-component VirA-like sensor kinase
- total range: NC\_002377.1 (145,694..148,183)
- total length: 2,490
- strand: plus
- protein product length: 829
- Links & Tools**
- GenBank View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)
- FASTA View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1](#)
- BLAST Genomic: [NC\\_002377.1 \(145,694..148,183\)](#)
- Graphical View: [NP\\_059797.1](#)
- BLAST Protein: [NP\\_059797.1](#)
- BLINK Results: [NP\\_059797.1](#)

Below the map, there are sections for **Bibliography** and **Related articles in PubMed**. The browser's taskbar at the bottom shows several open applications, including Firefox, Windows Media Center, and various office and utility programs.



# BLAST

## Basic Local Alignment Search Tool

BLINK precomputed BLAST

Home Taxonomy Report Multiple Alignment Blast Help

My NCBI [Sign In] [Register]

Pre-computed BLAST results for: [gi|16119781|ref|NP\\_396486.1](#) two component sensor kinase [Agrobacterium tumefaciens str. C58]  
Matching gis: [15163423;20141871;1019660](#)  
Total (score > 100) : 147086 hits in 146754 proteins in 6309 species  
Selected: 147086 hits in 146754 proteins in 6309 species Filter: **Min Score: 100** |  
Other views (Reports): [Taxonomy report](#) [Multiple Alignment](#) [Blast](#)  
[Reset all filters](#)

Choose Display Options

1203 Archaea 138285 Bacteria 13 Metazoa 1349 Fungi 554 Plants 6 Viruses 5676 The Others [reset selection](#)

Results: 1 - 100 [Next Page](#) [Last](#)

% hits 833 aa [reset selection](#)

blink	SCORE	ACCESSION	Length	Protein Description
	4166	<a href="#">AAK90927</a>	833	two component sensor kinase [Agrobacterium tumefaciens str. C58]
	4166	<a href="#">P18540</a>	833	RecName: Full=Wide host range virA protein; Short=WHR virA
	4166	<a href="#">AAA79282</a>	833	virA [Plasmid pTiC58]
	4159	<a href="#">NP_053380</a>	833	hypothetical protein pTi-SAKURA_p142 [Agrobacterium tumefaciens]
	4159	<a href="#">BAA87765</a>	833	tiorf140 [Agrobacterium tumefaciens]
	4153	<a href="#">AAA91590</a>	833	virA [Plasmid Ti]
	4153	<a href="#">gi 737127</a>	833	virA protein
	4153	<a href="#">CAA34777</a>	833	91.3 kDa protein [Agrobacterium tumefaciens]
	3800	<a href="#">CAA35780</a>	829	virA [Agrobacterium rhizogenes]
	3718	<a href="#">gi 227240</a>	869	virA gene
	3148	<a href="#">AAA88643</a>	829	virA [Plasmid Ti]

# 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

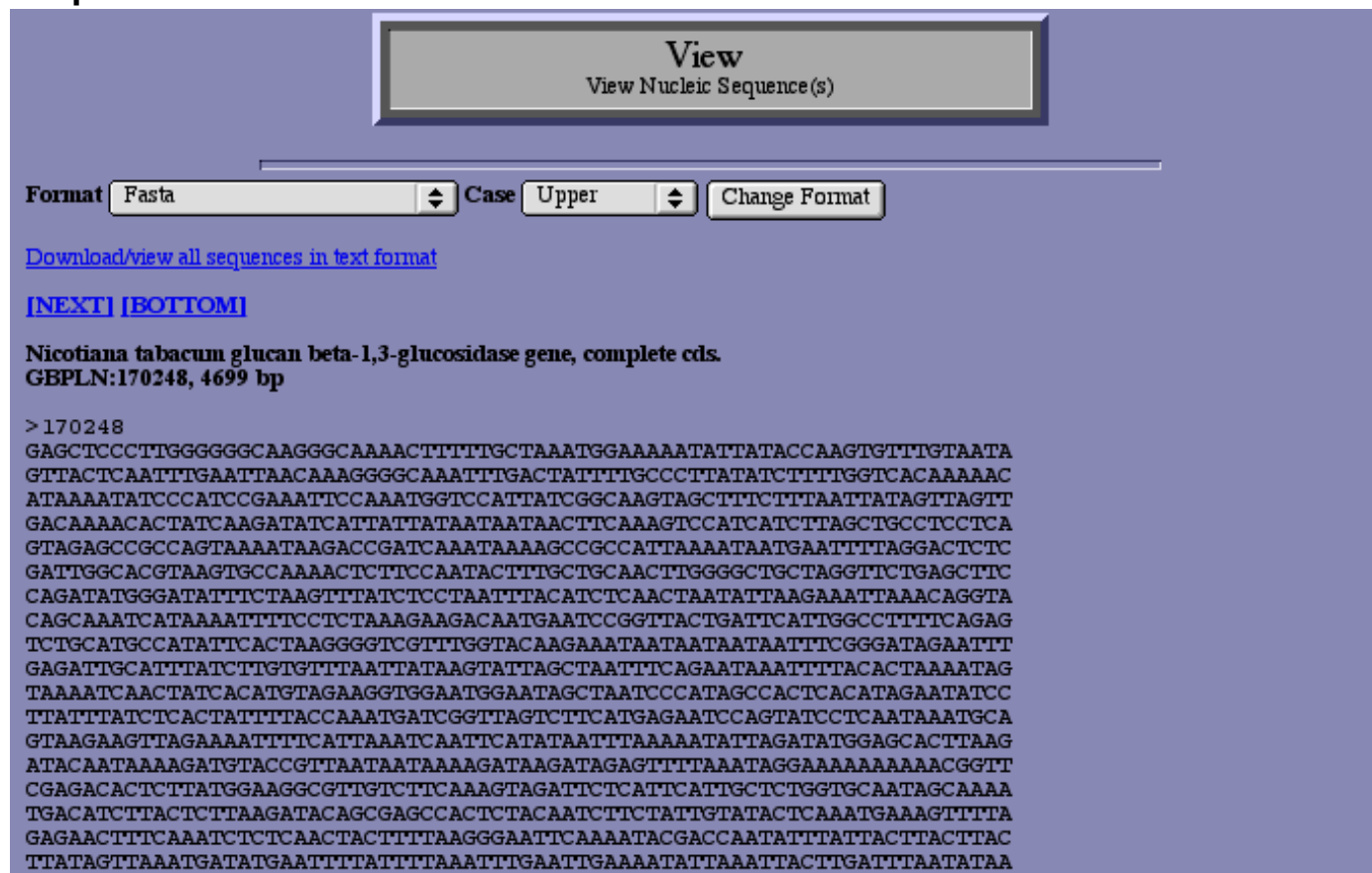
GBPLN:804655 **Hordeum vulgare L. beta-glucosidase (BGQ60) gene, complete cds.**  
 GBPLN:170248 **Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds.**

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  
GAGCTCCCTTGGGGGGCAAGGGCAAAACTTTTTGCTAAATGGAAAAATATTATACCAAGTGTGTAATA  
GTTACTCAATTTGAATTAACAAAGGGGCAAAATTTGACTATTTTGCCCTTATATCTTTTGGTCACAAAAAC  
ATAAAATATCCCATCCGAAATTCCAAATGGTCCATTATCGGCCAAGTAGCTTTCTTTAATTATAGTTAGTT  
GACAAAACACTATCAAGATATCATTATTATAAATAAATTCAAAGTCCATCATCTTAGCTGCCTCCTCA  
GTAGAGCCGCCAGTAAATAAAGACCGATCAAAATAAAGCCGCCATTAAAAAATGAATTTTAGGACTCTC  
GATTGGCACGTAAGTGCCAAAACCTTTCCAATACTTTGCTGCAACTTGGGGCTGCTAGGTTCTGAGCTTC  
CAGATATGGGATATTTCTAAGTTTTATCTCTAATTTTACATCTCAACTAATATTAAGAAAATTAACAGGTA  
CAGCAAATCATAAAATTTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCTTTTTCAGAG  
TCTGCATGCCATATTCCTAAGGGGTCGTTTGGTACAAGAAAATAAATAAATAAATTTTCGGGATAGAATTT  
GAGATTGCATTTATCTTGTGTTTTAATTATAAGTATTAGCTAATTTTCAGAATAAAATTTTACTAAAATAG  
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC  
TTATTTATCTCACTATTTTACCAAATGATCGGTTAGTCTTTCATGAGAATCCAGTATCCTCAATAAATGCA  
GTAAGAAGTTAGAAAATTTTCAATTAATCAATTCATATAAATTTAAAAATATTAGATATGGAGCACTTAAG  
ATACAATAAAAGATGTACCGTTAATAAATAAAGATAAGATAGAGTTTTAAATAGGAAAAAAAACCGTT  
CGAGACTCTTTATGGAAGGCGTTTCTTCAAAATGAGATTCTCATTCAATTGCTCTGGTGC AATAGCAAAA  
TGACATCTTACTCTTAAGATACAGCGAGCCACTCTACAATCTTCTATTGTATACTCAAATGAAAGTTTTA  
GAGAATTTCAAATCTCTCAACTACTTTTTAAGGGAATTCAAAATACGACC AATATTTATTACTTACTTAC  
TTATAGTTAAATGATATGAATTTTTATTTAAATTTGAATTGAAAATATTAATTTACTTGTATTAATATAA

# 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
GACAAAACACTATCAAGATATCATTATTATAATAATAAACTTCAAAAGTCCATCATCTTTAGCTGCCTCCTCA
GTAGAGCCGCCAGTAAAAATAAGACCAGATCAAATAAAAAGCCGCCATTAAAAATAATGAATTTTAGGACTCTC
GATTTGGCAGGTAAGTGGCAAAACTCTTCCAATACTTTTGTGCAACTTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATTTCTAAGTTTATCTCCTAATTTACATCTCAACTAATATTAAGAAATTAACAGGTA
CAGCAAATCATAAAATTTTCTCTAAAGAAGACAATGAATCCGGTTACTGATTCATTGGCCTTTTTCAGAG
TCTGCATGCCATATTCCTAAGGGGTCGTTTGGTACAAGAAATAATAATAAATTTTCGGGATAGAATTTT
GAGATTGCATTTATCTTGTGTTTAAATATAAGTATTAGCTAATTTTACAATAAAATTTTACTAAAATAG
TAAAATCAACTATCACATGTAGAAGGTGGAATGGAATAGCTAATCCCATAGCCACTCACATAGAATATCC
TTATTTATCTCACTATTTTACCAATGATCGGTTAGTCTTCATGAGAATCCAGTATCCTCAATAAATGCA
GTAAGAAGTTAGAAAAATTTTCATTAAATCAATTCATATAATTTAAAAATATTAGATATGGAGCACTTAAG
ATACAATAAAGATGTACCGTTAATAATAAAGATAAGATAGAGTTTAAATAGGAAAAAACAACGGTT
CGAGACACTCTTATGGAAGGCGTTGTCTTCAAAGTAGATTCTCATTCATTGCTCTGGTGCATAGCAAAA
TGACATCTTACTCTTAAGATACAGGAGCCACTCTACAATCTTCTATTGTATACTCAAAATGAAAGTTTTA
GAGAACTTTTCAAACTCTCAACTACTTTTAAAGGGAATTCAAAAATACGACCAATATTTATTACTTACTTAC
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

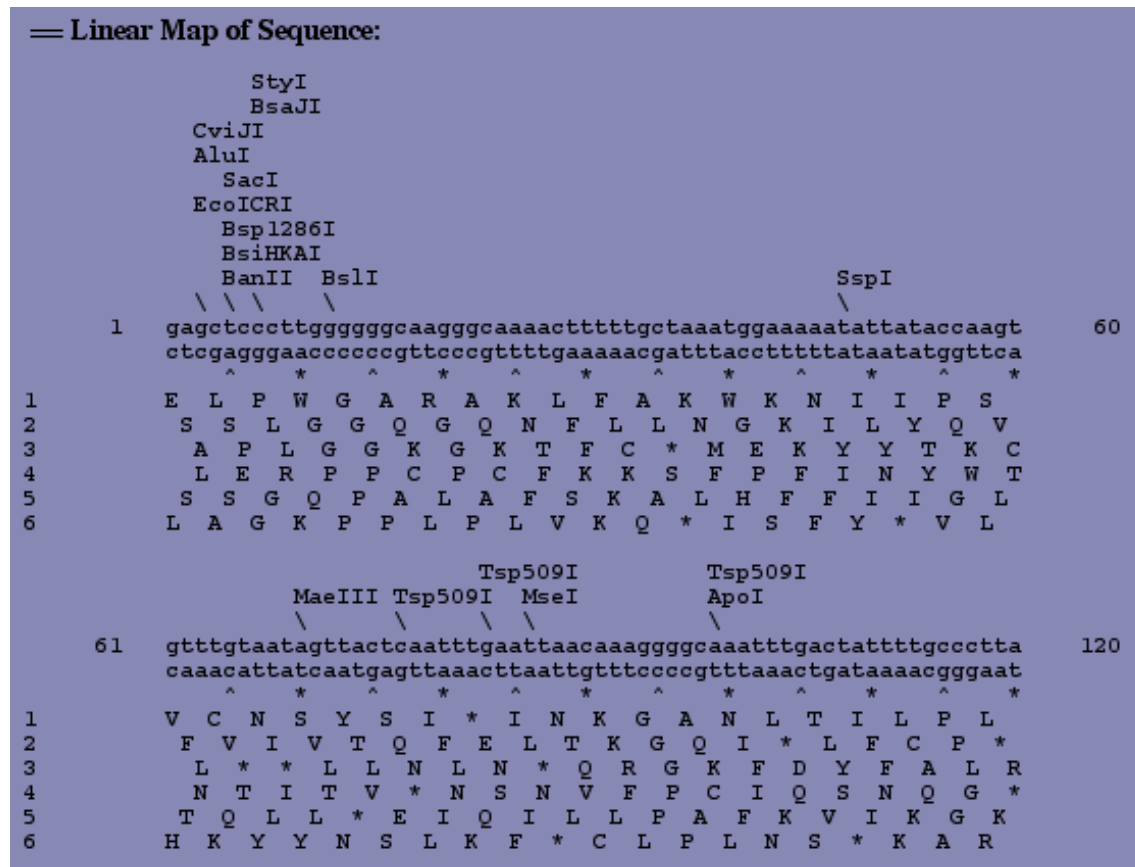
Nicotiana tabacum glucan beta-1,3-glucosidase gene, complete cds. Tran

```
>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/>



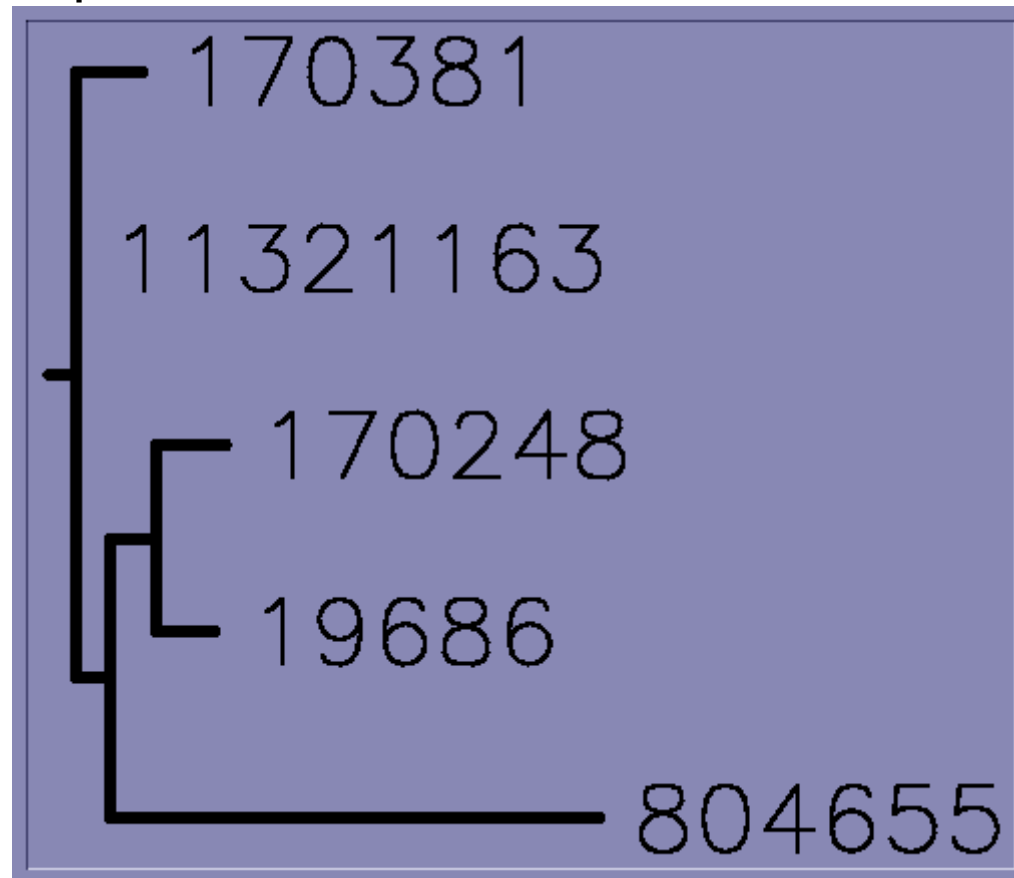







# Analytické nástroje

- <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, 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

# Další WWW zdroje




- Online Mendelian Inheritance in Man (OMIM)

Mirror sites: [us-east.omim.org](http://us-east.omim.org), [europe.omim.org](http://europe.omim.org)

**OMIM<sup>®</sup>**  
Online Mendelian Inheritance in Man<sup>®</sup>  
An Online Catalog of Human Genes and Genetic Disorders  
Updated 6 September 2012

Search OMIM   [Sample Searches](#)

Advanced Search: [OMIM](#), [Clinical Synopses](#), [OMIM Gene Map](#)

NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal medical or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions.  
OMIM<sup>®</sup> and Online Mendelian Inheritance in Man<sup>®</sup> are registered trademarks of the Johns Hopkins University.  
Copyright © 1966-2012 Johns Hopkins University.



# 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