

# CG020 Genomika

## 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](#)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Schéma předmětu

- **Kapitola 01**
  - Úvod do bioinformatiky
- **Kapitola 02**
  - Identifikace genů
- **Kapitola 03**
  - Přístupy reverzní genetiky
- **Kapitola 04**
  - Přístupy genetiky přímé



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Schéma předmětu

- **Kapitola 05**
  - Přístupy funkční genomiky
- **Kapitola 06**
  - Protein-protein interakce a jejich analýza
- **Kapitola 07**
  - Současné metody sekvenování DNA
- **Kapitola 08**
  - Struktura genomů



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Schéma předmětu

- **Kapitola 09**
  - Evoluce genomů
- **Kapitola 10**
  - Genomika a systémová biologie
- **Kapitola 11**
  - Praktické aspekty funkční genomiky
  - Modelové organismy
  - PCR
  - Zásady navrhování primerů



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Literatura

- Literární zdroje pro kapitulu 01:
  - **Bioinformatics and Functional Genomics**, 3rd Edition, Jonathan Pevsner, Wiley-Blackwell, 2015  
<http://www.bioinfbook.org/php/?q=book3>
  - **Úvod do praktické bioinformatiky**, Fatima Cvrčková, 2006, Academia, Praha
  - **Plant Functional Genomics**, ed. Erich Grotewold, 2003, Humana Press, Totowa, New Jersey



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Osnova

- Schéma předmětu
- Definice



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# GENOMIKA-co to je?

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

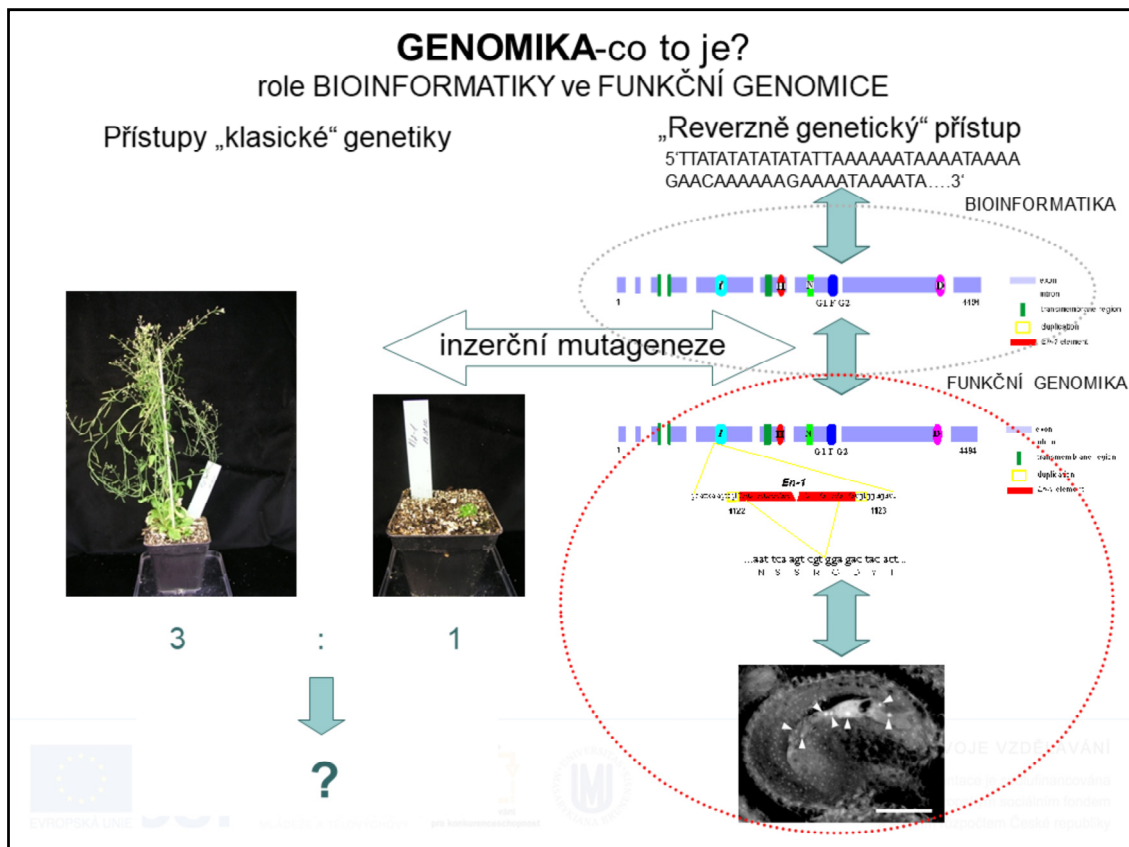
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

Genomics is a science discipline that is interested in the analysis of genomes. Genome of each organism is a complex of all genes of the respective organism. The genes could be located in cytoplasm (prokaryots) nucleus (in most eukaryotic organisms), mitochondria or chloroplasts (in plants).

The critical prerequisite of genomics is the knowledge of gene sequences.

Functional genomics is interested in function of individual genes.





With the knowledge of gene sequences (or the knowledge of the gene files in the individual organisms, i.e. the knowledge of genomes), **Reverse Genetics** appears that allows study their function.

In comparison to "classical" or **Forward Genetics**, starting with the phenotype, the reverse genetics starts with the sequence identified as a gene in the sequenced genome. The gene identification using approaches of **Bioinformatics** will be described later (see Lesson 02).

Reverse genetics uses a spectrum of approaches that will be described in the Lesson 03 that allow isolation of sequence-specific mutants and thus their phenotype analysis.

The necessity of having phenotype alterations in the forward genomics approach introduces important difference between those two approaches. Thus, the gene is no longer understood as a factor (*trait*) determining *phenotype*, but rather as a piece of DNA characterized by the unique *string of nucleotides*. i.e. **physical DNA molecule**.

# Osnova

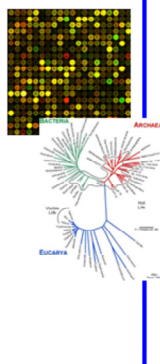
- Schéma předmětu
- Definice
- Role BIOINFORMATIKY v současném pojetí FUNKČNÍ GENOMIKY



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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.**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

## NIH WORKING DEFINITION OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

July 17, 2000

The following working definition of bioinformatics and computational biology were developed by the BISTIC Definition Committee and released on July 17, 2000. The committee was chaired by Dr. Michael Huerta of the National Institute of Mental Health and consisted of the following members:

### Bioinformatics Definition Committee BISTIC Members Expert Members

Michael Huerta (Chair) Gregory Downing  
Florence Haseltine Belinda Seto  
Yuan Liu

### Preamble

Bioinformatics and computational biology are rooted in life sciences as well as computer and information sciences and technologies. Both of these interdisciplinary approaches draw from specific disciplines such as mathematics, physics, computer science and engineering, biology, and behavioral science. Bioinformatics and computational biology each maintain close interactions with life sciences to realize their full potential. Bioinformatics applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and useful. Computational biology uses mathematical and computational approaches to address theoretical and experimental questions in biology. Although bioinformatics and computational biology are distinct, there is also significant overlap and activity at their interface.

### Definition

The NIH Biomedical Information Science and Technology Initiative Consortium agreed on the following definitions of bioinformatics and computational biology recognizing that no definition could completely eliminate overlap with other activities or preclude variations in interpretation by different individuals and organizations.

*Bioinformatics:* Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

*Computational Biology:* The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.

# 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.

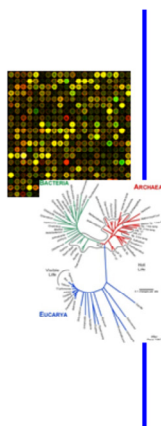
J. Pevsner,  
<http://www.bioinfbook.org/index.php>



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 genomy 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 **systemové biologie**
    - Matematické modelování genových regulačních sítí



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



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Spektrum on-line zdrojů

<b>EMBLet National Modes</b>		
Vienne BioCenter	Austria	<a href="http://www.at.emblnet.org/">http://www.at.emblnet.org/</a>
BBN	Belgium	<a href="http://www.be.emblnet.org/">http://www.be.emblnet.org/</a>
BioBase	Denmark	<a href="http://biobase.dk/">http://biobase.dk/</a>
CSC	Finland	<a href="http://www.fi.emblnet.org/">http://www.fi.emblnet.org/</a>
INFORMAGEN	France	<a href="http://www.infoblogen.fr/">http://www.infoblogen.fr/</a>
CRISOLINET	Germany	<a href="http://genome.zibb-helmholtz.de/biocom/">http://genome.zibb-helmholtz.de/biocom/</a>
IMBB	Greece	<a href="http://www.imbb.forth.gr/">http://www.imbb.forth.gr/</a>
HON	Hungary	<a href="http://www.hu.emblnet.org/">http://www.hu.emblnet.org/</a>
INCEI	Ireland	<a href="http://www.incei.tcd.ie/">http://www.incei.tcd.ie/</a>
JNN	Israel	<a href="http://dapsil.wellman.ac.il/bcd/inn.html">http://dapsil.wellman.ac.il/bcd/inn.html</a>
JIB-ADN	Italy	<a href="http://jib-www.ba.cnr.it/8000/BioWWW/Bio-WWW.htm">http://jib-www.ba.cnr.it/8000/BioWWW/Bio-WWW.htm</a>
CAS/C/CLAMN	Netherlands	<a href="http://www.cas.kun.nl/">http://www.cas.kun.nl/</a>
IBO	Norway	<a href="http://www.no.emblnet.org/">http://www.no.emblnet.org/</a>
IBB	Poland	<a href="http://www.ibb.wzpa.pl/">http://www.ibb.wzpa.pl/</a>
ISC	Portugal	<a href="http://www.igc.gulbenkian.pt/">http://www.igc.gulbenkian.pt/</a>
GeneBee	Russia	<a href="http://www.genebee.msu.ru/">http://www.genebee.msu.ru/</a>
CNB-CSC	Spain	<a href="http://www.es.emblnet.org/">http://www.es.emblnet.org/</a>
BNC	Sweden	<a href="http://www.se.emblnet.org/">http://www.se.emblnet.org/</a>
SIB	Switzerland	<a href="http://www.ch.emblnet.org/">http://www.ch.emblnet.org/</a>
SIGNET	UK	<a href="http://www.signet.dl.ac.uk/">http://www.signet.dl.ac.uk/</a>
<b>EMBLet Specialist Modes</b>		
MPS	Germany	<a href="http://www.mips.biochem.mpg.de/">http://www.mips.biochem.mpg.de/</a>
ICGB	Italy	<a href="http://www.icgb.internic.it/">http://www.icgb.internic.it/</a>
Pharmacia Uppsala	Sweden	<a href="http://www.gnu.com/">http://www.gnu.com/</a>
Faustmann-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-BC	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>
EMBL-EB	UK	<a href="http://www.embl.ac.uk/embrowser">http://www.embl.ac.uk/embrowser</a>
<b>EMBLet Associate Modes</b>		
IBBH	Argentina	<a href="http://iui.biot.unip.edu.ar/emblnet">http://iui.biot.unip.edu.ar/emblnet</a>
ANGS	Australia	<a href="http://www.angis.usc.edu.au/">http://www.angis.usc.edu.au/</a>
CEI	China	<a href="http://www.cei.cbi.cas.edu.cn/">http://www.cei.cbi.cas.edu.cn/</a>
CISB	Cuba	<a href="http://ibc.cigb.edu.cu/">http://ibc.cigb.edu.cu/</a>
CFDQ	India	<a href="http://falarjung.emblnet.org.in/">http://falarjung.emblnet.org.in/</a>
SANBE	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>



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



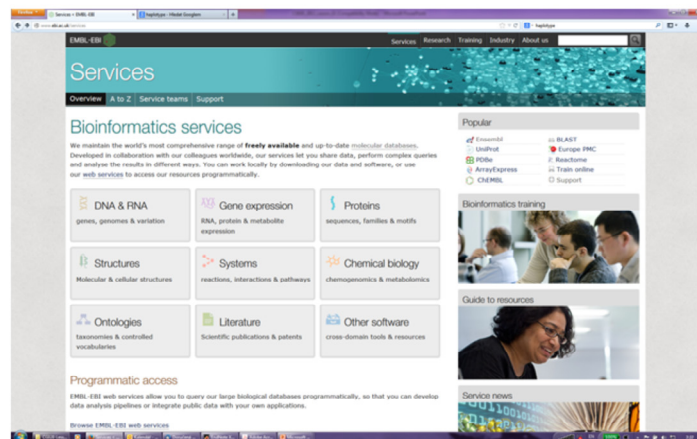
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

There are many of on-line resources that could be used.

# Spektrum on-line zdrojů

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

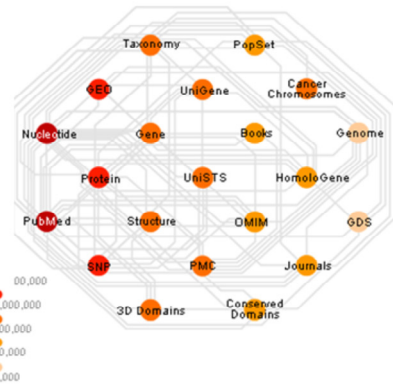
Nowadays, the resources are interconnected and could be accessed via dedicated web pages. Among the best and mostly used www resources integrating plenty of database resources belong www portal of European Bioinformatics Institute (EBI) in Europe (Germany) and National Center of Biotechnology Information (NCBI) in the USA (



# Spektrum on-line zdrojů

□ NCBI <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI homepage with a search bar and a navigation menu. The main content area includes a 'Welcome to NCBI' message, a 'Get Started' section with links to 'Tools', 'Downloads', 'HowTo', and 'Submissions', and a 'NCBI YouTube channel!' section. A 'Popular Resource' sidebar lists various databases like PubMed, Bookshelf, and BLAST.



EVROPSKÝM SOCIÁLNÍM FONDEM  
a  
státním rozpočtem České republiky

Nowadays, the resources are interconnected and could be accessed via dedicated web pages.

# Osnova

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

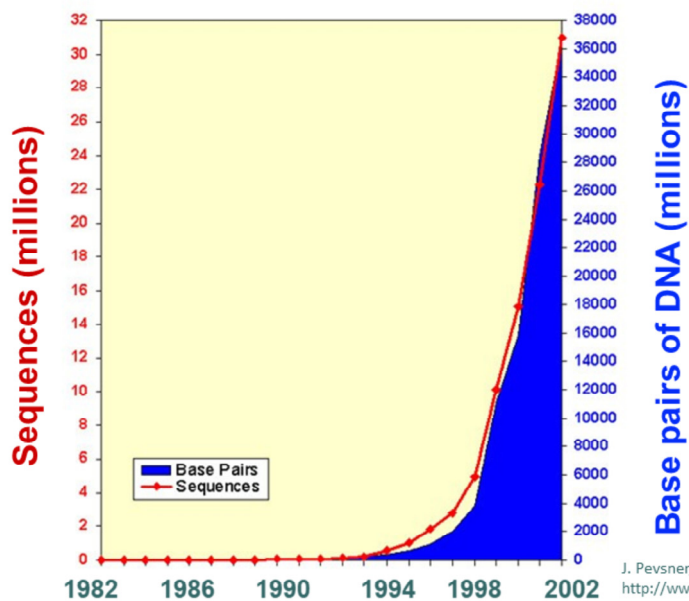
- zahrnují soubory primárních dat – sekvencí DNA a proteinů
  - Sekvence v databázích tzv. „Velké trojky“:
    - EMBL
      - <http://www.ebi.ac.uk/embl/>
    - GenBank,
      - <https://www.ncbi.nlm.nih.gov/>
    - DDBJ,
      - <http://www.ddbj.nig.ac.jp>
  - denně vzájemná výměna a zálohování dat
  - velká datová náročnost (kapacita i software)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

## Growth of GenBank



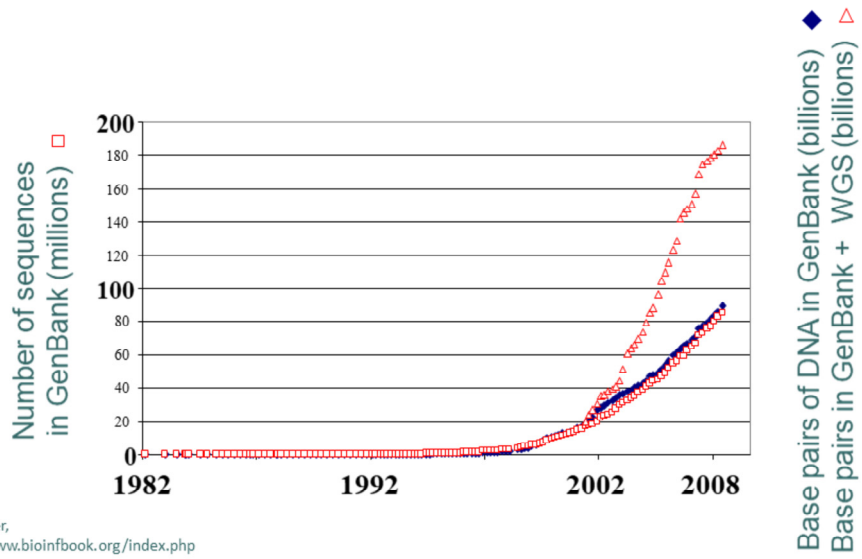
J. Pevsner,  
<http://www.bioinfbook.org/index.php>



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>

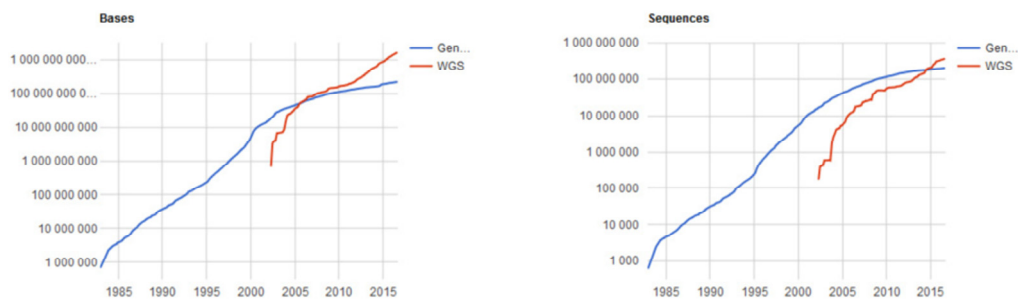


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Growth of GenBank

## Aug 2016



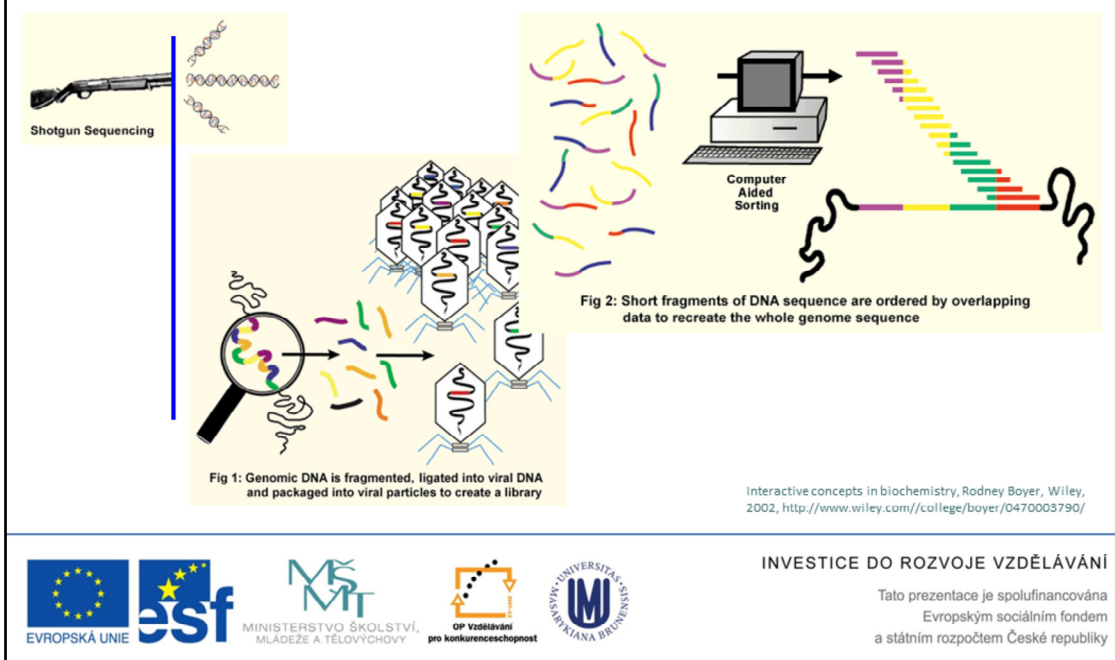
- Prosinec 1982 680 338 bp, 606 sekvencí
- Duben 2002  $19 \times 10^9$  bp,  $17 \times 10^6$  sekvencí + WGS  $692 \times 10^6$  bp, 172 768 sekvencí
- Srpen 2016  $218 \times 10^9$  bp,  $196 \times 10^6$  sekvencí + WGS  $1,6 \times 10^{12}$  bp,  $360 \times 10^6$  sekvencí



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# WGS

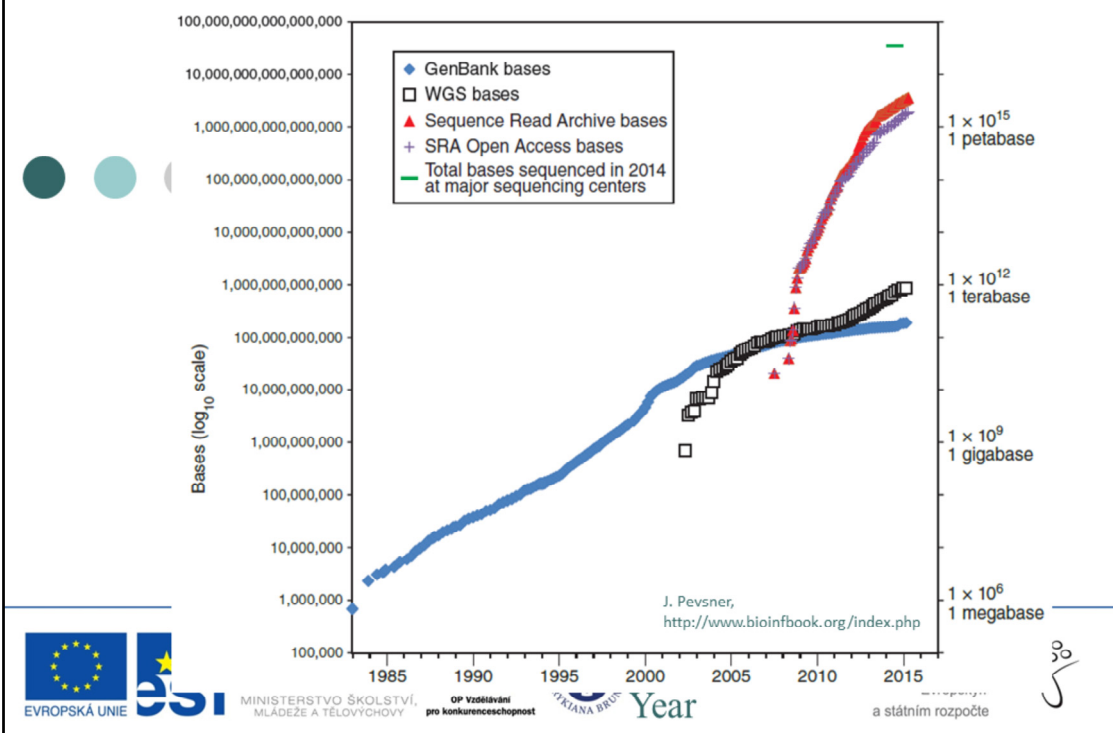


Shotgun sequencing allows a scientist to rapidly determine the sequence of very long stretches of DNA. The key to this process is fragmenting of the genome into smaller pieces that are then sequenced side by side, rather than trying to read the entire genome in order from beginning to end. The genomic DNA is usually first divided into its individual chromosomes. Each chromosome is then randomly broken into small strands of hundreds to several thousand base pairs, usually accomplished by mechanical shearing of the purified genetic material. Each of the short DNA pieces is then inserted into a DNA vector (a viral genome), resulting in a viral particle containing "cloned" genomic DNA (Fig. 1).

The collection of all the viral particles with all the different genomic DNA pieces is referred to as a library. Just as a library consists of a set of books that together make up all of human knowledge, a genomic library consists of a set of DNA pieces that together make up the entire genome sequence. Placing the genomic DNA within the viral genome allows bacteria infected with the virus to faithfully replicate the genomic DNA pieces. Additionally, since a little bit of known sequence is needed to start the sequencing reaction, the reaction can be primed off the known flanking viral DNA.

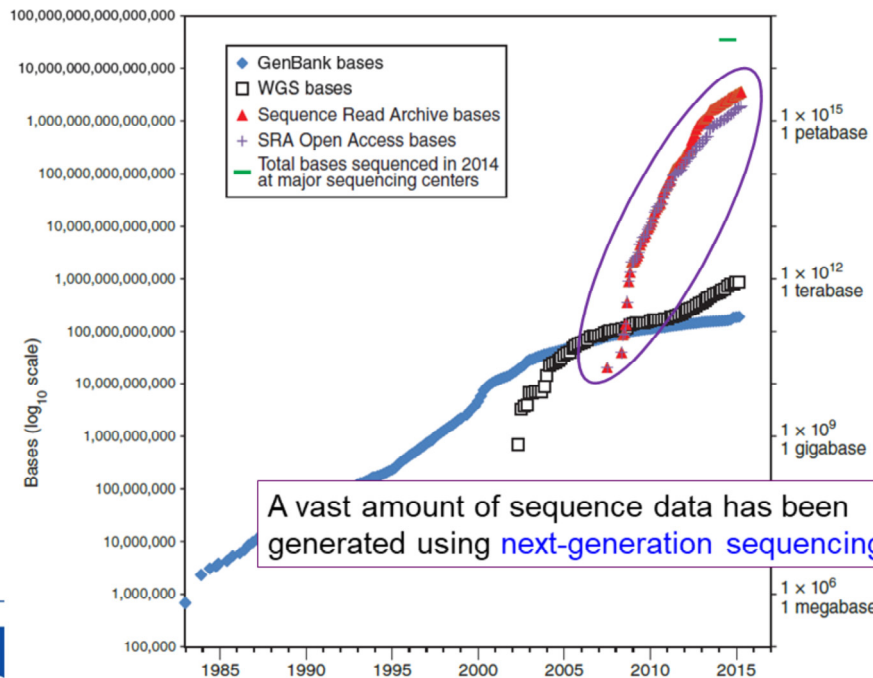
In order to read all the nucleotides of one organism, millions of individual clones are sequenced. The data is sorted by computer, which compares the sequences of all the small DNA pieces at once (in a "shotgun" approach) and places them in order by virtue of their overlapping sequences to generate the full-length sequence of the genome (Fig. 2). To statistically ensure that the whole genome sequence is acquired by this method, an amount of DNA equal to five to ten times the length of the genome must be sequenced. (Interactive concepts in biochemistry, Rodney Boyer, Wiley, 2002, <http://www.wiley.com/college/boyer/0470003790/>)

# Growth of DNA Sequence in Repositories





# Growth of DNA Sequence in Repositories



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

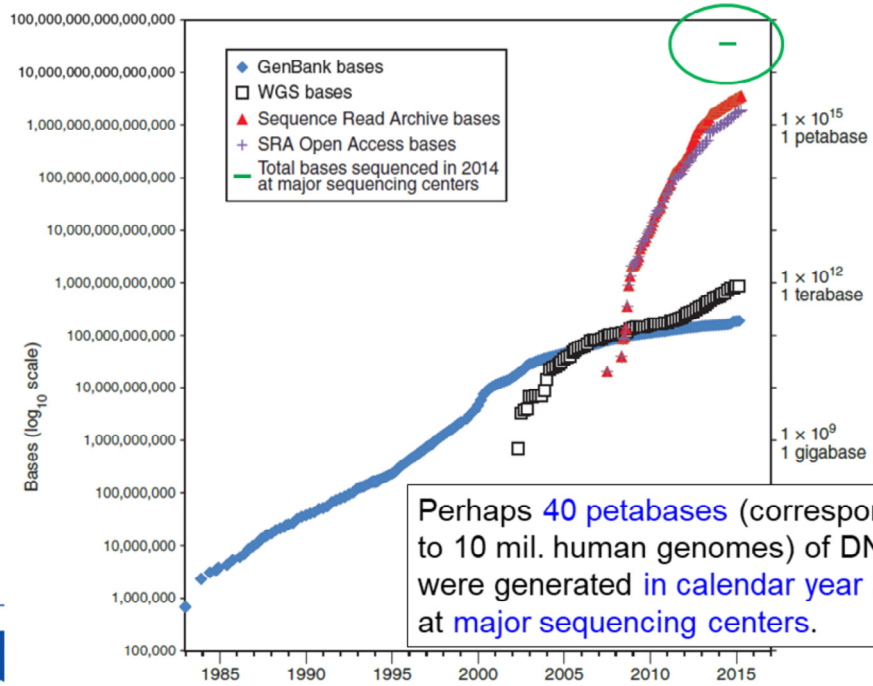
OP Vzdělávání  
pro konkurenceschopnost

EVROPSKÝ  
ROK  
2010

a státním rozpočte



# Growth of DNA Sequence in Repositories



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání  
pro konkurenceschopnost

EVROPSKÝ  
ROK  
2010

a státním rozpočte

# 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/>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

- Typy sekvencí v primárních databázích
  - Standardní nukleotidové sekvence získané kvalitním sekvencováním
  - **ESTs** (**E**xpressed **S**equences **T**ags)
  - **HGTS** (**H**igh **T**hroughput **G**enome **S**equencing)
    - neanotované „surové“ výsledky sekvenačních projektů
  - Referenční sekvence anotovaných genomů
  - **TPAs** (**T**hird **P**arty **A**notation)
    - sekvence anotované jinými než původními autory



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

GenBank (NCBI) <https://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 includes links for Tools, Downloads, How-To's, and Submissions. The 'Popular Resources' section lists various databases and tools like PubMed, Bookshelf, and BLAST.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

The screenshot displays the NCBI Gene database entry for the *uak* gene. Key information includes:

- Gene symbol:** *uak*
- Gene description:** non-component VWA-like sensor kinase
- Location:** *uak* [25]
- Gene type:** protein coding
- RefSeq status:** PROVISIONAL
- Organism:** *Agrobacterium tumefaciens* subsp. *Agrobacterium tumefaciens*, subsp. *Rhizobium radiobacter*
- Linkage:** Rhizobium, Rhizobiales, Alphaproteobacteria, Rhizobiales, Rhizobiales, Rhizobium/Agrobacterium group, Agrobacterium, Agrobacterium tumefaciens complex

The 'Genomic context' section shows the gene's location on chromosome 1 (NC\_023277.1) with coordinates 146,800 to 147,000. The 'Genomic regions, transcripts, and products' section shows the gene structure with exons and introns. The 'Related articles' section is highlighted with a yellow circle and lists four articles:

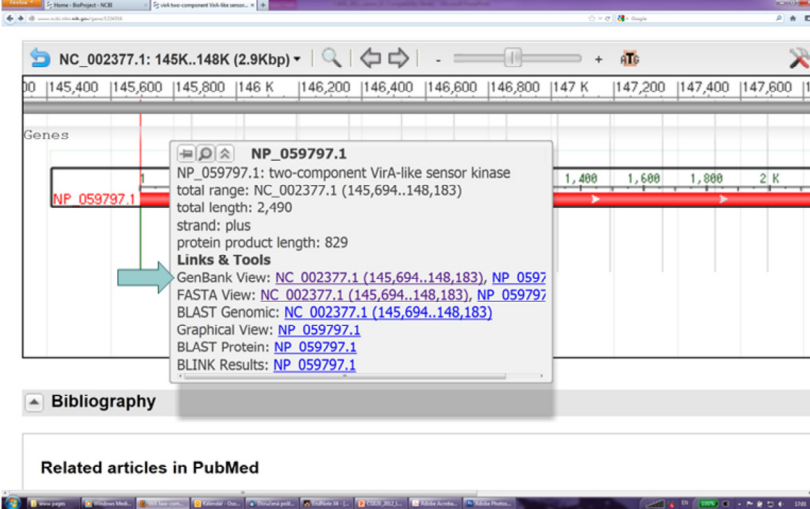
1. Sequences and organization of the *uak* gene in *Agrobacterium tumefaciens* and its relationship to the *uak* gene in *Rhizobium radiobacter*. Kohnenberger G, et al. J Euk Euk 2000 Jun; PMID 10940245
2. The *uak* gene encodes a VWA-like sensor kinase in *Agrobacterium tumefaciens*. Park WC, et al. Mol Microbiol 1993 Mar; PMID 8491915
3. Characterization of the *uak* locus of *Agrobacterium tumefaciens*, a transposon, insertion, and host range determinant. Lewis R, et al. EMBO J 1987 Sep; PMID 3098978
4. Analysis of the complete nucleotide sequence of the *uak* gene. Thompson DJ, et al. Nucleic Acids Res 1988 May 25; PMID 2837738



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze



The screenshot shows a web browser displaying a GenBank entry for NP\_059797.1. The entry details include:

- 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

Under the "Links & Tools" section, there are several links:

- GenBank View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1 \(145,694..148,183\)](#)
- FASTA View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1 \(145,694..148,183\)](#)
- 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 entry, there are sections for "Bibliography" and "Related articles in PubMed".



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze

NCBI Nucleotide

Search [Nucleotide]

Dir:

**Přístupový kód**

BC\_022377.1

**GeneBank Identifier**

BC\_022377.1

LOCUS BC\_022377 2490 bp DNA linear BCT 29-DEC-2003

DEFINITION Agrobacterium tumefaciens carotachrom plasmid Ti, complete sequence.

ACCESSION BC\_022377.1 REFSeq=94444.148193

VERSION BC\_022377.1 GI:10955224

KEYWORDS

SOURCE Agrobacterium tumefaciens (Rhizobium radiobacter);

Organism: Agrobacterium tumefaciens; Strain: Rhizobiales; Type: Agrobacterium.

FARRAND, S.K., SCHRAMMEYER, B., MOYKES, P.J. and

TITLE Octopine-type Ti plasmid sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2490)

AUTHORS Zhu, J., Oger, P.M., Schrammeyer, B., Mooykese, P.J., Farrand, S.K. and Williams, S.C.

TITLE Direct Submission

JOURNAL Submitted (07-MAR-2003) Microbiology, Cornell University, Ithaca, NY 14853, USA

COMMENTS PROVISIONAL [Accession](#): This record has not yet been subject to final NCBI review. The reference sequence was derived from [SRR11333](#).

FEATURES

Location/Qualifiers

1..2490

organism="Agrobacterium tumefaciens"

mol\_type="genomic DNA"

AB\_xref="taxon:258"

plasmid="Ti"

contig="contachromosomal"

octopine-type"

1..2490

gene="vira"

AB\_xref="GeneID:1224316"

1..2490

gene="vira"

note="Two-component regulator of vir regulon; ViraB is a transmembrane histidine kinase"

coding\_start=1

transl\_start=11

product="vira"

protein\_id="WP\_022377.1"

AB\_xref="GI:10955224"



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	<b>DNA</b>
NT_030059	Genomic contig	
Rs7079946	dbSNP (single nucleotide polymorphism)	
N91759.1	An expressed sequence tag (1 of 170)	<b>RNA</b>
NM_006744	RefSeq DNA sequence (from a transcript)	
NP_007635	RefSeq protein	<b>Protein</b>
AAC02945	GenBank protein	
Q28369	SwissProt protein	
1KT7	Protein Data Bank structure record	

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

Page 27

## 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

Page 27

# RefSeq

two-component ViA-like sensor kinase

NCBI Reference Sequences (RefSeq)

Genome Annotation

The following sections contain reference sequences that belong to a specific genome build. [Explain](#)

Reference assembly

Genomic

1. [NC\\_003065.3](#)

Range: 18031..18332  
Download: [GenBank](#), [FASTA](#), [Sequence Viewer](#), [Graphics](#)

mRNA and Protein(s)

1. [NP\\_396486.1](#) two component sensor kinase [Agrobacterium tumefaciens str. C58]

UniProtKB/Swiss-Prot: [E18640](#)

Conserved Domains (3) [summary](#)

<a href="#">cd00075</a>	HATPase_C: Histidine kinase-like ATPases. This family includes several ATP-binding proteins for example: histidine kinase, DNA gyrase B, topoisomerases, heat shock protein HSP90, phytochrome-like ATPases and DNA mismatch repair proteins.
<a href="#">cd00082</a>	HskA: Histidine Kinase A (dimerization/phosphoreceptor) domain: Histidine Kinase A dimers are formed through parallel association of 2 domains creating 4-helix bundles; usually these domains contain a conserved His residue and are activated via ...
<a href="#">PRK13637</a>	PRK13637: two-component ViA-like sensor kinase. Provisional

Location:580 - 694  
Blast Score: 202

Location:688 - 830  
Blast Score: 144

Location:14 - 833  
Blast Score: 2944

Related Sequences



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

## 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 \(145,694..148,183\)](#)  
FASTA View: [NC\\_002377.1 \(145,694..148,183\)](#), [NP\\_059797.1 \(145,694..148,183\)](#)  
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**

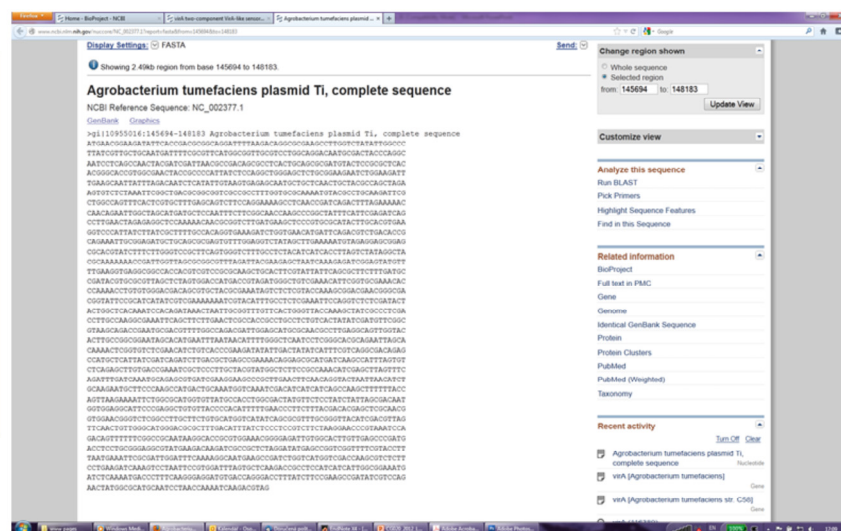
**Related articles in PubMed**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Primární databáze



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Sekundární databáze

- Databáze funkčních nebo strukturních motivů získaných srovnáváním primárních dat (sekvencí)
- PROSITE, <http://www.expasy.org/prosite/>

U: EAFAS, Basic page   Site Map   Search EAFAS   Contact us   Swiss-Prot   PROSITE   Proteomics tools

Search PROSITE: [input] [Go] [Clear]



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 [PROSITE](#) can be used to generate your own pattern. 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.

<b>Scan a protein for PROSITE matches</b> Enter a Swiss-Prot/TrEMBL accession number (AC) (for example: P01301) or a sequence identifier ID (for example: MDIC_BROWNE), or a PDB identifier, or paste your own protein sequence in the box below: <input type="text"/> [Go] [Clear]	<b>Search Swiss-Prot with a PROSITE entry</b> Enter a PROSITE accession number (for example: P08125S), or type your pattern in <a href="#">PROSITE</a> format: [input] [Go] [Clear]
<b>and specify which motifs to use:</b> Scan <input type="checkbox"/> patterns <input type="checkbox"/> profiles <input type="checkbox"/> rules ( <a href="#">User Manual</a> ) (You may also specify a PROSITE entry in the box to the right) <input type="checkbox"/> Find <a href="#">linked systems with a high probability of occurrence</a>	<b>and specify your search limits:</b> <ul style="list-style-type: none"><li>• The <input type="checkbox"/> Swiss-Prot <input type="checkbox"/> TrEMBL <input type="checkbox"/> TrEMBL/Expasy <input type="checkbox"/> PDB databases (You may also specify a pattern in the box to the left) <input type="checkbox"/> including signal regions</li><li>• The following data: [input]</li><li>• <input type="checkbox"/> <a href="#">FASTA</a> format - reports multiple hits with a connection, e.g. <a href="#">Blatt-sequence</a>, <a href="#">Drospheila</a>, <a href="#">Not available</a>, <a href="#">in PDB</a>.</li><li>• Sequences with at least [input] hits</li><li>• At most [input] matches</li></ul> <b>Advanced options:</b> <input type="checkbox"/> FASTA output <input type="checkbox"/> retrieve complete sequences allow at most [input] X sequence characters to match a conserved position in the pattern match [input] priority overlaps, no includes [input] the pattern, no <a href="#">help</a> <a href="#">Proteomics database</a> , <a href="#">CSP</a> <input type="checkbox"/> view a pattern, see <a href="#">help</a>
Your e-mail (optional): [input] (will not reach by e-mail) <input type="checkbox"/> plain text output [START THE SCAN] [RESET]	



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Sekundární databáze

- Databáze funkčních nebo strukturních motivů získaných srovnáváním primárních dat (sekvencí)
- PROSITE, <http://www.expasy.org/prosite/>

```
>PDOC00001 PS00001 SULFATION Tyrosine sulfation site [rule] [Warning: rule with a high probability of occurrence].
171 - 585 sksssstTctctaaa

>PDOC00004 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 RRSE

>PDOC00005 PS00005 PKC_PHOSPHO_SITE Protein kinase C phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].
148 - 150 ESR
144 - 146 TGR
171 - 173 RSR
219 - 221 RSR
369 - 371 TRS
440 - 442 RGR
513 - 515 RGR
585 - 587 RLR
602 - 604 TGR
612 - 614 TGR
716 - 718 RGR
726 - 728 RGR
747 - 749 TRR
784 - 786 RSR
804 - 806 RSR
864 - 866 RSR
868 - 870 RSR
921 - 923 RGR
957 - 959 RSR
960 - 962 TGR
974 - 976 TRR
997 - 999 RSR
1009 - 1011 TGR
1018 - 1020 RGR
1031 - 1033 TGR
1119 - 1121 RSR
```



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Sekundární databáze

- Databáze funkčních nebo strukturních motivů získaných srovnáváním primárních dat (sekvencí)
- PRINTS, <http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/>



PRINTS is a compilation 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 PROTEIN/FAMILY/ENTRY sequence. Usually the motifs do not overlap, but are scattered 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. [Reference](#)

#### New:

- [SPRINT](#) - Score & PRINTS-3 evolutionary PRINTS
- [comPRINTS](#) - Score & PRINTS automatic assignment
- [siteSite](#) - Search the integrated InterPro family database

#### Direct PRINTS access:

- [By accession number](#)
- [By PRINTS code](#)
- [By database code](#)
- [By ID](#)
- [By name](#)
- [By number of motifs](#)
- [By protein](#)
- [By query language](#)

#### PRINTS search:

- Search PRINTS with [NEW FingerprintScan](#)
- [FPrints](#)
- [U.FPRINTScan](#)
- [MULTISeq](#)
- FingerprintScan binaries and source are available: [patrick.scofield@bioinf.man.ac.uk](mailto:patrick.scofield@bioinf.man.ac.uk)

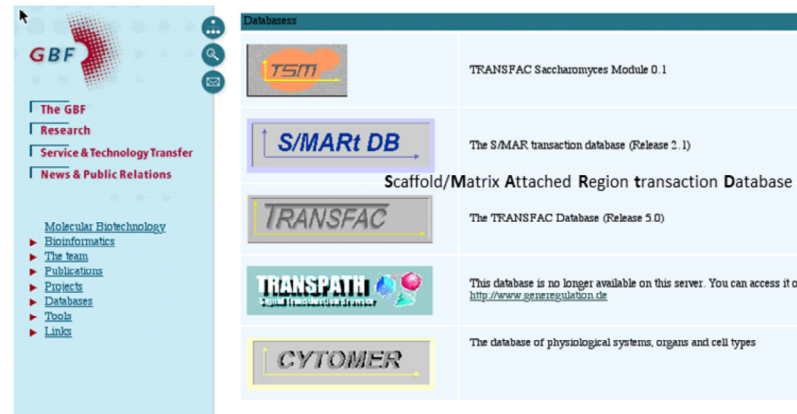


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Sekundární databáze

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



The screenshot shows the TRANSFAC website interface. On the left is a navigation menu for GBF (German Biotechnology Foundation) with categories like '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 lists several databases:

Database Name	Description
TSM	TRANSFAC Saccharomyces Module 0.1
S/MARt DB	The S/MAR transaction database (Release 2.1) Scaffold/Matrix Attached Region transaction Database
TRANSFAC	The TRANSFAC Database (Release 5.0)
TRANSPATI	This database is no longer available on this server. You can access it on <a href="http://www.gene-regulation.de">http://www.gene-regulation.de</a>
CYTOMER	The database of physiological systems, organs and cell types



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

S/MARt DB (scaffold/matrix attached region transaction database). This database collects information about S/MARs and the nuclear matrix proteins that are supposed be involved in the interaction of these elements with the nuclear matrix. <http://transfac.gbf.de/SMARTDB/index.html>)

# Strukturální databáze

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

The screenshot shows the PDB website homepage. At the top, it says "PROTEIN DATA BANK" with the RCSB logo and navigation links for Home, Contact, and Help. A welcome message states: "Welcome to the PDB, the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data." Below this are navigation links for "ABOUT PDB", "DATA UNIFORMITY", "RECENT FEATURES", "USER GUIDES", "FILE FORMATS", "EDUCATION", "STRUCTURAL GENOMICS", "PUBLICATIONS", and "SOFTWARE".

On the left side, there are links for "DEPOSIT data", "DOWNLOAD files", "Browse LINKS", "BETA TEST new features", and "BETA release files". Below these are "Current Holdings" statistics: "19623 Structures", "Last Update: 30-Dec-2002", and "PDB Statistics". A "Molecule of the Month" section features a 3D protein structure and the name "Cytochrome c".

The main content area includes a "Search the Archive" section with a search box, a "Find a structure" button, and checkboxes for "query by PDB id only", "match exact word", and "remove sequence homologues". There are also links for "SearchLite" and "Status Search".

On the right, there is a "PDB Mirrors" section listing various international mirrors such as "San Diego Supercomputer Center", "Rutgers University", "National Institute of Standards and Technology", "Cambridge Crystallographic Data Centre, UK", "National University of Singapore", "Osaka University, Japan", "Universidade Federal de Minas Gerais, Brazil", and "Max Delbrück Center for Molecular Medicine, Germany".

At the bottom of the page, there is a "News" section with a date "23-Dec-2002" and a holiday message: "Happy Holidays from the PDB! The PDB staff wish to extend our best wishes to the community for a happy holiday season and a wonderful new year!"



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Strukturální databáze

o **PDB** <http://www.rcsb.org/pdb/>

Structure Explorer - 1PSY



Structure Explorer - 1PSY

Title: The Structure Of Hot Range Controlling Region Of The Capsid Of Canine and Feline Parvovirus and Mutants  
Classification: Virus/Viral Protein  
Compound: Mol. Id.: 1; Molecule: Coat Protein Yp2; Chain: A; Fragment: Sequence Database Residues 190-231; Engineering: Yes; Mutation: Yes  
Exp. Method: X-ray Diffraction

View Structure

Summary Information

View Structure

Download Display File

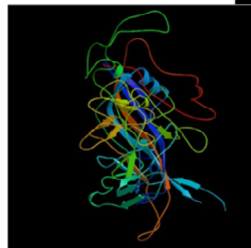
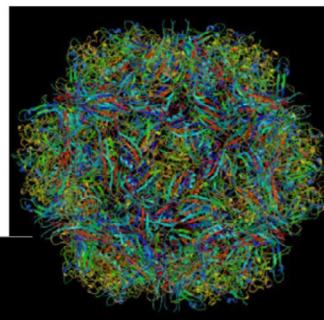
Structural Neighbors

Geometry

Other Sources

Sequence Details

Explore   
Search by  Search by



<http://www.rcsb.org/pdb/cgi/structure.cgi?job=graphics&pdb=1PSY&page=pdb-173561064329344&bio-1&opt-show&size=500> 12/20/2003

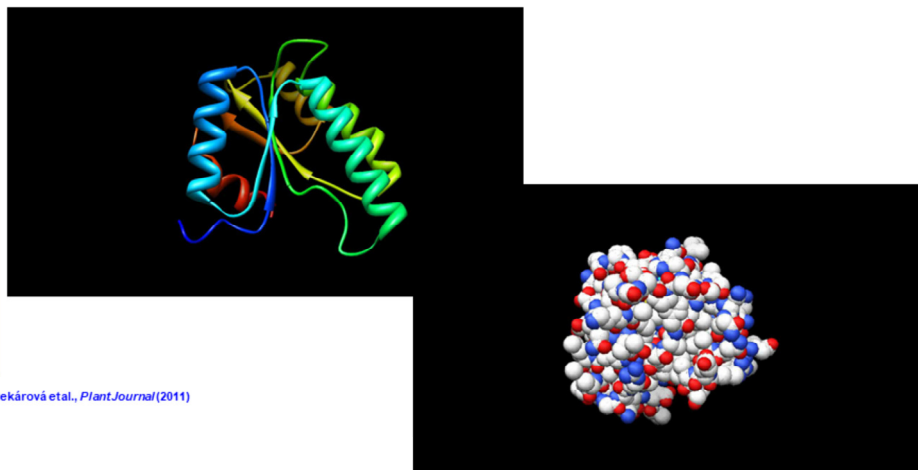


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Strukturální databáze

o **PDB** <http://www.rcsb.org/pdb/>



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



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



## 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 shows the Human Genome Browser interface. At the top, there's a navigation bar with 'Genomes', 'Genome Browser', 'Tools', 'Services', 'Downloads', 'My Data', and 'About Us'. Below this is a search form with fields for 'clade' (set to 'Human'), 'genome', 'assembly', and 'position'. A search button is visible. Below the search form, there's a section titled 'Human Genome Browser - hg19 assembly (sequences)'. It includes a 'Sample position queries' section with a list of requests and their corresponding descriptions. A small diagram of a human genome is visible on the right side of the page.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chr7p_g00212	Displays all of the unpaired contig g00212
20p13	Displays region for band p13 on chr 20
081.1:100000	Displays first million bases of chr 1, counting from p-arm telomere
chr3:100000-2000	Displays a region of chr3 that spans 2000 bases, starting with position 100000
RH1801:RH60175 15q11:15q13 rs154252/rs1600376	Displays region between genome landmarks, such as the STS markers RH1801 and RH60175, or chromosome bands 15q11 to 15q13, or SNPs rs154252 and rs1600376. This syntax may also be used for other range queries, such as between unpaired contigs, ESTs, mRNAs, refSeq, etc.
D18S3046	Displays region around STS marker D18S3046 from the Genethon/Manfield maps. Includes 100,000 bases on each side as well.
AK20414	Displays region of EST with GenBank accession AK20414 on BRCA1 cancer gene on chr 17
AC08101	Displays region of clone with GenBank accession AC08101
AF38211	Displays region of mRNA with GenBank accession number AF38211
FRNP	Displays region of genome with HSCO Gene Nomenclature Committee identifier FRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_056160	Displays the region of genome with protein accession number NP_056160
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs
homoexon cdistal	Lists mRNAs for causal homeobox genes
zinc_finger	Lists many zinc finger mRNAs
knapped_zinc_finger	Lists only knapped-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zeller	Lists mRNAs deposited by scientist named Zeller
Evans, J E	Lists mRNAs deposited by co-author J E Evans

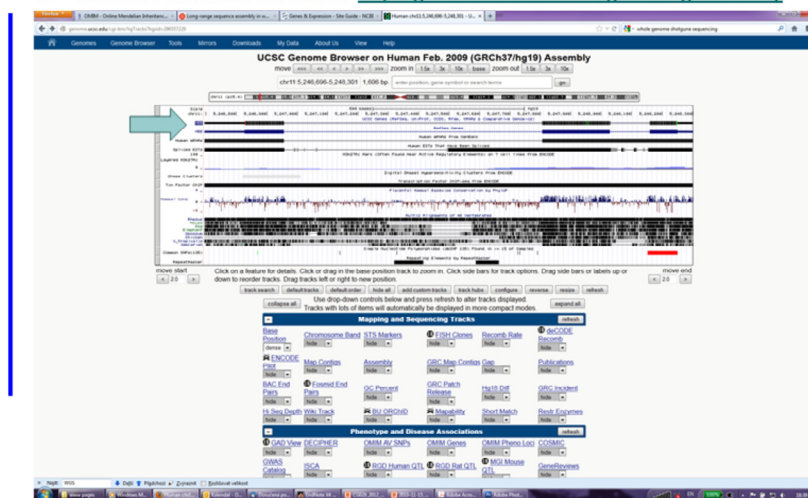


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

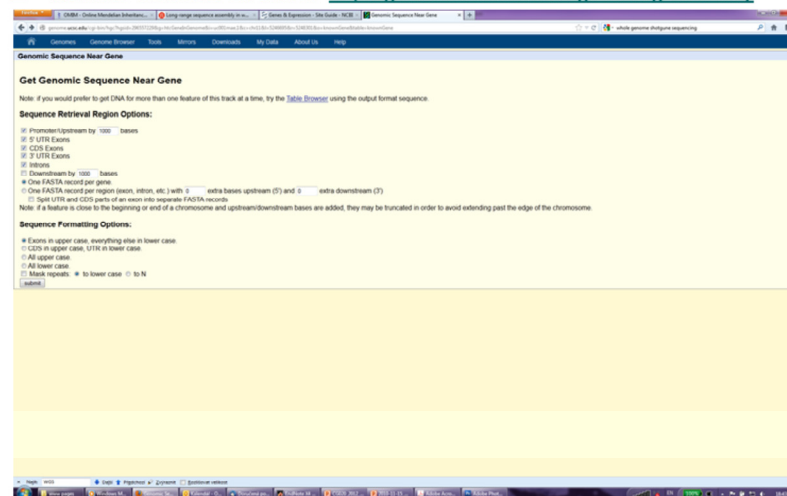
Human Genome Browser <http://genome.ucsc.edu/cgi-bin/hgGateway>

**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, HbA. 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 A - delta - beta 2 - beta 1 - gamma G - gamma B - epsilon. This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. [Entrez/Seq Attributes STATED](#)  
**Transcript, exon, coordinates, evidence:** [NM\\_000518.1](#) | [BC009186.1](#) | [CCO000002](#) | [M16766](#) | [M16767](#) | [M16768](#) | [M16769](#) | [M16770](#) | [M16771](#) | [M16772](#) | [M16773](#) | [M16774](#) | [M16775](#) | [M16776](#) | [M16777](#) | [M16778](#) | [M16779](#) | [M16780](#) | [M16781](#) | [M16782](#) | [M16783](#) | [M16784](#) | [M16785](#) | [M16786](#) | [M16787](#) | [M16788](#) | [M16789](#) | [M16790](#) | [M16791](#) | [M16792](#) | [M16793](#) | [M16794](#) | [M16795](#) | [M16796](#) | [M16797](#) | [M16798](#) | [M16799](#) | [M16800](#) | [M16801](#) | [M16802](#) | [M16803](#) | [M16804](#) | [M16805](#) | [M16806](#) | [M16807](#) | [M16808](#) | [M16809](#) | [M16810](#) | [M16811](#) | [M16812](#) | [M16813](#) | [M16814](#) | [M16815](#) | [M16816](#) | [M16817](#) | [M16818](#) | [M16819](#) | [M16820](#) | [M16821](#) | [M16822](#) | [M16823](#) | [M16824](#) | [M16825](#) | [M16826](#) | [M16827](#) | [M16828](#) | [M16829](#) | [M16830](#) | [M16831](#) | [M16832](#) | [M16833](#) | [M16834](#) | [M16835](#) | [M16836](#) | [M16837](#) | [M16838](#) | [M16839](#) | [M16840](#) | [M16841](#) | [M16842](#) | [M16843](#) | [M16844](#) | [M16845](#) | [M16846](#) | [M16847](#) | [M16848](#) | [M16849](#) | [M16850](#) | [M16851](#) | [M16852](#) | [M16853](#) | [M16854](#) | [M16855](#) | [M16856](#) | [M16857](#) | [M16858](#) | [M16859](#) | [M16860](#) | [M16861](#) | [M16862](#) | [M16863](#) | [M16864](#) | [M16865](#) | [M16866](#) | [M16867](#) | [M16868](#) | [M16869](#) | [M16870](#) | [M16871](#) | [M16872](#) | [M16873](#) | [M16874](#) | [M16875](#) | [M16876](#) | [M16877](#) | [M16878](#) | [M16879](#) | [M16880](#) | [M16881](#) | [M16882](#) | [M16883](#) | [M16884](#) | [M16885](#) | [M16886](#) | [M16887](#) | [M16888](#) | [M16889](#) | [M16890](#) | [M16891](#) | [M16892](#) | [M16893](#) | [M16894](#) | [M16895](#) | [M16896](#) | [M16897](#) | [M16898](#) | [M16899](#) | [M16900](#) | [M16901](#) | [M16902](#) | [M16903](#) | [M16904](#) | [M16905](#) | [M16906](#) | [M16907](#) | [M16908](#) | [M16909](#) | [M16910](#) | [M16911](#) | [M16912](#) | [M16913](#) | [M16914](#) | [M16915](#) | [M16916](#) | [M16917](#) | [M16918](#) | [M16919](#) | [M16920](#) | [M16921](#) | [M16922](#) | [M16923](#) | [M16924](#) | [M16925](#) | [M16926](#) | [M16927](#) | [M16928](#) | [M16929](#) | [M16930](#) | [M16931](#) | [M16932](#) | [M16933](#) | [M16934](#) | [M16935](#) | [M16936](#) | [M16937](#) | [M16938](#) | [M16939](#) | [M16940](#) | [M16941](#) | [M16942](#) | [M16943](#) | [M16944](#) | [M16945](#) | [M16946](#) | [M16947](#) | [M16948](#) | [M16949](#) | [M16950](#) | [M16951](#) | [M16952](#) | [M16953](#) | [M16954](#) | [M16955](#) | [M16956](#) | [M16957](#) | [M16958](#) | [M16959](#) | [M16960](#) | [M16961](#) | [M16962](#) | [M16963](#) | [M16964](#) | [M16965](#) | [M16966](#) | [M16967](#) | [M16968](#) | [M16969](#) | [M16970](#) | [M16971](#) | [M16972](#) | [M16973](#) | [M16974](#) | [M16975](#) | [M16976](#) | [M16977](#) | [M16978](#) | [M16979](#) | [M16980](#) | [M16981](#) | [M16982](#) | [M16983](#) | [M16984](#) | [M16985](#) | [M16986](#) | [M16987](#) | [M16988](#) | [M16989](#) | [M16990](#) | [M16991](#) | [M16992](#) | [M16993](#) | [M16994](#) | [M16995](#) | [M16996](#) | [M16997](#) | [M16998](#) | [M16999](#) | [M17000](#) | [M17001](#) | [M17002](#) | [M17003](#) | [M17004](#) | [M17005](#) | [M17006](#) | [M17007](#) | [M17008](#) | [M17009](#) | [M17010](#) | [M17011](#) | [M17012](#) | [M17013](#) | [M17014](#) | [M17015](#) | [M17016](#) | [M17017](#) | [M17018](#) | [M17019](#) | [M17020](#) | [M17021](#) | [M17022](#) | [M17023](#) | [M17024](#) | [M17025](#) | [M17026](#) | [M17027](#) | [M17028](#) | [M17029](#) | [M17030](#) | [M17031](#) | [M17032](#) | [M17033](#) | [M17034](#) | [M17035](#) | [M17036](#) | [M17037](#) | [M17038](#) | [M17039](#) | [M17040](#) | [M17041](#) | [M17042](#) | [M17043](#) | [M17044](#) | [M17045](#) | [M17046](#) | [M17047](#) | [M17048](#) | [M17049](#) | [M17050](#) | [M17051](#) | [M17052](#) | [M17053](#) | [M17054](#) | [M17055](#) | [M17056](#) | [M17057](#) | [M17058](#) | [M17059](#) | [M17060](#) | [M17061](#) | [M17062](#) | [M17063](#) | [M17064](#) | [M17065](#) | [M17066](#) | [M17067](#) | [M17068](#) | [M17069](#) | [M17070](#) | [M17071](#) | [M17072](#) | [M17073](#) | [M17074](#) | [M17075](#) | [M17076](#) | [M17077](#) | [M17078](#) | [M17079](#) | [M17080](#) | [M17081](#) | [M17082](#) | [M17083](#) | [M17084](#) | [M17085](#) | [M17086](#) | [M17087](#) | [M17088](#) | [M17089](#) | [M17090](#) | [M17091](#) | [M17092](#) | [M17093](#) | [M17094](#) | [M17095](#) | [M17096](#) | [M17097](#) | [M17098](#) | [M17099](#) | [M17100](#) | [M17101](#) | [M17102](#) | [M17103](#) | [M17104](#) | [M17105](#) | [M17106](#) | [M17107](#) | [M17108](#) | [M17109](#) | [M17110](#) | [M17111](#) | [M17112](#) | [M17113](#) | [M17114](#) | [M17115](#) | [M17116](#) | [M17117](#) | [M17118](#) | [M17119](#) | [M17120](#) | [M17121](#) | [M17122](#) | [M17123](#) | [M17124](#) | [M17125](#) | [M17126](#) | [M17127](#) | [M17128](#) | [M17129](#) | [M17130](#) | [M17131](#) | [M17132](#) | [M17133](#) | [M17134](#) | [M17135](#) | [M17136](#) | [M17137](#) | [M17138](#) | [M17139](#) | [M17140](#) | [M17141](#) | [M17142](#) | [M17143](#) | [M17144](#) | [M17145](#) | [M17146](#) | [M17147](#) | [M17148](#) | [M17149](#) | [M17150](#) | [M17151](#) | [M17152](#) | [M17153](#) | [M17154](#) | [M17155](#) | [M17156](#) | [M17157](#) | [M17158](#) | [M17159](#) | [M17160](#) | [M17161](#) | [M17162](#) | [M17163](#) | [M17164](#) | [M17165](#) | [M17166](#) | [M17167](#) | [M17168](#) | [M17169](#) | [M17170](#) | [M17171](#) | [M17172](#) | [M17173](#) | [M17174](#) | [M17175](#) | [M17176](#) | [M17177](#) | [M17178](#) | [M17179](#) | [M17180](#) | [M17181](#) | [M17182](#) | [M17183](#) | [M17184](#) | [M17185](#) | [M17186](#) | [M17187](#) | [M17188](#) | [M17189](#) | [M17190](#) | [M17191](#) | [M17192](#) | [M17193](#) | [M17194](#) | [M17195](#) | [M17196](#) | [M17197](#) | [M17198](#) | [M17199](#) | [M17200](#) | [M17201](#) | [M17202](#) | [M17203](#) | [M17204](#) | [M17205](#) | [M17206](#) | [M17207](#) | [M17208](#) | [M17209](#) | [M17210](#) | [M17211](#) | [M17212](#) | [M17213](#) | [M17214](#) | [M17215](#) | [M17216](#) | [M17217](#) | [M17218](#) | [M17219](#) | [M17220](#) | [M17221](#) | [M17222](#) | [M17223](#) | [M17224](#) | [M17225](#) | [M17226](#) | [M17227](#) | [M17228](#) | [M17229](#) | [M17230](#) | [M17231](#) | [M17232](#) | [M17233](#) | [M17234](#) | [M17235](#) | [M17236](#) | [M17237](#) | [M17238](#) | [M17239](#) | [M17240](#) | [M17241](#) | [M17242](#) | [M17243](#) | [M17244](#) | [M17245](#) | [M17246](#) | [M17247](#) | [M17248](#) | [M17249](#) | [M17250](#) | [M17251](#) | [M17252](#) | [M17253](#) | [M17254](#) | [M17255](#) | [M17256](#) | [M17257](#) | [M17258](#) | [M17259](#) | [M17260](#) | [M17261](#) | [M17262](#) | [M17263](#) | [M17264](#) | [M17265](#) | [M17266](#) | [M17267](#) | [M17268](#) | [M17269](#) | [M17270](#) | [M17271](#) | [M17272](#) | [M17273](#) | [M17274](#) | [M17275](#) | [M17276](#) | [M17277](#) | [M17278](#) | [M17279](#) | [M17280](#) | [M17281](#) | [M17282](#) | [M17283](#) | [M17284](#) | [M17285](#) | [M17286](#) | [M17287](#) | [M17288](#) | [M17289](#) | [M17290](#) | [M17291](#) | [M17292](#) | [M17293](#) | [M17294](#) | [M17295](#) | [M17296](#) | [M17297](#) | [M17298](#) | [M17299](#) | [M17300](#) | [M17301](#) | [M17302](#) | [M17303](#) | [M17304](#) | [M17305](#) | [M17306](#) | [M17307](#) | [M17308](#) | [M17309](#) | [M17310](#) | [M17311](#) | [M17312](#) | [M17313](#) | [M17314](#) | [M17315](#) | [M17316](#) | [M17317](#) | [M17318](#) | [M17319](#) | [M17320](#) | [M17321](#) | [M17322](#) | [M17323](#) | [M17324](#) | [M17325](#) | [M17326](#) | [M17327](#) | [M17328](#) | [M17329](#) | [M17330](#) | [M17331](#) | [M17332](#) | [M17333](#) | [M17334](#) | [M17335](#) | [M17336](#) | [M17337](#) | [M17338](#) | [M17339](#) | [M17340](#) | [M17341](#) | [M17342](#) | [M17343](#) | [M17344](#) | [M17345](#) | [M17346](#) | [M17347](#) | [M17348](#) | [M17349](#) | [M17350](#) | [M17351](#) | [M17352](#) | [M17353](#) | [M17354](#) | [M17355](#) | [M17356](#) | [M17357](#) | [M17358](#) | [M17359](#) | [M17360](#) | [M17361](#) | [M17362](#) | [M17363](#) | [M17364](#) | [M17365](#) | [M17366](#) | [M17367](#) | [M17368](#) | [M17369](#) | [M17370](#) | [M17371](#) | [M17372](#) | [M17373](#) | [M17374](#) | [M17375](#) | [M17376](#) | [M17377](#) | [M17378](#) | [M17379](#) | [M17380](#) | [M17381](#) | [M17382](#) | [M17383](#) | [M17384](#) | [M17385](#) | [M17386](#) | [M17387](#) | [M17388](#) | [M17389](#) | [M17390](#) | [M17391](#) | [M17392](#) | [M17393](#) | [M17394](#) | [M17395](#) | [M17396](#) | [M17397](#) | [M17398](#) | [M17399](#) | [M17400](#) | [M17401](#) | [M17402](#) | [M17403](#) | [M17404](#) | [M17405](#) | [M17406](#) | [M17407](#) | [M17408](#) | [M17409](#) | [M17410](#) | [M17411](#) | [M17412](#) | [M17413](#) | [M17414](#) | [M17415](#) | [M17416](#) | [M17417](#) | [M17418](#) | [M17419](#) | [M17420](#) | [M17421](#) | [M17422](#) | [M17423](#) | [M17424](#) | [M17425](#) | [M17426](#) | [M17427](#) | [M17428](#) | [M17429](#) | [M17430](#) | [M17431](#) | [M17432](#) | [M17433](#) | [M17434](#) | [M17435](#) | [M17436](#) | [M17437](#) | [M17438](#) | [M17439](#) | [M17440](#) | [M17441](#) | [M17442](#) | [M17443](#) | [M17444](#) | [M17445](#) | [M17446](#) | [M17447](#) | [M17448](#) | [M17449](#) | [M17450](#) | [M17451](#) | [M17452](#) | [M17453](#) | [M17454](#) | [M17455](#) | [M17456](#) | [M17457](#) | [M17458](#) | [M17459](#) | [M17460](#) | [M17461](#) | [M17462](#) | [M17463](#) | [M17464](#) | [M17465](#) | [M17466](#) | [M17467](#) | [M17468](#) | [M17469](#) | [M17470](#) | [M17471](#) | [M17472](#) | [M17473](#) | [M17474](#) | [M17475](#) | [M17476](#) | [M17477](#) | [M17478](#) | [M17479](#) | [M17480](#) | [M17481](#) | [M17482](#) | [M17483](#) | [M17484](#) | [M17485](#) | [M17486](#) | [M17487](#) | [M17488](#) | [M17489](#) | [M17490](#) | [M17491](#) | [M17492](#) | [M17493](#) | [M17494](#) | [M17495](#) | [M17496](#) | [M17497](#) | [M17498](#) | [M17499](#) | [M17500](#) | [M17501](#) | [M17502](#) | [M17503](#) | [M17504](#) | [M17505](#) | [M17506](#) | [M17507](#) | [M17508](#) | [M17509](#) | [M17510](#) | [M17511](#) | [M17512](#) | [M17513](#) | [M17514](#) | [M17515](#) | [M17516](#) | [M17517](#) | [M17518](#) | [M17519](#) | [M17520](#) | [M17521](#) | [M17522](#) | [M17523](#) | [M17524](#) | [M17525](#) | [M17526](#) | [M17527](#) | [M17528](#) | [M17529](#) | [M17530](#) | [M17531](#) | [M17532](#) | [M17533](#) | [M17534](#) | [M17535](#) | [M17536](#) | [M17537](#) | [M17538](#) | [M17539](#) | [M17540](#) | [M17541](#) | [M17542](#) | [M17543](#) | [M17544](#) | [M17545](#) | [M17546](#) | [M17547](#) | [M17548](#) | [M17549](#) | [M17550](#) | [M17551](#) | [M17552](#) | [M17553](#) | [M17554](#) | [M17555](#) | [M17556](#) | [M17557](#) | [M17558](#) | [M17559](#) | [M17560](#) | [M17561](#) | [M17562](#) | [M17563](#) | [M17564](#) | [M17565](#) | [M17566](#) | [M17567](#) | [M17568](#) | [M17569](#) | [M17570](#) | [M17571](#) | [M17572](#) | [M17573](#) | [M17574](#) | [M17575](#) | [M17576](#) | [M17577](#) | [M17578](#) | [M17579](#) | [M17580](#) | [M17581](#) | [M17582](#) | [M17583](#) | [M17584](#) | [M17585](#) | [M17586](#) | [M17587](#) | [M17588](#) | [M17589](#) | [M17590](#) | [M17591](#) | [M17592](#) | [M17593](#) | [M17594](#) | [M17595](#) | [M17596](#) | [M17597](#) | [M17598](#) | [M17599](#) | [M17600](#) | [M17601](#) | [M17602](#) | [M17603](#) | [M17604](#) | [M17605](#) | [M17606](#) | [M17607](#) | [M17608](#) | [M17609](#) | [M17610](#) | [M17611](#) | [M17612](#) | [M17613](#) |

# 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 homepage of the Arabidopsis Information Resource (TAIR). The page features a navigation menu with options like Search, Browse, Tools, Portals, Download, Submit, News, and ARIC Blocks. The main content area includes a welcome message, a list of news items such as 'Breaking News', '2012 MASC Report New Available', and 'New Protein Chip and Call Culture at ARIC', and a prominent call to action: 'Click here to try our new online submission form'. The website is designed with a clean, professional layout and includes social media links.

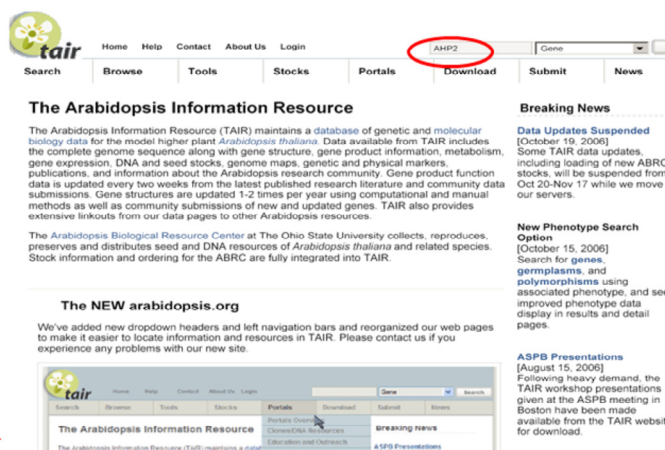


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Genomové zdroje

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



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

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

**The NEW arabidopsis.org**

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

**Breaking News**

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

**New Phenotype Search Option**  
[October 15, 2006]  
Search for genes, germplasms, and polymorphisms using associated phenotype, and see improved phenotype data display in results and detail pages.

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



## 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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Analytické nástroje

## □ Globální vs. lokální přiřazení

### Globální přiřazení

```
SLAV-----APATNIK-----PIQNYR-I-----AKSETQRYMVE  
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVE
```

### Lokální přiřazení

```
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVE  
-----NAPATNIKSECVRA-PIQNYRRVEHVRA-----
```

Cvrčková, Úvod do praktické bioinformatiky

- **Globální přiřazení** pouze u sekvencí, které jsou si **podobné a podobné délky** (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

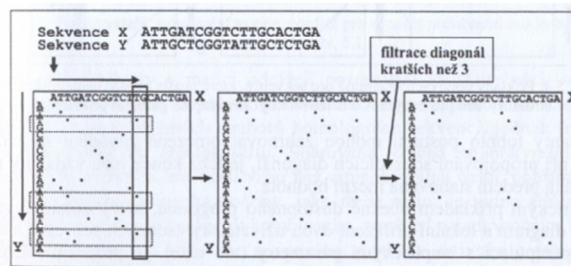


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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)

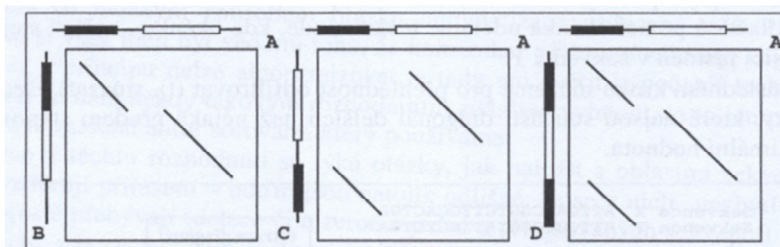


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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 srovnávání programem BLAST2 (viz dále)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

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

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

[Search](#)

```
acaccatcgc  
acaccatcgc cattatcacc atcgctttgg ggcgatgttg tgggttcca  
gcgatataat  
ataattaatt tattccacat gagatgatg atgatatact atgtattttt  
tttttttttt  
ttatttgtaa acotttaata taacaagaac tacaaaaaat gaaa
```

[Set subsequence](#) From:  To:

[Choose database](#)

Now: **BLAST!** or [Reset query](#) [Reset all](#)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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

Matice PAM 250

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



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

# BLAST

## Basic Local Alignment Search Tool

>gi|5016088|ref|NM\_001101.2| Length = 1793 actin, beta (ACTB), mRNA

Score = 1110 bits (560), Expect = 0.0  
Identities = 965/1100 (87%)  
Strand = Plus / Plus

Query: 156 gtgcacaacggctctggcatgtgcaaggccggattgccggagacgatgctccccggcc 215  
Sbjct: 101 gtgcacaacggctccggcatgtgcaaggccggcttcggggcgacgatgcccccgggcc 160

Query: 216 gtcttcccatcgattgtgggacgtccccgtcaccagggtgtgatggtggcatggccag 275  
Sbjct: 161 gtcttccctccatcgtggggcgcaccaggcaccaggcgtgatggtggcatgggtcag 220

Query: 276 aaggactcgtacgtgggtgatgaggccagagcaagcgtggtatcctcaccctgaagtac 335  
Sbjct: 221 aaggattcctatgtggggcagagccagagcaagagagggcctcctcaccctgaagtac 280

Query: 336 cccattgagcacggatcgtgaccaactgggacgatggagaagatctggcaccacacc 395  
Sbjct: 281 cccatcgagcacggatcgtcaccaactgggacgatggagaaatctggcaccacacc 340

E=expectancy value

ds..S=1213 E=0.0

>=200

250 1500

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



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

**Bibliography**

**Related articles in PubMed**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

BLINK is a link to the pre-computed BLAST search results for the respective sequence (see the next slide).

# BLAST

## Basic Local Alignment Search Tool

Pre-computed BLAST results for: [at16119781rvf/NP\\_396485.1](#) two component sensor kinase [Agrobacterium tumefaciens str. C58]  
 Matching gis: [15163423.20141874.1019660](#)  
 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 138295 Bacteria 13 Metazoa 1349 Fungi 554 Plants 6 Viruses 5676 The Others [reset selection](#)

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

% hits	Score	Accession	Length	Protein Description
833 aa				
4166	AM99527	833	two component sensor kinase [Agrobacterium tumefaciens str. C58]	
4166	P18548	833	RecName: Full=Wide host range virA protein; Short=WRB virA	
4166	AAA79262	833	virA [Plasmid pTIC58]	
4159	NP_053300	833	hypothetical protein pT1-GAMMA_p142 [Agrobacterium tumefaciens]	
4159	AAA07765	833	tiorf140 [Agrobacterium tumefaciens]	
4153	AAA91590	833	virA [Plasmid Ti]	
4153	gi1737127	833	virA protein	
4153	CAA34777	833	91.3 kDa protein [Agrobacterium tumefaciens]	
3800	CAA33380	829	virA [Agrobacterium rhizogenes]	
3718	gi1227240	849	virA gene	
3148	AAA88643	829	virA [Plasmid Ti]	



### INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# 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**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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



### INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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 (neznačená nic)
      - x(5) znamená 5 jakýchkoliv aminokyselin
      - x(3, 5) znamená 3 až 5 jakýchkoliv aminokyselin



### INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# BLAST

## Specializované verze

### □ Příklad vyhledávání pomocí PHI-BLAST

```
>gi|4758958|ref|NP_004148.1| Human cAMP-dependent protein kinase  
MSHIQIPPGLELLQGYTVEVLRQPPDLVEFAVEYFTRLREARAPASVLPAAATPRQSLGHPPPPEPGPDR  
VADAKGDSSESEDEDELEVVPVPSRFNRRVSVCAETYNPDEEBEDTDPRVIHPKTDEQRCRLQBEACKDILLF  
KNLDQEQLSQVLDAMFERIVKADHVIDQGDDGDNFYVIERGTYDILVTKDNQTRSVGQYDNRGSFGELA  
LMYNTPRAAITVATSEGSWGLDRVTFRRIIVKNNAKRRKMFESFIESVPLLSLEVSRMKIVDVIGEK  
IYKDGERRIITQGEKADSFYIIBSGEVSILIRSRTKSNKDGNGQBEVEIARCHKGQYFGBLALVINKPRAAS  
AYAVGDVKCLVMDVQAFERLLGPCMDIMKRNI SHYBQLVKMFSSVDLGNLGQ
```

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



#### 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....



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- o **Biology Workbench** <http://workbench.sdsc.edu/>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- o **Biology Workbench** <http://workbench.sdsc.edu/>

View  
View Nucleic Sequence(s)

Format: Fasta Case: Upper Change Format

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

[NEXT] [BOTTOM]

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

> 170248  
GAGCTCCCTTGGGGGGCAAGGGCAAAAACCTTTTGGCTAAATGGAAAAATATATACCAGTGTGTTGTAATA  
GTTACTCAATTTGAATTAACAAGGGGGCAATTTGACTATTTTGGCCCTTATATCTTTTGGTCACAAAAC  
ATAAATATCCCATCCGAAATTCAAATGGTCCATATCGGCAAGTAGCTTCTTTAATTTAGTTAGTT  
GACAAAACACTATCAAGATATCATTAATTAATAAATAACTTCAAGTCCATCATCTTAGCTGCTCCTCA  
GTTAGGCCCGCAGTAAATAGACCGATCAATAAAGCCCGCATTAATAATGAAATTTAGGACTCTC  
GATTTGGACGTAACTCCAAAACCTTTCAATACCTTTTCTCCAACTTTGGGGCTCTGAGTCTGACTTC  
CAGATATGGGATATTTCTAGTTTATCTCTTAATTTACATCTCAACTAATATTAAGAAATAAACAGGTA  
CAGCAATCATAAAATTTCTCTTAAGAAAGCAATGAATCCGGTTACTGATTCATTGGCCTTTTCAGAG  
TCATGATCCCATATTCCTAAGGGGTCGTTTGGTACAAAGAAATAATAATAATTTGGGATAGAAATTT  
GAGATTCATTTATCTTTGTTTAAATTAAGATTTAGCTAATTCAGAAATTTTTCATTAATTTAGTAAATAG  
TAAATCACTTTCATCTGAAAGTGAATGGATAGCTAATCCATGGCCACTCACTAGAAATTTCT  
TTATTTATCTACATTTTACCAATGATCGTTAGTCTTCAAGAGATCCAGTATCTCAATAAATGCA  
GTAGAAAGTTAGAAAATTTCTAATTAATCAATCATATAATTTAAAATATTAAGATAGGAGCTTAAG  
ATACATAAAGATGTACCGTTAATAATAAAGATAAGATAGATTTTAAATAGGAAAAAAAACGGTT  
CGAGACTCTTTATGGGAAGGGCTTGTCTCAAGTAGATTTCTATCATTTGCTCTGGTGCANTAGCAAAA  
TACACTTTAGCTTTAAGATTCAGCGAGCCACTTCAATCTCTTATTTATCTCAAGTGAAGTTTAA  
GGAACTTCAAACTTCACTACTTTTAAAGGAAATCAAAATACGACCAATTTATTTACTTACTTAC  
TTATAGTTAAATGATATGAATTTTAAATTTGAATTTGAAAATATTAATTTACTTGTATTAATATA



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# Analytické nástroje

- o **Biology Workbench** <http://workbench.sdsc.edu/>

**Regex pattern:**

ott. {1, 32}ott

0 sequences were searched

1 match was found

Matches are indicated in blue

```
>170248
GAGCTCCCTTGGGGGGCAAGGGCAAACCTTTGGCTAAATGGAAAAATATATACCAAGTGTGTAATA
GTATCTCAATTGAAATTAACAAAGGGCAAAATTTGACTATTTTGGCCCTTATATCTTTTGGTCACAAAAAC
ATAAATATCCCATCCGAATTCGAATGGTCCATATCGGCAAGTAGCTTTCTTTTAAATATAGTTAGTT
GACAAACCTATCCAGATATCTATTATTAATATATTAATTTCAAGGTCCTTCTTTAGTCCCTCTCA
GTAGAGCCGCGATAAATAGACCGATCAANTRAAAGCCGCCATTAATAATAGAAATTTAGGACTCTC
GATGGCACGTAAAGTCCAAAACCTCTCCAAATCTTGGCTGCAACTTGGGGCTGCTAGGTTCTGAGCTTC
CAGATATGGGATATCTAAGTTTATCTCCTAATTCATCTCAACTATATTAAGAAATTAACAGGTA
CGCAAACTATAAAATTTCTCTAAAGAGACAAATCCGGTTACTATTCATGSSCTTTCTAGAG
TCTGATGCCATTTACTAAGGGGCTGTTGGTACAAGAAATATAATAATTTCCGGATAGAAATTT
TAAATCAACTATACATGTAGAGGTGGAATGGAATAGTAATCCATAGCCACTACATAGAAATCC
TATTATCTACTATTTTACCAATGATCGGTTAGTCTTCATGAAATCCAGTATCCCAATTAATGCA
GTAGAAATTTAGAAATTTTCAATTAATCAATTTCAATTTTAAAAATTTAGATTTGGACACTTAG
ATACAATAAAGATGACCGTAAATTAATAAAGATAGATAGAGTTTAAATAGGAAAAAAAACGGTT
CGAGACTCTTATGGAGGGCTTGTCTCAAGGTAGATCTCATTCATGCTCTGGTCAATAGCAAAA
TGACATTTACTCTTAGATACAGCGACCTCTACAACTTCTATTTGTAATCTAAATGAAAGTTTAA
GAGAACTTCAAACTCTCACTACTCTTAGGGAAATCAAAATACGACCAATTTATTAATTTACTTAC
TTATGTTAAATGATAGAAATTTTATTTAAATTTGAAATGAAATTTAAATTTAGTTTAAATATAA
ACAATAGATATCGCTAAGTATTTACCAACAACATGGAGATACAGAAAGATTTATTTATTTGTAAGAT
GATTAAGCAGCTATTCATCTGGTTGTGAGGATGAAGAAAGTAACAGCTATTAATTTCTTTGTAAGT
```



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- **Biology Workbench** <http://workbench.sdsc.edu/>

Frame 1, 1 stop codon

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

>170248 Translated - Frame 1  
ELPWGARA K L FAKWKNI I P S V C N S Y S I \* I N K G A N L T I L P L

E L P W G A R A K L F A K W K N I I P S  
1 g a g c t c c o t t g g g g g c a a g g g c a a a a c t t t t g c t a a t g g a a a a t a t t a t a c c a a g t 60  
V C N S Y S I \* I N K G A N L T I L P L  
61 g t t t g t a a t a g t t a c t c a a t t t g a a t t a a c a a a g g g c a a a t t g a c t a t t t t g c c o t t a 120

Frame 2, 1 stop codon

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

>170248 Translated - Frame 2  
S S L G G Q Q N F L L N G K I L Y Q V F E L T R G Q I \* L F C P

S S L G G Q Q N F L L N G K I L Y Q V  
2 a g c t c c o t t g g g g g c a a g g g c a a a a c t t t t g c t a a t g g a a a a t a t t a t a c c a a g t 61  
F V I V T O F E L T K G O I \* L F C P  
62 t t t g t a a t a g t t a c t c a a t t t g a a t t a a c a a a g g g c a a a t t g a c t a t t t t g c c o t t a 120



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- o **Biology Workbench** <http://workbench.sdsc.edu/>

```
= Linear Map of Sequence:

      SbyI
      BsaJI
      CviJI
      AluI
      SacI
      EcoICRI
      Bsp1286I
      BsiHKAI
      BanII  BslI
      \ \ \ \ \
1 gagctcccttgggggcaagggaacaaacttttggtaaatgaaaaatattataccaagt 60
ctcgagggaacccccctcccgtttggaaaaagattaccttttataataggttca
      * * * * *
1 E L P W G A R A K L F A K W K N I I P S
2 S S L G G Q G Q N F L L N G K I L Y Q V
3 A P L G G K G K T F C * M E K Y Y T K C
4 L E R P P C P C F K K S F F F I N Y W T
5 S S G Q P A L A F S K A L H F F I I G L
6 L A G K P P L P L V K Q * I S F Y * V L

      Tsp509I
      MaeIII Tsp509I  MseI
      \ \ \ \ \
61 gtttgaatggttactcaattgaattaacaaagggaacaaattgactattttgcoccta 120
caaacattatcaatgagttaaacttaattgtttccocggttaaacgtataaacggggaat
      * * * * *
1 V C N S Y S I * I N K G A N L T I L P L
2 F V I V T Q F E L T K G Q I * L F C P *
3 L * * L L N L N * Q R G K F D Y F A L R
4 N T I T V * N S N V F P C I Q S N Q G *
5 T Q L L * E I Q I L L P A F K V I K G K
6 H K Y Y N S L K F * C L P L N S * K A R
```

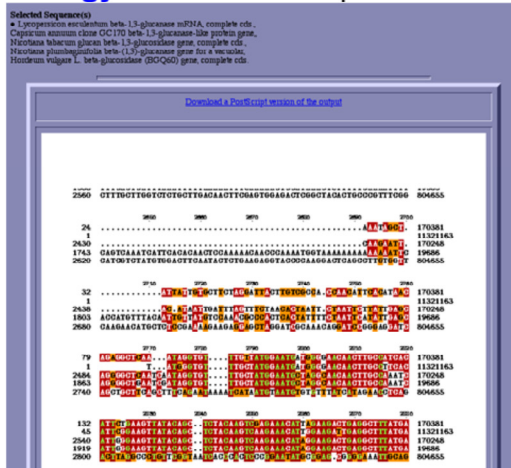


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- **Biology Workbench** <http://workbench.sdsc.edu/>

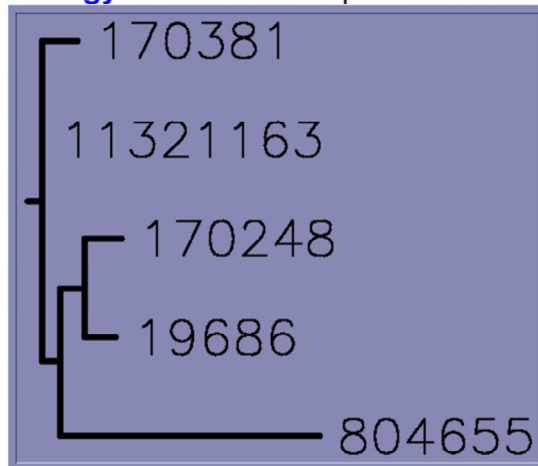


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- o **Biology Workbench** <http://workbench.sdsc.edu/>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

- Virtual PCR (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 (IUB 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 instability 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: BLAST in the database for: M. musculus

Primer 1  
Primer 2  
Primer 3  
Primer 4  
Primer 5  
Primer 6  
Primer 7  
Primer 8

Annealing temperature: 50

Do PCR! 



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



MASARYKOVA UNIVERZITA  
BRNO

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Analytické nástroje

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



## 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](#)



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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



# 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 includes the following information:

- Gene ID:** 65978, updated on 27-Aug-2011
- Official Symbol:** PHACTR4 (provided by HUGO)
- Official Full Name:** phosphatase and actin regulator 4 (provided by HUGO)
- Primary source:** [HGNC:2573](#)
- Location:** 1p11.42NC1\_A1
- See also:** [Ensembl:ENSG00000248138](#), [HGVS:HG1838](#), [MM:628725](#)
- Gene type:** protein coding
- RefSeq status:** REVIEWED
- Organism:** [Homo sapiens](#)
- Lineage:** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhina; Catarrhini; Hominoidea; Homo
- Also known as:** [FLJ13171](#), [MGC25818](#), [MGC34186](#), [Daf-2a68L7.205](#), [PP11-442NC1\\_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: 1 p11.4; Sequence: Chromosome 1, NC\_000001.10 (3894993-3895891)
- Genomic regions, transcripts, and products:** Includes a diagram of the gene structure on Chromosome 1 and a table of genomic features.



MINISTERSTVO  
MLÁDEŽE

JE VZDĚLÁVÁNÍ  
je spolufinancována  
kým sociálním fondem  
Česka a Evropské unie  
řídí Ministerstvo školství, mládeže a tělovýchovy  
Česka

# Další WWW zdroje

- **Online Mendelian Inheritance in Man (OMIM)** <http://www.omim.org/>



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# 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](#)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

# Diskuse



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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