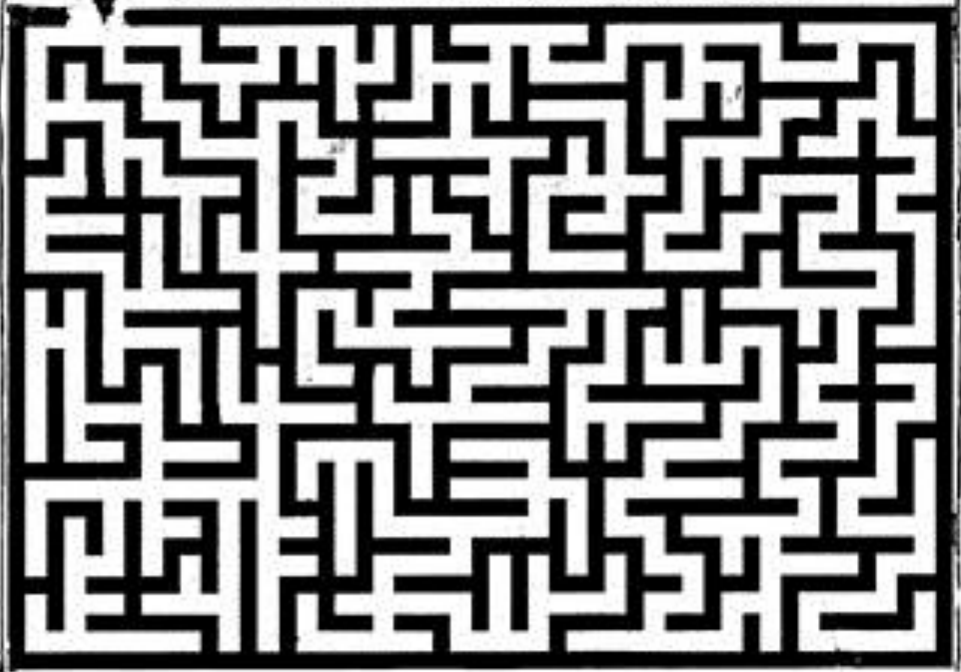


↓ **S námi nezabloudíte !**



RENČIN

Využití internetových zdrojů při studiu mikroorganismů

doc. RNDr. Milan Bartoš, Ph.D.

Bartos.Milan@atlas.cz

Přírodovědecká fakulta MU, 2017

Obsah přednášky

- 1) Práce se sekvenčními daty**
- 2) Základní veřejně dostupné databáze**
- 3) Práce se stránkami NCBI**
- 4) Jak se posuzuje podobnost sekvencí**
- 5) Prohledávač BLAST, BLAST2**
- 6) Mnohočetné přiřazení – program CLUSTAL**



Doporučená literatura

Cvrčková F. (2006):

**Úvod do praktické bioinformatiky, Academia
Praha**

<http://www.ncbi.nlm.nih.gov/>

Práce se sekvenčními daty

Sekvenční data = zápis primární sekvence makromolekul, tj. DNA (RNA) a proteinů

- **DNA a RNA se zapisují ve směru 5' - 3'**
- **Proteiny se zapisují od N-konce k C-konci**
- **Používají se jednopísmenkové kódy (podle IUPAC)**

Zkratky pro nukleové kyseliny

DNA, RNA			
Kód	Báze	Kód	Báze
A	Adenin	K	G, T (<u>k</u> eto)
C	Cytosin	M	A, C (<u>a</u> mino)
G	Guanin	B	C, G, T (ne A)
T	Tymin	D	A, G, T (ne C)
U	Uracil	H	A, C, T (ne G)
R	A, G (<u>p</u> urin)	V	A, C, G (ne T, U)
Y	C, T (<u>p</u> yrimidin)	N	cokoli (<u>a</u> ny)
S	G, C (<u>s</u> trong)	.	mezera
W	A, T (<u>w</u> eak)	-	

Zkratky pro proteiny

Kód	Zkratka	Amino kyselina	Kód	Zkratka	Amino kyselina
A	Ala	Alanin	P	Pro	Prolin
C	Cys	Cystein	Q	Gln	Glutamin
D	Asp	Aspartát	R	Arg	Arginin
E	Glu	Glutamát	S	Ser	Serin
F	Phe	Fenylalanin	T	Thr	Threonin
G	Gly	Glycin	V	Val	Valin
H	His	Histidin	W	Trp	Tryptofan
I	Ile	Izoleucin	Y	Tyr	Tyrosin
K	Lys	Lysin	X	Xxx	cokoli
L	Leu	Leucin	B	Asx	Asp, Asn
M	Met	Methionin	Z	Glx	Glp, Gln
N	Asn	Asparagin			

Způsoby zápisu

Surová data (raw data, raw format)

- Některé programy je umí přijmout a zpracovat
- Nejsou ale vhodné pro dlouhodobé uchování

Specializované formáty

- Základní veřejné databáze je umí převádět

Jednoduché formáty - FASTA

- Nejlépe bez mezer a speciálních znaků

>gi|291219937|ref|NM_001888.3| Homo sapiens crystallin, mu (CRYM),
transcript variant 1, mRNA

```
TTTCAAATGGGGAGTTTCCCTGCACAAGCTTTCTTGTCTGCCACTATGTGAGATATACCTT  
TCACCTTCTGCCGTGATTGTGAGGCCTCCTCAGCCACGTGGAAGTGTAAAACTCCTGGAA  
GAAAAGATCCTGCAATTT
```


FASTA a WORD

Na co si dát pozor

- **Uložit ve formátu „pouze text“**
- **Nepoužívat tabelátory a jiné cizí znaky**
- **Vypnout funkce „automatické opravy“ a „automatický text“ i funkce „inteligentní vyjímání a vkládání“**

Typ písma

Doporučuji formát písma „Courier New“
– každé písmeno zaujímá stejnou
plochu

Courier New 24

TTTCAAATGGGGAGTTTCCCTGCACAAGCTTTCTT
AAAGTTTACCCCTCAAAGGGACGTGTTTCGAAAGAA

Arial 24

TTTCAAATGGGGAGTTTCCCTGCACAAGCTTTCTT
AAAGTTTACCCCTCAAAGGGACGTGTTTCGAAAGAA

**Pozor, zkratky pro NA a proteiny jsou
v některých případech shodné!**

**Vstupní formáty pro počítačové zpracování
musí být specifikovány, aby program rozpoznal,
jde-li o NA nebo protein**



Molekulárně-biologické databáze

Evropský institut pro bioinformatiku ve Velké Británii (EBI)

EMBL, 1980

www.ebi.ac.uk

Národní centrum pro biotechnologické informace (NCBI) založené v rámci Národní lékařské knihovny (NLM) v USA

GenBank, 1982

www.ncbi.nlm.nih.gov

Centrum pro informační biologii (CIB) , jako oddělení Národního genetického institutu (NIG) v Japonsku

DDBJ, 1984

www.cib.nig.ac.jp

GenBank/EMBL/DDBJ

- **Vzájemně si vyměňují si informace**
- **Volně dostupné**
- **Přijímají nové sekvence z genomových center a pracovišť zabývajících se sekvenováním**



Sekvenci v databázích může zveřejnit kdokoli !

Databáze sekvencí proteinů

**Databáze SWISS-PROT založená na Univerzitě
v Ženevě v roce 1986**

Spravuje Švýcarský institut pro bioinformatiku (SIB)

www.expasy.org

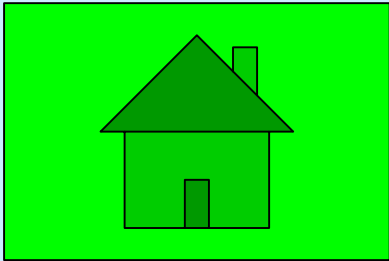
**Obsahuje automaticky doplňované překlady
sekvencí z EMBL**

Databáze PDB (The Protein Databank)

**Archivuje a analyzuje proteinové struktury a
komplexy informačních biomakromolekul**

<http://www.rcsb.org/pdb/home/home.do>

Práce s databází NCBI



www.ncbi.nlm.nih.gov

A screenshot of the NCBI website. The top navigation bar includes 'NCBI Resources' and 'How To'. A search bar is present with a dropdown menu set to 'All Databases'. The main content area is divided into a left sidebar with a 'Resource List (A-Z)' and a central 'Welcome to NCBI' section. A red box highlights the 'Get Started' section, which contains links for Tools, Downloads, How-To's, and Submissions. A red arrow points from the 'Domains & Structures' link in the sidebar to the 'Get Started' box. On the right, there are sections for 'Popular Resources' and 'NCBI Announcements'.

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

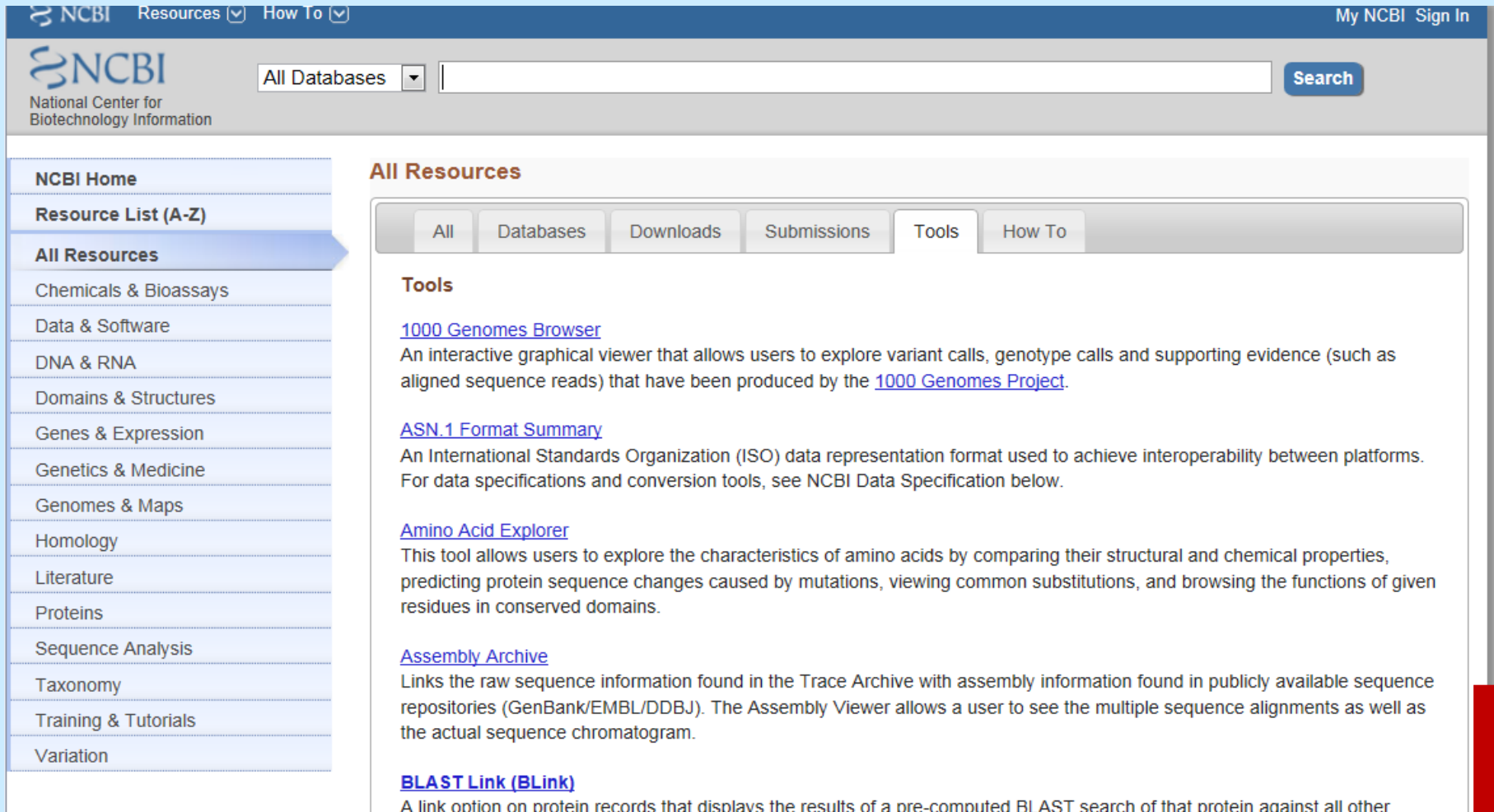
Popular Resources
PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI Announcements
New Microbial BLAST Page
12 Jun 2012
Now easier to use and with the familiar format and features of the standard NCBI BLAST services, including auto-complete
Sign up for the Fall Discovery Workshops!

Genomic Structural Variation
dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

1 2 3 4 5 6 7 8

Práce s databází NCBI



NCBI Resources ▾ How To ▾ My NCBI Sign In

NCBI
National Center for
Biotechnology Information

All Databases ▾ Search

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

All Resources

All Databases Downloads Submissions **Tools** How To

Tools

[1000 Genomes Browser](#)
An interactive graphical viewer that allows users to explore variant calls, genotype calls and supporting evidence (such as aligned sequence reads) that have been produced by the [1000 Genomes Project](#).

[ASN.1 Format Summary](#)
An International Standards Organization (ISO) data representation format used to achieve interoperability between platforms. For data specifications and conversion tools, see NCBI Data Specification below.

[Amino Acid Explorer](#)
This tool allows users to explore the characteristics of amino acids by comparing their structural and chemical properties, predicting protein sequence changes caused by mutations, viewing common substitutions, and browsing the functions of given residues in conserved domains.

[Assembly Archive](#)
Links the raw sequence information found in the Trace Archive with assembly information found in publicly available sequence repositories (GenBank/EMBL/DDBJ). The Assembly Viewer allows a user to see the multiple sequence alignments as well as the actual sequence chromatogram.

[BLAST Link \(BLink\)](#)
A link option on protein records that displays the results of a pre-computed BLAST search of that protein against all other

Práce s databází NCBI

bené položky Nástroje Nápověda

Identity Safe ▾

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

aligned sequence reads) that have been produced by the [1000 Genomes Project](#).

[ASN.1 Format Summary](#)

An International Standards Organization (ISO) data representation format used to achieve interoperability between platforms. For data specifications and conversion tools, see NCBI Data Specification below.

[Amino Acid Explorer](#)

This tool allows users to explore the characteristics of amino acids by comparing their structural and chemical properties, predicting protein sequence changes caused by mutations, viewing common substitutions, and browsing the functions of given residues in conserved domains.

[Assembly Archive](#)

Links the raw sequence information found in the Trace Archive with assembly information found in publicly available sequence repositories (GenBank/EMBL/DDBJ). The Assembly Viewer allows a user to see the multiple sequence alignments as well as the actual sequence chromatogram.

[BLAST Link \(BLink\)](#)

A link option on protein records that displays the results of a pre-computed BLAST search of that protein against all other protein sequences at NCBI.

[BLAST Microbial Genomes](#)

Performs a BLAST search for similar sequences from selected complete eukaryotic and prokaryotic genomes.

[BLAST RefSeqGene](#)

Performs a BLAST search of the genomic sequences in the [RefSeqGene](#)/LRG set. The default display provides ready navigation to review alignments in the Graphics display.

[BLAST Tutorials and Guides](#)

This page links to a number of BLAST-related tutorials and guides, including a selection guide for BLAST algorithms, descriptions of BLAST output formats, explanations of the parameters for stand-alone BLAST, directions for setting up stand-alone BLAST on local machines and using the BLAST URL API.

Práce s databází NCBI

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/BLAST/blastn suite **BLAST microbial genomes**

blastn blastp blastx tblastn

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [+](#)

From

To

Or, upload file [Procházet...](#) [+](#)

Job Title

Enter a descriptive title for your BLAST search [+](#)

Choose Search Set

Database Complete genomes Draft genomes [+](#) Genomes: 2096

Organism [+](#) Exclude [+](#)

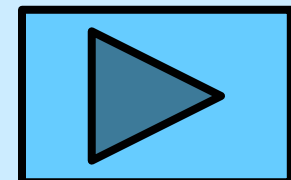
Optional Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [+](#)

Entrez Query

Optional Enter an Entrez query to limit search [+](#)

Program Selection

Dostali jste se na prohlédávač BLAST



Další zajímavé „Tools“

Vyhledávání STS

This interactive tool allows users to build E-utility URLs, either from a form or by hand, and then view their raw output. The tool provides a simple environment for testing E-utility URLs before including them in applications.

[E-Utilities](#)

Tools that provide access to data within NCBI's Entrez system outside of the regular web query interface. They provide a method of automating Entrez tasks within software applications. Each utility performs a specialized retrieval task, and can be used simply by writing a specially formatted URL.

[Ebot](#)

A tool that allows users to construct an E-utility analysis pipeline using an online form, and then generates a Perl script to execute the pipeline.

[Electronic PCR \(e-PCR\)](#)

A computational procedure that is used to identify sequence tagged sites (STSs) within DNA sequences. e-PCR looks for potential STSs in DNA sequences by searching for subsequences that closely match the PCR primers and have the correct order, orientation, and spacing that could represent the PCR primers used to generate known STSs.

[Frequency-weighted Link \(FLink\)](#)

FLink is a tool that enables you to link from a group of records in a source database to a ranked list of associated records in a destination database based on frequency-weighted statistics.

[Gene Expression Omnibus \(GEO\) BLAST](#)

Tool for aligning a query sequence (nucleotide or protein) to GenBank sequences included on microarray or SAGE platforms in the GEO database.

[Gene Plot](#)

A tool for pairwise comparison of two prokaryotic genomes that displays pairs of protein homologs that are symmetrical best hits between the two genomes.

[Genetic Codes](#)

Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree.

[Genome BLAST](#)

Další zajímavé „Tools“

Srovnání dvou prokaryotických genomů

This interactive tool allows users to build E-utility URLs, either from a form or by hand, and then view their raw output. The tool provides a simple environment for testing E-utility URLs before including them in applications.

[E-Utilities](#)

Tools that provide access to data within NCBI's Entrez system outside of the regular web query interface. They provide a method of automating Entrez tasks within software applications. Each utility performs a specialized retrieval task, and can be used simply by writing a specially formatted URL.

[Ebot](#)

A tool that allows users to construct an E-utility analysis pipeline using an online form, and then generates a Perl script to execute the pipeline.

[Electronic PCR \(e-PCR\)](#)

A computational procedure that is used to identify sequence tagged sites (STSs) within DNA sequences. e-PCR looks for potential STSs in DNA sequences by searching for subsequences that closely match the PCR primers and have the correct order, orientation, and spacing that could represent the PCR primers used to generate known STSs.

[Frequency-weighted Link \(FLink\)](#)

FLink is a tool that enables you to link from a group of records in a source database to a ranked list of associated records in a destination database based on frequency-weighted statistics.

[Gene Expression Omnibus \(GEO\) BLAST](#)

Tool for aligning a query sequence (nucleotide or protein) to GenBank sequences included on microarray or SAGE platforms in the GEO database.

[Gene Plot](#)

A tool for pairwise comparison of two prokaryotic genomes that displays pairs of protein homologs that are symmetrical best hits between the two genomes.

[Genetic Codes](#)

Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree.

[Genome BLAST](#)

Další zajímavé „Tools“

Tabulky genetických kódů

This interactive tool allows users to build E-utility URLs, either from a form or by hand, and then view their raw output. The tool provides a simple environment for testing E-utility URLs before including them in applications.

[E-Utilities](#)

Tools that provide access to data within NCBI's Entrez system outside of the regular web query interface. They provide a method of automating Entrez tasks within software applications. Each utility performs a specialized retrieval task, and can be used simply by writing a specially formatted URL.

[Ebot](#)

A tool that allows users to construct an E-utility analysis pipeline using an online form, and then generates a Perl script to execute the pipeline.

[Electronic PCR \(e-PCR\)](#)

A computational procedure that is used to identify sequence tagged sites (STSs) within DNA sequences. e-PCR looks for potential STSs in DNA sequences by searching for subsequences that closely match the PCR primers and have the correct order, orientation, and spacing that could represent the PCR primers used to generate known STSs.

[Frequency-weighted Link \(FLink\)](#)

FLink is a tool that enables you to link from a group of records in a source database to a ranked list of associated records in a destination database based on frequency-weighted statistics.

[Gene Expression Omnibus \(GEO\) BLAST](#)

Tool for aligning a query sequence (nucleotide or protein) to GenBank sequences included on microarray or SAGE platforms in the GEO database.

[Gene Plot](#)

A tool for pairwise comparison of two prokaryotic genomes that displays pairs of protein homologs that are symmetrical best hits between the two genomes.

[Genetic Codes](#)

Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree.

[Genome BLAST](#)

Další zajímavé „Tools“

Navrhování primerů pro PCR

[PSSM Viewer](#)

Allows users to display, sort, subset and download position-specific score matrices (PSSMs) either from CDD records or from Position Specific Iterated (PSI)-BLAST protein searches. The tool also can align a query protein to the PSSM and highlight positions of high conservation.

[Phenotype-Genotype Integrator \(PheGenI\)](#)

Supports finding human phenotype/genotype relationships with queries by phenotype, chromosome location, gene, and SNP identifiers. Currently includes information from dbGaP, the NHGRI GWAS Catalog, and GTeX. Displays results on the genome, on sequence, or in tables for download.

[Primer-BLAST](#)

The Primer-BLAST tool uses Primer3 to design PCR primers to a sequence template. The potential products are then automatically analyzed with a BLAST search against user specified databases, to check the specificity to the target intended.

[ProSplign](#)

A utility for computing alignment of proteins to genomic nucleotide sequence. It is based on a variation of the Needleman Wunsch global alignment algorithm and specifically accounts for introns and splice signals. Due to this algorithm, ProSplign is accurate in determining splice sites and tolerant to sequencing errors.

[PubChem Power User Gateway \(PUG\)](#)

PUG provides access to PubChem services via a programmatic interface. PUG allows users to download data, initiate chemical structure searches, standardize chemical structures and interact with the E-utilities. PUG can be accessed using either standard URLs or via SOAP.

[PubChem Standardization Service](#)

Standardization, in PubChem terminology, is the processing of chemical structures in the same way used to create PubChem Compound records from contributors' original structures. This service lets users see how PubChem would handle any structure they would like to submit.

[PubChem Structure Search](#)

PubChem Structure Search allows the PubChem Compound Database to be queried by chemical structure or chemical structure pattern. The PubChem Sketcher allows a query to be drawn manually. Users may also specify the structural query

Primer-BLAST

Primer-BLAST *A tool for finding specific primers*

► **NCBI/ Primer-BLAST:** Finding primers specific to your PCR template (using Primer3 and BLAST). [more...](#) [Tips for finding specific primers](#)

PCR Template [Reset page](#) [Save search parameters](#) [Retrieve recent results](#)

Enter accession, gi, or FASTA sequence (A refseq record is preferred) [Clear](#)

Range

Forward primer	<input type="text"/>	<input type="text"/>	Clear
Reverse primer	<input type="text"/>	<input type="text"/>	

Or, upload FASTA file

Primer Parameters

Use my own forward primer (5'->3' on plus strand) [Clear](#)

Use my own reverse primer (5'->3' on minus strand) [Clear](#)

PCR product size	Min <input type="text" value="70"/>	Max <input type="text" value="1000"/>		
# of primers to return	<input type="text" value="5"/>			
Primer melting temperatures (T _m)	Min <input type="text" value="57.0"/>	Opt <input type="text" value="60.0"/>	Max <input type="text" value="63.0"/>	Max T _m difference <input type="text" value="3"/> Clear

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section [Clear](#)

Exon junction span [Clear](#)

Exon junction match

Exon at 5' side Exon at 3' side

Prohlédněme si tuto stránku podrobně

Navrhněte primery pro identifikaci genu pro 16S rRNA *Borrelia burgdorferi* metodou PCR



- **Do zadávacího okénka pro sekvenci zadejte Acc. No. sekvence pro 16S rRNA, např. HQ433693.1**
- **Využijte DEFAULT nastavení nebo měňte parametry podle vlastního uvážení**




Ukázka výsledku

Primer-BLAST *Primer-Blast results*

▶ **NCBI/Primer-BLAST** : results: Job id=JSID_01_366985_130.14.18.128_9002 [more...](#)

Input PCR template [HQ433693.1](#) Borrelia burgdorferi strain QSYSP3 16S ribosomal RNA gene, partial sequence
Range 1 - 481
Specificity of primers primers may **not** be specific to the input PCR template as targets were found in selected database:All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences) ...[help on specific primers](#)
Other reports ▶[Search Summary](#)

▼ **Summary of primer pairs**



▼ **Detailed primer reports**

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GCGAAAGCCTGACGGAGCGA	Plus	20	322	341	59.77	65.00	3.00	0.00





Ukázka výsledku

▼ Detailed primer reports

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GCGAAAGCCTGACGGAGCGA	Plus	20	322	341	59.77	65.00	3.00	0.00
Reverse primer	ATTACCGCGGCTGCTGGCAC	Minus	20	478	459	60.39	65.00	6.00	2.00

Product length 157

Products on intended target

>[HQ433693.1](#) Borrelia burgdorferi strain QSYSP3 16S ribosomal RNA gene, partial sequence

product length = 157

```
Forward primer 1 GCGAAAGCCTGACGGAGCGA 20
Template 322 ..... 341

Reverse primer 1 ATTACCGCGGCTGCTGGCAC 20
Template 478 ..... 459
```

Products on potentially unintended templates

>[EU135595.1](#) Borrelia valaisiana strain QSYSP3 16S ribosomal RNA gene, partial sequence

product length = 157

```
Forward primer 1 GCGAAAGCCTGACGGAGCGA 20
Template 350 ..... 369
```



Vyhledejte sekvenci HQ433693.1 (16S rRNA *Borrelia burgdorferi*) a vyznačte na ní pozici nalezených primerů



- 1) Do vyhledávače BLAST zadejte „*Borrelia burgdorferi* 16S“
- 2) Najděte sekvenci HQ433693.1
- 3) Můžete do vyhledávače zadat taky přímo Acc. No.



Výsledek

AGCATGCAAGTCAAACGGGATGTAGCAATACATCTAGTGGCGAAC
GGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACT
ATTAGAAATAGTAGCTAATACCGAATAAAGTCAATTAATTTGTTA
ATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTG
CGTCTTATTAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATG
ATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAAGTGAAG
ACACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCG
CAATGG**GCGAAAGCCTGACGGAGCGA**CACTGCGTGAATGAAGAAG
GTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTG
TAGGAAATGACAAAGTGATGACGTTAATTTATGAATAAGCCCCGG
CTAATTAC**GTGCCAGCAGCCGCGGTAAT**ACG

Forward 322-341

5' - **GCGAAAGCCTGACGGAGCGA** - 3'

Reverse 478-459

5' - **ATTACCGCGGCTGCTGGCAC** - 3'

Další zajímavé „Tools“

Taxonomie

A utility for computing cDNA-to-Genomic sequence alignments. It is based on a variation of the Needleman-Wunsch global alignment algorithm and specifically accounts for introns and splice signals. Due to this algorithm, Splign is accurate in determining splice sites and tolerant to sequencing errors.

[TaxPlot](#)

A tool for comparing genomes on the basis of the protein sequences they encode. To use TaxPlot, one selects a reference genome and two species for comparison. Pre-computed BLAST results are then used to plot a point for each predicted protein in the reference genome, based on the best alignment with proteins in each of the two genomes being compared.

[Taxonomy Browser](#)

Supports searching the taxonomy tree using partial taxonomic names, common names, wild cards and phonetically similar names. For each taxonomic node, the tool provides links to all data in Entrez for that node, displays the lineage, and provides links to external sites related to the node.

[Taxonomy Common Tree](#)

Generates a taxonomic tree for a selected group of organisms. Users can upload a file of taxonomy IDs or names, or they can enter names or IDs directly.

[Taxonomy Statistics](#)

Displays the number of taxonomic nodes in the database for a given rank and date of inclusion.

[Taxonomy Status Reports](#)

Displays the current status of a set of taxonomic nodes or IDs.

[Variation Reporter](#)

A tool designed to search for and report human sequence variation data from [dbSNP](#) and [dbVar](#). Individual variations or batch files can be submitted in HGVS, GVF or BED formats. Related information will be retrieved and reported in a downloadable table containing variation identifiers, nucleotide and cytogenetic band locations on various genomic assemblies, allele type and minor allele frequencies, predicted functional consequences (missense, nonsense, frameshift, splice site, etc.), reported clinical significance, and relevant citations.

[VecScreen](#)

A system for quickly identifying segments of a nucleic acid sequence that may be of vector origin. VecScreen searches a

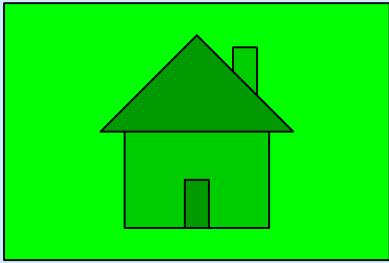
Kolik záznamů o sekvencích DNA a kolik záznamů o sekvencích proteinů je v databázi ohledně druhu *Thermus aquaticus*?



Ke konci června 2012 to bylo 338 záznamů o DNA a 562 (5 641) záznamů o proteinech



Práce s databází NCBI



www.ncbi.nlm.nih.gov

A screenshot of the NCBI website. The top navigation bar includes the NCBI logo, "Resources", and "How To". The main header features the NCBI logo, "National Center for Biotechnology Information", a search bar with a dropdown menu set to "All Databases", and a "Search" button. On the left is a vertical "Resource List (A-Z)" menu. The main content area is titled "Welcome to NCBI" and includes a "Get Started" section with a red arrow pointing to a list of links: Tools, Downloads, How-To's, and Submissions. Below this is a "Genomic Structural Variation" banner with a corn image and a video player. On the right is a "Popular Resources" list and an "NCBI Announcements" section with a date of 12 Jun 2012.

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

1 2 3 4 5 6 7 8

Popular Resources

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

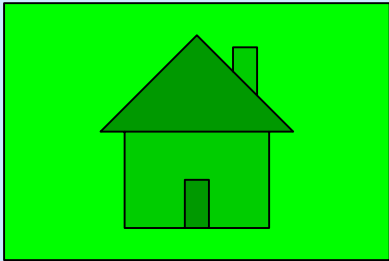
NCBI Announcements

New Microbial BLAST Page 12 Jun 2012

Now easier to use and with the familiar format and features of the standard NCBI BLAST services, including auto-complete

Sign up for the Fall Discovery Workshops!

Práce s databází NCBI



www.ncbi.nlm.nih.gov

A screenshot of the NCBI website homepage. The page features a navigation menu on the left, a search bar at the top, and several content sections. A red box highlights the 'Get Started' section, and a red arrow points from the 'Genetics & Medicine' menu item to it. The 'Genetic Structural Variation' section includes a video player with a progress bar.

NCBI Resources ▾ How To ▾ My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases ▾ Search

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

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

1 2 3 4 5 6 7 8

Popular Resources

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

NCBI Announcements

New Microbial BLAST Page 12 Jun 2012

Now easier to use and with the familiar format and features of the standard NCBI BLAST services, including auto-complete

Sign up for the Fall Discovery Workshops!

Jak s nástroji pracovat

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

All Resources

All Databases Downloads Submissions Tools How To

How To

- [Find bioassays in which a given drug is active](#)
- [Find bioassays that test a particular disease or protein target](#)
- [Submit data to NCBI](#)
- [Save text searches and set up automated searches with E-mail](#)
- [Download NCBI Software](#)
- [Retrieve all sequences for an organism or taxon](#)
- [Find the function of a gene or gene product](#)
- [Find expression patterns](#)
- [Find genes associated with a phenotype or disease](#)
- [Compare protein homologs between two microbial genomes](#)
- [View/download features around an object or between two objects on a chromosome](#)
- [Find sequenced genomes, including those in progress, for a taxonomic group](#)
- [Download the complete genome for an organism](#)
- [Display genomic annotation graphically](#)
- [Submit sequence data to NCBI](#)
- [Convert feature coordinates between genomic assemblies](#)
- [Determine conserved synteny between the genomes of two organisms](#)
- [Find a homolog for a gene in another organism](#)
- [Obtain the full text of an article](#)

uvidíme později

Porovnání proteinů u dvou genomů

The screenshot shows the NCBI website interface. At the top, there is a navigation bar with "NCBI Resources" and "How To" dropdown menus, and "My NCBI Sign In" on the right. Below this is a search bar with a dropdown menu set to "All Databases" and a "Search" button. On the left side, there is a vertical navigation menu with various categories like "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", and "Variation".

The main content area is titled "How to: Compare protein homologs between two microbial genomes". Below the title, a red-bordered box contains the text: "Starting with the [Prokaryotic Genome Project](#) homepage...".

FOR TWO ORGANISMS

1. Scroll down to find the genome of interest.
2. Click the NC_ accession link from the RefSeq column.
3. Click GenePlot (if available) from the BLAST homologs column of the resulting table interface.
4. Select the two organisms of choice and then click "Compare Selected Pair".

FOR THREE ORGANISMS

1. Proceed as in Steps 1 and 2 above.
2. Select TaxPlot from the BLAST homologs column of the resulting table interface.
3. Select two other organisms from the drop-down menus below the selected genome of interest.
4. Click the "compare" button located just below the graphical plot.

Návod

FOR TWO ORGANISMS

- 1) Scroll down to find the genome of interest.
- 2) Click the NC_ accession link from the RefSeq column.
- 3) Click GenePlot (if available) from the BLAST homologs column of the resulting table interface.
- 4) Select the two organisms of choice and then click "Compare Selected Pair".

FOR THREE ORGANISMS

- 1) Proceed as in Steps 1 and 2 above.
- 2) Select TaxPlot from the BLAST homologs column of the resulting table interface.
- 3) Select two other organisms from the drop-down menus below the selected genome of interest.
- 4) Click the "compare" button located just below the graphical plot.

Jak s nástroji pracovat

- [Download the complete genome for an organism](#)
- [Display genomic annotation graphically](#)
- [Submit sequence data to NCBI](#)
- [Convert feature coordinates between genomic assemblies](#)
- [Determine conserved synteny between the genomes of two organisms](#)
- [Find a homolog for a gene in another organism](#)
- [Obtain the full text of an article](#)
- [Find articles about a topic similar to that in a given article](#)
- [View the 3D structure of a protein](#)
- [Find a curated version of a sequence record \(NCBI Reference Sequence\)](#)
- [Align two or more 3D structures to a given structure](#)
- [Find published information on a gene or sequence](#)
- [Find transcript sequences for a gene](#)
- [Link from an object on a map to another resource](#)
- [Design PCR primers and check them for specificity](#)
- [Automate BLAST searches performed on NCBI servers](#)
- [Obtain genomic sequence for/near a gene, marker, transcript or protein](#)
- [Compare your sequence to the RefSeqGene/LRG standard](#)
- [Run BLAST software on a local computer](#)
- [Submit multiple query sequences in a single BLAST search](#)
- [Find the complete taxonomic lineage for an organism](#)
- [Generate a Common Tree for a set of taxa](#)
- [Complete an NCBI tutorial](#)
- [Find out what's new at NCBI](#)
- [Learn about an NCBI resource](#)
- [Learn about the basics of molecular biology and bioinformatics](#)
- [Download a large, custom set of records from NCBI](#)
- [Find human variations associated with a phenotype or disease \(clinical association\)](#)
- [View a mutation site in a 3D structure](#)
- [View all SNPs associated with a gene](#)
- [View genotype frequency data for a gene, disease or short genetic variation](#)

Databáze PubMed

The screenshot shows the NCBI website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' dropdown menus, and 'My NCBI Sign In' on the right. Below this is a search bar with a dropdown menu set to 'All Databases' and a 'Search' button. On the left side, there is a vertical menu with various categories: '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', and 'Variation'. A red arrow points from the 'Chemicals & Bioassays' menu item to the main content area. The main content area has a heading 'How to: Obtain the full text of an article' and a sub-heading 'Starting with an abstract in PubMed...'. It contains a list of instructions and a note about a YouTube tutorial.

NCBI Resources How To My NCBI Sign In

NCBI
National Center for
Biotechnology Information

All Databases Search

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

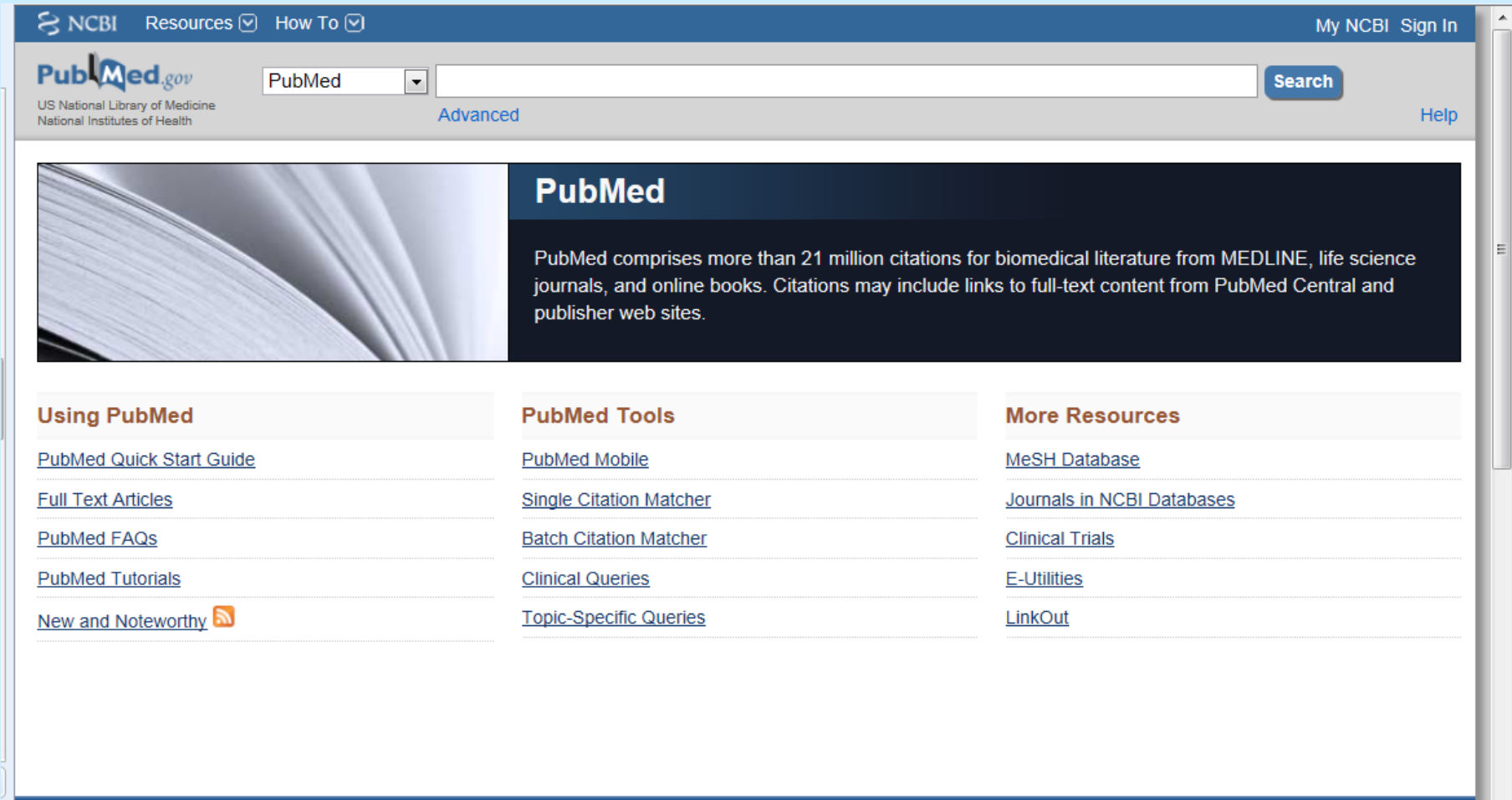
How to: Obtain the full text of an article

Please note that there is a [YouTube tutorial](#) about this.

Starting with an abstract in PubMed...

1. Search the [PubMed](#) with a search term, author name, or PubMed ID. Author name can be entered as follows: smith aj[au].
2. Click on the title of an entry of interest.
3. Look for icons in the upper-right-hand corner of the record:
 - Click on the [PubMed Central](#) link or a Publisher's link to access the full text of the article. Articles in [PubMed Central](#) are freely available. Articles on Publisher's websites are either freely available or can be accessed with a fee. Contact the specific publisher for questions about their site.
 - For PubMed records with no icons in the upper-right-hand corner, Loansome Doc can be accessed to order the article following these directions: [PubMed Help](#).

Databáze PubMed



The screenshot shows the PubMed website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links, and 'My NCBI Sign In' on the right. Below this is the PubMed logo and a search bar with a dropdown menu set to 'PubMed' and a 'Search' button. The main content area features a large image of a book and a dark blue box with the text: 'PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.' Below this are three columns of links: 'Using PubMed' (Quick Start Guide, Full Text Articles, FAQs, Tutorials, New and Noteworthy), 'PubMed Tools' (Mobile, Single Citation Matcher, Batch Citation Matcher, Clinical Queries, Topic-Specific Queries), and 'More Resources' (MeSH Database, Journals in NCBI Databases, Clinical Trials, E-Utilities, LinkOut).

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed

Advanced Help

PubMed

PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- [PubMed Quick Start Guide](#)
- [Full Text Articles](#)
- [PubMed FAQs](#)
- [PubMed Tutorials](#)
- [New and Noteworthy](#)

PubMed Tools

- [PubMed Mobile](#)
- [Single Citation Matcher](#)
- [Batch Citation Matcher](#)
- [Clinical Queries](#)
- [Topic-Specific Queries](#)

More Resources

- [MeSH Database](#)
- [Journals in NCBI Databases](#)
- [Clinical Trials](#)
- [E-Utilities](#)
- [LinkOut](#)

Najděte publikace o *Deinococcus radiodurans*

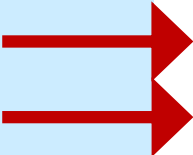
Kolik review databáze obsahuje?



- 1) Ke konci června 2012 jich bylo kolem 962**
- 2) Z toho review bylo 52**
- 3) Všimněte si, že jen některé jsou volně dostupné**



Jak s nástroji pracovat

- 
- [Download the complete genome for an organism](#)
 - [Display genomic annotation graphically](#)
 - [Submit sequence data to NCBI](#)
 - [Convert feature coordinates between genomic assemblies](#)
 - [Determine conserved synteny between the genomes of two organisms](#)
 - [Find a homolog for a gene in another organism](#)
 - [Obtain the full text of an article](#)
 - [Find articles about a topic similar to that in a given article](#)
 - [View the 3D structure of a protein](#)
 - [Find a curated version of a sequence record \(NCBI Reference Sequence\)](#)
 - [Align two or more 3D structures to a given structure](#)
 - [Find published information on a gene or sequence](#)
 - [Find transcript sequences for a gene](#)
 - [Link from an object on a map to another resource](#)
 - [Design PCR primers and check them for specificity](#)
 - [Automate BLAST searches performed on NCBI servers](#)
 - [Obtain genomic sequence for/near a gene, marker, transcript or protein](#)
 - [Compare your sequence to the RefSeqGene/LRG standard](#)
 - [Run BLAST software on a local computer](#)
 - [Submit multiple query sequences in a single BLAST search](#)
 - [Find the complete taxonomic lineage for an organism](#)
 - [Generate a Common Tree for a set of taxa](#)
 - [Complete an NCBI tutorial](#)
 - [Find out what's new at NCBI](#)
 - [Learn about an NCBI resource](#)
 - [Learn about the basics of molecular biology and bioinformatics](#)
 - [Download a large, custom set of records from NCBI](#)
 - [Find human variations associated with a phenotype or disease \(clinical association\)](#)
 - [View a mutation site in a 3D structure](#)
 - [View all SNPs associated with a gene](#)
 - [View genotype frequency data for a gene, disease or short genetic variation](#)

3D struktury proteinů

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

How to: View the 3D structure of a protein

Starting with...

A PDB CODE (e.g. 1B8G)

1. Go to the [Structure Home Page](#).
2. Enter the PDB code in the search box and press the Go button.
3. Click the structure image, and on the resulting page click the "Structure View in Cn3D" button.

A PDB-FORMAT FILE THAT IS NOT IN PDB

1. Go to the [VAST search page](#).
2. Enter or browse for the PDB file name and click the Submit button.
3. Click the "View 3D Structure" button on the next page.

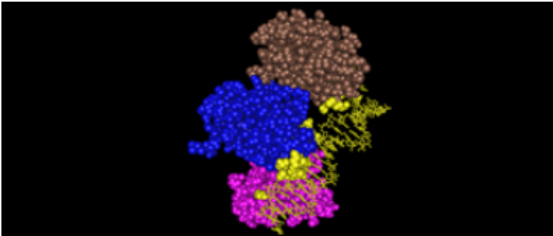
A PROTEIN ACCESSION NUMBER (e.g. NP_000240) OR SEQUENCE

1. Use the [Finding a Structural Template](#) guide to find the most appropriate PDB structure.
2. Continue with step 1 under "a PDB code" above.

3D struktury proteinů

NCBI Resources How To My NCBI Sign In

Structure Structure Search Limits Advanced Help



Structure

Three dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites, and more.

Using Structure

- [Search](#)
- [How to \(Quick Start\) Guides](#)
- [Help](#)
- [News](#)
- [FTP](#)
- [Publications](#)
- [Discover](#)

Structure Tools

- [Macromolecular Resources Overview](#)
- [CBLAST](#)
- [Cn3D](#)
- [IBIS](#)
- [VAST](#)

More Resources

- [PDB](#)
- [Protein](#)
- [CDD](#)
- [PubChem](#)
- [NCBI Structure Group Resources & Research](#)

Najděte strukturu mykobakteriální katalázy

Kolik záznamů najdete?



- 1) Heslo „catalase Mycobacterium“
- 2) Ke konci června 2012 jich bylo 46, všechny získané z krystalografických dat prostřednictvím paprsků X, žádná NMR



Jak s nástroji pracovat

- [Download the complete genome for an organism](#)
- [Display genomic annotation graphically](#)
- [Submit sequence data to NCBI](#)
- [Convert feature coordinates between genomic assemblies](#)
- [Determine conserved synteny between the genomes of two organisms](#)
- [Find a homolog for a gene in another organism](#)
- [Obtain the full text of an article](#)
- [Find articles about a topic similar to that in a given article](#)
- [View the 3D structure of a protein](#)
- [Find a curated version of a sequence record \(NCBI Reference Sequence\)](#)
- [Align two or more 3D structures to a given structure](#)
- [Find published information on a gene or sequence](#)
- [Find transcript sequences for a gene](#)
- [Link from an object on a map to another resource](#)
- [Design PCR primers and check them for specificity](#)
- [Automate BLAST searches performed on NCBI servers](#)
- [Obtain genomic sequence for/near a gene, marker, transcript or protein](#)
- [Compare your sequence to the RefSeqGene/LRG standard](#)
- [Run BLAST software on a local computer](#)
- [Submit multiple query sequences in a single BLAST search](#)
- [Find the complete taxonomic lineage for an organism](#)
- [Generate a Common Tree for a set of taxa](#)
- [Complete an NCBI tutorial](#)
- [Find out what's new at NCBI](#)
- [Learn about an NCBI resource](#)
- [Learn about the basics of molecular biology and bioinformatics](#)
- [Download a large, custom set of records from NCBI](#)
- [Find human variations associated with a phenotype or disease \(clinical association\)](#)
- [View a mutation site in a 3D structure](#)
- [View all SNPs associated with a gene](#)
- [View genotype frequency data for a gene, disease or short genetic variation](#)

Srovnání sekvence s referenčními

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

How to: Compare your sequence to the RefSeqGene/LRG standard

Starting with a sequence or sequences.

1. From the [RefSeqGene](#) homepage, click on [RefSeqGene BLAST](#) in the Tools section.
2. Submit your query sequence or multiple sequences.
3. Review the results as aligned to the RefSeqGene records by clicking on the Graphics in the Descriptions table.
4. If you submitted more than one query sequence and would like to review the alignment of a particular sequence, click on "Configure", select your chosen alignment and remove the check box in front of the alignments you don't want displayed. Then click on "Configure" at the bottom of the page to apply your revised selections.
5. If you identify any differences between your sequence and the RefSeqGene, you can evaluate whether others have reported sequence variation in that region by reviewing the variation annotated on the RefSeqGene.

Srovnání sekvence s referenčními

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI / BLAST / blastn suite **RefSeqGene Nucleotide BLAST**

blastn

Enter Query Sequence Search RefSeqGene using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

Or, upload file [Procházet...](#)

Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Reference genomic sequences (refseq_genomic) [+](#)

Organism Exclude [+](#)
Optional
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Models (XM/XP) Uncultured/environmental sample sequences
Optional

Entrez Query

Zkopírujte si níže uvedenou sekvenci a porovnejte ji s databází referenčních sekvencí. Komu patří?



- 1) ATGAGTGAAATGAAATGCCCTTATGACCATACCAACTTGACCATGAGTAATGGCGCGCCTGTTATTGACA
- 2) ACCAAAATTCAATGACCGCAGGTGCCAGAGGGCCACTGCTTGCCCAAGATTTATGGCTCAATGAAAAATT
- 3) AGCCGACTTTGCCCGTGAGGTCATTCCAGAACGCCGCATGCACGCCAAAGGCTCAGGGCGCATTGGCACA
- 4) TTCACGGTAACGCACGACATCACCCAATACACCCGTGCTAAGATTTTTAGTGAAGTTGGCAAAAAACTG
- 5) AGATGTTGCTCGTTTTACCACCGTAGCAGGGCAGCGGGGGGCGGGGACGCTGAGCGTGATATCCGTGG
- 6) TTTTGCCCTAAAATTCTACACCGAAGAGGGTAATTGGGACATGGTGGGTAATAACACGCCTGTTTTCTTT
- 7) TTAAGAGACCCAAAAAATTCCCTGATTTAAATAAAGCGGTCAAACGAGACCCACGCACCAACATGCGTT
- 8) CTGCCACCAATAACTGGGATTTTTGGACTGCTGCCAGAGGGCTTTCATCAGGTGACCATTGTGATGAG
- 9) CGACCGTGGCATTCTAAATCTTACCGTCATATGCACGGCTTTGGCTCGCACACTTATAGCTTTATCAAT
- 10) GCTGATAATGAACGCTTTTGGGTCAAATTTCACTTTCGCACCCAACAAGGCATTGAAAATCTAACCGATG
- 11) CCGAAGCTGAAATGGTGGTTGGTAAAGACCGTGAGAGCAATCAGCGTGATTTGTTTGATGCCATTGAGCG
- 12) TGGCGATTTCCAAAAATGGACAATGTATGTGCAAATCATGCCAGAAAACCGATGCCCAAATGTGCCTTAT
- 13) CACCCATTTGATTTAACCAAAGTGTGGCCAAAAGGCGACTATCCGCTCATTGAAGTGGGTGAGTTTGAGT
- 14) TAAATAAAAAATCCTGAAAATCTTTTTAGACGTTGAACAATCCGCTTTTGCCCAAGCAACCTAGTCCC
- 15) GGGCATCAGTGTGTCCCTGACCGCATGCTCCAAGCACGCTATTTAACTATGCTGATGCGCAGCGTTAT
- 16) CGTTTGGGCGTCAATCGTAACCAAATCCAGTGAATGCCCCACGCTGTCTGTGACTCAAACCAAAGAG
- 17) ACGGACAAGGGGCGAGTGGGCGATAACTATGGCGGTGCTCCGCACTATGAACCGAACAGTTTTGGACAATG
- 18) GCAAGACCAGCCGCATTTGGCTGAACCAGCATTAAAAATTCATGGCGATGCTAAGTTTTGGGATTATCGT
- 19) GAGAATGATGATGATTATTTAGCCAACCCAGAGCCTTGTGAGTTGATGAGCGATGAGCAAAAAACAGG
- 20) CGTTATTTGGTAATACGGCTCGTGCGATGGGCGATGCCCTGATTTATTAATAACCGCCATATCCGTAA
- 21) TTGCGATAAATGCCACCCTGATTATGCCATGGGTGTGGCCAAAGCGTTAGGCCTTACGGTTGAAGATGCC
- 22) AAAAATGCGTATGAGAGCGACCCTGCTCGCCATCTGCCAGCTTTTTATA

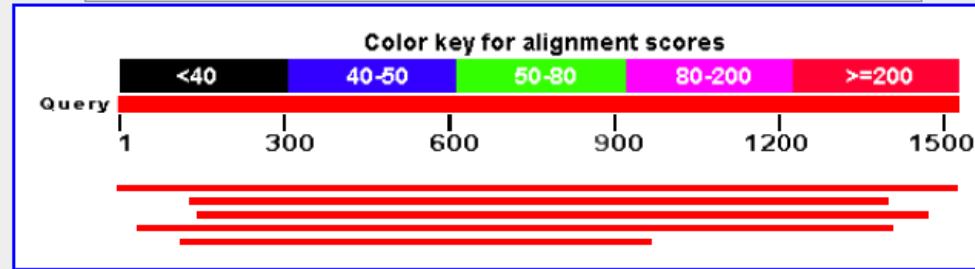
Mohlo by vám vyjít to, co je na následující stránce





Distribution of 5 BlastHits on the Query Sequence

Mouse over to see the define, click to show alignments



Descriptions

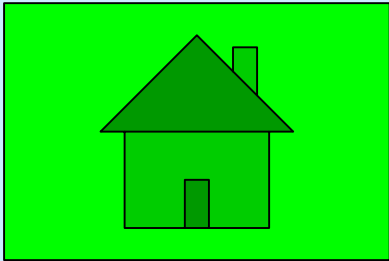
Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NC_014147.1	Moraxella catarrhalis RH4 chromosome, complete genome	2808	2808	100%	0.0	100%	
NC_015460.1	Gallibacterium anatis UMN179 chromosome, complete genome	763	763	83%	0.0	78%	
NC_009524.1	Psychrobacter sp. PRwf-1 chromosome, complete genome	695	695	87%	0.0	76%	
NC_014752.1	Neisseria lactamica 020-06 chromosome, complete genome	553	553	89%	7e-153	74%	
NC_010382.1	Lysinibacillus sphaericus C3-41 chromosome, complete genome	333	333	56%	1e-86	74%	



Práce s databází NCBI



www.ncbi.nlm.nih.gov

A screenshot of the NCBI website. The top navigation bar includes the NCBI logo, 'Resources', and 'How To'. A search bar is present with a dropdown menu set to 'All Databases'. The main content area is divided into a left sidebar with a 'Resource List (A-Z)' and a central 'Welcome to NCBI' section. A red rounded rectangle highlights the 'Get Started' section, which contains links for Tools, Downloads, How-To's, and Submissions. A red arrow points from the 'Genomes & Maps' link in the sidebar to the 'Get Started' box. Below this is a 'Genomic Structural Variation' banner with a corn cob image. On the right, there are sections for 'Popular Resources' and 'NCBI Announcements'.

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

All Databases Search

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

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

Popular Resources

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

NCBI Announcements

New Microbial BLAST Page 12 Jun 2012

Now easier to use and with the familiar format and features of the standard NCBI BLAST services, including auto-complete

Sign up for the Fall Discovery Workshops!

Pokyny pro vložení vlastních dat

NCBI Resources ▾ How To ▾ My NCBI Sign In

NCBI
National Center for
Biotechnology Information

All Databases ▾ Search

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

How to: Submit data to NCBI

Starting with...

SEQUENCE DATA

For guidance on the submission process for your sequence(s), please see [How To: Submit sequence data to NCBI](#).

Your data will be submitted to one of the following databases:

- [GenBank](#)
- [Sequence Read Archive \(SRA\)](#)
- [dbSNP](#)
- [dbVar](#)
- [GEO](#)

MICROARRAY DATA

If you have microarray data from clinical studies that require controlled access, you should submit your data to [dbGaP](#).

For all other microarray data, you should submit your data to [GEO](#) via [GEO's Submission page](#).

BIOASSAY DATA, SUBSTANCE OR SEQUENCE-BASED REAGENTS

BioAssay data and chemical substance information should be submitted to [PubChem](#) via their [PubChem Deposition Gateway](#).

Posuzování podobnosti sekvencí

Posuzování podobnosti sekvencí

Hledáme homologické sekvence vzniklé
v průběhu evoluce

Úkol: Jsou si podobnější sekvence A a B nebo B a C?

Výchozí sekvence

A = ATTGCTCTGT

B = ATAGCTCGGT

C = ATTGCACTGTAATGCCATGT

D = ATTGCTCTGAAATGCCCTGT

Posuzování podobnosti sekvencí

Přiložíme sekvence k sobě = přiřazení
(alignment)

A	=	A	T	T	G	C	T	C	T	G	T
B	=	A	T	A	G	C	T	C	G	G	T

pár nepár

C	=	A	T	T	G	C	A	C	T	G	T	A	A	T	G	C	C	A	T	G	T
D	=	A	T	T	G	C	T	C	T	G	A	A	A	T	G	C	C	C	T	G	T

Posuzování podobnosti sekvencí

Výpočet normalizované hodnoty podobnosti
(score)

A =	A	T	T	G	C	T	C	T	G	T
B =	A	T	A	G	C	T	C	G	G	T

hodnota páru

hodnota nepáru

$$S_{AB} = (8 \times 1 + 2 \times 0) / 10 = 0,80$$

počet pozic

počet párů
(match)

počet nepárů
(mismatch)

Posuzování podobnosti sekvencí

C = A T T G C A C T G T A A T G C C A T G T
| | | | | | | | | | | | | | | | | | |
D = A T T G C T C T G A A A T G C C C T G T

$$S_{CD} = (17 \times 1 + 3 \times 0)/20 = 0,85$$

0,85 > 0,80 → C a D jsou si podobnější

Globální a lokální přiřazení

Problém sekvencí odlišné délky nebo velmi odlišné sekvence stejné délky

Global alignment

- **Sekvence přiřadíme po celé délce i za cenu vnášení mezer**
- **Vhodné pouze u příbuzných sekvencí**
- **Vhodné pro mnohočetná přiřazení**

Local alignment

- **Sekvence přiřadíme jen tam, kde jsou velmi podobné, ostatní budeme ignorovat**
- **Vhodné pro nepříbuzné sekvence**
- **U podobných sekvencí odpovídá globálnímu přiřazení**

Globální a lokální přiřazení

Global alignment

SLAV-----APATNIK-----PIQNYR-I-----AKSETQRYMVIE
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVIE

Local alignment

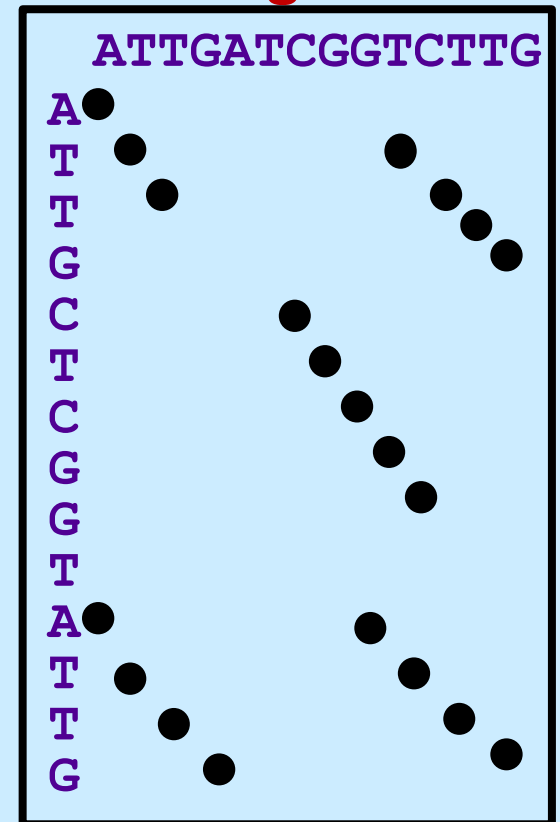
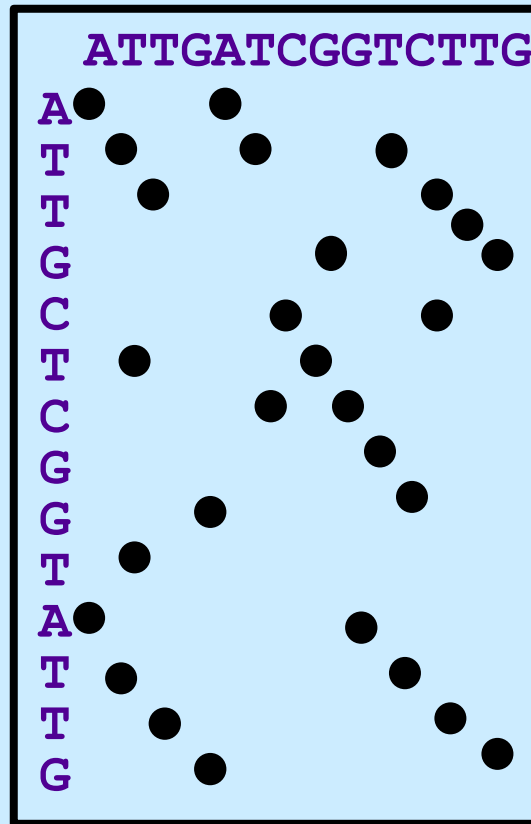
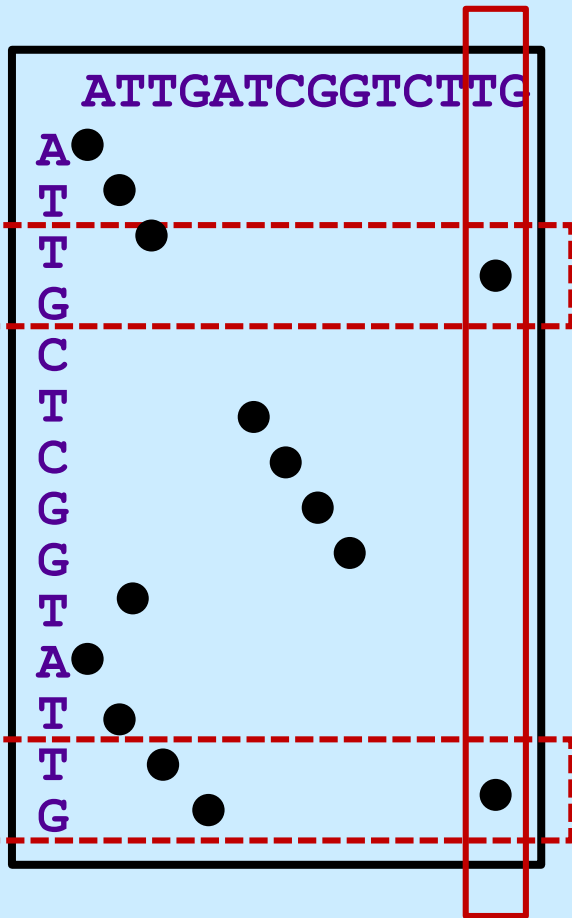
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVIE
-----NAPATNIKSECVRA-PIQNYRRVEHVRA-----

Bodový diagram

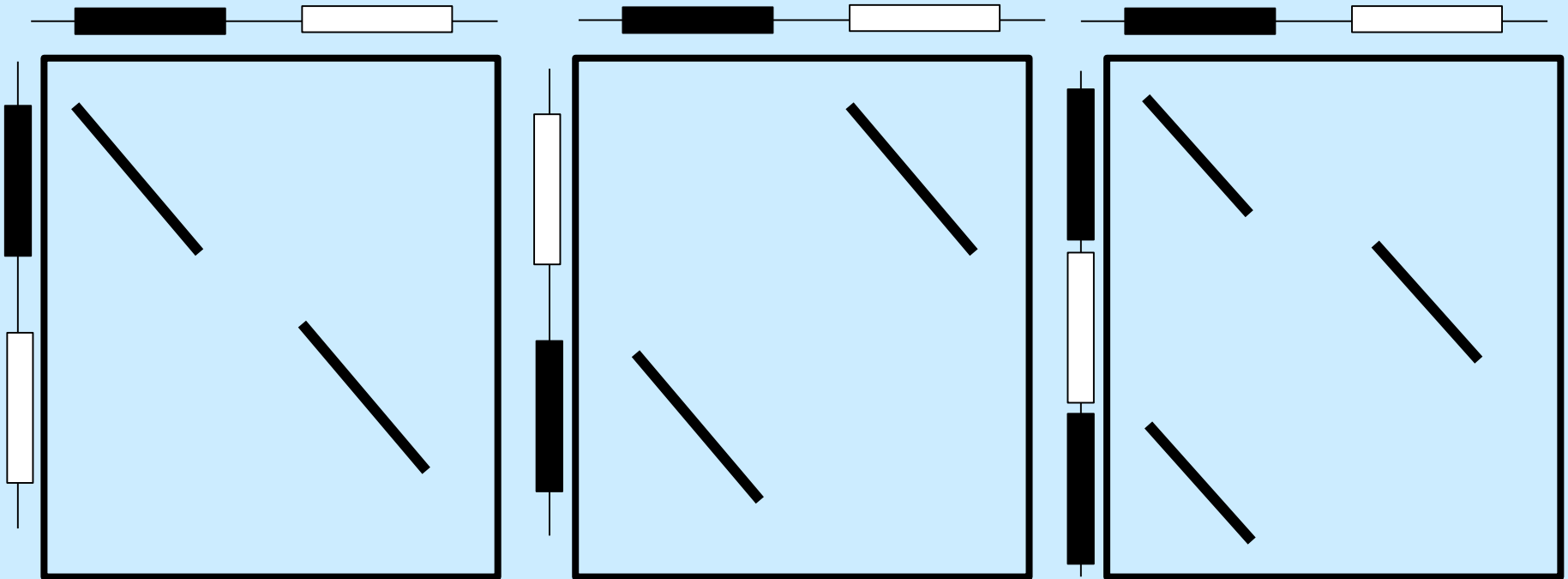
Grafická mapa podobností sekvencí, pomůcka pro volbu přiřazení

Filtrace krátkých diagonál

Nalezené shody



Výběr algoritmu přiřazení



Globální přiřazení je možné jen pro dvojici A-B

Prohledávače

FASTA

- **Modelový heuristický algoritmus**
- **Vytvořený v roce 1988**
- **Dnes už se málo používá, jsou výkonnější metody**

BLAST

- **Nejrozšířenější heuristický algoritmus**
- **Vytvořený v roce 1990**
- **Rychlejší než FASTA asi 6x**

BLAST

Basic Local Alignment Search Tool

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST website interface. At the top, there is a navigation bar with 'Home', 'Recent Results', 'Saved Strategies', and 'Help' buttons. A 'My NCBI' section includes 'Sign In' and 'Register' links. The main content area is titled 'NCBI/ BLAST Home' and contains a search box with the text 'BLAST finds regions of similarity between biological sequences. more...'. Below this is a 'New' banner for 'DELTA-BLAST, a more sensitive protein-protein search'. The 'BLAST Assembled RefSeq Genomes' section lists various species genomes for selection, including Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, and Microbes. The 'Basic BLAST' option is highlighted with a red box. Below this, there are options for 'nucleotide blast', 'protein blast', 'blastx', and 'tblastn', each with a brief description and the algorithms used. On the right side, there are sections for 'Your Recent Results', 'News', and 'Tip of the Day'.

Tento prohledávací nástroj prochází celou databází a už jsme jej několikrát použili

BLAST

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#)

Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontinuous megablast

[protein blast](#)

Search **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast, delta-blast

[blastx](#)

Search **protein** database using a **translated nucleotide** query

[tblastn](#)

Search **translated nucleotide** database using a **protein** query

[tblastx](#)

Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- ▣ Make specific primers with [Primer-BLAST](#)
- ▣ Search [trace archives](#)
- ▣ Find [conserved domains](#) in your sequence (cds)
- ▣ Find sequences with similar [conserved domain architecture](#) (cdart)
- ▣ Search sequences that have [gene expression profiles](#) (GEO)
- ▣ Search [immunoglobulins](#) (IgBLAST)
- ▣ Search using [SNP flanks](#)
- ▣ Screen sequence for [vector contamination](#) (vecscreen)
- ▣ [Align](#) two (or more) sequences using BLAST (bl2seq)
- ▣ Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay

available.

Mon, 04 Jun 2012 12:00:00 EST

[More BLAST news...](#)

Tip of the Day

[Use Genomic BLAST to see the genomic context](#)

If you are interested in the evolution of a particular gene or gene family it is often interesting to examine the intro-exon structure even across species.

[More tips...](#)



Využití variant BLAST

Program	Dotaz	Databáze	Úroveň srovnání	Použití
blastn	DNA	DNA	DNA	Hledání edentických sekvencí DNA
blastp	protein	protein	protein	Hledání homologických proteinů
blastx	DNA*	protein	protein	Hledání genů a homologických proteinů na nové DNA
tblastn	protein	DNA*	protein	Hledání genů u necharakterizovaných DNA
tblastx	DNA*	DNA*	protein	Studium struktury genů

*** Jsou srovnávány přeložené DNA sekvence ve všech čtecích rámcích**

Datové soubory

Jsou jednotné pro všechny zmíněné databáze

- **Každý záznam má přístupový kód – Accession Number** – proměnlivý počet písmen a číslic podle toho, přes kterou databázi byl přijat – je to jakési **rodné číslo**
- Publikací v GenBank získá **jedinečné číslo GI** (GenBank Identifier) – **číslo občanského průkazu**
- Autoři primárního záznamu jej mohou upravovat a vznikají tak **verze**, první má číslo 1
- Změnou verze se mění číslo GI
- Všechny verze se uchovávají

Hlavička záznamů

NCBI Resources How To

Nucleotide Nucleotide

přístupový kód **název** [Basic](#) [Advanced](#)

[Display Settings:](#) GenBank [Send to:](#)

Mycobacterium avium insertion element hot spot flanking region FR300

GenBank: AF369936.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS AF369936 312 bp DNA linear BCT 27-MAY-2001

DEFINITION Mycobacterium avium insertion element hot spot flanking region FR300.

ACCESSION AF369936

VERSION AF369936.1 GI:14210082

typ záznamu

verze

číslo GI

gb = GenBank, emb = EMBL, dbj = DDBJ

**Někdy sekvenuje daný úsek
nezávisle více různých skupin, pak
je v databázi v několika podobách
s různými přístupovými kódy a
často i pod různými názvy!**



Anatomie databázového záznamu

NCBI Resources How To My NCBI Sign In

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: GenBank Send to:

Mycobacterium avium insertion element hot spot flanking region FR300

GenBank: AF369936.1
[FASTA](#) [Graphics](#)

Go to:

LOCUS AF369936 312 bp DNA linear BCT 27-MAY-2001
DEFINITION Mycobacterium avium insertion element hot spot flanking region FR300.
ACCESSION AF369936
VERSION AF369936.1 GI:14210082
KEYWORDS .
SOURCE Mycobacterium avium
ORGANISM [Mycobacterium avium](#)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
avium complex (MAC).
REFERENCE 1 (bases 1 to 312)
AUTHORS Bartos,M., Svastova,P., Dvorska,L., Weston,R.T. and Pavlik,I.
TITLE Insertion element IS901 hot spot FR300
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 312)
AUTHORS Bartos,M., Svastova,P., Dvorska,L., Weston,R.T. and Pavlik,I.
TITLE Direct Submission
JOURNAL Submitted (13-APR-2001) Department of Bacteriology, Veterinary
Research Institute, Hudcova 70, Brno 621 32, Czech Republic
FEATURES Location/Qualifiers
source 1..312
/organism="Mycobacterium avium"
/mol_type="genomic DNA"
/db_xref="taxon:1764"
misc feature 1..312
/note="insertion element hot spot flanking region FR300;
contains hot spot for IS901 insertion"
ORIGIN
1 cagccagcgc aatgtcatcc cgaggtagag aagccagaac agccgaaga cgctccacgc
61 cgccacggcg cggcgccgga gcccgatgta gaggctgccc tgccgatcca cgcggttgat
121 ctgtctctcg atgctggcgg gcaagatcct cattggtggc ttcctttcgg tggggcgggc
181 ccggagtggc gccgtcgttg cgcacagtac aagcccggcc ggcggctacc gatcccaacc
241 acgtccggca cgcactacc tgcacggcag ggggctgtcg aaagggttcg ccggtgaacc
301 tgtcgcgagt tg
//

Change region shown
Customize view
Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence
Related information
Related Sequences
Taxonomy
Recent activity
Turn Off Clear
Mycobacterium avium insertion element hot spot flanking region FR300 Nucleotide
FR300 (2) Nucleotide
Neisseria gonorrhoeae strain PID2059 TraG3 (traG3), EppA (eppA), Ych1 (Nucleotide
Neisseria gonorrhoeae (22947) Nucleotide
Actinobacillus pleuropneumoniae in vivo induced promoter iViG; and Cps1B (Nucleotide
See more...

Anatomie databázového záznamu

Mycobacterium avium FR300

Neisseria gonorrhoeae

Program *bl2seq*

Porovnání dvou a více sekvencí

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search using [SNP flanks](#)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align](#) two (or more) sequences using BLAST (bl2seq)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- Search SRA [transcript and genomic libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)
- Needleman-Wunsch [Global Sequence Alignment Tool](#)
- Search [RefSeqGene](#)
- Search [WGS sequences](#) grouped by organism

BLAST is a registered trademark of the National Library of Medicine.

Program *bl2seq*

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI BLAST/ blastn suite **Align Sequences Nucleotide BLAST**

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide subjects using a nucleotide query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

From

To

Or, upload file Procházet...

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence Clear Subject subrange

From

To

Or, upload file Procházet...

Program Selection

Optimize for

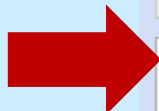
- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

Show results in a new window

+ Algorithm parameters



Výsledek porovnání dvou sekvencí

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign in] [Registered]

NCBI/ BLAST/ blastn suite-2sequences/ Formatting Results - YZXRUW9V11R

Edit and Resubmit Save Search Strategies Formatting options Download

Blast 2 sequences

Nucleotide Sequence (774 letters)

Query ID	lcl 31915	Subject ID	31917
Description	None	Description	None
Molecule type	nucleic acid	Molecule type	nucleic acid
Query Length	774	Subject Length	689
		Program	BLASTN 2.2.26+ Citation

Other reports: [Search Summary](#) [Taxonomy reports](#)

Graphic Summary

Distribution of 2 Blast Hits on the Query Sequence

Mouse-over to show defline and scores, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Query

1 150 300 450 600 750

Dot Matrix View

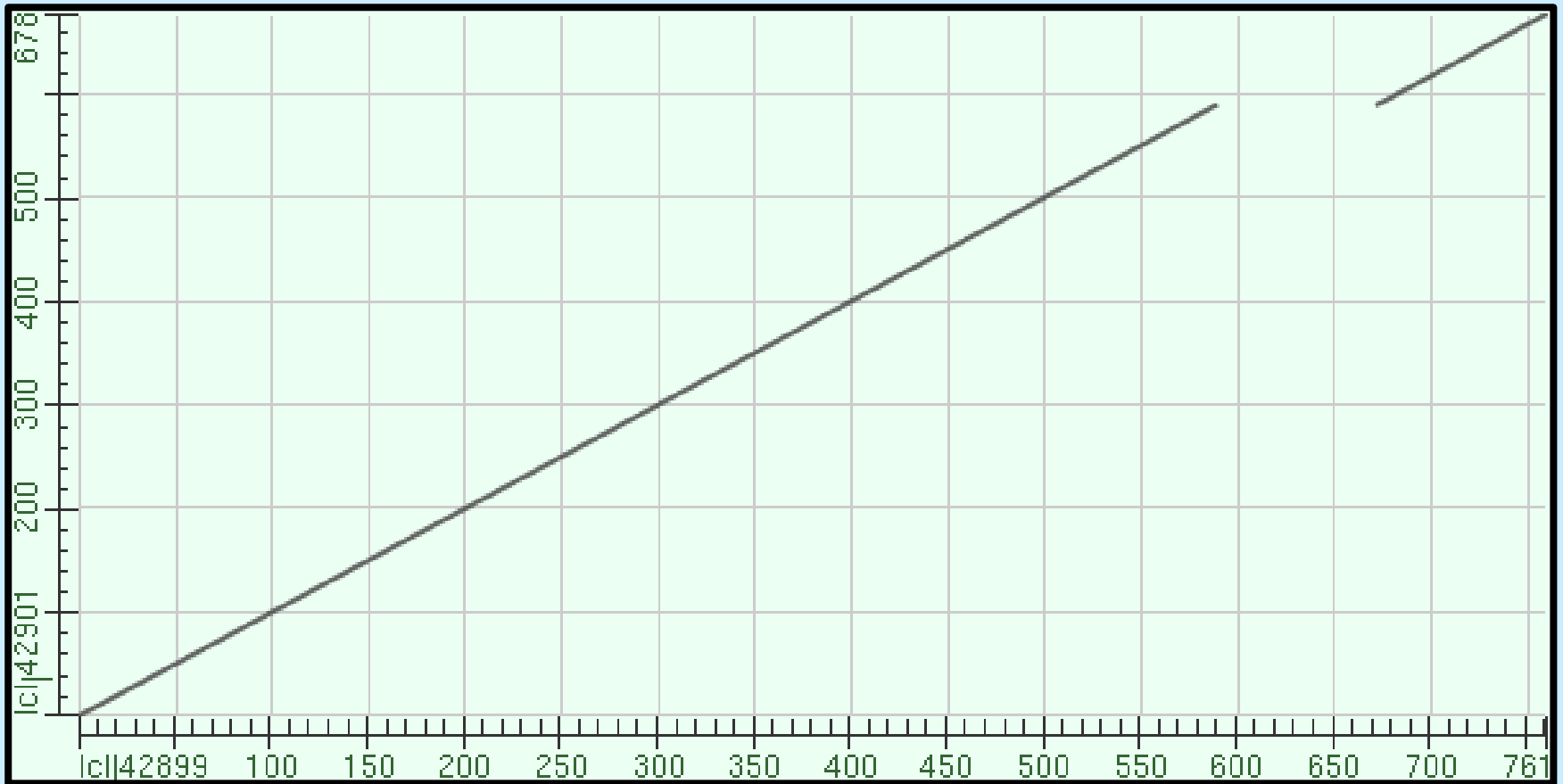
Descriptions

Legend for link resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Dot Matrix View

Plot of |c|42899 vs 42901



Výsledek porovnání dvou sekvencí

Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
31917		1057	1222	87%	0.0	100%	

Alignments

```
>lcl|31917
Length=689
```

Sort alignments for this subject sequence by:
E value [Score](#) [Percent identity](#)
[Query start position](#) [Subject start position](#)

Score = 1057 bits (572), Expect = 0.0
Identities = 584/590 (99%), Gaps = 0/590 (0%)
Strand=Plus/Plus

```
Query 1  GCTTTCGCTGTTGAGCGTCAGTACACGCAAGGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60
           |||
Sbjct 1  GCTTTCGCACATGAGCGTCAGTACACGCAAGGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60

Query 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAAATTCACCGGTCCCTAAAGTACTCTAG 120
           |||
Sbjct 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAAATTCACCGGTCCCTAAAGTACTCTAG 120

Query 121 ACTCCAGTCTGAAATGCAGTTCGCCAAGTTAAGCTCGGGGATTCACATCTCACTTAAAA 180
```

Score = 165 bits (89), Expect = 8e-45
Identities = 89/89 (100%), Gaps = 0/89 (0%)
Strand=Plus/Plus

```
Query 673 TCAGCAAAGAAAGCAAGCTTTCTCTCTGCTACCGTTCGACTTGCATGTGTTAAGCTGCC 732
           |||
Sbjct 590 TCAGCAAAGAAAGCAAGCTTTCTCTCTGCTACCGTTCGACTTGCATGTGTTAAGCTGCC 649

Query 733 GCCAGCGTTCAATCTGAGCCAGGATCAAC 761
           |||
Sbjct 650 GCCAGCGTTCAATCTGAGCCAGGATCAAC 678
```

Identities = frakce totožných pozic

Výsledek porovnání dvou sekvencí

Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
31917		1057	1222	87%	0.0	100%	

Alignments

```
>lcl|31917
Length=689

Sort alignments for this subject sequence by:
E value  Score  Percent identity
Query start position  Subject start position

Score = 1057 bits (572), Expect = 0.0
Identities = 584/590 (99%), Gaps = 0/590 (0%)
Strand=Plus/Plus

Query 1   GCCTTCGCTGTTGAGCGTCAGTACAITGCCAAGGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60
          |||
Sbjct 1   GCCTTCGCACATGAGCGTCAGTACAITCCCAAGGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60

Query 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAATTCTACCGGTCCTAAAGTACTCTAG 120
          |||
Sbjct 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAATTCTACCGGTCCTAAAGTACTCTAG 120

Query 121 ACTCCCACTGTAAGTGCAGTTCCCAAGTTAAGCTCGGGGATTTACATCTCACTTAAAA 180
```

```
Score = 165 bits (89), Expect = 8e-45
Identities = 89/89 (100%), Gaps = 0/89 (0%)
Strand=Plus/Plus

Query 673 TCAGCAAAGAAAGCAAGCTTTCTTCTGCTACCGITCGACTTGCATGTGTTAAGCCTGCC 732
          |||
Sbjct 590 TCAGCAAAGAAAGCAAGCTTTCTTCTGCTACCGITCGACTTGCATGTGTTAAGCCTGCC 649

Query 733 GCCAGCGTTCAATCTGAGCCAGGATCAAC 761
          |||
Sbjct 650 GCCAGCGTTCAATCTGAGCCAGGATCAAC 678
```

Score (zjištěná hodnota podobnosti) = pokud dosáhne zvolené mezní hodnoty (**cutoff**) program přiřazení zaznamená jako HSP (**high scoring pair**), jinak je opustí

Výsledek porovnání dvou sekvencí

Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
31917		1057	1222	87%	0.0	100%	

Alignments

```
>lcl|31917
Length=689
```



Sort alignments for this subject sequence by:
[E value](#) [Score](#) [Percent identity](#)
[Query start position](#) [Subject start position](#)

```
Score = 1057 bits (572), Expect = 0.0
Identities = 584/590 (99%), Gaps = 0/590 (0%)
Strand=Plus/Plus
```

```
Query 1  GCITTCGCTGTTGAGCGTCAGTACAITGCCAAGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60
          |||
Sbjct 1  GCITTCGCACATGAGCGTCAGTACAITCCCAAGGGGCTGCCTTCGCCTTCGGTATTCTCTC 60

Query 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAAATTTACCCGGTCCCTAAAGTACTCTAG 120
          |||
Sbjct 61  CACATCTCTACGCATTTCCACCGCTACACGTGGAAATTTACCCCTCCCTAAAGTACTCTAG 120

Query 121 ACTCCAGTCTGAAATGCAATGCCAAGTTAAGCTCGGGGATTTACATCTCACTTAAAA 180
```



```
Score = 165 bits (89), Expect = 8e-45
Identities = 89/89 (100%), Gaps = 0/89 (0%)
Strand=Plus/Plus
```

```
Query 673  TCAGCAAAGAAAGCAAGCTTTCTTCTGCTACCGTTCGACTTGCATGTGTTAAGCTGCC 732
          |||
Sbjct 590  TCAGCAAAGAAAGCAAGCTTTCTTCTGCTACCGTTCGACTTGCATGTGTTAAGCTGCC 649

Query 733  GCCAGCGTTCAATCTGAGCCAGGATCAAC 761
          |||
Sbjct 650  GCCAGCGTTCAATCTGAGCCAGGATCAAC 678
```

Expectancy, E-value (hodnota očekávatelnosti) = $8e^{-45}$ = 8×10^{-45} , průkazné jsou hodnoty pod 0,001

Něco navíc k procvičení BLAST



Prohledejte databázi a zjistěte, jakému organismu patří následující sekvence

GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCCAGTTATTCCGATTAACGCT
CGCACCCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
AATTAACGTCAATGATGCTATCTATTTAACAACATCCCTTCCTCATTACCGAAAGA
ACTTTACAACCCGAAGGCCTTCTTCATTCACGCGGCATGGCTGCGTCAGGGTTCCC
CCCATTGCGCAATATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCTCA
GTTCCAGTGTGGCTGGTCATCCTCTCAGACCAGCTAGAGATCGCAGGCTTGGTAGG
CCTTTACCCACCAACTACCTAATCCCCTTGGGCTCATCTTATGGCAGGTGGCCC
TAAGGTCCCACCCTTTCCTCCTCAGAGAATACGCGGTATTAGCTGCAGTTTCCCAC
AGTTATCCCCCTCCATAAGCCAGATTCCCAAGCATTACTCACCCGTCCGCCACTCG
TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCC
TGCCGCCAGCGTTCAATCTGAGCCAGGATCAACNTCTTTCTCCAAA

Měla by to být Pasteurella multocida





Porovnejte tyto dvě sekvence, patří stejnému druhu?

GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCCAGTTATTCCGATTAACGCT
CGCACCCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
AATTAACGTCAATGATGCTATCTATTTAACAACATCCCTTCCTCATTACCGAAAGA
ACTTTACAACCCGAAGGCCTTCTTCATTCACGCGG

GCTTTCGCGCATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAAAGCAGTTCCCAAGTTAAGCTCGGGGATTTCACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCGCAGTTATTCCGATTAACGCT
CGCACCCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
AATTAACGTCAATGATGCTATCTATTTAACAACATCCCTTCCTCATTACCGAAAGA
ACTTTACAACCCGAAGGCCTTCTTCATTCACGCGG

ANO, shoda 368/371, 99%



Mnohočetné přiřazení

Multiple alignment

- Jedním z příkladů využití je porovnávání více sekvencí současně

CLUSTAL

- CLUSTAL W = všeobecně dostupný
- CLUSTAL X = CLUSTAL W opatřený grafickým rozhraním pro Windows
- CLUSTAL OMEGA = poslední verze

<http://www.clustal.org>

Shrnutí

- 1) Práce se sekvenčními daty**
- 2) Základní veřejně dostupné databáze**
- 3) Práce se stránkami NCBI**
- 4) Jak se posuzuje podobnost sekvencí**
- 5) Prohledávač BLAST, BLAST2**
- 6) Mnohočetné přiřazení – program CLUSTAL**