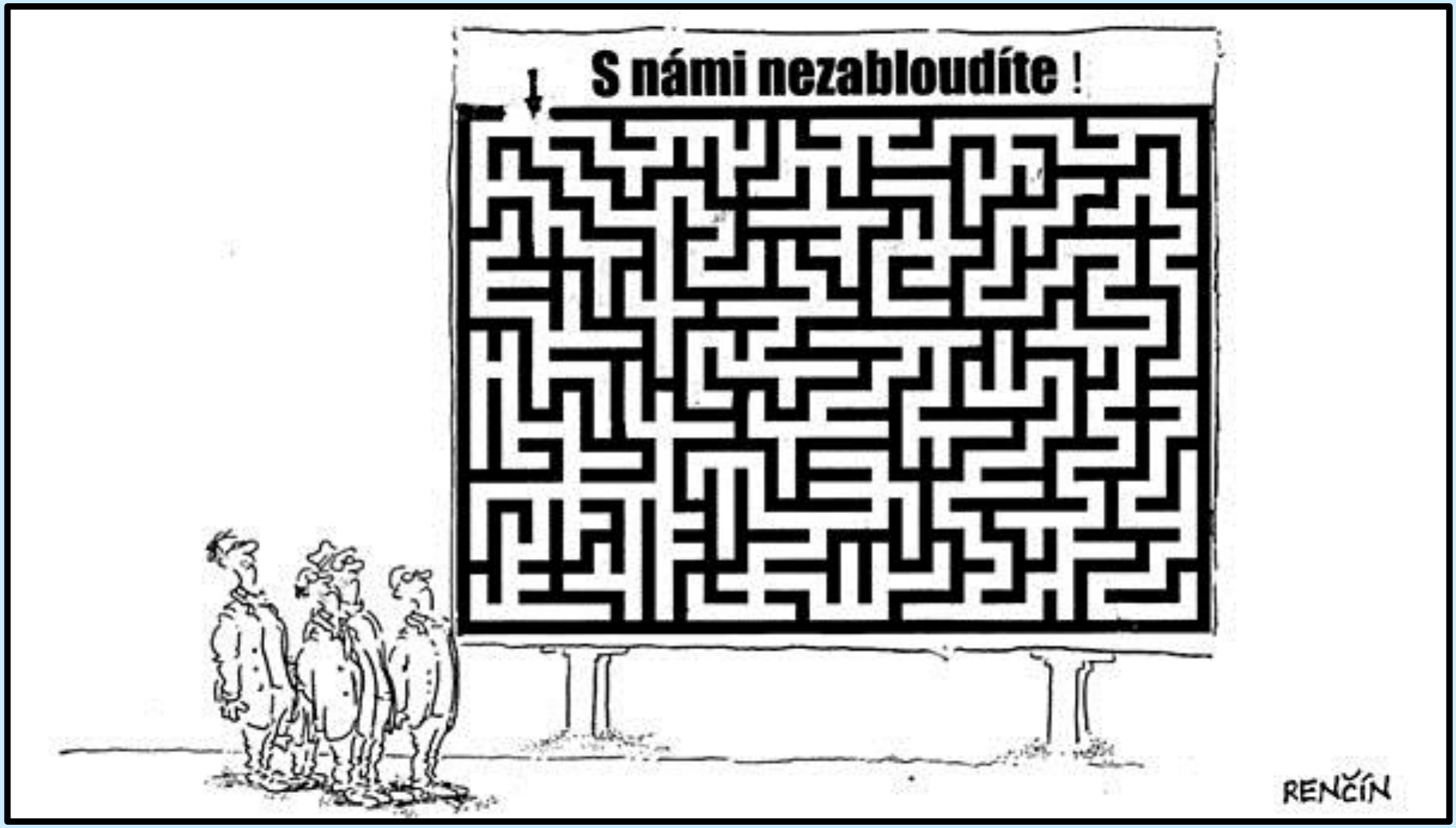

S námi nezabloudíte !
RENČÍN

# 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 (keto) |
| C | Cytosin | M | A, C (amino) |
| 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 (purin) | V | A, C, G (ne T, U) |
| Y | C, T (pyrimidin) | N | cokoli (any) |
| S | G, C (strong) | . | mezera |
| W | A, T (weak) | - | |

# 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 uchovávání

## 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
TCACCTTCTGCCGTGATTGTGAGGCCTCCTCAGCCACGTGGAACTGTAAAAACTCCTGGAA
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**

**AAAGTTTACCCCTCAAAGGGACGTGTTCGAAAGAA**

**Arial 24**

**TTTCAAATGGGGAGTTTCCCTGCACAAGCTTTCTT**

**AAAGTTTACCCCTCAAAGGGACGTGTTCGAAAGAA**

**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 inormační 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**

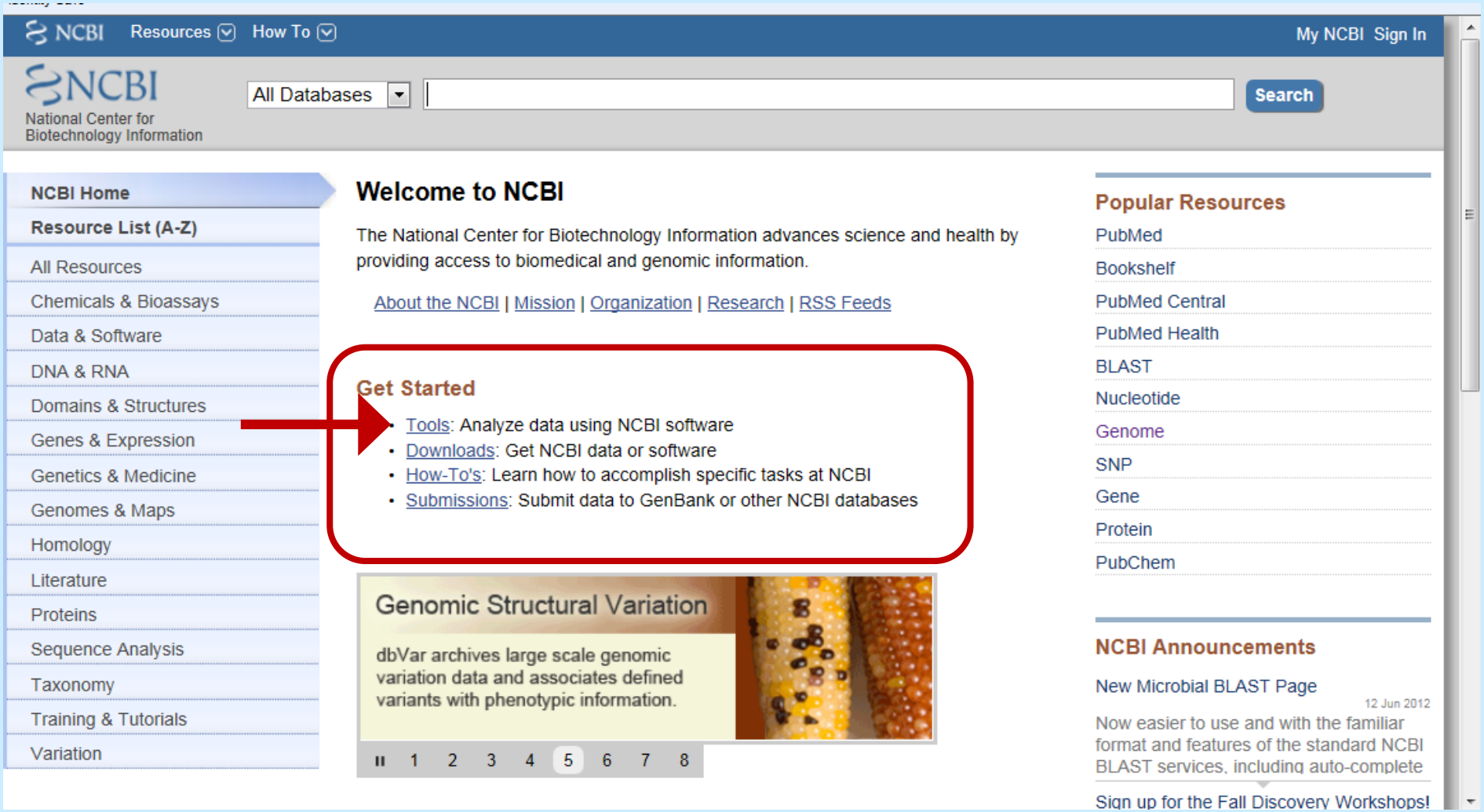
**Databáze PDB (The Protein Databank)**

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

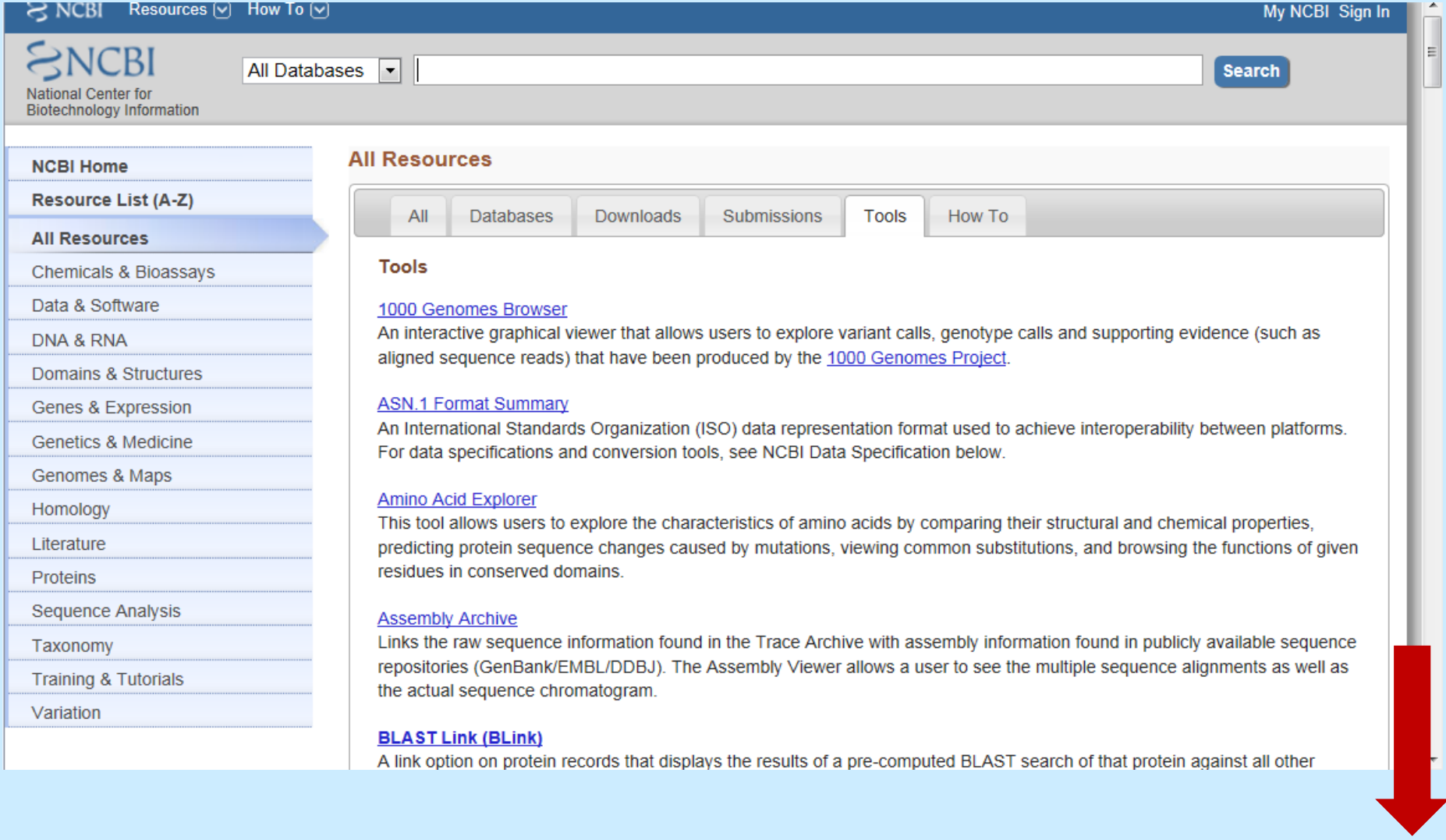
**http://www.rcsb.org/pdb/home/home.do**

# Práce s databází NCBI

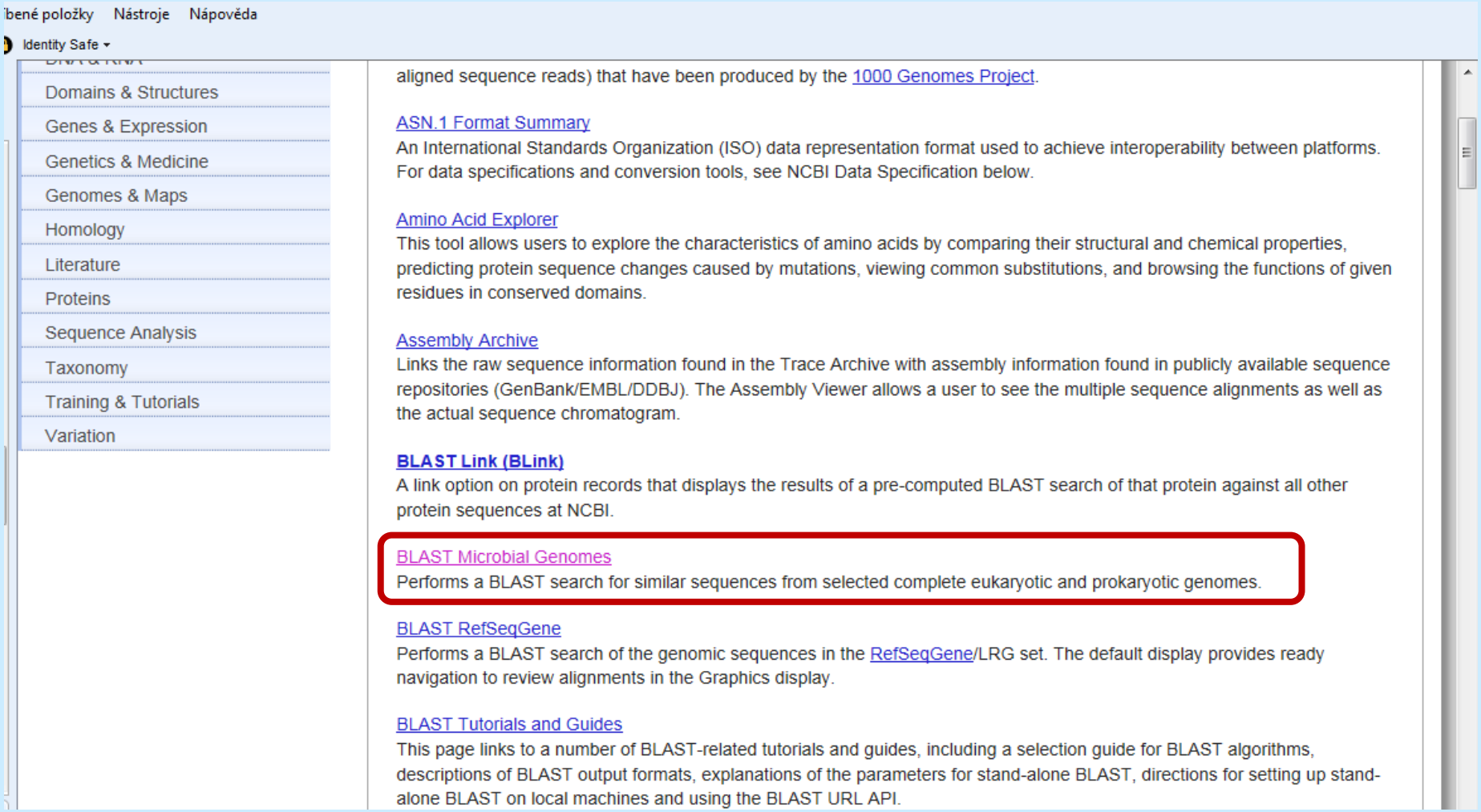
www.ncbi.nlm.nih.gov



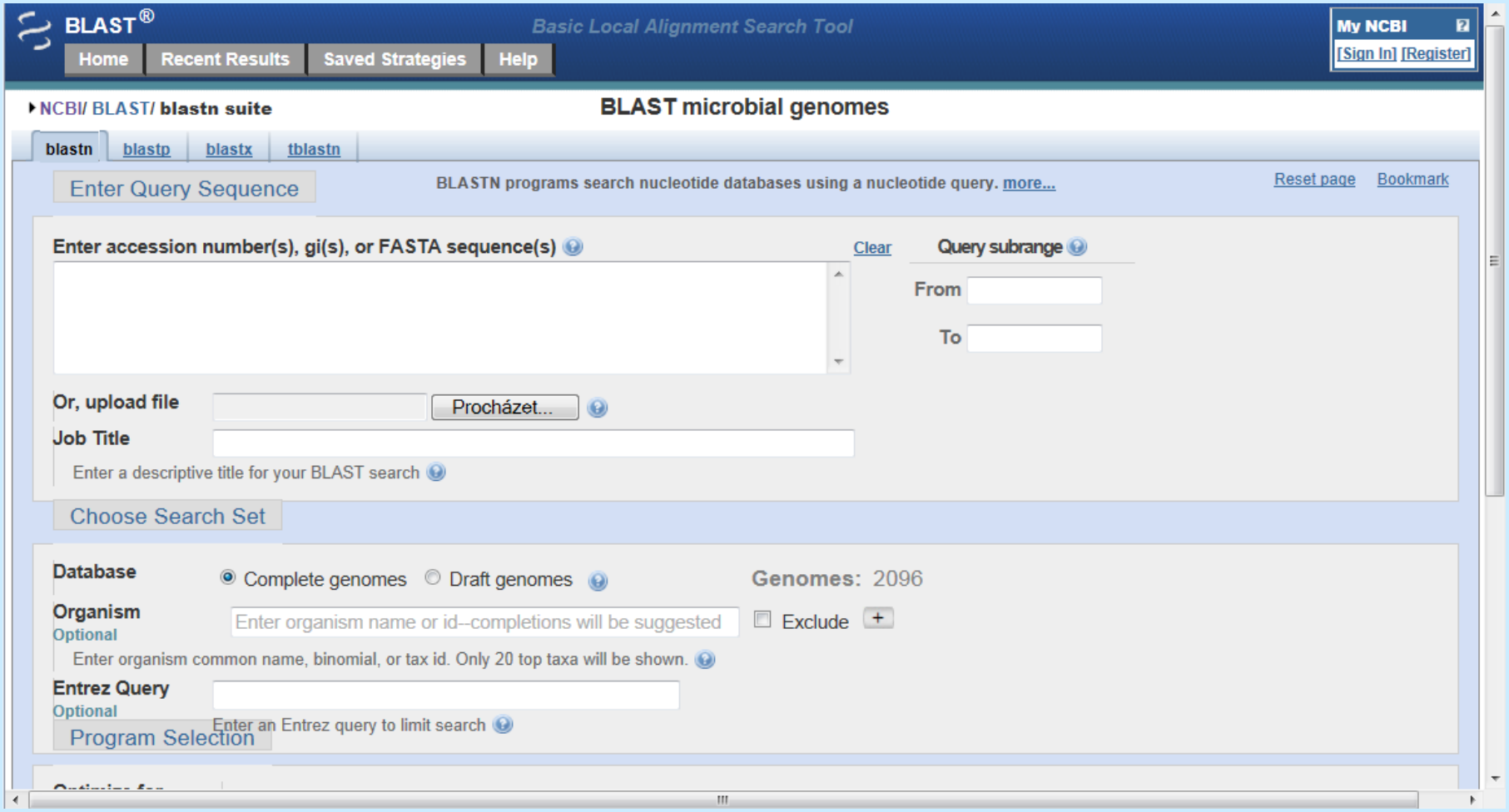
# Práce s databází NCBI



# Práce s databází NCBI



# Práce s databází NCBI




**Dostali jste se na prohledávač 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.

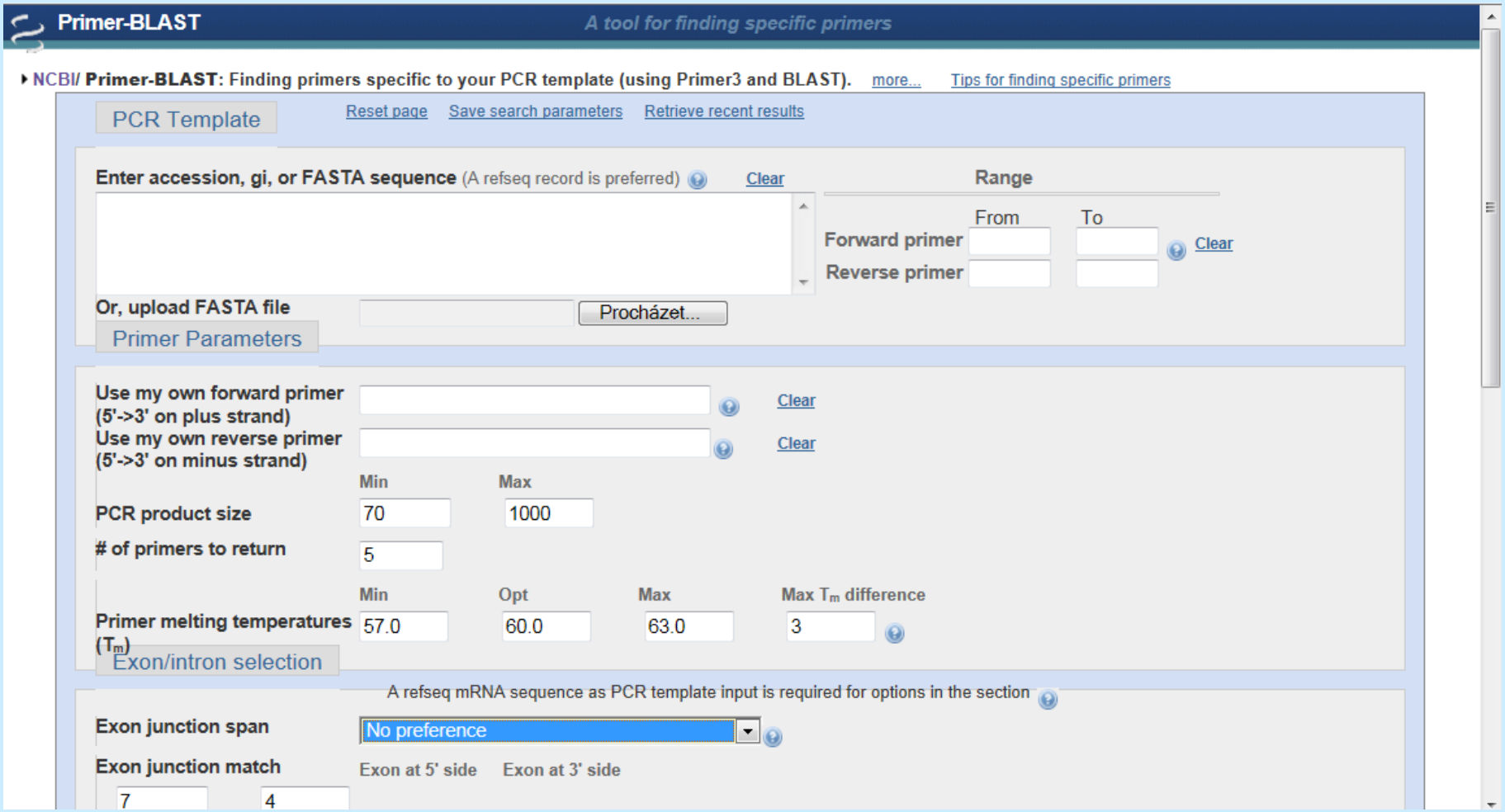
**ProSplign**
A utility for computing alignment of proteins to genomic nucleotide sequence. It is based on a variation of the Needleman Wunsch global alignment algorithm and specifically accounts for introns and splice signals. Due to this algorithm, ProSplign is accurate in determining splice sites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




**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***

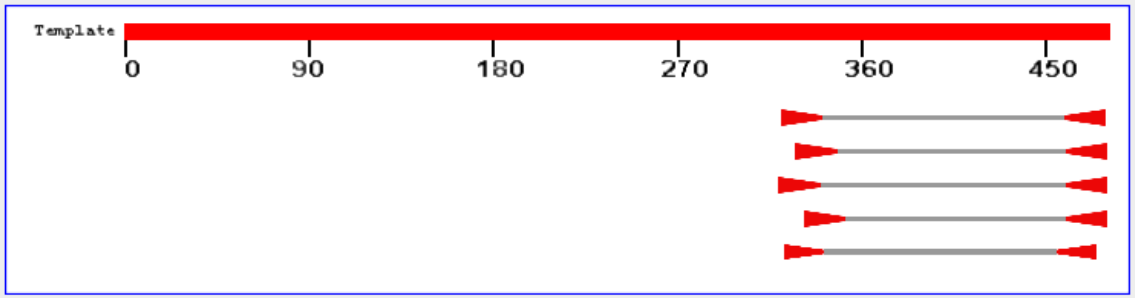
# Ukázka výsledku

**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
ATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAG
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.

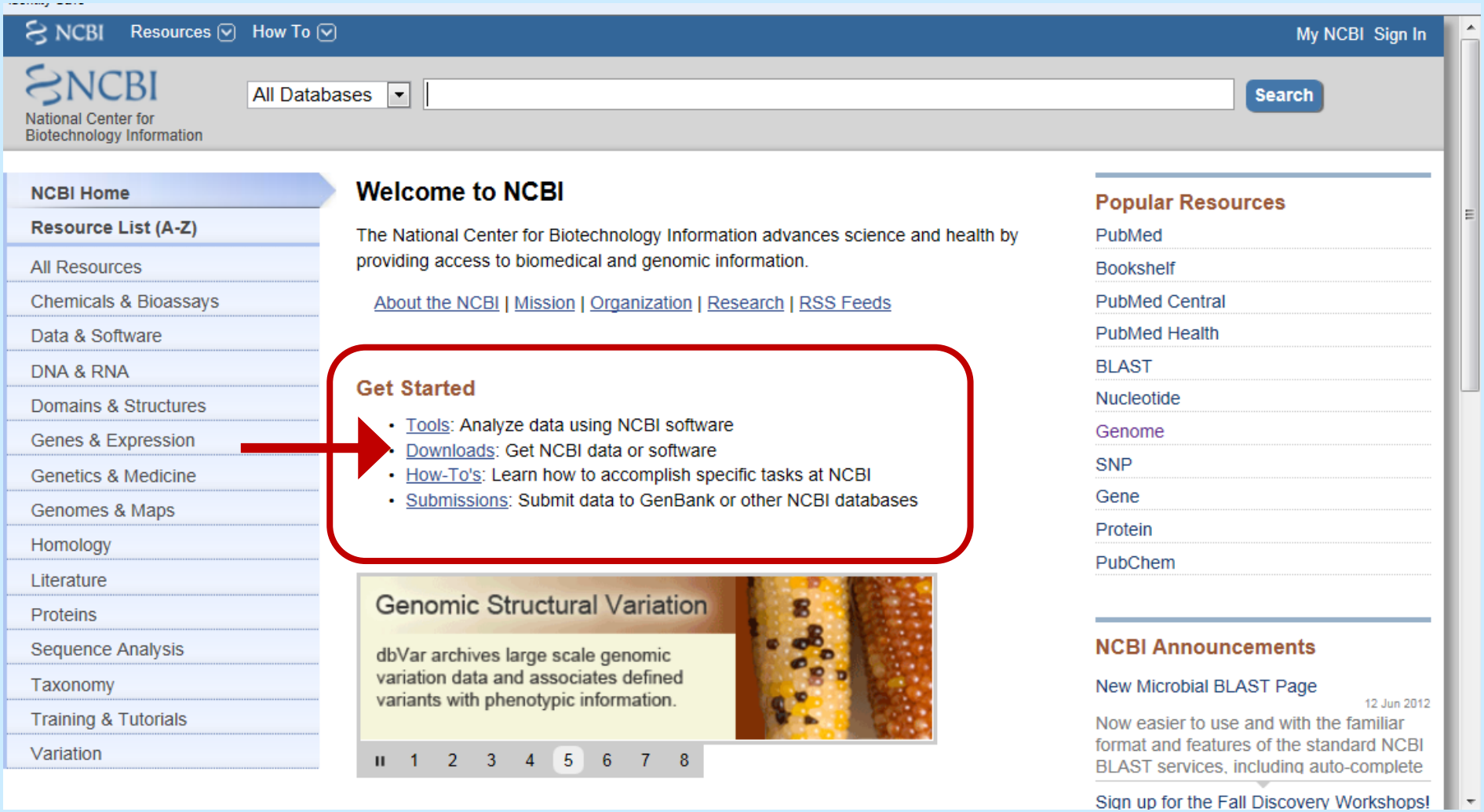
VecScreen
A system for quickly identifying segments of a nucleic acid sequence that may be of vector origin. VecScreen searches a

***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***

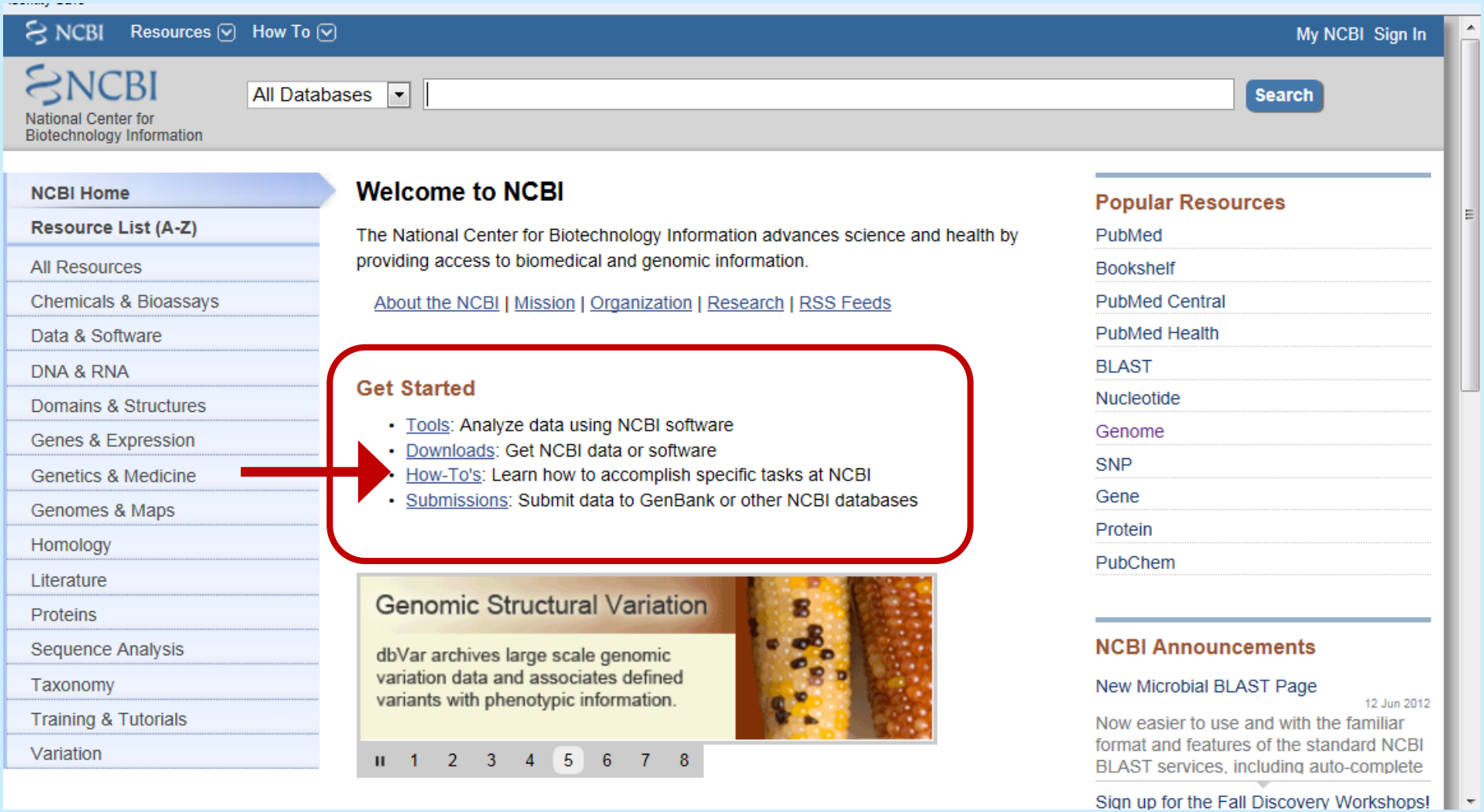
# Práce s databází NCBI

www.ncbi.nlm.nih.gov

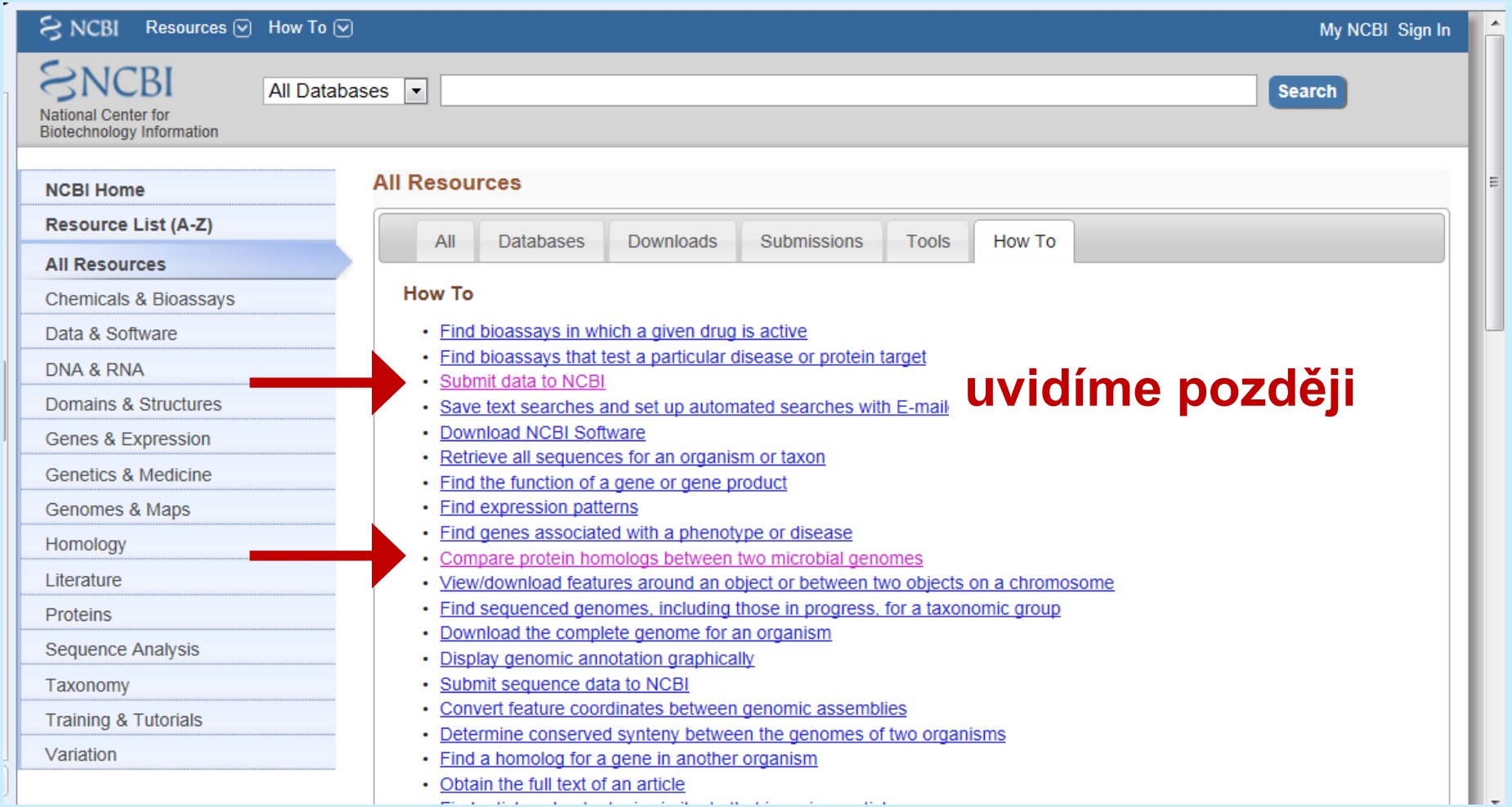


# Práce s databází NCBI

www.ncbi.nlm.nih.gov

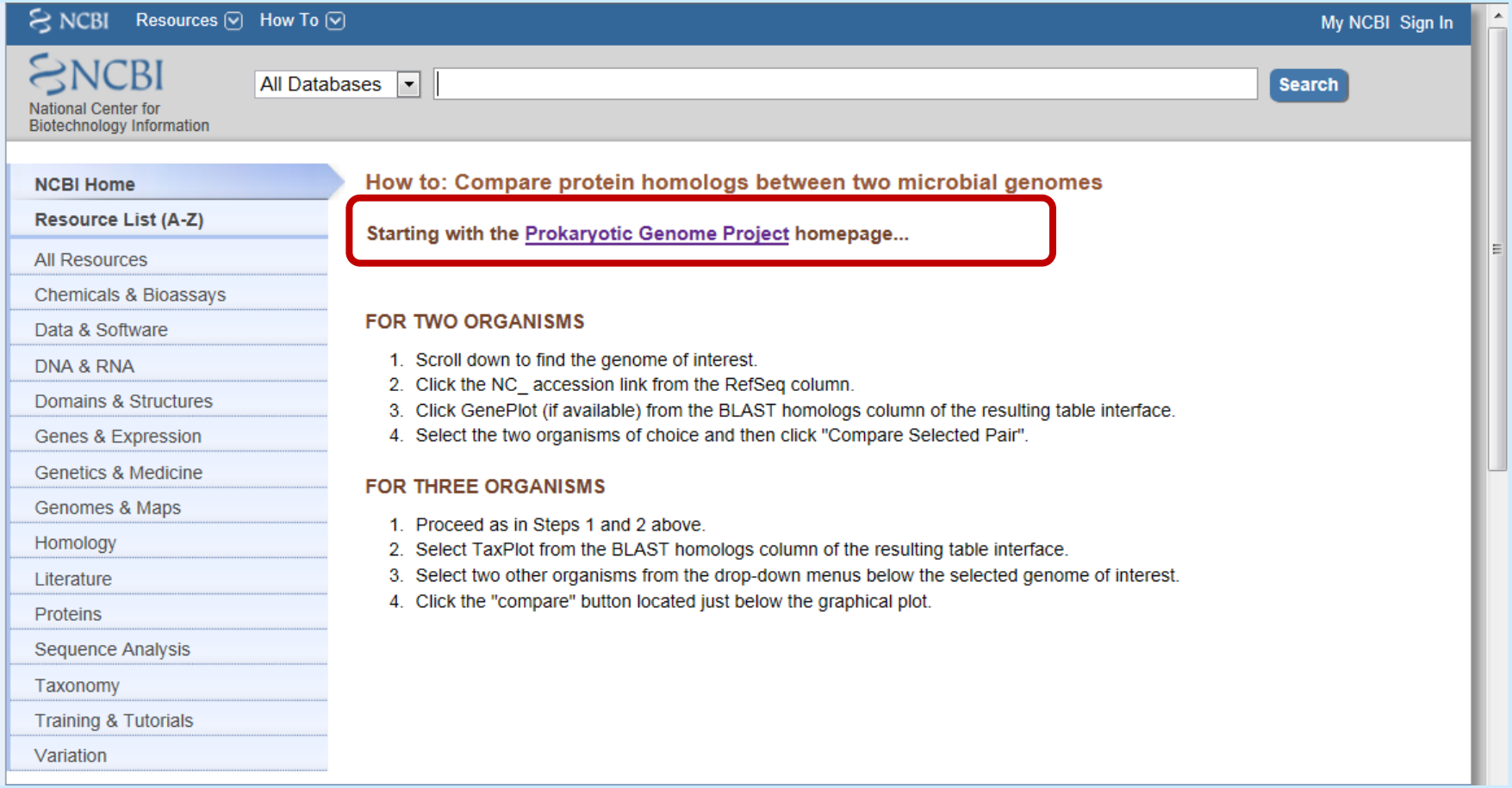


# Jak s nástroji pracovat




**uvidíme později**

# Porovnání proteinů u dvou genomů



# *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".

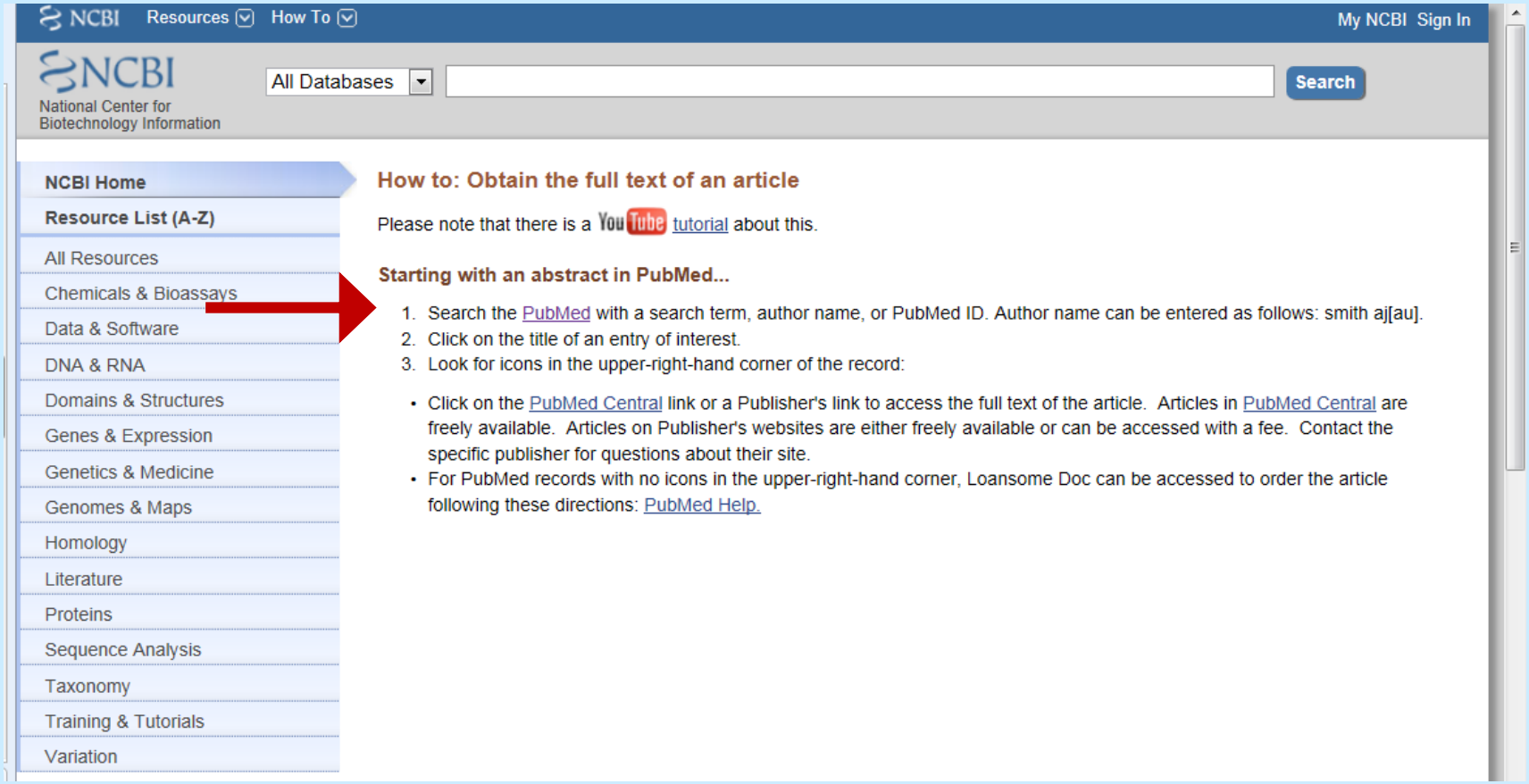
## FOR THREE ORGANISMS

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.

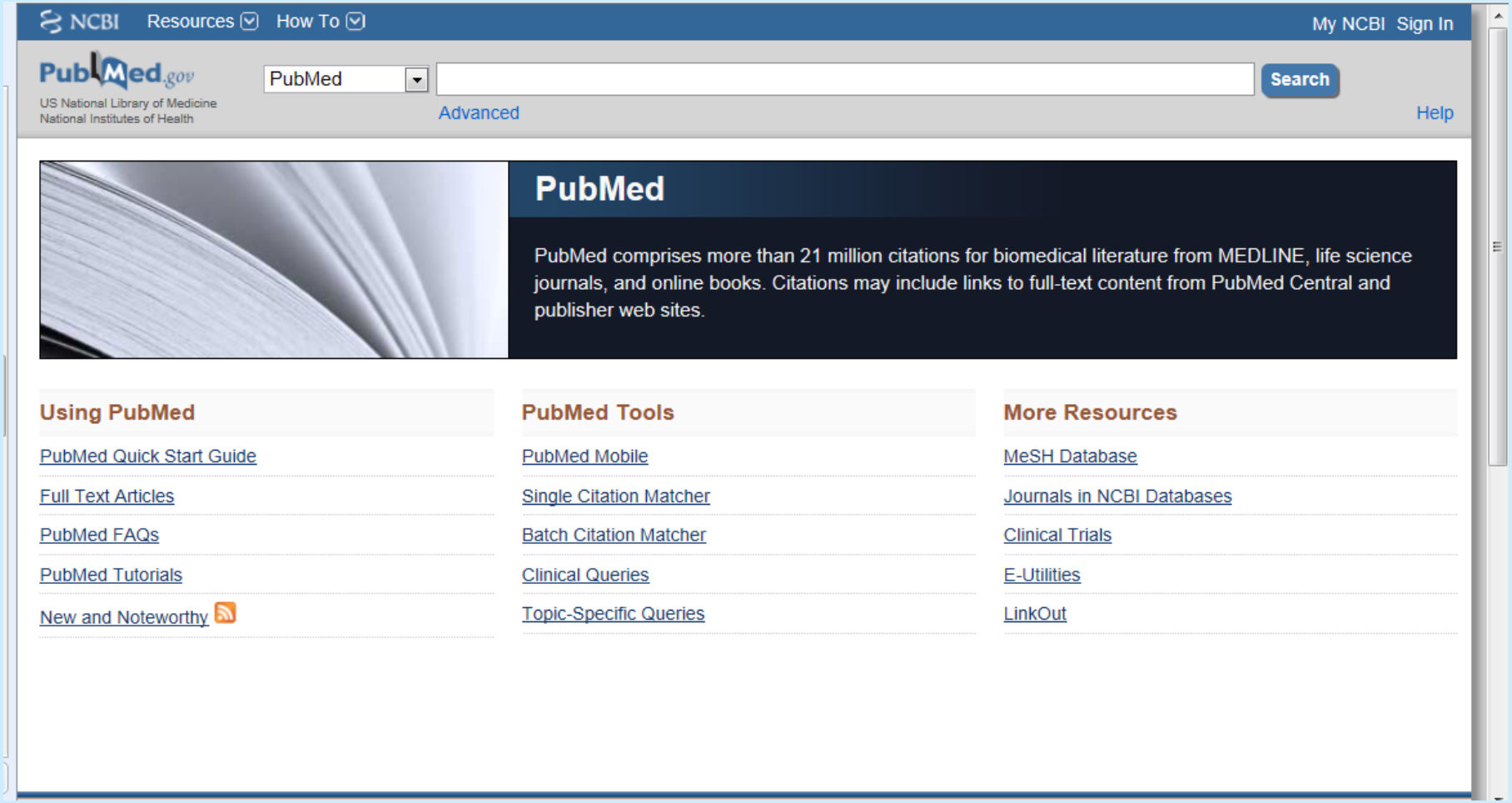
# Jak s nástroji pracovat

- 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



# Databáze PubMed

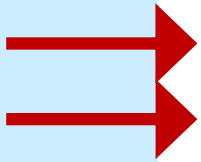


## *Najděte publikace o Deinococcus radiodurans*

## *Kolik review databáze obsahuje?*

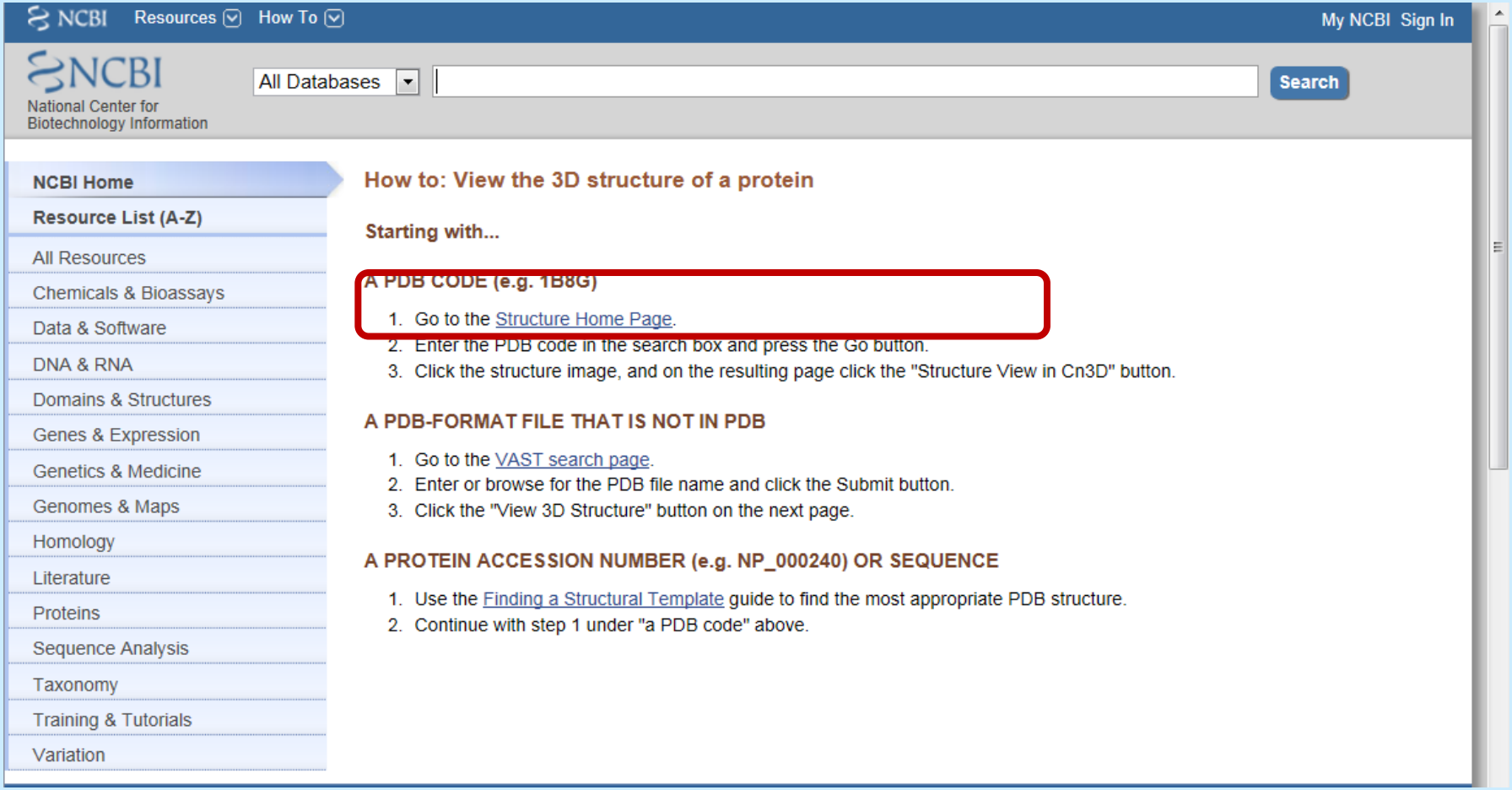
1) **Ke konci června 2012 jich bylo kolem 962**
2) **Z toho review bylo 52**
3) **Všimněte si, že jen některé jsou volně dostupné**

# Jak s nástroji pracovat

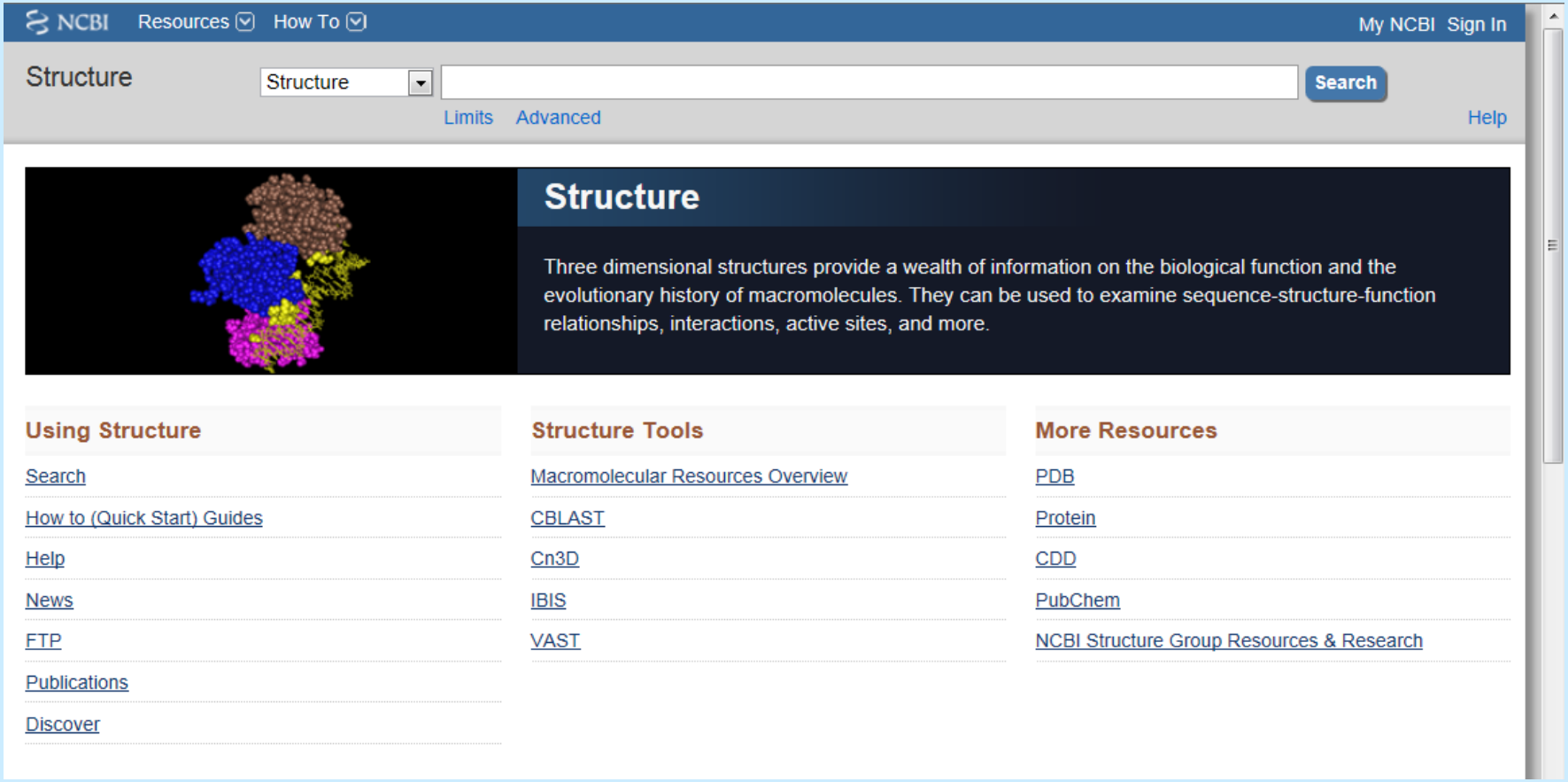


- Download the complete genome for an organism
- Display genomic annotation graphically
- Submit sequence data to NCBI
- Convert feature coordinates between genomic assemblies
- Determine conserved synteny between the genomes of two organisms
- Find a homolog for a gene in another organism
- Obtain the 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



# 3D struktury proteinů



***Najděte strukturu mykobakteriální katalázy***

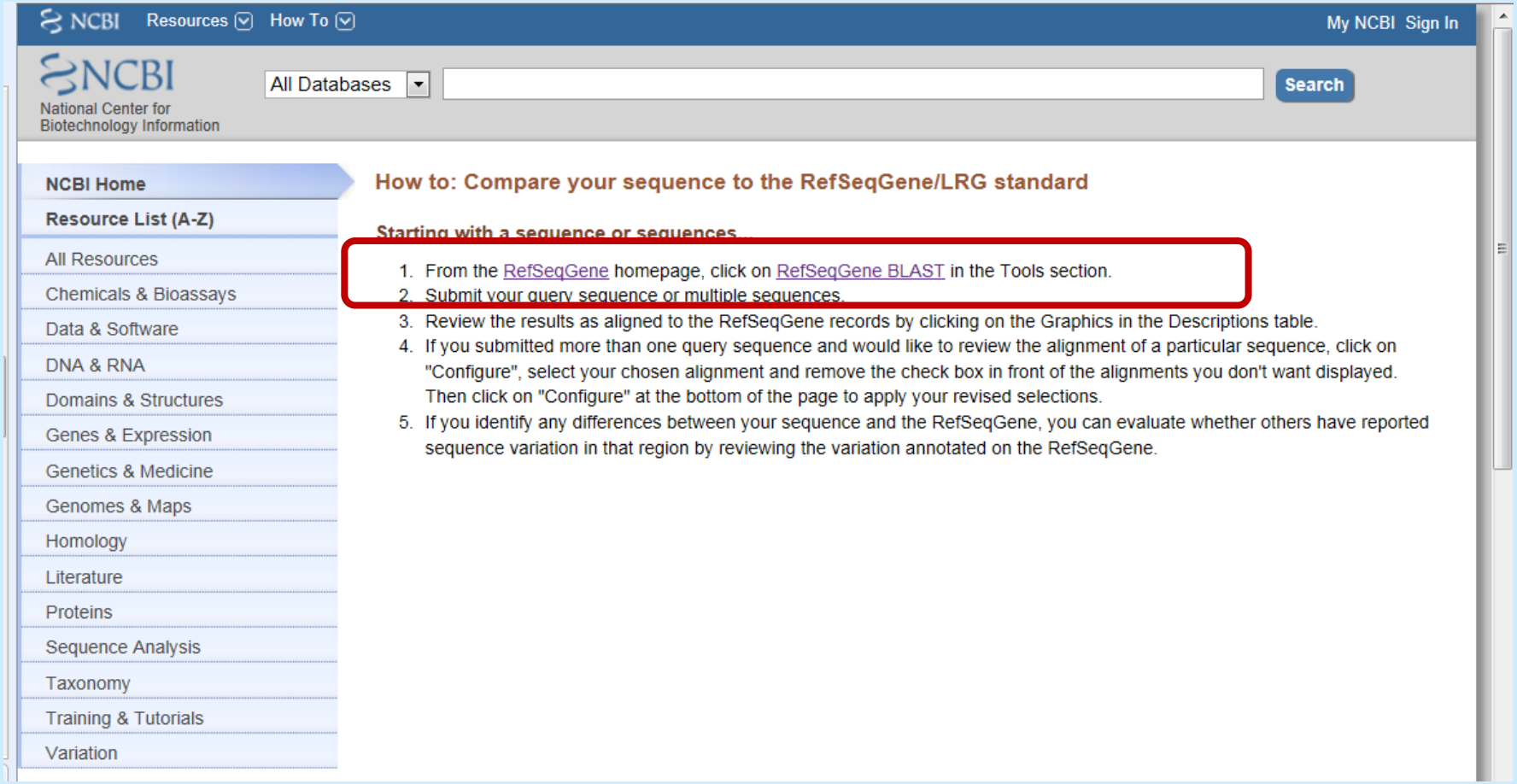
***Kolik záznamů najdete?***

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

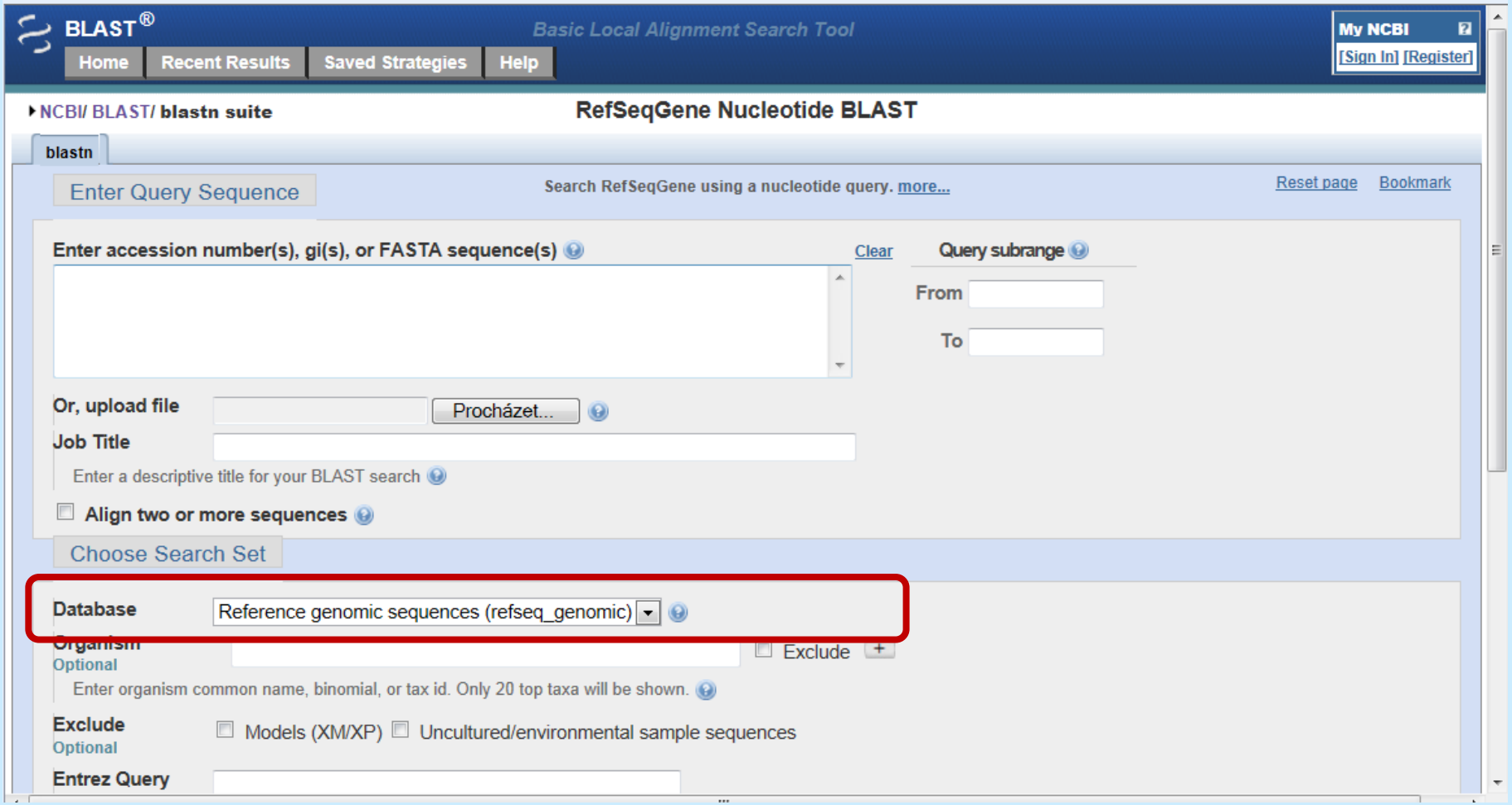
# Jak s nástroji pracovat

- 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



# Srovnání sekvence s referenčními



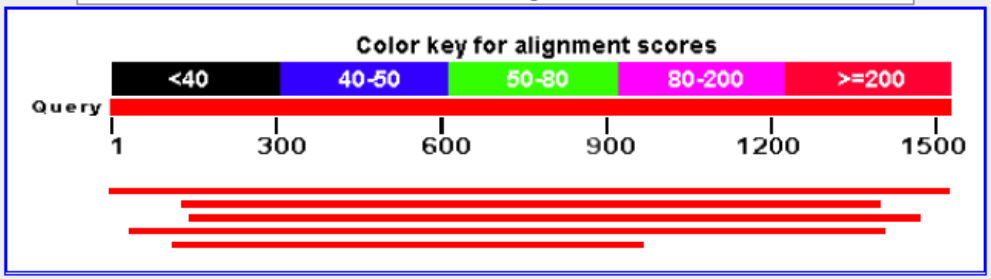
# *Zkopírujte si níže uvedenou sekvenci a porovnejte ji s databází referenčních sekvencí. Komu patří?*

1) ATGAGTGAAATGAAATGCCCTTATGACCATACCAACTTGACCATGAGTAATGGCGCGCCTGTTATTGACA
2) ACCAAAATTCAATGACCGCAGGTGCCAGAGGGCCACTGCTTGCCCAAGATTTATGGCTCAATGAAAAATT
3) AGCCGACTTTGCCCGTGAGGTCATTCCAGAACGCCGCATGCACGCCAAAGGCTCAGGCGCATTTGGCACA
4) TTCACGGTAACGCACGACATCACCCAATACACCCGTGCTAAGATTTTTAGTGAAGTTGGCAAAAAAACTG
5) AGATGTTCGCTCGTTTTACCACCGTAGCAGGCGAGCGGGGGGCGGCGGACGCTGAGCGTGATATCCGTGG
6) TTTTGCCCTAAAATTCTACACCGAAGAGGGTAATTGGGACATGGTGGGTAATAACACGCCTGTTTTCTTT
7) TTAAGAGACCCAAAAAAATTCCCTGATTTAAATAAAGCGGTCAAACGAGACCCACGCACCAACATGCGTT
8) CTGCCACCAATAACTGGGATTTTTGGACACTGCTGCCAGAGGCGTTTCATCAGGTGACCATTGTGATGAG
9) CGACCGTGGCATTCCTAAATCTTACCGTCATATGCACGGCTTTGGCTCGCACACTTATAGCTTTATCAAT
10) GCTGATAATGAACGCTTTTGGGTCAAATTTCACTTTCGCACCCAACAAGGCATTGAAAATCTAACCGATG
11) CCGAAGCTGAAATGGTGGTTGGTAAAGACCGTGAGAGCAATCAGCGTGATTTGTTTGATGCCATTGAGCG
12) TGGCGATTTCCCAAAATGGACAATGTATGTGCAAATCATGCCAGAAACCGATGCCCAAACTGTGCCTTAT
13) CACCCATTTGATTTAACCAAAGTGTGGCCAAAAGGCGACTATCCGCTCATTGAAGTGGGTGAGTTTGAGT
14) TAAATAAAAATCCTGAAAACTTCTTTTTAGACGTTGAACAATCCGCTTTTGCCCCAAGCAACCTAGTCCC
15) GGGCATCAGTGTGTCCCCTGACCGCATGCTCCAAGCACGCCTATTTAACTATGCTGATGCGCAGCGTTAT
16) CGTTTGGGCGTCAATCGTAACCAAATTCCAGTGAATGCCCCACGCTGTCCTGTGTACTCAAACCAAAGAG
17) ACGGACAAGGGCGAGTGGGCGATAACTATGGCGGTCGTCCGCACTATGAACCGAACAGTTTTGGACAATG
18) GCAAGACCAGCCGCATTTGGCTGAACCAGCATTAAAAATTCATGGCGATGCTAAGTTTTGGGATTATCGT
19) GAGAATGATGATGATTATTTTAGCCAACCCAGAGCCTTGTTTGAGTTGATGAGCGATGAGCAAAAACAGG
20) CGTTATTTGGTAATACGGCTCGTGCGATGGGCGATGCCCCTGATTTTATTAAATACCGCCATATCCGTAA
21) TTGCGATAAATGCCACCCTGATTATGCCATGGGTGTGGCCAAAGCGTTAGGCCTTACGGTTGAAGATGCC
22) AAAAATGCGTATGAGAGCGACCCTGCTCGCCATCTGCCCAGCTTTTTATA

# *Mohlo by vám vyjít to, co je na následující stránce*

## Distribution of 5 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments




## Descriptions

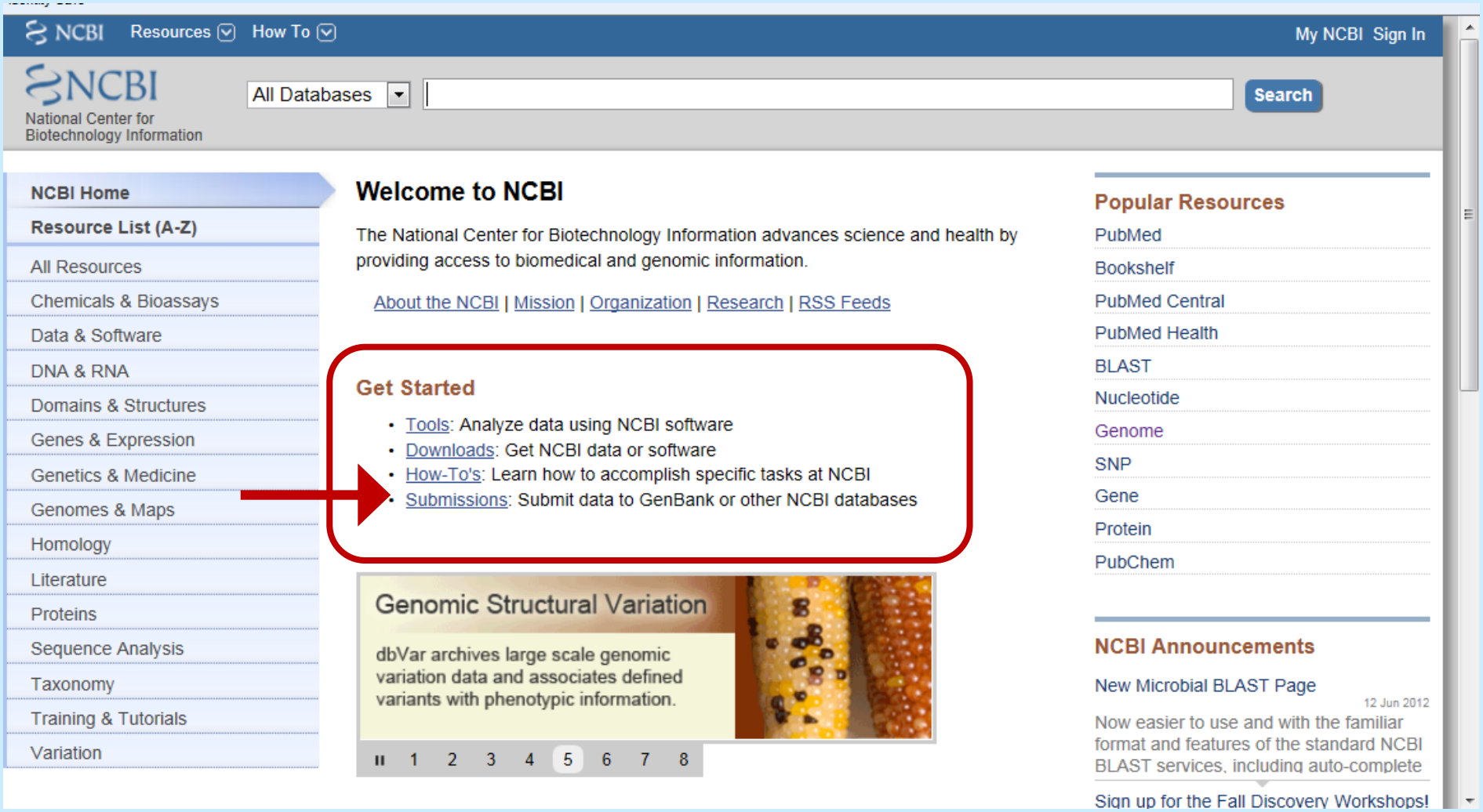
Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer PubChem BioAssay

**Sequences producing significant alignments:**

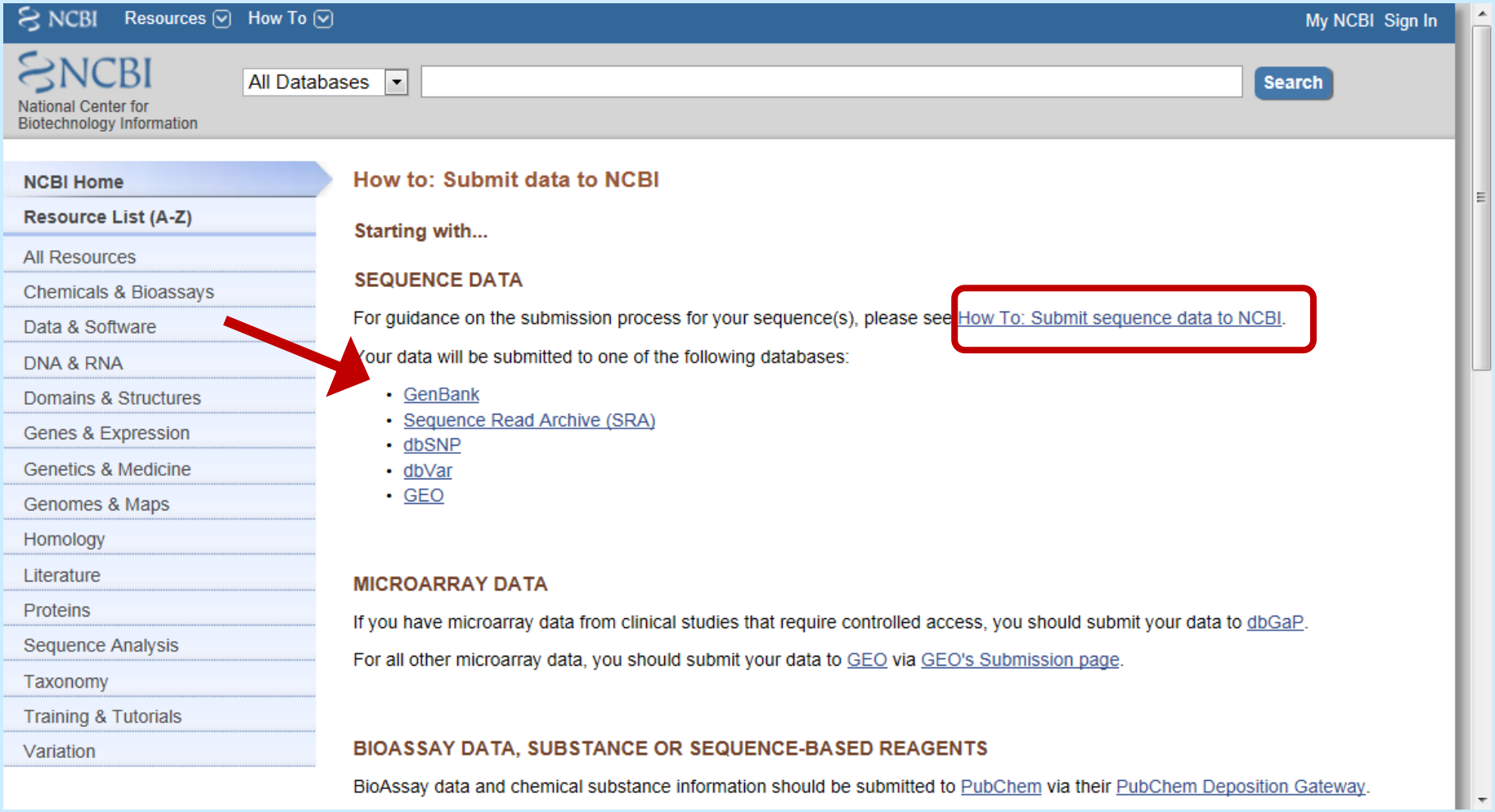
| Accession | Description | Max score | Total score | Query coverage | E value | Max ident | Links |
|---|---|---|---|---|---|---|---|
| NC_014147.1 | Moraxella catarrhalis RH4 chromosome, complete genome | 2808 | 2808 | 100% | 0.0 | 100% | |
| NC_015460.1 | Gallibacterium anatis UMN179 chromosome, complete 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



# Pokyny pro vložení vlastních dat



# *Posuzování podobnosti sekvencí*

# *Posuzování podobnosti sekvencí*

## Hledáme homologické sekvence vzniklé v průběhu evoluce

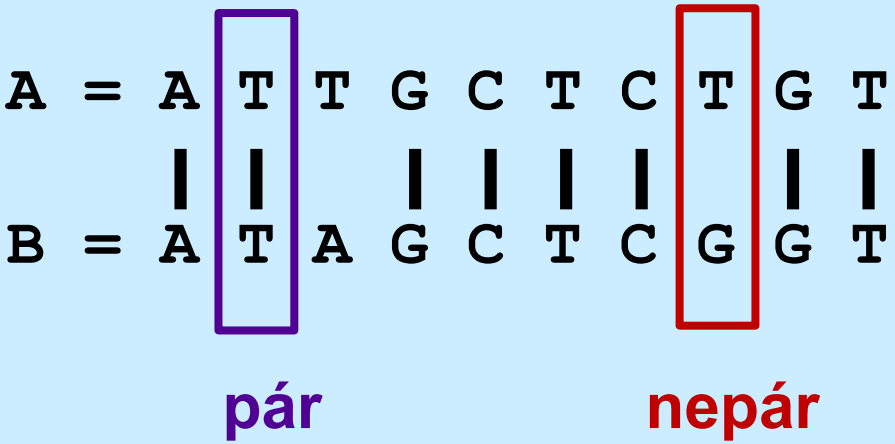
**Úkol: Jsou si podobnější sekvence A a B nebo B a C?**

## Výchozí sekvence

```
A = ATTGCTCTGT
B = ATAGCTCGGT
C = ATTGCACTGTAATGCCATGT
D = ATTGCTCTGAAATGCCCTGT
```

# *Posuzování podobnosti sekvencí*

**Přiložíme sekvence k sobě = přiřazení (alignment)**




```
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

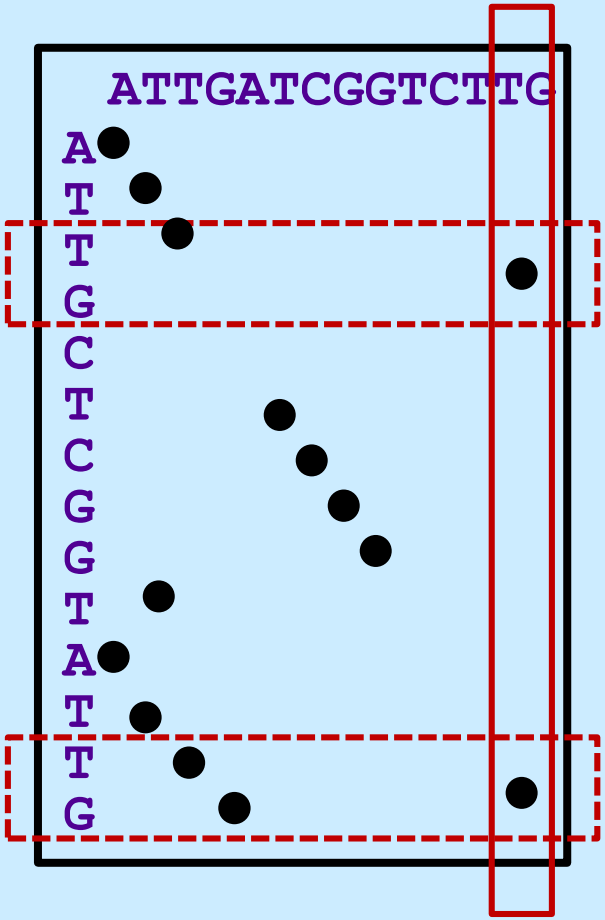
```
SLAV----------APATNIK-------PIQNYR-I------AKSETQRYMVIE
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVIE
```

## Local alignment

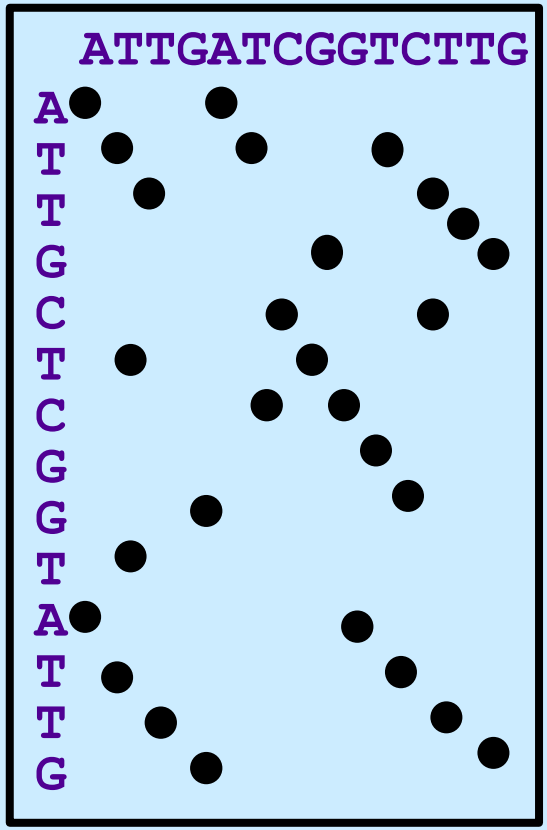
```
SLAVYTYIEFVRANAPATNIKSECVRAAPIQNYRRVEHVRATAKSETQRYMVIE
------------NAPATNIKSECVRA-PIQNYRRVEHVRA------------
```

# Bodový diagram

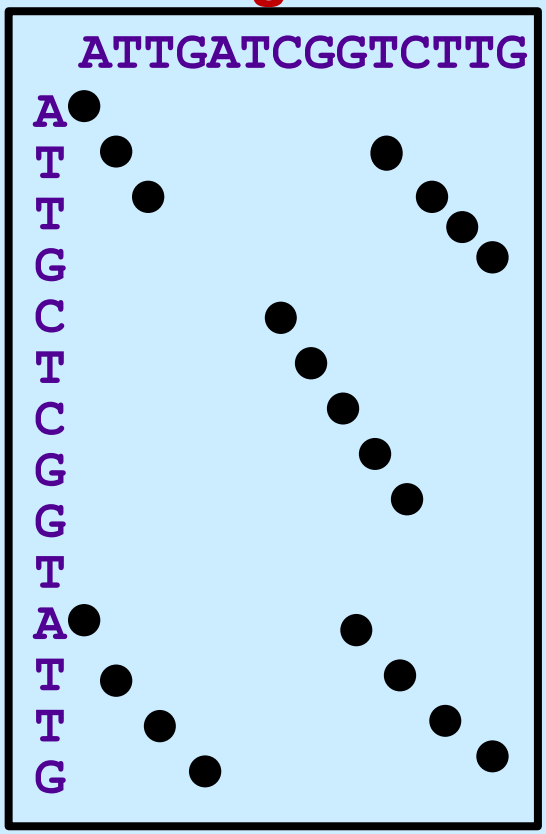
**Grafická mapa podobností sekvencí, pomůcka pro volbu přiřazení**



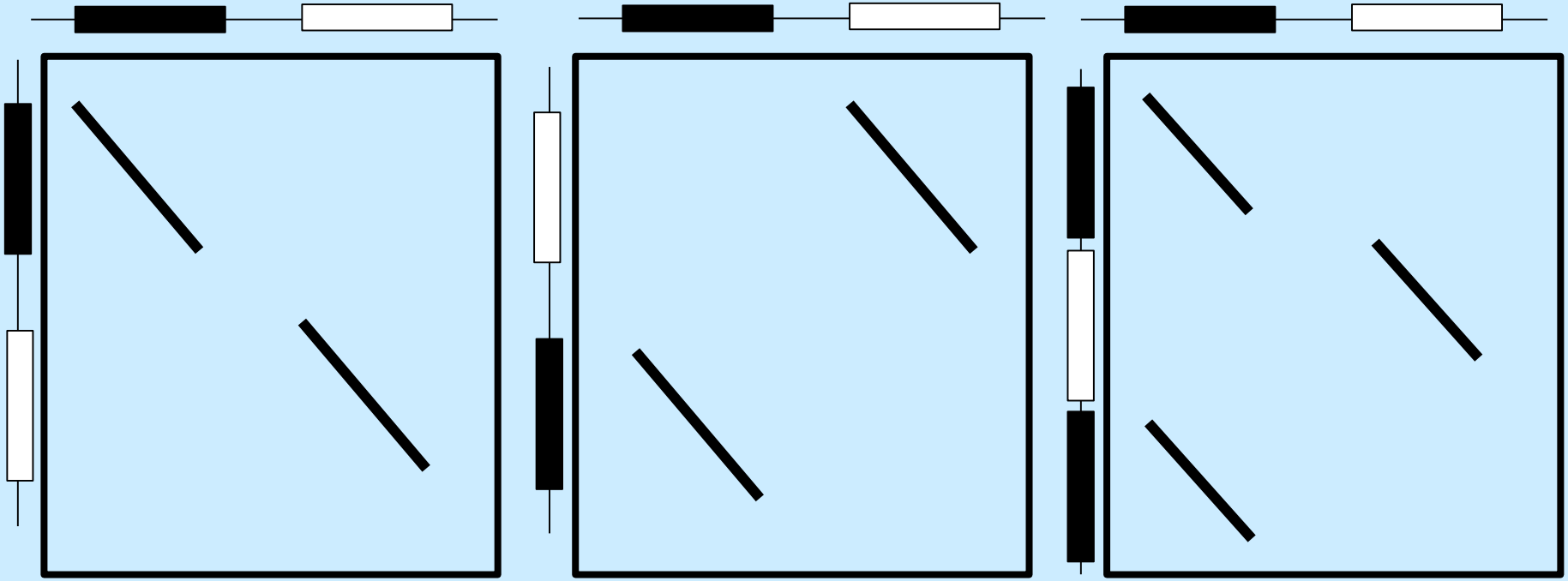

**Nalezené shody**




**Filtrace krátkých diagonál**



# *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

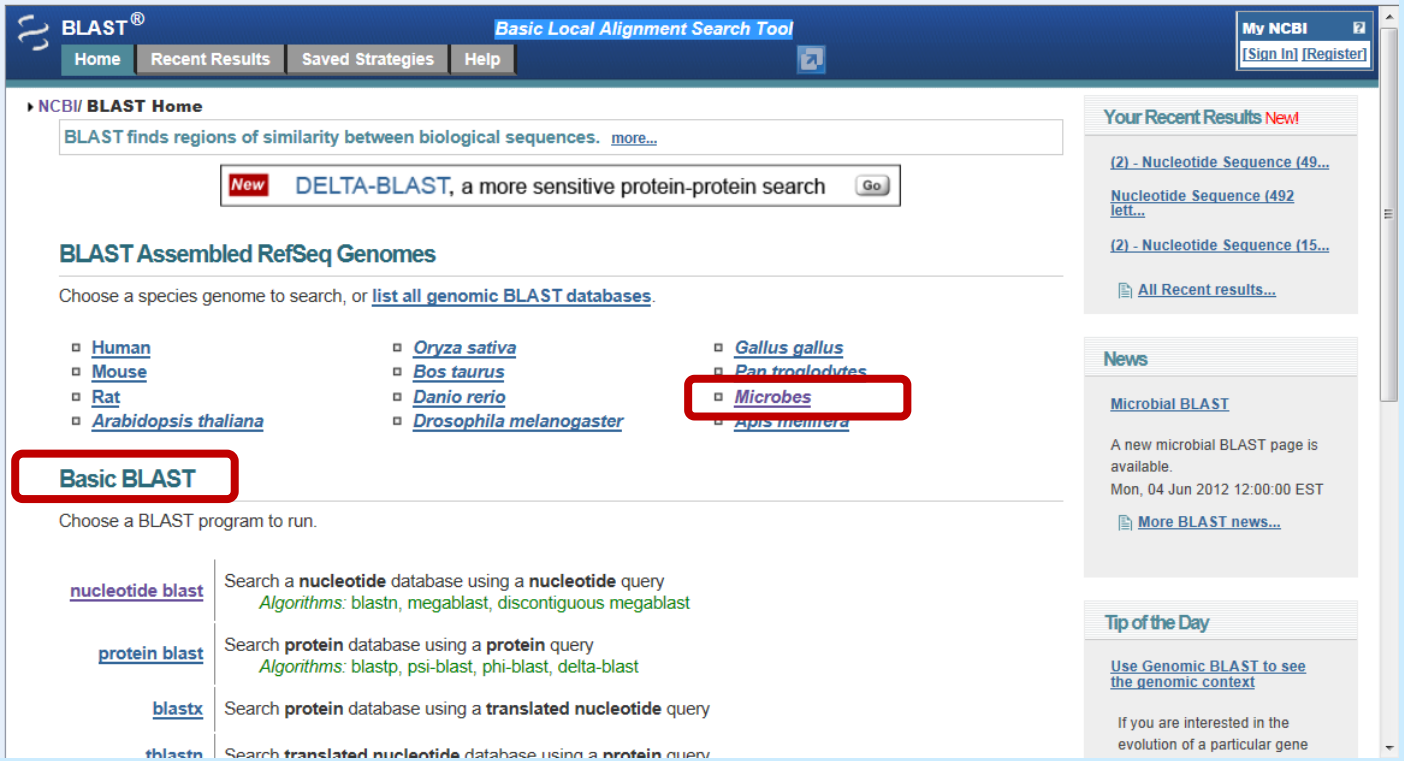
## BLAST

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

# BLAST

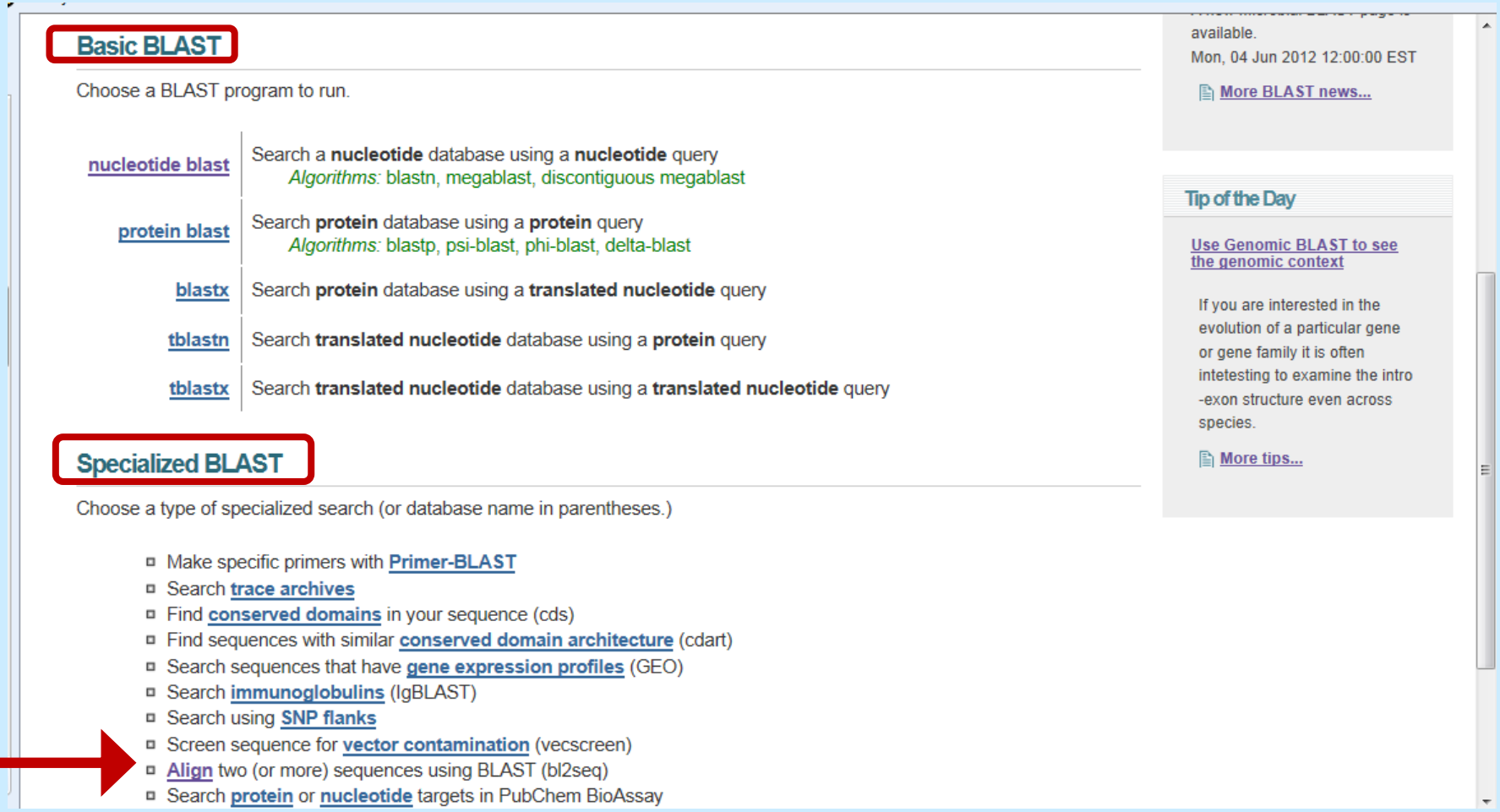
## Basic Local Alignment Search Tool

## http://blast.ncbi.nlm.nih.gov/Blast.cgi




**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<br>*Algorithms:* blastn, megablast, discontiguous megablast |
| protein blast | Search **protein** database using a **protein** query<br>*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 intetesting 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 |

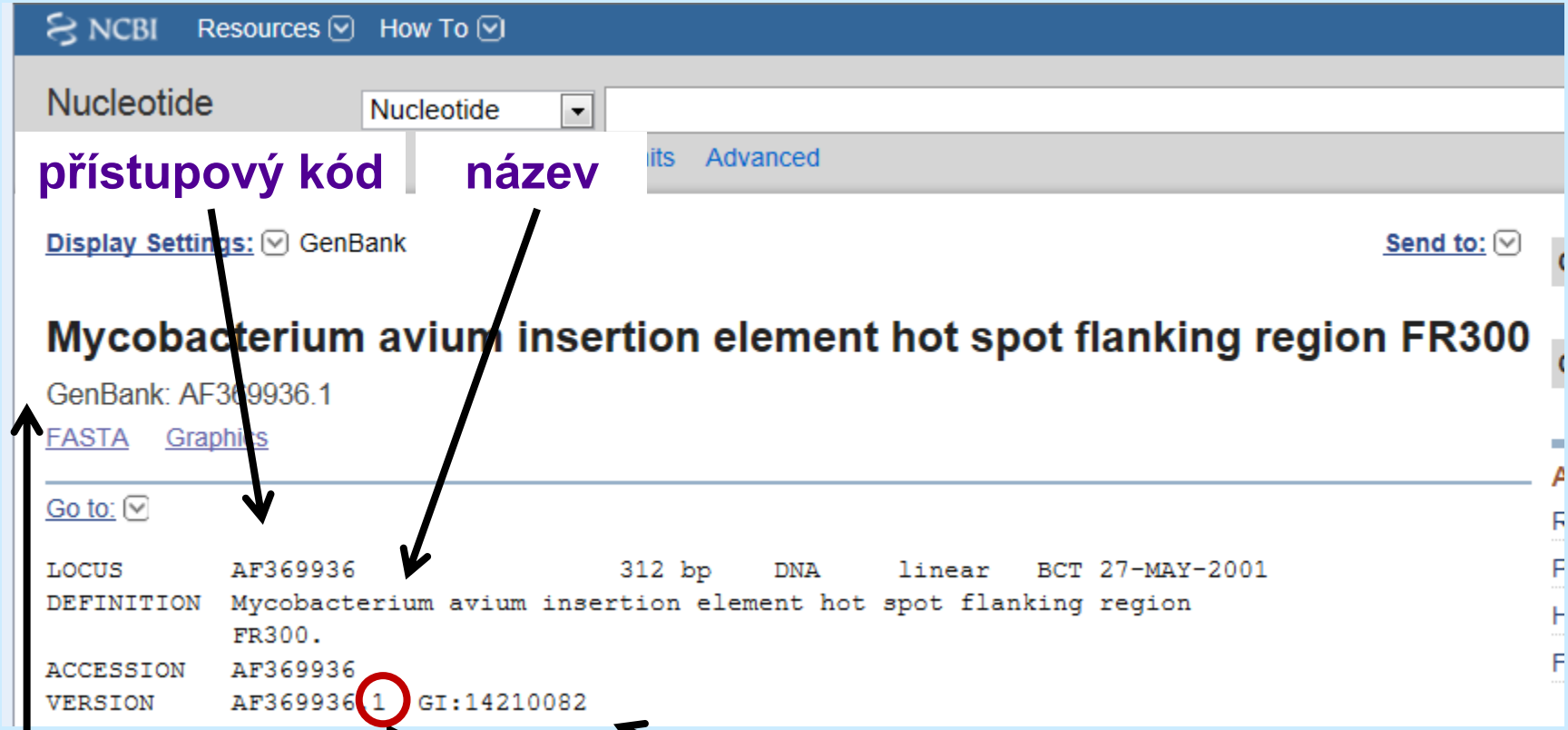
*** Jsou srovnávány přeložené DNA sekvence ve všech čtecích rámcích**

# *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**

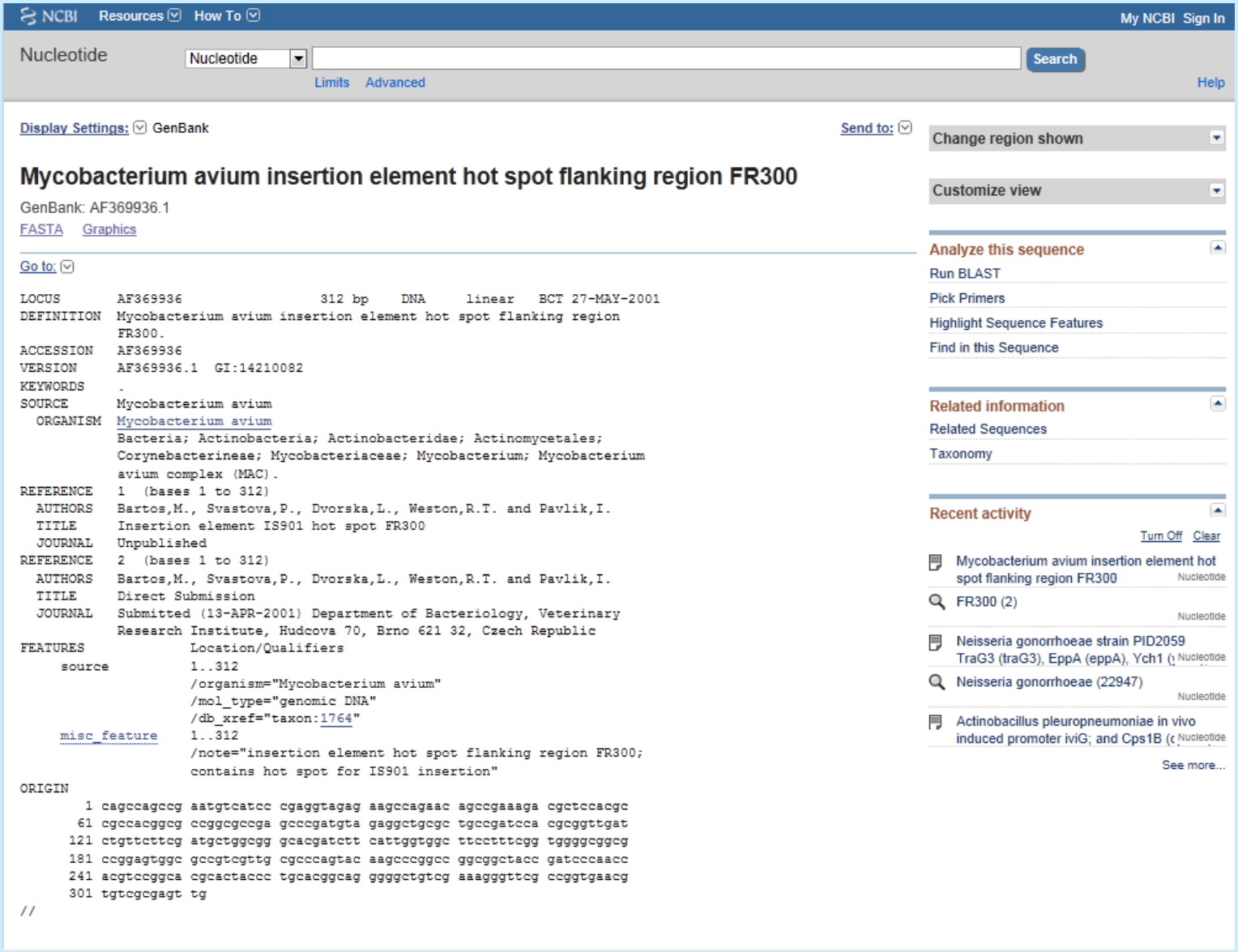
# Hlavička záznamů




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 cagccagccg aatgtcatcc cgaggtagag aagccagaac agccgaaaga cgctccacgc
       61 cgccacggcg ccggcgccga gcccgatgta gaggctgcgc tgccgatcca cgcggttgat
      121 ctgttcttcg atgctggcgg gcacgatctt cattggtggc ttcctttcgg tggggcggcg
      181 ccggagtggc gccgtcgttg cgcccagtac aagcccggcc ggcggctacc gatcccaacc
      241 acgtccggca cgcactaccc tgcacggcag ggggctgtcg aaagggttcg ccggtgaacg
      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...

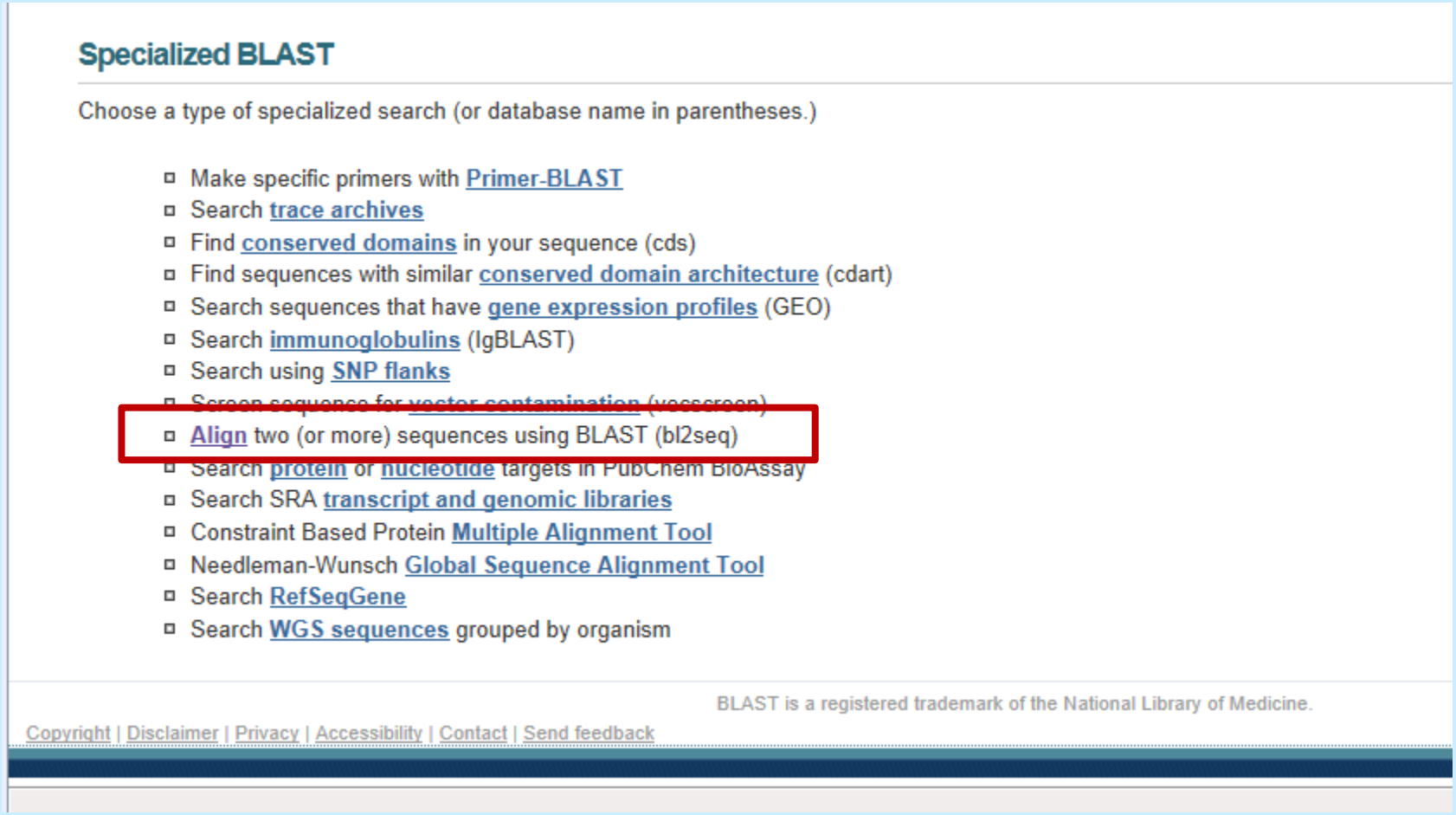
# *Anatomie databázového záznamu*

***Mycobacterium avium* FR300**

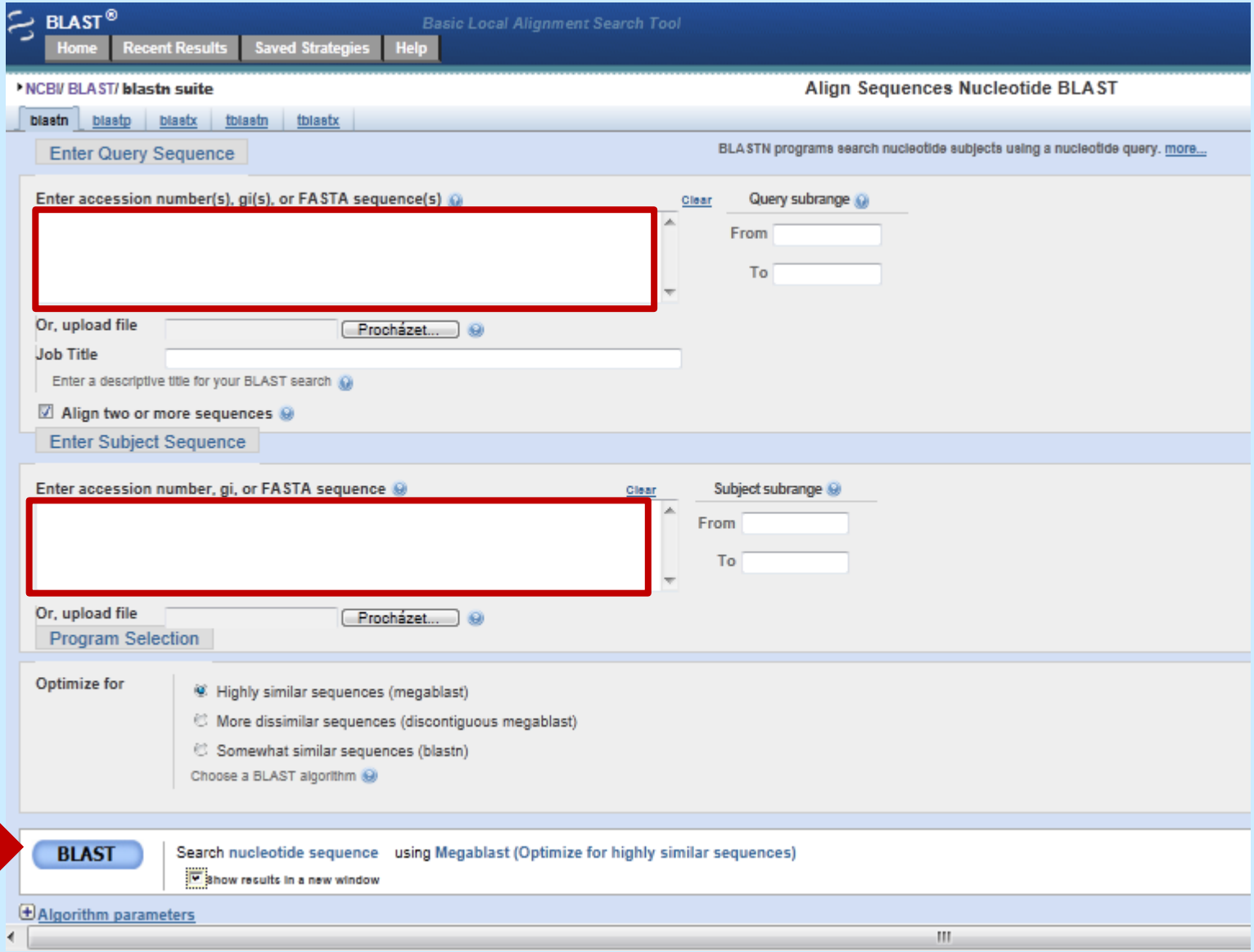
***Neisseria gonorrhoeae***

# Program bl2seq

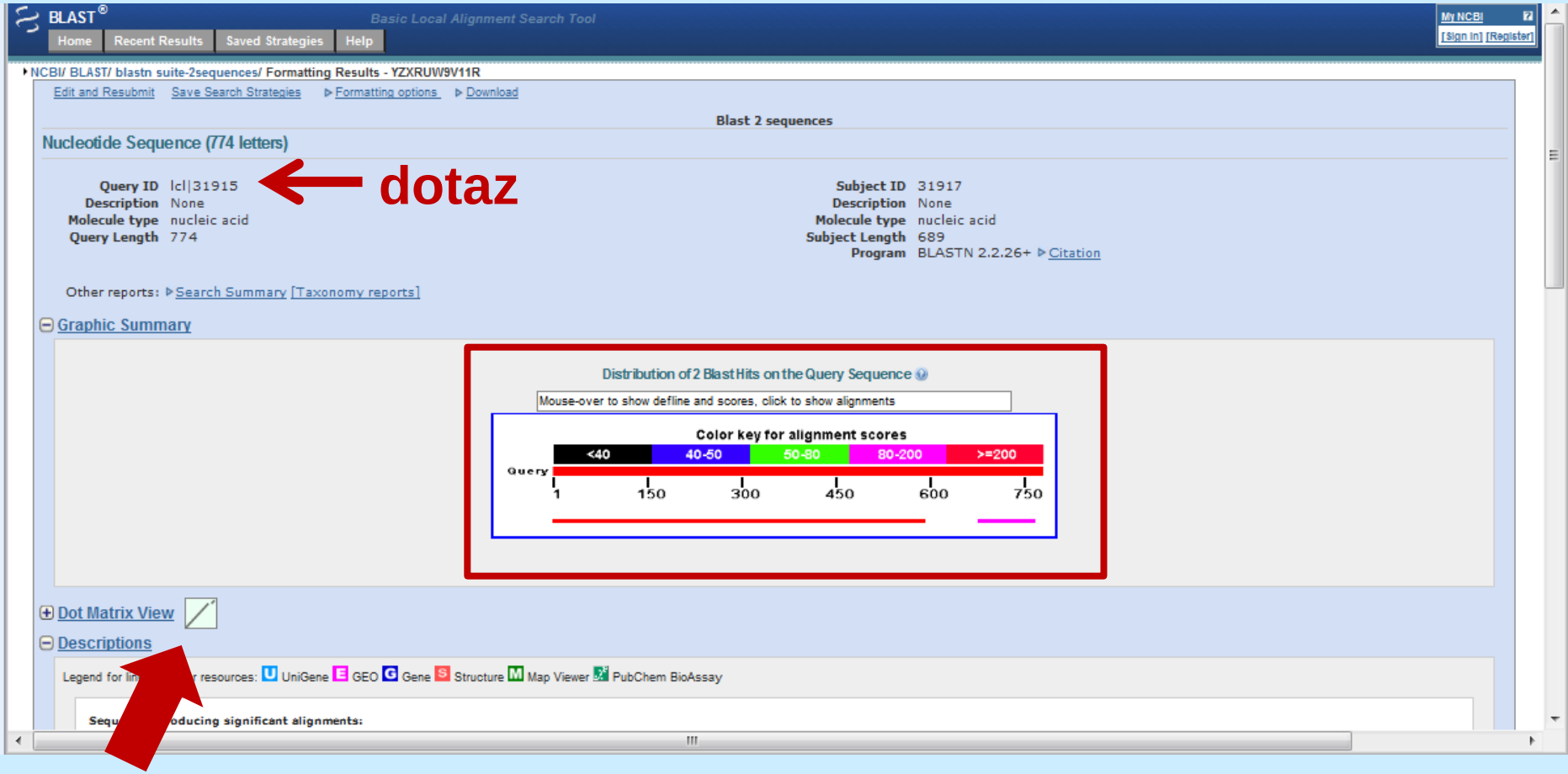
## Porovnání dvou a více sekvencí



# Program bl2seq

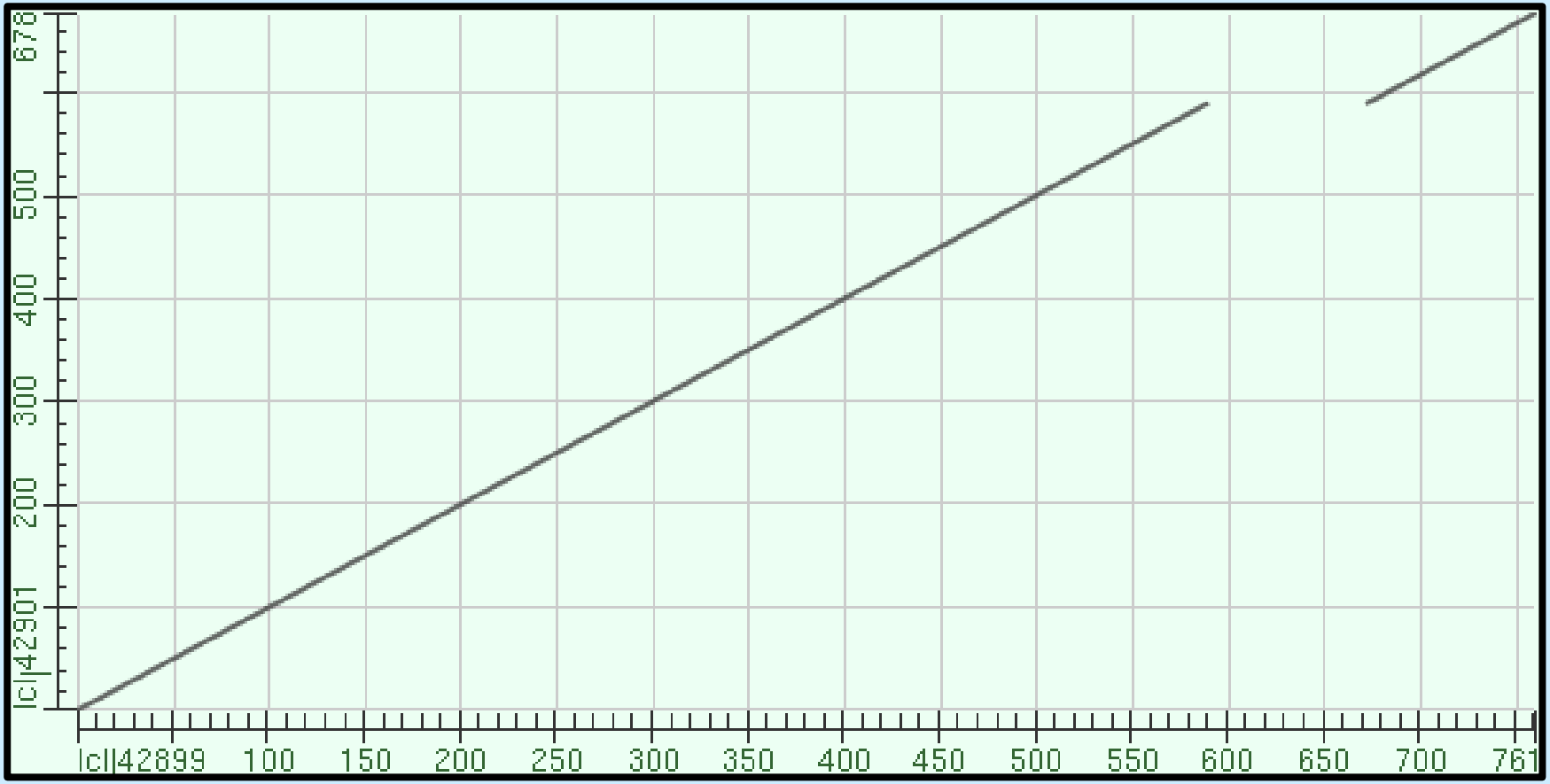


# Výsledek porovnání dvou sekvencí

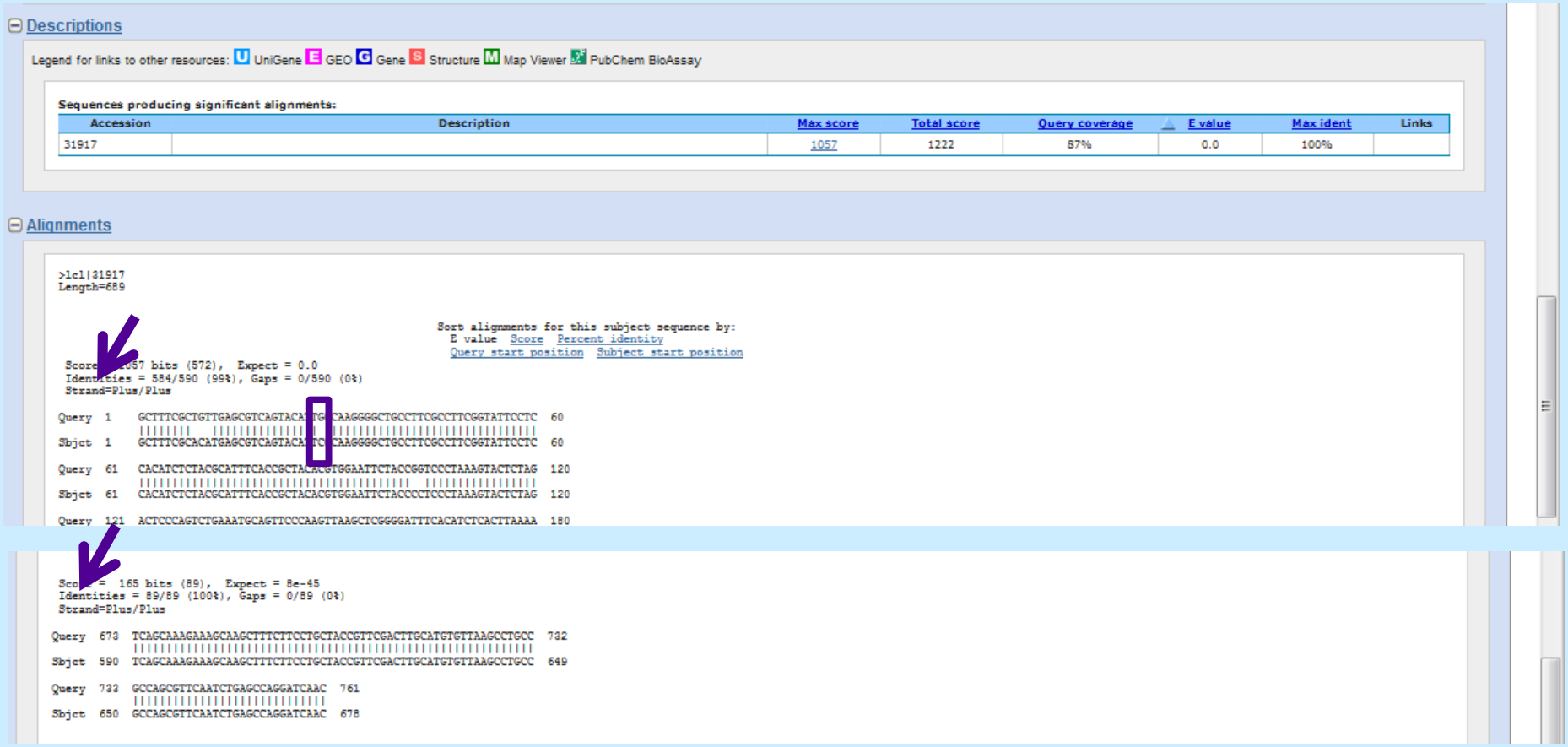


# *Dot Matrix View*

**Plot of lcl|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 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    GCTTTCGCTGTTGAGCGTCAGTACATGCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60
            ||||||||   ||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  1    GCTTTCGCACATGAGCGTCAGTACATTCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60

Query  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCGGTCCCTAAAGTACTCTAG  120
            ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Sbjct  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAGTACTCTAG  120

Query  121  ACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACATCTCACTTAAAA  180
```

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

```
Query  673  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  590  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  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 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    GCTTTCGCTGTTGAGCGTCAGTACATTGCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60
            ||||||||   ||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  1    GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60

Query  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCGGTCCCTAAAGTACTCTAG  120
            ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Sbjct  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAGTACTCTAG  120

Query  121  ACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACATCTCACTTAAAA  180
```

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

Query  673  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  732
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  590  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  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 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    GCTTTCGCTGTTGAGCGTCAGTACATTGCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60
            ||||||||   ||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  1    GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATTCCTC  60

Query  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCGGTCCCTAAAGTACTCTAG  120
            ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Sbjct  61   CACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAGTACTCTAG  120

Query  121  ACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACATCTCACTTAAAA  180
```

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

Query  673  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  590  TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCCTGCC  649

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

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

# *Něco navíc k procvičení BLAST*

# Prohledejte databázi a zjistěte, jakému organismu patří následující sekvence

GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCCCAGTTATTCCGATTAACGCT
CGCACCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
AATTAACGTCAATGATGCTATCTATTTAACAACATCCCTTCCTCATTACCGAAAGA
ACTTTACAACCCGAAGGCCTTCTTCATTCACGCGGCATGGCTGCGTCAGGGTTCCC
CCCATTGCGCAATATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCTCA
GTTCCAGTGTGGCTGGTCATCCTCTCAGACCAGCTAGAGATCGCAGGCTTGGTAGG
CCTTTACCCCACCAACTACCTAATCCCACTTGGGCTCATCTTATGGCAGGTGGCCC
TAAGGTCCCACCCTTTCCTCCTCAGAGAATACGCGGTATTAGCTGCAGTTTCCCAC
AGTTATCCCCCTCCATAAGCCAGATTCCCAAGCATTACTCACCCGTCCGCCACTCG
TCAGCAAAGAAAGCAAGCTTTCTTCCTGCTACCGTTCGACTTGCATGTGTTAAGCC
TGCCGCCAGCGTTCAATCTGAGCCAGGATCAACNTCTTTCTCCAAA

## Měla by to být Pasteurella multocida

# *Porovnejte tyto dvě sekvence, patří stejnému druhu?*

GCTTTCGCACATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAATGCAGTTCCCAAGTTAAGCTCGGGGATTTCACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCCCAGTTATTCCGATTAACGCT
CGCACCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
AATTAACGTCAATGATGCTATCTATTTAACAACATCCCTTCCTCATTACCGAAAGA
ACTTTACAACCCGAAGGCCTTCTTCATTCACGCGG

GCTTTCGCGCATGAGCGTCAGTACATTCCCAAGGGGCTGCCTTCGCCTTCGGTATT
CCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCTACCCCTCCCTAAAG
TACTCTAGACTCCCAGTCTGAAAAGCAGTTCCCAAGTTAAGCTCGGGGATTTCACA
TCTCACTTAAAAGTCCGCCTGCGTGCCCTTTACGCGCAGTTATTCCGATTAACGCT
CGCACCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGT
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