LOSCHMIDT
LABORATORIES

Úvod do bioinformatiky a bioinformatické databáze

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Osnova

- Organizační informace
- Studijní literatura
- Historie bioinformatiky
- Bioinformatické instituce
- Bioinformatické databáze
- Prohledávání databází

Organizační informace

- Kolokvium
    - Písemný test
    - Celkem 25 otázek s jednou i více správnými odpověďmi
    - Minimálně 17 správných odpovědí

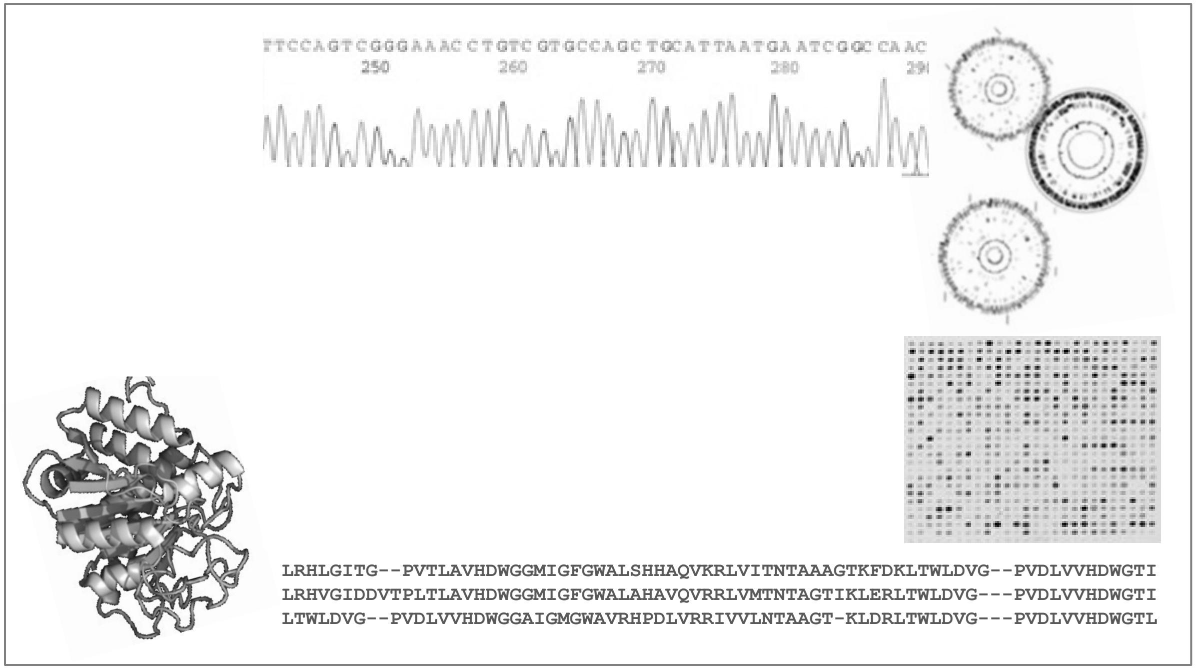
- Bi5000 Bioinformatika I – nukleové kyseliny
- Bi9060 Bioinformatika II – proteiny
- Bi9061 Bioinformatika – cvičení

Studijní literatura

- ☐ Xiong, J. **Essential Bioinformatics.** Cambridge University Press, New York, 2006.
- ☐ Claverie, J., and Notredame, C. **Bioinformatics for Dummies** $^{2\ ed}$**.** Wiley Publishing, Hoboken, 2006
- ☐ Cvrčková, F. **Úvod do praktické bioinformatiky.** Academia, Praha 2006.
- ☐ Misener, S., Krawetz S.A. **Bioinformatics: methods and protocols.** Humana Press, Totowa, New Jersey 2000.
- ☐ Attwood, T.K., Parry-Smith, D.J. **Introduction to bioinformatics.** Longman, Essex, 1999.
- ☐ Baxevanis, A.D., Ouellette, F.B.F. **Bioinformatics: a practical guide to the analysis of genes and proteins.** Wiley-Interscience, New York 1998.

Bioinformatika

- Informační technologie sloužící k získávání, uchovávání, analýze a distribuci informací týkajících se biomakromolekul



```
LRHLGITG--PVTLAVHDWGGMIGFGWALSHHAQVKRLVITNTAAAGTKFDKLTWLDVG--PVDLVVHDWGTI
LRHVGIDDVTPLTLAVHDWGGMIGFGWALAHAVQVRRLVMTNTAGTIKLERLTWLDVG---PVDLVVHDWGTI
LTWLDVG--PVDLVVHDWGGAIGMGWAVRHPDLVRRIVVLNTAAGT-KLDRLTWLDVG---PVDLVVHDWGTL
```

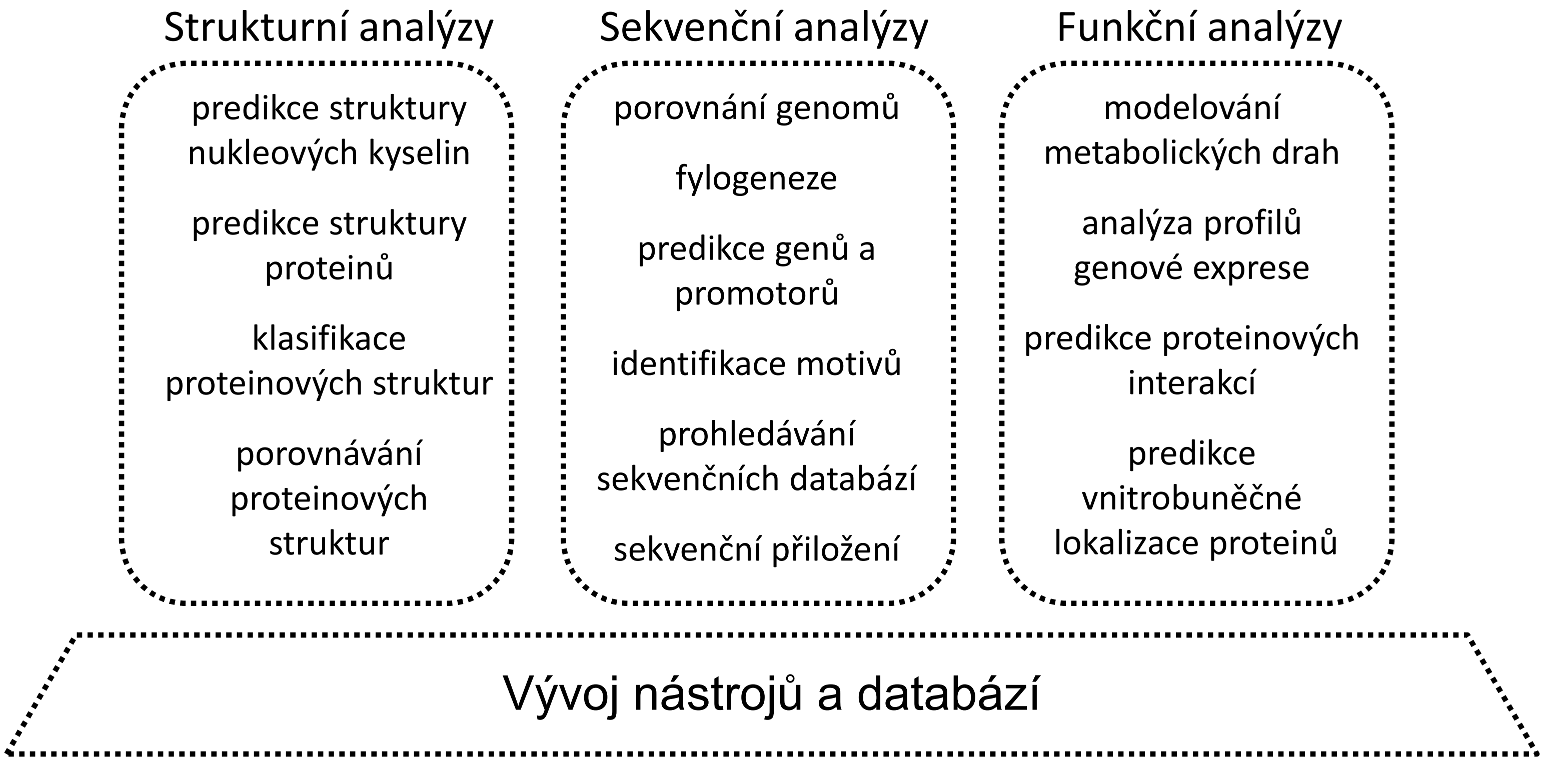
Bioinformatika

- Informační technologie sloužící k získávání, uchovávání, analýze a distribuci informací týkajících se biomakromolekul

Vývoj nástrojů a databází

Bioinformatika

- Informační technologie sloužící k získávání, uchovávání, analýze a distribuci informací týkajících se biomakromolekul



Strukturní analýzy

- predikce struktury nukleových kyselin
- predikce struktury proteinů
- klasifikace proteinových struktur
- porovnávání proteinových struktur

Sekvenční analýzy

- porovnání genomů
- fylogeneze
- predikce genů a promotorů
- identifikace motivů
- prohledávání sekvenčních databází
- sekvenční přiložení

Funkční analýzy

- modelování metabolických drah
- analýza profilů genové exprese
- predikce proteinových interakcí
- predikce vnitrobuněčné lokalizace proteinů

Vývoj nástrojů a databází

Bioinformatika

- Informační technologie sloužící k získávání, uchovávání, analýze a distribuci informací týkajících se biomakromolekul

Strukturní analýzy

- predikce struktury nukleových kyselin
- predikce struktury proteinů
- klasifikace proteinových struktur
- porovnávání proteinových struktur

Sekvenční analýzy

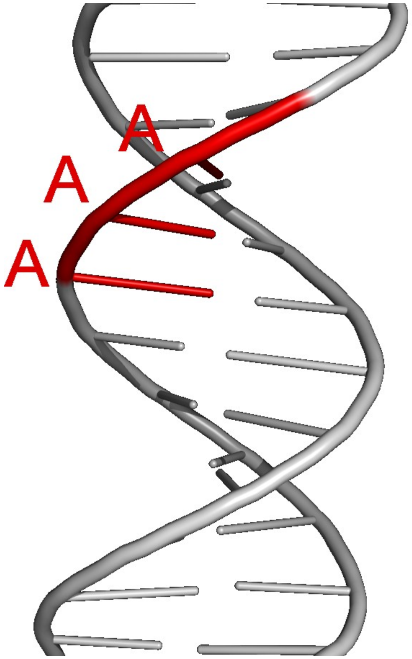
- porovnání genomů
- fylogeneze
- predikce genů a promotorů
- identifikace motivů
- prohledávání sekvenčních databází
- sekvenční přiložení

Funkční analýzy

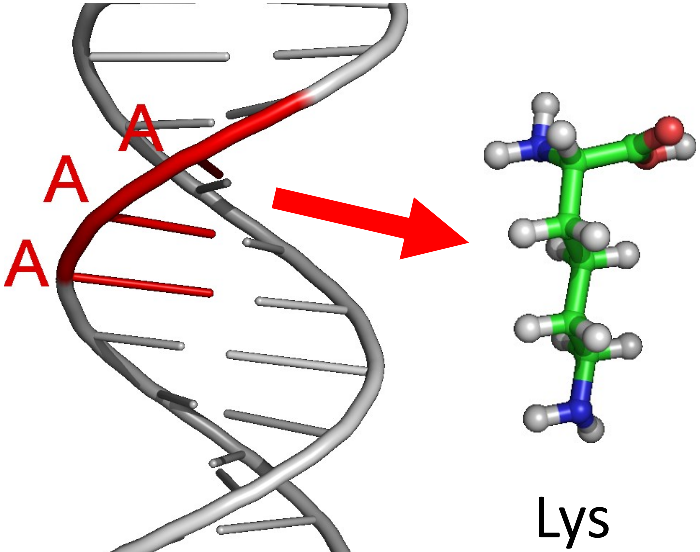
- modelování metabolických drah
- analýza profilů genové exprese
- predikce proteinových interakcí
- predikce vnitrobuněčné lokalizace proteinů

Vývoj nástrojů a databází

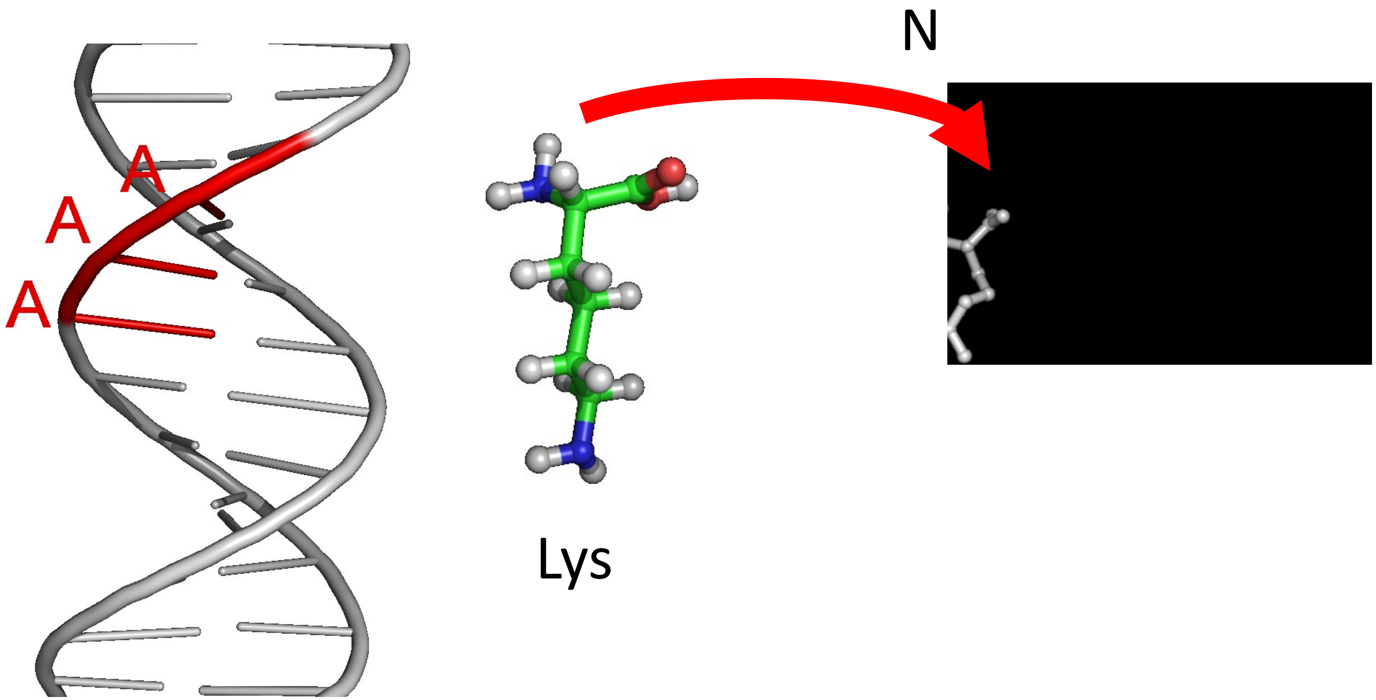
Nukleové kyseliny



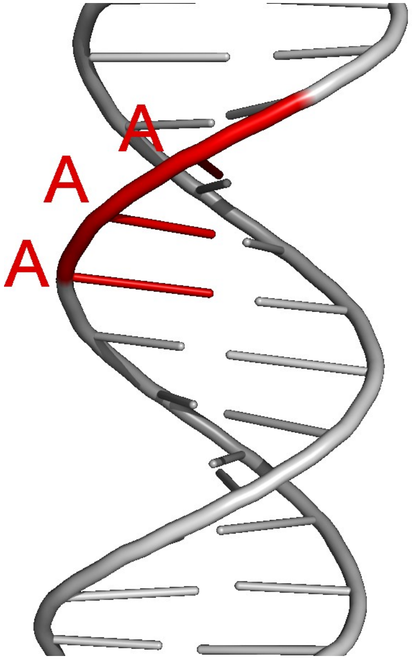
Nukleové kyseliny



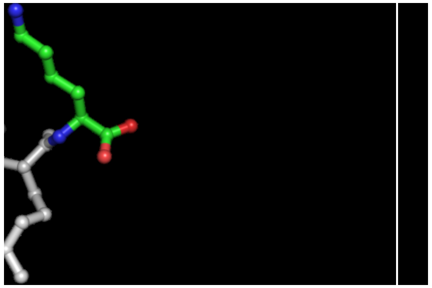
Nukleové kyseliny



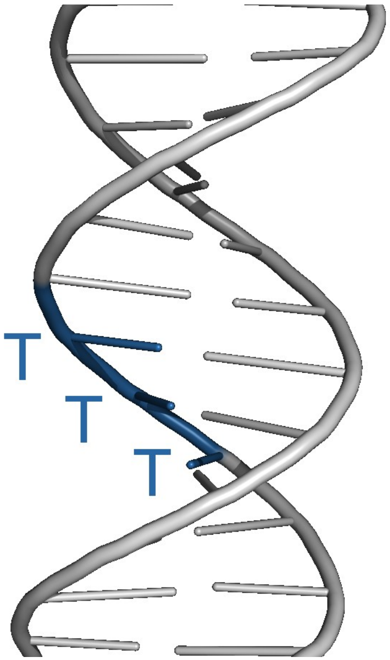
Nukleové kyseliny



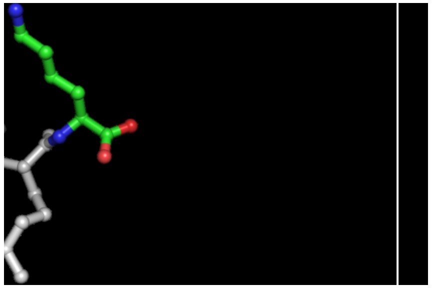
N-Lys



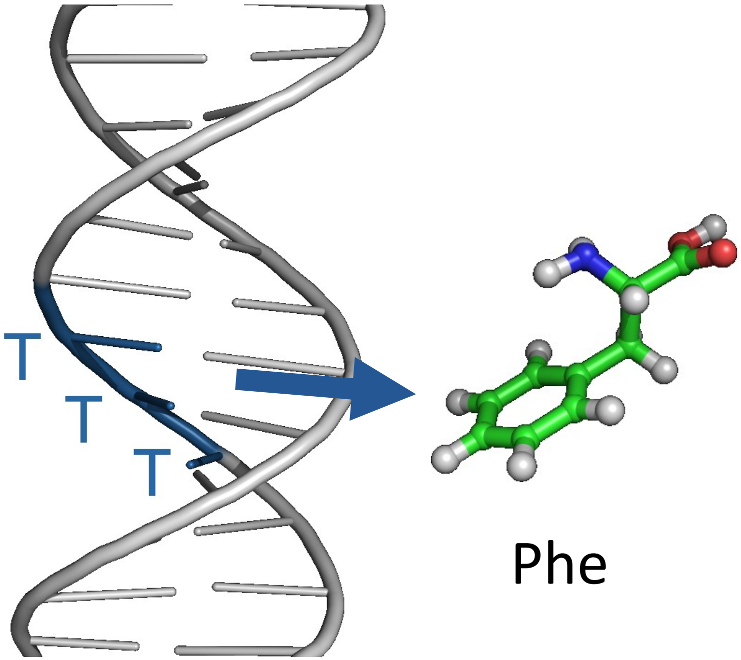
Nukleové kyseliny



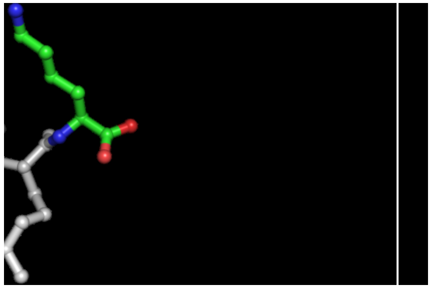
N-Lys



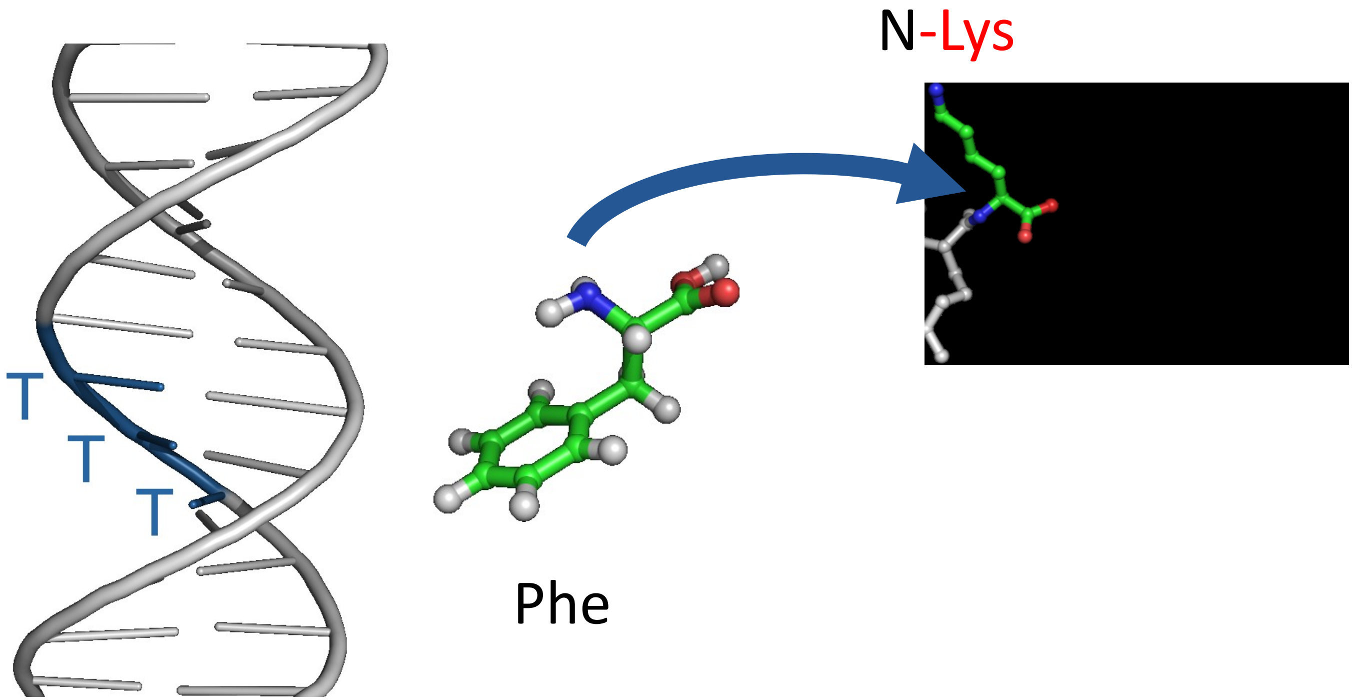
Nukleové kyseliny



N-Lys



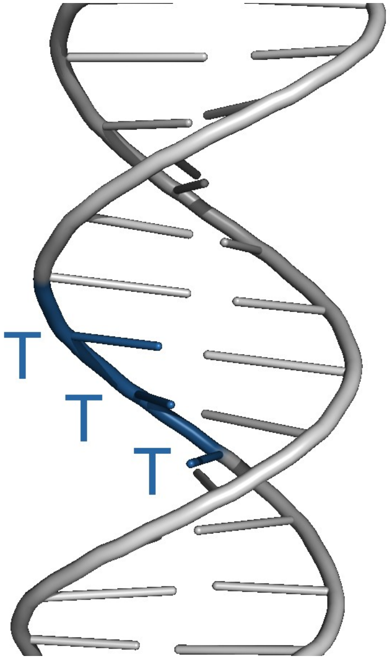
Nukleové kyseliny



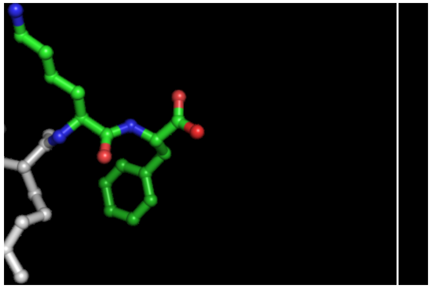
N-Lys

Phe

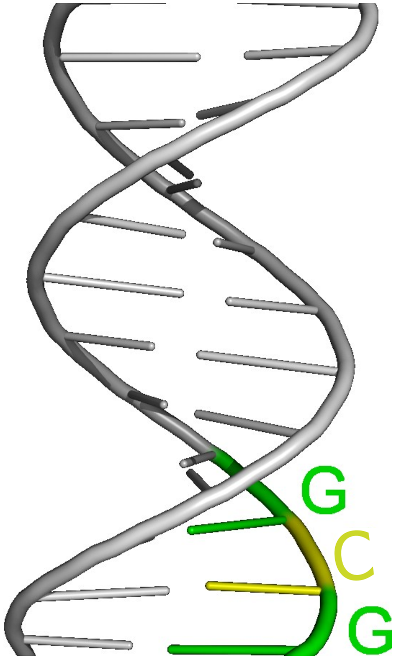
Nukleové kyseliny



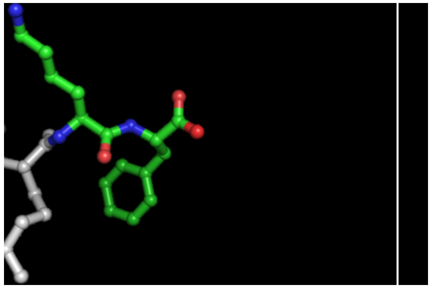
N-Lys-Phe



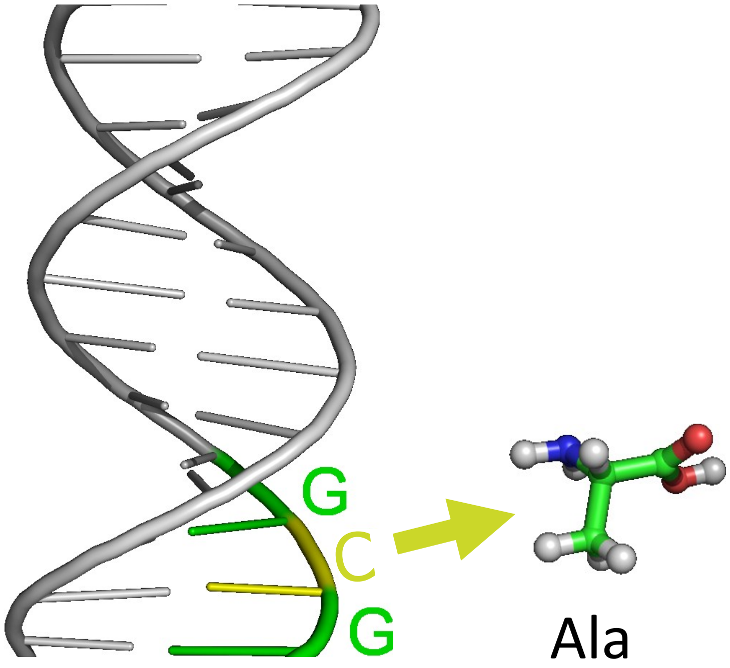
Nukleové kyseliny



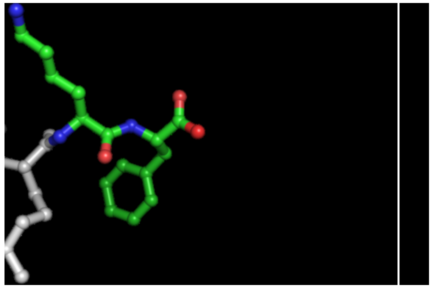
N-Lys-Phe



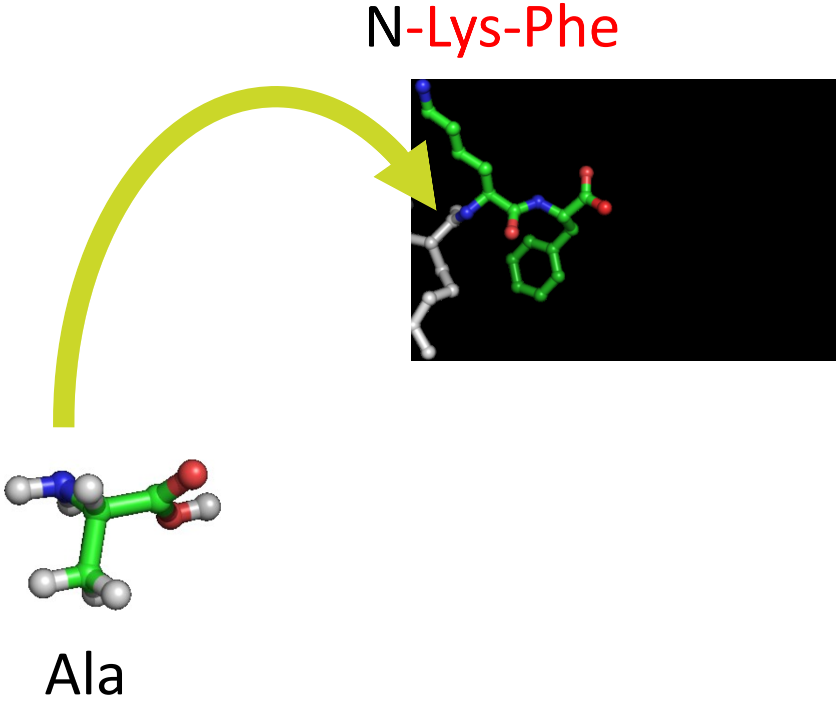
Nukleové kyseliny



N-Lys-Phe



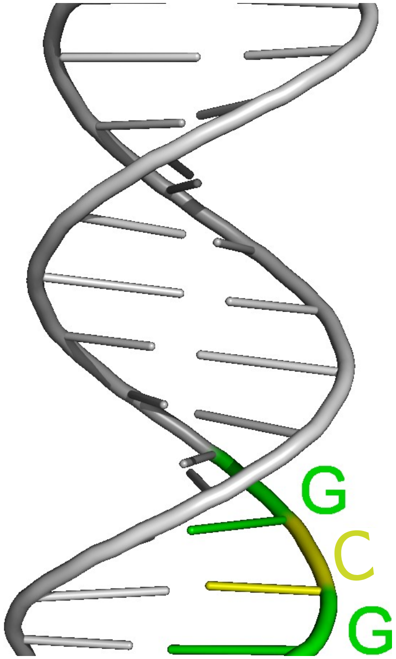
Nukleové kyseliny



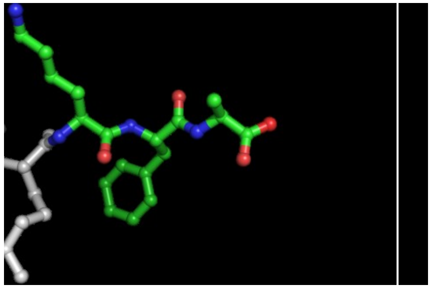
N-Lys-Phe

Ala

Nukleové kyseliny

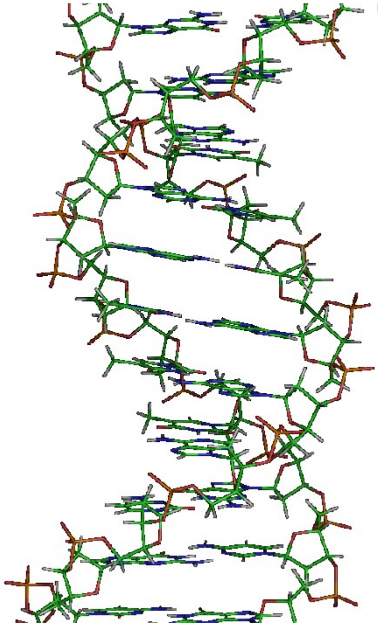


N-Lys-Phe-Ala

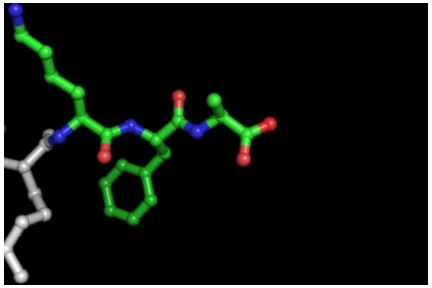


Nukleové kyseliny

5’-NCG-AAA-TTT-GCG-3’

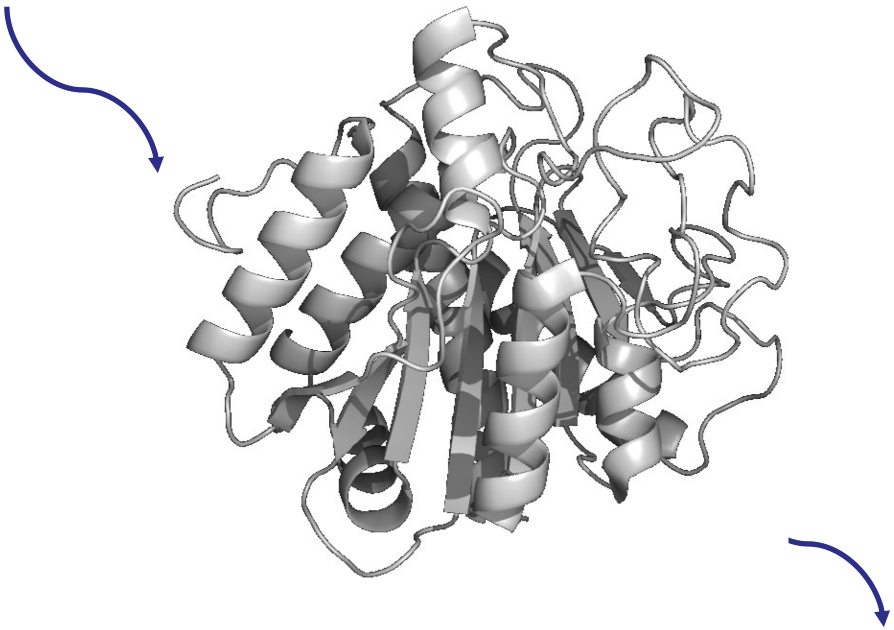


N-Lys-Phe-Ala



Proteiny

MSLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCA
GLGRLIACDLIGMGDSDKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVV
HDWGSALGFDWARRHRERVQGIAYMEAIAMPIEWADFPEQDRDLFQAFRS
QAGEELVLQD

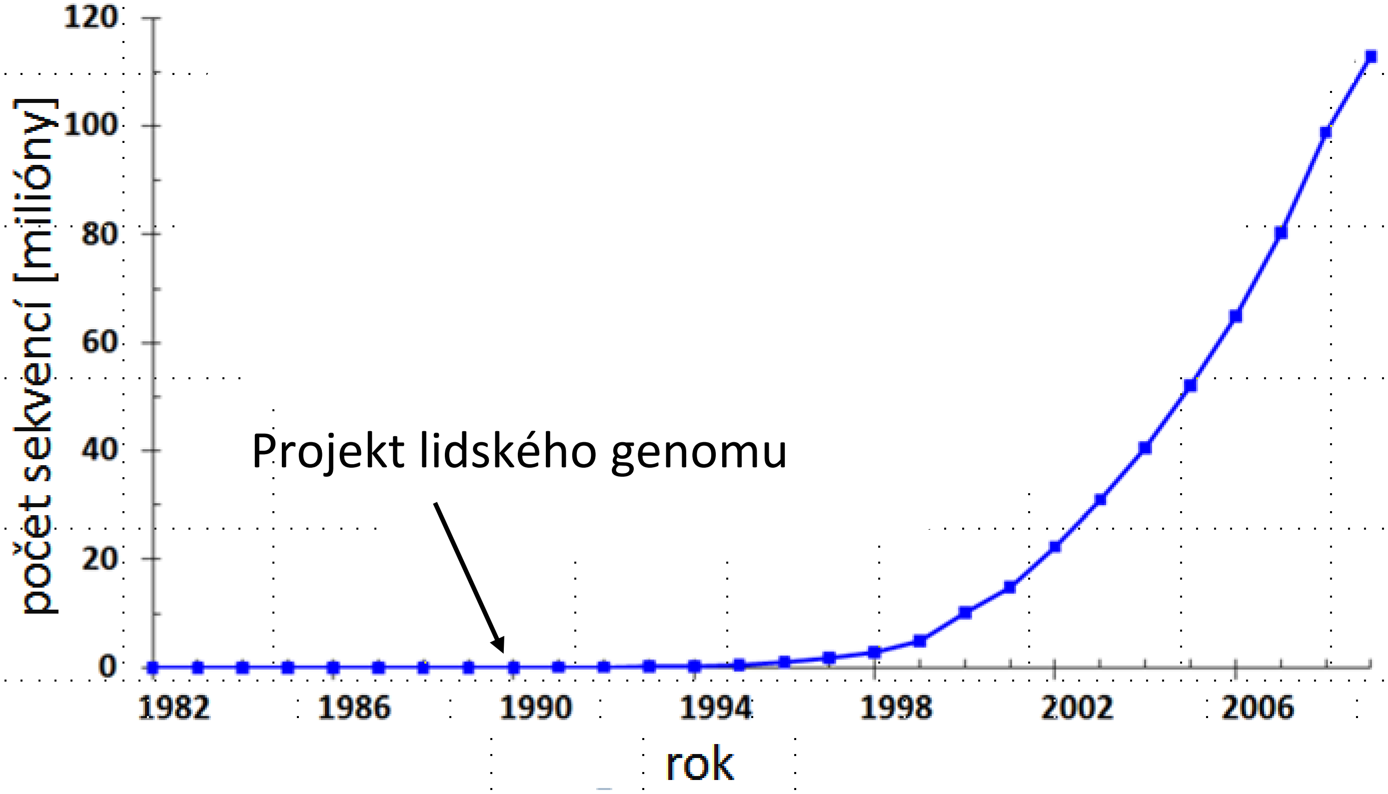


funkce

Historie bioinformatiky

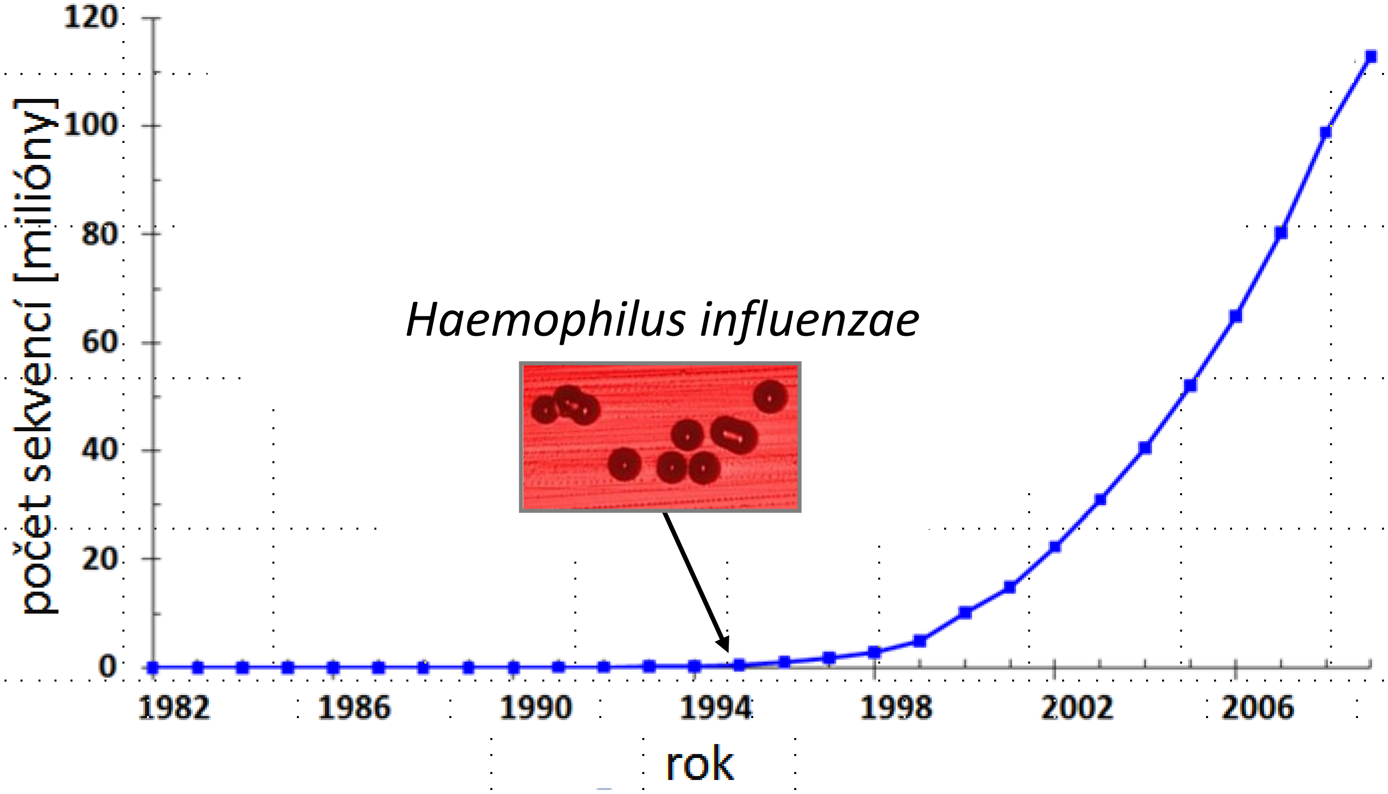
- Revoluce

Historie bioinformatiky



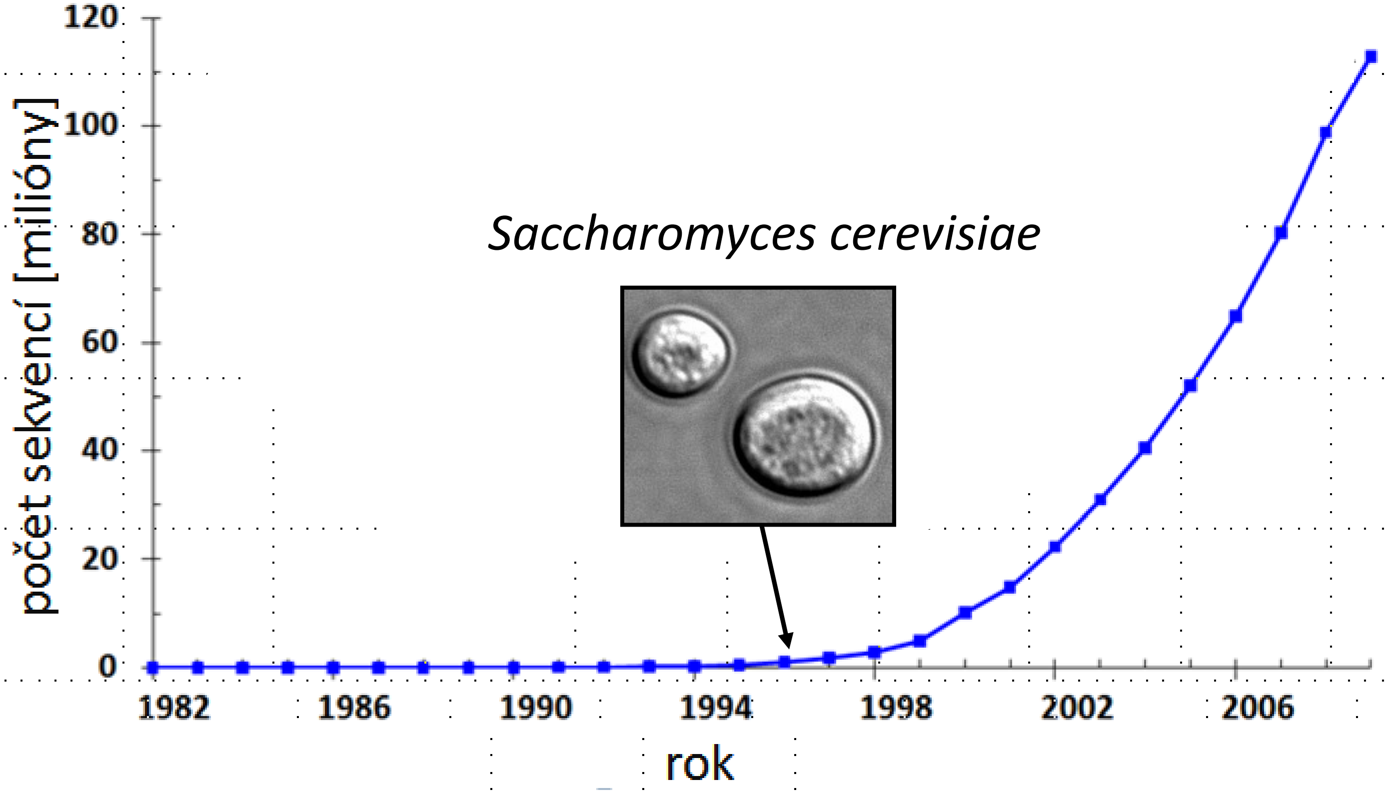
Projekt lidského genomu

Historie bioinformatiky

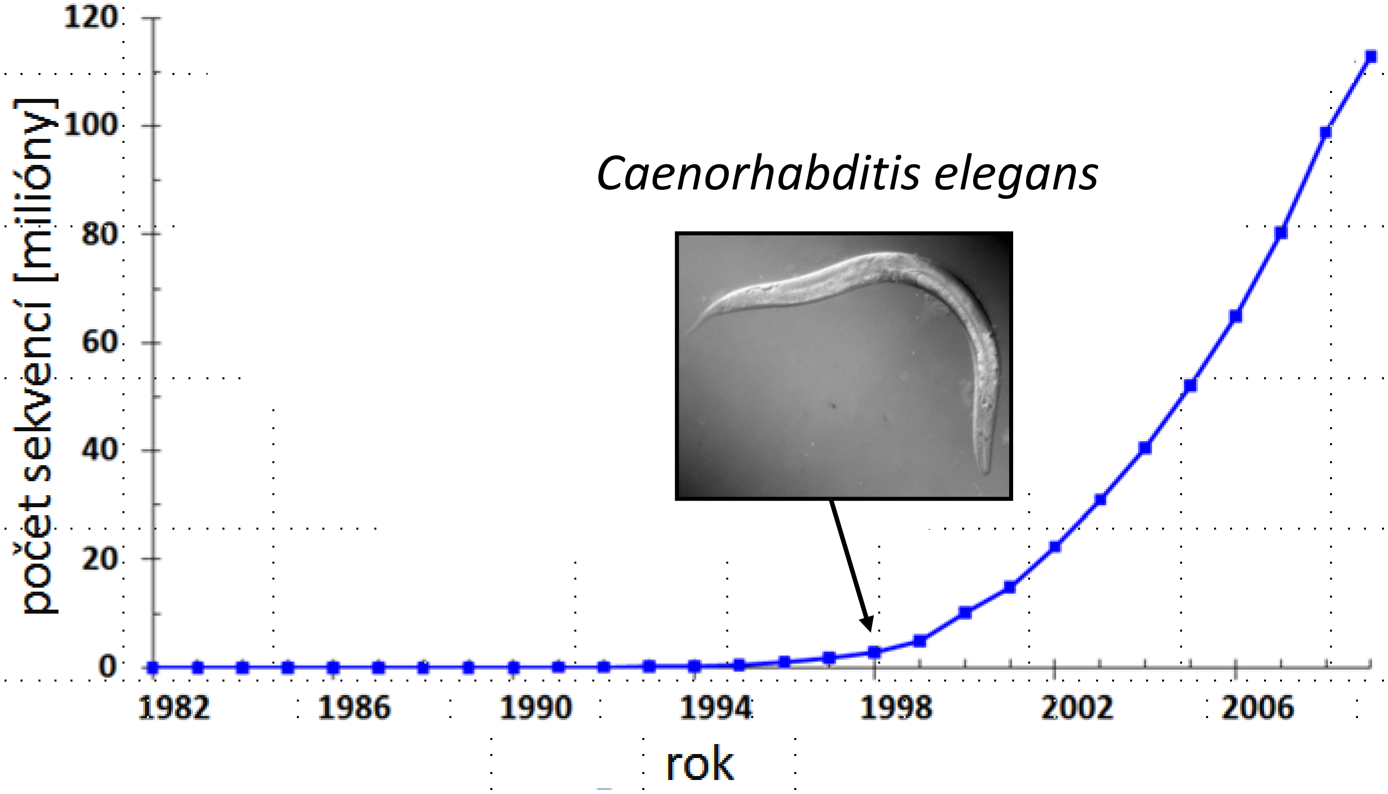


Haemophilus influenzae

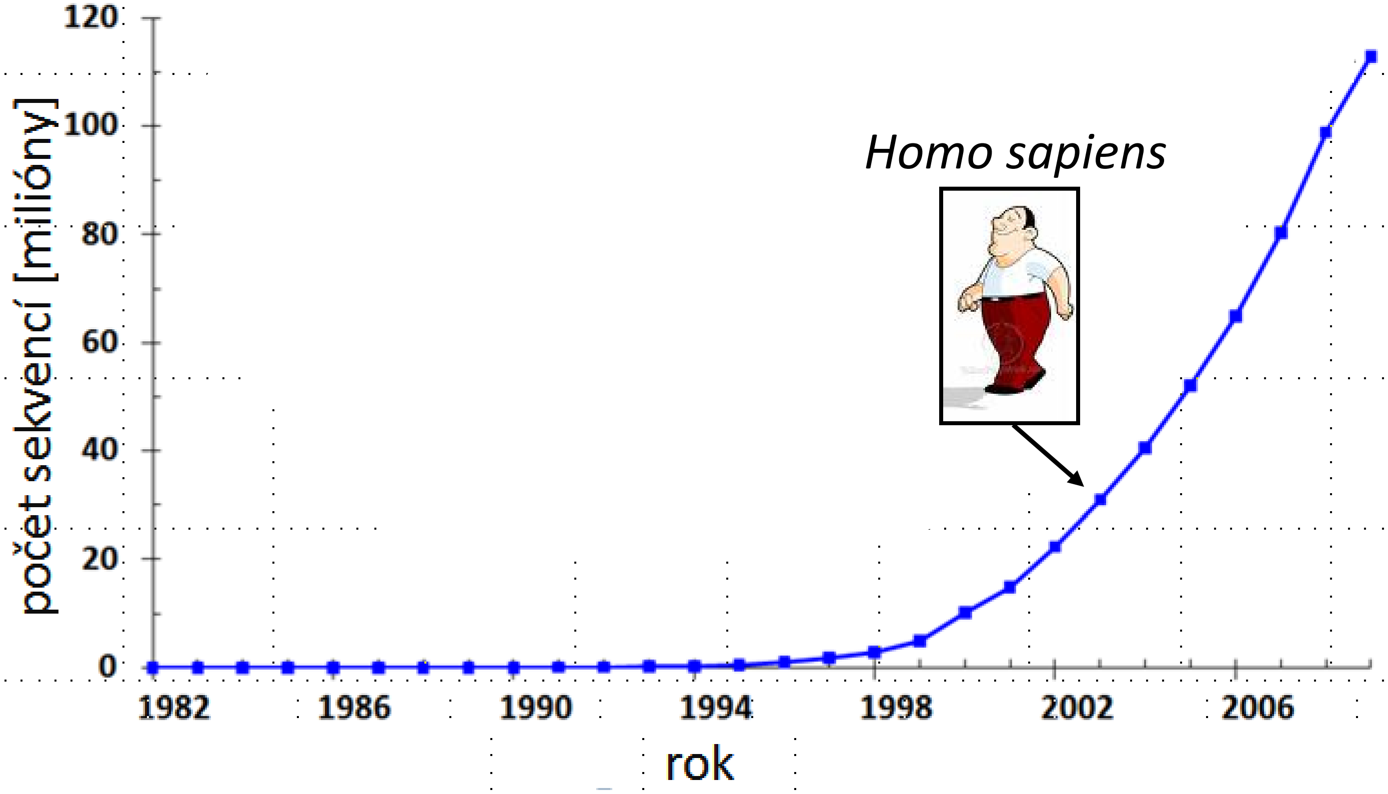
Historie bioinformatiky



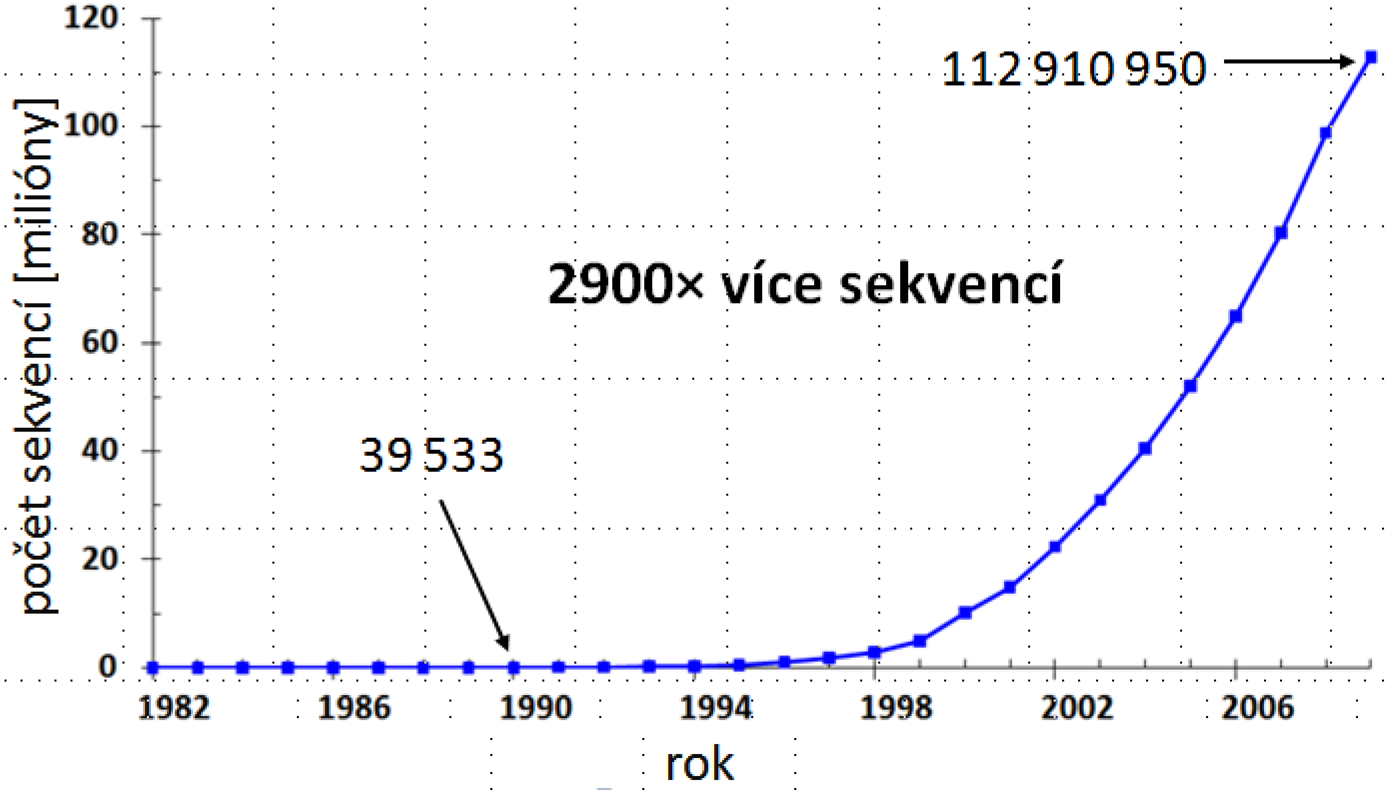
Historie bioinformatiky



Historie bioinformatiky



Historie bioinformatiky



Bioinformatické instituce

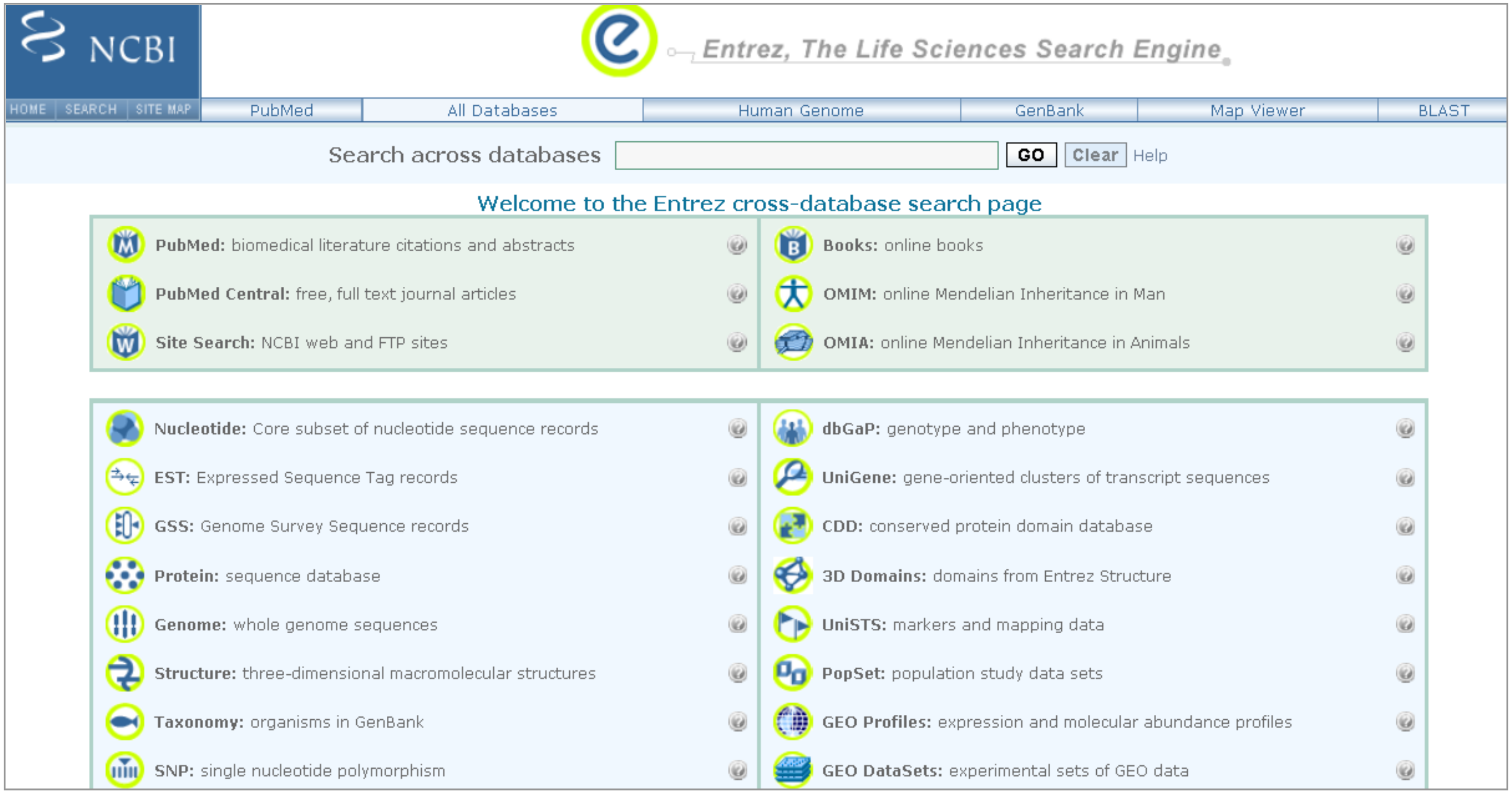
- ❑ National Center for Biotechnology Information (NCBI)
- ❑ European Bioinformatics Institute (EBI)

Bioinformatické instituce

- National Center for Biotechnology Information (NCBI)
    - oddělení National Library of Medicine při National Institutes of Health v USA
    - Poskytuje Databáze GenBank, PubMed, OMIM, Genome dbSNP, ...
    - Informace dostupné přes vyhledávací systém Entrez

Bioinformatické instituce

- National Center for Biotechnology Information (NCBI)



Entrez, The Life Sciences Search Engine

Search across databases

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts
PubMed Central: free, full text journal articles
Site Search: NCBI web and FTP sites

Books: online books
OMIM: online Mendelian Inheritance in Man
OMIA: online Mendelian Inheritance in Animals

Nucleotide: Core subset of nucleotide sequence records
EST: Expressed Sequence Tag records
GSS: Genome Survey Sequence records
Protein: sequence database
Genome: whole genome sequences
Structure: three-dimensional macromolecular structures
Taxonomy: organisms in GenBank
SNP: single nucleotide polymorphism

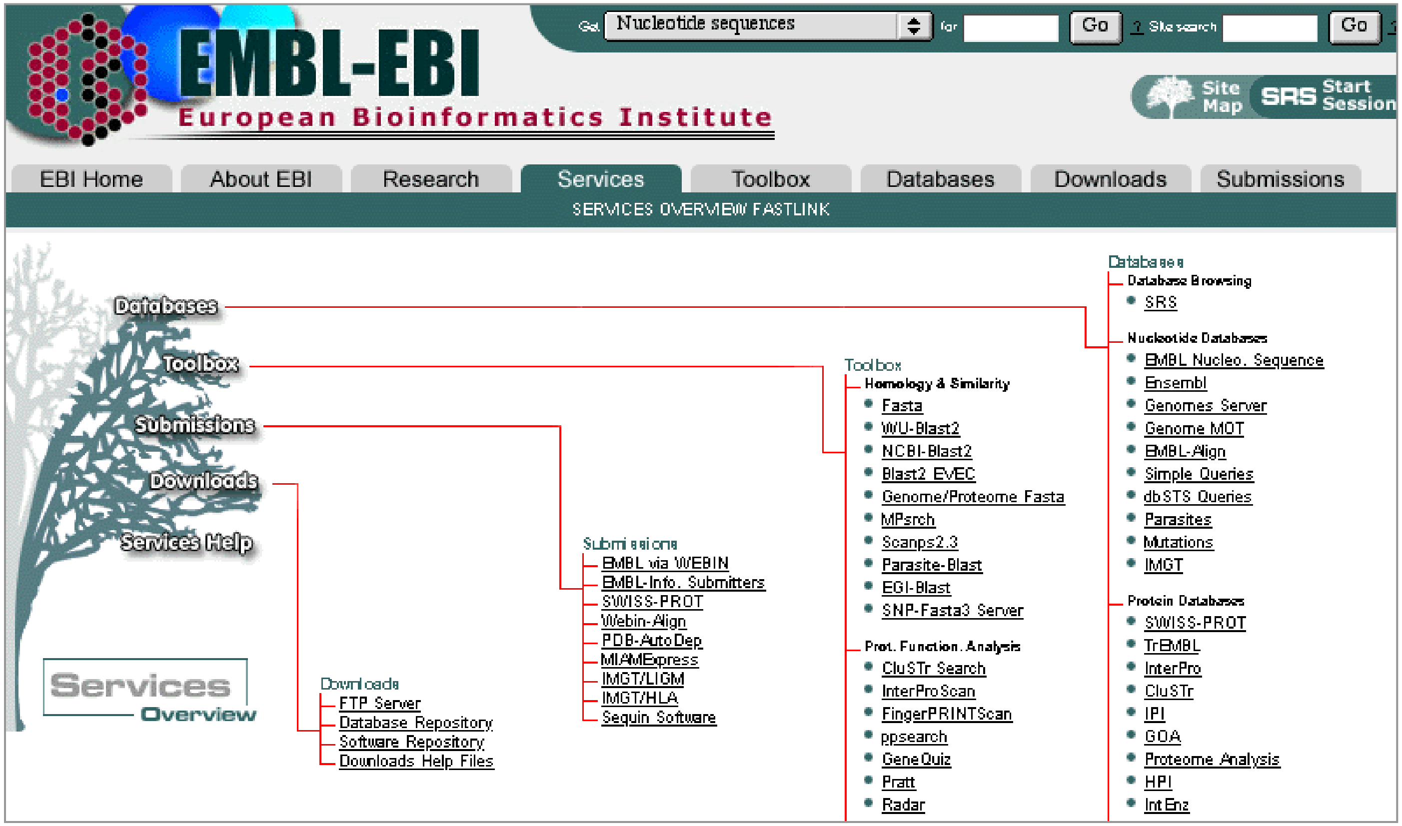
dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
3D Domains: domains from Entrez Structure
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance profiles
GEO DataSets: experimental sets of GEO data

Bioinformatické instituce

- European Bioinformatics Institute (EBI)
    - Součást European Molecular Biology Laboratory (EMBL), Wellcome Trust Genome Campus ve Velké Británii
    - Poskytuje databáze EMBL-Bank, UniProt, Ensembl, InterPro, ...
    - Informace dostupné přes vyhledávací systém SRS

Bioinformatické instituce

- European Bioinformatics Institute (EBI)



Downloads
- FTP Server
- Database Repository
- Software Repository
- Downloads Help Files

Submissions
- EMBL via WEBIN
- EMBL-Info. Submitters
- SWISS-PROT
- Webin-Align
- PDB-AutoDep
- MIAMExpress
- IMGT/LIGM
- IMGT/HLA
- Sequin Software

Toolbox

Homology & Similarity
- Fasta
- WU-Blast2
- NCBI-Blast2
- Blast2 EVEC
- Genome/Proteome Fasta
- MPsrch
- Scanps2.3
- Parasite-Blast
- EGI-Blast
- SNP-Fasta3 Server

Prot. Function. Analysis
- CluSTr Search
- InterProScan
- FingerPRINTScan
- ppsearch
- GeneQuiz
- Pratt
- Radar

Databases

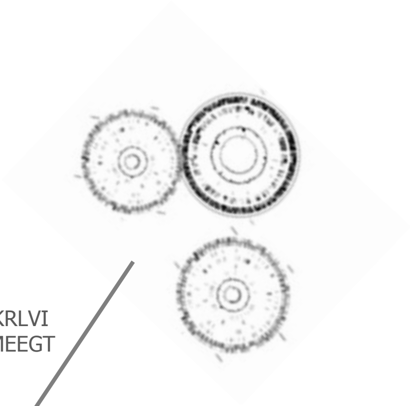
Database Browsing
- SRS

Nucleotide Databases
- EMBL Nucleo. Sequence
- Ensembl
- Genomes Server
- Genome MOT
- EMBL-Align
- Simple Queries
- dbSTS Queries
- Parasites
- Mutations
- IMGT

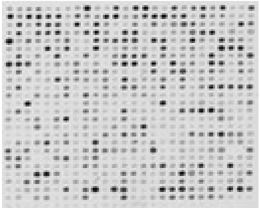
Protein Databases
- SWISS-PROT
- TrEMBL
- InterPro
- CluSTr
- IPI
- GOA
- Proteome Analysis
- HPI
- IntEnz

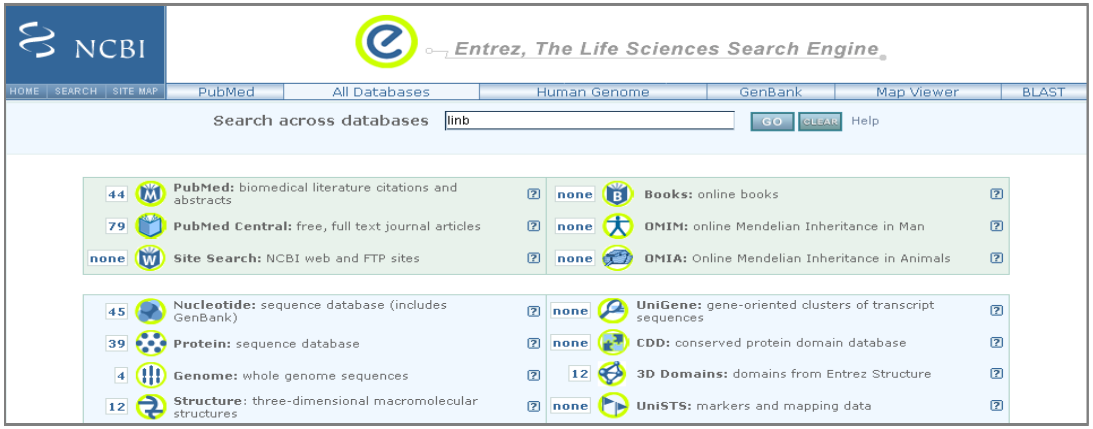
Bioinformatické databáze

- ❑ Bibliografie
- ❑ Nukleotidové sekvence
- ❑ Proteinové sekvence
- ❑ Proteinové struktury
- ❑ Genomy



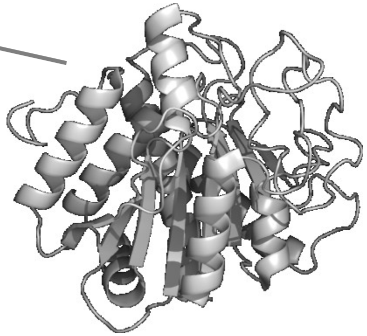
LRHLGITGPVTLAVHDWGGMIGFGWALSHHAQVKRLVI
TNTAAAGTKFDKLTWLDVGPVDLVVHDWGTISRMEEGT
WYLKLIRTTVWHQAIVLAEIGTVVCKTQENPA





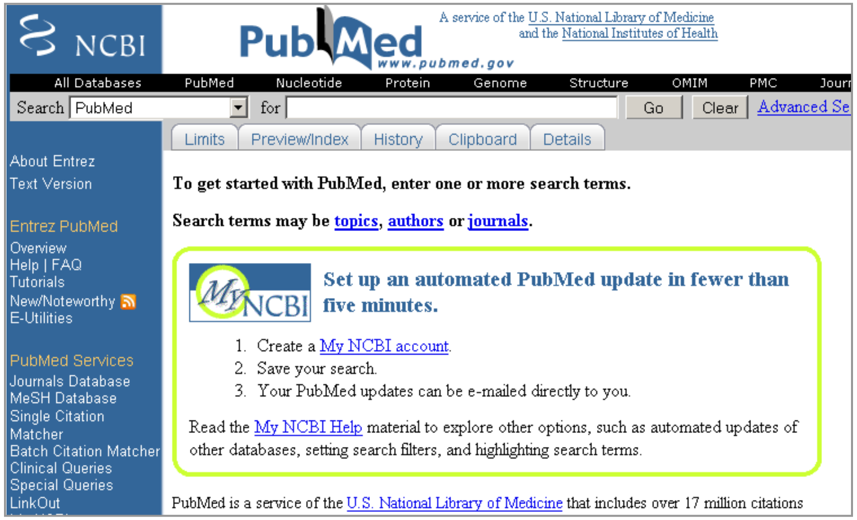
Search across databases: linb

44	**PubMed:** biomedical literature citations and abstracts	none	**Books:** online books
79	**PubMed Central:** free, full text journal articles	none	**OMIM:** online Mendelian Inheritance in Man
none	**Site Search:** NCBI web and FTP sites	none	**OMIA:** Online Mendelian Inheritance in Animals
45	**Nucleotide:** sequence database (includes GenBank)	none	**UniGene:** gene-oriented clusters of transcript sequences
39	**Protein:** sequence database	none	**CDD:** conserved protein domain database
4	**Genome:** whole genome sequences	12	**3D Domains:** domains from Entrez Structure
12	**Structure**: three-dimensional macromolecular structures	none	**UniSTS:** markers and mapping data



Bibliografické databáze

- PubMed
- Web of Science



A service of the U.S. National Library of Medicine and the National Institutes of Health

Search PubMed for

To get started with PubMed, enter one or more search terms.

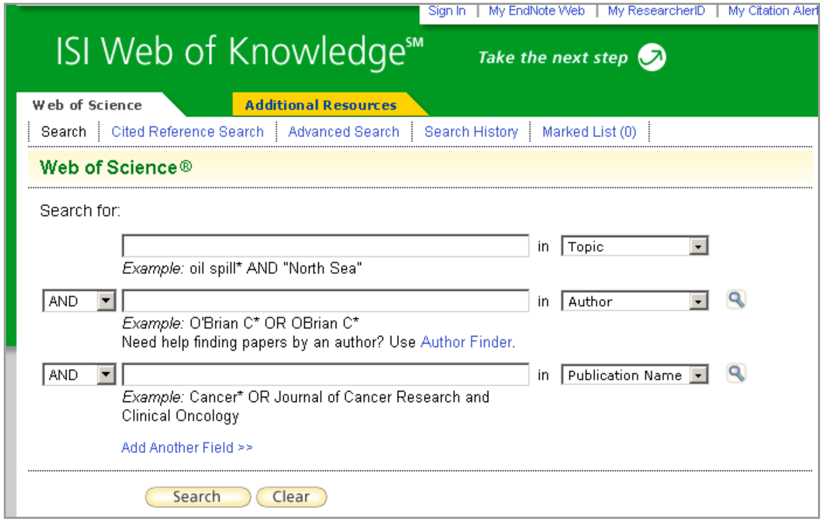
Search terms may be topics, authors or journals.

Set up an automated PubMed update in fewer than five minutes.

1. Create a My NCBI account.
2. Save your search.
3. Your PubMed updates can be e-mailed directly to you.

Read the My NCBI Help material to explore other options, such as automated updates of other databases, setting search filters, and highlighting search terms.

PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations



Web of Science®

Search for:

in Topic

Example: oil spill* AND "North Sea"

AND in Author

Example: O'Brian C* OR OBrian C*

Need help finding papers by an author? Use Author Finder.

AND in Publication Name

Example: Cancer* OR Journal of Cancer Research and Clinical Oncology

Bibliografické databáze

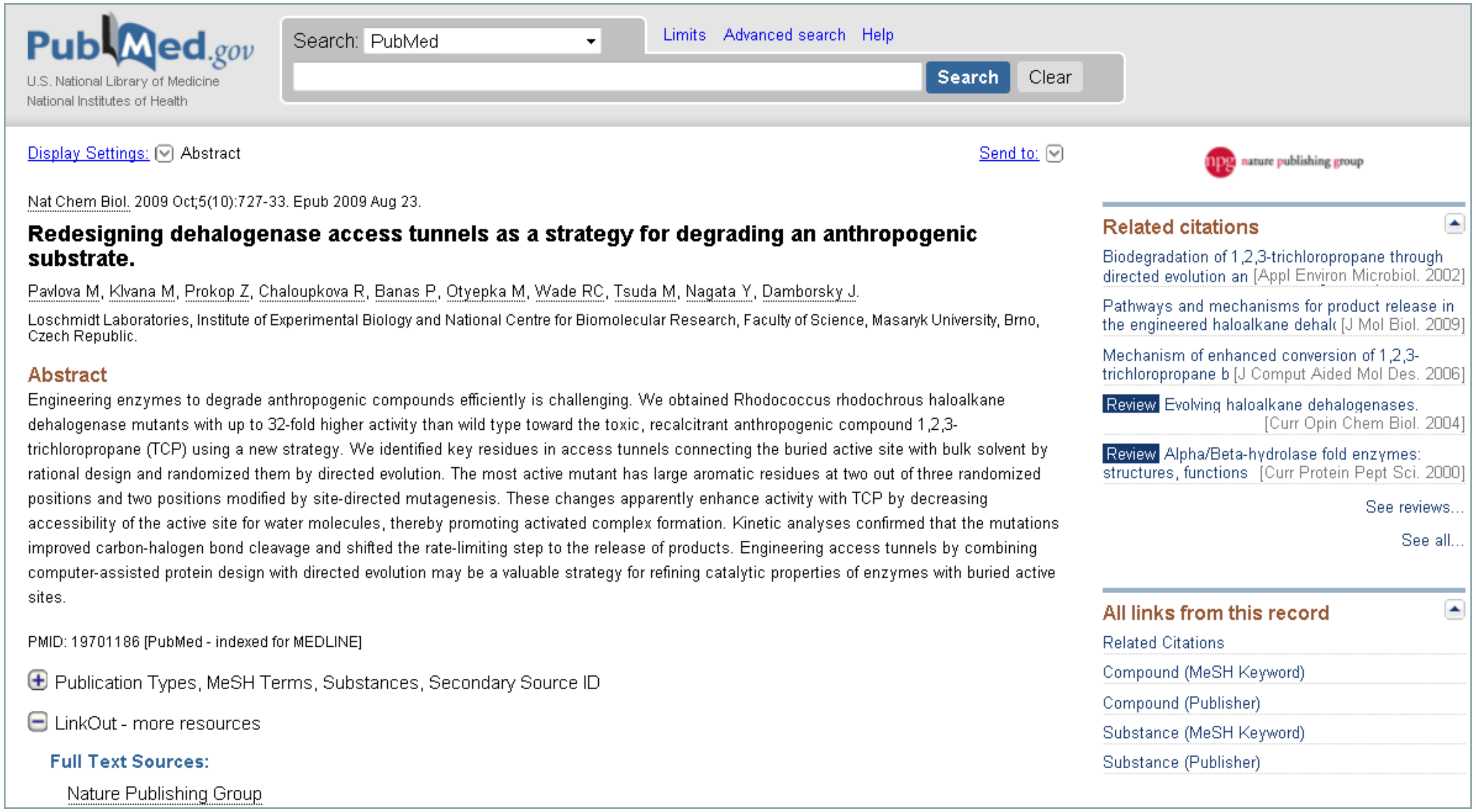
- PubMed
  - Provozováno National Library of Medicine
  - Obsahuje více než 22 milionů citací biomedicínské literatury
  - Integruje MEDLINE, časopisy z oblasti živých věd a online knihy
  - Prohledávání možné přes Entrez nebo DBGET
  - Obsahuje kromě abstraktů odkazy na plné texty dostupné přes PubMed Central nebo stránky nakladatelství

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

NCBI

Bibliografické databáze

- PubMed



Nat Chem Biol. 2009 Oct;5(10):727-33. Epub 2009 Aug 23.

Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate.

Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, Wade RC, Tsuda M, Nagata Y, Damborsky J.

Loschmidt Laboratories, Institute of Experimental Biology and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic.

Abstract

Engineering enzymes to degrade anthropogenic compounds efficiently is challenging. We obtained Rhodococcus rhodochrous haloalkane dehalogenase mutants with up to 32-fold higher activity than wild type toward the toxic, recalcitrant anthropogenic compound 1,2,3-trichloropropane (TCP) using a new strategy. We identified key residues in access tunnels connecting the buried active site with bulk solvent by rational design and randomized them by directed evolution. The most active mutant has large aromatic residues at two out of three randomized positions and two positions modified by site-directed mutagenesis. These changes apparently enhance activity with TCP by decreasing accessibility of the active site for water molecules, thereby promoting activated complex formation. Kinetic analyses confirmed that the mutations improved carbon-halogen bond cleavage and shifted the rate-limiting step to the release of products. Engineering access tunnels by combining computer-assisted protein design with directed evolution may be a valuable strategy for refining catalytic properties of enzymes with buried active sites.

PMID: 19701186 [PubMed - indexed for MEDLINE]

⊞ Publication Types, MeSH Terms, Substances, Secondary Source ID

⊟ LinkOut - more resources

Full Text Sources:

Nature Publishing Group

Related citations

Biodegradation of 1,2,3-trichloropropane through directed evolution an [Appl Environ Microbiol. 2002]

Pathways and mechanisms for product release in the engineered haloalkane dehalc [J Mol Biol. 2009]

Mechanism of enhanced conversion of 1,2,3-trichloropropane b [J Comput Aided Mol Des. 2006]

Review Evolving haloalkane dehalogenases. [Curr Opin Chem Biol. 2004]

Review Alpha/Beta-hydrolase fold enzymes: structures, functions [Curr Protein Pept Sci. 2000]

See reviews...

See all...

All links from this record

Related Citations

Compound (MeSH Keyword)

Compound (Publisher)

Substance (MeSH Keyword)

Substance (Publisher)

Bibliografické databáze

- Web of Science
  - Komerční databáze
  - Součást ISI Web of Knowledge
  - Používá se pro zjištění citovanosti a impaktního faktoru časopisů

Bibliografické databáze

- Web of Science

Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate

Author(s): Pavlova M (Pavlova, Martina)[1,2], Klvana M (Klvana, Martin)[1,2], Prokop Z (Prokop, Zbynek)[1,2], Chaloupkova R (Chaloupkova, Radka)[1,2], Banas P (Banas, Pavel)[3,4], Otyepka M (Otyepka, Michal)[3,4], Wade RC (Wade, Rebecca C.)[5], Tsuda M (Tsuda, Masataka)[6], Nagata Y (Nagata, Yuji)[6], Damborsky J (Damborsky, Jiri)[1,2]

Source: NATURE CHEMICAL BIOLOGY **Volume:** 5 **Issue:** 10 **Pages:** 727-733 **Published:** OCT 2009

Times Cited: 6 **References:** 50

Abstract: Engineering enzymes to degrade anthropogenic compounds efficiently is challenging. We obtained Rhodococcus rhodochrous haloalkane dehalogenase mutants with up to 32-fold higher activity than wild type toward the toxic, recalcitrant anthropogenic compound 1,2,3-trichloropropane (TCP) using a new strategy. We identified key residues in access tunnels connecting the buried active site with bulk solvent by rational design and randomized them by directed evolution. The most active mutant has large aromatic residues at two out of three randomized positions and two positions modified by site-directed mutagenesis. These changes apparently enhance activity with TCP by decreasing accessibility of the active site for water molecules, thereby promoting activated complex formation. Kinetic analyses confirmed that the mutations improved carbon-halogen bond cleavage and shifted the rate-limiting step to the release of products. Engineering access tunnels by combining computer-assisted protein design with directed evolution may be a valuable strategy for refining catalytic properties of enzymes with buried active sites.

Document Type: Article

Language: English

KeyWords Plus: SPHINGOMONAS-PAUCIMOBILIS UT26; HALOALKANE DEHALOGENASE; DIRECTED EVOLUTION; CYTOCHROME P450S; HETEROLOGOUS EXPRESSION; XENOBIOTIC COMPOUNDS; CATALYTIC MECHANISM; ESCHERICHIA-COLI; ENZYME; SPECIFICITY

Reprint Address: Damborsky, J (reprint author), Masaryk Univ, Fac Sci, Loschmidt Labs, Inst Expt Biol, CS-61137 Brno, Czech Republic

Cited by: 6

This article has been cited 6 times (from Web of Science).

Kourist R, Jochens H, Bartsch S, et al. The alpha/beta-Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering CHEMBIOCHEM 11 12 1635-1643 AUG 16 2010

Stsiapanava A, Dohnalek J, Gavira JA, et al. Atomic resolution studies of haloalkane dehalogenases DhaA04, DhaA14 and DhaA15 with engineered access tunnels ACTA CRYSTALLOGRAPHICA SECTION D-BIOLOGICAL CRYSTALLOGRAPHY 66 962-969 Part 9 SEP 2010

Brouk M, Derry NL, Shainsky J, et al. The influence of key residues in the tunnel entrance and the active site on activity and selectivity of toluene-4-monooxygenase JOURNAL OF MOLECULAR CATALYSIS B-ENZYMATIC 66 1-2 72-80 SEP 2010

[view all 6 citing articles]

Related Records:

Databáze nukleotidových sekvencí

- GenBank
- EMBL-Bank
- DDBJ

- Anotované kolekce veřejně dostupných nukleotidových sekvencí
- Data získaná z genomových center a odborných pracovišť
- Každodenní vzájemná synchronizace nových a aktualizovaných dat
- “Accession number” - jedinečný identifikátor záznamu, ve všech třech databázích

Databáze nukleotidových sekvencí

- GenBank
  - Založena v roce 1982, provozována NCBI
  - Přístupná prostřednictvím vyhledávacího systému Entrez nebo systému DBGET
  - Obsahuje více než 187.000.000 sekvencí (2015)
  - Nové sekvence možné vložit pomocí BankIt nebo Sequin

Databáze nukleotidových sekvencí

- ❑ EMBL-Bank
  - Založena v roce 1980, Provozována EBI
  - Přístupná prostřednictvím vyhledávacího systému SRS či DBGET
  - Obsahuje více než 608.000.000 sekvencí (2015)
  - Nové sekvence možné vložit pomocí Webin nebo Sequin

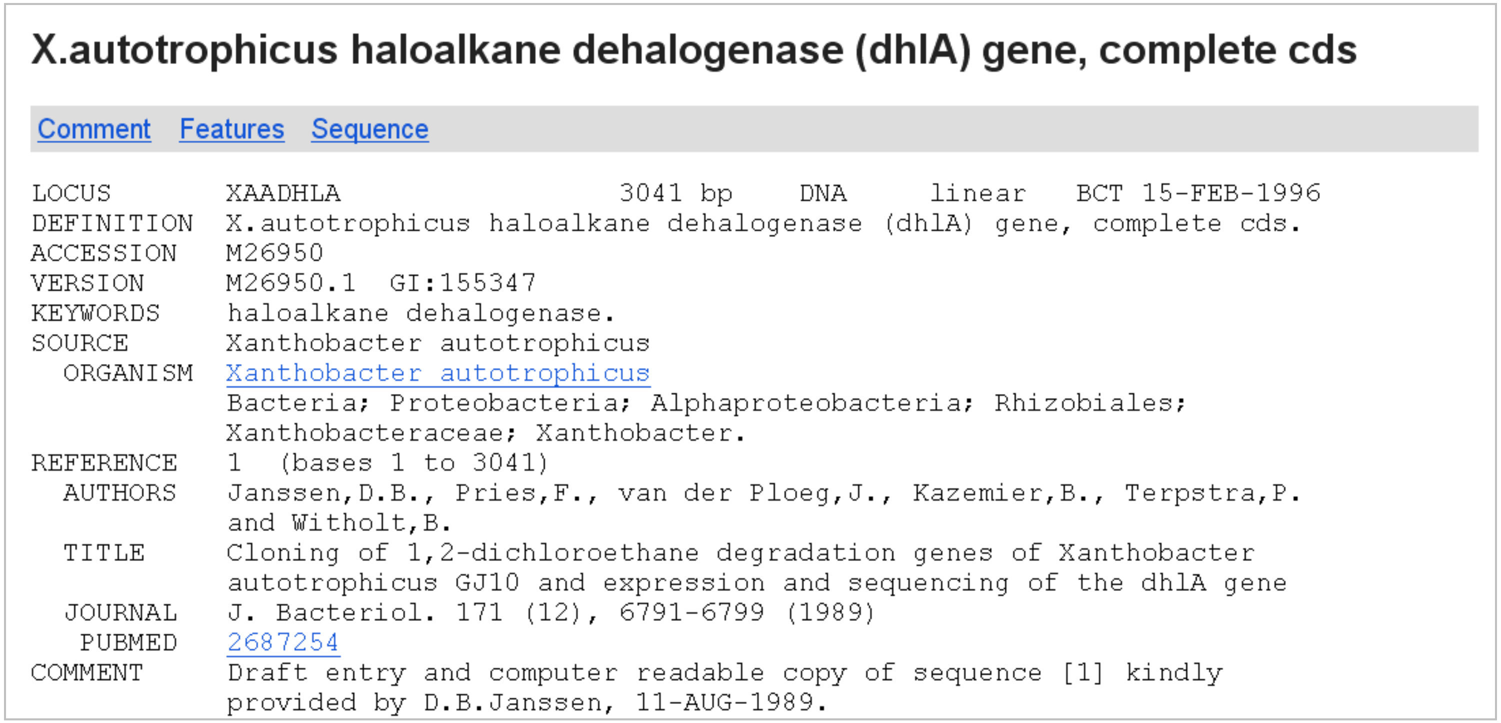
EMBL
NUCLEOTIDE
SEQUENCE
DATABASE

Databáze nukleotidových sekvencí

- DNA Data Bank of Japan (DDBJ)
  - Založena v roce 1984, provozována National Institute of Genetics
  - Obsahuje více než 253.000.000 sekvencí (2015)
  - Nové sekvence možné vložit pomocí Sakura nebo Sequin

Ukázka záznamu v GenBank

- Hlavička
  - Základní informace o záznamu
  - Lokus, definice, přístupový kód, klíčová slova, organizmus, reference,..



X.autotrophicus haloalkane dehalogenase (dhlA) gene, complete cds

```
LOCUS       XAADHLA                 3041 bp    DNA     linear   BCT 15-FEB-1996
DEFINITION  X.autotrophicus haloalkane dehalogenase (dhlA) gene, complete cds.
ACCESSION   M26950
VERSION     M26950.1  GI:155347
KEYWORDS    haloalkane dehalogenase.
SOURCE      Xanthobacter autotrophicus
  ORGANISM  Xanthobacter autotrophicus
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Xanthobacteraceae; Xanthobacter.
REFERENCE   1  (bases 1 to 3041)
  AUTHORS   Janssen,D.B., Pries,F., van der Ploeg,J., Kazemier,B., Terpstra,P.
            and Witholt,B.
  TITLE     Cloning of 1,2-dichloroethane degradation genes of Xanthobacter
            autotrophicus GJ10 and expression and sequencing of the dhlA gene
  JOURNAL   J. Bacteriol. 171 (12), 6791-6799 (1989)
   PUBMED   2687254
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            provided by D.B.Janssen, 11-AUG-1989.
```

Ukázka záznamu v GenBank

- Charakteristiky
  - Popis jednotlivých oblastí genu
  - Promotor, RBS (ribozóm vazebné místo), CDS (kódující sekvence), ...

```
     gene            918..1931
                     /gene="dhlA"
     promoter        918..946
                     /gene="dhlA"
                     /note="putative"
     promoter        945..974
                     /gene="dhlA"
                     /note="putative"
     RBS             986..998
                     /gene="dhlA"
     CDS             999..1931
                     /gene="dhlA"
                     /codon_start=1
                     /transl_table=11
                     /product="haloalkane dehalogenase"
                     /protein_id="AAA88691.1"
                     /db_xref="GI:155348"
                     /translation="MINAIRTPDQRFSNLDQYPFSPNYLDDLPGYPGLRAHYLDEGNS
                     DAEDVFLCLHGEPTWSYLYRKMIPVFAESGARVIAPDFFGFGKSDKPVDEEDYTFEFH
                     RNFLLALIERLDLRNITLVVQDWGGFLGLTLPMADPSRFKRLIIMNACLMTDPVTQPA
                     FSAFVTQPADGFTAWKYDLVTPSDLRLDQFMKRWAPTLTEAEASAYAAPFPDTSYQAG
```

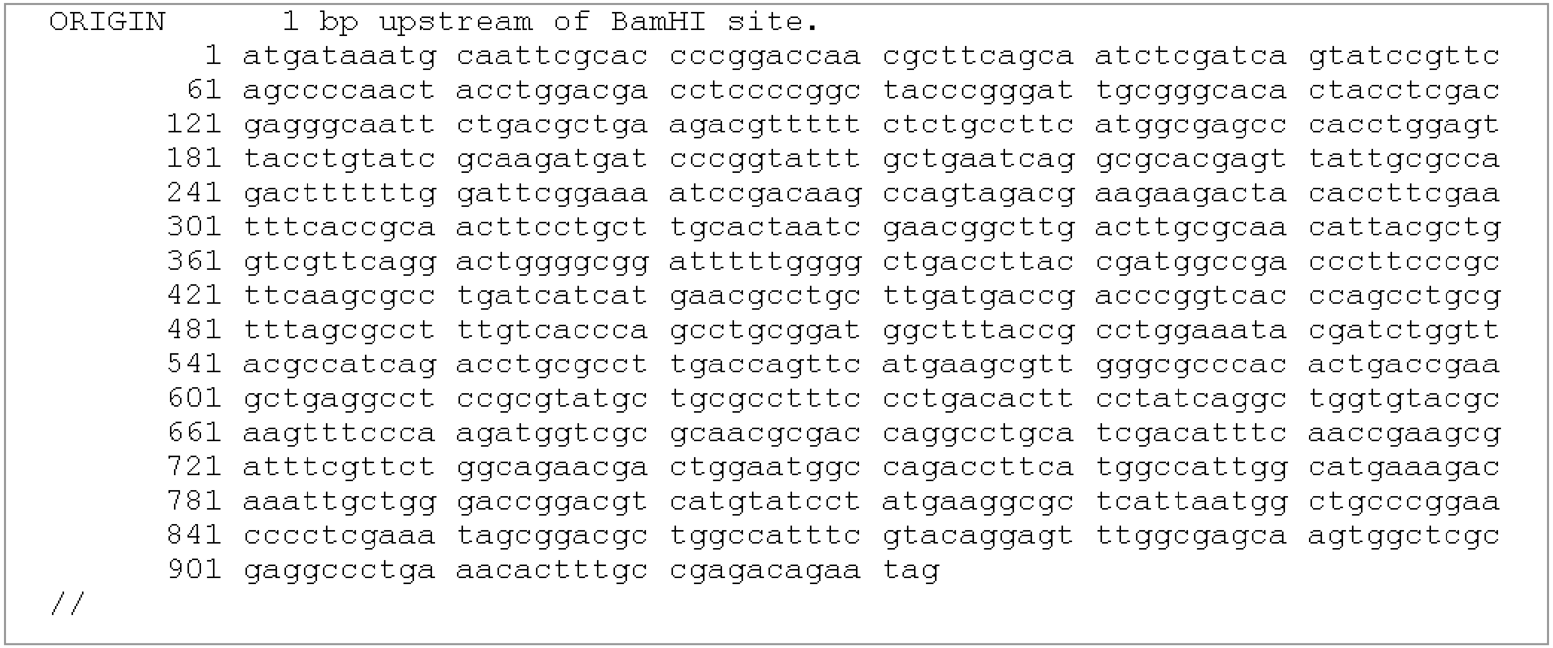
Ukázka záznamu v GenBank

- Charakteristiky

```
FEATURES             Location/Qualifiers
     source          1..3041
                     /organism="Xanthobacter autotrophicus"
                     /mol_type="genomic DNA"
                     /strain="GJ10"
                     /db_xref="taxon:280"
     CDS             complement(316..924)
                     /note="ORF 1; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein"
                     /protein_id="AAA88690.1"
                     /db_xref="GI:1197026"
                     /translation="MSTFFEPENGMKQNAKTERILDVALELLETEGEFGLTMRQVATQ
                     ADMSLSNVQYYFKSEDLLLVAMADRYFQRCLTTMAEHPPLSAGRDQHAQLRALLRELL
                     GHGLEISEMCRIFREYWAIATRNETVHGYLKSYYRDLAEVMAEKLAPLASSEKALAVA
                     VSLVIPYVEGYSVTAIAMPESIDTISETLTNVVLEQLRISNS"
     gene            918..1931
                     /gene="dhlA"
     promoter        918..946
                     /gene="dhlA"
                     /note="putative"
     promoter        945..974
                     /gene="dhlA"
                     /note="putative"
     RBS             986..998
                     /gene="dhlA"
     CDS             999..1931
                     /gene="dhlA"
                     /codon_start=1
                     /transl_table=11
                     /product="haloalkane dehalogenase"
                     /protein_id="AAA88691.1"
                     /db_xref="GI:155348"
                     /translation="MINAIRTPDQRFSNLDQYPFSPNYLDDLPGYPGLRAHYLDEGNS
                     DAEDVFLCLHGEPTWSYLYRKMIPVFAESGARVIAPDFFGFGKSDKPVDEEDYTFEFH
                     RNFLLALIERLDLRNITLVVQDWGGFLGLTLPMADPSRFKRLIIMNACLMTDPVTQPA
                     FSAFVTQPADGFTAWKYDLVTPSDLRLDQFMKRWAPTLTEAEASAYAAPFPDTSYQAG
                     VRKFPKMVAQRDQACIDISTEAISFWQNDWNGQTFMAIGMKDKLLGPDVMYPMKALIN
                     GCPEPLEIADAGHFVQEFGEQVAREALKHFAETE"
```

Ukázka záznamu v GenBank

- Sekvence



```
ORIGIN      1 bp upstream of BamHI site.
        1 atgataaatg caattcgcac cccggaccaa cgcttcagca atctcgatca gtatccgttc
       61 agccccaact acctggacga cctccccggc tacccgggat tgcgggcaca ctacctcgac
      121 gagggcaatt ctgacgctga agacgttttt ctctgccttc atggcgagcc cacctggagt
      181 tacctgtatc gcaagatgat cccggtattt gctgaatcag gcgcacgagt tattgcgcca
      241 gacttttttg gattcggaaa atccgacaag ccagtagacg aagaagacta caccttcgaa
      301 tttcaccgca acttcctgct tgcactaatc gaacggcttg acttgcgcaa cattacgctg
      361 gtcgttcagg actggggcgg atttttgggg ctgaccttac cgatggccga cccttcccgc
      421 ttcaagcgcc tgatcatcat gaacgcctgc ttgatgaccg acccggtcac ccagcctgcg
      481 tttagcgcct ttgtcaccca gcctgcggat ggctttaccg cctggaaata cgatctggtt
      541 acgccatcag acctgcgcct tgaccagttc atgaagcgtt gggcgcccac actgaccgaa
      601 gctgaggcct ccgcgtatgc tgcgcctttc cctgacactt cctatcaggc tggtgtacgc
      661 aagtttccca agatggtcgc gcaacgcgac caggcctgca tcgacatttc aaccgaagcg
      721 atttcgttct ggcagaacga ctggaatggc cagaccttca tggccattgg catgaaagac
      781 aaattgctgg gaccggacgt catgtatcct atgaaggcgc tcattaatgg ctgcccggaa
      841 cccctcgaaa tagcggacgc tggccatttc gtacaggagt ttggcgagca agtggctcgc
      901 gaggccctga aacactttgc cgagacagaa tag
//
```

Databáze proteinových sekvencí

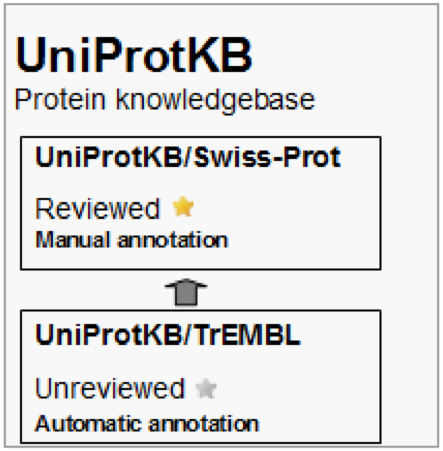
- UniProtKB
- nr Protein Database

Databáze proteinových sekvencí

- UniProtKB
  - Spolupráce EBI, Swiss Institute of Bioinformatics a Protein Information Resource
  - Centrální úložiště proteinových sekvencí a funkčních informací
  - Kvalitní anotace - informace o funkci proteinu a jednotlivých aminokyselin, experimentální informace, biologické ontologie, klasifikace, odkazy do dalších databází
  - Indikace kvality anotace (manuální vs. automatická)

Databáze proteinových sekvencí

- UniProtKB/Swiss-Prot
  - Vysoká kvalita manuálních anotací
  - ☺ Manuální anotace – spolehlivé informace
  - ☹ 549.000 sekvencí (2015)



UniProtKB
Protein knowledgebase

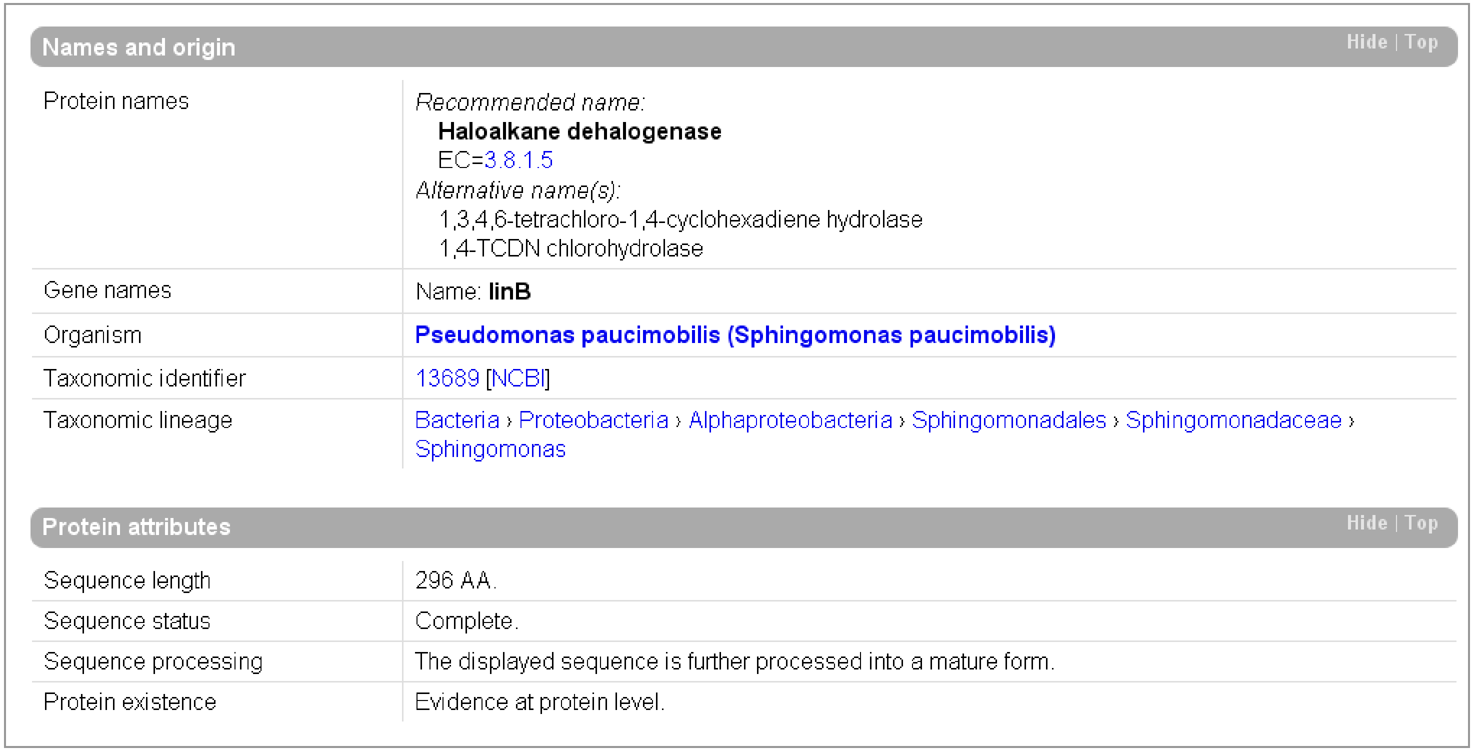
UniProtKB/Swiss-Prot
Reviewed ★
Manual annotation

UniProtKB/TrEMBL
Unreviewed ★
Automatic annotation

- UniProtKB/TrEMBL
  - Sekvence konceptuální translací kódujících sekvencí EMBL-Bank
  - ☹ Automatická anotace – nižší kvalita, možnosti chyb
  - ☺ 50.825.000 sekvencí (2015)

Ukázka záznamu v UniProtKB

- Názvy a zdroj proteinu
- Vlastnosti proteinu



Names and origin

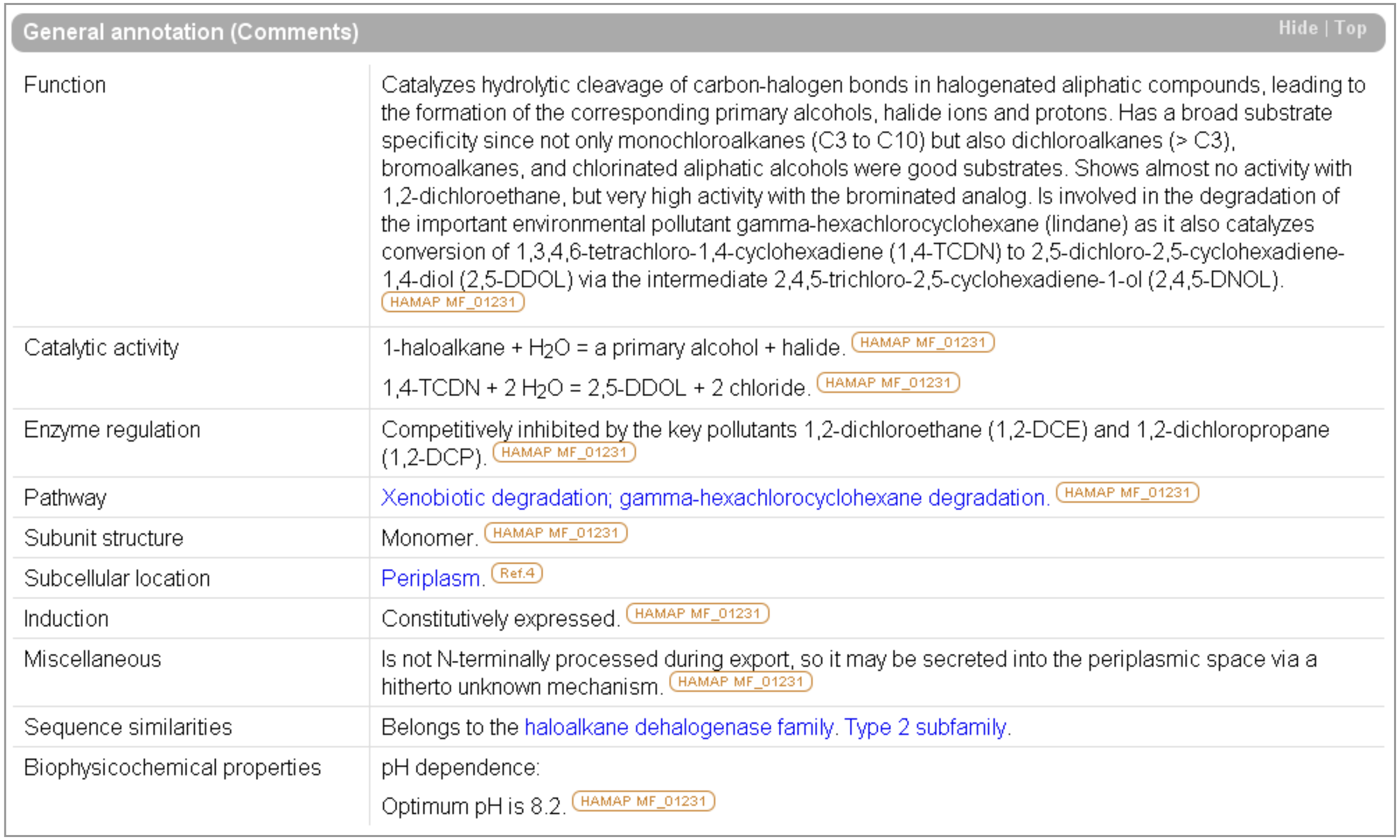
Protein names	*Recommended name:* **Haloalkane dehalogenase** EC=3.8.1.5 *Alternative name(s):* 1,3,4,6-tetrachloro-1,4-cyclohexadiene hydrolase 1,4-TCDN chlorohydrolase
Gene names	Name: **linB**
Organism	**Pseudomonas paucimobilis (Sphingomonas paucimobilis)**
Taxonomic identifier	13689 [NCBI]
Taxonomic lineage	Bacteria › Proteobacteria › Alphaproteobacteria › Sphingomonadales › Sphingomonadaceae › Sphingomonas

Protein attributes

Sequence length	296 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level.

Ukázka záznamu v UniProtKB

- Obecná anotace



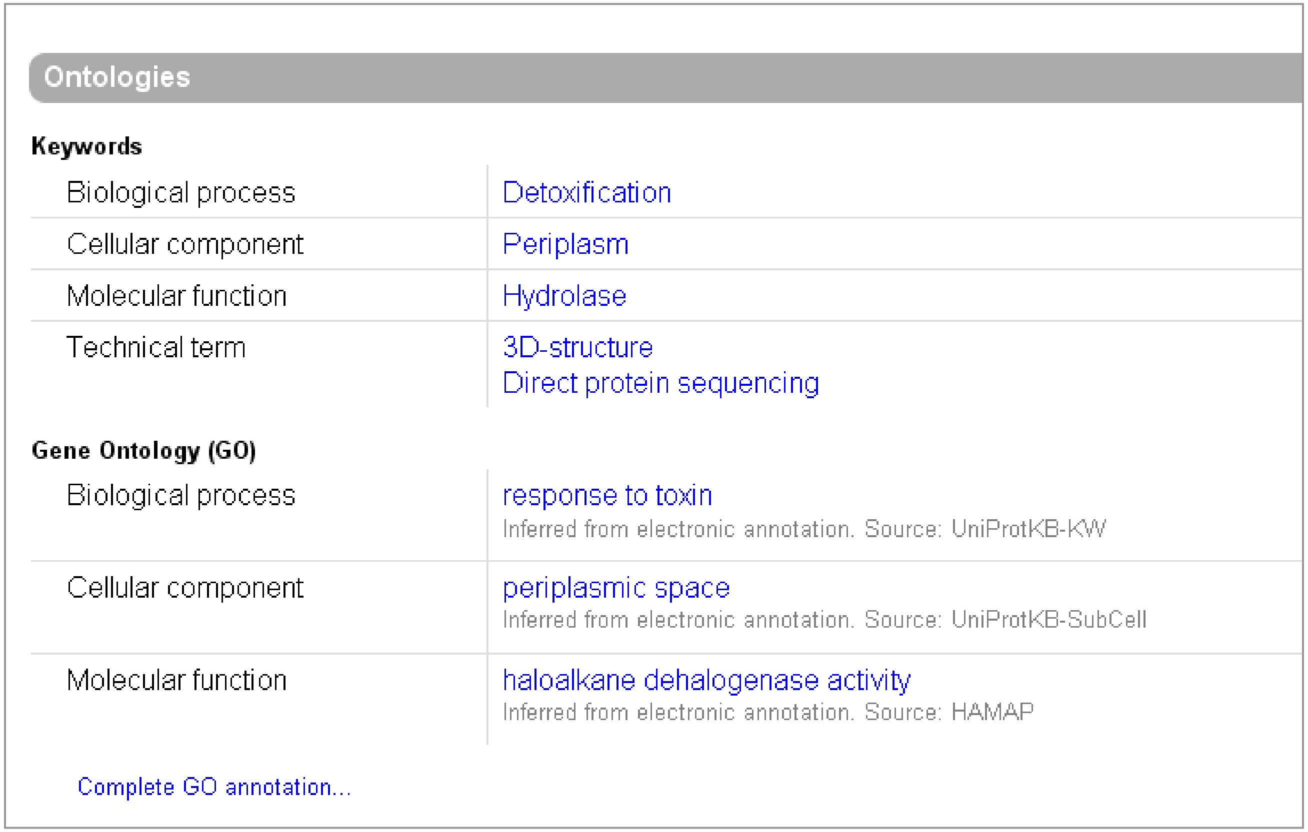
General annotation (Comments)

Hide | Top

Function	Catalyzes hydrolytic cleavage of carbon-halogen bonds in halogenated aliphatic compounds, leading to the formation of the corresponding primary alcohols, halide ions and protons. Has a broad substrate specificity since not only monochloroalkanes (C3 to C10) but also dichloroalkanes (> C3), bromoalkanes, and chlorinated aliphatic alcohols were good substrates. Shows almost no activity with 1,2-dichloroethane, but very high activity with the brominated analog. Is involved in the degradation of the important environmental pollutant gamma-hexachlorocyclohexane (lindane) as it also catalyzes conversion of 1,3,4,6-tetrachloro-1,4-cyclohexadiene (1,4-TCDN) to 2,5-dichloro-2,5-cyclohexadiene-1,4-diol (2,5-DDOL) via the intermediate 2,4,5-trichloro-2,5-cyclohexadiene-1-ol (2,4,5-DNOL). HAMAP MF_01231
Catalytic activity	1-haloalkane + H_2O = a primary alcohol + halide. HAMAP MF_01231 1,4-TCDN + 2 H_2O = 2,5-DDOL + 2 chloride. HAMAP MF_01231
Enzyme regulation	Competitively inhibited by the key pollutants 1,2-dichloroethane (1,2-DCE) and 1,2-dichloropropane (1,2-DCP). HAMAP MF_01231
Pathway	Xenobiotic degradation; gamma-hexachlorocyclohexane degradation. HAMAP MF_01231
Subunit structure	Monomer. HAMAP MF_01231
Subcellular location	Periplasm. Ref.4
Induction	Constitutively expressed. HAMAP MF_01231
Miscellaneous	Is not N-terminally processed during export, so it may be secreted into the periplasmic space via a hitherto unknown mechanism. HAMAP MF_01231
Sequence similarities	Belongs to the haloalkane dehalogenase family. Type 2 subfamily.
Biophysicochemical properties	pH dependence: Optimum pH is 8.2. HAMAP MF_01231

Ukázka záznamu v UniProtKB

- Ontologie



Ontologies

Keywords

Biological process	Detoxification
Cellular component	Periplasm
Molecular function	Hydrolase
Technical term	3D-structure Direct protein sequencing

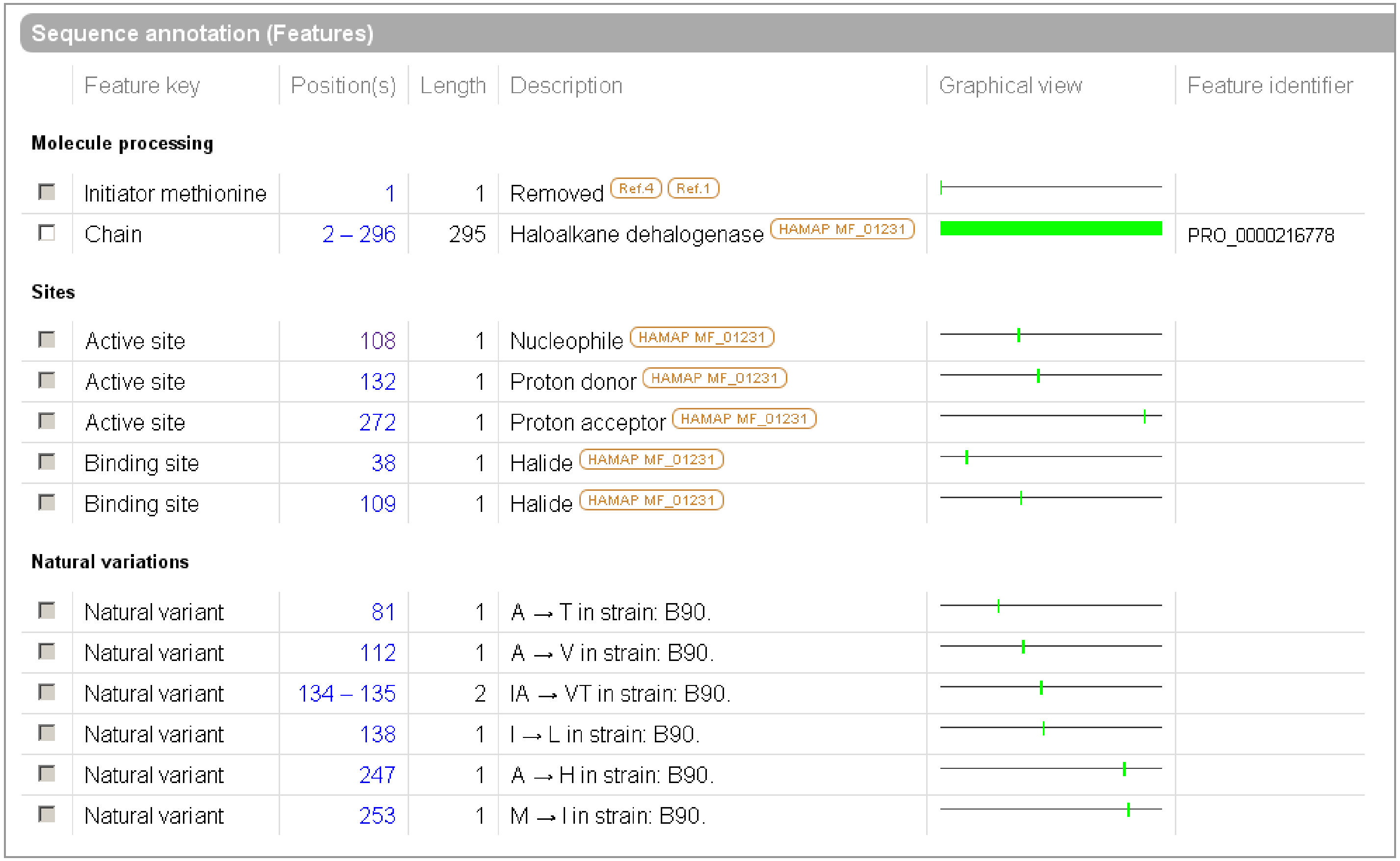
Gene Ontology (GO)

Biological process	response to toxin Inferred from electronic annotation. Source: UniProtKB-KW
Cellular component	periplasmic space Inferred from electronic annotation. Source: UniProtKB-SubCell
Molecular function	haloalkane dehalogenase activity Inferred from electronic annotation. Source: HAMAP

Complete GO annotation...

Ukázka záznamu v UniProtKB

- Anotace sekvence

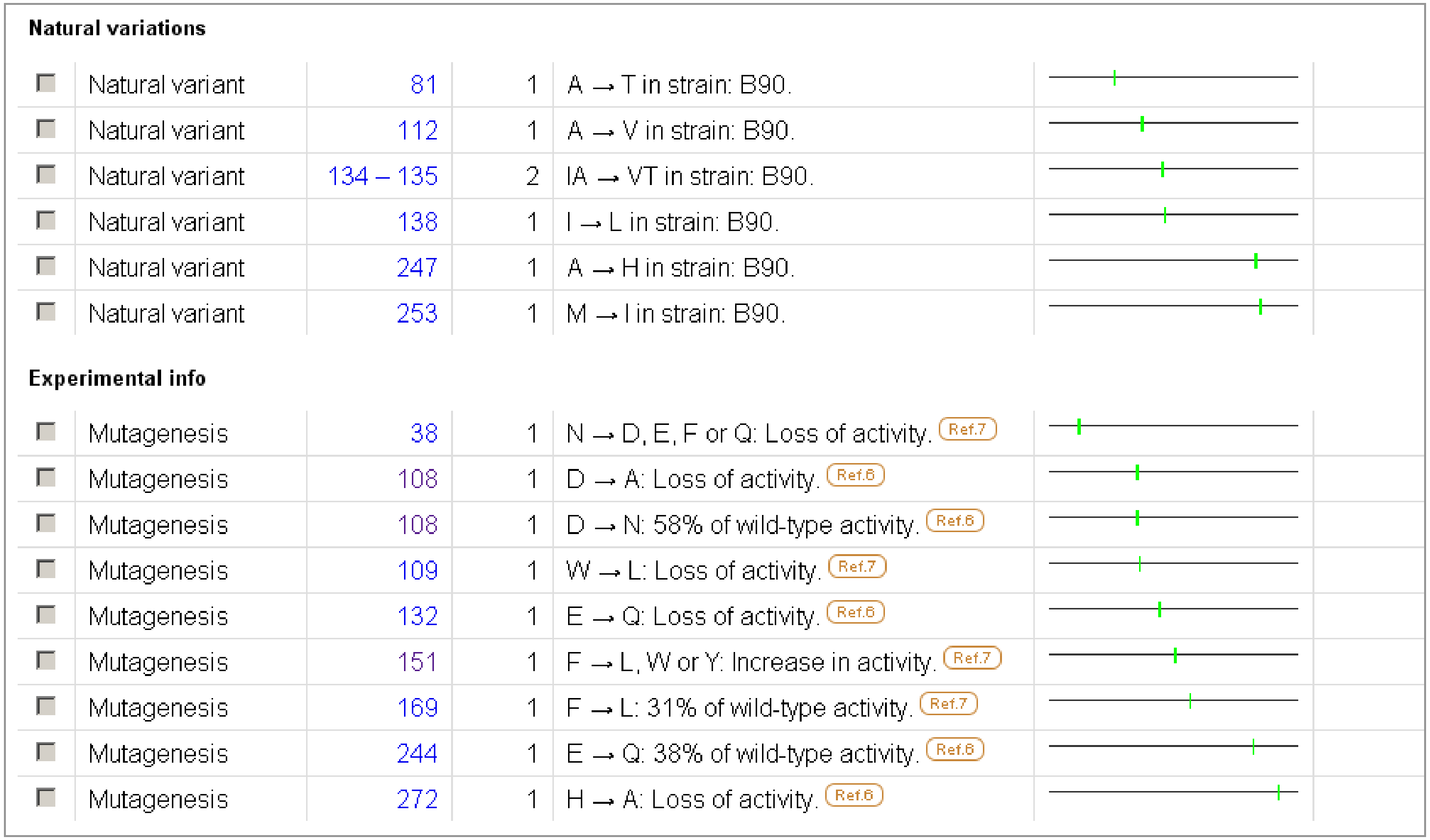


Sequence annotation (Features)

	Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Molecule processing						
☐	Initiator methionine	1	1	Removed Ref.4 Ref.1		
☐	Chain	2 – 296	295	Haloalkane dehalogenase HAMAP MF_01231		PRO_0000216778
Sites						
☐	Active site	108	1	Nucleophile HAMAP MF_01231		
☐	Active site	132	1	Proton donor HAMAP MF_01231		
☐	Active site	272	1	Proton acceptor HAMAP MF_01231		
☐	Binding site	38	1	Halide HAMAP MF_01231		
☐	Binding site	109	1	Halide HAMAP MF_01231		
Natural variations						
☐	Natural variant	81	1	A → T in strain: B90.		
☐	Natural variant	112	1	A → V in strain: B90.		
☐	Natural variant	134 – 135	2	IA → VT in strain: B90.		
☐	Natural variant	138	1	I → L in strain: B90.		
☐	Natural variant	247	1	A → H in strain: B90.		
☐	Natural variant	253	1	M → I in strain: B90.		

Ukázka záznamu v UniProtKB

- Anotace sekvence



Natural variations

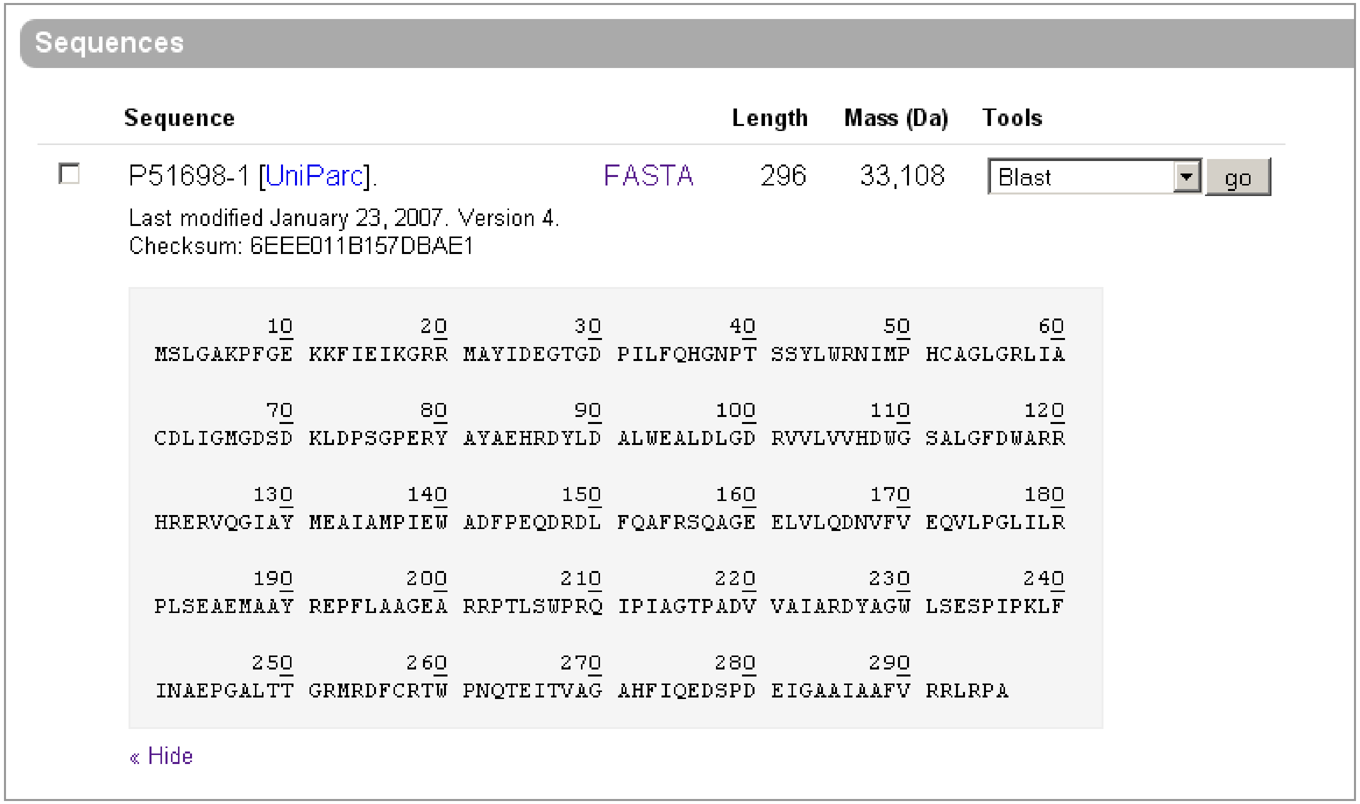
☐	Natural variant	81	1	A → T in strain: B90.
☐	Natural variant	112	1	A → V in strain: B90.
☐	Natural variant	134 – 135	2	IA → VT in strain: B90.
☐	Natural variant	138	1	I → L in strain: B90.
☐	Natural variant	247	1	A → H in strain: B90.
☐	Natural variant	253	1	M → I in strain: B90.

Experimental info

☐	Mutagenesis	38	1	N → D, E, F or Q: Loss of activity. Ref.7
☐	Mutagenesis	108	1	D → A: Loss of activity. Ref.6
☐	Mutagenesis	108	1	D → N: 58% of wild-type activity. Ref.6
☐	Mutagenesis	109	1	W → L: Loss of activity. Ref.7
☐	Mutagenesis	132	1	E → Q: Loss of activity. Ref.6
☐	Mutagenesis	151	1	F → L, W or Y: Increase in activity. Ref.7
☐	Mutagenesis	169	1	F → L: 31% of wild-type activity. Ref.7
☐	Mutagenesis	244	1	E → Q: 38% of wild-type activity. Ref.6
☐	Mutagenesis	272	1	H → A: Loss of activity. Ref.6

Ukázka záznamu v UniProtKB

- Sekvence



Sequences

	Sequence		**Length**	**Mass (Da)**	**Tools**
☐	P51698-1 [UniParc]. Last modified January 23, 2007. Version 4. Checksum: 6EEE011B157DBAE1	FASTA	296	33,108	Blast

```
        10         20         30         40         50         60
MSLGAKPFGE KKFIEIKGRR MAYIDEGTGD PILFQHGNPT SSYLWRNIMP HCAGLGRLIA

        70         80         90        100        110        120
CDLIGMGDSD KLDPSGPERY AYAEHRDYLD ALWEALDLGD RVVLVVHDWG SALGFDWARR

       130        140        150        160        170        180
HRERVQGIAY MEAIAMPIEW ADFPEQDRDL FQAFRSQAGE ELVLQDNVFV EQVLPGLILR

       190        200        210        220        230        240
PLSEAEMAAY REPFLAAGEA RRPTLSWPRQ IPIAGTPADV VAIARDYAGW LSESPIPKLF

       250        260        270        280        290
INAEPGALTT GRMRDFCRTW PNQTEITVAG AHFIQEDSPD EIGAAIAAFV RRLRPA
```

« Hide

Ukázka záznamu v UniProtKB

- Reference

[1] **"Cloning and sequencing of a dehalogenase gene encoding an enzyme with hydrolase activity involved in the degradation of gamma-hexachlorocyclohexane in Pseudomonas paucimobilis."**
Nagata Y., Nariya T., Ohtomo R., Fukuda M., Yano K., Takagi M.
J. Bacteriol. 175:6403-6410(1993) [PubMed: 7691794] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA], PROTEIN SEQUENCE OF 2-16.
Strain: UT26.

[2] Nagata Y., Nariya T., Ohtomo R., Fukuda M., Yano K., Takagi M.
Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases
Cited for: SEQUENCE REVISION.

[3] **"Cloning and characterization of lin genes responsible for the degradation of hexachlorocyclohexane isomers by Sphingomonas paucimobilis strain B90."**
Kumari R., Subudhi S., Suar M., Dhingra G., Raina V., Dogra C., Lal S., van der Meer J.R., Holliger C., Lal R.
Appl. Environ. Microbiol. 68:6021-6028(2002) [PubMed: 12450824] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA].
Strain: B90.

[4] **"Two different types of dehalogenases, LinA and LinB, involved in gamma-hexachlorocyclohexane degradation in Sphingomonas paucimobilis UT26 are localized in the periplasmic space without molecular processing."**
Nagata Y., Futamura A., Miyauchi K., Takagi M.
J. Bacteriol. 181:5409-5413(1999) [PubMed: 10464214] [Abstract]
Cited for: PROTEIN SEQUENCE OF 2-10, SUBCELLULAR LOCATION.

[5] **"Purification and characterization of a haloalkane dehalogenase of a new substrate class from a gamma-hexachlorocyclohexane-degrading bacterium, Sphingomonas paucimobilis UT26."**
Nagata Y., Miyauchi K., Damborsky J., Manova K., Ansorgova A., Takagi M.
Appl. Environ. Microbiol. 63:3707-3710(1997) [PubMed: 9293022] [Abstract]
Cited for: CHARACTERIZATION.
Strain: UT26.

Databáze proteinových sekvencí

- ❑ nr Protein Database
  - Databáze proteinových sekvencí NCBI
  - Kolekce sekvencí získaných konceptuální translací kódujících oblastí GenBank/EMBL-Bank/DDBJ a dále sekvencí z UniProtKB, PRF a RCSB PDB
  - ☹ většinou automatická anotace – nižší kvalita, možnost chyb
  - ☹ chybí indikace původu anotace
  - ☺ více než 25.000.000 sekvencí (2015)

Formáty sekvencí

- Prostá sekvence

```
DQLTEEQIAEFKEAFSLFDK
```

Formáty sekvencí

- ❑ Prostá sekvence
- ❑ **GenBank**

```
LOCUS       AAU03518      237 bp    DNA             PLN       04-FEB-1995
DEFINITION  Aspergillus awamori internal transcribed spacer 1 (ITS1) and 18S
            rRNA and 5.8S rRNA genes, partial sequence.
ACCESSION   U03518
BASE COUNT       41 a      77 c      67 g      52 t
ORIGIN
        1 aacctgcgga aggatcatta ccgagtgcgg gtcctttggg cccaacctcc catccgtgtc
       61 tattgtaccc tgttgcttcg gcgggcccgc cgcttgtcgg ccgccggggg ggcgcctctg
      121 ccccccgggc ccgtgcccgc cggagacccc aacacgaaca ctgtctgaaa gcgtgcagtc
      181 tgagttgatt gaatgcaatc agttaaaact ttcaacaatg gatctcttgg ttccggc
//
```

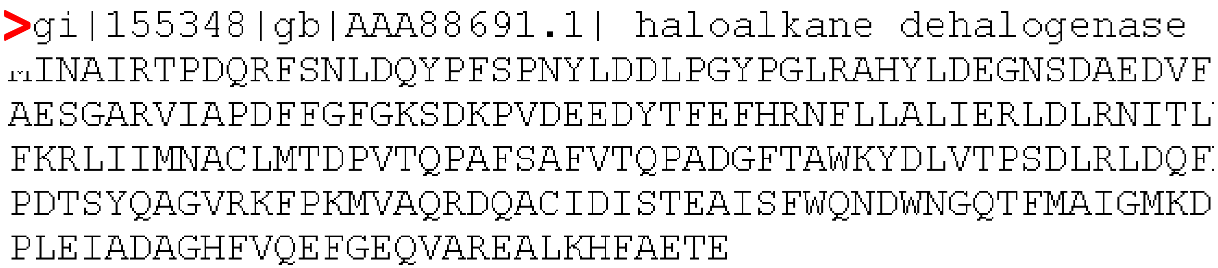
Formáty sekvencí

- Prostá sekvence
- GenBank
- **EMBL**

```
ID   AA03518    standard; DNA; FUN; 237 BP.
XX
AC   U03518;
XX
DE   Aspergillus awamori internal transcribed spacer 1 (ITS1) and 18S
DE   rRNA and 5.8S rRNA genes, partial sequence.
XX
SQ   Sequence 237 BP; 41 A; 77 C; 67 G; 52 T; 0 other;
     aacctgcgga aggatcatta ccgagtgcgg gtcctttggg cccaacctcc catccgtgtc       60
     tattgtaccc tgttgcttcg gcgggcccgc cgcttgtcgg ccgccggggg ggcgcctctg      120
     ccccccgggc ccgtgcccgc cggagacccc aacacgaaca ctgtctgaaa gcgtgcagtc      180
     tgagttgatt gaatgcaatc agttaaaact ttcaacaatg gatctcttgg ttccggc          237
//
```

Formáty sekvencí

- ❑ Prostá sekvence
- ❑ GenBank
- ❑ EMBL
- ❑ **FASTA**



```
>gi|155348|gb|AAA88691.1| haloalkane dehalogenase
MINAIRTPDQRFSNLDQYPFSPNYLDDLPGYPGLRAHYLDEGNSDAEDVF
AESGARVIAPDFFGFGKSDKPVDEEDYTFEFHRNFLLALIERLDLRNITL
FKRLIIMNACLMTDPVTQPAFSAFVTQPADGFTAWKYDLVTPSDLRLDQF
PDTSYQAGVRKFPKMVAQRDQACIDISTEAISFWQNDWNGQTFMAIGMKD
PLEIADAGHFVQEFGEQVAREALKHFAETE
```

Nejčastěji používaný formát

Databáze proteinových struktur

- Worldwide Protein Data Bank (wwPDB)
  - Světový depositář proteinových struktur, obsahuje rovněž struktury nukleových kyselin a biomolekulárních komplexů
  - Research Collaboratory for Structural Bioinformatics (RCSB PDB), Protein Data Bank Europe (PDBe), Protein Data Bank Japan (PDBj), Biological Magnetic Resonance Data Bank (BioMagResBank)
  - Obsahuje více než 84.000 struktur (2012)
  - Struktury získané rentgenovou krystalografií (88%) a nukleární magnetickou rezonancí (11 %)

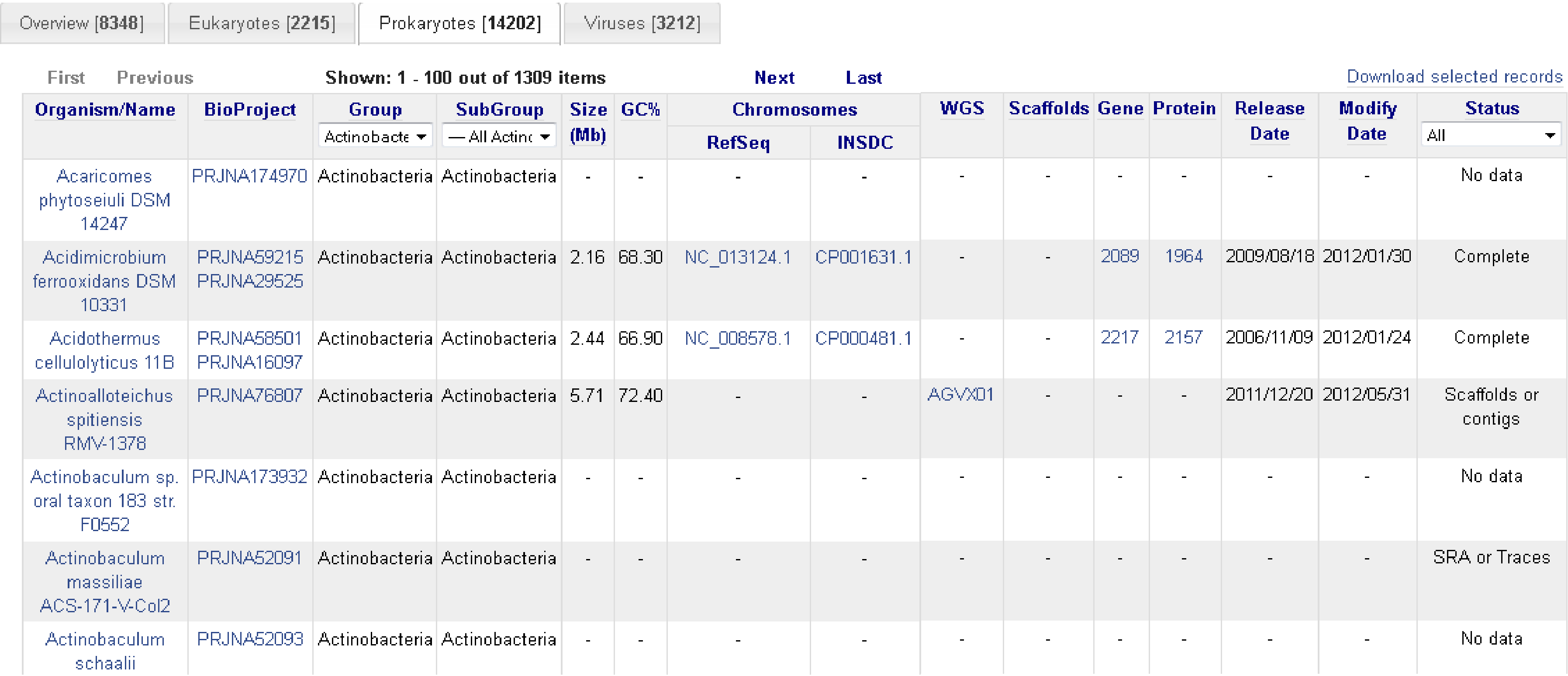
Databáze genomů

- Entrez Genome
- Ensembl
- Genomes OnLine Database GOLD

- Informace o zdrojovém organizmu
- Nukleotidové a proteinové sekvence
- Geny v kontextu genomu
- Anotace a analýza genomů

Ukázka záznamu v Entrez Genome

- Přehled prokaryotických genomů

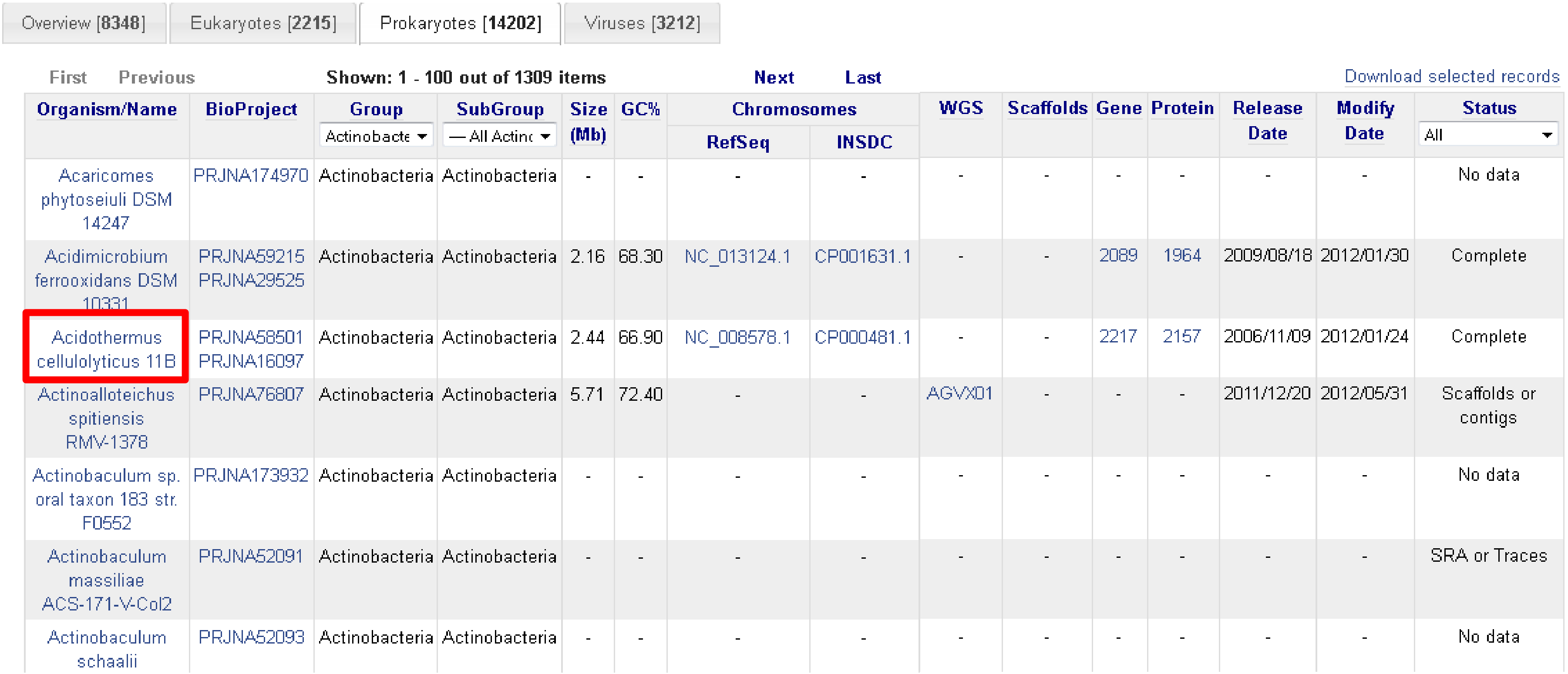


Shown: 1 - 100 out of 1309 items

Organism/Name	BioProject	Group	SubGroup	Size (Mb)	GC%	Chromosomes		WGS	Scaffolds	Gene	Protein	Release Date	Modify Date	Status
		Actinobacte ▾	— All Actin ▾			RefSeq	INSDC							All ▾
Acaricomes phytoseiuli DSM 14247	PRJNA174970	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data
Acidimicrobium ferrooxidans DSM 10331	PRJNA59215 PRJNA29525	Actinobacteria	Actinobacteria	2.16	68.30	NC_013124.1	CP001631.1	-	-	2089	1964	2009/08/18	2012/01/30	Complete
Acidothermus cellulolyticus 11B	PRJNA58501 PRJNA16097	Actinobacteria	Actinobacteria	2.44	66.90	NC_008578.1	CP000481.1	-	-	2217	2157	2006/11/09	2012/01/24	Complete
Actinoalloteichus spitiensis RMV-1378	PRJNA76807	Actinobacteria	Actinobacteria	5.71	72.40	-	-	AGVX01	-	-	-	2011/12/20	2012/05/31	Scaffolds or contigs
Actinobaculum sp. oral taxon 183 str. F0552	PRJNA173932	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data
Actinobaculum massiliae ACS-171-V-Col2	PRJNA52091	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	SRA or Traces
Actinobaculum schaalii	PRJNA52093	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data

Ukázka záznamu v Entrez Genome

- Přehled prokaryotických genomů

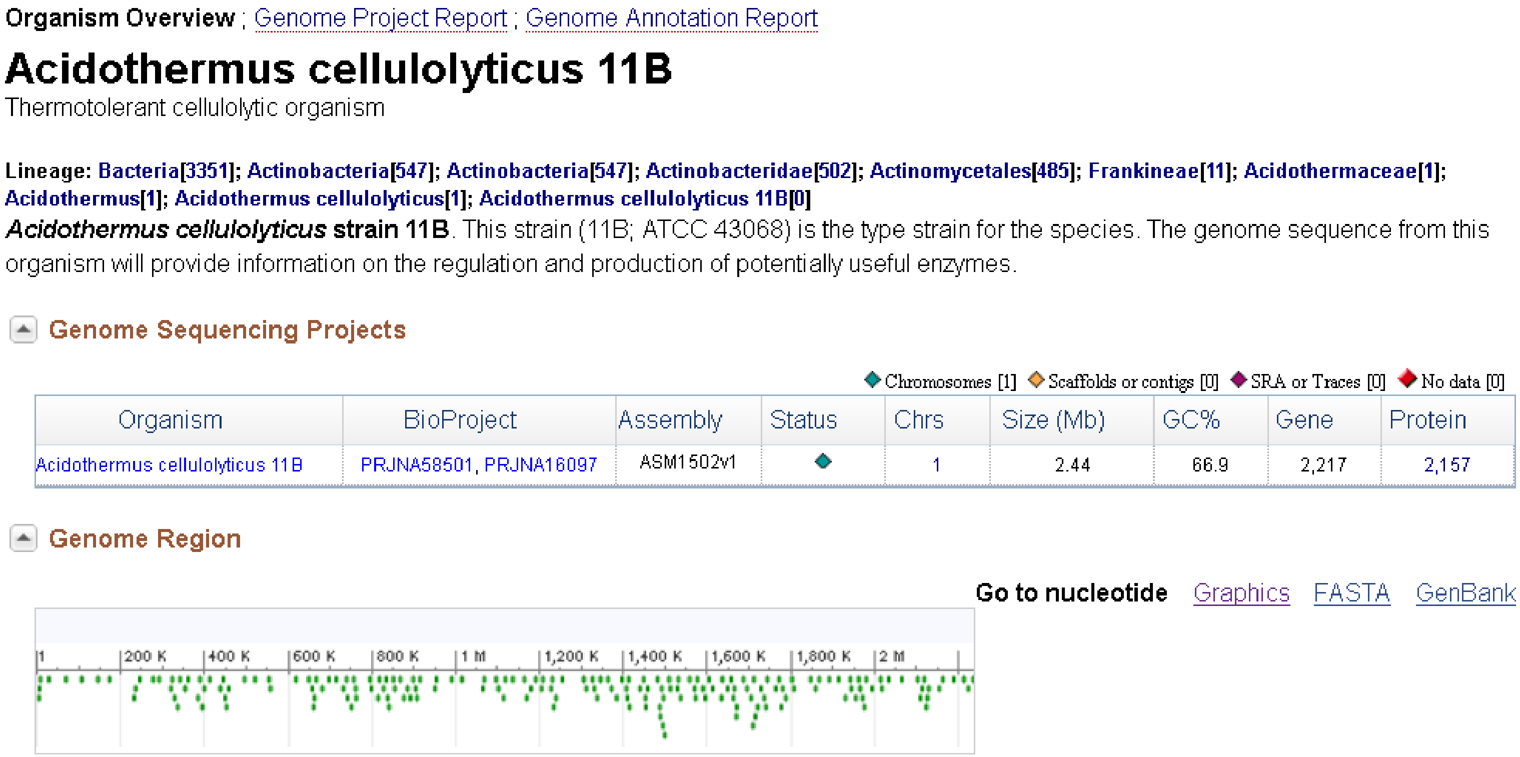


Shown: 1 - 100 out of 1309 items

Organism/Name	BioProject	Group	SubGroup	Size (Mb)	GC%	Chromosomes		WGS	Scaffolds	Gene	Protein	Release Date	Modify Date	Status
		Actinobacte ▾	— All Actinc ▾			RefSeq	INSDC							All ▾
Acaricomes phytoseiuli DSM 14247	PRJNA174970	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data
Acidimicrobium ferrooxidans DSM 10331	PRJNA59215 PRJNA29525	Actinobacteria	Actinobacteria	2.16	68.30	NC_013124.1	CP001631.1	-	-	2089	1964	2009/08/18	2012/01/30	Complete
Acidothermus cellulolyticus 11B	PRJNA58501 PRJNA16097	Actinobacteria	Actinobacteria	2.44	66.90	NC_008578.1	CP000481.1	-	-	2217	2157	2006/11/09	2012/01/24	Complete
Actinoalloteichus spitiensis RMV-1378	PRJNA76807	Actinobacteria	Actinobacteria	5.71	72.40	-	-	AGVX01	-	-	-	2011/12/20	2012/05/31	Scaffolds or contigs
Actinobaculum sp. oral taxon 183 str. F0552	PRJNA173932	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data
Actinobaculum massiliae ACS-171-V-Col2	PRJNA52091	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	SRA or Traces
Actinobaculum schaalii	PRJNA52093	Actinobacteria	Actinobacteria	-	-	-	-	-	-	-	-	-	-	No data

Ukázka záznamu v Entrez Genome

- Informace o genomu



Organism Overview ; Genome Project Report ; Genome Annotation Report

Acidothermus cellulolyticus 11B

Thermotolerant cellulolytic organism

Lineage: Bacteria[3351]; Actinobacteria[547]; Actinobacteria[547]; Actinobacteridae[502]; Actinomycetales[485]; Frankineae[11]; Acidothermaceae[1]; Acidothermus[1]; Acidothermus cellulolyticus[1]; Acidothermus cellulolyticus 11B[0]

***Acidothermus cellulolyticus* strain 11B**. This strain (11B; ATCC 43068) is the type strain for the species. The genome sequence from this organism will provide information on the regulation and production of potentially useful enzymes.

Genome Sequencing Projects

◆ Chromosomes [1] ◆ Scaffolds or contigs [0] ◆ SRA or Traces [0] ◆ No data [0]

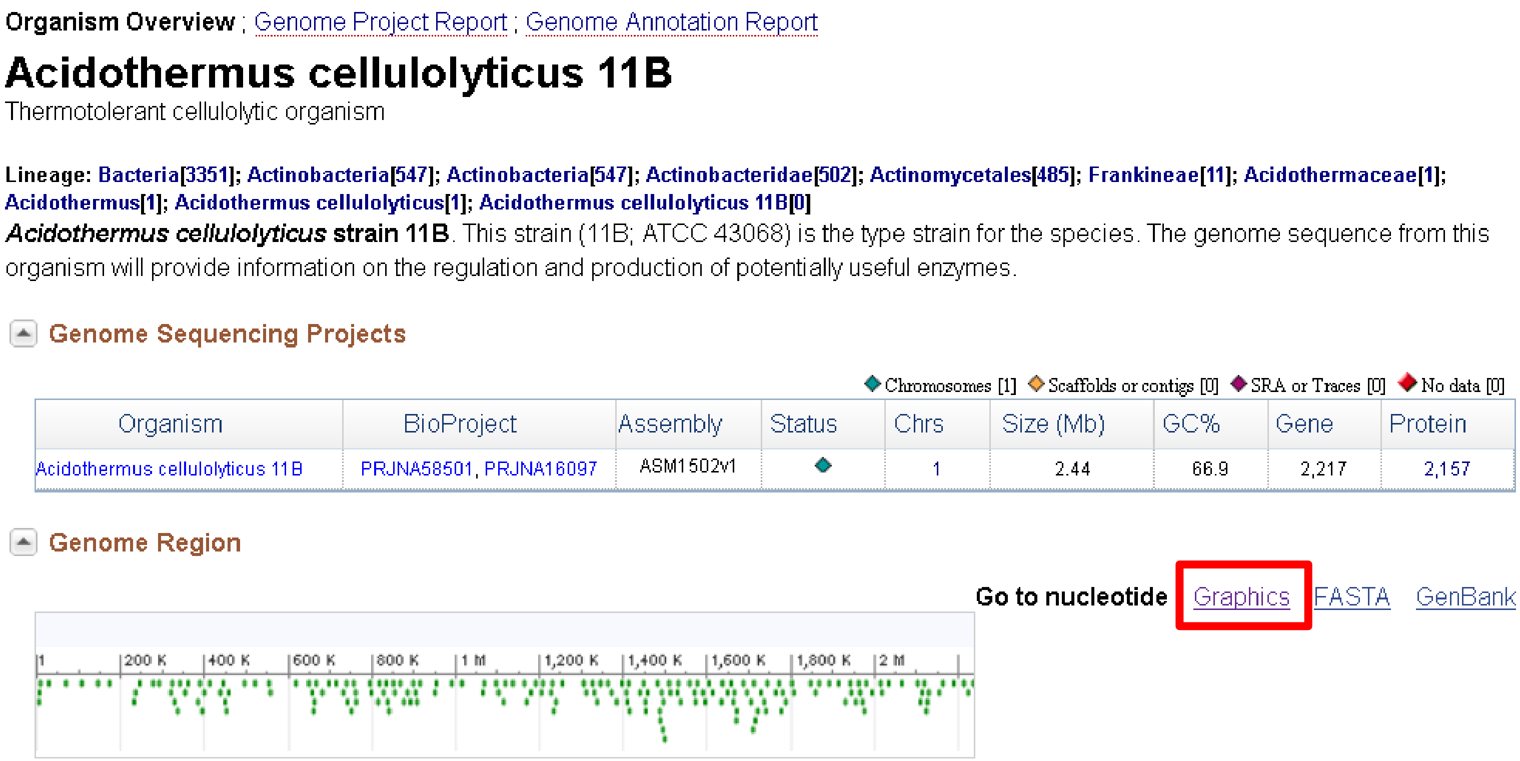
Organism	BioProject	Assembly	Status	Chrs	Size (Mb)	GC%	Gene	Protein
Acidothermus cellulolyticus 11B	PRJNA58501, PRJNA16097	ASM1502v1	◆	1	2.44	66.9	2,217	2,157

Genome Region

Go to nucleotide Graphics FASTA GenBank

Ukázka záznamu v Entrez Genome

- Informace o genomu



Organism Overview ; Genome Project Report ; Genome Annotation Report

Acidothermus cellulolyticus 11B

Thermotolerant cellulolytic organism

Lineage: **Bacteria[3351]; Actinobacteria[547]; Actinobacteria[547]; Actinobacteridae[502]; Actinomycetales[485]; Frankineae[11]; Acidothermaceae[1]; Acidothermus[1]; Acidothermus cellulolyticus[1]; Acidothermus cellulolyticus 11B[0]**

***Acidothermus cellulolyticus* strain 11B**. This strain (11B; ATCC 43068) is the type strain for the species. The genome sequence from this organism will provide information on the regulation and production of potentially useful enzymes.

Genome Sequencing Projects

◆ Chromosomes [1] ◆ Scaffolds or contigs [0] ◆ SRA or Traces [0] ◆ No data [0]

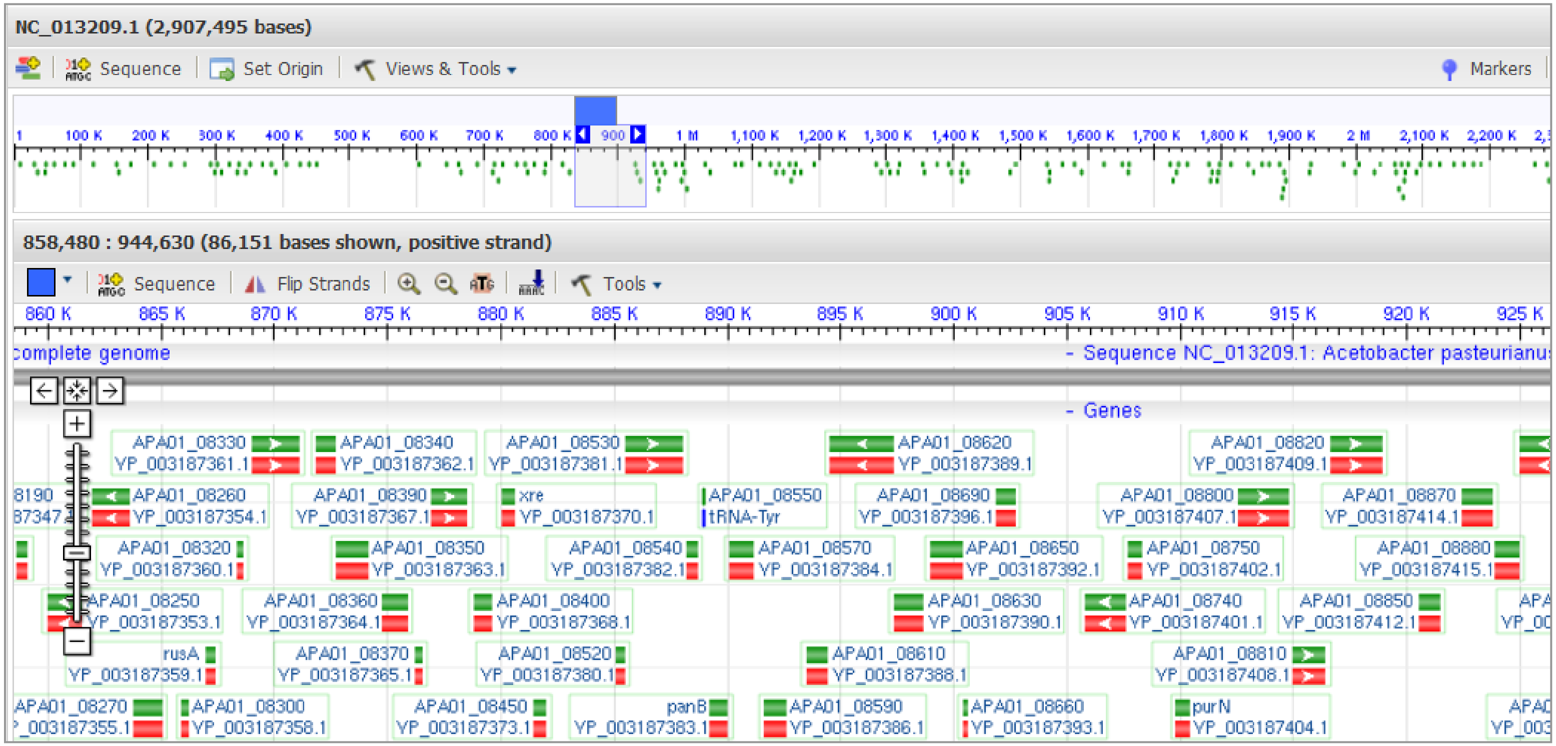
Organism	BioProject	Assembly	Status	Chrs	Size (Mb)	GC%	Gene	Protein
Acidothermus cellulolyticus 11B	PRJNA58501, PRJNA16097	ASM1502v1	◆	1	2.44	66.9	2,217	2,157

Genome Region

Go to nucleotide Graphics FASTA GenBank

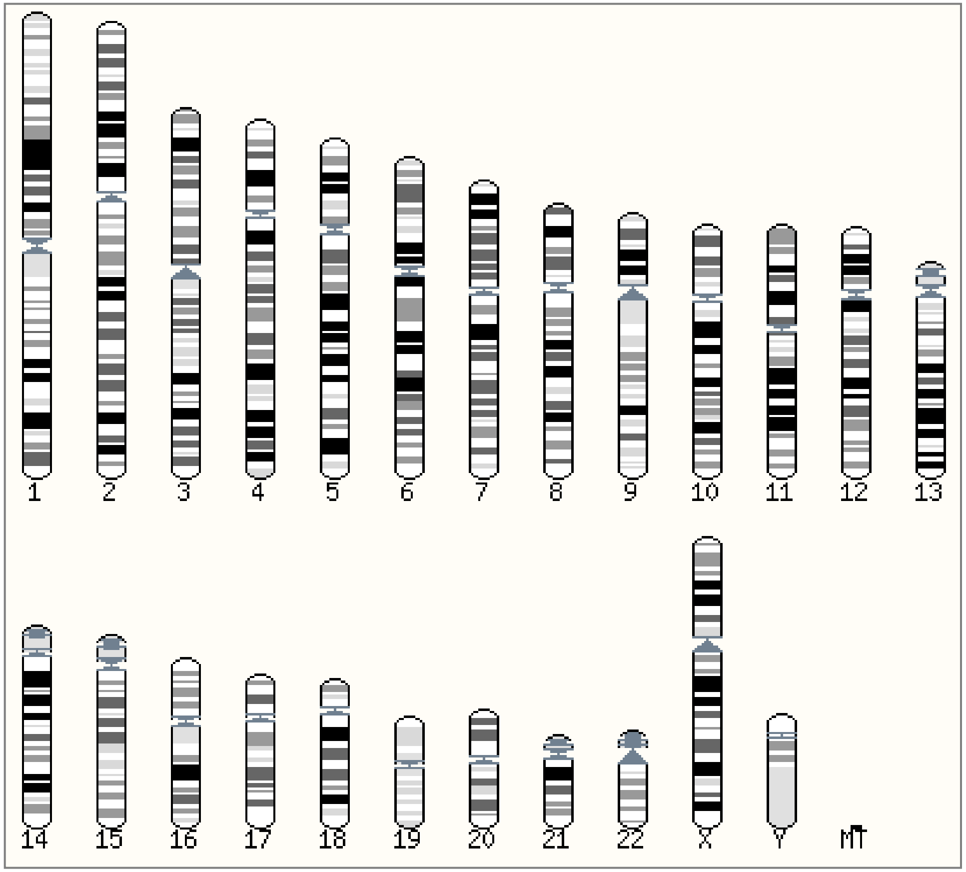
Ukázka záznamu v Entrez Genome

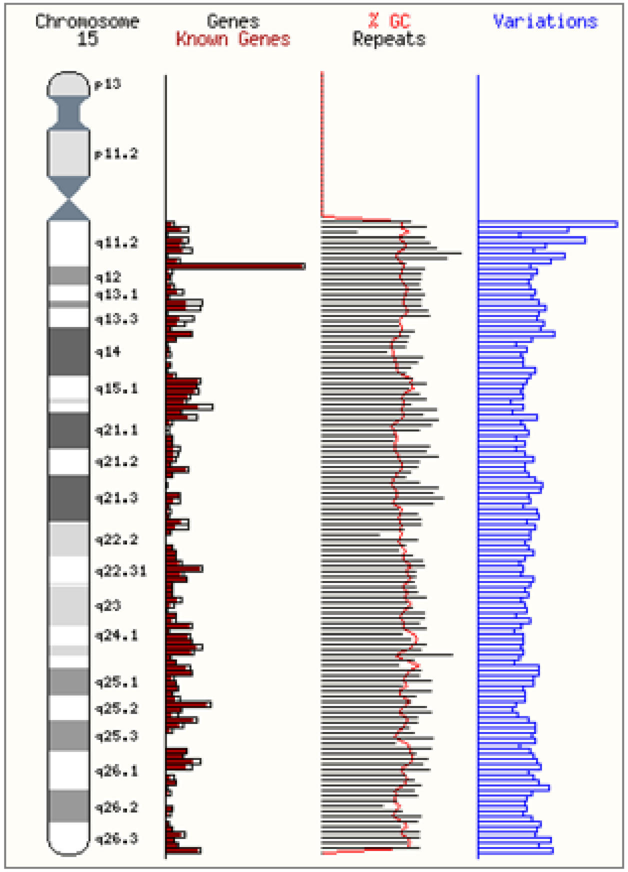
- Grafické znázornění genomu



Ukázka záznamu v Ensembl

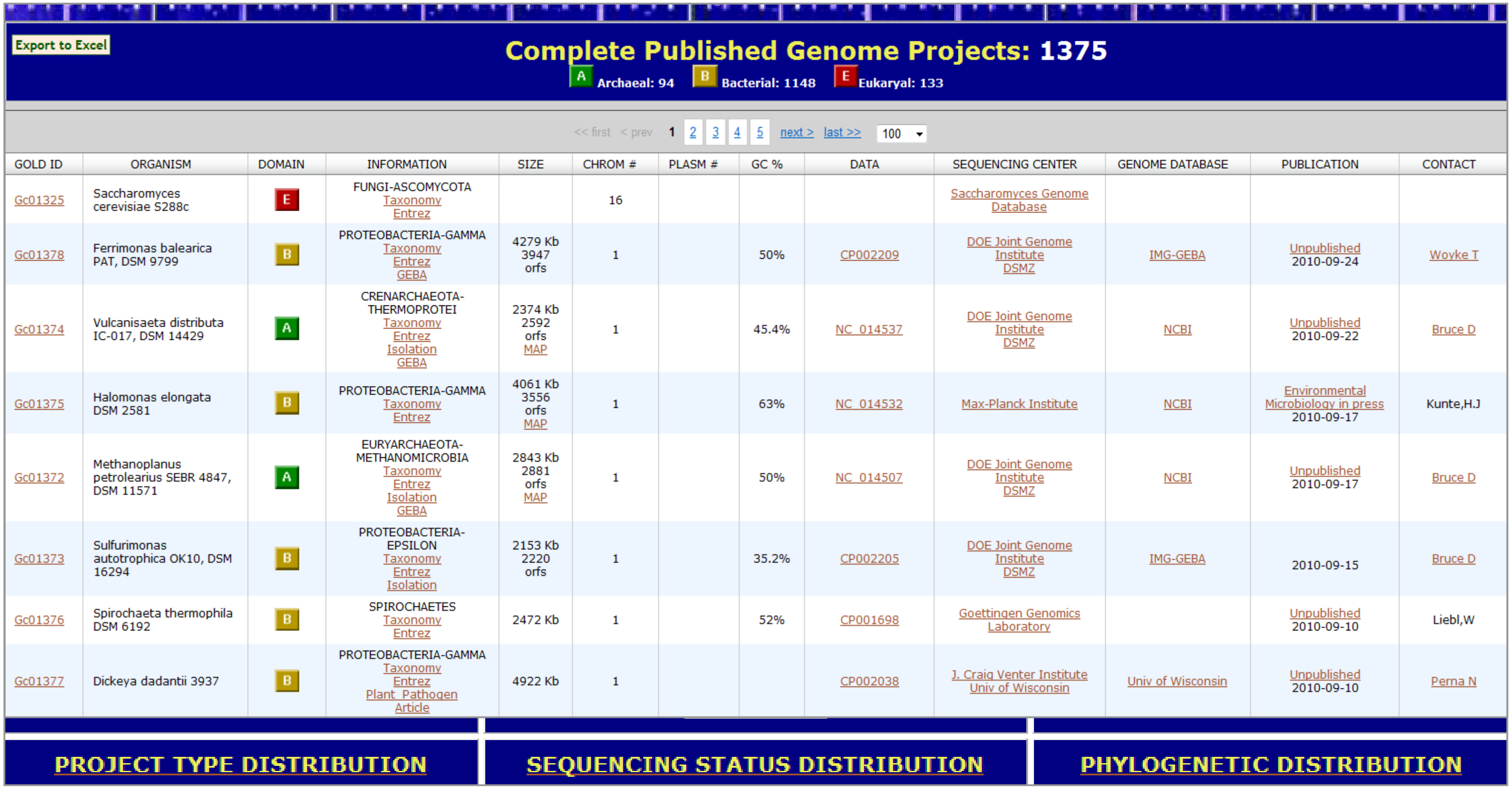
- Anotace eukaryotických genomů





Ukázka záznamu v GOLD

- Seznam dokončených a probíhajících genomových projektů



Export to Excel

Complete Published Genome Projects: 1375

A Archaeal: 94 B Bacterial: 1148 E Eukaryal: 133

<< first < prev 1 2 3 4 5 next > last >> 100

GOLD ID	ORGANISM	DOMAIN	INFORMATION	SIZE	CHROM #	PLASM #	GC %	DATA	SEQUENCING CENTER	GENOME DATABASE	PUBLICATION	CONTACT
Gc01325	Saccharomyces cerevisiae S288c	E	FUNGI-ASCOMYCOTA Taxonomy Entrez		16				Saccharomyces Genome Database			
Gc01378	Ferrimonas balearica PAT, DSM 9799	B	PROTEOBACTERIA-GAMMA Taxonomy Entrez GEBA	4279 Kb 3947 orfs	1		50%	CP002209	DOE Joint Genome Institute DSMZ	IMG-GEBA	Unpublished 2010-09-24	Woyke T
Gc01374	Vulcanisaeta distributa IC-017, DSM 14429	A	CRENARCHAEOTA-THERMOPROTEI Taxonomy Entrez Isolation GEBA	2374 Kb 2592 orfs MAP	1		45.4%	NC_014537	DOE Joint Genome Institute DSMZ	NCBI	Unpublished 2010-09-22	Bruce D
Gc01375	Halomonas elongata DSM 2581	B	PROTEOBACTERIA-GAMMA Taxonomy Entrez	4061 Kb 3556 orfs MAP	1		63%	NC_014532	Max-Planck Institute	NCBI	Environmental Microbiology in press 2010-09-17	Kunte,H.J
Gc01372	Methanoplanus petrolearius SEBR 4847, DSM 11571	A	EURYARCHAEOTA-METHANOMICROBIA Taxonomy Entrez Isolation GEBA	2843 Kb 2881 orfs MAP	1		50%	NC_014507	DOE Joint Genome Institute DSMZ	NCBI	Unpublished 2010-09-17	Bruce D
Gc01373	Sulfurimonas autotrophica OK10, DSM 16294	B	PROTEOBACTERIA-EPSILON Taxonomy Entrez Isolation	2153 Kb 2220 orfs	1		35.2%	CP002205	DOE Joint Genome Institute DSMZ	IMG-GEBA	2010-09-15	Bruce D
Gc01376	Spirochaeta thermophila DSM 6192	B	SPIROCHAETES Taxonomy Entrez	2472 Kb	1		52%	CP001698	Goettingen Genomics Laboratory		Unpublished 2010-09-10	Liebl,W
Gc01377	Dickeya dadantii 3937	B	PROTEOBACTERIA-GAMMA Taxonomy Entrez Plant_Pathogen Article	4922 Kb	1			CP002038	J. Craig Venter Institute Univ of Wisconsin	Univ of Wisconsin	Unpublished 2010-09-10	Perna N

PROJECT TYPE DISTRIBUTION SEQUENCING STATUS DISTRIBUTION PHYLOGENETIC DISTRIBUTION

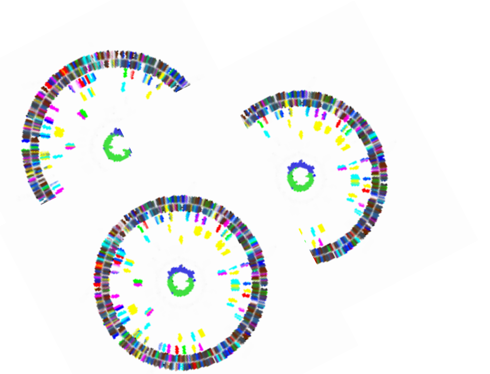
Problémy bioinformatických databází

- Vysoká redundance dat
- Chyby v sekvencích
- Chyby a nepřesnosti v anotacích
- Propagace chyb během automatických anotací

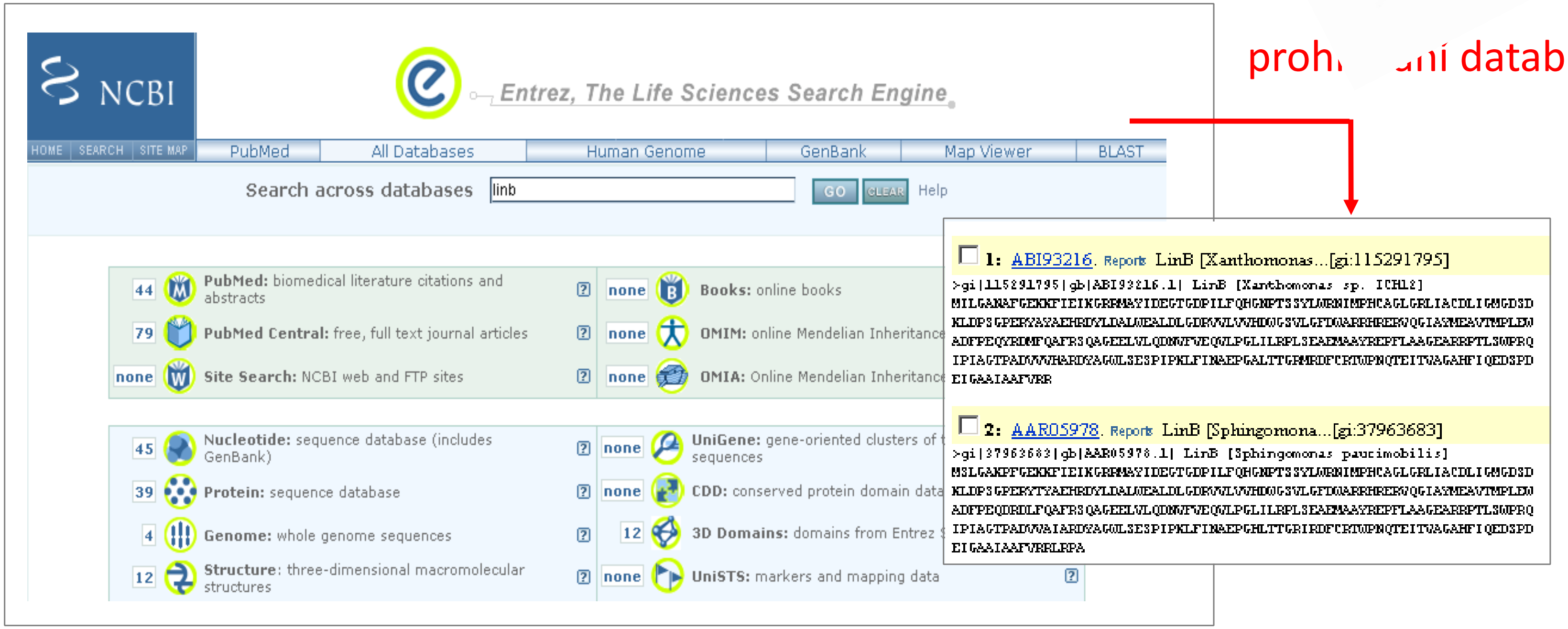
Prohledávání databází

- ❑ Textové prohledávání
- ❑ Sekvenční prohledávání

vložení dat



proh ní databáze



Search across databases linb

44	**PubMed:** biomedical literature citations and abstracts	?
79	**PubMed Central:** free, full text journal articles	?
none	**Site Search:** NCBI web and FTP sites	?
none	**Books:** online books	
none	**OMIM:** online Mendelian Inheritance	
none	**OMIA:** Online Mendelian Inheritance	
45	**Nucleotide:** sequence database (includes GenBank)	?
39	**Protein:** sequence database	?
4	**Genome:** whole genome sequences	?
12	**Structure:** three-dimensional macromolecular structures	?
none	**UniGene:** gene-oriented clusters of t sequences	
none	**CDD:** conserved protein domain data	
12	**3D Domains:** domains from Entrez S	
none	**UniSTS:** markers and mapping data	?

```
☐ 1: ABI93216. Reports LinB [Xanthomonas...[gi:115291795]
>gi|115291795|gb|ABI93216.1| LinB [Xanthomonas sp. ICH12]
MILGANAFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCAGLGRLIACDLIGMGDSD
KLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVVHDWGSVLGFDWARRHRERVQGIAYMEAVTMPLEW
ADFPEQYRDMFQAFRSQAGEELVLQDNVFVEQVLPGLILRPLSEAEMAAYREPFLAAGEARRPTLSWPRQ
IPIAGTPADVVVHARDYAGWLSESPIPKLFINAEPGALTTGRMRDFCRTWPNQTEITVAGAHFIQEDSPD
EIGAAIAAFVRR

☐ 2: AAR05978. Reports LinB [Sphingomona...[gi:37963683]
>gi|37963683|gb|AAR05978.1| LinB [Sphingomonas paucimobilis]
MSLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCAGLGRLIACDLIGMGDSD
KLDPSGPERYTYAEHRDYLDALWEALDLGDRVVLVVHDWGSVLGFDWARRHRERVQGIAYMEAVTMPLEW
ADFPEQDRDLFQAFRSQAGEELVLQDNVFVEQVLPGLILRPLSEAEMAAYREPFLAAGEARRPTLSWPRQ
IPIAGTPADVVAIARDYAGWLSESPIPKLFINAEPGHLTTGRIRDFCRTWPNQTEITVAGAHFIQEDSPD
EIGAAIAAFVRRLRPA
```

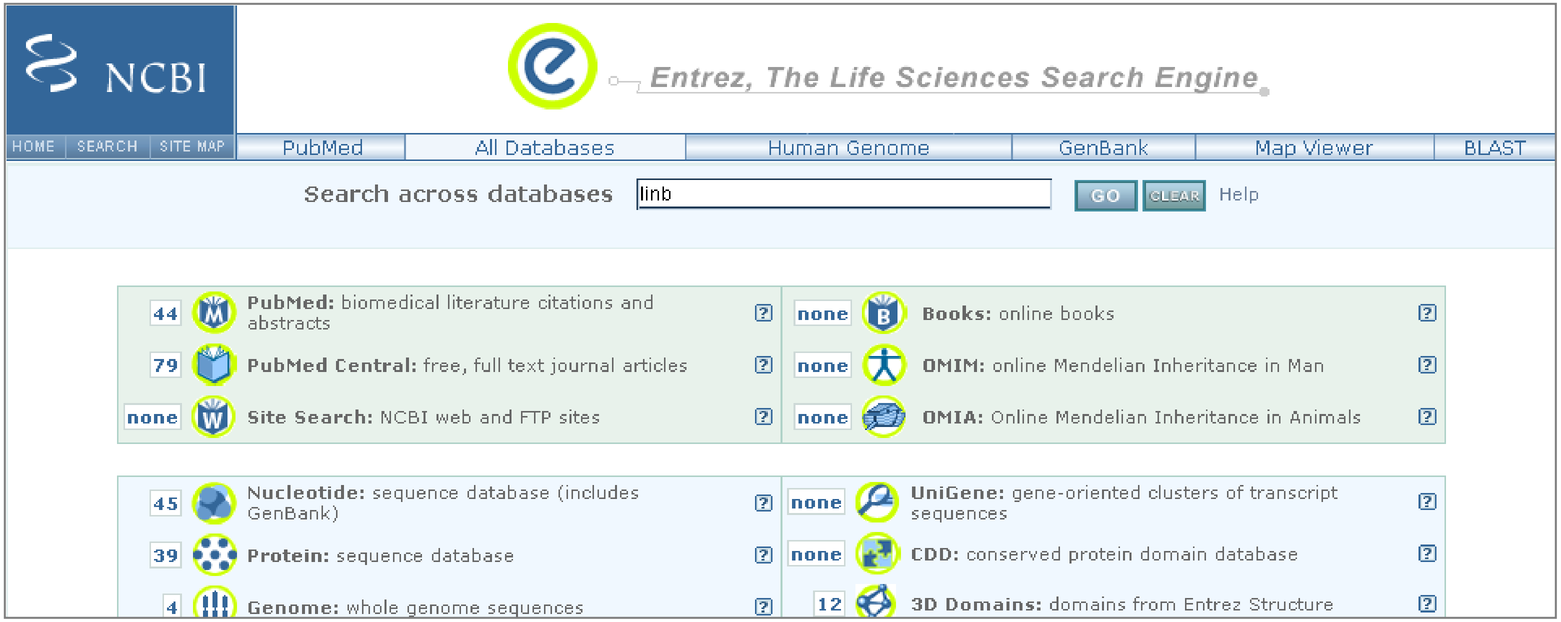
Textové prohledávání databází

- Entrez
- SRS
- DBGET

- Integrují data z různých databází
- Umožňují prohledání mnoha databází současně
- Umožňují formulaci dotazů pomocí logických operátorů

Textové prohledávání databází

- Entrez
  - Vyhledávací systém pro databáze NCBI
  - Integruje data ze 40 různých databází, pouze NCBI



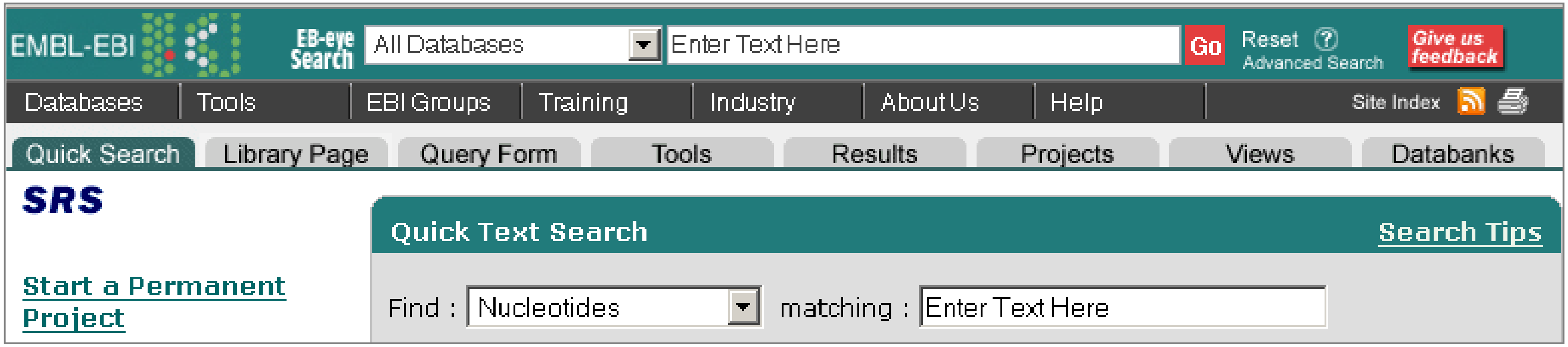
Entrez, The Life Sciences Search Engine

Search across databases: linb

44	**PubMed:** biomedical literature citations and abstracts	none	**Books:** online books
79	**PubMed Central:** free, full text journal articles	none	**OMIM:** online Mendelian Inheritance in Man
none	**Site Search:** NCBI web and FTP sites	none	**OMIA:** Online Mendelian Inheritance in Animals
45	**Nucleotide:** sequence database (includes GenBank)	none	**UniGene:** gene-oriented clusters of transcript sequences
39	**Protein:** sequence database	none	**CDD:** conserved protein domain database
4	**Genome:** whole genome sequences	12	**3D Domains:** domains from Entrez Structure

Textové prohledávání databází

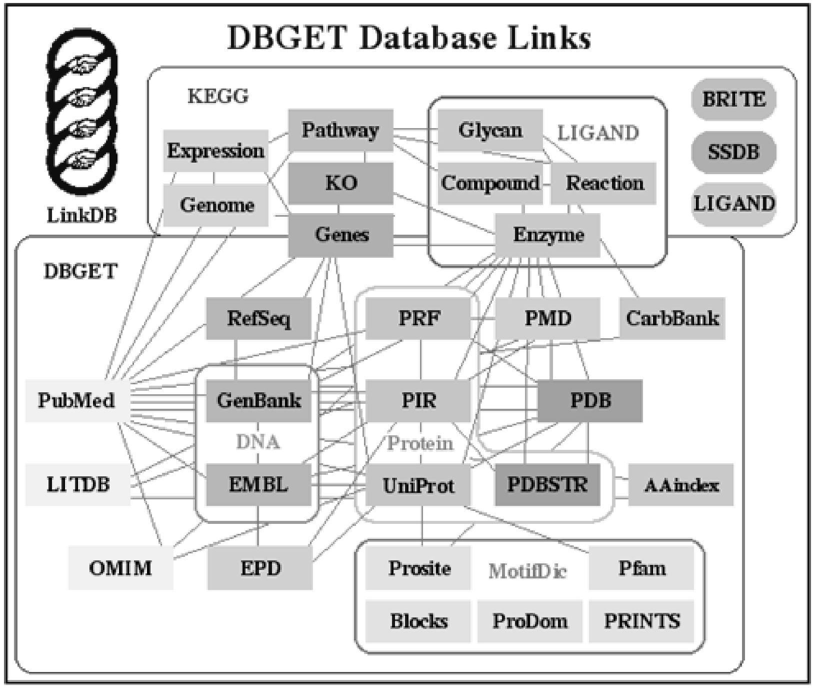
- SRS
  - Vyhledávací systém pro databáze EBI
  - Umožňuje prohledávat i databáze jiných institucí, databáze uživatelů či databáze výsledků vybraných výpočetních nástrojů



Textové prohledávání databází

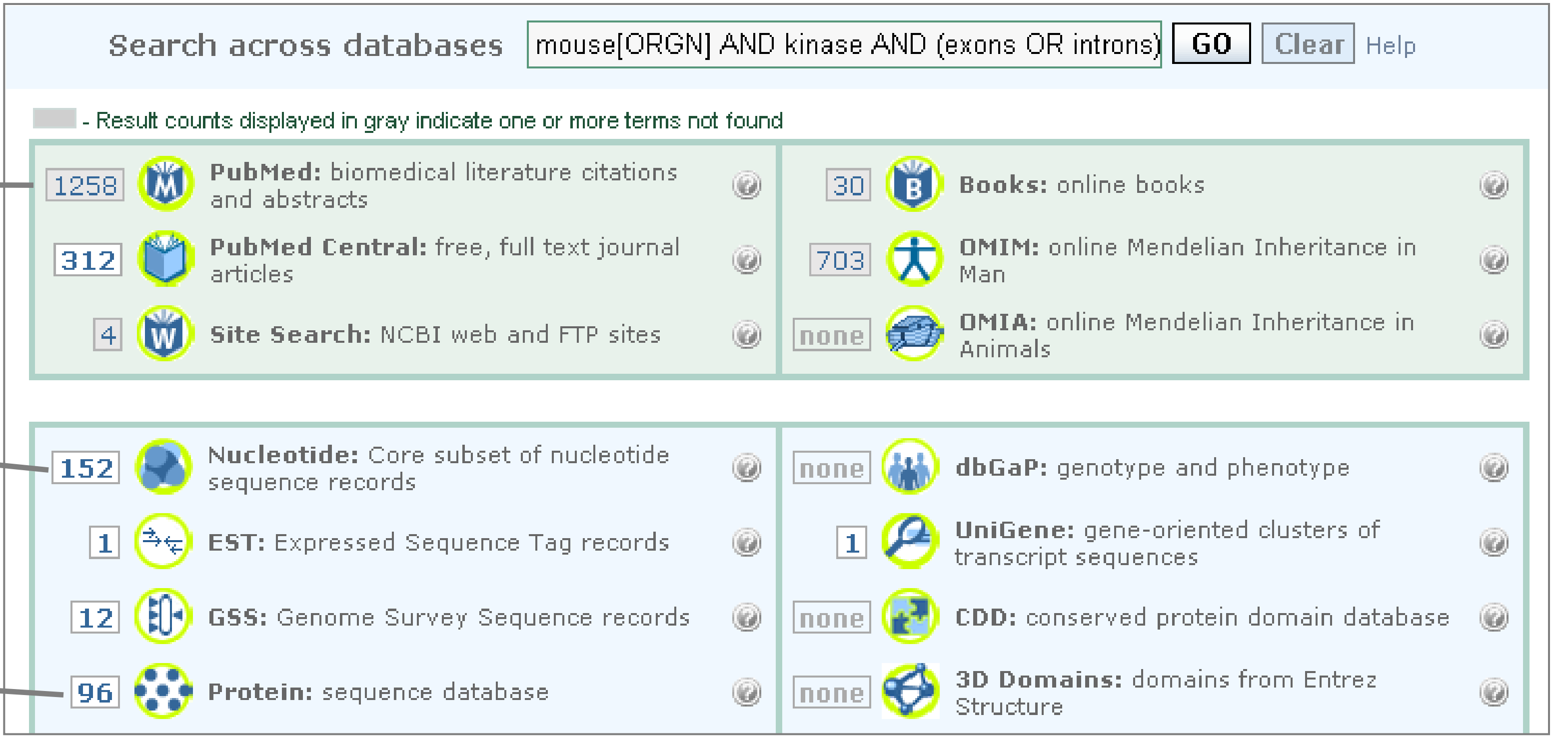
DBGET

- Vyhledávací systém pro databáze LinkDB
- Umožňuje mj. prohledávat databázi metabolických drah KEGG



Ukázka textového prohledávání

- Vyhledávání na základě klíčových slov



Search across databases mouse[ORGN] AND kinase AND (exons OR introns)

▨ - Result counts displayed in gray indicate one or more terms not found

1258	**PubMed:** biomedical literature citations and abstracts	30	**Books:** online books
312	**PubMed Central:** free, full text journal articles	703	**OMIM:** online Mendelian Inheritance in Man
4	**Site Search:** NCBI web and FTP sites	none	**OMIA:** online Mendelian Inheritance in Animals
152	**Nucleotide:** Core subset of nucleotide sequence records	none	**dbGaP:** genotype and phenotype
1	**EST:** Expressed Sequence Tag records	1	**UniGene:** gene-oriented clusters of transcript sequences
12	**GSS:** Genome Survey Sequence records	none	**CDD:** conserved protein domain database
96	**Protein:** sequence database	none	**3D Domains:** domains from Entrez Structure

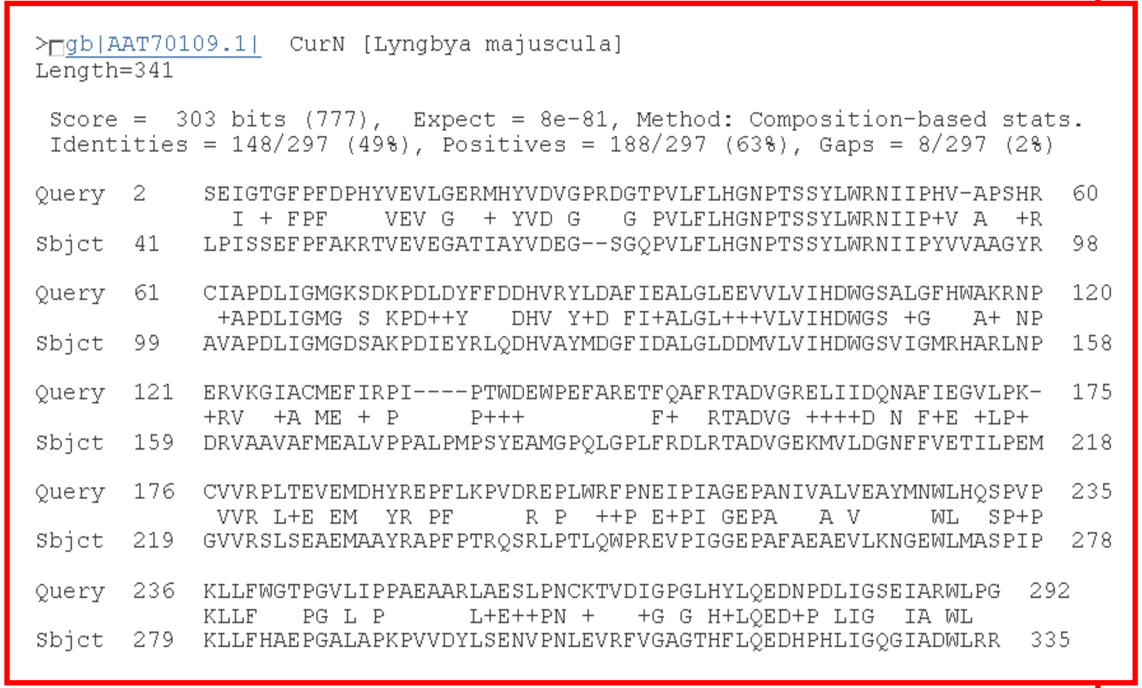
1258

152

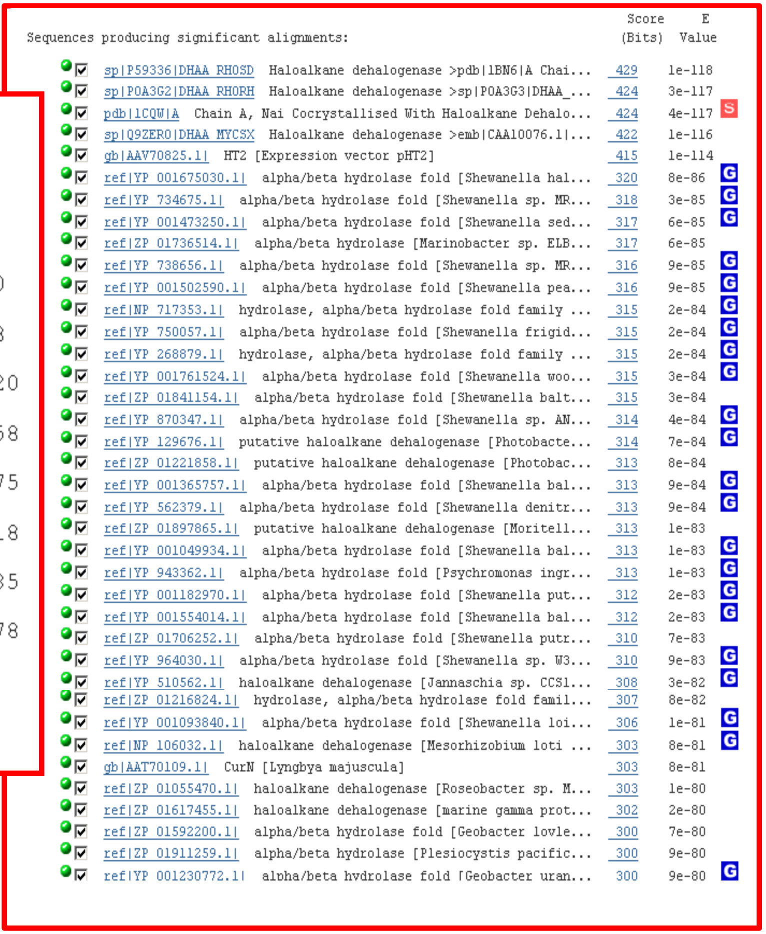
96

Ukázka sekvenčního prohledávání

- Vyhledávání na základě sekvenční podobnosti



```
>☐gb|AAT70109.1|  CurN [Lyngbya majuscula]
Length=341

 Score =  303 bits (777),  Expect = 8e-81, Method: Composition-based stats.
 Identities = 148/297 (49%), Positives = 188/297 (63%), Gaps = 8/297 (2%)

Query  2      SEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYLWRNIIPHV-APSHR  60
                I + FPF    VEV G   + YVD G   G PVLFLHGNPTSSYLWRNIIP+V A  +R
Sbjct  41     LPISSEFPFAKRTVEVEGATIAYVDEG--SGQPVLFLHGNPTSSYLWRNIIPYVVAAGYR  98

Query  61     CIAPDLIGMGKSDKPDLDYFFDDHVRYLDAFIEALGLEEVVLVIHDWGSALGFHWAKRNP  120
               +APDLIGMG S KPD++Y   DHV Y+D FI+ALGL+++VLVIHDWGS +G   A+ NP
Sbjct  99     AVAPDLIGMGDSAKPDIEYRLQDHVAYMDGFIDALGLDDMVLVIHDWGSVIGMRHARLNP  158

Query  121    ERVKGIACMEFIRPI----PTWDEWPEFARETFQAFRTADVGRELIIDQNAFIEGVLPK-  175
              +RV  +A ME + P     P+++         F+  RTADVG ++++D N F+E +LP+
Sbjct  159    DRVAAVAFMEALVPPALPMPSYEAMGPQLGPLFRDLRTADVGEKMVLDGNFFVETILPEM  218

Query  176    CVVRPLTEVEMDHYREPFLKPVDREPLWRFPNEIPIAGEPANIVALVEAYMNWLHQSPVP  235
               VVR L+E EM  YR PF     R P  ++P E+PI GEPA   A V     WL  SP+P
Sbjct  219    GVVRSLSEAEMAAYRAPFPTRQSRLPTLQWPREVPIGGEPAFAEAEVLKNGEWLMASPIP  278

Query  236    KLLFWGTPGVLIPPAEAARLAESLPNCKTVDIGPGLHYLQEDNPDLIGSEIARWLPG  292
              KLLF   PG L P      L+E++PN +   +G G H+LQED+P LIG  IA WL
Sbjct  279    KLLFHAEPGALAPKPVVDYLSENVPNLEVRFVGAGTHFLQEDHPHLIGQGIADWLRR  335
```



```
                                                                            Score    E
Sequences producing significant alignments:                                 (Bits)  Value

● ☑ sp|P59336|DHAA_RHOSD  Haloalkane dehalogenase >pdb|1BN6|A Chai...  429   1e-118
● ☑ sp|P0A3G2|DHAA_RHORH  Haloalkane dehalogenase >sp|P0A3G3|DHAA_...  424   3e-117
● ☑ pdb|1CQW|A  Chain A, Nai Cocrystallised With Haloalkane Dehalo...  424   4e-117 S
● ☑ sp|Q9ZER0|DHAA_MYCSX  Haloalkane dehalogenase >emb|CAA10076.1|...  422   1e-116
● ☑ gb|AAV70825.1|  HT2 [Expression vector pHT2]                       415   1e-114
● ☑ ref|YP_001675030.1|  alpha/beta hydrolase fold [Shewanella hal...  320   8e-86  G
● ☑ ref|YP_734675.1|  alpha/beta hydrolase fold [Shewanella sp. MR...  318   3e-85  G
● ☑ ref|YP_001473250.1|  alpha/beta hydrolase fold [Shewanella sed...  317   6e-85  G
● ☑ ref|ZP_01736514.1|  alpha/beta hydrolase [Marinobacter sp. ELB...  317   6e-85
● ☑ ref|YP_738656.1|  alpha/beta hydrolase fold [Shewanella sp. MR...  316   9e-85  G
● ☑ ref|YP_001502590.1|  alpha/beta hydrolase fold [Shewanella pea...  316   9e-85  G
● ☑ ref|NP_717353.1|  hydrolase, alpha/beta hydrolase fold family ...  315   2e-84  G
● ☑ ref|YP_750057.1|  alpha/beta hydrolase fold [Shewanella frigid...  315   2e-84  G
● ☑ ref|YP_268879.1|  hydrolase, alpha/beta hydrolase fold family ...  315   2e-84  G
● ☑ ref|YP_001761524.1|  alpha/beta hydrolase fold [Shewanella woo...  315   3e-84  G
● ☑ ref|ZP_01841154.1|  alpha/beta hydrolase fold [Shewanella balt...  315   3e-84
● ☑ ref|YP_870347.1|  alpha/beta hydrolase fold [Shewanella sp. AN...  314   4e-84  G
● ☑ ref|YP_129676.1|  putative haloalkane dehalogenase [Photobacte...  314   7e-84  G
● ☑ ref|ZP_01221858.1|  putative haloalkane dehalogenase [Photobac...  313   8e-84
● ☑ ref|YP_001365757.1|  alpha/beta hydrolase fold [Shewanella bal...  313   9e-84  G
● ☑ ref|YP_562379.1|  alpha/beta hydrolase fold [Shewanella denitr...  313   9e-84  G
● ☑ ref|ZP_01897865.1|  putative haloalkane dehalogenase [Moritell...  313   1e-83
● ☑ ref|YP_001049934.1|  alpha/beta hydrolase fold [Shewanella bal...  313   1e-83  G
● ☑ ref|YP_943362.1|  alpha/beta hydrolase fold [Psychromonas ingr...  313   1e-83  G
● ☑ ref|YP_001182970.1|  alpha/beta hydrolase fold [Shewanella put...  312   2e-83  G
● ☑ ref|YP_001554014.1|  alpha/beta hydrolase fold [Shewanella bal...  312   2e-83  G
● ☑ ref|ZP_01706252.1|  alpha/beta hydrolase fold [Shewanella putr...  310   7e-83
● ☑ ref|YP_964030.1|  alpha/beta hydrolase fold [Shewanella sp. W3...  310   9e-83  G
● ☑ ref|YP_510562.1|  haloalkane dehalogenase [Jannaschia sp. CCS1...  308   3e-82  G
● ☑ ref|ZP_01216824.1|  hydrolase, alpha/beta hydrolase fold famil...  307   8e-82
● ☑ ref|YP_001093840.1|  alpha/beta hydrolase fold [Shewanella loi...  306   1e-81  G
● ☑ ref|NP_106032.1|  haloalkane dehalogenase [Mesorhizobium loti ...  303   8e-81  G
● ☑ gb|AAT70109.1|  CurN [Lyngbya majuscula]                           303   8e-81
● ☑ ref|ZP_01055470.1|  haloalkane dehalogenase [Roseobacter sp. M...  303   1e-80
● ☑ ref|ZP_01617455.1|  haloalkane dehalogenase [marine gamma prot...  302   2e-80
● ☑ ref|ZP_01592200.1|  alpha/beta hydrolase fold [Geobacter lovle...  300   7e-80
● ☑ ref|ZP_01911259.1|  alpha/beta hydrolase [Plesiocystis pacific...  300   9e-80
● ☑ ref|YP_001230772.1|  alpha/beta hydrolase fold [Geobacter uran...  300   9e-80  G
```

Problémy prohledávání databází

- Textové vyhledávání
  - ☹ chybné, nepřesné či obecné anotace
  - ☹ synonyma
  - ☹ velký počet falešně pozitivních a falešně negativních výsledků
- Sekvenční vyhledávání
  - ☹ podmínka evoluční příbuznosti
  - ☹ větší nároky na uživatele
  - ☹ potenciálně falešně pozitivní i falešně negativní výsledky

- Vhodné je oba přístupy kombinovat

Reference

- Claverie, J-M., & Notredame, C. (2006) **Bioinformatics for Dummies** (2nd ed.) Wiley Publishing, Hoboken, p. 436.
- Xiong, J. (2006) **Essential Bioinformatics**, Cambridge University Press, New York, p. 352.

- **ENTREZ tutorial:** http://www.ncbi.nlm.nih.gov/Entrez/tutor.html
- **SRS documentation:** http://srs.ebi.ac.uk/srs/doc/index.html
- **NCBI handbook:** http://www.ncbi.nlm.nih.gov/books/NBK21101/
- **UniProtKB manual:** http://www.uniprot.org/manual/

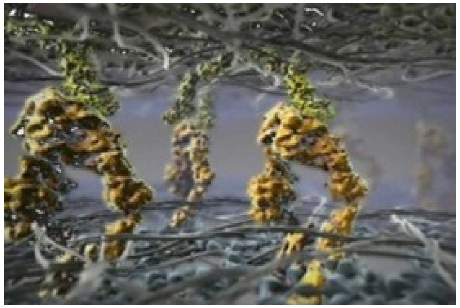
- **NCBI:** http://www.ncbi.nlm.nih.gov/
- **EBI:** http://www.ebi.ac.uk/
- **Pubmed:** http://www.ncbi.nlm.nih.gov/pubmed
- **Web of Science:** http://apps.isiknowledge.com

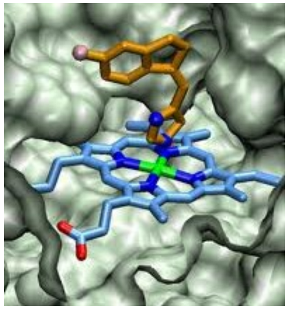
Reference

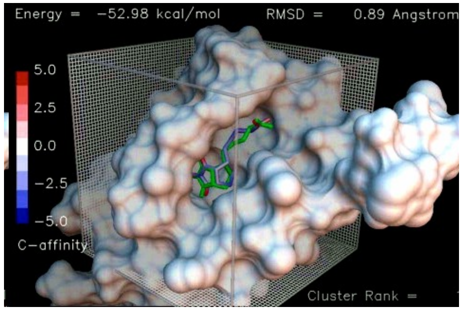
- **GenBank:** http://www.ncbi.nlm.nih.gov/genbank/
- **EMBL-EBI:** http://www.ebi.ac.uk/embl/
- **DDBJ:** http://www.ddbj.nig.ac.jp/
- **UniProt:** http://www.uniprot.org/
- **nrdb:** http://www.ncbi.nlm.nih.gov/protein/
- **wwPDB:** http://www.wwpdb.org/
- **Entrez Genome:** http://www.ncbi.nlm.nih.gov/genome
- **Ensembl:** http://www.ensembl.org
- **GOLD:** http://www.genomesonline.org/
- **Entrez:** http://www.ncbi.nlm.nih.gov/sites/gquery
- **SRS:** http://srs.ebi.ac.uk
- **DBGET:** http://www.genome.jp/dbget/

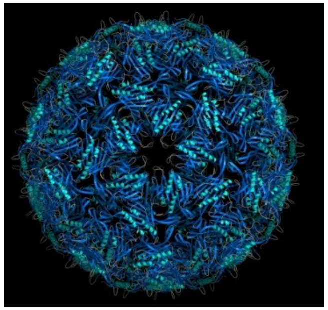
Strukturní biologie Bi9410+9410c

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: Mgr. Jan Brezovský, Ph.D.
- Osnova:
  - struktura, stabilita a dynamika biologických makromolekul
  - makromolekulární interakce a komplexy
  - stanovení a předpověď struktury, identifikace důležitých oblastí
  - stanovení vlivu mutace na strukturu a funkci proteinu
  - aplikace v biologickém výzkumu, návrhu léčiv a biokatalyzátorů



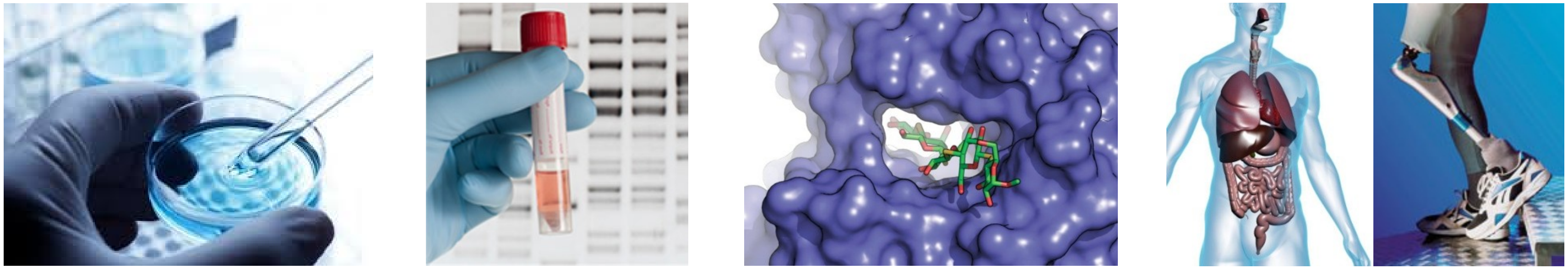






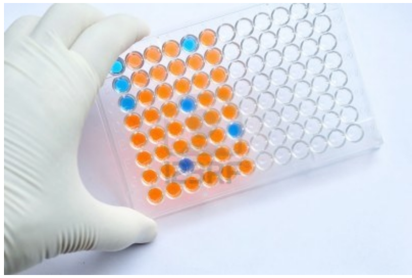
Proteinové inženýrství Bi7410

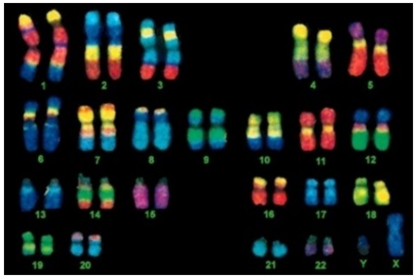
- Období: jaro
- Rozsah: přednáška 1 hodina/týden
- Vyučující: Mgr. Radka Chaloupková, Ph.D.
- Osnova:
    - strukturně-funkční vztahy proteinů
    - metody exprese a purifikace rekombinantních proteinů
    - metody strukturní a funkční analýzy proteinů
    - racionální design, semi-racionální design a řízená evoluce
    - příklady využití proteinového inženýrství

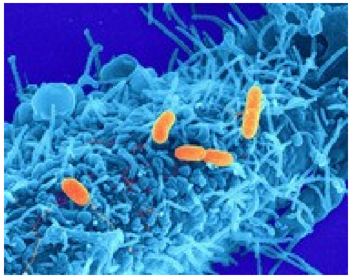


Molekulární biotechnologie Bi7430

- Období: podzim (každoročně)
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Přednášky: Doc. Prokop, Dr. Dvořák, Dr. Bidmanová
- Cvičení: Dr. Bidmanová, Dr. Beerens, Dr. Štěpánková, Mgr. Buryška, Mgr. Chrást
- Osnova:
  - proteinové a metabolické inženýrství
  - molekulární diagnostika a moderní vakcíny
  - buněčná a genová terapie a regenerativní medicína
  - molekulární biotechnologie v průmyslu a zemědělství







Mikrobiologické exkurze Bi6161

- Období: jaro
- Rozsah: 4 dvou až pětihodinové exkurze
- Vyučující: Mgr. Šárka Bidmanová, Ph.D.
- Exkurze:
    - Pivovar Starobrno – http://www.starobrno.cz/
    - Erba Lachema – https://www.erbalachema.com/
    - Čistírna odpadních vod – http://www.vodarenska.cz/
    - Kompostárna – http://www.kompostarna-blansko.cz/

Biotechnologické exkurze Bi7171

- Období: podzim
- Rozsah: 4 jednodenní exkurze (8.2.-11.2.2015)
- Vyučující: Mgr. Šárka Bidmanová, Ph.D.
- Exkurze:
  - Biotechnologické centrum INBIT – www.jic.cz/inbit
  - Bioveta – www.bioveta.cz
  - BioVendor – www.biovendor.cz
  - Contipro Group – www.contipro.com