Genetika a taxonomie virů

MVDr. Jana Kvičerová, Ph.D.

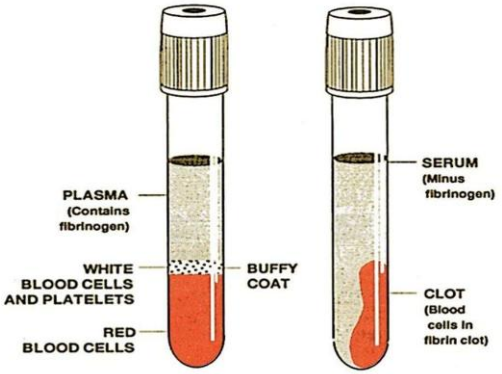
Katedra parazitologie
Přírodovědecká fakulta JU
České Budějovice

2.-3.12. 2024

Zásady odběru vzorků pro izolaci nukleových kyselin virů

Druhy vzorků (tkání):

- krev
- trus / stolice
- moč
- sliny
- stěry
- orgány (slezina, ledviny, játra, plíce, srdce, mozek)



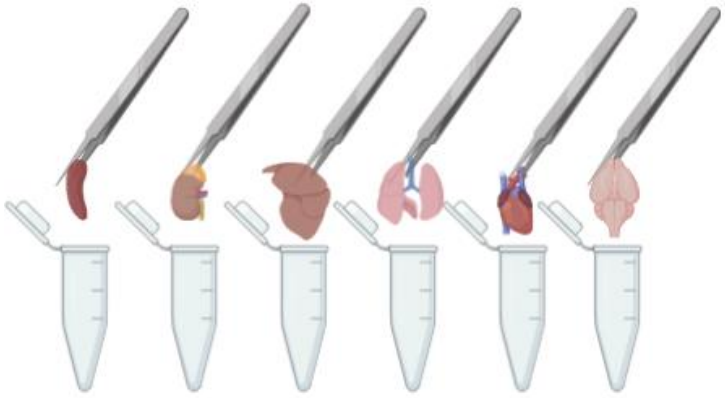
V humánní medicíně:

- krev

- stolice
- stěry

U zvířat:

- krev
- trus
- **orgány (postmortálně)**



Zásady odběru vzorků pro izolaci nukleových kyselin virů

Obecné zásady odběru vzorků pro izolaci NK virů:

- vzorky **co nejčerstvější**

  in vivo – není problém

  post mortem – co nejdříve

- **včasné zpracování** (krev na sérum – co nejdříve) či **fixace/konzervace**
- **sterilní odběr** (sterilní nástroje, sterilní zkumavky, případně prostředí flowboxu)
- **použití rukavic** (vždy), případně respirátoru
- **(biohazard boxy)**

Zásady odběru vzorků
pro izolaci nukleových kyselin virů

Krevní sérum:

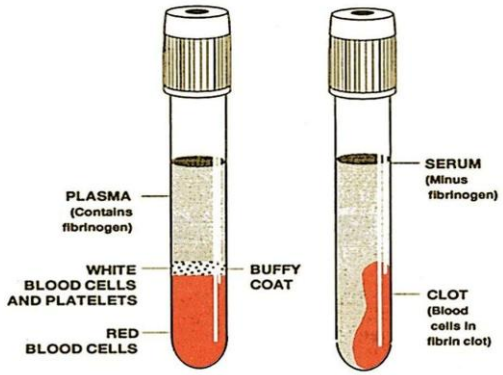
- odběr plné žilní krve, její vysrážení a centrifugace

- **centrifugace 5-10 min při 1500 otáčkách/min**

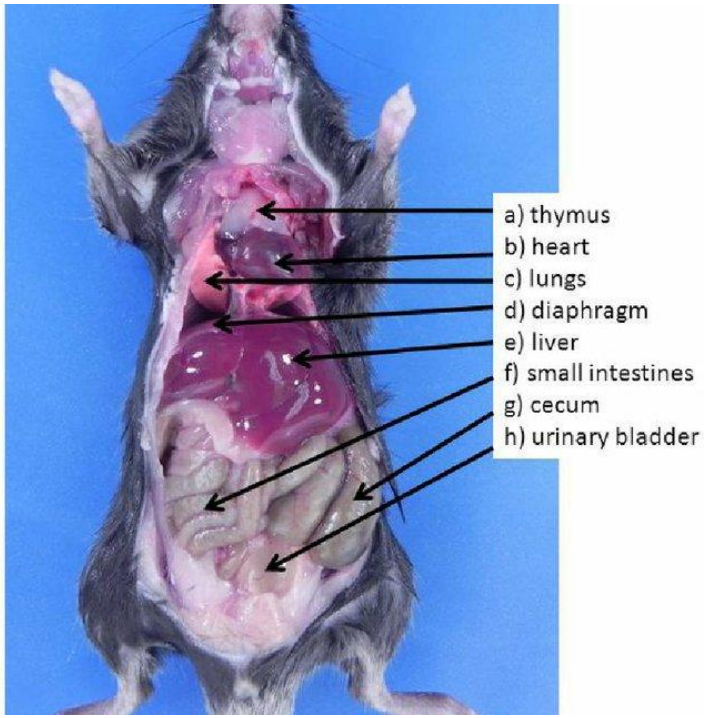
- **odebrat sérum** (tj. žlutavou tekutinu vyloučenou na povrch)

- umístění séra do sterilní mikrozkumavky a **zamrazení**

→ odstranění krevních elementů, fibrinogenu, a většiny koagulačních faktorů



Orgány a zásady jejich odběru (*Mus musculus*)



<u>Sterilně vyjmout:</u> 1. slezina, 2. ledviny, 3. játra, 4. srdce, 5. plíce, 6. mozek

Média pro fixaci vzorků tkání pro virovou diagnostiku

- **bez fixačního média** (tzv. nasucho) – u čerstvých odběrů

výhody: finanční, časová úspora při odběru, bez nutnosti eliminace média před izolací nukleových kyselin
nevýhody: nutnost okamžitého zamrazení

- **RNAlater**

výhody: bez nutnosti okamžitého zamrazení (např. při odběrech vzorků v terénu v delší vzdálenosti od laboratoře)
nevýhody: finanční, časové, nutnost eliminace před izolací NK

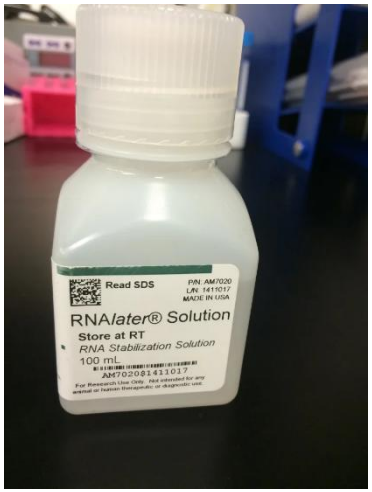
- **DNA/RNA Shield**

výhody: bez nutnosti okamžitého zamrazení (např. při odběrech vzorků v terénu v delší vzdálenosti od laboratoře),
lze uchovávat až 30 dní při pokojové teplotě (!)
nevýhody: finanční, časové, nutnost eliminace před izolací NK

RNAlater


RNA Later Solution

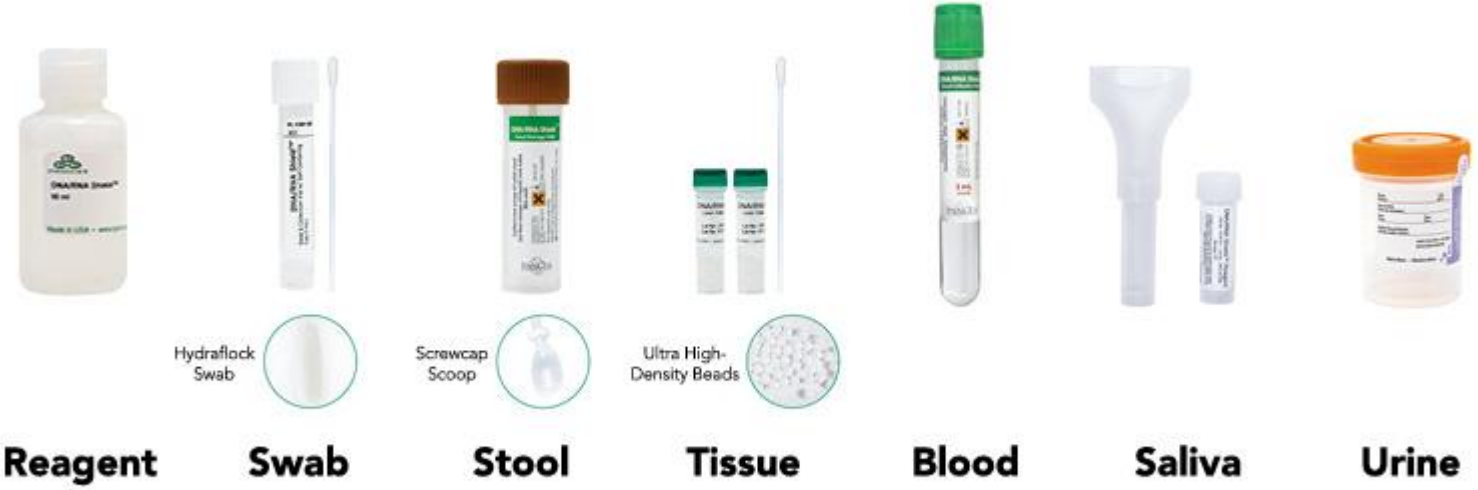

SIGMA
RNAlater™


Read SDS
RNAlater® Solution
Store at RT
RNA Stabilization Solution
100 mL


71969
100 ml
RNA Stabilizing Solution
SRL
71969
RNA Stabilizing


invitrogen
RNAlater™ Soln.

DNA/RNA Shield



DNA/RNA Shield

DNA/RNA Shield™

ZYMO RESEARCH
The Beauty of Science is to Make Things Simple

Catalog Nos: R1100-50 (50 ml)
R1100-250 (250 ml)
R1200-25 (25 ml) 2X concentrate
R1200-125 (125 ml) 2X concentrate
Storage: Reagent stable -80°C to 70°C.

Features

- Eliminates cold-chain. Ensures DNA/RNA stability during sample transport/storage at ambient temperatures.
- Inactivates infectious agents *(viral, bacterial, fungal and parasitic)*.
- Easy sample processing. DNA/ RNA can be isolated directly without sample precipitation or reagent removal *(compatible with most DNA and RNA purification kits/ high-throughput workflows)*.

Sample Stability

Samples in DNA/RNA Shield™ are stable for long periods of time before subsequent purification of high-quality DNA/RNA.

Temperature	Time
-20°C and below	Indefinite
4°C – 25°C (ambient)	Minimum 30 days
35-40°C	Up to 7 days

Instructions for Sample Storage/Transport

Unsure of sample type? Just add 9 volumes of DNA/RNA Shield™. If sample is viscous, add more DNA/RNA Shield™.

Suggested Volumes

DNA/RNA Shield™	300 µl	600 µl
Cell pellets	10^6	10^7
Tissue	30 mg	60 mg
Biological liquid	100 µl	200 µl

DNA/RNA Shield™ (2X concentrate):
(1) Use only for immediate sample processing/purification.
(2) For sample storage, dilute with nuclease-free water (1:1) prior to use.

DNA/RNA Purification

Samples in DNA/RNA Shield™ are directly compatible with:
- All Zymo Research Purification Kits.
- Most kits and workflows from Qiagen, Roche, Thermo-Fisher, Macherey Nagel, etc.
- High-throughput automated workflows from Hamilton, Tecan, bioMérieux, PerkinElmer, Eppendorf, Promega, etc.

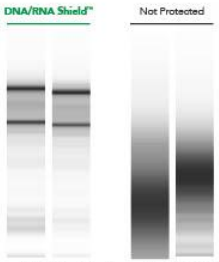
Recommended Purification Kits for DNA/RNA Shield™:

Product / Format	Size	Cat. No.
Quick-DNA/RNA™ Miniprep Plus *Spin Column*	50 preps	D7003
Quick-DNA/RNA™ MagBead *Magnetic Bead*	96 preps 384 preps	R2130 R2131

ZYMO RESEARCH CORP.
Phone: (949) 679-1190 • Toll Free: (888) 882-9682 • Fax: (949) 266-9452 • info@zymoresearch.com • www.zymoresearch.com

Protection from freeze-thaw damage

Are your samples protected from the stress caused by freeze-thaw cycling? DNA/RNA Shield™ is not only beneficial for sample transport, but also for the long-term storage of biological samples. DNA/RNA Shield™ protects DNA/RNA from multiple freeze-thaw cycles, even in the most complex of samples (i.e., whole blood).



High quality RNA from blood stored in DNA/RNA Shield™ that was freeze-thawed from -80°C to room temperature.

Pathogen Inactivation

DNA/RNA Shield abides by Center for Disease Control's (CDC) guidelines for pathogen inactivation.

Validated organisms by various research groups:

Bacteria	Viruses	Yeast & Eukaryotes
B. subtilis	Parvovirus	*C. albicans*
E. faecalis	Chikungunya Virus	*C. neoformans*
E. coli	Dengue Virus	*S. cerevisiae*
L. fermentum	Ebolavirus	*P. malariae*
L. monocytogenes	Herpes Simplex Virus-1	
M. tuberculosis	Herpes Simplex Virus-2	
P. aeruginosa	Influenza A	
S. enterica	Rhinovirus	
S. aureus	MERS-coronavirus	
S. pneumoniae	West Nile Virus	
X. fastidiosa		

COMPLIANT CDC Centers for Disease Control and Prevention

CERTIFIED BY Eurovir® Hygiene Labor Antivirale Validierung & Rabies

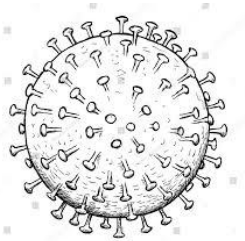
DNA/RNA Shield™ Collection Devices

Product	Description	Recommended processing
Fecal collection tube R1101	20 x 76 mm container prefilled with 9 ml DNA/RNA Shield™ for the direct collection of up to 1 gram or 1 ml stool. Collection spoon is included in the container screwcap.	ZymoBIOMICS™ DNA/RNA D4301 (spin-column); D4302 (magbead)
Swab & collection tube R1106, R1107, R1108, R1109	12 x 80 mm self-centering screwcap container filled with DNA/RNA Shield™ (1 ml, 2 ml) and a sterile swab for specimen collection.	
Saliva collection kit R1210	Saliva collection tube with funnel and tube filled with 2 ml of DNA/RNA Shield™ for the direct collection of 2 ml saliva.	*Quick*-DNA/RNA™ Plus D7001 (spin-column); R2130 (magbead)
Collection & lysis tube R1102, R1103, R1104, R1105	2 ml screw-cap tube with DNA/RNA Shield™ and Bashing Beads (lysis tubes: microbe, tissue, pathogen) for the collection and homogenization of tough-to-lyse samples.	
Blood collection tube R1150	16 x 100 mm evacuated blood tube filled with 6 ml of DNA/RNA Shield™ for the direct collection of 3 ml whole-blood (human or animal).	*Quick*-DNA/RNA™ Blood Tube R1151 (spin-column); R2130 (magbead)

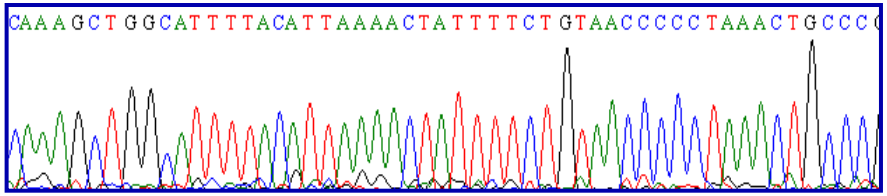
For bulk reagent, large volume, and custom device requests, please contact Zymo Research directly.

ZYMO RESEARCH CORP.
Phone: (949) 679-1190 • Toll Free: (888) 882-9682 • Fax: (949) 266-9452 • info@zymoresearch.com • www.zymoresearch.com

Ver. 2.0.0

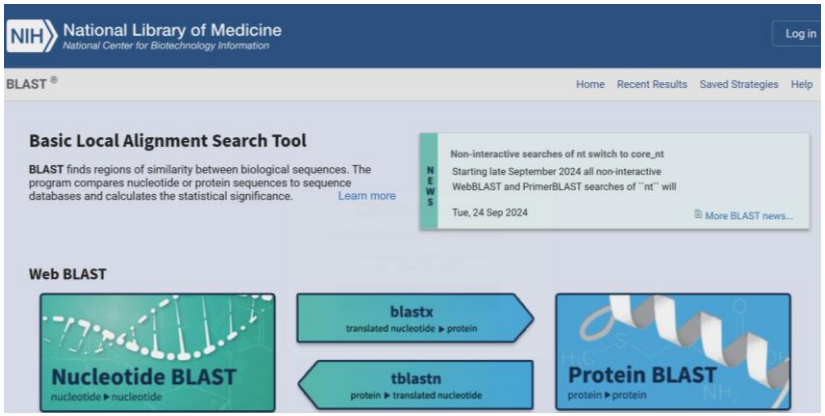


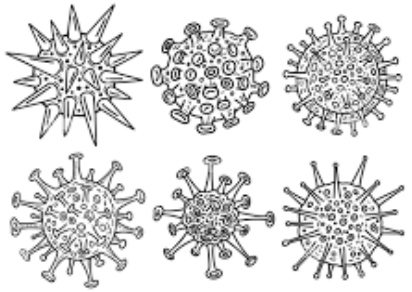
Izolace DNA / RNA, PCR, sekvenování



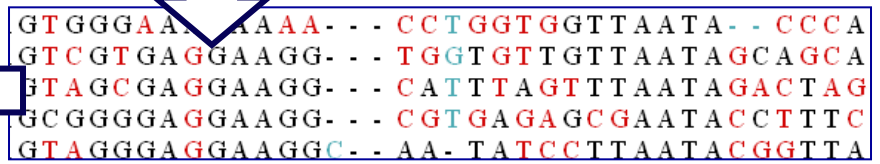
Organismus, který chci fylogeneticky charakterizovat (měl bych vědět proč)

Vyhledání homologických sekvencí pro další taxony (měl bych vědět, pro které)

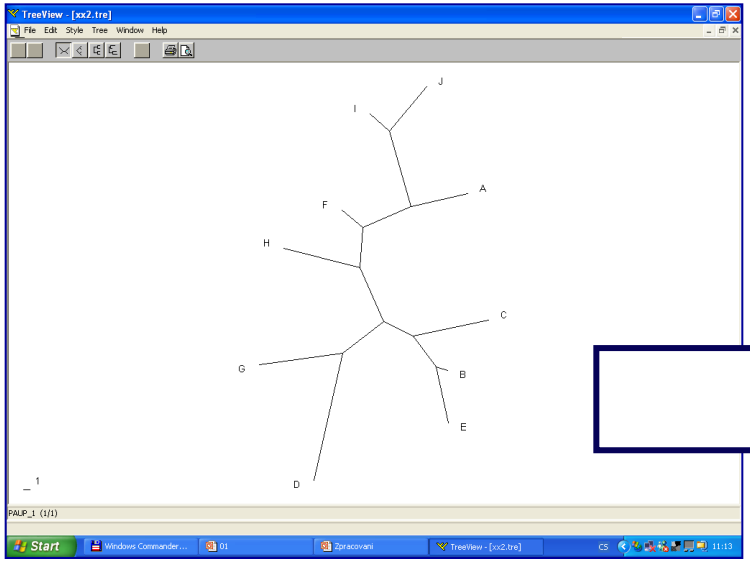




Vytvoření alignmentu

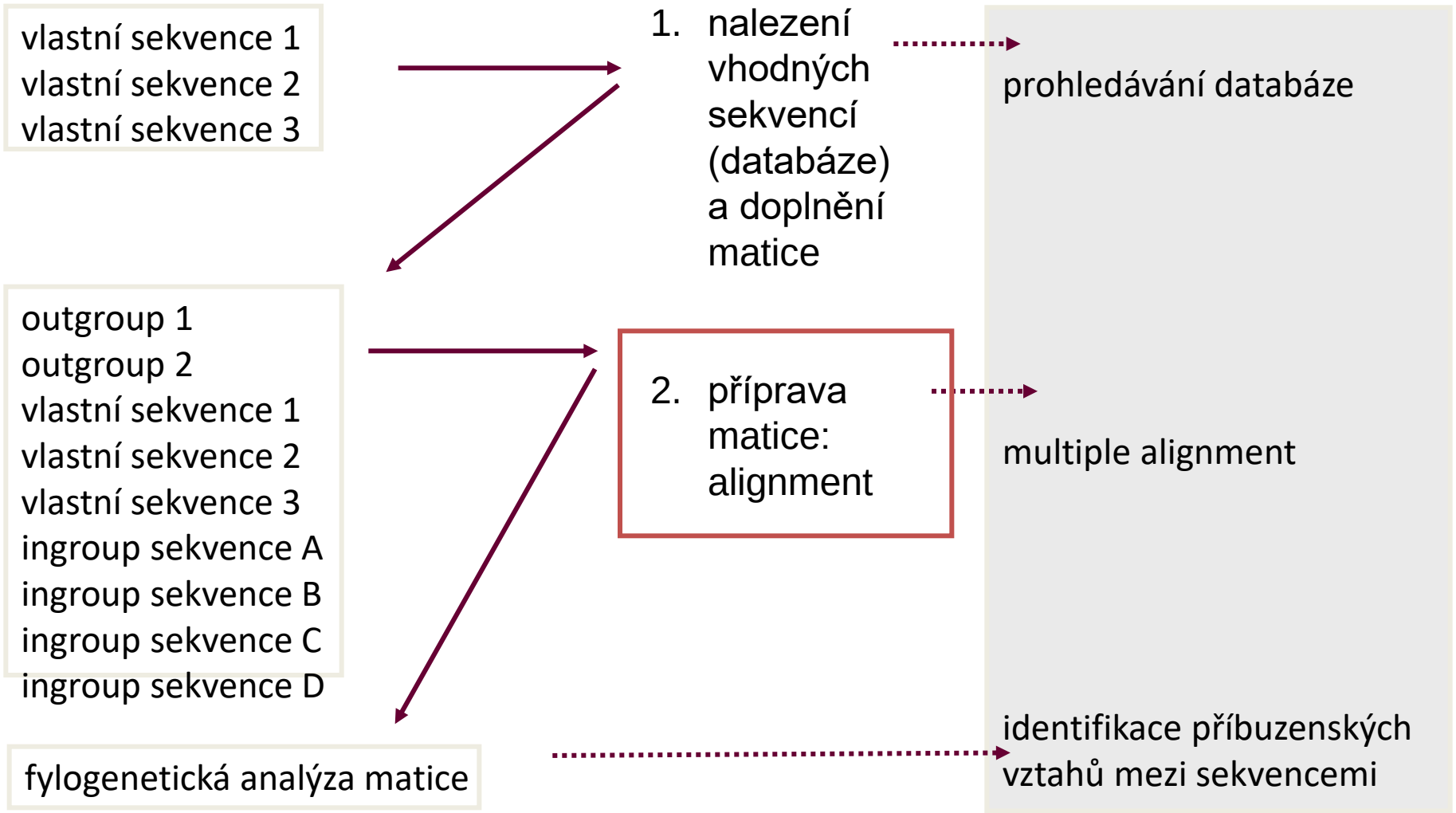


Fylogenetická analýza zvolenou metodou



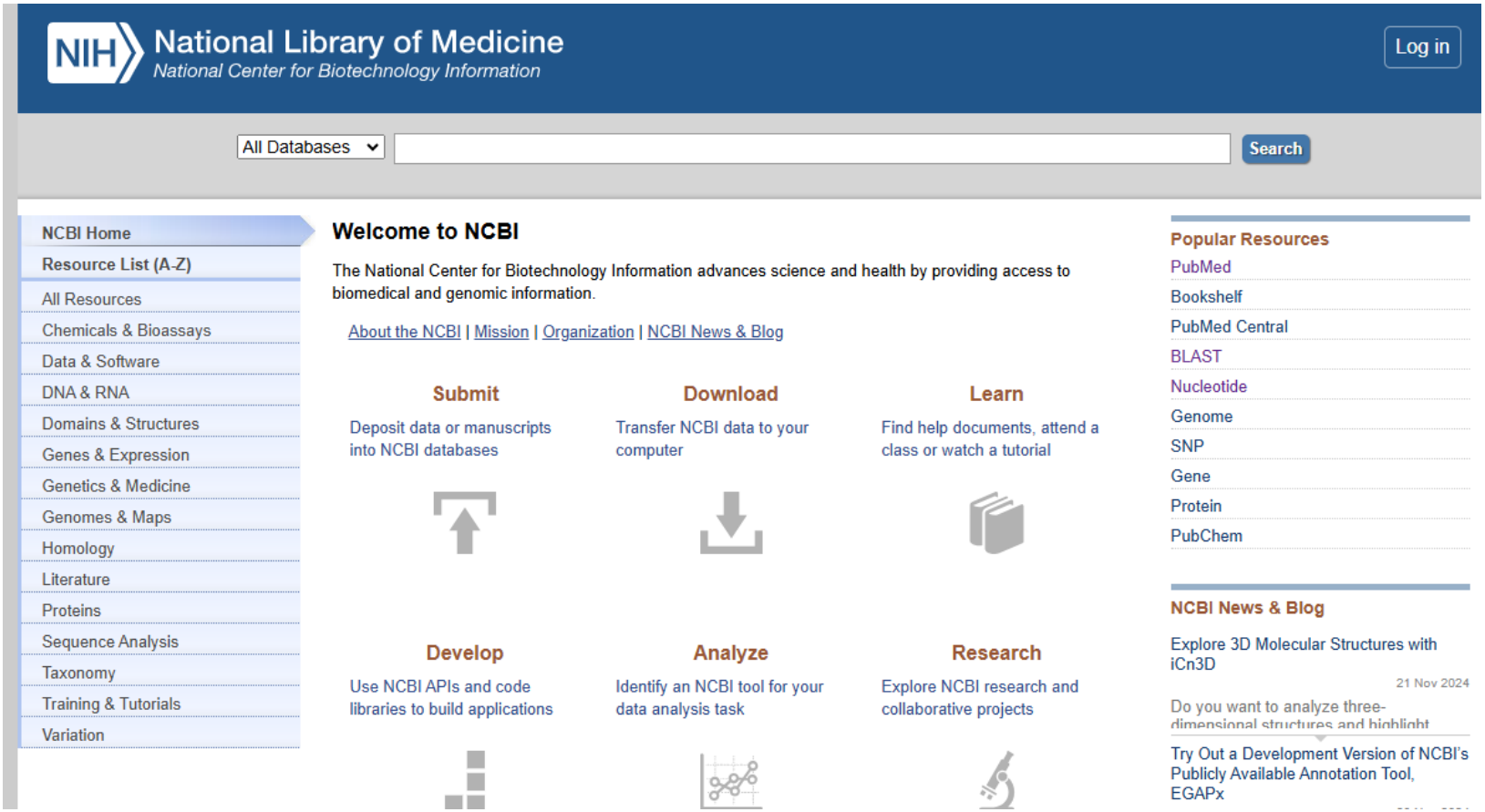
To hlavní a podstatné:
interpretace fylogenetických vztahů

Fylogenetická analýza



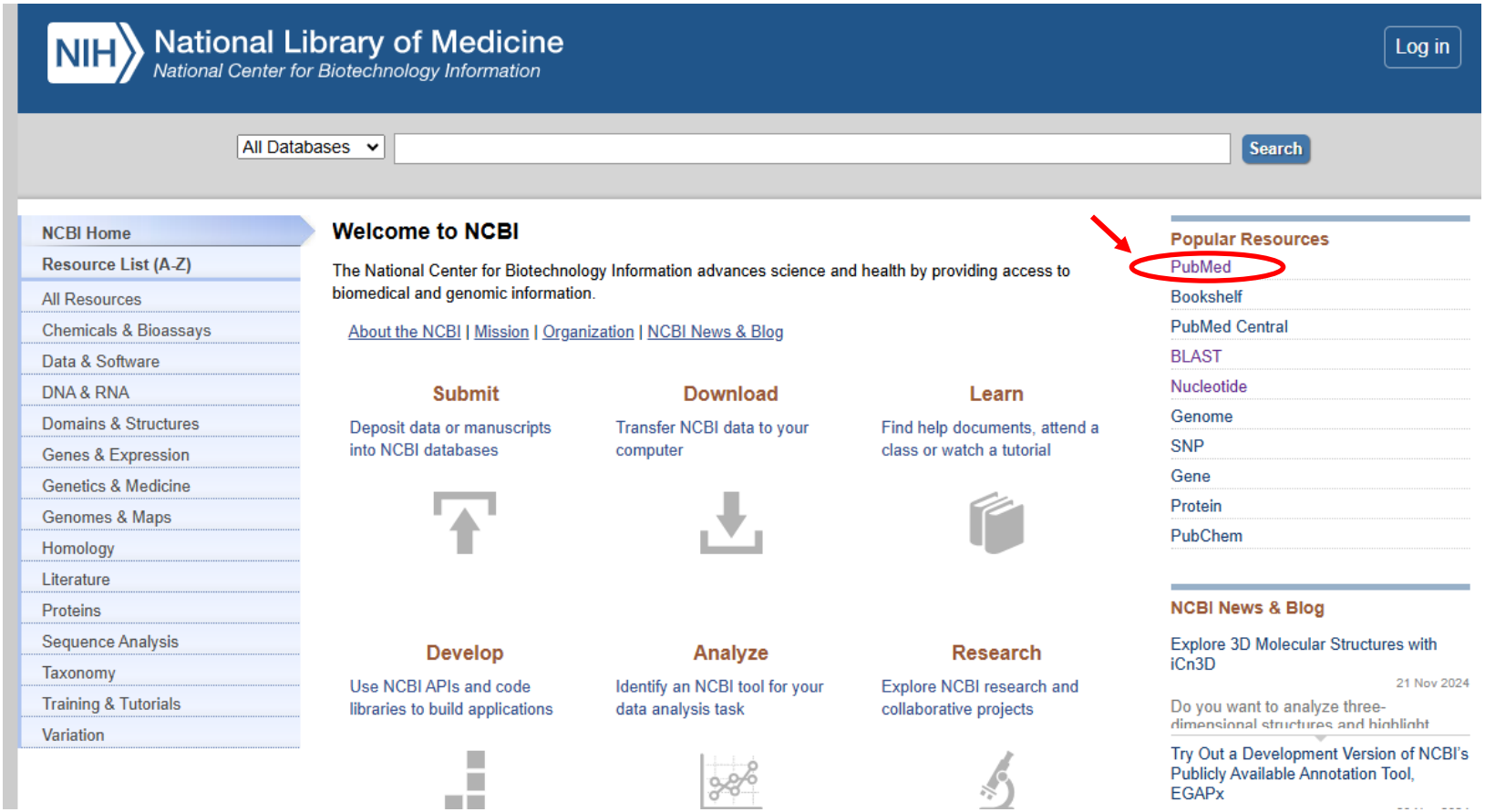
Databáze GenBank

www.ncbi.nlm.nih.gov



Databáze GenBank – PubMed

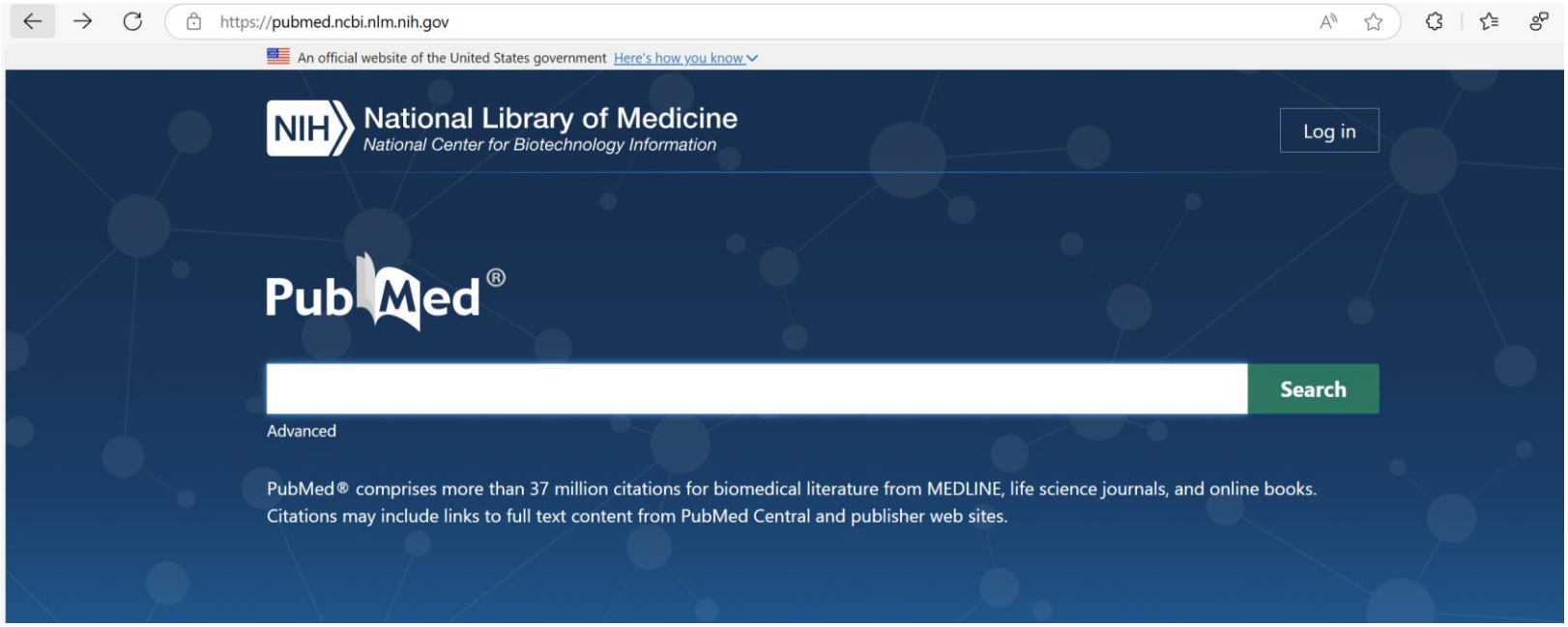
→ vyhledávání publikací



www.ncbi.nlm.nih.gov

Databáze GenBank – PubMed

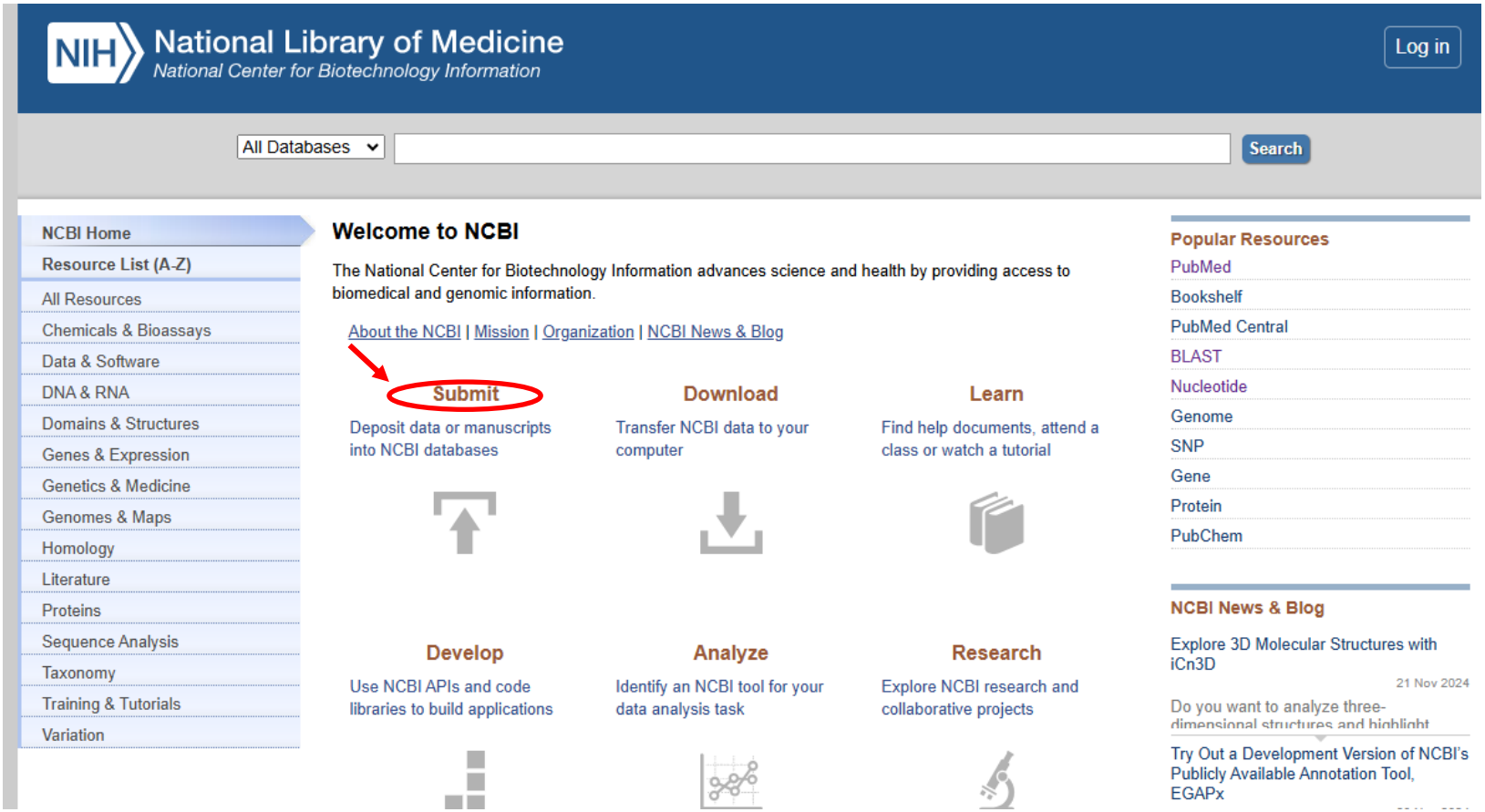
→ vyhledávání publikací



www.ncbi.nlm.nih.gov

Databáze GenBank – Submit (BankIt, Sequin, SRA…)

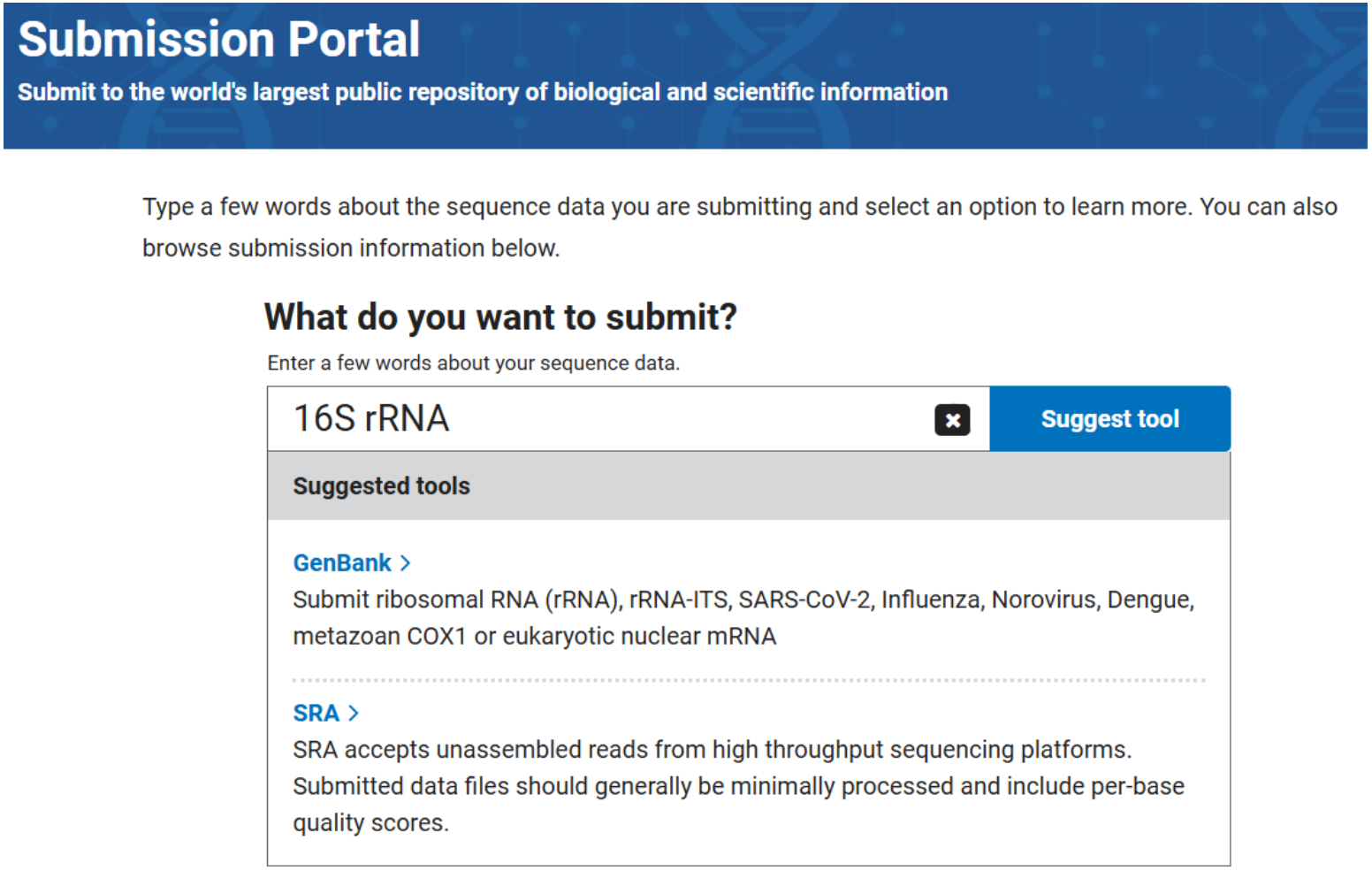
→ vkládání sekvencí (po přihlášení heslem a uživatelským jménem)



www.ncbi.nlm.nih.gov

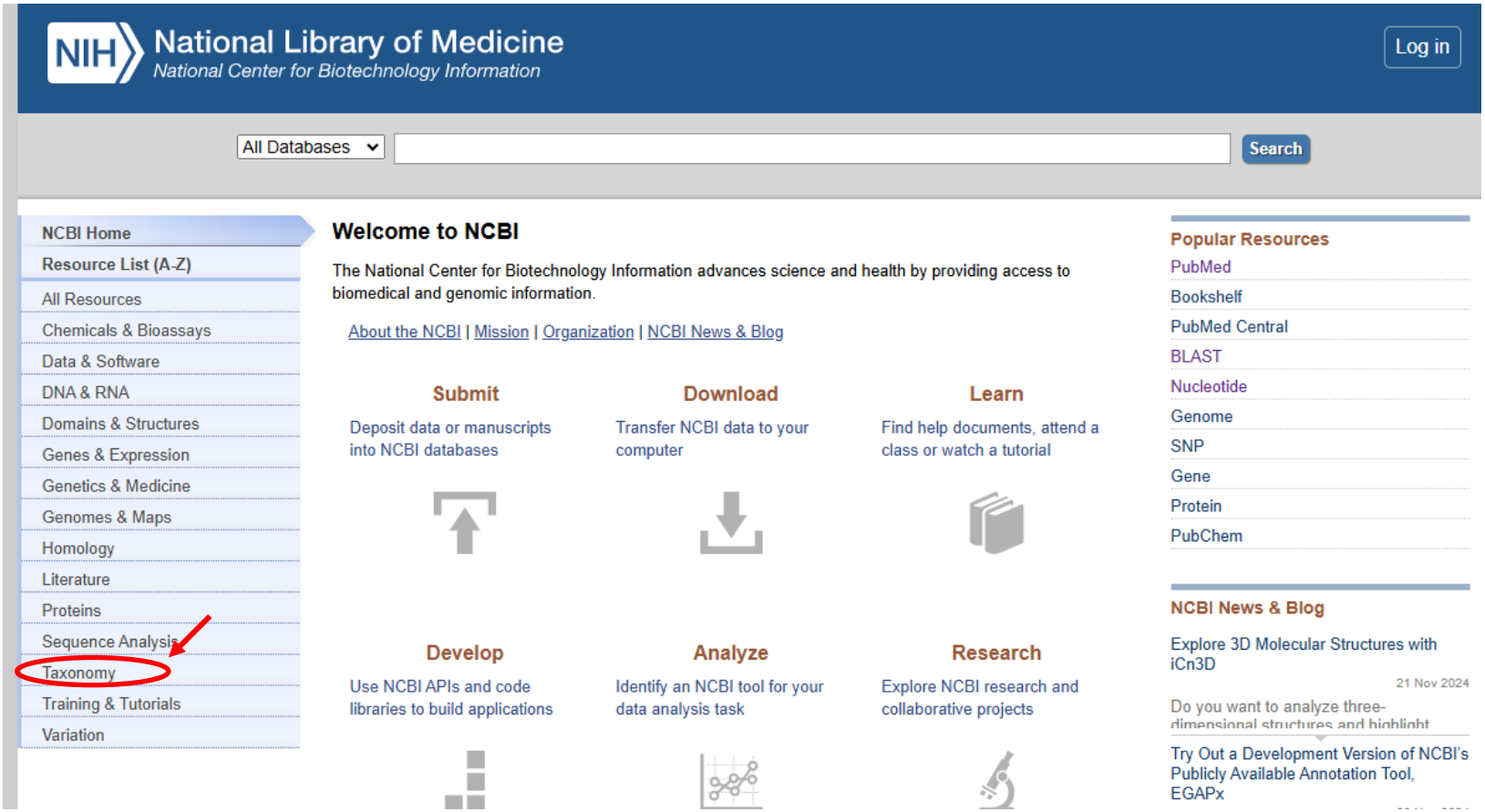
Databáze GenBank – Submit (BankIt, Sequin, SRA…)

→ vkládání sekvencí (po přihlášení heslem a uživatelským jménem)



Databáze GenBank – Taxonomy

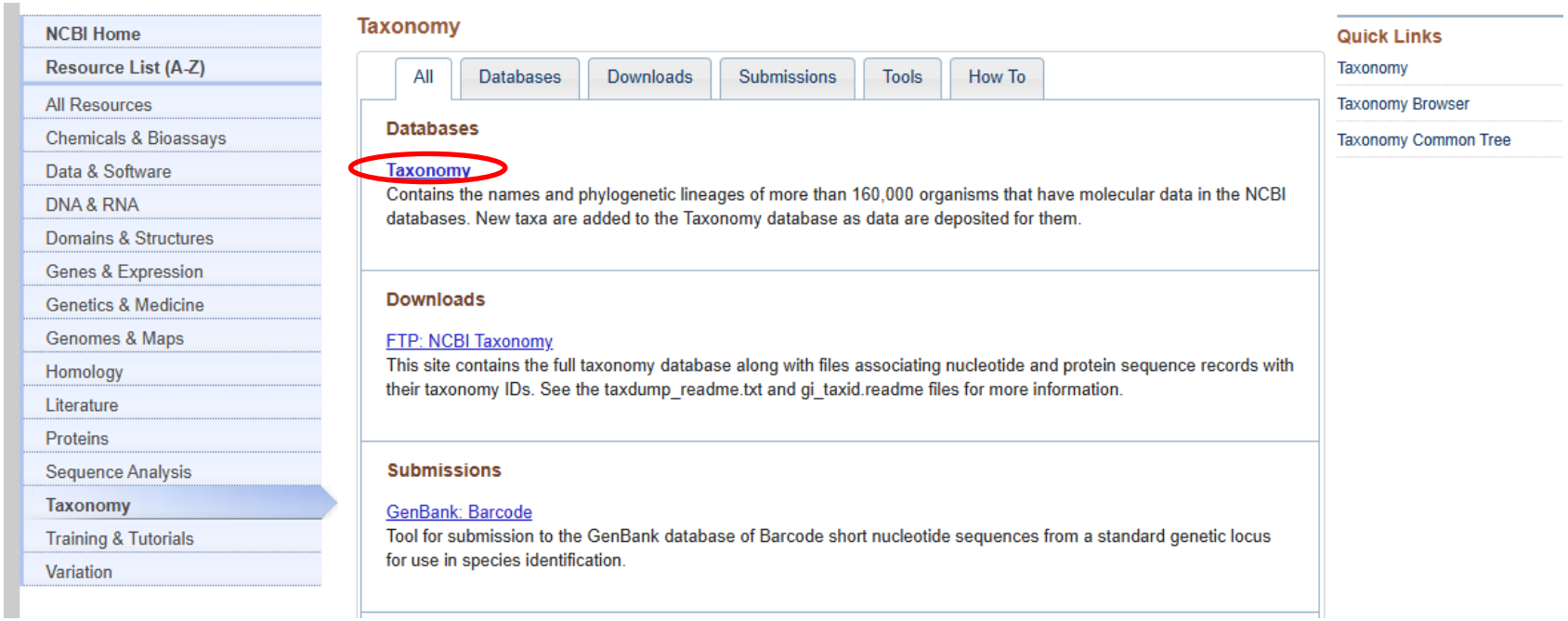
→ vyhledávání sekvencí podle názvu taxonu (viru)



www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

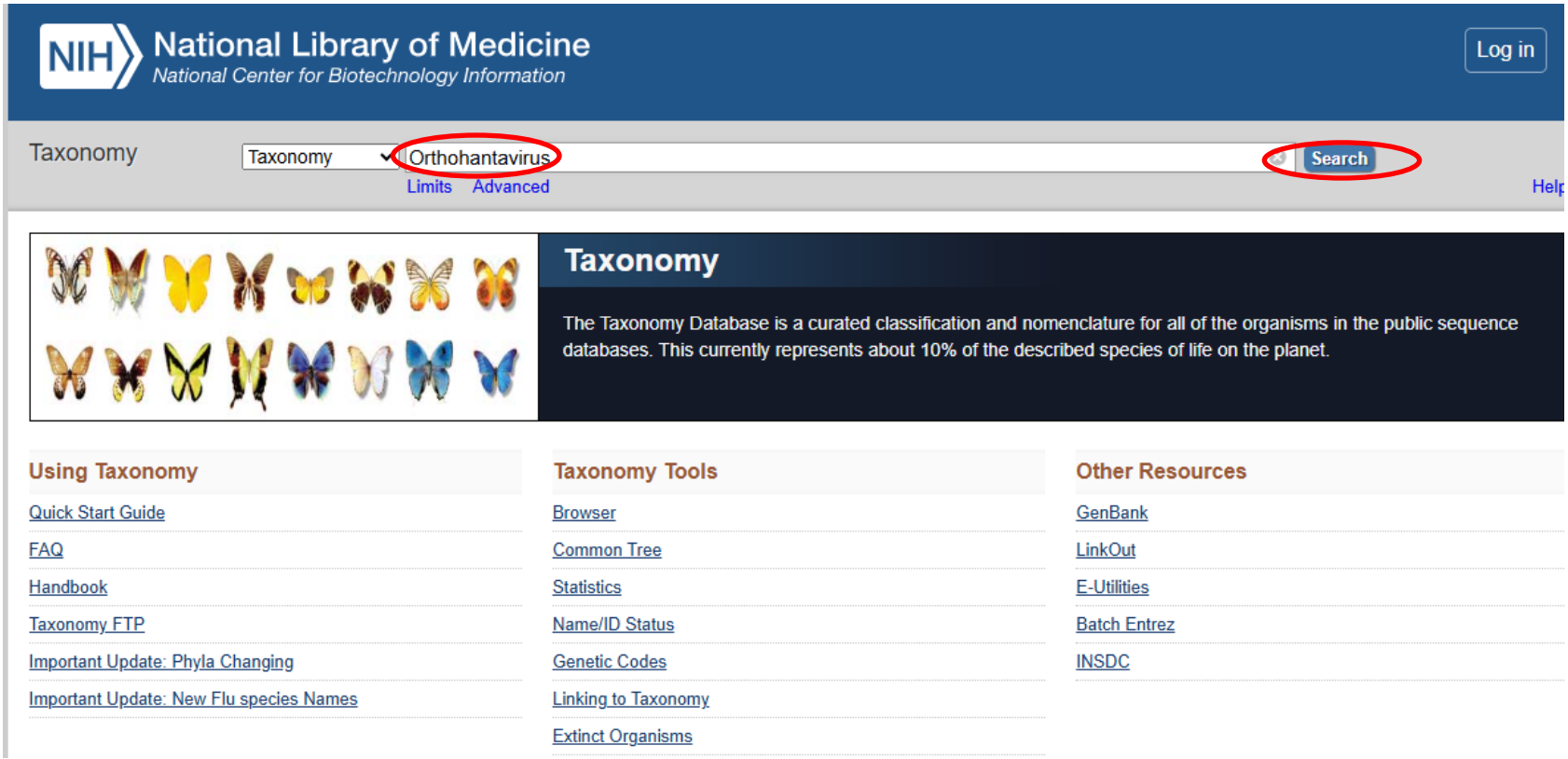
→ vyhledávání sekvencí podle názvu taxonu (viru)



www.ncbi.nlm.nih.gov

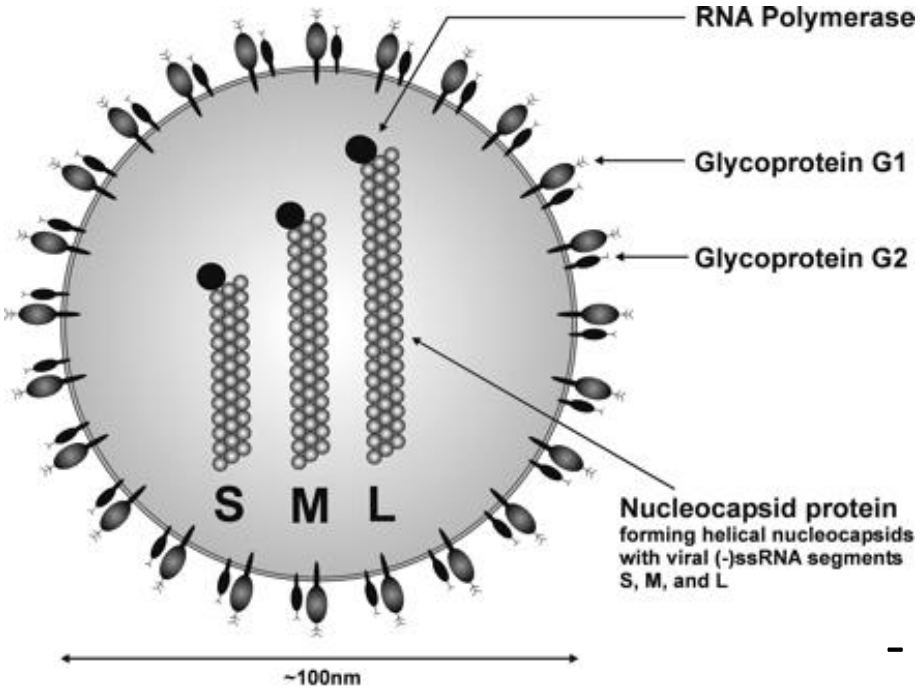
Databáze GenBank – Taxonomy

→ vyhledávání sekvencí podle názvu taxonu (viru)



www.ncbi.nlm.nih.gov

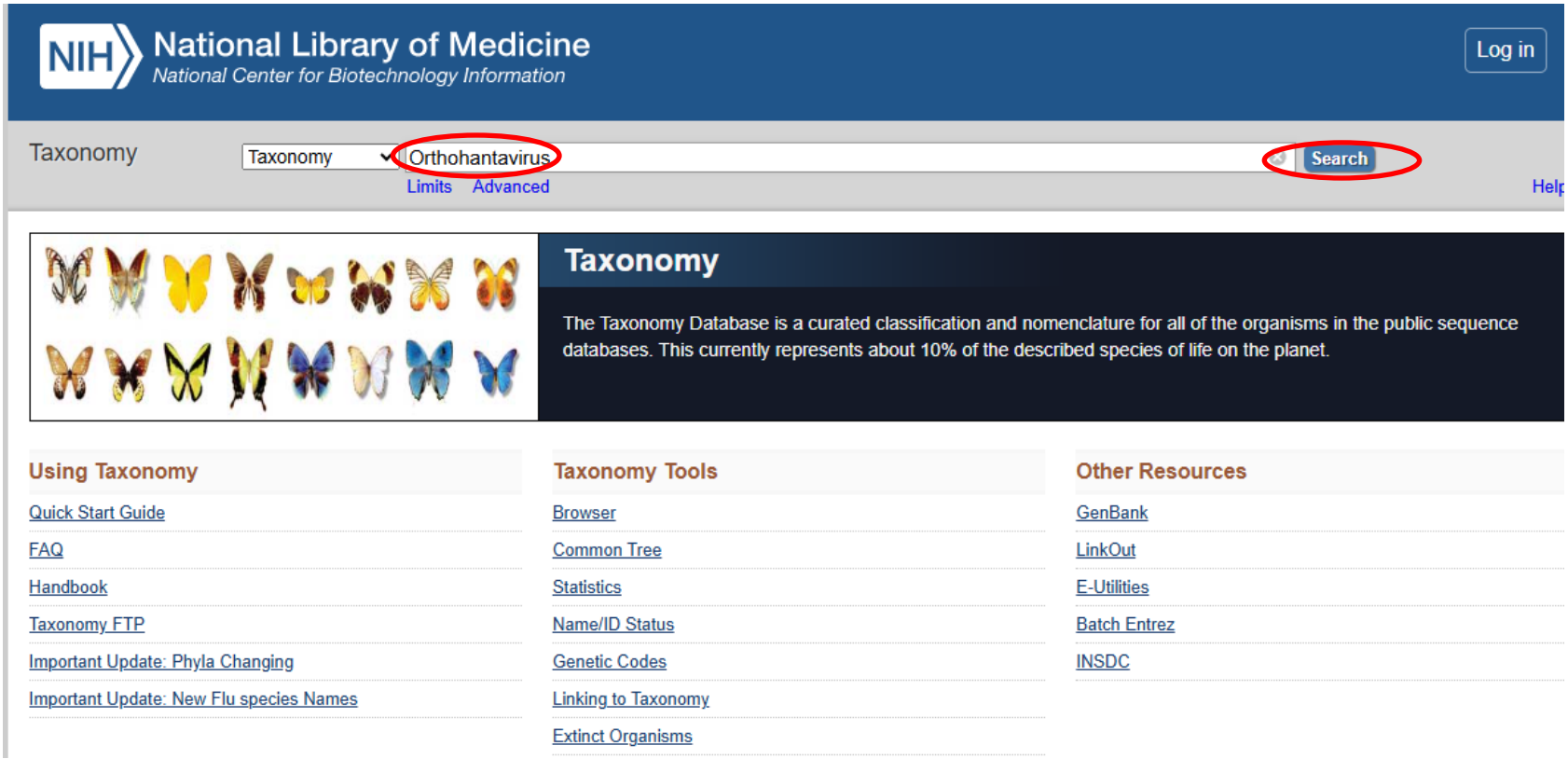
rod *Orthohantavirus* (*Elliovirales*: *Hantaviridae*)



- ssRNA
- negativní polarita
- segmentovaný genom (S, M a L segment)
- M segment → reassortment
- zoonotický potenciál

Databáze GenBank – Taxonomy

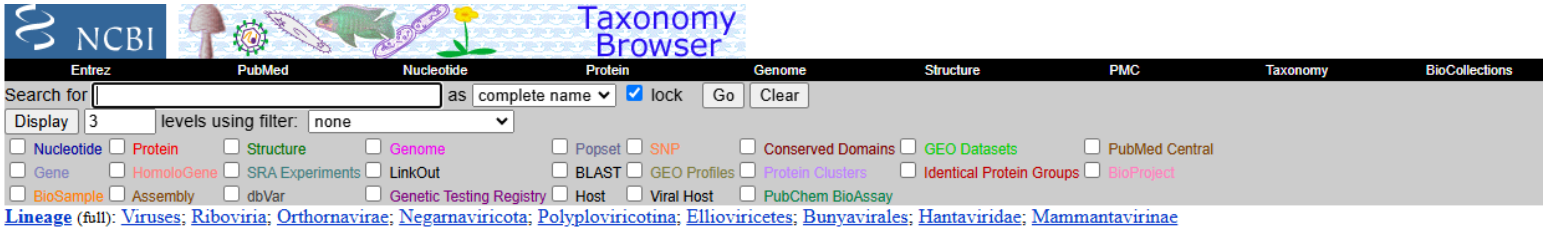
→ vyhledávání sekvencí podle názvu taxonu (viru)



www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ vyhledávání sekvencí podle názvu taxonu (viru)

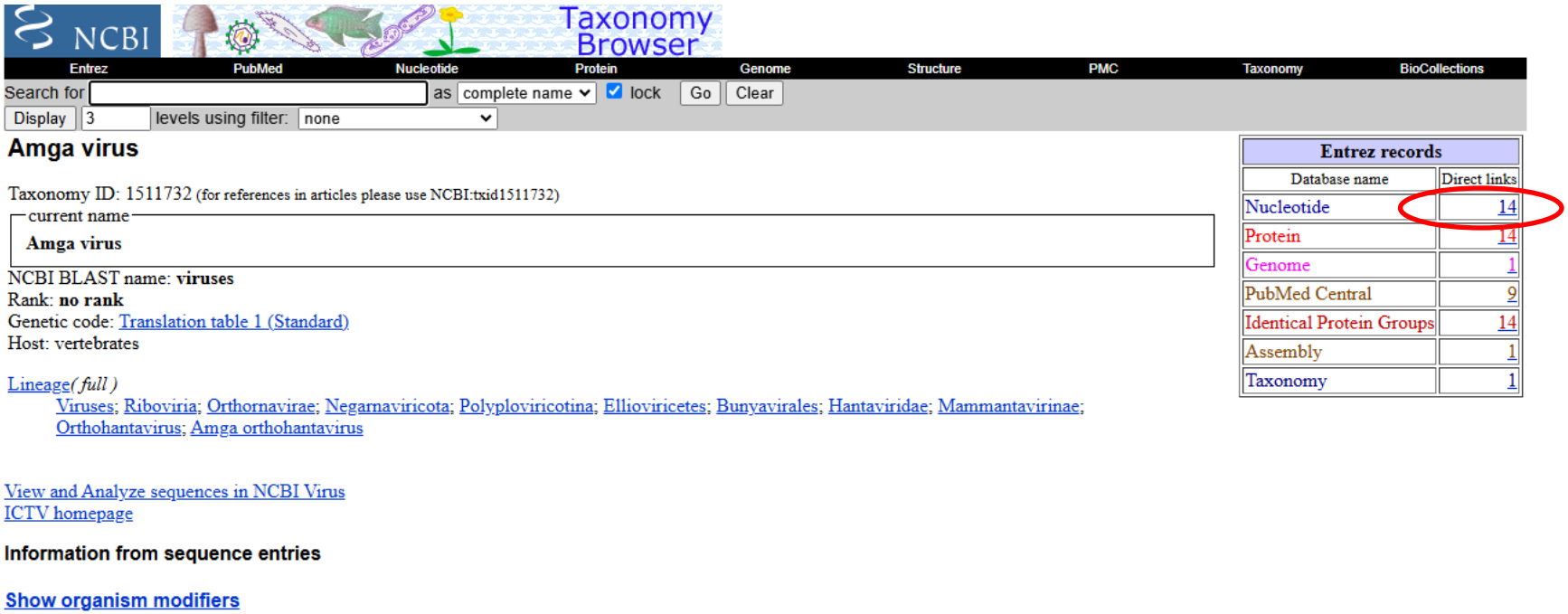


- Orthohantavirus *Click on organism name to get more information.*
  - Amga orthohantavirus
    - Amga virus
  - Orthohantavirus andesense
    - Araraquara virus
    - Araraquara-like virus strain P5/Cajuru
    - Bermejo virus
    - Castelo dos Sonhos virus
    - Central Plata virus
      - Hantavirus Ac1Hu
      - Hantavirus Bu3Hu
      - Hantavirus Ce20
      - Hantavirus Ce22
      - Hantavirus Pe4Hu
      - Hantavirus Sa63
      - Hantavirus U89
    - Hu39694 virus
    - Jabora hantavirus
    - Juquitiba-like virus strain On576
    - Juquitiba-like virus strain PB 1002
    - Juquitiba-like virus strain PB 1033
    - Juquitiba-like virus strain PB 981
    - Lechiguanas virus
    - Maciel virus
    - Neembucu hantavirus
    - Oran virus
    - Pergamino virus
    - Tunari virus
  - Orthohantavirus asamaense
    - Asama virus
  - Orthohantavirus asikkalaense
  - Orthohantavirus bayoui
    - Catacamas virus

www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ vyhledávání sekvencí podle názvu taxonu (viru)



www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ získávání a stahování sekvencí (data mining)

Items: 14

1. Amga virus strain MSB148347 nucleocapsid gene, partial cds
   582 bp linear cRNA
   Accession: KM201419.1 GI: 808178698
   Protein Taxonomy
   GenBank FASTA Graphics

2. Amga virus strain MSB148558 nucleocapsid gene, complete cds
   1,627 bp linear cRNA
   Accession: KM201411.1 GI: 808178696
   Protein Taxonomy
   GenBank FASTA Graphics

3. Amga virus strain MSB148347 glycoprotein precursor, gene, partial cds
   729 bp linear cRNA
   Accession: KM201420.1 GI: 808178694
   Protein Taxonomy
   GenBank FASTA Graphics

4. Amga virus strain MSB148457 glycoprotein precursor, gene, partial cds
   683 bp linear cRNA
   Accession: KM201417.1 GI: 808178692
   Protein Taxonomy
   GenBank FASTA Graphics

5. Amga virus strain MSB148436 glycoprotein precursor, gene, partial cds
   727 bp linear cRNA
   Accession: KM201415.1 GI: 808178690
   Protein Taxonomy
   GenBank FASTA Graphics

www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ získávání a stahování sekvencí

Amga virus strain MSB148347 RNA-dependent RNA polymerase gene, partial cds

GenBank: KM201421.1

FASTA Graphics

Go to:

```
LOCUS       KM201421               1302 bp    cRNA    linear   VRL 31-JAN-2016
DEFINITION  Amga virus strain MSB148347 RNA-dependent RNA polymerase gene,
            partial cds.
ACCESSION   KM201421
VERSION     KM201421.1
KEYWORDS    .
SOURCE      Amga virus
  ORGANISM  Amga virus
            Viruses; Riboviria; Orthornavirae; Negarnaviricota;
            Polyploviricotina; Ellioviricetes; Bunyavirales; Hantaviridae;
            Mammantavirinae; Orthohantavirus; Amga orthohantavirus.
REFERENCE   1  (bases 1 to 1302)
  AUTHORS   Kang,H.J., Gu,S.H., Cook,J.A. and Yanagihara,R.
  TITLE     Amga virus, a newly identified hantavirus in the Laxmann's shrew
            (Sorex caecutiens)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1302)
  AUTHORS   Kang,H.J., Gu,S.H., Cook,J.A. and Yanagihara,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2014) Department of Tropical Medicine and Medical
            Microbiology and Pediatrics, John A. Burns School of Medicine,
            University of Hawaii at Manoa, 651 Ilalo Street, Honolulu, HI
            96813, USA
FEATURES             Location/Qualifiers
     source          1..1302
                     /organism="Amga virus"
                     /mol_type="viral cRNA"
                     /strain="MSB148347"
                     /isolation_source="lung"
                     /host="Sorex caecutiens"
                     /db_xref="taxon:1511732"
                     /geo_loc_name="Russia"
                     /collection_date="14-Aug-2006"
```

Co lze zjistit:

- druh viru
- sekvenovaný lokus
- Accession number
- (hostitel)
- (tkáň hostitele)
- (lokalita sběru)
- (datum sběru)

www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ získávání a stahování sekvencí (data mining)

```
     CDS             <1..>1302
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase"
                     /protein_id="AKD00010.1"
                     /translation="SFLSRVIYKHYKSLISEVTTCFFLFEKGLHGNVNEEAKIHLETV
                     EWALKFRAKEEKYGSHLVEHGYRISDLYLNPSLVEQQMYCQDVVELGAFELNNMLLSK
                     TQVVGNSIINKHWNLPYFSQTRNISLKGMSGSIQEDGHLSASVTLIEAIRYLQNSRHN
                     PSLLQLYEETRTAKAQARIVRKYQRTEADRGFFITTLPTRCRLEIIEDYYDAIAKNVP
                     EEYISYGGERKILNIQQALEKALRWASGESHLELSTGKLIPMKRKLMYVSADATKWSP
                     GDNSAKFRRFTAVLHNGLRDDKLRNCVIDALINIYKTDFFMSRKLKKYIGNMESLDEH
                     VKAFLSFFPDGNSGEVHGNWLQGNLNKCSSLFAVGMSLLFKRVWKELFPELECFFEFA
                     HHSDDGLFIYGYLEPVDDGTDWFMYVTQQIQAGNHHWYSVNT"
ORIGIN
        1 tcattcttat ctagagtgat atataagcat tataaaagct tgatttcaga agtgactaca
       61 tgtttttttc tatttgagaa gggattacat gggaatgtaa atgaggaggc aaaaatacat
      121 ttggaaactg tagaatgggc actaaagttc cgggcaaagg aagagaaata tggttcccac
      181 ttggtagaac atggatatag gatcagtgat ttgtatttga acccatctct ggttgagcaa
      241 caaatgtatt gtcaagatgt agtggaacta ggagcatttg aactaaataa tatgctgcta
      301 tcgaagacac aagttgttgg aaattctatt attaataagc attggaactt accatatttt
      361 agtcaaacgc ggaatattag tttaaaggga atgtctggtt caatccaaga agatggacat
      421 ctatctgctt cagtaactct tatagaagca atccggtatt tacaaaattc acgtcacaac
      481 cctagtctcc tccagttata tgaagagacc agaacagcaa aagcacaggc tagaattgtt
      541 aggaaatatc agagaactga ggcagataga ggattcttca taacgactct ccctaccaga
      601 tgtagactag aaatcattga ggactattat gatgctatag caaagaatgt cccagaagaa
      661 tatatttcat atggtggtga gcgaaagata ttaaatattc agcaagcact tgaaaaagct
      721 ttaagatggg cttctggtga gagtcatcta gagctctcaa cgggtaagct gataccaatg
      781 aaacggaaat tgatgtatgt aagcgcagat gctacaaaat ggtcacctgg tgataattca
      841 gctaaattcc gtcgattcac tgctgtttta cataatggat tgagggatga taaattgcga
      901 aattgtgtga tagatgcact aatcaatatt tataagactg atttttttat gtctagaaaa
      961 ctcaaaaagt atataggaaa tatggagagt ttagatgagc atgtaaaggc cttcttgtcg
     1021 ttttttccag atggaaattc tggtgaagta catggtaatt ggttacaggg gaatctaaac
     1081 aaatgctctt ccttgtttgc tgtaggtatg tctttacttt ttaaaagagt ctggaaagaa
     1141 ttatttccgg aactggaatg ctttttttgaa tttgcacatc attcagatga tgggcttttc
     1201 atatatggtt atcttgaacc agttgatgat ggaacagact ggtttatgta tgtcacacag
     1261 caaattcagg caggtaatca tcattggtac agtgtaaata ca
//
```

Lze zjistit:

- nukleotidovou sekvenci
- aminokyselinovou sekvenci

www.ncbi.nlm.nih.gov

Databáze GenBank – Taxonomy

→ stahování sekvencí (formát FASTA)

FASTA ▾ Send to: ▾

Amga virus strain MSB148347 RNA-dependent RNA polymerase gene, partial cds

GenBank: KM201421.1

GenBank Graphics

```
>KM201421.1 Amga virus strain MSB148347 RNA-dependent RNA polymerase gene, partial cds
TCATTCTTATCTAGAGTGATATATAAGCATTATAAAAGCTTGATTTCAGAAGTGACTACATGTTTTTTTC
TATTTGAGAAGGGATTACATGGGAATGTAAATGAGGAGGCAAAAATACATTTGGAAACTGTAGAATGGGC
ACTAAAGTTCCGGGCAAAGGAAGAGAAATATGGTTCCCACTTGGTAGAACATGGATATAGGATCAGTGAT
TTGTATTTGAACCCATCTCTGGTTGAGCAACAAATGTATTGTCAAGATGTAGTGGAACTAGGAGCATTTG
AACTAAATAATATGCTGCTATCGAAGACACAAGTTGTTGGAAATTCTATTATTAATAAGCATTGGAACTT
ACCATATTTTAGTCAAACGCGGAATATTAGTTTAAAGGGAATGTCTGGTTCAATCCAAGAAGATGGACAT
CTATCTGCTTCAGTAACTCTTATAGAAGCAATCCGGTATTTACAAAATTCACGTCACAACCCTAGTCTCC
TCCAGTTATATGAAGAGACCAGAACAGCAAAAGCACAGGCTAGAATTGTTAGGAAATATCAGAGAACTGA
GGCAGATAGAGGATTCTTCATAACGACTCTCCCTACCAGATGTAGACTAGAAATCATTGAGGACTATTAT
GATGCTATAGCAAAGAATGTCCCAGAAGAATATATTTCATATGGTGGTGAGCGAAAGATATTAAATATTC
AGCAAGCACTTGAAAAAGCTTTAAGATGGGCTTCTGGTGAGAGTCATCTAGAGCTCTCAACGGGTAAGCT
GATACCAATGAAACGGAAATTGATGTATGTAAGCGCAGATGCTACAAAATGGTCACCTGGTGATAATTCA
GCTAAATTCCGTCGATTCACTGCTGTTTTACATAATGGATTGAGGGATGATAAATTGCGAAATTGTGTGA
TAGATGCACTAATCAATATTTATAAGACTGATTTTTTTATGTCTAGAAAACTCAAAAAGTATATAGGAAA
TATGGAGAGTTTAGATGAGCATGTAAAGGCCTTCTTGTCGTTTTTTCCAGATGGAAATTCTGGTGAAGTA
CATGGTAATTGGTTACAGGGGAATCTAAACAAATGCTCTTCCTTGTTTGCTGTAGGTATGTCTTTACTTT
TTAAAAGAGTCTGGAAAGAATTATTTCCGGAACTGGAATGCTTTTTTGAATTTGCACATCATTCAGATGA
TGGGCTTTTCATATATGGTTATCTTGAACCAGTTGATGATGGAACAGACTGGTTTATGTATGTCACACAG
CAAATTCAGGCAGGTAATCATCATTGGTACAGTGTAAATACA
```

www.ncbi.nlm.nih.gov

Taxonomie virů

Jak zjistím, které taxony potřebuji v datasetu?

Jak zjistím druhy virů, které se nacházejí v dané čeledi / řádu?

→ ICTV (International Committee on Taxonomy of Viruses)

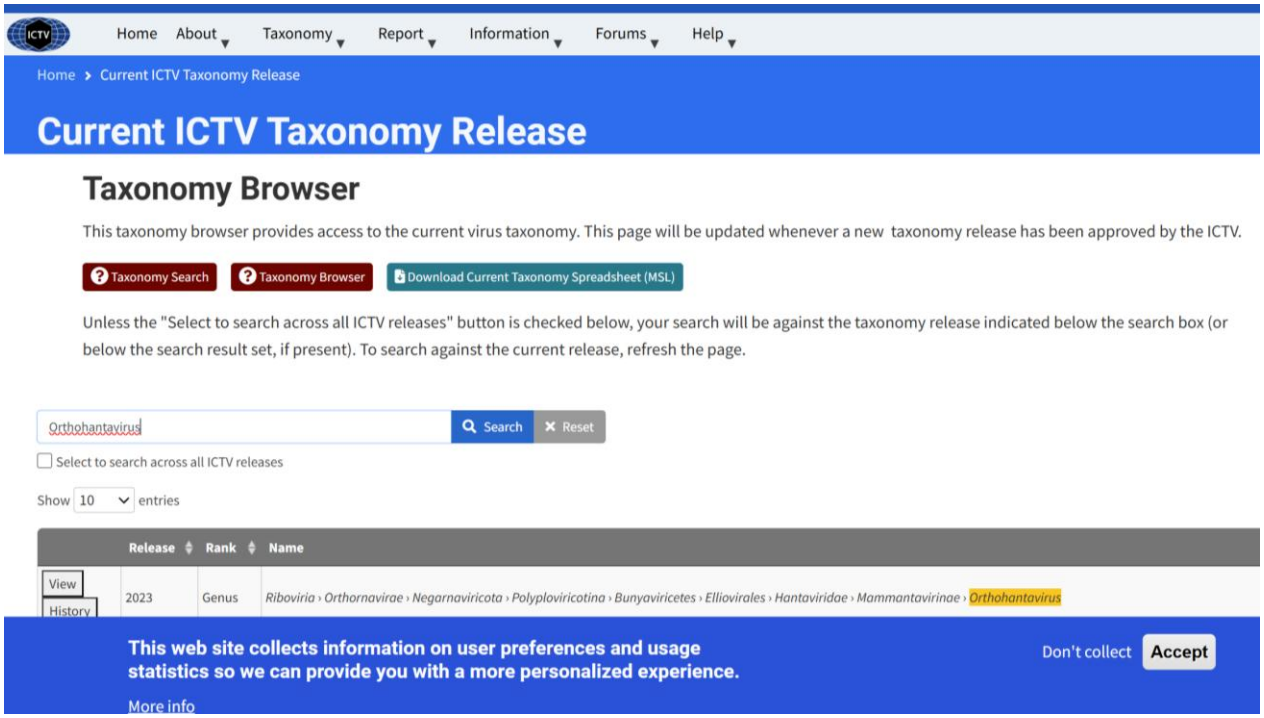
www.ictv.global

Taxonomie virů

Jak zjistím, které taxony potřebuji v datasetu?

Jak zjistím druhy virů, které se nacházejí v dané čeledi / řádu?

→ ICTV (International Committee on Taxonomy of Viruses)



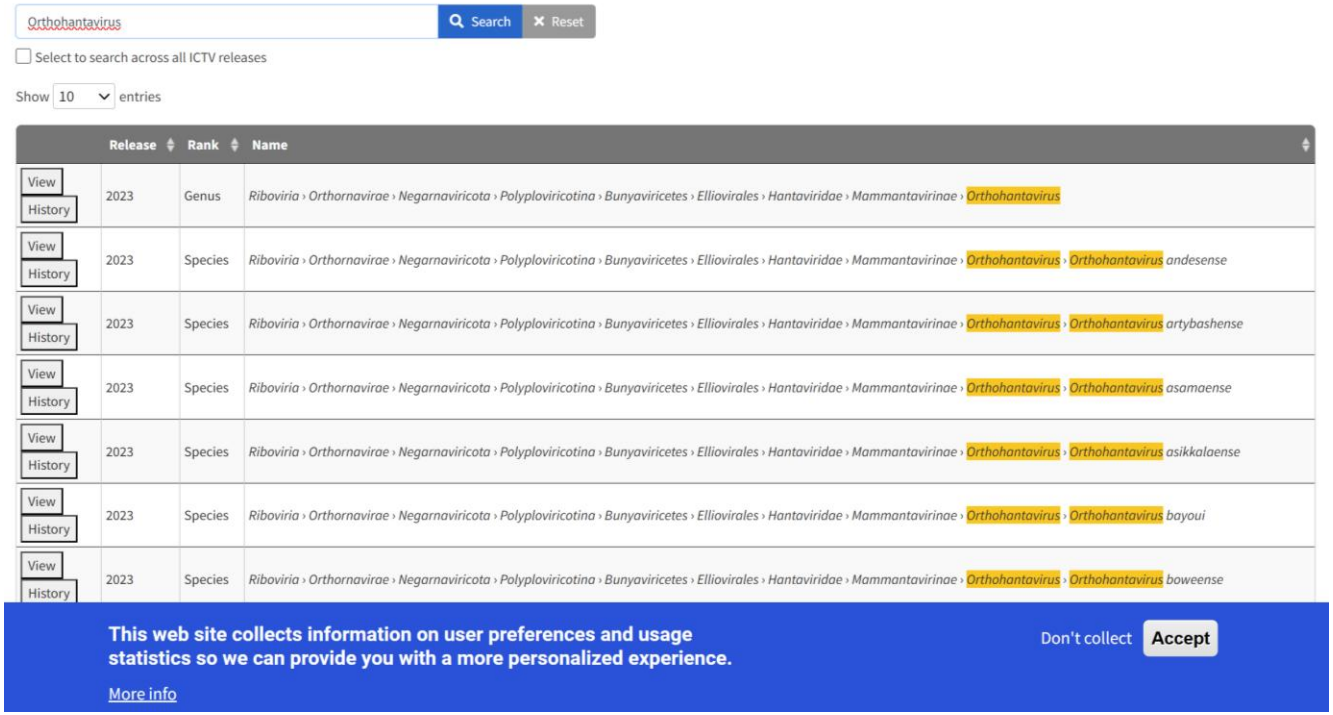
www.ictv.global

Taxonomie virů

Jak zjistím, které taxony potřebuji v datasetu?

Jak zjistím druhy virů, které se nacházejí v dané čeledi / řádu?

→ ICTV (International Committee on Taxonomy of Viruses)



www.ictv.global

Databáze GenBank – Taxonomy

→ stahování sekvencí (formát FASTA)

FASTA ▾ Send to: ▾

Amga virus strain MSB148347 RNA-dependent RNA polymerase gene, partial cds

GenBank: KM201421.1

GenBank Graphics

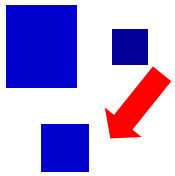
```
>KM201421.1 Amga virus strain MSB148347 RNA-dependent RNA polymerase gene, partial cds
TCATTCTTATCTAGAGTGATATATAAGCATTATAAAAGCTTGATTTCAGAAGTGACTACATGTTTTTTTC
TATTTGAGAAGGGATTACATGGGAATGTAAATGAGGAGGCAAAAATACATTTGGAAACTGTAGAATGGGC
ACTAAAGTTCCGGGCAAAGGAAGAGAAATATGGTTCCCACTTGGTAGAACATGGATATAGGATCAGTGAT
TTGTATTTGAACCCATCTCTGGTTGAGCAACAAATGTATTGTCAAGATGTAGTGGAACTAGGAGCATTTG
AACTAAATAATATGCTGCTATCGAAGACACAAGTTGTTGGAAATTCTATTATTAATAAGCATTGGAACTT
ACCATATTTTAGTCAAACGCGGAATATTAGTTTAAAGGGAATGTCTGGTTCAATCCAAGAAGATGGACAT
CTATCTGCTTCAGTAACTCTTATAGAAGCAATCCGGTATTTACAAAATTCACGTCACAACCCTAGTCTCC
TCCAGTTATATGAAGAGACCAGAACAGCAAAAGCACAGGCTAGAATTGTTAGGAAATATCAGAGAACTGA
GGCAGATAGAGGATTCTTCATAACGACTCTCCCTACCAGATGTAGACTAGAAATCATTGAGGACTATTAT
GATGCTATAGCAAAGAATGTCCCAGAAGAATATATTTCATATGGTGGTGAGCGAAAGATATTAAATATTC
AGCAAGCACTTGAAAAAGCTTTAAGATGGGCTTCTGGTGAGAGTCATCTAGAGCTCTCAACGGGTAAGCT
GATACCAATGAAACGGAAATTGATGTATGTAAGCGCAGATGCTACAAAATGGTCACCTGGTGATAATTCA
GCTAAATTCCGTCGATTCACTGCTGTTTTACATAATGGATTGAGGGATGATAAATTGCGAAATTGTGTGA
TAGATGCACTAATCAATATTTATAAGACTGATTTTTTTATGTCTAGAAAACTCAAAAAGTATATAGGAAA
TATGGAGAGTTTAGATGAGCATGTAAAGGCCTTCTTGTCGTTTTTTCCAGATGGAAATTCTGGTGAAGTA
CATGGTAATTGGTTACAGGGGAATCTAAACAAATGCTCTTCCTTGTTTGCTGTAGGTATGTCTTTACTTT
TTAAAAGAGTCTGGAAAGAATTATTTCCGGAACTGGAATGCTTTTTTGAATTTGCACATCATTCAGATGA
TGGGCTTTTCATATATGGTTATCTTGAACCAGTTGATGATGGAACAGACTGGTTTATGTATGTCACACAG
CAAATTCAGGCAGGTAATCATCATTGGTACAGTGTAAATACA
```

www.ncbi.nlm.nih.gov

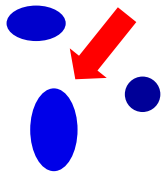
Dataset pro fylogenetické analýzy

reprezentativní výběr taxonů do datasetu, **tentýž gen**

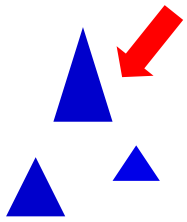
čeleď čtverečků



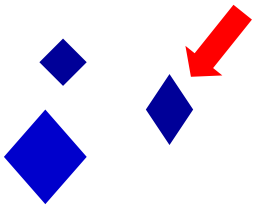
čeleď koleček



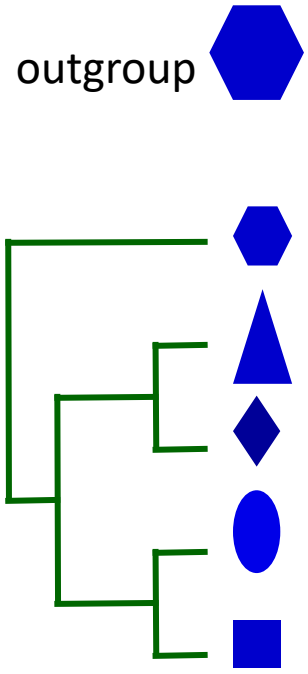
čeleď trojúhelníků



čeleď kosočtverců



outgroup



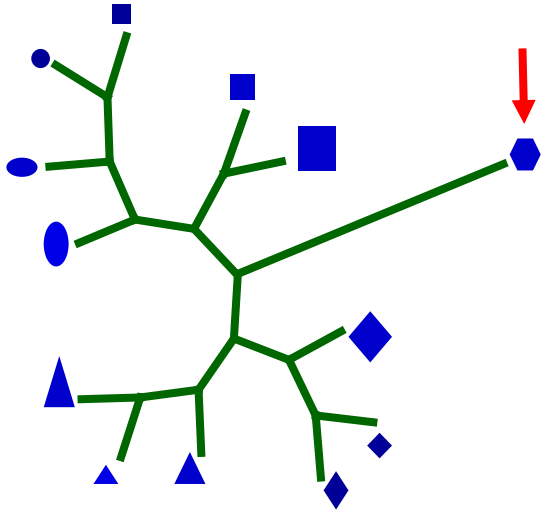
Výběr outgroupu pro fylogenetické analýzy

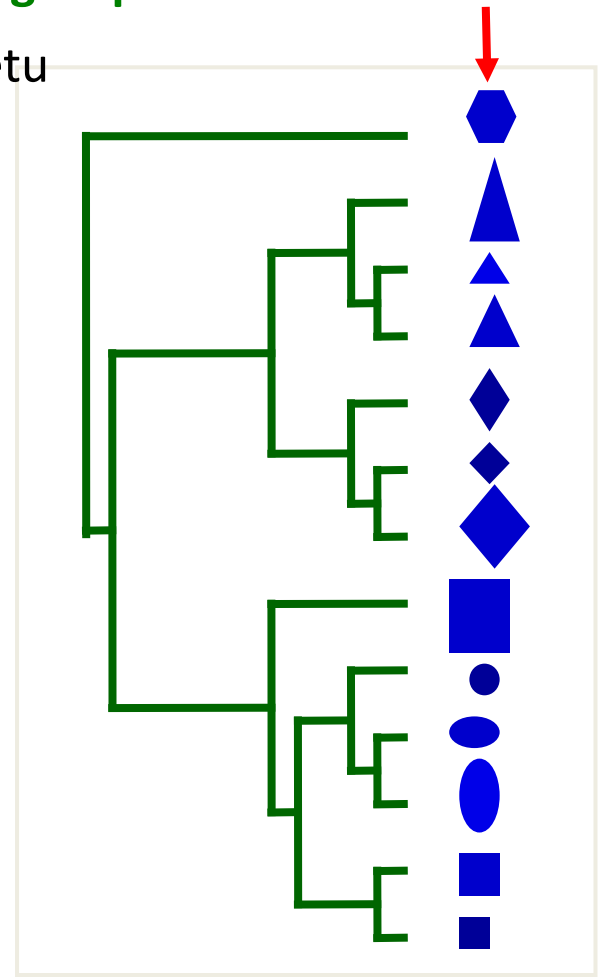
Výběr outgroupu (tj. kořene) - požadavky na outgroup:

- sekvence **stejného genu** jako u taxonů v datasetu
- **vzdálenější, ale příbuzný organismus**

(outgroup musí být out-group)

(neměl by ale být příliš vzdálený)





1. Praktická úloha – tvorba datasetu

Postup:

1. Vyhledat sekvence TÉHOŽ GENU (tj. například L segment či jeho synonymum, RNA dependent RNA polymerase) virů rodu *Orthohantavirus*
dataset musí být VŽDY složen ze sekvencí stejného genu

2. Vyhledané a vybrané sekvence ve formátu FASTA zkopírovat pod sebe (Ctrl+C, Ctrl+V) do Poznámkového bloku

3. Pokud je v GenBanku k danému druhu viru více sekvencí téhož genu, vybíráme:
- nejdelší sekvenci / sekvence
- sekvence pocházející z různých hostitelů
- sekvence pocházející z různých geografických oblastí

4. Vybrat vhodný outgroup a vložit ho do datasetu

www.ncbi.nlm.nih.gov

1. Praktická úloha – tvorba datasetu

Hantavirový genom:

L segment = RNA dependentní RNA polymeráza (RdRp)
GenBank: *RdRp*, RNA-dependent RNA polymerase, segment L, L gene

M segment = glykoproteiny G1 a G2 (Gn a Gc)
GenBank: *Gc protein*, *M polyprotein*, *glycoprotein gene*, *glycoprotein precursor*

S segment = nukleokapsidový protein N
GenBank: *nucleocapsid protein*, *nucleoprotein S*, *nucleoprotein gene*

Tvorba datasetu pro fylogenetické analýzy

Soubor Úpravy Formát Zobrazení Nápověda

```
>FJ495100.1 Tula virus strain TULV/Sestrze/Mag98_02 segment L RNA dependent RNA polymerase gene, partial cds
CAGGAGACAATTCTGCAAAATTCAGAAGGTTTACTCAAACATTGCATGATGGTTTACGAGATGATAAACT
GAAAAGGTGTGTTGTAGATGCTTTACGGAACATATATGAAACTGATTTTTTTATGTCTAGAAAGCTACAT
AGGTATATTGATGGTATGGACGATTTGTCTGAATTTGTAGAGGATTTTTTGTCATTTTTCCCAAACAAGG
TGTCTGCTGCCATAAAAGGAAATTGGCTTCAGGGAAATCTAAACAAATGTTCATCATTATTTGGTGCAGC
AGTTTCCCTGTTATTTAGAAAAGTCTGGAACCTTCTTTATCCAGAACTGGATTGCTTTTTTGAGTTTGCA
CATCATTCT

>MK605685.1 Orthohantavirus tulaense isolate TULV/Magr1100/Beskydy/CZ/2014 segment L RNA-dependent RNA polymerase gene, partial cds
CCAGGGGATAACTCAGCAAAATTCCGAAGATTTACCCAGACATTACATGATGGTTTGCGAGATGATAAAC
TGAAAAGGTGTGTAGTTGATGCATTGCGGAATATATATGAAACTGATTTTTTTATGTCTCGAAAGTTACA
TAGGTATATTGATGGTATGGATGATTTATCTGAATTTGTAGAAGACTTTTTATCATNCTTCCCAAATAAG
GTATCTGCTGCTATTAAAGGGAATTGGCTTCAGGGAAACCTAAACAAGTGCTCATCATTATTTGGCGCAG
CAGTTTCCCTGCTATTTAGAAAAATTTGGAGCCTCCTTTATCCAGAATTAGACTGCTTTTTTGAG

>KC522413.1 Tula virus isolate JiTr/Opava/12 RNA-dependent RNA polymerase gene, partial cds
ACGAAGTGGTCTCCAGGAGATAATTCAGCAAAATTTCGAAGATTTACTCAGGCATTACATGATGGTCTGC
GAGATGATAAATTAAAAAGGTGTGTAGTGGATGCATTGCGGAATATATATGAAACTGACTTTTTTATGTC
TAGAAAGTTACATAGGTATATAGATGGTATGGATGATTTGTCTGAATTTGTAGAAGACTTTTTATCATTC
TTCCCGAATAAGGTATCTGCTGCTATTAAAGGAAATTGGCTTCAAGGAAATCTAAACAAGTGCTCATCAT
TATTTGGTGCAGCAGTTTCCTTGCTATTTAGAAAGATTTGGAGCCTCCTTTATCCAGAATTAGACTGCTT
TTTTGAATTTGCCCACCACTCGAACGACGCA

>FJ495102.1 Tula virus strain TULV/Sred ob Dravi/Ms51_97 segment L RNA dependent RNA polymerase gene, partial cds
CAGGAGATAATTCAGCAAAATTCCGAAGATTCACTCAGGCATTGCATGATGGTTTACGAGATGATAAATT
GAAGAGATGTGTAGTGGATGCTTTGCGGAATATATATGAAACTGATTTTTTTATGTCAAGAAAGCTACAT
AGATATATTGATGGTATGGATGATTTGTCTGAATTTGTAGAAGACTTTTTGTCATTCTTCCCAAATAAGG
TATCTGCTGCTATTAAAGGAAATTGGCTCCAAGGAAACTTAAATAAGTGTTCATCGTTGTTTGGTGCAGC
AGTTTCCTTGTTATTTAGAAAGATTTGGAGTCTCCTTTATCCGGAATTGGATTGCTTTTTTGAGTTTGCA
CACCATTCT

>ON243802.1 Tula orthohantavirus isolate 56AA RNA-dependent RNA polymerase gene, partial cds
ACTCAGACATTACACGATGGTTTACGGGATGATAAATTAAAAAGGTGCGTAGTAGATGCATTACGGAATA
TATATGAGACTGATTTTTTTATGTCTAGAAAATTACATAGATACATTGATGGTATGGATGATTTATCTGA
GTTTGTAGAAGATTTTTTTATCATTCTTTCCAAATAAGGTATCCGCTGCTATTAAAGGAAACTGGCTTCAA
GGAAACCTAAATAAATGCTCATCATTATTTGGTGCAGCAGTTTCCCTGCTATTTAGGAAAATTTGGAGTC
TGCTTT

>HQ728459.1 Tula virus isolate GER/152/Arv segment L RNA dependent RNA polymerase gene, partial cds
GATGCTACTAAATGGTCGCCAGGTGATAACTCCGCAAAATTTAGGAGGTTTACTCAGGCACTACATGATG
GTTTAAGAGATGATAAATTAAAGAGATGTGTGGTTGATGCACTGAGAAACATATATGAGACTGACTTTTT
TATGTCTAGAAAATTACATAGATACATTGATGGTATGGATGACTTGTCCGAATTTGTGGAGGACTTTCTT
TCATTTTTCCCAAATAAAGTATCAGCTGCTATTAAAGGGAATTGGCTACAAGGCAATTTGAATAAGTGCT
CTTCATTATTTGGTGCAGCTGTTTCCCTGTTATTCAGAAAAATATGGAATCTTCTTTACCCAGAATTAGA
TTGTTTCTTTGAATTTGCACACCACTCTGATGATGCATTATTTATTTATGGTTATTTAGAACCTACAGAT
GACGGA
```

www.ncbi.nlm.nih.gov

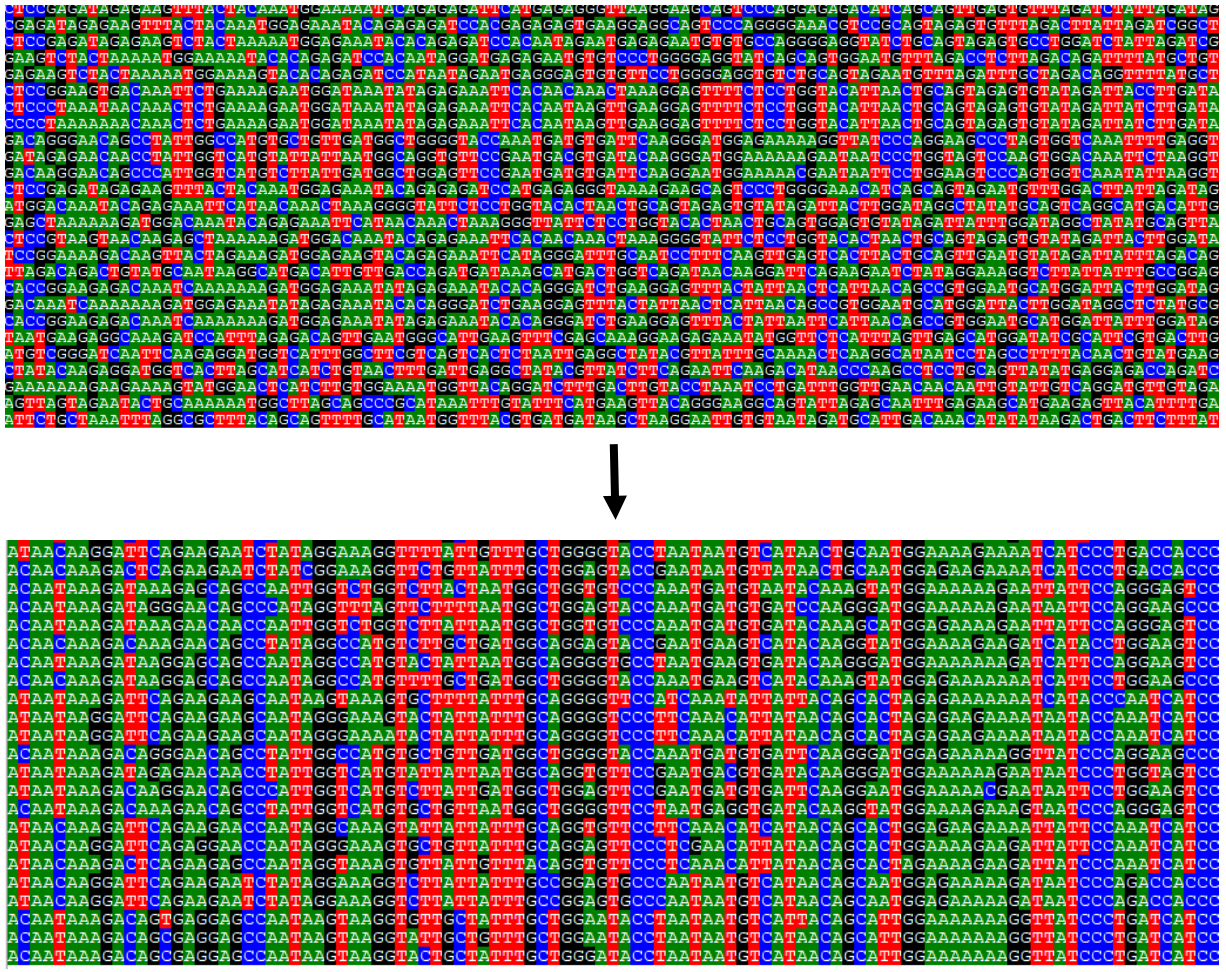
Úprava datasetu pro fylogenetické analýzy

```
>FJ495100
CAGGAGACAATTCTGCAAAATTCAGAAGGTTTACTCAAACATTGCATGATGGTTTACGAGATGATAAACT
GAAAAGGTGTGTTGTAGATGCTTTACGGAACATATATGAAACTGATTTTTTTATGTCTAGAAAGCTACAT
AGGTATATTGATGGTATGGACGATTTGTCTGAATTTGTAGAGGATTTTTTGTCATTTTTCCCAAACAAGG
TGTCTGCTGCCATAAAAGGAAATTGGCTTCAGGGAAATCTAAACAAATGTTCATCATTATTTGGTGCAGC
AGTTTCCCTGTTATTTAGAAAAGTCTGGAACCTTCTTTATCCAGAACTGGATTGCTTTTTTGAGTTTGCA
CATCATTCT

>MK605685
CCAGGGGATAACTCAGCAAAATTCCGAAGATTTACCCAGACATTACATGATGGTTTGCGAGATGATAAAC
TGAAAAGGTGTGTAGTTGATGCATTGCGGAATATATATGAAACTGATTTTTTTATGTCTCGAAAGTTACA
TAGGTATATTGATGGTATGGATGATTTATCTGAATTTGTAGAAGACTTTTTATCATNCTTCCCAAATAAG
GTATCTGCTGCTATTAAAGGGAATTGGCTTCAGGGAAACCTAAACAAGTGCTCATCATTATTTGGCGCAG
CAGTTTCCCTGCTATTTAGAAAAATTTGGAGCCTCCTTTATCCAGAATTAGACTGCTTTTTTGAG

>KC522413
ACGAAGTGGTCTCCAGGAGATAATTCAGCAAAATTTCGAAGATTTACTCAGGCATTACATGATGGTCTGC
GAGATGATAAATTAAAAAGGTGTGTAGTGGATGCATTGCGGAATATATATGAAACTGACTTTTTTATGTC
TAGAAAGTTACATAGGTATATAGATGGTATGGATGATTTGTCTGAATTTGTAGAAGACTTTTTATCATTC
TTCCCGAATAAGGTATCTGCTGCTATTAAAGGAAATTGGCTTCAAGGAAATCTAAACAAGTGCTCATCAT
TATTTGGTGCAGCAGTTTCCTTGCTATTTAGAAAGATTTGGAGCCTCCTTTATCCAGAATTAGACTGCTT
TTTTGAATTTGCCCACCACTCGAACGACGCA

>FJ495102
CAGGAGATAATTCAGCAAAATTCCGAAGATTCACTCAGGCATTGCATGATGGTTTACGAGATGATAAATT
GAAGAGATGTGTAGTGGATGCTTTGCGGAATATATATGAAACTGATTTTTTTATGTCAAGAAAGCTACAT
AGATATATTGATGGTATGGATGATTTGTCTGAATTTGTAGAAGACTTTTTGTCATTCTTCCCAAATAAGG
TATCTGCTGCTATTAAAGGAAATTGGCTCCAAGGAAACTTAAATAAGTGTTCATCGTTGTTTGGTGCAGC
AGTTTCCTTGTTATTTAGAAAGATTTGGAGTCTCCTTTATCCGGAATTGGATTGCTTTTTTGAGTTTGCA
CACCATTCT

>ON243802
ACTCAGACATTACACGATGGTTTACGGGATGATAAATTAAAAAGGTGCGTAGTAGATGCATTACGGAATA
TATATGAGACTGATTTTTTTATGTCTAGAAAATTACATAGATACATTGATGGTATGGATGATTTATCTGA
GTTTGTAGAAGATTTTTTATCATTCTTTCCAAATAAGGTATCCGCTGCTATTAAAGGAAACTGGCTTCAA
GGAAACCTAAATAAATGCTCATCATTATTTGGTGCAGCAGTTTCCCTGCTATTTAGGAAAATTTGGAGTC
TGCTTT

>HQ728459
GATGCTACTAAATGGTCGCCAGGTGATAACTCCGCAAAATTTAGGAGGTTTACTCAGGCACTACATGATG
GTTTAAGAGATGATAAATTAAAGAGATGTGTGGTTGATGCACTGAGAAACATATATGAGACTGACTTTTT
TATGTCTAGAAAATTACATAGATACATTGATGGTATGGATGACTTGTCCGAATTTGTGGAGGACTTTCTT
TCATTTTTCCCAAATAAAGTATCAGCTGCTATTAAAGGGAATTGGCTACAAGGCAATTTGAATAAGTGCT
CTTCATTATTTGGTGCAGCTGTTTCCCTGTTATTCAGAAAAATATGGAATCTTCTTTACCCAGAATTAGA
TTGTTTCTTTGAATTTGCACACCACTCTGATGATGCATTATTTATTTATGGTTATTTAGAACCTACAGAT
GACGGA
```

www.ncbi.nlm.nih.gov

Alignment

→ seřazení sekvencí obsažených v datasetu



Alignment

→ různé programy, různé algoritmy

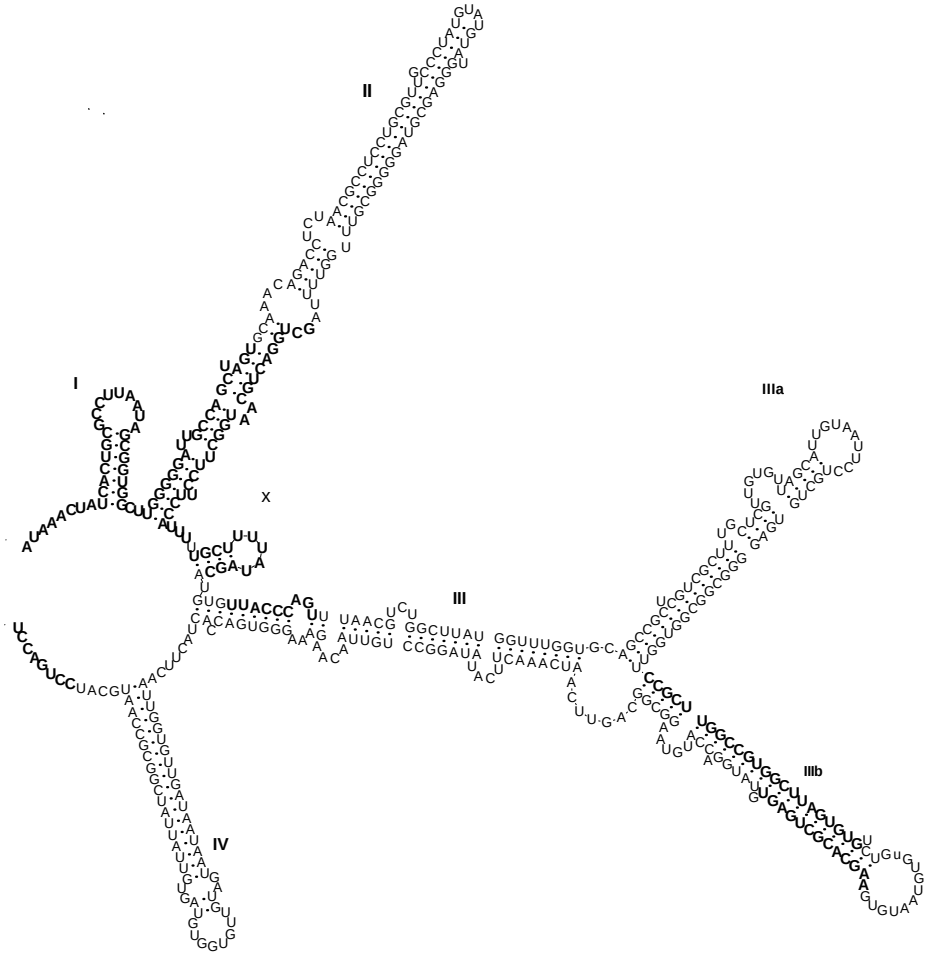
BioEdit
Clustal W
T-Coffee
MUSCLE
MegAlign – součást balíčku DNASTAR
MAFFT – online
SeaView – možnost konkatenace

konkatenovaný alignment = spojený alignment 2 a více různých genů

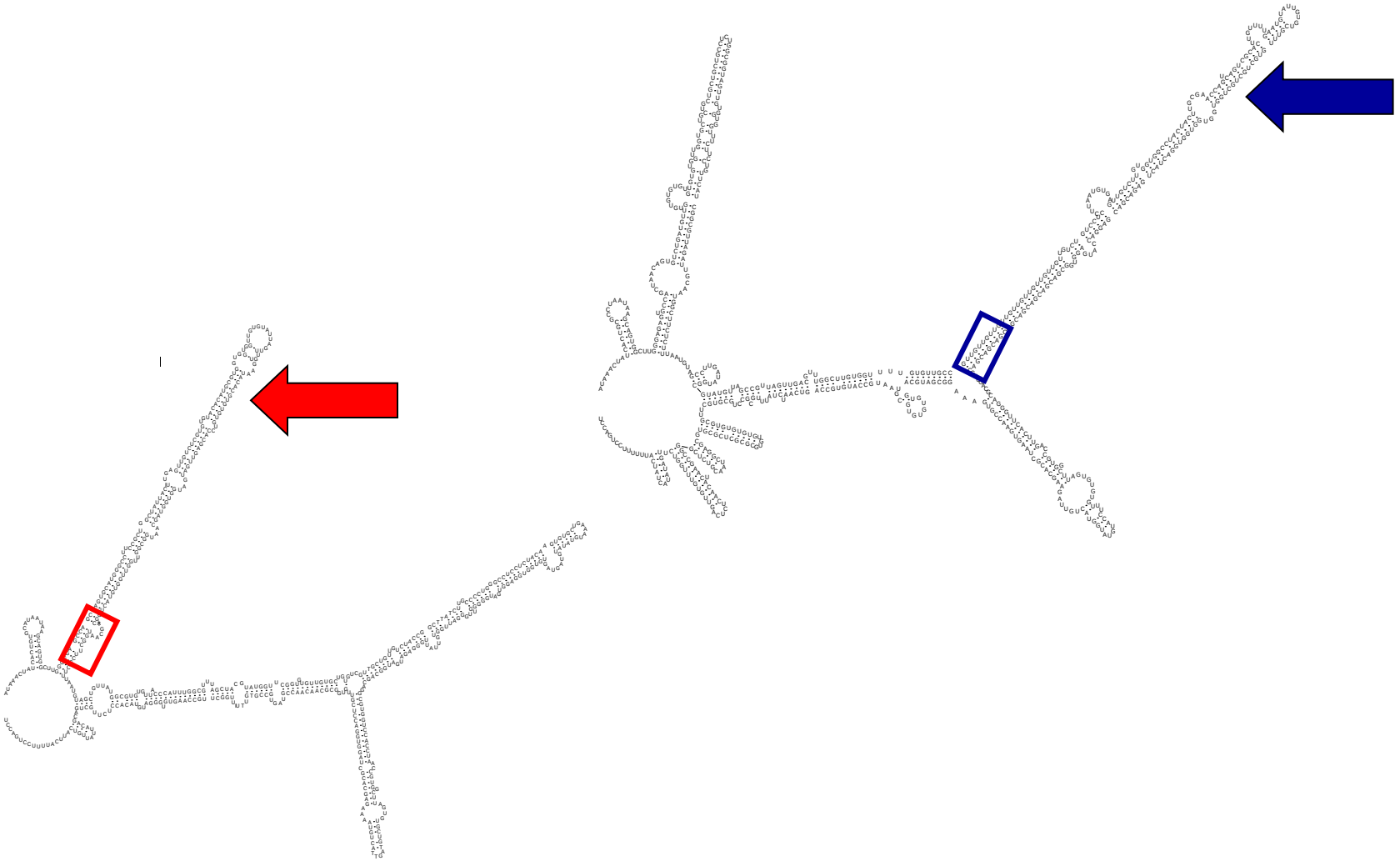
→ další možnosti alignování:

- využití „předalignovaných“ souborů z databází
- alignování zohledňující sekundární strukturu

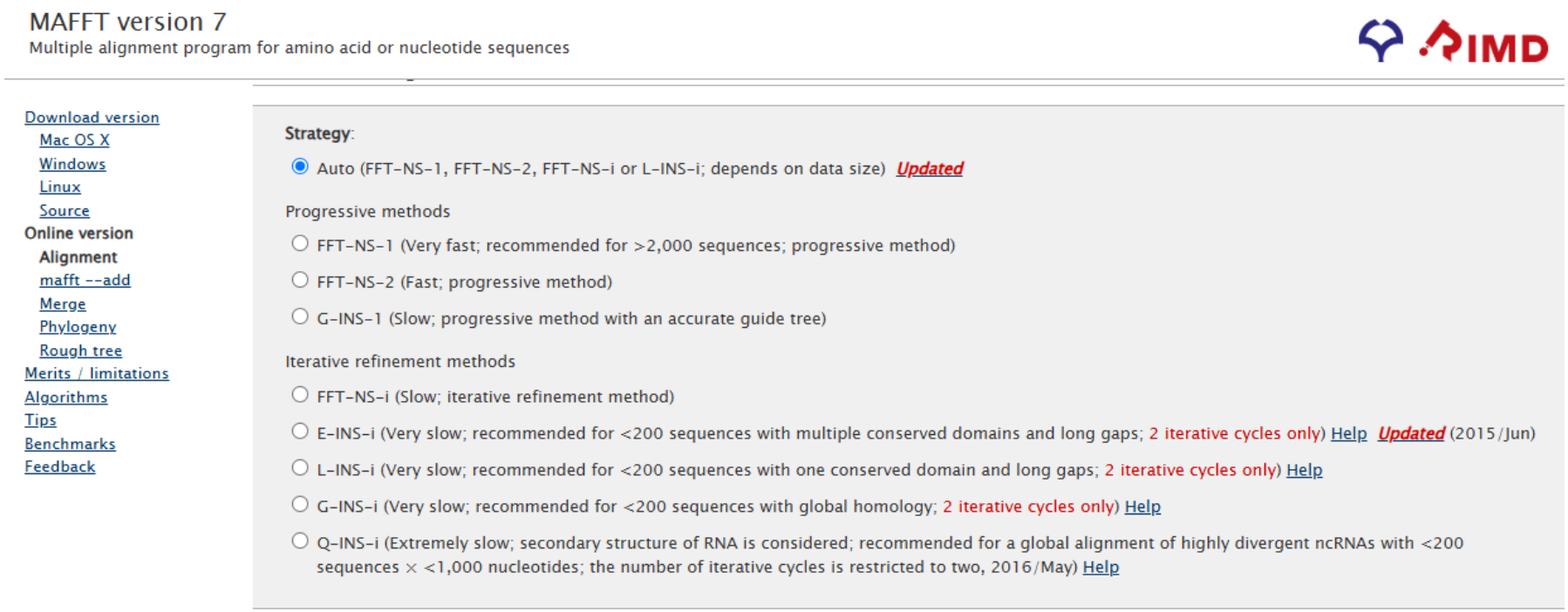
Alignování za pomoci sekundární struktury



Alignování za pomoci sekundární struktury



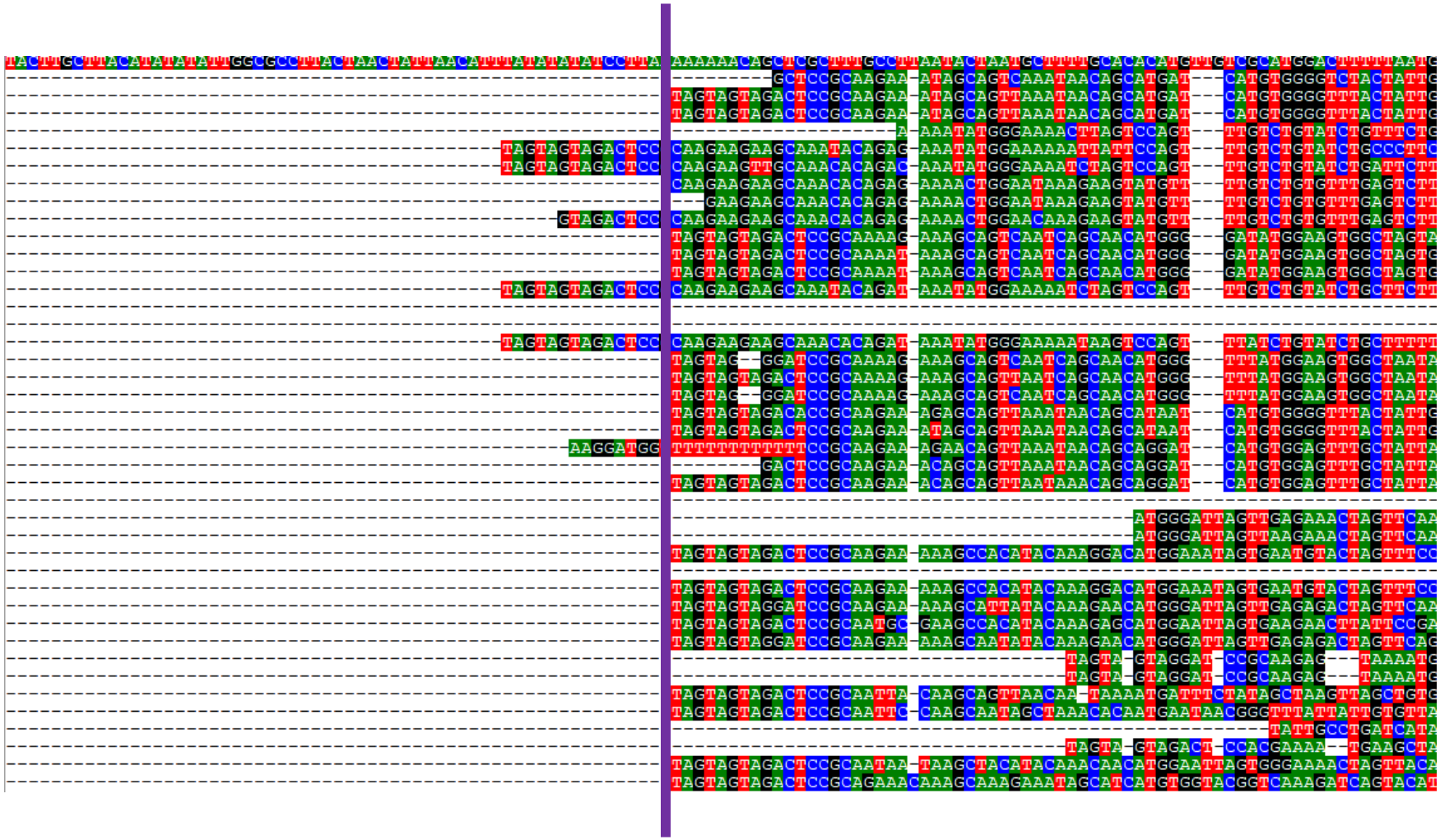
MAFFT



https://mafft.cbrc.jp/alignment/server/index.html

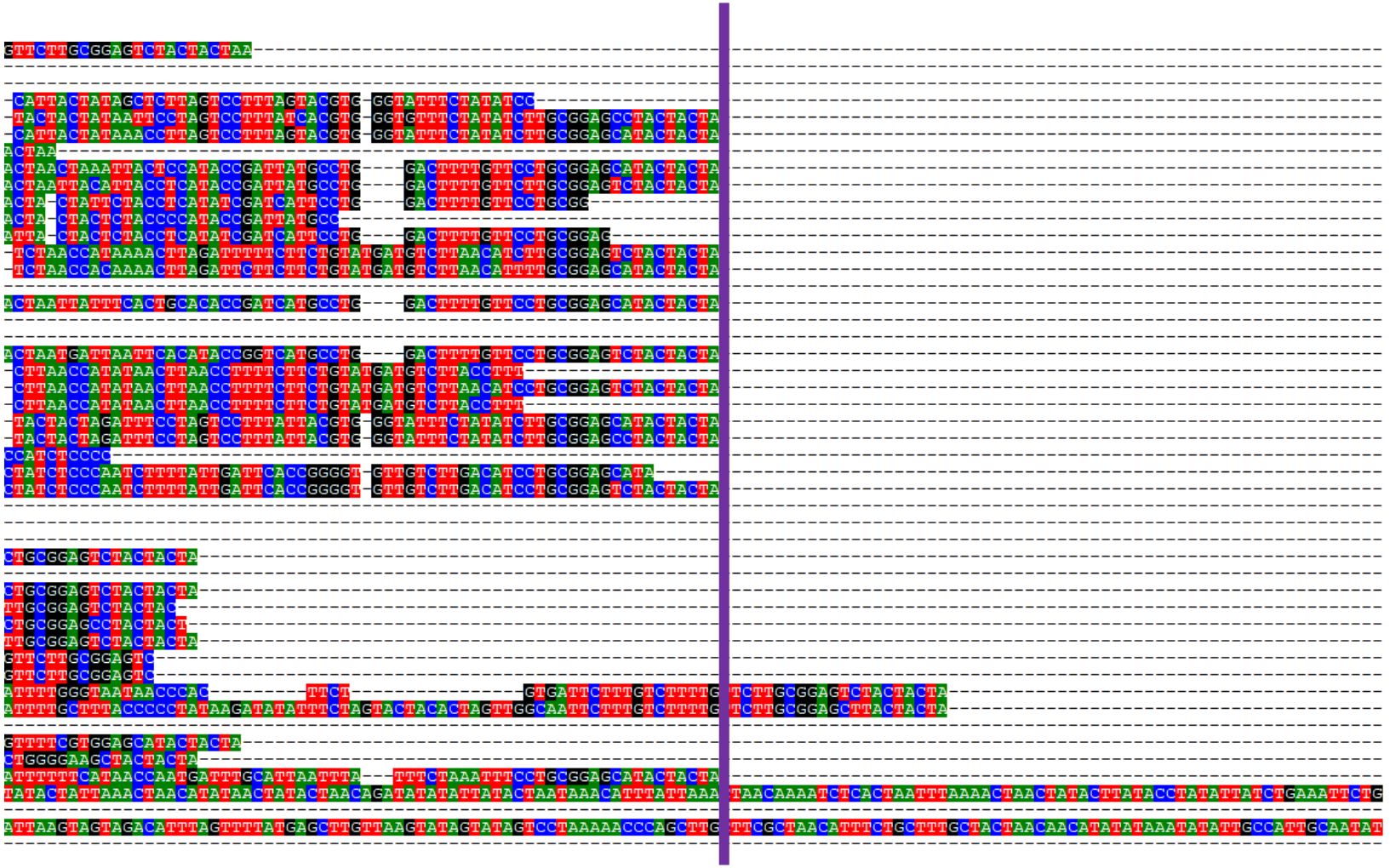
Alignment

→ ořezání (trimování) alignmentu na jednotnou délku (začátek i konec)



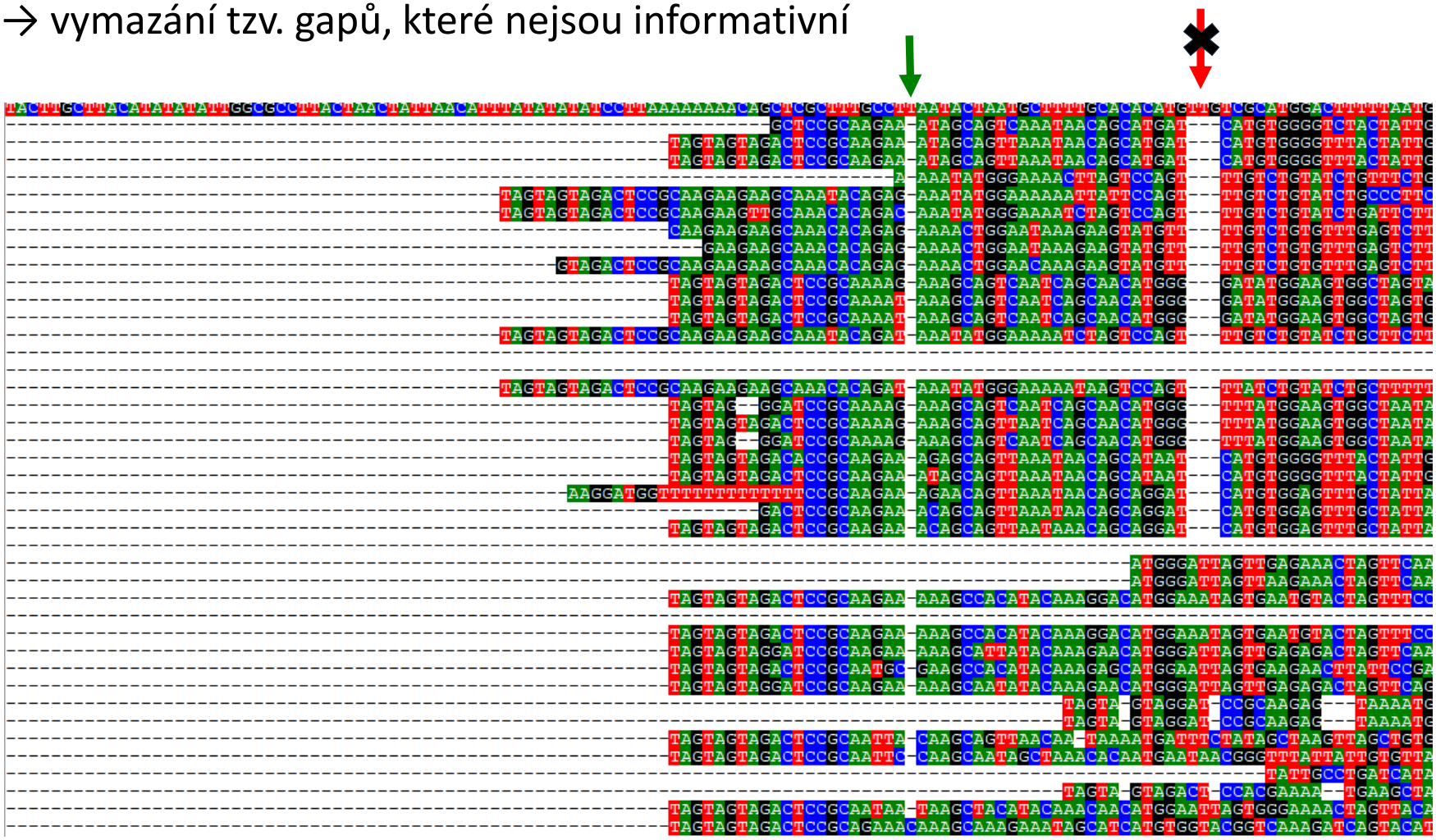
Alignment

→ ořezání (trimování) alignmentu na jednotnou délku (začátek i konec)



Alignment

→ vymazání tzv. gapů, které nejsou informativní

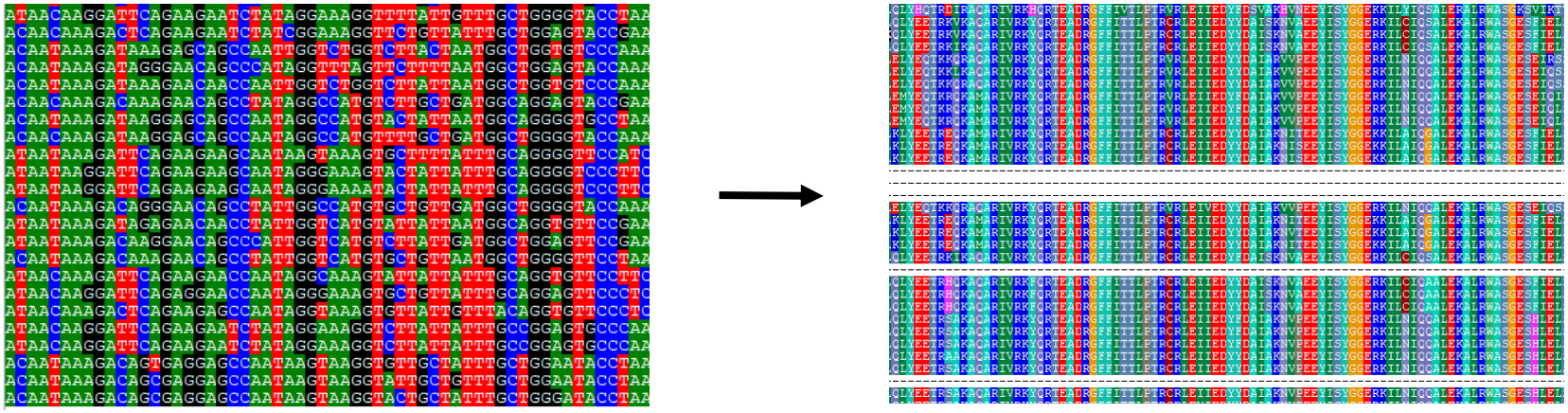


Alignment

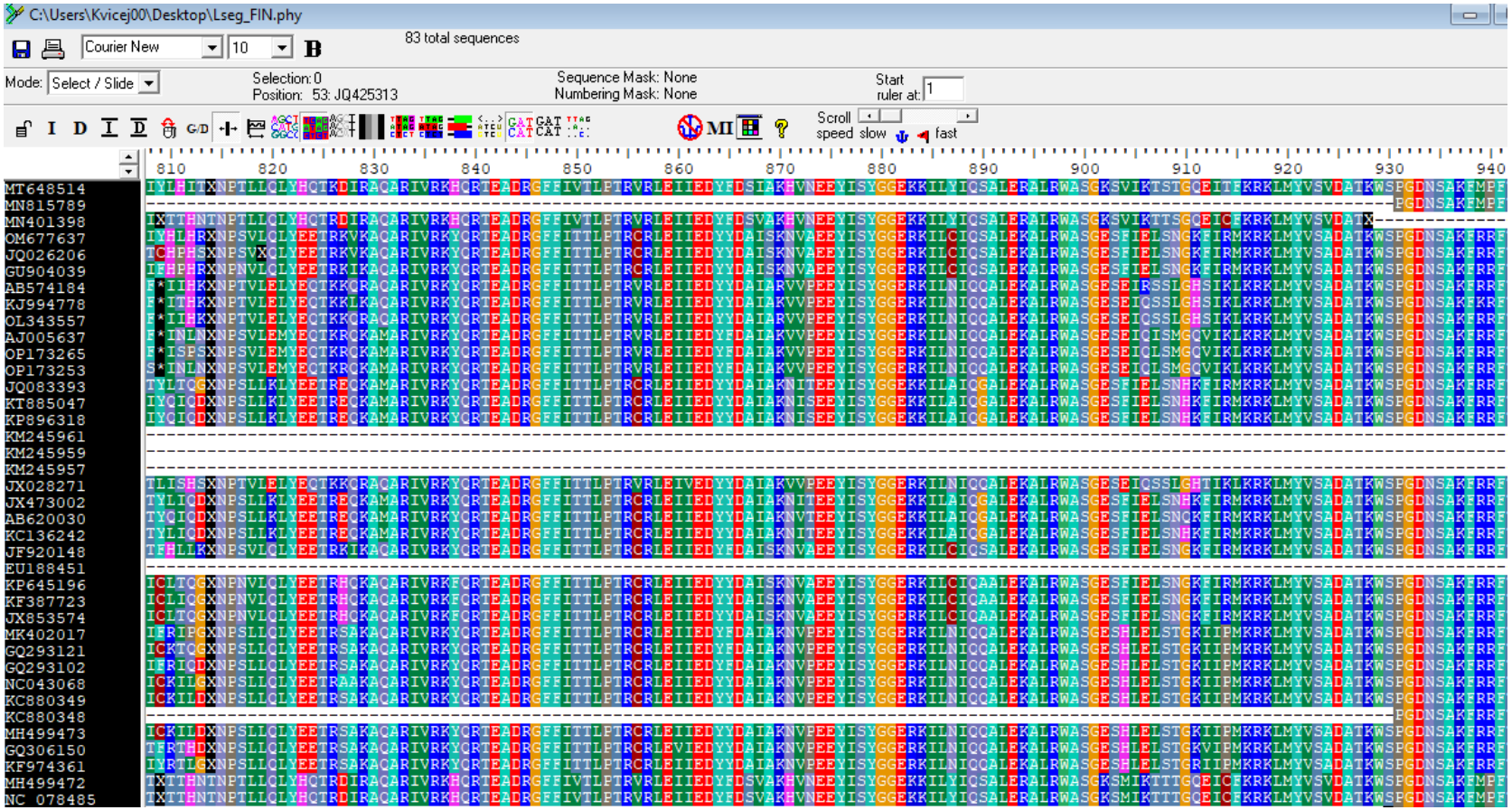
Ořezávání alignmentu:

Geny nekódující protein – ořezáváme v nukleotidovém formátu

Geny kódující protein – ořezáváme v aminokyselinovém formátu
(aby nedošlo k porušení čtecího rámce)



Alignment


C:\Users\Kvicej00\Desktop\Lseg_FIN.phy
Courier New
10
83 total sequences
Mode: Select / Slide
Selection: 0
Position: 53: JQ425313
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
810 820 830 840 850 860 870 880 890 900 910 920 930 940
MT648514
MN815789
MN401398
OM677637
JQ026206
GU904039
AB574184
KJ994778
OL343557
AJ005637
OP173265
OP173253
JQ083393
KT885047
KP896318
KM245961
KM245959
KM245957
JX028271
JX473002
AB620030
KC136242
JF920148
EU188451
KP645196
KF387723
JX853574
MK402017
GQ293121
GQ293102
NC043068
KC880349
KC880348
MH499473
GQ306150
KF974361
MH499472
NC_078485

Alignment

Ukládání alignmentu:

- v různých formátech, v závislosti na účelu dalšího použití

FASTA (.fas) – možnost otevření jak v programu alignmentu,
tak v textové podobě (Poznámkový blok)
– možnost mazání či přidávání nových sekvencí
do již zalignovaného souboru (→ usnadnění práce)

PHYLIP (.phy) – pro maximum likelihood analýzy (PhyML, RAxML apod.)

NEXUS (.nex) – pro Bayesovskou analýzu (MrBayes)
a analýzy maximum parsimony (PAUP)

2. Praktická úloha – alignment

Úloha – tvorba a úprava alignmentu v programu BioEdit:

- načíst dataset, který mám uložený jako .txt, do programu BioEdit
- Accessory Application – ClustalW Multiple alignment
- Run ClustalW – OK

    → načte sekvence a porovnává každou s každým, řadí dle podobnosti

- převod do aminokyselin pomocí Ctrl+T, zpět do nukleotidů také Ctrl+T
- ořezat začátek a konec alignmentu na vhodnou délku
- zkontrolovat alignment, vymazat gapy
- uložit alignment ve formátech FASTA a PHYLIP (File – Save As)

Fylogenetické analýzy

Fylogenetická analýza = klasifikace organismů založená na evoluční historii

Účel:

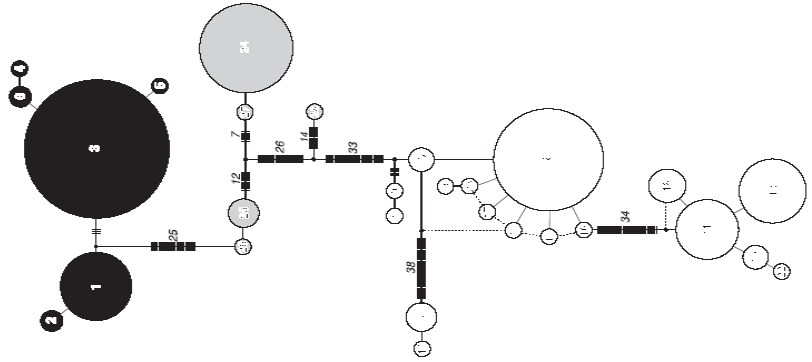
- zjišťování příbuznosti a vztahů mezi organismy / taxony
- evoluční interpretace fylogenetických vztahů
- taxonomie
- koevoluce

→ pro pozorovaná data hledáme adekvátní vysvětlení

Výstup:

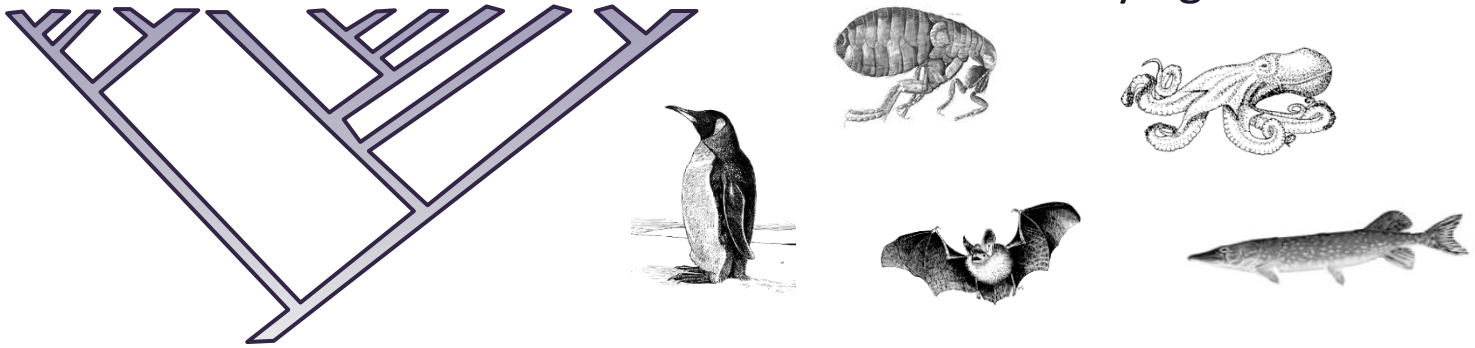
grafické zobrazení pomocí **fylogenetického stromu** (**fylogramu**)

Fylogenetické analýzy



vnitrodruhové vztahy
(populační genetika)

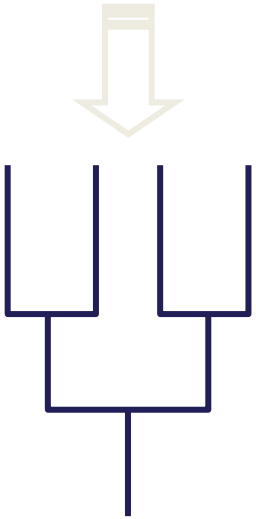
mezidruhové vztahy



fylogenetické vztahy

Fylogenetické analýzy

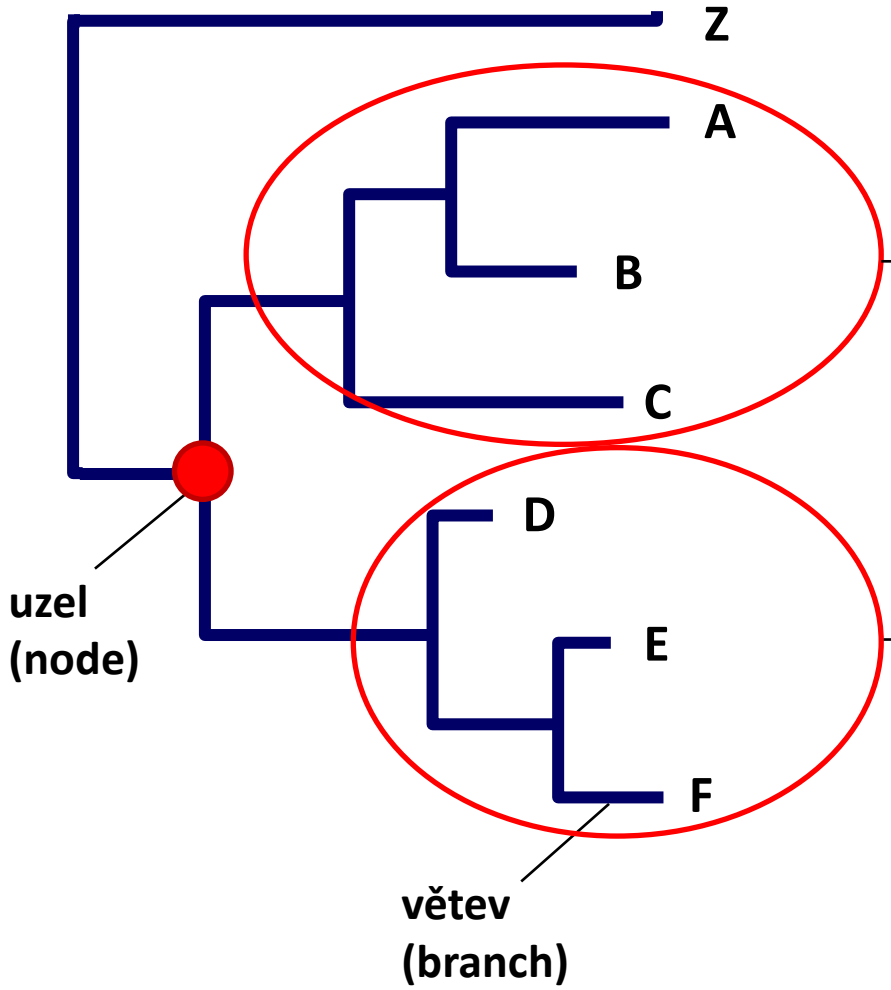
GAATCATCCC
GACCAAACCTA
GAATCATCCC
GACCAAACCTA



?

Fylogenetický strom

A-F taxony
Z outgroup (kořen)



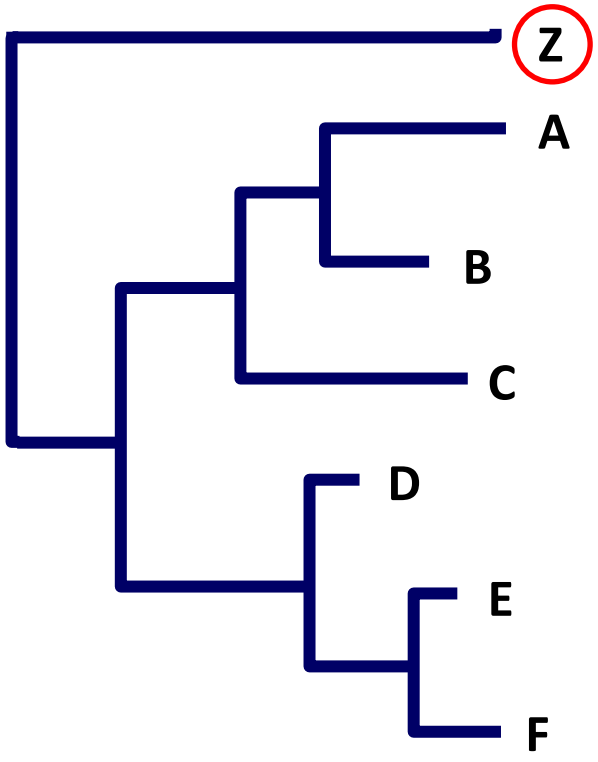
skupina (klastr) B,
sesterská skupina ke skupině A

skupina (klastr) A
(group, cluster, clade)

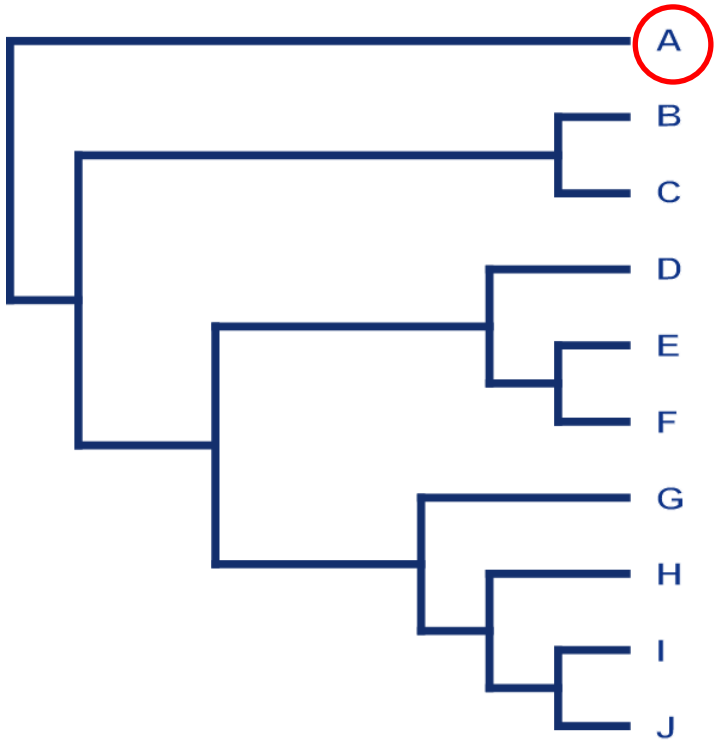
dvojcestné větve (dichotomie)
uspořádání větví = topologie

Zobrazení fylogenetických stromů

**ZAKOŘENĚNÝ,
S DÉLKOU VĚTVÍ**

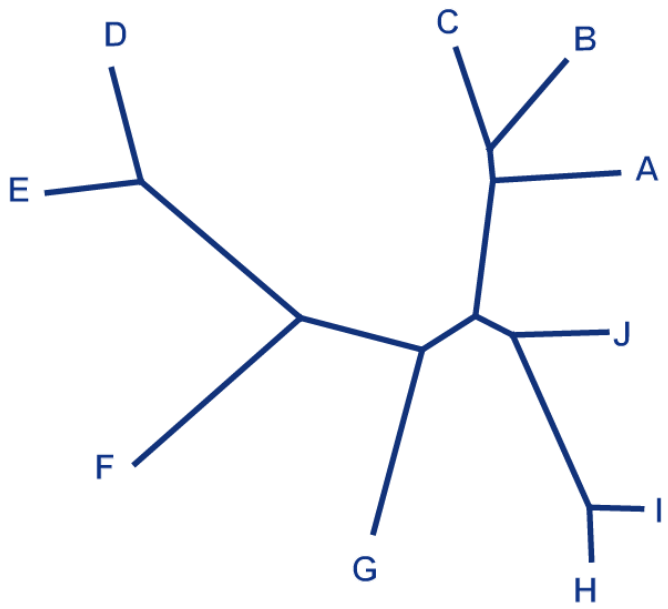


**ZAKOŘENĚNÝ,
BEZ DÉLKY VĚTVÍ**



Zobrazení fylogenetických stromů

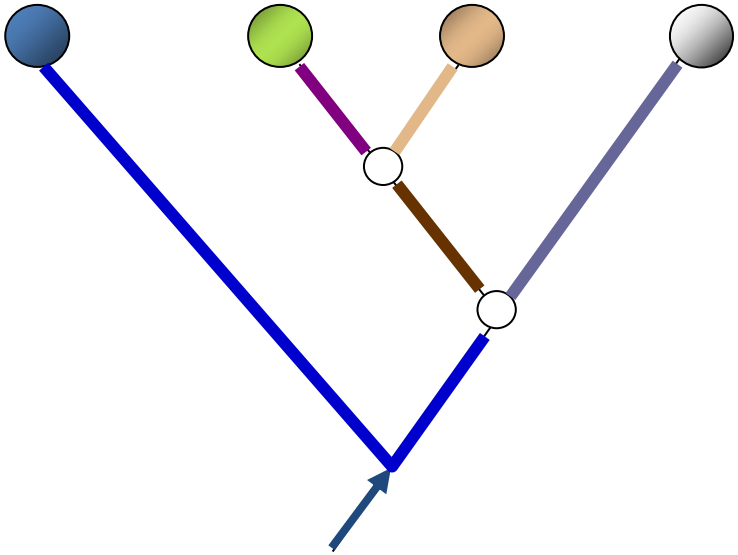
NEZAKOŘENĚNÝ

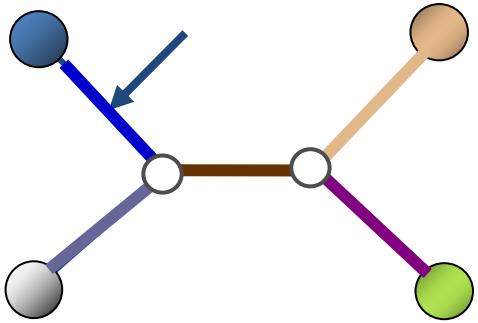


NEWICK FORMÁT

(A ((B C) ((D (E F))(G (H (I J))))))

Fylogenetický strom



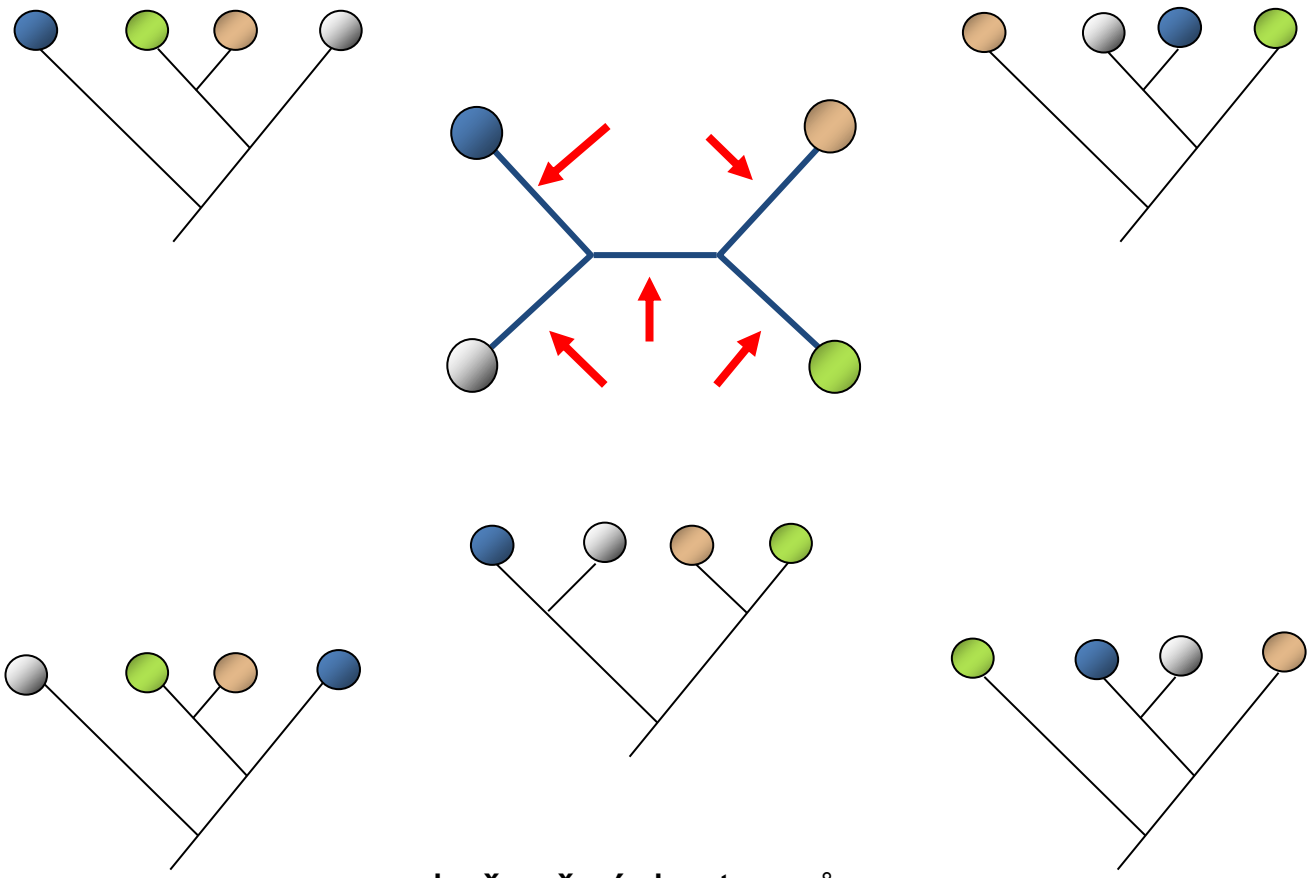


n počet taxonů ve stromu

n-2 uzlů

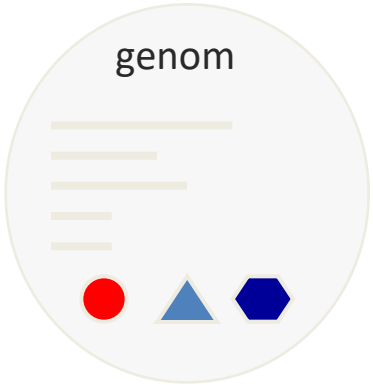
2n-3 větví

Fylogenetický strom



zakořeněných stromů
je více než nezakořeněných

Zdroje fylogenetické informace

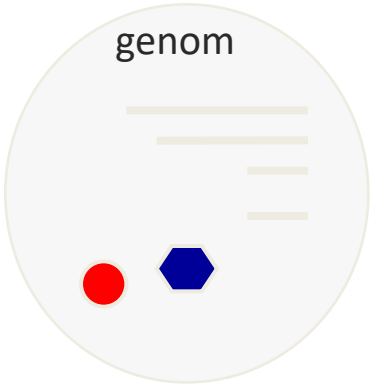


amplikonová data z NGS

„multilokusové“ informace (RFLP, RAPD, AFLP, mikrosatelity)

pořadí genů

sekvence nukleotidů



gen A — gen B — gen C

gen A — gen C — gen B

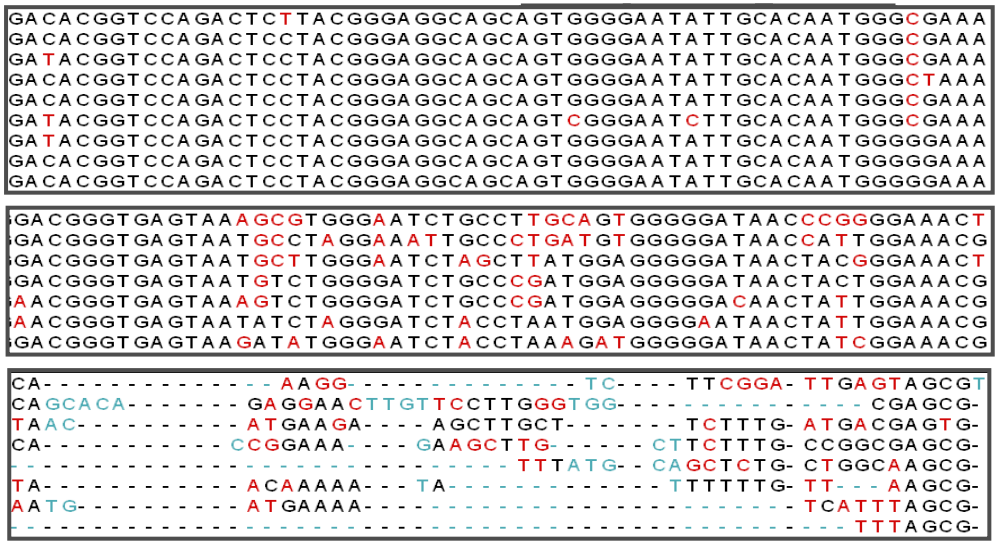
GAATCATCCGGACCAAACTTA

GAATCATCCCGACCAAACCTA

Fylogenetická data

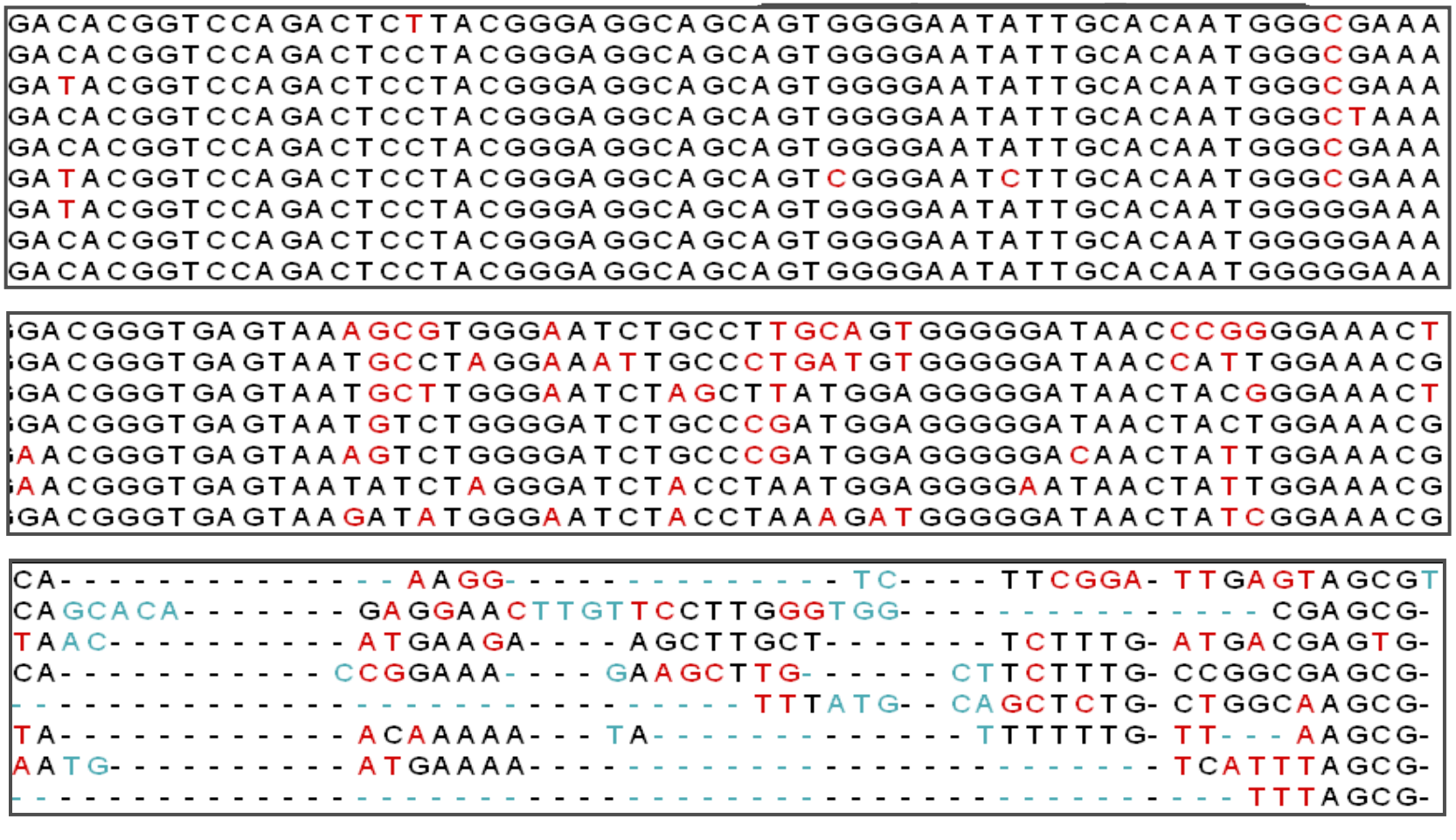
Požadavky na fylogenetická data:

- informativnost (tj. přiměřená variabilita mezi taxony)
- nezávislost
- homologie
- dostatečné množství



Fylogenetická data

Informativnost (přiměřená variabilita mezi taxony):



Fylogenetická analýza

Několik přístupů (algoritmů):

→ pro pozorovaná data hledáme adekvátní vysvětlení

• parsimonie → co nejúspornější uspořádání dat

MP (Maximum parsimony) – počet substitucí

• pravděpodobnostní přístup → co nejpravděpodobnější uspořádání dat na základě daných předpokladů

ML (Maximum likelihood) – počet substitucí „na pozici“

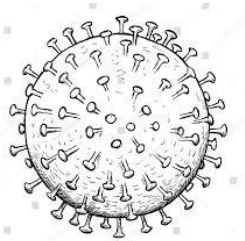
BI (Bayesian inference)

je nutné „znát“ (stanovit) model, jak probíhá evoluce

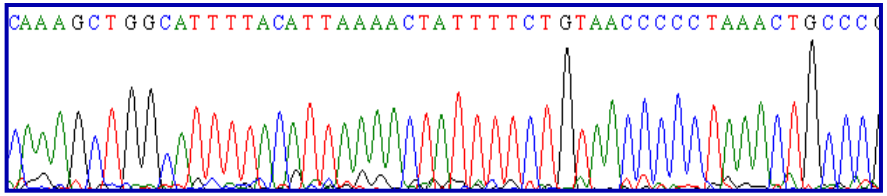
→ výběr vhodného modelu molekulární evoluce

Evoluční modely:

→ korekce na opakované substituce

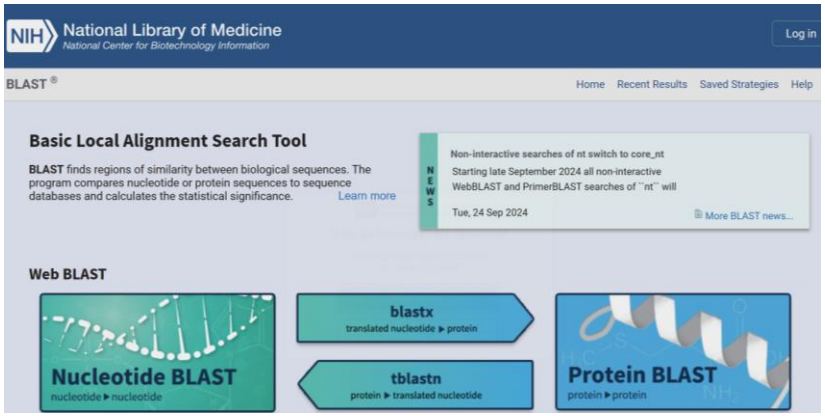


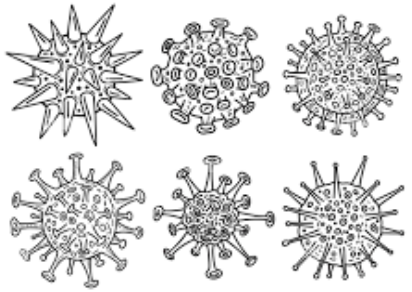
Izolace DNA / RNA, PCR, sekvenování



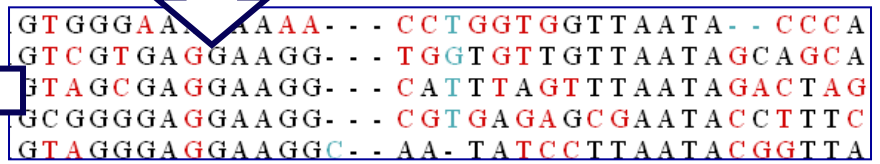
Organismus, který chci fylogeneticky charakterizovat (měl bych vědět proč)

Vyhledání homologických sekvencí pro další taxony (měl bych vědět, pro které)

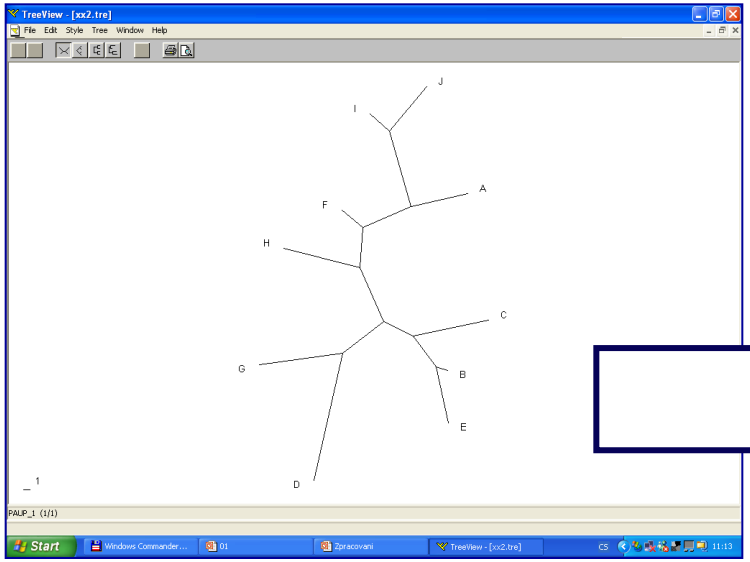




Vytvoření alignmentu



Fylogenetická analýza zvolenou metodou



To hlavní a podstatné:
interpretace fylogenetických vztahů

Maximální parsimonie (MP)

Maximální parsimonie (Maximum parsimony, MP)

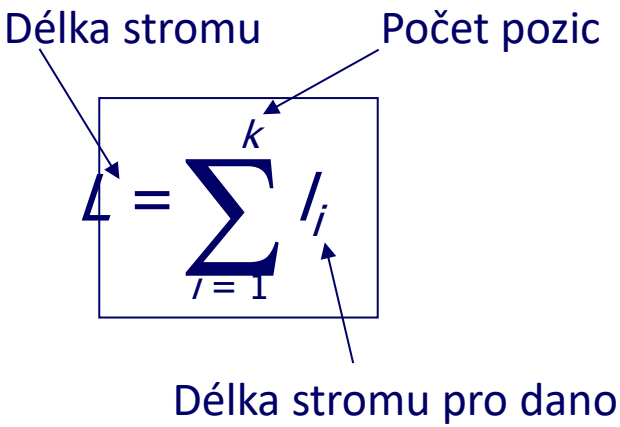
- pracuje s délkou stromu a hledá nejkratší kladogram
(tj. co nejúspornější uspořádání dat)

Znaky (pozice alignmentu):

- neinformativní (konstantní)
- variabilní neinformativní
- **variabilní informativní**

→ délka stromu je množstvím
všech změn na všech pozicích alignmentu

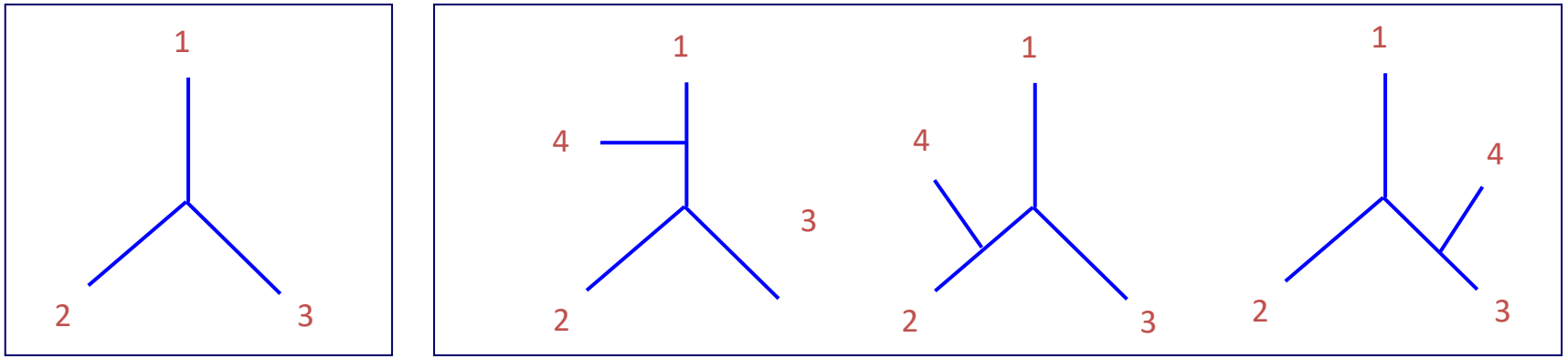
→ sestavení všech stromů



Maximální parsimonie (MP)

Například:

- Výchozí strom z libovolných 3 taxonů
- Sestavení tří možných stromů s přidáním 4. taxonu
- Z každého stromu vzniklého v předešlém kroku se postaví pět dalších stromů přidáním 5. taxonu
- Pokračujeme až do přidání posledního taxonu



3 taxony → 1 nezakořeněný strom
4 taxony → 3 nezakořeněné stromy
→ a pak to prudce roste...
počet možných kombinací roste jako faktoriál *n*

Maximální parsimonie (MP)

No. taxa	Number of unrooted trees	Number of rooted trees
	1	1
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10 395
8	10 395	135 135
9	135 135	2 027 025
10	2 027 025	34 459 425

Maximální parsimonie (MP)

Maximální parsimonie (Maximum parsimony, MP)

→ sestavení co nejkratšího stromu (výchozího)

→ jeho následné zlepšování pomocí přeskupování větví (branch swapping) heuristické metody (různé algoritmy)

Výhody MP:

- jednoduchá, pochopitelná, rychlá
- minimální množství předpokladů o evoluci
- dobře prostudována matematicky

Problémy MP:

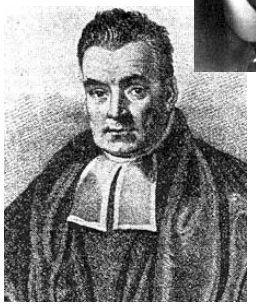
- předpoklad parsimonie je zcela jistě nesprávný pro sekvence s rychlou evolucí

Pravděpodobnostní metody

- pojem pravděpodobnost nelze definovat, je značně subjektivní

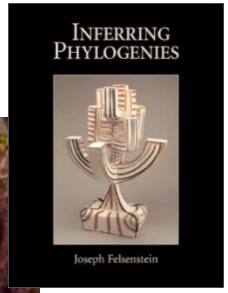
R.A. Fischer (evoluční biolog) → likelihood pro biologická data

Thomas Bayes (matematik) → pravděpodobnost



pravděpodobnost (probability) ≠ věrohodnost (likelihood)

Joseph Felsenstein – přišel na to, že Fischerův přístup by šel napasovat na fylogenetiku (kniha „Inferring Phylogenies“)



máme data (tj. pozorovaný jev) → chceme ho adekvátně vysvětlit (hypotéza)
→ pravděpodobnost mezi daty a fylogenezí

Pravděpodobnostní metody

pravděpodobnost (probability) ***vs.*** **věrohodnost (likelihood)**

vysvětlení (*hypotéza*) pozorovaný jev (*data*)

- někdo skládal mince aby to bylo hezké, a takhle to dopadlo.
- roztrhl se pytlik s mincemi a takhle to padlo náhodu
- spodní strana mince je těžší a pod stolem jsou magnety

probability(H|D)

že tato hypotéza způsobila tato data

probability(D|H)

že se objeví takováto data, pokud nastane předpokládaná hypotéza

likelihood

Pravděpodobnostní metody

Fylogenetika → hledáme hypotézu s maximální věrohodností

v likelihoodu jsou délky větví velmi důležité

Hypotéza = topologie vč. délky větví a modelu

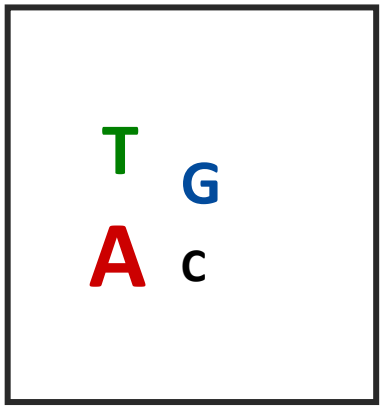
→ vezmu data, prohledávám topologie, délky větví v každé topologii
na každé úrovni najdu nejvěrohodnější a z nich vyberu ten nejvěrohodnější
(na nejvyšší úrovni)

Model:
- evoluce na jednotlivých pozicích je nezávislá
- evoluce v jednotlivých liniích je nezávislá

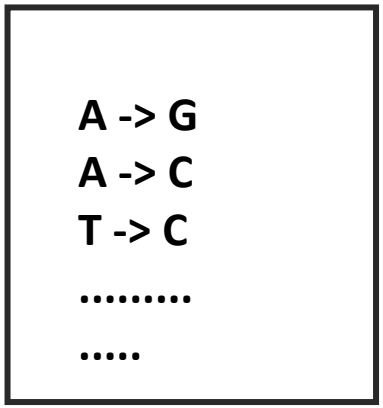
G	A	C	T	C	A	T	C	C	...*m*
G	C	A	T	C	A	T	T	C	
G	C	A	T	C	A	T	G	C	
G	C	A	T	T	A	T	T	C	
G	G	A	T	C	A	T	T	C	

Evoluční modely

frekvence nukleotidů
(base composition)



transition
probabilities



rychlost substitucí
na různých pozicích matice
(distribution of rates)

Výběr vhodného evolučního modelu:

různé volně stažitelné nebo online běžící programy, např.

SMS: Smart Model Selection
www.atgc-montpellier.fr/sms/

Evoluční modely

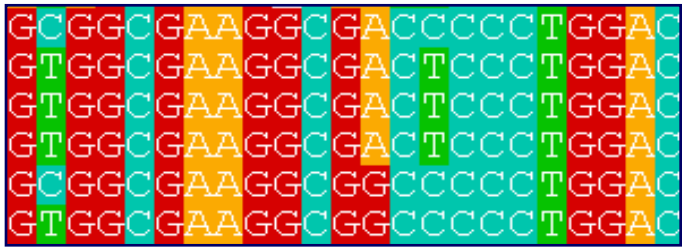
SMS: Smart Model Selection

www.atgc-montpellier.fr/sms/

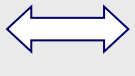
Maximální věrohodnost (ML)

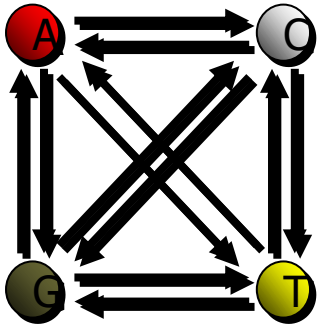
Maximální věrohodnost (Maximum likelihood, ML)

→ hledání hypotézy s maximální věrohodností

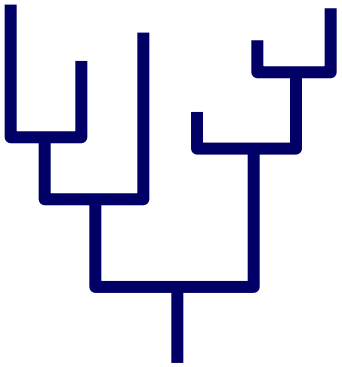


matice





model



topologie (včetně délky větví)

data

hypotéza

Maximální věrohodnost (ML)

Maximální věrohodnost (Maximum likelihood, ML)

→ hledání hypotézy s maximální věrohodností

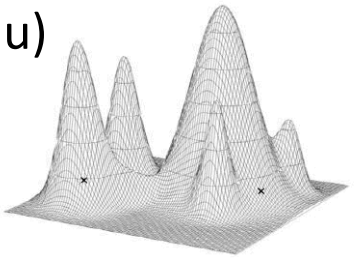
- evoluce na jednotlivých pozicích je nezávislá
- evoluce v jednotlivých liniích je nezávislá

Výhody ML:

→ jednoznačně lepší, pravděpodobnější stromy než MP

Nevýhody ML:

- hledání ML stromu je „optimality criterion“ proces – stejný problém jako u MP
- nastavení parametrů analýzy, výběr vhodného modelu
- výpočetně náročná (hledání nejlepší kombinace modelu a stromu)
- časově náročná



Maximální věrohodnost (ML)

→ různé programy

- volně stažitelné, online běžící, součást balíčků (např. Geneious)

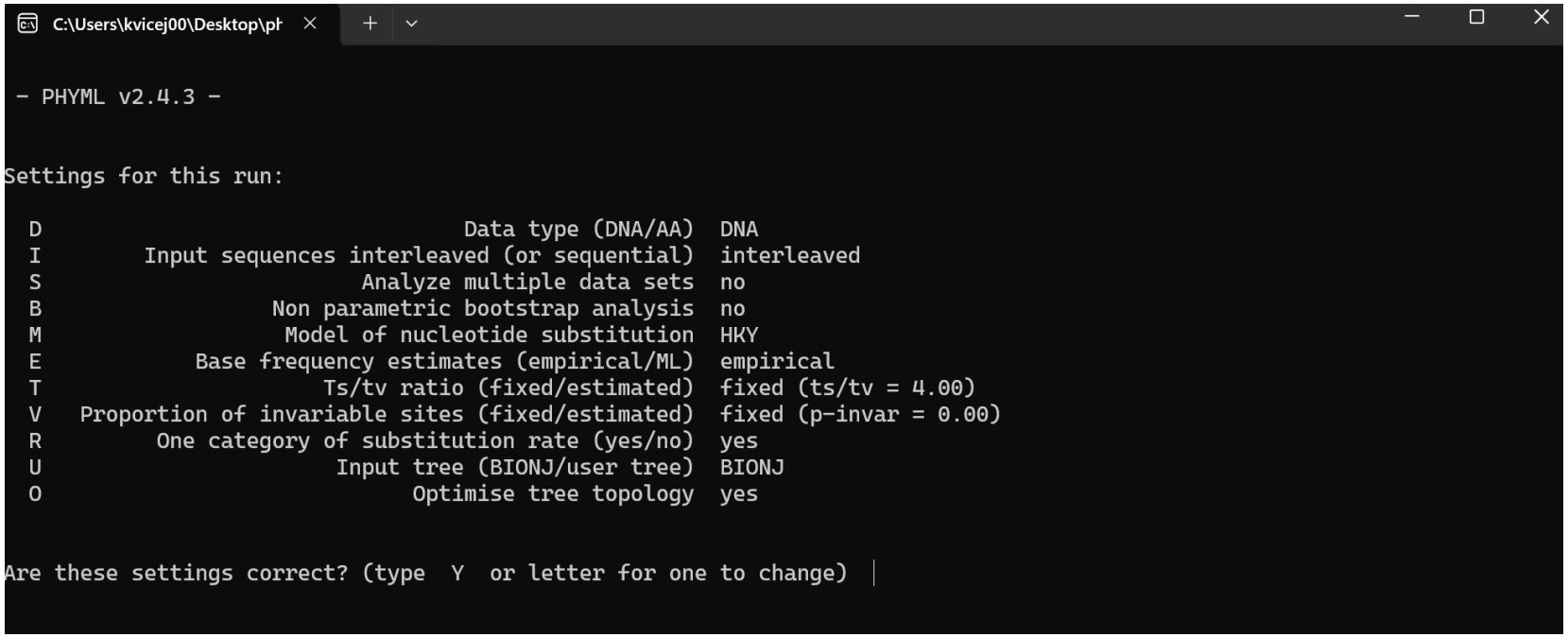
PhyML – zdarma, běží poměrně rychle a efektivně

RAxML

→ načítá soubory ve formátu PHYLIP (.phy)

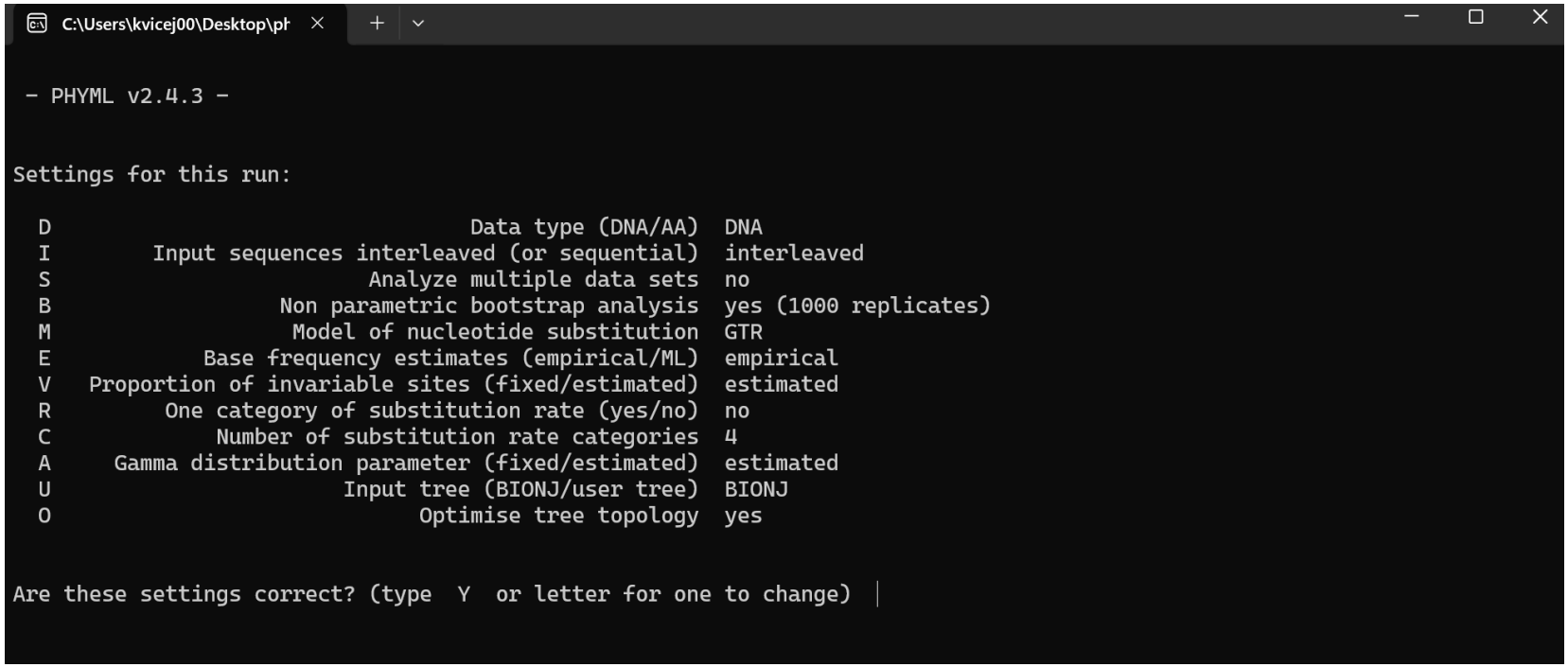
Maximální věrohodnost (ML)

PhyML



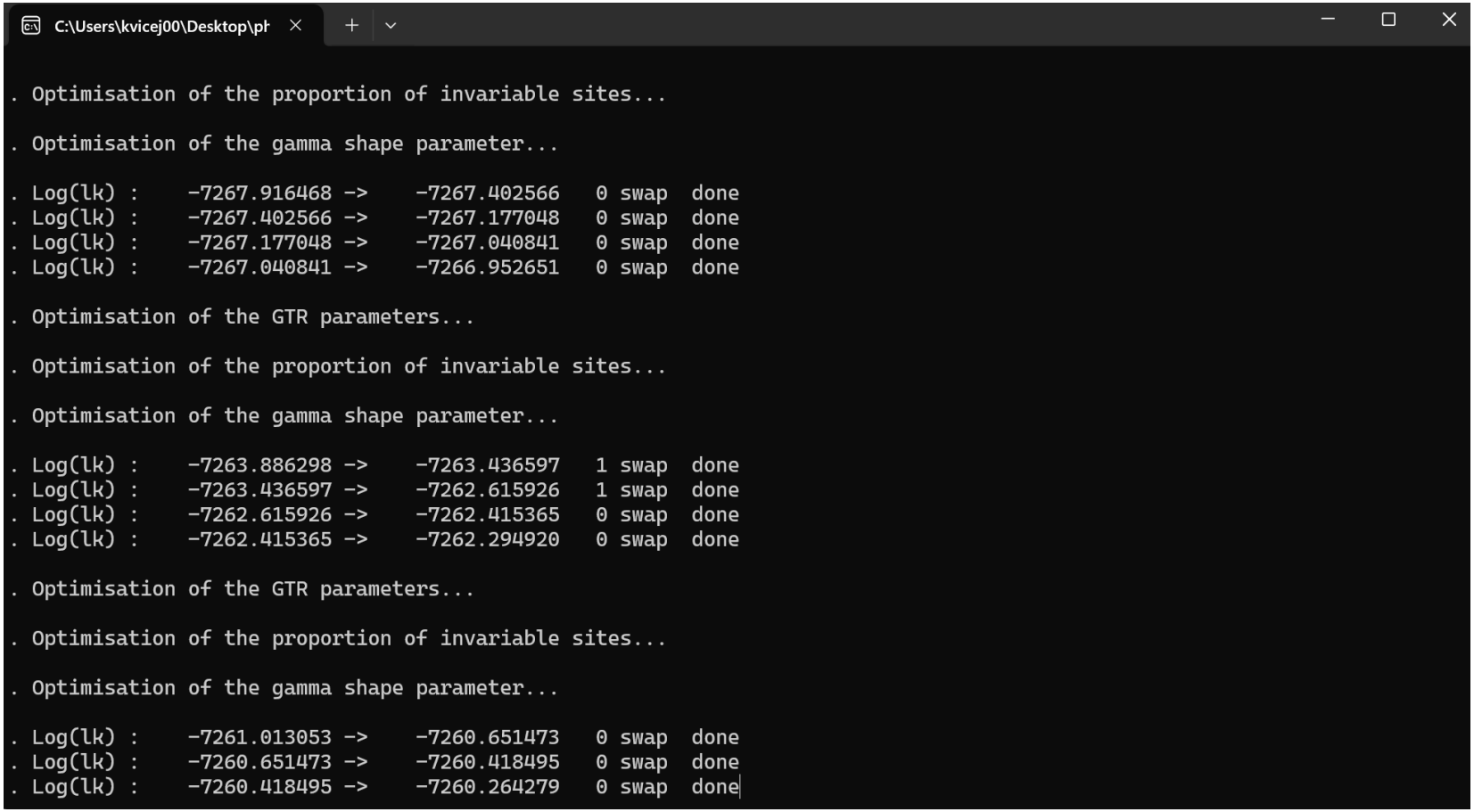
Maximální věrohodnost (ML)

PhyML



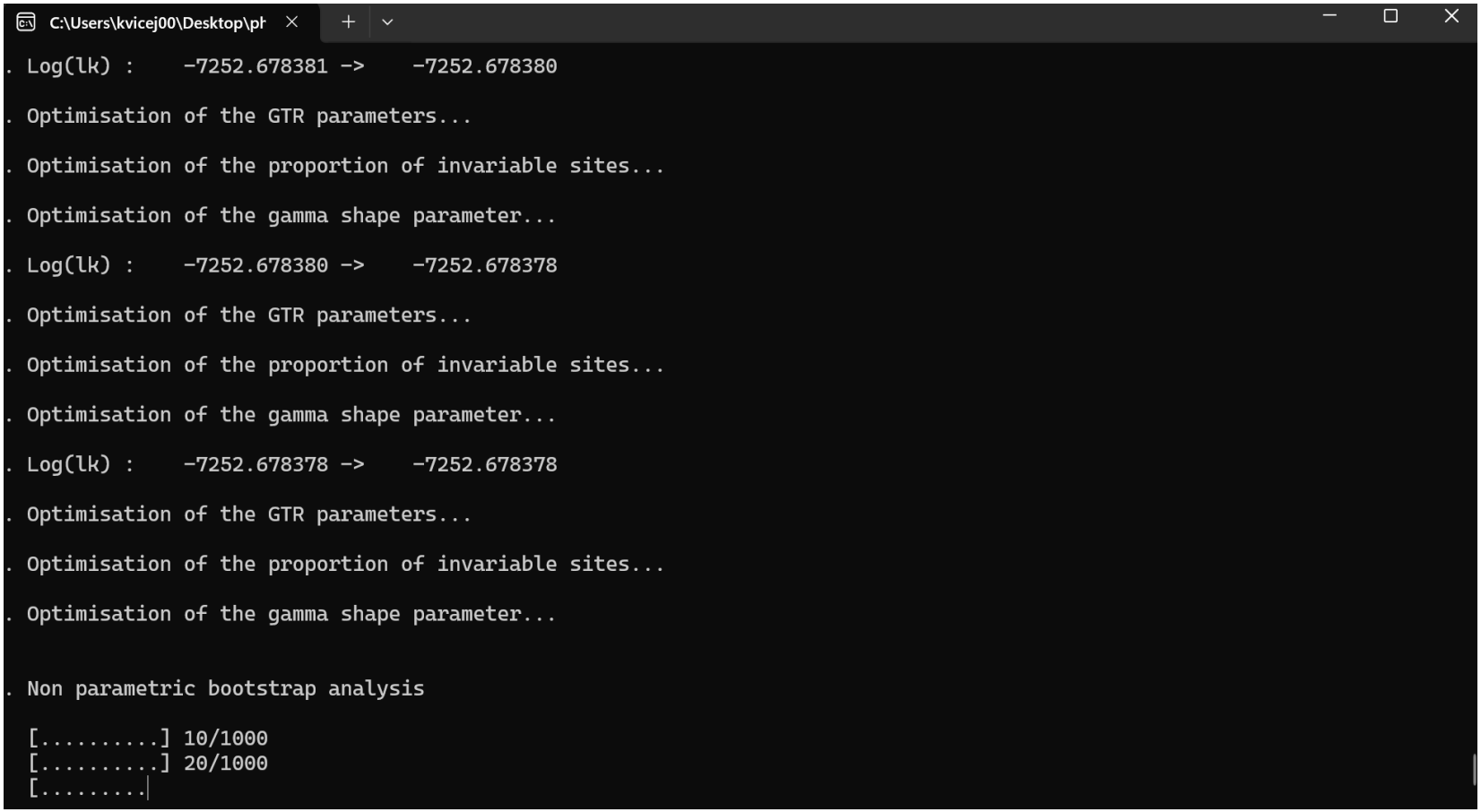
Maximální věrohodnost (ML)

PhyML



Maximální věrohodnost (ML)

PhyML



Bayesovská analýza (BI)

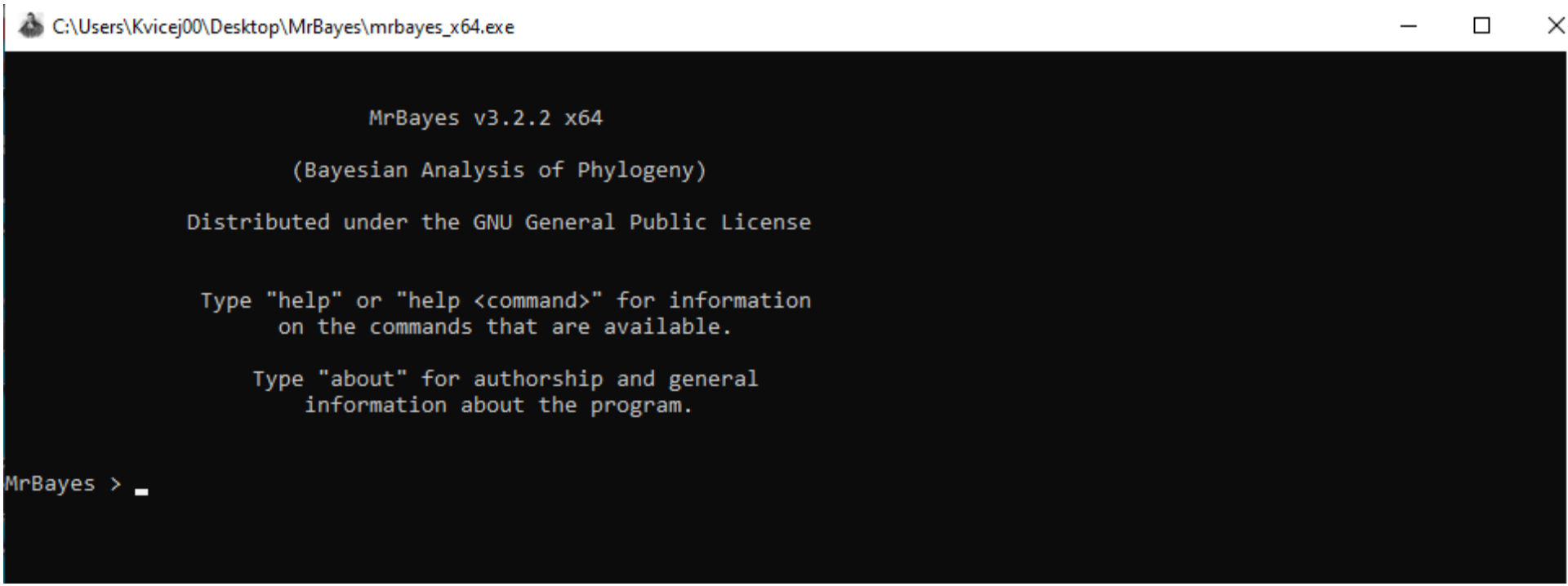
Bayesovská analýza (Bayesian Inference, BI)

→ načítá soubory ve formátu NEXUS (.nex)

Thomas Bayes

Bayesovská analýza (BI)

Mr. Bayes



Bayesovská analýza (BI)

Mr. Bayes

```
                    (Bayesian Analysis of Phylogeny)

            Distributed under the GNU General Public License


             Type "help" or "help <command>" for information
                   on the commands that are available.

                 Type "about" for authorship and general
                     information about the program.


MrBayes > execute L_zkus.nex

   Executing file "L_zkus.nex"
   DOS line termination
   Longest line length = 117
   Parsing file
   Expecting NEXUS formatted file
   Reading data block
      Allocated taxon set
      Allocated matrix
      Defining new matrix with 31 taxa and 346 characters
      Data is Dna
      Missing data coded as -
      Taxon  1 -> EU837270H
      Taxon  2 -> JX119008A
      Taxon  3 -> KX064268S
      Taxon  4 -> MK605659D
      Taxon  5 -> AJ410618S
      Taxon  6 -> MK605680K
      Taxon  7 -> MK605661K
      Taxon  8 -> KM192209S
      Taxon  9 -> KU529945S
      Taxon 10 -> JX316008T
      Taxon 11 -> KC522413T
      Taxon 12 -> MK386156T
      Taxon 13 -> KF974361S
      Taxon 14 -> NC043068A
      Taxon 15 -> JX028271M
      Taxon 16 -> KX815401P
      Taxon 17 -> AB574184P
      Taxon 18 -> MF077572P
      Taxon 19 -> MF077567P
      Taxon 20 -> MF077564P
      Taxon 21 -> MN832779P
      Taxon 22 -> KF776606P
      Taxon 23 -> 801F
      Taxon 24 -> 783F
      Taxon 25 -> MN026176P
```

Bayesovská analýza (BI)

Mr. Bayes

```
      Taxon  2 -> JX119008A
      Taxon  3 -> KX064268S
      Taxon  4 -> MK605659D
      Taxon  5 -> AJ410618S
      Taxon  6 -> MK605680K
      Taxon  7 -> MK605661K
      Taxon  8 -> KM192209S
      Taxon  9 -> KU529945S
      Taxon 10 -> JX316008T
      Taxon 11 -> KC522413T
      Taxon 12 -> MK386156T
      Taxon 13 -> KF974361S
      Taxon 14 -> NC043068A
      Taxon 15 -> JX028271M
      Taxon 16 -> KX815401P
      Taxon 17 -> AB574184P
      Taxon 18 -> MF077572P
      Taxon 19 -> MF077567P
      Taxon 20 -> MF077564P
      Taxon 21 -> MN832779P
      Taxon 22 -> KF776606P
      Taxon 23 -> 801F
      Taxon 24 -> 783F
      Taxon 25 -> MN026176P
      Taxon 26 -> MF077566P
      Taxon 27 -> MN026178P
      Taxon 28 -> MN026173P
      Taxon 29 -> KF425497D
      Taxon 30 -> KF177176D
      Taxon 31 -> MK605679D
      Successfully read matrix
      Setting default partition (does not divide up characters)
      Setting model defaults
      Seed (for generating default start values) = 1733088771
      Setting output file names to "L_zkus.nex.run<i>.<p|t>"
   Exiting data block
   Reached end of file

MrBayes > lset nst=6 rates=invgamma ngammacat=4

   Setting Nst to 6
   Setting Rates to Invgamma
   Setting Ngammacat to 4
   Successfully set likelihood model parameters

MrBayes > mcmc ngen= 10000000
```

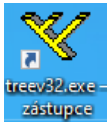
Bayesovská analýza (BI)

Mr. Bayes

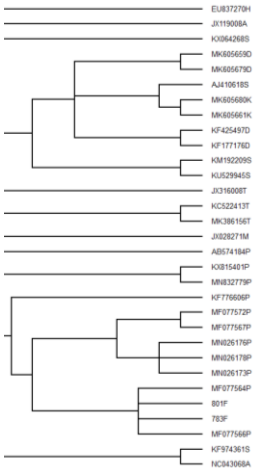
```
Chain results (10000000 generations requested):

    0 -- [-7001.037] (-7423.560) (-7310.681) (-7471.296) * [-7262.191] (-7463.806) (-7017.065) (-7446.010)
  500 -- [-4520.078] (-4568.591) (-4567.438) (-4851.601) * (-4710.812) (-4842.278) [-4640.954] (-4755.677) -- 0:00:00
 1000 -- [-4191.650] (-4311.517) (-4349.056) (-4403.381) * (-4296.386) (-4245.397) [-4264.047] (-4304.791) -- 2:46:39
 1500 -- [-4062.136] (-4172.653) (-4192.854) (-4137.050) * (-4207.548) [-4071.215] (-4170.401) (-4113.387) -- 1:51:05
 2000 -- [-3999.329] (-4055.607) (-4131.093) (-4062.699) * (-4147.245) [-3995.171] (-4062.825) (-4011.408) -- 2:46:38
 2500 -- (-4004.056) [-3957.566] (-4076.370) (-3994.656) * (-4120.848) [-3953.068] (-3957.554) (-4004.623) -- 2:13:18
 3000 -- (-3980.809) [-3932.391] (-4011.878) (-3978.856) * (-4088.130) [-3933.328] (-3931.256) (-3972.958) -- 1:51:04
 3500 -- (-3957.155) [-3910.655] (-3971.395) (-3947.220) * (-4044.541) [-3926.653] (-3918.208) (-3925.643) -- 2:22:48
 4000 -- [-3921.520] (-3921.043) (-3963.030) (-3919.616) * (-4012.472) (-3911.313) (-3911.143) [-3912.661] -- 2:04:57
 4500 -- (-3925.067) (-3903.955) [-3915.394] (-3909.239) * (-3964.335) (-3915.977) [-3906.873] (-3895.661) -- 2:28:04
 5000 -- (-3929.571) (-3906.815) [-3898.958] (-3897.281) * (-3966.163) (-3916.646) (-3911.149) [-3888.063] -- 2:13:16

Average standard deviation of split frequencies: 0.111959

 5500 -- (-3930.309) (-3910.432) [-3909.824] (-3911.371) * (-3937.099) (-3904.676) (-3915.333) [-3890.150] -- 2:31:25
 6000 -- (-3910.707) (-3916.071) (-3916.499) [-3887.104] * (-3919.871) (-3893.397) (-3908.263) [-3899.397] -- 2:18:48
 6500 -- [-3893.685] (-3911.698) (-3903.295) (-3898.103) * (-3922.663) (-3881.209) [-3903.879] (-3891.859) -- 2:33:44
 7000 -- (-3895.051) (-3901.029) [-3899.044] (-3897.073) * (-3909.663) [-3886.384] (-3910.816) (-3887.637) -- 2:22:45
 7500 -- (-3904.609) (-3903.049) (-3898.748) [-3895.102] * (-3914.970) (-3886.157) (-3917.214) [-3886.752] -- 2:35:26
 8000 -- (-3892.106) (-3893.815) [-3904.499] (-3899.277) * (-3911.833) [-3887.747] (-3910.232) (-3903.989) -- 2:25:43
 8500 -- [-3902.334] (-3894.355) (-3902.532) (-3888.610) * [-3905.619] (-3888.486) (-3903.637) (-3904.968) -- 2:17:08
 9000 -- (-3902.992) [-3889.185] (-3898.782) (-3880.044) * [-3891.998] (-3891.271) (-3888.769) (-3910.201) -- 2:28:00
 9500 -- (-3902.893) (-3902.981) (-3897.013) [-3899.451] * (-3889.090) [-3900.555] (-3900.639) (-3906.465) -- 2:20:13
10000 -- (-3895.596) [-3888.290] (-3900.646) (-3906.445) * (-3895.943) [-3892.994] (-3903.430) (-3908.482) -- 2:29:51

Average standard deviation of split frequencies: 0.106066

10500 -- (-3895.464) (-3896.796) [-3899.754] (-3916.701) * (-3905.663) [-3886.282] (-3925.605) (-3898.088) -- 2:22:42
11000 -- (-3892.916) [-3892.540] (-3898.288) (-3907.304) * [-3895.142] (-3889.422) (-3901.840) (-3896.691) -- 2:31:20
11500 -- (-3891.138) [-3887.290] (-3897.505) (-3898.465) * [-3891.534] (-3906.512) (-3896.863) (-3896.493) -- 2:24:45
12000 -- (-3886.158) (-3895.316) (-3899.951) [-3895.952] * (-3901.181) (-3909.347) (-3896.408) [-3883.249] -- 2:32:35
12500 -- [-3887.538] (-3902.739) (-3902.760) (-3912.968) * [-3892.912] (-3897.436) (-3906.132) (-3899.721) -- 2:26:29
13000 -- (-3895.558) (-3902.703) [-3903.714] (-3914.507) * (-3890.671) [-3909.579] (-3903.474) (-3897.025) -- 2:20:50
13500 -- [-3891.386] (-3903.404) (-3898.820) (-3912.319) * (-3901.203) [-3909.100] (-3904.095) (-3894.503) -- 2:27:56
14000 -- [-3893.204] (-3889.766) (-3901.741) (-3916.616) * (-3892.042) (-3906.833) (-3900.489) [-3882.750] -- 2:22:39
14500 -- (-3907.001) [-3889.357] (-3882.366) (-3921.408) * (-3892.004) (-3916.012) [-3896.471] (-3885.342) -- 2:29:12
15000 -- (-3903.163) [-3885.333] (-3885.999) (-3920.101) * [-3888.005] (-3912.557) (-3907.089) (-3893.594) -- 2:24:13

Average standard deviation of split frequencies: 0.079817

15500 -- (-3911.243) [-3890.943] (-3901.495) (-3904.671) * [-3880.738] (-3896.484) (-3902.559) (-3896.746) -- 2:30:18
16000 -- (-3908.134) (-3896.962) [-3891.923] (-3910.491) * [-3890.882] (-3909.738) (-3893.035) (-3904.252) -- 2:25:36
16500 -- (-3901.485) (-3896.219) [-3900.147] (-3921.860) * [-3885.614] (-3912.660) (-3889.887) (-3907.672) -- 2:31:15
17000 -- (-3894.608) [-3892.851] (-3902.664) (-3905.756) * (-3890.610) (-3896.670) (-3895.539) [-3904.515] -- 2:26:48
17500 -- (-3904.587) [-3894.477] (-3886.298) (-3920.261) * [-3902.131] (-3900.916) (-3908.223) (-3910.062) -- 2:32:06
18000 -- (-3893.863) (-3892.201) [-3880.665] (-3910.568) * (-3913.601) [-3890.487] (-3903.432) (-3908.082) -- 2:27:52
18500 -- (-3893.014) [-3884.708] (-3890.133) (-3892.454) * (-3919.170) (-3898.925) (-3901.211) [-3909.553] -- 2:23:52
19000 -- (-3899.257) [-3893.612] (-3899.712) (-3909.788) * (-3914.812) (-3897.487) (-3910.528) [-3898.014] -- 2:28:50
```

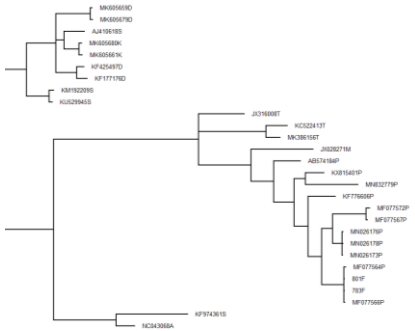
Vizualizace a úprava fylogenetických stromů

TreeView



- vizualizace stromu ve formátu **.tre** vytvořeného fylogenetickým programem
- možnost vizualizace různých typů stromů (s délkou větví, bez délky větví)
- možnost zobrazení statistických podpor uzlů
- možnost nastavení outgroupu / outgroupů a zakořenění stromu
- možnost uložení ve formátu **.emf** umožňujícím grafické úpravy v dalších programech
- možnost tisku hrubého, graficky neupraveného stromu





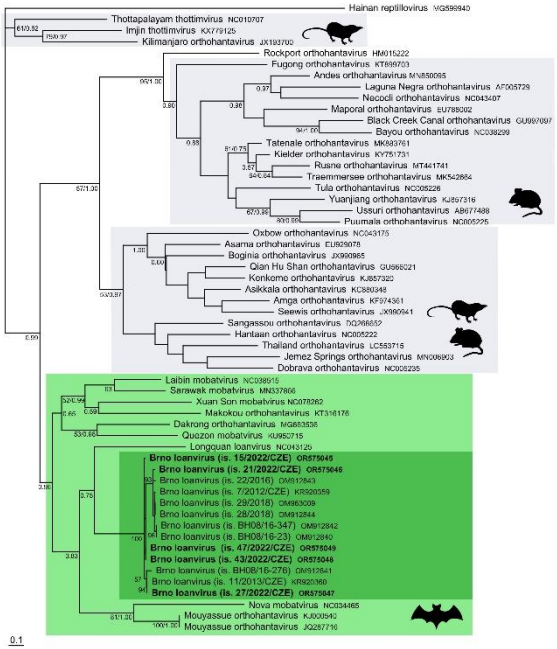
Vizualizace a úprava fylogenetických stromů

Finální grafické úpravy fylogenetických stromů:

FigTree

Adobe Illustrator

- vizualizace hrubého, graficky neupraveného stromu, ve formátech **.emf**, **.ai** nebo **.pdf**
- možnost textových úprav (názvy taxonů, velikost a typ písma)
- možnost barevných úprav a vkládání obrázků



Odkazy

- https://www.ncbi.nlm.nih.gov
- https://ictv.global
- https://mafft.cbrc.jp/alignment/server/index.html
- https://phylipweb.github.io/phylip/software.html
- https://molbiol-tools.ca/Phylogeny.htm