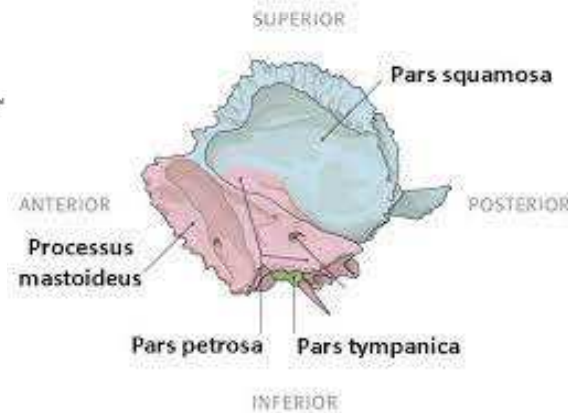
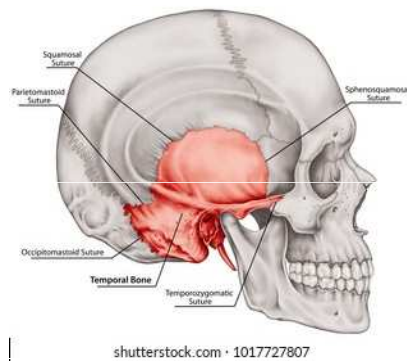


ARCHAICKÁ DNA

Paleogenetika = studium vymřelých organismů prostřednictvím archaické DNA (aDNA, *ancient DNA*)

zdroje aDNA: zuby, kosti (kost skalní), muzejní exponáty atd., ale také např. vaječné skořápky, koprolity, rozžvýkaná březová kůra, půda, proteinové obaly hnid vši šatních



další zdroje: keratin, chitin, osteokalcin, kolagen, hemoglobin, pigmenty, rostlinné polysacharidy → proteomické metody

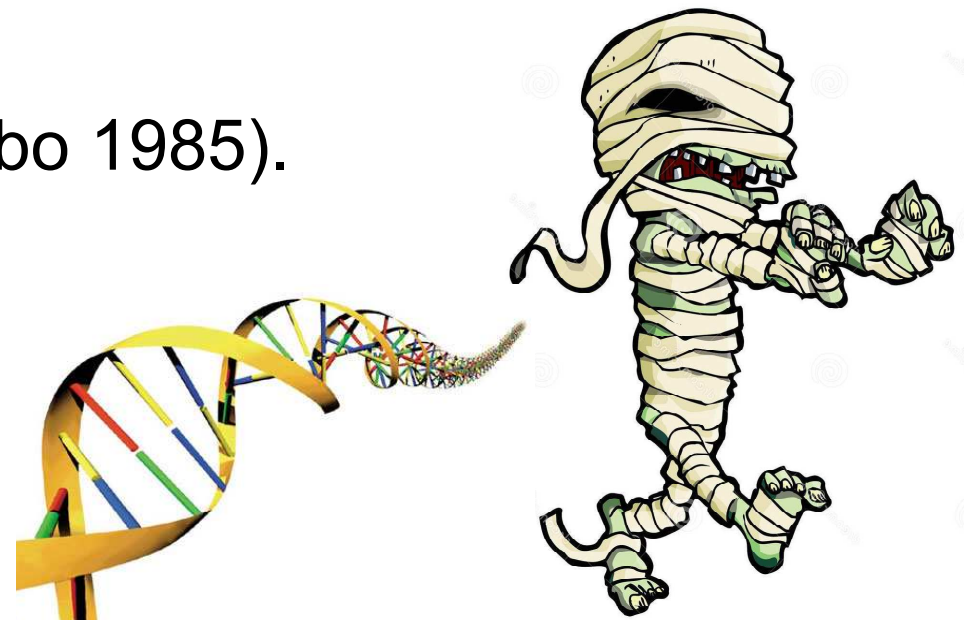
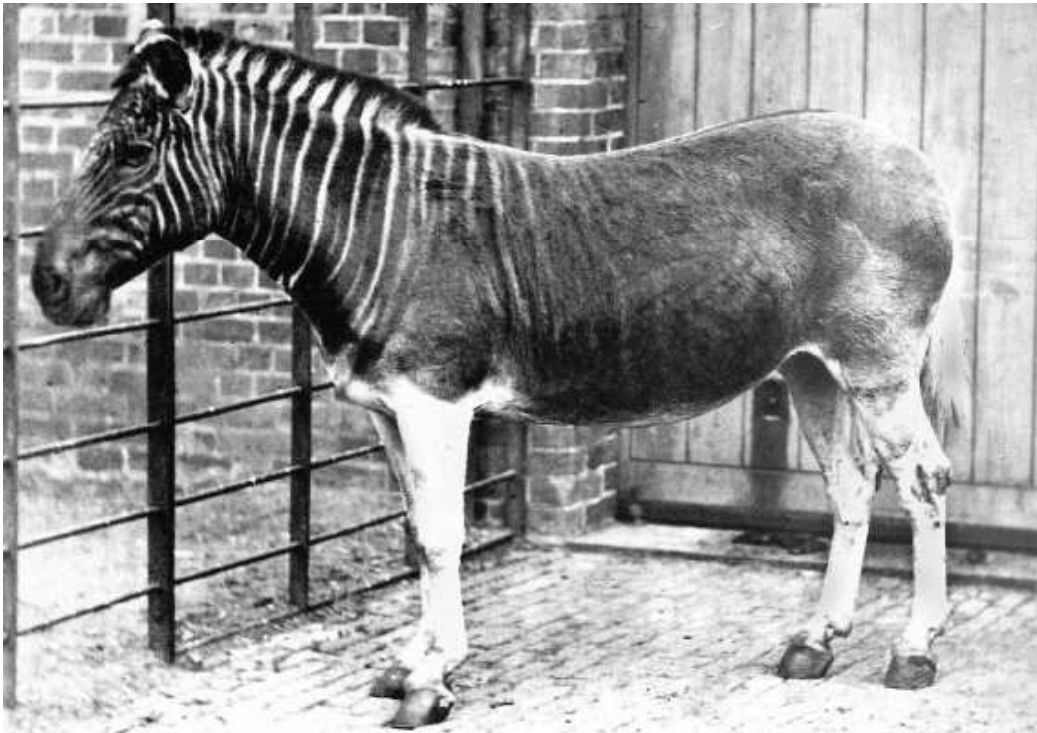
= paleoproteomika

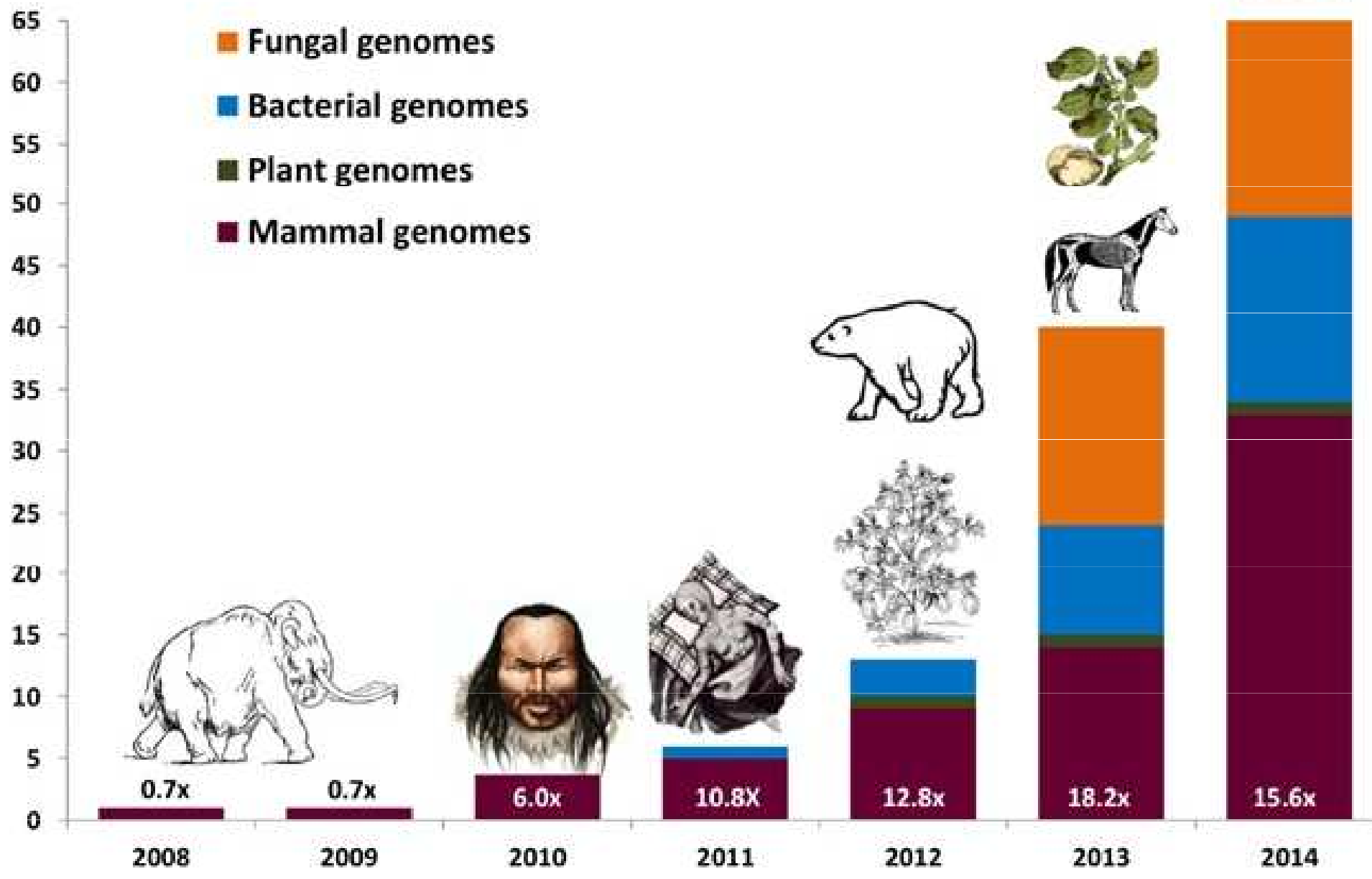
Výhoda: nedochází k fragmentaci proteinů ⇒ větší časové škály

Nevýhoda: proteiny neposkytují tolik informací jako DNA

První sekvence aDNA: zebra kvaga (*Equus quagga*)
150 let (Higuchi et al. 1984)

egyptská mumie: 2400 let (Pääbo 1985).

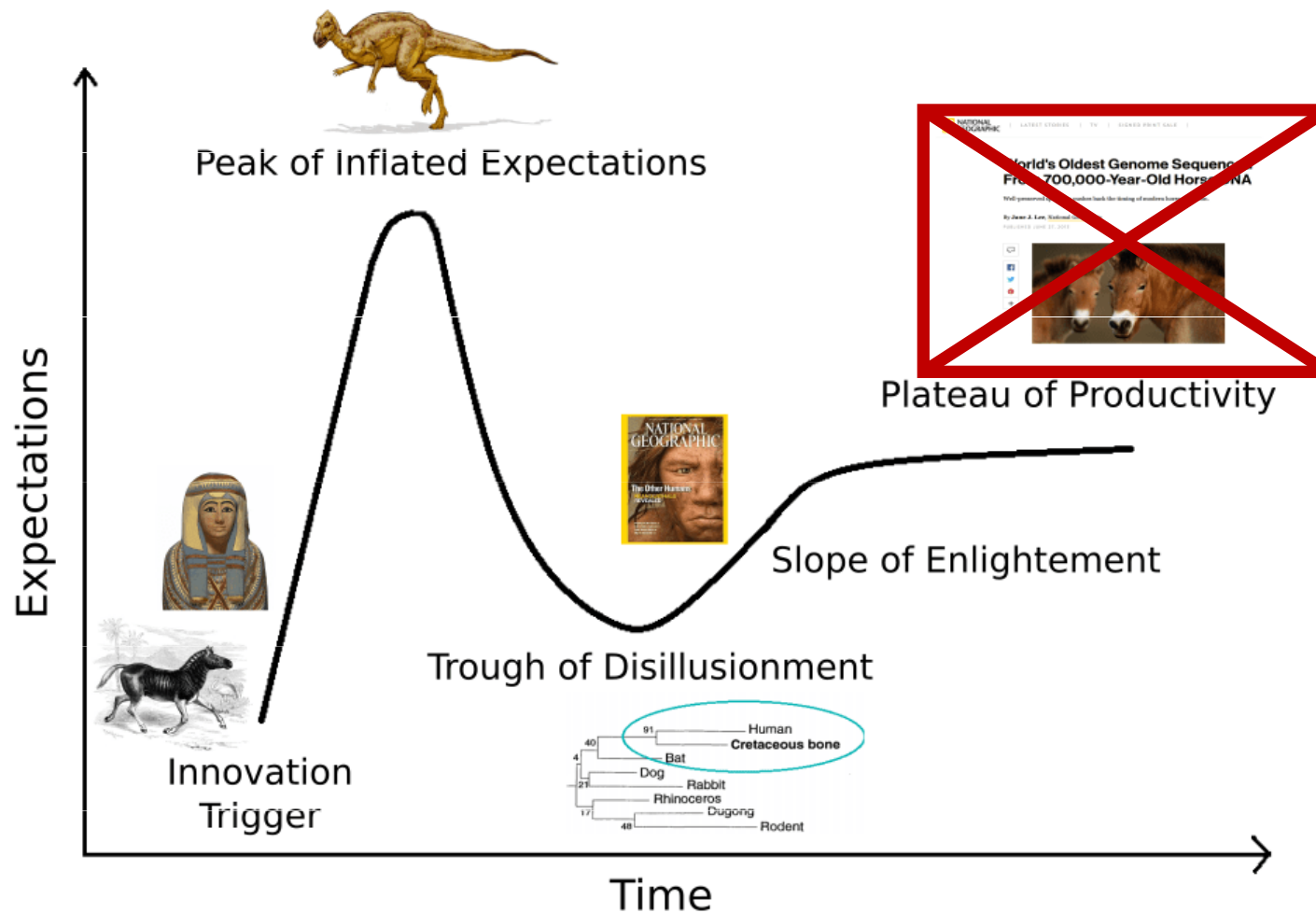




Stáří aDNA:

většinou max. tisíce až desítky tisíc let

The Hype Cycle of Ancient DNA



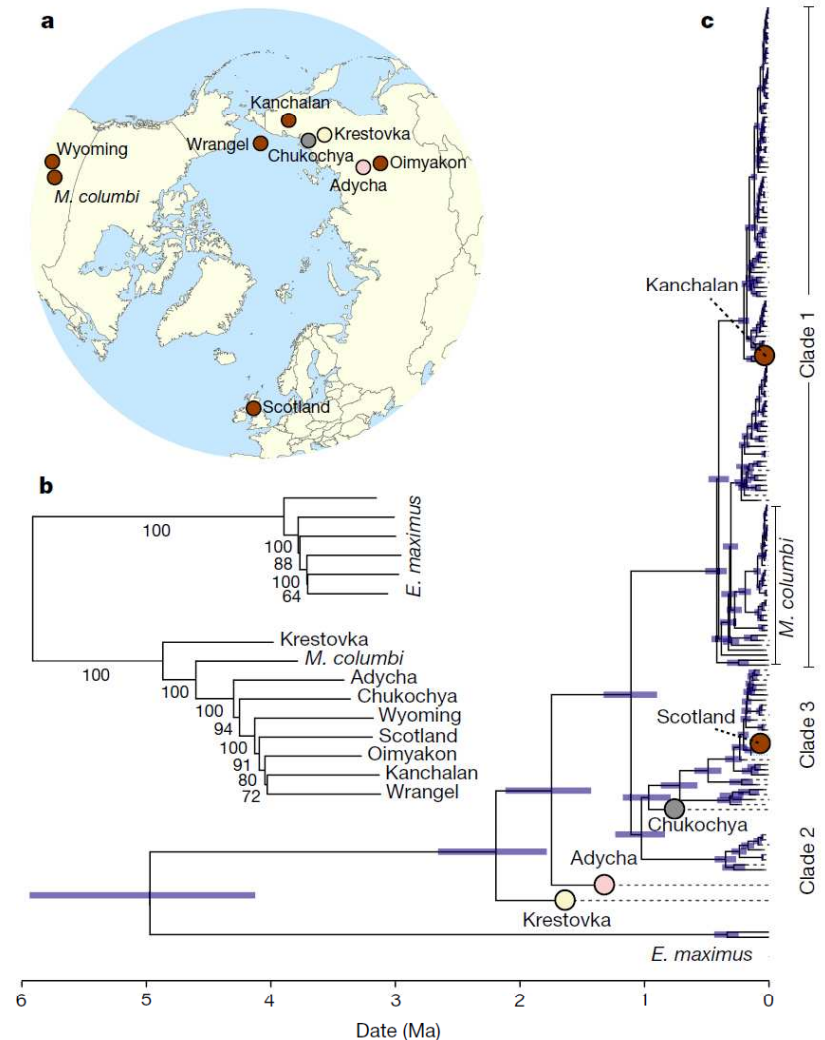
mamut ze sibiřského permafrostu:
 1,10 – 1,65 mil. let
 van der Valk et al., Nature 2021



DNA from multiple mammoth species is illuminating a complex evolutionary picture.

MAMMOTH GENOMES SHATTER RECORD FOR OLDEST ANCIENT DNA

Permafrost-preserved teeth, up to 1.65 million years old, identify a new kind of mammoth in Siberia.



Million-year-old DNA sheds light on the genomic history of mammoths

<https://doi.org/10.1038/s41586-021-03224-9>

Received: 3 July 2020

Accepted: 11 January 2021

Published online: 17 February 2021

Tom van der Valk^{1,2,3,17}✉, Patrícia Pečnerová^{2,4,5,17}, David Díez-del-Molino^{1,2,4,17}, Anders Bergström⁶, Jonas Oppenheimer⁷, Stefanie Hartmann⁸, Georgios Xenikoudakis⁸, Jessica A. Thomas⁸, Marianne Dehasque^{1,2,4}, Ekin Sağlıcan⁹, Fatma Rabia Fidan⁹, Ian Barnes¹⁰, Shanlin Liu¹¹, Mehmet Somel⁹, Peter D. Heintzman¹², Pavel Nikolskiy¹³, Beth Shapiro^{14,15}, Pontus Skoglund⁶, Michael Hofreiter⁸, Adrian M. Lister¹⁰, Anders Götherström^{1,16,18} & Love Dalén^{1,2,4,18}✉

Check for updates

environmentální DNA: 2 mil. let (Kjaer et al., Nature 2022)

A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA


<https://doi.org/10.1038/s41586-022-05453-y>

Received: 30 September 2021

Accepted: 18 October 2022

Published online: 7 December 2022

Open access

 Check for updates

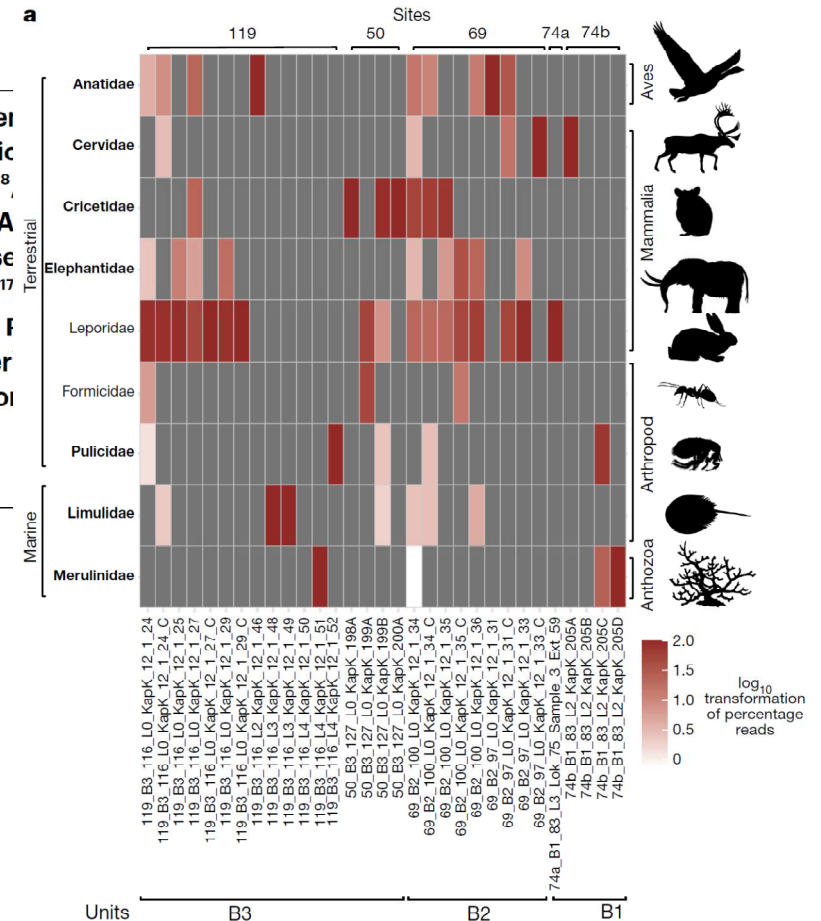
Kurt H. Kjær^{1,27}✉, Mikkel Winther Pedersen
Thorfinn S. Korneliussen¹, Christian S. Mic
Anthony H. Ruter¹, Astrid M. A. Schmidt^{7,8},
Ian Snowball¹¹, John C. Gosse¹², Inger G. A
Magnus Rasmussen¹⁴, Morten E. Jørgense
Jeppe Å. Kristensen^{15,16}, Morten Bjerager¹⁷
PhyloNorway Consortium^{*,**}, Alexandra I
Antonio Fernandez-Guerra¹, Chris Bowler
John J. Welch³, Alan J. Hidy²³, Martin Siko
Nicolaj K. Larsen¹ & Eske Willerslev^{1,2,26}✉

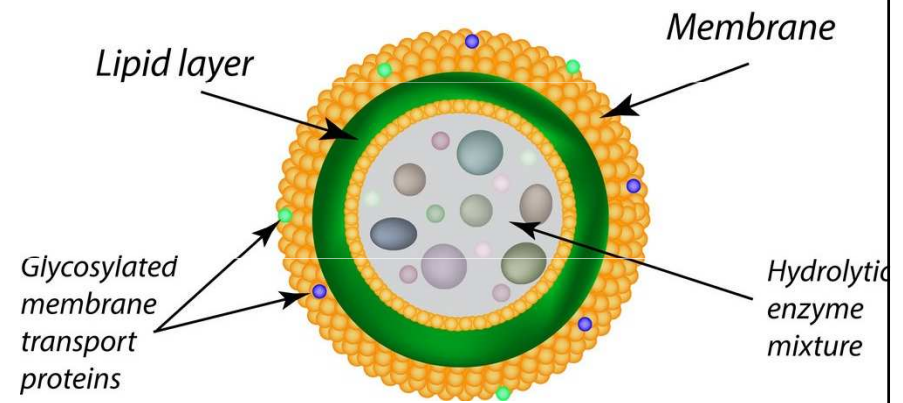
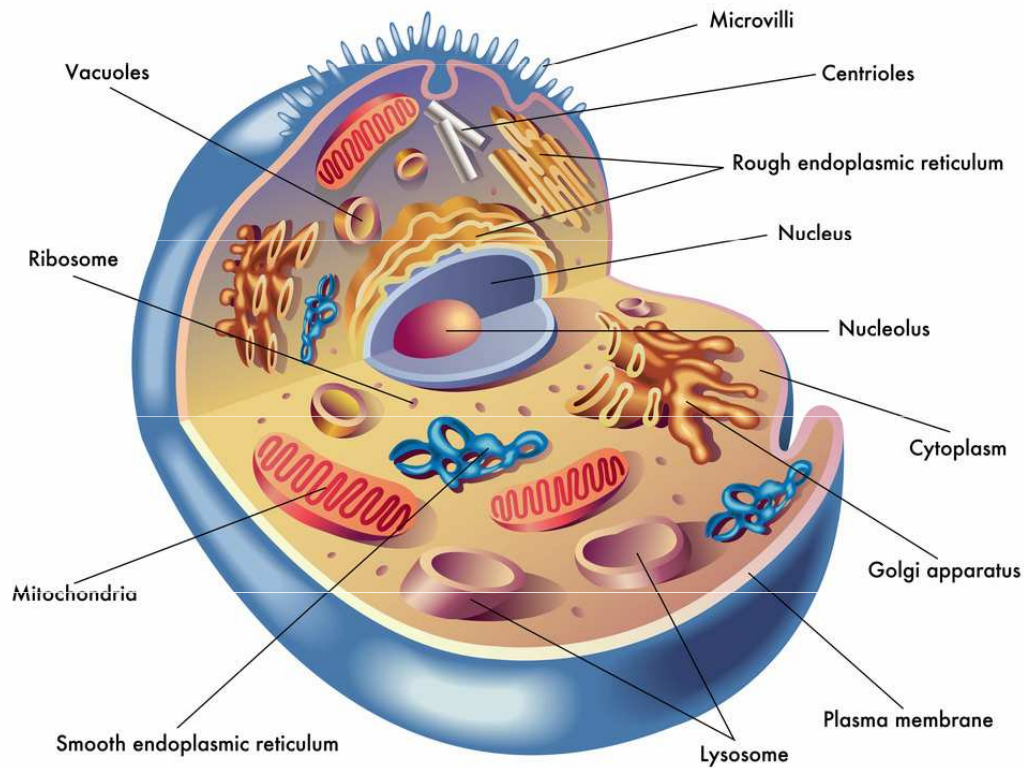
proteinové sekvence:

2019: nosorožec, 1,77 mil., Dmanisi

2016: pštros, 3,8 mil., Laetoli

analýza proteinů vzorků až z rané jury (ale ne sekvence)





Degradace DNA ihned po zániku buňky:

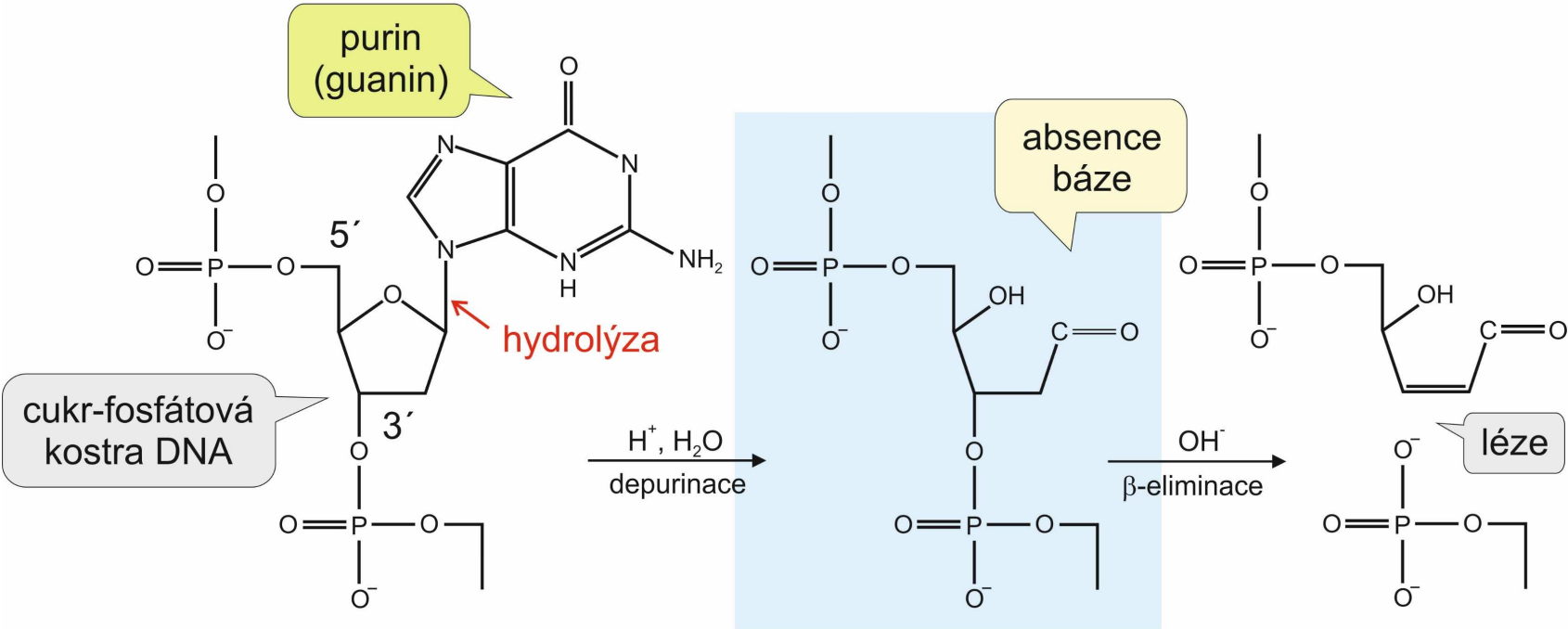
→ fragmentace

→ poškození blokující replikaci DNA

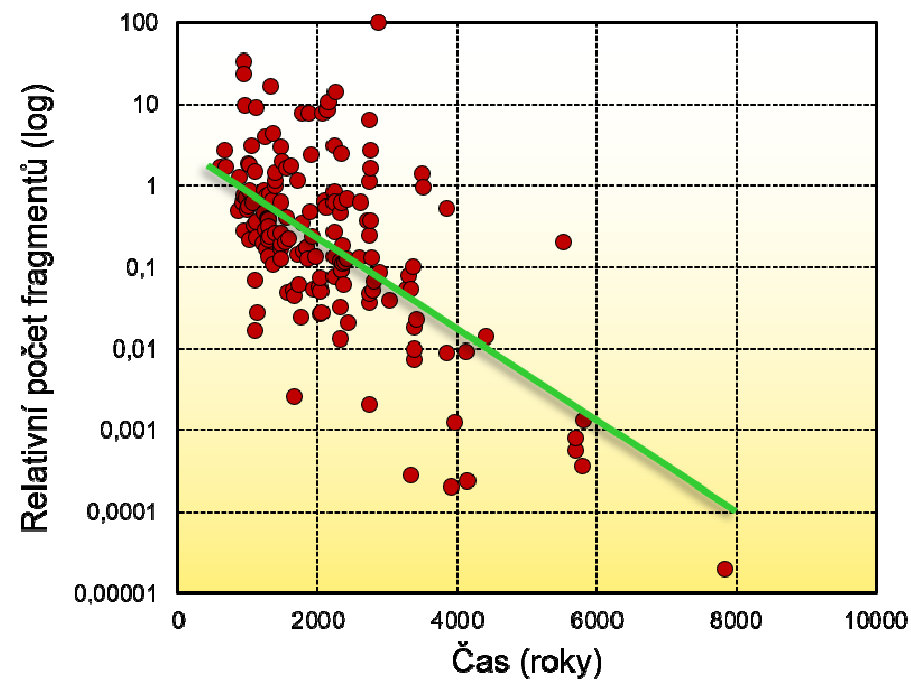
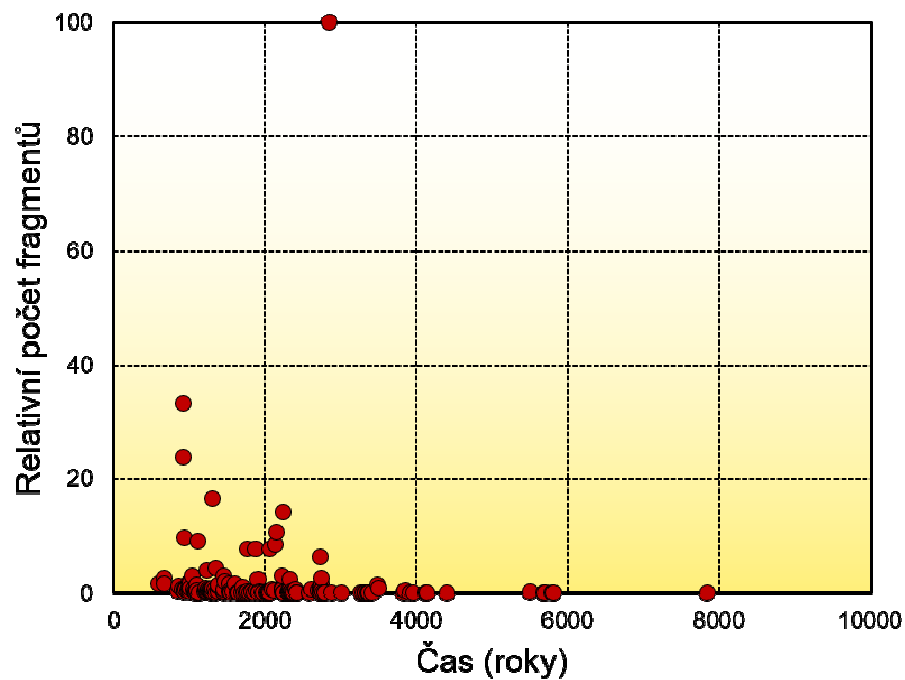
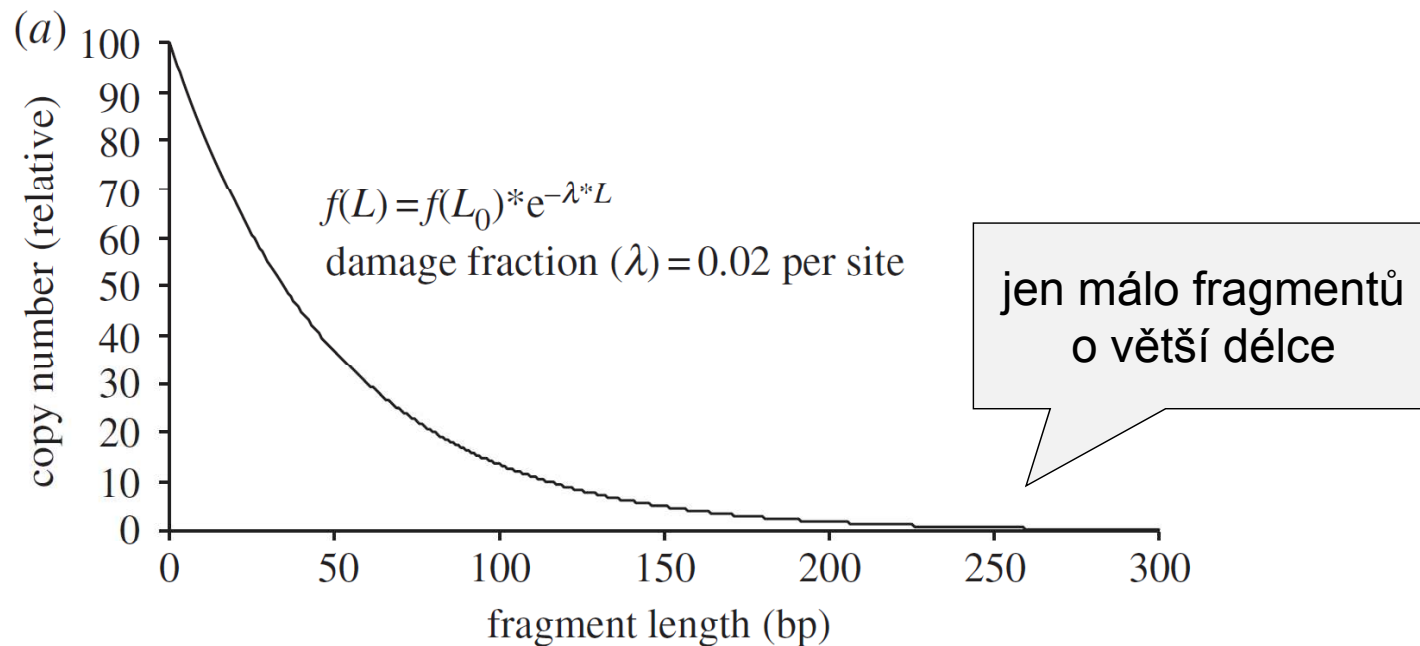
→ poškození způsobující inkorporaci nesprávných bází

Fragmentace:

většinou < 100 bp



Délky fragmentů se zmenšují zhruba exponenciálně



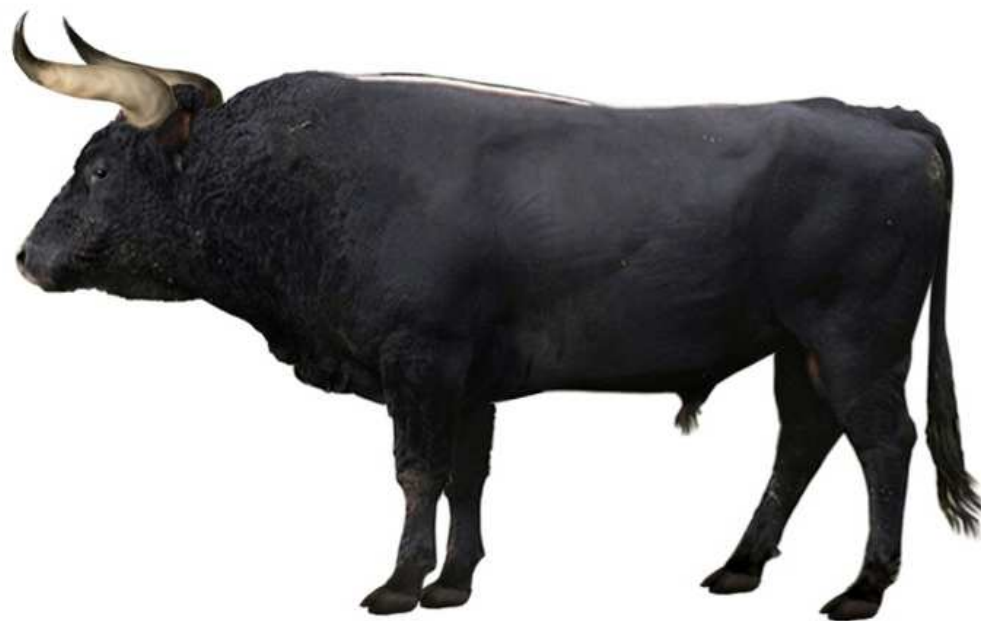
Pruvost et al. (2007):

2 cca. 3200 let staré vzorky téhož jedince pratora (*Bos primigenius*):

- 1) 1947, muzejní sbírky, 54 let → žádná použitelná aDNA
- 2) 2004 → autentická aDNA

To znamená, že 99 % DNA degradováno během pár desítek let

⇒ rychlost rozkladných procesů byla v muzejním prostředí více než 70× vyšší než v průběhu předchozích ~3200 let v půdě!



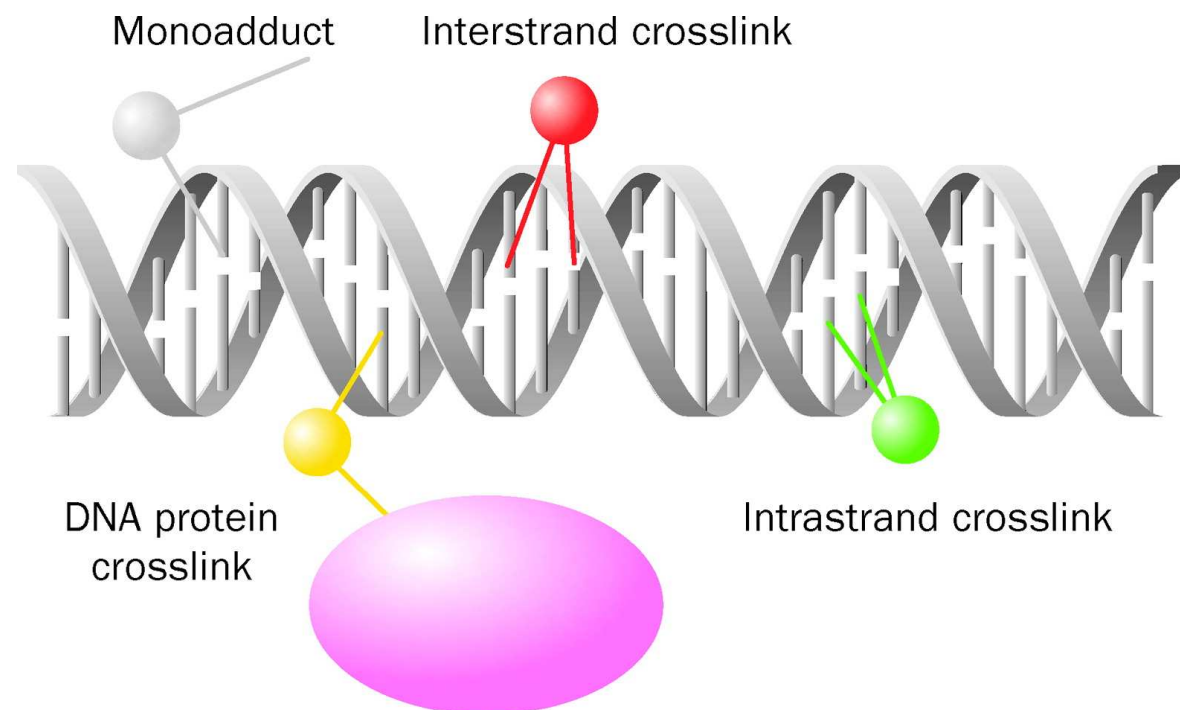
Blokující poškození:

Blokování postupu DNA polymerázy při PCR

modifikace nukleotidů

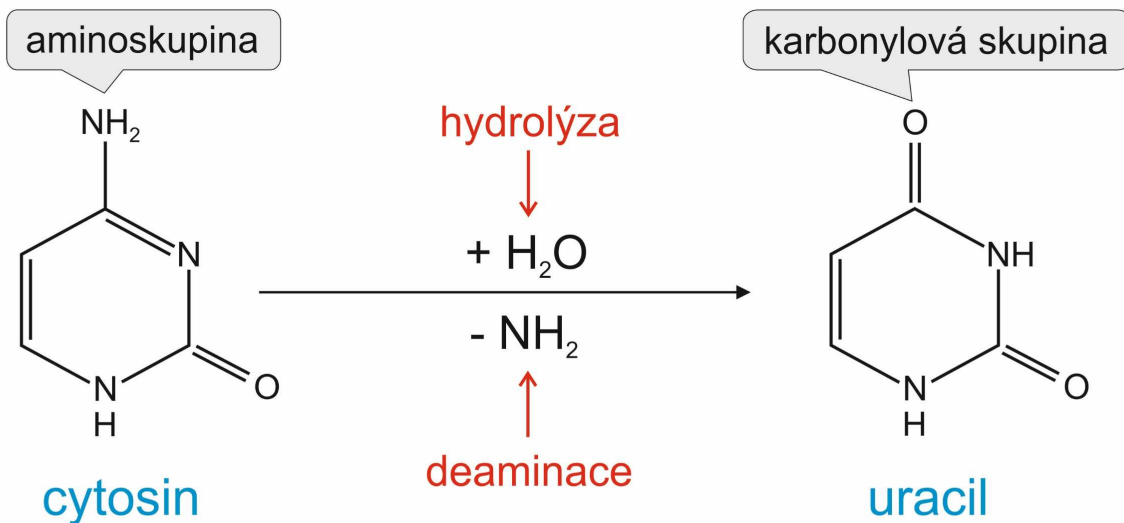
příčné vazby (*crosslinks*):

v rámci jednoho řetězce
mezi řetězci téže molekuly
mezi různými molekulami
mezi DNA a proteinem

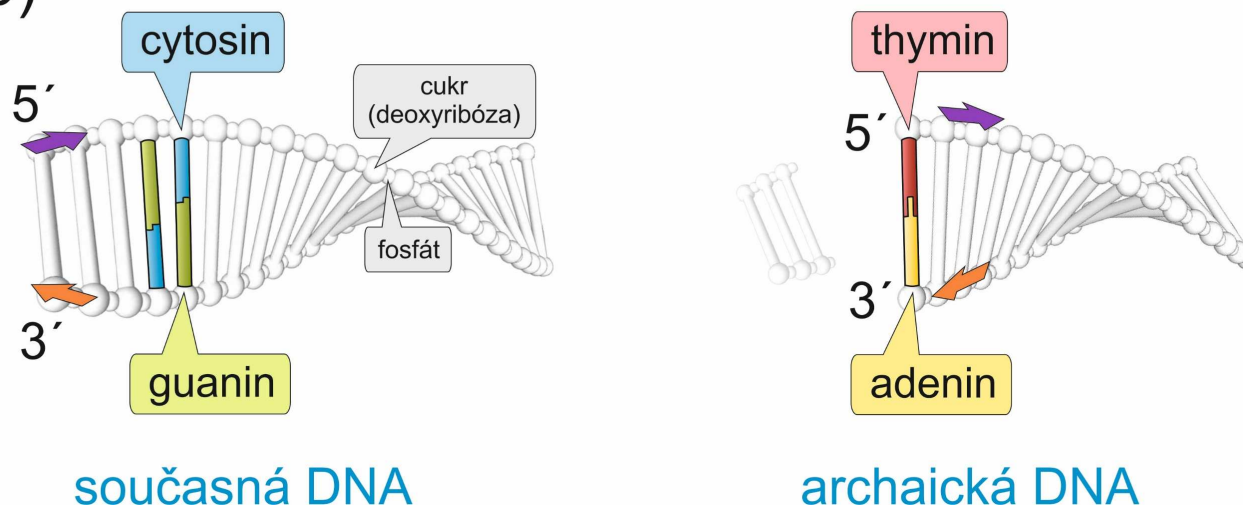


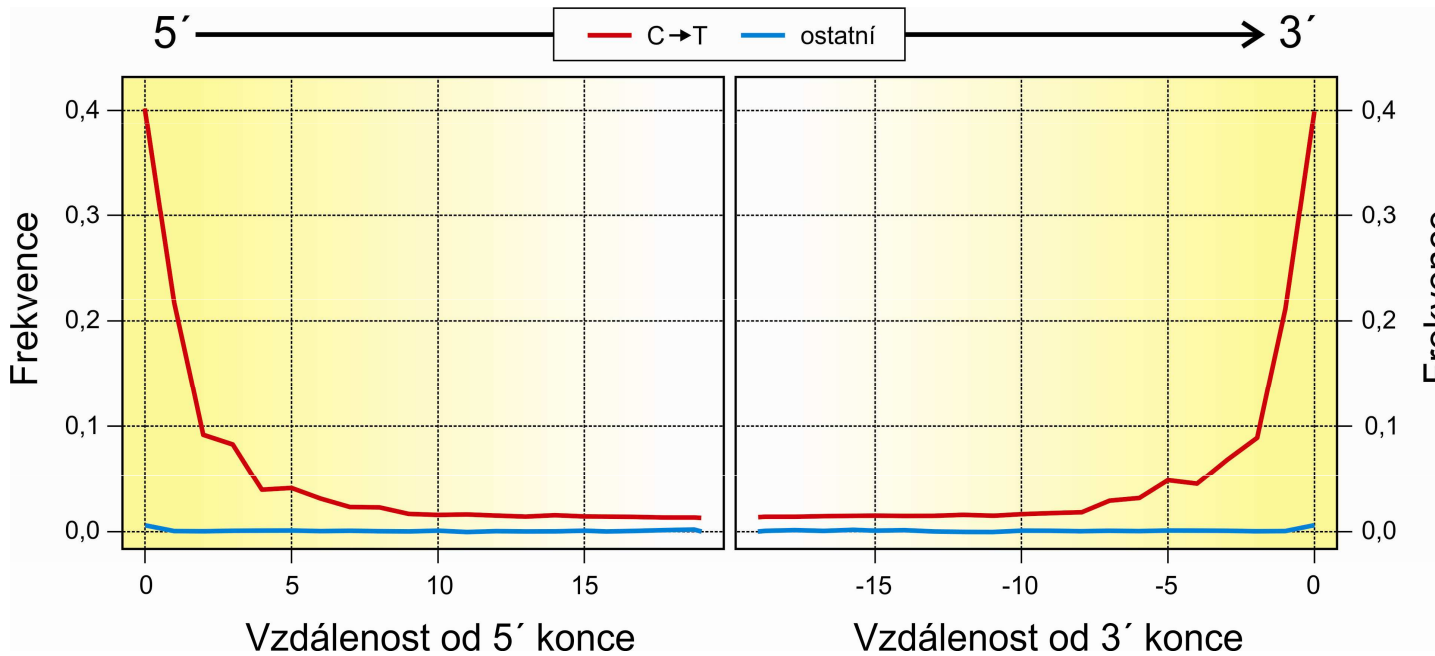
Inkorporace chybného nukleotidu:

a)

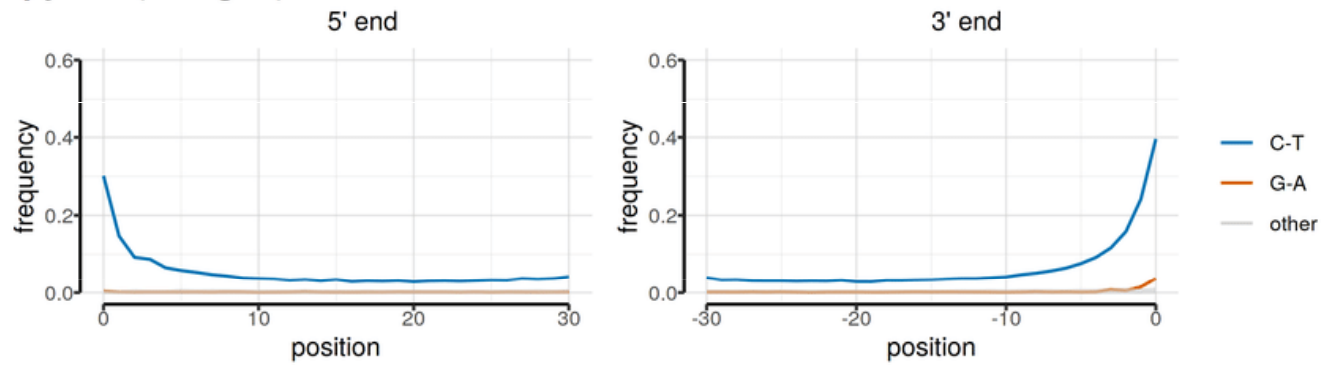


b)

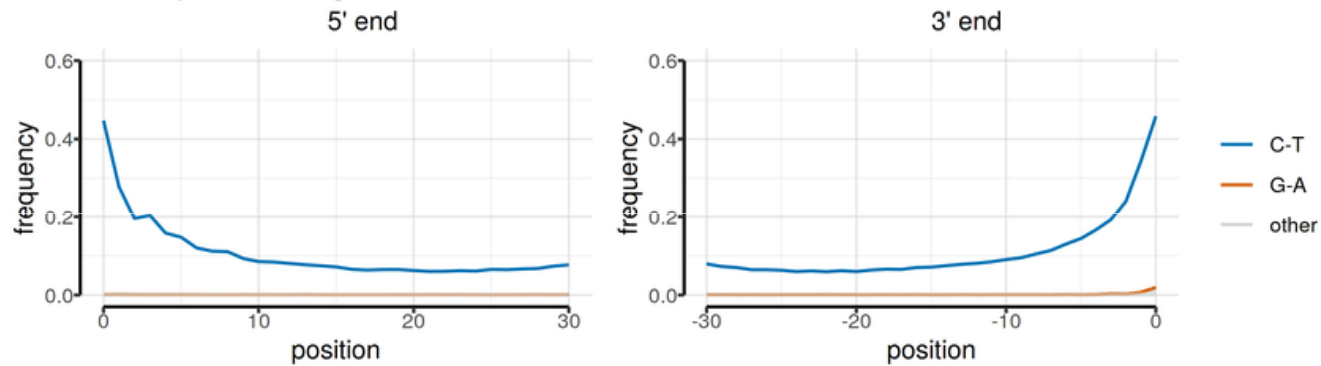




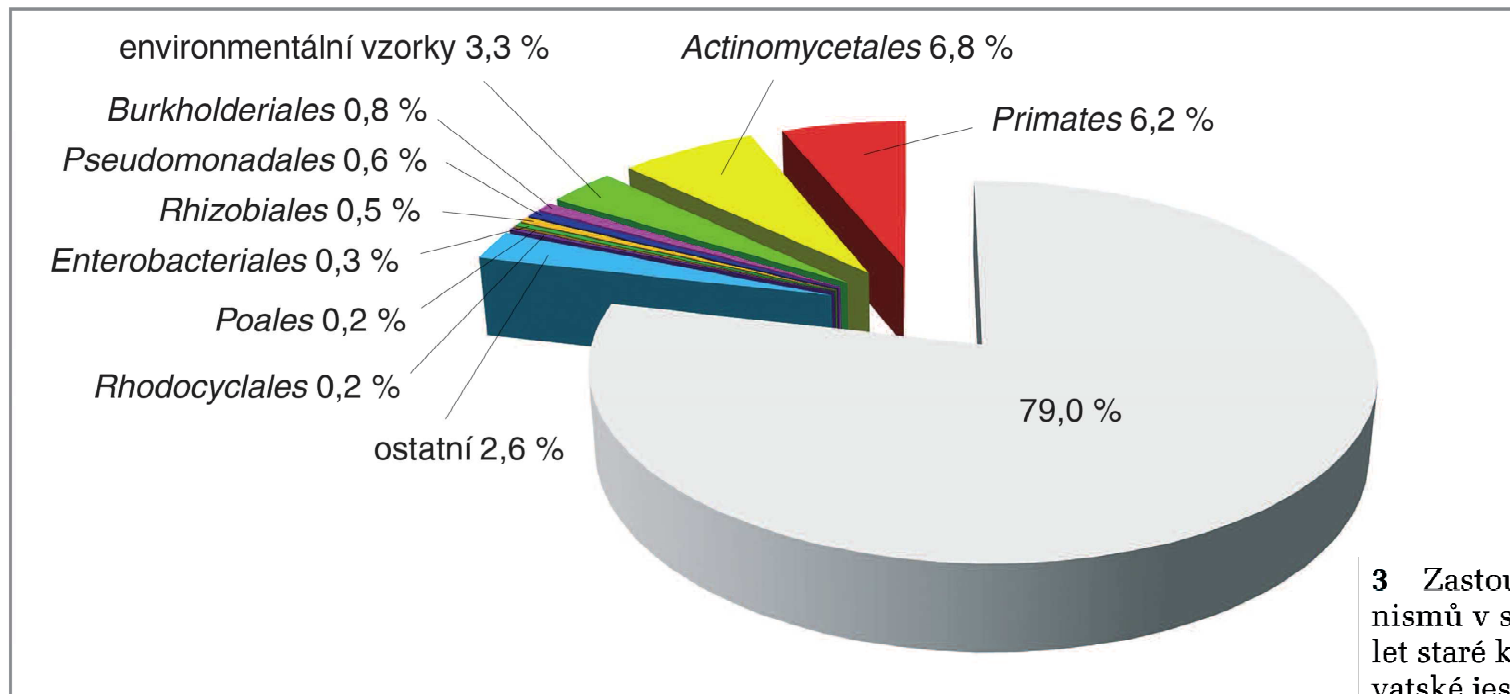
Spy 94a (shotgun)



Mezmaiskaya 2 (shotgun)



Kontaminace



3 Zastoupení nejhojnějších řádů organismů v sekvencích získaných z 38 tisíc let staré kosti neandertálce Vi-80 z chorvatské jeskyně Vindija. Největší podíl v sekvenovaných úsecích DNA měly půdní bakterie řádu *Actinomycetales* (6,8 %) a primáti (6,2 %). Průměrná shoda s databázovými sekvencemi primátů byla 98,8 % (u ostatních řádů 92–98 %), což ukazuje na blízkou taxonomickou příbuznost identifikovaných sekvencí primátů ve srovnání s ostatními řády. Skupina tzv. environmentálních vzorků, což jsou neidentifikované organismy vyskytující se ve vzorcích půdy, mořské vody, obsahu střeva apod., tvořila 3,3 % sekvencí. Za povšimnutí stojí, že naprostou většinu (79 %) sekvencí se nepodařilo přiřadit k žádné databázové sekvenci. Upraveno podle: R. E. Green a kol. (2006)



PRÁCE SE SOUBORY

DNA databáze:

EMBL (European Molecular Biology Laboratory) – European Bioinformatics Institute, Hinxton, UK: <http://www.ebi.ac.uk/embl/>

GenBank – NCBI (National Center for Biotechnology Information), Bethesda, Maryland, USA: <http://www.ncbi.nlm.nih.gov/Genbank/>

DDBJ (DNA Data Bank of Japan) – National Institute of Genetics, Mishima, Japan: <http://www.ddbj.nig.ac.jp/>

Proteinové databáze:

SWISS-PROT – University of Geneva & Swiss Institute of Bioinformatics: <http://www.expasy.ch/sprot/> a <http://www.ebi.ac.uk/swissprot/>

PIR (Protein Information Resource) – NBRF (National Biomedical Research Foundation, Washington, D.C., USA) & Tokyo University & JIPID (Japanese International Protein Information Database, Tokyo) & MIPS (Martinsried Institute for Protein Sequences, Martinsried, Germany): <http://www-nbrf.georgetown.edu/>

PRF/SEQDB (Protein Resource Foundation) – Ósaka, Japan: <http://www.prf.or.jp/en/os.htm>

PDB (Protein Data Bank) – University of New Jersey, San Diego & Super-computer Center, University of California & National Institute of Standards and Technology: <http://www.rcsb.org/pdb/>

Formáty souborů:

FASTA:

>H_sapiens

```
ATGACCCCAATACGCAAAATTAACCCCTAATAAAATTAATTAACCACTCATTTCATCGACCTCCCCACCC  
CATCCAACATCTCCGCATGATGAACTTCGGCTCACTCCTTGGCGCCTGCCTGATCCTCAAATCACCAC  
AGGACTATTCCTAGCCATACTACTCACCAGACGCCTCAACCGCCTTTTCATCAATCGCCCACATCACT  
CGAGACGTAAATTATGGCTGAATCATCCGCTACCTTCACGCCAATGGCGCCTCAATATTCTTTATCTGCC  
TCTTCCTACACATCGGGCGAGGCCTATATTACGGATCATTTCTCTACTCAGAAACCTGAAACATCGGCAT
```

...

>P_troglod

```
ATGACCCCGACACGCAAAATTAACCCACTAATAAAATTAATTAATCACTCATTTTATCGACCTCCCCACCC  
CATCCAACATTTCCGCATGATGGA ACTTCGGCTCACTTCTCGGCGCCTGCCTAATCCTTCAAATTACCAC  
AGGATTATTCCTAGCTATACTACTCACCAGACGCCTCAACCGCCTTCTCGTCGATCGCCCACATCACC  
CGAGACGTAAACTATGGTTGGATCATCCGCTACCTCCACGCTAACGGCGCCTCAATATTTTTTTATCTGCC  
TCTTCCTACACATCGGCCGAGGTCTATATTACGGCTCATTTCTCTACCTAGAAACCTGAAACATTGGCAT
```

...

>P_paniscus

```
ATGACCCCAACACGCAAAATCAACCCACTAATAAAATTAATTAATCACTCATTTTATCGACCTCCCCACCC  
CATCCAATATTTCCACATGATGAACTTCGGCTCACTTCTCGGCGCCTGCCTAATCCTTCAAATCACCAC  
AGGACTATTCCTAGCTATACTACTCACCAGACGCCTCAACCGCCTTCTCATCGATCGCCCACATTACC  
CGAGACGTAAACTATGGTTGAATCATCCGCTACCTTCACGCTAACGGCGCCTCAATACTTTTCATCTGCC  
TCTTCCTACACGTCGGTCGAGGCCTATATTACGGCTCATTTCTCTACCTAGAAACCTGAAACATTGGCAT
```

...

Formáty souborů:

GenBank:

ORIGIN

```
1  tgaaatgaag atattctctt ctcaagacat caagaagaag gaactactcc ccaccaccag
61  cacccaaagc tggcattcta attaaactac ttcttgtgta cataaattta catagtacaa
121 tagtacatth atgtatatcg tacattaaac tattttcccc aagcatataa gcaagtacat
181 ttaatcaatg atataggcca taaaacaatt atcaacataa actgatacaa accatgaata
241 ttataactaat acatcaaatt aatgctttaa agacatatct gtgttatctg acatacacca
301 tacagtcata aactcttctc ttccatatga ctatcccctt ccccatthtg tctattaatc
361 taccatcctc cgtgaaacca acaaccgcgc caccaatgcc cctcttctcg ctccggggcc
421 attaaacttg ggggtagcta aactgaaact ttatcagaca tctgggttct acttcagggc
481 catcaaatgc gttatcgccc atacgttccc cttaaataag acatctcgat ggtatcgggt
541 ctaatcagcc catgaccaac ataactgtgg tgtcatgcat ttggtattht thttatthtg
601 cctactthca tcaacatagc cgtcaaggca tgaaaggaca gcacacagtc tagacgcacc
661 tacgggtgag aatcattagt ccgcaaaacc caatcaccta aggctaatta thcatgcttg
721 ttagacataa atgctactca ataccaaatt ttaactctcc aaacccccca acccctcct
781 cttaatgcca aacccccaaa aactaagaa cttgaaagac atatattht aactatcaaa
841 ccctatgtcc tgatcgattc tagtagttcc caaaatatga thcatattht agtacttgta
901 aaaatthtac aaaatcatgc tccgtgaacc aaaactctaa tcacactcta thacgcaata
961 aatattaaca agttaatgta gcttaataac aaagcaaagc actgaaaatg cttagatgga
1021 taatthtatc cca
```

//

Formáty souborů:

PHYLIP (“interleaved” format):

6 1120

H_sapiens	ATGACCCCAA	TACGCAAAAT	TAACCCCCTA	ATAAAATTAA	TTAACCACTC
P_troglod	ATGACCCCGA	CACGCAAAAT	TAACCCACTA	ATAAAATTAA	TTAATCACTC
P_paniscus	ATGACCCCAA	CACGCAAAAT	CAACCCACTA	ATAAAATTAA	TTAATCACTC
G_gorilla	ATGACCCCTA	TACGCAAAAC	TAACCCACTA	GCAAAACTAA	TTAACCACTC
P_pygmaeus	ATGACCCCAA	TACGCAAAAC	CAACCCACTA	ATAAAATTAA	TTAACCACTC
H_lar	ATGACCCCCC	TGCGCAAAAC	TAACCCACTA	ATAAAACTAA	TCAACCACTC
	ATTCATCGAC	CTCCCCACCC	CATCCAACAT	CTCCGCATGA	TGAAACTTCG
	ATTTATCGAC	CTCCCCACCC	CATCCAACAT	TTCCGCATGA	TGGAACTTCG
	ATTTATCGAC	CTCCCCACCC	CATCCAATAT	TTCCACATGA	TGAAACTTCG
	ATTCATTGAC	CTCCCTACCC	CGTCCAACAT	CTCCACATGA	TGAAACTTCG
	ACTCATCGAC	CTCCCCACCC	CATCAAACAT	CTCTGCATGA	TGGAACTTCG
	ACTTATCGAC	CTTCCAGCCC	CATCCAACAT	TTCTATATGA	TGAAACTTTG

Formáty souborů:

NEXUS (PAUP*, “interleaved”):

```
#NEXUS
begin data;
dimensions ntax=6 nchar=1120;
format datatype=DNA interleave datatype=DNA missing=? gap=-;
matrix
P_troglod   ATGACCCCGACACGCAAAATTAACCCACTAATAAAATTAATTAATCACTC
P_paniscus  ATGACCCCAACACGCAAAATCAACCCACTAATAAAATTAATTAATCACTC
H_sapiens   ATGACCCCAATACGCAAAATTAACCCCTAATAAAATTAATTAACCACTC
G_gorilla   ATGACCCCTATACGCAAAACTAACCCACTAGCAAAACTAATTAACCACTC
P_pygmaeus  ATGACCCCAATACGCAAAACCAACCCACTAATAAAATTAATTAACCACTC
H_lar       ATGACCCCCTGCGCAAAACTAACCCACTAATAAAACTAATCAACCACTC

P_troglod   ATTTATCGACCTCCCCACCCCATCCAACATTTCCGCATGATGGAACTTCG
P_paniscus  ATTTATCGACCTCCCCACCCCATCCAATATTTCCACATGATGAAACTTCG
H_sapiens   ATTCATCGACCTCCCCACCCCATCCAACATCTCCGCATGATGAAACTTCG
G_gorilla   ATTCATTGACCTCCCTACCCCGTCCAACATCTCCACATGATGAAACTTCG
P_pygmaeus  ACTCATCGACCTCCCCACCCCATCAAACATCTCTGCATGATGGAACTTCG
H_lar       ACTTATCGACCTTCCAGCCCATCCAACATTTCTATATGATGAAACTTTG

end;
```

Formáty souborů:

Clustal:

```
P_troglod ATGACCCCGACACGCAAAATTAACCCACTAATAAAATTAATTAATCACTCATTATCGAC
P_paniscus ATGACCCCAACACGCAAAATCAACCCACTAATAAAATTAATTAATCACTCATTATCGAC
H_sapiens ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTATCGAC
G_gorilla ATGACCCCTATACGCAAAACTAACCCACTAGCAAAACTAATTAACCACTCATTATCGAC
P_pygmaeus ATGACCCCAATACGCAAAACCAACCCACTAATAAAATTAATTAACCACTCACTCATCGAC
H_lar ATGACCCCCCTGCGCAAAACTAACCCACTAATAAAACTAATCAACCACTCACTTATCGAC
*****          *****          *****  ***          *****  *****  ** *****  * **  ***
```

```
P_troglod CTCCCCACCCCATCCAACATTTCCGCATGATGGAACTTCGGCTCACTTCTCGGCGCCTGC
P_paniscus CTCCCCACCCCATCCAATATTTCCACATGATGAAACTTCGGCTCACTTCTCGGCGCCTGC
H_sapiens CTCCCCACCCCATCCAACATCTCCGCATGATGAAACTTCGGCTCACTCCTTGGGCGCCTGC
G_gorilla CTCCCTACCCCGTCCAACATCTCCACATGATGAAACTTCGGCTCACTCCTTGGTGCCTGC
P_pygmaeus CTCCCCACCCCATCAAACATCTCTGCATGATGGAACTTCGGCTCACTTCTAGGCGCCTGC
H_lar CTTCCAGCCCCATCCAACATTTCTATATGATGAAACTTTGGTTCACTCCTAGGCGCCTGC
** **  *****  ** **  ** **          *****  *****  ** *****  ** **  *****
```

Formáty souborů:

FASTQ:

Line 1 begins with a '@' character and is followed by a sequence identifier and an *optional* description (like a [FASTA](#) title line).

Line 2 is the raw sequence letters.

Line 3 begins with a '+' character and is *optionally* followed by the same sequence identifier (and any description) again.

Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

Illumina sequence identifiers:

@HWUSI-EAS100R:6:73:941:1973#0/1

HWUSI-EAS100R	the unique instrument name
6	flowcell lane
73	tile number within the flowcell lane
941	'x'-coordinate of the cluster within the tile
1973	'y'-coordinate of the cluster within the tile
#0	index number for a multiplexed sample (0 for no indexing)
/1	the member of a pair, /1 or /2 (<i>paired-end or mate-pair reads only</i>)

BLAST (*Basic Local Alignment Search Tool*):

Nucleotide BLAST: nucleotide → nucleotide

blastx: nucleotide → protein

tblastn: protein → nucleotide

Protein BLAST: protein → protein

megablast

blastn

discontiguous megablast

BLAST

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

Using BLAST+ in Docker and on the cloud: [Webinar](#) on December 9, 2020.

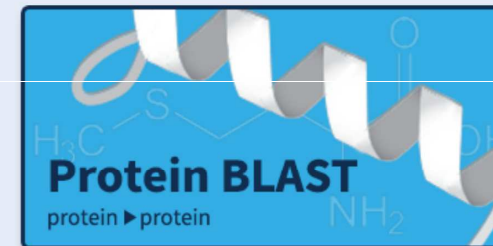
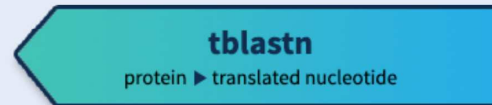
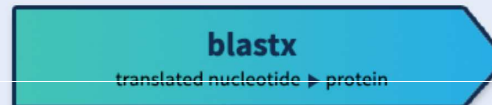
NEWS

In this webinar, the NCBI BLAST team will demonstrate containerized BLAST+ in Docker that is ready to use locally and in the cloud.

Wed, 02 Dec 2020 12:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)


[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
CAAAAACACTAAGAACCTTGAAGACATATACTATTAACATCTAACCCCTATGTCCTGATCAATTCTAGTAGTT
CAAAAAATATGACTTATATTTTAGTCTTGTAAAAATTTTGCAAAAAATGCCCCATAAGCCAAAACTCTAAT
TATACCCCTATACGCAATAAACAAATAGTAAGTAAATGTAGCTTAAATAAAAAGCAAAGCCTGAAAAATGCTTAG
ATGGATAATTTTATCCATAAACACAAAGGTTTGGTC
```

Clear Query subrange
From
To

New columns added to the Description Table
Click 'Select Columns' or 'Manage Columns'. 

Or, upload file

Soubor nevybrán

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Nucleotide collection (nr/nt)

Organism

exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude

Models (XM/XP) Uncultured/environmental sample sequences

Limit to

Sequences from type material

Entrez Query

[YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search

Program Selection

Optimize for

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

Choose a BLAST algorithm

BLAST

Search database **Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

Show results in a new window

[+ Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) [YouTube](#)

NCBI
National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA



[Policies and Guidelines](#) | [Contact](#)

[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▾

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

Job Title	Nucleotide Sequence
RID	WX41A1JA013 <small>Search expires on 12-09 03:27 am</small> Download All ▾
Program	BLASTN ? Citation ▾
Database	nt See details ▾
Query ID	lcl Query_50051
Description	None
Molecule type	dna
Query Length	950
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions

[Graphic Summary](#)

[Alignments](#)

[Taxonomy](#)

Sequences producing significant alignments

[Download](#) ▾

New [Select columns](#) ▾

Show

100 ▾

[?](#)

select all 100 sequences selected

[GenBank](#)

[Graphics](#)

[Distance tree of results](#)

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1375 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	878	1756	100%	0.0	100.00%	1062	EU106210.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1373 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	878	1751	100%	0.0	100.00%	1062	EU106208.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1372 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	874	1747	100%	0.0	99.79%	1062	EU106207.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1391 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1653	95%	0.0	99.79%	1015	EU106216.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1388 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1739	100%	0.0	99.79%	1062	EU106214.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1387 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1712	100%	0.0	99.79%	1062	EU106213.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1374 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1717	100%	0.0	99.79%	1062	EU106209.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1364 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1734	100%	0.0	99.79%	1062	EU106204.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1360 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1706	100%	0.0	99.79%	1062	EU106200.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1371 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1739	100%	0.0	99.79%	1062	EU106199.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1383 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1712	100%	0.0	99.79%	1062	EU106198.1
<input checked="" type="checkbox"/>	Mus cypriacus isolate MM1358 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gen...	Cypriot mouse	872	1706	100%	0.0	99.79%	1062	EU106195.1

100 sequences selected

[Download](#) [GenBank](#) [Graphics](#) Sort by: E value Next Previous Descriptions

Mus cypriacus isolate MM1375 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial

Sequence ID: [EU106210.1](#) Length: 1062 Number of Matches: 2

[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 475 [GenBank](#) [Graphics](#) Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
878 bits(475)	0.0	475/475(100%)	0/475(0%)	Plus/Plus
Query 1	TGTA AACCTGAAATGAAGATATTTCTTTCTCAAGACATCAAGAAGAAGGAAC TTATTCCC			60
Sbjct 1	TGTA AACCTGAAATGAAGATATTTCTTTCTCAAGACATCAAGAAGAAGGAAC TTATTCCC			60
Query 61	CACCACCAACACCCAAAGCTGGTATTCTAGTTAAACTACTTCTTGAGTACATAAATTTAC			120
Sbjct 61	CACCACCAACACCCAAAGCTGGTATTCTAGTTAAACTACTTCTTGAGTACATAAATTTAC			120
Query 121	ATAGTACATTAGTACATTTATGTATATCGTACATTA AATTATATTTCCCAAGCATATAAG			180
Sbjct 121	ATAGTACATTAGTACATTTATGTATATCGTACATTA AATTATATTTCCCAAGCATATAAG			180
Query 181	CACGTAAATTA AATTAATGACATAGCACATAAAAACGATATTTAACATAAAAATACTACACA			240
Sbjct 181	CACGTAAATTA AATTAATGACATAGCACATAAAAACGATATTTAACATAAAAATACTACACA			240
Query 241	ACATGAATATTATATTA AATACATTAAGTTAATGCTTTAAAGACATATCTGTGTTATCTG			300
Sbjct 241	ACATGAATATTATATTA AATACATTAAGTTAATGCTTTAAAGACATATCTGTGTTATCTG			300
Query 301	ACATACACCATAAAGTCATAAACCTTCTCTTCCATATGACTATCCCCTTCCCATTGG			360
Sbjct 301	ACATACACCATAAAGTCATAAACCTTCTCTTCCATATGACTATCCCCTTCCCATTGG			360
Query 361	TCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCACCTATGCCCTCTTCTCG			420
Sbjct 361	TCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCACCTATGCCCTCTTCTCG			420
Query 421	CTCCGGGCCCATTA AACTTGGGGGTAGCTAAACTGAACTTTATCAGACATCTGG			475
Sbjct 421	CTCCGGGCCCATTA AACTTGGGGGTAGCTAAACTGAACTTTATCAGACATCTGG			475

Range 2: 588 to 1062 [GenBank](#) [Graphics](#) Next Match Previous Match First Match

Sekvence conversion:

Complementary Sequence Conversion Tool

GenBank:

původně Los Alamos National Laboratory

90's → NCBI (*National Center for Biotechnology Information*)

20 000 přímých podání, 200 000 hromadných podání denně

každých 10 měsíců zdvojnásobení počtu

GenBank

ncbi.nlm.nih.gov/nuccore/?term=mus+cypriacus

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Create alert Advanced Help

COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>

Get the latest research information from NIH: <https://www.nih.gov/coronavirus>

Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Species Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

Animals (29) Customize ...

Molecule types genomic DNA/RNA (29) Customize ...

Source databases INSDC (GenBank) (29) Customize ...

Sequence Type Nucleotide (29)

Genetic compartments Mitochondrion (28)

Sequence length Custom range...

Release date Custom range...

Revision date Custom range...

Items: 1 to 20 of 29

<< First < Prev Page 1 of 2 Next > Last >>

- [Mus cypriacus mitochondrial partial cytb gene for cytochrome b](#)
1. 1,140 bp linear DNA
Accession: FR751074.1 GI: 323713991
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- [Mus cypriacus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial](#)
2. 1,062 bp linear DNA
Accession: EU106281.1 GI: 157266050
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Mus cypriacus isolate MM1377 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial](#)
3. 1,062 bp linear DNA
Accession: EU106280.1 GI: 157266049
[PubMed](#) [Taxonomy](#)

Find related data

Database: Select

Find items

Search details

"Mus cypriacus"[Organism] OR mus cypriacus[All Fields]

Search See more...

Recent activity

Turn Off Clear

- mus cypriacus (29) Nucleotide
- mus cypriacus control region (0)

Sem zadejte hledaný výraz

CES 20:09
CSQ 07.12.2020

Mus cypricus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial

GenBank: EU106281.1

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

Go to:

LOCUS EU106281 1062 bp DNA linear ROD 16-NOV-2007
DEFINITION Mus cypricus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial.

ACCESSION EU106281
VERSION EU106281.1
KEYWORDS .
SOURCE mitochondrion Mus cypricus (Cypriot mouse)

ORGANISM [Mus cypricus](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus.

REFERENCE 1 (bases 1 to 1062)
AUTHORS Macholan,M., Vyskocilova,M., Bonhomme,F., Krystufek,B., Orth,A. and Vohralik,V.

TITLE genetic variation and phylogeography of free-living mouse species (genus Mus) in the Balkans and the Middle East

JOURNAL Mol Ecol 16 (22), 4774-4788 (2007)
PUBMED [17908218](#)

REFERENCE 2 (bases 1 to 1062)
AUTHORS Macholan,M., Vyskocilova,M., Bonhomme,F., Krystufek,B., Orth,A. and Vohralik,V.

TITLE Direct Submission
JOURNAL Submitted (20-AUG-2007) Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, Acad. Sci. Czech Rep., Veveri 97, Brno CZ-60200, Czech Republic

FEATURES
Location/Qualifiers
source 1..1062
/organism="Mus cypricus"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/isolate="MM1381"
/db_xref="taxon:468371"
/country="Cyprus: Paramytha, 12 km N Limassol"
/note="type: CY17"
[tRNA](#) <1..37
/product="tRNA-Thr"
[tRNA](#) 38..105
/product="tRNA-Pro"
[D-loop](#) 106..982
gap 476..587
/estimated_length=112
[tRNA](#) 983..1049
/product="tRNA-Phe"

ORIGIN
1 tgtaaacctg aaatgaagat atctcttct caagacatca agaagaagga acttattccc
61 caccaccaac acccaagct ggtattctag ttaactact cttgagtac ataaatttac
121 atagtacatt agtacattta tgtatatcgt acattaaatt atattcccca agcatataag
181 cacgtaaat aaattaatga catagcacat aaaacgata ttaacataaa atactacaca
241 acatgaatat tatattaat acattaagtt aatgctttaa agacatctc gtgttatctg
301 acatacaca taaagtcata aacccttctc ttccatatga ctatcccctt ccccatcttg
361 tctattaatc taccatctc cgtgaaacca acaaccgcc cacctatgcc cctctctcgc
421 ctccgggcc attaaactg gggtagcta aactgaaact ttatcagaca tctgg
[gap 112 bp] [Expand Ns](#)
588 cat ttggtattt
601 tttatttgg tctactttca tcaacatagc cgtcaaggca tgaaggaca gcacacagtc
661 tagacgacc tacggtgaag aatcattagt ctcataacc caatcaccca aggctaatta
721 ttcagcttg ttagacataa aattattcaa taccagattt taactctcca aacccccccc
781 accccatcc tcttaatgcc aaacccaaa aacactaaga acttgaaga catatactat
841 taactatcta accctatgct ctgatcaatt ctgtagttc aaaaaatag acttatattt
901 tagttctgt aaaaatttg caaataatg ccccataagc caaactcta attataacct
961 attacgcaat aaacaatag agttaatgt agcttaataa aaagcaagc actgaaaatg
1021 cttagatgga taattttatc ccataaacac aaaggtttgg tc

//

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

PubMed

Taxonomy

PopSet

Recent activity

[Turn Off](#) [Clear](#)

Mus cypricus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-P Nucleotide

mus cypricus (29) Nucleotide

mus cypricus control region (0) Nucleotide

On the Probability of Fixation of Mutant Genes in a Population

The frequency of multiple paternity suggests that sperm competition is common in...

[See more...](#)

Mus cypriacus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial

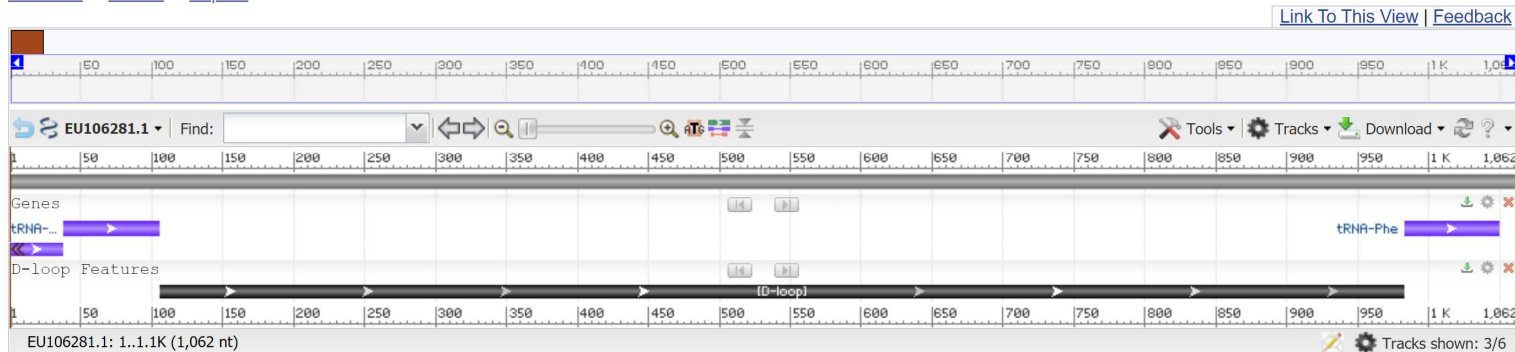
GenBank: EU106281.1

[GenBank](#) [Graphics](#) [PopSet](#)

```
>EU106281.1 Mus cypriacus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro
gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial
TGTAACCTGAAATGAAGATATTCTTCTCAAGACATCAAGAAGAAGGAAC TTATCCCCACCACCAAC
ACCCAAAGCTGGTATTCTAGTTAAACTACTTCTTGAGTACATAAATTTACATAGTACATTAGTACATTTA
TGTATATCGTACATTAATTATATTCCCCAAGCATATAAGCACGTAATAATAATTAATGACATAGCACAT
AAAACGATATTTAACATAAAAATACTACACAACATGAATATTATATTAATACATTAAGTTAATGCTTTAA
AGACATATCTGTGTTATCTGACATACACCATAAAGTCATAAACCCCTTCTTCCATATGACTATCCCTT
CCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCACCTATGCCCTCTTCTCG
CTCCGGGCCATTAAACTTGGGGTAGCTAAACTGAACTTTATCAGACATCTGGNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CGTCAAGGCATGAAAGGACAGCACAGCTAGACGCACCTACGGTGAAGAATCATTAGTCTCATAACC
CAATCACCCAAGGCTAATTATTCATGCTTGTAGACATAAAATATTCAATACCAGATTTAACTCTCCA
AACCCCCCACCCCATCCTCTTAATGCCAAACCCCAAAAACACTAAGAACTTGAAAGACATATACTAT
TAACTATCTAACCTATGCTCTGATCAATTCTAGTAGTTCAAAAAATAGACTTATATTTAGTCTTGT
AAAAATTTGCAAAATGATGCCCATAGCCAAACTCTAATTATACCCTATTACGCAATAAACAATAGT
AAGTTAATGTAGCTTAATAAAAAGCAAAGCACTGAAAATGCTTAGATGGATAATTTATCCATAAACAC
AAAGGTTTGGTC
```

Mus cypriacus isolate MM1381 tRNA-Thr gene, partial sequence; and tRNA-Pro gene, D-loop, and tRNA-Phe gene, complete sequence; mitochondrial

GenBank: EU106281.1

[GenBank](#) [FASTA](#) [PopSet](#)


Change region shown ▾

Customize view ▾

Analyze this sequence ▾

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information ▾

PubMed

Taxonomy

PopSet

Recent activity ▾

[Turn Off](#) [Clear](#)

Mus cypriacus isolate MM1381 tRNA-Thr

Analyze this sequence ▾

Run BLAST

Pick Primers

Related information ▾

PubMed

Taxonomy

PopSet

Recent activity ▾

[Turn Off](#) [Clear](#)

 Mus cypriacus isolate MM1381 tRNA-Thr gene, partial
sequence; and tRNA-Pro gene,... Nucleotide

 mus cypriacus (29) Nucleotide

mus cypriacus control region (0)

European Nucleotide Archive (ENA)

The **European Nucleotide Archive (ENA)** is a repository providing free and unrestricted access to annotated [DNA and RNA sequences](#). It also stores complementary information such as experimental procedures, details of [sequence assembly](#) and other [metadata](#) related to [sequencing projects](#).^[1] The archive is composed of three main databases: the [Sequence Read Archive](#), the Trace Archive and the EMBL Nucleotide Sequence Database (also known as EMBL-bank). (Wikipedia)

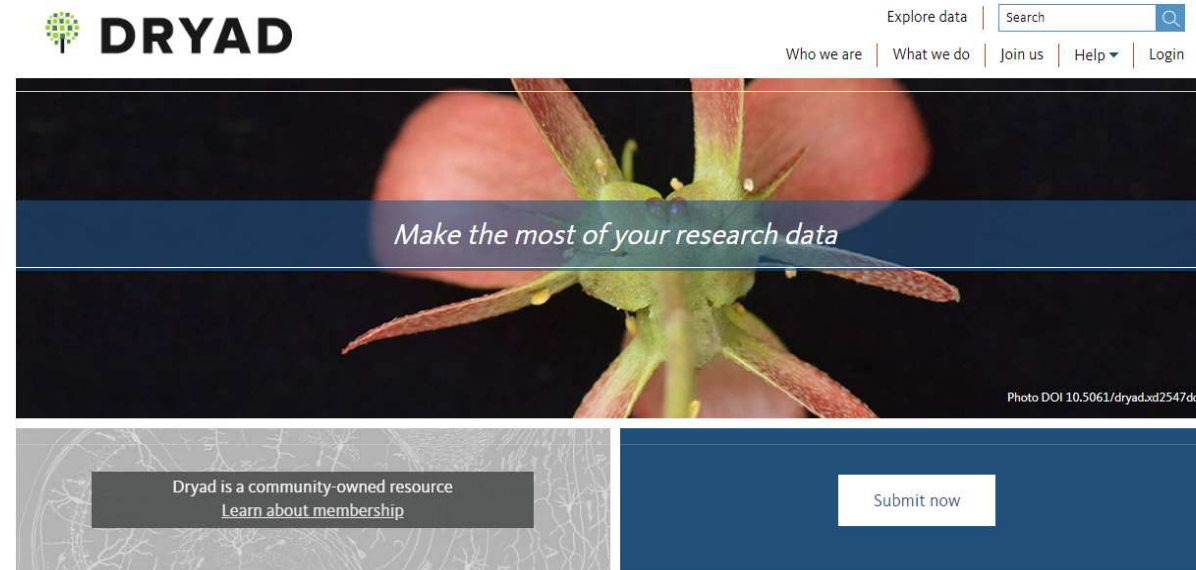
The screenshot shows the ENA website homepage. At the top, there is a teal header with the ENA logo (European Nucleotide Archive) and a search bar. The search bar has two input fields: "Enter text search terms" and "Enter accession". Below the search bar, there are navigation links: Home, Submit, Search, Rulespace, About, and Support. A yellow banner below the header contains the text: "The ENA Advanced Search API changed on 2023-05-02! Details here." Below this, a blue banner says: "We recommend that you subscribe to the ENA-announce mailing list for updates on services." A yellow banner below that provides information for SARS-CoV-2 data submissions, including contact details for virus-dataflow@ebi.ac.uk and a link to the Drag-and-Drop Data Submission Service. The main content area is titled "European Nucleotide Archive" and contains a paragraph describing the archive's purpose and access methods. Below this, there are four teal buttons: Submit, Search, Rulespace, and Support. To the right, there is a "Tweets from @ENASequences" section with a "Follow" button and a tweet from the European Nucleotide Archive (ENA) celebrating the 15th anniversary of the GWAS Catalog.

DRYAD

Dryad is an international [open-access repository](#) of research data, especially data underlying scientific and medical publications (mainly of evolutionary, genetic, and ecology biology). (Wikipedia)

od roku 2008

úložiště: github.com/CDL-Dryad/dryad



How it works



Login

Use your ORCID. If your institution is a [Dryad member](#), connect to your existing credentials.



Submit

Whether or not your data are related to an article, [upload](#) your data files and receive a citable DOI.



Review

Our [curators](#) will check through your submission to ensure the data are usable. They may contact you with advice or



Cite

[Cite](#) and promote your data publication!

Seřazení sekvencí:

Sekvence 1 TTGTACGACGG
 Sekvence 2 TTGTACGACG

```

TTGTACGACGG   TTGT---ACGACGG
|||||         |||   |||
TTGTACGACG    TTGTACGACG
    
```

$$GP = g + hl$$

g - gap penalty
h - gap extension penalty
l - gap length

Gap penalty:

Sekvence 1 ACTTGCTGCTTC
 Sekvence 2 ACGTGCTGCTC

Path 1
 ACTTG-TGCTTC
 ||| ||| |||
 ACGTGCTGCTC

Path 2
 ACTTGCTGCTTC
 || | ||| ||| |
 AC--GTGCTGCTC

g = penalizace za výskyt mezery (1×)

h = extenze za každou „pomlčku“

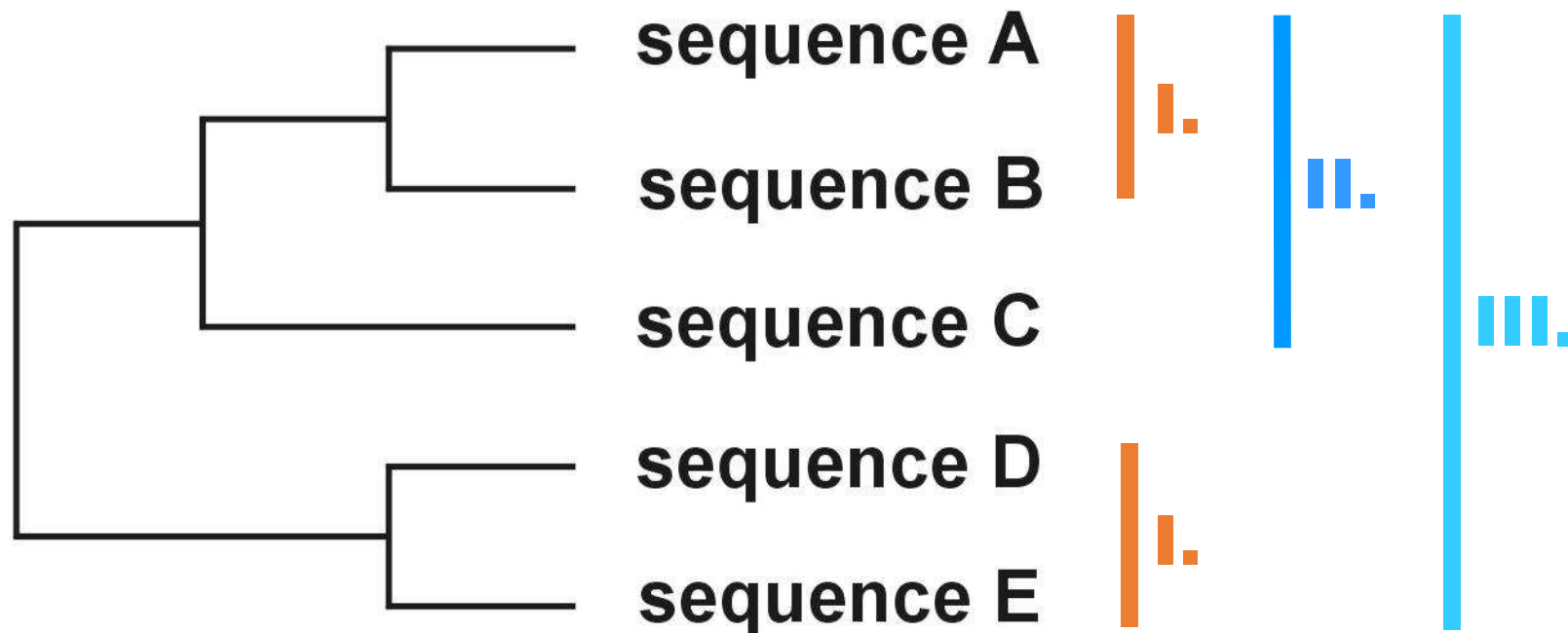
l = délka mezery (= počet „pomlček“)

Př.: GC-----TTAA
l = 5, *g* = *x*, *h* = 5*x*

Progresivní seřazení - ClustalX

3 phases:

1. Alignment of sequence pairs → pairwise distances
2. Construction of „guide tree“ (NJ)
3. Alignment of all sequences according to the guide tree



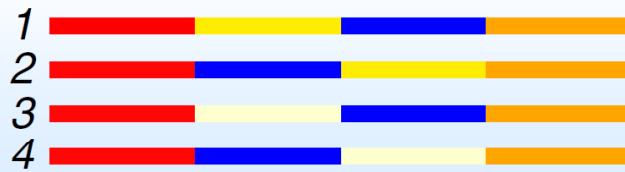
Další programy:

Clustal Omega – online

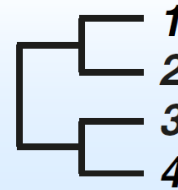
MUSCLE, MAFFT, Geneious

programy pro seřazení součástí mnoha populačněgenetických balíků

Align-free methods



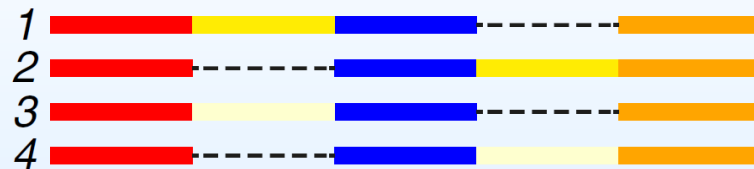
homologous sequence set



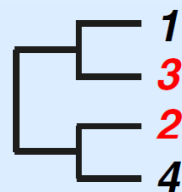
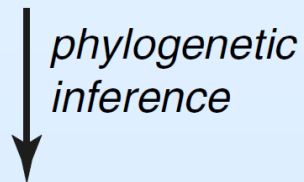
reference phylogeny

A

classical approach



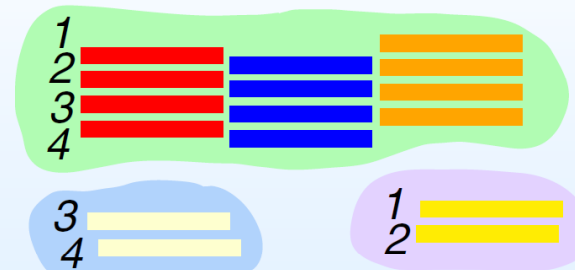
multiple sequence alignment



phylogenetic tree

B

alternative approach



alignment-free method

