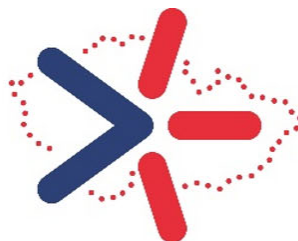




Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Analýza sekvenčních dat

Vratislav Peška – vpeska@ibp.cz

NPO-4 Pokročilé metody v genomice a
proteomice, NPO_MUNI_MSMT-16606/2022

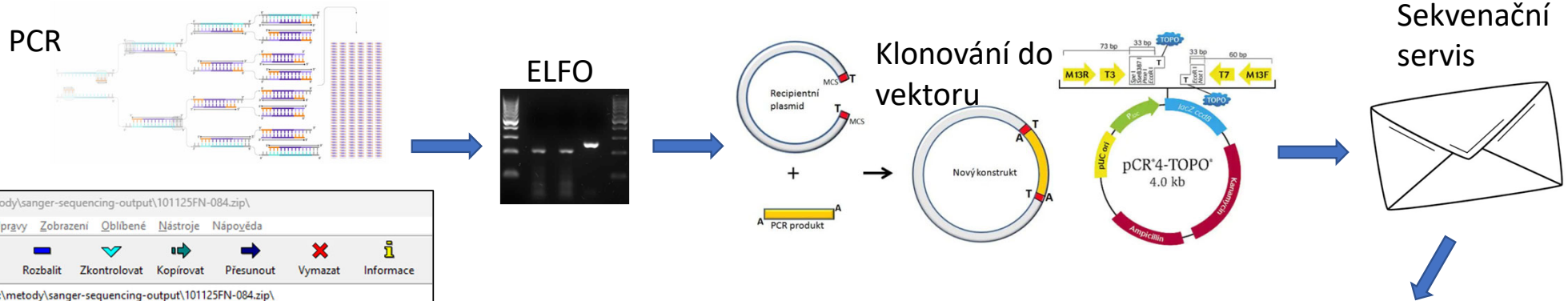


CEITEC

Central European Institute of Technology
BRNO | CZECH REPUBLIC

MUNI
SCI

Sangerova metoda sekvenace - výsledky

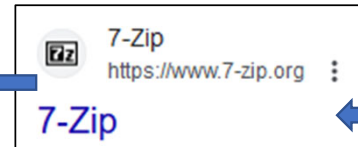


D:\metody\sanger-sequencing-output\101125FN-084.zip\

Soubor Úpravy Zobrazení Oblíbené Nástroje Nápořádá

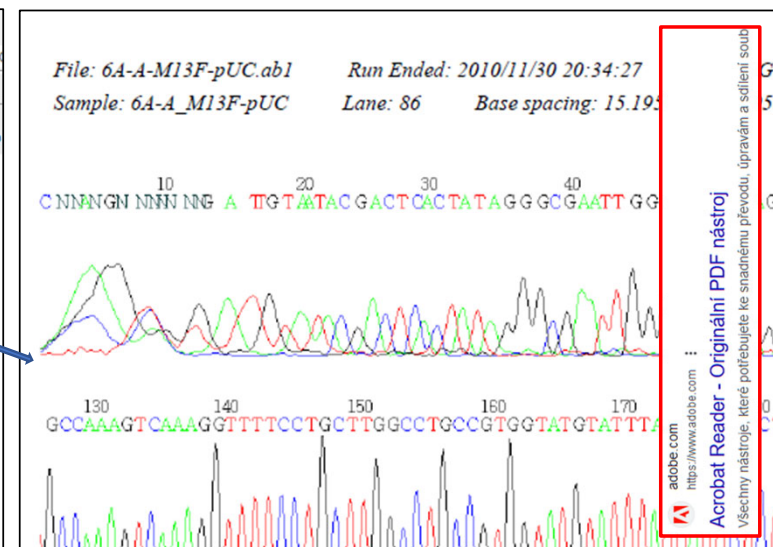
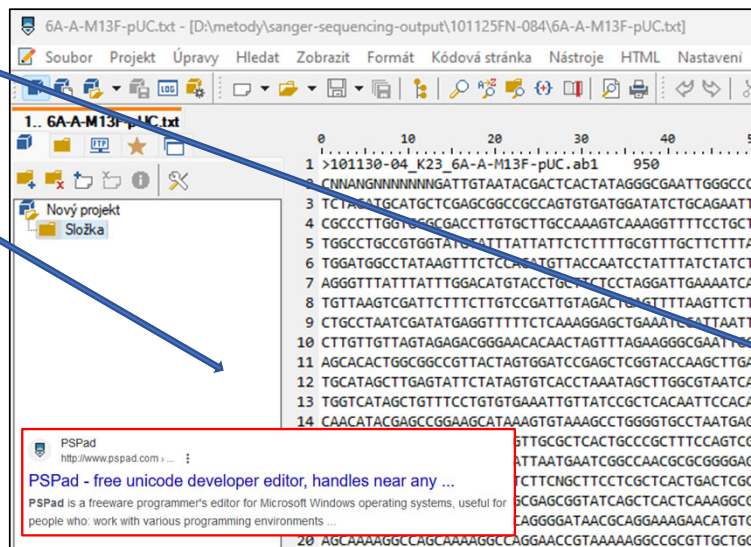
Přidat Rozbalit Zkontrolovat Kopírovat Přesunout Vymazat Informace

Název	Velikost	Komprimovan...	Změněn	Vytvoř
blast_result	362 475	78 618		
6A-A-M13F-pUC.ab1	280 084	144 420	2010-11-30 21:01	
6A-A-M13F-pUC.pdf	101 963	96 718	2010-11-30 21:05	
6A-A-M13F-pUC.phd.1	10 882	4 190	2010-11-30 21:01	
6A-A-M13F-pUC.scf	104 702	41 167	2010-11-30 21:01	
6A-A-M13F-pUC.txt	1 006	409	2010-11-30 21:05	
6B-B-M13F-pUC.ab1	280 132	152 381	2010-11-30 21:01	
6B-B-M13F-pUC.pdf	85 305	79 982	2010-11-30 21:06	
6B-B-M13F-pUC.phd.1	10 891	3 963	2010-11-30 21:01	
6B-B-M13F-pUC.scf	104 754	35 778	2010-11-30 21:01	
6B-B-M13F-pUC.txt	1 006	407	2010-11-30 21:06	
6B-B-M13R-pUC.ab1	280 372	147 458	2010-11-30 21:01	
6B-B-M13R-pUC.pdf	109 177	103 352	2010-11-30 21:06	
6B-B-M13R-pUC.phd.1	10 875	4 261	2010-11-30 21:01	
6B-B-M13R-pUC.scf	104 990	43 318	2010-11-30 21:01	
6B-B-M13R-pUC.txt	1 006	414	2010-11-30 21:06	
7A-1-M13F-pUC.ab1	265 665	136 828	2010-11-30 22:47	
7A-1-M13F-pUC.pdf	131 394	125 558	2010-11-30 22:47	
7A-1-M13F-pUC.phd.1	9 201	3 380	2010-11-30 22:47	
7A-1-M13F-pUC.scf	91 166	45 602	2010-11-30 22:47	
7A-1-M13F-pUC.txt	882	394	2010-11-30 22:47	
7B-5-M13F-pUC.ab1	279 740	144 380	2010-11-30 22:47	
7B-5-M13F-pUC.pdf	123 112	117 721	2010-11-30 22:48	
7B-5-M13F-pUC.phd.1	10 706	3 984	2010-11-30 22:47	
7B-5-M13F-pUC.scf	104 394	46 502	2010-11-30 22:47	
7B-5-M13F-pUC.txt	1 006	437	2010-11-30 22:48	

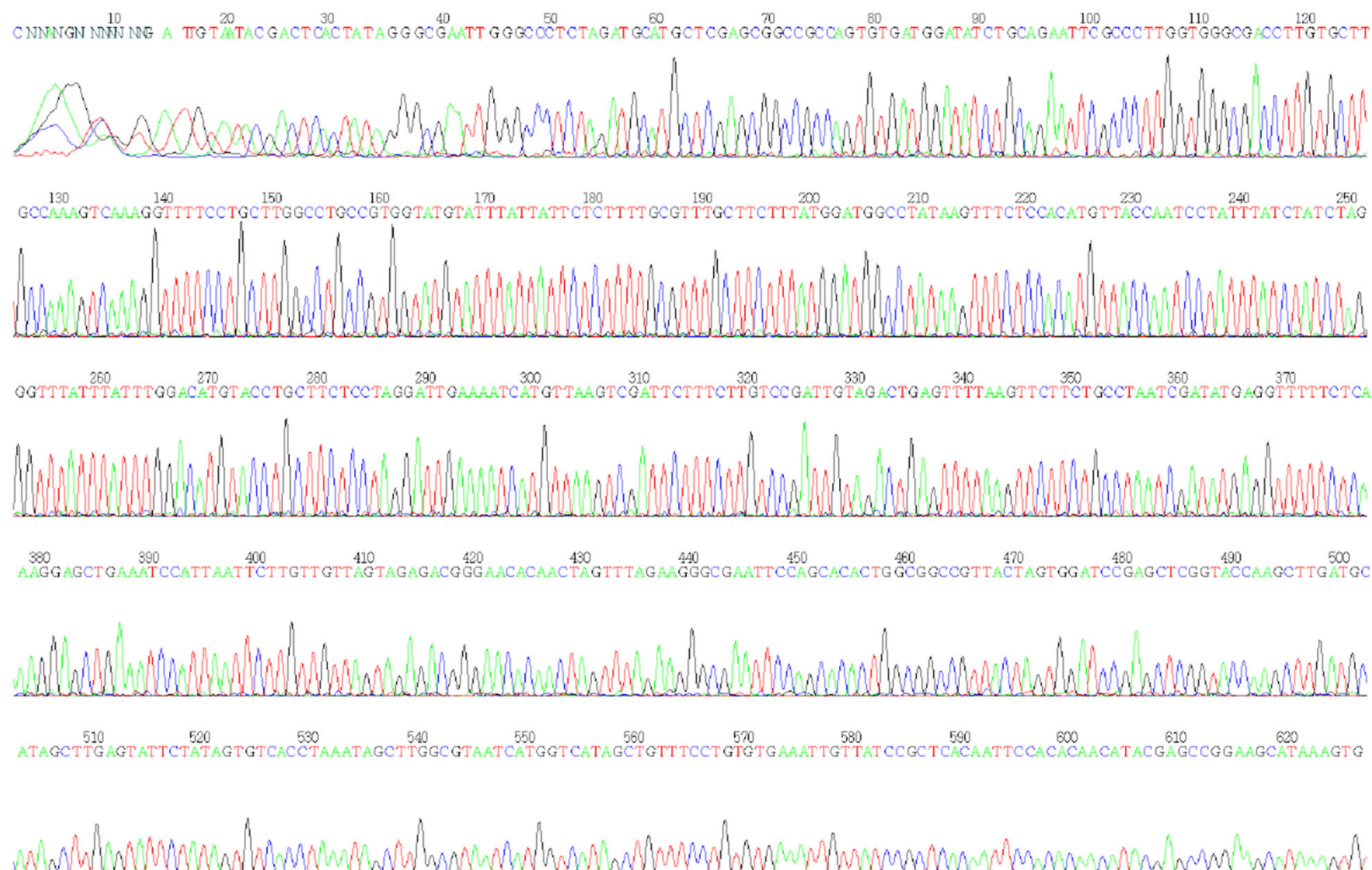


Tento počítač > DATA (D:) > metody > sanger-sequencing-output

Název	Datum změny	Typ	Velikost
101125FN-084.zip	02.12.2010 14:05	Soubor ZIP	4 177 kB



File: 6A-A-M13F-pUC.ab1 Run Ended: 2010/11/30 20:34:27 Signal G:169 A:232 C:259 T:281
Sample: 6A-A_M13F-pUC Lane: 86 Base spacing: 15.195725 950 bases in 11613 scans Page 1 of 2



Geneious
https://www.geneious.com : **\$900 / rok** **FREE** na Metacentrum

Geneious | Bioinformatics Software for Sequence Data Analysis

Geneious Prime is the world's leading bioinformatics software platform for molecular biology and sequence analysis. Enterprises involved in antibody ...

ACADEMIC GROUP	
ACTIVATIONS	PER YEAR
2	\$900
5	\$2,250
10	\$4,250
Custom tiers are available	
BUY >	

Start Your 30-day Geneious Prime Free Trial

No credit card is required.

EMAIL *

[Get Started](#)

By clicking the button I agree to Geneious' [Terms of Service](#) and [Privacy Policy](#).
Protected by reCAPTCHA (Google [Privacy Policy](#) and [Terms of Service](#) apply).

DNASTAR **\$2600 / rok**

> Supported

	Lasergene Molecular Biology	Lasergene Genomics	Lasergene Protein	DNASTAR Lasergene
Pricing Annual pricing starting at:	\$699/year See Pricing	\$1,599/year See Pricing	\$799/year See Pricing	\$2,599/year See Pricing

QIAGEN Solutions Products and Services Resources Support About Request a Trial Product Login Contact

Secondary Analysis

\$\$\$ / rok

QIAGEN CLC Workbench Premium REQUEST TRIAL	QIAGEN CLC Genome Finishing Module REQUEST TRIAL
QIAGEN CLC Genomics Workbench REQUEST TRIAL	QIAGEN CLC Microbial Genomics Module REQUEST TRIAL
QIAGEN CLC Main Workbench REQUEST TRIAL	QIAGEN CLC Genomics Server REQUEST DEMO
QCI Secondary Analysis REQUEST DEMO	FREE na Metacentrum
Interpretation and Visualization	
QIAGEN Ingenuity Pathway Analysis (IPA) REQUEST TRIAL	QIAGEN OmicSoft Suite REQUEST TRIAL

Availability

Geneious release R7 (version 7.1.5): module [gei](#)

- Geneious release R7 (version 7.0.3): module [g](#)
- a license purchased by CERIT-SC Centre allow

CLCbio Genomics Workbench version 9.5.3: mo

- upgrade of a license purchased by Centre CER

CLCbio Genomics Workbench version 7.0: mod

- upgrade of a license purchased by Centre CER

CLCbio Genomics Workbench version 6.5.1: mo

- a license purchased by Centre CERIT-SC (avai

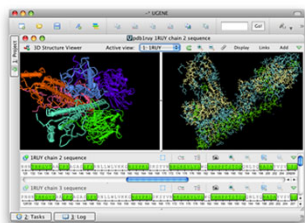
Unipro UGENE - Integrated Bioinformatics **FREE**

Home Documentation Learn Community Downloads Support Contacts

Unipro UGENE v. 46.0 February, 2023

UGENE is a free open-source bioinformatics software for Windows, macOS, and Linux.

[Download UGENE](#)



FREE

BioEdit

Biological sequence alignment editor for Win95/98/NT/2K/XP

Copyright © 1997-2017 Tom Hall

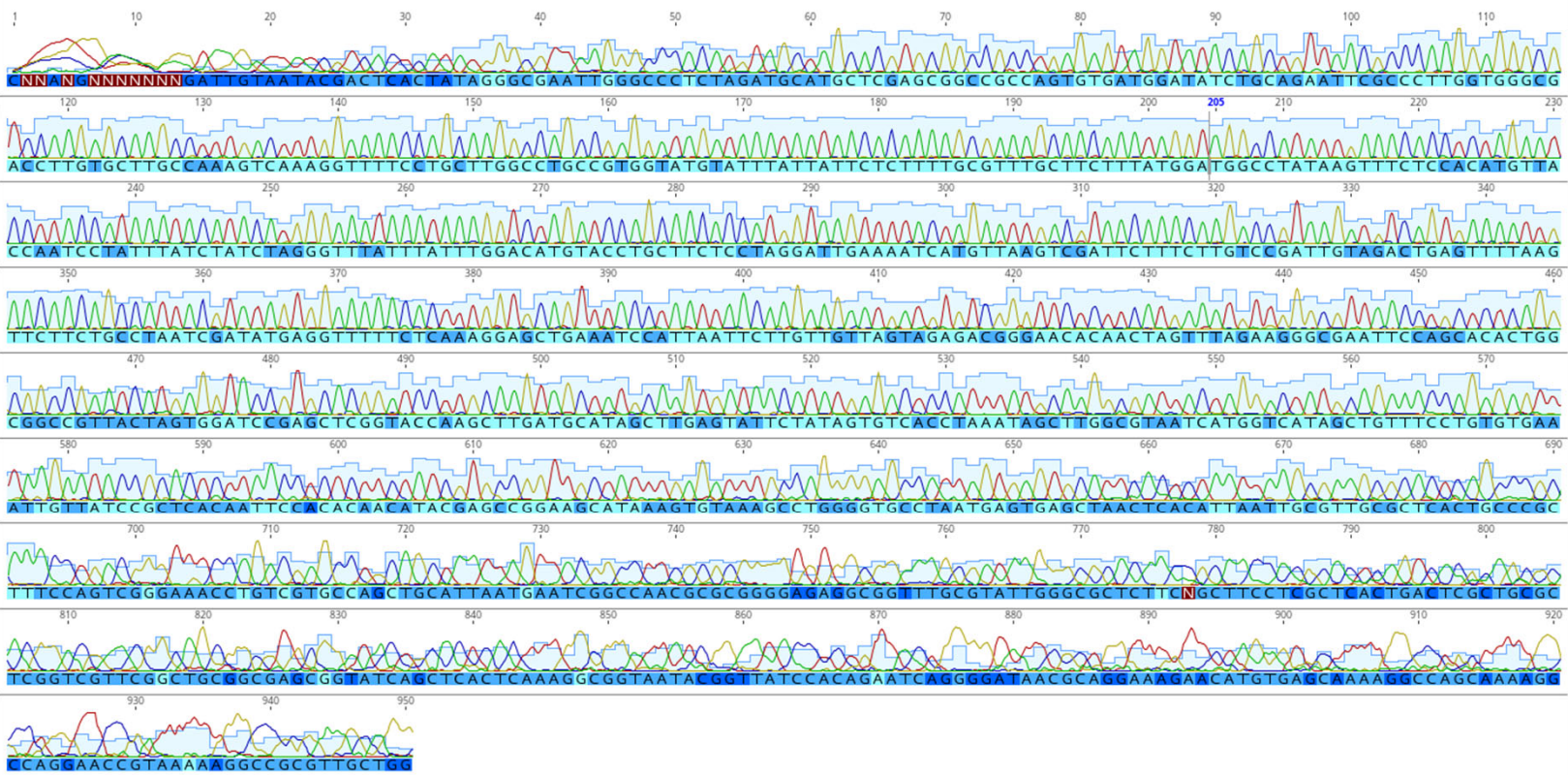
Back Forward Add Export BLAST Workflows Align/Assemble Tree Primers Cloning Help

Search Everywhere

Local 1			1 of 1 selected		
	Name	Description	Modified	Created	
☒	6A-A-M13F-pUC.ab1	-	30 Nov 2010 8:01 p...	30 Nov 2010 8:01 p...	No

Sequence View Dotplot (Self) Chromatogram Text View Lineage Info

Extract R.C. Translate Add/Edit Annotation Allow Editing Annotate & Predict Save



100% 40 100 50

Graphs

☒ Show Graphs

☒ Chromatogram

☒ A ☒ G ☒ C ☒ T

☒ Quality ☐ Highlight

☐ GC Content

☐ Show AT Graph

☐ Frame plot

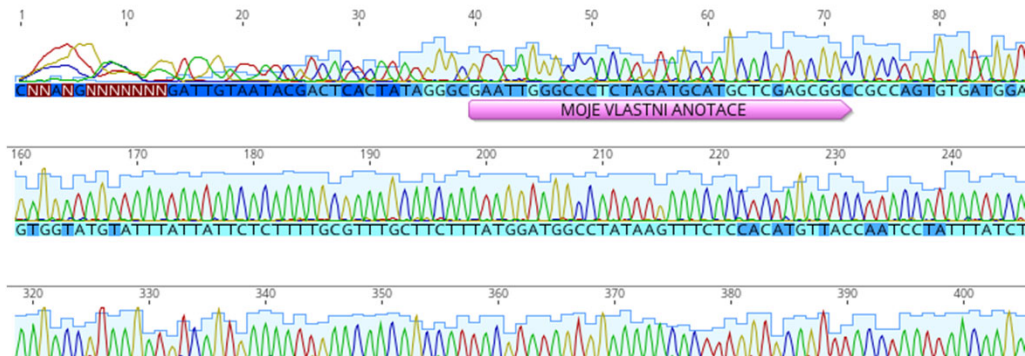
Sliding Window Size: 48

☐ Protein Coding Prediction

Window Size: 200

Step Size: 3

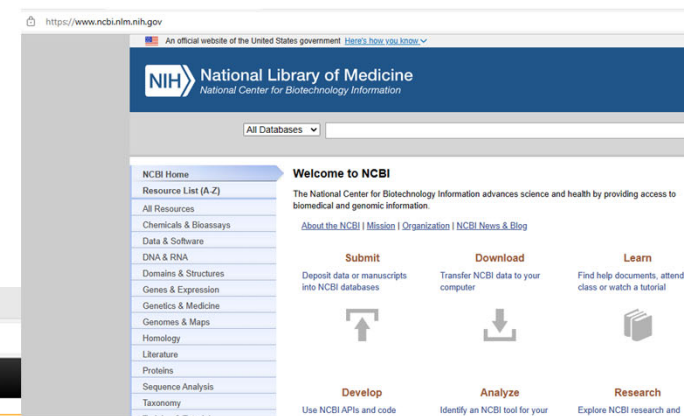
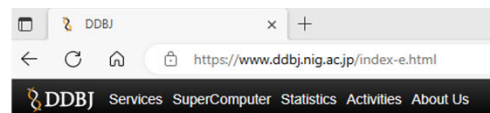
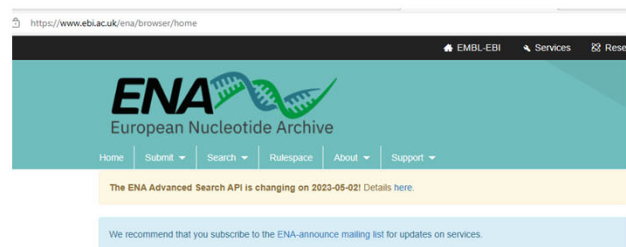
Based on the EMBOSS 6.5.7 tool tcode.



DATABÁZE - Velká trojice EMBL/NCBI/DDBJ



Data type	DDBJ	EMBL-EBI	NCBI
Next generation reads	Sequence Read Archive	European Nucleotide Archive (ENA)	Sequence Read Archive
Capillary reads	Trace Archive		Trace Archive
Annotated sequences	DDBJ		GenBank
Samples	BioSample		BioSample
Studies	BioProject		BioProject



(April 10th 10:00 - April 11th 14:00) Suspension of DDBJ		Filter this table
Tools	Description	
Amino Acid Explorer	Explores amino acid properties, substitutions and functions	
Assembly Archive	Links the raw sequence information found in the Trace Archive with assembly information found in GenBank/EMBL/DDBJ	
Basic Local Alignment Search Tool (BLAST)	Finds regions of local similarity between biological sequences	
Batch Entrez	Retrieves records specified in an uploaded file of identifiers	
BioAssay Services	Tools that summarize the biological test results in the PubChem database	
BLAST Link (BLINK)	Displays the results of a pre-computed BLAST search of a protein against all other protein sequences at NCBI	
BLAST Microbial Genomes	Finds regions of local similarity between query sequences and sequences from complete microbial genomes	
BLAST RefSeqGene	Finds regions of local similarity between query sequences and genomic sequences in the RefSeqGene/LRG set	
CDTree	Classifies protein sequences and investigates their evolutionary relationships	
CGV	Compare genomes based on assembly-assembly alignments	
Cn3D	Displays and manipulates 3-dimensional structures and alignments from the Structure database	
COBALT	Performs protein multiple sequence alignments	
Concise Microbial Protein BLAST	Finds regions of local similarity between query proteins and proteins from complete microbial (prokaryotic) genomes	
Conserved Domain Architecture Retrieval Tool (CDART)	Displays the functional domains that make up a given protein sequence	
Conserved Domain Search Service (CD Search)	Identifies the conserved domains present in a protein sequence	
Digital Differential Display (DDD)	Identifies genes with significantly different expression levels by comparing EST profiles	
Electronic PCR (e-PCR)	Identifies sequence tagged sites (STSs) within DNA sequences	
Frequency-weighted Link (FLink)	Links a group of records in a source database to a ranked list of associated records in a destination database based on frequency-weighted statistics	
Gene Expression Omnibus (GEO) BLAST	Finds regions of local similarity between query sequences and GenBank sequences included on microarray or SAGE platforms in the GEO database	
Genetic Codes	Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree	
Genome BLAST	Finds regions of local similarity between query sequences and genome sequences	

Featured data resources



AlphaFold DB

Database for protein structure predictions for numerous species

CC-BY



BioModels

A repository of peer-reviewed, published, computational models.

Web API | CC0



ChEMBL

An open data resource of binding, functional and ADMET bioactivity data.

Web API | CC-BY



Featured tools



Clustal Omega

Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

Web API



HMMER

Fast sensitive protein homology searches using profile hidden Markov models (HMMs) for querying against both sequence and HMM target databases.

Web API



Annotation Platform

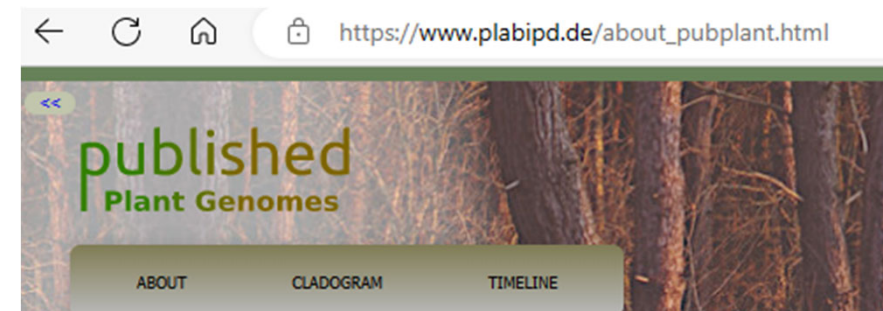
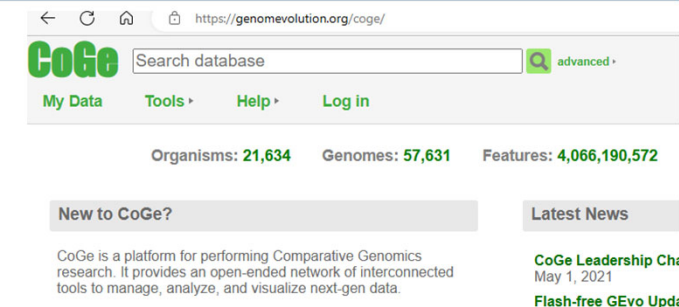
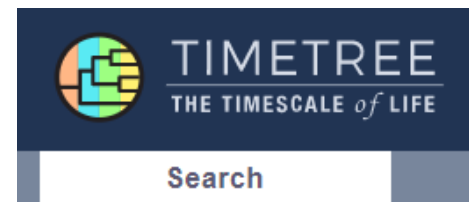
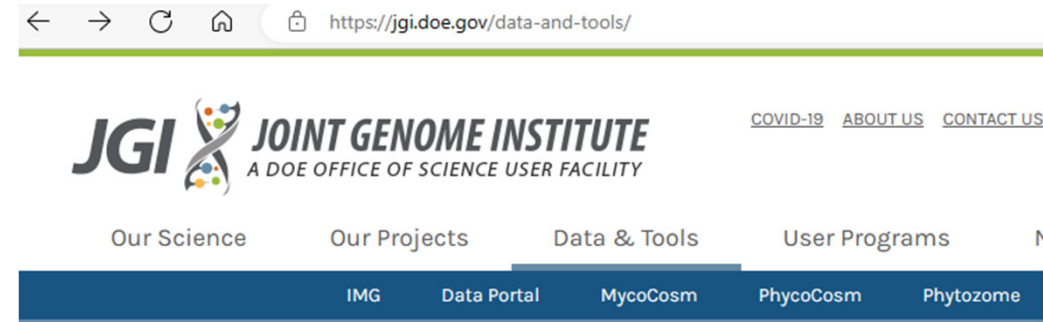
Consolidating text-mined and curated annotations

Web API

NCBI databáze

Assembly
 BioCollections
 BioProject (formerly Genome Project)
 BioSample
 Bookshelf
 ClinVar
 ClinicalTrials.gov
 Computational Resources from NCBI's
 Structure Group
 Consensus CDS (CCDS)
 Conserved Domain Database (CDD)
 Database of Genomic Structural Variation
 (dbVar)
 Database of Genotypes and Phenotypes
 (dbGaP)
 Database of Short Genetic Variations (dbSNP)
 GenBank
 Gene
 Gene Expression Omnibus (GEO) Database
 Gene Expression Omnibus (GEO) Datasets
 Gene Expression Omnibus (GEO) Profiles
 GeneReviews
 Genes and Disease
 Genetic Testing Registry (GTR)
 Genome
 Genome Reference Consortium (GRC)
 Glycans
 HIV-1, Human Protein Interaction Database
 Identical Protein Groups
 Influenza Virus
 Journals in NCBI Databases

MeSH Database
 MedGen
 NCBI C++ Toolkit Manual
 NCBI Education Page
 NCBI Glossary
 NCBI Handbook
 NCBI Help Manual
 NCBI Pathogen Detection Project
 National Library of Medicine (NLM) Catalog
 Nucleotide Database
 Online Mendelian Inheritance in Man (OMIM)
 PopSet
 Protein Clusters
 Protein Database
 Protein Family Models
 PubChem BioAssay
 PubChem Compound
 PubChem Substance
 PubMed
 PubMed Central (PMC)
 RefSeqGene
 Reference Sequence (RefSeq)
 Retrovirus Resources
 SARS CoV
 Sequence Read Archive (SRA)
 Structure (Molecular Modeling Database)
 Taxonomy
 Third Party Annotation (TPA) Database
 Trace Archive
 Viral Genomes
 Virus Variation



← ↻ 🏠 🔒 <https://www.genome.jp/kegg/>

KEGG Databases Tools Auto annotation Kanehisa Lab



KEGG Search Help

» Japanese

KEGG Home Release notes

KEGG: Kyoto Encyclopedia of Genes and Genomes

← ↻ 🏠 ⚠️ Nezabezpečeno | geneontology.org

GENE ONTOLOGY Unifying Biology

About Ontology Annotations Downloads Help

ONTOLOGY ANNOTATION GO-CAM TOOLS &

← ↻ 🏠 🔒 <https://www.expasy.org/resources/uniprotkb-swiss-prot>

e.g. [BLAST](#), [UniProt](#), [MSH6](#), [Albumin](#)...

UniProtKB/Swiss-Prot

← ↻ 🏠 ⚠️ Nezabezpečeno | flybase.org

GTGGGCAATCGTAAGATAGCCAAATATTATTATTGTTTCAGATACTCAC
AGGCGACGACGTCAGATGCTTTTGGAGTCTTTTGAATCAGTGAAATT
AATATGAGGCGGCAAGCGACTTTTGGGTTGATGATGATGATGATGATGAT
AATAATAAAAACAACACAGTGCACACACAGCCGGGCATCTTCATAGA

FlyBase

Home Tools Downloads Links Community

Proteins & Proteomes Database

← ↻ 🏠 🔒 <https://wormbase.org/#012-34-5>

WormBase

Version: WS287 | ALLIANCE of GENOME RESOURCES

About WormBase Home Downloads Community Support

← 🔒 <https://www.arabidopsis.org>

tair Home Help Contact About Us Subscribe Login Register

Search Browse Tools Portals Download Submit

The Arabidopsis Information Resource

NEW ARABIDOPSIS COMMUNITY Group

Free to Join.
Open to All.



Jin CY Liao

Supported by NAASC

← 🔒 <https://www.girinst.org/repbase/>

giri REPBASE

Subscription News!

Browse Search Repeat Masking Download

OXFORD ACADEMIC Journals

You are here: [NAR Journal Home](#) » [Database Summary Paper Categories](#)

NAR Database Summary Paper Category List

Nucleotide Sequence Databases
 RNA sequence databases
 Protein sequence databases
 Structure Databases
 Genomics Databases (non-vertebrate)
 Metabolic and Signaling Pathways
 Human and other Vertebrate Genomes
 Human Genes and Diseases
 Microarray Data and other Gene Expression Databases
 Proteomics Resources
 Other Molecular Biology Databases
 Organelle databases
 Plant databases
 Immunological databases
 Cell biology

► [Compilation Paper](#)
 ► [Category List](#)
 ► [Alphabetical List](#)
 ► [Category/Paper List](#)
 ► [Search Summary Papers](#)

Oxford University Press is not responsible for the content of external internet sites

OXFORD ACADEMIC Journals Books

DATABASE

The Journal of Biological Databases and Curation

Volumes Submit Alerts About

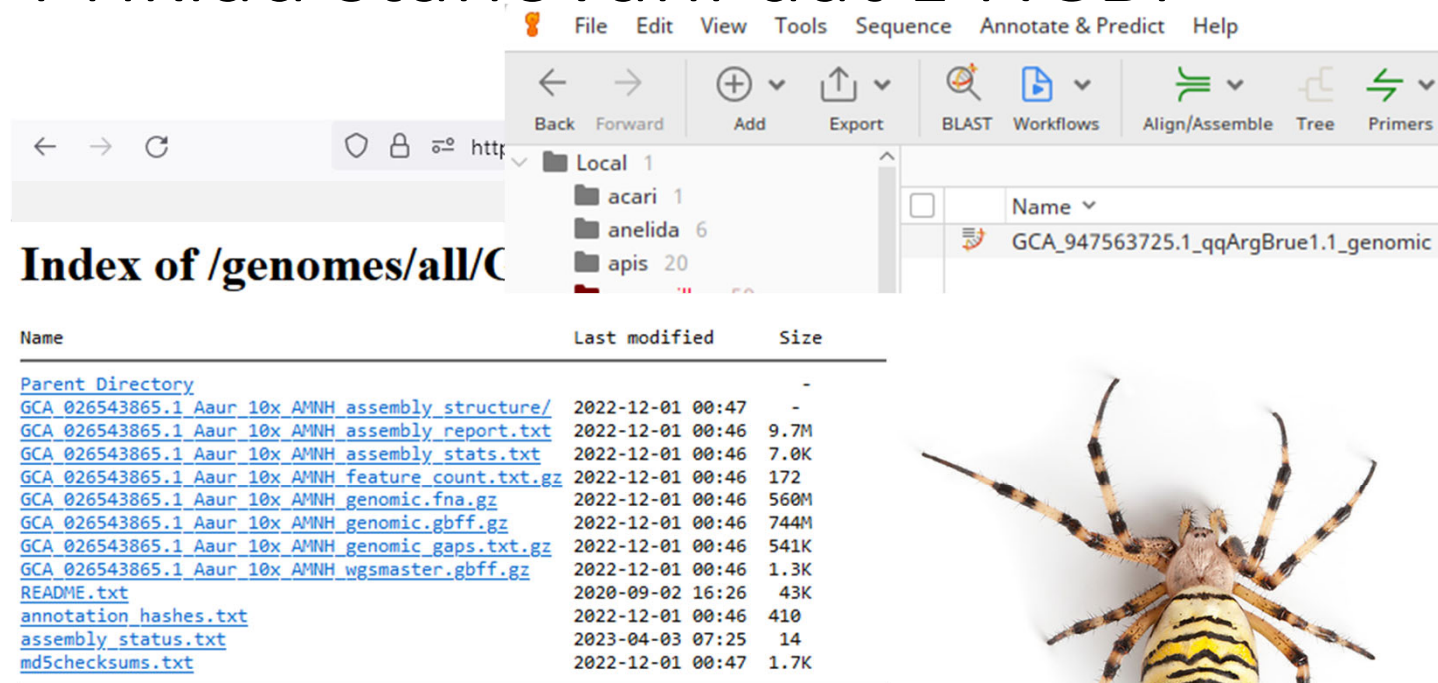


Latest Volume
 Volume 2023
 2023

Impact Factor
4.462
 5 year Impact Factor
4.776

Editor-in-Chief
 David Landsman

Příklad stahování dat z NCBI



Name	Last modified	Size
Parent Directory		-
GCA_026543865.1_Aaur_10x_AMNH_assembly_structure/	2022-12-01 00:47	-
GCA_026543865.1_Aaur_10x_AMNH_assembly_report.txt	2022-12-01 00:46	9.7M
GCA_026543865.1_Aaur_10x_AMNH_assembly_stats.txt	2022-12-01 00:46	7.0K
GCA_026543865.1_Aaur_10x_AMNH_feature_count.txt.gz	2022-12-01 00:46	172
GCA_026543865.1_Aaur_10x_AMNH_genomic.fna.gz	2022-12-01 00:46	560M
GCA_026543865.1_Aaur_10x_AMNH_genomic.gbff.gz	2022-12-01 00:46	744M
GCA_026543865.1_Aaur_10x_AMNH_genomic_gaps.txt.gz	2022-12-01 00:46	541K
GCA_026543865.1_Aaur_10x_AMNH_wgsmaster.gbff.gz	2022-12-01 00:46	1.3K
README.txt	2020-09-02 16:26	43K
annotation_hashes.txt	2022-12-01 00:46	410
assembly_status.txt	2023-04-03 07:25	14
md5checksums.txt	2022-12-01 00:47	1.7K

[HHS Vulnerability Disclosure](#)





National Library of Medicine

National Center for Biotechnology Information

All Databases 

homo sapiens

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts
into NCBI databases



Download

Transfer NCBI data to your
computer



Learn

Find help documents, attend a
class or watch a tutorial



Develop

Use NCBI APIs and code
libraries to build applications



Analyze

Identify an NCBI tool for your
data analysis task



Research

Explore NCBI research and
collaborative projects





Search NCBI

homo sapiens



Search

Results found in 32 databases

TAXONOMY

Was this helpful?



Homo sapiens

Human (*Homo sapiens*) is a species of primate in the family *Hominidae* (great apes).

Taxonomy ID: 9606



Genomes

Browse and download



Genes

Browse and download



Genome Data Viewer

Browse the reference genome



BLAST

Search the reference sequence

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Homo sapiens (human)  Enter one or more taxonomic names

 Filters

Download 

Select columns

1 082 genomes

Rows per page

20 

1-20 of 1 082



☐ Assembly	GenBank	RefSeq	Scientific name 	Modifier	Annotati...	Size (	Action
☐ GRCh38.p14 	GCA_000001405.29	GCF_000001405.40	Homo sapiens (human)		NCBI RefS...	3,09	
☐ HuRef	GCA_000002125.2	GCF_000002125.1	Homo sapiens (human)	male (sex)		2,84	
☐ CHM1_1.1	GCA_000306695.2	GCF_000306695.2	Homo sapiens (human)	CHM1 (isolate)	NCBI RefS...	3,09	
☐ T2T-CHM13v2.0	GCA_009914755.4	GCF_009914755.1	Homo sapiens (human)		NCBI RefS...	3,11	
☐ WGS 	GCA_000002115.2		Homo sapiens (human)			2,86	

Search NCBI ...

Datasets Taxonomy **Genome** Gen



Genom

reference

Download

datasets

curl

NCBI RefSeq sequence GCF_000

Submitted GenBank sequence GCA_000

Taxon *Homo s*

Synonym hg38

Assembly type haploid-v

Submitter Genome

Date Feb 3, 20

View the [legacy Assembly page](#)

Assembly statistics

Download Package

1 Genome available for download

Select the files you want

Select file source

- ☒ All (2)
- ☐ RefSeq only (1)
- ☐ GenBank only (1)

Select file types

- ☒ Genome sequences (FASTA)
- ☐ Annotation features (GTF)
- ☒ Annotation features (GFF)
- ☐ Sequence and annotation (GBFF)
- ☒ Transcripts (FASTA)
- ☒ Genomic coding sequences (FASTA)
- ☒ Protein (FASTA)
- ☐ Sequence report (JSONL)
- ☒ Assembly data report (JSONL)

Your selected data will be downloaded as a ZIP archive

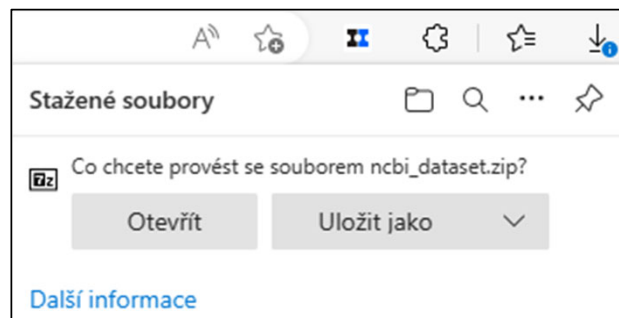
Estimated file size is 2 GB

Name your file

ncbi_dataset.zip

Cancel

Download



ncbi_dataset.zip

ncbi_dataset

ncbi_dataset

README.md

» Tento počítač » DATA (D:) » metody » ncbi_dataset » ncbi_dataset » data » GCF_000001405.40

Název	Datum změny	Typ	Velikost
cds_from_genomic.fna	06.04.2023 15:20	Soubor FNA	357 020 kB
GCF_000001405.40_GRCh38.p14_genomic.fna	06.04.2023 15:20	Soubor FNA	3 261 464 kB
genomic.gff	06.04.2023 15:20	Soubor GFF	1 480 761 kB
protein.faa	06.04.2023 15:20	Soubor FAA	103 388 kB
rna.fna	06.04.2023 15:20	Soubor FNA	725 186 kB

Drop files here to import

Ambiguous Sequences

What type of sequences are in GCF_000001405.29_GRCh38.p14_genomic.fna? Geneious thinks this data is probably nucleotide sequences.

☒ Nucleotide sequences

☐ Protein sequences

☐ Apply to all ambiguous sequences

Reset to Defaults

OK Cancel

36% - Importing Files

Analyzing file contents

Cancel

Grouping Sequences

How do you want to store the sequences from GCF_000001405.29_GRCh38.p14_genomic.fna in Geneious Prime?

This file contains a large number of sequences so creating a list is highly recommended. Large numbers of separately stored sequences may cause Geneious to run slowly and use more memory.

Sequences can be grouped into or extracted from lists at any time using the Sequence menu.

☐ Remember my preference

Keep sequences separate

Create sequence list

← Back
→ Forward
⊕ Add
⬆ Export
🔍 BLAST
📄 Workflows
≡ Align/Assemble
🌳 Tree
↔ Primers
🔄 Cloning
❓ Help
🔍 Search Everywhere

Local 1

- acari 1
- anelida 6
- apis 20
- aspergillus 59
- basidiomycota 569
- chordata 34
- from PETR 4698 (4 unread)
- heteroptera 1
- human 1**
- leptopilina 1
- Pisum_sativum 7
- pyrApt 14 (1 unread)
- ranunculaceae 5
- Results_extended 378 (4 u
- spiders 94
- spodoptera 12
- TRB1-proteins 57
- yeasts 232
- Sample Documents 311 (1
- Reference Features 941

0 of 1 se

☐	Name	Description
	GCA_000001405.29_GRCh38.p14_genomic	-

☒	Name	Description
☒	GCA_000001405.29_GRCh38.p14_genomic	-

Sequence View
Lengths Graph
Text View
Lineage
Info

← → ↺ Extract R.C. Translate ↗ Add/Edit Annotation ✎ Allow Editing ↗ Annotate & Predict ▾ 💾 Save



All Sequences ▾
0 %

Annotations and Tracks

Filter

This sequence has no c

Sequence View | Lengths Graph | Text View

← → ↺ Extract R.C. Translate ↗ A

1 20 Mbp 40 Mbp

CM000663.2 (Homo sapiens chromosome 1,...
KI270706.1 (Homo sapiens chromosome 1 u...
KI270707.1 (Homo sapiens chromosome 1 u...
KI270708.1 (Homo sapiens chromosome 1 u...
KI270709.1 (Homo sapiens chromosome 1 u...
KI270710.1 (Homo sapiens chromosome 1 u...
KI270711.1 (Homo sapiens chromosome 1 u...
KI270712.1 (Homo sapiens chromosome 1 u...
KI270713.1 (Homo sapiens chromosome 1 u...
KI270714.1 (Homo sapiens chromosome 1 u...
CM000664.2 (Homo sapiens chromosome 2,...
KI270715.1 (Homo sapiens chromosome 2 u...
KI270716.1 (Homo sapiens chromosome 2 u...
CM000665.2 (Homo sapiens chromosome 3,...
GL000221.1 (Homo sapiens chromosome 3 ...
CM000666.2 (Homo sapiens chromosome 4,...
GL000008.2 (Homo sapiens chromosome 4 ...
CM000667.2 (Homo sapiens chromosome 5,...
GL000222.1 (Homo sapiens chromosome 5 ...

GCF_000001405.40

Nový ✂️ 📄 📄 📄 📄 🗑️ ↕️ Seřadit ▾ ☰ Zobrazit ▾ ⋮

← → ↺ ↻

Domů

Vratislav – MU

Disk Google ➦

ownCloud ➦

Plocha ➦

Stažené sou ➦

Dokumenty ➦

Obrázky ➦

Název

Datum změny

cds_from_genomic.fna 06.04.2023 15:20

GCF_000001405.40_GRCh38.p14_genomic.fna 06.04.2023 15:20

genomic.gff 06.04.2023 15:20

protein.faa 06.04.2023 15:20

rna.fr

1.19% - Importing Files

Reading GFF data

Cancel

Download

 datasets

curl

Actions

Download RefSeq

BLAST against this genome

See in Genome Data Viewer

See more files on FTP

BioProject

PRJNA31257

The Human Genome

maintained by the

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Genome Reference Consortium

Index of /genomes/all/GCF/000/001/405/GCF_000001405.40_GRCh38.p14

Name	Last modified	Size
Parent Directory		-
Annotation_comparison/	2023-03-21 10:15	-
GCF_000001405.40_GRCh38.p14_assembly_structure/	2023-03-30 02:36	-
GRCh38_major_release_seqs_for_alignment_pipelines/	2021-03-02 22:14	-
RefSeq_transcripts_alignments/	2023-03-21 10:15	-
GCF_000001405.40-RS_2023_03_annotation_report.xml	2023-03-21 10:15	280K
GCF_000001405.40_GRCh38.p14_assembly_regions.txt	2023-03-21 10:15	44K
GCF_000001405.40_GRCh38.p14_assembly_report.txt	2023-03-21 10:15	77K
GCF_000001405.40_GRCh38.p14_assembly_stats.txt	2023-03-21 10:15	90K
GCF_000001405.40_GRCh38.p14_cds_from_genomic.fna.gz	2023-03-21 10:15	27M
GCF_000001405.40_GRCh38.p14_feature_count.txt.gz	2023-03-21 10:15	1.5K
GCF_000001405.40_GRCh38.p14_feature_table.txt.gz	2023-03-21 10:15	8.6M
GCF_000001405.40_GRCh38.p14_genomic.fna.gz	2023-03-21 10:15	928M
GCF_000001405.40_GRCh38.p14_genomic.gbff.gz	2023-03-21 10:15	1.2G
GCF_000001405.40_GRCh38.p14_genomic.gff.gz	2023-03-21 10:15	67M
GCF_000001405.40_GRCh38.p14_genomic.gtf.gz	2023-03-21 10:15	49M
GCF_000001405.40_GRCh38.p14_genomic.gaps.txt.gz	2023-03-21 10:15	11K
GCF_000001405.40_GRCh38.p14_protein.faa.gz	2023-03-21 10:15	27M
GCF_000001405.40_GRCh38.p14_protein.gpff.gz	2023-03-21 10:15	158M
GCF_000001405.40_GRCh38.p14_pseudo_without_product.fna.gz	2023-03-21 10:15	8.9M
GCF_000001405.40_GRCh38.p14_rm.out.gz	2023-03-21 10:15	183M
GCF_000001405.40_GRCh38.p14_rm.run	2023-03-21 10:15	874
GCF_000001405.40_GRCh38.p14_rna.fna.gz	2023-03-21 10:15	129M
GCF_000001405.40_GRCh38.p14_rna.gbff.gz	2023-03-21 10:15	461M
GCF_000001405.40_GRCh38.p14_rna_from_genomic.fna.gz	2023-03-21 10:15	94M
GCF_000001405.40_GRCh38.p14_translated_cds.faa.gz	2023-03-21 10:15	17M
README_GCF_000001405.40-RS_2023_03	2023-03-21 10:15	2.1K
README_patch_release.txt	2023-03-21 10:15	1.5K
all_alt_scaffold_placement.txt	2023-03-21 10:15	50K
annotation_hashes.txt	2023-03-21 10:15	411
assembly_status.txt	2023-04-06 05:34	14
md5checksums.txt	2023-04-04 13:58	193K

[HHS Vulnerability Disclosure](#)

Assembly statistics

RefSeq	GenBank
--------	---------

File Transfer Protocol – FTP (FTPS, SFTP)

Using your Browser as an FTP Client

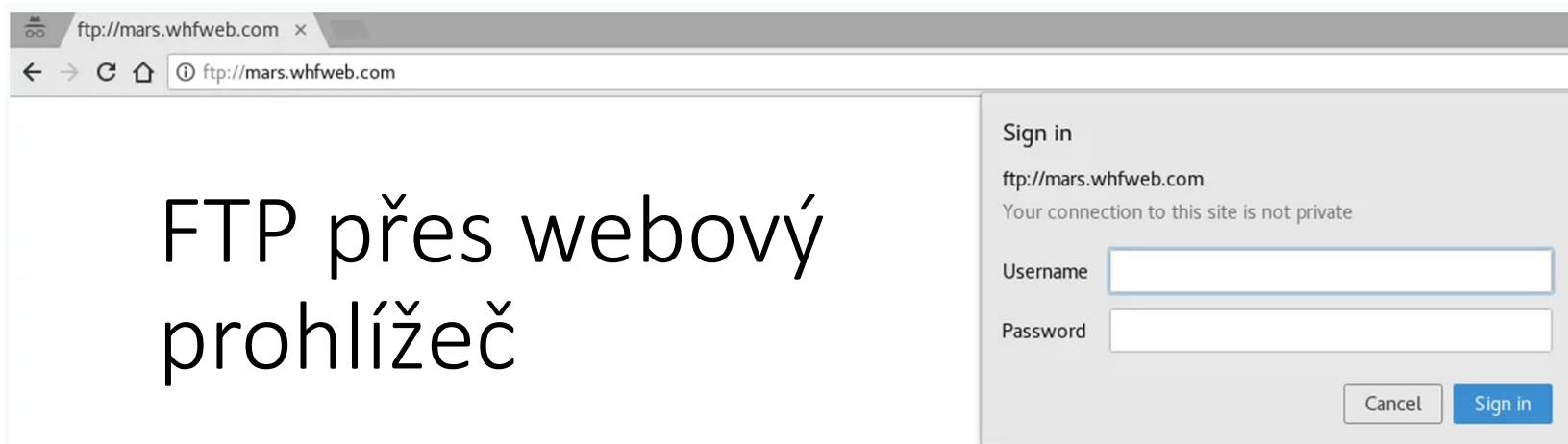
1. Open your Browser, in our example I'll use Chrome

2. In your Address Bar, you can enter: **ftp://Host**

* Note: If you're not certain what your FTP details are you can read our article from [here](#).

In our example, I'll use mars.whfweb.com for **Hostname**, so I'll connect by typing **ftp://mars.whfweb.com**

3. If you didn't directly use your FTP User and its Password in the URL you will be prompted for them



FTP Software

From sources across the web



FileZilla

GNU General Public License



Commander One

Proprietary software



CuteFTP

Proprietary software



WinSCP

GNU General Public License



Transmit

Proprietary software



Core FTP

Freeware



Cyberduck

GNU General Public License



SmartFTP

Proprietary software



Filezilla server

GNU General Public License



30 more



Feedback

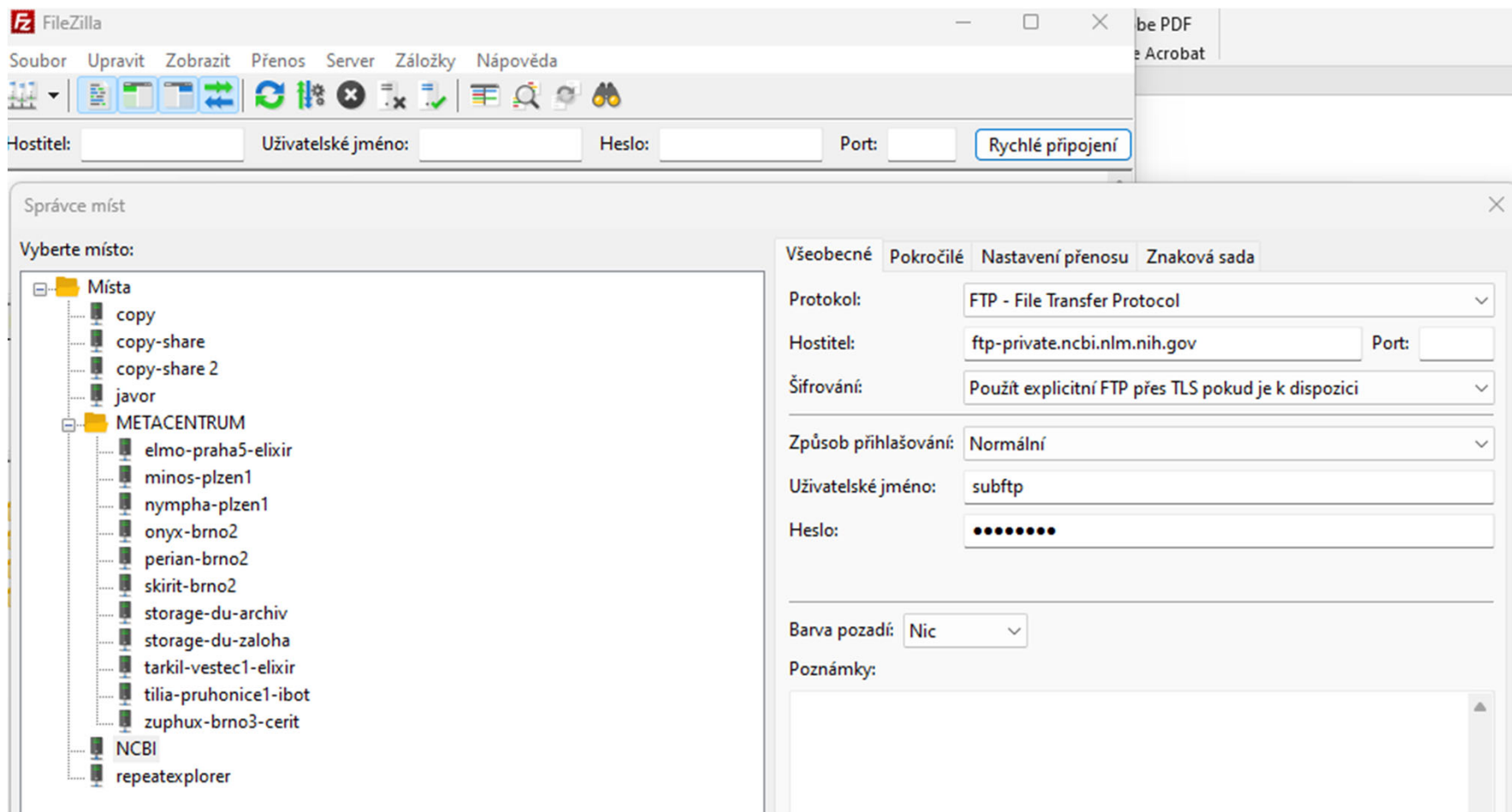
Co musím znát:

Hostitel –

Uživatelské jméno –

Heslo –

(port + protokol)





Hostitel: Uživatelské jméno: Heslo: Port:

tav: Připojení navázáno, čekání na uvítací zprávu...
 tav: Nezabezpečený server, nepodporuje FTP přes TLS.
 tav: Přihlášen
 tav: Načítání výpisu složky „/uploads/vpeska_sci.muni.cz_ARrmq53e“...
 tav: Výpis složky „/uploads/vpeska_sci.muni.cz_ARrmq53e“ proběhl úspěšně

Místní složka: D:\PAVOUCI2024\

- + PAVOUCI2024
- + scripts
- System Volume Information
- + trfi-fagopyrum-20230307

Vzdálená složka: /uploads/vpeska_sci.muni.cz_ARrmq53e

- ? /
- ? uploads
 - vpeska_sci.muni.cz_ARrmq53e

Název souboru

- ..
- literatura
- old
- opraveni
- list-conflict-of-interest_2023.pdf
- C1-final.pdf
- 2305560_CZ_f.pdf

Název souboru

- ..

Prázdný výpis složky

Hostitel: Uživatelské jméno: Heslo: Port: [Rychlé připojení](#)

stav: Připojování k 130.14.250.5:21...
stav: Připojení navázáno, čekání na uvítací zprávu...
stav: Nezabezpečený server, nepodporuje FTP přes TLS.
stav: Přihlášen
stav: Spouští se odesílání D:\9-10-analyza_sekvenci_metody2022.pptx na server

Místní složka:	Vzdálená složka:
D:\	/uploads/vpeska_sci.muni.cz_ARrmq53e
 D: (DATA)	 /
 SRECYCLE.BIN	 uploads
 AnyMP4 Temp	 vpeska_sci.muni.cz_ARrmq53e
 downloads	
Název souboru	Název souboru
 trfi-fagopyrum-20230307	 ..
 Vratislav	
 .dropbox.device	
 ~\$9-10-analyza_sekvenci_metody2022.pptx	
 Bločky pro PFGE.docx	
 prihlaska_do_gsgm.doc	
 20230323_metacentrum_handout.docx	
 9-10-analyza_sekvenci_metody2022.pptx	
 BFU-AVCR.pptx	
	Prázdný výpis složky
/ybrán 1 soubor. Celková velikost: 9 735 531 bajtů	Prázdna složka.

Server nebo místní soubor

 subftp@ftp-private.ncbi.nlm.nih.gov

☐ D:\9-10-analyza_sekvenci_metody2022.pptx

00:00:03 uplynulo

00:00:08 zbývá

 37.6%

3 670 016 bajtů (973,4 KiB/s)

Soubory ve frontě (1)

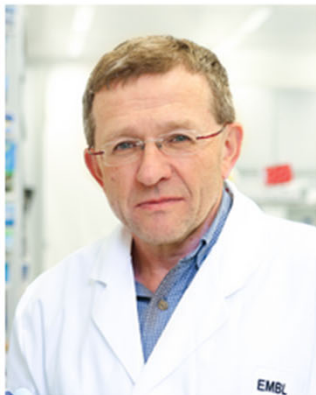
Neúspěšné přenosy

Úspěšné přenosy

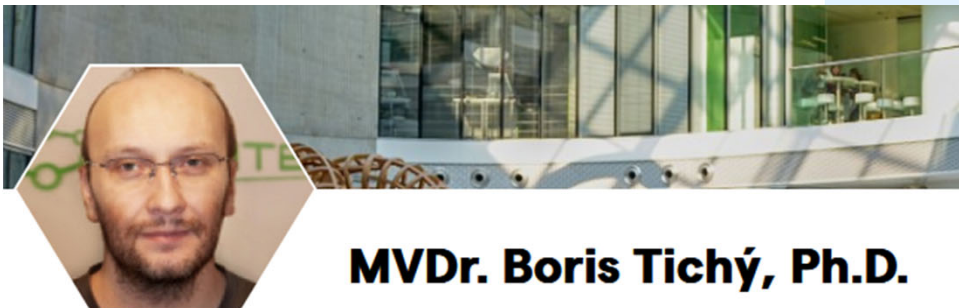
Sekvenační servis nebo centrální laboratoř (CF – Core Facility)



People



Vladimir Benes
Head of GeneCore
Room V106
Phone: +49 6221 387-8486
benes@embl.de



MVDr. Boris Tichý, Ph.D.

Vedoucí sdílené laboratoře

Researcher ID [C-1314-2009](#)



[Services](#) [Applications](#) [Technology](#) [Resources](#) [About](#) [Contact Us](#)

Whole Genome Sequencing

Whole Genome Sequencing (WGS) provides a deep insight into the DNA sequence of humans, animals, plants, and microorganisms. With data analysis at the individual or population level, SNP/INDEL/CNV/SV and other variants of the genome can be fully analysed. Our sequencing analysis enables the identification of somatic and germline mutations as well as customized patterns of cancers and other diseases.

[Request A Quote](#)

Contact us to discuss how we can help you achieve your research goals.

Human
Whole Genome Sequencing



Animal & Plant
Whole Genome Sequencing



Microbial
Whole Genome Sequencing



Sequencing Platform

PACBIO PacBio Sequel II/IIe System

System specifications:
N50>15kb, read lengths up to 25kb (CCS),
>99.999% (QV50) consensus accuracy,
coverage across high-CC/repeat regions. HiFi
reads for PacBio Sequel IIe only.

[LEARN MORE](#)



OmethION
System specifications:
N50>2Mb, high yields for
real-time and accessible



Jak vypadá produkt poskytnutý sekvenačním servisem nebo centrální laboratoří (CF – core facility)

```
TEL103_S11_L002_R2_001.fastq.gz
TEL103_S11_L002_R1_001.fastq.gz
TEL102_S10_L001_R2_001.fastq.gz
TEL102_S10_L001_R1_001.fastq.gz
TEL100_S9_L001_R2_001.fastq.gz
TEL100_S9_L001_R1_001.fastq.gz
TEL97_S14_L002_R2_001.fastq.gz
TEL97_S14_L002_R1_001.fastq.gz
SME_S6_L001_R2_001.fastq.gz
SME_S6_L001_R1_001.fastq.gz
SCENE_S13_L002_R2_001.fastq.gz
SCENE_S13_L002_R1_001.fastq.gz
RACR_S7_L001_R2_001.fastq.gz
RACR_S7_L001_R1_001.fastq.gz
PYR_S20_L002_R2_001.fastq.gz
PYR_S20_L002_R1_001.fastq.gz
PpWT1_S1_L001_R2_001.fastq.gz
PpWT1_S1_L001_R1_001.fastq.gz
PARA_S12_L002_R2_001.fastq.gz
PARA_S12_L002_R1_001.fastq.gz
P_S16_L002_R2_001.fastq.gz
P_S16_L002_R1_001.fastq.gz
L_S17_L002_R2_001.fastq.gz
L_S17_L002_R1_001.fastq.gz
IPS_S19_L002_R2_001.fastq.gz
IPS_S19_L002_R1_001.fastq.gz
CHX302_S8_L001_R2_001.fastq.gz
CHX302_S8_L001_R1_001.fastq.gz
H_S18_L002_R2_001.fastq.gz
H_S18_L002_R1_001.fastq.gz
FTAR_S5_L001_R2_001.fastq.gz
FTAR_S5_L001_R1_001.fastq.gz
FESC_S15_L002_R2_001.fastq.gz
FESC_S15_L002_R1_001.fastq.gz
DKC MOCK_S4_L001_R2_001.fastq.gz
DKC MOCK_S4_L001_R1_001.fastq.gz
DKC_IP7_S3_L001_R2_001.fastq.gz
DKC_IP7_S3_L001_R1_001.fastq.gz
DKC_IP6_S2_L001_R2_001.fastq.gz
```

- Přes webové rozhraní
- Externí disk
- FTP
- Filesender
- Aspera
- Etc.

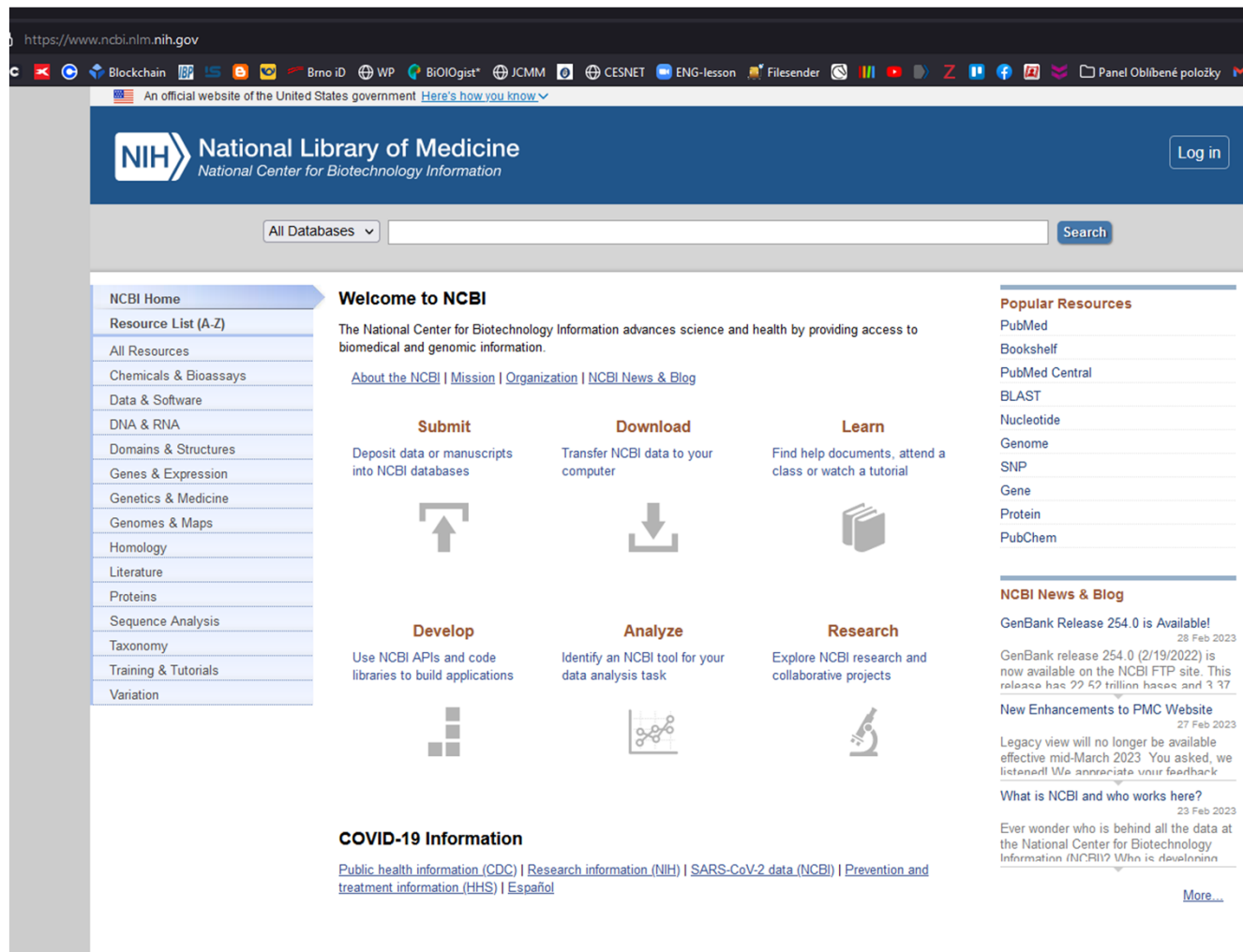
```
1 597 274 KB
1 506 485 KB
3 567 597 KB
3 392 515 KB
2 088 595 KB
2 013 927 KB
1 443 726 KB
1 385 953 KB
1 980 845 KB
1 881 276 KB
1 948 463 KB
1 847 953 KB
1 737 364 KB
1 662 259 KB
5 008 736 KB
4 754 863 KB
2 933 673 KB
2 712 568 KB
1 945 228 KB
1 835 490 KB
1 987 326 KB
1 886 706 KB
1 692 458 KB
1 609 402 KB
4 271 811 KB
4 085 774 KB
2 363 943 KB
2 226 033 KB
1 635 235 KB
1 559 576 KB
2 784 510 KB
2 675 555 KB
1 756 909 KB
1 691 095 KB
1 444 264 KB
1 318 614 KB
2 597 897 KB
2 380 177 KB
1 933 748 KB
```

```
PAM42230_pass_135f8216_ed2e97e7_162.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_161.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_160.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_159.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_158.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_157.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_156.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_155.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_154.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_153.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_152.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_151.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_150.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_149.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_148.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_147.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_146.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_145.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_144.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_143.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_142.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_141.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_140.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_139.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_138.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_137.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_136.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_135.fastq.gz
PAM42230_pass_135f8216_ed2e97e7_134.fastq.gz
```

53,5 GB z 53,5 GB v 2 855 z 2 855

```
19 897 KB
20 124 KB
19 839 KB
20 616 KB
18 159 KB
19 752 KB
20 038 KB
20 243 KB
20 654 KB
19 804 KB
20 518 KB
20 268 KB
19 866 KB
17 925 KB
20 533 KB
19 538 KB
20 242 KB
20 534 KB
20 272 KB
21 058 KB
19 007 KB
19 980 KB
20 086 KB
19 480 KB
20 256 KB
20 059 KB
20 506 KB
20 014 KB
19 998 KB
```

Nahrávání hrubých NGS dat do SRA NCBI



The screenshot shows the NCBI National Library of Medicine homepage. The browser address bar displays <https://www.ncbi.nlm.nih.gov>. The page header includes the NIH logo, the text "National Library of Medicine" and "National Center for Biotechnology Information", and a "Log in" button. Below the header is a search bar with a dropdown menu set to "All Databases" and a "Search" button. The main content area is divided into several sections:

- NCBI Home**: A vertical sidebar menu on the left with links to "Resource List (A-Z)", "All Resources", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", "Genes & Expression", "Genetics & Medicine", "Genomes & Maps", "Homology", "Literature", "Proteins", "Sequence Analysis", "Taxonomy", "Training & Tutorials", and "Variation".
- Welcome to NCBI**: A central section with the text "The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." and links to "About the NCBI", "Mission", "Organization", and "NCBI News & Blog".
- Submit**: A section with the text "Deposit data or manuscripts into NCBI databases" and an icon of an upward arrow.
- Download**: A section with the text "Transfer NCBI data to your computer" and an icon of a downward arrow.
- Learn**: A section with the text "Find help documents, attend a class or watch a tutorial" and an icon of a book.
- Develop**: A section with the text "Use NCBI APIs and code libraries to build applications" and an icon of a stack of blocks.
- Analyze**: A section with the text "Identify an NCBI tool for your data analysis task" and an icon of a network diagram.
- Research**: A section with the text "Explore NCBI research and collaborative projects" and an icon of a microscope.
- COVID-19 Information**: A section with the text "Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español".
- Popular Resources**: A section on the right with links to "PubMed", "Bookshelf", "PubMed Central", "BLAST", "Nucleotide", "Genome", "SNP", "Gene", "Protein", and "PubChem".
- NCBI News & Blog**: A section on the right with news items, including "GenBank Release 254.0 is Available!" dated 28 Feb 2023, "GenBank release 254.0 (2/19/2022) is now available on the NCBI FTP site. This release has 22.52 trillion bases and 3.37 million genes." dated 27 Feb 2023, "New Enhancements to PMC Website" dated 27 Feb 2023, "Legacy view will no longer be available effective mid-March 2023. You asked, we listened! We appreciate your feedback." dated 27 Feb 2023, and "What is NCBI and who works here?" dated 23 Feb 2023. A "More..." link is at the bottom.



3rd-Party login is now mandatory

If you log into an account with an unretired NCBI password, you will be redirected to our Password Retirement Wizard.

- The wizard will guide you through the steps of linking a 3rd-party account if needed and retiring your password.
- For more information, read our [FAQs](#) and [Transition Tips](#).

Log in



eRA Commons



Google Account



ORCID



Login.gov



Microsoft



NIH Account



NCBI Account



vpeska

Search

Other Tools

TSA

Submit computationally assembled, transcribed RNA sequences after submitting unassembled reads to SRA. [Learn more >](#)

GEO

Submit RNA-seq, ChIP-seq, and other types of gene expression and epigenomics datasets. [Learn more >](#)

BioProject & BioSample

Choose a tool above if submitting sequence data. [Learn more >](#)

What You Should Expect

How to submit

BioProject

Sample types

Sample metadata

You should register BioProject or BioSamples separately from your data **only in the following situations**:

1. Large and long-term projects where samples are collected over a course of year or more
2. If an NCBI curator instructed you to register a separate BioProject or BioSamples
3. If you are submitting an annotated genome before submitting the reads or the unannotated genome

[Next >](#)

[Submit](#)

BioProject

New submission

ATTN: to update an existing record or recent submission, you can use "Manage data" to make some changes yourself. If you cannot make the desired change there, then please [email your request](#) with your BioProject accession or Submission ID included. **Do not** create new submission to update an existing submission!

Short description and brief instructions

+

5 submissions				
Submission	Title	Group	Status	Updated
SUB10948809	Acalles testensis Metagenome		✓ BioProject: Processed PRJNA796755 : Evolution of telomeres and telomerases in plants (TaxID: 932851) Manage data	05:00
SUB10948813	Acalles testensis Metagenome		⚙️ Unfinished at the Review & Submit step ✖ BioProject: Error Similar projects already exist: PRJNA796755	Jan 13 2022
SUB6671801	Telomere sequence identification in Zostera genus		✓ BioProject: Processed PRJNA594842 : Telomere sequence identification in Zostera genus Manage data	Jan 01 2021
SUB5615852	Identification of telomerase RNA in plants		✓ BioProject: Processed PRJNA542932 : Identification of telomerase RNA in plants Locus Tag Prefixes: • None (SAMN11639744) • None (SAMN11639746) locustagprefix.txt Manage data	Jun 16 2020
SUB4980599	Comparative study of repeats in onion, garlic, and wild garlic		✓ BioProject: Processed PRJNA512235 : Comparative study of repeats in onion, garlic, and wild garlic Manage data	May 14 2019

Formát sekvenčních dat

- FASTQ
- FASTA
- SAM/BAM
- GFF
- BED
- VCF



FASTQ [fa:stkju:]

Kontrolní otázka:

100 řádků FASTQ, kolik je to readů a co z toho vyplývá ohledně párovosti dat?

- Nukleotidová sekvence + odpovídající Phred quality scores (Q)
- textový formát (4 řádky/1sekvence)
- přípona obvykle *.fq nebo *.fastq (často GNU zip, *.fastq.gz)
- sekvence + kvalita zastoupena jedním znakem ASCII
- Malý soubor v notepadu, velký soubor nahlížení pomocí less/more, head/tail (linux terminal – lt)
- FASTQC / MULTIQC



Phred score (Q)

- 10x záporný logaritmus o základu 10 pravděpodobnosti chyby

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call			Base call accuracy
10	1 in 10	P = 0.1	10^{-1}	90%
20	1 in 100	P = 0.01	10^{-2}	99%
30	1 in 1000	P = 0.001	10^{-3}	99.9%
40	1 in 10,000	P = 0.0001	10^{-4}	99.99%
50	1 in 100,000	P = 0.00001		99.999%
60	1 in 1,000,000	P = 0.000001		99.9999%

$$Q = -10 \log_{10} P$$

$$x = \log_a y$$

$$a^x = y$$

Převodní tabulka různých Q_s

[illegible]

S - Sanger Phred+33, raw reads typically (0, 40)
X - Solexa Solexa+64, raw reads typically (-5, 40)
I - Illumina 1.3+ Phred+64, raw reads typically (0, 40)
J - Illumina 1.5+ Phred+64, raw reads typically (3, 41)
with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)
(Note: See discussion above).
L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)
P - PacBio Phred+33, HiFi reads typically (0, 93)

Hlavičky readů

```
@HWI-ST193:542:C2H0GACXX:8:1101:4404:2179 1:Y:0:ACACGA  
ATGCNTTTTATAATCAAAGCAGAAGCTTTATGCTAGCTAGCATATAAT
```

+

```
<<<<<#2<:5>:944>:??AAAAAAAAABAAAABBBBBB??????????
```

@ - začátek readů (hlavička)

@HWI-ST193 - Jméno stroje (od výrobce)

542 - run ID

C2H0GACXX - flowcell id

8 - flowcell lane

1101 - tile number

4404 - x-coordinate

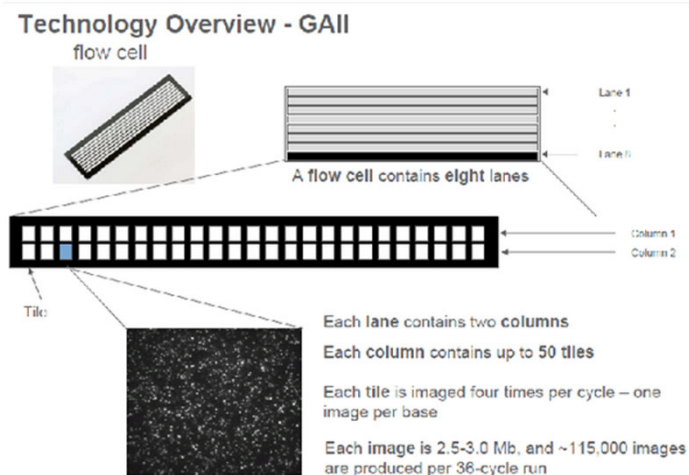
2179 - y-coordinate

1 - část páru

Y - filter status (Y/N)

0 - status kontrol. Bit. (0 nebo sudá čísla)

ACACGA - sekvence indexu



FASTQC

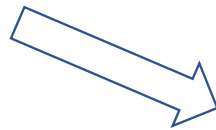
- První náhled na NGS data (short reads, illumina)
- Výstup - HTML formát
- Možnost spouštět z terminalu (linux) i v rámci GALAXY (web)

```
@A00180:10:H7NKS0MXX:1:1101:16821:1000 1:N:0:GGACTT+GTCGTTTCG
TGCAGCAGTAATGAGGAACCACTTCCTCCCTCCAGCCGCTCTAAATACCTCAGAACAAATAGGATCATATAATCCCTAGTCTGAAGT
+
8FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:19090:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGAAGAAGAAAGCACAAGTATTTACGCCTATCCTTCATATTTCCGCAAGGTAACATATCTCGTTTCATATCGAGATTATATAGAATCT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF -FFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:19325:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGGAAGTTATGCAGGGGCATCCTGTATTATTAATAGAGCACCTACTCTTCATAGATTAGGTATACAGGCGTTCCAACTATTTTAGTGG
+
FF-88F8FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF -FFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:30897:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGAGTACATCAACAAAAGAAACCTAACTGCCCTACCGCAAACCGGTAGAGTACCTTCCCCAAAAGTATTACTCCAGTCAATATAAGG
+
8FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:19090:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGAAGAAGAAAGCACAAGTATTTACGCCTATCCTTCATATTTCCGCAAGGTAACATATCTCGTTTCATATCGAGATTATATAGAATCT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF -FFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:19325:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGGAAGTTATGCAGGGGCATCCTGTATTATTAATAGAGCACCTACTCTTCATAGATTAGGTATACAGGCGTTCCAACTATTTTAGTGG
+
FF-88F8FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF -FFFFFFFFF
@A00180:10:H7NKS0MXX:1:1101:30897:1016 1:N:0:GGACTT+GTCGTTTCG
TGCAGAGTACATCAACAAAAGAAACCTAACTGCCCTACCGCAAACCGGTAGAGTACCTTCCCCAAAAGTATTACTCCAGTCAATATAAGG
+
8F888FFF -FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```



FASTQC summary + basic statistics

Dobrá data



Summary

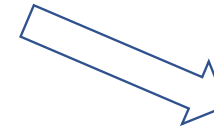
- ✓ [Basic Statistics](#)
- ✓ [Per base sequence quality](#)
- ✓ [Per tile sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ✓ [Per base sequence content](#)
- ✓ [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ✓ [Sequence Duplication Levels](#)
- ✓ [Overrepresented sequences](#)
- ✓ [Adapter Content](#)

Basic Statistics

Measure	Value
Filename	good_sequence_short.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Sequences flagged as poor quality	0
Sequence length	40
%GC	45

Analýza (mapování,
sestavování
genomu)

Špatná data



Summary

- ✓ [Basic Statistics](#)
- ✗ [Per base sequence quality](#)
- ✗ [Per tile sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ! [Per base sequence content](#)
- ! [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ! [Sequence Duplication Levels](#)
- ! [Overrepresented sequences](#)
- ✓ [Adapter Content](#)

Basic Statistics

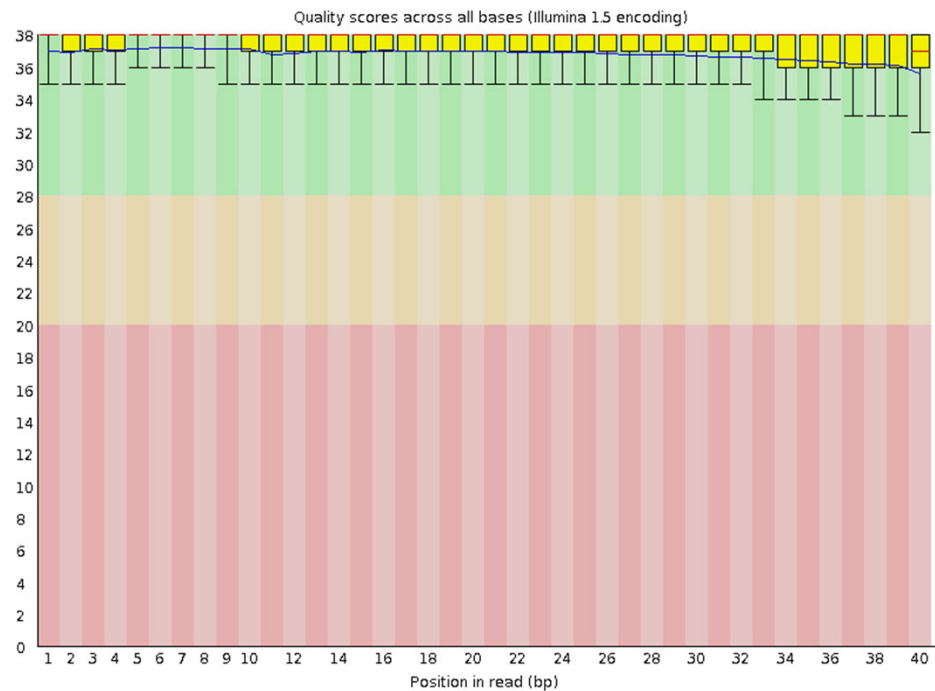
Measure	Value
Filename	bad_sequence.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	395288
Sequences flagged as poor quality	0
Sequence length	40
%GC	47

Pre-processing
Trimming
Filtrování

Per base sequence quality

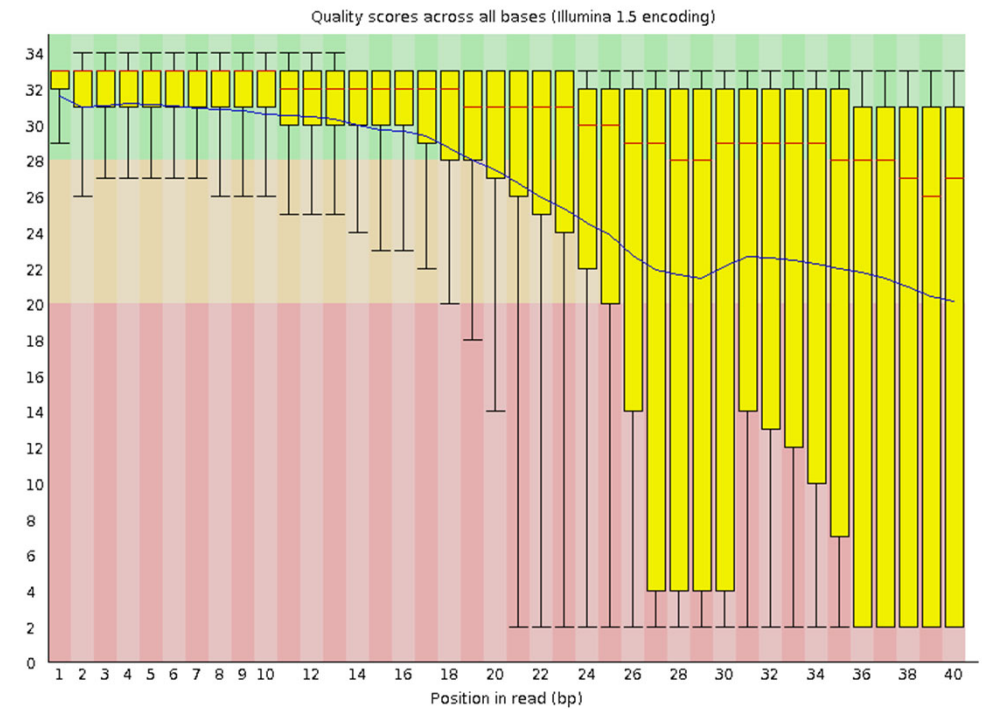
Dobrá data

✓ Per base sequence quality



Špatná data

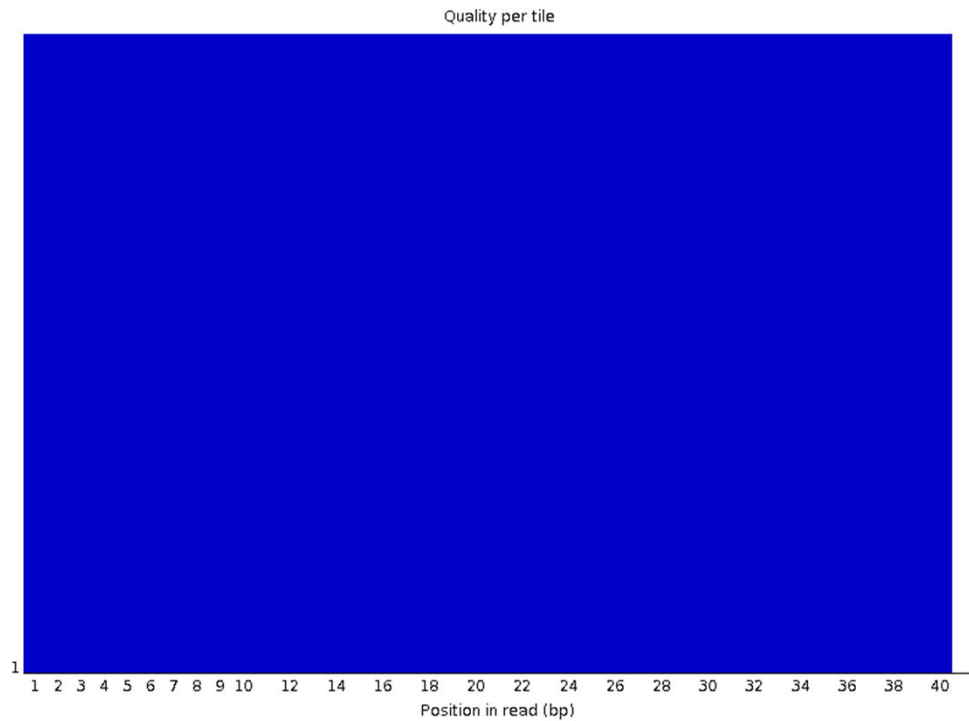
✗ Per base sequence quality



Per tile sequence quality

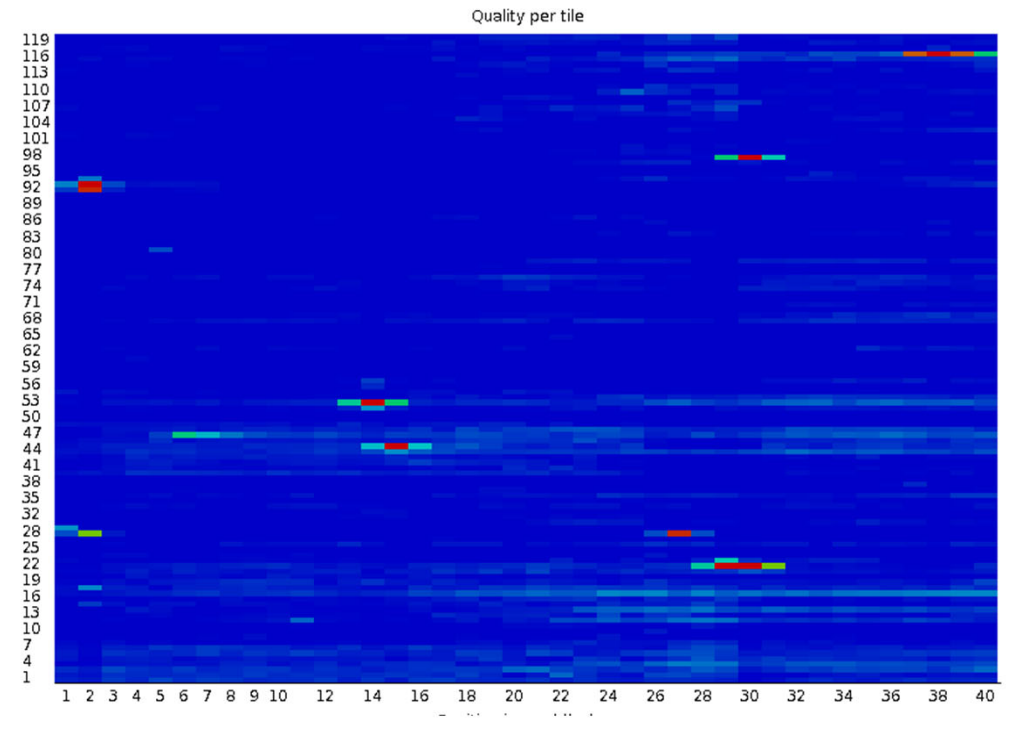
Dobrá data

✅ Per tile sequence quality



Špatná data

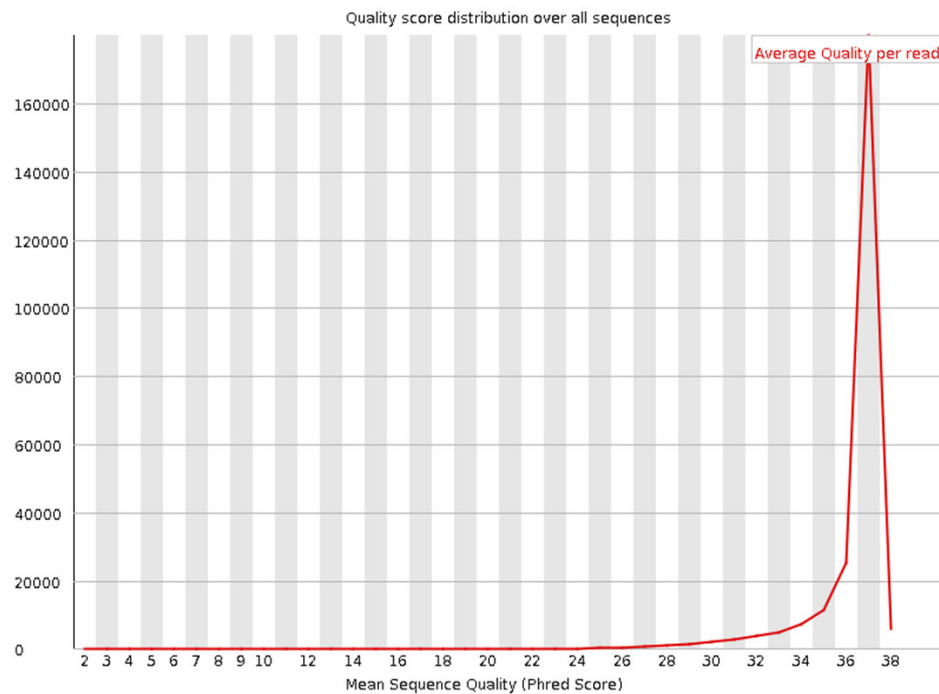
❌ Per tile sequence quality



Per sequence quality scores

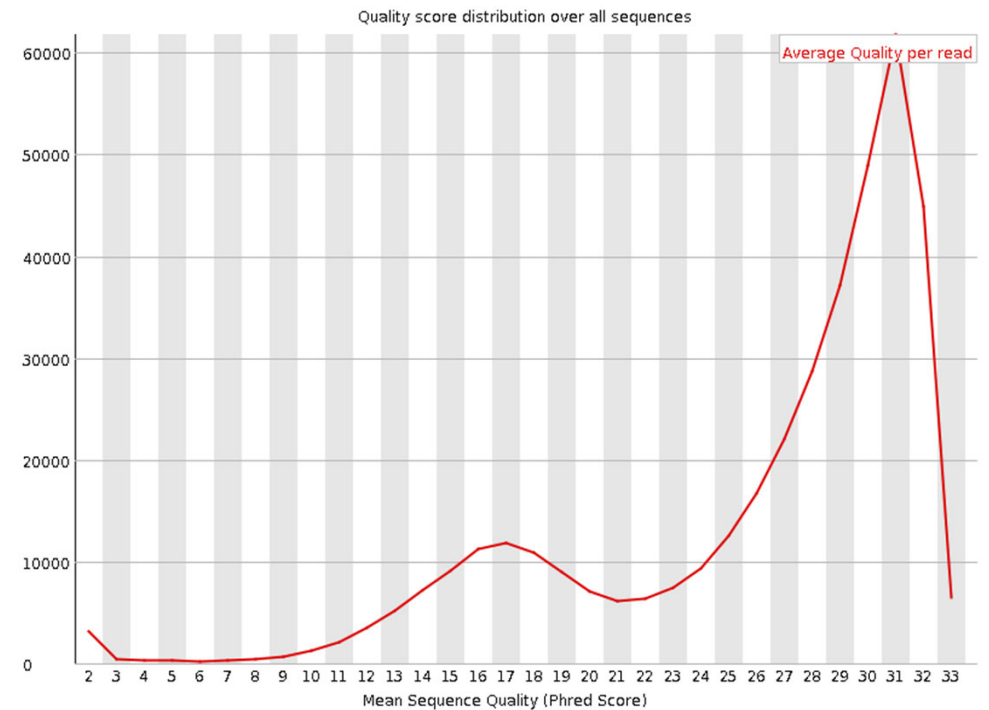
Dobrá data

✅ Per sequence quality scores



Špatná data

✅ Per sequence quality scores

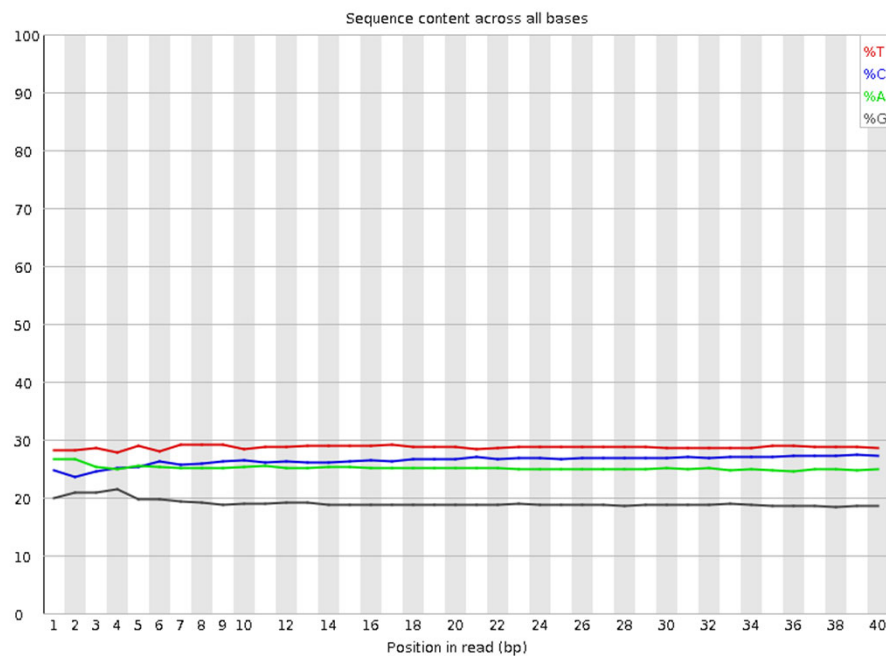


Per base sequence content

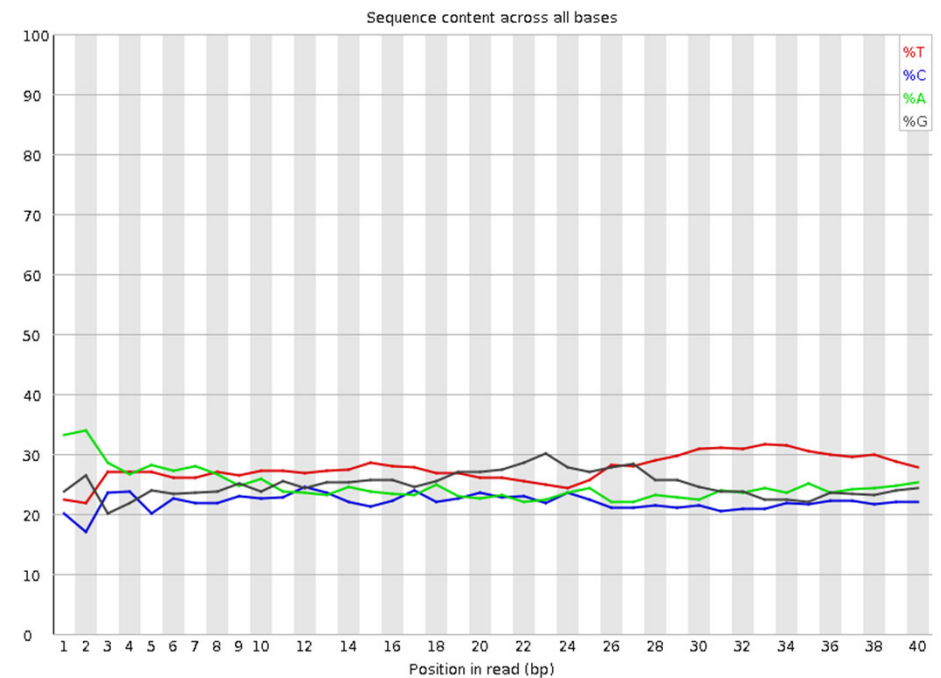
Dobrá data

Špatná data

✔ Per base sequence content



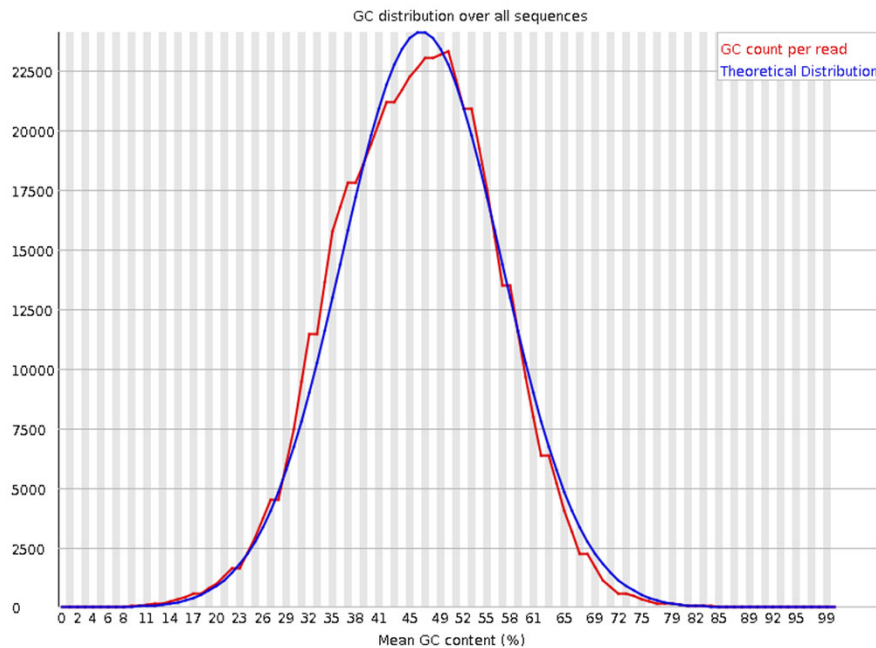
⚠ Per base sequence content



Per sequence GC content

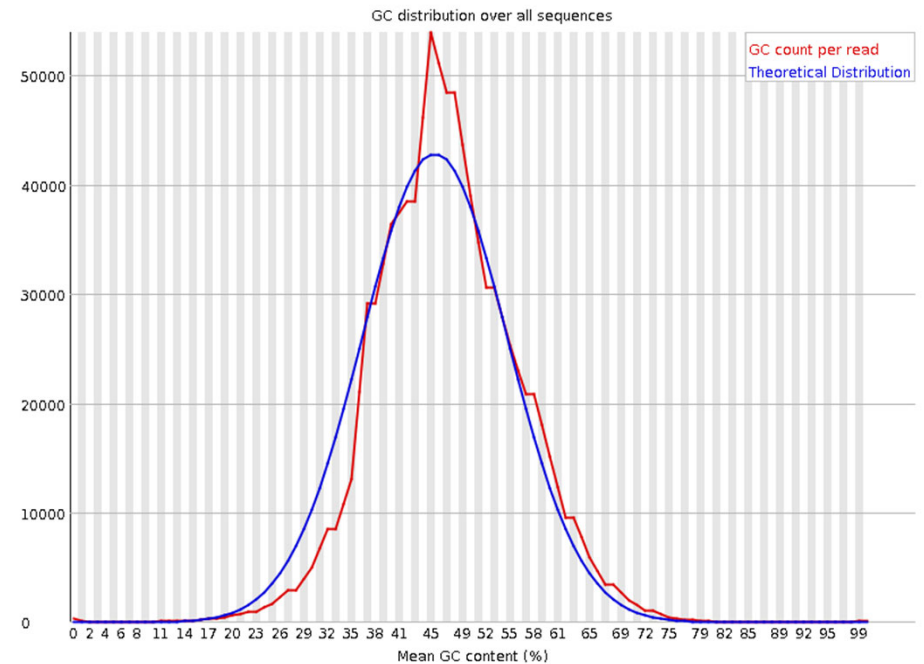
Dobrá data

✓ Per sequence GC content



Špatná data

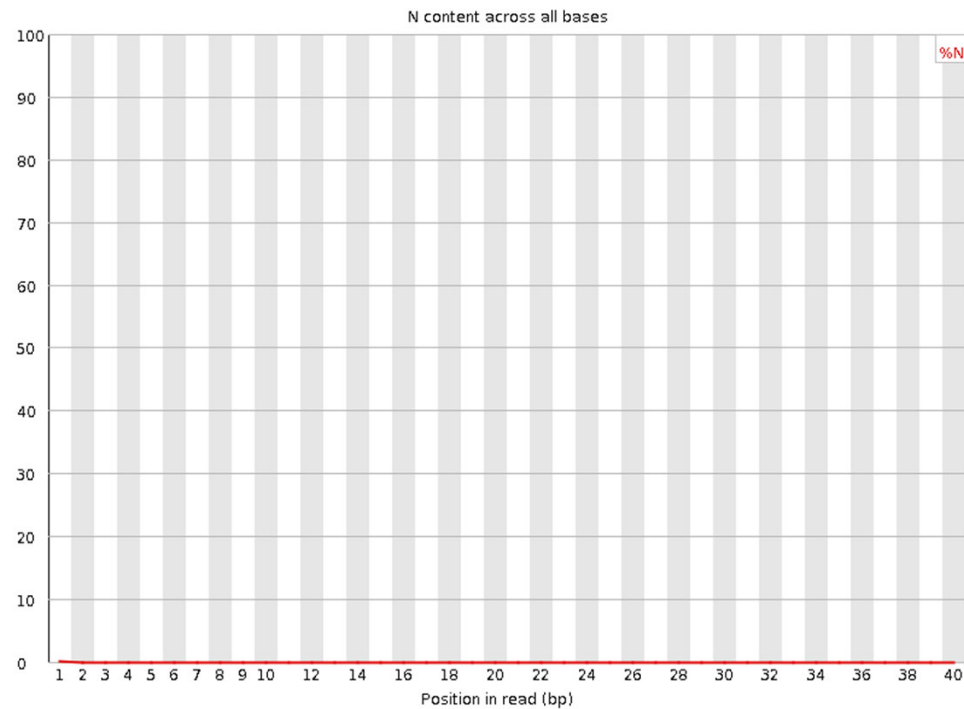
⚠ Per sequence GC content



Per base N content

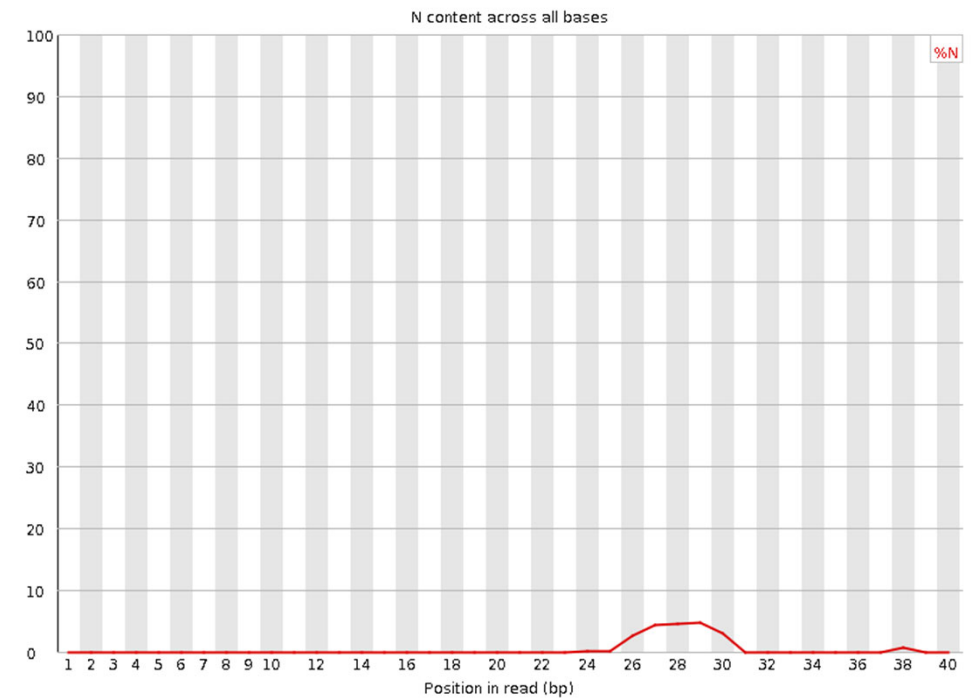
Dobrá data

✓ Per base N content



Špatná data

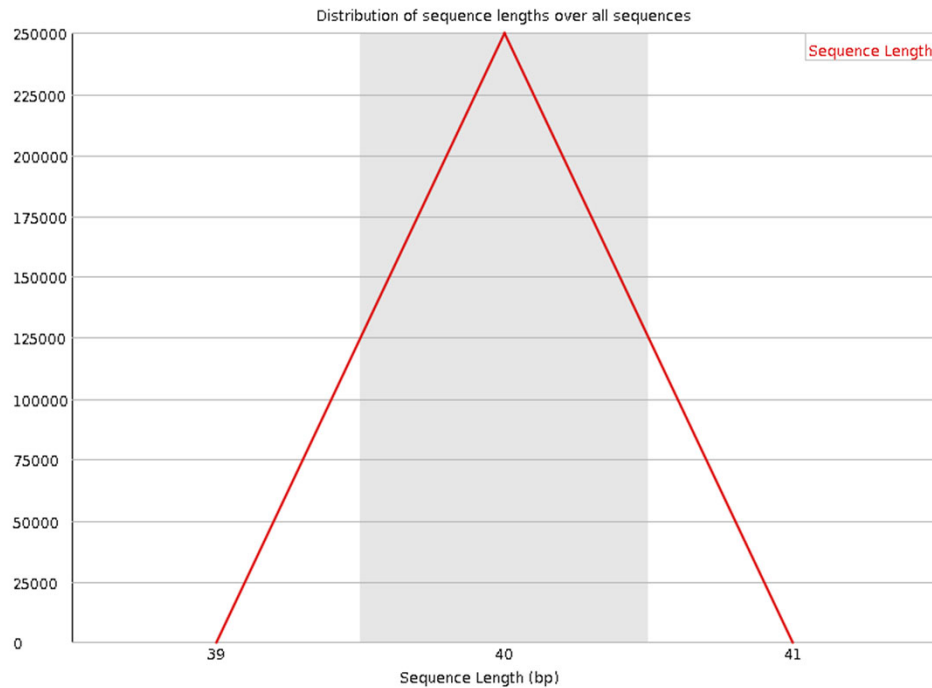
✓ Per base N content



Sequence Length Distribution

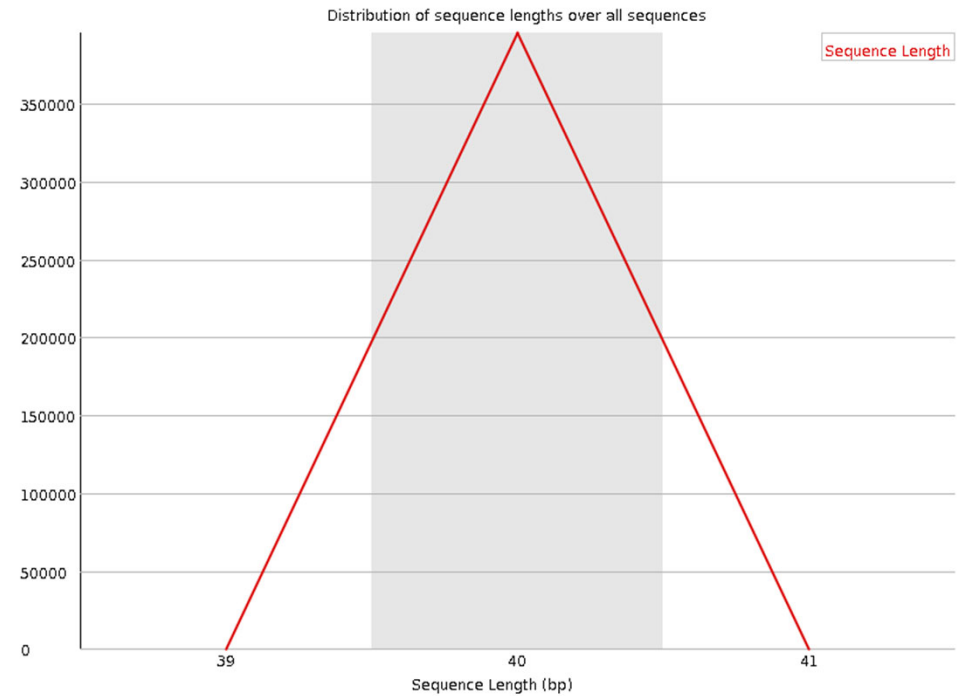
Dobrá data

✔ Sequence Length Distribution



Špatná data

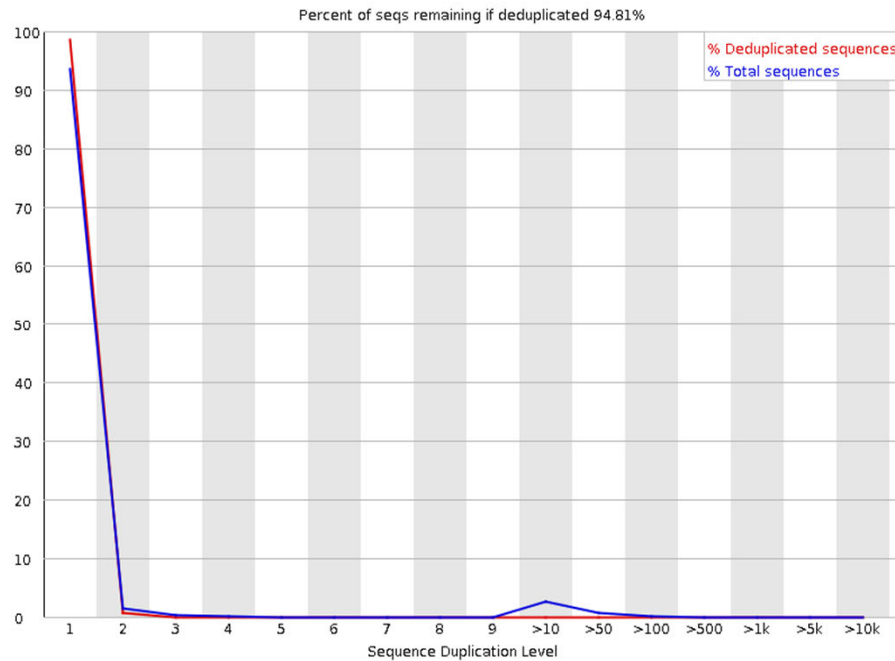
✔ Sequence Length Distribution



Sequence Duplication Levels

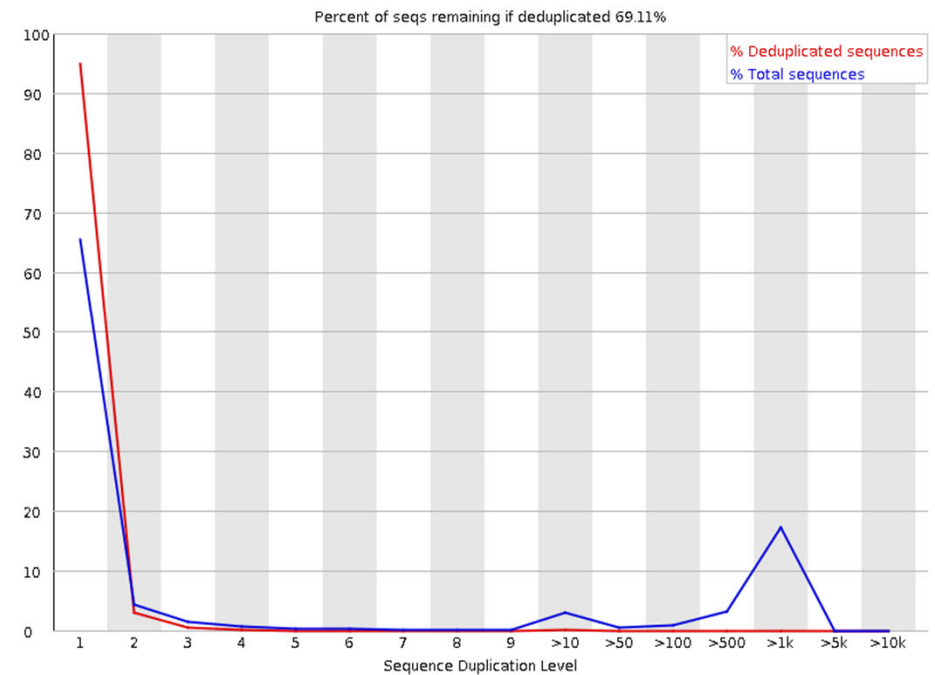
Dobrá data

✔ Sequence Duplication Levels



Špatná data

⚠ Sequence Duplication Levels



Overrepresented sequences

Dobrá data



Overrepresented sequences

No overrepresented sequences

Špatná data



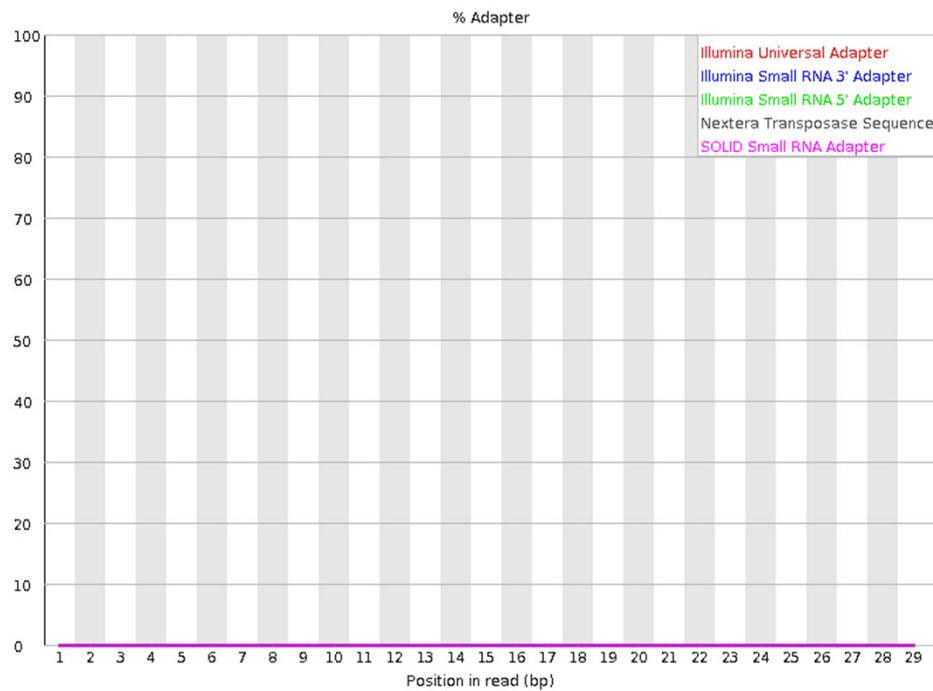
Overrepresented sequences

Sequence	Count	Percentage	Possible Source
AGAGTTTTATCGCTTCCATGACGCAGAAAGTTAACACTTTC	2065	0.5224039181558763	No Hit
GATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATG	2047	0.5178502762542754	No Hit
ATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGA	2014	0.5095019327680071	No Hit
CGATAAAAATGATTGGCGTATCCAACCTGCAGAGTTTTAT	1913	0.4839509420979134	No Hit
GTATCCAACCTGCAGAGTTTTATCGCTTCCATGACGCAGA	1879	0.47534961850600066	No Hit
AAAAATGATTGGCGTATCCAACCTGCAGAGTTTTATCGCT	1846	0.4670012750197325	No Hit
TGATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCAT	1841	0.46573637449150995	No Hit
AACCTGCAGAGTTTTATCGCTTCCATGACGCAGAAAGTTAA	1836	0.46447147396328753	No Hit
GATAAAAATGATTGGCGTATCCAACCTGCAGAGTTTTATC	1831	0.4632065734350651	No Hit
AAATGATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTC	1779	0.45005160794155147	No Hit
ATGATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCA	1779	0.45005160794155147	No Hit
AATGATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCC	1760	0.4452449859343061	No Hit
AAAATGATTGGCGTATCCAACCTGCAGAGTTTTATCGCTT	1729	0.4374026026593269	No Hit
CGTATCCAACCTGCAGAGTTTTATCGCTTCCATGACGCAG	1713	0.43335492096901496	No Hit
ATCCAACCTGCAGAGTTTTATCGCTTCCATGACGCAGAAG	1708	0.43209002044079253	No Hit
CAGAGTTTTATCGCTTCCATGACGCAGAAAGTTAACACTTT	1684	0.42601849790532476	No Hit
TGCAGAGTTTTATCGCTTCCATGACGCAGAAAGTTAACACT	1668	0.4219708162150128	No Hit
CAACCTGCAGAGTTTTATCGCTTCCATGACGCAGAAAGTTA	1668	0.4219708162150128	No Hit
TATCCAACCTGCAGAGTTTTATCGCTTCCATGACGCAGAA	1630	0.4123575722005221	No Hit
GTCATGGAAGCGATAAACTCTGCAGGTTGGATACGCCAA	1620	0.40982777114407726	No Hit
AACCTCTGCGTCATGGAAGCGATAAACTCTGCAGGTTGG	1616	0.4088158507214993	No Hit
GCAGAGTTTTATCGCTTCCATGACGCAGAAAGTTAACACTT	1580	0.39970856691829754	No Hit
TGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGACG	1569	0.3969257857562082	No Hit
GGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGACGC	1542	0.39009532290380683	No Hit
ATAAAAATGATTGGCGTATCCAACCTGCAGAGTTTTATCG	1481	0.37466353645949285	No Hit
ACCTGCAGAGTTTTATCGCTTCCATGACGCAGAAAGTTAAC	1479	0.37415757624820384	No Hit

Adapter Content

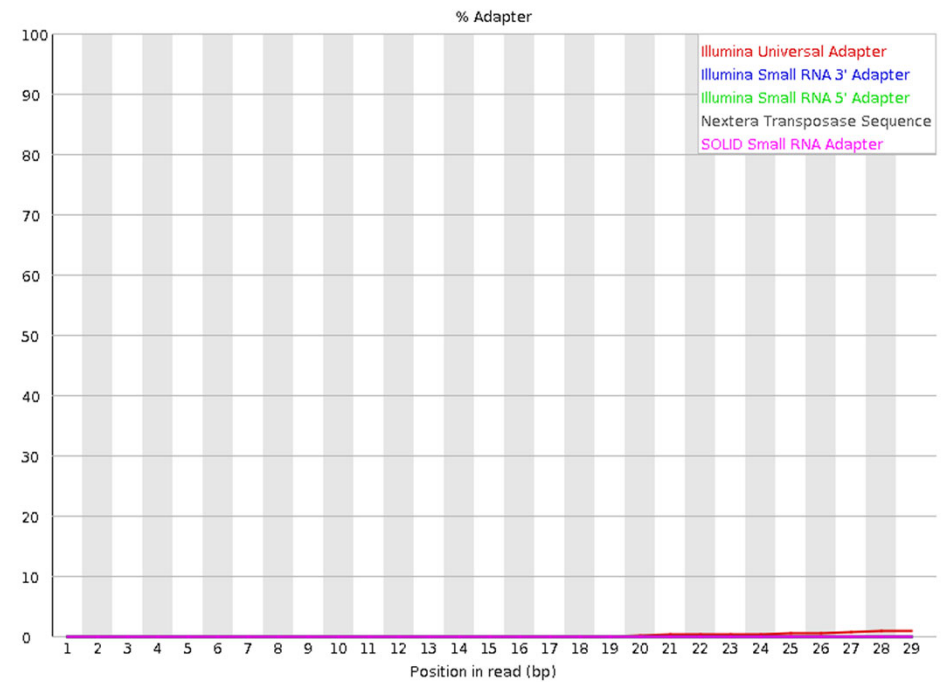
Dobrá data

✓ Adapter Content



Špatná data

✓ Adapter Content



GALAXY

FASTA [fa:stei]

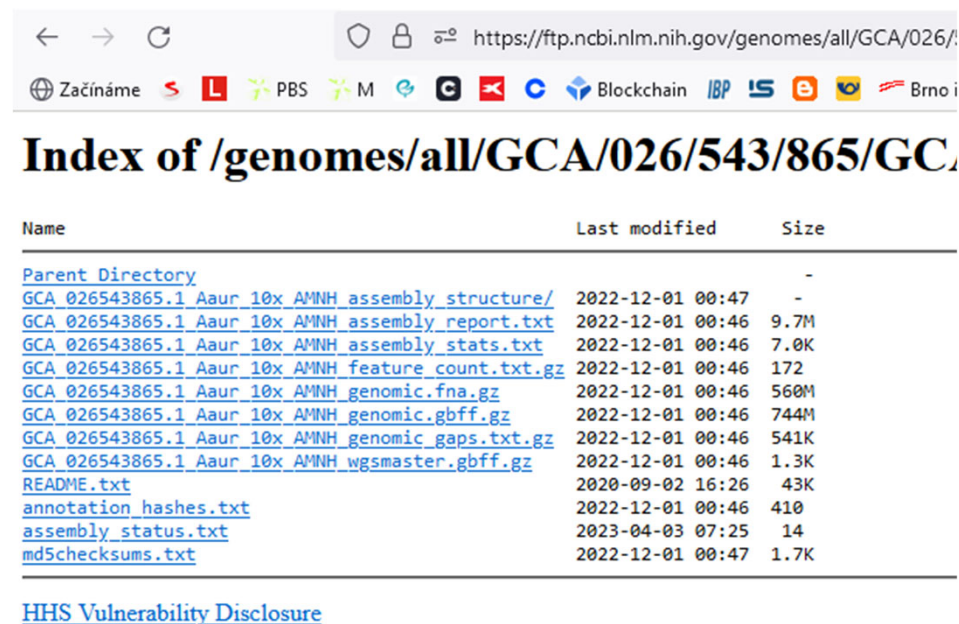
- Textový soubor, přípony *.fa, *.fas, *.fasta, *.fna
- Řádky – lines – command line
- Zalamování (UNIX/DOS)
- NCBI příklad
- multifasta

>hlavička-1

GATCGATCG

>hlavička-2

ATCGATCGATCG



Name	Last modified	Size
Parent Directory	-	-
GCA_026543865.1_Aaur_10x_AMNH_assembly_structure/	2022-12-01 00:47	-
GCA_026543865.1_Aaur_10x_AMNH_assembly_report.txt	2022-12-01 00:46	9.7M
GCA_026543865.1_Aaur_10x_AMNH_assembly_stats.txt	2022-12-01 00:46	7.0K
GCA_026543865.1_Aaur_10x_AMNH_feature_count.txt.gz	2022-12-01 00:46	172
GCA_026543865.1_Aaur_10x_AMNH_genomic.fna.gz	2022-12-01 00:46	560M
GCA_026543865.1_Aaur_10x_AMNH_genomic.gbff.gz	2022-12-01 00:46	744M
GCA_026543865.1_Aaur_10x_AMNH_genomic_gaps.txt.gz	2022-12-01 00:46	541K
GCA_026543865.1_Aaur_10x_AMNH_wgsmaster.gbff.gz	2022-12-01 00:46	1.3K
README.txt	2020-09-02 16:26	43K
annotation_hashes.txt	2022-12-01 00:46	410
assembly_status.txt	2023-04-03 07:25	14
md5checksums.txt	2022-12-01 00:47	1.7K

[HHS Vulnerability Disclosure](#)

SAM (BAM) - sekvence, mapování, kvalita...

- Textový – TAB-delimited/TAB-separated. *.sam *.bam
- SAM = sequence alignment map (BAM – binární SAM tj. zapsaný formou jedniček a nul)

Hlavička je nepovinná a skládá se z řádků:

@HD – (řádek hlavička) **VN:(verze formátu)** **SO:(alignment seřazen podle – unknown/unsorted/queryname/coordinate)** **GO:(group order – none/query/reference)**

@SQ – (slovník referenční sekvence, může být více @SQ řádků) **SN: jméno referenční sekvence** (každá SQ musí mít unikátní SN tag) **LN: (délka referenční sekvence 1 – (1²³-1))** **AS:** identifikátor genomového assembly **M5:MD5** kontrolní součet sekvence **SP:** (species) **UR:** (URI – http;; ftp: file-systém path)

@RG – (read group information , může být více @RG řádků) **ID:(unikátní pro každý @RG)** **CN:**jméno sekvenačního centra **DS:**description **DT:**date run **ISO8601** **FO:**flow order **KS:**key sequence **LB:**library **PG:**program of processing **PI:**median insert size **PL:**platform model **PU:**platform unit unikátní identifikátor **SM:**sample name

@PG – (info o programu pro vytvoření BAM/SAM) **ID:**unikátní identifikátor **PN:**jméno programu **CL:**command line **PP:**previous @PG-ID **DS:**popis programu **VN:**verze programu

@CO Line – (jednořádkový textový komentář)

Sekce alignmentu

11 povinných polí (sloupců) + další volitelné (pole SEQ * = sekvence není uložena)

QNAME: query name (=FASTQ) **FLAG:** **RNAME:**jméno referenční sekvence (jedna z SN z @SQ řádku) **POS:**první shoda v referenci vlevo **MAPQ:**kvalita mapování **CIGAR:**charakter shody **RNEXT:**jméno reference dalšího namapovaného readu **PNEXT:**pozice dalšího namapovaného readu **TLEN:**délka pozorované shody **SEQ:** sekvence readů (=FASTQ) **QUAL:**quality scores(=FASTQ)

@HD VN:1.6 SO:coordinate										
@SQ SN:ref LN:45										
QNAME: FLAG: RNAME: POS: MAPQ: CIGAR: SEQ:										
r001	99	ref	7	30	8M2I4M1D3M	=	37	39	TTAGATAAAGGATACTG	*
r002	0	ref	9	30	3S6M1P1I4M	*	0	0	AAAAGATAAGGATA	*
r003	0	ref	9	30	5S6M	*	0	0	GCCTAAGCTAA	* SA:Z:ref.29.-.6H5M.17.0;

GFF/GTF - general feature format/general transfer format (GFF2=GTF, GFF3)

- Textový soubor – TAB-delimited, přípony *.gff *.gtf
- Popisuje geny a jiné >>features<< v DNA, RNA, proteinových sekvencích
- Každý feature má svůj řádek (8 povinných polí/sloupců, které nesmí být prázdné – >> . <<)

seqname source feature start end score strand frame {attributes comments}

seqname – jméno sekvence, může být shoda s EMBL/Genbank/DDBJ

source – program, který identifikoval feature

feature – gen, exon, repeat, transcript...

start – pozice 1. nt

end – pozice posledního nt

score – floating point value

strand – + (forward) – (reverse)

frame – 0/1/2/.

attributes

comments

```
#!/genome-build GRCh38
#!/genome-date 2013-12
#!/genome-build-accession NCBI:GCA_000001405.15
#!/genebuild-last-updated 2014-08
1 havana gene 11869 14409 . + . gene_id "ENSG00000223972"
1 havana exon 11869 14409 . + . gene_id "ENSG00000223972"
1 havana exon 11869 12227 . + . gene_id "ENSG00000223972"
1 havana exon 12613 12721 . + . gene_id "ENSG00000223972"
```

BED - Browser Extensible Data

- Textový soubor – TAB-delimited, *.bed
- Primární účel - anotace genomových dat
- Mohou mít hlavičku s řádky:

browser

track

#

číslo pole	Název pole	definice
1	chrom	Chromosome (e.g. chr3, chrY, chr2_random) or <i>scaffold</i> (e.g. scaffold10671) name
2	chromStart	Start anotace (1. nt na chromozomu je 0)
3	chromEnd	konec anotace
4	name	jméno řádku
5	score	Score 0 - 1000
6	strand	DNA strand orientation (positive ["+"] or negative ["-"] or "." if no strand)
7	thickStart	nastavení zobrazení
8	thickEnd	nastavení zobrazení
9	itemRgb	barva anotace
10	blockCount	počet bloků na řádku
11	blockSizes	velikost bloků - musí korespondovat s 10
12	blockStarts	start bloků

```
browser position chr7:127471196-127495720
browser hide all
track name="ItemRGBDemo" description="Item RGB demonstration" visibility=2 itemRgb="On"
chr7 127471196 127472363 Pos1 0 + 127471196 127472363 255,0,0
chr7 127472363 127473530 Pos2 0 + 127472363 127473530 255,0,0
chr7 127473530 127474697 Pos3 0 + 127473530 127474697 255,0,0
chr7 127474697 127475864 Pos4 0 + 127474697 127475864 255,0,0
chr7 127475864 127477031 Neg1 0 - 127475864 127477031 0,0,255
chr7 127477031 127478198 Neg2 0 - 127477031 127478198 0,0,255
chr7 127478198 127479365 Neg3 0 - 127478198 127479365 0,0,255
chr7 127479365 127480532 Pos5 0 + 127479365 127480532 255,0,0
chr7 127480532 127481699 Neg4 0 - 127480532 127481699 0,0,255
```

VCF (BCF) - Variant Call Format

Kontrolní otázka:

Jaký je rozdíl mezi SNP a SNV?

- Textový soubor, TAB-delimited, rozšíření GFF3
- Primárně pro anotaci včetně varianility (nejčastější SNPs, malé INDELs)

• gVCF

VCF

```
##fileformat=VCFv4.2
##contig=<ID=2,length=51304566>
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
```

} metadata

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE1	SAMPLE2	SAMPLE3	SAMPLE4	SAMPLE5	SAMPLE6	SAMPLE7
2	81170	.	C	T	.	.	AC=9;AN=7424	GT:DP:GQ	0/0:4:12	0/0:3:9	0/1:1:3	0/1:9:24	1/0:4:12	0/0:5:15	0/0:4:12
2	81171	.	G	A	.	.	AC=6;AN=7446	GT:DP:GQ	0/1:4:12	0/0:3:9	0/0:1:3	0/0:9:24	0/1:4:12	0/1:5:15	0/0:4:12
2	81182	.	A	G	.	.	AC=5;AN=7506	GT:DP:GQ	0/0:5:15	0/0:4:12	0/0:5:15	0/0:9:24	0/0:4:12	0/0:4:12	0/0:4:12
2	81204	.	T	G	.	.	AC=2;AN=7542	GT:DP:GQ	1/0:5:15	0/0:9:27	0/0:10:30	0/0:15:39	0/0:9:27	1/0:13:39	0/1:14:42

} data

BCF

2	81170	.	C	T	.	.	AC=9;AN=7424	GT:0/0:0/0:0/1:0/1:1/0:0/0:0/0	DP:4:3:1:9:4:5:4	GQ:12:9:3:24:12:15:12
2	81171	.	G	A	.	.	AC=6;AN=7446	GT:0/1:0/0:0/0:0/0:0/1:0/1:0/0	DP:4:3:1:9:4:5:4	GQ:12:9:3:24:12:15:12
2	81182	.	A	G	.	.	AC=5;AN=7506	GT:0/0:0/0:0/0:0/0:0/0:0/0:0/0	DP:5:4:5:9:4:4:4	GQ:15:12:15:24:12:12:12
2	81204	.	T	G	.	.	AC=2;AN=7542	GT:1/0:0/0:0/0:0/0:0/0:1/0:0/1	DP:5:9:10:15:9:13:14	GQ:15:27:30:39:27:39:42

Jaký je rozdíl mezi SNP a SNV?

- single nucleotide variant (SNV)
- single nucleotide polymorphism (SNP)
- single-base substitute (real-time PCR, microarrays, NGS)
- SNV variatna na úrovni 1 nukleotidu v genomu populace
- SNP variatna na úrovni 1 nukleotidu v genomu zárodečné linie alespoň u 1% populace.

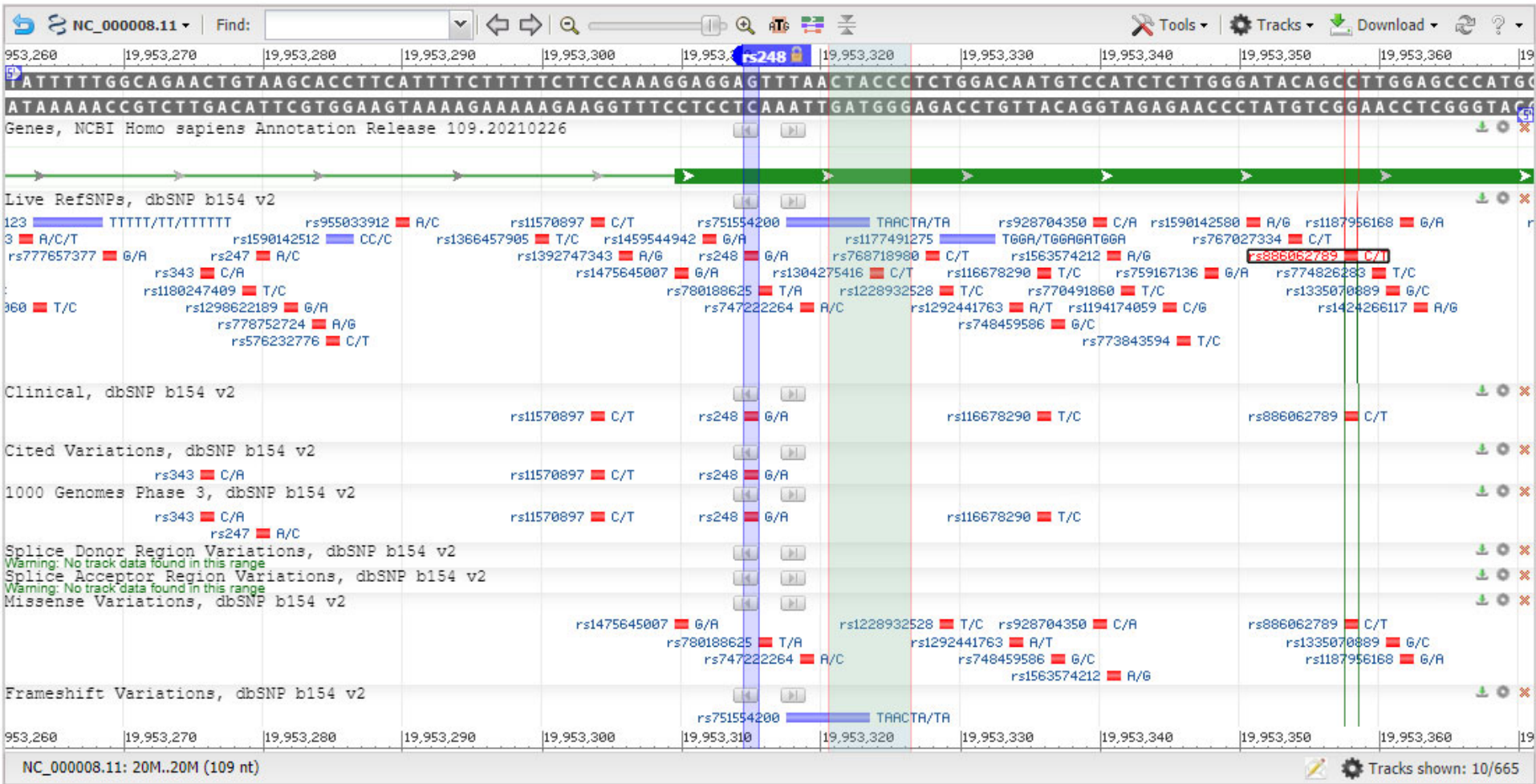
Genomic regions, transcripts, and products

Top ▲ ?

Choose placement

GRCh38.p12 (NC_000008.11)

See rs248 in Variation Viewer



Command line introduction



Type cmd and press enter

```
Command Prompt
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.
C:\Users\21286>
```

dir (+ enter)
cd
cd ..
G:
del
mkdir
rmdir
ipconfig
netstat
ping
systeminfo
cls
color 0A
title

šipky nahoru dolu
tabulator

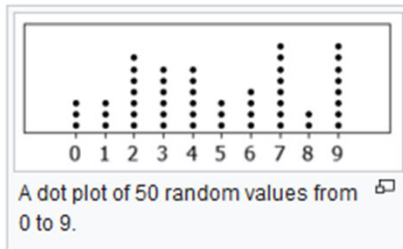
[Is there a Windows
command line with smart
bash-like autocompletions /
command history? - Super
User](#)

key F7

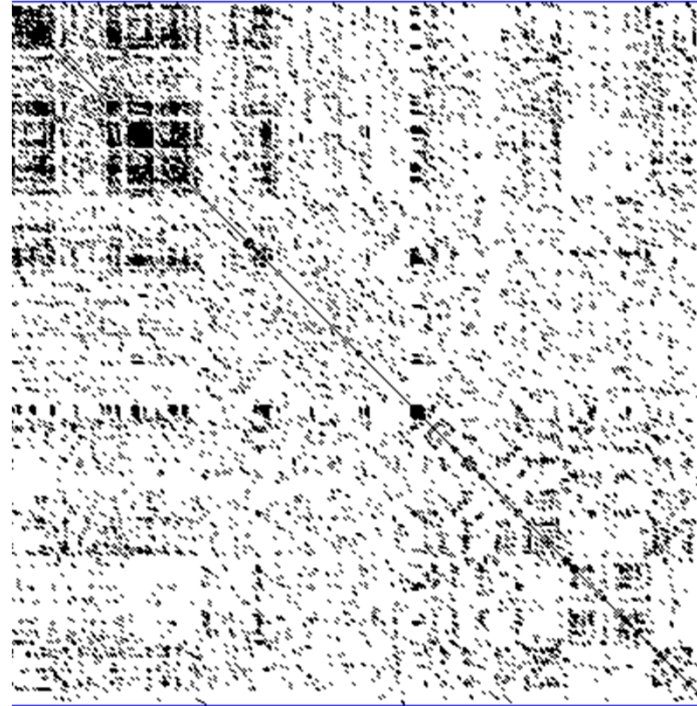
označ text + right-click

Dot plot (dotter)

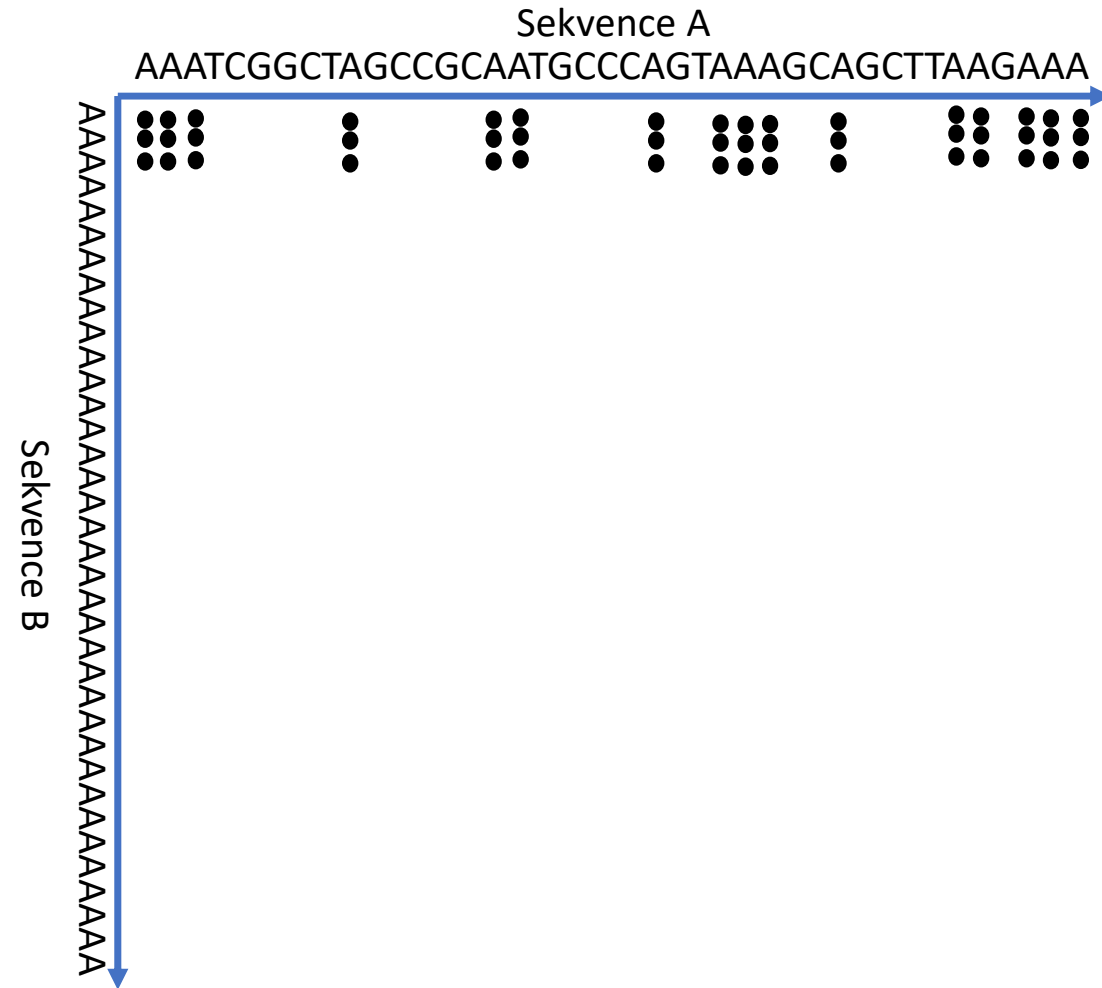
STATISTIKA



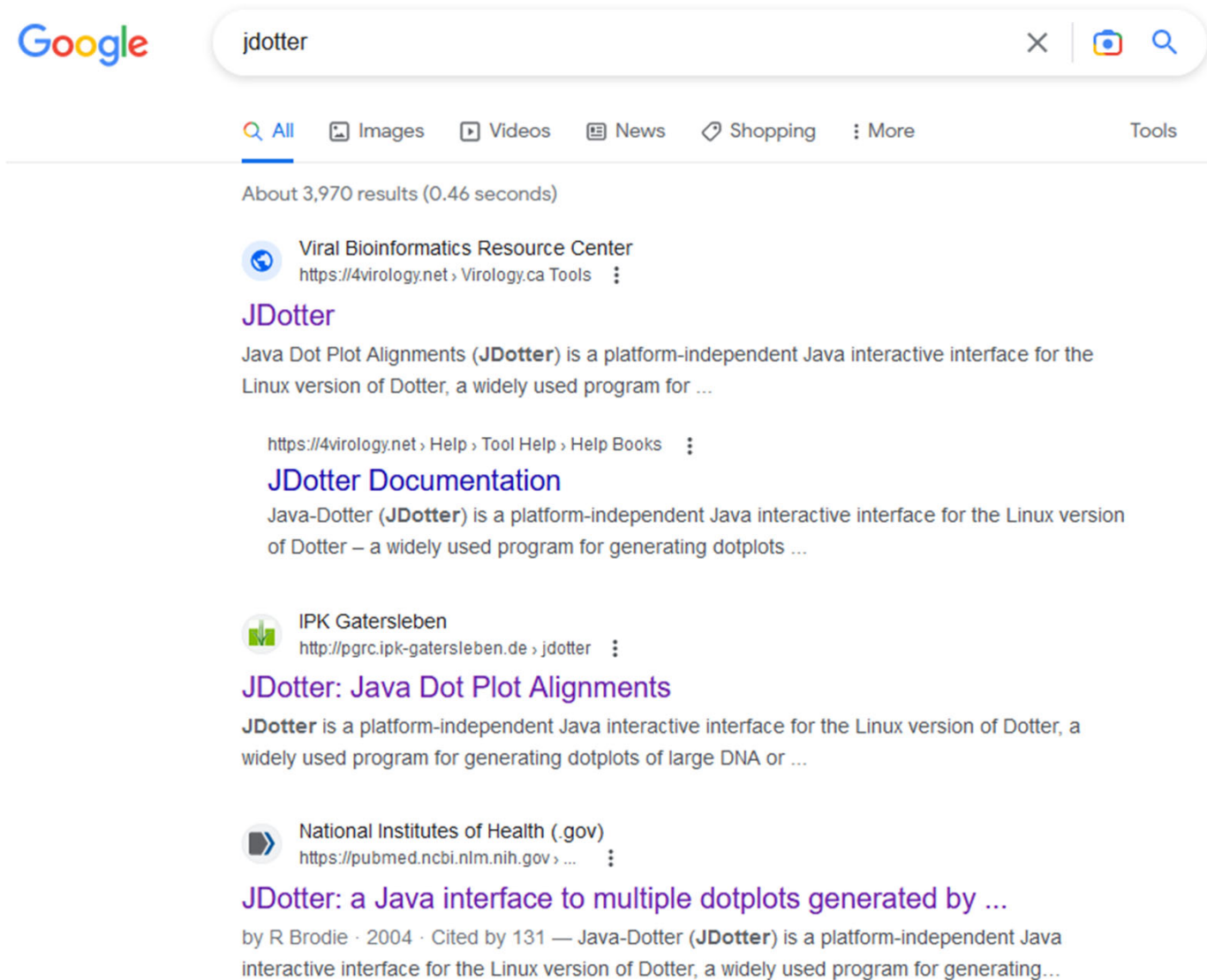
BIOINFORMATIKA



Dotplot



JDotter



The image is a screenshot of a Google search results page for the query "jdotter". The Google logo is in the top left. The search bar contains "jdotter" with a clear button (X) and image search icons. Below the search bar are tabs for "All", "Images", "Videos", "News", "Shopping", "More", and "Tools". The "All" tab is selected. The results show "About 3,970 results (0.46 seconds)".

Result 1:

- Icon: A blue circular icon with a white 'v'.
- Title: Viral Bioinformatics Resource Center
- URL: <https://4virology.net> › Virology.ca Tools
- Snippet: **JDotter**
Java Dot Plot Alignments (**JDotter**) is a platform-independent Java interactive interface for the Linux version of Dotter, a widely used program for ...
- More links: <https://4virology.net> › Help › Tool Help › Help Books

Result 2:

- Icon: A green circular icon with a white 'v'.
- Title: IPK Gatersleben
- URL: <http://pgrc.ipk-gatersleben.de> › jdotter
- Snippet: **JDotter: Java Dot Plot Alignments**
JDotter is a platform-independent Java interactive interface for the Linux version of Dotter, a widely used program for generating dotplots of large DNA or ...

Result 3:

- Icon: A blue circular icon with a white 'v'.
- Title: National Institutes of Health (.gov)
- URL: <https://pubmed.ncbi.nlm.nih.gov> › ...
- Snippet: **JDotter: a Java interface to multiple dotplots generated by ...**
by R Brodie · 2004 · Cited by 131 — Java-Dotter (**JDotter**) is a platform-independent Java interactive interface for the Linux version of Dotter, a widely used program for generating...

BLAST

 **U.S. National Library of Medicine**
National Center for Biotechnology Information

BLAST[®] » blastn suiteHome

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file Nevybrán žádný soubor [?](#)

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 Optional ☐ exclude [Add organism](#)

Virtualizace – cesta k linuxu (windows)

The image shows the Oracle VM VirtualBox logo, which consists of a blue cube with the letters 'VM' on its top face and 'VirtualBox' on its side. To the right of the cube, the word 'VirtualBox' is written in a large, bold, sans-serif font. Above the cube, the word 'ORACLE' is written in a smaller, red, sans-serif font, and below it, 'VM' is written in a smaller, black, sans-serif font.

Linux Basics: Convert IMG Files To VDI ...
unixmen.com

Unixmen

Linux Basics: Convert IMG Files To VDI Format For Oracle Virtual...

Images may be subject to copyright. [Learn More](#)

Related images

[See more](#)

The image shows the Oracle VM VirtualBox logo, which consists of a blue cube with the letters 'VM' on its top face and 'VirtualBox' on its side. To the right of the cube, the word 'VirtualBox' is written in a large, bold, sans-serif font. Above the cube, the word 'ORACLE' is written in a smaller, red, sans-serif font, and below it, 'VM' is written in a smaller, black, sans-serif font.

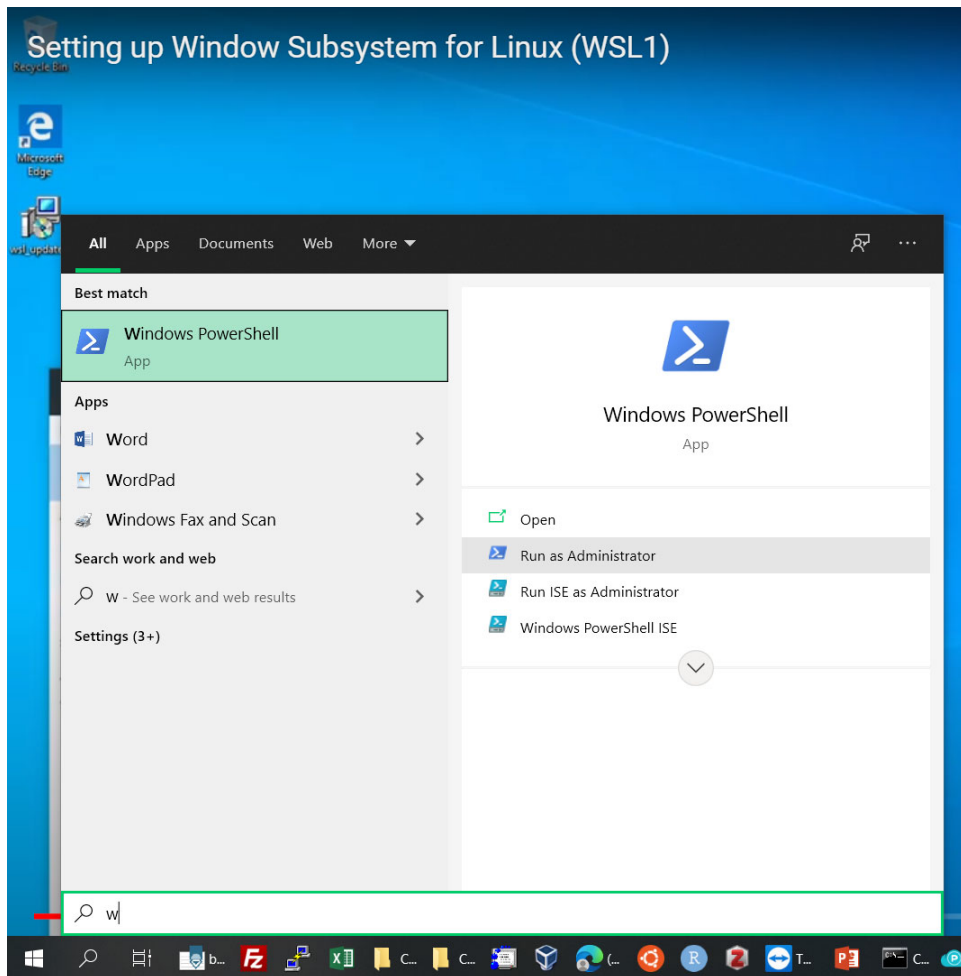
The image shows the Oracle VM VirtualBox logo, which consists of a blue cube with the letters 'VM' on its top face and 'VirtualBox' on its side. To the right of the cube, the word 'VirtualBox' is written in a large, bold, sans-serif font. Above the cube, the word 'ORACLE' is written in a smaller, red, sans-serif font, and below it, 'VM' is written in a smaller, black, sans-serif font.

The image shows the Oracle VM VirtualBox logo, which consists of a blue cube with the letters 'VM' on its top face and 'VirtualBox' on its side. To the right of the cube, the word 'VirtualBox' is written in a large, bold, sans-serif font. Above the cube, the word 'ORACLE' is written in a smaller, red, sans-serif font, and below it, 'VM' is written in a smaller, black, sans-serif font.

The image shows a Windows taskbar with several open applications. From left to right, the icons are: a file explorer window showing 'C:\...', a folder icon, a 'virt...' window, a terminal window, a 'me...' window, a 'Co...' window, a 'Scr...' window, and a help window. The taskbar also shows the system tray with icons for network, volume, and power.

CES 10:36 PM
US 4/25/2021

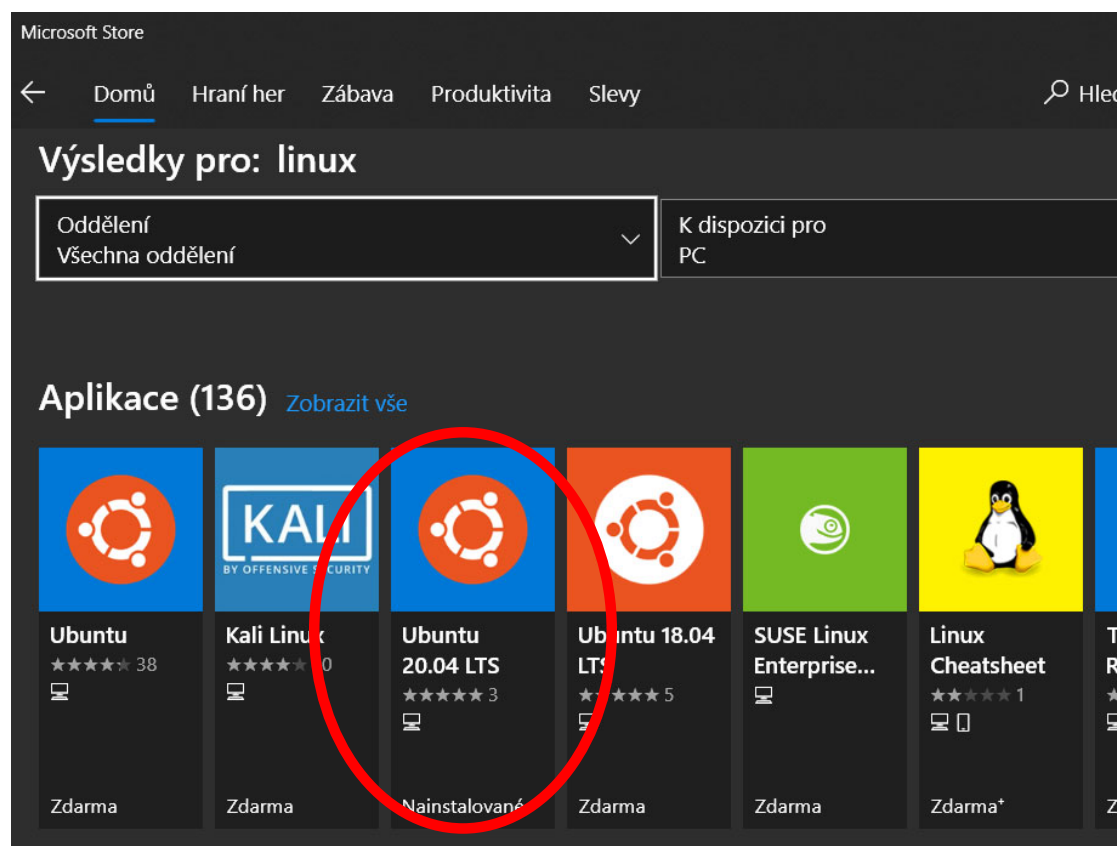
WSL – window for subsystem linux



A screenshot of the Windows System Information window. The window title is 'System Information'. The left sidebar shows a tree view with 'System Summary' selected. The main area displays a table of system information.

Item	Value
OS Name	Microsoft Windows 10 Pro
Version	10.0.19042 Build 19042
Other OS Description	Not Available
OS Manufacturer	Microsoft Corporation
System Name	ZAM-71200803-NB
System Manufacturer	Hewlett-Packard
System Model	HP ProBook 650 G1
System Type	x64-based PC
System SKU	H5G79EA#BCM
Processor	Intel(R) Core(TM) i5-4200M CPU @ 2.50GHz, 2
BIOS Version/Date	Hewlett-Packard L77 Ver. 01.43, 1/25/2018
SMBIOS Version	2.7
Embedded Controller V...	22.60
BIOS Mode	Legacy
BaseBoard Manufacturer	Hewlett-Packard
BaseBoard Product	1993
BaseBoard Version	KBC Version 16.3C
Platform Role	Mobile
Secure Boot State	Unsupported
PCR7 Configuration	Binding Not Possible
Windows Directory	C:\WINDOWS
System Directory	C:\WINDOWS\system32
Boot Device	\Device\HarddiskVolume1
Locale	United States
Hardware Abstraction L...	Version = "10.0.19041.906"
User Name	ZAM-71200803-NB\21286
Time Zone	Central Europe Daylight Time

nainstalovat aplikaci (virtuální systém)



Ubuntu 20.04 LTS
download VS + instalace VS + restart celého počítače

MetaCentrum

The screenshot shows the MetaCentrum website in a web browser. The browser's address bar displays the URL: `https://metavo.metacentrum.cz/osobniv3/wayf/proxy.jsp?locale=cs&target=https%3A%2F%2Fperun.cesnet.cz%2Ffed%2Fregistrar%2F%3Fvo...`. The browser's bookmark bar includes links to 'investice', 'aukce', 'cloud', 'knihy', 'Google', 'Editorial System - Y...', 'Third Call | EASI-Ge...', 'Internet věcí je tady...', 'ISSN 1802-2685', 'Bookmarks', and 'ISISApp'.

The website's header features the 'e-Infrastruktura CESNET' logo and a navigation menu with links: 'Síť', 'Výpočty', 'Úložiště', 'Spolupráce', 'Multimédia', 'Bezpečnost', and 'Identita'. Below the header, the 'metacentrum' logo is visible.

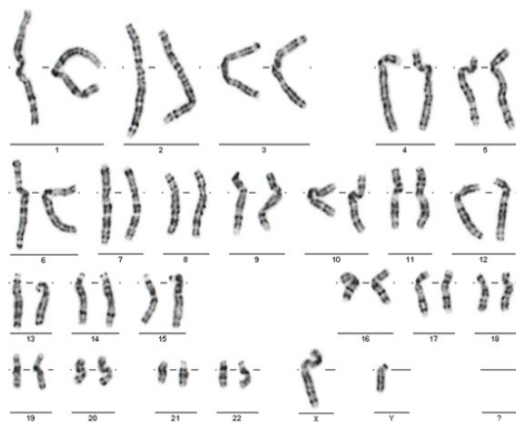
The main content area is titled 'Odkud přicházíte ?'. It contains a paragraph explaining the registration process for academic staff, employees, and students of research institutions in the Czech Republic. The text mentions the use of the 'eduid.cz' system and the role of the 'CESNET' organization in verifying academic identity.

On the left side, there is a vertical menu with the following items: 'O MetaCentru VO', 'Aktuality', 'Dokumentace a služby', 'Přihláška', 'Můj účet', 'Stav zdrojů', 'Uživatelská podpora', 'Semináře', 'Mapa portálu', and 'Interní část'. Below this menu is a search bar with the text 'Hledat'.

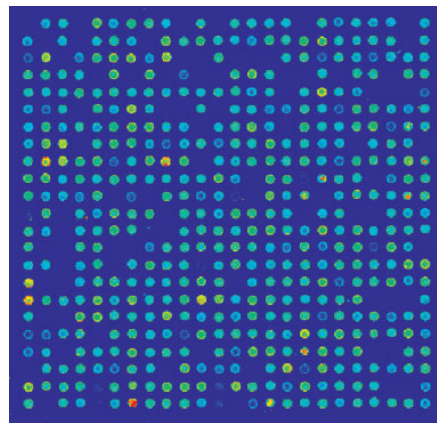
The main content area also features two large boxes. The left box, titled 'eduid.cz', contains the text 'Mám účet v organizaci zapojené do eduid.cz'. The right box, titled '?', contains the text 'Moje organizace není v eduid.cz a potřebuji ověřit svoji náhradní identitu'.

At the bottom, there is a section titled 'Přímé odkazy na přihlášení přes vybrané instituce z eduid.cz'. This section includes logos and names for 'MUNI Masarykova univerzita', 'Univerzita Karlova', and 'Západočeská univerzita v Plzni'.

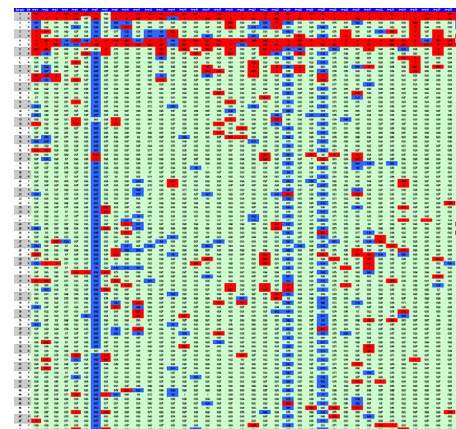
Prenatální diagnostika



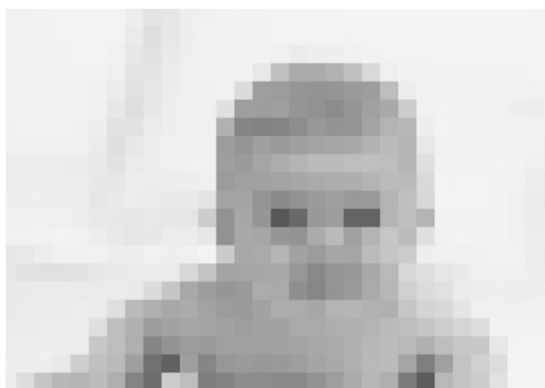
Klasický karyotyp (rozlišení ~ 5 Mb)



Array-CGH (rozlišení ~ 0,1 Mb)

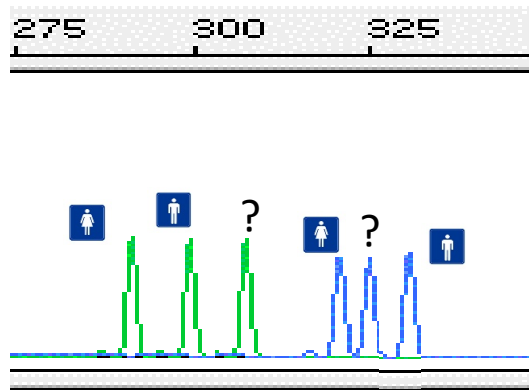


NGS (rozlišení- jednotlivé báze)



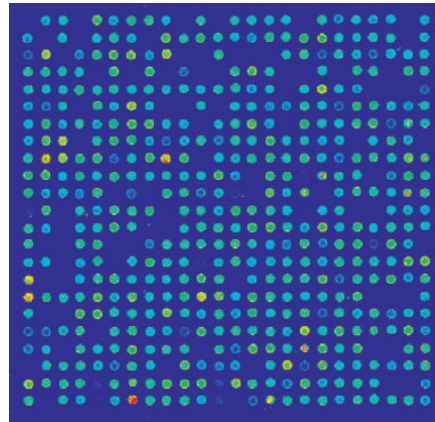
Doba odezvy (TAT)

v den odběru



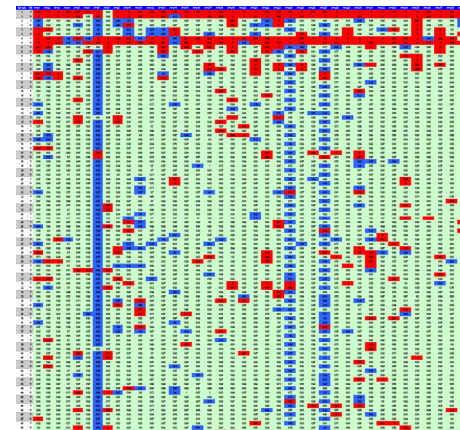
QF-PCR

1-2 týdny

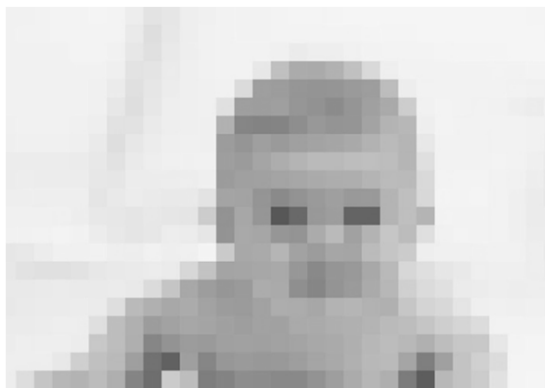


Array-CGH (rozlišení ~ 0,1 Mb)

2-3 týdny

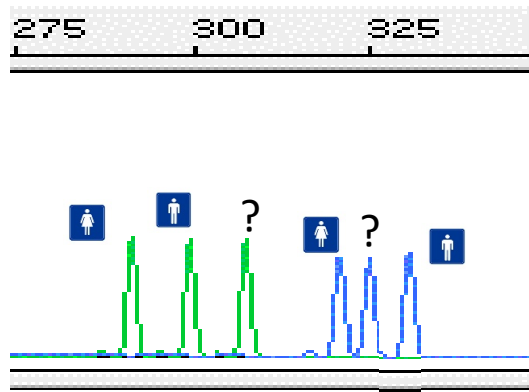


NGS (rozlišení- jednotlivé báze)



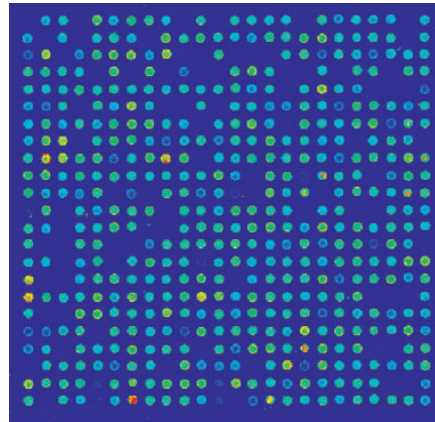
Cena \$\$\$

\$



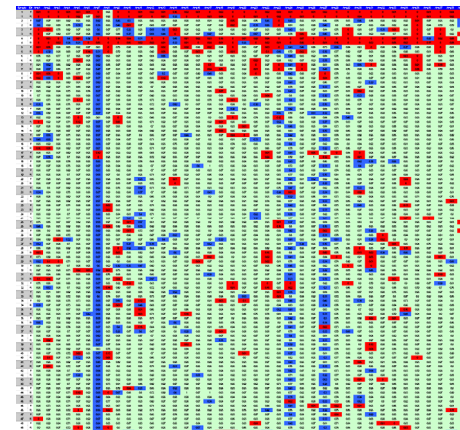
QF-PCR

\$\$

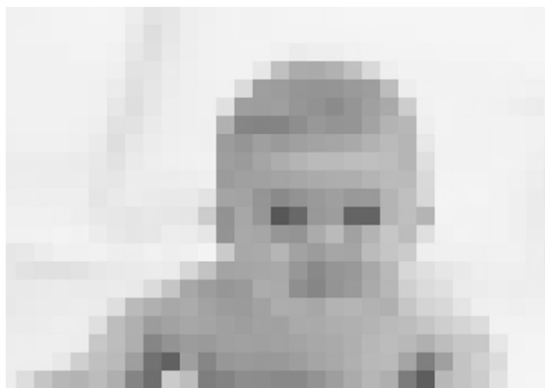


Array-CGH (rozlišení ~ 0,1 Mb)

\$\$\$

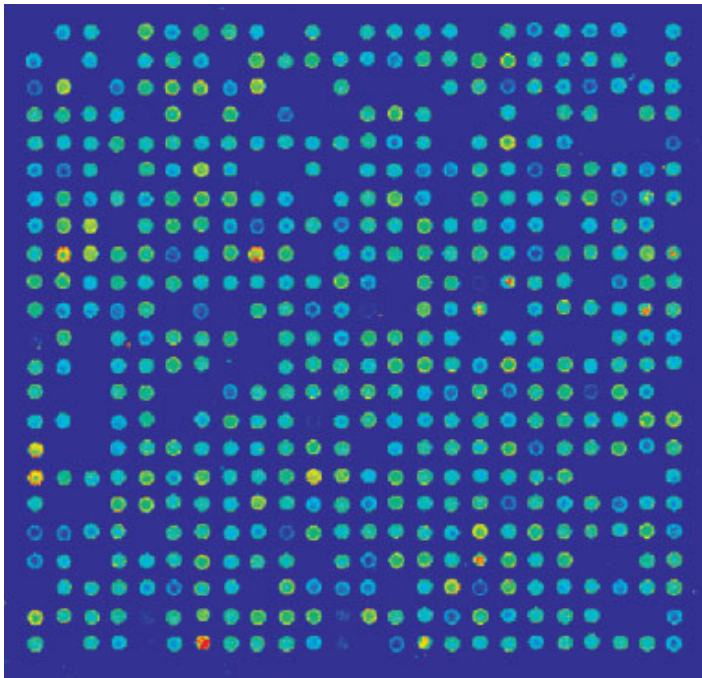


NGS (rozlišení- jednotlivé báze)

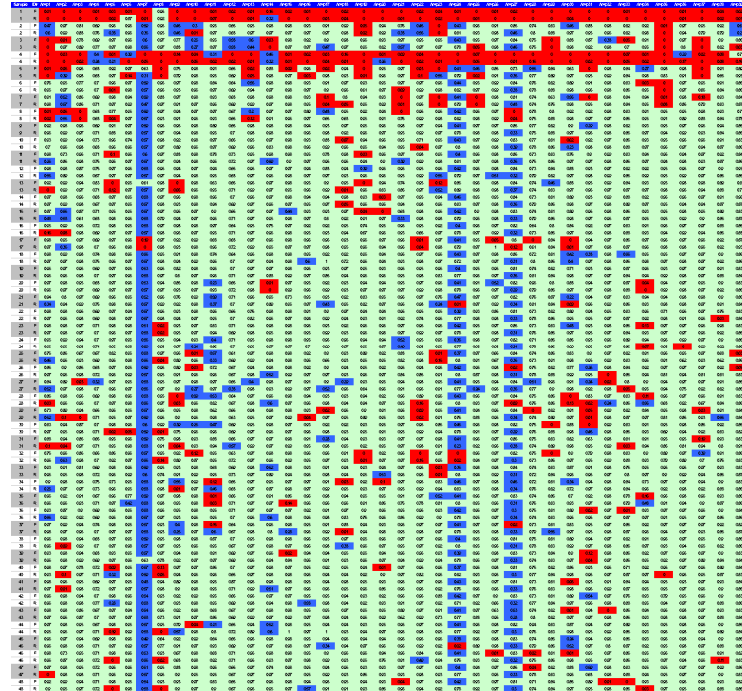


„Netušíme, kde a co hledat“

array CGH



NGS



NGS v prenatální diagnostice

- **Panely genů pro specifické fenotypové skupiny**
 - např. vrozené srdeční vady, RASopatie, skeletální vady
- WES (popř. WGS) – u nás Heredity panel (klinický exom 3332 genů)
- trend WES + CNV (2v1)
- **NIPT** – neinvazivní prenatální testování

RASopatie – sy Noonanové

Syndrom Noonanové

Noonan-like (CBL syndrom)

Noonan-like sy with loose anagen hair

Costello syndrom

Hereditární gingivální fibromatóza

Jaffe Campanacci syndrom

Kardio-facio-kutánní syndrom (CFC)

Legius syndrom (NF1-like)

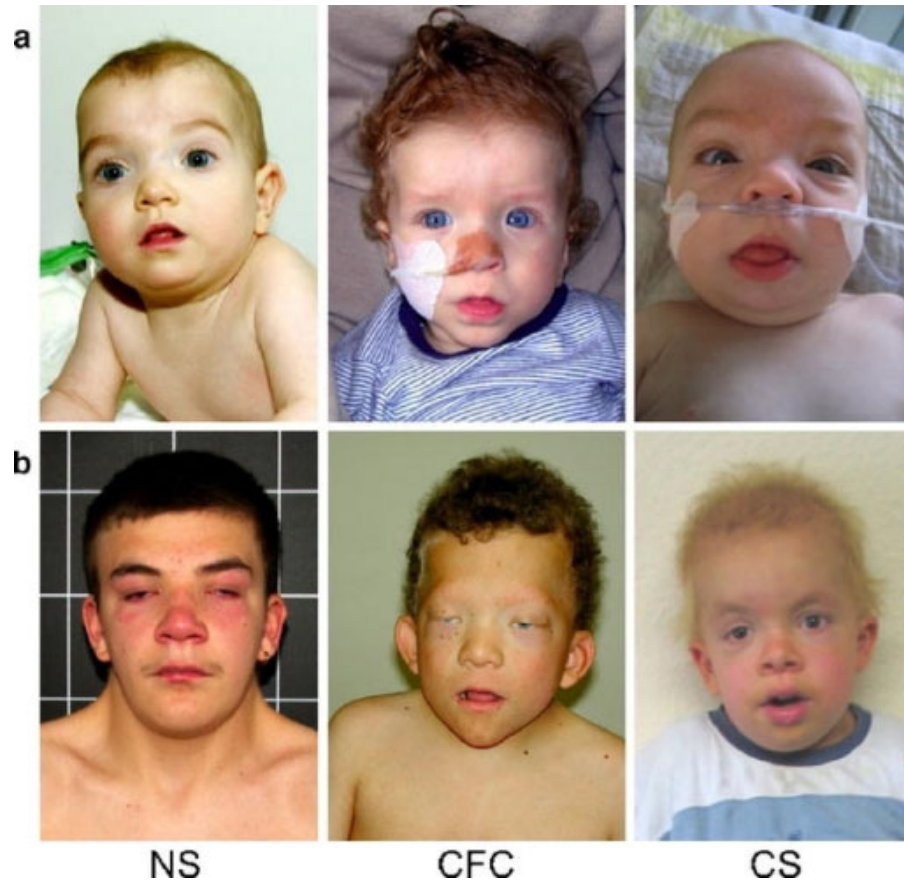
LEOPARD syndrom (Multiple lentigines)

Neurofibromatóza typu 1

Neurofibromatóza typu 2

Neurofibromatosa-Noonan (NFNS)

Watson syndrom



Kratz *et al.*, J Mol Med. **2007** Mar; 85(3): 227–235.

RASopatie – sy Noonanové

Rasopathy genes	
<i>PTPN11</i>	<i>SHOC2</i>
<i>SOS1</i>	<i>A2ML1</i>
<i>RAF1</i>	<i>LZTR1</i>
<i>RIT1</i>	<i>RASA2</i>
<i>BRAF</i>	<i>SOS2</i>
<i>KRAS</i>	<i>MAP2K2</i>
<i>NRAS</i>	<i>HRAS</i>
<i>MAP2K1</i>	<i>SPRED1</i>
<i>RRAS</i>	<i>NF1</i>
<i>CBL</i>	<i>NF2</i>

Indikace pro prenatální diagnostiku of RASopatií:

- zvýšené šíjové projasnění (NT) nebo cystická hygroma (karyotyp, aCGH v normě) v kombinaci s:
 - » *hydrops fetalis*
 - » srdeční anomálie
 - » polyhydramnion a/nebo pleurální výpotek
 - » specifické faciální anomálie (hypertelorismus and micrognathia)
 - » ledvinové anomálie

RA Sopaťie – sy Noonanov 

Klinick  popis

matka 32 let

UZ:

- 13/14/16/19 tg NT:2,3/3,1/1,8 mm
- 14/16 tg jugular lymfatick  vaky
- 19 tg anom lie ren ln ho duplexu a hypertrofie prav  komory



Noonan syndrome

IVF

13 gw - NT 6 mm

17 gw - total fetal hydrops with subcutaneous infiltration, ascites

Mutation in *RIT1*: c.319A>G, p.Met107Val *de novo*

***RIT1* (*RIT1*)** ~5 % cases (Aoki *et al.*, 2016), causal gene since 2013

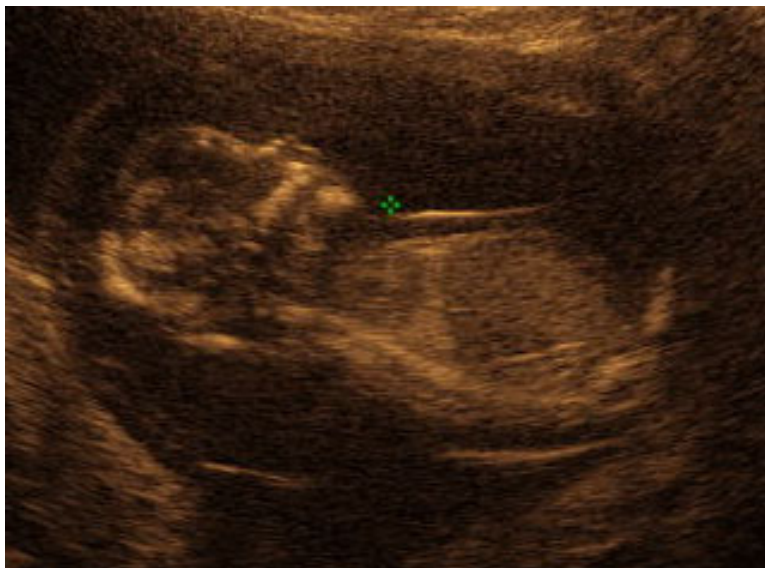
Milosavljević *et al.*, 2016:

UZ: zvišene NT, lymfatické vaky, duplikace renálneho systému, polyhydramnion, hydrops s lymfodémom, hydrothorax, ascites.

Ve 26. t. g. intrauterinní úmrtí plodu.

Pitva plodu prokázala edém, kraniofaciální abnormality charakteristické pro sy Noonanové (hypertelorismus, nízko posazené uši, široký krk).

- 14 gw - NT 8 mm, lymphatic sacs, hygroma colli, renal pelvis bilat 2.5 mm, susp. CHD
- 16 gw - NT 5 mm, lymphatic sacs, agenesis ductus venosus,
- 19 gw NT 4 mm, dilatation of renal pelvis 8 mm, hypertelorism, low set earlobes, CHD
- ***RAF1* c.770C>T *de novo*** (several publications)



RASopatie – sy Noonanové

Molekulární diagnostika:

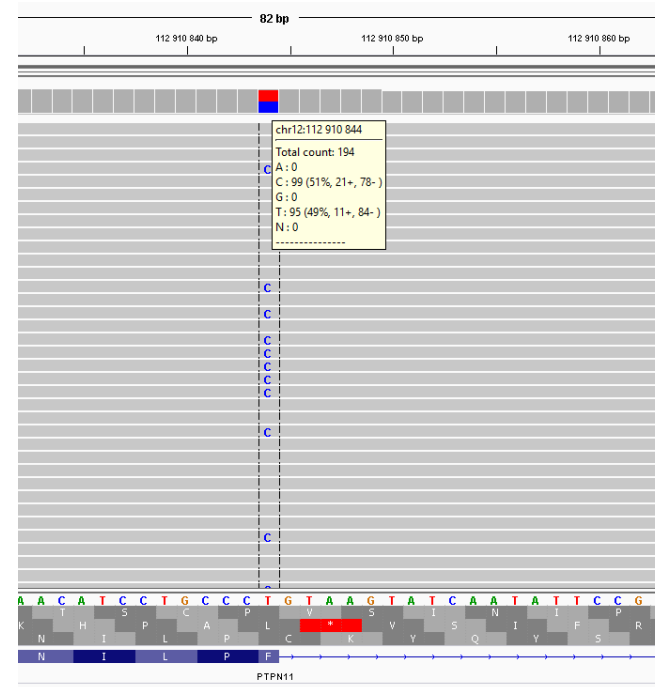
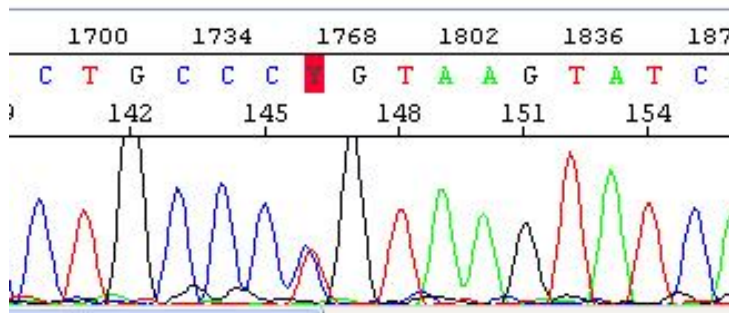
QF-PCR norma (13, 18, 21, X a Y)

Karyotyp norma

array-CGH – žádná patogenní varianta



RASopatie NGS panel

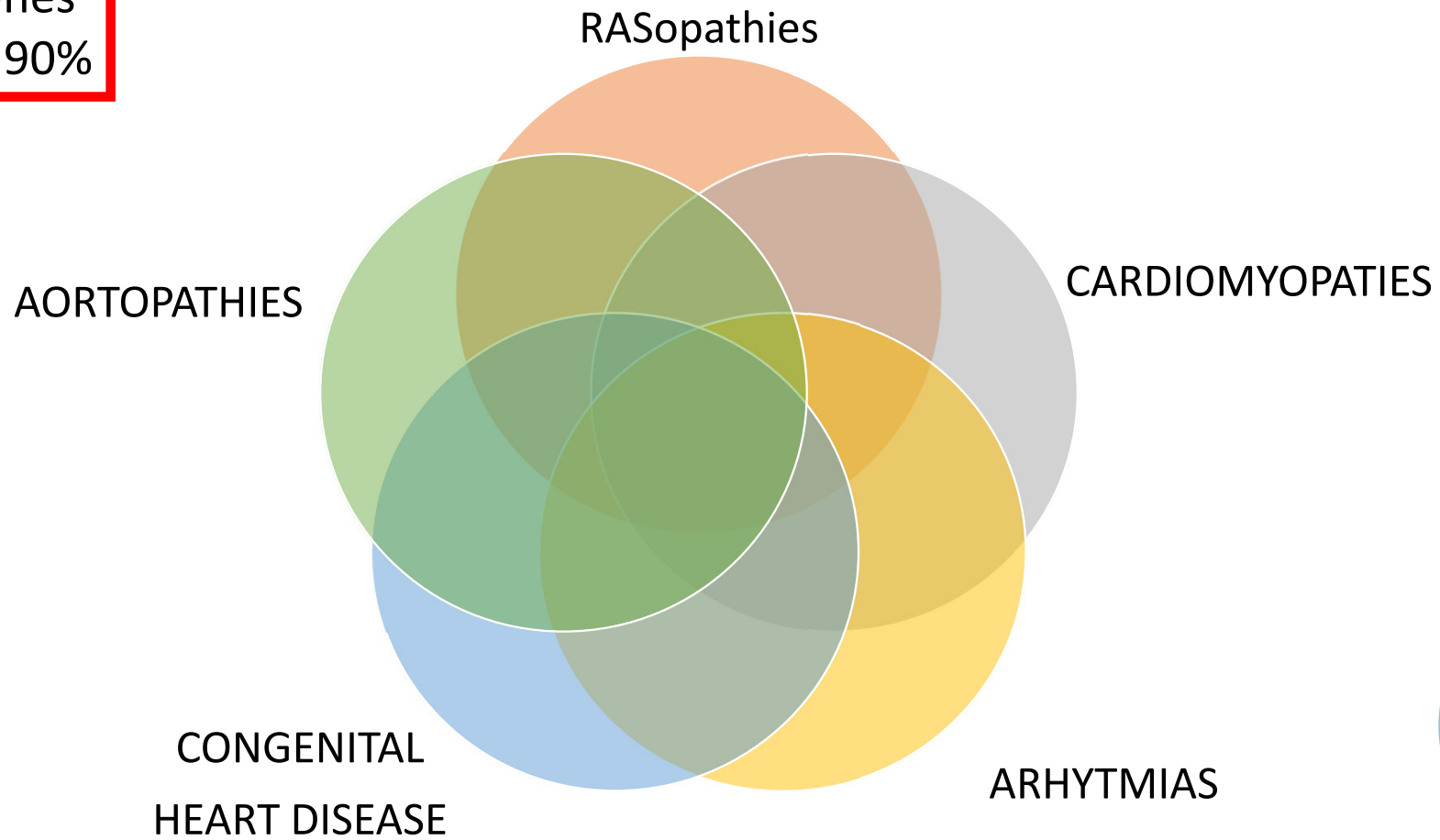


Kauzální mutace:

PTPN11: c.853T>C (p.Phe285Leu) – de novo

Cardio panel

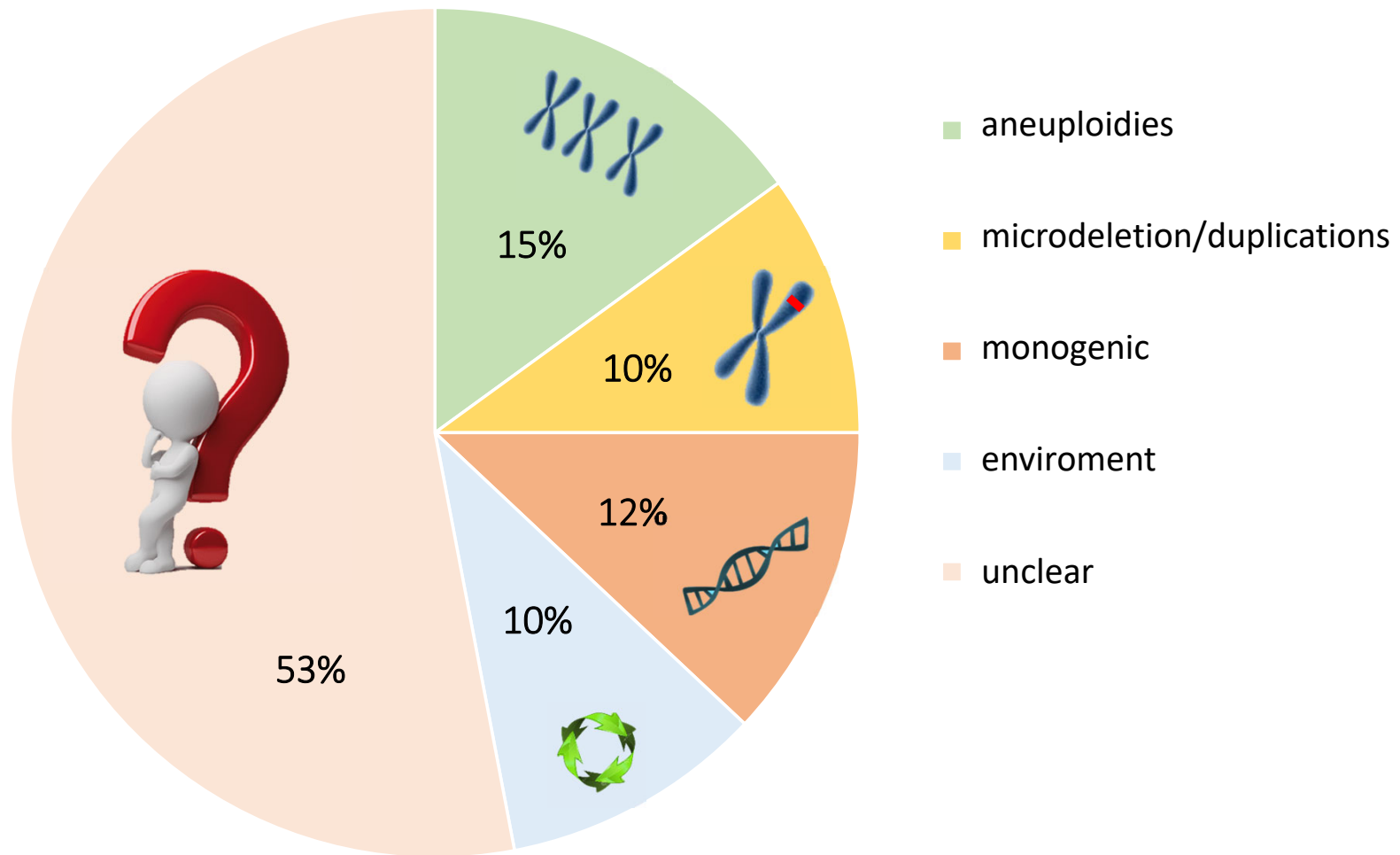
230 genes
CHD ~ 90%



RASopathies

21 genes

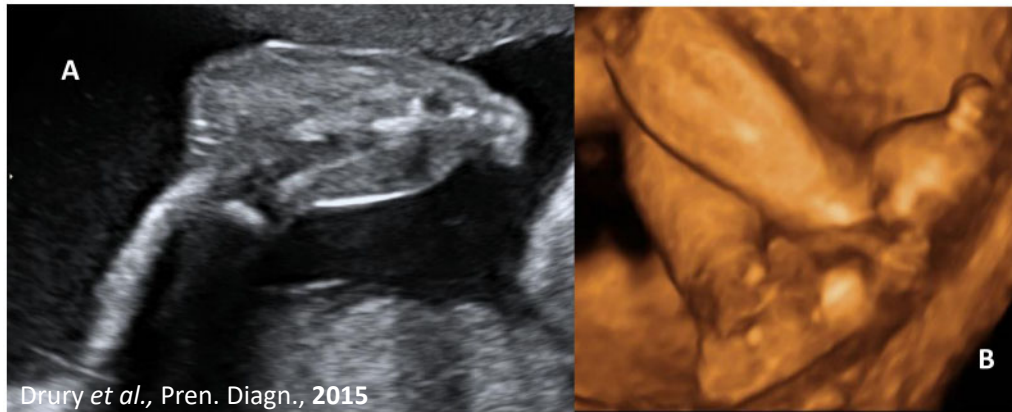
Ethiology of CHD



WES – celoexomové sekvenování



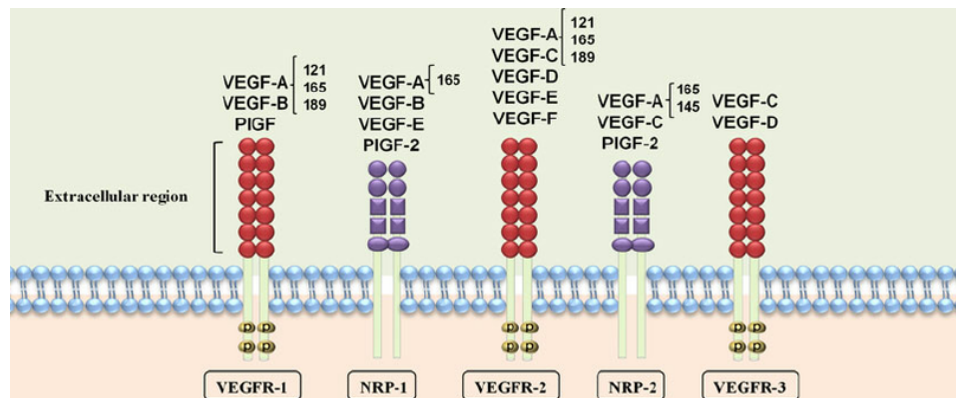
WES – celoexomové sekvenování



Mutace v genu **FLT4** c.3075G>A p.(Met1025Ile)

vascular endothelial growth factor 3

- regulace vývoje struktur lymfatického systému

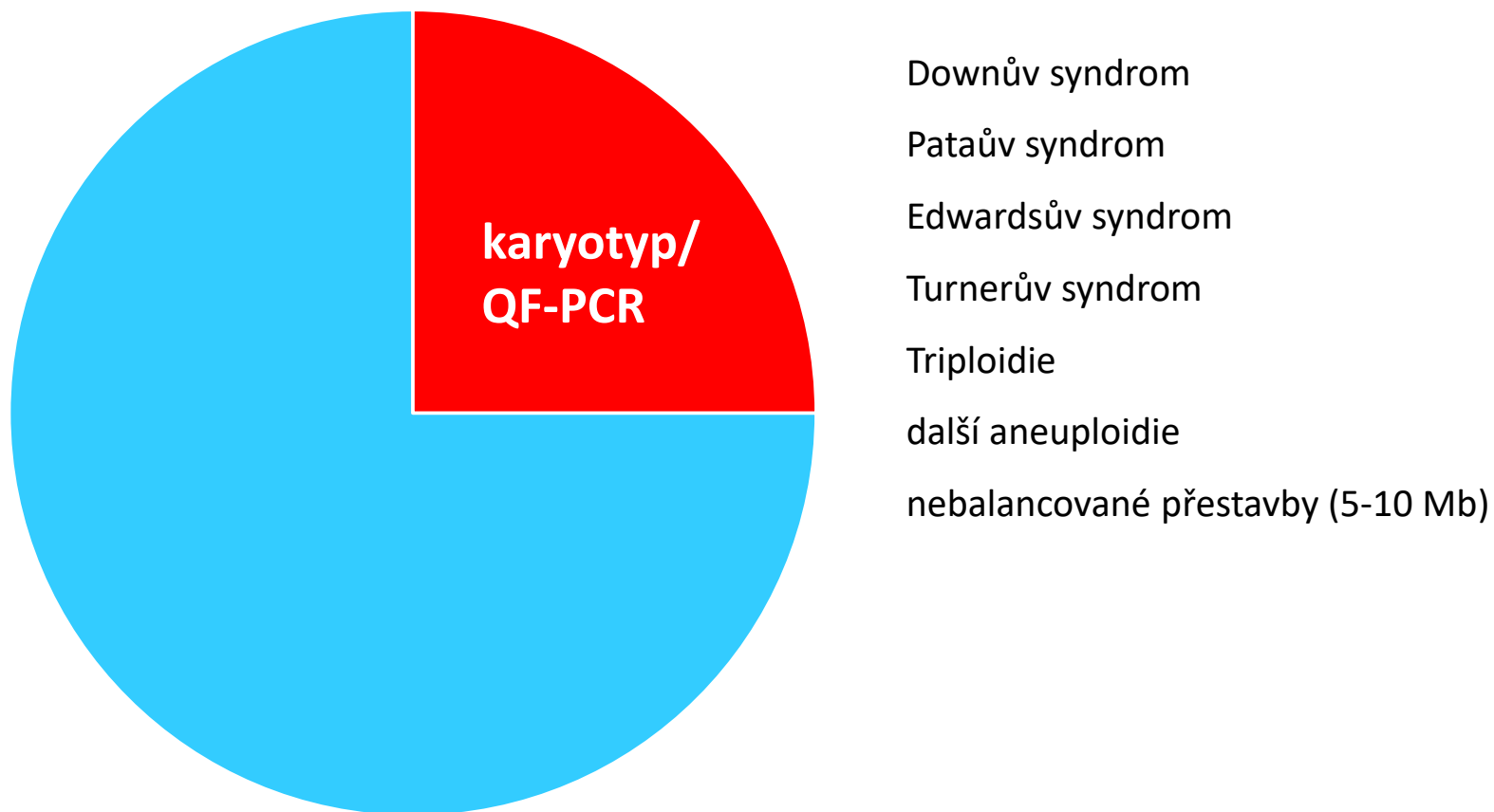


Milroy syndrom



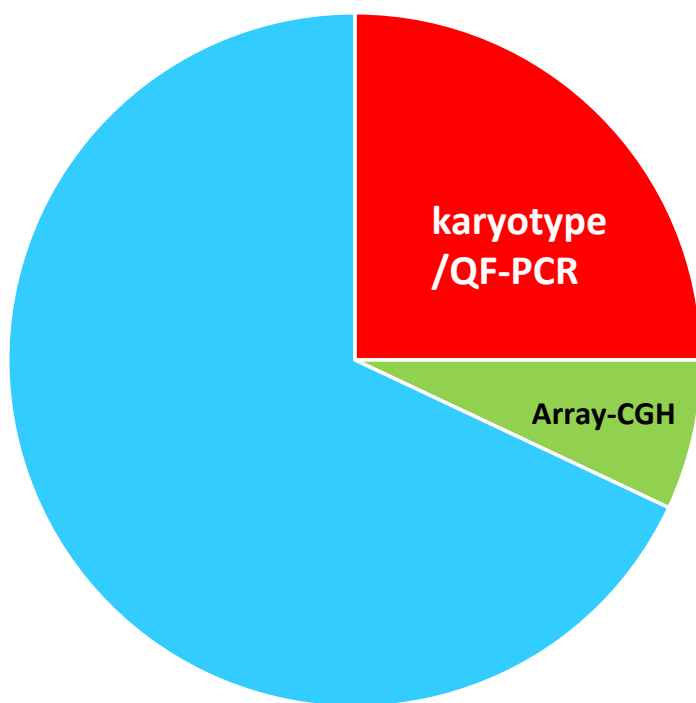
KARYOTYP/QF-PCR

diagnostická výtěžnost



KARYOTYP/QF-PCR + array-CGH

diagnostická výtěžnost



Downův syndrom

Patauův syndrom

Edwardsův syndrom

Turnerův syndrom

Triploidie

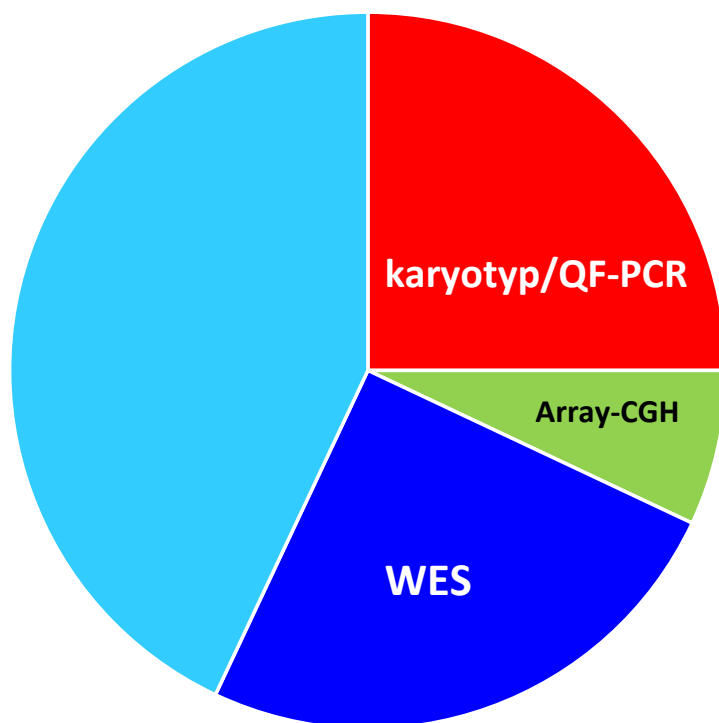
další aneuploidie

nebalancované přestavby (5-10 Mb)

Mikrodelece a mikroduplikace (10 kb)

UPD, AOH

KARYOTYP/QF-PCR + array-CGH + WES diagnostická výtěžnost



Downův syndrom

Patauův syndrom

Edwardsův syndrom

Turnerův syndrom

Triploidie

další aneuploidie

nebalancované přestavby (5-10 Mb)

Mikrodelece a mikroduplikace (10 kb)

UPD, AOH

Monogenní nemoci

Interpretace SNP variant

© American College of Medical Genetics and Genomics

ACMG STANDARDS AND GUIDELINES

**Genetics
in Medicine**

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD¹, Nazneen Aziz, PhD^{2,16}, Sherri Bale, PhD³, David Bick, MD⁴, Soma Das, PhD⁵, Julie Gastier-Foster, PhD^{6,7,8}, Wayne W. Grody, MD, PhD^{9,10,11}, Madhuri Hegde, PhD¹², Elaine Lyon, PhD¹³, Elaine Spector, PhD¹⁴, Karl Voelkerding, MD¹³ and Heidi L. Rehm, PhD¹⁵; on behalf of the ACMG Laboratory Quality Assurance Committee

Databáze

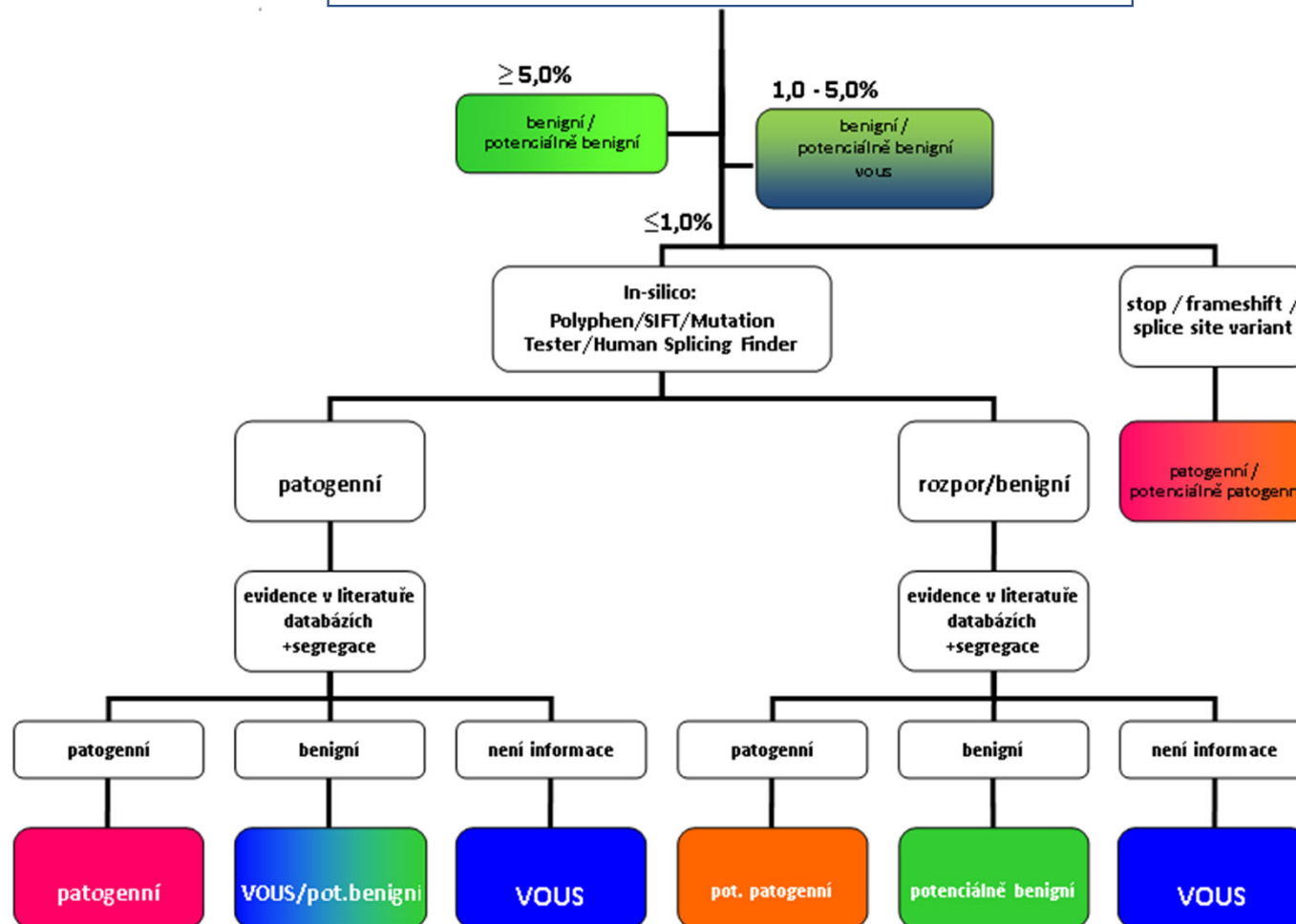
- aktuální verze!!! (hg38 a GRCh38)
- UCSC, NCBI, Ensembl - Genome Browser

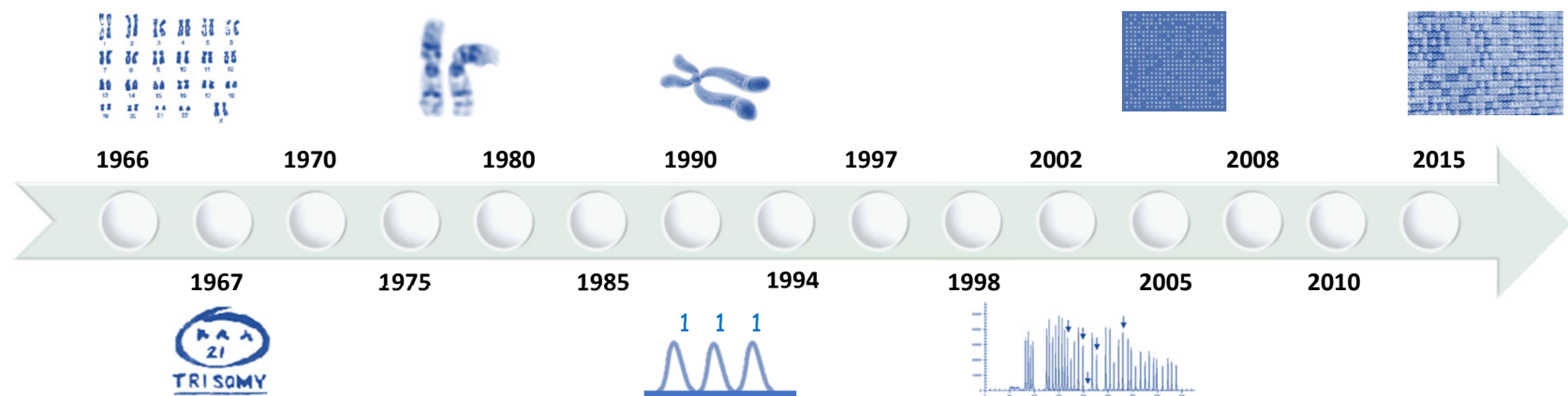
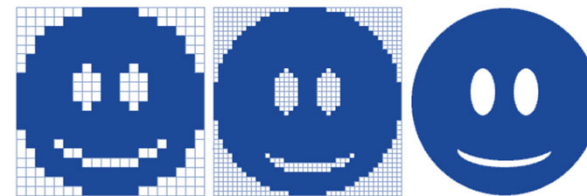
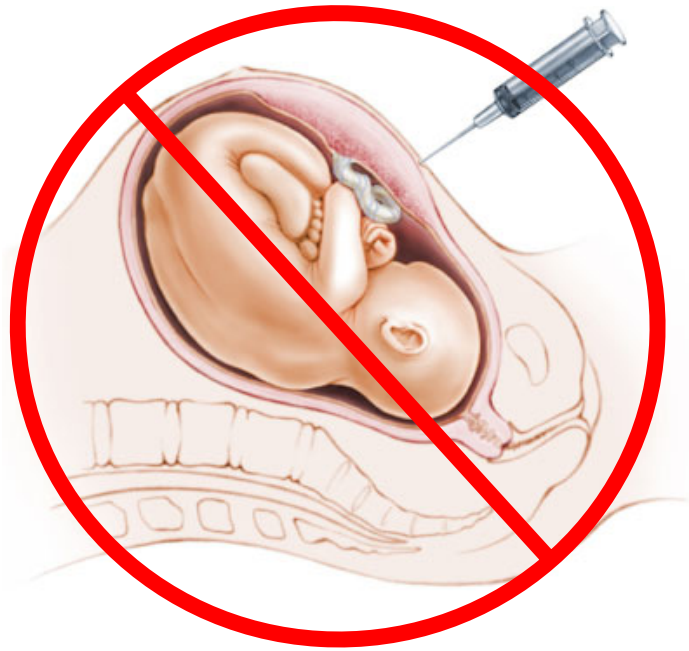


- LOVD
- HGMD
- ClinVar
- OMIM + dbSNP
- UniProt



Klasifikace variant





Další zdroje fetálního materiálu pro prenatální diagnostiku

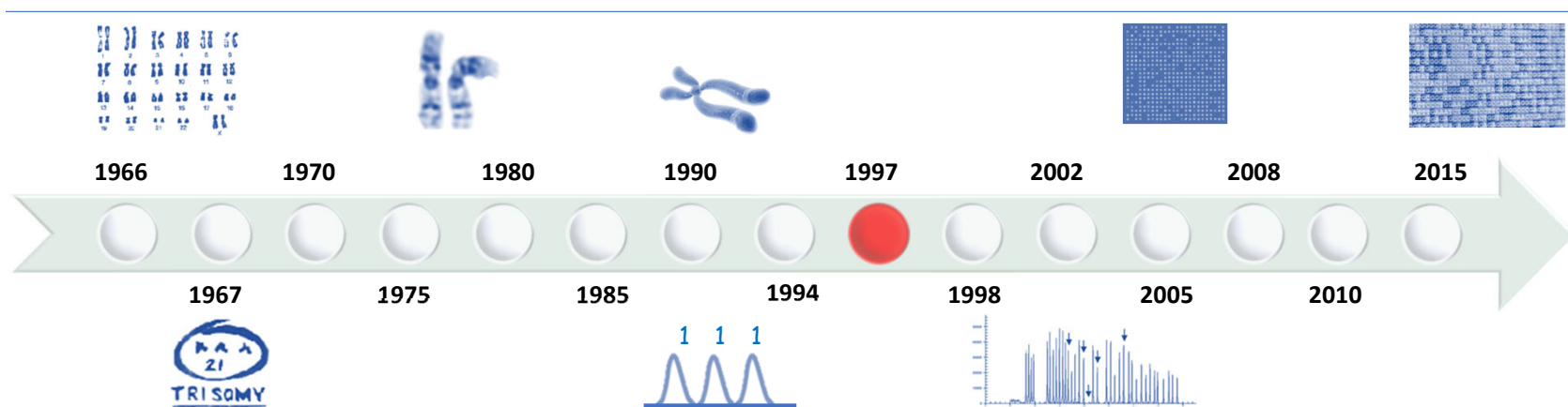
Volná fetální DNA v maternální cirkulaci

Detekovatelná od 5. tg

Po porodu mizí za 30-60 minut

3 – 25 % celkové cirkulující DNA

Původ z trofoblastu



Volná fetální DNA cirkulující v krvi matky

určení pohlaví

- pro X vázaná onemocnění, př. hemofilie, DMD
- při nejasnosti genitálu na UZ

RhD genotypizace plodu

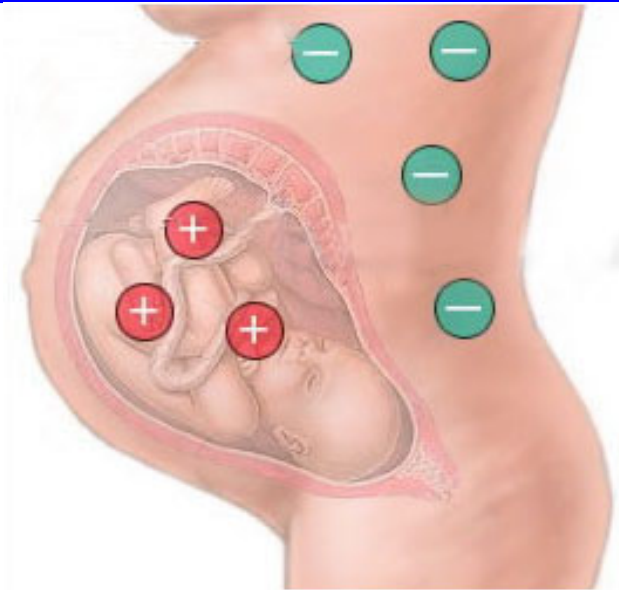
určování otcovství

monogenní onemocnění

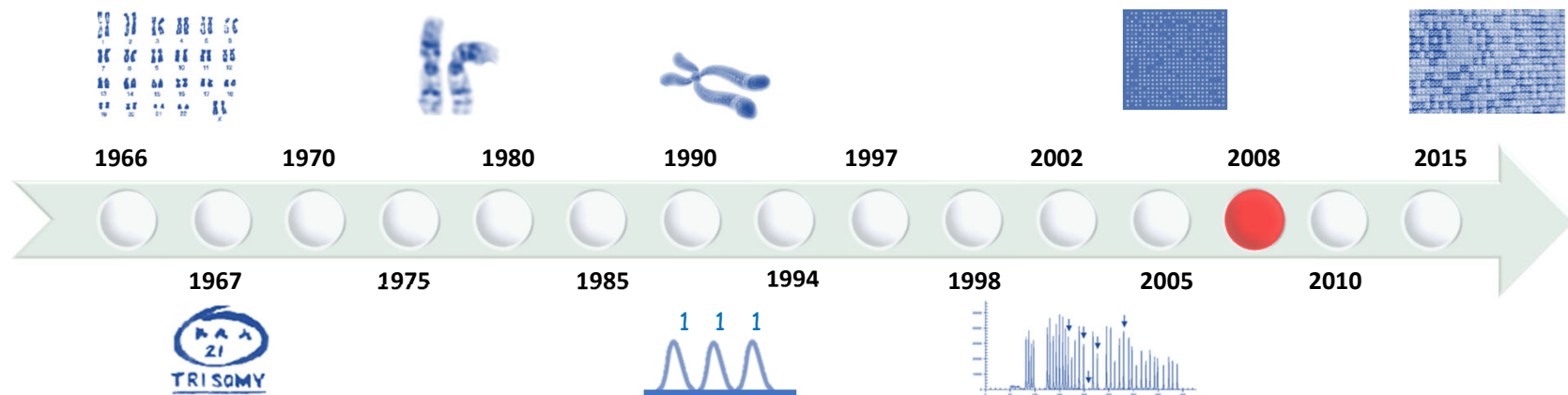
AD: achondroplázie, thanatoformní dysplázie, Huntingtonova chorea, myotonická dystrofie

AR: beta-talasemie, cystická fibróza, kongenitální adrenální hyperplázie

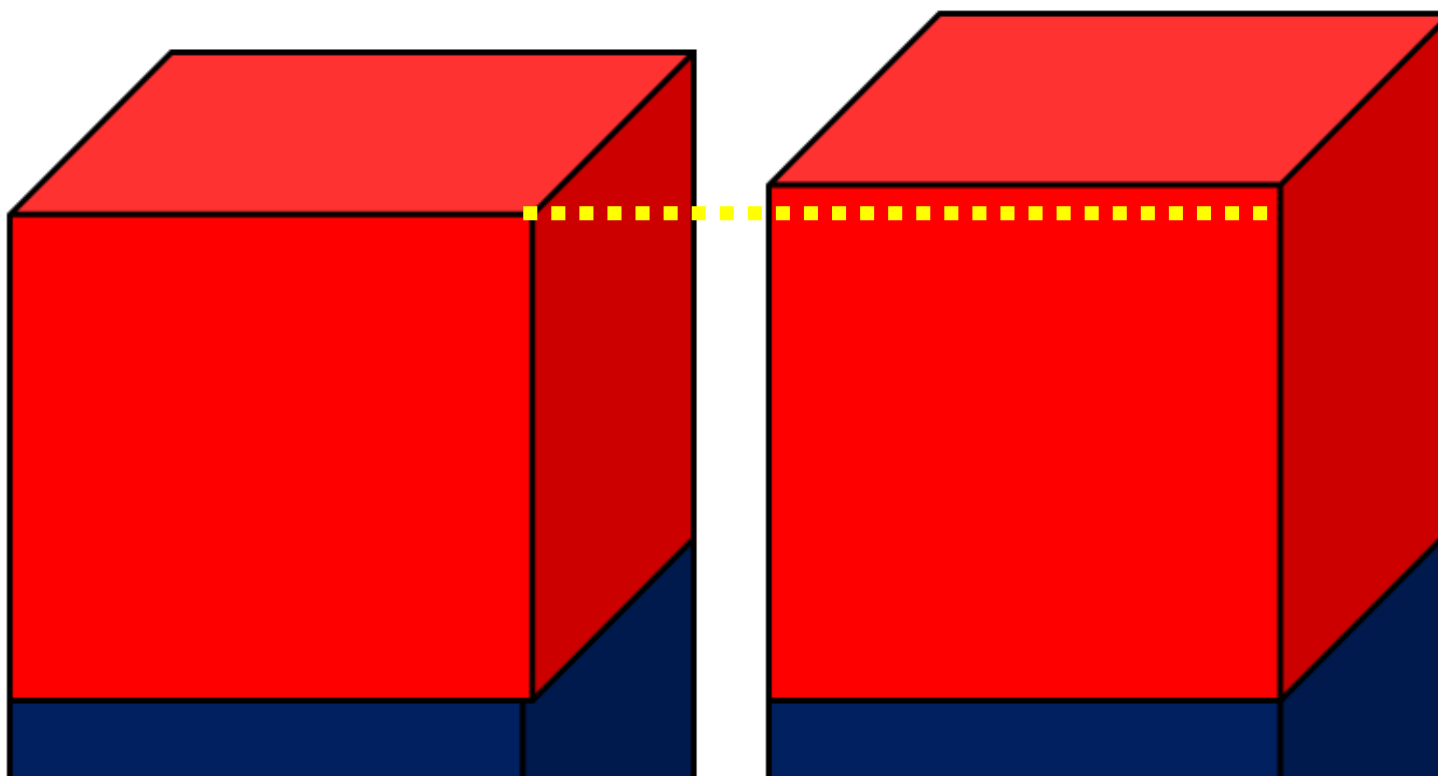
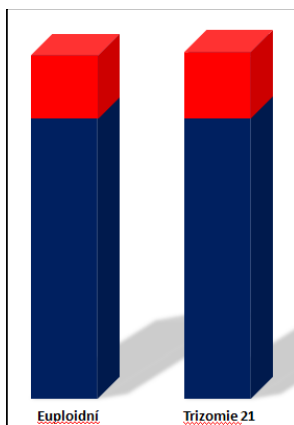
X vázené: hemofilie, retinitis pigmentosa

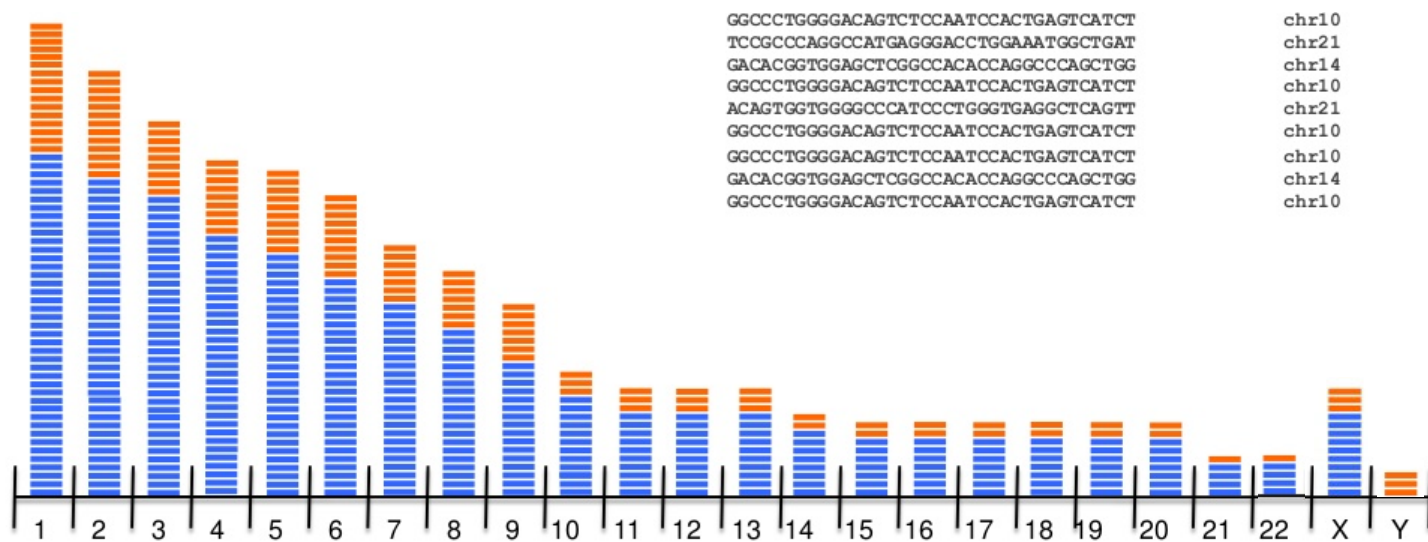
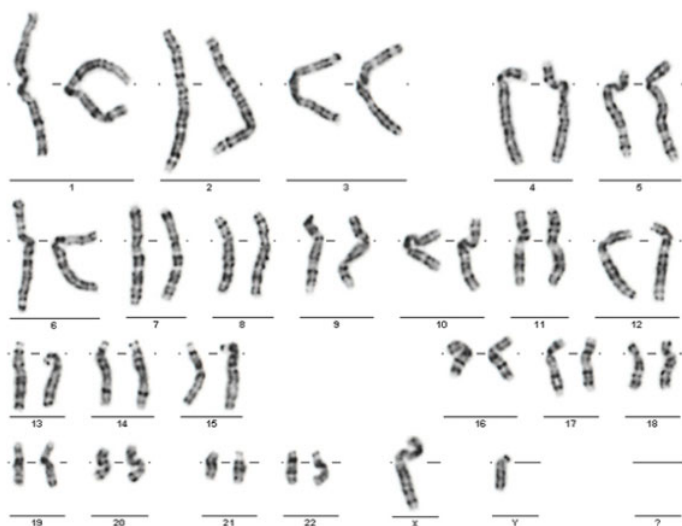


NIPT aneuploidii



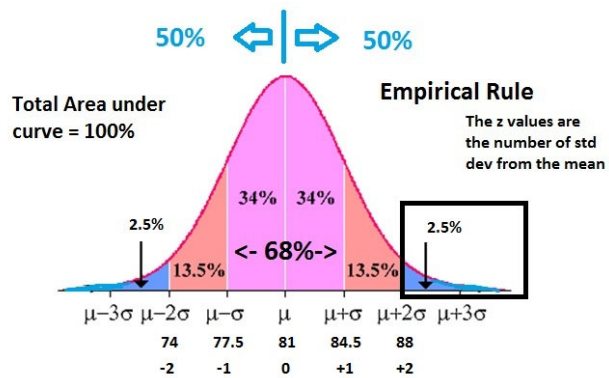






GGCCCTGGGGACAGTCTCCAATCCACTGAGTCATCT	chr10
TCCGCCCAGGCCATGAGGGACCTGGAAATGGCTGAT	chr21
GACACGGTGGAGCTCGGCCACACCAGGCCAGCTGG	chr14
GGCCCTGGGGACAGTCTCCAATCCACTGAGTCATCT	chr10
ACAGTGGTGGGGCCATCCCTGGGTGAGGCTCAGTT	chr21
GGCCCTGGGGACAGTCTCCAATCCACTGAGTCATCT	chr10
GGCCCTGGGGACAGTCTCCAATCCACTGAGTCATCT	chr10
GACACGGTGGAGCTCGGCCACACCAGGCCAGCTGG	chr14
GGCCCTGGGGACAGTCTCCAATCCACTGAGTCATCT	chr10

Chromozóm 21



z- score

NCV – normalised chromosome value

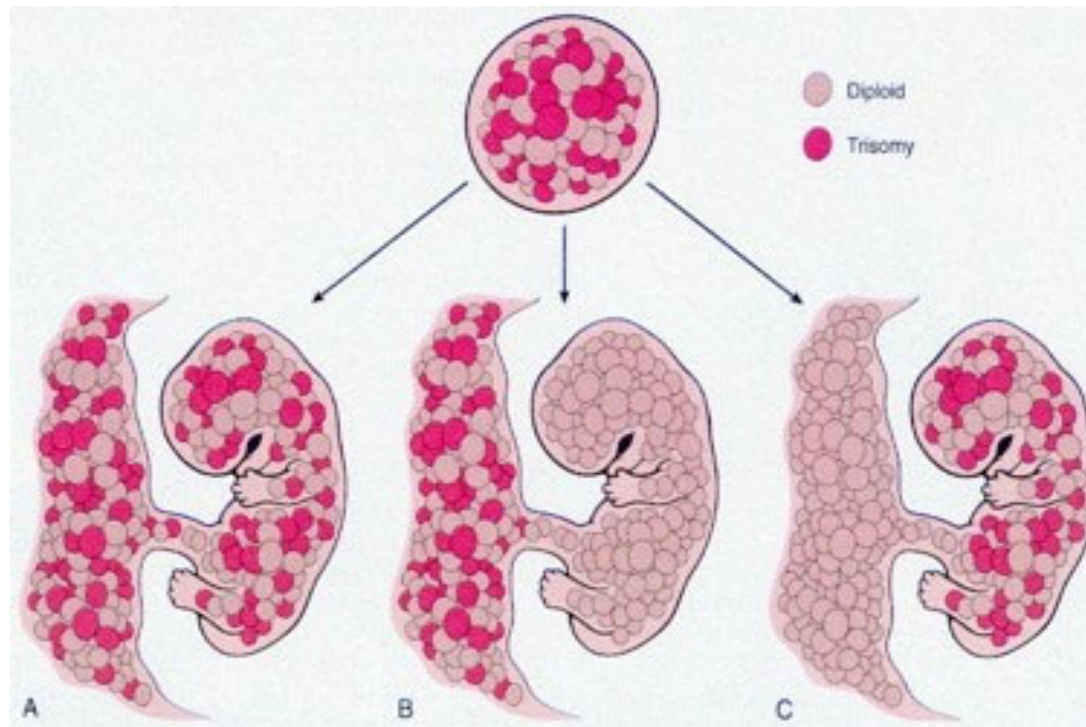
z- score

NCV – normalised chromosome value

Falešně pozitivní / negativní / zkreslené výsledky NIPT

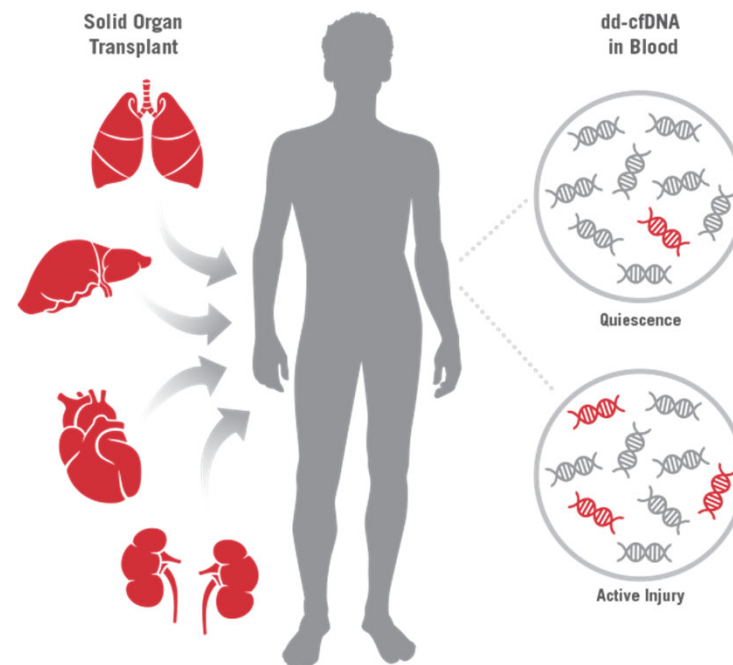
- fetoplacentární mozaicismus

1-2% všech těhotenství



Falešně pozitivní / negativní / zkreslené výsledky NIPT

- fetoplacentární mozaicismus
- „mizející“ dvojčata (8 týdnů; 0,23%)
- rozdílný karyotyp u vícčet
- maternální karyotyp, mozaicismus
- terapie (transfúze, transplantace...)
- malignita (1 na 1000 těhotenství)





Forenzní analýza DNA

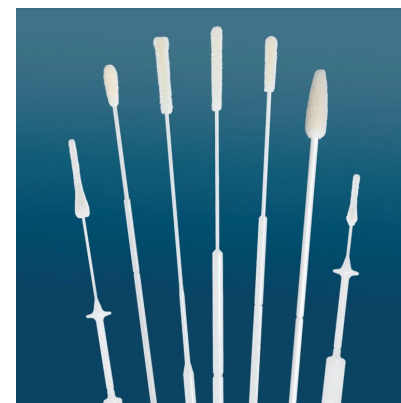


Forenzní analýza DNA – vzorky

- Stěry (FLOQ) – bukální stěry úplně izolováno od všeho ostatního
- Věcné stopy – laboratoř (kousek látky, případně dodatečný stěr) – krev vata, jinak FLOQ



– izolace



- Izolační kity např. QIAGEN



QIAamp DNA Investigator Kit (50)

Cat. No. / ID: 56504

For 50 DNA preps: 50 QIAamp MinElute Columns, Proteinase K, Carrier RNA, Buffers, Collection Tubes (2 ml)

[Copy order details](#)

\$342.00

[Log in](#) to see your account pricing.

– kvantifikace, amplifikace, vyhodnocení

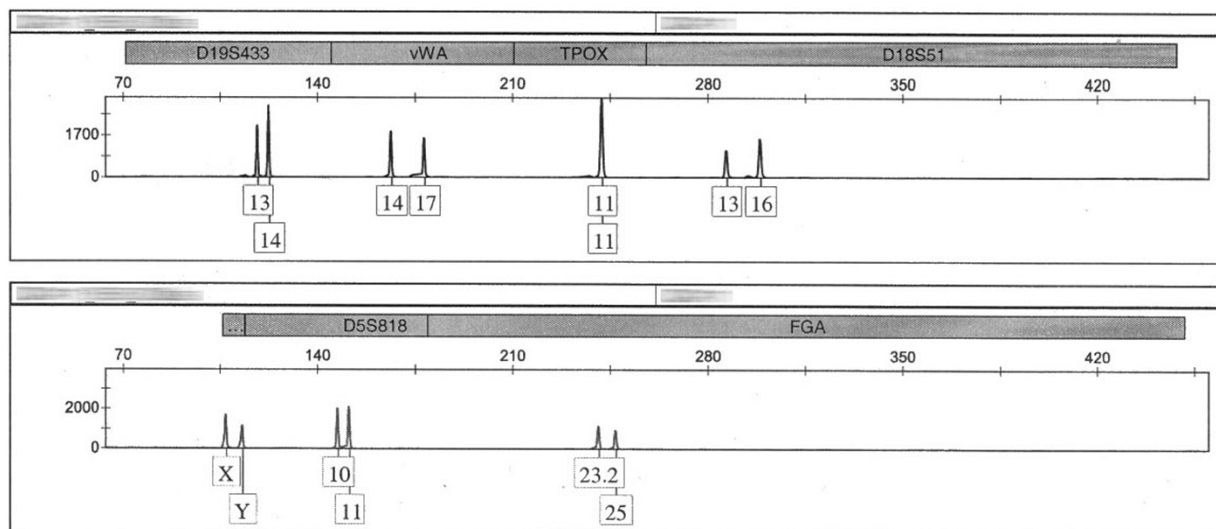
Forenzní analýza DNA - kvantifikace

- Je potřeba 5pg/ul (17ul do reakce)
- Stanovení pomocí qPCR (**Silné vzorky**, tj. krev, sliny, sperma – více pod kontrolou ruční příprava vzorků, Slabé vzorky, tj. ze zbraní, páčidel, klik atd. pomocí pipetovacích robotů)
- Celková lidská DNA + Y DNA
- Když pod limit -> konec
- Když nad limit -> amplifikace 23+1 STR markerů

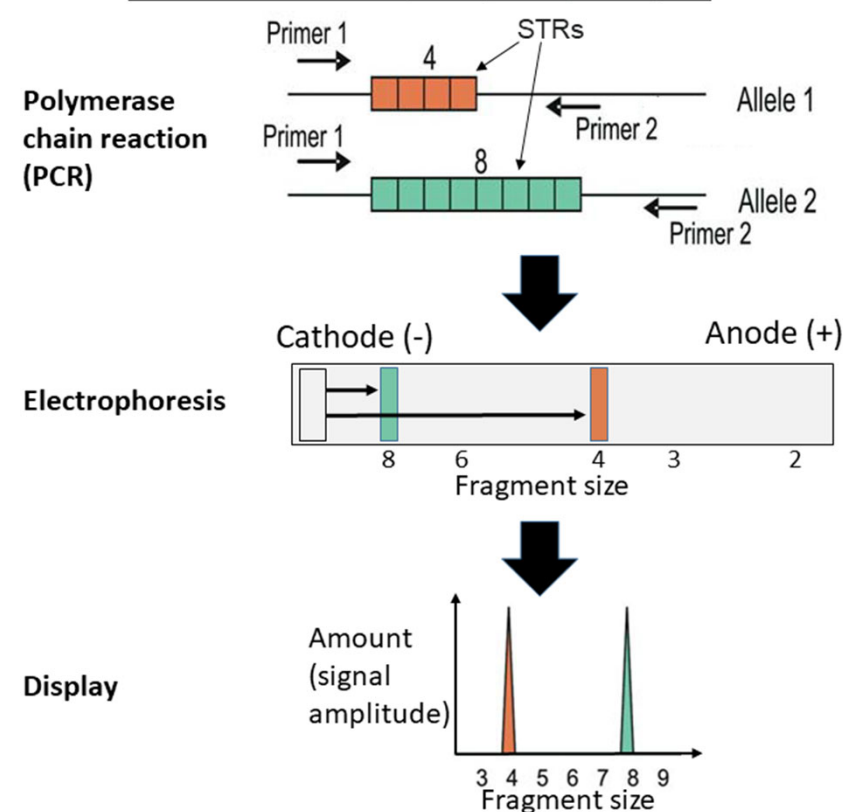


Forenzní analýza DNA - amplifikace

- Analýza 23+1 STR markerů
- Znásilnění – vylučovací metoda Y
- mtDNA – vylučovací metoda v ženské linii



Short Tandem Repeat (STR) analysis



Forenzní analýza DNA - vyhodnocení

- geneMapper
- CODIS – combined DNA index system (FBI)
- Databáze DNA PČR (jména nezávisle na CODIS)



[Home](#) / [Glossary](#)

CODIS

CODIS is an acronym for Combined DNA Index System, which is a computer software program that operates local, state, and national databases of DNA profiles from convicted offenders, unsolved crime scene evidence, and missing persons.

DNA (Deoxyribonucleic Acid)

Combined DNA Index System (CODIS)

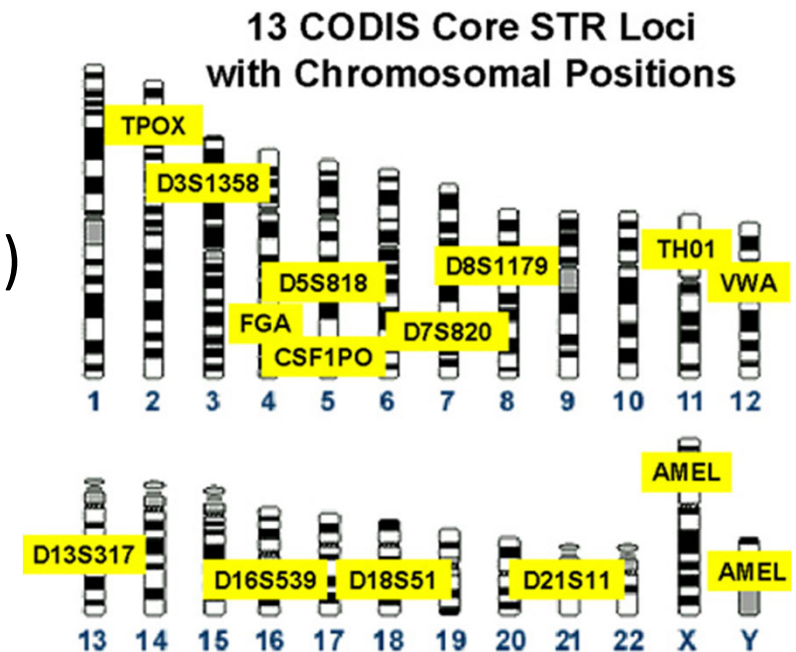
Forensic sciences

[FAQs](#)

[Glossary](#)

[Related Links](#)

[Multimedia](#)



Závěr:

Seznámili jsme se s metodami analýzy (např. genomových) sekvencí v základním i aplikovaném výzkumu.

Uvedli jsme si, kde genomové sekvence získat (sekvenací/stahováním z databází) a jak pracovat s velkými objemy datových souborů

Probrané databáze: GenBank, SRA, a přehled několika dalších

Vysvětlili jsme si funkci a strukturu formátů, ve kterých jsou sekvenční data uchovávána a zpracovávána

Probrali jsme sw Geneious, FASTQC, JDotter, RepeatExplorer, úvod práce v linuxu a příkazové řádce

Zběžně jsme se seznámili s forenzní analýzou DNA a sekvenční prenatální diagnostikou