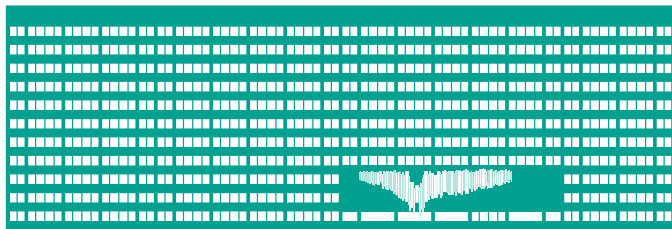


VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

VSB TECHNICAL
UNIVERSITY
OF OSTRAVA



www.vsb.cz

Algoritmy pro Bioinformatiku

Úvod do molekulární biologie

Michal Vašínek

VŠB – Technická univerzita Ostrava

FEI/EA404

michal.vasinek@vsb.cz

21. září 2022

100111100101011110101
ACCGTTACGGA CTTA
CATGGGTAGGGAGGGG
01101100110100111011





- michal.vasinek@vsb.cz
- Katedra informatiky / FEI - EA404
- Konzultační hodiny: Úterý 13:30-15:00
- www.olgen.cz



- homel.vsb.cz/~vas218/bio
- Klasifikovaný zápočet
- Projekt - 40 bodů
- Témata v průběhu semestru
- Úlohy na cvičení - 60 bodů
- Skupina v Teams.



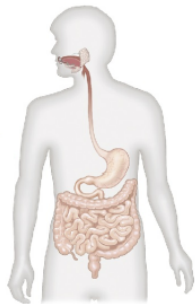
- Budete schopni porozumět algoritmům, které jsou součástí programů, které se používají v klinickém vyšetřování pacientů.
 - Zvládnete nástroje pro analýzu dat z Next-Generation Sequencingu (NGS).
 - Porozumíte různým typům výstupů současných nástrojů.
 - Budete rozumět procesu analýzy exprese genů.
 - Budete vědět odkud pochází Adam a Eva a proč si to myslíme.



- Základní pojmy z molekulární biologie.
- Podobnosti sekvencí, nalezení sekvence na referenční sekvenci.
- Struktury pro indexaci genomu.
- Biologické databáze.
- Sekvenační zařízení, fundamentální problémy.
https://www.youtube.com/watch?v=1_mER5qmaVk
- Predikce struktury RNA.
- Predikce genů v genomu.
- Sestavení neznámého genomu.
- Fylogenetická analýza.



- Buňka
- DNA, RNA, Protein
- Genom, Chromozom, Gen
- Centrální dogma
- Mutace



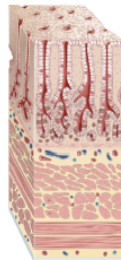
organismus



orgánová
soustava



orgán



buněčná
tkáň



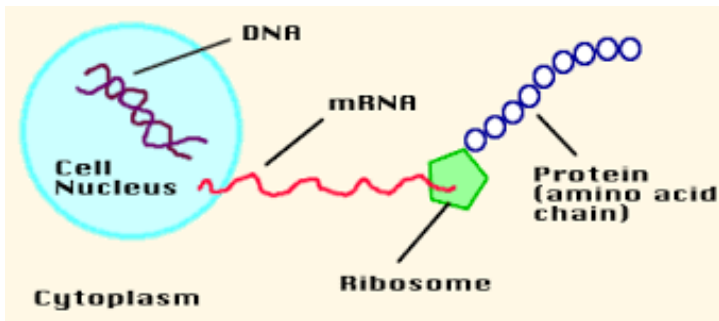
buňka

© Encyclopædia Britannica, Inc.

1

¹Je odhadováno, že člověk je složen z 37.2 bilionů buněk.

- Plní dva základní typy úkolů:
 - Provádí chemické reakce nezbytné k udržení života organismu.
 - Předává informace nezbytné k udržení života dalším generacím buňek.
- Základními zprostředkovateli úkolů výše jsou:
 - Proteiny - účastní se chemických reakcí.
 - DNA - uložení a poskytování informací.
 - RNA - Meziprodukt při syntéze proteinů z DNA.





Protein je řetězec aminokyselin.

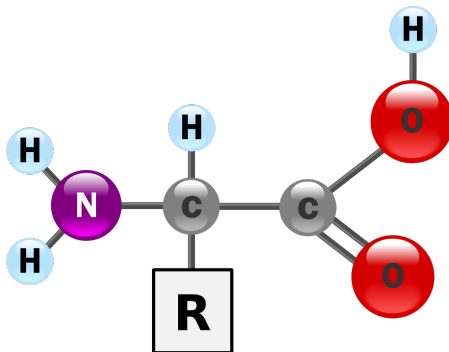
- Z infromatického hlediska je protein sekvencí symbolů z abecedy o 20 symbolech.
- Délka proteinu se různí od 20 až po několik tisíc aminokyselin.²
- Proteiny tvoří 3D struktury, které jsou základní stavební prvky organismu a jsou zodpovědné za provedení většiny chemických reakcí uvnitř buněk.
- Hormony, enzymy, kolagen, hemoglobin, ...

²Titin - je třetí nejhojněji zastoupený protein v lidském těle. Tělo dospělého jedince obsahuje cca 0.5 kg titinu. Titin je nejdelší známý protein s délkou cca 33 tisíc aminokyselin.



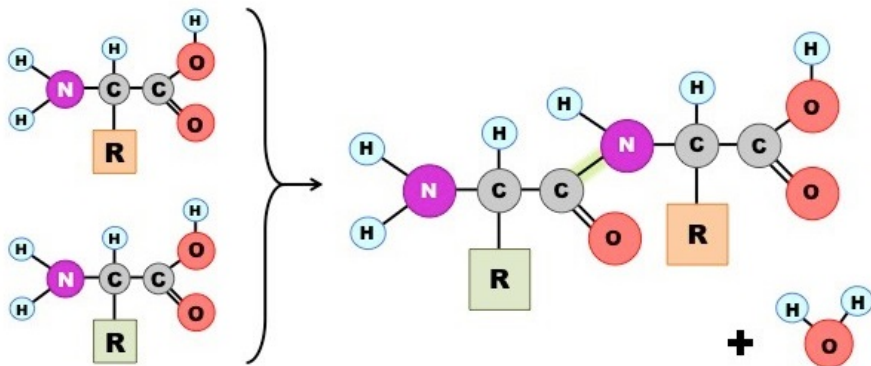
Každá aminokyselina se skládá ze tří funkčních skupin:

- Amino skupina
- Karboxylová skupina
- R skupina - postranní řetězec aminokyseliny





Protein se utváří spojením amino skupiny s karboxylovou skupinou za vzniku tzv. peptidové vazby.





Rozlišujeme čtyři struktury proteinů:

- Primární - rozlišení na úrovni aminokyselin.
- Sekundární - komplementární vazby mezi aminokyselinami.
- Terciární struktura - 3D struktura proteinu.
- Kvartérní struktura - interakce více proteinů, formování proteinových komplexů.

3

³<https://www.youtube.com/watch?v=wwTv8TqWC48>

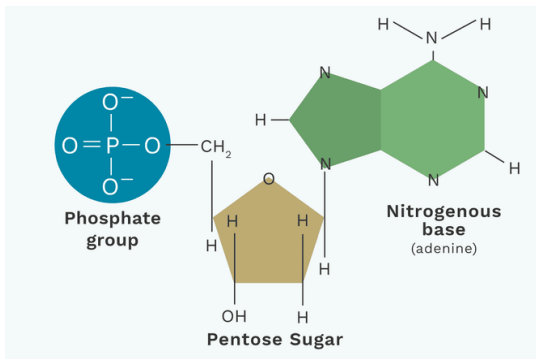


- DNA obsahuje informace potřebné k chodu buňky.
- Je tvořena dvěma vlákny, které jsou spleteny do dvoušroubovicové struktury.
- Každé vlákno je řetězcem propojených malých molekul označovaných pojmem nukleotidy.





- DNA - deoxyribonukleová kyselina
- Její páteř je tvořena trojicí molekul označovanou pojmem nukleotid:
 - Fosfátová skupina navázána na 5' konci.
 - Cukr (deoxy-)ribóza, která spojuje fosfátové skupinu a váže na sebe dusíkatou bází.
 - Báze - nositel informace.

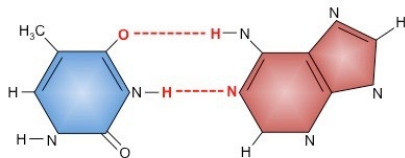




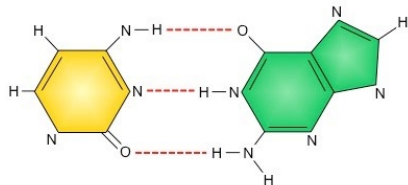
- Existuje pět různých nukleotidů:
 - adenine (A)
 - cytosin (C)
 - guanin (G)
 - thymin (T)
 - uracil (U)
- A,C,G a T se vyskytují v DNA.
- V RNA je místo thyminu uracil.



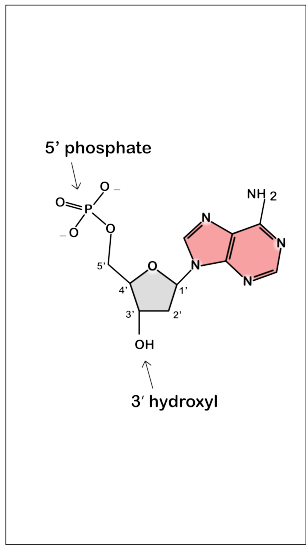
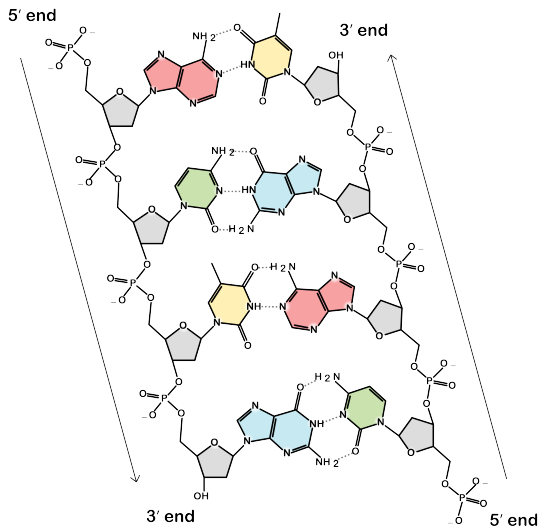
- Vlákna DNA drží pohromadě pomocí speciálního typu vazby: vodíkového můstku.
- Vazba A - T - dvě vodíkové vazby,
- Vazba C - G - tři vodíkové vazby.



Thymin - Adenine

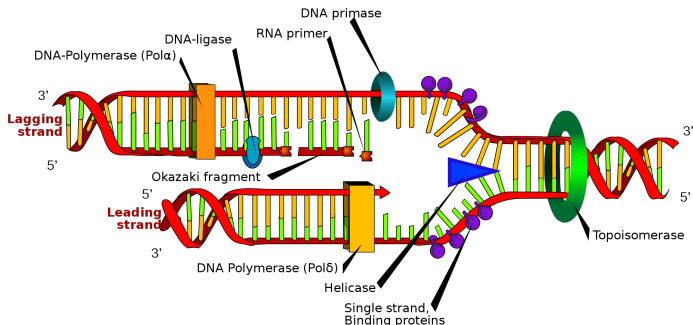


Cytosin - Guanin





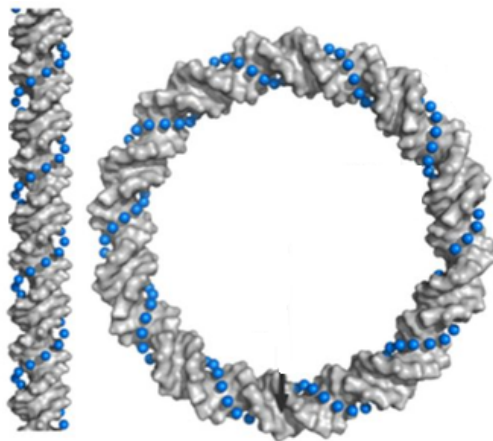
- Obě vlákna DNA obsahují stejnou informaci, liší se pouze směrem a jsou vzájemně komplementární.
- Komplementarita je základem replikace DNA.
- Replikace - vlákna jsou rozpletena a každé zvlášť je doplněno volnými nukleotidy.⁴



⁴<https://www.youtube.com/watch?v=TNKWgcFPHqw>

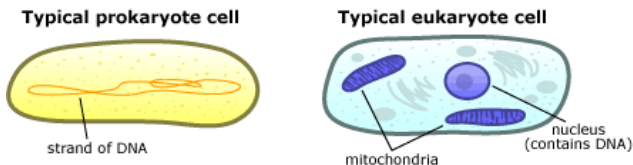


- DNA se obvykle vyskytuje v lineární podobě.
- U některých jednodušších organismů je DNA v kruhové podobě.
Například. *E.coli*



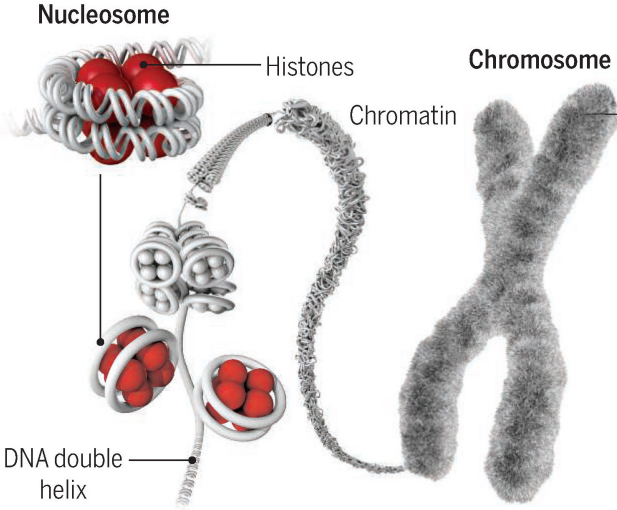


- V přírodě jsou dva typy buněk a od nich odvozené organismy:
- **Prokaryoty** jednobuněčné organismy, buňky nemají jádro. DNA volně plave v cytoplazmě. Například bakterie.
- **Eukaryoty** jedno a více buněčné organismy. DNA je umístěna v jádře buňky. Živočichové a rostliny.





- DNA řetězce jsou namotány kolem speciálního proteinu zvaného histone a společně tvoří strukturu nazývanou chromozom.
- DNA ze všech chromozomů se pak nazývá genomem daného organismu.
- U většiny organismů platí, že každá buňka obsahuje ve svém jádře tu samou DNA. Malé rozdíly kvůli mutacím.
- Lidský genom se skládá z 23 párů chromozomů. Od každého rodiče jeden pár.





- Gen je úsek DNA v němž je zakódován předpis pro vytvoření proteinu nebo RNA molekuly.
- Nemusí se vždy jednat o souvislý úsek DNA.
- U člověka se odhaduje cca 30 - 35 tisíc genů.
- Jeden gen může kódovat více než jenom jeden protein, dochází k různému prostorovému uspořádání.



- Lidský genom je složen přibližně z 3.2 mld. párů bazí, neboli 3.2 Gbp.
- Amoeba dubia (jednobuněčný organismus) 670 Gbp.
- Velikost genomu není to samé co složitost organismu.
- Lidský genom zabírá na disku cca 1GB, Windows 10 složka System32 6GB.
- Některé známé genomy:
 - Nejmenší virový genom - virus Porcine circovirus - 1,8 knt.
 - Nejmenší bakteriální genom - bakterie Nasuia deltocephalinicola - 112 kbp a 137 genů.
 - Virus HIV - 9.7 knt.

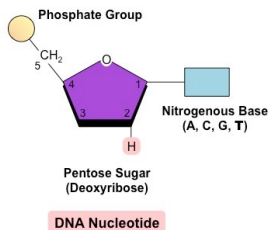
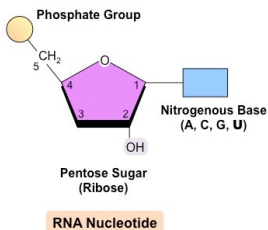


- E.coli - prokaryotický organismus
 - Délka genomu: 4.5 Mbp ,
 - Počet genů: 4 tisíce.
 - Průměrná délka genu: 1000 bp.
 - 90% genomu je tvořeno geny.
- Člověk:
 - Délka genomu: 3,2 Gbp
 - Počet genů 30-35 tisíc.
 - Odhadovaná průměrná délka 1000-2000 bp.
 - Pouze přibližně 2,5% lidského genomu je tvořeno geny. Zbytek "Junk DNA"⁵.
- SARS-COV2
 - Délka genomu: 30 kbp
 - Počet genů 11.

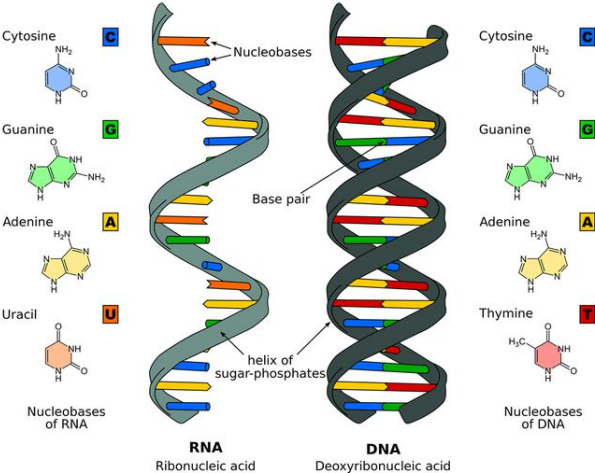
⁵Junk se doslova překládá jako haraburdí.



- Podobná DNA, je schopná ukládání a přenosu informací.
- Jednovláknová.
- Místo Thyminu(T) je Uracil (U), Uracil je také komplementární k Adeninu (A).
- Dodatečná OH skupina umožňuje RNA vytvářet více vodíkových můstků a tak utvářet složitější prostorové struktury než DNA.



RNA ve srovnání s DNA



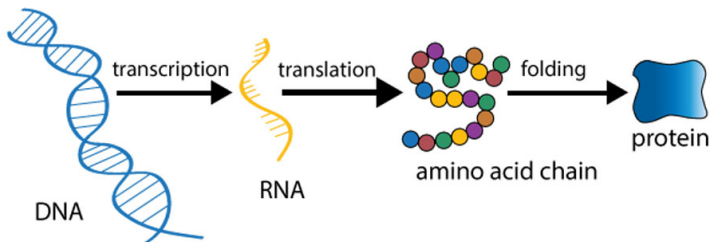


- mRNA - meziprodukt při tvorbě proteinů.
- transferová RNA (tRNA), ribosomální (rRNA), krátké RNA snoRNA, microRNA, siRNA, nebo dlouhé ncRNA.
- microRNA nebo též miRNA je cca 22 nt dlouhá RNA molekula.
- Je komplementární k některým úsekům mRNA.
- Tím, že se naváže na mRNA znemožní vytvoření proteinu => reguluje tvorbu proteinů⁶.
- Pokud virus vloží svůj kód do DNA aby se replikoval, navázáním miRNA lze potlačit schopnost viru se replikovat.

⁶<https://www.youtube.com/watch?v=t5jroSCBBwk>



- Centrální Dogma nám říká, jakým způsobem obdržíme protein z konkrétního genu. Tomuto procesu se říká exprese genu.
- Exprese genu se skládá ze dvou kroků⁷:
 - Transkripce: DNA → mRNA
 - Translace: mRNA → Protein
 - Následně dochází k po-translačním úpravám.
- Každá aminokyselina v proteinu je zakódována trojicí nukleotidů, tzv. codonem.



⁷<https://www.youtube.com/watch?v=gG7uCskU0rA>



		Second letter				
		U	C	A	G	
First letter	U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA STOP UAG STOP	UGU Cys UGC UGA STOP UGG Trp	U C A G
	C	CUU CUC Leu CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC Arg CGA CGG	U C A G
	A	AUU Ile AUC AUA AUG Met	ACU ACC Thr ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU GUC Val GUA GUG	GCU GCC Ala GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC Gly GGA GGG	U C A G

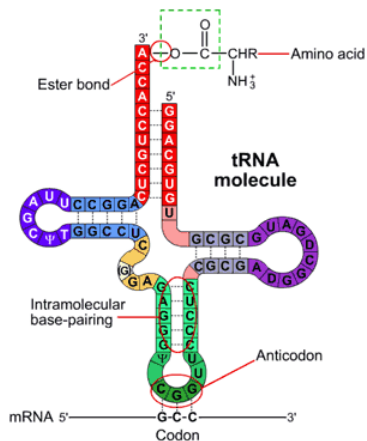


- Enzym RNA polymeráza dočasně rozplete dvouvláknovou strukturu DNA.
- Transkripce DNA do RNA začne v místě, kde se vyskytuje start codon.
- Během transkripce dojde k nahrazení thyminu (T) za uracil (U).
- Když polymeráza dosáhne stop codonu, transkripce se ukončí.



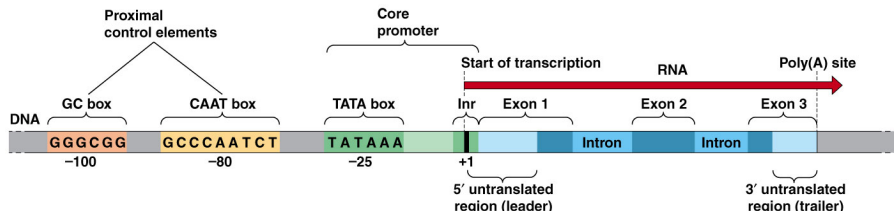
- Ve fázi translace dochází k vytvoření proteinu.
- Existuje $4^3 = 64$ různých kodonů, ale je pouze 20 aminokyselin => mnoho je kódováno více kodony.
- Všechny organismy používají stejný genetický kód.
- Aminokyselina je dopravena na ribosom pomocí tRNA, která je komplementární na právě zpracovávaný kodon.

- Existuje 61 různých tRNA, které odpovídají různým neterminálním kodonům.





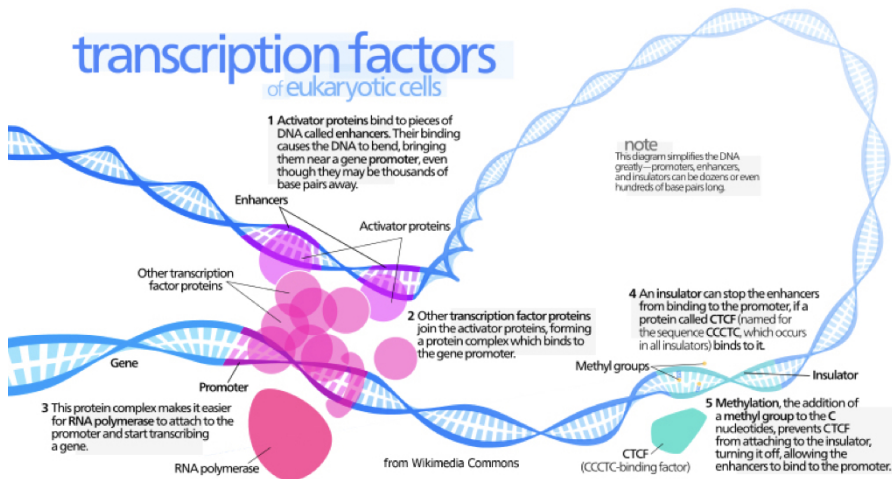
- Gen je složitá struktura, na jehož expresi se podílí celá řada proteinů.
- RNA polymeráza nasedá na DNA sekvenci v místě tzv. TATA boxu.
- TATA box je předcházen dalšími regulačními místy, na která se vážou proteiny ulehčující nebo ztěžující nasednutí polymerázy.
- Pouze exonové oblasti jsou přeloženy do proteinu, introny jsou v průběhu transkripce odstřiženy (obvykle začíná GT a končí AG).



© 2012 Pearson Education, Inc.



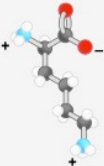
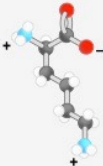
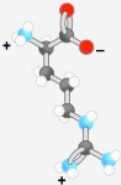
transcription factors of eukaryotic cells

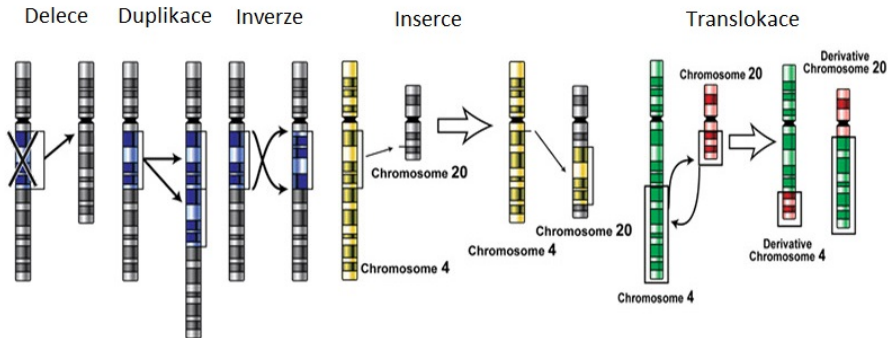




- Přestože dělení buněk a replikace DNA je téměř bezchybný proces může v něm občas dojít k chybě. Tyto změny v DNA označujeme pojmem mutace.
- Mutace a přirozený výběr jsou hlavním důvodem evoluce organismů.
- Některé mutace mohou vést k vytvoření odlišných vlastností, které umožní organismu lépe se adaptovat na prostředí.
- Mutace je však také zdrojem negativních projevů jako třeba rakovinná onemocnění.



	No mutation	Point mutations		
		Silent	Missense	Nonsense
DNA	TTC	TTT	TCC	ATC
mRNA	AAG	AAA	AGG	UAG
Protein	Lys	Lys	Arg	STOP
				





- V lednu 2020 vyšel článek Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag ⁸
- Přebrala média (dnes již články staženy z webu).
- Nesmyslnost lze snadno ověřit s použitím bioinformatických nástrojů.
<https://blast.ncbi.nlm.nih.gov/>

⁸<https://www.biorxiv.org/content/10.1101/2020.01.30.927871v1>



www.biocomicals.com

Thank you for your attention

Michal Vašínek

VŠB – Technická univerzita Ostrava

FEI/EA404

michal.vasinek@vsb.cz

21. září 2022