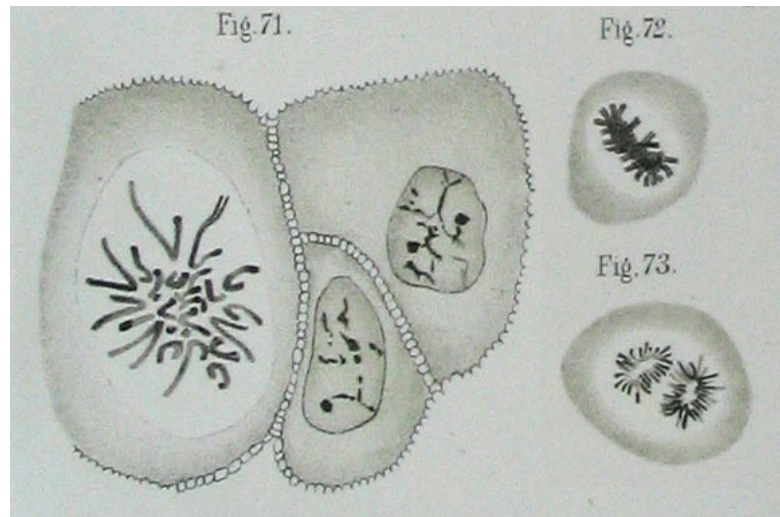


# Biológia pre informatikov

Askar Gafurov

22.9.2022



Walther Flemming, 1881

## Hlavné postavy

### Deoxyribonukleová kyselina (DNA)

Obsahuje genetickú informáciu prenášanú z generácie na generáciu.

Dlhý reťazec nukleotidov z množiny  $\{A, C, G, T\}$

(adenín, cytozín, guanín, tymín).

Informácia uložená v symbolickej, digitálnej forme.

### Ribonukleová kyselina (RNA)

Blízka príbuzná DNA, tymín T nahradený uracylom U

### Proteíny (bielkoviny)

Katalyzujú biochemické reakcie v bunke (enzýmy),

prenášajú signály v rámci bunky/medzi bunkami,

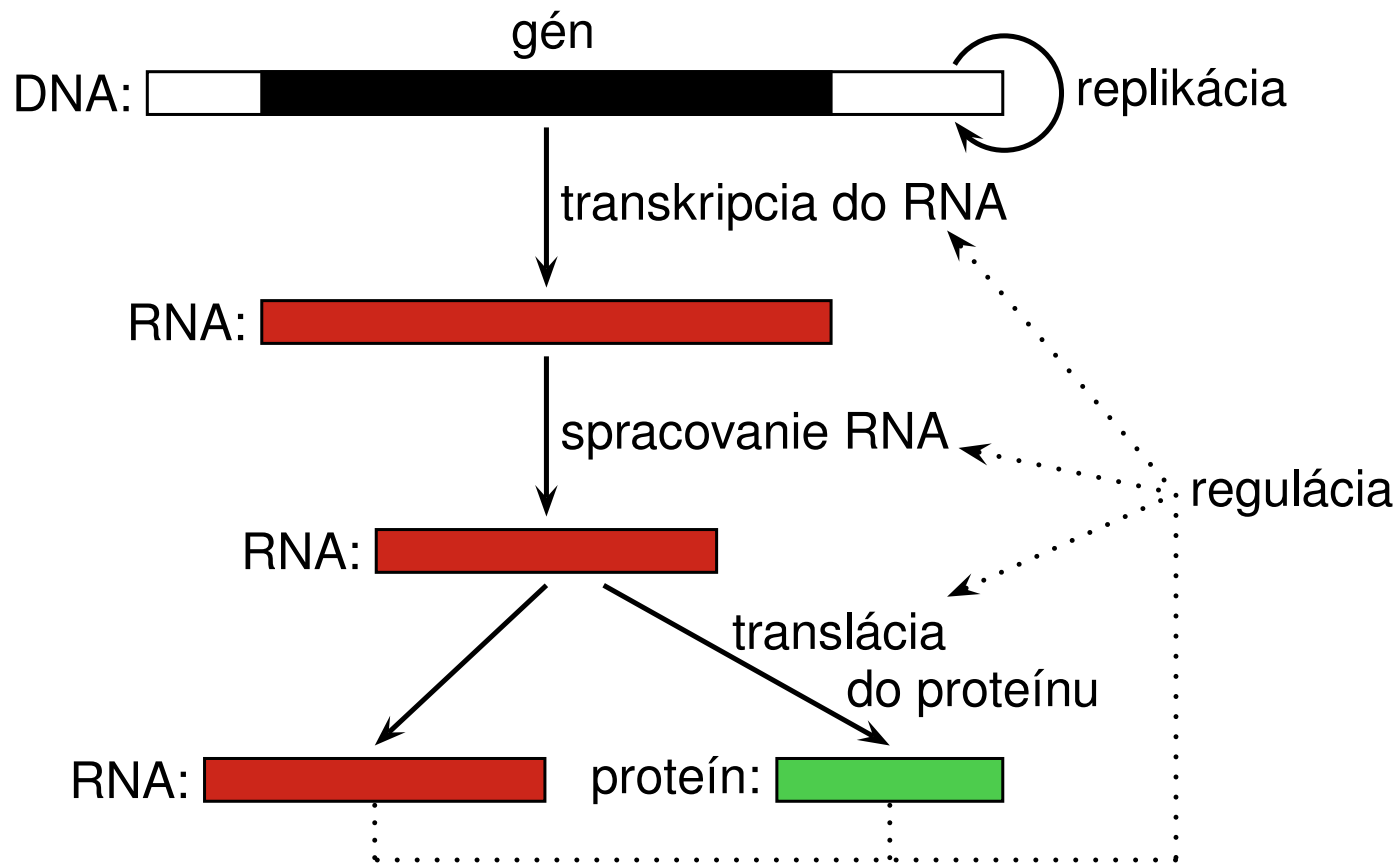
sú dôležité pre stavbu bunky a pohyb.

Reťazec aminokyselín (20 rôznych aminokyselín).

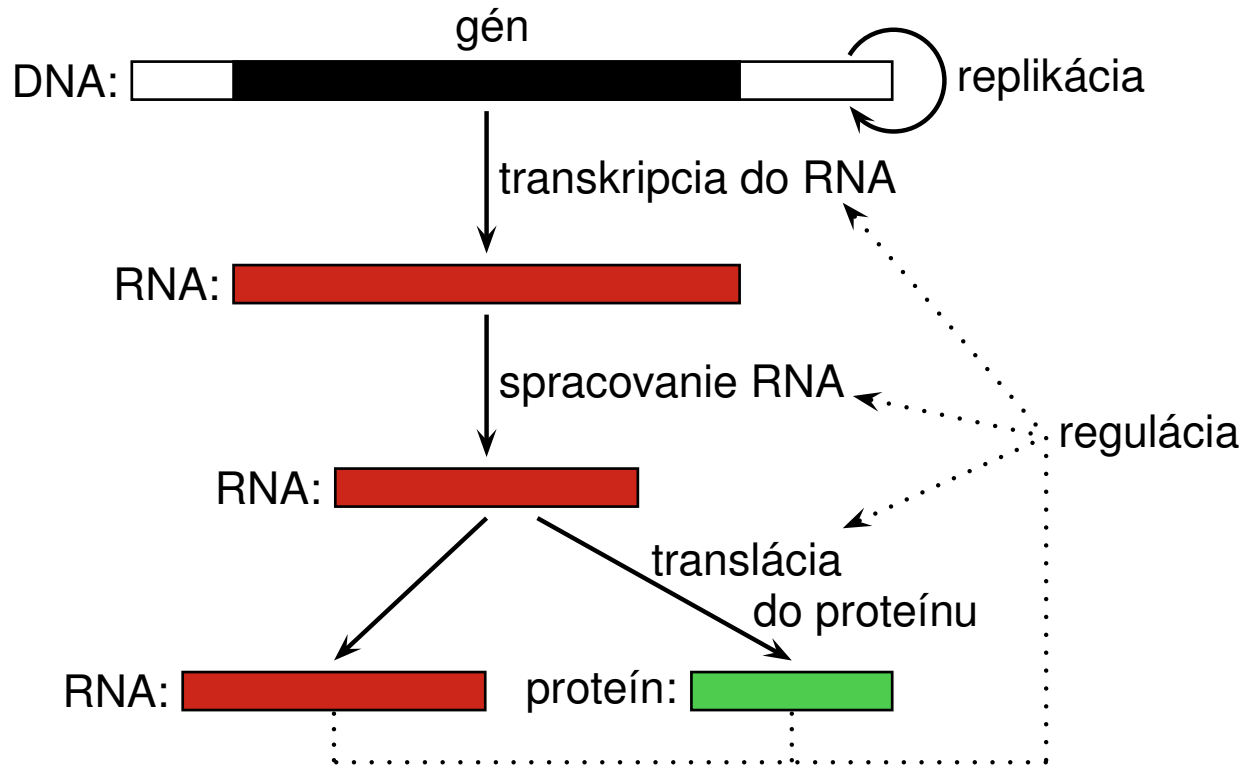
## Aká informácia je uložená v DNA?

**Gény:** Predpisy na tvorbu proteínov a funkčných RNA molekúl.

**Riadenie ich expresie:** kedy a koľko sa má tvoriť.



## Centrálna dogma (Francis Crick 1958,1970)



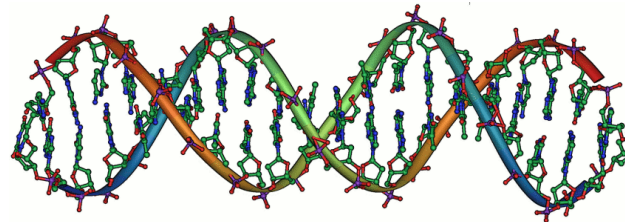
*“The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid.”*

## DNA, chromozómy

**DNA:** dve komplementárne vlákna, strands (páry A-T, C-G),  
v opačnej orientácii (konce sa nazývajú 5' a 3').

Napr. ACCATG je komplementárny s CATGGT.

Tvar dvojitej špirály:



Dvojvláknová štruktúra poskytuje redundanciu, možnosť opravy pri poškodení jedného vlákna.

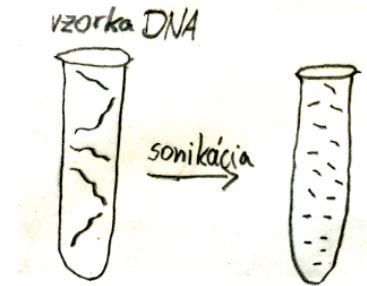
Pri delení bunky sa dvojvláknová DNA rozdelí a ku každému vláknu sa doplní komplement (DNA replikácia).

**Chromozóm:** Súvislý úsek dvojvláknovej DNA a podporných proteínov.

Ľudský genóm má 22 párov chromozómov plus dva pohlavné,  
spolu 3GB.

## Technológia: sekvenovanie DNA

- Postup na zisťovanie poradia báz v chromozómoch genómu.
- Chromozómy sa nasekajú na krátke kúsky, každý sa sekvenuje zvlášť  
napr. Sangerovým sekvenovaním.  
– využíva prírodné enzýmy, napr. DNA polymerázu



## Sangerovo sekvenovanie (Sanger sequencing)

Sekvenujeme AGCTAGGACT (zobrazená sprava doľava)

Primer AGT + enzýmy + nukleotidy + modifikované ofarbené nukleotidy

Výsledky sekvenovacej reakcie:

```

TCAGGATCGA
AGTCCTAGC TCAGGATCGA
              AGTCCTA
              TCAGGATCGA
              AGTCCTAGCT
              TCAGGATCGA
              AGTCCT
TCAGGATCGA TCAGGATCGA
AGTCC      AGTCCT
              TCAGGATCGA
              AGTCCTAG
              TCAGGATCGA
              AGTC
    
```

Na géli zoradíme podľa dĺžky:

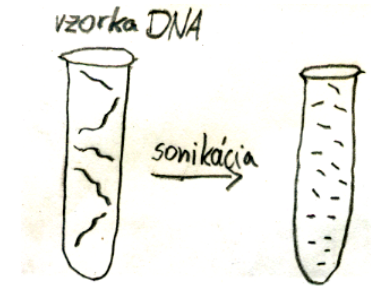
```

AGTCCTAGCT
AGTCCTAGC
AGTCCTAG
AGTCCTA
AGTCCT
AGTCC
AGTC
AGTC
    
```

Odčítaním farieb dostaneme komplementárne vlákno: AGTCCTAGCT

## Technológia: sekvenovanie DNA

- Postup na zisťovanie poradia báz v chromozómoch genómu.
- Chromozómy sa nasekajú na krátke kúsky, každý sa sekvenuje zvlášť napr. Sangerovým sekvenovaním.
  - využíva prírodné enzýmy, napr. DNA polymerázu



- **Bioinformatický problém:** skladanie celej sekvencie z kúskov.
- Dostupnosť genómov umožňuje katalogizovať gény a iné funkčné úseky, hľadať podobnosti a rozdiely medzi druhmi a jedincami.

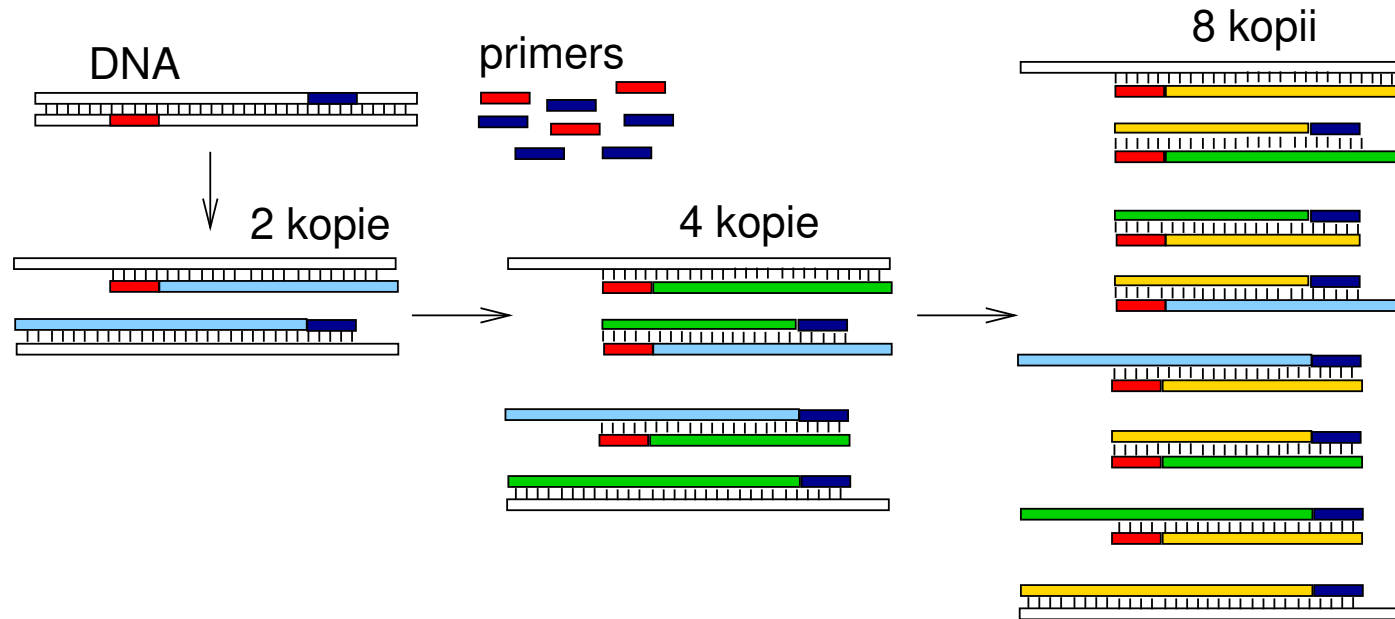


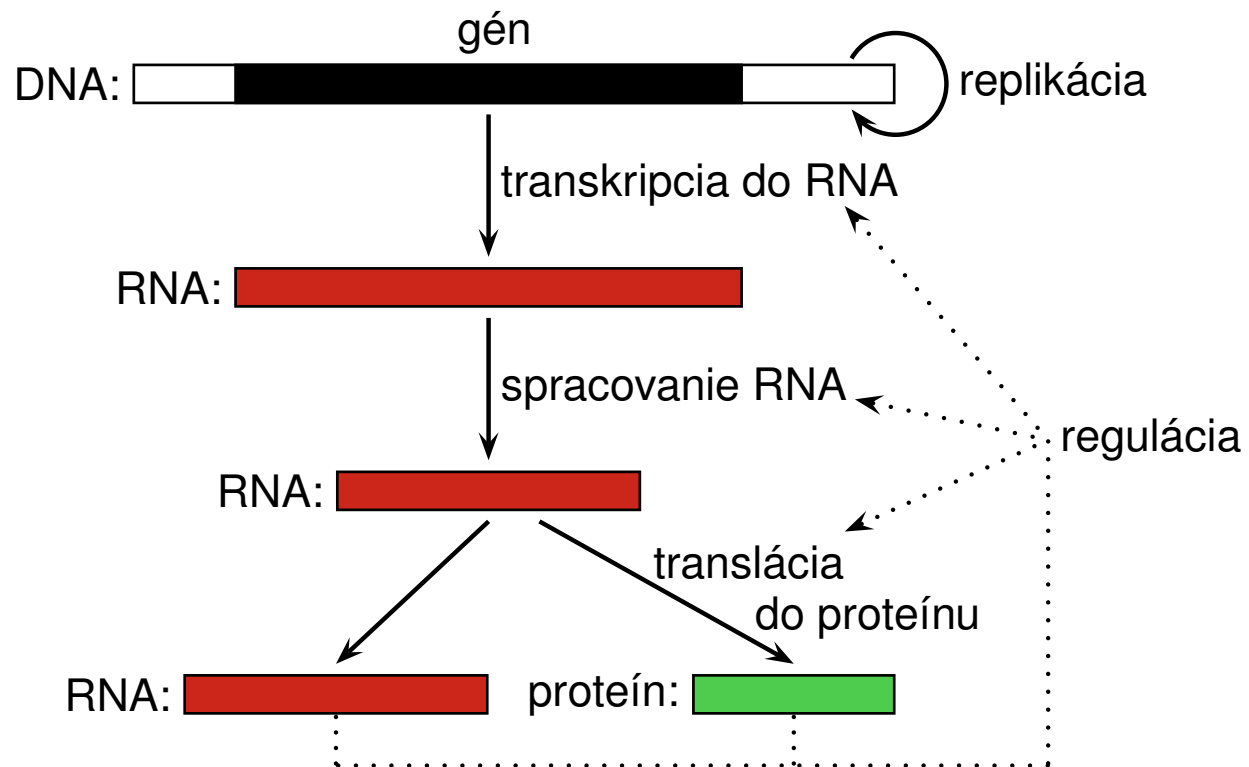
## PCR (polymerase chain reaction)

Zvolíme si dva krátke úseky DNA (primers)

PCR testuje či sú v DNA blízko seba (stovky, tisíce báz)

Ak áno, namnoží úsek medzi nimi

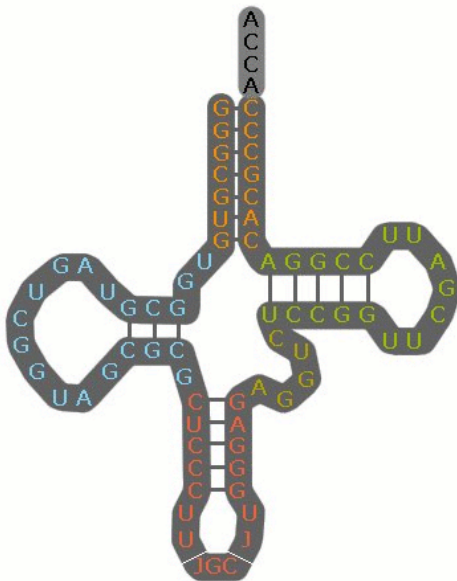




# RNA

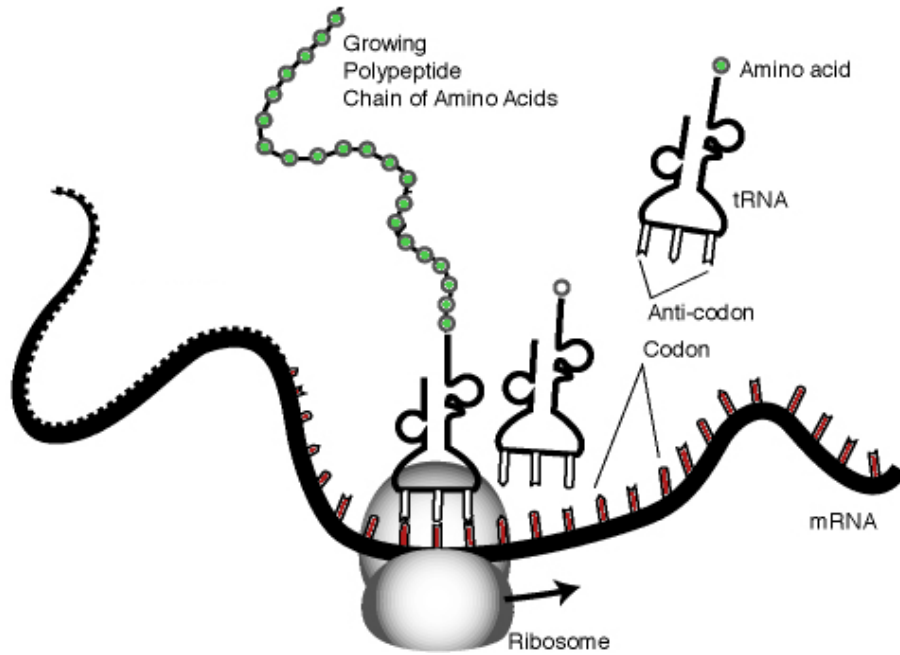
## Ako sa líši od DNA?

- obsahuje ribózu namiesto deoxyribózy
- obsahuje uracil namiesto tymínu (bázy A,C,G,U)
- jednovláknové reťazce, zvyčajne kratšie
- zložitá sekundárna štruktúra: spárované komplementárne úseky

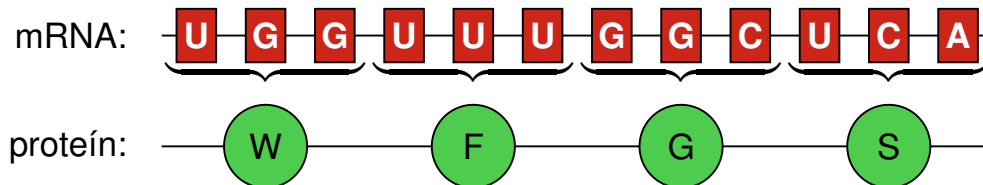


transferová RNA (tRNA)

# Translácia



Kodón (trojica nukleotidov) určuje 1 aminokyselinu



## Genetický kód

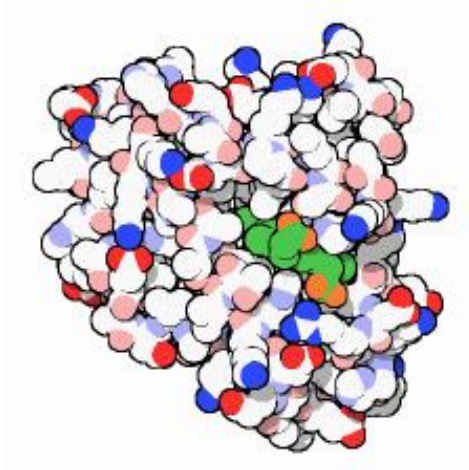
Ala / A	GCT, GCC, GCA, GCG	Leu / L	TTA, TTG, CTT, CTC, CTA, C
Arg / R	CGT, CGC, CGA, CGG, AGA, AGG	Lys / K	AAA, AAG
Asn / N	AAT, AAC	Met / M	ATG
Asp / D	GAT, GAC	Phe / F	TTT, TTC
Cys / C	TGT, TGC	Pro / P	CCT, CCC, CCA, CCG
Gln / Q	CAA, CAG	Ser / S	TCT, TCC, TCA, TCG, AGT, A
Glu / E	GAA, GAG	Thr / T	ACT, ACC, ACA, ACG
Gly / G	GGT, GGC, GGA, GGG	Trp / W	TGG
His / H	CAT, CAC	Tyr / Y	TAT, TAC
Ile / I	ATT, ATC, ATA	Val / V	GTT, GTC, GTA, GTG
START	ATG	STOP	TAA, TGA, TAG

## Proteíny

Reťazce 20 rôznych aminokyselín s rôznymi chemickými vlastnosťami:

Aminokyselina	Postranný reťazec	Jeho vlastnosti
Alanín (A)	-CH <sub>3</sub>	hydrofóbny
Arginín (R)	-(CH <sub>2</sub> ) <sub>3</sub> NH-C(NH)NH <sub>2</sub>	bázický
Asparagín (N)	-CH <sub>2</sub> CONH <sub>2</sub>	hydrofilný
Kyselina asparágová (D)	-CH <sub>2</sub> COOH	kyslý
Cysteín (C)	-CH <sub>2</sub> SH	hydrofóbny
Kyselina glutámová (E)	-CH <sub>2</sub> CH <sub>2</sub> COOH	kyslý
Glutamín (Q)	-CH <sub>2</sub> CH <sub>2</sub> CONH <sub>2</sub>	hydrofilný
Glycín (G)	-H	hydrofilný
Histidín (H)	-CH <sub>2</sub> -C <sub>3</sub> H <sub>3</sub> N <sub>2</sub>	bázický
Izoleucín (I)	-CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	hydrofóbny
Leucín (L)	-CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	hydrofóbny
Lyzín (K)	-(CH <sub>2</sub> ) <sub>4</sub> NH <sub>2</sub>	bázický
Metionín (M)	-CH <sub>2</sub> CH <sub>2</sub> SCH <sub>3</sub>	hydrofóbny
Fenylalanín (F)	-CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	hydrofóbny
Prolín (P)	-CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> -	hydrofóbny
Serín (S)	-CH <sub>2</sub> OH	hydrofilný
Treonín (T)	-CH(OH)CH <sub>3</sub>	hydrofilný
Tryptofán (W)	-CH <sub>2</sub> C <sub>8</sub> H <sub>6</sub> N	hydrofóbny
Tyrozín (Y)	-CH <sub>2</sub> -C <sub>6</sub> H <sub>4</sub> OH	hydrofóbny
Valín (V)	-CH(CH <sub>3</sub> ) <sub>2</sub>	hydrofóbny

## Štruktúra proteínov

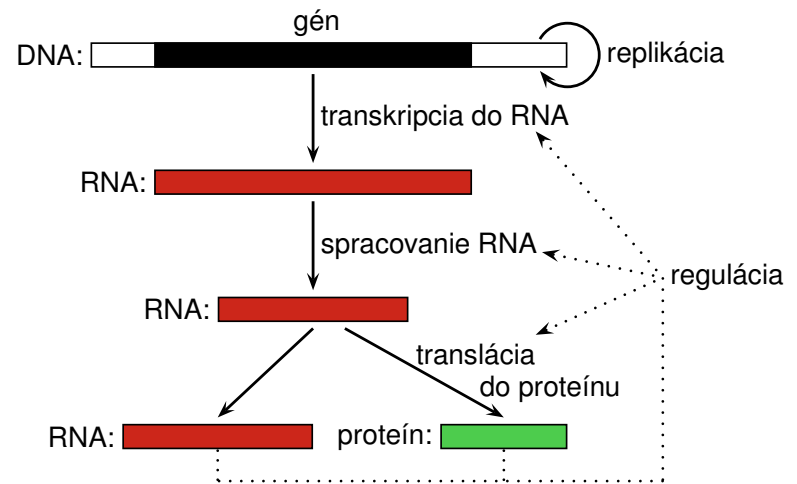
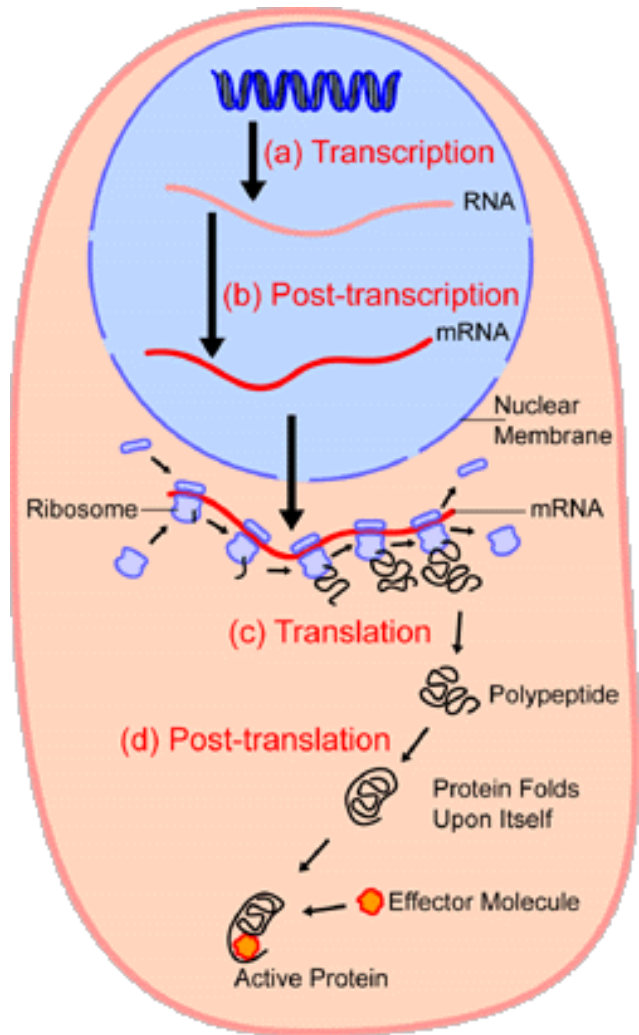


Myoglobín, prvý proteín so známou štruktúrou (Kendrew a kol. 1958).

Proteíny sa vyskytujú poskladané v určitej stabilnej štruktúre, prípadne prechádzajú medzi niekoľkými stavmi.

Hydrofóbne aminokyseliny neinteragujú s vodou, zväčša sa vyskytujú vo vnútri štruktúry.

Štruktúra proteínu určuje jeho funkciu.



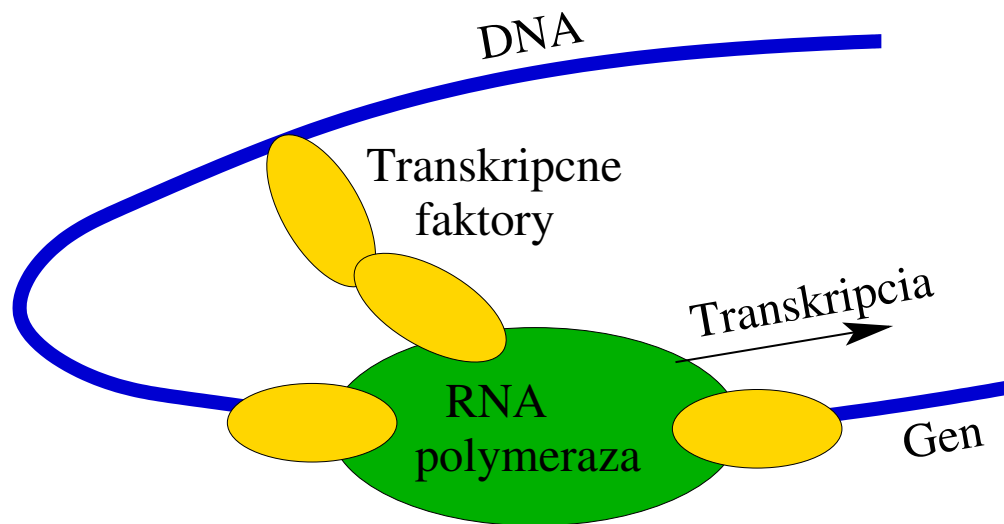


## Regulácia expresie

Bunky v rôznych tkanivách toho istého organizmu zdieľajú ten istý genóm, vyzerajú a fungujú však veľmi rôzne.

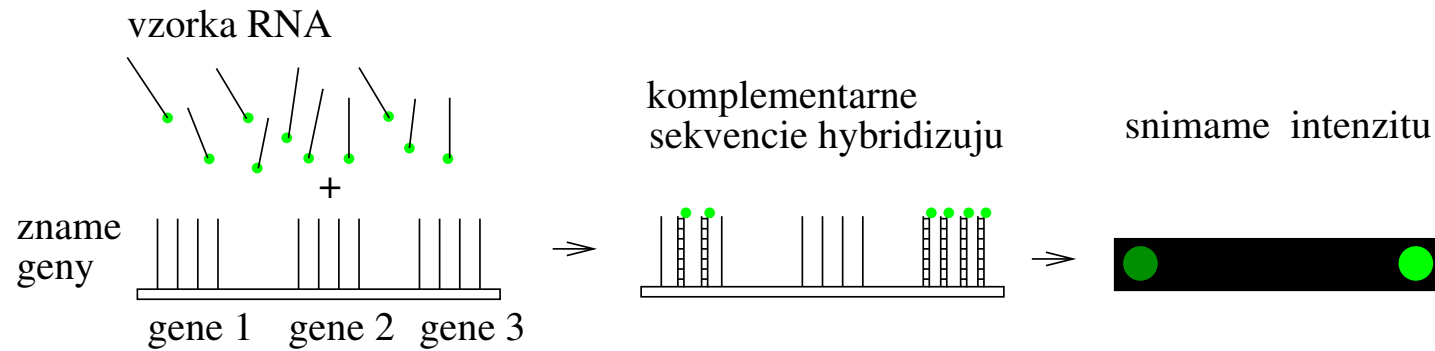
Niektoré proteíny sa tvoria len za určitých okolností, alebo v premenlivom množstve.

Regulácia začatia transkripcie pomocou transkripčných faktorov:



**Bioinformatický problém:** zisti, ktoré faktory ovplyvňujú ktorý gén, kde presne sa viažu.

## Technológia: microarray



Meranie množstva mRNA prítomnej v bunke pre **veľa génov** naraz.

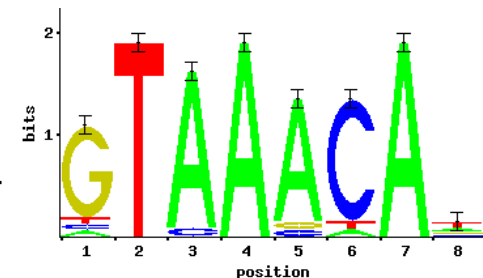
Zopakujeme za rôznych podmienok, študujeme korelácie medzi génmi.

Môžu byť dôsledkom spoločného regulátora (transkripčného faktoru).

### Bioinformatický problém:

niekoľko ko-regulovaných génov,

nájdí motív, ku ktorému sa môže viazať spoločný transkripčný faktor (**motif finding**)



## Príklad microarray dát

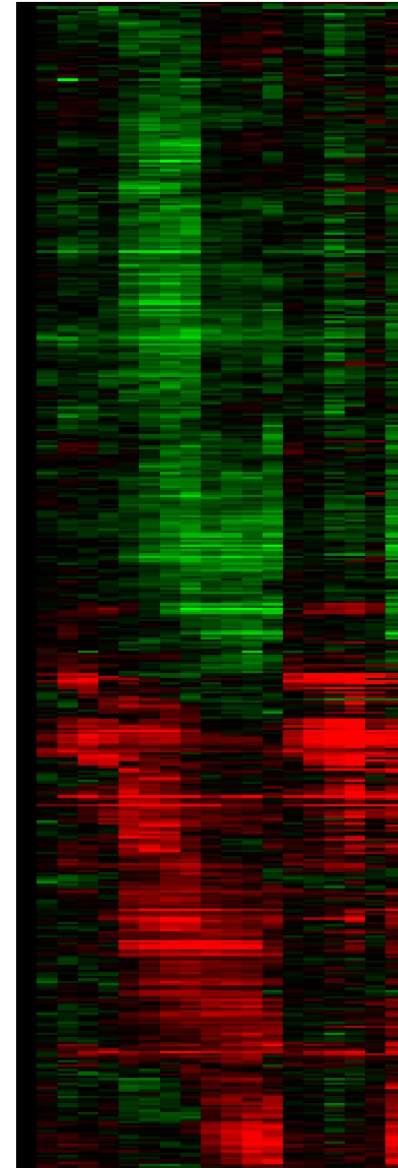
Pomer expresie génu v meranej a kontrolnej vzorke  
fg/bg

Červená:  $fg > bg$

Zelená:  $fg < bg$

517 génov

19 experimentov



## Mutácie DNA

V DNA občas dochádza k zmenám, mutáciám (napr. pod vplyvom prostredia, či chybou pri replikácii).

### Typy mutácií:

substitúcia, substitution (jedna báza sa zmení na inú),  
inzercia, insertion (vloží sa niekoľko nových báz),  
delécia, deletion (vynechá sa niekoľko báz),  
zmeny väčšieho rozsahu (napr. translokácie).

### Bioinformatické problémy:

Ktoré sekvencie vznikli z spoločného predka mutovaním?

(hľadanie homológov, homology search)

Ktoré bázy v dvoch príbuzných sekvenciách si navzájom zodpovedajú?

(sequence alignment, zarovnávanie sekvencií)

## Populačná genetika

Mutácie sa šíria v populácii z rodičov na potomkov.

Nebezpečné mutácie rýchlejšie vymiznú, prospešné sa rýchlejšie ujmú (prírodný výber, natural selection).

**Polymorfizmus:** genetický rozdiel medzi organizmami v rámci druhu.

Vedie k rozdielnosti vo fenotype, napr. výzor, dedičné choroby.

Sekvenovaním viacerých jedincov toho istého druhu získame prehľad o polymorfizme.

**Bioinformatický problém:**

Nájdí polymorfizmus zodpovedný za určitý znak (napr. chorobu).

# Evolúcia

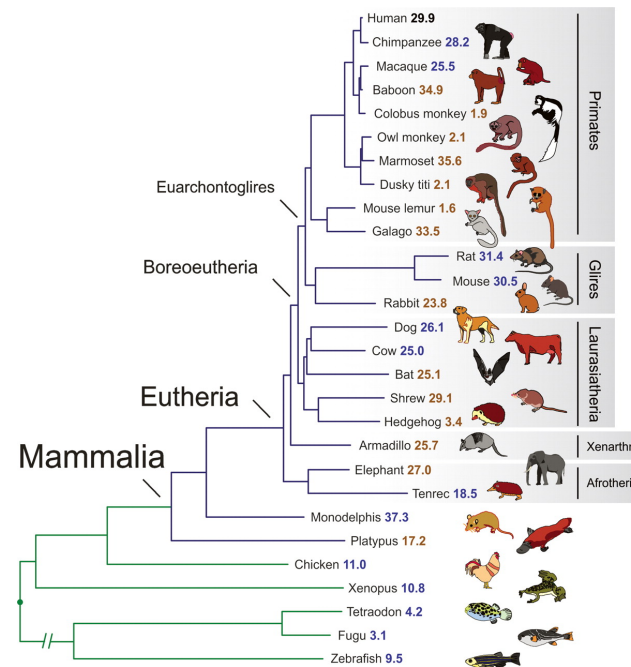
## Vznik nových druhov (speciation):

Po rozdelení populácie na viacero oddelených častí nedochádza k výmene genetického materiálu.

Hromadia sa zmeny až kým nie je možné párenie: vznik nových druhov.

## Bioinformatický problém:

Na základe dnešných sekvencií určí strom reprezentujúci vývoj druhov (fylogenetický strom, phylogenetic tree)



## Prokaryotické vs. eukaryotické organizmy

**Prokaryoty:** baktérie, jednoduché jednobunkové organizmy.

Nemajú jadro (DNA priamo v cytoplazme),  
majú kruhový chromozóm (a prípadné kratšie plasmidy),  
jednoduchšia štruktúra génu atď.

**Eukaryoty:** živočíchy, rastliny, huby, niektoré jednobunkové organizmy.

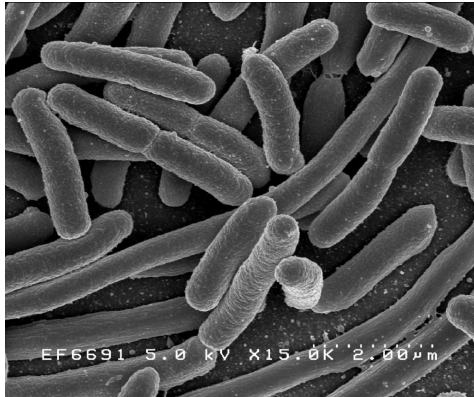
Bunka obsahuje jadro s DNA, viacero organel.

Mitochondrie a chloroplasty sú pohltené prokaryoty, ktoré sa stali časťou eukaryotickej bunky.

Dlhší genóm v niekoľkých lineárnych chromozómoch.

## Modelové organizmy

Dôležité pre biologický výskum, vieme o nich viac než o príbuzných druhoch.  
Poznatky širšie aplikovateľné.



**Escherichia coli:** baktéria žijúca v črevách. Jednoduchá manipulácia, delenie každých 20 min. Štúdium základných životných procesov: DNA replikácia, expresia génov, atď. Genóm s 4000 génmi, 4.6MB.



**Saccharomyces cerevisiae:** pekárske droždie. Jednoduchý eukaryotický organizmus. Genóm s 6000 génmi, 13MB. Delenie každé 2 hodiny. Štúdium špecificky eukaryotických javov.



## Modelové organizmy



**Arabidopsis thaliana:** malá kvitnúca rastlina, 6-týždňový životný cyklus. Skúmanie javov špecifických pre rastliny.

**Caenorhabditis elegans:** malý červ, nematód, žijúci v pôde. Štúdium vývinu (ontogenéza, development), diferenciácie buniek.

**Drosophila melanogaster:** vínna muška. Štúdium genetiky, gény riadiace vývin jedinca.

**Stavovce:** žaba *Xenopus laevis* (veľké, ľahko manipulovateľné vajíčka), akvariijná ryba *Danio rerio* (priehľadné embryá), myš *Mus musculus* (existuje veľa plemien so špeciálnymi vlastnosťami).

## Dostupné dáta

- DNA sekvencie: celé genómy, ich časti
- Ich anotácia: súradnice génov a iných funkčných častí
- Sekvencie RNA, ich štruktúra
- Sekvencie proteínov, ich funkcia a štruktúra
- Merania množstva RNA/proteínu v bunke
- ...

Dáta založené na experimentoch alebo výsledky výpočtových metód

Veľa chýb (v oboch prípadoch)

## Ďalšie informácie

- Zvelebil, Baum: Understanding Bioinformatics, kap. 1
- Vysokoškolské učebnice molekulárnej biológie
- Anglická wikipédia
- Tutoriály na stránke predmetu