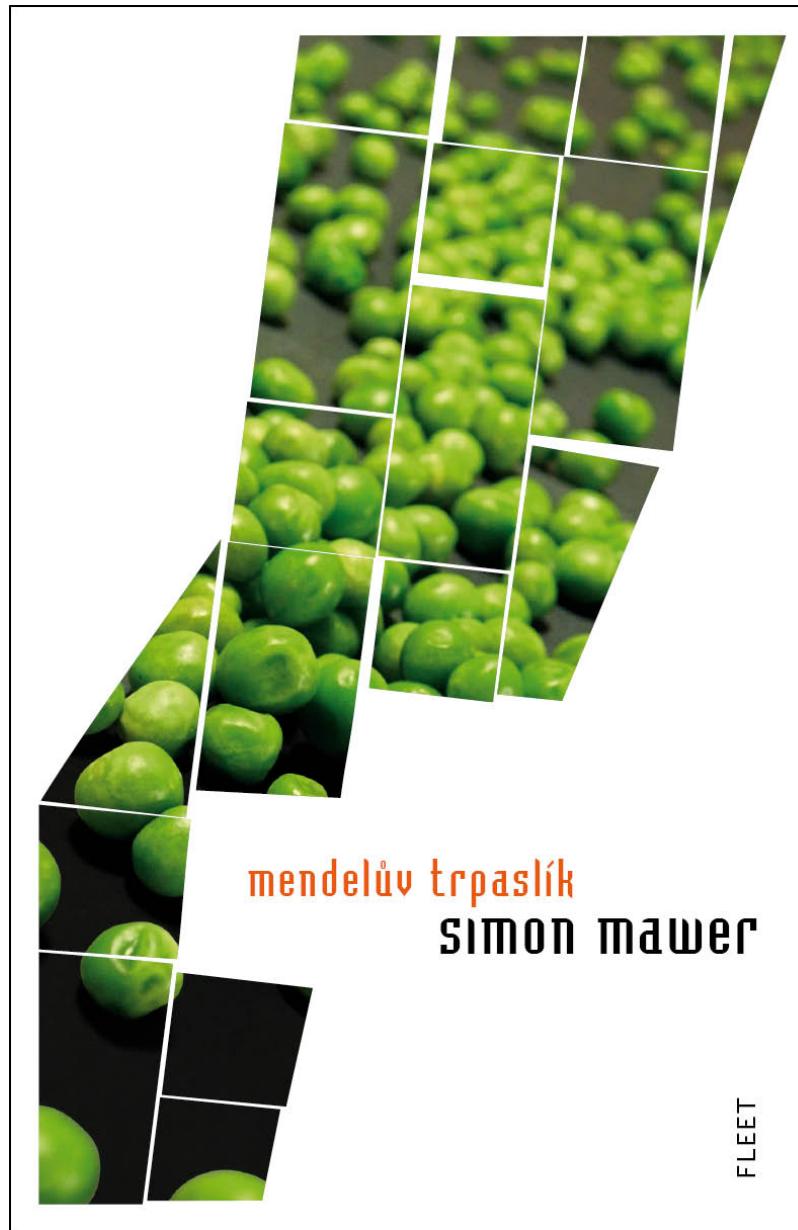


Úvod do genomiky



mendelův trpaslík
SIMON MAWER

FLEET

10.12.2015

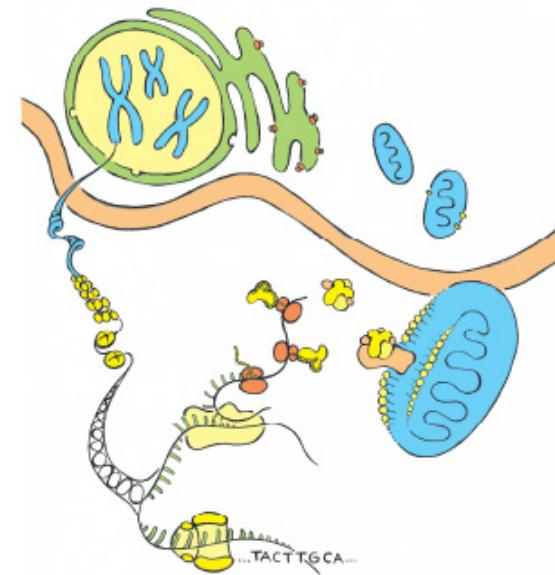
GENETIKA

D. PETER SNUSTAD
MICHAEL J. SIMMONS

MASARYKOVÁ UNIVERZITA

Genomika

Nosek Brejová Neboháčová Baráth
Bhatia-Kišsová Valent Kollár Tomáška



www.createspace.com/4520295 (f)
www.createspace.com/4535137 (čb)

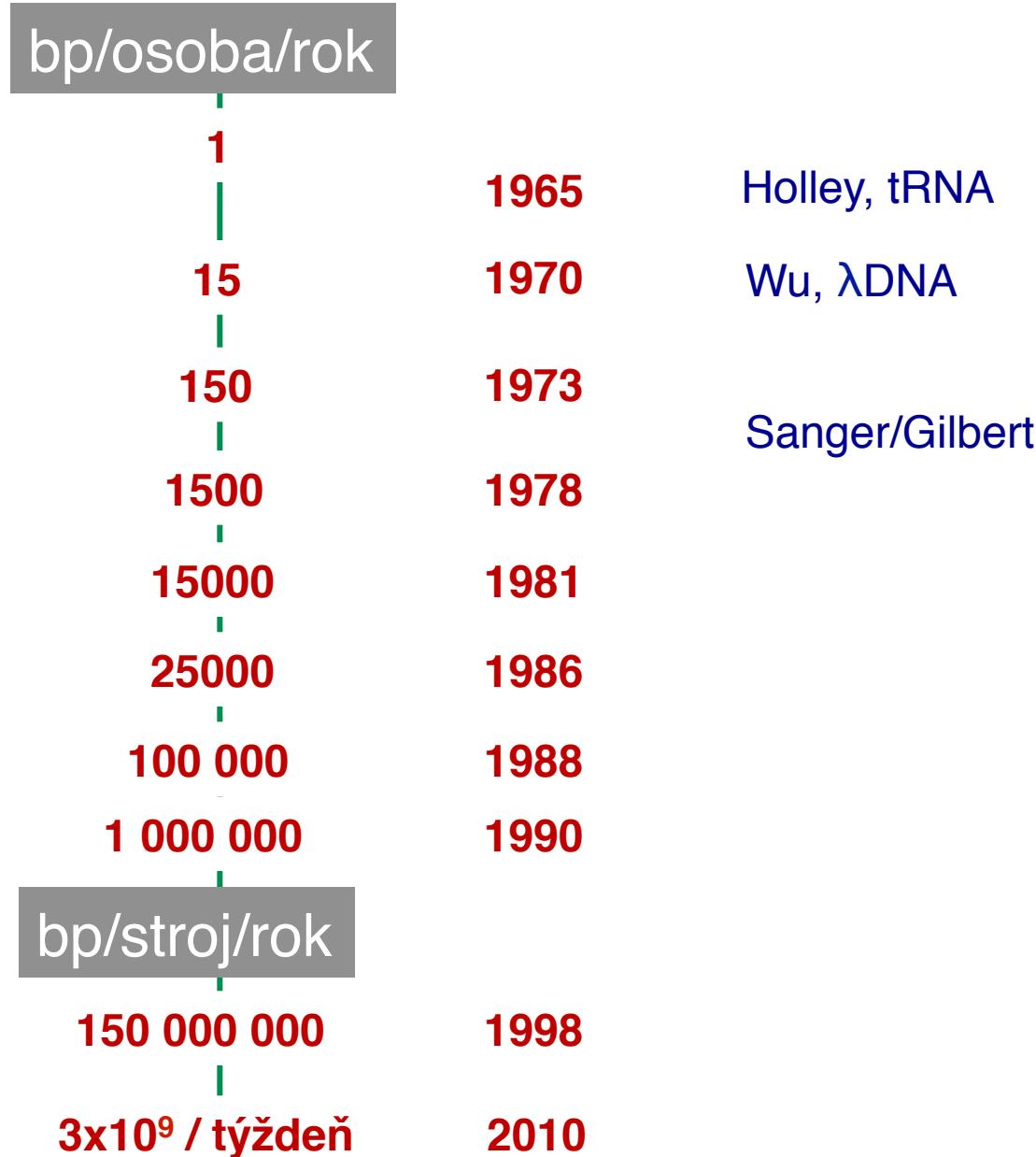
Prehl'ad prednášky

- ▶ Čo je genomika?
- ▶ Porovnanie genetických, cytologických a fyzikálnych máp chromozómov
- ▶ Pozičné klonovanie génov
- ▶ Ľudský genómový projekt
- ▶ Komparatívna (porovnávacia) genomika

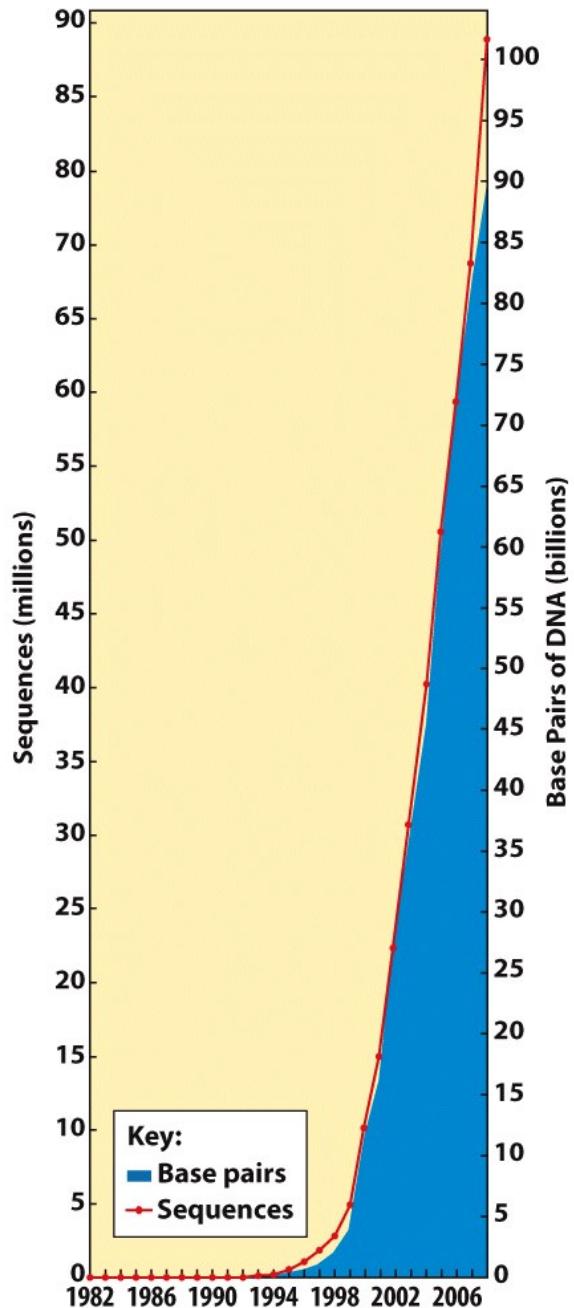
Genóm: kompletná genetická informácia daného organizmu

Genomika: disciplína zameraná na štruktúru (**štruktúrna genomika**), funkciu (**funkčná genomika**) a porovnávaciu analýzu
(komparatívna genomika)
genómov

Zvyšovanie efektívnosti sekvenovania DNA



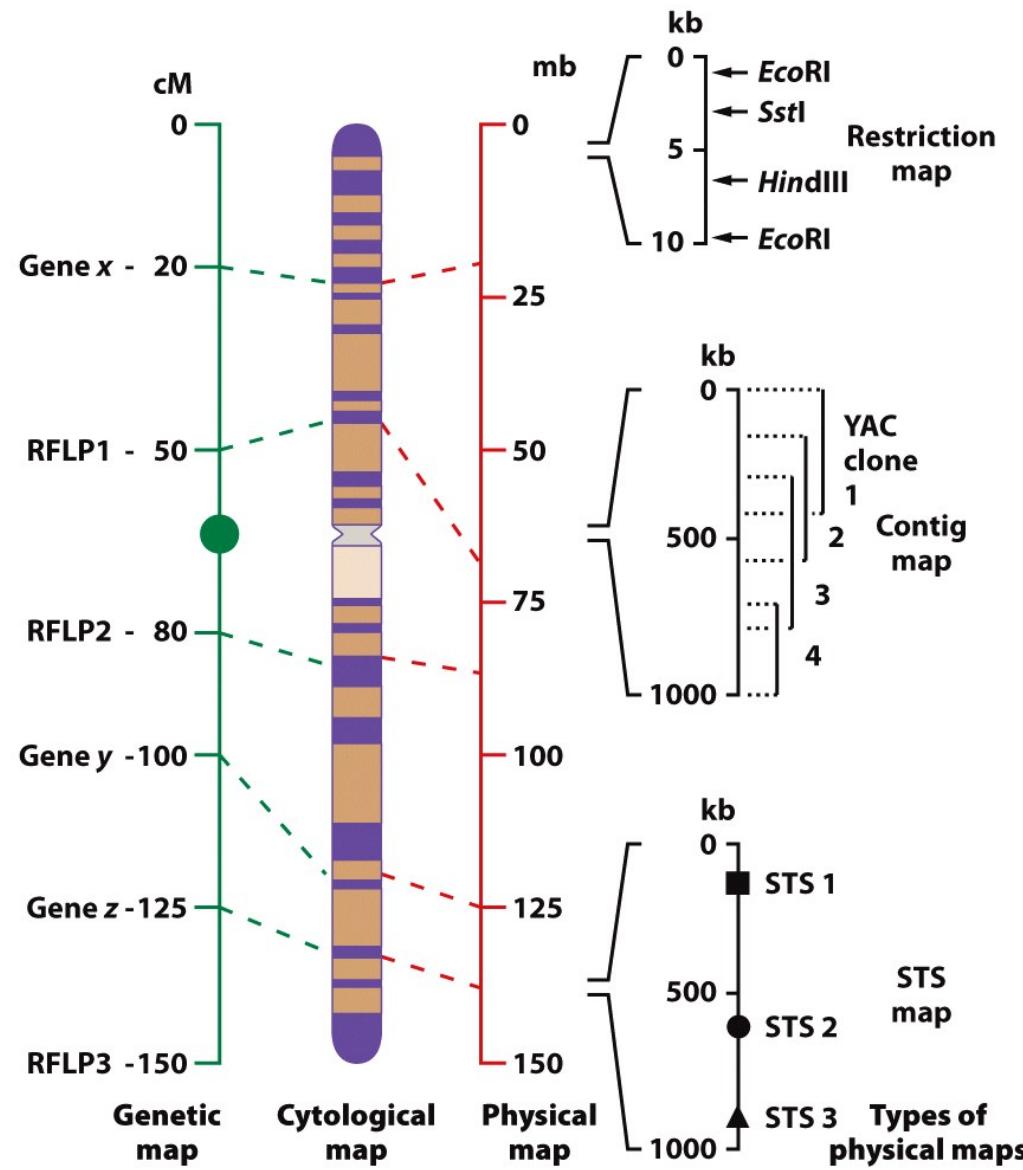
Nárast počtu „prečítaných“ nukleotidov prebieha podľa Moorovho zákona



Kompletne sekvenované genómy

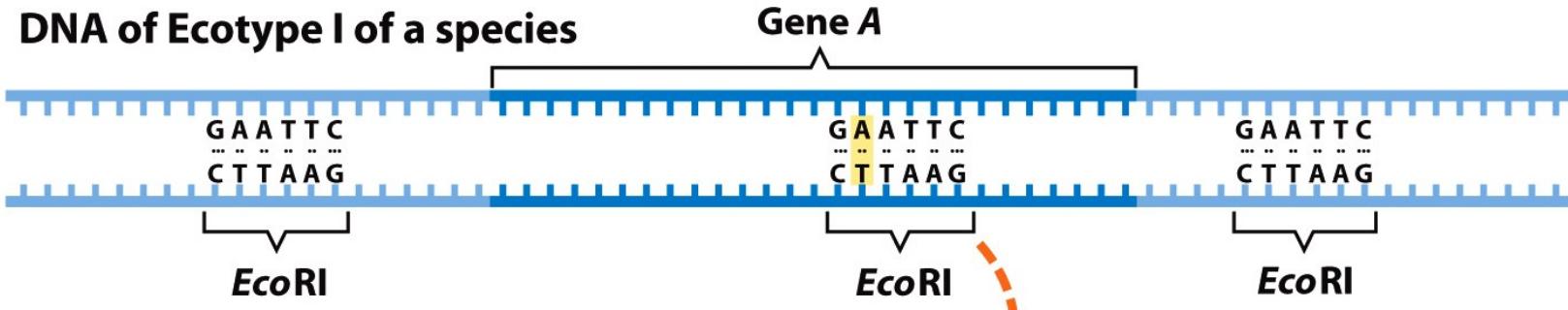
- ▶ >2000 vírusov
- ▶ >1500 plazmidov
- ▶ ~1500 mitochondriálnych genómov
- ▶ ~150 chloroplastových genómov
- ▶ ~120 genómov *Archaea*
- ▶ ~700 genómov *Eubacteria*
- ▶ ~50 eukaryotických genómov, včítane genómu človeka

Porovnanie genetických, cytogenetických a fyzikálnych máp chromozómov



Polymorfizmus restrikčných fragmentov (RFLP) ako nástroj fyzikálneho mapovania

Mutational origin of an RFLP

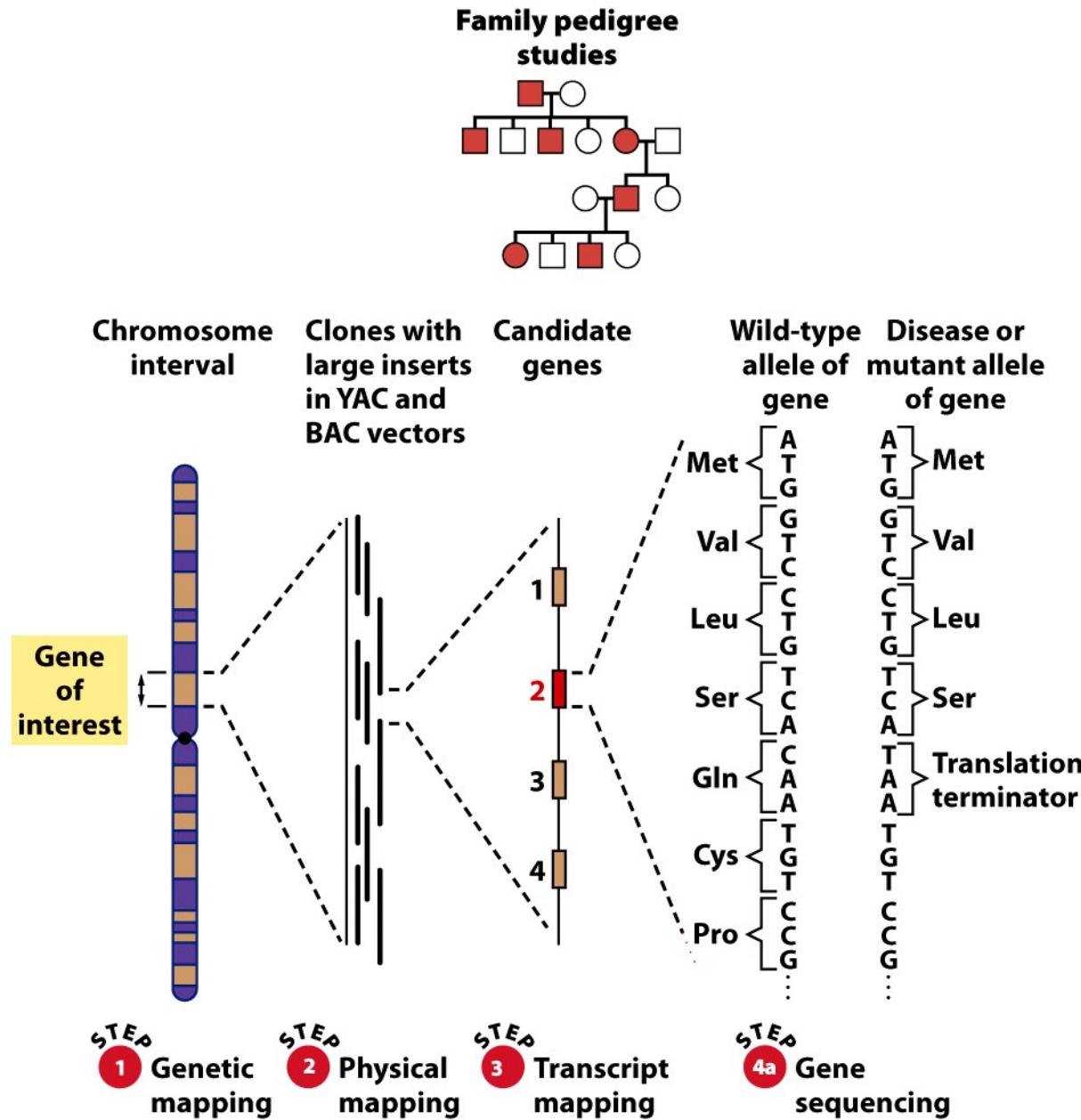


Súhrn 1

- ▶ Genetické mapy sú založené na frekvencii rekombinácie medzi markérovými génnimi
- ▶ Cytologické mapy sú konštruované na základe pozície markérov v blízkosti oblasti vizualizovanej svetelnou mikroskopiou
- ▶ Fyzikálne mapy sú predstavované vzdialenosťou medzi markérmi meranou v počte párov báz.

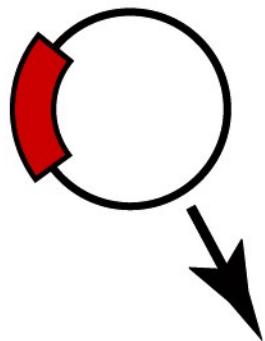
Pozičné klonovanie má za cieľ identifikovať gén s využitím informácií o jeho polohe v príslušnej mape

Schéma pozičného klonovania



Ako je možné dokázať, že identifikovaný gén je zodpovedný za patologický fenotyp?

Štandardný
(WT) gén



Plazmidový
vektor

MUTANT

→
ŠTANDARDNÝ
FENOTYP

Súhrn 2

- ▶ Kombináciou konštrukcie detailných genetických, cytologických a fyzikálnych máp spolu so stratégiami pozičného klonovania je možné identifikovať príslušný gén v genóme

L'udský genómový projekt (The Human Genome Project, HUGO)

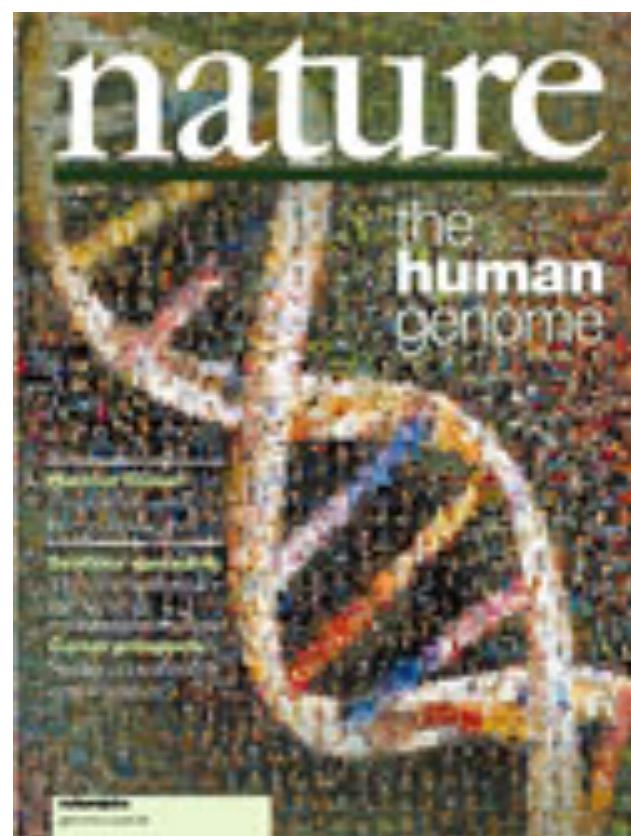
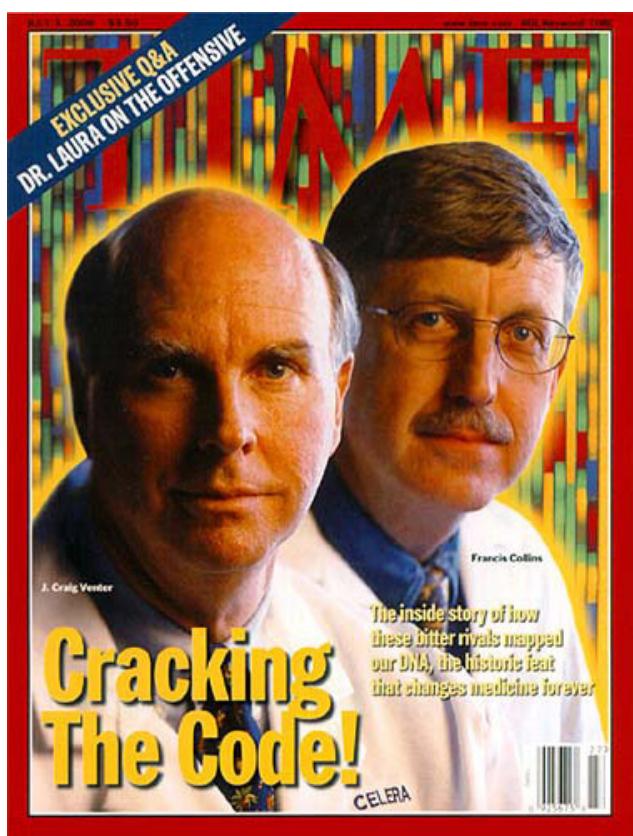
CIELE:

- Zmapovať všetky l'udské gény
- Skonštruovať detailnú fyzikálnu mapu celého l'udského genómu
- Stanoviť poradie nukleotidov všetkých 24 l'udských chromozómov

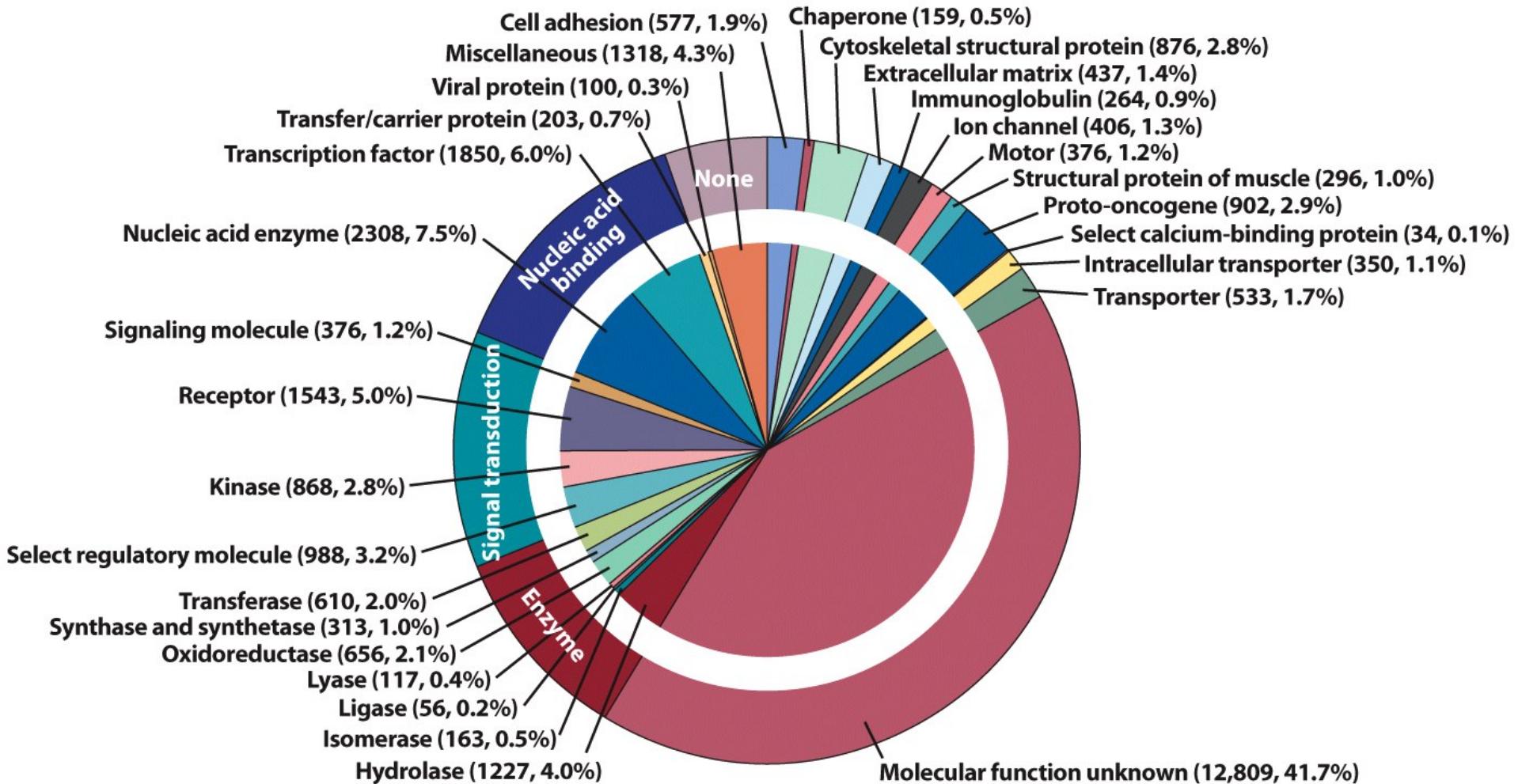
ALTERNATÍVA:

- „*Shot-gun*“ sekvenovanie (Craig Venter)
<http://www.youtube.com/watch?v=vg7Y5EeZsjk>

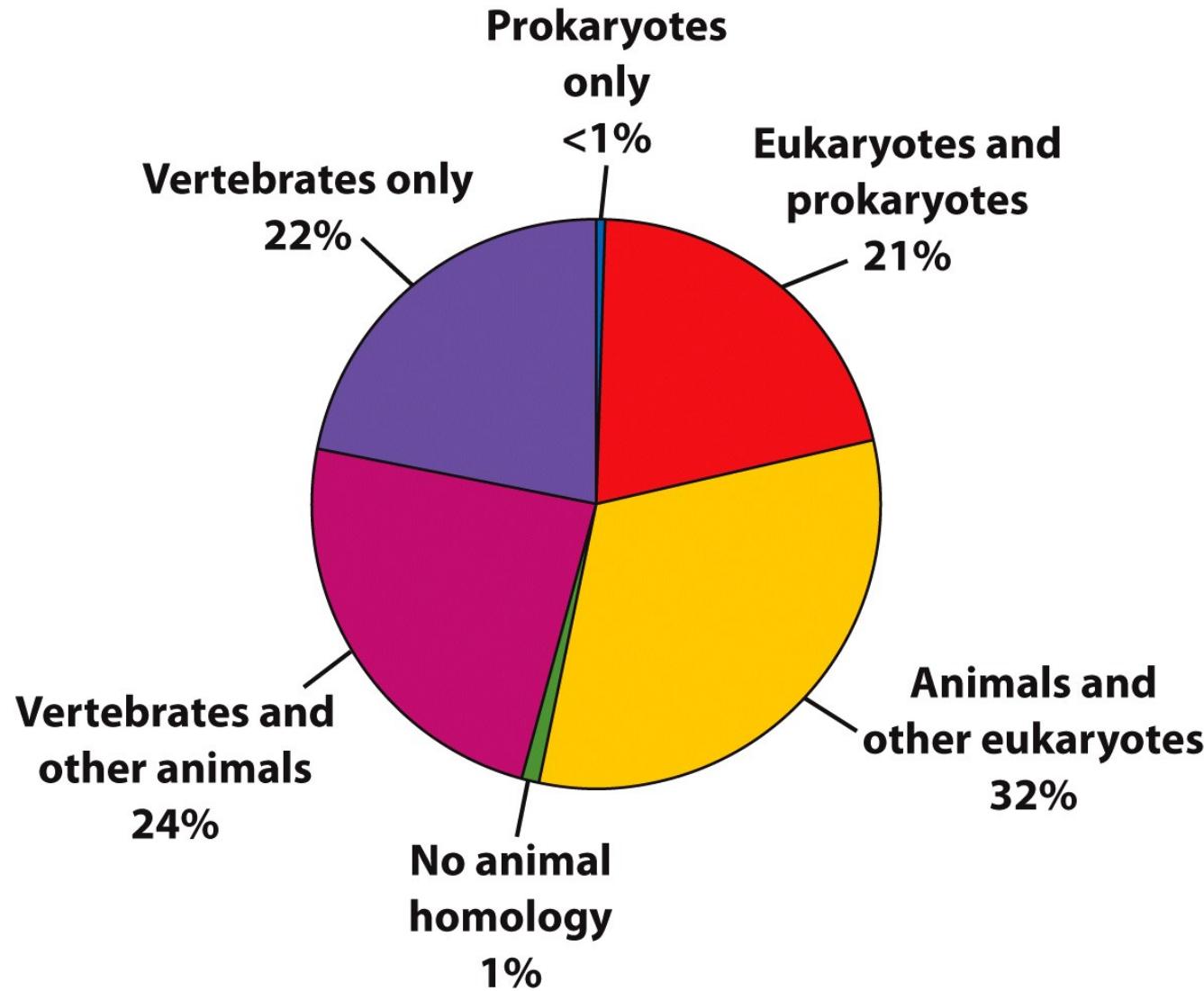
Február 2001: Prvé verzie ľudského genómu



Funkčná klasifikácia ľudských proteín-kódujúcich génov



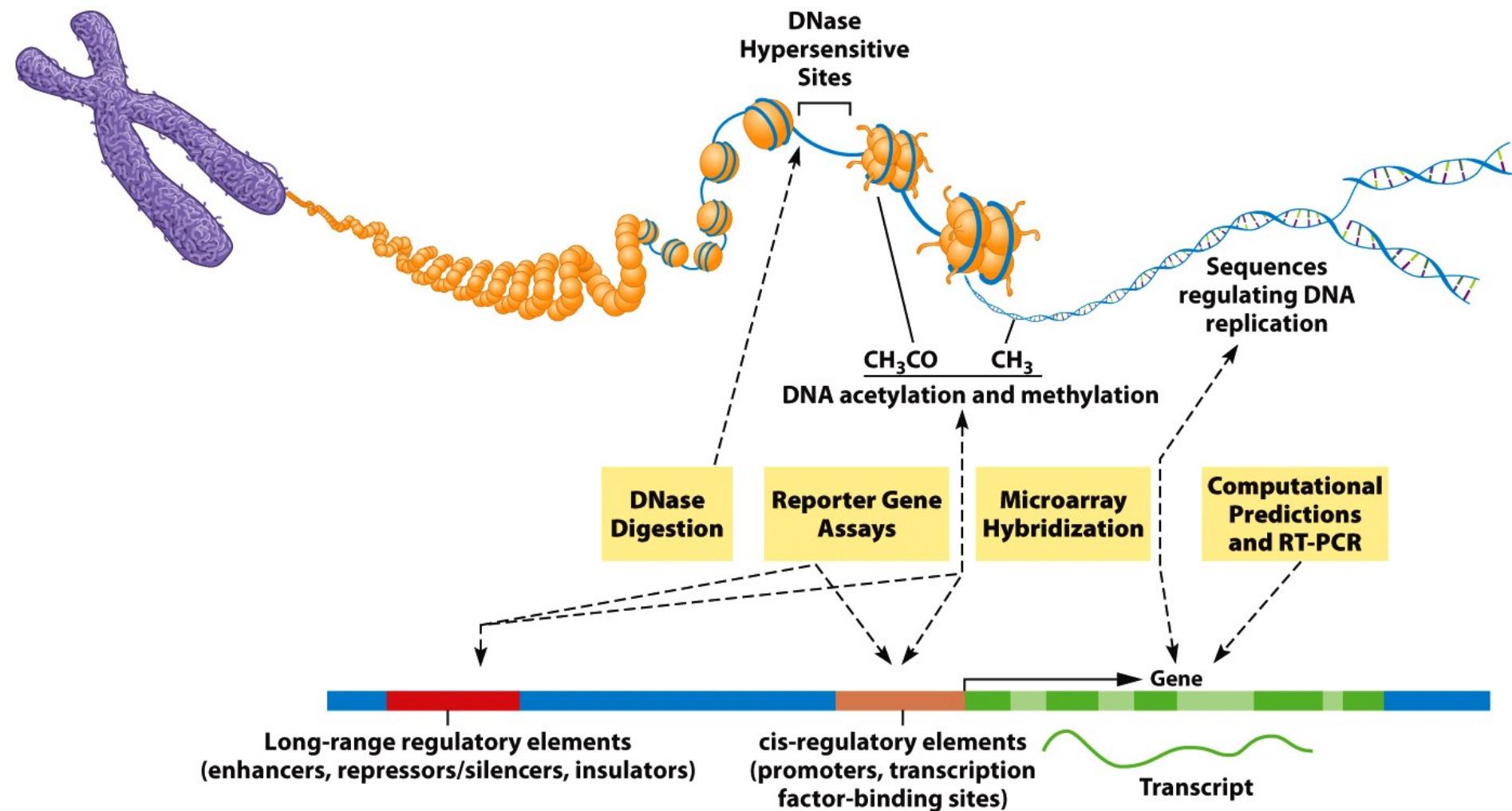
Homológia (podobnosť) ľudských proteínov s proteínmi rôznych skupín organizmov



Fakty o ľudskom genóme

- ▶ V priemere 1 proteín-kódujúci gén na 145 000 bp (pozor: klastre génov)
- ▶ ~22000 proteín kódujúcich génov; a zatiaľ ľažko odhadnuteľný počet génov pre RNA (tRNA, rRNA, ďalšie nekódujúce RNA)
- ▶ Priemerná dĺžka génu: 27000 bp (najdlhší gén pre dystrofín: 2 400 000 bp)
- ▶ Priemerný počet exónov na 1 gén: 9 (najviac v géne pre kolagén VII: 115)
- ▶ 1.1% genómu tvoria exóny, 24% intróny (najdlhší intrón 104 000 bp) a 75% medzigénová DNA
- ▶ 44% medzigénovej DNA je derivovaná z transpozónov

Projekt ENCODE (*EN*Cyclopedia Of DNA Elements): Identifikácia negénových funkčných elementov



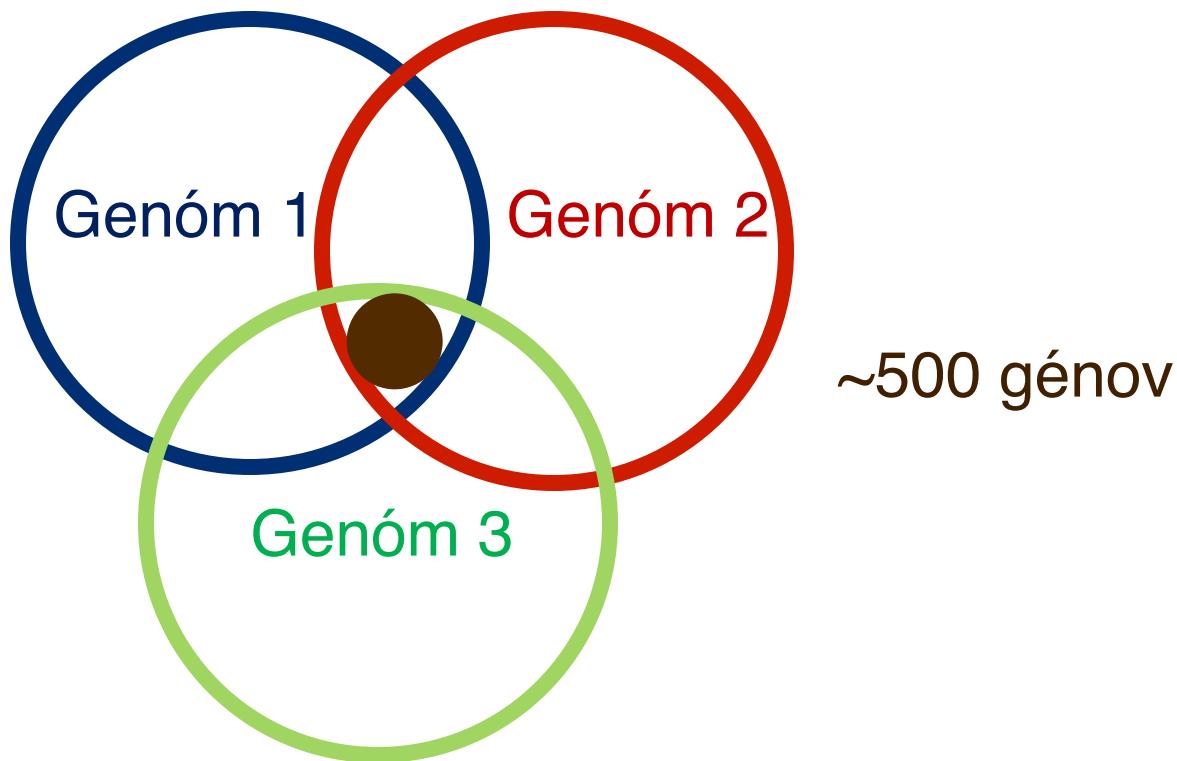
Komparatívna genomika

- ▶ Porovnanie poradia nukleotidov genómov rôznych organizmov umožňuje konštrukciu fylogenetických stromov a sledovanie molekulárnych zmien vedúcich k jednotlivým fylogenetickým líniám.
- ▶ BIOINFORMATIKA: spracovanie genomických, proteomických, ... –omických dát informatickými nástrojmi

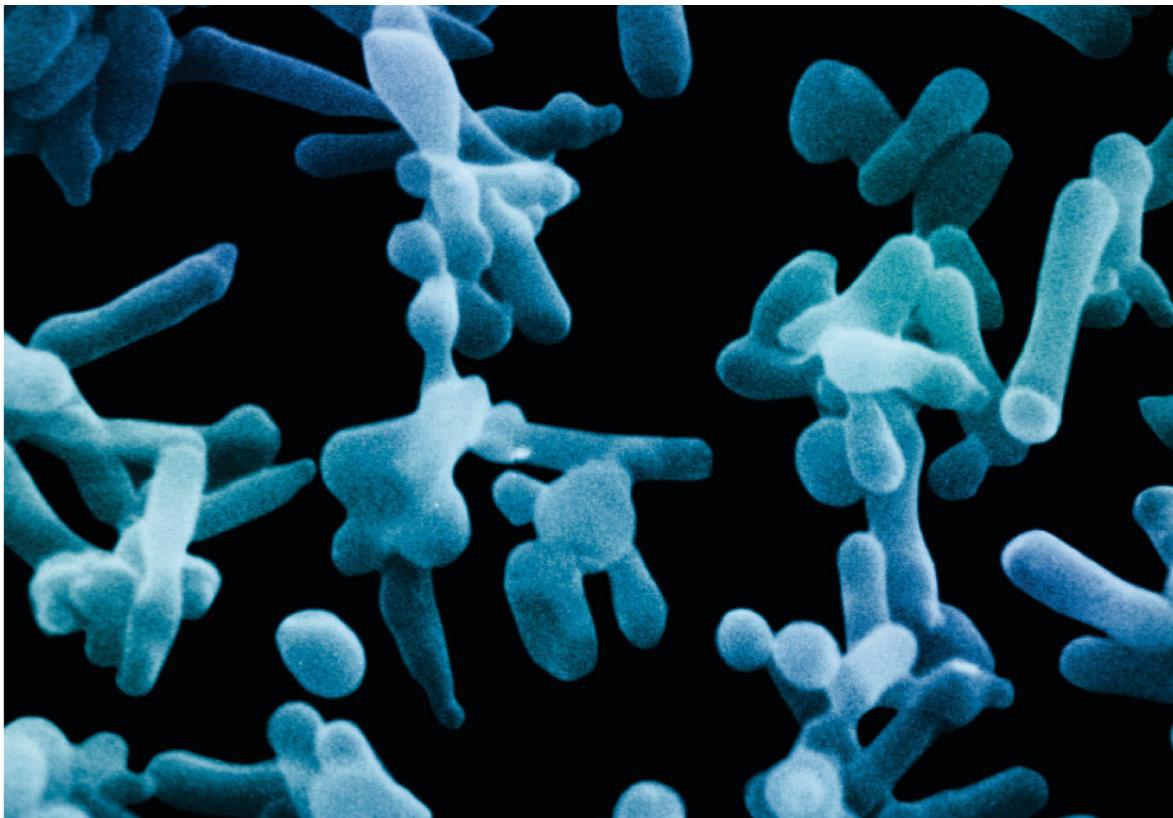
Veľkosť genómov môže byť odlišná aj u veľmi príbuzných organizmov

Size and Gene Content of Selected Prokaryotic Genomes		
Species	Genome Size in Nucleotide Pairs	Predicted Number of Genes
Archaea		
<i>Archaeoglobus fulgidus</i>	2,178,400	2,486
<i>Methanosaarcina acetivorans</i>	5,751,492	4,721
<i>Nanoarchaeum equitans</i>	490,885	582
<i>Pyrococcus furiosus</i>	1,908,256	2,228
<i>Sulfolobus solfataricus</i>	2,992,245	3,033
<i>Thermoplasma volcanium</i>	1,584,804	1,548
Eubacteria		
<i>Bacillus subtilis</i>	4,214,630	4,225
<i>Bordetella parapertussis</i>	4,773,551	4,467
<i>Bradyrhizobium japonicum</i>	9,105,828	8,373
<i>Buchnera aphidicola</i>	615,980	550
<i>Chlamydia pneumoniae</i> , strain AR39	1,229,853	1,167
<i>Escherichia coli</i> , strain K12 MG1655	4,639,675	4,467
<i>Escherichia coli</i> , strain O157 EDL933	5,528,970	5,463
<i>Haemophilus influenzae</i> Rd KW20	1,830,138	1,789
<i>Legionella pneumophila</i> , strain Paris	3,503,610	3,136
<i>Mycobacterium tuberculosis</i> , strain CDC	4,403,837	4,293
<i>Mycobacterium genitalium</i>	580,076	525
<i>Neisseria meningitidis</i> Z2491	2,184,406	2,208
<i>Pseudomonas syringae</i> strain DC3000	6,397,126	5,660
<i>Rickettsia typhi</i>	1,111,496	919
<i>Salmonella typhimurium</i>	4,857,432	4,622
<i>Staphylococcus aureus</i> , strain MW2	2,820,462	2,712
<i>Streptomyces coelicolor</i>	8,667,507	7,912
<i>Ureaplasma parvum</i> ATCC 700970	751,719	653
<i>Yersinia pestis</i> , strain KIM	4,600,755	4,240

Genomika môže pomôcť odpovedať na otázku, aká je genetická výbava „minimálnej“ bunky?



Aká je genetická výbava „minimálnej“ bunky?



0.4 μm

Mycoplasma genitalium: 477 génov

Genómy eukaryotov

► *Saccharomyces cerevisiae*

- Prvá kompletná sekvencia eukaryotického genómu (1996)
- Dĺžka genómu: 12,068 kb
- 5885 proteín-kódajúcich génov , 140 rRNAs, 40 snRNA

► Ďalšie eukaryotické genómy

- *Caenorhabditis elegans* (1998)
- *Drosophila melanogaster* (2000)
- *Arabidopsis thaliana* (2000)
- *Homo sapiens* (2004)
- ...

Genómy eukaryotov sa odlišujú veľkosťou i počtom génov

Size and Predicted Gene Content of Selected Eukaryotic Genomes				
Species	Common Name	Genome Size in Nucleotide Pairs	Predicted Number of Genes*	Gene Density (bp/gene) [†]
Protists				
<i>Encephalitozoon cuniculi</i>	microsporidian	2,497,519	2,029	1,200
<i>Plasmodium falciparum</i>	malaria protozoan	22,820,308	5,361	4,300
Fungi				
<i>Candida glabrata</i>	yeast (infectious)	12,280,357	5,272	2,500
<i>Saccharomyces cerevisiae</i>	baker's yeast	12,057,909	6,268	1,900
Nematode				
<i>Caenorhabditis elegans</i>	roundworm	100,291,841	20,516	4,900
Insects				
<i>Anopheles gambiae</i>	mosquito	278,253,050	14,707	18,900
<i>Apis mellifera</i>	honeybee	197,657,892	29,832	6,600
<i>Drosophila melanogaster</i>	fruit fly	131,000,899	13,792	9,500
Plant				
<i>Arabidopsis thaliana</i>	mouse ear cress	119,186,496	28,152	4,200
Vertebrates				
<i>Canis familiaris</i>	dog	2,359,826,366	18,201	129,700
<i>Danio rerio</i>	zebra fish	1,571,018,465	23,524	66,800
<i>Gallus gallus</i>	chicken	1,054,180,845	17,709	59,600
<i>Homo sapiens</i>	human	2,851,330,913	22,287	127,900
<i>Mus musculus</i>	mouse	2,932,368,526	25,396	115,500
<i>Ornithorhynchus anatinus</i>	platypus	1,840,000,000	18,527	99,500
<i>Pan troglodytes</i>	chimpanzee	2,928,563,828	21,098	139,000
<i>Rattus norvegicus</i>	Norwegian rat	2,571,104,688	22,159	115,800
<i>Takifugu rubripes</i>	Japanese pufferfish	329,140,338	20,796	15,800
<i>Tetraodon nigroviridis</i>	spotted pufferfish	402,240,326	27,918	14,400

Blast: nástroj na analýzu sekvencí DNA a proteínov

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

 NCBI

Nucleotide Protein Translations Retrieve results for
an RID

What is MegaBLAST?

ATGAGAGAAATTCTTCATATTCAAGGAGGTCAAGTGCAGAACCCAGATCGGAGCTAAGTTCTC

Search

Load query file from disk

Set subsequence

Choose database

Return alignment endpoints only

Now: or

The MegaBLAST web site at NCBI.

Blast porovnáva sekvenciu DNA, alebo proteínu so sekvenciami v databáze

Query: atgagagaaattttcatattcaaggaggtcagtgcggaaaccagatcgagactaaggttctggaaagtta-tttgcggcgagcacgttattgtacaaaccg 100
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

M84706.1 atgagagaaattttcatattcaaggaggtcagtgcggaaaccagatcgagactaaggttctggaaagtta-tttgcggcgagcacgttattgtacaaaccg 227
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

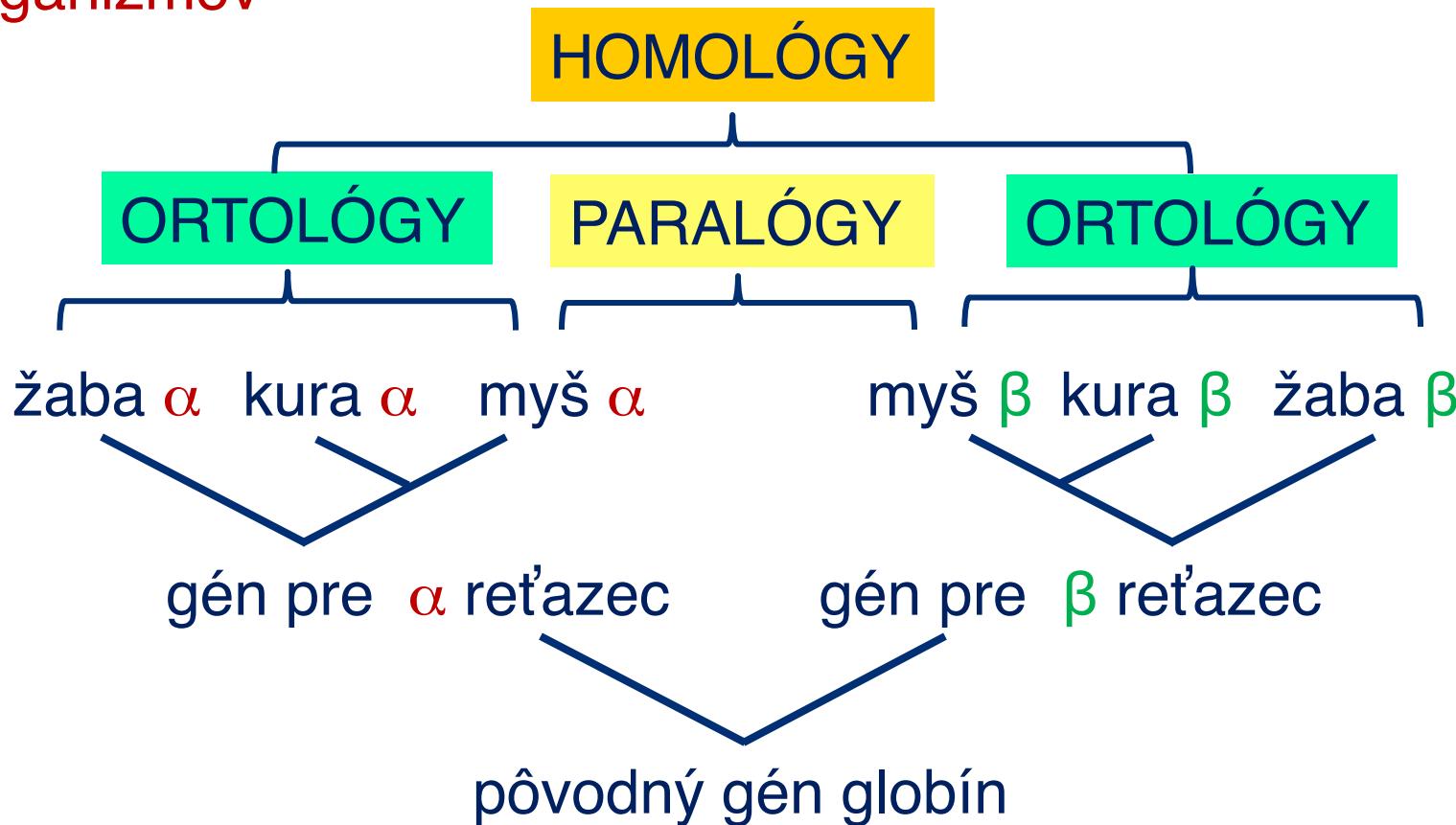
Query: atgagagaaattttcatattcaaggaggtcagtgcggaaaccagatcgagactaaggttctggaaagtta-tttgcggcgagcacgttattgtacaaaccg 93
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

NM_122291.2 atgcgagagattttcacatacaagggtggccatgcggaaaccagatcgagactaaggttctggaaagtta-aggttgcgcgagcacggatcgt 172
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Nucleotide alignments of the query sequence with two sequences identified by the MegaBLAST search.

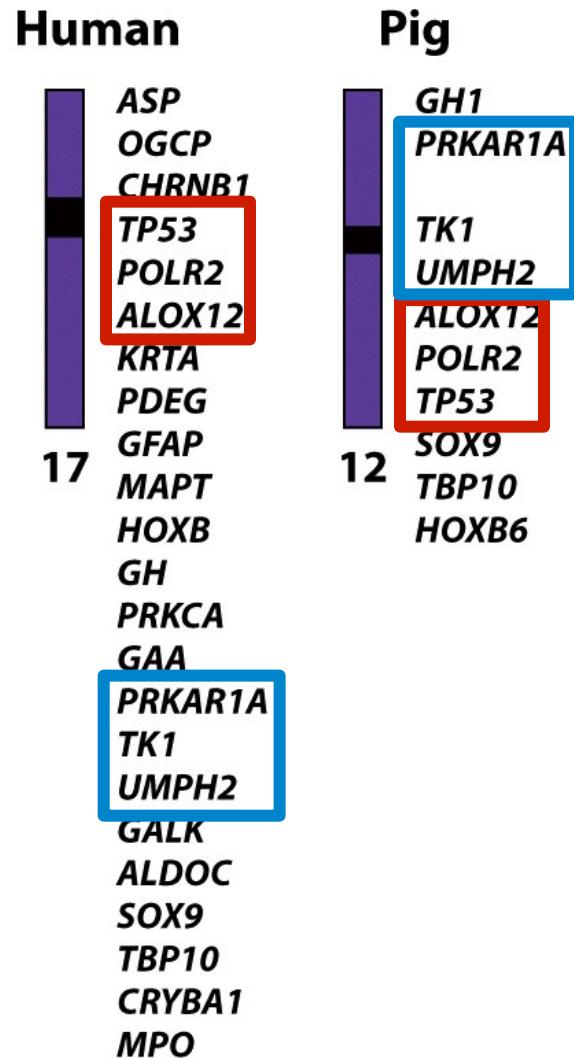
Homologické gény majú podobnú sekvenciu DNA, pretože majú spoločnú evolučnú história

- ▶ PARALÓGY: Homologické gény v tom istom organizme
- ▶ ORTOLÓGY: Homologické gény v dvoch rôznych druhoch organizmov

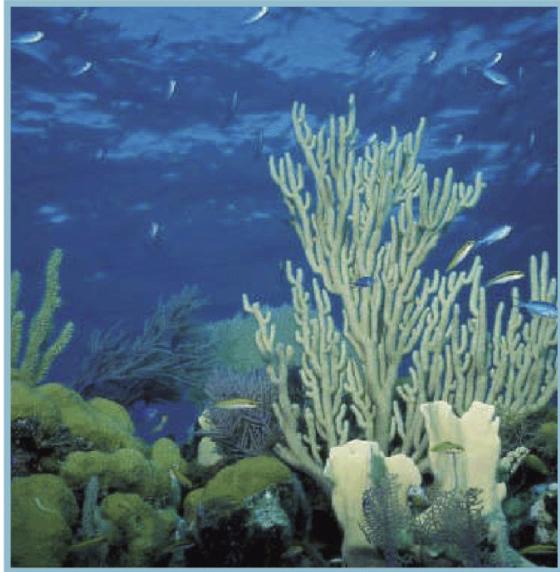


SYNTÉNIA: konzervovanie blokov genómov medzi rôznymi druhmi organizmov: príklad u cicavcov

Homo sapiens *Sus scrofa*



Metagenomika umožňuje identifikáciu nových foriem bez toho, aby ich bolo nevyhnutné množiť v laboratóriu



Sekvenovanie celej extrahovanej DNA s cieľom zistiť, aké gény sa nachádzajú vo vzorke

Zistovanie funkcie a diverzity identifikovaných génov

Neexistuje viac domén života?

Pre-genomic era



Microscopy



rRNA-based molecular taxonomy

Cultivation-independent shotgun genomics

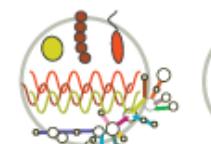
Next-generation sequencing



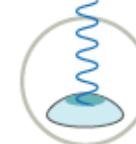
Shotgun metagenomics



Single-cell genomics



Meta-transcriptomics



Single-molecule sequencing

Targeted approaches for finding new branches

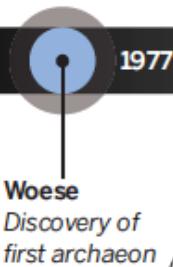
Large-scale assemblies of

- Environmental nucleic acid sequence
- Single cells lacking amplifiable rRNA genes

Advanced analyses

- K-mer
- tRNA structure
- Codon usage
- Phylogeny
- Detection of non-canonical bases

Single-molecule sequencing
of environmental nucleic acids



1990
Woese
Three domains

2004

2005

Now



Synthetic life parallel
to the existing domains?



Is there a
fourth domain?

Organizmy:

1. Preskúmané (*explored*)
2. Nepreskúmané (*unexplored*)
3. Neidentifikované (*undiscovered*)

Komparatívna genomika je nástrojom pre
rekonštrukciu evolúcie

Ako vyzeral genóm spoločného predka primátov?



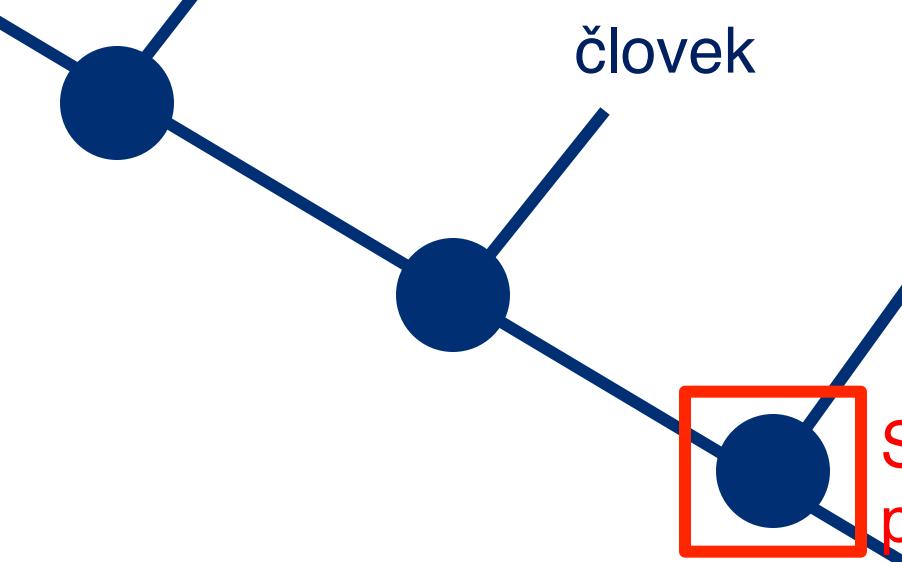
šimpanz

bonobo

človek

gorila

Spoločný
predok (SP)



Komparatívna genomika umožňuje odhalovať aj tajomstvá vyhynutých organizmov

Homologické gény/proteíny:

človek

GTCCAA**A**GATCCCCATCGATC

šimpanz

GTCC**A**GATCCCTATCGACC

gorila

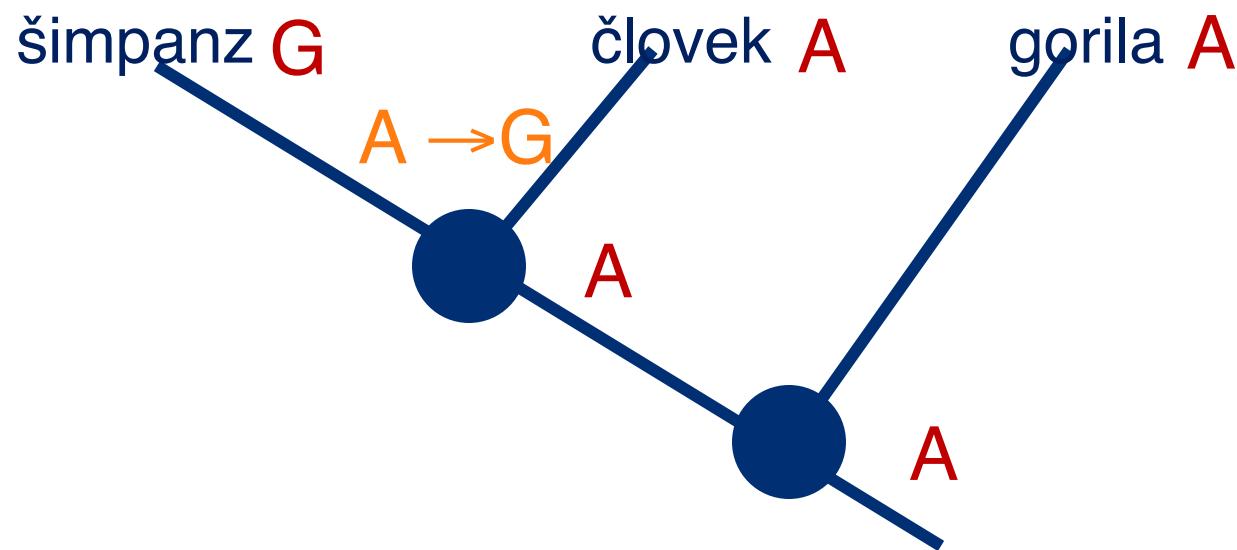
GTCCAA**A**GATCCCTATCGATC

SP

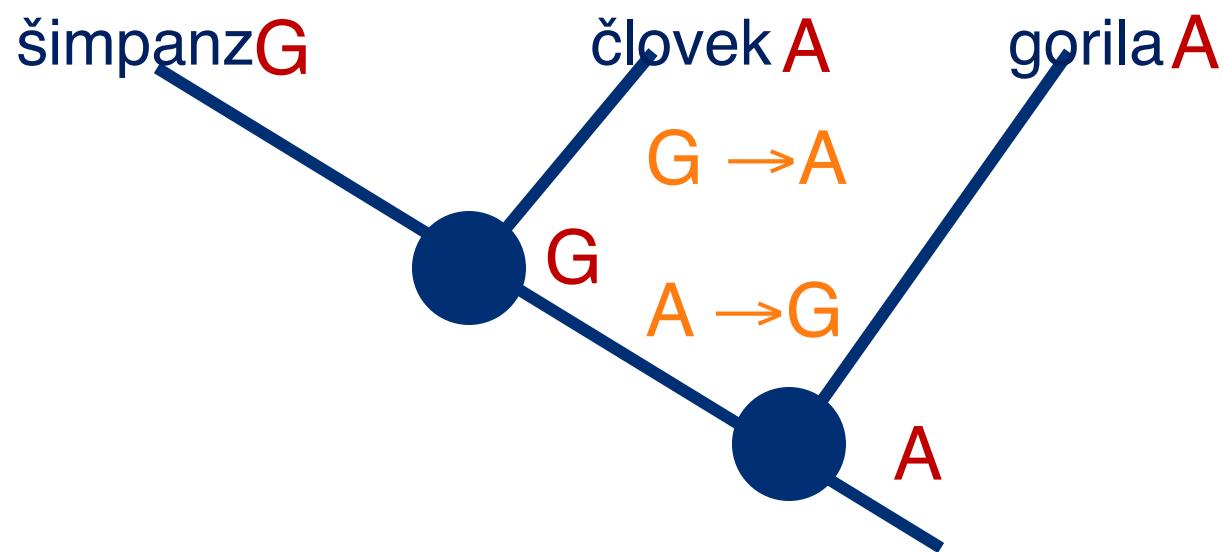
GTCCAA**X**GATCCC**X**ATCG**A**C

SP=spoločný predok

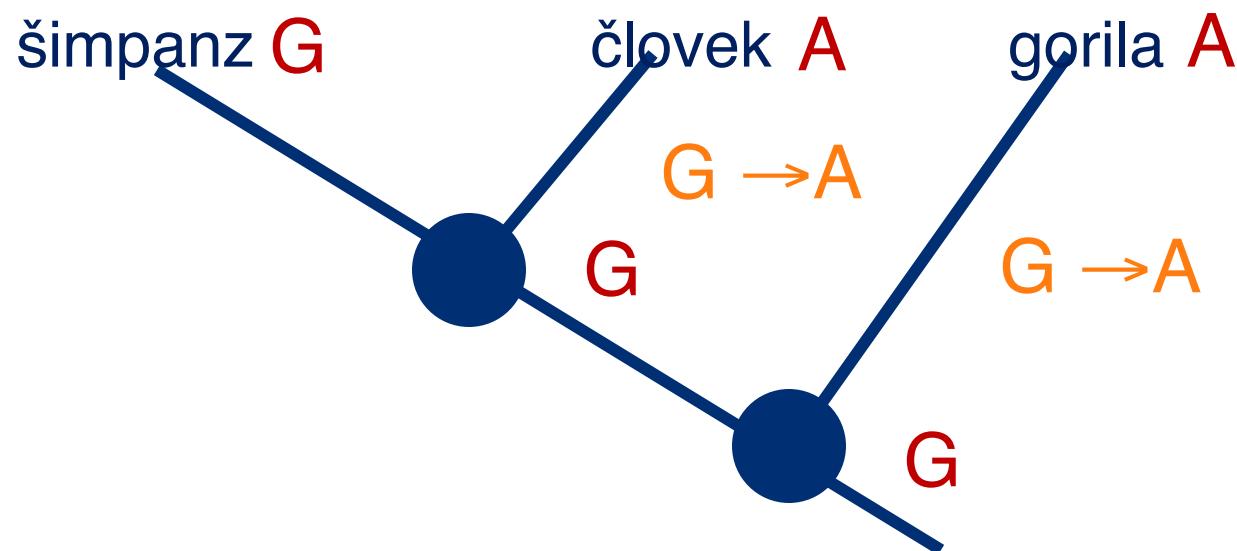
Fylogenetický strom vybraných primátov



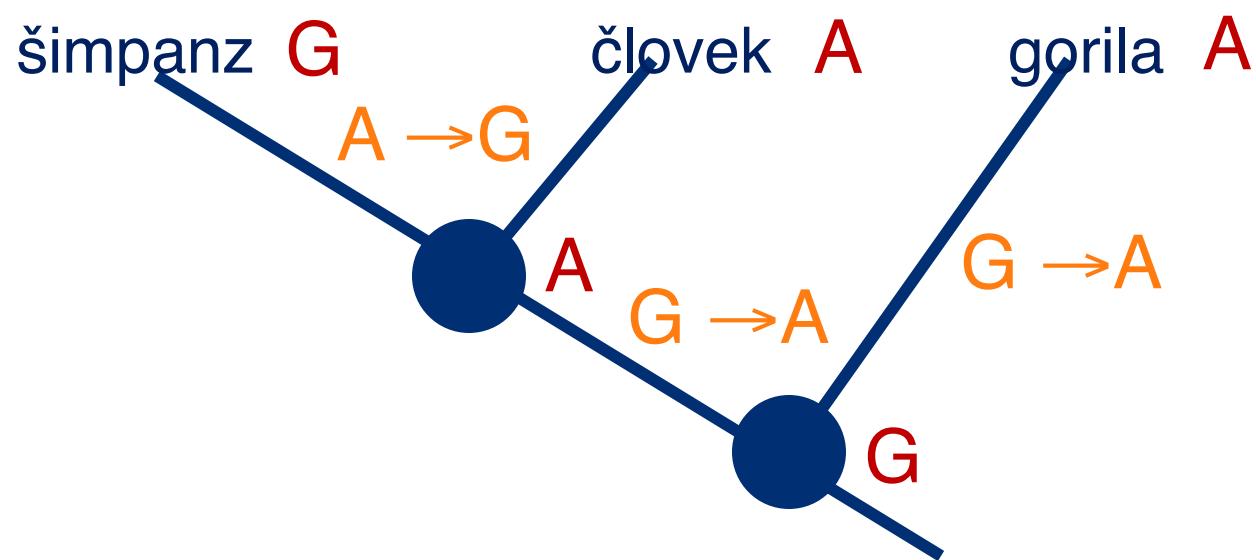
Fylogenetický strom vybraných primátov



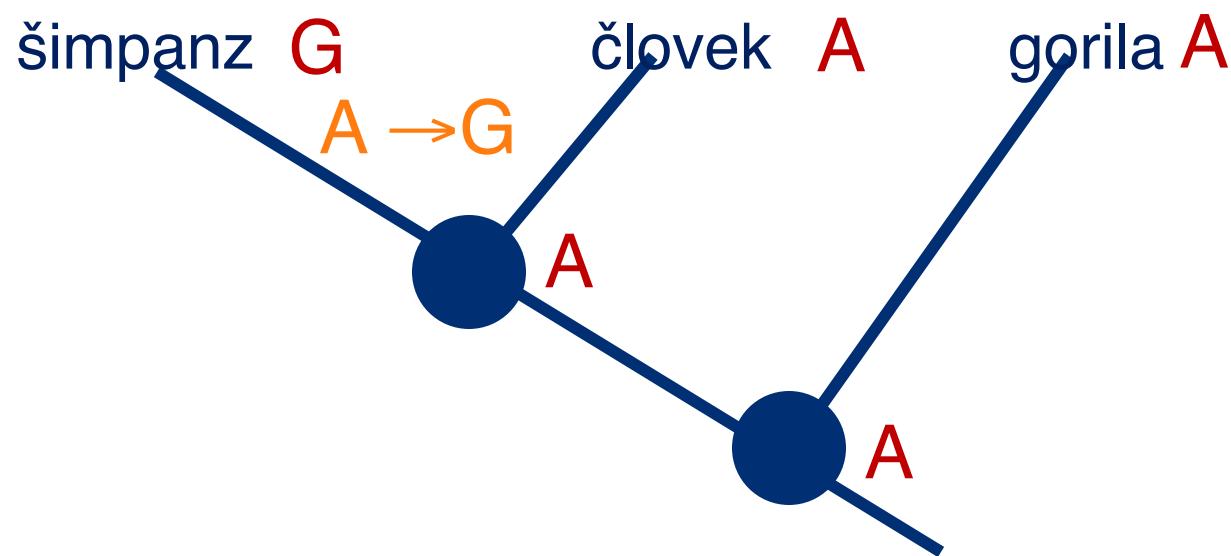
Fylogenetický strom vybraných primátov



Fylogenetický strom vybraných primátov



Parsimónia: najjednoduchšie vysvetlenie fylogenetických údajov



Molekulárna taxonómia

Homologické gény/proteíny:

človek

GTCCAAGATCCCCATCGATC

šimpanz

GTCCAGGATCCCTATCGACC

gorila

GTCCAAGATCCCTATCGATC

SP

GTCCAXGATCCCXATCGAXC

SP=spoločný predok

Molekulárna taxonómia

Homologické gény/proteíny:

človek

GTCCAAGATCCCCATCGATC

šimpanz

GTCCAGGATCCCTATCGACC

gorila

GTCCAAGATCCCTATCGATC

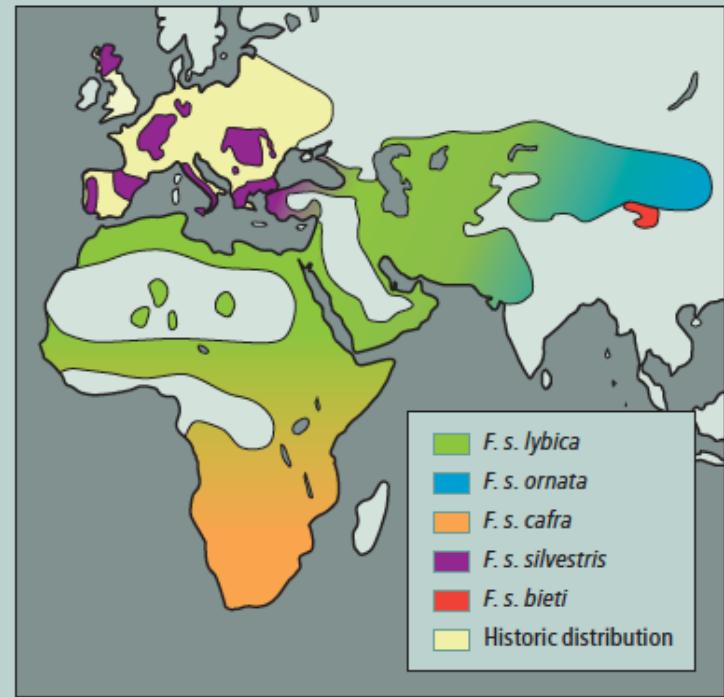
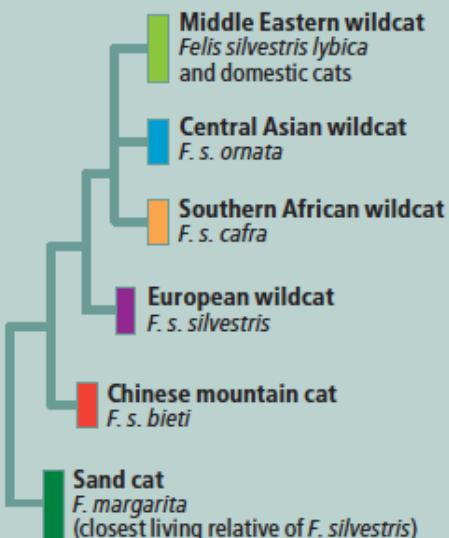
SP

GTCCAAGATCCCTATCGATC

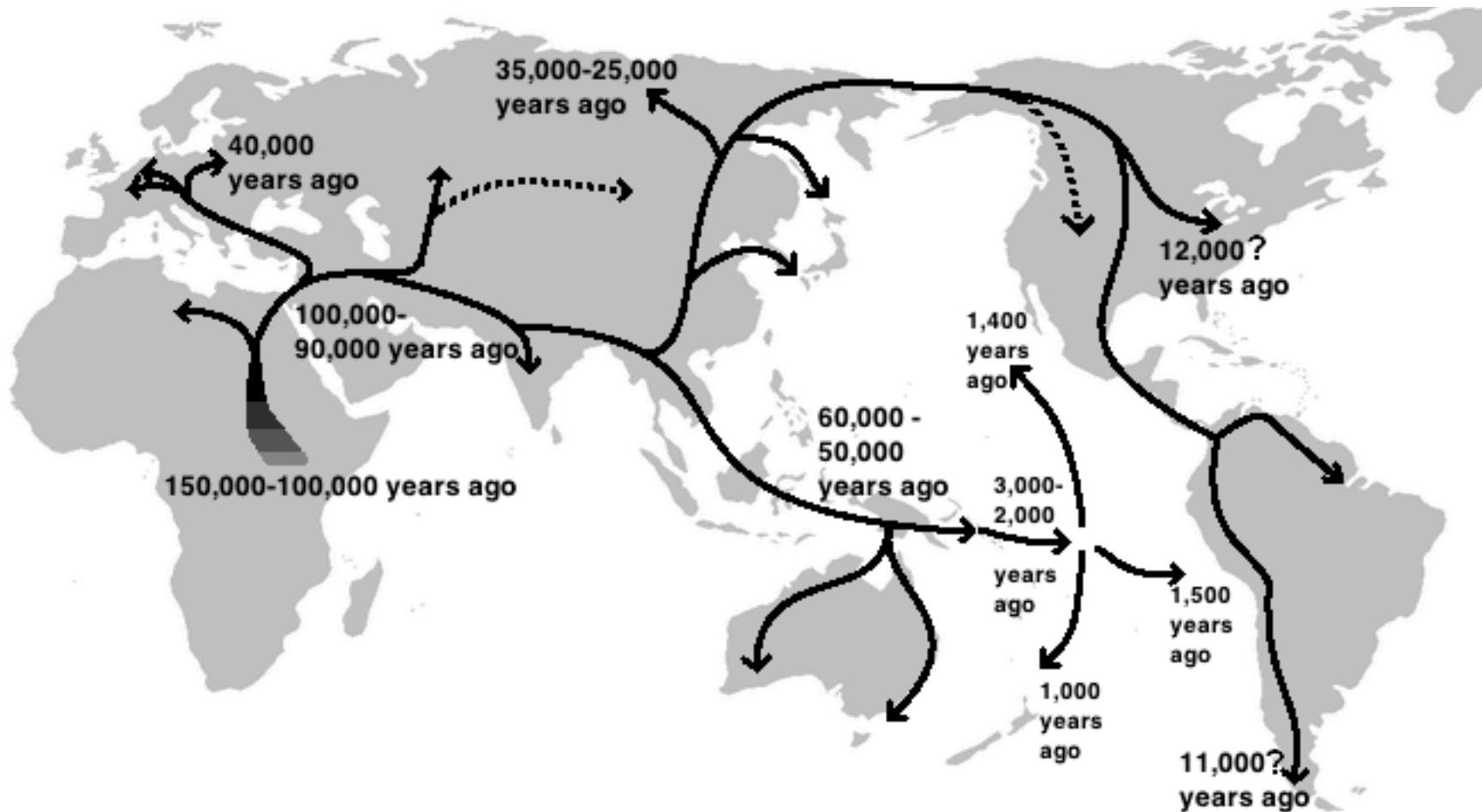
SP=spoločný predok

Last Universal Common Ancestor, LUCA

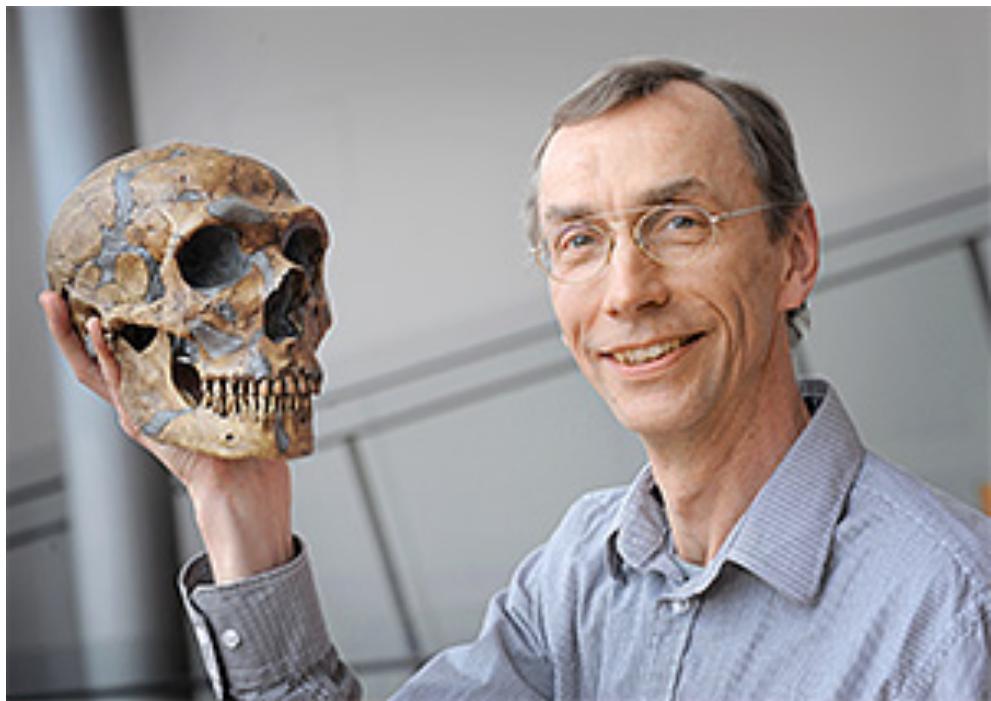
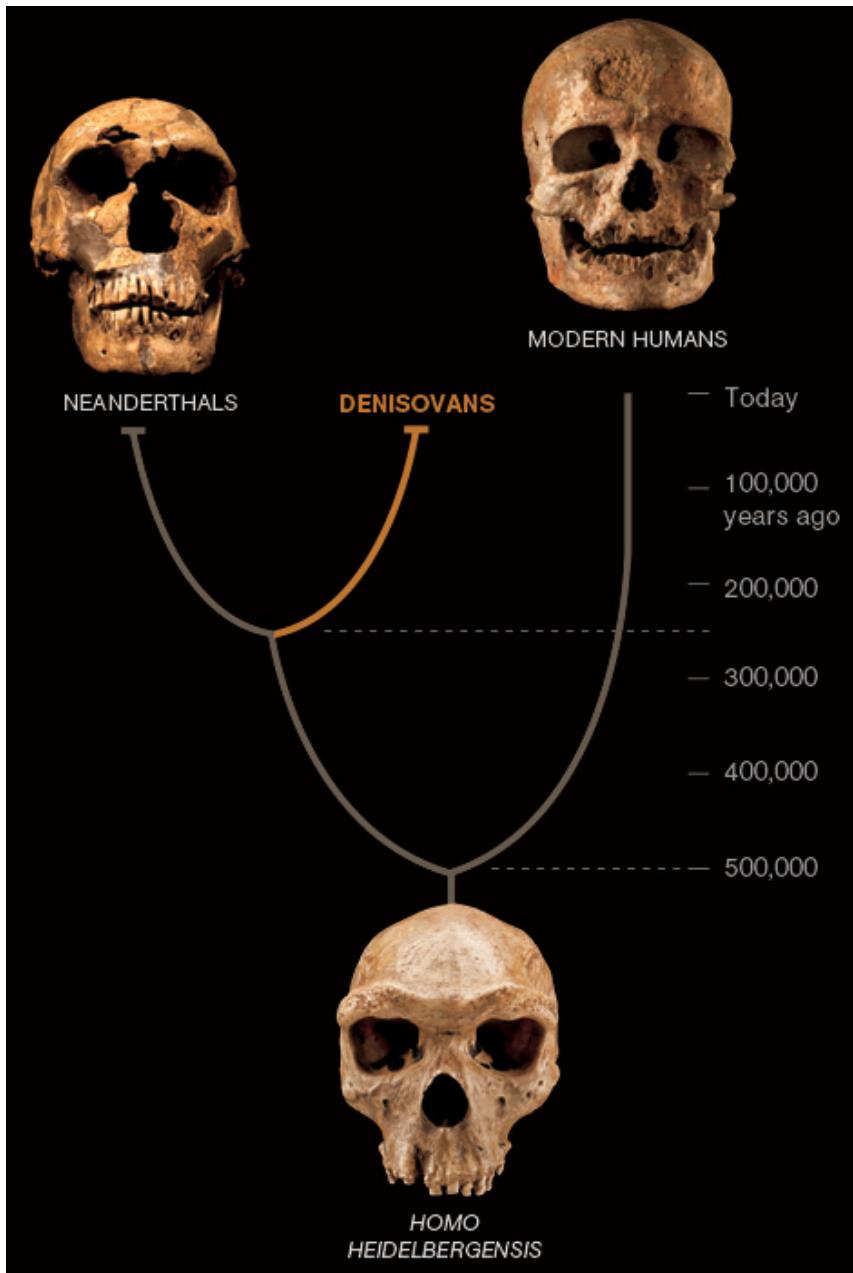
Ako je možné vystopovať evolučného predka mačky domácej?



Genomické analýzy umožňujú rekonštruovať pravdepobné cesty migrácie *H. sapiens*

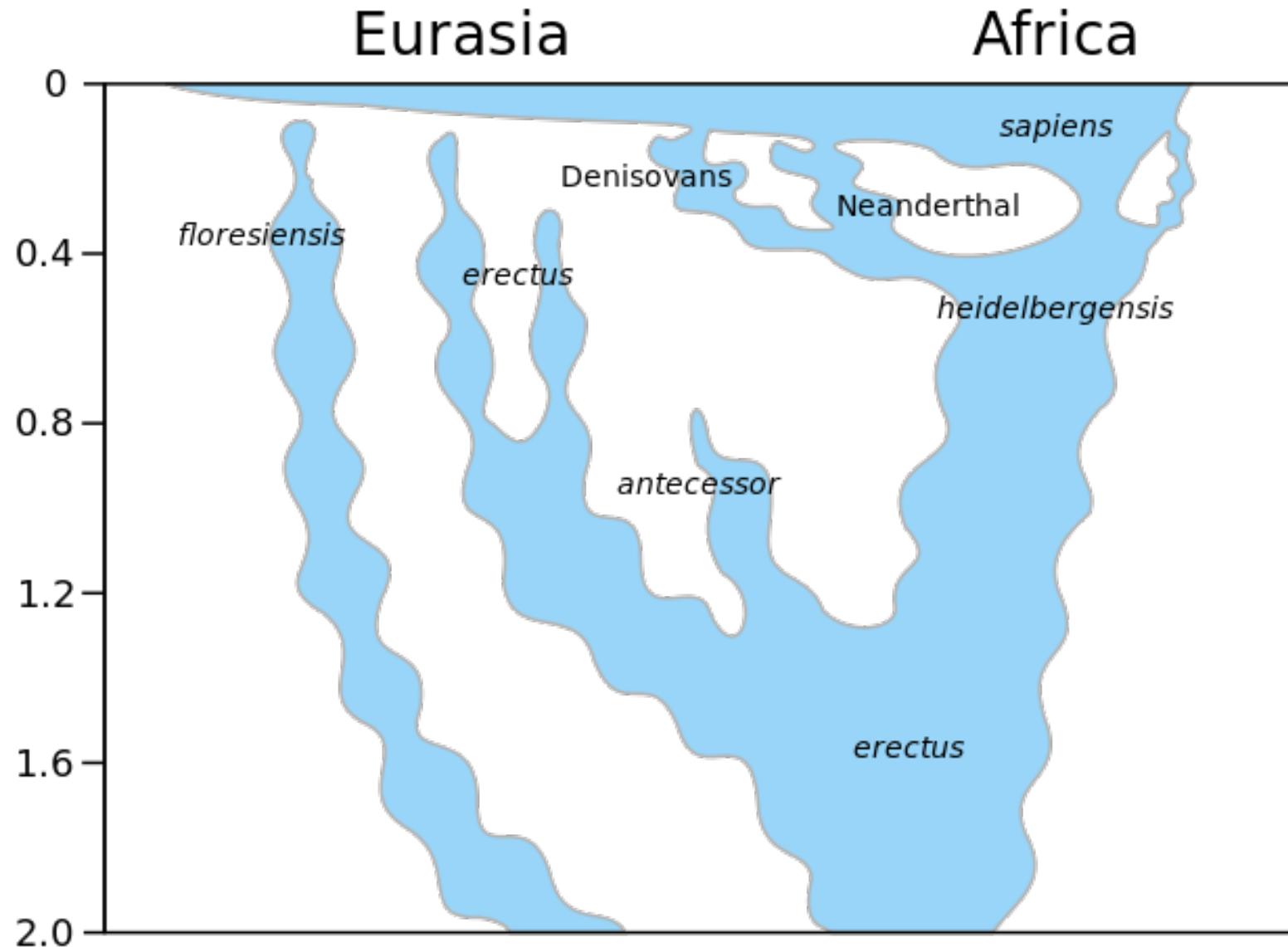


Paleogenomické techniky umožňujú odhaliť sekvencie genómov vyhynutých živočíchov, včítane starobylých foriem rodu *Homo*



Svante Paabo

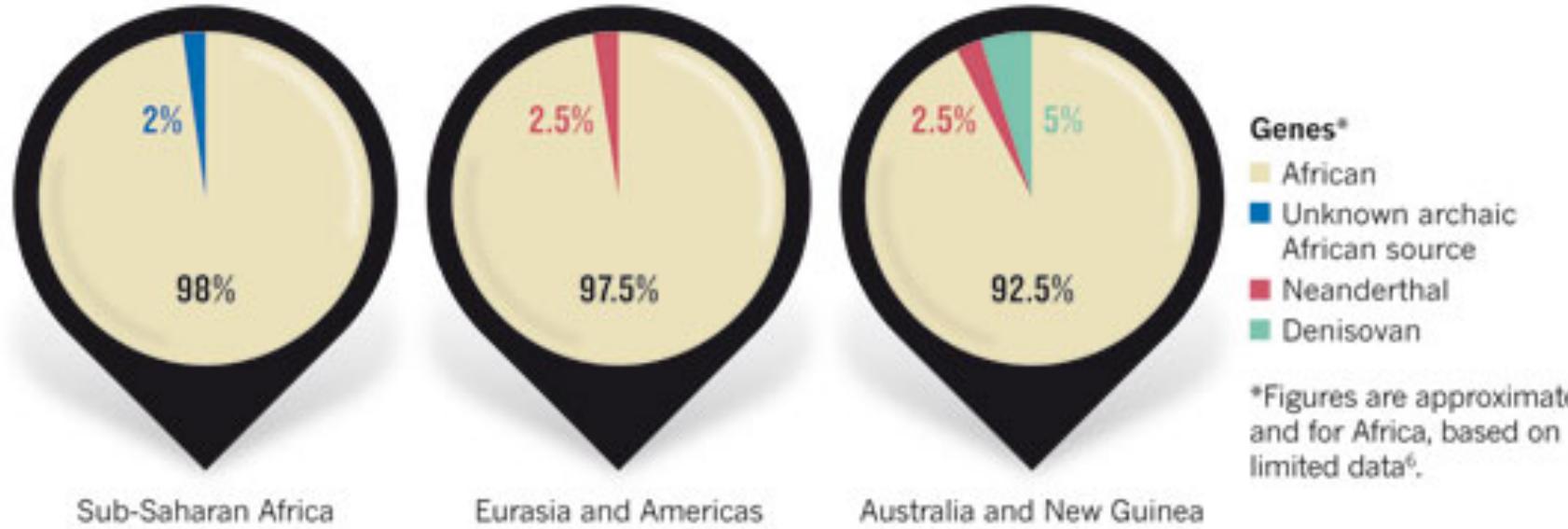
Porovnanie sekvenčí vyhynutých druhov rodu *Homo* viedlo k odhaleniu príspevkov do genetickej výbavy *H. sapiens*



Porovnanie sekvenčí vyhynutých druhov rodu *Homo* viedlo k odhaleniu príspevkov do genetickej výbavy *H. sapiens*

PATCHWORK PLANET

Most people's genomes contain remnants of archaic DNA from ancient interbreeding³⁻⁶.



Prečo africké populácie *H. sapiens* nemajú v genóme stopy po neandertálcoch a denisovanoch?

Problémy, ktoré rieši moderná genetika sú nielen výzvou, ale aj príležitostou pre potešenie z hry pri odhalovaní ich riešení



Ján Amos Komenský
(1592–1670)

**Veľa zdaru za všetkých prednášajúcich a
vedúcich cvičení kurzu Genetika 1!**