Projekt lidského genomu

Úkolem je zjistit kompletní genomovou sekvenci.

Problém:

Genom je velký!

3 miliardy párů bazí

1000 telefonních seznamů

Jak jsme genom sekvencovali?

Co lze nalézt v genomu?

Jaké je využití znalosti genomu?

Jak si stojíme ve srovnání s nejbližšími příbuznými?

Které další projekty z HGP vycházejí?

Genomová sekvenace

6/ 25/ 04

1128 genomových projektů:

199 kompletních (včetně 28 eukaryontních)

508 prokaryotických genomů před dokončením

421 eukaryotických genomů před dokončením

nejmenší: archaebacterium *Nanoarchaeum equitans* **500 kb**

Bacillus anthracis (anthrax) **5228 kb**

S. cerivisiae (kvasinka) **12,069 kb**

Arabidopsis thaliana **115,428 kb**

Drosophila melanogaster (octomilka) **137,000 kb**

Anopheles gambiae **278,000 kb**

Oryza sativa (rýže) **420,000 kb**

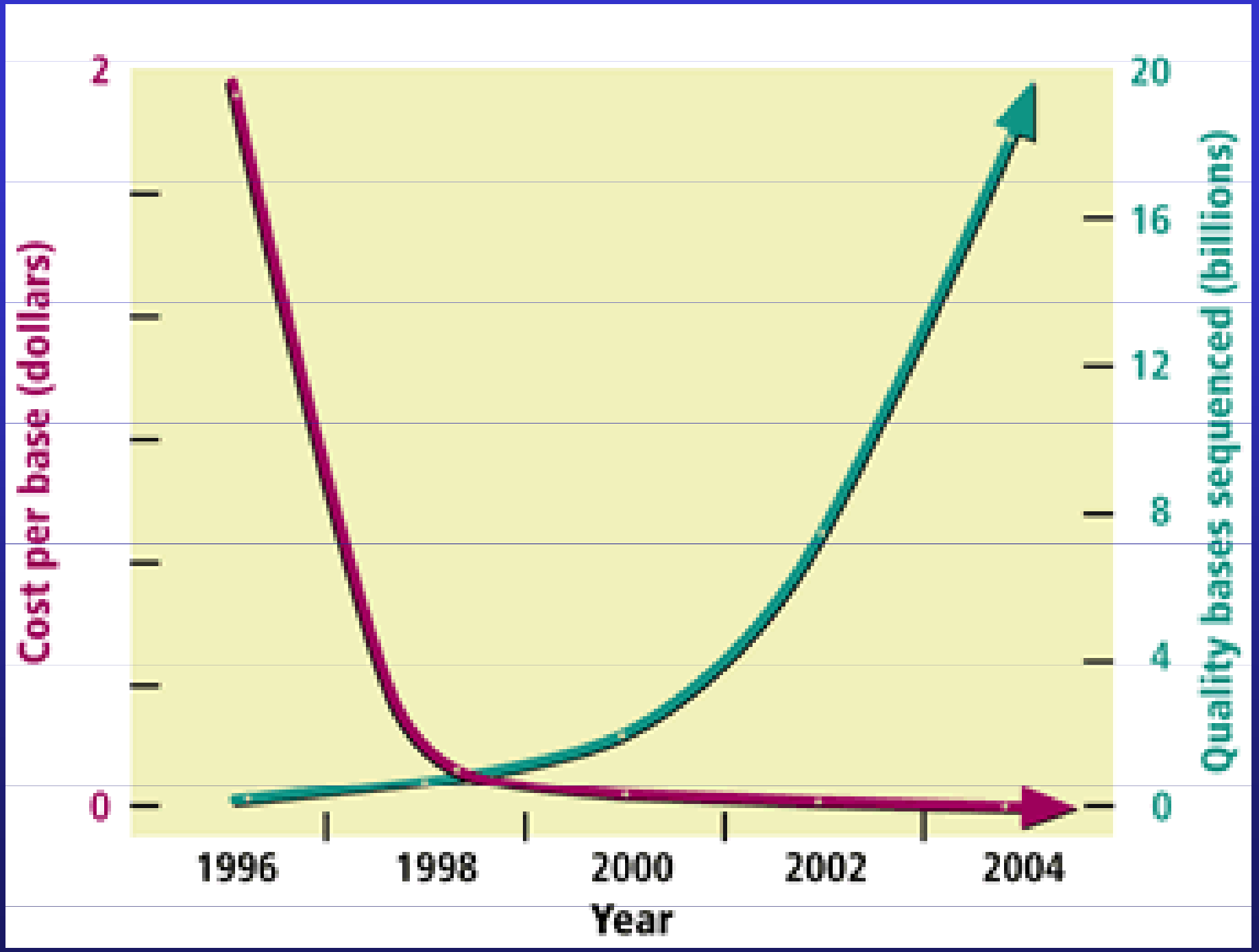
Mus musculus (myš) **2,493,000 kb**

Homo sapiens (člověk) **2,900,000 kb**

http:// www. genomesonline. org/

S. cerevisiae
200x
H. sapiens
200x
A. dubia

1980 - $10/bp
2001 - $0.1/bp
2006 - $0.01/bp


2
0
Cost per base (dollars)
20
16
12
8
4
0
Quality bases sequenced (billions)
1996
1998
2000
2002
2004
Year

Pracoviště, které osekvencovaly 85% genomové sekvence

1.**Whitehead Institute for Biomedical Research**, Center for Genome Research, Cambridge, MA
2. **The Sanger Centre**, Cambridge, UK
3. **Washington University Genome Sequencing Center**, St. Louis, MI
4. **US Department of Energy**, JGI, Walnut Creek, CA
5. **Baylor College of Medicine Human Genome Sequencing Center**, Houston, TX

USA, UK, Japan, Germany, China, France

Jak?

HGC: 9 neznámých lidí

- 5x vzorek krve
- 3x spermie
- 1x 987SK buňky

Celera: 2 muži, 3 ženy

Afroamerican
Asian - Chinese
2 Caucasians
Hispanic - Mexican

Genomy individuálních lidí (J. D. Watson, etc.)

Sekvenace genomu

Genom: 3 Gb

Štepit genom na větší kusy DNA

Klonovat do BACs: 100 kb

Mapování BAC klonů podél chromosomů

Znovu štěpit

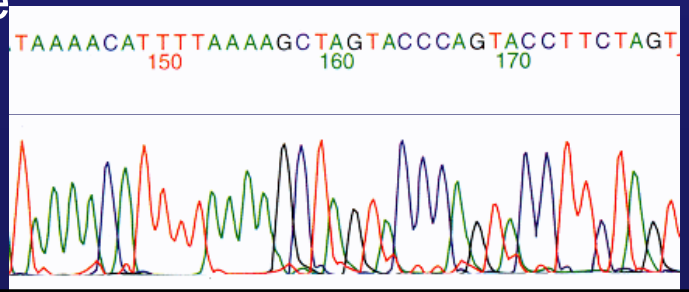
Dokončit sekvenci

…TTGTAAGTGAGAACAGGACGTATGTGGTTTTCTACTCCTGTGTT…

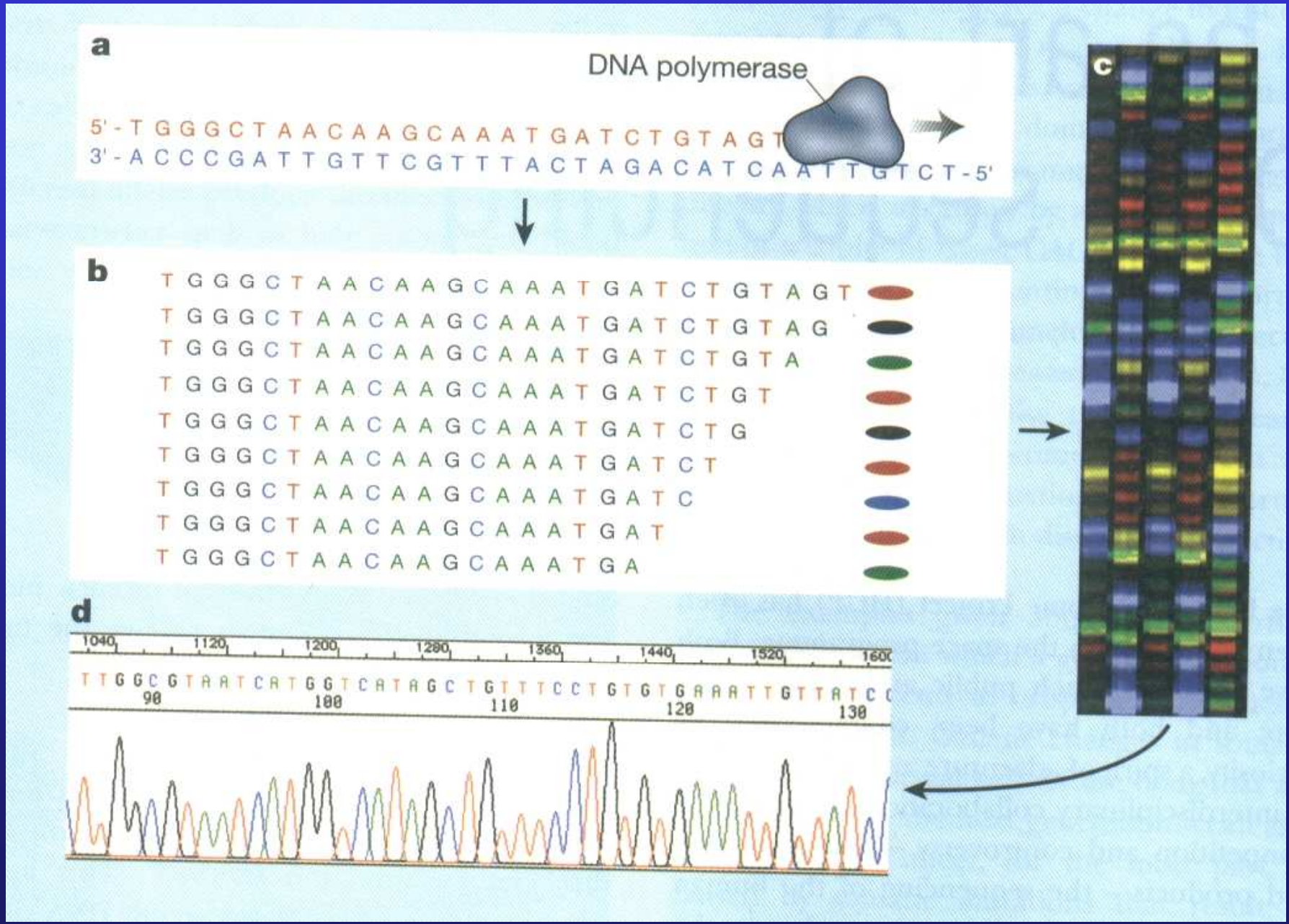
uspořádat BAC sekvence

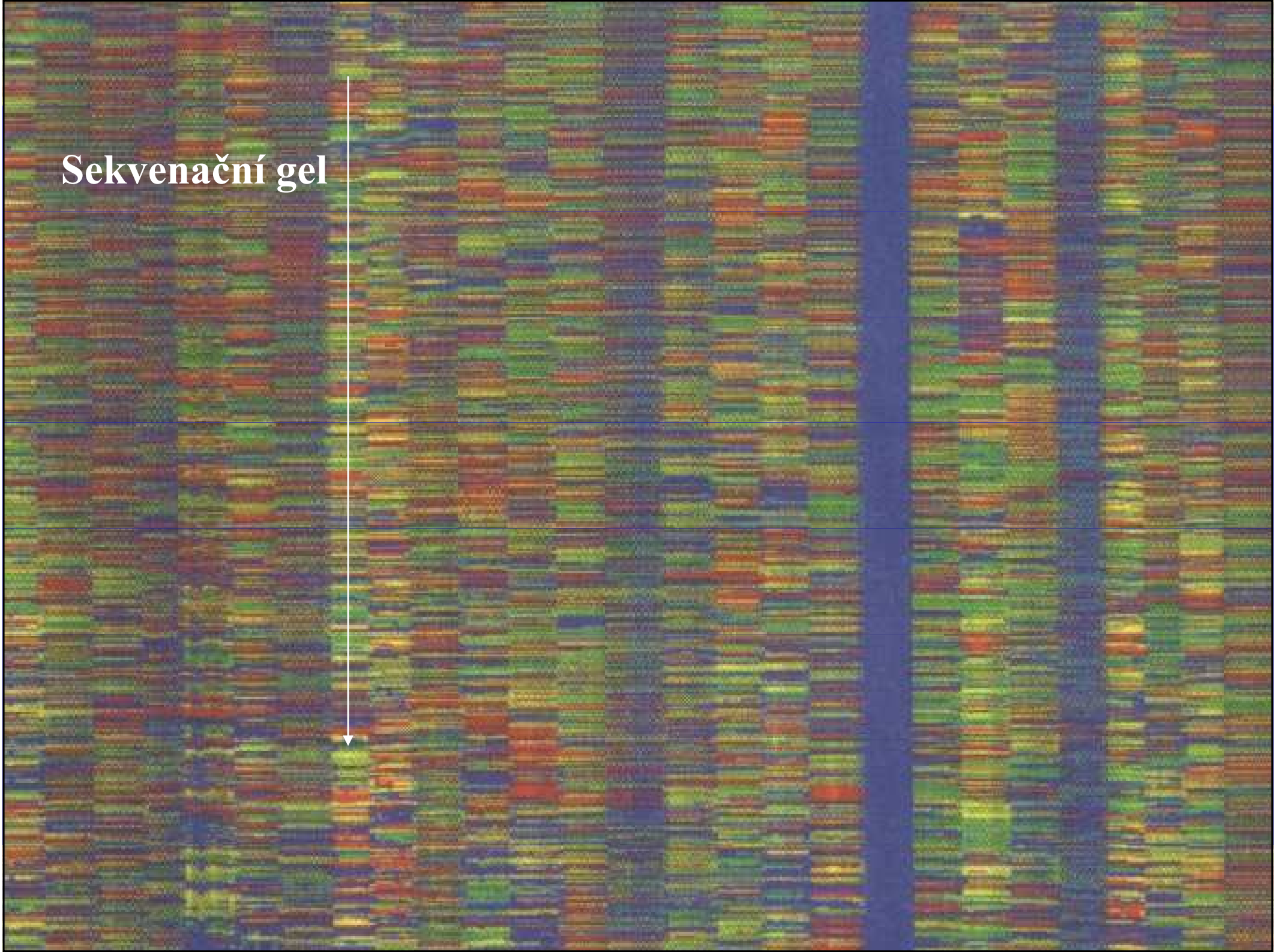
TTGTAAGTGAGAACA
AGAACAGGACGTATGTGGT
TGTGGTTTTCTACTCC
CTACTCCTGTGTT

Sekvenace



Princip sekvenace




Sekvenační gel

Kompletace lidské genomové sekvence

International Human Genome Sequencing Consortium.
Finishing the euchromatic sequence of the human genome.
Nature 2004 Oct 21;431(7011):931-45.

„The current genome sequence (Build 35) contains **2.85** billion nucleotides interrupted by only **341** gaps.
It covers approximately **99%** of the euchromatic genome and is accurate to an error rate of approximately **1 event per 100,000** bases.
Human genome seems to encode only **20,000-25,000** protein-coding genes“

2.85 miliard nt a 341 neosekvencovaných oblastí.
1 chyba na 100 000 nt.
20 000-25 000 genů kódujících proteiny.

Kousek sekvence chromosomu 21

TCACAATTTAGACATCTAGTCTTCCACTTAAGCATATTTAGATTGTTTCCAGTTTTCAGCTTTTATGACTAAATCTTCTAAAATTGTTTTTCCCTAAATGTATATTTTAATTTGTCTCAGGAGTAGAATTTCTGAGTCATAAAGCGGT
CATATGTATAAATTTTAGGTGCCTCATAGCTCTTCAAATAGTCATCCCATTTTATACATCCAGGCAATATATGAGAGTTCTTGGTGCTCCACATCTTAGCTAGGATTTGATGTCAACCAGTCTCTTTAATTTAGATATTCTAGTACAT
ACAAAATAATACCTCAGTGTAACCTCTGTTTGTATTTCCCTTGATTAACTGATGCTGAGCACATCTTCATGTGCTTATTGACCATTAATTAGTCTTATTTGTTAAATGTCTCAAATATTTTATACAGTTTTACATTGTGTTATTCATT
TTTTAAAAAATTCATTTTAGGTTATATGTATGTGTGTGTCAAAGTGTGTGTACATCTATTTGATATATGTATGTCTATATATTCTGGATACCATCTCTGTTTCATGCATTGCATATATATTTGCCTATTTAGTGGTTTATCTTTTCAT
TTTCTTTTGGTATCTTTTCATTAGAAATGTTATTTATTTTGAGTAAGTAACATTTAATATATTCTGTAACATTTAATGAATCATTTTATGTTATGTTTAGTATTAAATTTCTGAAAACATTCTATGTATTCTACTAGAATTGTCATAA
TTTTATCTTTTATATACATTGATATTTTTATGTCAAATATGTAGGTATGTGATATTATGCACATGGTTTTAATTCAGTTAATTGTTCTTCCAGATGTTTGTACCATTCCAACATCATTTAAATCATTAAATGAAAAGCCTTTCCTTAC
TAGCTAGCCAGCTTTGAAAATCCATTCATAGGGTTTGTGTTAATATATTTTTGTTCTTTTTTTTCCTTTCTACTGATCTCTTTATATTAATACCTACTGTGGCTTTATATGAAGTCATGGAATAATACGTAGTAAGCCCTCTAACACT
GTTCTGTTACTGTTGTTATTGTTTTCTCAGGGTACTTTGAAATATTCGAGATTTTATTATTTTTTAGTAGCCTAGATTTCAAGATTGTTTTGACGATCAATTTTTGAATCAATTGTCAATATTTTTAGTAATAAAATGATGATTTTTG
ATTGGAAATACATTAAATCTATAAGCCAAATTGGAGATTATTGATATATTAACAAAAATGAGTTTTCCAGTCCATGAATGTATGCACATTATAAAATTCATTCTTAAGTATGTCATTTTTTAAGTTTTAGTTTCAGCAGTATATGTTT
GTTACATAGGTAAACTCCTGTCATGGGGGTTAGTTGTACAGGTTATTTTATCATCCAGGCATAAAGCCCAGTACCCAGTAGTTATCTTTTCTGCTCCTCTCCCTCCTGTCACCCTCCACTCTCAAGTAGACCCCAGTTTCTGTTGTTC
TCTTCTTTGCATTAATGACTTCTCATCATTTAGATTGCACTTGTAAGTGAGAACAGGACGTATGTGGTTTTCTACTCCTGTGTTAGTTTGCTAAGGATAACCACCTCCATCTCCATCCATGTTCCCACAAAAGACATGATCTCCTTTT
TTATGGCTGCATATTATTCCATGGTATATATGTACCACATTTTCTTTATCCAATCTGTCATTGATGGACATTTAGGTTGTTTCCACATCATTGCCGTTGTAAATACTGCTGCAGTGAATATTCGTGTGTATGTCTTTATGGTAGAATG
ATTTATATTCCTCTGGGTATATTTCCAAGTAATGGGATGGTTGGGTCAAATGGTAATTCTGCTTTTAGCTTTTTGAGGAATTGCCATATTGCCTTTCACAACGGTTGAACTAATTTATACTCCCAAGAGTGTATAAGTTGTTCCTTTT
TCTCTGCAACCTCGACATCACCTGTTATTTATGACTTTTATATAATAGCCATTCTGCTGGTCTGAGATGGTATCTCATTATGATTTTGATTTGCATTTCTCTAATGCTCAGTGATATTGAGCTTGGCTGCATATATGTCTTCTTTTAA
AAATATCTGTTCATGTCCTTTGCCTAATTTATAACGGGGTTGTTTGTTTTTCTCTTGTAAATTTGTTTAAGTTCCTTATAGATTCTAGGTATTAAACCTTTTTTCAGAGGCGTGGCTTGCAAATATTTTCTCCCATTCTATAGGTTGT
CTGTTTATTCTGTTGATAGTTTCCCTTGCTGTGCAGAAGCTCTTAACTTTAATTAGATCCGACTTGTCAATTTTTGCTTTGGTCGCAATTGCTTTTGATGTTATTGTCGTGAAATCTTTGCTAGTTCTTAGGTCCAGGATGATATTGC
CCAAGTTGTCTTCCAGGGCTTTTATAATTTTGGATTTTACATTTAAGTCTTAATATATTTATTAAATTTGTTAGGGTTTCAGGATACAAGGACAATATAGCAGCAAACAATGTAAAAGTAAAATCTGAAAAATAATAGAAAACAGTTT
AATTGAACACTTTACCATTATGTAATGCCCTTCTTTGTCTTTCCTGATCTTTGTTGGTTTGAAGTTCAAAAAAGACAAACTTAATGGTACAATAGGTATTGTAGATTTCAGGACTTTCTGTATAAAATATTTTGTATATATGAATAGA
TCATTTTTTATTTCCAGTCTTTAAACATTTTCTTAACATTTTCTTCTATTGCTTCACTTCACTCGCTAGGACCATCAGGACAGTGTTGAACAGAAATTGTCAGACTGATCATCACAACTTTTTCTAGATTTTAGAAGGAAATTTTTCT
TTATTTCAACATAAAGCAGCATGTTAATGCCAAGTTTTAATATGTGTTATCAGATTGAAATTTTTTTGTATATTTCTACATTACCAAGAATTTTTAGCAAGAGTTTTTGTTGAGTTTTAATTTAAAAATCATTTGTTAATTTCATCTG
ATTTTTTTATTTCTCTTTTTACCTTAAGAGATTAAACTGACTACAGATTGAATATAAACAAACAAACAAACAAACAAAAACTCTAAAATGCTGTGGATCAACACCACTTAGTAATTTGTATACTTGGATTCAATTTGCTGAAATTTTG
TTAGACATTTTTGCGTCGATATTTATGAGGGATGTTGATCTGTAAAAGTATTAAAATGCCTTTGACAGATTTTGATAGCAGTGTTATTCTGGCCTAATAAATCAAACTGAGGTATGATCCTTCCTTTTCTATTTCTTAATAGCATTTT
TAAAATTGGTGGTTTTTTCCTTCCTTAGTGAAATTTACCAGCAAAGTAACAGGCCTTATATTTCTCTTGTGGAAATATTTTAATTTCAAATTAATGGTATTTTGTTCTTGTAGGGTGGTAATTTTCTCTGTGTTTGGTCTTAATGGAC
TCTTAGCTGATCACCCAGTTACTCAGCGAGGTCTCTTCACTCTGGAAGAGCTGGAACTCCAGTGTGTTTTAGTGCAGCATGACCACGGGTATTACCGTTCAACATTTAGGCTTTATCAGTGATAACTATTTGTCCTCATGGAGTTTTT
GCCGCTGGGCCTACACAGTTTAGGCTTCAGCTTAGAACACATAATGAATTCTTATGCAGATTTCTGCCCACCTTTGACCTTTCATGATTTCCTCTTCTTGGGTAAGCTGCCTTATTAATCTGATACACTTCAGCAGTCCAGAACTACA
CTCTTTCCCTTCTCTGCTCTTGGAGATGACTCTTTTGTCTGAGATTCACTTTGCTGTGCTGAAAAAGAAAAGTGCTTCAAGGAAGATACCAAGGAAAATCACAGGGCTCATTTATGTATTTCTCTTCTTTCAAGGACTACAGCTTTGT
GTTGCCTATGTTCAATTTCTGAAAATAATTAGAGCATATATACTCTGTGTGAGAAGGCAAATCCAGACAGTTAGTTTGTATGACTAGAAGCAGAAGTCTACATGGAGAATTTTACTTAACTGTGTTATAGTTTCTTTAATTATTTCAA
GAGTATGTTTAATGTTCCACAGATCTCATTCTATAAATCTTTATCATCTTAGAGCTCTGATACTATTTAGAATTACTATTCCTTCAAATAAGAGATTAGAAACAGGGTTATATTTGGGGTAGGTTGACTTACTTTTCTGGGAACCAAA
GCATATTAAATTGACCAGTTTTAACACACTTCTATGTATGCACAAAGATATATATTTACATTCTGCAAAATCATTCTTTCCTTTTTGAATTTGAAAAGGATCTTTGGTATACAGATATTCAATAGCCAGCCTGAAGATTCATTTGAAT
TCATTTAATGTTTAGATTCACTACATGAAATGATCCAGAAGAGAGTACTCAAATATAAGTATCTATAACGATGGAAATATACATCTCCACTGCCCAAGATGGTAGTCATGAGTCAATATTGATCATGTGAGACGTGGCAAGTGTTACT
CAGGGTCTCAATATTTAAATGTATTAAGCTTTAATTAATGTAAATTTGAATTTAGCAAAACATGTATAGCTTGTGGTTACTGTTTTATTCAGTGCCAATATAGAACATTTCCATGATTACAGAAAGTTATCTTAGAATACTCAGTTCT
GGACTATTTTATCTGGCTAAATTAAATGTTAAAATATTACAAATTCATCTTCAGGCTGGCTGTTGAATATTTTTATAGCAAAAGTCATTTATAAATTTAAAACTCAAATAATTATCTTTTTCAATATGTAAAATATGTCTTTACATAT
TCTACTCCCTTCTTACATACATATTCTGATGTAACATAGGTATTCTCTTATTCATGCACACTGAAATGACAACATAAATAATTTTACTAAGTGTCACCATATAAAAAACTTTGAACAAAATCAGATTATATCACTGTGGATATTTCTA
TTTTGAACTAACTTAGATGATAATTTTAATCTATATCCTAGATGAACTTTAAATCAATAAAATCTCTCAATGGTGTTATAAATCTCAAGCCATTAGCCACTGATTATCCCATTTTTATTCTTTTCATATTAATTTTATTGCCATGTAT
GAATGCTGTAGCATCCATGTTTAAATACTAGTTAACAAAATGCACTGGCATCAGATACAATAAGGATGAAATGAGATATAATTAGGACTCTGGTAACACACATAAAATTGGAAAGATACCCTGAAATTCAAGCCAAGAAGATATTTAT
CCAGCTTATTTTATTTTGAGACAGAGTCTTGCTCTCTCACTCAGGCTGGAGTGCAGTGGACCATTCTAGGCTCGCTCCAACCTCTGTCTCCCAAATTGAAGTAATTCTCGTGCCTCAATCTCCCGAGTAGCTGGGATTACAGGCATGT
GTCACCAAGCCTGGCTGATTTTTGTAGTTTTAGTAGAGACGGGGTTTCACCATGATGGCCAGGCTGGTCTTGAACTCCTGGCCTCAAGTGACTGGAACACCTCGGCCTCCTAAAGTGCTGGGATTACAGACGAGAGCCACTGAACAGC
TTTGATCCAACTTATTTGGATGAATGAGTTACATATTTTACATTAAATCTGTTATTGTGATAATTCTTCATGTTATTTTCCATGTATAGATTTATATATAATGTAATTTTAATTTTTTTTCACCGGAGAGTATAAACAACAATTATTT
TATAAACAGGATAATAAAAATAAGACAAAAATTGTTGAAATGTCTTCATTTGACTACTAACTTTTTACATGTTTGTTACTTTGAAGCTGTTATCAATACTTGTGATGTATTACAATTAAGTAAAGATTTAAAGATGCCATTTTTAACT
TATTATGACACAAAGTCTATAAATTCTTATATTTTGAGATTTGTATTTAAATAACTTGTGAAATTTAATTTTAAAATAAAATTTCTTCTATGGATTGGTCTTCAATCGAGGCATAAAAAGGAATATAACAGTGTGGCACTATAACTTC
TATATTGAATTTCTATATTATTTAACACAATTATAATTTTGCTAATGAATTGTAATGTTTTTAAAAAGCTAGGTGAATTTTATTAAATTCATTACATGGCGATAACACAGAGAAAACATTTTGGGGATTCTTTTAAAATGGTATGTAC
AAAAGCTTAAAAGTTGTTATGTAGTGGCAGAGATAAAAAAGTAAAACAAAAAAAAGCTTAAAAGTTTGCTTTACTATTTATAGGCTCATAAGTGTAAGTGTGCCAGAAAATGAAAAAGAAAGGAGAGAAATTATAAATAACTGTGTGG
AAAACACAGATAAAGCATAAAGATAGAATATAAAGATAGAAGCATTTTAATATGAGGCAGTGATGGCTTTTTGAAGAATCCCAACTAAGGACCTACTTTTAGTTAATAAATAATATGTTTCTAATCCCTATATTGTCCACAGCAACCT
TTTTAGGACATGGAGCAGTGACTATGAGTGCCAGAAGGCAAGAGTAGAAGCAATTGTAAAATCATGAACACTAGTTTGTAAAATCCTCACTGAGATATAATATCTGTTTGCCTCTACCTTAGAATTATTAATGTCTTGAGGGCTGGGA


PHASE TWO : INTERPRETATION
SHENEMAN The Star-Ledger
I THINK I FOUND A CORNER PIECE.
3 BILLION PIECES
GENOME

Co lze nalézt v genomu?

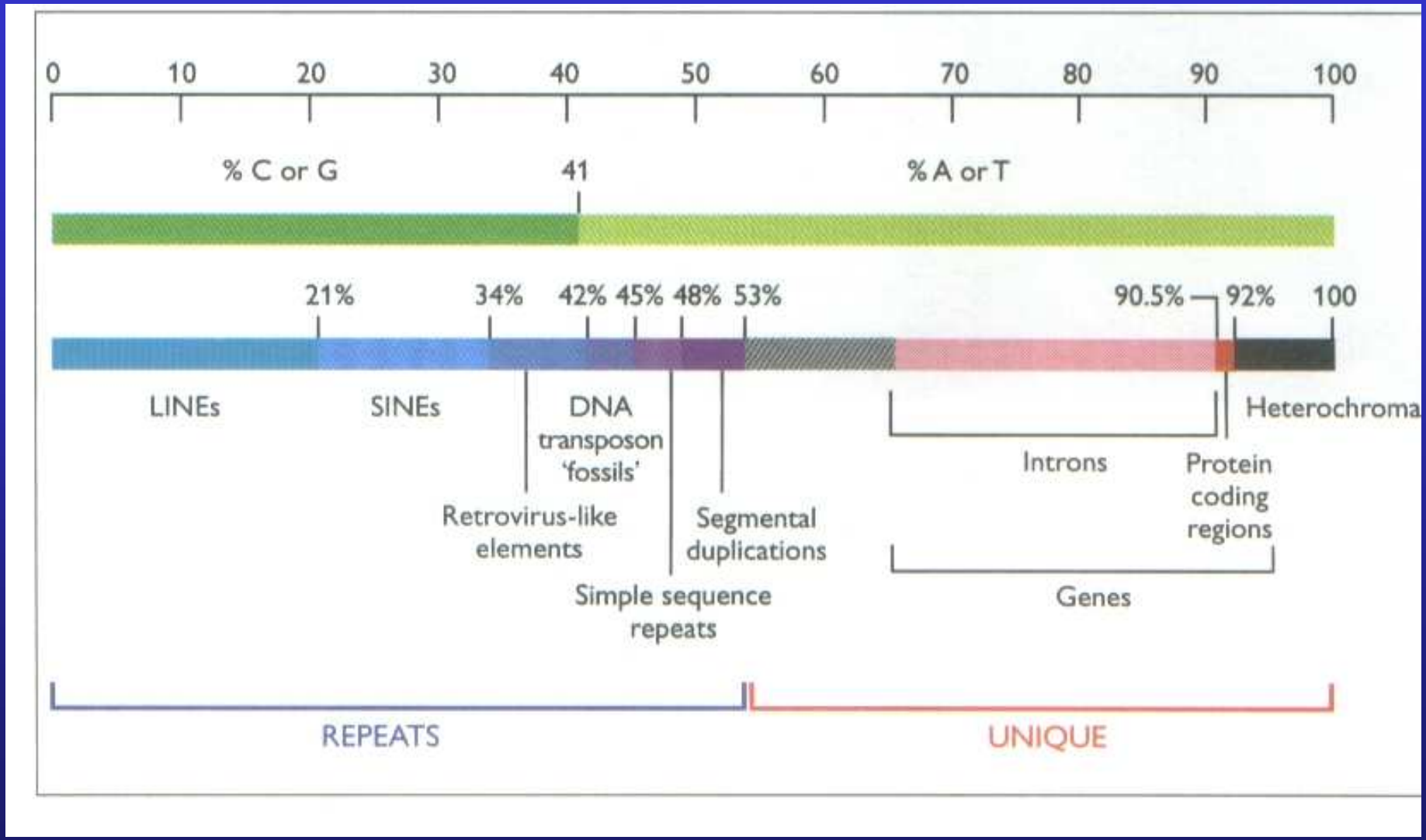
Geny (tj. protein kódující oblasti)

jen <2% genomu kóduje proteiny

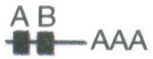
- geny pro nekódující RNA (rRNA, tRNA, miRNAs, atp.)
- strukturální sekvence (scaffold attachment regions)
- pseudogeny

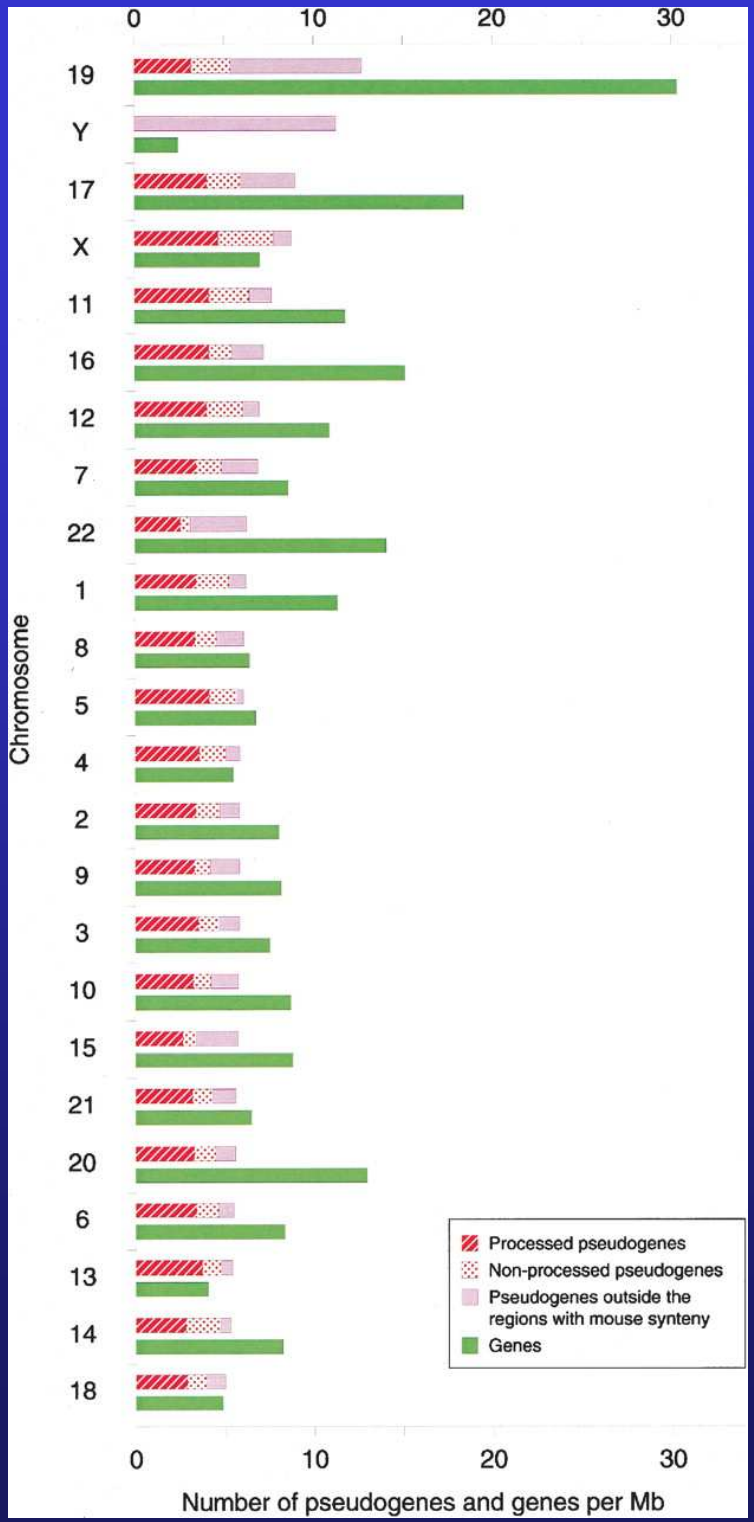
Regulační sekvence

- “junk” (zahrnující transposony, retroviry, atp.)


0
10
20
30
40
50
60
70
80
90
100
% C or G
41
% A or T
21%
34%
42%
45%
48%
53%
90.5%
92%
100
LINEs
SINEs
DNA transposon 'fossils'
Retrovirus-like elements
Segmental duplications
Simple sequence repeats
Introns
Protein coding regions
Heterochroma
Genes
REPEATS
UNIQUE

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous		6–8 kb	850,000	21%
SINEs	Non-autonomous		100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Non-autonomous		1.5–3 kb		
DNA transposon fossils	Autonomous		2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

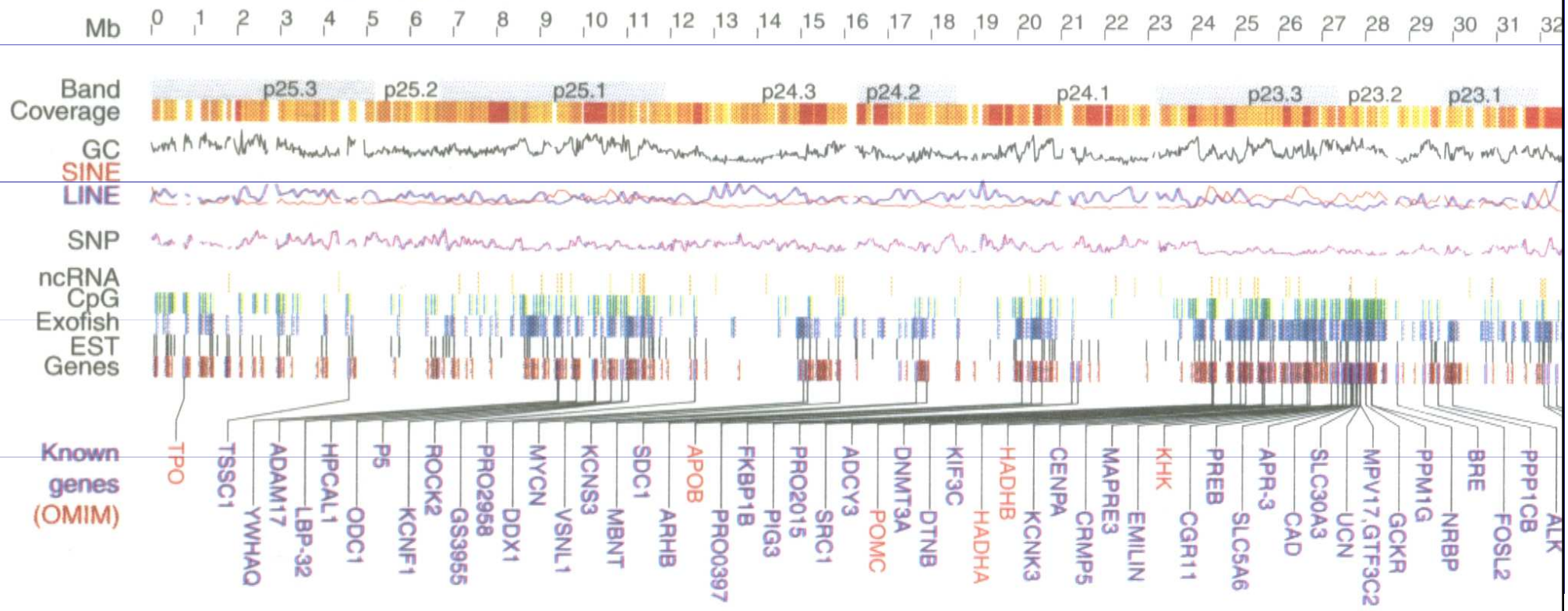


Pseudogeny

- 70 % procesované pseudogeny
- 30 % neprocesované pseudogeny
- ~20,000

Torrents et al. Genome Res. 2003 13: 2559-67.

Chromosome 2



Shrnutí

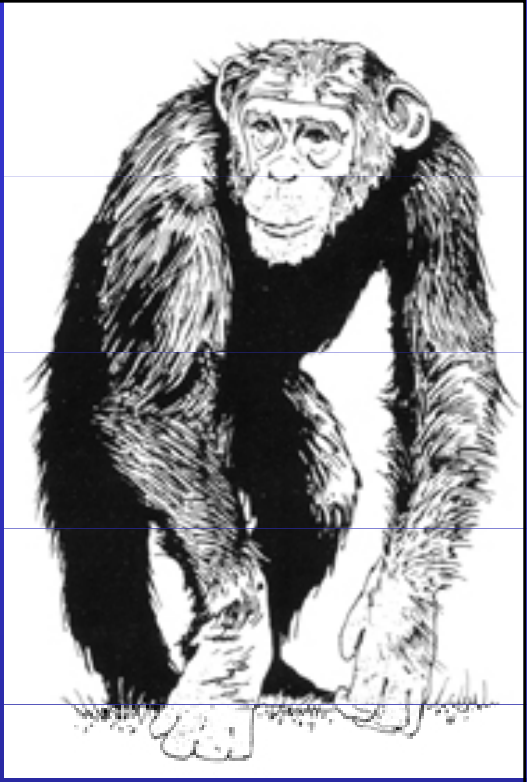
- Variabilní distribuce řady parametrů (GC, CpG islands, repetice)
- 20.000-25.000 protein-kódujících genů
- Proteom je mnohem komplexnější než u bezobratlých
- Horizontální transfer genů vs. ztráta genu u bezobratlých
- 20.000-30.000 pseudogenů
- V genomu se vyskytuje asi 10 miliónů SNP

Jaké je využití znalosti genomu?

Aplikace znalostí lidského genomu

- **Geny podmiňující gen. choroby** – poziční klonování (30 genů)
- **Paralogní geny** (achromatopsie, CNGA3, CNGB3); (971 známých genů => 286 paralogních genů)
- **Cíle zásahu medikamentů** – recentní kompendium = 483 cílů, 18 nově identifikovaných; (Alzheimer's disease, β-amyloid is generated by processing APP by BACE; BACE2 in obligatory Down's syndrom region of chromosome 21)
- **Obecná biologie** – hořká chuť – nová rodina G-proteinových receptorů

Jaké jsou příbuzné genomy?



Sekvenace genomu šimpanze

Chimpanzee Sequencing and Analysis Consortium

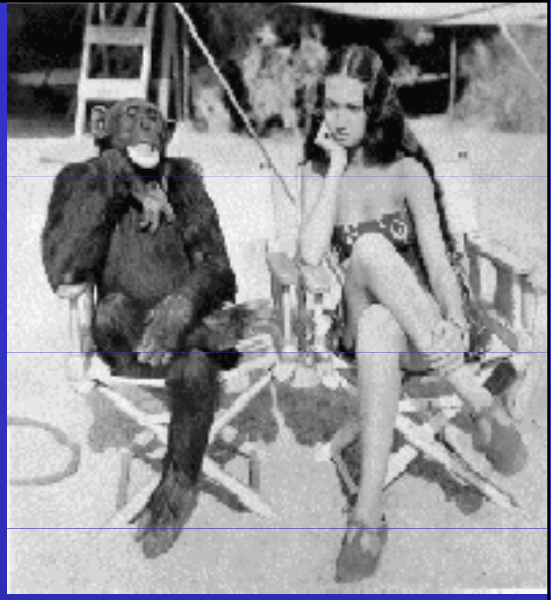
Initial sequence of the chimpanzee genome and comparison with the human genome.

Nature **2005** Sep 1;437(7055):69-87.

Thirty-five million single-nucleotide changes, five million insertion/deletion events, and various chromosomal rearrangements.
98,6 % identitity to human genome sequence
Differences in gene/exon structures

35 miliónů záměn nt, 5 miliónů inzercí, delecí a dalších změn
96% identita s lidskou genomovou sekvencí
Změny ve struktuře genů popř. exonů

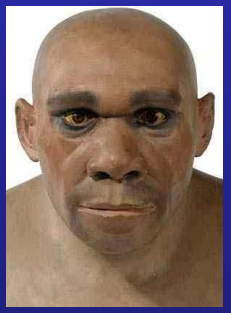
Rozdíly mezi lidmi a dalšími primáty



	Lidé	Primáti
Definitivní		
HIV progrese v AIDS	častá	vzácná
Symptomatologie chřipky A	středně těžká až závažná	lehká
Komplikace u hepatitidy B/C	středně těžké až závažné	lehké
Malárie *(P. falciparum)*	citliví	rezistentní
Menopauza	obligátní	vzácná
Pravděpodobné		
E. coli K99 gastroenteritida	rezistentní	sensitivní?
Rozvoj m. Alzheimer	kompletní	částečný
Koronární ateroskleróza	častá	vzácná
Karcinomy	časté	vzácné

Genetické rozdíly mezi současným člověkem (Člověk moudrý, *Homo sapiens*) a vybranými organismy na celogenomové úrovni a odhadovaná evoluční vzdálenost od posledního společného předka

Druh	Odhadovaná doba uplynulá od výskytu společného předka	Přibližná příbuznost na úrovni DNA
Člověk moudrý (*Homo sapiens*), kterýkoliv jiný jedinec		~99,9 %
Člověk neandertálský (*Homo neandertalensis*)	0,5 mil. let	≈99,5 %
Šimpanz učenlivý (*Pan troglodytes*)	6 mil. let	~96 %
Makak rhesus (*Maccaca mulata*)	25 mil. let	~93 %
Potkan obecný (*Rattus norvegicus*)	75 mil. let	~40 % genomu je sekvenčně příbuzných (nikoliv shodných)



Neandertálci

- Hominidé nejvíce příbuzní moderním lidem
- Objevili se asi před 500.000 lety
- Evropa and západní Asie
- Vyhynuli před 30.000 lety

http://www.msnbc.msn.com/id/13154583/

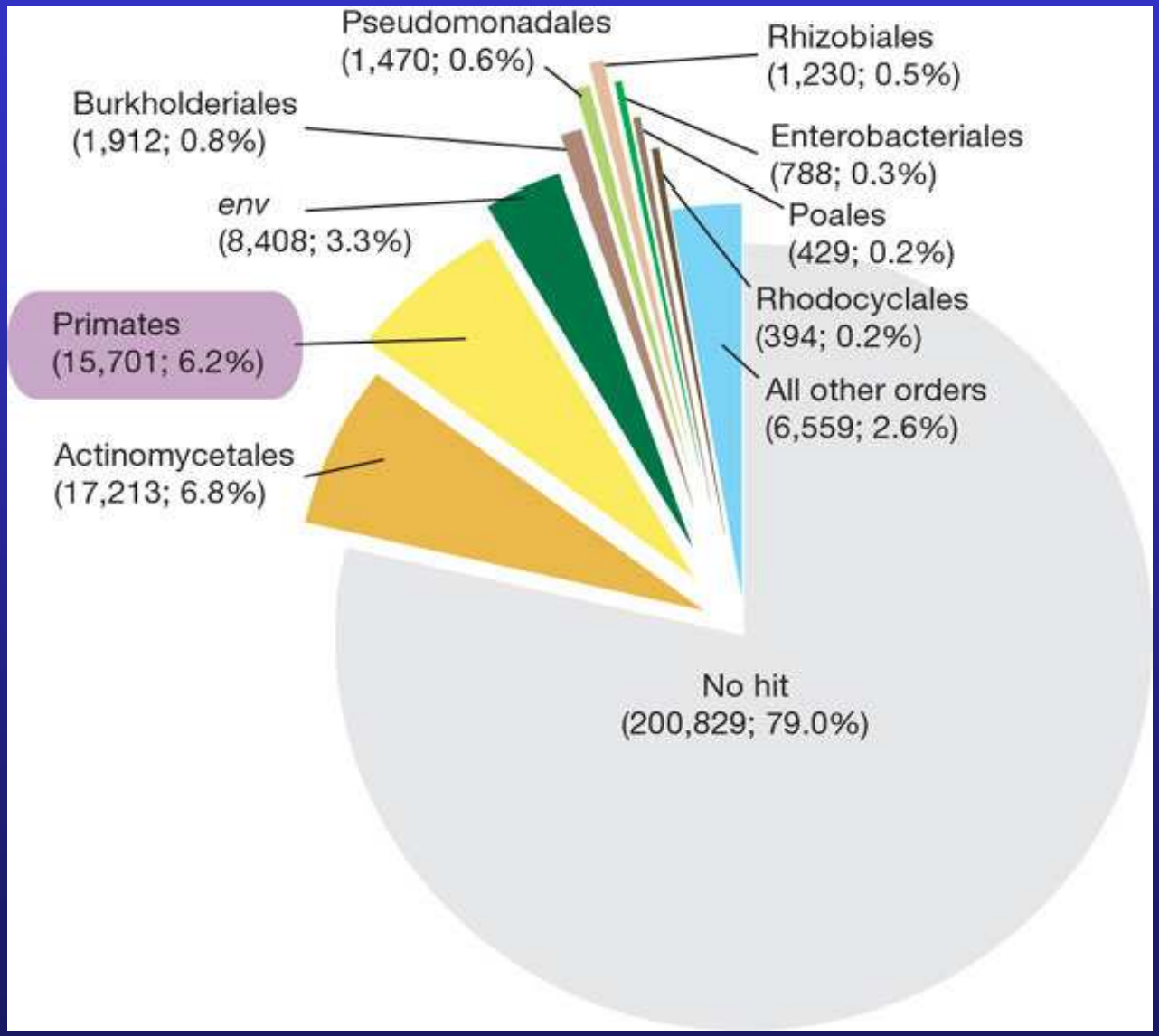
DNA

- Kost nalezena v roce 1980 v chrorvatské jeskyni
- Radioizotopové datování:

38,310 ± 2,130 let

http://www.nature.com/nature/journal/v444/n7117/pdf/444254a.pdf

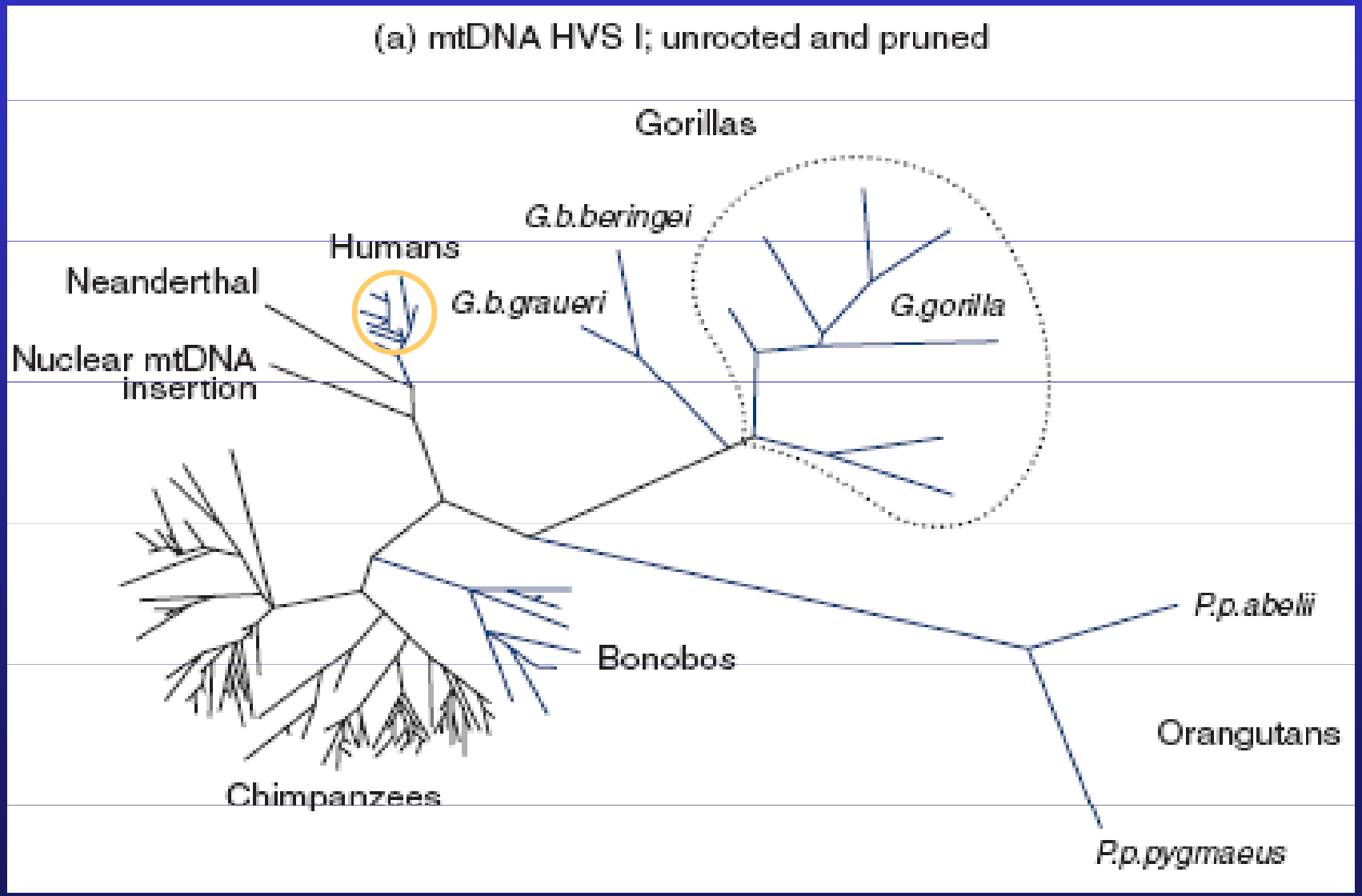
Kost Vi-80 (z jeskyně Vindija)

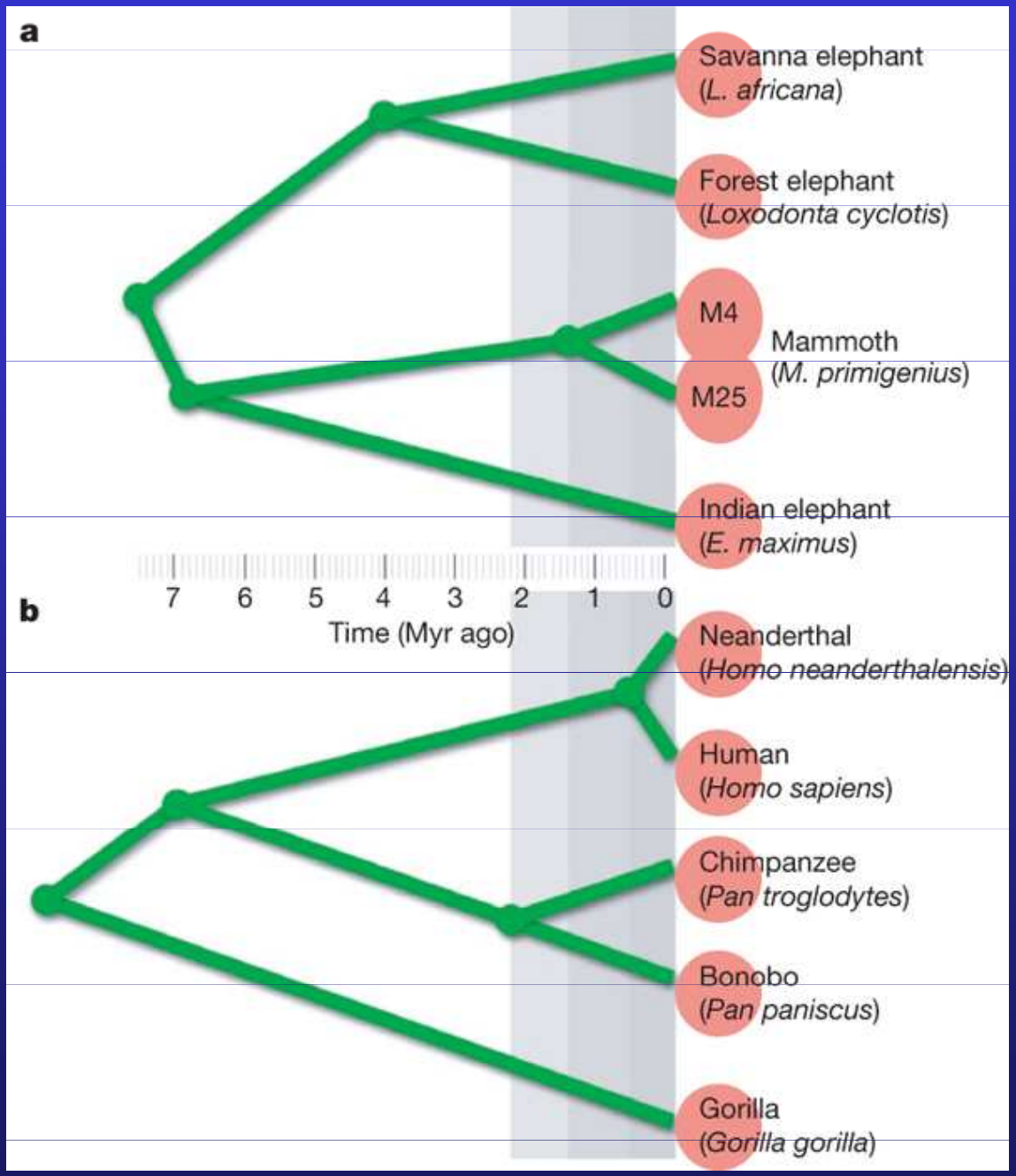


Genetická odlišnost: 0,5%

Neandertálci a lidé dnešního typu se zřejmě vůbec nekřížili

Genetická diverzita současného člověka



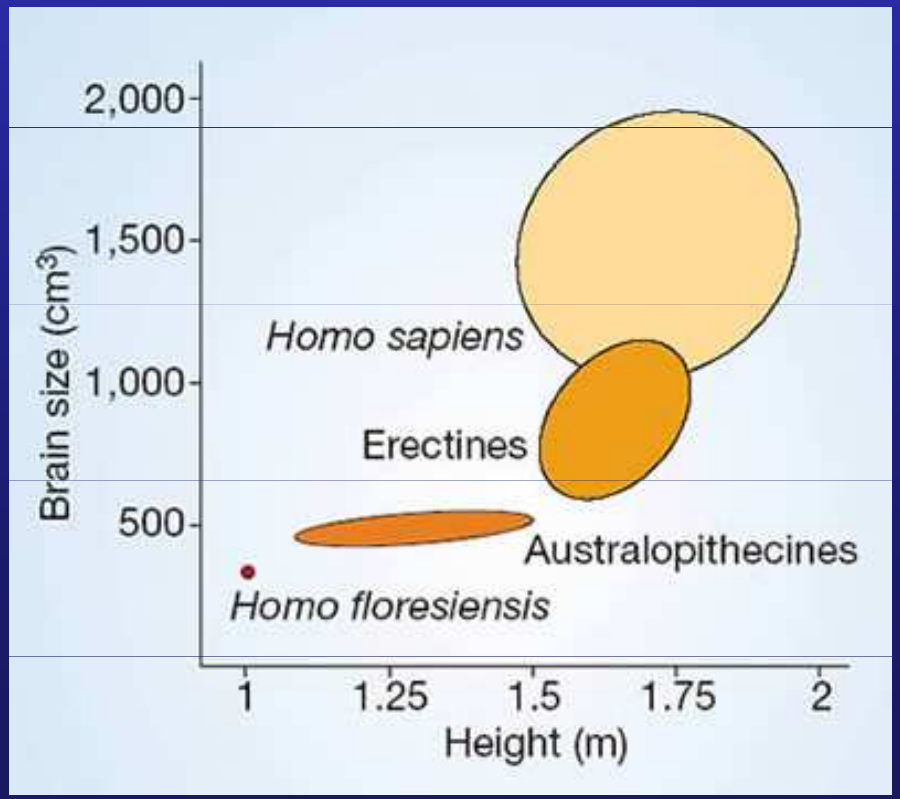


Srovnání fylogenetických stromů

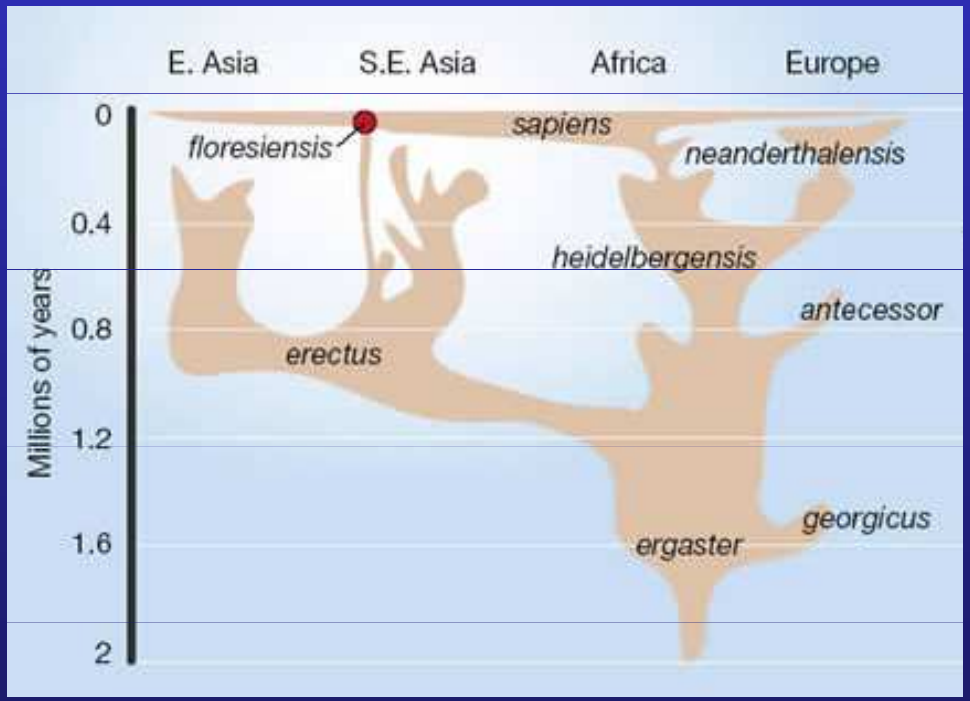
Homo floresiensis

žil před 94.000 - 13.000 lety

Velikost mozku

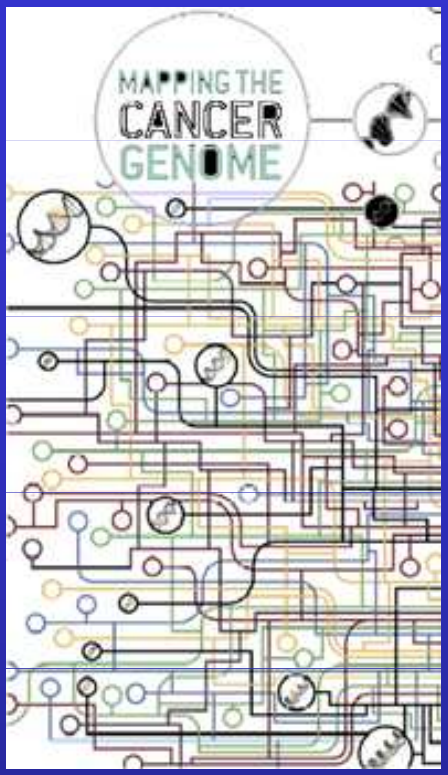


Homo floresiensis



H. floresiensis was part of the Asian dispersals of the descendants of *H. ergaster* and *H. erectus*.

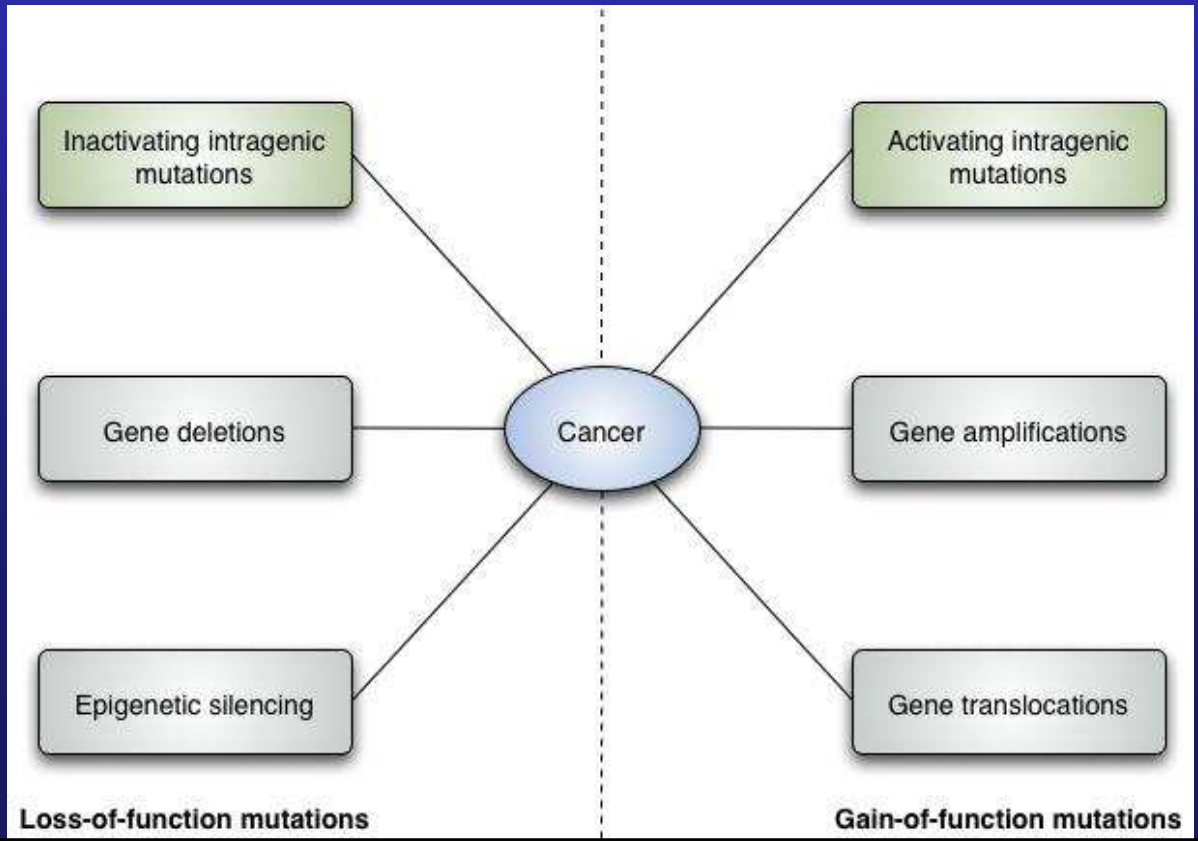
Které další projekty z HGP vycházejí?



The Cancer Genome Atlas (**TCGA**)

➢Jde o vysokokapacitní sekvenaci (tj. sekvenaci téměř celého genomu) mnoha nádorových vzorků jednoho typu nádoru od mnoha různých pacientů.

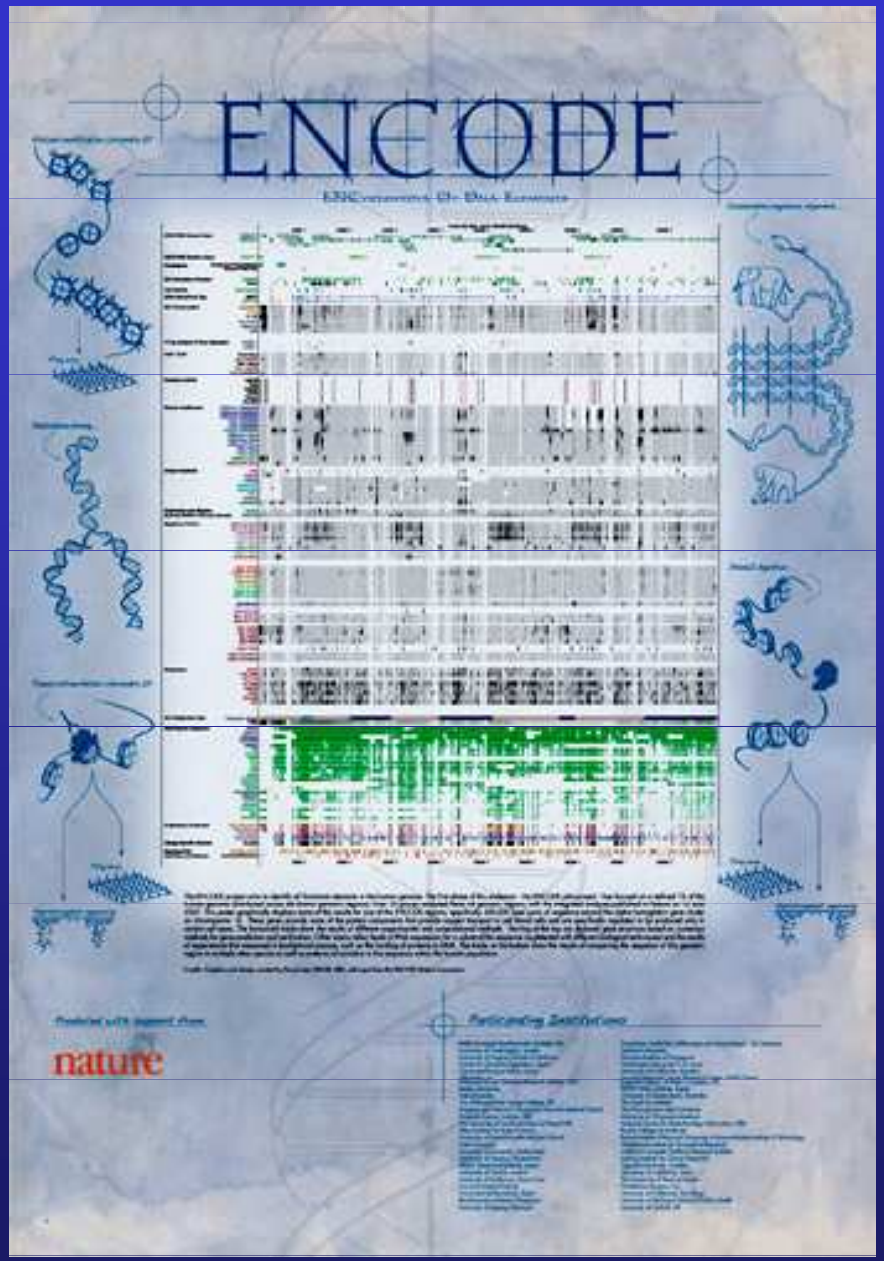
➢Tohle vše je plánováno pro mnoho typů nádorů. Jde tedy o jakýsi frontální útok na odhalení genetického pozadí nádorových onemocnění.



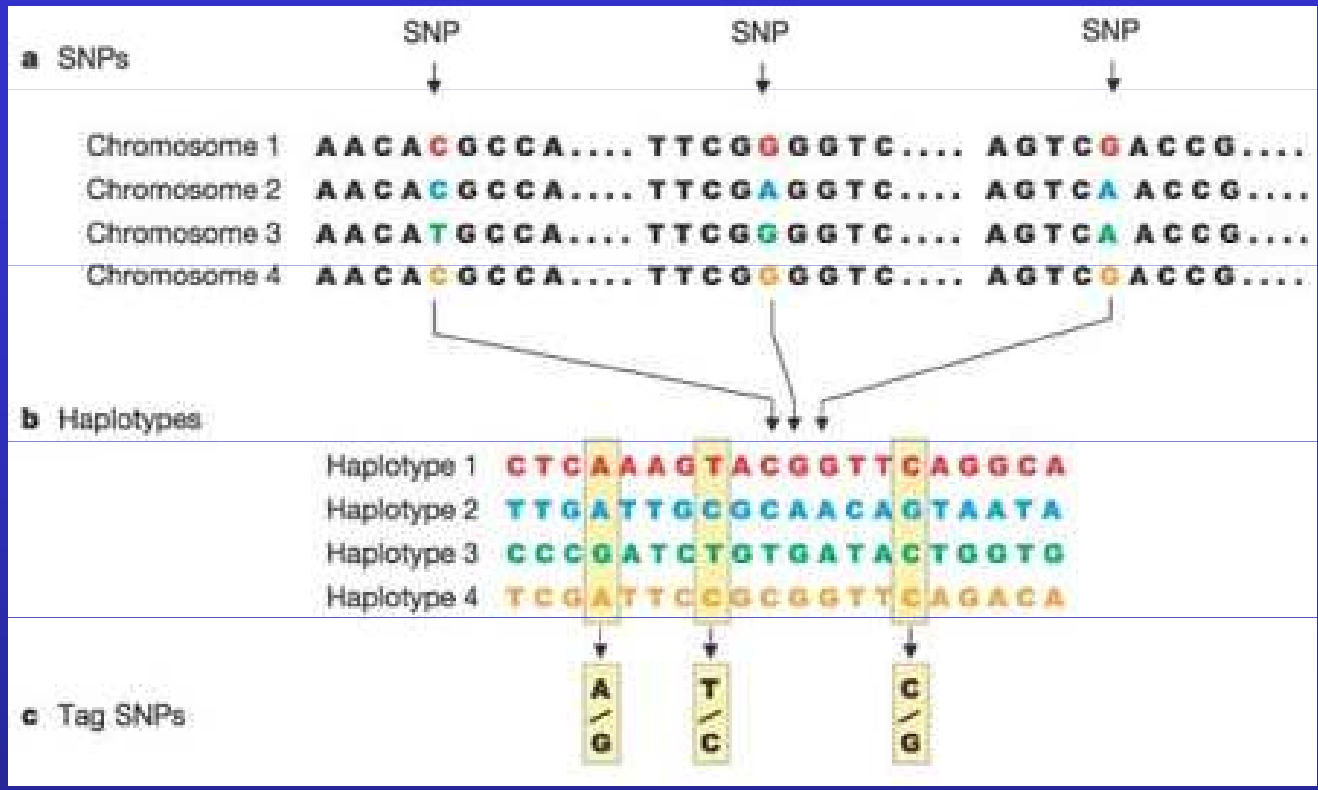
ENCODE
(*Encyclopedia of DNA Elements*)

- Transkripčně aktivní jsou také podstatné části genomu (molekuly DNA), o kterých se dosud soudilo, že jsou nefunkční a jsou pouhým „balastem“. Přitom tato DNA tvoří okolo 98 % veškeré lidské DNA.

- Znamená to tedy, že i když RNA kódovaná touto „nefunkční“ DNA není přepisována do bílkovin, tvoří se v takovém množství a na tak rozsáhlé části DNA, že nějakou její funkci lze oprávněně očekávat.



Projekt HapMap



- Vzorky DNA od z 269 lidí z Afriky, Japonska, Číny a USA
- Bylo identifikováno okolo 10 miliónů míst, ve kterých se lidé čtyř různých populací liší nejčastěji. (To přibližně odpovídá shodě 99,9 % mezi kterýmikoliv dvěma osobami.)
- Tato místa se označují jako SNP (jednonukleotidové polymorfismy).
- Platí přitom, že jednotlivé SNP se dědí po určitých blocích (haplotypech). Z toho vyplývá možnost definovat genom individuálního člověka jako kombinaci určitých haplotypů a pro zjištění genotypu konkrétní osoby tedy není třeba mapovat všech 10 miliónů jeho genomových míst, ale stačí genotypovat jen 300.000 – 600.000 klíčových SNP.

"We think it has something to do with your genome."