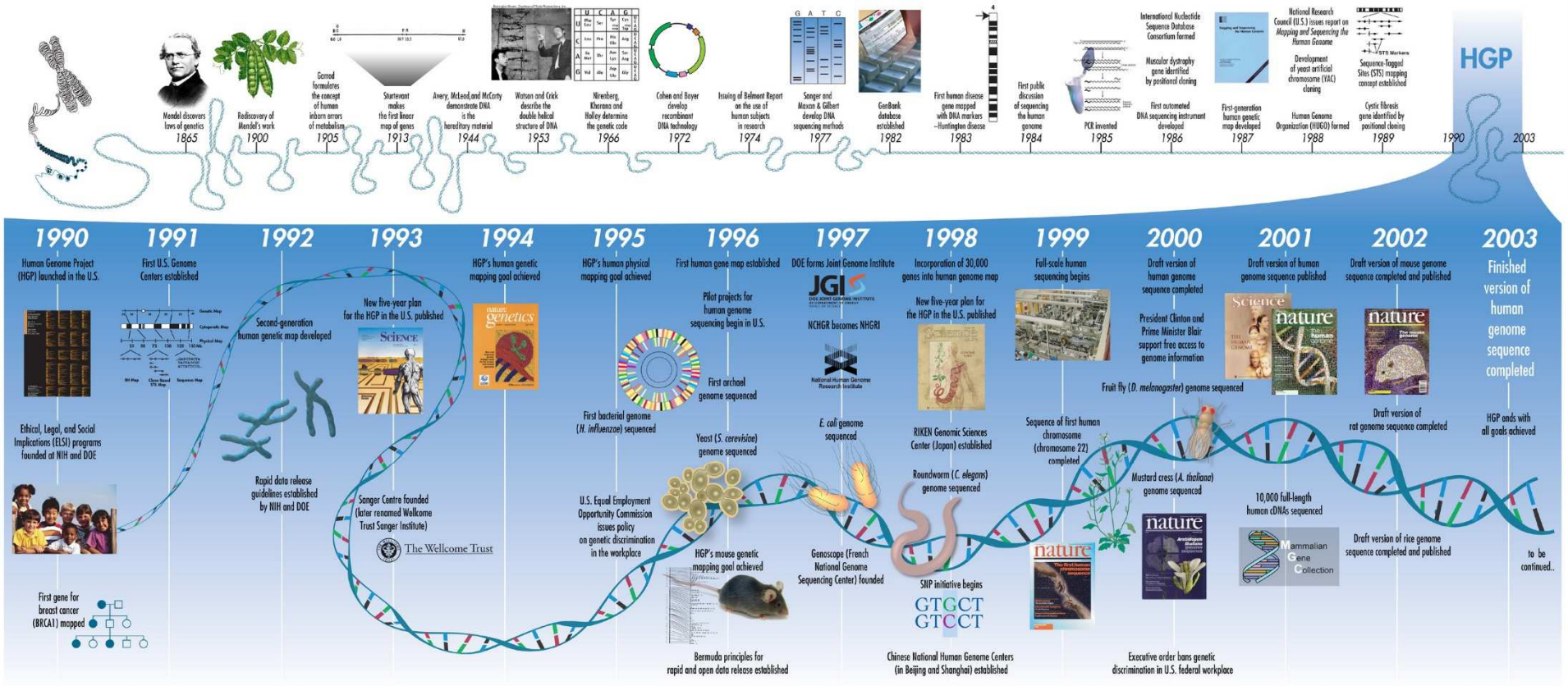
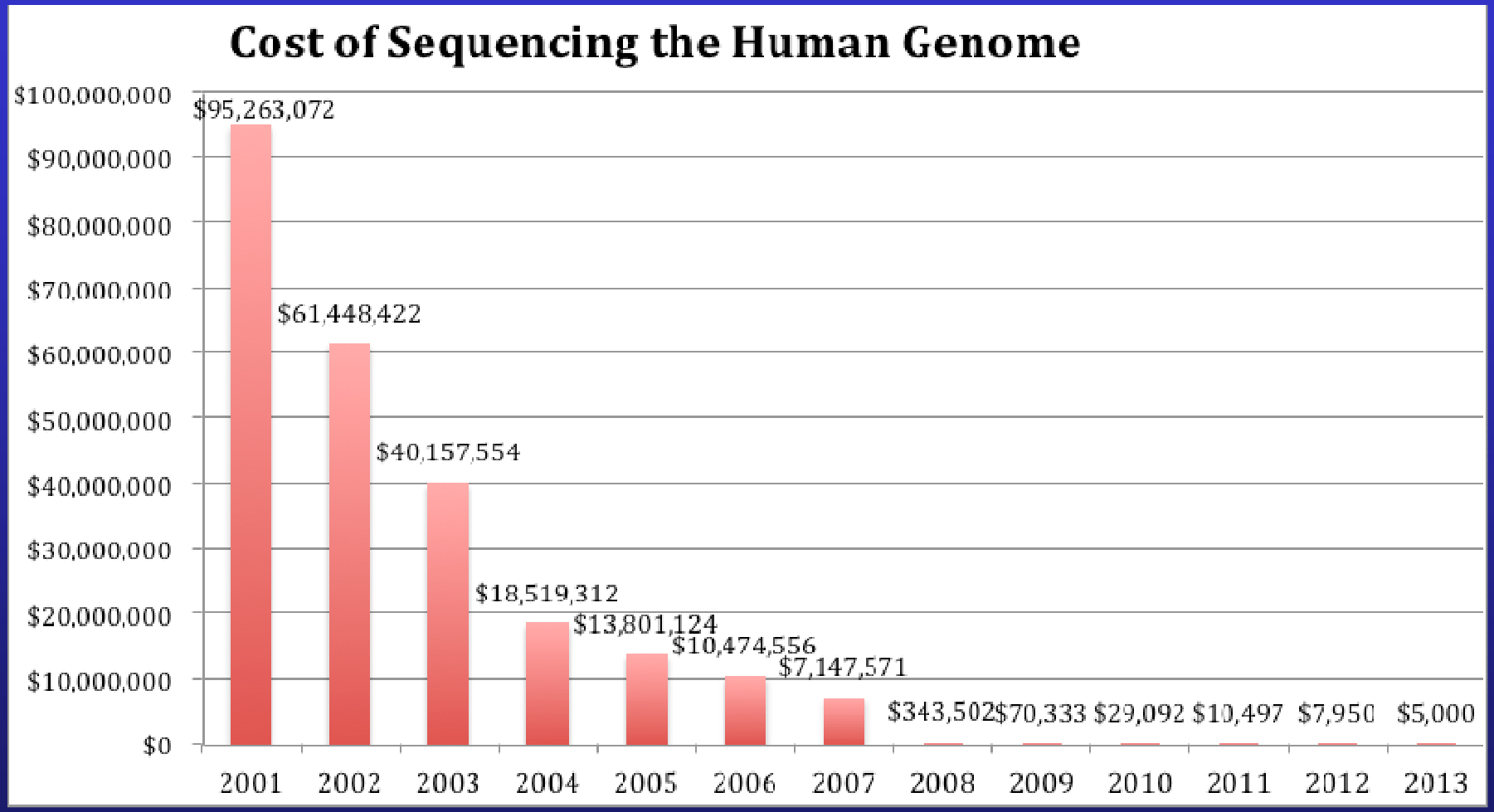
Lidský genom

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
CTCTTTCCCTTCTCTGCTCTTGGAGATGACTCTTTTGTCTGAGATTCACTTTGCTGTGCTGAAAAAGAAAAGTGCTTCAAGGAAGAT
GTTGCCTATGTTCAATTTCTGAAAATAATTAGAGCATATATACTCTGTGTGAGAAGGCAAATCCAGACAGTTAGTTTGTATGACTAG
GAGTATGTTTAATGTTCCACAGATCTCATTCTATAAATCTTTATCATCTTAGAGCTCTGATACTATTTAGAATTACTATTCCTTCAA
GCATATTAAATTGACCAGTTTTAACACACTTCTATGTATGCACAAAGATATATATTTACATTCTGCAAAATCATTCTTTCCTTTTTG
TCATTTAATGTTTAGATTCACTACATGAAATGATCCAGAAGAGAGTACTCAAATATAAGTATCTATAACGATGGAAATATACATCTC
CAGGGTCTCAATATTTAAATGTATTAAGCTTTAATTAATGTAAATTTGAATTTAGCAAAACATGTATAGCTTGTGGTTACTGTTTTA
GGACTATTTTATCTGGCTAAATTAAATGTTAAAATATTACAAATTCATCTTCAGGCTGGCTGTTGAATATTTTTATAGCAAAAGTCA
TCTACTCCCTTCTTACATACATATTCTGATGTAACATAGGTATTCTCTTATTCATGCACACTGAAATGACAACATAAATAATTTTAC
TTTTGAACTAACTTAGATGATAATTTTAATCTATATCCTAGATGAACTTTAAATCAATAAAATCTCTCAATGGTGTTATAAATCTCA
GAATGCTGTAGCATCCATGTTTAAATACTAGTTAACAAAATGCACTGGCATCAGATACAATAAGGATGAAATGAGATATAATTAGGA
CCAGCTTATTTTATTTTGAGACAGAGTCTTGCTCTCTCACTCAGGCTGGAGTGCAGTGGACCATTCTAGGCTCGCTCCAACCTCTGT
GTCACCAAGCCTGGCTGATTTTTGTAGTTTTAGTAGAGACGGGGTTTCACCATGATGGCCAGGCTGGTCTTGAACTCCTGGCCTCAA
TTTGATCCAACTTATTTGGATGAATGAGTTACATATTTTACATTAAATCTGTTATTGTGATAATTCTTCATGTTATTTTCCATGTAT
TATAAACAGGATAATAAAAATAAGACAAAAATTGTTGAAATGTCTTCATTTGACTACTAACTTTTTACATGTTTGTTACTTTGAAGC
TATTATGACACAAAGTCTATAAATTCTTATATTTTGAGATTTGTATTTAAATAACTTGTGAAATTTAATTTTAAAATAAAATTTCTT
TATATTGAATTTCTATATTATTTAACACAATTATAATTTTGCTAATGAATTGTAATGTTTTTAAAAAGCTAGGTGAATTTTATTAAA
AAAAGCTTAAAAGTTGTTATGTAGTGGCAGAGATAAAAAAGTAAAACAAAAAAAAGCTTAAAAGTTTGCTTTACTATTTATAGGCTC
AAAACACAGATAAAGCATAAAGATAGAATATAAAGATAGAAGCATTTTAATATGAGGCAGTGATGGCTTTTTGAAGAATCCCAACTA
TTTTAGGACATGGAGCAGTGACTATGAGTGCCAGAAGGCAAGAGTAGAAGCAATTGTAAAATCATGAACACTAGTTTGTAAAATCCT

1000 telefonních seznamů

Genetika → Genomika




Cost of Sequencing the Human Genome
$100,000,000
$90,000,000
$80,000,000
$70,000,000
$60,000,000
$50,000,000
$40,000,000
$30,000,000
$20,000,000
$10,000,000
$0
$95,263,072
$61,448,422
$40,157,554
$18,519,312
$13,801,124
$10,474,556
$7,147,571
$343,502
$70,333
$29,092
$10,497
$7,950
$5,000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

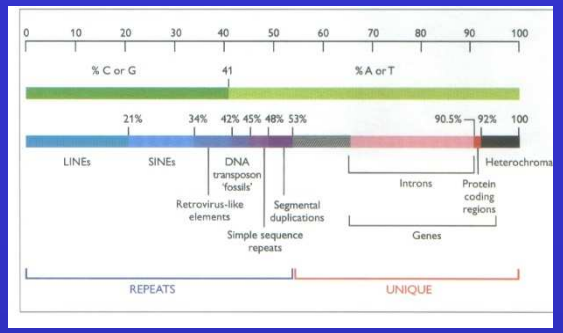
Completion of the genome

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

A recent study noted more than 160 euchromatic gaps of which 50 gaps were closed.[10] However, there are still numerous gaps in the heterochromatic parts of the genome which is much harder to sequence due to numerous repeats and other intractable sequence features.

The sequence rainbow. The output from an automated DNA sequencing machine used to determine the complete human genome sequence. Each vertical line shows the sequence of bases in a given stretch of DNA. Each of the four different bases is labelled with one of the four coloured dyes. The order of the bases is analysed by a computer and assembled to give the continuous base sequence for a segment of the genome. [Sanger Centre/Wellcome Photo Library]



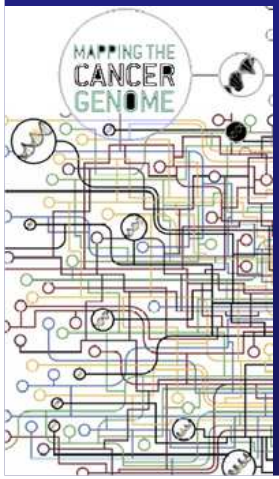
How the genome was sequenced?

What's in a genome?

- **Geny podmiňující gen. choroby** – pozční 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

What are application to Medicine and Biology?

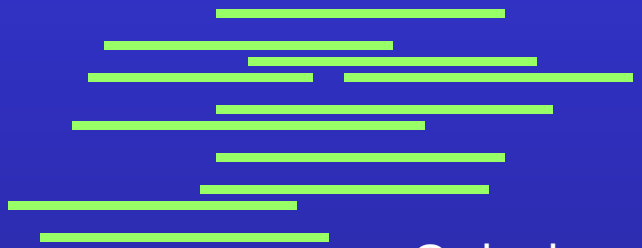
What are genetic differences between Modern Man and other species?



What are other projects that follow HGP?

Genome Sequencing

Genome: 3 Gb

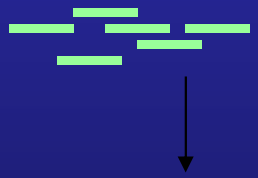


Cut genome into large pieces

Clone into BACs: 100 kb

Order based on sequence features (*markers*) = mapping

Cut again



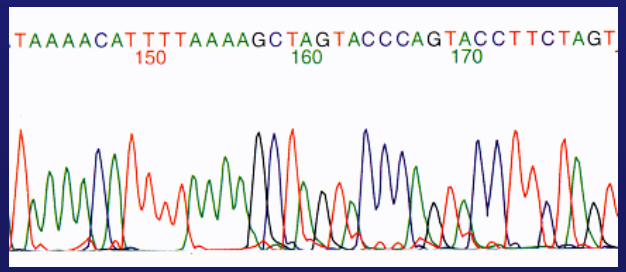
Assemble entire sequence

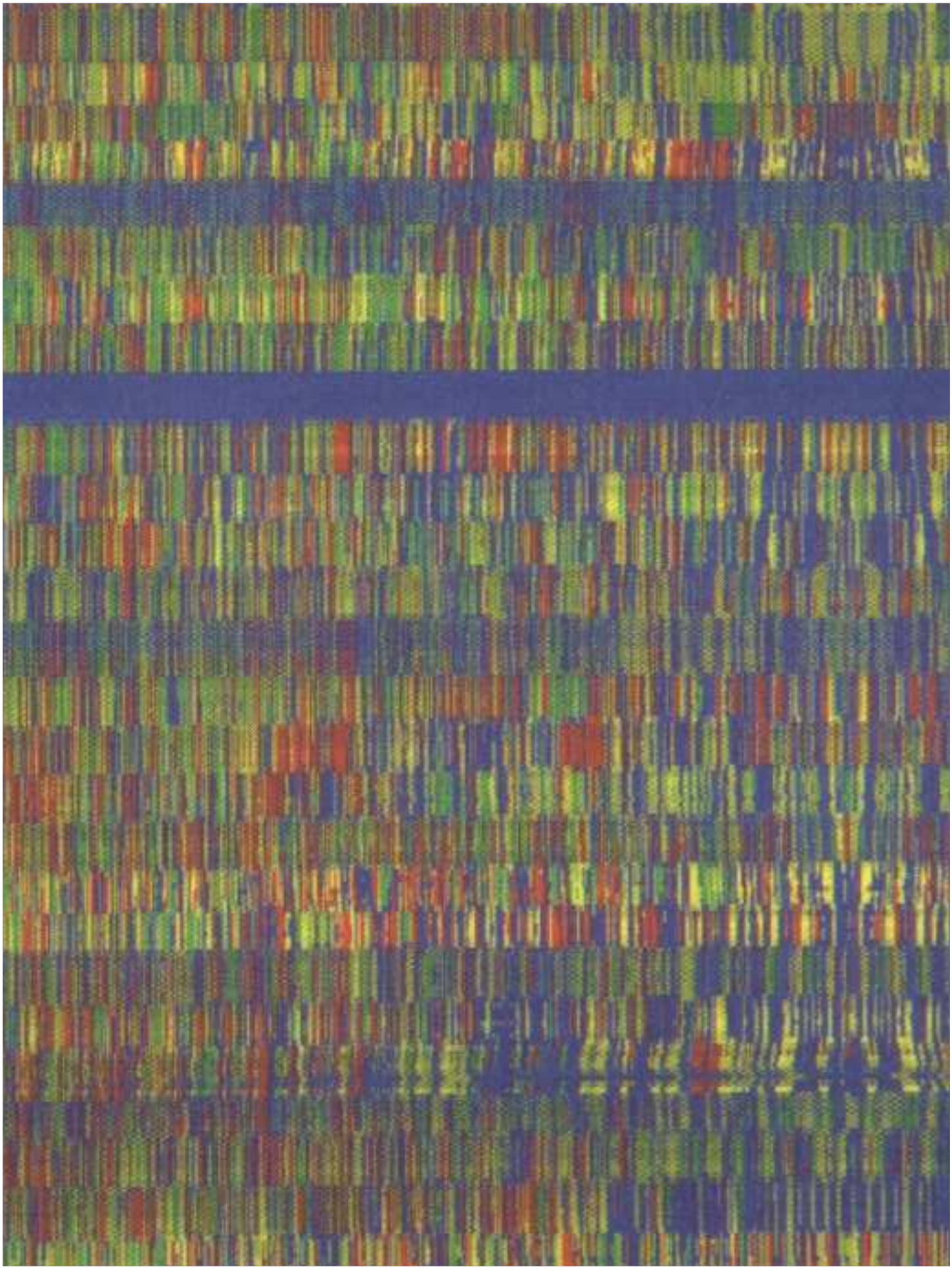
…TTGTAAGTGAGAACAGGACGTATGTGGTTTTCTACTCCTGTGTT…

Assemble each BAC

TTGTAAGTGAGAACA
AGAACAGGACGTATGTGGT
TGTGGTTTTCTACTCC
CTACTCCTGTGTT

Sequence





What does the sequence mean?

A very small piece of chromosome 21

What's in a genome?

Genes (i. e., protein coding)

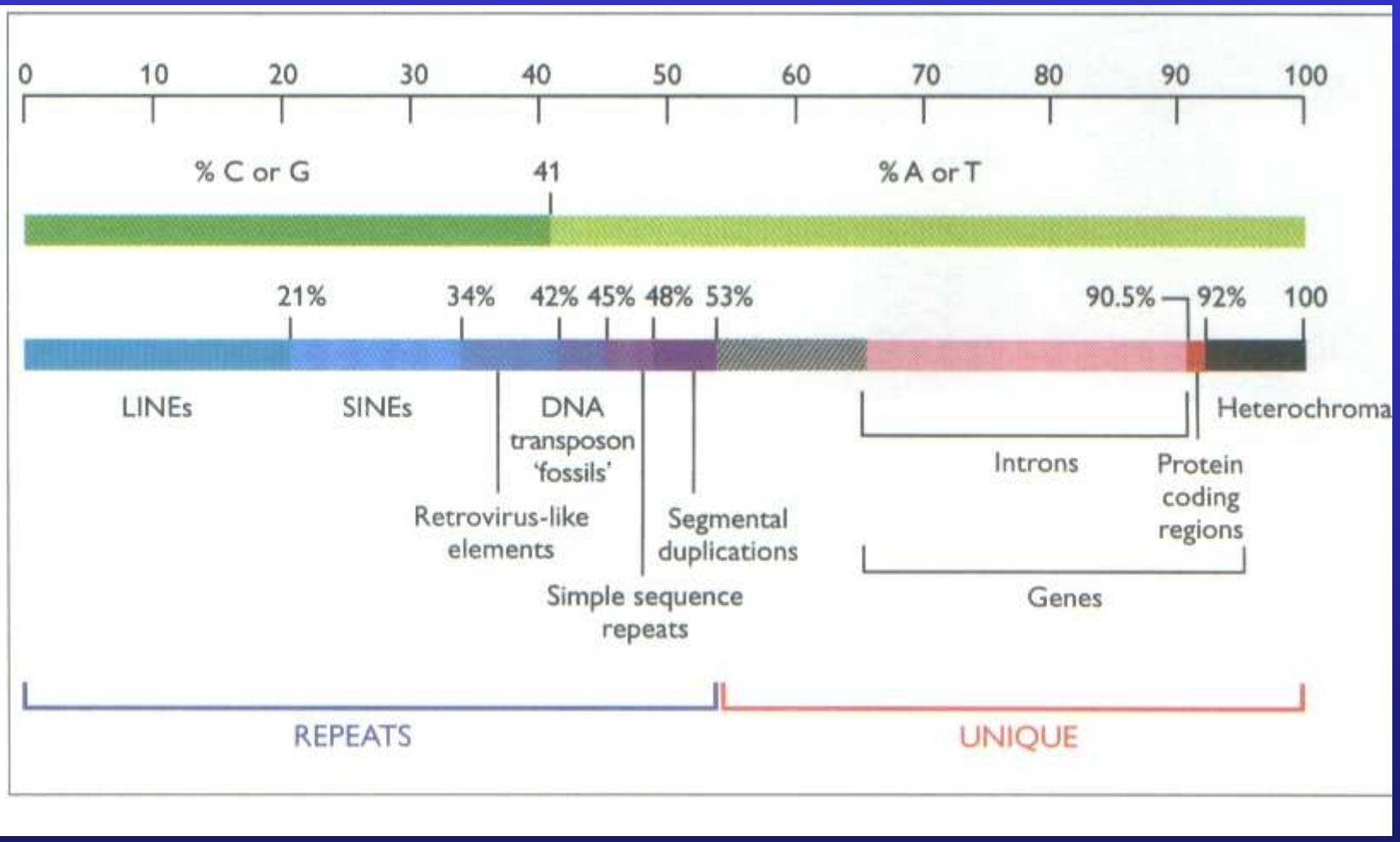
But. . . only <2% of the human genome encodes proteins 21.000±1.000 protein-coding genes

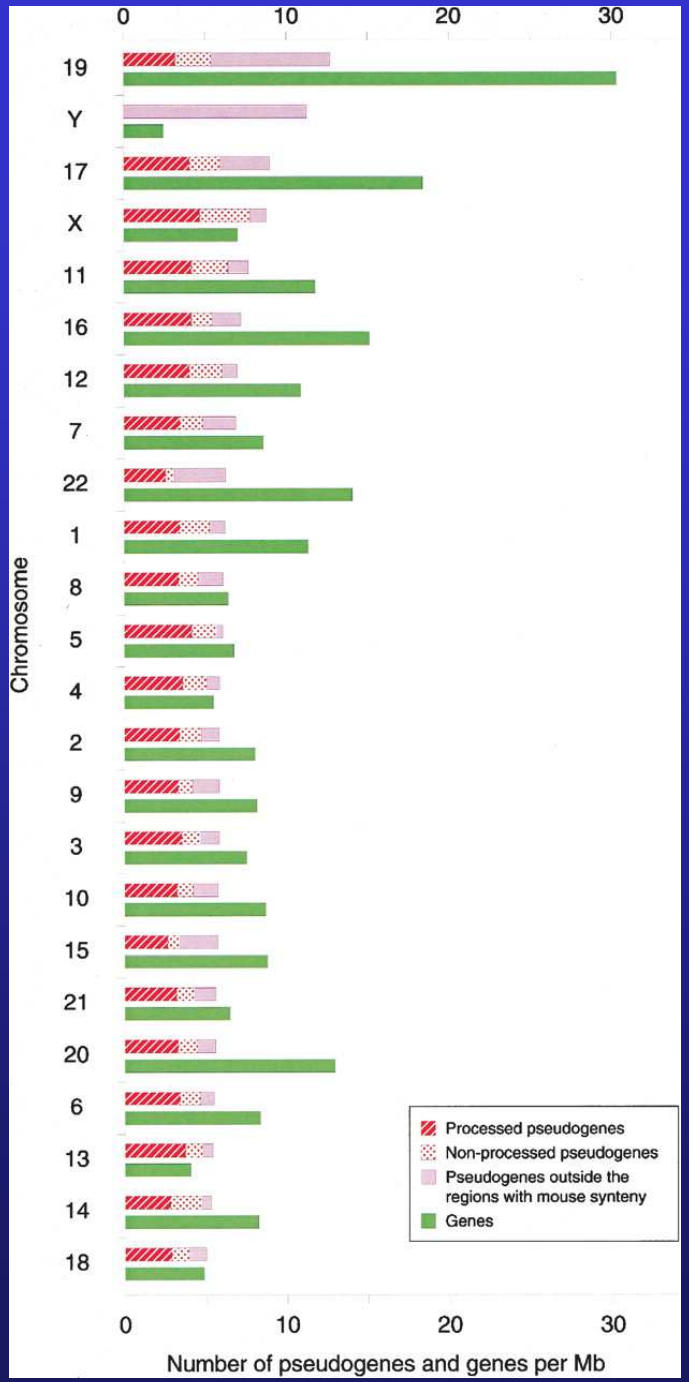
Other than protein coding genes, what is there?

- genes for noncoding RNAs (rRNA, tRNA, miRNAs, etc.)
- structural sequences (scaffold attachment regions)

Regulatory sequences

- "junk" (including transposons, retroviral insertions, etc.)

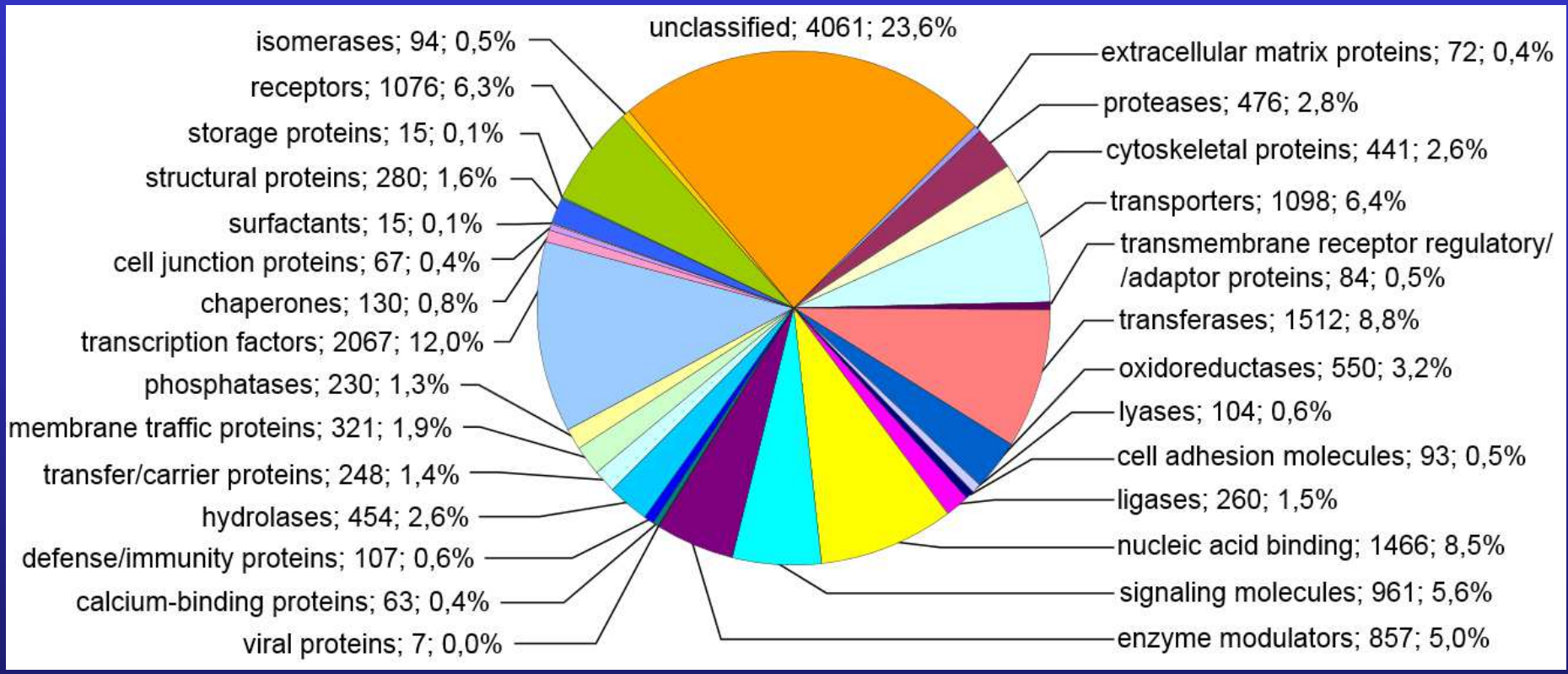

0
10
20
30
40
50
60
70
80
90
100
% C or G
41
% A or T
21%
34%
42%
45%
48%
53%
90.5%
92%
100
LINEs
SINEs
DNA transposon 'fossils'
Retrovirus-like elements
Segmental duplications
Simple sequence repeats
Introns
Protein coding regions
Heterochroma
Genes
REPEATS
UNIQUE



Pseudogenes

- 70% of processed pseudogenes (retrotransposon genes)
- 30 % unprocessed pseudogenes (duplicated genes)
- ~20,000

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


unclassified; 4061; 23,6%
extracellular matrix proteins; 72; 0,4%
proteases; 476; 2,8%
cytoskeletal proteins; 441; 2,6%
transporters; 1098; 6,4%
transmembrane receptor regulatory/ /adaptor proteins; 84; 0,5%
transferases; 1512; 8,8%
oxidoreductases; 550; 3,2%
lyases; 104; 0,6%
cell adhesion molecules; 93; 0,5%
ligases; 260; 1,5%
nucleic acid binding; 1466; 8,5%
signaling molecules; 961; 5,6%
enzyme modulators; 857; 5,0%
isomerases; 94; 0,5%
receptors; 1076; 6,3%
storage proteins; 15; 0,1%
structural proteins; 280; 1,6%
surfactants; 15; 0,1%
cell junction proteins; 67; 0,4%
chaperones; 130; 0,8%
transcription factors; 2067; 12,0%
phosphatases; 230; 1,3%
membrane traffic proteins; 321; 1,9%
transfer/carrier proteins; 248; 1,4%
hydrolases; 454; 2,6%
defense/immunity proteins; 107; 0,6%
calcium-binding proteins; 63; 0,4%
viral proteins; 7; 0,0%

What are application to Medicine and Biology?

- Geny podmiňující gen. Choroby

  - rok 2000 – 1300 genů pro choroby s jednoduchou *mendelovskou dědičností*
  - rok 2010 – 2900 genů
  - Zbývá asi 1800

  - 1100 genů asociovaných s 165 častými chorobami(včetně chorob infekčních)
  - (např. IBD 70-100 genů)

What are application to Medicine and Biology?

- 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)

Mutation rate in humans

1. Error rate in DNA replication (1.0×10^{-10} per bp)

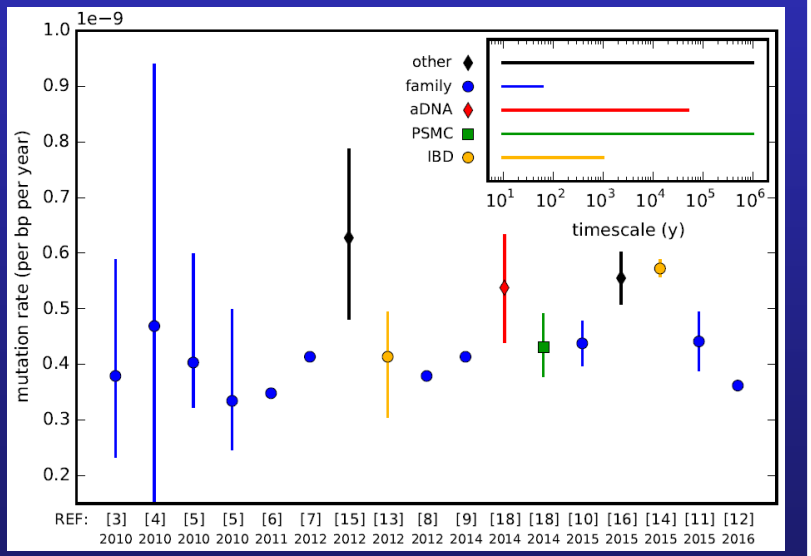
Given that the human genome is 3.2×10^9 bp, there are about **30 cell generations** between zygote and egg cells and about **400 cell divisions** between zygote and mature sperm. Thus, in males, the sperm cells have about 128 new mutations and the haploid egg genome has about 10 new mutations for a total of **138** new mutations in every new zygote.

2. Phylogenetic method (humans vs. chimpanzees)

112-160 new mutations in every new generation.

3. Direct method (human families)

56-103 new mutations in every new generation,
but only 60-80.% of the genome sequence is reliable



More than 100 new mutations in every new generation.

number of mutations throughout humanity per generation

$$\approx 10\text{–}100 \ \frac{\text{mutations}}{\text{generation}} \times 7 \times 10^9 \text{ people} \approx 10^{11}\text{–}10^{12} \text{ mutations per generation globally}$$

$$\approx 3 \times 10^9 \ \frac{\text{base-pairs}}{\text{human genome}} \Longrightarrow$$

tens to hundreds of people with *de novo* appearance of any specific mutation

$$\frac{(3\times10^9)^2 \text{ possible 2 bp combinations}}{10^{12} \text{ mutations/generation}} \Longrightarrow$$

$> 10^6$ generations for humanity to randomly reach a specific 2 base-pair mutation

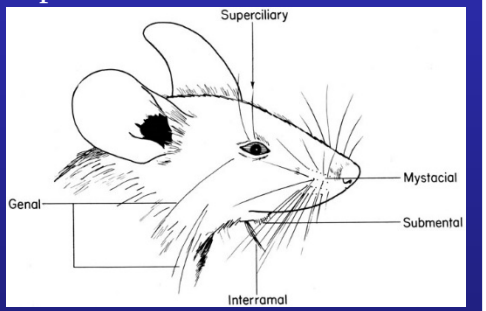
We find that each single base pair mutation is explored dozens of times in every generation but that a specific combination of two base pairs will require an unrealistic number of generations to occur at random.

Mclean et al., 2011 Nature

Search for complete deletion of sequences otherwise highly conserved between chimpanzees and other mammals. **510 such deletions** were identified in humans, which fall almost exclusively in non-coding regions and are enriched near genes involved in steroid hormone signalling and neural function.

One deletion removes a **sensory vibrissae** and **penile spine** enhancer from the human androgen receptor (AR) gene, a molecular change correlated with anatomical **loss of androgen-dependent sensory vibrissae and penile spines in the human lineage**.

Another deletion removes a forebrain subventricular zone enhancer near the tumour suppressor gene growth arrest and DNA-damage-inducible, gamma (GADD45G), a loss correlated with expansion of specific brain regions in humans.

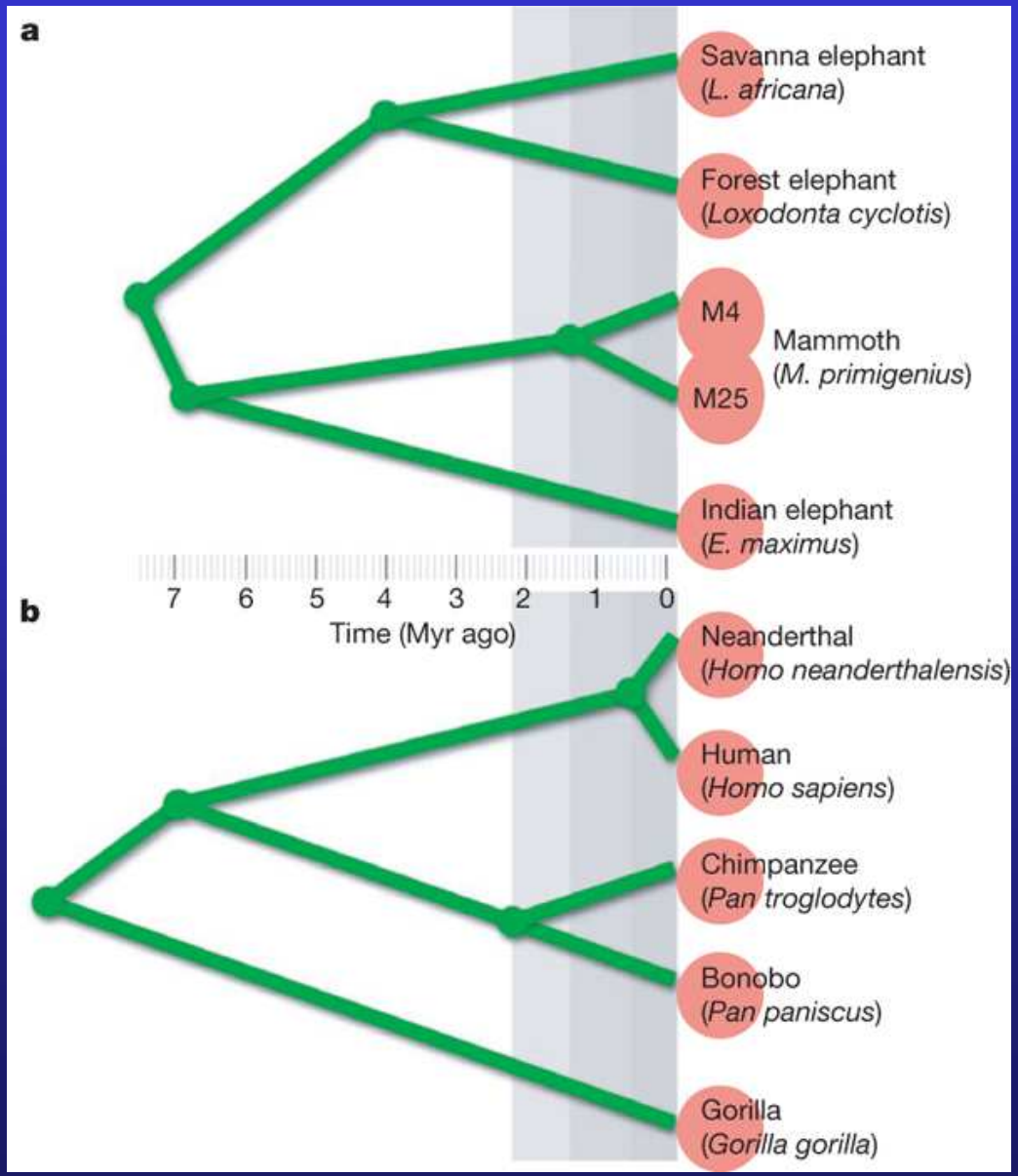


The loss of genetic information may have an important role in human evolution!

Chimpanzee Sequencing and Analysis Consortium. **Initial sequence of the chimpanzee genome and comparison with the human genome.**

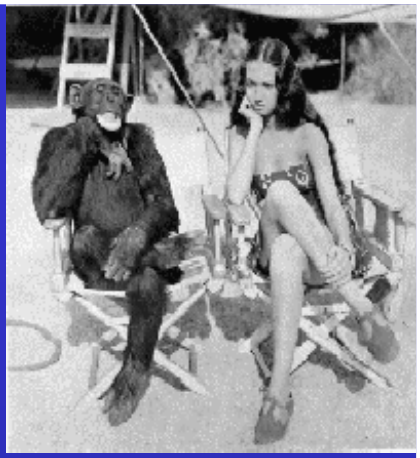
Nature **2005** Sep 1;437(7055):69-87.

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



Phylogenetic trees

Apparent differences between humans and great apes in the incidence or severity of medically important conditions (excluding differences explained by obvious anatomical differences).

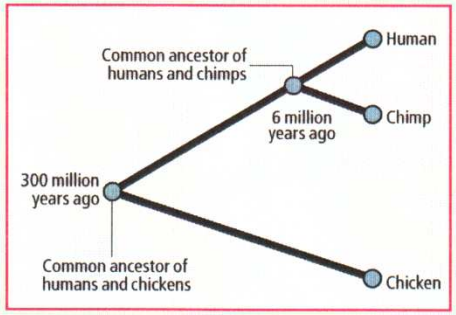


Medical Condition	Humans	Great Apes
Definite		
HIV progression to AIDS	Common	Very rare
Influenza A symptomatology	Moderate to severe	Mild
Hepatitis B/C late complications	Moderate to severe	Mild
P. falciparum malaria	Susceptible	Resistant
Menopause	Universal	Rare
Likely		
E. coli K99 gastroenteritis	Resistant	Sensitive?
Alzheimer's disease pathology	Complete	Incomplete
Coronary atherosclerosis	Common	Uncommon
Epithelial cancers	Common	Rare

[EXPERIMENT]

SCANNING THE GENOME

To find the parts of our genome that make us human, the author wrote a computer program that searched for the DNA sequences that have changed the most since humans and chimpanzees diverged from their last common ancestor. Topping the list was a 118-letter snippet of code known as human accelerated region 1 (HAR1). This region of the genome changed very little for most of vertebrate evolution, with chimp and chicken sequences differing by just two letters. Human and chimp HAR1s, however, differ by 18 letters, suggesting that HAR1 acquired an important new function in humans.



T G A A A **C** G G A G G A G A **C G** T T A C
A G C A A **C G** T **G** T C A **G** C T G A A A T
G A T **G** G G **C** G T A G A C **G** C A **C** G T C
A G C **G** G **C** G G A A A T **G** G T T T C T A
T C A A A A T **G** A A A G T **G** T T T A G A
G A T T T T C C T C A A **G** T T T C A

Changes in human sequence relative to that of the chimp

T G A A A T G G A G G A G A A A T T A C
A G C A A T T T A T C A A C T G A A A T
T A T A G G T G T A G A C A C A T G T C
A G C A G T **G** G A A A **T** A G T T T C T A
T C A A A A T T A A A G T A T T T A G A
G A T T T T C C T C A A A T T T C A

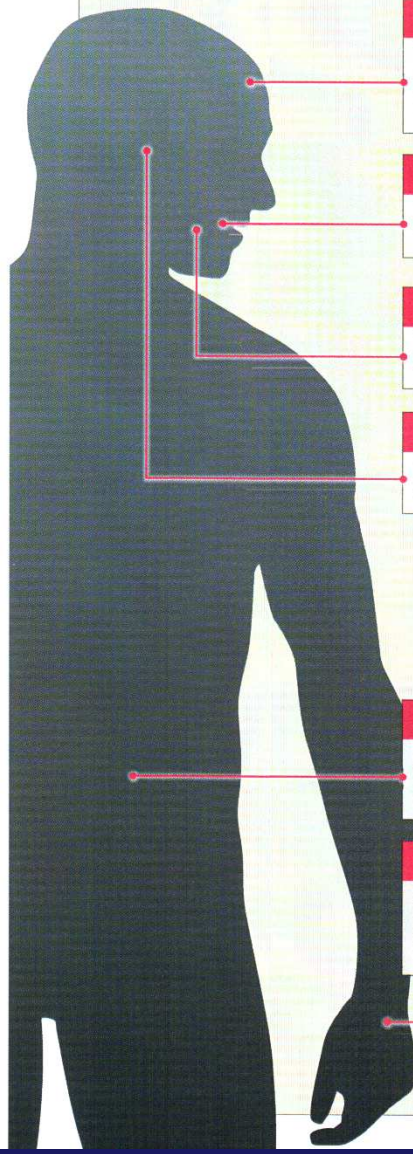
Changes in chimp sequence relative to that of the chicken

T G A A A T G G A G G A G A A A T T A C
A G C A A T T T A T C A A C T G A A A T
T A T A G G T G T A G A C A C A T G T C
A G C A G T A G A A A C A G T T T C T A
T C A A A A T T A A A G T A T T T A G A
G A T T T T C C T C A A A T T T C A

[FINDINGS]

DISTINCTIVE DNA

Efforts to uncover uniquely human DNA have yielded a number of sequences that are distinctive in humans as compared with chimpanzees. A partial list of these sequences—and some of their functions—follows below.



SEQUENCE: HAR1

What it does: Active in the brain; may be necessary for development of the cerebral cortex, which is especially large in humans. Possibly also involved in sperm production.

SEQUENCE: *FOXP2*

What it does: Facilitates formation of words by the mouth, enabling modern human speech.

SEQUENCE: *AMY1*

What it does: Facilitates digestion of starch, which may have enabled early humans to exploit novel foods.

SEQUENCE: *ASPM*

What it does: Controls brain size, which has more than tripled over the course of human evolution.

SEQUENCE: *LCT*

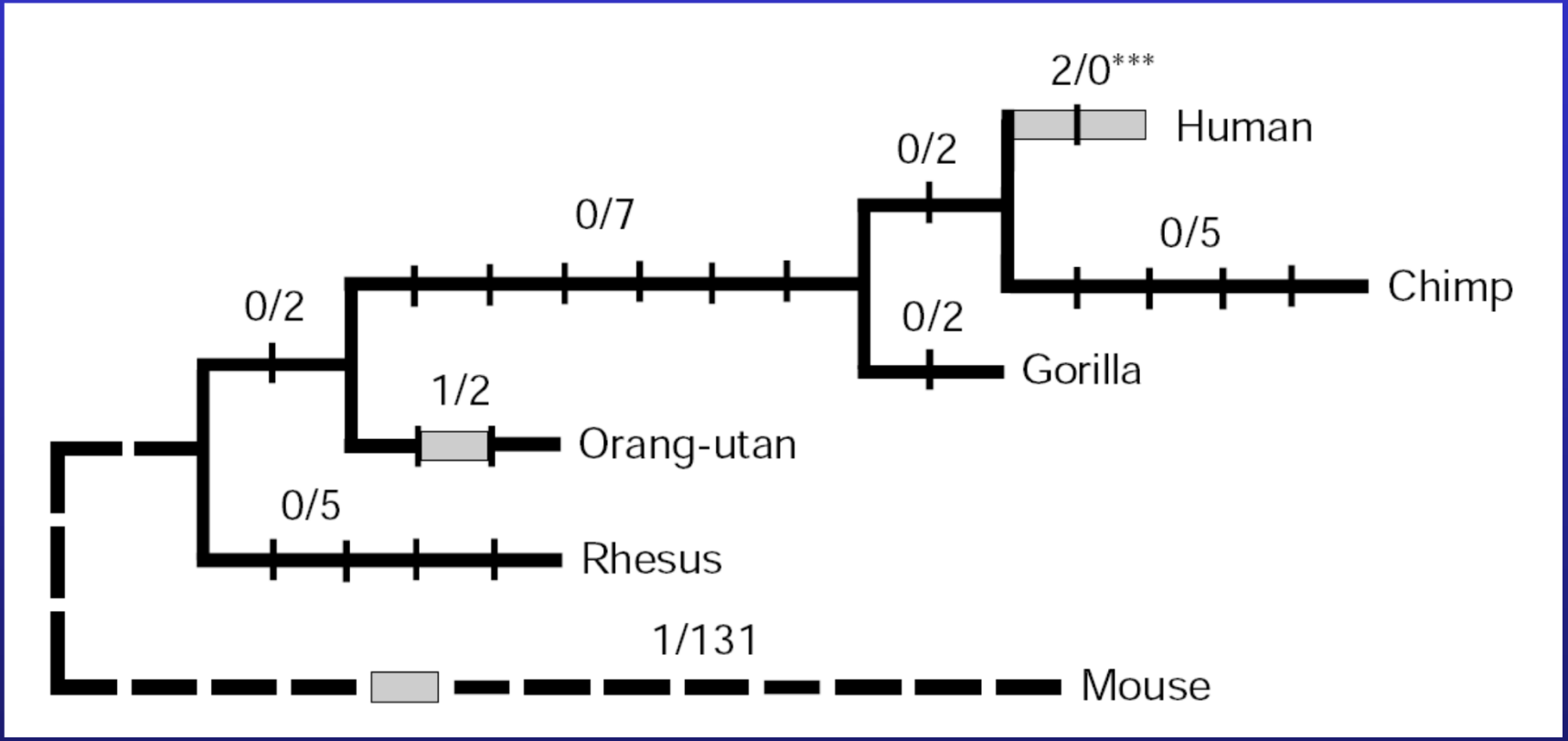
What it does: Permits digestion of milk sugar in adulthood, allowing people to make milk from domesticated animals a dietary staple.

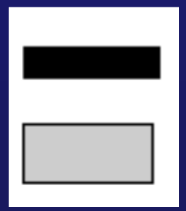
SEQUENCE: HAR2

What it does: Drives gene activity in the wrist and thumb during development, an activity that may have given the hand enough dexterity to make and use complex tools.

BRAIN SHAPERS: Changes to certain genome sequences can have dramatic effects on the brain. Mutation of the *ASPM* gene, for example, leads to markedly reduced brain size (*middle*) compared with a normal brain (*top*), suggesting that this gene played a key role in the evolution of large brain size in humans. Malfunctions in the neurons in which HAR1 is active during development, meanwhile, can lead to a severe disorder in which the brain's cerebral cortex fails to fold properly (*bottom*), hinting that HAR1 is essential for the formation of a healthy cortex.

Evolution of FOXP2



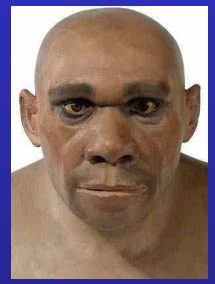


Base substitution

Amino acid substitution

Genetic differences between Modern Man (*Homo sapiens*) and other species

Species	Time from last common ancestor	Genome sequence identity
Modern Man (*Homo sapiens*),		~99,9 %
Neanderthal Man (*Homo neandertalensis*)	0,5 mil. years	≈99,5 %
Chimpanzee (*Pan troglodytes*)	6 mil. years	~96 %
Rhesus macaque (*Maccaca mulata*)	25 mil. years	~93 %
Rat (*Rattus norvegicus*)	75 mil. years	~40 % of the genome is sequentially related (not identical)

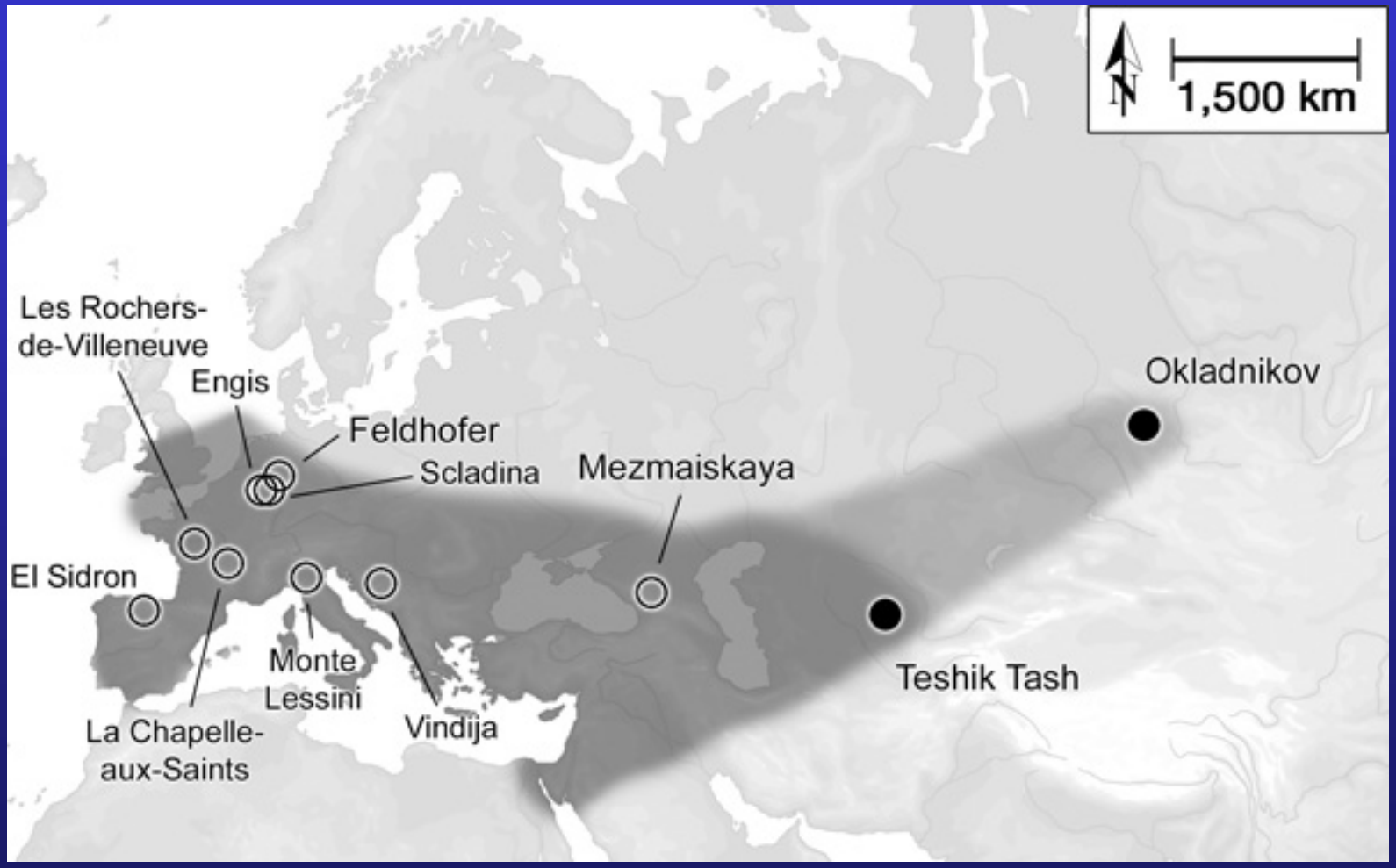


http://www.msnbc.msn.com/id/13154583/

Neanderthal Man (*Homo neandertalensis*)

- an extinct member of the Homo genus
- appeared in Europe 500.000 years ago
- Europe and West Asia
- extinct about 30,000 years ago
- interbreeding took place with modern humans between roughly 80,000 to 50,000 years ago in the Middle East

Neanderthal Man (*Homo neandertalensis*)



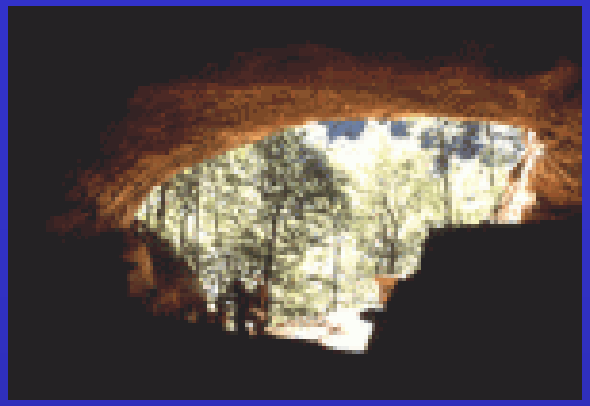
Neanderthal Man DNA

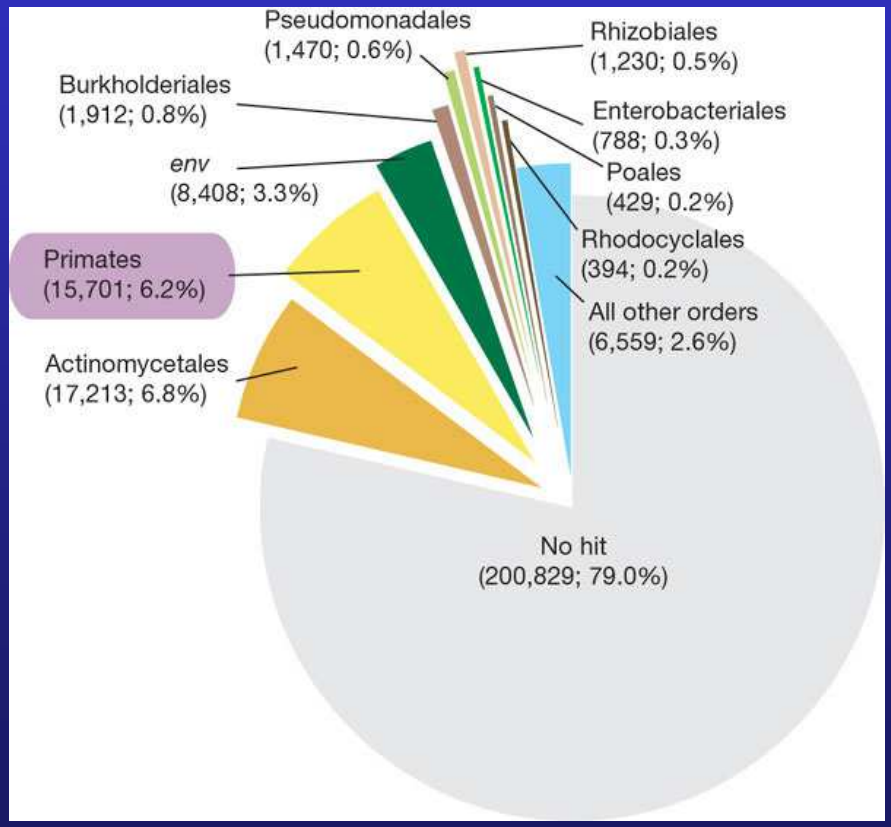
- Bone fouund in 1980 in Vindija Cave, Croatia
- Radioisotope dating:

38,310 ± 2,130 let

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

Bone Vi-80 (from Vindija Cave)



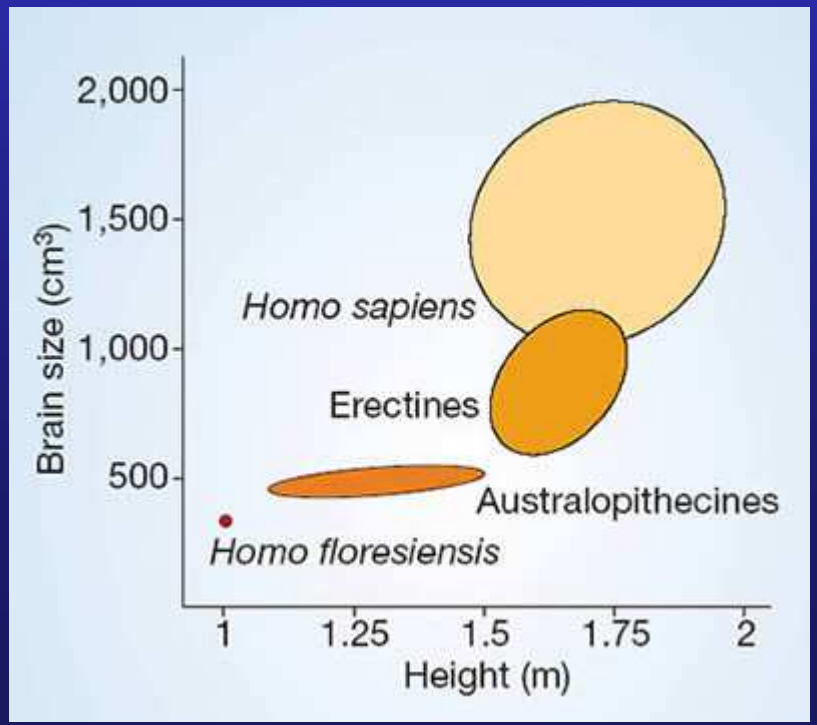


Genetic identity:
99,5%

1–4% of the genome of people from Eurasia having been contributed by Neanderthals

Homo floresiensis
the Hobbit

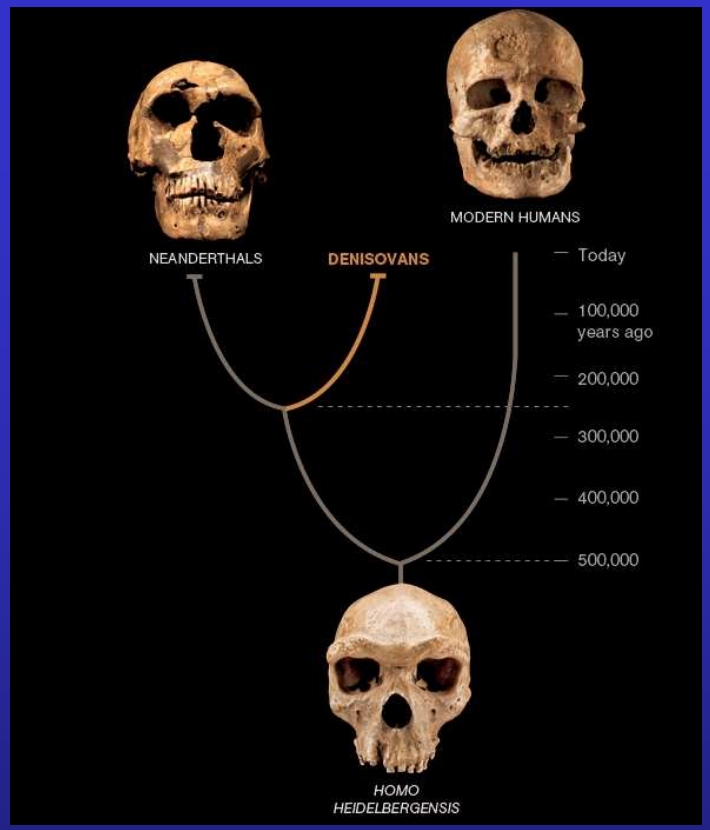
94.000 - 13.000 years



The **Denisovan** is an extinct subspecies of human in the genus *Homo*. Pending its status as either species or subspecies it currently carries the temporary name ***Homo* sp. *Altai***, and ***Homo sapiens* ssp. *Denisova.***

In March 2010, scientists announced the discovery of a finger bone fragment of a juvenile female who lived about 41,000 years ago, found in the remote Denisova Cave in the Altai Mountains in Siberia.

Analysis of the mitochondrial DNA (mtDNA) of the finger bone showed it to be genetically distinct from the mtDNAs of Neanderthals and modern humans.



Denisovans shared a common origin with Neanderthals, that they ranged from Siberia to Southeast Asia, and that they lived among and interbred with the ancestors of some modern humans, with about 3% to 5% of the DNA of Melanesians and Aboriginal Australians deriving from Denisovans

African Cradle

Most paleoanthropologists and geneticists agree that modern humans arose some 200,000 years ago in Africa. The earliest modern human fossils were found in Omo Kibish, Ethiopia. Sites in Israel hold the earliest evidence of modern humans outside Africa, but that group went no farther, dying out about 90,000 years ago

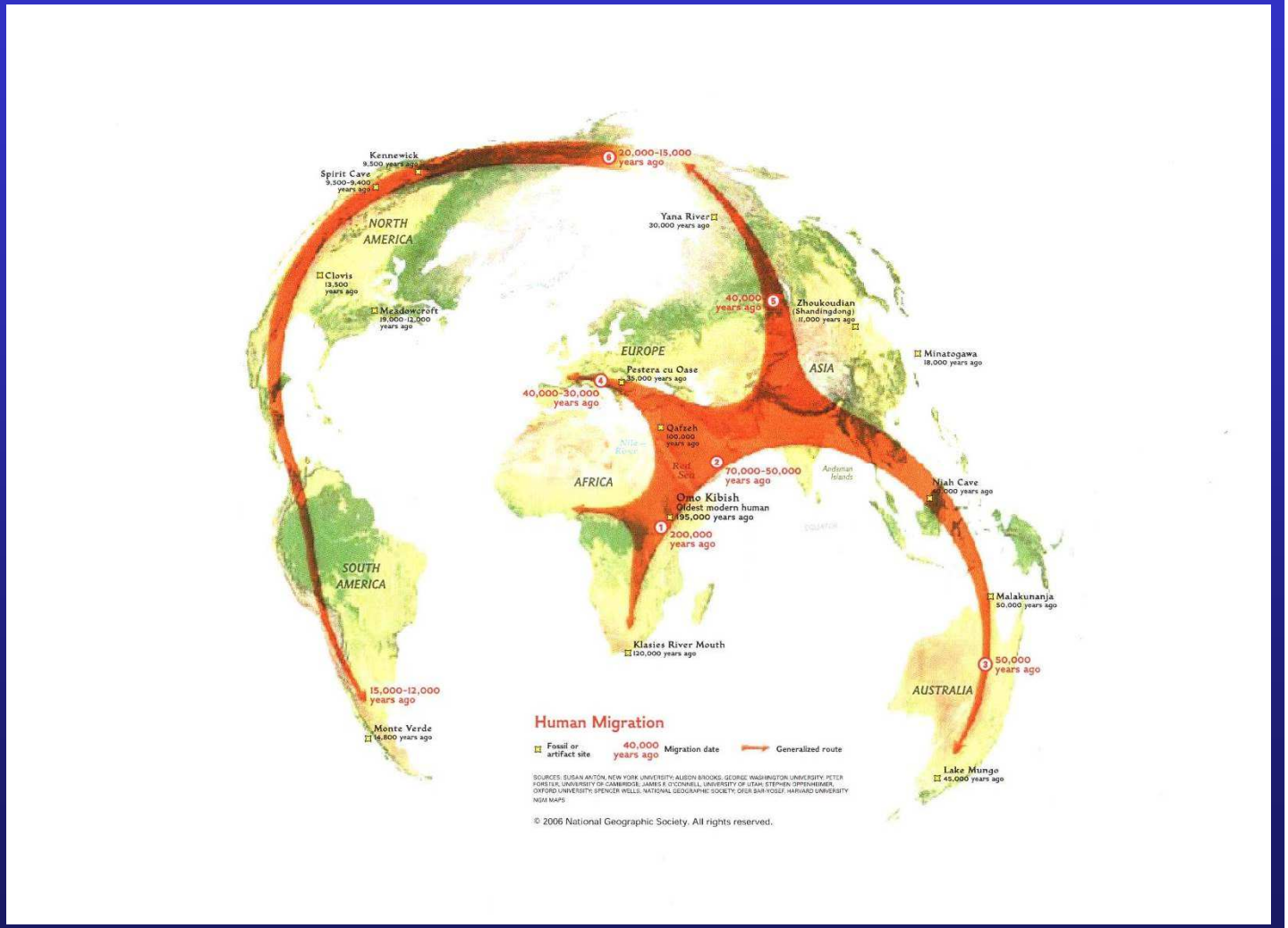
Most paleoanthropologists and geneticists agree that modern humans arose some 200,000 years ago in Africa.

Genetic data show that a small group of modern humans left Africa for good 70,000 to 50,000 years ago.

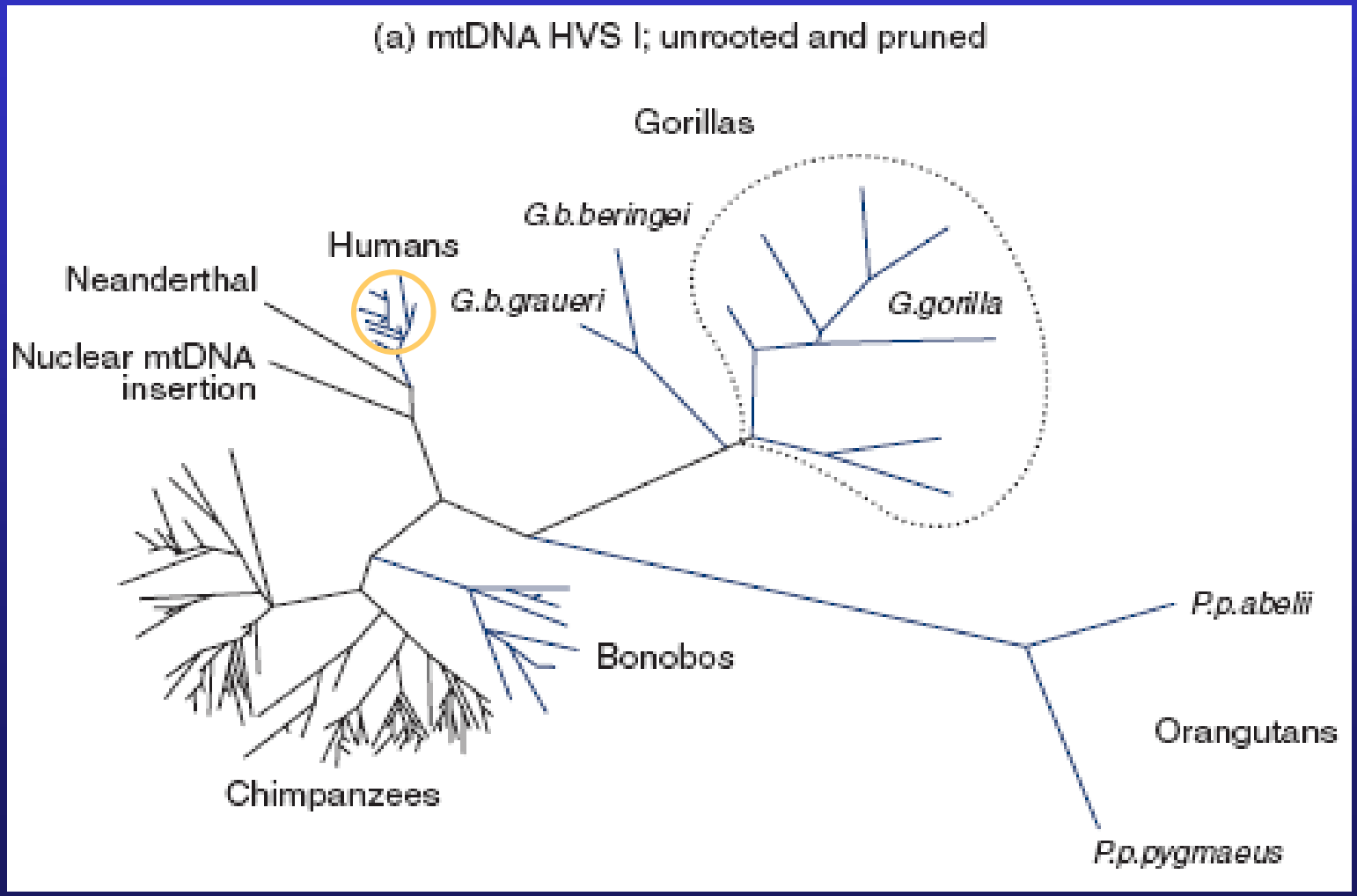
Modern humans followed a coastal route along southern Asia and reached Australia nearly 50,000 years ago.

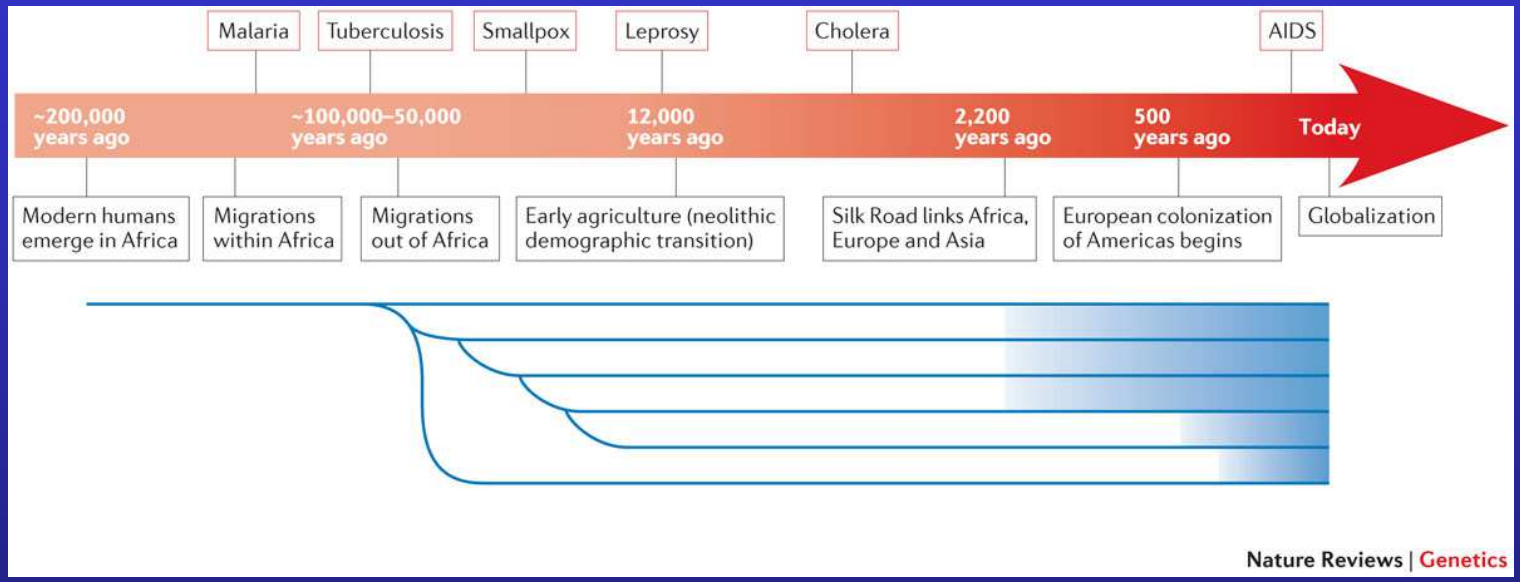
Genetic data show that the DNA of today's western Eurasians resembles that of people in India. It's possible that an inland migration from Asia seeded Europe between 40,000 and 30,000 years ago.

Genetic evidence suggests it was between 20,000 and 15,000 years ago, when sea levels were low and land connected Siberia to Alaska

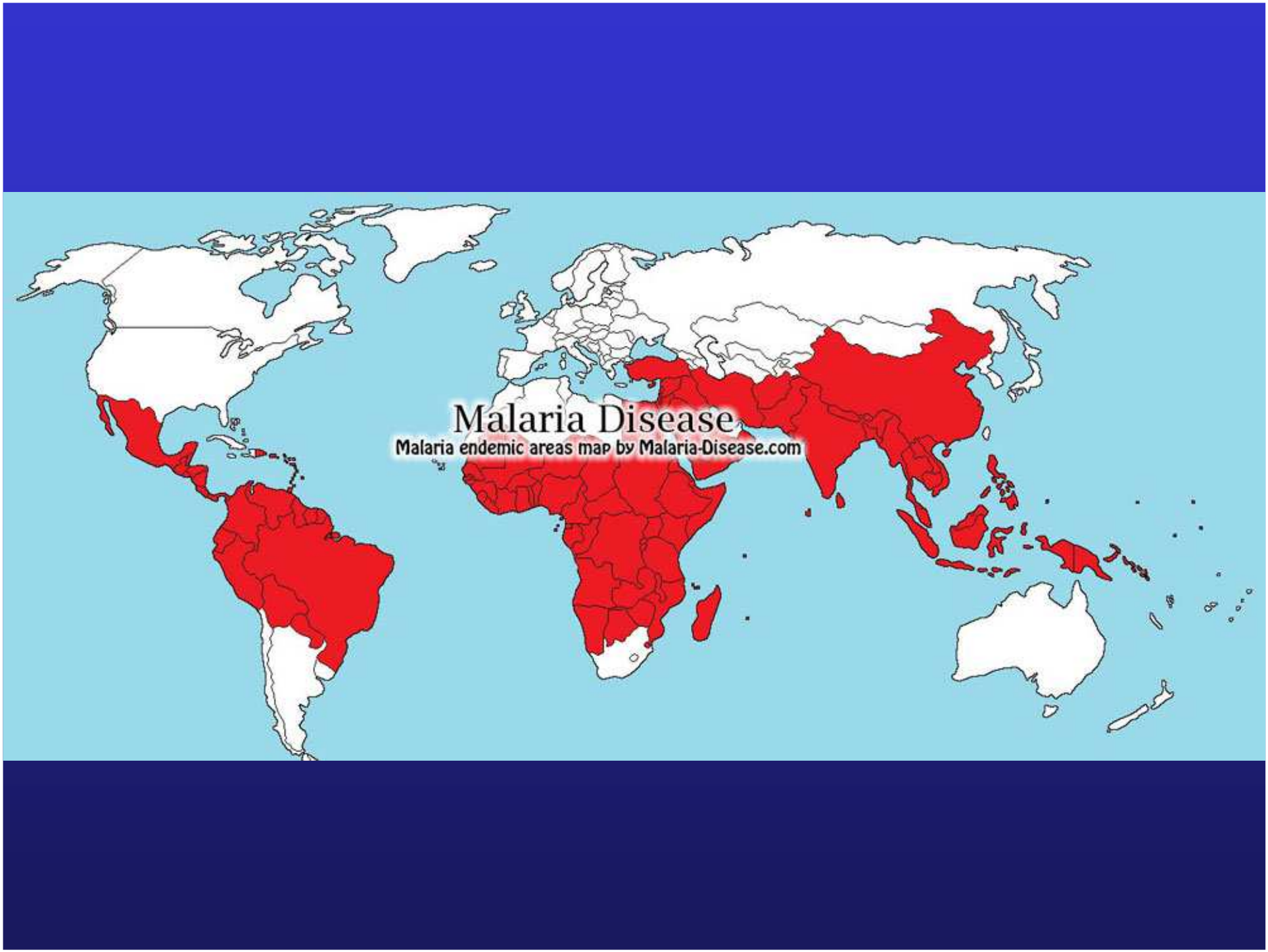


Genetic diversity of Modern Man





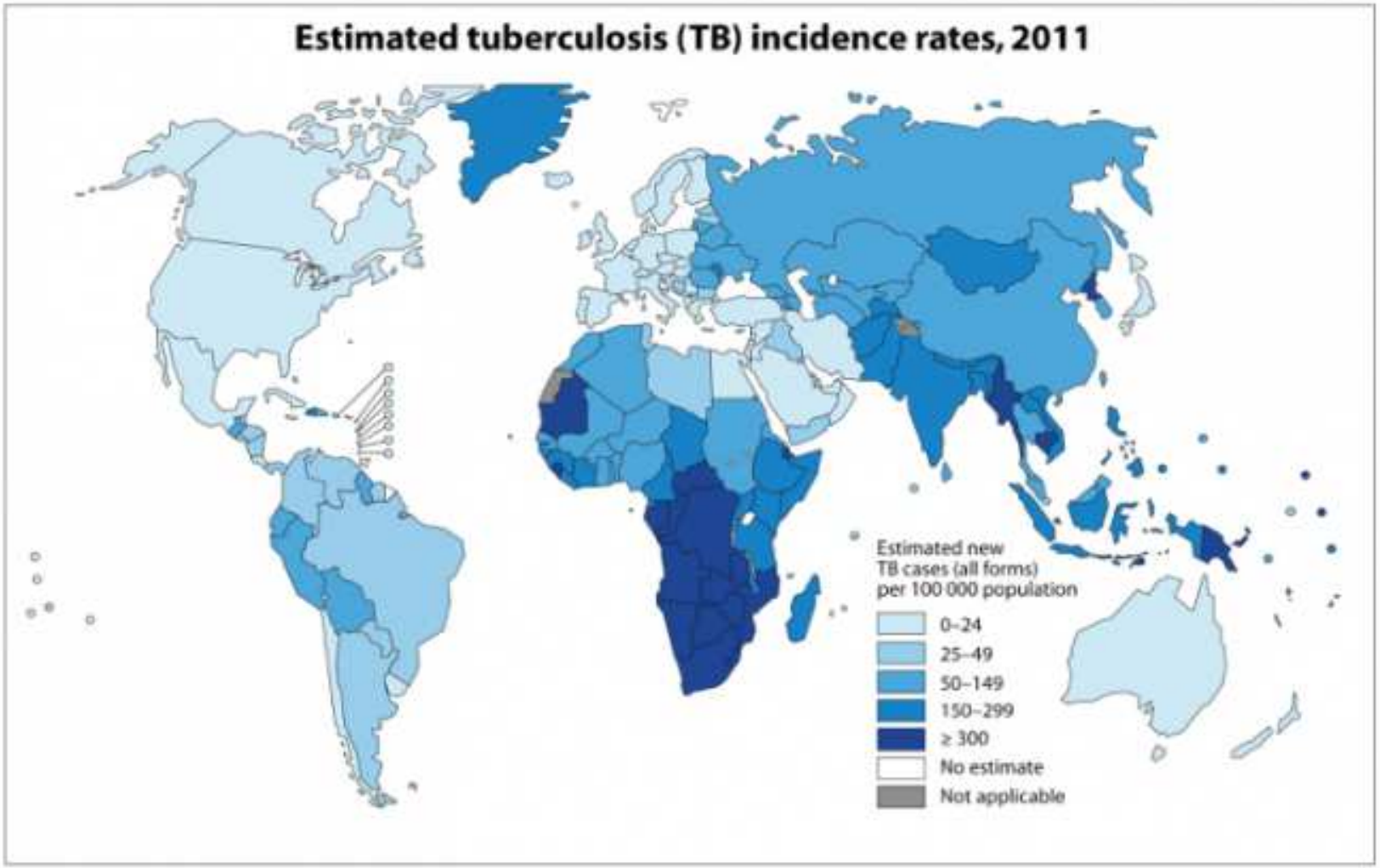
Key events in recent human evolution (boxes outlined in black) are juxtaposed with **the estimated ages of infectious disease emergence** (boxes outlined in red). The fragmentation of the human lineage into genetically and geographically distinct populations (blue lines) accelerates with migration out of Africa. Later, these populations started mixing more (blue shaded regions between the populations) along trade routes (such as the Silk Road), through colonization and through high rates of global travel nowadays.


Malaria Disease
Malaria endemic areas map by Malaria-Disease.com

Common Erythrocyte Variants That Affect Resistance to **Malaria**

Gene	Protein	Function	Reported Genetic Associations with Malaria
FY	Duffy antigen	Chemokine receptor	FY*O allele completely protects against *P. vivax* infection.
G6PD	Glucose-6-phosphatase dehydogenase	Enzyme that protects against oxidative stress	G6PD deficiency protects against severe malaria.
GYPA	Glycophorin A	Sialoglycoprotein	GYPA-deficient erythrocytes are resistant to invasion by *P. falciparum.*
GYPB	Glycophorin B	Sialoglycoprotein	GYPB-deficient erythrocytes are resistant to invasion by *P. falciparum.*
GYPC	Glycophorin C	Sialoglycoprotein	GYPC-deficient erythrocytes are resistant to invasion by *P. falciparum.*
HBA	α-Globin	Component of hemoglobin	α^+ Thalassemia protects against severe malaria but appears to enhance mild malaria episodes in some environments.
HBB	β-Globin	Component of hemoglobin	HbS and HbC alleles protect against severe malaria. HbE allele reduces parasite invasion.
HP	Haptoglobin	Hemoglobin-binding protein present in plasma (not erythrocyte)	Haptoglobin 1-1 genotype is associated with susceptibility to severe malaria in Sudan and Ghana.
SCL4A1	CD233, erythrocyte band 3 protein	Chloride/bicarbonate exchanger	Deletion causes ovalocytosis but protects against cerebral malaria.

Estimated tuberculosis (TB) incidence rates, 2011



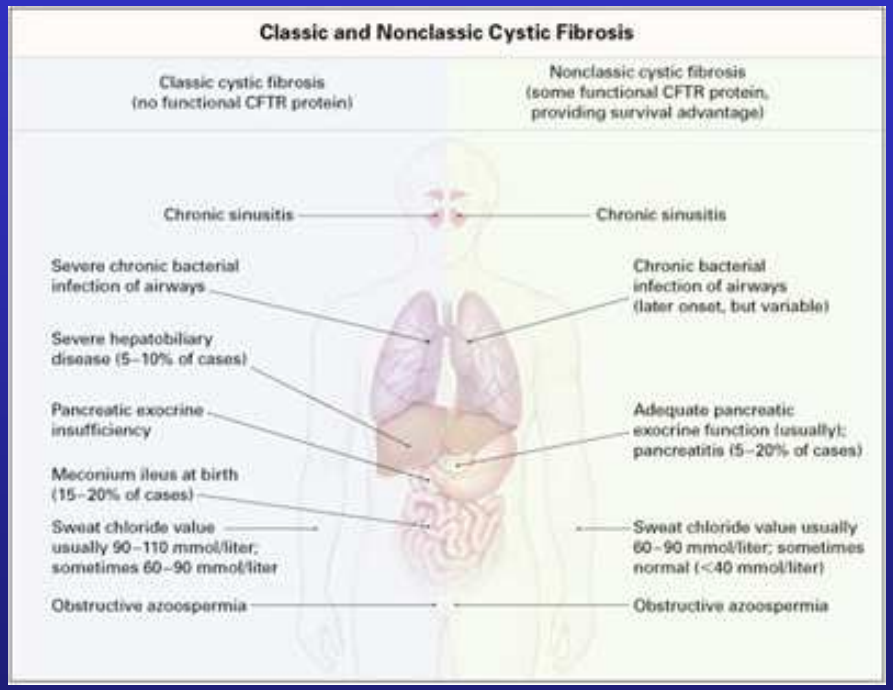
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

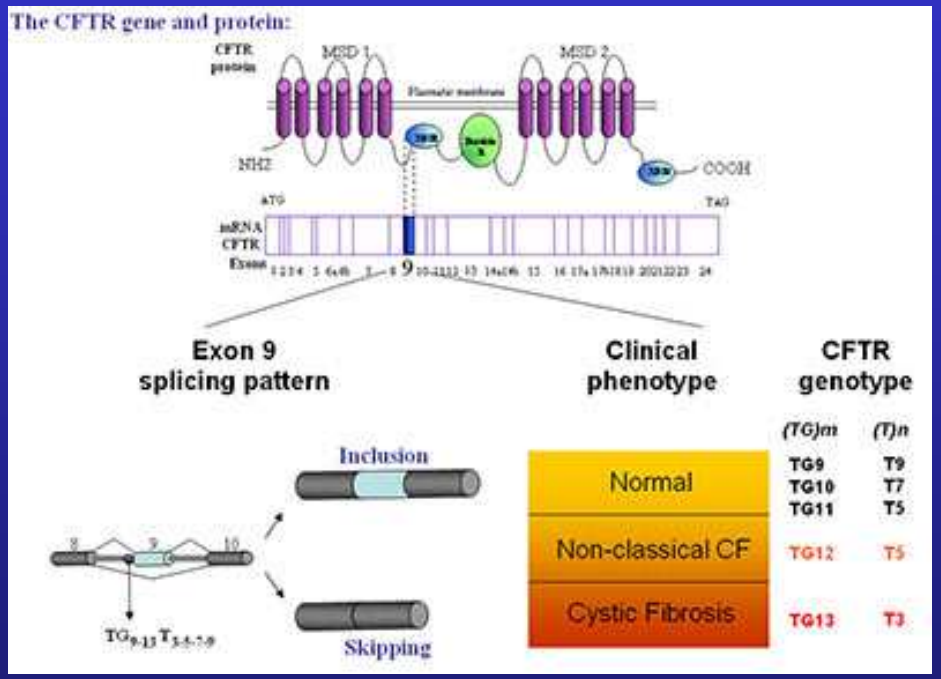
Source: *Global Tuberculosis Report 2012*. WHO, 2012.

World Health Organization

Cystic fibrosis

1 in 3,000 children are born with CF, and 2% of people carry one mutant gene

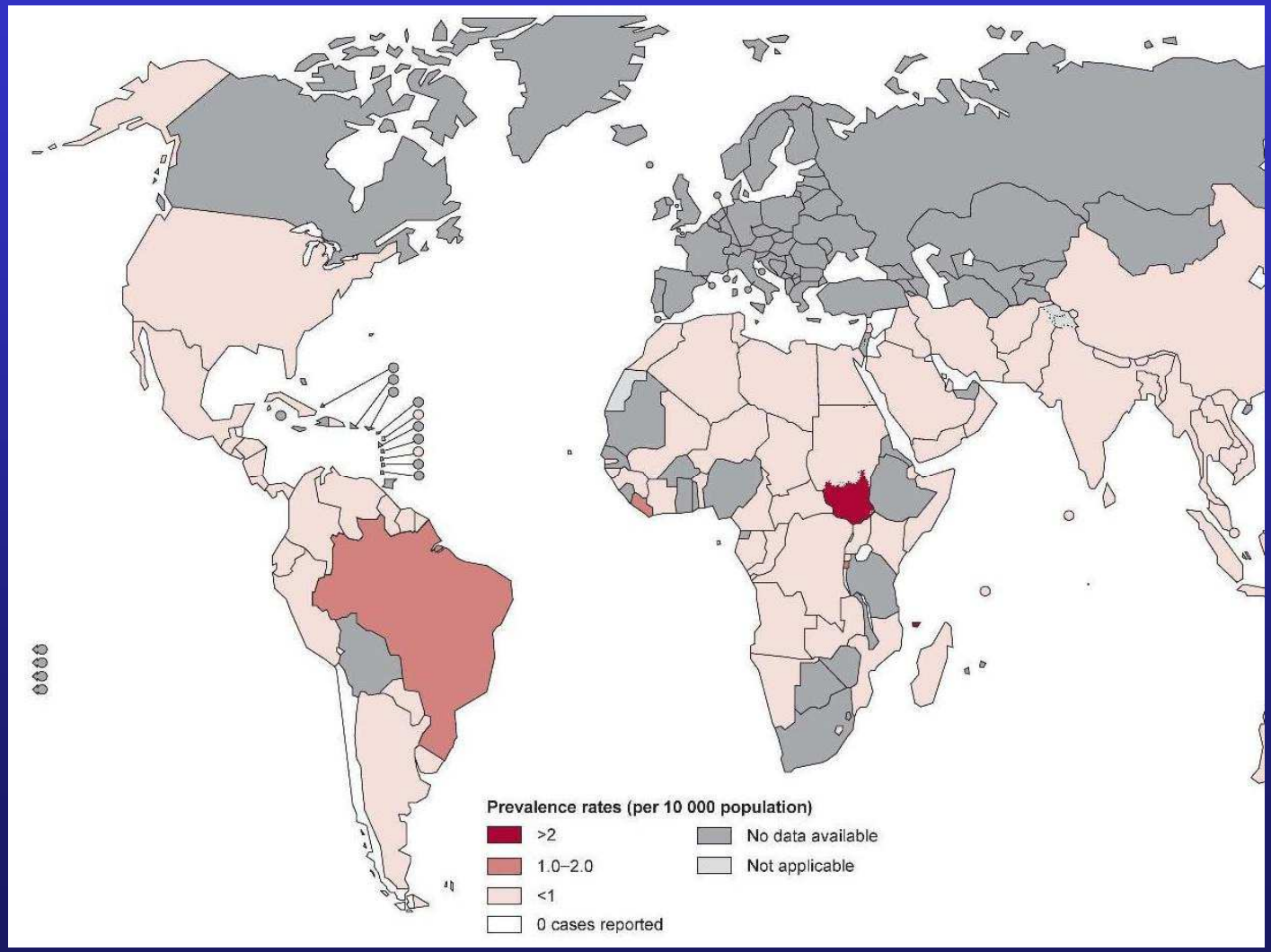




Cystic fibrosis gene protects against tuberculosis

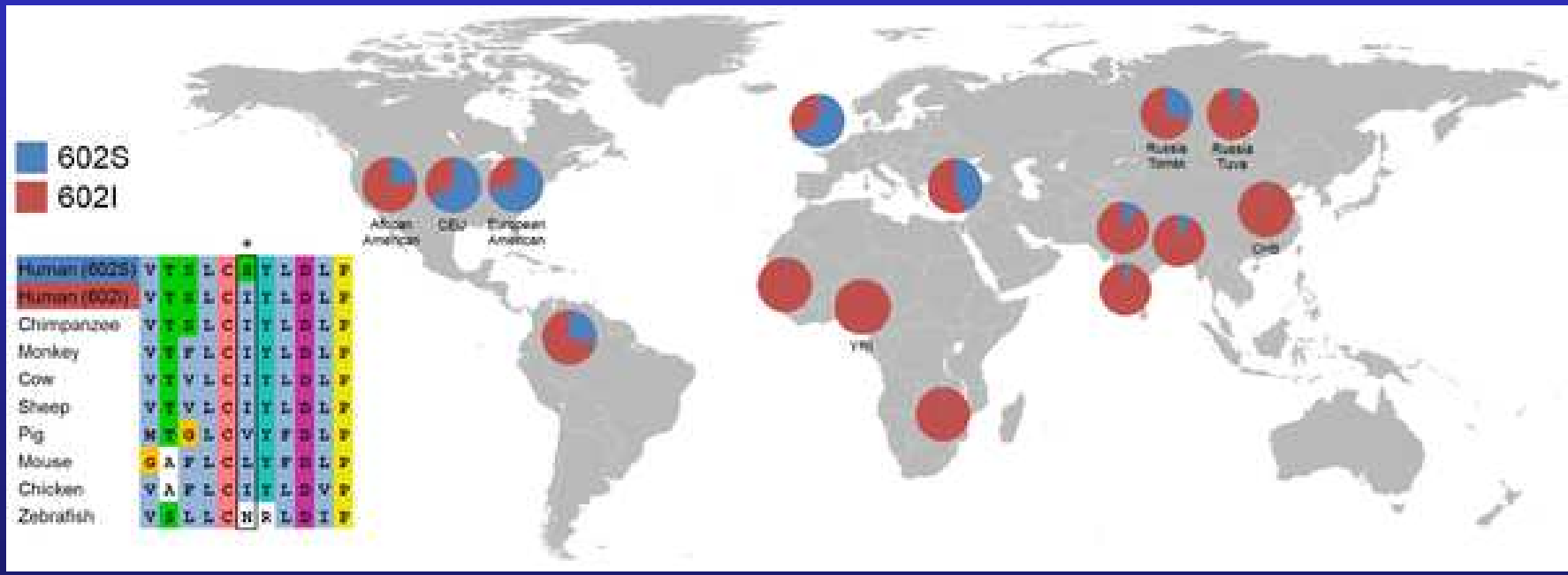
Between 1600 and 1900, TB caused 20% of all deaths in Europe

There were nearly 200,000 cases of leprosy around the world at the beginning of 2012.



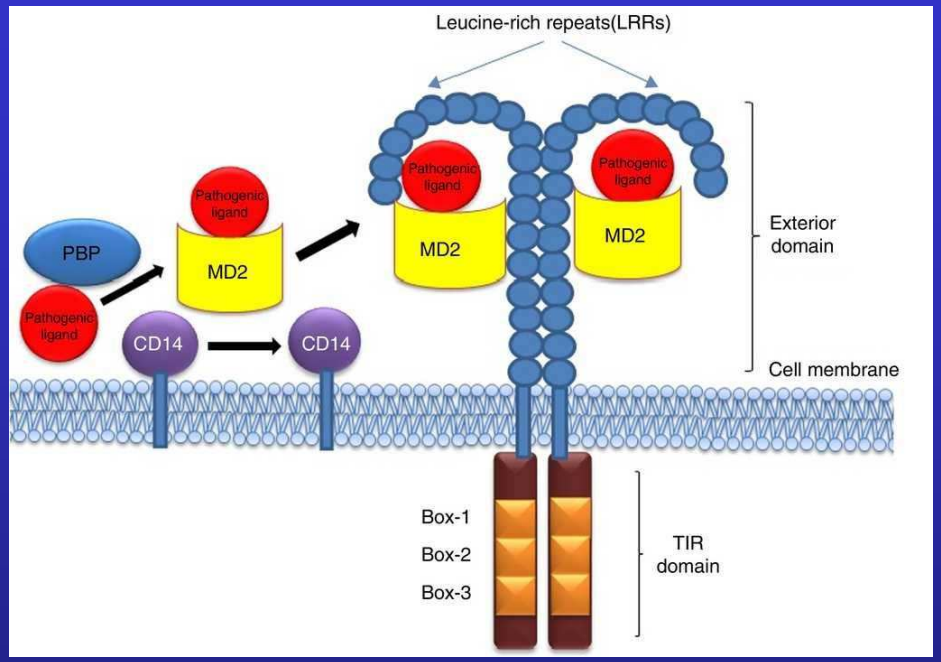
Leprosy and the Adaptation of Human Toll-Like Receptor 1. Population differentiation at TLR1 I602S.

The **protective dysfunctional 602S** allele is rare in Africa but expands to become the dominant allele among individuals of **European descent**. This supports the hypothesis that this locus may be under selection from mycobacteria or other pathogens that are recognized by TLR1 and its co-receptors.



Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, et al. (2010) Leprosy and the Adaptation of Human Toll-Like Receptor 1. PLoS Pathog 6(7): e1000979. doi:10.1371/journal.ppat.1000979
http://127.0.0.1:8081/plospathogens/article?id=info:doi/10.1371/journal.ppat.1000979

PLOS | PATHOGENS

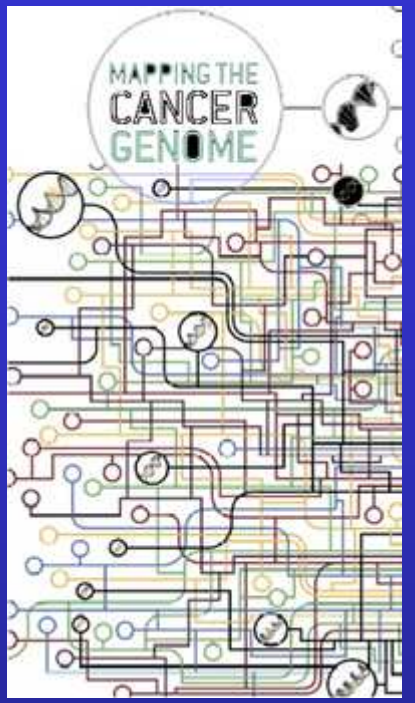


TLR2 is one of the toll-like receptors and plays a role in the immune system. TLR2 is a membrane protein, a receptor, which is expressed on the surface of certain cells and recognizes foreign substances and passes on appropriate signals to the cells of the immune system.

The presence of the TLR2 Arg753Gln polymorphism was significantly associated with pneumonia in AML patients.

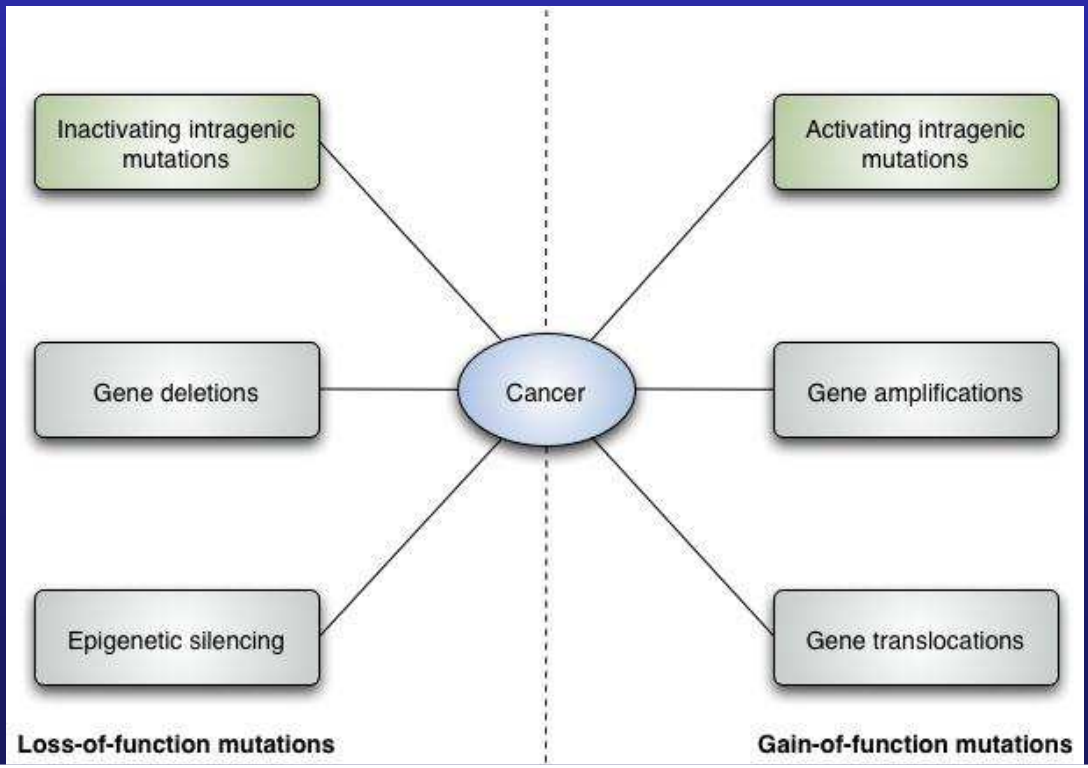
The presence of the TLR2 Arg753Gln polymorphism was significantly associated with syphilis.

The Cancer Genome Atlas (**TCGA**)

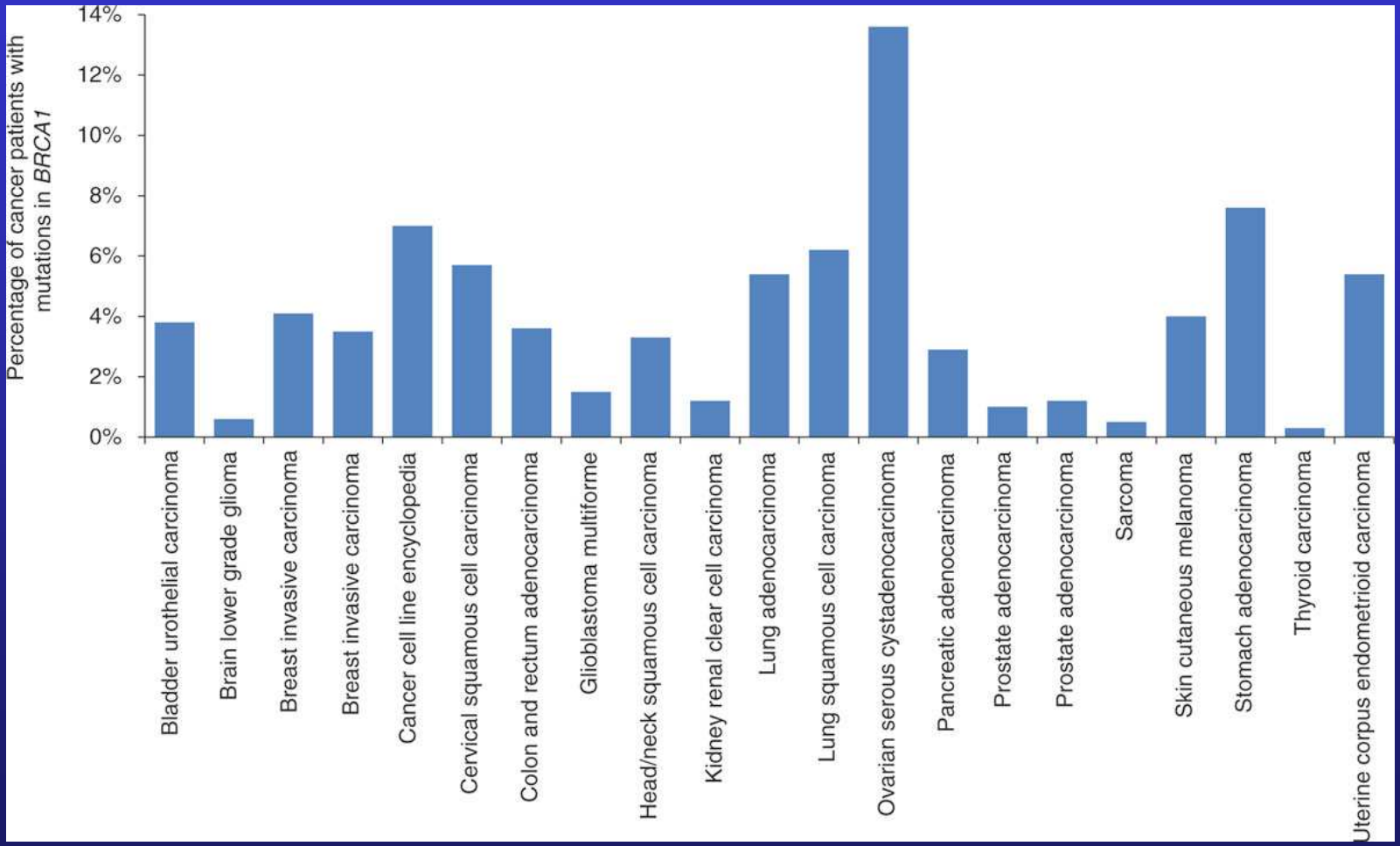


- is a comprehensive effort to accelerate our understanding of the molecular basis of cancer through the application large-scale genome sequencing.

- Many tumor types and many tumor samples



The Cancer Genome Atlas (TCGA)

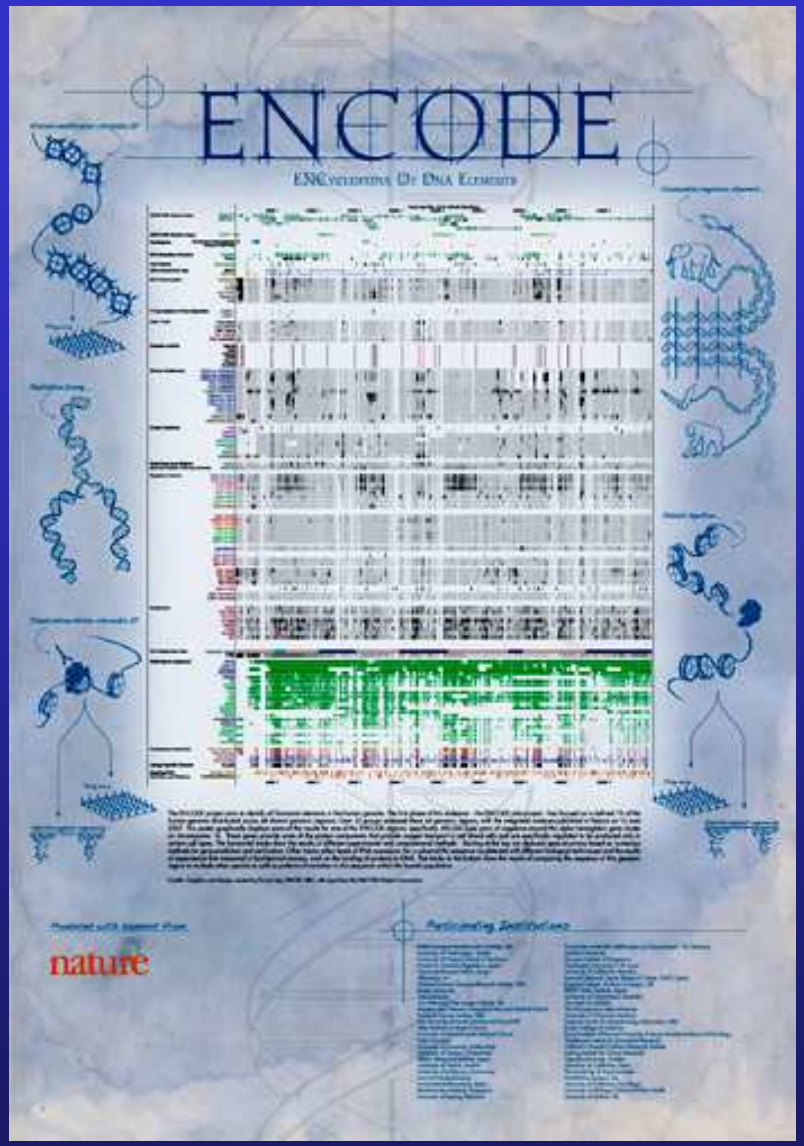


BRCA1 is a human tumor suppressor gene

ENCODE
(*Encyclopedia of DNA Elements*)

- project to identify all functional elements in the human genome sequence

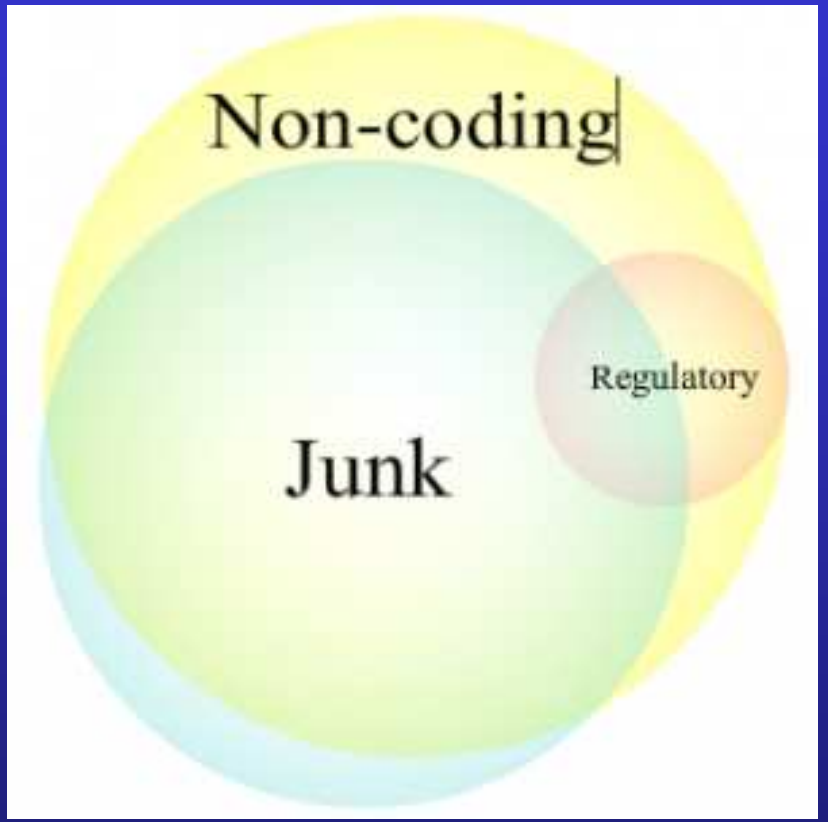
- the genome is pervasively transcribed, such that the majority of its bases can be found in primary transcripts, including non-protein-coding transcripts



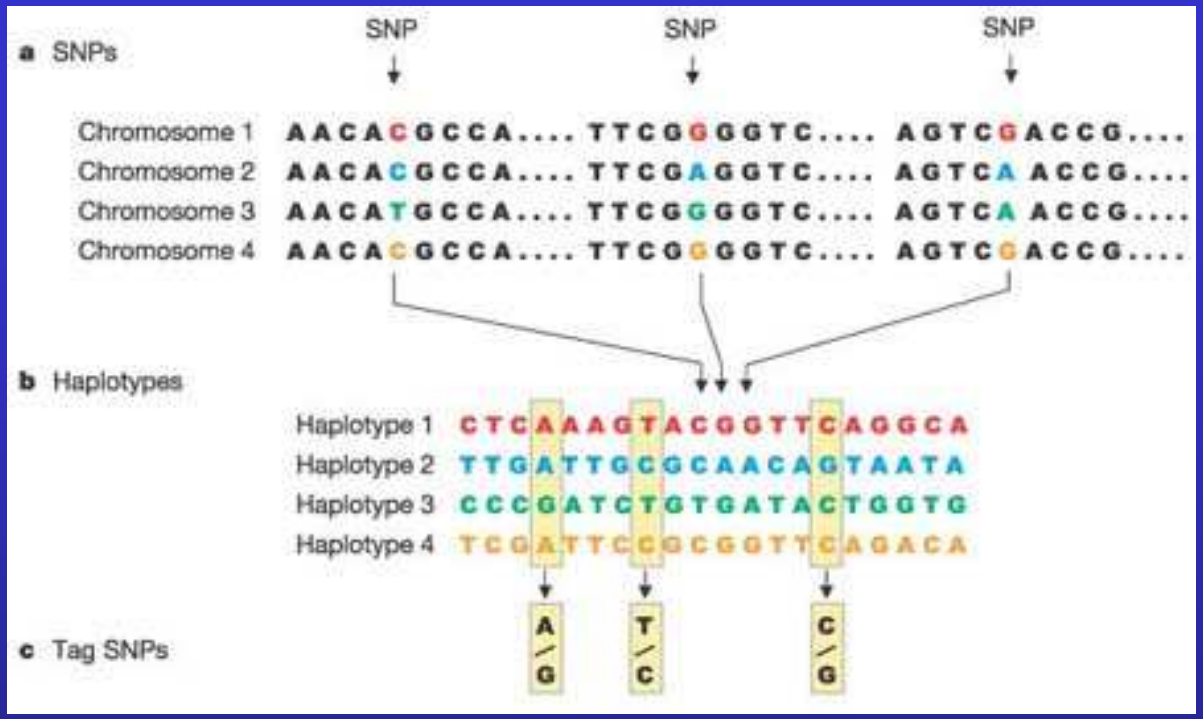
ENCODE
(*Encyclopedia of DNA Elements*)

For the last decade, geneticists have run a seemingly endless stream of "genome-wide association studies" (GWAS), attempting to understand the genetic basis of disease. They have thrown up a long list of SNPs – variants at specific DNA letters—that correlate with the risk of different conditions.

The ENCODE team have mapped *all* of these to their data. They found that just 12 percent of the SNPs lie within protein-coding areas. They also showed that compared to random SNPs, the disease-associated ones are 60 percent more likely to lie within functional, non-coding regions, especially in promoters and enhancers. This suggests that many of these variants are controlling the activity of different genes, and provides many fresh leads for understanding how they affect our risk of disease.

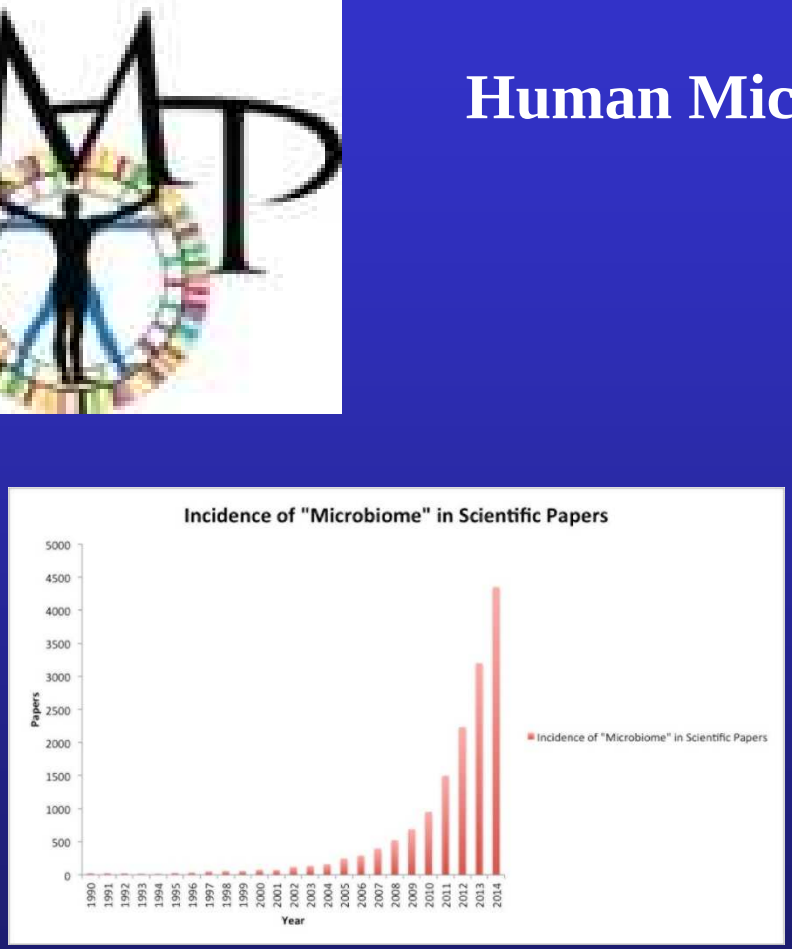


HapMap Project



- The International HapMap Project is a multi-country effort to identify and catalog genetic similarities and differences in human beings. Using the information in the HapMap, researchers will be able to find genes that affect health, disease, and individual responses to medications and environmental factors.
- DNA samples from 269 people from Africa, Japan, China and USA
- 10 millions nucleotide sites where people most frequently differ – SNPs (99,9%)
- SNPs are present in haplotypes. Haplotype is a set of single-nucleotide polymorphisms (SNPs) on a single chromatid (half a chromosome pair) that are statistically associated (300.000 – 600.000 haplotypes).

Human Microbiome Project

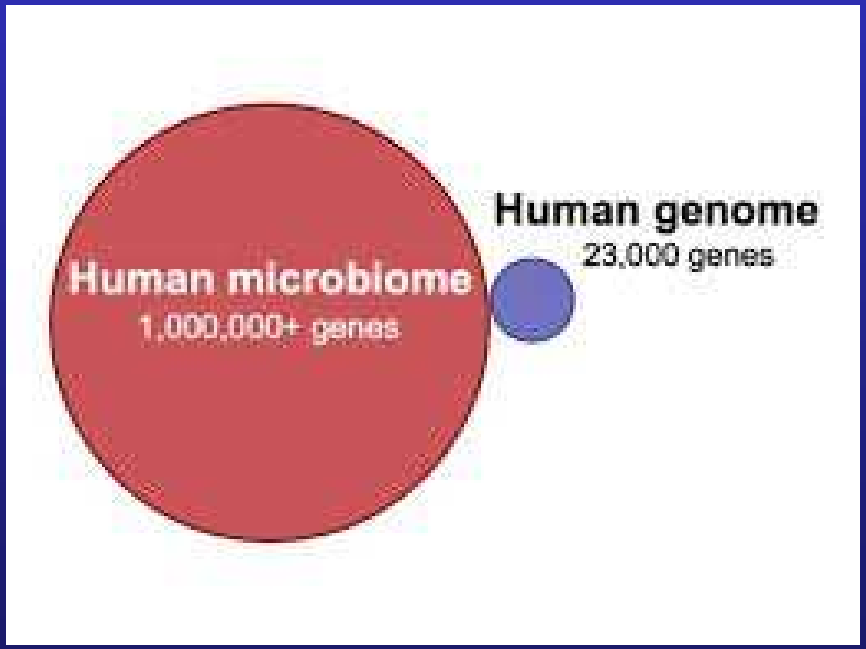


Within the body of a healthy adult, microbial cells are estimated to outnumber human cells by a factor of ten. These communities, however, remain largely unstudied, leaving almost entirely unknown their influence upon human development, physiology, immunity, and nutrition.

The Human Microbiome: Our Second Genome

21 000 human genes, 1 000 000 bacterial genes

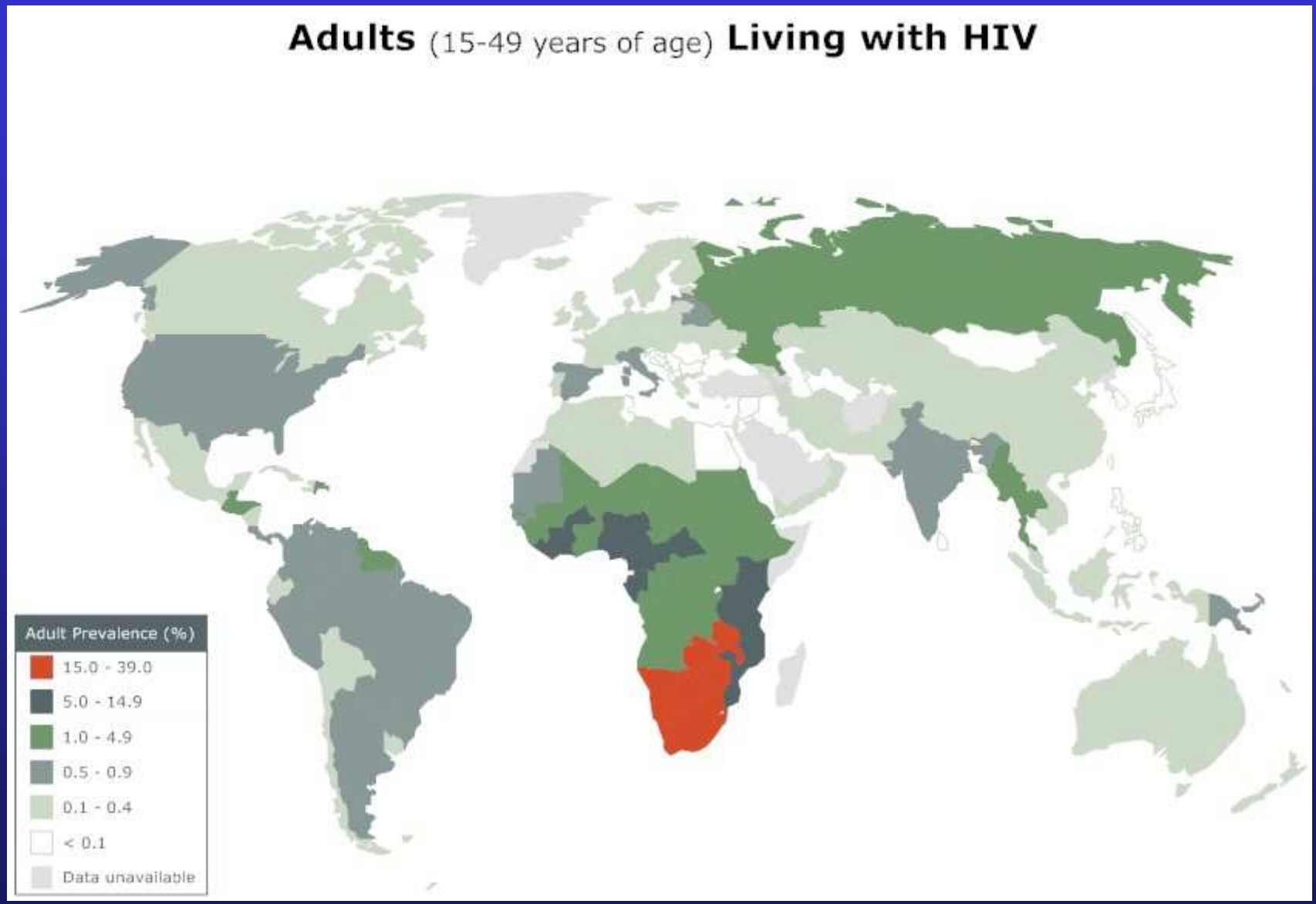
The human microbiome is a source of genetic **diversity**, a **modifier** of disease, an essential component of **immunity**, and a functional entity that influences **metabolism** and modulates drug interactions.



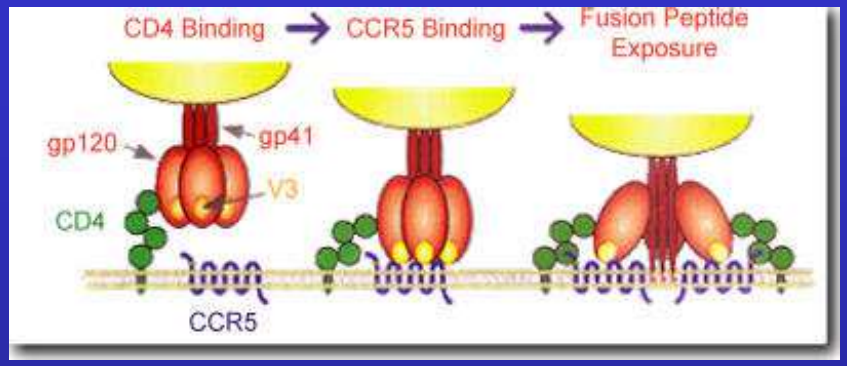
Pilot HM projects

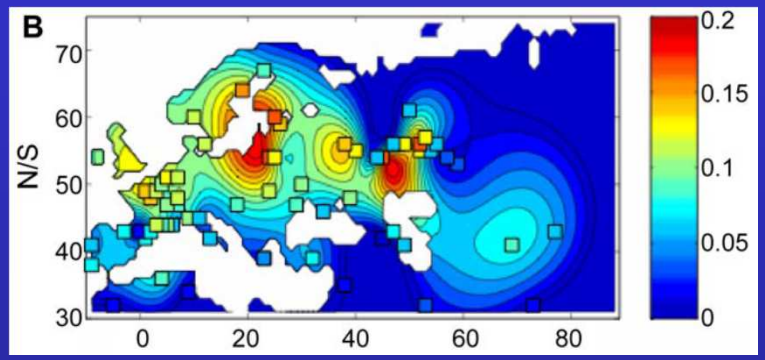
	Disease/symptoms
skin	psoriasis
	acne vulgaris
	atopic dermatitis
GIT	obesity
	morbus Crohn
	esophageal cancer
	necrotizing Enterocolitis
	colitis ulcerosa
	irritable bowel syndrome
Urogenital tract	bacterial vaginosis
	STD
Systemic	imunodeficiences
	febriles

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


Adults (15-49 years of age) Living with HIV
Adult Prevalence (%)
15.0 - 39.0
5.0 - 14.9
1.0 - 4.9
0.5 - 0.9
0.1 - 0.4
< 0.1
Data unavailable

HIV epidemic





The CCR5 locus shows that historical epidemics have been important in shaping the genomes of humans and other primatespecies. It has been projected that if the HIV epidemic continues for another 100 years, it will leave a signature on the human genome at the **CCR5 locus** and related HIV resistance loci.

Although higher HLA-C expression protects against HIV progression, it also increases risk of the inflammatory disorder Crohn's disease, which highlights the potential for health repercussions of pathogen-driven selection.

The **human leukocyte antigen (HLA)** system is the locus of genes that encode for proteins on the surface of cells that are responsible for regulation of the immune system in humans.

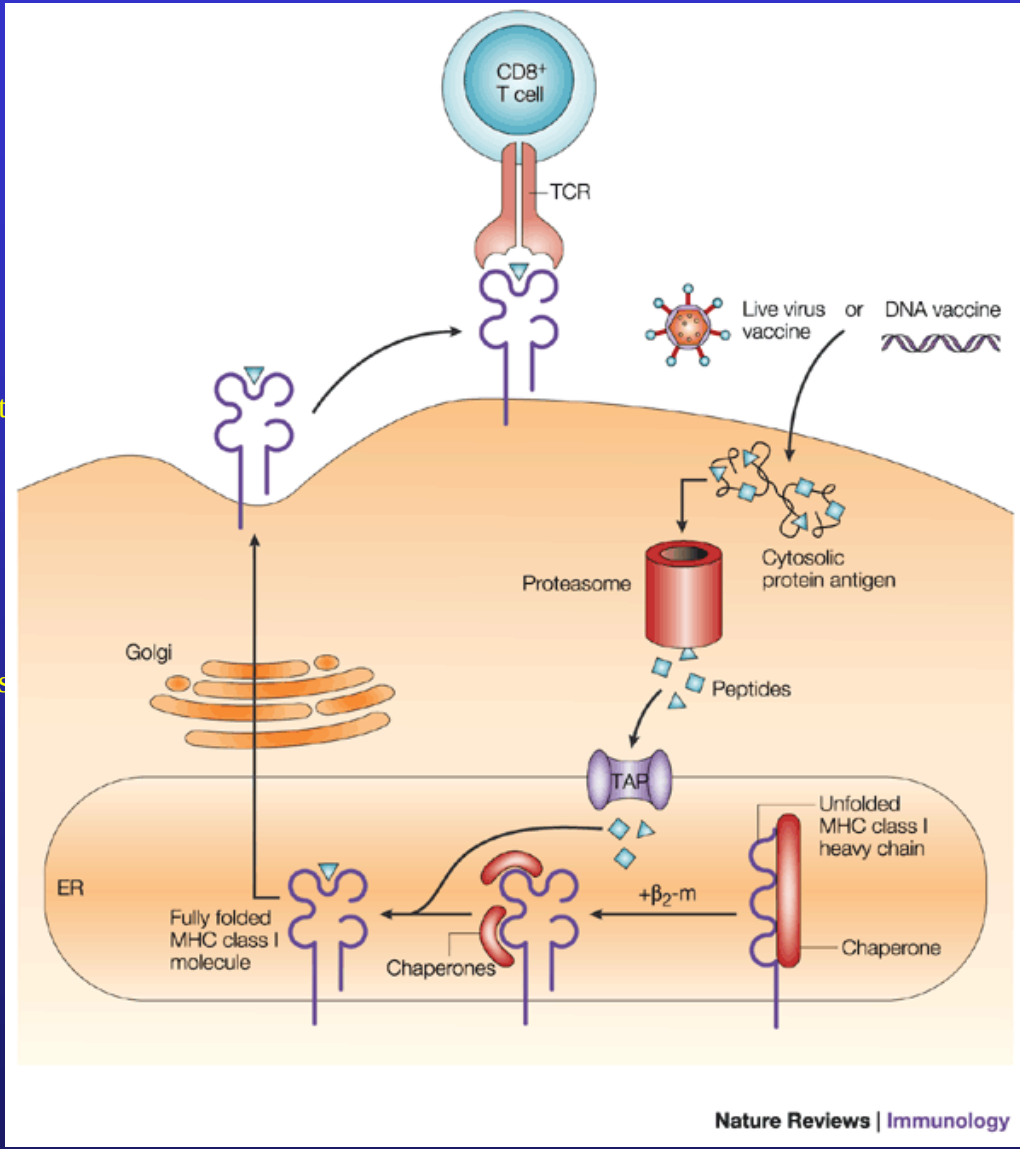
The HLA genes are the human versions of the major histocompatibility complex (MHC) genes that are found in most vertebrates

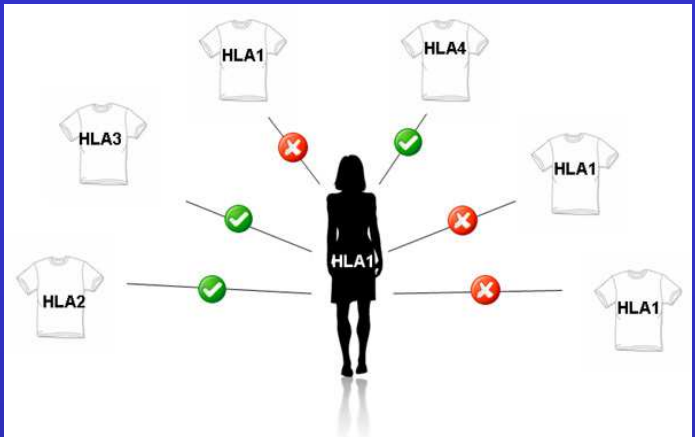
HLAs corresponding to MHC class I (**A**, **B**, and **C**) present peptides from inside the cell.

HLAs corresponding to MHC class II (**DP**, **DM**, **DOA**, **DOB**, **DQ**, and **DR**) present antigens from outside of the cell to T-lymphocytes.

HLAs corresponding to MHC class III encode components of the complement system.

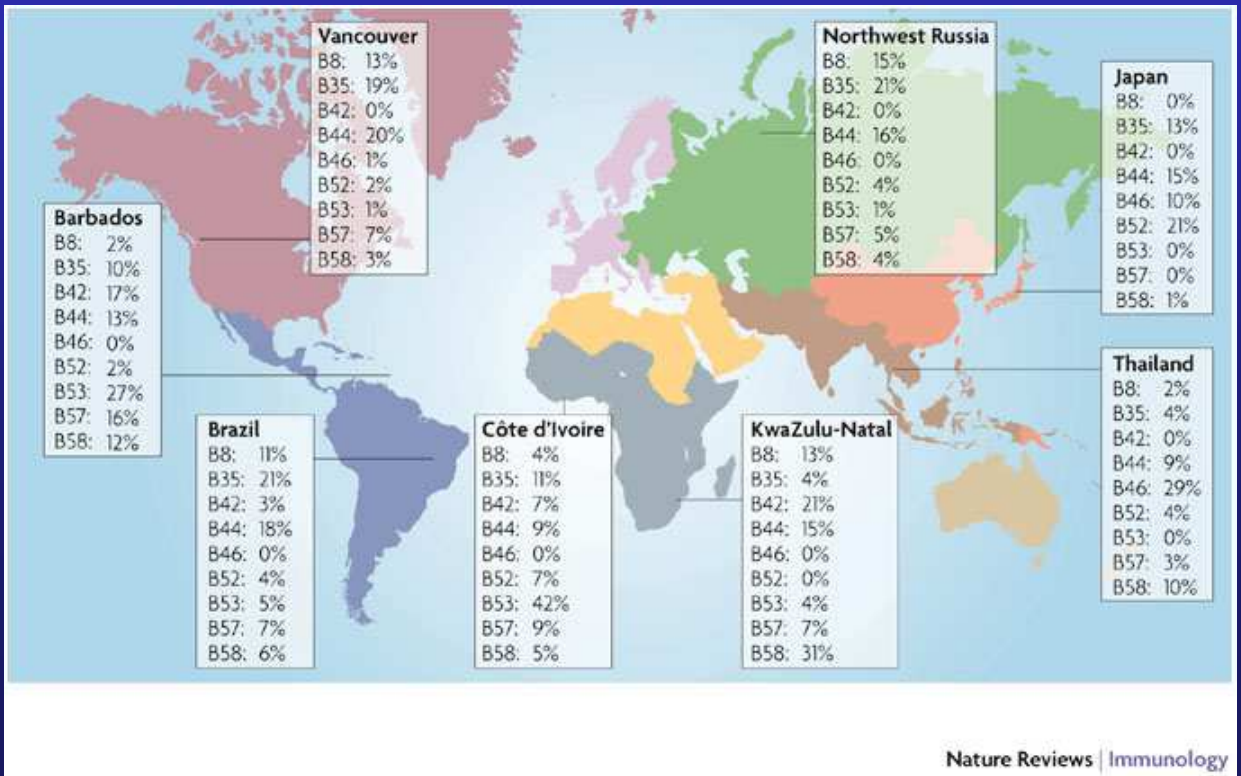
HLAs are important in disease defense. They are the major cause of organ transplant rejections. They may protect against or fail to protect (if down-regulated by an infection) against cancers.[1] Mutations in HLA may be linked to autoimmune disease (examples: type I diabetes, coeliac disease). HLA may also be related to people's perception of the odor of other people, and may be involved in mate selection, as at least one study found a lower-than-expected rate of HLA similarity between spouses in an isolated community.[2]





The MHC represents the most polymorphic gene cluster in humans, and more than **2,700** alleles have been described for the most variable gene, HLAB.

Increased diversity at *HLA* class I genes (compared to the genome average) is observed in populations living in geographic regions where pathogen diversity is also high.



Diverse distribution of HLA-B alleles worldwide

Inflammatory bowel disease

Caused by autoimmune attacks on the gastrointestinal system.

163 distinct loci have been significantly associated with IBD.

Seven of the eight **leprosy** susceptibility loci are also associated with increased IBD risk.

Coeliac disease

Coeliac disease is a strongly heritable (~80%) inflammatory intestinal disorder triggered by gluten consumption.

Coeliac disease occurs at 1–2% in Europe and up to 6% in North African Sahrawi.

Individuals who are homozygous for the coeliac risk allele (~22% of the European population) have stronger activation of the NOD2 pathway and a **3–5-fold higher pro-inflammatory cytokine response** to lipopolysaccharide.

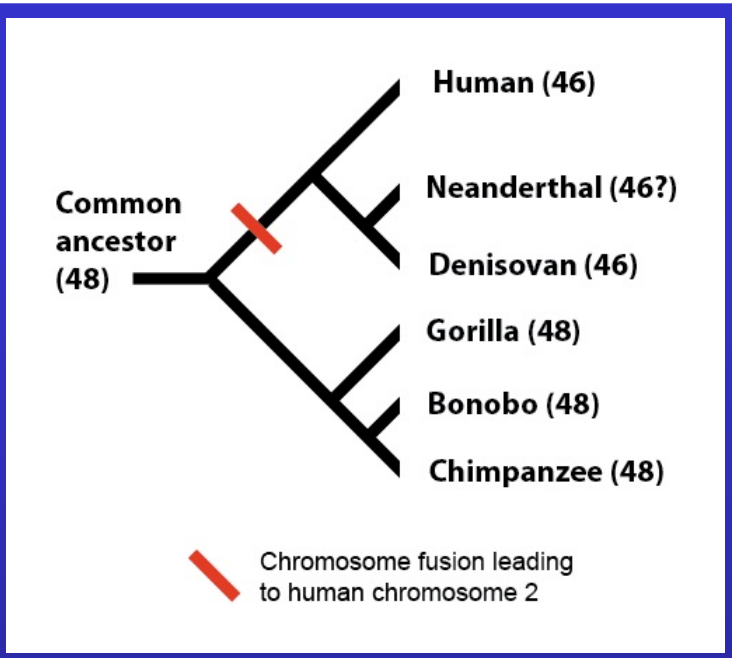
Better protection against bacterial infection may have conferred a selective advantage that outweighed the increased risk of coeliac disease risk.

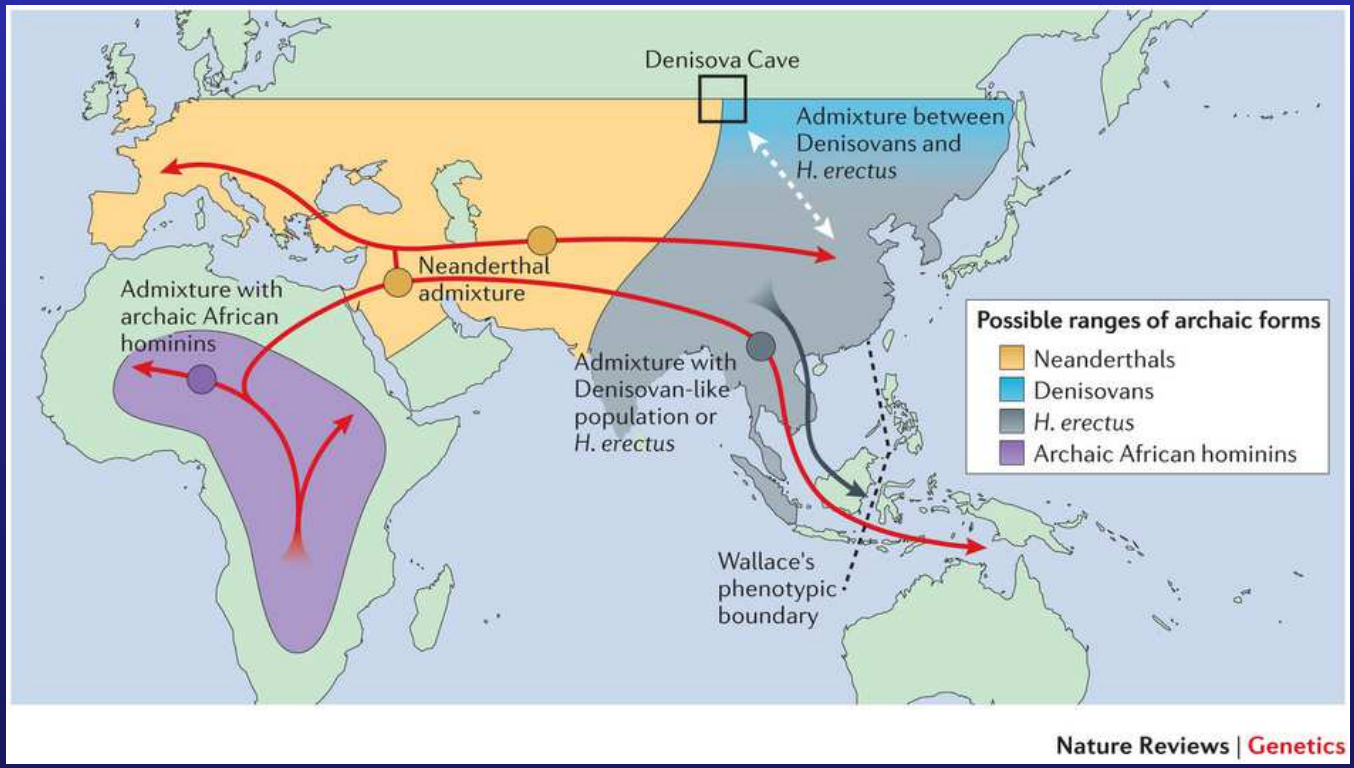
Non-autoimmune disease: kidney disease

African Americans suffer from kidney disease — including focal segmental glomerulosclerosis (FSGS) and hypertension-attributed end-stage kidney disease (H-ESKD) — at higher rates than European Americans.

Two independent coding variants in *APOL1* that are strongly associated with FSGS (odds ratio = 10.5) and H-ESKD (odds ratio = 7.3).

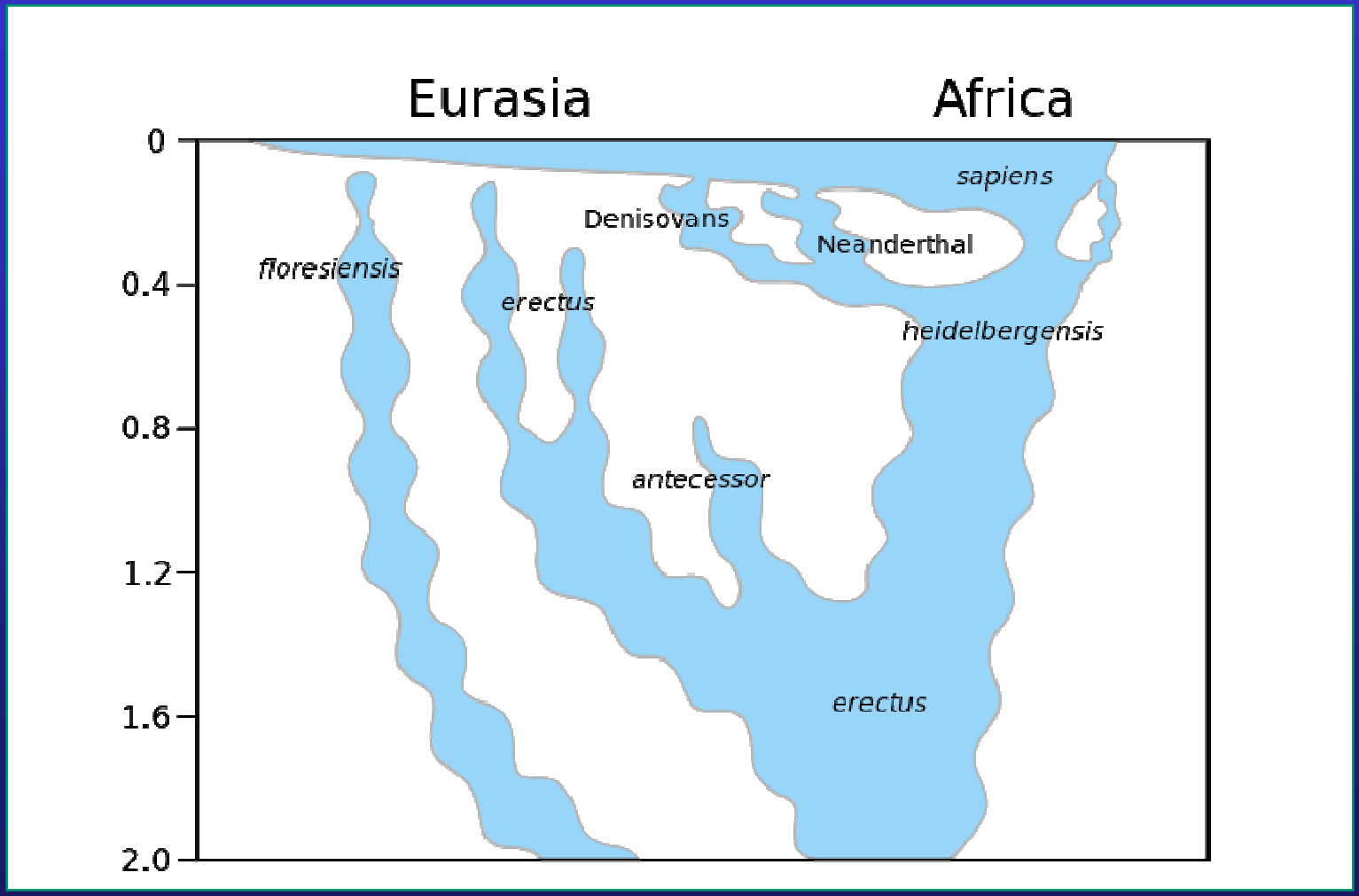
In vitro assays showed that the kidney disease-associated variants lyse ***T. brucei rhodesiense***, which is the trypanosome parasite that causes the most acute, virulent form of sleeping sickness.

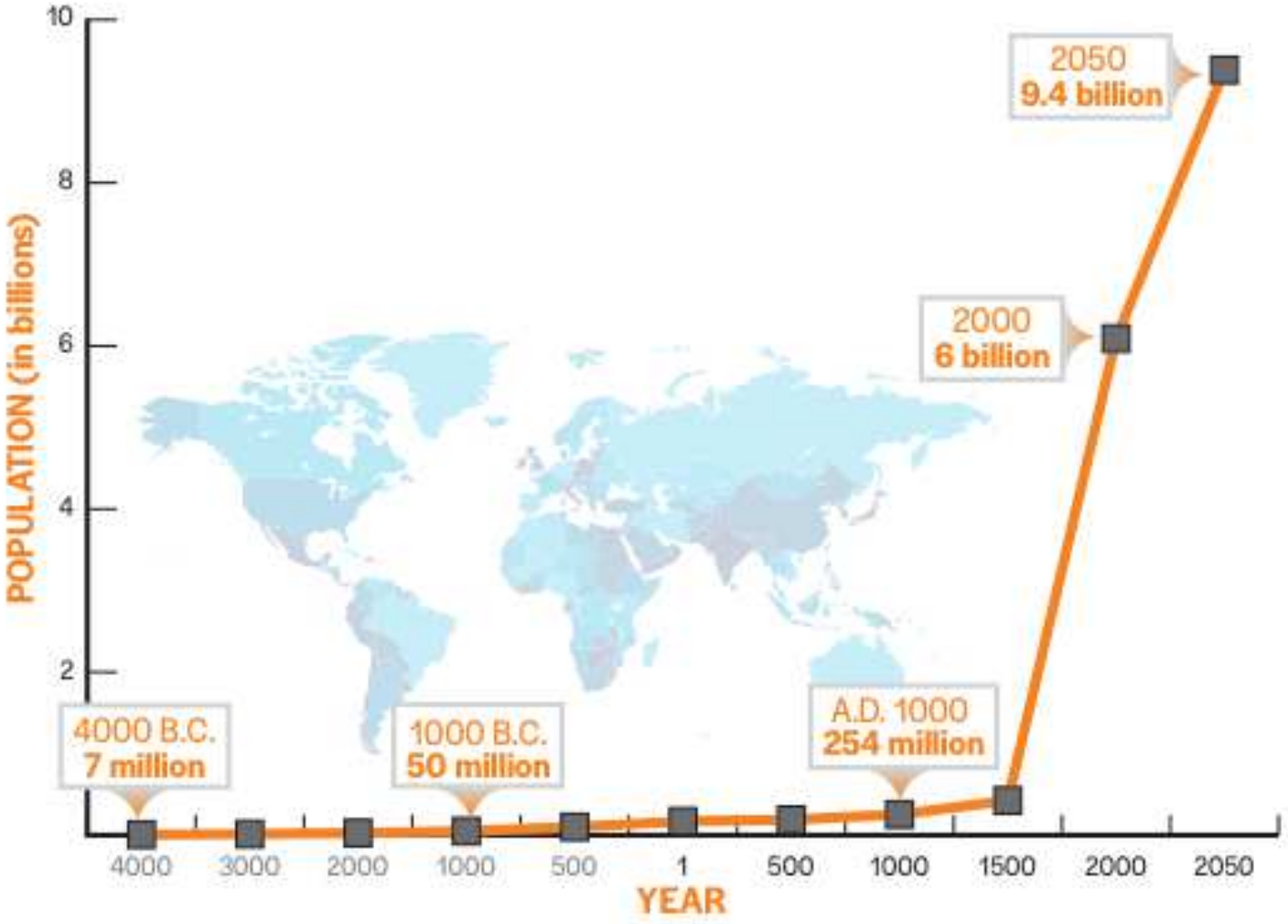




The **Wallace Line** or **Wallace's Line** is a faunal boundary line drawn in 1859 by the British naturalist Alfred Russel Wallace that separates the ecozones of Asia and Wallacea, a transitional zone between Asia and Australia.

Denisovans shared a common origin with Neanderthals, that they ranged from Siberia to Southeast Asia, and that they lived among and interbred with the ancestors of some modern humans, with about 3% to 5% of the DNA of Melanesians and Aboriginal Australians deriving from Denisovans




POPULATION (in billions)
10
8
6
4
2
2050
9.4 billion
2000
6 billion
A.D. 1000
254 million
1000 B.C.
50 million
4000 B.C.
7 million
4000
3000
2000
1000
500
1
500
1000
1500
2000
2050
YEAR