

SAMUEL DE CHAMPLAIN



Champlain's map of
Quebec, ca. 1610

GÉNOME QUÉBEC



Projets par secteur
Santé
Environnement
Foresterie
Enjeux éthiques
Agriculture
Développement de nouvelles technologies



THE PERSONAL GENOME

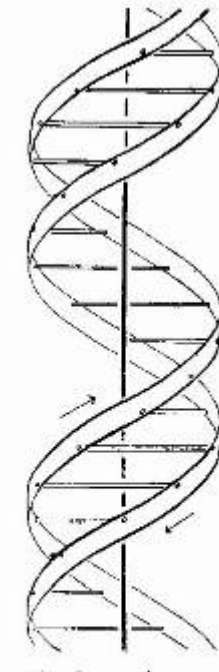
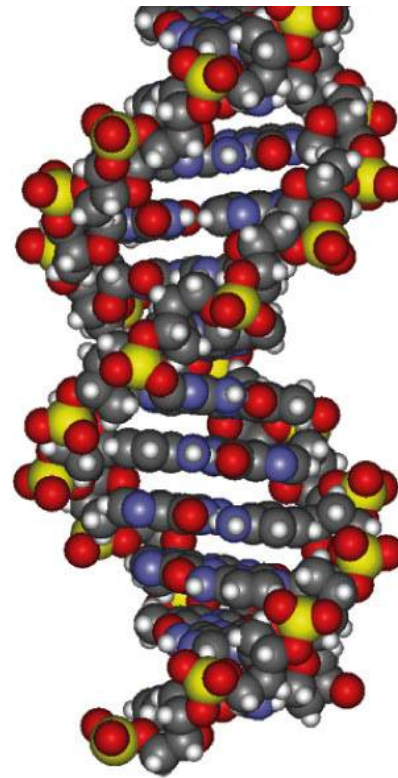
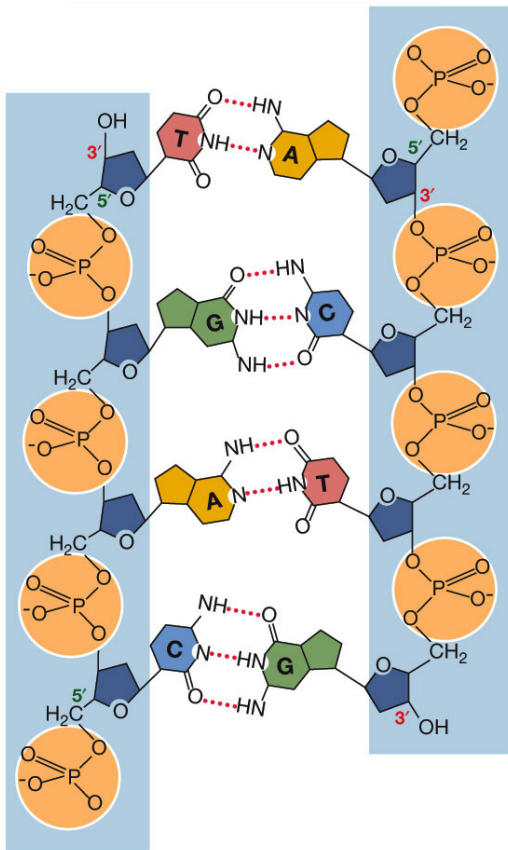
A. *Genes and genomes*

B. The human genome project

C. The human genome

D. The human genome and individuality

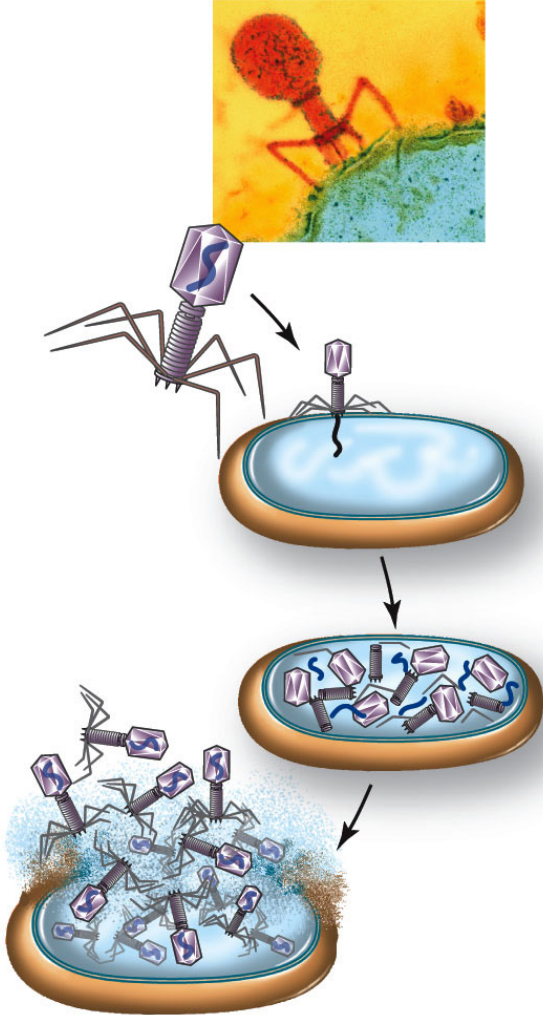
DNA: DEOXYRIBONUCLEIC ACID



DNA AS A SECULAR ICON



DNA AS THE GENETIC MATERIAL



GENES DETERMINE PHENOTYPE



Genes: DNA in two cell nuclei fusing in the fertilized egg

Determine

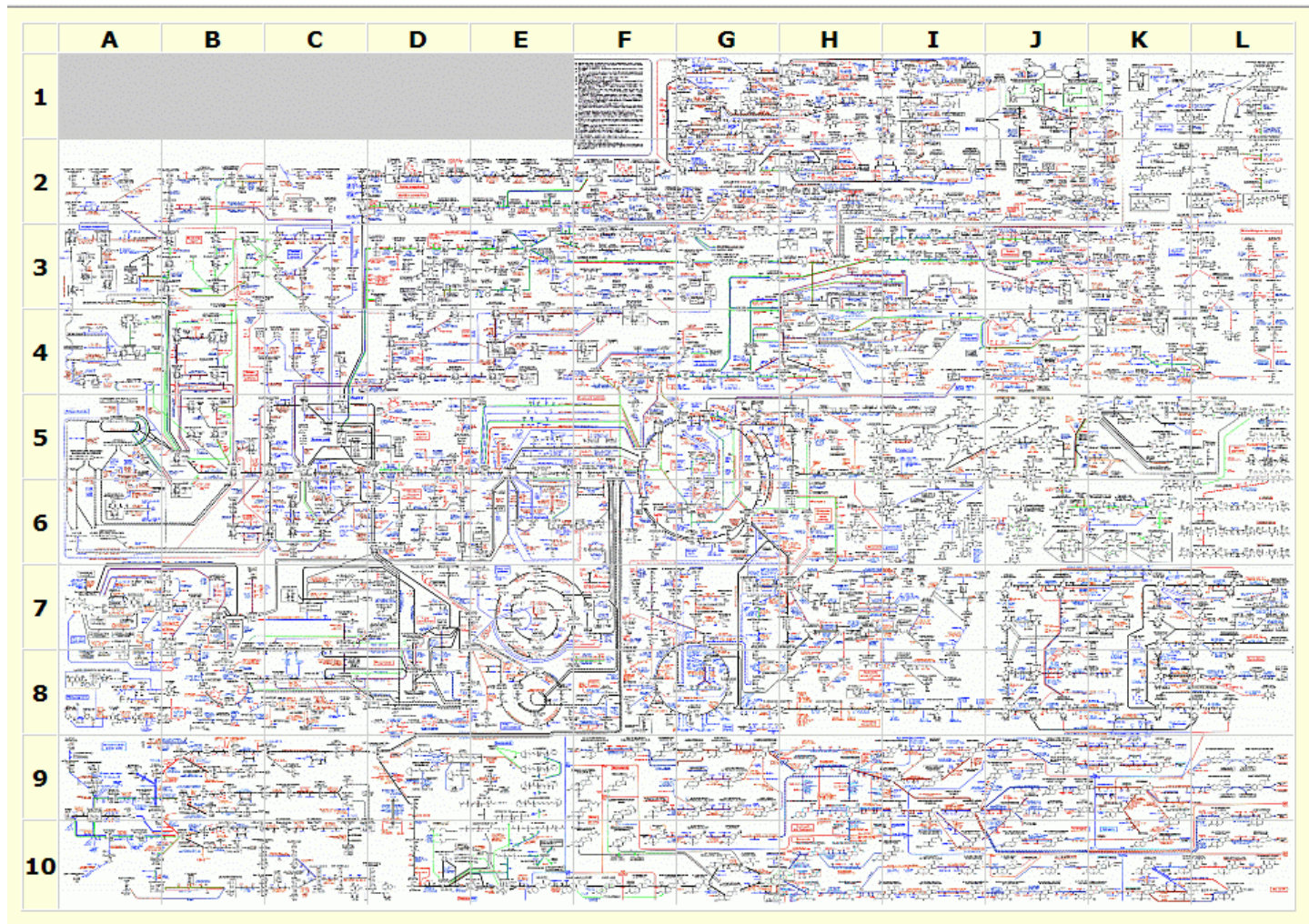


Phenotype: The characteristics of an individual

GENES ARE NOT DESTINY

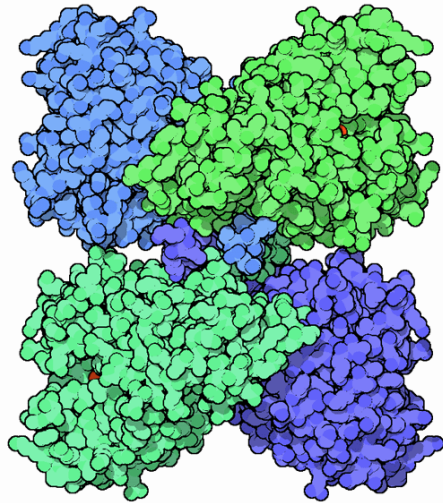


PHENOTYPE IS DETERMINED BY PROTEINS

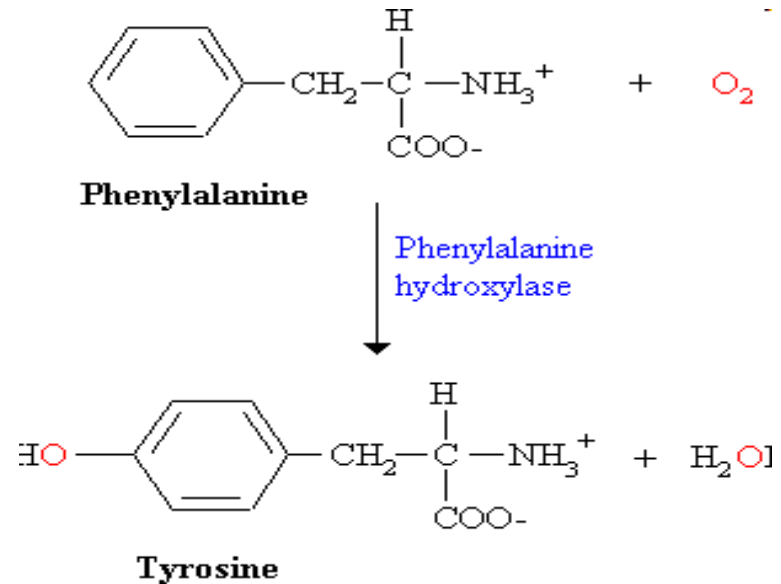


A unique protein is needed for every chemical transformation

ONE CHEMICAL TRANSFORMATION

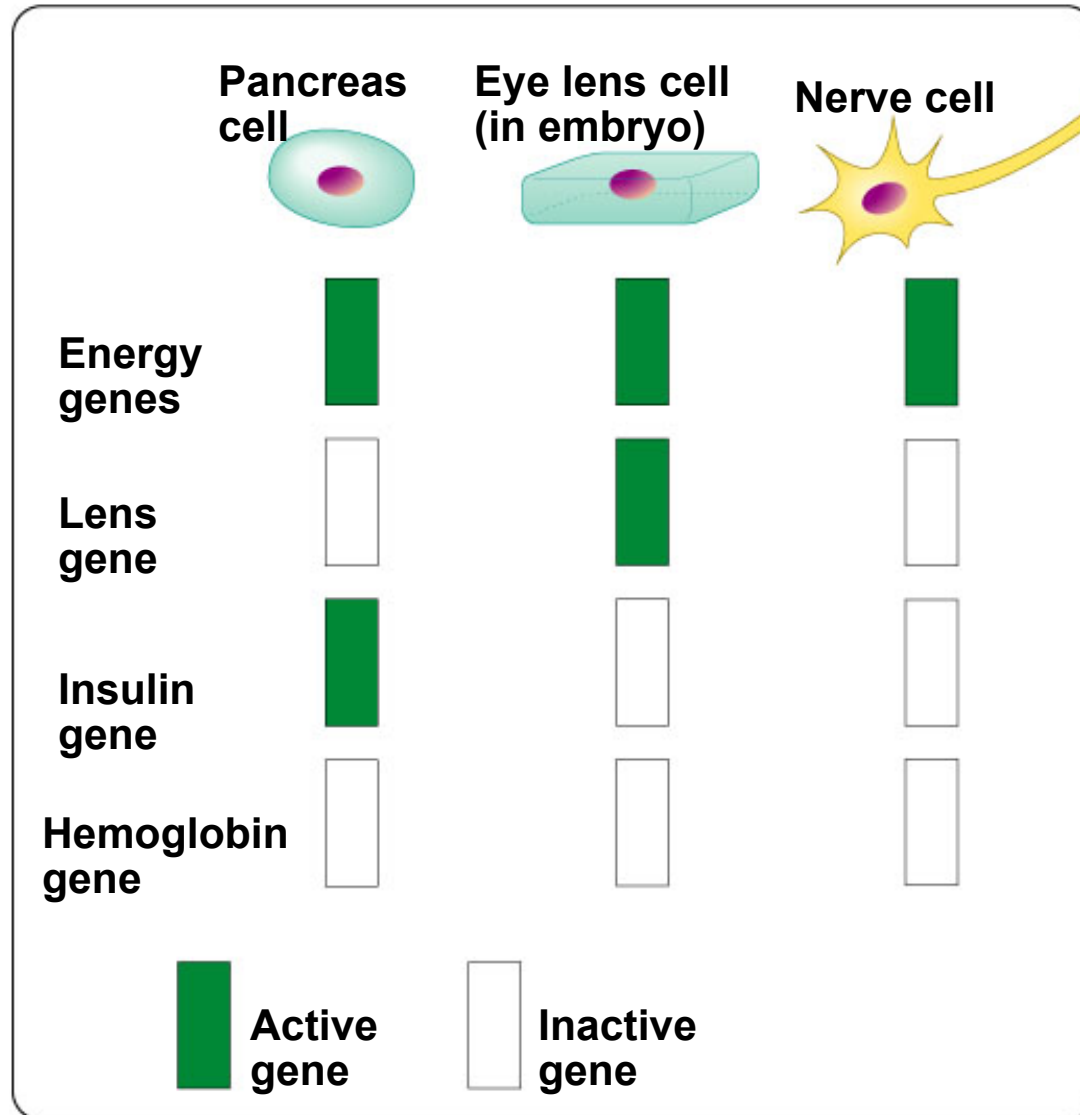


**Phenylalanine hydroxylase
protein is much larger
than phenylalanine**

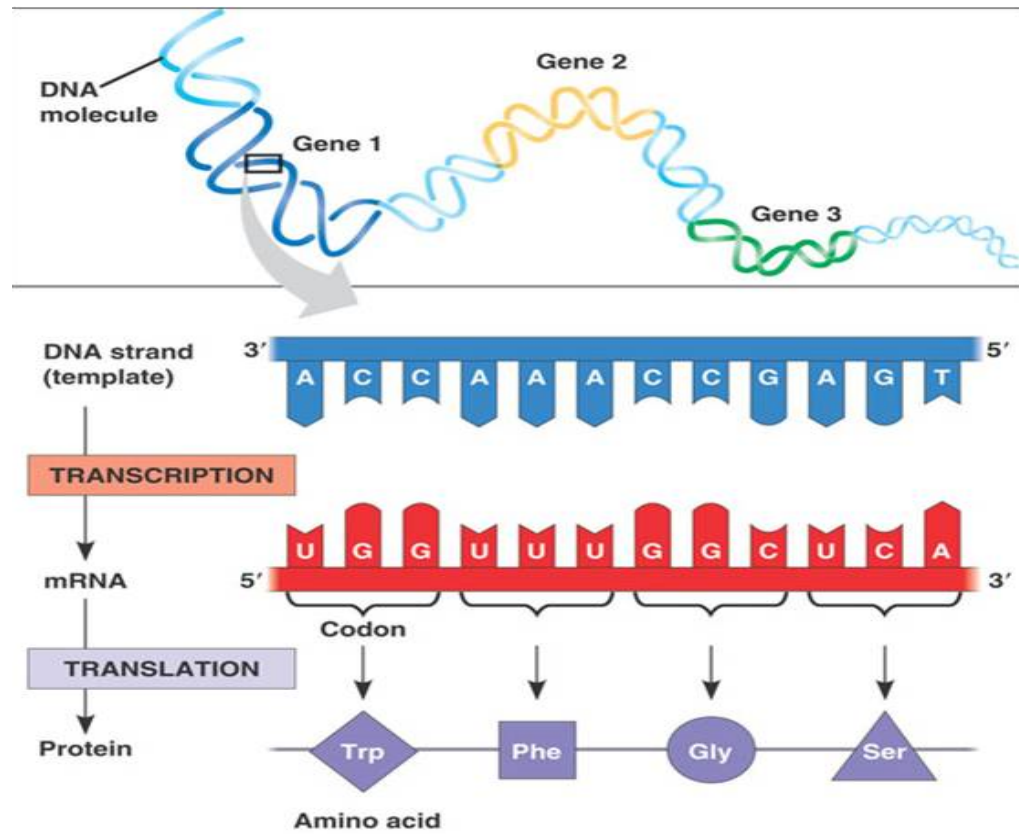


GENES AND PHENOTYPE INSIDE THE BODY

All cell types have all genes (complete DNA) but selectively express them



GENES AND PROTEINS



THE GENETIC CODE: RELATING DNA TO PROTEIN

		Second letter					
		U	C	A	G		
First letter	U	UUU Phenylalanine UUC Phenylalanine UUA Leucine UUG Leucine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine UAA Stop codon UAG Stop codon	UGU Cysteine UGC Cysteine UGA Stop codon UGG Tryptophan	Third letter	U
	C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine		C
	A	AUU Isoleucine AUC Isoleucine AUA Isoleucine AUG Methionine; start codon	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGU Serine AGC Serine AGA Arginine AGG Arginine		A
	G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartic acid GAC Aspartic acid GAA Glutamic acid GAG Glutamic acid	GGU Glycine GGC Glycine GGA Glycine GGG Glycine		G

THE INFORMATION CONTENT OF DNA IS ITS BASE SEQUENCE

DNA base sequence of the gene for human insulin

1

```
AGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAG  
CAGATCACTGTCCTTCTGCCATGGCCCTGTGGATGCGCCT  
CCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGA  
CCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTC  
ACACCTGGTGGGAAGCTCTCTACCTAGTGTGCGGGGAACGA  
GGCTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAG  
GACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCC  
TGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGG  
GTCCCTGCAG AGCGTGGCATTGTGGAACAATGCTGTACC  
AGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTA  
GACGCAGCCCGCAGGCAGCCCCACACCCGCCGCCTCCT  
GACCGAGAGAGATGGAATAAAGCCCTTGAACCAGCAAAA
```

469

DNA CAN BE MUTATED

Normal DNA sequence: ATC**G**GTAACT
TAG**C**CAATTGA

Mutated DNA sequence: ATC**A**GTAACT
TAC**T**CAATTGA

THE PERSONAL GENOME

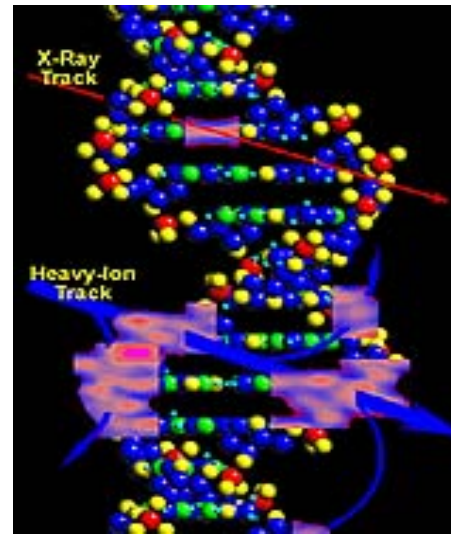
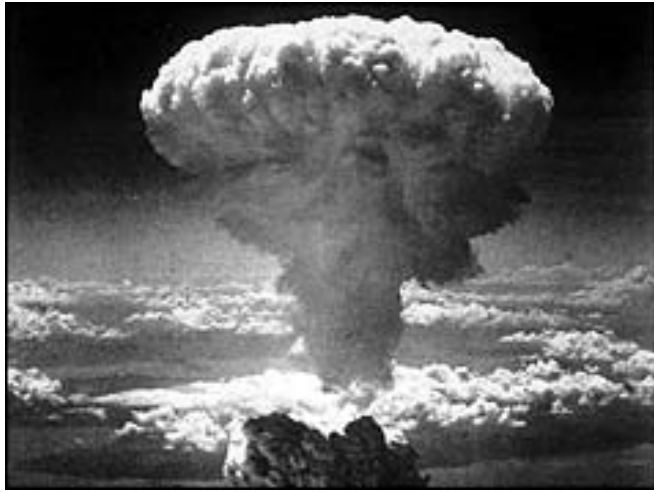
A. Genes and genomes

B. *The human genome project*

C. The human genome

D. The human genome and individuality

DNA DAMAGE BY RADIATION



Proposal: Identify gene mutations in Japanese people exposed to the bombs by comparing their entire DNA sequence – the **genome** -- to non-exposed people

Problem: We don't know the sequence of all human DNA

Solution: The Human Genome Project

TWO CHALLENGES FOR THE HUMAN GENOME PROJECT

- DNA molecules are huge: e.g., humans have 3.3 billion bases in 23 separate molecules (chromosomes)
- get smaller in If are, how smaller do they smaller put the right fragments order sequenced?

(If smaller fragments are sequenced, how do they get put in the right order?)

GENOME TECHNOLOGIES

Late 1960s: Sequencing a bacterial virus (5386 bp)

1 million years to complete human genome (~3,000 Mbp)

Late 1980s: sequencing techniques improved

1,000's of years to complete human genome

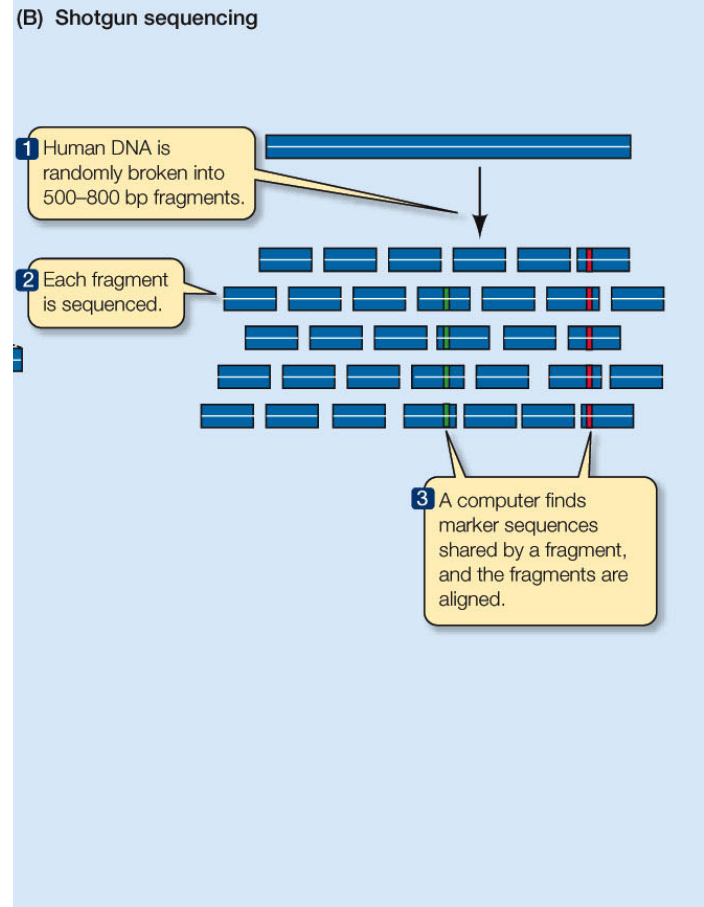
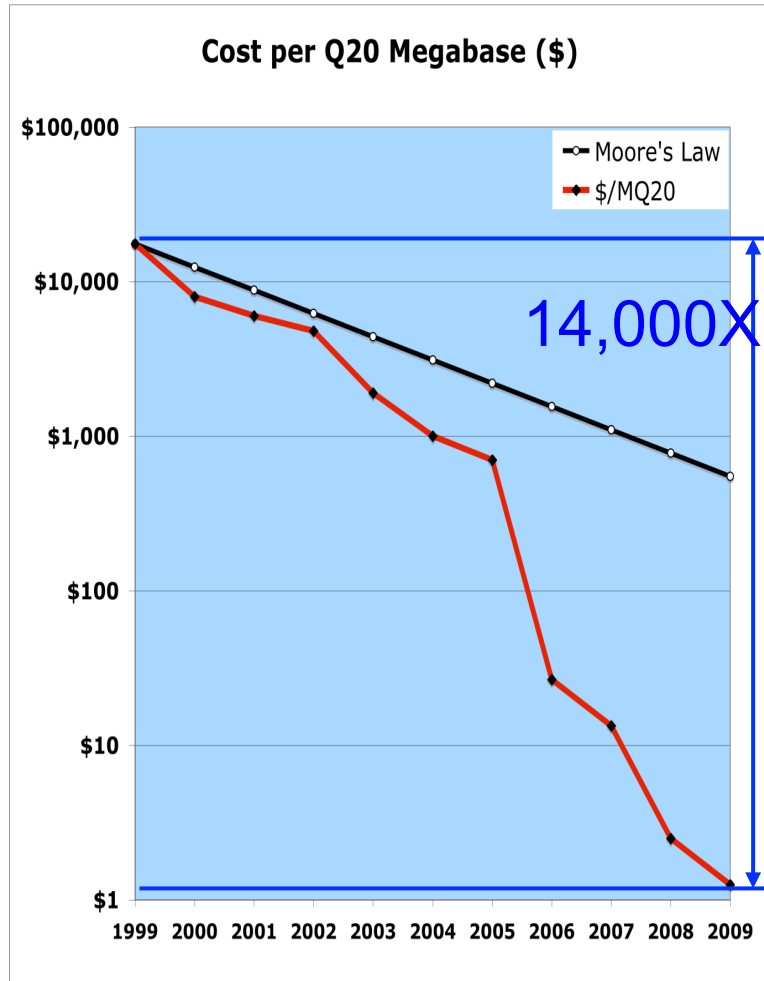
1990s: Human genome project: still faster methods

10 years to complete human genome

2010s: New sequencing technologies

Less than 1 week and \$1000?

GENOME SEQUENCING

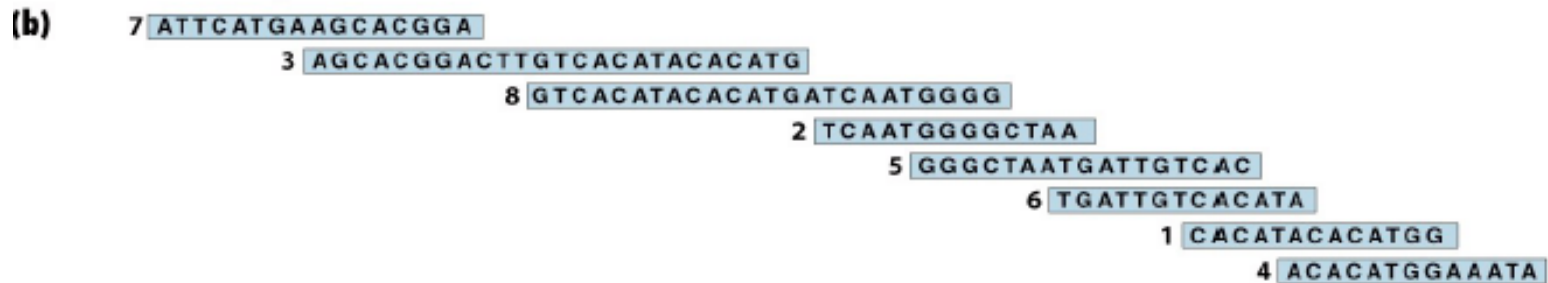


SEQUENCING THE GENOME

(a) Sequence reads

- Read 1 CACATACACATGG
- Read 2 TCAATGGGGCTAA
- Read 3 AGCACGGACTTGTACATACACATG
- Read 4 ACACATGGAAATA
- Read 5 GGGCTAATGATTGTCAC
- Read 6 TGATTGTCACATA
- Read 7 ATTCATGAAGCACGGA
- Read 8 GTCACATACACATGATCAATGGGG

Use computer to assemble sequence reads



Assembled sequence

(c) ATTCATGAAGCACGGACTTGTACATACACATGATCAATGGGGCTAATGATTGTCACATACACATGGAAATA

SEQUENCED GENOMES: MODEL ORGANISMS

(b) *Escherichia coli*



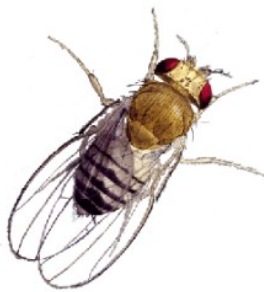
Bacterium



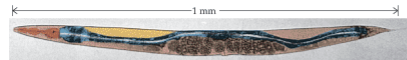
Yeast



Rice



Fruit fly



Worm



Mouse

GENOMES SEQUENCED

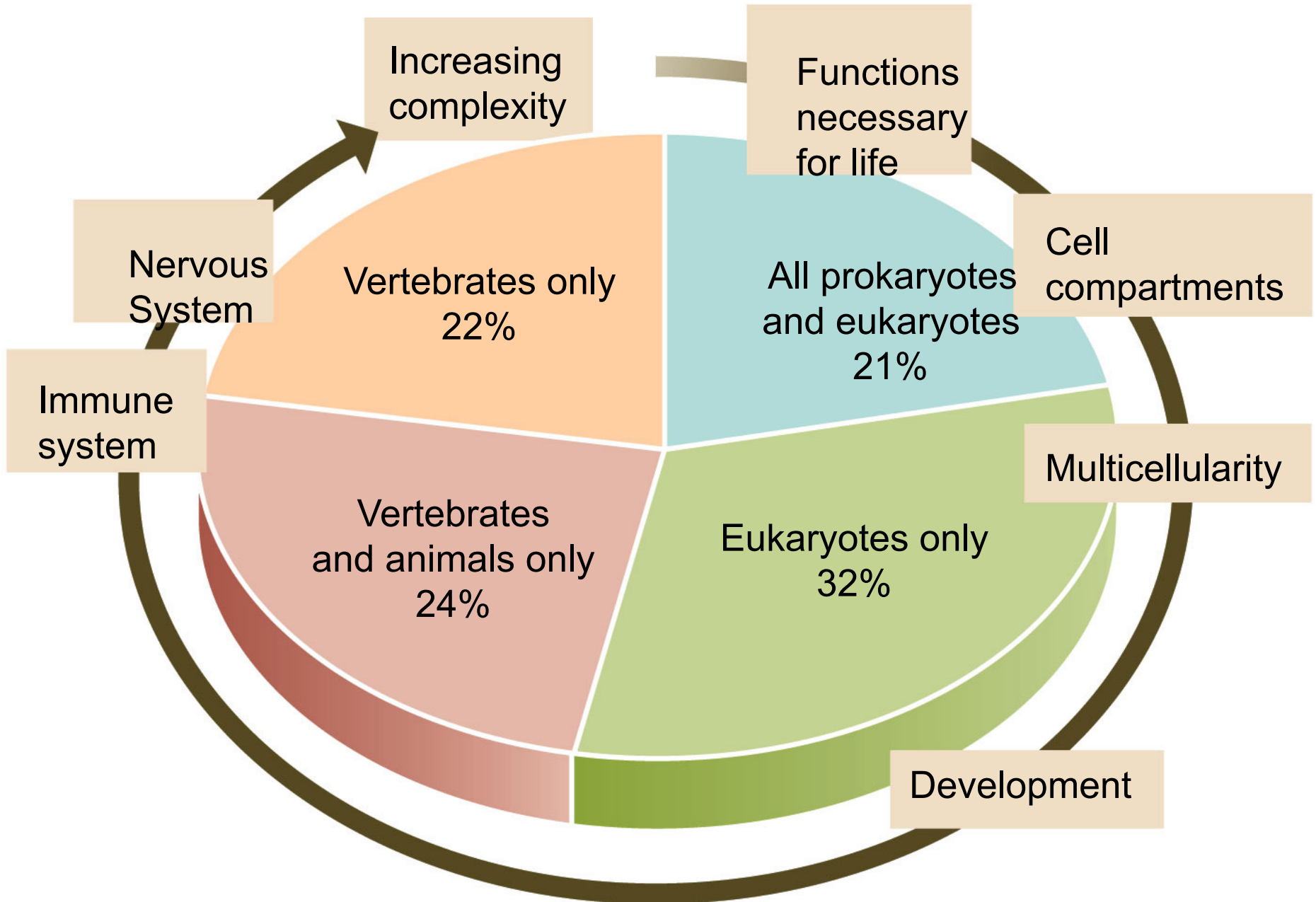
ORGANISM	HAPLOID GENOME SIZE (Mb)	NUMBER OF GENES	PROTEIN-CODING SEQUENCE
Bacteria			
<i>M. genitalium</i>	0.58	485	88%
<i>H. influenzae</i>	1.8	1,738	89%
<i>E. coli</i>	4.6	4,377	88%
Yeasts			
<i>S. cerevisiae</i>	12.5	5,770	70%
<i>S. pombe</i>	12.5	4,929	60%
Plants			
<i>A. thaliana</i>	115	28,000	25%
Rice	390	37,544	12%
Animals			
<i>C. elegans</i>	100	19,427	25%
<i>D. melanogaster</i>	123	13,379	13%
Pufferfish	342	27,918	10%
Chicken	1,130	25,000	3%
Human	3,300	24,000	1.2%

Mb = millions of base pairs

RESULTS FROM MODEL ORGANISMS

- Genomes can have many regions that do not encode proteins; in many cases, their functions are not known
- In some cases, not all genes are essential to life: there is a “minimal genome”
- Genes can be interrupted by “nonsense” stretches of DNA called introns; this requires cutting and splicing when the gene is expressed
- Genomes are not completely stable: certain regions of DNA move from place to place on the larger DNA
- There is a lot of similarity between organisms in their DNAs: the fruit fly and mouse are 75% similar in their protein-encoding genes

EVOLUTION OF THE GENOME



HUMAN GENOME SEQUENCE, 2000



Craig Venter

Francis Collins

The draft human genome: 2000; final first genome: 2003

THE PERSONAL GENOME

A. Genes and genomes

B. The human genome project

C. *The human genome*

D. The human genome and individuality

HUMAN GENOME SEQUENCE

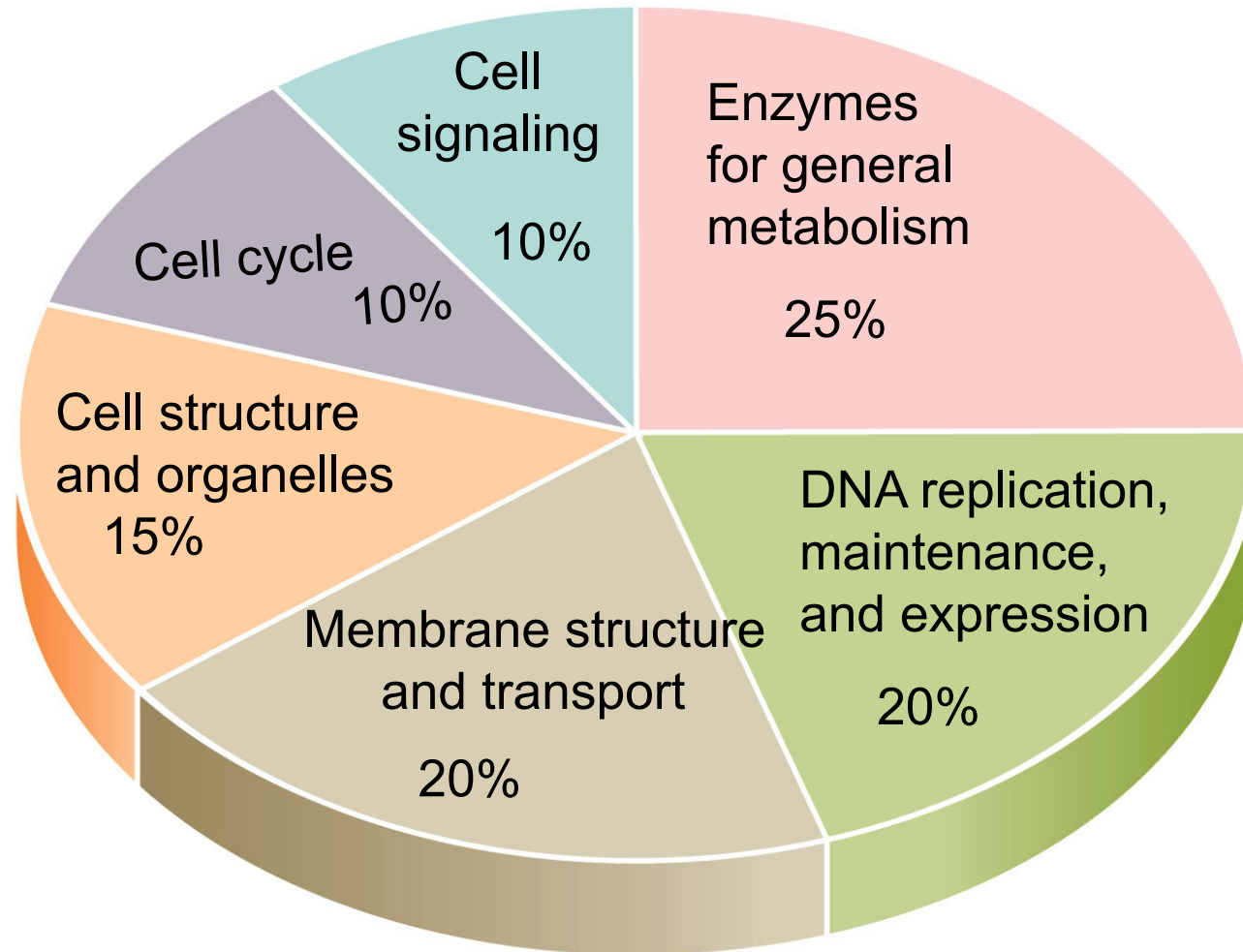
~3,000 bp (0.0001%) of Human Genome Sequence

```
TGCCGCGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGAAAGCTGGTGGCAGCGGGTCTGGGTCTGGCGGACCTGA
CGCGAAGGAGGCTTAGCAAGCTCTCCGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTGTCTCCAGCGTTGCCAATCGACCTAAAGAGAGG
CCGCGACTGTCGCCACCTGCGGGATGGGCTGGTCTGGGCGGTAAGGACACGGACTGGAAGGAGCGCGCGGAGGGAGGGAGGCTGGGAGTC
AGAATCGGAAAAGGGAGGTGCGGGGCGCCGAGGGAGCGAAGGAGGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGTGCGTACTGGGTGGA
GAAAGCCGCTAGACCAAATTTGGGCGCCGACCAGGCAGCACTCGGCTTTAACCCTGGCAGTGAAGGCGGGGAAAAGCAAAAGGAAGGGTGG
TGTGCGGAGTAGGGGTGGTGGGGGAAATGGAAGCAAATGCATCACAGCAGGTCAGAGAAAAAGGTTGAGCGGCAGGCACCCAGAGTAGTAG
GTCCTTTGGCATTAGGAGCTTGAGCCAGACGGCCCTAGCAGGGACCCAGCGCCCGAGAGACCATGCAGAGGTGCGCTCTGGAAAAGGCCAGCGT
TGTCCTCAAACCTTTTTTCAGGTGAGAAGTGGCCAAACCGAGCTTCGGAAGACACGTCGCCACGAAAGAGGAGGGCGTGTATGGGTGGGT
TGGGGTAAAGCAATAAGCAGTTTTTAAAAAGATGCCCTATCATTCATTGTTTTGAAAGAAAATGTGGGTATTGTAGAATAAAACAGAAAGCATA
AGAAGAGATGGAAGAATGCAACTGAAGCTGATTGAATAGAGAGCCACATCTACTTGCAACTGAAAAGTTAGAATCTCAAGACTCAAGTACGCTACT
ATGCACTTGTATTTATTTTCTAAGAACTAAAAATACCTGTTAAATAAGTACCTAAGTATGGTTTATTTGGTTTTCCCTTCATGCCCTGG
ACACTGATTGCTTCTTGCCACATACAGGTGCCATGCCTGCATATAGTAAGTGCTCAGAAAAACATTTCTTGACTGAATTCAGCCAACAAAAAT
TTGGGTAGGTAGAAAAATATATGCTTAAAGTATTTATTTGTTATGAGACTGGATATATCTAGTATTTGTCACAGGTAAATGATTCCTCAAATAATG
AAAGCAAATTTGTTGAAATATTTATTTGAAAAAGTTACTTCACAAGCTATAAAATTTAAAAAGCCATAGGAATAGATACCGAAGTTATATCCAA
CTGACATTTAATAAATGTATTCAATAGCCTAATGTGATGAGCCACAGAAGCTTGCAAACCTTAATGAGATTTTTTAAAAATAGCATCTAAGTTCGG
AATCTTAGGCAAAGTCTTGTAGATGTAGCACTTCATATTTGAAGTGTCTTTGGATATTGCATCTACTTTGTTCTCTGTTATTATACTGGTGTGA
ATGAATGAATAGGTACTGCTCTCTCTTGGGACATTACTTGACACATAAATFACC CAATGAATAAGCATACTGAGGTATCAAAAAAGTCAAATATGT
TATAAATAGCTCATATATGTGTGTAGGGGGGAAGAAATTTAGCTTTCACATCTCTCTTATGTTTAGTTCTCTGCATGTGCAGTTAATCCTGGAAC
TCCGGTGCTAAGGAGAGACTGTTGGCCCTTGAAGGAGAGCTCTCCCTGTGGATGAGAGAGAAGGACTTTACTCTTTGGAATTATCTTTTTGTGT
TGATGTTATCCACCTTTTGTACTCCACCATAAAAAATCGGCTTATCTATTGATCTGTTTTCTAGTCTTATAAAGTCAAATGTTAATTGGCAT
AAATATAGACTTTTTTAGCAGAGAACTTTGAGGAACCTAAATGCCAACAGCTAAAAATGCAGTTTCAGAAAGATGAATATTTTATGATA
GTTCTAAATACTAATGAACTTTAAAAATAGCTTACTATTGATCTGTCAAAGTGGGTTTTATATAATTTCTTTTTACAAATCACCTGCACATTT
AATATAGGTTAAAAAATGCTATCAGGCTGGTTTGC AAAAGAAATGATTACAAAGGCTGCTAAGTGTGTTAAGAGCATACTCATTTCTGTTCTCC
AAAAATTTTCAAAAGTCTTTAAGAATAGGTATGTTTTTAAAAAGTTAAGTTCTTATTTATAGCAACTGACAATCACCTAAAAATACCAATGA
TTACAAACTTCTCTGGCCTTCTGGACTGCAATTCAAAAGTGTA AAAAACATATTTTCTGCATTAAGTTAGGCAGTATTGCTTACTTTTCAA
GTGGTAGGCTTTGGAGTCAGATTATTTGATTGAGATCTTACATCTACTGTTTAGTAGCTCTGTTGCTGAGGCAGGTCCTTAAACATCTCTGTG
TGTGACTTGACTTTTAAAAATTTGAGACTGTCATAGGGGTTAATCCCTTGAGAAAATGAATGTGAAAAGTTAGCTTAATGTTAACGCTATTATT
ATGGATTACCATATTTTACATTCATCACAGTACATGCACCTTGTTAATAAAGATGCTCAATTCATCTTTGAGTATAATTTTGTGACTCTCAAT
CTGGATATGCAATGAGTGGCCCTGATGAGAATTTAATTTATGAAAAATTTGTTTTCACATGGCCTTACCAGATATACAGGAAACCGTCACATG
TTTCTATTGTATGTTTAAATGCCCTTAGAATTTAATTTCTGAATAGGATCCCTTCAGTTTGAGAGTCATAAAAGAGTAAAATTTATATGAT
```

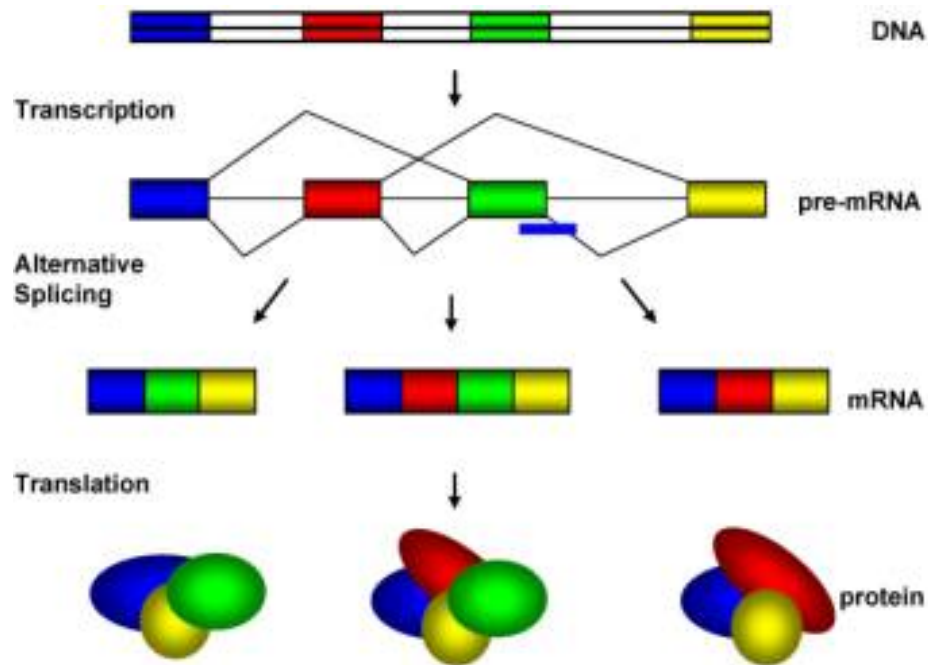
THE HUMAN GENOME: A USER'S GUIDE

- Of the 3.3 billion base pairs of DNA, about 1.5% encode proteins
- About 3% are regions apparently involved in regulating the expression of protein-coding genes
- There are 23,000 protein-coding genes, less than twice that of the fruit fly or worm. But there are far more proteins in a human than genes (150,000)
- About 50% of the genome is made up of movable sequences and repetitive sequences, most of whose functions are not known
- Over 99.8% of the genome sequence is the same in all people
- Most of the differences between people are at single base pairs

THE HUMAN GENOME: PROTEIN-CODING REGIONS

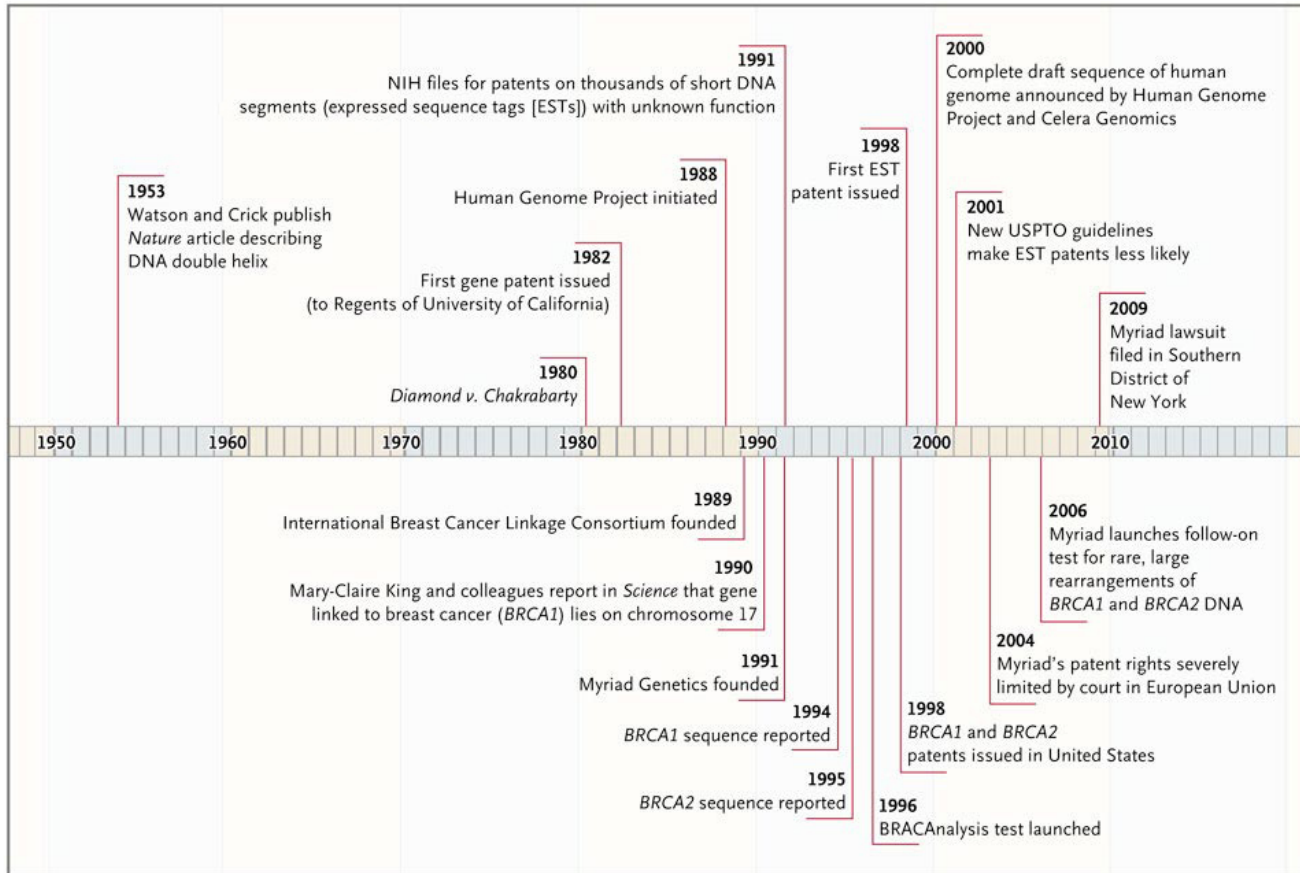


SHUFFLING OF GENE SEGMENTS LEADS TO A DIVERSITY OF PROTEINS



Average human gene: 3-4 splice variants

CAN A GENE SEQUENCE BE PATENTED?



2010 ruling: In light of DNA's unique qualities as a physical embodiment of information, none of the structural and functional differences cited by Myriad between native *BRCA1/2* DNA and the isolated *BRCA1/2* DNA claimed in the patents-in-suit render the claimed DNA 'markedly different'

THE PERSONAL GENOME

- A. Genes and genomes
- B. The human genome project
- C. The human genome
- D. ***The human genome and individuality***

VARIATIONS IN THE HUMAN GENOME

- Single nucleotide polymorphisms (SNP)
 - 1 every few hundred bp, mutation rate* $\approx 10^{-9}$ per generation
- Short indels (=insertion/deletion)
 - 1 every few kb, mutation rate variable
- Short tandem repeats (STR) repeat number
 - 1 every few kb, mutation rate $\leq 10^{-3}$

TGCATTGCGTAGGC
TGCATTCCGTAGGC

TGCATT---TAGGC
TGCATTCCGTAGGC

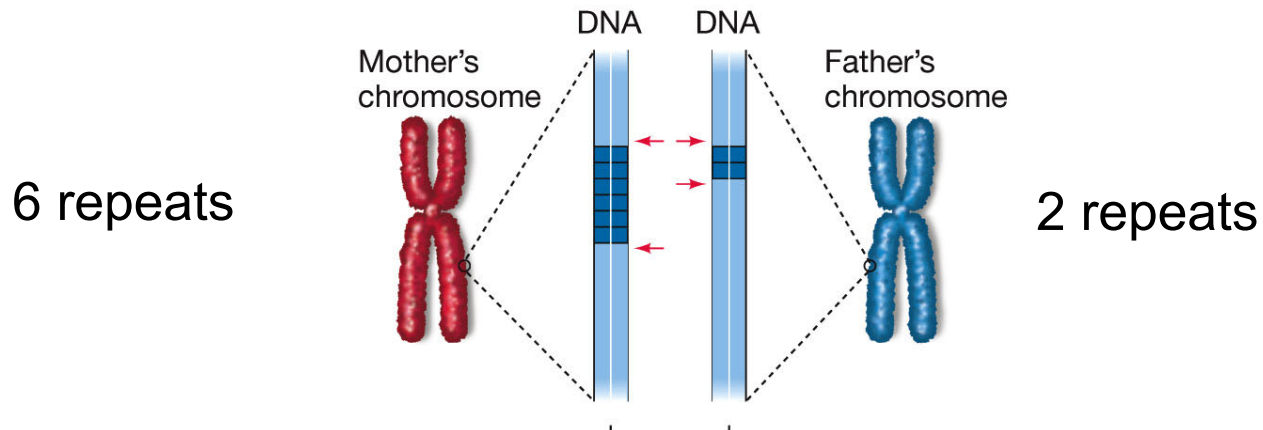
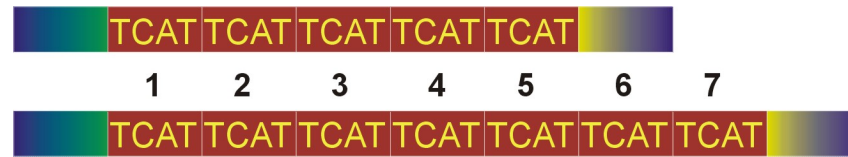
TGCTCATCATCATCAGC
TGCTCATCA-----GC

Unrelated humans differ by 6 million bases, ~99.8% identical overall, with coding regions 99.9999% identical

USES OF HUMAN GENOMIC VARIATION

- Identify individuals
- Relate genetic variants to phenotypes (diseases)
- Trace human evolution

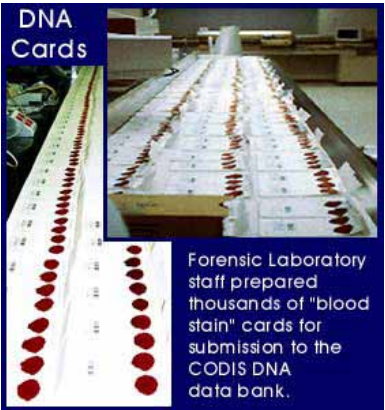
IDENTIFYING INDIVIDUALS: SHORT TANDEM REPEATS



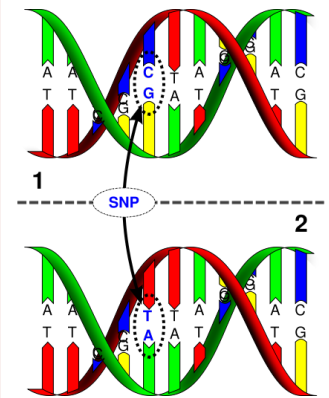
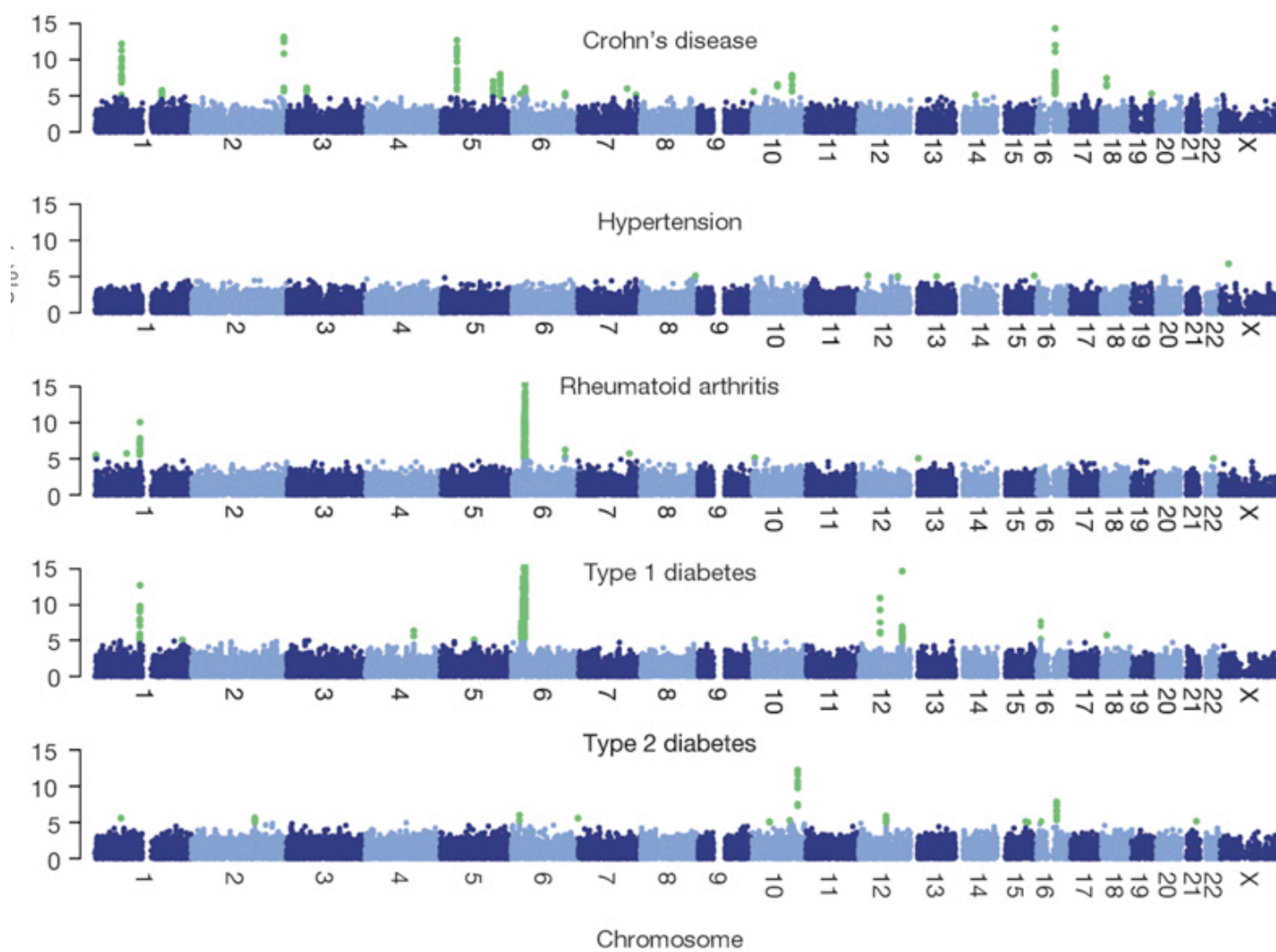
Repeat frequencies vary: the gene for 6 might be 80%
and the gene variant for 2 might be 20%

There are many genetic locations scattered on
the genome that have variable repeats

IDENTIFYING PEOPLE BY STR-DNA



RELATE VARIANTS TO PHENOTYPES: DNA-SNPs



RELATE VARIANTS TO PHENOTYPES

- *1000 genomes project*: Sequence 1000 genomes of people from many ethnic groups
 - Relate SNPs to diseases
- e.g., Age-related macular degeneration: an STR variant in a protein-coding region accounts for 50% of disease
- **But**: For most phenotypes the contribution of a single STR variant is low: 2-20%
 - There must be many genes involved for a phenotype:
The Dark matter of genetics

RELATE VARIANTS TO PHENOTYPES



23andMe genetics just got personal.

[log in](#) | [claim codes](#) | [blog](#) | [we're hiring!](#) | [help](#)

Search 23andMe

welcome

how it works

genetics 101

store

about us

Get the latest on your DNA with \$399 and a tube of saliva.

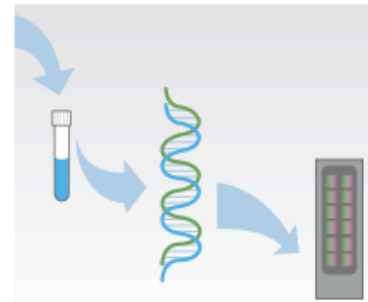
Here's what you do:



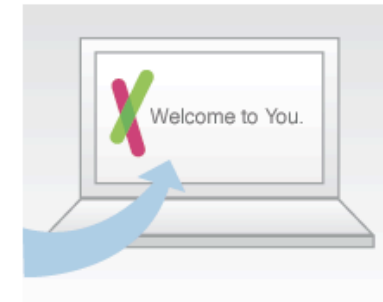
1. Order a kit (\$399 USD) from our [online store](#).



2. [Claim your kit](#), spit into the tube, and send it to the lab.



3. Our CLIA-certified lab analyzes your DNA in 8-10 weeks.



4. [Log in](#) and start exploring your genome.

Our high-density, custom genome scan includes:



Health and Traits

Discover how your genes influence your health and traits. Get your data on over 90 traits and diseases, with more topics



buy \$399 USD

try a demo

RELATE VARIANTS TO PHENOTYPES

Clinical Reports

Research Reports (83)

Show data for:

Disease Risks ?

	Psoriasis
	Parkinson's Disease
	Venous Thromboembolism
	Type 2 Diabetes
	Rheumatoid Arthritis

[See all 10 risk reports...](#)

Carrier Status ?

Alpha-1 Antitrypsin Deficiency	Variant Absent
BRCA Cancer Mutations (Selected)	Variant Absent
Bloom's Syndrome	Variant Absent
Cystic Fibrosis (Delta F508 mutation)	Variant Absent
G6PD Deficiency	Variant Absent

[See all 8 carrier status...](#)

Traits ?

Alcohol Flush Reaction	Does Not Flush
Bitter Taste Perception	Can Taste
Earwax Type	Wet
Eye Color	Likely Blue
Lactose Intolerance	Likely Tolerant

[See all 10 traits...](#)

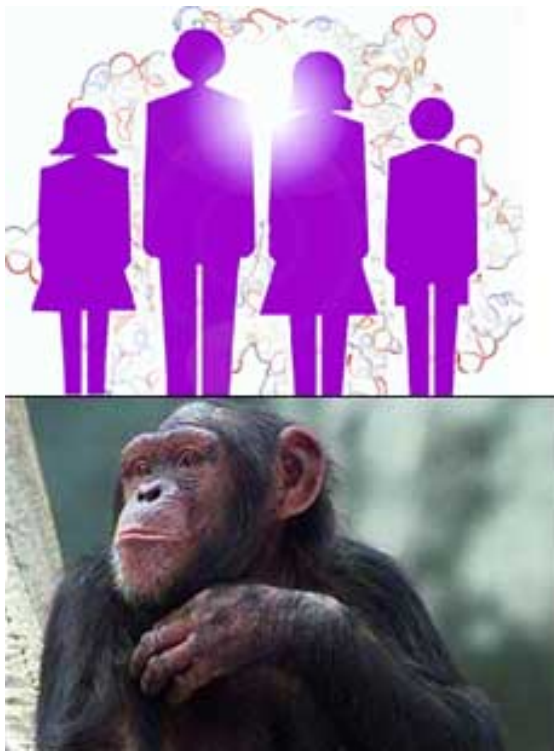
Drug Response ?

Warfarin (Coumadin) Sensitivity	Typical
---------------------------------	---------

[See new and recently updated reports »](#)

The genotyping services of 23andMe are performed in LabCorp's CLIA-registered laboratory. The results presented here have not been cleared or approved by the FDA but have been analytically validated according to CLIA standards.

GENOME EVOLUTION: CHIMPS/HUMANS



- Body shape and thorax
- Cranial properties (brain case and face)
- Relative brain size
- Relative limb length
- Long ontogeny and lifespan
- Small canine teeth
- Skull balanced upright on vertebral column
- Reduced hair cover
- Elongated thumb and shortened fingers
- Dimensions of the pelvis
- Presence of a chin
- S-shaped spine
- Language
- Advanced tool making
- Brain topology

6-8 million years

2% genome sequence different

NEANDERTHAL GENOME



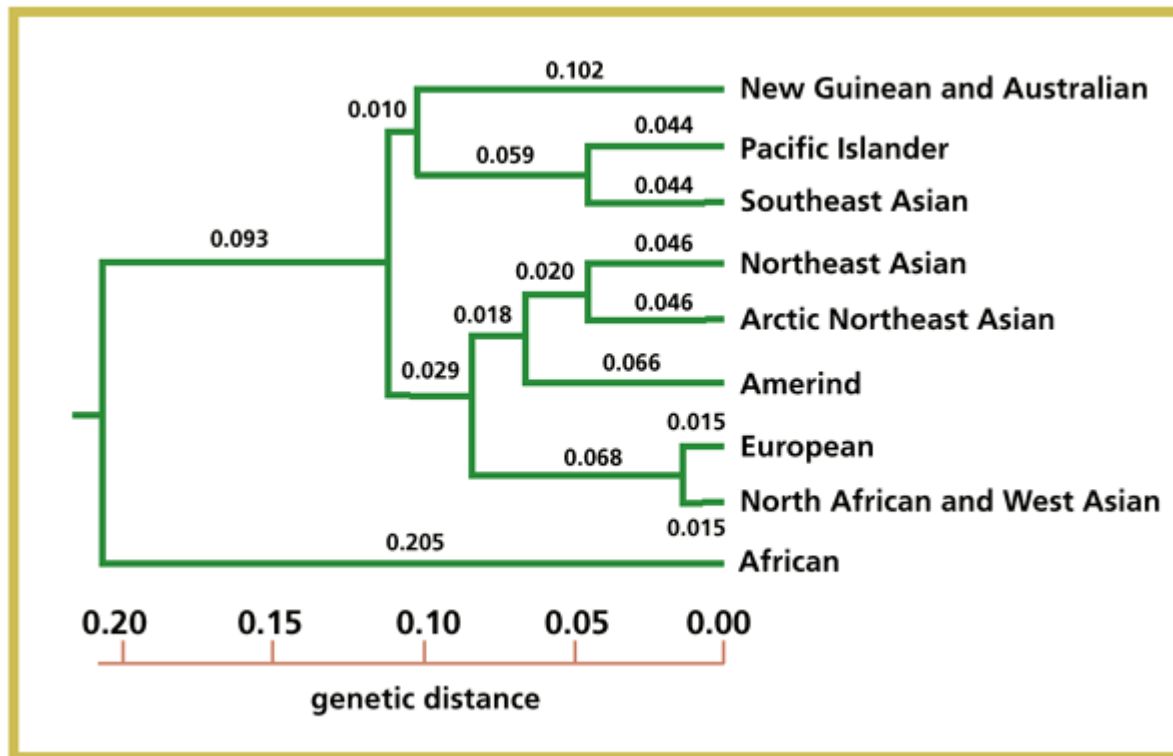
Lived 600,000 to 30,000 years ago

1-4% seq. contained in human genome
(interbred in Middle East)

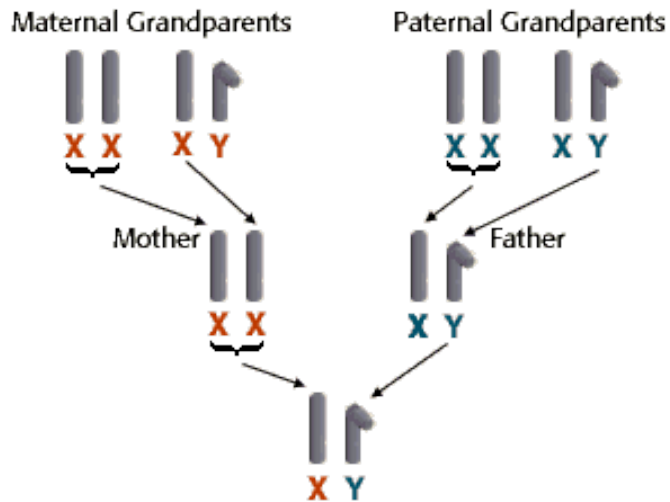
Genes for red hair, short stature, pale
skin, speech, hearing

Sequenced DNA from bones of
3 females, 38,000 years old, Croatia

RELATING HUMANS BY DNA SEQUENCES



Y CHROMOSOME DNA: INHERITED THROUGH THE MALE



Y-chromosomal “Adam”

Circa 60,000 years ago

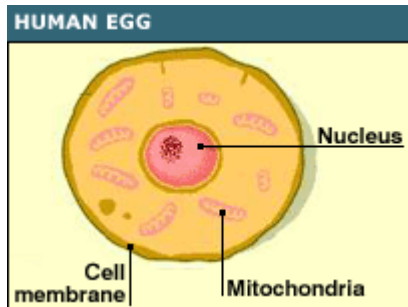
Most recent common male ancestor of all living humans

Lived in Africa

He lived as part of a group of early humans

But only his Y-chromosome survives today

MITOCHONDRIAL DNA: INHERITED THROUGH THE FEMALE



Mitochondrial “Eve”

Lived 60,000 to 250,000 years ago

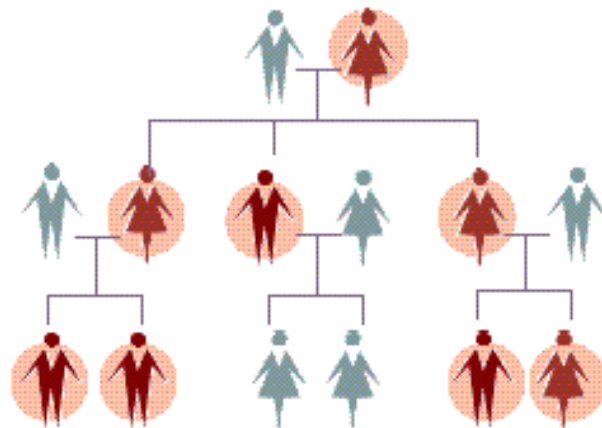
Most recent common female ancestor of all living humans

Lived in or around modern-day

Tanzania in Africa

She was part of a group of early humans

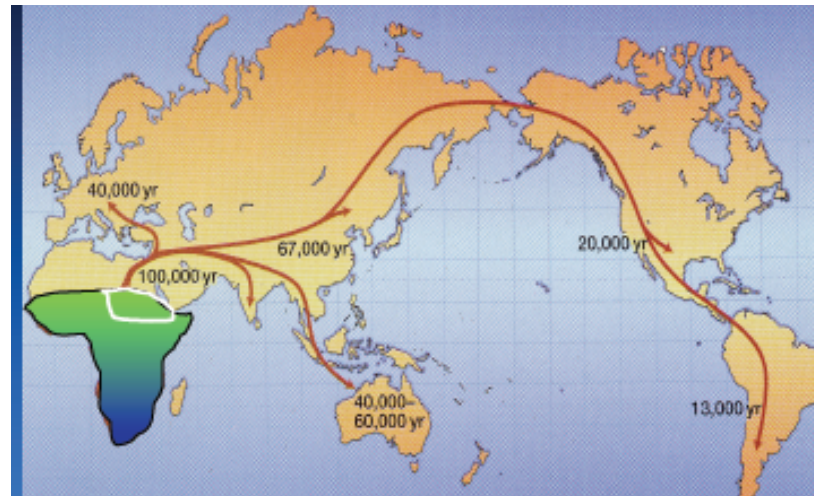
But only her mitochondrial DNA survives today



Mitochondrial DNA Inheritance

DNA VARIANTS AND HUMAN MIGRATIONS

- *Fully modern humans (like us) in Africa by around 60,000 years ago*
- *Early human migrations*
 - Southeastern Africa and spread throughout the continent
 - By 10,000 years ago, modern human beings had spread all over the globe
 - Genographic project traces human migrations



THE PERSONAL GENOME

- A. Genes and genomes
- B. The human genome project
- C. The human genome
- D. The human genome and individuality