Genomes

San Luis Potosi State University (UASLP) Mexico
Molecular Biology Course, Faculty of Medicine post-graduate program

Dr. Christian A. García-Sepúlveda

Viral & Human Genomics BSL-3 Laboratory

Last updated November 21, 2024 v1

Genomes

A genome is all the genetic information of an organism.

A genome is an organism's complete set of DNA, including all of its genes as well as its hierarchical, three-dimensional structural configuration.

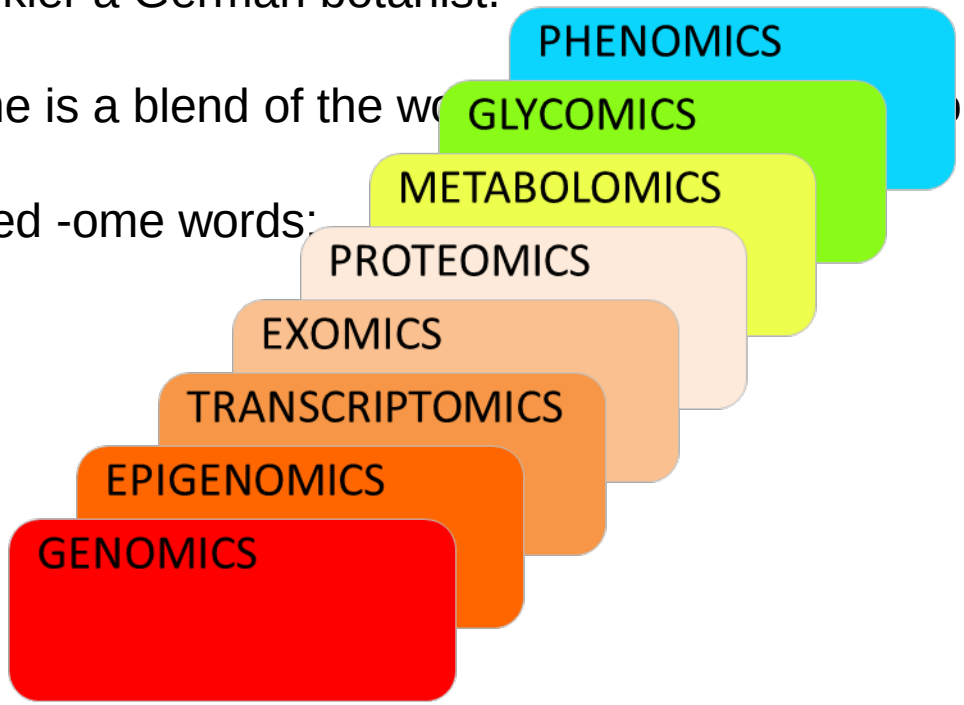
Consists of nucleotide sequences of DNA or RNA.

Term created in 1920 by Hans Winkler a German botanist.

Oxford Dictionary suggest the name is a blend of the wo ome.

Fits systematically with a few related -ome words:

- Biome
- Rhysome
- Nucleosome
- Chromosome
- Ribosome
- Replisome
- Etceterasome



Genome sizes

C-value Paradox: Genome size does not correlate with organism complexity.

Non-coding DNA proportion: humans have ~98% non-coding DNA, while some simple organisms have less.

Viral Genomes: Extremely compact, ranging from a few thousand to over a million base pairs, with minimal non-coding DNA.

Gene Density: Smaller genomes often have higher gene density, while larger genomes tend to have more repetitive and non-coding sequences.

Comparative Insight: Genome size and structure provide insights into evolutionary history, adaptation, and complexity.

organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mbp	4,300	1
budding yeast *S. cerevisiae*	12 Mbp	6,600	16
fission yeast *S. pombe*	13 Mbp	4,800	3
amoeba *D. discoideum*	34 Mbp	13,000	6
nematode *C. elegans*	100 Mbp	20,000	12 (2n)
fruit fly *D. melanogaster*	140 Mbp	14,000	8 (2n)
model plant *A. thaliana*	140 Mbp	27,000	10 (2n)
moss *P. patens*	510 Mbp	28,000	27
mouse *M. musculus*	2.8 Gbp	20,000	40 (2n)
human *H. sapiens*	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known animal RNA virus)	1.7 Kb	1	ssRNA
HIV-1	9.7 kbp	9	2 ssRNA (2n)
influenza A	14 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
Pandoravirus salinus (largest known viral genome)	2.8 Mbp	2500	1 dsDNA
organelles			
mitochondria - *H. sapiens*	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
mitochondria – *S. cerevisiae*	86 kbp	8	1
chloroplast – *A. thaliana*	150 kbp	100	1
bacteria			
C. ruddii (smallest genome of an endosymbiont bacteria)	160 kbp	182	1
M. genitalium (smallest genome of a free living bacteria)	580 kbp	470	1
H. pylori	1.7 Mbp	1,600	1
Cyanobacteria S. elongatus	2.7 Mbp	3,000	1
methicillin-resistant *S. aureus* (MRSA)	2.9 Mbp	2,700	1
B. subtilis	4.3 Mbp	4,100	1
S. cellulosum (largest known bacterial genome)	13 Mbp	9,400	1
archaea			
Nanoarchaeum equitans (smallest parasitic archaeal genome)	490 kbp	550	1
Thermoplasma acidophilum (flourishes in pH<1)	1.6 Mbp	1,500	1
Methanocaldococcus (Methanococcus) jannaschii (from ocean bottom hydrothermal vents; pressure >200 atm)	1.7 Mbp	1,700	1
Pyrococcus furiosus (optimal temp 100°C)	1.9 Mbp	2,000	1
eukaryotes - multicellular			
pufferfish *Fugu rubripes* (smallest known vertebrate genome)	400 Mbp	19,000	22
poplar *P. trichocarpa* (first tree genome sequenced)	500 Mbp	46,000	19
corn *Z. mays*	2.3 Gbp	33,000	20 (2n)
dog *C. familiaris*	2.4 Gbp	19,000	40
chimpanzee *P. troglodytes*	3.3 Gbp	19,000	48 (2n)
wheat *T. aestivum* (hexaploid)	16.8 Gbp	95,000	42 (2n=6x)
marbled lungfish *P. aethiopicus* (largest known animal genome)	130 Gbp	unknown	34 (2n)
herb plant *Paris japonica* (largest known genome)	150 Gbp	unknown	40 (2n)

Genomes

Viroid and viral genomes are either DNA or RNA based, are compact and overlapped.

Prokaryote (Bacterial) genomes are DNA based, mainly circularized and single.

Bacteria usually have one or two chromosomes containing all essential genetic material.

Bacteria also contain smaller extrachromosomal plasmid molecules that carry additional non-essential genetic information.

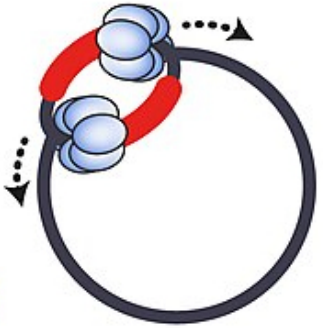
Eukaryote genomes are diploid and have nuclear and endosymbiont components.

- All eukaryotes have mitochondrial genome.
- Algae and plants also contain chloroplast genome.

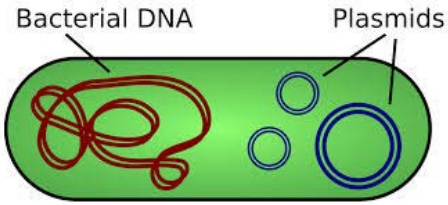
The scientific literature term 'genome' is usually restricted to the large chromosomal DNA molecules in bacteria and eukaryotes.

Prokaryotic genomes

Compact and Efficient:
Prokaryotic genomes are typically smaller and more compact than eukaryotic genomes, ranging from about 0.5 to 10 Mb in size.



Circular DNA
Most prokaryotes have a single, circular chromosome, although some species may possess multiple chromosomes or linear chromosomes.

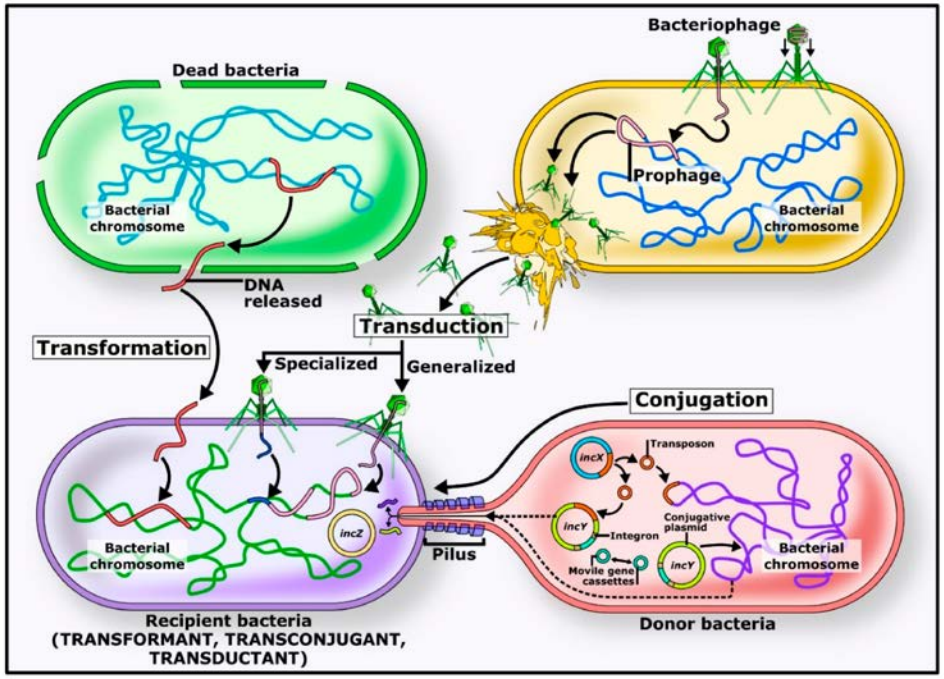


Replication Origin
Replication usually starts at a single origin of replication (OriC) and proceeds bidirectionally.

Plasmids
In addition to the main chromosome, prokaryotes often carry extrachromosomal DNA in the form of plasmids, which can confer advantageous traits such as antibiotic resistance.

Prokaryotic genomes

Horizontal Gene Transfer (HGT)
Prokaryotes frequently acquire genetic material through HGT mechanisms like transformation, transduction, and conjugation, contributing to genetic diversity and adaptation.



Lack of Introns
Prokaryotic genomes generally lack introns, leading to a high proportion of coding DNA.

High Mutation Rate
Prokaryotic genomes tend to have higher mutation rates, facilitating rapid evolution in response to environmental changes.

Transposable Elements
Prokaryotic genomes often contain mobile genetic elements like transposons, which contribute to genome plasticity.

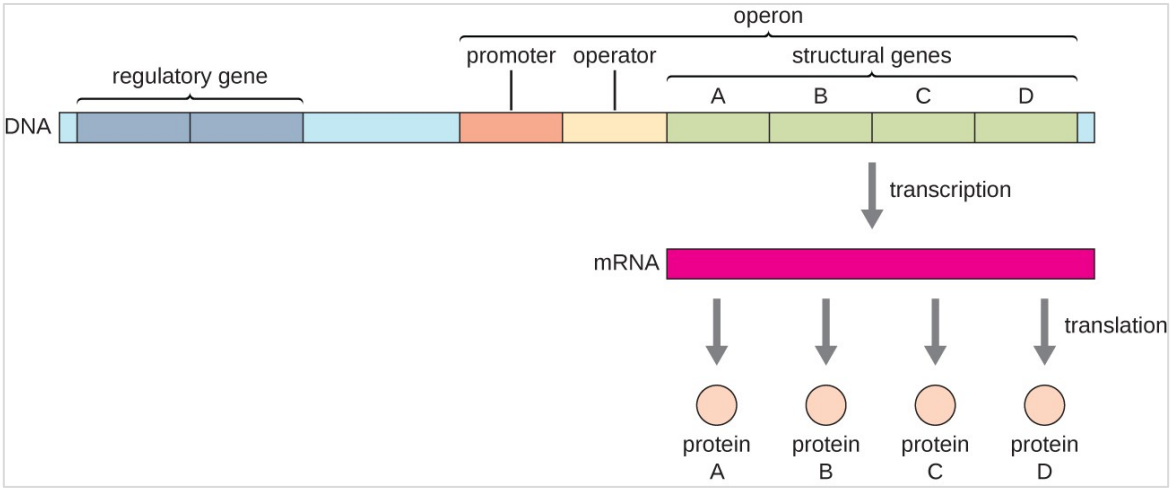
Prokaryotic genomes

Minimal Non-Coding Regions

Non-coding DNA is minimal compared to eukaryotic genomes, optimizing genetic information density.

Operons

Genes are often organized into operons, allowing coordinated expression of functionally related genes.



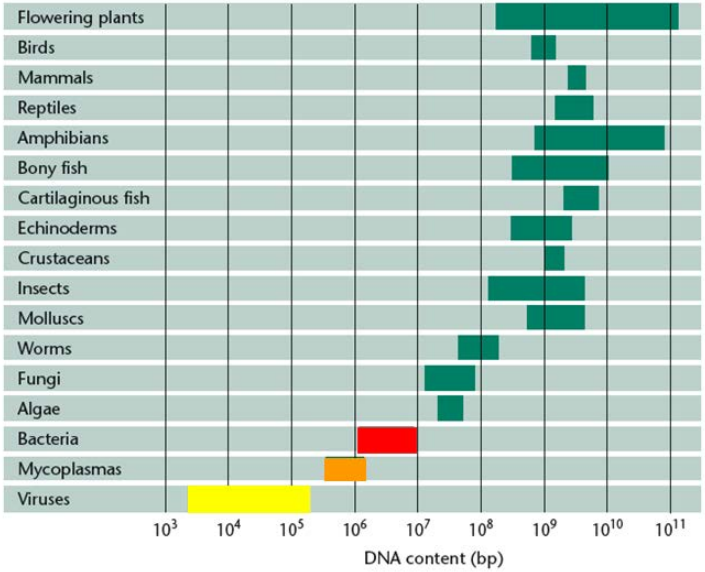
Functional Redundancy

Despite their compact size, prokaryotic genomes often encode redundant systems for critical functions, enhancing survival under diverse conditions.

Eukaryotic genomes

Large and Complex

Eukaryotic genomes are significantly larger and more complex than prokaryotic genomes, ranging from tens of millions to billions of base pairs.

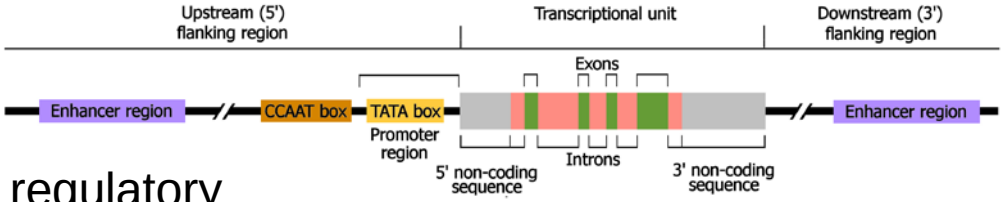


Linear Chromosomes

Eukaryotic genomes are organized into multiple linear chromosomes housed within a nucleus.

Introns and Exons

Genes are interrupted by non-coding sequences (introns), requiring splicing to generate functional mRNAs.



Regulatory Elements

Eukaryotic genomes contain extensive regulatory elements, including promoters, enhancers, and silencers, to finely tune gene expression.

Eukaryotic genomes

Repetitive DNA

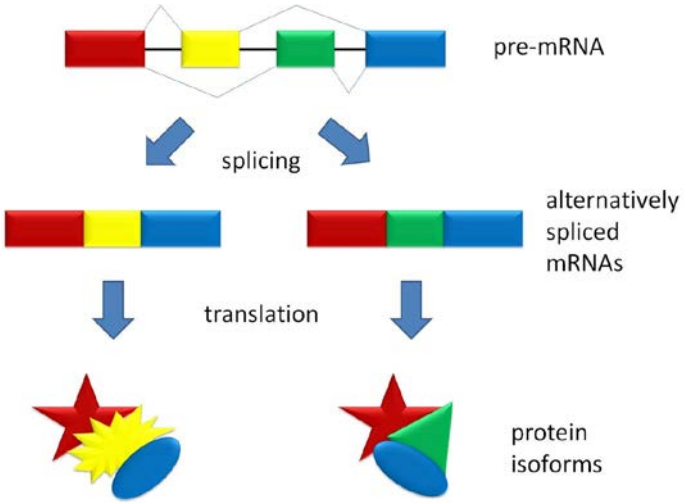
A large portion of eukaryotic genomes consists of repetitive DNA sequences, including tandem repeats and transposable elements.

Epigenetics

DNA methylation and histone modifications regulate gene expression without altering the underlying DNA sequence.

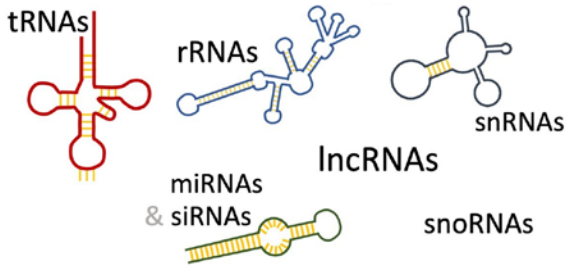
Alternative Splicing

Eukaryotes utilize alternative splicing to produce multiple protein isoforms from a single gene, increasing proteome diversity.



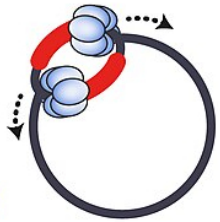
Non-Coding RNA

Significant portions of eukaryotic genomes transcribe non-coding RNAs, such as microRNAs and long non-coding RNAs, with regulatory roles.

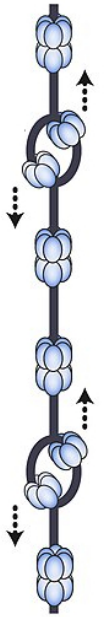


Eukaryotic genomes

Genomic flexibility (Expansion/contraction)
Genome size does not always correlate with organismal complexity, a phenomenon known as the "C-value paradox.“



Multiple Origins of Replication
Eukaryotic chromosomes have multiple origins of replication to ensure timely DNA replication.



Mitochondrial and Chloroplast Genomes
Eukaryotes possess organelle genomes that are typically circular and inherited maternally.

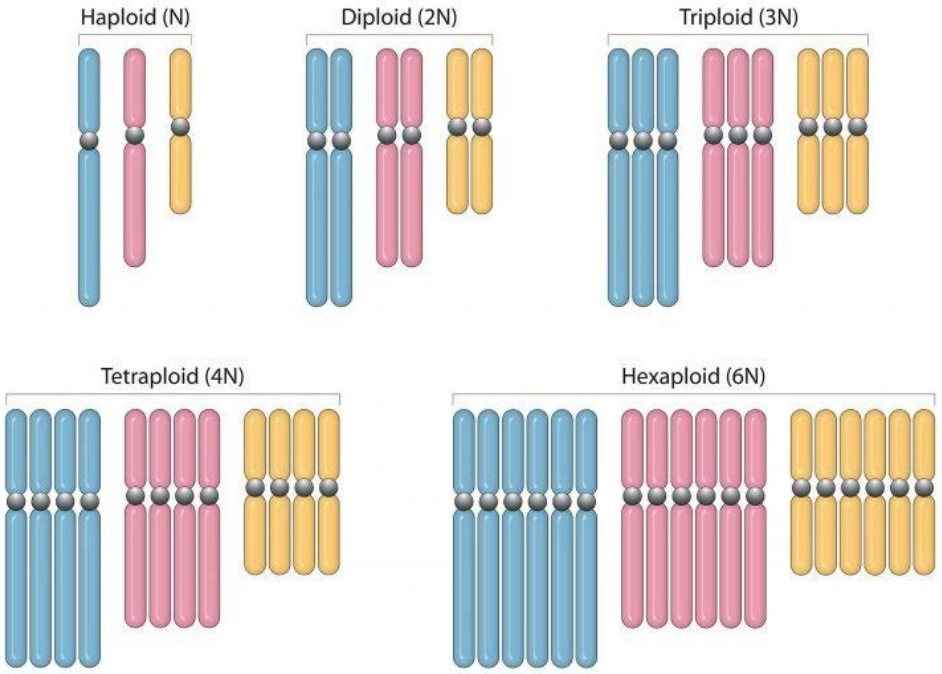
Dynamic Chromatin
Chromatin structure, regulated by nucleosomes and chromatin remodeling complexes, influences genome accessibility and transcription.

Gene Families and Duplications
Eukaryotic genomes often include gene families and duplicated genes, which can evolve new functions over time.

Eukaryotic genomes

Polyploidy

Some eukaryotic organisms exhibit polyploidy, possessing multiple sets of chromosomes, which can drive evolution and adaptation.

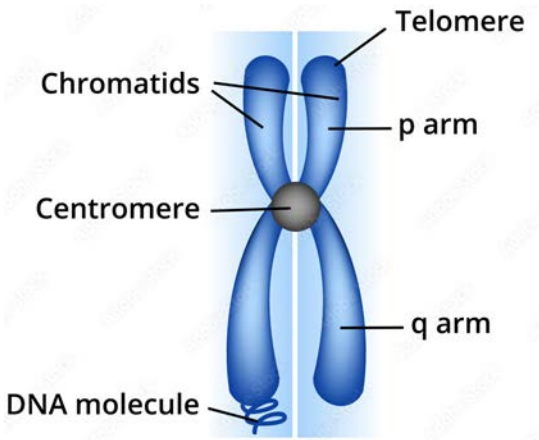


Pseudogenes

Eukaryotic genomes contain pseudogenes—non-functional remnants of previously active genes.

Telomeres and Centromeres

Chromosomes have specialized structures, such as telomeres for protecting ends and centromeres for segregation during cell division.



Endosymbionts

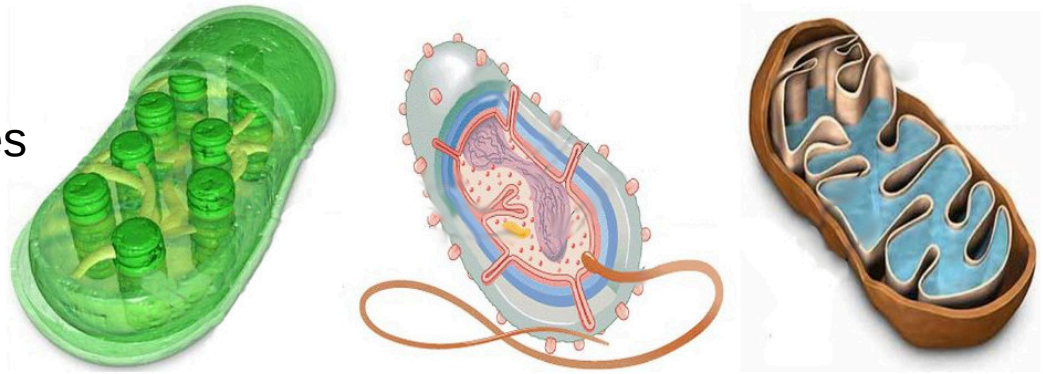
Organisms that live within the body or cells of another, often in mutualistic relationships (e.g., nitrogen-fixing bacteria in legumes, algae in corals).

~2.2 billion years ago archaea absorbed a bacterium leading to mitochondria in eukaryotic cells.

~1 billion years ago eukaryote absorbed cyanobacteria leading to chloroplasts.

~100 million years ago amoebas (*Paulinella*) and marine algae (e.g., *Braarudosphaera bigelowii*) absorbed cyanobacterial endosymbionts leading to chromatophores and nitroplasts.

Rhopalodiaceae diatoms absrobed cyanobacterial endosymbionts (diazoplasts), leading to early stages of organelle evolution.



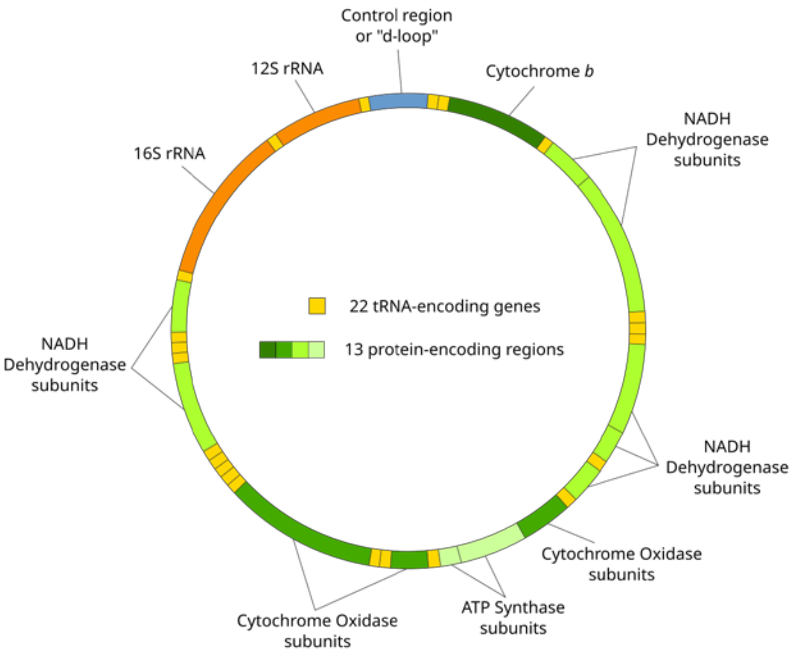
Mitochondrial genome (mtDNA)

mtDNA is a small portion of total cellular DNA

First significant part of the human genome to be sequenced.

Contains 16,569 base pairs encoding 13 proteins.

Genetic code differs slightly from nuclear DNA.



Animal mtDNA evolves faster than nuclear DNA, making it vital for phylogenetics and evolutionary studies.

Animals have 37 genes in their mitochondrial DNA: 13 for proteins, 22 for tRNAs, and 2 for rRNAs.

Mitochondrial genomes for animals average about 16,000 bp in length.

en.wikipedia.org/wiki/Mitochondrial_DNA

Mitochondrial genome

The vast majority of the proteins in the mitochondria (approx. 1500 different types in mammals) encoded by nuclear DNA.

Most thought to be of bacterial origin but transferred to the eukaryotic nucleus during evolution.

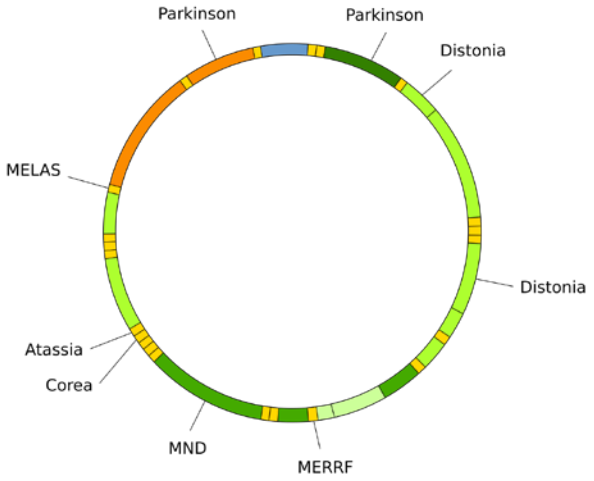
Transferring mitochondrial genes to the nucleus has several advantages:

- Addiction strategy to prevent run-away mitochondrion

Some genes are retained in mtDNA:

- Difficulty of targeting cytosolic hydrophobic protein products to mitochondrion

Mitochondrial DNA linked to several human diseases



en.wikipedia.org/wiki/Mitochondrial_DNA

Mitochondrial genome

Across all organisms, there are 6 main mitochondrial genome types

Classified by structure (i.e. circular versus linear), size, presence of introns or plasmid like structures, and whether the genetic material is a singular molecule or collection of homogeneous or heterogeneous molecules.

Genome Type[14]	Kingdom	Introns	Size	Shape	Description
1	Animal	No	11–28 kbp	Circular	Single molecule
2	Fungi, Plant, Protista	Yes	19–1000 kbp	Circular	Single molecule
3	Fungi, Plant, Protista	No	20–1000 kbp	Circular	Large molecule and small plasmid like structures
4	Protista	No	1–200 kbp	Circular	Heterogeneous group of molecules
5	Fungi, Plant, Protista	No	1–200 kbp	Linear	Homogeneous group of molecules
6	Protista	No	1–200 kbp	Linear	Heterogeneous group of molecules

en.wikipedia.org/wiki/Mitochondrial_DNA

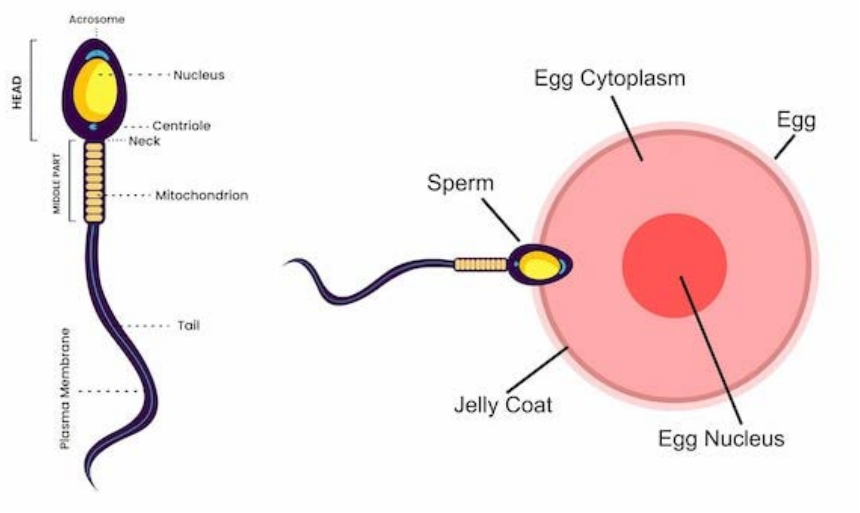
Mitochondrial genome

In most multicellular organisms, mtDNA is inherited from the mother (maternally inherited).

Mechanisms for this include:
Simple dilution (ovum contains approx. 200,000 mtDNA molecules, sperm 5).
Degradation of sperm mtDNA in the male genital tract and in the fertilized egg.
Failure of sperm mtDNA to enter the egg.

Uniparental inheritance of mtDNA inheritance seen in most animals, most plants and also in fungi.

Used to trace maternal lineage as only ovum provides mitochondria to zygote.



en.wikipedia.org/wiki/Mitochondrial_DNA

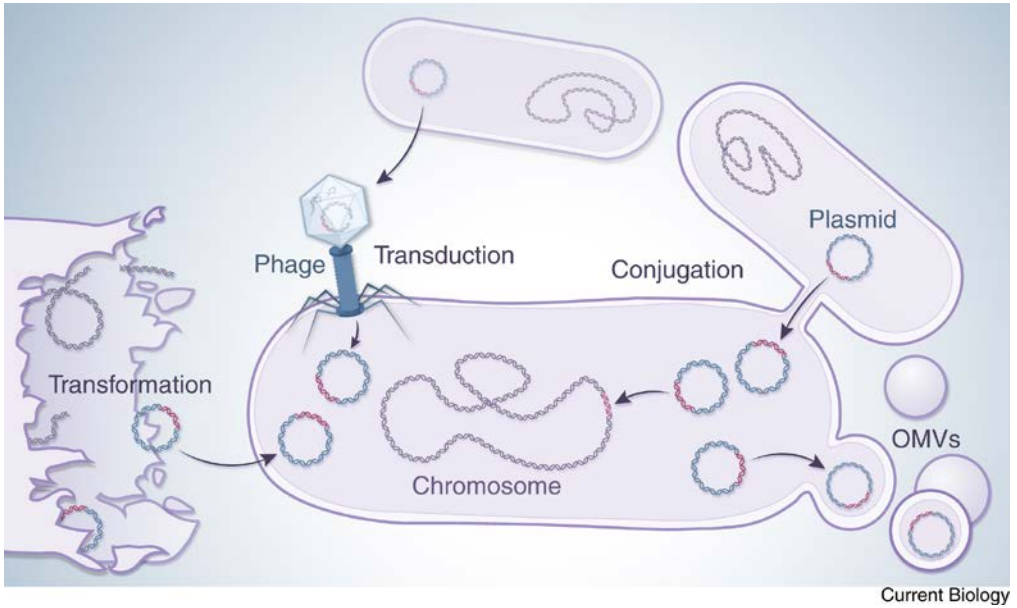
Plasmids

Small, extrachromosomal DNA molecule separate from genophore.

Typically, circular and double-stranded.

Capable of independent replication within the cell.

Found in bacteria, occasionally in archaea and rarely in eukaryotes.



Often carry antibiotic resistance or virulence factors.

Plasmids are considered replicons, DNA units capable of replicating autonomously.

Like viruses respond to selfish gene hypothesis but not generally classified as life.

Transmitted from one bacterium to another through conjugation, transduction, transformation or through Outer Membrane Vesicles (OMVs).

Wein T, Dagan T. Plasmid evolution. Curr Biol. 2020 Oct 5;30(19):

Functional Classification of Plasmids

Fertility F-plasmids
Contain *tra* genes which allow bacterial conjugation through a sex pili.

Resistance (R) plasmids
Contain antibiotic resistance genes, first discovered in 1959.

***Col* plasmids**
Contain genes coding for *bacteriocins*, proteins that can kill other bacteria.

Degradative plasmids
Enable the digestion of unusual substances like toluene or salicylic acid.

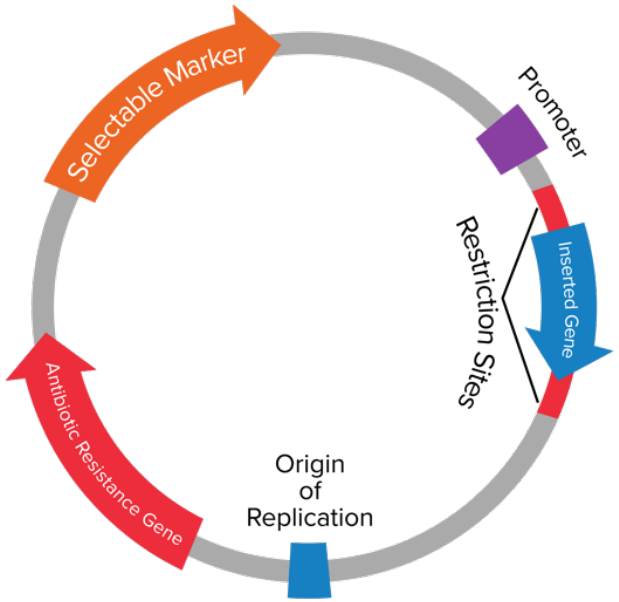
Virulence plasmids
Turn the bacterium into a pathogen. Like Ti plasmid in *Agrobacterium tumefaciens*.

Bacteria under selective pressure will keep plasmids containing virulence factors if they benefit survival, removal of the selective pressure can lead to the loss of a plasmid.

Wein T, Dagan T. Plasmid evolution. Curr Biol. 2020 Oct 5;30(19):

Plasmids as Genetic Engineering Vectors

Artificially constructed plasmids may be used as vectors in genetic engineering (molecular cloning vectors).



Commonly used to clone and amplify exogenous (recombinant) DNA sequences.

The most-commonly used bacterial cloning vectors.

Contain a site that allows DNA fragments to be ligated at several restriction sites.

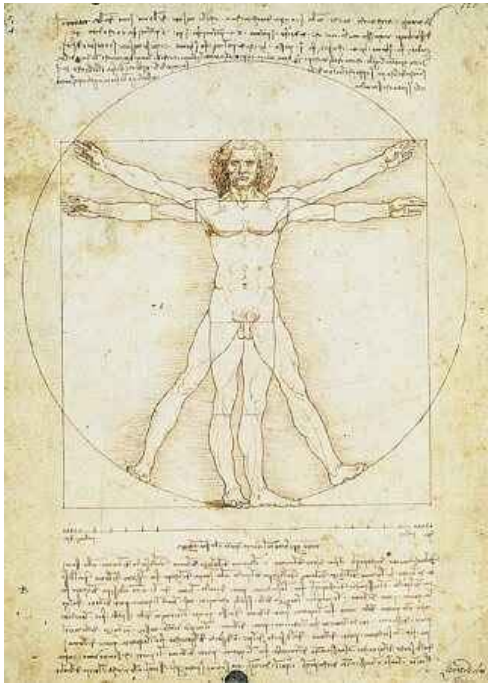
Engineered plasmids introduced into bacteria (competent cells) by transformation.

Contain a selectable marker (antibiotic resistance gene) which confers the ability to survive and proliferate in a selective growth medium containing that antibiotic.

Selective media only allows transformed bacteria to grow.

Subhash C. et al. Architecture of the Escherichia coli nucleoid. PLoS Genet (2019) 15(12)

Eukaryote Genomes



Compared to prokaryotes a thousand times larger, heavier (# genes) and complex.

When comparing genomes (size, number of genes, complexity) of eukaryotes, linearity is lost...genomic complexity does not equate to organic complexity among eukaryotes.

Ploidy, generally diploid with the exception of gametes.

Less spatial restrictions, DNA stored in the nucleus (from 10 to 200 microns in diameter).

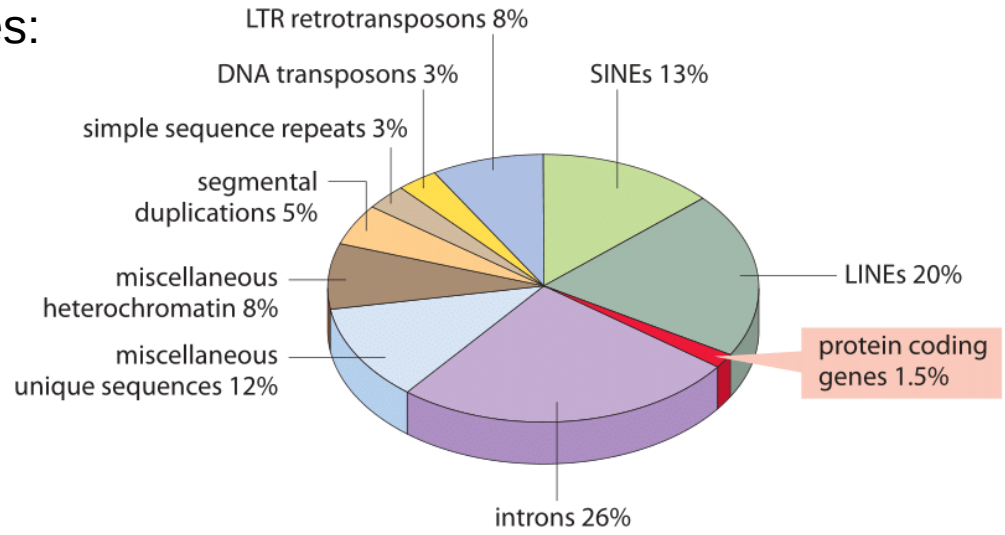
Compaction, they make use of various hierarchical levels of compaction...from nucleosomes to chromosomes.

Greater number of genes, repetitive sequences, RNA genes, sequences of unknown utility... eukaryotic philosophy allows flexibility and innovation, the prokaryote is minimalist.

Human genome

The eukaryote nuclear genome includes:

- Protein-coding genes
- Non-coding genes
- Regulatory sequences
- Junk DNA with no evident function
- Endoretroviral sequences (hERVs)
- Transposable elements



Short interspersed nuclear elements (SINEs)

- Non-autonomous, non-coding transposable elements (TEs)
- About 100 to 700 base pairs in length.
- A class of retrotransposons, DNA elements that amplify themselves in eukaryotic genomes through RNA intermediates.

Long interspersed nuclear elements (LINEs)

- Non-LTR (long terminal repeat) retrotransposons widespread in eukaryote genomes.
- Contain an internal Pol II promoter to initiate transcription into mRNA and encode one or two proteins (ORF1 or ORF2).
- ORF1 exhibits RNA/DNA binding activity.
- ORF2 is reverse transcriptase and endonuclease.
- The most abundant transposable element within the human genome.
- The only active lineage found in humans belongs to the LINE-1 class (L1Hs).
- The human genome contains an estimated 100,000 truncated and 4,000 full-length LINE-1 elements.
- Accumulation of random mutations, LINEs have degenerated and no longer transcribed or translated.

www.genome.gov/human-genome-project

Viroids

Free naked ssRNA, 300 bp than phytopathogens.

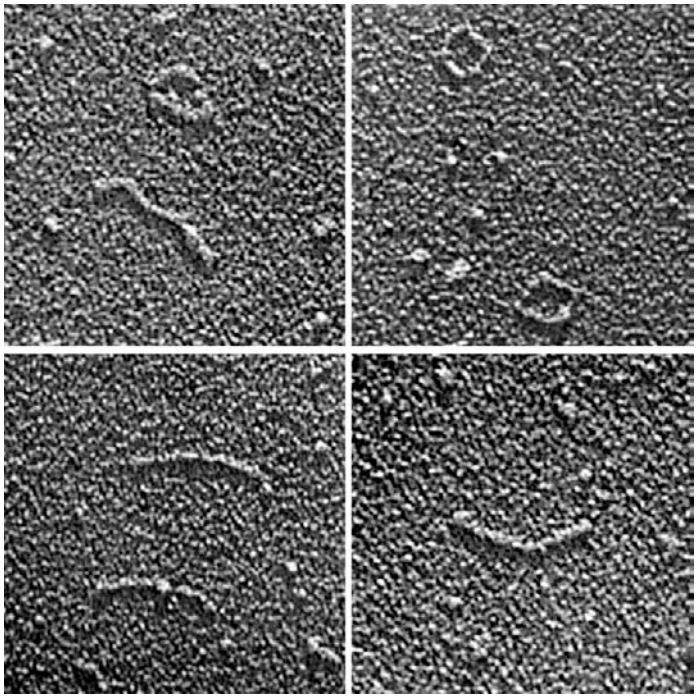
They do not code for proteins.

Best candidates for primordial genetic material.

They also accumulate mutations that can be followed phylogenetically.

Under certain periodic conditions two fragments can join together to form larger fragments (linear recombination).

Some recombinants have different pathological, invasive or interactive properties.



Subviral particles

Hepadnaviruses (HBV) are the smallest enveloped (membrane) viruses observed in animals with virion diameters of up to 42 nm.

Infected hepatocytes also secrete non-infectious Subviral Particles (SVP) that lack genetic material and are shaped like spheres or filaments.

The production of SVP exceeds 100,000 or 1,000,000 particles per cell.

We don't know much about it... Probably resulting from replicative machinery.

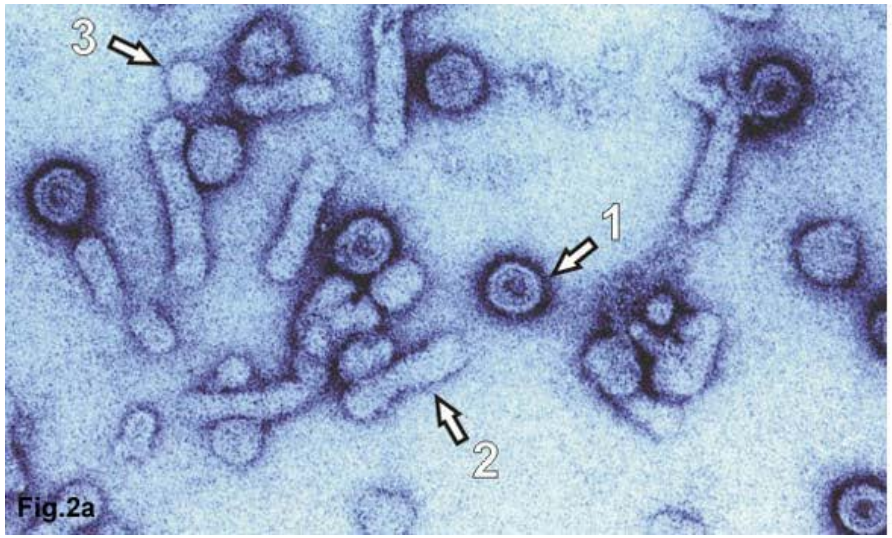


Fig.2a

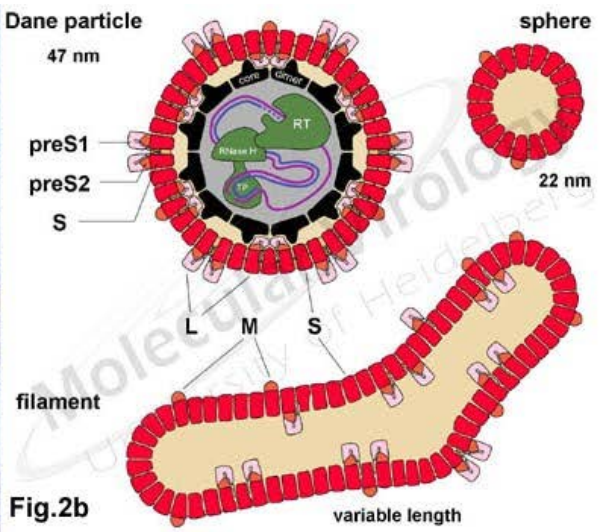


Fig.2b

Prions

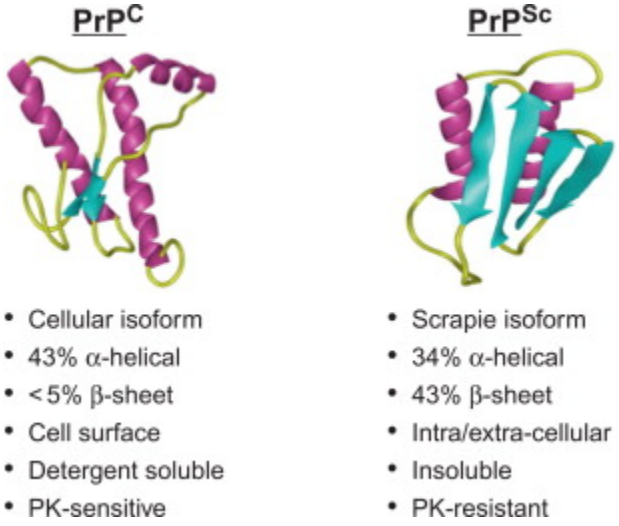
Stanley Prusiner coined the term in the early 1980s.

Neurological diseases caused by infectious agents resistant to nucleic acid destruction processes.

Initially controversial, he won the 1997 Nobel Prize in Physiology or Medicine.

Diseases called spongiform enceflaopathies due to histological appearance.

Protein particles with the ability to catalyze irreversible changes on other proethins, leading to their accumulation inside cells...which kills them.



Sanger (Dideoxy terminator) Sequencing

Method of DNA sequencing that involves electrophoresis.

Use of chain-terminating dideoxynucleotides by DNA polymerase during PCR.

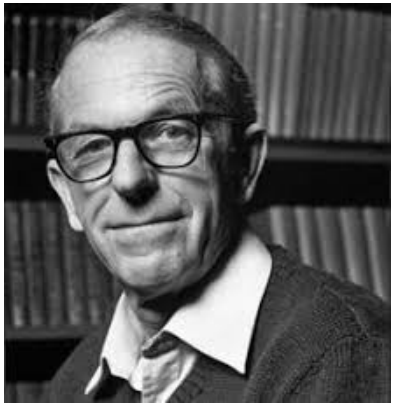
Developed by Frederick Sanger and colleagues in 1977.

The most widely used sequencing method for approximately 40 years

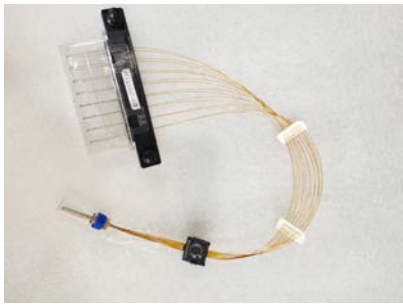
Automated instrument using slab gel electrophoresis and fluorescent labels was first commercialized by Applied Biosystems in March 1987.

Slab gels were replaced with automated capillary array electrophoresis.

Still has advantage over short-read sequencing technologies (Illumina) in that it can produce DNA sequence reads of > 500 nucleotides at very low error rates.

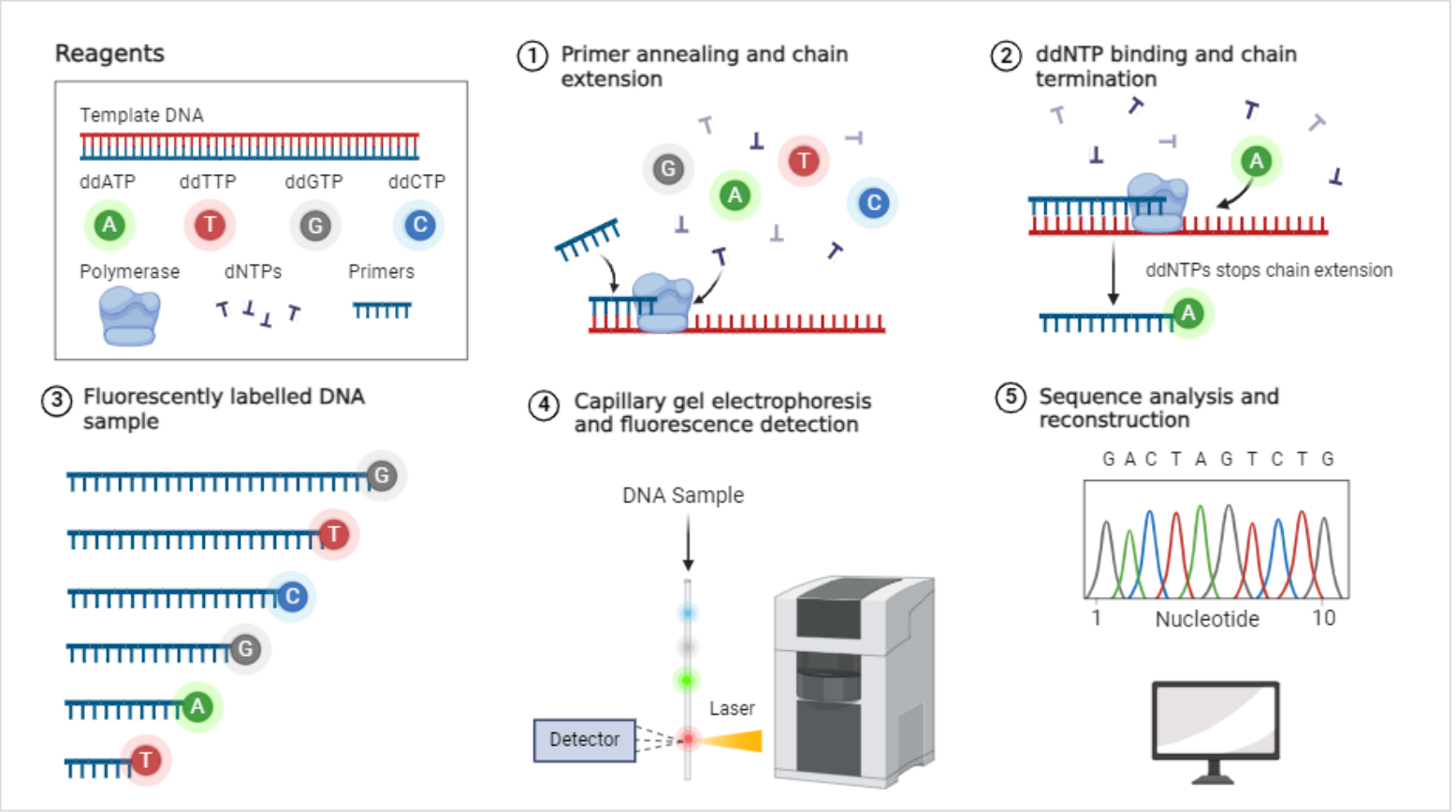


Prism 377 Applied Biosystems with banana for scale.



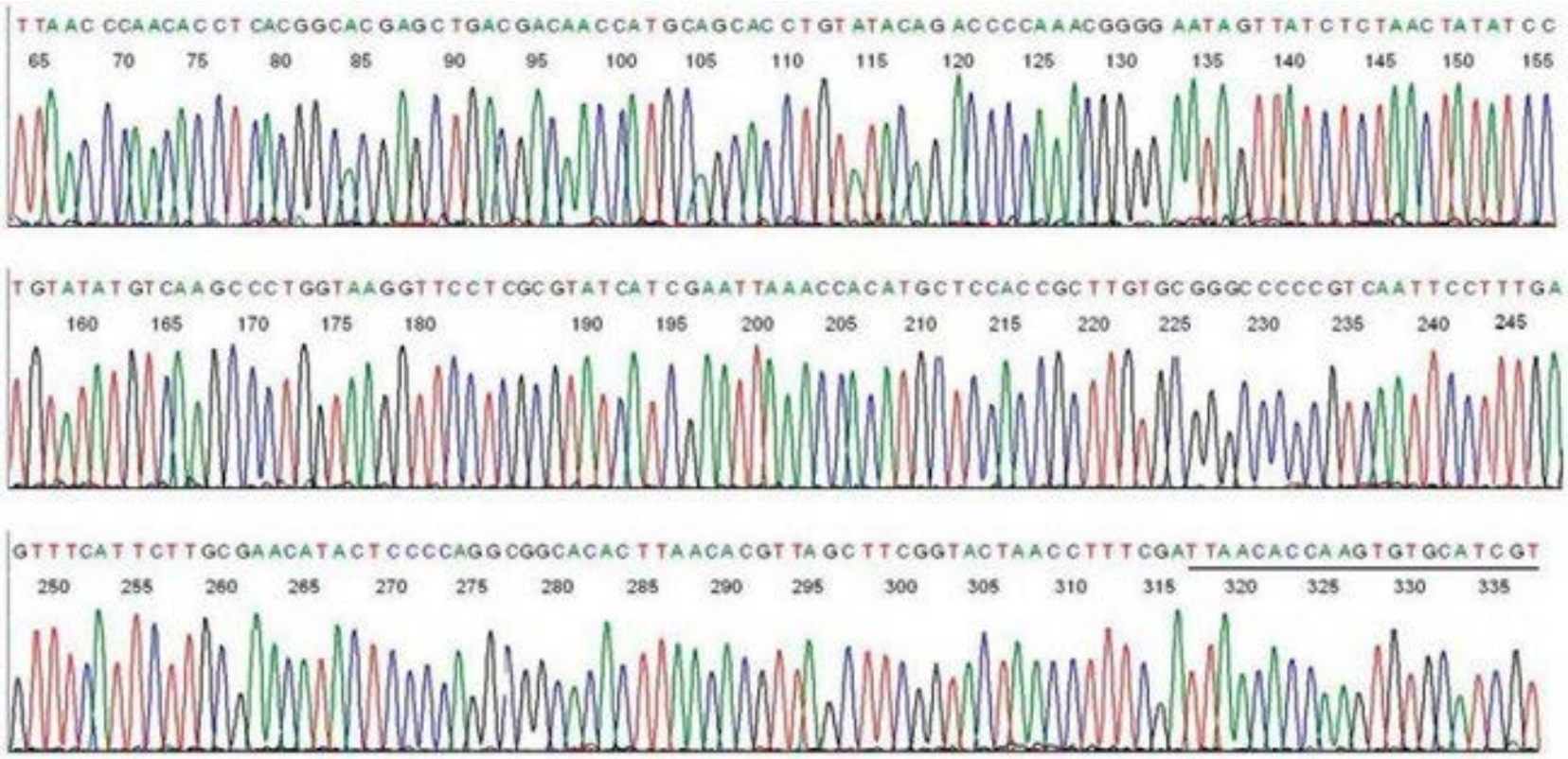
Sanger (Dideoxy terminator) Sequencing

Sanger sequencing steps



Sanger (Dideoxy terminator) Sequencing

Typical read



Next Generation Sequencing

High-throughput sequencing (HTS) methods.

AKA "next-generation" or "second-generation" sequencing (NGS) methods.

Allow entire genomes to be sequenced at once.

Fragmenting the genome into small pieces, randomly sampling fragments, and sequencing using one of a variety of technologies.

Can be "short-read" and third-generation "long-read" sequencing methods.

Applications include:

- Exome sequencing
- Genome sequencing
- Transcriptome profiling (RNA-Seq),
- DNA-protein interactions (ChIP-sequencing)
- Epigenome characterization

en.wikipedia.org/wiki/Mitochondrial_DNA

Next Generation Sequencing Technologies

Method	Read length	Accuracy	Reads per run	Time per run	Cost per 1 Bbp (USD)	Advantages	Disadvantages
Single-molecule real-time sequencing (Pacific Biosciences)	N50: 30,000 bp >100,000 bp	87%	4 million	30 minutes to 20 hours	$7.2 to $43.3	Fast	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (Ion Torrent sequencing)	600 bp	99.60%	up to 80 million	2 hours	$66.8 to $950	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (454)	700 bp	99.90%	1 million	24 hours	$10,000	Long read size. Fast.	Runs are expensive. Homopolymer errors.
Sequencing by synthesis (Illumina)	MiniSeq, NextSeq: 75–300 bp MiSeq: 50–600 bp HiSeq 2500: 50–500 bp HiSeq 3/4000: 50–300 bp HiSeq X: 300 bp	99.90%	1–25 million 130-00 million 300 million – 2 billion 2.5 billion 3 billion	1 to 11 days	$5 to $150	High sequence yield	Equipment can be very expensive. Requires high concentrations of DNA.
Combinatorial probe anchor synthesis (cPAS- BGI/MGI)	BGISEQ-50: 35-50 bp MGISEQ 200: 50-200 bp BGISEQ-500 & -2000: 50-300 bp	99.90%	50: 160 million 300 million 1300 million 375 million	1 to 9 days	$5 to $120		
Sequencing by ligation (SOLiD sequencing)	35 to 50 bp	99.90%	1.2 to 1.4 billion	1 to 2 weeks	$60 to 130	Low cost per base.	Slower than other methods. Palindromic sequences difficult.
Nanopore Sequencing	Depends on library prep up to 2,272,580 bp	92–97%	User selectable	1 min to 48 hrs	$7 to 100	Longest individual reads.	Lower throughput than other machines, Single read accuracy in 90s.
GenapSys Sequencing	150 bp	99.90%	1 to 16 million	24 hours	$667	Low-cost of instrument	
Chain termination (Sanger sequencing)	400 to 900 bp	99.90%	2	20 minutes to 3 hours	$2,400,000	Useful for many applications.	Most expensive and impractical. Ttime-consuming cloning or PCR.

en.wikipedia.org/wiki/Mitochondrial_DNA

Genomics

Genomics is an interdisciplinary field of molecular biology focusing on the structure, function, evolution, mapping, and editing of genomes.

Genetics refers to the study of individual genes and their roles in inheritance.

Genomics refers to the collective characterization and quantification of all of an organism's genes, their interrelations and influence on the organism.

Genomics also involves the sequencing and analysis of genomes through uses of high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes.

Genomics also studies intragenomic (within the genome) phenomena such as:

- Epistasis (effect of one gene on another)
- Pleiotropy (one gene affecting more than one trait)
- Heterosis (hybrid vigour)
- And interactions between loci and alleles within the genome.

History of genomic characterization



Shleiden / Schwann
Cellular biology
(1839)


Brown
Nucleous
(1833)


Mendel
Inheritance
(1866)


Miescher
Nuclein
(1869)


Levene
A, G, C, T
(1910)


Morgan
Cromosomas
Drosophila
(1911)

Mendel (de Vries)
Rediscovery
(1900)


Avery-MacLeod-McCarty
Transforming principle DNA
(1944)


Astbury
Crystallography
(1943)


Griffith
Transforming principle
(1928)


Chargaff
A=T
(1948)


Hershey-Chase
Phage
(1952)


Franklin-Wilkins
DNA structure
(1952)


Kornberg
DNA Pol
(1955)


Watson-Crick
DNA 3D
(1953)


Cohen & Boyer
Cloning
(1973)


Sanger
Sequencing
(1975)


Mullis
PCR
(1983)

Phage FX174
(1980)

DOE HUGO
(1986)

H. influenza
(1995)

S. cervisiae
(1996)

C. elegans
(1998)

Chrom. 22
(1999)

D. melanogaster
(2000)

H. sapiens
(2001/3)

1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030

History of genomic characterization

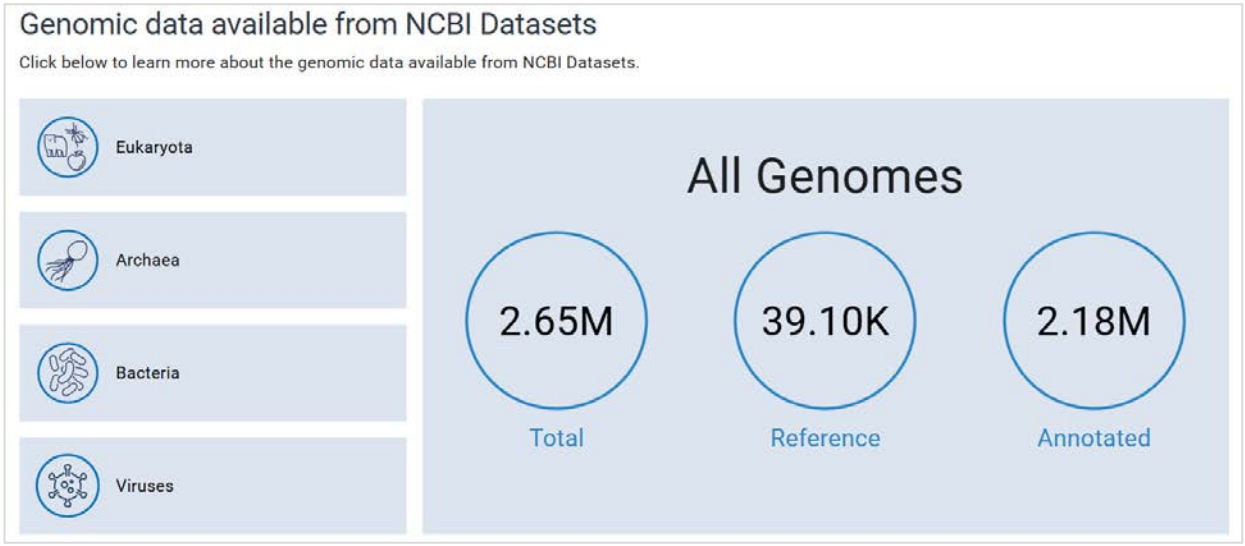
Mitochondrion, the first complete genome sequence (16,568 bp) in 1981.

Chloroplast genome sequenced in 1986.

S. cerevisiae Chr III (315 kb) first eukaryotic chromosome sequenced in 1992.

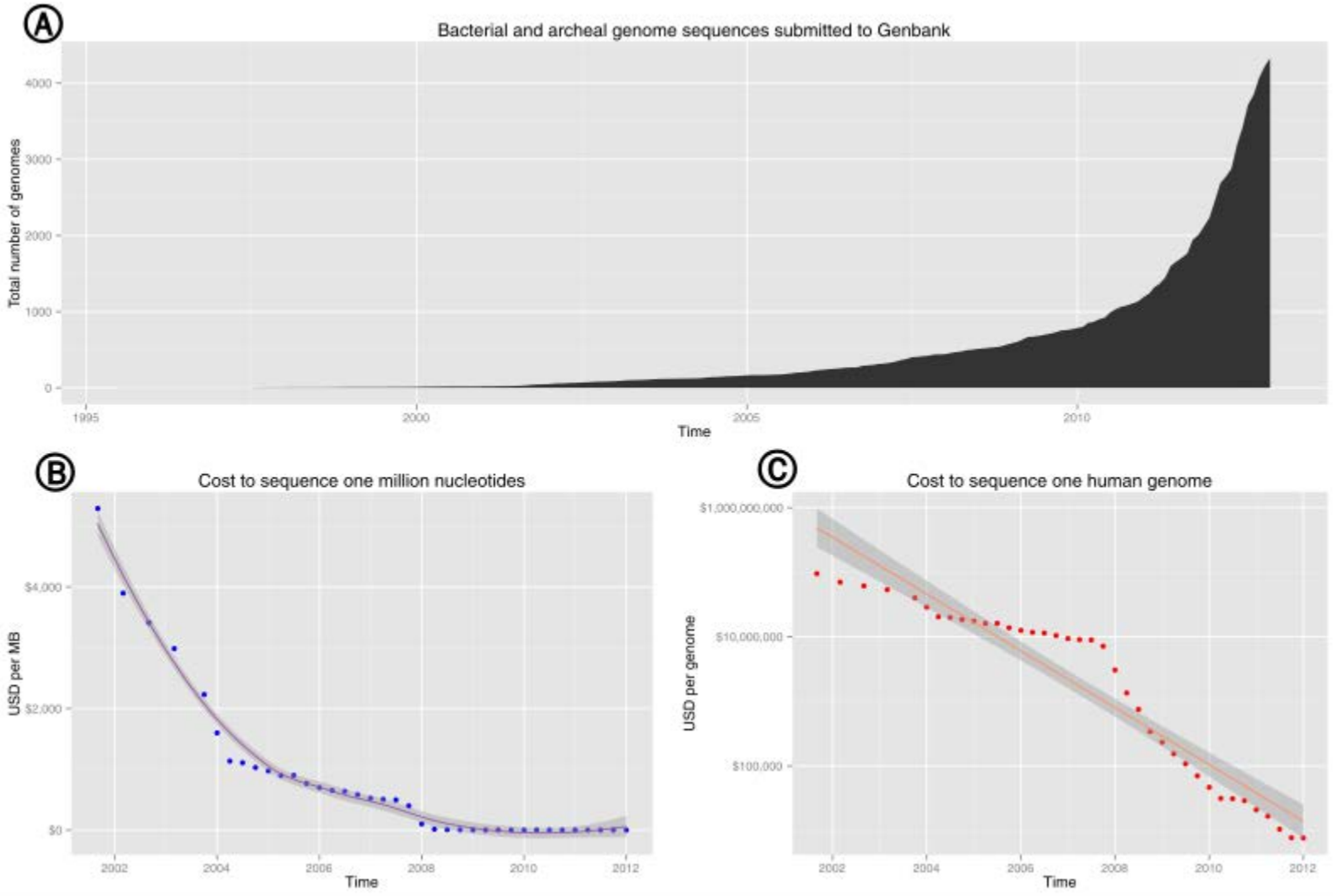
Haemophilus influenzae, first free-living organism sequenced (1.8 Mb) in 1995.

S. cerevisiae first eukaryotic genome sequenced in 1996.



www.ncbi.nlm.nih.gov/datasets/genome/

Evolution of genome sequencing and cost



The Human Genome Project

Started in October 1990

First draft reported in April 2003 (92%), 8% consisting of repetitive sequences.

First complete Telomere-to-Telomere (T2T) human genome sequence in March 2022.

Human Genome Project made every part of the human genome sequence publicly available.

The “Bermuda Principles” set out the rules for the rapid release of sequence data.

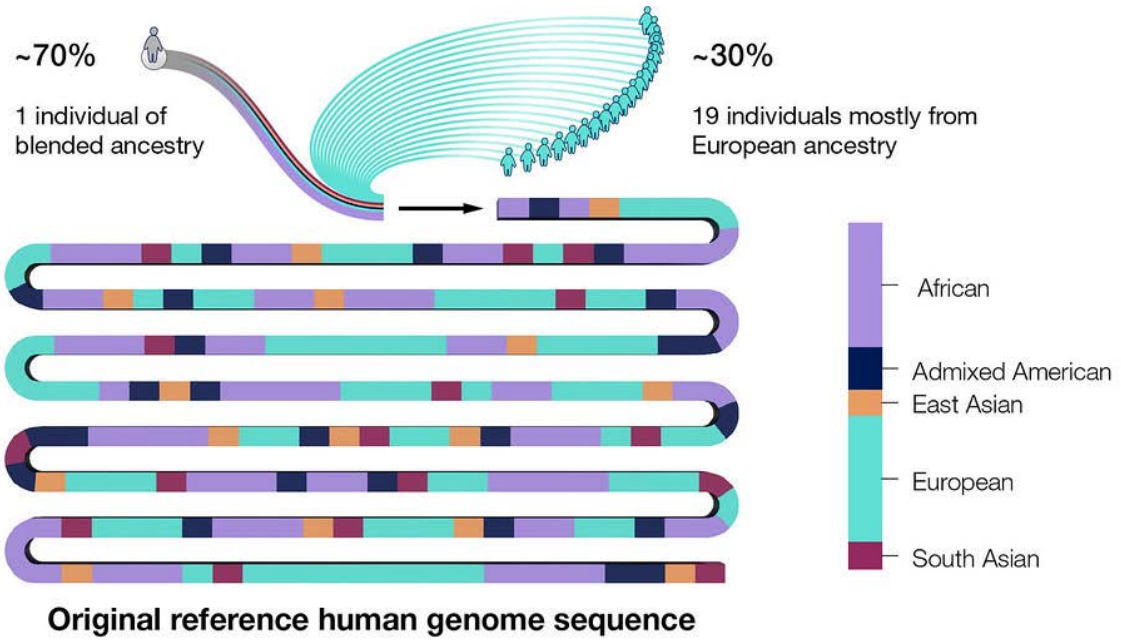
www.genome.gov/human-genome-project

The Human Genome Project

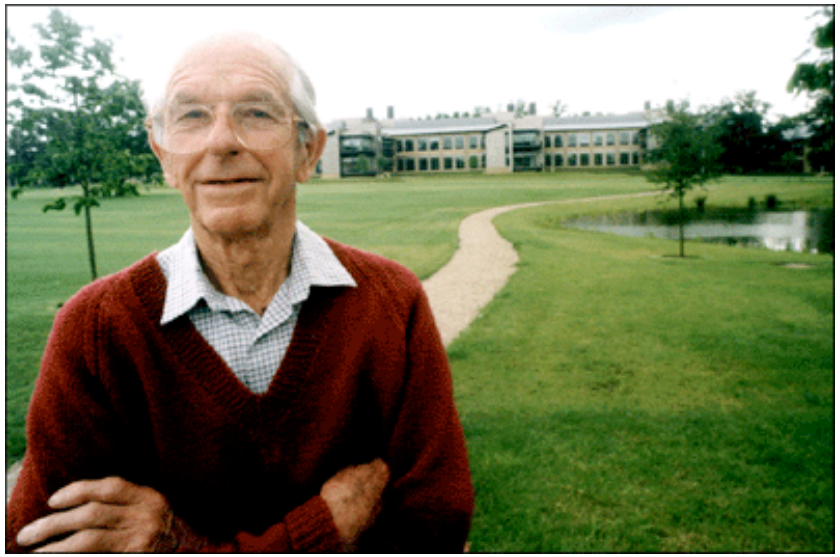
Actually, a patchwork of multiple people whose identities were anonymous.

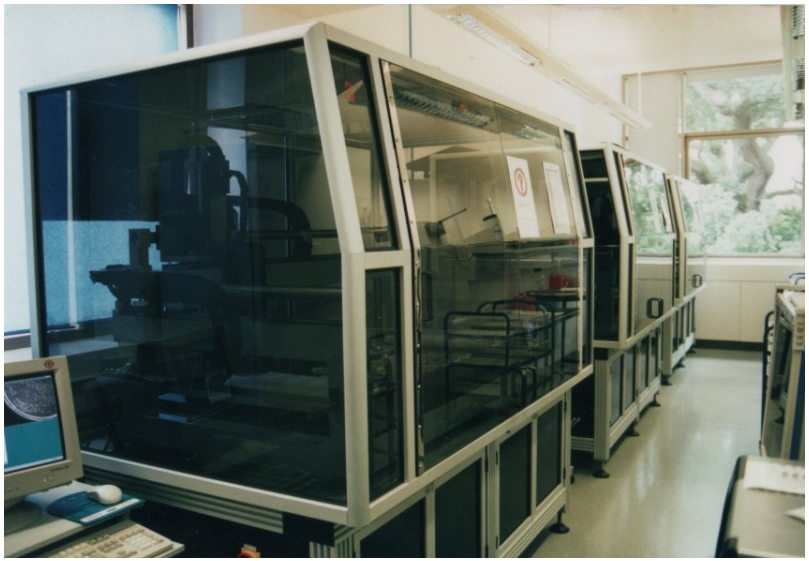
However, the majority of the sequence came from one person of blended ancestry.

70% of the reference human genome sequence was generated from that single individual's DNA, with the remaining 30% coming from a combination of 19 other individuals of mostly European ancestry.



The Human Genome Project





The Human Genome Project

Anti-Human Genome Project Campaign

Led by Dr. Martin Rechsteiner in the early 1990s; 55 scientists from 33 institutions wrote to NIH Acting Director Dr. William Raub opposing the HGP.

Questioned its $3 billion cost, timeline, and healthcare relevance.

Highlighted scepticism about whole-genome sequencing vs. targeted approaches.

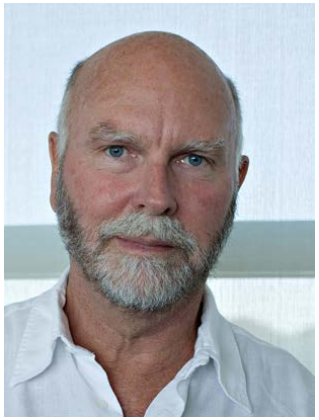
Raised enduring questions about genomics' role in healthcare.

Dr. Elke Jordan of the National Center for Human Genome Research defended HGP.

www.genome.gov/human-genome-project

Celera Genomics Private Initiative

In 1992 J. C. Venter founded The Institute for Genomic Research (TIGR) and in 1998 he founded Celera in association with ABI.

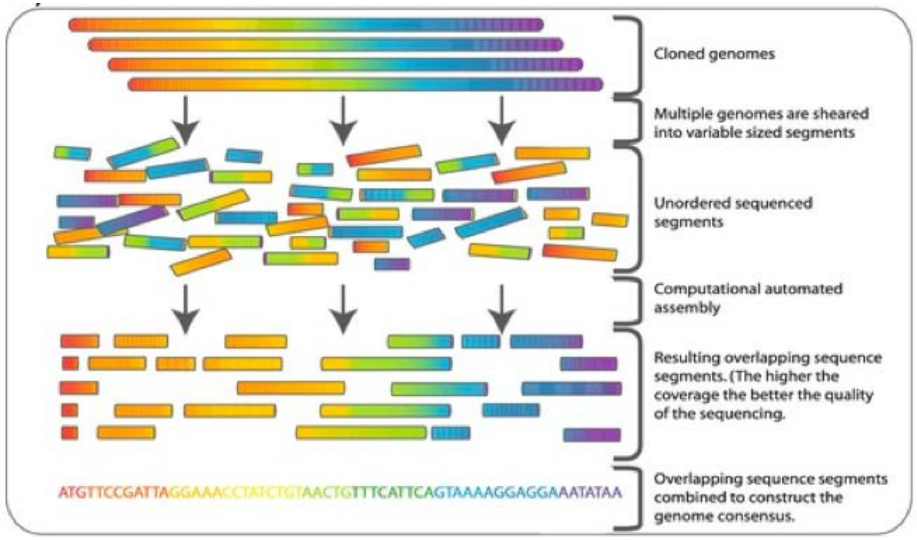


ABI provides a little more than 300 96 capillary sequencers.

Their intention is to patent genes.

Celera means speed.

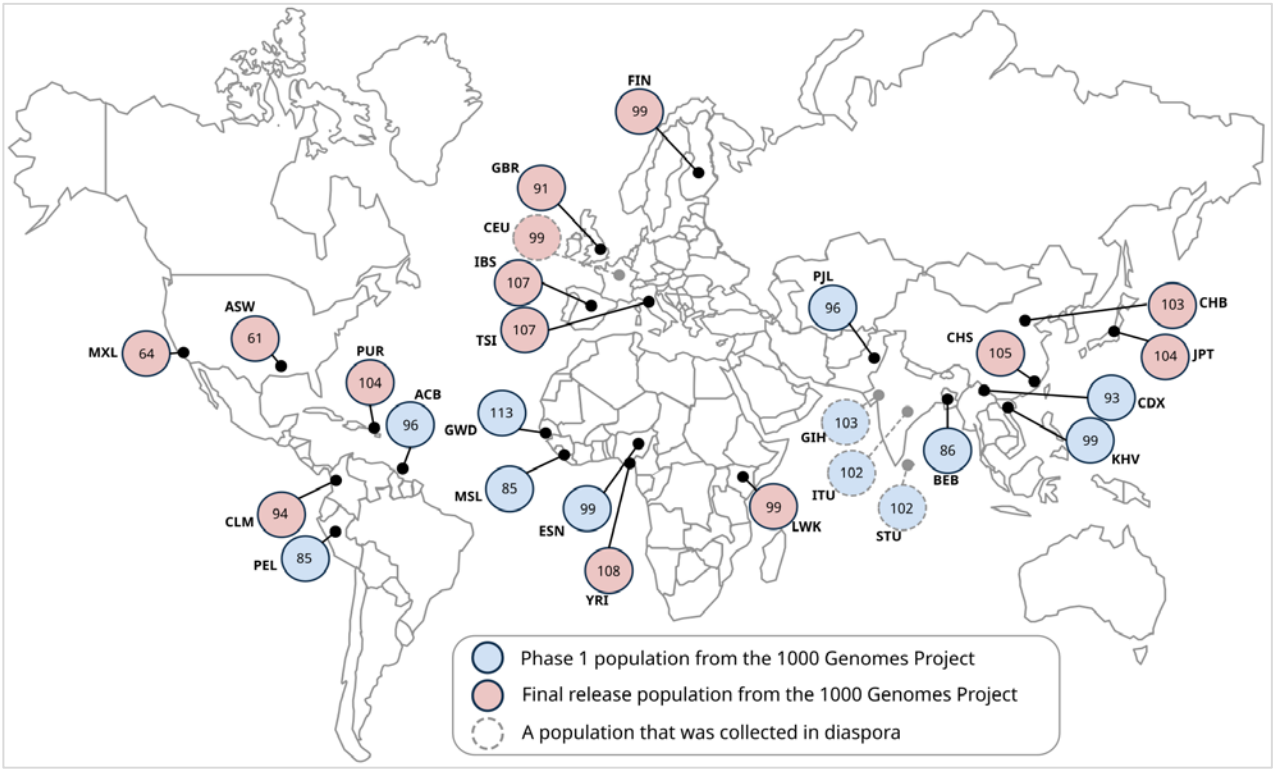
Shotgun method



The 1000 Genomes Project (1KGP)

Conducted from 2008 to 2015 to create a detailed catalog of human genetic variation.

Sequenced genomes of over 1,000 anonymous healthy participants from diverse ethnic groups.



The 1000 Genomes Project Consortium. 'A global reference for human genetic variation.' Nature, 1 October 2015.

The 1000 Genomes Project (1KGP)

Involved multidisciplinary teams from China, Italy, Japan, Kenya, Nigeria, Peru, the UK, and the US.

- 2010 Pilot phase completed, results published in Nature.
- 2012 Sequencing of 1,092 genomes announced.
- 2015 Project completion and findings published.

Identified many rare genetic variations and analyzed 8 variation classes.

Produced a refined human genome map accessible through public databases.

Data hosted by the International Genome Sample Resource to support ongoing research.

https://www.internationalgenome.org/

The 1000 Genomes Project Consortium. 'A global reference for human genetic variation.' Nature, 1 October 2015.

The 1000 Genomes Project (1KGP)

Involved multidisciplinary teams from China, Italy, Japan, Kenya, Nigeria, Peru, the UK, and the US.

- 2010 Pilot phase completed, results published in Nature.
- 2012 Sequencing of 1,092 genomes announced.
- 2015 Project completion and findings published.

Identified many rare genetic variations and analyzed 8 variation classes.

Produced a refined human genome map accessible through public databases.

Data hosted by the International Genome Sample Resource to support ongoing research.

The 1000 Genomes Project Consortium. 'A global reference for human genetic variation.' Nature, 1 October 2015.

International HapMap Project

Developed a haplotype map (HapMap) of the human genome to identify genetic variants influencing health, disease, and responses to drugs/environmental factors.

Data made freely available for research.

- Began in October 2002 and spanned three phases:
- Phase I (2005) Initial dataset published.
- Phase II (2007) Extended dataset released.
- Phase III (2010) Final results published.

Global collaboration including Canada, China, Hong Kong, Japan, Nigeria, the UK, and the US.

An informatics audit revealed security flaws in the legacy HapMap site that required NCBI to take it down immediately.

NCBI was planning to decommission this site in the near future anyway as the 1,000 genomes (1KG) project has established itself as a research standard.

www.genome.gov/10001688/international-hapmap-project

Mexican Biobank Genome-Wide Association Study (GWAS)

Latin America remains underrepresented in genomics research.

Genotyped 6,057 individuals from 898 localities across Mexico using 1.8 million genome-wide markers linked to trait and disease data.

Ancestry deconvolution revealed Indigenous, colonial, and postcolonial demographic dynamics across Mesoamerican regions.

Variations in runs of homozygosity among genomic regions highlighted distinct demographic histories and distributions of rare deleterious variants.

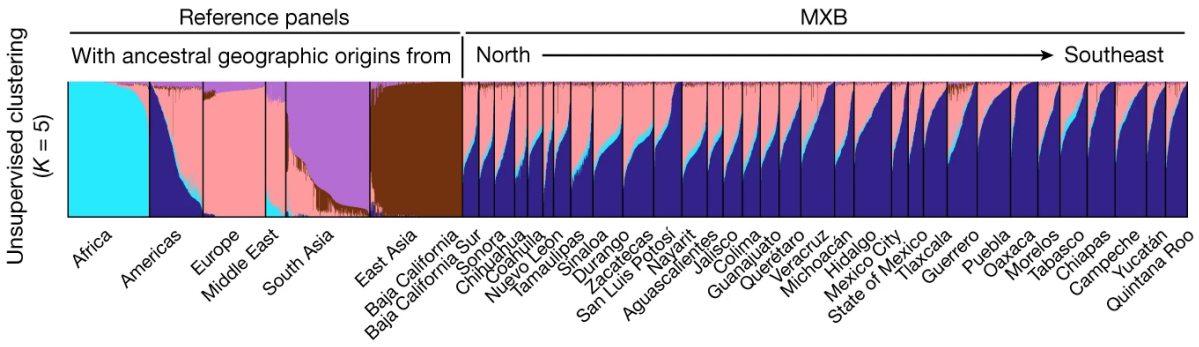
GWAS for 22 traits showed better predictions using the Mexican Biobank compared to the UK Biobank.

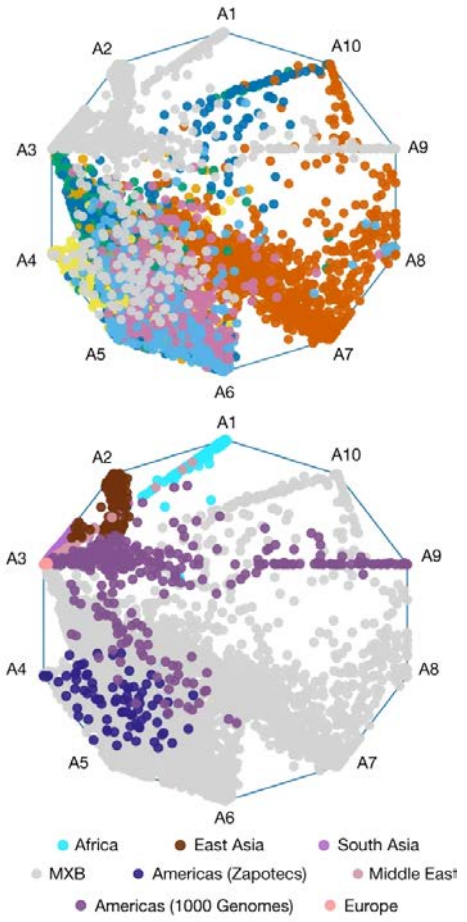
The study advances understanding of Mexico's genetic history and trait architecture, supporting precision and preventive medicine globally.

www.nature.com/articles/s41586-023-06560-0.pdf

Mexican Biobank Genome-Wide Association Study (GWAS)

Mexico regionalized into Mesoamerican regions according to anthropological and archaeological context.





www.nature.com/articles/s41586-023-06560-0.pdf

How to read the genome and build a human

Viral & Human Genomics Laboratory

CDC Commissioned Biosafety Level 3 (BSL-3) High Biocontainment Facility

Laboratorio de Genómica Viral y Humana

Instalaciones de Alta Contención Biológica Nivel de Bioseguridad 3 (BSL-3) CDC-certificadas

Facultad de Medicina UASLP
San Luis Potosí, México

www.genomica.uaslp.mx

GenomicaUASLP

YouTube GenomicaUASLP