

NGS Analysis for Monogenic Disease in African Populations

Human Genetics Basics Part 1

Segun Fatumo PhD

Associate Professor, Department of Non-Communicable Disease Epidemiology, LSHTM, UK
Head, The African Computational Genomics Group, MRC/UVRI and LSHTM, Uganda

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



segun.fatumo@lshtm.ac.uk

Outlines

Part 1

- How did we get here ? (Segun)
- Human Genomics Project till Date (Segun)
- African genomics (Segun)
- Basics of human genetics and terminologies (Valentina)

Part 2

- Brainstorming – Two examples (Segun)
- Different modes of inheritance and Mendelian inheritance (Valentina)
- Locus and/or allelic heterogeneity, compound heterozygosity, consanguinity, homo/heterozygosity (Valentina)



From DNA to Genome

Until the early 1970's, DNA was the most difficult cellular molecule for biochemists to analyze.

The term “genetics”
William Bateson

Watson and Crick
DNA model

Sequence
alignment

PDB (Protein Data
Bank)

GenBank
database

1906

1955

1960

1965

1970

1975

1980

1985

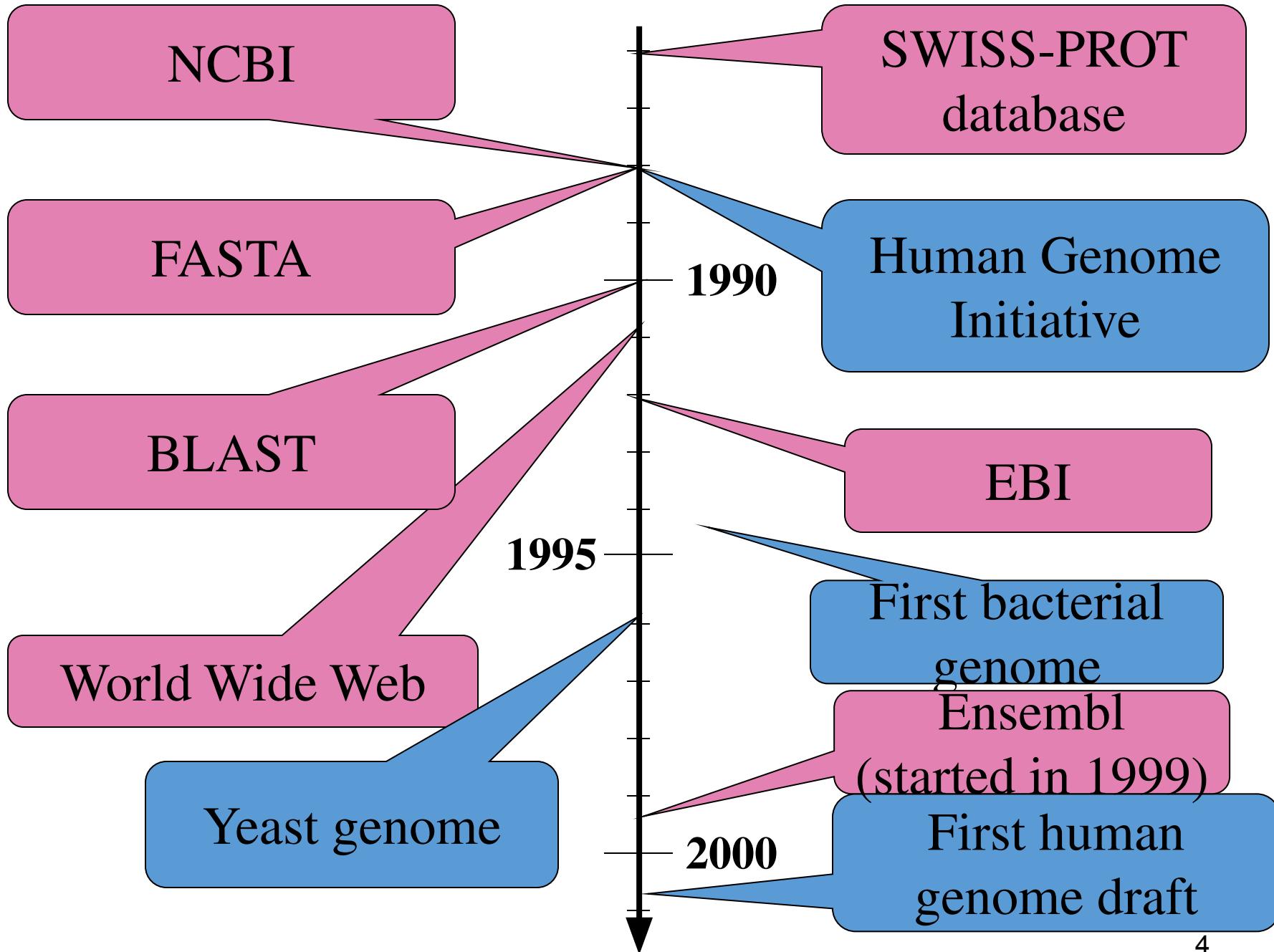
Sanger sequences
insulin protein

Dayhoff’s Atlas

ARPANET
(early Internet)

Sanger dideoxy DNA
sequencing

PCR (Polymerase
Chain Reaction)



The Human Genome Project

(is one of the greatest scientific feats in history)

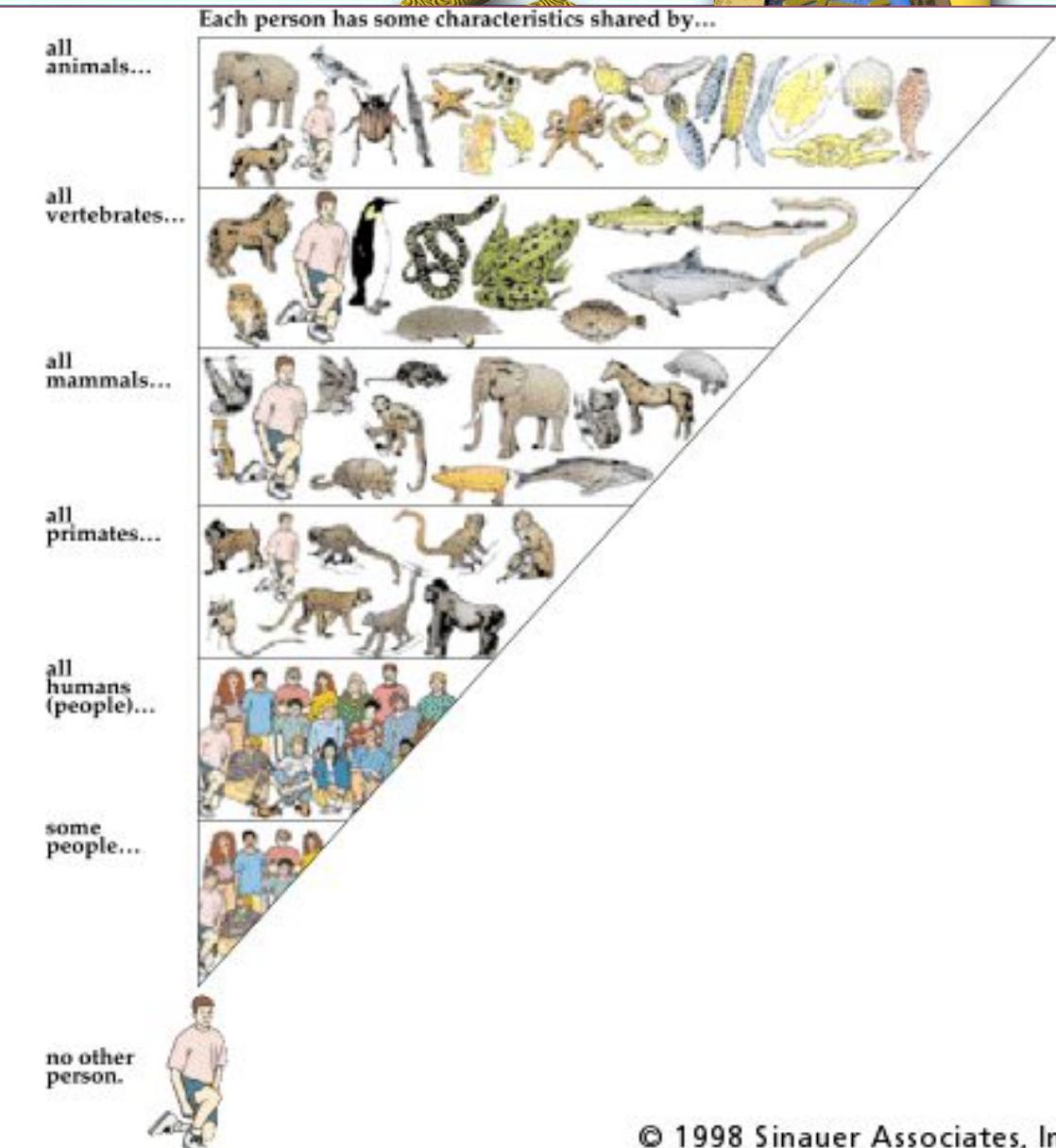


The Human Genome

Trillions of cells

Each cell:

- 46 human chromosomes
- 2 meters of DNA
- 3 billion DNA subunits (the bases: A, T, C, G)
- Approximately 21 000 genes code for proteins that perform most life functions



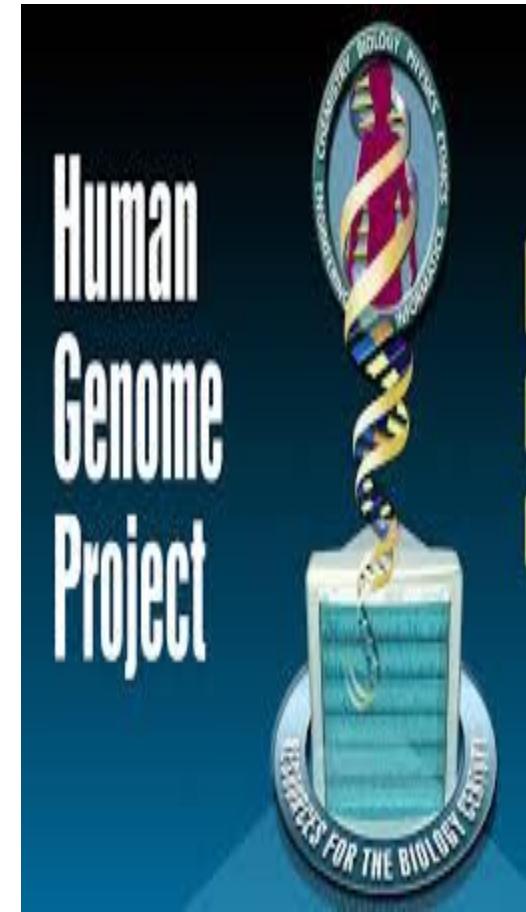
Human Genome Project

Goals:

- identify all the approximate 20,000 genes in human DNA,
- determine the sequences of the 3 billion chemical base pairs that make up human DNA,
- store this information in databases,
- improve tools for data analysis,
- transfer related technologies to the private sector, and
- address the ethical, legal, and social issues (ELSI) that may arise from the project.

Milestones:

- 1990: Project initiated as joint effort of U.S. Department of Energy and the National Institutes of Health
- June 2000: Completion of a working draft of the entire human genome (covers >90% of the genome to a depth of 3-4x redundant sequence)
- February 2001: Analyses of the working draft are published
- April 2003: HGP sequencing is completed and Project is declared finished two years ahead of schedule



The Completion of the Human Genome Sequence

- June 2000 White House announcement that the majority of the human genome (80%) had been sequenced (working draft).
- Working draft made available on the web July 2000 at genome.ucsc.edu.
- Publication of 90 percent of the sequence in the February 2001 issue of the journal *Nature*.
- Completion of 99.99% of the genome as finished sequence on July 2003.

26 June 2000



Human Genome Project: The Competition

- In 1998, Craig Venter, founder of the Institute for Genomic Research in Maryland, USA, announced that he had formed a new private company (later to become Celera Genomics) to take on the task of human genome sequencing
- Craig felt that the Human Genome Project was taking too long, proving too costly and that it was getting bogged down by non-essential discussions, such as who was going to take credit for it.



- Bill Clinton wanted biotech entrepreneur Craig Venter (left) and Francis Collins (centre) of the US National Institutes of Health to patch up their differences. Credit: Ron Sachs/Shutterstock

Human Genome Project

WHAT

- DoE/NIH project, formally launched 1990, completed 2003
- \$3 billion, 1000 scientists, 50 countries
- Primary goals:
 - Sequence a human genome
 - Identify and map all human genes (physical and functional)
- ELSI issues (5%)

WHY

- Better understand genetics and disease
- Find disease-causing variants
- Identify new treatments
- Enable comparative genomics/evolutionary studies, functional genomics
- Technique development



Only ~3% of genome codes for protein
Remainder is regulatory or of unknown function (junk)
e.g repetitive sequence, possibly viral DNA

Humans v. humans

Humans are, on average, 99.9% identical to every other human...



...but $0.001 * 3 \text{ billion} = 3 \text{ million points of variation.}$

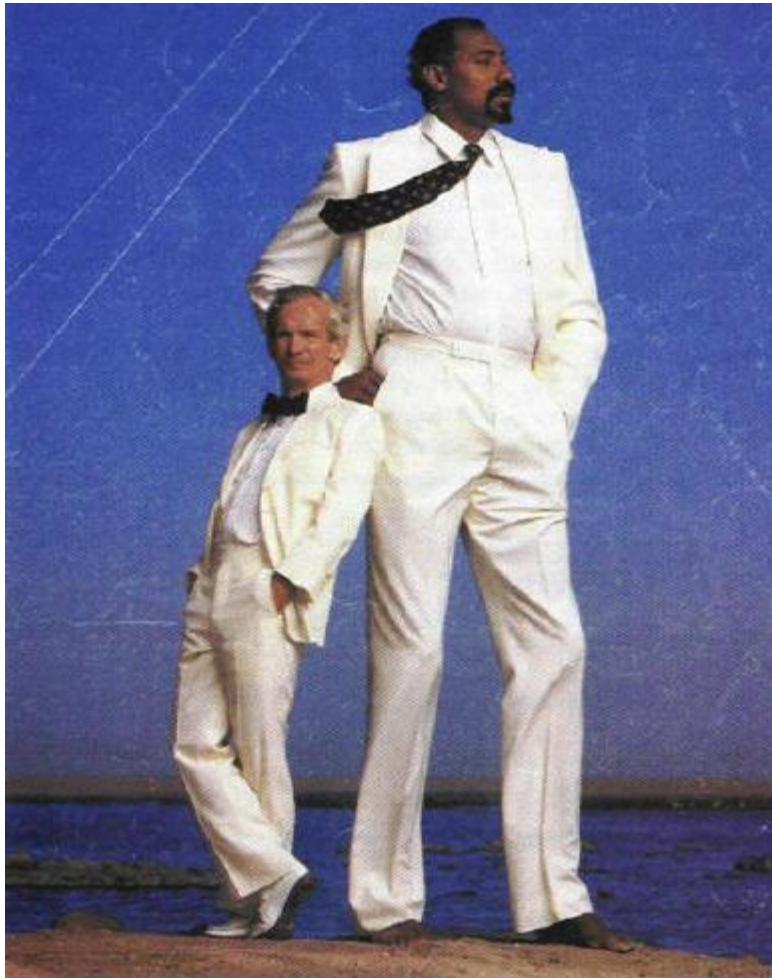
These differences are what make us unique.



Some Facts

- In human beings, 99.9% bases are same
- Remaining 0.1% makes a person unique
 - Different attributes / characteristics / traits
 - how a person looks
 - diseases he or she develops
- These variations can be:
 - Harmless (change in phenotype)
 - Harmful (diabetes, cancer, heart disease, Huntington's disease, and hemophilia)
 - Latent (variations found in coding and regulatory regions, are not harmful on their own, and the change in each gene only becomes apparent under certain conditions e.g. susceptibility to heart attack)

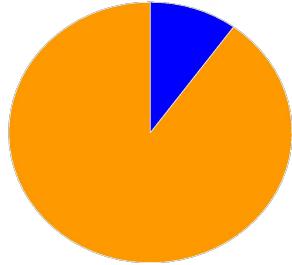
Genetic variations underlie phenotypic differences



Wilt Chamberlain,
a famous NBA basketball player
(7 feet, 1 inch; 275 pounds)

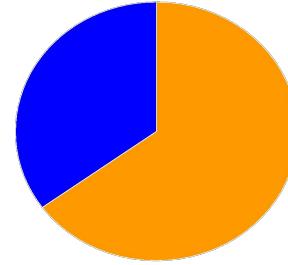
Willie Shoemaker,
a famous horse racing jockey
(4 feet, 11 inches;
barely 100 pounds).

Genetic variations cause inherited diseases



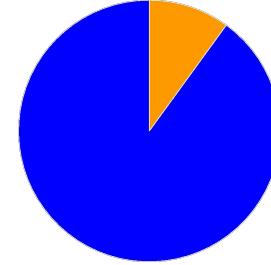
Genetic Diseases

- Cystic fibrosis
- Down syndrome
- Sickle cell disease
- Turner syndrome



Complex Diseases

- Alzheimer disease
- Cardiovascular Disease
- Diabetes (type 2)
- Parkinson Disease



Environmental Diseases

- Influenza
- Hepatitis
- Measles

- Environment - Genes

Humans v. other organisms

Percent of genes shared between humans and...

Chimp: 98%

Gorilla: 96%

Mouse: 92%

Fruit Fly: 44%

Yeast: 26% Plants: ~18%

National Academy of Science

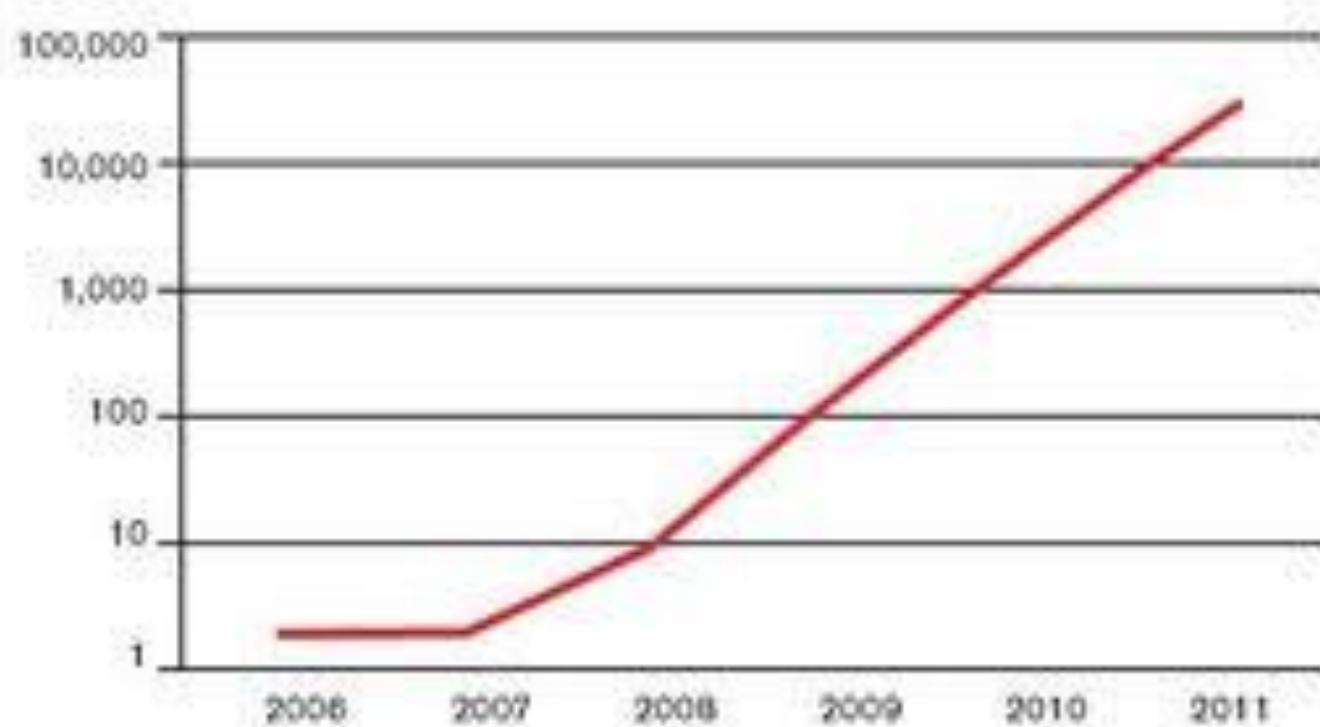


www.sciencemag.org

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



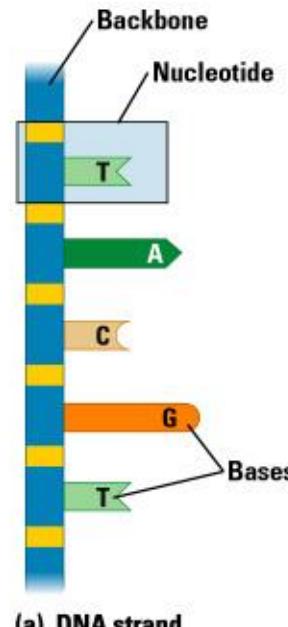
Number of Human Genomes Sequenced



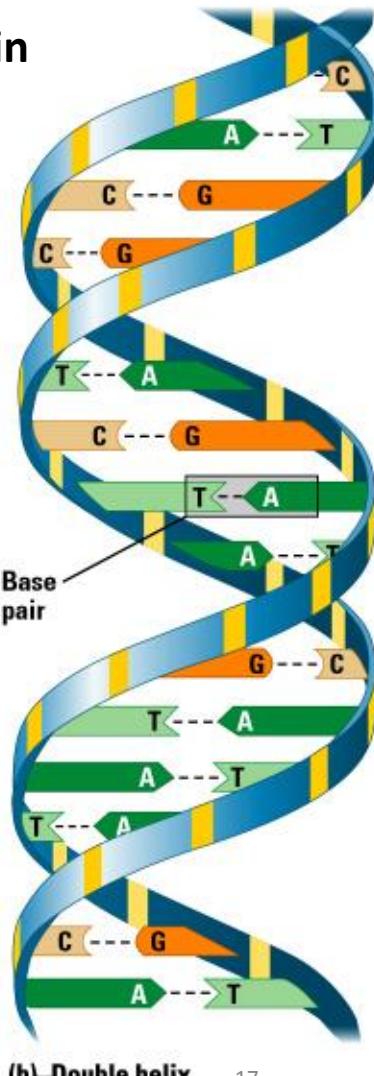
The Genome is Who We Are on the inside!

- Chromosomes consist of DNA
 - molecular strings of A, C, G, & T
 - base pairs, A-T, C-G
- Genes
 - DNA sequences that encode proteins
 - less than 3% of human genome

Information coded in DNA



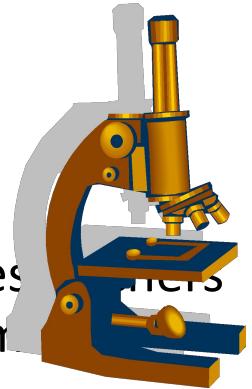
(a) DNA strand



(b) Double helix

How do scientists find genes?

- The genome is so large that useful information is hard to find.



- **computational microscope** to help scientists search the genome.

- Just as you would use “google” to find something on the internet, researchers can use the “**UCSC Genome Browser**” to find information in the human genome.

<http://genome.ucsc.edu>

5000 bases per page

CACACTTGCATGTGAGAGCTTCAATATCTAAITTAATGTTGAATCATTATTAGAACAACAGAGAGCTAACTGTTATCCCATCCGTACTTTATTCTTATG AGAAAAAATACAGTGATTCC
AAGTTACCAAGTTAGTGTGCTGCTTATAAATGAAGTAATATTAAAAGTTGTGCATAAGTTAAAATTAGAACAATAAAACTTCATCCTAAAACCTGTGTGTTGCTTAAATAATC
AGAGCATCTGC TACTTAATTGGTGTGGGTGCACAATAGATGTTAATGAGATCCTGTCATCTGTCTGTTTTTATTGTAAAACAGGAGGGTTTAATACTGGAGGAACAA
CTGATGTACCTGAAAAGAGA AGAGATTAGTTAATGATTGAGGGTTGTCTGTCTTAGTACGTTTATTCTCTAGGTACTATTGATTGTAAAATAGAATTATC
CTCATTAAATGAAAATCACACAGGAGAACAGAAAACCTATGAGATAGATGAACGTTGTGAGTGGCATGGTTAATTGTTGAGAAGAAGCAGTGCCTGCCAGAACATACAAAT
GAAATTCACTGTTATTGAGTAGAGTAGAATACAGTGTGTTCCCTGTGAAGTTCATAACCAAGAATTAGTAGTGGATAGGTAGGCTGAATAACTGACTTCCTATC ATTTTCAGGTT
CTGCGTTGATTTTTACATATTAAATTCTTGATCCACATTAAGCTCAGTTATGTATTCCATTAAATGAAAAAAATAGGCACGTGCAAATGTCAGATCACTGCCTGTGGT
CATTGGGTAGAGATTGTGGAGCTAAGTGGCTTAATCAAATGTCAGCTTTTTTTCTTATAAAATAGGTTTAATATGAGTTTAAATAAAATTAGAAAAGGCAA
ATTACTCAATATATAAGGTATTGCAATTGTAATAGGTAGGTATTCTAGTTATGGTGGGATATTACAGACTATAATTCCAATGAAAAAAACTTAAAAATGCTAGTGA
TTGCACACTTAAACACCTTTAAAAGCATTGAGAGCTTATAAATTAAATGAGTGTAAAACCAAAATTGAGAGAAAAGAAGAACCAGAGAGGTAAGGATAACCTTAC
AGTGCAATTGCCGATCTCTACAAATATTAAATTGACAGTTCAGGTGAATGAGAAAGAACCAAACCAAGACTAGCATATGTTGTCTTCAAGGAGCCCTCCCT
AAAAGATTGAGATGACCAAATCTTAACTCTCAGCATAAGGTGAACCAAGACAGACCTAAAGCAGTGGTAGCTGGATCCACTACTGGGTTGTGTCGGCTGACTCAGGTAACTC
CAAGAATTGACATTTTTAAGGTGGCTTACTCATACTGCCAGGTATTAGGGAGAACAAATCTGAATGCTTATAAAATACCCATAAGCTAAATCTTACAAATATTCTCAAG
AACACAGTGA ACAAGGCAAATAAGTAAAATCAACAAAACATGAACACATAATTAGACACACAAAGACTCAAACATTGAAATACCCAGAGAAAGATAATAAATAT
TTTACTCTTAAAATTAGTTAAAGCTTAAACTAATTGAGAGAAA AACTATGTTAGTATTATATTGAGTGAATAAGCAAACATTAAATACAGGTACTAGAATTCTTCAATTGAGATGTA
TAAATATAATAGATAATTACCAACAGATTAGTACCTGAAAGGATAATTAAATACTGAAATACAGGTACTAGAATTCTTCAATTGAGATGTA AAAAATGAAA
TTAATGCAAAAATAAGGGCACAAAAGAAATGAGTAATTGATCAGAAAATGTTAAATAAAATGAAATTGCAATTAAAAGCATTGTCATCCAAGTAGATG
TCTATTAAATAGTTCTCATATCCAGTAATGTAATTATTATTCCCTCTCATGCACTTCAGATTCTGGGTAATCTTAGACATCAGTTGTCTTTATTATTCTTACTAC
ATTATTATTGCAATGATATTTTAATTCTGACATTCTGGAGATTGCTGTAAGGTTATTAAATACTTATGTTATTCTTACTCATAGATAGGATATTCTTATTATT
ATTGACATTCTTGGTTCTCTGTACTCTTCTAGTGTATTATCTTACTCATAGATAGGATATTCTTATTATTAAATATTGGTATTCTGGTTCTCAGCC
ATCTATTGTCAGTGTCTTATTAGCATTATTATAAAATAAGATTATTCTCTAAACATGAGAATCTTATTCTCCCAGTAATTGAAATTGCAATGCCATCTGCCATG
TACAGCATGGTTGGCTGTTCTCTTAACTTTATTAGTTGGAGTACCTGTGAAAGTTGTTATAGGTAACCTGTGTCACCAAGGGTTGTACAGATCA
TTTGTACCTAGGTACCAAGTACTCAACAATTATTCTGCTCTGTCTCCTGTCACCCACTCTCAAGTAGACTCCGGTGTCTGTTCCATTCTGTGTCATGTTCT
ATAATTAGTCCCCACTTGTAAAGTGAAGAACATGCACTATTGTTTTGTTCTGTGTTAAATTGCCAGTATAATGCCCTCAGCTCATGTTACTGCAAGAA
CATGATCTCATCTTATTAGCTCATGGTGTATATACCAACTTCTTATCTAAACTCTTATTGATGAGCATTGAGGTGGATTCTATGTCATTGTCATATTGTC
AACATTGTCATGTGTTATGGTAGATGATATTCTCTGGGTATATGCACTGAAATGCGATTGCTGGTTGGAATGGTAGTTCTGCTTTATCTTGTGAGGAATTGCCATG
CTGTTCCACAATAGTGAACTAACTACCTCCACTAACAGTGTCAAGTGTCTTCTCCACAACCTGCCAGCATCTGTATTGACATTAAATAGTGCATTAACT
GGTATGAAATTATATTCTATTGTTTAATTGCAATTCTCTAAATGATCAGTGTATTGAGTTGTTTTCACATGCTGTGGCTGCATGTATGTCCTTAAAGTGTCTGTT
CATGACTTTGCCACATTAAATGGGTGTTCTCTGTAAATTGTTAAATTCTTATAGGTGCTGGATTAGACATTGTCAGACGCATAGTTGCAAAATGTTCTCCATT
TGTAGGTTGTGTTATTGTTAATAGTTCTTGTATGCAAGCTTAAATAGTTAATGAGATCTGATATGTTAGGCTTGTCCCCACCAAATCTCATCTGAATTATA
TCTCCATAATCACACATGGAGAGACCAGGTGGAGGTAATTGAATCTGGGGTGTGTTACCCATGCTGTTGTGATAGTGAATGAGTTCTCAGGAGATCTAATGGTTTATGAGG
GGCTCTCCAGTTGCCGTGACTCTCTCTGCCGCTTGTGAAAAGGTGCAATTGCGCCCTTACCTCTTATAATTGTAAGTTCCTGAGGCCTCCAGCATG
CTTCAAGTCATTAACCTTTCTTATAAATTACTCAGTCTGTTCTTATAGCAGTGTGAAAATGGACTAATGAAGTCCCATTGAAATTGCTTGTGCAATTG
TTGACATCTTGTGAAATCCTGCCGTCTAAGTACAGGACGGTATTGCCAGGGTTTCTAATTGTTGTTGCAATTAGTGTAAATCCATCTGAGTTG
TTTTGTATATTGTAAGGAAGGGTCCAGTTCAATCTTGTGATAGGCTAGTTAGTATTCCAGTACCTATTGAAAAGACAGTCTTCCCATGCCGTTGTGAGTT
ATTGATGATCAGATAATTCTGATAGCTGTGCTTATTCTGGTTCTTATTCTGTATTGTTATGTCCTGTGTTGTGCAATTAGCCCTGAGT
ATAGTTGAGGTGAGATAGCCTGATGCTTCCAGTTGTTCTTCTTAAGATTGCCATTGCTTGTGTTGCAATTG
GAATGTCATTGAGTTGATAGAAATGCAATTGAAATTGATTGCACTGAGTATGCCCTTAAATGATATTGATTGTCATTGTTGTTGCAATTAGTGT
TATCCTCTGATTCTTGTGCACTGTTGTAATTCTCAT TGAGAGATTCTCAGCTCCCTGGTTAGTTGATTACCTAGTATTGAAAAGACAGTCTTCC
TGCCTCTGATTGACTGC CAGCTGGTACTGTTGGTTATAGAAATGCTAGTGTATTGTCATTGACATG ATTCTCTTCTAAAACCTTGTGAAAGTTTT
TTGGGGCTGAGACTATGGGTTCTAGATATAGAATCATGTCAGCTCAAATAGGGATAATTCTCCTCTTGTGATGCCCTTATTG
GCTGGATTCCATGTGAATAGGAGT CATGAGAGAGGGCATCAAATCACACATCAAACACTAACCTGAAATGCTAGTATTGATTCTTGTGAA
ATTCTCTTGTGAAATTGGAATGGGAT

How much data make up the human genome?

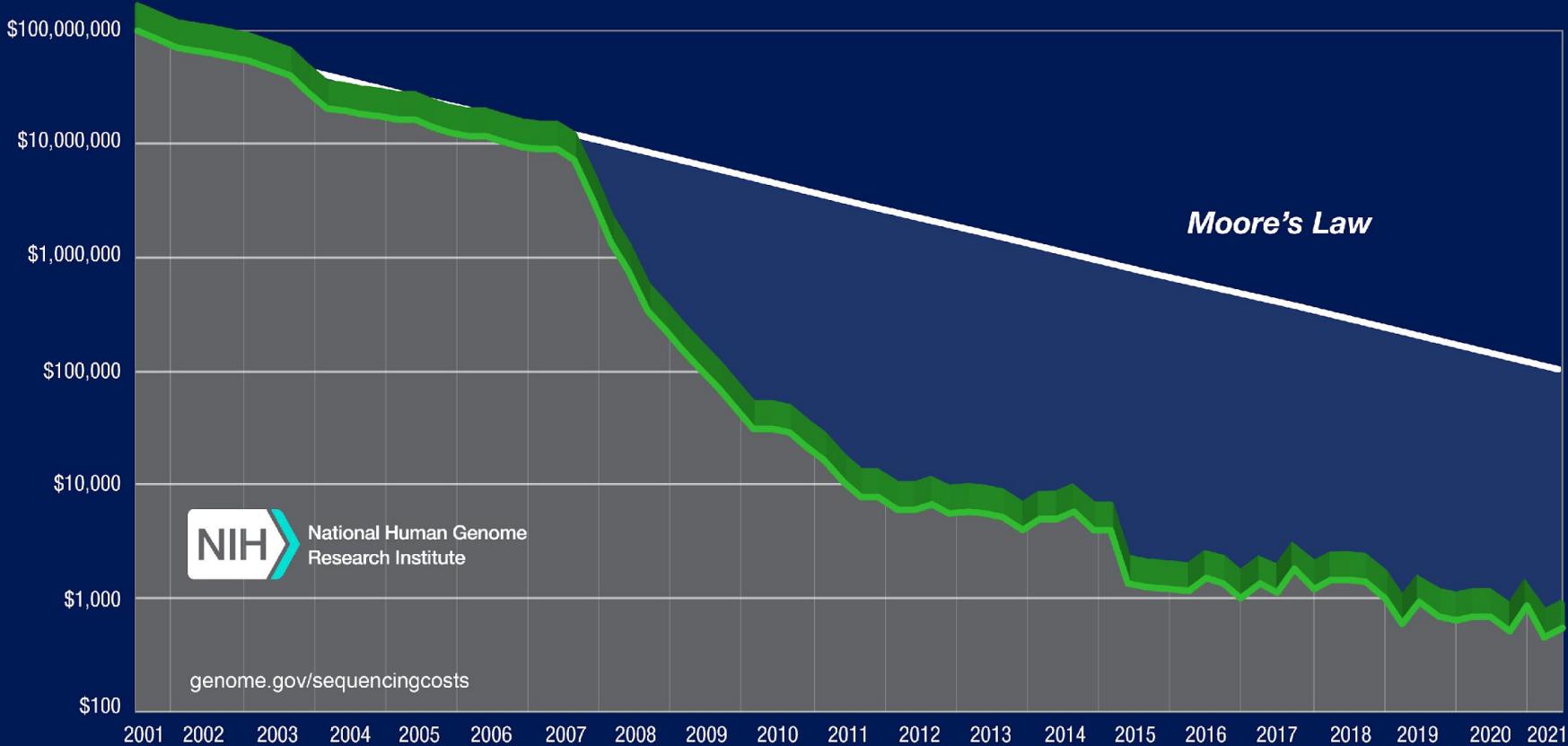
- 3 pallets with 40 boxes per pallet x 5000 pages per box x 5000 bases per page = 3,000,000,000 bases!
- Even on the computer – it is a lot of data



How does the human genome stack up?

Organism	Genome Size (Bases)	Estimated Genes
Human (<i>Homo sapiens</i>)	3 billion	30,000
Laboratory mouse (<i>M. musculus</i>)	2.6 billion	30,000
Mustard weed (<i>A. thaliana</i>)	100 million	25,000
Roundworm (<i>C. elegans</i>)	97 million	19,000
Fruit fly (<i>D. melanogaster</i>)	137 million	13,000
Yeast (<i>S. cerevisiae</i>)	12.1 million	6,000
Bacterium (<i>E. coli</i>)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9700	9

Cost per Human Genome



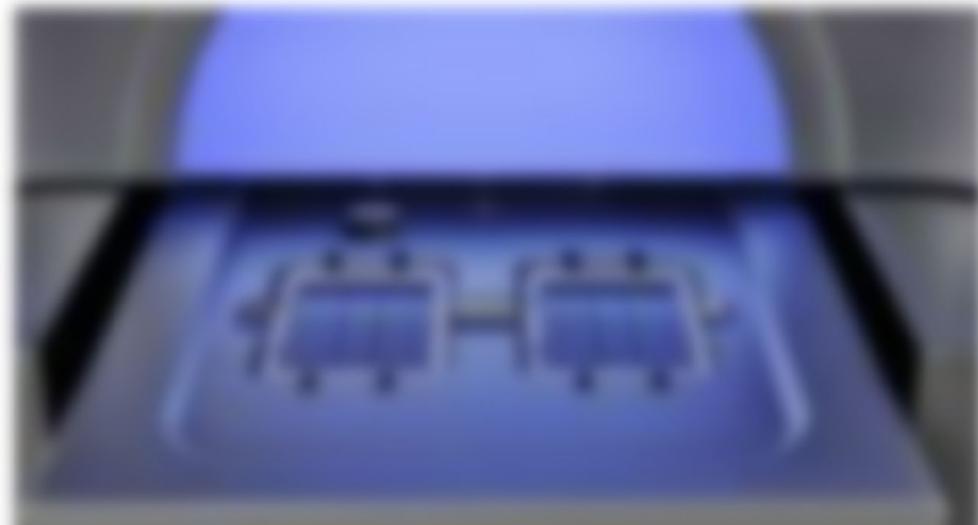
LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



[Live Now](#)[Markets](#)[Economics](#)[Industries](#)[Technology](#)[Politics](#)[Wealth](#)[Pursuits](#)[Opinion](#)[Businessweek](#)[Equality](#)[Green](#)[CityLab](#)[Crypto](#)[More](#) :[Markets](#)
[Prognosis](#)

Illumina Aims to Push Genetics Beyond the Lab With \$200 Genome

- New machine pushes company toward goal of \$100 genome sequence
- Could be boon for gene-targeting drugs, Regeneron exec says



• LIVE ON BLOOMBERG
[Watch Live TV](#) >
[Listen to Live Radio](#) >

B

LONDON
SCHOOL OF
HYGIENE
& TROPICAL
MEDICINE



Human Genetic Variation



- Multi-national, multi-sector project
2002- 2009
- Goal of identifying “tag” SNPs that are informative about genotype for other polymorphic loci
- Associations among SNPs and frequencies vary across populations
- Included Nigerian (Yoruban), Han Chinese, Japanese, and European participants



- Multi-national project, initiated 2008
- Sequence the genomes of ~1,000 individuals from different ethnic groups
- Goal of identifying >95% of genetic variation
 - Structural variation
 - Rare variation in genes
 - Population allele frequencies
 - Haplotypes and linkage disequilibrium patterns
- Samples from HapMap populations, plus Kenya, Italy, Peru, Native American, Asian- American, Mexican-American, African- American.

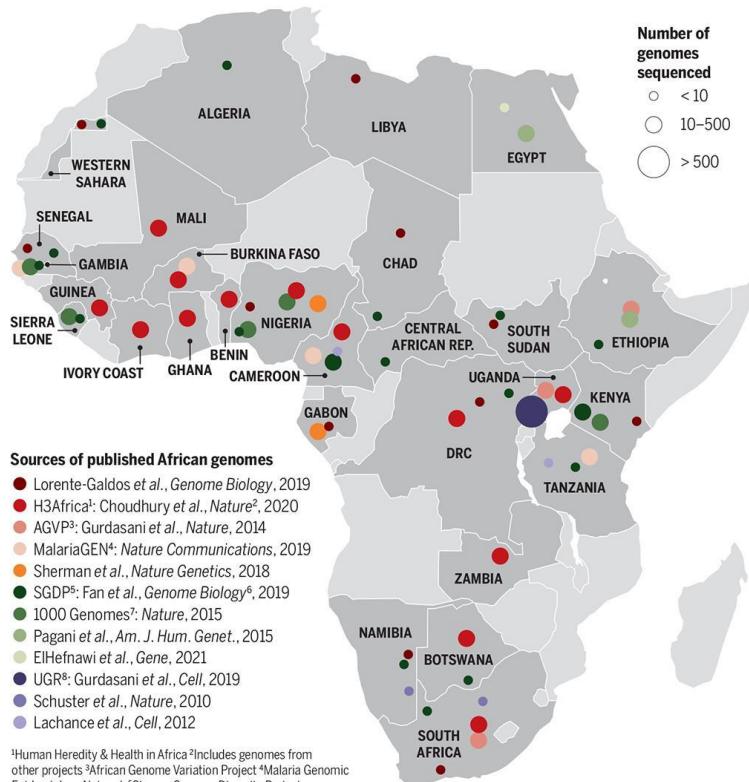


Celebrating 20 Years of Human Genome



Tallying African genomes

Researchers have only just begun to sample the genomes of Africa's 2000 ethnic groups and populations. A handful of whole-genome sequences in 2010 has grown to thousands from multiple projects, many of which are captured on the map below. Their distribution reveals huge gaps in genomic sampling across the continent.



Sources of published African genomes

- Lorente-Galdos *et al.*, *Genome Biology*, 2019
- H3Africa¹; Choudhury *et al.*, *Nature*, 2020
- AGVP²; Gurdasani *et al.*, *Nature*, 2014
- MalariaGEN³; *Nature Communications*, 2019
- Sherman *et al.*, *Nature Genetics*, 2018
- SGDP⁴; Fan *et al.*, *Genome Biology*⁴, 2019
- 1000 Genomes⁵; *Nature*, 2015
- Pagani *et al.*, *Am. J. Hum. Genet.*, 2015
- ElHefnawi *et al.*, *Gene*, 2021
- UGR⁶; Gurdasani *et al.*, *Cell*, 2019
- Schuster *et al.*, *Nature*, 2010
- Lachance *et al.*, *Cell*, 2012

¹Human Heredity & Health in Africa ²Includes genomes from other projects ³African Genome Variation Project ⁴Malaria Genomic Epidemiology Network ⁵Simons Genome Diversity Project

⁶Some data first reported in Mallick *et al.*, *Nature*, 2016. ⁷1000 Genomes Project Consortium ⁸Uganda Genome Resource

Elizabeth Pennisi Science 2021;371:556-559

<https://science.sciencemag.org/content/371/6529/556>

<https://science.sciencemag.org/content/371/6529/556>

Genomes arising

Africans have begun to study their continent's rich human diversity—but what comes after current grants end?

ELIZABETH PENNISI | [Authors Info & Affiliations](#)

SCIENCE • 5 Feb 2021 • Vol 371, Issue 6529 • pp. 556-559 • DOI: 10.1126/science.371.6529.556

 478



e Letters (0)



Volunteers in rural Uganda provided blood samples and health information for the biggest genomics effort in Africa, the Uganda Genome Resource.

PHOTO: GEORGINA MURPHY

I
S
I
C

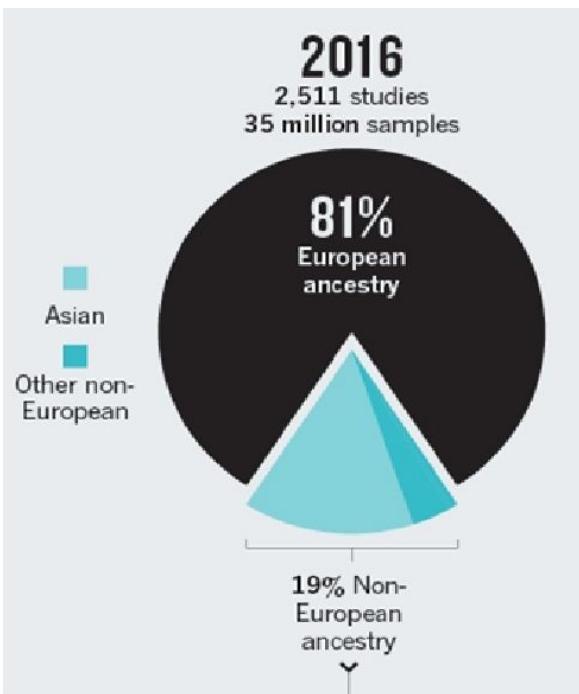
In 1987, 10-year-old Segun Fatumo was on the streets of Lagos, Nigeria, hawking palm oil, yams, and pepper each day after school to help put food on the table. In the evenings, he and his family crowded into a two-room dwelling without running water or electricity. He knew nothing of the plan being hatched by U.S. and U.K. geneticists to sequence the human genome.

<https://science.sciencemag.org/content/371/6529/556>

Despite repeated calls and warning, genomics eurocentric bias is on the rise

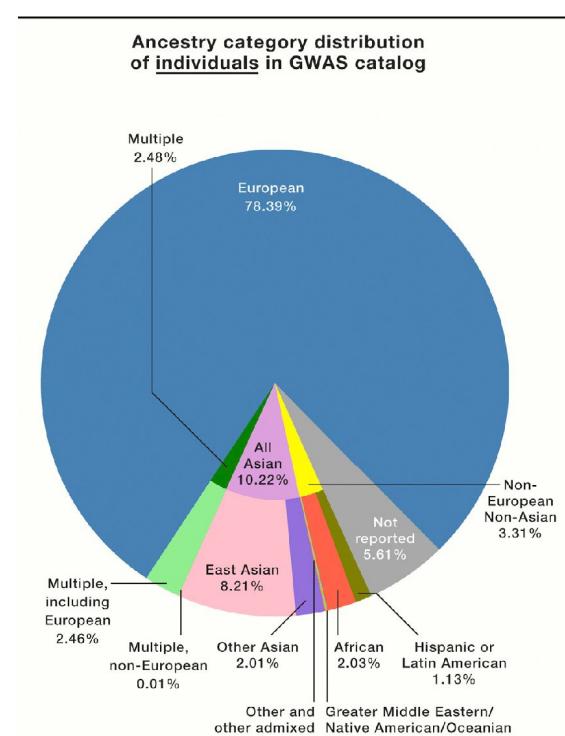
2016

- Europeans: 81%
- Africans: 3%



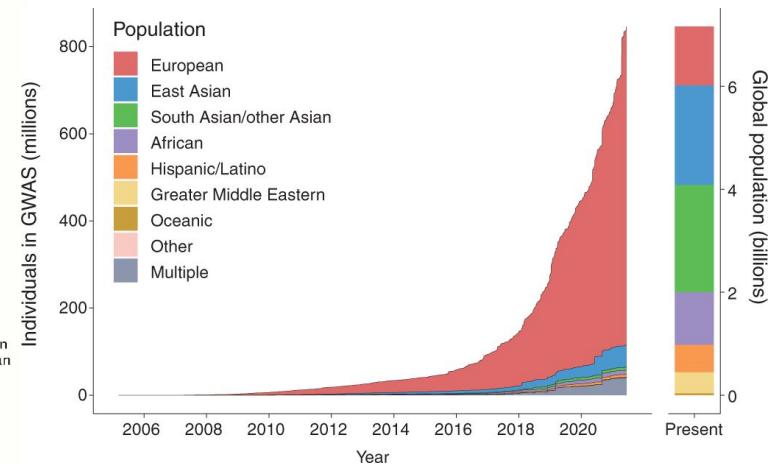
2019

- Europeans: 78.39%
- Africans: 2.03%



2022

- Europeans: 86%
- Africans: 1.1%

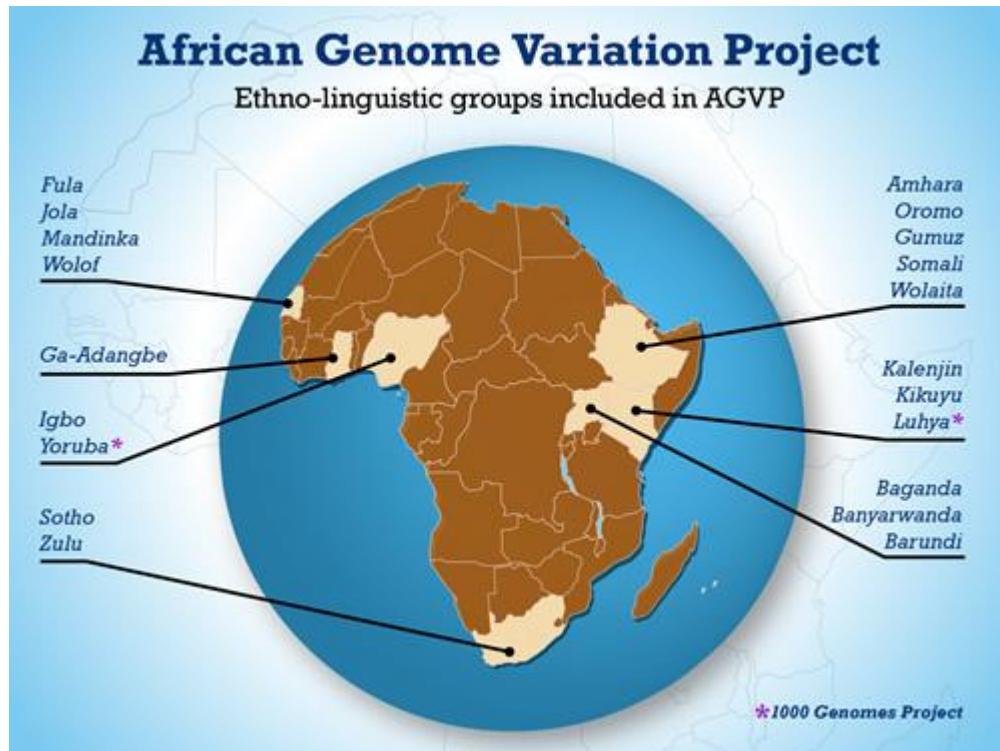


Popejoy, A., Fullerton, S.
2016 *Nature* **538**, 161–164

Sirugo et al., 2019.
Cell, 177(1), 26-31.

Fatumo et al., 2022 *Nature Medicine*, 28(2), 243-250.

The African Genome Variation Project



- AGVP represents dense genotypes from 1,481 individuals and additional whole-genome sequences from 320 individuals across sub-Saharan Africa.
- Found genetic difference in regions within Africa. For example, **people from South Africa are less likely to carry a genetic mutation that offers protection against malaria than those from other parts Africa**

Sequence three million genomes across Africa

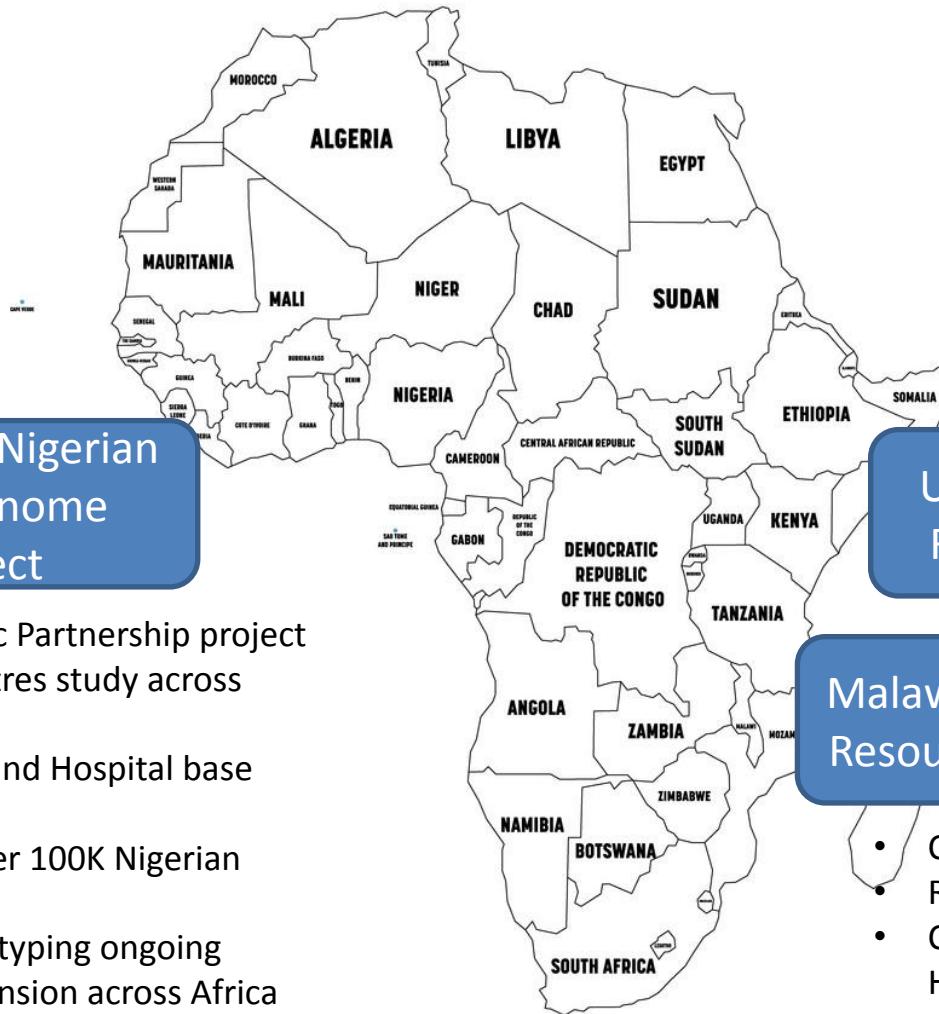
Capture the full scope of variation to improve health care, equity and medical research globally.

Ambroise Wonkam 



Christian Happi at Redeemer's University in Ede, Nigeria, plans to sequence human genomes. Credit: ACEGID

- To improve representation of Africans in genomic studies, we must engage, recruit participants and analyse data from indigenous populations.



NCD-GHS' Nigerian 100K genome Project

- Private-Public Partnership project
 - Multiple centres study across Nigeria
 - Community and Hospital base studies
 - Recruited over 100K Nigerian Participants
 - WGS & Genotyping ongoing
 - Plan for expansion across Africa

Uganda Genome Resource (UGR)

- Cohort population of 22K
 - 5000 Participants genotyped
 - Additional 2000 WGS
 - Plan for expansion
 - Proteomics, Metabolomics, Epigenomics and Single-cell genomics

Malawi Genome Resource (MGR)

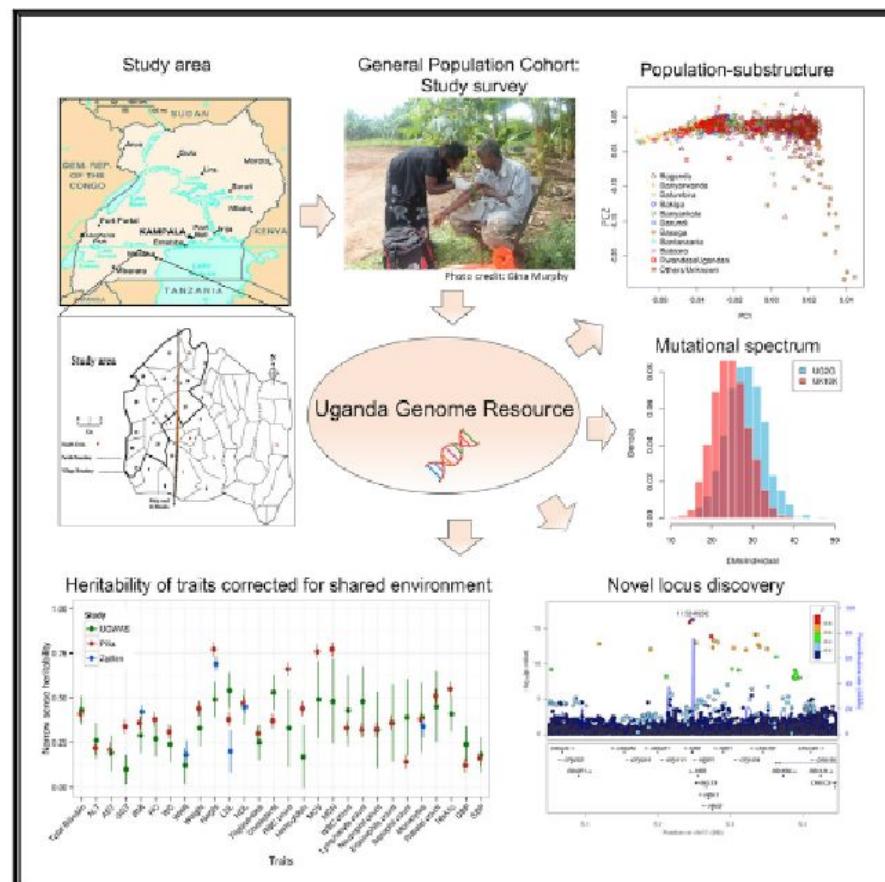
- Cohort population of 25K
 - Rural and Urban community
 - Currently genotyping ~7000 on H3Africa Array
 - Plan for expansion

Largest ever genome study of Africans

Cell

Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa

Graphical Abstract



Authors

*Deepti Gurdasani,*Tommy Carstensen,
*Segun Fatumo, ..., Pontiano Kaleebu,
Ines Barroso, Manj S. Sandhu

*co-first Authors

Correspondence

cts@sanger.ac.uk (C.T.-S.),
motala@ukzn.ac.za (A.M.),
rotimic@mail.nih.gov (C.R.),
pontiano.kaleebu@mrcuganda.org (P.K.),
ib1@sanger.ac.uk (I.B.),
mss31@cam.ac.uk (M.S.S.)

In Brief

Genome-wide data from Ugandans reveal insights into their ancestry, trait heritability, and loci associated with metabolic parameters, thereby providing a diverse resource for the study of African population genetics.

High-depth African genomes inform human migration and health

<https://doi.org/10.1038/s41586-020-2859-7>

Received: 10 May 2019

Accepted: 7 August 2020

Published online: 28 October 2020

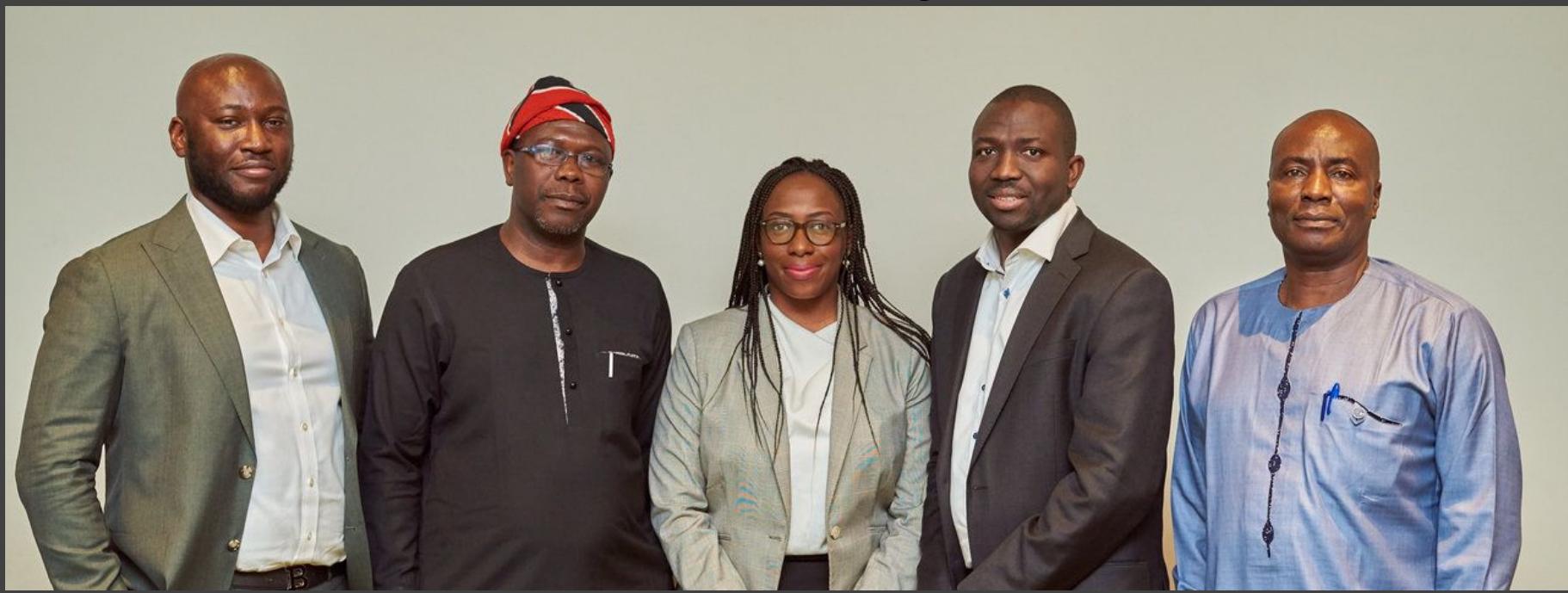
Open access



Ananyo Choudhury¹, Shaun Aron¹, Laura R. Botigué², Dhriti Sengupta¹, Gerrit Botha³, Taoufik Bensellak⁴, Gordon Wells^{5,6,49}, Judit Kumuthini^{5,6}, Daniel Shriner⁷, Yasmina J. Fakim^{8,9}, Anisah W. Ghoorah⁹, Eileen Dareng^{10,11}, Trust Odia¹², Oluwadamilare Falola¹², Ezekiel Adebiyi^{12,13}, Scott Hazelhurst^{1,14}, Gaston Mazandu³, Oscar A. Nyangiri¹⁵, Mamana Mbiyavanga³, Alia Benkahla¹⁶, Samar K. Kassim¹⁷, Nicola Mulder³, Sally N. Adebamowo^{18,19}, Emile R. Chimusa²⁰, Donna Muzny²¹, Ginger Metcalf²¹, Richard A. Gibbs^{21,22}, TrypanoGEN Research Group*, Charles Rotimi⁷, Michèle Ramsay^{1,23}, H3Africa Consortium*, Adebowale A. Adeyemo⁷✉, Zané Lombard²³✉ & Neil A. Hanchard²²✉

The African continent is regarded as the cradle of modern humans and African genomes contain more genetic variation than those from any other continent, yet only a fraction of the genetic diversity among African individuals has been surveyed¹. Here we performed whole-genome sequencing analyses of 426 individuals—comprising 50 ethnolinguistic groups, including previously unsampled populations—to explore the breadth of genomic diversity across Africa. We uncovered more than 3 million previously undescribed variants, most of which were found among individuals from newly sampled ethnolinguistic groups, as well as 62 previously unreported loci that are under strong selection, which were predominantly found in genes that are involved in viral immunity, DNA repair and metabolism. We observed complex patterns of ancestral admixture and putative-damaging and novel variation, both within and between populations, alongside evidence that population from Zambia were a likely intermediate site along the routes of expansion of Bantu-speaking populations. Pathogenic variants in genes that are currently characterized as medically relevant were uncommon—but in other genes, variants denoted as ‘likely pathogenic’ in the ClinVar database were commonly observed. Collectively, these findings refine our current understanding of continental migration, identify gene flow and the response to human disease as strong drivers of genome-level population variation, and underscore the scientific imperative for a broader characterization of the genomic diversity of African individuals to understand human ancestry and improve health.

To improve representation of Africans in genomics, we are creating the Nigerian 100K Genome Project



1

Meet the NCD-GHS
Consortium Leads

2

Abasi Ene-Obong/CEO 54GENE
Babatunde Salako/CEO NIMR
Omolola Salako/Consultant Oncologist, CMUL

3

Segun Fatumo/Associate Prof, LSHTM
Oyekanmi Nash/Director, CGRI-NABDA

4

5

African scientists must stop waiting for the government money that would never come. Time to think about an alternative – eg Private/Public Partnership

Comment | [Published: 09 May 2022](#)

Promoting the genomic revolution in Africa through the Nigerian 100K Genome Project

[Segun Fatumo](#)✉, [Aminu Yakubu](#), [Olubukunola Oyedele](#), [Jumi Popoola](#), [Delali Attiogbe Attipoe](#), [Golibe Eze-Echesi](#), [Fatima Z Modibbo](#), [Nabila Ado-Wanka](#), [54gene Team](#), [NCD-GHS Consortium](#), [Omolola Salako](#), [Oyekanmi Nashiru](#), [Babatunde L Salako](#), [Colm O'Dushlaine](#) & [Abasi Ene-Obong](#)✉

Nature Genetics **54**, 531–536 (2022) | [Cite this article](#)

4137 Accesses | **4** Citations | **338** Altmetric | [Metrics](#)

Conclusion

- Human Genomics advances for 20 years since the first human genome was published.
- Since then, advances in genome technologies have resulted in whole-genome sequencing and microarray-based genotyping of millions of human genomes and several databases to house them.
- However, human genetic and genomic studies are predominantly based on populations of European ancestry.
- Africa is still lagging behind, but great initiatives such as H3Africa, NCD-GHS, TopMed, eMERGE, PAGE are helping to fill the gaps



diagnostics

an Open Access Journal by MDPI

IMPACT
FACTOR
3.992

Indexed in:
PubMed

Genomic Diagnosis in Next - Generation Sequencing

Guest Editors

Dr. Segun Fatumo, Dr. Manuel Corpas

Deadline

24 March 2023

[mdpi.com/si/145436](https://www.mdpi.com/si/145436)

Special Issue

Invitation to submit

https://www.mdpi.com/journal/diagnostics/special_issues/R021T940FS

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



As a reminder

Genetics: the study of genes, variation and heredity in living organisms (eg Human)

1. Genes

2. Genetic variation

3. Heredity

Thank
you

