

# NGS and population genomics



[vincenza.colonna@igb.cnr.it](mailto:vincenza.colonna@igb.cnr.it)

Consiglio Nazionale delle Ricerche

# NGS

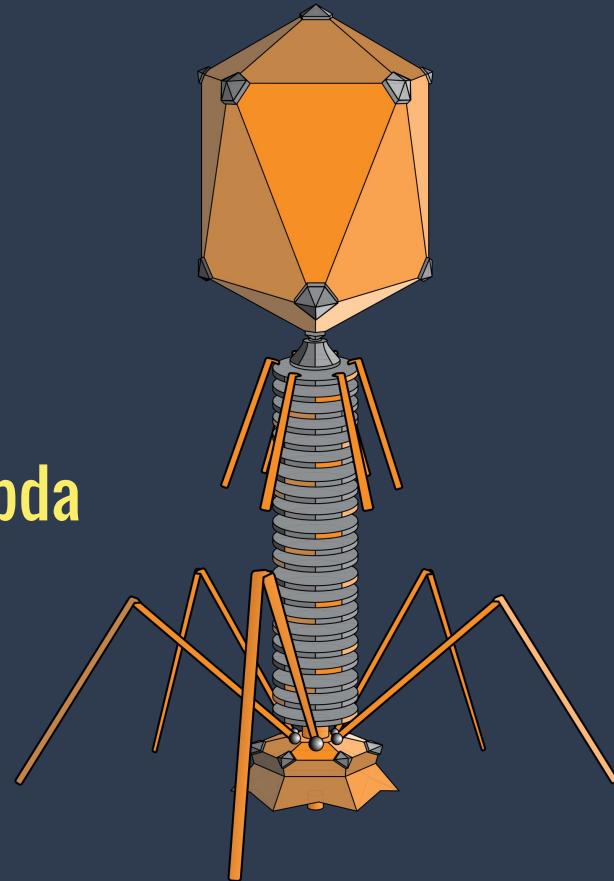
*Size Matters*

# First Whole Genome Sequences

1977 - 5,386bp - single-stranded - phage  $\varphi$ X174

1982 - 48,502bp - double-stranded - phage lambda

1983 - 39,936bp - double stranded - phage T7

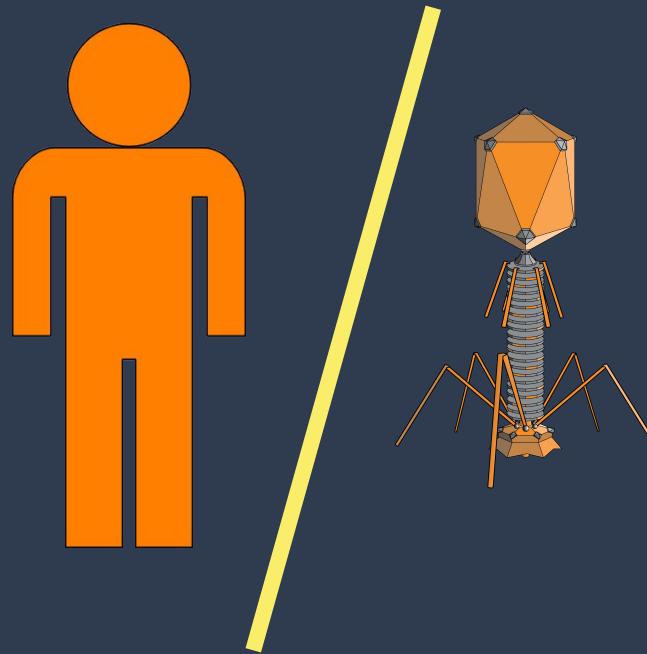


# DNA size matters

human / phage  $\varphi$ X174 = 668,399.6

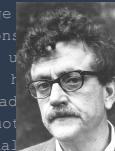
human/ phage lambda = 74,223.74

human/ phage T7 = 90,144.23

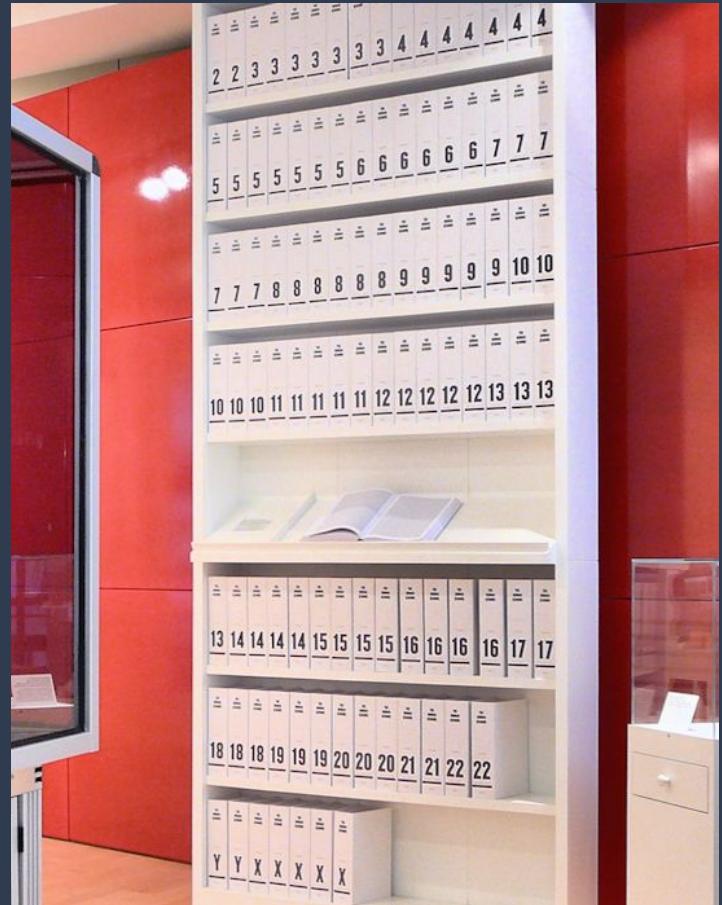


# 5,386 characters from “Galapagos”

"That was another thing people used to be able to do, which they can't do anymore: enjoy in their heads events which hadn't happened yet and might never occur. My mother was good at that. Someday my father would stop writing science fiction, and write something a whole lot of people wanted to read instead. And we would get a new house in a beautiful city, and nice clothes, and so on. She used to make me wonder why God had ever gone to all the trouble of creating reality. Quoth Mandarax:Imagination is as good as many voyages - and how much cheaper!- GEORGE WILLIAM CURTIS"-I'll tell you what the human soul is, Mary,' he whispered, his eyes closed. 'Animals don't have one. It's the part of you that knows when your brain isn't working right. I always knew, Mary. There wasn't anything I could do about it, but I always knew."-"Why so many of us knocked us major chunks of our brains with alcohol from time to time remains an interesting mystery. It may be that we were trying to give evolution a shove in the right direction - in the direction of smaller brains."-Colds and babies were both caused by germs which loved nothing so much as a mucous membrane."-"Why so many of us knocked us major chunks of our brains with alcohol from time to time remains an interesting mystery. It may be that we were trying to give evolution a shove in the right direction - in the direction of smaller brains."-"What humanity was about to lose, though, except for one tiny colony on Santa Rosalia, was what the trackless sea could never lose, so long as it was made of water, the ability to heal itself."-"This woman was so ugly and stupid, she probably never should have been born. And yet Wait was the second person to have married her"-Human beings used to be molecules which could do many, many different sorts of dances, or decline to dance at all --as they pleased. My mother could do the waltz, the tango, the rumba...."-What made marriage so difficult back then was yet again that instigator of so many other sorts of heartbreak: the oversize brain. That cumbersome computer could hold so many contradictory opinions on so many different subjects all at once, and switch from one opinion or subject to another one so quickly, that a discussion between a husband and wife under stress could end up like a fight between blindfolded people wearing roller skates""That was another thing people used to be able to do, which they can't do anymore: enjoy in their heads events which hadn't happened yet and might never occur. My mother was good at that. Someday my father would stop writing science fiction, and write something a whole lot of people wanted to read instead. And we would get a new house in a beautiful city, and nice clothes, and so on. She used to make me wonder why God had ever gone to all the trouble of creating reality. Quoth Mandarax:Imagination is as good as many voyages - and how much cheaper!- GEORGE WILLIAM CURTIS"-I'll tell you what the human soul is, Mary,' he whispered, his eyes closed. 'Animals don't have one. It's the part of you that knows when your brain isn't working right. I always knew, Mary. There wasn't anything I could do about it, but I always knew."-"Why so many of us knocked us major chunks of our brains with alcohol from time to time remains an interesting mystery. It may be that we were trying to give evolution a shove in the right direction - in the direction of smaller brains."-Colds and babies were both caused by germs which loved nothing so much as a mucous membrane."-"Why so many of us knocked us major chunks of our brains with alcohol from time to time remains an interesting mystery. It may be that we were trying to give evolution a shove in the right direction - in the direction of smaller brains."-"What humanity was about to lose, though, except for one tiny colony on Santa Rosalia, was what the trackless sea could never lose, so long as it was made of water, the ability to heal itself."-"This woman was so ugly and stupid, she probably never should have been born. And yet Wait was the second person to have married her"-Human beings used to be molecules which could do many, many different sorts of dances, or decline to dance at all --as they pleased. My mother could do the waltz, the tango, the rumba...."-What made marriage so difficult back then was yet again that instigator of so many other sorts of heartbreak: the oversize brain. That cumbersome computer could hold so many contradictory opinions on so many different subjects all at once, and switch from one opinion or subject to another one so quickly, that a discussion between a husband and wife under stress could end up like a fight between blindfolded people wearing roller skates""That was another thing people used to be able to do, which they can't do anymore: enjoy in their heads events which hadn't happened yet and might never occur. My mother was good at that. Someday my father would stop writing science fiction, and write something a whole lot of people wanted to read instead. And we would get a new house in a beautiful city, and nice clothes, and so on. She used to make me wonder why God had ever gone to all the trouble of creating reality. Quoth Mandarax:Imagination is as good as many voyages - and how much cheaper!- GEORGE WILLIAM CURTIS"-I'll tell you what the human soul is, Mary,' he whispered, his eyes closed. 'Animals don't have one. It's the part of you that knows when your brain isn't working right. I always knew, Mary. There wasn't anything I could do about it, but I always knew."-"Why so many of us knocked us major chunks of our brains with alcohol from time to time remains an interesting mystery. It may be that we were trying to give evolution a shove in the right direction - in the direction of smaller brains."-Colds and babies were both caused by germs which loved nothing so much as a mucous membrane."



3,600,000,000 characters  
from the human genome



# Before the Human Genome Project

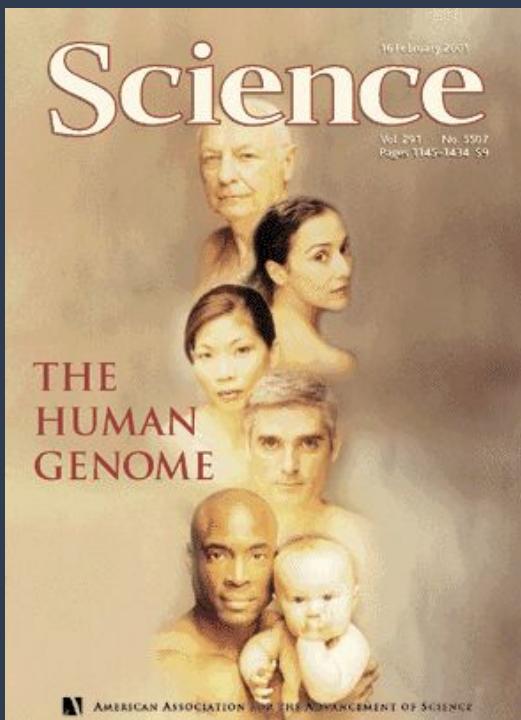
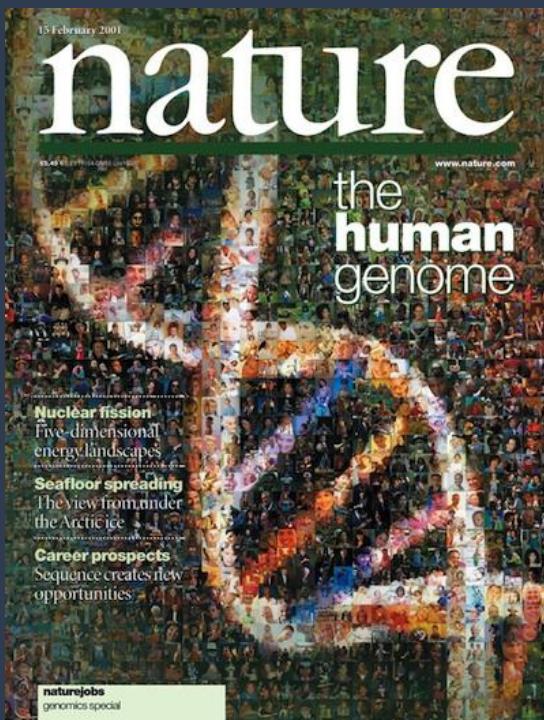


*“...somebody working on your car to make a car repairing without fundamentally knowing how the car is put together”*

Eric Green

(Nature Podcast, 03 Oct 2015)

# Year 2000



# Human Genome Project

vs

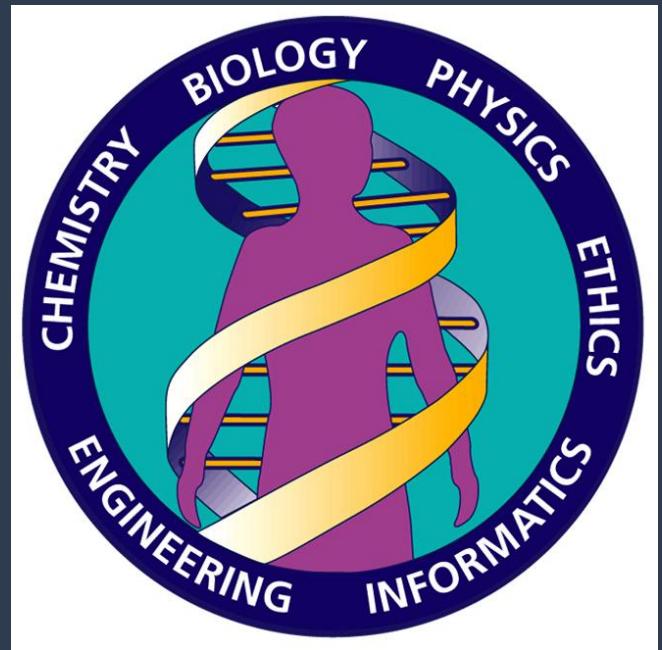
# Celera Genomics

- Launched in 1990
- Publicly funded: \$3 billion
- Data publicly available
- Launched in 1998
- Privately funded: \$0.3 billion
- Relied upon data made available by the publicly funded project
- Requested patent applications on 6,500 whole or partial genes

# Lesson #1: Embrace partnerships

> 2000 researchers  
many disciplines

- strong leadership from the funders
- shared sense of the importance of the task
- cede individual achievements for the collective good



## Lesson #2: Maximize data sharing

1996 - Rapid and public release of DNA sequence data (Bermuda Principles)

2013 - International Framework for Responsible Sharing of Genomic and Health-related Data (Global Alliance for Genomics and Health)

2014 - Almost all large-scale genomic data generated or analysed using NIH funds should be shared (Genomic Data Sharing Policy, NIH)



# Lesson #3: Plan for data analysis

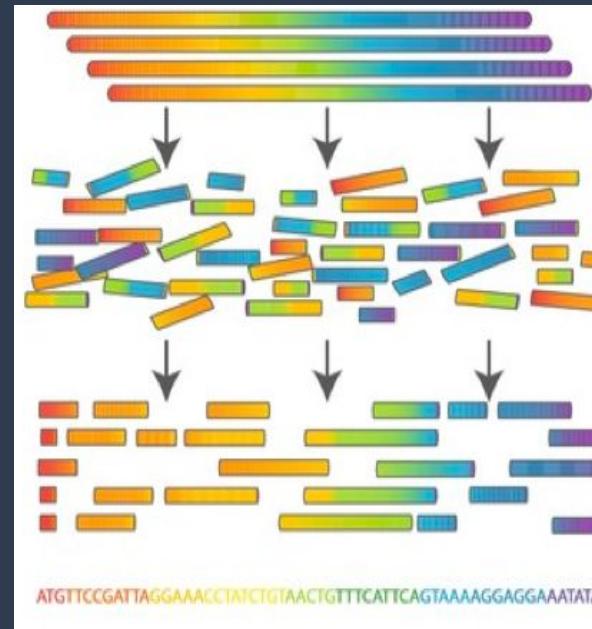
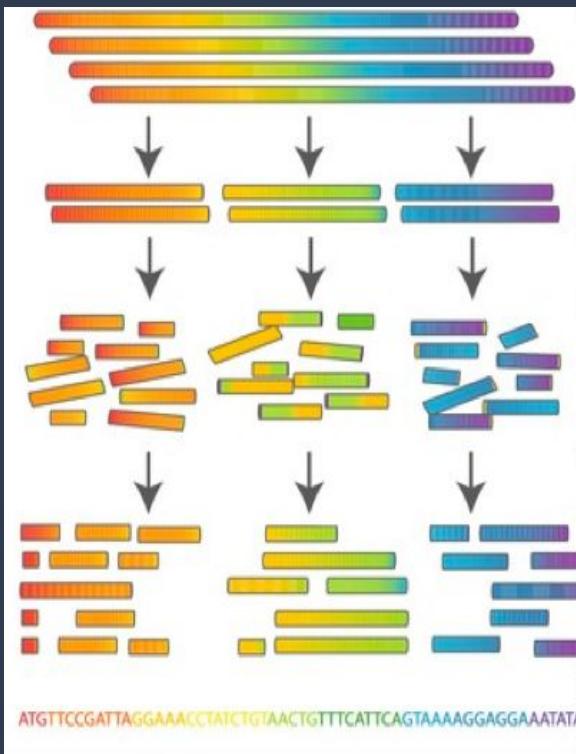
Early design of plans for data analysis can inform strategies for data generation



## Hierarchical shotgun

vs

## Whole genome shotgun sequencing

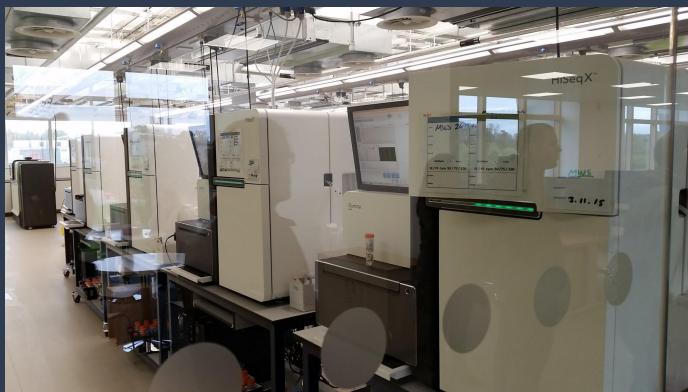


# Lesson #4: Prioritize technology development

HiSeq X System can sequence a human genome at 30x coverage or greater for significantly less than \$1000 in consumables. When used at scale, the HiSeq X Ten delivers a \$1000 genome, inclusive of instrument depreciation, DNA extraction, library preparation, and estimated labor for a typical high-throughput genomics laboratory.



# Wellcome Trust Sanger Institute sequencing facility



# GENOMES ON PRESCRIPTION

*The first clinical uses of whole-genome sequencing show just how challenging it can be.*

BY BRENDAN MAHER



# Lesson #5: Address the societal implications of advances

## ELSI (ethical, legal and social implications) research

Supported by about 5% of the NIH budget for the HGP

The largest ever investment in bioethics research

First large-scale research project to include a component dedicated to examining broader societal issues:

- how to protect people's privacy
- how to prevent discrimination.

# Lesson #6: Be audacious yet flexible



## Lesson #7: Game changer

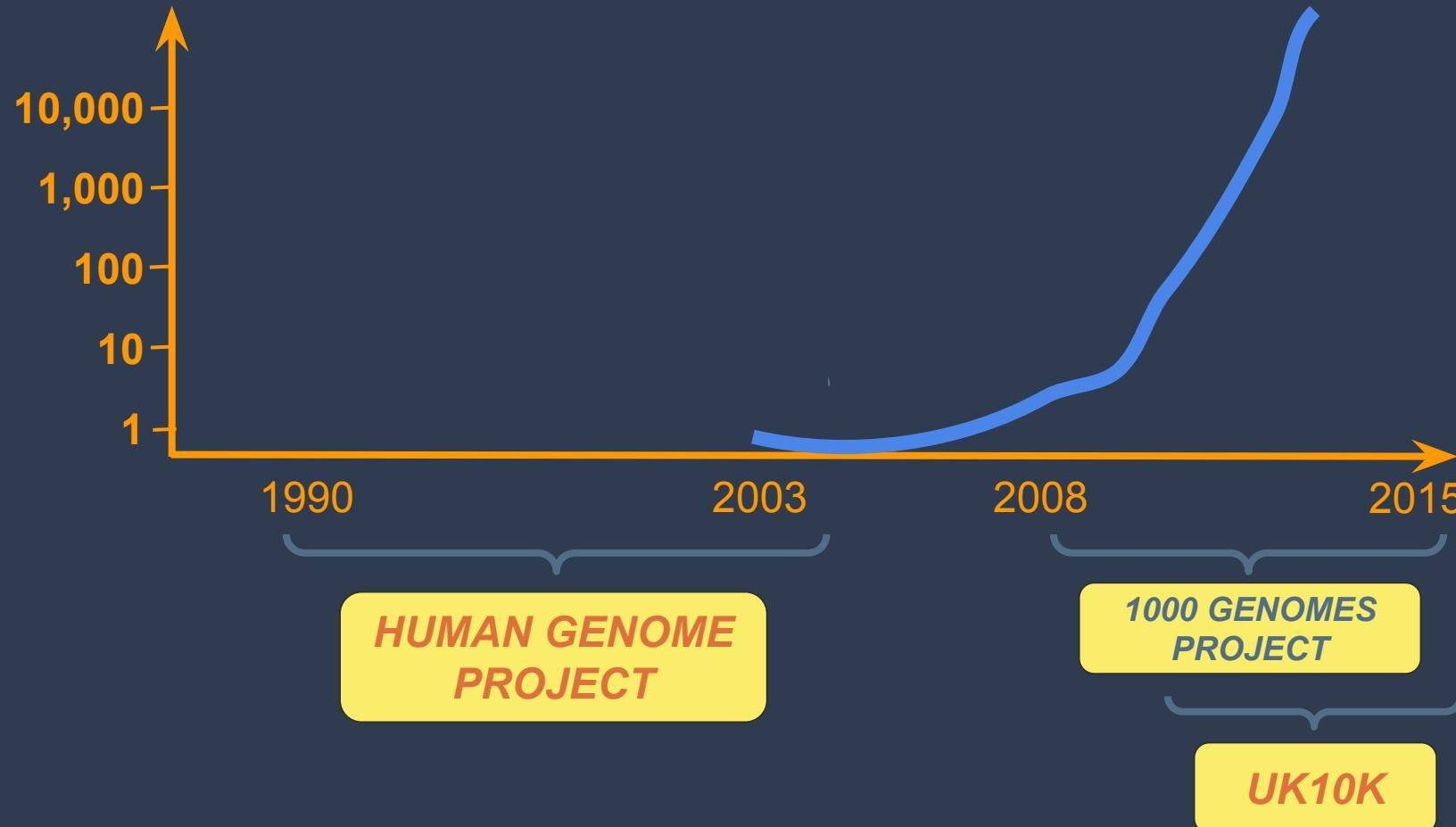
*"In the early 1990s — whether it was while leading the NIH's effort in the HGP or working on the front line of the project — none of us foresaw that a major legacy of the HGP would be a new way of doing science"*

Eric D. Green, James D. Watson & Francis S. Collins

Human Genome Project: Twenty-five years of big biology (Nature)



# of publicly available sequences





# Population genomics

*Happily Overwhelmed*

# The 1000 Genomes Project timeline

Pilot	low coverage, exome, trios	180 inds	4 pops	30 M variants
Phase 1	exome, low coverage	1000 inds	13 pops	49 M variants
Phase 2-3	method improvement	2500 inds	26 pops	88 M variants

# It was fun...

Top stories

<http://www.guardian.co.uk/science>



1000 Genomes Project completes its first map of human genetic variation

Thousands of human genomes are being used to catalogue the full diversity of human DNA in the 1000 Genomes Project

14 comments





# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## END OF THE BEGINNING

Final phase of 1000 Genomes Project maps human genetic variation in open-access resource PAGES 52, 60 & 75

REVIEWS  
**AUTUMN BOOKS SPECIAL**  
*Fizz*, wars, geopolity, mind and matter  
PAGE 34

MAUNA KEA OBSERVATORY  
**MOUNTAIN DIFFICULTIES**  
*Is the Thirty Meter Telescope a step too far?*  
PAGE 24

THE HUMAN GENOME  
**25 YEARS OF BIG BIOLOGY**  
*Three major players reflect on lessons learned*  
PAGE 29

NATURE.COM/NATURE  
1 October 2015 £10  
Vol 526, No. 7571  
ISSN 0028-1811  
406  
TIFF Review  
9 770028053897

# 1000 Genomes Project Phase 3

2,504 individuals  
26 populations

88 M variants

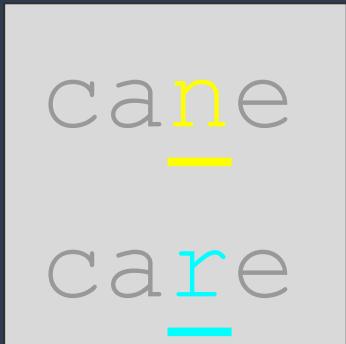
<http://www.nature.com/collections/dcfqmlgsrw/>

# Aims

- Discover population level human genetic variations of all types (95% of variation > 1% frequency)
- Define haplotype structure in the human genome
- Develop sequence analysis methods, tools, and other reagents that can be transferred to other sequencing projects



# Variants = Alleles != Mutations

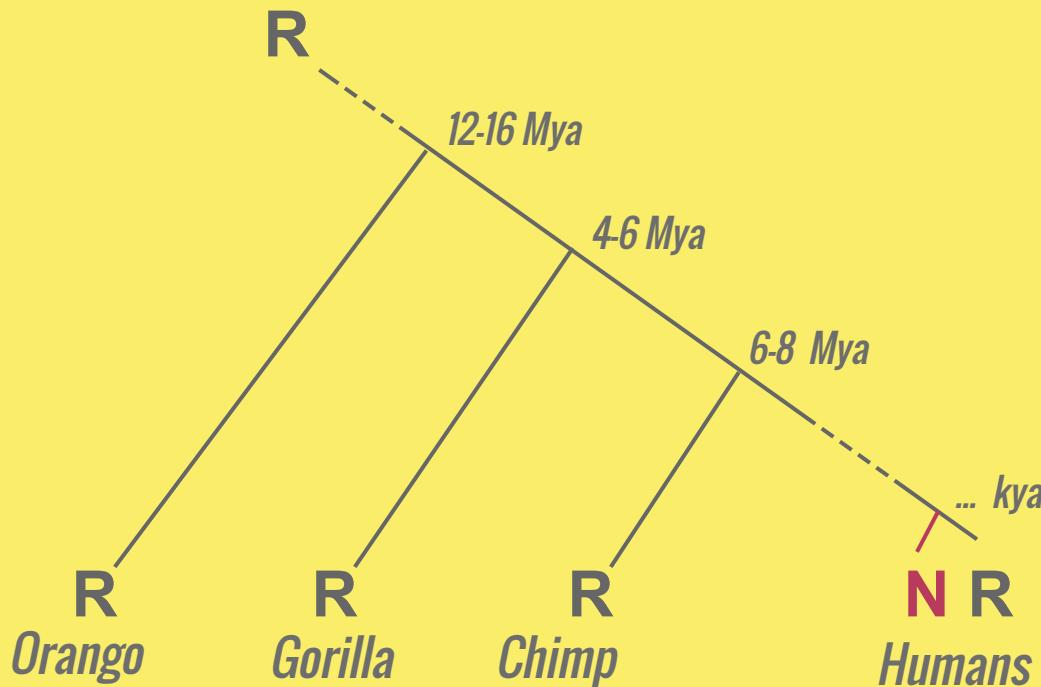


cane  
—  
care  
—

n e r are genetic variants or alleles

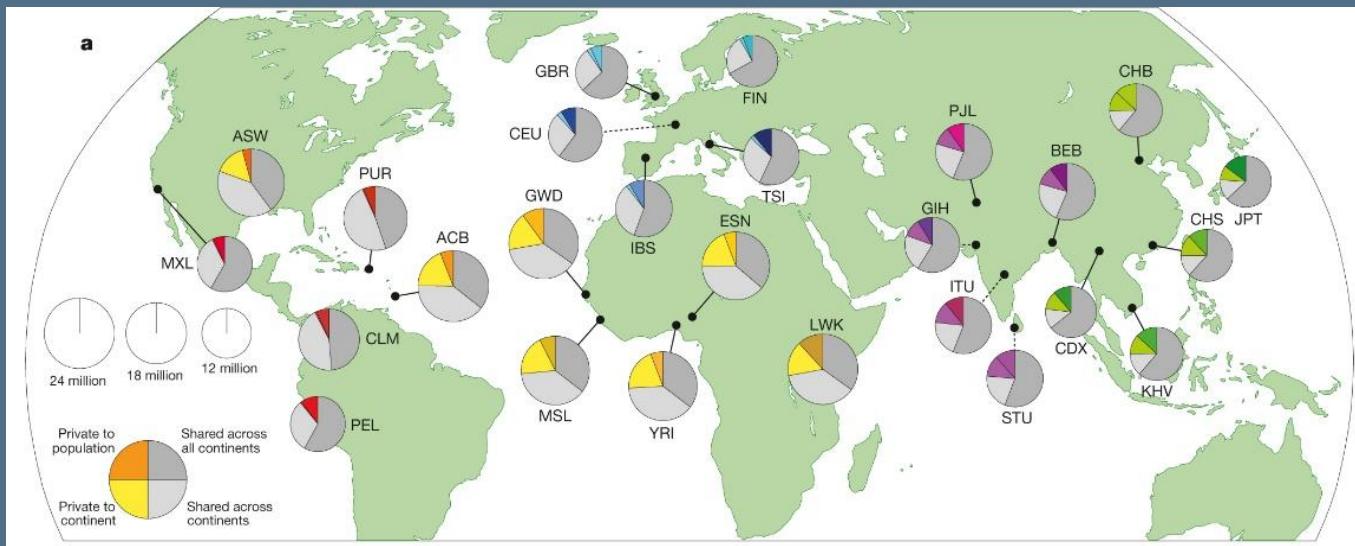
n originated from a mutation of r

# The concept of derived allele



# Lessons from the 1000 Genomes project

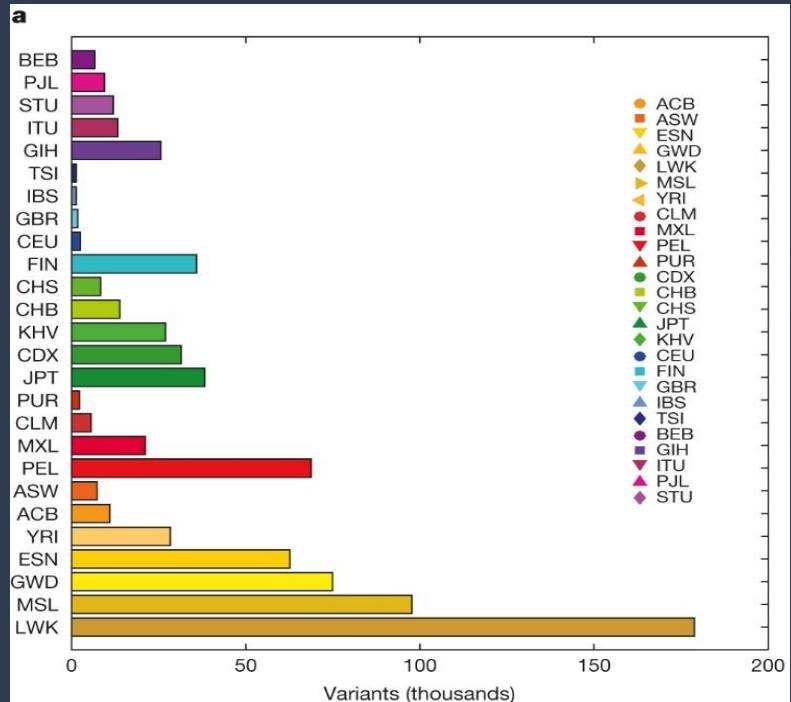
# Population sampling



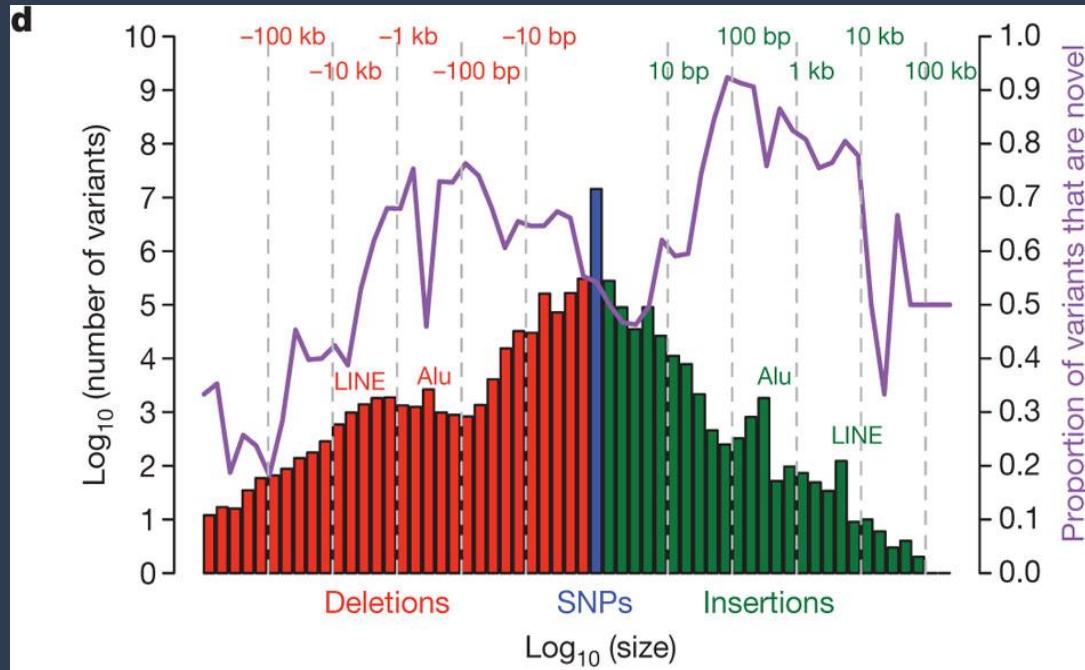
A Auton et al. *Nature* 526, 68-74 (2015) doi:10.1038/nature15393

**nature**

Africans have the highest  
number of variants  
compared to the  
reference sequence (P3)



# Properties of the variants found (Pilot)



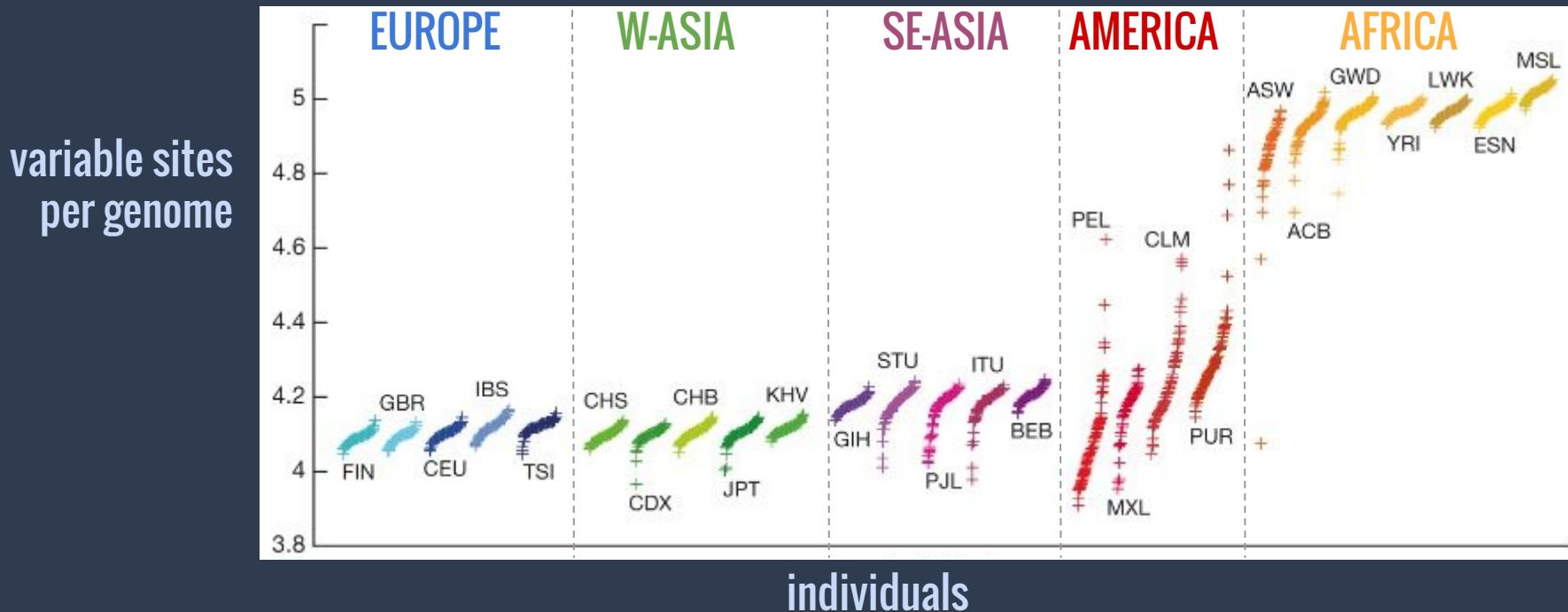
The 1000 Genomes Project Consortium et al. Nature 467, 1061-1173 (2010) doi:10.1038/nature09534

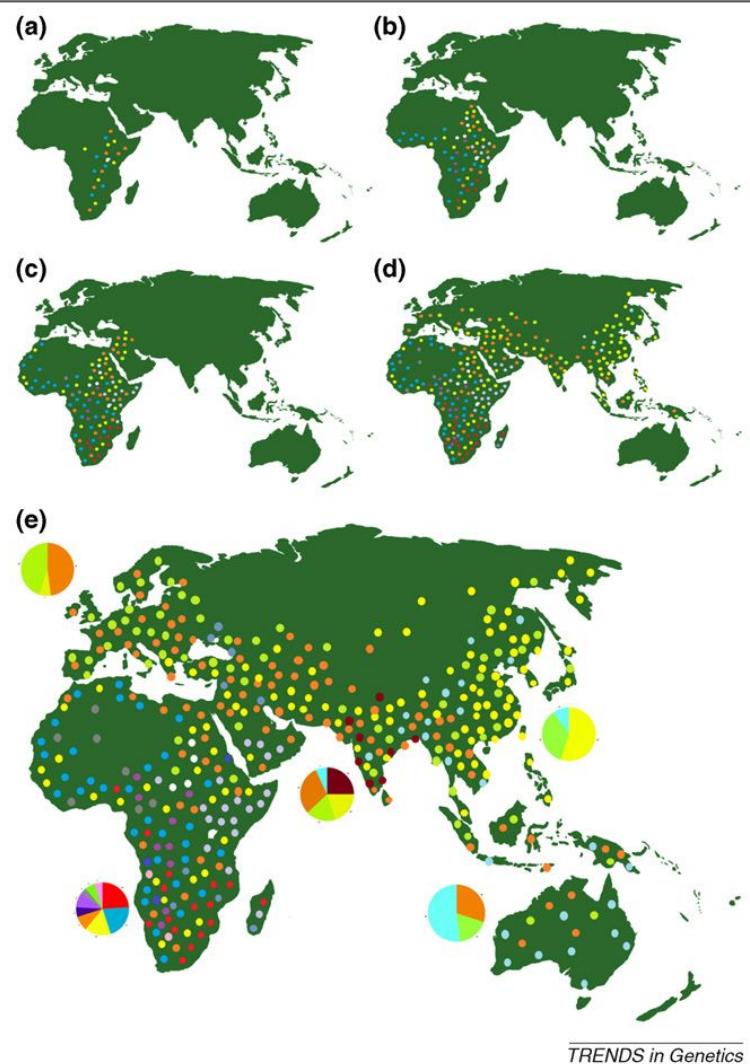
nature

# A typical genome (P3)

- **4.1 million to 5.0 million** sites differ from the reference human genome
- 
- **>99.9%** of variants consist of **SNPs and short indels** (~ 5-10 million bases)
- 
- **0.1% of structural variants** affect more bases of sequence: 2,100 to 2,500 structural variants (~20 million bases)
  - ~1,000 large deletions, ~160 copy-number variants, ~915 Alu insertions, ~128 L1 insertions, ~51 SVA insertions, ~4 NUMTs, and ~10 inversions

# 4.3 M differences on average between two individuals

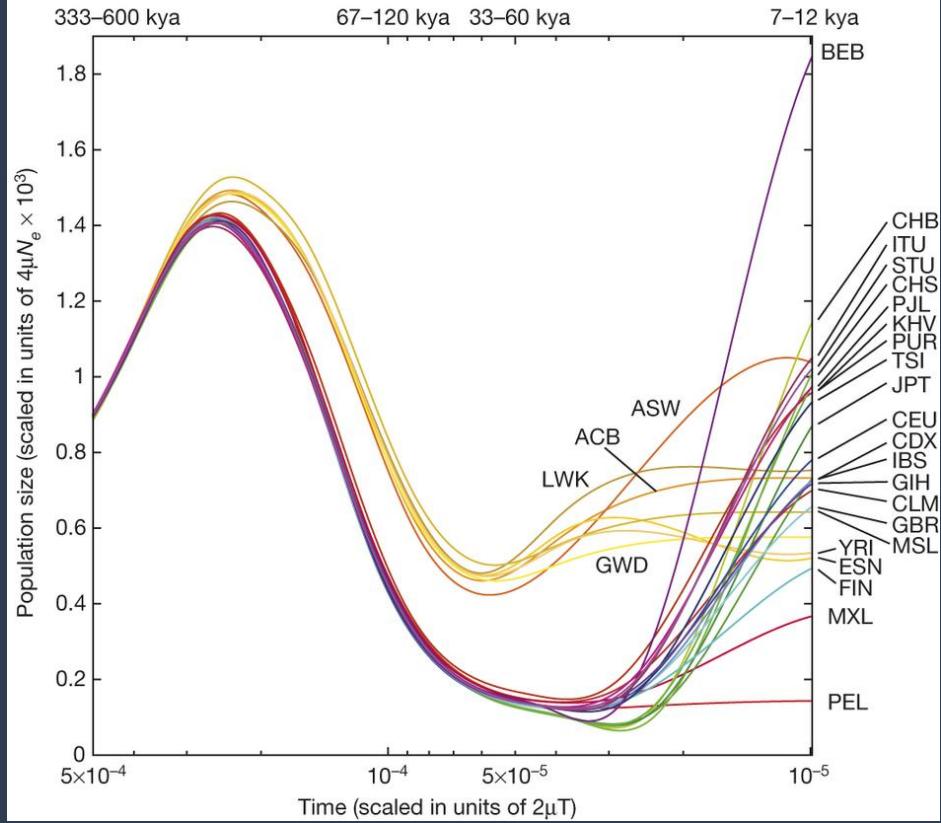




# Why there is more genetic diversity in Africans ?

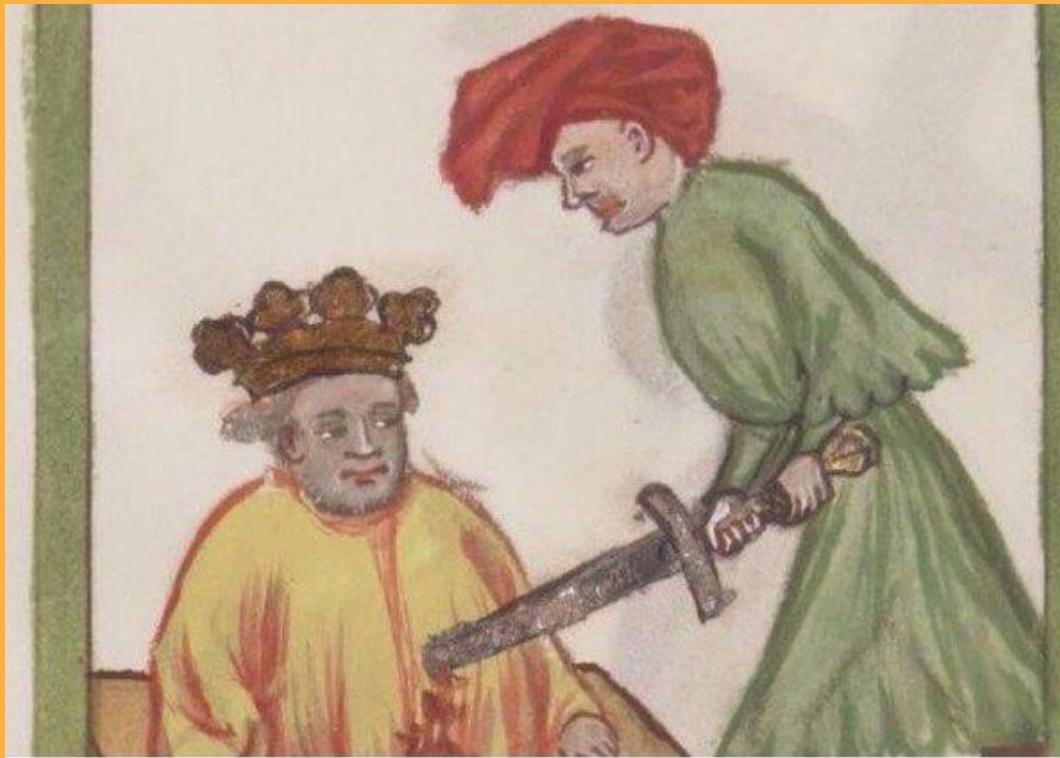
Barbujani & Colonna Trends in Genetics 2010 Jul;26(7):285-95

**b** Time, assuming  $\mu = 1.25 \times 10^{-8}$  to  $1.5 \times 10^{-8}$  per bp per generation and 20–30 years per generation



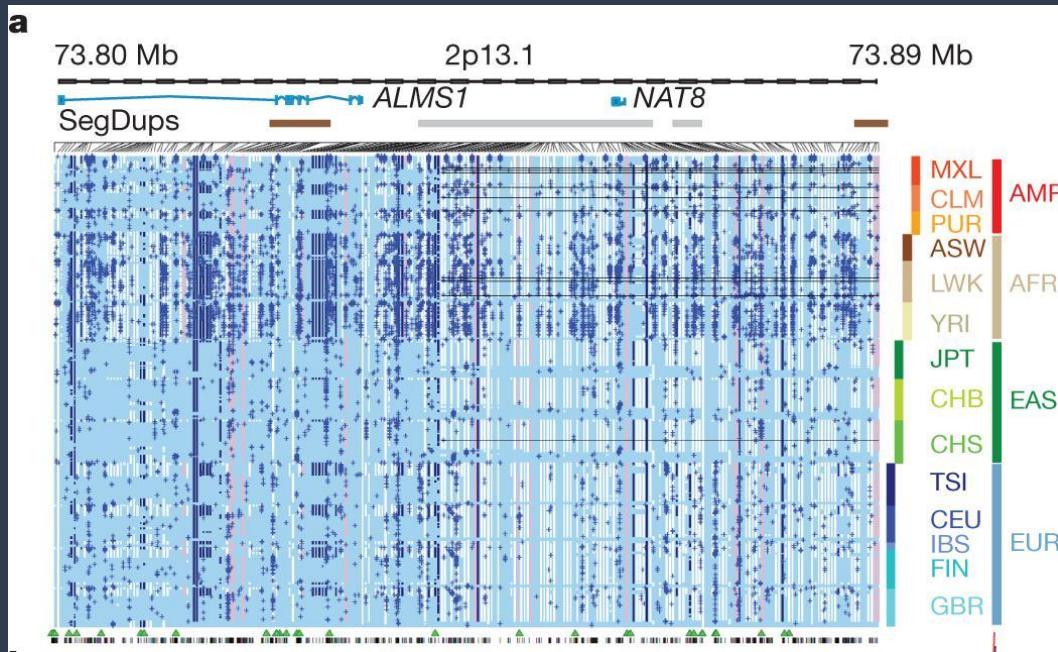
Highest effective population size in Africans

# I don't care



When somebody's roasting you but you're already dead inside.

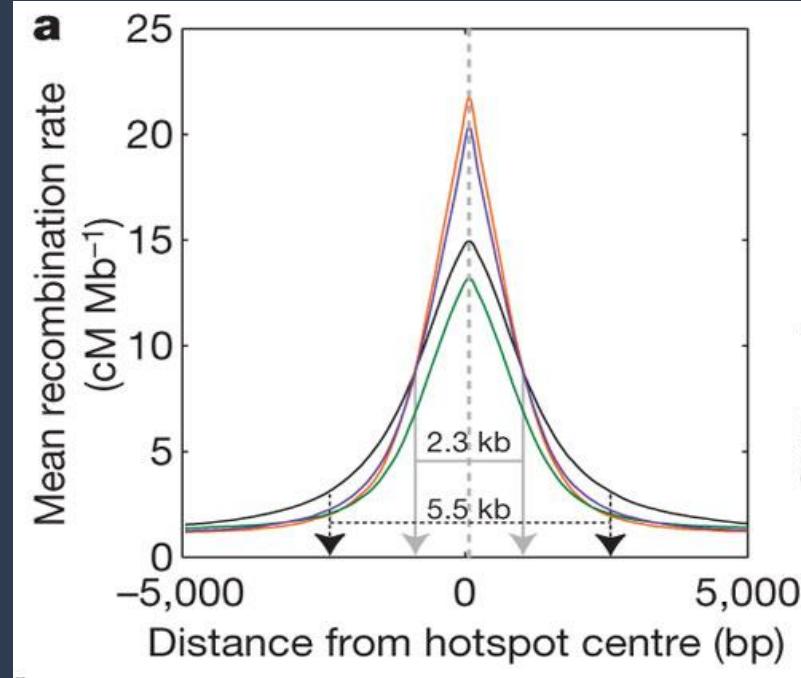
# The distribution of rare and common variants (P1)



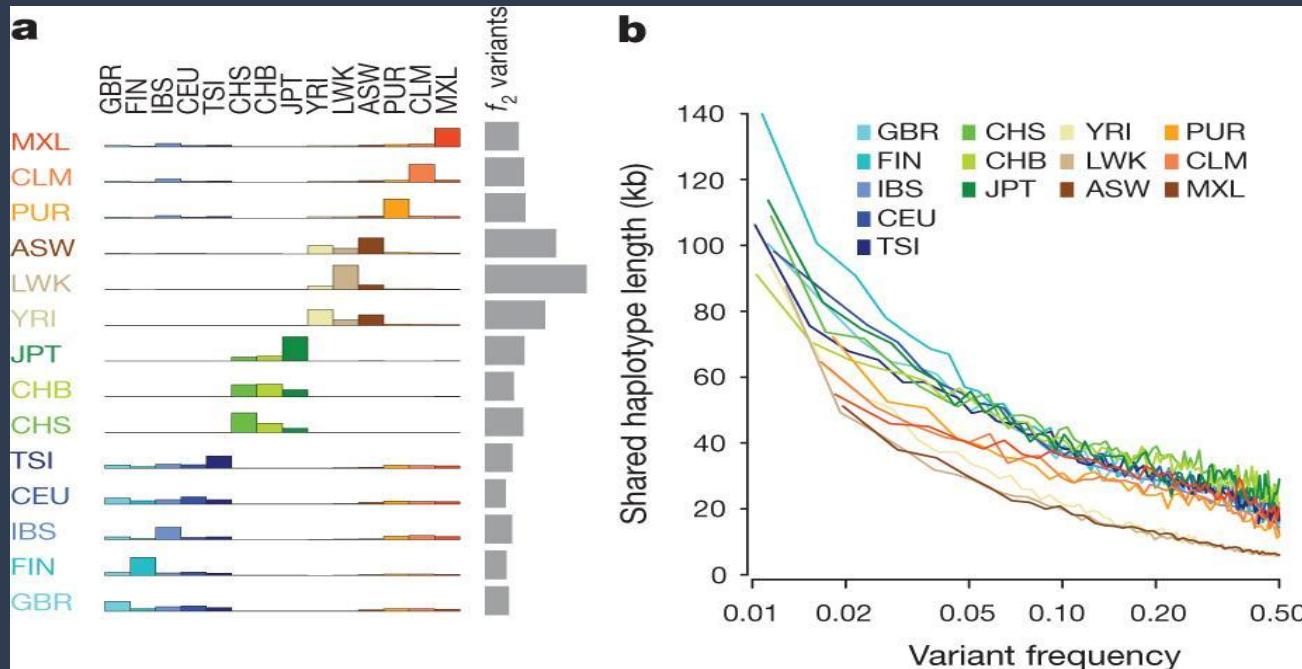
The 1000 Genomes Project Consortium Nature 491, 56-65 (2012) doi:10.1038/nature11632

nature

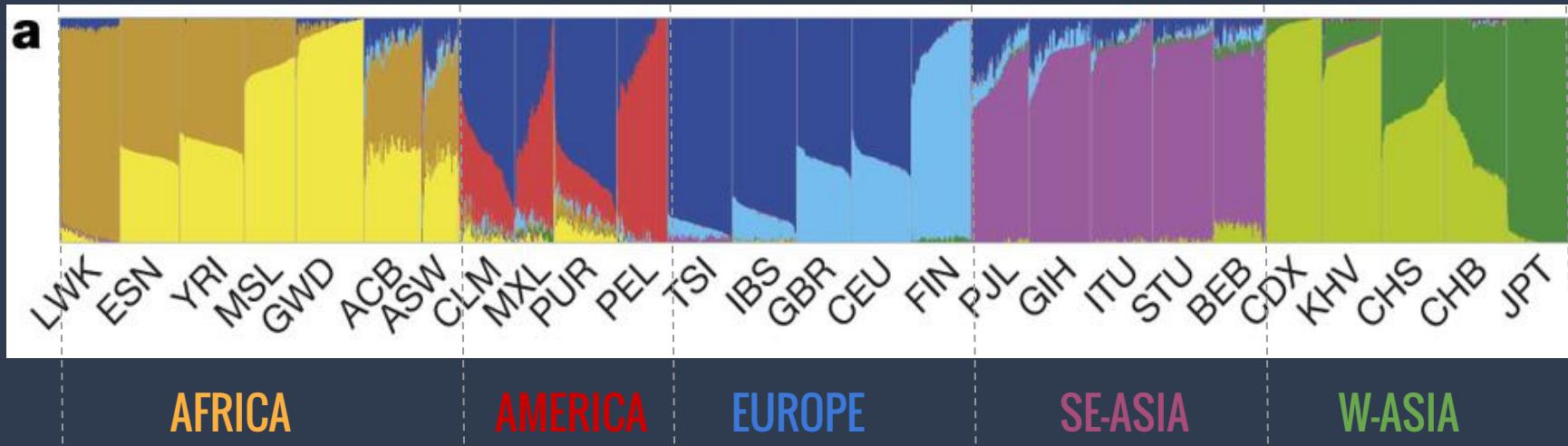
Recombination take place every 2.3 kb on average (Pilot)



# Allele sharing within populations (P1)



# Population structure (P3)



A Auton et al. Nature 526, 68-74 (2015) doi:10.1038/nature15393

# Take home message?

# Future Challenges

Rare variants are often restricted to closely related groups. Many more rare variants are still to be identified.

Research bottleneck has shifted from generation of data to analysis and interpretation.

Make sense of the non-coding regions of DNA

Tease out the links between genetic variation and clinical symptoms.