Population genomics in 2022

#EMBOpopgen

Andrea Manica

UNIVERSITY OF CAMBRIDGE

Dept of Zoology

1

Overview

•**Type of data**

•**Demography**
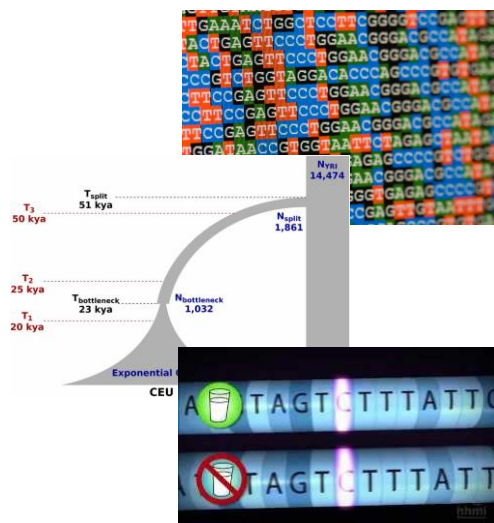
•**Selection**

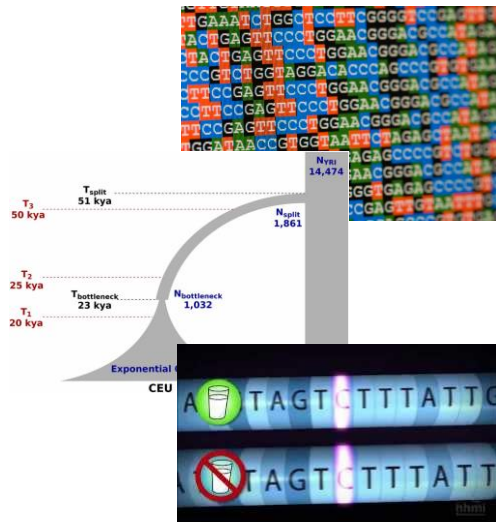2

Overview

•**Type of data**

•**Demography**

•**Selection**

3

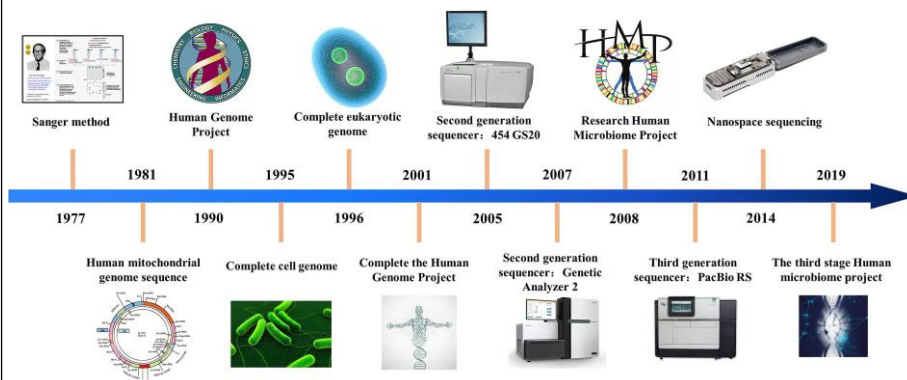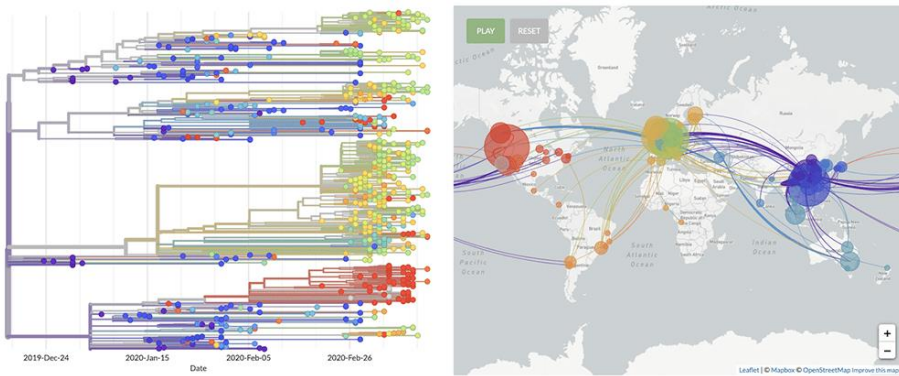# Technological advances over the years

Image from wikipedia

4

The cost of a genome

Image from NIH

5

Genomics on a big scale: an example
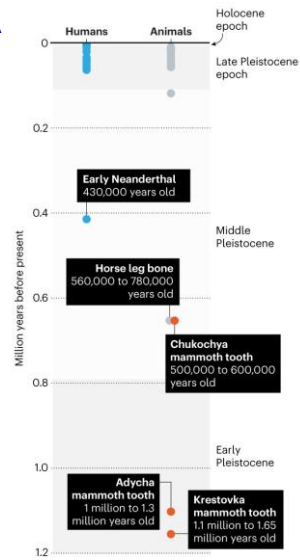

Tracing covid through genomics

Image from UKRI

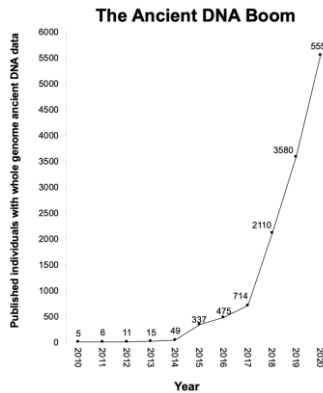6

Ancient DNA

The Ancient DNA Boom

Callaway (2018) Nature

But beware of the ethical issues of aDNA from human remains

Destructive sampling

Callaway (2021) Nature

7

Environmental DNA
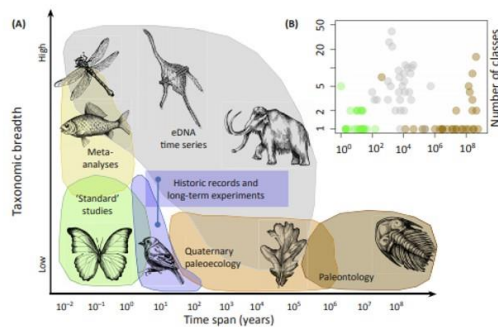
But we need extensive databases for eDNA!!!
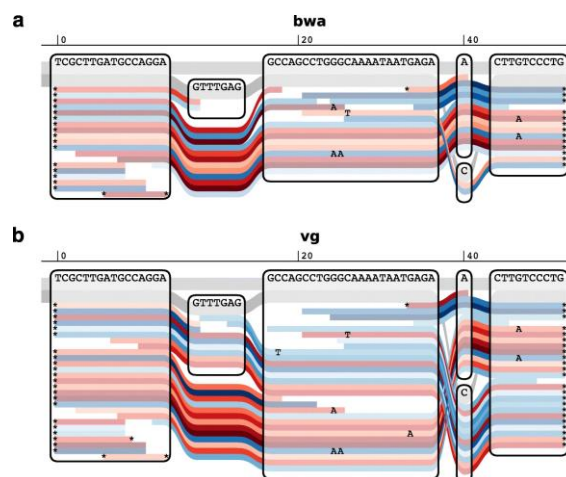
8

Challenges: non-model species

Long and linked reads have dramatically decreased the cost of new reference genomes (~US$ 3k)

A solution for building the extensive databases needed for eDNA?

Low coverage and **genotype likelihoods**

9

Challenges: biases from reference

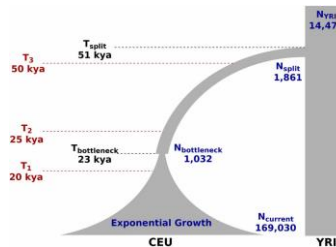Graph based methods hold a lot of promise

10

Overview

•Type of data

•**Demography**

N_YRI
14,474

T_3
50 kya

T_split
51 kya

N_split
1,861

T_2
25 kya

T_bottleneck
23 kya

N_bottleneck
1,032

T_1
20 kya

•Selection

Exponential Growth

N_current
169,030

CEU   YRI

11

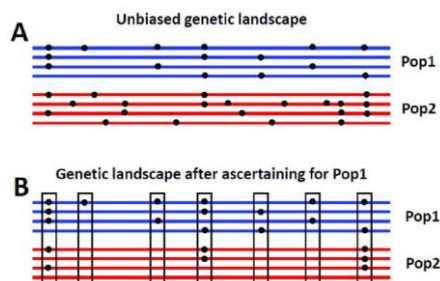
More of the same?

Many approaches based on metrics that describe SNP frequencies (especially the Site Frequency Spectrum)

A   Unbiased genetic landscape

Pop1

Pop2

B   Genetic landscape after ascertaining for Pop1
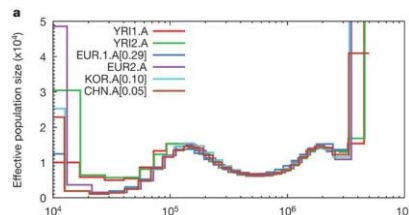
Pop1

Pop2

Very challenging to model ascertained data
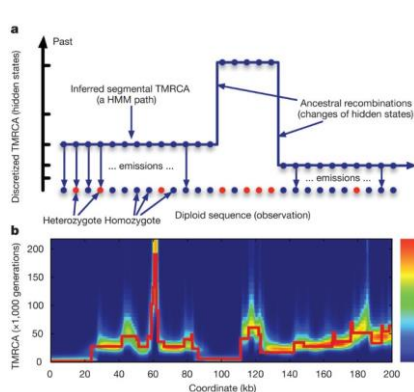
Whole genomes solve the problem of **ascertainment** present in SNP chips!!!
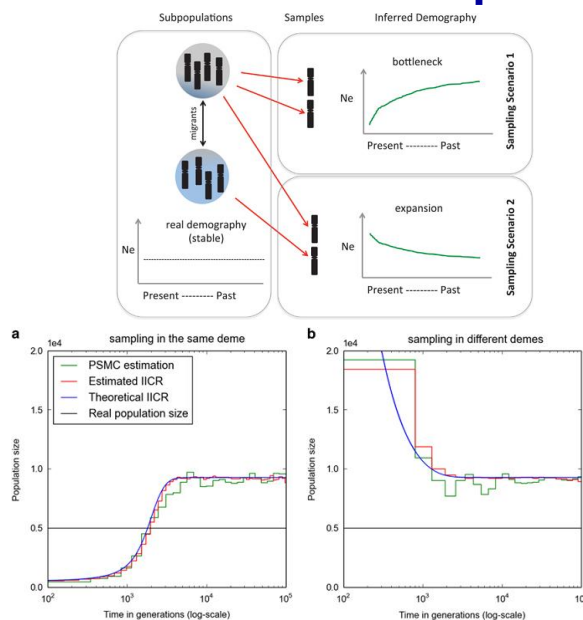
12

The power of whole genomes

But it assumes a single panmictic population (same assumption of Bayesian Skyline Plots)

Be careful about the assumptions behind methods, they can affect your interpretation

Li & Durbin (2011) Nature

13

Beware of assumptions

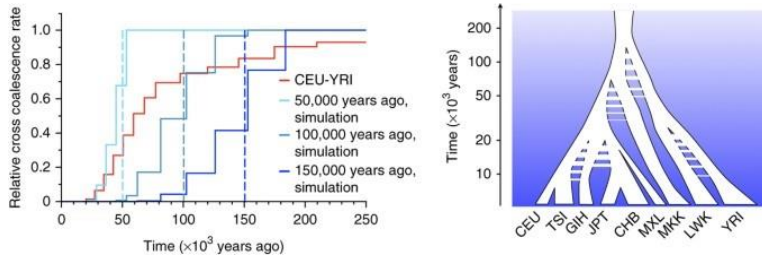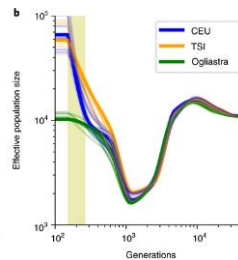Mazet et al (2016) Heredity

14

The power of whole genomes

MSMC extends to multiple genomes (but only a few, and need to be phased…)

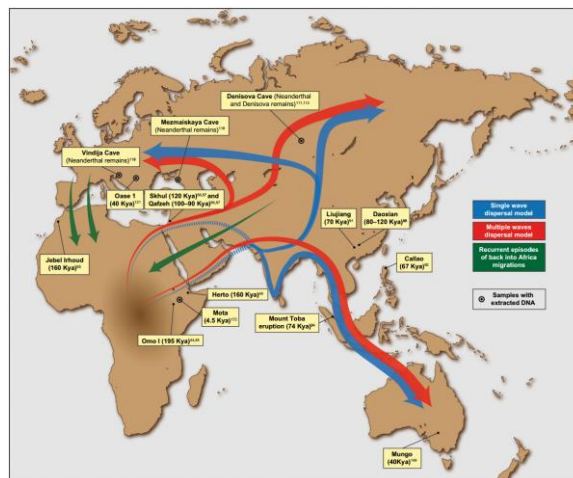Split times assume a clean split without migrations

Methods coming out to deal with 1000s of genomes (SMC++, Relate)

Schiffel & Durbin (2014) Nat Gen, Chiang et al. (2018) Nat Gen

15

But does more data mean better answers?

The story of 3 papers…

16

But does more data mean better answers?

ARTICLE **One wave only!**

doi:10.1038/nature18299

A genomic history of Aboriginal Australia

ARTICLE **One wave (but it could have been two…)**

doi:10.1038/nature18964

The Simons Genome Diversity Project: 300 genomes from 142 diverse populations

A list of authors and affiliations appears at the end of the paper.

LETTER **Two waves!**

doi:10.1038/nature19792

Genomic analyses inform on migration events during the peopling of Eurasia

A list of authors and affiliations appears at the end of the paper

17

Overview

- •Type of data

- •Demography

- •Selection

18

More of the same?

Selection scans and GWAS already possible with SNP chips

19

Rare alleles

Gene = SLC2A9; Chr = 4; Phenotype = Urate; Gene effect size = −0.6
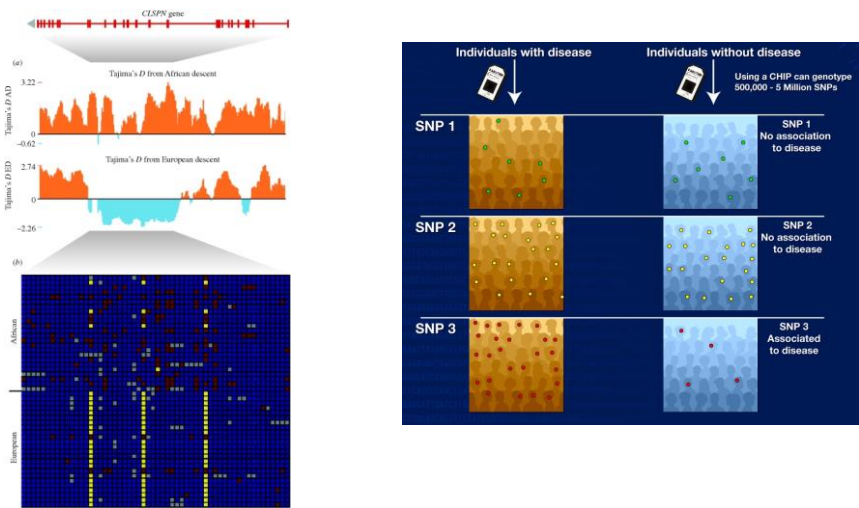
Rare variants from 70k exomes
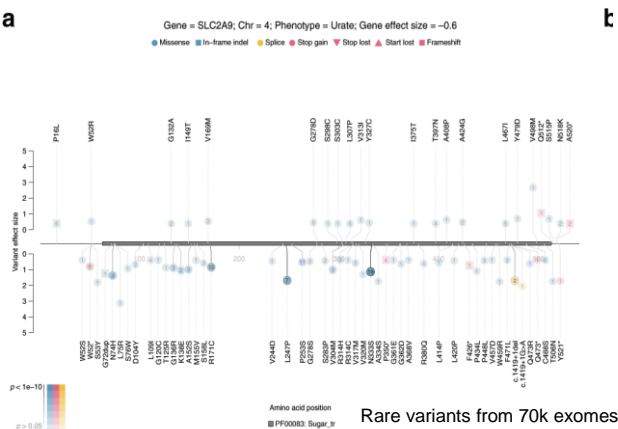
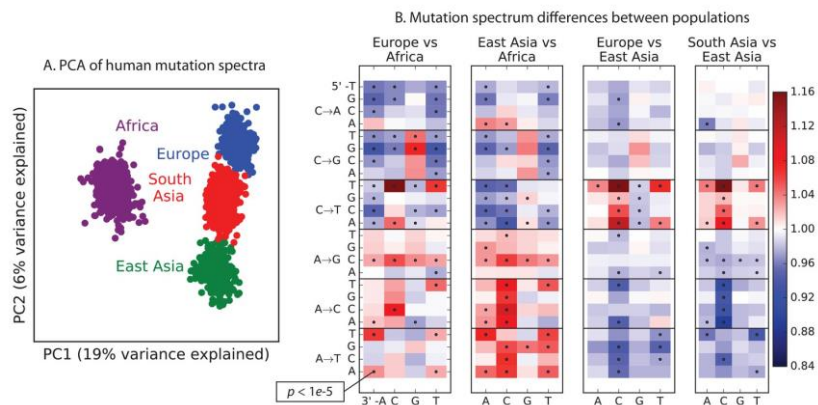But still a big need for functional validation of candidates!
False positives are a big issue in selection studies.

Cirulli et al (2020) Nature

20

Mutation spectrum

A. PCA of human mutation spectra

B. Mutation spectrum differences between populations
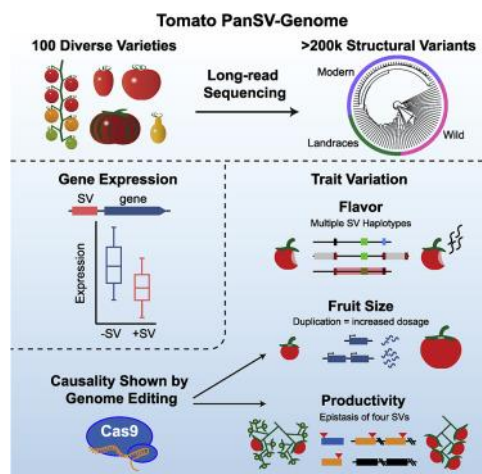
Harris & Pritchard (2017) Elife

21

Structural variation

Long read sequencing helps a lot with structural variation!

Alonge et al (2020) Cell

22

Metagenome

Phenotypes might be greatly affected by other organisms

Zhang et al (2016) IEEE proceedings

23

Big data and Machine Learning

A black box?
Not necessarily

Trends in Genetics

CellPress
REVIEWS

Review

Opening the Black Box: Interpretable Machine Learning for Geneticists

Christina B. Azodi,[1,2,*] Jiliang Tang,[3] and Shin-Han Shiu [1,4,*]
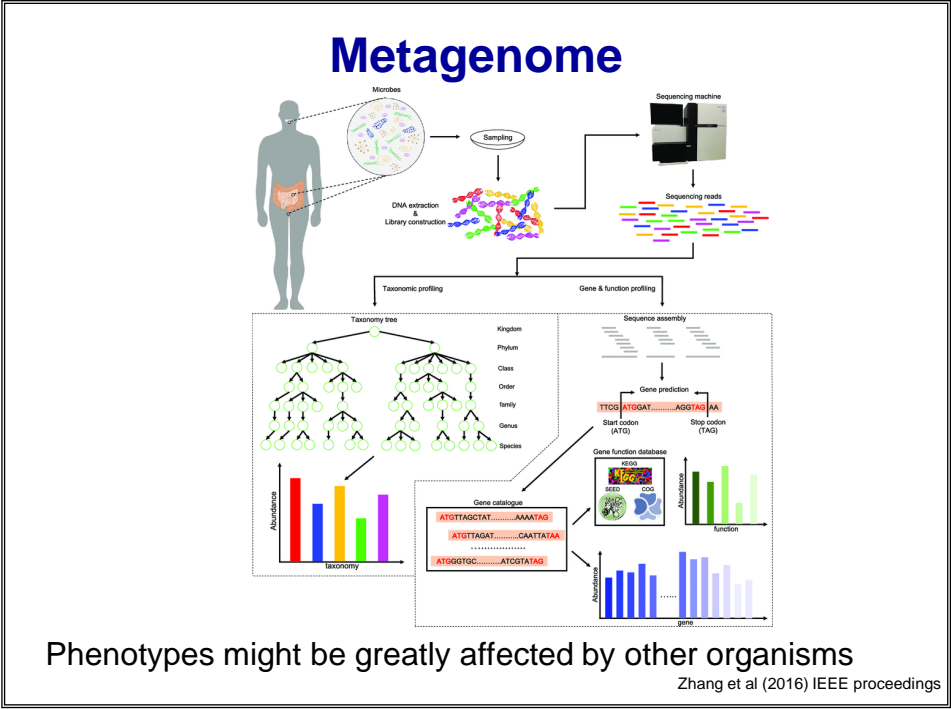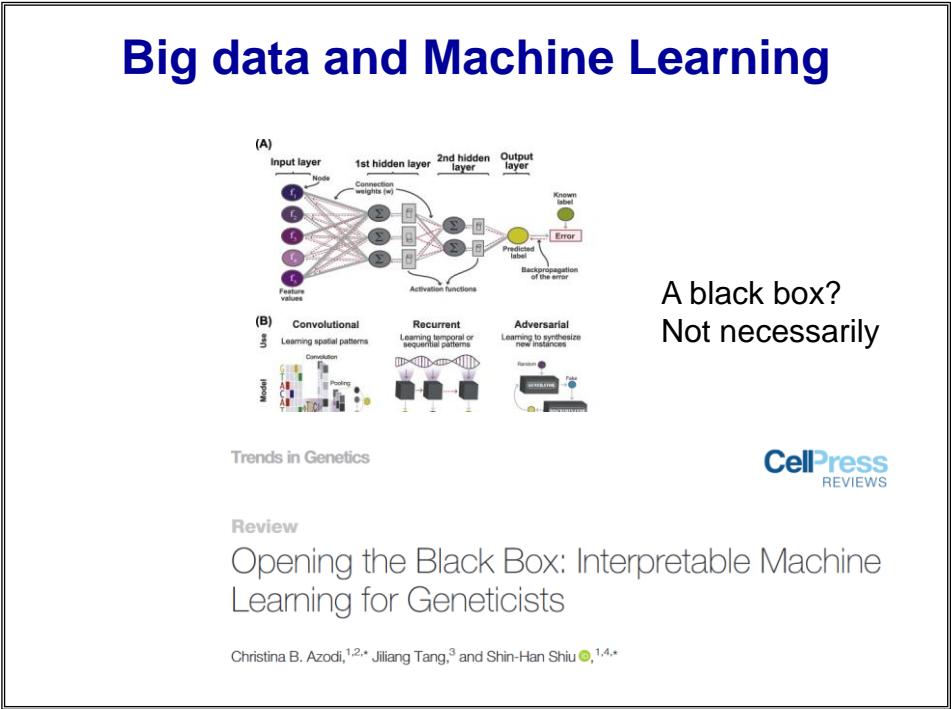
24

Summary

•**Type of data**

•**Demography**

•**Selection**
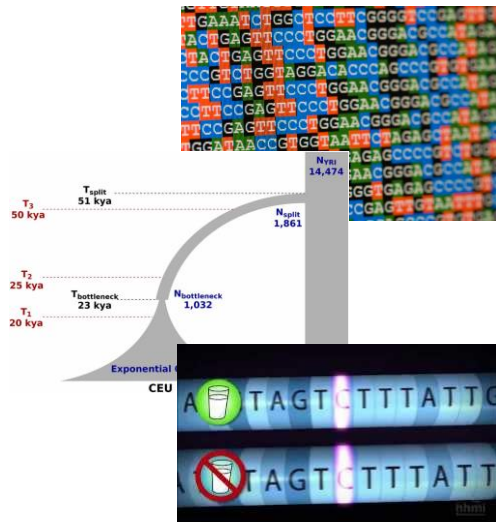
25

EMBO Population Genomics 2022

Amy Goldberg — Duke University, US
Anders Albrechtsen — Copenhagen University, DK
Nina Jablonski — The Pennsylvania State University, US
Andrea Manica — University of Cambridge, UK
Anna-Sapfo Malaspinas — University of Lausanne, CH
Benjamin Peter — Max Planck Institute for Evolutionary Anthropology, DE
Dan Schrider — University of North Carolina at Chapel Hill, US
Leo Speidel — University College London and the Francis Crick Institute, UK
Daniel Wegmann — University of Fribourg, CH
Kirstie Whitaker — The Alan Turing Institute, UK

26