

Population genomics in 2021



Andrea Manica



UNIVERSITY OF
CAMBRIDGE

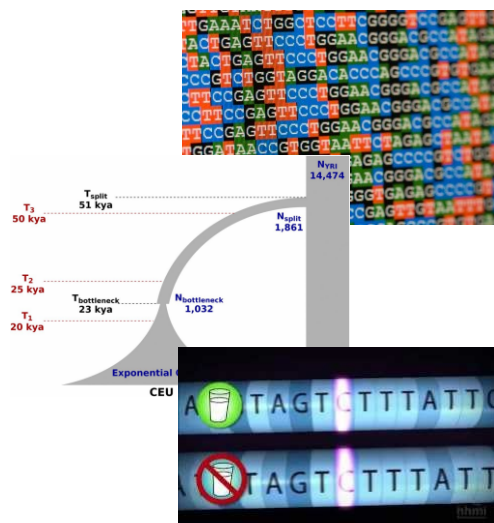
Dept of Zoology

Overview

- Type of data

- Demography

- Selection

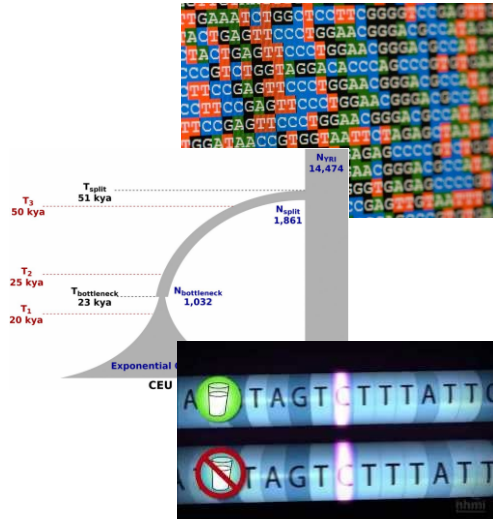


Overview

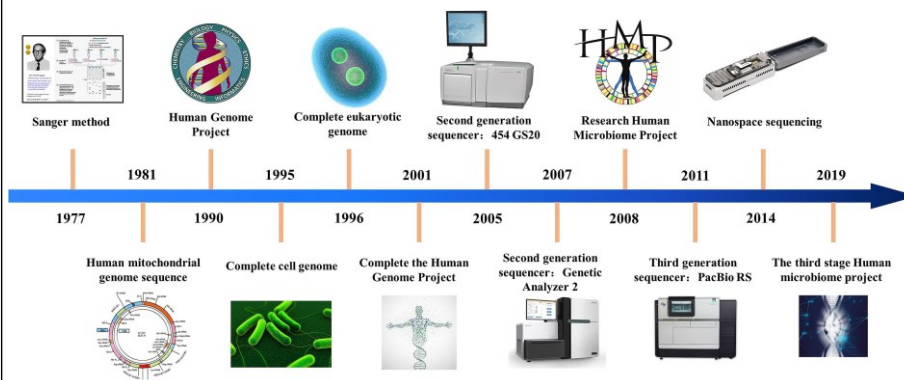
•Type of data

•Demography

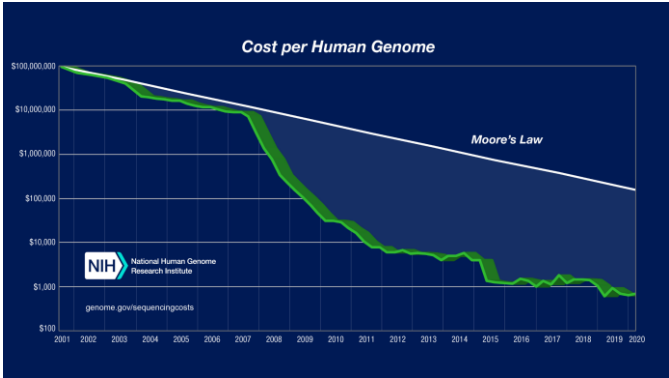
•Selection



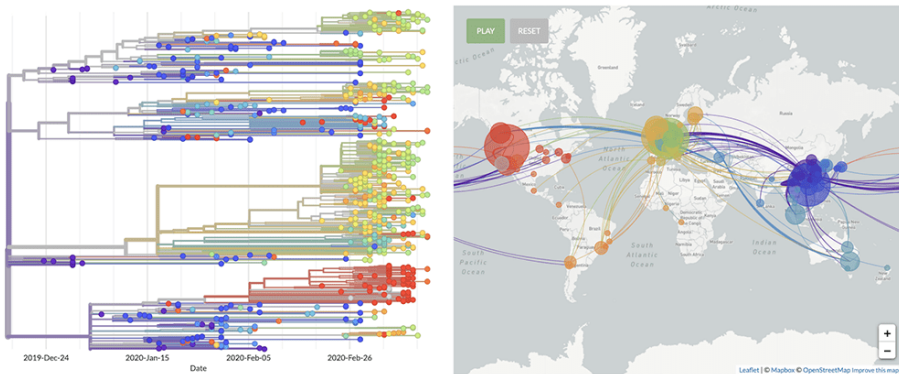
Technological advances over the years



The cost of a genome



Genomics on a big scale: an example



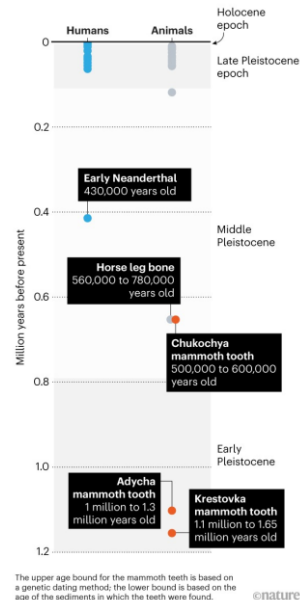
Tracing covid through genomics

Ancient DNA

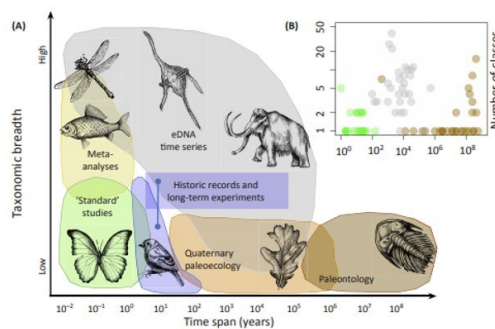


But beware of the ethical issues of aDNA from human remains

Destructive sampling: capture vs shotgun sequencing



Environmental DNA



But we need extensive databases for eDNA!!!

Challenges: non-model species

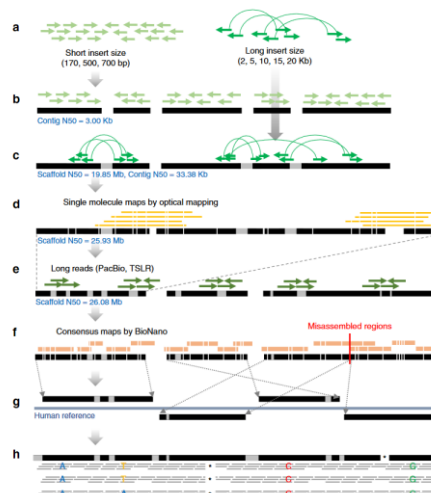


Long and linked reads have dramatically decreased the cost of new reference genomes (~US\$ 3k)

A solution for building the extensive databases needed for eDNA?

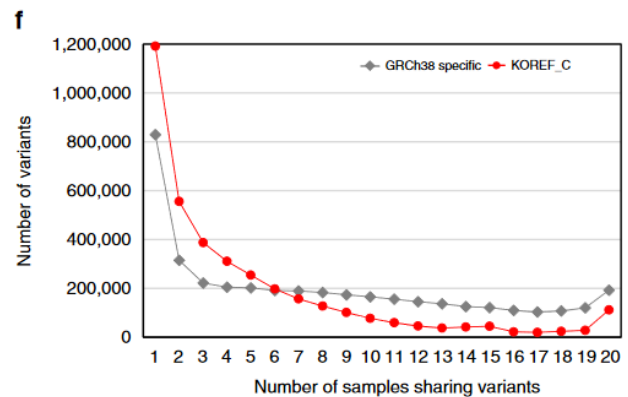
Low coverage and **genotype likelihoods**

Challenges: biases from reference



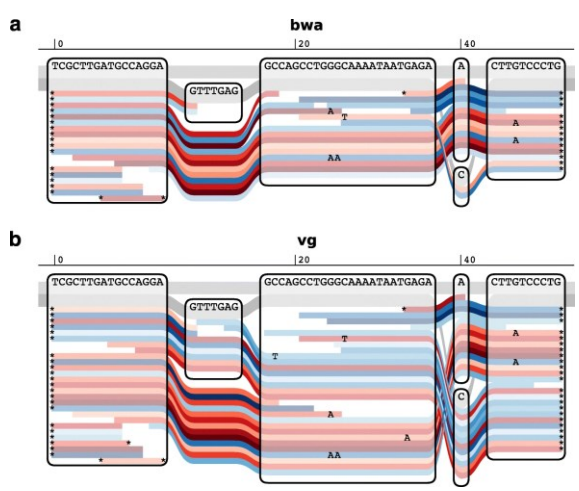
A Korean reference genome

Challenges: biases from reference



The only advantage is storage saving...

Challenges: biases from reference



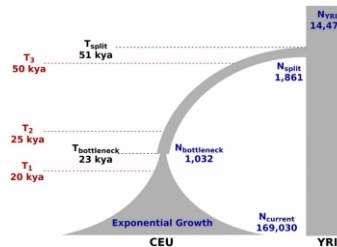
Graph based methods hold a lot of promise

Overview

- Type of data

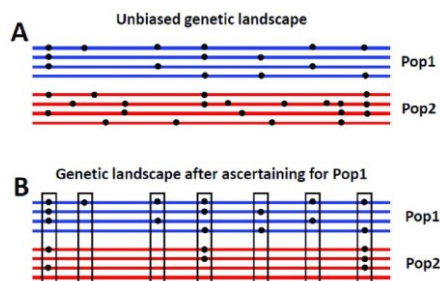
- Demography

- Selection



More of the same?

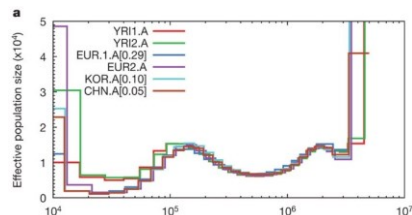
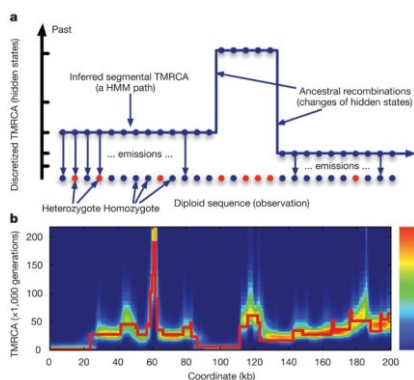
Many approaches based on metrics that describe SNP frequencies (especially the Site Frequency Spectrum)



Very challenging to model ascertained data

Whole genomes solve the problem of **ascertainment** present in SNP chips!!!

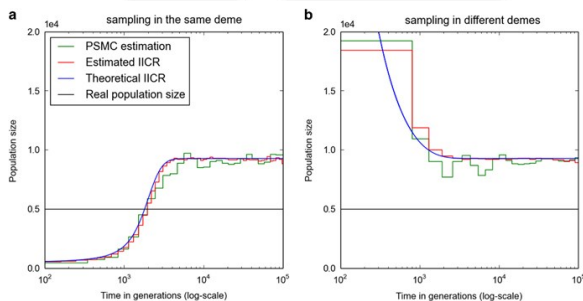
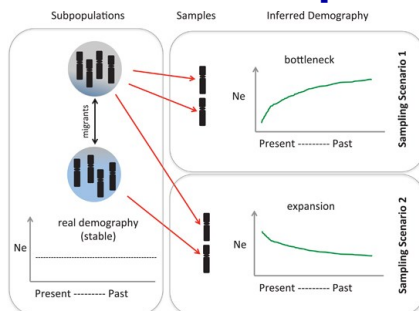
The power of whole genomes



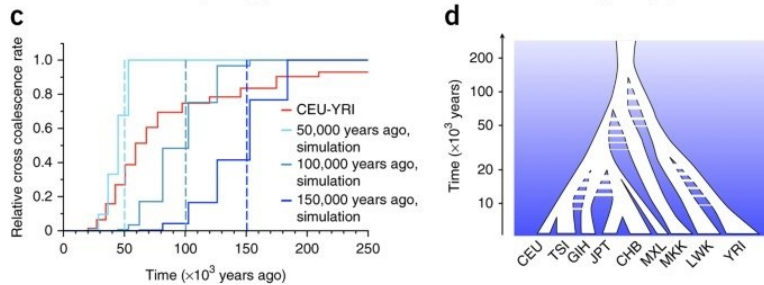
But it assumes a single panmictic population (same assumption of Bayesian Skyline Plots)

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

Beware of assumptions



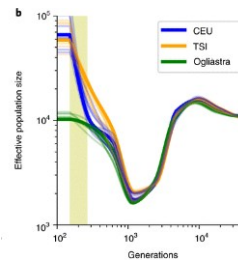
The power of whole genomes



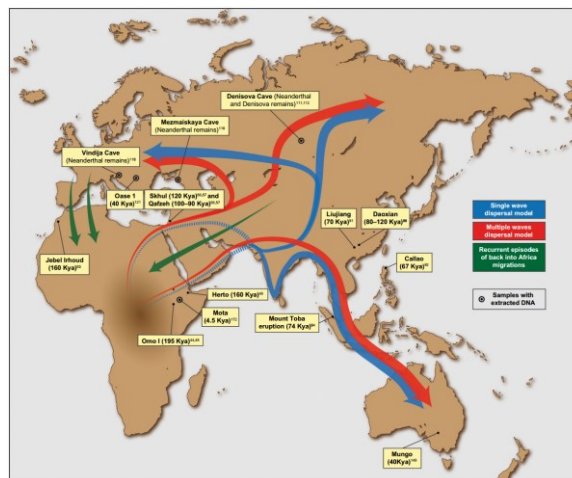
MSMC extends to multiple genomes (but only a few, and need to be phased...)

Split times assume a clean split without migrations

Methods coming out to deal with 1000s of genomes



But does more data mean better answers?



The story of 3 papers...

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.

Overview

•Type of data

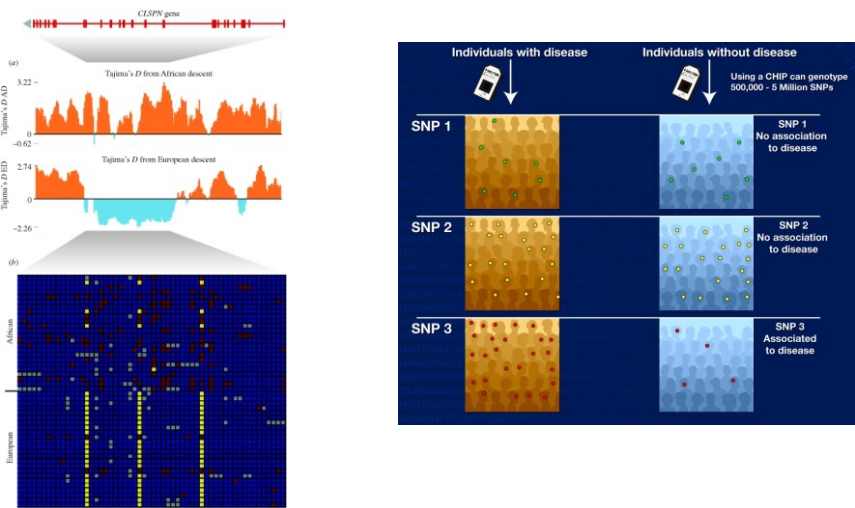
•Demography

•Selection

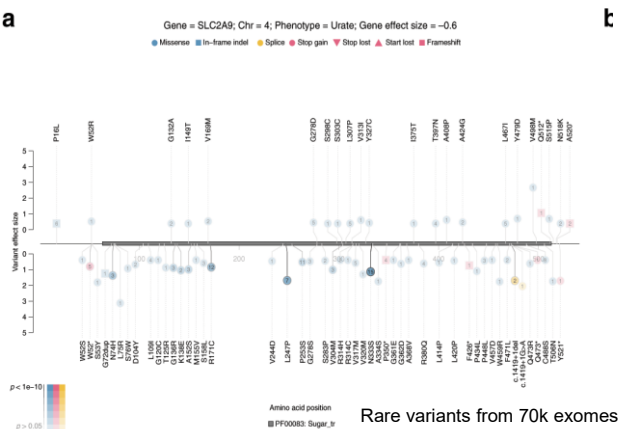


More of the same?

Selection scans and GWAS already possible with SNP chips

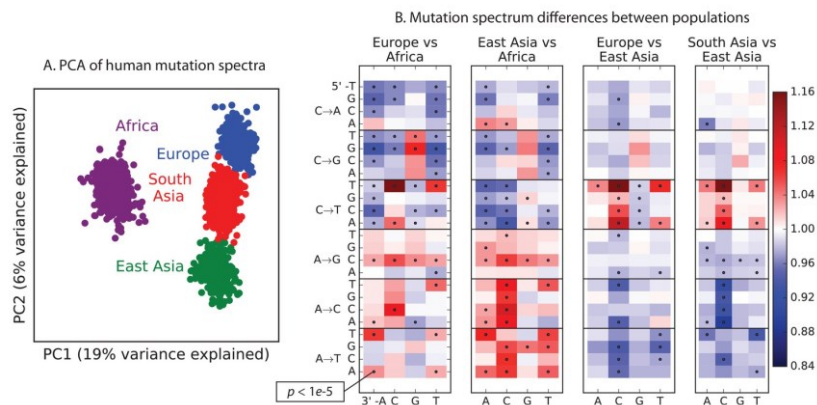


Rare alleles

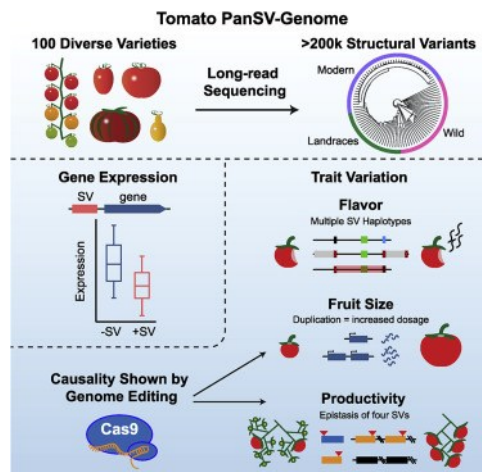


But still a big need for functional validation of candidates!
False positives are a big issue in selection studies.

Mutation spectrum

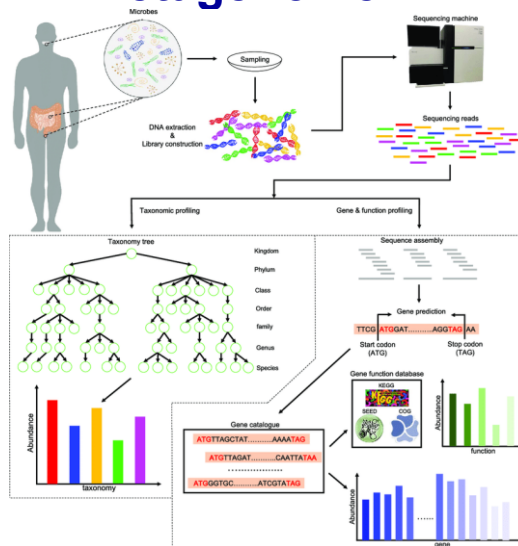


Structural variation



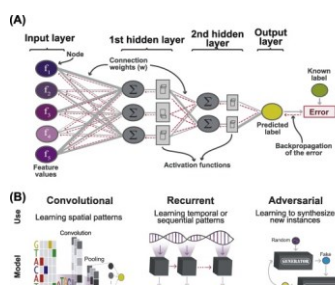
Long read sequencing helps a lot with structural variation!

Metagenome



Phenotypes might be greatly affected by other organisms

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

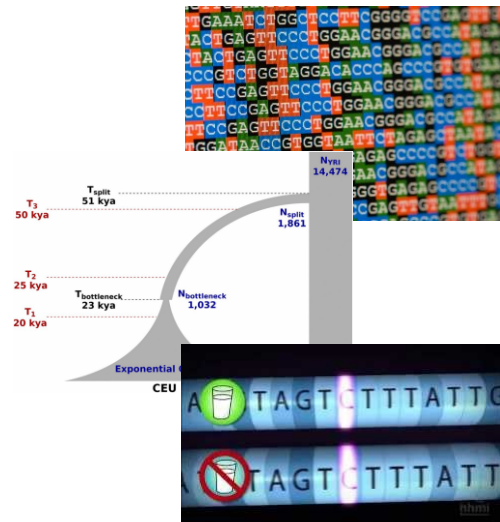
Christina B. Azodi,^{1,2,*} Jiliang Tang,³ and Shin-Han Shiu,^{1,4,*}

Summary

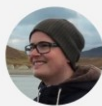
•Type of data

•Demography

•Selection



EMBO Population Genomics 2020/2021



Anders Albrechtsen
Copenhagen University, DK



Anna-Sapfo Malaspinas
University of Lausanne, CH



Andrea Manica
University of Cambridge, UK



Benjamin Peter
Max Planck Institute for Evolutionary
Anthropology, DE



Dan Schrider
University of North Carolina at Chapel
Hill, US



Nicole Soranzo
Wellcome Sanger Institute, UK



Leo Speidel
University College London and the
Francis Crick Institute, UK



Lucy van Dorp
University College London, UK



Daniel Wegmann
University of Fribourg, CH