Modern brains and bones:

Genomic analysis of derived *Homo sapiens* traits

Alejandro G. Muñoz Andirkó
Cognitive Science and Language PhD. program
26-07-2021

UNIVERSITAT DE BARCELONA

Cognitive Biology of Language Group

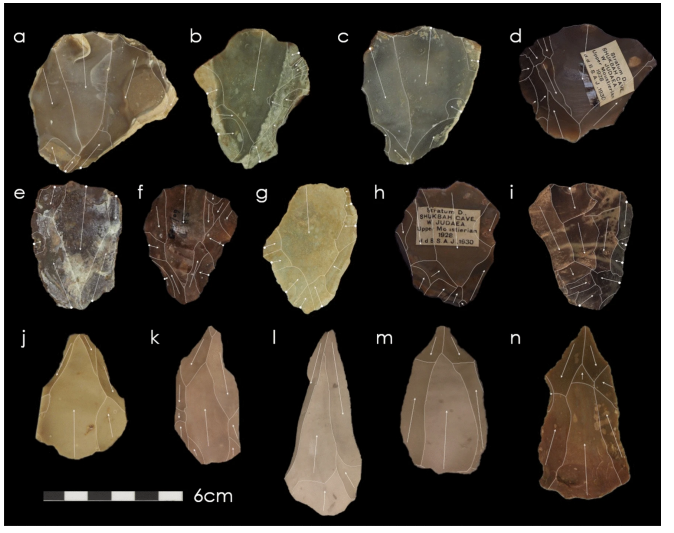
Chapter 1

Human evolution in light of paleogenomics

The paradox of *sapiens* uniqueness

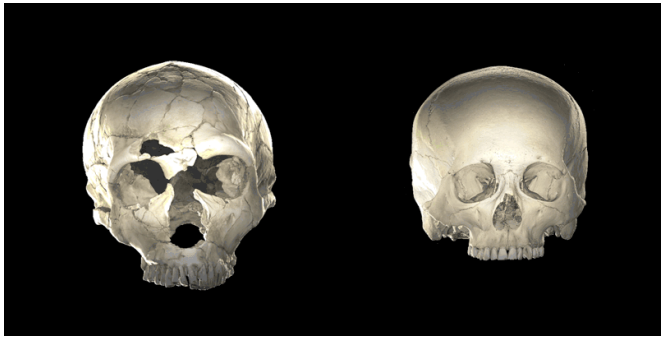
More and more **similarities** between Neanderthals and us

- Several admixture events with both Neanderthals and Denisovans
- Similar levels of cultural complexity:
  - Lithics
  - Pigment use
  - Cave art
  - Body ornamentation



(Nubian Levallois technology associated with Neanderthals, Blinkhorn et al. 2021)

The paradox of *sapiens* uniqueness



(Credit: Phillip Gunz, CC BY-NC-ND 4.0)

... Yet there are important **differences**

- Anatomical differences in endocasts: diverging parietal, cerebellar and (possibly) subcortical morphology
- The existence of deserts of introgression that resisted admixture
- Regions under positive selection

How do we reconcile these two views?

Integrating paleogenomic data with large-scale -omics databases that take advantage of **current variation** in modern populations

Using clinical data from **pathologies** known to be caused by **targeted, restricted** genetic disruption, and affecting phenotypes that have changed over human evolution

... all while taking into account that this data can't be really understood without a **temporal dimension** of variation

Chapter 2

Effects of high-frequency *Homo sapiens*-specific variants in brain gene expression

Questions

- Often, too much emphasis is put on **missense** variants, but is there a role for **regulation** in human evolution, as predicted?
- Can we infer **which tissues and genomic regions** are particularly affected by *Homo sapiens*-specific variation in gene expression?

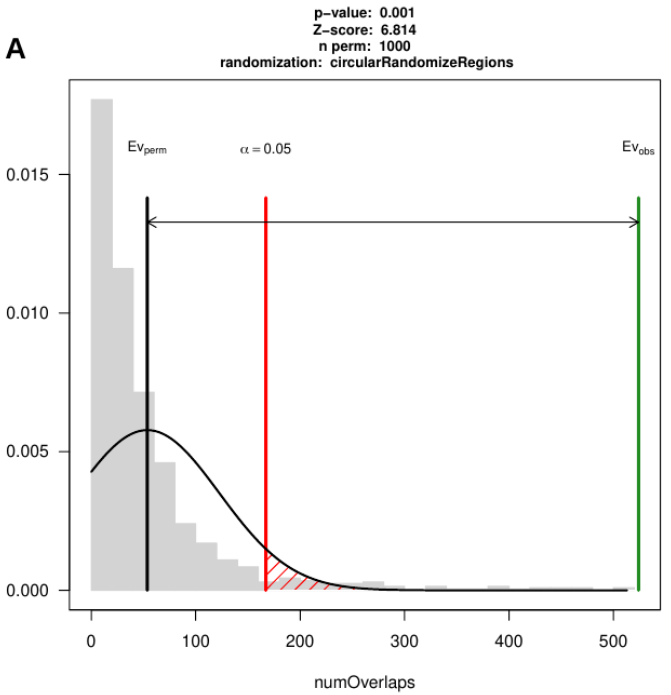
How can we adress these questions?

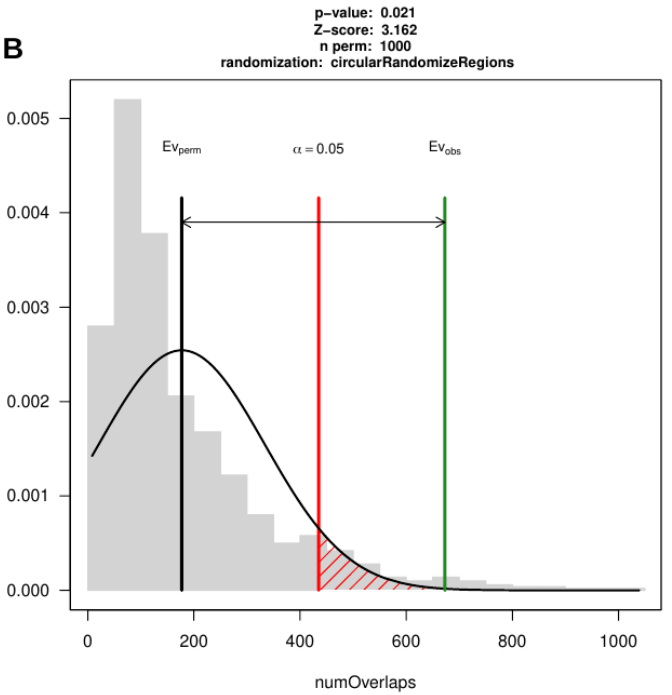
- GTEx single-tissue expression quantitative trait loci (**eQTLs**) across **15 brain structures**.
- Filtered through **Kuhlwilm & Boeckx (2019)**, an exhaustive catalog of variants differentiating modern humans from Neanderthals and Denisovans.
- Regions identified as under **positive selection** in two independent studies.

Results summary

Permutation tests

High-frequency eQTLs are **overrepresented** in genomic regions under positive selection.





Results summary

Tissue-specificity

- In terms of eQTL variance across tissues, the **pituitary** and **cerebellum** accumulate more eQTLs than expected by chance (against control set: non-derived & non high-frequency eQTLs).
- Mendelian randomization shows correlation, not shared eQTL-top GWAS hit signal (in 10 UKBiobank brain region volume studies).

Others

- No directional skewness.
- Overrepresentation of specific functional categories in high frequency derived eQTLs relative to control: **NMD**, **5'-UTR**, **non coding transcript** variants.

Chapter 3

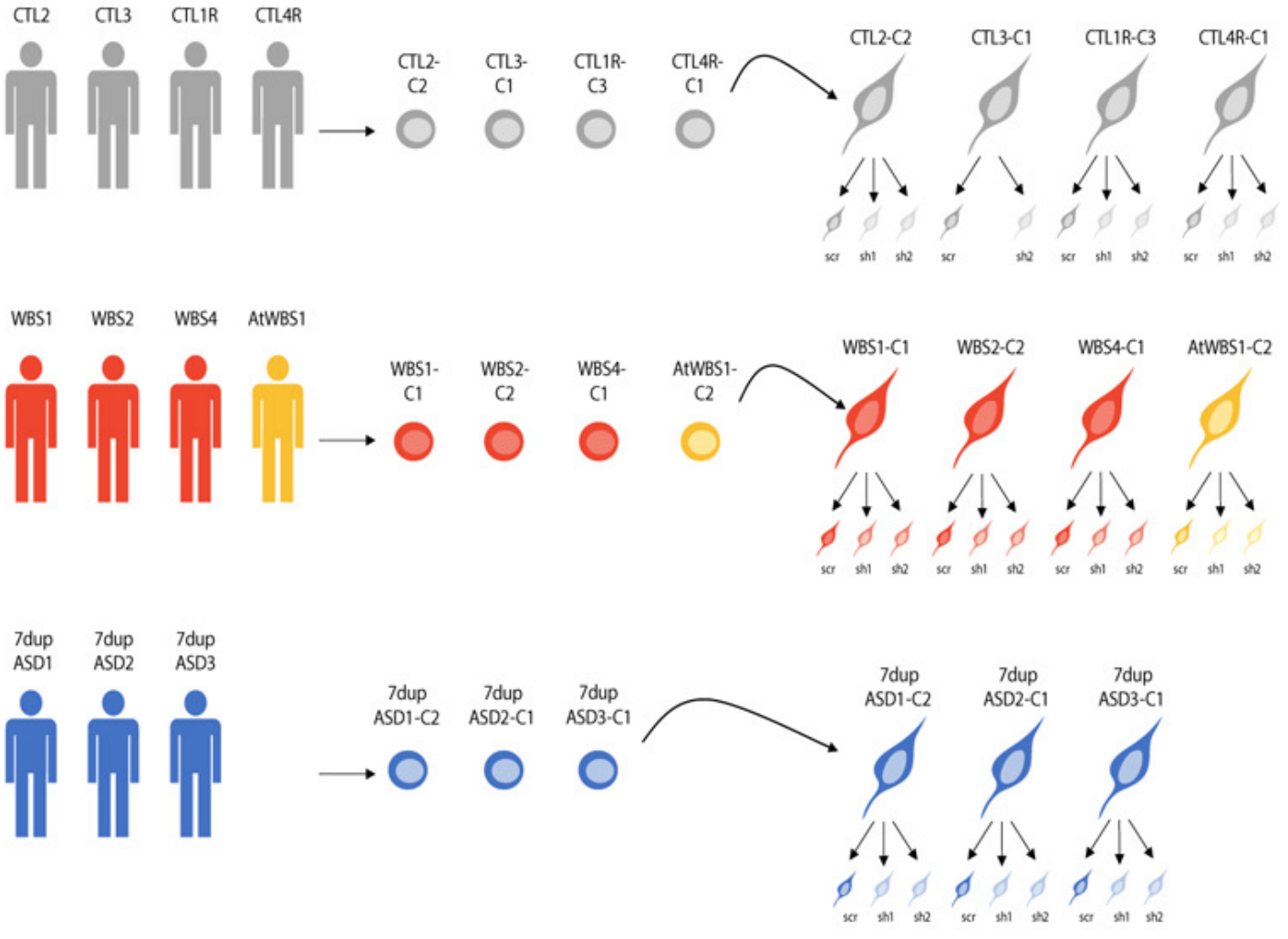
BAZ1B, Williams-Beuren syndrome and the evolution of the human face

Williams-Beuren syndrome

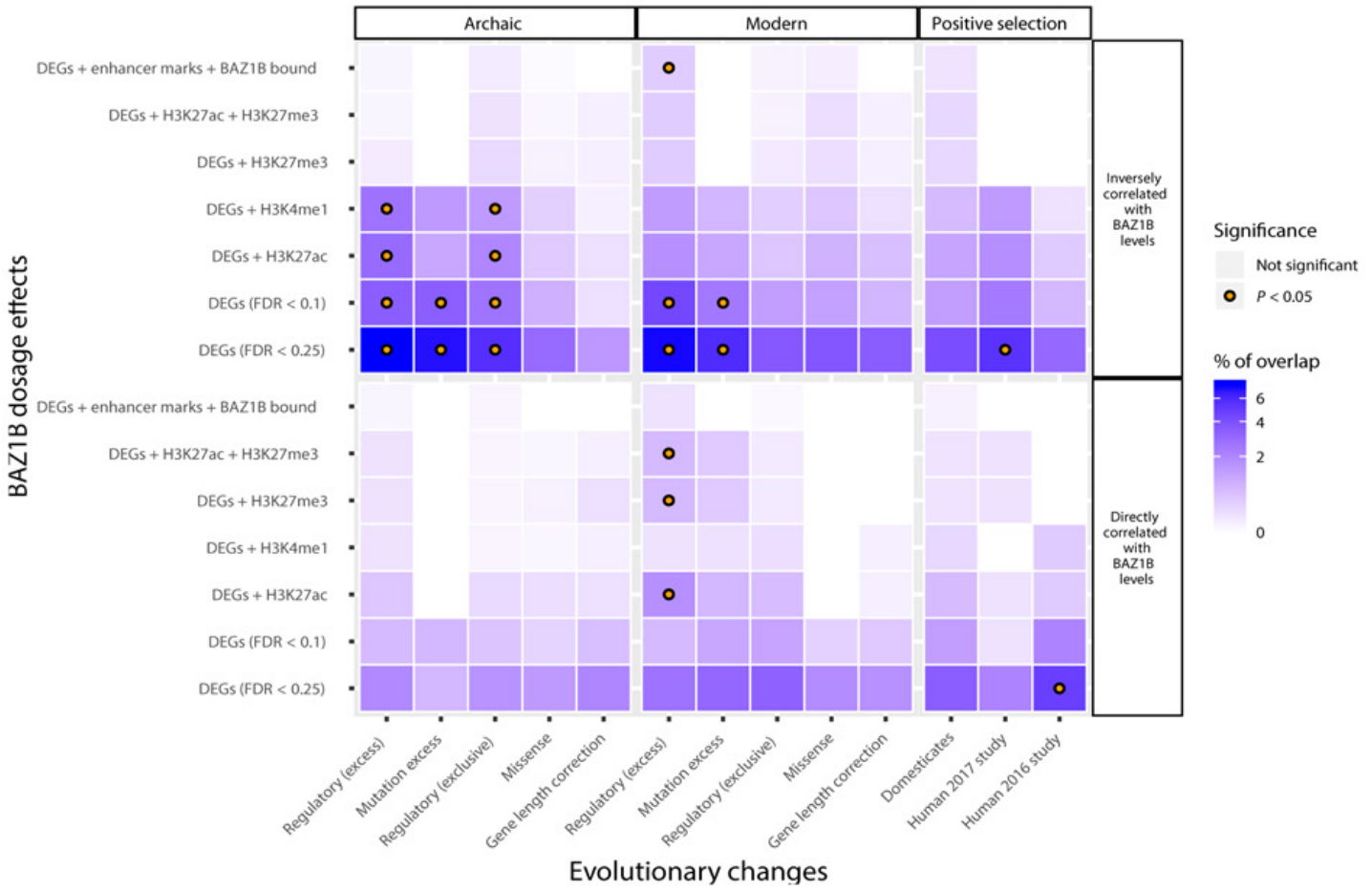
- **Neurocristopathy** caused by the deletion of an specific region in chromosome 7.
- Typically results in a characteristic **craniofacial morphology** reminiscent of *Homo sapiens*-derived traits: eg., retraction of the lower and mid face.
- **Sociocognitive changes** as well: exaggerated social tendencies, increased trust and friendliness, reduced reactive aggression, relatively intact language.
- All of this has led to parallels between Williams-Beuren and the "domestication syndrome", and a potential way to test the **"self-domestication" hypothesis** [1].
- Along with **Giuseppe Testa's lab**, we set to provide the first experimental validation of the hypothesis through the effects of **BAZ1B**.

[1]: Though my current stance on the term and the hypothesis overall has changed, as stated in the introduction.

Cohort and experimental design



BAZ1B-bound genes



Summary of results

The most stringent category of genes following BAZ1B dosage is tied to an **excess of regulatory changes** in *sapiens*.

Neanderthal/Denisovan enrichment in **inversely correlated** gene lists.

Homo sapiens-specific enrichment in regulatory variants **directly correlated** with BAZ1B.

Chapter 4

Temporal mapping of *Homo sapiens* variants

Questions

Most human evolution studies disregard the temporal dimension of their results.

Simulation-based age estimates are usually dependent on unkown demographic parameters.

Can we derive a temporal dimension of variants in an evolutionary scale?

Albers and McVean (2020)

- Non-parametric method - theoretically agnostic to demography and selection.
- Estimated ages of 45 million of variants.
- Hidden Markov model + pairwise coalescent-based TMRCA estimation, using recombination, mutation and joint-model clocks.

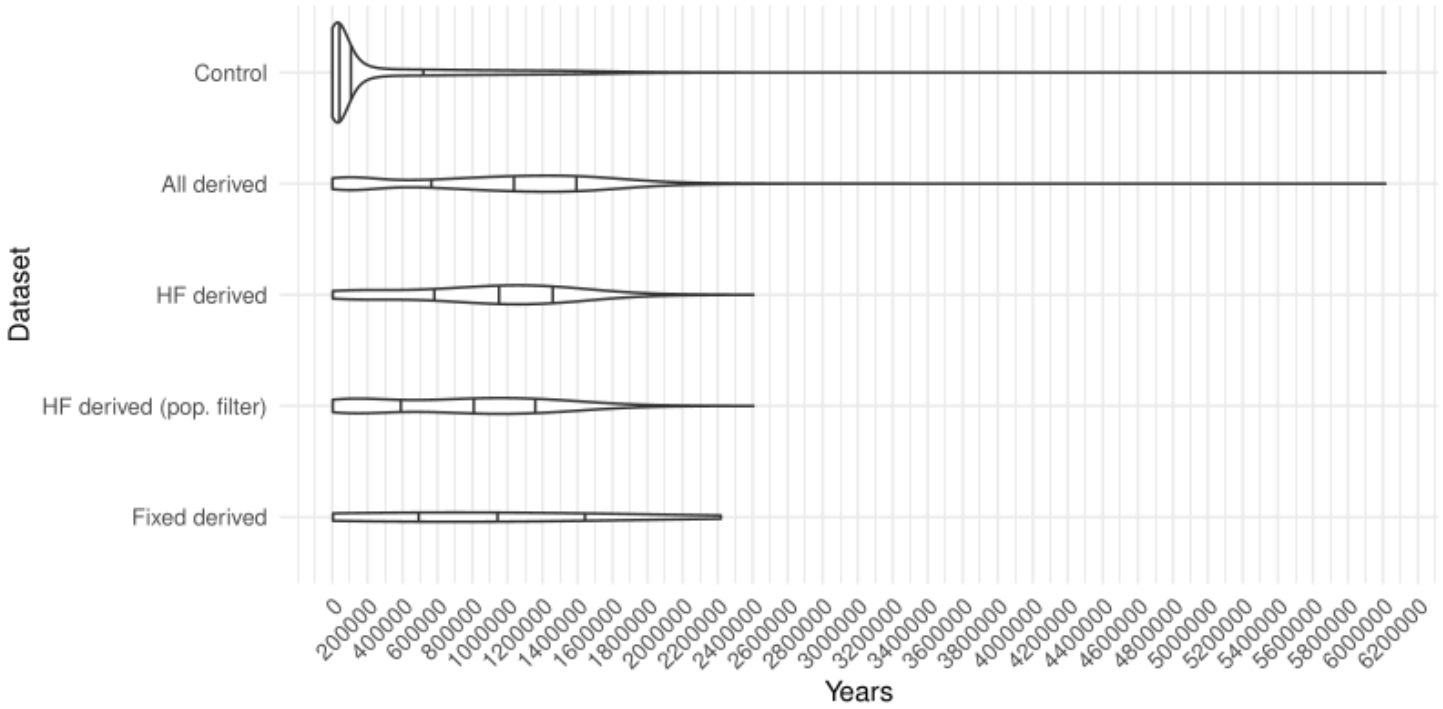
Downstream analysis

- **ExPecto**: deep learning, variant-based *in silico* predictor of expression variation [1].
- **Time-sensitive GO** enrichment analysis.
- **Case example**: *BAZ1B* bound genes.

[1]: Through a convolutional neural network train on "cell-type-specific model for 2,002 genome-wide histone marks, TF-binding and chromatin accessibility profiles"

Summary of results

Two mode distribution in derived variant sets



Summary of results

ExPecto:

We found genes related to cerebellar Purkinje neurons (0-60k), glutamate system (300-500k), dopamine system (500-800k)


Type of tissue:
Brain−related
Not brain−related
0
−2
−4
Sum of variant predicted expression (logFC)
0−300kya
300−500kya
500−1mya
Time window

Summary of results

ExPecto:

We found genes related to cerebellar Purkinje neurons (0-60k), glutamate system (300-500k), dopamine system (500-800k)

GO enrichment:

Facial shape category found in 300-500k, consistent with our vision of the archaeological record, among other things.

BAZ1B:

At the same time, older variation in genes of the regulatory network of *BAZ1B* than expected initially.

Conclusion

Each of these three studies **complement each other** in perspective, and ask similar questions at heart - that is, trying to determine what really differentiates *Homo sapiens* from Neanderthals and Denisovans.

While the answers in this thesis are partial, they aim at all points to derive information from open data while **doing justice to the increasing complexity of our species' history**.

An **interdisciplinary view** of human evolution is key to derive conclusions from all dimensions of the paleogenetic data.

Thank you!

Extra slides

eQTL

Q: Permutations and high frequency

- Both selection studies rely on frequency parameters. Their results give by definition **regions with overall higher frequency than expected**
- But it makes sense that we would find - that's a fair point.

Some solutions:

- Controls with random non-eQTL variants at similar frequencies.
- Working backwards: from the regions under positive selection, which span large windows, identify the diverging eQTL, as Racimo did, and people now do with gene regulation inferences of aDNA.

Q: Permutations and clumping

- High overlap might be driven by a few eQTL.

Possible solution:

- Maybe including a variant outside the region if it is in LD with an eQTL inside.

Q: Types of functional variants in eQTL project

- Control: non-HF GTEx variants in brain tissues.

BAZ1B

Q: The "domestication syndrome"

Promises

- Comparative genomics across different species models
- A localized, testable hypothesis, centered on neural crest cells
- An expansion of the term of "domestication" as a biological process, rather than a cultural practice

and practical problems

- Not consistent across species (except for reduced reactive aggression)
- Unknown relationship with other components of domestication (eg., glutamatergic system)
- Pleiotropy and the overall general role of neural crest cells in early development

Q: Lists used

From suplementary table 2A. Both lists apply to Neanderthal/Denisovan compared to modern humans and viceversa.

- **Regulatory (excess)**: Top 10% genes harboring high frequency regulatory mutations as defined in Kuhlwilm and Boeckx (2018).
- **Mutation excess**: Top 10% genes harboring high frequency mutations as defined in Kuhlwilm and Boeckx (2018).
- **Regulatory (exclusive)**: Genes harboring lineage-specific high frequency regulatory mutations (only in Neanderthal/Denisovans but not modern humans or viceversa).
- **Missense**: Genes harboring high frequency missense mutations.
- **Gene length correction**: Top 5% genes harboring high frequency mutations divided by their genomic lengths.

Temporal mapping

Q: Why are derived variants so old?

Some proposals:

- We don't have enough archaic genomes to reliably determine ancestrality.
- The model by Albers & McVean is actually is actually reflecting demographic history (despite their claims).
- Mutation rate, recombination or the populations used for the pairwise coalescence clock are skewing the picture.

> Although several high-resolution recombination maps exist for European-descent populations, the recombination landscape of African populations remains relatively understudied. Given that there is high genetic divergence among groups in Africa, **it is possible that recombination hotspots also diverge significantly**.
>
> van Eeden et al. 2021

Q: Why those time divisions?

- Originally organized that way due to computational limits to what we could do we ExPecto, with what we thought was the added advantage of a finer resolution.
- A new, revised version will most likely include only three periods to help with clarity:

    1. **0-300 thousand** years ago.
    2. **300-500 thousand** years ago.
    3. **500-1,000 thousand** years ago.