



Detecting recent selective sweeps

Joana Meier

Tree of Life Programme, Wellcome Sanger Institute

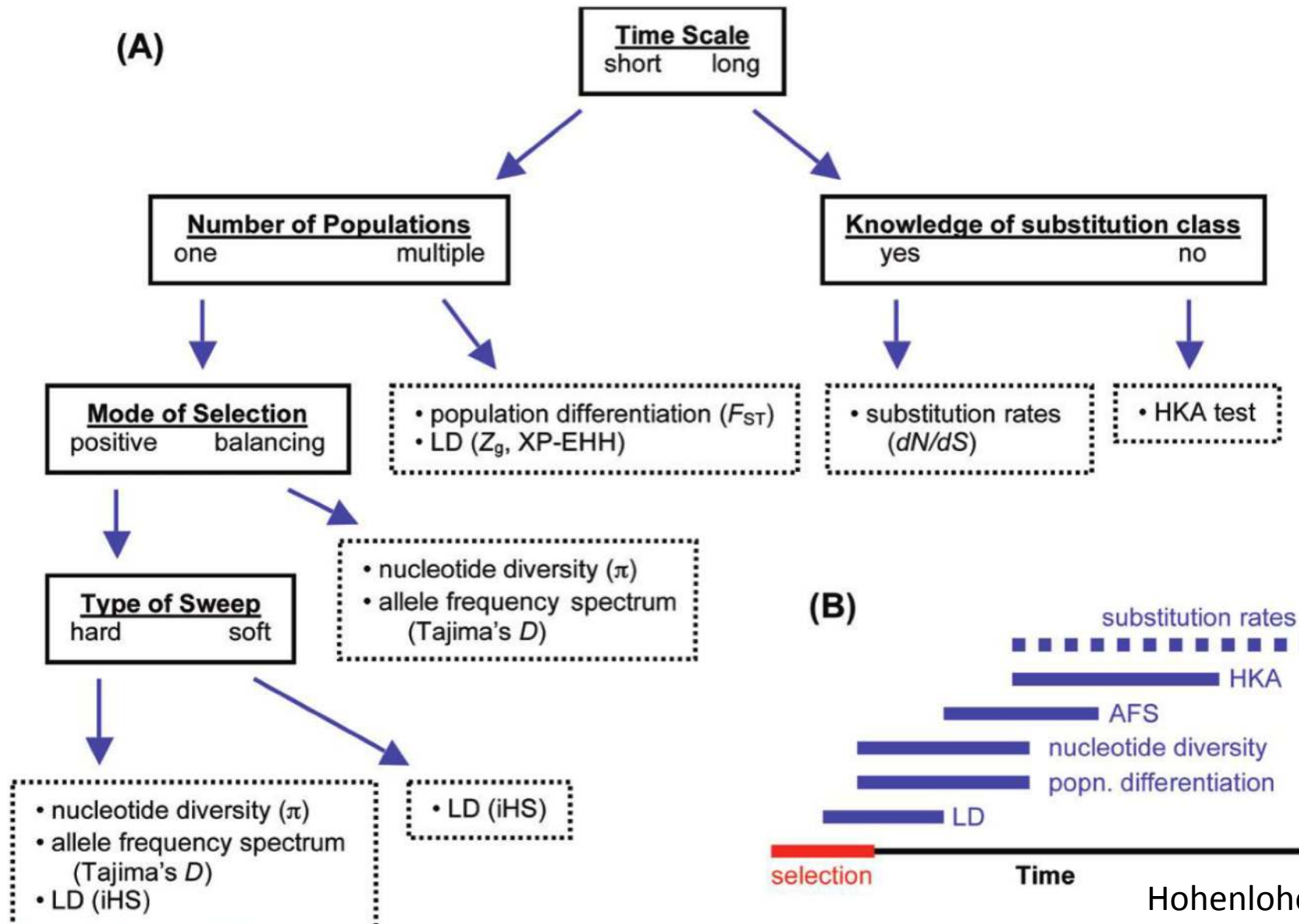
THE
ROYAL
SOCIETY

The
Branco Weiss
Fellowship
Society in Science



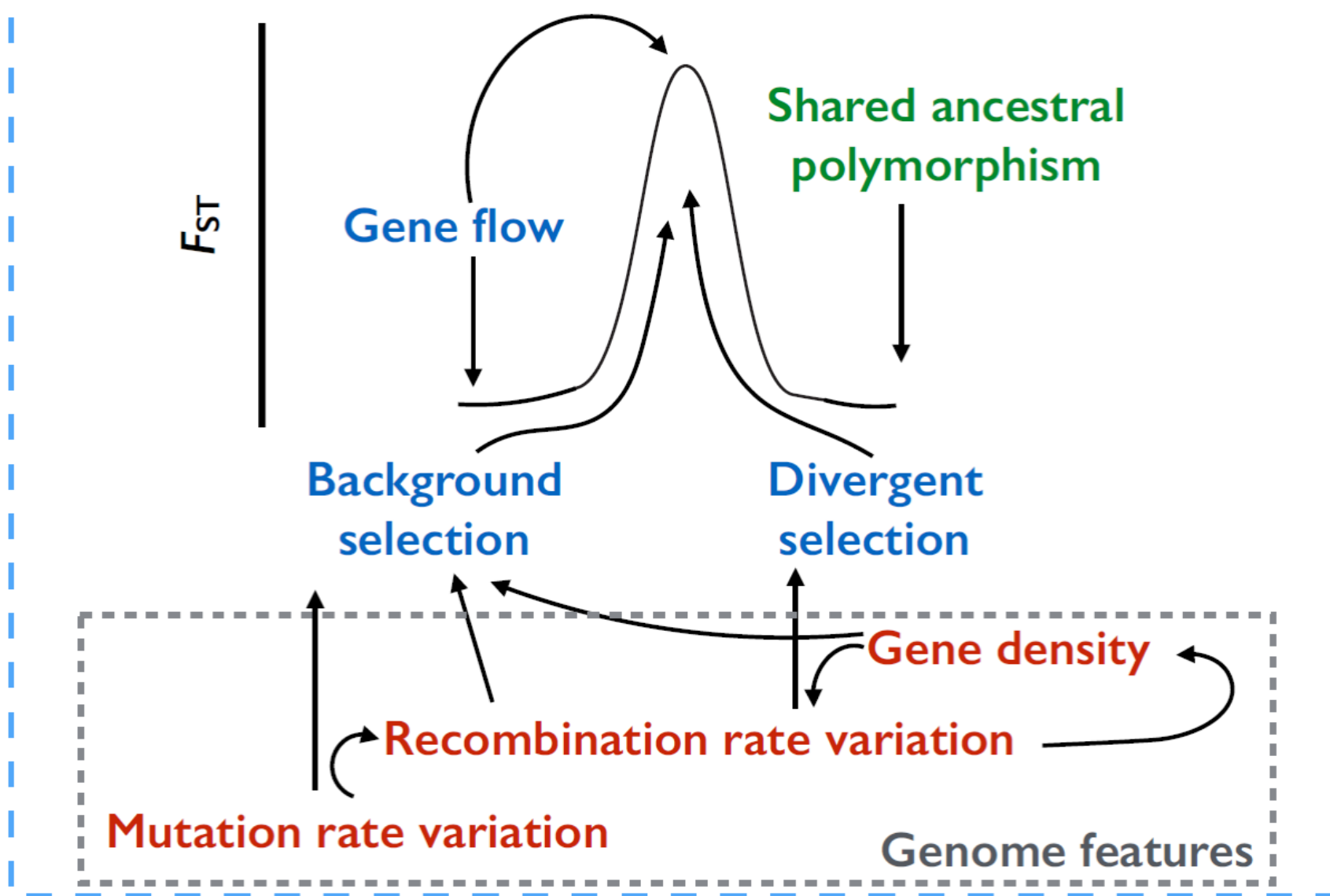
UNIVERSITY OF
CAMBRIDGE

Detecting selection

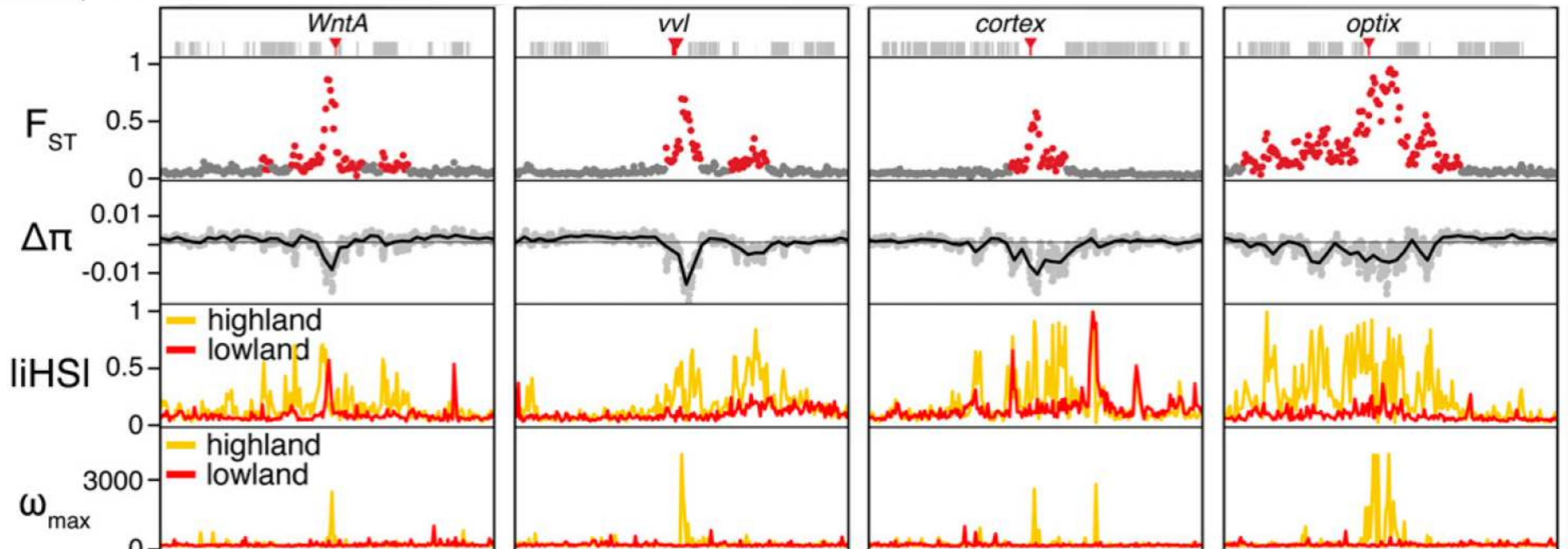
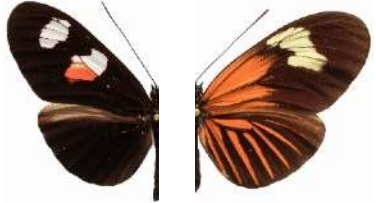


Confounding factors

Demographic history



Combining different genome scans to test for selection



Genetic signatures of selective sweeps

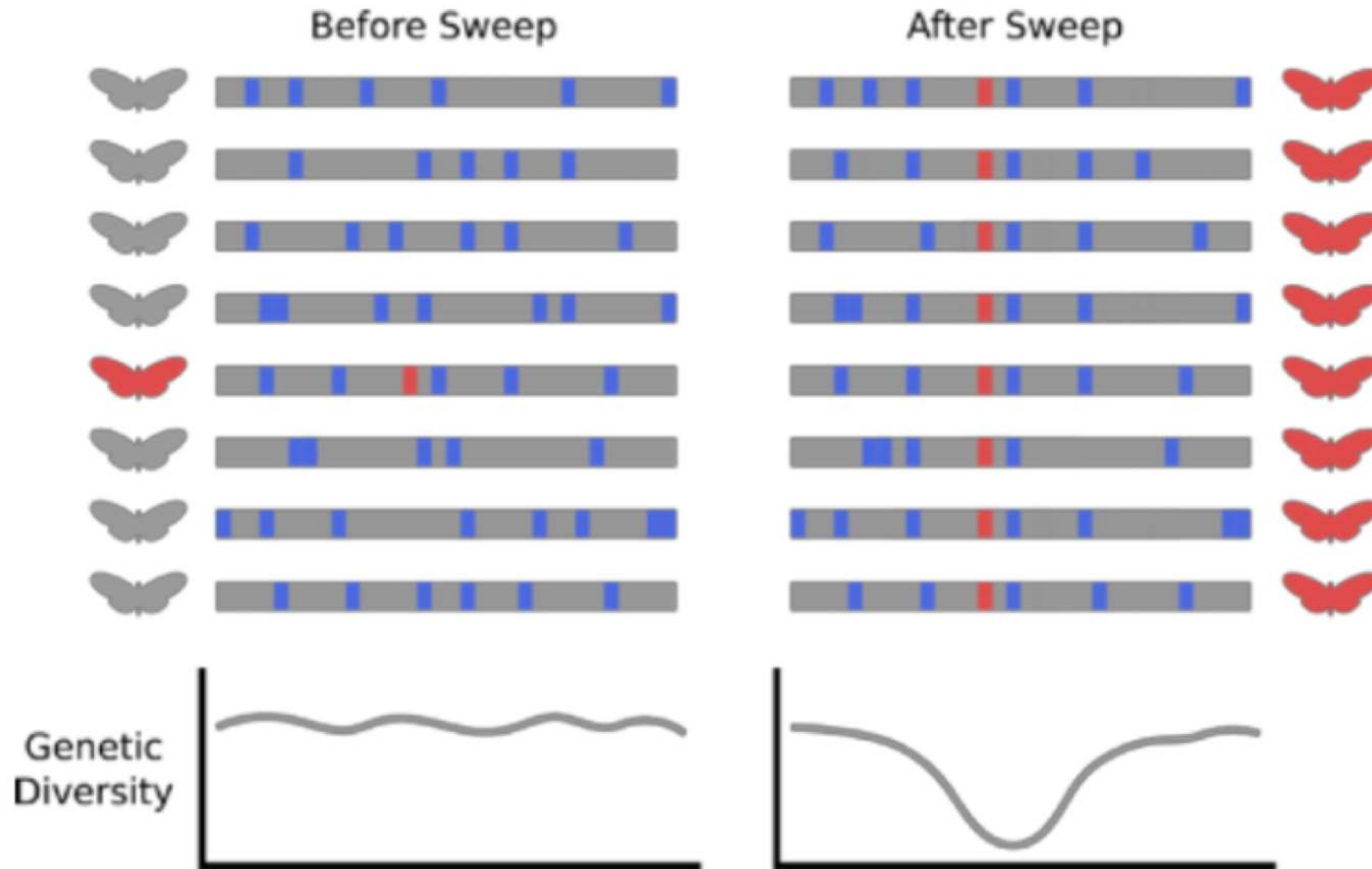
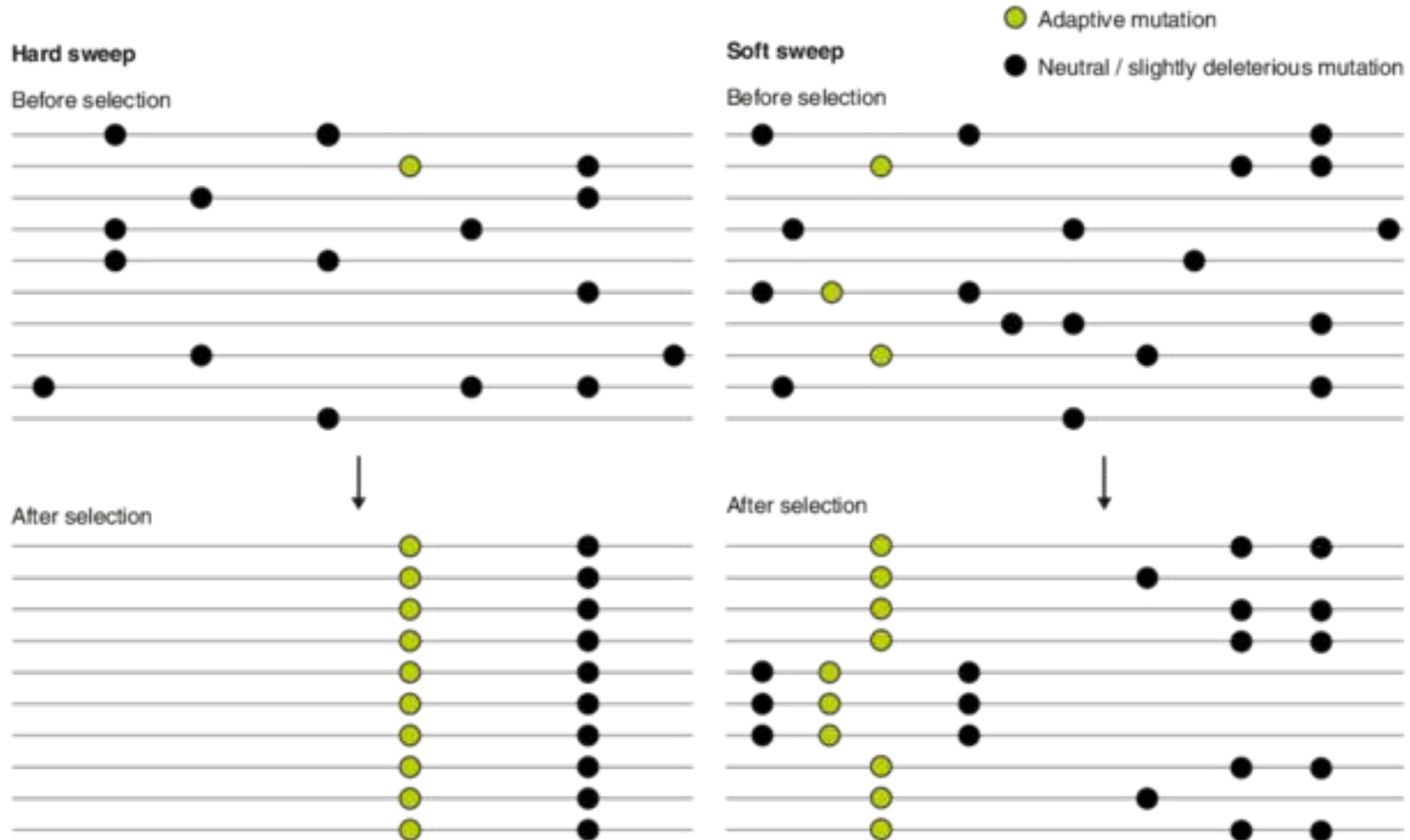


Figure by Chris Jiggins

Hard sweep and soft sweep

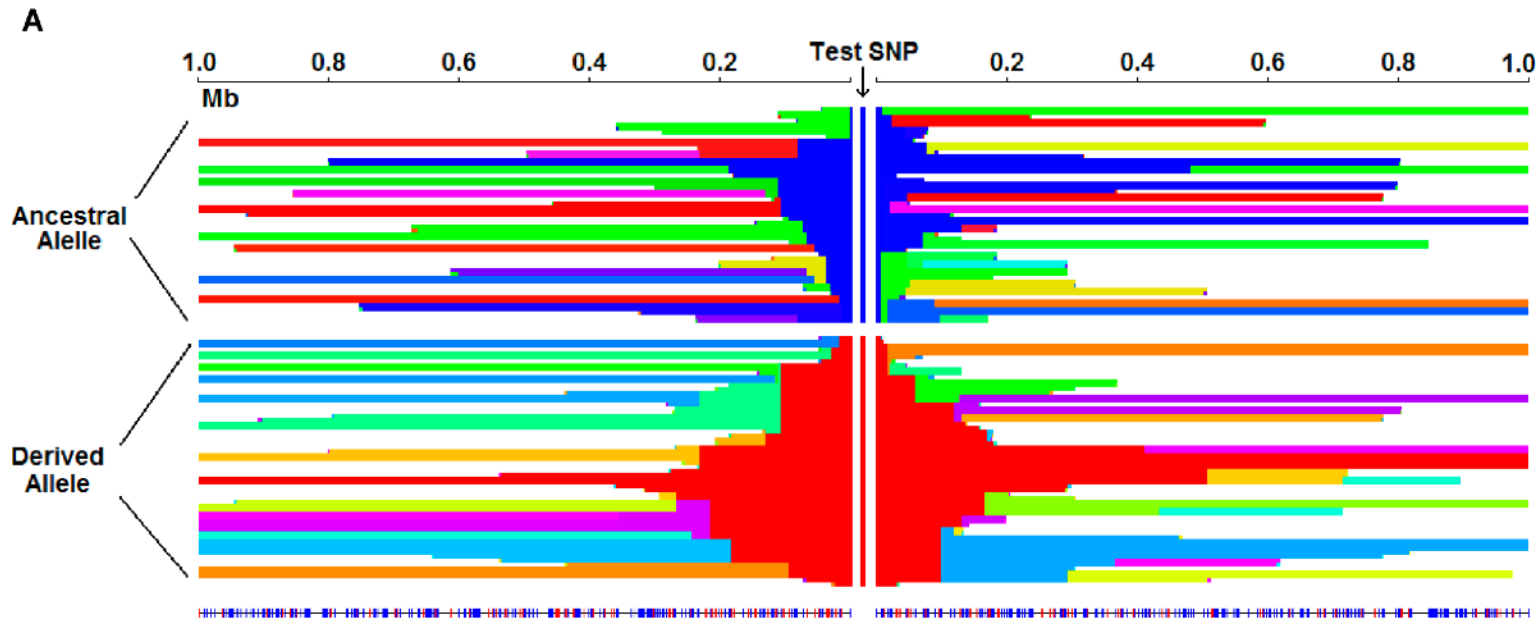
de novo mutations versus selection on standing variation



Statistics to find selective sweeps

- SFS-based methods, e.g. Fay & Wu's H or SweepFinder/SweeD
- Tajima's D: genomic intervals
- Extended haplotype statistics
 - measures the decay (of linkage around a specific site) due to both recombination and mutations
 - Key software: selscan, Rpackage rehh
 - Not affected by variation in recombination rate

Extended haplotype statistics: iHS and XP-EHH



iHS: within a population

iHS (integrated haplotype score) compares haplotype lengths **within a population**

-> an allele under selection will lead to increased haplotype length relative to other haplotypes in the same region

-> useful to detect **ongoing/incomplete sweeps**

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8765611/>

XP-EHH: between populations

XP-EHH (cross population extended haplotype homozygosity) compares haplotype lengths **between populations**

-> a population that had a sweep has increased haplotype lengths relative to the haplotypes in the other population in the same region

-> most powerful with **complete sweeps** restricted to one population

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research



Cite this article: Ravinet M, Elgvin TO, Trier C, Aliabadian M, Gavrillov A, Sætre G-P. 2018 Signatures of human-commensalism in the house sparrow genome. *Proc. R. Soc. B* **285**: 20181246.
<http://dx.doi.org/10.1098/rspb.2018.1246>

Signatures of human-commensalism in the house sparrow genome

Mark Ravinet¹, Tore Oldeide Elgvin^{1,2}, Cassandra Trier¹, Mansour Aliabadian³, Andrey Gavrillov⁴ and Glenn-Peter Sætre¹

¹Centre for Ecological and Evolutionary Synthesis, and ²Natural History Museum, University of Oslo, Oslo, Norway

³Ferdowsi University of Mashhad, Mashhad, Iran

⁴Institute of Zoology, Ministry of Education and Science of the Republic of Kazakhstan, Astana, Kazakhstan

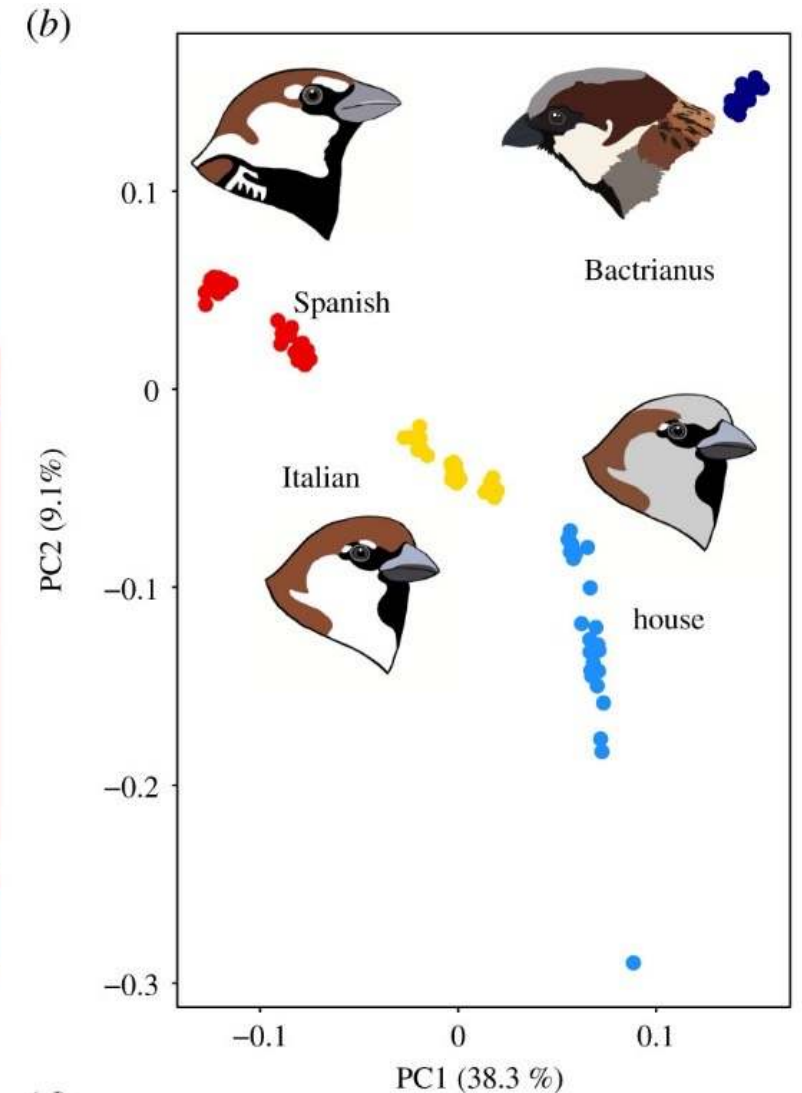
 MR, 0000-0002-2841-1798; G-PS, 0000-0002-6236-4905

House sparrows (*Passer domesticus*) are a hugely successful anthrodependent species; occurring on nearly every continent. Yet, despite their ubiquity and familiarity to humans, surprisingly little is known about their origins.

House sparrow subspecies



Mark Ravinet



The bactrianus and the European house sparrow split about 11 kya

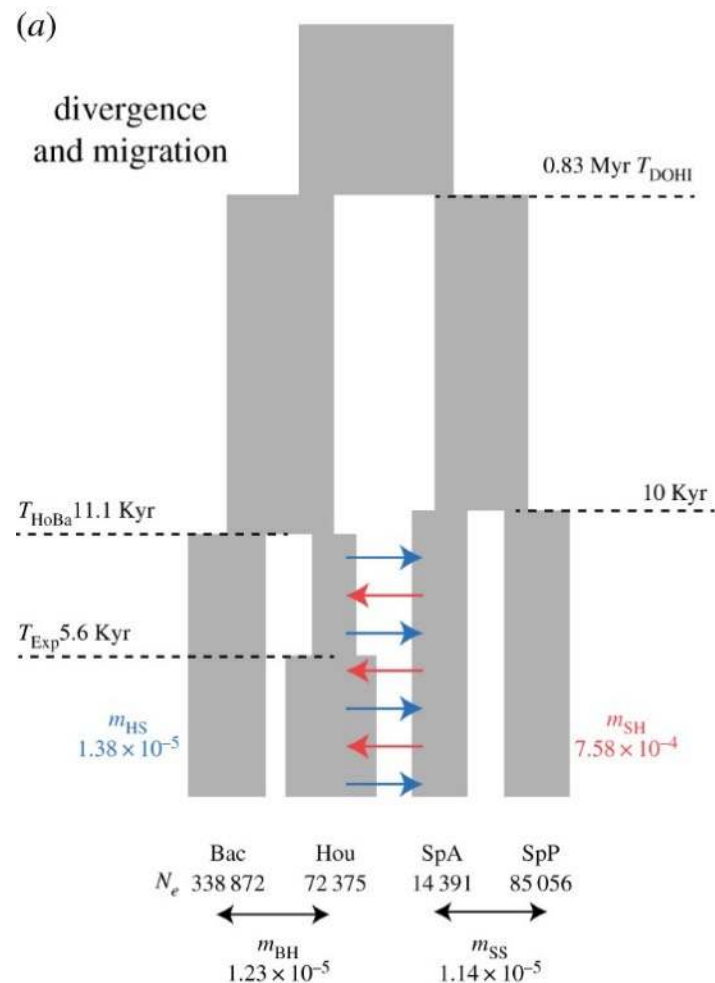


Photo by Askar Isabekov

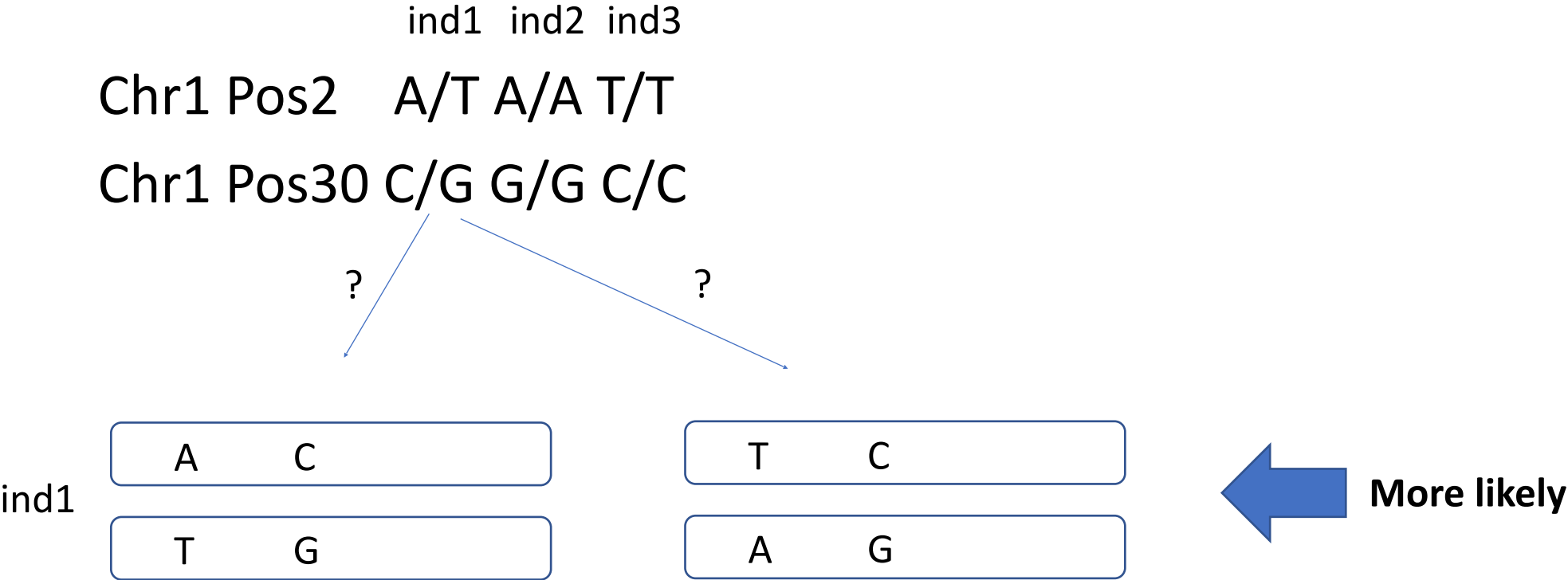


Photo by Askar Isabekov

Outline of this activity

- Phasing the dataset
- Genome scans for selective sweeps with extended haplotype statistics (iHS and XP-EHH)
- Identifying genes near the strongest sweep

Phasing the data

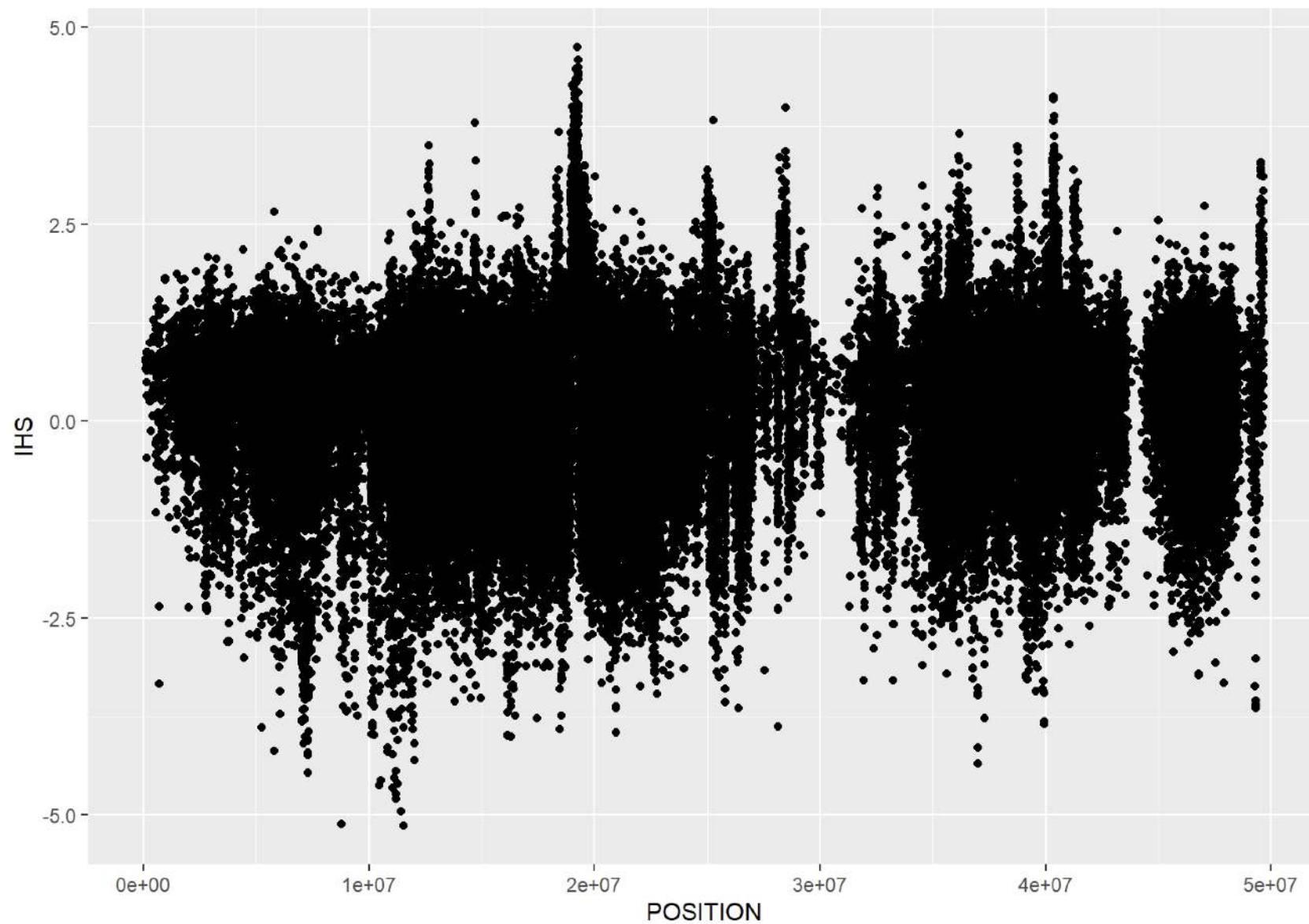


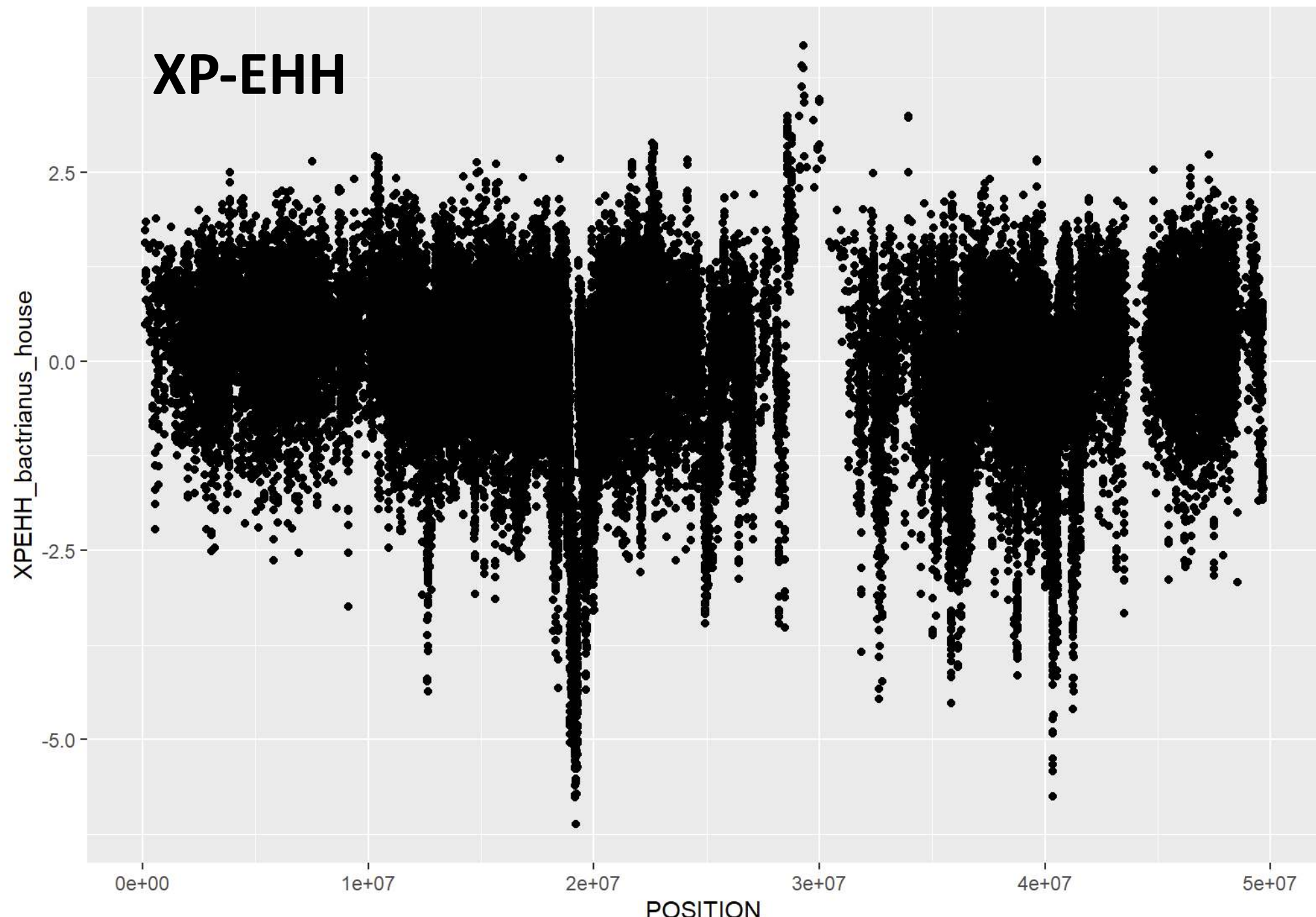
Additional information that can be used:

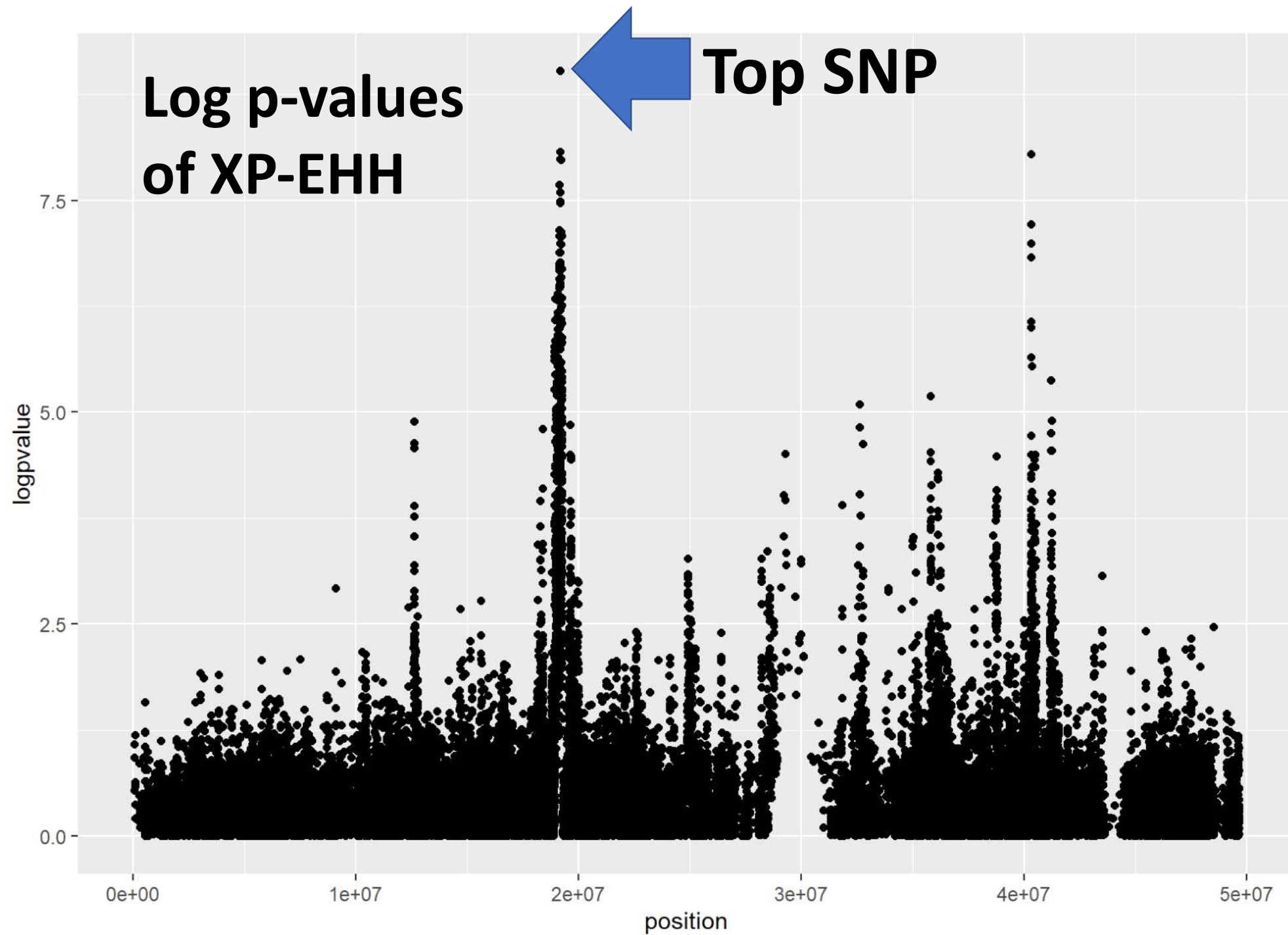
- Trios (rarely available)
- Reference panel (mostly only for humans)
- Recombination rates (e.g. In SHAPEIT or BEAGLE)
- reads spanning the sites (e.g. Whatsap and SHAPEIT4)

**Time to do the first exercise
and phase our data**

iHS of the house sparrow

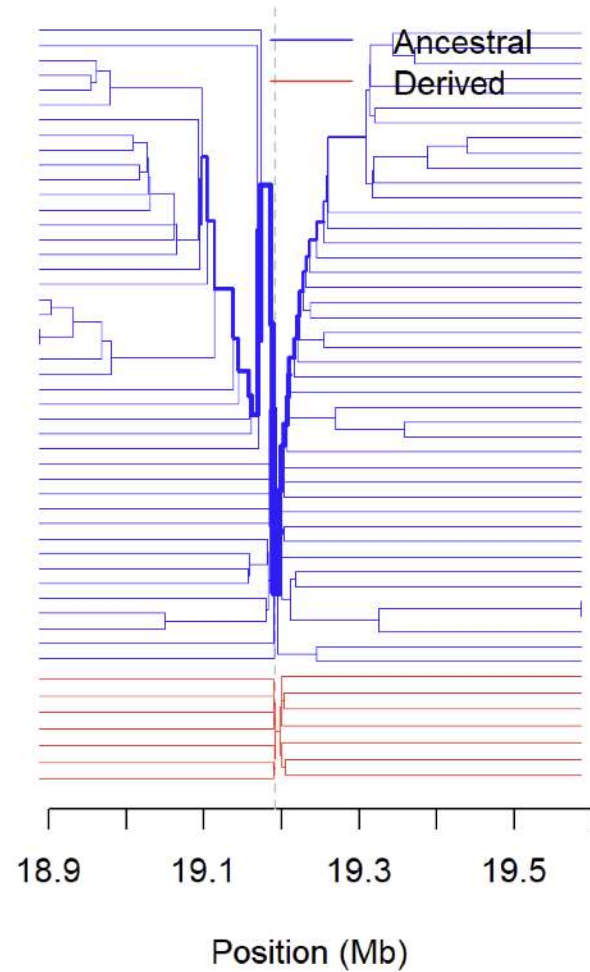






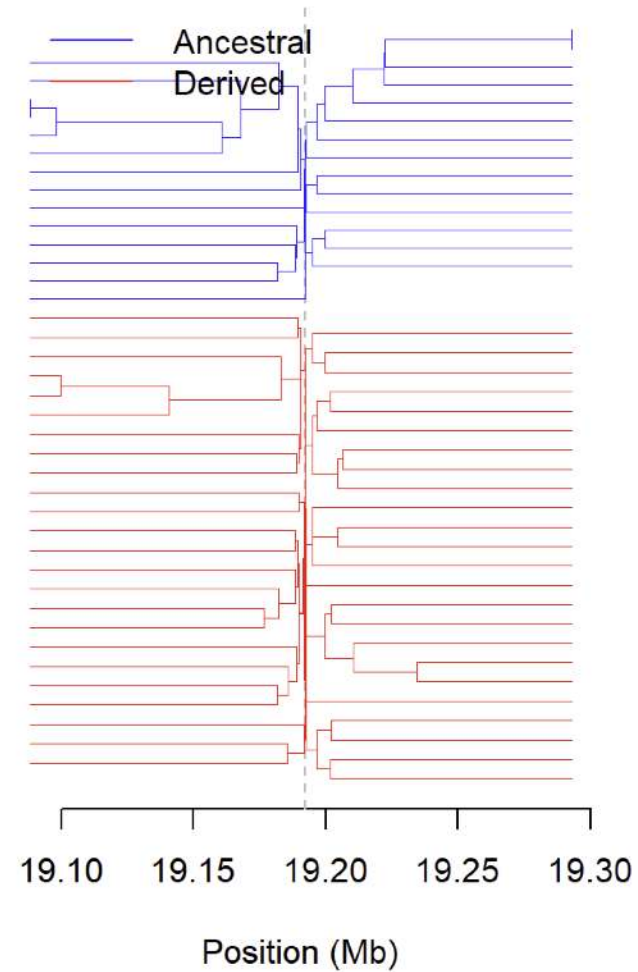
European House Sparrow

Haplotype furcations around '28429'



Bactrianus House Sparrow

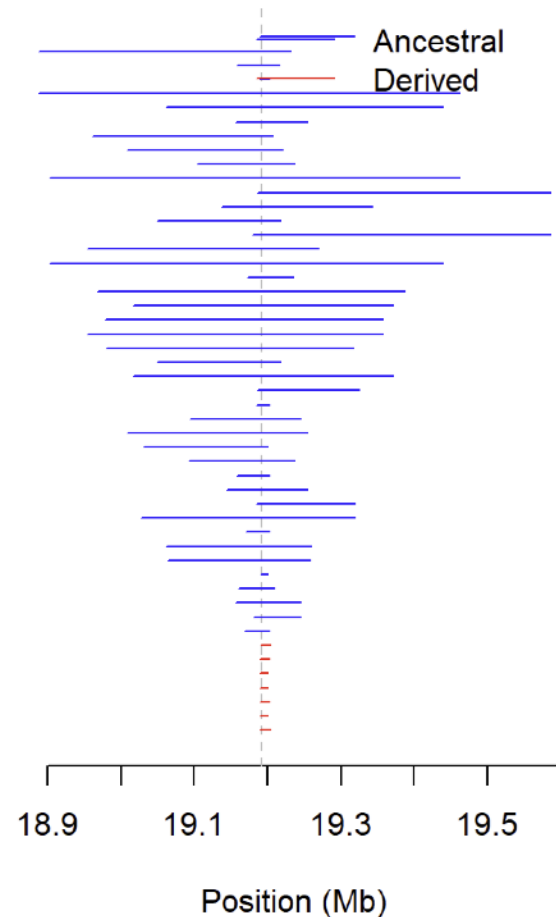
Haplotype furcations around '36191'



One of the alleles has very long haplotypes in the house sparrow, whereas in bactrianus the two alleles have similarly long haplotypes

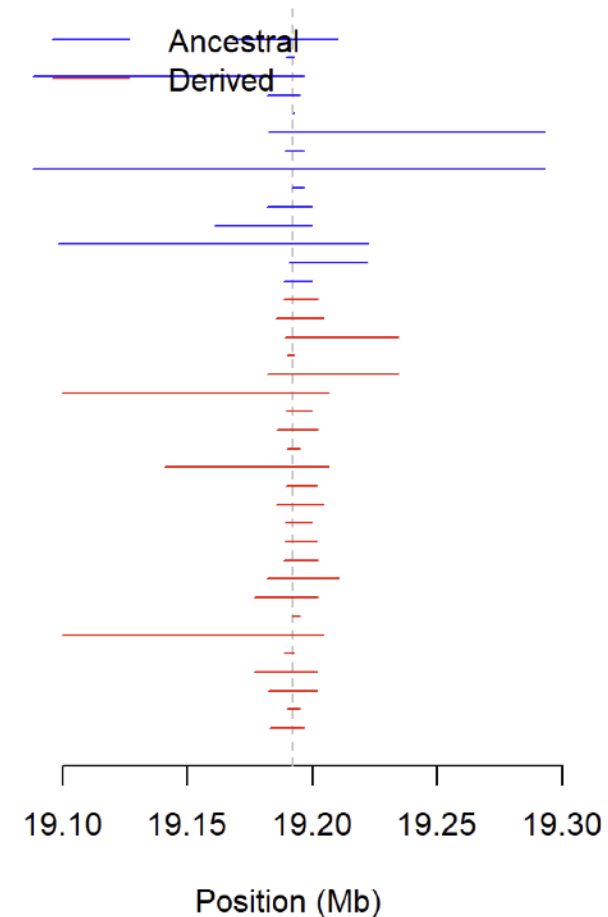
European House Sparrow

Haplotype length around '28429'

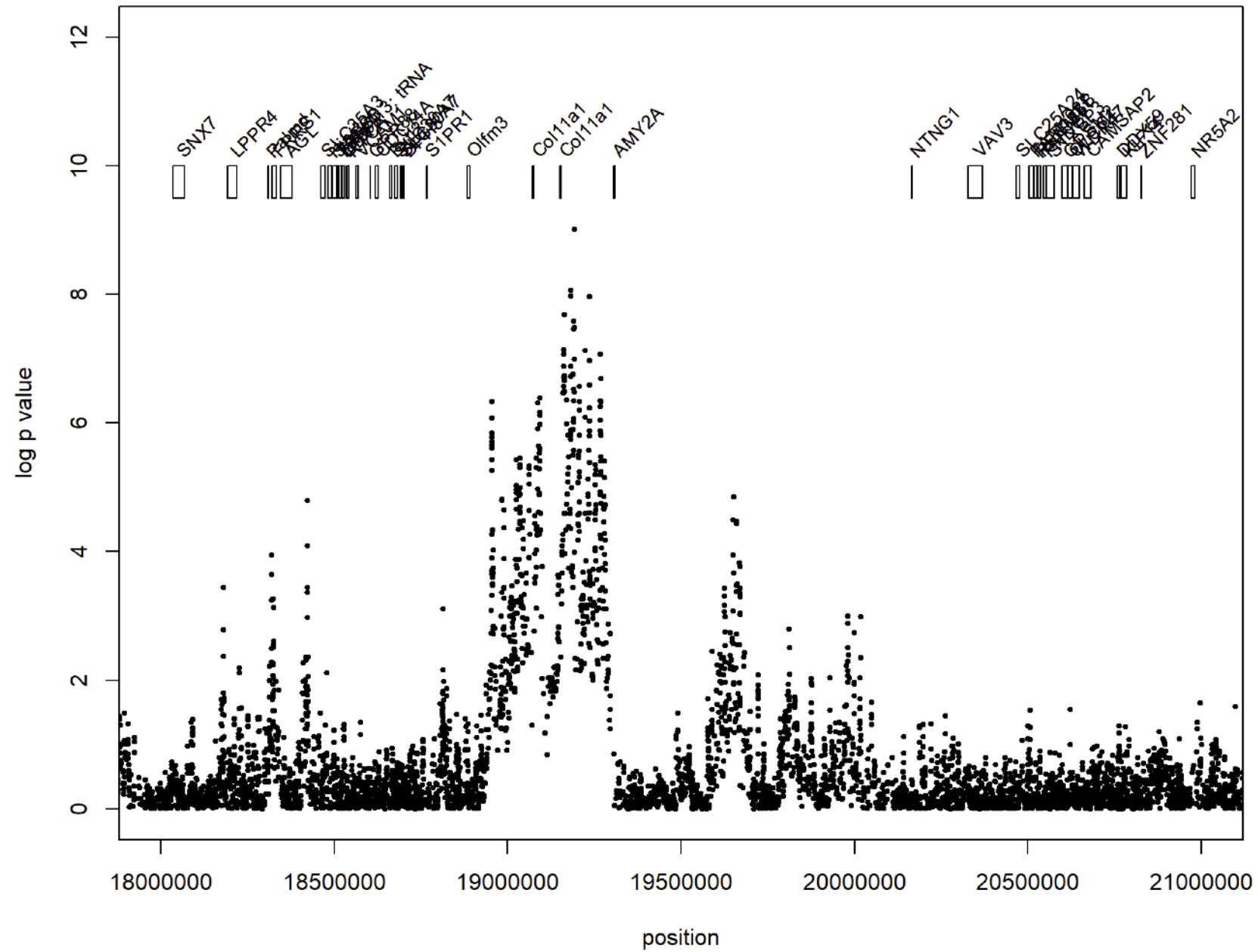


Bactrianus House Sparrow

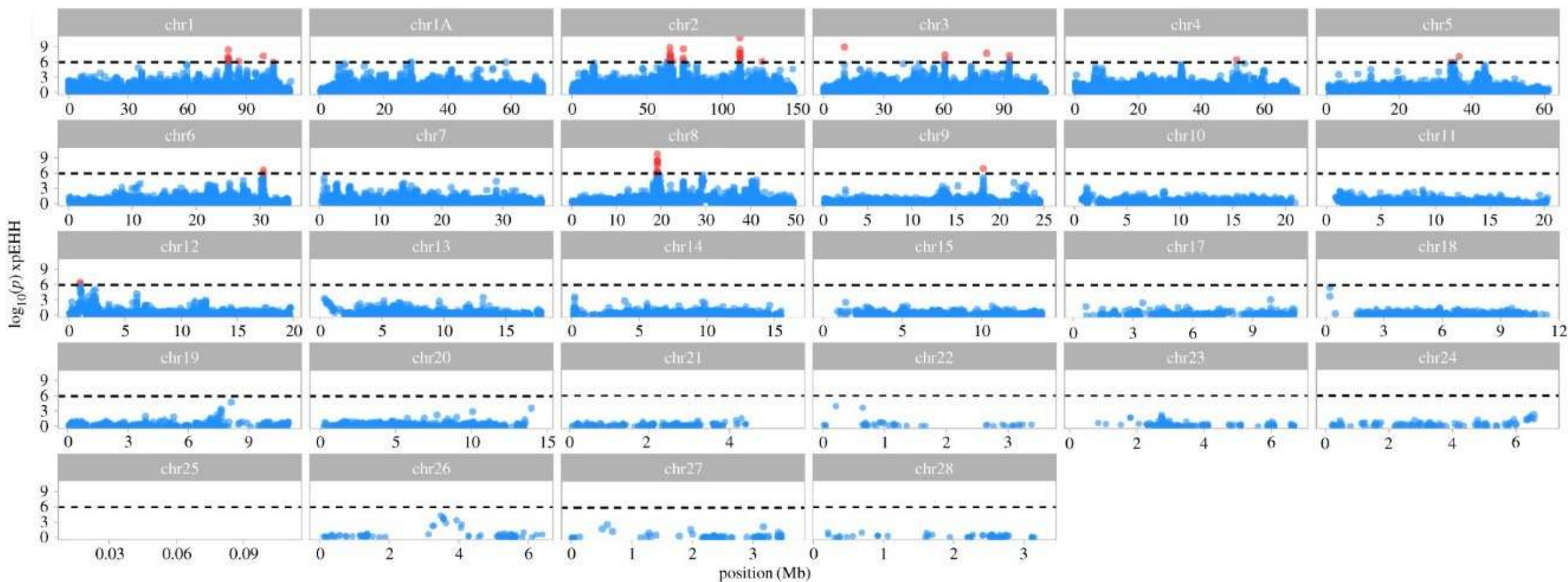
Haplotype length around '36191'



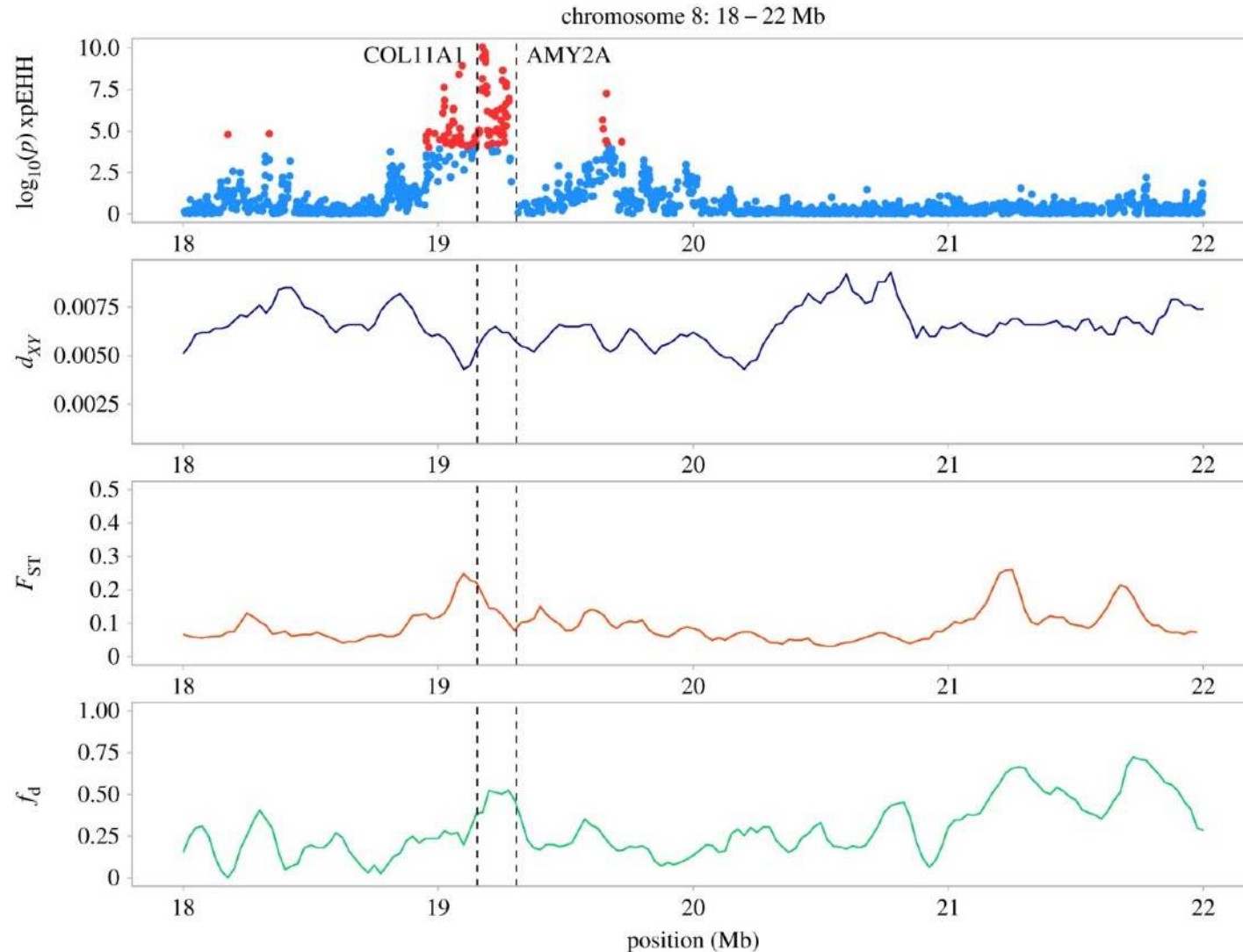
Now let's check if there are any genes nearby



**There are sweep signals on other chromosomes,
but the chr8 sweep is the strongest one**



**The region on chromosome 8 shows the strongest sweep signal
but the signal in d_{XY} and F_{ST} is not strong**



***COL11A*—collagen type XII alpha**

associated with craniofacial development

in humans, Col11a1 mutations are associated with the Marshall Syndrome



AMY2A—amylase alpha2

part of the amylase gene family associated with adaptation to a higher starch diet in both humans and dogs

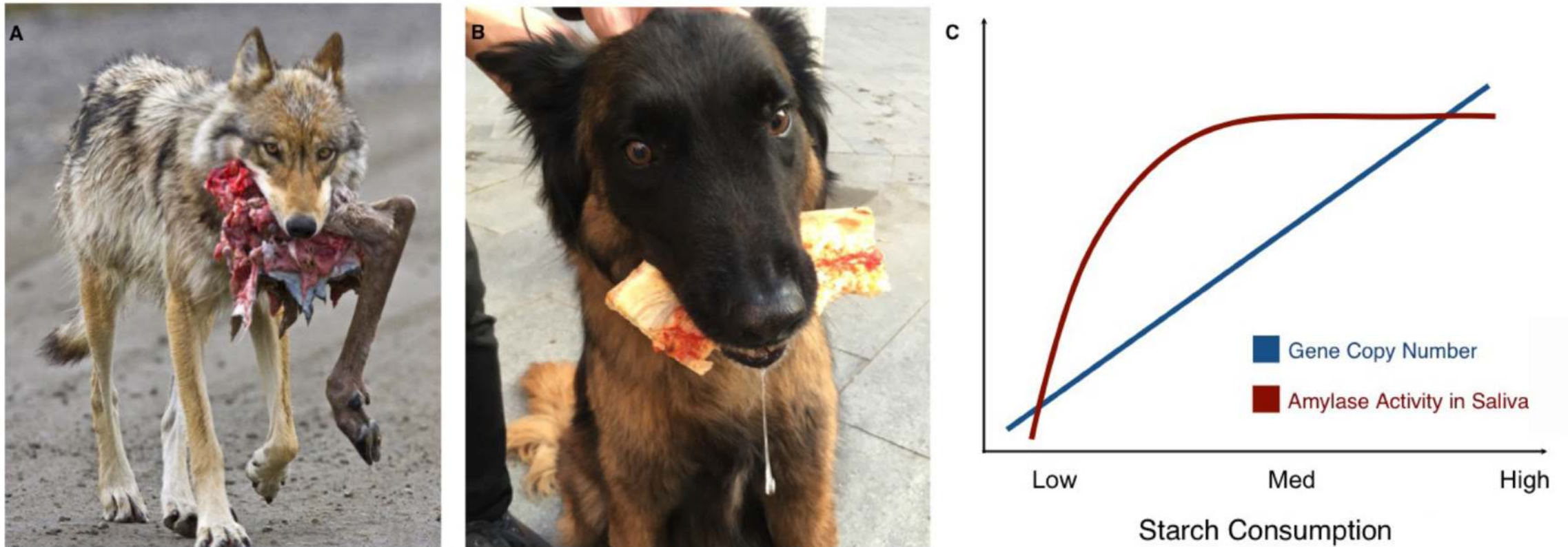


Figure by Mareike C Janiak



Information on the genome status at the Sanger

Information on genome status of species at the Tree of Life Programme, Wellcome Sanger Institute

<https://portal.tol.sanger.ac.uk>

Information on the data quality once data exists:

<https://tolqc.cog.sanger.ac.uk>