

# Extra exercises for most classes based on a dataset from Wildebeest (only if there is time)

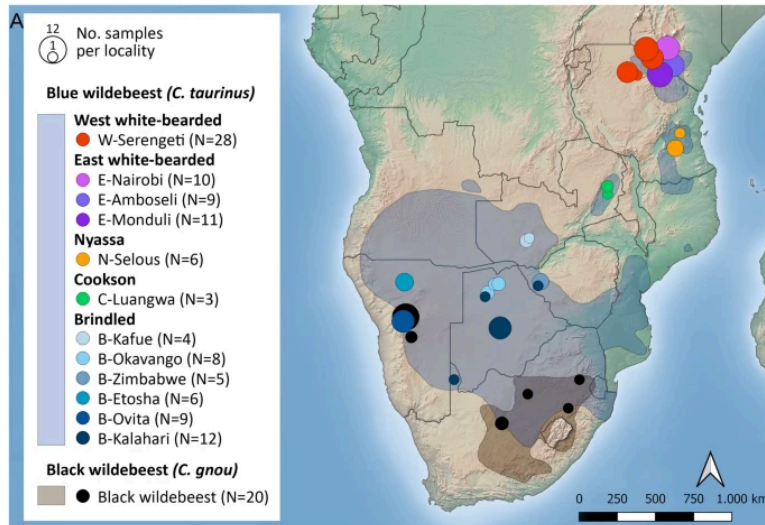
Below we introduce a dataset you can try to apply the tools/code from the class exercises if you have extra time.

## List of content

Description of the dataset	2
Exercise on diversity (Monday Feb 10)	3
Exercise on LD	3
Exercise on population structure and admixture	3
Exercise on FST	4
Exercise on ancient introgression	4
Exercise on selection	4
Exercise on relatedness/inbreeding	4

## Description of the dataset

The wildebeest is one of the most iconic antelopes in Africa and includes two distinct species. The blue wildebeest (*Connochaetes taurinus*), well known for its spectacular annual migration, is widely spread across Eastern and Southern Africa, and classically partitioned into five subspecies. Their sister species, black wildebeest (*C. gnou*) has a much more restricted natural distribution and underwent a severe bottleneck due to overhunting and persecution in the 1900s. The two species can interbreed and produce fertile hybrid offspring.



The dataset is a subset of the whole genome resequencing data from this paper which was published in 2024: <https://www.nature.com/articles/s41467-024-47015-y> (see figure from the paper above). It consists of approximately 4 million randomly thinned bi-allelic SNPs based on whole genome resequencing of 29 wildebeest individuals (see table below). These wildebeest samples were collected from five locations representing the black wildebeest and three subspecies of the blue wildebeest. As an outgroup, genotypes of two hartebeest individuals at the same SNP loci are also provided in a separate file. You will receive PLINK binary format genotypes and a list of detailed sampling localities for the samples. Sequencing reads of both wildebeest and hartebeest have been mapped to the reference genome of wildebeest.

Population	Subspecies	Species	n
Black	Black	black wildebeest	5
B-Etosha	Brindled	blue wildebeest	5
B-Kafue	Brindled	blue wildebeest	5
N-Selous	Nyassa	blue wildebeest	5
W-Serengeti	Western white-bearded	blue wildebeest	10

**You can find a set of plink files with the data in the directory here:**

/home/pls394/popgen\_2025/extra\_exercise/data/wildebeest.\*

## Exercise on diversity (Monday Feb 10)

*Objective:* Utilize PLINK and R to assess genetic diversity across different wildebeest populations.

*Questions/tasks:*

Q1. Try to estimate heterozygosity for each population and use this to compare heterozygosity across black wildebeest and various populations of blue wildebeest. Are there observable differences? If so, do these observations align with the historical background of the two wildebeest species? And do you get comparable estimates to those in the paper? (if not try to explain why)

Q2. Try to estimate local nucleotide diversity ( $\pi$ ) along the chromosome HiC\_scaffold\_1? Do you see any fluctuations? What evolutionary process can influence  $\pi$  locally and in what manner?

*Hints that may help:*

You may need to partition the PLINK file according to population groups (column “Population” in the table above).

## Exercise on LD

TBA

## Exercise on population structure and admixture

TBA

## Exercise on FST

TBA

## Exercise on ancient introgression

TBA

## Exercise on selection

TBA

## Exercise on relatedness/inbreeding

TBA