The Data

For this work, we utilized data collected from the Amboseli baboon population during the period of 2000-2013 1. This dataset includes microbiome profiles of the gut microbiome, based on 16S rRNA gene sequencing2. The profiles were constructed using fecal samples collected opportunistically. Each sample is accompanied by metadata detailing the environmental conditions at the time of defecation, as well as information about the baboon from which the sample was taken.

The dataset was divided into three parts: the training dataset, the test dataset, and the validation dataset. The train dataset contained 6,096 samples from 80 baboons, the test dataset contained XX samples from XX, and the validation dataset contained XX samples from XX.

**Metadata**

The metadata for each sample includes several elements. Temporal information covers the collection date, month, and hydrological year. At the individual level, the data specifies the baboon's unique identifier (*baboon\_id*), age at the time of collection, and sex. Social group data details the baboon’s social group affiliation during sampling, the group's size, and the diet composition, which is analyzed using 30-day sliding windows and Principal Component Analysis (PCA) of clr-transformed data, resulting in 13 principal components. Environmental context is also provided, including the season (*dry* or *wet*) and the amount of rainfall during the collection month.

Bibliography

1. Björk, J. R. *et al.* Synchrony and idiosyncrasy in the gut microbiome of wild baboons. *Nat. Ecol. Evol.* **6**, 955–964 (2022).

2. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).