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.

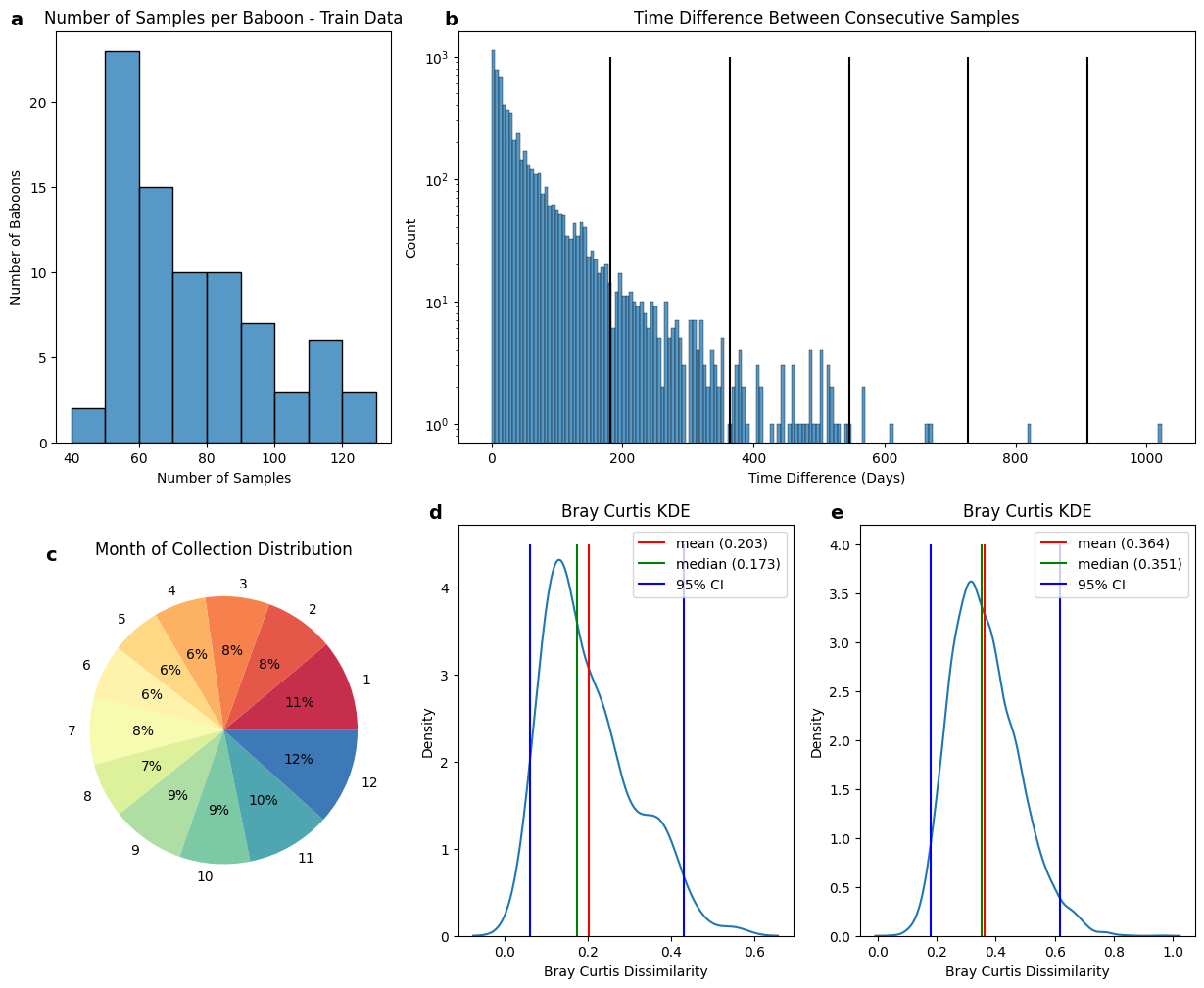
**Metadata**

The metadata for each sample encompasses several essential elements. Temporal information includes 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 provides details on the baboon’s social group affiliation during sampling, the group's size, and diet composition, which is assessed using 30-day sliding windows and Principal Component Analysis (PCA)3 of center log ratio (clr)-transformed data, resulting in 13 principal components. Environmental context is also included, covering the season (dry or wet) and the amount of rainfall during the collection month.

Analysis of the training data reveals that each baboon has an average of 76.2 samples **(Fig 1a)**. The median time difference between consecutive samples from the same baboon is 22 days, with 282 samples having a preceding sample collected more than 180 days earlier **(Fig 1b).** Although the wet season spans 42% of the year (November to May), 61% of the samples were collected during this period **(Fig 1c)**.

**Sample data**

For each sample, there is a measurement of the abundance rates of the 61 most prevalent genera across all microbiome samples. A calculation of the Bray-Curtis dissimilarity score over 102 samples taken on the same day from the same baboon presented a right-tailed bell curve distribution with a mean score of 0.203 and a median of 0.173 **(fig 1d)**. Since both samples share the same metadata, this distribution can be considered the expected limit for the accuracy of any prediction using our data. An analysis of the Bray-Curtis dissimilarity score between sample and the mean of the predeceasing samples presented a right-tailed bell curve distribution with a mean score of 0.364 and a median of 0.351 **(fig 1e)**, suggesting the potential for predictions based on previous samples.



**Figure 1. Analasys results of the train data**

**(a)**Histogram of number of samples collected per baboon in the train data. **(b)**Histogram of the time differences between the collection date of every 2 subsequent samples of the same baboon. **(c)**Pie chart of the distribution of samples' collection month. **(d)**Kernel Density Estimation (KDE) smoothed histogram of the Bray- Curtis dissimilarity score between every two samples collected on the same date from the same baboon. **(e)**KDE smoothed histogram of the Bray- Curtis dissimilarity score between sample and the mean of previous samples of the same baboon.

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).

3. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441 (1933).