Supplementary Material

# Supplementary Tables

Supplementary Table 1. Anatomical representation data from several experimental and archaeological rabbit assemblages expressed in %RA (relative abundance) used in the PCA and ML analyses.

A screenshot of a computer

Description automatically generated

Supplementary Table 2. Taphonomic and age of death data of the leporid assemblages used in the HCPC and K-means clustering analysis. The values represent percentages per NISP. The used variables are % cut marks (CM); % burning (BU); % digested (DI); % tooth/beak marks (TB); % complete long limbs (CLL); % tubes (T); % of adult leporids (AD).

A table of numbers and letters

Description automatically generated

Supplementary Table 3. Correlation values between the features (anatomical portions) used in the ML analysis. High correlation (>0.7) is observed between MC (metacarpal) and VB (vertebrae), as well as between metacarpals (MC) and metatarsals (MT). RI: ribs, SC: scapulae, INM: innominates, RA: radii, CA: carpals/tarsals, FM: femora, TA: tibiae.

A white sheet with black text

Description automatically generated

Supplementary Table 4. Size and average silhouette width of the clusters estimated using a) HCPC and b) K-means clustering.

1. A table with numbers and letters

   Description automatically generated b) A table with numbers and text

   Description automatically generated

# Supplementary Figures

A diagram of a group of individuals

Description automatically generated

**Supplementary Figure 1.** PCA biplot showing the distribution of samples across the two principal components, with labeled samples colored according to their respective groups. Confidence ellipses illustrate the variability within groups.



**Supplementary Figure 2.** Correlation matrix of relative abundance variables used to determine the highly correlated variables (>0.7) and eliminate them from the ML analysis.

A) B)



**Supplementary Figure 3.** A) 95% confidence intervals of accuracy and kappa scores resulting from the resamples of the different ML models used in the analysis. B) Box-and-whisker plot showing the distribution, spread and outliers of the performance metrics for each model. The smallest ranges, which reflect higher consistency in performance, are obtained for the GBM (Gradient Boosting Model). SVM: Support Vector Machine, RF: Random Forest.

A) B)



**Supplementary Figure 4.** A) Principal Component Analysis of variables. The first two components explain 45.50% and 19.16% of the variance, respectively. Along the first dimension (Dim 1), which explains the largest variance, there is a clear separation of variables that are typically higher in human-influenced assemblages to the right. B) Hierarchical clustering on the factor map showing the clusters and structure of the data.

A) B)



**Supplementary Figure 5. A)** Silhouette plot of the cluster analysis performed using HCPC showing the quality of clustering. Negative values indicate the point might have been assigned to the wrong cluster. Values close to 0 indicate that the point is on or very close to the decision boundary between two neighboring clusters. Values close to 1 indicate the point is well-clustered. Bars are colored by cluster. B) Silhouette plot of the cluster analysis performed using K-means clustering showing the slightly higher quality of clustering compared to HCPC.