## Statistical analysis

Alpha diversity analysis was performed using the “Shannon-Wiener” and “Simpson” diversity indexes We considered three taxonomical levels, namely family, genus and species. Each sample was considered as an isolated ecosystem. Index distributions are reported in Figure XXX\_alpha\_distr and averaged in Table XXX\_alpha.

Samples can also be grouped by province of origin, thus allowing investigate differences between groups. Distributions of diversity indexes are reported in Figure XXX\_alpha\_prov. Moreover, we performed a standard ANOVA to formally ascertain the presence of statistically significant differences between provinces. Resulting p-values are also reported in Table XXX\_alpha.

**Table XXX\_alpha. Average values for Shannon and Simpson richness indexes for all samples, computed at three different taxa levels. Significance test p-values computed via ANOVA grouping the samples by province.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Taxa level** | | |
| **Richness index** | | **Family** | **Genera** | **Species** |
| Shannon | Average | 0.684 | 1.038 | 0.46 |
| p-value | 0.108 | 0.128 | 0.322 |
| Simpson | Average | 0.350 | 0.557 | 0.245 |
| p-value | 0.083 | 0.267 | 0.285 |



Results revealed no statistically significant differences for both indexes at three taxonomic levels (family, genus and species), with all p-values ≥ 0.05. The Shannon index shows a clearly non-normal distribution (Fig. 4). In fact, it appears to be bimodal, *i.e.* as if it was the union of two distributions, one with lower and one with higher average diversity.

**Fig XXX\_alpha\_prov. Distribution of alpha diversity computed using Shannon’s and Simpson’s indexes. Colours indicate different provinces of origin. The three boxes report indexes computed at different taxonomic levels.**

**Table 4. p-values from ANOVA.**

|  |  |  |
| --- | --- | --- |
| **Level** | **Shannon\_pvalue** | **Simpson\_pvalue** |
| Family | 0.219 | 0.205 |
| Genus | 0.924 | 0.955 |
| Species | 0.857 | 0.792 |

Beta diversity also considers information about biological richness, but measures the variation between pairs of different communities using the Bray-Curtis index. Once the (square, symmetrical) diversity matrix is obtained we applied Multi Dimensional Scaling (MDS) to allow data exploration. Results are reported in Figure XXX\_beta.

**Fig 6. Diversity analysis - beta diversity.** Beta diversity computed using Bray–Curtis index and then transformed via Multi-Dimensional Scaling for milk samples by province.