

Modelling how the diet drives the gut microbiota and how we can make prediction on the gut microbiota changing the diet.

We explored two ways.

1. The first one is the development of a model in a simplified system (i.e. *C. elegans*) which microbiota can be easily monitored. The idea was to study the species that are found in *C. elegans* and the metabolism of those species to parametrize for each species the growth rate on the diet. We discussed the possibility of build up a new experiment or to study already published data. This methodology can be good to understand really which exact species is driven by which diet, but we criticized the difficulty to scale up to a different system (i.e. human) where there are bigger communities.
2. The second one is to explore the human gut microbiota communities in terms of species abundance distribution and frequency abundance distribution. The idea is to review already published data about how gut species abundances are distributed in different diets and then compare it with the species abundance distributions found in the model simulations. With this methodology the aim is to predict how the diet drives the microbiota abundance distribution (if it does!), but not taxonomically naming the species.

For the moment the first one was discarded

Regarding the second activity my timeline was:

1. Starting evaluating AFD and SAD for the data of the gut microbiota that I already got.
Deadline 15/12/24
2. Review experimental work that contains gut microbiota data driven by diet (it is not necessary that are in time but could be nice...) Deadline 12/01/24

Does the diet variate the gut microbiota?

Hypothesis:

1. The diet variates the abundance distribution of the gut microbiota

Questions:

- More polysaccharides mean more dilution so less diversity?
- An animal-based diet means more specialized microbiota so less diversity?
- Different diets correspond to different shapes in gut microbiota abundance distributions?
- Different diets correspond to the same shape in gut microbiota abundance distribution but to the different slope?

The first data I used belong to the paper: "Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake." Meslier V, et al. Gut 2020

<doi:10.1136/gutjnl-2019-320438>

The experiment consisted of 82 subjects with the obese condition. It was conducted for 8 weeks, there is the timeline from 0, 4, 8 weeks. In the initial time they were all the same. They started from a diet that was Western diet and then the half of them were under Mediterranean diet.

To the best of my understanding, they used both metagenomics (total genomics amplification) and 16 S (amplification of just some regions of the genome). By the way, the data that I have (Edoardo Pasolli gave me) are Metagenomic data

What statistical analysis they made?

Variations in dietary and clinical variables at specific time points compared with baseline values between the MedD and ConD groups were evaluated by two-way analysis of variance with repeated measures and Tukey's post hoc test. Non-parametric Wilcoxon signed-rank test (testRelations function of momr R package) was performed to compare means between ConD and MedD subjects at each time point, while the post hoc Nemenyi test for multiple comparisons following the Friedman test was used within each group. Pairwise Spearman's rank correlations were used to estimate the overall similarity of the microbiome and metabolome within the MedD and ConD groups and between time points (baseline vs 4 weeks and 4 weeks vs 8 weeks).

I plotted the Species Abundances distribution SADs dividing by the diets (Fig. 1). We observe on the x-axis, the log of abundances, so on the left the less abundant species, and on the right the more abundant species. On the y-axis, we have the prob. density, so for the low abundant species, it is high (10^4) while for the high abundant species it is low (10^{-2}).

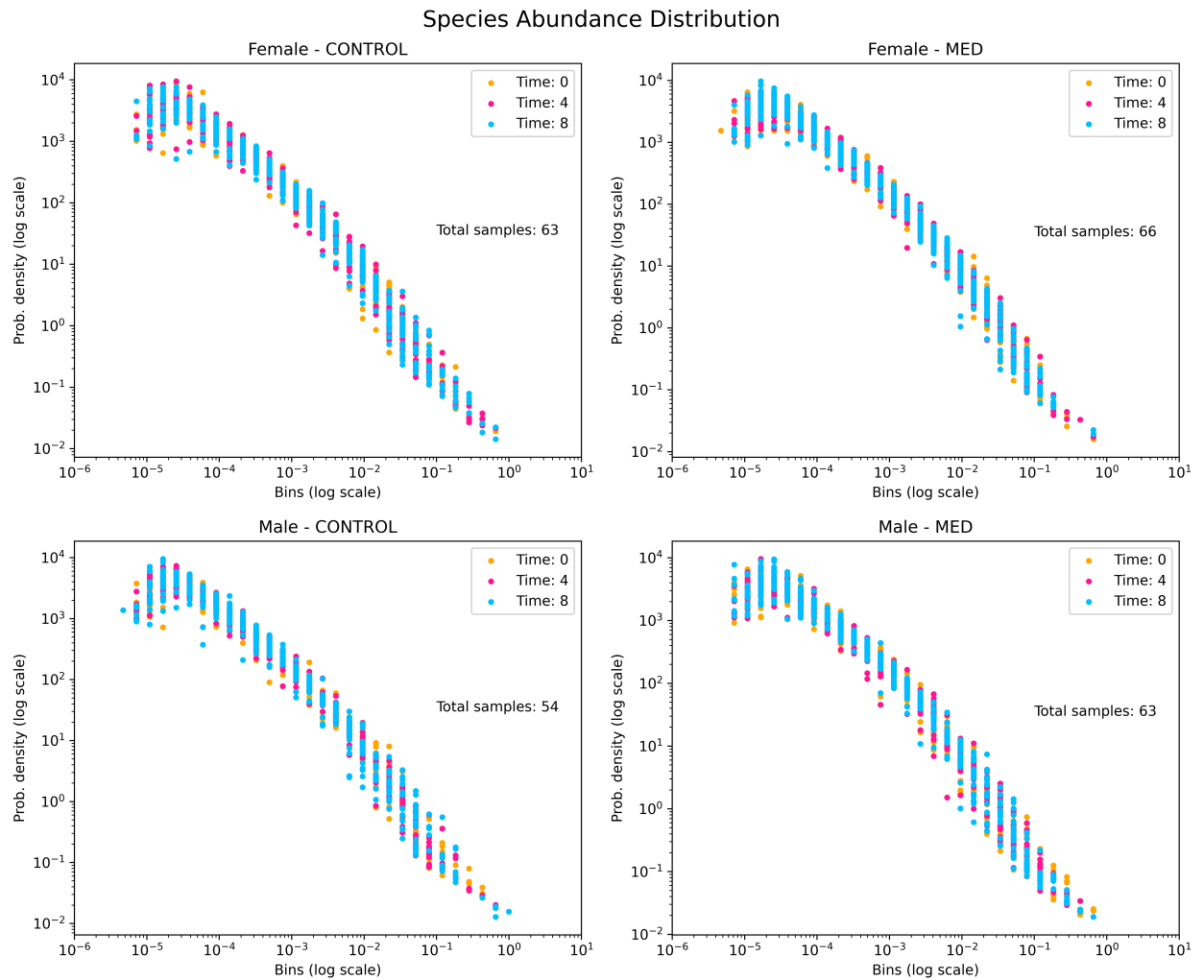


Fig. 1: Species Abundance distribution of all the species across samples divided in male, female, control and mediterranean diet.

Then, I found curated Metagenomic Data, an R package that enable to easily download metagenomic data <DOI:10.18129/B9.bioc.curatedMetagenomicData>. The curatedMetagenomicData package provides standardized, curated human microbiome data for novel analyses.

From this dataset I checked from data driven by diet and I found the data from De Filippis et al., 2019 "Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets" <DOI:10.1016/j.chom.2019.01.004>. This experiment includes 90 subject, healthy, that went through three different diets: vegan, vegetarian and omnivore.

Again, I plotted the abundance distributions (Fig. 2). We observe on the x-axis, the log of abundances, so on the left the less abundant species, and on the right the more abundant species. On the y-axis, we have the prob. density, so for the low abundant species, it is high (10^2) while for the high abundant species it is low (10^{-5}).

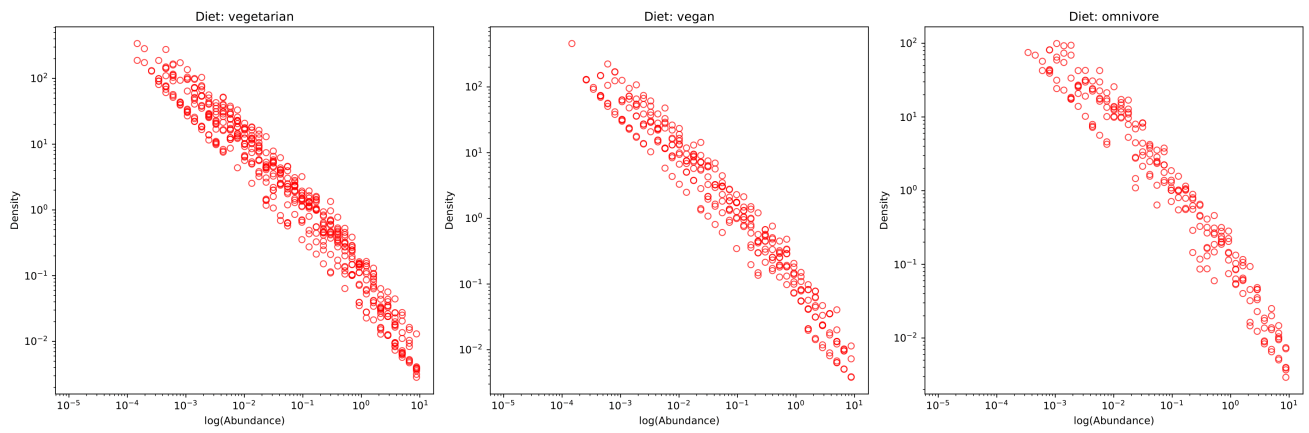


Fig. 2: The Species Abundance Distribution of all the species across different samples, divided in vegetarian, vegan and omnivore.

Then, I evaluated the abundances by ordering the species from the most abundant to the less abundant with a Rank Abundance Plot (Fig. 3)

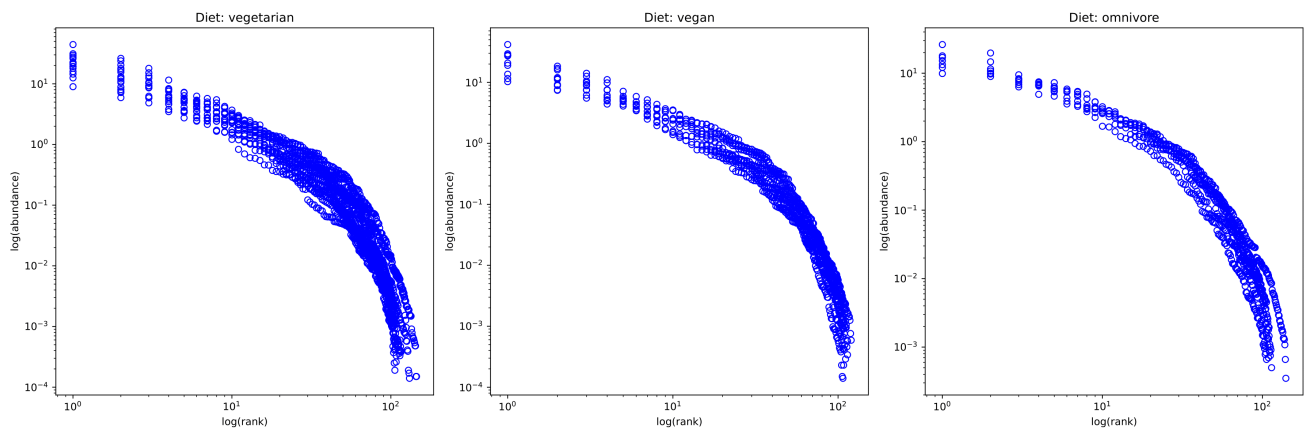


Fig. 3: The rank abundance plot of all the species across different samples, divided in vegetarian, vegan and omnivore.

From those two datasets I didn't see visually a difference in the shape of the abundance distribution. The conclusion at this point is that the shape is always the same regardless of the diet. Looking at different shapes of SADs in different individuals, is it enough to say that there is no difference?

Then, I tried to see if there is a difference in the slope of the curves regarding the diet. I tried to use the code from Ser-Giacomi et al., 2018 "Ubiquitous abundance distribution of non-dominant plankton across the global ocean" but I was not able to fit my data.

The main differences between mine and his data is that:

1. mine are in percentage, his are in absolute abundances
2. each individual in my data has less than 200 species, while in his there are thousands
3. I selected for individuals that have at least 100 species, from 90 individuals the dataset was reduced to 30

Creating a code in python using the idea of Ser-Giacomi to fit my data on a power law distribution and see if the slope changes between diets.

2. The diet variates the taxonomy of the gut microbiota and not the abundances

- Are there some microbial groups that segregate in different diets?

I tried to explore if the species classifies depending on the diet type and if they order in the bidimensional space. Clustering analysis and different techniques of ordination were conducted but not all of them enabled classification and ordination

For example, the CAP, Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology, enables us to see segregation. CAP Canonical analysis of principal coordinates based on discriminant analysis: this method uses the diet as a constraint, and we can see that some species segregates into different diets (Fig. 4).

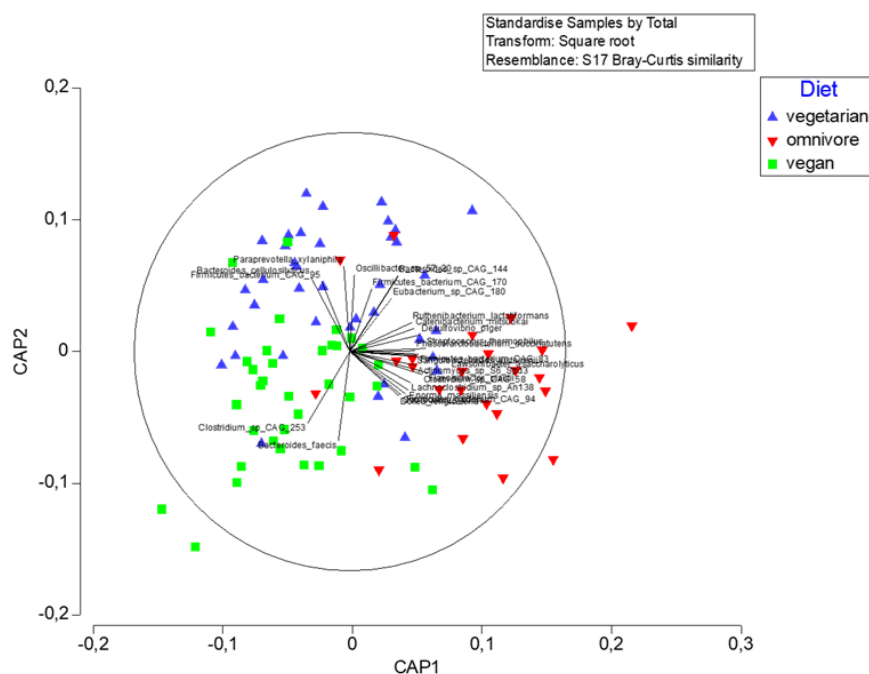


Fig. 4: CAP analysis based on the species and on the individuals with the constraint of the diet. The blue, red and green triangles represent the individuals that were under a vegetarian, omnivore and vegan diet, respectively.

We then tried to plot the species abundances in a three phase diagram (Fig. 5). We mediated all the individuals for the three type of diet and then we plotted the abundance of all the species across the three type of diet. In the middle there are the species that occur in the three diets, while in the angles the one that occur more in one of the three diets. The vegan and the omnivore share less species than the omnivore and vegetarian and the most of the species are shared between vegetarian and vegan.

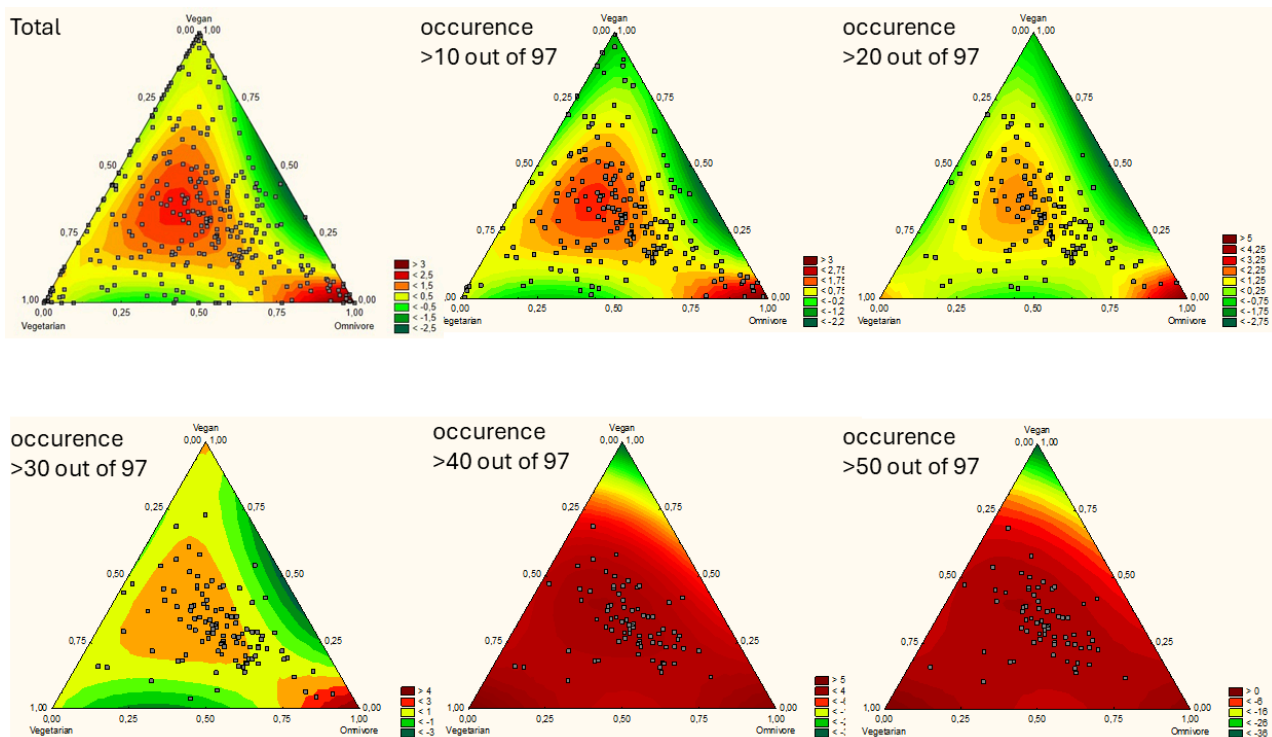


Fig. 5: Three phases diagram based on the abundance of the species inside the three group. From the first to the sixth triangle, from left up to right down, the occurrence of the species in the number of individuals increases. The little squares represent the species.

I had a look to other papers about the topic, and they use usually as ordination not just PCA but nMDS or PCA with other methods (coinertia analysis) and they obtain segregation. For example, in De Filippis et al., 2019 "High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome", they calculated the Pairwise Spearman correlations were calculated between microbial genera, dietary and metabolome datasets and then the correlation plots were visualized and clustered in R. And moreover, for the ordination: In order to explore the relationship between the genus-level microbiota and the dietary datasets, we carried out principal components analysis (PCA) on the individual datasets and then integrated them using coinertia analysis (CIA),³⁹ which allows identification of the shared biological trends within the two datasets. Again, Rampelli et al 2015, "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota" did the hierarchical clustering based on the spearman correlation coefficients of the KEGG Orthology (KO) gene count, filtered for subject presence ≥ 100 in at least 20% subjects. and Hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the KO gene counts, filtered for subject presence in at least 50% of subjects.

3. The health condition variates the abundance distribution of the gut microbiota

No progress

More: Lorenzo (colleague of Onofrio) had some suggestions because he also works on gut microbiota:

- oscillation of the distributions (?)

- studying the variability of each species across the samples (?)