Community Ecology: Final Project

Spring 2014

Yoav Ram, 066245259

Table of Contents

Part 1: Community data analysis 2
Species richness 3
Species diversity 4
Beta diversity 6
Ordination 8
Sites ordination 8
Species ordination 8
Assembly rules 11
Part 2: Open questions 13
Question 4: What limits the number of mammal species in Israel? 13
Question 1: Deterministic or stochastic community structure? 14
References 15

Part 1: Community data analysis

I analyzed data from the Human Microbiome Project (http://www.hmpdacc.org/). I downloaded the Mother file of the Phase I (May 1) 16S variable regions 1-3 OTU (operational taxonomic unit) counts (http://downloads.hmpdacc.org/data/HMMCP/finalData/hmp1.v13.hq.otu.counts.bz2). This dataset contains 2,799 samples and 27,655 OTUs. The OTUs are described in a corresponding lookup file (http://downloads.hmpdacc.org/data/HMMCP/finalData/hmp1.v13.hq.otu.lookup.bz2). The samples metadata (including body site, sex, etc. for each sample) was retrieved (http://www.hmpdacc.org/doc/ppAll_V13_map.txt) and merged with the occurrences data. After filtering out unclassified OTUs, samples with zero counts, and OTUs with zero counts, 2,774 samples and 8,289 OTUs remained in the dataset.

The samples represent 173 individuals (85 males and 88 females), who were sampled between 1 and 3 times in five major body sites (see Fig. 1), divided to up to 18 sub-sites.

I analyzed the data using *R* with the *vegan* package. The code (analysis.R), this document (final_project.docx), and an *IPython notebook* that was used to clean and format the dataset (munge.ipynb) are available online as a *git* repository at https://github.com/yoavram/CommunityEcologyProject.

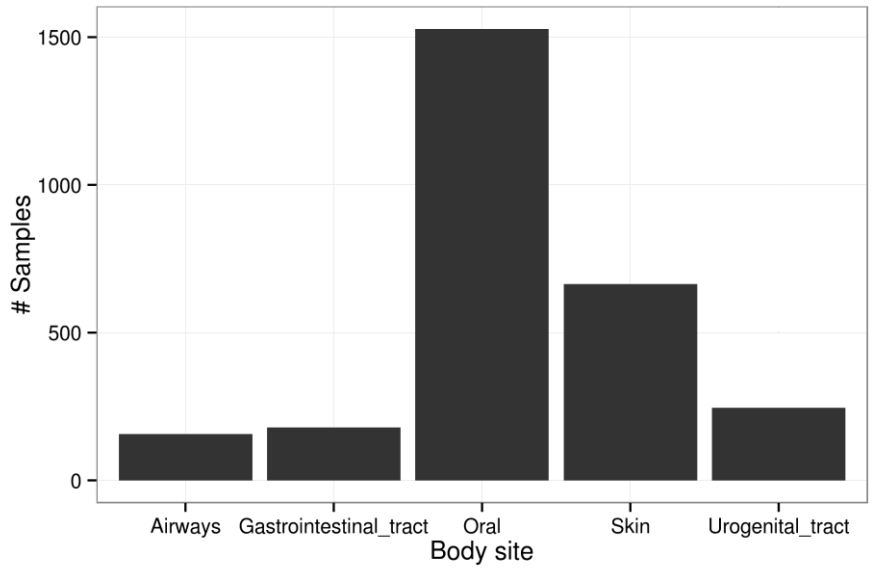

Figure 1 - Number of samples per body site, summed over all sampled individuals.

Species richness

The species richness analysis is summarized in Figs. 2-3. Overall, the GI tract seems to have the richest communities, followed by the oral, skin, airways, and finally, the UG tract. Interestingly, the two richest body sites also have the widest distribution of richness by sample, which may suggest that the sampling effort is insufficient in terms of the number of people sampled.

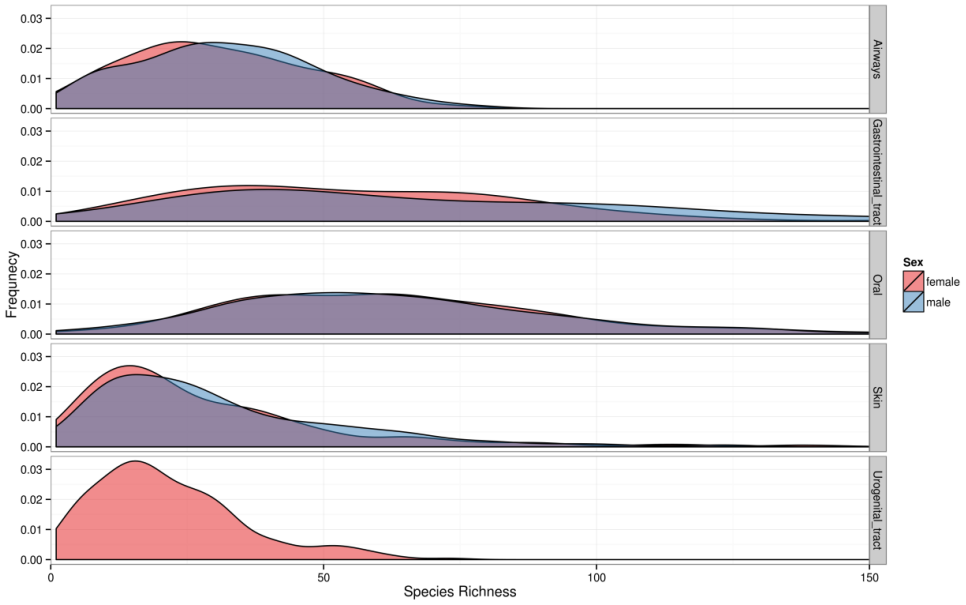

Figure 2 – Species richness distribution by sex (females in red, males in blue) and body sites (from top to bottom: airways, GI tract, oral, skin, UG tract). The richness distribution is similar between the sexes (confirmed by rarefaction which is not shown and consistent with similar number of male and female participants), but very different between body sites, where the GI tract and oral sites have a much wider distribution compared to the airways, skin and UG tract. See Fig. 3 for rarefaction curves.

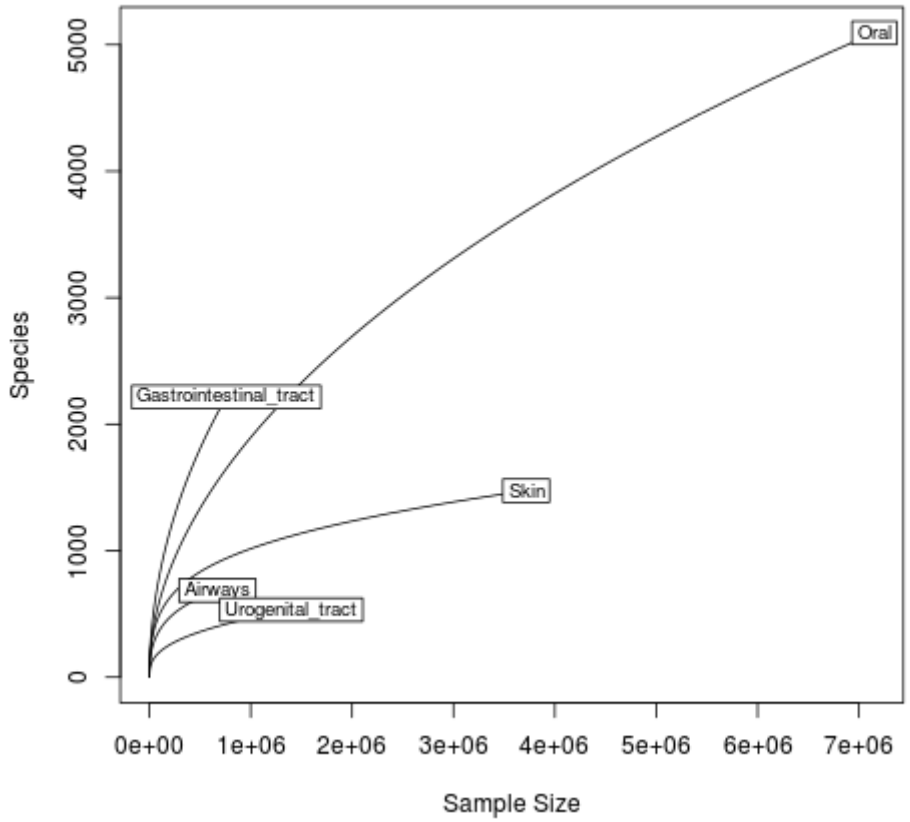

Figure 3 – Rarefaction curves for species richness by body site. This figure suggests that the GI tract would exhibit the highest species richness if the sampling effort was comparable.

Species diversity

Fig. 4 shows the distribution of true diversity (Shannon index) by sex and body site. Similar to species richness, there seems to be little effect of sex on diversity, and a significant effect of body site, with greater diversity for the "richer" body sites (see Fig. 3) and lower diversity for the "poorer" body sites.

The diversity profile (Fig. 5) is consistent with this observation, and the trend of diversity as a function of the Hill number is similar in all body sites, rapidly declining as the effect of rare species decreases.

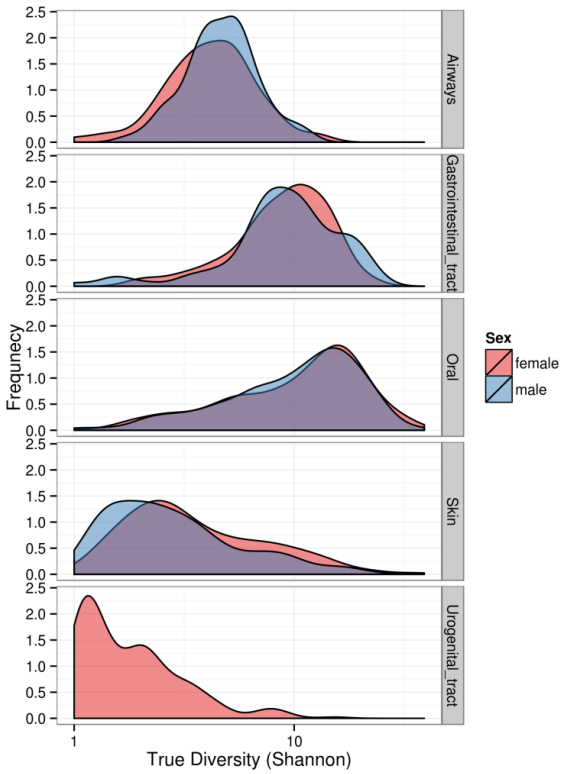

Figure 4 – Species "true" diversity by body site and sex. The distribution of true diversity (Shannon index) by sex (red for female, blue for male) and body site (from top to bottom: airways, GI tract, oral, skin, and UG tract).

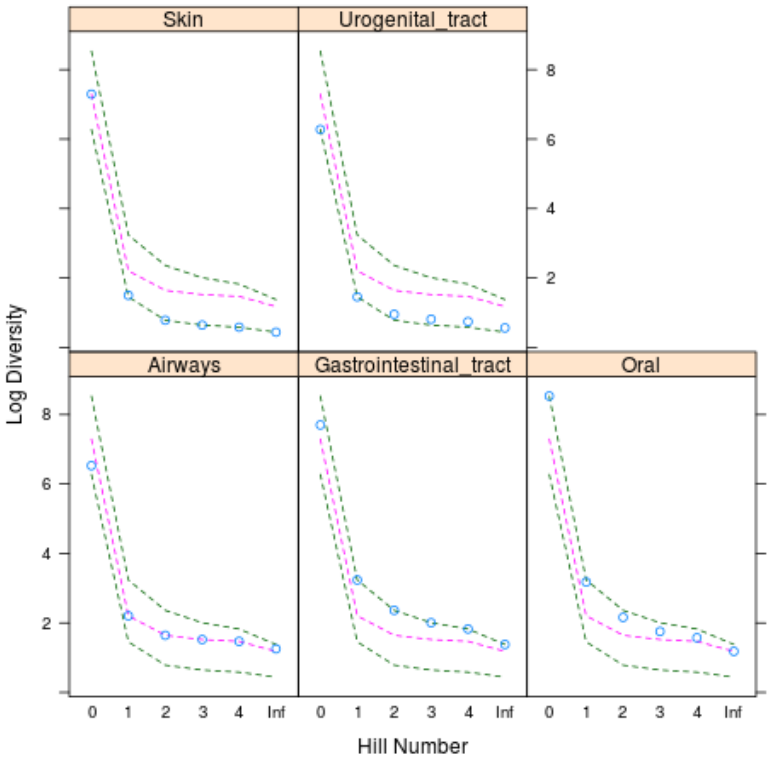

Figure 5 – Diversity profile by body site. The log of the diversity as a function of Hill numbers for the different body sites.

Beta diversity

I analyzed the beta diversity of the different body sites. The triplot in Fig. 6 shows that in this dataset beta diversity is mainly influenced by species that occur in one body site but not in the other (b and c), rather than by species that occur in both body sites (a).

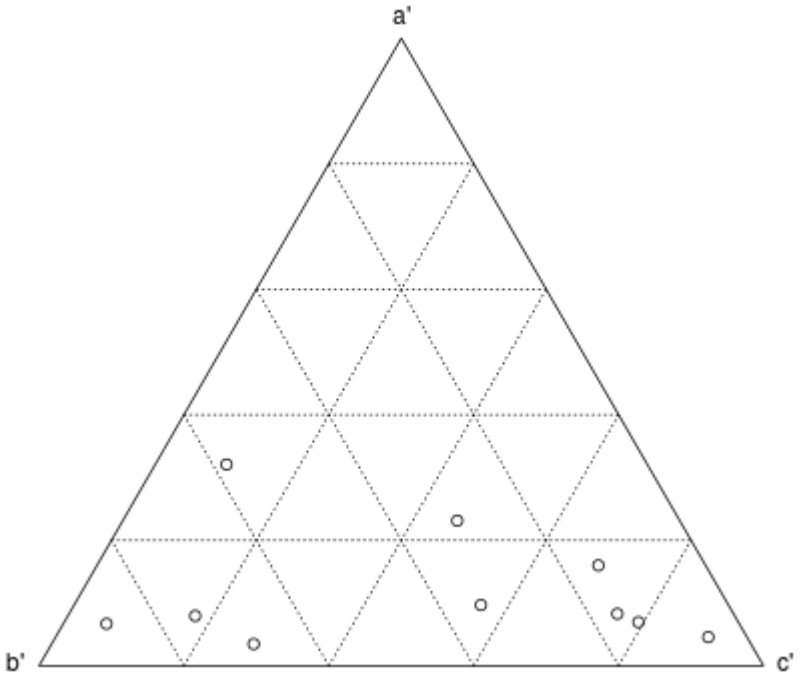

Figure 6 – Beta diversity triplot by body site.

Next, I analyzed three different beta diversity indices: Jaccard $\left(\frac{a}{a+b+c}\right)$, Simpson $\left(\frac{\min(b,c)}{\min(b,c)+a}\right)$, and Lennon $\left(\frac{2|b-c|}{2a+b+c}\right)$. These indices were chosen because they present relatively different features. Jaccard's index (Fig. 7, top panel) shows that the oral body sites share more species than the other body sites, on average, but there is a lot of variance. Simpson's index (Fig. 7, middle panel) is similar, but inverse, and puts less weight on species that occur only in one sample but not in another sample. Lennon's index (Fig. 7, bottom panel) checks the symmetry in segregation of species between two samples; the results are all around 0.5 but it seems that the oral samples have an increased variance, meaning that some samples had more unique species.

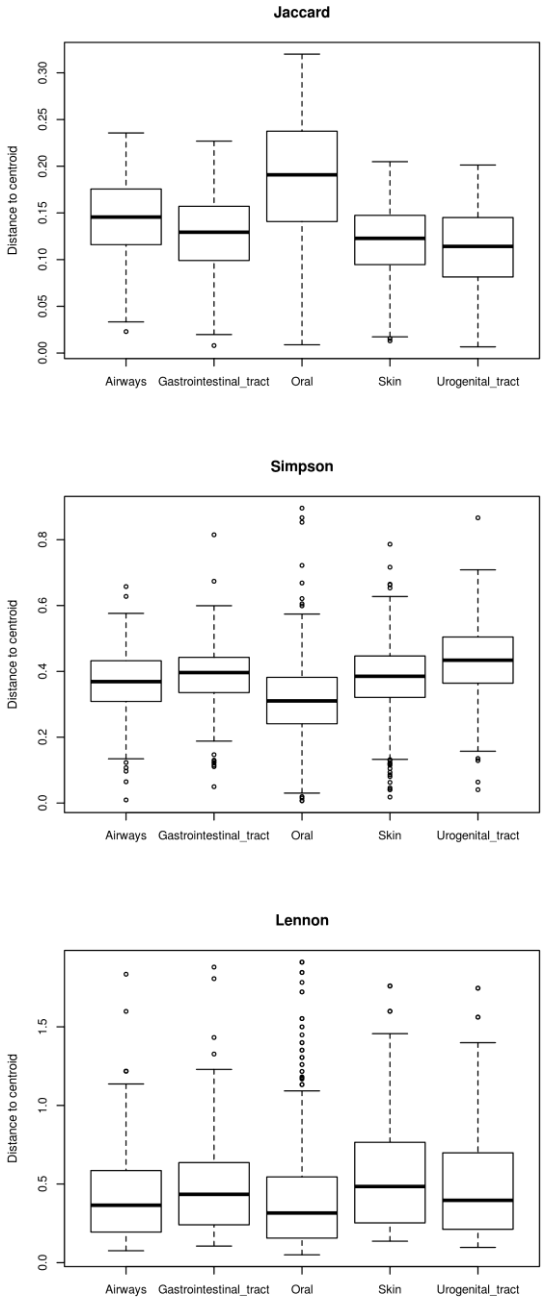

Figure 7 – Beta diversity by body site. The figure shows boxplots of three different beta diversity indices (Jaccard – top panel; Simpson – middle panel; Lennon – bottom panel; see text for description of indices) across five body sites.

Ordination

I performed nMDS analysis on the occurrence matrix to compute the species and sites ordination. The stress level was ~0.2.

Sites ordination

The sites ordination (Fig. 8) clearly shows that samples are clustered by body sites (marker color), but not sex (marker shape), with the only significant overlap between the skin and airways samples. Consistent with our observation from the beta diversity analysis, the variance in the oral samples is the largest, demonstrated by the relatively many green markers far from the centroid of the green cluster. However, this can be a result of an increased sampling effort (Fig. 1).

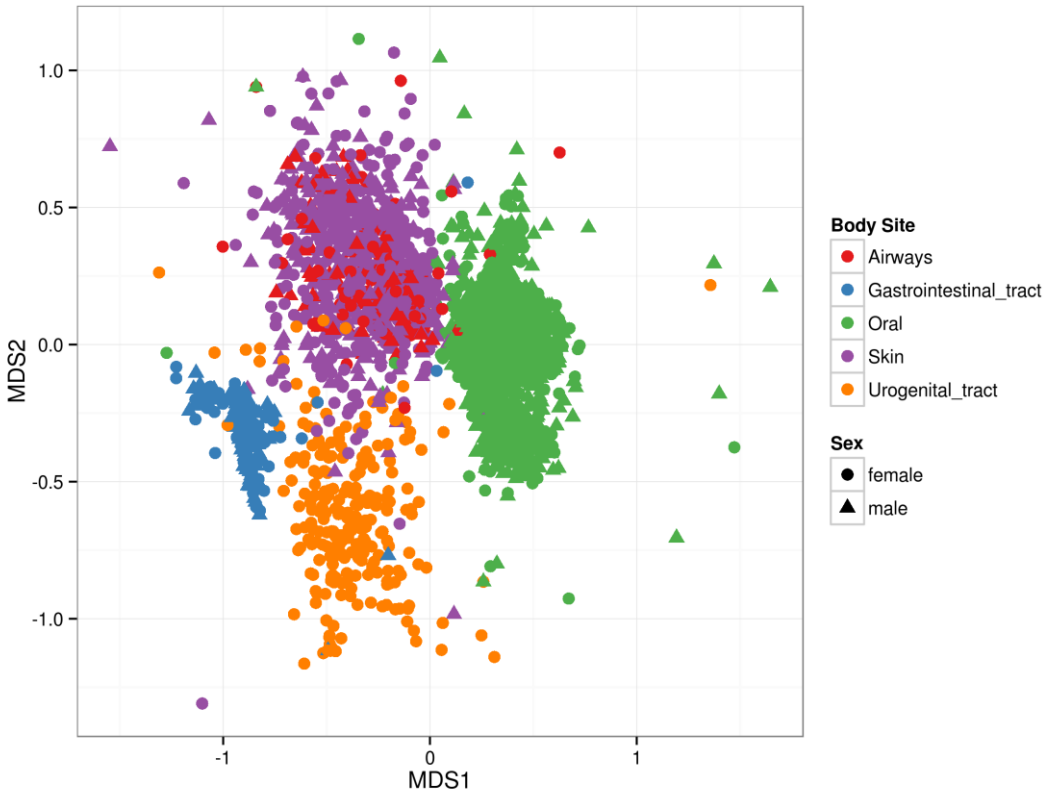

Figure 8 – Site ordination (nMDS). Sites are designated by body site (marker color) and sex (marker shape).

Species ordination

The species ordination (Fig. 9) shows that there are 3-4 distinct species clusters. The left and right clusters are composed of frequent species (species that appear in many samples, see marker intensity in top panel); the species in the left cluster are

mostly bacteroidetes, while the species in the right cluster are from all bacterial phylums (see color composition in the bottom panel). There is another cluster in the middle of the ordination plot (or maybe two clusters, one in the middle and one at the bottom); this cluster is composed of rare species of all phylums.

It seems that the species ordination has caught a glimpse of something – especially interesting is the left, homogeneous cluster – but the separation is lacking, either due to high dimensionality of the underlying data, the complex and somewhat arbitrary concept of OTUs, and/or the possibly messy and stochastic composition of bacterial communities.

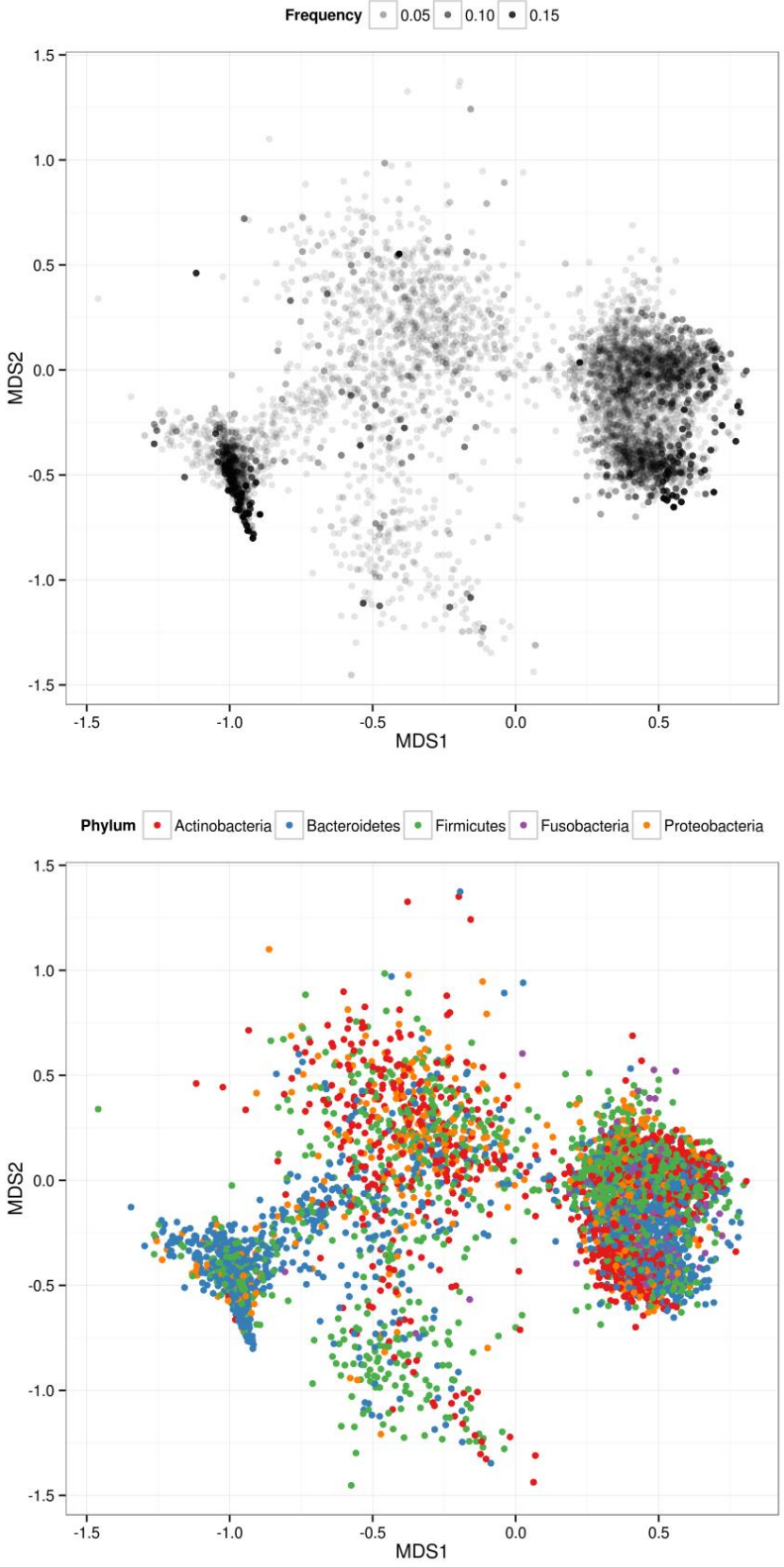

Figure 9 – Species ordination (nMDS). The top panel shows the species by their frequency in the samples (the darker the marker the more frequent the species); the bottom panel shows the species by their phylum (denoted by different colors; only the five most frequent phylums are shown). See legends above panel

Assembly rules

I analyzed the correlation in the species composition between the different samples to study which body sites have similar bacterial communities. Fig. 10 shows a correlation map in which each row and column correspond to a sample and samples are ordered by body sites. A correlation map between the different bacterial genuses did not yield interesting results, as the majority of the correlation coefficients were very low.

First, it is clear that the main diagonal is composed of high-correlation blocks and therefore there is a high correlation between samples from the same body site. This correlation is especially strong in samples from the oral body sites (marked "O" in the bottom and right axes), except for the gums samples (supra- and subgingival plaque; marked "Sub" and "Sup" in the top axis), which have a lower correlation with the other oral sites, possibly because the gums contain a repository of cavity causing bacteria.

Second, the airways samples (marked "A") have a relatively high correlation with the skin samples (marked "S") and to some extent with the oral and UG tract samples. This is in accordance with the overlap between the airways and skin samples in the ordination analysis (Fig. 8; red marks for airways, purple marks for skin).

Third, the GI tract samples (marked "G"), despite their high diversity (Figs. 4-5), have very low correlation with samples from other body sites, in accordance with their tight clustering in the site ordination analysis (Fig. 8; blue marks).

This trend can be the result of strict regulation by the host due to required function of the gut microbiome or due to the gut being a unique ecological niche. On the other hand, the airways and the skin are more suceptible to infection by random microbes and therefore present higher correlation with other body sites.

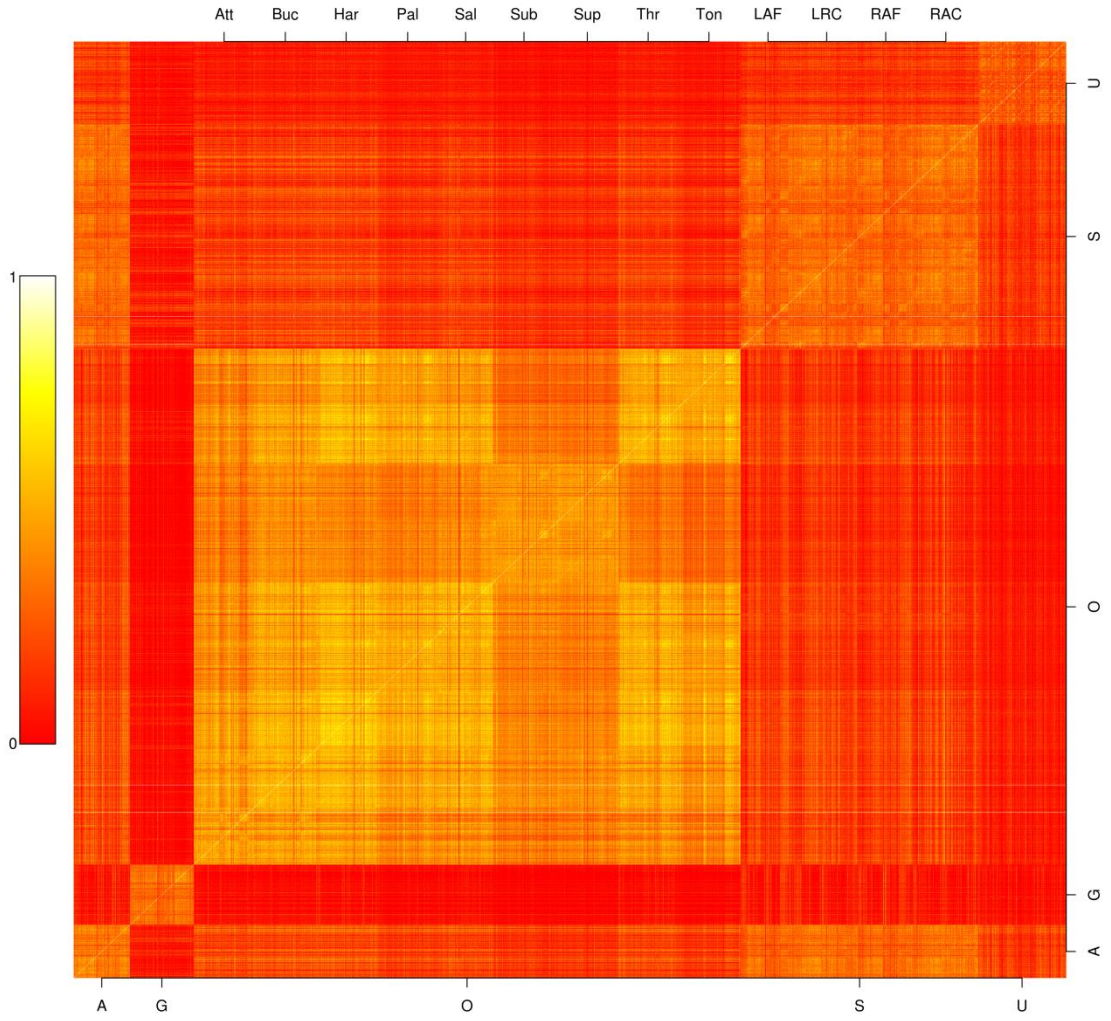

Figure 10 – Correlation between samples. Low correlation in red, high correlation in yellow (see colorbar on the left). The samples are ordered by body sites (bottom and right axes: A – airways; G – GI tract; O – oral; S – skin; U – UG tract) and sub-sites (top axis, oral: Att - Attached Keratinized gingiva; Buc - Buccal mucosa; Har - Hard palate; Pal - Palatine Tonsils; Sal – Saliva; Sub - Subgingival plaque; Sup - Supragingival plaque; Thr – Throat; Ton - Tongue dorsum; skin: LAF - Left Antecubital fossa; LRC - Left Retroauricular crease; RAF - Right Antecubital fossa; RAC - Right Retroauricular crease).

Part 2: Open questions

Question 4: What limits the number of mammal species in Israel?

The number of species in Israel is likely limited by several factors. In my opinion, the most important factor is the small area of Israel: with ~22,000 km^2 of which 2% is water and 55% is desert [1], and a dense urban and rural population in the temperate regions, there is very limited natural landscape for wildlife. The effect of limited space should be most evident in large mammals such as wolves, foxes, and deer, and I expect that human activities have reduced the species richness in the last 10,000 years, and more so since the foundation of the state of Israel.

An additional important limitation might be the dominance of human-dependent or human-facilitated species such as the house mouse, the roof rat, the mole rat, the golden jackal, the red fox, the wolf, the hyena, bats, the mountain gazelle, and of course the domestic cat and dog [2]. Species that enjoy from a positive interaction with humans have a competitive advantage over wild species that have a negative or null interaction with humans, causing the exclusion of the latter species and thereby a reduction in species richness. This effect may be aggravated by invading species [3] such as the nutria from South America [4], the Norwegian rat (the common or brown rat), and the northern palm squirrel from central Asia; however, it is not clear if invading species increase species richness by colonizing empty niches or reduce species richness by excluding local native species.

On the other hand, the number of mammal species is likely enhanced by the geographic and climate heterogeneity in Israel, both along the north-south temperature and precipitation gradient, and along the west-east geological and altitude gradient. This heterogeneity provides a diversity of ecological niches which can be occupied, in theory, by multiple different species. However, this niche occupation relies on migration, as the small mammal population sizes in Israel cannot not support a high rate of speciation.

Nevertheless, Israel is a land bridge between Africa, Arabia, and Eurasia. This location makes Israel a home for seasonal migrators, which probably doesn't affect

mammal species; however, this unique location allowed many mammal species to pass through Israel while migrating between the continents, which could have facilitated a large number of mammal species. Indeed, there is evidence that species from tropical Africa and Asia, Palearctic Eurasia,south central Asia, and the Sahara desert [5] were established in Israel at some point, suggesting migration from a wide range of areas and habitats.

Question 1: Deterministic or stochastic community structure?

In order to consider if the community structure of the human microbiome is deterministic or stochastic, first consider Fig. 8 above: given a sample of microbes from a random person and a random body site, one can predict which body site was sampled. This suggests a correlation between the environment (body site) and species abundance, which is indicative of a deterministic community structure.

A possible test of the community structure can be conducted by studying the microbial composition of ancient human fecal matter. The oldest examples are dated to the Middle Palaeolithic, some 50,000 years ago, and are attributed to Neanderthals from Spain [6]. If the microbial composition of ancient fecal matter is similar to that of contemporary humans – taking into account a similar diet and the use of antimicrobial drugs, among other factors – then it is likely that the community structure of the GI tract is determined by deterministic processes.

Further support has been given by a study from the Alm lab [7]. The researchers collected data on food intake and lifestyle, as well as daily salivary and fecal samples, from two individuals (the postdoc and the PI conducting the research) over a period of a year. The analysis of the community structure dynamics over time demonstrates that the microbial communities are stable for months. However, drastic lifestyle events (the postdoc's travel to Southeast Asia and the PI's Salmonella infection) rapidly changed the microbial community. Nevertheless, after return to previous conditions (the postdoc's return to the USA; the PI's recovery from the infection), the community rebounded, at least to a certain degree, as some species were replaced by similar ones. These findings strengthen the deterministic view: the stability to small perturbations (minor changes in diet and lifestyle over weeks and months) and

the replacement of species by similar species after major perturbations both suggest that the community structure is functional and deterministic rather than stochastic.

References

1. Area of Districts, Sub-Districts, Natural Regions and Lakes. *Israel Central Bureau of Statistics*. 11 September 2012.
2. Mendelssohn H and Yom-Tov Y (1999) A Report of Birds and Mammals Which Have Increased Their Distribution and Abundance in Israel Due to Human Activity, *Israel Journal of Zoology*, 45:1, 35-47.
3. Global invasive species database. Retrieved 1 July 2014.
4. Carter J and Leonard BP (2002) A Review of the Literature on the Worldwide Distribution, Spread of, and Efforts to Eradicate the Coypu (*Myocastor coypus*). *Wildlife Society Bulletin*, 30:1, 162-175.
5. Kolska Horwitz L and Tchernov E (1990) Cultural and Environmental Implications of Hippopotamus Bone Remains in Archaeological Contexts in the Levant. *Bulletin of the American Schools of Oriental Research*, 280, 67-76.
6. Sistiaga A, Mallol C, Galván B, Summons RE (2014) The Neanderthal Meal: A New Perspective Using Faecal Biomarkers. *PLoS ONE* 9(6): e101045.
7. David LA, Materna AC, Friedman J, Baptista MIC, Blackburn MC, Perrotta A, Erdman SE, Alm EJ (2014) Host lifestyle affects human microbiota on daily timescales. *Submitted*. Data deposited as EBI Study: ERP006059.