WORKSHOP IN COMPUTATIONAL BIOSKILLS (76552)
FINAL ASSIGNMENT - ARAD TAL AND ITAMAR NINI

RESEARCH QUESTION

In the original project conducted by Tal Raviv within the Moran Yassour Lab, the study explored the development of the infant gut microbiome across diverse lifestyles. Raviv's project encompassed two primary questions: firstly, it investigated the potential relationship between maternal lifestyle and the composition of the infant microbiome; secondly, it compared the microbiome of the Bedouin community with that of other communities. In our project, our aim was to investigate whether the similarity between infant microbiomes and their respective mothers' microbiomes varies across different communities.

DATA

In the study, data were gathered from mothers and their infants shortly after childbirth in three distinct communities: Finland, Boston, and the Israeli Bedouin community. The dataset comprised two separate sheets - one containing metadata and the other detailing the composition of the microbiome. Samples were collected from various sources: infant samples were acquired from stool samples, while maternal samples were obtained from rectal, vaginal, or stool samples. The metadata spreadsheet comprises 2,140 samples listed as rows, with various properties represented as columns. These properties include the sample's source, family, time elapsed after childbirth when the sample was collected, the community to which the clients belong, the predominant bacteria in the sample along with their respective abundances, as well as the Shannon index and Chao1 values, which serve as indicators of microbiome diversity within each sample. The second spreadsheet features different bacterial species listed as rows, with each column corresponding to a specific sample. The values in this sheet represent the percentage of a particular bacterium within the respective sample.

EXPLORATORY PLOT

Raviv presented PCA results for infants and mothers separately. As we want to compare infants and mother microbiomes, we decided to apply the same process to the combined dataset. This enabled us to explore similar and different trends in comparison to the PCA results obtained from the separate datasets. We choose those color and shape as we want maintain consistency with Raviv's plots.
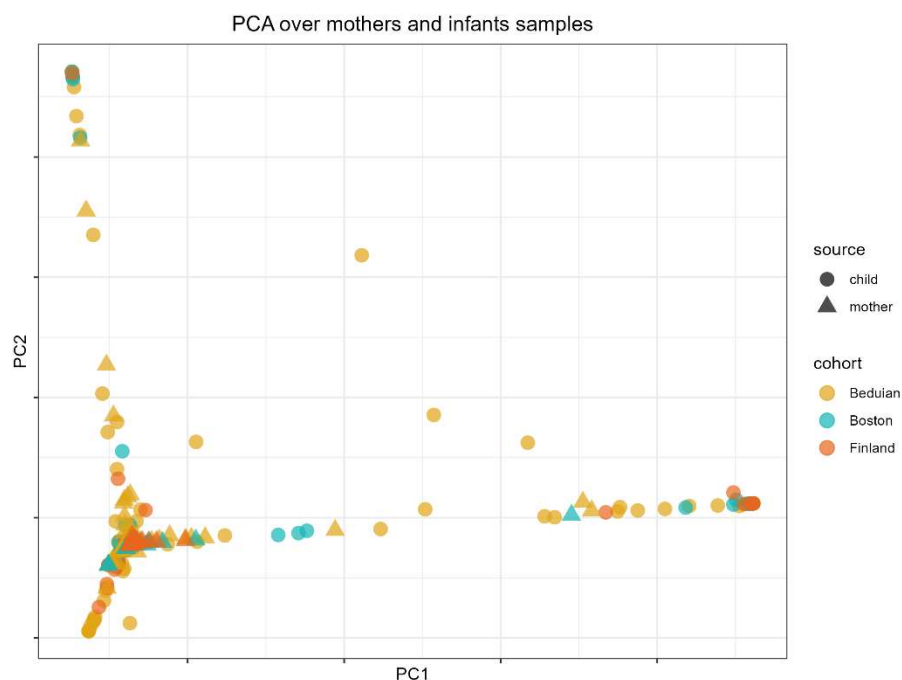
FIGURE 1. PRINCIPAL COMPONENT ANALYSIS

PCA applied both maternal and infant data. The dataset was refined by selecting the earliest rectal or toll sample from each individual. Vaginal samples were depleted. Communities are distinguished by color, while source attribution is represented by shape.

EXPLANATORY PLOT

We created this plot to answer whether the differences between mothers and infants across varies communities are significant. Box plots enable us to easily grasp different statistical properties and compare them with the other groups. It is also enabled to show the pairwise p-value easily. As in the previous figure, the colors maintain consistency.
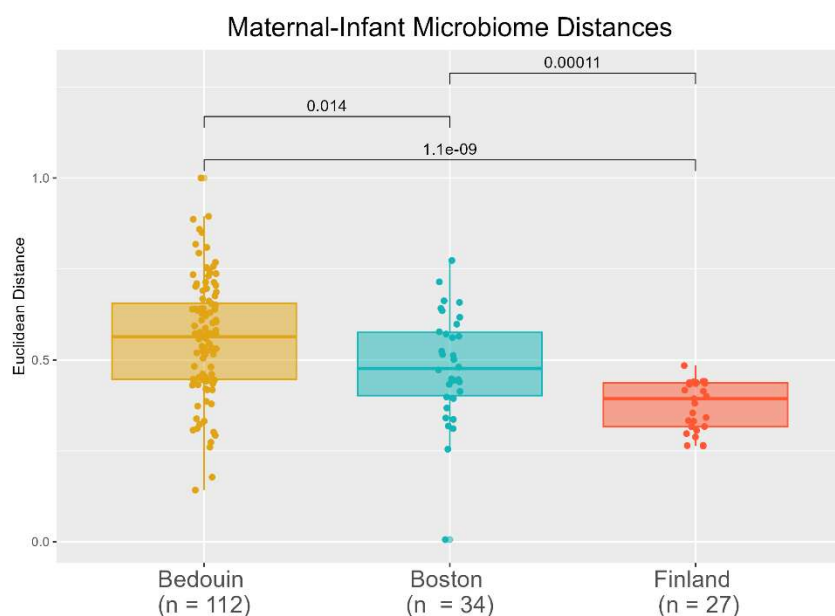

FIGURE 2. EUCLIDEAN DISTANCE BETWEEN MOTHER AND INFANT MICROBIOME SAMPLES.

For each mother-infant pair the mean Euclidean distance between their microbiome samples has been calculated and normalized by dividing the it by the maximum distance of a single pair within the 3 groups. P-value were calculated using pairwise mean comparison. N indicates mother-infant pairs amount at each group.