

Workshop in Computational Bioskills - Final assignment

Noam Rosenmann And Tsori kislev

Research Project Background:

The research project focuses on analyzing microbiome samples collected from mothers and their infants from a Bedouin community attending Soroka Medical Center. The primary goal is to investigate the microbial diversity and composition in relation to different factors such as birth mode, breastfeeding status and time points.

Research Question:

The central research question of this study is: How does microbial diversity and composition vary with birth mode (C-Section vs. Vaginal), breastfeeding status (Breastfeeding vs. Formula) and time point (e.g., different stages of infant development)?

Data Description:

The dataset includes microbiome samples from mothers and infants collected at Soroka Medical Center. Samples were taken from various sources (vaginal, oral, skin, rectal) and sequenced using MetaPhlAn to identify microbial taxa.

The data consists of two main files:

- `samples_metadata_and_essentials.tsv`: This file contains metadata and essential details about the samples. It includes information on the read depth before and after processing, percentage of unclassified reads, diversity indices (Shannon and Chao1), and dominant species/genus in each sample.
- `species_mpa.txt`: This file provides the microbial composition data, specifically the abundance of various bacterial taxa in each sample, generated by the MetaPhlAn software.

Data Structure:

- **Rows:** Each row represents an individual sample.
- **Columns:**
 - In `samples_metadata_and_essentials.tsv`: Includes columns such as `read_counts_in_bam`, `read_counts_after_preprocessing`, `percentage_unclassified_by_metaphlan`, and diversity metrics `shannon` and `chao1`.
 - In `species_mpa.txt`: Includes columns for microbial taxa and their counts.

The dataset includes multiple samples collected from different time points and birth modes. The units of measurement include read counts, percentages, and diversity indices.

Exploratory Plot:

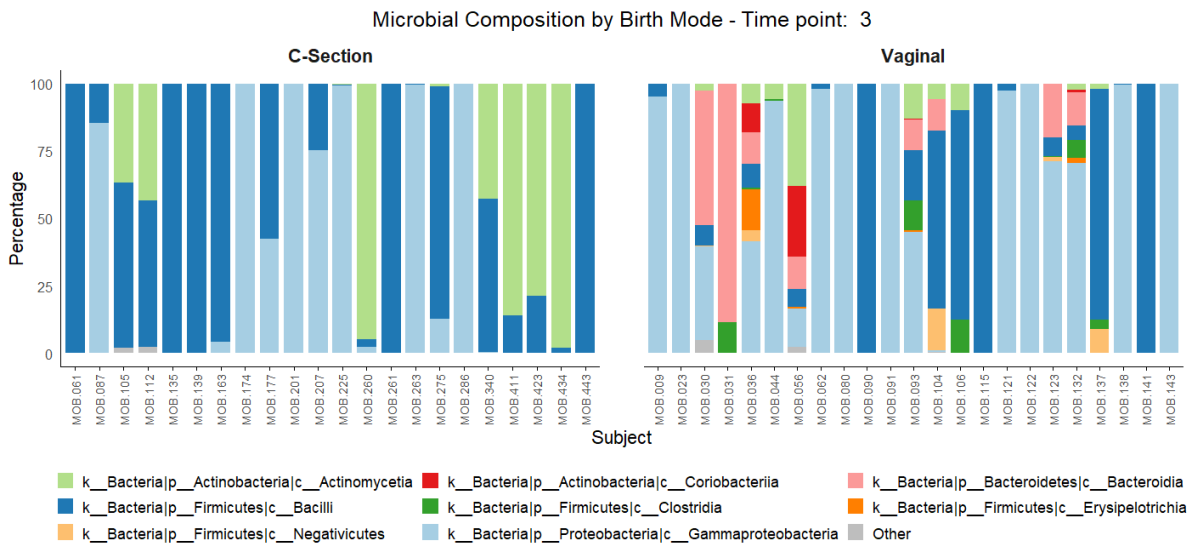


Figure 1: Microbial Composition by Birth Mode at Time Point 3.

The stacked bar plot represents the relative abundance of various microbial taxa at the class level across individual subjects, separated by birth mode (C-Section on the left and Vaginal on the right).

Discussion: This exploratory plot visualizes the microbial composition of infants at Time Point 3, with the subjects grouped by their mode of delivery (C-Section or Vaginal). The diversity and composition of the microbial communities differ notably between the two groups. Infants delivered via C-section exhibit a more homogeneous microbial composition, while infants delivered vaginally show a more varied microbial community. This suggests that vaginal birth may expose infants to a broader range of microbes, potentially contributing to a more diverse microbiome.

The choice of a stacked bar plot effectively highlights these differences in microbial composition and allows for a detailed comparison of the relative abundance of microbial classes between the two birth modes. The use of distinct colors for each microbial class further aids in visual distinction, making it easier to identify trends and variations across individual subjects.

Explanatory Plot:

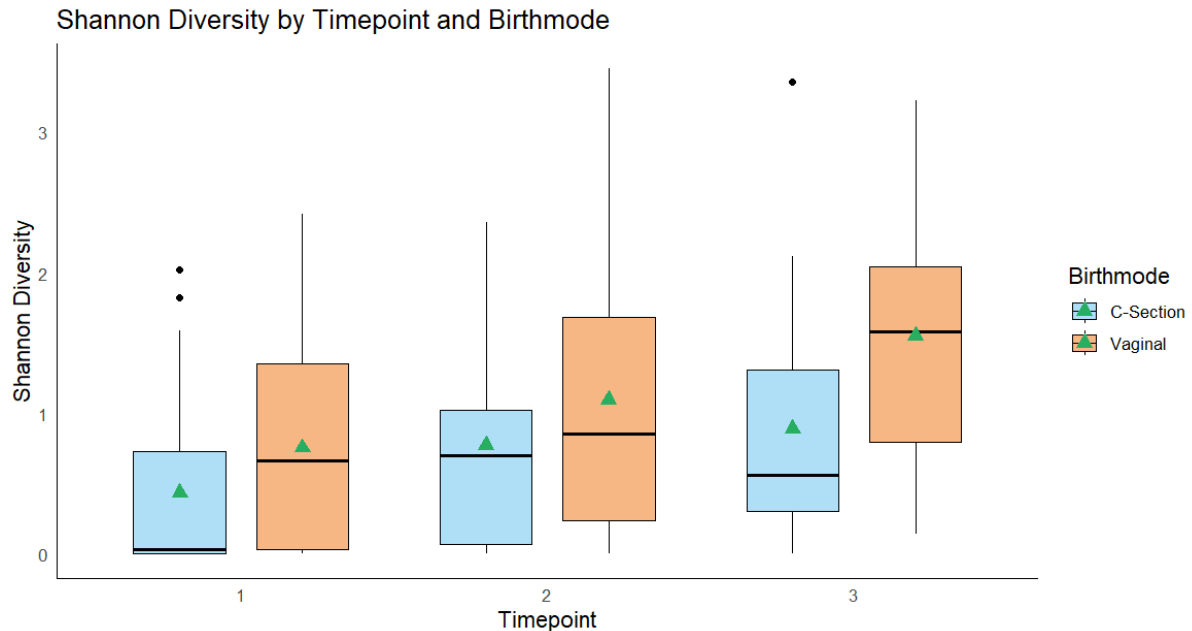


Figure 2: Shannon Diversity by Timepoint and Birthmode.

The box plot displays Shannon diversity indices across three time points, categorized by birth mode (C-Section and Vaginal). The green triangles represent the mean diversity for each group, while the boxes show the interquartile range (IQR), with the whiskers extending to 1.5 times the IQR.

Discussion: This plot illustrates the relationship between Shannon diversity (a measure of microbial diversity) and birth mode across three time points. The data suggests that, on average, infants delivered vaginally tend to have higher microbial diversity than those delivered via C-section at all time points. Additionally, there is a noticeable increase in microbial diversity across time points for both groups, indicating that microbial diversity develops as the infants age, regardless of birth mode.

The boxplot effectively conveys these differences, with clear distinctions between the groups and timepoints. The use of color coding for birth mode helps in visualizing the trends, while the inclusion of median values and outliers provides additional insight into the distribution and range of diversity scores.