Supplementary Material

1 SUPPLEMENTARY TABLES AND FIGURES

1.1 Tables

Q1: Conceptually, what information/knowledge can we gain or would we like to obtain from doing microbiome research?
* Influence of food on obesity * How drugs change the gut microbiome * Effect of past diseases (how gut microbiome changes after having had certain disease e.g. salmonella) * Effects on plant growth/production * Role in diseases * Define and find differences in different health and disease statuses. * Effect on health * Influence of food and other factors on microbiome * Microbiome effects on plants (yield more, more productive agriculture) * Impact of bacteria metabolites on treatment efficiency * Role of gut microbiome in autoimmunity (e.g. how microbiome can influence t-cells) * Effect on immune response * Precision medicine * Influence on psychology (mood, depression, . . .) * Who has a dog? (based on skin microbiome) * Changes during day or different seasons * Who lives with whom? (instead of dna, use microbiome to see who was in contact with whom) * Composition of microbiome with varying diet * Microbiome and survival * Composition of microbiome with respect to body combacting diseases * How ”quickly” does diet have effect on microbiome? * Link between diet and longevity? * Use of fecal therapy * Relation between microbiome and MS * How do drugs change the microbiome, and at the same time affect other clinical variables * Diversity and richness

Table S1. Answers provided by the workshop participants to question 1.

Q2: Which data is required or relevant to obtain this knowledge?
* Time * Location * Gender * Age * Exposure data (what did you eat, did you use drugs) * Species (mice, human, other animal) * Disease status (after treatment, before treatment) * Time series (longitudinal sequencing, how dynamic, how quickly does it change) * Which bacteria? + count (absolute) * Diet * Genome * Timestamps * Location of microbiome sample in the body/plant) * Baseline characteristics (e.g. age, gender) * Environmental information * Epidemiological information * Diet * Clinical trial data * Sequencing (amplicon, whole gen.) * Longitudinal information * Treatment and other clinical variables * Disease status/clinical outcome * Metabolic information * Hormonal changes * Microbiomes of contacts * Lifestyle * Health status * Gender * Time and age * Immune response over time

Table S2. Answers provided by the workshop participants to question 2.

Q3: To answer questions of Q1: which specific aspects can be retrieved from the OTU/ASV abundance table?
* Most present taxonomies in collected samples * Taxonomic abundance * Variability of abundances * p-values * Effect sizes (e.g. diff. in relative or absolute abundance between two conditions) * Co-abundance (association measures) * Diversity indices * Absolute/relative abundance (important difference) * Metabolites * Pro/anti-inflammatory profile * Integrated exploration of raw data (of all data listed before) * Variance between different environments/treatments (how different are they) * Variance after re-sampling/removing noisiness, duplicates * Effects of baseline characteristics * Alpha and Beta diversity * Correlation between taxa * Alpha diversity * Biomarker discovery * Relative abundance * Interaction network * Identify pathogens * Metabolite classification * Phylogenetic level * Fat content * Metabolism

Table S3. Answers provided by the workshop participants to question 3.

Q4: Given the aspects you wrote down before, can you think about methods needed and or used (statistically, visually) to obtain this information.
* PCoA (and related ordination methods) * Statistical modeling + differential abundance analysis (DESeq2, EdgeR, ANCOM2) * Venn diagrams (to display; most present taxonomies, compare taxonomic ranks between different samples) * Differential abundance methods (parametric/ non-parametric/ semi-parametric) * (stacked) Bar charts (used to display metabolites) * Dendrograms * Custom visual designs * Parallel coordinates (to display different types of information) * Sample clustering (tsne, UMAP, STAD) $\rightarrow$ scatterplot, node link * Heatmaps (OTU abundance vs timepoints) * Upset (replacement for venn diagram if more than three categories)

Table S4. Answers provided by the workshop participants to question 4.

Q5: When you think about your own research, I'm interested in the platforms, tools, packages you have used, or are using currently to analyze the microbiome. Can you list these up?
* Phyloseq * Microbiome * DESeq2 (diff abundance) * Limma (diff abundance) * EdgeR (diff abundance) * Piur (probabilistic index models) * HMP (Human microbiome project) (data) * phyloseq (R-package) * MicrobiomeHD (datasets + methods) * PICRUST * ggplot2 (for plotting) * D3 * HTML/CSS/JS Svelte * Databases (SQL, graph, document store) * Webserver * Pen + paper

Table S5. Answers provided by the workshop participants to question 5.

1.2 Figures

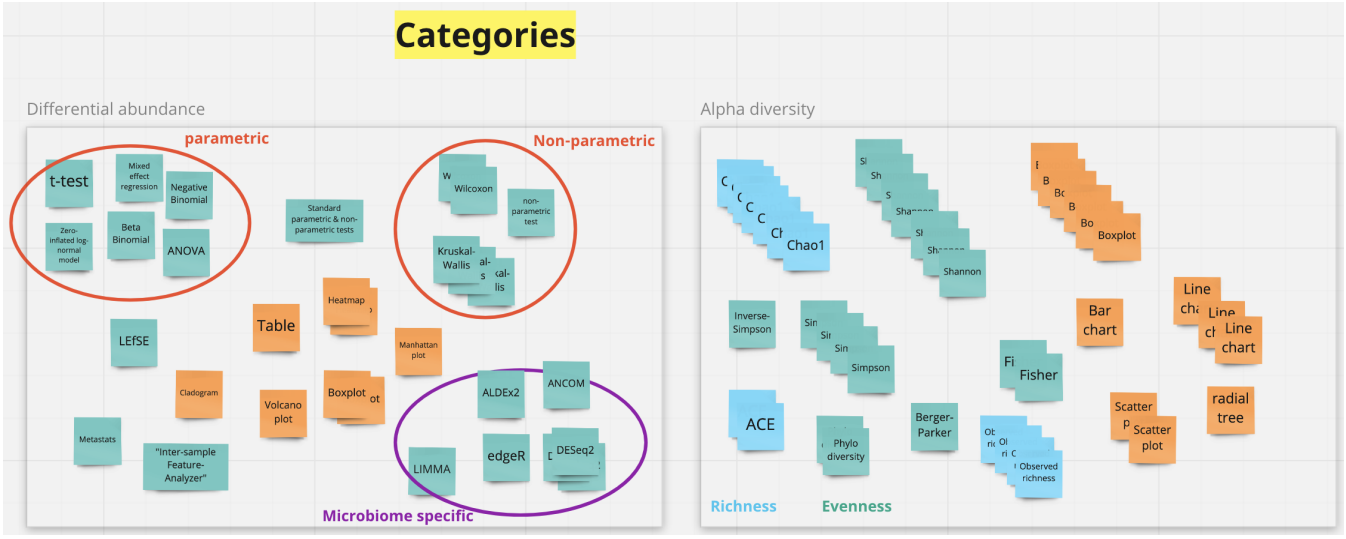

Figure S1. A reproduction of the initial card sort game, displaying how cards are sorted within 2 categories based on frequency of occurrence. The actual exercise were carried out using post its and markers.

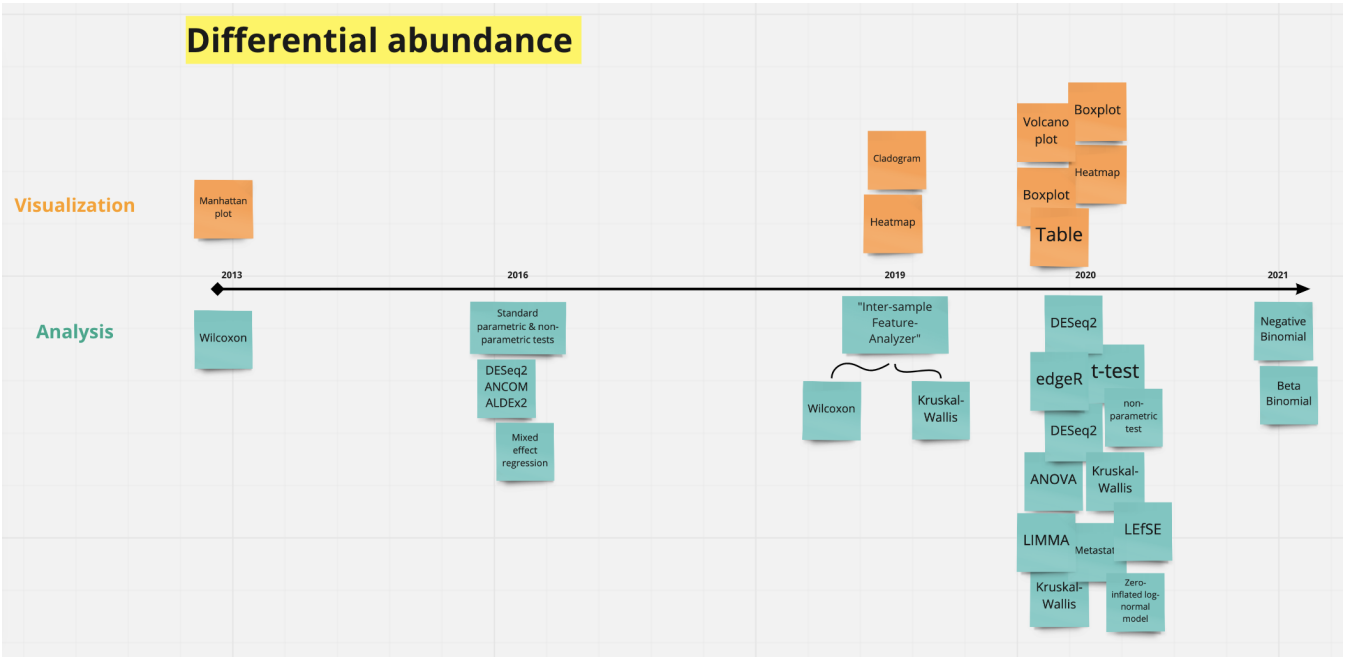

Figure S2. A reproduction of a history map exercise conducted to answer question 2 (i.e. How did the methods used to capture these microbiome aspects develop or change over time?) on differential abundance. The actual exercises were carried out using pen and paper.