H3ABioNet

Pan African Bioinformatics Network for H3Africa

16SrRNA Intermediate Bioinformatics Online Course: Int_BT_2019

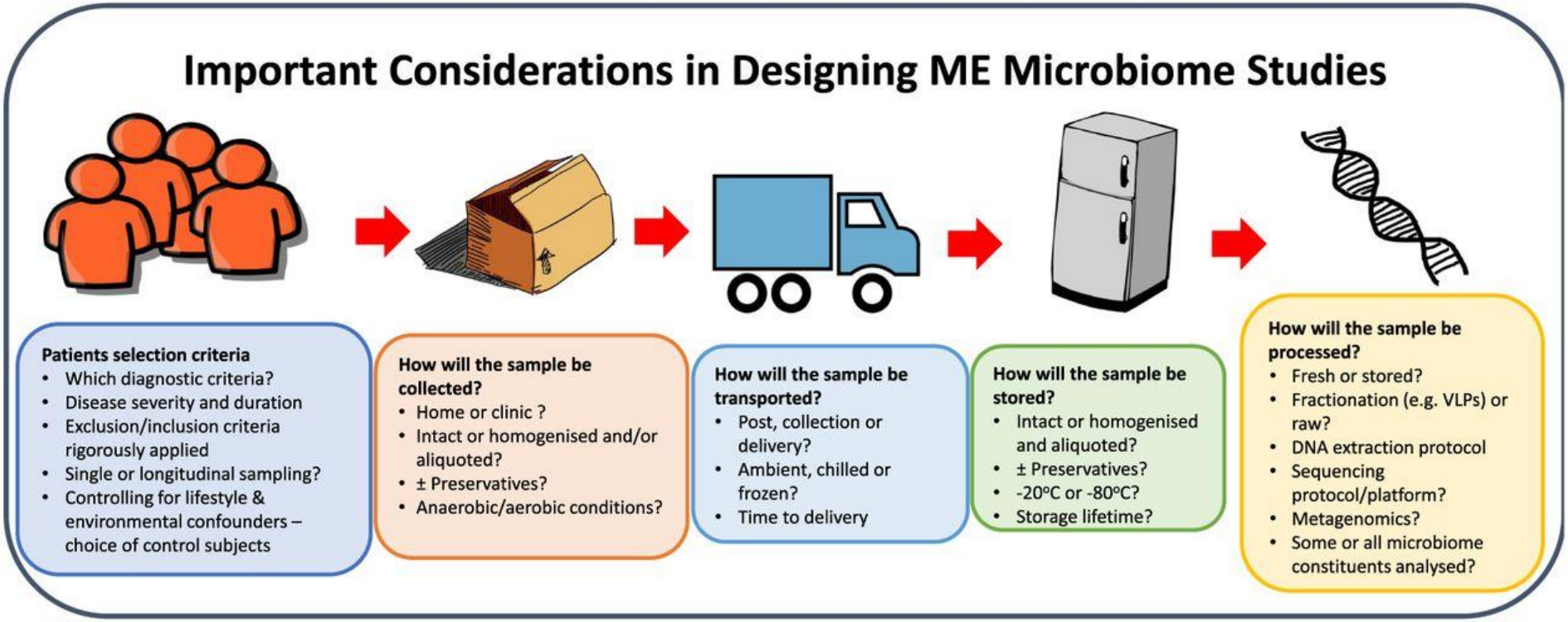
Module 3:

Sample collection, extraction and library prep for 16S NGS analyses

Part 3.2

16S rRNA high throughput sequencing: study design and sample collection

16S rRNA high throughput sequencing: study design

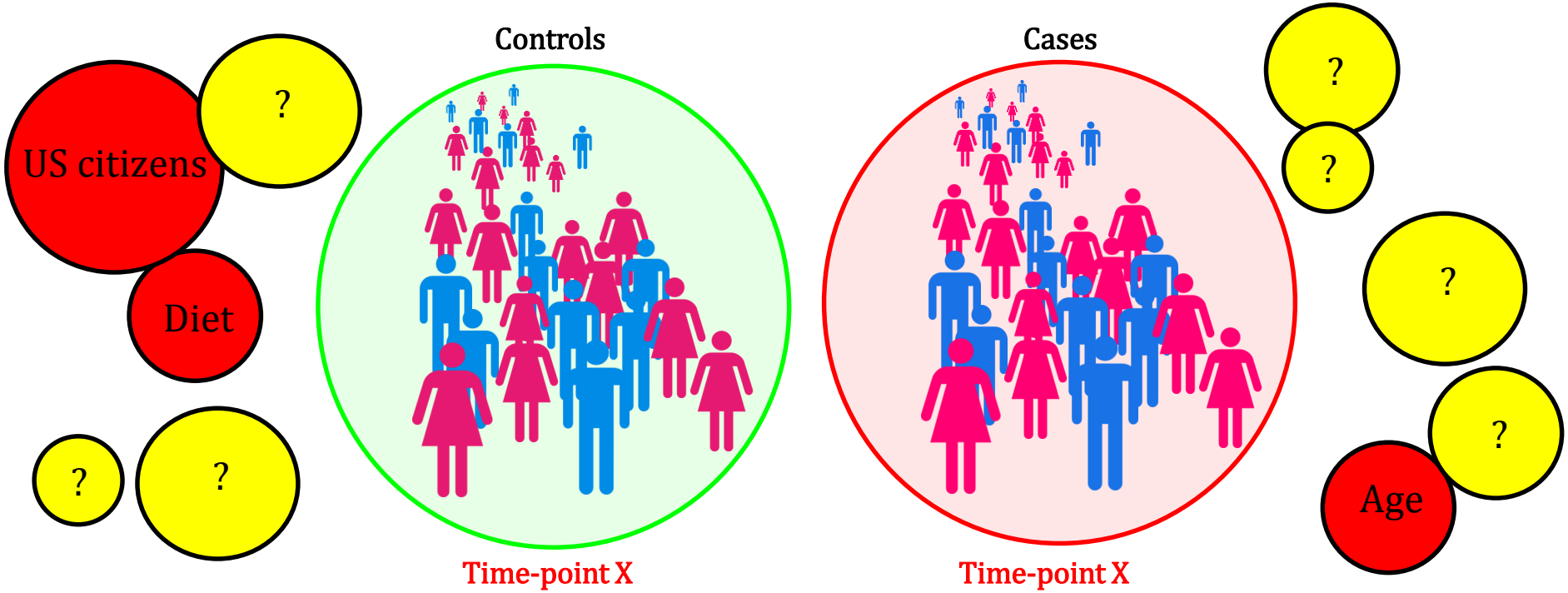


Recommendations:

- Consistent use of verified diagnostic criteria involving patients of similar disease severity and duration. Archiving and access to as detailed patient clinical metadata as possible
- Healthy subjects (parents/carers) living in same environment/household as patients can help control for certain environmental confounders of microbiome analysis
- Longitudinal sampling best suited for identifying temporal changes associated with changes in environment, behaviour and/or lifestyle
- Consider recent use of drugs/medications/antibiotics, probiotics and laxatives, and confounding (GI)-disorders (IBS, IBD, obesity) as exclusion criteria
- Minimise sampling handling, aliquot and freeze at source and keep frozen until processed, ideally within 36h
- Long term storage at -80°C to preserve as much sample integrity as possible
- Standardize all sample processing procedures including DNA/RNA extraction, sequencing and bioinformatics pipelines

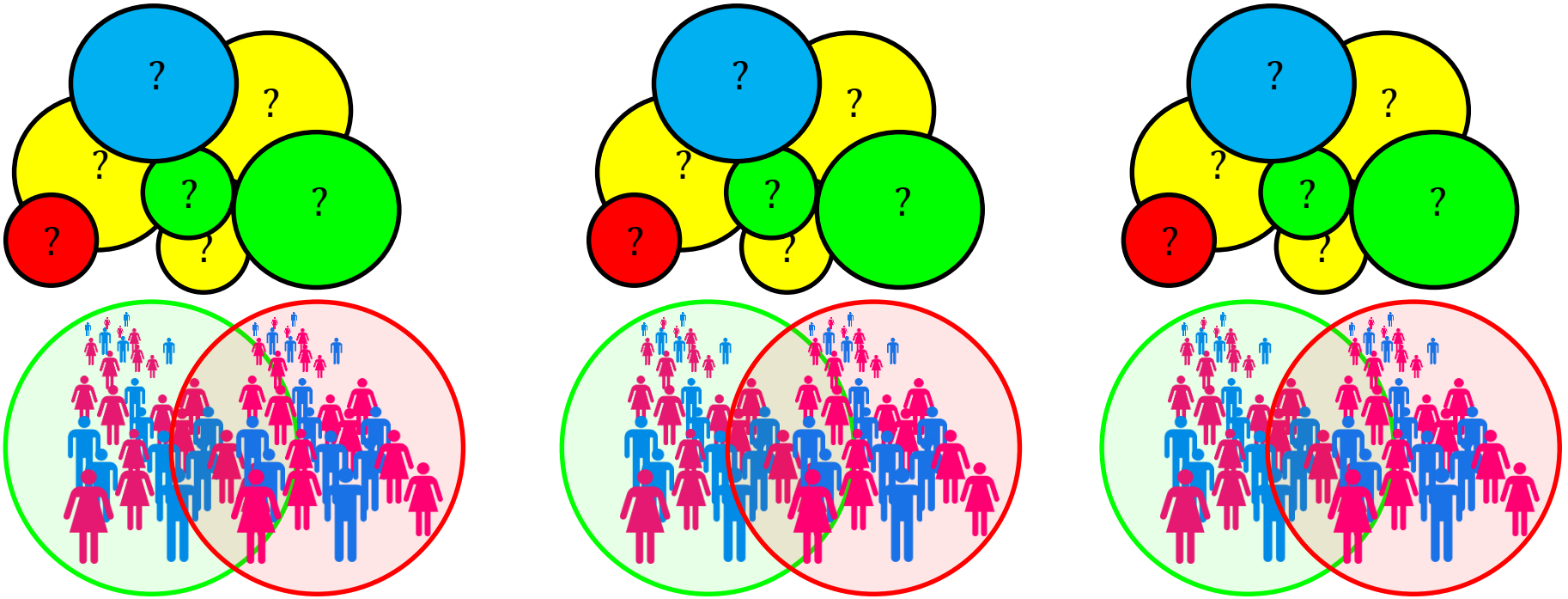
http://www.clinsci.org/content/132/5/523

16S rRNA high throughput sequencing: study design

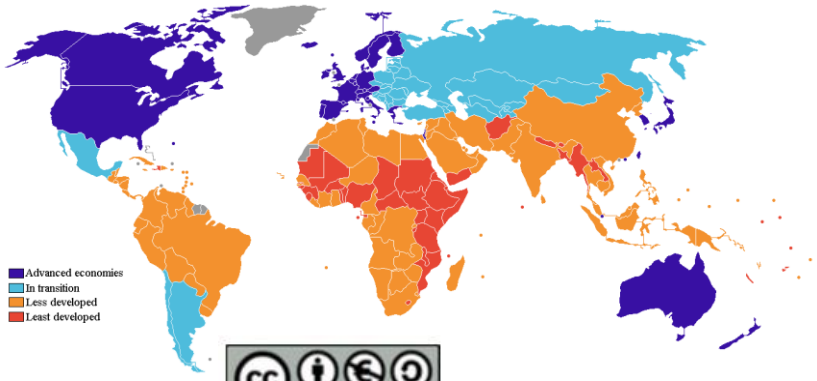


Vandeputte et al. (2017) *FEMS Microbiology Reviews* 41:S154–S167

16S rRNA high throughput sequencing: study design



Vandeputte et al. (2017) *FEMS Microbiology Reviews* 41:S154–S167



16S rRNA high throughput sequencing: study design

Review Article on From Microbe to Microbiome: New Implication in Respiratory & Critical Care Medicine

Insights into study design and statistical analyses in translational microbiome studies

Jyoti Shankar

J. Craig Venter Institute, Rockville, Maryland, USA

Correspondence to: Jyoti Shankar, MBBS, PhD. J. Craig Venter Institute, Rockville, Maryland, USA. Email: jyoti.shankar@gmail.com.

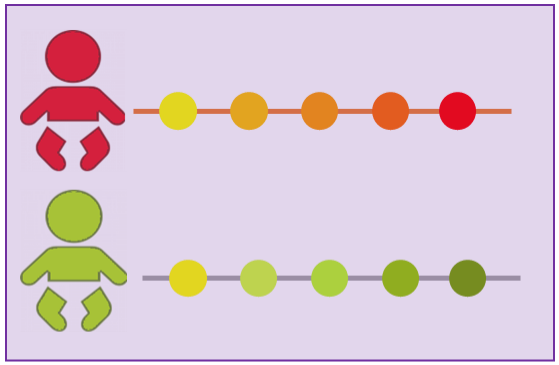
- Because studies cannot include entire populations, it is crucial to define the target population of interest and then draw a representative study sample to ensure that the findings from the study are generalizable
- Given the complex disease targets that translational studies attempt to understand, it is inevitable that diseased populations of interest are heterogeneous in their clinical phenotypes.
- While heterogeneous phenotypes could enable investigators to understand several facets of a disease spectrum, the presence of heterogeneity dilutes the statistical estimates of effect sizes of the microbiome.
- This problem is compounded by the fact that typical effect sizes of individual members of the microbiome are weak.
- The dilution of effect size estimates is even more acute for diseases with several complex phenotypes.
- If there is prior evidence for substantial clinical heterogeneity or if there are theoretically defined subpopulations with different disease characteristics within the population of interest, it is prudent to prioritize specific aspects of the disease to study and recruit a relatively homogeneous study population. For example, in a study examining the role of the microbiome in the development of pneumonia, it could be beneficial to focus on the most common type of pneumonia in the patient population rather than trying to profile all of the different etiological types.

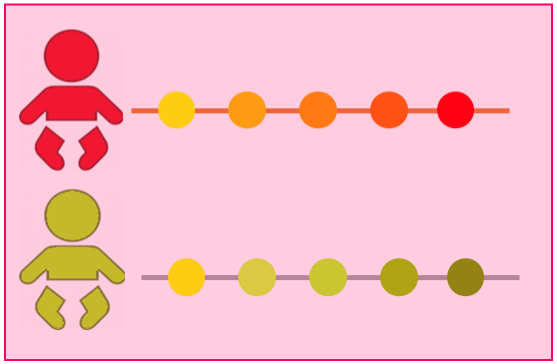
16S rRNA high throughput sequencing: study design

Step 1

Design your experiment based on your research questions (think about your study design)

Case-control set 1





Case-control set 2

Plate 1

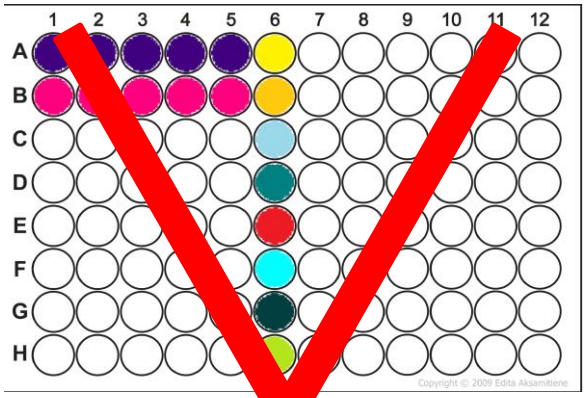


Plate 2

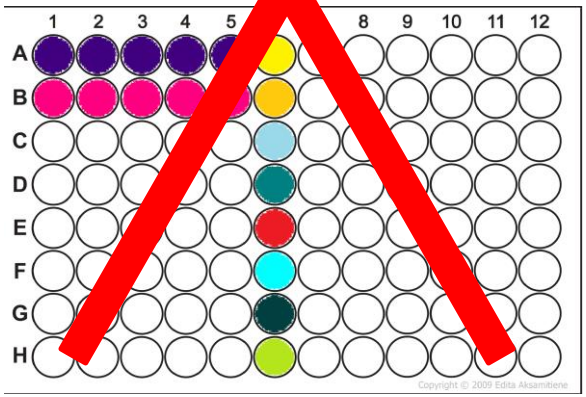


Plate 1

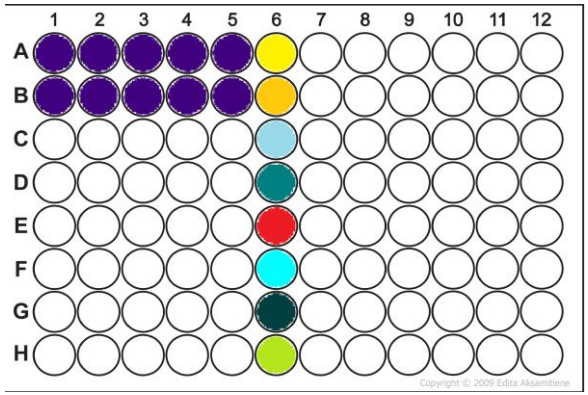
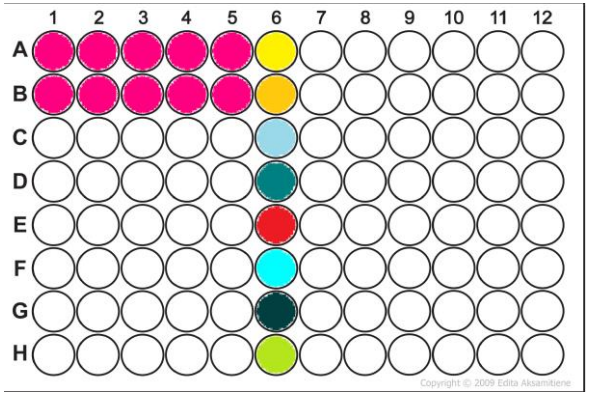


Plate 2



16S rRNA high throughput sequencing: sample collection

- Ideally, faecal material intended for microbiome monitoring needs to be **frozen immediately** after sampling in order to **stop the growth of residing bacteria and potential contaminants** and to conserve baseline microbial abundances.

- Subsequently, samples should be stored at –80◦C until DNA extraction.
- As sampling is often performed in the comfort of the participants' home, the latter could cause a significant logistical burden.
- Furthermore, faecal microbiome monitoring efforts risk to suffer from selection biases and drop-out associated to personal aversion towards faecal sampling—especially when sampling procedures are experienced as overly laborious.

Vandeputte et al. (2017) *FEMS Microbiology Reviews* 41:S154–S167

16S rRNA high throughput sequencing: sample collection

	Metagen-omics	Transcri-ptomics	Metabol-omics	Advantages	Disadvantages	Quality observed microbiome composition*
Freezing (–20°C)	✓	✓	✓	Culturing possible Long-term storage	Need for equipment and electricity Cold chain management necessary	*****
No buffer (4°C)	<24 [a,b]	✗	<2h[c]	Slows down bacterial growth Slows down fermentation Culturing possible	Need for equipment and electricity Limited amount of time (hours)	*****
No buffer (RT)	<24h[a,d]	✗	✗	No need for equipment and electricity Affordable (e.g. dry swabs, tubes)	Limited amount of time (hours) With swabs: initial weight not determinable	*
TE buffer (RT)	<24h[b]	✗	✗	No need for equipment and electricity	Quantitative analyses need upfront weighing step Limited amount of time (days)	*****
RNA later (RT)	<24h[e,f]	<6d[g]	✗	No need for equipment and electricity	Lower DNA yield[e,h] Lower DNA purity[f] Expensive (relative compared to freezing) Quantitative analyses need upfront weighing step Limited amount of time (days)	**
Ethanol (95%) (RT)	<2d[i]	<2d[i]	<4d[j]	No need for equipment and electricity	Lower DNA yield[k] Flammable reagent (increased shipping costs) Removal of buffer necessary Quantitative analyses need upfront weighing step Limited amount of time (days)	*****
RTTVs (RT)	✓	<28d[l]	✗	No need for equipment and electricity High DNA and RNA recovery[l] Immediate homogenization	Quantitative analyses need upfront weighing step Limited amount of time (weeks) Expensive (relative compared to freezing)	***
Carry Blair (RT)	✓	✓	✓	No need for equipment and electricity Affordable Culturing possible	Overgrowth gram negative bacteria likely Limited amount of time (hours-days) With swabs: lower DNA yield and initial weight not determinable	*****
FTA cards (RT)	✓	✗	✗	No need for equipment and electricity	Initial weight not determinable Small amount of sample, not easy to process	**
FOBT/FIT (RT)	✓	✗	✗	No need for equipment and electricity Large sample collections available from colon cancer screenings	Initial weight not determinable Small amount of sample	*****

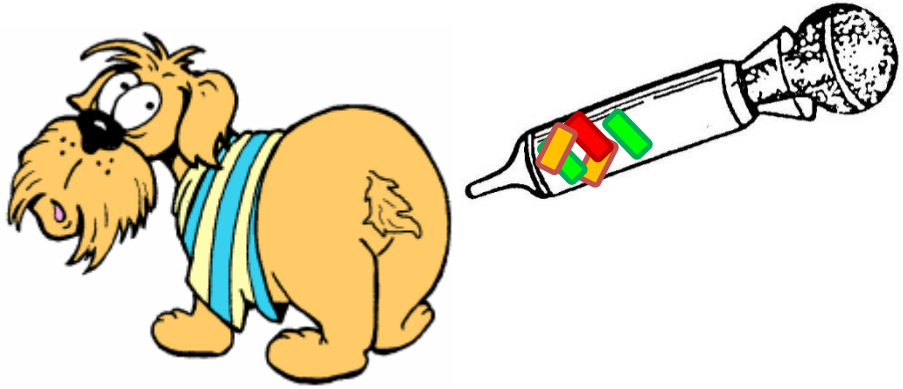
Vandeputte et al. (2017) *FEMS Microbiology Reviews* 41:S154–S167

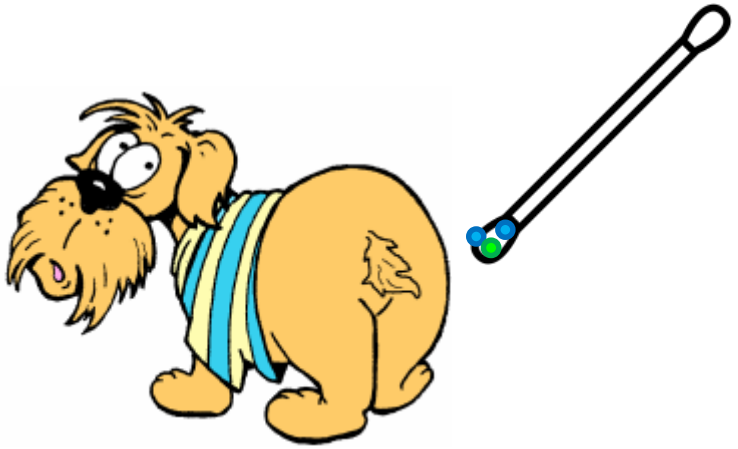
16S rRNA high throughput sequencing: sample collection

- Whichever storage method is chosen, it is crucial that samples are all treated the same.
- Buffers added to preserve specimens may contain a “microbiota composition” of their own. This will result in additional background “profiles” added to specimens.
- *In case specimens are for example stored without the buffer, whilst controls are stored with, false difference will be detected between the two groups under study.*

16S rRNA high throughput sequencing: sample collection

- Sample collection also need to be performed in the exact same manner for all samples under study, for example:
  - If stool samples are collected via aspiration, microbiota background may be introduced to the stool specimens via the solution used to aspirate the sample. Caution should therefore be taken when comparing microbiota profiles from aspirated samples and those collected during passing of stool.
  - If, for example, a swab is used to collect the specimen, the same supplier and product should be used throughout the project.





16S rRNA high throughput sequencing: study design and sample collection

In summary:

- Design your study carefully, by considering things such as participants selection criteria, sample collection and transportation procedures, as well as sample storage and processing.
- Large longitudinal cohort studies (with suitable controls) compared to cross-sectional studies may provide a better insight into our microbial communities when studying their associations with health and disease.
- Collect as much environmental, clinical and demographic data to correct for confounders or to study mediation effects between our bacterial communities and for example disease states.
- Be aware that different sampling methods, sampling buffers, and storage procedures may impact on bacterial profiles obtained from high throughput 16S rRNA gene sequencing.