

Key concepts & alpha diversity

Demo

Microbiome data exploration

Lecture

Key concepts in microbiome data science

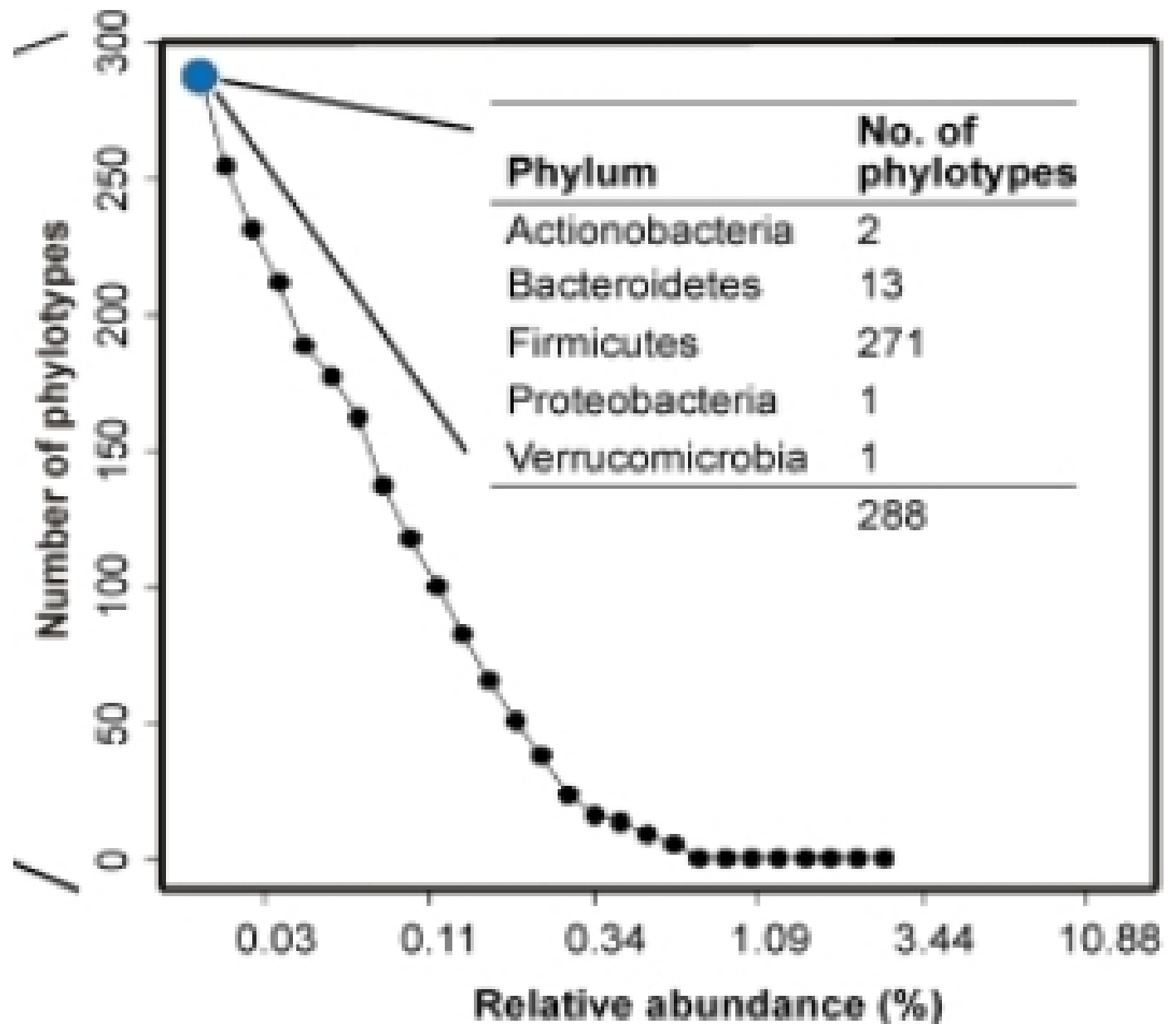
Practical

Alpha diversity: estimation, analysis, and visualization

Gut microbiome species distribution

B

Jalanka-Tuovinen et al. PloS One 2011



Richness:

- Number of types
- Estimates of true richness based on finite sample sizes (Howard Sanders 1968); see e.g. Chao1

Evenness:

distribution of sizes (even or uneven?)

Diversity:

richness & evenness

Dominance

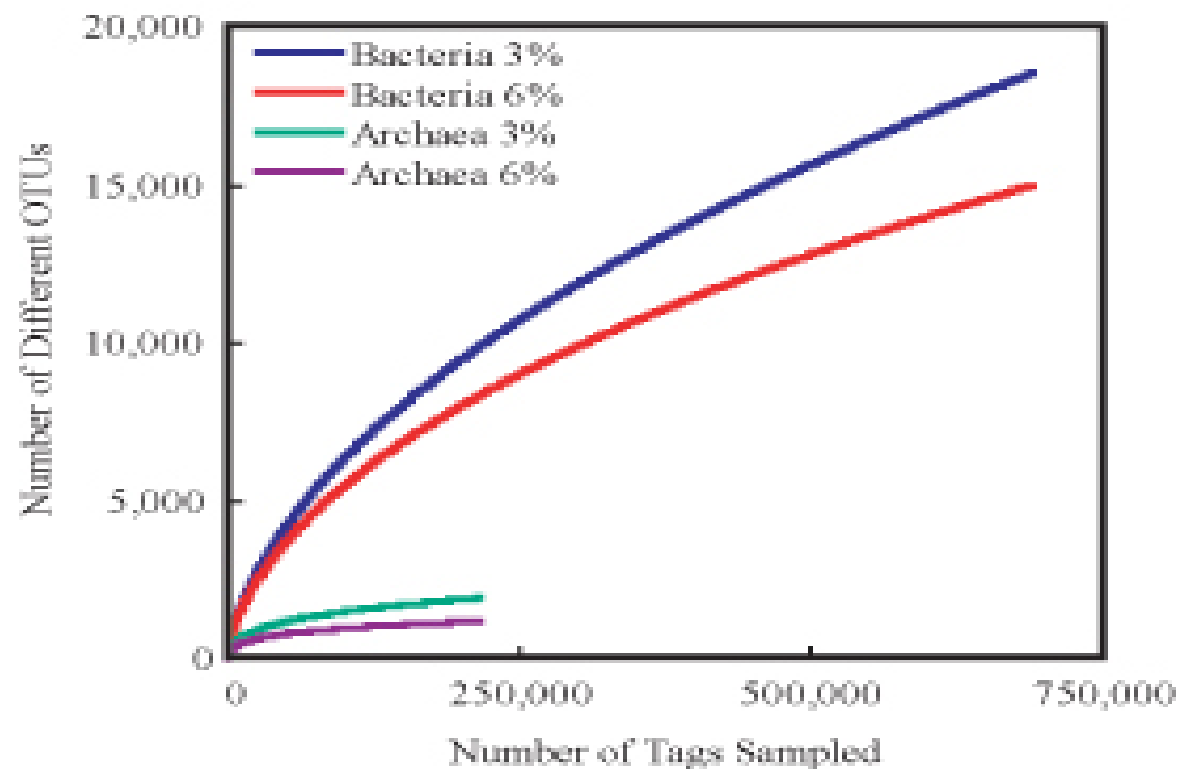
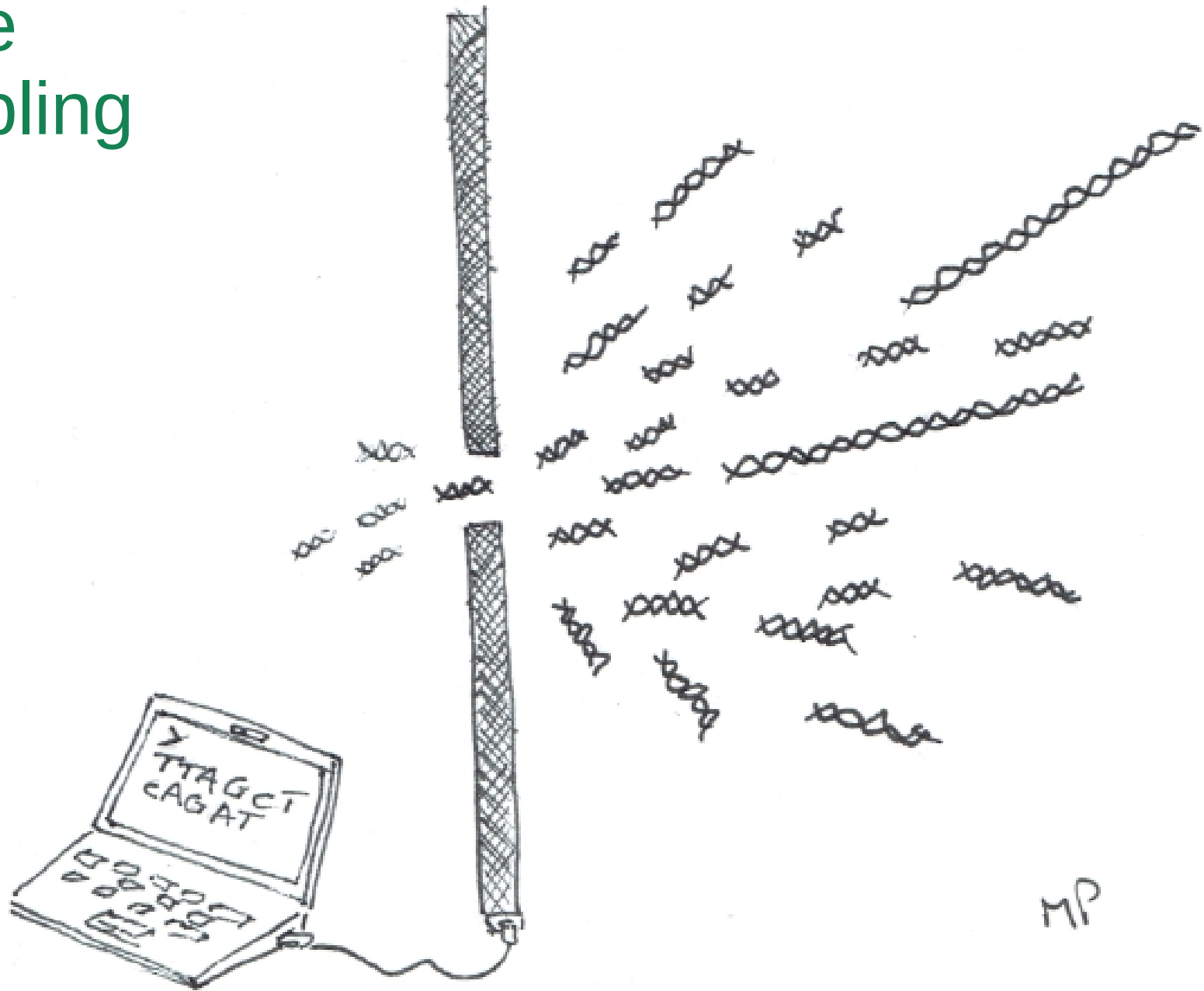
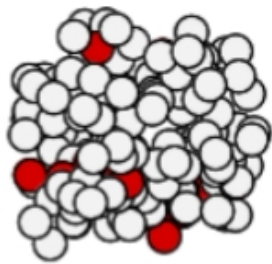


Figure 2

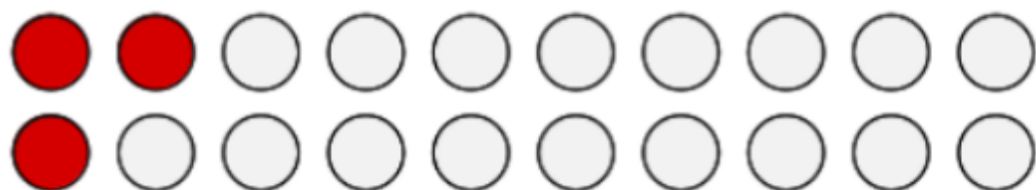
Finite sampling



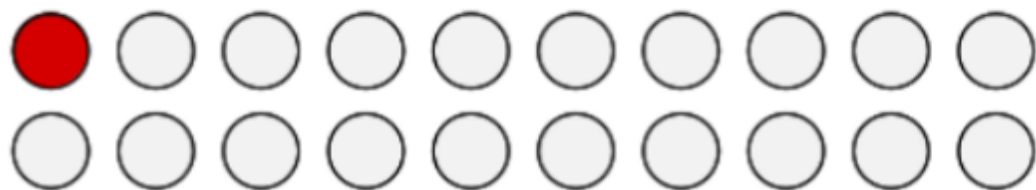
The Poisson distribution



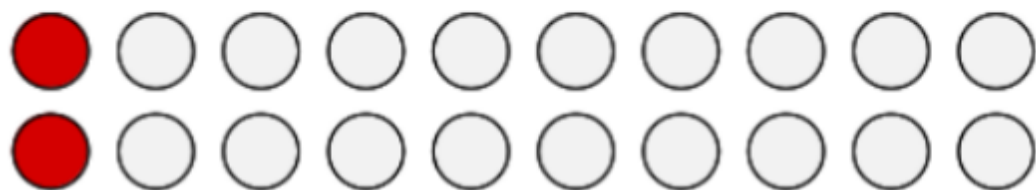
- This bag contains very many small balls, 10% of which are red.
- Several experimenters are tasked with determining the percentage of red balls.
- Each of them is permitted to draw 20 balls out of the bag, without looking.



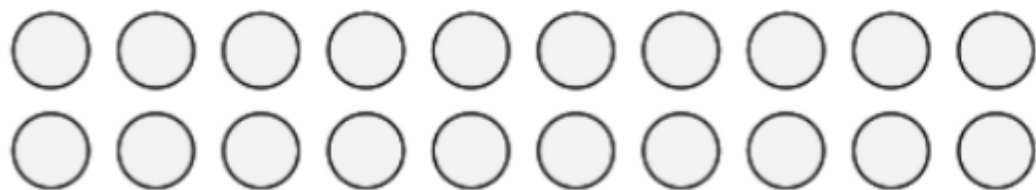
$$3 / 20 = 15\%$$



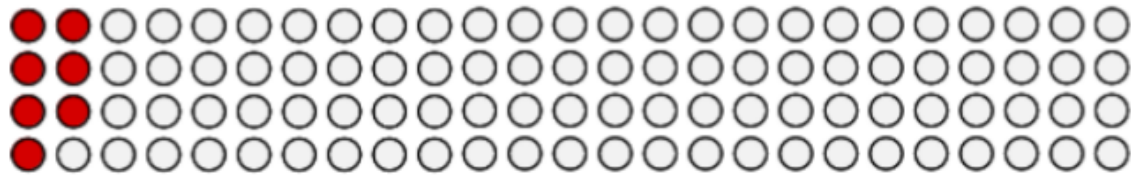
$$1 / 20 = 5\%$$



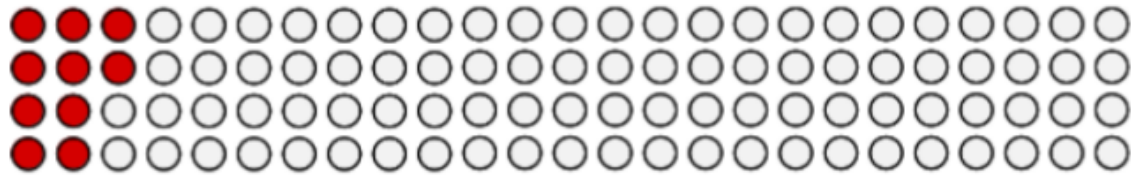
$$2 / 20 = 10\%$$



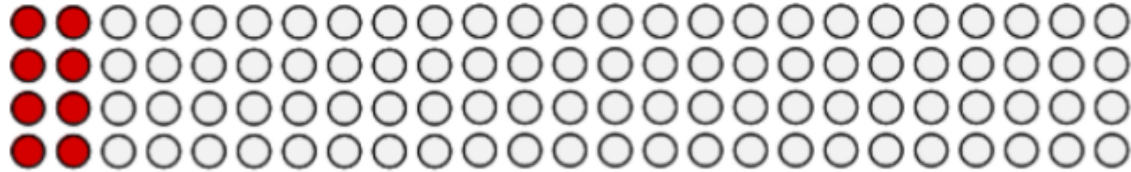
$$0 / 20 = 0\%$$



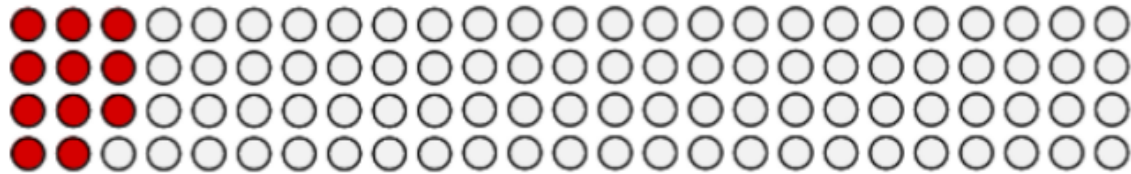
$$7 / 100 = 7\%$$



$$10 / 100 = 10\%$$



$$8 / 100 = 8\%$$

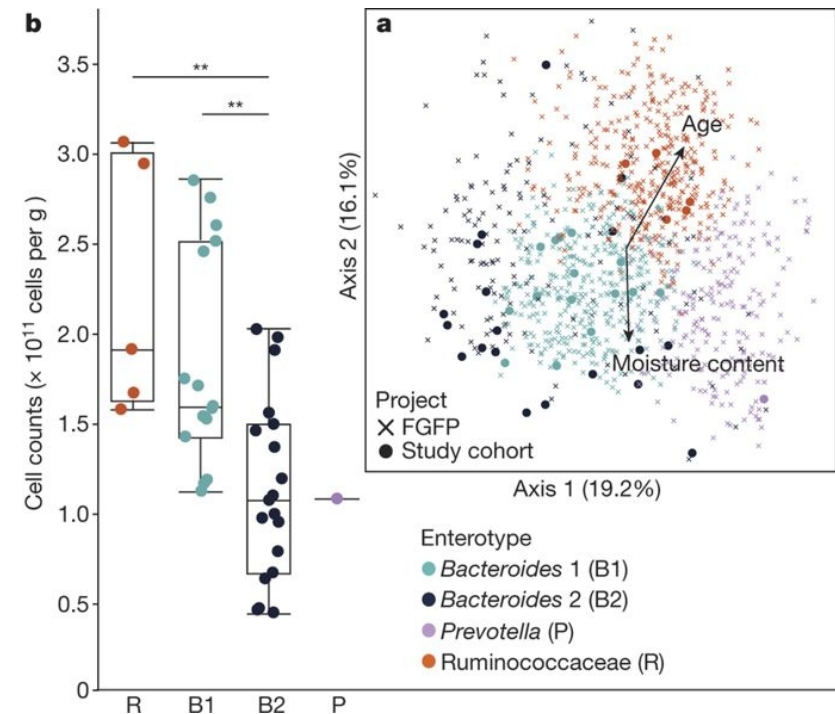
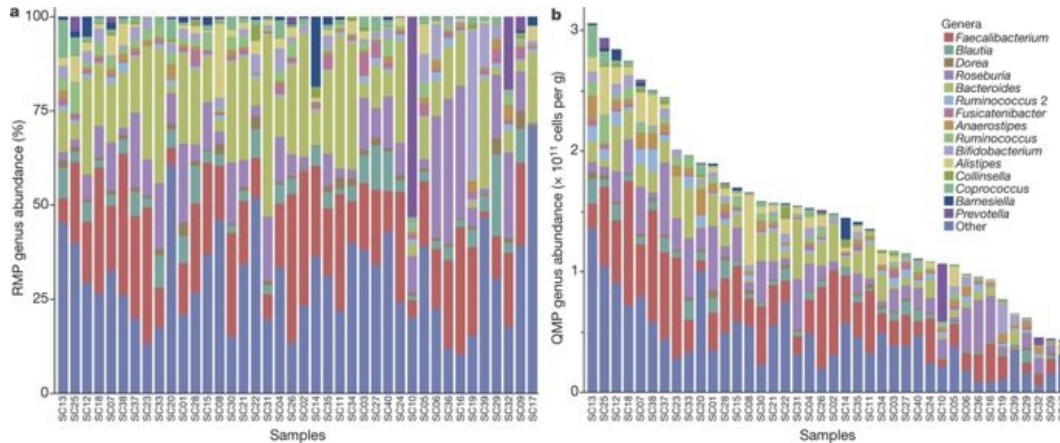


$$11 / 100 = 11\%$$

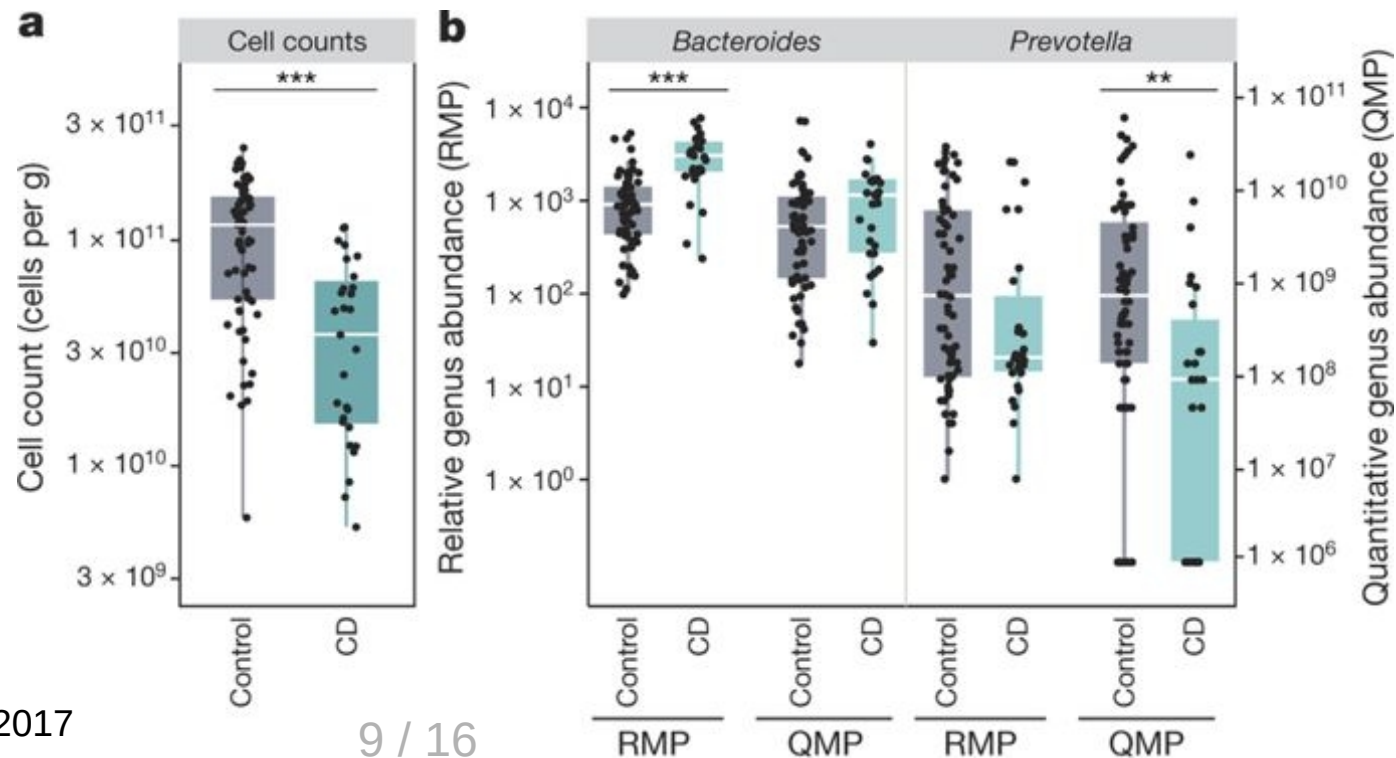
Poisson distribution: Counting uncertainty

expected number of red balls	standard deviation of number of red balls	relative error in estimate for the fraction of red balls
10	$\sqrt{10} = 3$	$1 / \sqrt{10} = 31.6\%$
100	$\sqrt{100} = 10$	$1 / \sqrt{100} = 10.0\%$
1,000	$\sqrt{1,000} = 32$	$1 / \sqrt{1000} = 3.2\%$
10,000	$\sqrt{10,000} = 100$	$1 / \sqrt{10000} = 1.0\%$

Relative versus absolute abundance: quantitative microbiome profiling



RMP vs. QMP:
drastic effect
on conclusions!




Rarefaction

If sample A has been sampled deeper than sample B, we the counts can be expected to be higher.

Solution: Rarefy to even sampling depth

Limitation: Abundant taxa may distort the ratios.

Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible

Paul J. McMurdie, Susan Holmes 

Shannon diversity

$$H' = - \sum_{i=1}^S p_i \ln p_i$$

True richness

$$\exp(H)$$

Inverse Simpson

$$DI = \frac{N(N-1)}{\sum n(n-1)}$$

KEY 

N = Total number of individuals collected

n_i = Number of individuals of a species

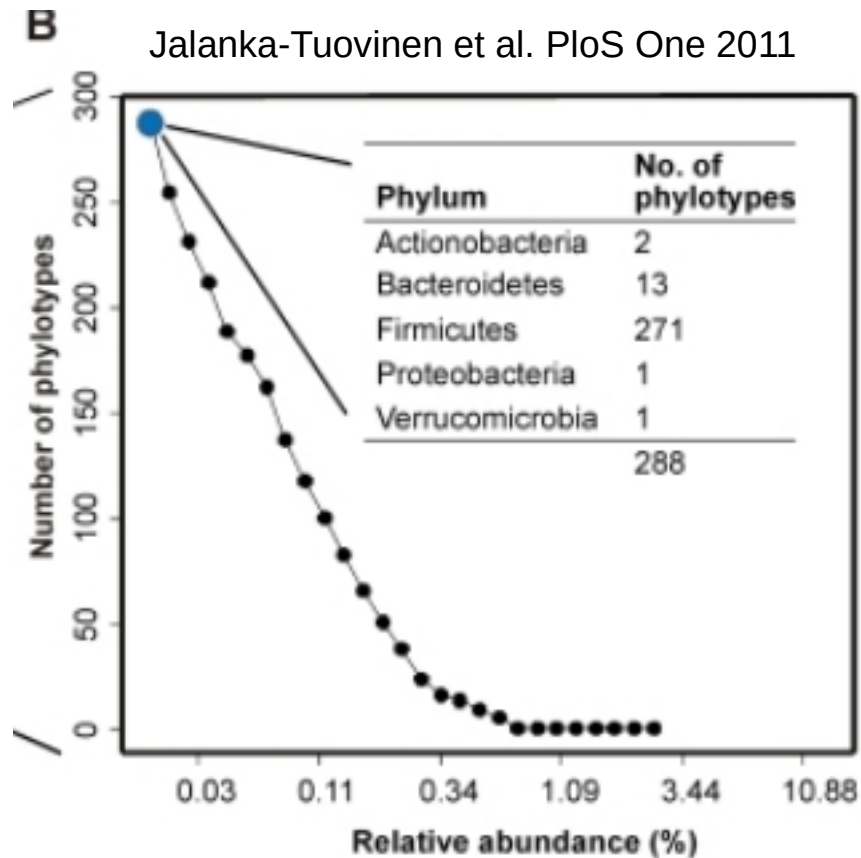
DI = Simpson Diversity Index

How likely it is to pick two members of the same species at random?

(Alpha) Diversity

True diversity, or the effective number of types, refers to the number of equally abundant types needed for the average proportional abundance of the types to equal that observed in the dataset of interest.

$${}^qD = \left(\sum_{i=1}^R p_i^q \right)^{1/(1-q)}$$



R: richness (number of distinct types)
 p_i : proportion of type i

Order of diversity:

$q = 0$: Species Richness

$q = 1$: Shannon diversity

$q = 2$: (Inverse) Simpson diversity

$q \neq 1$: Renyi entropy

Evenness (Equitability)

$$H / \ln(S)$$

H: Shannon diversity

S: Species richness

Alpha & beta diversity

alpha(x)

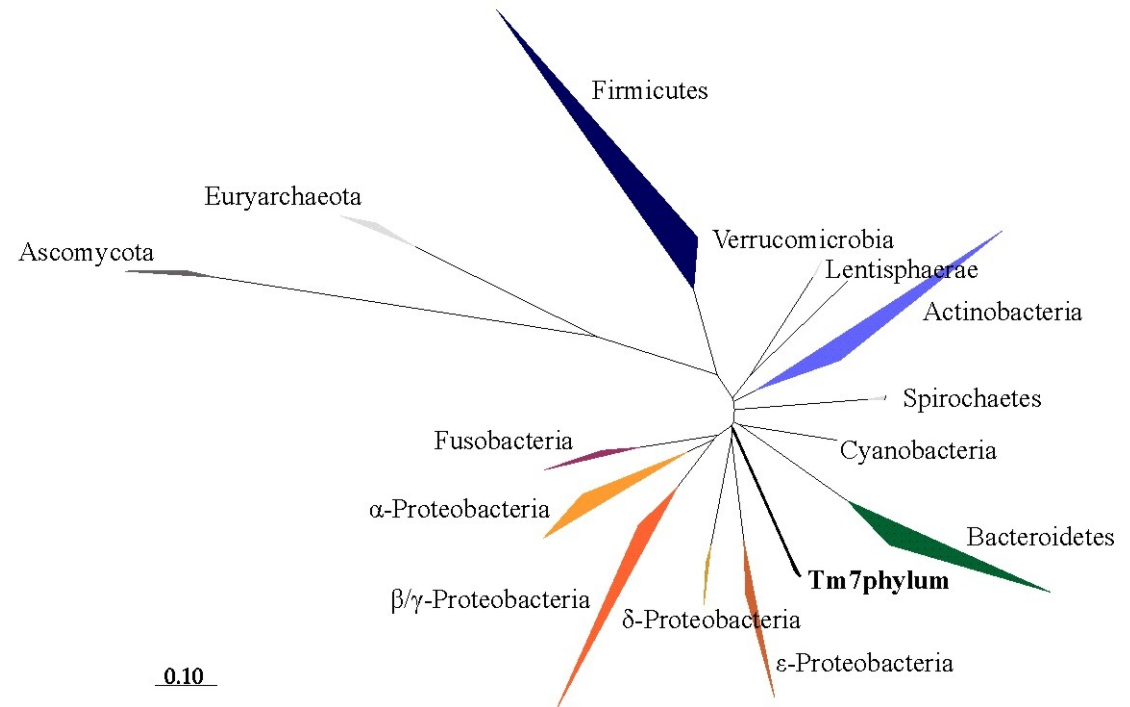
diversity(x)

evenness(x)

dominance(x)

rarity(x)

readcount(x)



$$\text{clr}(\mathbf{x}) = \left[\ln \frac{x_i}{g(\mathbf{x})}; \dots; \ln \frac{x_D}{g(\mathbf{x})} \right]$$

$$\text{alr}(\mathbf{x}) = \left[\ln \frac{x_i}{x_D}; \dots; \ln \frac{x_{D-1}}{x_D} \right]$$