# KMC Bioinformatics Workshop

## Alpha diversity

Kiel, September 25, 2024 – Ana Schaan

IKMB
Institute of Clinical Molecular Biology

# Alpha diversity in Microbiome Analysis

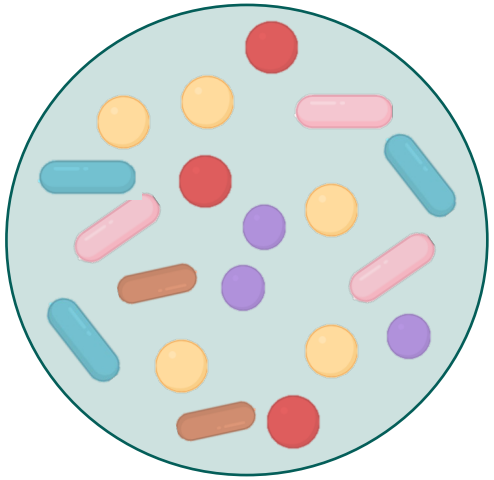
- Part 1: Understanding the concepts
  - → Why Alpha diversity matters
  - → How we measure it (diversity indexes)
  - → Rarefaction
- Part 2: Hands-on Practice
  - → Quick intro to R and to the JupyterHub
  - → Calculating and visualizing Alpha diversity with real GMbC data

# Alpha diversity: Within sample diversity

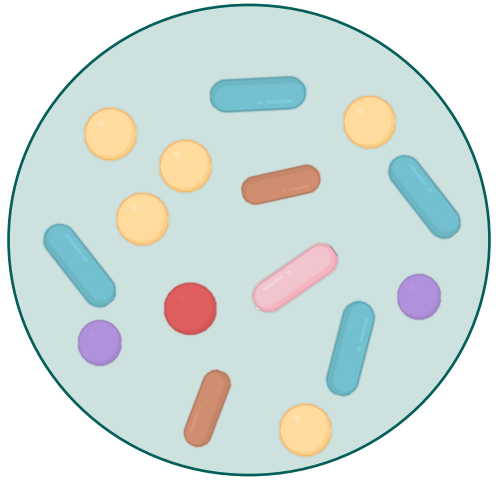
- Definition
  - → Local-scale diversity
  - → Describes **species richness** (number of species) and **evenness** (how evenly they are distributed) within an ecological unit, e.g. the human gut, a pond.
- Importance
  - → Understanding ecosystem functionality and resilience
  - → Captures how diverse a microbial population is within a given environment

# Exploring microbial communities across different environments

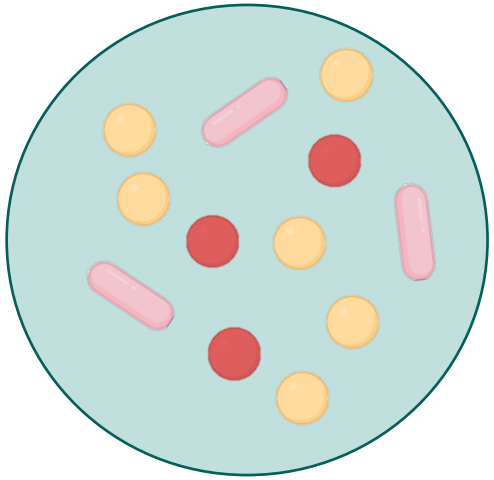
Sample 1

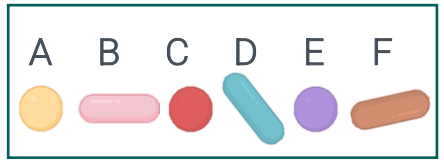


Sample 2



Sample 3





# Key concepts for alpha diversity

## Species richness

Measure of the number of different species present in a community.

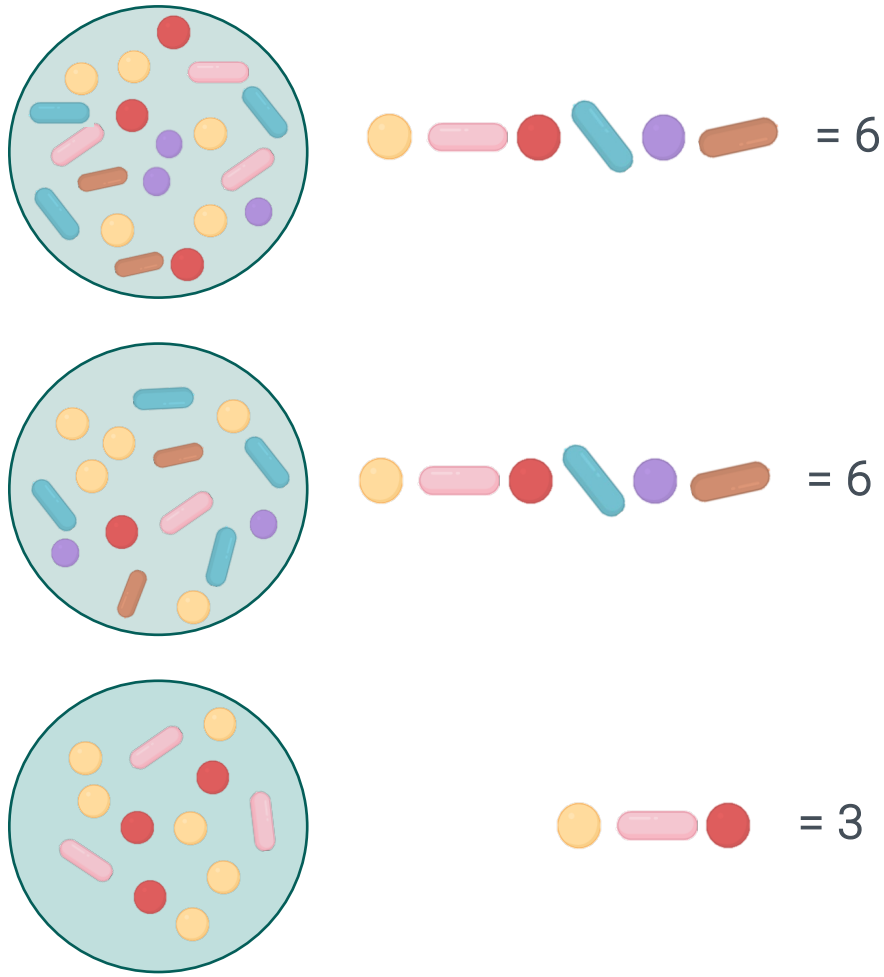
The simplest wat to describe biodiversity: focuses on counts.

# Key concepts for alpha diversity

## Species richness

Measure of the number of different species present in a community.

The simplest wat to describe biodiversity: focuses on counts.



# Key concepts for alpha diversity

## Evenness

Measures how similar the abundances of different species are.

**High evenness:** species are similarly distributed.

**Low evenness:** One of a few species dominate the community.

# Key concepts for alpha diversity

## Evenness

Measures how similar the abundances of different species are.

**High evenness:** species are similarly distributed.

**Low evenness:** One of a few species dominate the community.

| | | | | | |
|---|---|---|---|---|---|
| 6 | 3 | 3 | 3 | 3 | 2 |
| 5 | 1 | 1 | 4 | 2 | 2 |
| 6 | 3 | 3 | | | |

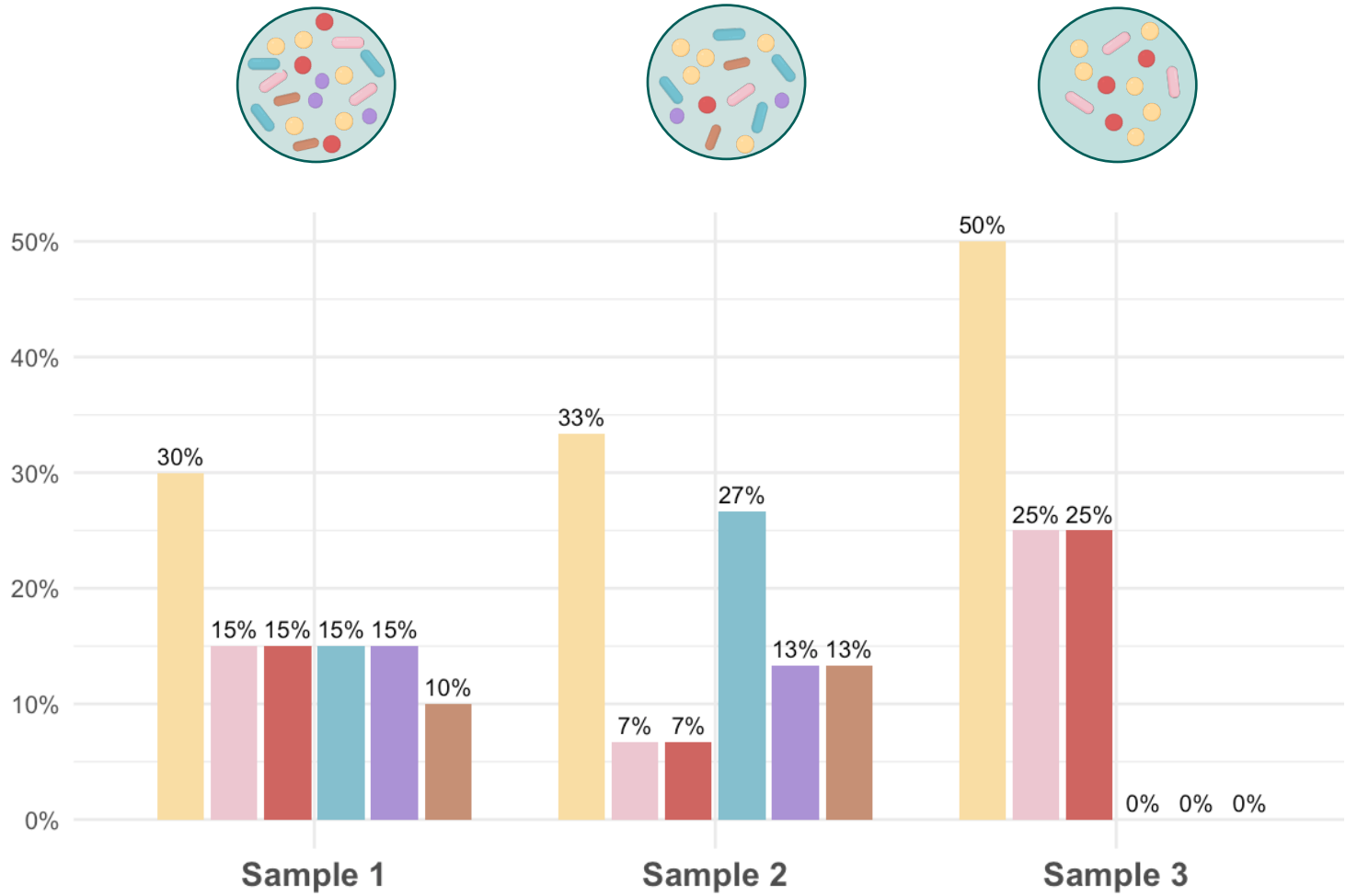
# Key concepts for alpha diversity

## Evenness

Measures how similar the abundances of different species are.

**High evenness:** species are similarly distributed.

**Low evenness:** One of a few species dominate the community.



# Key concepts for alpha diversity

## Singletons and doubletons

**Singletons:** A species that is represented only once in a community or sample. Provide insight into potential undetected species!

**Doubleton:** A species that is represented exactly two times in a community or sample. Also informs about rarity of species in a community.

Richness estimators such as **Chao1** take this into consideration.

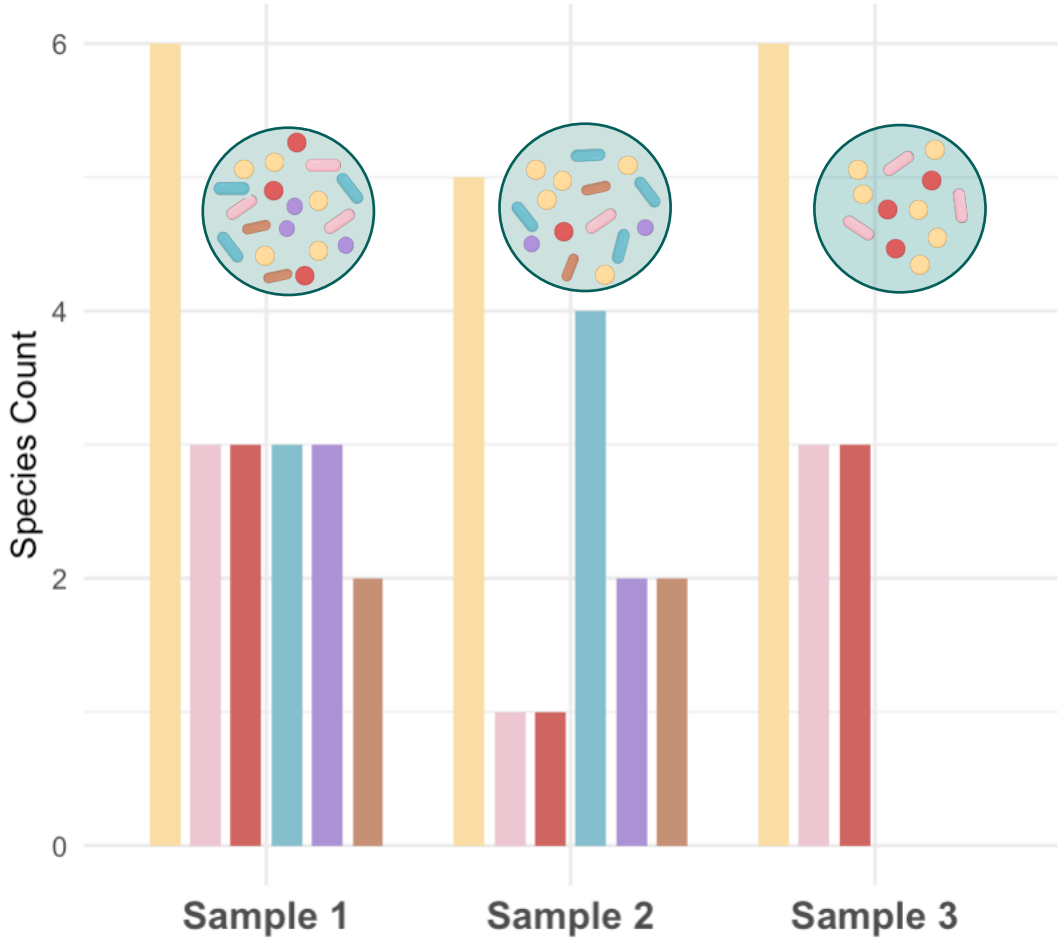
# Key concepts for alpha diversity

## Singletons and doubletons

**Singletons:** A species that is represented only once in a community or sample. Provide insight into potential undetected species!

**Doubleton:** A species that is represented exactly two times in a community or sample. Also informs about rarity of species in a community.

Richness estimators such as **Chao1** take this into consideration.



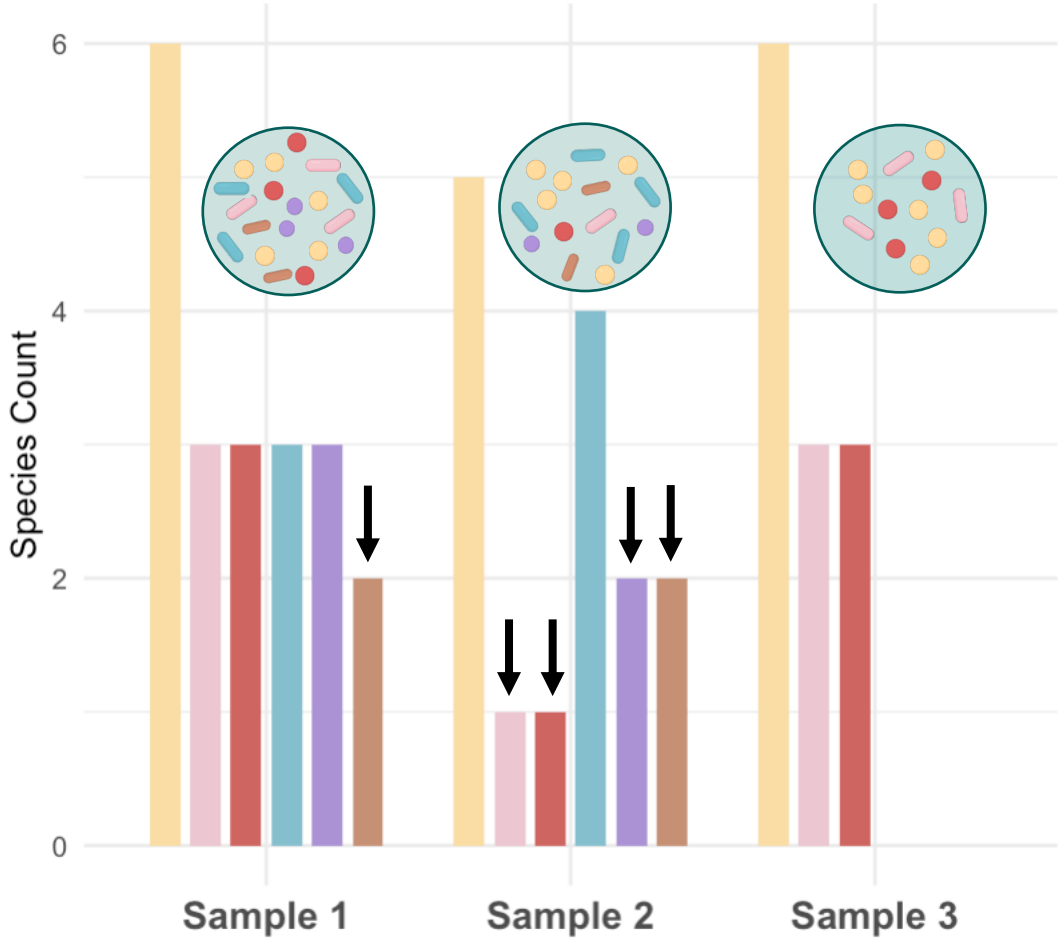
# Key concepts for alpha diversity

## Singletons and doubletons

**Singletons:** A species that is represented only once in a community or sample. Provide insight into potential undetected species!

**Doubleton:** A species that is represented exactly two times in a community or sample. Also informs about rarity of species in a community.

Richness estimators such as **Chao1** take this into consideration.



# Key metrics used to calculate alpha diversity

## Observed Number of Species

Total count of distinct species in sample.

$S = Number\ of\ unique\ species\ observed$

## Shannon Diversity Index

Measures richness and evenness.

$$H' = -\sum(p\_i * ln(p\_i))$$

$p_i$ is the proportion of species $i$

## Simpson's Diversity Index

Measures the probability that two randomly selected individuals belong to the same species.

$$D = 1 - \sum pi^2$$

$p_i$ is the proportion of species $i$

## Chao1 Richness Estimator

Estimator of richness that accounts for rare species, including those that are absent from the sample.

$$S = S\_obs + (f_1^2) / (2 * f_2)$$

$S_{obs}$ is the species count,
$f_1$ is the number of singletons,
$f_2$ is the number of doubletons

# Key metrics used to calculate alpha diversity

## Observed Number of Species

Total count of distinct species in sample.

$S = Number\ of\ unique\ species\ observed$

## Shannon Diversity Index

Measures richness and evenness.

$$H' = -\sum(pi * ln(pi))$$

$p_i$ is the proportion of species *i*

## Simpson's Diversity Index

Measures the probability that two randomly selected individuals belong to the same species.

$$D = 1 - \sum pi^2$$

$p_i$ is the proportion of species *i*

## Chao1 Richness Estimator

Estimator of richness that accounts for rare species, including those that are absent from the sample.

$$S = S\_obs + (f_1{}^2) / (2 * f_2)$$

$S_{\_obs}$ is the species count,
$f_1$ is the number of singletons,
$f_2$ is the number of doubletons

# Key metrics used to calculate alpha diversity

## Observed Number of Species

Total count of distinct species in sample.

$S = Number\ of\ unique\ species\ observed$

## Shannon Diversity Index

Measures richness and evenness.

$H' = -\sum(p\_i * ln(p\_i))$

$p_i$ is the proportion of species *i*

## Simpson's Diversity Index

Measures the probability that two randomly selected individuals belong to the same species.

$$D = 1 - \sum pi^2$$

$p_i$ is the proportion of species *i*

## Chao1 Richness Estimator

Estimator of richness that accounts for rare species, including those that are absent from the sample.

$S = S\_obs + (f_1{}^2) / (2 * f_2)$

$S_{obs}$ is the species count,
$f_1$ is the number of singletons,
$f_2$ is the number of doubletons

# Key metrics used to calculate alpha diversity

## Observed Number of Species

Total count of distinct species in sample.

$S = Number\ of\ unique\ species\ observed$

## Shannon Diversity Index

Measures richness and evenness.

$$H' = -\sum(p\_i * ln(p\_i))$$

$p_i$ is the proportion of species *i*

## Simpson's Diversity Index

Measures the probability that two randomly selected individuals belong to the same species.

$$D = 1 - \sum pi^2$$

$p_i$ is the proportion of species *i*

## Chao1 Richness Estimator

Estimator of richness that accounts for rare species, including those that are absent from the sample.

$$\mathcal{S} = S\_obs + (f_1^2) / (2 * f_2)$$

$S_{\_obs}$ is the species count,
$f_1$ is the number of singletons,
$f_2$ is the number of doubletons

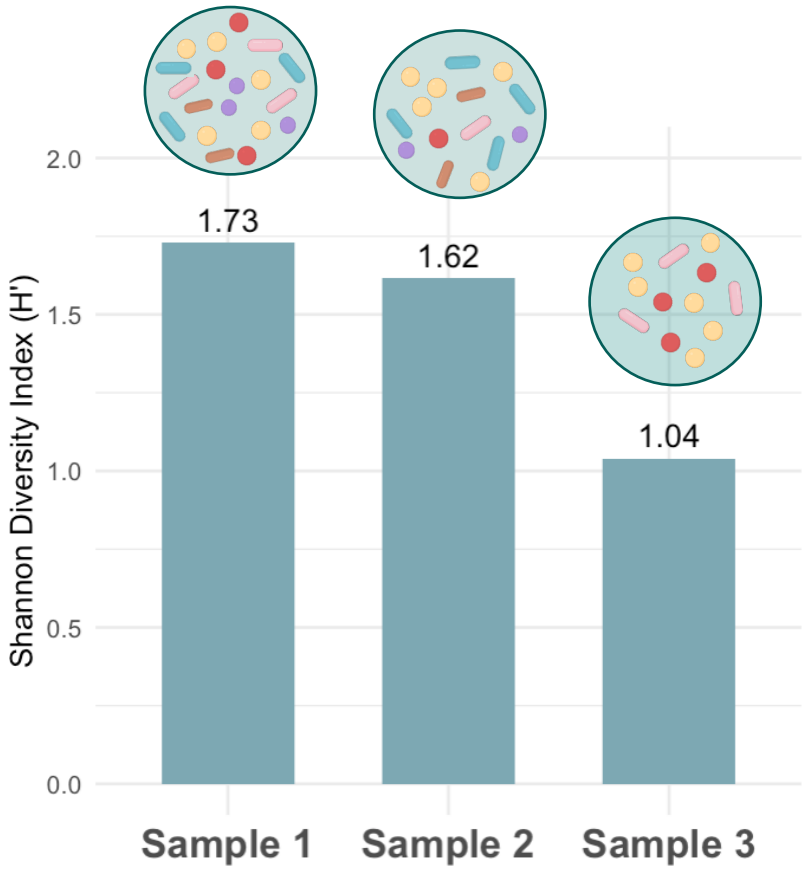
# The Shannon Diversity Index (H')

## Definition:

A measure that accounts for both species richness and evenness of species in a community. It quantifies the uncertainty in predicting the species of a randomly chosen individual.

$$H' = -\sum(pi * ln(pi))$$

$p_i$ is the proportion of species *i*



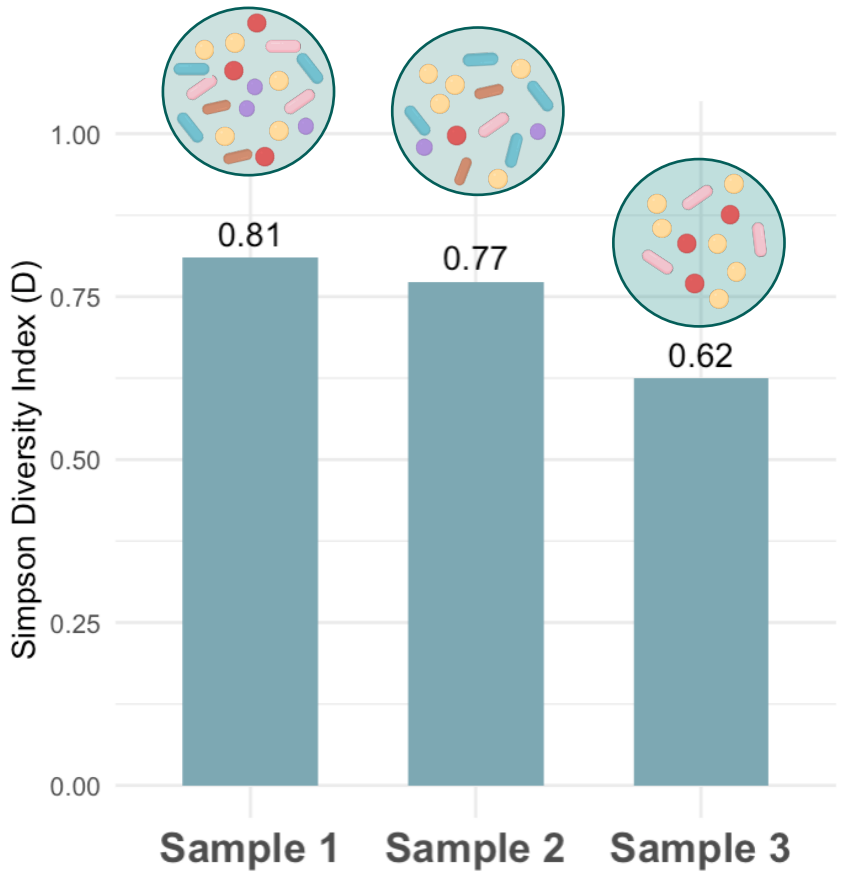
# Simpson’s diversity index

## Definition:

Measures the probability that two individuals randomly selected will belong to the same species. Emphasizes dominance (more weight to common species).

$$D = 1 - \sum pi^2$$

$p_i$ is the proportion of species *i*



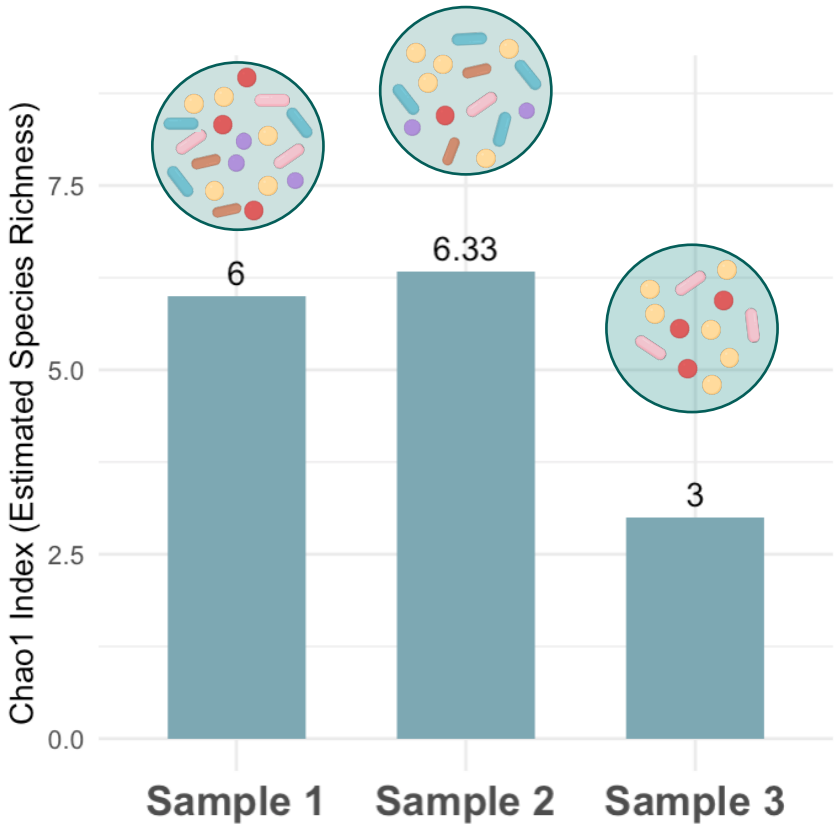
# Chao1 Richness Estimator

## Definition:

Estimate of species richness, accounting for rare species. Considers **singletons** and **doubletons** to estimate the total number of species, including those that may not have been observed.

$$S = S\_obs + (f_1^2) / (2 * f_2)$$

$S_{obs}$ is the species count,
$f_1$ is the number of singletons,
$f_2$ is the number of doubletons



# Other alpha diversity metrics

## Pielou's Evenness (J')

Evenness in species distribution

$$J' = \frac{H'}{ln(S)}$$

## Faith's Phylogenetic Diversity (PD)

Measures branch length of a phylogenetic tree. Incorporates evolutionary relationships!

$$PD = \Sigma(branch\ lengths)$$

## ACE (Abundance-based Coverage Estimator)

Estimates richness by accounting for rare species

$$(\frac{f_{rare}}{C\_ACE})S\_ACE = S\_abund + S\_rare +$$

# Challenges in Alpha Diversity Estimation

- Technical and biological factors can influence the calculation of alpha diversity measurements.

- Common challenges:

  → Differences in sequencing depth

  → Data sparcity

# Impact of Sequencing Depth on Alpha diversity

- Sequencing depth: Total number of reads obtained from a sample.
- Higher sequencing depth leads to a better detection of rare species

- Major issue:

→ Inconsistent sequencing depths across samples can lead to inaccurate or incomparable alpha diversity measurements

# Using rarefaction to balance sequencing depth and diversity

- Rarefaction:
  - → A technique used to standardize the number of reads across samples
  - → Random subsampling of each sample to the same sequencing depth
  - → Prevents that diversity estimates are not inflated by samples with higher depths

# Using rarefaction to balance sequencing depth and diversity

- Rarefaction:
  - → A technique used to standardize the number of reads across samples
  - → Random subsampling of each sample to the same sequencing depth
  - → Prevents that diversity estimates are not inflated by samples with higher depths