

OTU Methods in Numerical Taxonomy

Presenter: PhD Student Quynh Nhu Nguyen - Heidelberg University Hospital, Germany

21.07.2024

CONTENTS

1. Case study
2. Introduction to Numerical Taxonomy
3. Core Techniques of Numerical Taxonomy
4. Summary

1. Case study

Research paper/report

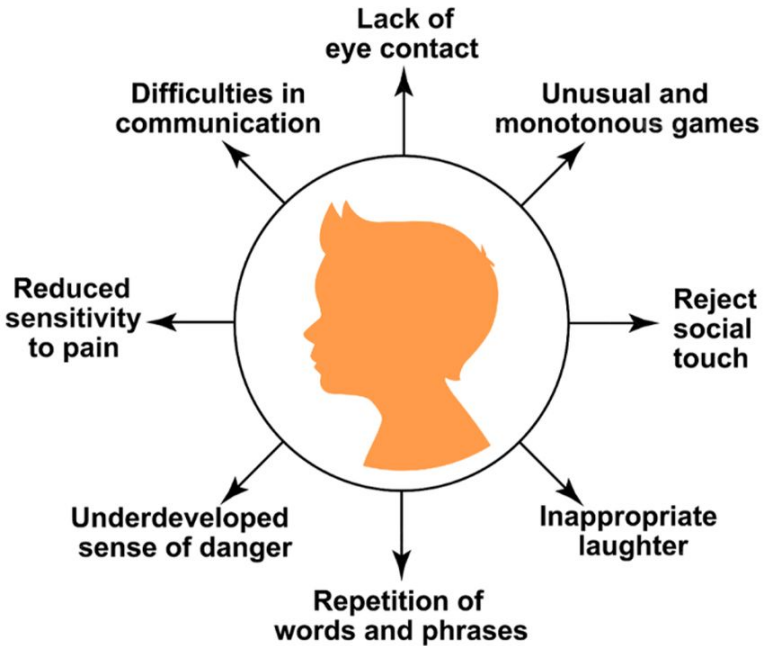
Altered gut microbial profile is associated with abnormal metabolism activity of Autism Spectrum Disorder

Zhou Dan, Xuhua Mao, Qisha Liu, Mengchen Guo, Yaoyao Zhuang, Zhi Liu, Kun Chen, Junyu Chen, Rui Xu, Junming Tang, Lianhong Qin, Bing Gu, Kangjian Liu, Chuan Su, Faming Zhang, Yankai Xia, Zhibin Hu & Xingyin Liu

Pages 1246-1267 | Received 29 Jul 2019, Accepted 11 Mar 2020, Published online: 21 Apr 2020

<https://doi.org/10.1080/19490976.2020.1747329>

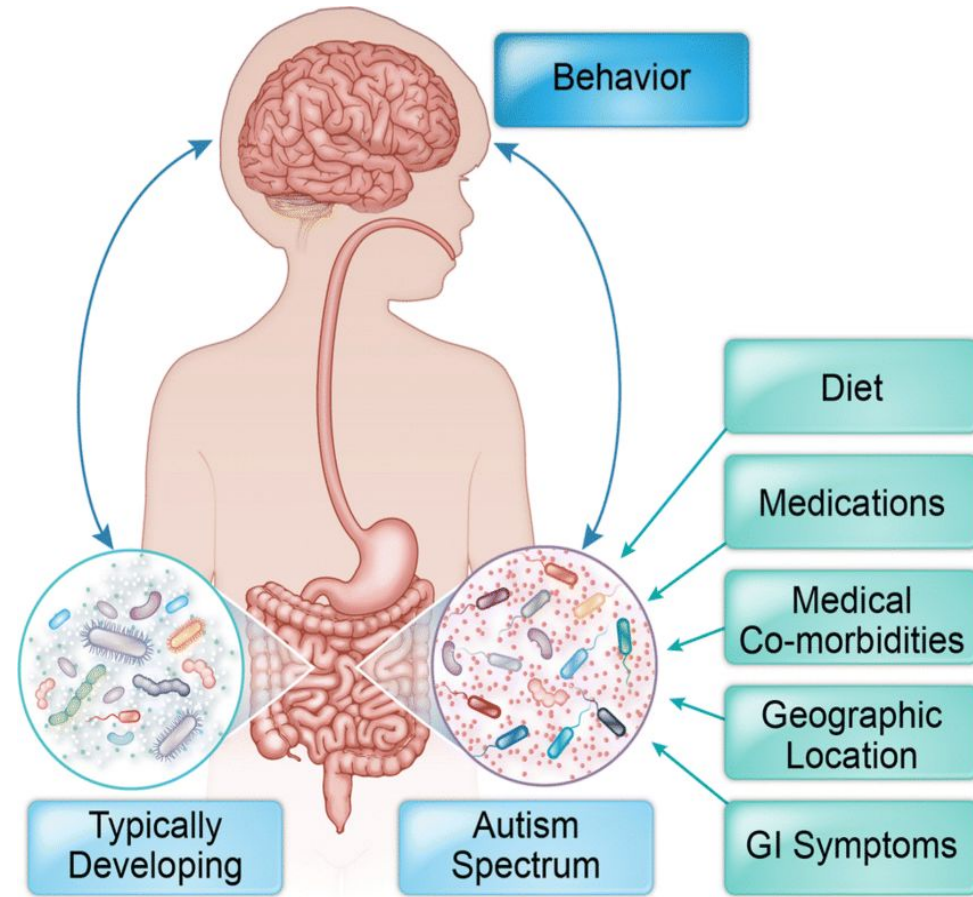
Autism Spectrum Disorder (ASD)



- Severe neurodevelopmental disorder
- Key symptoms: impaired social interaction, stereotyped behaviors, restricted interests
- Increasing prevalence: Affects 1 in 59 children in the U.S. and is a leading mental disability in China

Gut Microbiota and ASD

- GI symptoms often co-occur with ASD symptoms (gaseousness, diarrhea, constipation)
- Gut microbiota can influence gastrointestinal physiology, immune function, and behavior
- The gut-microbiome-brain axis plays a crucial role



Study Objectives

1. **Assess Gut Microbiota Diversity:** Evaluate changes across ages in ASD vs. normal (TD) children.
2. **Investigate Gut Microbiota and Metabolites:** Explore the relationship with fecal metabolites, particularly in children with chronic constipation.
3. **Understand GI Symptoms:** Analyze interactions between ASD and GI symptoms, focusing on chronic constipation.

Methodology

Participants:

- 143 children, ages 2–13 years old (ASD and TD groups)

Techniques Used:

- **16S rRNA Sequencing:** To evaluate gut microbial populations
- **Metagenomics Analysis:** On 30 C-ASD and 30 TD participants
- **Liquid Chromatography-Mass Spectrometry (LC/MS):** For metabolite analysis

Methodology

Table 1 of 1

Table 1. Characteristics of study participants.

Characteristic	TD	ASD	P value
Subjects (n)	143	143	
Male/Female	130/13	127/16	0.5567 (Chi-square test)
Age range, years	1-11	1-13	
2-3(years)	33	31	
4-6(years)	82	79	
7-11(years)	28	32	
12-13(years)	0	1	
Age (mean), years	4.937 ± 0.155	5.189 ± 0.170	0.2751 (T test)
Constipation	0	52	
Diarrhea	0	5	

OTU (Operational Taxonomic Unit)

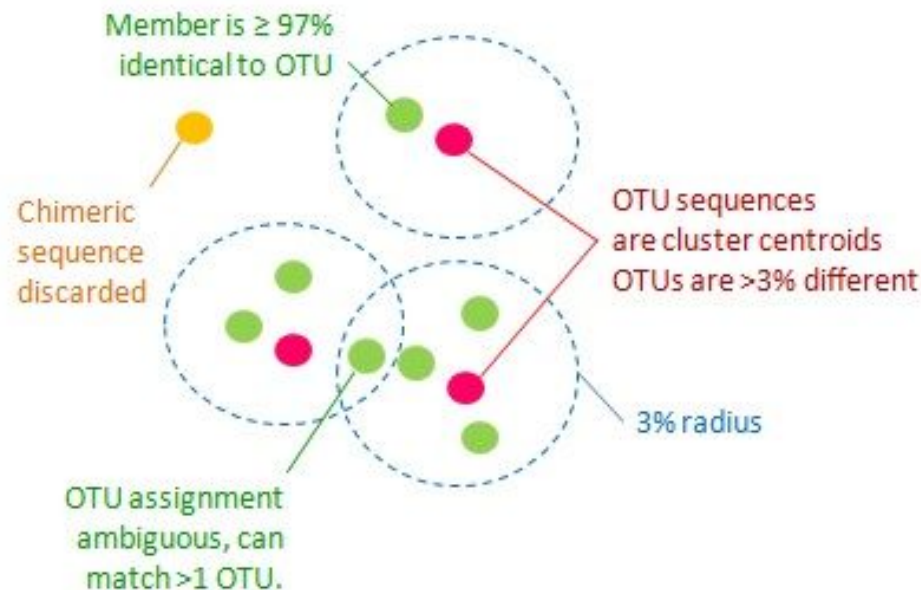
- A method to classify groups of closely related individuals.
- Used to categorize bacteria based on DNA sequence similarity.
- Clustering was performed using Uparse 7.1 (<http://drive5.com/uparse>) with sequences similarity $\geq 97\%$.

Richness

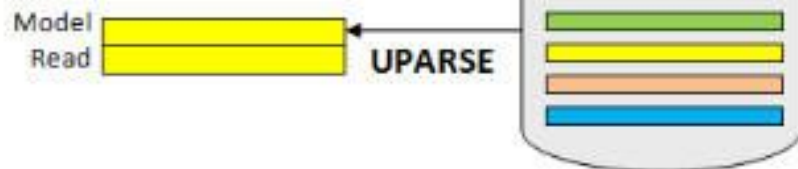
- Refers to the number of different species present in a sample.
- Higher richness means more diverse species.

UPARSE-OTU algorithm

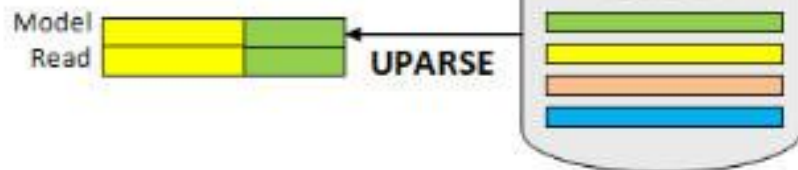
(usearch tool)



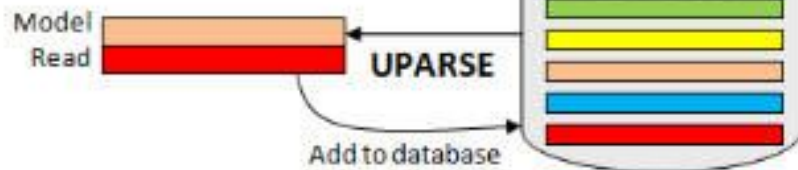
a. Model $<3\%$, assign to OTU.

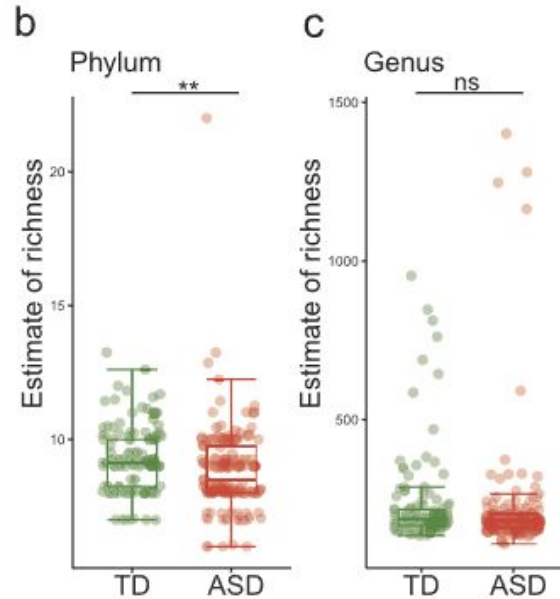
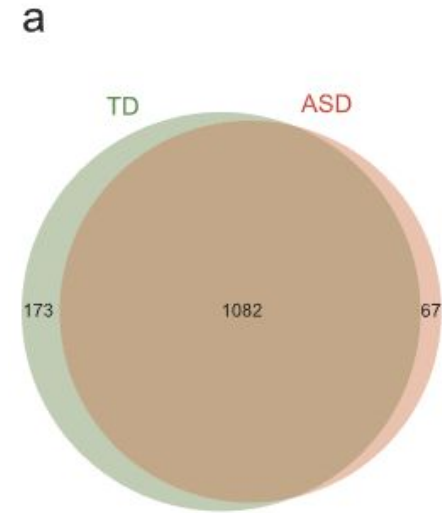


b. Model is chimeric, discard.



c. Model $\geq 3\%$, new OTU.





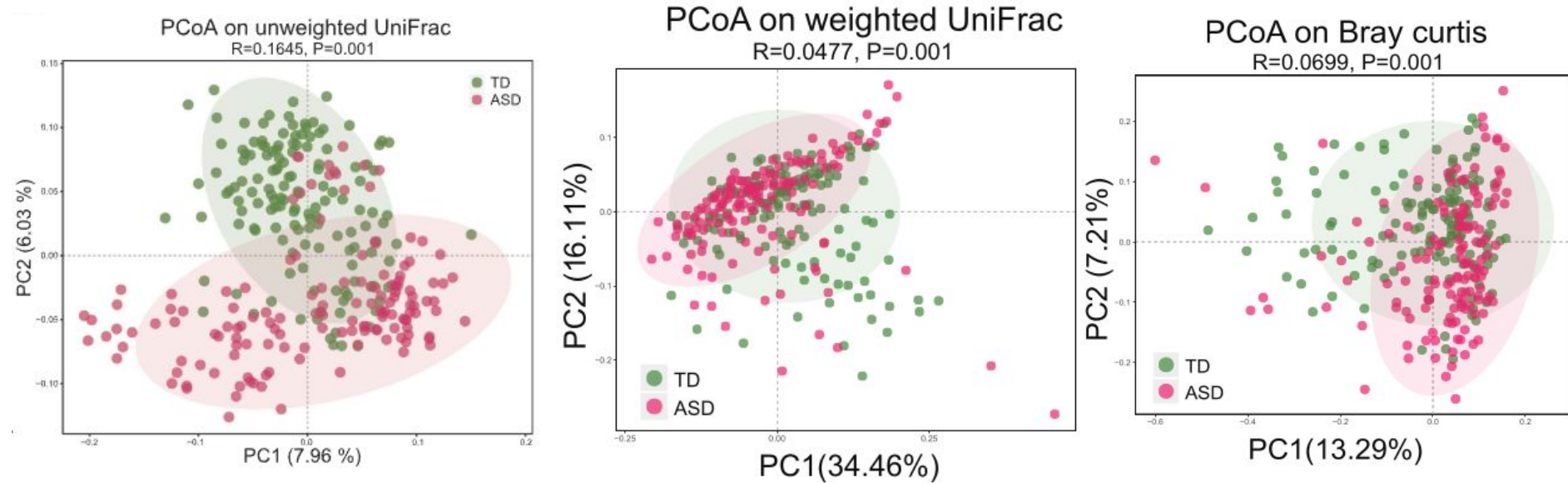
(a) Venn diagram of the observed OTUs in TD and ASD

- 173 unique OTUs in the TD group
- 67 unique OTUs in the ASD group
- 1082 OTUs shared by both groups

(b) (c) Species Richness

- Significantly lower at the phylum level in the ASD group compared to the TD group
- No significant differences at the genus level between the ASD and TD groups

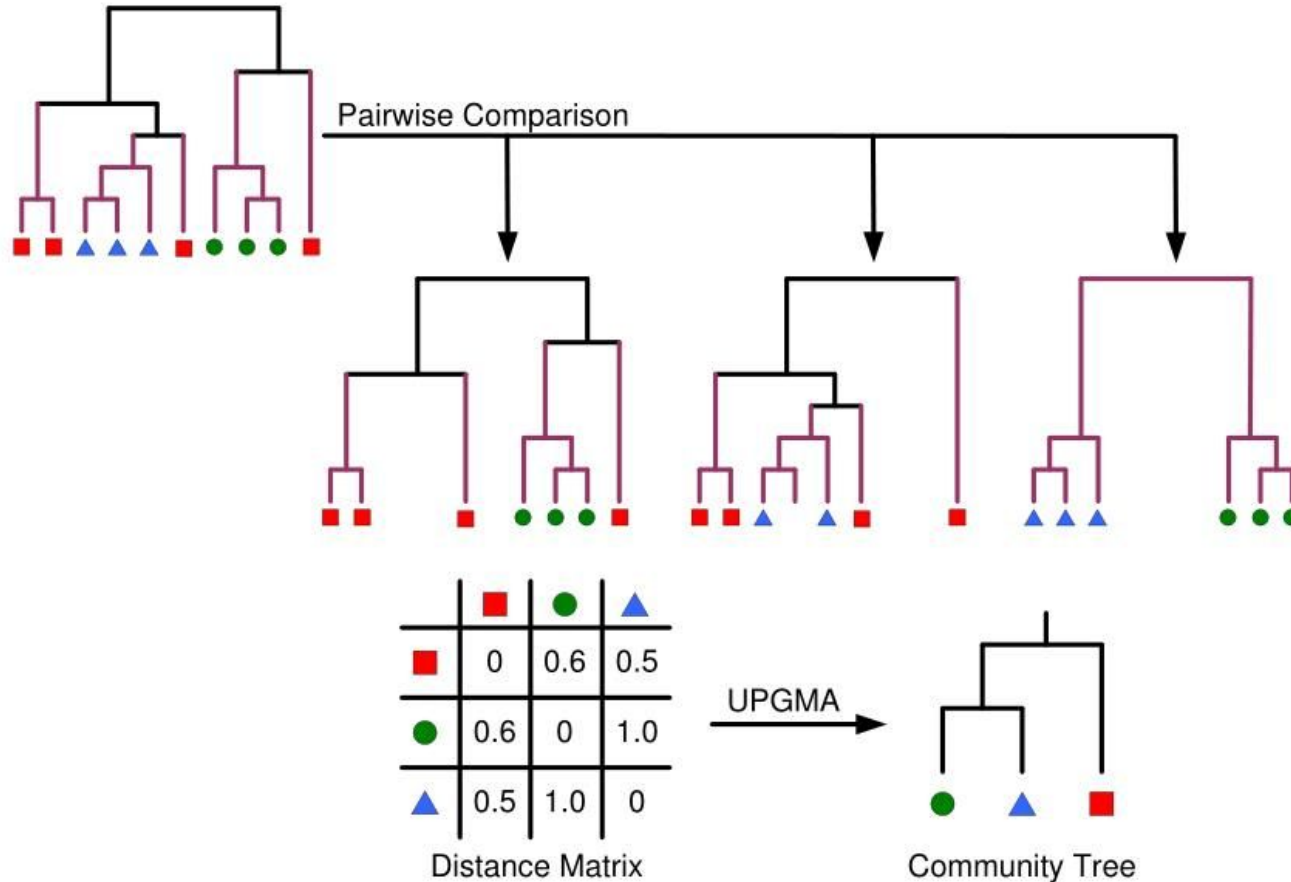
Assess Microbial Community Composition Differences



UniFrac

- **Definition:** A measure of distance between microbial communities that incorporates **phylogenetic information**.
- **Calculation:** The UniFrac distance between two communities (A and B) is the fraction of branches in a phylogenetic tree that lead to members of either community A or B, but not both.
 - **Presence/Absence:** This measure only considers whether an OTU (Operational Taxonomic Unit) is present or absent in the community, ignoring its abundance.
 - **UniFrac Distance:** Focuses on the shared evolutionary history between communities by examining the unique branches in the phylogenetic tree.

UniFrac: Community Distance Matrix



Weighted UniFrac

Definition: An extension of UniFrac that takes into account the abundance of OTUs.

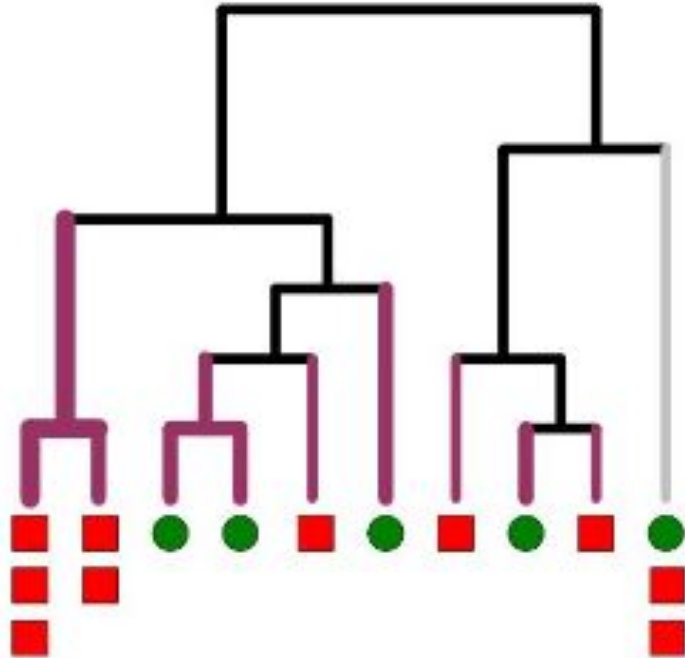
Calculation: Weighted UniFrac between communities A and B is defined as:

Components:

- b_i is the length of the i th branch,
 - A_T is the overall abundance of OTU in community A,
 - A_i is the number of OTU in community
 - Similar terms apply for community B.
- $$wUF(A, B) = \sum_i b_i |A_i/A_T - B_i/B_T|$$

Abundance Consideration: Unlike UniFrac, Weighted UniFrac incorporates how many of each OTU are present, not just whether they are present.

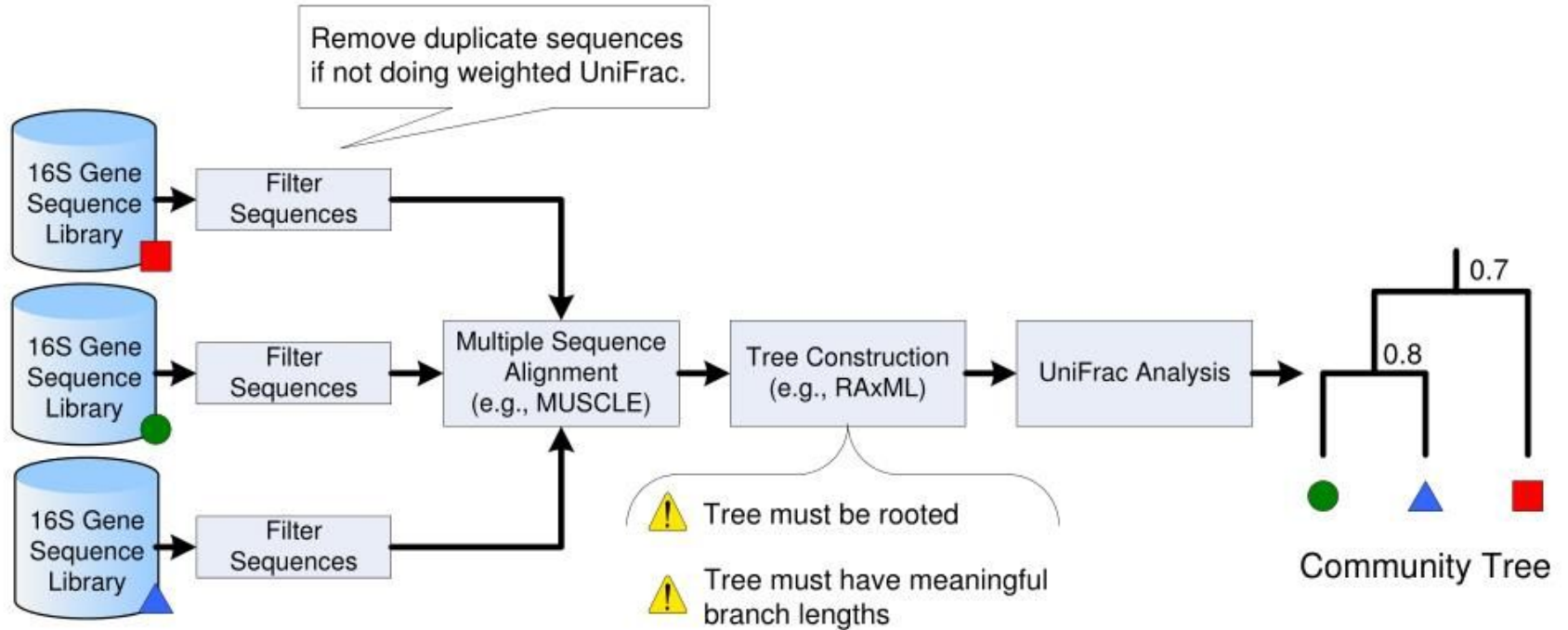
Weighted UniFrac



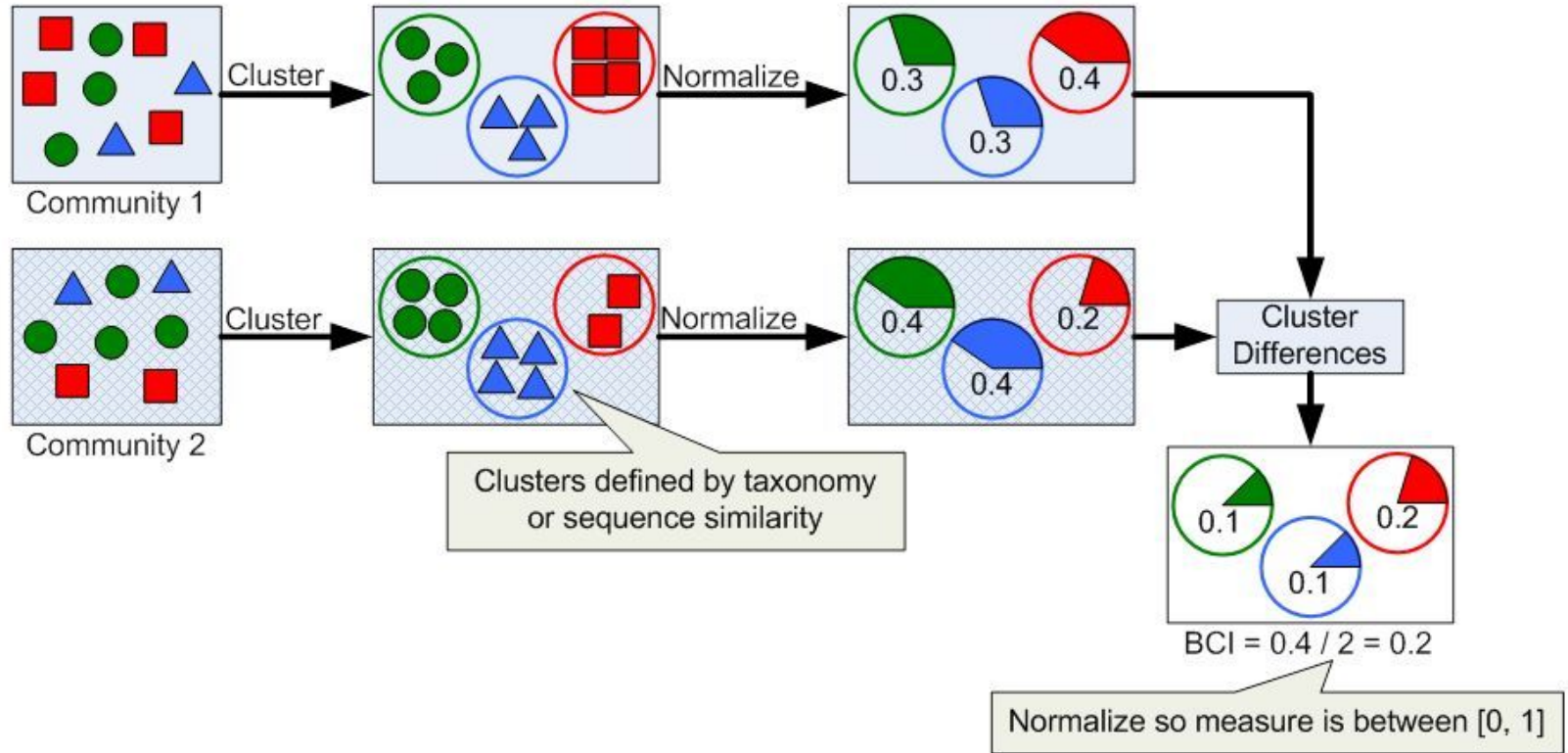
Since there are twice as many ■ sequences, they get half the weight.

What about the relative abundance of sequences (i.e., evenness)?

UniFrac: A practical look



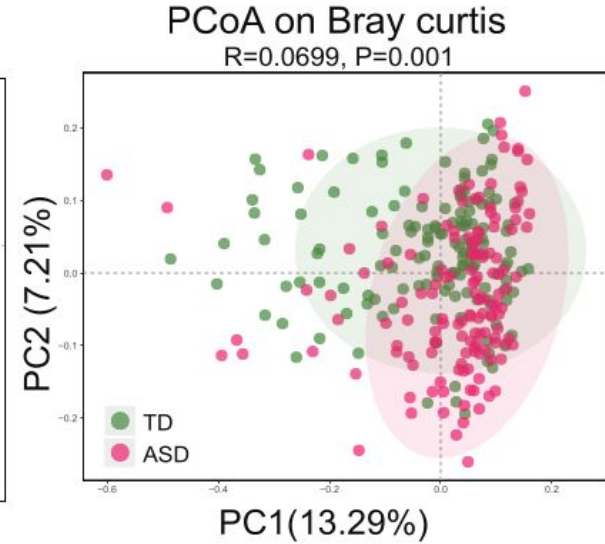
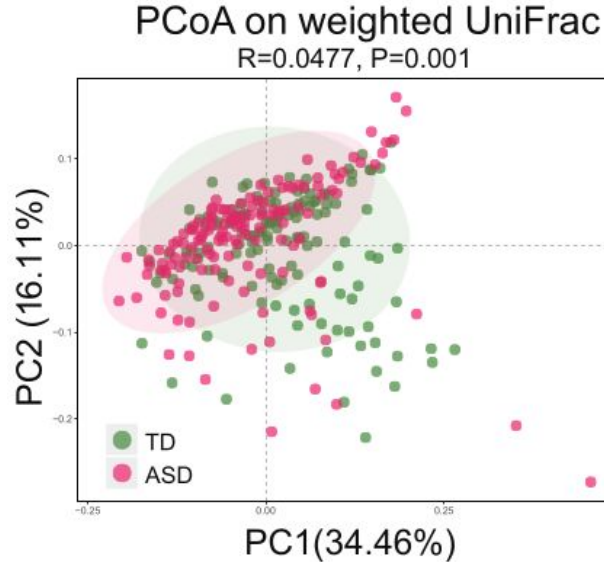
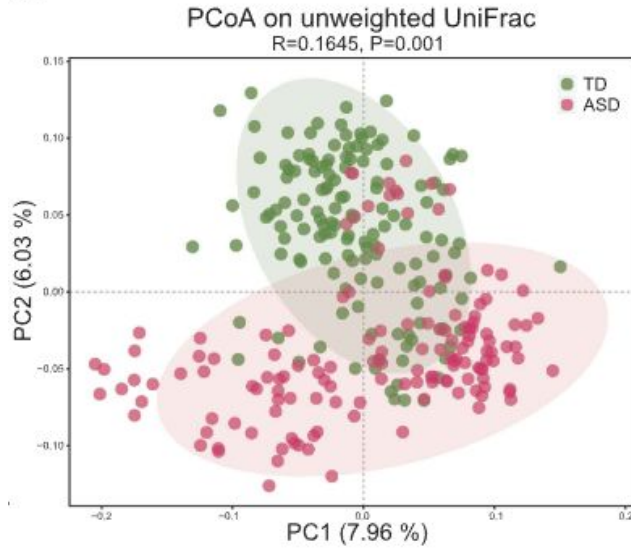
Bray-Curtis Index (BCI): Distance Measure



BCI vs UniFrac

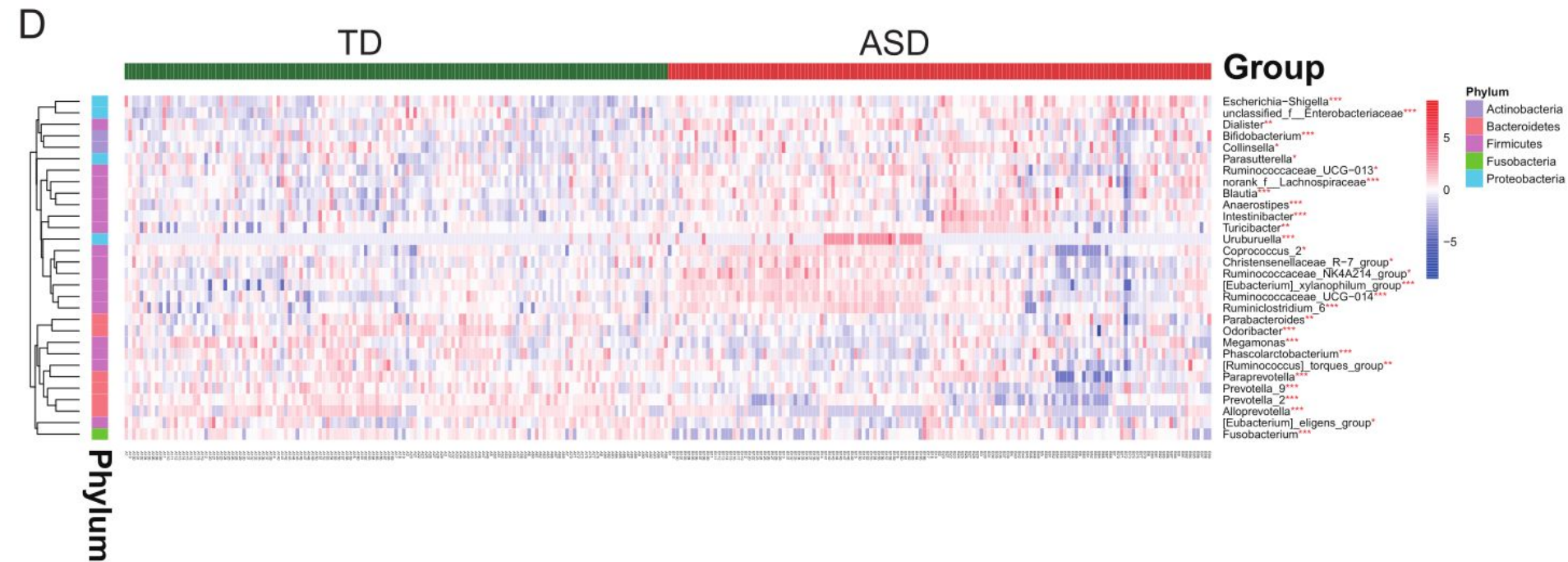
- They are different, equally valid measures
- BCI: taxon-based method which considers richness and evenness
- UniFrac: phylogenetic-based method which considers evolutionary divergence
- Which to prefer depends on the questions you are trying to answer

Assess Microbial Community Composition Differences



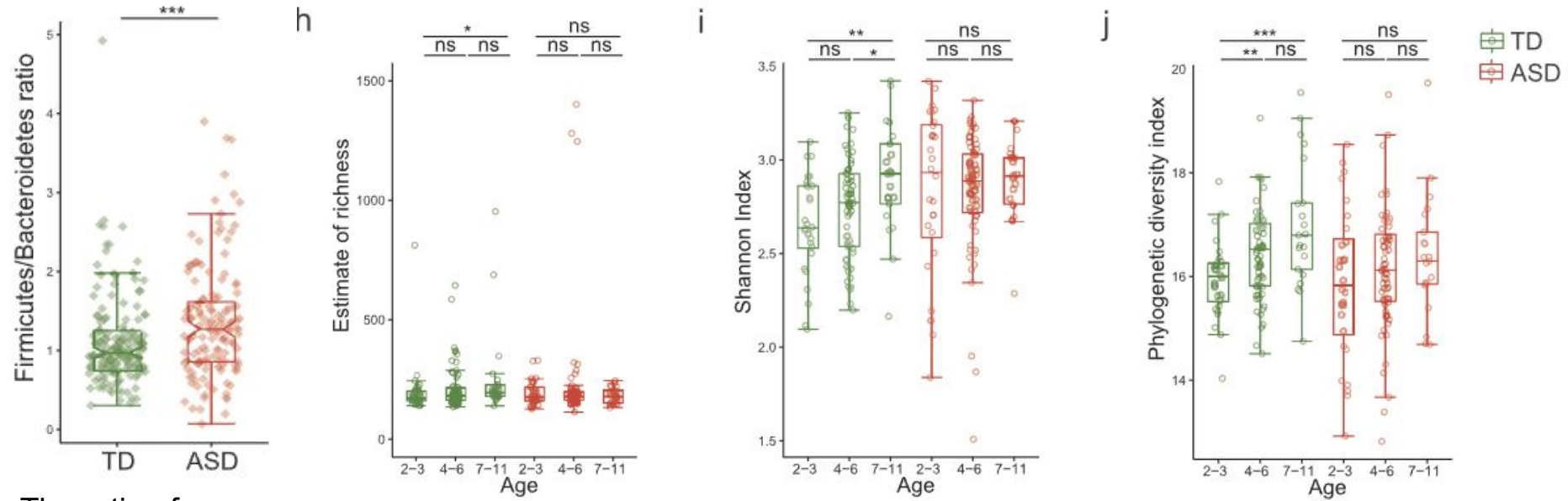
Heterogeneity in ASD Group: The PCoA analysis showed that the microbial communities in the ASD group were more heterogeneous, meaning there was greater variability or diversity in the composition of microbial communities within this group.

Significant Differences: The ASD group's microbial community composition was significantly different from that of the TD group. This means that, overall, the ASD and TD groups have distinct microbial community structures. [DOI: 10.1080/19490976.2020.1747329](https://doi.org/10.1080/19490976.2020.1747329)



Heatmap of the 30 most significant different genera detected in all of samples. Dendrograms for hierarchical cluster analysis grouping genera and sample locations are shown at the left and at the top, respectively.

Alpha diversity indices of genus for TD and ASD according to age from 2 to 11



The ratio of Firmicutes/Bacteroidetes of ASD showed significantly higher than that of TDs.

Alpha diversity

- **Richness of Species**

- In the TD group, species richness increased from ages 2–3 to 7–11
- The ASD group showed no significant change in species richness with age

- **Shannon Index**

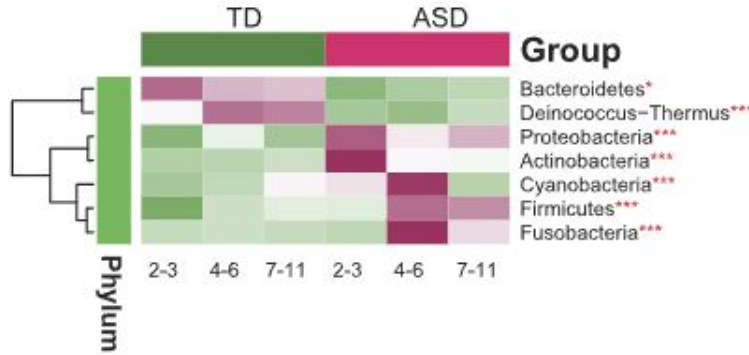
- In the TD group, the Shannon index increased from ages 2–3 to 7–11
- The ASD group showed stable Shannon index values across different ages

- **Phylogenetic Diversity (PD) Index**

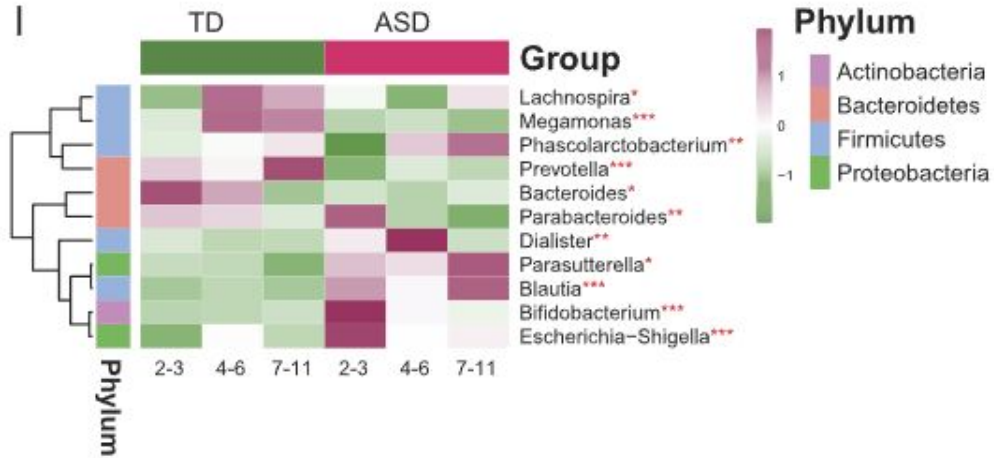
- In the TD group, the PD index increased from ages 2–3 to 4–6 and 7–11
- The ASD group displayed no significant change in the PD index with age

Relative abundance of phylum (k) and genus (l) genera with age growth between two groups

k



l



Microbial Profiles

- **Phylum-level: Bacteroidetes**
 - Decreased with age in the TD group
 - Remained low in the ASD group regardless of age
- **Phylum-level: Cyanobacteria**
 - Increased with age in the TD group
 - Did not show significant changes in the ASD group
- **Genus-Level Changes**
 - **Prevotella**, **Lachnospira**, and **Megamonas** became more abundant from ages 2–3 to 7–11 in the TD group. This change was not observed in the ASD group
 - **Bacteroides** increased at age 2–3 in the TD group but decreased with age; ASD group had lower abundance levels at all ages

Summary of the Research

Gut Dysbiosis in ASD

- **Levels Affected:** Phylum, genus, and species levels
- **Alpha-Diversity:** No significant age-related change in α -diversity in ASD; TD group showed increased diversity with age

Gut Microbiota Composition

- **ASD vs. TD**
 - **Reduced Genera:** Prevotella and Megamonas were less common in ASD
 - **Increased Genera:** Escherichia-Shigella, Dialister, and Bifidobacterium were more abundant in ASD

2. Introduction to Numerical Taxonomy

Numerical taxonomy is a method for classifying species based on quantitative or numerical traits. It aims to organize and analyze biodiversity by systematically measuring and comparing specific characteristics of organisms. This approach was developed by Sneath and Sokal.

Key Aspects of Numerical Taxonomy

- **Process**

- **Character Determination**

- Begins with identifying the presence or absence of selected characters in the organisms being studied

- **Number of Characters**

- Typically involves at least 60 different characters to ensure comprehensive analysis

- **Types of Data Used**

- **Morphological:** Physical characteristics such as size, shape, and structure

- **Biochemical:** Chemical compositions and metabolic processes

- **Genotypic:** characters like size of the genome, G+C%, 16s rRNA sequencing (Ribotyping)

- **Physiological:** Functional aspects such as growth rates and environmental responses

Principles of Numerical taxonomy

1. The greater the content of information in the taxa, and more the characters taken into consideration, the better a classification system will be
2. Every character should be given equal weightage in creating new taxa
3. For comparison purpose, the similarity between any two entities is considered
4. Correlation of characters differs in the groups of organisms under study. Thus distinct taxa can be recognized
5. Phylogenetic conclusions can be drawn from the taxonomic structure of a group and from character correlations, assuming some evolutionary mechanisms and pathways
6. The science of taxonomy is viewed and practiced as an empirical science.
7. Phenetic similarity is the basis of classifications

Mathematical Equations

Mathematical formula used are-

Pairing affinity = (No of characters same in both organisms/ Total no of characters analyzed) x 100

$$\text{The simple matching coefficient } (S_{SM}) = \frac{a + d}{a + b + c + d}$$

$$\text{The Jaccard coefficient } (S_J) = \frac{a}{a + b + c}$$

a = number of characters coded as present (1) for both organisms
 b and c = numbers of characters differing (1,0 or 0,1) between the two organisms
 d = number of characters absent (0) in both organisms
Total number of characters compared = $a + b + c + d$

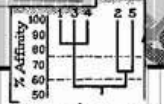
		Organism B	
		1	0
Organism A	1	a	b
	0	c	d



Character Matrix

	1	2	3	4	5
1	3	0	4	5	1
2	2	7	3	2	5
3	9	2	8	8	4
4	4	5	3	2	6
5	6	9	6	7	9

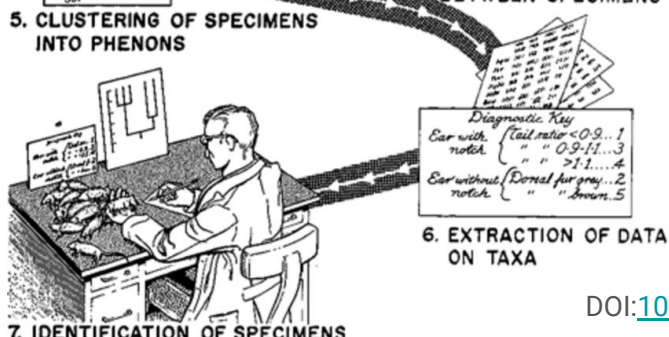
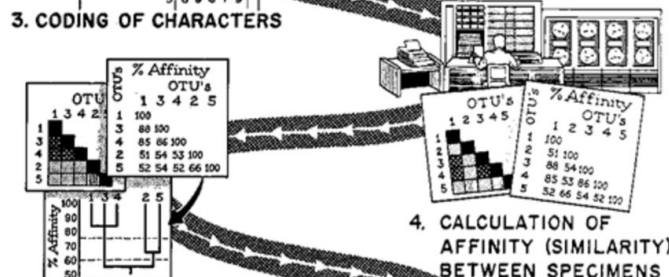
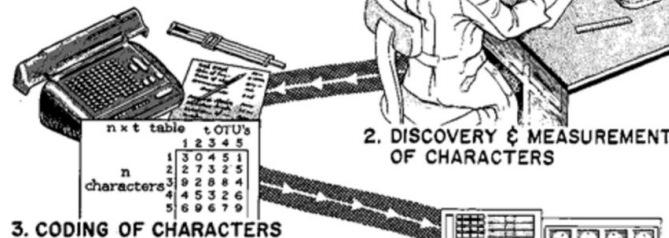
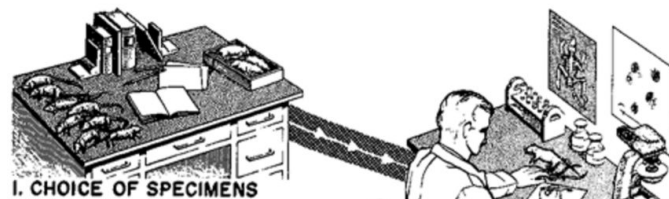
Methodologists



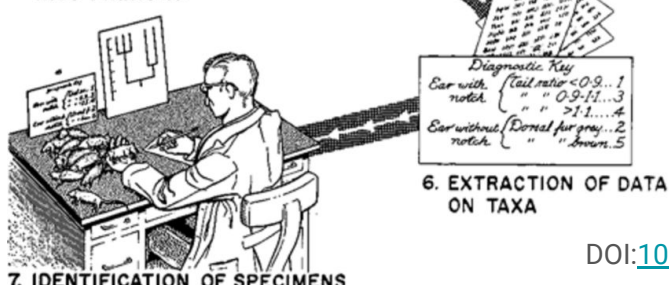
Taxon Specialists



Diagnostic Key
 Ear with notch { Tail ratio < 0.9...1
 " " 0.9-1.1...3
 " " > 1.1...4
 Ear without notch { Dorsal fur gray...2
 " " Brown...5



5. CLUSTERING OF SPECIMENS INTO PHENONS



7. IDENTIFICATION OF SPECIMENS

3. Core Techniques of Numerical Taxonomy

1. **Measuring Traits:**

Quantitative traits such as height, weight, size, or other biological characteristics are recorded numerically.

2. **Creating a Trait Matrix:**

Data is organized into a matrix with rows representing individuals or species and columns representing traits.

3. **Quantitative Analysis:**

Principal Component Analysis (PCA): Reduces the number of traits to highlight the main components of the data.

Cluster Analysis: Groups individuals or species based on trait similarity using methods like hierarchical clustering and k-means.

Distance Analysis: Measures differences or similarities between objects based on traits, e.g., Euclidean distance, Bray-Curtis distance.

5. **Building Phylogenetic Trees:**

Results are often displayed as dendrograms, showing relationships based on trait similarity.

6. **Using Quantitative Indices:**

Richness Index: Measures the number of species

Shannon Index: Assesses species abundance and evenness

Phylogenetic Diversity (PD) Index: Measures evolutionary divergence between groups

4. SUMMARY

- Numerical taxonomy techniques such as trait measurement, PCA, cluster analysis, and distance analysis provide insights into microbial diversity and relationships.
- The study highlights how numerical taxonomy methods can elucidate differences in gut microbiota composition between ASD and TD groups, as well as the impact of gastrointestinal symptoms.