



Network Science

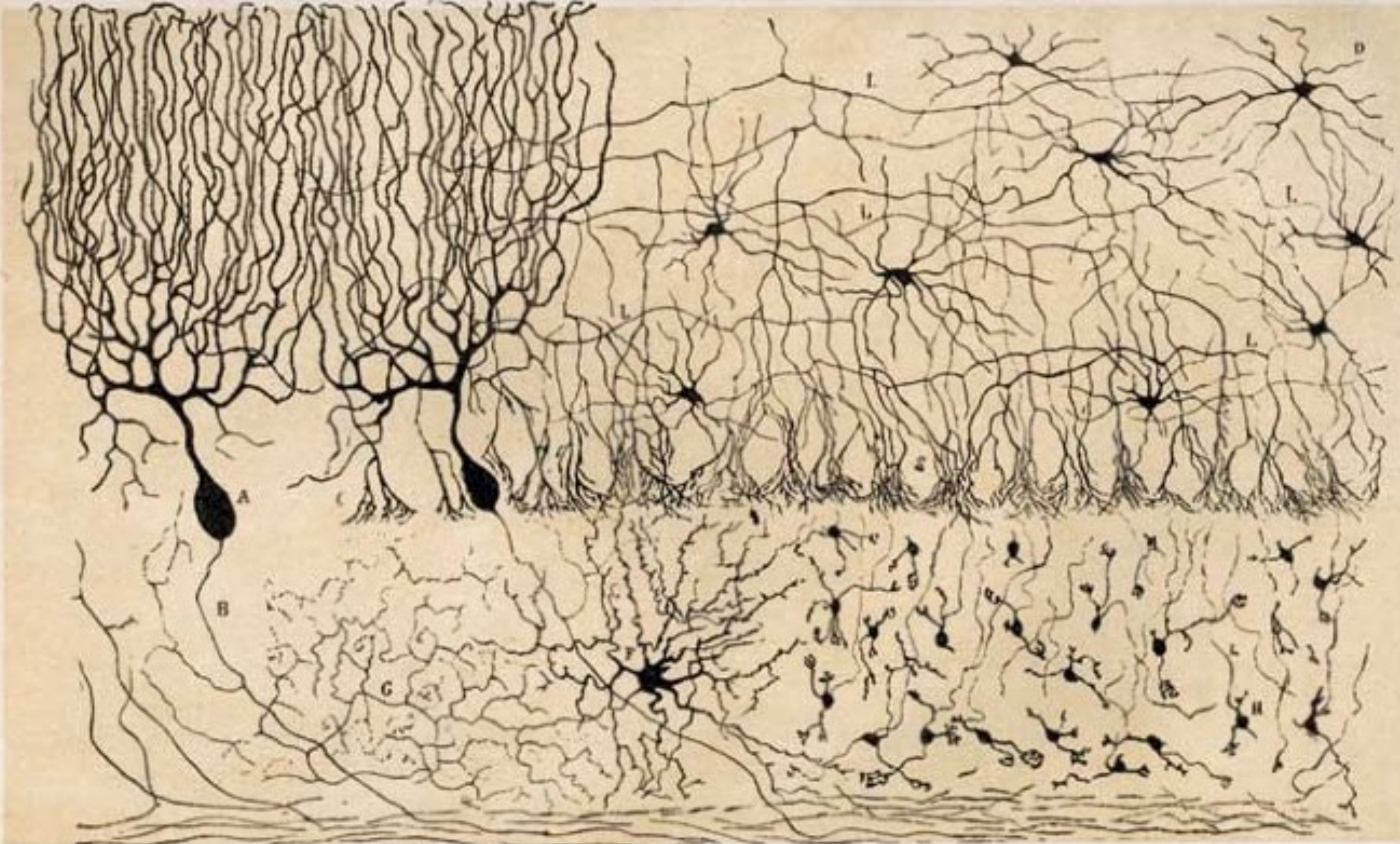
Basics

Savvas Paragkamian, post doc

What is a network?

From omics to networks?

Examples



600 πΧ Ομφαλος Δελφών
Δίκτυο Ανθρωπινων Σχέσεων

http://en.wikipedia.org/wiki/Oren_Patashnik



From Prof I. Antoniou

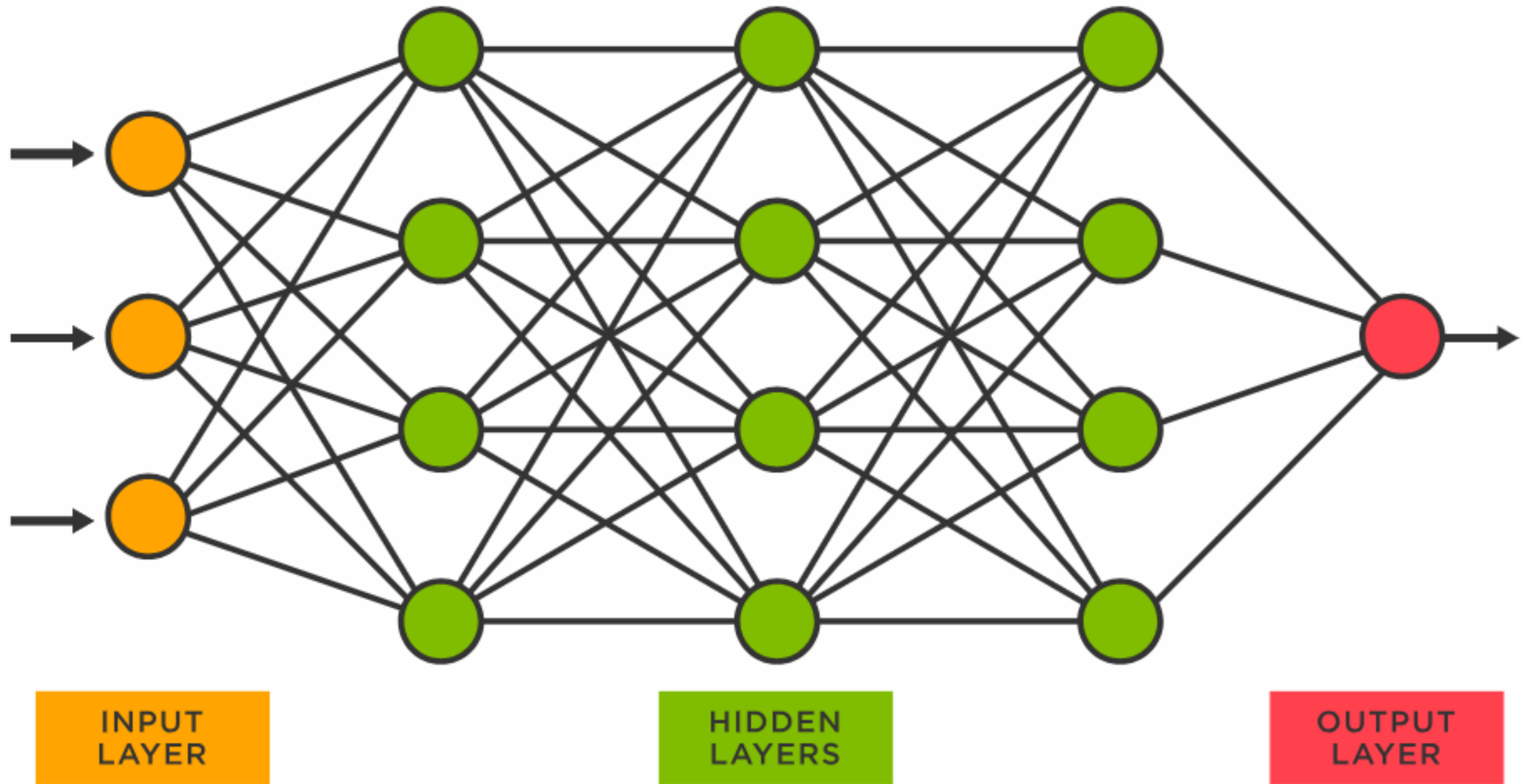
Complexity

Less is more

More Is Different

Large Language Models
Diffusion models
Deep Neural Networks

ChatGPT
AlphaFold
EVO



A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

Backfed Input Cell

Input Cell

Noisy Input Cell

Hidden Cell

Probablistic Hidden Cell

Spiking Hidden Cell

Output Cell

Match Input Output Cell

Recurrent Cell

Memory Cell

Different Memory Cell

Kernel

Convolution or Pool

Perceptron (P)



Feed Forward (FF)



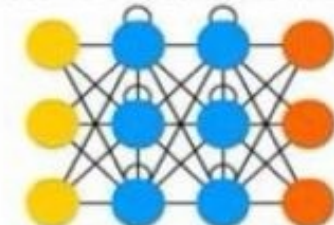
Radial Basis Network (RBF)



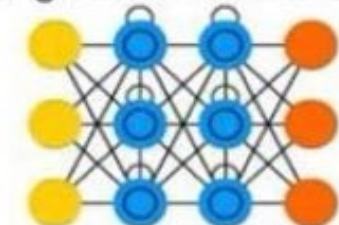
Deep Feed Forward (DFF)



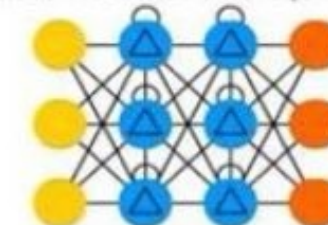
Recurrent Neural Network (RNN)



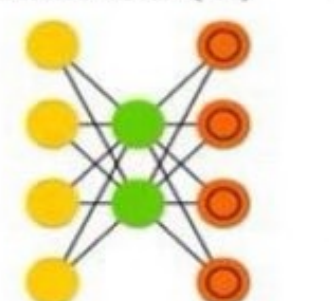
Long / Short Term Memory (LSTM)



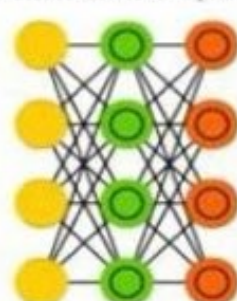
Gated Recurrent Unit (GRU)



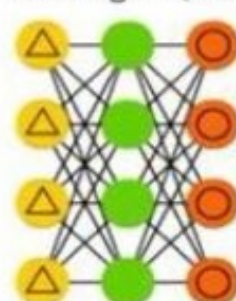
Auto Encoder (AE)



Variational AE (VAE)



Denoising AE (DAE)



Sparse AE (SAE)



Markov Chain (MC)



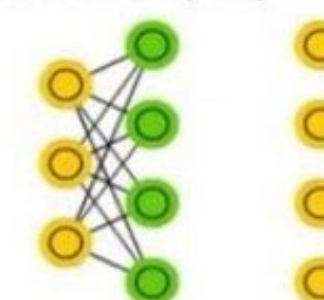
Hopfield Network (HN)



Boltzmann Machine (BM)



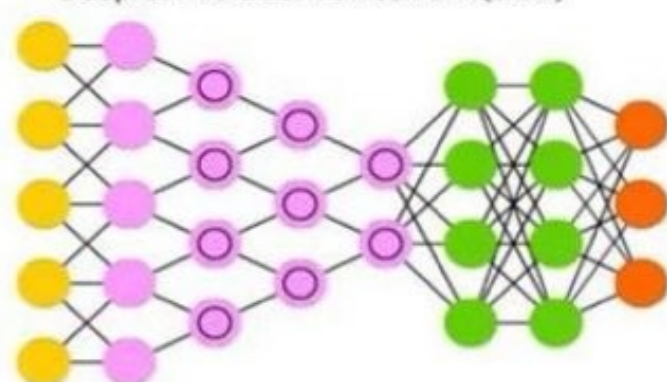
Restricted BM (RBM)



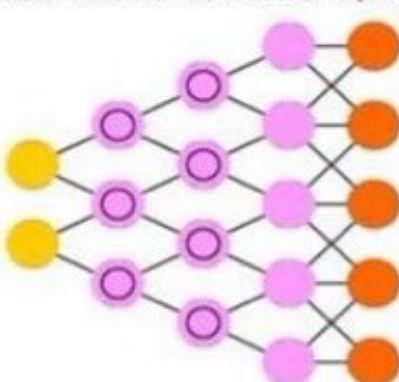
Deep Belief Network (DBN)



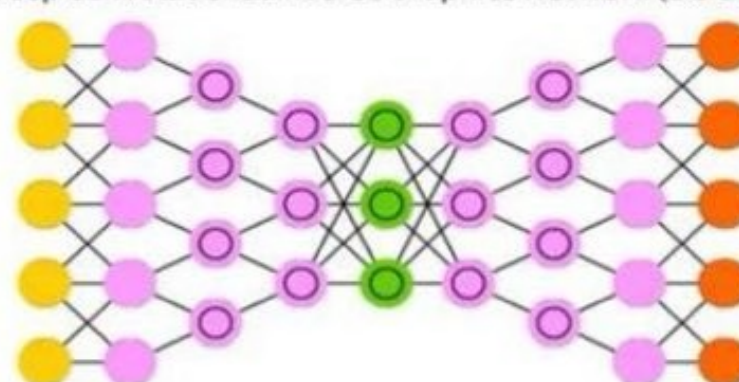
Deep Convolutional Network (DCN)



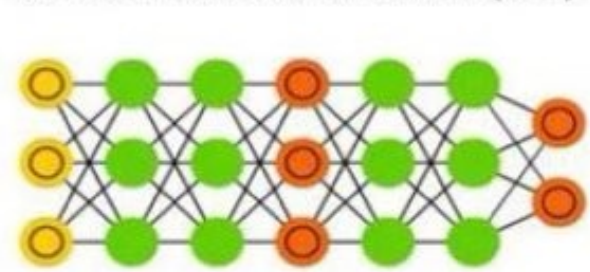
Deconvolutional Network (DN)



Deep Convolutional Inverse Graphics Network (DCIGN)



Generative Adversarial Network (GAN)



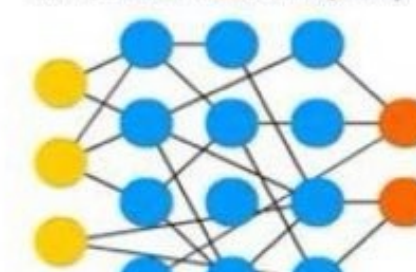
Liquid State Machine (LSM)



Extreme Learning Machine (ELM)



Echo State Network (ESN)



Deep Residual Network (DRN)



Kohonen Network (KN)



Support Vector Machine (SVM)



Neural Turing Machine (NTM)



Types of Networks

Unipartite

Bipartite

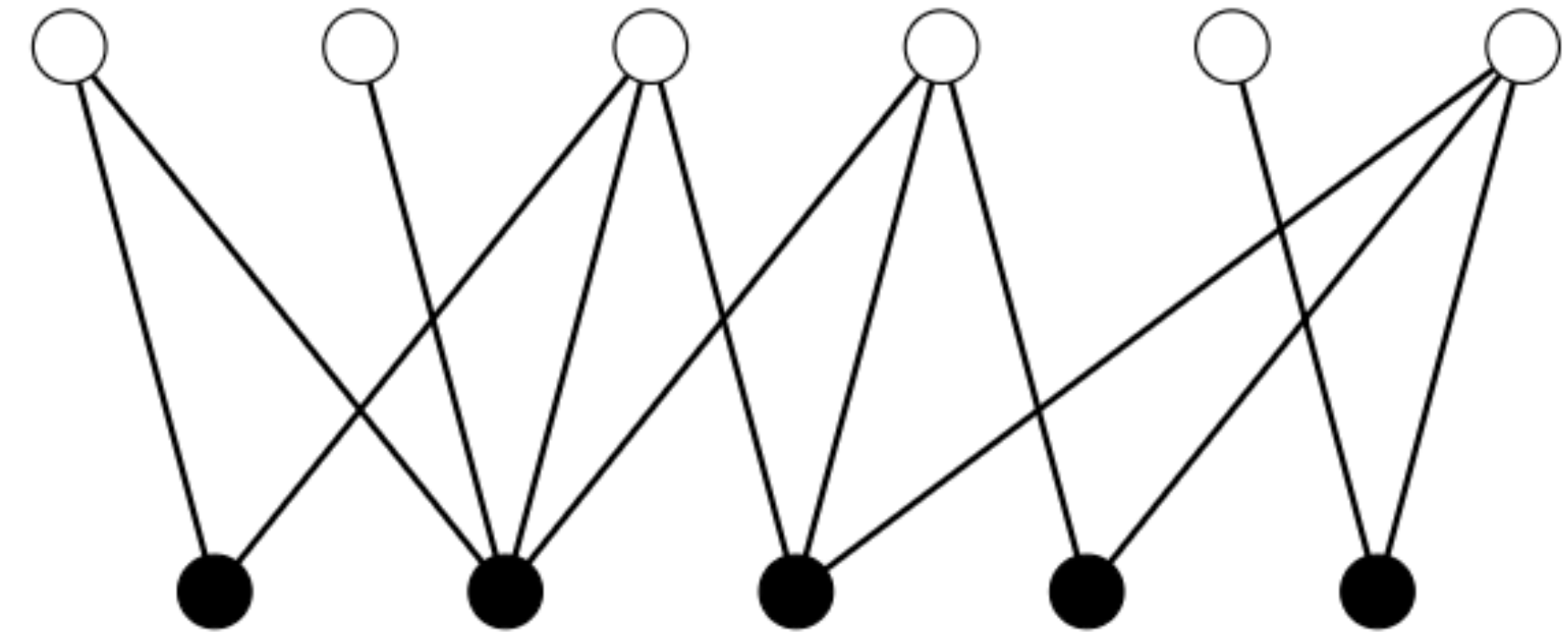
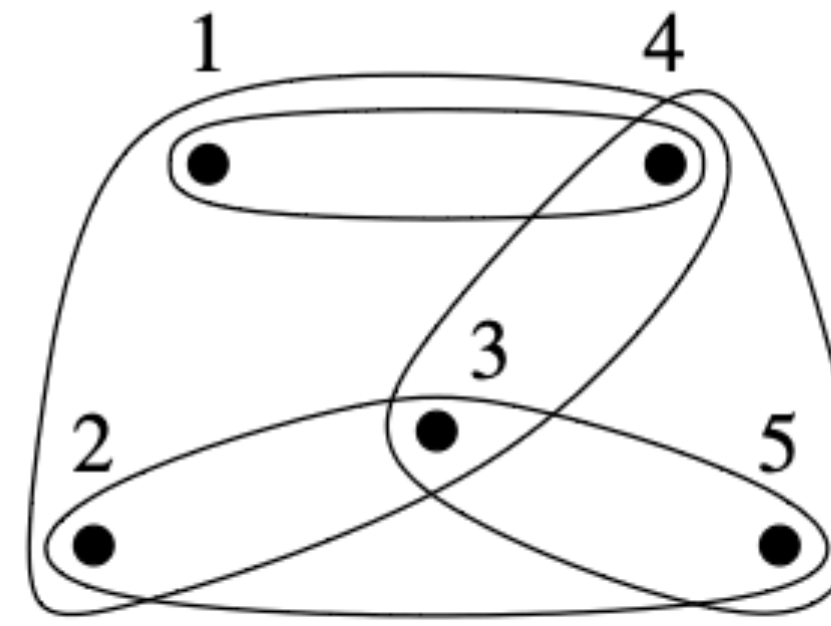
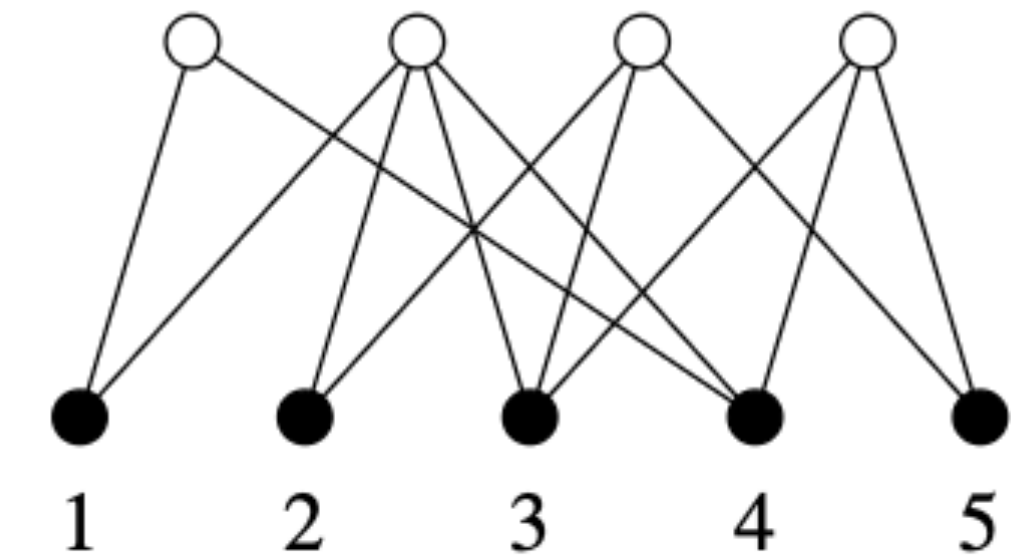


Figure 6.5: A small bipartite network. The open and closed circles represent two types of nodes and edges run only between nodes of different types. It is common to draw bipartite networks with the two sets of nodes arranged in lines, as here, to make the bipartite structure clearer. See Fig. 4.2 on page 50 for another example.

Hypergraph

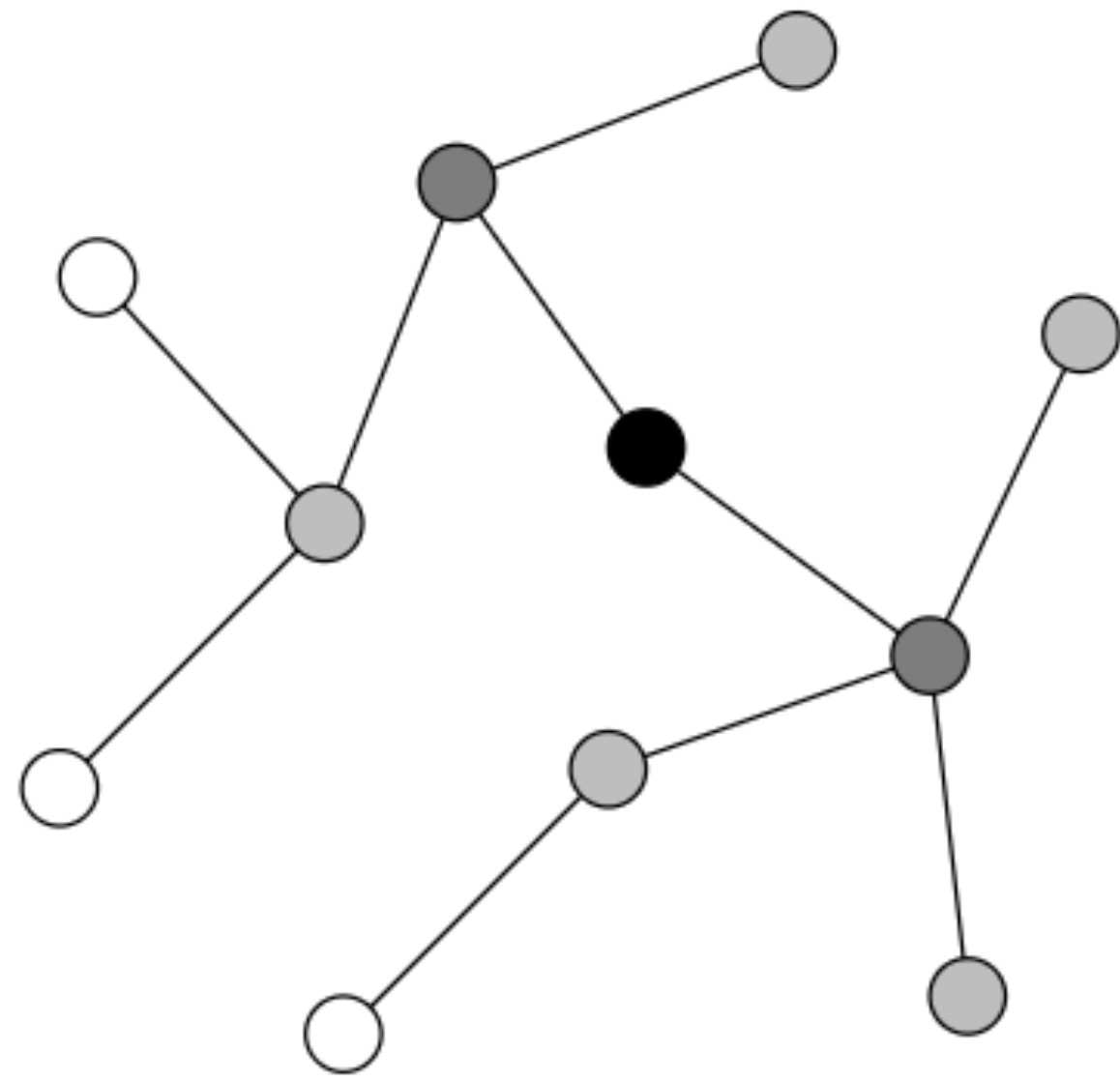


(a)

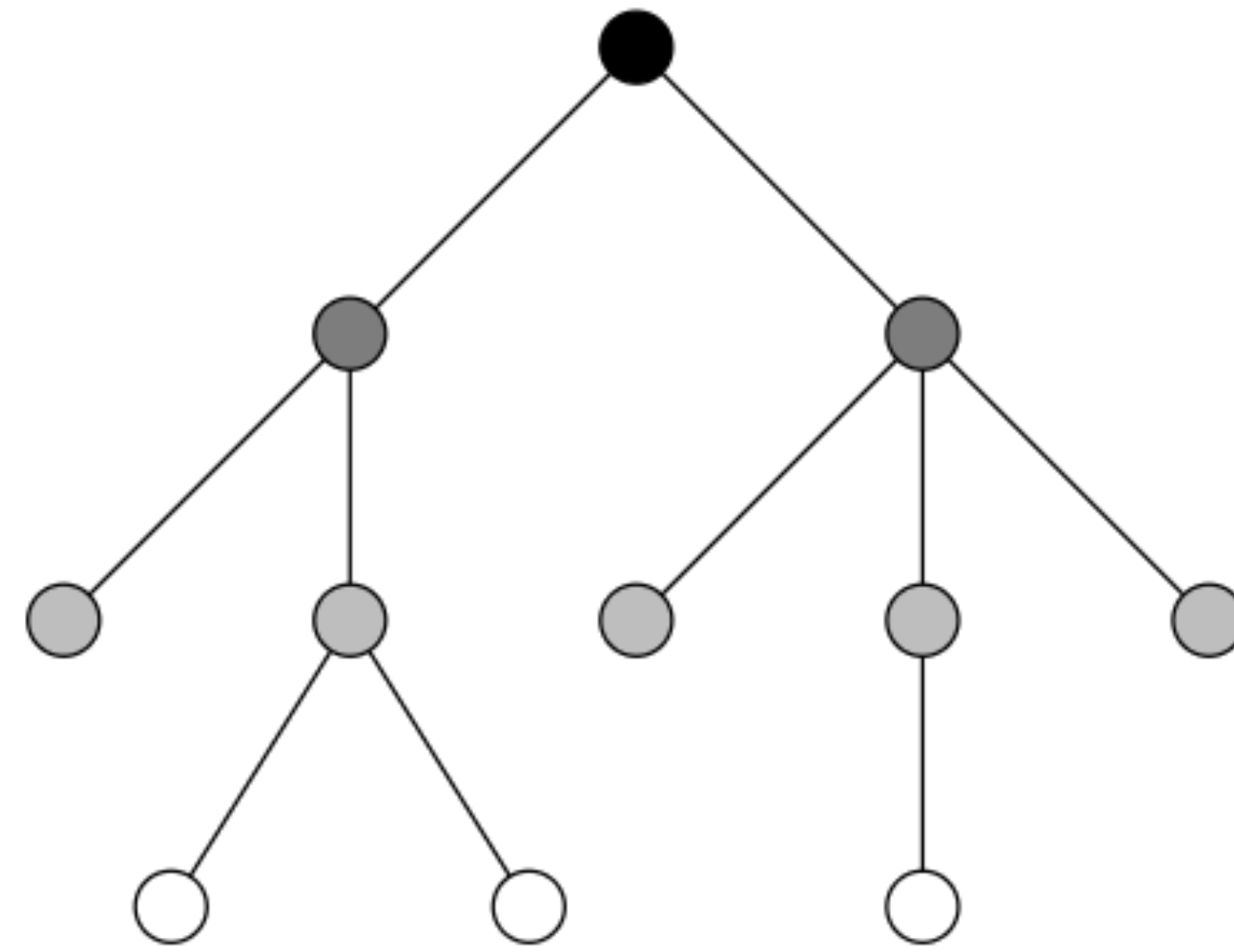


(b)

Figure 6.4: A hypergraph and corresponding bipartite graph. These two networks convey the same information—the membership of five nodes in four different groups. (a) The hypergraph representation in which the groups are represented as hyperedges, denoted by the loops circling sets of nodes. (b) The bipartite representation in which we introduce four new nodes (open circles at the top) representing the four groups, with edges connecting each of the original five nodes (bottom) to the groups to which it belongs.

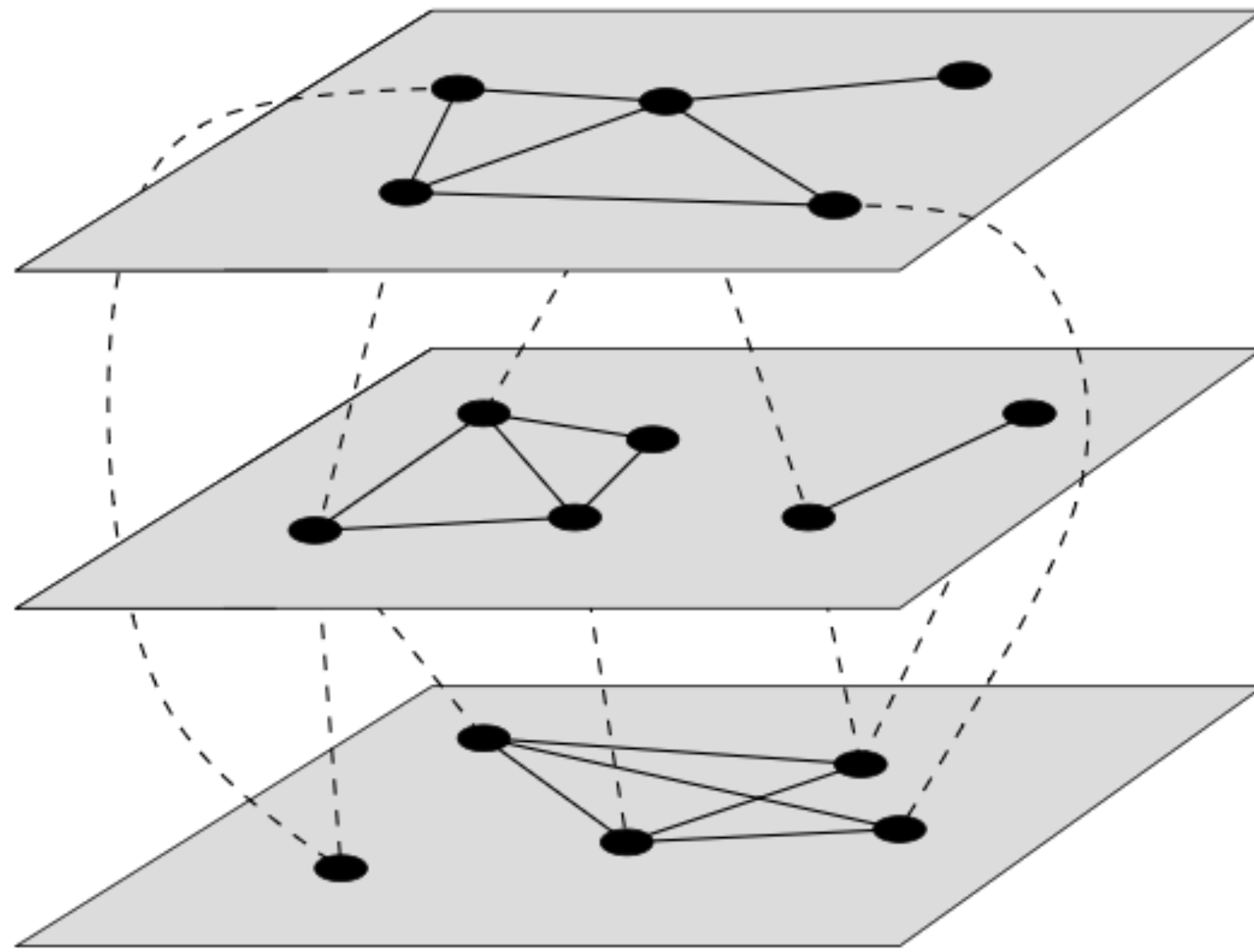


(a)

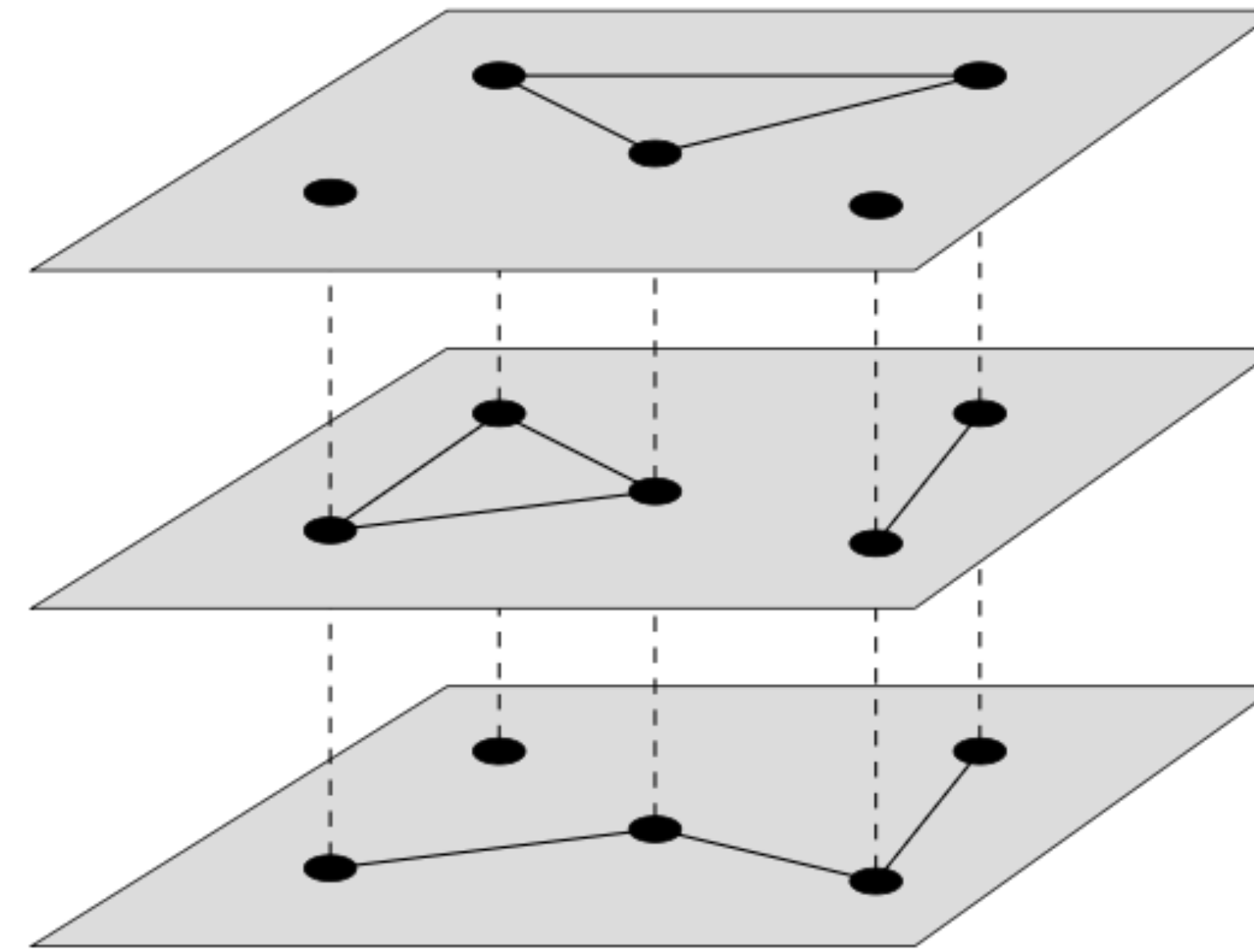


(b)

Figure 6.8: Two sketches of the same tree. The two panels here show two different depictions of a tree, a network with no closed loops. In (a) the nodes are positioned on the page in any convenient position. In (b) the tree is laid out in a “rooted” fashion, with a root node at the top and branches leading down to “leaves” at the bottom.



(a)

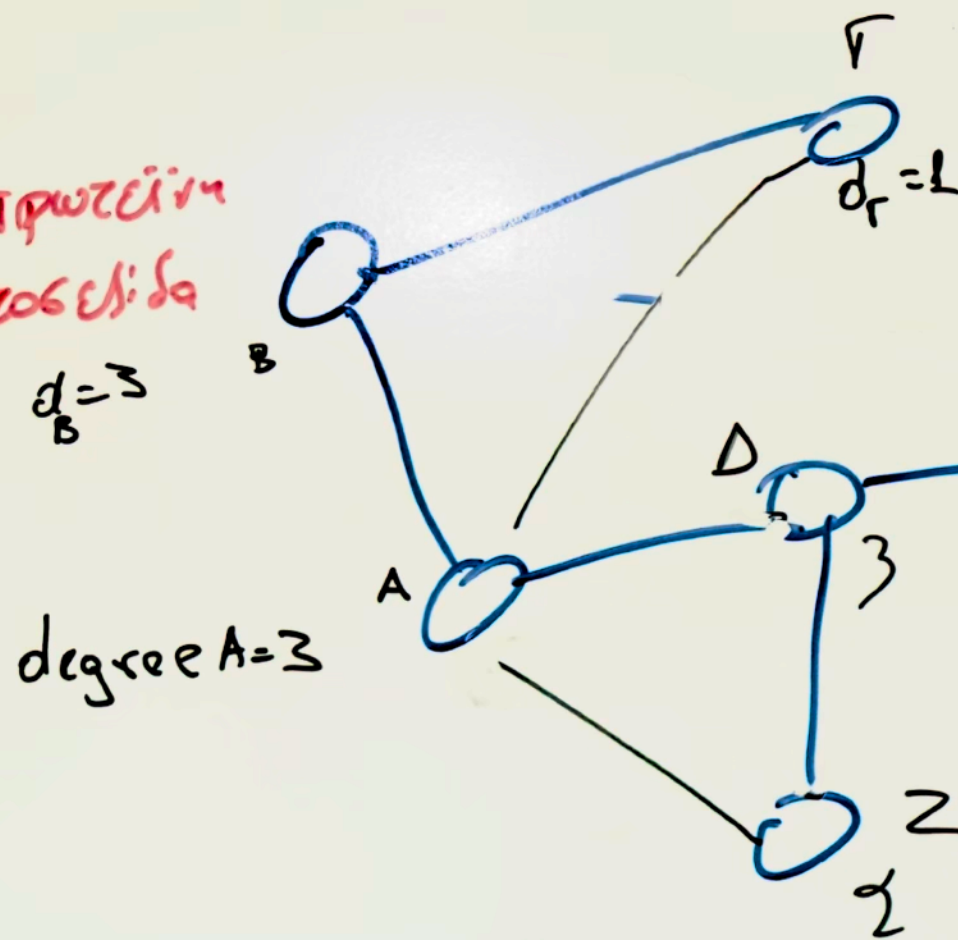


(b)

Figure 6.7: Multilayer and multiplex networks. (a) A multilayer network consists of a set of layers, each containing its own network, plus interlayer edges connecting nodes in different layers (dashed lines). An example is a transportation network with layers corresponding to airlines, trains, buses, and so forth. (b) A multiplex network is a special case of a multilayer network in which the nodes represent the same set of objects or people in each layer. For instance, a social network with several different types of connections could be represented as a multiplex network with one layer for each type. Dynamic or temporal networks are another example, where the layers represent snapshots over time of the structure of a single, time-varying network. In principle one can include interlayer edges in a multiplex network, as here, to represent the equivalence of nodes in different layers, although in practice these are often omitted.

Summary of 21/05/2025

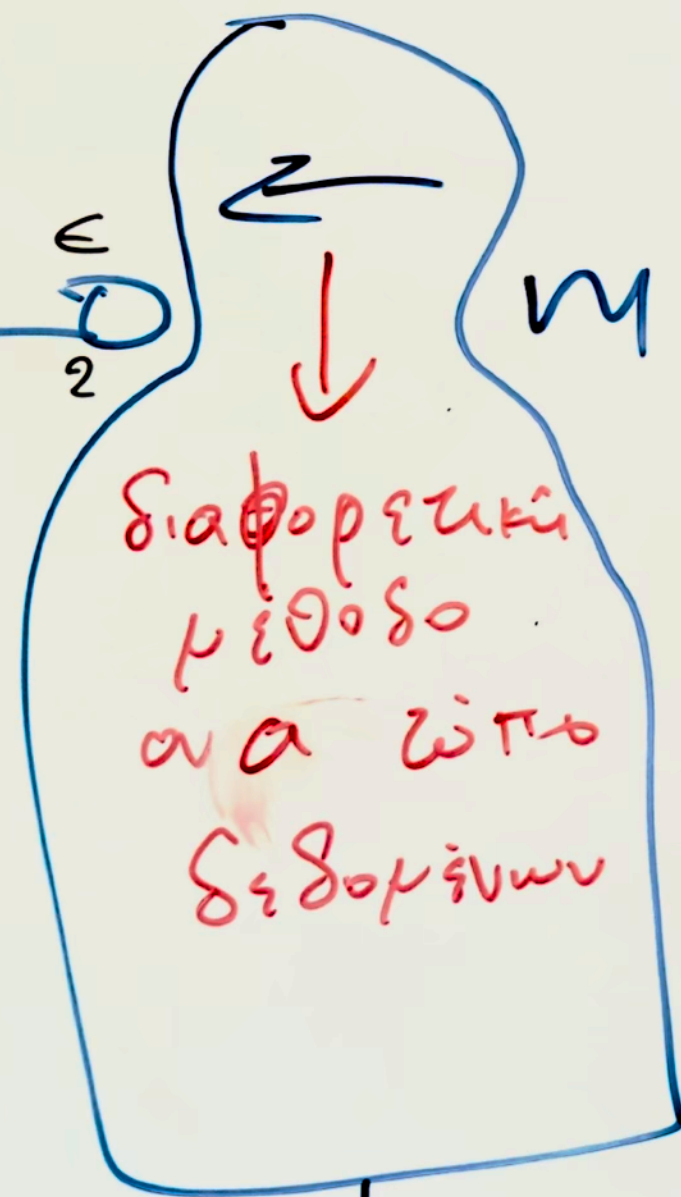
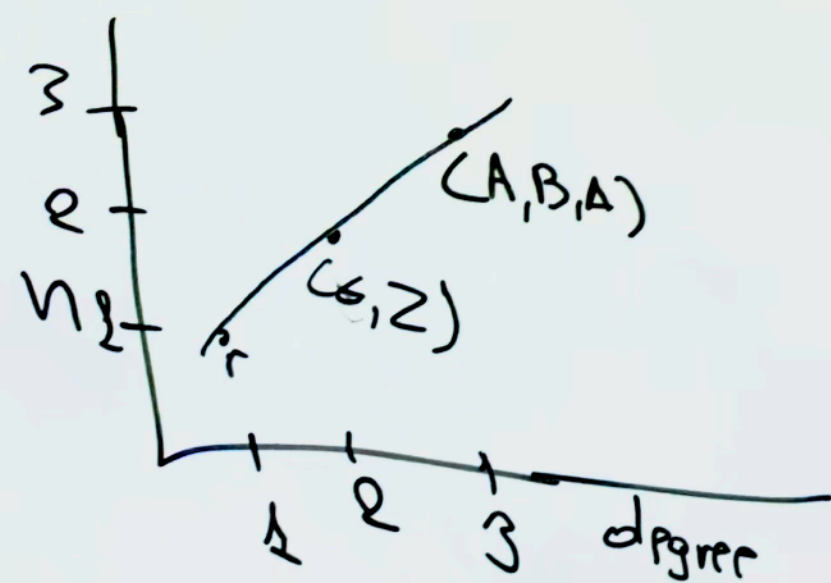
Πρώτη
162061/162062



degree A=3

- Τι είναι ο κόμβος
- Τι είναι η σύνδεση

Κατανομή Βαθμίου



Γιατί ο βιολόγος
δεν είναι
ευχαριστημένος

Pareto Principle 20-80

	g-1	g-2	g-3	g-4
συ. 1
συ. 2
συ. 3

Text
Entities
word 1
word 2

Mining
Score
score

Ποσο
ποσα
είναι
οι λέξεις

Context
↓
LLM

Μονοπάτια

$Z \rightarrow \Gamma = A \rightarrow B \rightarrow \Gamma (3), D \rightarrow B \rightarrow \Gamma (4), A \rightarrow B \rightarrow \Gamma (4)$

Average Shortest Path Six degrees of Separation

28/05/2025

Rmarkdown - Network Basics

15 minutes Break

Soil Metagenome - From sequences to microbe co-occurrences

- Take a look at Metagenomics course by Dr Lagkouvardos



Average TREC team member's statistics

240

field days

67

sampling sites

18

countries



378

total expedition days
for AML in 2023–24



Hummus consumed
several kilos

Cereal bars eaten
around 1500



Times of a sampling video shoots:

Earliest

04:15

Latest

02:10

163kg

carried on a video shooting day

1 theme song

For further info please
contact AML team



The success of this complex international project is a testament to the dedication of all teams involved. Credit: Savvas Parakamian/HCMR, Gerals Pfister, Joanna Zukowska, and Kinga Lubowiecka, Creative Team/EMBL

An expedition where land meets sea



TREC expedition stops 2024

<u>Malaga, Spain</u>	Late February
<u>Mallorca, Spain</u>	Early March
<u>Barcelona, Spain</u>	Mid-March
<u>Banyuls, France</u>	Late March
<u>Villefranche-sur-Mer, France</u>	Early April
<u>Pisa, Italy</u>	Mid-April
<u>Naples, Italy</u>	Late April
<u>Calabria, Italy</u>	Early May
<u>Lesina, Italy</u>	Mid-May
<u>Chioggia, Italy</u>	Late May
<u>Split, Croatia</u>	Early June
<u>Kotor, Montenegro</u>	Mid-June
<u>Athens, Greece</u>	Early-July
<u>Thrace, Greece</u>	Mid-July

Follow the expedition via
our interactive map
<https://trec.embl.de/itinerary.cgi>



TREC 2023 IN NUMBERS

8

months of expedition

65

land-sea interface and selected
species sampling sites (EMBL)

99

sampling sites at sea (*Tara*)

33.000

samples in total

5

sites with advanced mobile
lab support



LEGEND

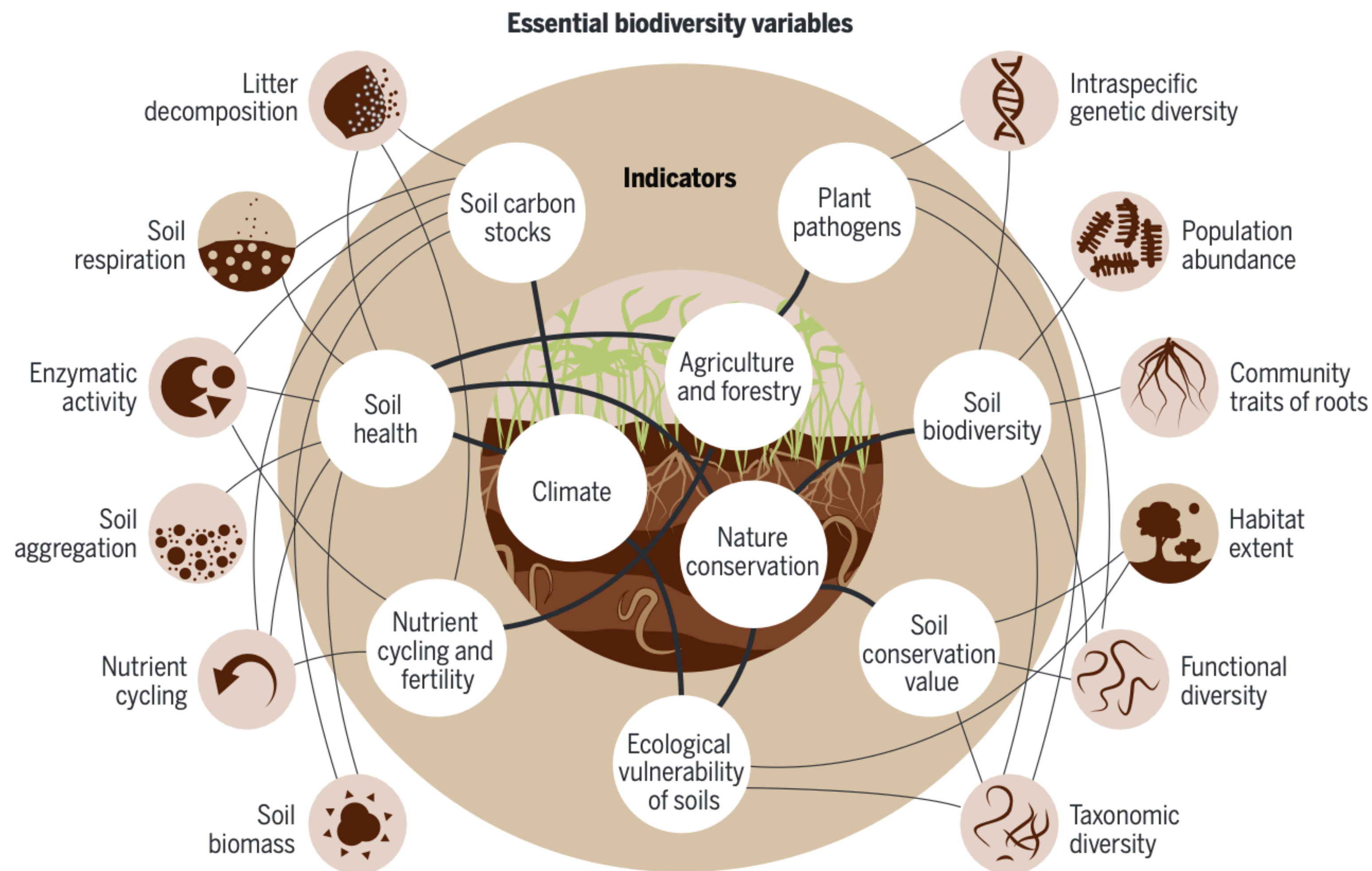
- Common stopovers TREC / Tara EUROPA
- Port calls Tara EUROPA
- Sampling Coastal Sites
- 2023
- 2024

The TREC expedition began in Roscoff, France, in spring 2023 and will conclude in Thessaloniki, Greece, in July 2024. During this period, researchers from EMBL, the Tara OceanS consortium, together with the Tara Ocean Foundation, and numerous European collaborating institutes and organisations will be working at 120 sampling sites across 21 European countries.

- TREC <https://www.embl.org/about/info/trec/>
- Salamina <https://www.youtube.com/watch?v=FcYOZWYtmms>
- Psatha <https://www.youtube.com/watch?v=BQ2hJBeSBwQ>

Linking soil biodiversity to policy

Links between global soil essential biodiversity variables (EBVs) (outer ring) are prioritized by the Soil Biodiversity Observation Network (SoilBON) and policy sectors (center) through the use of soil ecological indicators (inner ring; table S3). Thin lines correspond to links between EBVs and soil indicators; thicker lines refer to links between each soil indicator and specific policy sectors. The EBVs for soil systems are proposed as a holistic system approach (table S2), where soil organisms are intertwined with relevant soil chemical, physical, and functional properties, contributing to overall societal well-being. See table S1 for further information on links to specific policy targets and policies. See table S2 for details of the EBVs.



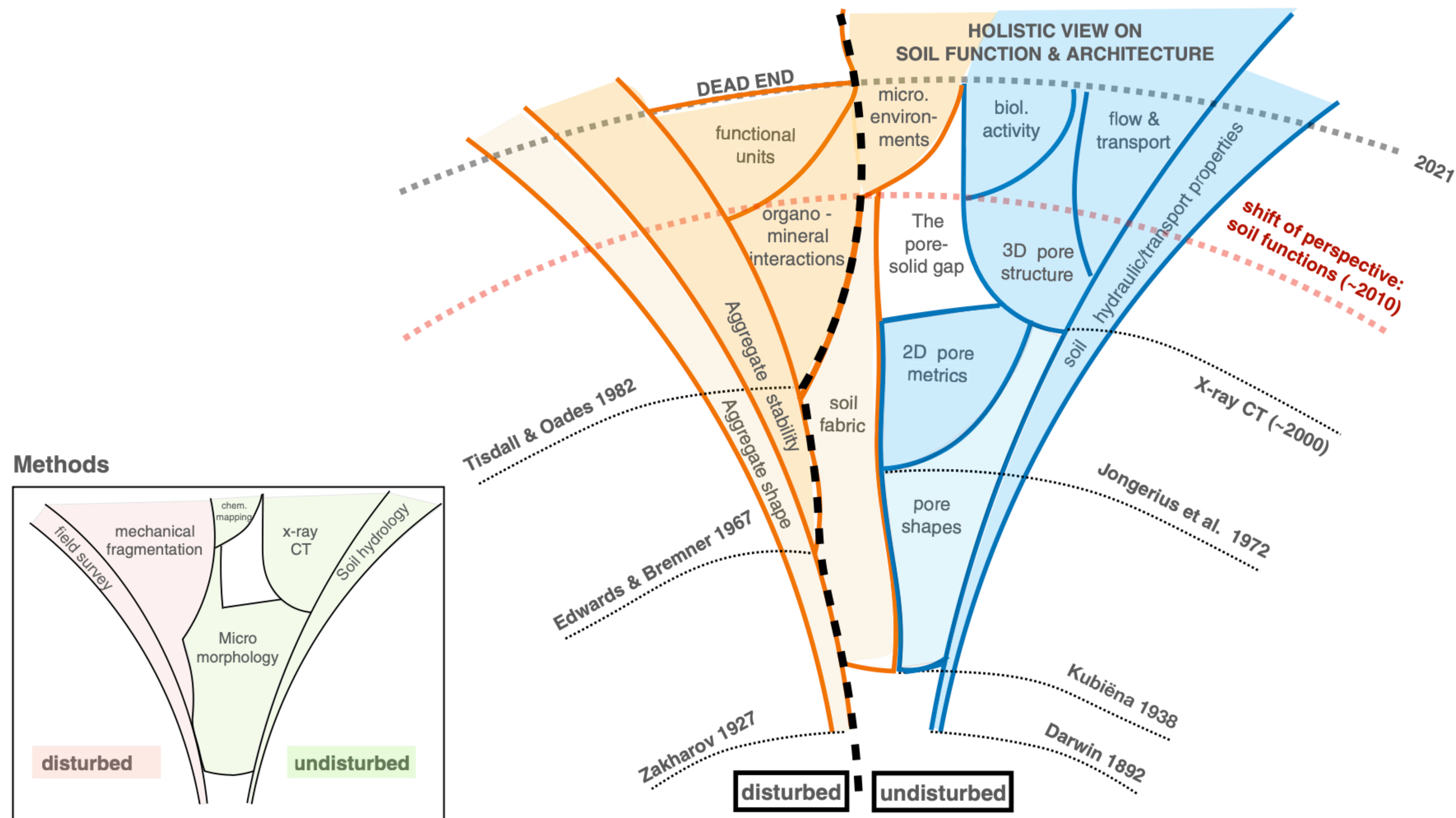


FIGURE 1 Historical development of exploring soil structure. The pore perspective is in blue, the solid perspective in ochre. Dark colours and light colours indicate quantitative and qualitative analysis, respectively. New branches of developments with respect to research foci are marked by milestone publications or new technical developments

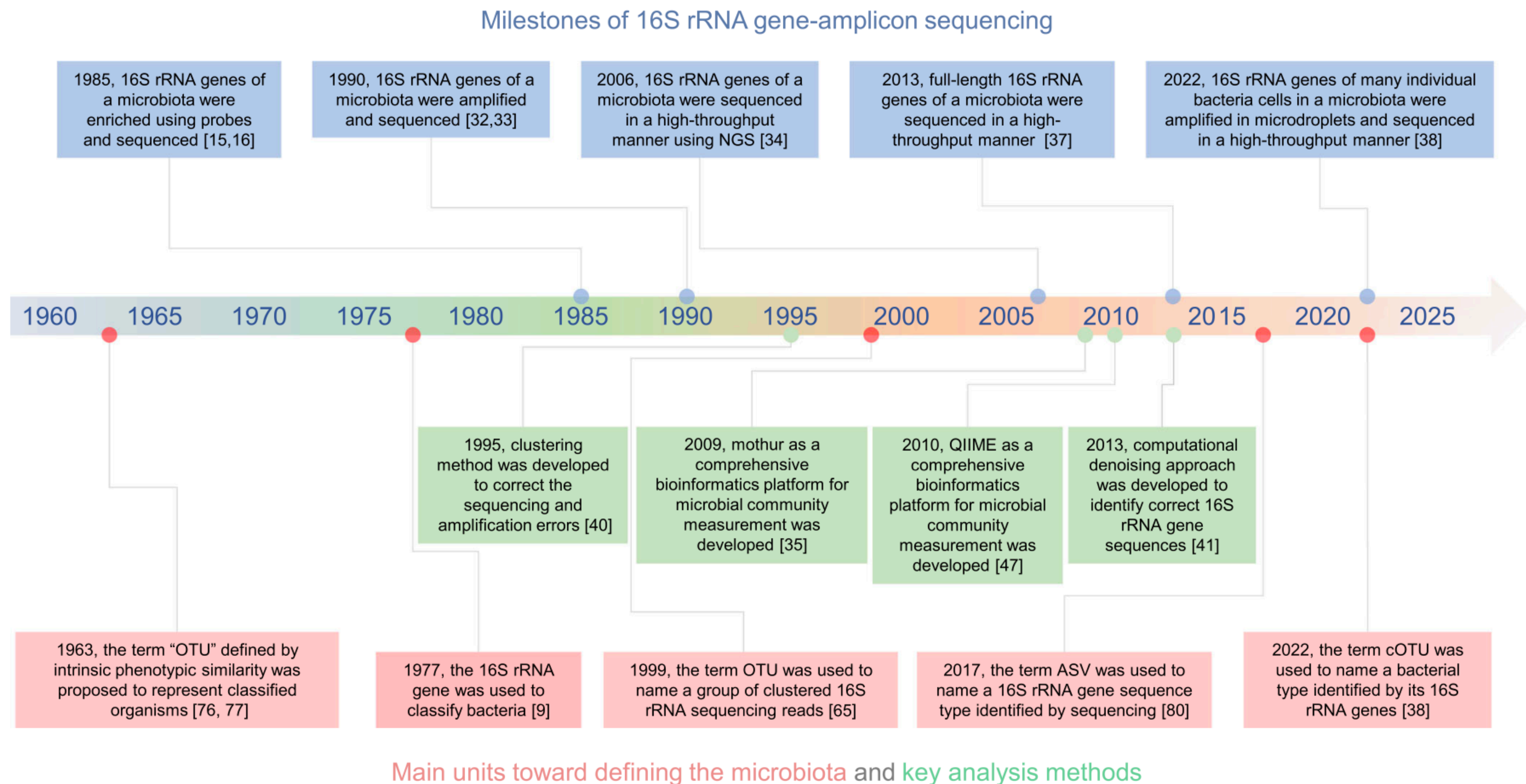


FIGURE 1 History of 16S rRNA gene-amplicon sequencing, key analysis methods, and related classification units. ASV, amplicon sequence variants; cOTU, cell-based operational taxonomic unit; NGS, next-generation sequencing; OTU, operational taxonomic unit; QIIME, quantitative insights into microbial ecology; rRNA, ribosomal RNA.

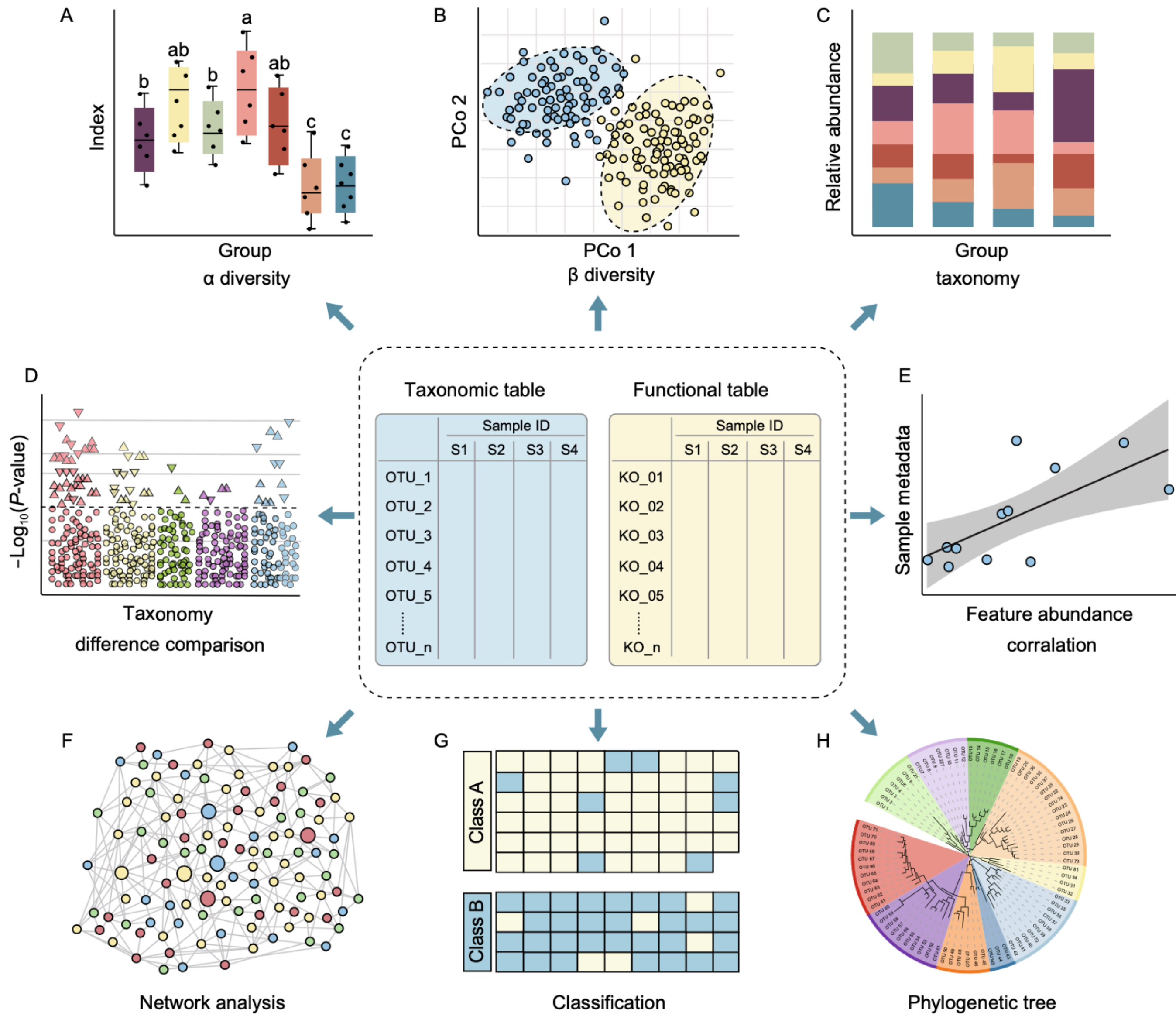


Figure 3. Overview of statistical and visualization methods for feature tables. Downstream analysis of microbiome feature tables, including alpha/beta-diversity (A/B), taxonomic composition (C), difference comparison (D), correlation analysis (E), network analysis (F), classification of machine learning (G), and phylogenetic tree (H). Please see Table 2 for more details.

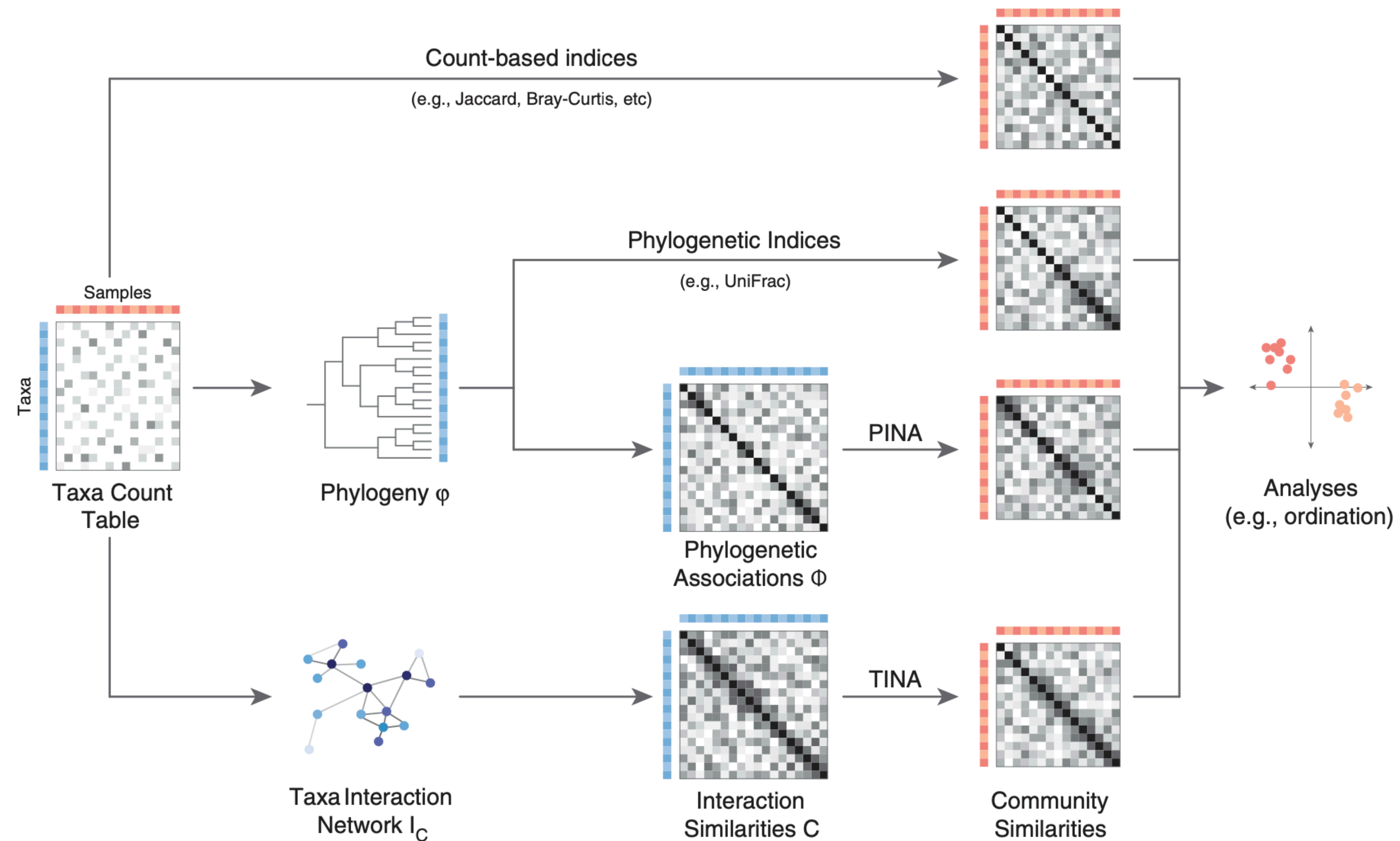


Figure 1 Overview of different approaches to quantifying community similarity. Based on a taxa-sample count table, traditional count-based indices such as Jaccard and Bray–Curtis quantify community similarity from the overlap in taxa composition (upper branch). In contrast, phylogenetic indices such as UniFrac take into account taxa relationships, quantifying community similarity as shared evolutionary history, based on taxa phylogeny (middle branch). Our proposed Taxa Interaction-Adjusted (TINA) and Phylogenetic Interaction-Adjusted (PINA) indices, in contrast, take into account similarities on a taxa co-occurrence network, codified in an interaction similarity matrix C , or in terms of cophenetic phylogenetic distances, represented in a phylogenetic association matrix Φ .

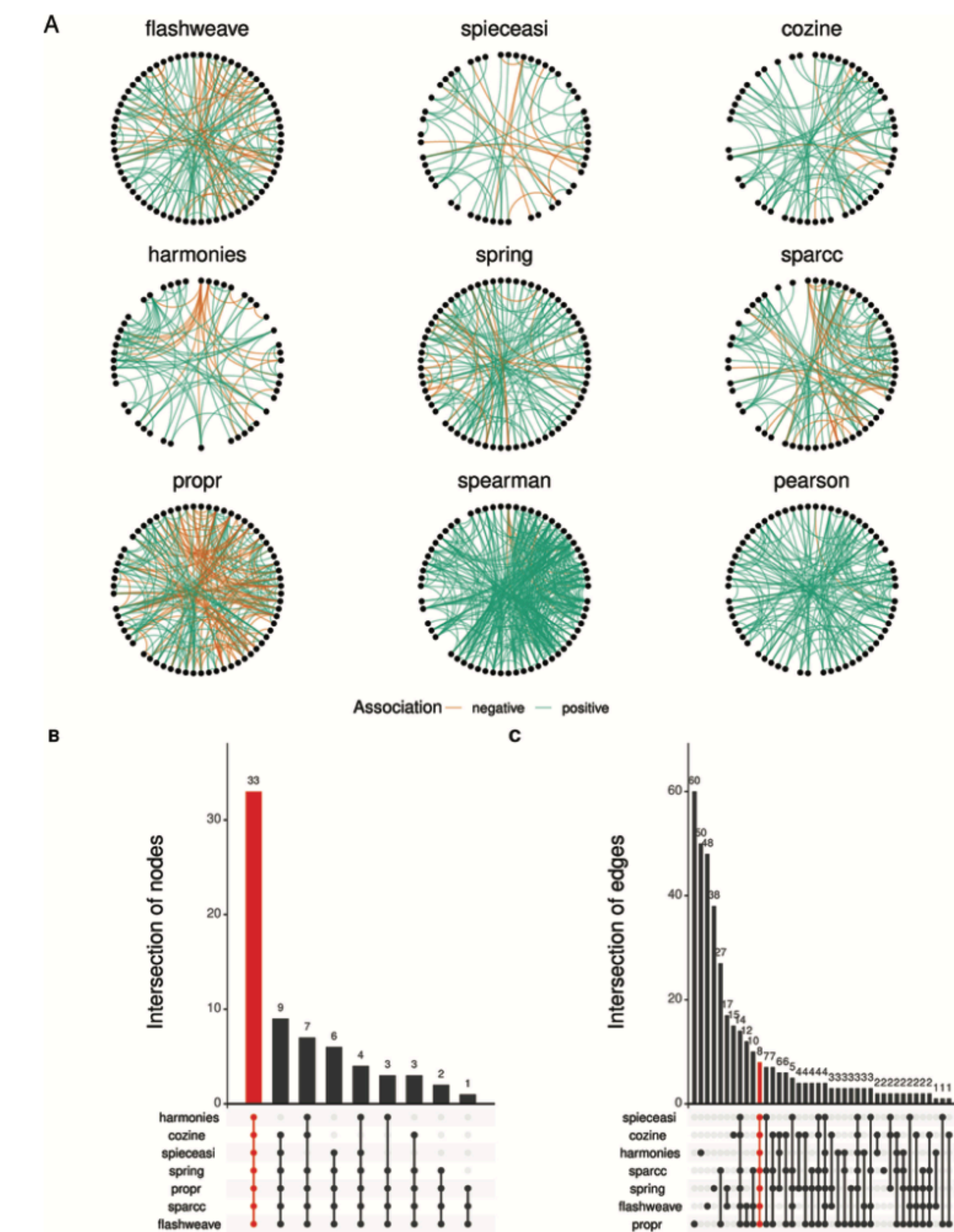


FIG 4 Networks generated using different network inference methods show notable differences in terms of edge-density and connectivity. (A) The nine different networks generated by the different network inference methods (excluding mLDM). The nodes for each network (representing taxa) are arranged in the same positions in a circular layout, and the differences in the connections can be directly visualized and compared. The green links are positive associations, and the orange links represent negative associations. The networks look dissimilar and vary widely in terms of connectivity, and it is notable that the correlation-based methods generally produce networks with higher edge-densities. A threshold of 0.3 was set for the correlation-based methods (sparcc, propr, spearman, and pearson), and a threshold of 0.01 was set for the direct association methods (flashweave, spieceasi, cozzine, harmonies, and spring). (B) The node overlap Upset plot indicates that all the networks have a large proportion of common nodes involved in connections (33 out of 68). Conversely (C), the edge overlap Upset plot shows that a very small fraction of these connections are actually shared (8 out of 202). The data used in this analysis were the healthy stool samples from the FMT data set. mLDM is not shown in the comparisons because the algorithm failed to converge for the particular network combination used here (default setting of the MiCoNE pipeline).



 | Computational Biology | Research Article

Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation

Dileep Kishore,^{1,2,3} Gabriel Birzu,^{4,5} Zhenjun Hu,¹ Charles DeLisi,^{1,4,6} Kirill S. Korolev,^{1,2,4} Daniel Segre^{1,2,4,6,7}

AUTHOR AFFILIATIONS See affiliation list on p. [25](#).

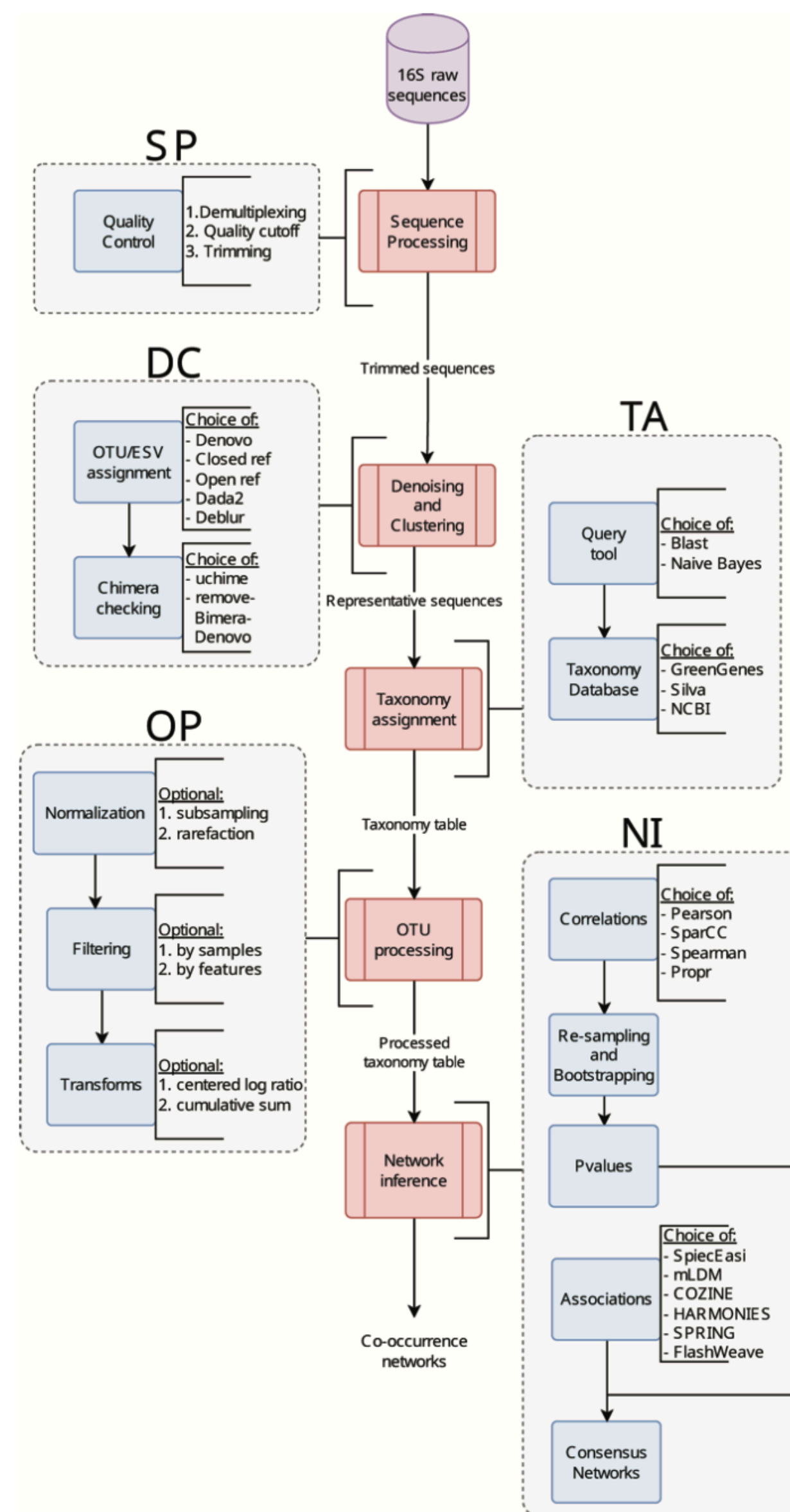


FIG 1 The workflow of the MiCoNE pipeline. The steps of the workflow can be broken down into five major groups: (SP) sequence processing, (DC) denoising and clustering, (TA) taxonomy assignment, (OP) OTU and ESV processing, and (NI) network inference. Each step incorporates several processes (blue boxes), each of which, in turn, has several alternative algorithms for the same task (indicated by the text to the right of the blue boxes). Each arrow describes the data that is being passed from one step to another. The inputs to the pipeline are 16S rRNA sequencing reads, and the final output is the consensus network generated from the inferred co-occurrence networks. For details on each process and the different outputs, see Methods.

Microbiome Datasets Are Compositional: And This Is Not Optional

Gregory B. Gloor^{1}, Jean M. Macklaim¹, Vera Pawlowsky-Glahn² and Juan J. Egozcue³*

¹ Department of Biochemistry, University of Western Ontario, London, ON, Canada, ² Departments of Computer Science, Applied Mathematics, and Statistics, Universitat de Girona, Girona, Spain, ³ Department of Applied Mathematics, Universitat Politècnica de Catalunya, Barcelona, Spain

16s amplicon workflow

*Or How I Learned to
Stop Worrying and
Love the Errors*

- 1. Current Challenges and Pitfalls in Soil Metagenomics**
- Pitfalls in the statistical analysis of microbiome amplicon sequencing data
- 3. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform**
- The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis
- 5. Biases in Prokaryotic Community Amplicon Sequencing Affected by DNA Extraction Methods in Both Saline and Non-saline Soil**
- Examining Sources of Error in PCR by Single-Molecule Sequencing, Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies
- Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies
- 8. Groundtruthing Next-Gen Sequencing for Microbial Ecology–Biases and Errors in Community Structure Estimates from PCR Amplicon Pyrosequencing**
- Examining Sources of Error in PCR by Single- Molecule Sequencing
- Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies
- Batch effects removal for microbiome data via conditional quantile regression
- Parsing ecological signal from noise in next generation amplicon sequencing
- Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil
- Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform
- 15.Ten quick tips for effective dimensionality reduction**
- 16.Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform**
- Handling of targeted amplicon sequencing data focusing on index hopping and demultiplexing using a nested metabarcoding approach in ecology
- Normalization and microbial differential abundance strategies depend upon data characteristics
- Analysis of microbial compositions/ a review of normalization and differential abundance analysis
- Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data
- Applications and Comparison of Dimensionality Reduction Methods for Microbiome Data
- Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units
- 23.Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold**
- Uniform Manifold Approximation and Projection (UMAP) Reveals Composite Patterns and Resolves Visualization Artifacts in Microbiome Data
- 25.Blocking Factors and Hypothesis Tests in Ecology: Is Your Statistics Text Wrong?**
- 26.Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations**
- 27.Parsing ecological signal from noise in next generation amplicon sequencing**
- 28.Batch effects removal for microbiome data via conditional quantile regression**
- Exact sequence variants should replace operational taxonomic units in marker-gene data analysis
- Amplicon Sequence Variants Artificially Split Bacterial Genomes into Separate Clusters
- 31.A field guide for the compositional analysis of any-omics data**
- A family of interaction-adjusted indices of community similarity
- Analysis of compositions of microbiomes with bias correction
- Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data
- 35.Microbiome Datasets Are Compositional: And This Is Not Optional**
- Microbiome differential abundance methods produce different results across 38 datasets
- Normalization and microbial differential abundance strategies depend upon data characteristics
- Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS)/ application to microbial communities
- 39.From hairballs to hypotheses–biological insights from microbial networks**
- Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation
- Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data
- 42.Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies**

Σφάλματα (Errors) που οδηγούν σε Απάτες (Delusions) Ατομικές-Συλλογικές

Syntactic Processing Errors	Observation Errors	Τυχαια, Ακριβειας (Accuracy) Τυπική Αποκλιση Συστηματικά, Σαφηνειας (Precision) Εντροπια
	Approximation Errors	Analytic Approximations Asymptotic Approximations
	Computation Errors (Finite Representation)	Rounding Στρογγυλευση $\pi = 3.1415926535... \approx 3.1416$ Truncation Αποκοπη $\pi \approx 3.1415926535... \approx 3.1415$
	Logical Errors	Processing Errors, Paradoxes
	Statistical Errors	Estimation-Decision Errors
		Testing Hypothesis Errors, Validation Errors
Semantic Processing Errors	Perception Errors (Σφάλματα Αντιληψης), Hallucinations	
	Deception (Αυταπατη), Cognitive Bias, Glamour,	
	Πλανε (Illusions, Misinterpretation, Ontological Errors, Λαθος Πλαισιο)	
Significance Errors	Εμμονη σε Subjective Υποθεσεις-Beliefs-Representations of Reality που είναι Εσφαλμενες	

ΣΧΟΛΙΟ 1: **Λαθος** είναι το Σφάλμα που Λανθανει της Προσοχης και για το οποιο δεν εχουμε Επιγνωση

ΣΧΟΛΙΟ 2: Τα Σφάλματα μπορεί να ωφειλονται και σε Υλικές Δυσλειτουργίες (Hardware Malfunctions), όπως Brain Deficits (Ανοια, Μωρια, Χημικές Παρεμβασεις-Αλλοιωσεις).

Slide from
Prof I. Antoniou

ISLAND SAMPLING DAY

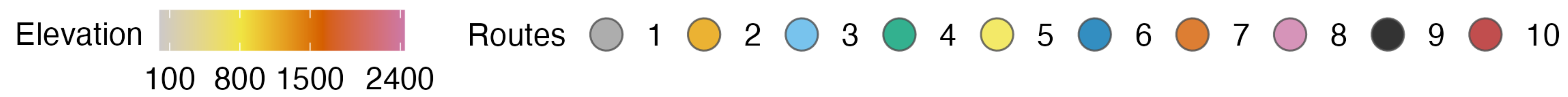
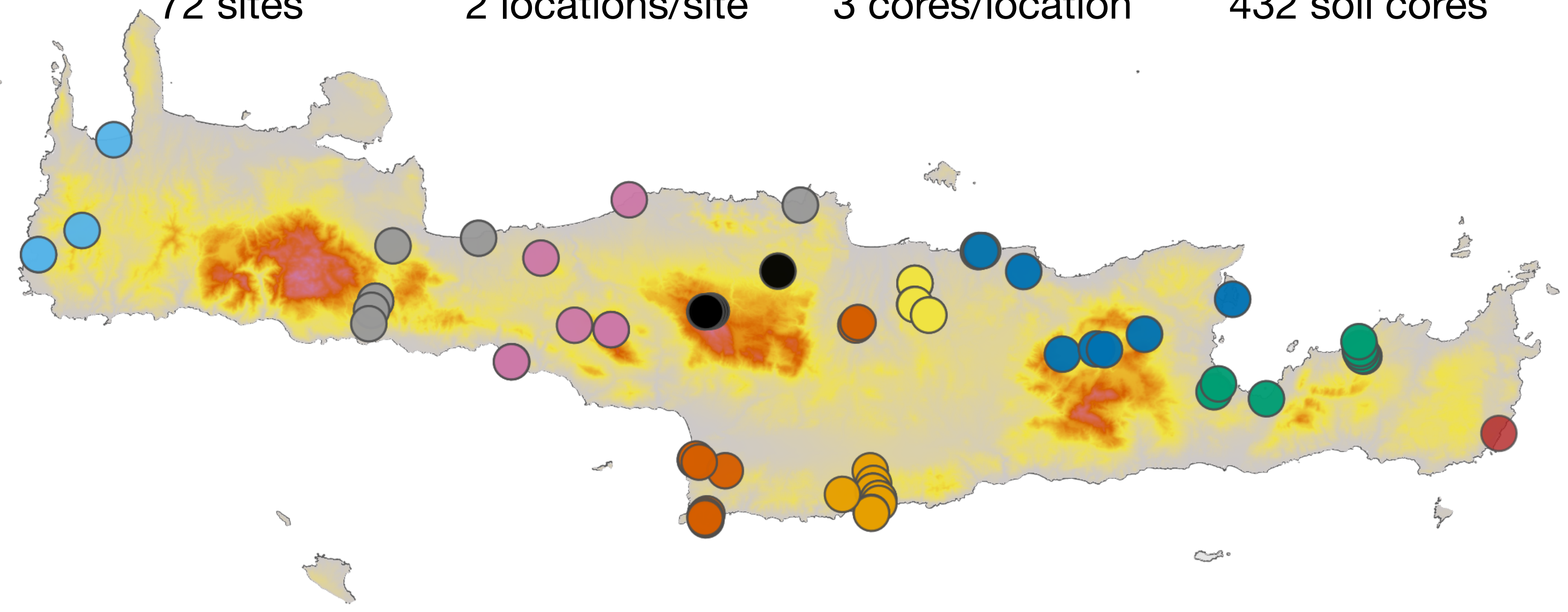
CRETE

1 day
72 sites

26 participants
2 locations/site

10 routes
3 cores/location

2400 km
432 soil cores



Island Sampling Day 2016

One Day, One Island - 15th June 2016, Crete - GSC18

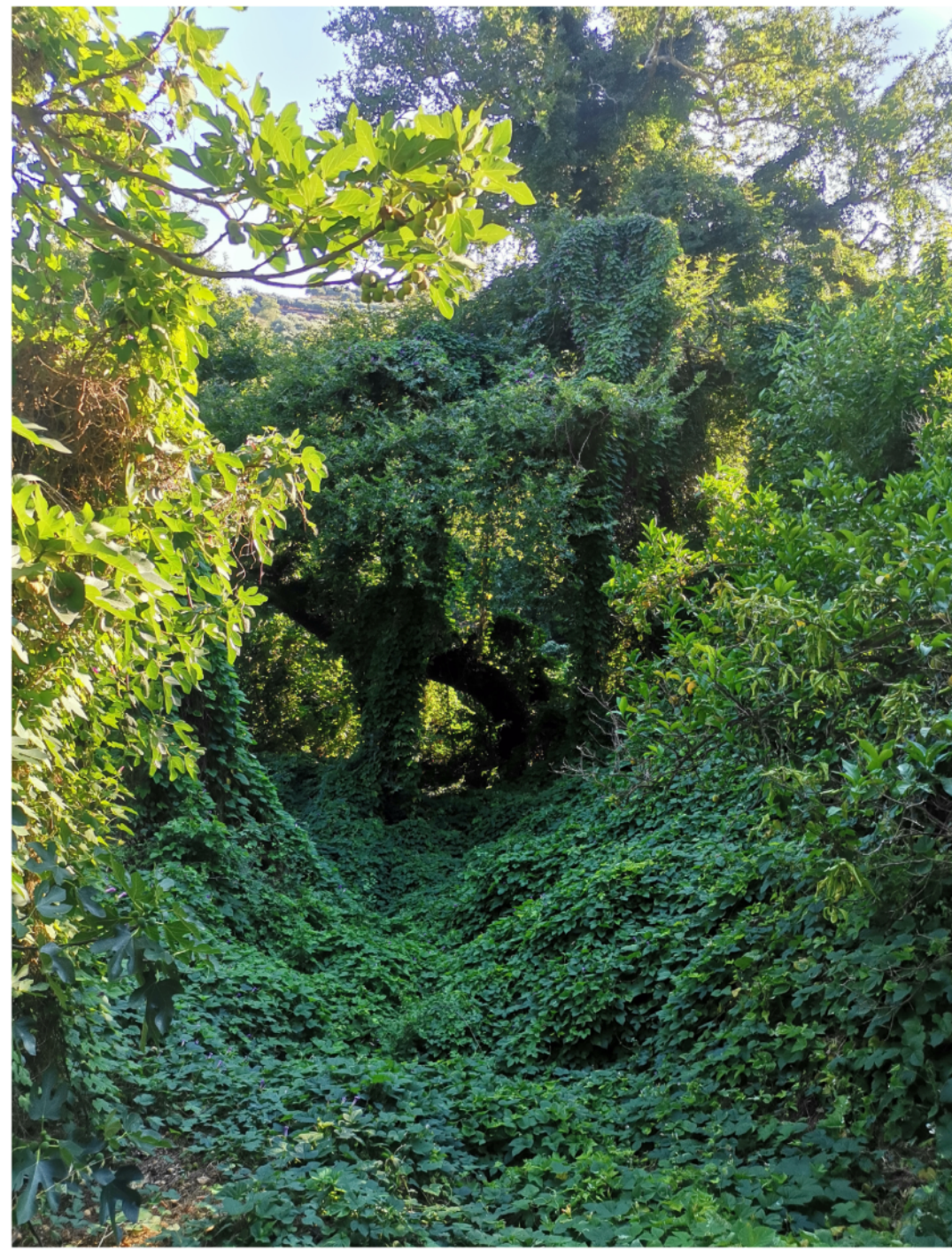


16s rRNA amplicon study

26 participants, 10 teams/routes

72 predefined locations across the island of Crete

Second time point in 2022, same locations



Sampling

Topsoil cores

72 sites (2 sub-sites in each)

Each sub-site 3 replicates

DNA - Chemistry - HCMR

Designed for Citizen science by experts

Single day event to avoid seasonality

Capture as much ecosystem diversity as possible



Sequences

- **Open data with immediate release**
- Available at ENA project PRJEB21776
- 140 samples available (Illumina HiSeq 2500) using Amplicon 16s rRNA V3V4 regions
- 3 Samples NOT sequenced
- Dr Lynn Schriml, University of Maryland

Reads	60.6 x 10 ⁶
Ns	76.000
Filtered	50.1 x 10 ⁶
denoisedF	45.9 x 10 ⁶
denoisedR	48.5 x 10 ⁶
merged	35.0 x 10 ⁶
nonchim	33.3 x 10 ⁶


Metadata

Onsite measurements
GPS coordinates
nearest plant
Etc

total nitrogen
water content
total organic carbon
pH

Dr Stephanie A Yarwood, University of
Meryland

FAIR Data by Design - Findable - Accessible
- Interoperable - Reusable

 European Nucleotide Archive	
Home	Submit ▾
Search ▾	Rulespace
About ▾	Support
Biosample: SAMEA104726343 ↗	
metagenome from Crete soil	
<hr/>	
Organism:	soil metagenome
Sample Accession:	SAMEA104726343
Sample Title:	Crete soil metagenome
Location:	35.3518526 N 24.3609229 E
Center Name:	INSTITUTE FOR GENOME SCIENCES,
Sample Alias:	3
Checklist:	ERC000022
Tag:	terrestrial terrestrial_medium_confidence metagenome datahub
Geographic Location (Depth):	5 cm
DNA Concentration:	7.81
Environment (Material):	soil
Sample Volume Or Weight For DNA Extraction:	0.2596

ISD Crete - Reproducible analysis

GET

API

sequences (fastq)
metadata (xml)



INFER



a pipeline for eDNA metabarcoding analysis

ANALYZE



Taxonomic assignment

- **DADA2**

- Taxonomy - Silva 138
- ASVs = 239000
- Taxa = 3102

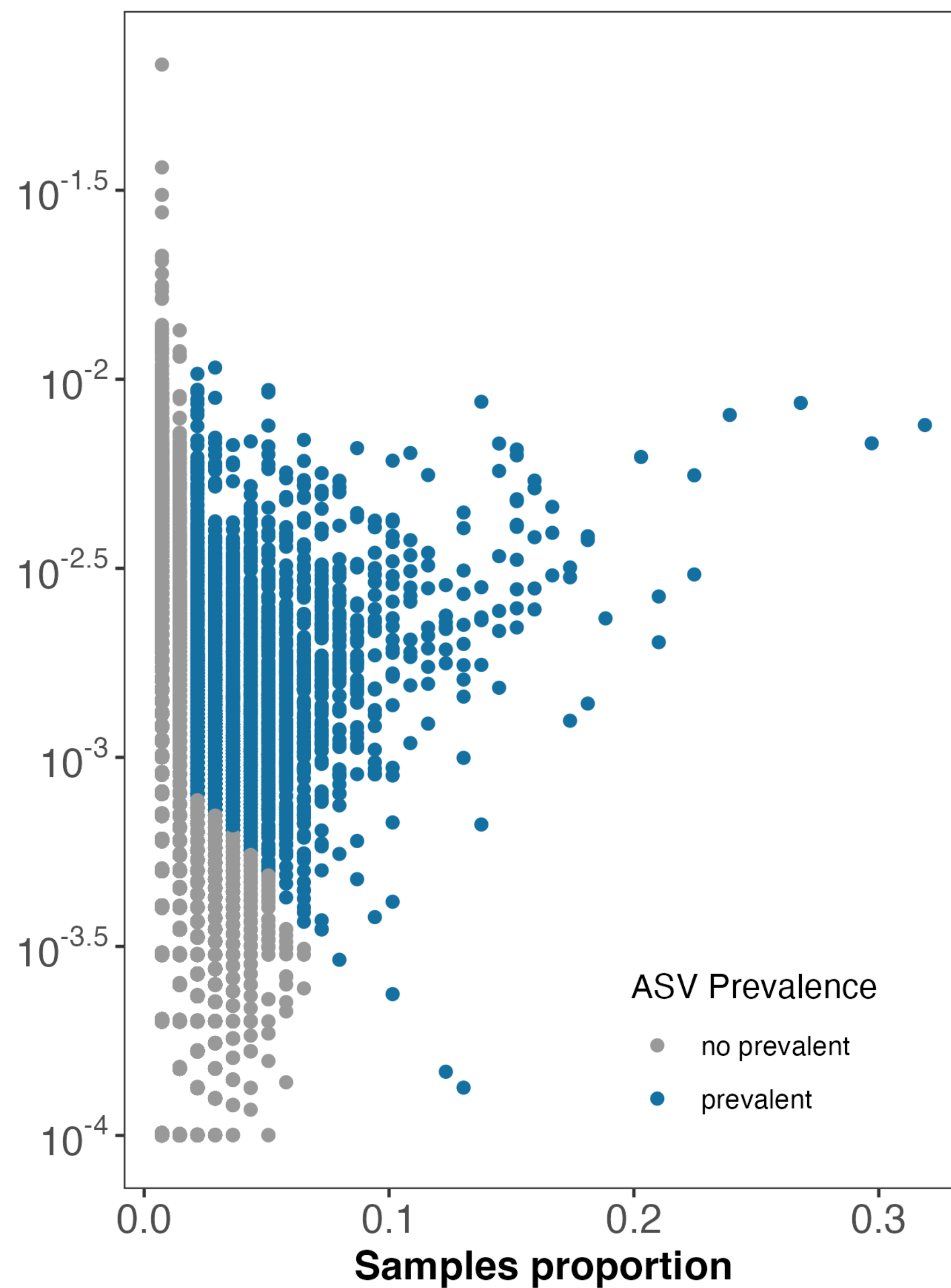
- **PEMA (VSEARCH)**

- Taxonomy - Silva 132
- OTUS 6890
- Taxa = 1057

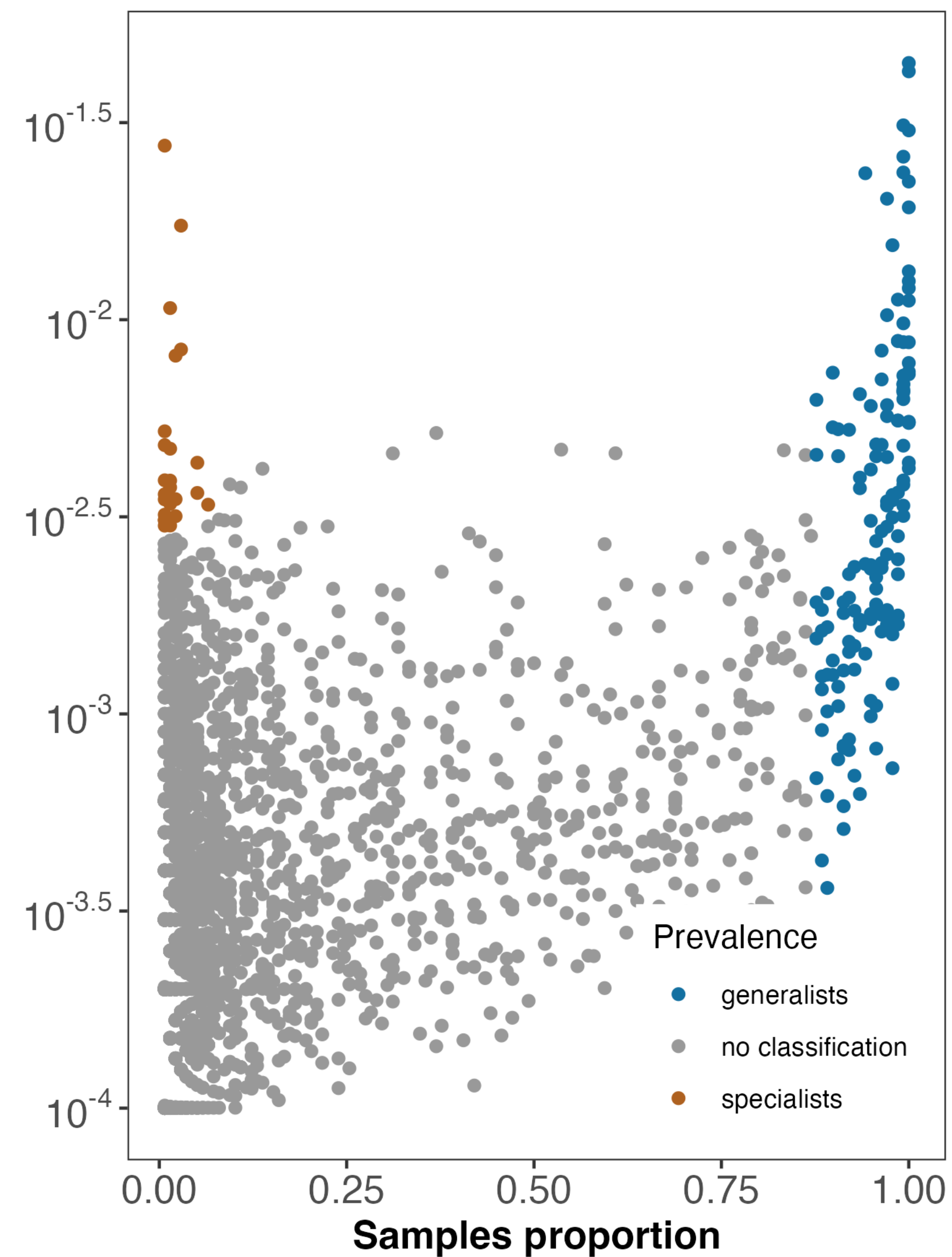
Classification depth	DADA2		PEMA	
	ASVs	Taxa	OTUs	Taxa
Kingdom	1974	2	284	2
Phylum	4034	33	121	15
Order	38517	193	1224	135
Class	24157	83	978	62
Family	71355	287	2319	218
Genus	90137	1166	1920	582
Species	9120	1338	44	43
Total	~ 239000	3102	6890	1057

A

Mean relative abundance

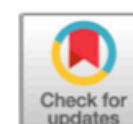
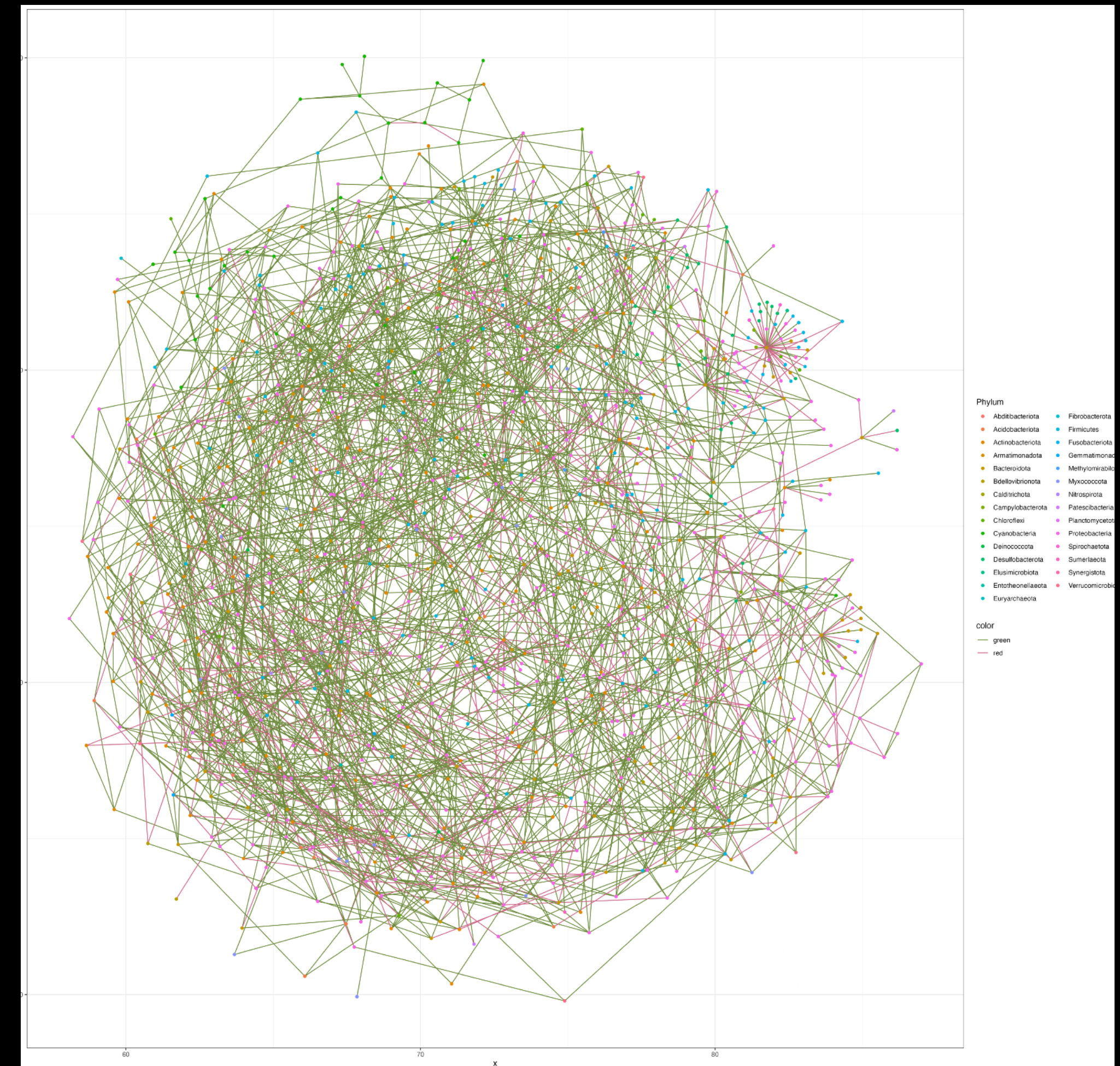
**B**

Mean relative abundance



Networks

- Network analysis (FlashWeave v0.19.2)
- Focus on Richtis gorge



Computational Biology | Research Article

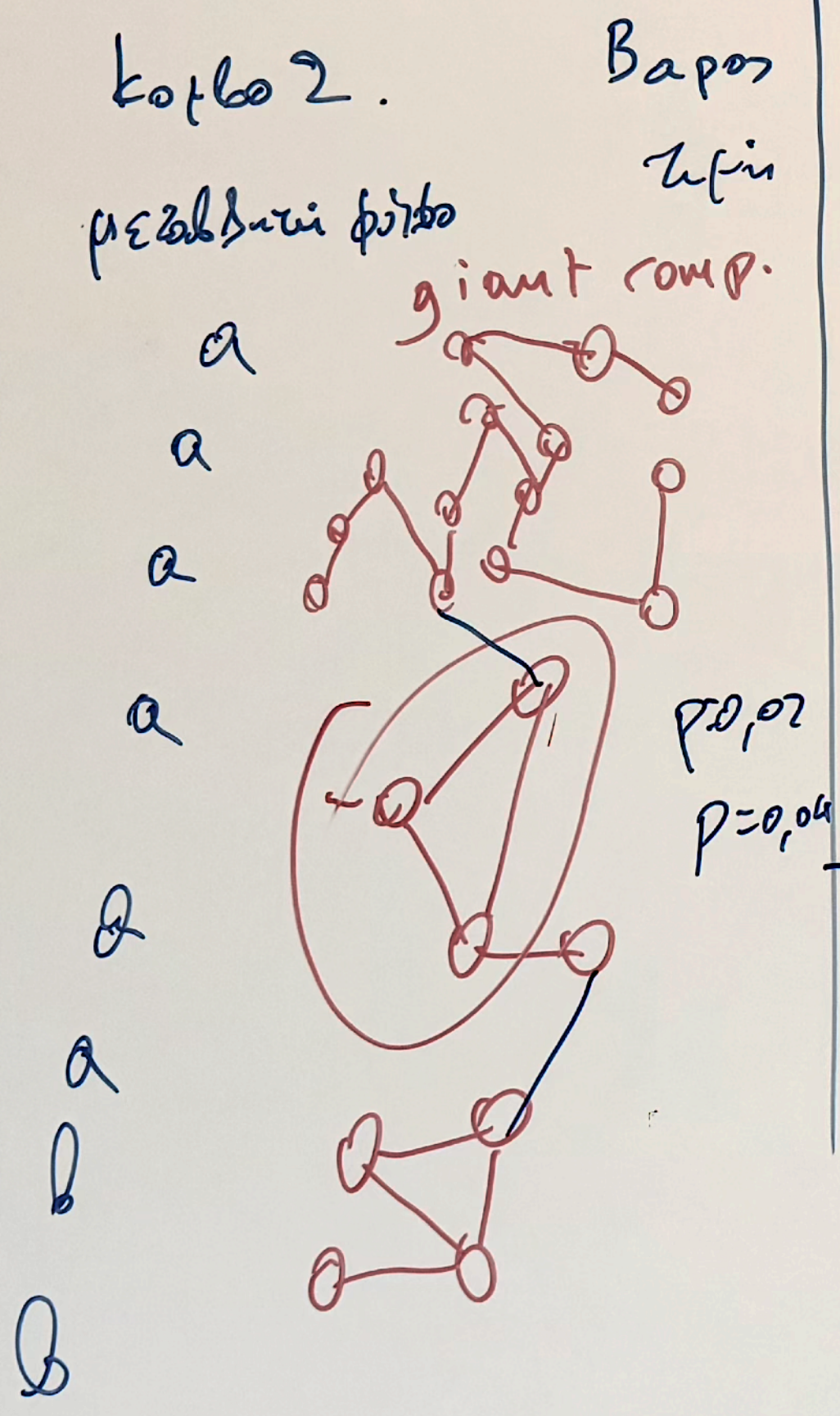
Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation

Dileep Kishore,^{1,2,3} Gabriel Birzu,^{4,5} Zhenjun Hu,¹ Charles DeLisi,^{1,4,6} Kirill S. Korolev,^{1,2,4} Daniel Segre^{1,2,4,6,7}

AUTHOR AFFILIATIONS See affiliation list on p. 25.

κομβος 1
 είδος ορ.
 1
 1
 1
 2
 2
 3
 3
 3

Edgelist long format



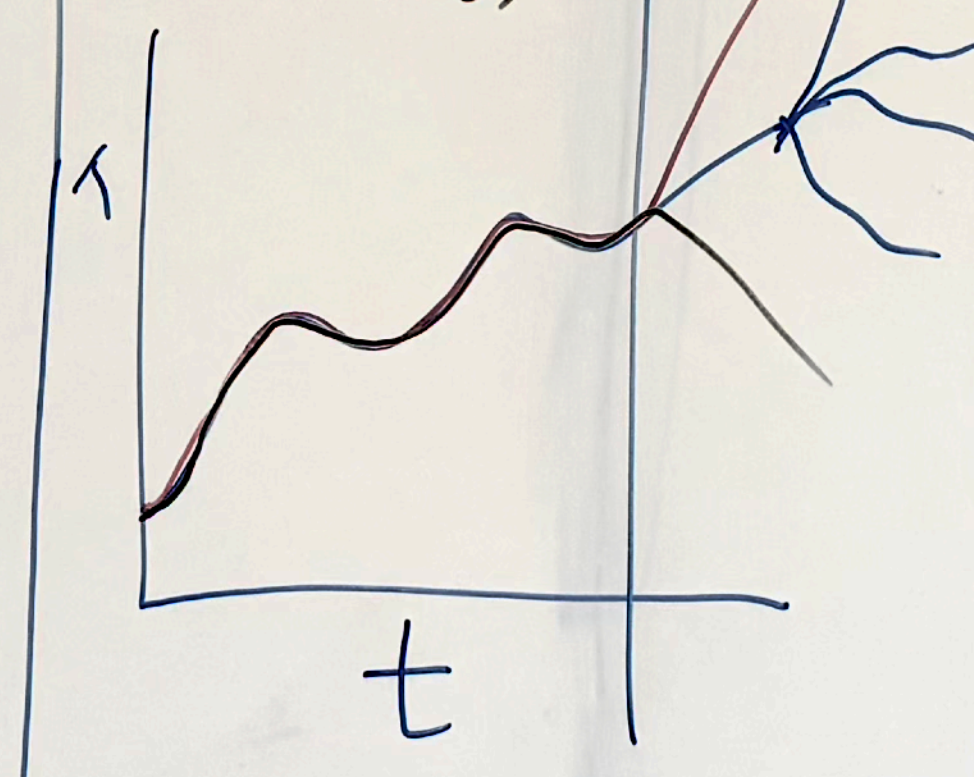
Wide

Μεταβλητή/Sample

είδος	a	b	c	d
1	1			
2		1		
3			1	
				1

1 2 3 4

Ευαίσθητα ορική ανάλυση
 Χαος



Bioinformatics
 Κατηγορίες εργασιών
 σε Metagenomics

Ποιο είναι το ερώτημα

TL	Quil>
Flan	
Πω>	

$$g(x) = f(x) + p$$

