

MB590-012
Microbiome Analysis
Core Microbiomes

Dr. Christine Hawkes

NC STATE UNIVERSITY


Final project - SRA data

- Rachel has posted two GitHub issues with helpful info on
 - Converting .sra -> .fastq
 - Transferring sra files that downloaded in individual folders
- Upcoming deadlines
 - Optional: can use your project data in class on 3/9 for exploratory analysis practicum
 - Presentations 4/20
 - Final Rmd and knitted file 4/27



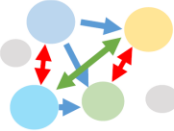
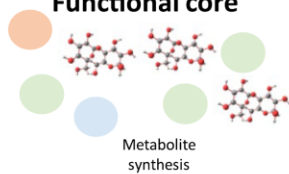
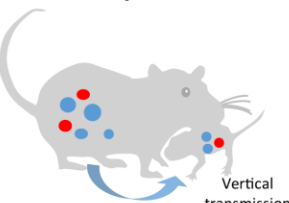
Best practices for learning R

- Look at the usage and arguments for R functions
 - <https://www.rdocumentation.org/>
 - <https://rdr.io/>
 - Package CRAN manuals, vignettes, Github pages
- Example: `vegan::envfit`
 - <https://www.rdocumentation.org/packages/vegan/versions/2.4-2/topics/envfit>
 - <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Example: microbiome package
 - <https://rdr.io/github/microbiome/microbiome/man/>

What is a core microbiome?

- Simplest definition: taxa (OTUs/ASVs) that are shared by two or more samples in the same host, site, ecosystem, or treatment (or other group)
 - Identification of core members
 - Occupancy
 - Abundance
 - Both occupancy and abundance
 - Stability over time → Temporal core
 - Network structure → Ecological core
 - Function → Functional core
 - Adaptation → Host-adapted core
- 
- Common core

What is a core microbiome?

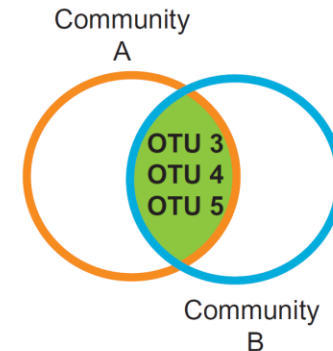
Term	Definition	Criteria
Common core 	<p>The component of the microbiome that is found across a considerable proportion of hosts within a defined host population or species</p>	<ul style="list-style-type: none"> • High prevalence/occupancy frequency across host population/species • Can be identified using occupancy-abundance curves • (Optional) Common in host species of interest but not in other closely related species • Rare (non-prevalent) taxa cannot be core
Temporal core 	<p>A temporally stable or predictable component of the microbiota</p>	<ul style="list-style-type: none"> • Taxa that demonstrate stable or predictable dynamics over time, either within a single host or across host population/species • Within individuals, rare (non-prevalent) taxa can be core
Ecological core 	<p>The component of the microbiome that is disproportionately important for shaping the organisation and diversity of the ecological community</p>	<ul style="list-style-type: none"> • Removal or introduction results in large cascading effects on ecological structure and diversity • May form interaction hubs in ecological networks • May increase community stability • Rare (non-prevalent) taxa can be core (e.g. predators or ecosystem engineers)
Functional core 	<p>The component of the microbiome that performs essential biological functions to the host, usually in respect to their biochemical, physiological or ecological services to the host</p>	<ul style="list-style-type: none"> • A set of genes or taxa that are linked to a measureable facet of host function • Natural variation in host function does not affect host fitness OR • Natural variation in function does affect host fitness but phylogenetically distinct taxa can perform function • Likely to represent facultative symbionts • Can be horizontally or vertically acquired • Rare (non-prevalent) taxa can be core
Host-adapted core 	<p>A set of microbes that has co-evolved with the host species or sub-population and whose presence increases host fitness in at least some ecological contexts</p>	<ul style="list-style-type: none"> • Taxa that are linked to a measureable facet of host function • Natural variation in host function affects host fitness in at least some ecological contexts • Are not functionally redundant (other taxa cannot perform same function) • Are expected to be vertically transmitted • Likely to represent obligate or near-obligate symbionts • Very rare (non-prevalent) taxa unlikely to be core, but host-adapted cores be restricted to certain populations or ecological conditions

Common core microbiome

Occupancy (“Prevalence”)

shared OTU occurrences across communities
1 = present, 0 = below detection

List of observed OTUs		Occurrences in community A	Occurrences in community B	Shared occurrences A & B
	OTU 1	1	0	
	OTU 2	0	1	
	OTU 3	1	1	X ●
	OTU 4	1	1	X ●
	OTU 5	1	1	X ●

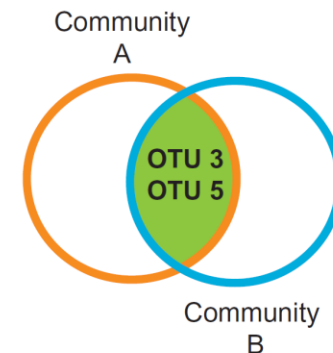


Core = 3 shared OTUs

Relative Abundance (“Detection”)

similar OTU abundances across communities

List of observed OTUs		Abundances community A	Abundances community B	Similar abundances A & B
	OTU 1	0.4	0	
	OTU 2	0	0.1	
	OTU 3	0.1	0.1	X ●
	OTU 4	0.2	0.5	
	OTU 5	0.3	0.3	X ●



Core = 2 shared OTUs

Note: Rel abund used as a threshold

Common core microbiome

`microbiome::core(ps_object,`

`detection=x/100,`
`prevalence=y/100)`

Rel abund threshold

0/100 = any

1/100 = >1%

0.1/100 = >0.1%

Returns a filtered ps object

Occupancy threshold

99/100 = >99% of samples

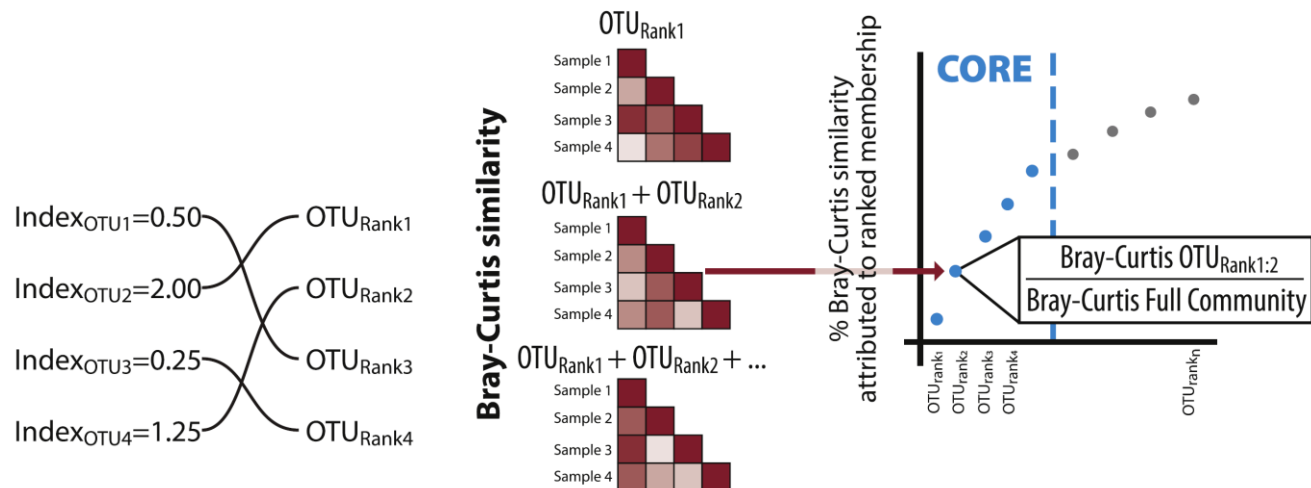
50/100 = >50% of samples

Common core microbiome: abundance-occupancy

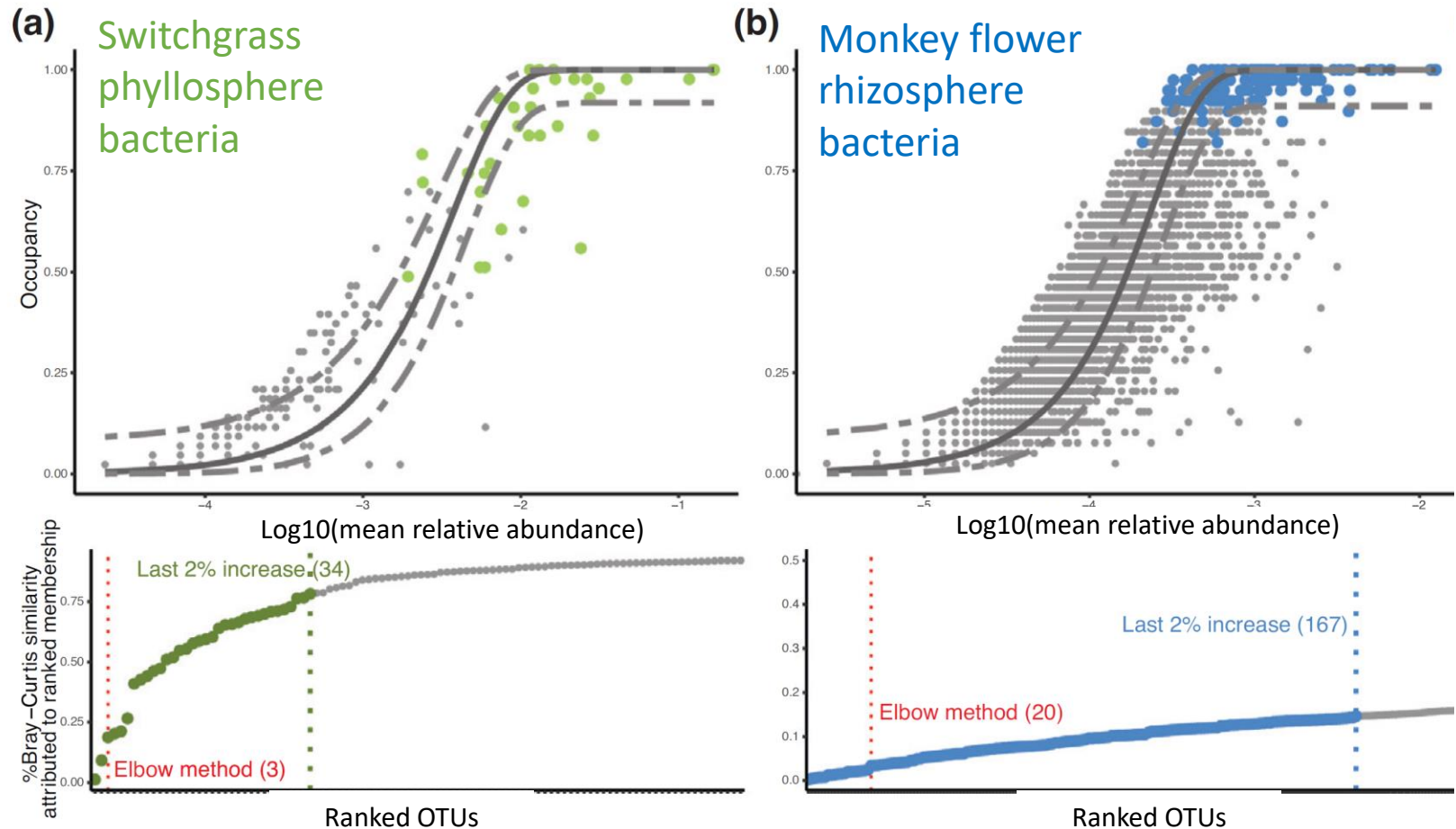
Abundance	Site 1		Site 2		Site-specific occupancy	Site 1 Site 2		Replication consistency	Site 1 Site 2	
	Rep1	Rep2	Rep1	Rep2		OTU1	OTU2		OTU1	OTU2
	OTU1	11	0	0	13	0.5	0.5		0	0
	OTU2	10	7	15	40	1	1		1	1
	OTU3	0	0	0	32	0	0.5		0	0
	OTU4	17	4	0	23	1	0.5		1	0
	OTUn	65	9	24	18	1	1		1	1

$$\text{Index}_{\text{OTU}_i} = \frac{\sum (\text{Site-specific occupancy}) + \sum (\text{Replication consistency})}{2}$$

Identification of Core Taxa: uses contributions to BC similarity of total community after ranking by index



Common core microbiome: abundance-occupancy

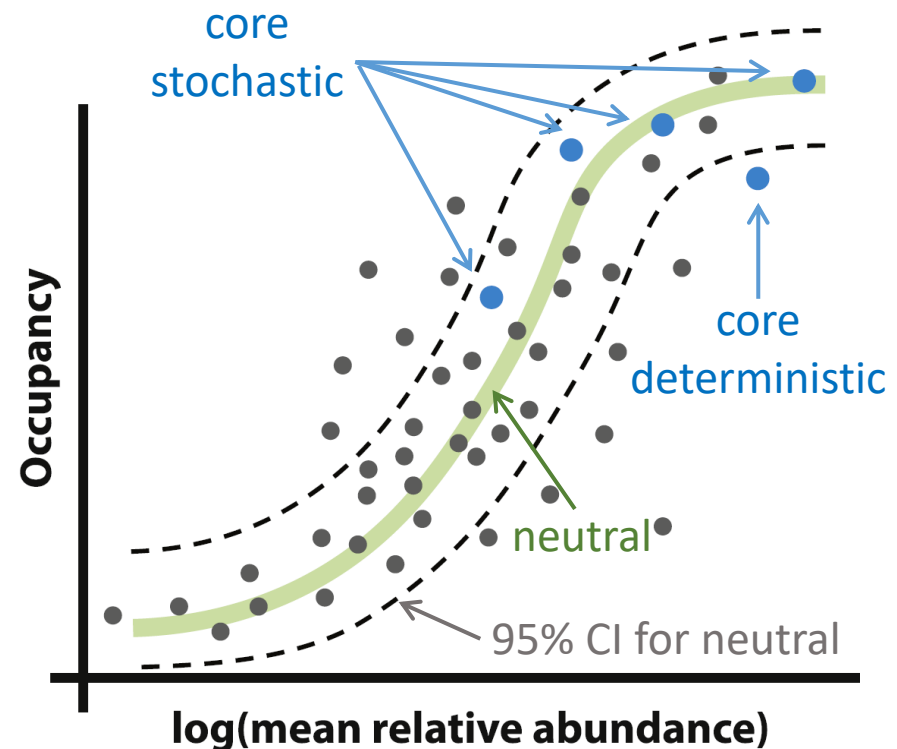


Cumulative method: keep adding taxa until increase in BC similarity explained is <2%

Elbow method: find breakpoint in graph that maximizes slope differences in each curve

Common core microbiome: neutral models and abundance-occupancy

- Fit neutral model to abundance-occupancy distribution
- Allows for identification of core members that are
 - consistent with neutral model = stochastic assembly
 - outside neutral model = deterministic assembly



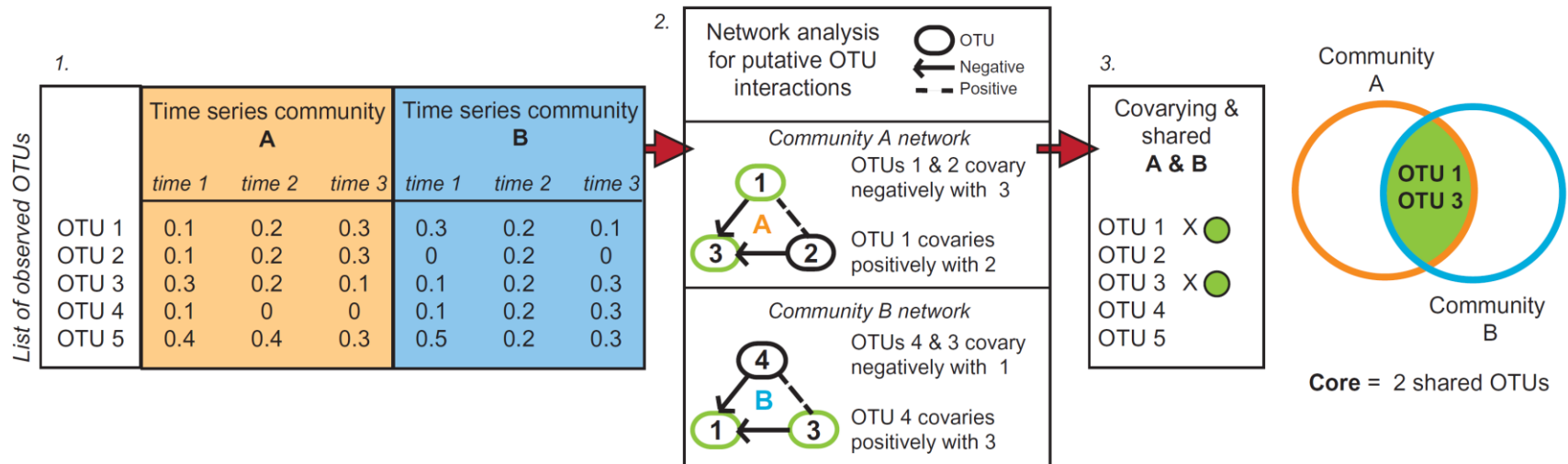
Common core: abundance & occupancy

	Pros	Cons
Occurrence only	<ul style="list-style-type: none">• computationally simple• commonly used• includes rare taxa	<ul style="list-style-type: none">• arbitrary cutoffs• affected by seq depth• no abundance info
Abundance only	<ul style="list-style-type: none">• computationally simple• abundance may better reflect function	<ul style="list-style-type: none">• arbitrary cutoffs• affected by seq depth• affected by low sample nums• affected by standardization/rarefaction
Occupancy-abundance curves	<ul style="list-style-type: none">• based on macroecological theory• potentially differentiates between deterministic and stochastic core members	<ul style="list-style-type: none">• arbitrary cutoffs• assumes microbial interactions and functions are similar to animals/plants

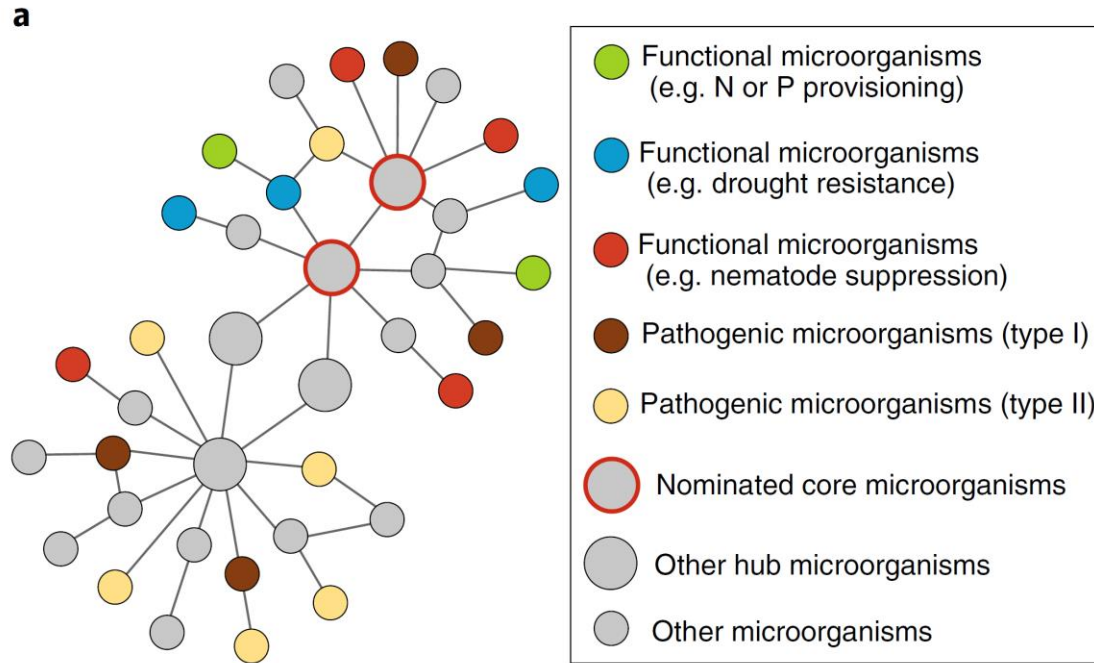
Ecological core microbiome: network connectivity

Network Connectivity

OTUs co-varying within a community and shared across communities



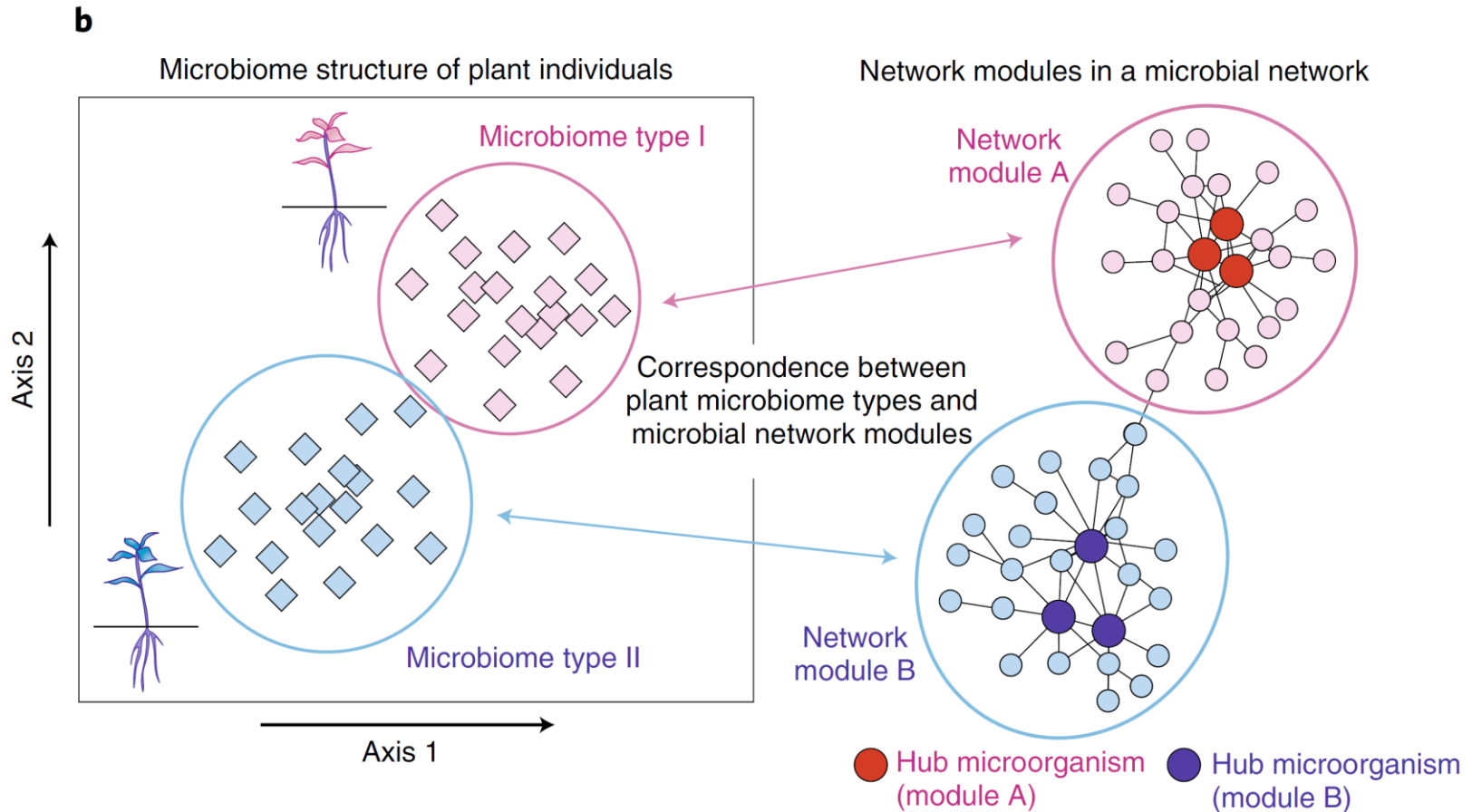
Ecological core microbiome: network hubs as candidate core members



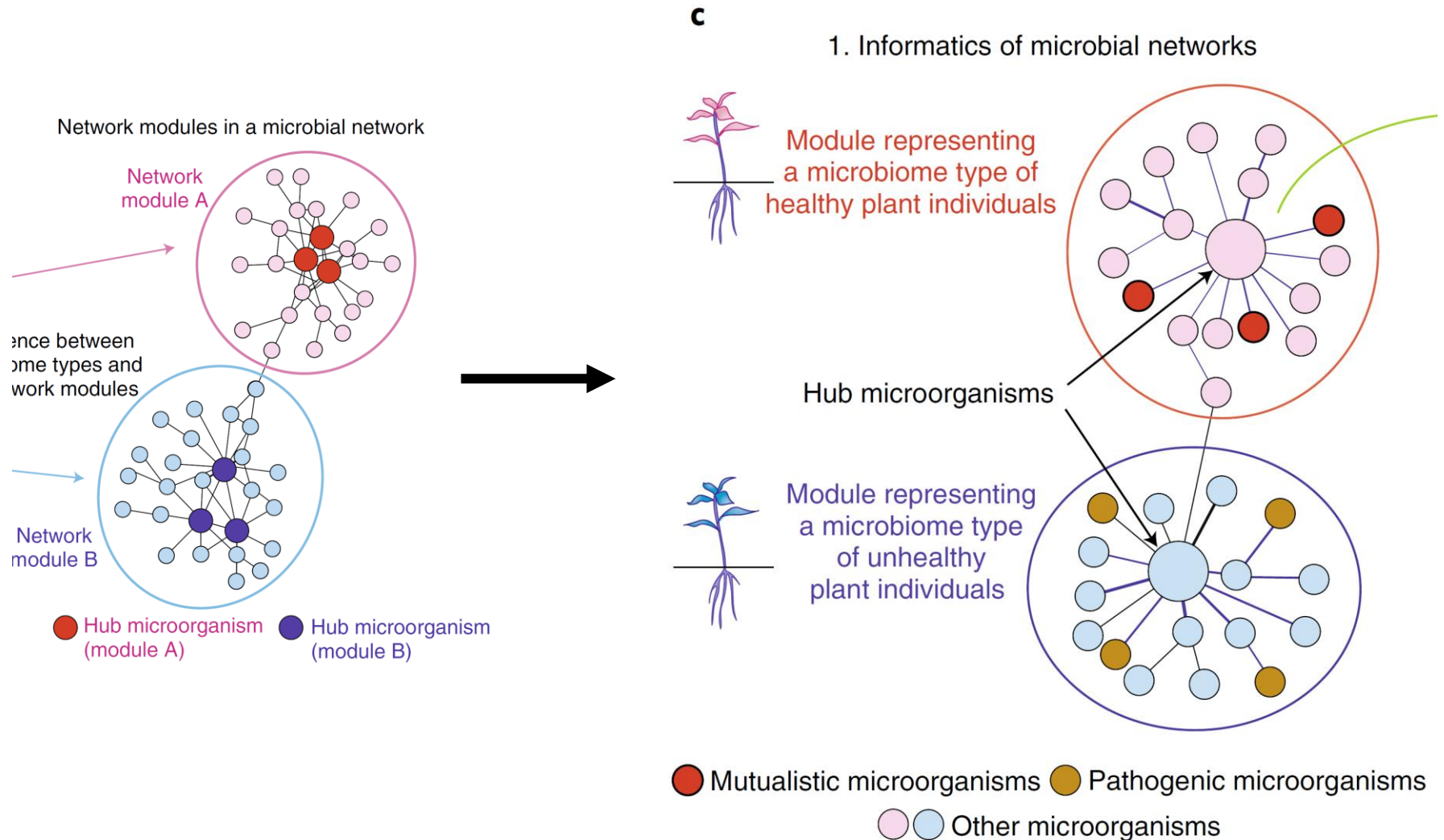
Hub taxa: defined from network topology as either (1) nodes with higher than average connectivity (= high coexistence with other taxa) or (2) nodes that bridge between modules; often considered “keystone” taxa

Module: section of network with high density of connections and sparse connections to other such groups

Ecological core microbiome: assume connections=interactions



Ecological core microbiome: further work to confirm interactions and functions



Problems with core microbiome

- Taxa found in all conditions may be generalists and/or good dispersers
- High abundance/prevalence does not necessarily equate with function
- Focus on shared taxa in specific hosts or habitats ignores broader microbial distributions/ecology

Today's practical coding

- Download `wk8_MicAnal_coremic.html`
- Selection of core taxa via occupancy/prevalence and abundance/detection using `microbiome::core`
- Example of how to build an abundance-occupancy curve
- Data from Oono et al. 2017 on leaf fungi
- Exercises: split and transform data, compare various settings for core definitions, identify core for `soilrep` data

