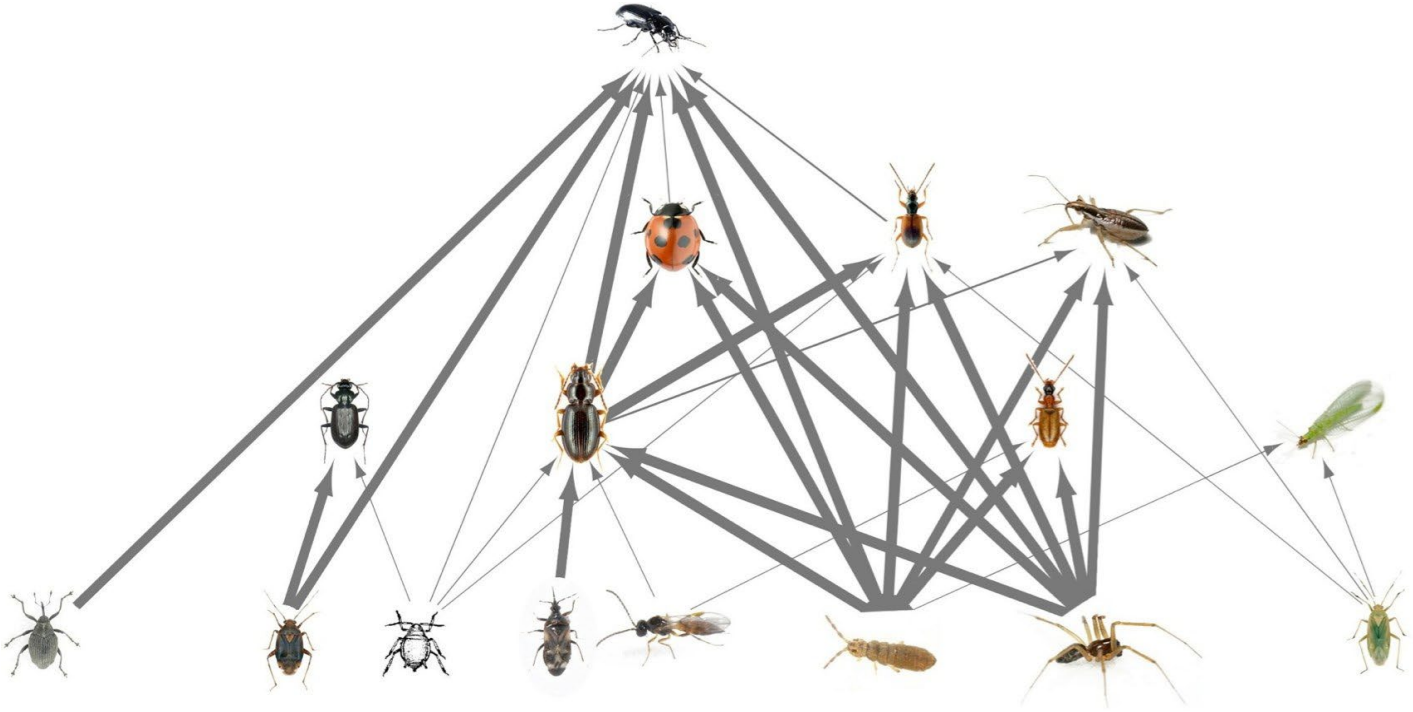
Exploring the web of dark interactions



David A. Bohan (David.Bohan@inrae.fr)
Directeur de Recherche, INRAE UMR Agroécologie, Centre Bourgogne Franche- Comté, Dijon.

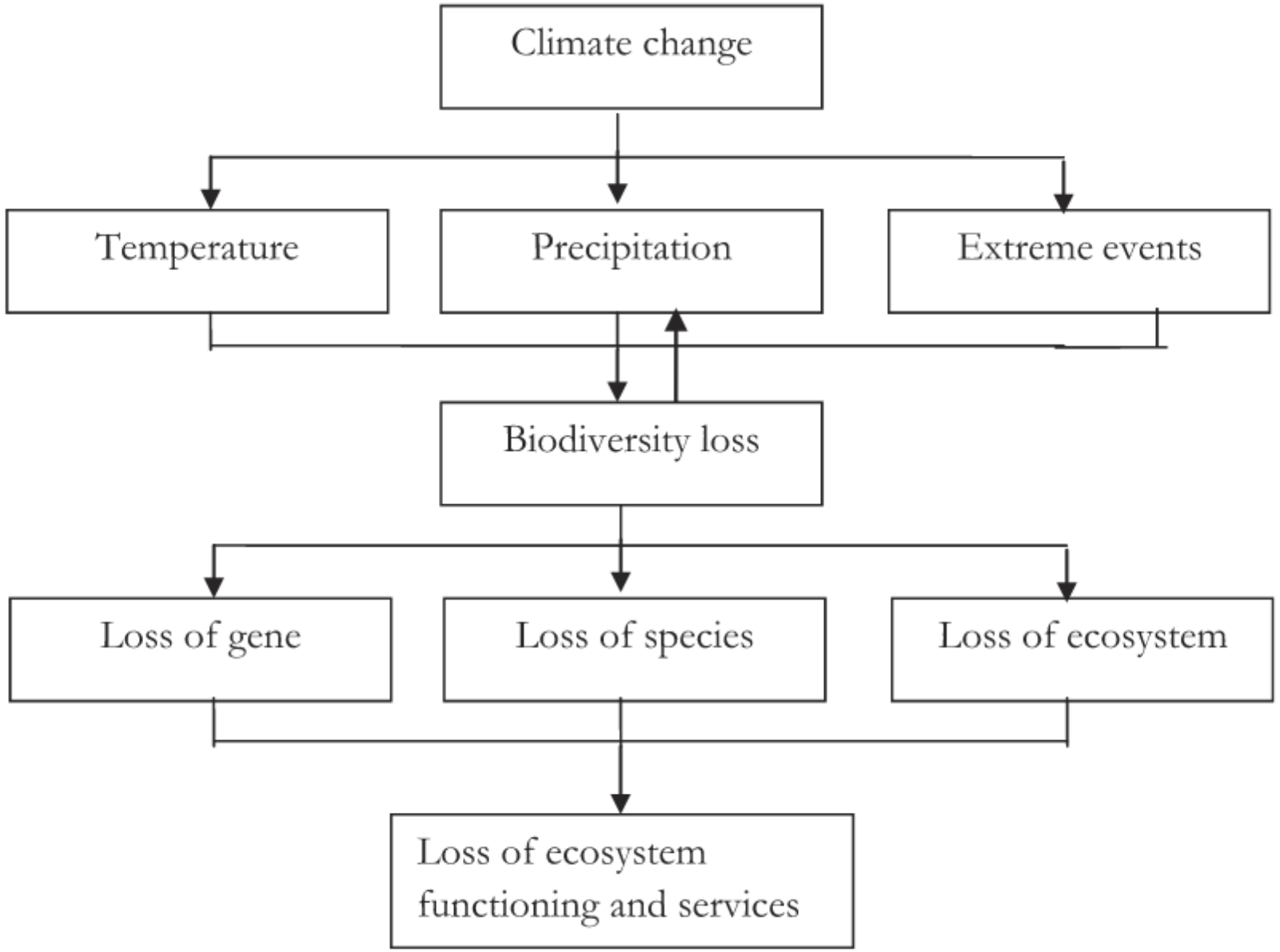
InfIntE: Inference of Interactions using Explainable machine learning

Count data by sample (OTU Table)

	a	...	v	w	x	y	z
OTU1	7	...	25	0	113	50	0
OTU2	55	...	33	0	0	22	87
OTU3	40	...	19	1	4	0	12
OUT4	0	...	155	78	0	47	92
OTU5	27	...	0	3	14	9	0
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
OTUX	110	...	78	5	9	17	39

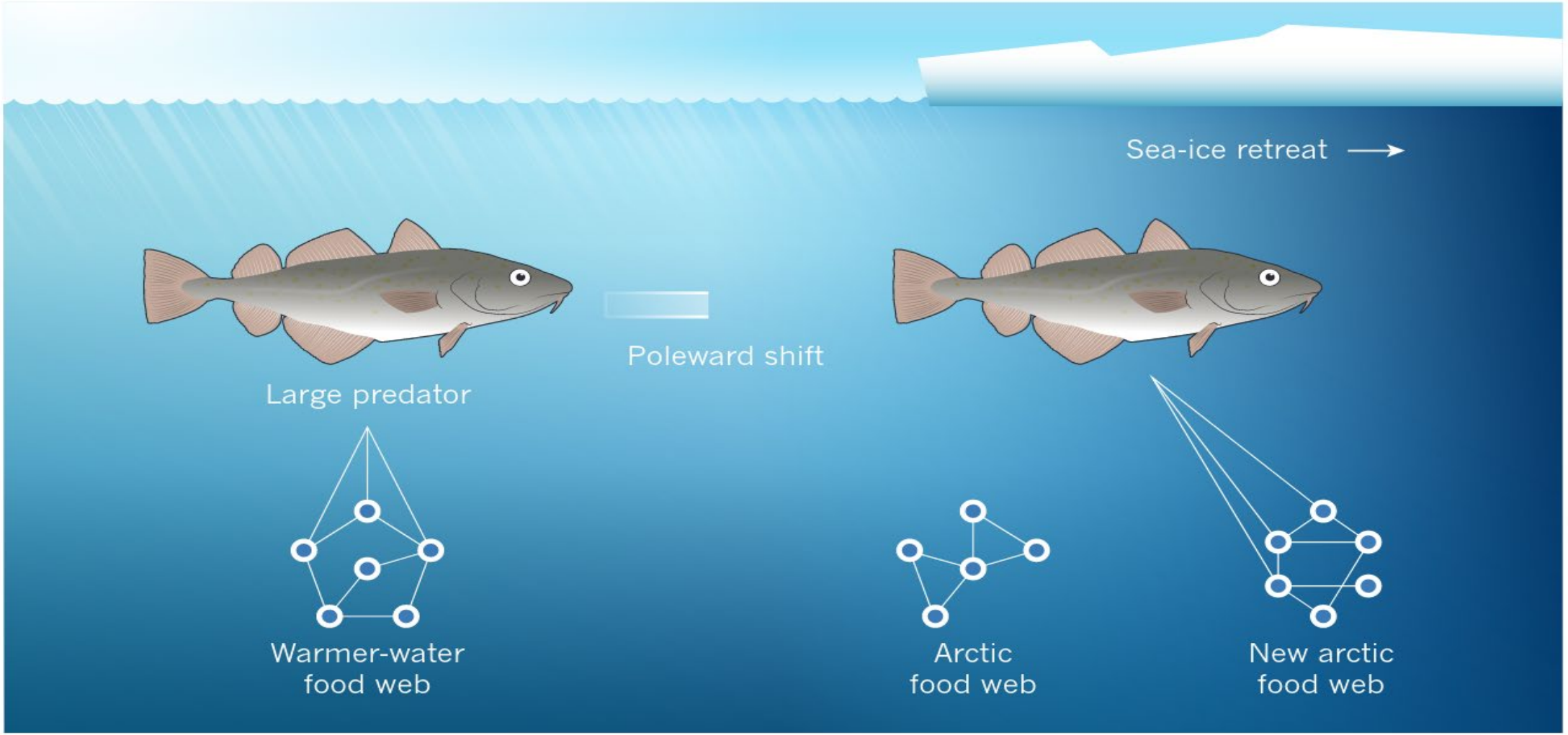
Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

Our world is changing



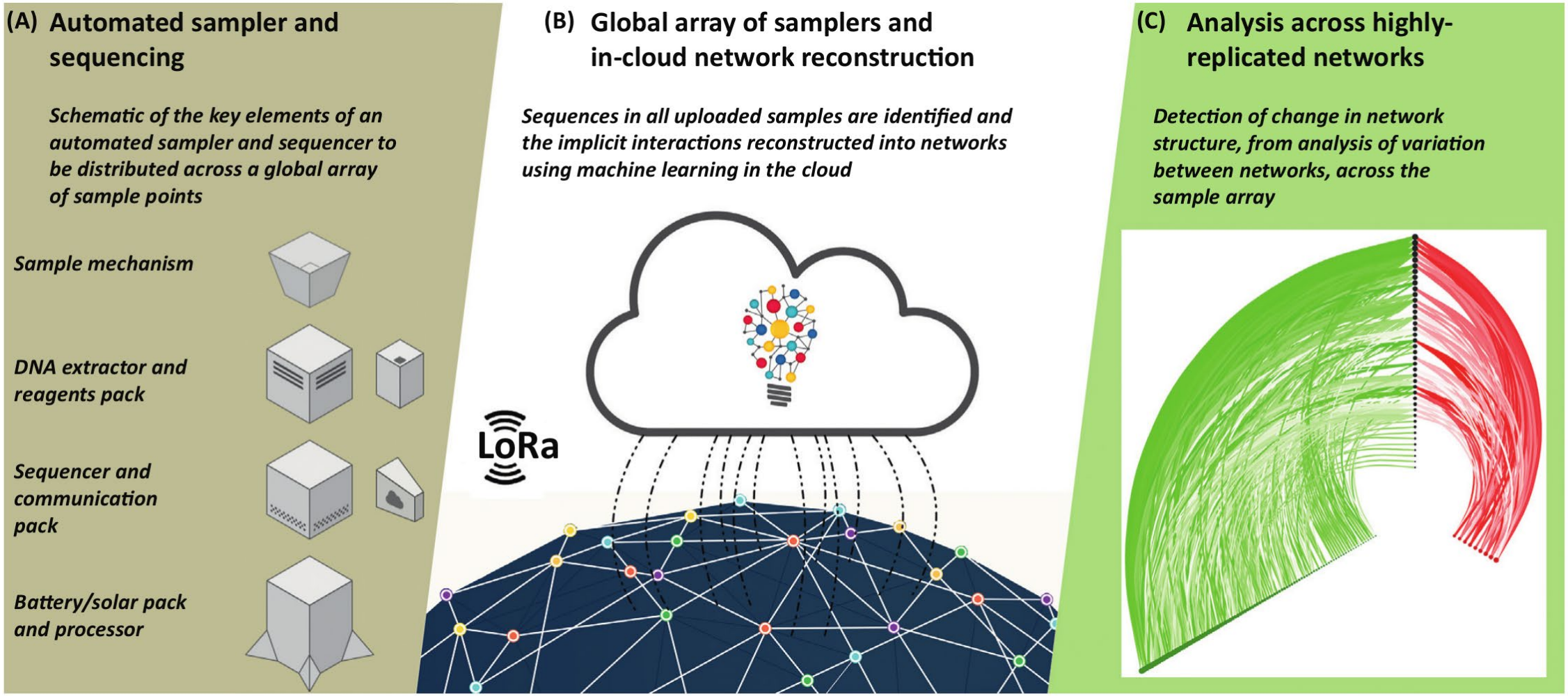
Sintayehu (2018) Impact of climate change on biodiversity and associated key ecosystem services in Africa: a systematic review. Ecosystem Health and Sustainability

Why study interactions?



Blanchard et al. 2015. A rewired food web. Nature (inspired in Kortsch et al. 2015)

The Next-Generation Global Biomonitoring Project



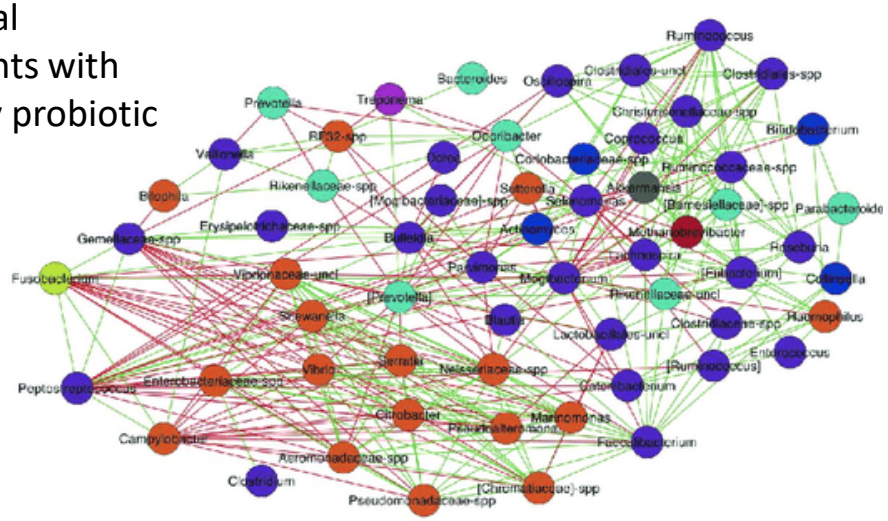
Implement Global Biosurveillance

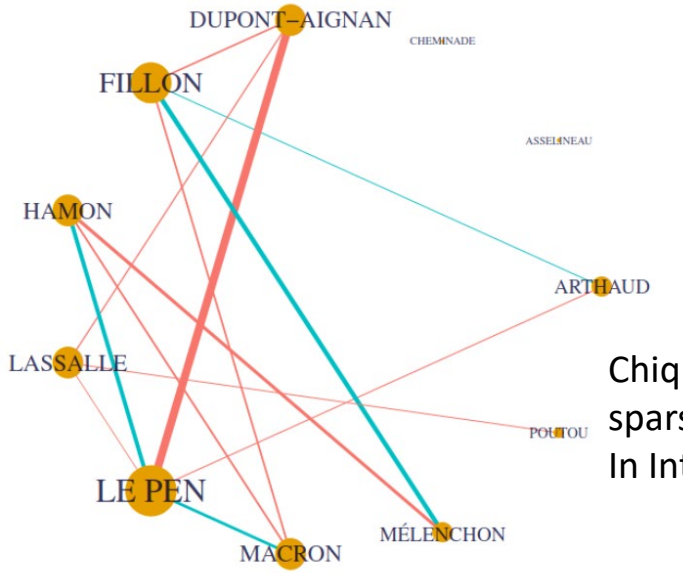
The ultimate goal of all BIOSCAN research is to lay the foundation for a comprehensive global biosurveillance system, providing real-time insight into shifts in biodiversity patterns. Effective global monitoring of biodiversity could help us to understand the shifting distributions and abundances of species, and to reduce the environmental impact of human activities.

Bohan et al. 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution.*

Interaction networks

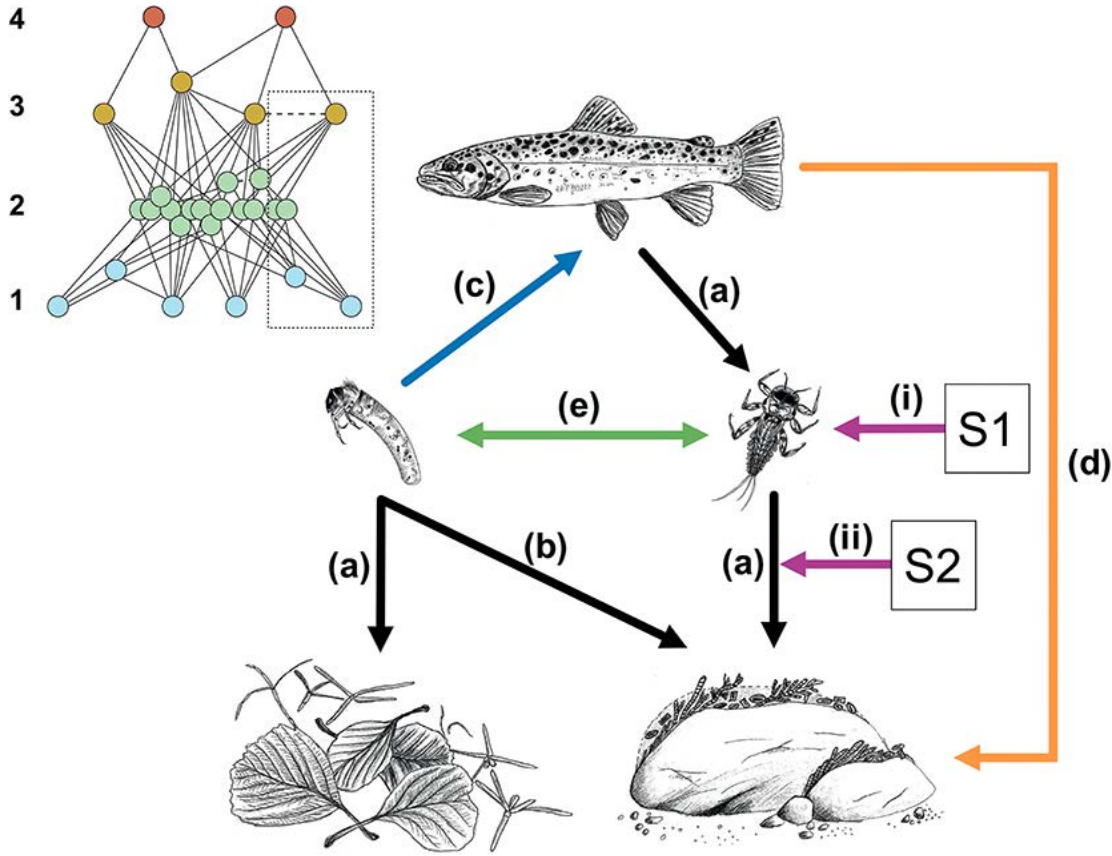
Hibberd et al. (2017) Intestinal microbiota is altered in patients with colon cancer and modified by probiotic intervention. BMJ Open Gastroenterology.





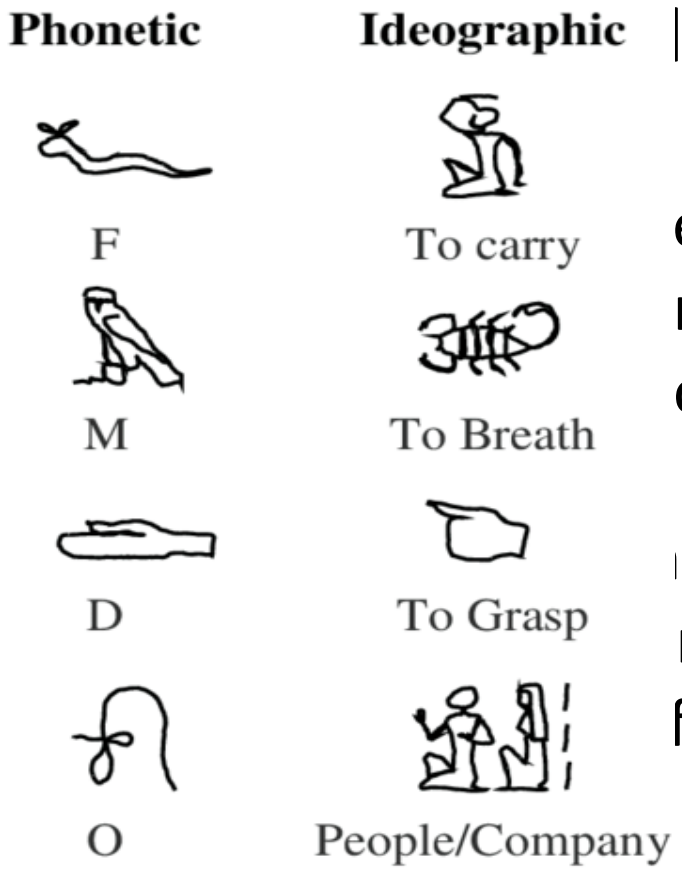
Replication

Chiquet et al. (2019) Variational inference for sparse network reconstruction from count data. In International Conference on Machine Learning



Bruder et al. (2019) The Importance of Ecological Networks in Multiple-Stressor Research and Management Front. Environ. Sci.

Two different strategies to reconstruction: correlation and logic



Jean-François Champollion (1790-1832)

Logic translation

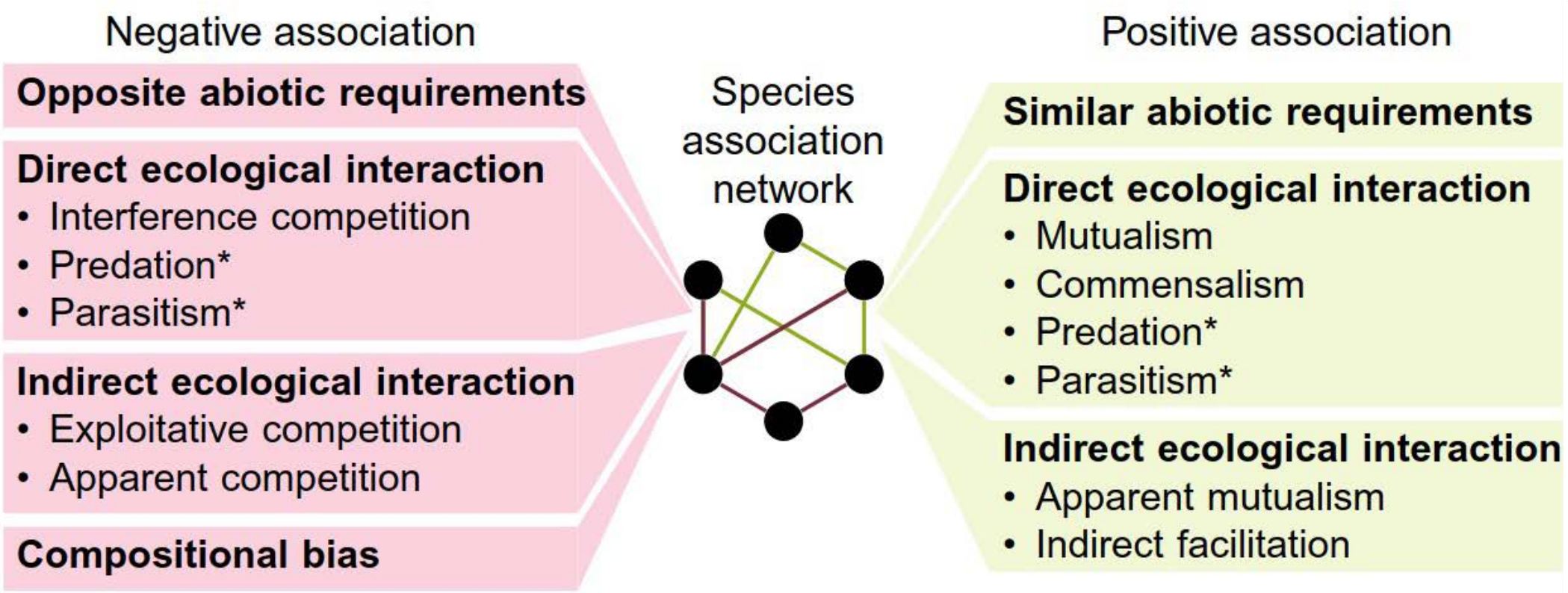
- Each sentence has a subject, object, nouns, and verbs...
- Language structure is used for the translation
- Uses previous knowledge to make an explainable translation

Translation

een the positions and
rds might allow to
oglyphs and Demotic
d useful
rmation (explanation)
for the correlations

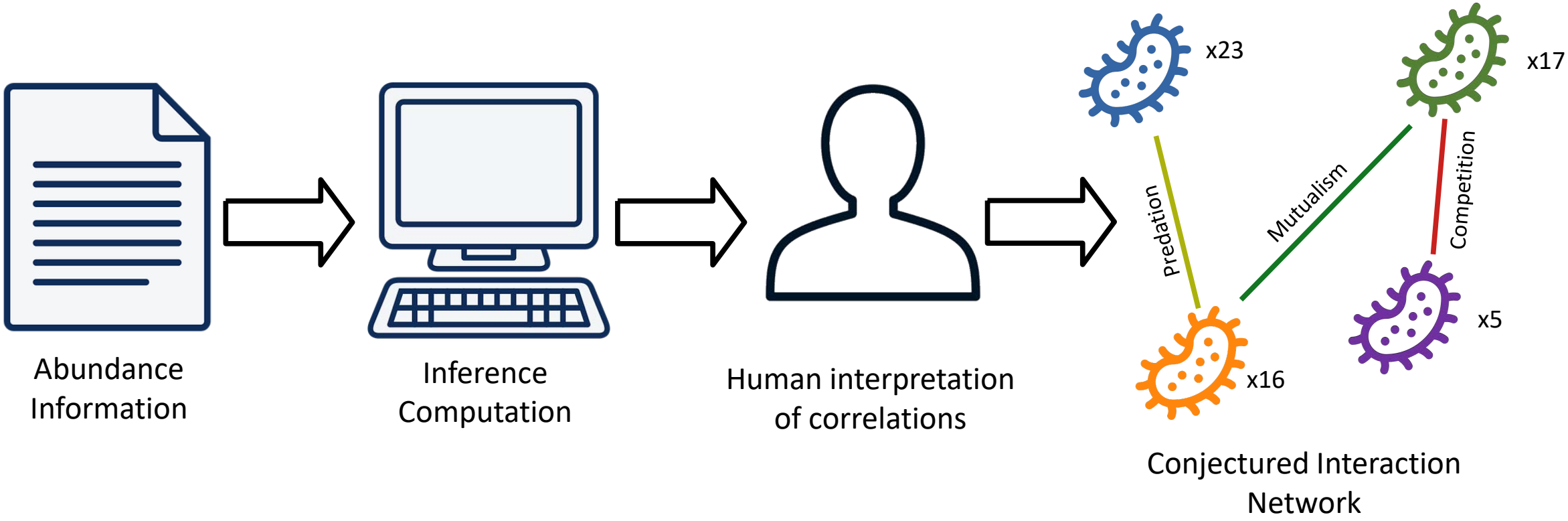
Elliot. (2017). Visual Analytics for Management: Translational Science and Applications in Practice.

What is the difference between and association and an interaction?

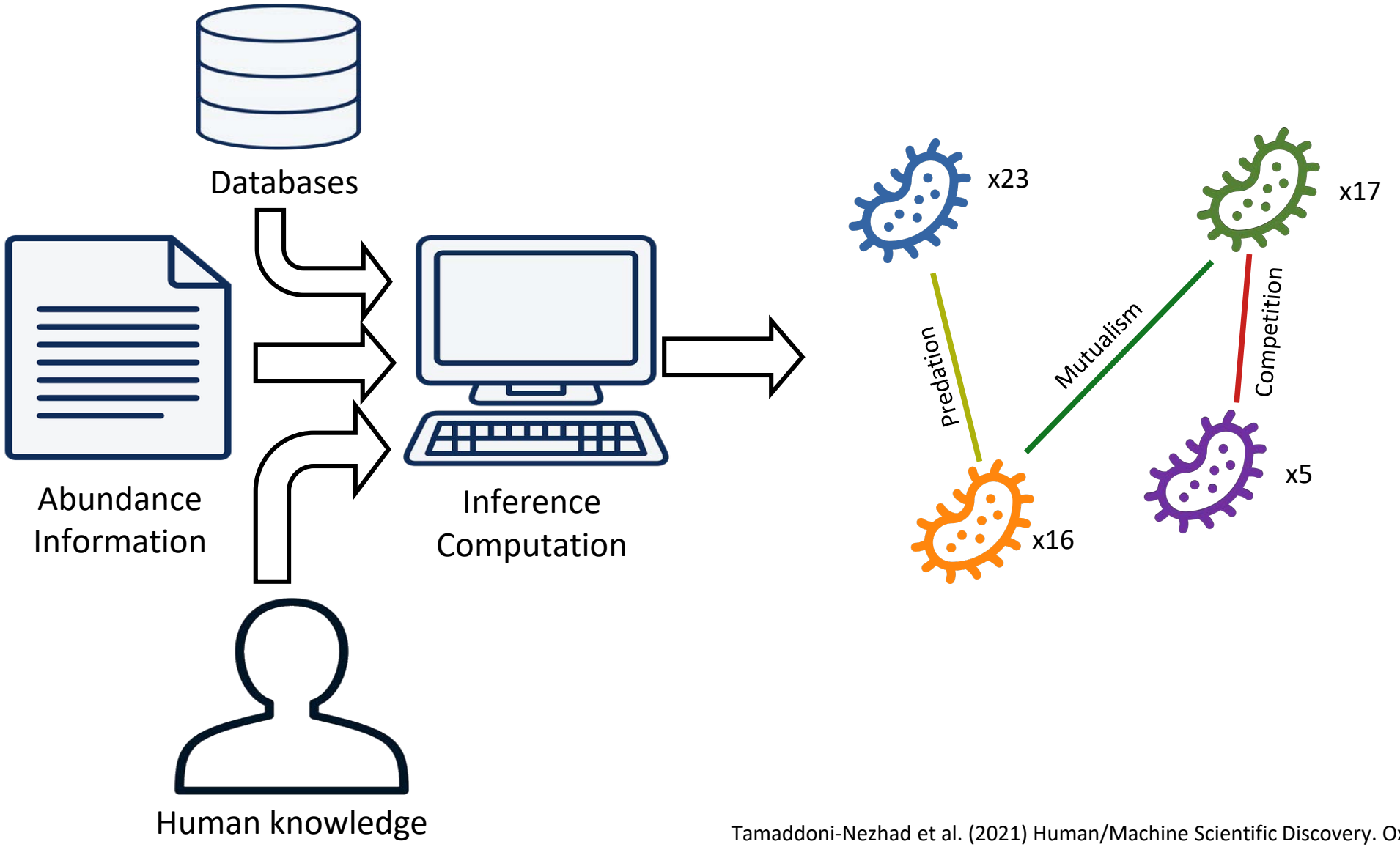


Derocles et al. (2018). Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into Ecological Network Analysis. *Advances in Ecological Research*

Interaction Inference using Correlation-based Methods

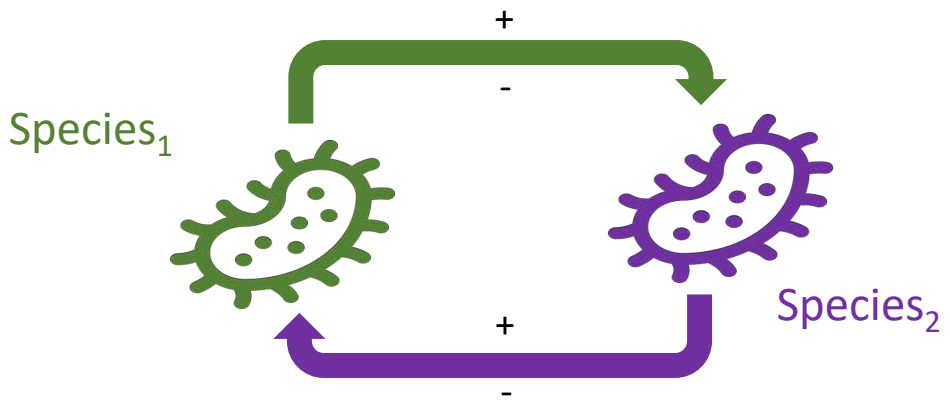


Interaction Inference using Explainable Machine Learning



Tamaddoni-Nezhad et al. (2021) Human/Machine Scientific Discovery. Oxford University Press

Defining logical interaction rules (hypotheses)

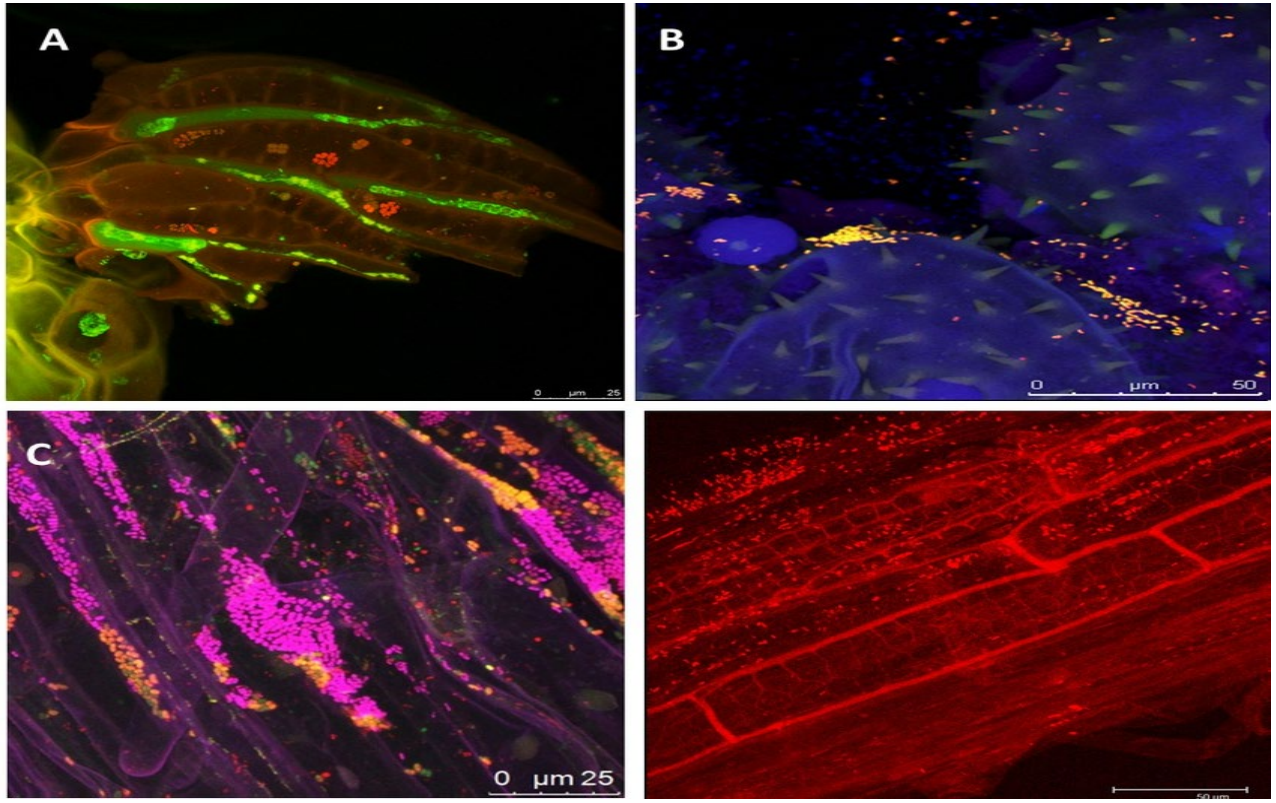


Interaction type	Effect on abundance of Species$_1$	Effect on abundance of Species$_2$	Nature of interaction
Mutualism	Up (+)	Up (+)	Mutual benefits to both species
Competition	Down (-)	Down (-)	Species have negative effect on each other
Predation/Parasitism	Up (+)	Down (-)	Predator/parasite develops at the expense of the prey/host
Commensalism	Up (+)	Null	Species$_1$ benefits while Species$_2$ is not affected
Amensalism	Down (-)	Null	Species$_2$ has a negative effect on Species$_1$, but Species$_2$ is not affected

Derocles et al. 2018. Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into Ecological Network Analysis. *Advances in Ecological Research*

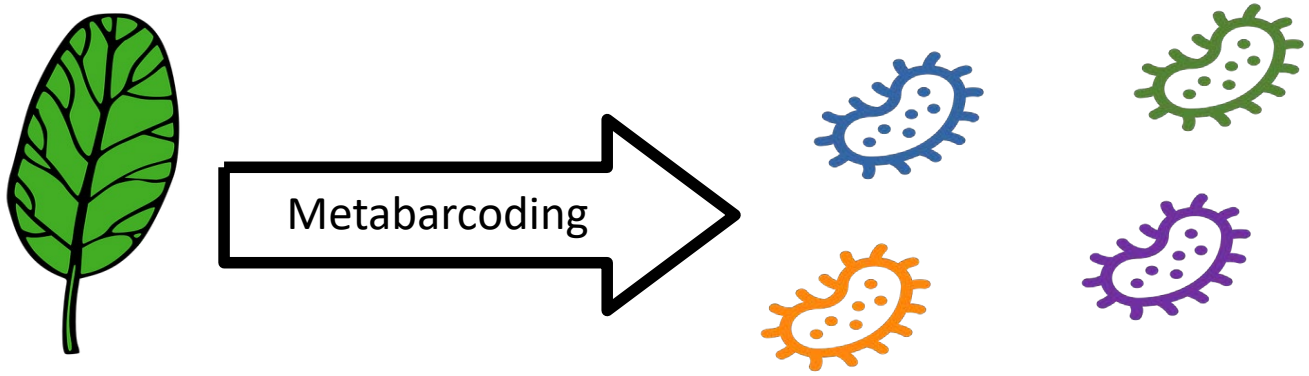
Microorganisms as a study case

- Microorganisms play a major role in all ecosystems
- There is still a lack of knowledge about microbial communities
- Interactions are key for defining microbial communities, but are not visible
- eDNA offers the possibility of obtaining new information



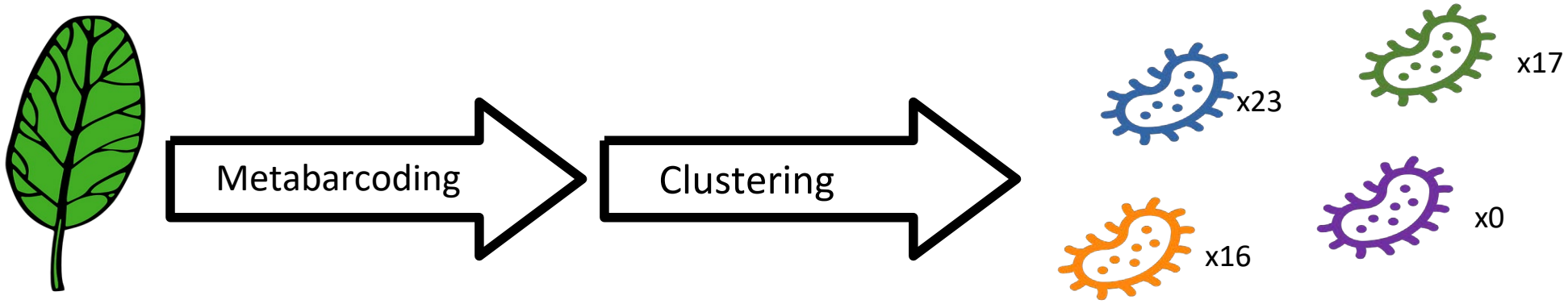
Berg et al. (2013) Next-Generation Bio-Products Sowing the Seeds of Success for Sustainable Agriculture. *Agronomy*

Learning from Next Generation Sequencing data



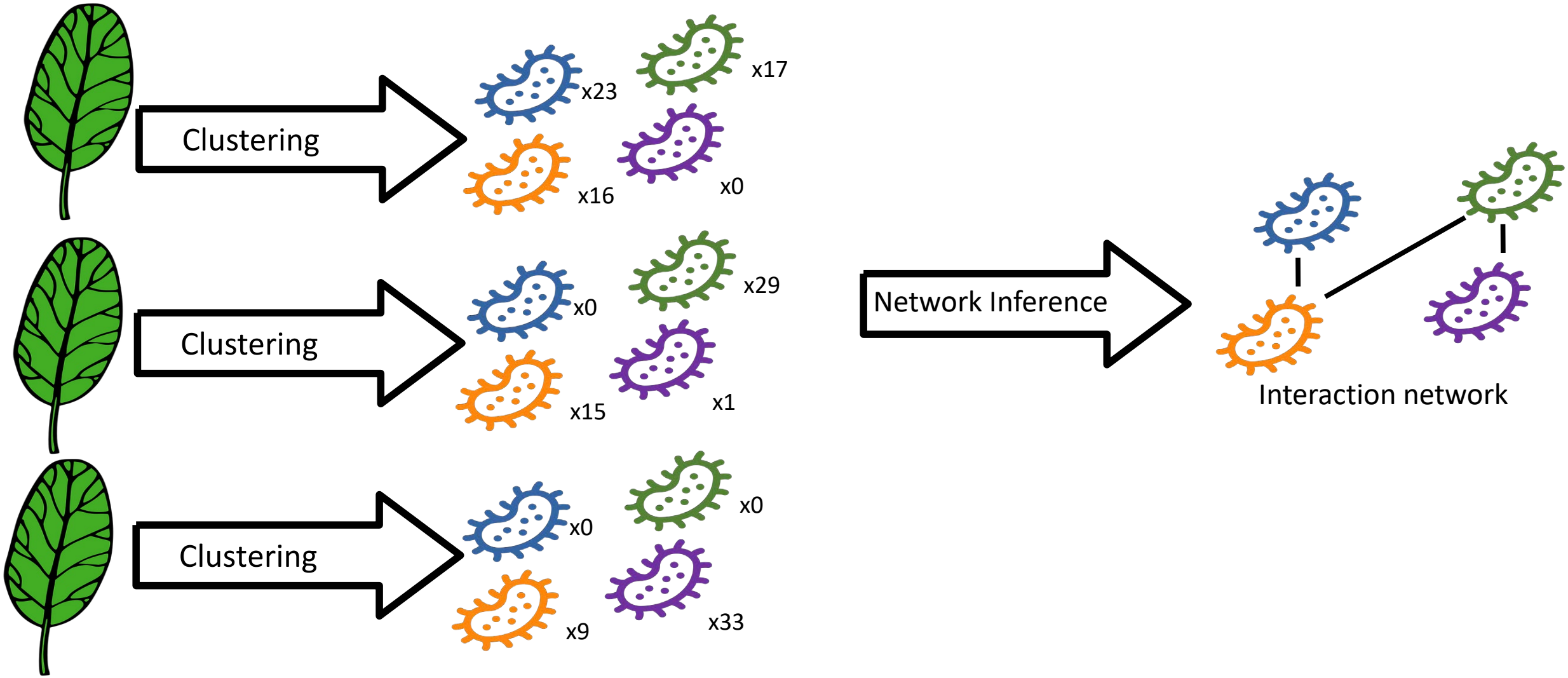
Taxa identification

Learning from Next Generation Sequencing data



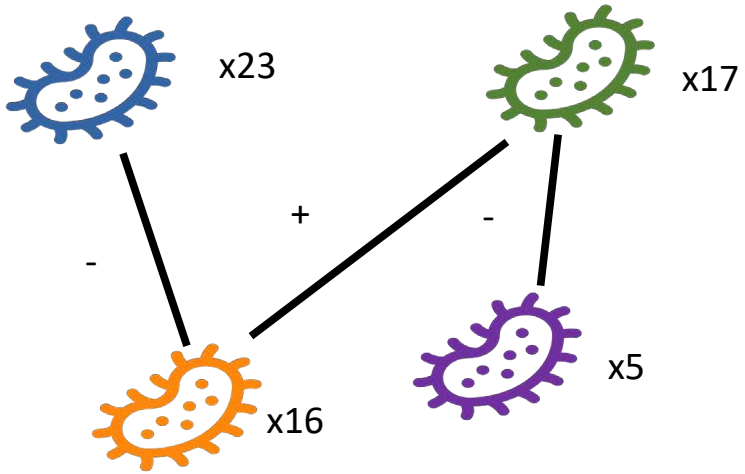
- Clustering groups of sequences into Operational Taxonomic Units (OTUs), which represent taxonomic groups (e.g. species)
- Amplicon Sequence Variants (ASVs) are a high resolution version of the OTUs

Learning from Next Generation Sequencing data



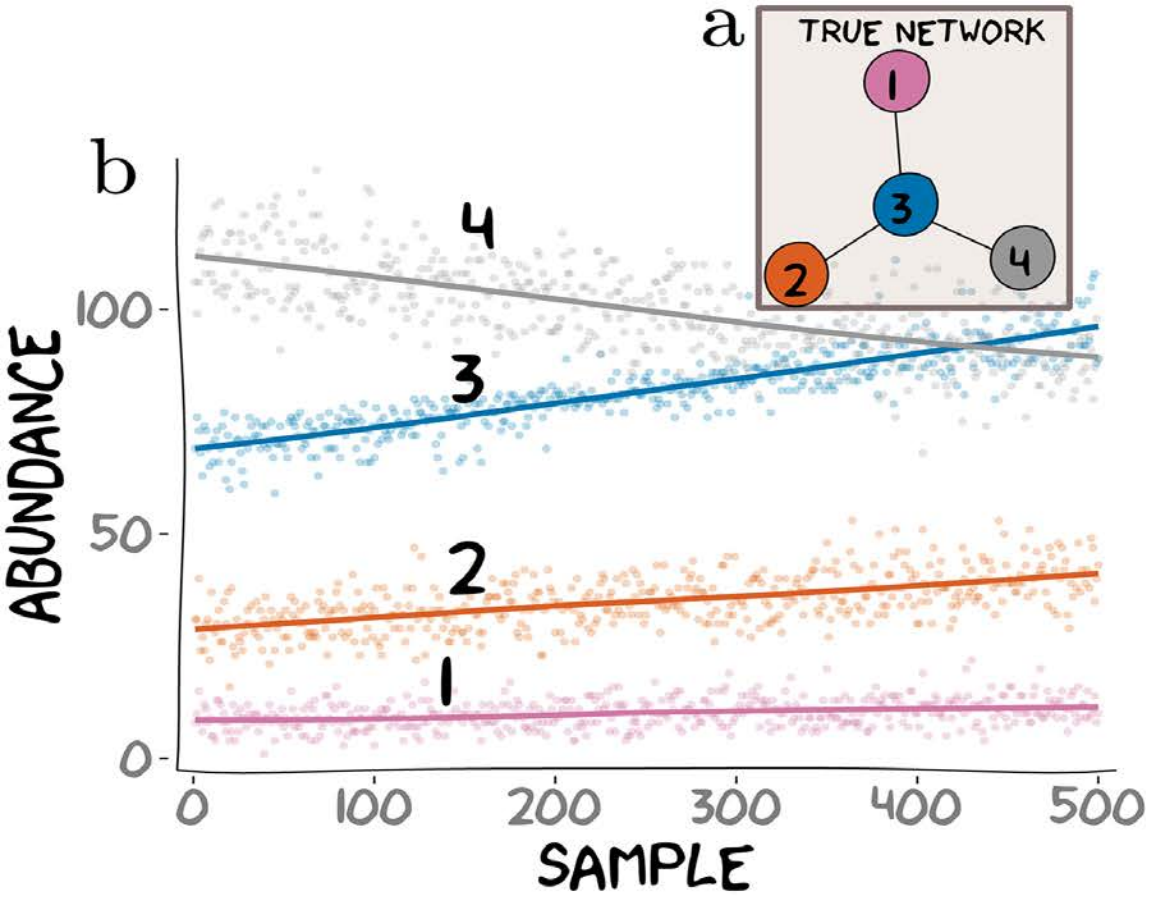
Using correlation between abundances to infer networks

- There are numerous network inference tools that can be applied to metabarcoding systems
- These include SparCC and SPIEC-EASI, amongst others
- They detect positive and negative associations (links) that require interpretation and experimental validation



Interaction Network

Using correlation between abundances to infer networks



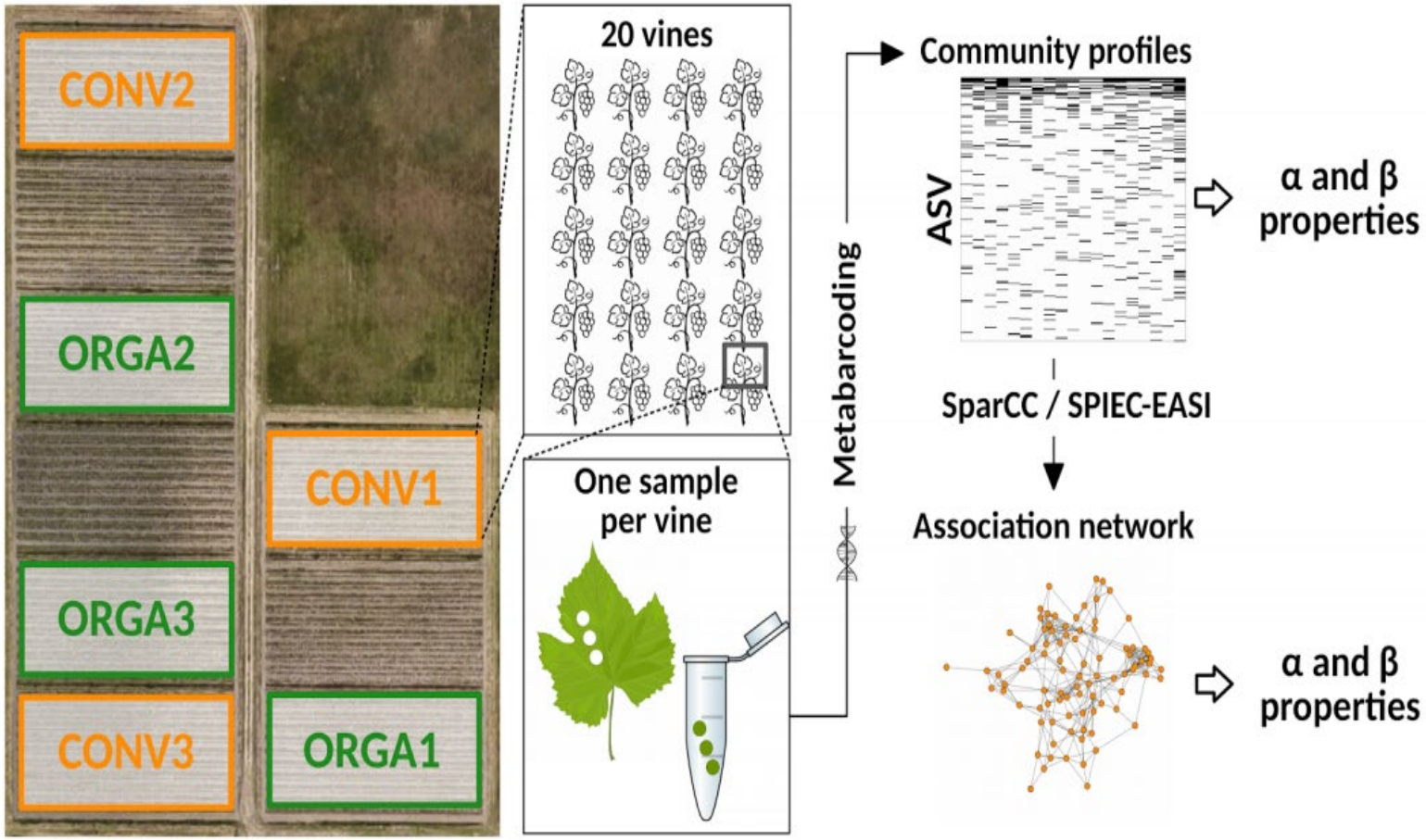
Kurtz et al. (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLoS computational biology*

Testing correlation tools for biomonitoring

- Microbial interaction networks change depending on the cropping system
- Are networks reconstructed from the same cropping system samples repeatable?
- Can reconstructed interaction networks capture the change in practice between cropping systems?

Experimental design

- 3 organic and 3 conventional plots
- 20 vines sampled per plot
- ASV Clustering performed by DADA2



Barroso-Bergada et al. (2020) Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. *Molecular Ecology Resources*

Network properties

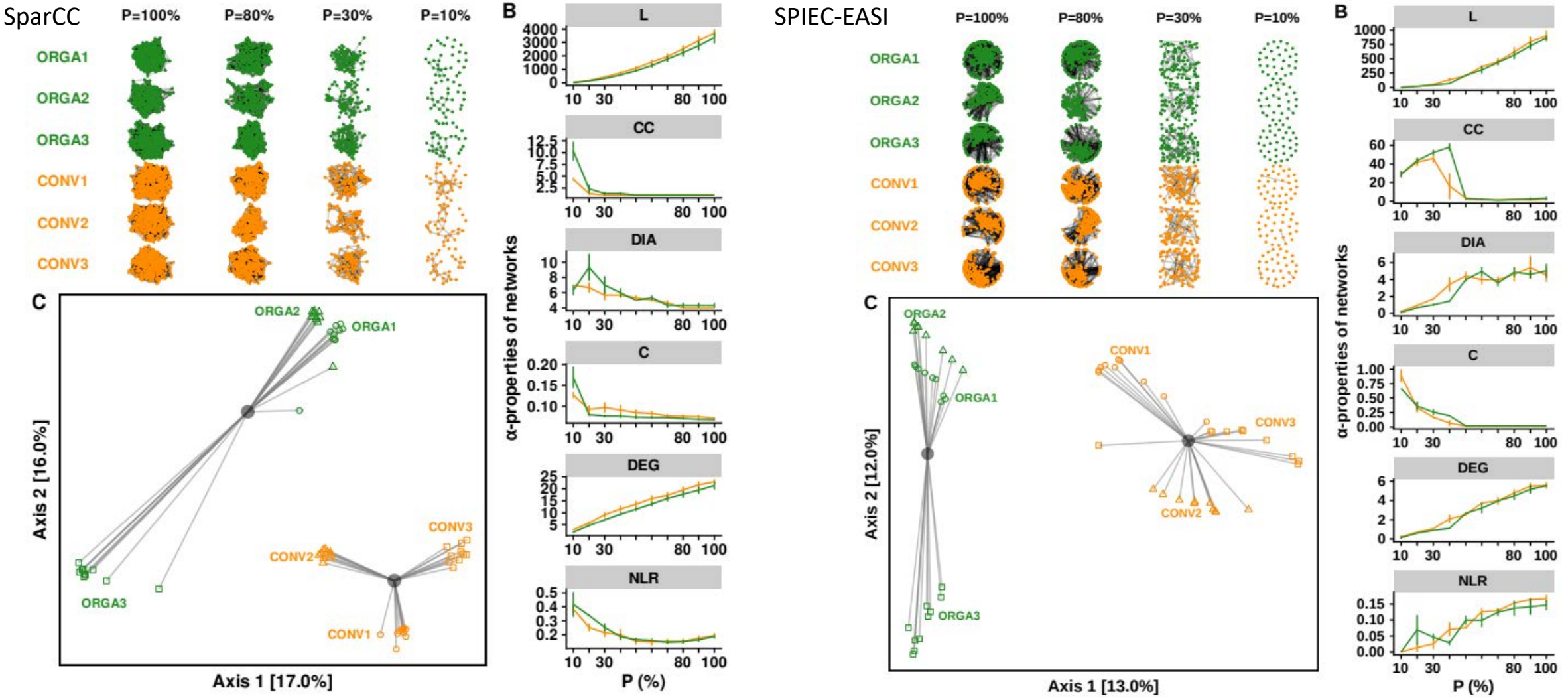
Network α-properties

Number of links (L)	Total number of links	-
Connectance (C)	Fraction of the total number of possible links actually realized	Coleman & Moré, 1983
Number of connected components (CC)	Number of groups of nodes connected together	Martinez, 1992
Diameter (DIA)	The longest of all the shortest paths between two nodes	Barabási et al, 2000
Mean node degree (DEG)	Mean number of links per node	Martinez, 1992
Proportion of negative links (NLR)	Proportion of links for which the SparCC correlation is negative	Faust et al, 2015

Network β-properties

Topological dissimilarity (Schieber's D)	Dissimilarity of global and local network structure	Schieber et al, 2017
Association dissimilarity (βWN)	Overall dissimilarity of associations	Poisot et al, 2012
Association dissimilarity (βOS)	Dissimilarity of associations between shared ASVs	Poisot et al, 2012
Association dissimilarity (βST)	Dissimilarity of associations due to ASV turnover	Poisot et al, 2012

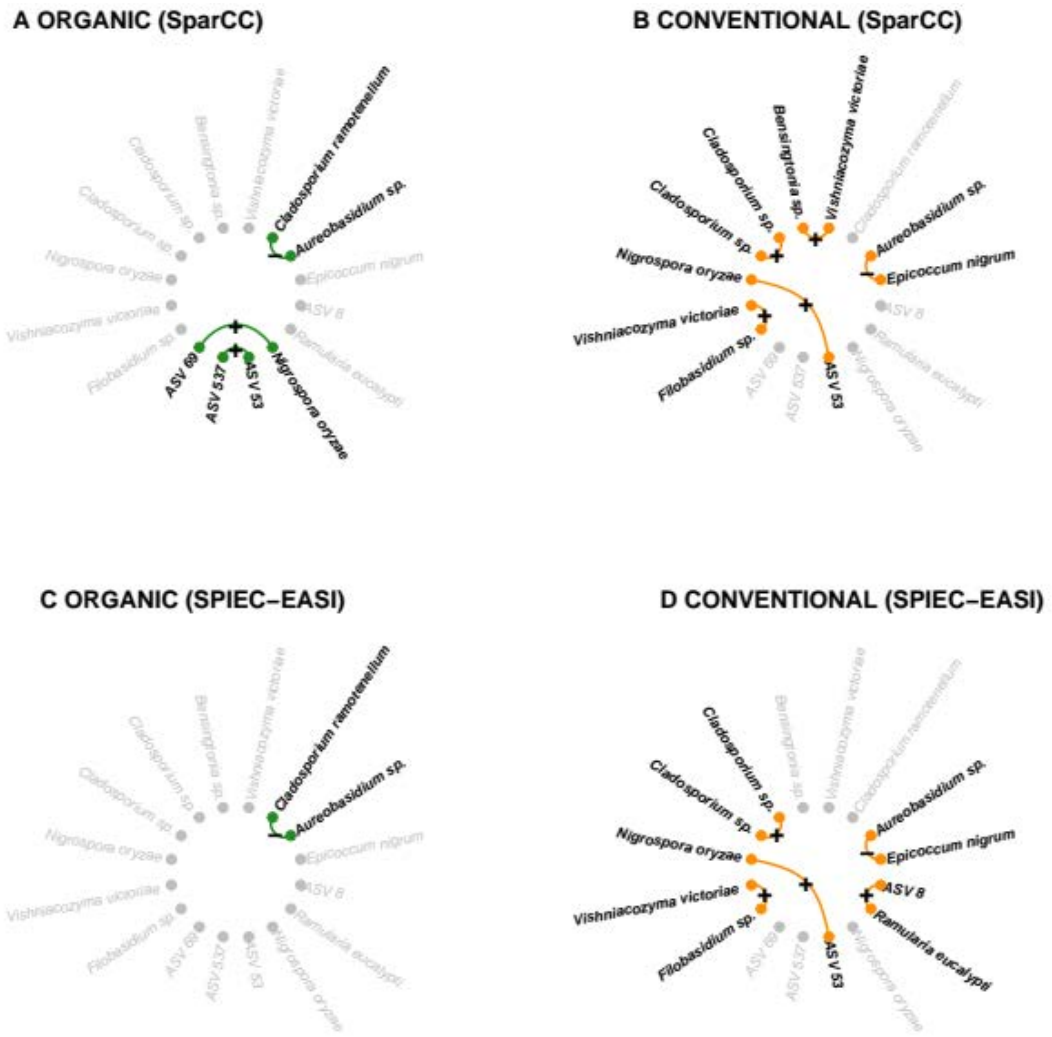
Network β-properties differentiate cropping systems



Barroso-Bergada et al. (2020) Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. *Molecular Ecology Resources*

Few consensus associations were shared between networks

- Each network had several hundred ASVs and associations
- A maximum of 5 associations were found to be shared by networks from the same experimental conditions



Barroso-Bergada et al. (2020) Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. *Molecular Ecology Resources*

Conclusions

- Network α-properties do not differentiate cropping systems
- Some β-properties differed between cropping systems
- Networks inferred from replicate cropping systems had few consensus associations
- Associations obtained by SparCC and SPIEC-EASI were classified into positive and negative correlations

Barroso-Bergada et al. (2020) Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. *Molecular Ecology Resources*

Explainable microbial network inference: The logic process

Deduction

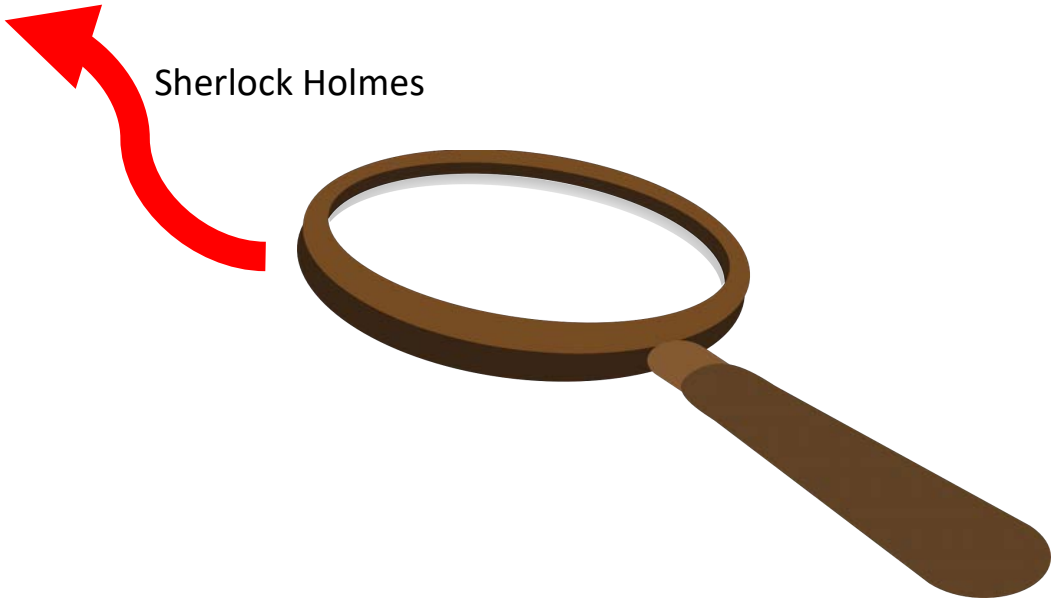
Rule: An interaction produces a change in abundance
Case: Two taxa interact
Result: The abundance of at least one taxa changes

Induction

Case: Two taxa interact
Result: The abundance of at least one taxa changes
Rule: An interaction produces a change in abundance

Abduction

Result: The abundance of at least one taxa changes
Rule: An interaction produces a change in abundance
Case: Two taxa interact



Explainable microbial network inference: The logic process

Result: The abundance of at least one taxa changes

	x	y	z
Species1 (S1)	0	11	300
Species2 (S2)	121	27	0
Species3 (S3)	0	0	43

Logic Statements:

$$abundance(x, y, s2, down)$$
$$presence(s1, x, no)$$

Rule: An interaction produces a change in abundance

$$effect\_up(s1, s2) \quad \text{if:} \quad \begin{cases} abundance(x, y, s2, up) \\ presence(s1, x, no) \\ presence(s1, y, yes) \end{cases}$$

$$effect\_down(s1, s2) \quad \text{if:} \quad \begin{cases} abundance(x, y, s2, down) \\ presence(s1, x, no) \\ presence(s1, y, yes) \end{cases}$$

Case: Two taxa interact

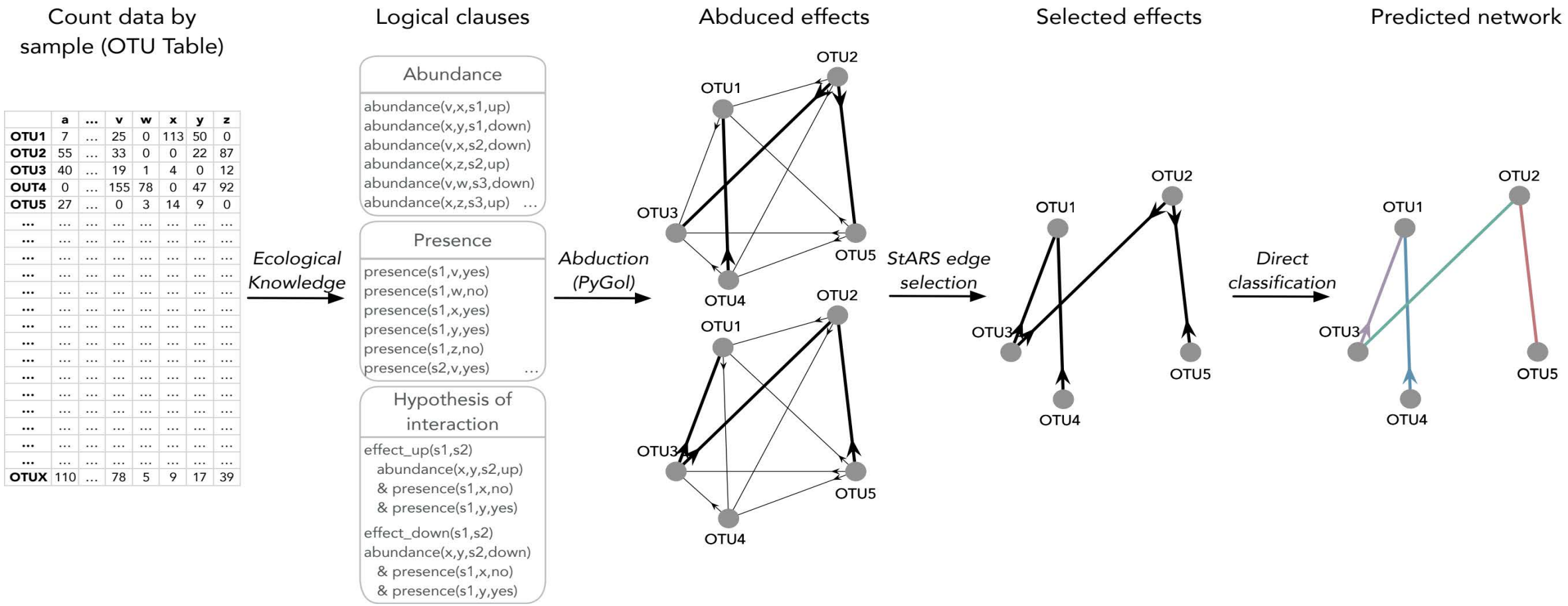
$$effect\_up(s1, s2)$$
$$effect\_down(s1, s2)$$

235
- 157

I statistic = 78

Compression: Amount of examples supporting the abduced fact

InfIntE: Inference of Interactions using Explainable machine learning



	a	…	v	w	x	y	z
OTU1	7	…	25	0	113	50	0
OTU2	55	…	33	0	0	22	87
OTU3	40	…	19	1	4	0	12
OUT4	0	…	155	78	0	47	92
OTU5	27	…	0	3	14	9	0
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
OTUX	110	…	78	5	9	17	39

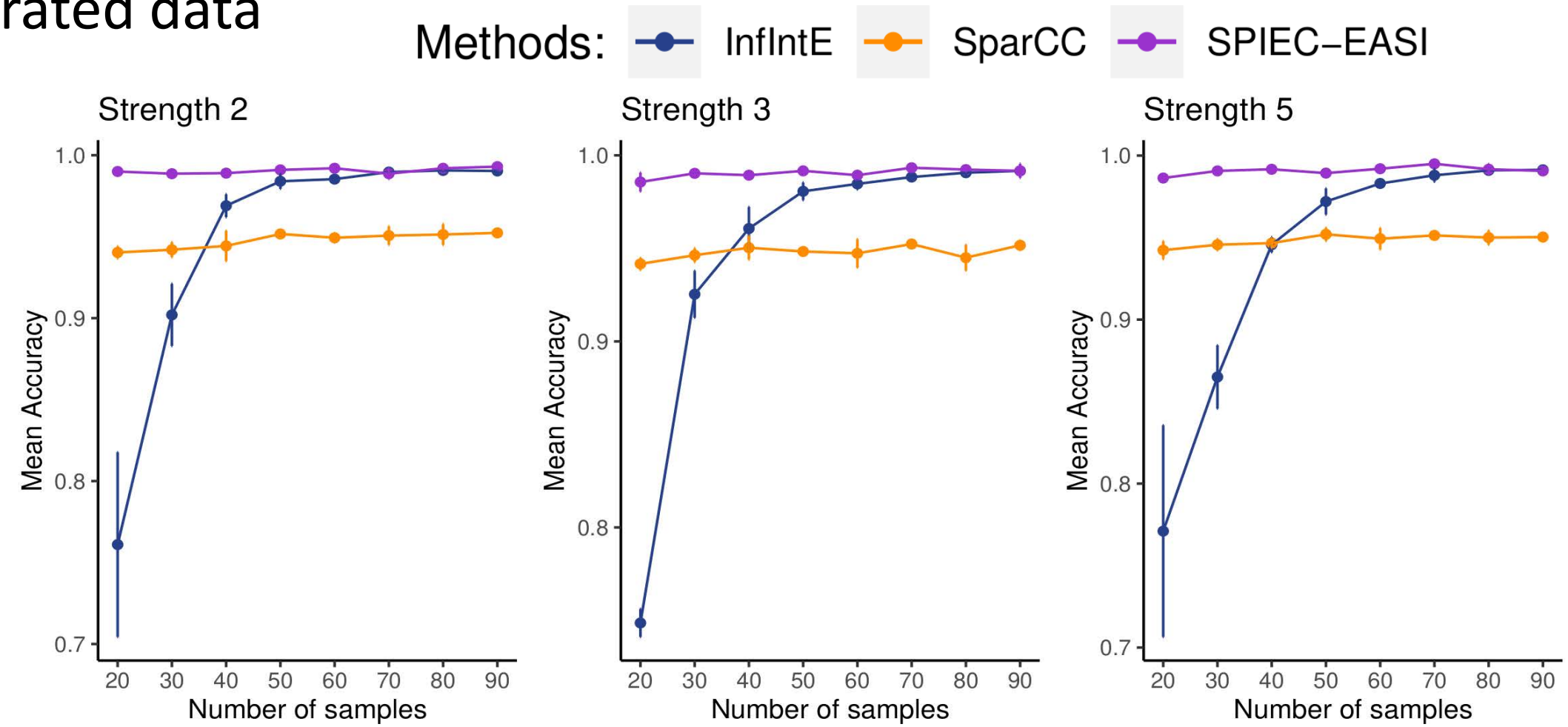
Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

Testing InfIntE

- Two stage process – i) using simulated data; and, ii) using real data…
- Use of computer generated data for ecological interactions, as proposed by Weiss et al. (2016)
- Different strengths of interaction
- 4 types of interactions simulated (commensalism, competition, mutualism and amensalism)
- Tested on real metabarcoding data from grapevine leaves in different vineyards
- Vineyards were affected by downy mildew, caused by *Plasmopara viticola*

Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

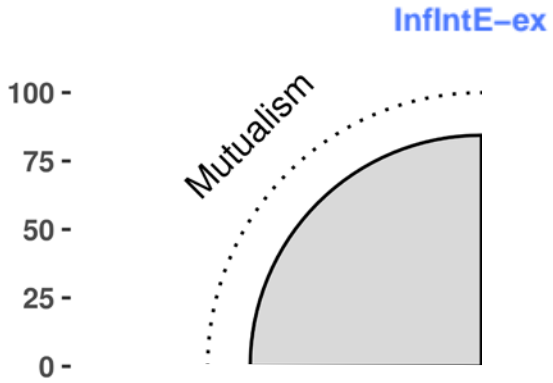
Accuracy of interaction selection, as evaluated using computer generated data



$$Accuracy = \frac{TP + TN}{N}$$

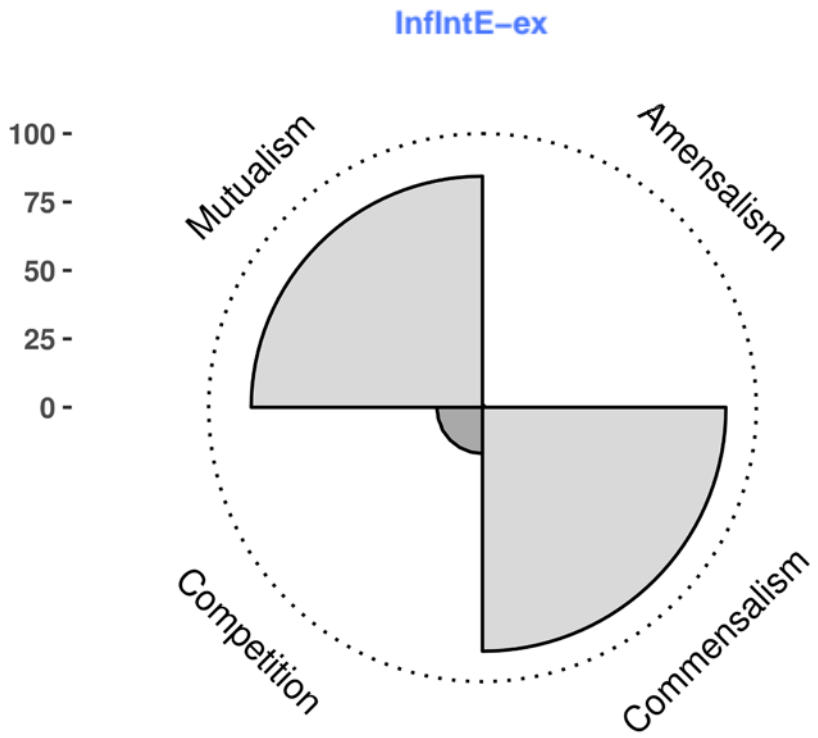
Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

Accuracy of interaction classification for 60 samples



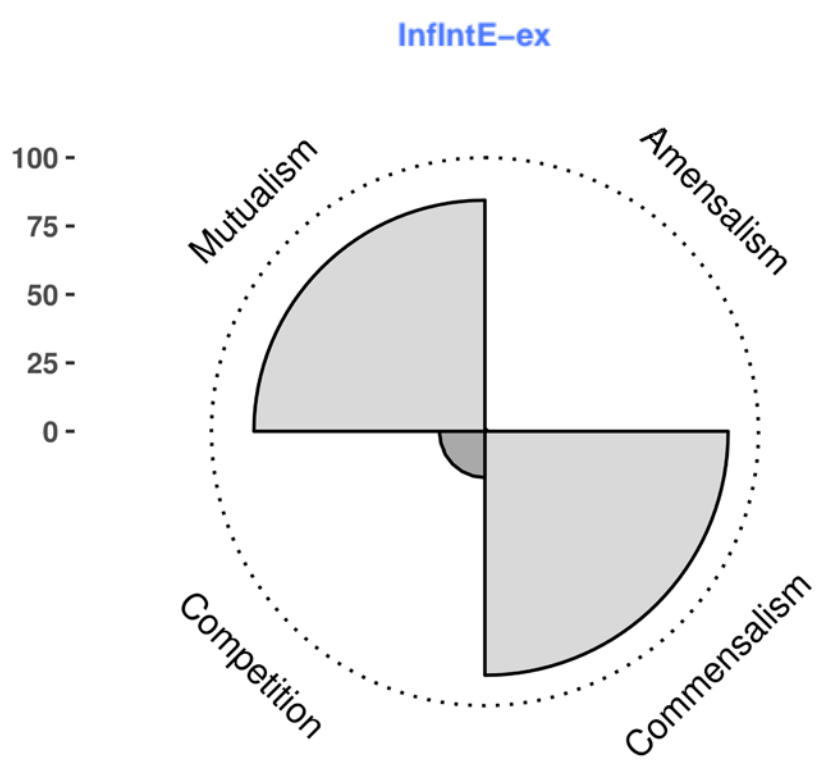
Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

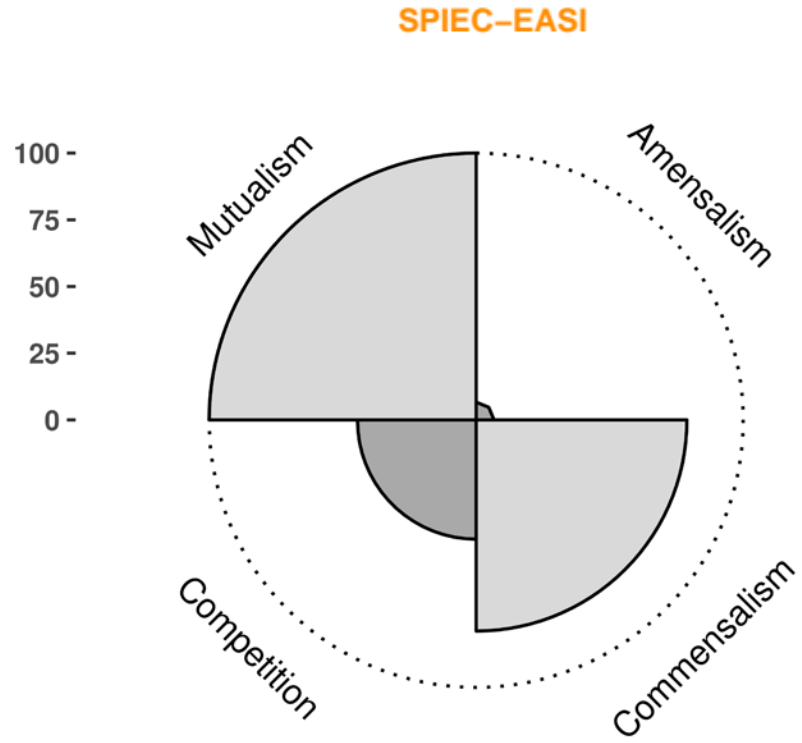
Accuracy of interaction classification for 60 samples



Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

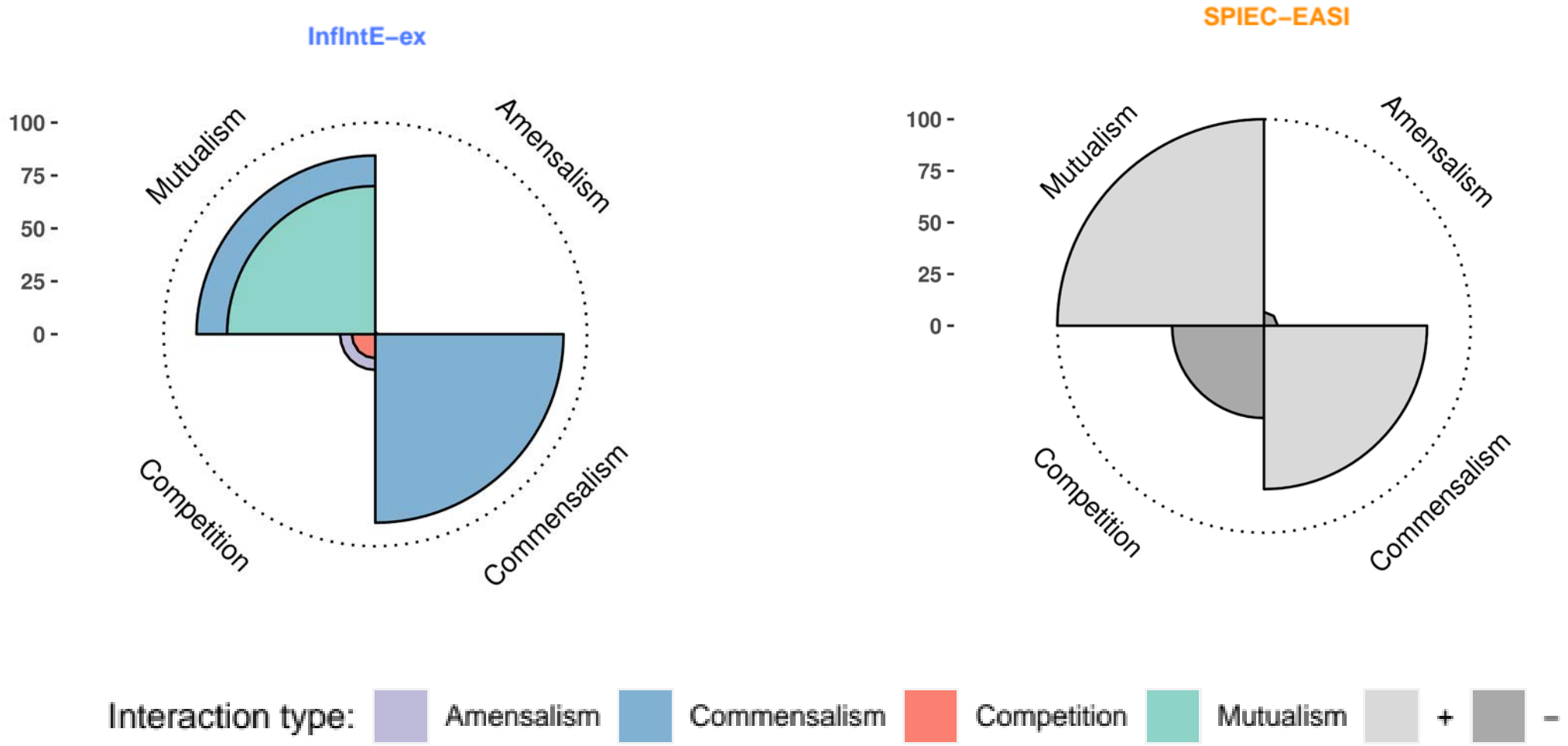
Accuracy of interaction classification for 60 samples





Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

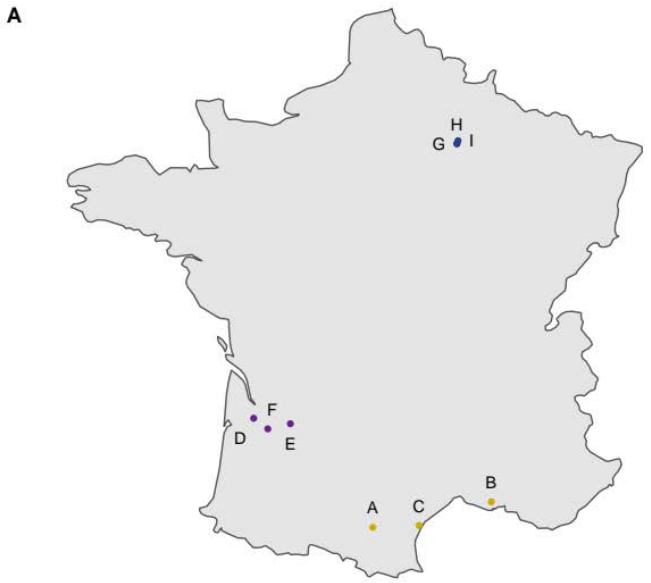
Accuracy of interaction classification for 60 samples



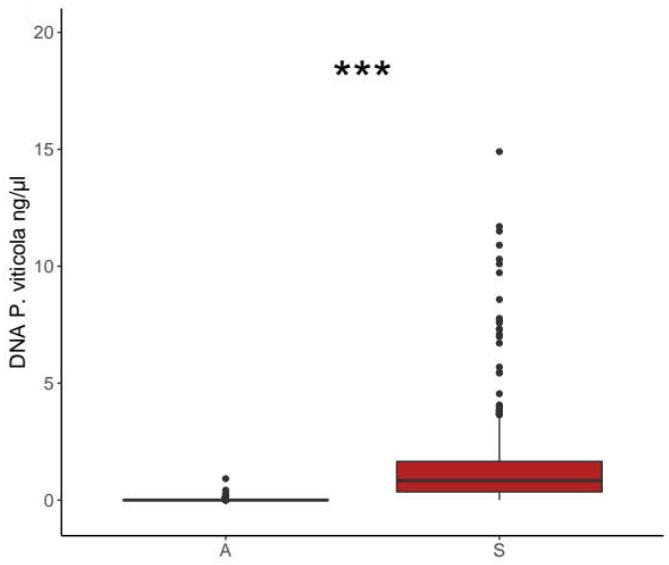
Barroso-Bergada et al. in press. Unravelling the web of dark interactions: explainable inference of the diversity of microbial interactions. *Advances in Ecological Research.*

Using real metabarcoding data

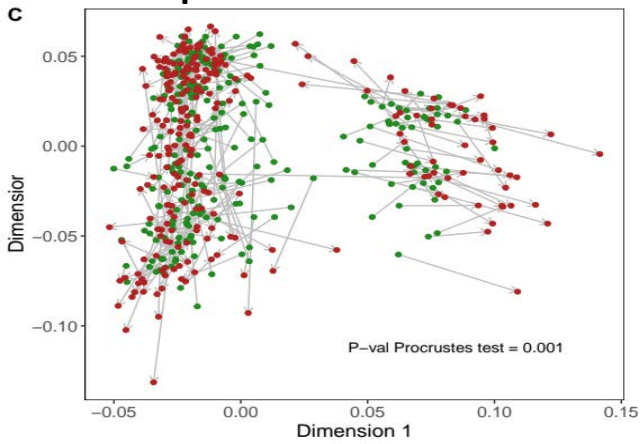
Vineyard locations

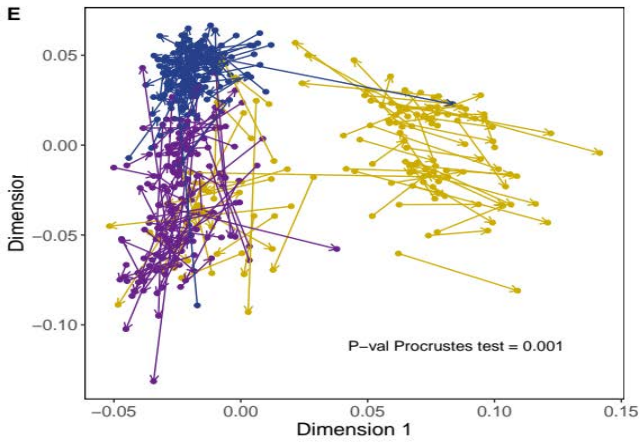


Pathogen concentration



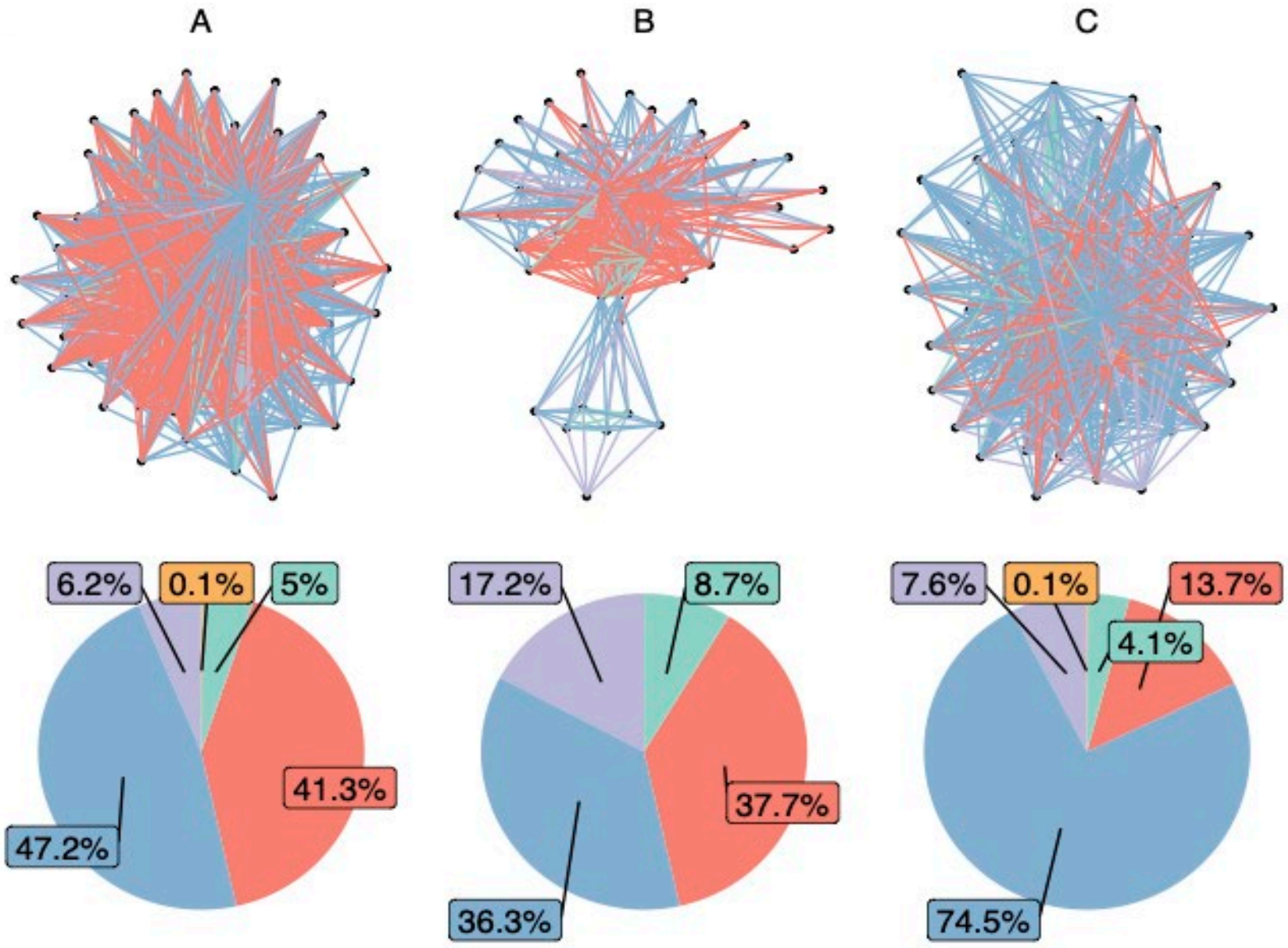
Sample ordination





Networks reconstructed using real data

Interaction Type	Effect on OTU1	Effect in OTU2
Amensalism	down	null
Commensalism	up	null
Competition	down	down
Mutualism	up	up
Predation	up	down



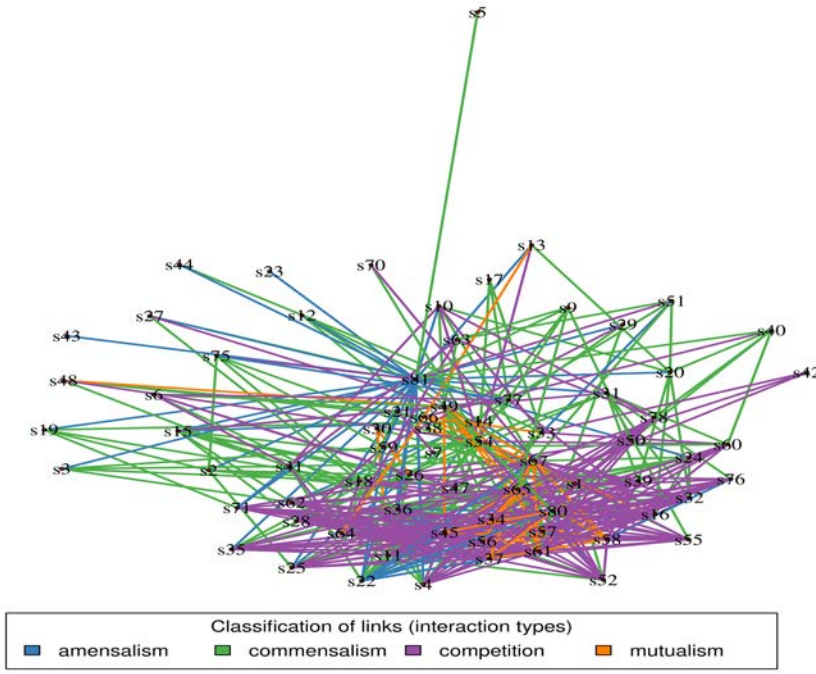
Validating a sub-network:

- Negative interactions with the pathogen *P. viticola*
- Searched using keywords:
  - *P. viticola* antagonist
  - Biocontrol

Name	Plot	Interaction	Bibliography against plasmopara	Bibliography biocontrol
Cladosporium delicatulum	I	competition		Kohl et al. 2019 ; Baharvandi et al. 2015; Venkateswarulua et al. 2018
Mycosphaerella tassiana	I	competition		
Alternaria rosae	A	amensalism	Mussetti et al. 2006	
Alternaria alternata*	I	competition	Mussetti et al. 2006, 2007	
Alternaria brasicae	B	competition	Mussetti et al. 2006 ; Duhan et al. 2021	
Aureobasidium pullulans*	I	competition	Harm et al. 2011; Ruhmann et al. 2013	
Filobasidium chernovii	Ix2	competition		
Filobasidium magnum*	D	competition		
Fusarium prolifetarum*	A	competition	Ghule et al. 2018 ; Bakshi et al. 2001	
Fusarium equiseti*	B,E	amensalism, competition	Ghule et al. 2018	
Phlebia rufa	E	amensalism		Silva et al. 2013; Hiscox et al. 2018
Sporobolomyces roseus	Ix3	competition		janisiewicz et al. 1994; Filonow et al. 1996; Pauvert et al. 2020?
Sporobolomyces pararoseus*	A,G	competition		Li et al. 2017
Vishniacozyma victoriae	B,C	amensalism, competition		Gramisci et al. 2018; Lutz et al. 2020; Pauvert et al. 2020?
Vishniacozyma carnescens	D	amensalism		Cordero et al. 2017; Becker et al. 2020

Conclusions

- InfIntE: Interaction inference tool based on explainable machine learning
- Identified key network comparison measures
- Potential use for identifying biocontrol agents and biomonitoring

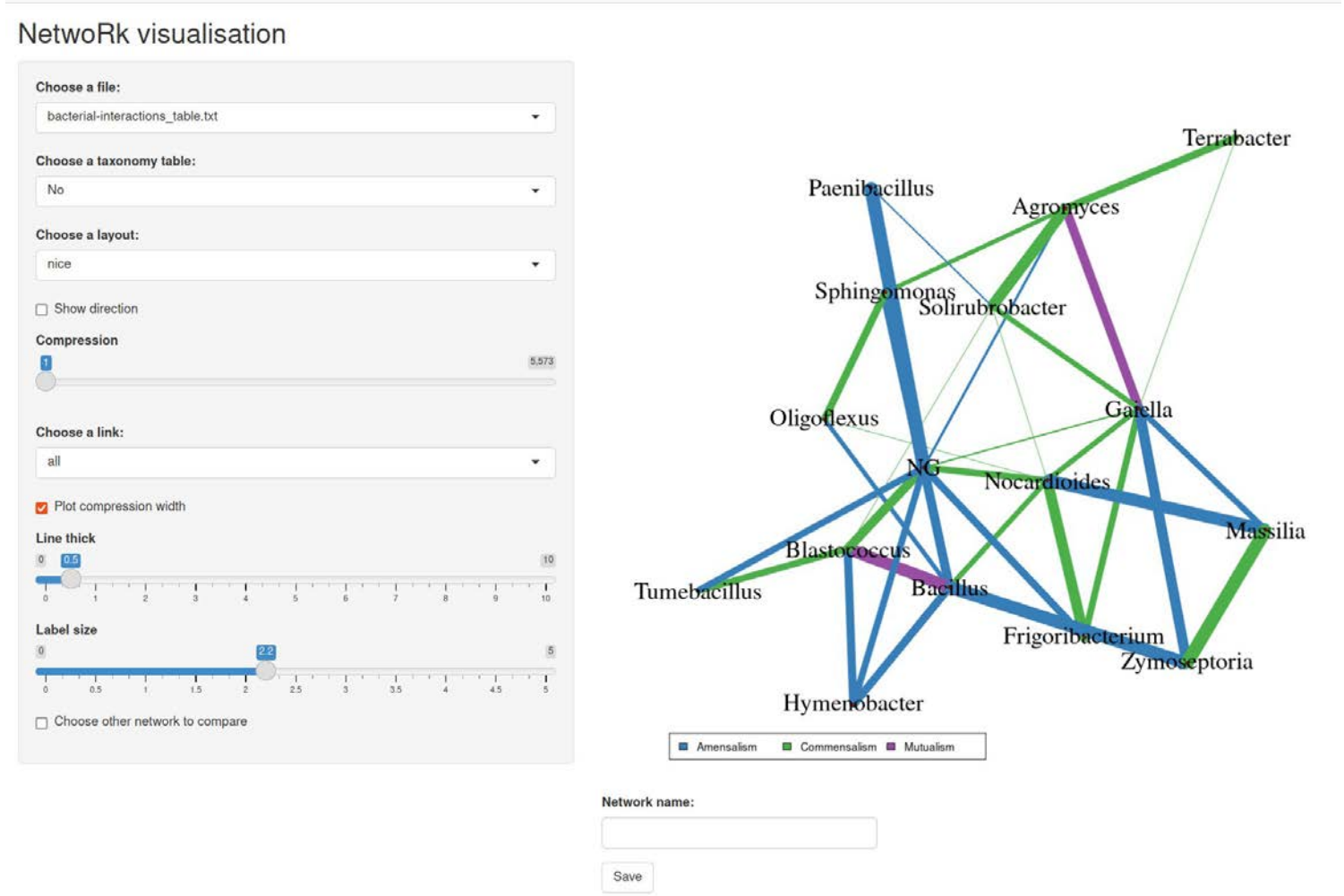


Perspectives

- Introduction of background knowledge (databases) and new interaction hypothesis to InfIntE
- Large scale testing of the InfIntE R package
- Application to automated biomonitoring

Sharing InfIntE

- Coded in an R package
- Easy to install
- Tutorial Vignette
- Complemented by a network visualization app



Which method should I choose?

Correlation based

Pros

- Robust to biases such as composition or abiotic effects
- Can work with smaller sample sizes
- Fast
- Widely used

Cons

- Requires interpretation
- Does not use domain or background knowledge

InfIntE

Pros

- Explainable detection and classification of interaction types, using composition
- Accuracy comparable to statistical methods
- Reconstruction of networks of diverse interaction types
- Flexibility: new uses of interaction hypotheses and background knowledge

Cons

- Needs larger sample sizes than correlation
- Longer execution times (but improving)

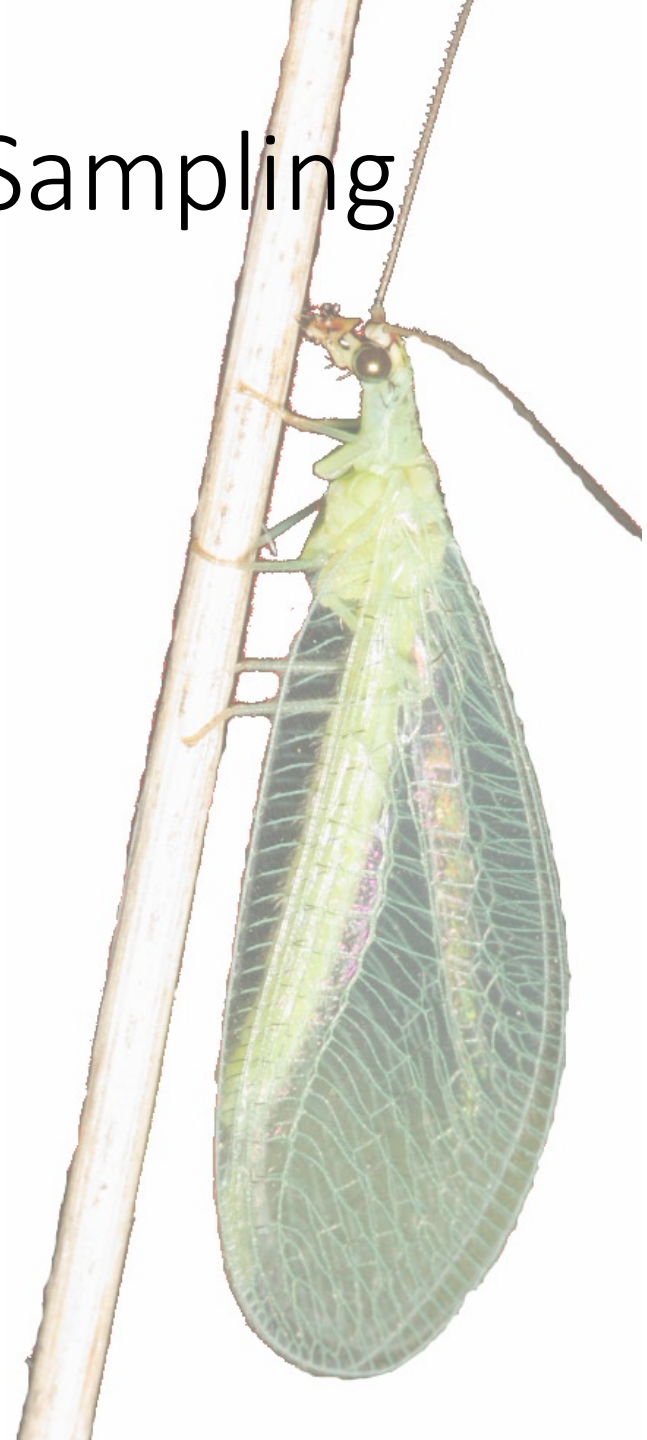
So, could we learn macro- networks?

The FSE data

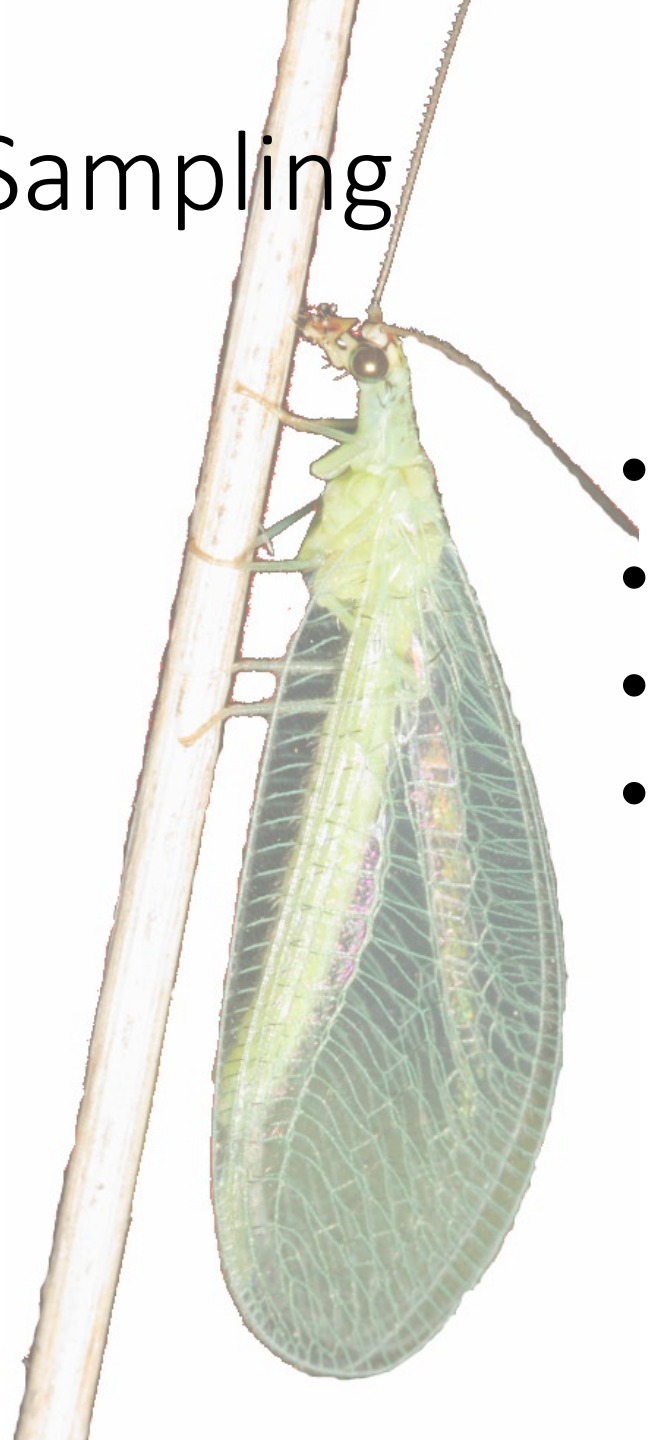
- 256 fields
- Split-field design

Betterave | Maïs | Colza de printemps | Colza d'hiver

Sampling

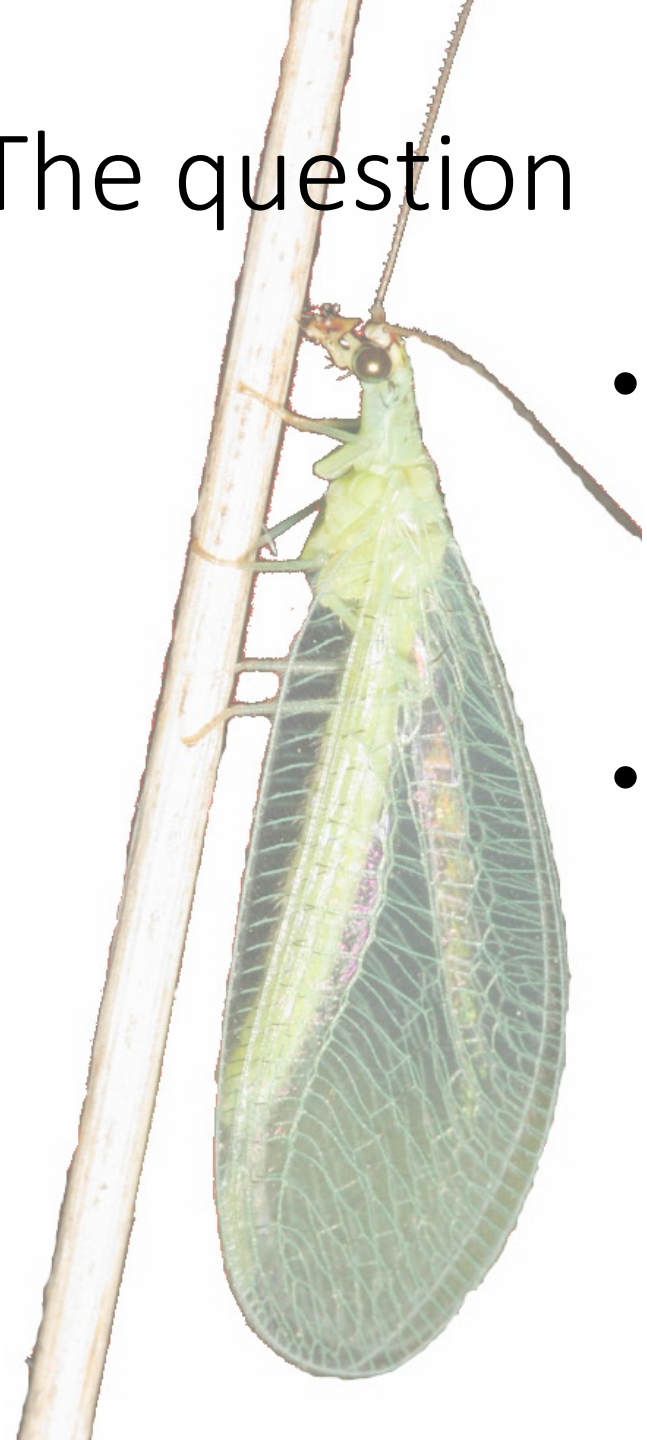


Sampling



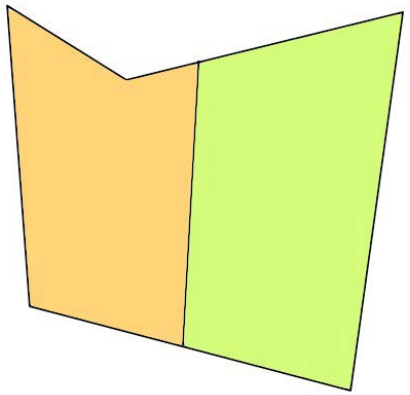
- 1.5 million weed plants counted
- 1 ton (dried) plant biomass sorted
- > 2.5 million invertebrates trapped
- > 1400 km of pollinator transects walked

The question



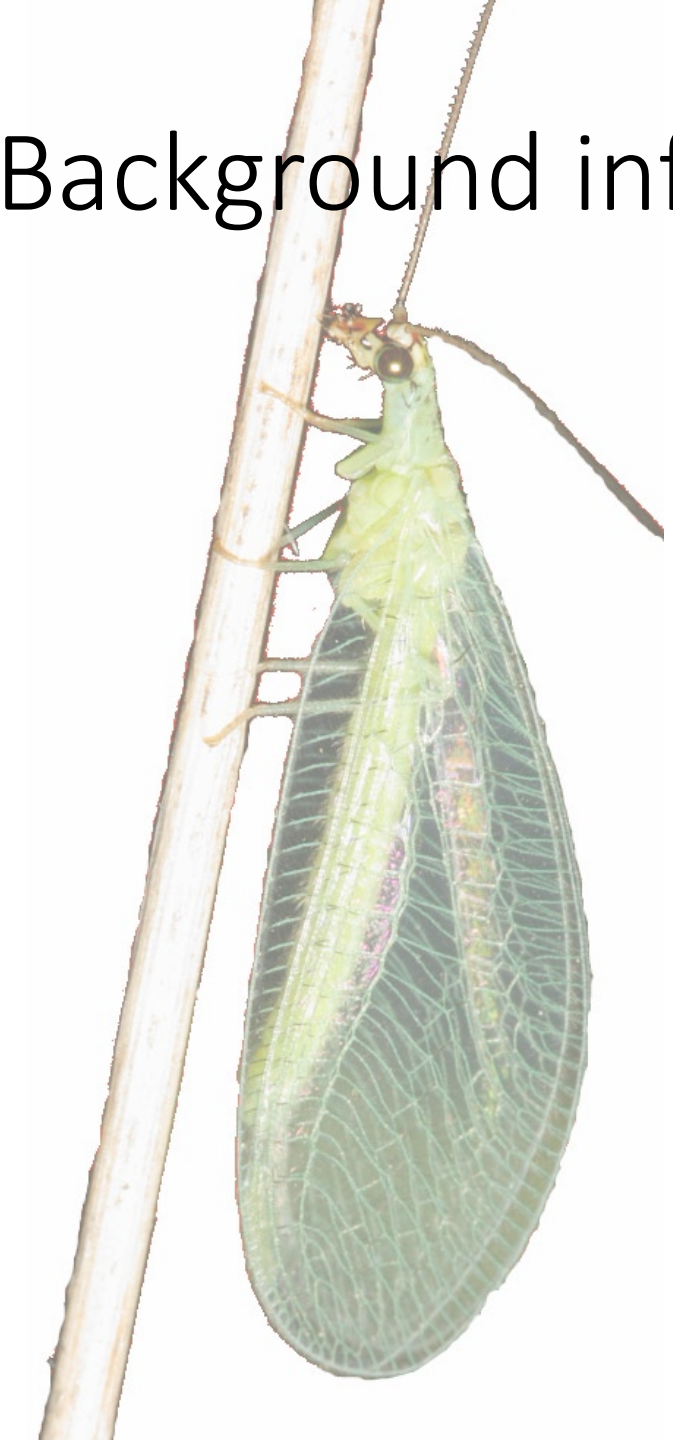
- Can we construct food webs synthetically, using currently available data?
- Are functional explanations of the ecosystem useful?

Trophic model - data



- Herbicide removes food and shelter - something to eat and somewhere to sit
- Species Y will move to new habitat or die - measured as R_Y, where $R_Y = \log_{10}(Y_{GM}/Y_C)$
- Species X, which feeds on Y, will change with Y as R_X
- *Expectation* that: R_X is 'trophically' related to R_Y

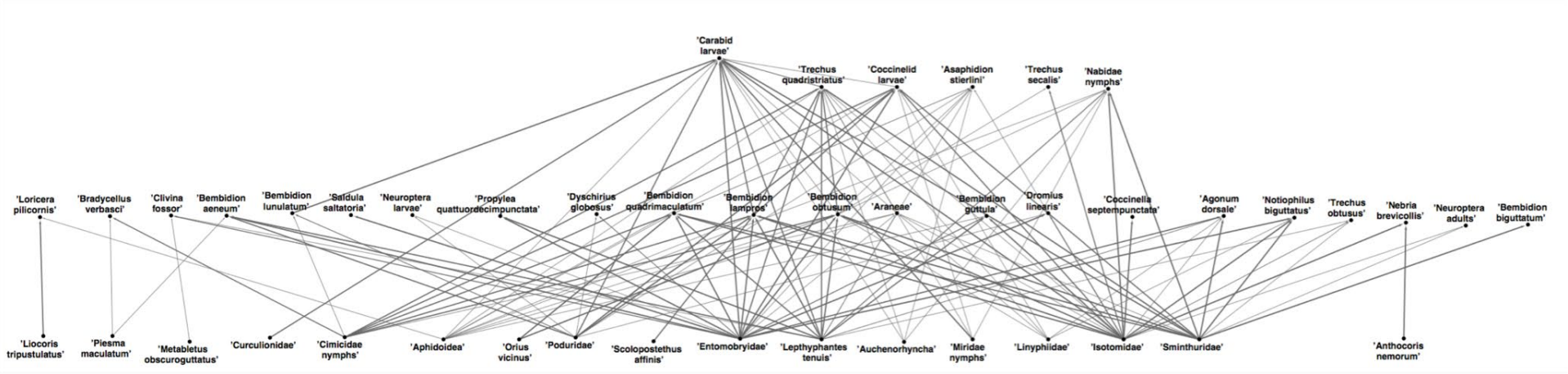
Background information



- Appropriate mouthparts for feeding
- In any sample Y and X should co-occur
- Big things eat small things
- With this set of ‘rules’ we ‘learn’ food webs

Codified in Inductive Logic Programming system, Progol5.0

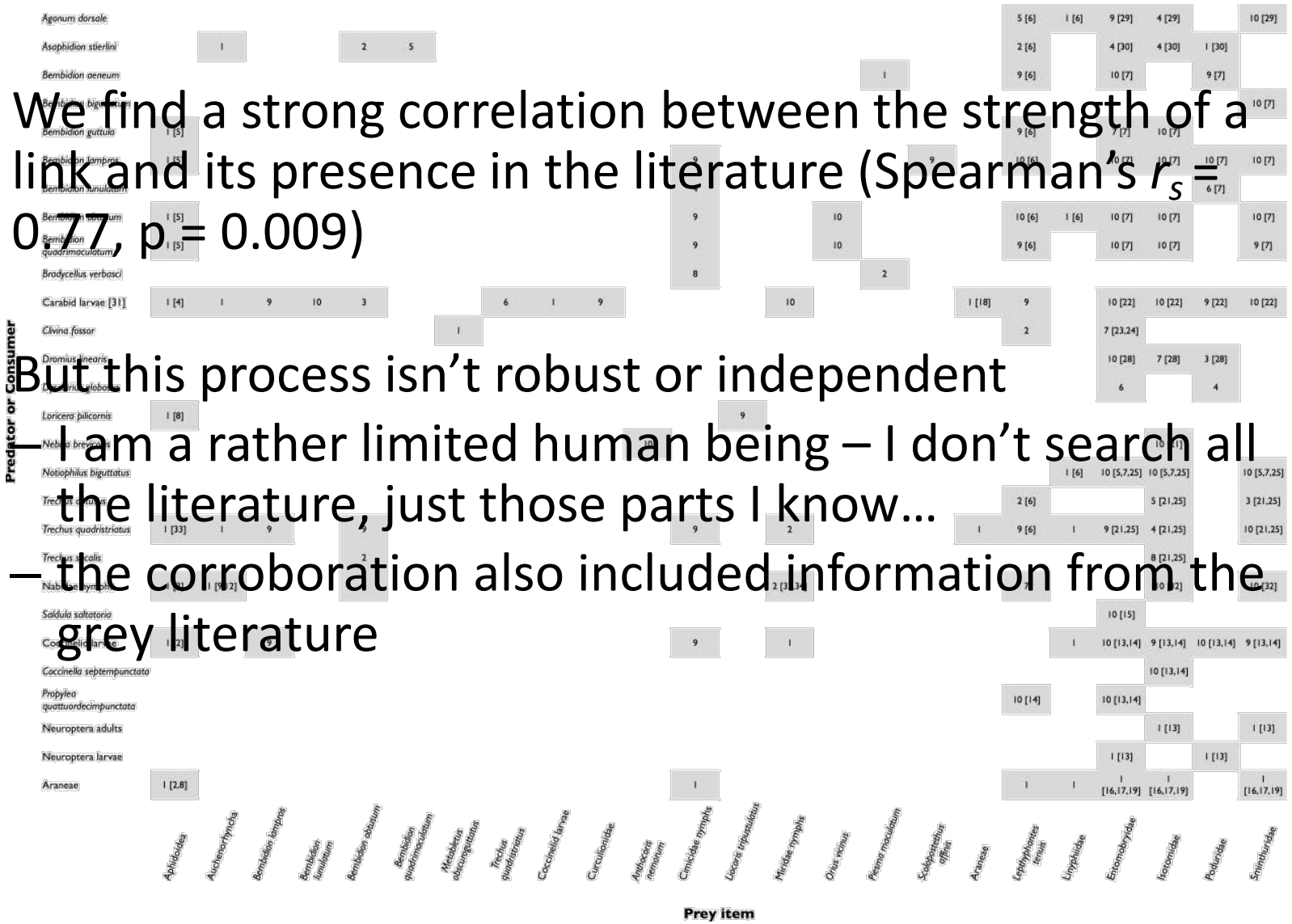
```
abundance(X, S, up):-
    predator(X),
    co occurs(S, X, Y),
    bigger than(X, Y),
    abundance(Y, S, up),
    eats(X, Y).
```



- 45 invertebrate species or taxa (~25%), but about 74% of the individuals were linked
- The literature corresponds with learnt links
- Collembola important prey. Carabid beetles were the dominant predators. Carabid larvae predators of a wide variety of prey.
- Lots of intraguild predation

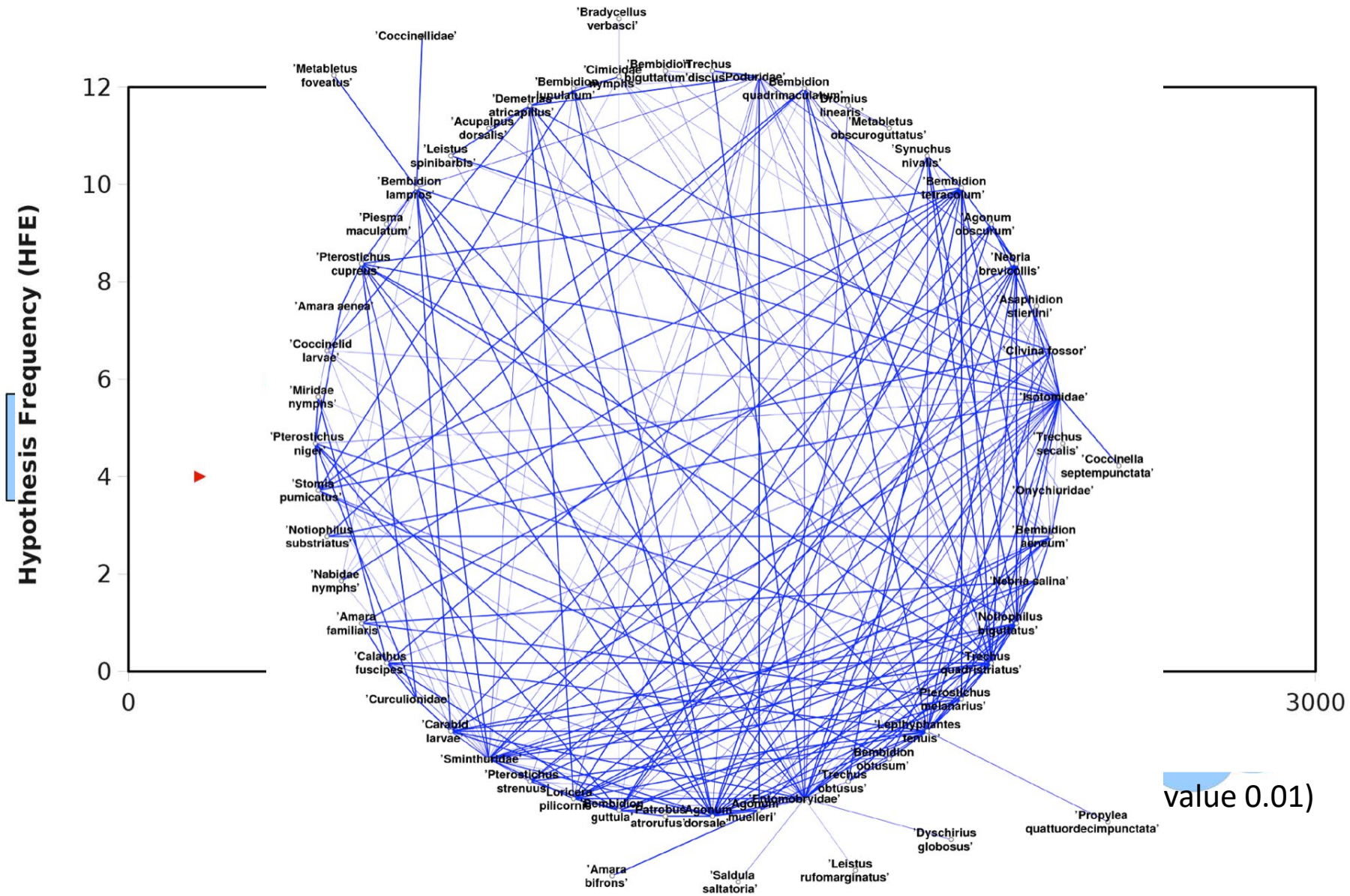
Bohan et al. (2011) *PLoS ONE* and Tamaddoni-Nezhad et al. (2012) *Machine Learning*

Validation



- We find a strong correlation between the strength of a link and its presence in the literature (Spearman's r_s = 0.77, p = 0.009)
- But this process isn't robust or independent
  - I am a rather limited human being – I don't search all the literature, just those parts I know…
  - the corroboration also included information from the grey literature

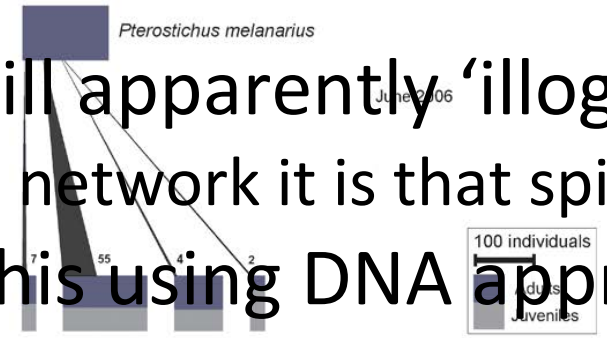
Automatic literature verification

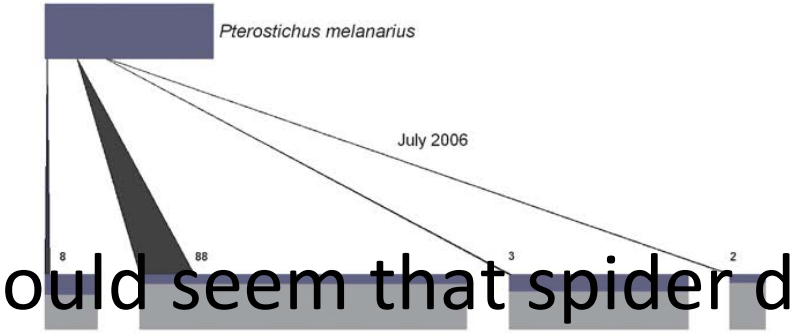


value 0.01)

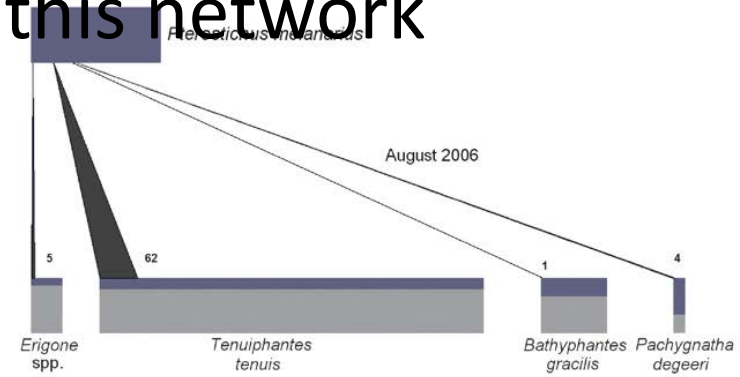
Problems

- There are still apparently ‘illogical’ links…
  - » In the vortis network it is that spiders act as prey
- We tested this using DNA approaches





- and… it would seem that spider do indeed act as prey in this network

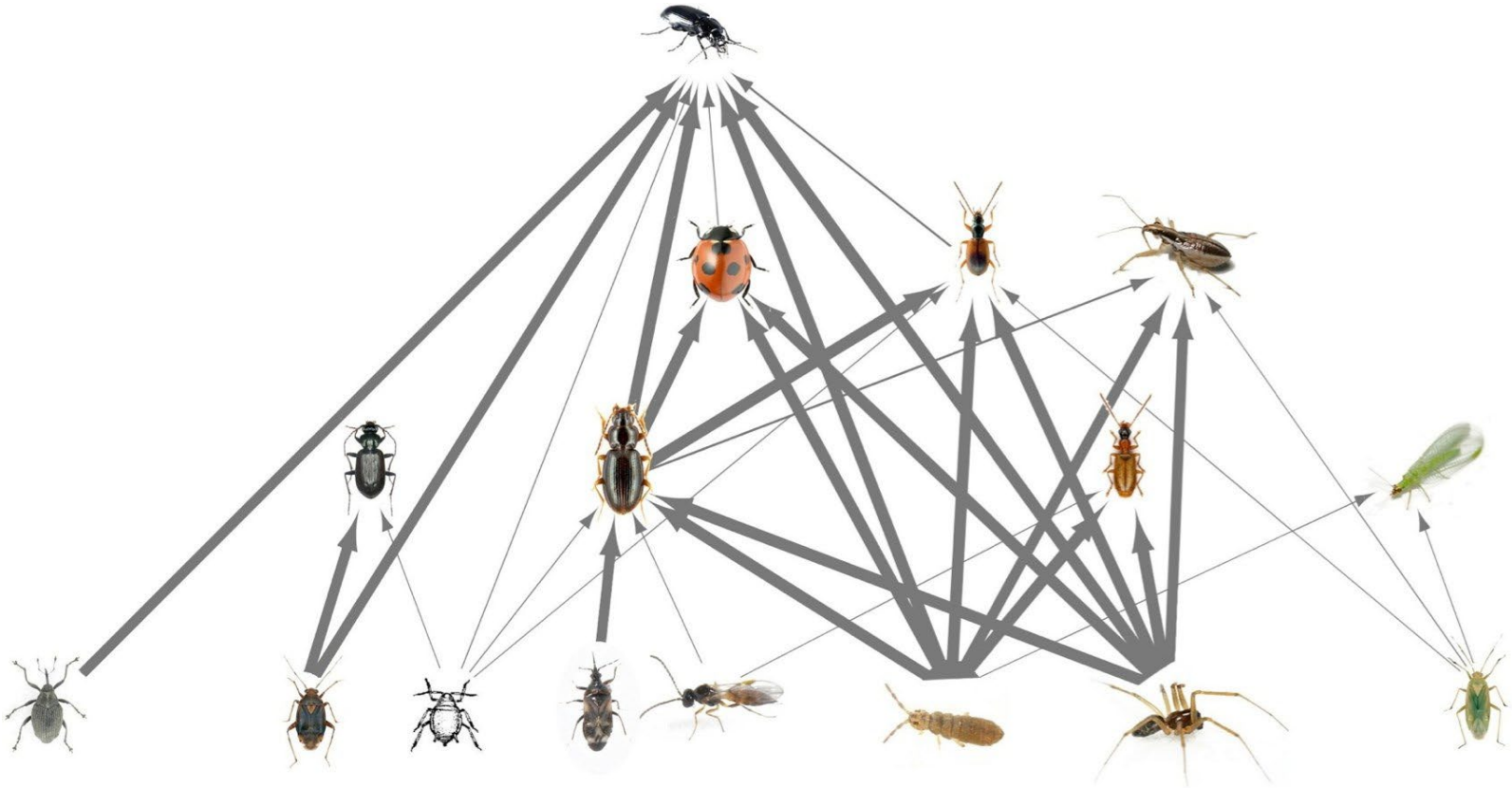


Davey et al. (2013). *Journal of Applied Ecology*. doi: 10.1111/1365-2664.12008

Functional grouping...

- Trophic behavior; Body size; Activity pattern.
- 182 species present ~ 17 functional types.
- We recreated a data-set of R-values, this time for functional groups rather than species

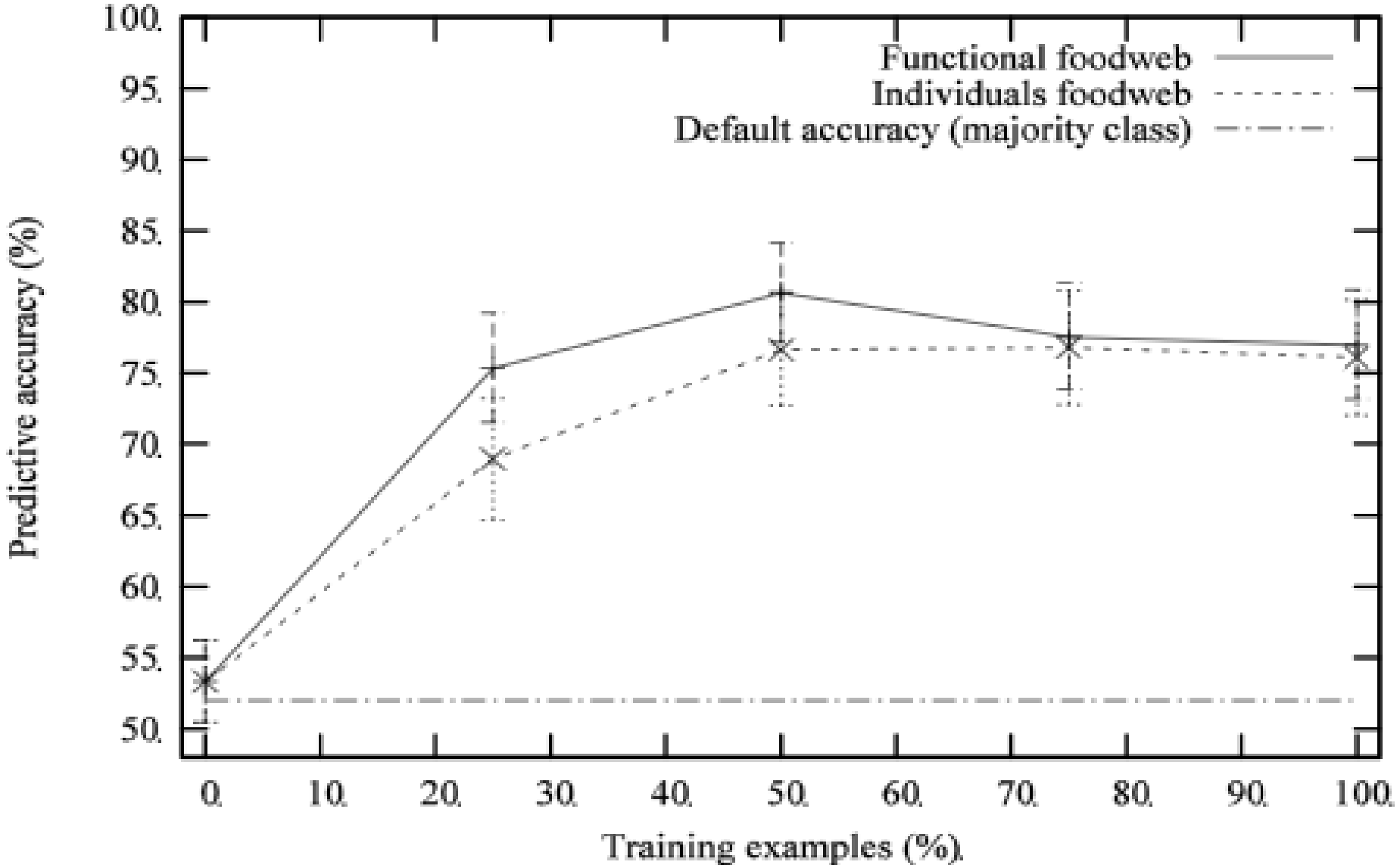
Hawes et al. (2009) *Basic and Applied Ecology* 10, 34-42. Caron-Lormier et al. (2009) *Ecological Modelling* 220, 1935–1949



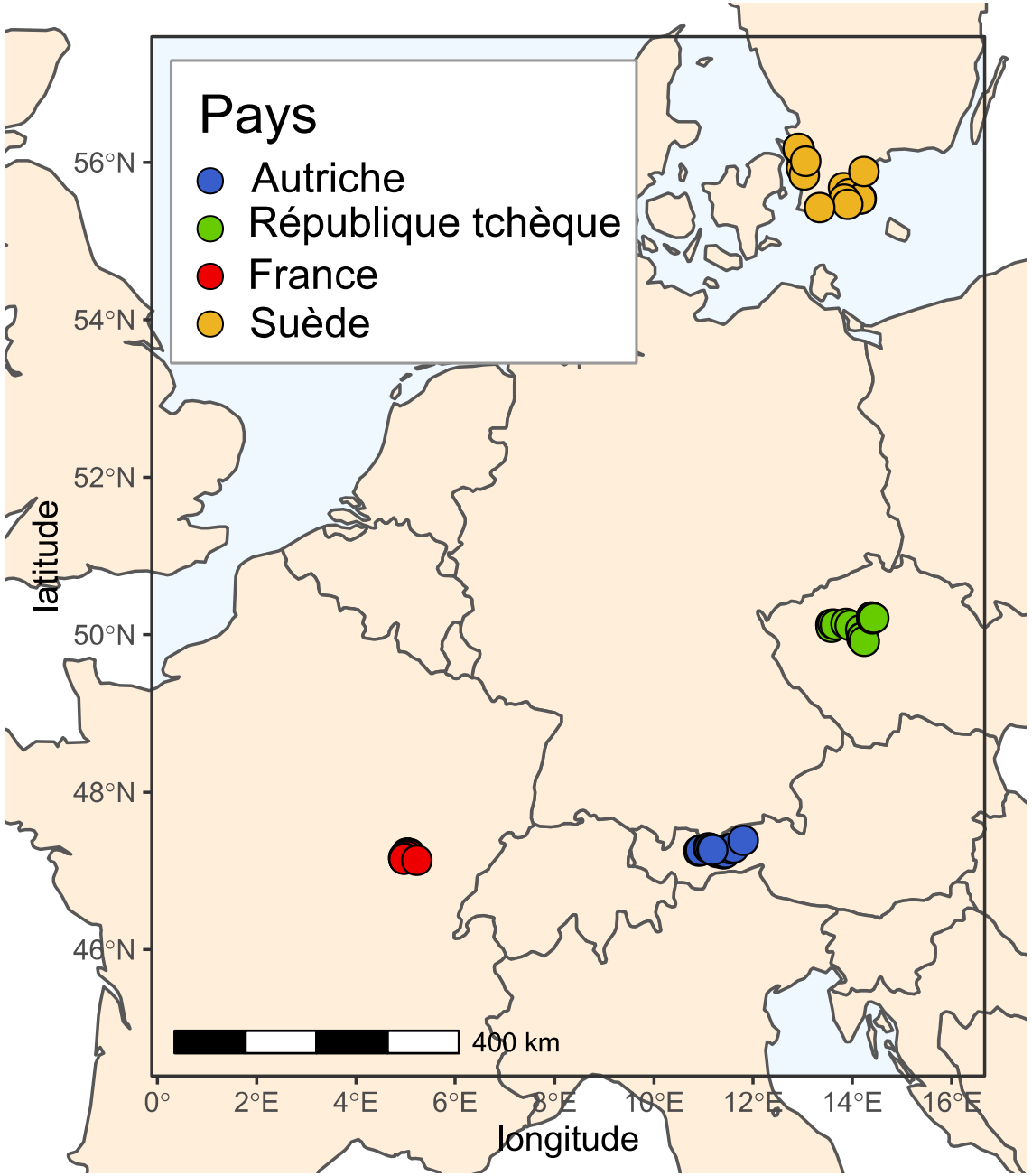
Preserves the structure found previously - detritivores at bottom, carabids at the top and lots of intra guild predation

Tamaddoni-Nezhad et al. 2013. Construction and Validation of Food Webs Using Logic-Based Machine Learning and Text Mining. *Advances in Ecological Research*.

Overall learning curves for functional vs individual food-webs (Vortis)



Tamaddoni-Nezhad et al. 2013. Construction and Validation of Food Webs Using Logic-Based Machine Learning and Text Mining. *Advances in Ecological Research*.


Pays
Autriche
République tchèque
France
Suède
56°N
54°N
52°N
50°N
48°N
46°N
latitude
400 km
0°
2°E
4°E
6°E
8°E
10°E
12°E
14°E
16°E
longitude

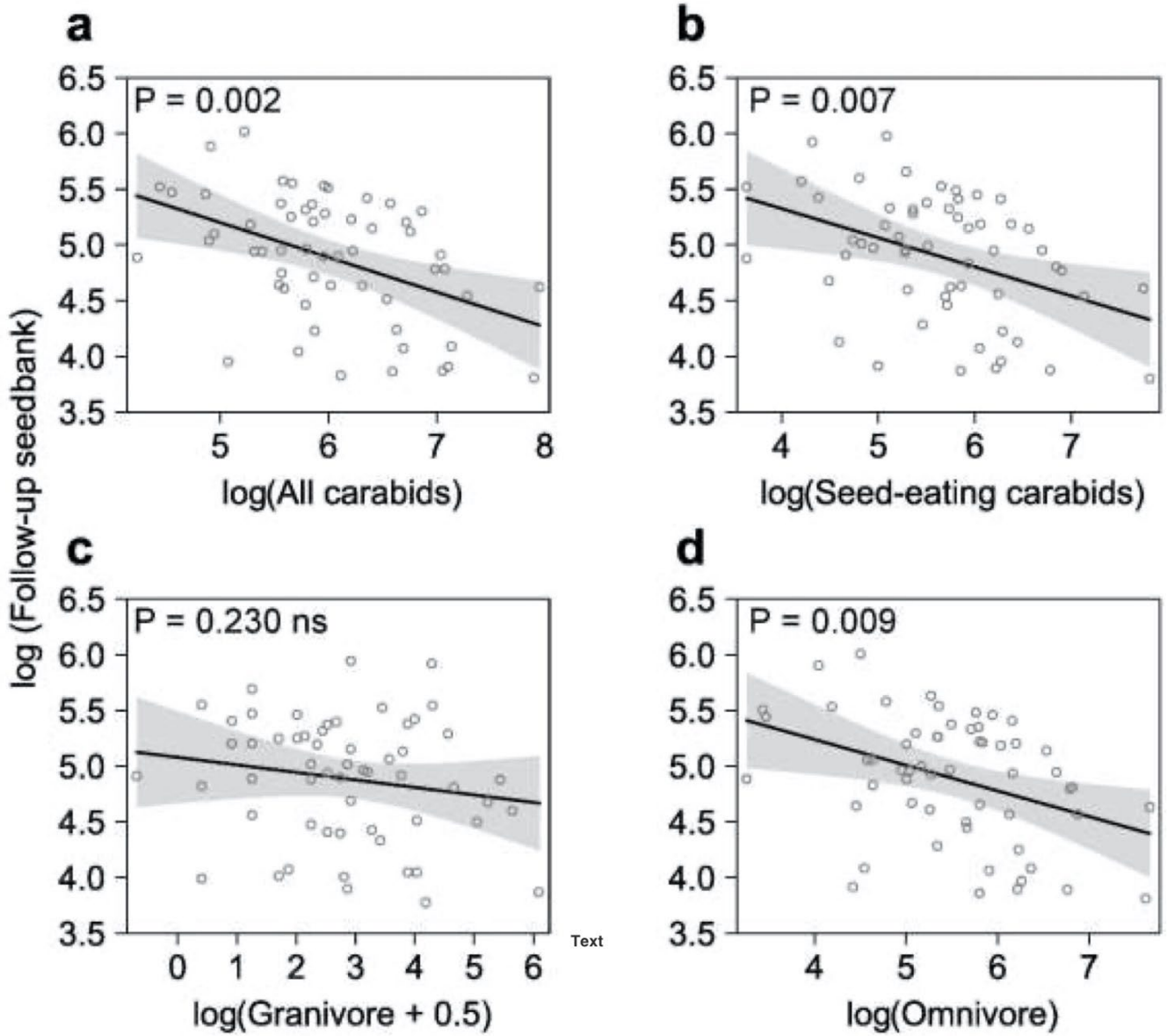
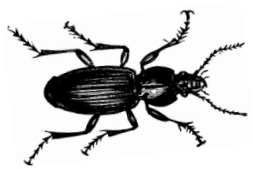
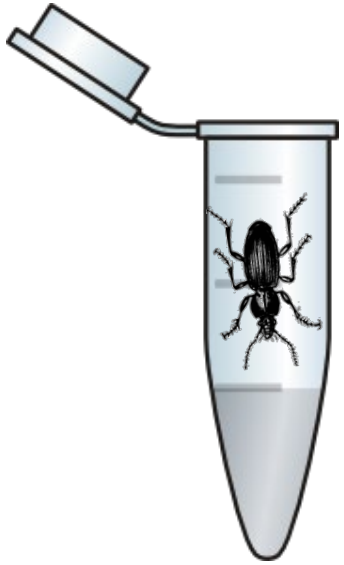
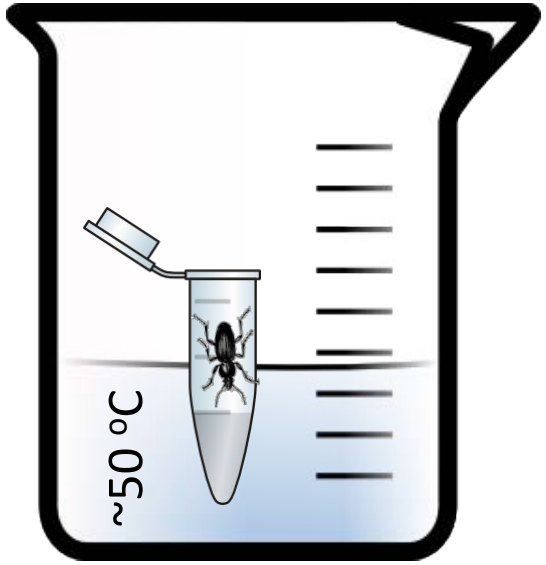


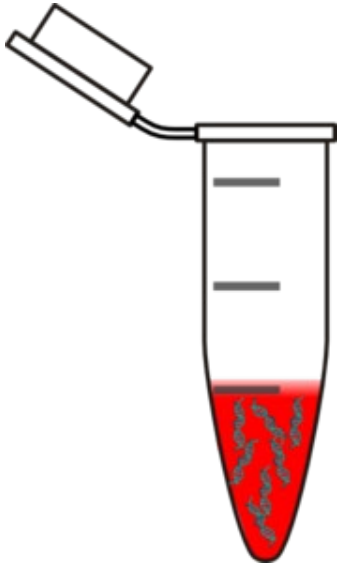
Figure 1. Multiple linear regression model fits to test the weed seedbank regulation by carabids. The log(Follow-up seedbank) is plotted against the log-transformed AD of: (a) all occurring carabid species; (b) seed-eating; (c) granivore; and, (d) omnivore carabids in session 2. The line represents the fixed-effect prediction and associated 95% confidence intervals (shaded), with the open circles being the partial residuals. Negative slopes indicate a regulatory effect of carabid AD on the seedbank. The plot is created using R version 3.6.1 and the package effects.

Do carabids eat weed seeds?



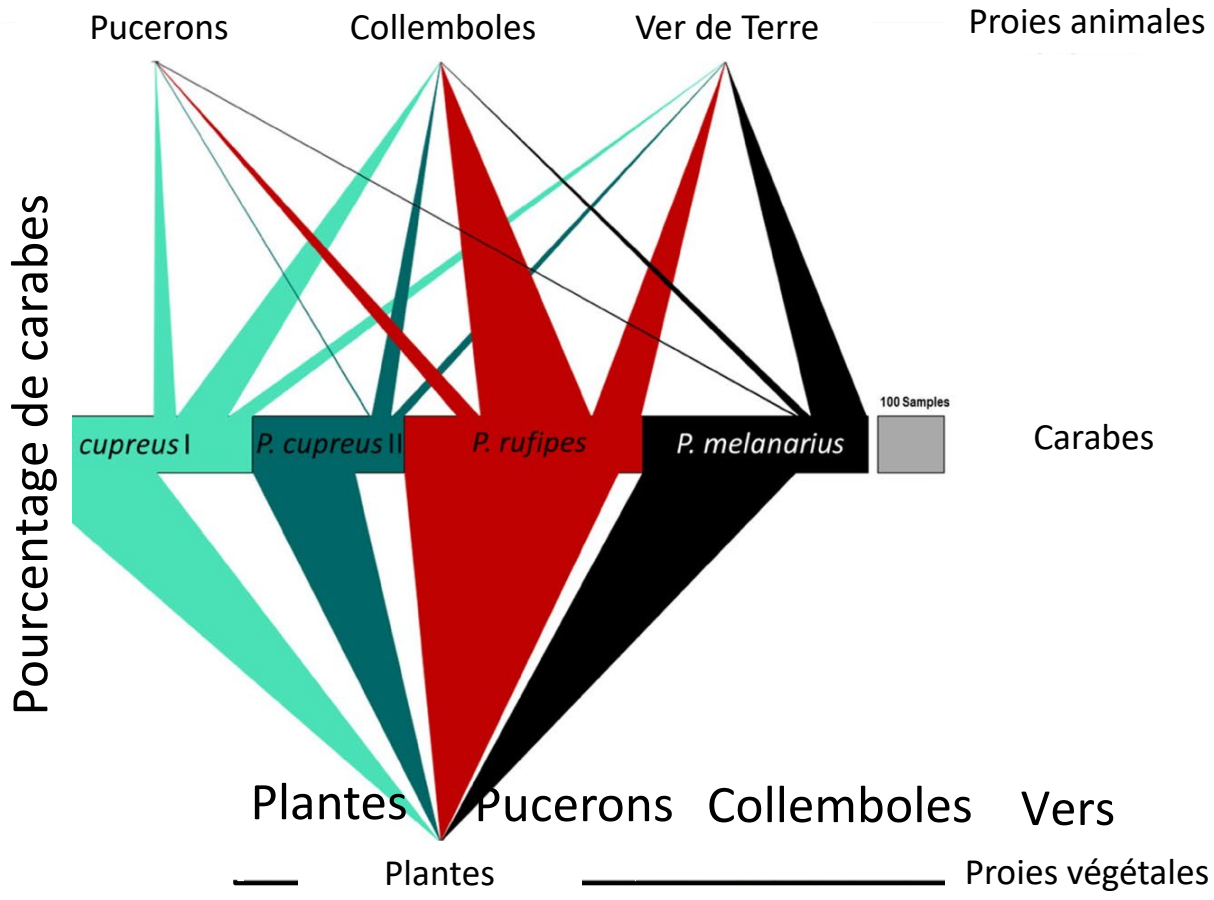


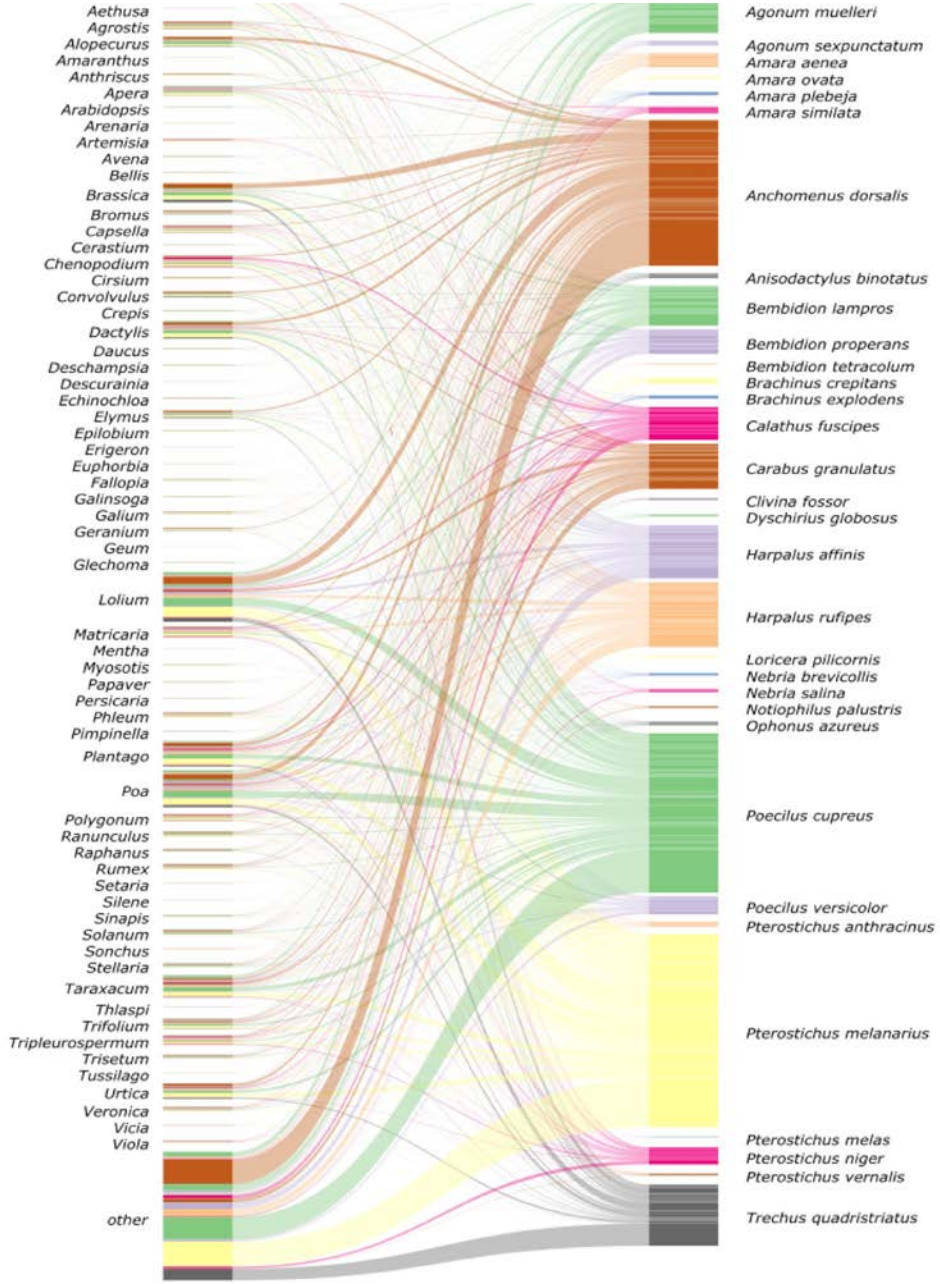




L’interaction trophique carabes / adventices

- Le régime alimentaire des carabes





BioAWARE

Resilient weed seed Biological Control