Learning ecological networks – benefits and perspectives

David A. Bohan (and colleagues)

UMR 1347 Agroécologie

INRA Centre de Dijon

CARTABLE – Colloque Apprentissage de Réseaux : de la Théorie aux Applications en Biologie et Ecologie.
12 -14 octobre 2016, Toulouse

Big picture

- We want to modify the world to do what we want:
  - More biodiversity or more crop productivity
- New management
  - Mitigate problems
  - Conservation
  - Production
  - Invasions / Diseases

This talk…

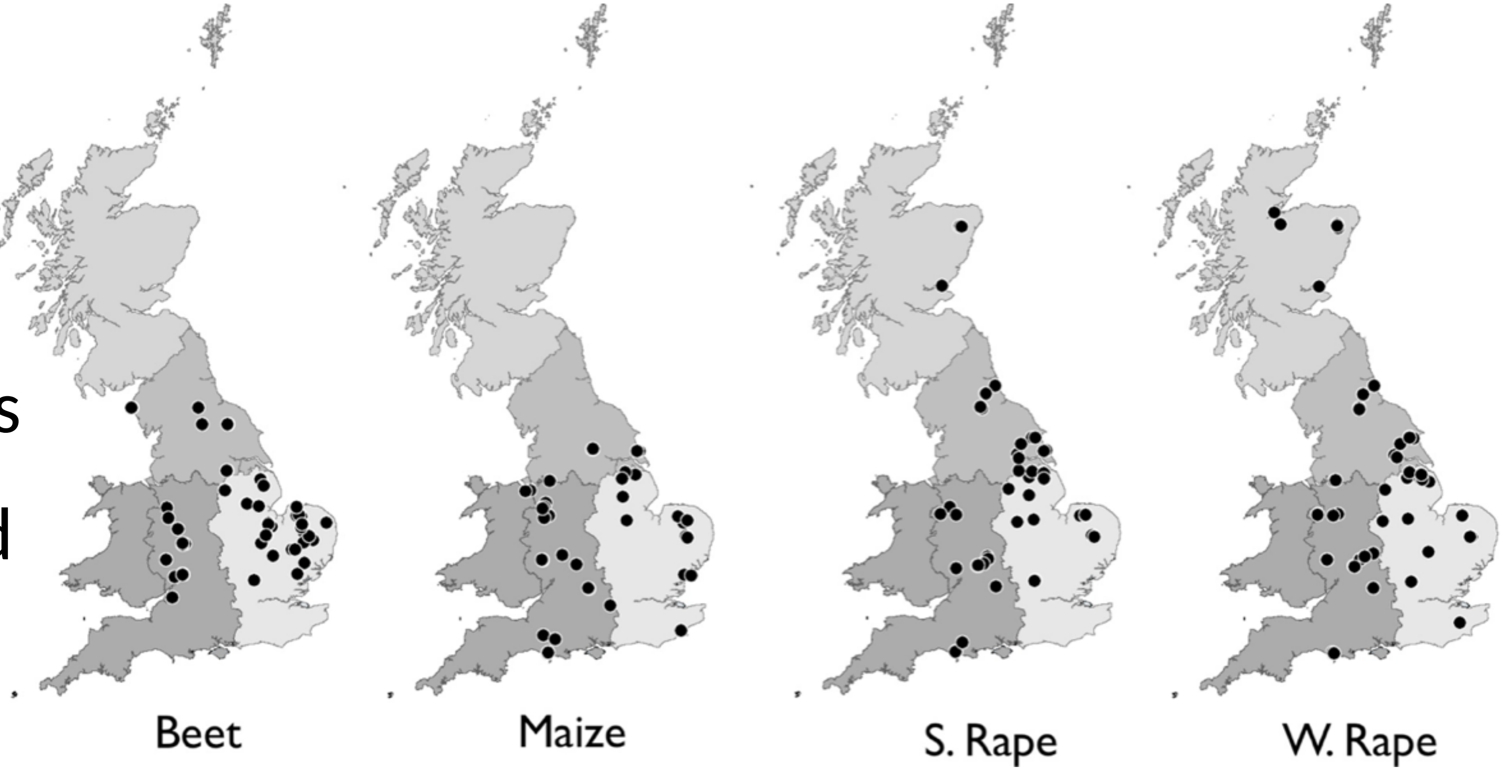
- What I want to do is tell a story
- I won’t detail learning approaches
- What I will try and do is identify interesting points and questions, and try to illustrate the importance of our work
- I’ll also try to identify some key gaps

Question

- Why can’t we detect, measure, understand and predict ecosystem change?

The FSE data

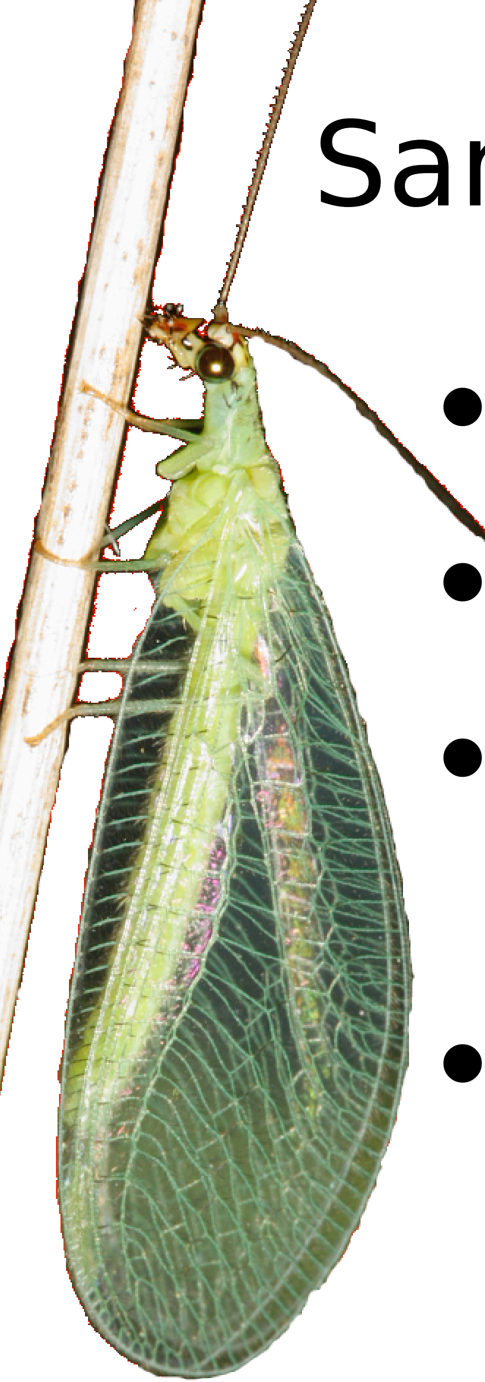
- 256 fields
- Split-field design



Sampling

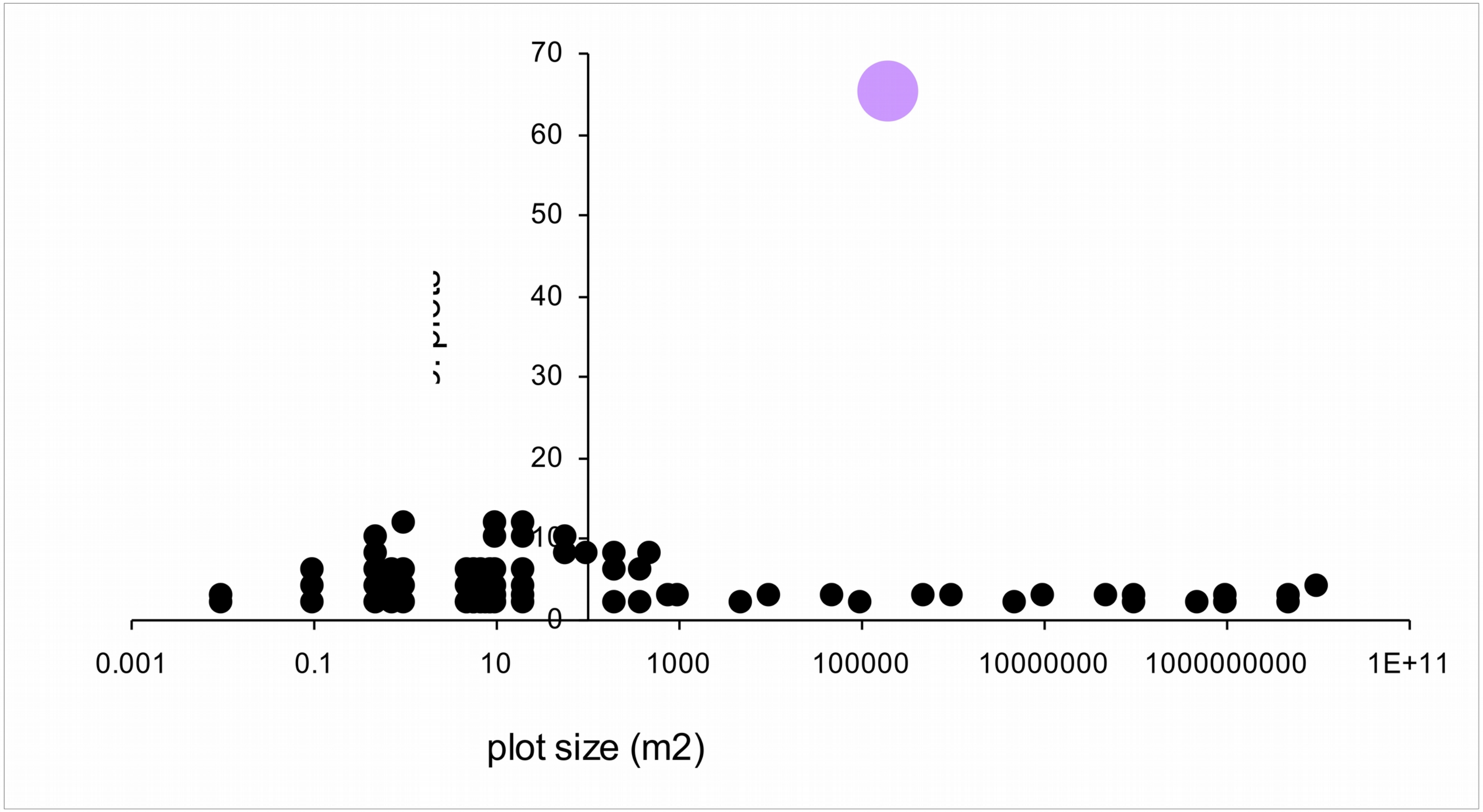
Biomass & Final count, *t*
Seedbank, *t*, *t*+1, *t*+2
Seedling, *t*, *t*+1, *t*+2
Early Spring, *t*
After-herbicide, *t*
Seed rain, *t*

Sampling



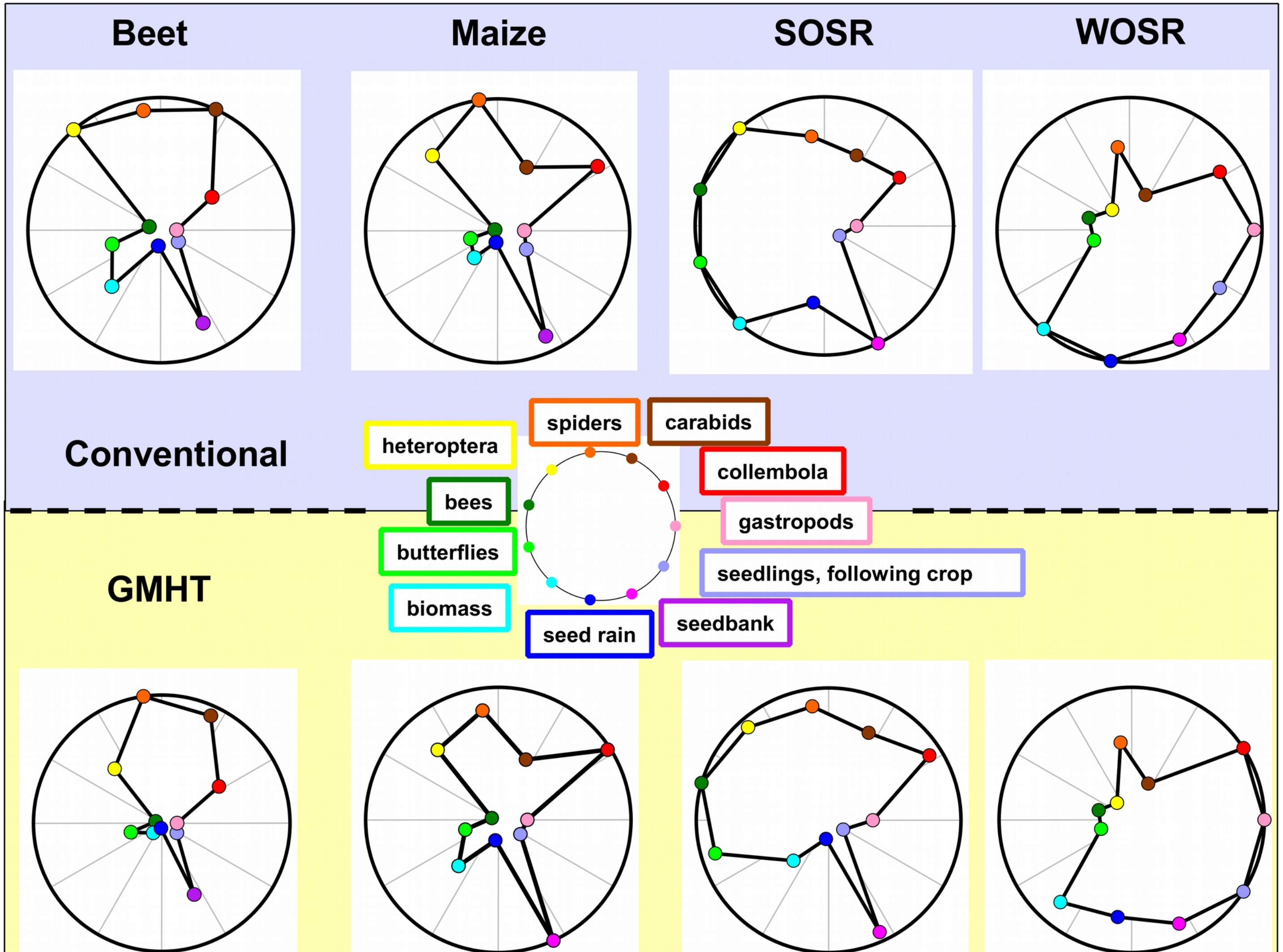
- 1.5 million weed plants counted
- 1 ton (dried) plant biomass sorted
- > 2.5 million invertebrates trapped
- > 1400 km of pollinator transects walked

Size of the FSE



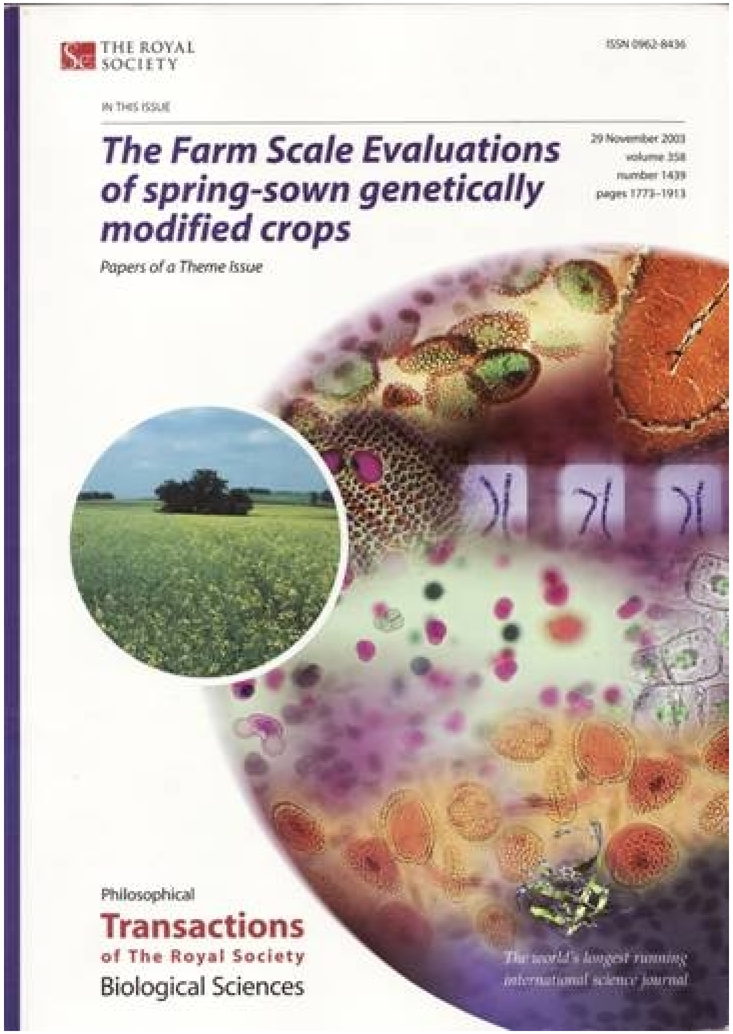
Size of the FSE versus 82 comparable ecological experiments. After Perry 03

Results



Beet
Maize
SOSR
WOSR
Conventional
GMHT
heteroptera
spiders
carabids
collembola
bees
gastropods
butterflies
seedlings, following crop
biomass
seed rain
seedbank

Impact of FSE




Editorial

A victory for reason

An astonishing ecology experiment has blown the lid off farming practices

IF BIG biology means sequencing genomes, and big physics is about testing theories in particle accelerators, what does big ecology involve? Answer: tramping around hundreds of muddy fields counting weeds and insects while enduring the wrath of protesters.

For the past four years, critics have queued up to take shots at Britain's farm-scale evaluation of three genetically modified crops, claiming it was at best a politically motivated irrelevance and at worst downright dangerous. Just about the only thing those for and against the trial could ever agree on was that this was ecological research on an unprecedented scale. But now the results are in (see pages 21 and 46). So who are the winners and losers in the debate about the science?

The most obvious loser is the biotech industry. Contrary to its repeated claims that herbicide-resistant crops are kinder to wildlife, two out of the three tested were worse than conventional crops, while doubts remain about the third. The green groups opposed to the trial are winners in some respects but losers in others. Winners because of what the trial discovered, losers because throughout they claimed the experiment was biased and the researchers were in the pocket of the biotech industry. It wasn't, and they weren't. As the findings make clear, this was impartial research carried out without fear or favour.

The British government, too, is both winner and loser. It loses because it is hard to see how it can now fulfil its ambition of opening Britain up to GM crops any time soon. It wins because it can share the credit for backing an evidence-based approach that has delivered genuinely impressive findings. The experiment will not help the cause of high-tech crop science, but what a superb advert it is for British ecology.

Clear winners include the scientists who carried out the trial, often in the face of adversity, and Brian Johnson of English Nature, the British government adviser whose concerns about herbicide-resistant crops led to the trial. Trust in government scientists has been at a low ebb since BSE. The success and evident integrity of the trial can only raise it.

Policy makers, the public and the media now have to face up to the trial's compelling evidence that the herbicide treatments farmers use, as well as the crops they choose to grow, can dramatically influence farmland wildlife. The implications go way beyond the GM debate. The research proves that countries like Britain do not have to blow in the wind. They can use science to choose the level of farmland biodiversity they want and then set policies that reward farmers for adopting the practices that deliver it. The end of the farm trial could, if all concerned get it right, signal the beginning of a new and rational way of managing land. The British government – and Europe – should seize the chance. ◆

www.newscientist.com

25 October 2003 | NewScientist | 3

"They (policy-makers) can use science to choose the level of farmland biodiversity they want…

New Scientist Editorial, 25-Oct-03

Interpretation

Table 2. Half-field whole-season mean counts of bees in conventional (C) and GMHT beet, maize and spring oilseed rape, and their respective treatment effects.
(Multiplicative treatment ratio, $R = 10^d$, where d is the mean of the differences between GMHT and C treatments on the logarithmic scale; confidence limits for R are back-transformed from those for d. CI, confidence interval.)

			geometric mean			
crop and taxa	period	*n*	C	GMHT	*R* (95% CI)	*p*-value
beet						
total bees	year	20	3.62	1.55	0.55 (0.31–0.99)	0.05*
Apis mellifera	year	7	4.73	0.55	0.27 (0.20–0.36)	0.03*
bumble-bees	year	18	2.58	1.07	0.58 (0.31–1.07)	0.09
long-tongued bees	year	5	1.71	0.00	0.37 (0.31–0.44)	0.10
maize						
total bees	year	15	1.14	2.09	1.44 (0.58–3.57)	0.41
A. mellifera	year	3	0.71	4.24	3.07 (0.01–1136.41)	0.49
bumble-bees	year	14	1.02	2.12	1.55 (0.65–3.65)	0.32
spring oilseed rape						
total bees	year	62	44.28	36.52	0.83 (0.66–1.05)	0.13
A. mellifera	year	51	10.95	9.16	0.85 (0.57–1.28)	0.44
bumble-bees	year	62	27.38	21.58	0.80 (0.63–1.00)	0.06
long-tongued bees	year	38	2.68	2.02	0.82 (0.61–1.10)	0.16

* $p < 0.05$.

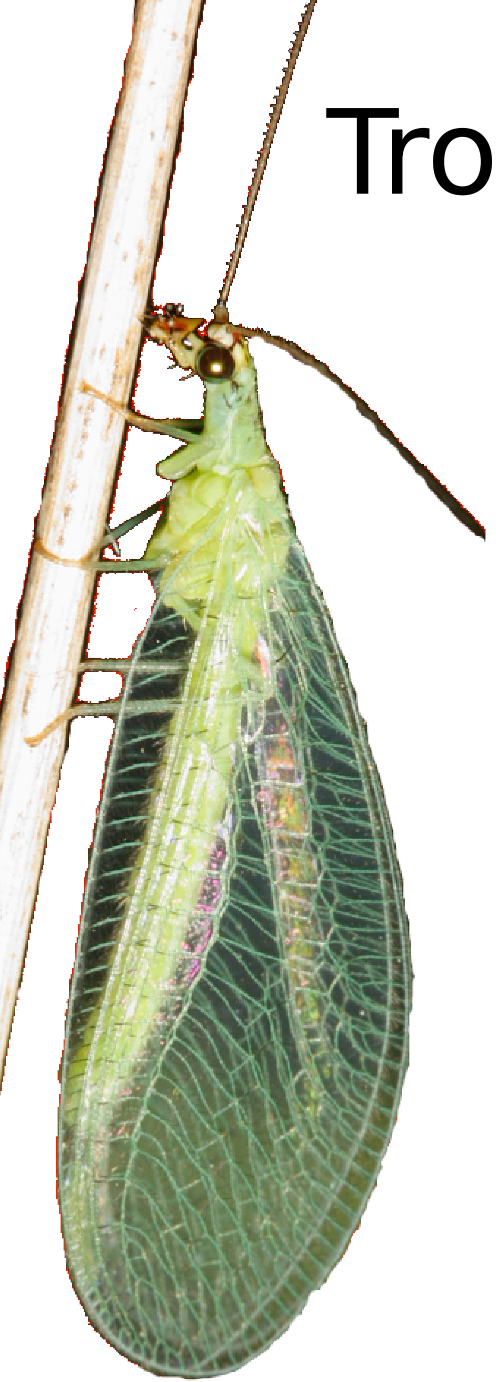
Interpretation

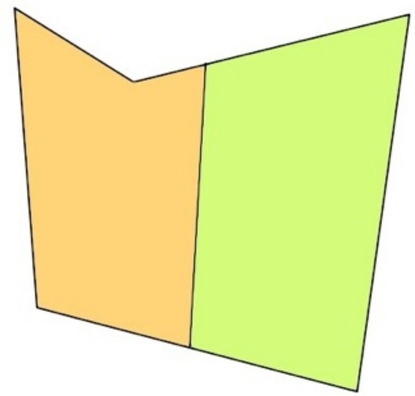
- Changes like this seen in all systems with any new management
  - Will all be interpreted as ‘bad’
- I would argue that we hadn’t really answered the question of whether the ecosystem had changed

Benefits of networks

- The world is complex and is not well reflected by simple approach of the FSE
- Networks are one way of representing, analysing and understanding this complexity
- Problem is that they are 'expensive' to produce
  - This is where 'learning' from data really comes into play.

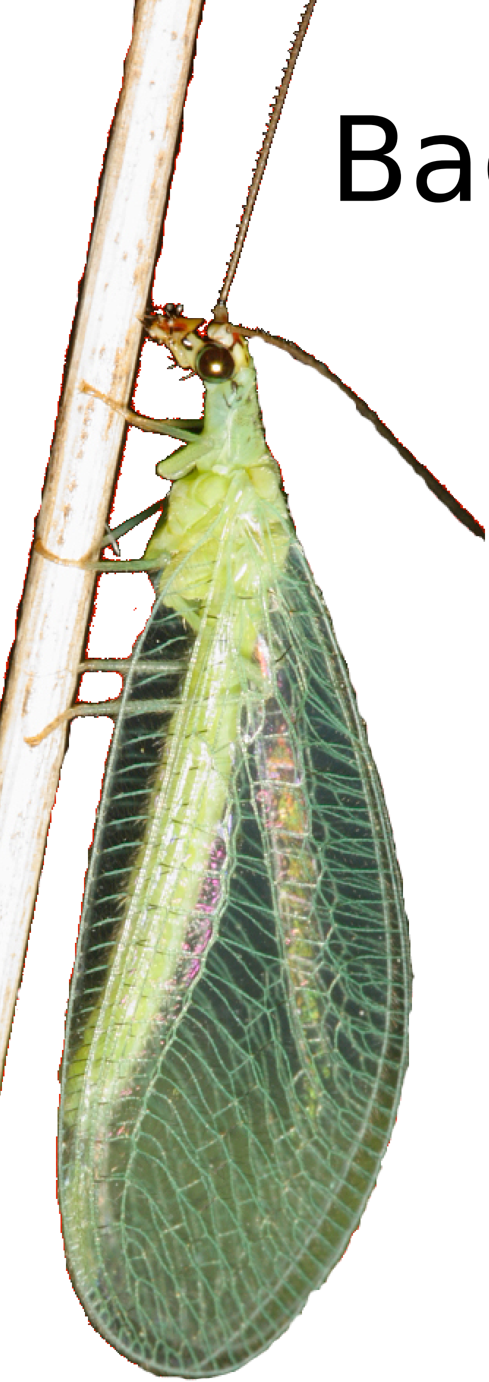
Trophic model - data



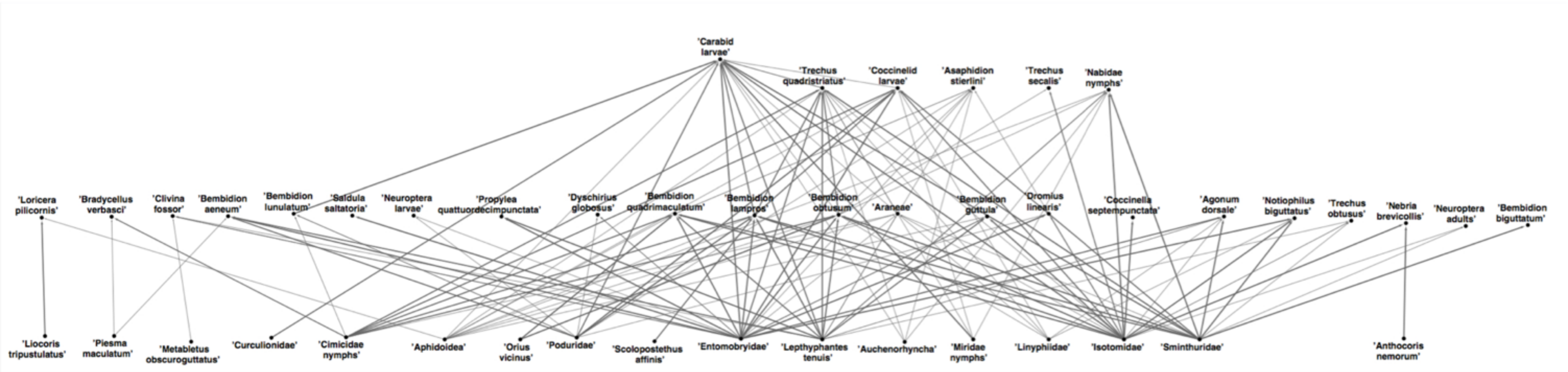


- Herbicide removes food and shelter - something to eat and somewhere to sit
- Species Y will move to new habitat or die - measured as R_Y, where $R_Y = \log_{10}(Y_{GM}/Y_C)$
- Species X, which feeds on Y, will change with Y as R_X
- Expectation that: R_X is correlated with R_Y

Background information



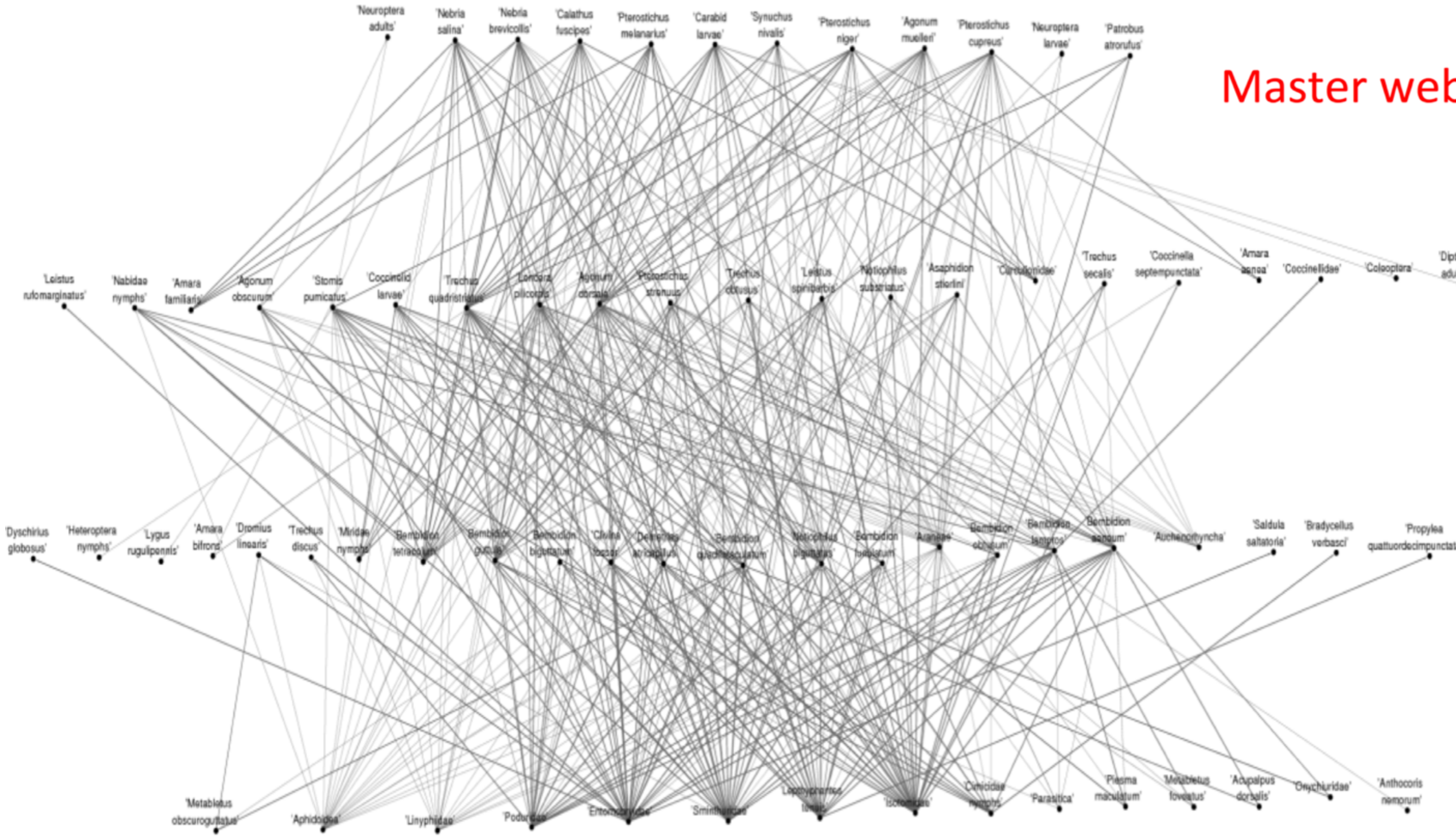
- Appropriate mouthparts for feeding
- In any sample Y and X should co-occur
- Big things eat small things
- With this set of ‘rules’ we ‘learn’ food webs



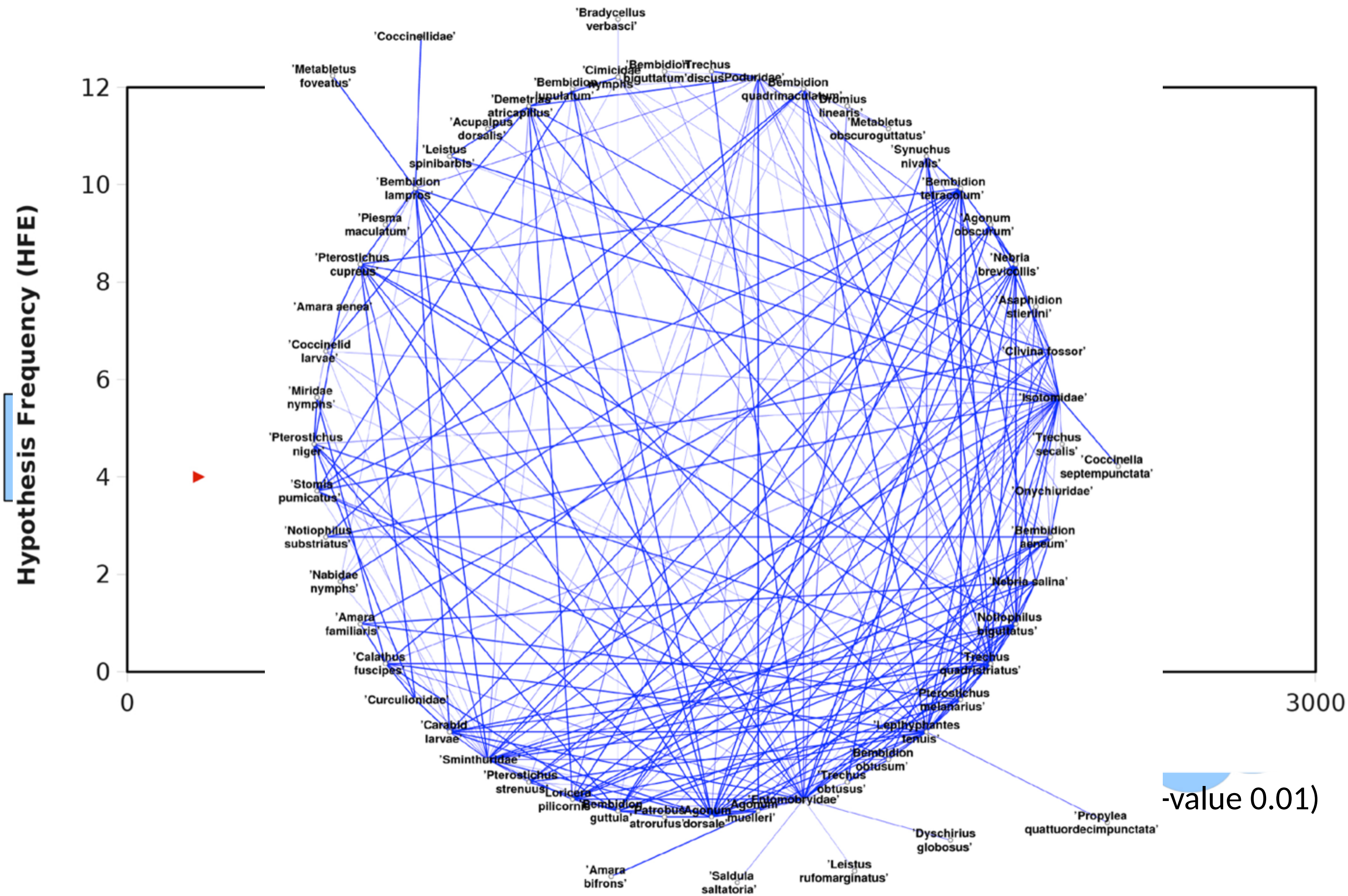
- 45 invertebrate species or taxa (~25%), but about 74% of the individuals were linked
- Collembola important prey. Carabid beetles were the dominant predators. Carabid larvae predators of a wide variety of prey.
- Lots of intraguild predation

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

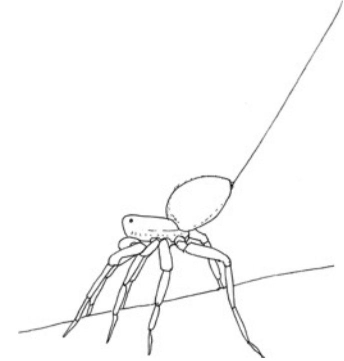
Master web


'Neuroptera adults'
'Nebria salina'
'Nebria brevicollis'
'Calathus fuscipes'
'Pterostichus melanarius'
'Carabid larvae'
'Synuchus nivalis'
'Pterostichus niger'
'Agonum muelleri'
'Pterostichus cupreus'
'Neuroptera larvae'
'Patrobus atrorufus'
'Leistus rufomarginatus'
'Nabidae nymphs'
'Amara familiaris'
'Agonum obscurum'
'Stomis pumicatus'
'Coccinelid larvae'
'Trechus quadristriatus'
'Loricera pilicornis'
'Agonum dorsale'
'Pterostichus strenuus'
'Trechus obtusus'
'Leistus spinibarbis'
'Notiophilus substriatus'
'Asaphidion stierlini'
'Trechus secalis'
'Coccinella septempunctata'
'Amara aenea'
'Coccinellidae'
'Coleoptera'
'Diptera adults'
'Dyschirius globosus'
'Heteroptera nymphs'
'Lygus rugulipennis'
'Amara bifrons'
'Dromius linearis'
'Trechus discus'
'Miridae nymphs'
'Bembidion biguttatum'
'Demetrias atricapillus'
'Bembidion quadrimaculatum'
'Notiophilus biguttatus'
'Bembidion lampros'
'Bembidion aeneum'
'Auchenorrhyncha'
'Saldula saltatoria'
'Bradycellus verbasci'
'Propylea quattuordecimpunctata'
'Metabletus obscuroguttatus'
'Aphididae'
'Linyphiidae'
'Poduridae'
'Lepthyphantes tenuis'
'Parasitica'
'Piesma maculatum'
'Metabletus foveatus'
'Acupalpus dorsalis'
'Onychiuridae'
'Anthocoris nemorum'

Automatic literature verification

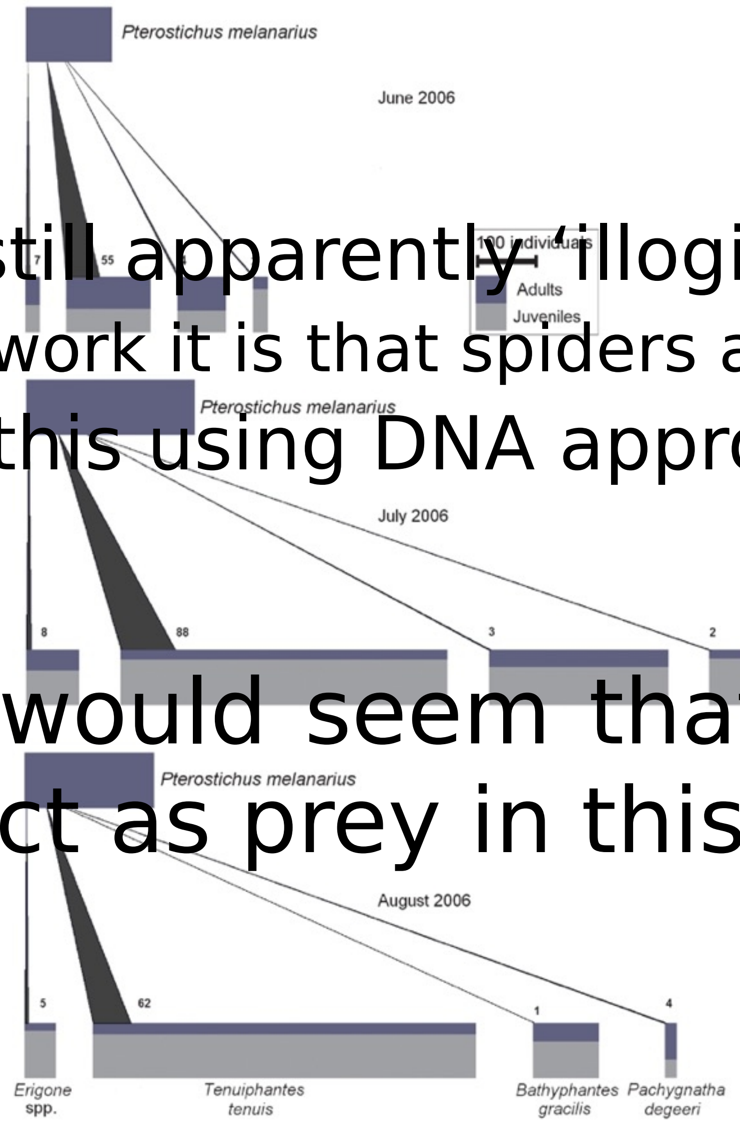


Problems



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

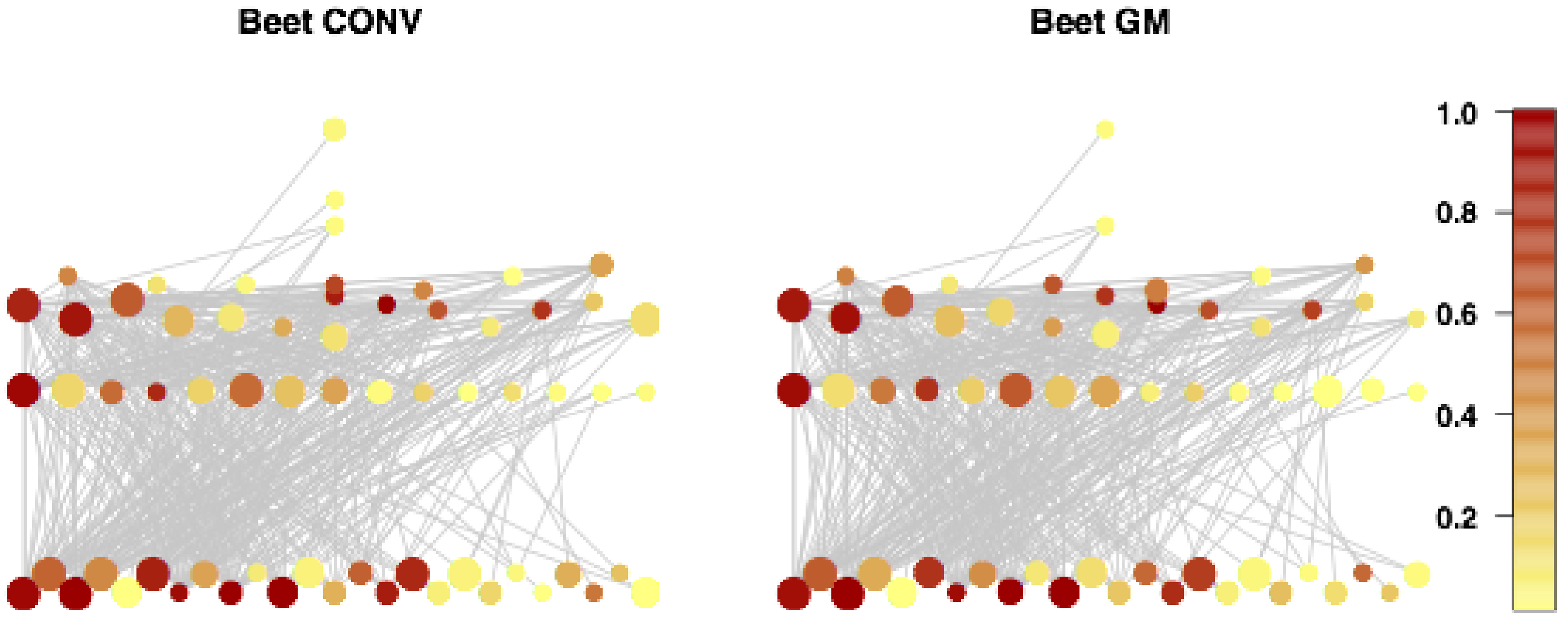
- and… it would seem that spider do indeed act as prey in this network



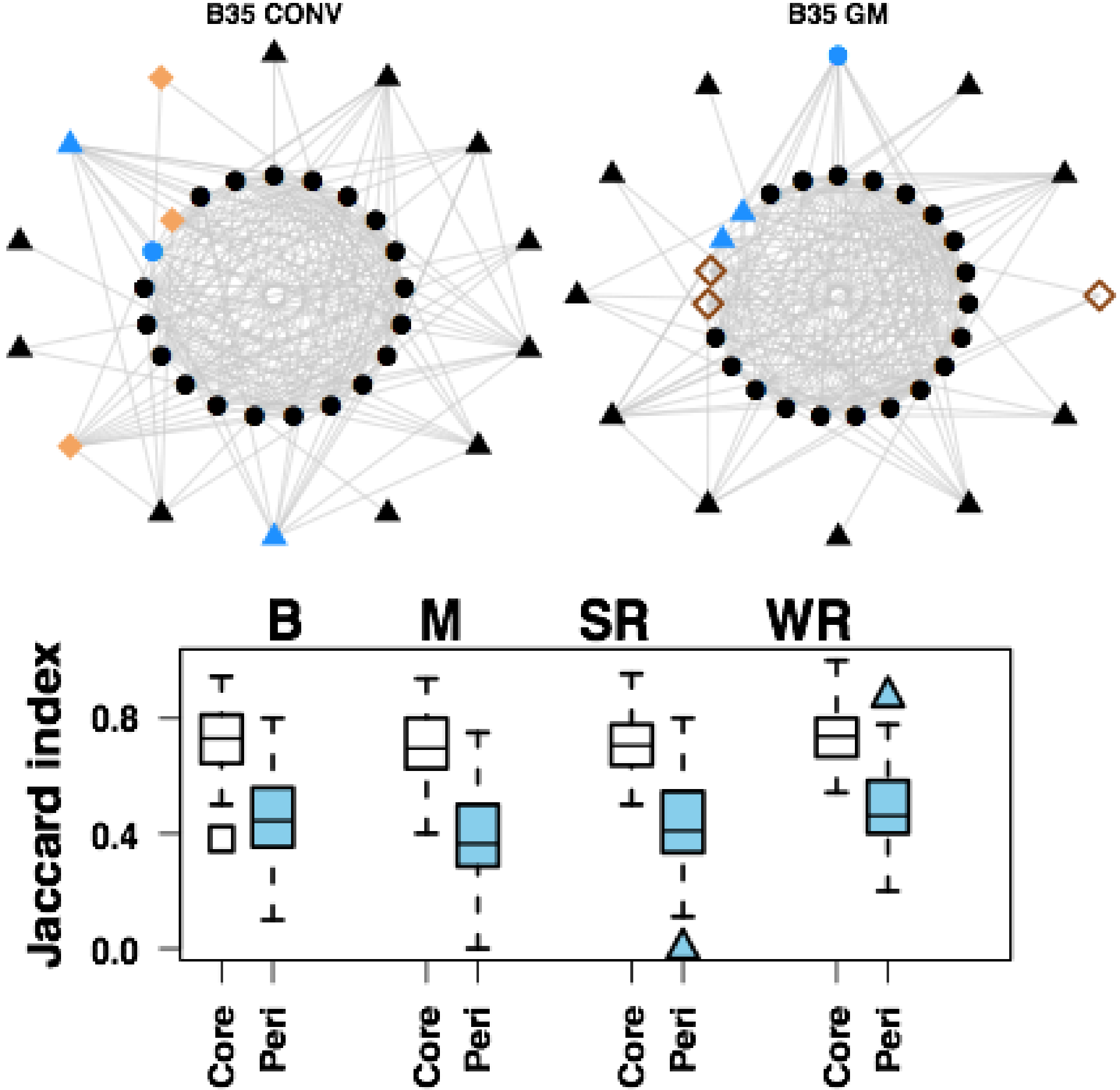
Davey et al. (2013). *Journal of Applied Ecology*. doi: 10.1111/1365-2664.12008

Key point

- The learning is not ***just*** doing the expected reconstruction – text mining the data
- It is suggesting a reduced set of hypotheses for testing
- Moreover it is suggesting "Novel" hypotheses that were validated
- It is doing genuine science

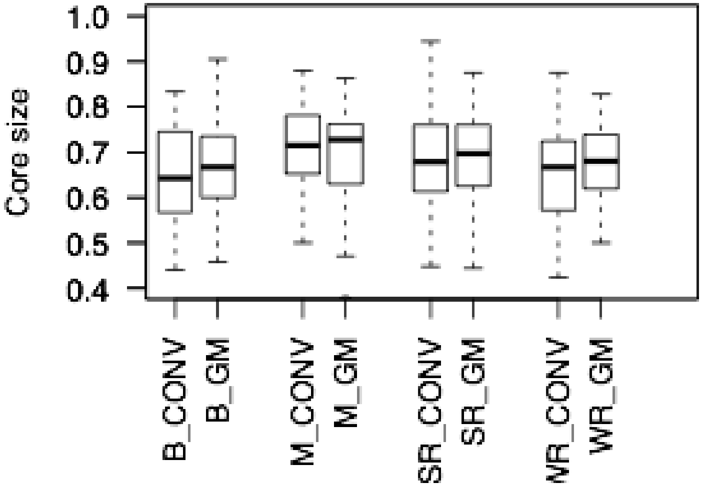


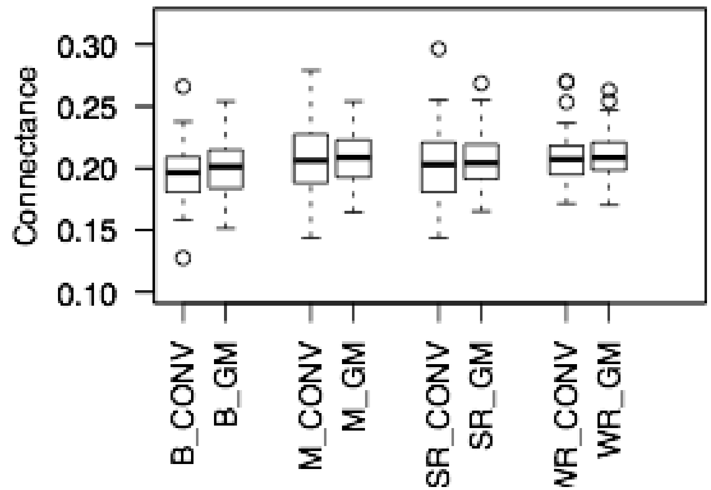
Beet CONV
Beet GM
1.0
0.8
0.6
0.4
0.2

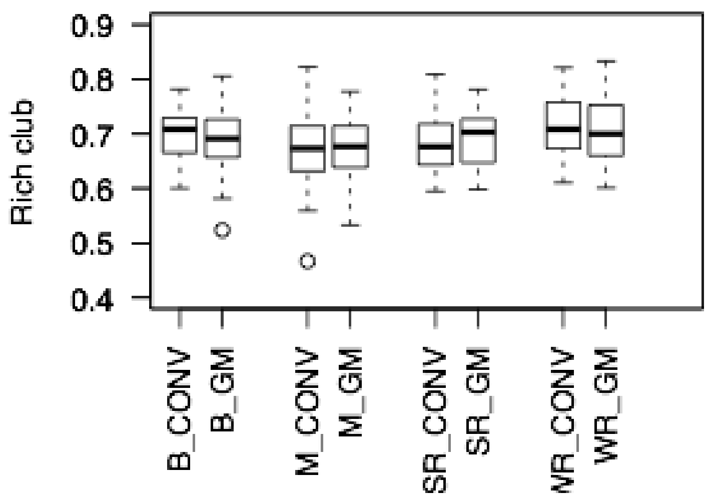


B35 CONV
B35 GM
B
M
SR
WR
Jaccard index
0.8
0.4
0.0
Core
Peri
Core
Peri
Core
Peri
Core
Peri

Network statistics







Key point

- Learning from existing data can produce management-important networks

- We can test for ecosystem change due to management

Next Generation Biomonitoring

- Current biomonitoring approaches use indicators (chemical species, biological species, behaviour) as proxies of effect
- But:
  - Limited accuracy (don’t capture complexity)
  - Costs (limited scale and slow)
  - Generality (system specific and don’t add to ecological

done with Corinne Vacher, Alireza Tamaddoni-Nezhad, Alex Dumbrell and Guy Woodward

Next Generation Biomonitoring

- The approaches reflect methods and techniques that date to the 1950s
- We believe that a new biomonitoring approach is long overdue that can
    - solve these problems
    - produce indicator values (fits in current biomonitoring framework)

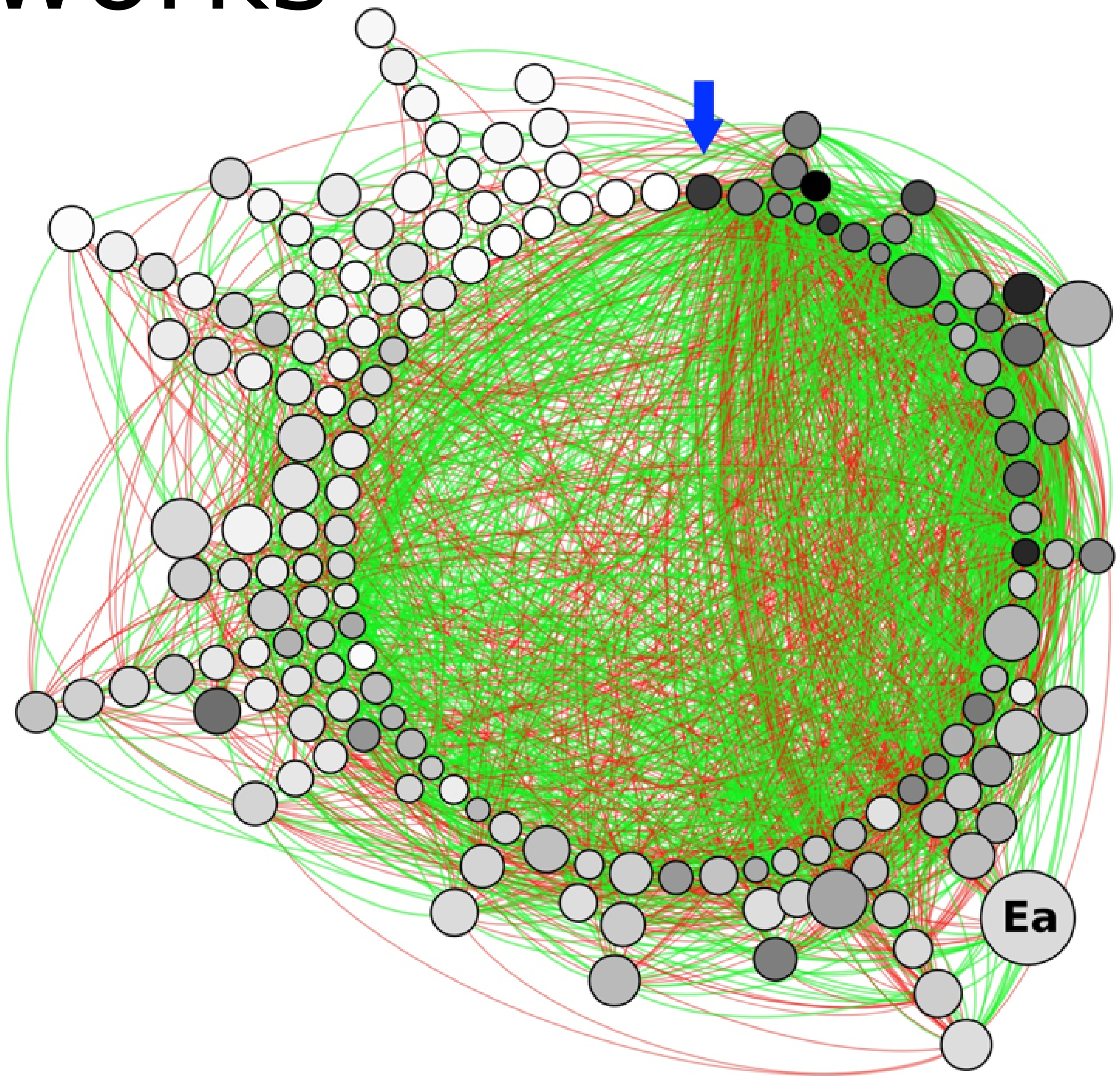
Next Generation Biomonitoring

- The idea is really simple
- Nucleic acids (DNA/RNA) are ubiquitous, and nucleic acid sample data contains information we can learn on:
    - species (OTUs)
    - ecological functions
- Next generation sequencing

Case study – oak tree pathogens

- Corinne Vacher was to be discussing this
- Aim to build microbial interaction networks, on oak tree leaves, subject to invasion by a pathogen - *Erysiphe alphitoides*
- DNA sampled from leaves across a number of trees
- NGS performed
- Co-occurrence network built

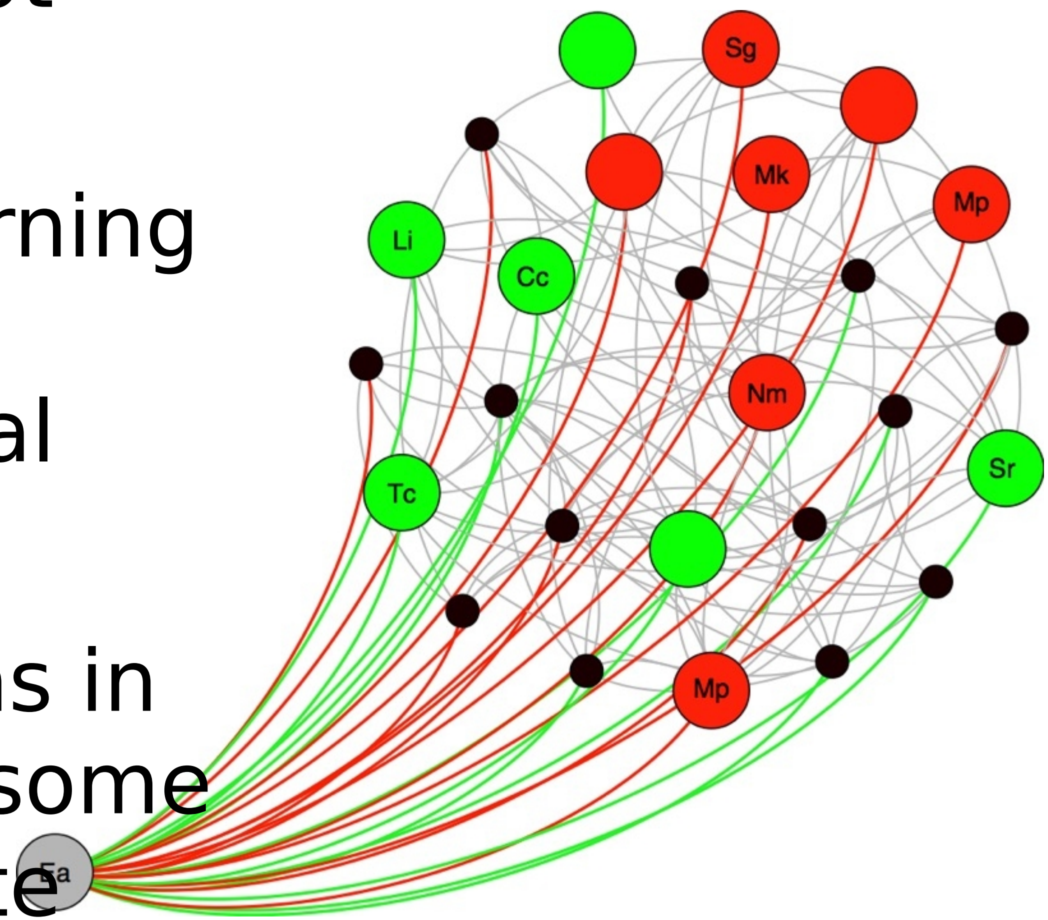
Reconstructing NGS networks



Jakuschkin, Fievet, Robin and Vacher (2016) Microbial Ecology.

Reconstructing NGS networks

- But, this is not function
- Bayesian learning and environmental preference
- What happens in invasion; do some OTUs facilitate invasion; and are



Jakuschkin, Fievet, Robin and Vacher (2016) Microbial Ecology.

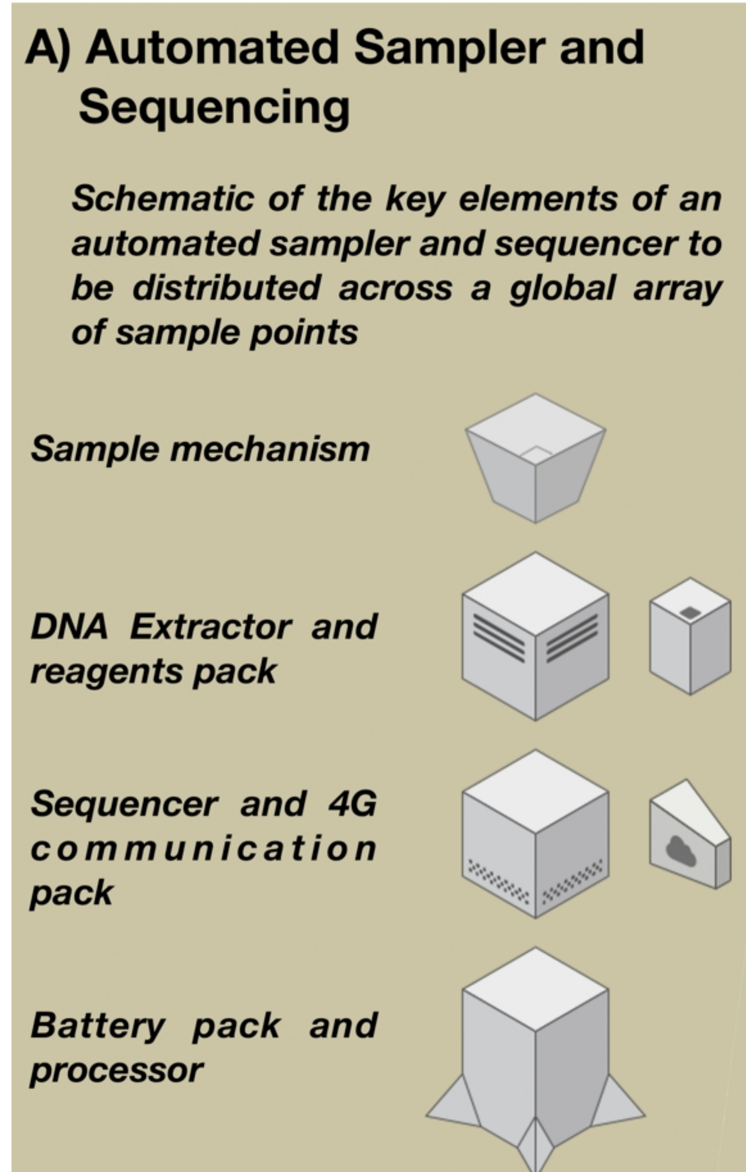
Key point

- Learning from NGS data can reconstruct networks of OTUs
- Moreover, it can also start to recover the mechanistic processes (the functions) structuring the community

- It does ecology

Global next-generation biomonitoring

- What we envision are automatic samplers
- Based on existing technology

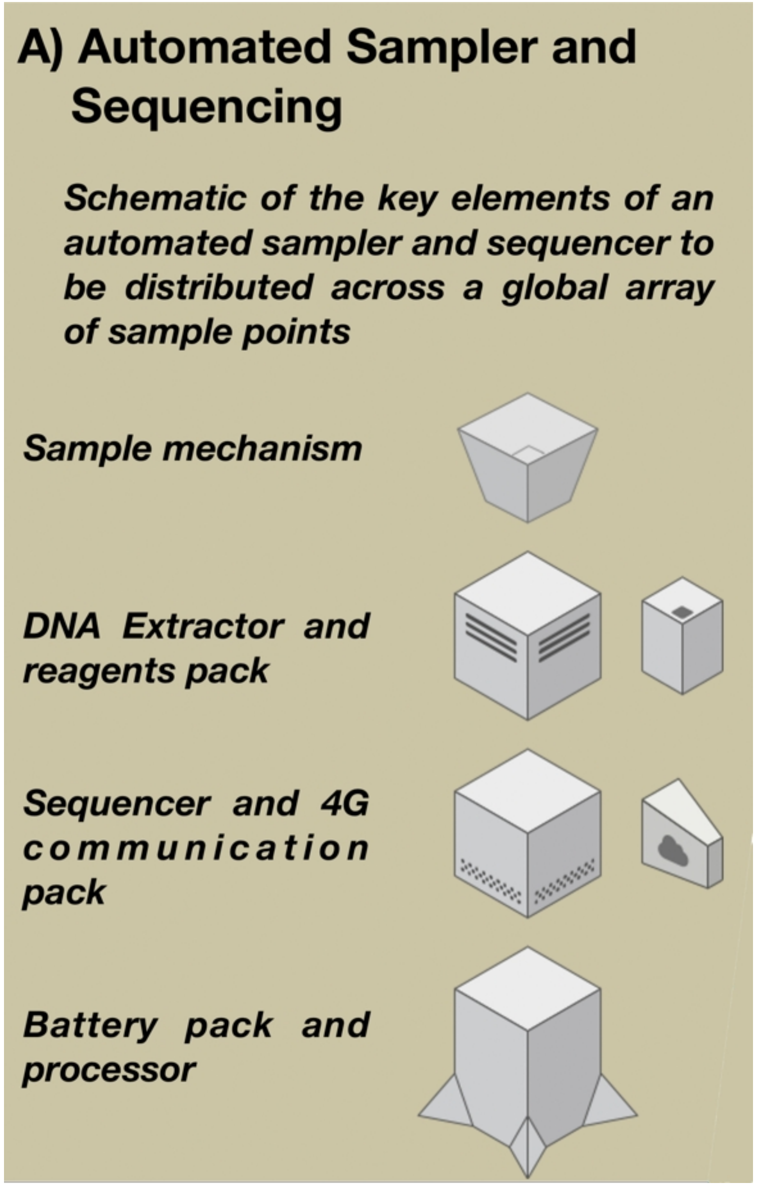


Global next-generation biomonitoring

- Existing technology – MinION
- Sampler possibly the biggest hurdle

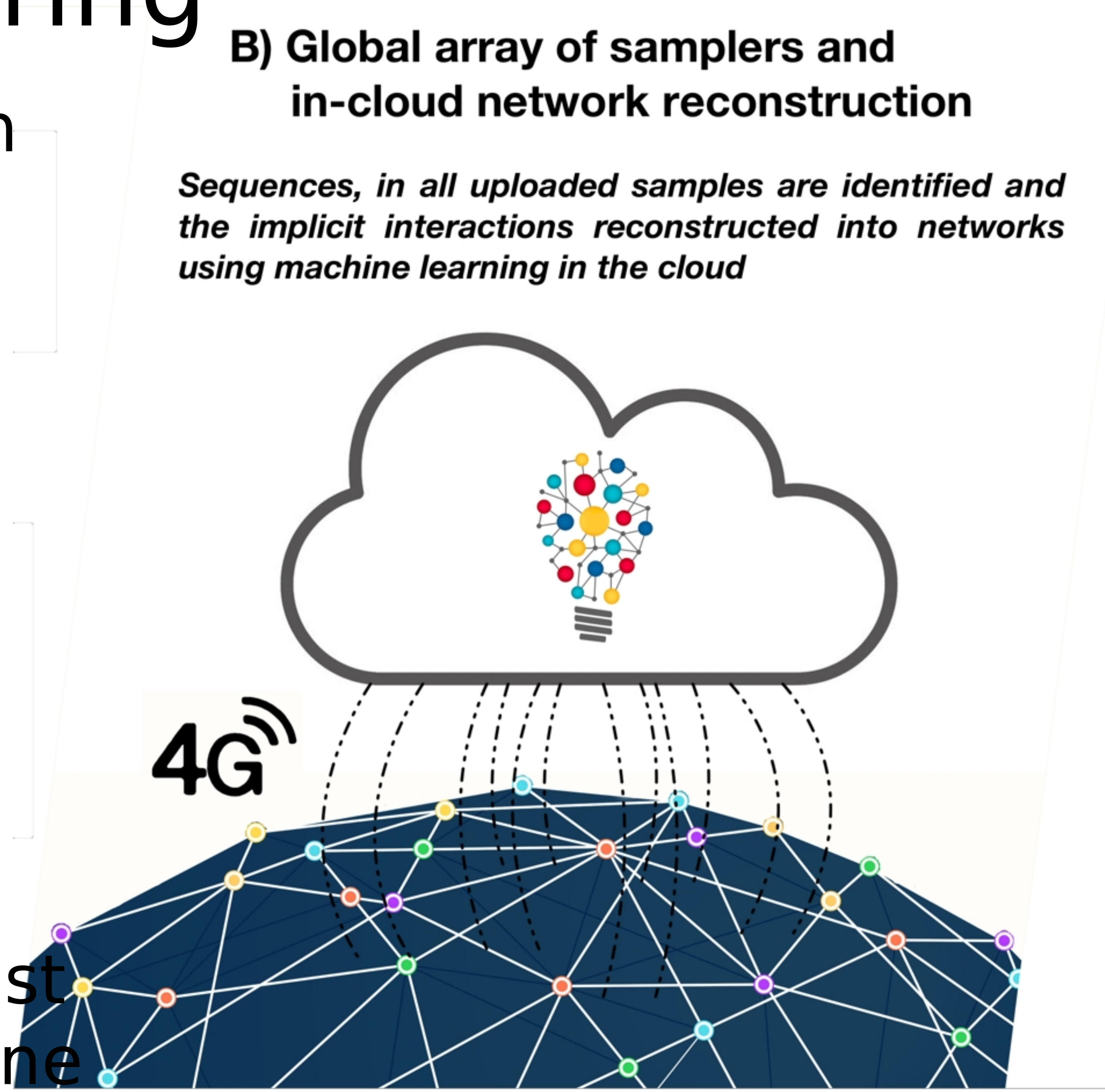
Global next-generation biomonitoring

- What we envision are automatic samplers
- Based on existing technology
- Sampling hourly, daily, weekly… replicate samples



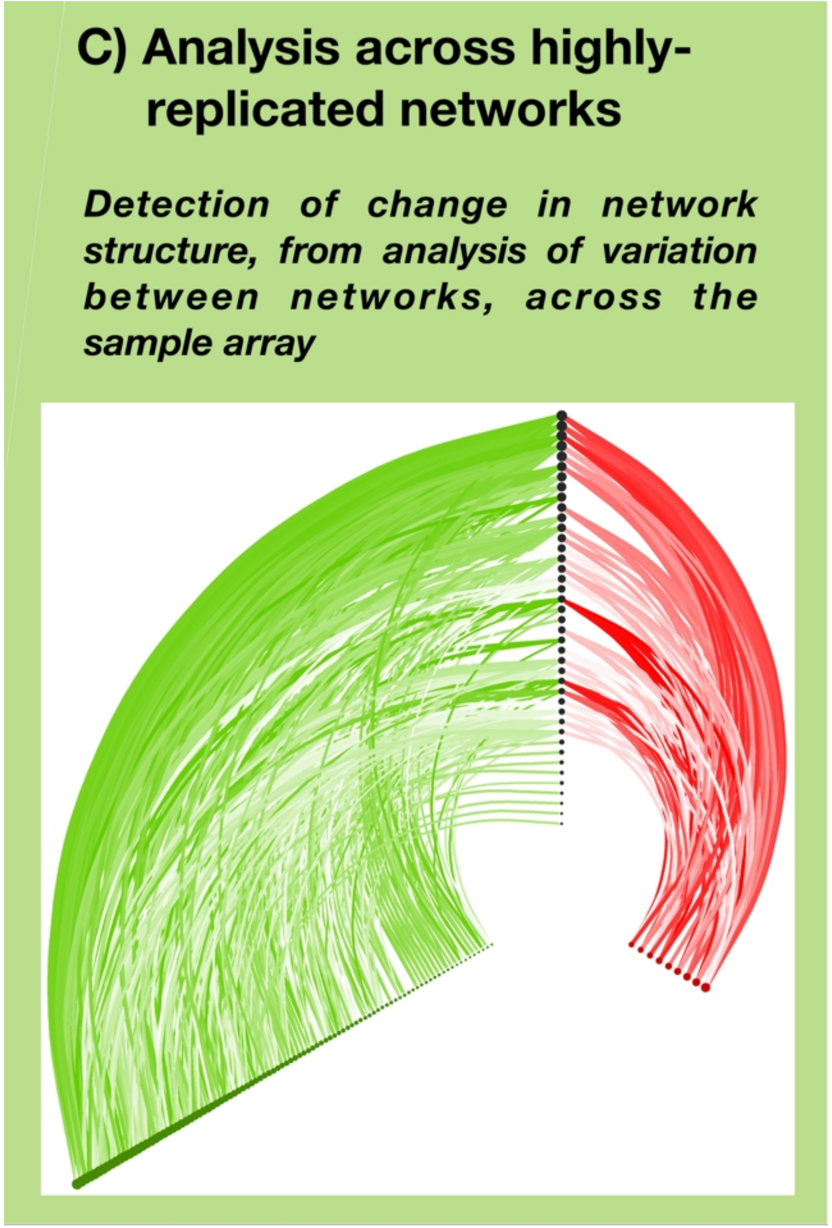
Global next-generation biomonitoring

- Thousands in ecosystems around the globe
- NGS data uploaded to the cloud
- Validity checked
- Parsed against existing, online



Global next-generation biomonitoring

- Learning in the cloud in real-time…
- This already exists in the form of *Google Translate*, *Now* and Apple's *Siri*
- Replicated networks and automatic analysis



Global next-generation biomonitoring

- Comparison between between networks should lead to understanding
- Variation over all sets of replicate networks → natural network variation
- Change greater than this variation indicates something has happened (automatically)

- Humans then interpret this using ecological theory

Global next-generation biomonitoring

- Our belief is that this biomonitoring would be:
  - Universal
  - General
  - Rapid and sensitive
- Lead to a revolution in our understanding of ecosystem change and management

Problems with NGB

- Cost
  - economies of scale for the equipment
  - reuse existing infrastructure
  - human intervention
  - Satellite-based remote sensing
- Policy – no network based policy
- Technical – NGS sequence databases are currently poor

Things to do….

- Can we learn eco-network structure and function from NGS data
- Which learning methods work best for this framework
- Can we learn and analyse networks in real-time
- We need to develop the statistics of this
    - Detect change across replicate networks
    - Power – how many samplers for measuring a given effect size

Conclusions

- Learning methods are great!
- They do real science and real ecology
- We learn community structure, change and function
- Moreover, we can do this from generic nucleic acid-based data
- We could therefore build a