LIFEPLAN:

A planetary inventory of life – a new synthesis built on Big Data combined with novel statistical methods

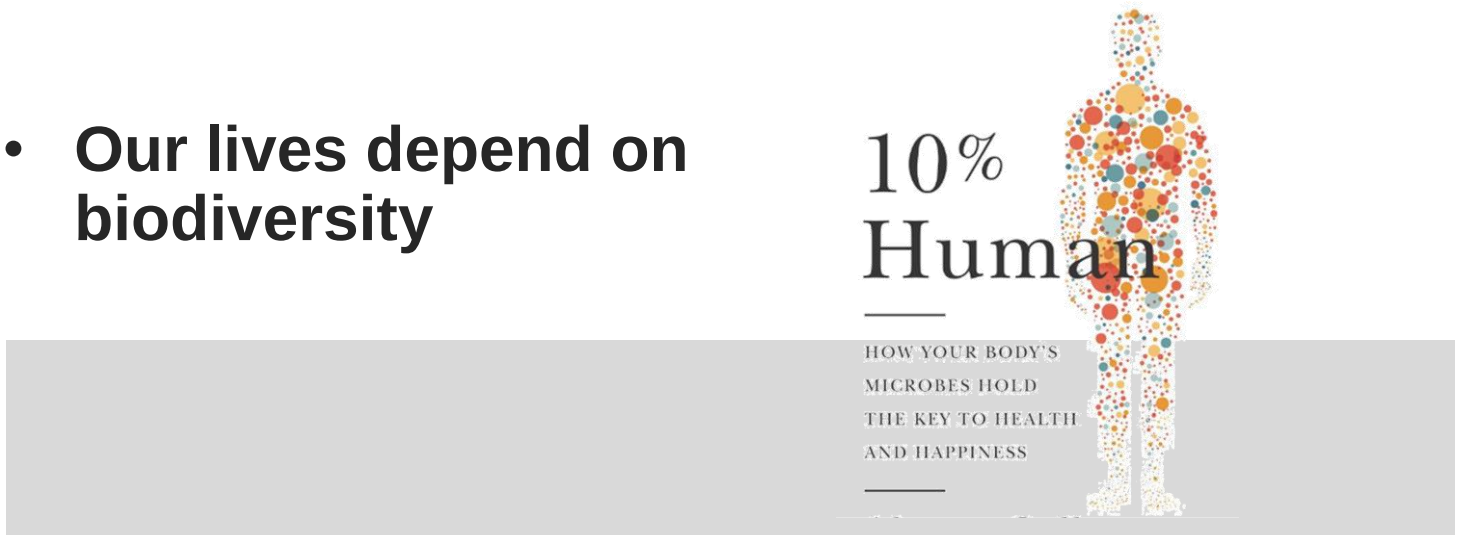
Prof. Otso Ovaskainen, University of Helsinki, Finland

Prof Tomas Roslin, Swedish Univ. of Agric. Sciences, Sweden

Prof. David Dunson, Duke University, USA

We need a predictive understanding of biodiversity

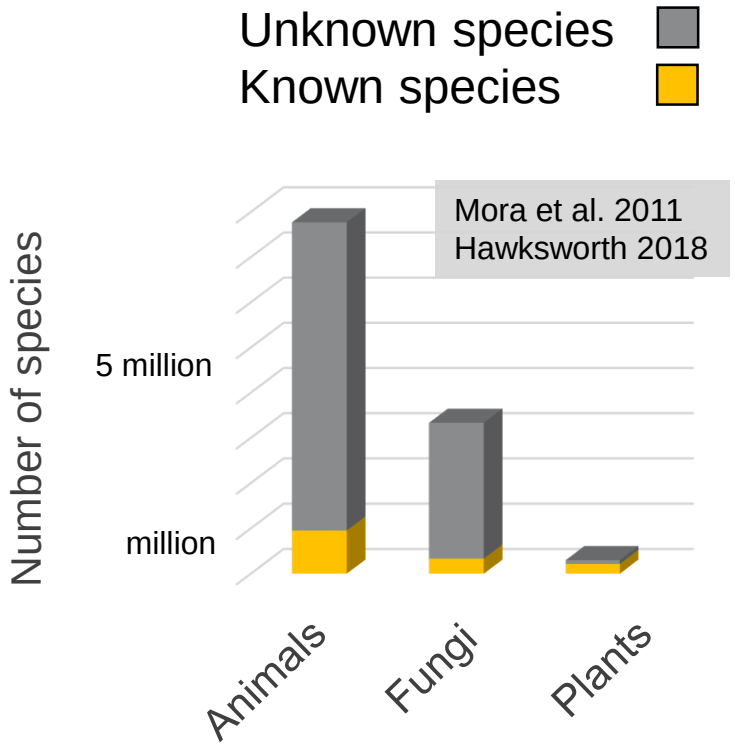
- **Our lives depend on biodiversity**



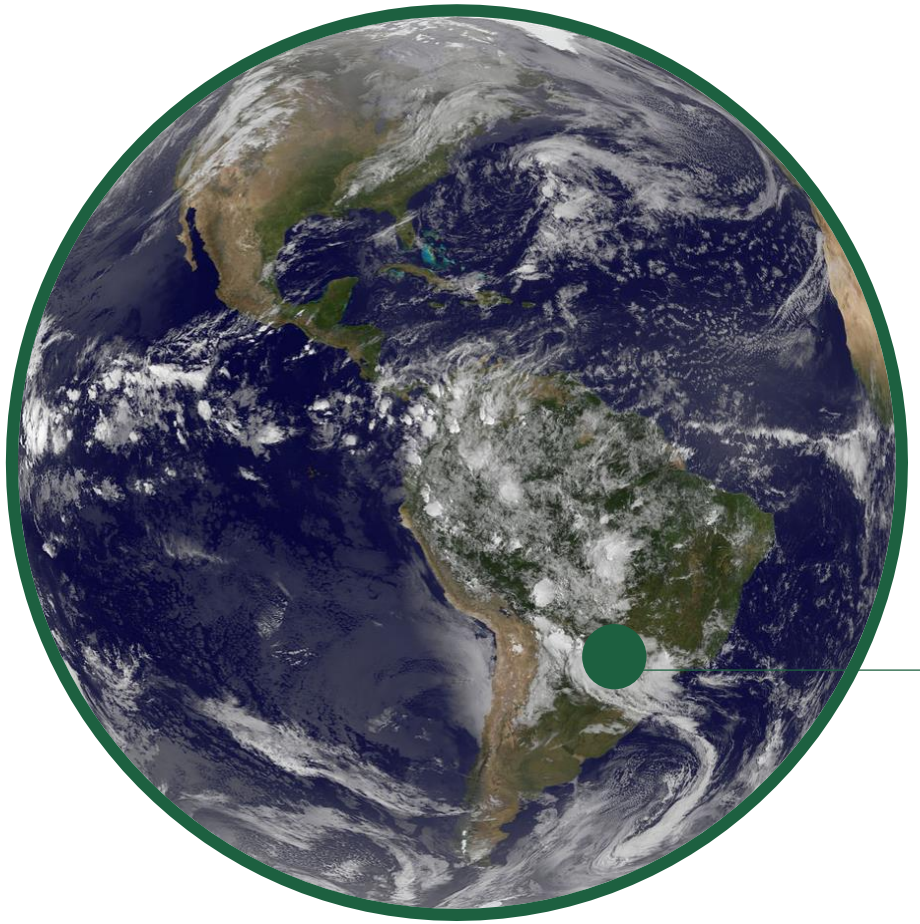
- **Global change is fundamentally altering the biodiversity**

United Nations (2019): ***The health of ecosystems on which we depend is deteriorating more rapidly than ever.***

- **Current state of biodiversity poorly known**



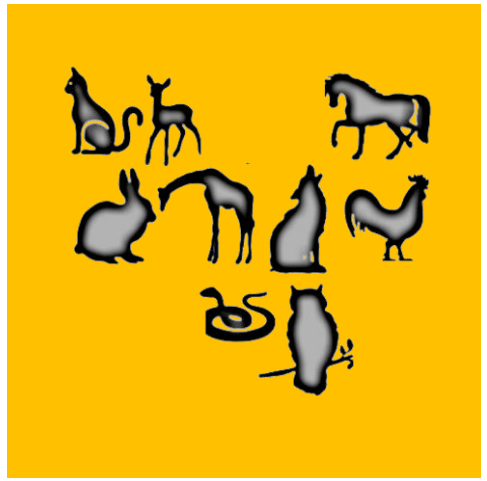
LIFEPLAN will generate a predictive understanding of global biodiversity and its drivers



LIFEPLAN WILL:

Predict current state

Predict future state given environmental change



Lifeplan

Methods for Big Data statistics

Scalable Bayesian inference, spatiotemporal modelling

Statistical, mathematical and theoretical ecology

Distributed experiments, biodiversity research

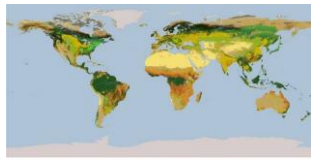
Globally distributed sampling design

Transformative understanding of life on Earth

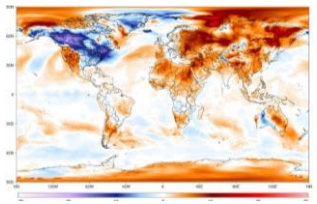
The Big AIM:
Global Joint Species Distribution Models

LAND COVER AND CLIMATIC DATA

MODIS

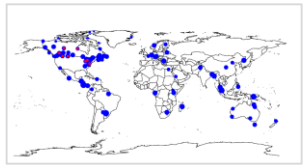


WORLD-CLIM

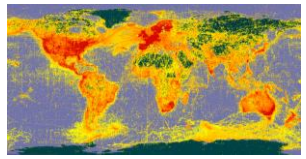


BIODIVERSITY DATA

LIFEPLAN DATA



GBIF AND OTHER SOURCES



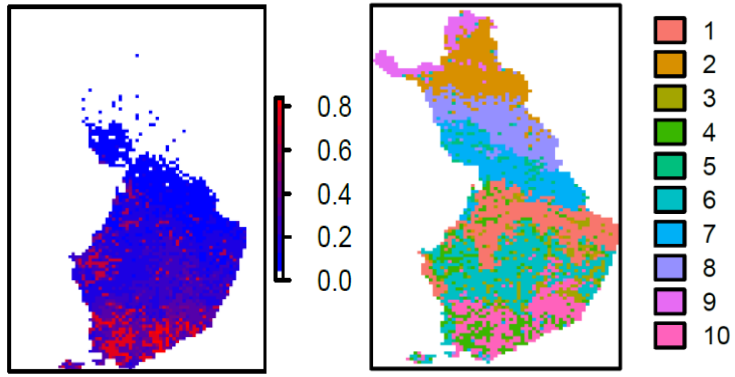
GLOBAL JOINT SPECIES DISTRIBUTION MODELS

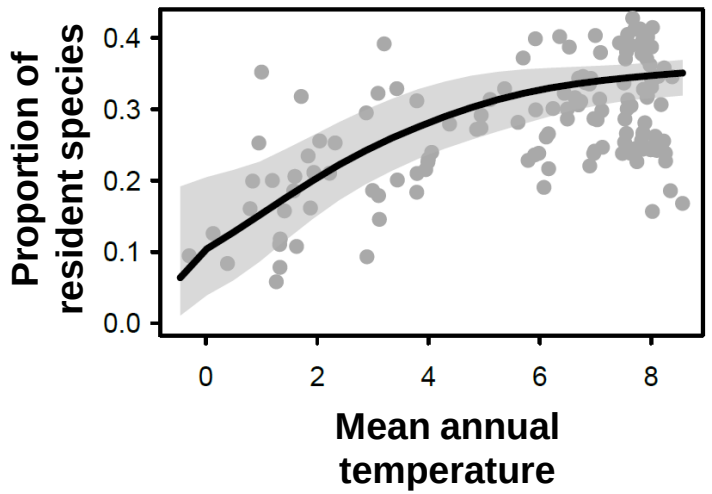
Current distribution of biodiversity

Response to environmental change

SPECIES LEVEL

COMMUNITY LEVEL



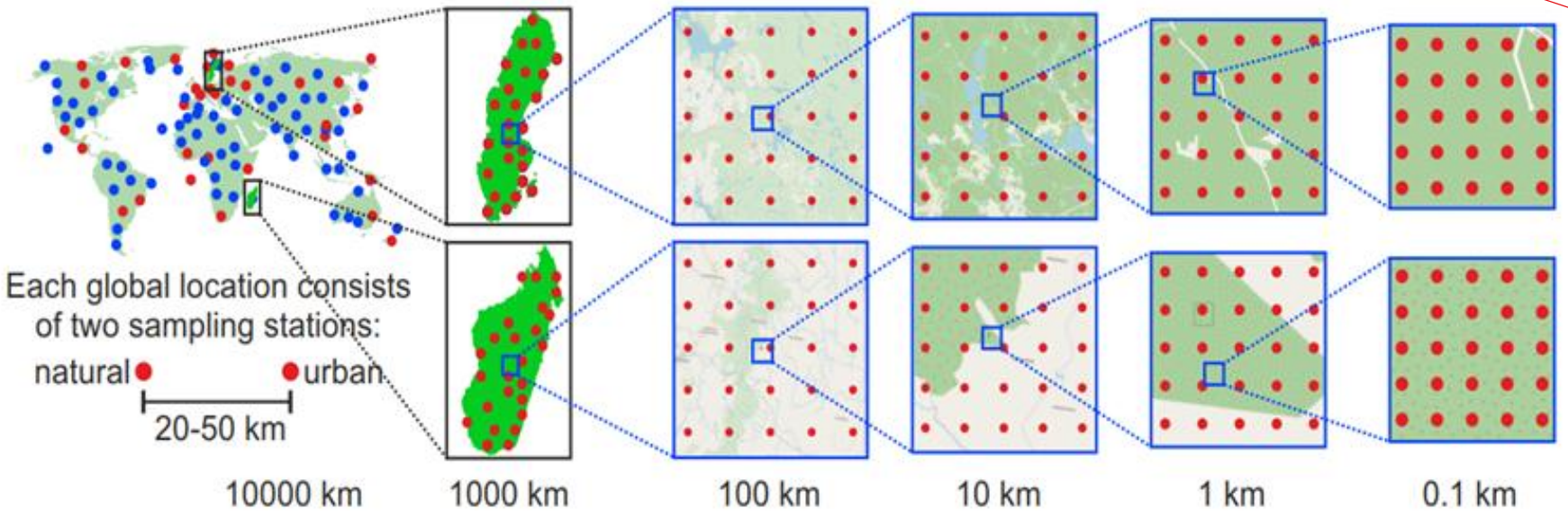


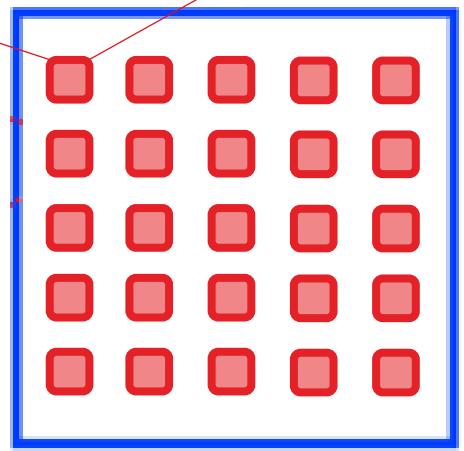
MODEL FITTING

INFERENCE AND PREDICTION

Globally Relevant, Unbiased Community-level Data Across Hierarchical Scales

- **Automated methods**
- **Systematic sampling**
- **Independent of expertise**





Grid of 5 x 5 sampling stations

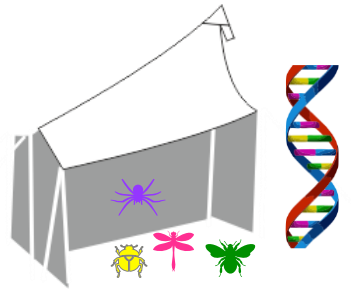
Global Malaise Trap Pilot Data (unpublished)

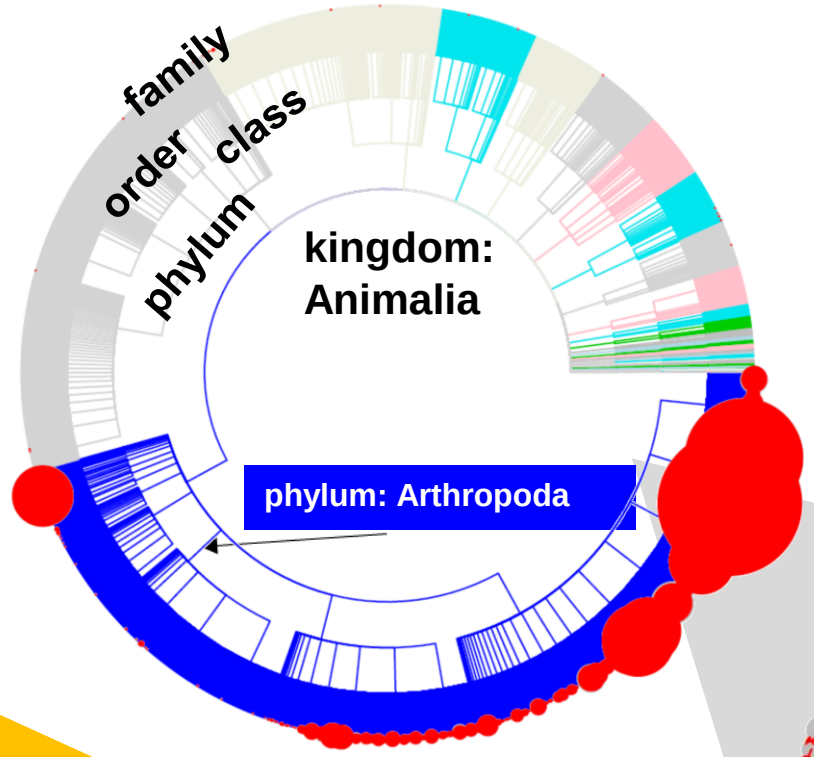
Biggest systematically collected insect data in the world!

2 million individuals

200,000 species

Insects captured by malaise trap are identified from DNA

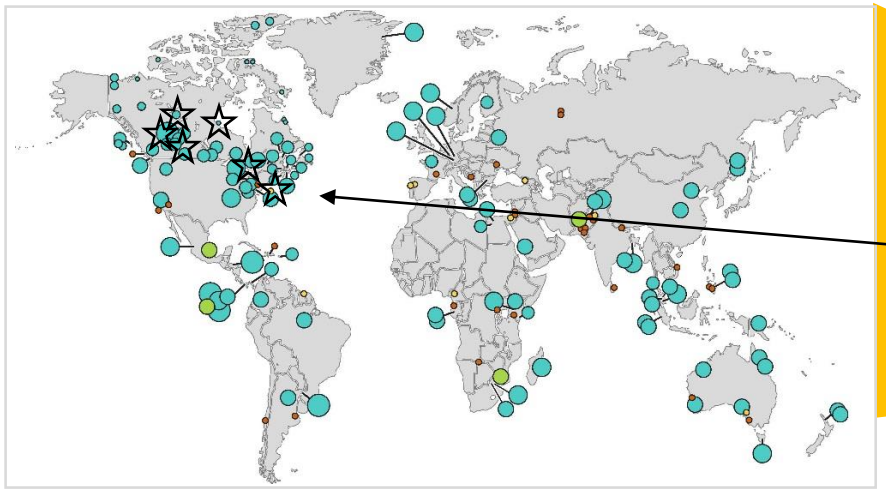




Reliable taxonomic placement

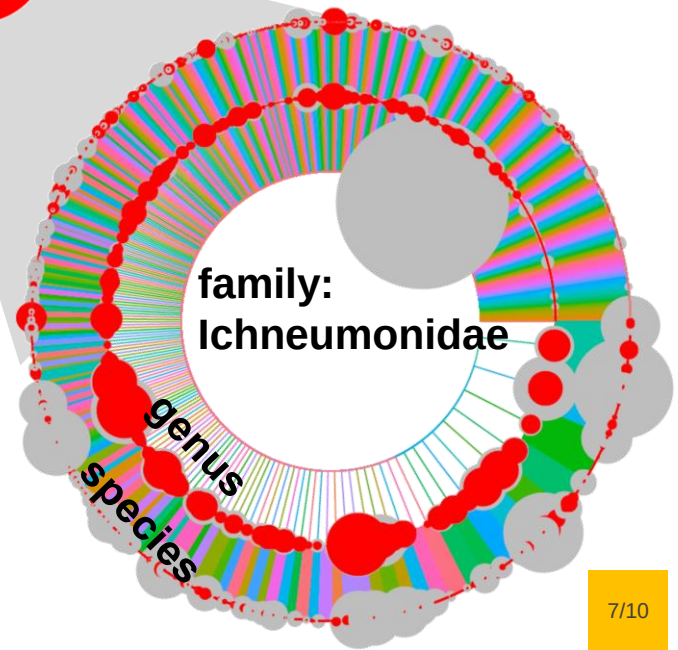
Reliable taxonomic placement currently not possible

Spatial coverage is already global, but with some gaps



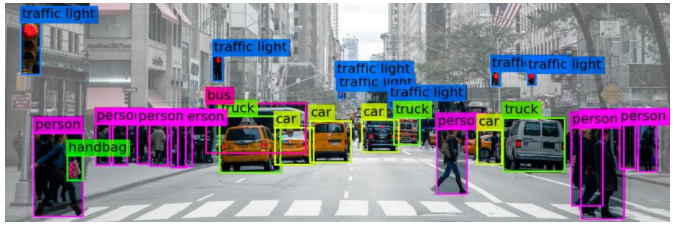
☆ Sampling locations where we recorded *D. insulare*

Diadegma insulare,
a parasitoid of the Cabbage Moth *Plutella xylostella*

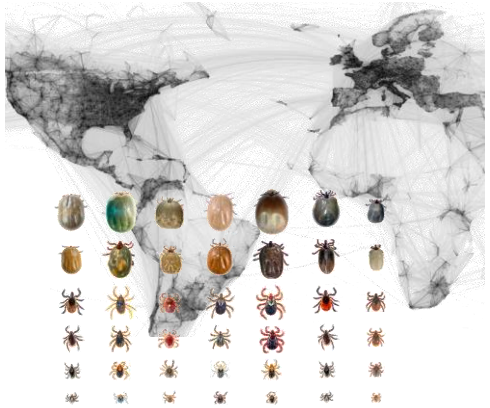


Statistical Methods for Big Ecological Data are Lacking

Interpretable models to understand, explain and predict

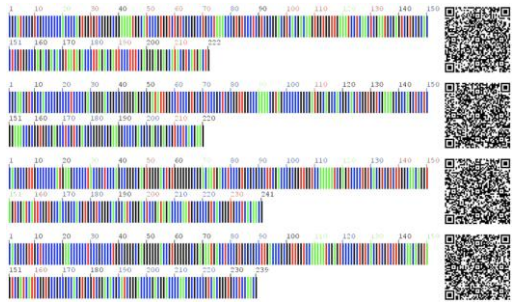


Huge data, many sources

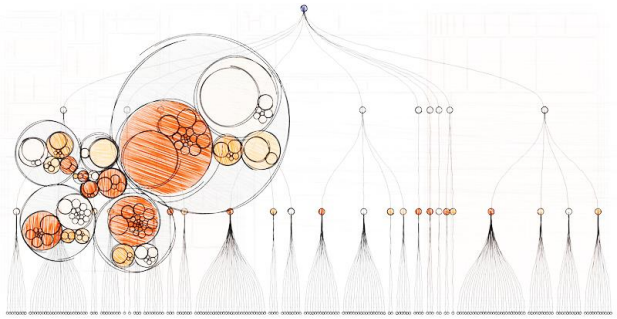


a unique combination of

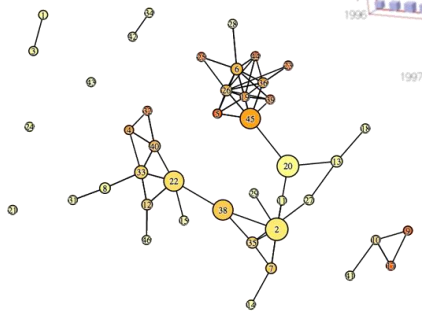
Entity resolution

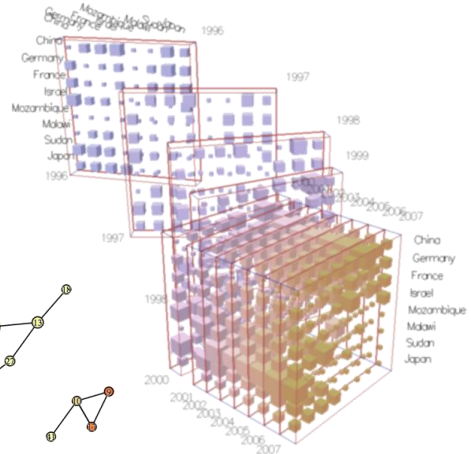


Complex structure



Sparsity and imbalance





The Synergy in Our Past Collaborations

BIG DATA STATISTICS

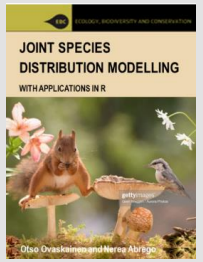
400 papers on Bayesian statistics

Recent highlight:
2019 **Mitchell Prize** - best applied paper in Bayesian statistics internationally

STATISTICAL ECOLOGY

170 papers on statistical ecology

Recent highlight:
Book on joint species distribution modelling (CUP)



COMMUNITY ECOLOGY

110 papers on community ecology

Recent highlight:
Science paper on globally distributed experiment

7 joint papers and manuscripts

12 joint papers and manuscripts

2 joint papers

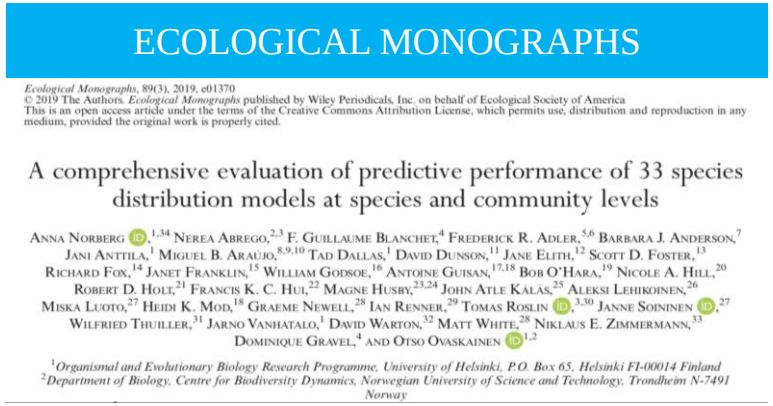
Highly cited by ISI (145 citations since 2017)


ECOLOGY LETTERS

Ecology Letters, (2017) doi: 10.1111/ele.12757

IDEA AND PERSPECTIVE

How to make more out of community data? A conceptual framework and its implementation as models and software


ECOLOGICAL MONOGRAPHS

Ecological Monographs, 89(3), 2019, e01370
© 2019 The Authors. *Ecological Monographs* published by Wiley Periodicals, Inc. on behalf of Ecological Society of America
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels

ANNA NORBERG,[1,34] NEREA ABREGO,[2,3] F. GUILLAUME BLANCHET,[4] FREDERICK R. ADLER,[5,6] BARBARA J. ANDERSON,[7] JANI ANTTILA,[1] MIGUEL B. ARAÚJO,[8,9,10] TAD DALLAS,[1] DAVID DUNSON,[11] JANE ELITH,[12] SCOTT D. FOSTER,[13] RICHARD FOX,[14] JANET FRANKLIN,[15] WILLIAM GODSOE,[16] ANTOINE GUISAN,[17,18] BOB O'HARA,[19] NICOLE A. HILL,[20] ROBERT D. HOLT,[21] FRANCIS K. C. HUI,[22] MAGNE HUSBY,[23,24] JOHN ATLE KÁLÁS,[25] ALEKSI LEHIKOINEN,[26] MISKA LUOTO,[27] HEIDI K. MOD,[18] GRAEME NEWELL,[28] IAN RENNER,[29] TOMAS ROSLIN,[3,30] JANNE SOININEN,[27] WILFRIED THUILLER,[31] JARNO VANHATALO,[1] DAVID WARTON,[32] MATT WHITE,[28] NIKLAUS E. ZIMMERMANN,[33] DOMINIQUE GRAVEL,[4] AND OTSO OVASKAINEN[1,2]

[1]*Organismal and Evolutionary Biology Research Programme, University of Helsinki, P.O. Box 65, Helsinki FI-00014 Finland*
[2]*Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim N-7491 Norway*

Transformative statistical methods for Big Data

Model uncertainty indicates data gaps

Real-world big data challenges

Transformative statistics

Novel model structures

The synergy of LIFEPLAN

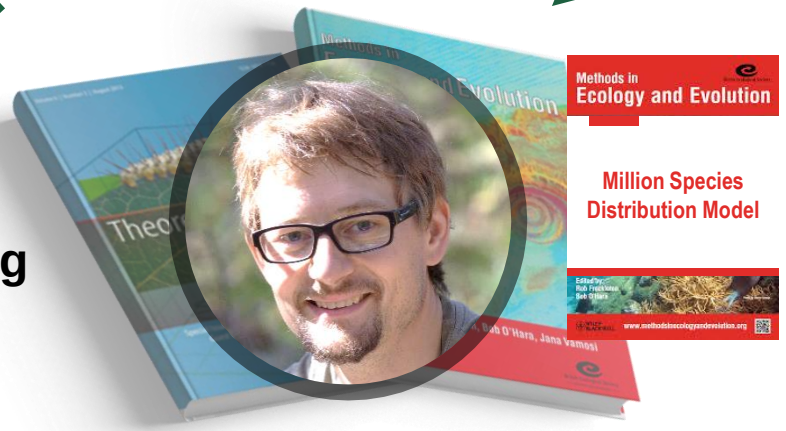
A new understanding of life on Earth

Synthesizing insights

Data and hypotheses

Globally distributed sampling design

Integrative understanding of life on earth



DELIVERABLES, NOVELTY AND RELATION TO WORK BY OTHERS

KEY DELIVERABLES (12)

ECOLOGICAL FRAMEWORK (13)

ECOLOGICAL HYPOTHESES (14)

GLOBAL SAMPLING PROJECTS (15)

JSDM: MODEL STRUCTURES (16)

JSDM: DATA TYPES (17)

JSDM: PREDICTION (18)

SPECIES IDENTIFICATION: DNA (19)

SPECIES IDENTIFICATION: AUDIO (20)

SPECIES IDENTIFICATION: IMAGE (21)

HMSC: PRESENT STATE (22)

STRUCTURE AND MANAGEMENT

WORK PACKAGES (23)

BUDGET (24)

HOST SUPPORT (25)

MANAGEMENT (26)

TRAINING (27)

TIMELINE (28)

DATA STORAGE AND SHARING (29)

FEASIBILITY AND PILOT STUDIES

ACCESS TO GLOBAL SAMPLING (30)

GSSP PILOT DATA (31)

AUDIO PILOT DATA (32)

HOW MUCH SAMPLING IS ”ENOUGH”? (33)

PROCESSES FROM PATTERNS? (34)

OUR RECENT WORK: 2019 (35)

OUR RECENT WORK: SUBMITTED (36)

METHODS

CALIBRATION: FUNGI (37)

CALIBRATION: MALAISE (38)

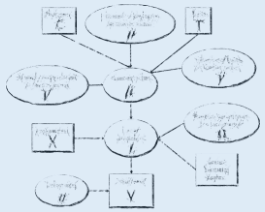
CALIBRATION: AUDIO (39)

CALIBRATION: CAMERA (40)

COLLABO-RATORS (41)

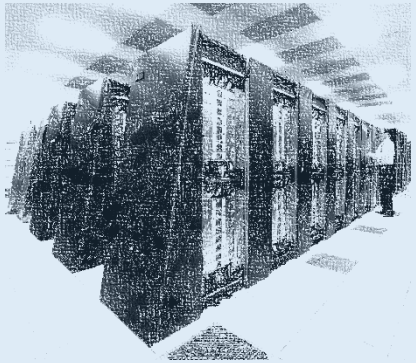
GLOBAL BIRD LISTENING PROJECT (42)

Key Deliverables of LifePlan



New Generation of Statistical Methods

- For huge dimensional, highly structured, sparse and imbalanced data
- Widely applicable across sciences, engineering and technology



A Transformative Data Resource

- Unbiased data across the globe on millions of species
- Spatial resolution from 100 meters to global scale

Transformative new understanding of biodiversity

- Global Joint Species Distribution Models will move community ecology towards a predictive science
- LIFEPLAN will address long-standing unsolved ecological hypotheses

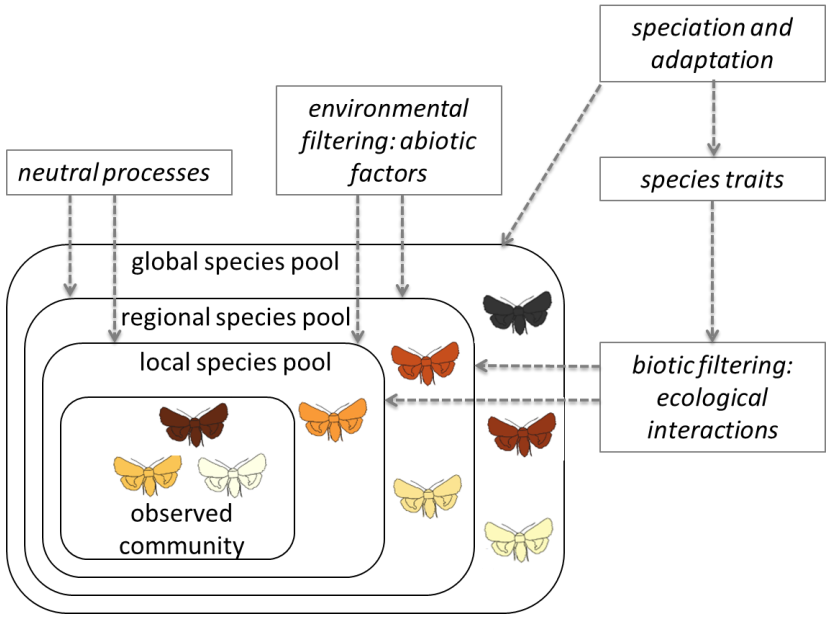
Training of “Scientific Data Scientists”

- Fundamentally different from “industry” data scientists
- Tools to build & implement realistic models & algorithms driven by scientific knowledge

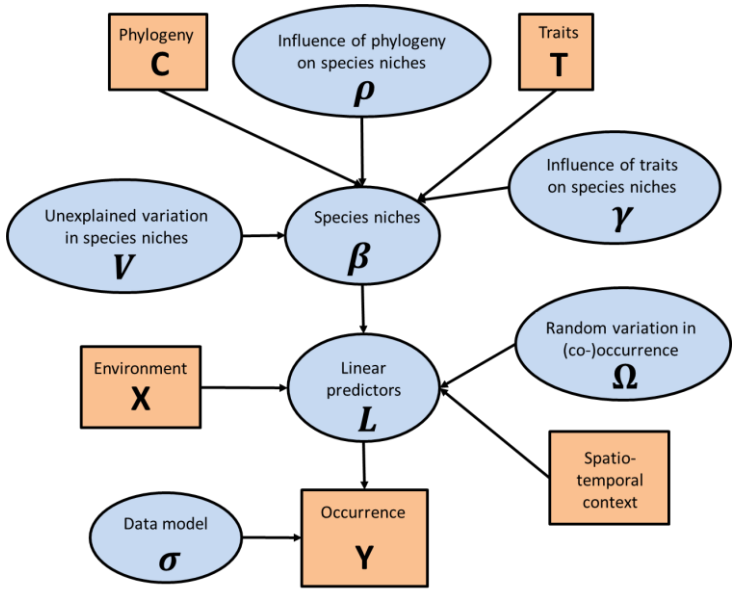
Ecological framework: community assembly processes

Starting point: present state-of-the-art

Conceptual framework….



…and its statistical implementation



LIFEPLAN WILL

- **Provide unprecented global data across taxa for testing and revising this framework**
- **Reveal the scale-dependency of assembly processes from 100 m to global scale**
- **In doing so: establish the principles for how communities are formed around the world**

LIFEPLAN will test long-standing ecological hypotheses

HYPOTHESIS: For micro-organisms "everything is everywhere but environment selects" **(ref).**

LIFEPLAN CAN TEST THIS BECAUSE: **We will have global data both before (air) and after (soil) the selection**

HYPOTHESIS: Species range sizes are smaller towards the Equator, creating more diversity among sites ("Rapoport's rule").

LIFEPLAN CAN TEST THIS BECAUSE: **We will have systematic data on ALL taxa from ALL latitudes *and* we can deal with sampling bias**

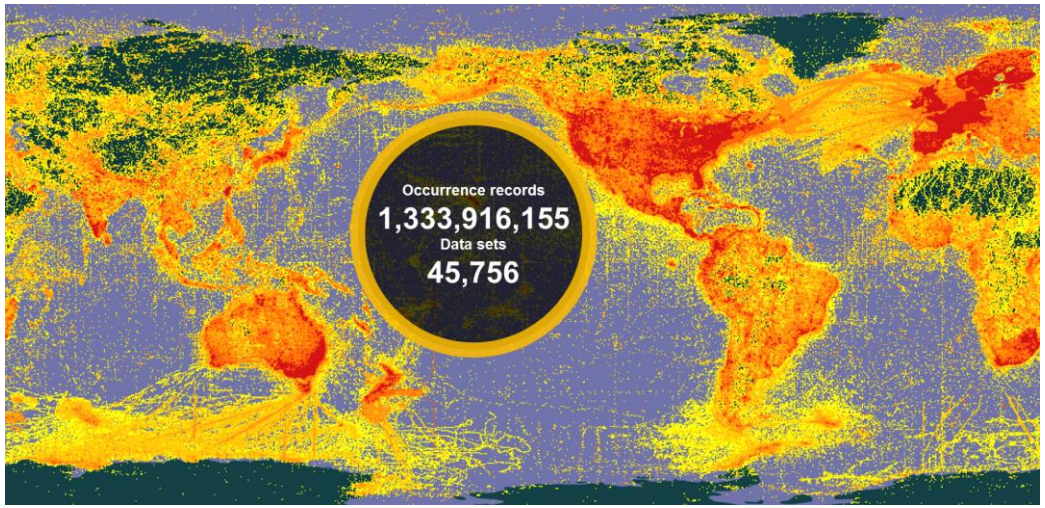
HYPOTHESIS: Species richness in megadiverse but "hard" groups can be efficiently deduced from the presence or richness of "indicator" species **(Caro 2010).**

LIFEPLAN CAN TEST THIS BECAUSE: **We will have systematic data on ALL taxa from all sites *and* we can deal with sampling bias.**

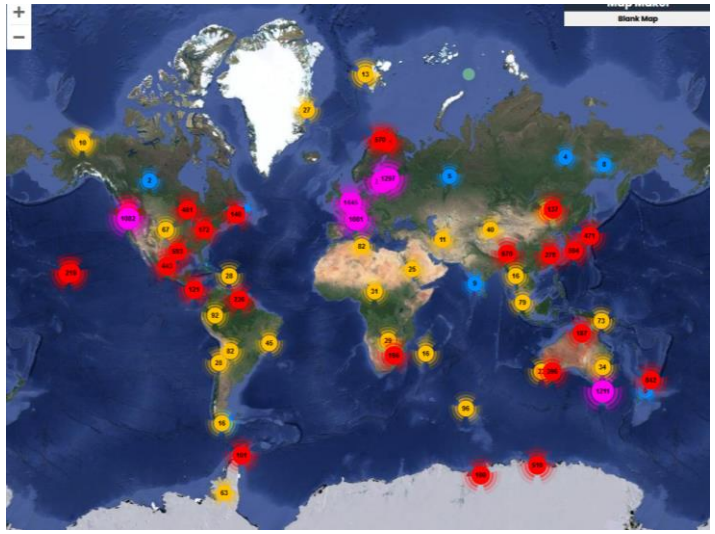
.

Examples of Global Biodiversity Databases

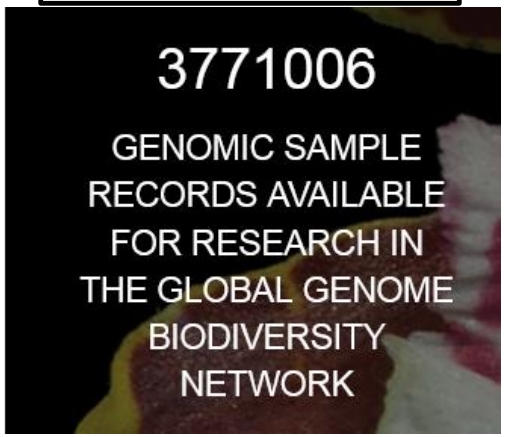
Global Biodiversity Information Facility (GBIF)



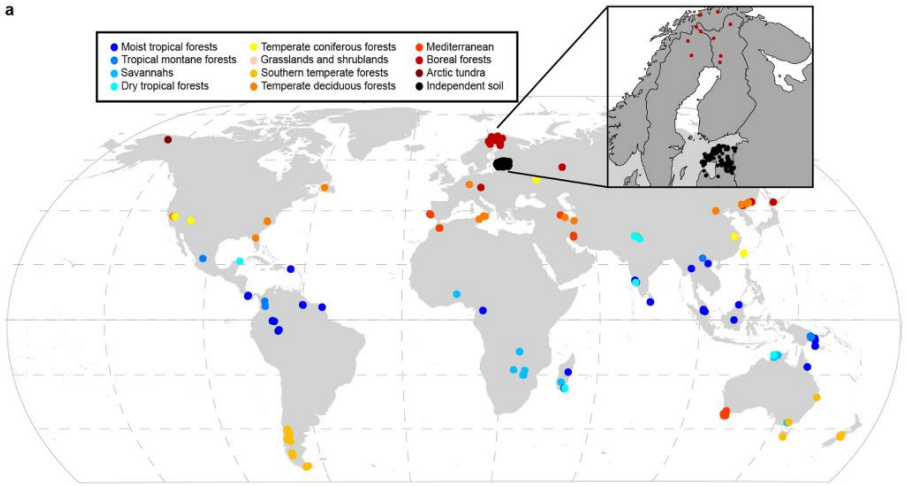
Global Fungi (Petr Baldrian et al.)



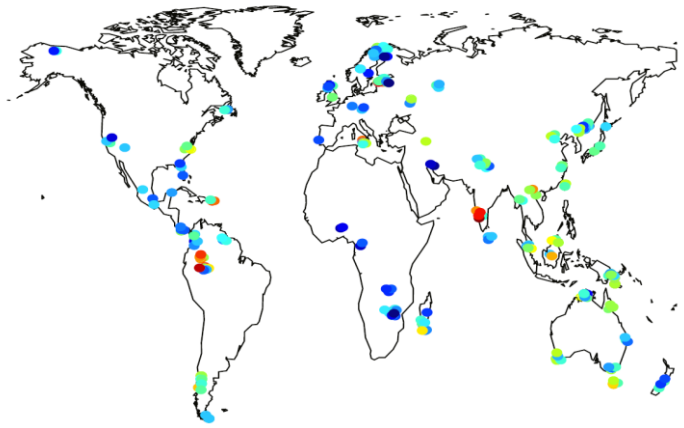
GLOBAL GENOME INITIATIVE



Global Soil Microbiomes (Mohammad Bahram et al.)



Global Soil Fungi (Leho Tedersoo et al.)



How “joint” is a joint species distribution model?

Ovaskainen et al: Hierarchical Modelling of Species Communities (HMSC)

Harris: Multivariate Stochastic Neural Network (MISTN)

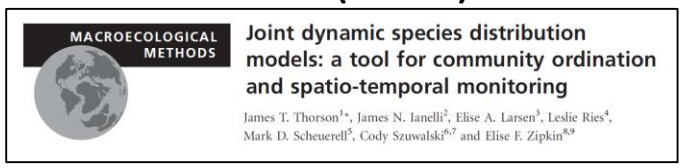
Hui et al: Bayesian ordination and regression analysis (BORAL)

Hui et al: Species Archetype Model (SAM)

Clark et al.: Generalised joint attribute modelling (GJAM)

Golding and Harris: Bayesian community ecology analysis (BC)

Thorson et al: Joint Dynamic Species Distribution Models (JDSDM)



Niche structure

N-1: joint modelling of species niches

N-2: dependency on traits

N-3: dependency on phylogeny

Biotic intetarctions

B-1: Estimation of association networks

B-2: dependency on traits

B-3: dependency on phylogeny

MODEL	N-1	N-2	N-3	B-1	B-2	B-3
HMSC	x	x	x	x		
BORAL	x	x		x		
GJAM				x		
JDSDM	x			x		
MISTN				x		
SAM	x					
BC				x		

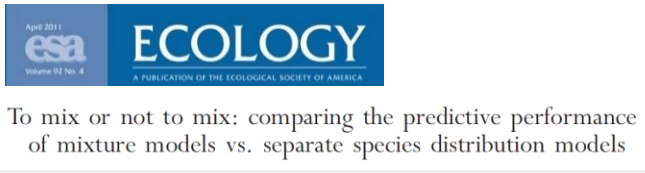
Data types utilized by Joint Species Distribution Models

Ovaskainen et al: Hierarchical Modelling of Species Communities (HMSC)

Harris: Multivariate Stochastic Neural Network (MISTN)

Hui et al: Bayesian ordination and regression analysis (BORAL)

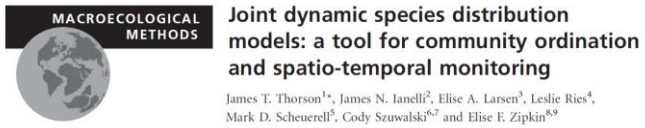
Hui et al: Species Archetype Model (SAM)



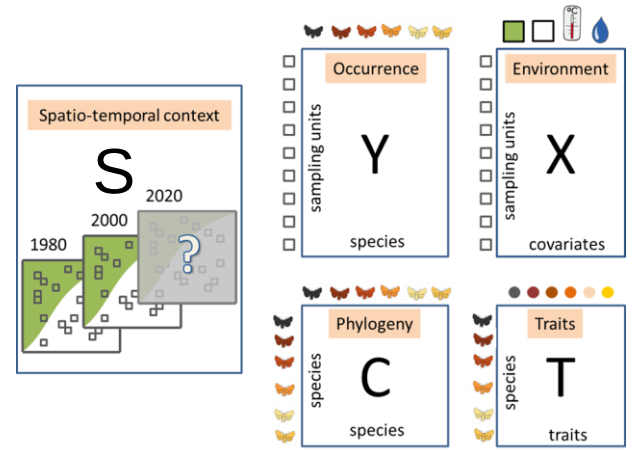
Clark et al.: Generalised joint attribute modelling (GJAM)

Golding and Harris: Bayesian community ecology analysis (BC)

Thorson et al: Joint Dynamic Species Disribution Models (JDSDM)



Y: Species
X: Environment
T: Traits
C: Phylogeny
S: Space, time, …



MODEL	Y	X	T	C	S
HMSC	x	x	x	x	x
BORAL	x	x	x		
GJAM	x	x	(x)		
JDSDM	x	x			x
MISTN	x	x			
SAM	x	x			
BC	x	x			

Predictive performance of Joint Species Distribution Models

Ovaskainen et al: Hierarchical Modelling of Species Communities (HMSC)

Harris: Multivariate Stochastic Neural Network (MISTN)

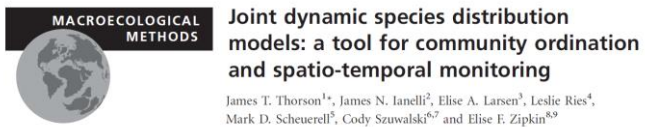
Hui et al: Bayesian ordination and regression analysis (BORAL)

Hui et al: Species Archetype Model (SAM)

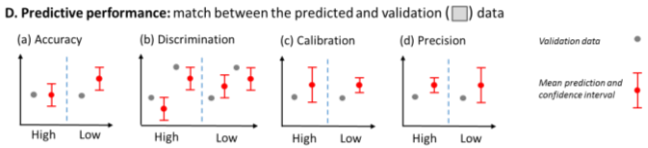
Clark et al.: Generalised joint attribute modelling (GJAM)

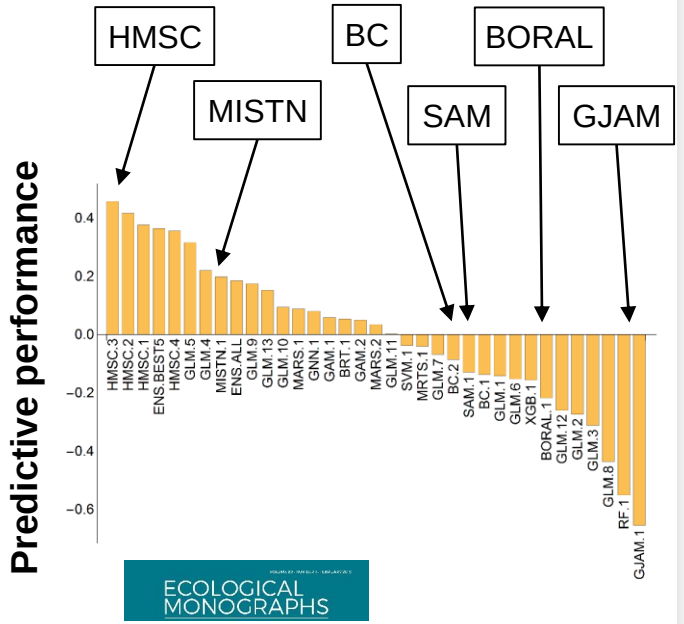
Golding and Harris: Bayesian community ecology analysis (BC)

Thorson et al: Joint Dynamic Species Disribution Models (JDSDM)



Predictive performance was evaluated for data on six taxonomical groups, for interpolation and extrapolation, for individual species and community features





JDSDM

Was excluded from comparison as does not apply to species rich communities

DNA-based species identification

Ovaskainen et al: PROTAX

Methods in Ecology and Evolution
Methods in Ecology and Evolution 2017, 8, 398–407 doi: 10.1111/2041-210X.12721

TECHNOLOGICAL ADVANCES AT THE INTERFACE BETWEEN ECOLOGY AND STATISTICS

Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding

Panu Somervuo[*,1], Douglas W. Yu[*,2,3], Charles C.Y. Xu[2,4], Yinqiu Ji[2], Jenni Hultman[5], Helena Wirta[6] and Otso Ovaskainen[1,7]

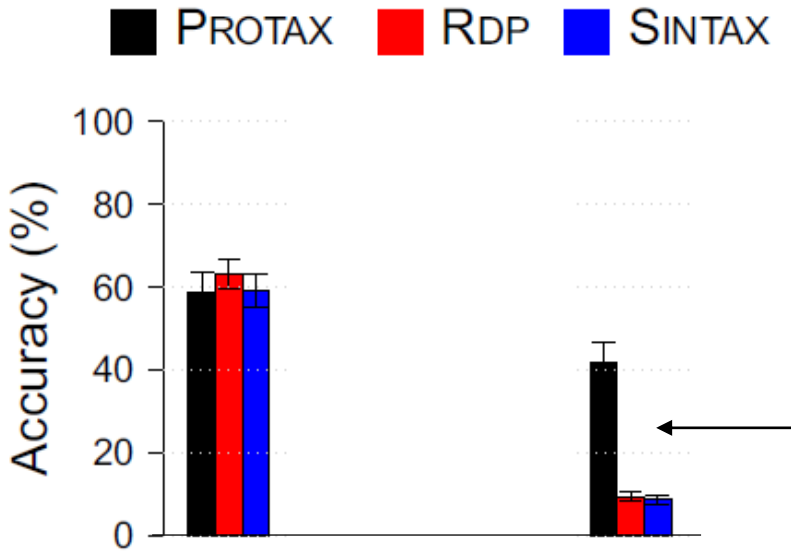
Edgar: SINTAX

SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences

Wang et al: RDP

Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy[▽†]

PROTAX is the only method that explicitly models species not known to science and species for which there are no reference sequences



Comparison of predictive accuracy based on fungal mock-community data. *Abarenkov et al. 2018, New Phytologist*

Fungi and arthropods

Audio-based species identification

Ovaskainen et al: ASI


ECOLOGY LETTERS

Ecology Letters, (2018) 21: 1244–1254 doi: 10.1111/ele.13092

METHODS

Animal Sound Identifier (ASI): software for automated identification of vocal animals

Katz et al: monitoR

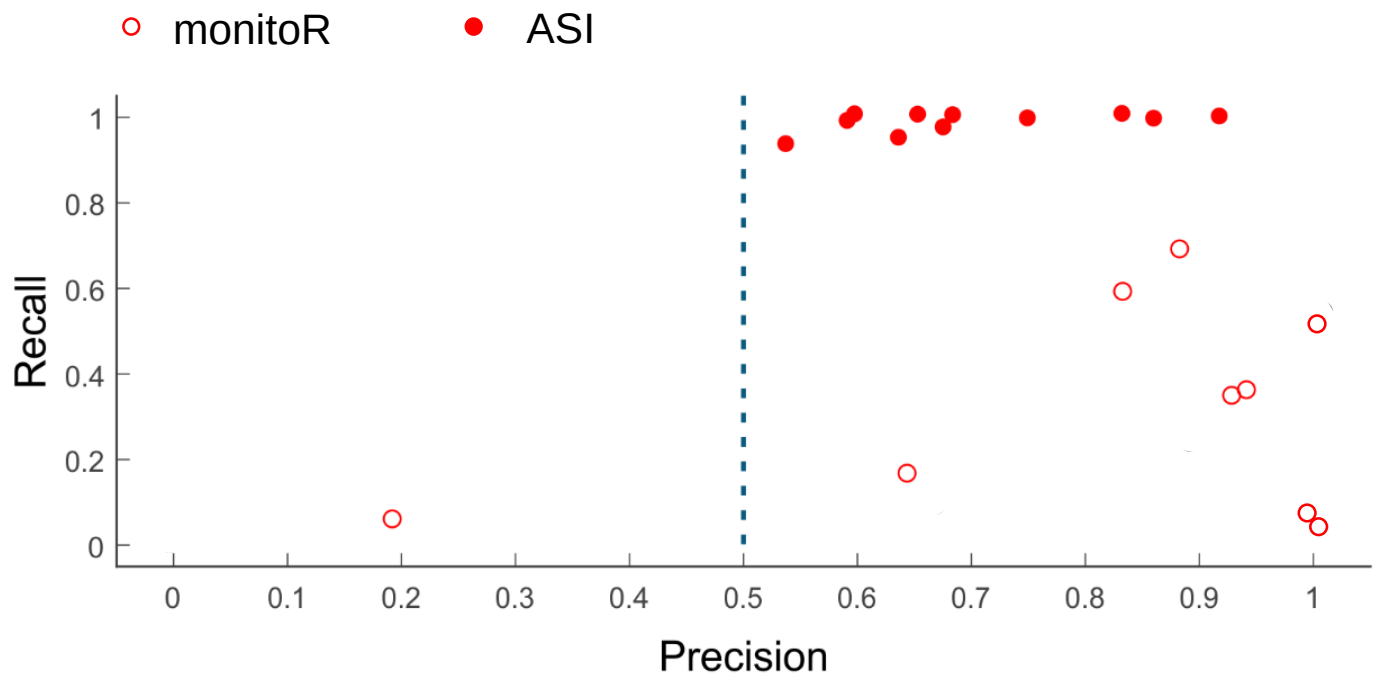

Articles

Tools for automated acoustic monitoring within the R package monitoR

Jonathan Katz, Sasha D. Hafner & Therese Donovan

Pages 197-210 | Received 18 Sep 2015, Accepted 28 Dec 2015, Published online: 05 Feb 2016

ASI does not require a-priori templates, but generates them automatically from the training data



Comparison of predictive accuracy based on bird vocalizations.
Ovaskainen et al. 2018, Ecology Letters

State-of-the-art in automated species identification from camera trap data

PNAS

Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning

Mohammad Sadegh Norouzzadeh[a], Anh Nguyen[b], Margaret Kosmala[c], Alexandra Swanson[d], Meredith S. Palmer[e], Craig Packer[e], and Jeff Clune[a,f,1]

2018

REVIEW

Journal of Animal Ecology — British Ecological Society

A computer vision for animal ecology

Ben G. Weinstein

2018

RESEARCH ARTICLE

2019

Methods in Ecology and Evolution — British Ecological Society

A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images

Colin J. Torney[1*] | David J. Lloyd-Jones[2*] | Mark Chevallier[1] | David C. Moyer[3] | Honori T. Maliti[4] | Machoke Mwita[4] | Edward M. Kohi[4] | Grant C. Hopcraft[5]

SCIENTIFIC REPORTS

OPEN

Insights and approaches using deep learning to classify wildlife

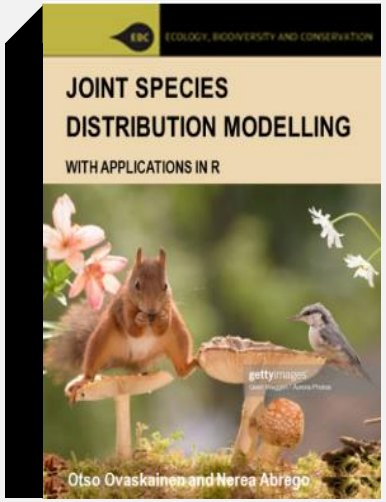
Zhongqi Miao[1,2], Kaitlyn M. Gaynor[1], Jiayun Wang[1], Ziwei Liu[2], Oliver Muellerklein[2,3], Mohammad Sadegh Norouzzadeh[4], Alex McInturff[1], Rauri C. K. Bowie[5], Ran Nathan[6], Stella X. Yu[2,3] & Wayne M. Getz[1,7]

2019

Current state of HMSC: Hierarchical Modelling of Species Communities

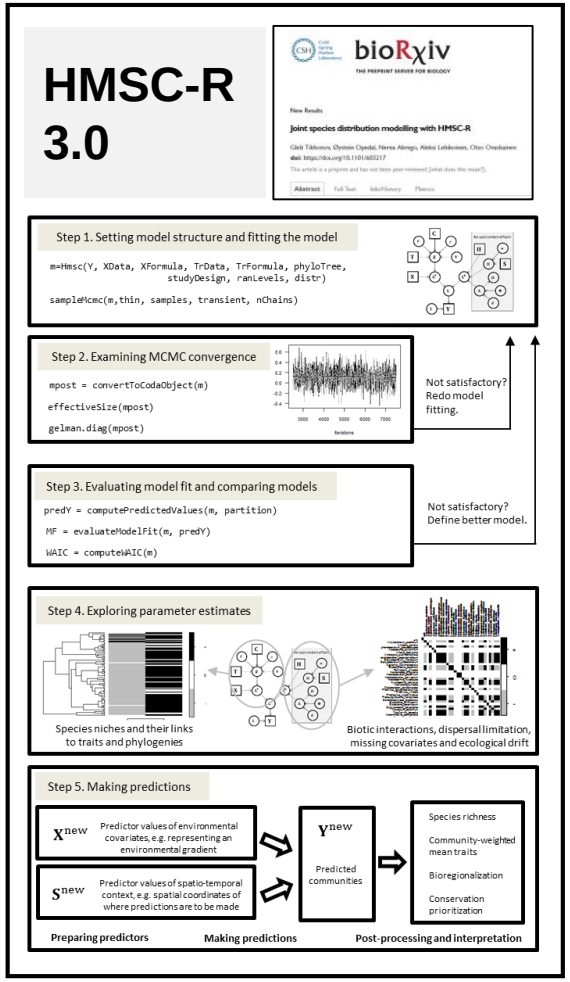
Selected journal papers and book

Tikhonov, G., Duan, L., Abrego, N., Newell, G., White, M., Dunson, D. and Ovaskainen, O.
Computationally efficient joint species distribution modelling of big spatial data.

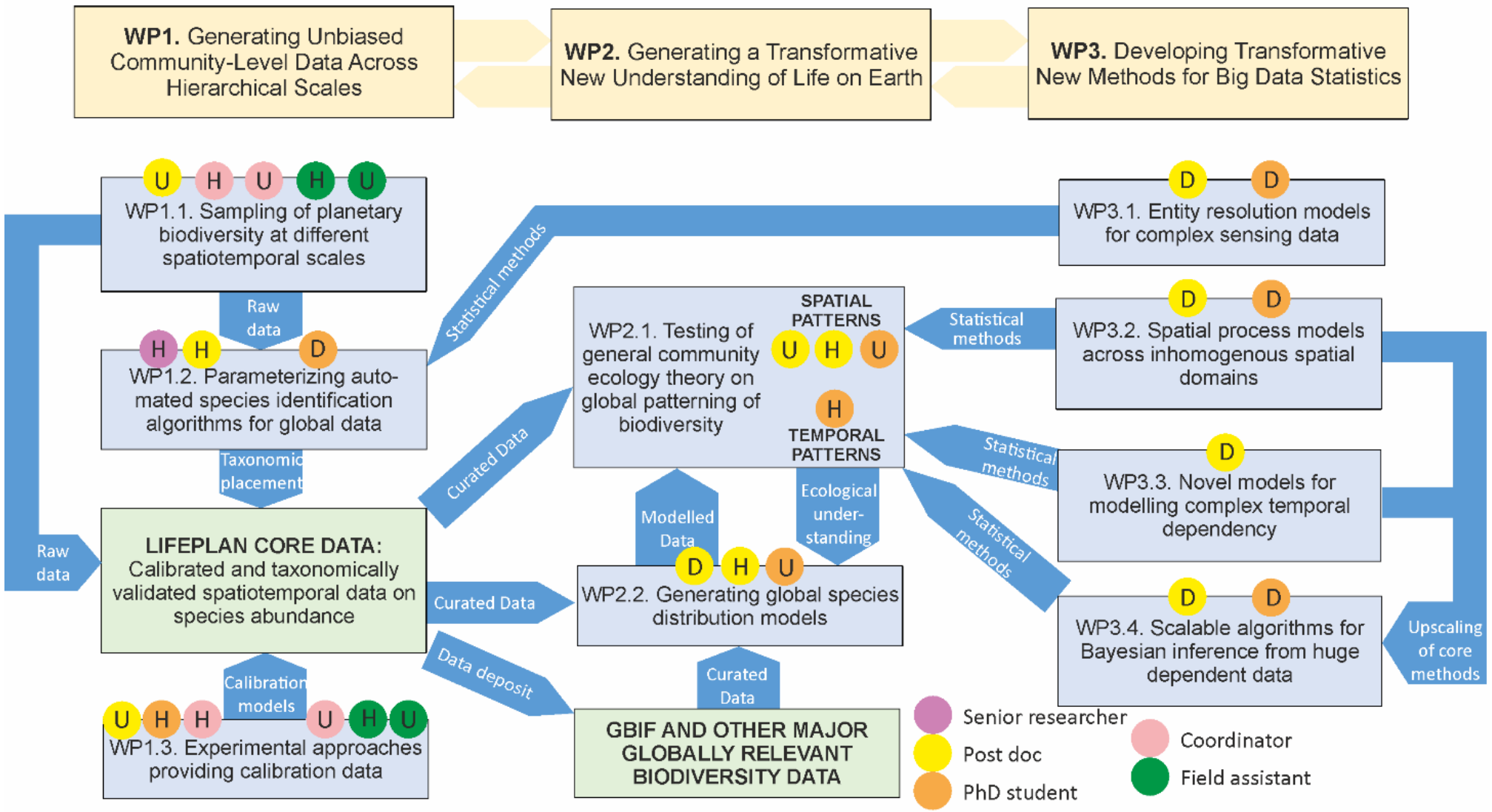


Cambridge University Press (in press)

Software



The structure of the LIFEPLAN project



Budget details

Cost category			Corresponding PI	2nd PI	3rd PI	Total in euro
		PI name	Ovaskainen (OO)	Roslin (TR)	Dunson (DD)	
		Host institution	Helsinki Univ.	Swedish Univ. Agric. Sci	Duke Univ.	
Direct cost	Personnel	Pi	0	0	919,221	919,221
		Senior staff	469,360	0	0	469,360
		Postdocs	586,287	558,724	651,590	1,796,601
		Students	375,296	410,751	357,844	1,143,891
		Other	173,837	934,897	0	1,108,734
	Total direct costs for personnel (in euro)		1,604,780	1,904,372	1,928,655	5,437,807
	Travel		210,000	210,000	130,418	550,418
	Equipment		1,054,500	1,054,500	36,502	2,145,502
	Other goods and services	Consumables	810,000	810,000	5,259	1,625,259
		Publications	33,000	33,000	33,000	99,000
		Other: Tuition remission and Audit Fees, field work costs in Madagascar as services purchased through local collaborating organizations	65,200	40,000	133,062	238,262
		Total Other Direct Costs (in euro)	2,172,700	2,147,500	338,241	4,658,441
A – Total Direct Costs (i + ii)			**3,777,480**	**4,051,872**	**2,266,896**	**10,096,248**
B – Indirect Costs (overheads) 25% of Direct Costs			**944,370**	**1,012,968**	**566,724**	**2,524,062**
C1 – Subcontracting Costs (no overheads)			**0**	**0**	**0**	**0**
C1 – Subcontracting Costs with no overheads			**0**	**0**	**0**	**0**
Total Estimated Eligible Costs (A + B + C)			**4,721,850**	**5,064,840**	**2,833,620**	**12,620,310**
Total Requested Grant			**4,721,850**	**5,064,840**	**2,833,620**	**12,620,310**

Global Sampling Infrastructure: 2,1M€

- Sampling stations (n=200, each 9.820€)
  - Cyclone Sampler (2700€)
  - Malaise Trap (240€)
  - Ten AudioMoth recorders (850€)
  - Six Browning Strike Force Game Cameras (810€)
  - Eight 2TB external hard drives (520€)
  - Consumables (2200€)
  - Shipping and customs (2500€)
- Sample management system 25,000€

Field-work and sequencing: 2,9M€.

- Logistics coordinator and field team of two assistants in Sweden (934,897€)
- Field team in Madagascar, services through local collaborating organizations (personnel 173,837€, other costs 55,200€)
- Transport and accommodation of field teams (120,000€)
- Sequencing 64,800 samples: 1.62M€
- Permits and documentation according to the Nagoya protocol: 30,000€

Scientific personnel and resources: 5,1M€

LIFEPLAN HELSINKI TEAM

- Senior researcher, 6-year (469,360€),
- Tree 3-year post docs (586,287€)
- Two 4-year PhD students (375,296€)

LIFEPLAN UPPSALA TEAM

- Three 3-year post docs (558,724€)
- Two 4-year PhD students (410,751€).

LIFEPLAN DUKE TEAM

- PI (30% contribution; 919,221€)
- Five 2-year post docs (651,590€)
- Four 3-year PhD students (salary 357,844€ and tuition remission fees 120,062€)

RESOURCES

- Travel for team members and collaborators (430,418€)
- Computational resources (equipment 36,502€ and consumables 5,259€)
- Publication costs (99,000€)
- Audit fees (33,000€)

Support from the host institutions and infrastructures

Additional support from the host institutions

- Helsinki contributes a full-time coordinator for the general management of the project.
- Uppsala contributes a full-time lawyer to deal with sample transport permits.
- Duke contributes computational clusters (ca. 60,000€) and storage space (ca. 60,000€).

Relevant infrastructures

- “Thriving Nature” was selected as the 2019 profiling action of University of Helsinki, with specific emphasis on ecological big data. This will provide 2 tenure track positions (Ecological Data Sciences and Ecological Networks), and four staff scientists (statisticians / bioinformaticians).
- The Swedish Science for Life Laboratory offers outstanding facilities for the high-throughput DNA work
- Duke has world class computing facilities and support, including access to the Duke Computing Cluster, and possibilities to leverage immense resources for Artificial Intelligence ($100 million to a new AI centre).

Management of LIFEPLAN

OUR PREVIOUS EXPERIENCE IN MANAGING GRANTS

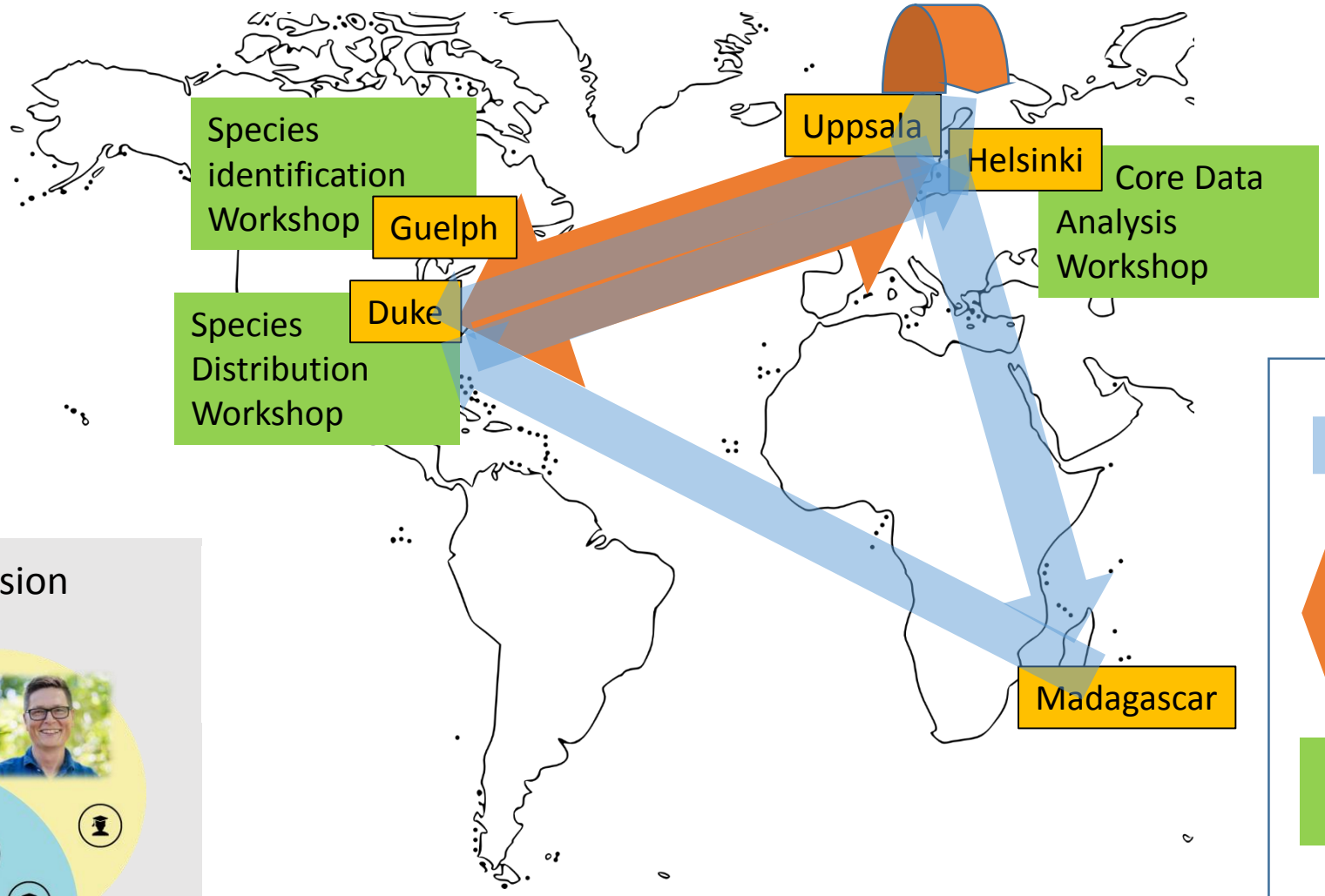
- OO was a director of a national Centre of Excellence, with 60 researchers
- TR has successfully lead ten international sampling projects
- DD has directed several interdisciplinary grants ranging from neurosciences to tech industry applications

PI	Grants as PI	Grants as co-PI	Total
OO	7.4M€	9.4M€	16.8M€
TR	3.6M€	7.9M€	11.5M€
DD	6.4M€	3.1M€	9.5M€
Total	**17.4M€**	**10.4M€**	**37.8M€**

LIFEPLAN MANAGEMENT

- Overall project management from Helsinki (meetings, travels, etc.).
  - Full-time project coordinator (host support)
- Global sampling coordination from Uppsala
  - Nagoya-savvy lawyer (host support)
- Data management coordinated by Duke
  - Extensive computational resources (host support)
  - Detailed database management plan in place

Training the next generation of scientists

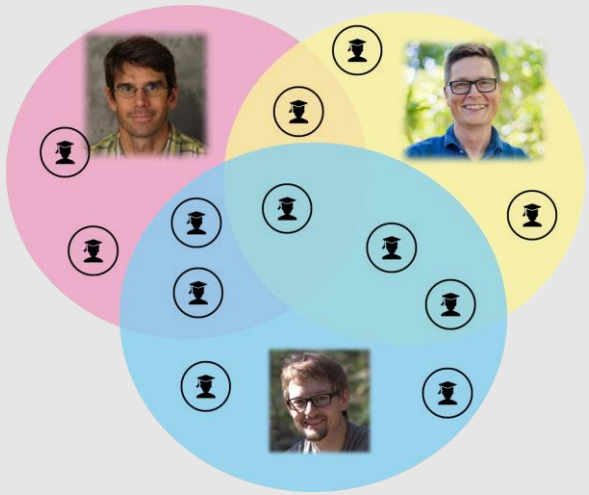


Annual meetings switch places

Longterm work visits maximize interaction

Cross-WP Workshops bring key expertise together around tasks

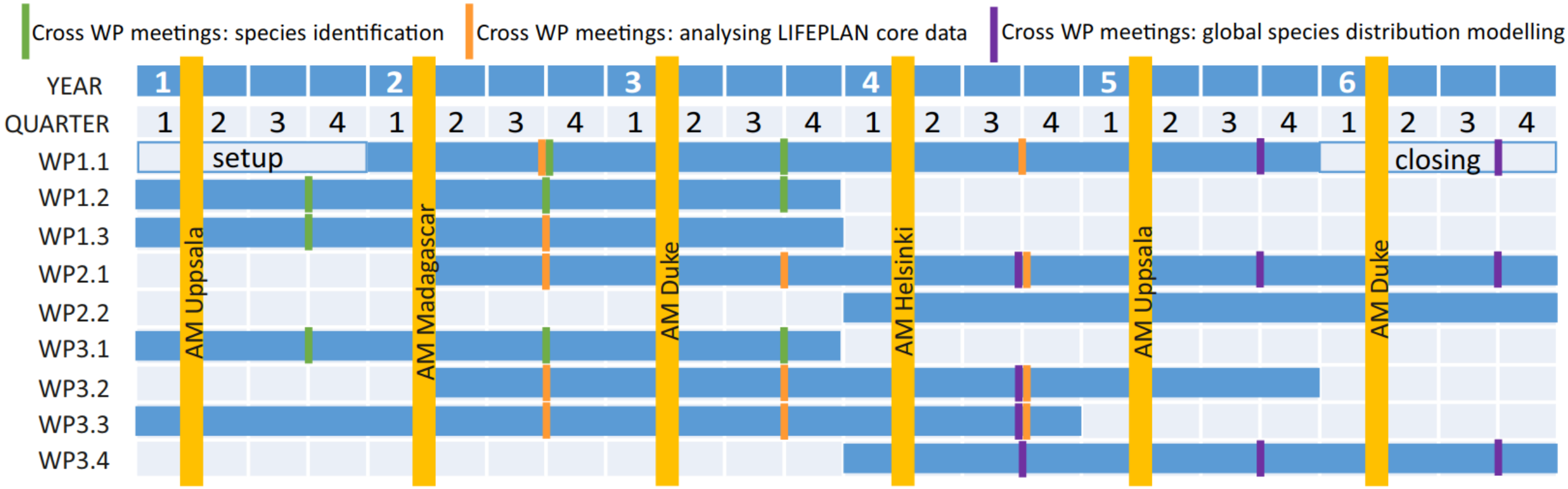
Shared supervision



Previous example of ⟷: One-month visit by PhD student OO → DD resulted in

Tikhonov, G., Duan, L., Abrego, N., Newell, G., White, M., Dunson, D. and Ovaskainen, O. Computationally efficient joint species distribution modelling of big spatial data. *Ecology,* in press.

LIFEPLAN timeline



Data storage & sharing

Back-up data repository

Funds committed by Duke University
for data storage and back-up.
(letter from Dean Mohammed Noor)

Duke

Main processing site

Central data repository

Organized by sampling site & data type

From physical samples
to digital DNA sequences

Uppsala

Helsinki

High-speed connections to
rapidly move data from/to Europe

Individual sites

Initial data collection and storage

Local data storage

Audio recordings for birds

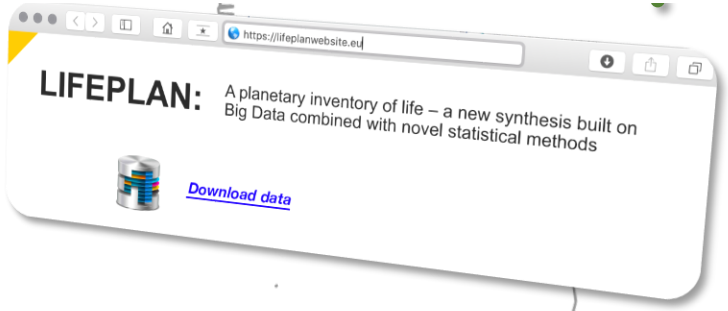
Numerical data for ecosystem processes

Image files for mammals

Physical samples for fungal spores, soil, insects

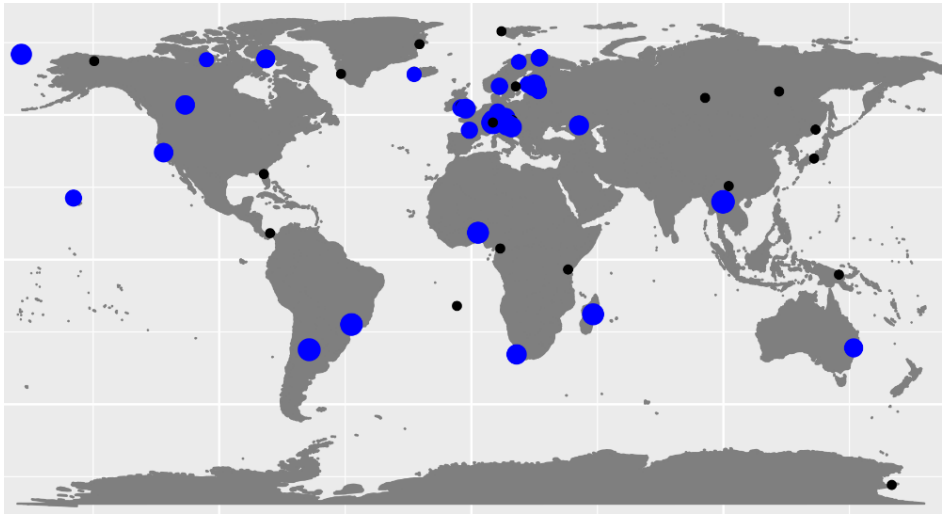
Dedicated LIFEPLAN website

Describes and provides free data access to researchers

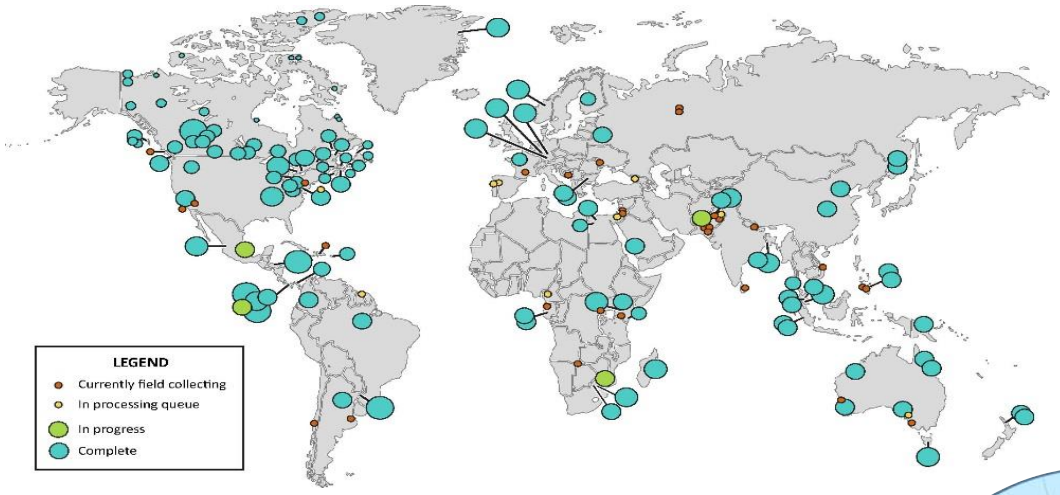


We have access to Global Sampling

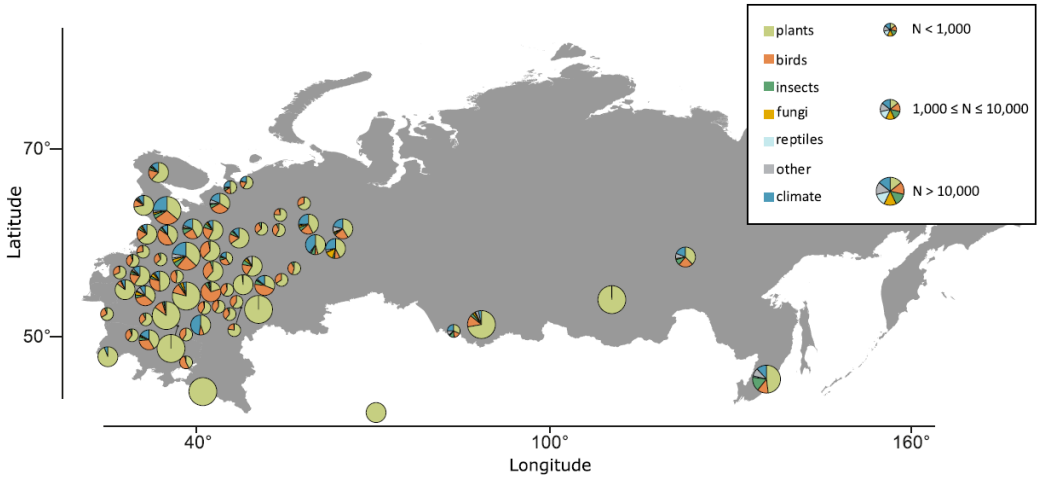
Global Spore Sampling Project



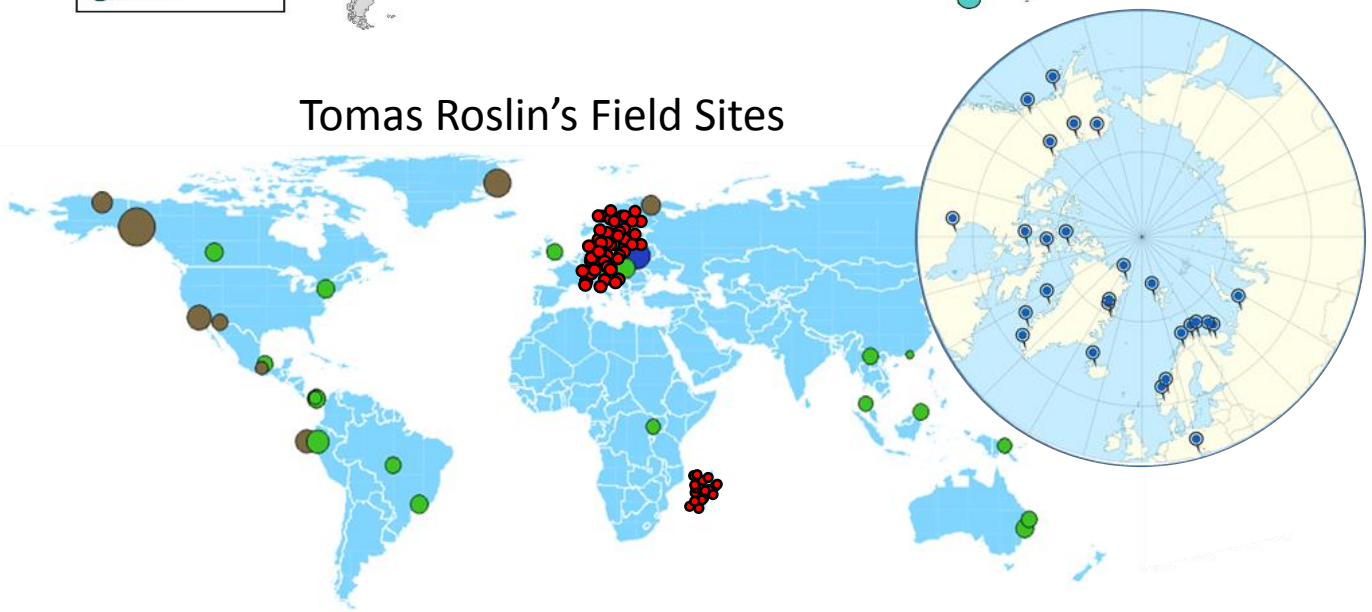
Global Malaise Trap Project



Russian National Parks Project



Tomas Roslin's Field Sites



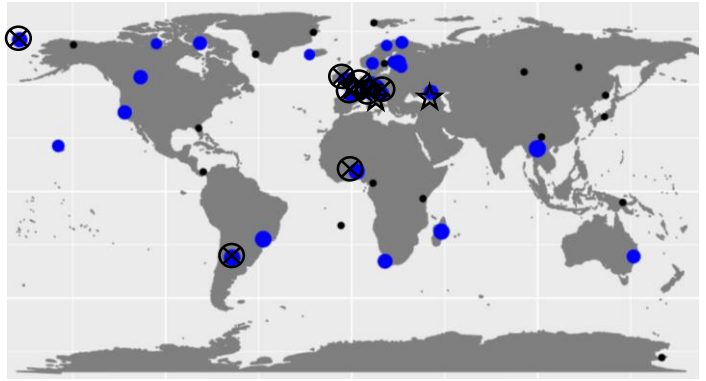
Global Spore Sampling Project pilot data (unpublished)

Cyclone sampler

- **336 samples**
- **31 locations**
- **100 million sequences**
- **160,000 species**

Locations from which the pilot data were acquired

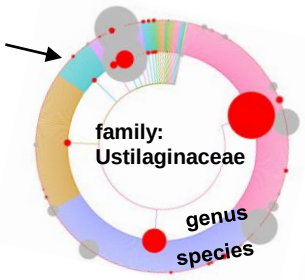
Other locations with ongoing sampling

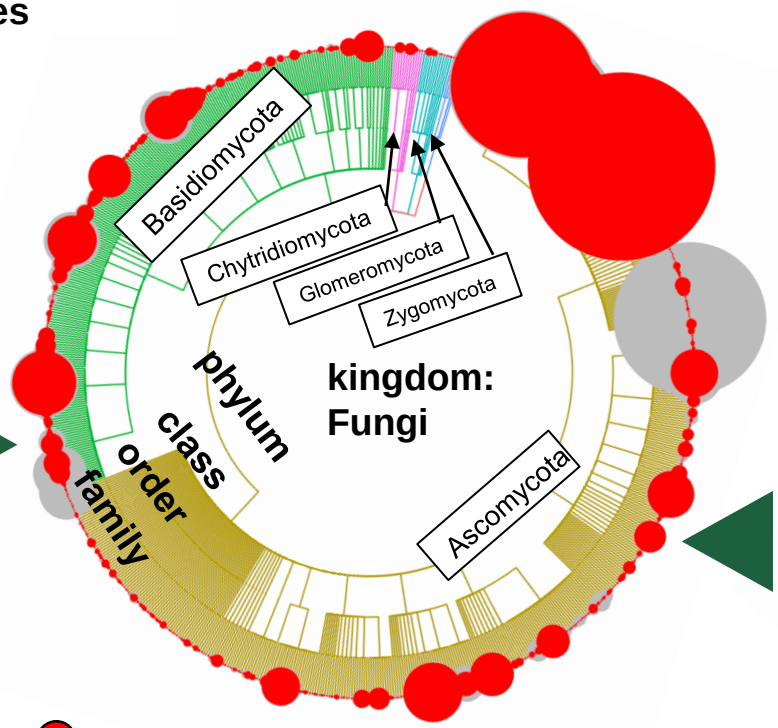


Sampling locations where we recorded *U. maydis*

Sampling locations where we recorded *A. fumigatus*

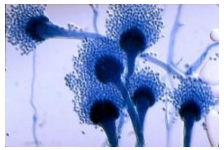
Corn smut is a plant disease caused by the pathogenic fungus *Ustilago maydis* that causes smut on maize and teosinte

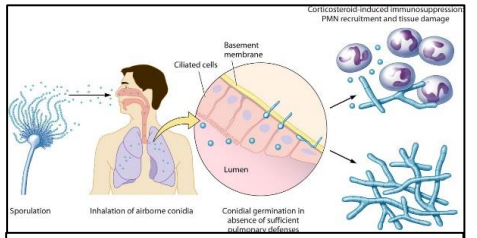




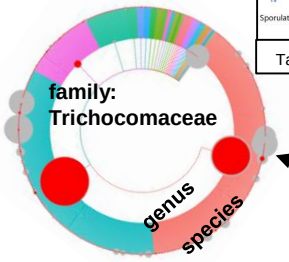
Reliable taxonomic placement

Reliable taxonomic placement currently not possible





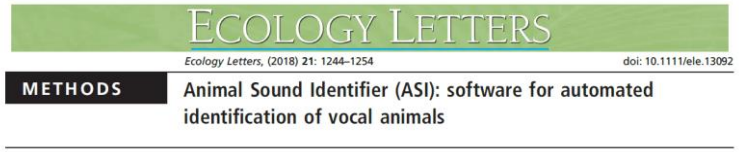
Taylor and Keller (2009). Clinical Microbiology Reviews.



Aspergillus fumigatus causes disease in individuals with immunodeficiency

Audio Pilot data

We used

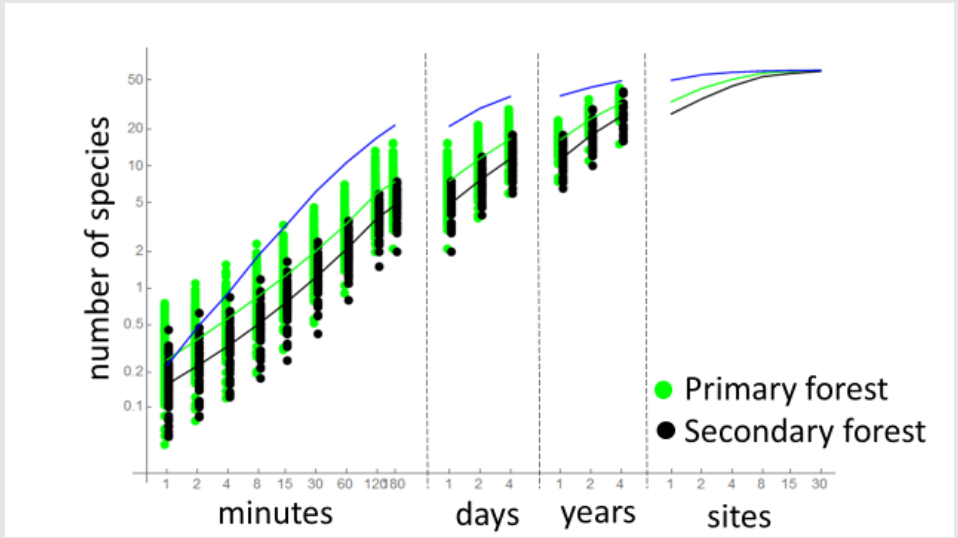


to score 600,000 one-minute audio segments for the presence-absence of 60 diurnal bird species in the Amazon.

We analysed the data to understand….

….the structure of beta-diversity over space and time

Past forest fragmentation can be HEARD for decades

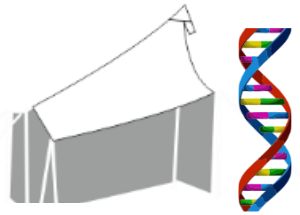


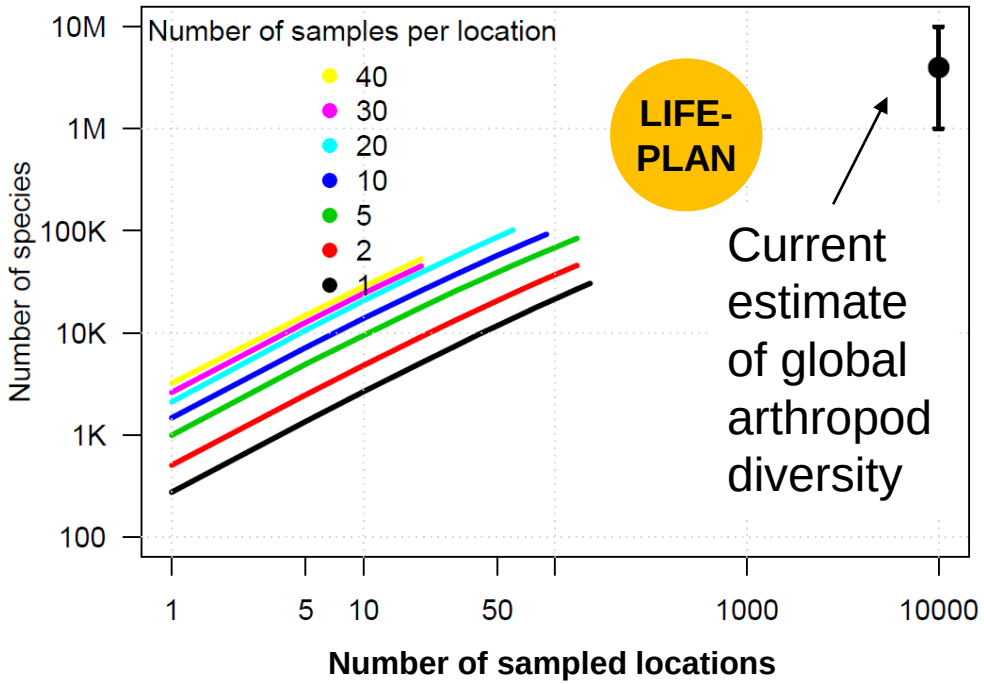
…self-similarity of bird vocalization

[UNDER PREPARATION]

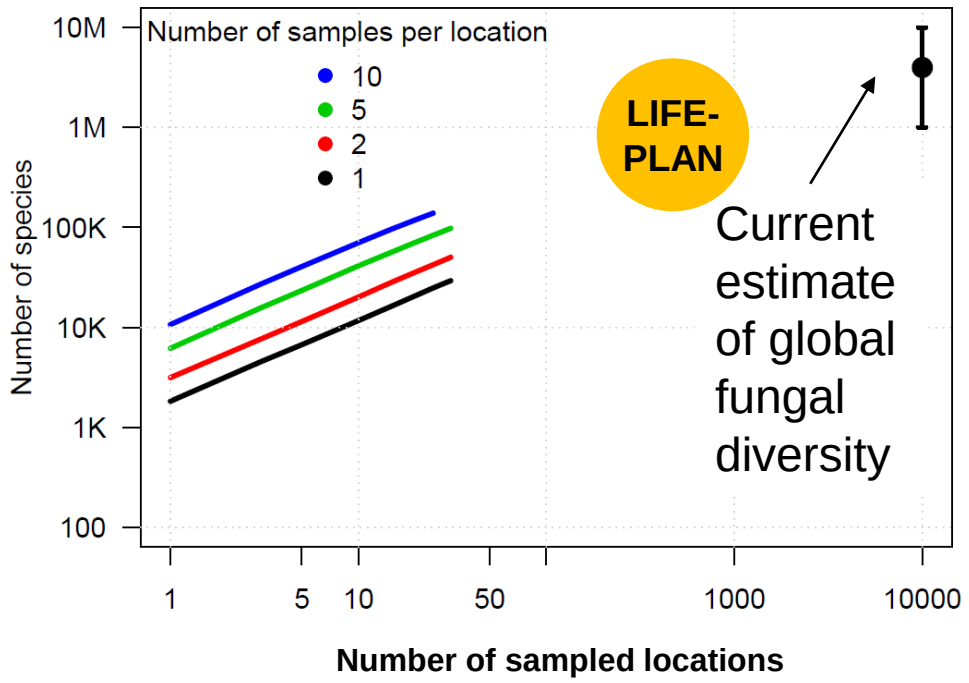
How much sampling is needed to represent Global Biodiversity?

Pilot data from the Global Malaise Trap Project





Pilot data from the Global Spore Sampling Project

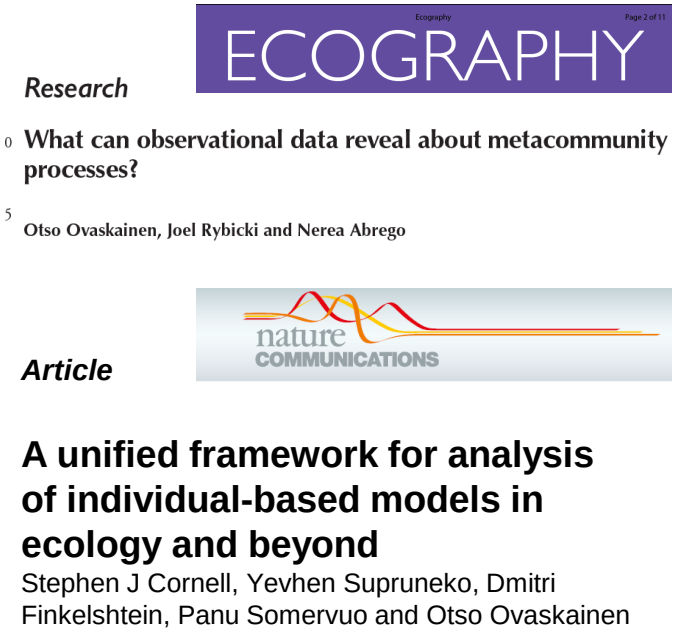


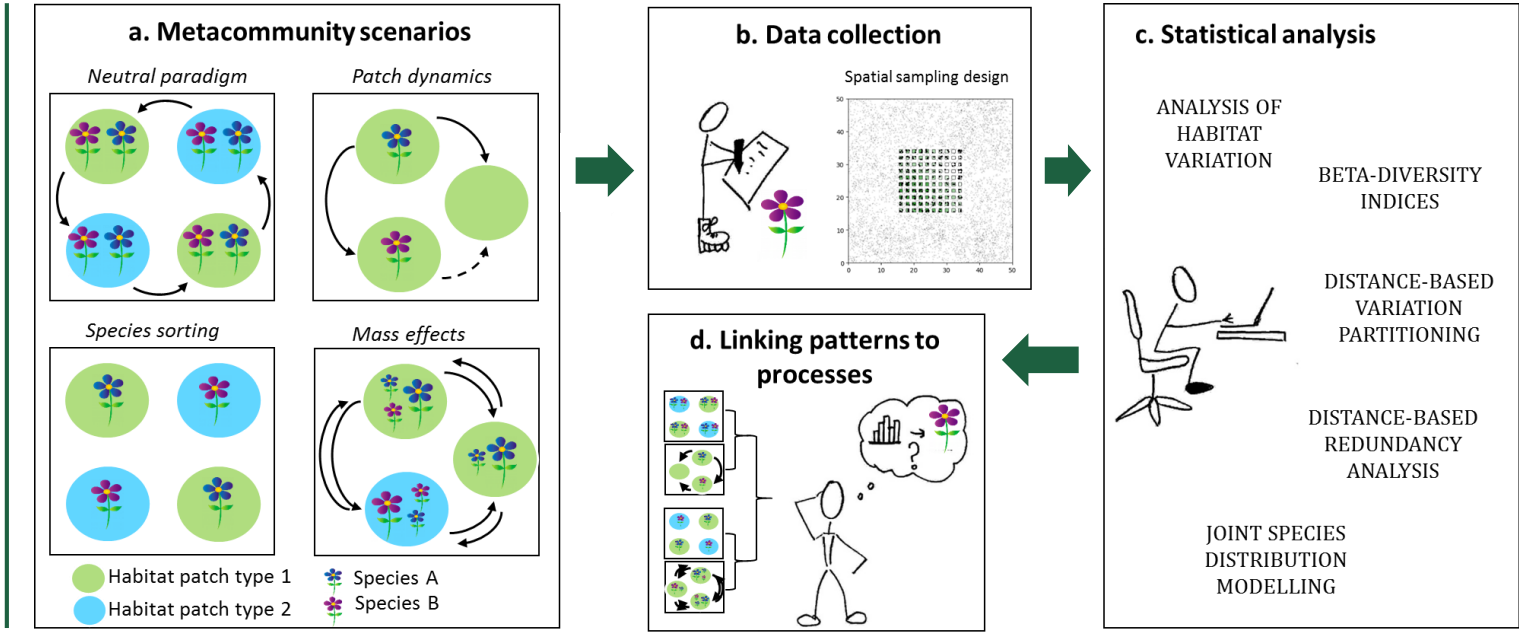
LIFEPLAN will generate 48 samples for each of 450 locations

Is it possible to infer ecological processes from the data on biodiversity patterns?

- Theoretically no, because different processes can result in identical patterns
- Yet our methods are aimed at linking observational data on to the underlying community assembly processes

Papers in press providing tools to link data on patterns to underlying processes:





Recent research activity: published in 2019 or in press

1. **OO, DD, TR.** A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels***. Ecological Monographs.***
2. **OO, TR.** SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. ***Molecular Ecology Resources.***
3. **OO, TR.** Spatio-temporal scaling of biodiversity in acoustic tropical bird communities. ***Ecography.***
4. **OO, DD.** Computationally efficient joint species distribution modelling of big spatial data. ***Ecology.***
5. **OO.** Joint Species Distribution Modelling – With Applications in R. ***Cambridge University Press.***
6. **OO.** A unified framework for analysis of individual-based models in ecology and beyond. ***Nature Communications.***
7. **OO.** Soil fertility in boreal forest relates to root-driven nitrogen retention and carbon sequestration in the mor layer. ***New Phytologist.***
8. **OO.** Handbook for standardised measurement of macrofungal functional traits; a start with basidiomycete wood fungi. ***Functional Ecology.***
9. **OO.** Long-term shifts in water quality show scale-dependent bioindicator responses across Russia - insights from 40 year-long bioindicator monitoring program. ***Ecological Indicators.***
10. **OO.** Joint species movement modelling: how do traits influence movements? ***Ecology.***
11. **OO.** Scaling up the effects of inbreeding depression from individuals to metapopulations. ***Journal of Animal Ecology.***
12. **OO.** Experimentally induced community assembly of polypores reveals the importance of both environmental filtering and assembly history. ***Fungal Ecology.***
13. **OO.** Adaptation to local climate in multi-trait space: evidence from silver fir (Abies alba Mill.) populations across a heterogeneous environment. ***Heredity.***
14. **OO.** What can observational data reveal about metacommunity processes? ***Ecography.***
15. **OO.** Morphological traits predict host-tree specialization in wood-inhabiting fungal communities. ***Fungal Ecology.***
16. **OO.** The microbiome of the Melitaea cinxia butterfly shows marked variation but is only little explained by the traits of the butterfly or its hostplant. ***Environmental Microbiology.***
17. **OO.** Metapopulation models. Encyclopedia of Ecology, ***Elsevier.***
18. **OO.** Species distribution models. Handbook of Environmental and Ecological Statistics, ***Chapman & Hall/CRC.***
19. **OO.** Temporal sampling and abundance measurement influences support for occupancy-abundance relationships. ***Journal of Biogeography.***
20. **TR.** An ecosystem-wide reproductive failure with more snow in the Arctic. PLoS Biology.
21. **TR.** Spatial variability in a plant-pollinator community across a continuous habitat: high heterogeneity in the face of apparent uniformity. ***Ecography.***
22. **TR.** A quantitative framework for investigating the reliability of empirical network construction. ***Methods in Ecology and Evolution.***
23. **TR.** Landscape connectivity explains interaction network patterns at multiple scales. ***Ecology.***
24. **TR.** Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. ***Molecular Ecology Resources.***
25. **TR.** Flower-visitor communities of an arcto-alpine plant– global patterns in species richness, phylogenetic diversity and ecological functioning. ***Molecular Ecology.***
26. **TR.** Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. ***Ecography.***
27. **TR.** Assessing changes in arthropod predator-prey interactions through DNA-based gut content analysis – variable environment, stable diet. ***Molecular Ecology.***
28. **TR.** Finding flies in the mushroom soup: host specificity of fungus-associated communities revisited with a novel molecular method. ***Molecular Ecology.***
29. **TR.** Impacts of urbanization on insect herbivory and plant defenses in oak trees. ***Oikos.***
30. **TR.** Special issue on species interactions, ecological networks and community dynamics: untangling the entangled bank using molecular techniques. ***Molecular Ecology.***
31. **DD.** Bayesian sparse linear regression with unknown symmetric error. ***Information and Inference.***
32. **DD.** The Hastings algorithm at fifty. ***Biometrika.***
33. **DD.** On posterior consistency of tail index for Bayesian kernel mixture models. ***Bernoulli.***
34. **DD.** Comparing and weighting imperfect models using D-probabilities. ***Journal of the American Statistical Association.***
35. **DD.** Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods. ***Biometrika.***
36. **DD.** Intrinsic Gaussian processes on complex constrained domains. ***Journal of the Royal Statistical Society Series B.***
37. **DD.** Common and individual structure of brain networks***. The Annals of Applied Statistics .***
38. **DD.** Symmetric Bilinear regression for signal subgraph estimation. ***IEEE Transactions on Signal Processing.***
39. **DD.** Tensor network factorizations: Relationships between brain structural connectomes and traits. ***NeuroImage.***
40. **DD.** Nonparametric Bayes models of fiber curves connecting brain regions. ***Journal of the American Statistical Association.***

Co-authors omitted to compact list

Recent research activity: submitted manuscripts

1. **OO, TR.** Monitoring fungal communities with the Global Spore Sampling Project.
2. **OO, TR.** Early and late phenological events are constrained by local differentiation in reaction norms.
3. **OO, TR.** Chronicles of Nature Calendar: A long-term and large-scale multitaxon database on phenology.
4. **OO, TR.** Host-plant specialization of root-associated fungi along elevation: Higher specialization of endophytes than mycorrhizal fungi.
5. **OO.** Defaunation is a key driver of functional loss in a tropical biodiversity hotspot.
6. **OO.** Refining predictions of metacommunity dynamics by modelling species non-independence.
7. **OO.** Bioregions: combining biological and environmental data for management and scientific understanding.
8. **OO.** Movement syndromes of a Neotropical frugivorous bat inhabiting a heterogeneous landscape in Brazil.
9. **OO.** The relative importance of local and regional processes to metapopulation dynamics.
10. **OO.** Co-occurrences of tropical trees: disentangling abiotic and biotic forces.
11. **OO.** Joint species distribution modelling with HMSC-R.
12. **OO.** Ten precautionary principles to ensure ecological soundness of conservation translocations of red-listed wood-inhabiting fungi.
13. **OO.** I. Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude.
14. **OO.** Spatial synchrony is related to the rate of environmental change in Finnish moth communities
15. **OO.** Maternal effects and environmental filtering shape seed fungal communities in oak trees.
16. **OO.** Ilkka Aulis Hanski. 14 February 1953 - 10 May 2016.
17. **OO**. Forest and connectivity loss drive changes in movement behavior of bird species.
18. **OO.** The ghost of the hawk: top predator shaping bird communities in space and time
19. **OO.** Landscape of fear hypothesis explains spatio-temporal occurrence and co-occurrence of large and medium mammals.
20. **OO.** Fragmented tropical forests lose quantity and quality of mutualistic interactions.
21. **OO.** Habitat fragmentation and species diversity in competitive species communities.
22. **OO.** Predicting parasite associations and community composition using joint species distribution models.
23. **OO.** DNA barcoding and modelling illuminate complex host-parasitoid dynamics.
24. **TR.** Local management actions override farming systems in determining dung beetle species richness, abundance and biomass and associated ecosystem services
25. **TR.** Climate and host genotype jointly shape tree phenology, disease levels and insect attacks.
26. **TR.** Host plant phenology, insect outbreaks and herbivore communities – the importance of timing
27. **TR.** Heated rivalries: Shifting phenology modifies competition for pollinators among arctic plants.
28. **TR.** Murderous but sensitive: parasitoids indicate major climate-induced shifts in Arctic communities.
29. **TR.** Land-use intensity affects the potential for apparent competition within and between habitats.
30. **TR.** Compound- and context-dependent effects of antibiotics on greenhouse gas emissions from livestock.
31. **TR.** Contrasting latitudinal patterns in diversity and stability in a high-latitude species-rich moth community
32. **TR.** Birds of a feather advance their breeding together.
33. **TR.** Can school children support ecological research? Lessons from the ‘Oak bodyguard’ citizen science project.
34. **TR**. Land-use intensity affects the potential for apparent competition within and between habitats.
35. **DD.** Removing the influence of a group variable in high-dimensional predictive modelling.
36. **DD.** Bayesian inferences on uncertain ranks and orderings
37. **DD**. Modular Bayes screening for high-dimensional predictors.
38. **DD**. Bayesian distance clustering.
39. **DD**. Bayesian factor analysis for inference on interactions.
40. **DD.** Random orthogonal matrices and the Cayley transform.
41. **DD.** Supervised multiscale dimension reduction for spatial interaction networks.
42. **DD.** Monte Carlo simulation on the Stiefel manifold via polar expansion.
43. **DD.** Bayesian cumulative shrinkage for infinite factorizations.
44. **DD.** Maximum pairwise Bayes factors for covariance structure testing.
45. **DD**. Classification via local manifold approximation.
46. **DD.** Geodesic distance estimation with spherelets.
47. **DD.** Efficient manifold and subspace approximation with spherelets.
48. **DD.** Reducing over-clustering via the powered Chinese restaurant process.
49. **DD.** Targeted random projection for prediction from high-dimensional features
50. **DD.** Estimating densities with nonlinear support using Fisher-Gaussian kernels.
51. **DD.** Constrained Bayesian inference through posterior projections.
52. **DD.** Bayesian modular and multiscale regression..
53. **DD.** Nonparametric Bayesian graphical model for counts.
54. **DD.** Bayesian time-aligned factor analysis of paired multivariate time series
55. **DD.** Multivariate mixed membership modelling: Inferring domain-specific risk profiles.
56. **DD.** Efficient posterior sampling for high-dimensional imbalanced logistic regression.
57. **DD.** Bayesian mosaic: Parallelizable composite posterior.
58. **DD.** Efficient entropy estimation for stationary time series.

Co-authors omitted to compact list

Calibration data for fungal DNA

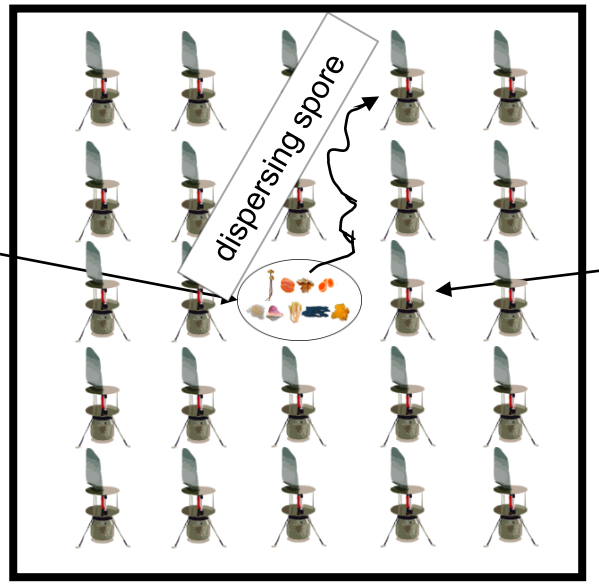
AIM: Convert samples to estimates of spore density and local spore production rate

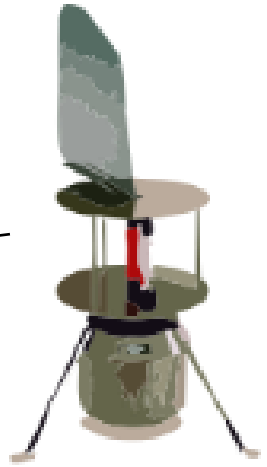
DONE: Convert sequence number to estimates of DNA abundance ”...corresponds to 0.029 ng of fungal eDNA per cubic meter of air….” (Ovaskainen et al., manuscript)

ONGOING: Collected soil and air samples at distances of 1km, 10km, 100km and 500 km to understand how ”regional” the samples are.

TO DO IN LIFEPLAN: Sampling at various distances and directions from known point source.

Fungal fruit-bodies with total spore production rate measured





Grid of cyclone samplers

Calibration data for Malaise traps

AIM: Convert samples to estimates of insect density (number / ha)

DONE: Convert mass-sequenced samples into species-specific estimates of DNA abundance

TO DO IN LIFEPLAN: mass mark-recapture of insects released at different distances from Malaise traps

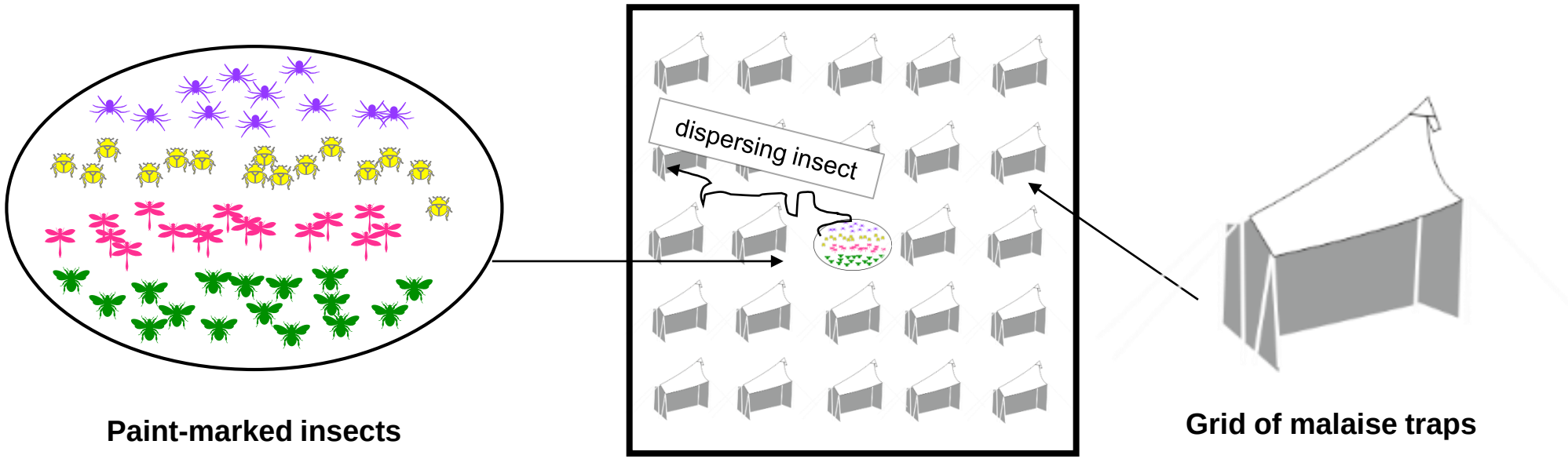
DOI: 10.1111/1755-0998.13057

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes

Yinqiu Ji[1*] | Tea Huotari[2*] | Tomas Roslin[2,3] | Niels Martin Schmidt[4,5] | Jiaxin Wang[1] | Douglas W. Yu[1,6,7] | Otso Ovaskainen[8,9]



Paint-marked insects

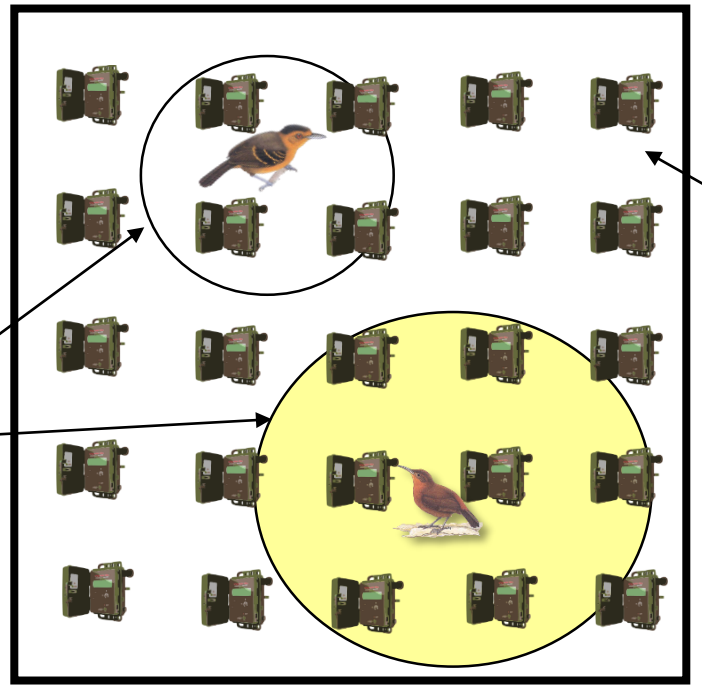
Grid of malaise traps

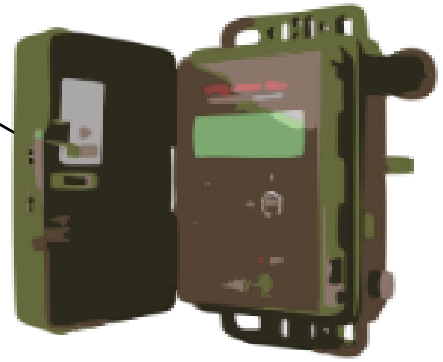
Calibration data for audio recorders

AIM: Convert audio-recordings on birds into estimates of species density (number of individuals / ha) around the sampling area.

TO DO IN LIFEPLAN: Grid recorders to identify detection radius, combined with territory mapping to obtain calibration data on species density

Recorders within the detection radius from the singing birds





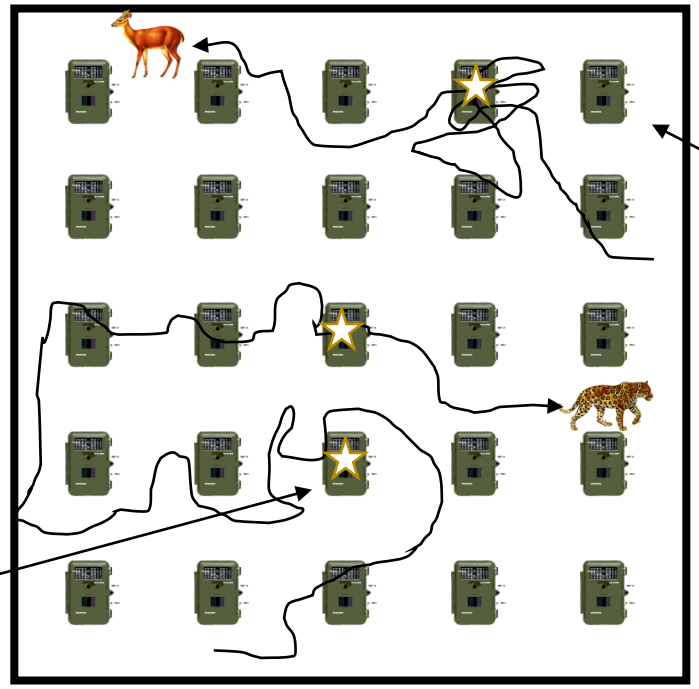
Grid of audio recorders

Calibration data for camera traps

AIM: Convert camera-trap images on mammals into estimates of species density (number / ha)

TO DO IN LIFEPLAN: Detect GPS-collared individuals with a grid of camera traps

Setting a GPS-collard to a jaguar in Pantanal (Brazil). Photo by OO.



Camera traps that detect the GPS-collared individual

Grid of camera traps

LIFEPLAN key collaborators

Prof. Paul Hebert
Director of the Centre for Biodiversity Genomics, University of Guelph, Canada

Integration LIFEPLAN DNA-based species identification methods into the BOLD

Key collaborator for high-throughput sequencing

Access to Global Malaise Trapping Data as an important pilot data

Prof. Brian Fisher
California Academy of Sciences, USA

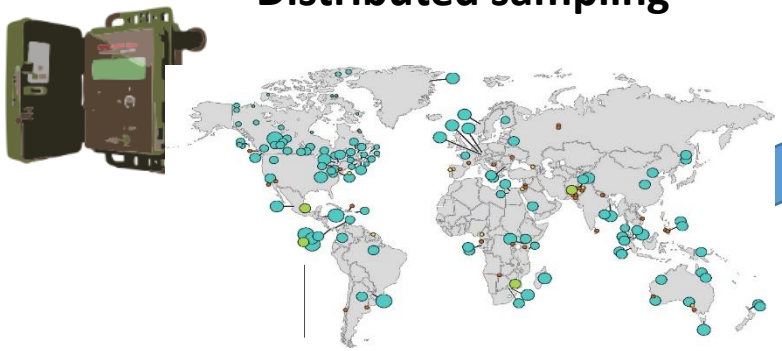
Logistics in Madagascar. Fisher maintains the Madagascar Biodiversity Center in Antananarivo, and holds unparalleled experience in working in this region.

Prof. Fredrik Ronquist
Head of the Department of Bioinformatics and Genetics, Natural History Museum, Sweden

Access to the pilot data generated by the Comparative Insect Biomics project, a one-year sampling campaign run in Sweden and Madagascar

Global Bird Listening Project

Distributed sampling



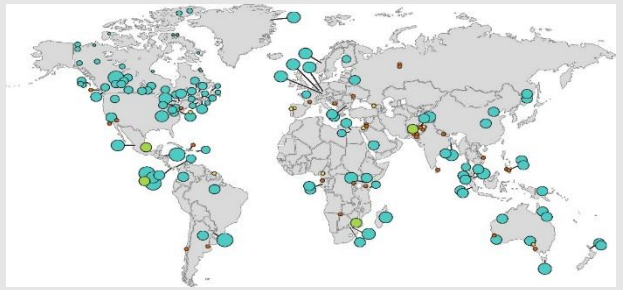
Global audio database

1,000,000,000 audio files of 10 seconds each

Species identifications and their reliability estimates

Parameterization of classification models

Generating training data



Distributed annotation of classification features

Classification features extracted by machine learning methods