Biogeography of Polynesian Pteridophytes in a Global Context

Joel Nitta, Alex White, Warren Wagner, & Eric Schuettpelz

National Museum of Natural History, Smithsonian Institution

3rd Annual Digital Data Conference 🌎

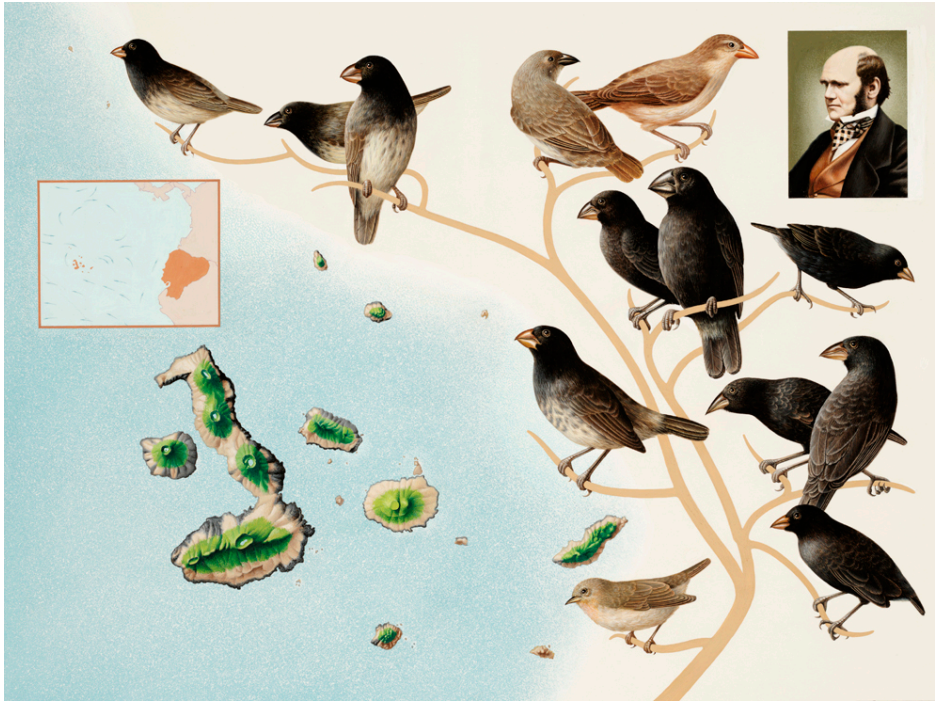
2019.06.10

https://joelnitta.com

Ever since Darwin...

Islands have been used as "natural experiments" to study evolution.

- Few species
- Many replicates
- Clear hypotheses

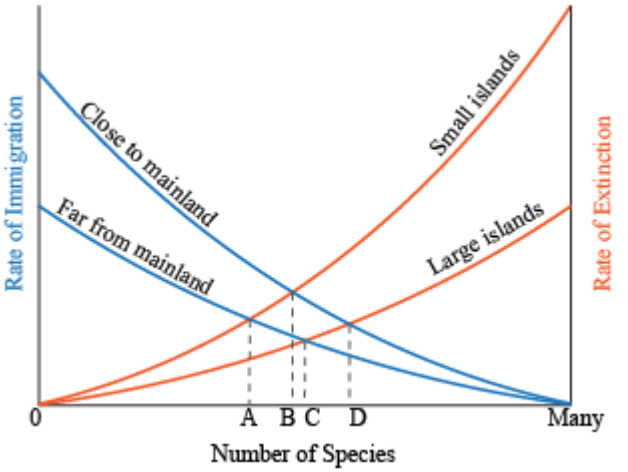


natgeoeducationblog.files.wordpress.com

Theory of Island Biogeography

Species richness
as a factor of

- island size
- island isolation

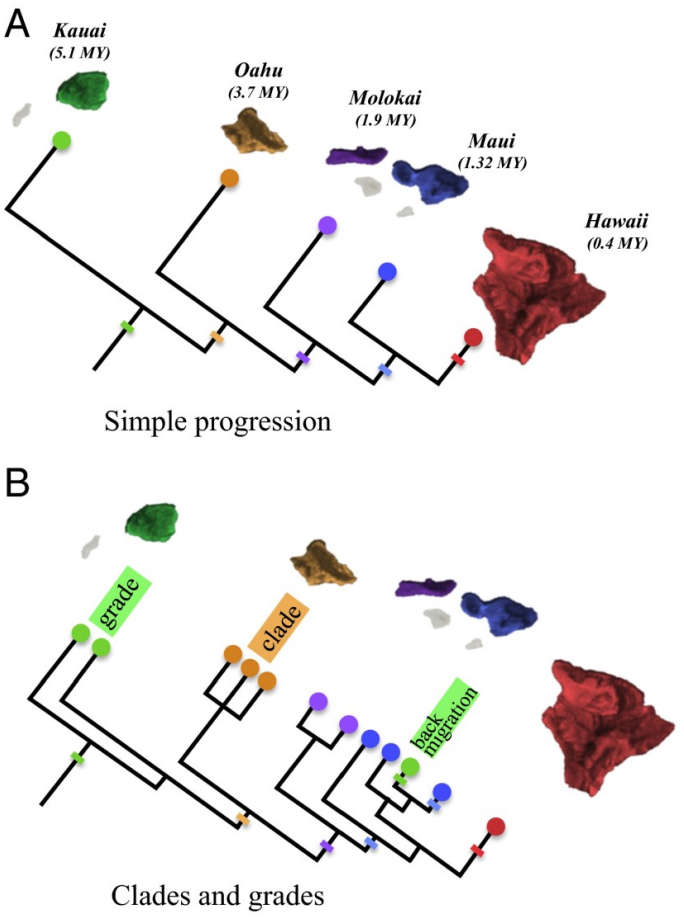


media1.shmoop.com

MacArthur & Wilson (1963, 1967)

Progression rule

Phylogeny recapitulates geology



Wagner & Funk (1995)

Shaw & Gillespie (2016) *Proc Natl Acad Sci USA* 113

However...

Most island studies focus on small clades within a single archipelago

... and that archipelago is usually Hawaii

Few studies investigating patterns at **broader scales**, in other archipelagos

Pteridophytes as a study system

Pteridophytes (ferns and lycophytes)...

- are over-represented on islands
- play important ecological roles
- need more study

fernsoftheworld.com

Pteridophytes as a study system

Pteridophytes (ferns and lycophytes)...

- are over-represented on islands
- play important ecological roles
- need more study

fernsoftheworld.com

Pteridophytes as a study system

Pteridophytes (ferns and lycophytes)...

- are over-represented on islands
- play important ecological roles
- need more study

fernsoftheworld.com

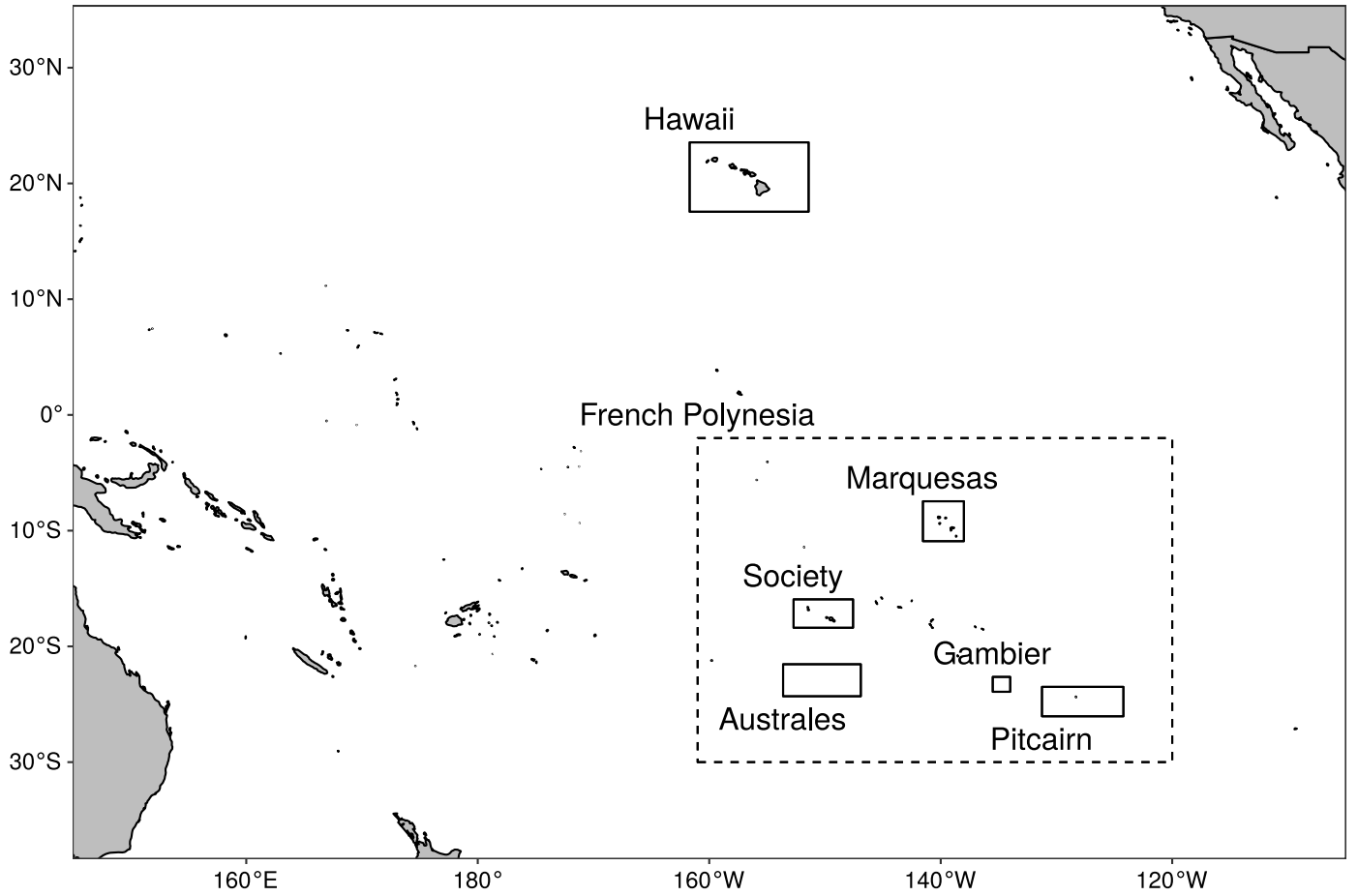
Pteridophytes as a study system

Pteridophytes (ferns and lycophytes)...

- are over-represented on islands
- play important ecological roles
- need more study

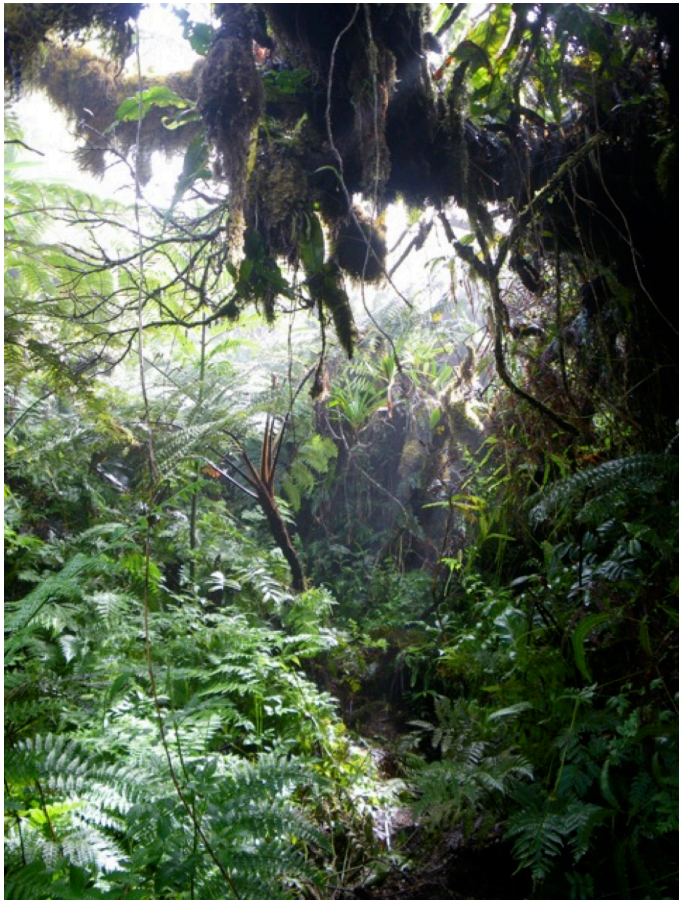
fernsoftheworld.com

Pteridophytes of Polynesia



Pteridophytes of Polynesia

- ca. 413 native spp. total
- 141/182
  endemic/native to HI
- 99/251
  endemic/native to FP
- only ca. 20 native spp. in common



Goals

Trace dispersal **into** Polynesia

Model diversification **within** Polynesia

Goals

Trace dispersal **into** Polynesia ⬅ **Global scale**

Model diversification **within** Polynesia ⬅ **Regional scale**

Need to assemble a dataset that can be used for **both**

Data sources

DNA sequence data: GenBank

Occurrence data:

- GBIF
- Floras
- Collections

Data sources

DNA sequence data: **GenBank**

Occurrence data:

- **GBIF**
- **Floras**
- **Collections**

Four sources of names that need to be harmonized

Taxonomic name resolution strategy

Use Catalog of Life as taxonomic standard

- single taxonomic concept
- 13,994 accepted taxa
- **43,599 synonyms**

GenBank and GBIF: exact match on genus + species

Floras and collections: fuzzy match on full scientific name

Drop any records with names that can't be unambiguously resolved

Occurrence data cleaning

- All occurrences of pteridophytes (ferns and lycophytes) on GBIF:
  9,422,314 initial records.

- Use only records with GPS points, identified to extant species:
  6,552,924 records kept.

- Remove unusual records with CoordinateCleaner:
  6,427,135 records kept.

- Remove records with names that can't be resolved:
  6,370,661 records kept.

Phylogenetic analysis

- Download *rbcL* for all pteridophytes on GenBank*: 11,343 sequences / 5,024 species.
- Resolve names, keep single best sequence per species: 4,150 species.
- Infer phylogeny using maximum likelihood

* `gbfetch` R package, https://github.com/joelnitta/gbfetch

Biogeographic analysis

- Infer **historical** biogeographic movements: DEC model in `BioGeoBears` (Matzke 2013)
- Infer **extant** biogeographic structure: GoM model in `Ecostructure` (White et al. *Nature Comm.*, in press)

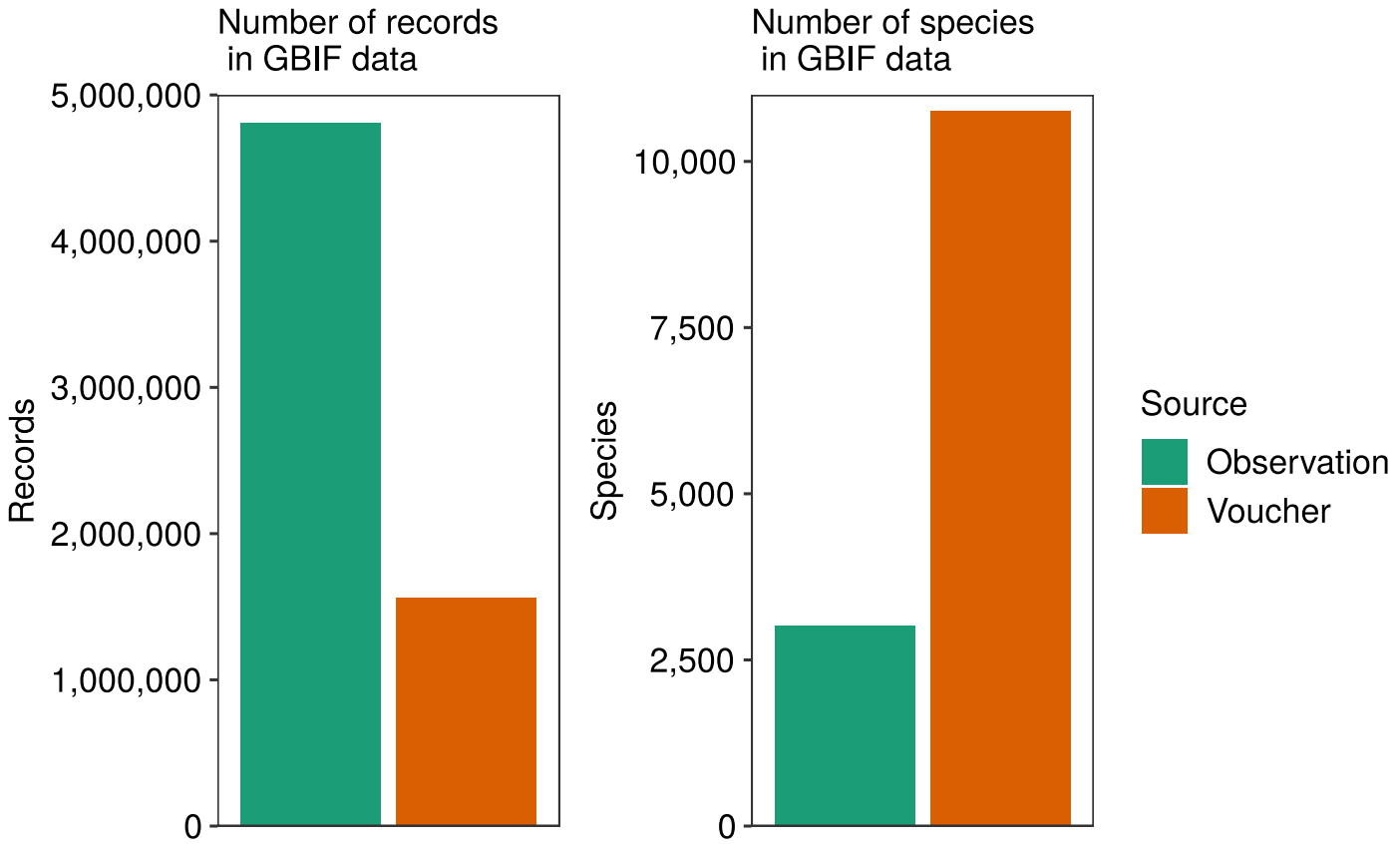
Results

GBIF has good taxonomic representation of pteridophtyes

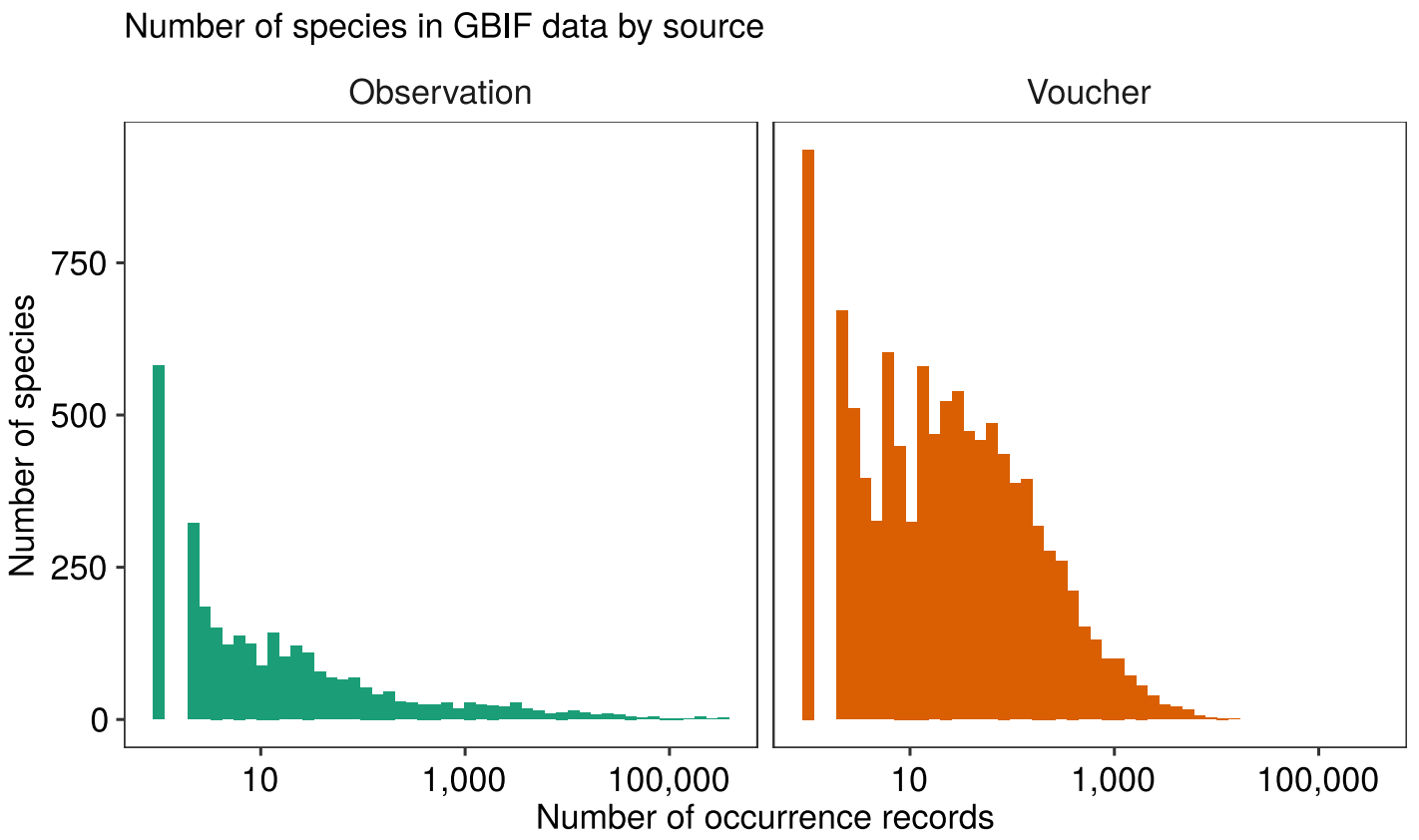
Source	Species	Genera	Families
GBIF data	10789	338	51
Catalog of Life	12996	340	51

How do vouchered records compare to observations? 🤔

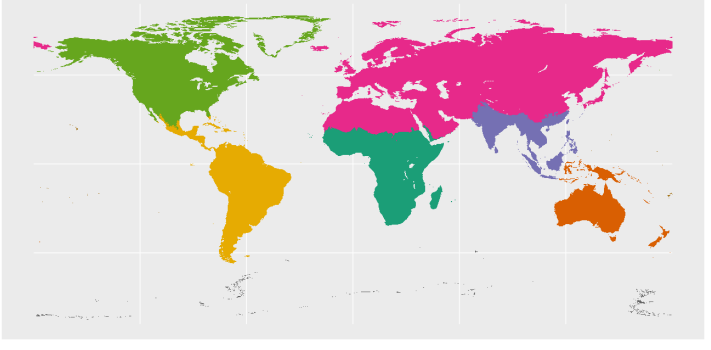
There are many more observation records, representing fewer species

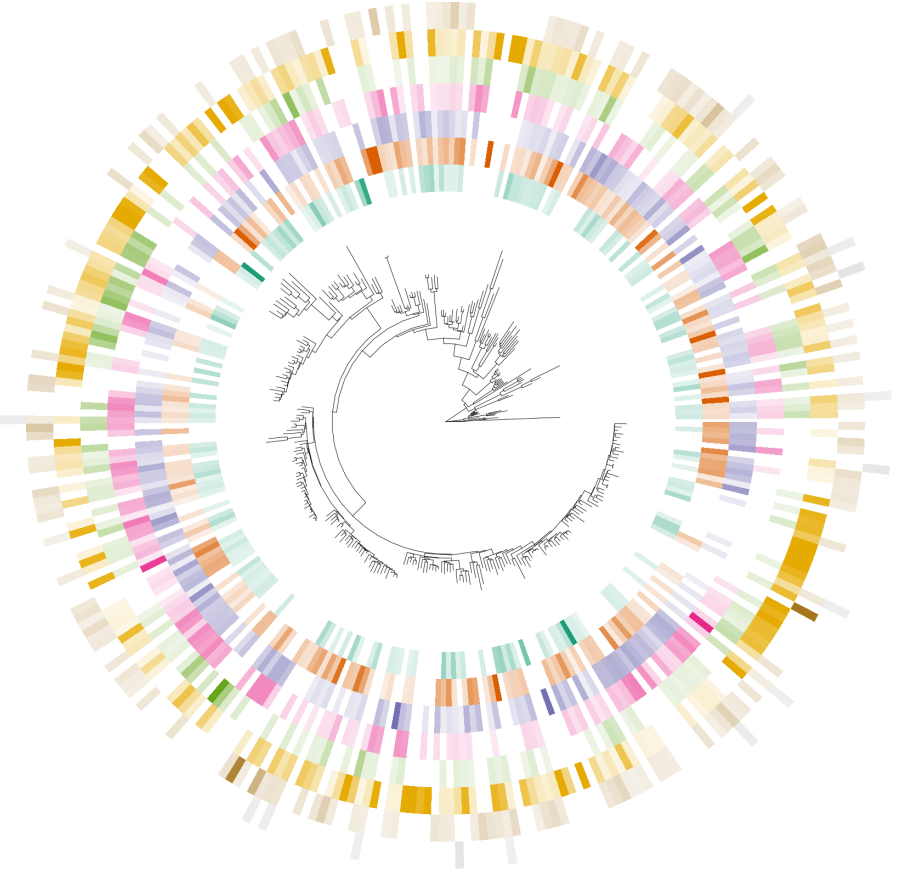


Observation records are taxonomically skewed



Mapping realms onto tree reveals NW/OW clades





rbcL genus-level tree. Colors weighted by relative number of species per realm in each genus. Realms after Dinerstein et al. (2017) *BioSci*: 67.

Classification of biological regions faces two challenges

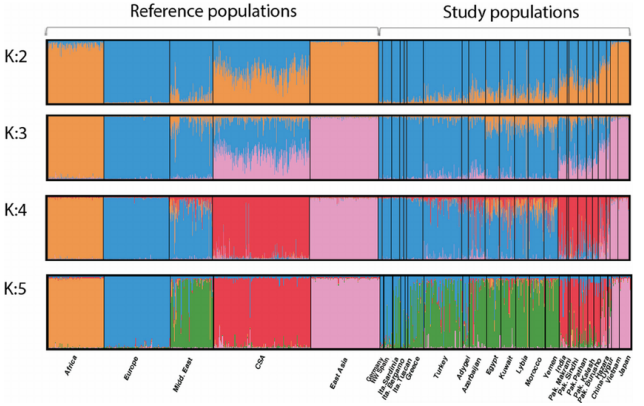
- Either classification must be made *a priori*
- Or, clustering algorithms only allow for **discrete** membership

A new method overcomes these challenges

`ecostructure` (White et al. *Nature Comm.*, in press)

- **estimates** clusters from the data
- allows for membership in clusters to be **continuous**

STRUCTURE (Pritchard et al. 2000) uses a similar framework to assign individuals continuously to genetic groups



Bulbl et al. 2015

ecostructure

The data

G = 10,751 species

N = 10,181 map cells

M_{NxG}

locations	species	g_1	g_2	g_3	g_4	g_5	⋯ G
	n_1	1	1	0	0	0	
	n_2	0	0	1	1	1	
	n_3	1	1	1	1	0	
	⋮						
	N						

The model

$$m_{ng} \sim Ber\,(p_{ng})$$

$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\,\theta_{kg}$$

$$\sum_{k=1}^{K} \omega_{nk} = 1 \qquad 0 \le \theta_{kg} \le 1$$

ecostructure

A toy example with $K = 2$

M_{NxG}

species

locations

	g_1	g_2	g_3
n_1	**1**	**1**	0
n_2	**1**	**1**	0
n_3	0	**1**	**1**
n_4	0	**1**	**1**

=

	k_1	k_2
n_1	.9	.6
n_2	.9	.1
n_3	.6	.9
n_4	.1	.9

×

	g_1	g_2	g_3
k_1	1	1	~0
k_2	~0	1	1

Locations memberships in clusters (theta)

Cluster contribution to species' ranges (omega)

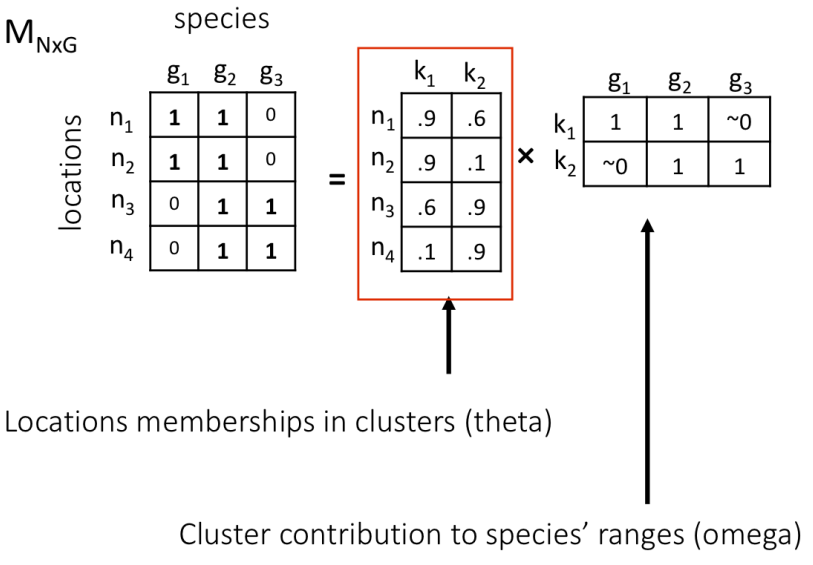
$$m_{ng} \sim Ber\,(p_{ng})$$

$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\,\theta_{kg}$$

$$\sum_{k=1}^{K} \omega_{nk} = 1 \qquad 0 \le \theta_{kg} \le 1$$

ecostructure

A toy example with $K = 2$

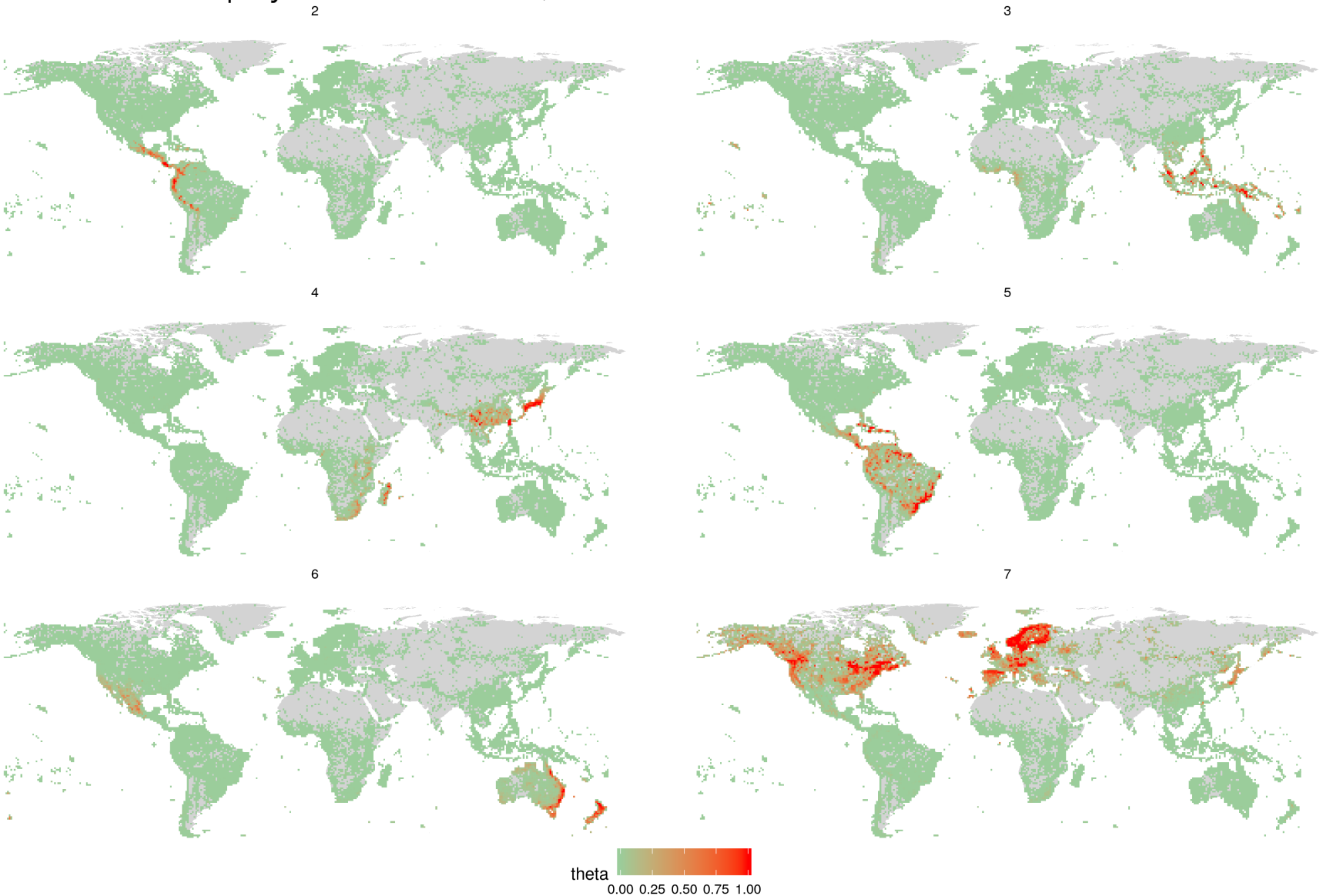


$$m_{ng} \sim Ber\,(p_{ng})$$

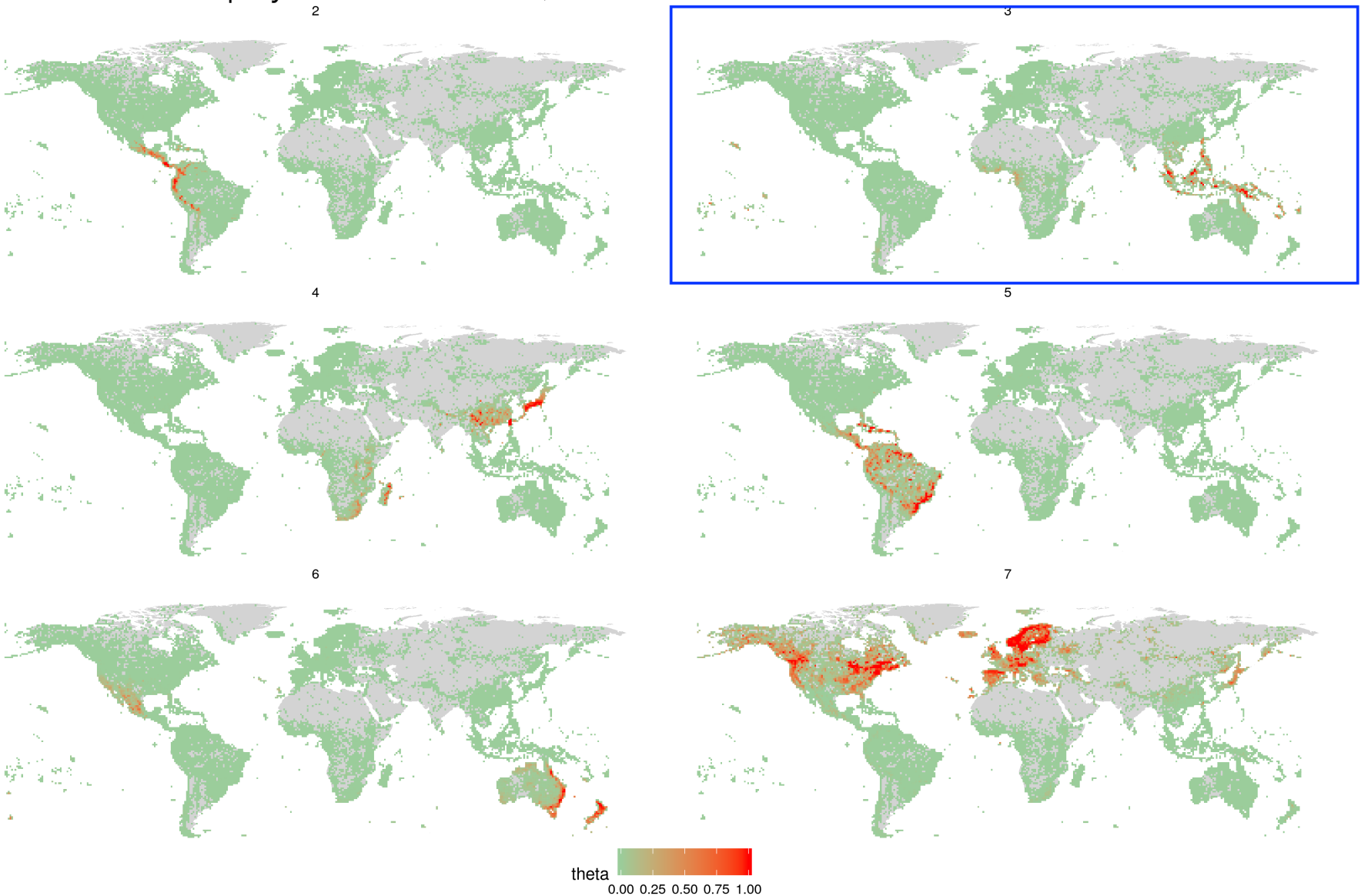
$$p_{ng} = \sum_{k=1}^{K} \omega_{nk}\,\theta_{kg}$$

$$\sum_{k=1}^{K} \omega_{nk} = 1 \qquad 0 \le \theta_{kg} \le 1$$

Pteridophytes of the world, $k = 7$

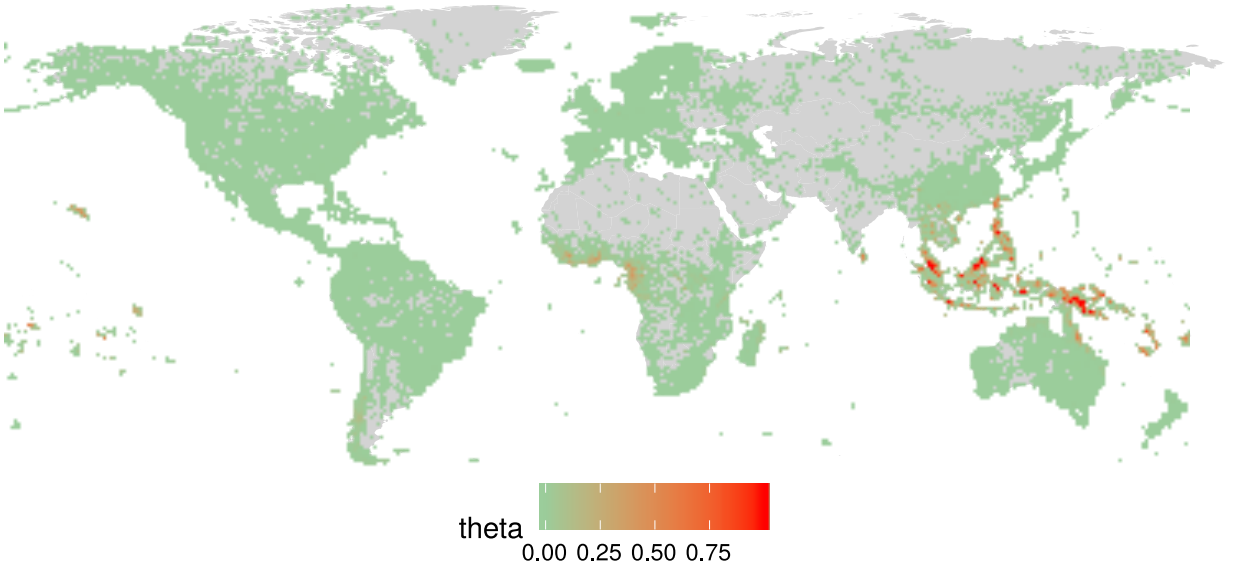


Pteridophytes of the world, $k = 7$



Pacific pteridophytes have affinities with SE Asia

Motif 3



Conclusions

- **Single** source of names with **comprehensive synonymy** is key to resolving taxonomy
- Occurrence data are common in GBIF but should be used with caution
- New world / old world splits a common theme in pteridophyte evolution
- Pacific pteridophytes have affinities with SE Asia

Thank you!

Peter Buck Fellowship

Smithsonian Institution DNA Barcode Network

Pacific Tropical Botanical Garden

Jean-Yves Meyer (Délégation à la recherche, FP), Tom Ranker (UH), Ken Wood (NTBG), David Lorence (NTBG), Tim Flynn (NTBG), Greg Plunkett (NYBG), Mike Balick (NYBG), Ann Kitalong (Belau National Museum)

Extra Slides

iNaturalist records not as common as I thought