Evaluating geographic patterns of morphological disparity in ferns using deep neural networks

Alex White

Postdoctoral Fellow

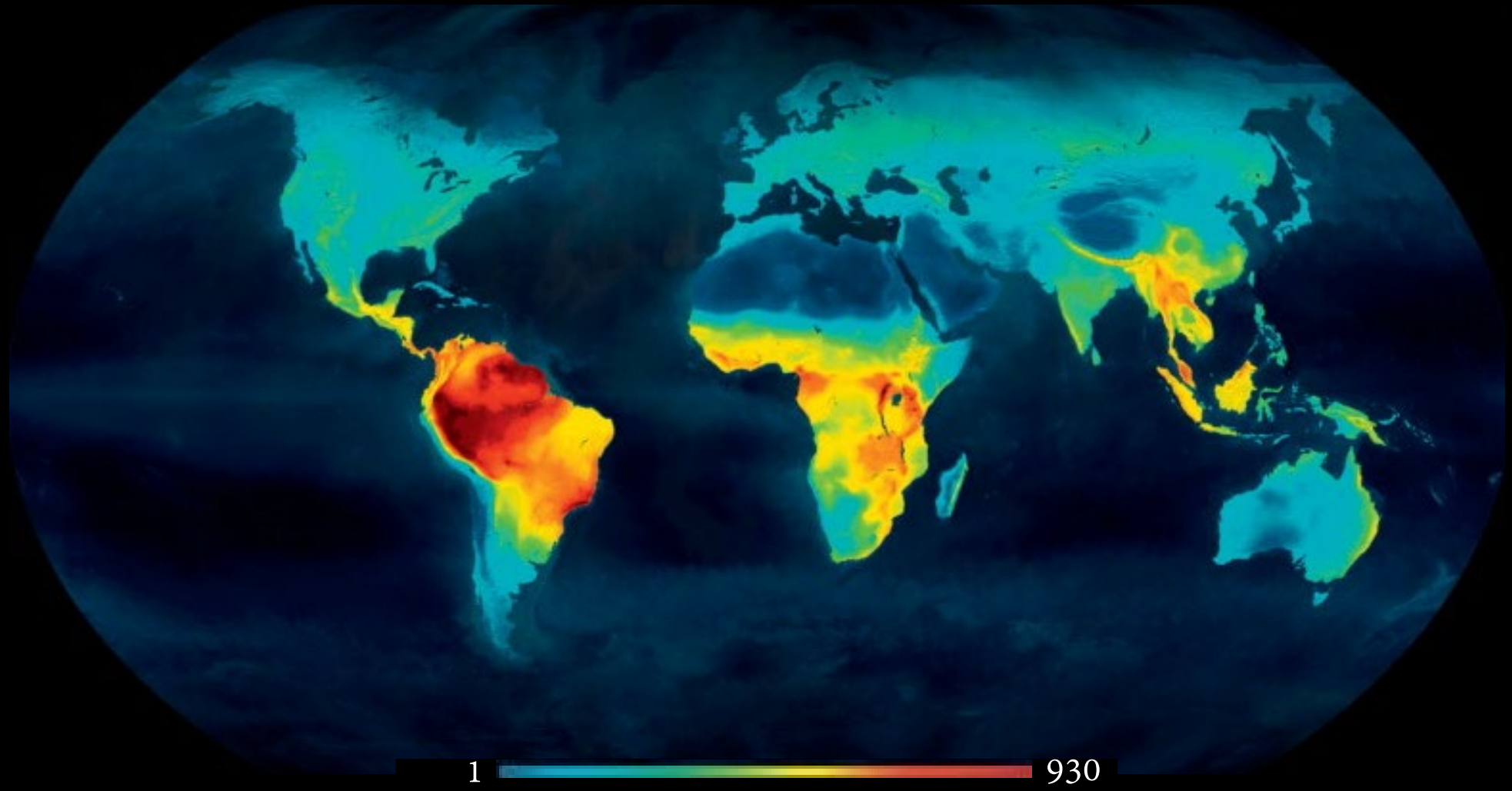
Smithsonian Institution

DATA SCIENCE LAB

NATIONAL MUSEUM *of* NATURAL HISTORY

Photo credit: Masato Shimono, Sandra Richard

The number of species in any place on earth varies widely by latitude



Mannion et al., 2014

The latitudinal diversity gradient is pervasive across taxa:

birds

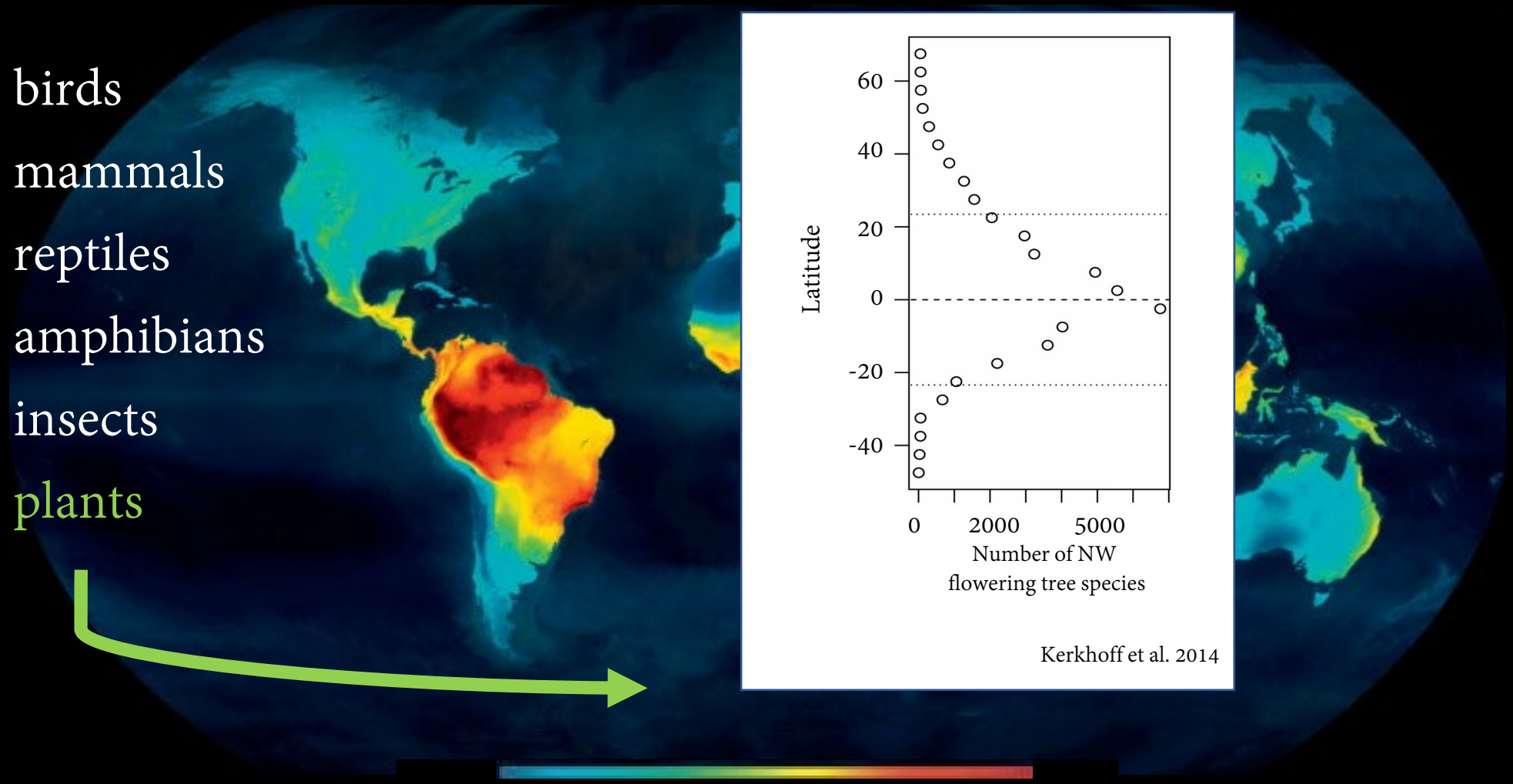
mammals

reptiles

amphibians

insects

plants



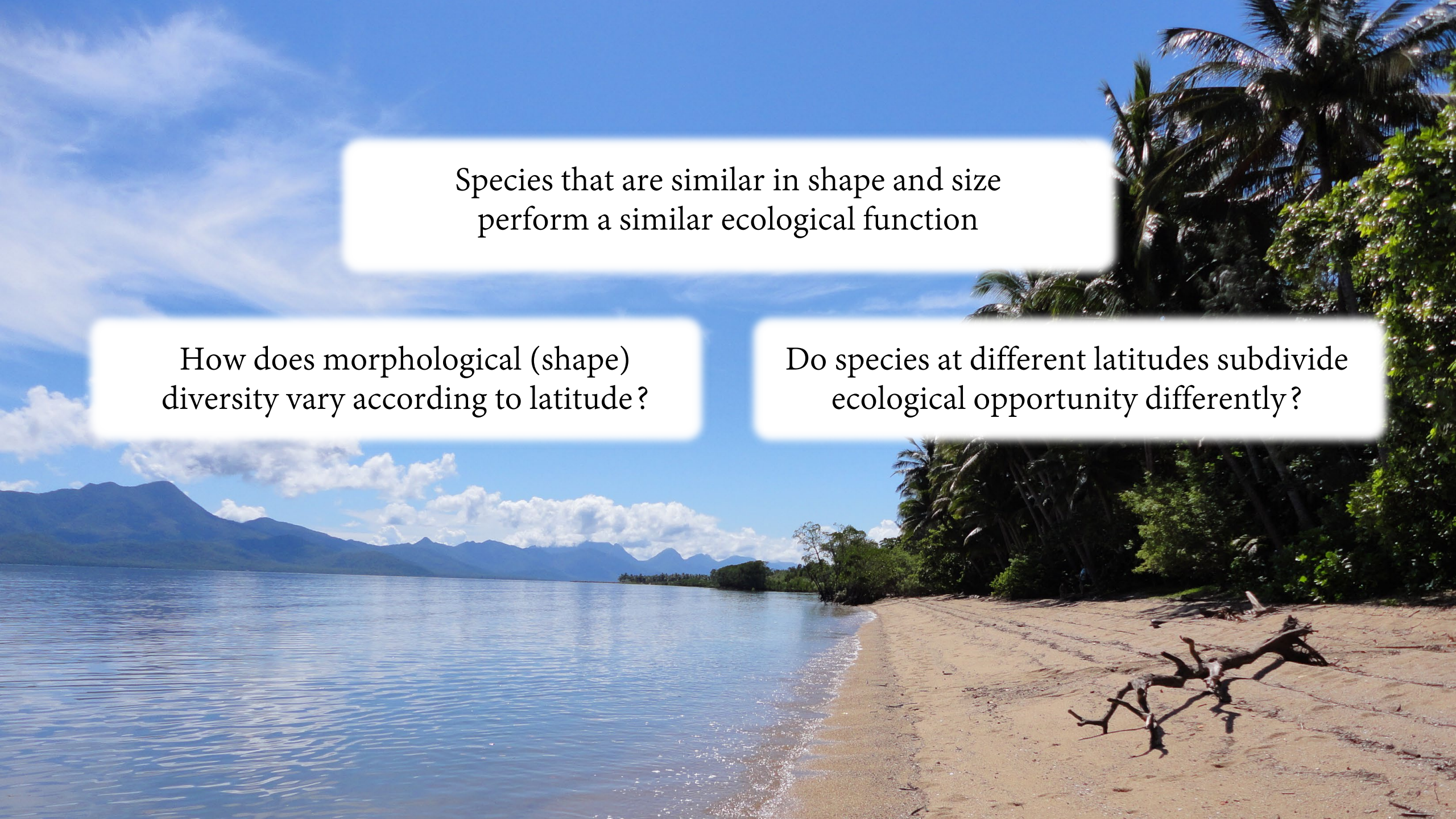
Explanations for the latitudinal diversity gradient generally invoke:

1) Differences in diversification rate
2) Differences in age and area
3) Differences in ecological limits (accommodation)

Species that are similar in shape and size perform a similar ecological function

6 different species (*Zosterops*)

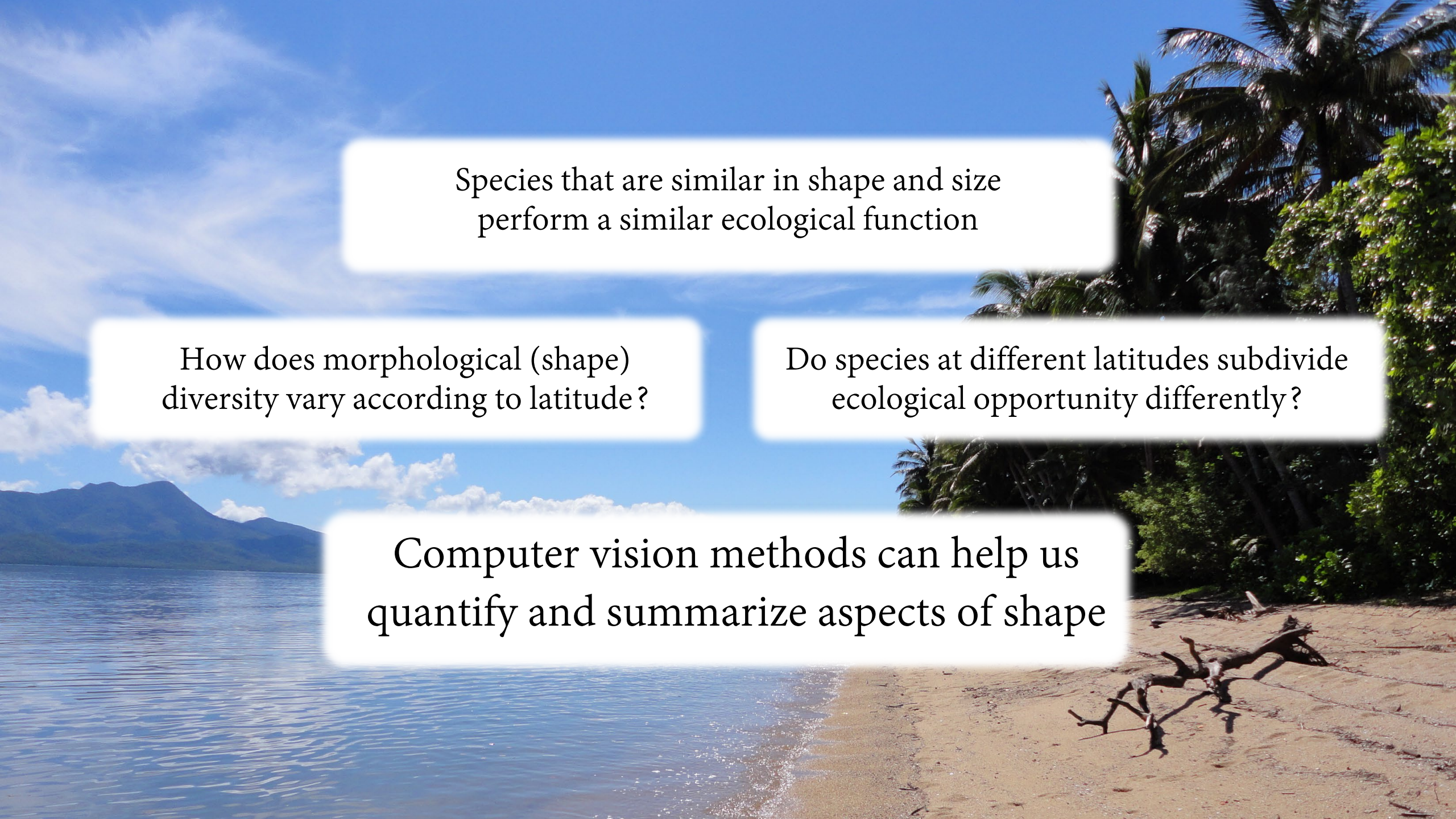
Wells 2017



Species that are similar in shape and size perform a similar ecological function

How does morphological (shape) diversity vary according to latitude?

Do species at different latitudes subdivide ecological opportunity differently?



Species that are similar in shape and size perform a similar ecological function

How does morphological (shape) diversity vary according to latitude?

Do species at different latitudes subdivide ecological opportunity differently?

Computer vision methods can help us quantify and summarize aspects of shape

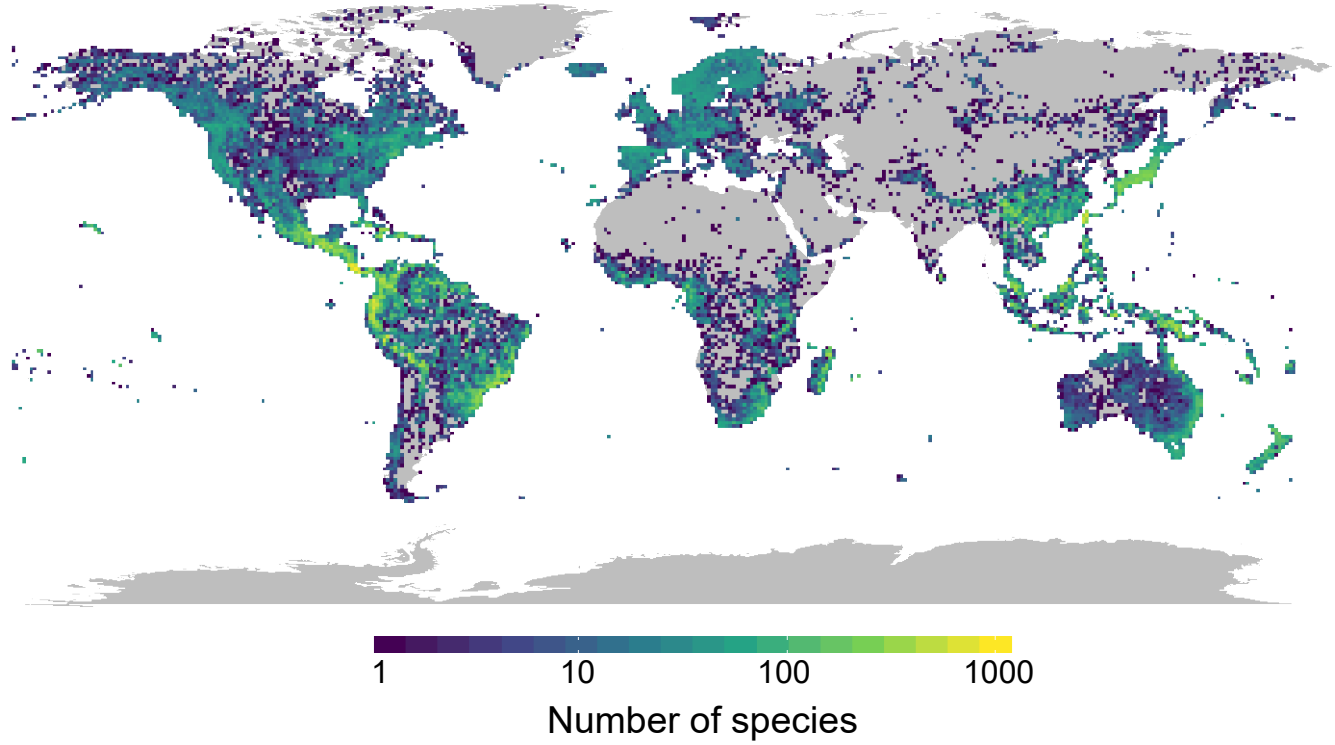
We focus our study on ferns

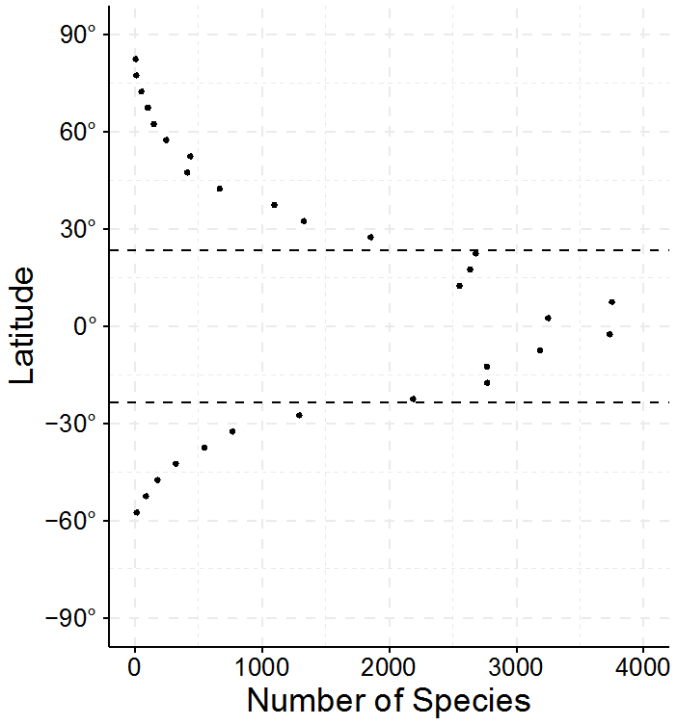
~12,000 species

distributed globally

Fern species numbers vary according to latitude

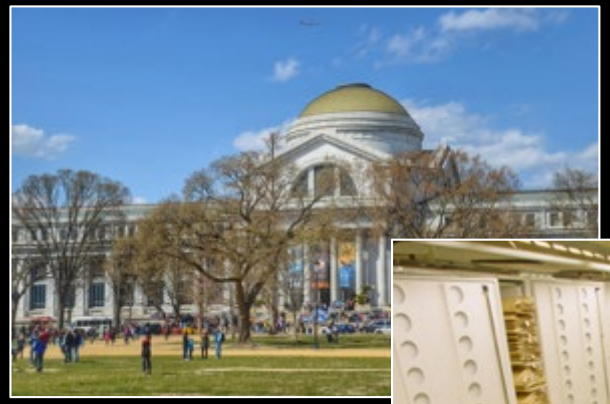
Estimated from 1.4 million museum records (GBIF)





Ferns are highly diverse in both size and shape

Ferns are well represented in digitized collections, including the US National Herbarium



Our data:

All available IDigBio specimens for ferns and lycophytes

~517,000 images

~8600 species

70% global diversity

~320 genera

95% generic diversity

Our computer vision based approach:

Build a convolutional neural net and train it to label fern specimens:

- by genus
  - >500 specimens (86 genera)
  - >50 specimens (269 genera)
- by species
  - >50 specimens (1425 species)

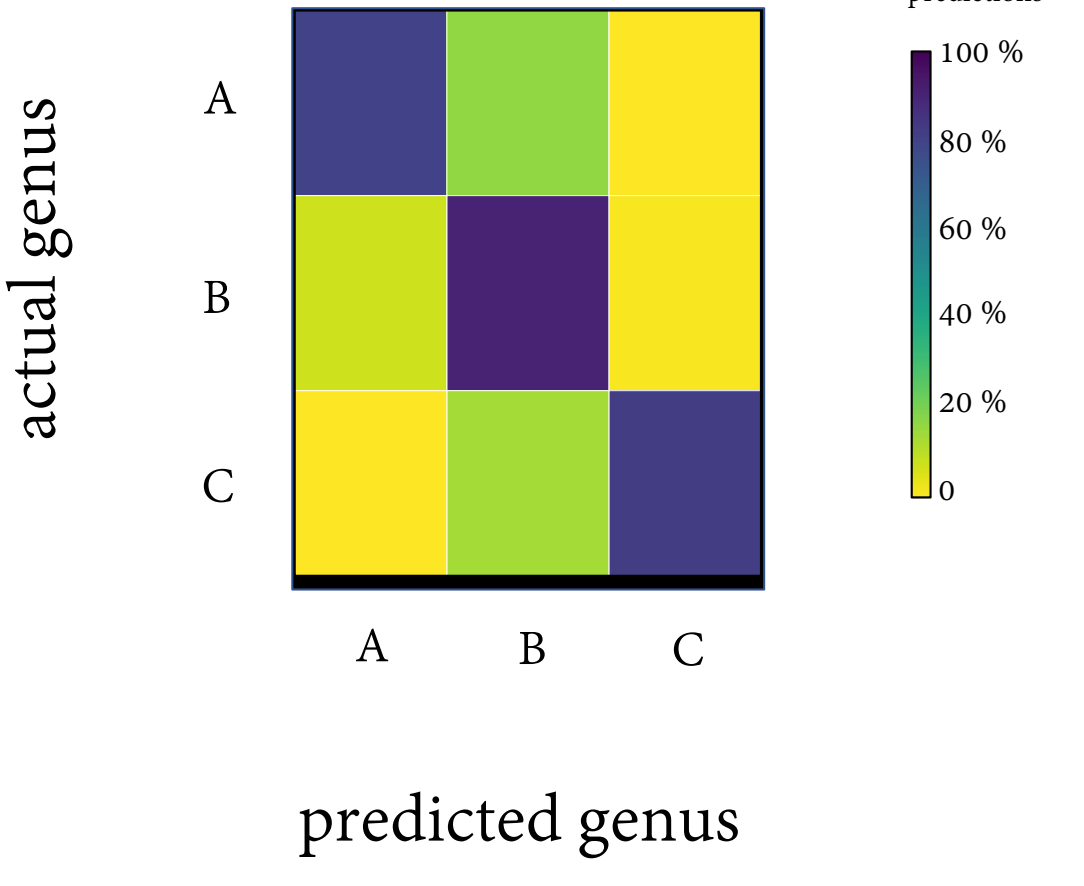
80% of data used for training neural network
20% of data set aside for validation

We validate the model by feeding images with known labels through the network

example genus A

example genus B

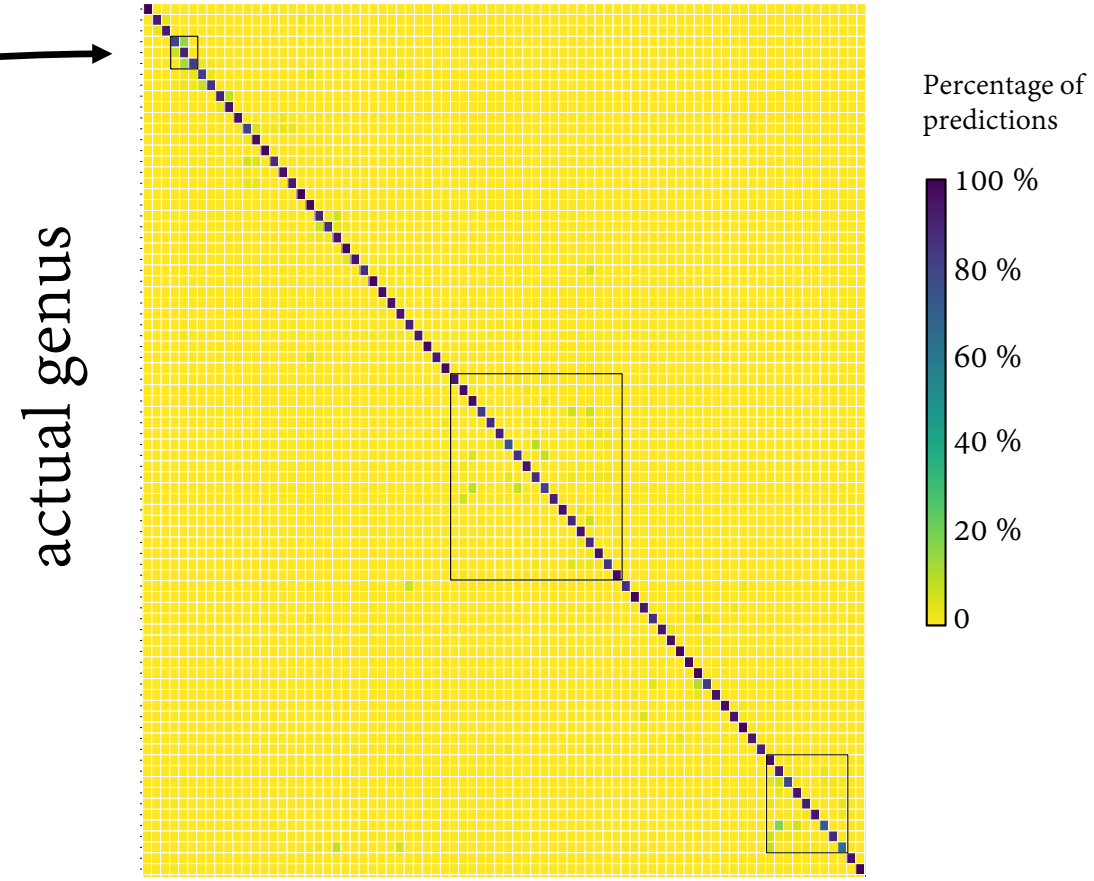
example genus C



FernNet is 97% accurate at genus ID

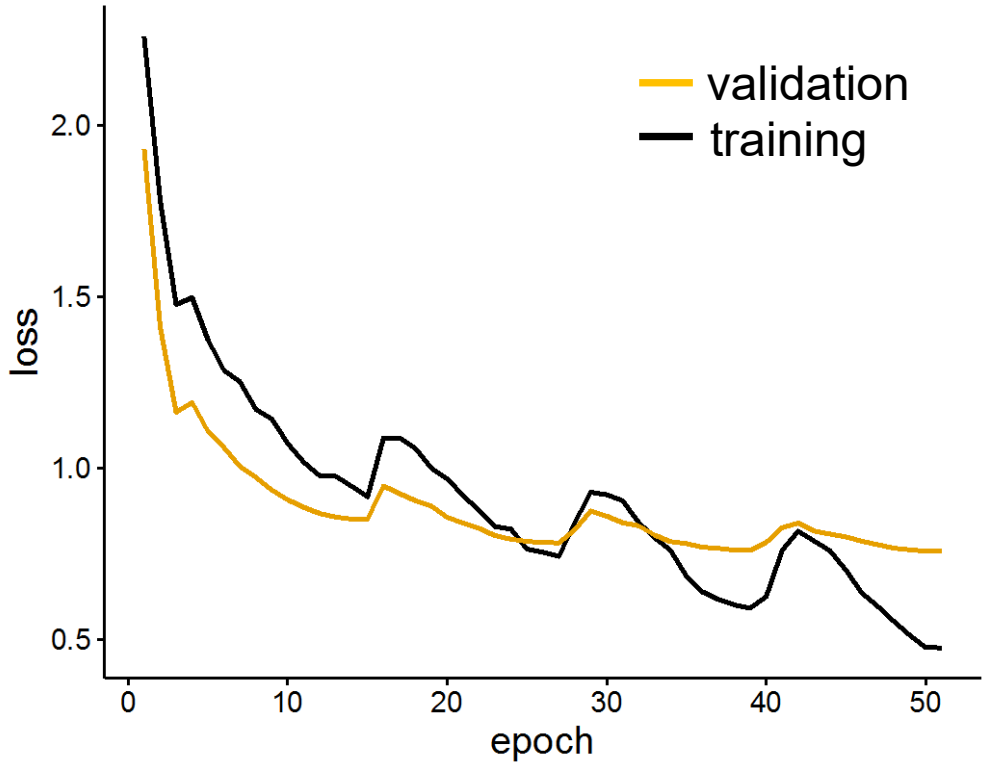
Boxes contain examples of genera within the same family

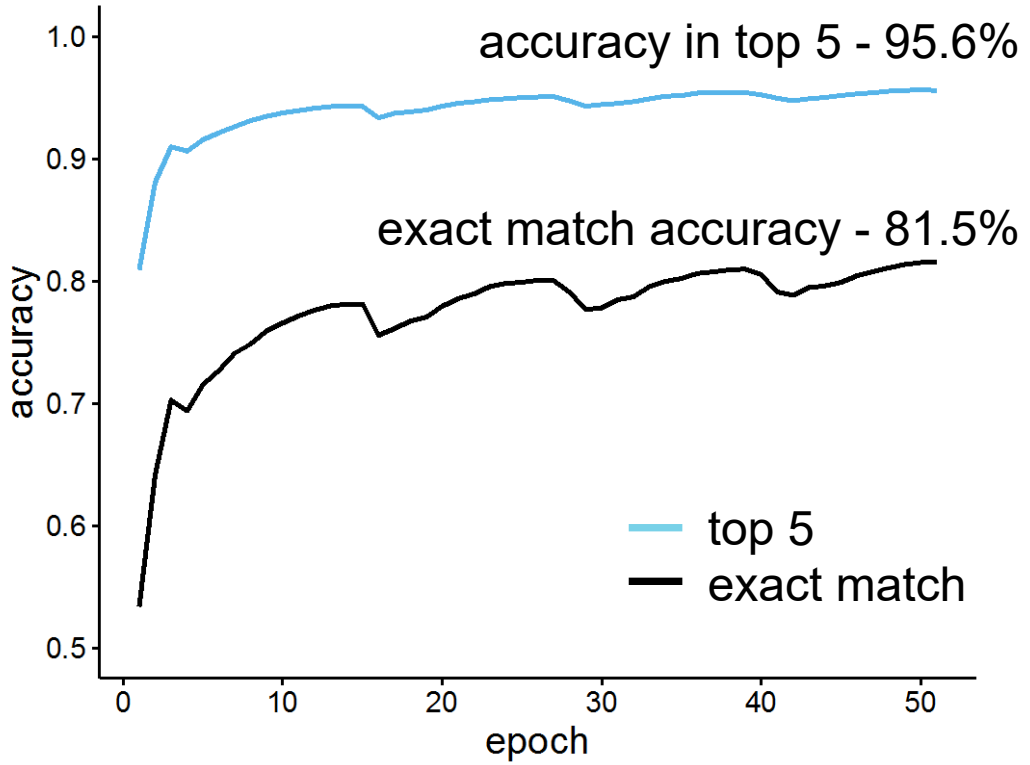
3 genera in the tree fern family Cyatheaceae



Confusion is most often between closely related genera

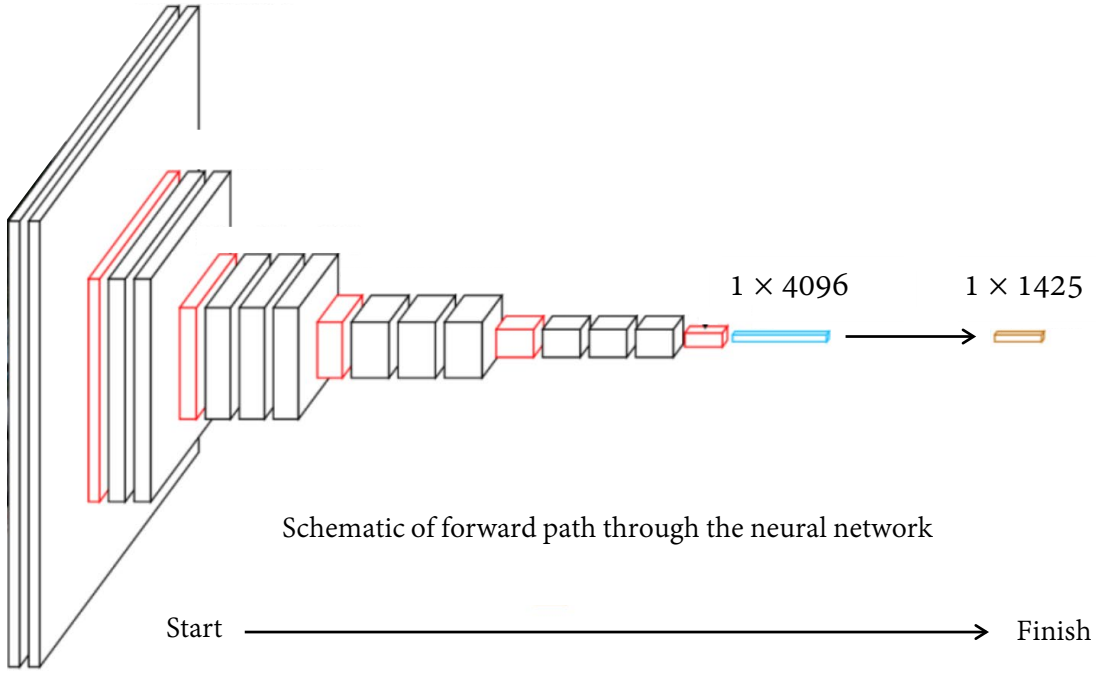
FernNet is highly accurate for species ID (1425 species)





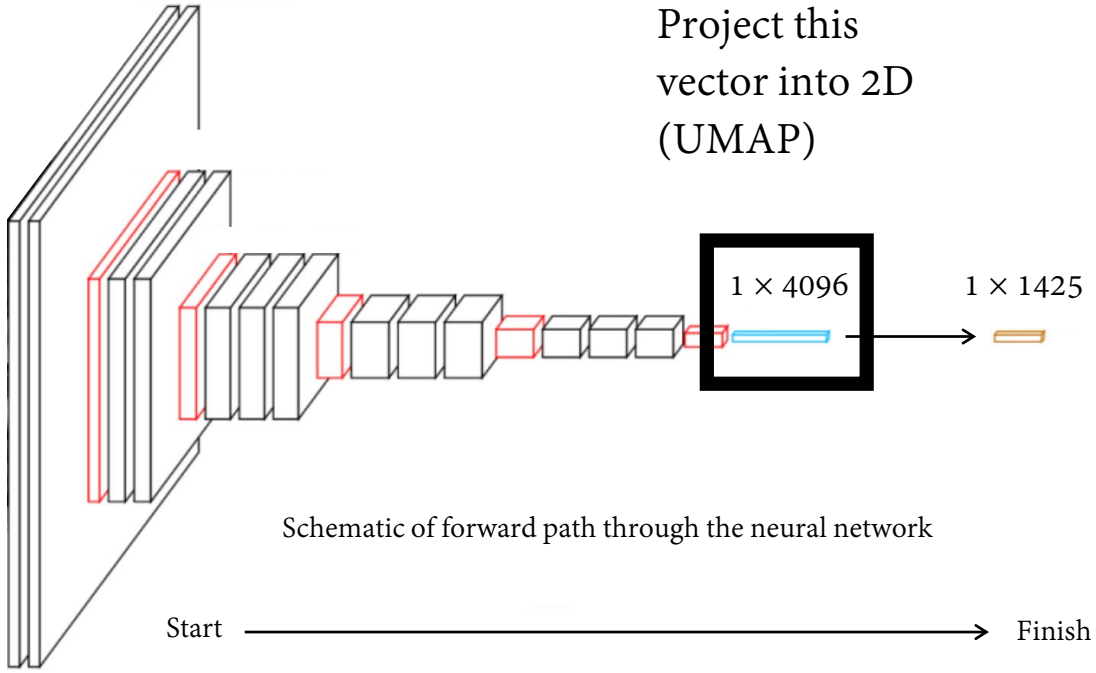
What pixels is FernNet using to make these identifications?

We use the penultimate network layer to quantify shape space occupation



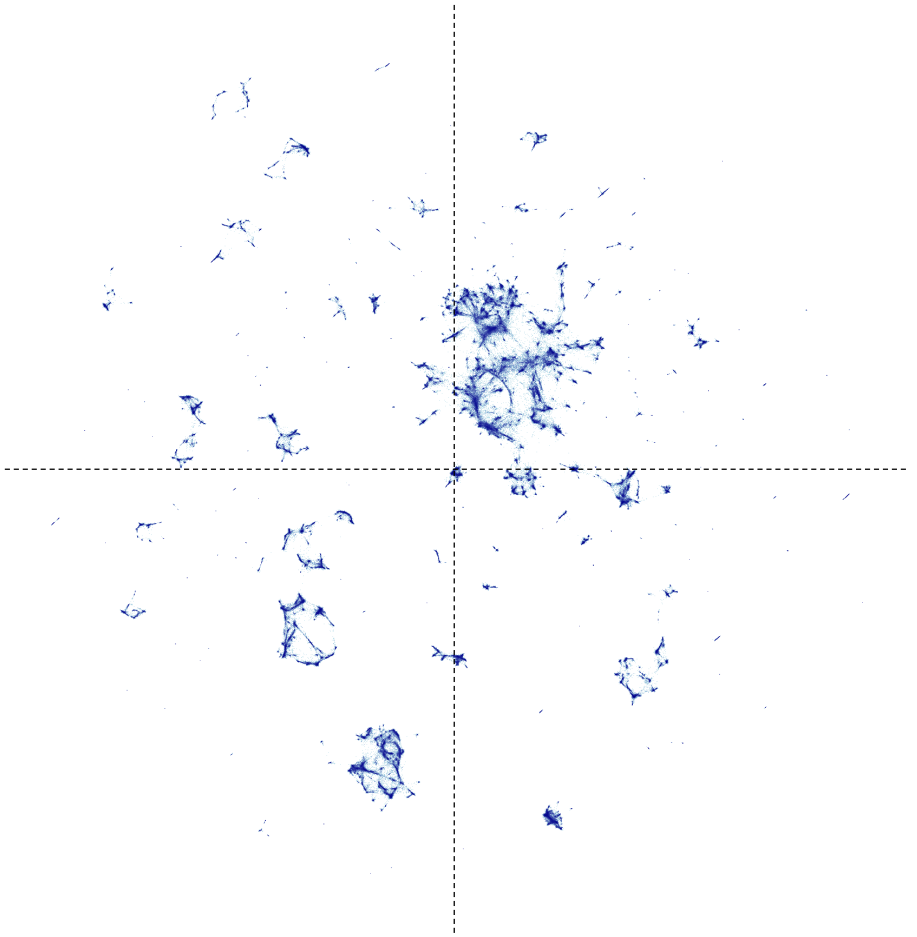
Simonyan and Zisserman 2014

We use the penultimate network layer to quantify shape space occupation

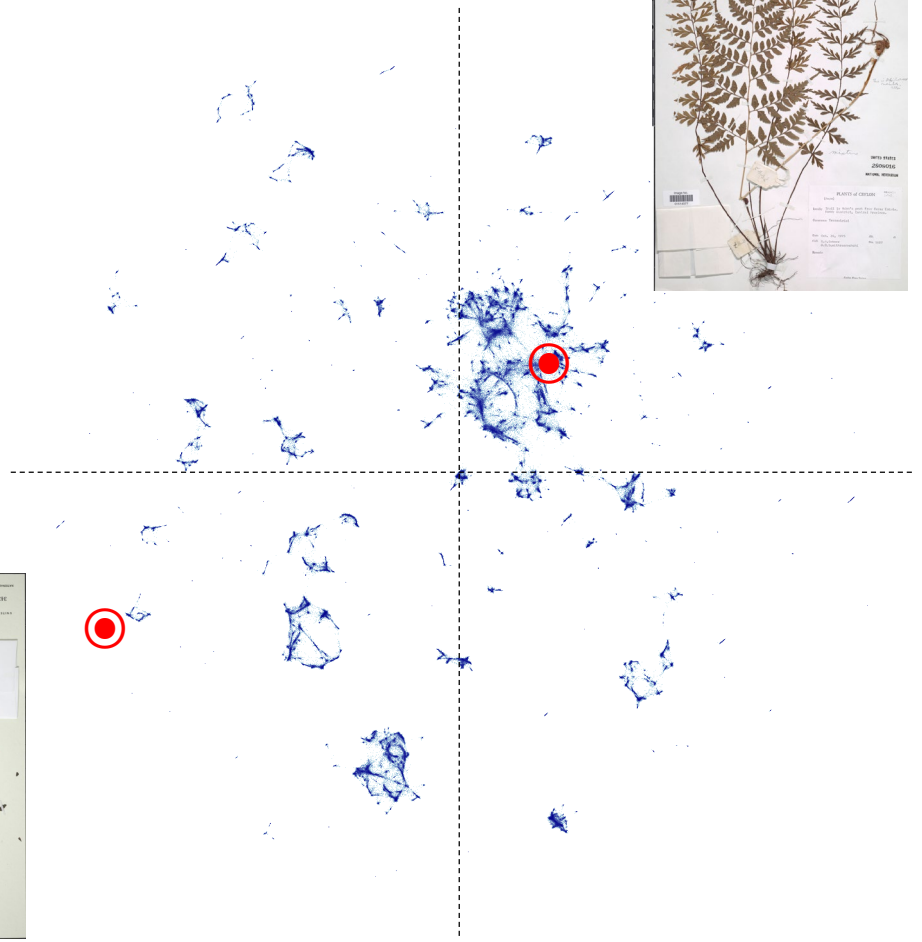


Simonyan and Zisserman 2014

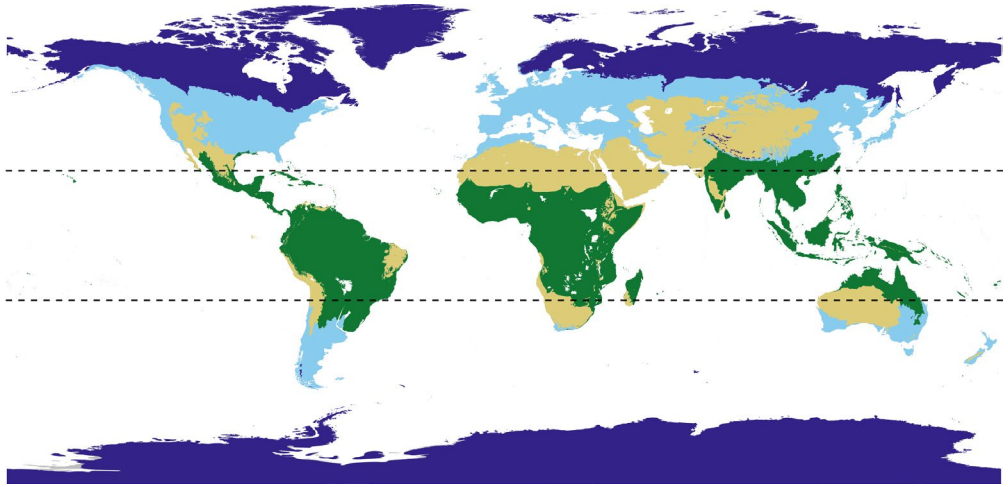
Ferns specimens generate a complex morphological landscape

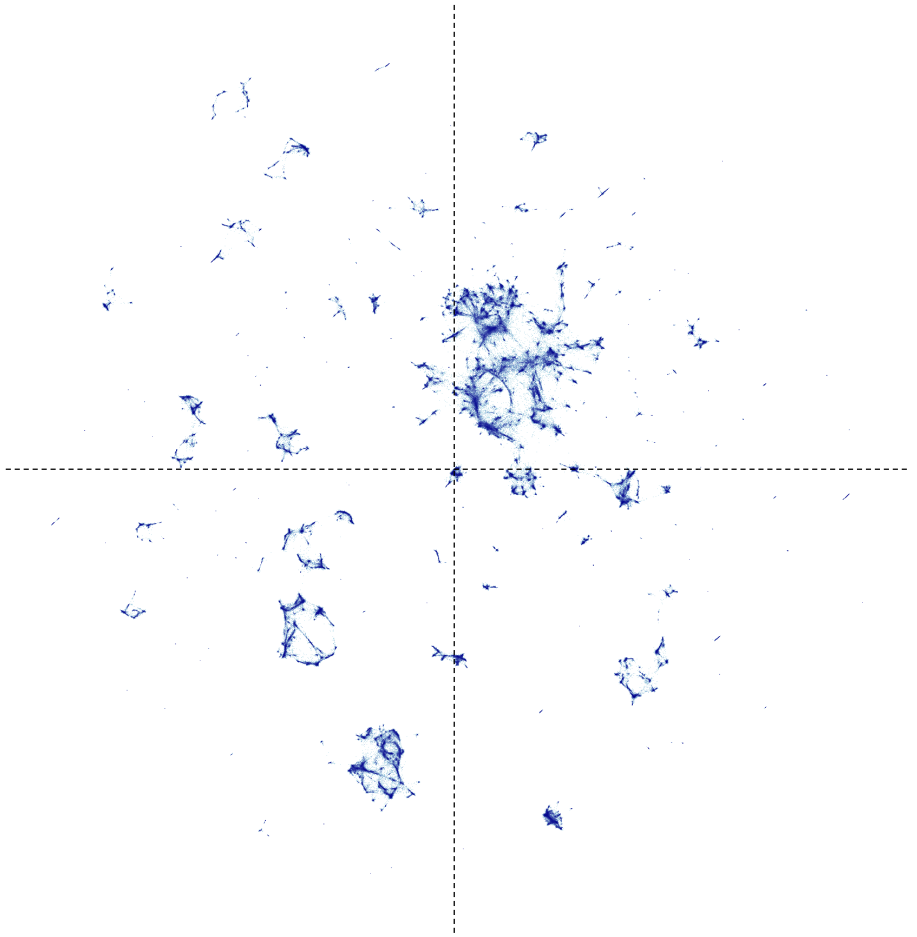


Ferns specimens generate a complex morphological landscape

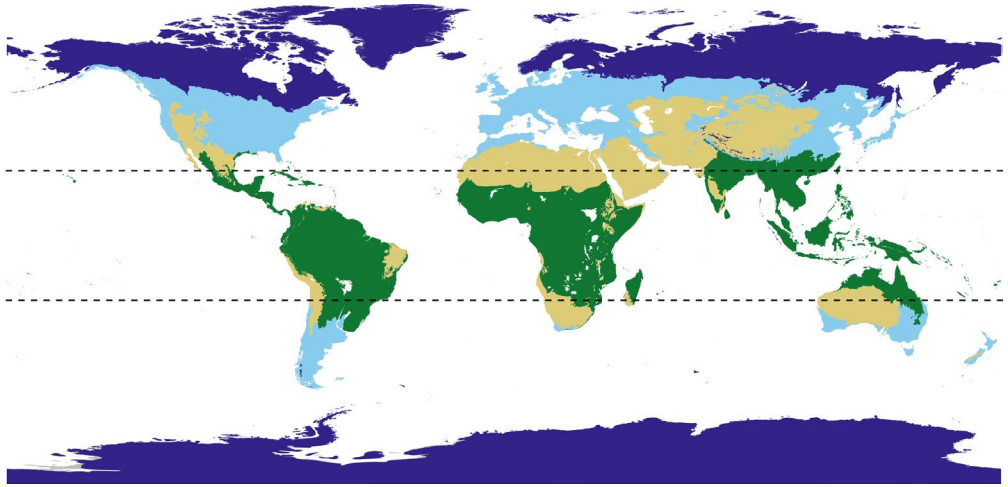


Ferns specimens generate a complex morphological landscape

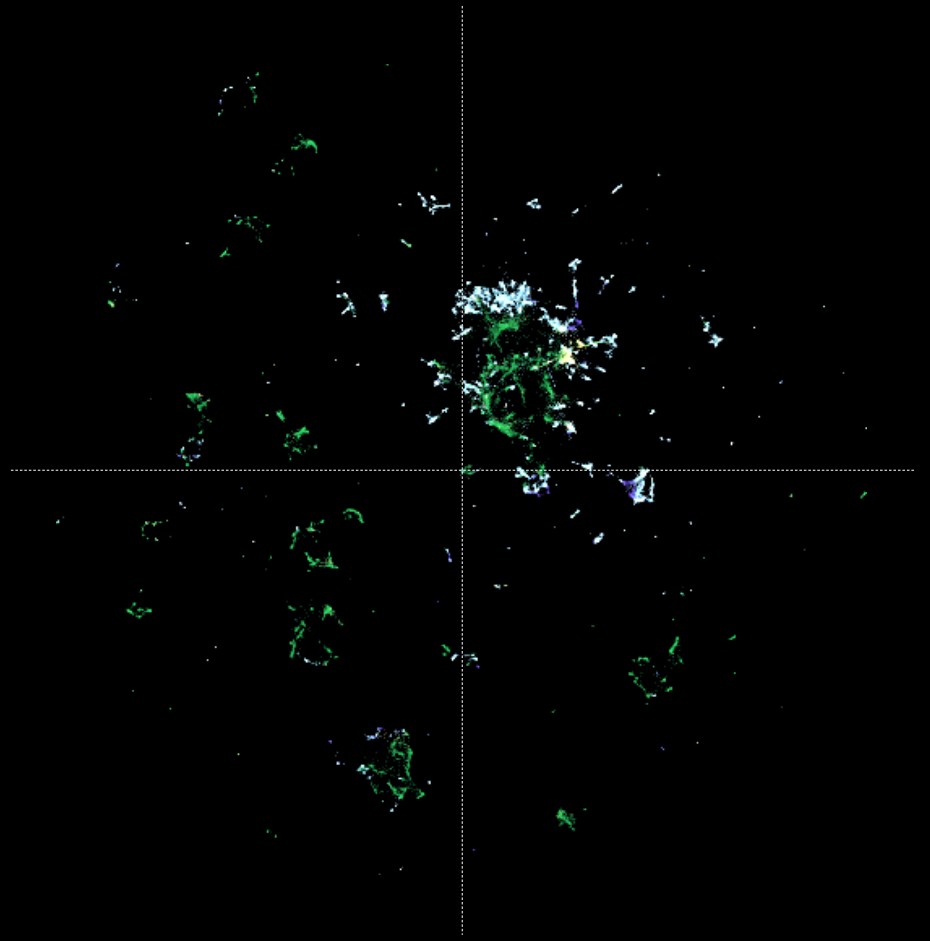


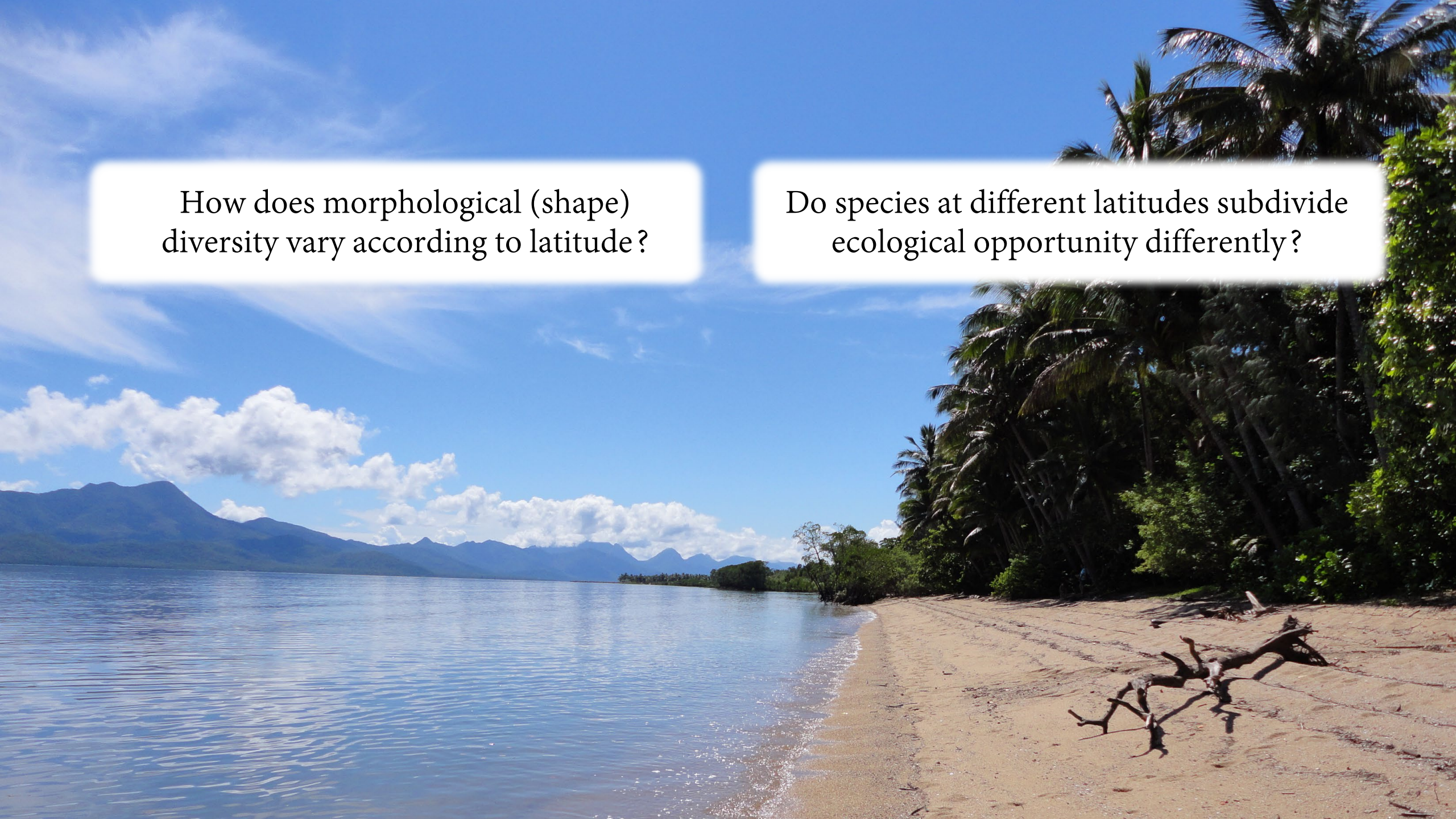


Ferns specimens generate a complex morphological landscape

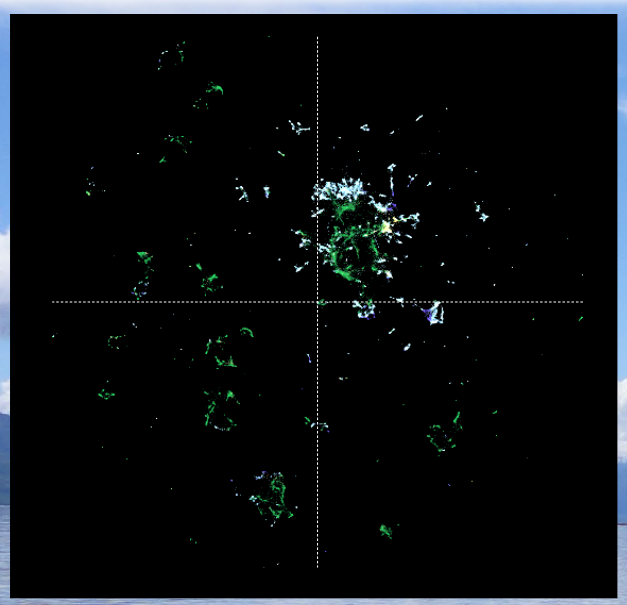


…with clusters of variation based on latitude and climate




How does morphological (shape) diversity vary according to latitude?
Do species at different latitudes subdivide ecological opportunity differently?

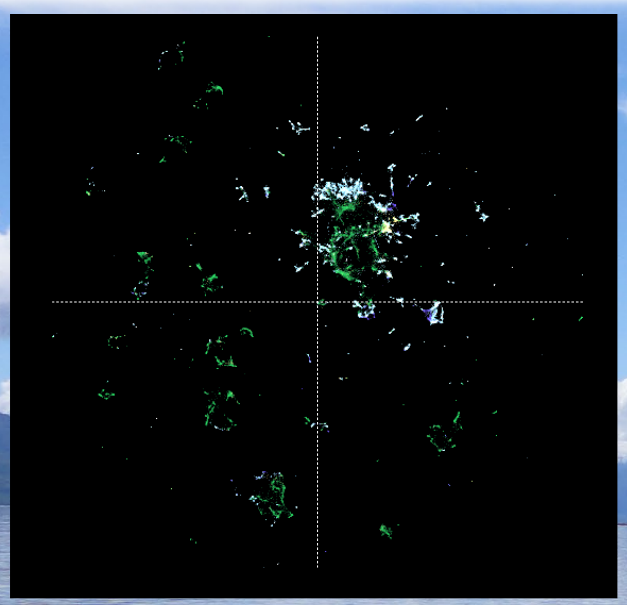
How does morphological (shape) diversity vary according to latitude?

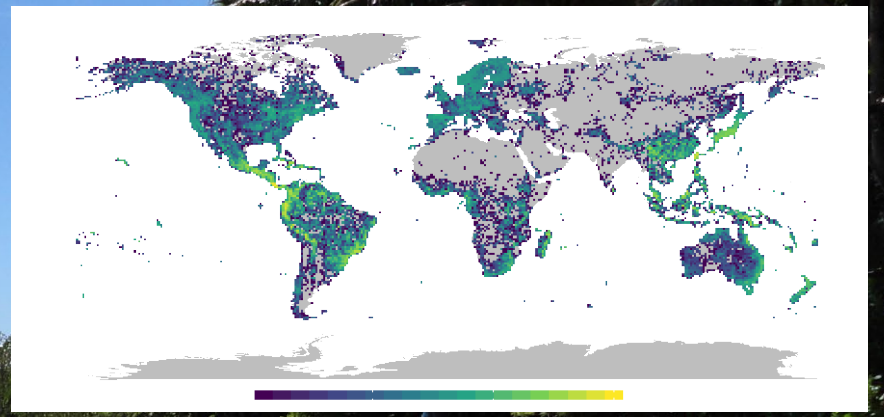


Do species at different latitudes subdivide ecological opportunity differently?

How does morphological (shape) diversity vary according to latitude?

Do species at different latitudes subdivide ecological opportunity differently?





Many contributors, many thanks

Rebecca Dikow
Mike Trizna
Paul Frandsen
Larry Dorr
Eric Schuettpelz

DATA SCIENCE LAB

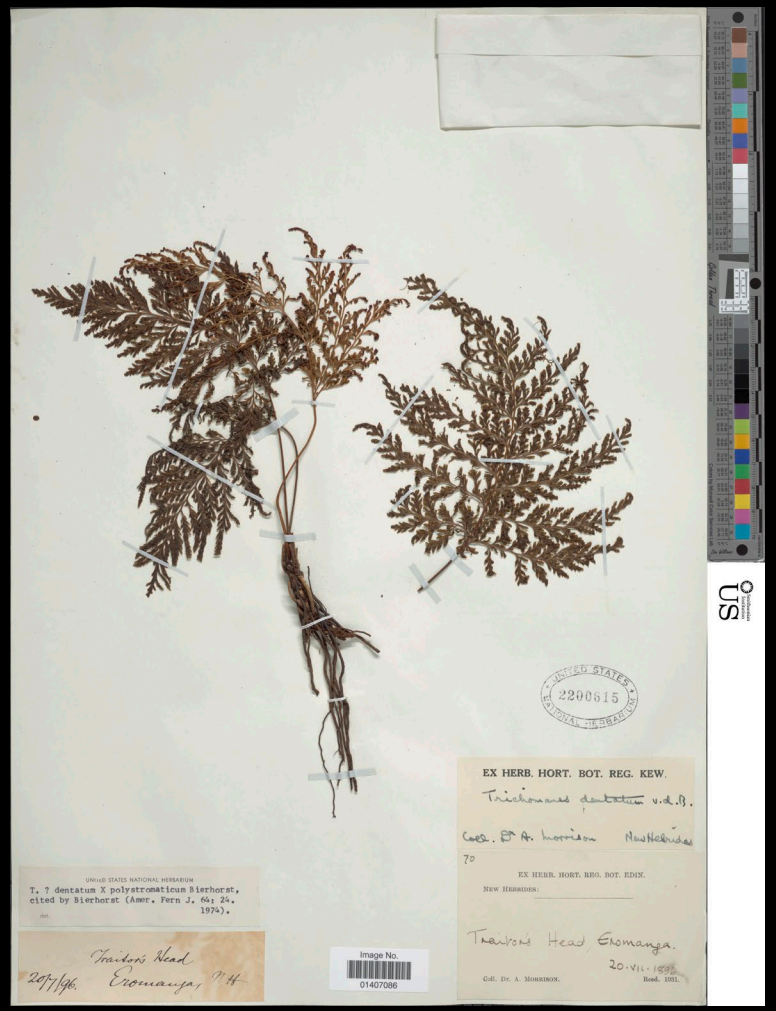
NATIONAL MUSEUM of NATURAL HISTORY

SMITHSONIAN INSTITUTION

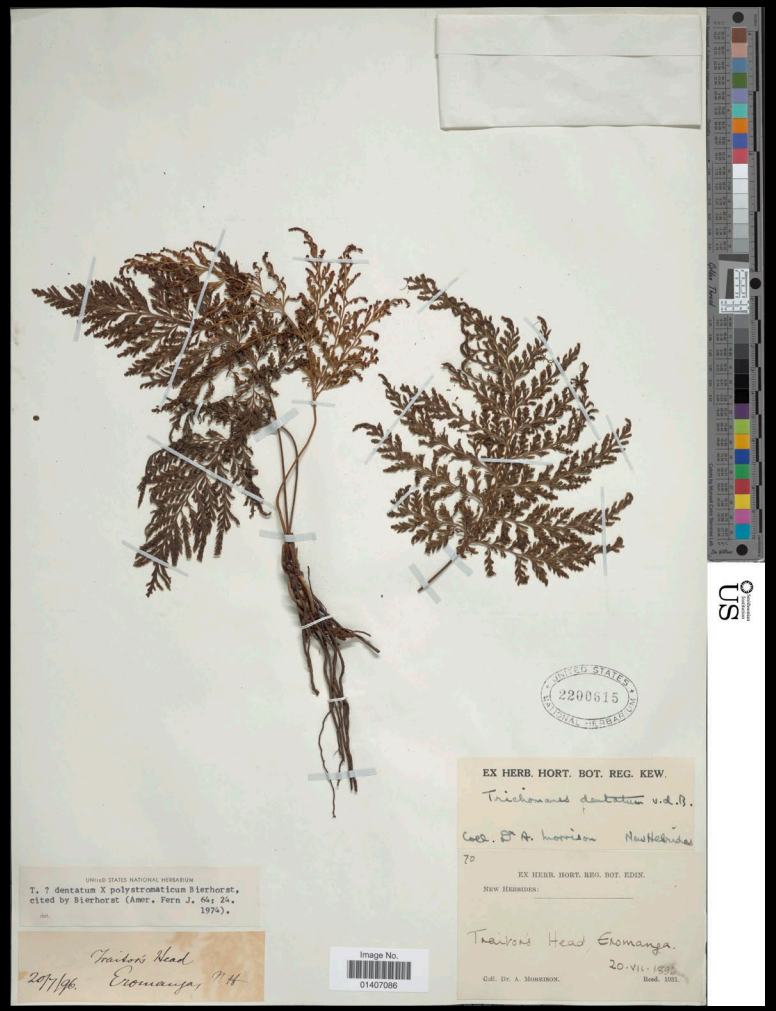
Manuela Dal Forno
Karen Golinski
Joel Nitta
Kathryn Picard
Mirian Tsuchiya

Smithsonian OCIO
NMNH Department of Botany

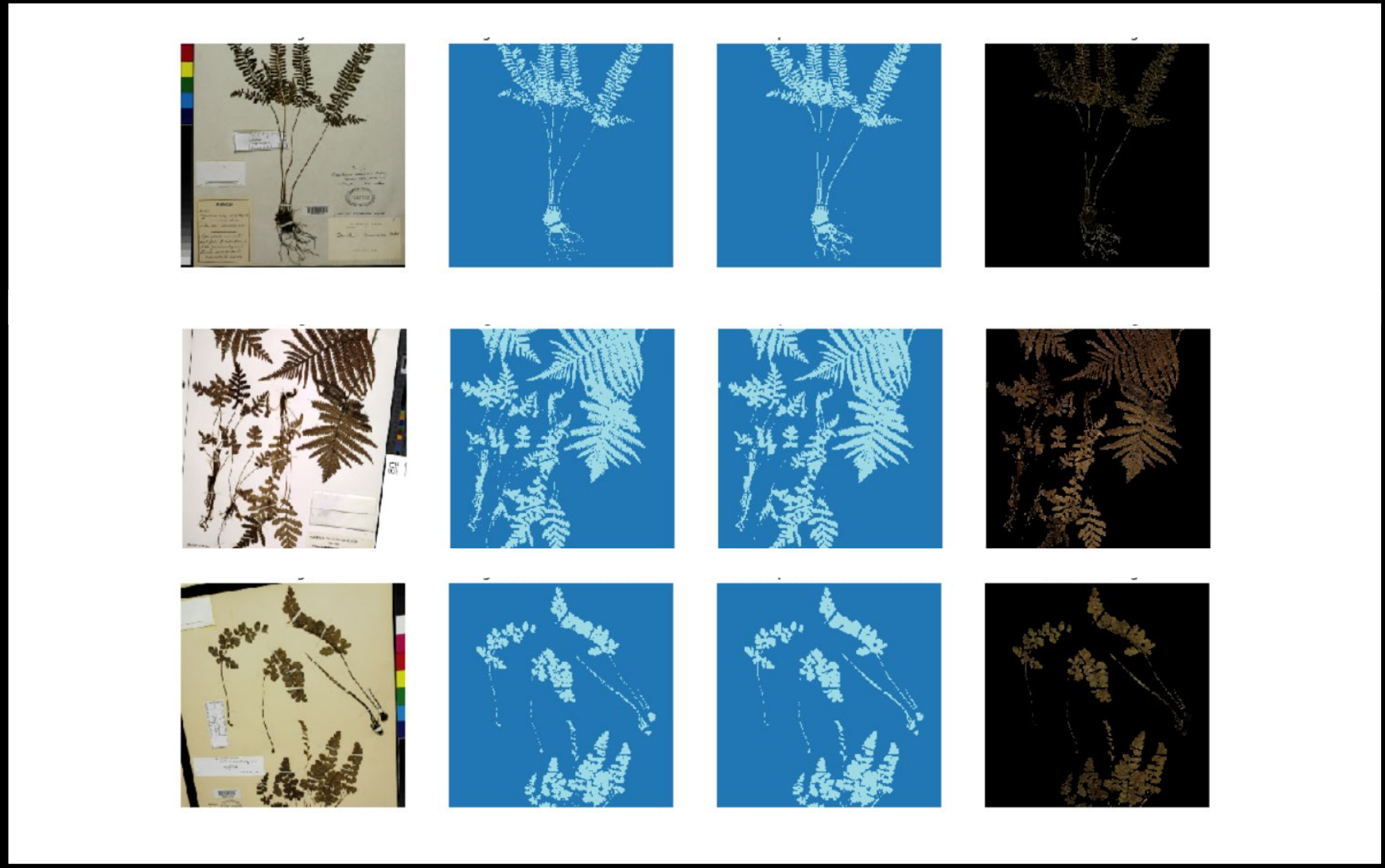
Extraneous objects could impact estimates of shape



So we are building masks to focus only on the plant in the image

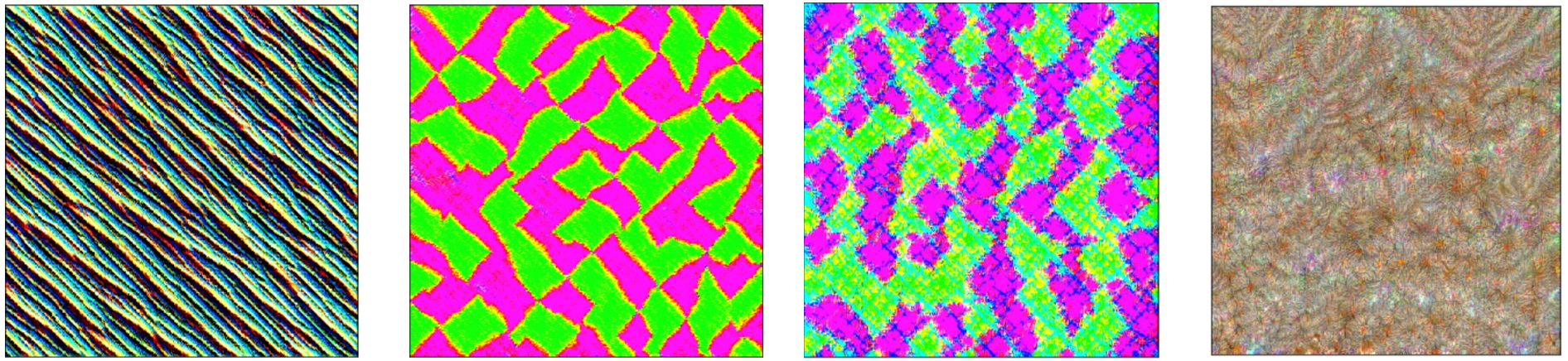


So we are building masks to focus only on the plant in the image



What are the convolutional filters looking at?

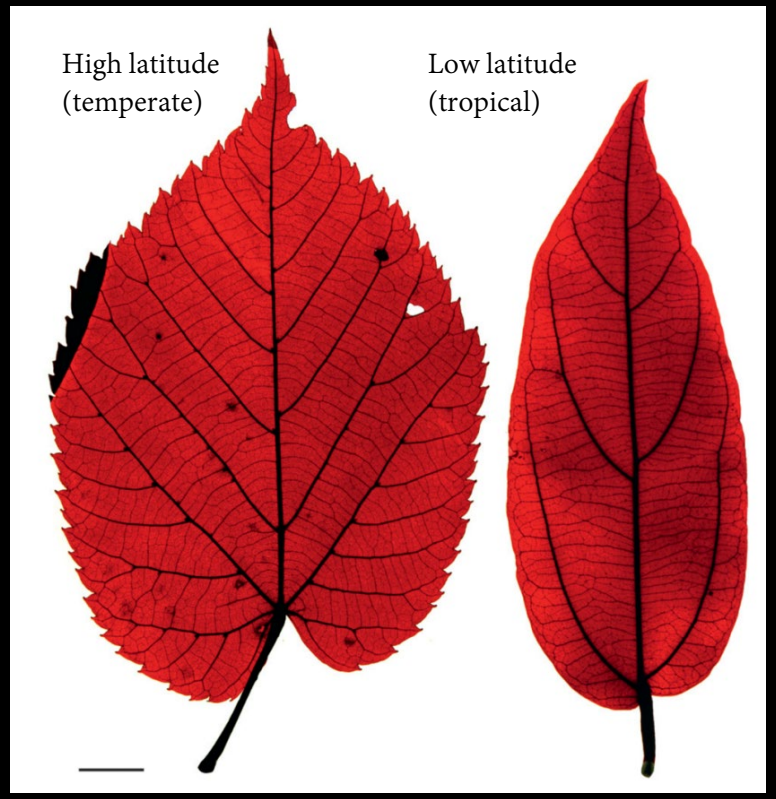
4 example "feature maps"



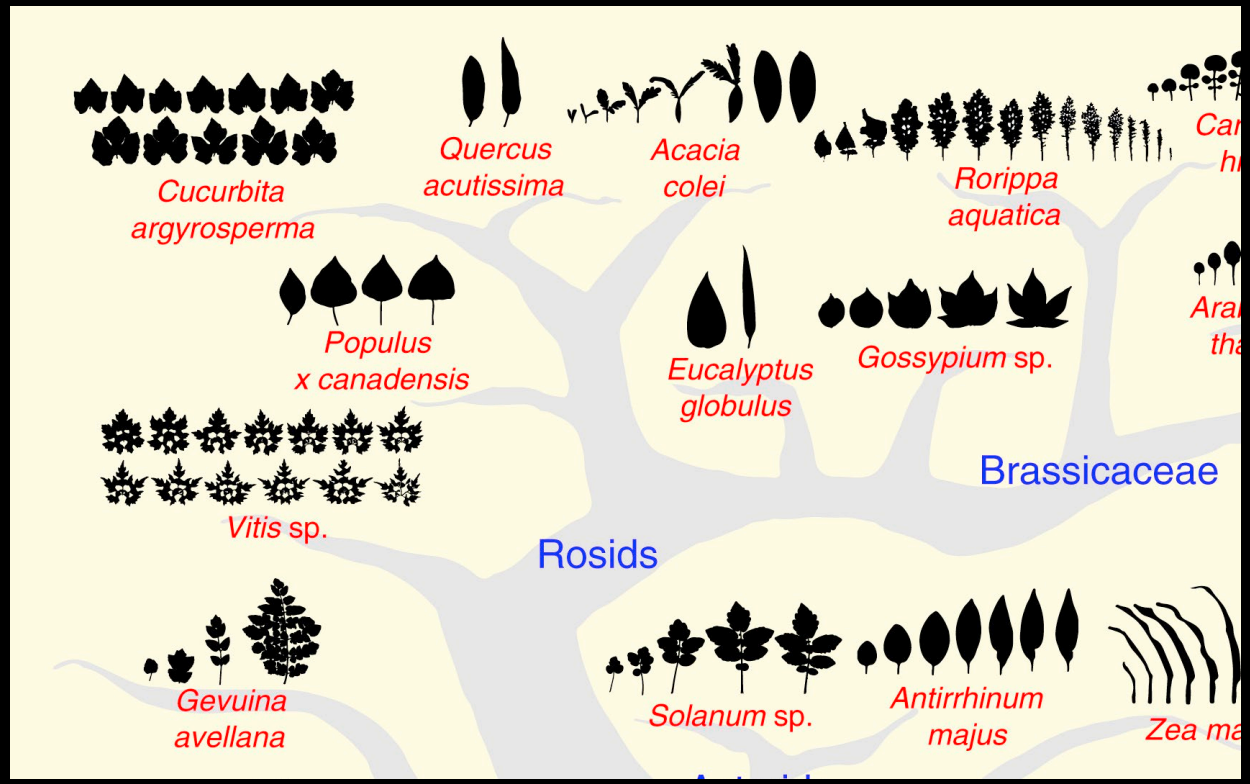
early layer filters → late layer filters

Plant leaf shapes are diverse and complex, with variations by latitude

Computer vision methods can help us quantify and summarize leaf shape



Givnish and Kriebel 2017



Chitwood and Sinha 2016