www.ala.org.au

ALA and ALA4R

January 2017

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Ben Raymond alaben@untan.gl
Dave Martin david.martin@csiro.au

ALA – *sharing biodiversity knowledge*

>$50 million investment

- $8.2M NCRIS (2007-2011)
- $30M SS EIF (2010-2012)
- $2.8 M CRIS (2013-2015)
- $5.7M NCRIS2013 (2013-2015)
- $4.6M NCRIS 2015

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Partners – founding & beyond

- founding partners and contributors primarily biological collections and museums
- increased contribution & use by citizen science, government, industry

National Research infrastructure

- NCRIS established in 2006
- Currently within Education Dept
- Related environmental infrastructures TERN (terrestrial ecological info/data capture) & IMOS (marine)

Open source & open access

- ALA driving cultural shift regarding open access to data
- ALA at forefront of accessibility to public sector information

A world-leading collaborative e-infrastructure integral to growing biodiversity knowledge

>57 million records
>1100 data sets
>410 spatial layers
>3.7 billion records downloaded
>3.5 thousand users/day

Data

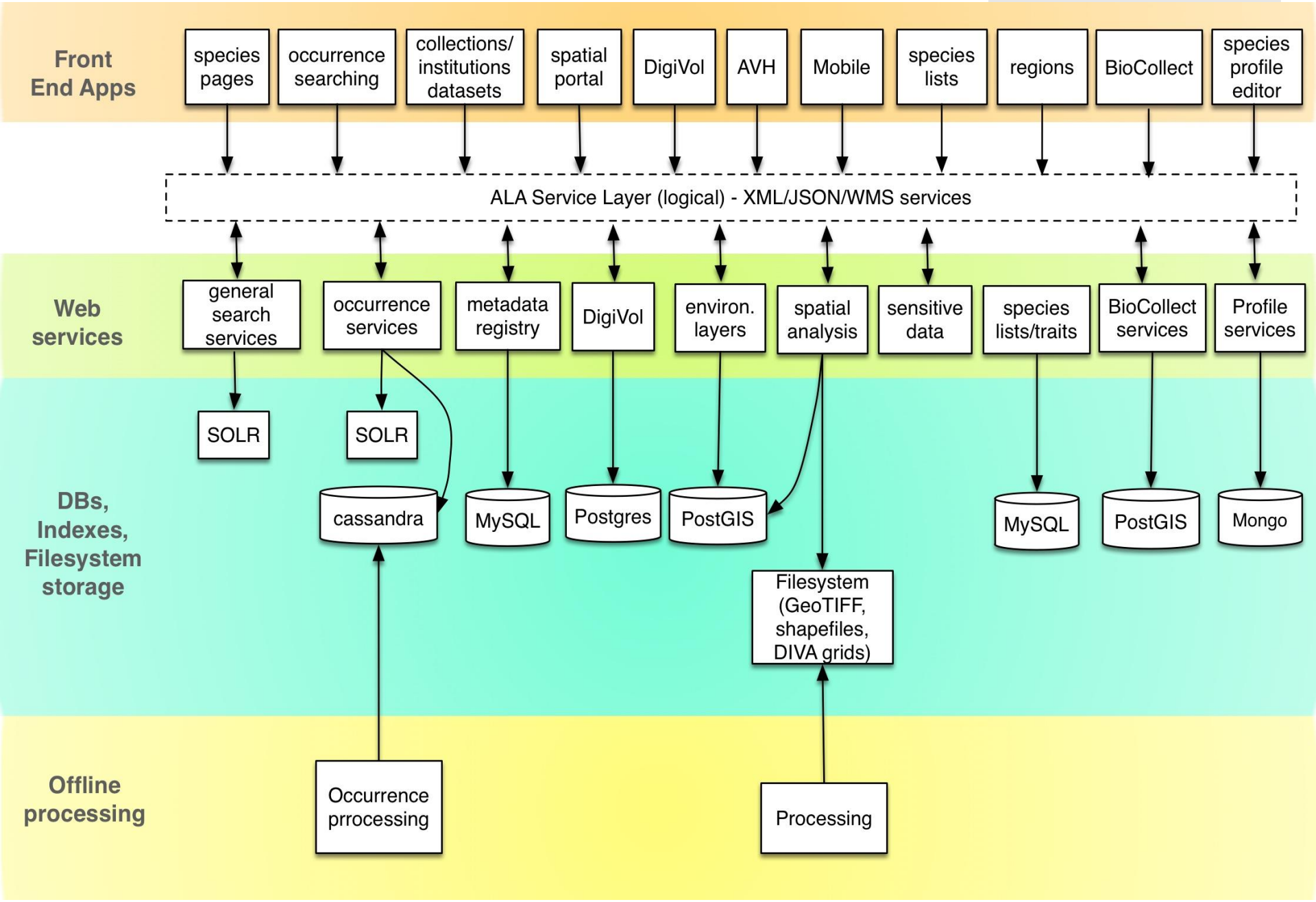
- specimens
- occurrence
- images, sounds
- literature
- sequences
- more coming……

System

- data capture & aggregation
- data management
- data discovery
- data visualisation
- data analysis & reporting

Production installations

- Spain
- France
- ICMBio Brazil
- Costa Rica
- UK (Scotland)
- Argentina
- Portugal
- Sweden
- Australia


Front End Apps
species pages
occurrence searching
collections/ institutions datasets
spatial portal
DigiVol
AVH
Mobile
species lists
regions
BioCollect
species profile editor
ALA Service Layer (logical) - XML/JSON/WMS services
Web services
general search services
occurrence services
metadata registry
DigiVol
environ. layers
spatial analysis
sensitive data
species lists/traits
BioCollect services
Profile services
DBs, Indexes, Filesystem storage
SOLR
SOLR
cassandra
MySQL
Postgres
PostGIS
MySQL
PostGIS
Mongo
Filesystem (GeoTIFF, shapefiles, DIVA grids)
Offline processing
Occurrence prrocessing
Processing

ALA4R

- motivation:
  - easy R access to ALA data and resources
  - support data-driven biodiversity science
  - also use compute-at-node where sensible
- subset of the full ALA API (http://api.ala.org.au/)
- first ALA4R release (github) 2014, CRAN release 2016
- reconfigurable: use with other national installations using ALA infrastructure

ALA4R

- core functionality
  - name searching
    - names lookup, partial matching, fulltext searching
  - species information
    - taxonomy, species profiles, images
  - occurrences
    - with environmental/contextual data
    - "offline" mode for large downloads
    - species lists
    - species by site matrices
  - environmental
    - environmental/contextual data

ALA4R examples

- some examples: see R script
- also

```
vignette("ALA4R")
```

Other examples: soundscape

- Ebbe Nielsen challenge round 1 winner, 2015
- https://peterneish.github.io/gbif-soundscape/
- overview:
  - get species list in region of interest
  - download sound recordings and images of those species
  - play back to produce a soundscape of the region
- R source

Other examples: Eucalypt height

- More info and R source
- ALA4R::species_info returns an information profile that includes descriptive text
- for plants, e.g.
  - "tree up to 90 metres”
  - "grows to 15 metres”
  - "may reach 30 - 40 metres in height”
  - "growing to a maximum height of 4 metres”

- can we use this as a source of trait (height) information?

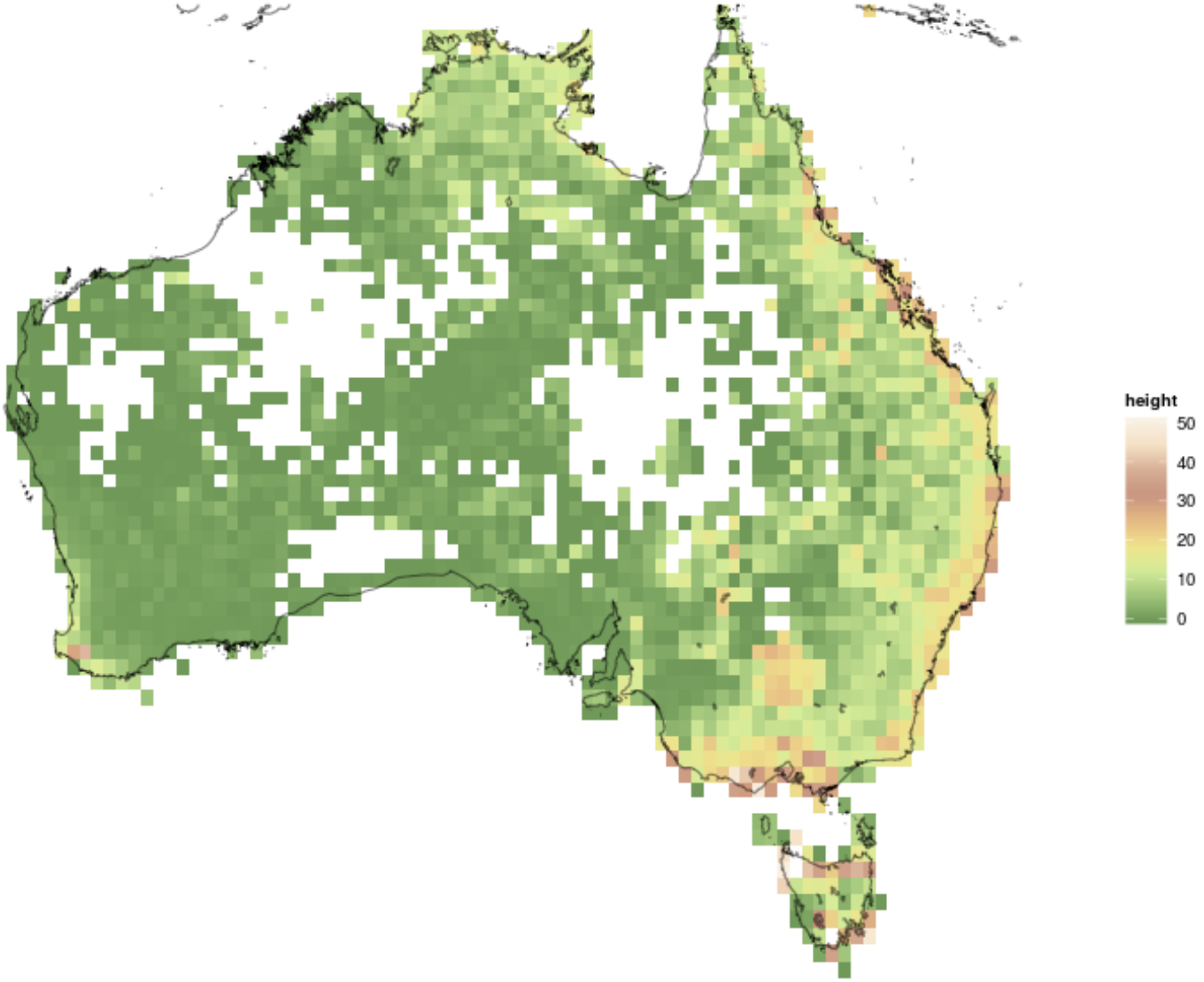
Other examples: Eucalypt height

- download gridded eucalypt data:

```
ss <- sites_by_species("genus:Eucalyptus", …)
```

- for each species, run the `species_info` function and extract height information
- given heights of species, and species within grid cells, map height across grid cells

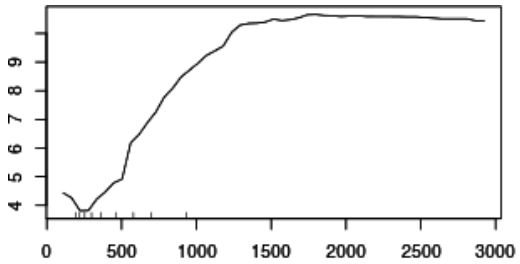
Other examples: Eucalypt height

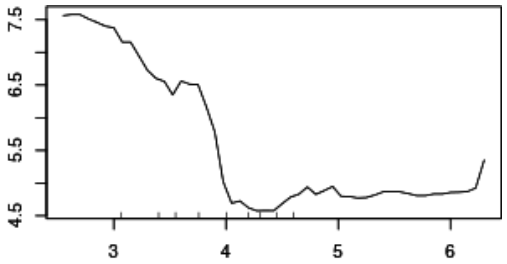


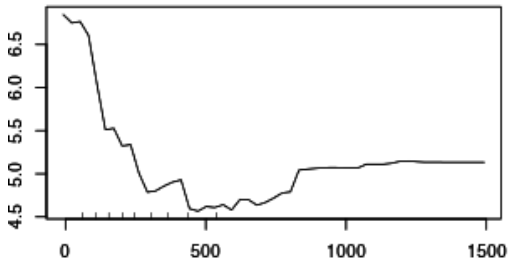
Other examples: Eucalypt height

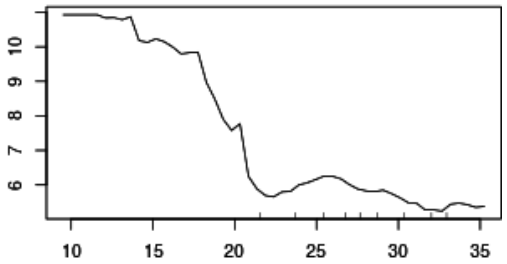
- using a random forest, model height as function of environmental variables
  - precipitation
  - solar radiation
  - elevation
  - maximum temperauture
  - wind speed

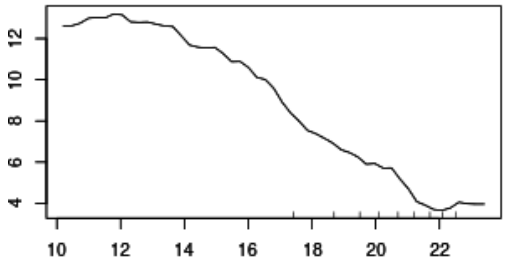
Other examples: Eucalypt height



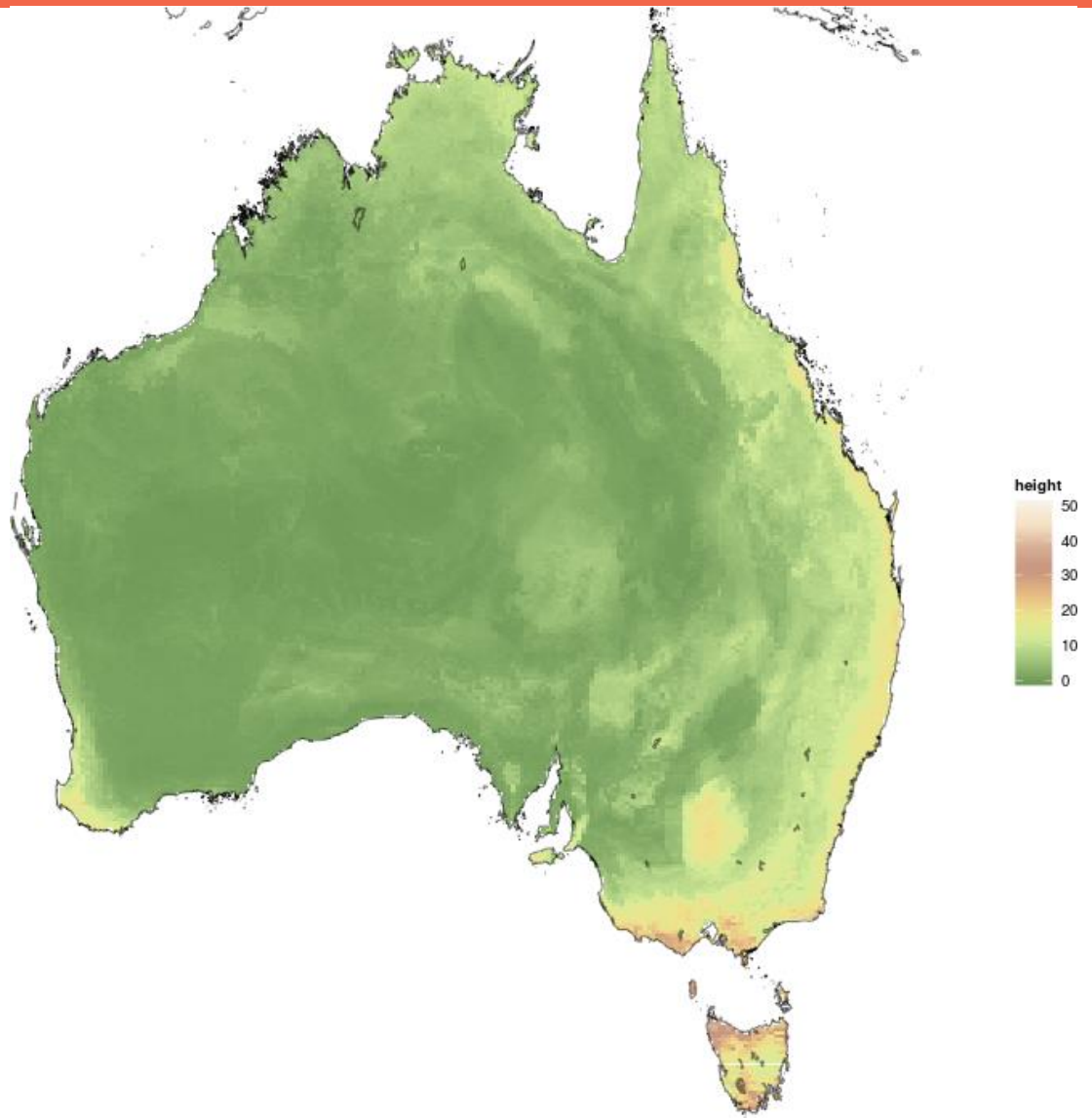








Other examples: Eucalypt height

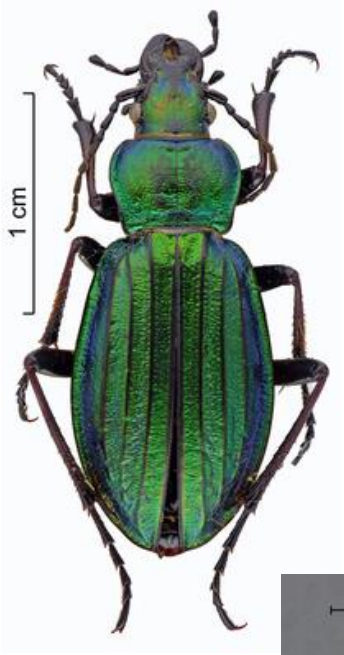


Colours from images

- More info including R source

- many sources of species photographs
- can we extract colour information from photos?
  - evolution of colour traits
  - colour variation with taxonomy
  - link with occurrence records to reconstruct spatial variations in colour
  - environmental drivers of colour variation

Typical beetle images

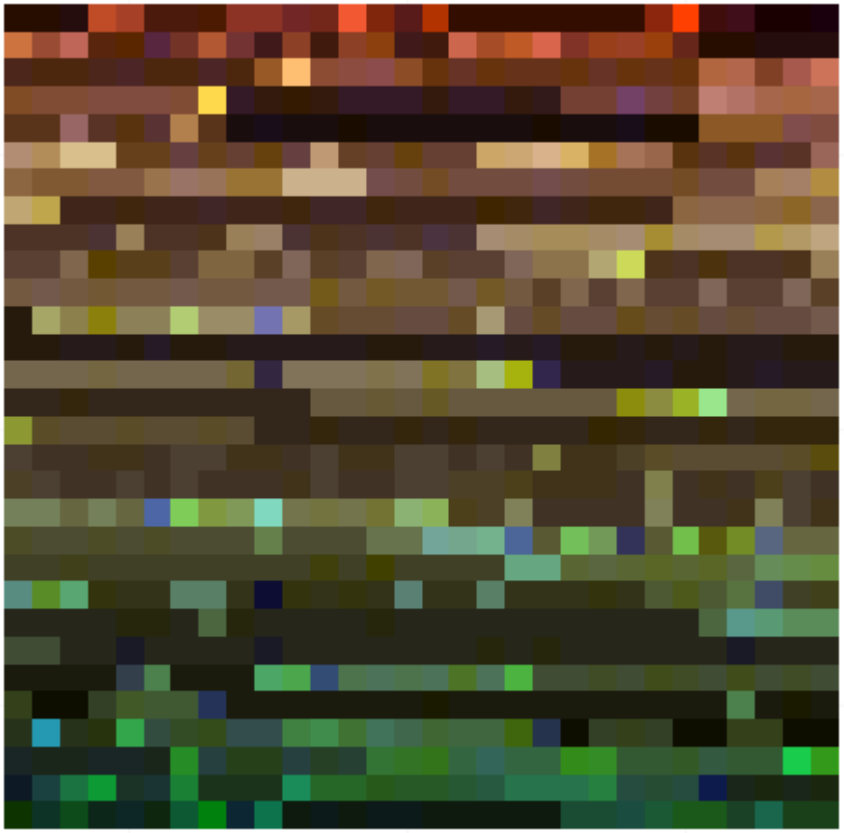


Colours from images: processing

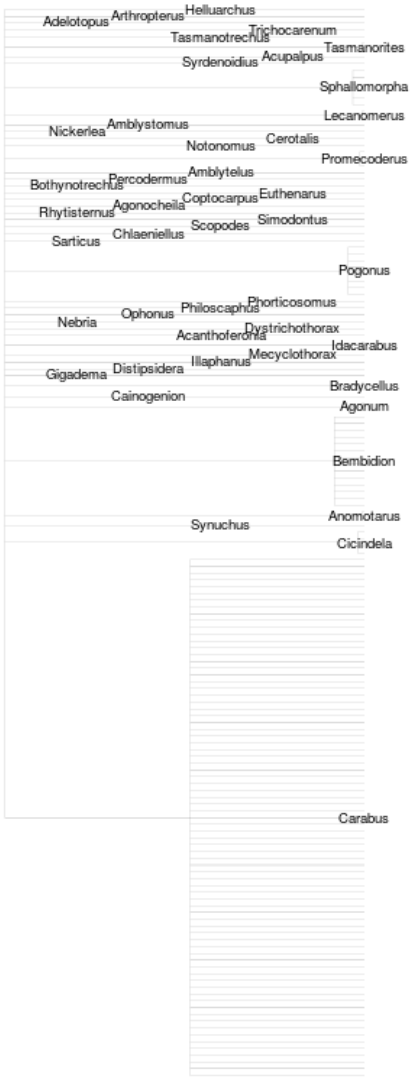
- list of species in the family Carabidae (ground beetles)
- find up to 10 images per species
  - look for `PRESERVED_SPECIMEN` as the `basisOfRecord` attribute
- manual filtering of unsuitable images
  - for simplicity, use only images with near-white backgrounds
  - discard near-white colours

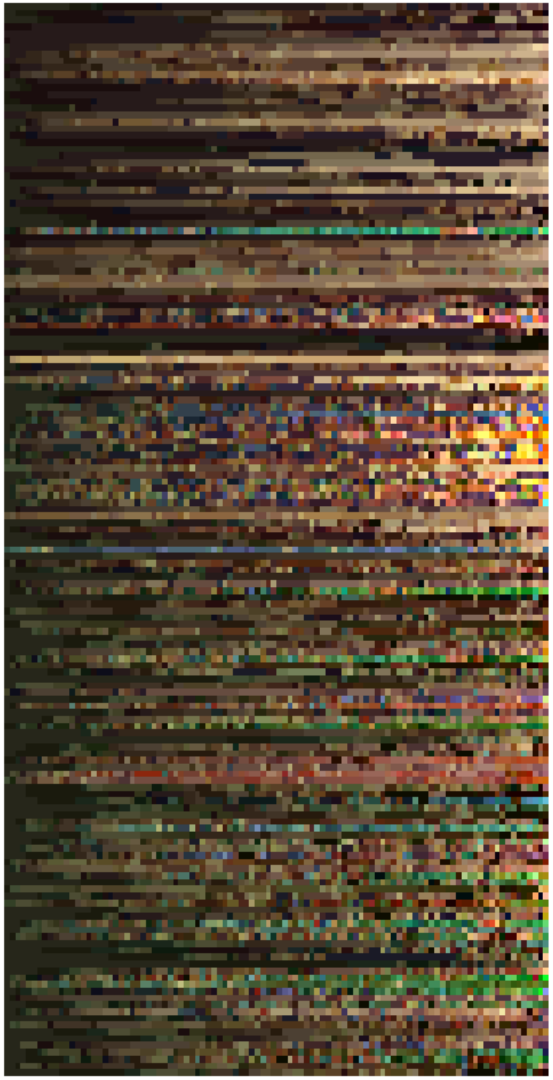
Colours from images: palette

- build a global colour palette (i.e. a matrix of all of the colours present in all images)

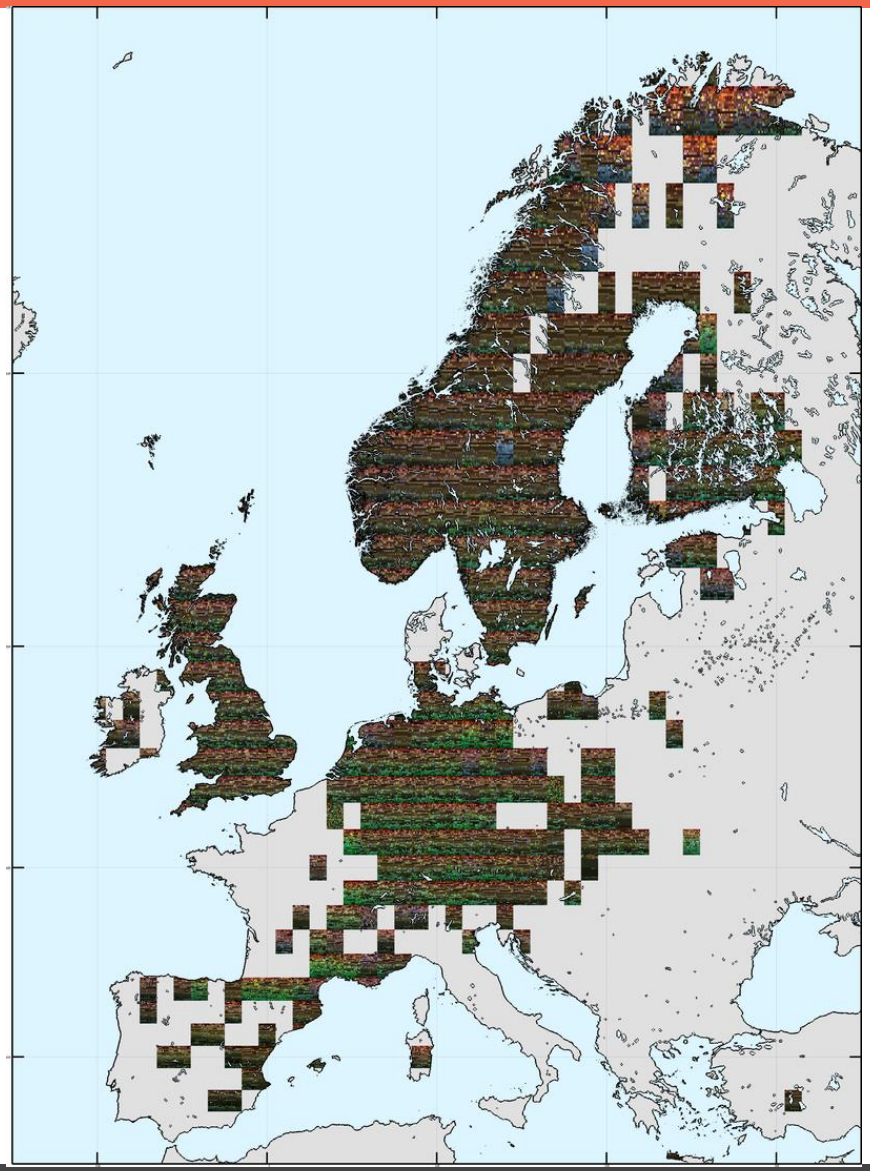


Colours by taxonomy



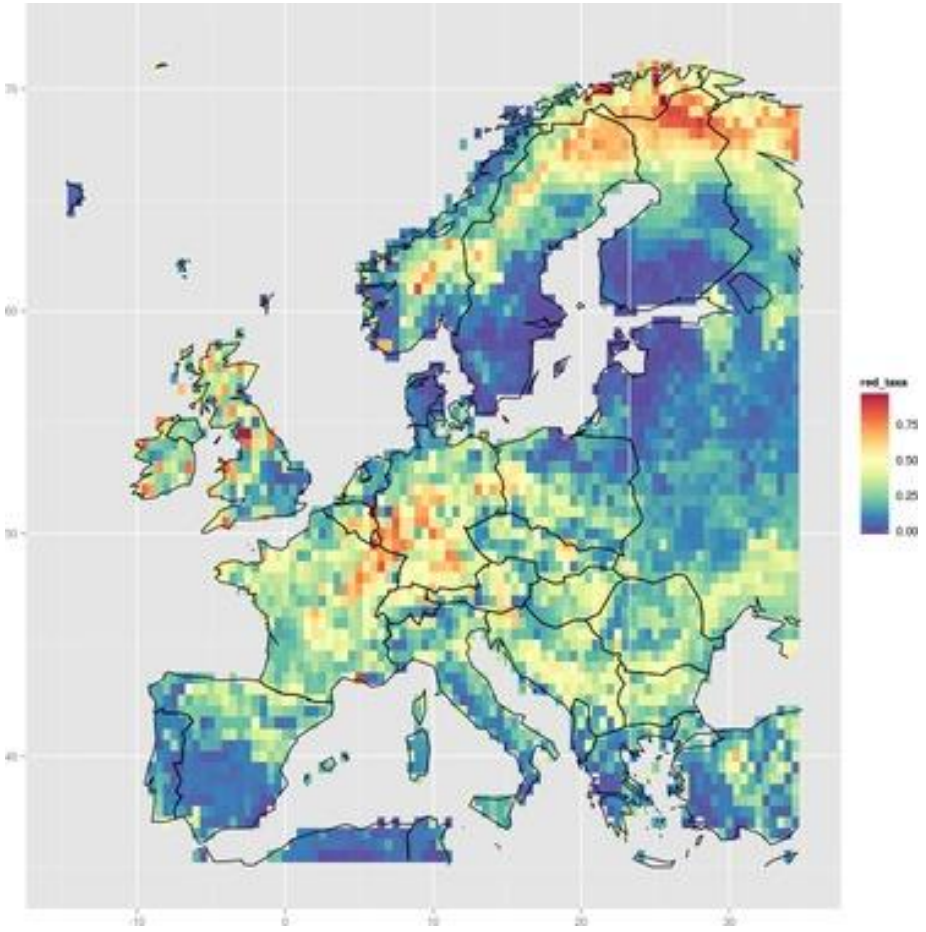


Colours across space



Colours as function of environment

- probability of red colouration



Other image sources

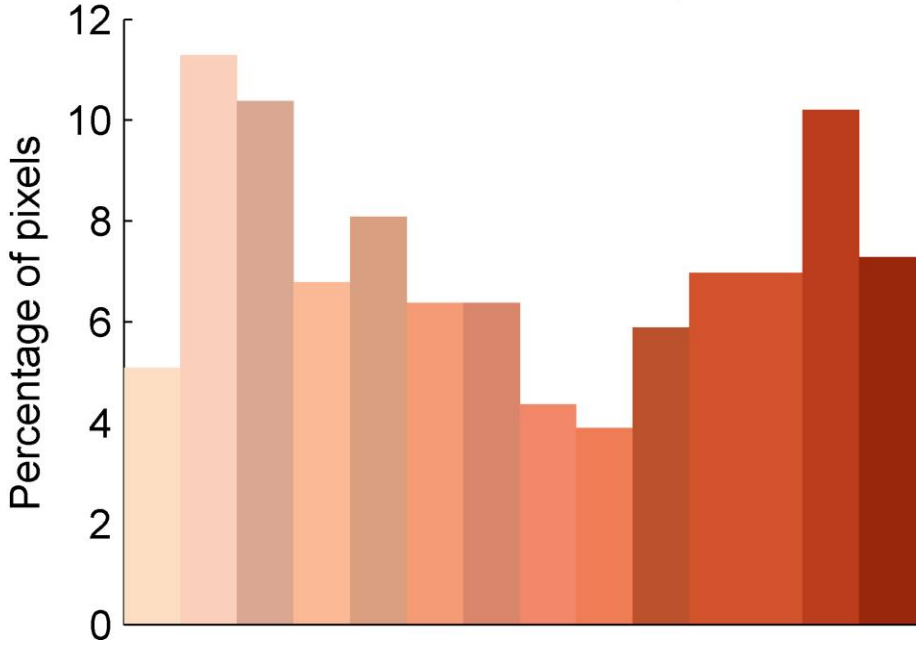
- colours of Grevillea flowers
- images from ALA/other species information pages
- typically photos of live plants, not specimens

Processing

- for each photo, extract representative flower colour palette

Grevillea aneura

Extracted flower colour palette

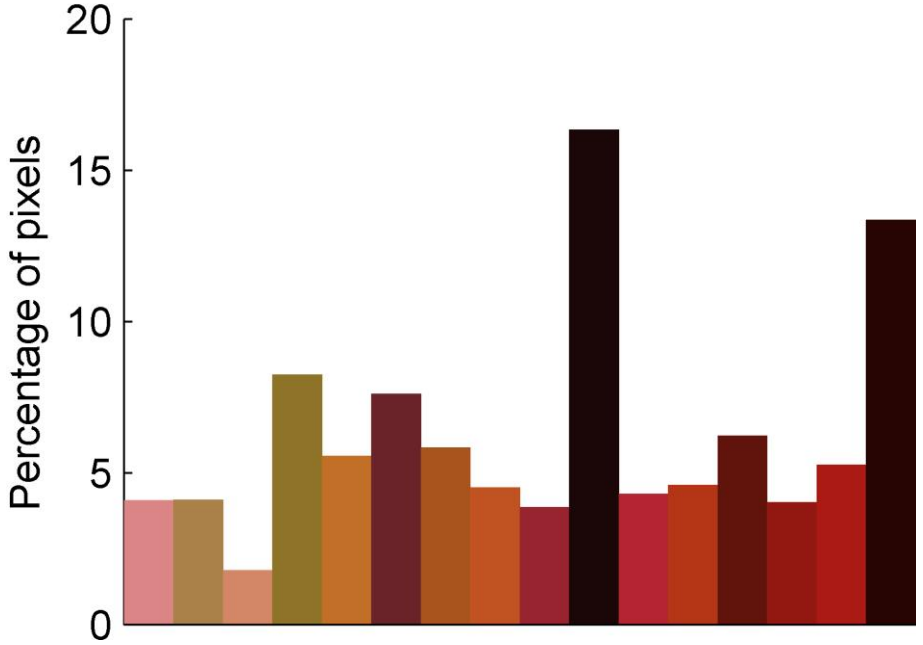


Processing

- for each photo, extract representative flower colour palette

Grevillea calliantha

Extracted flower colour palette

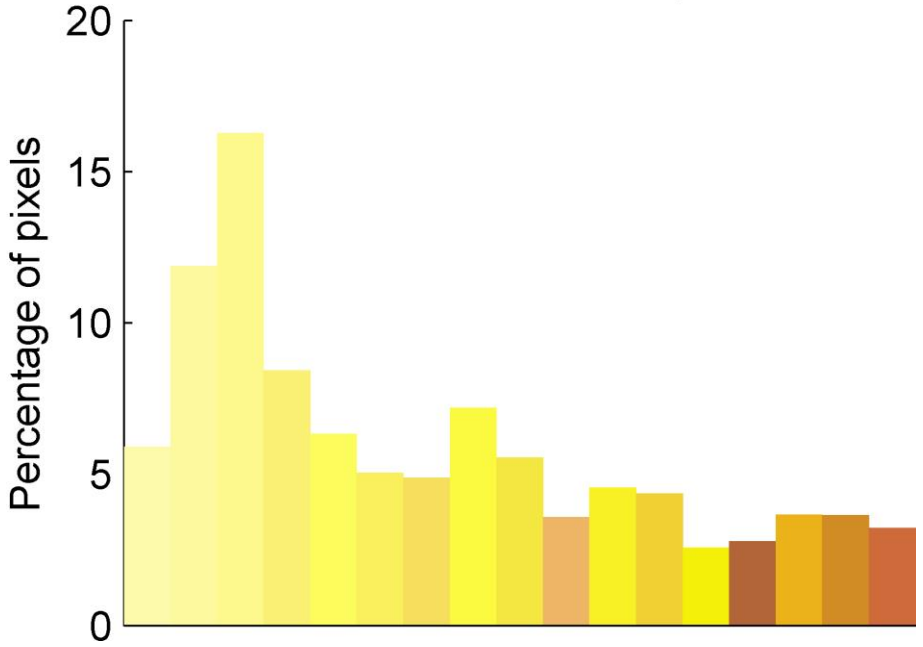


Processing

- for each photo, extract representative flower colour palette

Grevillea tenuiloba

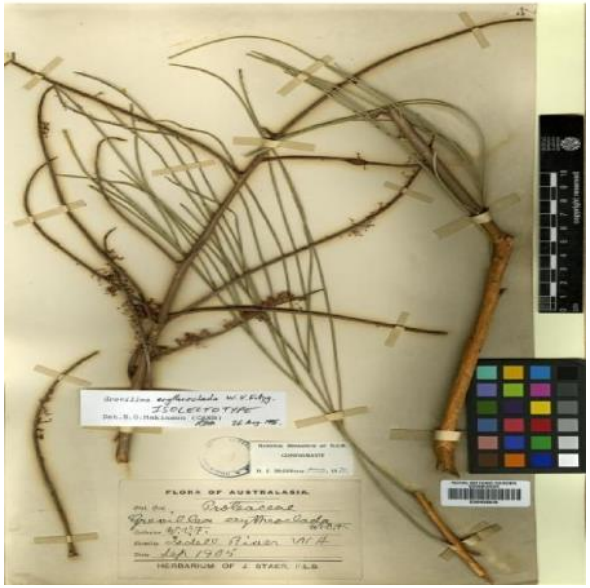
Extracted flower colour palette



Processing

- for each photo, extract representative flower colour palette

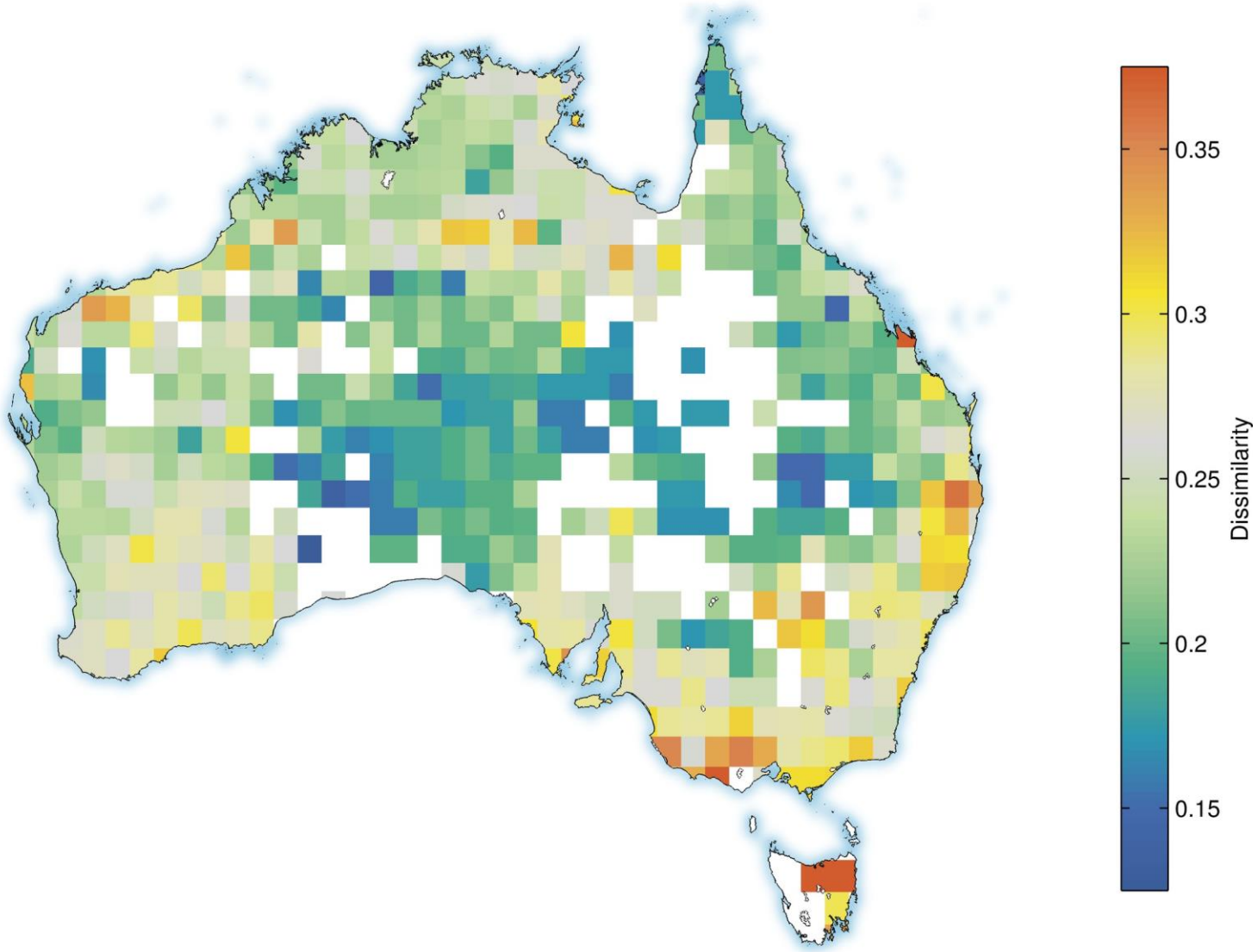
Grevillea erythroclada



Grevillea striata

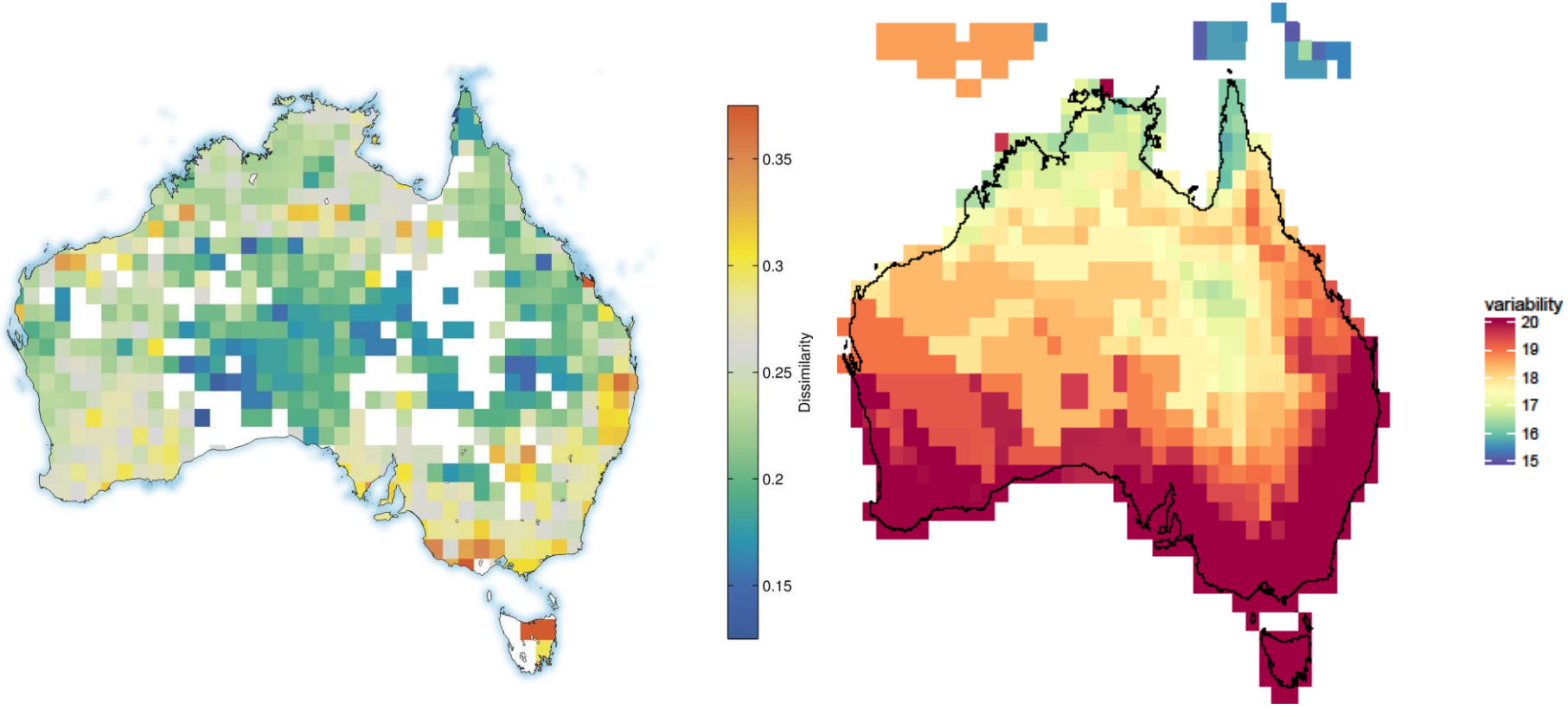
Results: Grevillea flowers

Within-species colour variability



Large-scale processes?

- within-species colour variability in Grevillea (left) and honeyeaters (right)



Colour processing

- manual processing of images is tedious
- deep learning methods for image processing
- open-source network for image segmentation: https://github.com/torrvision/crfasrnn

Colour processing

- not fully automatic, but promising …

And we're done!

Thanks!

alaben@untan.gl