Using DNA barcoding and machine learning to investigate macroecological relationships between plants and arthropods across South Africa.

R. D. Stewart[1,2], K. Perez[3], P. D. N. Hebert[3], N. Bard[5], T. Sethusa[4], M. van der Bank[2], T. J. Davies[2,5]

[1] Sol Plaatje University | [2] African Centre for DNA Barcoding (ACDB), University of Johannesburg | [3] Centre for Biodiversity Genomics | [4] South African National Biodiversity Institute (SANBI) | [5] University of British Columbia

Relevance

- What is climate change doing?
  - Baseline data
  - Is it altering ecological relationships?
  - timing shift and range shifting
- Challenges
  - There are millions of species!
  - Lack base line data on insects
  - Lack baseline data with the flowering times

Insect Diversity Overview

- Described insect diversity.
  - About 1 million.
  - Estimates range between 8-100 million.

    (Mora et al., 2011)

- Challenges with describing insect biodiversity.
  - Time consuming | taxonomic expertise | funding.

Importance of Insects

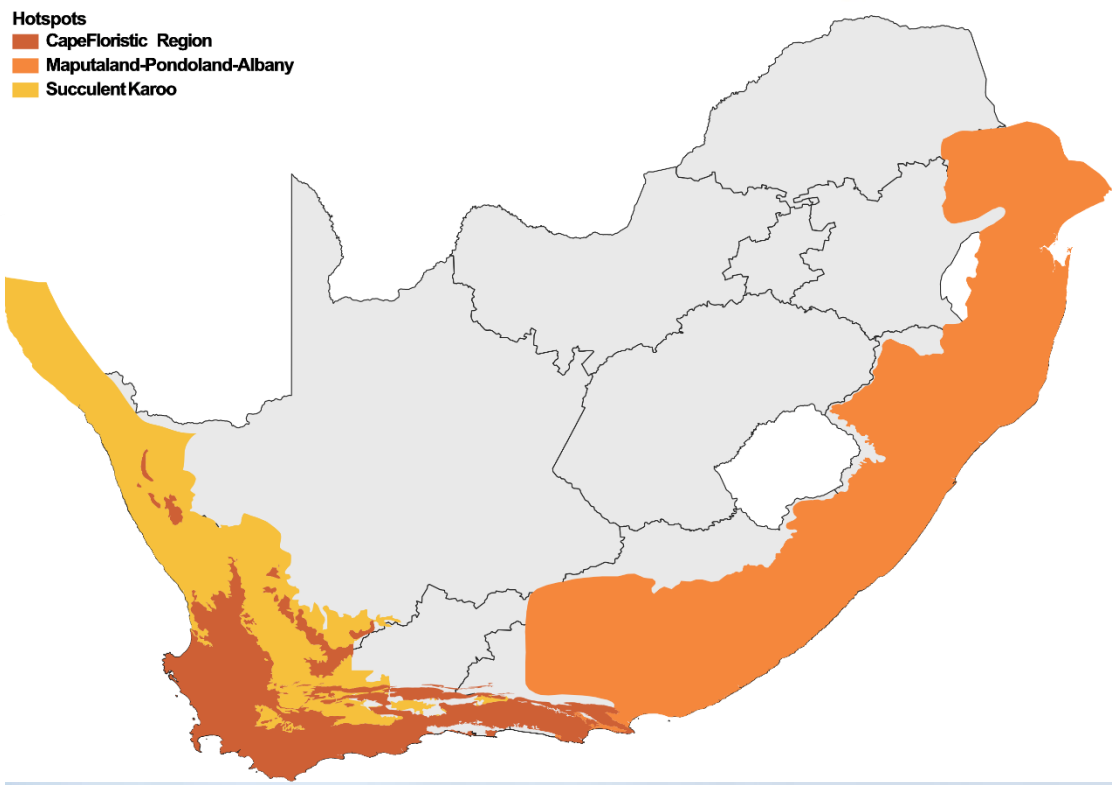
- Important ecosystem roles
  - The Good – Pollination.
  - The Bad – Plant pests.
  - The Ugly – Vectors for pathogens (Polyphagous Shot Hole Borer beetle – PSHB).

- Key ecological indicators for climate change
  - Species richness | abundance | predator-prey ratios | sensitive species | specialists vs. generalists.

- Insect apocalypse (Hallmann et al., 2017)

South African Biodiversity

- Mega biodiversity, estimated
  - 21,000 plant species.
  - 44,000 insect species in Southern Africa.



- Three biodiversity hotspots
  - Cape Floristic Region – 6,210 endemic plants.
  - Succulent Karoo – 2,542 endemic plants.
  - Maputaland-Pondoland-Albany – 1,900 endemic plants.

Insects in South Africa

The last study done on insects was done in southern Africa in:

- 2016 – Scholtz
- 1995 – Scholtz & Chown
- 1985 – Scholtz & Holm

Nearly 40 years!

44,000 species, 7750 genera, 569 families, and 25 orders

New ways to update this baseline data

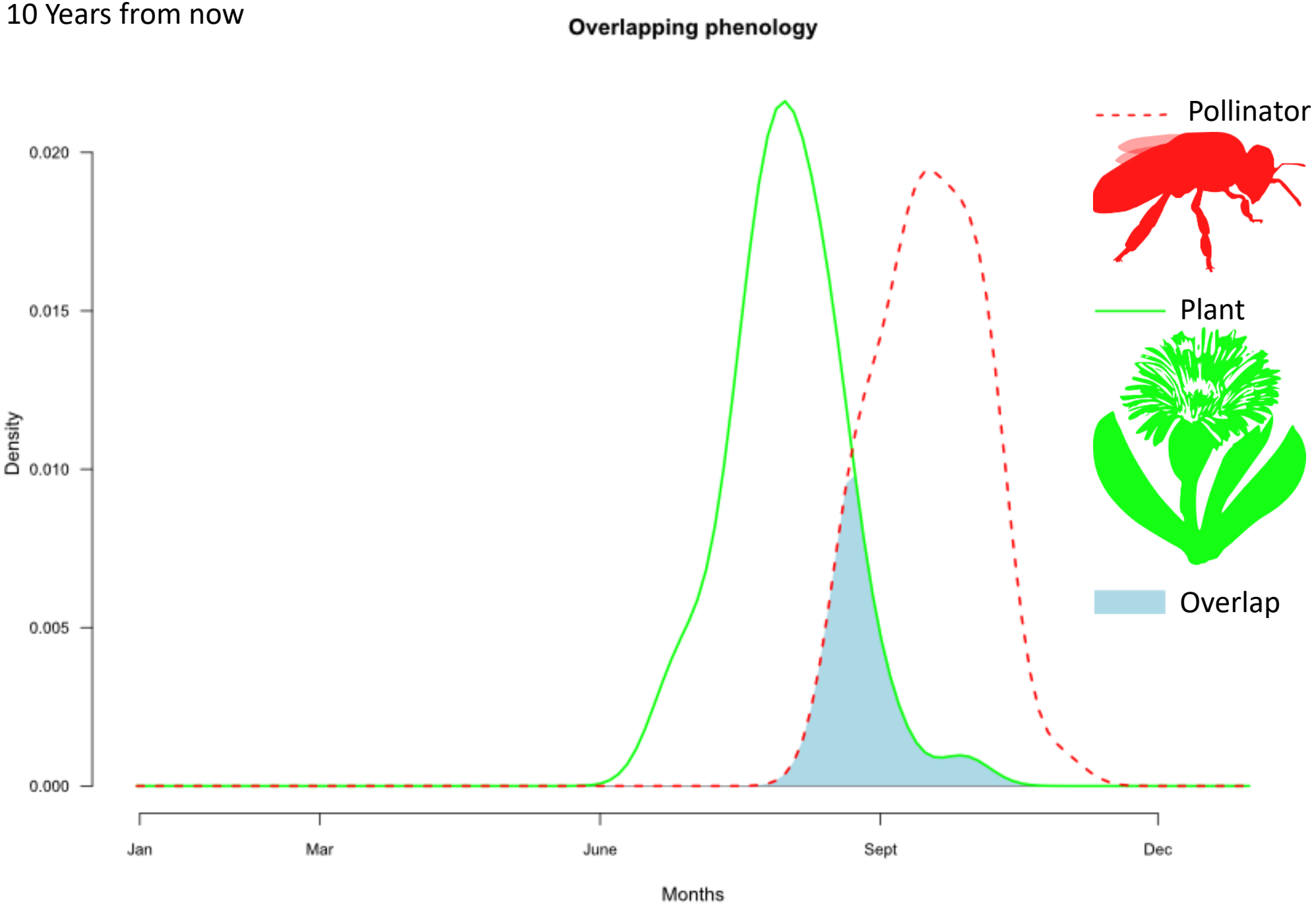
Flowering Phenology

~Represents the timing of biological events.

- Flowering time
  - Crucial part of a plant's life cycles
- UN Sustainable Development Goals
- Fingerprint to climate change (Fei et al., 2017)
- Ecological interactions
  - Pollination
  - Insect Armageddon (Hallmann et al., 2017)


SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS

10 Years from now



Insects in South Africa

Prosoeca marinusi

Babiana framesii

J. van Rooyen

Phenology

~Represents the timing of biological events.

- Flowering time
  - Crucial part of a plant's life cycles
- UN Sustainable Development Goals
- Fingerprint to climate change (Fei et al., 2017)
- Ecological interactions
  - Pollination
  - Insect Armageddon (Hallmann et al., 2017)

Plants and Insects

- Plant community defines the niche space for insects – expect insect species richness to reflect the plant species richness.

- Large scale plant diversity is well-studied
  - Climate | soil properties | species interactions.

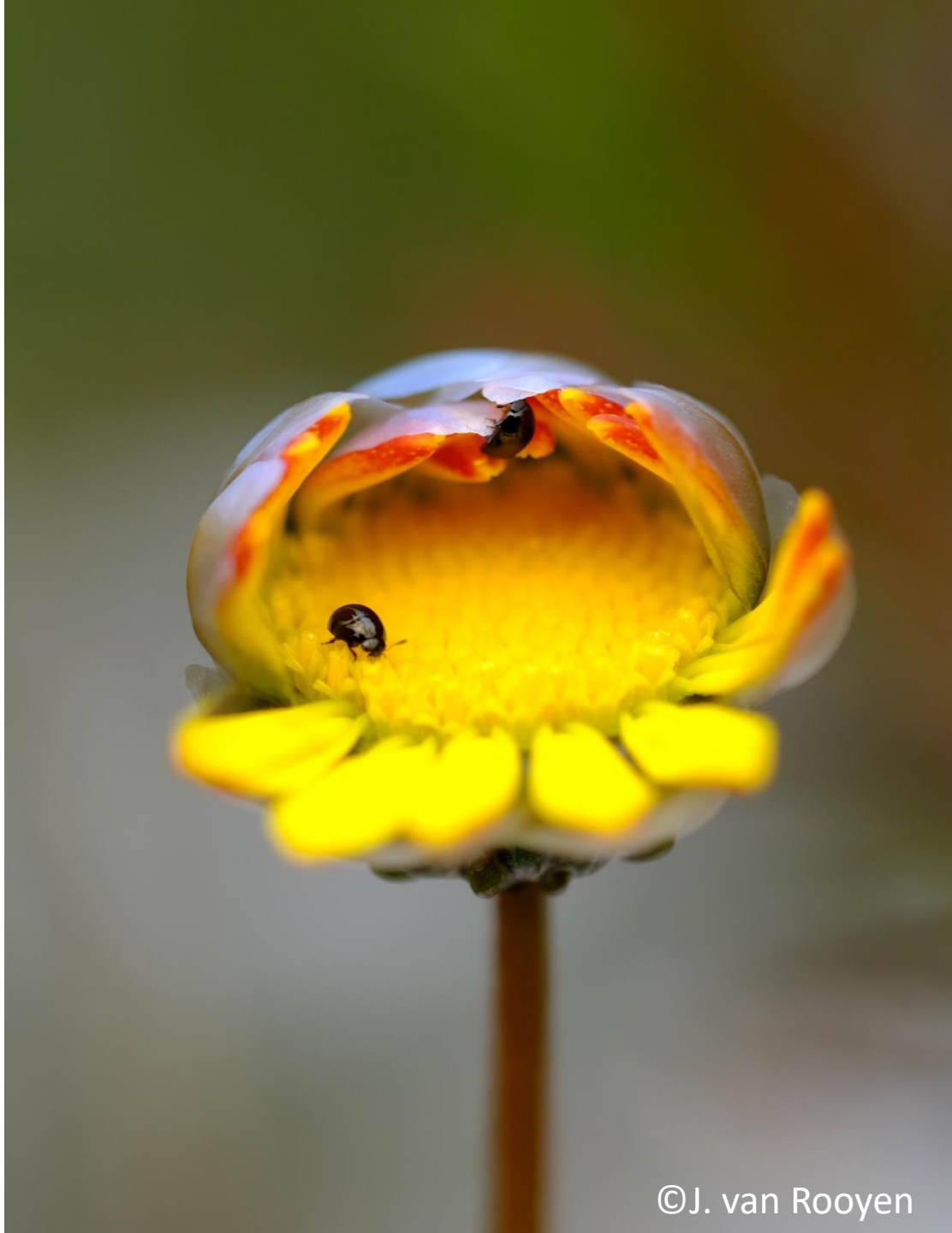
- Lack of macroecological studies for insects.

- What will happen to the plant-insect interactions in SA?

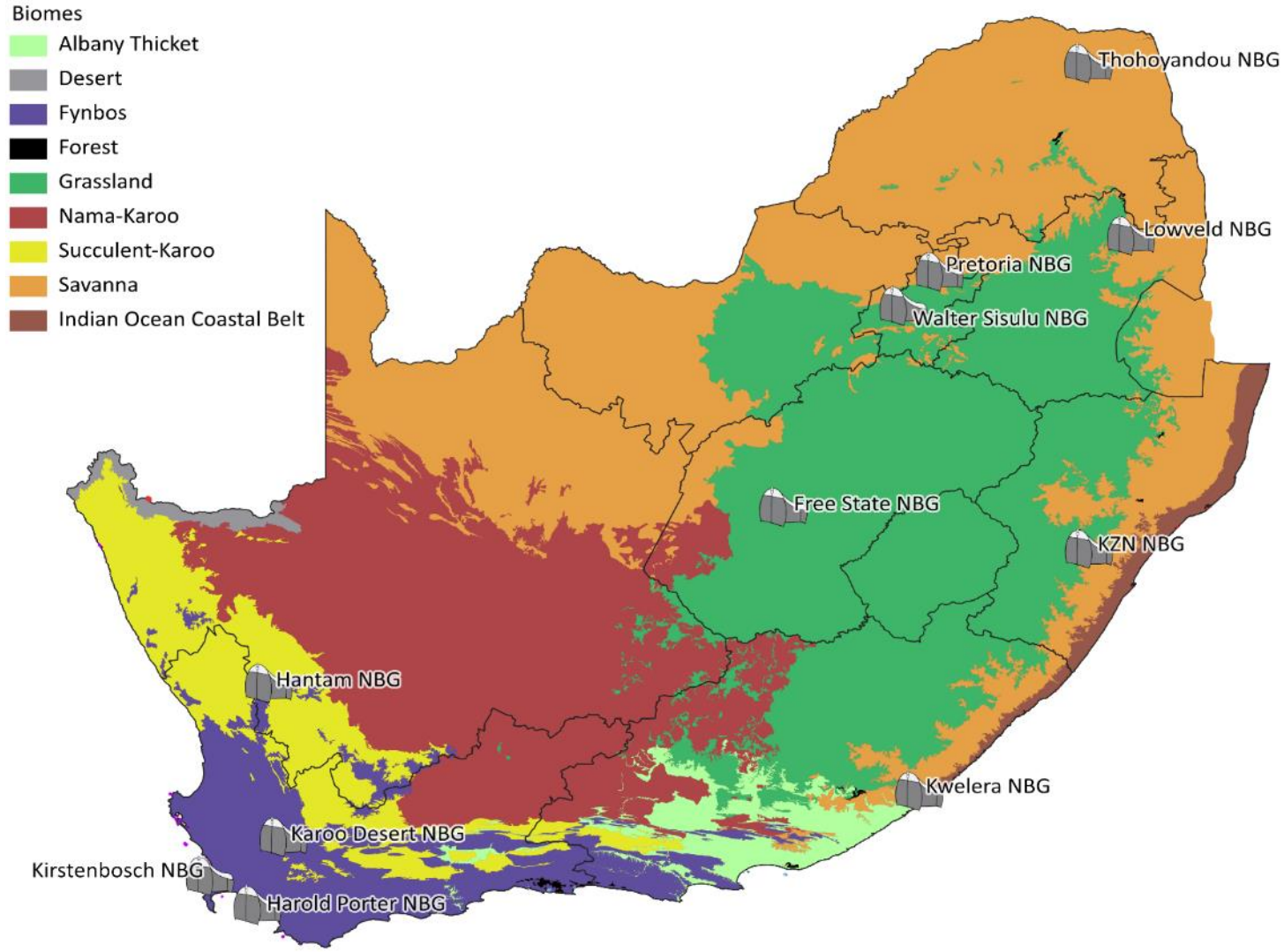
Aims and Objectives

To investigate macroecological relationships between plants and arthropods across South Africa.

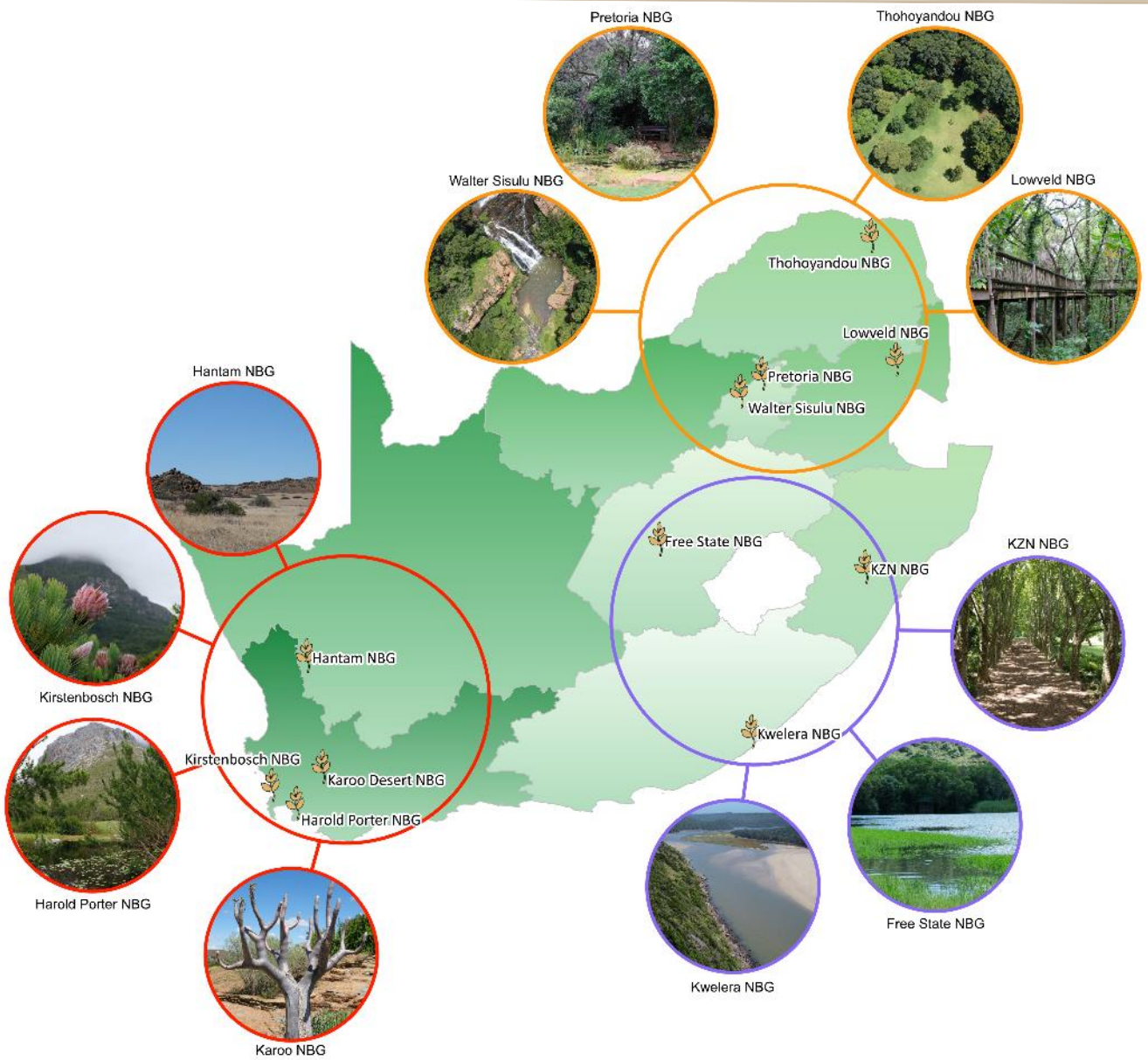
- Generate baseline data on insect diversity in South Africa and explore the diversity across different biomes.
- Find the flowering period of all the plants in different parts of South Africa.
- Investigate the relationship between plants and insect diversity.



Methods – Insect Composition

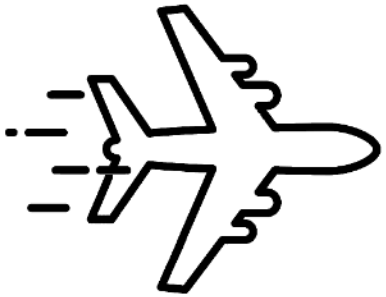


Methods – Insect Composition



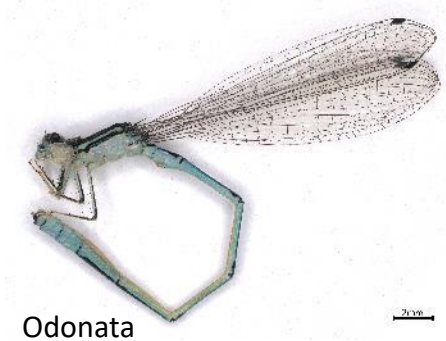
Methods – Insect Composition

R.D. Rattray

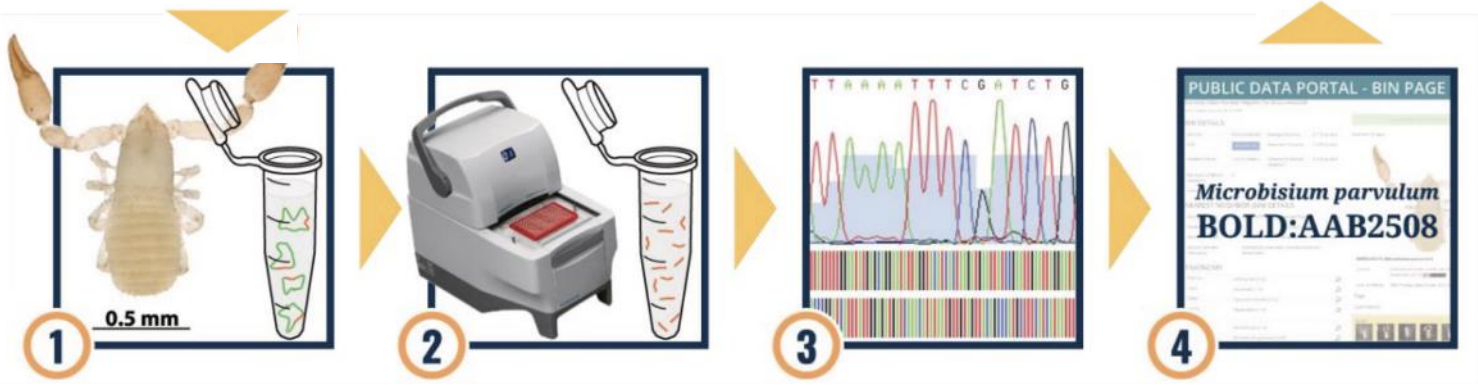



Amphipoda


Pseudoscorpiones


Odonata


Araneae



DNA Barcoding

- It is a molecular technique that uses a **short**, **standardised region** of DNA (*COI*) to identify and classify species.
- It is great because it can distinguish between species when you **don't have morphological characters** or a species ID.
- It is widely used in biodiversity research, **conservation**, and **ecological monitoring**
- "**Dark Taxa**" – Species with barcode sequence data but **missing formal taxonomic classification**.

Informatics and phylogenetics

All informatics were performed in R using several packages, including:

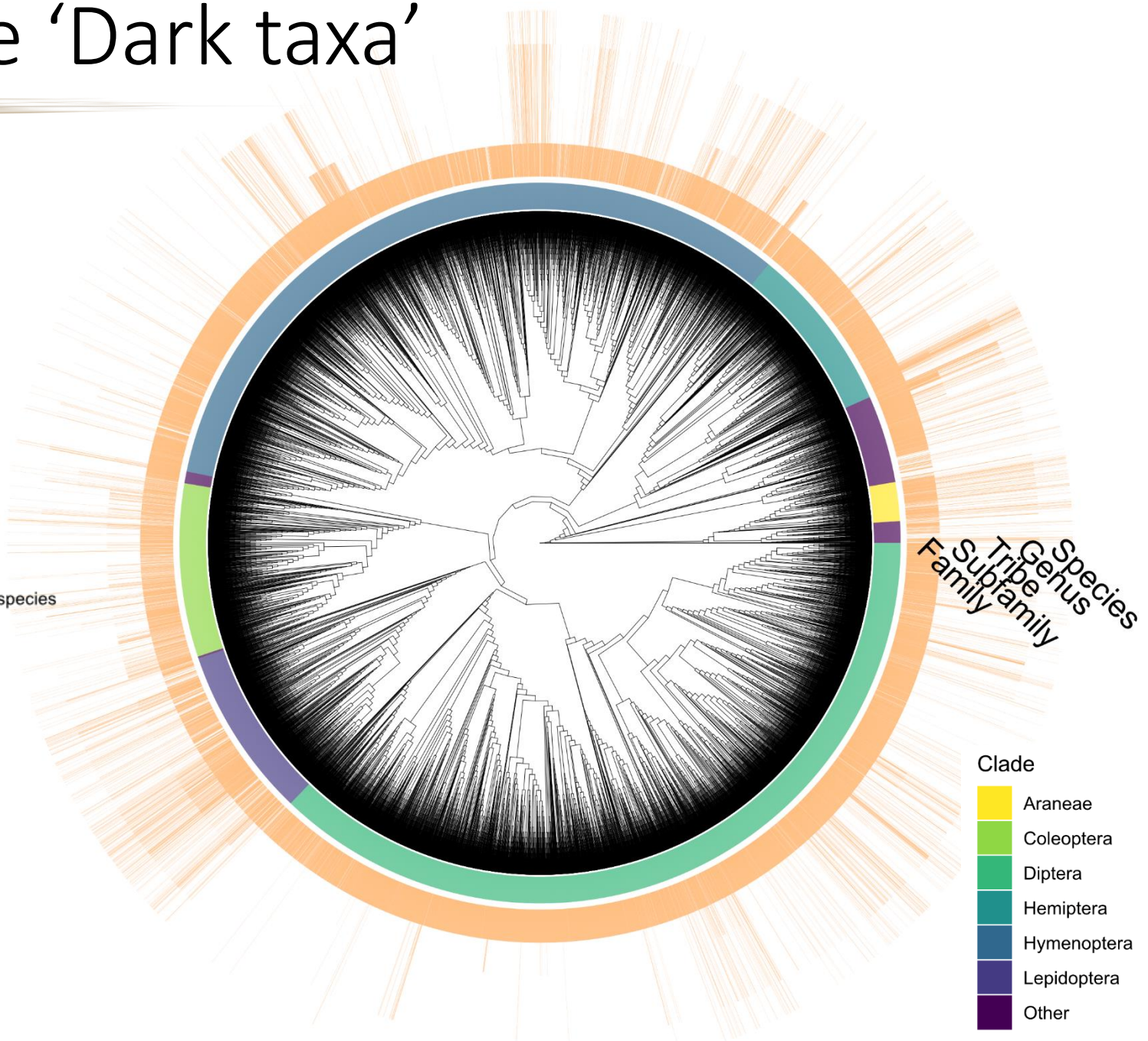
- Species estimates, Mantel tests, phylogenetic rarefaction curves, assessed the β diversity and investigated the phylogenetic turnover.

- Performed phylogenetics analyses on insects (29,919 unique species proxies) and Plants (ca. 14,000 unique species).

Results: The Numbers

- 22 Traps
- 337,809 Specimens
- 29,919 unique BINs
- 37 unique orders
- 522 unique families
- 40,994 specimens to genus
- 18,251 specimens to species

Results: Where are the ‘Dark taxa’



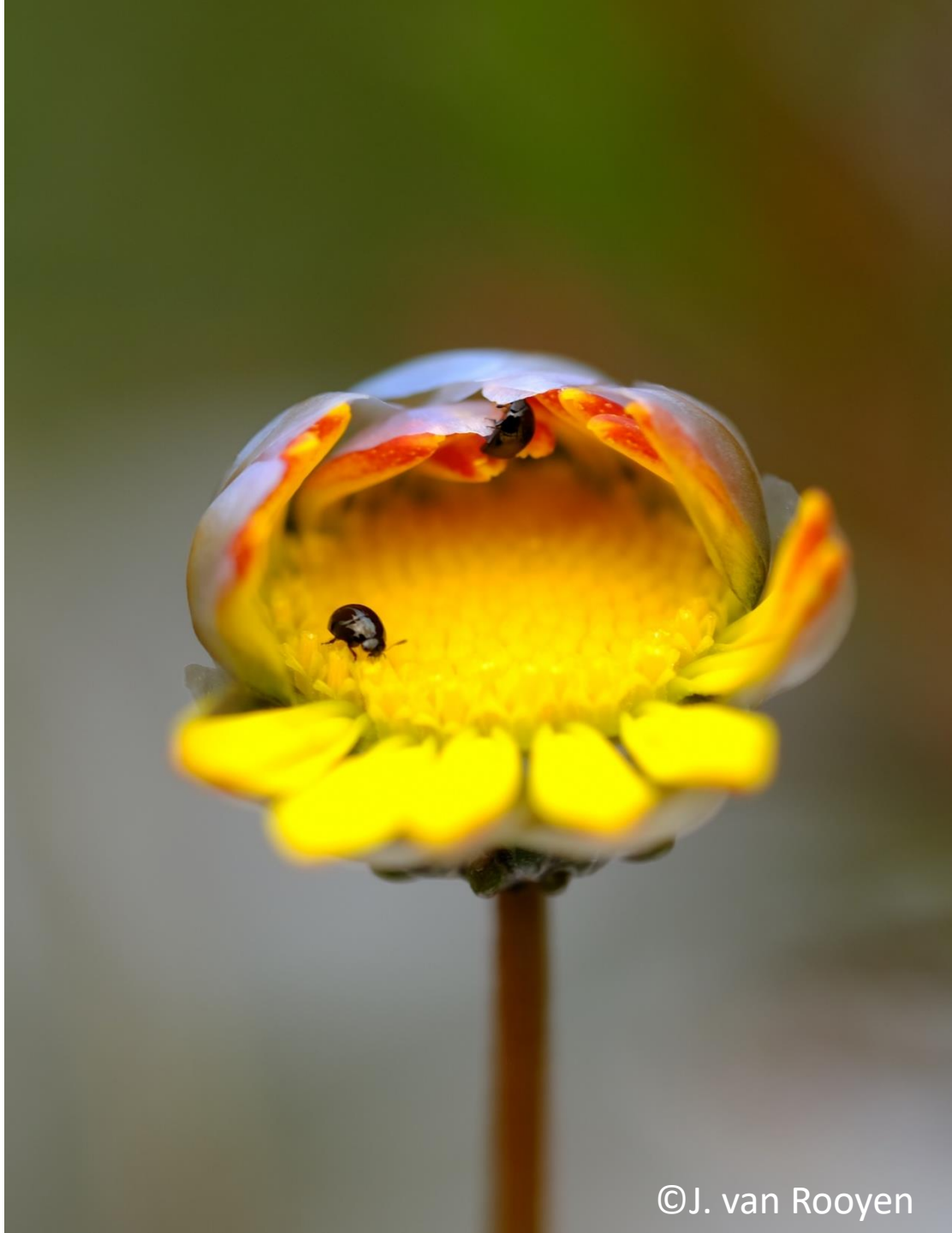
Take Home Message

- Plant species richness doesn’t really predict the insect species richness.
- ‘Dark taxa’ are growing rapidly in South Africa – we’ll need solutions.
- We have a lot more to learn about insects than we thought!

Aims and Objectives

To investigate macroecological relationships between plants and arthropods across South Africa.

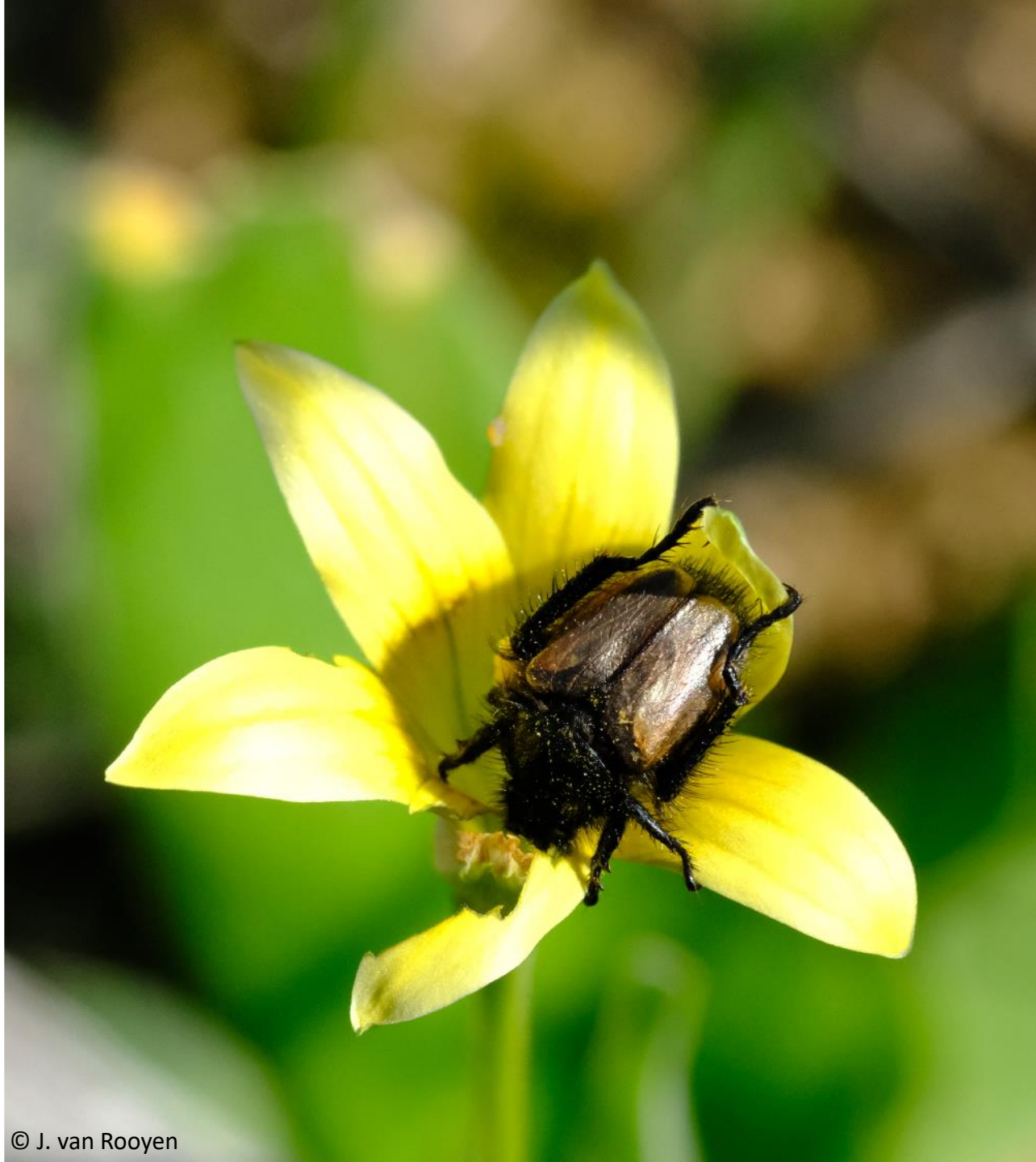
- Generate baseline data on insect diversity in South Africa and explore the diversity across different biomes.
- Find the flowering period of all the plants in different parts of South Africa.
- Investigate the relationship between plants and insect diversity.

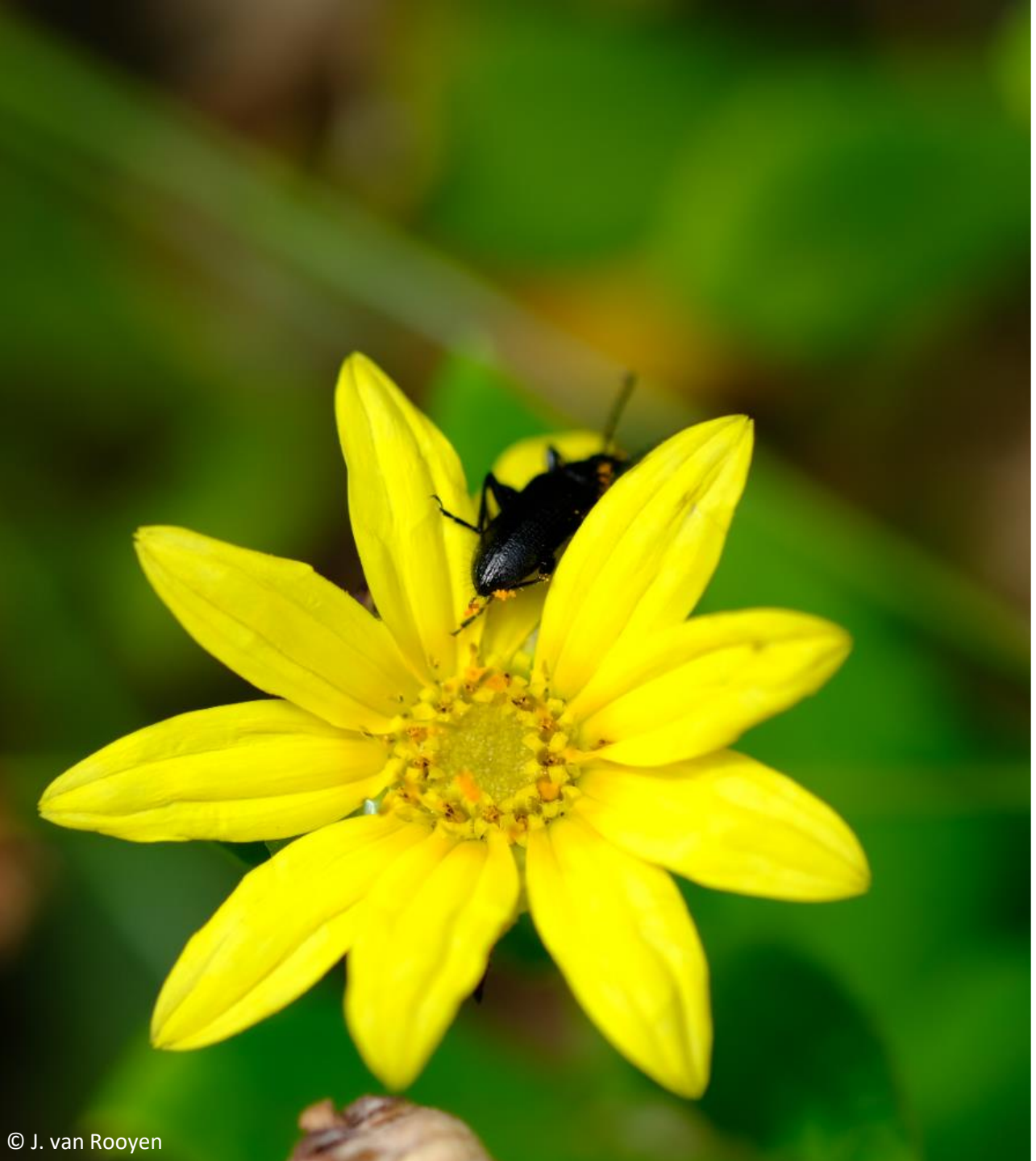


South Africa

~Is the third most biodiverse country in the world – Making it one of the richest countries in the world

- Ca. 21,000 Plants. (Zengeya et al., 2020)
- Not enough taxonomists.
- Phenological data gap.
- Need to accelerate data collection.





Machine learning

"In the age of big data, machine learning is the key to unlocking the secrets hidden within the numbers."

- Hal Varian

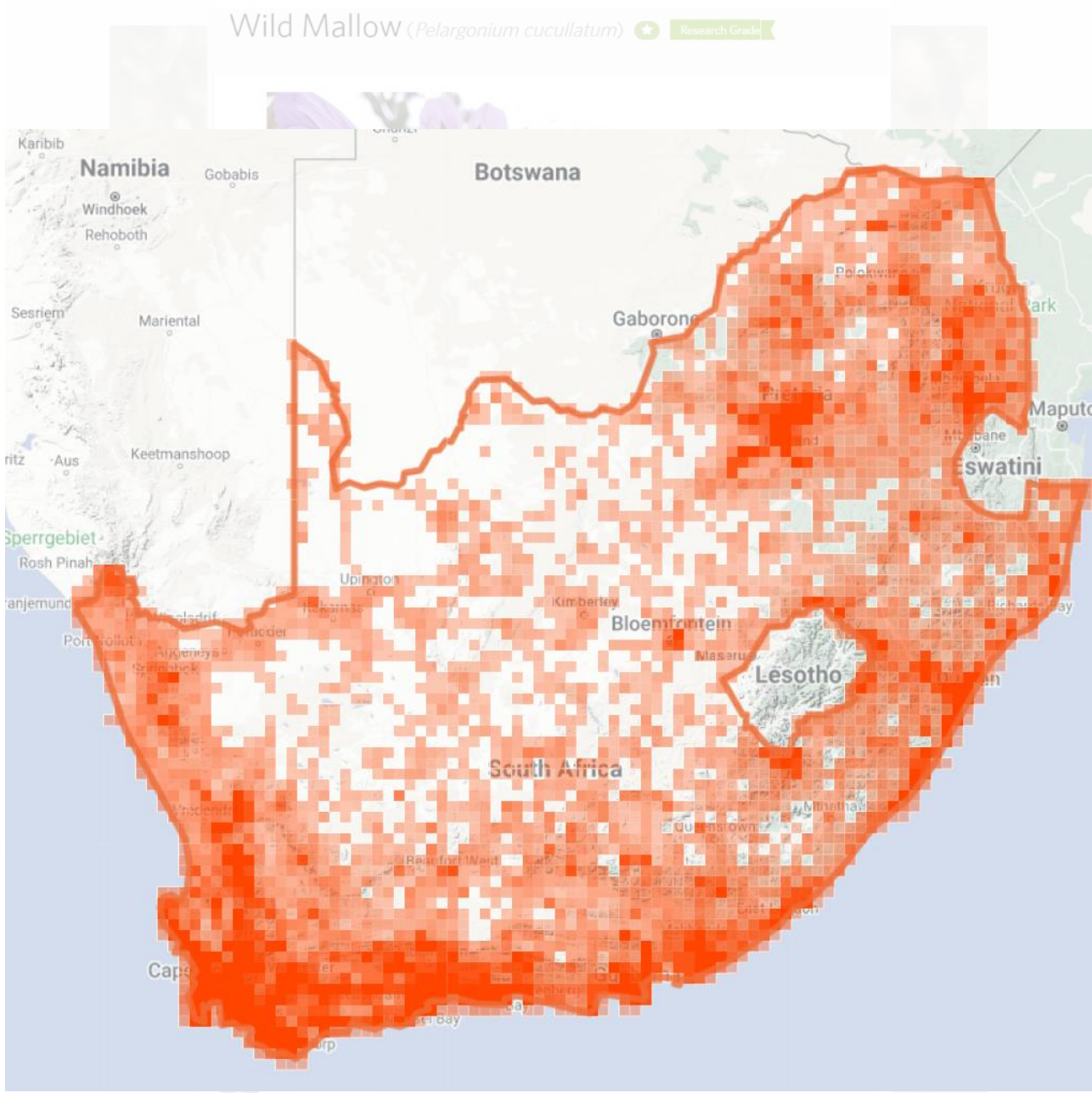
- Manual image categorization can be time-consuming.
- Convolutional neural networks (CNNs).
- Has been used to classify herbarium specimen for single species.

iNaturalist

- A community that takes pictures of species for identification.
  - Non-standardized images.

- What is a Research Grade observation?
  - Suggested ID (ML identification), two more suggested ID's (confirmation).

- 1,797,903 research grade observations in SA.



Training Data – Primary Model

5000

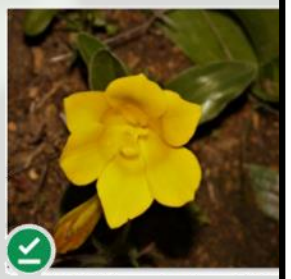

2018-09-09_Heliophila_lactea_76075005.jpg


2018-09-11_Cephalophyllum_pillans_16417621.jpg


2018-09-11_Nemesia_calcarata_18243710.jpg

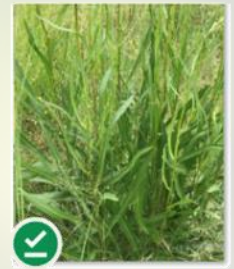

2018-09-12_Babiana_dregei_165299[illegible]jpg


2018-09-12_Pauridia_gracilipes_16894


2018-09-13_Moraea_kamiesensis_21[illegible]

Flowering

5000 Not Flowering


2020-01-24_Senecio_pentactinus_37993608.jpg


2018-05-23_Indigofera_alternans_14443827.jpg


2021-07-24_Oxalis_reflexa_88505062.jpg


2022-05-31_Pharnaceum_microphyllum_120553439.jpg


2011-09-30_Fockea_capensis_11301956.jpg


2021-08-10_Euphorbia_avasmontana_90872535.jpg

Training Data – Secondary Model

1000 Flowering


2008-01-16_Disa_longicornu_63338580.jpg


2008-02-08_Gladiolus_vinosomaculatus_10842644.jpg


2008-03-06_Utricularia_livida_37452276.jpg


2008-09-02_Foveolina_dichotoma_105029874.jpg


2008-09-03_Passerina_filiformis_10877226.jpg


2008-09-10_Podalyria_leipoldtii_56144288.jpg


2008-10-09_Gibbaeum_petrense_41321479.jpg


2008-10-25_Aspalathus_galeata_11059128.jpg

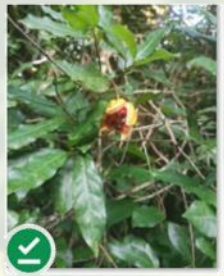

2008-10-29_Kotschya_parvifolia_85059


2008-10-30_Aristea_cuspidata_109166

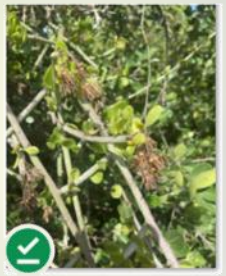

2008-11-01_Aspalathus_tylodes_34033


2008-11-03_Ixia_rouxii_10968278.jpg

168 Fruiting


2022-03-10_Xylotheca_kraussiana_112508569.jpg

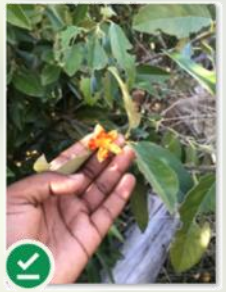

2022-04-06_Euphorbia_nutans_110662945.jpg


2022-04-10_Senecio_viminalis_111294116.jpg


2022-04-29_Cysticapnos_vesicaria_113427960.jpg


2022-04-30_Diospyros_lycioides_113707236.jpg


2022-04-30_Hedera_helix_113759727.jpg

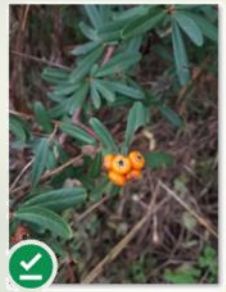

2022-04-30_Kiggelaria_africana_113787260.jpg


2022-04-30_Kiggelaria_africana_115724118.jpg


2022-05-01_Diospyros_whyteana_114147941.jpg


2022-05-02_Apodytes_dimidiata_114591974.jpg


2022-05-02_Kiggelaria_africana_114614997.jpg


2022-05-08_Pyracantha_angustifolia_116086163.jpg

Determining the Phenology

Download Images
(1,8 million)

For the code scan below:

https://github.com/rossdstewart/ML-Phenology-Code

Image categorization

Extract the date
from each image.

Extract the
Phenological pattern.

Abrus_precatorius

Example: *Abrus precatorius*

F – Flowering; NF – Not-Flowering; F2 – Flowering 2; Fu - Fruiting

F=54.47%; NF=45.53%
F2=14.85%; **Fu=85.15%**

F=12.19%; **NF=87.81%**

F=13.82%; **NF=86.18%**

F=2.48%; **NF=97.52%**

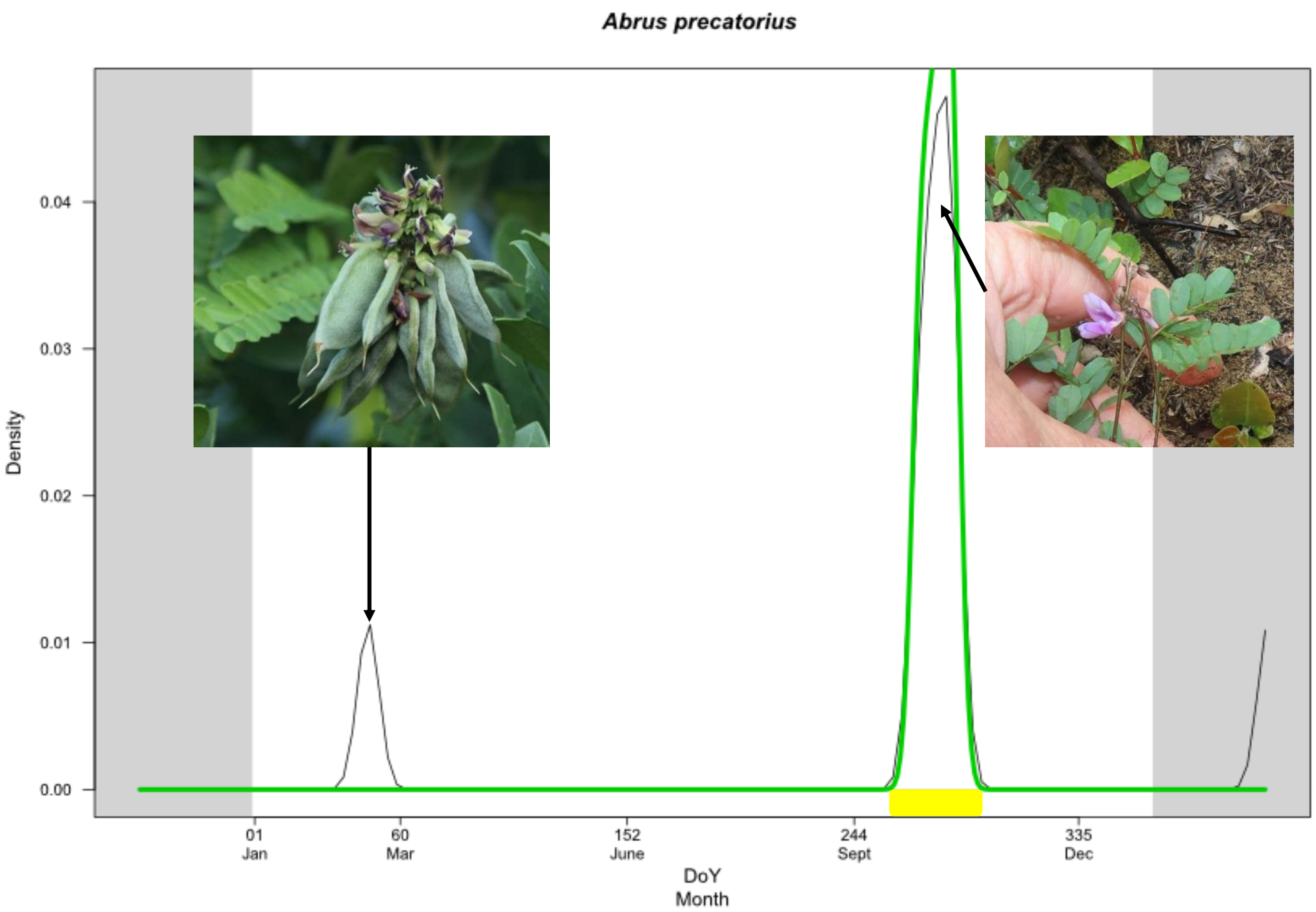
F=64%; NF=36%
F2=99.5%; Fu=0.5%

F=0.4%; **NF=99.6%**

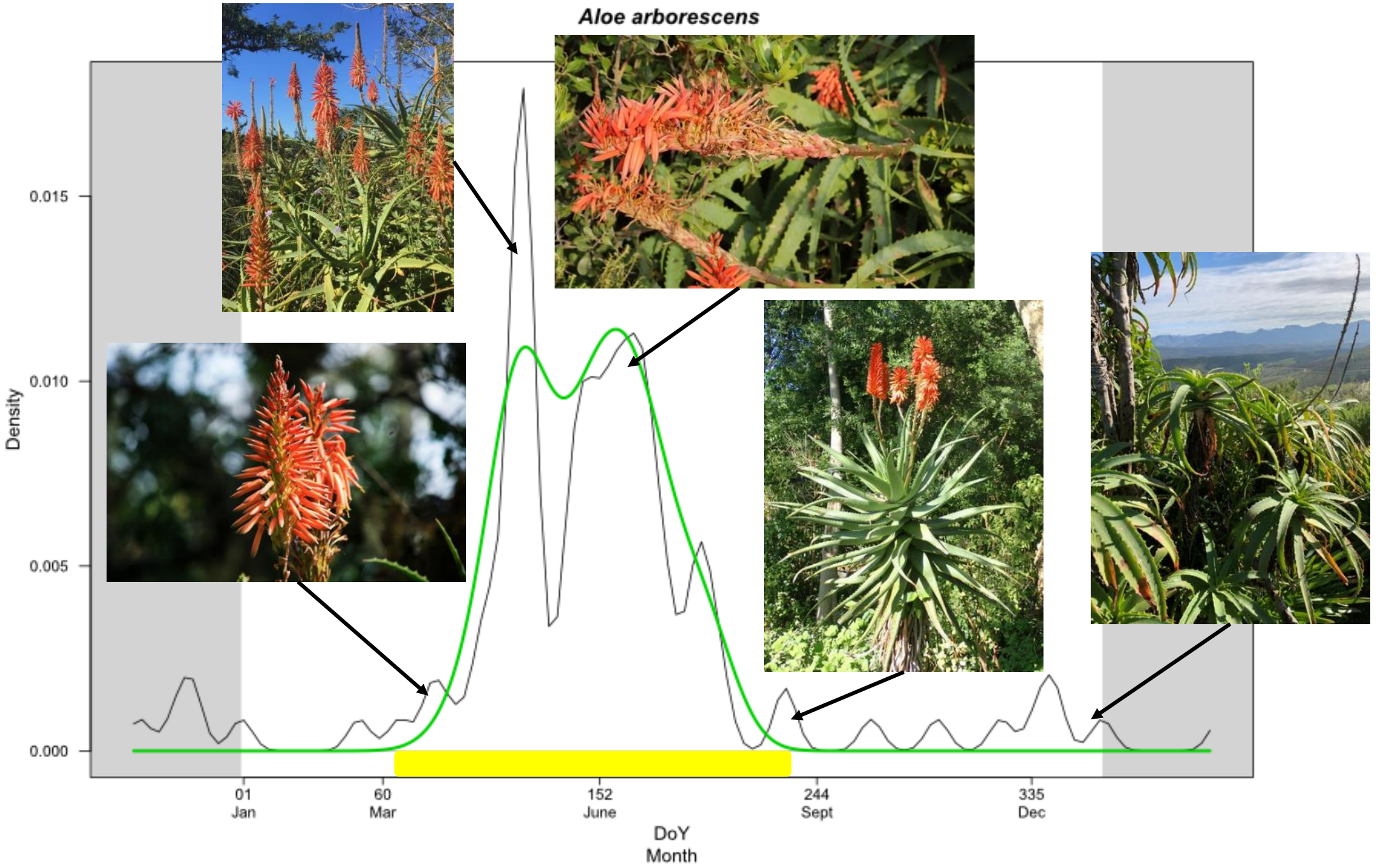
F=86.19%; NF=13.81%
F2=99.95%; Fu=0.05%

F=29.53%; **NF=70.47%**

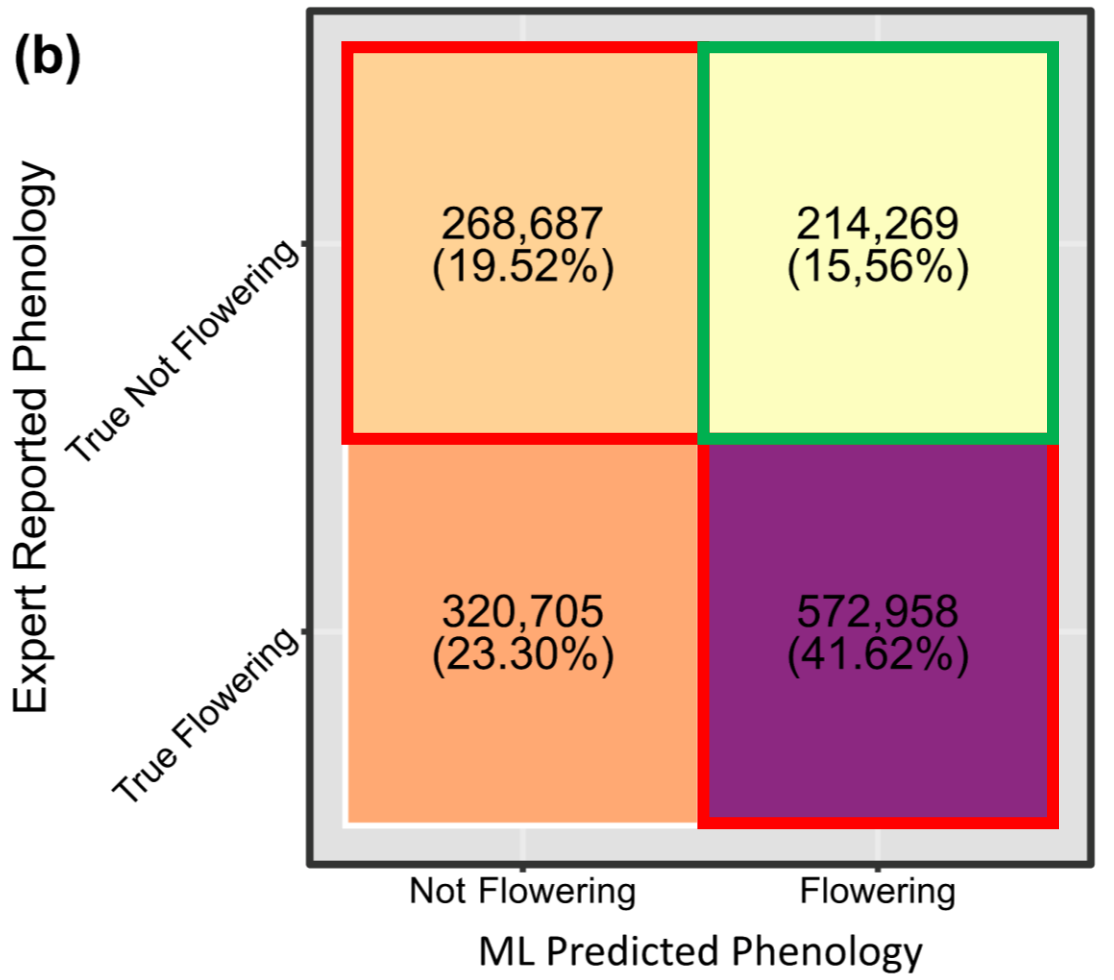
F=98.11%; NF=1.89%
F2=95.89%; Fu=4.11%


Abrus precatorius
Density
0.04
0.03
0.02
0.01
0.00
01
Jan
60
Mar
152
June
244
Sept
335
Dec
DoY
Month

n=7


Aloe arborescens
0.015
0.010
0.005
0.000
Density
01
Jan
60
Mar
152
June
244
Sept
335
Dec
DoY
Month

n=77

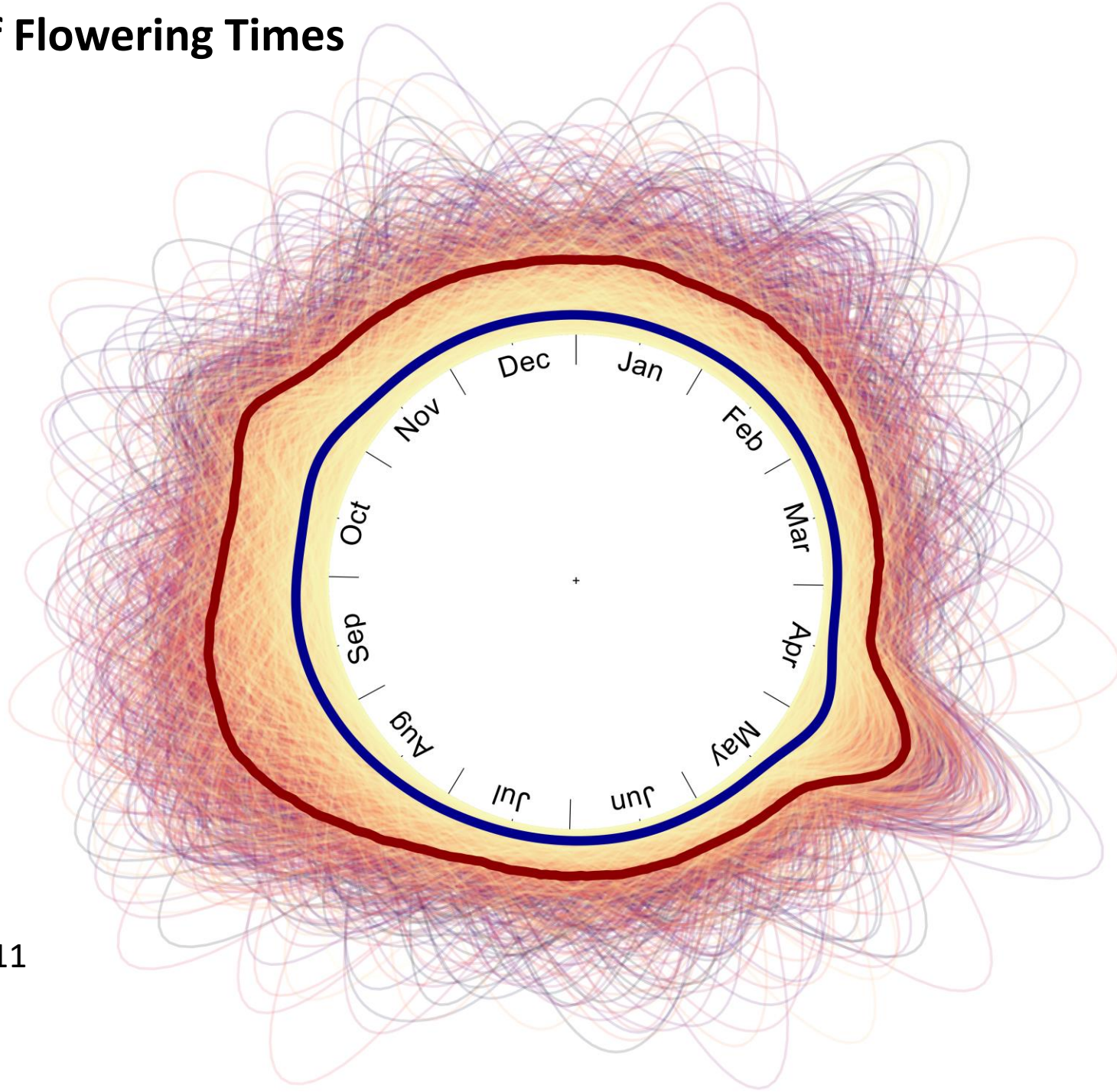


(a) – Accuracy of two-step prediction method

(b) – Accuracy of predicted phonologies

Manning & Goldblatt (2012)

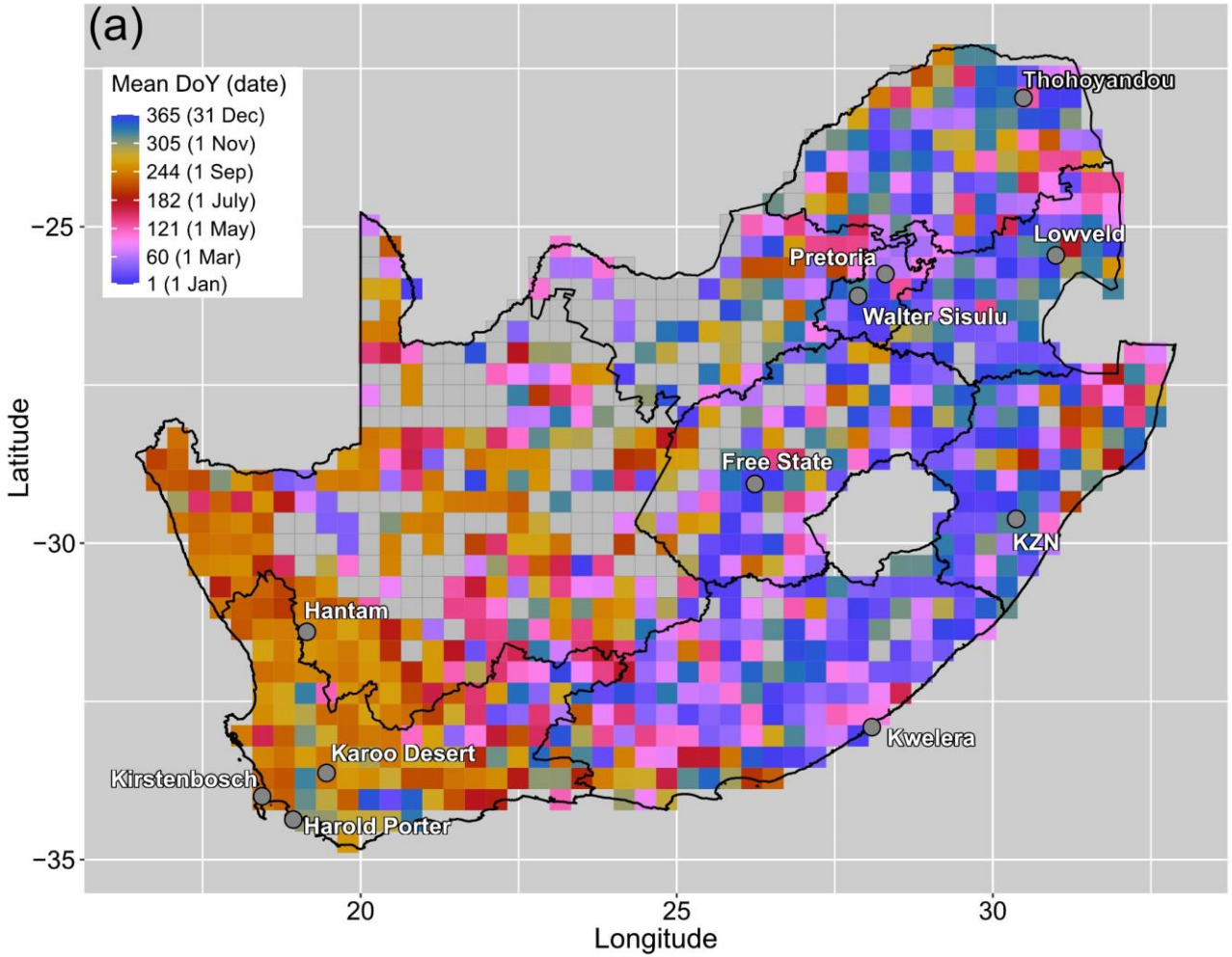
Density Diagram of Flowering Times



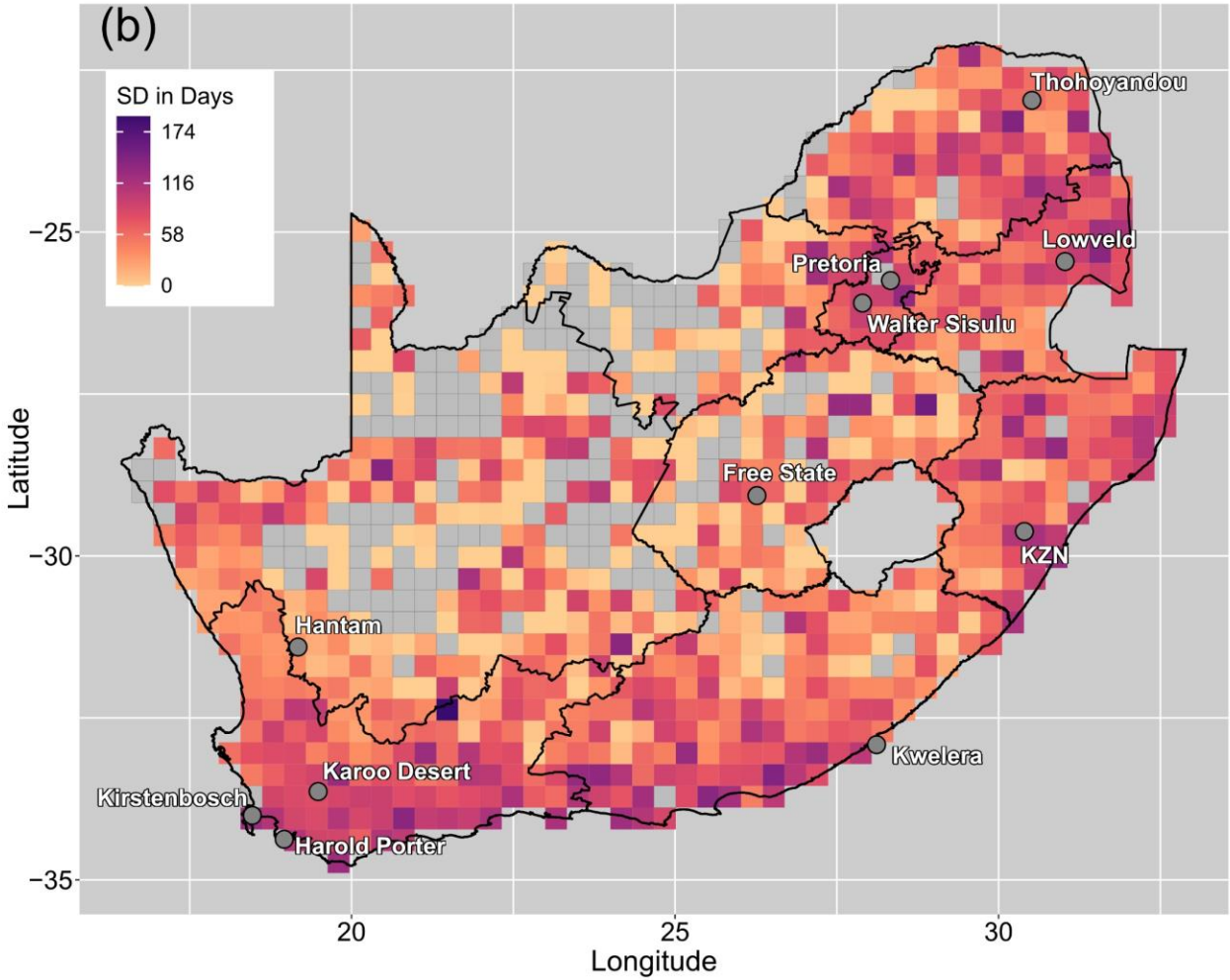
Species with > 5 records
Species = 6,986
Flowering images = 497,911

95% quantile Flowering time
Mean Flowering time

Flowering Times of South Africa



(a) – Mean phenology of South Africa



(b) – Variation of phenology of South Africa

Take Home

"All models are wrong, but some of them are useful."
- George Box

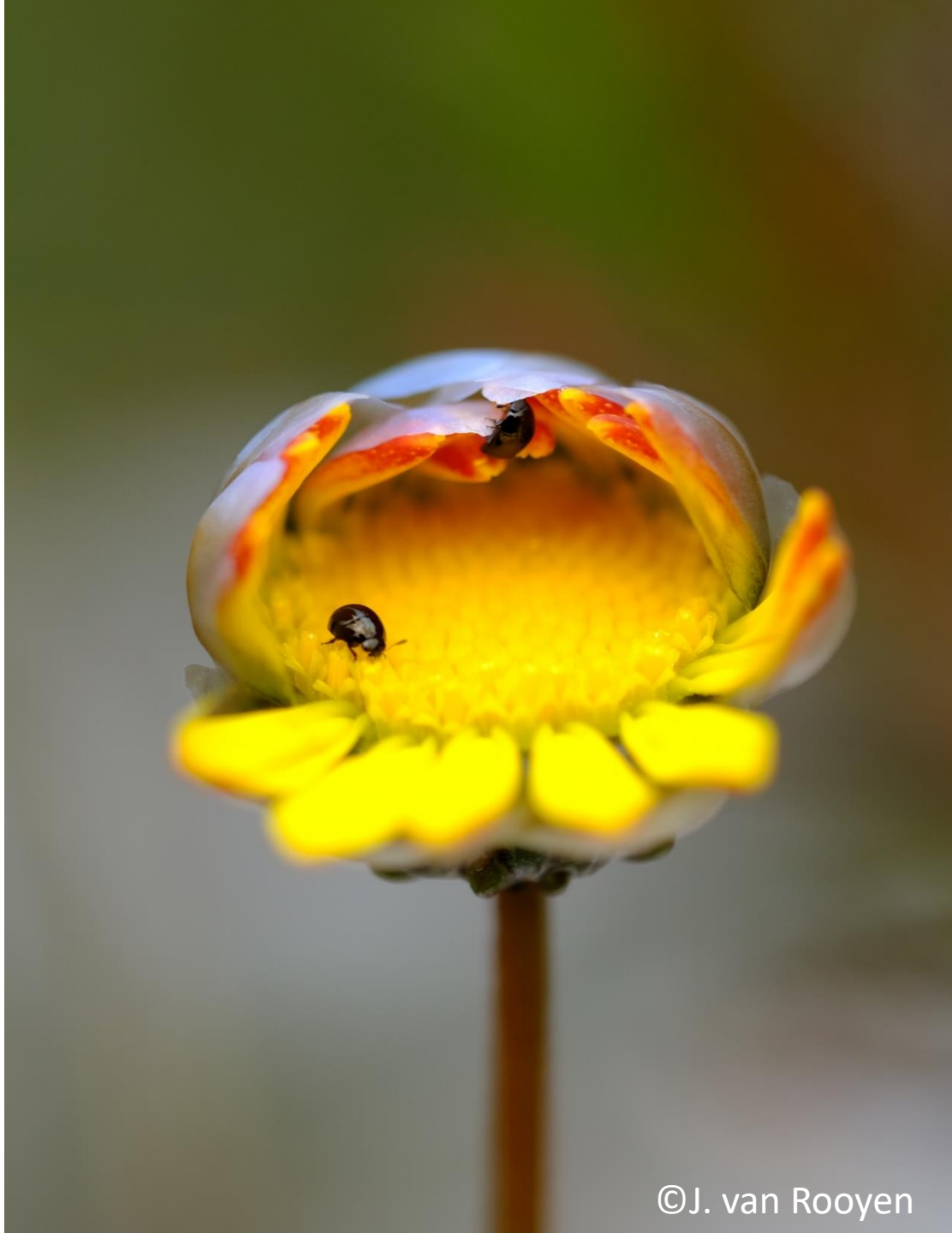
- We have successfully used Machine Learning to determine plant flowering patterns at a large scale, and we can start investigating which species are at risk of extinction due to climate change.

- I want to encourage people to contribute to iNaturalist

Aims and Objectives

To investigate macroecological relationships between plants and arthropods across South Africa.

- Generate baseline data on insect diversity in South Africa and explore the diversity across different biomes.
- Find the flowering period of all the plants in different parts of South Africa.
- Investigate the relationship between plants and insect diversity

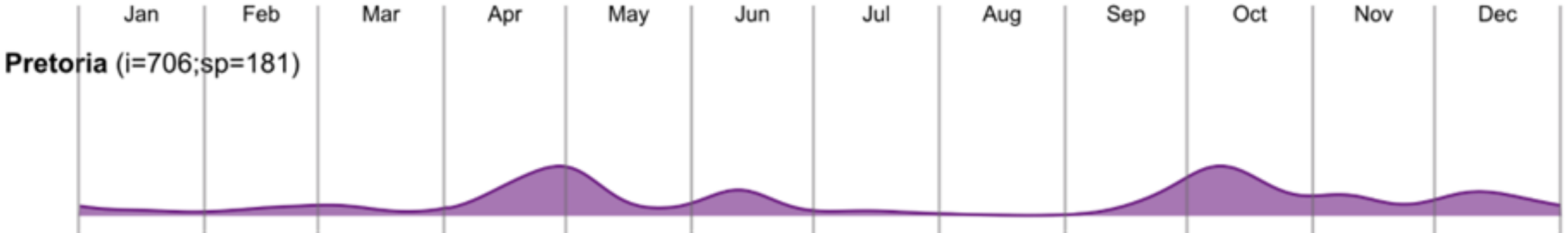


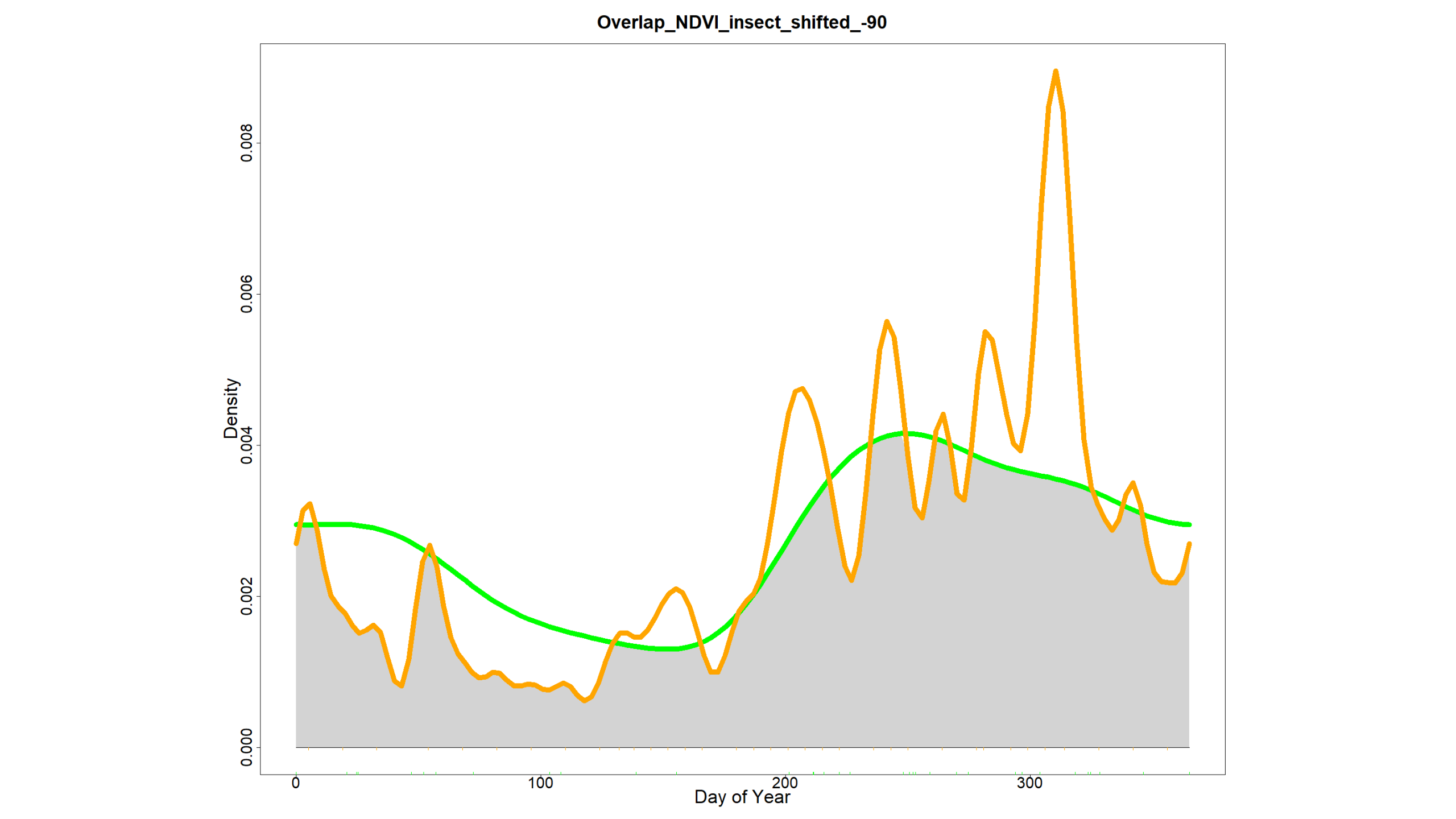
Pretoria Garden

Numbers:

- Specimen: 45,210
- Species (BINs): 5,539
- ID’ed species: 324

Cultivated	**Natural**
Specimen: 22,350	Specimen: 22,860
Species (BINs): 3,617	Species (BINs): 3,946
ID’ed species: 214	ID’ed species: 246




Overlap_NDVI_insect_shifted_-90
Density
0.000
0.002
0.004
0.006
0.008
0
100
200
300
Day of Year

Thank you to Everyone for Helping



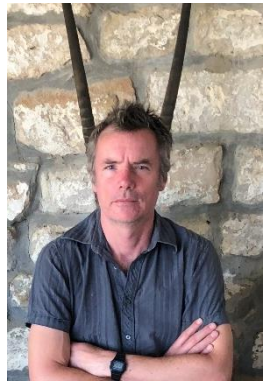












The End

J. van Rooyen