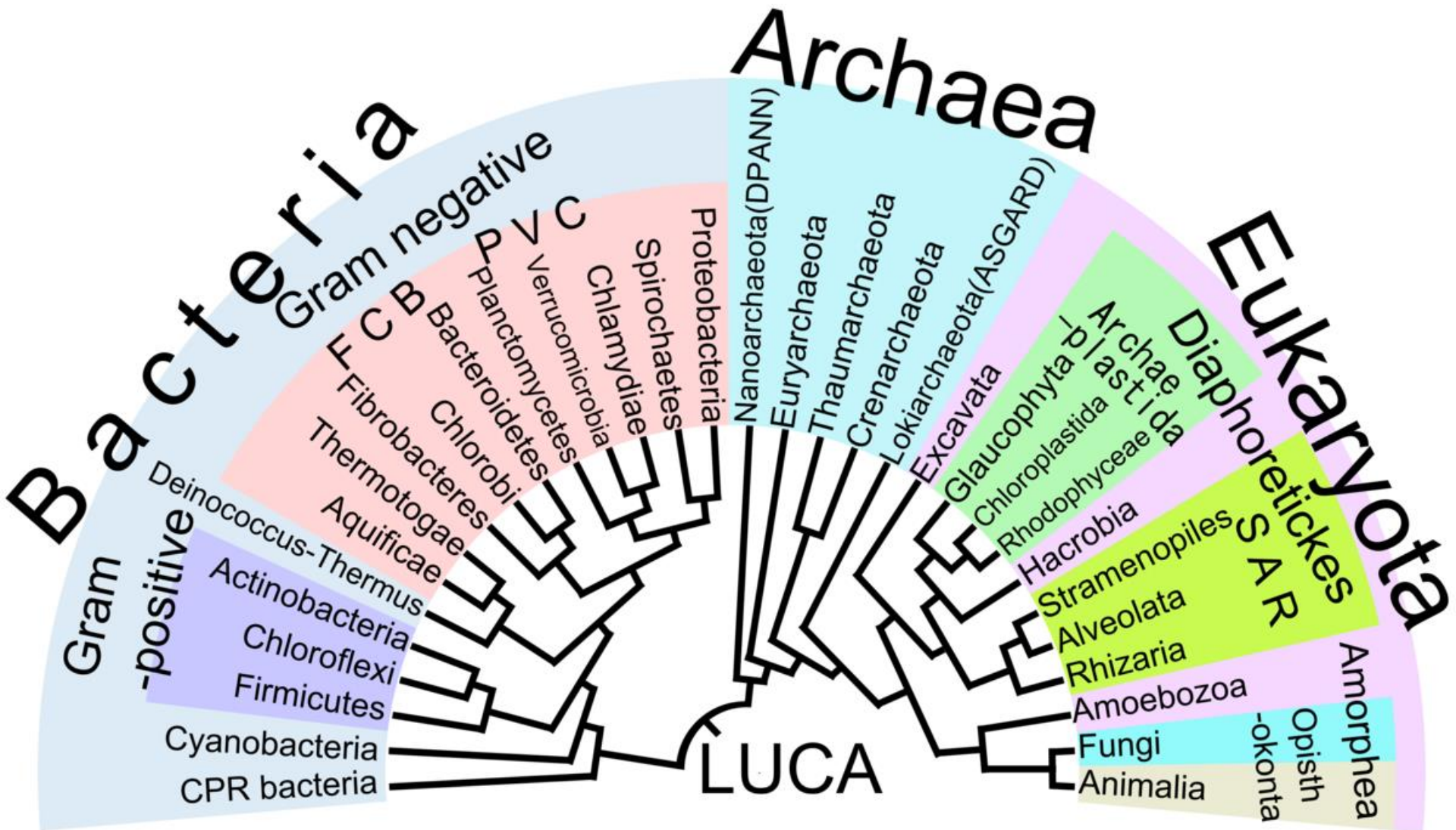# Building "trees" of South African shrubs
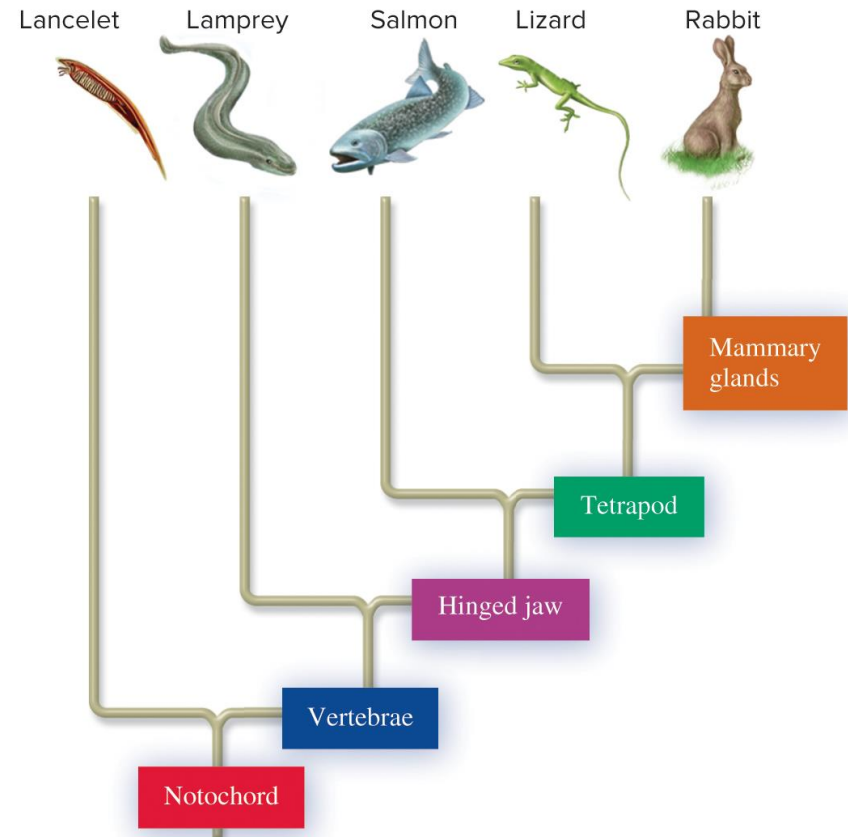
THE TREE OF LIFE

Building a phylogeny

- **Phylogeny** = evolutionary relationships among organisms
- Group organisms based on **shared**, **derived** features

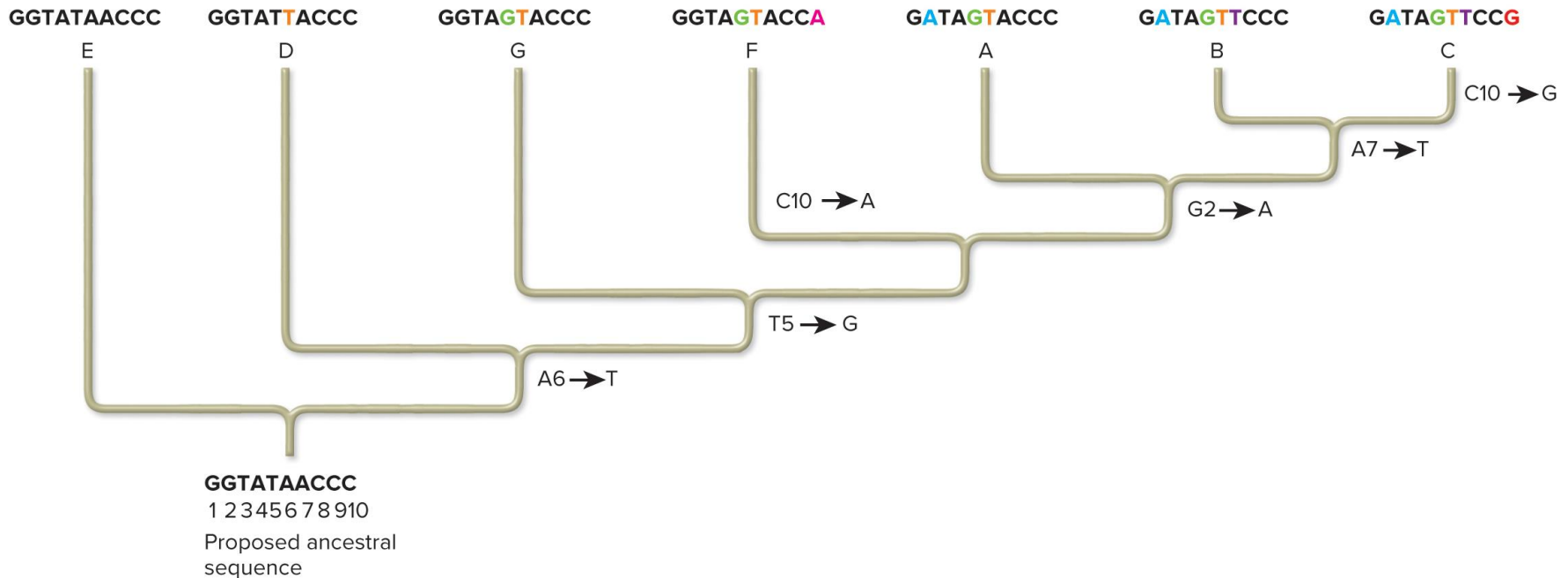|	Lancelet	Lamprey	Salmon	Lizard	Rabbit
Notochord	Yes	Yes	Yes	Yes	Yes
Vertebrae	No	Yes	Yes	Yes	Yes
Hinged jaw	No	No	Yes	Yes	Yes
Tetrapod	No	No	No	Yes	Yes
Mammary glands	No	No	No	No	Yes

(a) Characteristics among species

(b) Cladogram based on morphological traits

Building a phylogeny

- Often built using molecular data (most commonly DNA, also RNA, protein)

GGTATAACCC — E
GGTATTACCC — D
GGTAGTACCC — G
GGTAGTACCA — F
GATAGTACCC — A
GATAGTTCCC — B
GATAGTTCCG — C

C10 → G
A7 → T
C10 → A
G2 → A
T5 → G
A6 → T

GGTATAACCC
1 2 3 4 5 6 7 8 9 10
Proposed ancestral
sequence

Note: A, T, G, and C refer to nucleotide bases, and the numbers refer to the position of the base in the nucleotide sequences. For example, A6 refers to an adenine at the sixth position.
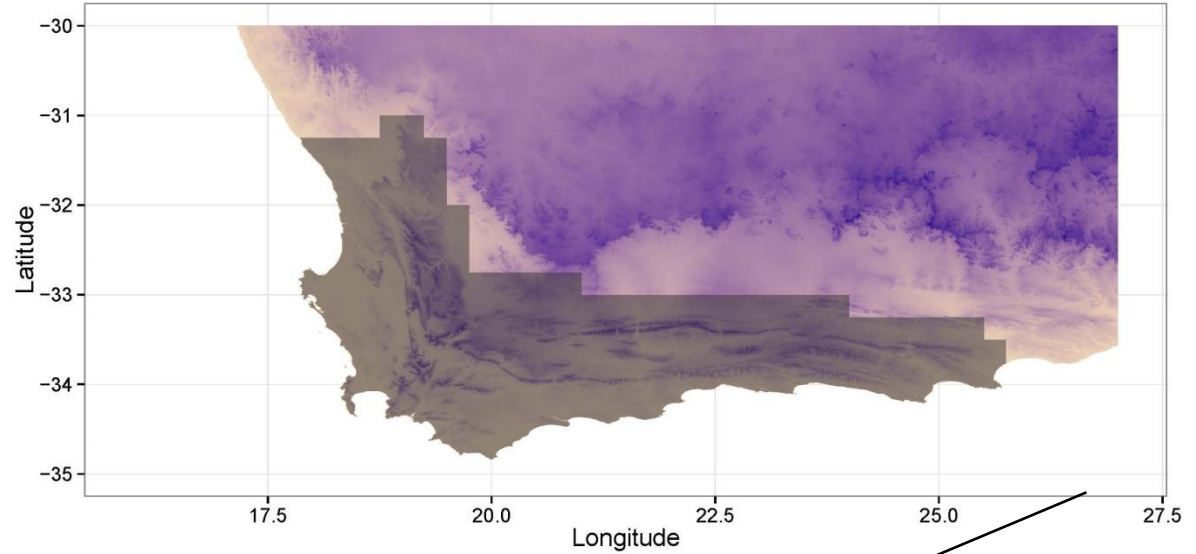
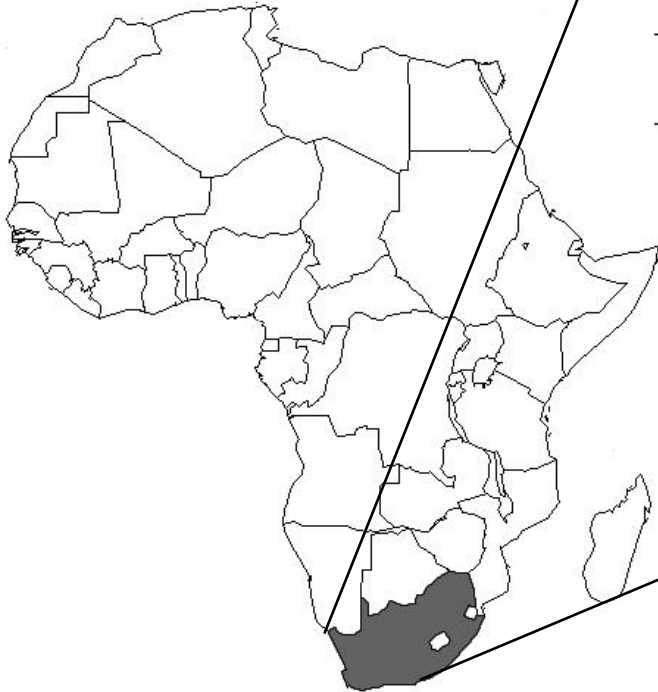# Biodiversity Hotspots

Myers et al. (2010)
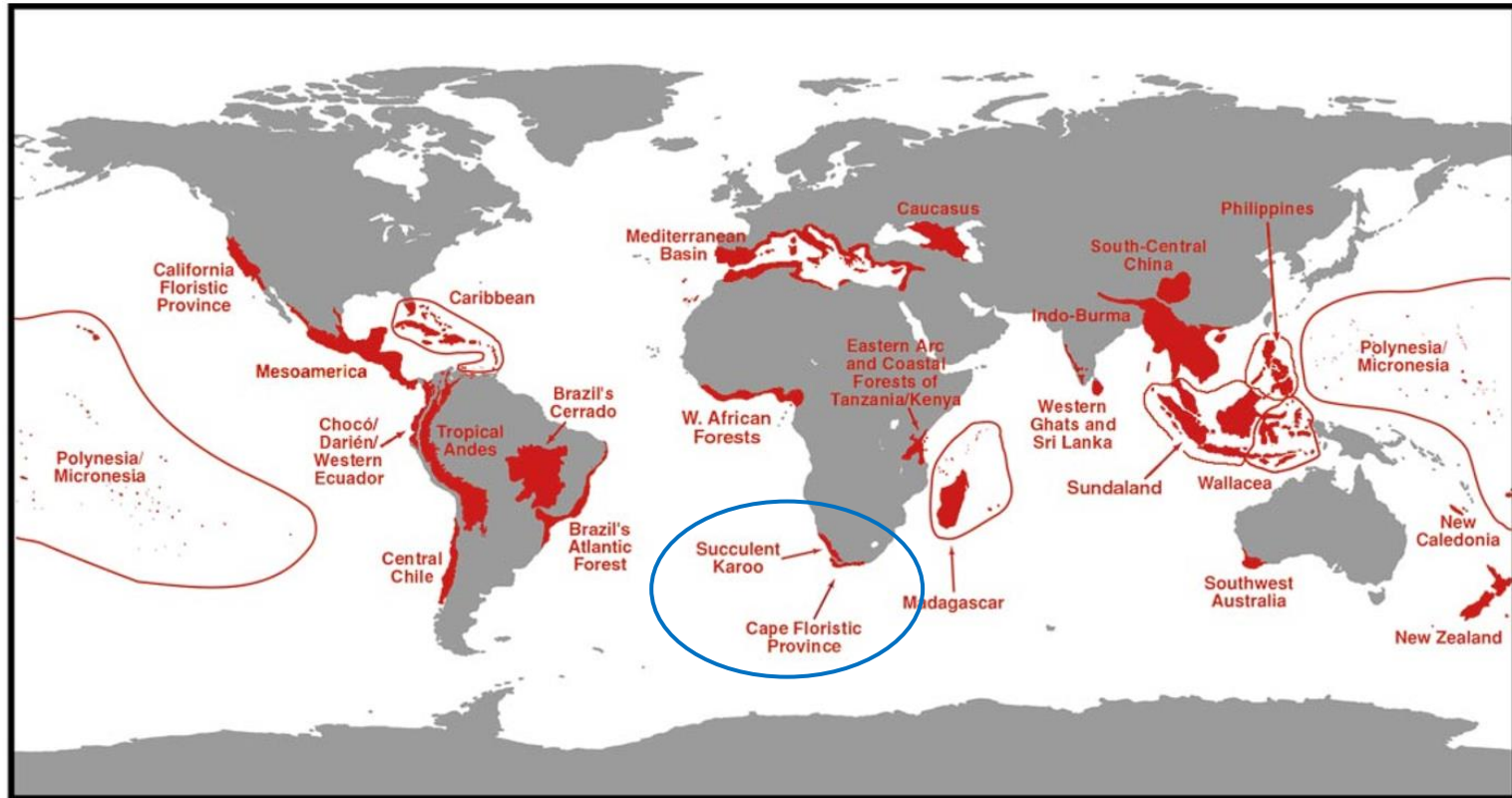
South Africa

> 35 languages spoken (11 official)

Cape Floristic Region

> 9000 plant species
~ 70% endemic

Protea

- Approximately 112 species
- Mostly shrubs
- 60% found in CFR

Species richness
- 29 - 45
- 25 - 28
- 16 - 24
- 12 - 15
- 8 - 11
- 5 - 7
- 3 - 4
- 0 - 2

Valente et al. (2010)

Protea

Protea

Protea

P. venusta   P. coronata   P. punctata   P. nubigena   P. canaliculata   P. subulifolia   P. acaulos

P. obtusifolia   P. caffra   P. comptonii   P. dracomontana   P. curvata   P. longifolia

P. grandiceps   P. eximia   P. tenax   P. cynaroides   P. susannae   P. neriifolia   P. denticulata

P. piscina   P. lorifolia   P. scolopendriifolia   P. scabra   P. cryophila   P. laurifolia   P. aurea potbergensis   P. cordata   P. compacta   P. repens   P. repens   P. nitida
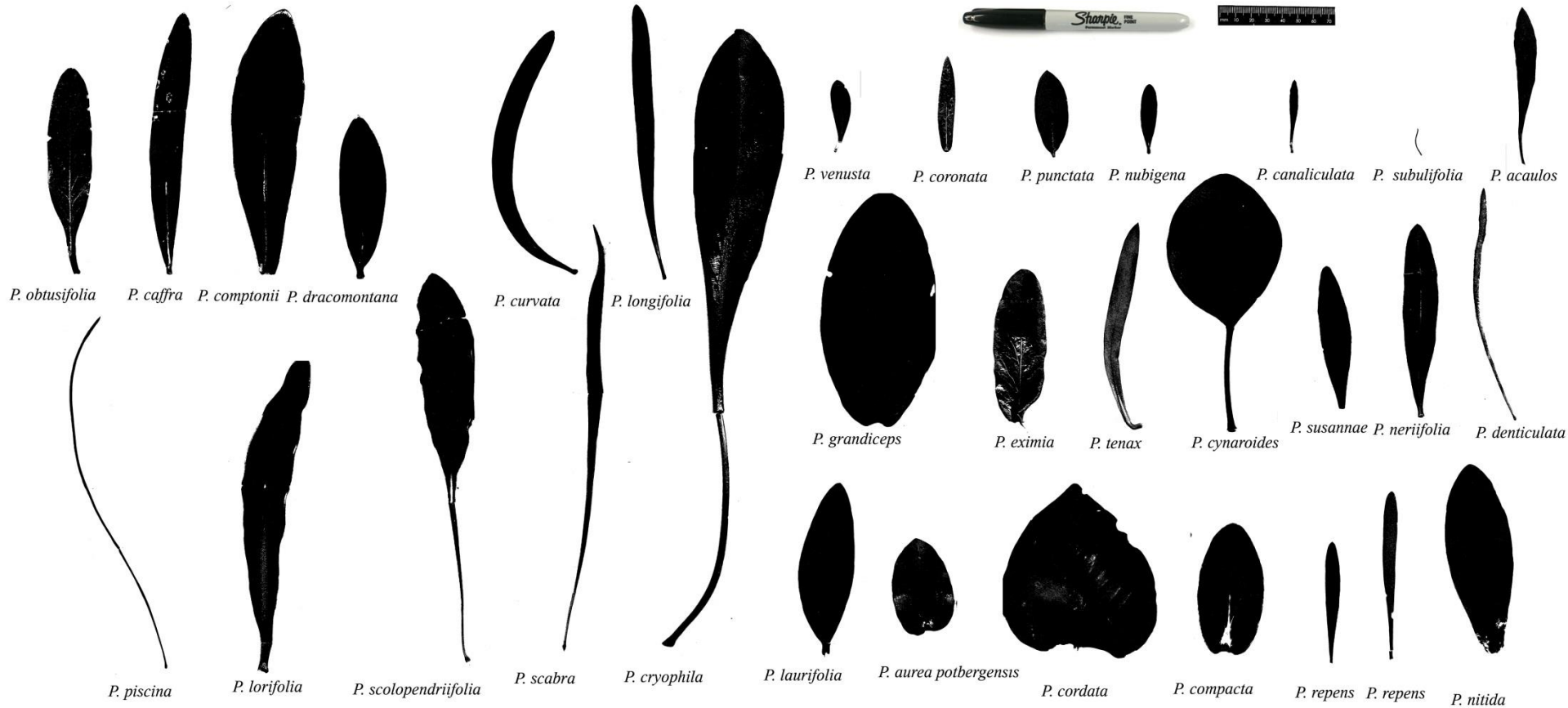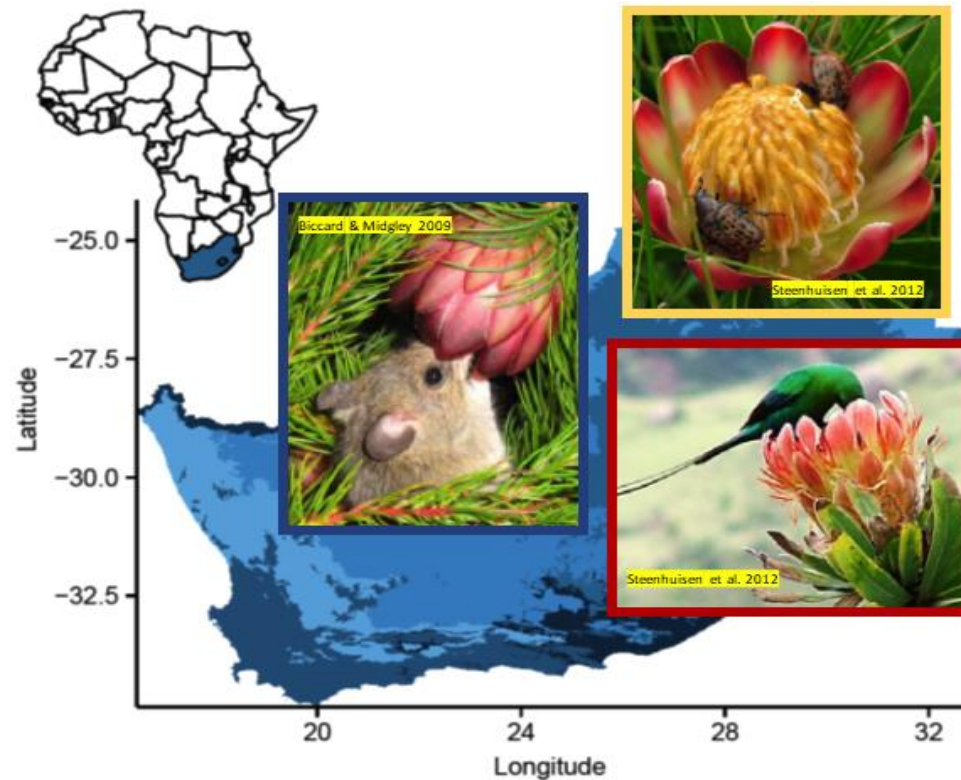
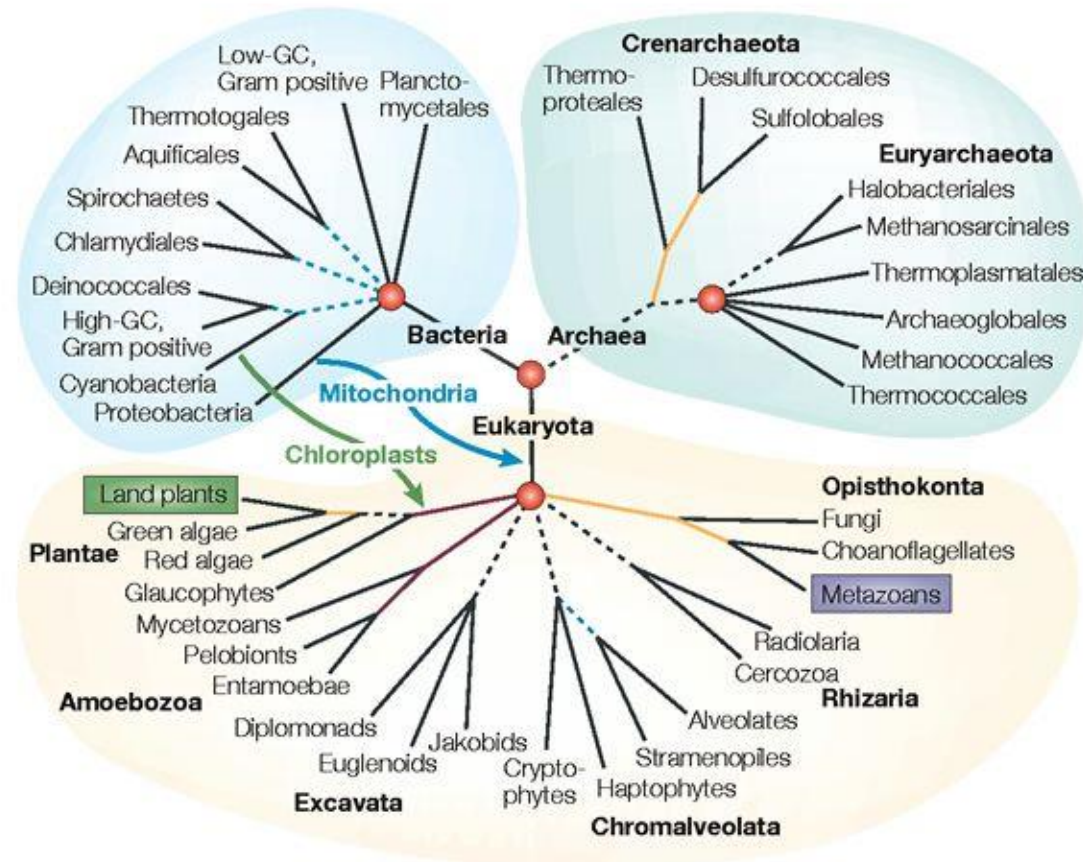Photo Credit: Jane Carlson

Major questions:

- How are species related to each other?
- How have different pollination syndromes evolved?

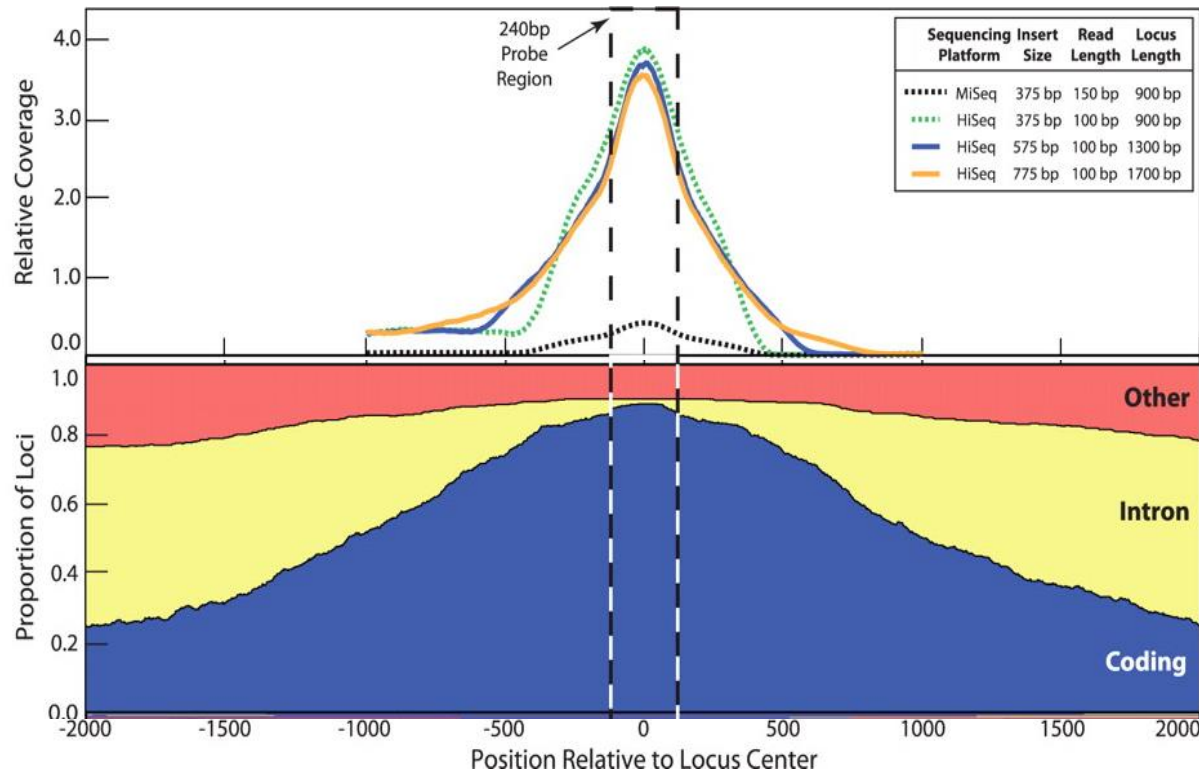# Phylogenomics

Use whole or reduced **genomes** to infer evolutionary relationships
Genome = ALL of the DNA!

Nature Reviews | Genetics

Delsuc *et al.* (2005)

Anchored Phylogenomics

- Capture of **hundreds** of gene sequences at once

Lemmon *et al.* (2012), Buddenhagen *et al.* In revision
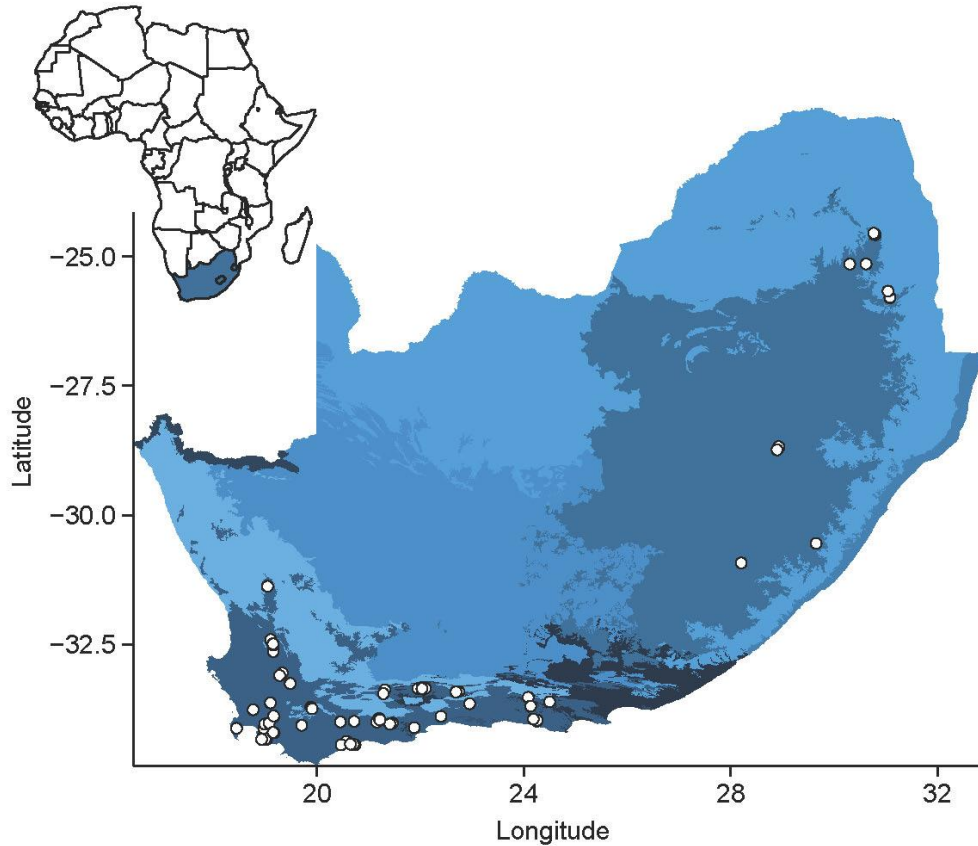
Protea Data Collection

Protea Data Collection

Protea Data Collection

Protea Data Collection

- 164 DNA samples

- 59 *Protea* species

  (+ 6 outgroup)

- 498 loci captured (average length 551bp)

- About 275,000bp each

Mitchell *et al.* (2017) *The American Journal of Botany*

That's a LOT of data…

- HPC enables us to analyze lots of data using specific software!

- We will be using RAxML (**R**andomized **Ax**elerated **M**aximum **L**ikelihood) to build phylogenies

- Use program (TreeGraph2 or FigTree) to look at output

- Goals: build evolutionary trees for subset of species, explore different approaches

amkozlov/**raxml-ng**

RAxML Next Generation: faster, easier-to-use and more flexible

Let's use the cluster to build phylogenies!

- 3 different "scripts" to run
- Follow along with me as we use OnDemand / BOSE to…
  - Build the "best" tree
  - Build "bootstrap" trees (other possibilities)
  - Decide how confident we are on how species are related

Protea aurea   *Protea caffra*   *Protea cordata*   *Protea cynaroides*

Protea gaguedi   *Protea lanceolata*   *Protea laurifolia*   *Protea longifolia*

Protea magnifica   *Protea montana*   *Protea nitida*   *Protea punctata*
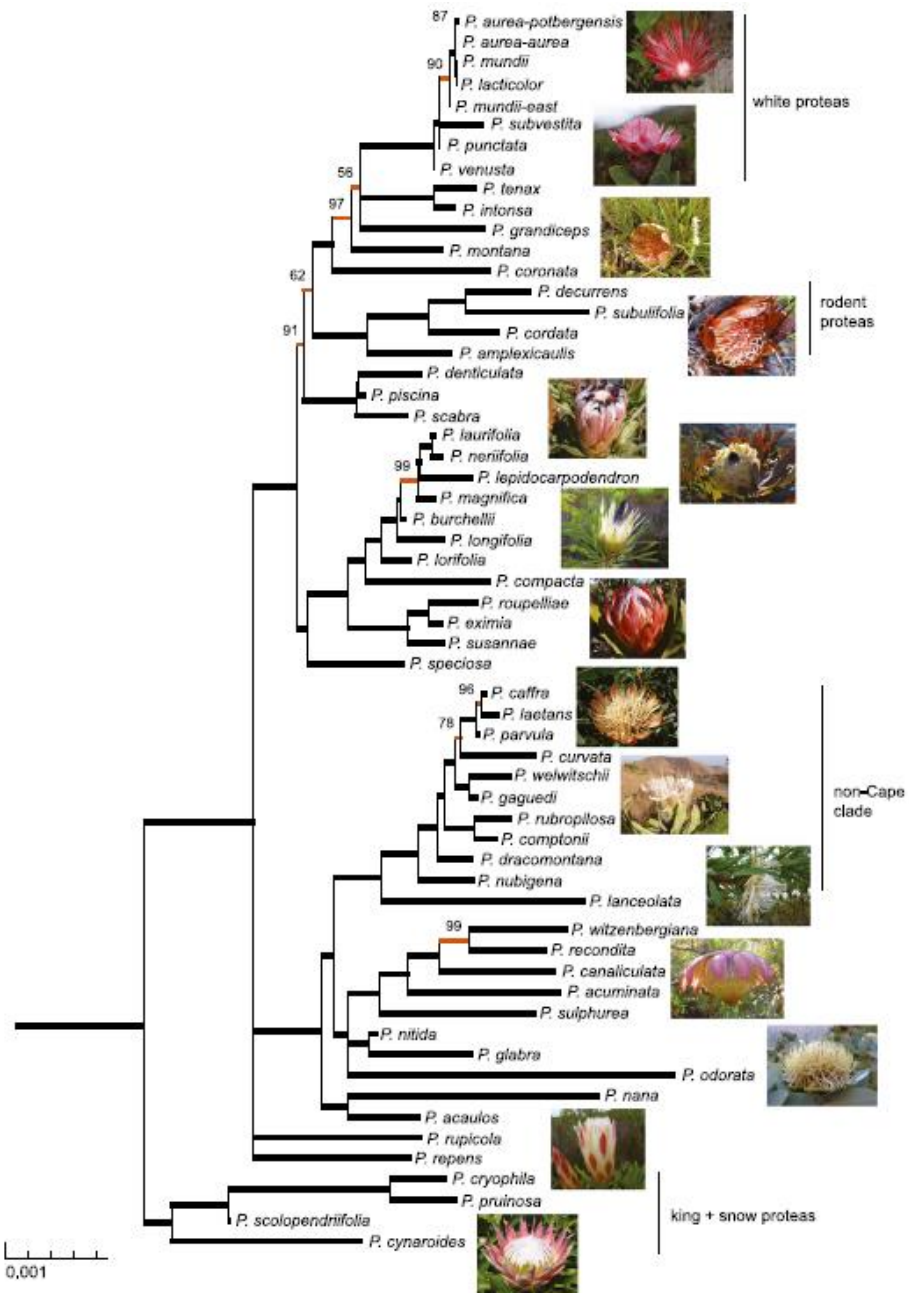
Protea repens   *Protea sulphurea*   *Protea susannae*

87 P. aurea-potbergensis
P. aurea-aurea
90 P. mundii
P. lacticolor
P. mundii-east
P. subvestita
P. punctata
P. venusta
56 P. tenax
97 P. intonsa
P. grandiceps
P. montana
P. coronata
62 P. decurrens
P. subulifolia
91 P. cordata
P. amplexicaulis
P. denticulata
P. piscina
P. scabra
P. laurifolia
P. neriifolia
99 P. lepidocarpodendron
P. magnifica
P. burchellii
P. longifolia
P. lorifolia
P. compacta
P. roupelliae
P. eximia
P. susannae
P. speciosa
96 P. caffra
P. laetans
78 P. parvula
P. curvata
P. welwitschii
P. gaguedi
P. rubropilosa
P. comptonii
P. dracomontana
P. nubigena
P. lanceolata
99 P. witzenbergiana
P. recondita
P. canaliculata
P. acuminata
P. sulphurea
P. nitida
P. glabra
P. odorata
P. nana
P. acaulos
P. rupicola
P. repens
P. cryophila
P. pruinosa
P. scolopendriifolia
P. cynaroides

white proteas

rodent proteas

non-Cape clade

king + snow proteas

0,001

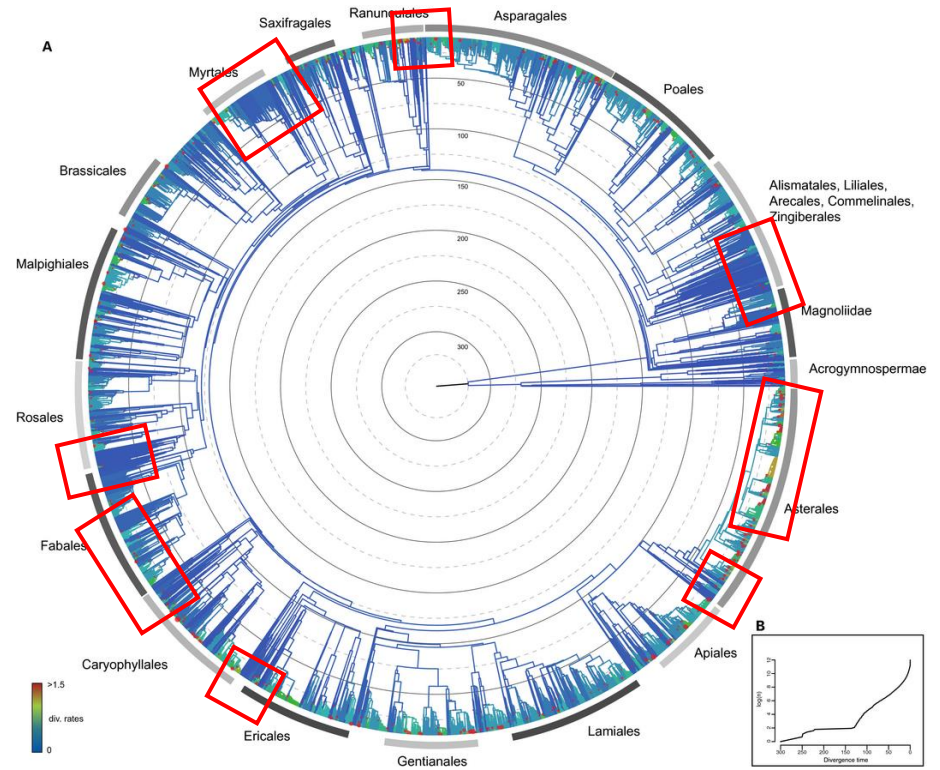Mitchell et al. (2017) *The American Journal of Botany*
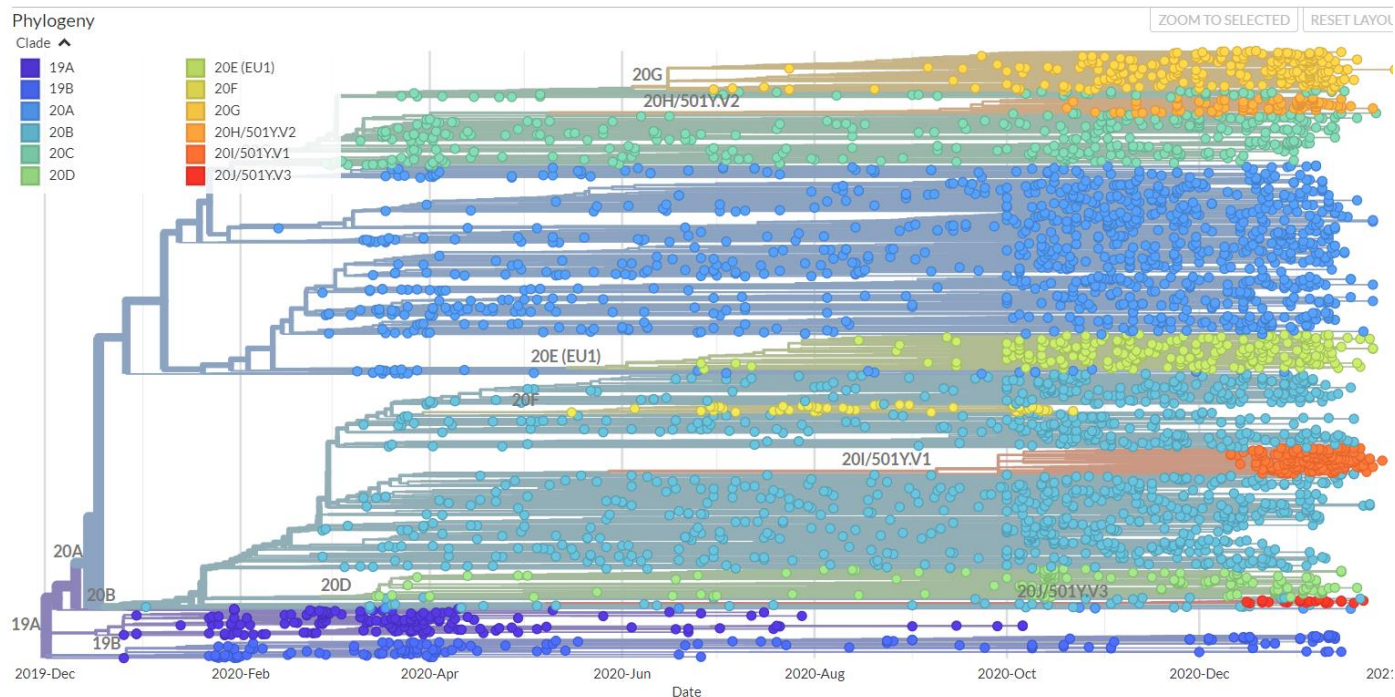
Plant Diversity

Byng et al. (2018)

Smith and Brown (2018)

How do we use phylogenies?

COVID-19 origin and variants

Where did SARS-CoV-2 originate? Where did different strains originate, and where have they spread

nextstrain.org/sars-cov-2

Hadfield et al. 2018. Bioinformatics 34: 23.

What other questions can we ask/answer using evolutionary trees?