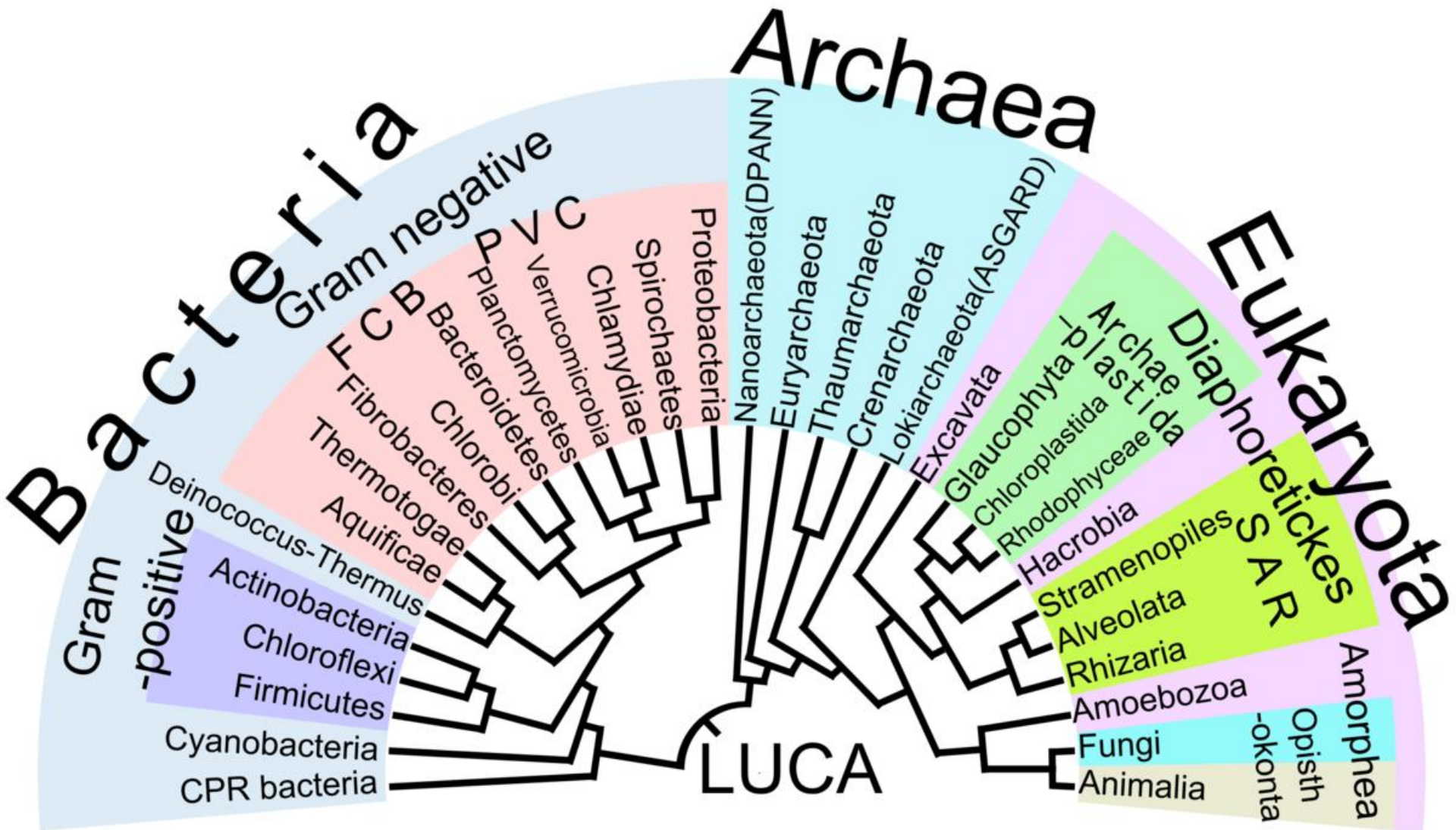


Building “trees” of South African shrubs



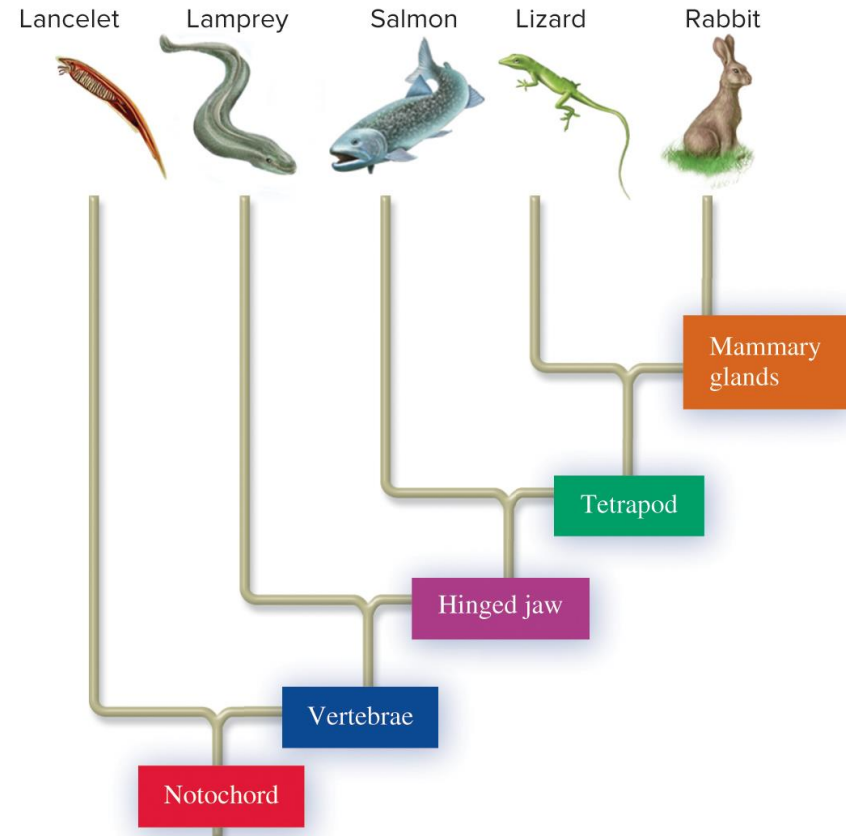


Building a phylogeny

- **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

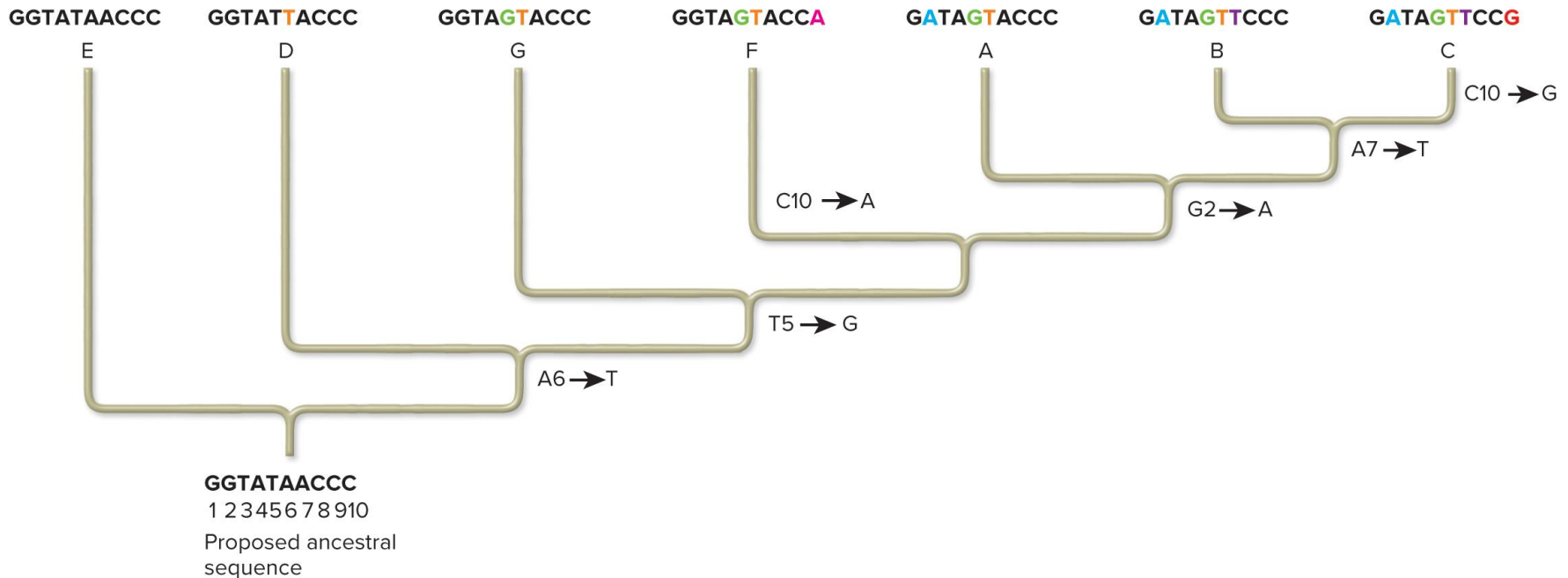
(a) Characteristics among species



(b) Cladogram based on morphological traits

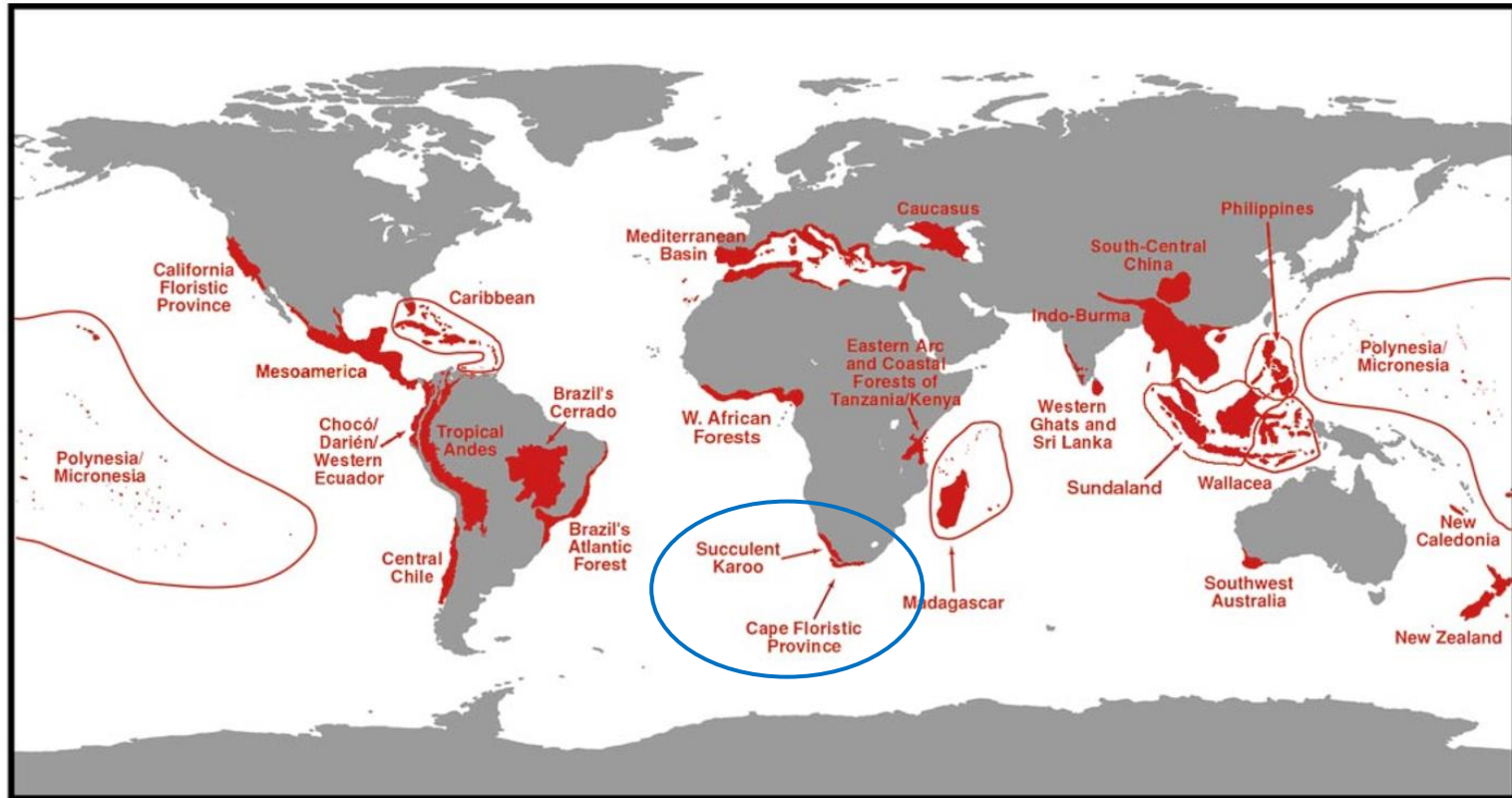
Building a phylogeny

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



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



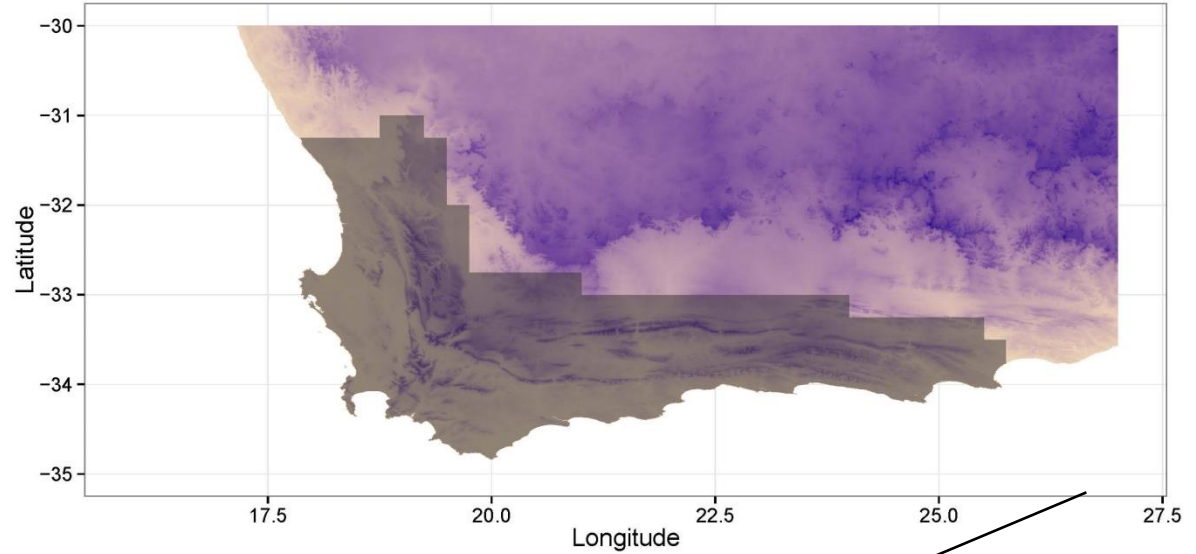
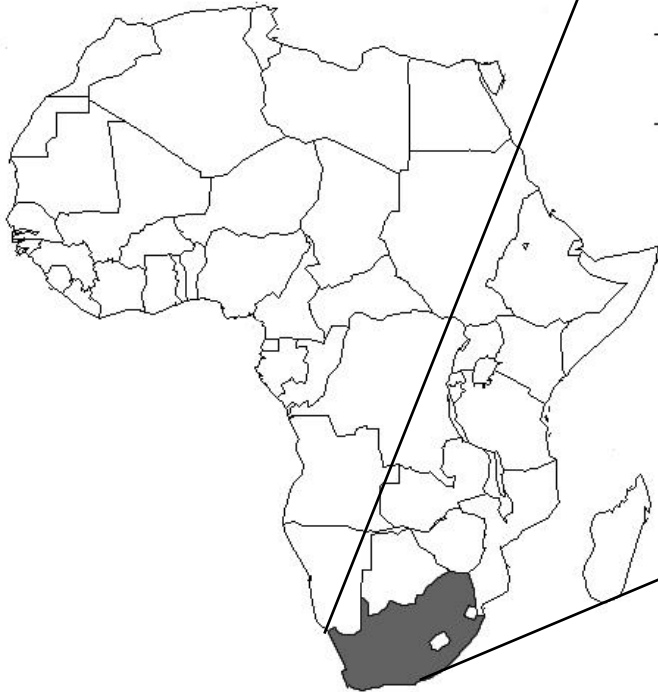
South Africa

> 35 languages spoken (11 official)



Cape Floristic Region

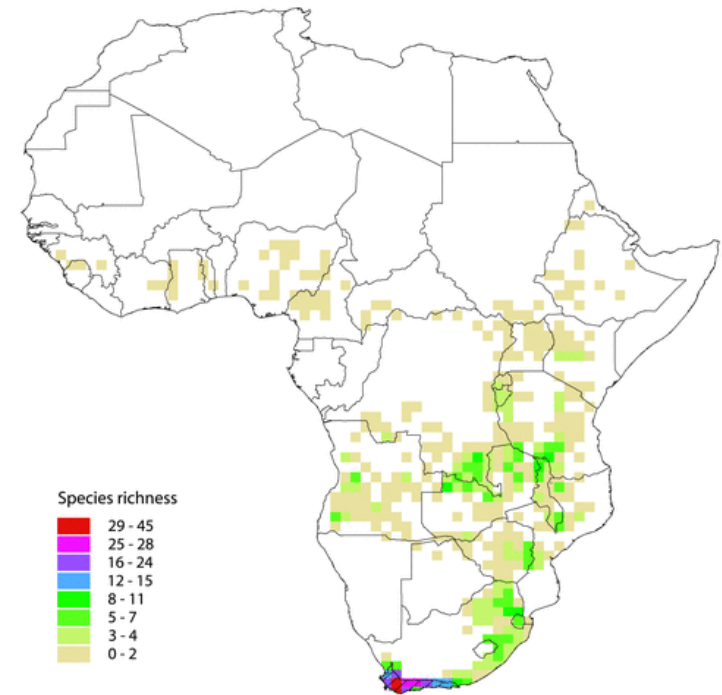
> 9000 plant species
~ 70% endemic





Protea

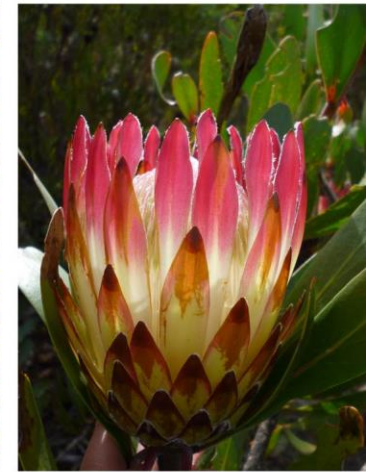
- Approximately 112 species
- Mostly shrubs
- 60% found in CFR



Protea



Protea



Protea

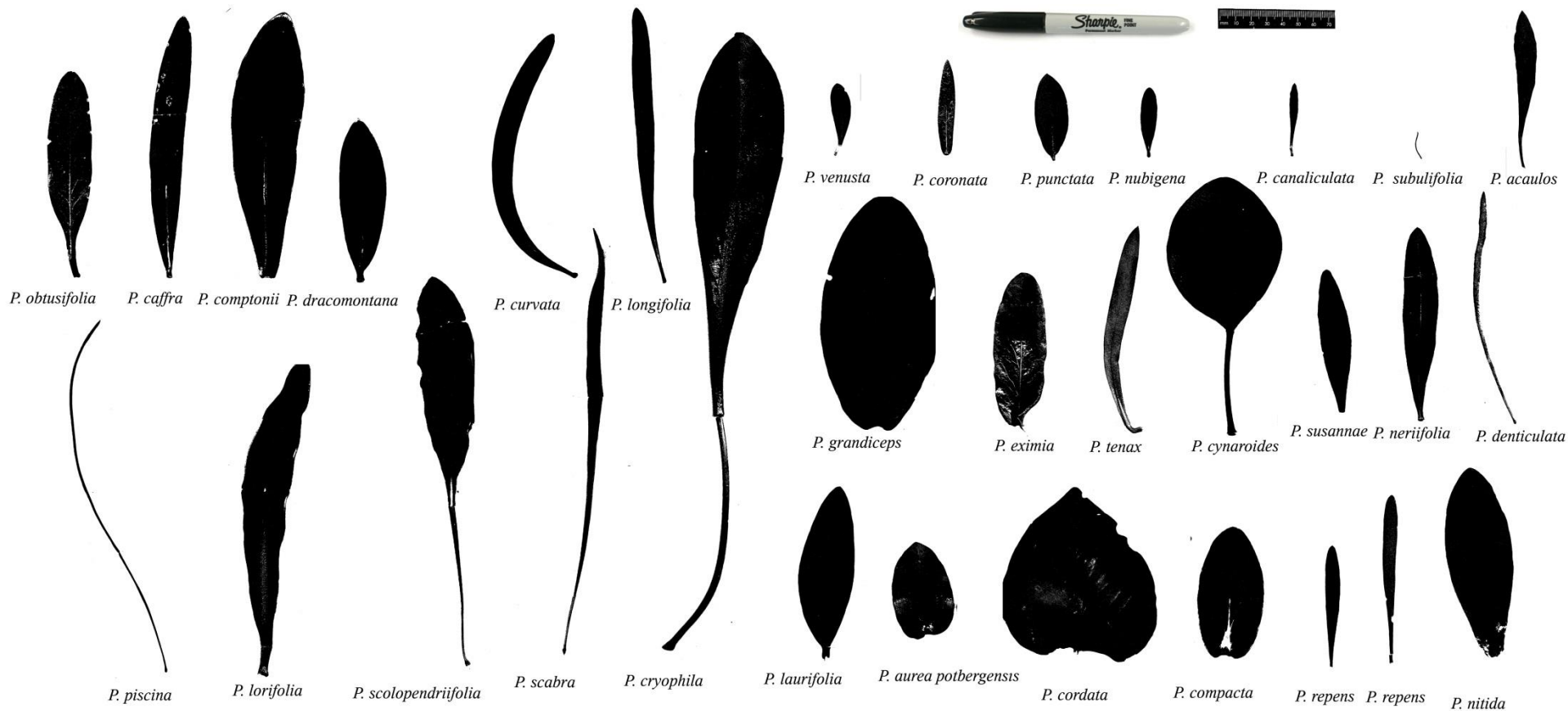
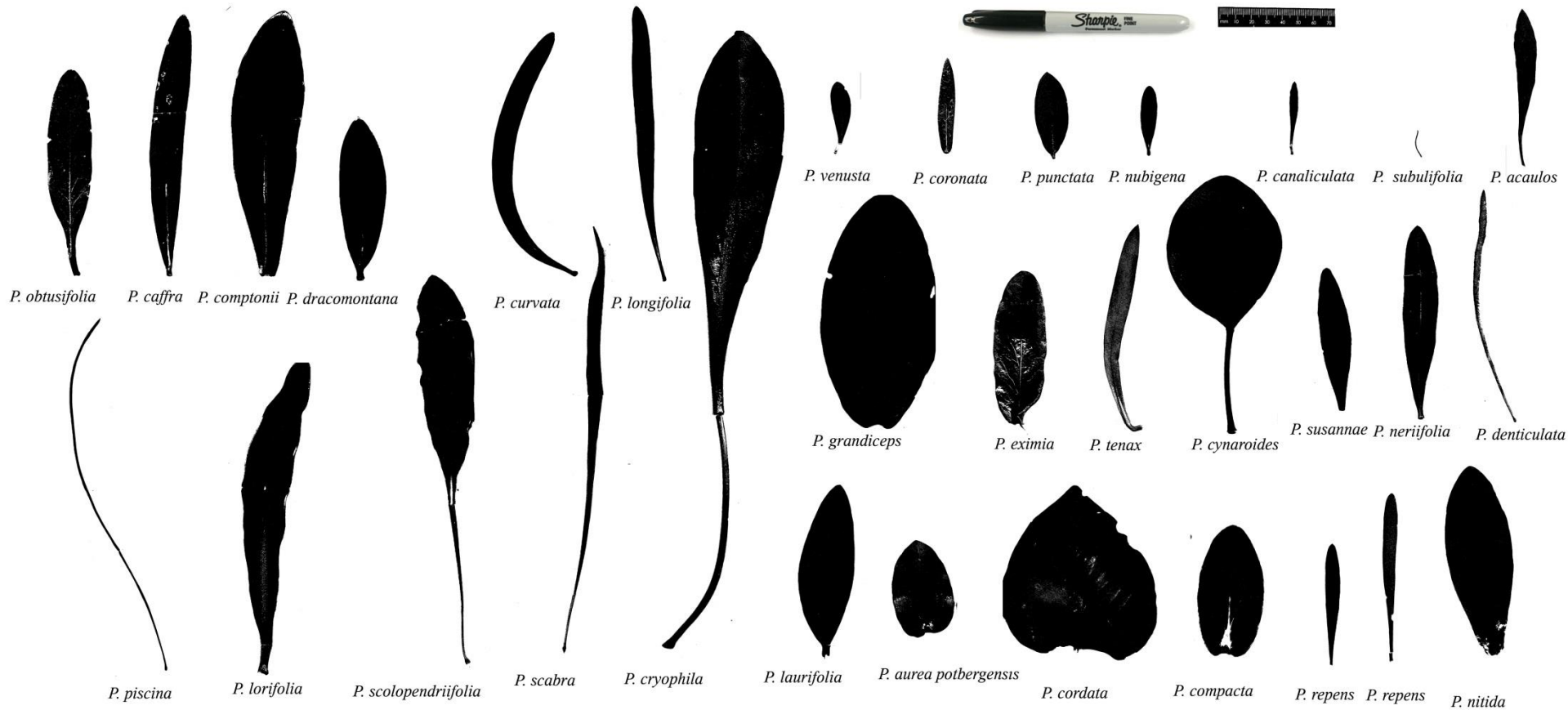


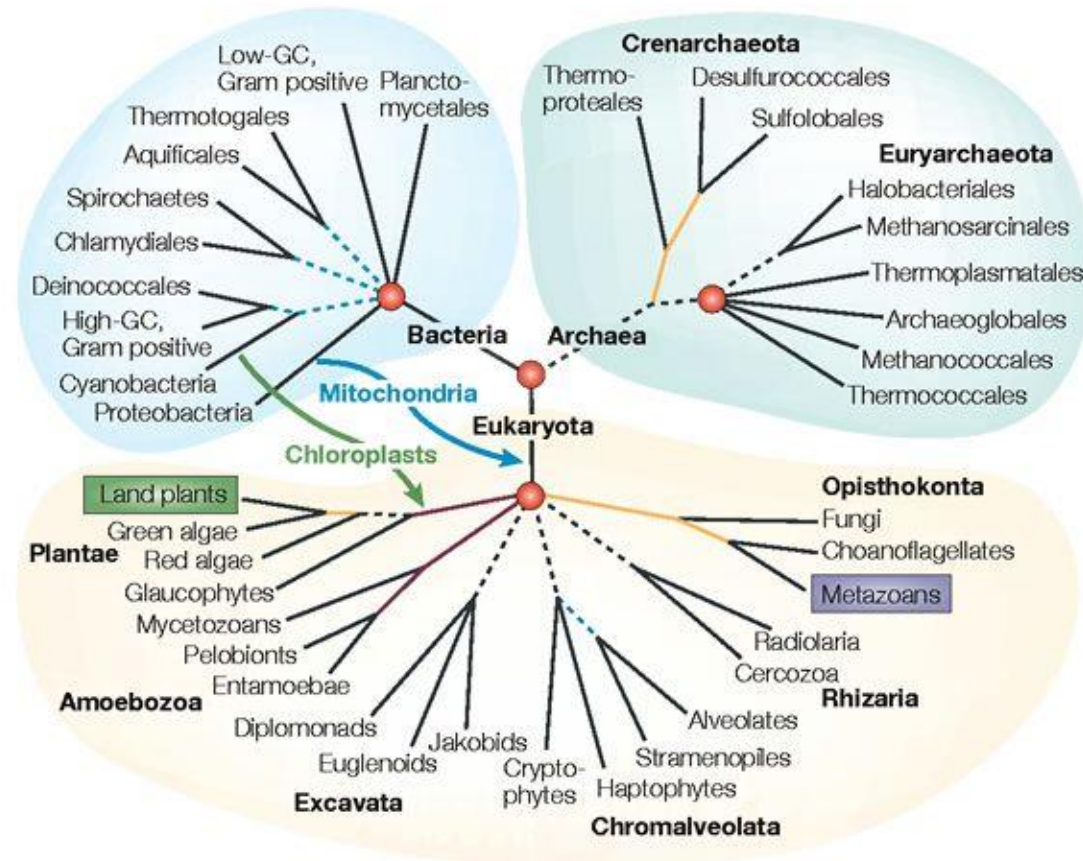
Photo Credit: Jane Carlson

Major question: How are species related to each other?



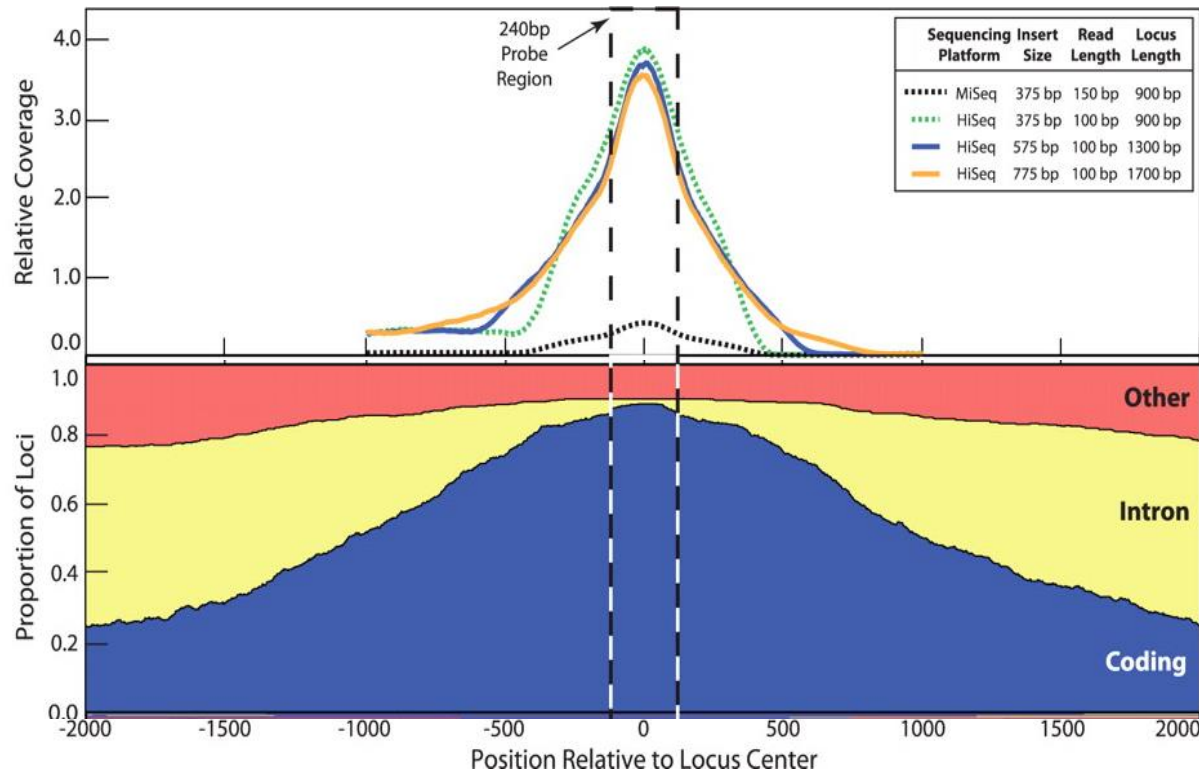
Phylogenomics

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

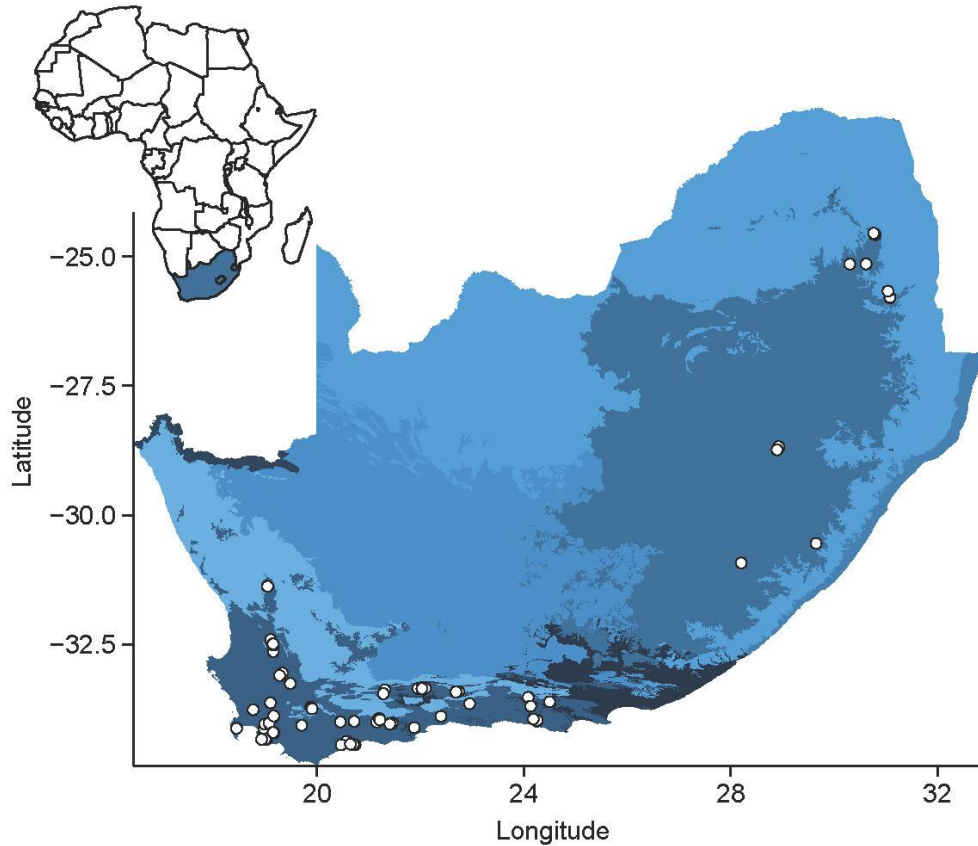


Anchored Phylogenomics

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



Protea Data Collection



- 164 DNA samples
- 59 *Protea* species
(+ 6 outgroup)
- 498 loci captured (average length 551bp)
- About 275,000bp each

That's a LOT of data...

- HPC enables us to analyze lots of data using specific software!
- We will be using RAxML (**R**andomized **A**xelerated **M**aximum **L**ikelihood) to build phylogenies
- TreeGraph2 to look at the phylogenies that we build
- Goals: build evolutionary trees for subset of species, explore different approaches

amkozlov/**raxml-ng**

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



Tree
Graph**2**

Let's use the cluster to build phylogenies!

- 3 different “scripts” to run
- Follow along with me as we use WinSCP, Putty, etc. to use the cluster 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

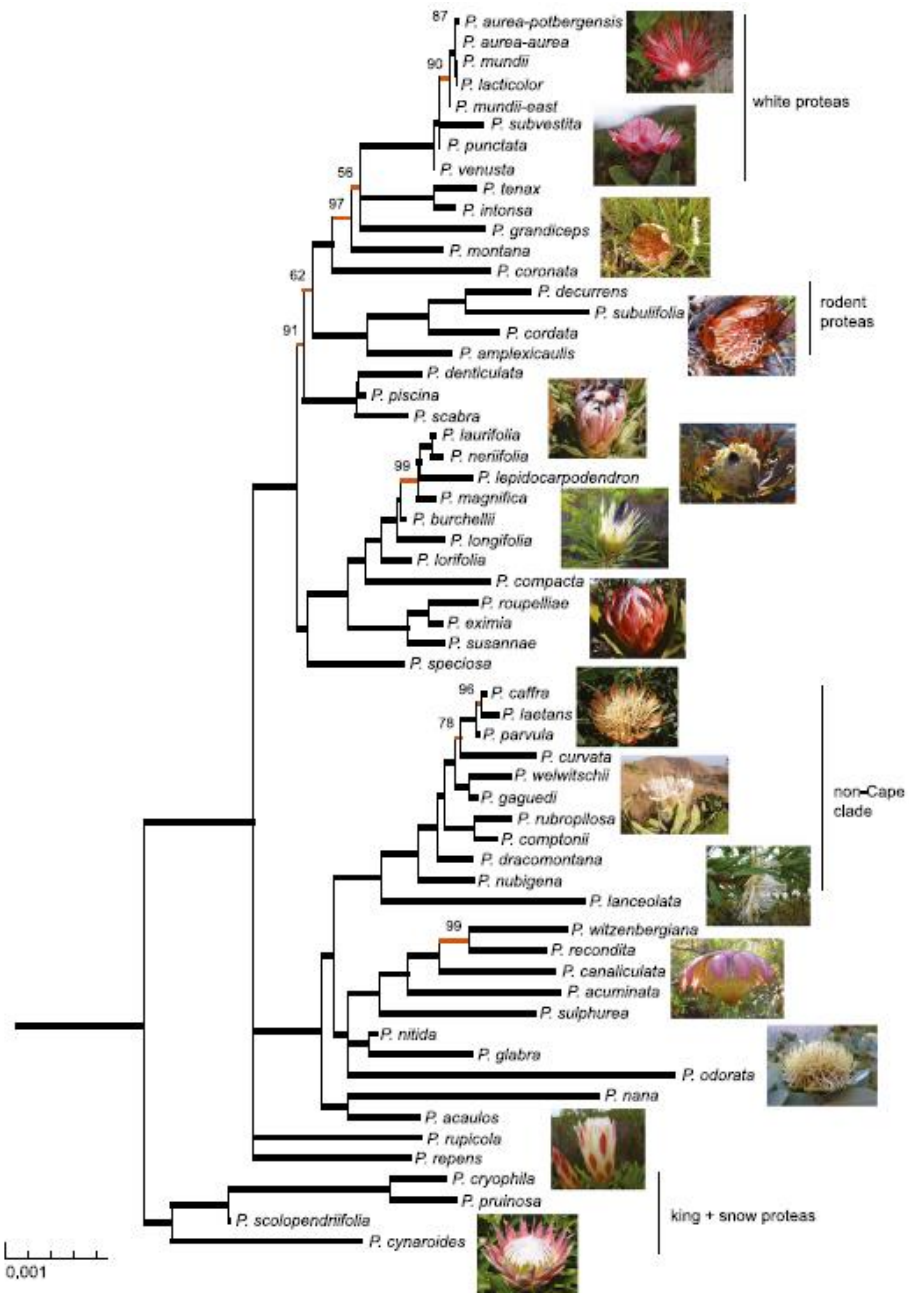


Protea sulphurea



Protea susannae

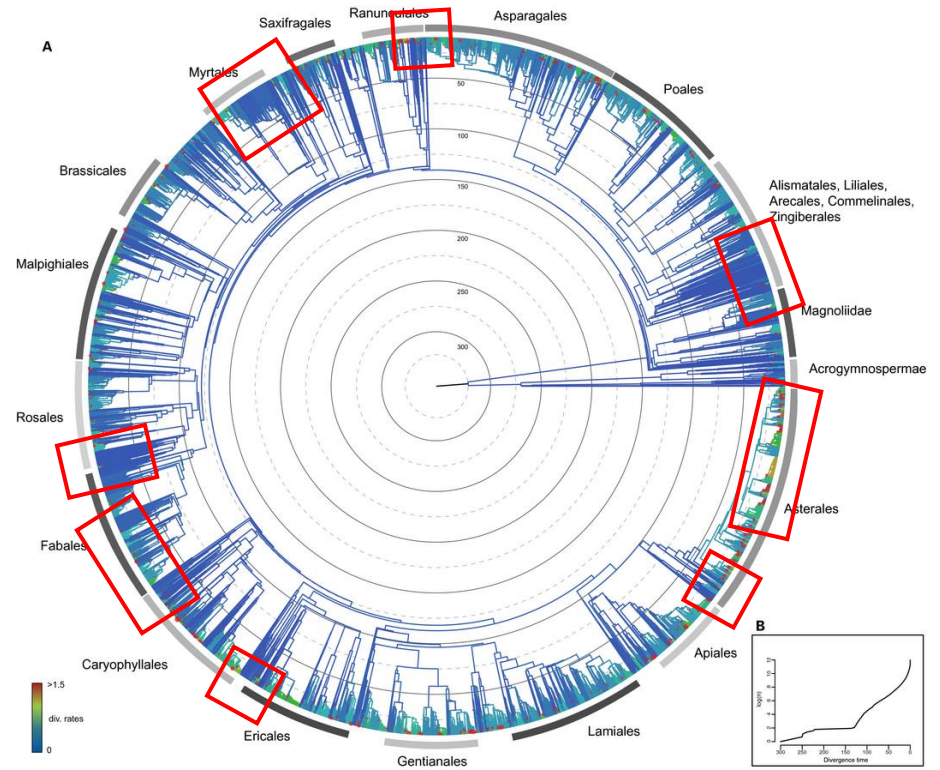




Plant Diversity



Byng et al. (2018)



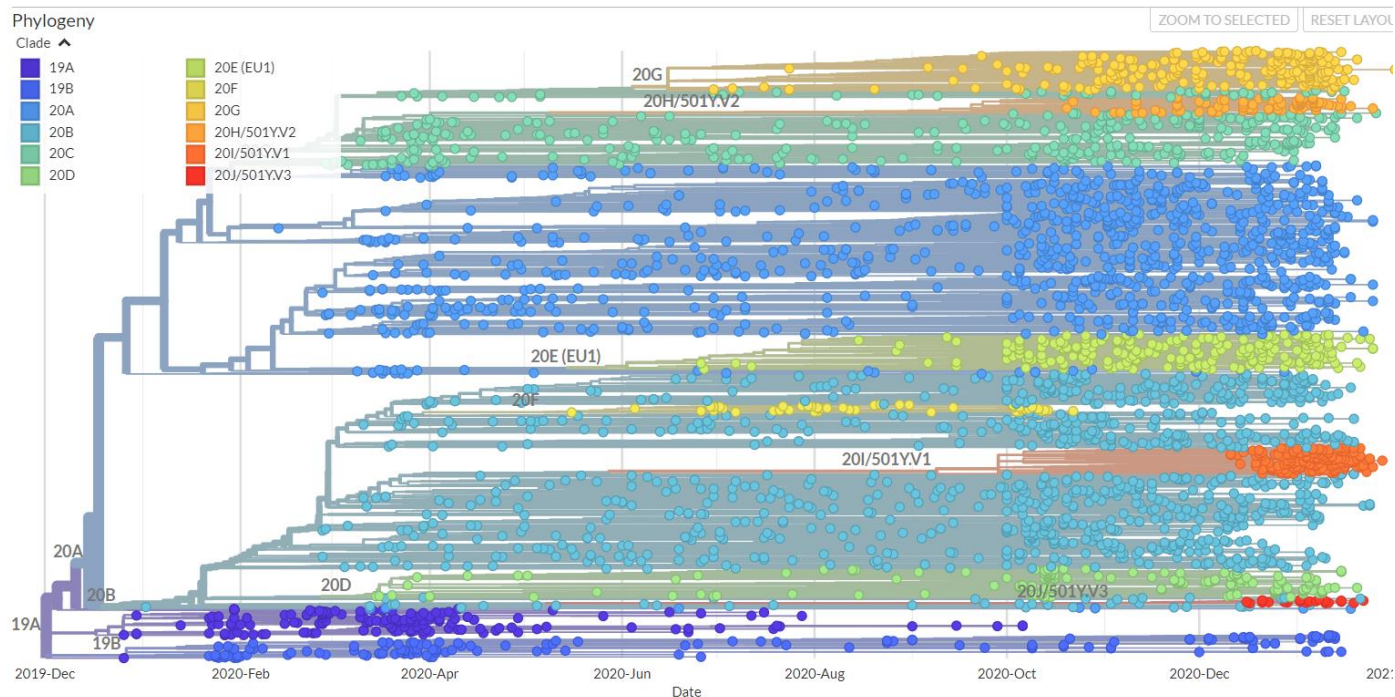
Smith and Brown (2018)

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



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