

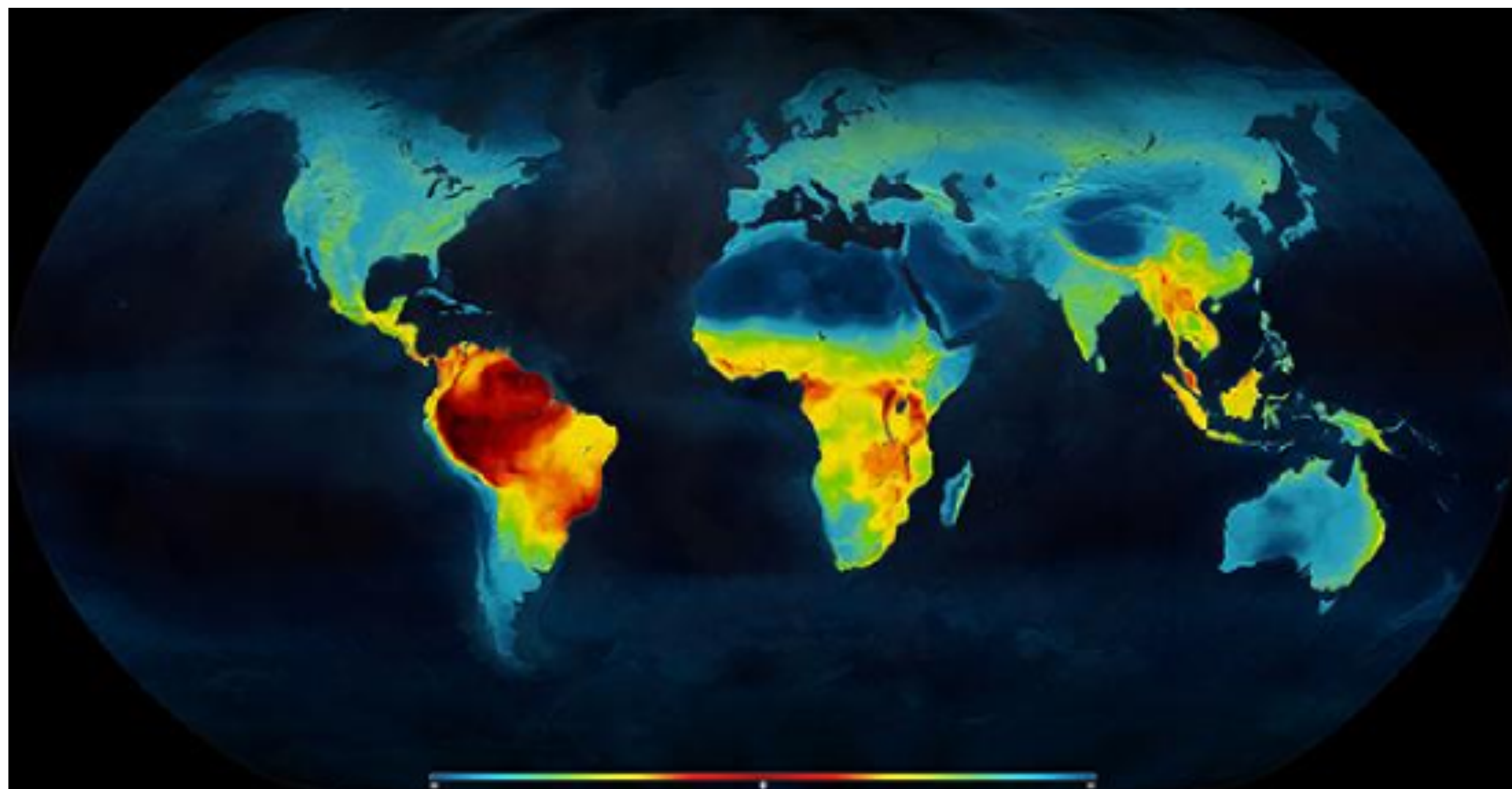
BIOL 5480

Dealing with Data

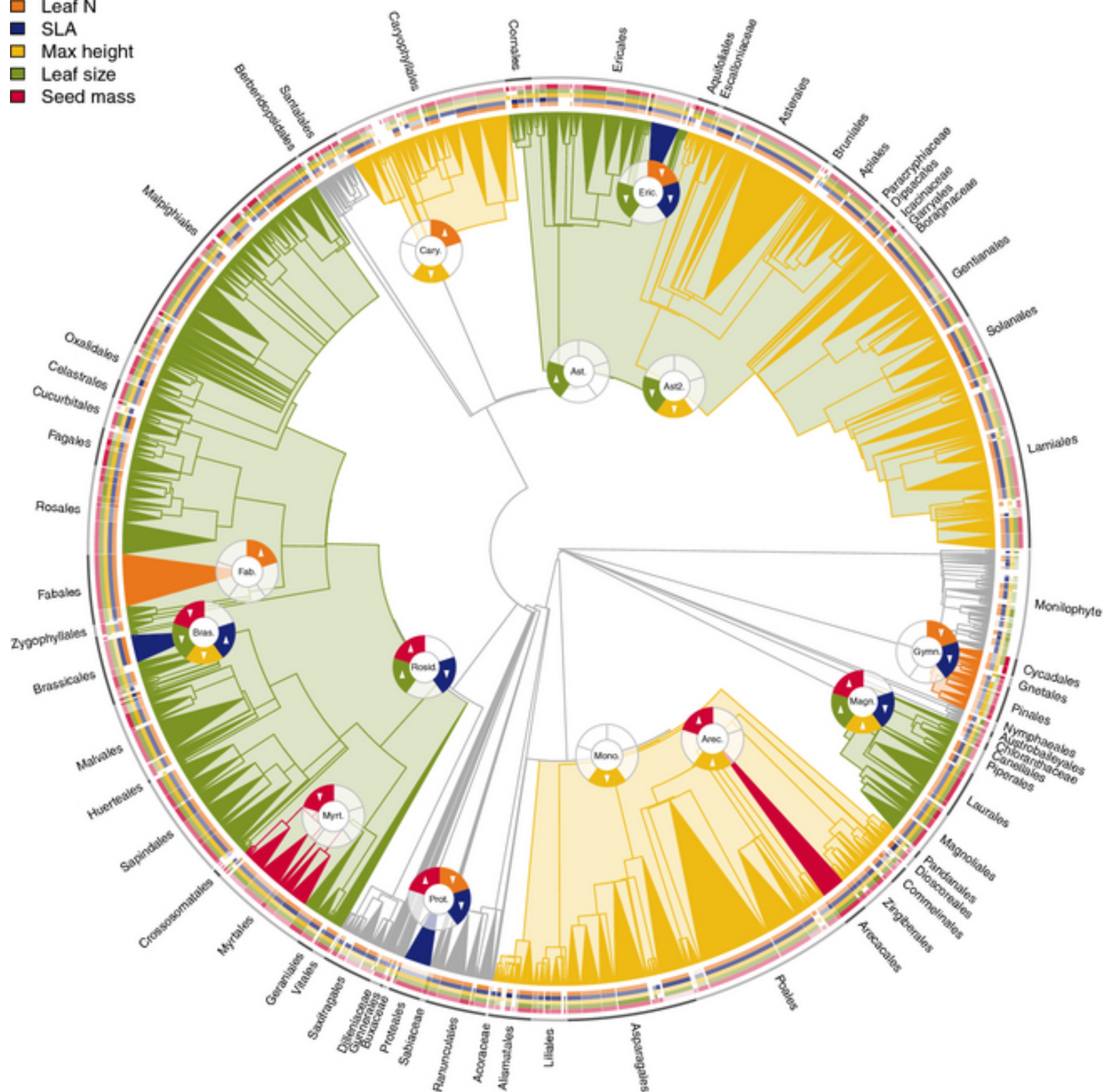
Matt Pennell – UBC Zoology

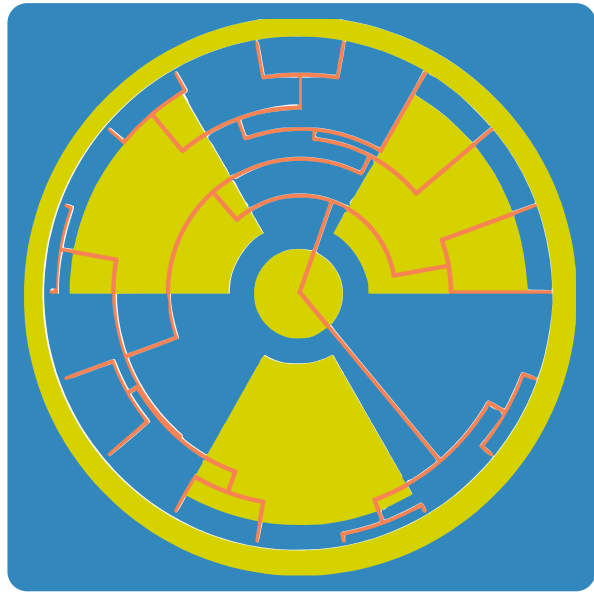
[pennell@zoology.ubc.ca](mailto:pennell@zoology.ubc.ca)

Biodiversity Research Centre 208



- Leaf N
- SLA
- Max height
- Leaf size
- Seed mass





**geiger** — fit macroevolutionary models in R

<https://github.com/mwpennell/geiger-v2>



**arbutus** — goodness-of-fit for phylo models

<https://github.com/mwpennell/arbutus>

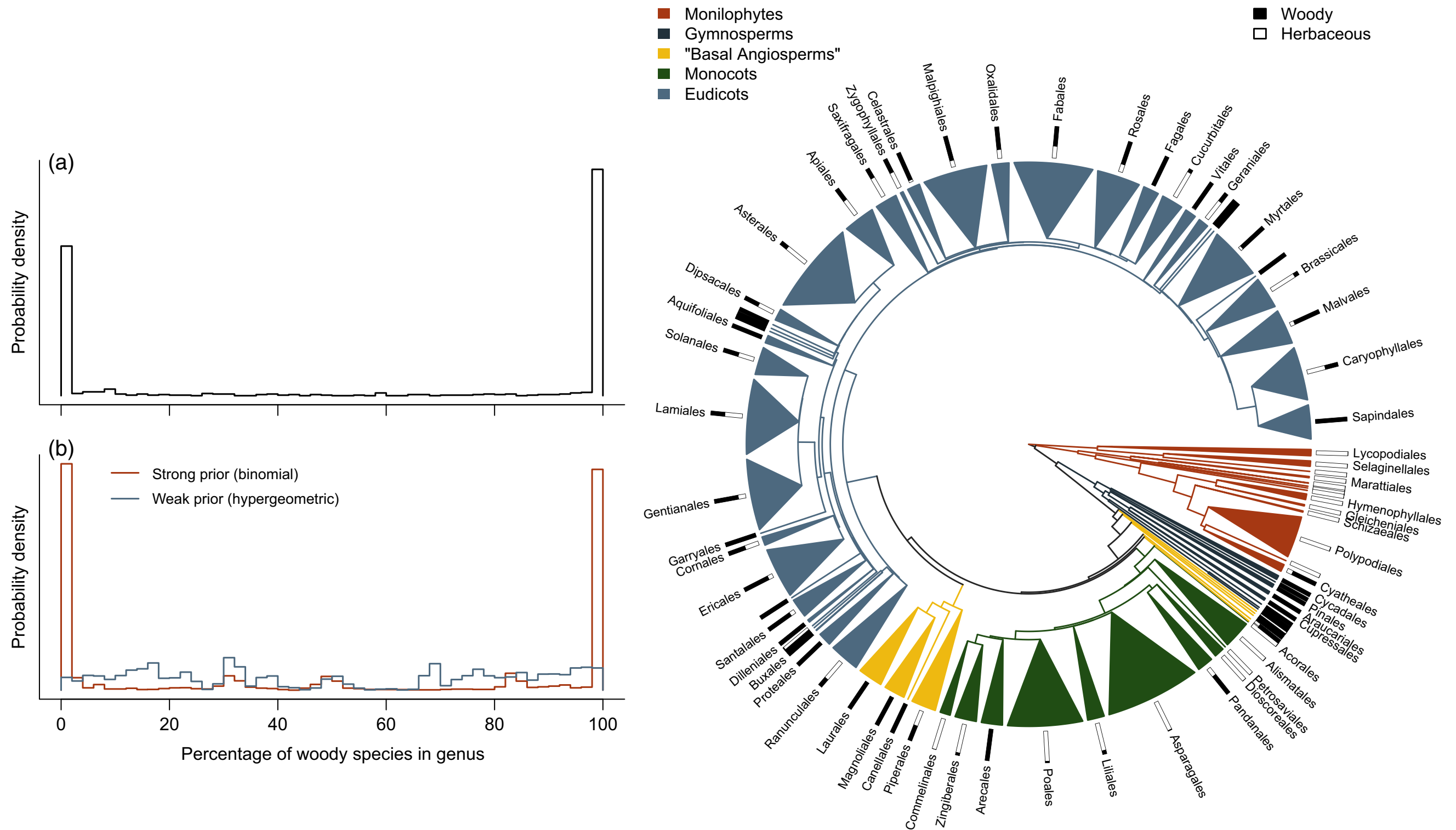


**chromer** — access data on chromosome counts

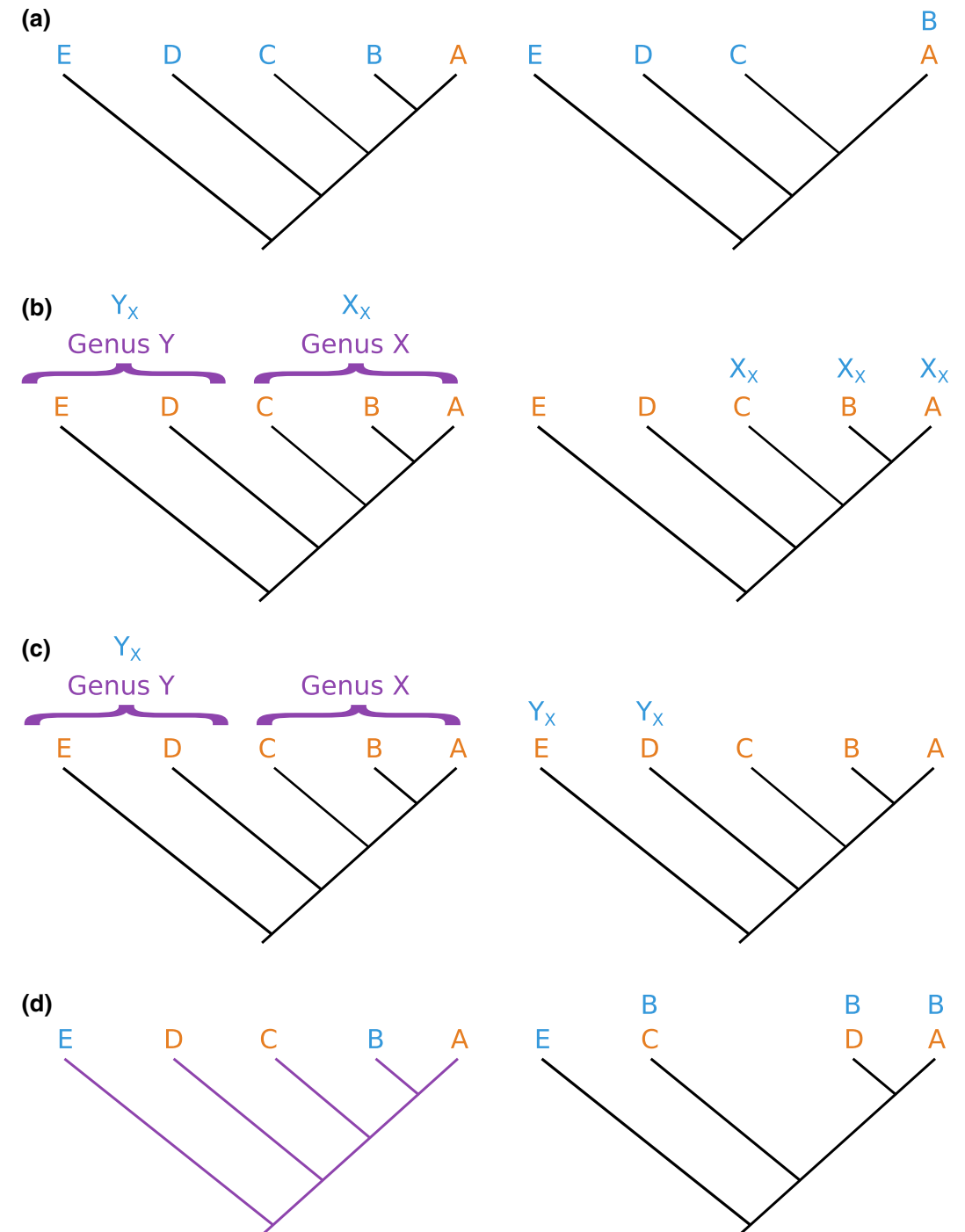
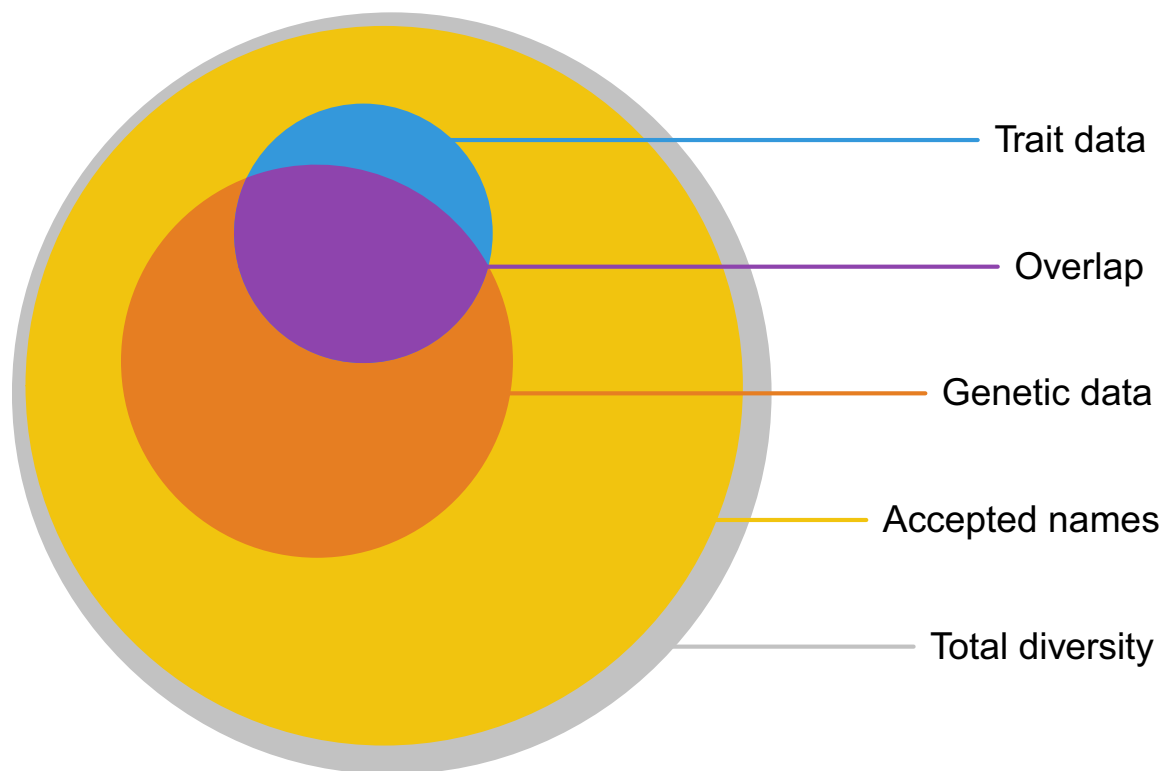
<https://github.com/ropensci/chromer>



# Sampling bias in trait databases



# Matching traits with phylogeny



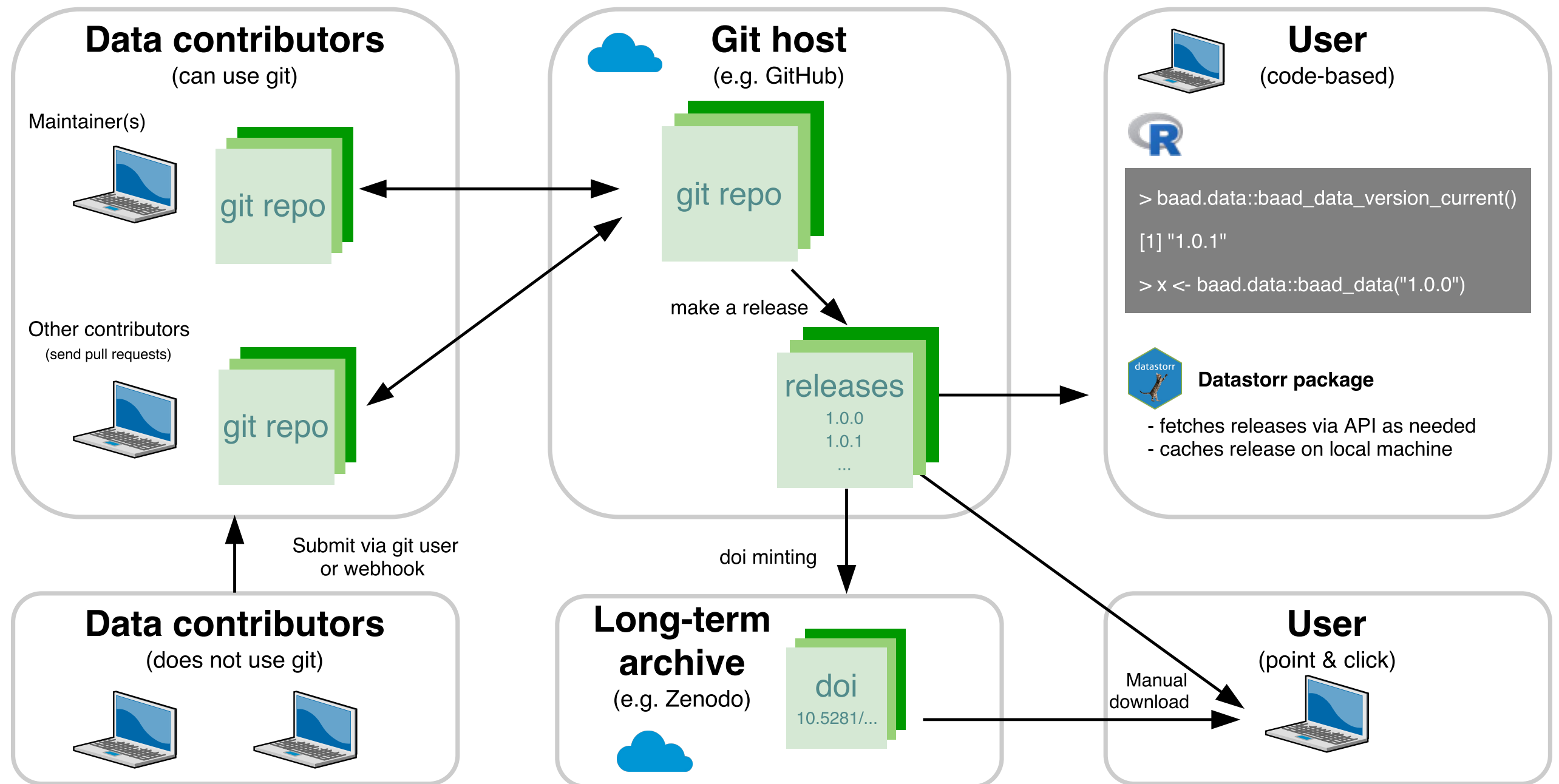
# Taxonomy — the hardest problem in biodiversity science

```
> library(taxonlookup)
> taxon <- plant_lookup()
> head(taxon)
```

	genus	family	order	group
1	Acorus	Acoraceae	Acorales	Angiosperms
2	Albidella	Alismataceae	Alismatales	Angiosperms
3	Alisma	Alismataceae	Alismatales	Angiosperms
4	Astonia	Alismataceae	Alismatales	Angiosperms
5	Baldellia	Alismataceae	Alismatales	Angiosperms
6	Burnatia	Alismataceae	Alismatales	Angiosperms



# Managing “living datasets”





## Course topics:

Version control

Project organization / management

Principles of tidy data

Combining and separating datasets

Automated workflows

Reproducible analyses

Data sharing how-tos

The importance of metadata

## Not course topics:

Statistics

Advanced R coding

Plotting

Making R packages

Open Science / Open Data / Open Access