



# R-based Data Downloads

Shelly Gaynor  
University of Florida



# Github: soltislab/BotanyENMWorkshop/

The screenshot shows the GitHub interface for the repository **soltislab / BotanyENMWorkshops**. The repository is public and has 10 stars, 6 forks, and 2 unwatchers. The main branch is **main**, with 5 branches and 2 tags. The repository description is "Using Digitized Collections-Based Data in Research: Applications for Ecology, Phylogenetics, and Biogeography Botany 2025." The repository includes a README.md file, which is highlighted with a "minor edit" link. The repository also has a .gitattributes file and a .gitignore file. The repository is tagged with "r-markdown" and "ecological-niche-modelling".

**Repository Details:**

- Repository: **BotanyENMWorkshops** (Public)
- Stars: 10
- Forks: 6
- Unwatch: 2
- Edit Pins: 1

**Files and Commits:**

File	Commit Message	Commit Time
Botany2025	Adding updated RMD and HTML	2 days ago
.gitattributes	tracking asc files with lfs	3 weeks ago
.gitignore	remove and ignore .Rproj.user	3 weeks ago
README.md	<a href="#">minor edit</a>	2 days ago

**About:** Using Digitized Collections-Based Data in Research: Applications for Ecology, Phylogenetics, and Biogeography Botany 2025.

[soltislab.github.io/BotanyENMWorksh...](https://soltislab.github.io/BotanyENMWorksh...)

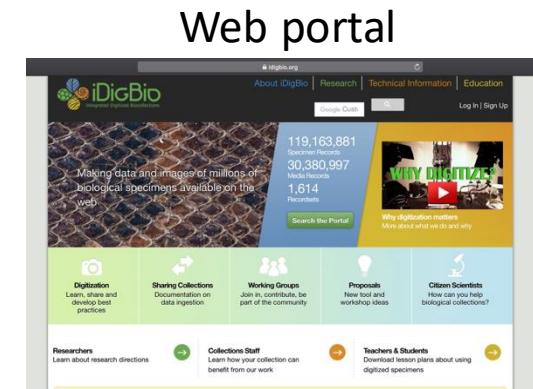
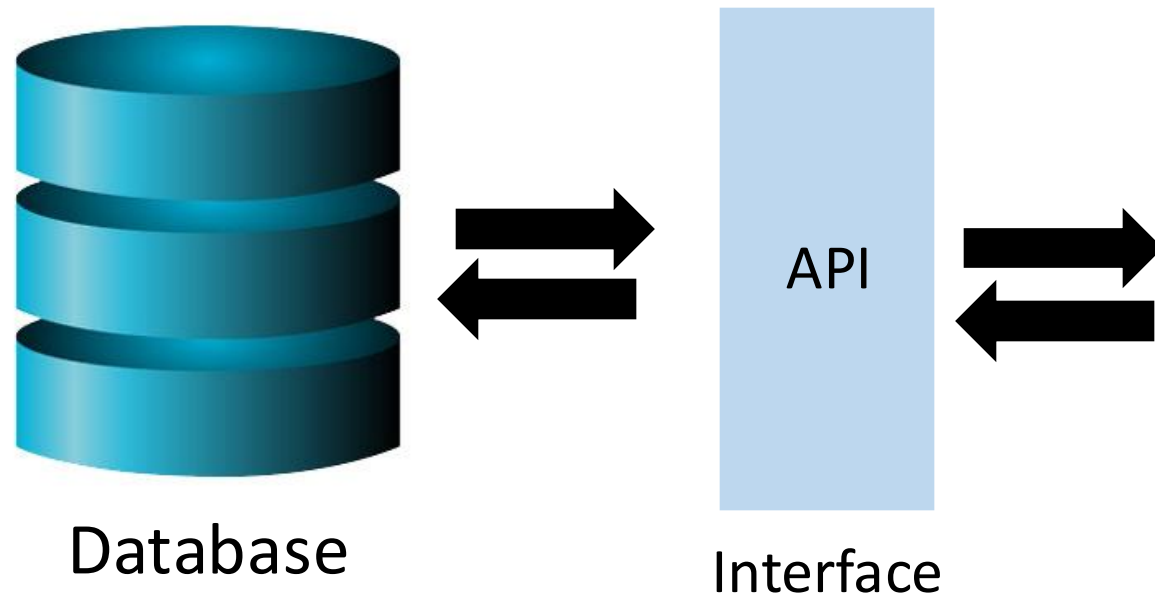
[r-markdown](#) [ecological-niche-modelling](#)


[Readme](#)

See the code: <https://tinyurl.com/bworksho>

# API = Application Programming Interface

- Allows users to interact with a system





Making data and images of millions of biological specimens available on the web

121,428,342

Specimen Records

31,871,262

Media Records

1,621

Recordsets

[Search the Portal](#)

**Why digitization matters**  
More about what we do and why

**Digitization**

Learn, share and develop best practices

**Sharing Collections**

Documentation on data ingestion

**Working Groups**

Join in, contribute, be part of the community

**Proposals**

New tool and workshop ideas

**Citizen Scientists**

How can you help biological collections?

**Researchers**

Learn about research directions

**Collections Staff**

Learn how your collection can benefit from our work

**Teachers & Students**

Download lesson plans about using digitized specimens



# iDigBio API

- Multiple ways to access the API:

API Name	Info
Search API	ridigbio R package <100,000 records
Download API	>100,000 records
Record API	Single record
Media API	Single record



[biodiversity-specimen-data/specimen-data-use-case](https://github.com/biodiversity-specimen-data/specimen-data-use-case)

# iDigBio API

- Multiple ways to access the API:

API Name	Info
Search API	ridigbio R package <100,000 records
Download API	>100,000 records
Record API	Single record
Media API	Single record



[biodiversity-specimen-data/specimen-data-use-case](https://github.com/biodiversity-specimen-data/specimen-data-use-case)

# GBIF API

- Multiple ways to access the API:

API Name	Info
Registry API	Create, edit, update and search for information about datasets
Species API	Taxonomy API
Occurrence API	Record API
Maps API	Show maps of GBIF
News API	Search papers published using GBIF



# Load Packages

```
library(ridigbio)  
library(gatoRs)  
library(leaflet)
```





# Downloading data using ridigbio

- First, we are searching for the species *Galax urceolata*
- Next, download occurrence records for the family Diapensiaceae

Search for the species *Galax urceolata*.

```
iDigBio_GU <- idig_search_records(rq=list(scientificname="Galax  
urceolata"))
```

Search for the family Diapensiaceae.

```
iDigBio_GU_family <- idig_search_records(rq=list(family="Diapen  
siaceae"), limit=1000)
```



# Records only in North America

```
rq_input <- list("scientificname"=list("type"="exists"),  
               "family"="Diapensiaceae",  
               geopoint=list(  
                 type="geo_bounding_box",  
                 top_left=list(lon = -98.16, lat = 48.92),  
                 bottom_right=list(lon = -64.02, lat = 23.06)  
               )  
             )
```

Search using the input you just made

```
iDigBio_GU_family_USA <- idig_search_records(rq_input, limit=100)
```

## Search Records

[Help](#)[Reset](#)☐ Must have media☐ Must have map point

Filters

Mapping

Sorting

Download

Lat/Lon Bounds

[Clear](#)☒ Rectangle ☐ Circle

Lat:

Lon:

NorthWest

48.92

-98.16

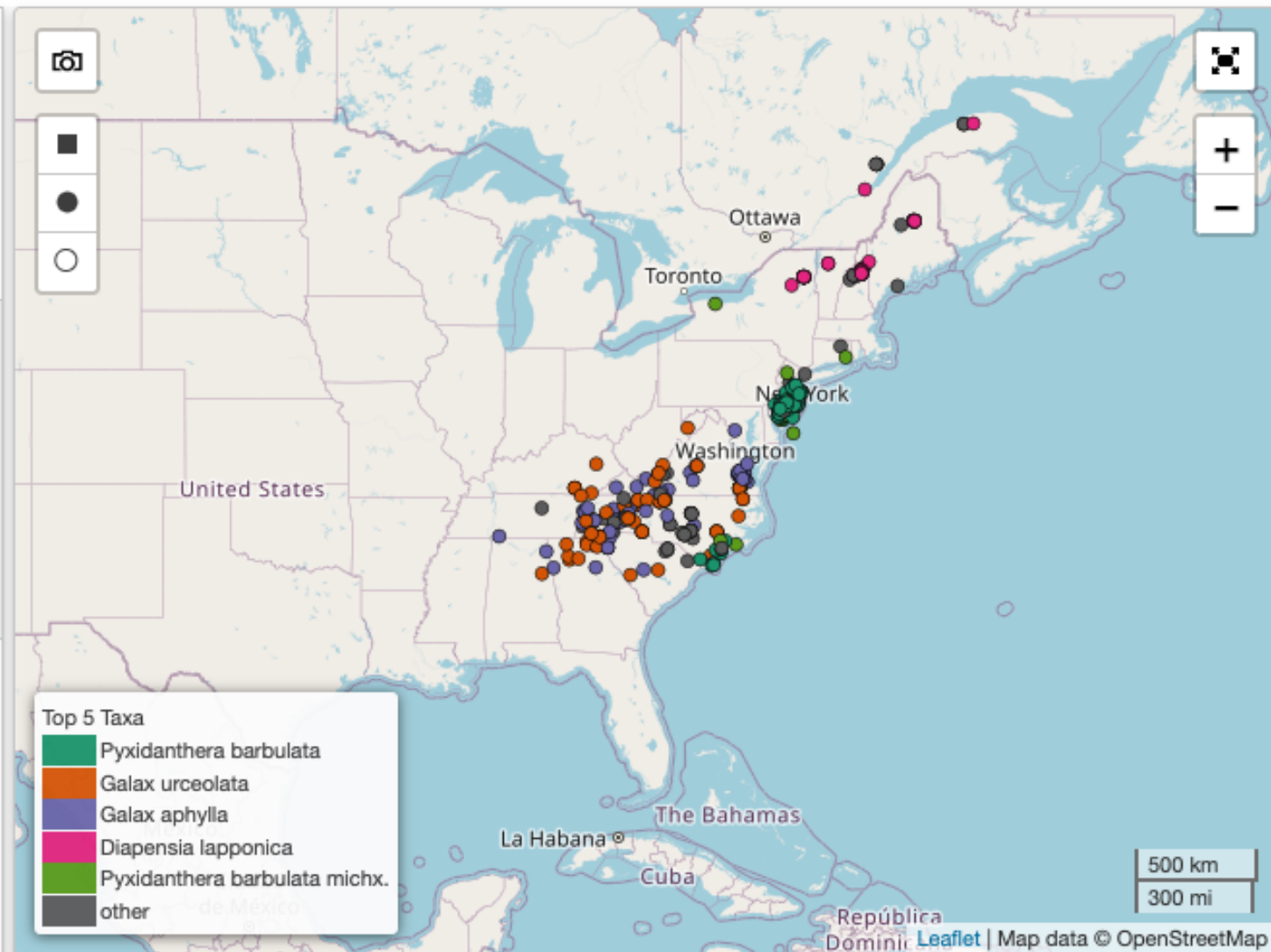
Lat:

Lon:

SouthEast

23.06

-64.02



List

Labels

Media

Recordsets

Total: 483

# Save as csv

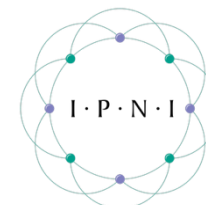
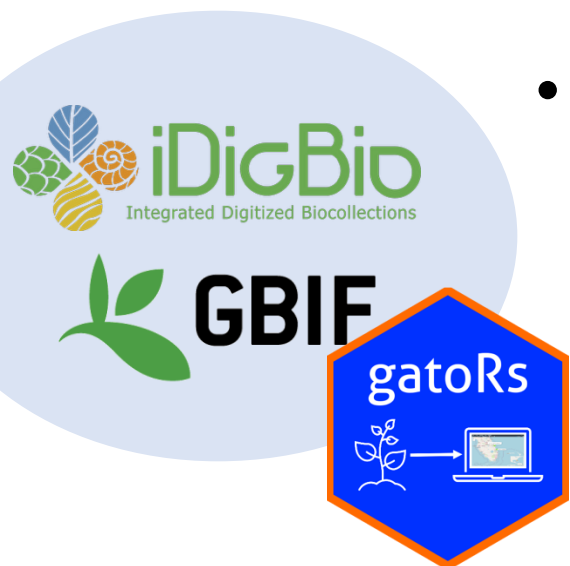
```
write.csv(iDigBio_GU,  
          file = "data/01_download/iDigBio_GU_2025_06_27.csv",  
          row.names = FALSE)  
write.csv(iDigBio_GU_family,  
          file = "data/01_download/iDigBio_GU_family_2025_06_27.csv",  
          row.names = FALSE)
```

# Data download using gatoRs

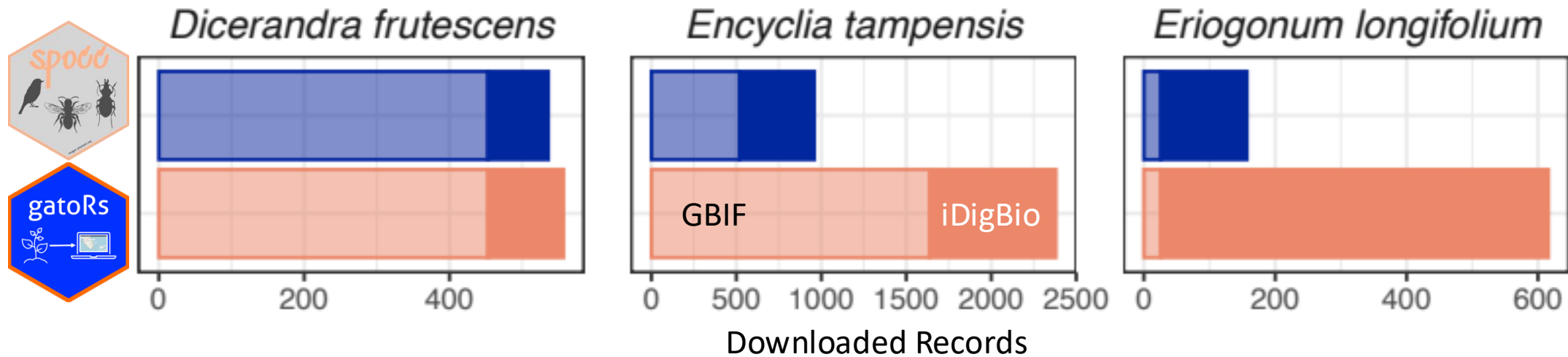


Natalie Patten

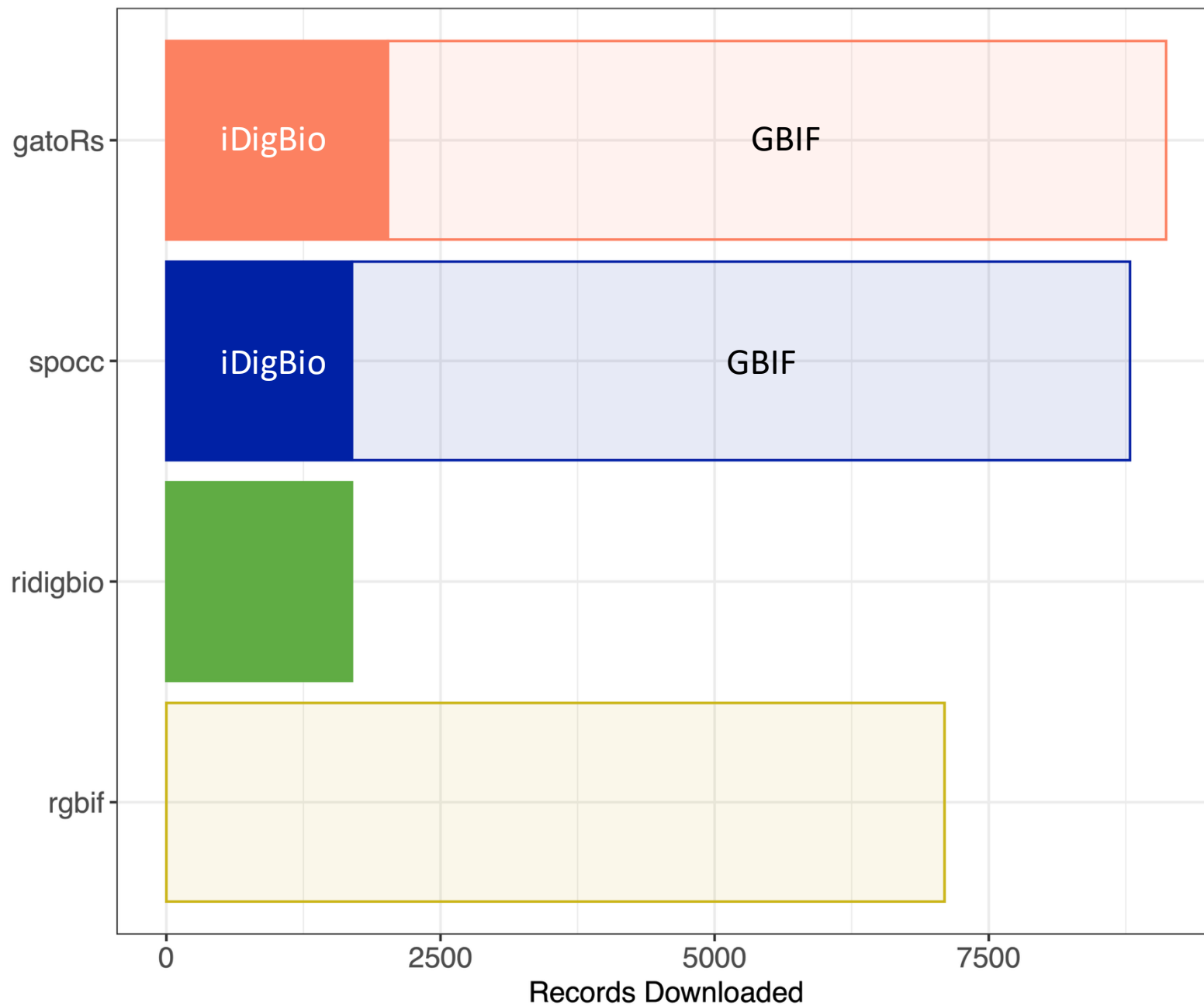
- To pull data from GBIF and iDigBio for a set of synonyms
- Identifying synonyms:
  - Taxonomic Name Resolution Service
    - <https://tnrs.biendata.org/>
    - Used in soltislab/BotanyENMWorkshops 2020
  - R package taxize
    - 20 sources for synonyms
    - <https://docs.ropensci.org/taxize/>



# Why should I use gatoRs?



Patten et al. 2024. Applications in Plant Science.



*Galax urceolata*





*Shortia galacifolia*



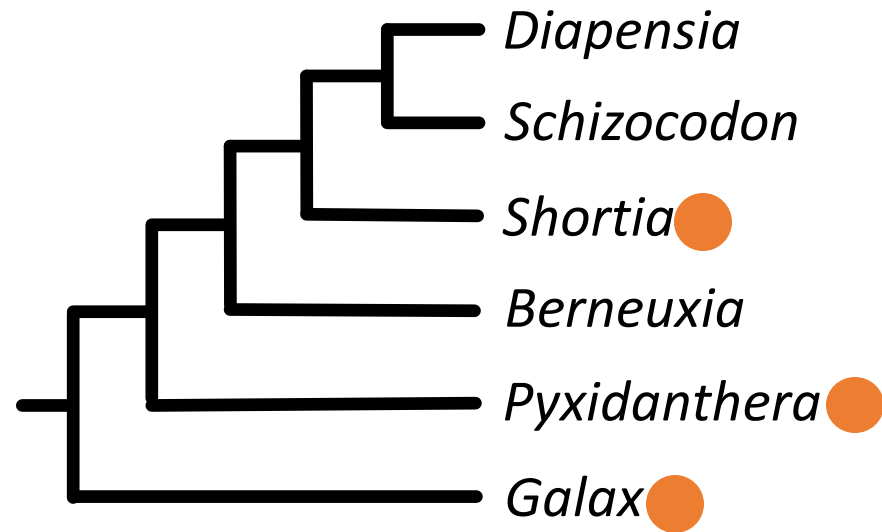
*Galax urceolata*



*Pyxidanthera barbulata*



*Pyxidanthera brevifolia*



Diapensiaceae  
Ericales

# Data download using gator\_download

Make synonym lists

Object

List of strings

```
Shortia_galacifolia <- c("Shortia galacifolia", "Sherwoodia galacifolia")
Galax_urceolata <- c("Galax urceolata", "Galax aphylla")
Pyxidanthera_barbulata <- c("Pyxidanthera barbulata", "Pyxidanthera barbulat  
a var. barbulata")
Pyxidanthera_brevifolia <- c("Pyxidanthera brevifolia", "Pyxidanthera barbu  
lata var. brevifolia")
```

# Data download using gators\_download

```
gators_download(synonyms.list = Shortia_galacifolia,  
                write.file = TRUE,  
                filename = "data/download/raw/Shortia_galacifolia_raw_20230605.csv")  
gators_download(synonyms.list = Galax_urceolata,  
                write.file = TRUE,  
                filename = "data/download/raw/Galax_urceolata_raw_20230605.csv")  
gators_download(synonyms.list = Pyxidanthera_barbulata,  
                write.file = TRUE,  
                filename = "data/download/raw/Pyxidanthera_barbulata_raw_20230605.csv")  
gators_download(synonyms.list = Pyxidanthera_brevifolia,  
                write.file = TRUE,  
                filename = "data/download/raw/Pyxidanthera_brevifolia_raw_20230605.csv")
```

Synonym list

Save csv file



# Quick-look at downloaded files

Read in downloaded data frame

```
rawdf <- read.csv("data/download/raw/Shortia_galacifolia_raw_20230605.csv")
```

Inspect the data frame

**What columns are included?**

```
names(rawdf)
```

```
## [1] "scientificName"      "genus"
## [3] "specificEpithet"     "infraspecificEpithet"
## [5] "ID"                  "occurrenceID"
## [7] "basisOfRecord"       "eventDate"
## [9] "year"               "month"
## [11] "day"                "institutionCode"
## [13] "recordedBy"         "country"
## [15] "county"            "stateProvince"
## [17] "locality"          "latitude"
## [19] "longitude"         "coordinateUncertaintyInMeters"
## [21] "informationWithheld" "habitat"
## [23] "aggregator"
```

## Where are these points?

The error message here indicates many points do not have long/lat values (more in 02).

```
leaflet(rawdf) %>%  
  addMarkers(label = paste0(rawdf$longitude, ", ", rawdf$latitude)) %>%  
  addTiles()
```

