Using databases for exploring research questions

ALINA AVANESYAN

LAB MEETING 11/30/2020

Data, Databases, and Research Questions: Overview

- What type of the data do you want to collect?
  - Qualitative data (images, list of species, species origin, plant life form, insect life cycle, species occurrence, etc.)
  - Quantitative data (body length, wing size, stylet length, leaf thickness, trichome density, leaf area, etc.)
  - Combination of both?

- What research (or other) questions do you want to investigate?
  - Collecting measurements
  - Species identification
  - Conducting review, meta-analysis, etc.
  - Something else?

Today's talk: Outline

- Examples of available public databases
- Data you could extract
- Questions you might be interested in

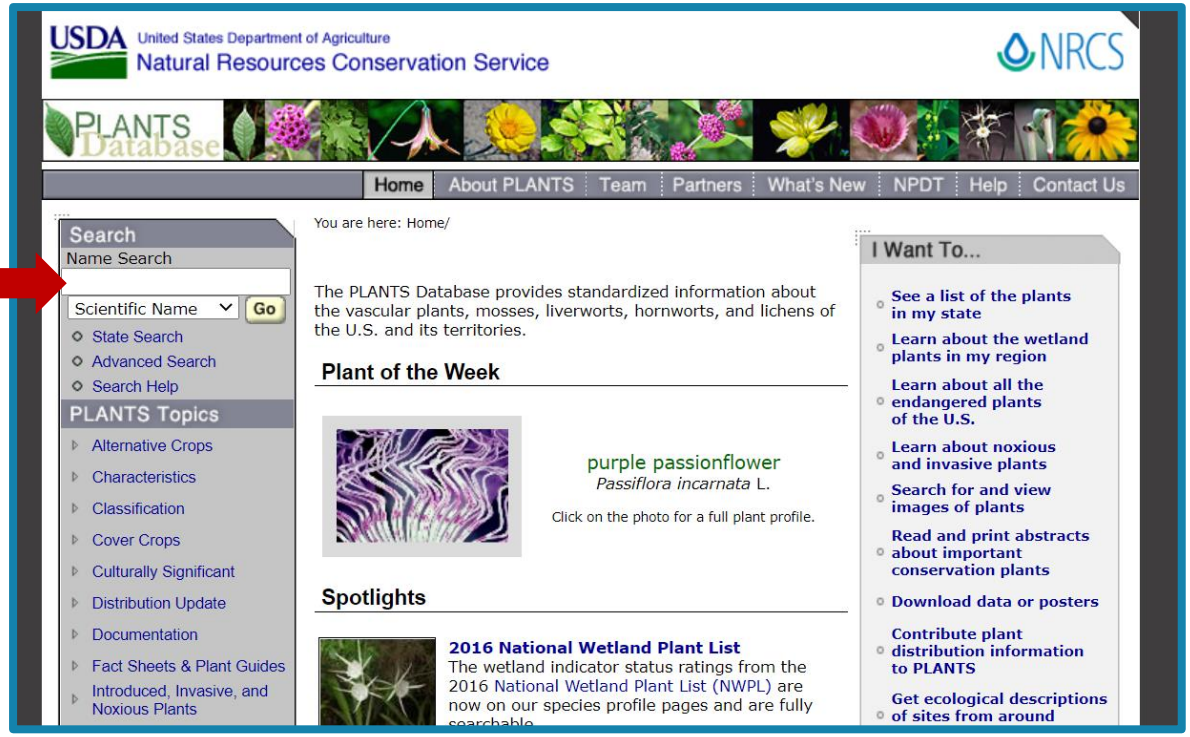
- Plant databases 
- Various species (including plants and insects) 
- Specific insect databases 

USDA plants database

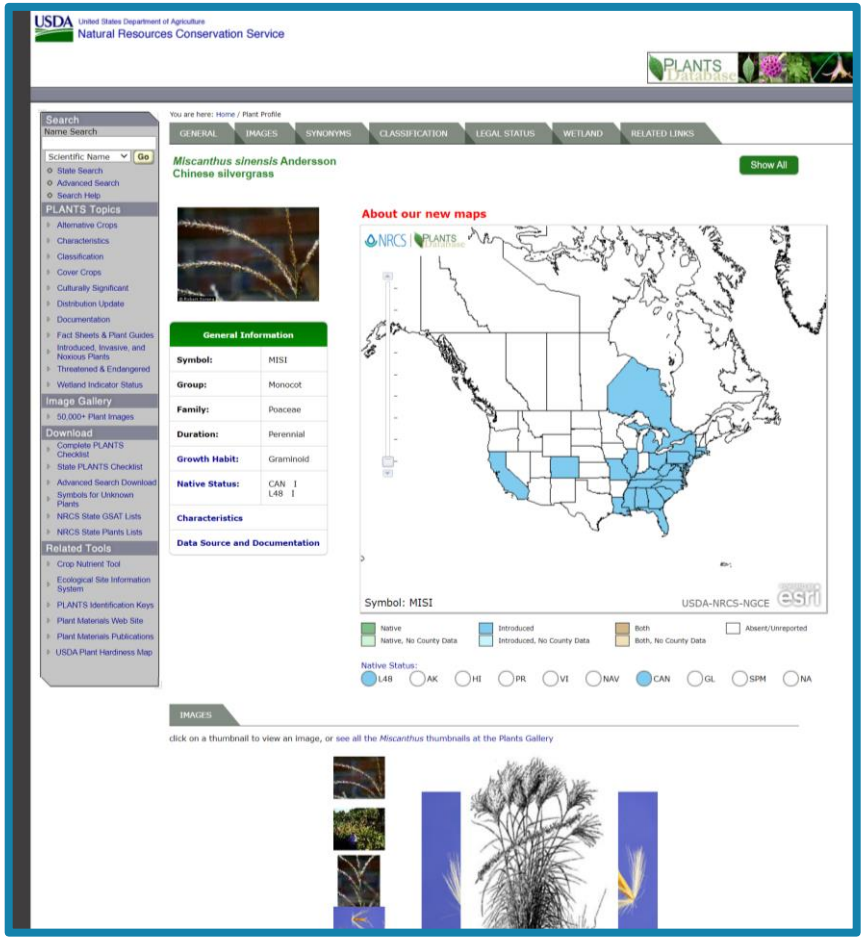
- https://plants.sc.egov.usda.gov/java/

Common name

Scientific name



USDA plants database



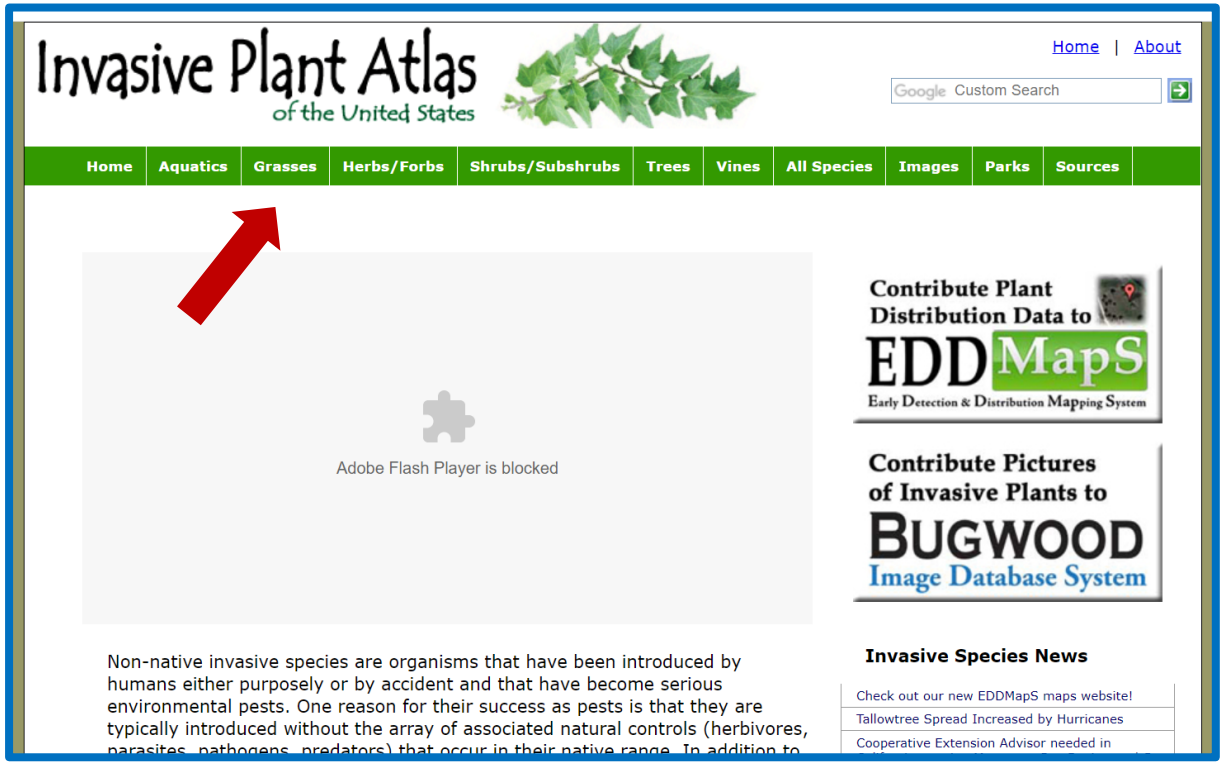
- Plant origin
- Species distribution
- Life form
- Taxonomy
- Images
- Other characteristics

The Invasive Plant Atlas of the US

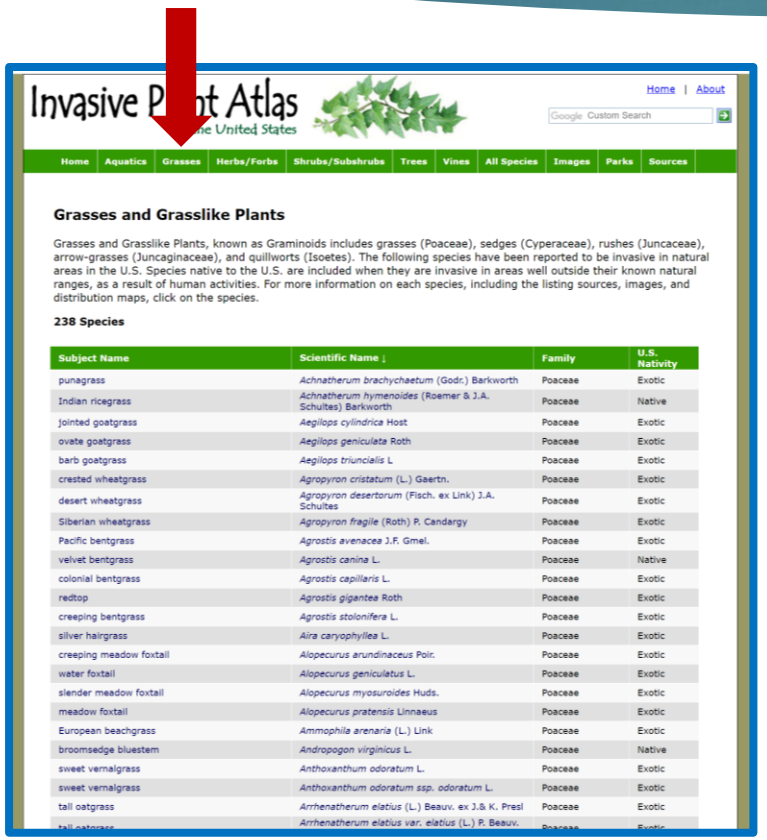
- https://www.invasiveplantatlas.org/

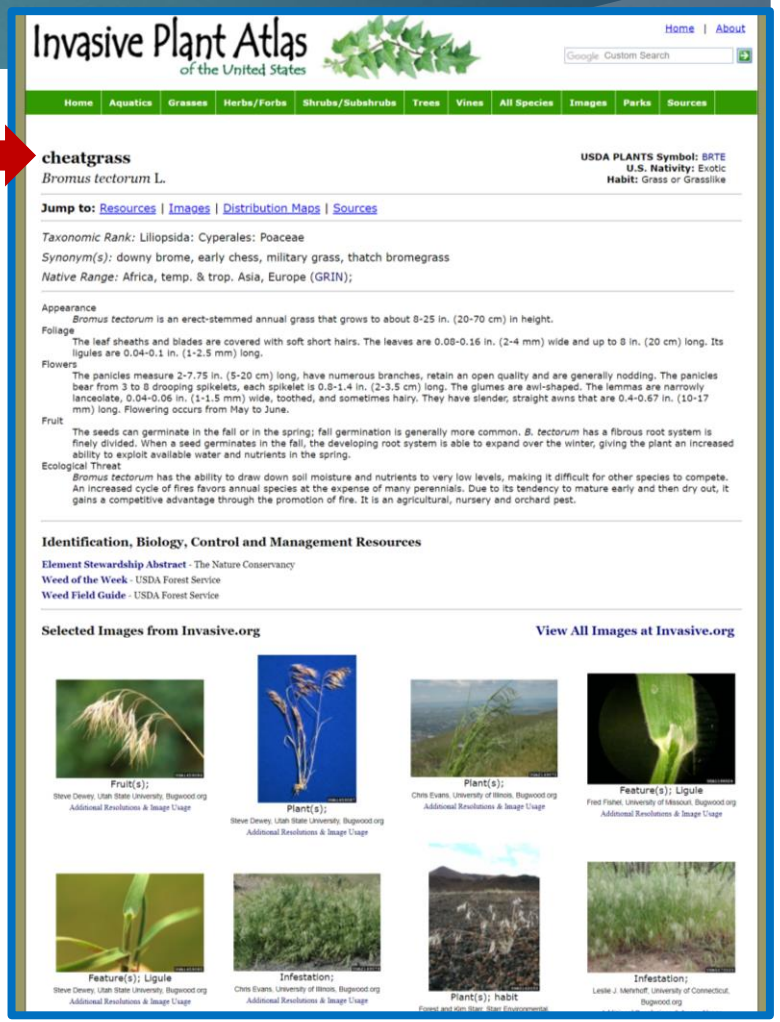
Life form

All species



The Invasive Plant Atlas of the US





- Appearance
- Flowers/seeds size
- Native range
- Ecological threat

TRY Plant Database

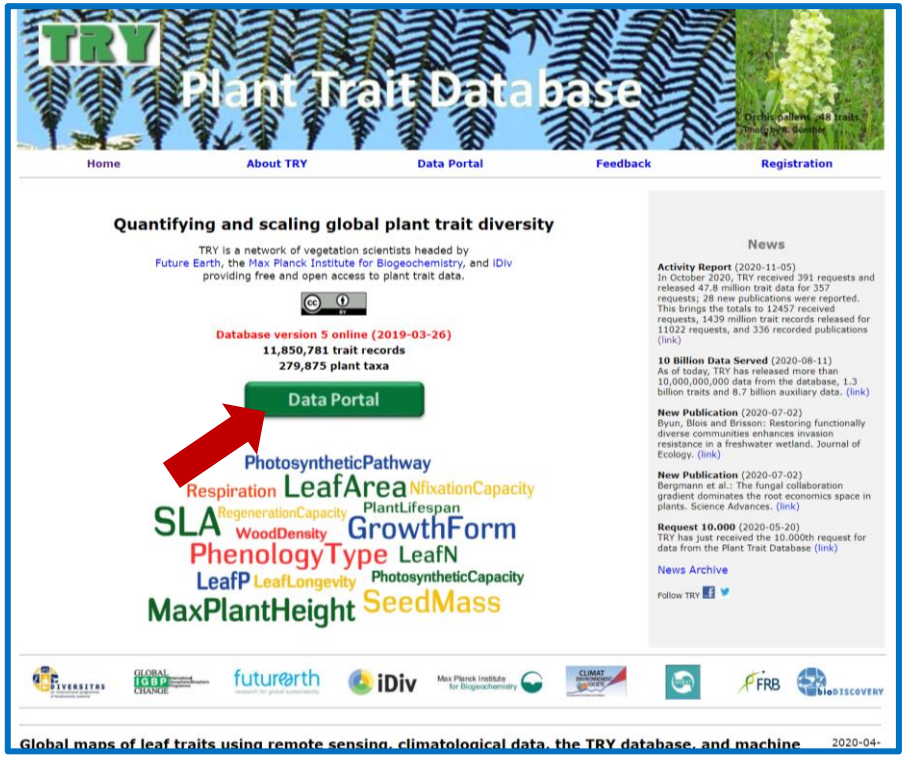
- https://www.try-db.org/TryWeb/Home.php

Quantitative plant traits

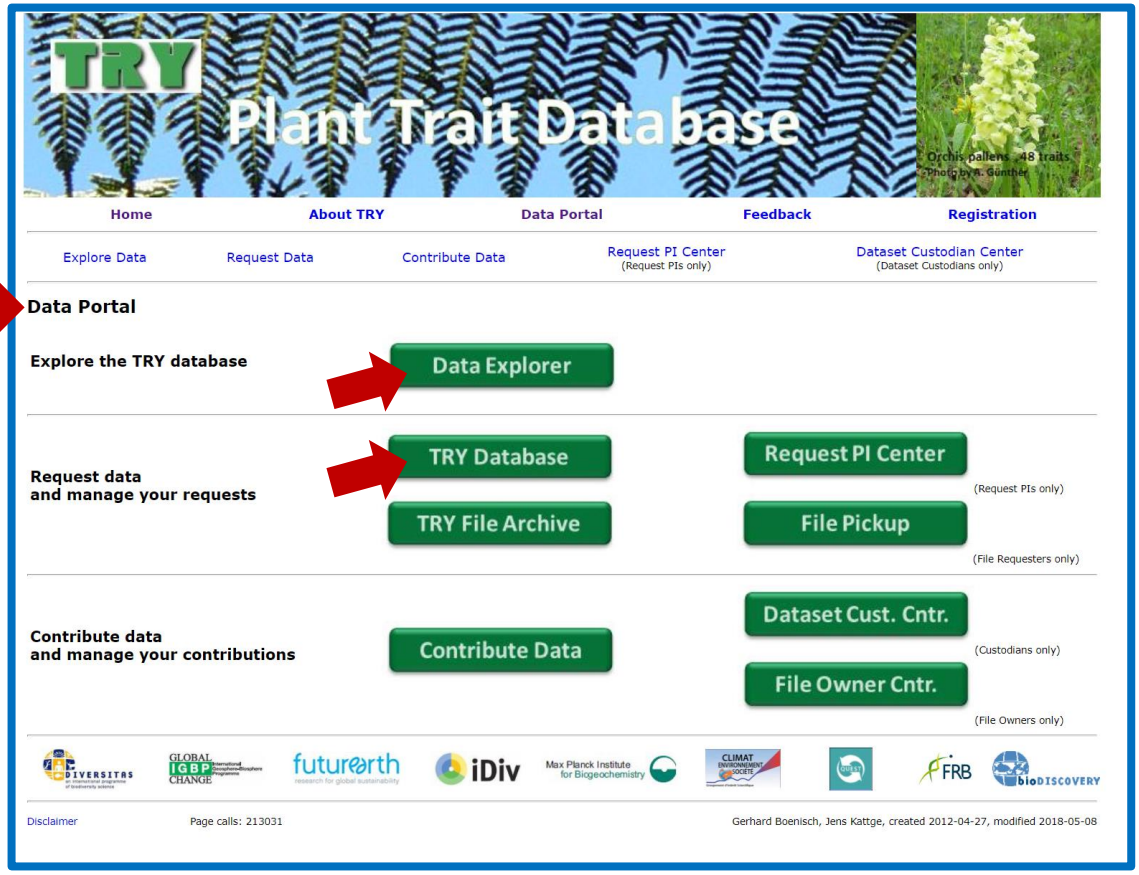
Woody and non-woody plants

Data need to be requested (easy and quick)

Data are in txt-format: can be opened in Excel (if data are not too large) or in Linux (if data are too large)



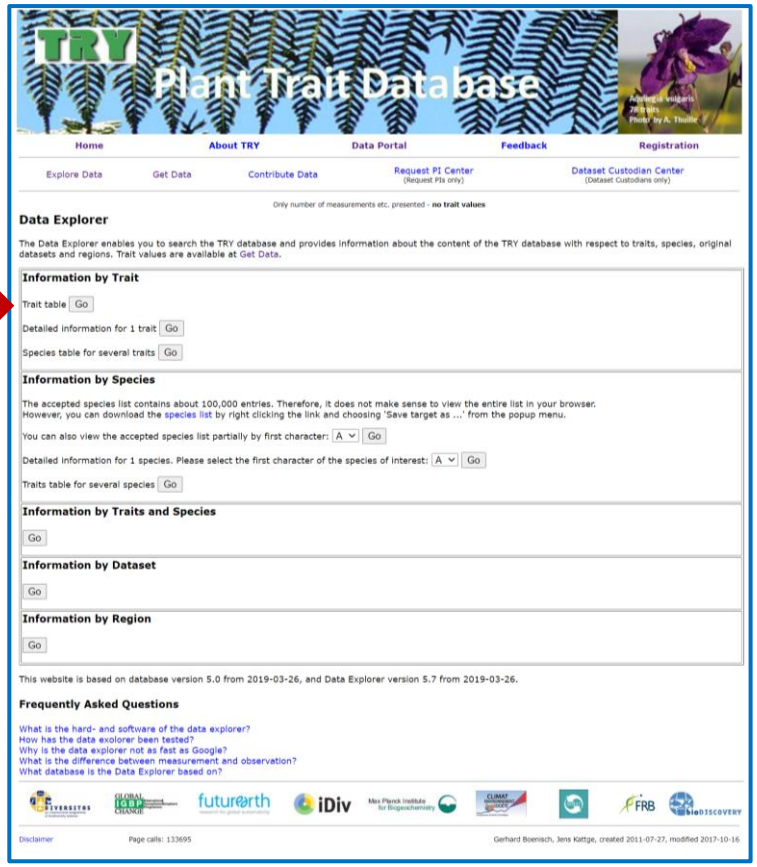
TRY Plant Database

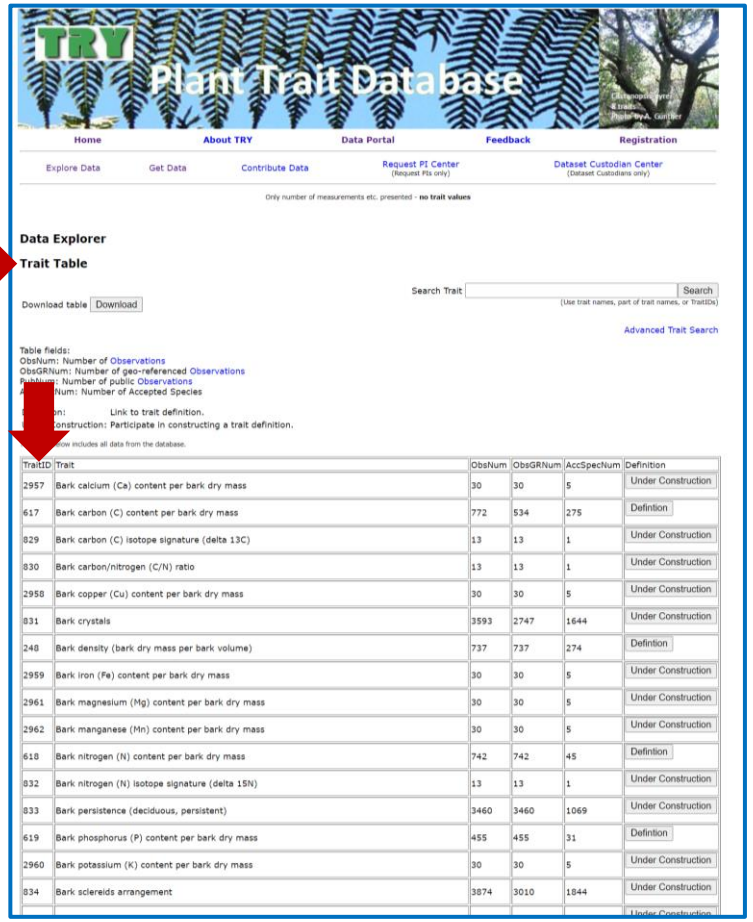


- Exploring data
- Requesting data

TRY Plant Database

- Exploring data

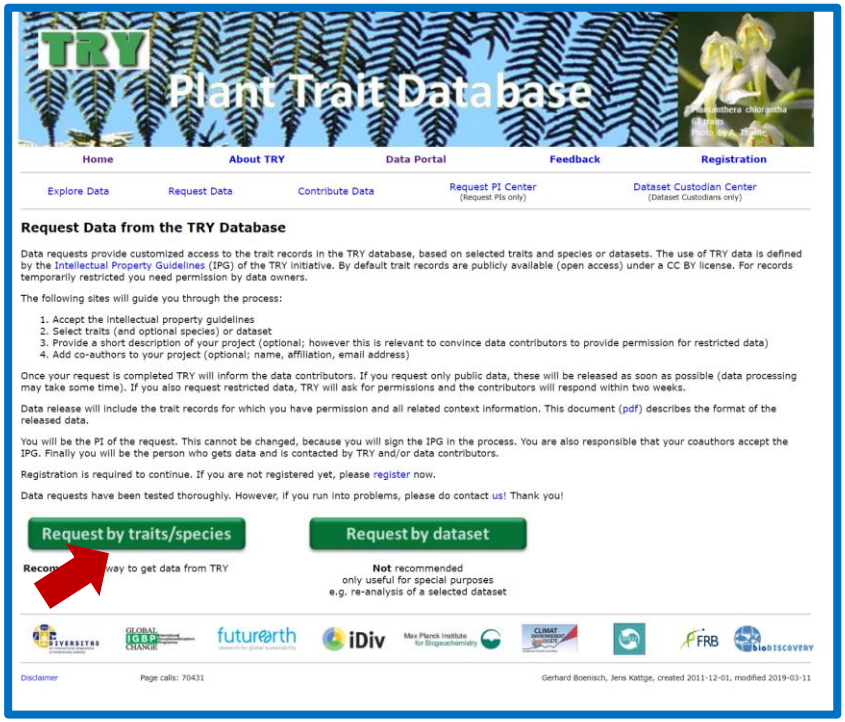




TraitID	Trait	ObsNum	ObsGRNum	AccSpecNum	Definition
2957	Bark calcium (Ca) content per bark dry mass	30	30	5	Under Construction
617	Bark carbon (C) content per bark dry mass	772	534	275	Definition
829	Bark carbon (C) isotope signature (delta 13C)	13	13	1	Under Construction
830	Bark carbon/nitrogen (C/N) ratio	13	13	1	Under Construction
2958	Bark copper (Cu) content per bark dry mass	30	30	5	Under Construction
831	Bark crystals	3593	2747	1644	Under Construction
248	Bark density (bark dry mass per bark volume)	737	737	274	Definition
2959	Bark iron (Fe) content per bark dry mass	30	30	5	Under Construction
2961	Bark magnesium (Mg) content per bark dry mass	30	30	5	Under Construction
2962	Bark manganese (Mn) content per bark dry mass	30	30	5	Under Construction
618	Bark nitrogen (N) content per bark dry mass	742	742	45	Definition
832	Bark nitrogen (N) isotope signature (delta 15N)	13	13	1	Under Construction
833	Bark persistence (deciduous, persistent)	3460	3460	1069	Under Construction
619	Bark phosphorus (P) content per bark dry mass	455	455	31	Definition
2960	Bark potassium (K) content per bark dry mass	30	30	5	Under Construction
834	Bark sclereids arrangement	3874	3010	1844	Under Construction

TRY Plant Database

- Requesting data



- Example of the request:

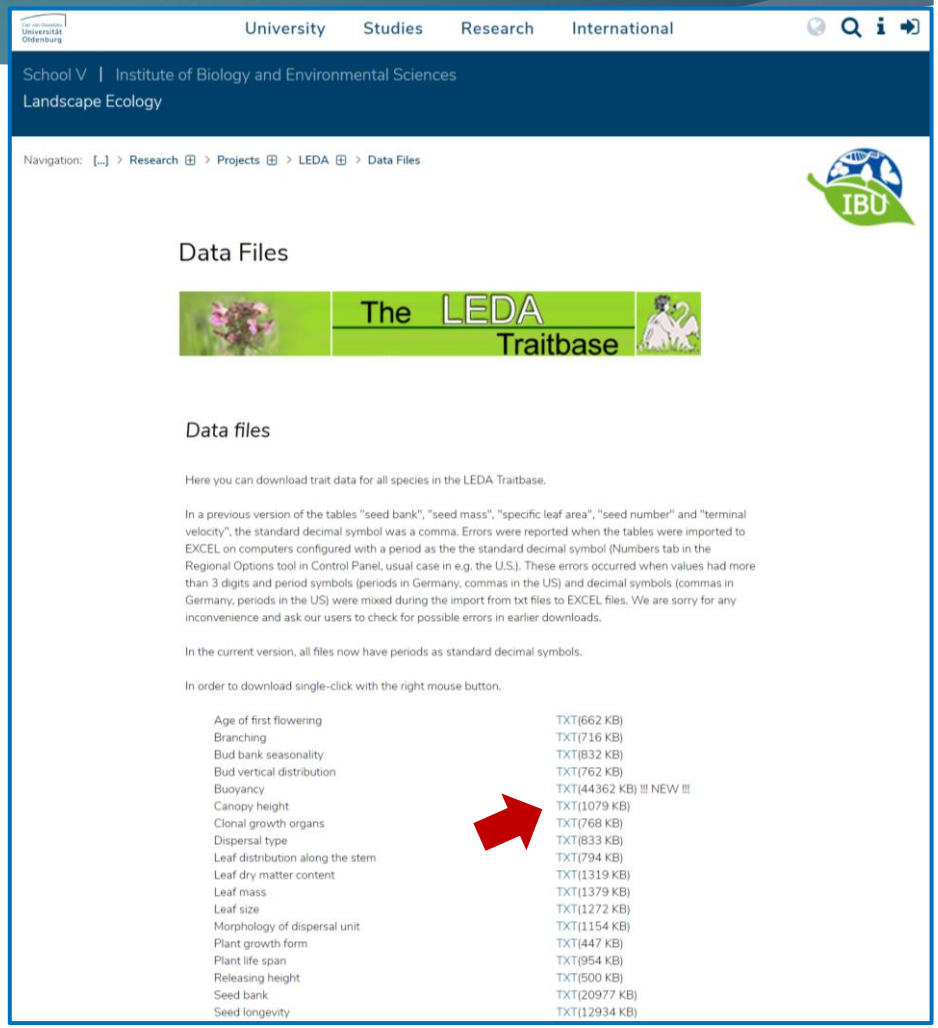
TRY Data Request 10895
Only public data were requested.
Title:
Host plant usage by the spotted lanternfly, Lycorma delicatula
Authors:
Alina Avanesyan (University of Maryland, Department of Entomology)
Olivia Shaffer (Frostburg State University, Frostburg MD)
William Lamp (University of Maryland, Dept of Entomol)
Trait List:
38, 3064, 24, 618, 2957, 617, 838, 837, 3404
Species List:

Description:
Lycorma delicatula is a highly invasive insect pest of fruit crops and trees in the eastern US. Nymphs and adults pose a significant economic threat to many woody tree species in MD including native and economically-important trees and woody plants. In this project we focus on exploring the basic behavioral, morphological, and physiological mechanisms which drive food plant selection of this invasive insect in its introduced range.

The LEDA Traitbase

- https://uol.de/en/landeco/research/leda/data-files/

Downloadable datasets



R package:
TR8: Extract traits data for plant species

- https://cran.r-project.org/web/packages/TR8/vignettes/TR8.pdf

- retrieves traits data for plant species from the following publicly available databases:
  - Biolflor http://www.ufz.de/biolflor/index.jsp
  - Ecological Flora of the British Isles http://www.ecoflora.co.uk/
  - LEDA traitbase http://www.leda-traitbase.org/LEDAportal/
  - Ellenberg values for Italian Flora
  - Flowering period for Italian Flora (data retrieved from http://luirig. altervista.org/)
  - Mycorrhizal intensity database
  - MycoFlor database
  - Catminat database
  - BROT

R package:
TR8: Extract traits data for plant species

- Example of retrieving data:

```
##writeLines('PATH="${RTOOLS40_HOME}\\usr\\bin;${PATH}"', con = "~/.Renviron")

install.packages("TR8",type= "binary", dependencies = TRUE)
install.packages("XML", type = "binary")
library(TR8)

##See available traits
print(available_tr8)

## a vector containing a list of plant species names
intro_species<-c("Acer platanoides","Acer pseudoplatanus","Ailanthus altissima", "Betula pendula", "Prunus avium", "Populus
alba", "Platanus occidentalis", "Rosa multiflora", "Vitis vinifera")
## a vector of traits (split, takes too long with all 5)
to_be_downloaded<-c("le_area","woodiness")
to_be_downloaded2<-c("leaf_thick","C.N.Ratio","Height")
## now run tr8 and store the results in the my_traits object
intro_traits<-tr8(species_list = intro_species,download_list = to_be_downloaded)
intro_traits2<-tr8(species_list = intro_species,download_list = to_be_downloaded2)

##see downloaded data
print(intro_traits)
print(intro_traits2)
```

Plant Databases: Possible Applications

- Plant community composition

  (Dylan's project)
- Plant traits of insect host plants

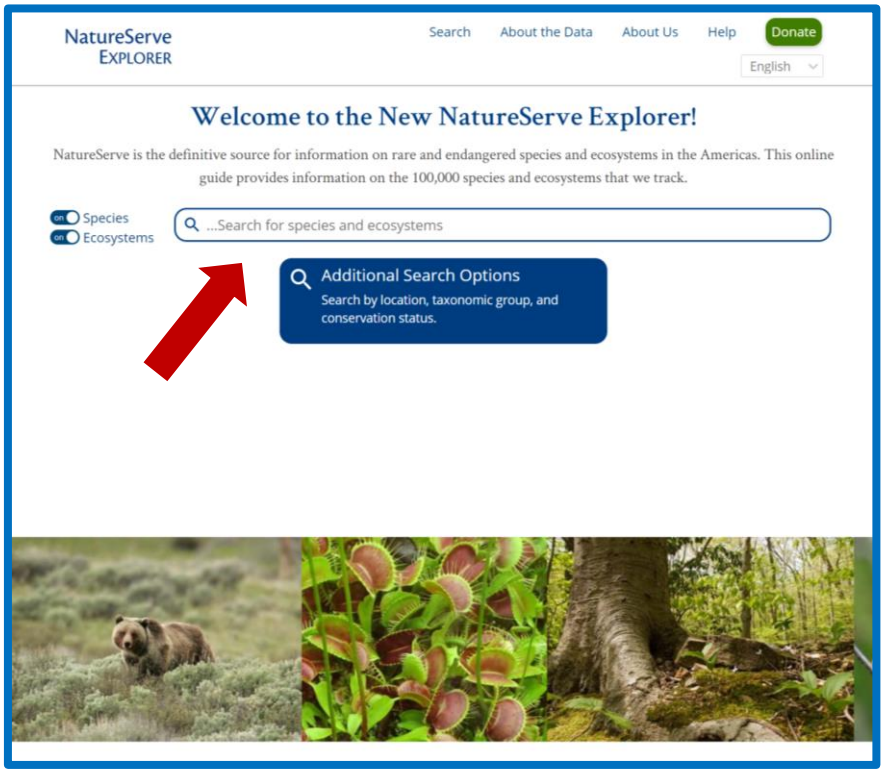
  (Olivia's project)
- Plant DNA detection from insect guts

  (identification of SLF host plants)
- Feeding preference of insect herbivores

  (meta-analysis of grasshopper feeding preferences for exotic vs. native host plants)

The NatureServe Explorer

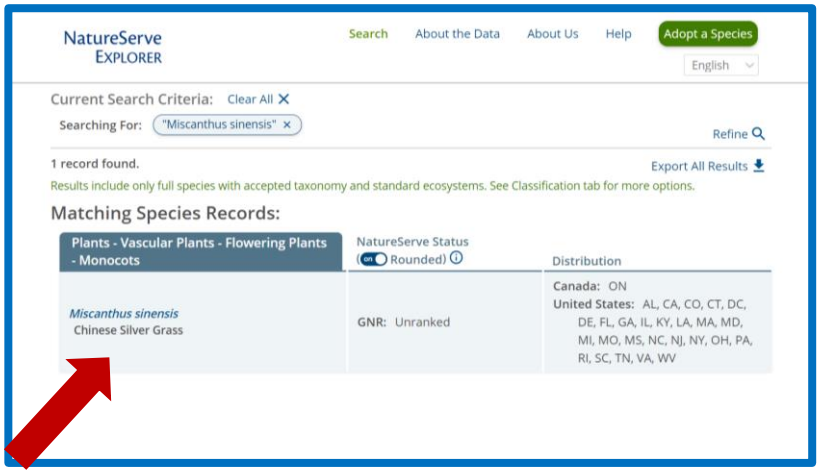
- https://explorer.natureserve.org/

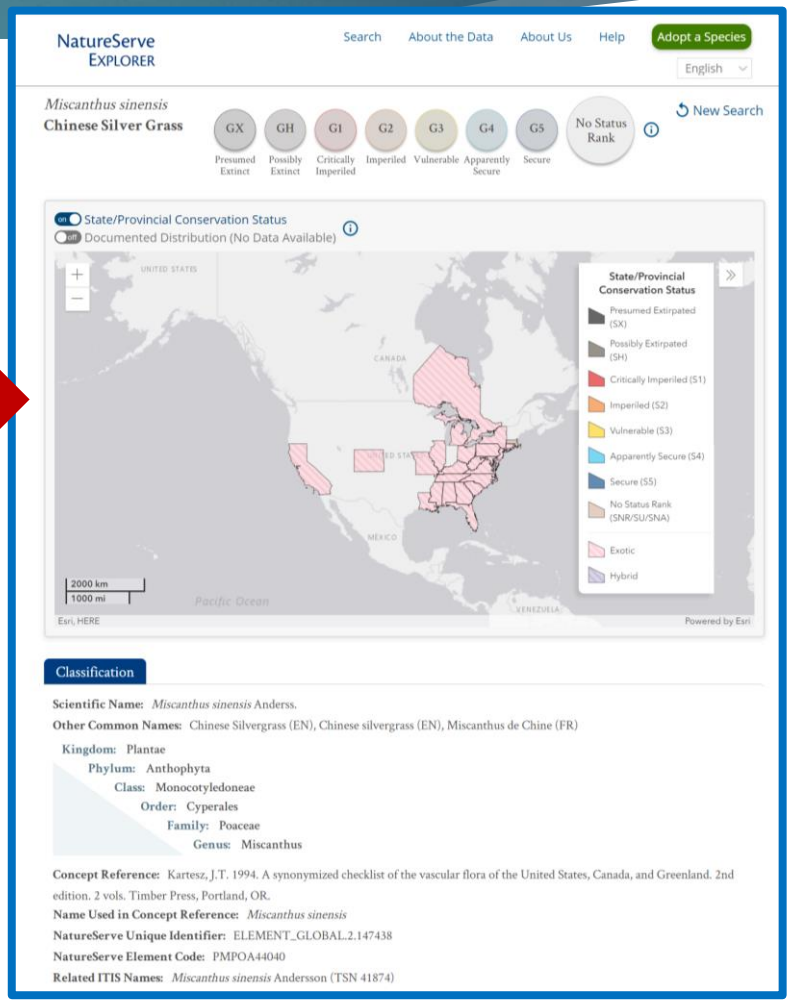


searchable database

helpful for retrieving U.S. Invasive Species Impact Rank

The NatureServe Explorer

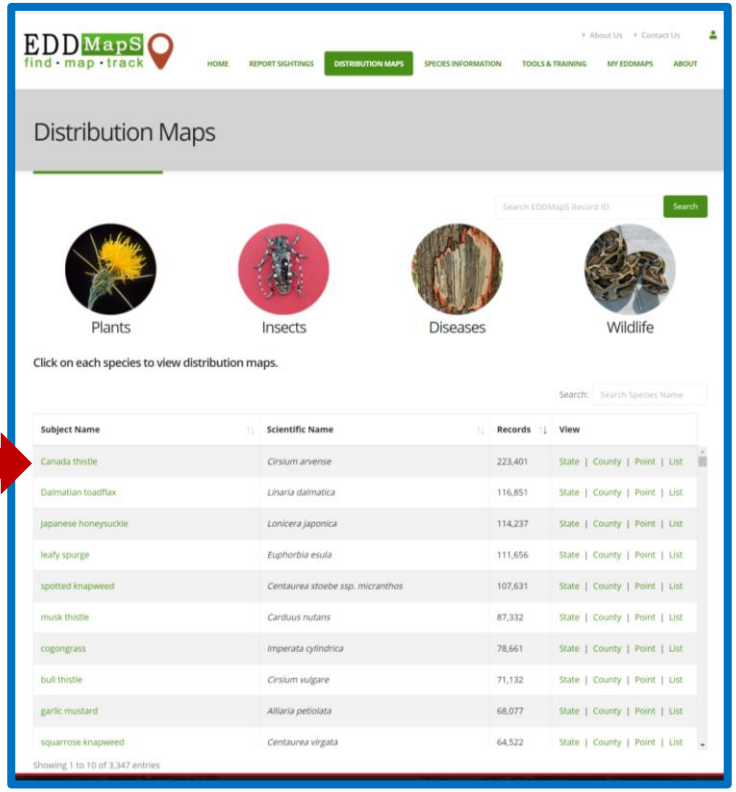


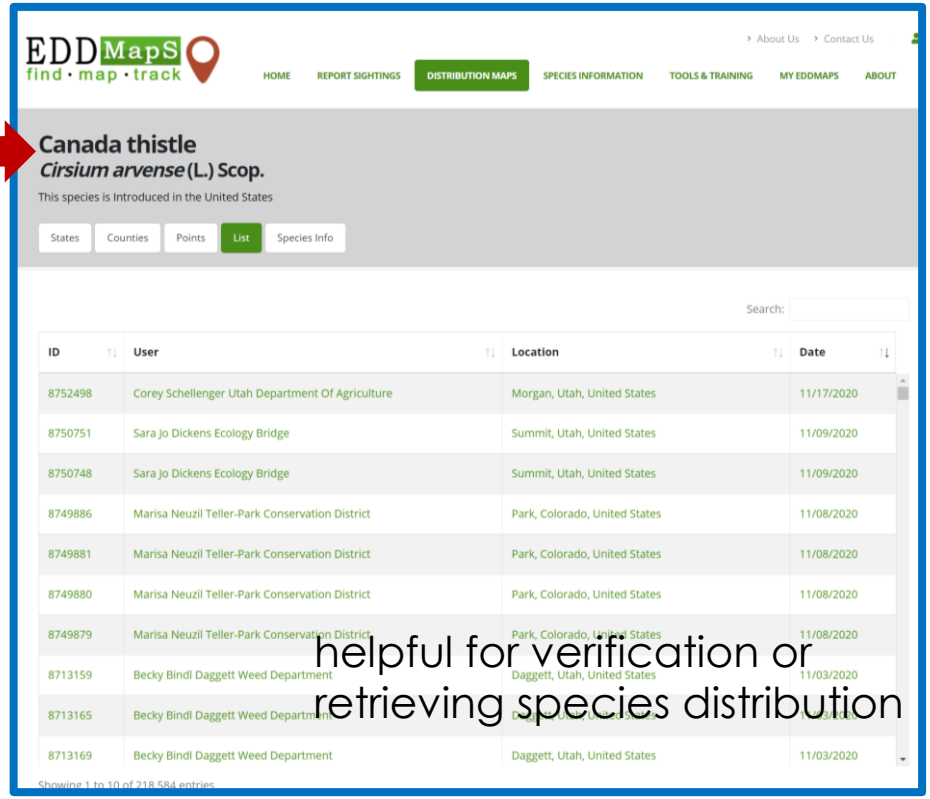


- Taxonomy
- Conservation status for each state

EDDMapS

- https://www.eddmaps.org/distribution/



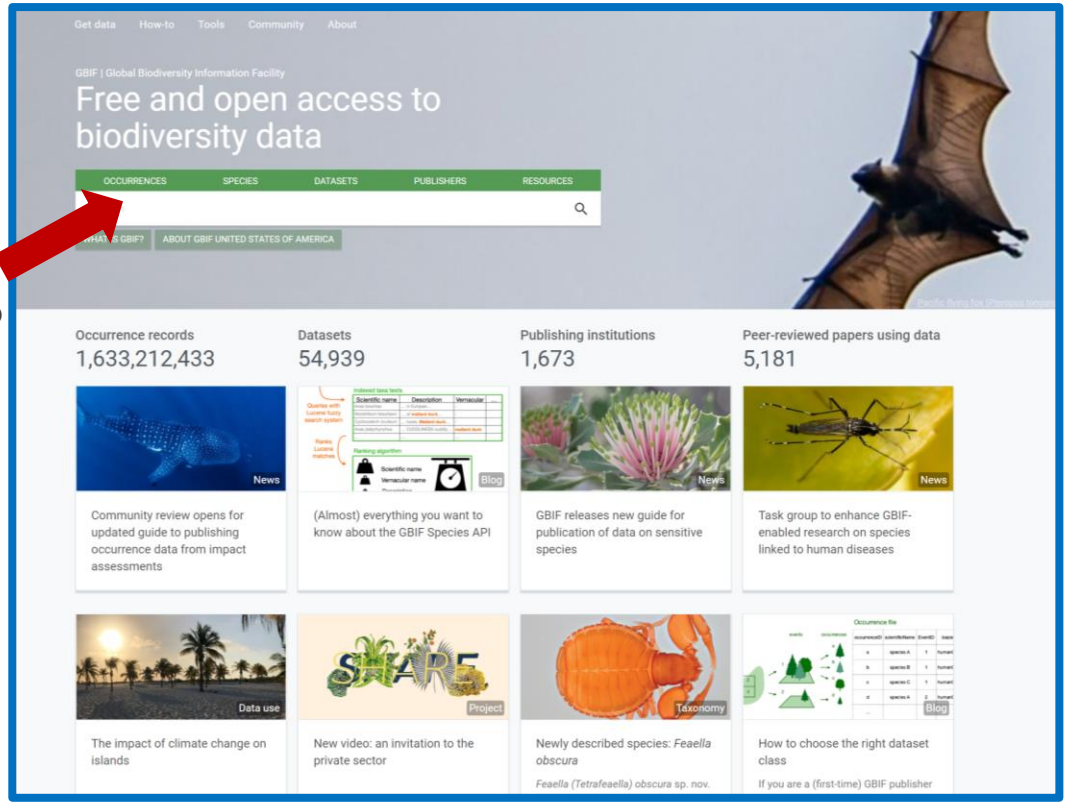


helpful for verification or retrieving species distribution

The Global Biodiversity Information Facility

- https://www.gbif.org/

Multiple options to search

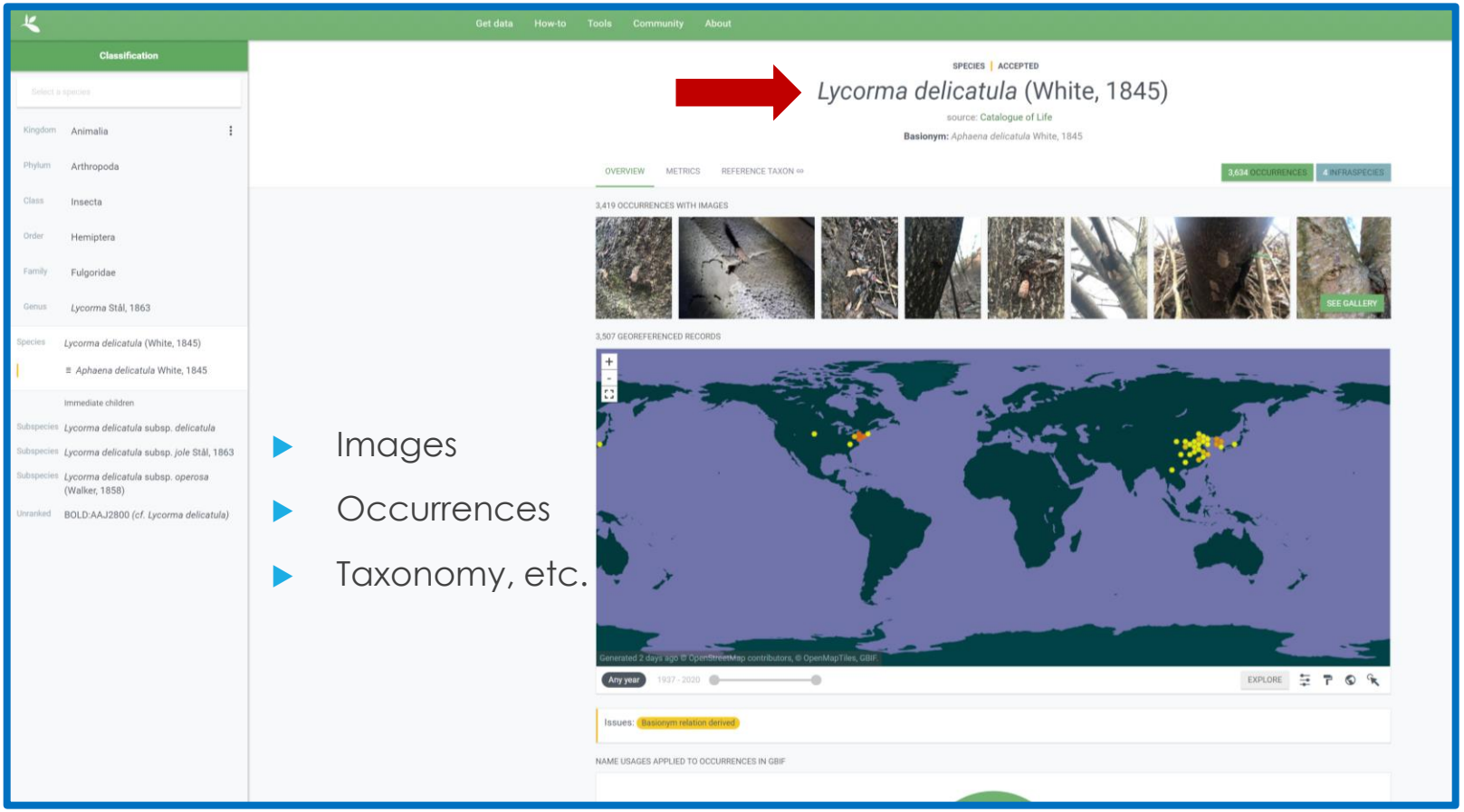


Helpful for exploring invasive species occurrences and distribution

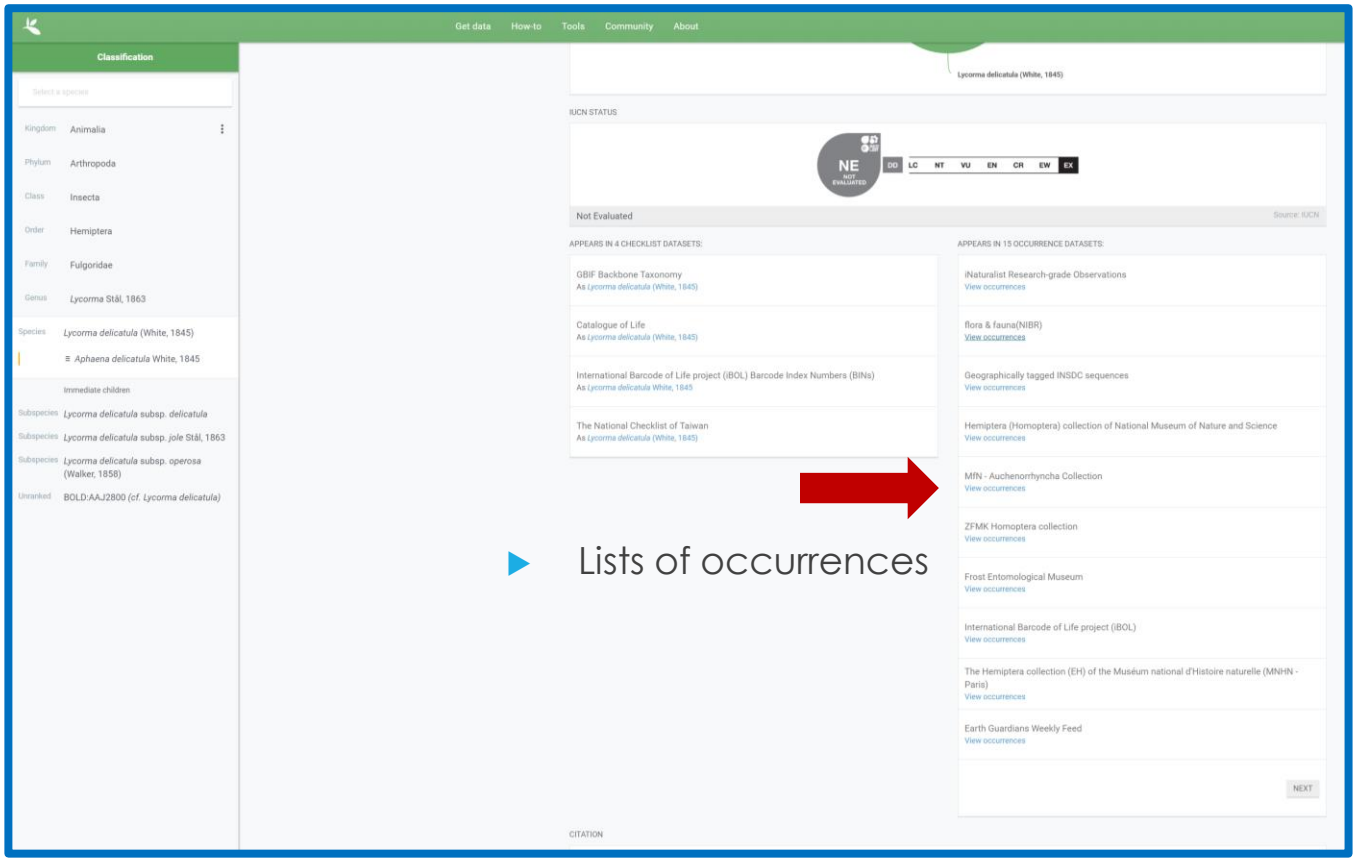
It also has downloadable datasets

The Global Biodiversity Information Facility

- Images
- Occurrences
- Taxonomy, etc.

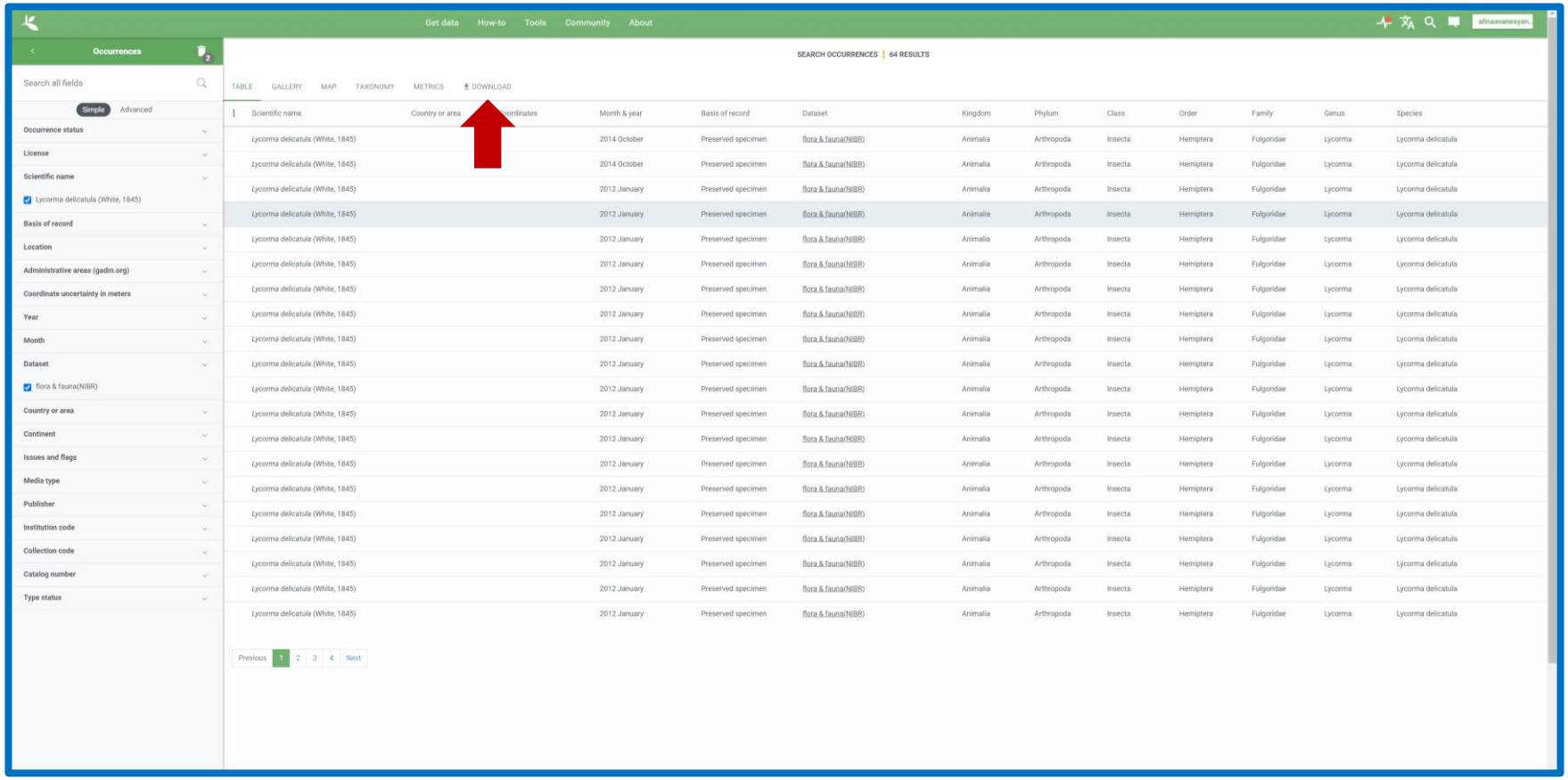


The Global Biodiversity Information Facility



- Lists of occurrences

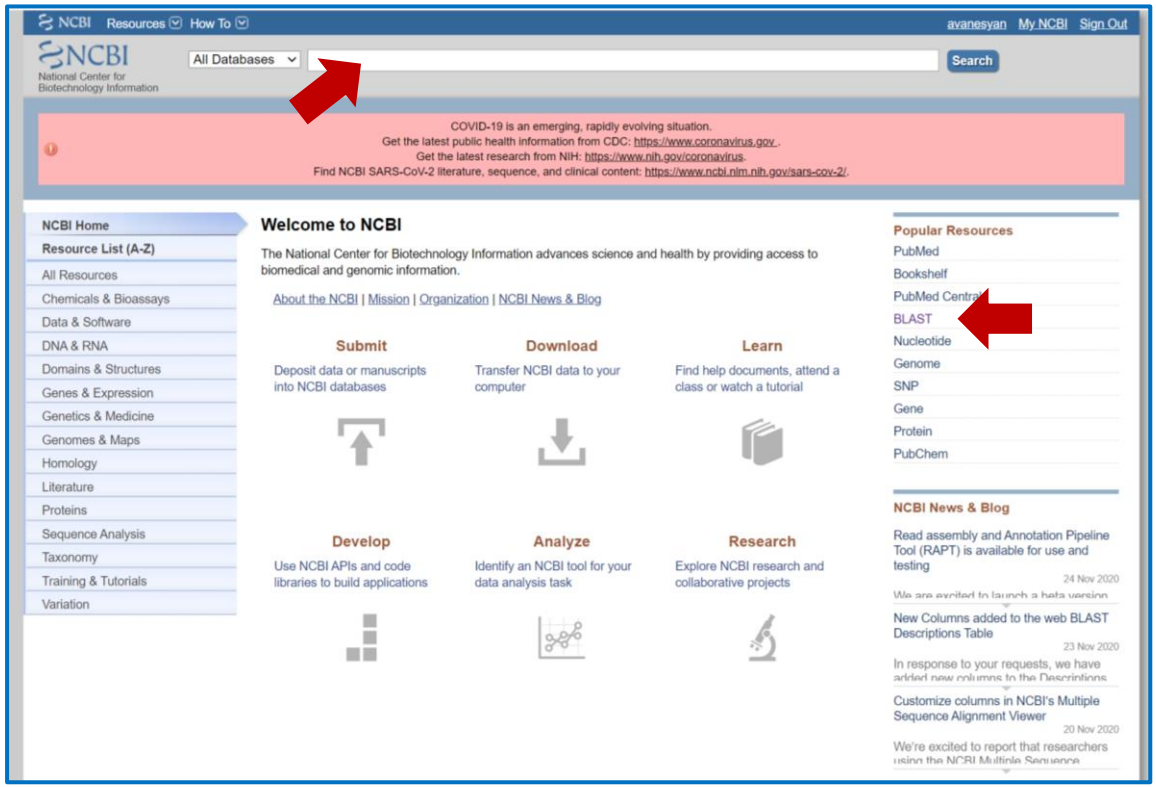
The Global Biodiversity Information Facility



- downloadable results

The NCBI GenBank database

- https://www.ncbi.nlm.nih.gov/



Species identification

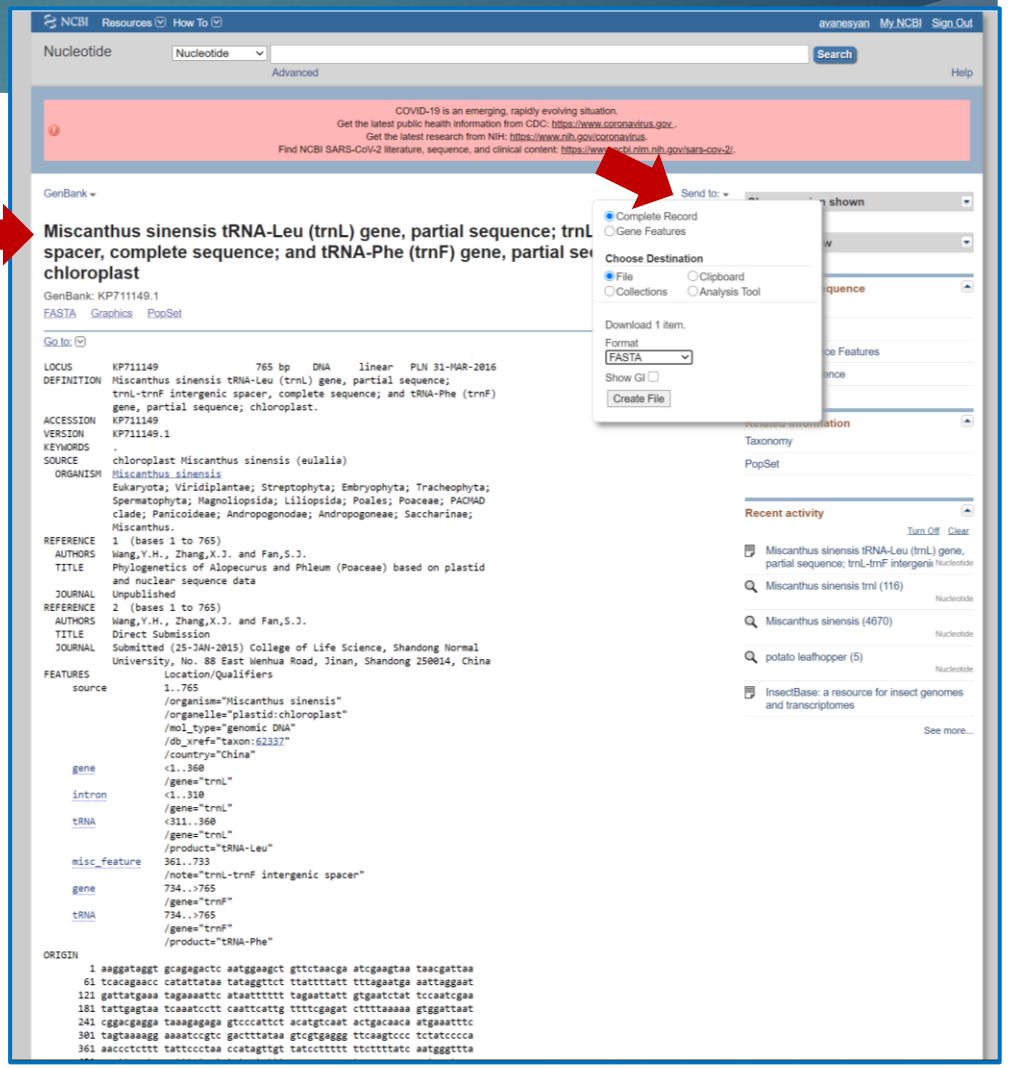
Retrieving sequences for a certain DNA locus

The NCBI GenBank database

Nucleotide

Species name and DNA locus

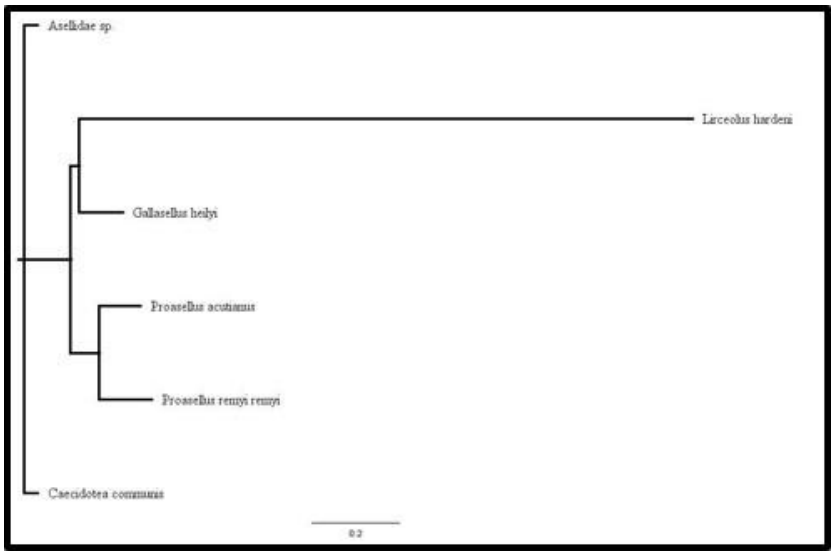
The NCBI GenBank database



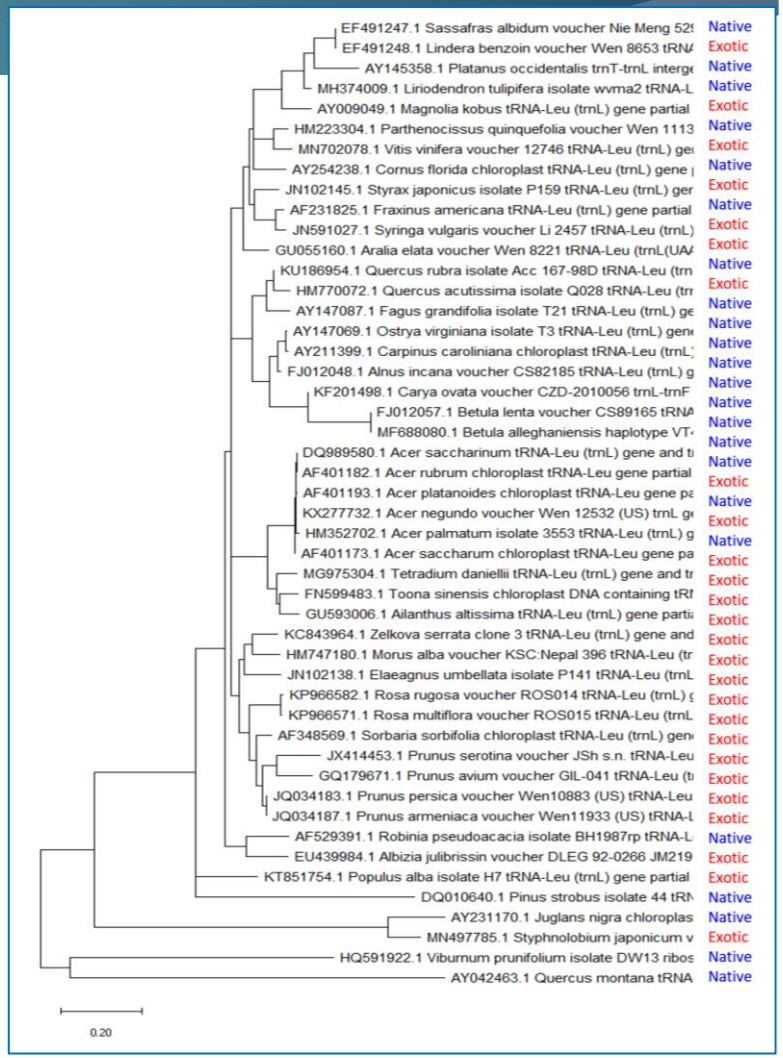
Downloadable sequences

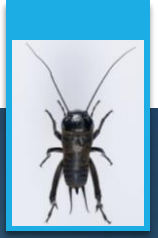
The NCBI GenBank database

- Possible applications



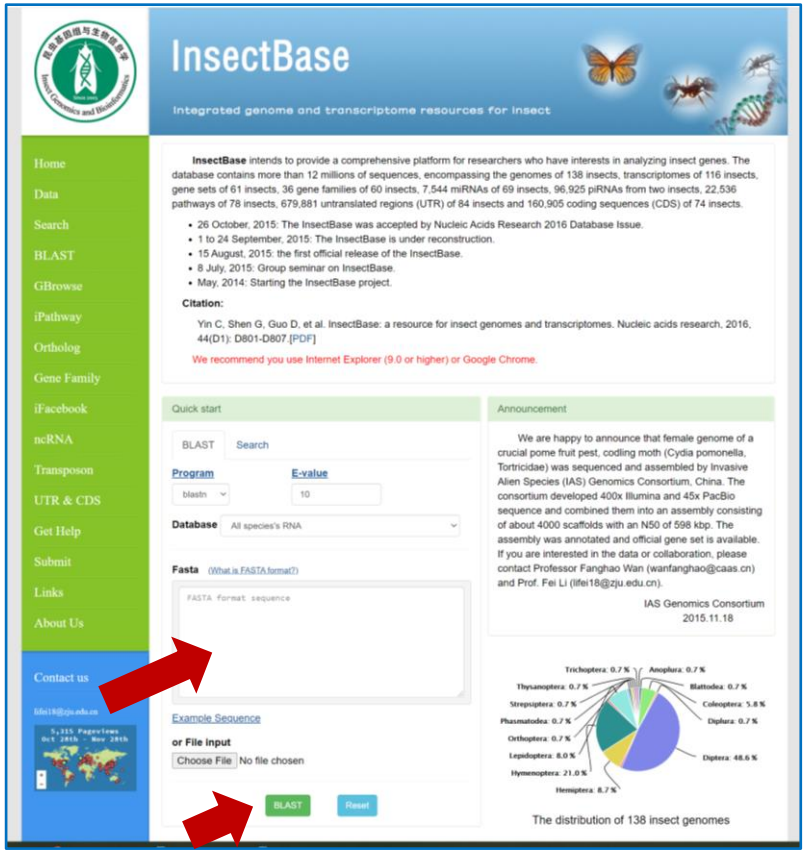
- Phylogenetic relationships of isopods (Nina's project)
  - Phylogenetic relationships of SLF host plants (Olivia's project)



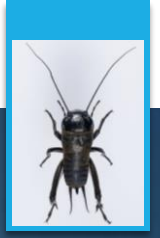


InsectBase

- http://www.insect-genome.com/

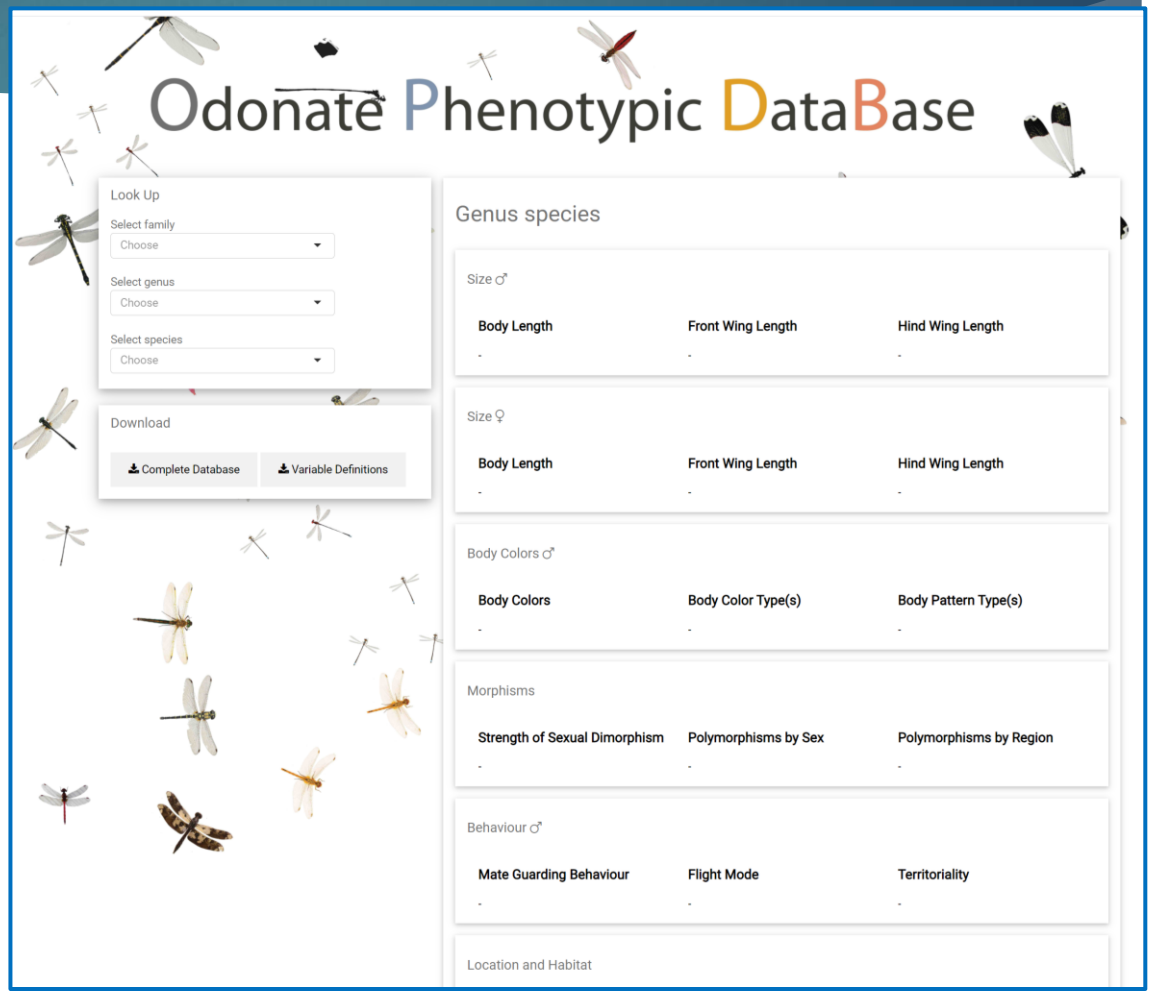


Very similar to the NCBI GenBank database

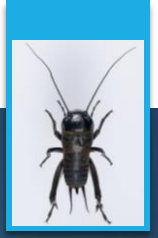


Odonate Phenotypic DataBase

- http://www.odonatephenotypicdatabase.org/shiny/odonates/?_inputs_&choose_species=%22%22

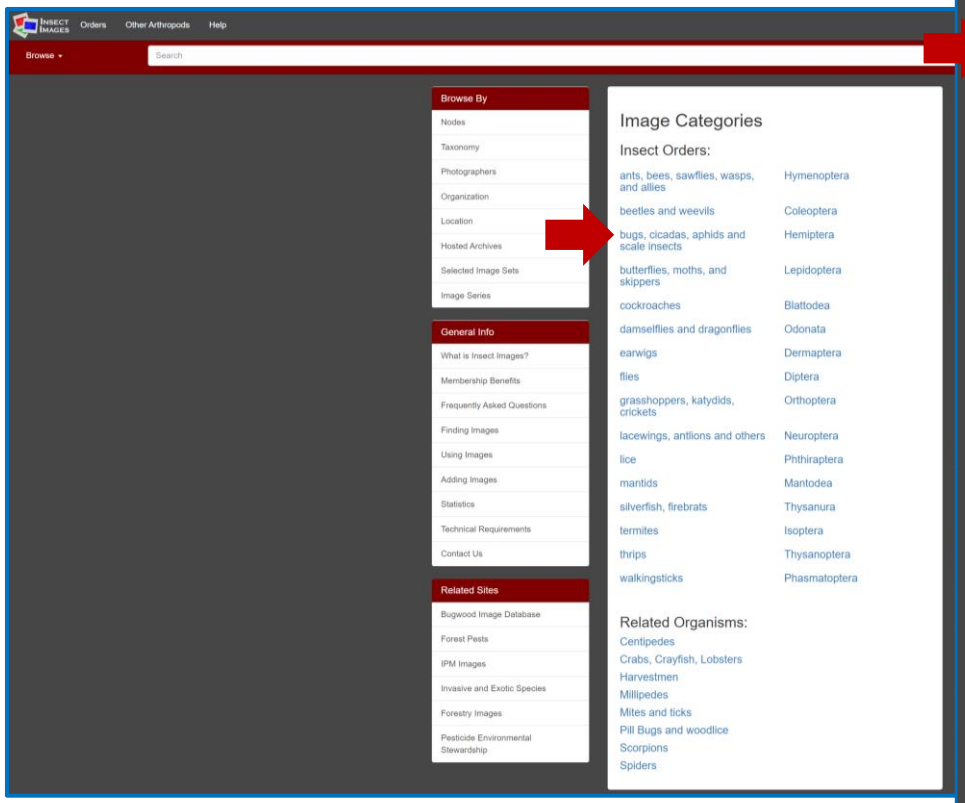


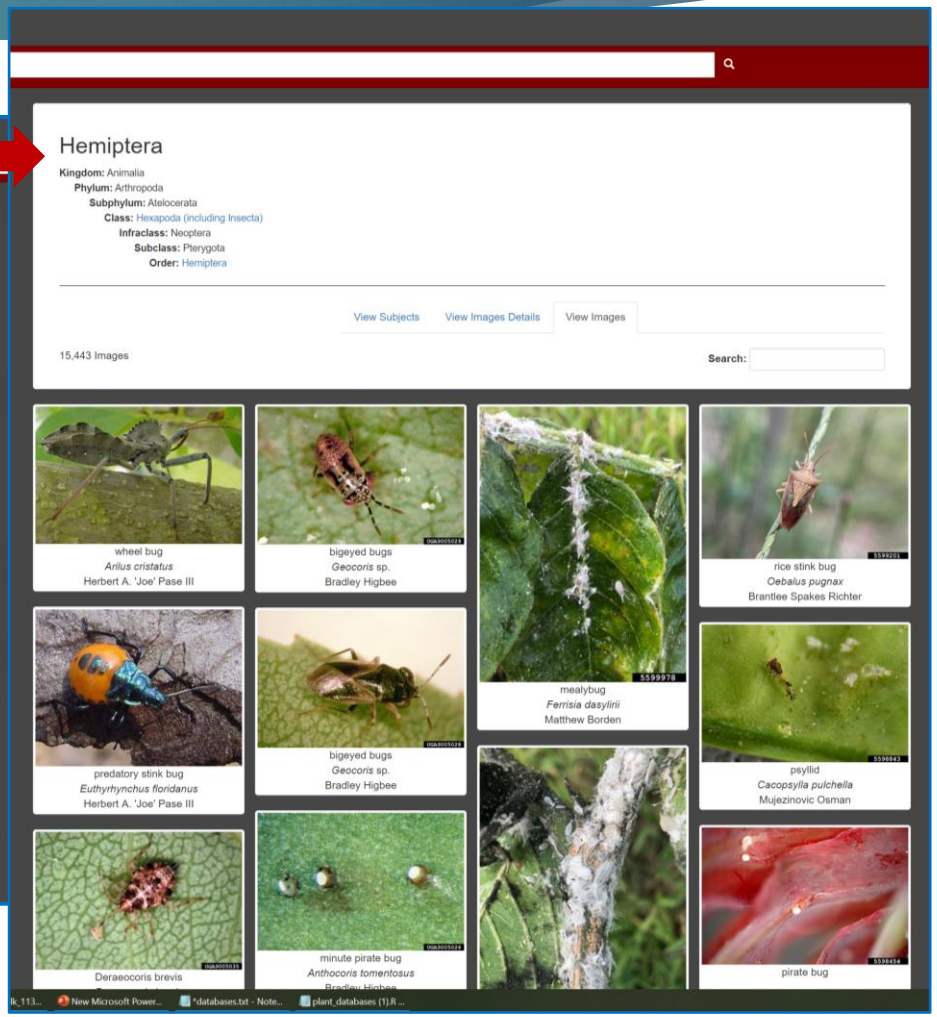
Waller, J. T., Willink, B., Tschol, M., & Svensson, E. I. (2019). The odonate phenotypic database, a new open data resource for comparative studies of an old insect order. *Scientific Data*, *6*(1), 1-6.

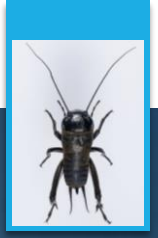


Insect Images

- https://www.insectimages.org/

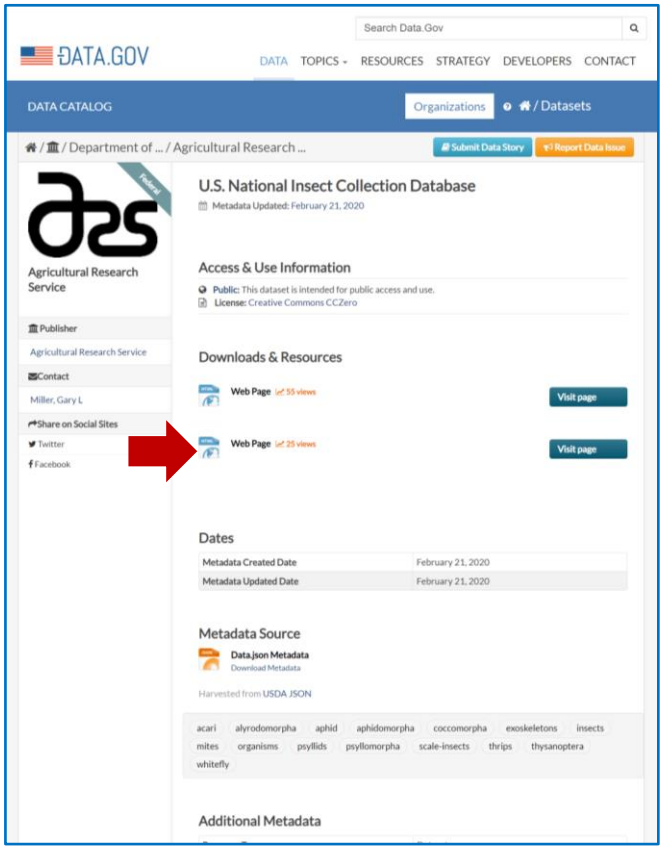




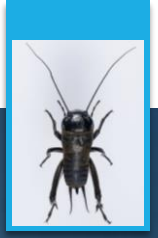


U.S. National Insect Collection Database

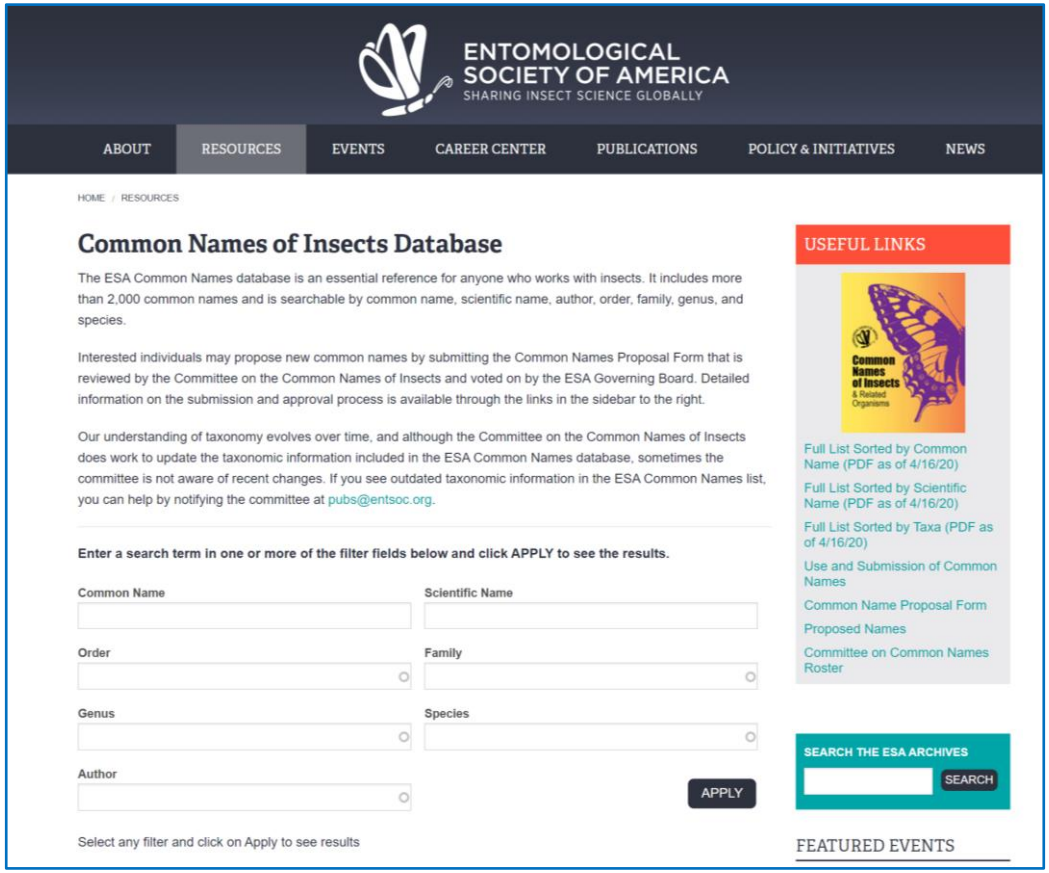
- https://catalog.data.gov/dataset/u-s-national-insect-collection-database



ESA Common Names of Insects Database



- https://www.entsoc.org/common-names



Additional helpful resources

- BugGuide

  https://bugguide.net/node/view/15740

- Illinois Natural History Survey: Insect Collection

  https://insect.inhs.illinois.edu/data/

- SCALETOOL / Species Traits Databases

  http://scales.ckff.si/scaletool/?menu=6

Thank you!

Happy Data Mining!