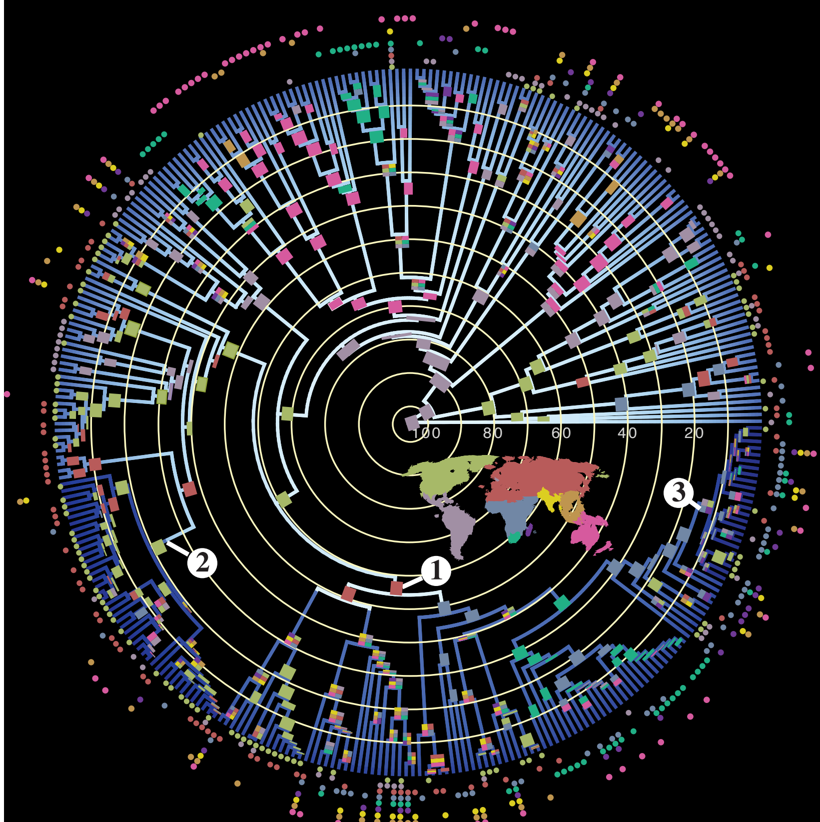
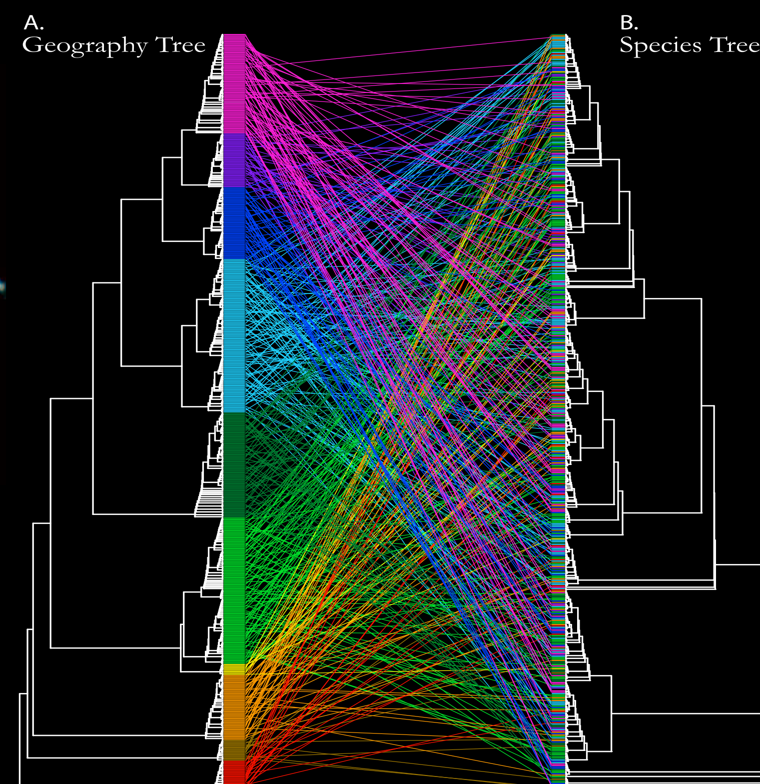
Harvesting and harnessing data for biogeographical research



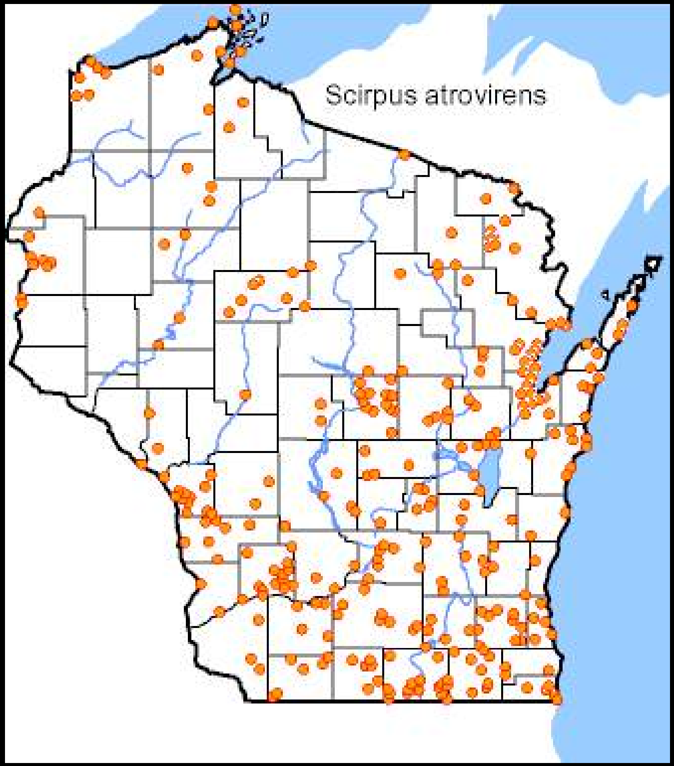


How do we know what grows where?

- inventories and surveys
  - natural areas, preserves, state forests, private properties
  - development permits
- herbarium/museum specimens
  - collectors often target specific places or taxa
  - records of what, where, when, and by whom

How do we know what grows where?

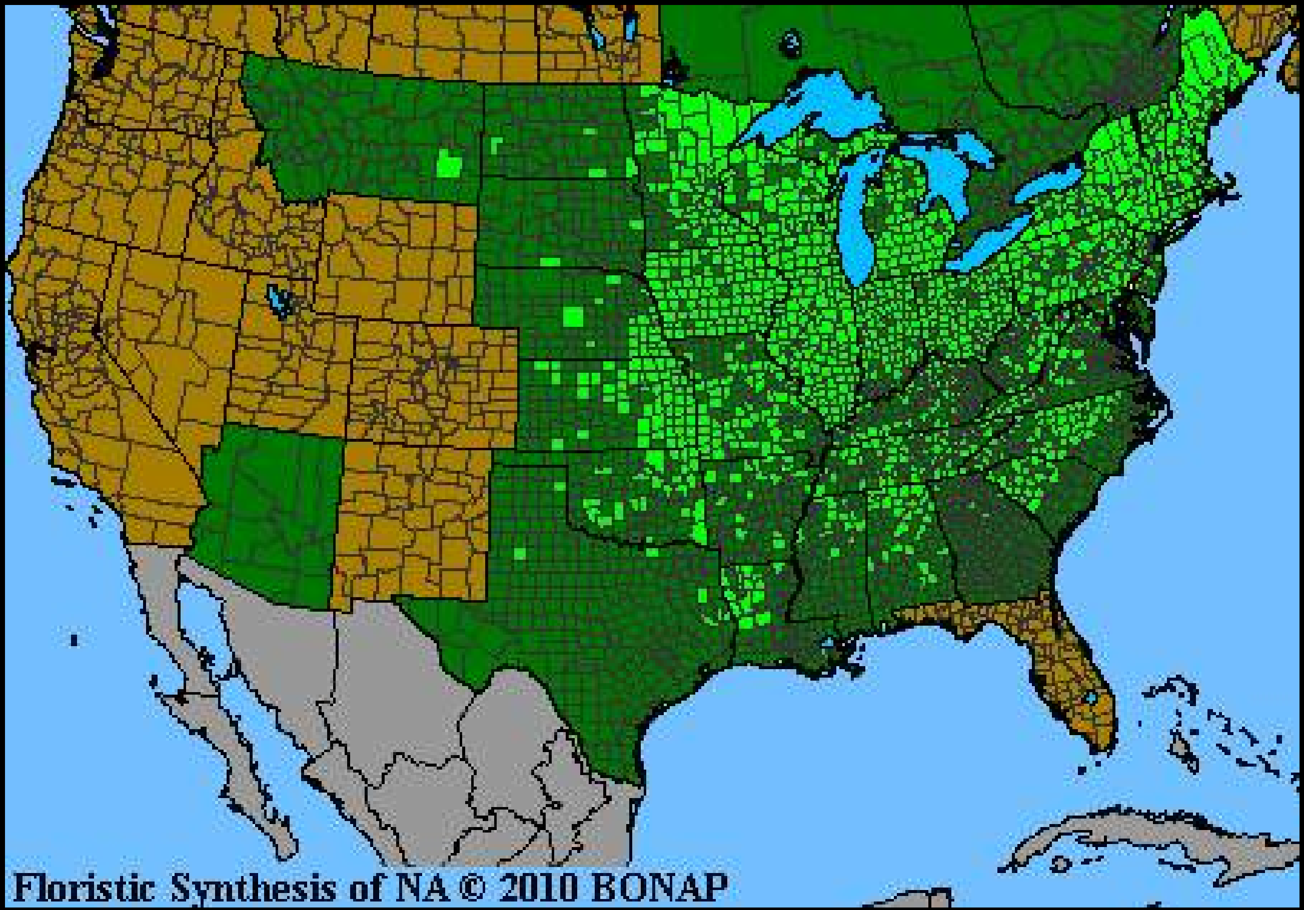
∑ Local observations = State Flora



black
bulrush

How do we know what grows where?

∑ State Floras = National Flora



black bulrush

How do we know what grows where?

$\sum$ National Floras = Global Flora

black
bulrush

Who manages these data?

Wisconsin State Herbarium · University of Wisconsin

United States Environmental Protection Agency

U.S. Department of the Interior · March 3, 1849

Forest Service · U S · Department of Agriculture

Missouri Botanical Garden

Michigan Natural Features Inventory

USDA

The Field Museum

Kew · Royal Botanic Gardens

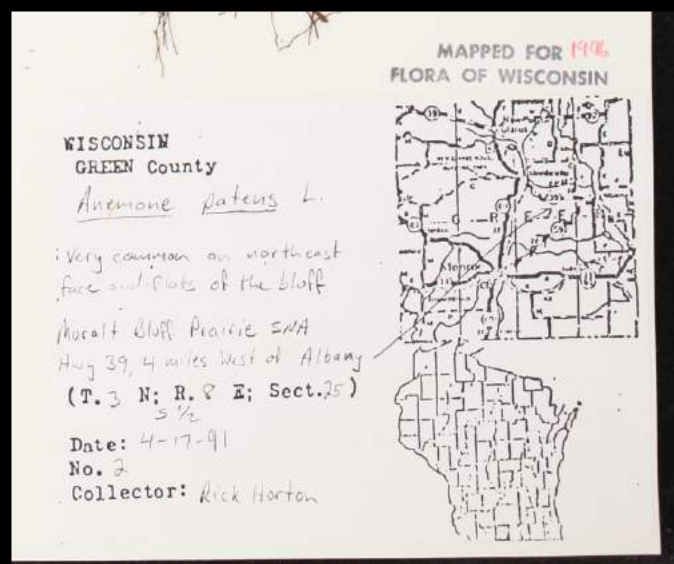
Wisconsin Dept. of Natural Resources

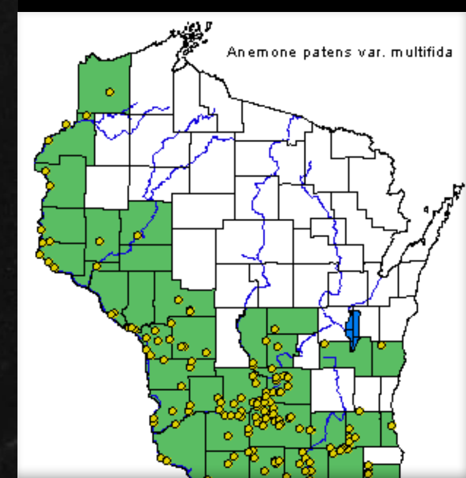
The New York Botanical Garden

Benefits of Digitization

- Preservation of rare/fragile specimens
- Easier to query a database than physical collections (e.g., WISFLORA)
- Efficient data sharing
- Collation yields powerful data repositories

Digitization of Distribution Data





Examples of data repositories

1. Wisconsin Flora – more on this later
2. North America Digital Flora
3. Biota of North America Project
4. Global Biodiversity Information Facility

North America Digital Flora

- Pros:
  - Great for generating species lists
  - Can query by species, attribute, or geography
  - Incorporates interesting GIS layers
- Cons:
  - Limited to wetland species
  - Geography search lacks precision

BONAP

- Pros:
  - Comprehensive of NA diversity
  - Great, easy to use distribution maps
- Cons:
  - Can't query the data
  - Can't download data
  - North America only
  - Poor resolution in Canada

GBIF

- Pros
  - Global data of plants, animals, fungi
  - Database is fully searchable
  - Data are downloadable
- Cons
  - Data are incomplete (WIS not in yet)
  - Some records are suspect

What can you do with these data?

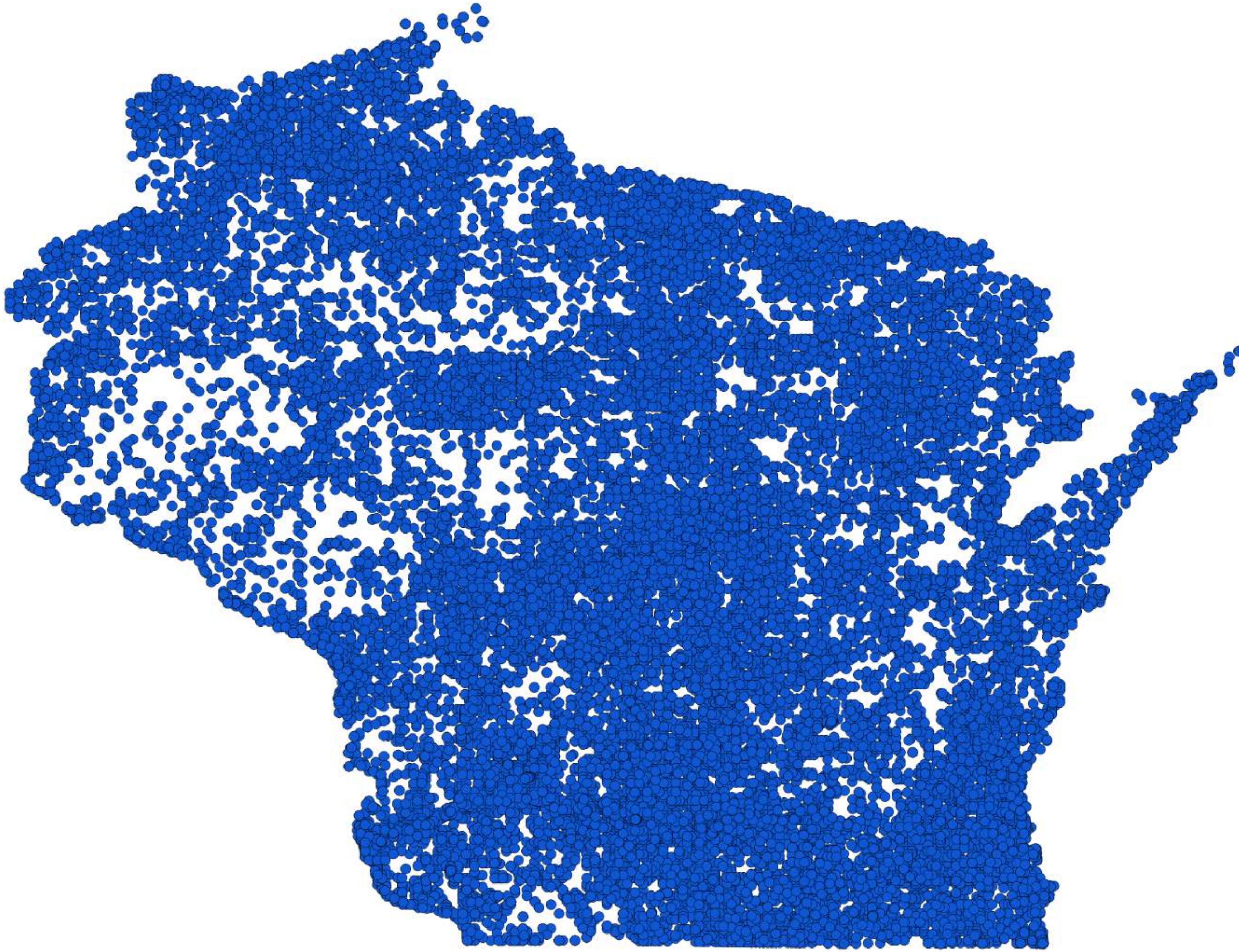
- Test biogeographic patterns in Wiscosnin
- Historical biogeography
- Characterize species distributions
- Obtain species “climatic envelope”
- Identify areas exhibiting high levels of endemism
- Predict responses to global climate change
- Track and predict spread of invasive species

All without spending $1

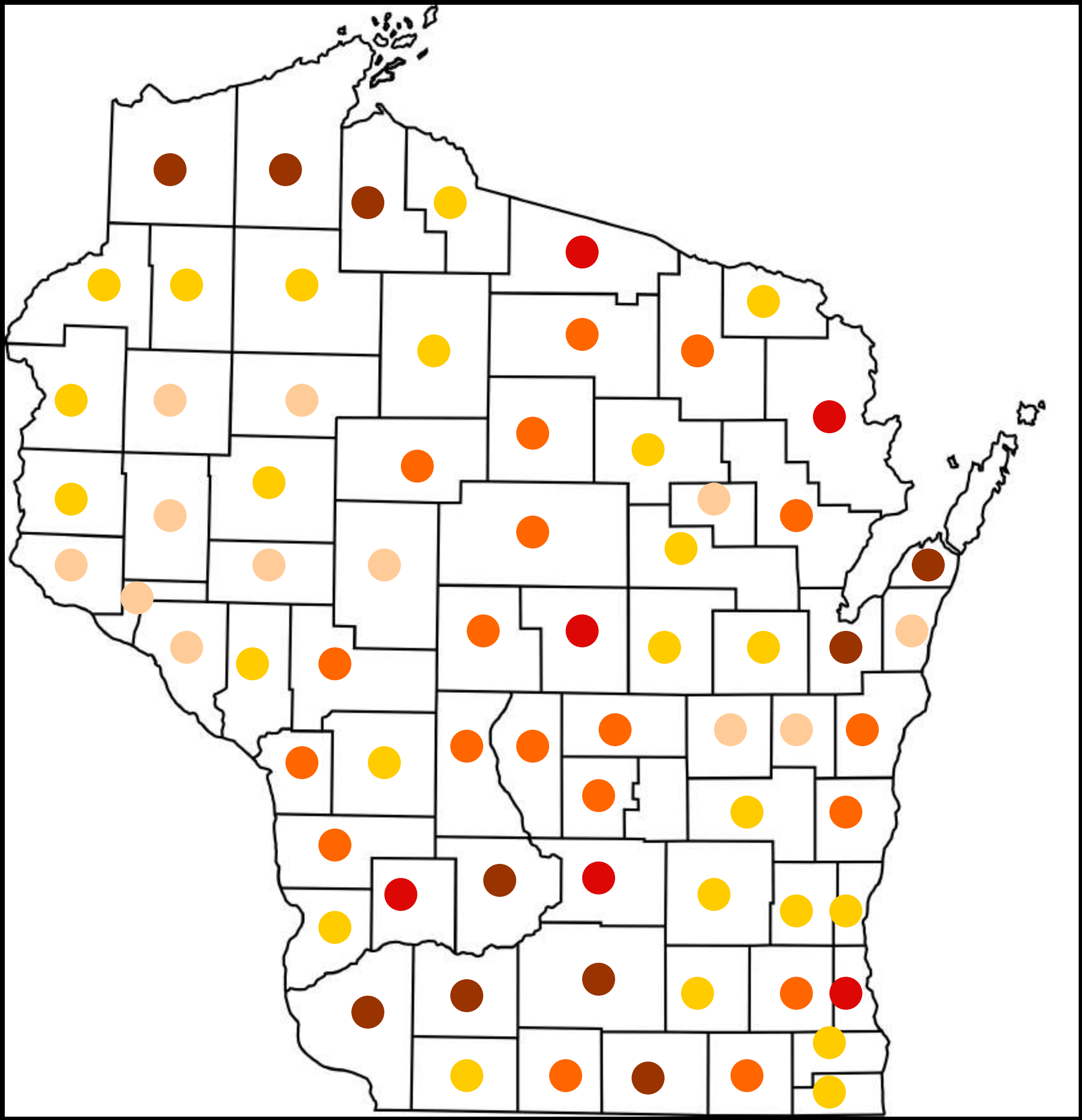
Wisconsin Flora Mapping Project

2,570 species—
1,890 native
680 introduced

WIS Herbarium—
300,000 specimens
200,000 databased



Wisconsin Flora Mapping Project

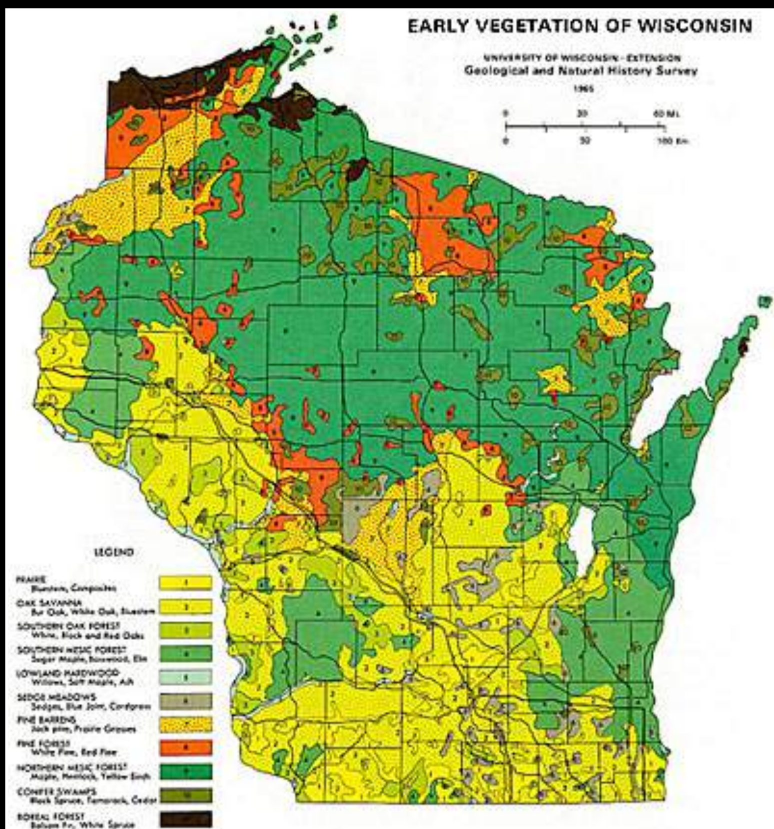


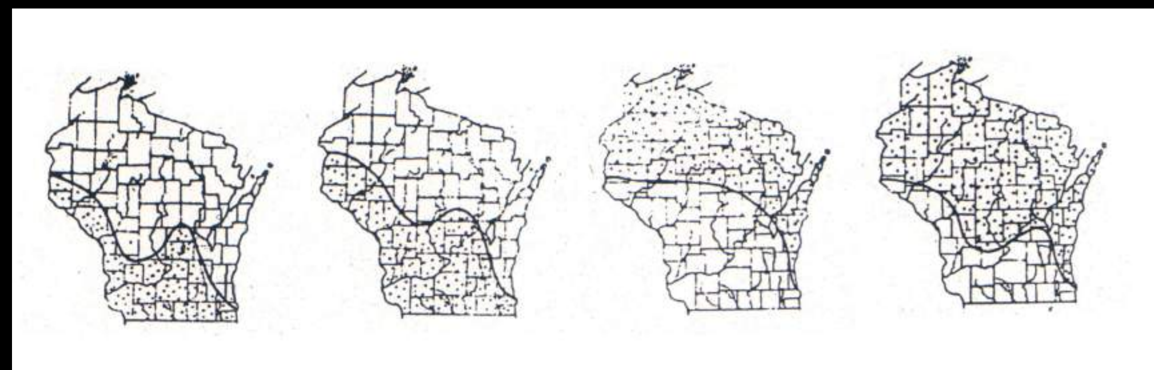
Wisconsin is not equally inventoried

- > 10,000 specimens
- 8,000 - 6,000
- 6,000 - 4,000
- 2,000 - 4,000
- < 2,000

Wisconsin Flora Mapping Project

The tension zone separating the two provinces is based on the upper and lower limits of the southern and northern species, respectively.



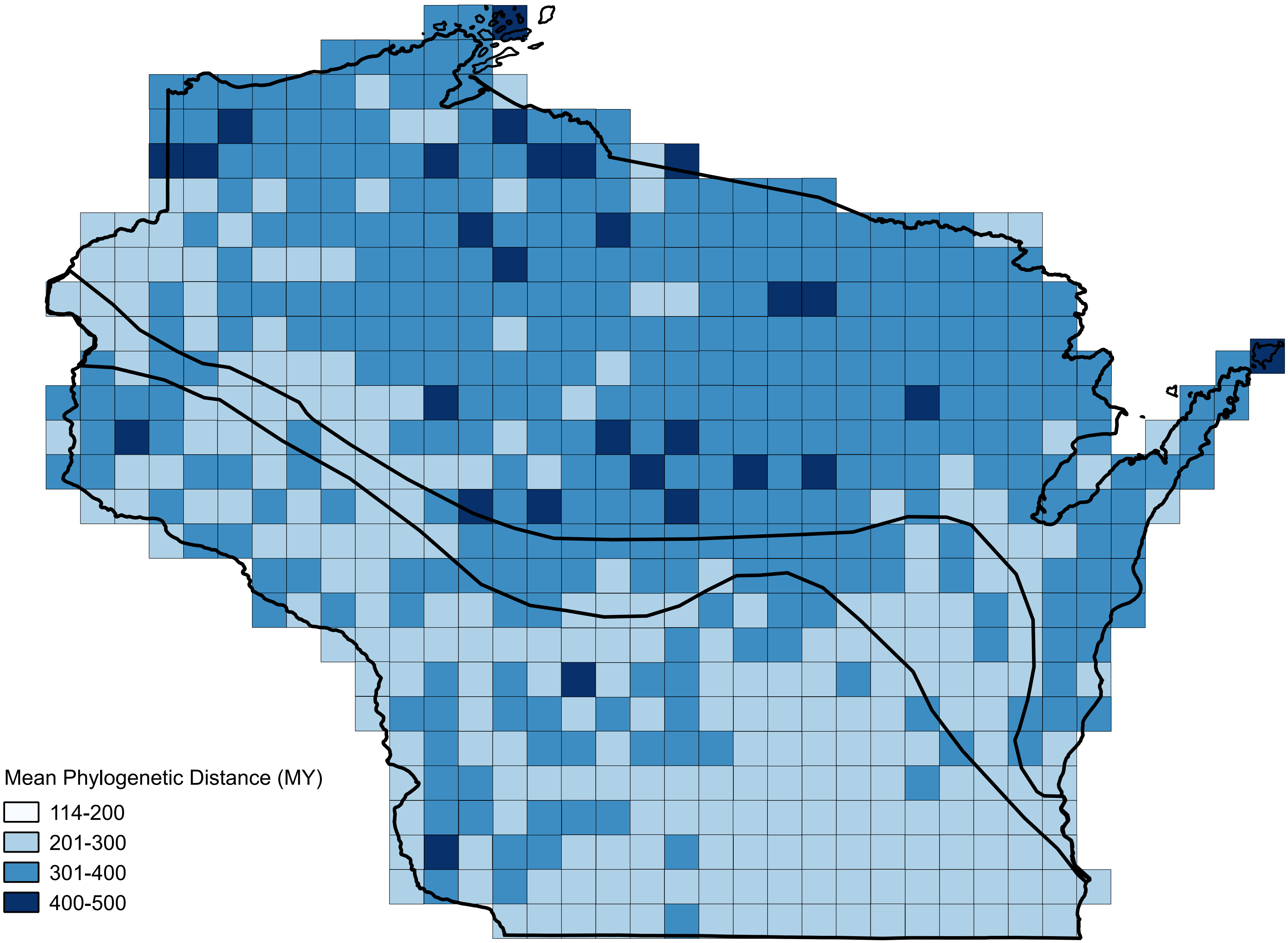


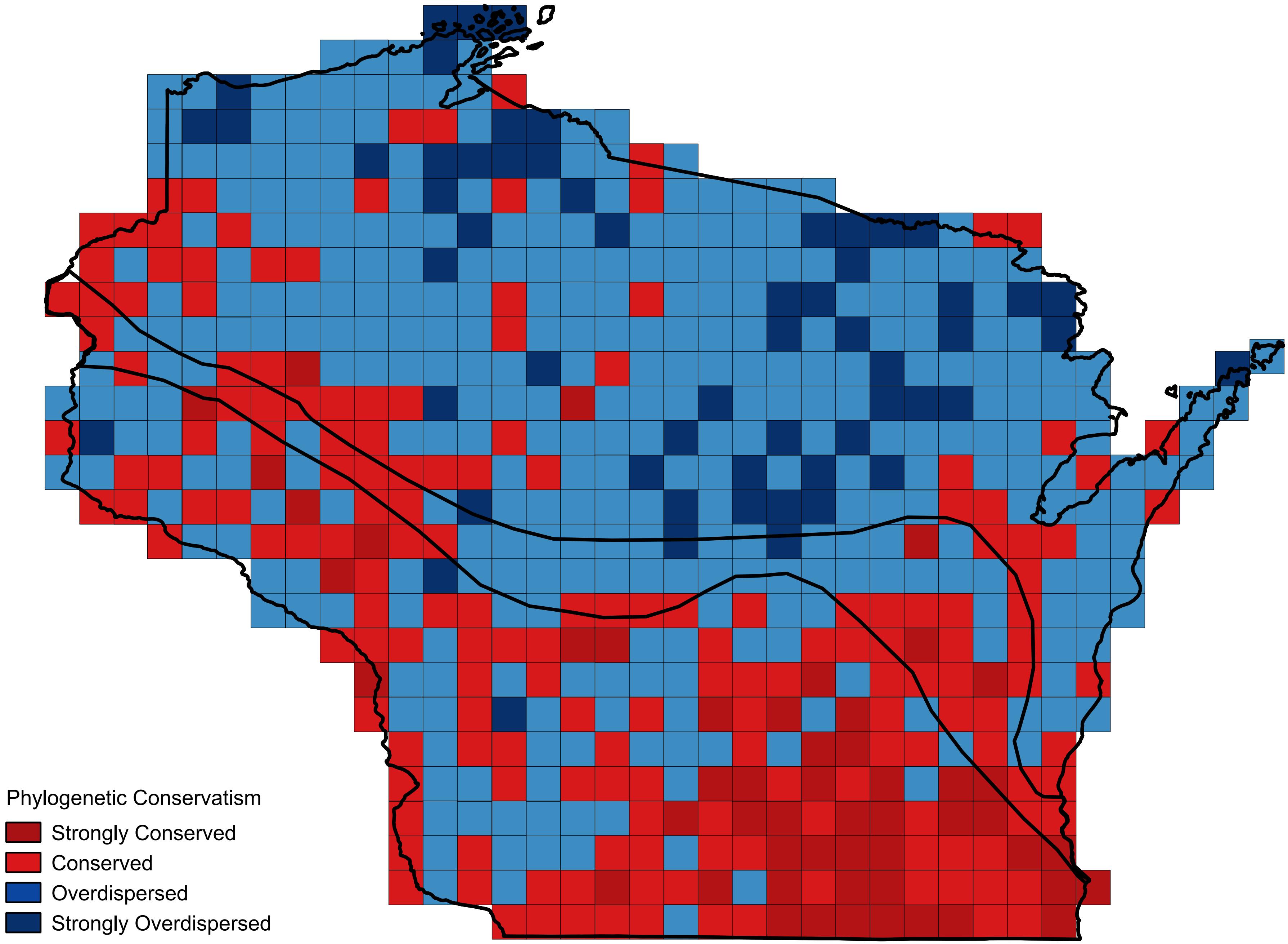
Golden cassia

Wild indigo

Ram’s head ladyslipper

Stemless ladyslipper


Mean Phylogenetic Distance (MY)
114-200
201-300
301-400
400-500



Phylogenetic Conservatism
Strongly Conserved
Conserved
Overdispersed
Strongly Overdispersed

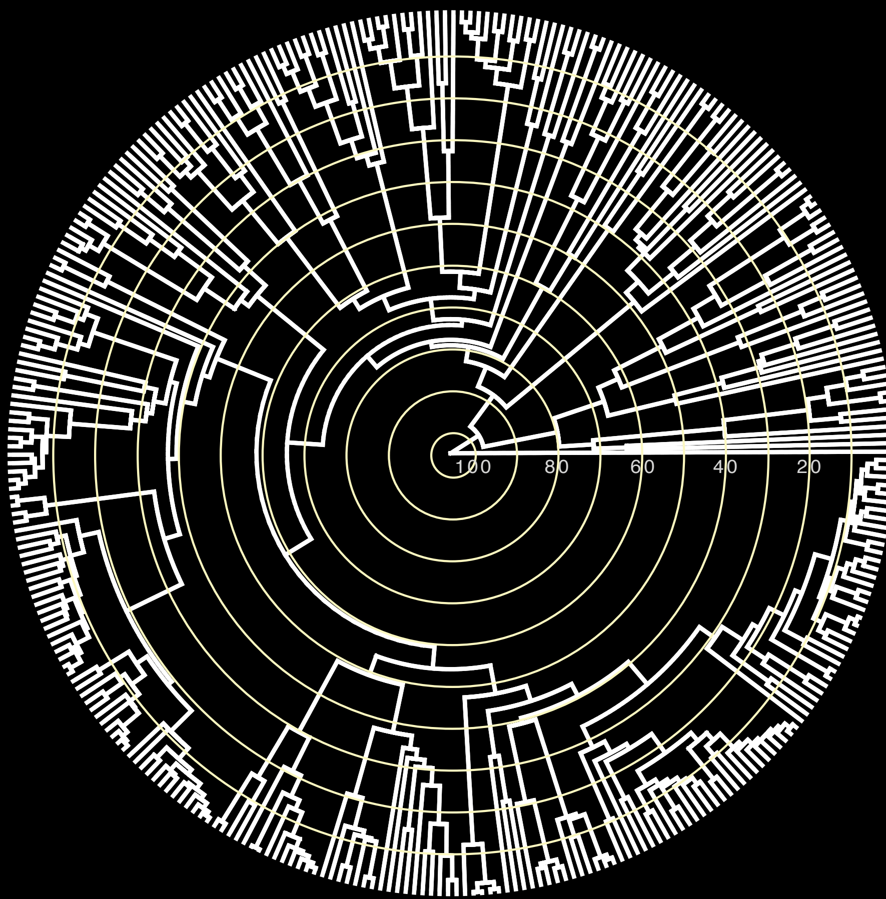
Historical Biogeography

- If you wanted, you could build a phylogeny of 300,000 organisms using 187 billion base pairs from Genbank

e.g., family Cyperaceae (sedges, bulrushes) – Spalink et al. 2016

Historical Biogeography

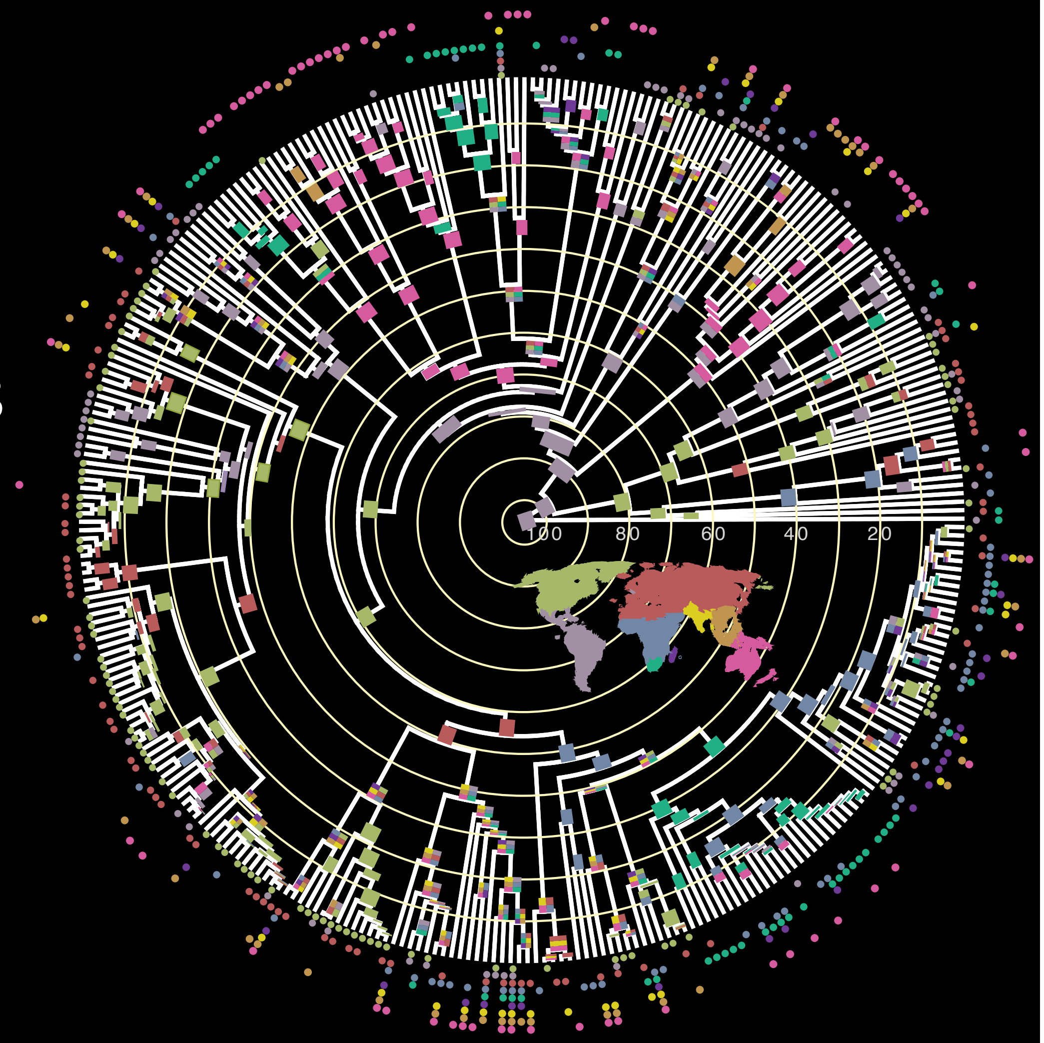
- Build a phylogeny
  - Date the phylogeny



Data from Genbank , analyzed using BEAST

Historical Biogeography

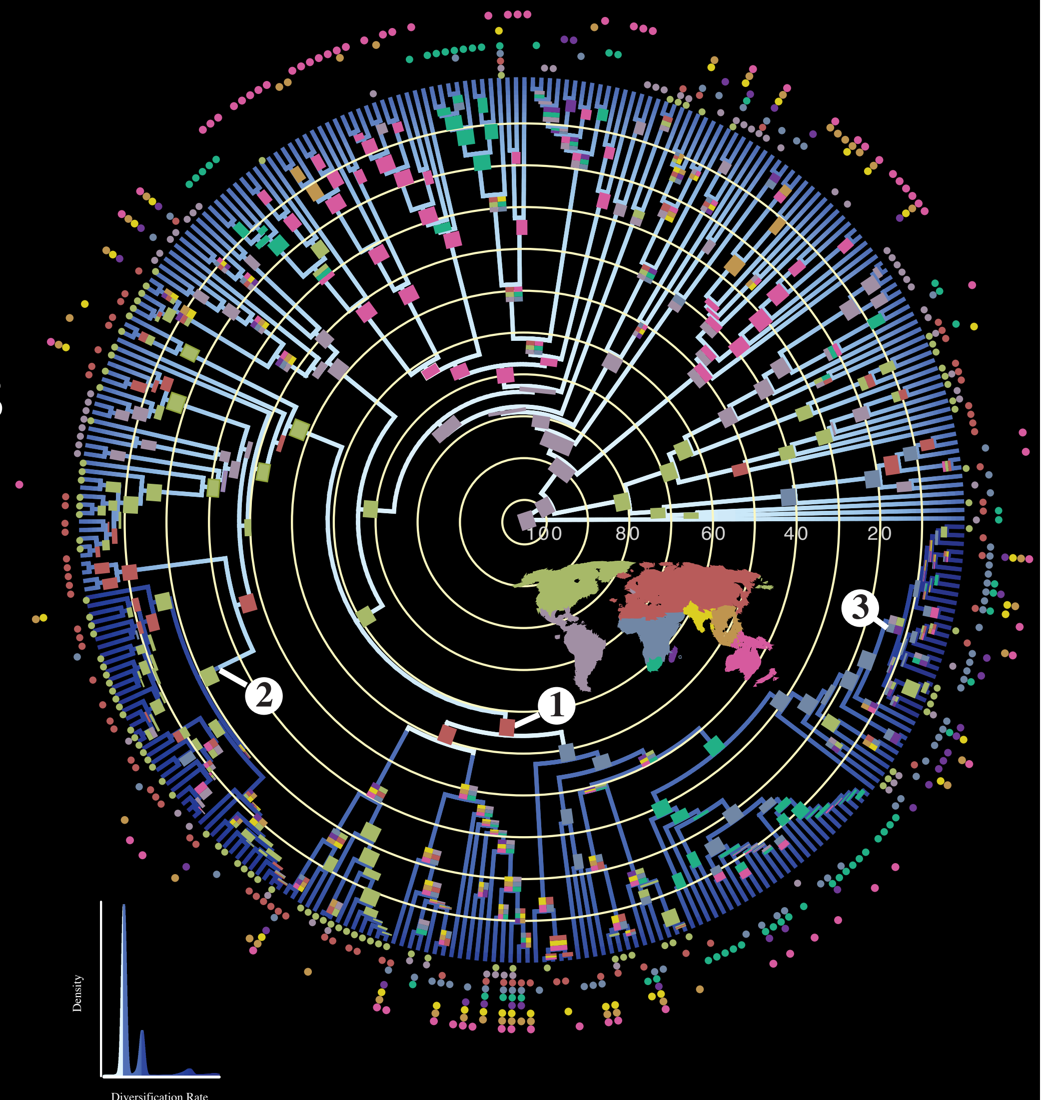
- Build a phylogeny
  - Date the phylogeny
  - Reconstruct areas inhabited by ancestors



Data from Genbank + eMonocot, analyzed using BEAST and BioGeoBEARS

Historical Biogeography

- Build a phylogeny
  - Date the phylogeny
  - Reconstruct areas inhabited by ancestors
  - Measure rates of diversification
  - Track evolution of morphological traits across time and space



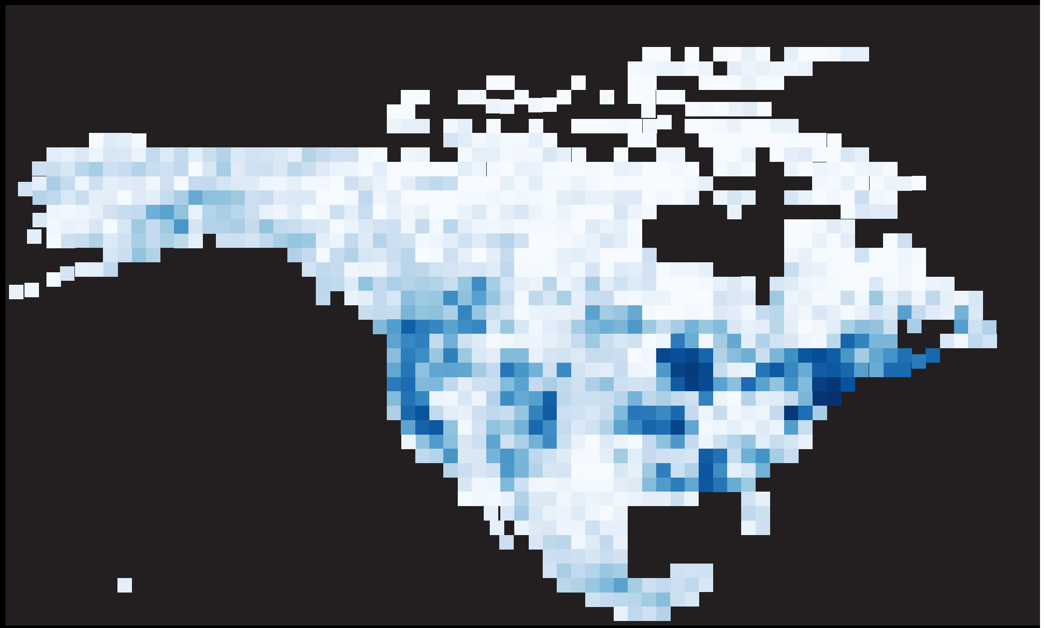
Data from Genbank + eMonocot, analyzed using BEAST,BioGeoBEARS and BAMM

High Resolution Species Distributions

- Characterize distributions of species for 1.45 million species using 460 million geo-referenced occurrences

High Resolution Species Distributions

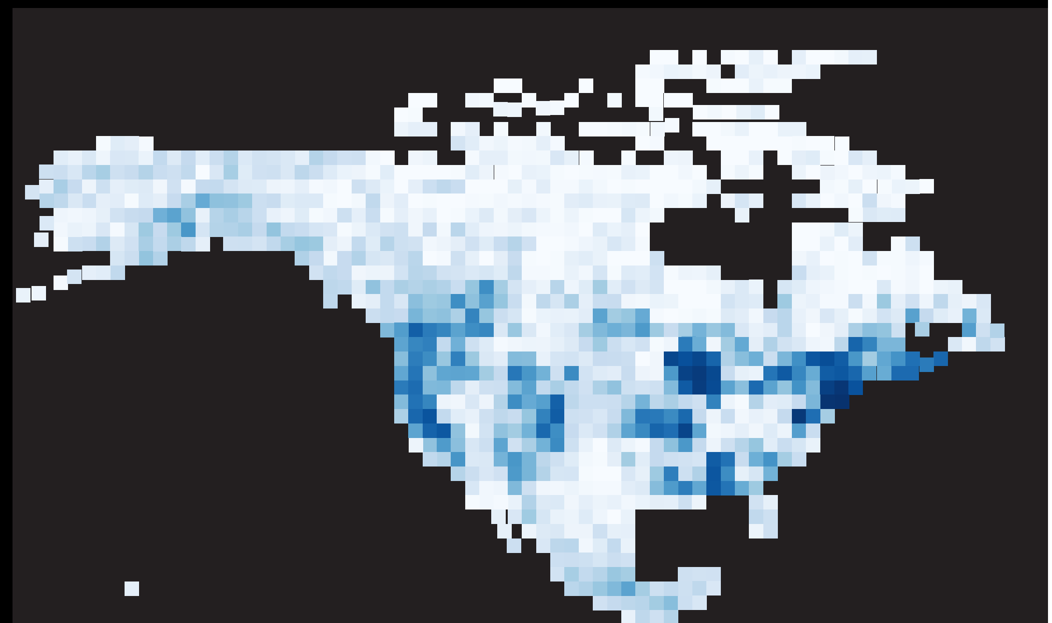
- Characterize distributions of species for 1.45 million species using 460 million geo-referenced occurrences
  - Identify species rich areas

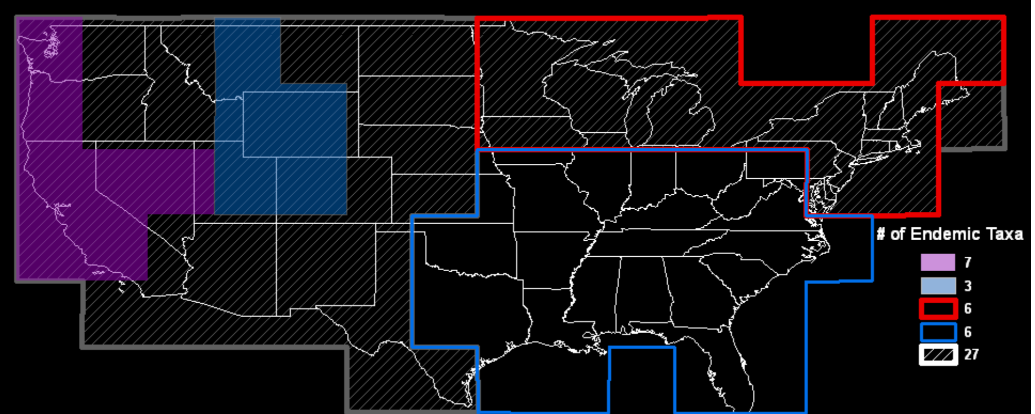


Data from GBIF, analyzed in QGIS

High Resolution Species Distributions

- Characterize distributions of species for 1.45 million species using 460 million geo-referenced occurrences
  - Identify species rich areas
  - Define regions with high levels of rare or endemic taxa





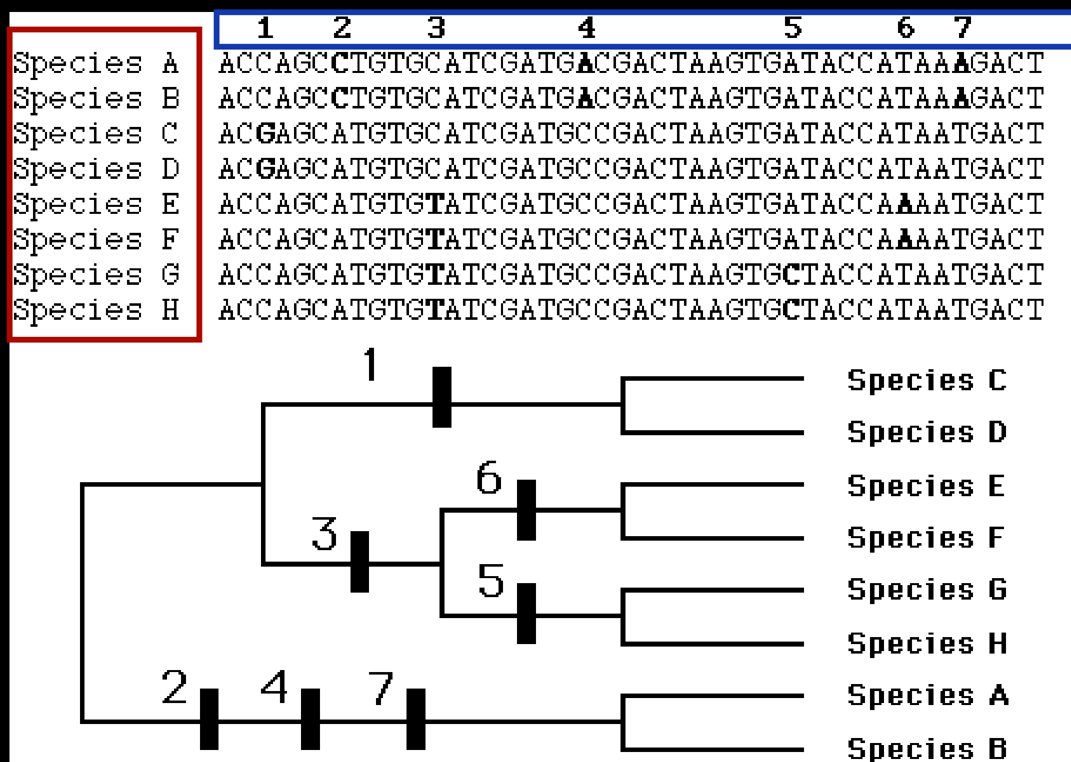
Data from GBIF, analyzed in QGIS and PAUP* - available in systematics lab

Parsimony Analysis of Endemicity (PAE)

Parsimony: typically used in phylogenetics…

Species

Characters

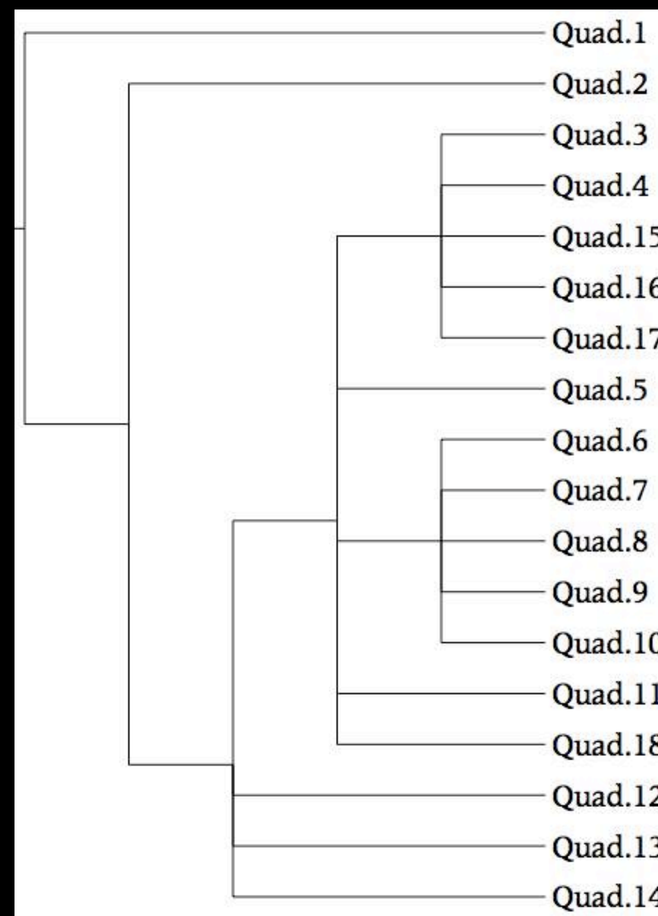


Parsimony Analysis of Endemicity (PAE)

…but, can be applied to PAE

	Species A	Species B	Species C	Species D	Species E	Species F	Species G	Species H
Area 1	1	0	1	0	0	0	0	0
Area 2	1	0	1	0	0	0	0	0
Area 3	0	1	0	1	0	0	0	0
Area 4	0	1	0	1	0	0	0	0
Area 5	0	0	0	0	0	0	0	0
Area 6	0	0	0	0	1	1	0	0
Area 7	0	0	0	0	1	1	0	0
Area 8	0	0	0	0	1	1	0	0
Area 9	0	0	0	0	1	1	0	0
Area 10	0	0	0	0	1	1	0	0
Area 11	0	0	0	0	0	0	0	0
Area 12	1	0	0	0	0	0	0	0
Area 13	1	0	0	0	0	0	0	0
Area 14	1	0	0	0	0	0	0	0
Area 15	0	1	0	1	0	0	0	0
Area 16	0	1	0	1	0	0	0	0
Area 17	0	1	0	1	0	0	0	0
Area 18	0	0	0	0	0	0	0	0

where geographic Areas are treated as taxa, and the presence/absence of the species serve as Characters

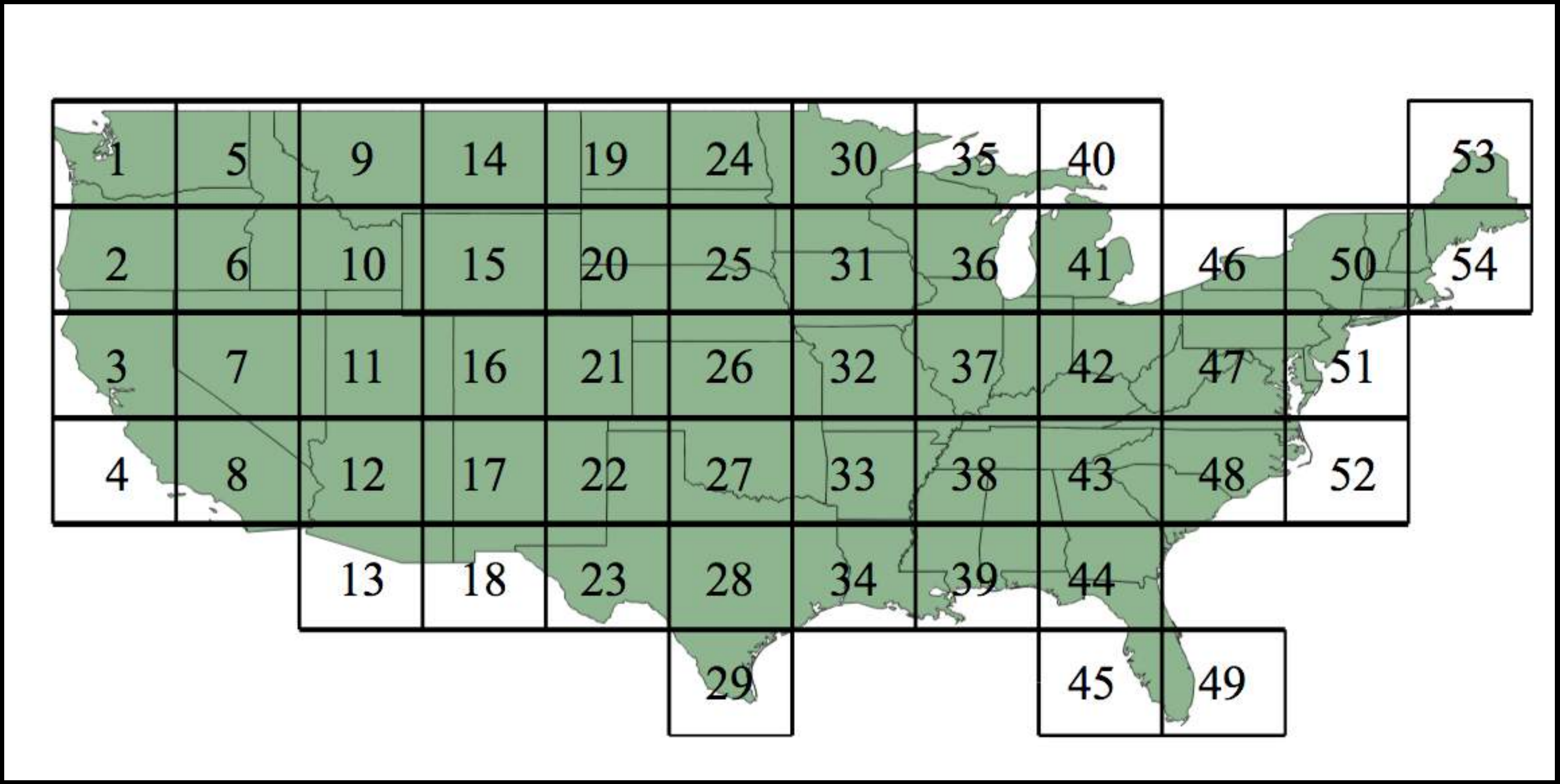


Case Study: finding areas of endemism in North American sedges

Step 1: Get distributional data of all species in N Am

Case Study: finding areas of endemism in North American sedges

Step 2: Define geographic region and divide into quadrats



Case Study: finding areas of endemism in North American sedges

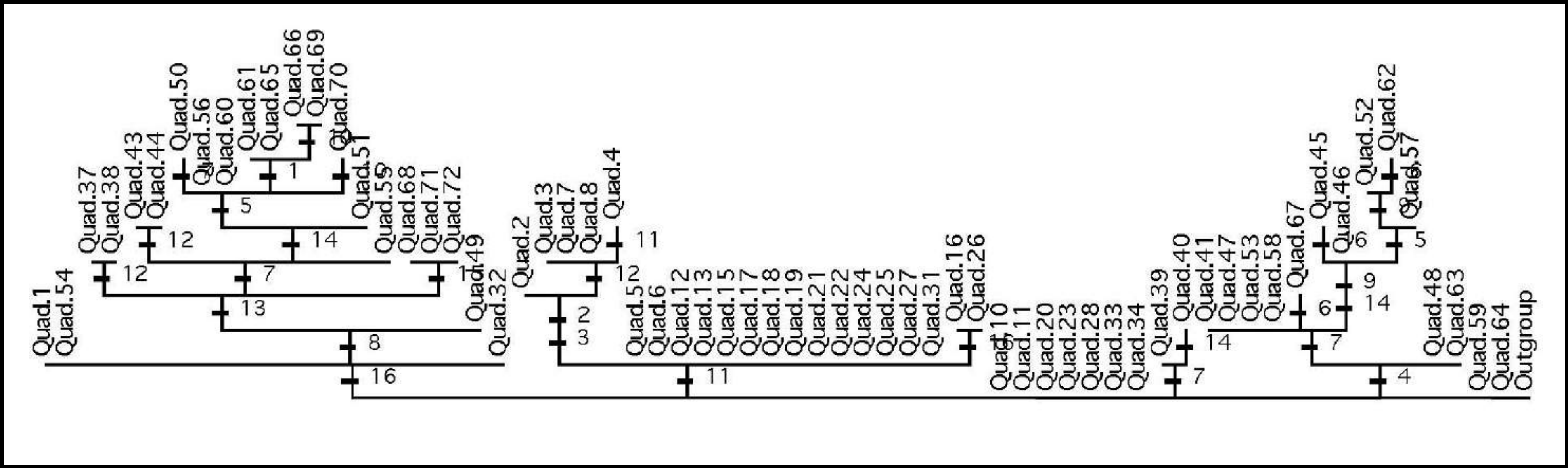
Step 3: Construct data matrix: absence = 0, presence = 1

Taxa	Characters	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
		Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Sci	Eri	Eri	Eri	Eri	Eri	Eri	Eri	Eri	Eri	Eri	Kot	Kot	Kot	Car
1	Quad.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
2	Quad.2	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
3	Quad.3	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
4	Quad.4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	Quad.5	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0
6	Quad.6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0
7	Quad.7	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
8	Quad.8	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
9	Quad.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	Quad.10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	1	0
11	Quad.11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
12	Quad.12	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Quad.13	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Quad.14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	Quad.15	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	1	0
16	Quad.16	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	1	0	1	0	0	0	1	1	1	1	0
17	Quad.17	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0
18	Quad.18	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	Quad.19	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	Quad.20	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	Quad.21	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1	1	1	0
22	Quad.22	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	1	1	0
23	Quad.23	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0
24	Quad.24	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	Quad.25	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
26	Quad.26	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
27	Quad.27	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0
28	Quad.28	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
29	Quad.29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	Quad.30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	Quad.31	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
32	Quad.32	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
33	Quad.33	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
34	Quad.34	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	Quad.35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	Quad.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	Quad.37	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	1	0	1	1	1	1	0	0	0	0
38	Quad.38	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	0	0
39	Quad.39	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	Quad.40	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	Quad.41	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Taxa	Characters	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
		Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car	Car
1	Quad.1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
2	Quad.2	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
3	Quad.3	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1
4	Quad.4	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
5	Quad.5	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
6	Quad.6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
7	Quad.7	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
8	Quad.8	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1
9	Quad.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
10	Quad.10	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	1
11	Quad.11	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1
12	Quad.12	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
13	Quad.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1
14	Quad.14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
15	Quad.15	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
16	Quad.16	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
17	Quad.17	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1
18	Quad.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1
19	Quad.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
20	Quad.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1
21	Quad.21	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1
22	Quad.22	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1
23	Quad.23	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1
24	Quad.24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0
25	Quad.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
26	Quad.26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1
27	Quad.27	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1
28	Quad.28	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1
29	Quad.29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
30	Quad.30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
31	Quad.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
32	Quad.32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1
33	Quad.33	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
34	Quad.34	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
35	Quad.35	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0
36	Quad.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
37	Quad.37	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
38	Quad.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
39	Quad.39	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
40	Quad.40	0	1	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
41	Quad.41	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0

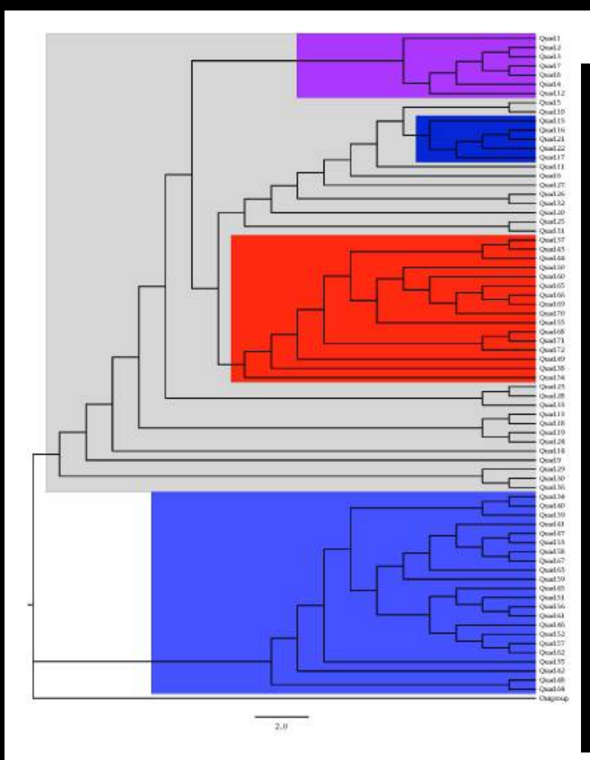
Case Study: finding areas of endemism in North American sedges

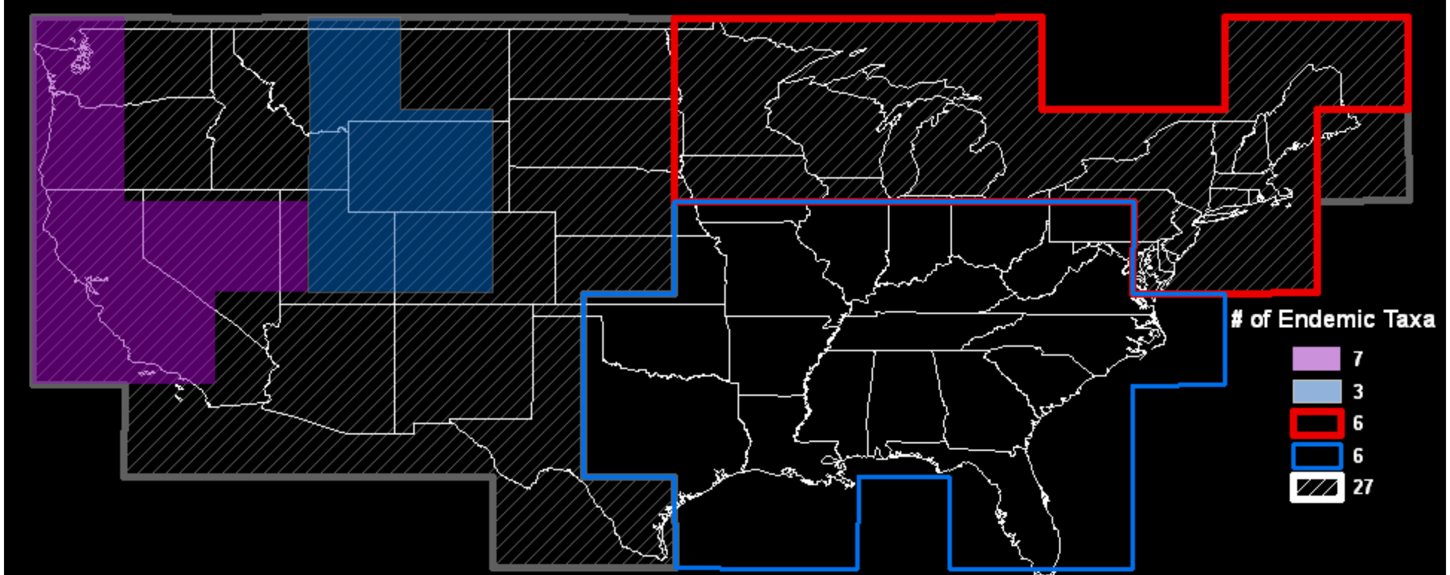
Step 4: Conduct parsimony analysis



Case Study: finding areas of endemism in North American sedges

Step 5: Find clades consisting of 2+ endemic taxa with congruent distributions.





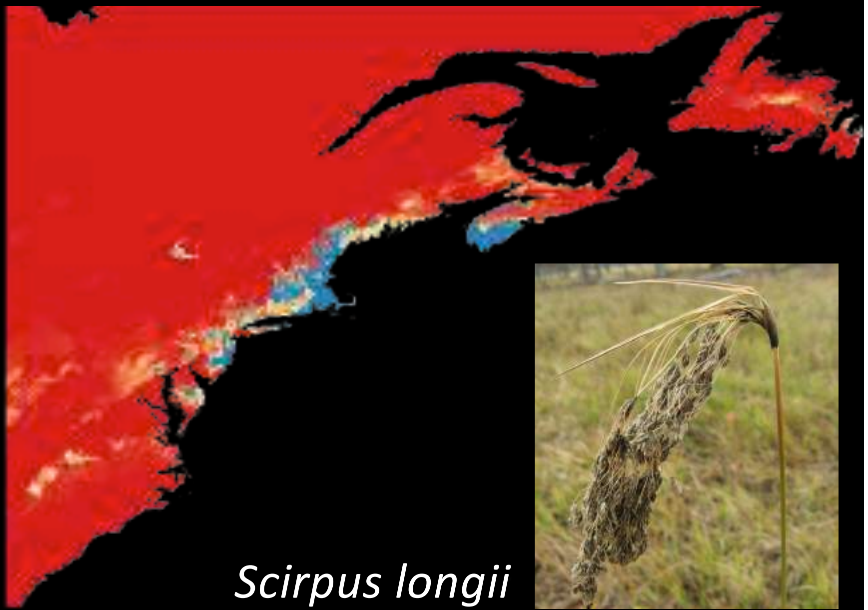
Ecological Niche Modeling

- Develop ecological niche models using waypoint and climate data (BioClim)

Scirpus longii

Ecological Niche Modeling

- Develop ecological niche models using waypoint and climate data
  - Identify suitable areas for endangered species

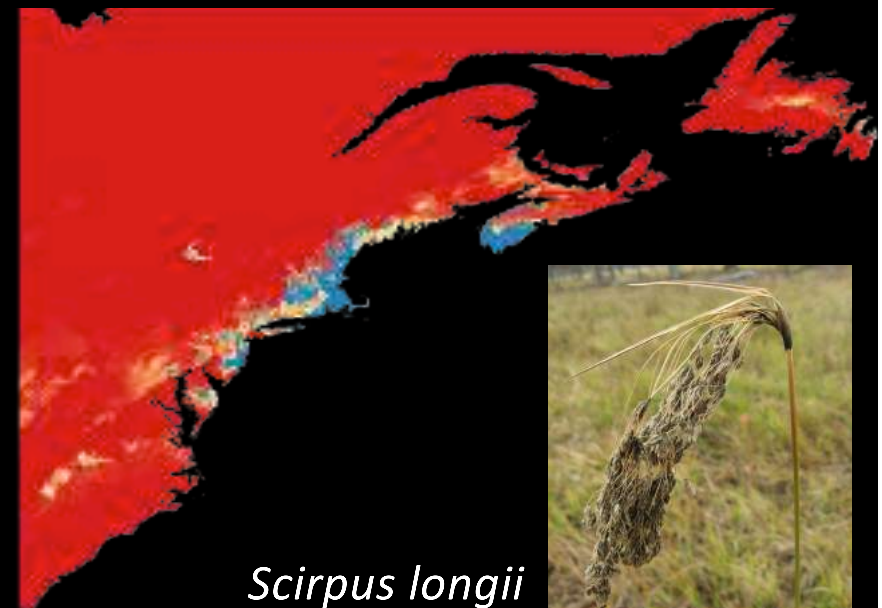


Scirpus longii

Data from GBIF and BioClim, analyzed with Maxent

Ecological Niche Modeling

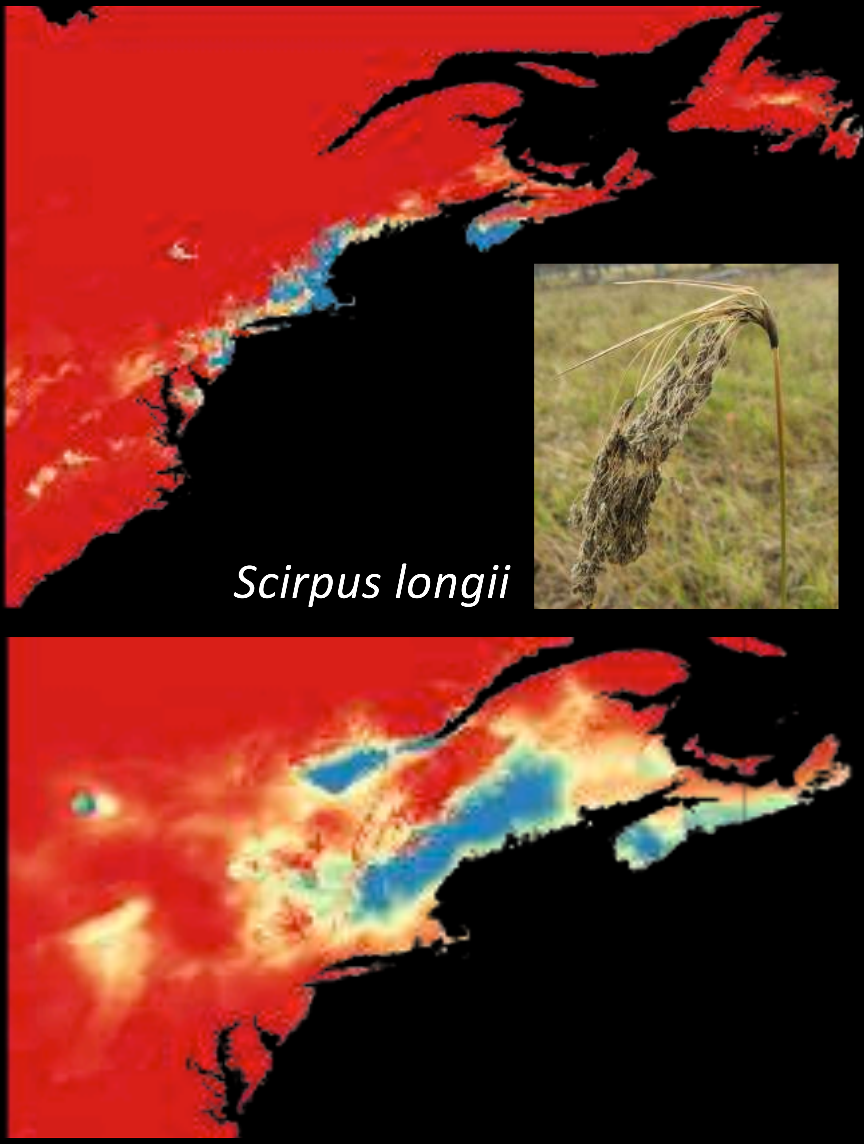
- Develop ecological niche models using waypoint and climate data
  - Identify suitable areas for endangered species
  - Reconstruct distributions during Last Glacial Maximum



Data from GBIF and BioClim, analyzed with Maxent

Without spending $1, you can:

- Develop ecological niche models using waypoint and climate data
  - Identify suitable areas for endangered species
  - Reconstruct distributions during Last Glacial Maximum
  - Predict species distributions in 50 years



Data from GBIF and BioClim, analyzed with Maxent

Track spread of invasive species

- Where did they originate?
- When did they get here?
- Where are they going?

Garlic Mustard and Purple Loosestrife: 1870

Data from GBIF, analyzed in QGIS

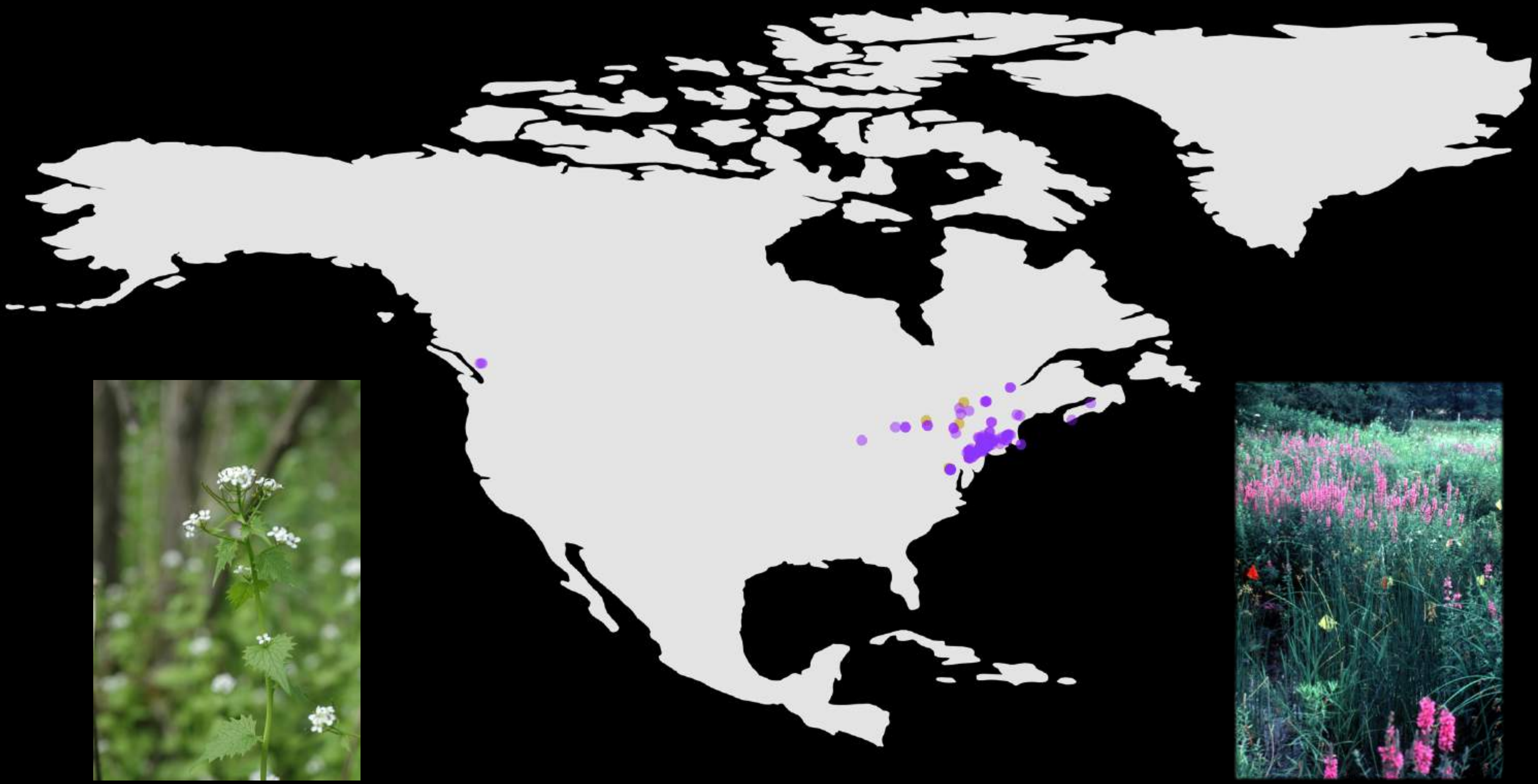
Garlic Mustard and Purple Loosestrife: 1900

Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1910

Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1920



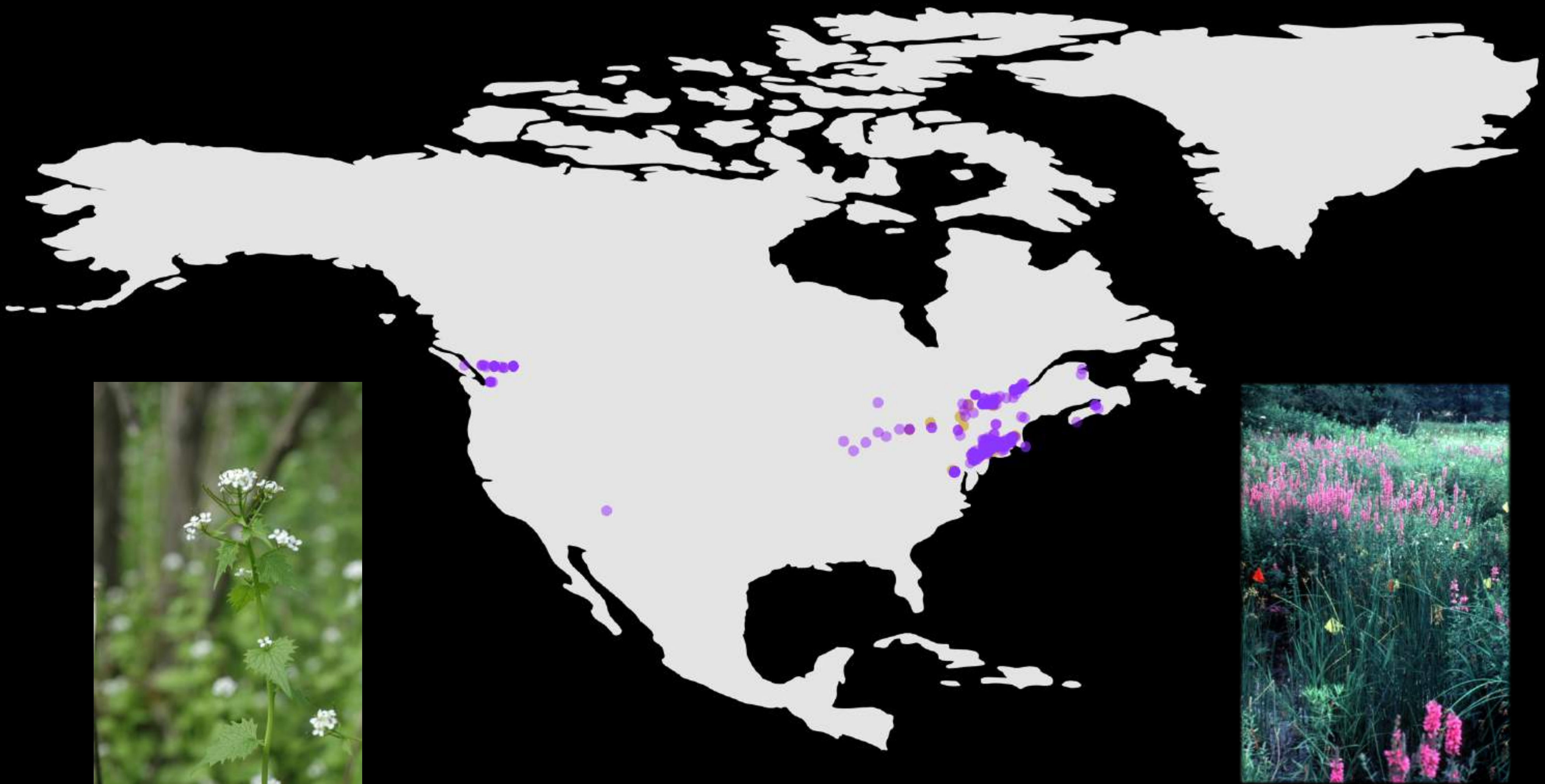
Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1930

Data from GBIF, analyzed in QGIS

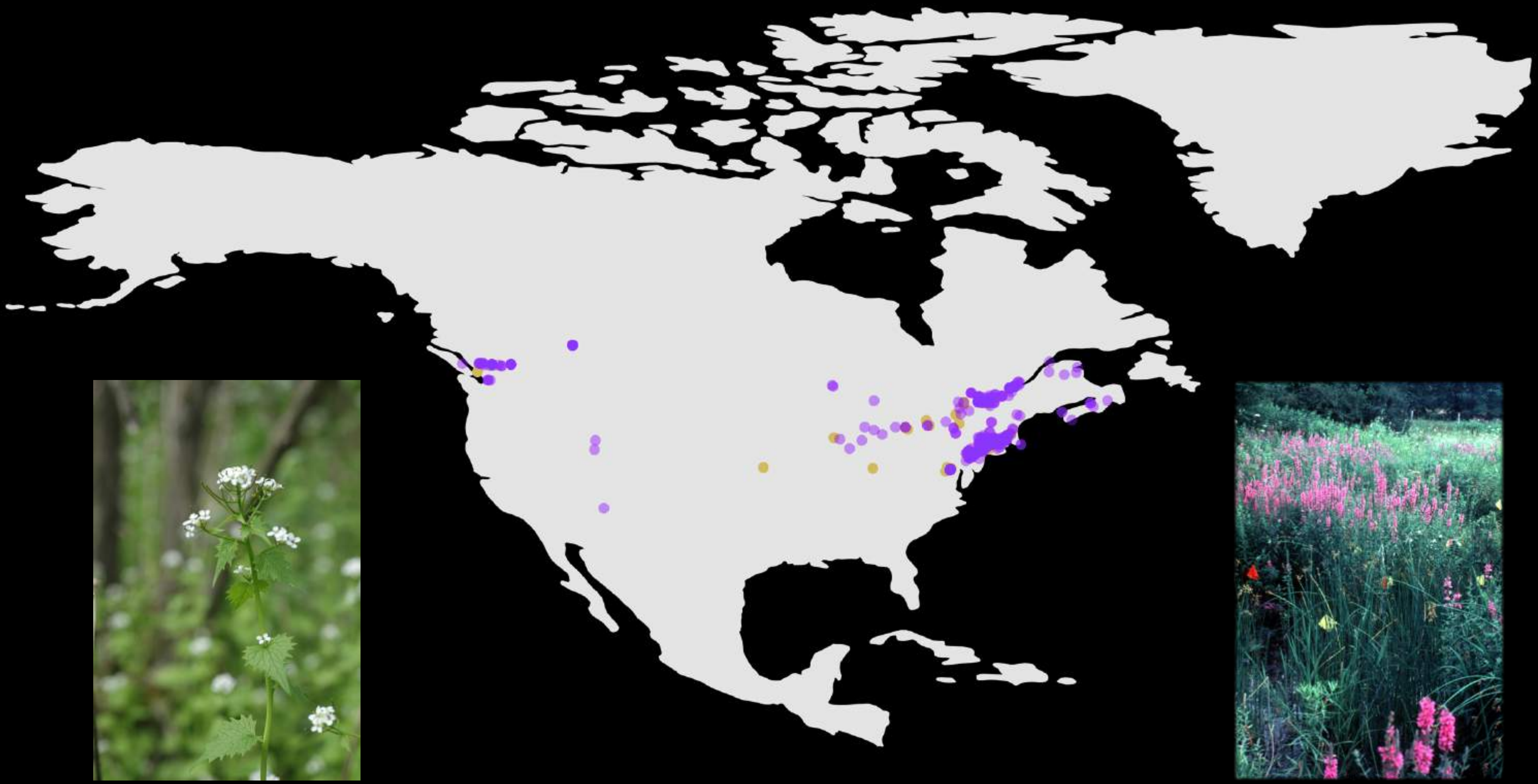
1927 Roadmap

Garlic Mustard and Purple Loosestrife: 1940



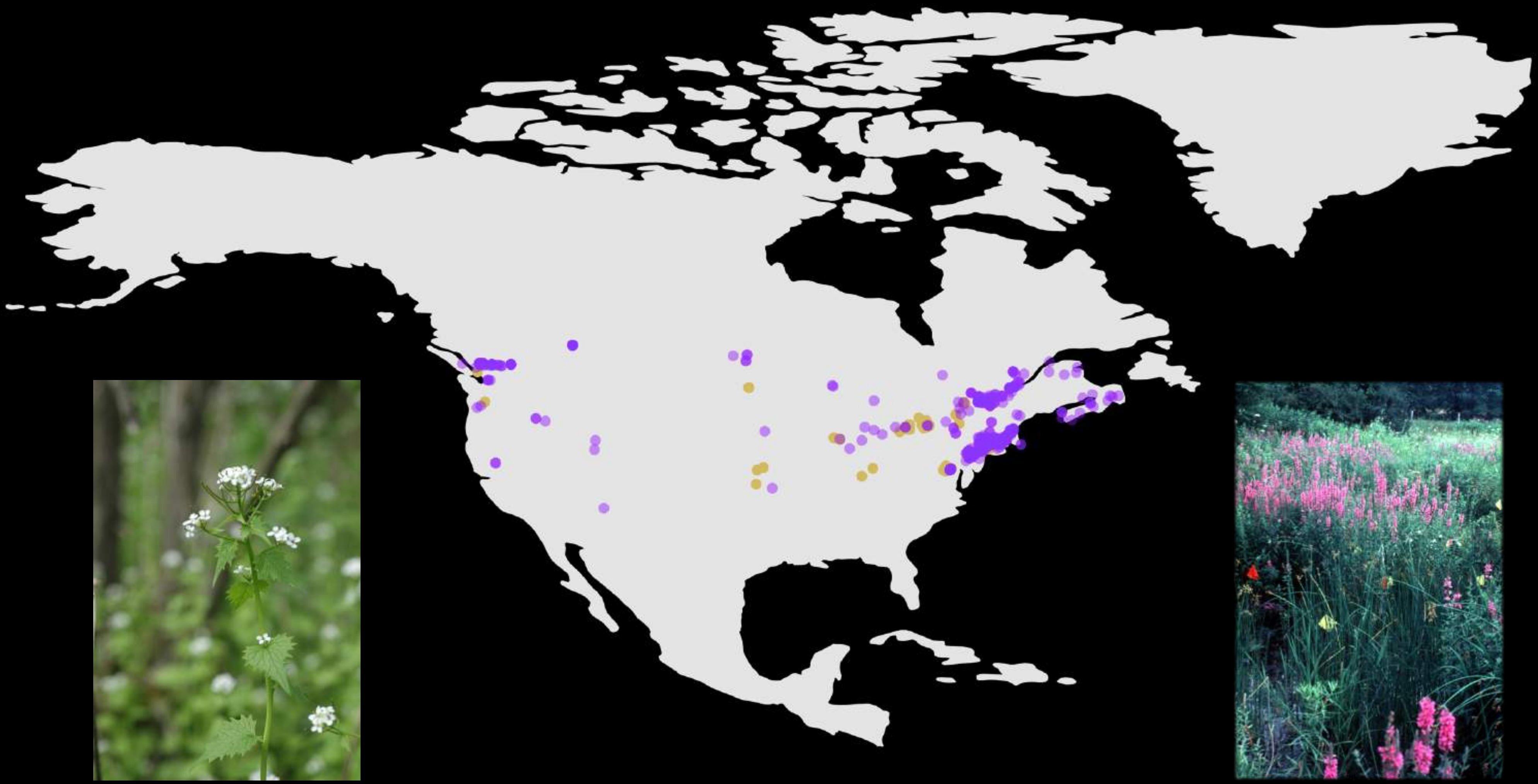
Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1950



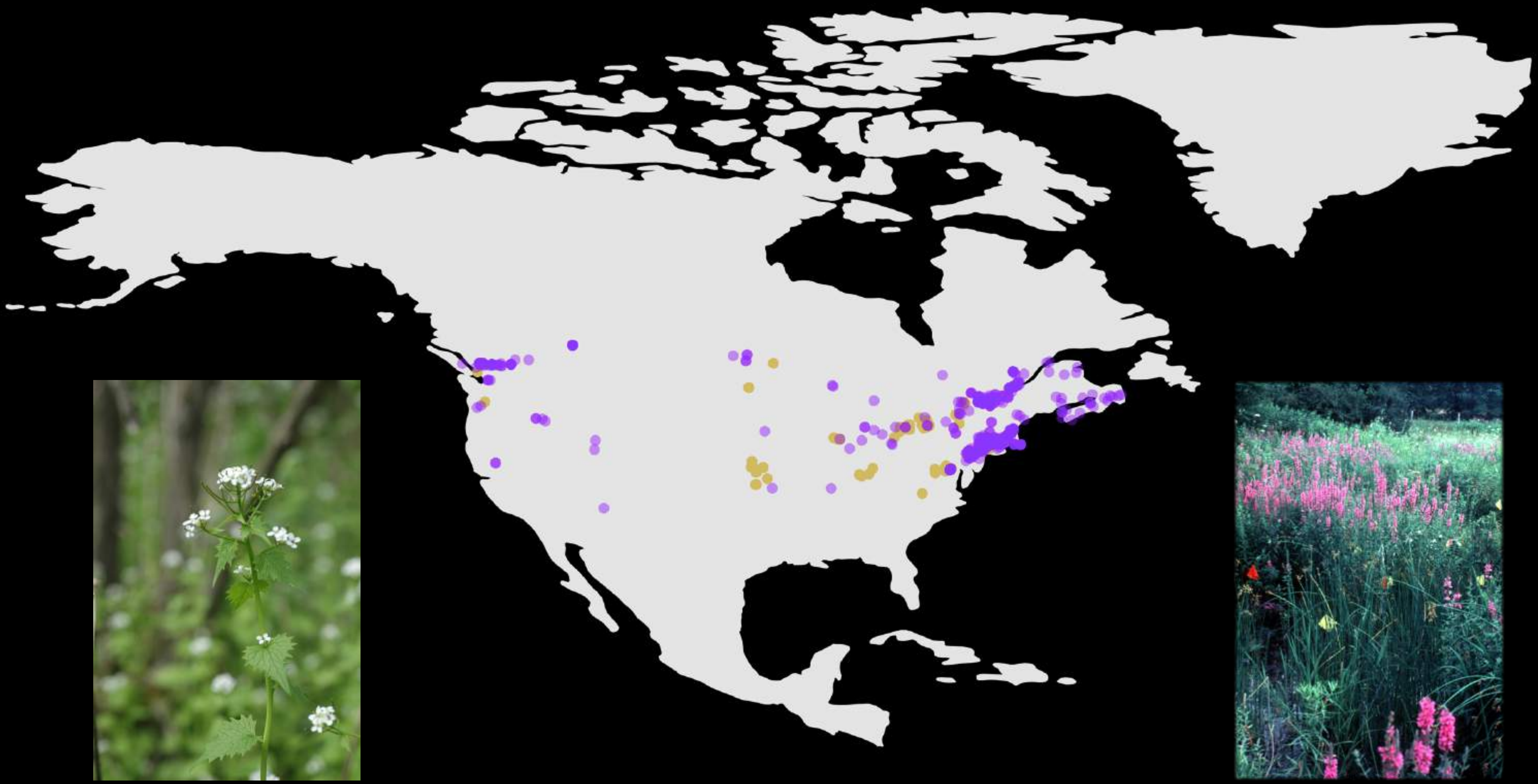
Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1960



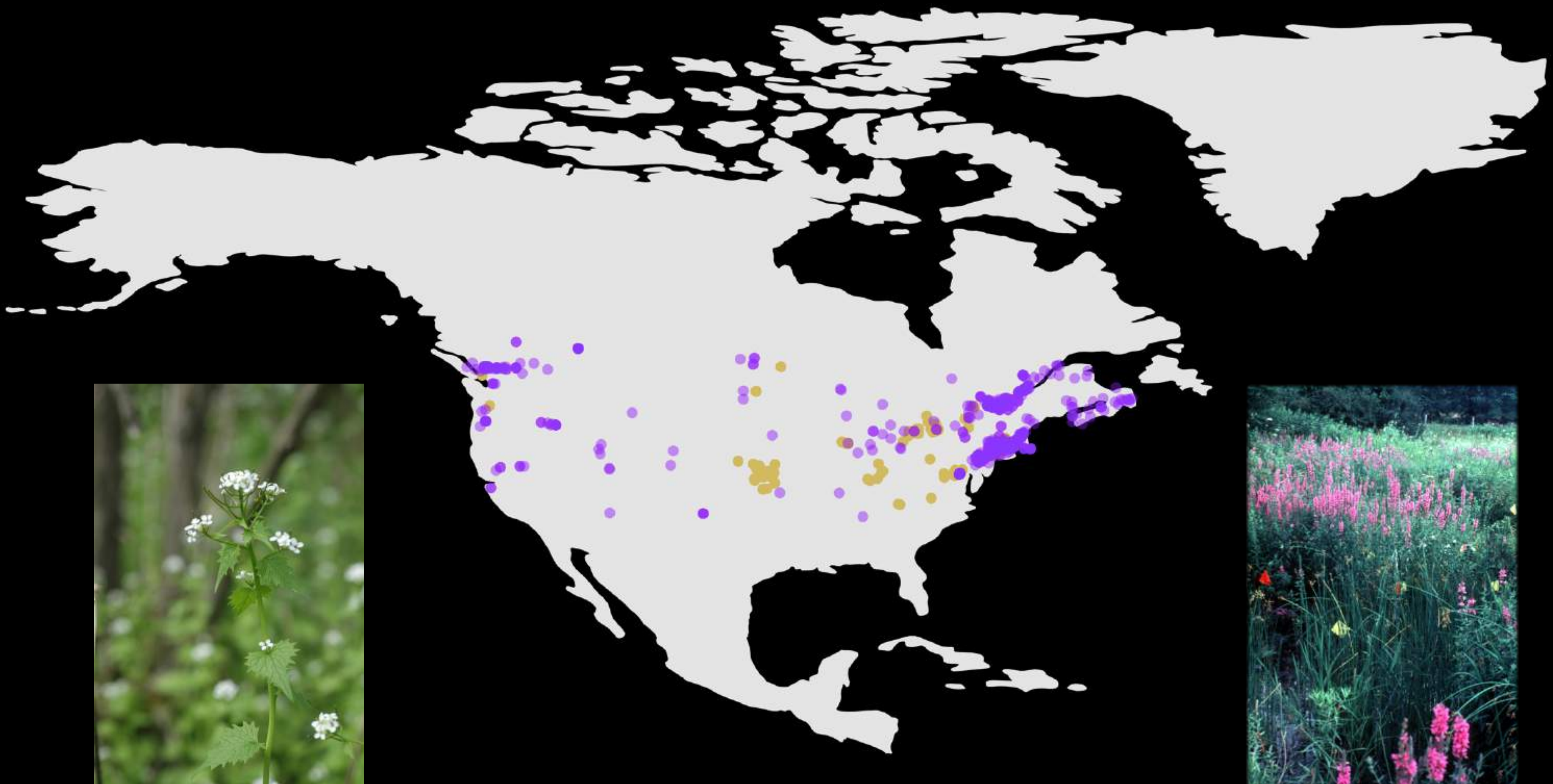
Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1970



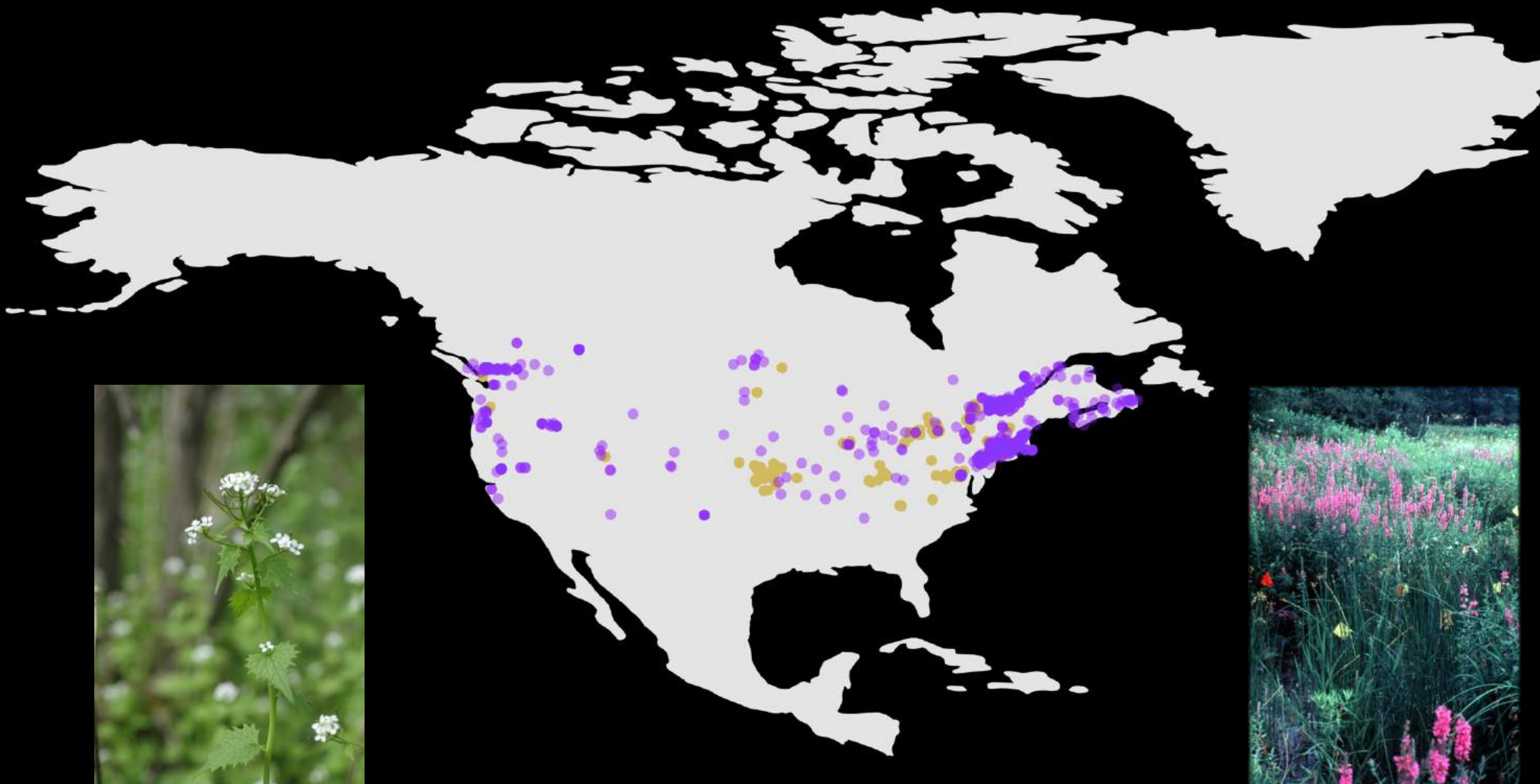
Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: 1980



Data from GBIF, analyzed in QGIS

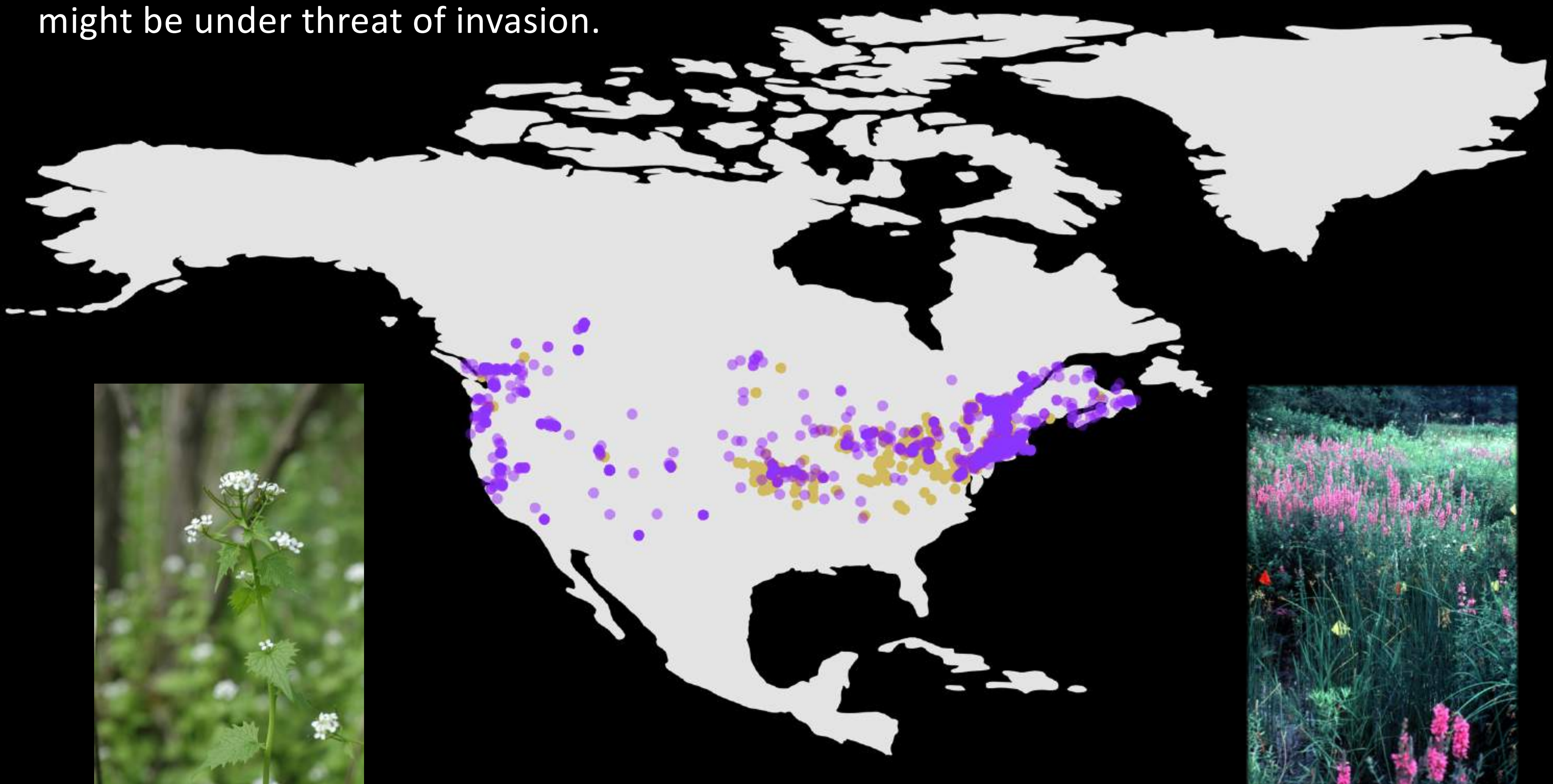
Garlic Mustard and Purple Loosestrife: 1990



Data from GBIF, analyzed in QGIS

Garlic Mustard and Purple Loosestrife: Today

Using ecological niche modeling, you could predict which regions might be under threat of invasion.

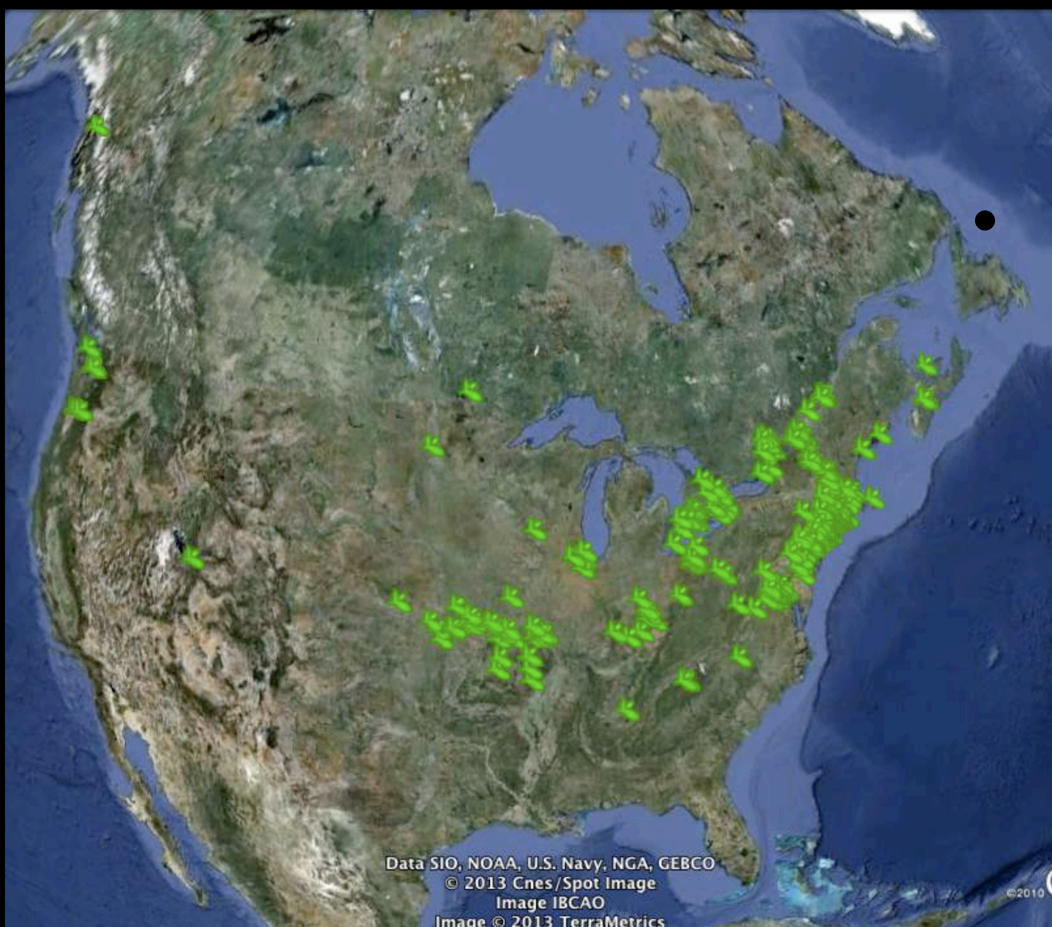


Data from GBIF, analyzed in QGIS

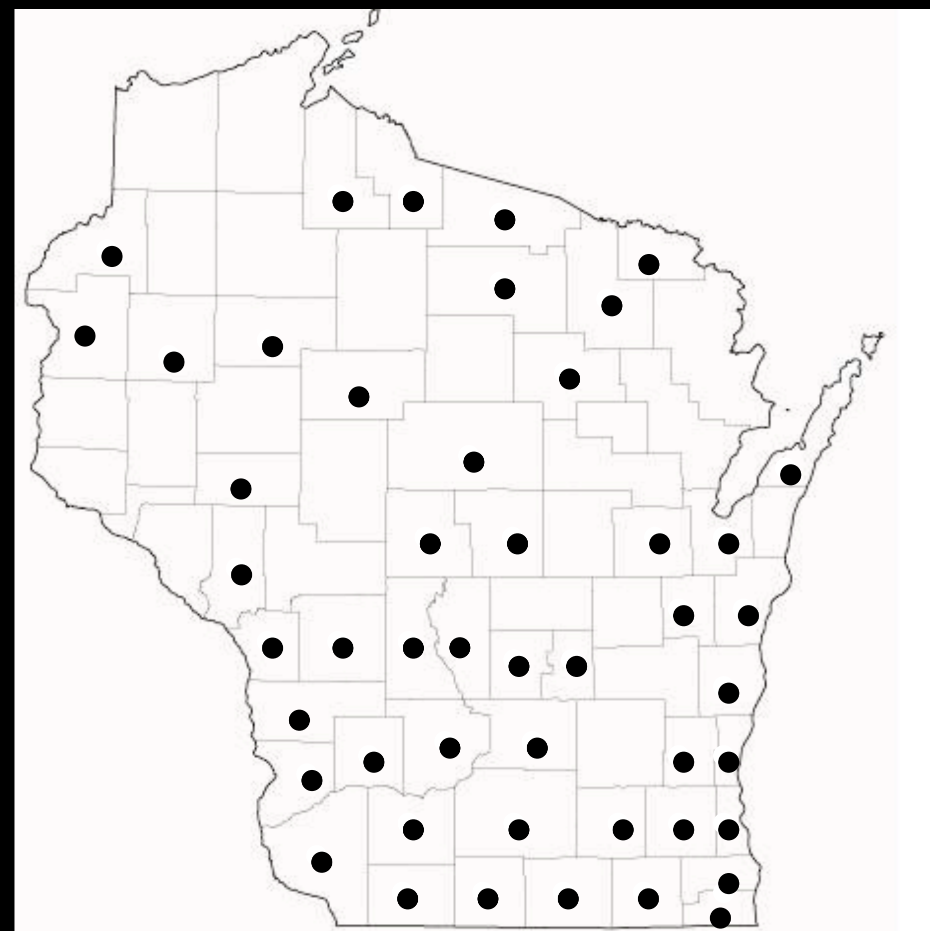
Some Cautions:

Distribution of *Alliaria petiolata* according to:

GBIF

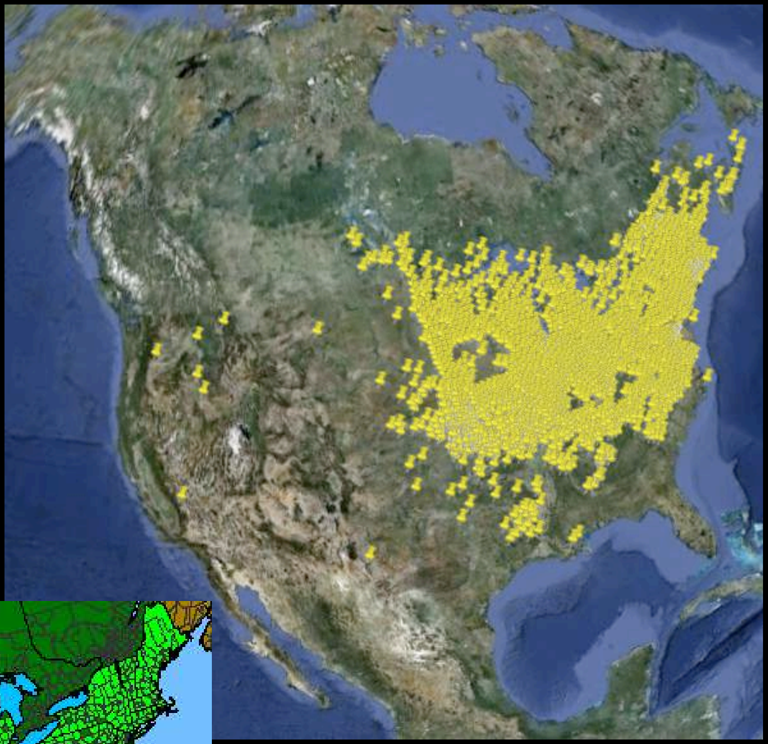


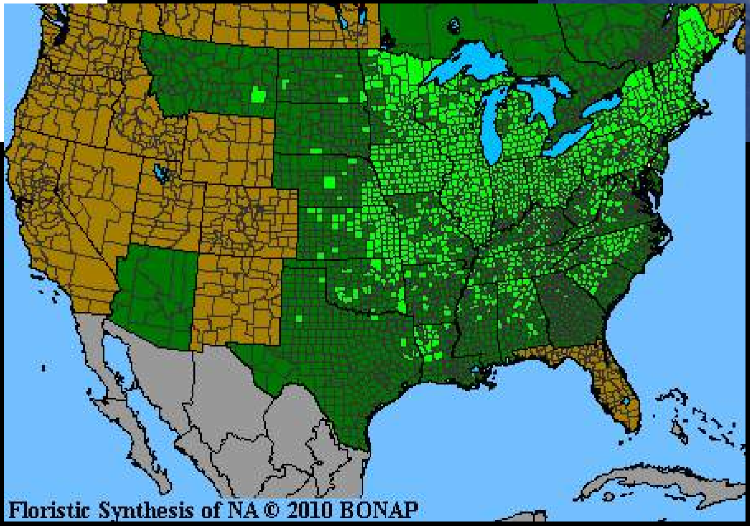
WIS Flora



Some Cautions:

How do these maps of the distribution of *Scirpus atrovirens* differ? Why do they differ?





Additional Resources

- panbiogeography analysis (martitracks)
- simulate origin and spread of species (biogeosim)
- forecast species distributions (biomod)
- spatial analysis of diversity (biodiverse)
- list of free phylogenetic software here
- simulate historical island biogeography (shiba)
- spatial analysis in macroecology (SAM)
- predict and analyze distributions (GARP)