BIEN

Botanical Information and Ecology Network

National Center for Ecological Analysis and Synthesis

BRIAN J. ENQUIST

THE UNIVERSITY OF ARIZONA, TUCSON AZ USA

THE SANTA FE INSTITUTE, SANTA FE NEW MEXICO, USA

Outline

1) **Meeting Outline**
2) **Introduction and Motivation**
3) **What is BIEN?**
   - Why do we need it?
   - How does it work?
   - What does it contain?
   - What can I do with it?
4) **Conclusions**

Introduction

Some Big Questions in Ecology:

- What grows where and why?
- What drives species richness patterns?
- How have traits evolved within/across taxa?
- What controls functional diversity?
- How will climate change impact communities and diversity?
- Where to prioritize land purchase for conservation?

BIEN approach - Our informatics goals should be guided by the science and science needs

The BIEN working group (2008 – 2017)

Ecologists, Informaticians, Plant Taxonomists,
Computer Scientists, Evolutionary Biologists

Network of 50+ people at 20+ institutions

Richard Condit, STRI, Panama and CTFS
Robert K. Peet, UNC
Brad Boyle, U. Arizona
Steven Dolins, Bradley University
Mark Schildhauer, NCEAS
Barbara Thiers, NYBG
Jens C. Svenning, Aarhus University, Denmark
Brian McGill, UMaine
John Donoghue, UArizona
Peter Jorgensen, Missouri Botanical Garden
Martha Narro, iPlant
Jim Regetz, NCEAS
Cyrille Violle, UArizona, CNRS
Aaron Marcuse-Kubitza, NCEAS
Bill Piel, Yale
Nathan Kraft, UMaryland
Naia Morueta-Holme, Uaarhus, Denmark
Nick Spensor, New Zealand Landcare
Susan Wiser, New Zealand Landcare
Jeff Ott, UNC
Barbara Dobrin, UArizona
Sandy Andelman, TEAM Conservation Int'l
Lindsey Sloat, UArizona
Kristine Engemann Jensen, U. Aarhus, Denmark
Brody Sandel, Aarhus University, Denmark
Irena Simova, Charles University, Czech Rep.
Benjamin Blonder, UArizona
Cory Merow, Yale
Brian Maitner, UArizona

Introduction

Big Data in Ecology? (e.g. Big BAD Data . . . .)

What we need:

- Conglomeration
- Standardization
- Error checking
- Open accessibility

This is where BIEN comes in!

See Enquist et al. (2016) Peer J

Introduction

BIEN Goals

(1) To address big science questions – a need to merge botanical data **herbarium**, **plot** (abundance), and **trait** data for plants in the Americas.

(2) Technology development goals - a need to establish an informatics workflow for continuing to assemble and integrate botanical observation data (plots, specimens, traits) for BIEN and other projects.

(3) Longer-term program development - to contribute to more permanent technical solutions to the integration of vegetation, botanical, & ecological data

Enquist et al. (2016) Peer J

BIEN Deliverables

Botanical Informatics (*Done and ongoing*)

- Tools for botanical data scrubbing, standardization, & feedback
- Integrated database integrating 'primary' plant observation data trait measures, specimens, ecological plots (abundance).
- Repeatable workflow for integration and standardization of botanical observation data.

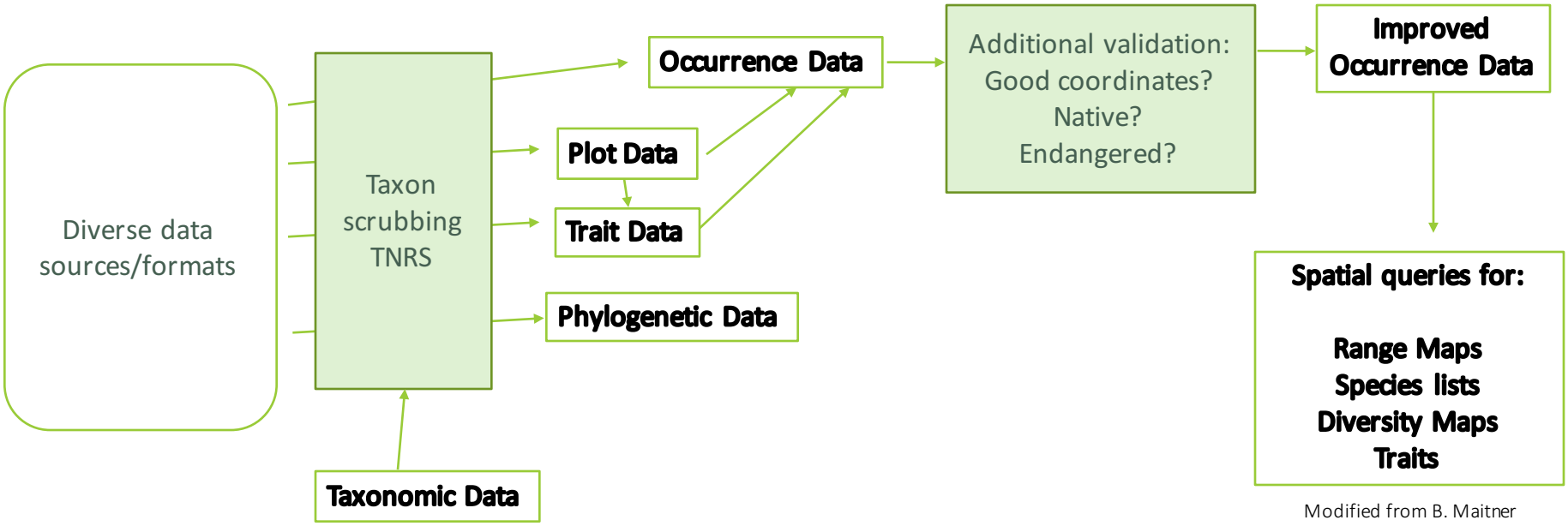
Enquist et al. (2016) Peer J

BIEN Deliverables

Derived Products (*Done and ongoing*)

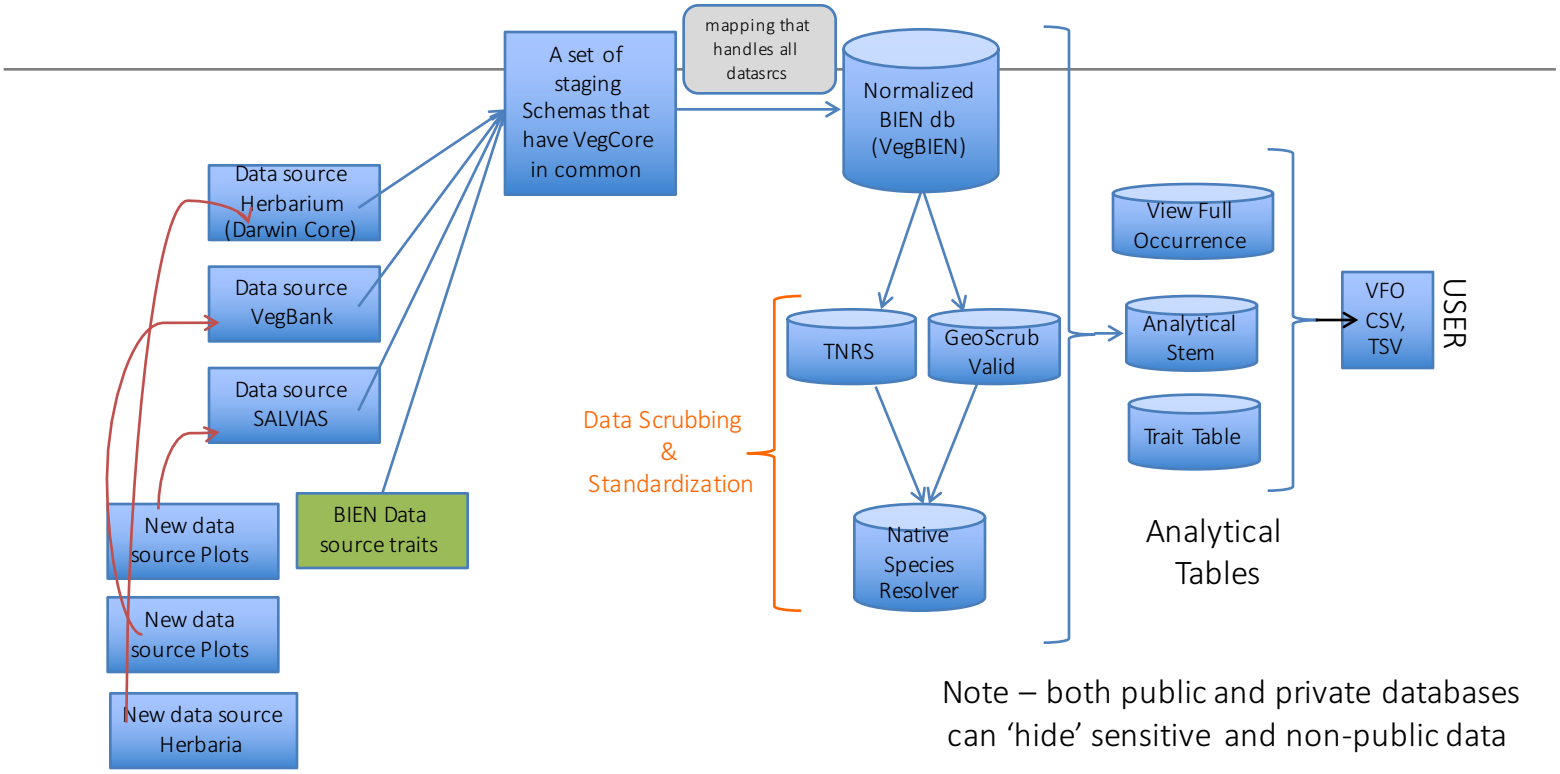
- Standardized species-list for plants of the New World
- Species level phylogeny for the Embryophytes of the New World
- Geographic range maps for all plants (Embryophytes) in the New World

The BIEN database



Modified from B. Maitner

BIEN Data Workflow



Note – both public and private databases can 'hide' sensitive and non-public data

BIEN Data Scrubbing and Standardization Steps

Validation underlying BIEN

- Taxonomic name resolution (TNRS) and API Boyle et al. (2013) BMC Bioinformatics
- Geographic name resolution (GNRS), API under development
- Geovalidation
- Detection and flagging of suspected cultivated specimens (Native Species Resolver, NSR and API) http://bien.nceas.ucsb.edu/bien/tools/nsr/
- Normalization and indexing of data sources
  Metadata pertaining to data sources, citations, data ownership are linked to the observations they provide
- Standardization of plot methodology metadata, citations

Development of a repeatable work flow to ingest, standardize, clean/scrub botanical data

How widespread is naming and taxonomic error?

Observational records are plagued with various sources of naming and taxonomic errors

March, 2013 – accessed 3 million botanical observations, from GBIF

- Able to satisfactorily standardize 2.4 million records consisting of 592,984 plant names
- After correcting those 592,984 species names for **spelling errors, and standardizing for taxonomy and synonymy differences**, the species name count was reduced to 297,558 distinct accepted species names.
- So, ~ 50% (!) of the original species names were erroneous in some way.

Scrub and standardize your plant names for free

tnrs.iplantcollaborative.org

How widespread is geographic error?

Observational records are plagued with various sources of location errors

Of ~4.2M observations that we were able to submit for geographic scrubbing and standardization . . . .

- **2.2M observations could not be checked** with geographic place names, indicating that ~50% of observations cannot be definitively assessed for their geographic coordinate accuracy.

- **Of the records with some geographic information**, 1.4M records had correct geographic coordinates but ~ 80,000 records (~36%) had various levels of errors associated with their geographic coordinates.

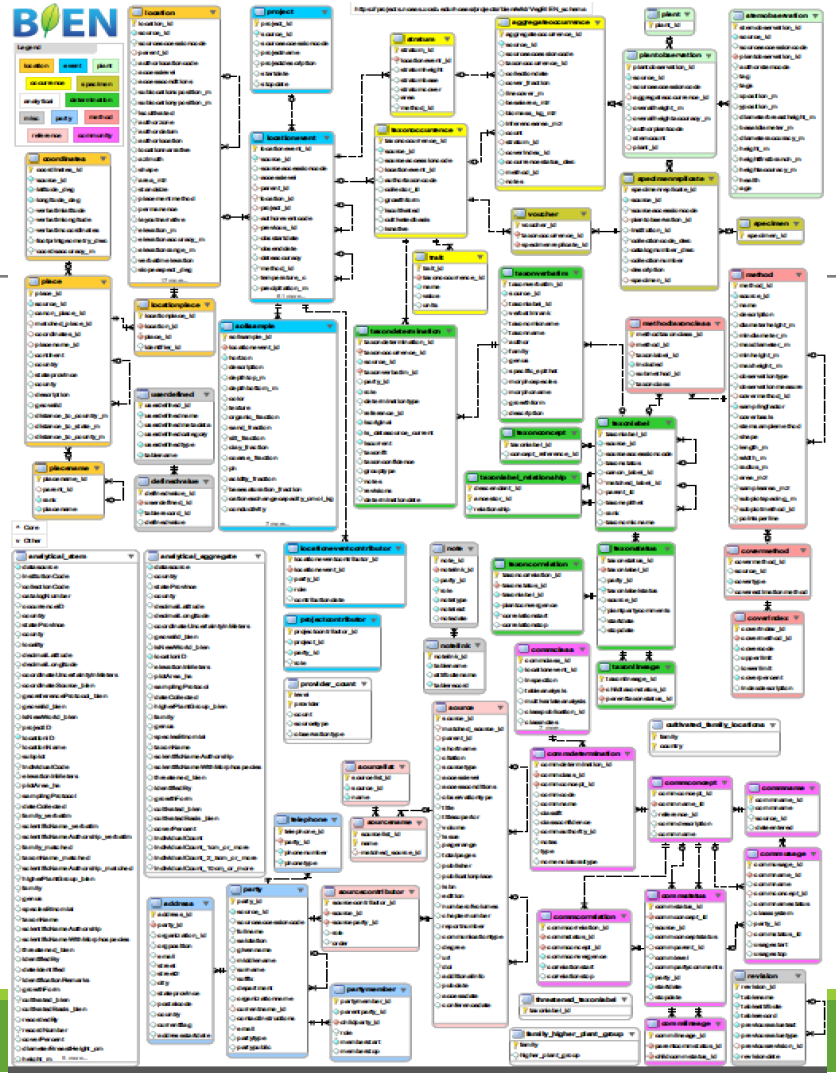
BIEN Workflow

Brad Boyle, UofA

Susan Wiser
Bob Peet
Rick Condit
Steve Dolins

Aaron Marcuse-Kubitza
NCEAS

Modified by MPS from B. Boyle



BIEN 3.0 Database Schema

A perpetual motion machine?

Individual data sources—or the entire database—can be loaded and re-loaded *rapidly*, allowing updates as data sources are modified or grow, or new sources are acquired.

Aaron Marcuse-Kubitza
NCEAS

Brad Boyle, UofA

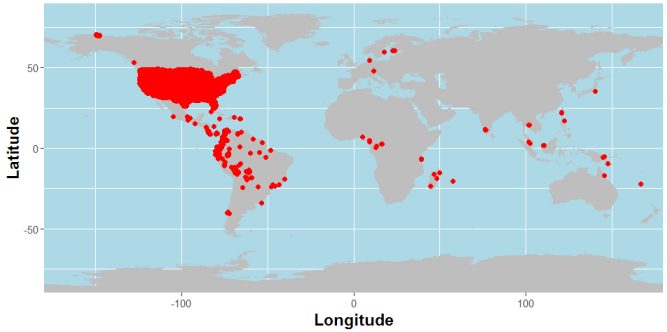
Data Sources

From BIEN2 (17)		New Sources (13)	
GBIF	REMIB	BRIT (Texas)	ACAD (Canada*)
SpeciesLink	ARIZ	TEX (Texas)	HIBG (Canada)
U	NY	HVAA (Chile)	JBM (Canada)
NCU	UNCC	NVS (N. Zealand)	MT (Canada)
MO	SALVIAS		QFA (Canada)
CVS	Madidi		TRT (Canada)
CTFS	VegBank		TRTE (Canada)
FIA	TEAM		UBC (Canada)
BIEN2 traits		BIEN3 traits	WIN (Canada)

* Canadian herbaria from Canadensys

The BIEN database

	Number observed within BIEN 3.2
Total Observations (Global)	81,693,397
Specimens	64,269,636
Plot observations	17,423,761
Plots	114,182
Species	378,554
Trait measurements	172,315
Ranges	81,274
Taxonomic Data	374,685



BIEN 3.2 Plot locations

BIEN3
>10.2 million valid and high quality georeferenced observations in New World

BIEN 3.2 Observation locations

BIEN Trait data

Trait	Units	Number of Species	Number of Records
Leaf Photosynthesis	µmol·s⁻¹·m⁻²	871	1,061
Flowering date	date	677	4,529
Flowering month	month	4,059	8,128
Height	m	4,508	16,224
Leaf area	mm²	7,735	3,374
Leaf Cmass	Percent, mg·g⁻¹	714	2,329
Leaf dry mass	g	1,981	18,738
Leaf dry matter content (LDMC)	mg·g⁻¹	1,761	12,309
Leaf lifespan (LLS)	months	699	800
Leaf Narea	g·m⁻²	1,717	2,153
Leaf Nmass	Percent, mg·g⁻¹	4,110	7,348
Leaf Parea	g·m⁻²	670	756
Leaf Pmass	Percent, mg·g⁻¹	2,080	4,796
Leaf Amass	µmol·g⁻¹·s⁻¹	818	1,007
seed mass	g	10,507	36,517
Specific leaf area (SLA)	m²·kg⁻¹	5,783	30,378

Trait measurements	172,315

16 traits

Literature compilation

Cyrille Violle

BIEN Geographic Range Maps

- Maps for most species were produced using Maximum Entropy distribution modeling and other range estimate algorithms.
- An enormous computational challenge
- Utilized the Texas Advanced Computing Center (TACC)

Geographic ranges for 'all' New World Embryophyte species

Liverworts
Mosses
Ferns
Gymnosperms
Angiosperms

~ 88,000 species range maps

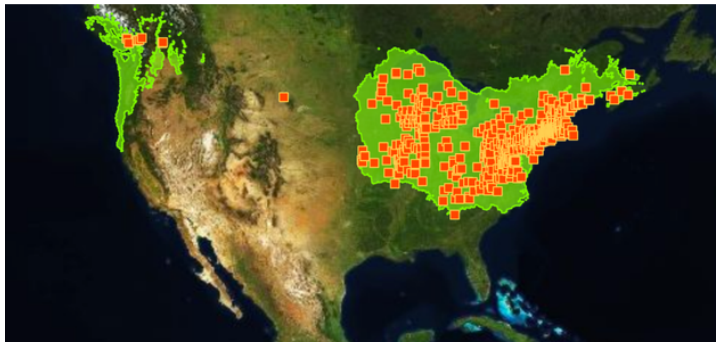
'Ranger' at TACC

Cory Merow
Brian McGill
Jens Svenning
John Donohugh
Naia Morueta-Holme
Nathan Casler et al.

BIEN Geographic Range Maps

What we have done

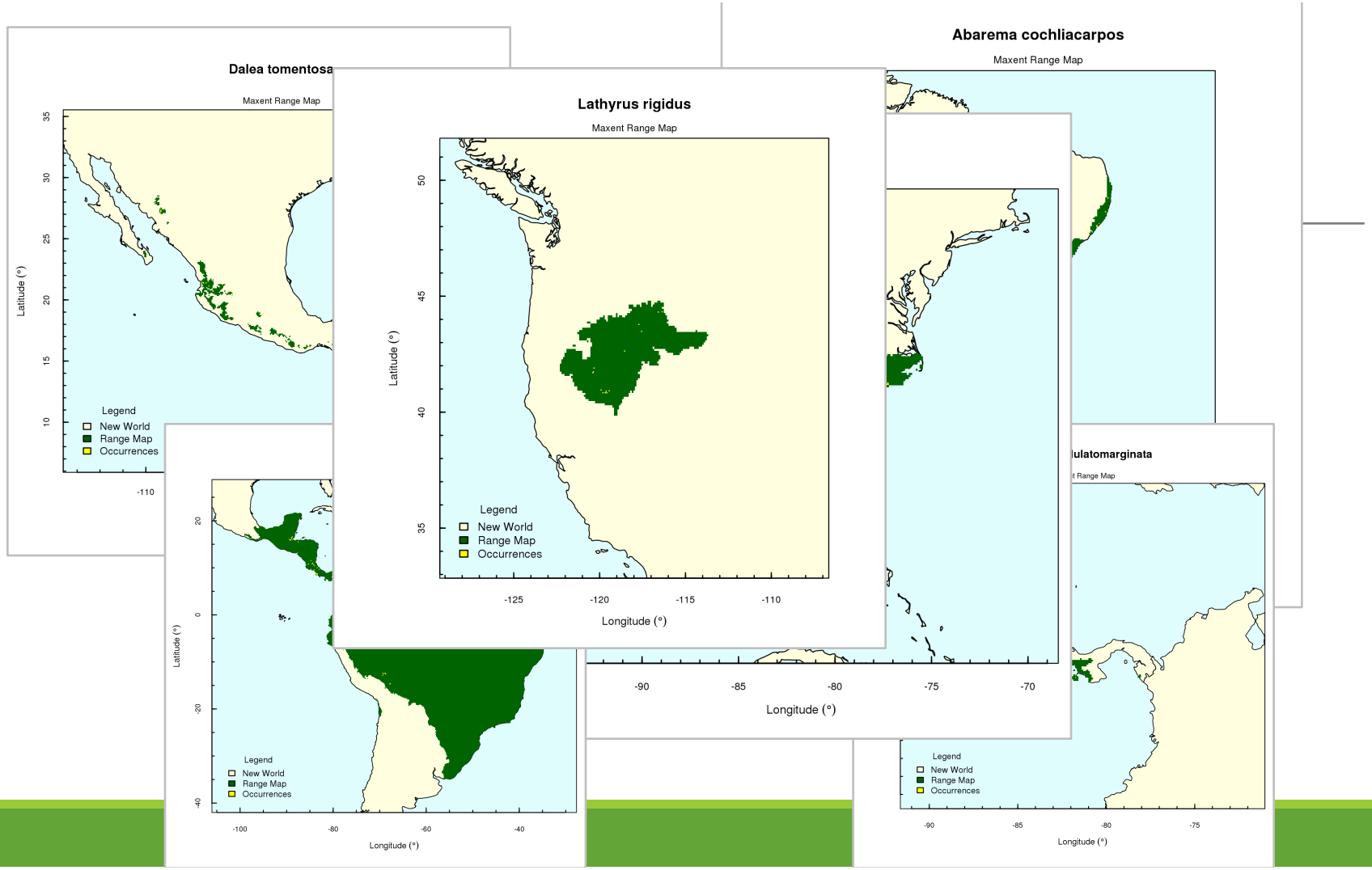
- New World Land Plants
- Decision tree for range modeling
  - cell (very rare species)
  - Convex hull (rare species)
  - SDM (common species)
- Geographic distributions
  - Continuous maps
  - Binary maps



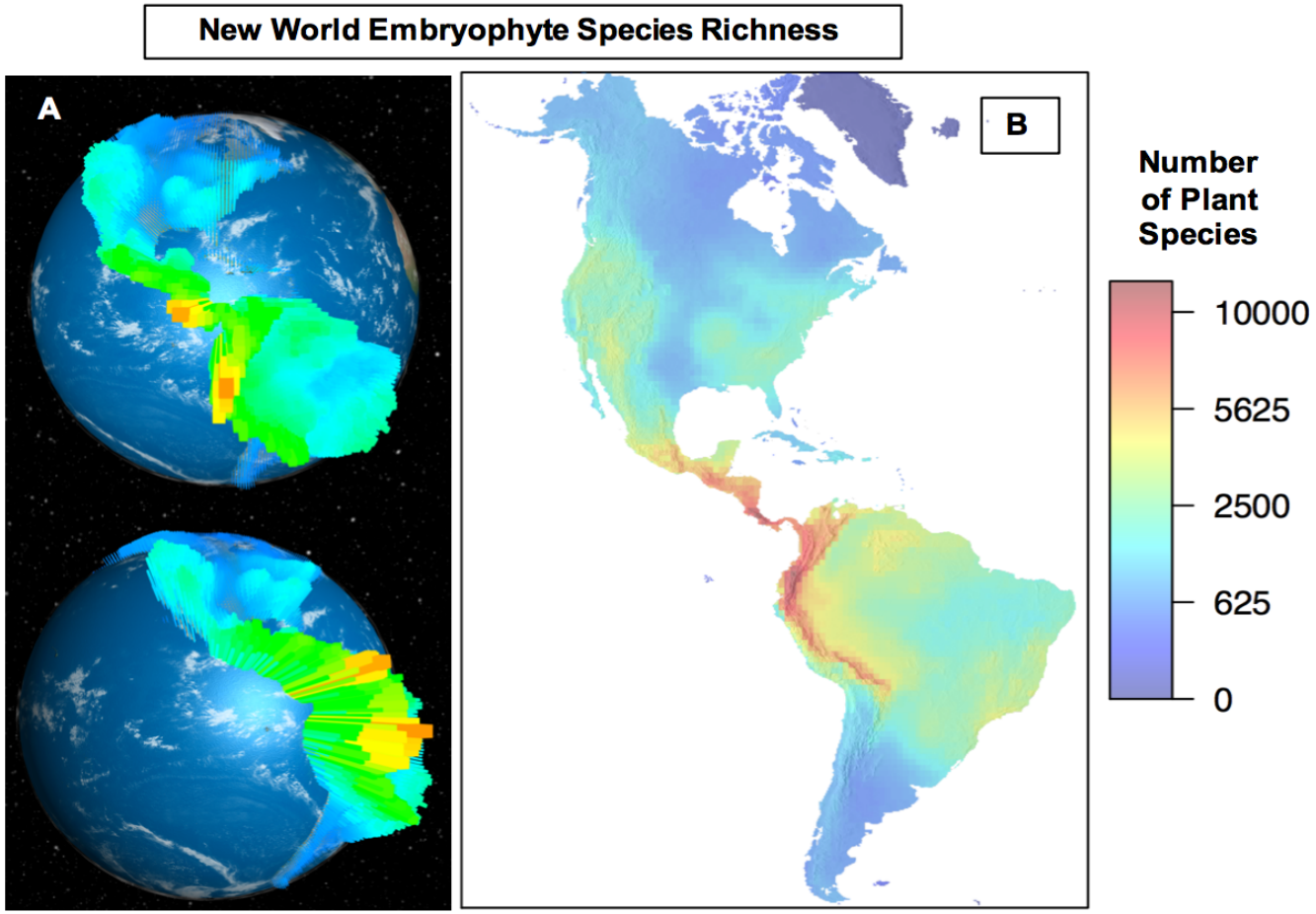
What we are doing

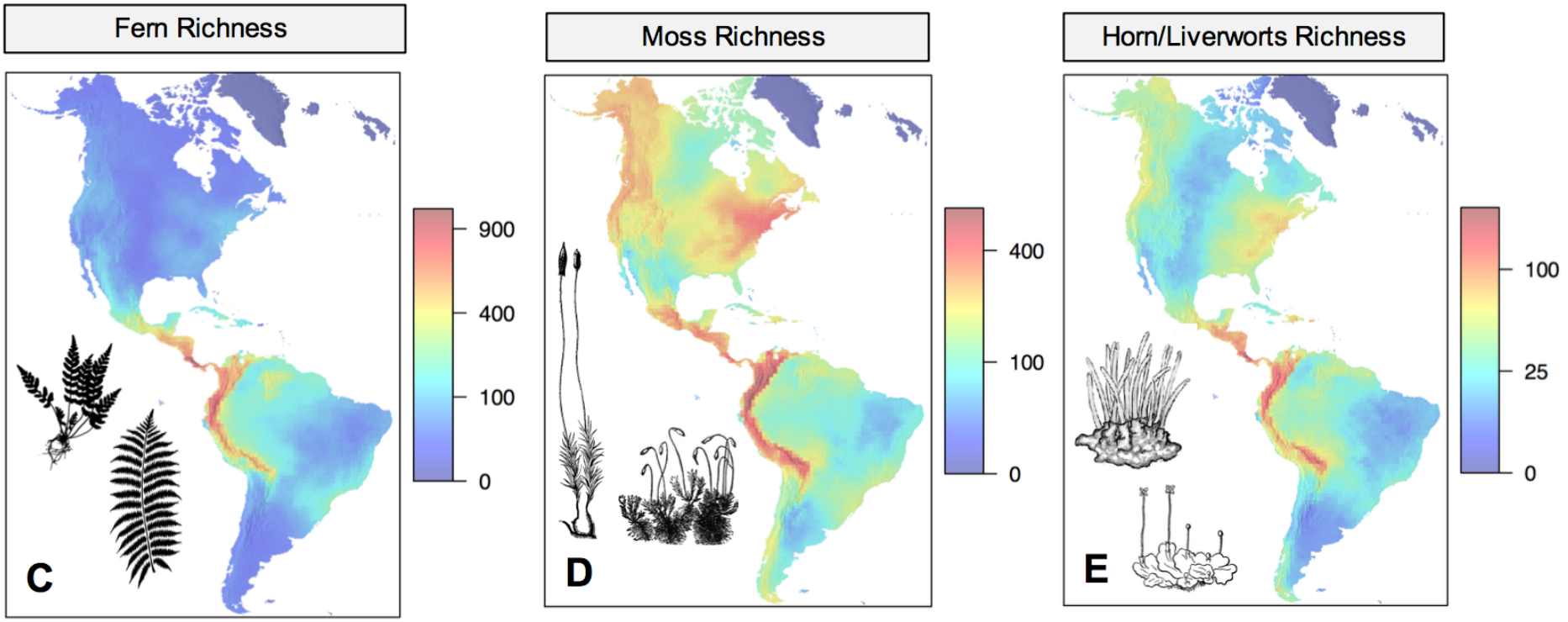
- Cleaning nonnative records
- Perfecting model complexity
- Improving thresholding

Cory Merow
Brian McGill
Jens Svenning
John Donohugh
Naia Morueta-Holme
Nathan Casler et al.


Dalea tomentosa
Maxent Range Map
Latitude (°)
Legend
New World
Range Map
Occurrences
Lathyrus rigidus
Maxent Range Map
Latitude (°)
Longitude (°)
Legend
New World
Range Map
Occurrences
Abarema cochliacarpos
Maxent Range Map
Longitude (°)
ulatomarginata
Longitude (°)
Legend
New World
Range Map
Occurrences

www.bien3.org/richness




Fern Richness
900
400
100
0
C
Moss Richness
400
100
0
D
Horn/Liverworts Richness
100
25
0
E

BIEN Geographic Range Maps

What we will do

1. Niche vs Distribution Models
2. Past ranges
3. Future ranges

w/ Brian McGill, Cory Merow, Jens Svenning, Naia Morueta-Holme, Nathan Casler et al.

Lee Hannah, Patrick Roehrdanz and the SPARC group

Spring 2017 onwards . . .

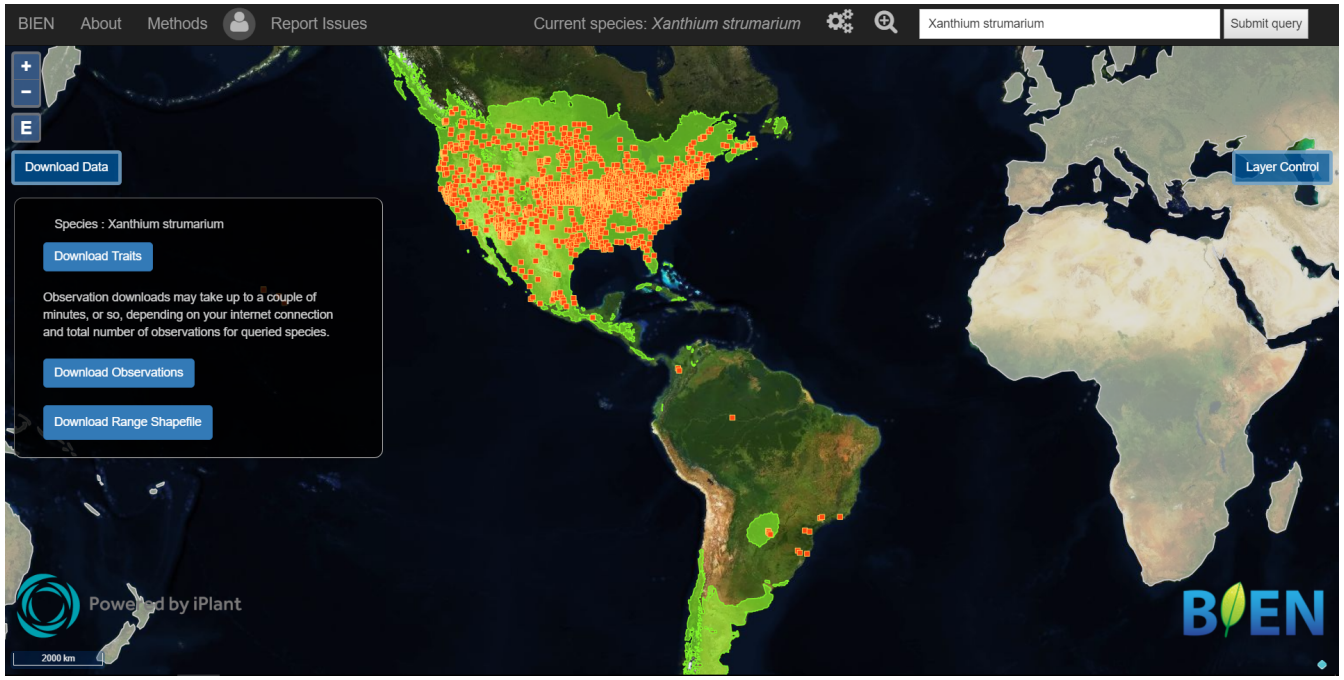
CONSERVATION INTERNATIONAL

The BIEN database

Accessing BIEN:

1) BIEN3.org
2) RBIEN
3) POSTGRESQL

BIEN3.org



Daniel Guaderama

RBIEN: What is it?

A set of R functions to work with the BIEN database

- Easier than learning SQL
- Much easier than learning BIEN structure
- Core goal of the package: make BIEN data easy to access

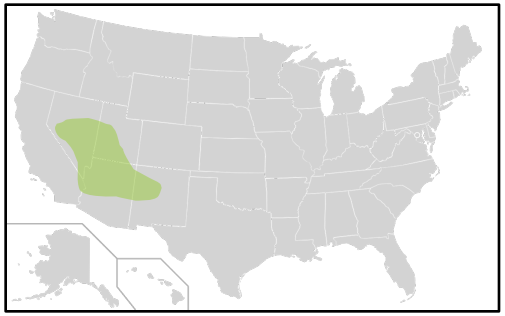
Brian Maitner

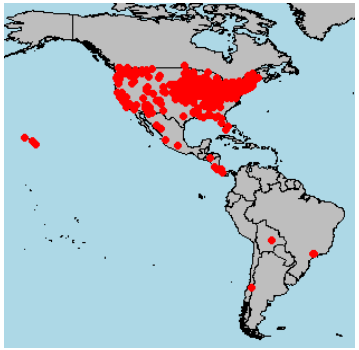
Maitner et al. (ms in prep)

RBIEN: What is it?

9 R function families:

1. species lists
2. occurrence records
3. plot data
4. species range maps
5. trait data
6. taxonomic information
7. phylogenetic information
8. metadata
9. custom queries





Learn more! - https://cmerow.github.io/RDataScience/3_3_RBIEN_tutorial.html

Maitner et al. (ms in prep)

Outreach

Tools to discover the diversity that surrounds you

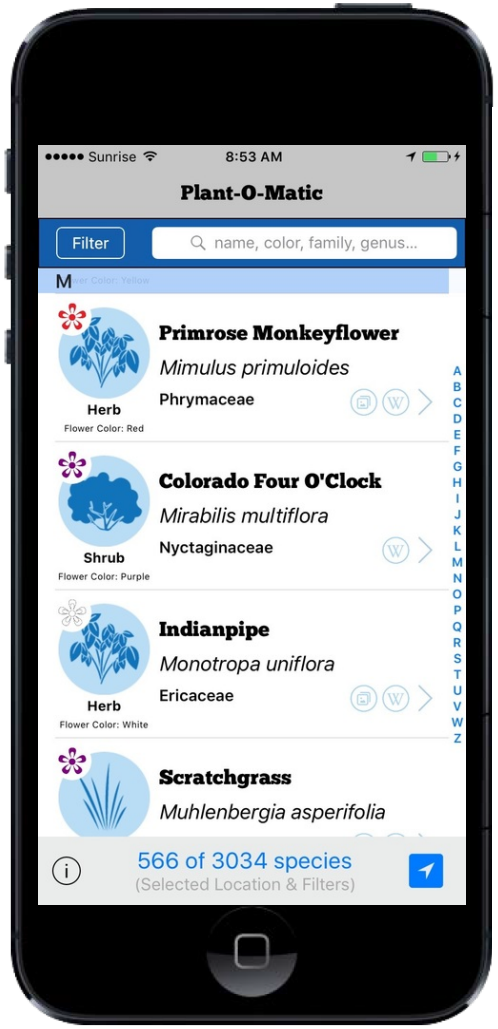
Deliver biodiversity information to the public

Plant-O-Matic: A dynamic and mobile field guide to all plants of the Americas

w/ Greg Goldsmith

Goldsmith et al. (2016) *Methods in Ecology and Evolution*

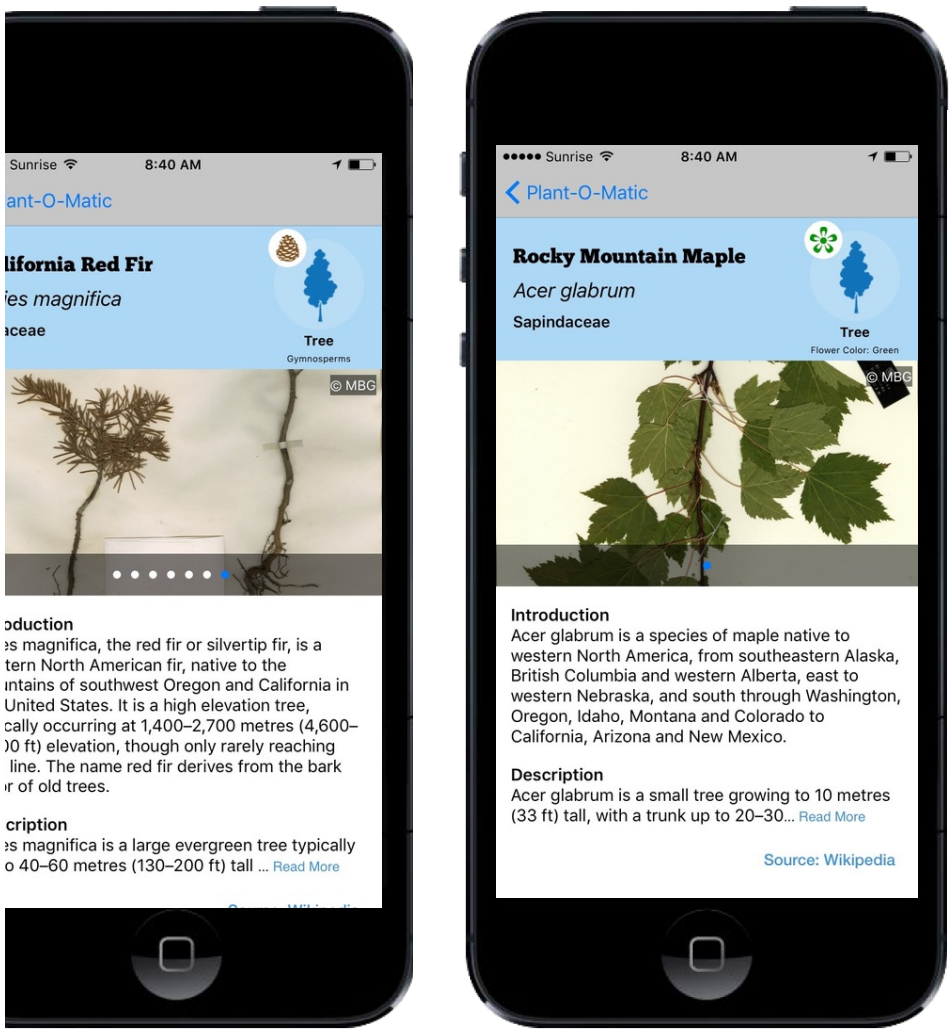
Plant-O-Matic: A dynamic and mobile field guide to all plants of the Americas



- For anywhere in the New World app will deliver a searchable species list to the user
- Names are all standardized via TNRS
- Diversity that is likely around you
- Future updates will allow more discovery. Search on habit (trees, lianas, herbs, cacti etc.)
- Does not require cellular data connection

Will deliver a species list in the middle of the Amazon

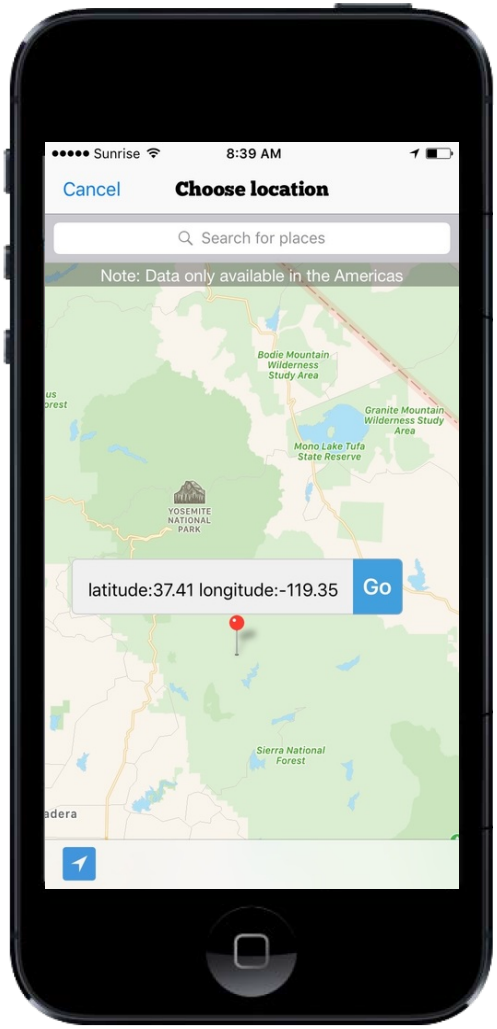
Goldsmith et al. (2016) *Methods in Ecology and Evolution*



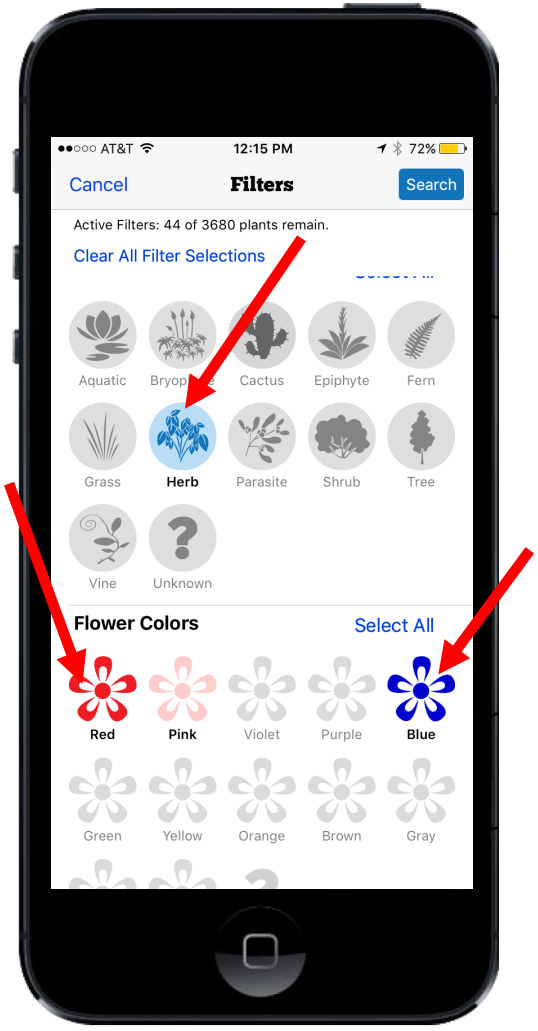
Plant-O-Matic: A dynamic and mobile field guide to all plants of the Americas

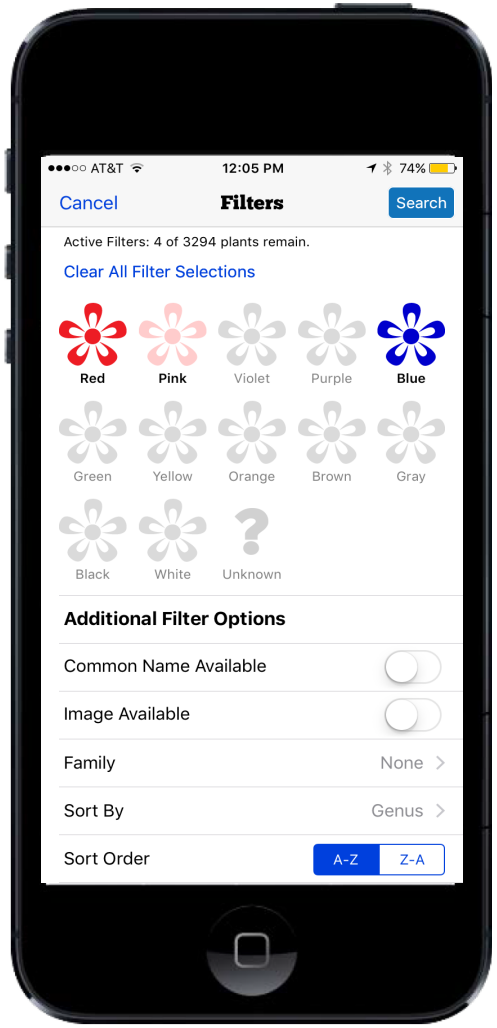
- Developing API queries to image repositories
  - TROPICOS API images
    Herbarium specimens and other images
  - Additional txon descriptions from Wikipedia
- Links to read more and view more images (photos, range maps, etc.) in Wikipedia

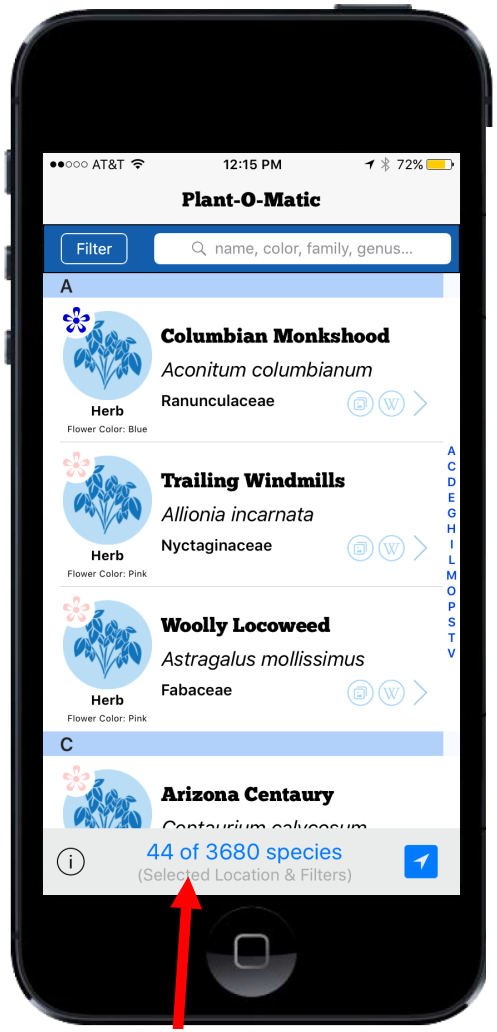
Goldsmith et al. (2016) *Methods in Ecology and Evolution*



Search any location in the Americas to generate species list per 100km^2







Filter the specie
habit* (tree, h
etc.), flower co
common nam

*when data are availa

Search for he
red and blue
(44 species
3,680 Embryc

Conclusions

Much progress toward a BIEN workflow . . .

- To reliably use the streams of botanical data now becoming available

- To integrate different sources of botanical data

- To discover and analyze patterns and the processes generating botanical diversity

Goals

Highlight new developments, collaborations, and connections

Highlight new science and research directions

What are our next short- and long-term science and informatics goals?