

Synopsis of current BIEN and Enquist projects managed by Martha

iPlant 2014

Ecology and Biodiversity

The problem

- Community created a database of plant specimen and plot data spanning North and South America.
 - Hundreds of thousands of species
 - 12 million records of species occurrence observations
 - 140,000 trait measurements
- Lack of standardization of species names, inflated measures of species richness and made it nearly impossible to do reliable research.

BIEN - NCEAS - iPlant Collaboration

- Develop a Taxonomic name Resolution Service.
- Compute range models for all BIEN species having reliable location data.
- Compute a phylogenetic tree of all species in database. (in progress)
- Redesign BIEN database and generalize import scripts. (in progress)
- Community members use products to pursue their research interests.

Results

Taxonomic Name Resolution Service

- Web Service resolves large batches of names against four taxonomic sources
- Averaging 1000 hits per month, 57% returning visits, used in 80 countries
- BIEN database contains 200,915 species, 30% fewer than before resolving the names.
- Resolved 27,317 plant names in NCBI and ITIS, increasing name intersection by 15%.

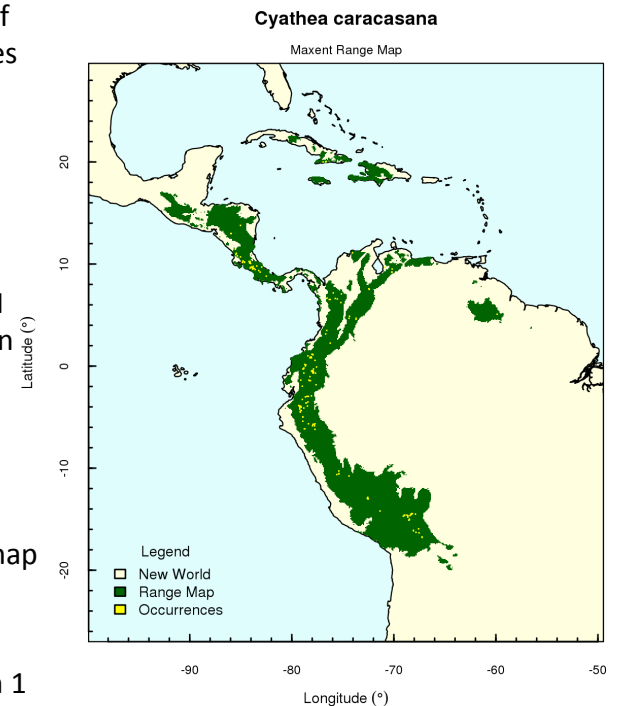
Range Modeling

- Computed range models for 88,824 species.
- Then, computed a species richness map for the Americas.

Scalable science

- TNRS processes 1.2 million records in 1 day, 6 hours (depending on server load)
- 100,000 CPU hours to compute ranges, 12 TB output

iPlant enabled these communities to pursue their science at global scale.

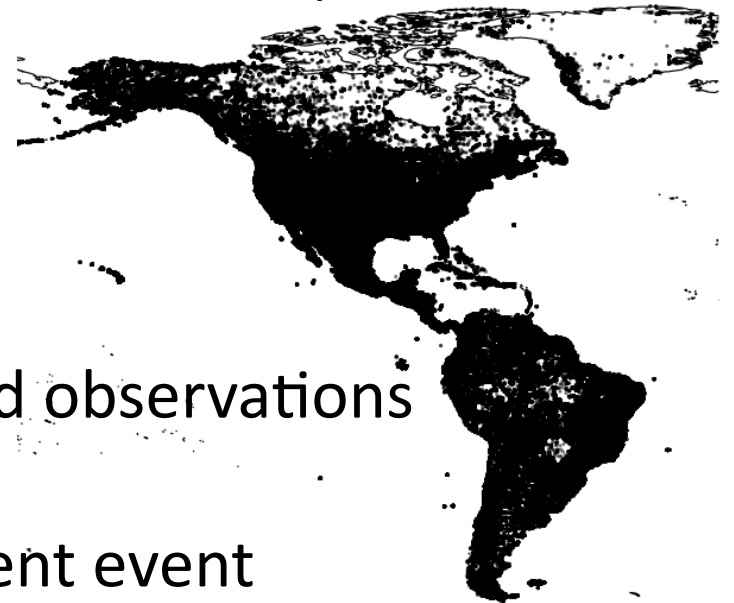


Range distribution of *Cyathea caracasana* based on Maxent modeling.

J. Donoghue et. al, unpublished.

BIEN Database 3.0

- A schema for merging geo-referenced observations of individuals and species from specimens, vegetation inventories, and regional checklists, with measurements of species-level traits such as size, growth form, wood density, specific leaf area, etc.
- Workflow for loading data, including data cleanup (taxon name, geographic name and location)
- Primary focus is the Americas
- 30 data sources
- 83,794, 197 observations
- 374,268 species
- 10.2 million validly geo-referenced observations
- August 2014 enter maintenance
- September 2014 public engagement event



Future Forest Distribution Model

(part of Ecological Modeling)

- Forecasting and visualizing American Western tree species' responses to climate
- Modeling the species distributions of 100 trees at high geographic spatial resolution (1 km) under future climate scenarios.
 - Using downscaled CMIP5 climate data
 - Three increasingly complex modeling approaches (maxent, improved maxent, demographic)
- “Range Modeling” ECS request included this work

Range Modeling

- ECS Request combined two modeling efforts
- BIEN species distribution models (maxent)
 - Run the models using improved BIEN3 data
 - Create a workflow others can use
- Future Forest Distribution Models

BIEN Summary

Feb. 2014

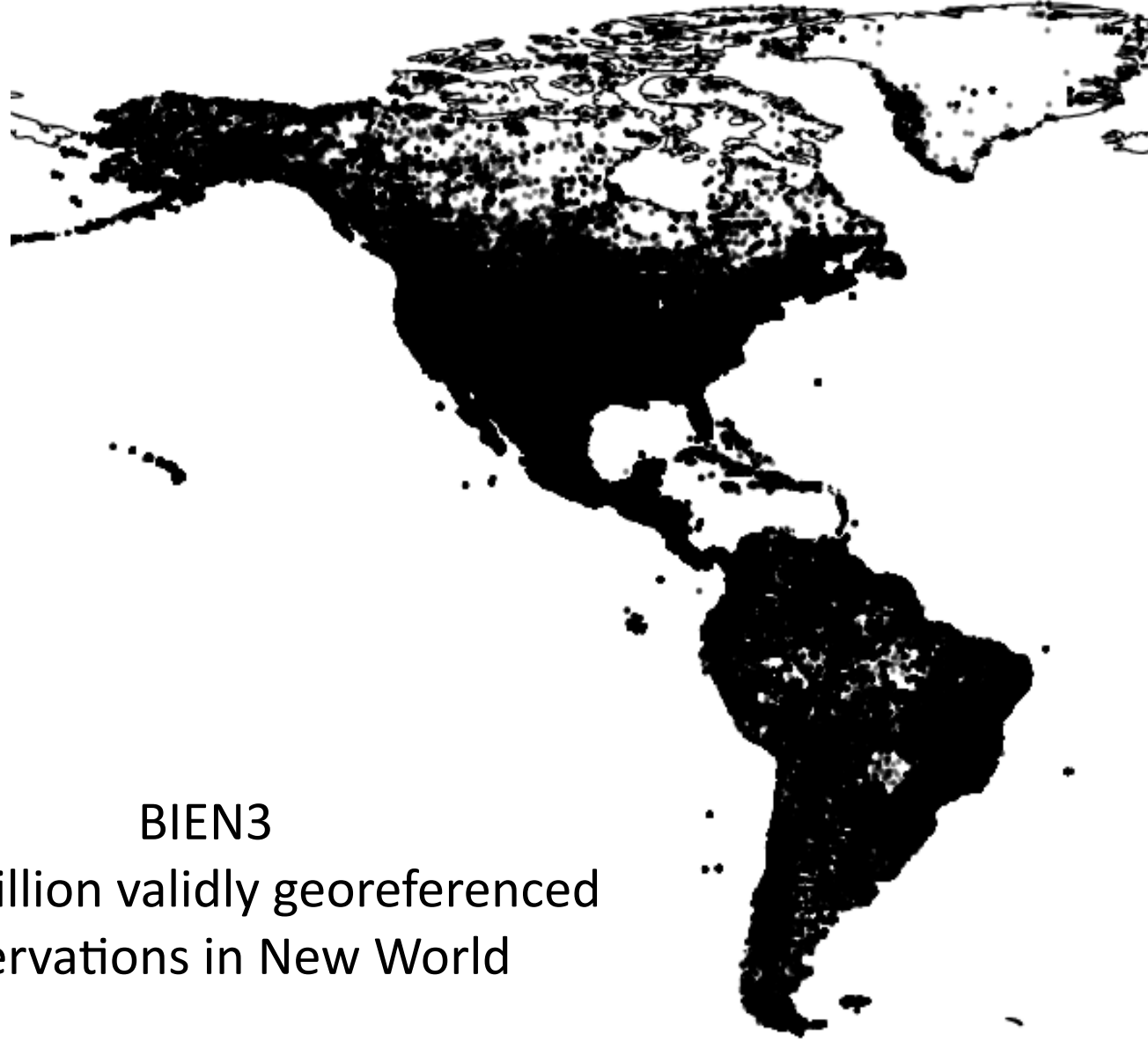
Data Sources

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			WIN (Canada)

* Canadian herbaria from Canadensys

Data Statistics

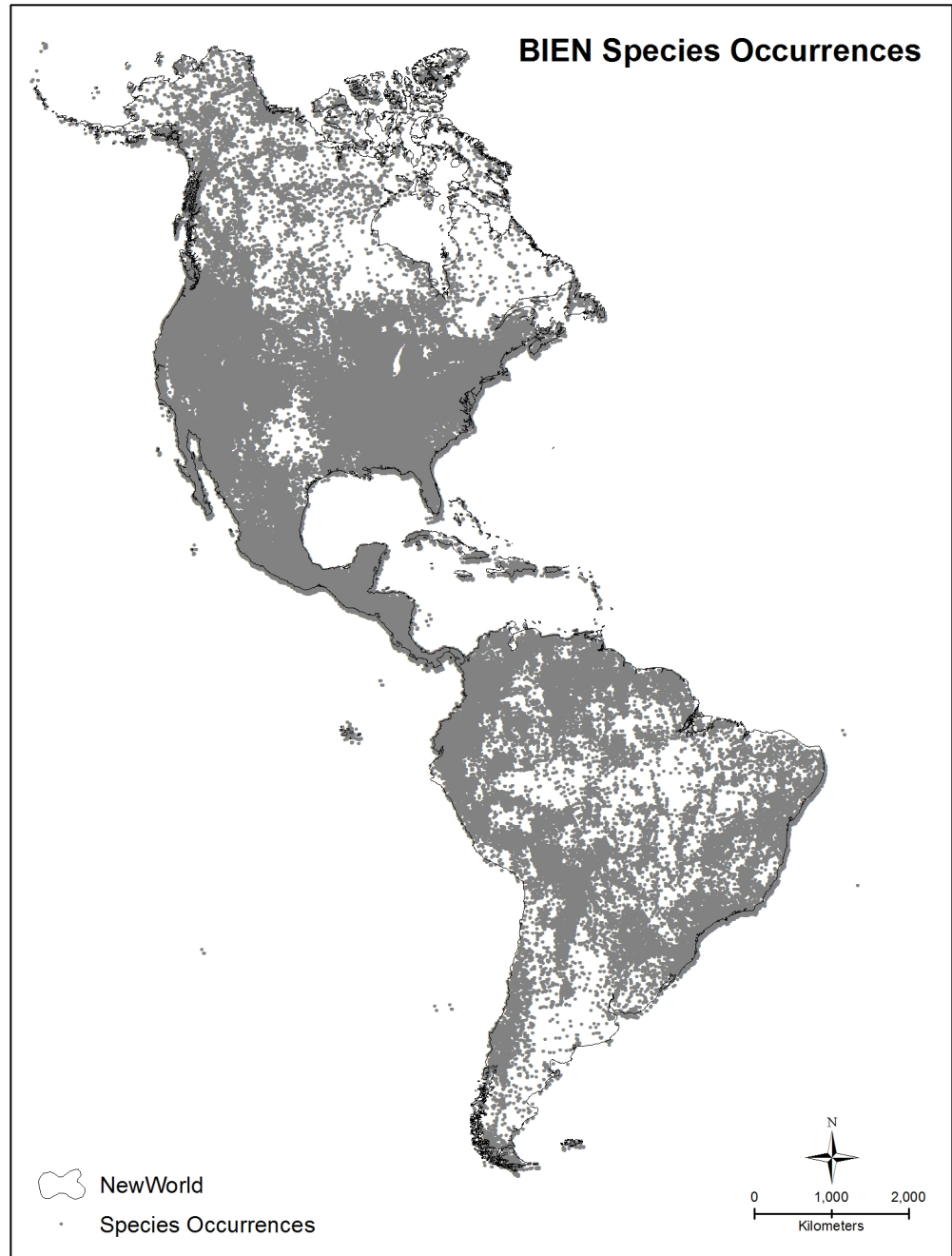
	BIEN2	BIEN3	Proportional increase
Observations	12,030,729	83,794,197	7x
Specimens	9,345,197	64,139,441	6.9x
Plot observations	2,685,532	19,505,960	7.3x
Plots	329,741	591,886	1.8x
Species	209,848	374,268	1.8x



BIEN3

>10.2 million validly georeferenced
observations in New World

BIEN2:
3.5 million



BIEN3 Database (beta)

beta = data are loaded in normalized database, still need validation

Deliverable	Due Date
Schema	✓
Workflow to load data	✓
Data loaded in normalized BIEN 3 database (30 sources)	✓
Import new datasets: through VegBank, SALVIAS, GBIF	✓

BIEN3 Database (1.0)

1.0 = ready for scientists to use

Deliverable	Due Date
Data validation (data are loaded correctly)	April , Aug
Data quality: Cultivated and introduced specimens	✓
Serve the data	Sept
Enter maintenance mode	August

BIEN products available through iPlant

To enable large-scale analyses

Deliverable	Due Date
Species distribution modeling workflow	Aug 2014
BIEN 3.0 database	Aug 2014
Public release Host a virtual event to engage community users	Sept 2014
Integrate iPlant species phylogenies with BIEN species phylogenies Dependent on Stephen Smith's and Tandy Warnow's work to create the perpetually updating iPlant phylogenetic trees of plant species.	Nov 2014
Finalize and automate the workflow to create the BIEN phylogenies.	Feb 2015
Make the BIEN phylogenies and workflows accessible through iPlant.	Feb 2015