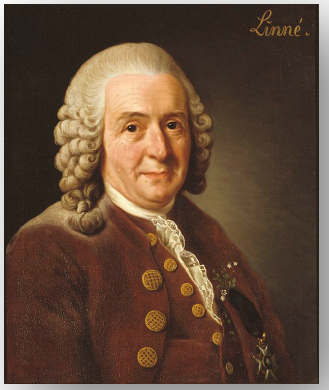
Prospects for the use of digitized specimens in systematics and comparative biology, with examples from plants

Michael Donoghue, Patrick Sweeney, and Beth Spriggs
Yale University, Peabody Museum of Natural History

I. What is systematics? Comparative biology? Phylogenetic biology?
II. How do digitized specimens facilitate research in these areas?
III. Some obvious uses (briefly!)
IV. Some less obvious uses, with examples mainly from *Viburnum*

Systematics after Linnaeus



Carl Linnaeus

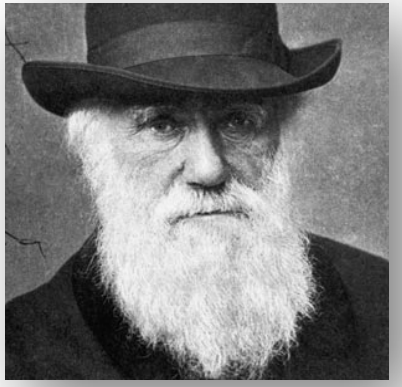
Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

↓

Classification/naming
("taxonomy")

Systematics after Darwin



Charles Darwin

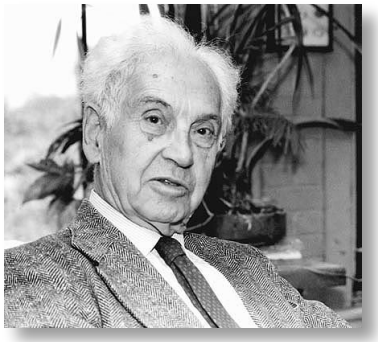
Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

Classification/naming ("taxonomy")

Evolutionary history (phylogeny)

Systematics after the "modern synthesis"



Ernst Mayr

Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

- Classification/naming ("taxonomy")
- Evolutionary history (phylogeny)
- "Biosystematics" (pop. genetics)

Systematics after Hennig

Willi Hennig

Inventories/floras

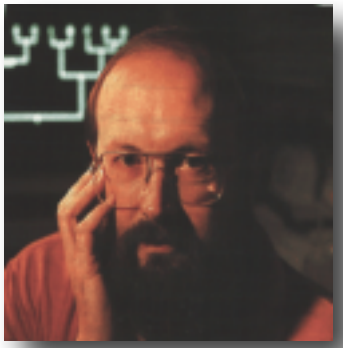
Collections ⟷ **Comprehensive description of organisms**

Evolutionary history (phylogeny)

"Biosystematics" (pop. genetics)

Classification/naming ("taxonomy")

Systematics after Felsenstein (etc.)



Joe Felsenstein

Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

Evolutionary history (phylogeny)

Classification/naming ("taxonomy")

Speciation, hybridization ("biosystematics")

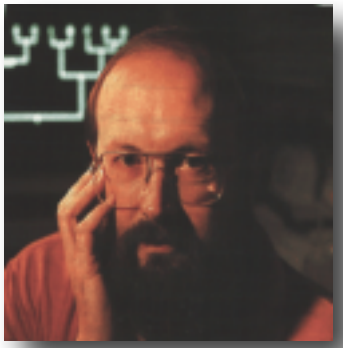
Character evolution (adaptation)

Dates and rates of evolution

Historical biogeography

Coevolution/diversification

Systematics after Felsenstein (etc.)



Joe Felsenstein

Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

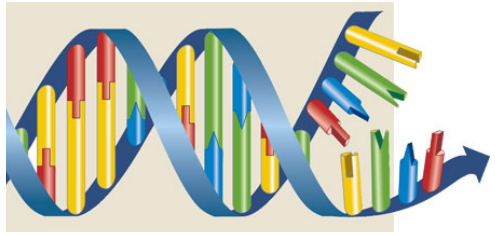
"Phylogenetic comparative biology"
(multiple clades)

Evolutionary history (phylogeny)

- Speciation, hybridization ("biosystematics")
- Character evolution (adaptation)
- Dates and rates of evolution
- Historical biogeography
- Coevolution/diversification

Classification/naming ("taxonomy")

Systematics after DNA sequencing



Inventories/floras

Collections ⟷ **Comprehensive description of organisms**

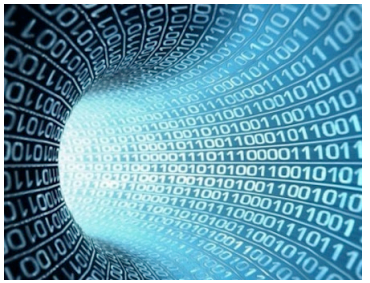
infer tree

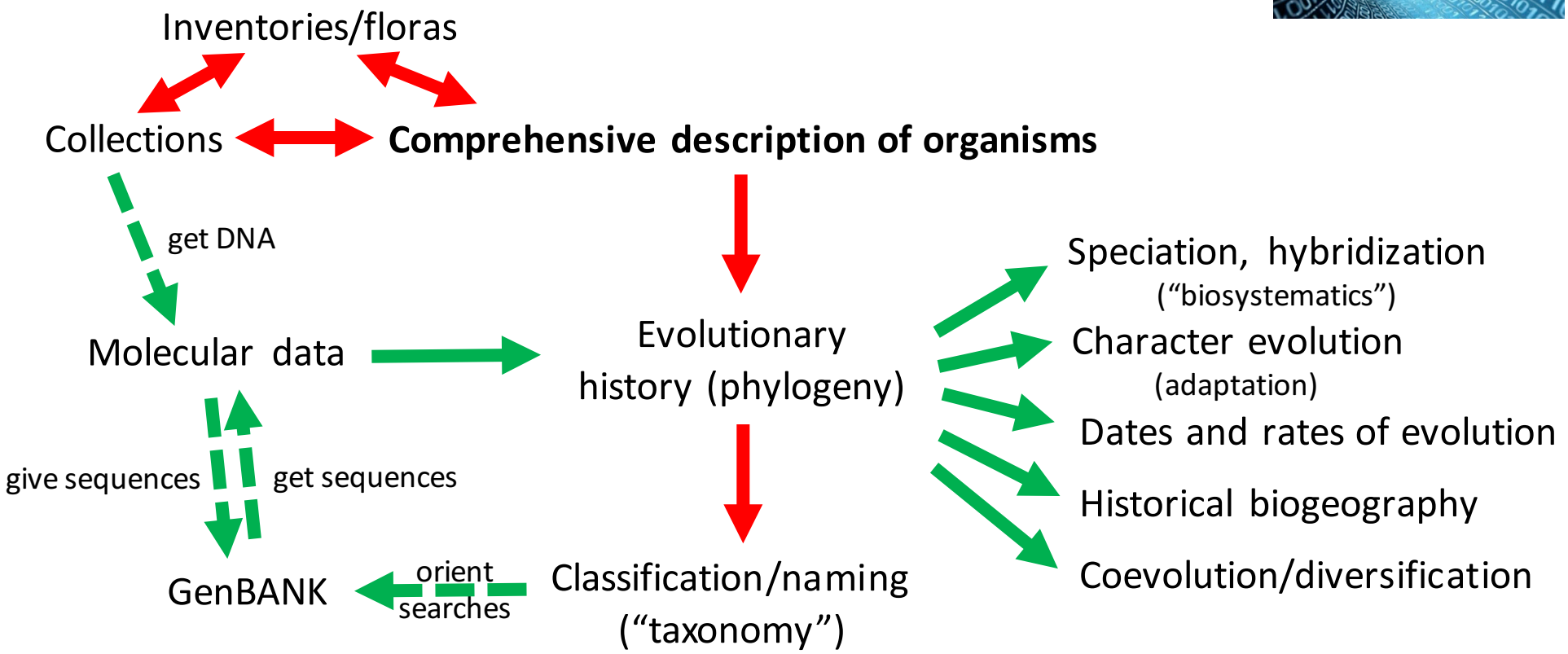
Molecular data → infer tree → Evolutionary history (phylogeny)

Evolutionary history (phylogeny) → Classification/naming ("taxonomy")

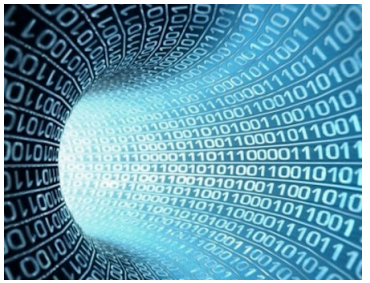
Evolutionary history (phylogeny) →
- Speciation, hybridization ("biosystematics")
- Character evolution (adaptation)
- Dates and rates of evolution
- Historical biogeography
- Coevolution/diversification

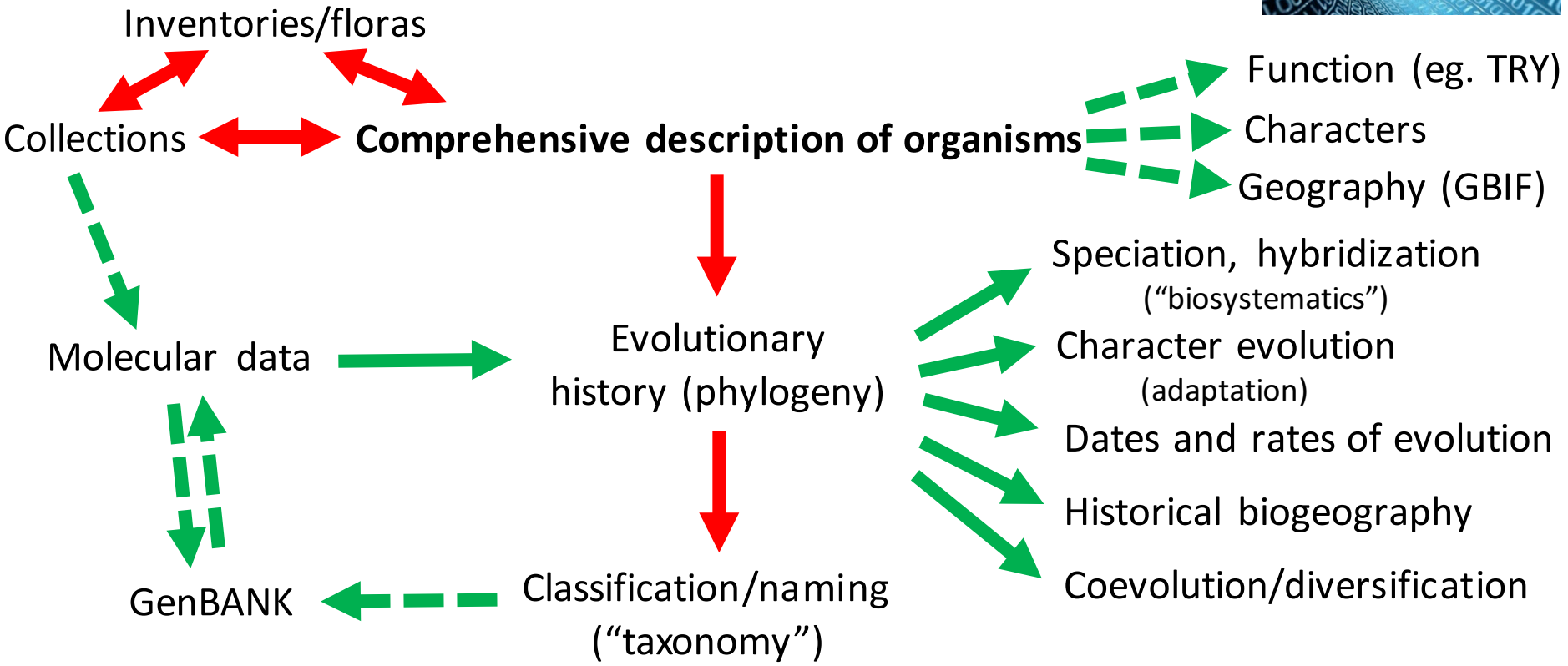
Systematics after DNA sequencing and informatics



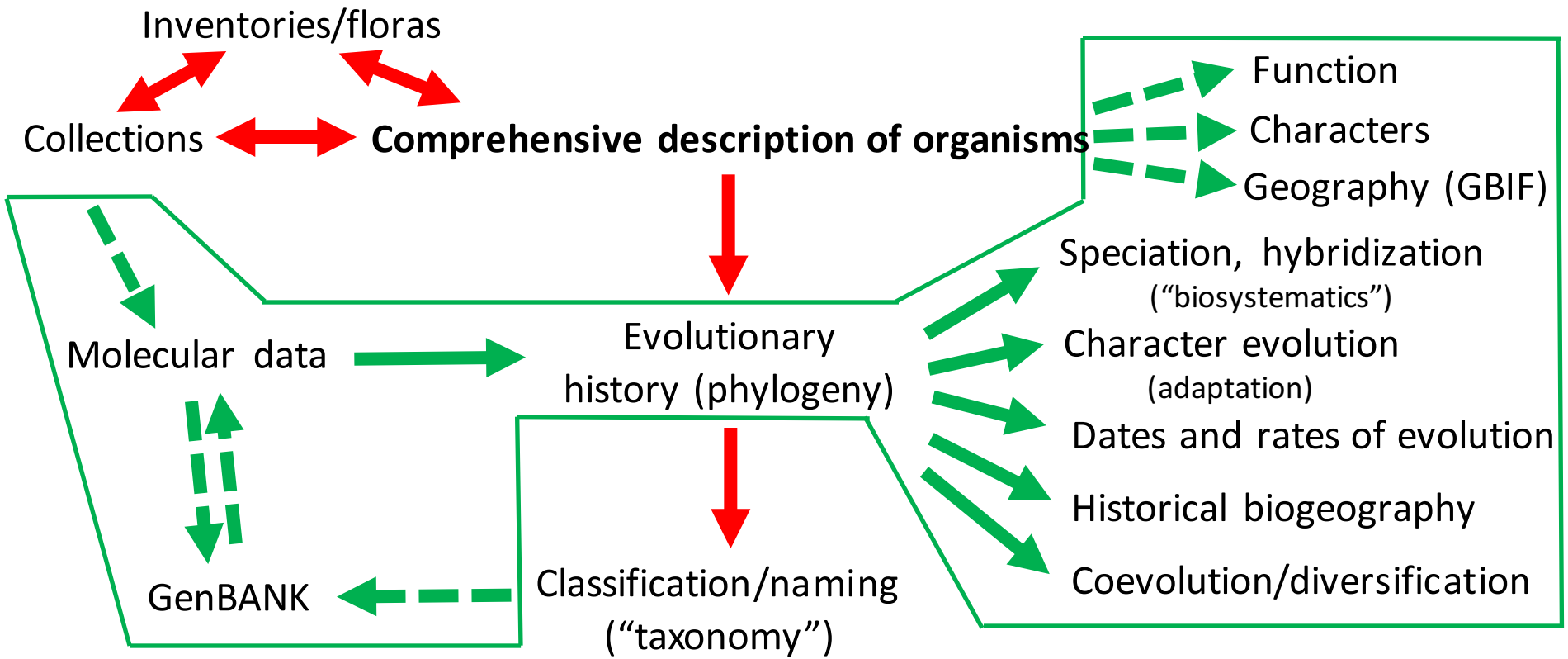


Systematics after DNA sequencing and informatics



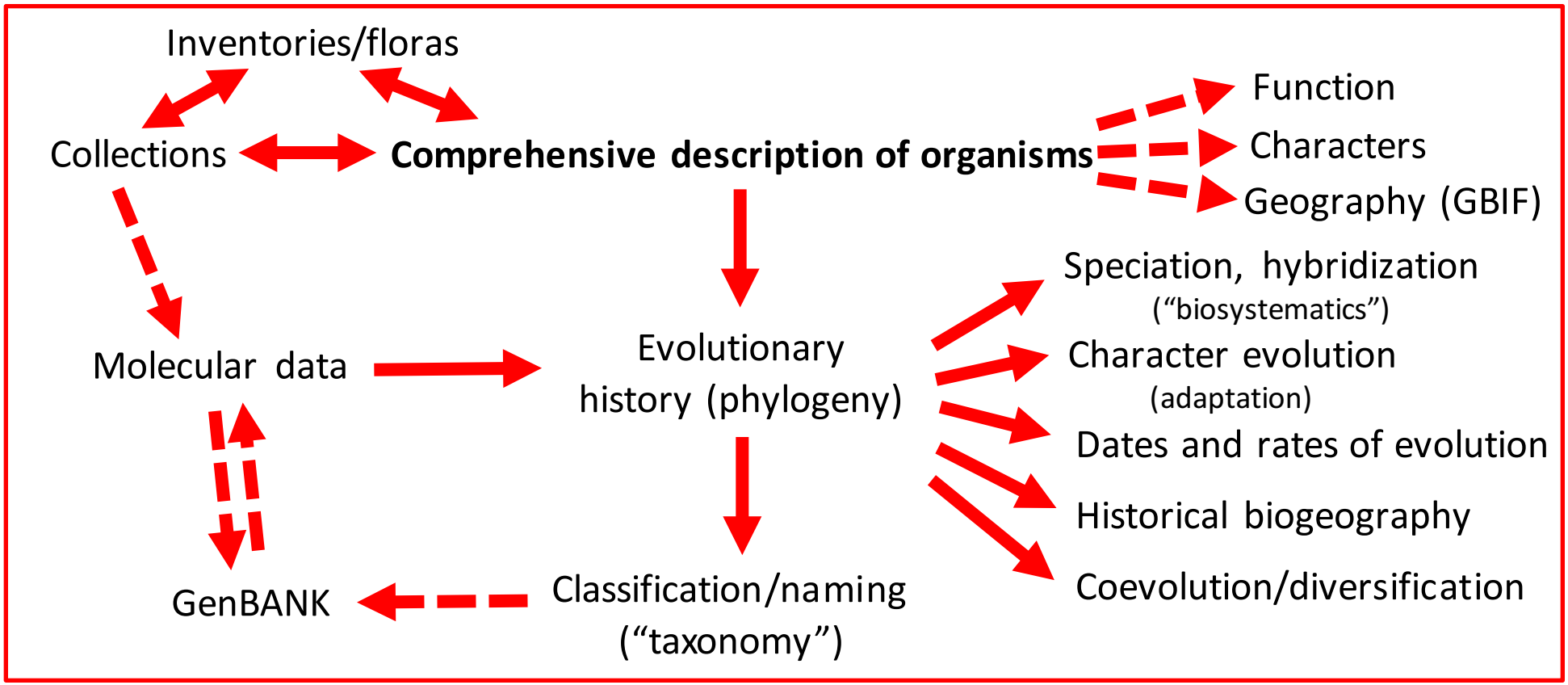


Systematics after DNA sequencing and informatics



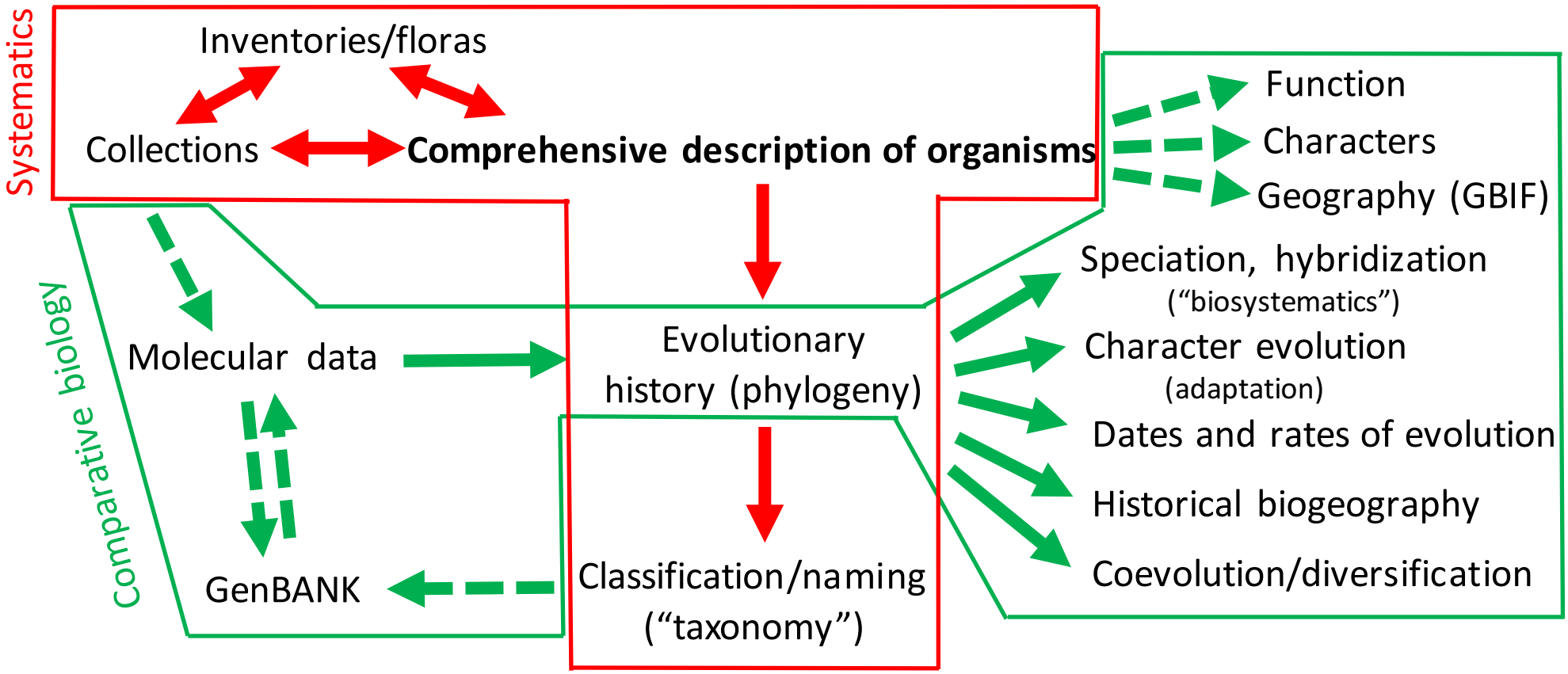
Modern phylogenetic comparative biology
(multiple clades or very large ones)

Is all of this systematics, or what?



Option 1 – YES, it's all systematics, with sub-disciplines, e.g., taxonomy, comparative biology, etc.

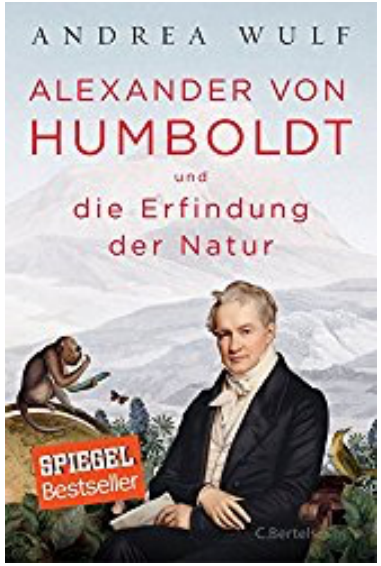
Is all of this systematics, or what?



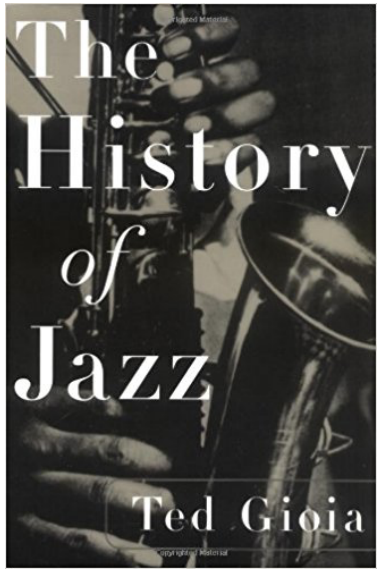
Option 2 – NO, maybe it's better to split it into "Systematics" and "Comparative Biology"

The distinction is a bit psychological – style of thought and motivation

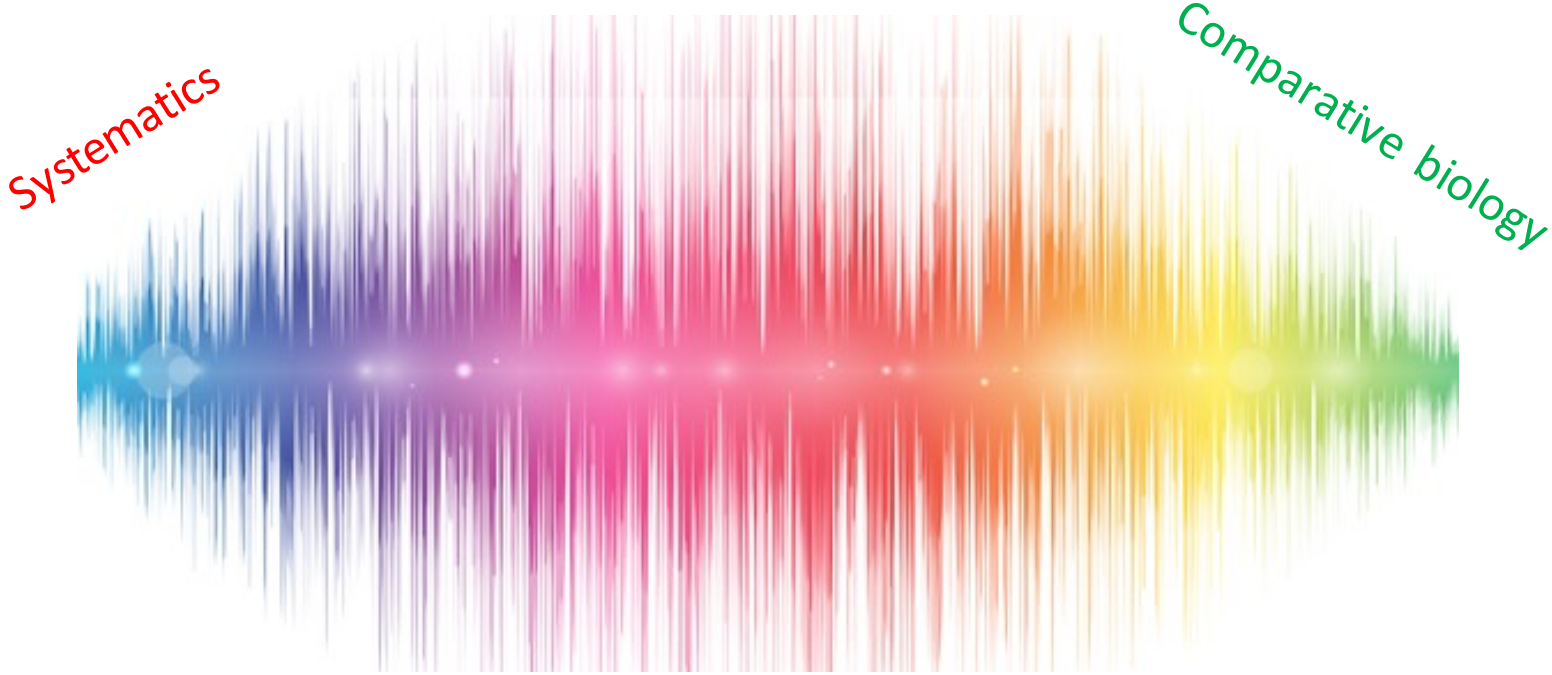
Systematists are obsessed with a particular clade; every detail is absolutely fascinating, and has to be integrated. They collect and classify. They're a bit obsessive-compulsive ("sorting of objects", "rules for arranging"). Systematists are are writing clade "biographies."



Comparative biologists want to generalize across multiple clades; they're interested in a phenomenon or a process, not a particular clade. They don't collect or classify. They have a bit of an attention deficit problem. Comparative biologists are writing "historical analyses."



Of course, it's a spectrum, many people are in the middle, and there's no value judgement implied here!

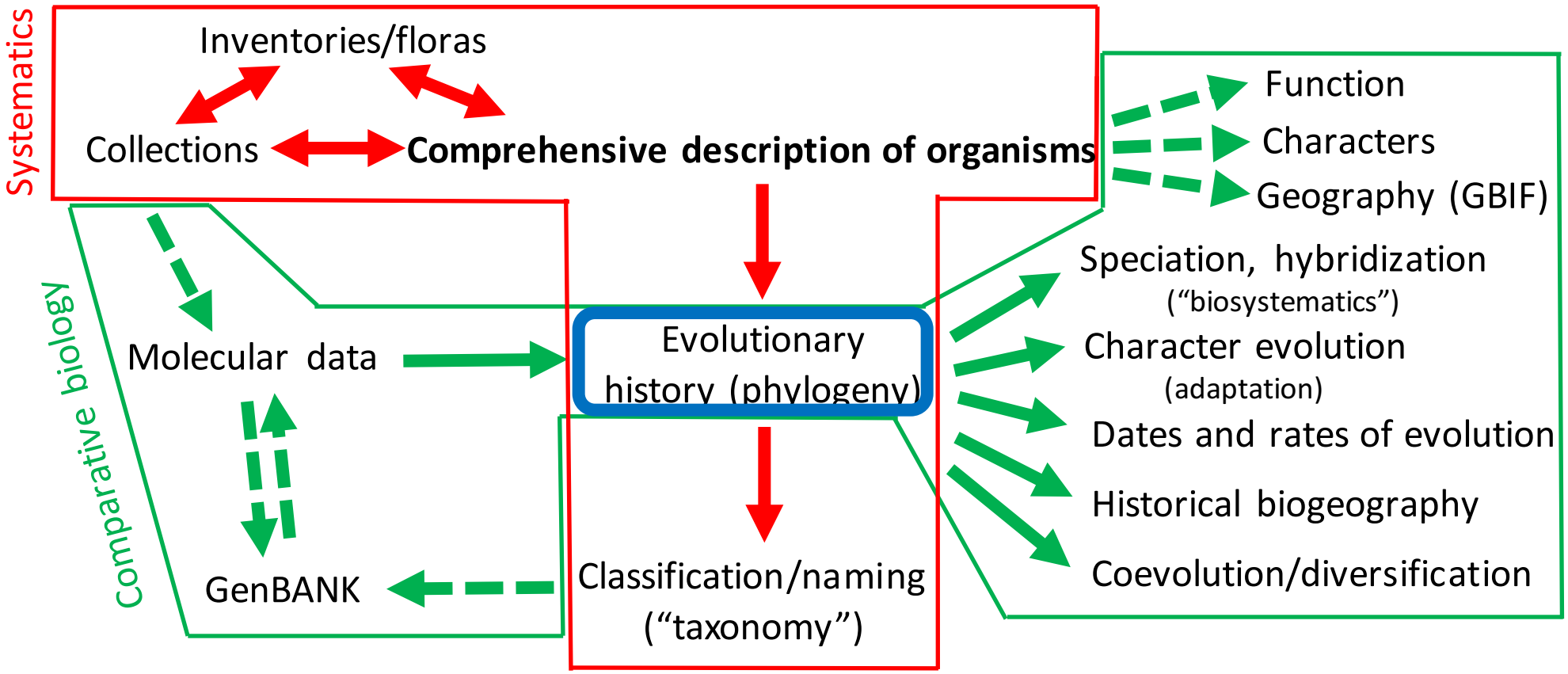


Systematists can switch clades.

Comparative biologists can switch phenomena/processes.

And, a single person can do both!

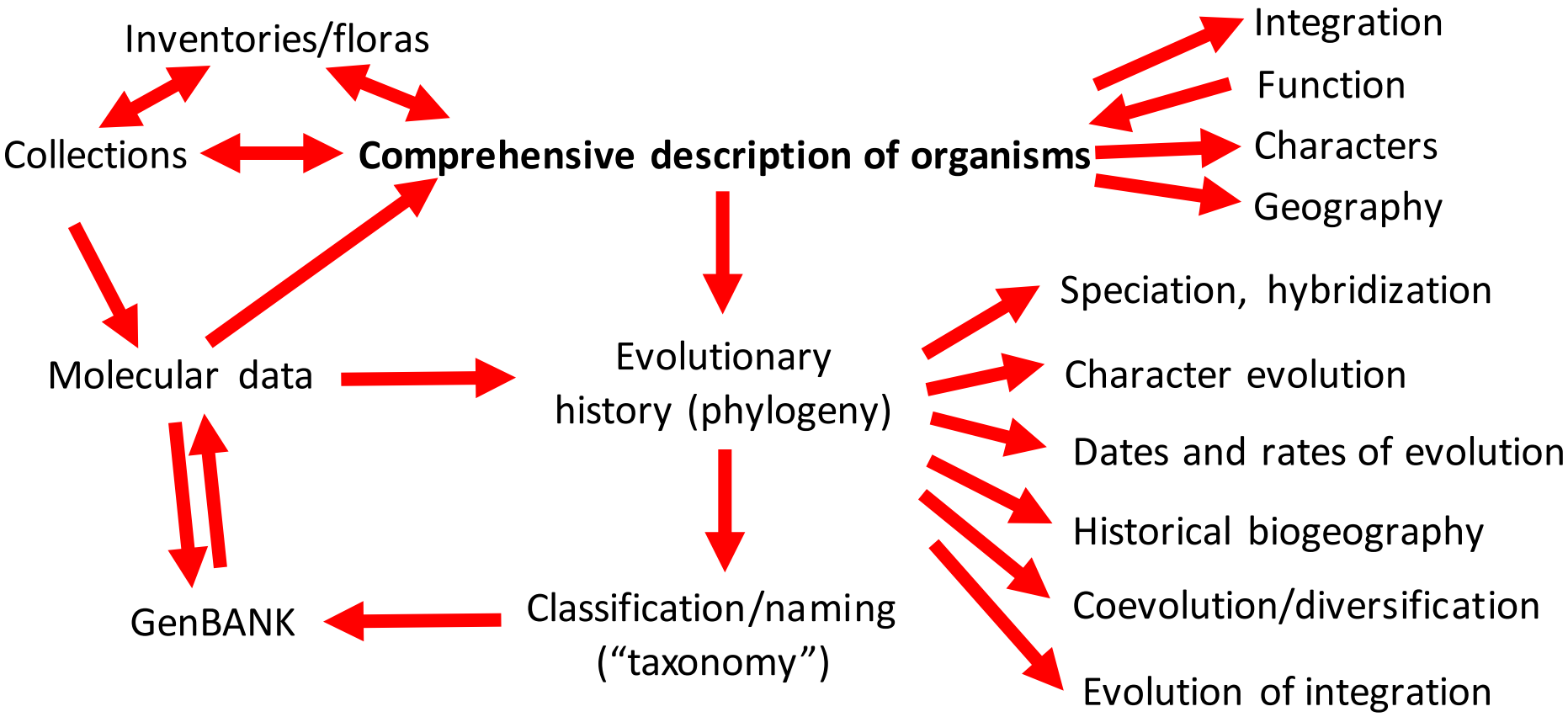
Fortunately, they have something very obvious in common, namely:



So, maybe we could call the whole thing PHYLOGENETIC BIOLOGY?!?

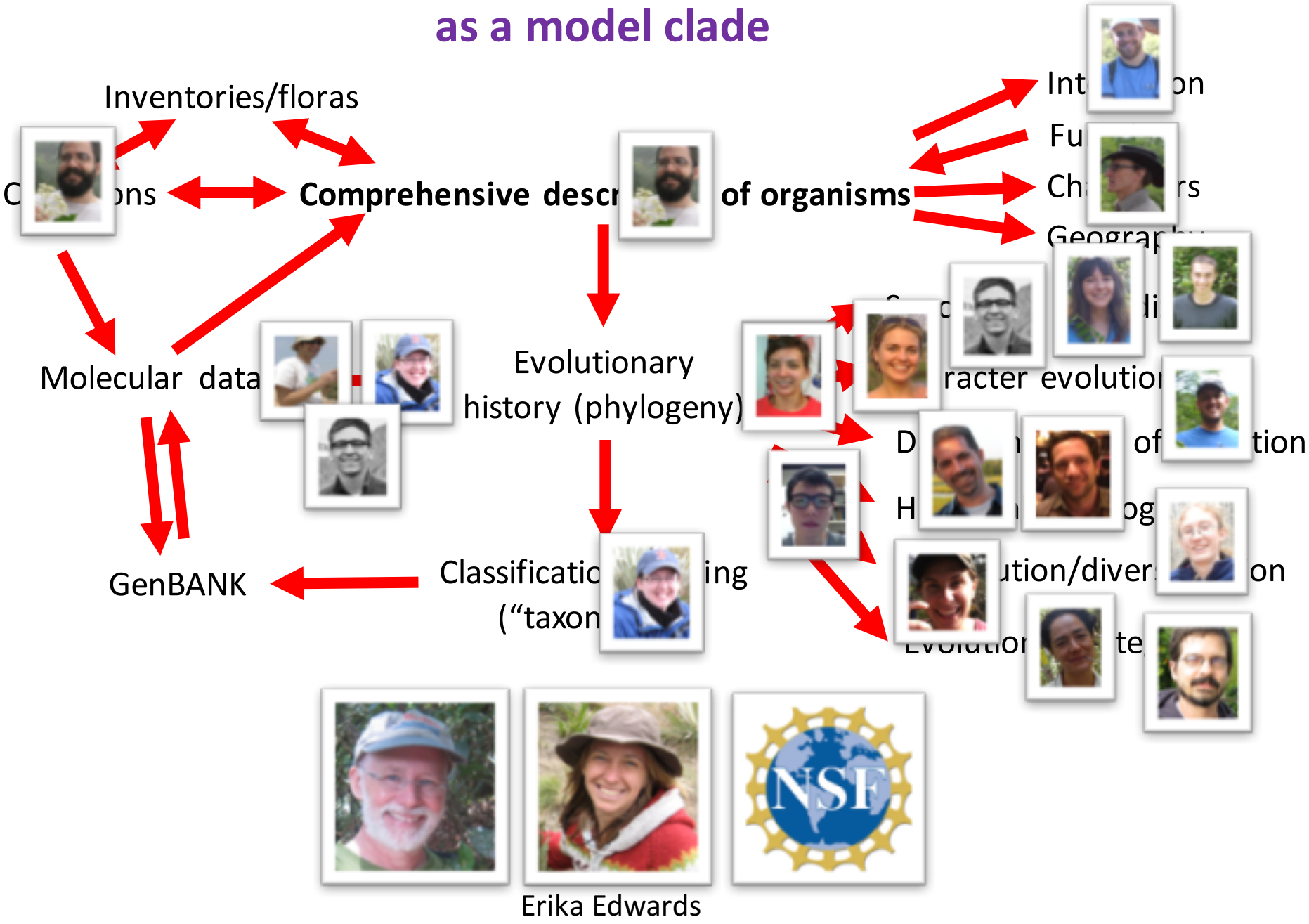
(everything in evolutionary biology that centers around phylogeny, broadly conceived to include reticulation too)

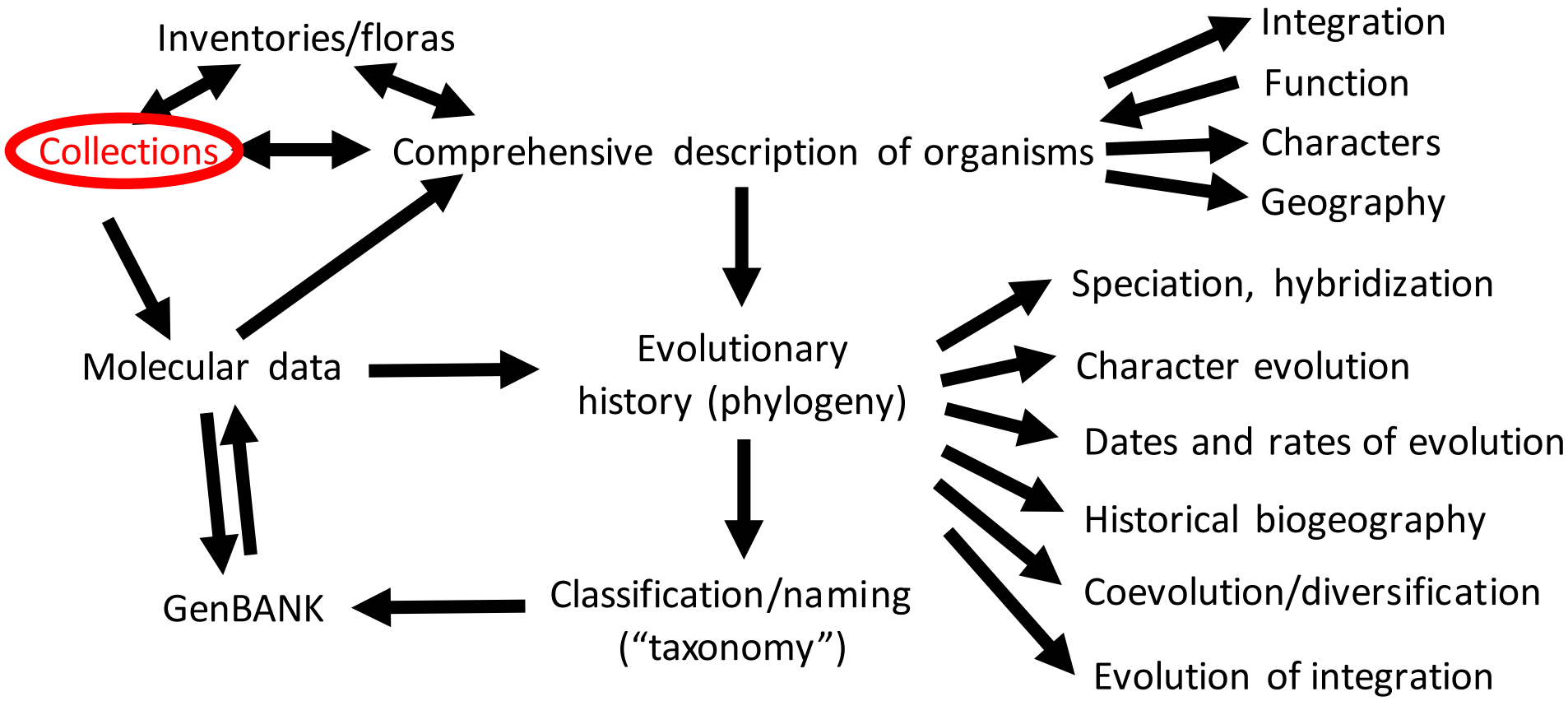
What if people (as a team, over many years) did all of this stuff for a particular clade?

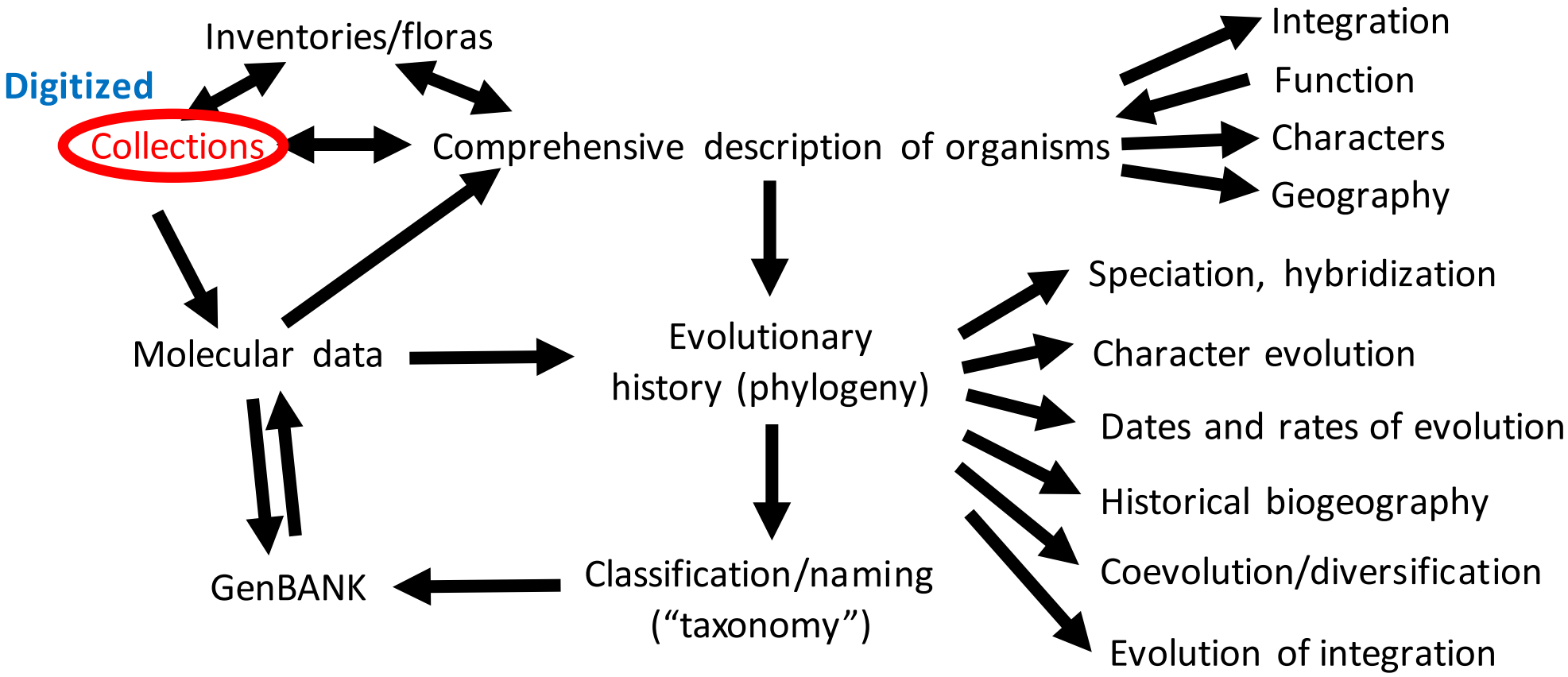


Then it would be a MODEL CLADE (model lineage)!

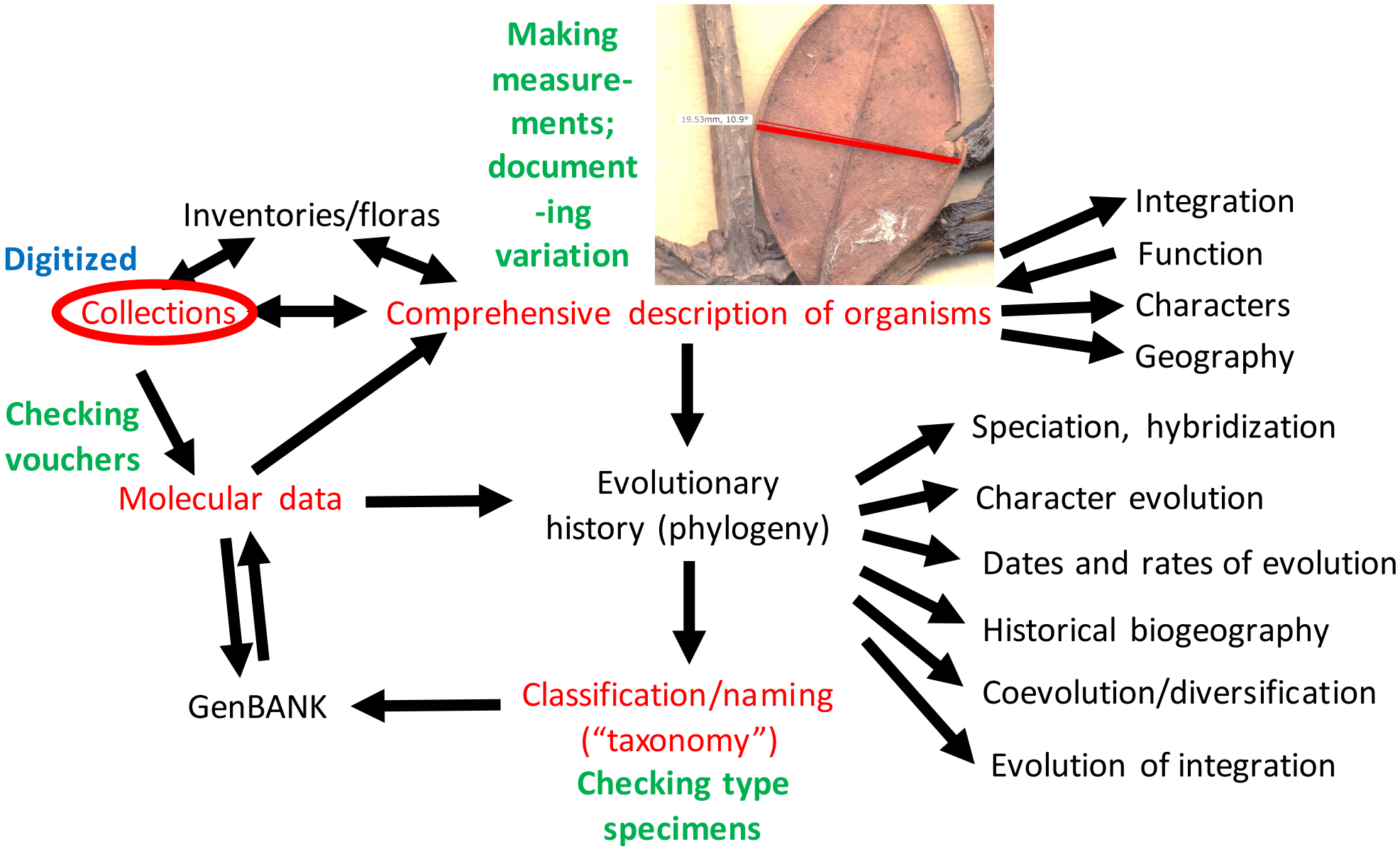
We've been trying to develop *Viburnum* (~165 species of woody flowering plants) as a model clade



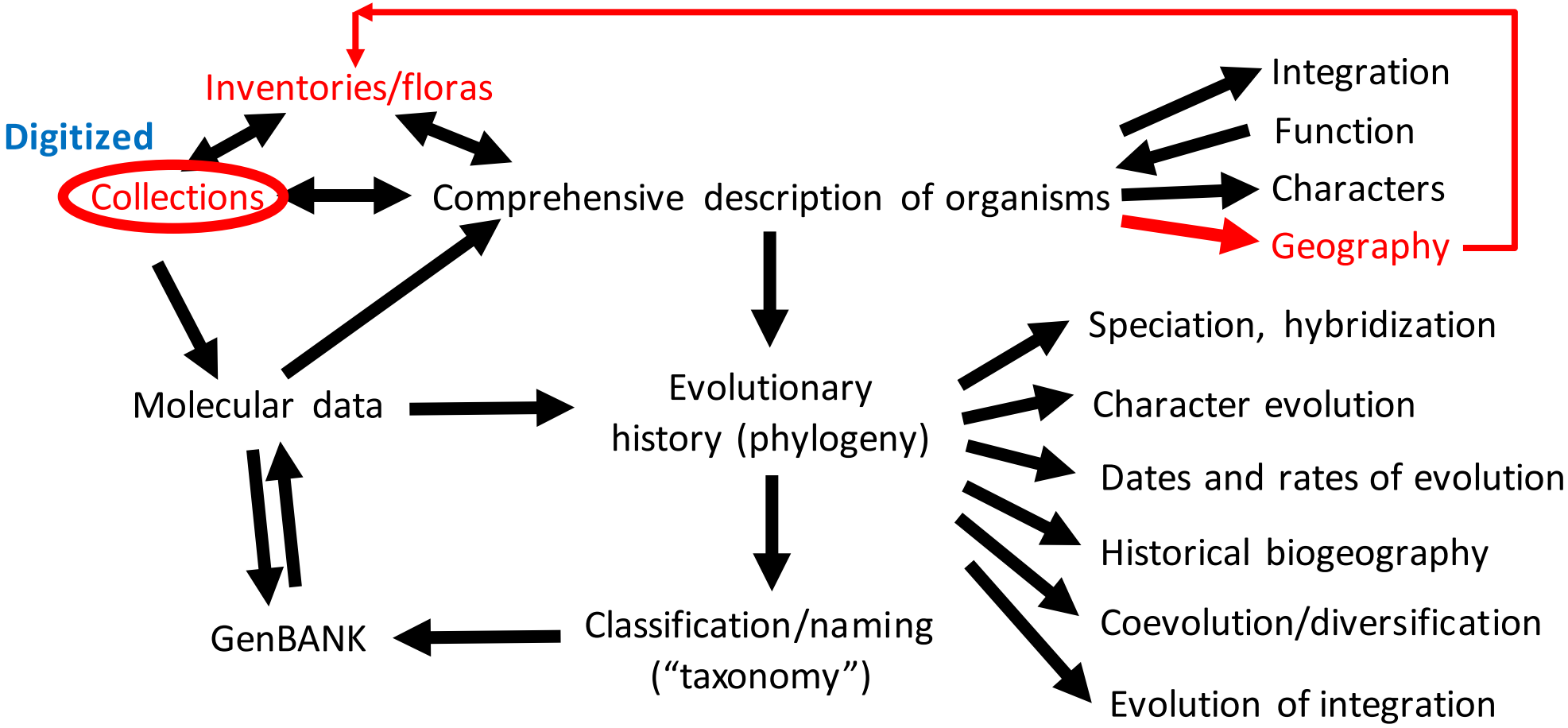

Inventories/floras
Collections
Comprehensive description of organisms
Integration
Function
Characters
Geography
Molecular data
Evolutionary history (phylogeny)
Speciation, hybridization
Character evolution
Dates and rates of evolution
Historical biogeography
Coevolution/diversification
Evolution of integration
GenBANK
Classification/naming ("taxonomy")



Important note: having only label data and images cuts out a bunch of key uses of specimens, e.g., DNA extraction, chemical/isotopic analyses, internal anatomy, tiny parts (e.g., pollen), many "functional traits", etc.

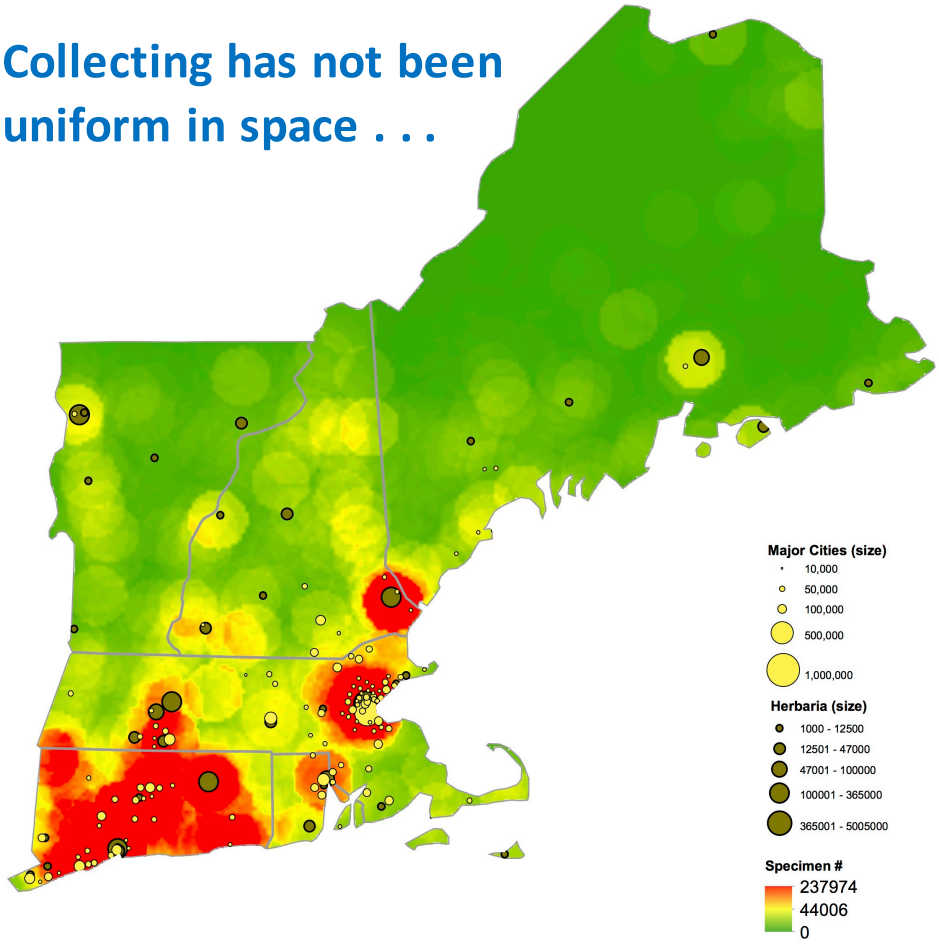


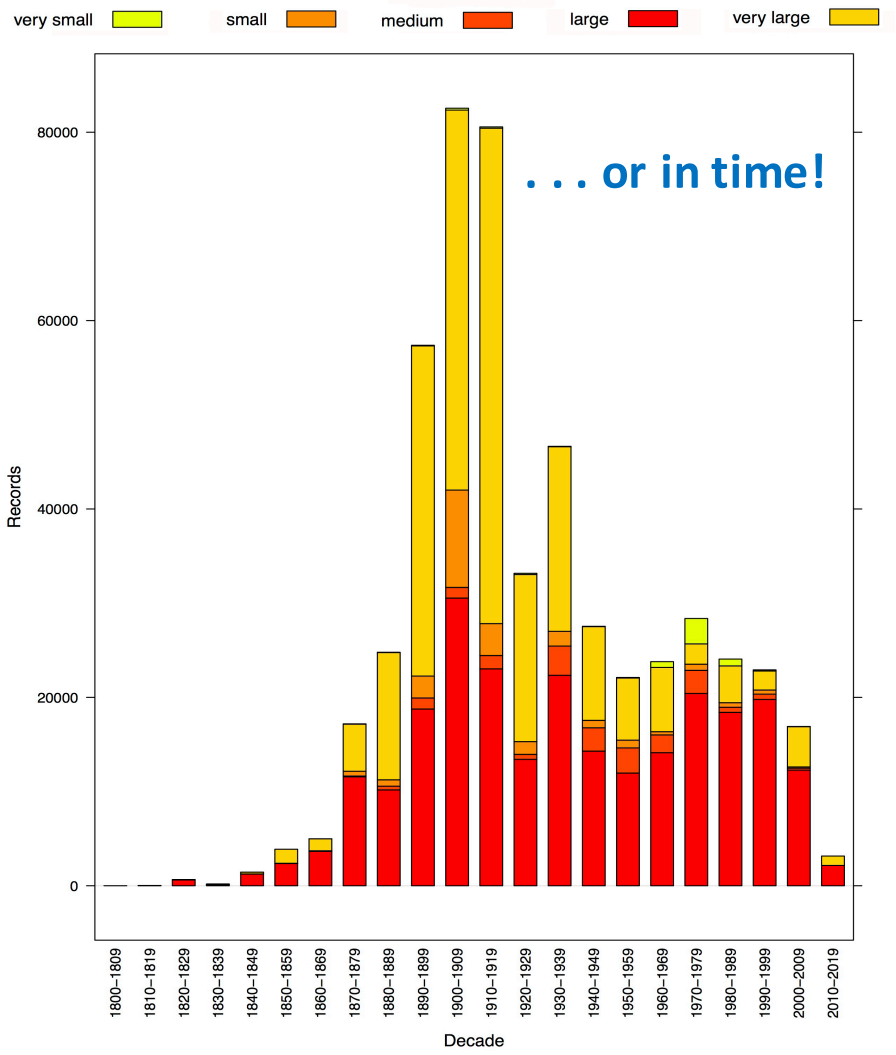
However, the ready accessibility of digital data -- and the ability to aggregate -- brings with it huge benefits!


Inventories/floras
Digitized
Collections
Comprehensive description of organisms
Integration
Function
Characters
Geography
Molecular data
Evolutionary history (phylogeny)
Speciation, hybridization
Character evolution
Dates and rates of evolution
Historical biogeography
Coevolution/diversification
Evolution of integration
GenBANK
Classification/naming ("taxonomy")

“Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes”, Patrick Sweeney (PI)

Collecting has not been uniform in space . . .



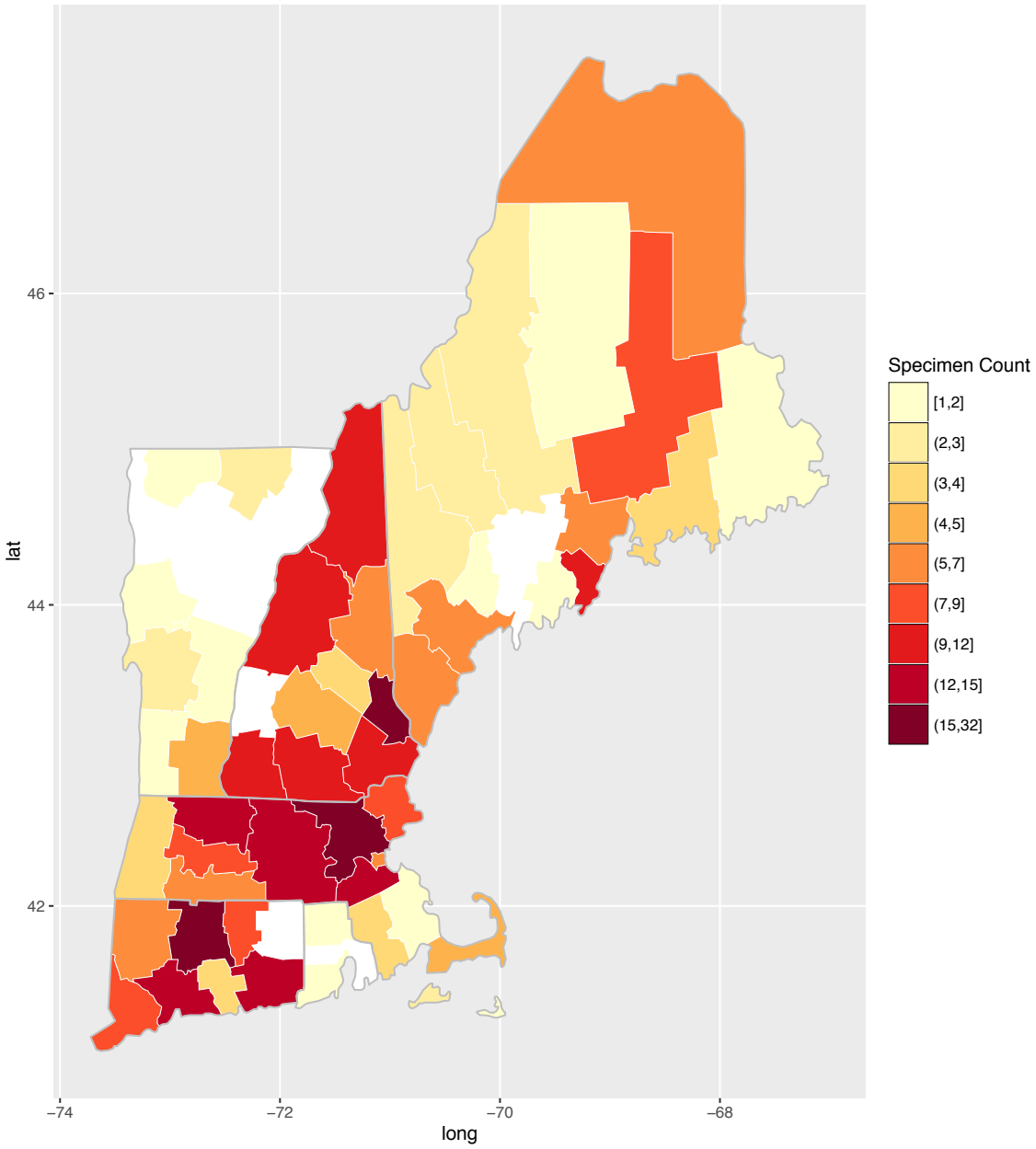


The NEVP TCN includes 27 New England herbaria, and has already generated 950,761 images

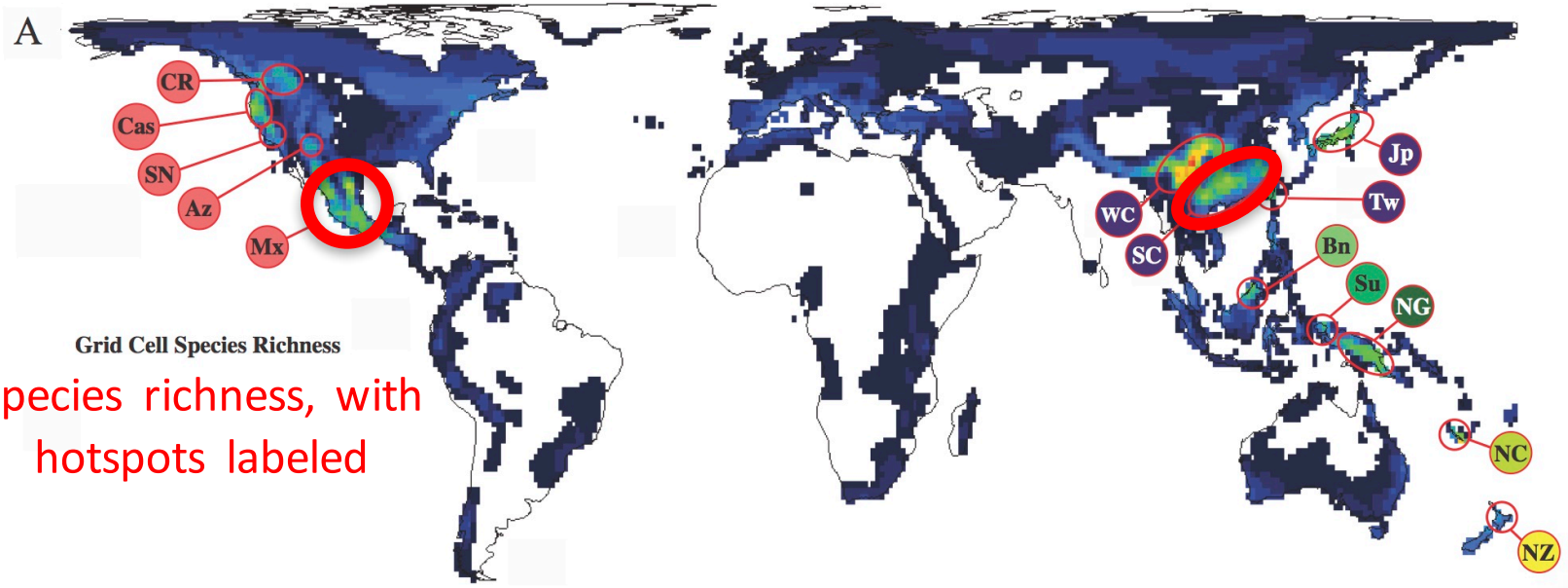
Taraxacum officinale (dandelion)

529 records, from 78% of the NE Counties, but only 11% of the Towns

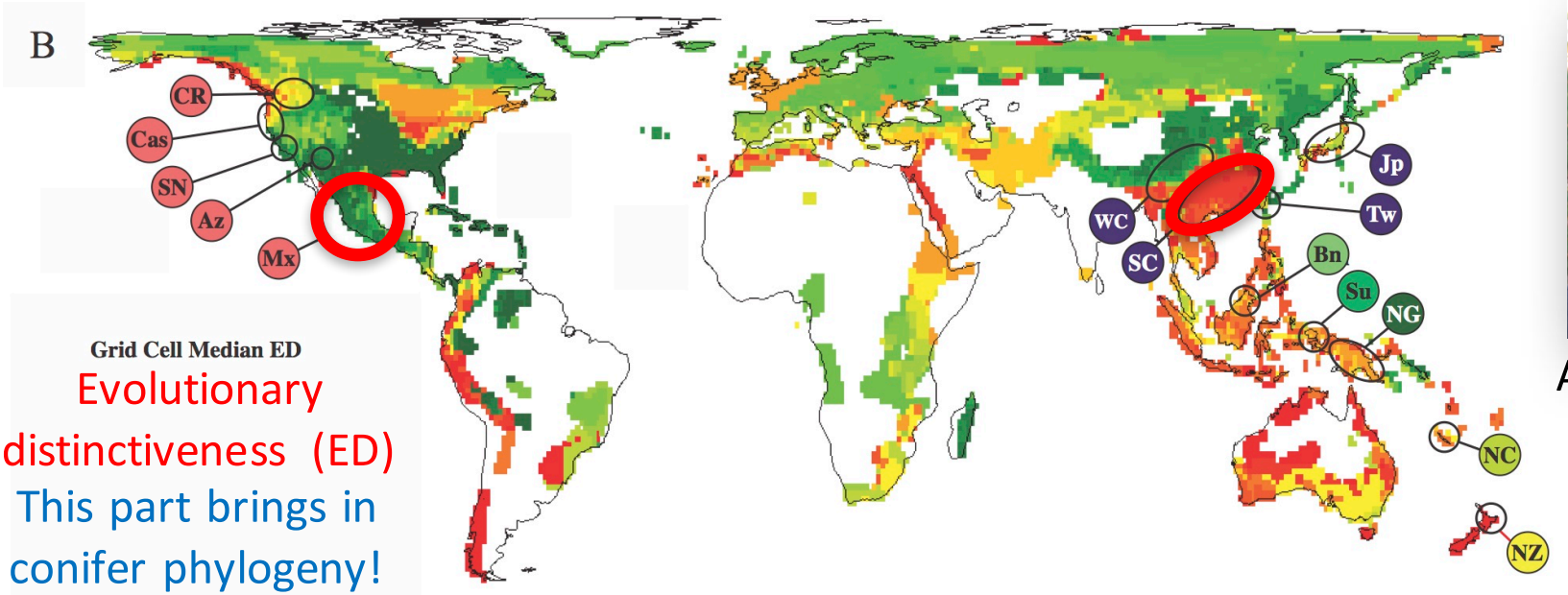
Patrick Sweeney



Based on maps/niche models for 630 species of conifers!



Species richness, with hotspots labeled



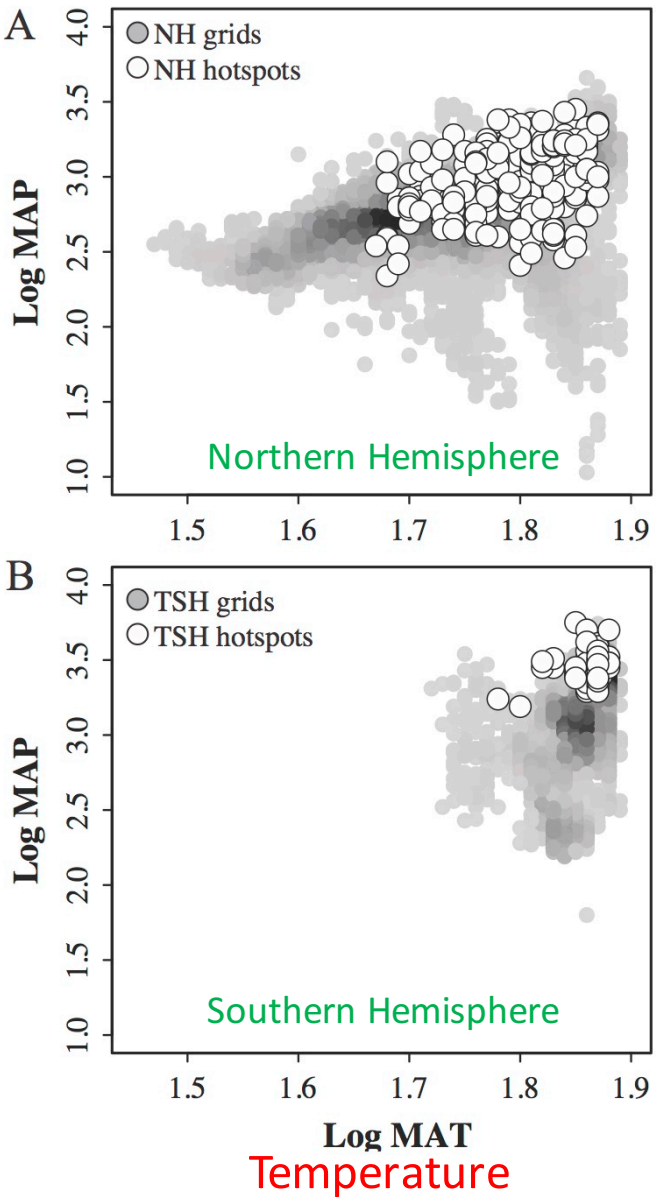
Evolutionary distinctiveness (ED)

This part brings in conifer phylogeny!

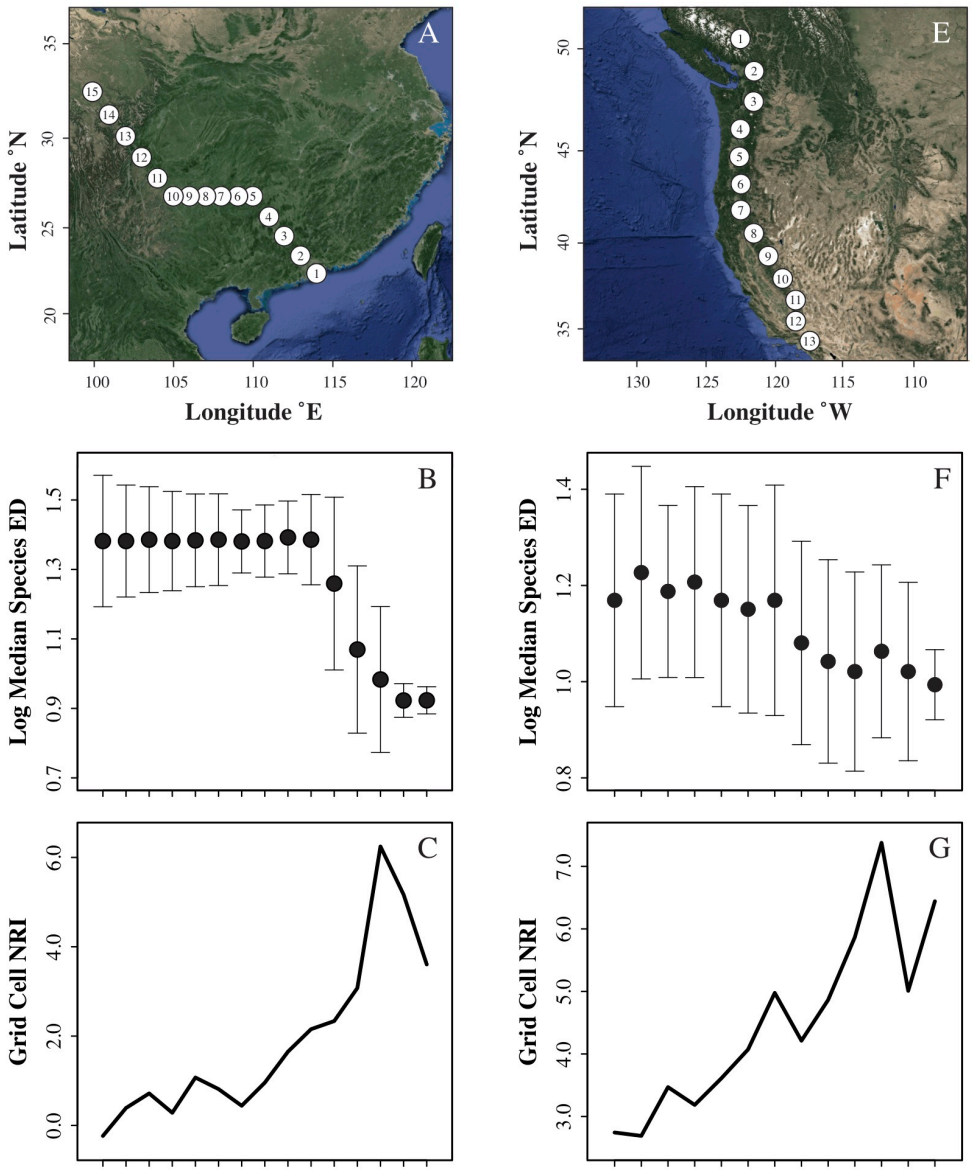
Andrew Leslie
Brown U.

Andrew Leslie, Sarah Mathews, Michael Donoghue, Jeremy Beaulieu, Walter Jetz et al. - Conifer diversity hotspots along a museum-to-cradle continuum. MS in prep.

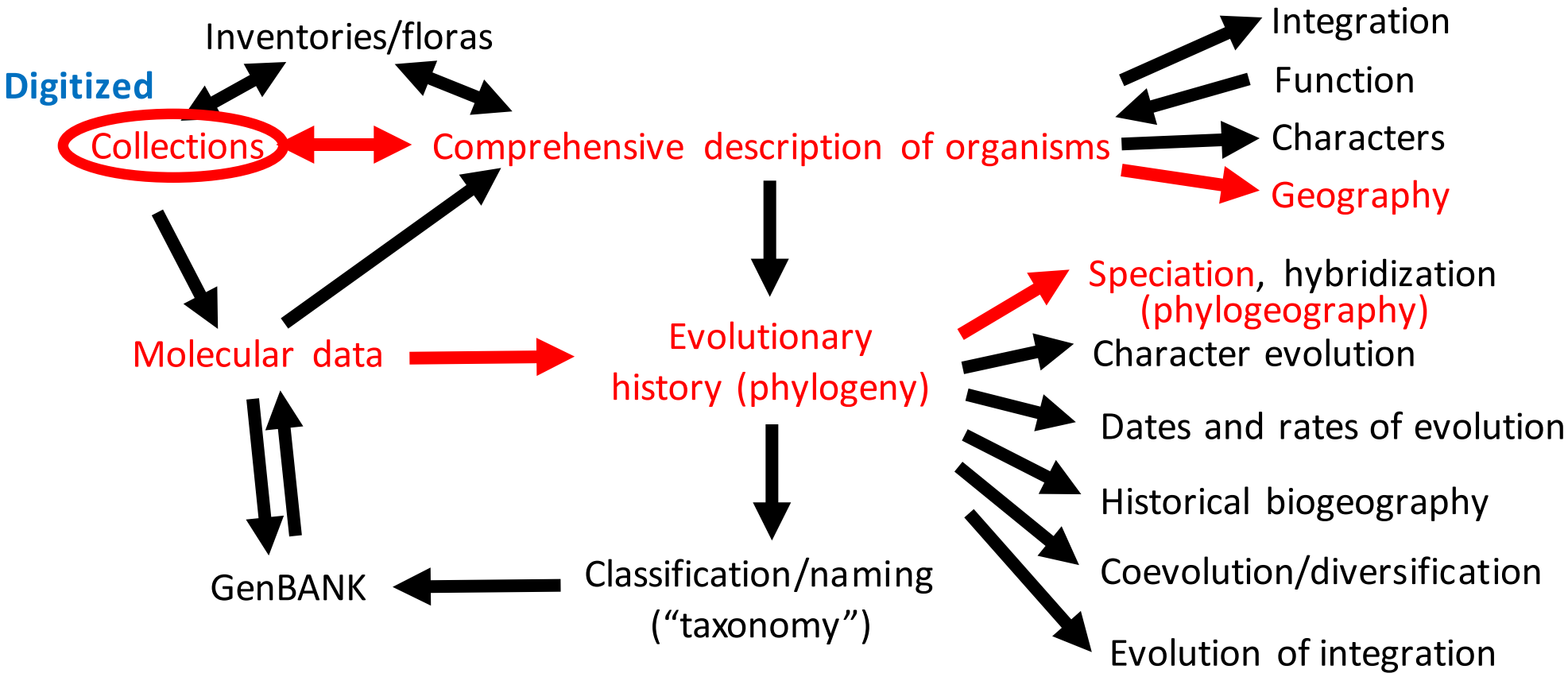
Conifers hotspots are in wetter warmer places



Changes in ED and NRI along two transects

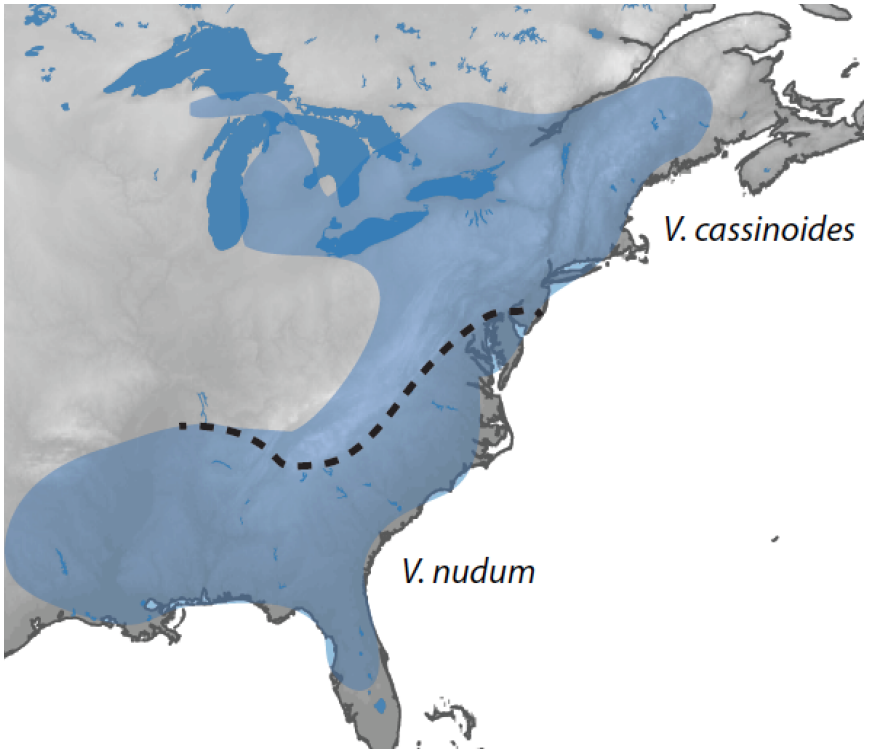


Andrew Leslie, Sarah Mathews, Michael Donoghue, Jeremy Beaulieu, Walter Jetz et al. - Conifer diversity hotspots along a museum-to-cradle continuum. MS in prep.


Inventories/floras
Digitized
Collections
Comprehensive description of organisms
Integration
Function
Characters
Geography
Molecular data
Evolutionary history (phylogeny)
Speciation, hybridization (phylogeography)
Character evolution
Dates and rates of evolution
Historical biogeography
Coevolution/diversification
Evolution of integration
GenBANK
Classification/naming ("taxonomy")

Current status of the *Viburnum nudum* species complex

Geographic range

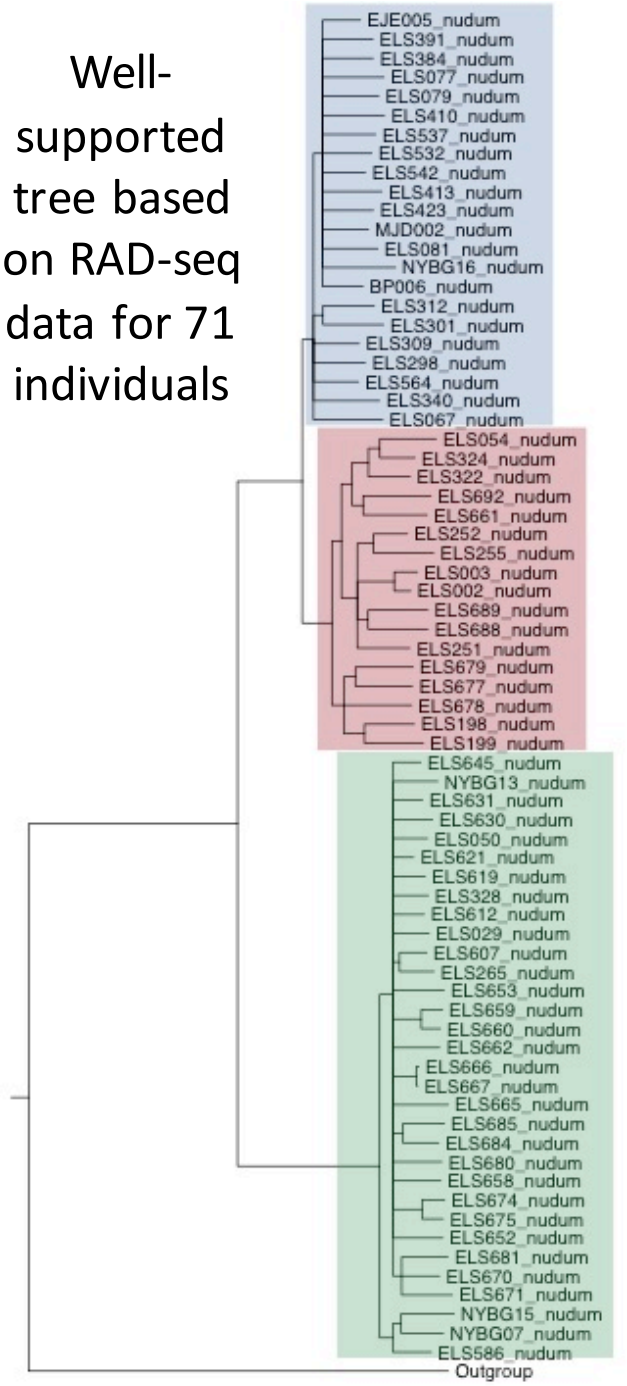


The complex is usually divided into:

Southern=*V. nudum*

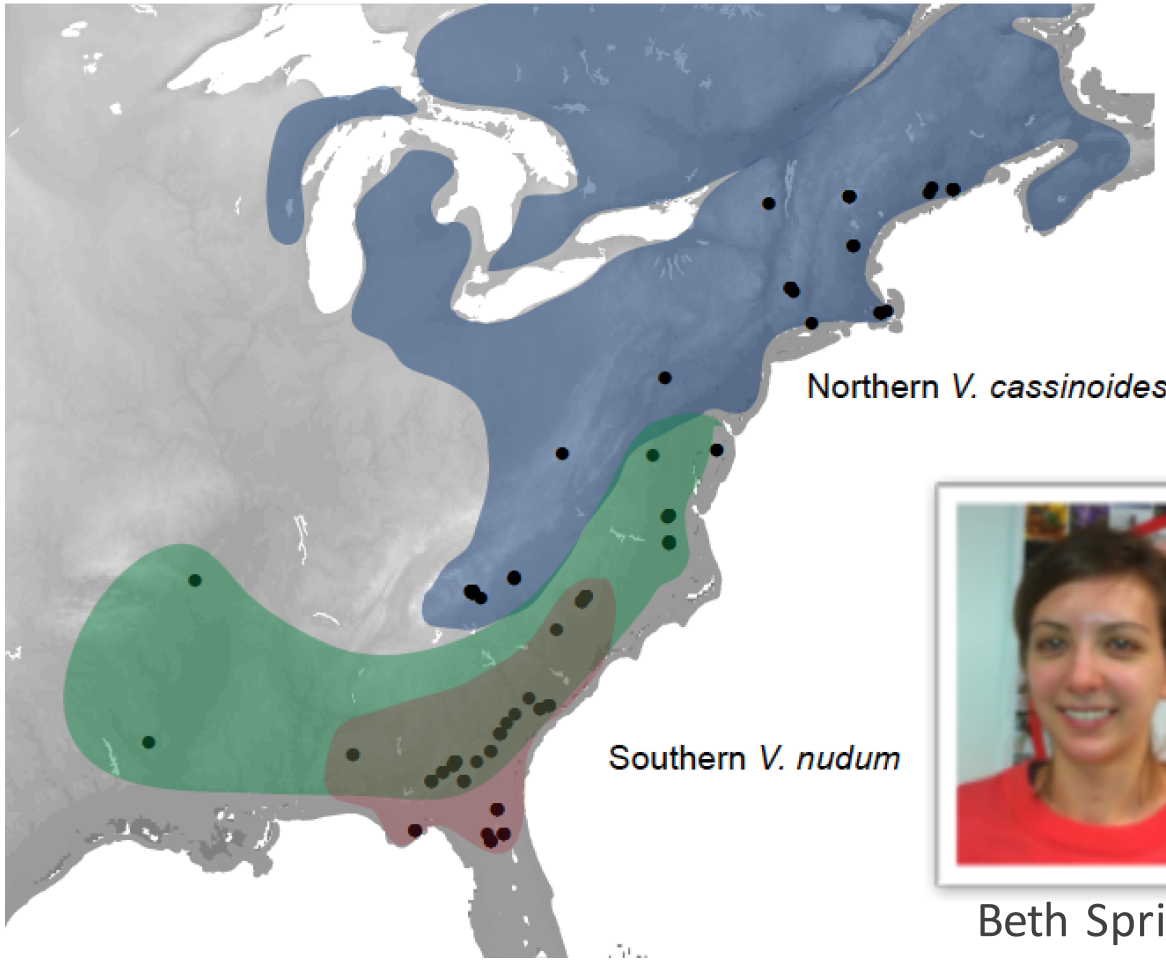
Northern=*V. cassinoides*

Well-supported tree based on RAD-seq data for 71 individuals



There are three well-supported clades

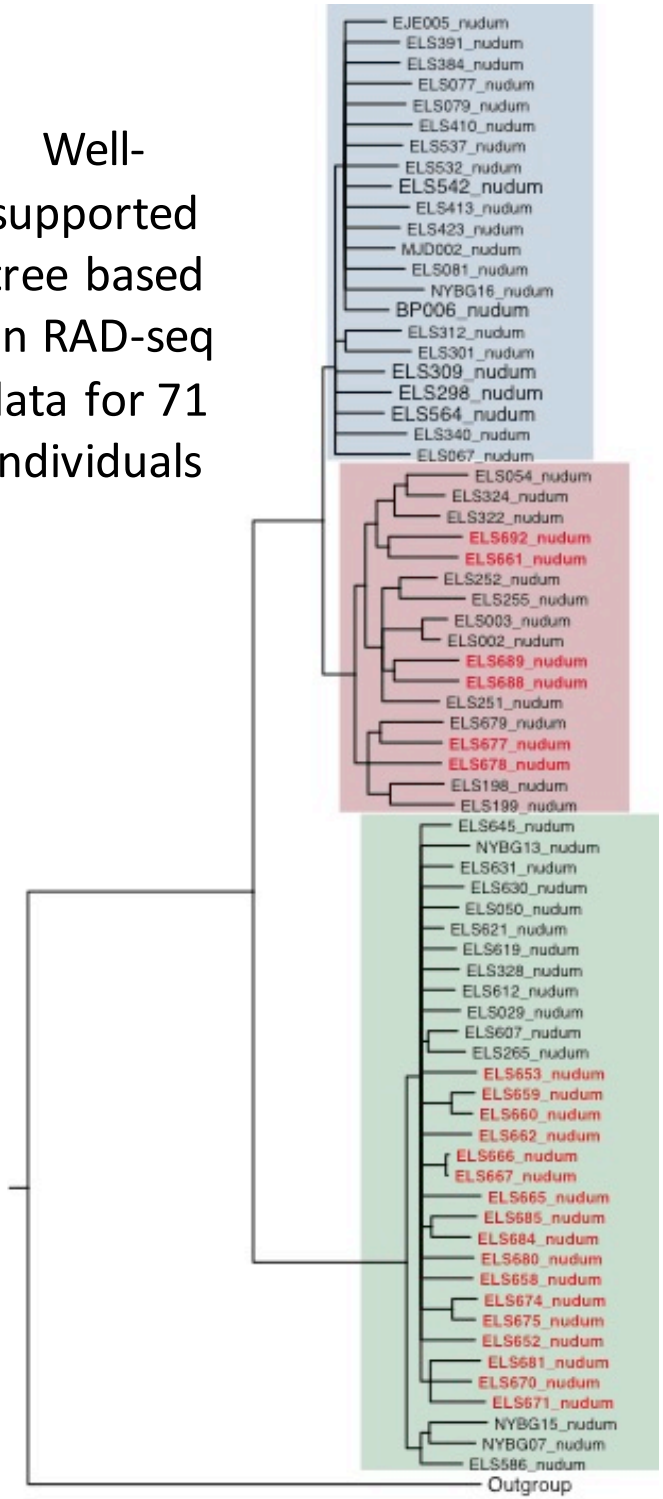
Traditional '*V. nudum*' – in the south – is clearly paraphyletic



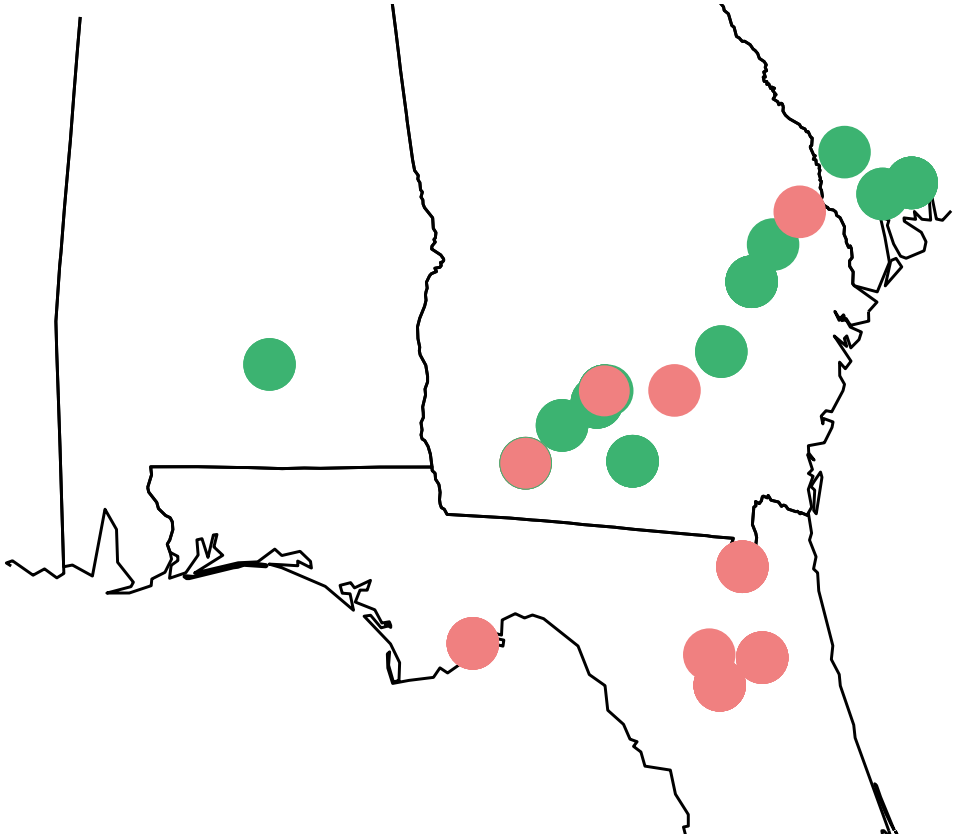
Beth Spriggs

139 individuals were collected from 45 populations

Well-supported tree based on RAD-seq data for 71 individuals



Individuals from the red and the green clades occur in close proximity in Georgia. They're basically sympatric!



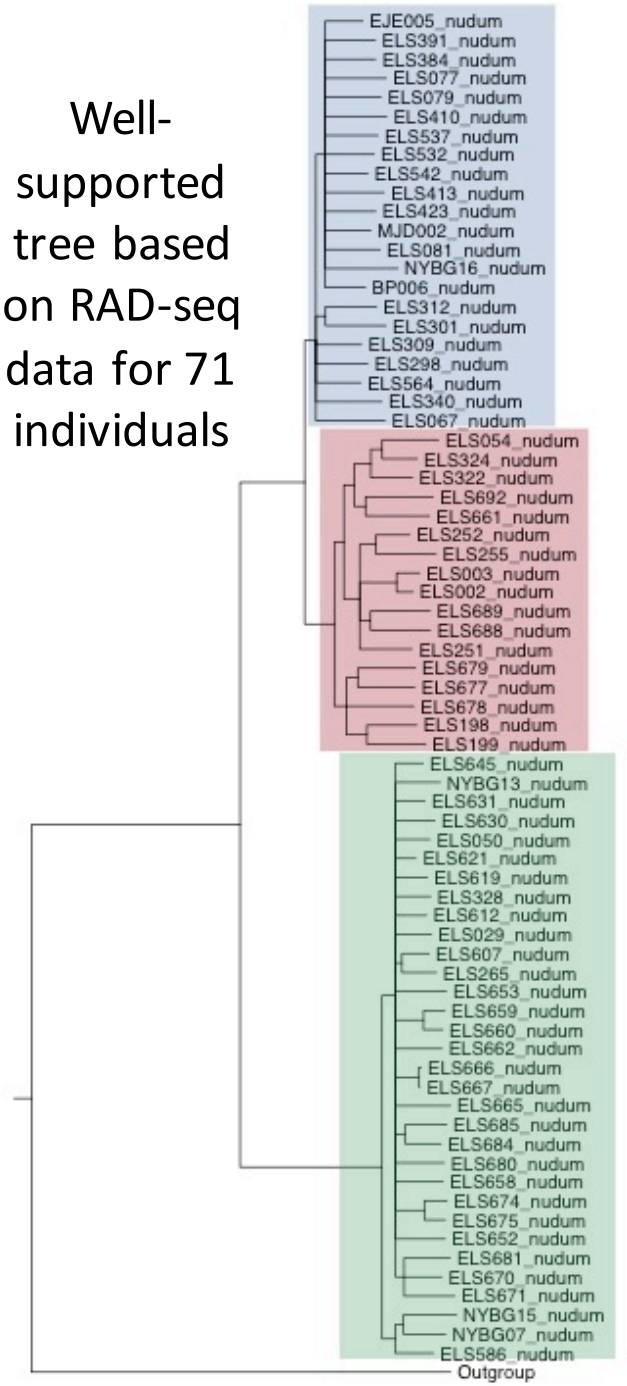
But, they differ in habitat – the green ones are in bald-cypress swamps!

Members of the green and red clades turn out to differ in leaf shape, and can be ID'd from herbarium specimens

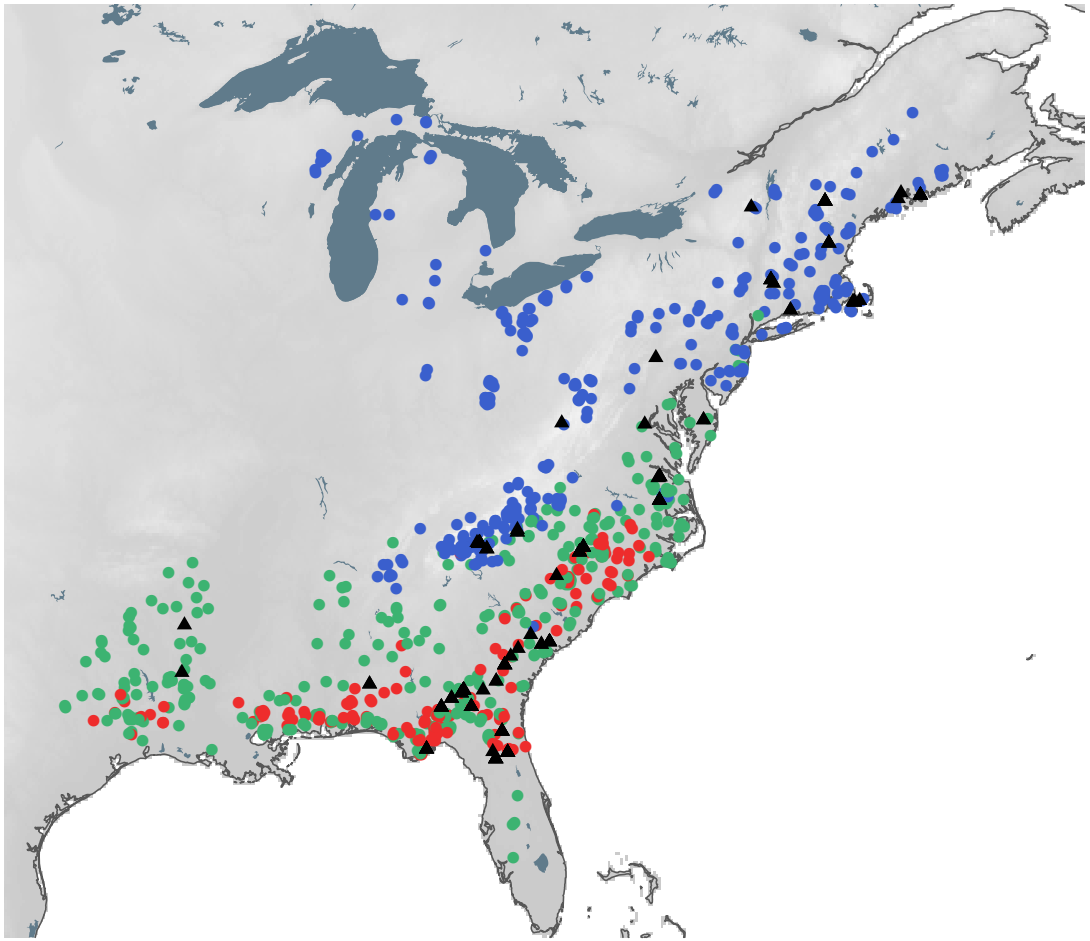
Green type

Red type

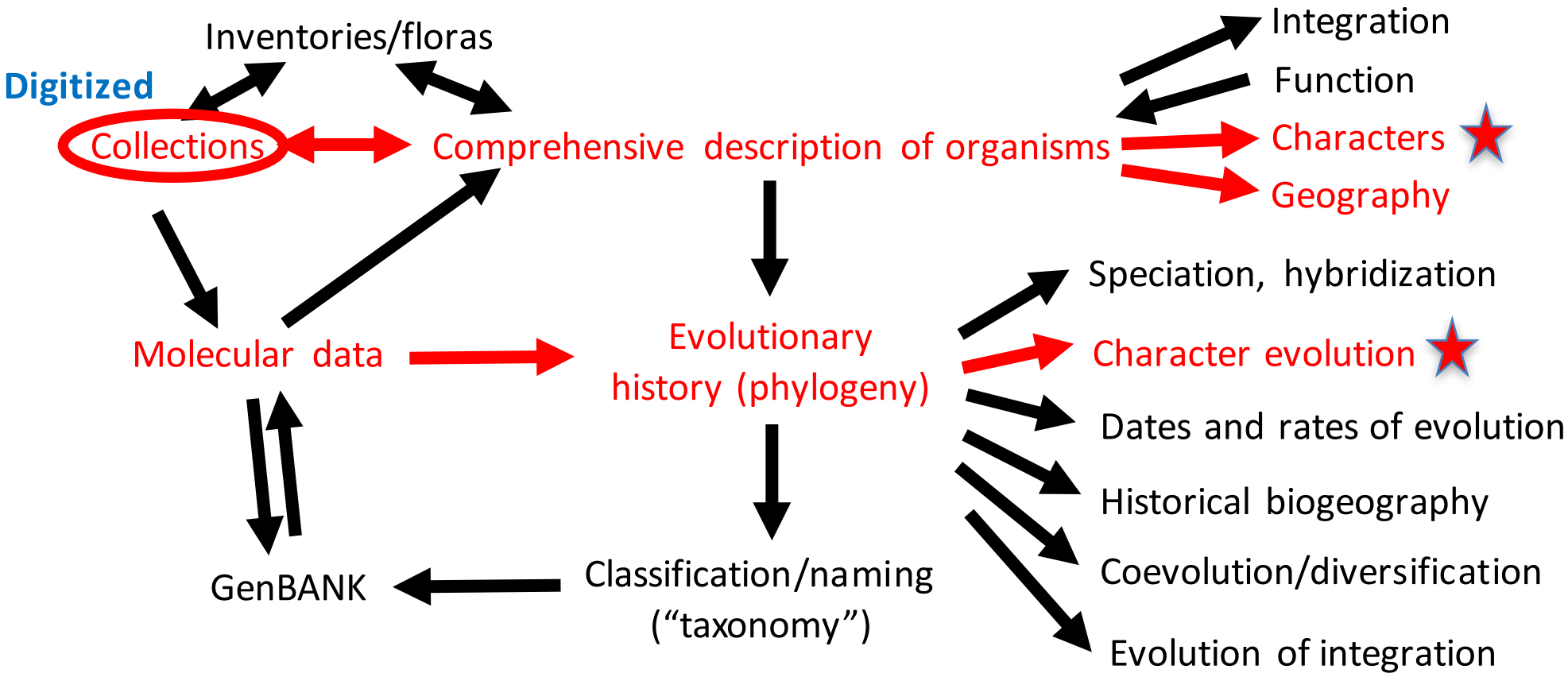
Well-supported tree based on RAD-seq data for 71 individuals



Now we can draw a proper range map based on >1000 images of specimen

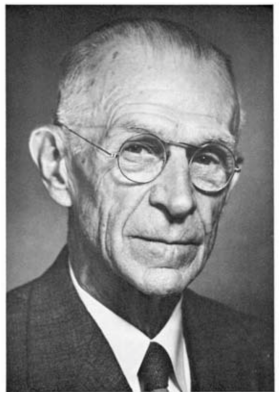


Images from: NYBG, University of Connecticut, Louisiana State University, Ohio State University, Florida State University, University of Wisconsin, University of North Carolina, University of Texas, and SERNEC.

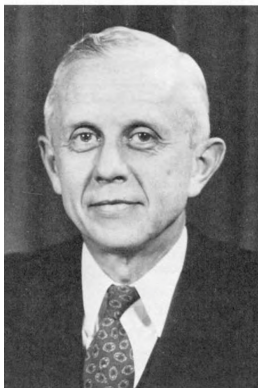

Inventories/floras
Digitized
Collections
Comprehensive description of organisms
Integration
Function
Characters
Geography
Molecular data
Evolutionary history (phylogeny)
Speciation, hybridization
Character evolution
Dates and rates of evolution
Historical biogeography
Coevolution/diversification
Evolution of integration
GenBANK
Classification/naming ("taxonomy")

"The Climatic Distribution of Certain Types of Angiosperm Leaves"

Irving W. Bailey



Edmund W. Sinnott



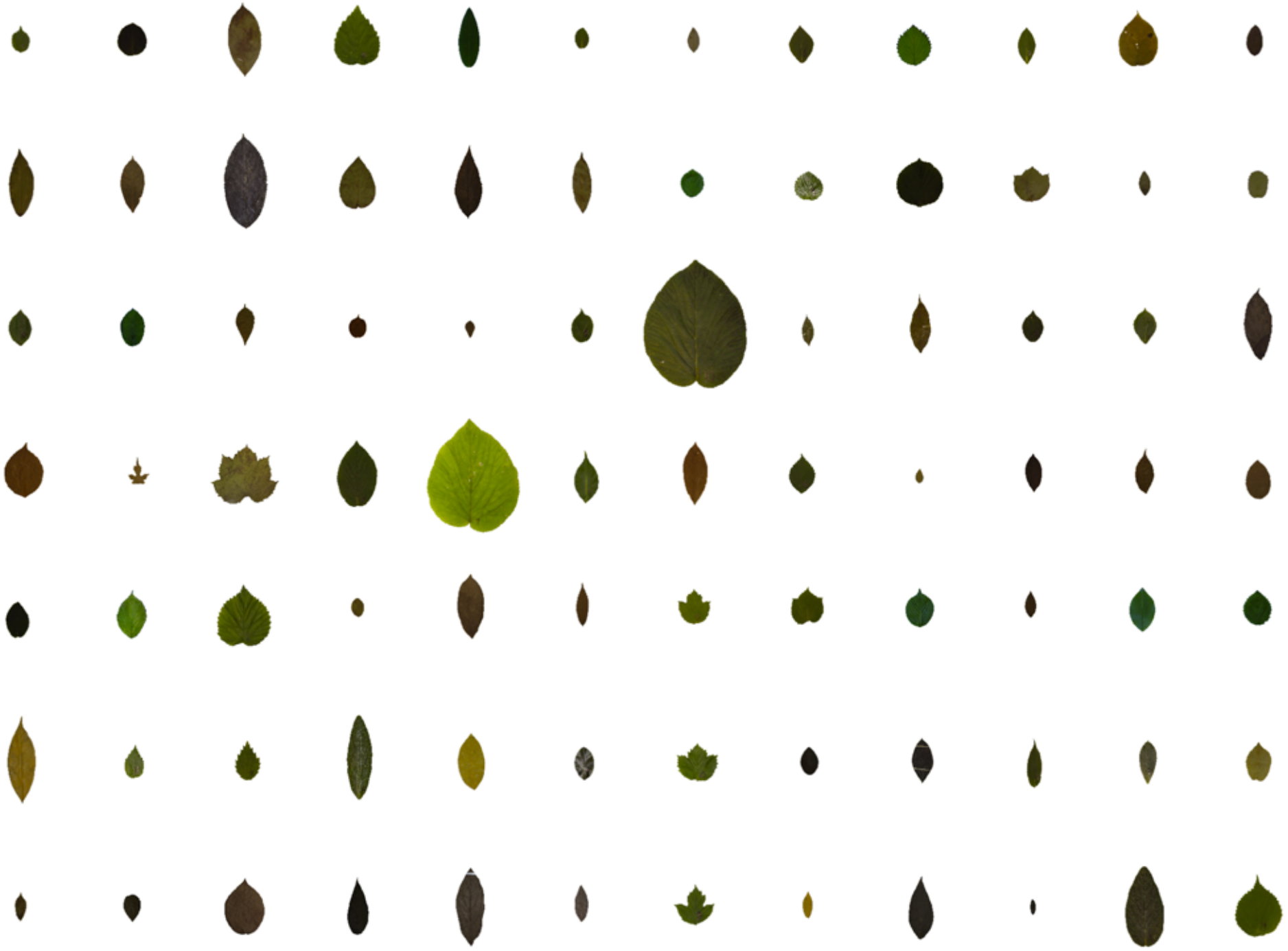
	Entire %		
	Trees, percent.	Shrubs, percent	Herbs, percent.
Mesophytic Cold-temperate (E. C. N. A.) . .	10	14	23
Lowland-tropical. (Brazil)	90	87	62

"entire" leaf margins in tropical forests

"lobed" and "toothed" leaf margins in temperate forests

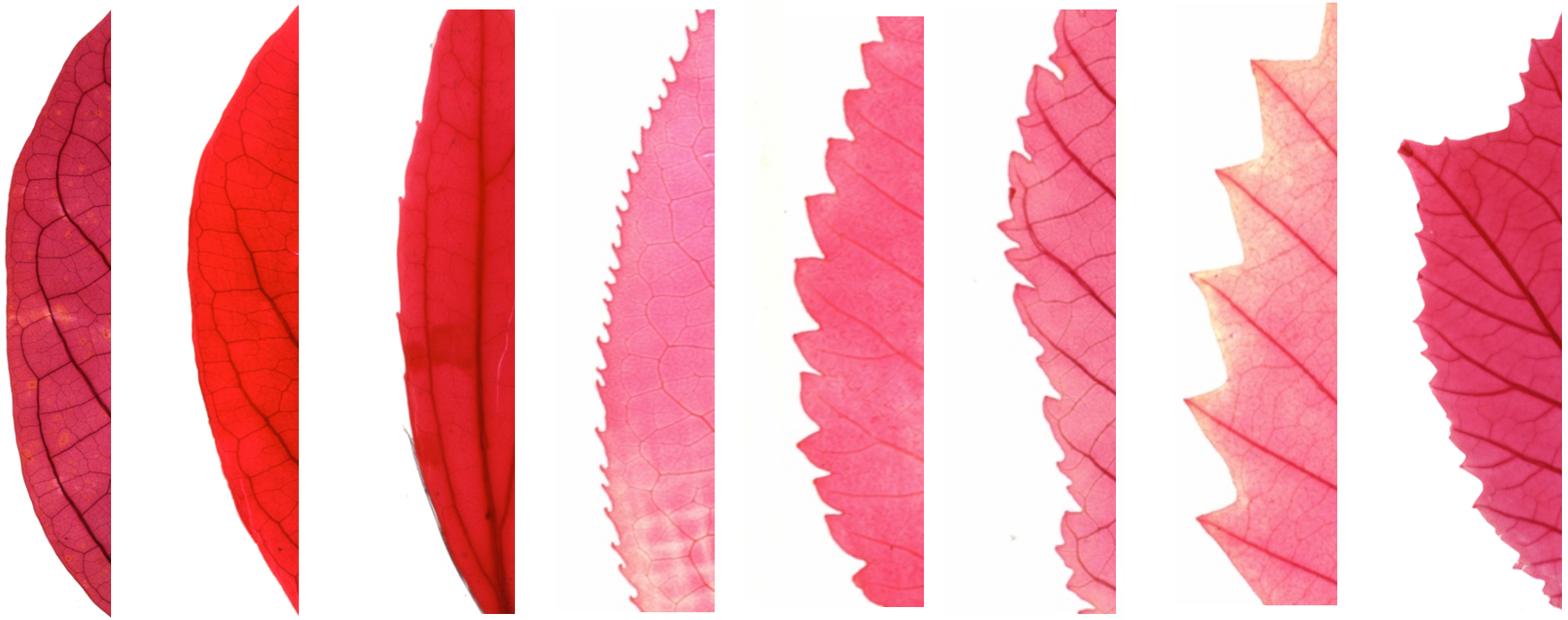
American Journal of Botany, Vol. 3, No. 1 (Jan., 1916), pp. 24-39

Viburnum leaf form diversity is pretty impressive!

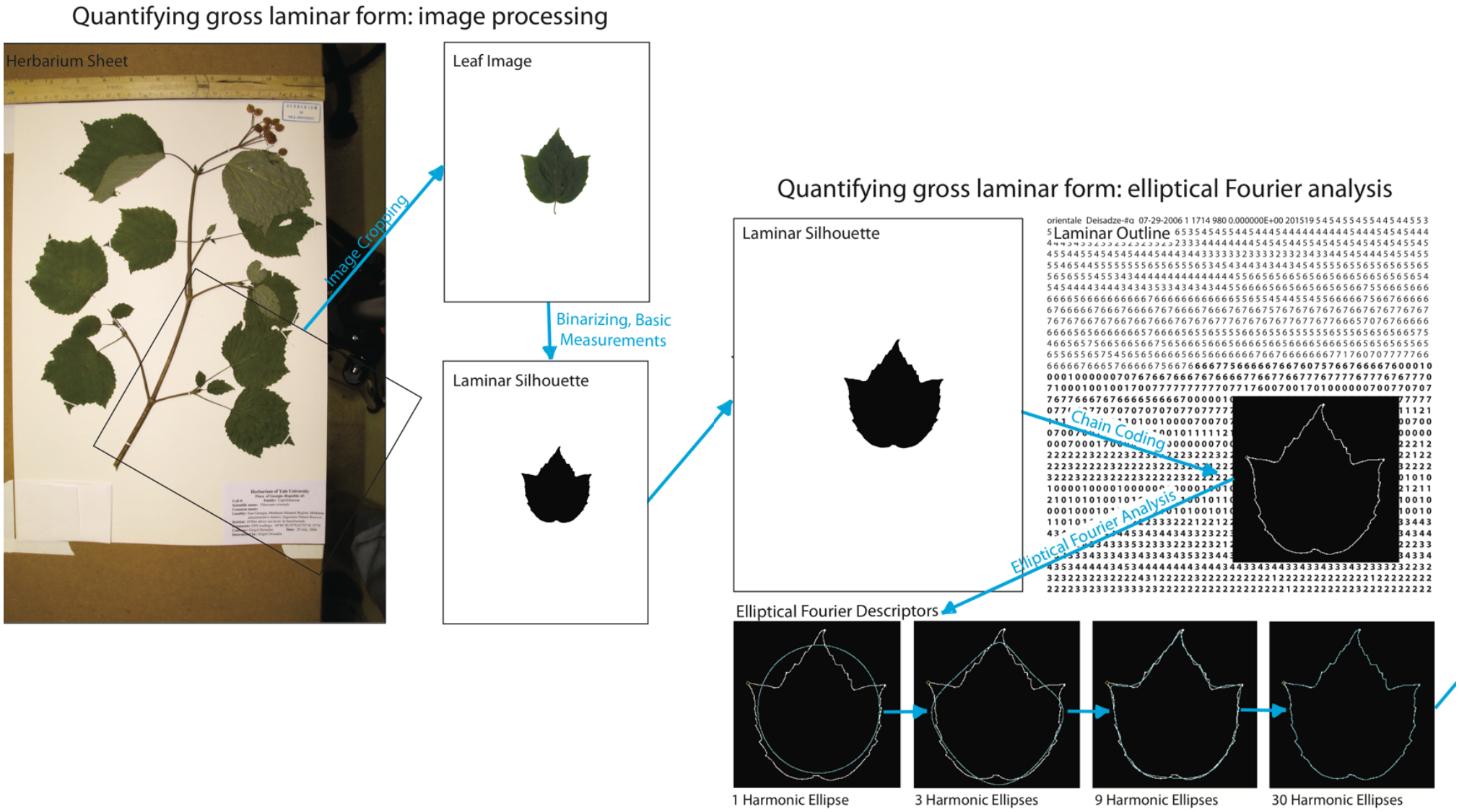


Viburnum species differ a lot in leaf shape

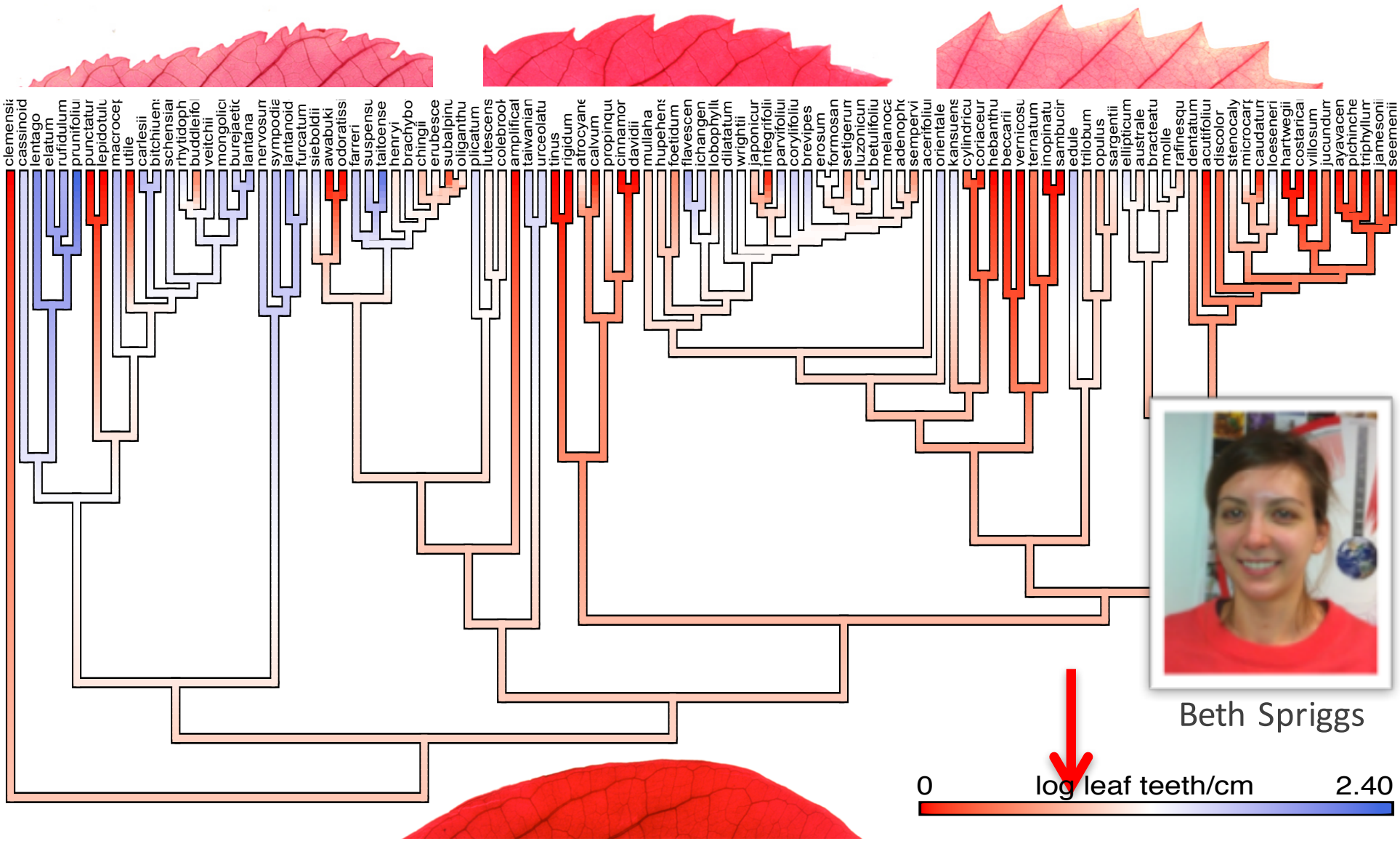
. . . and also in leaf margins!



Work-flow for extracting leaf silhouettes from herbarium specimens and carrying out Elliptical Fourier Analysis



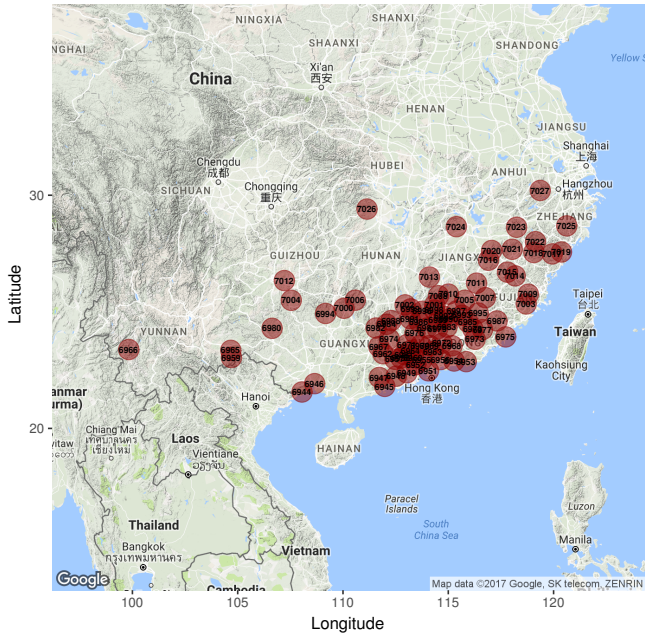
Leaf teeth/lobes evolved independently in multiple lineages (~10), with the occupation of cold temperate forests



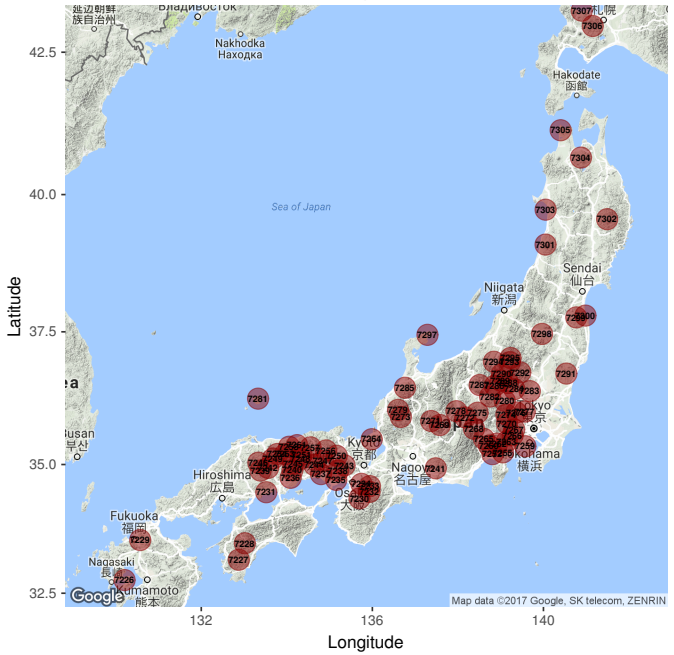
Viburnum probably started out with entire leaf margins (i.e., with few or no teeth)

We have good range maps of almost all *Viburnum* species

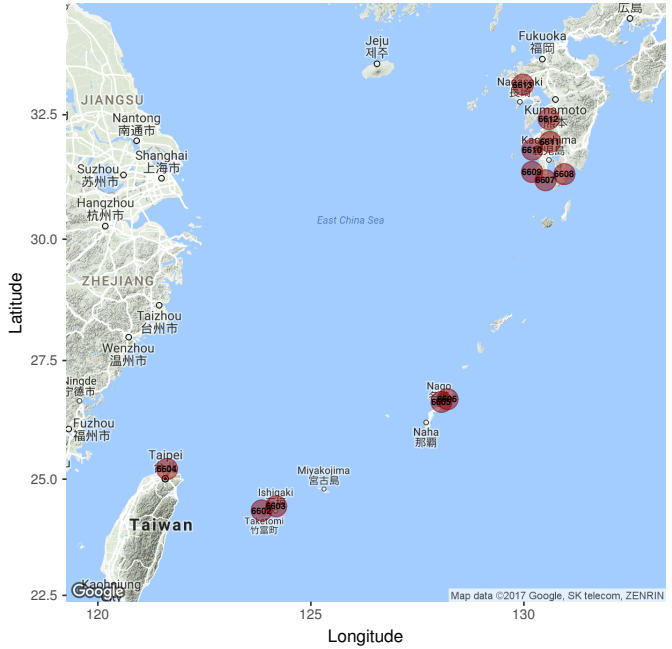
V. sempervirens



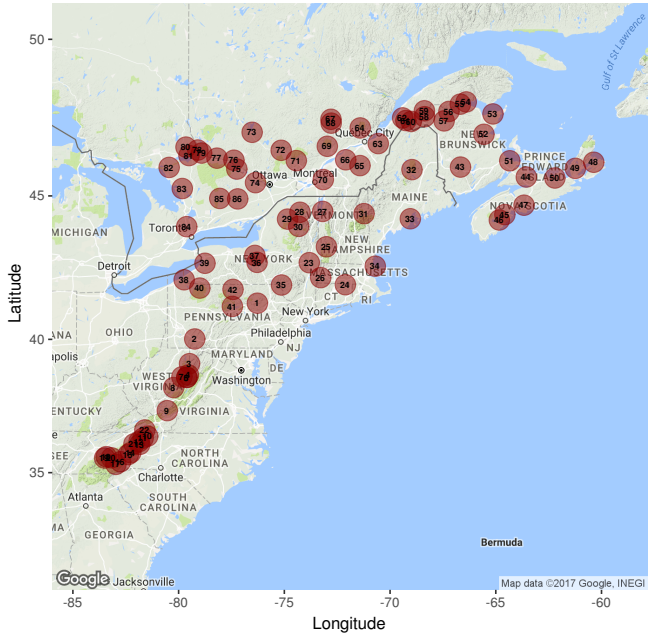
V. wrightii



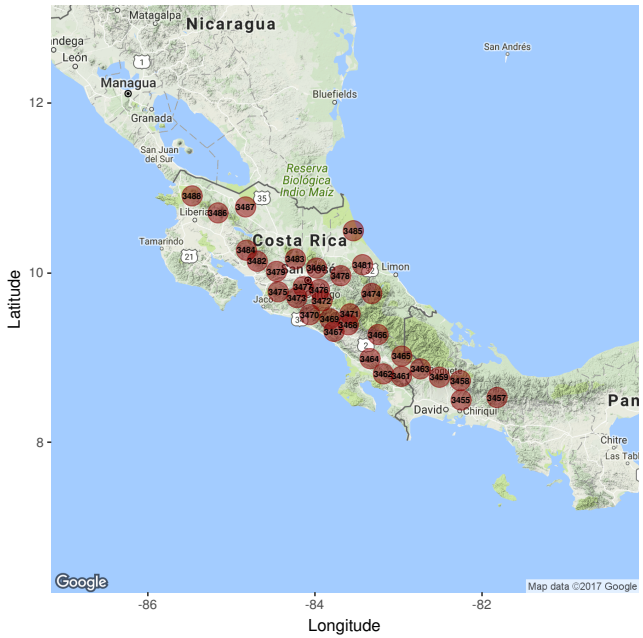
V. japonicum



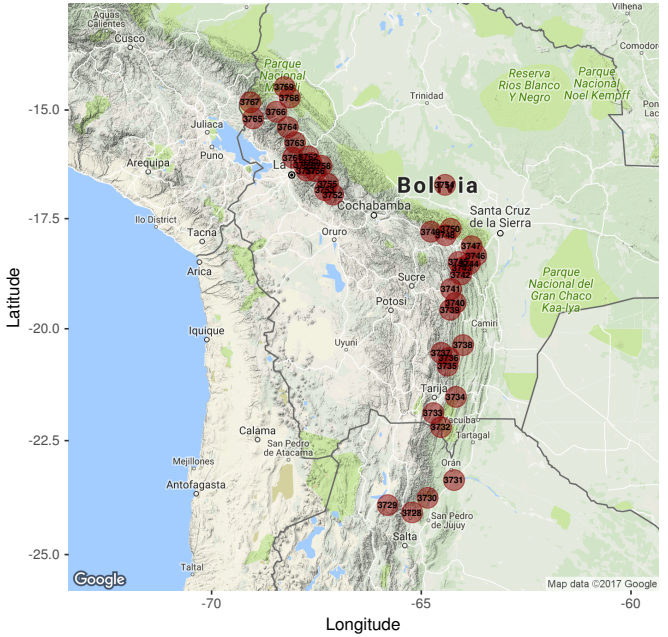
V. lantanoides



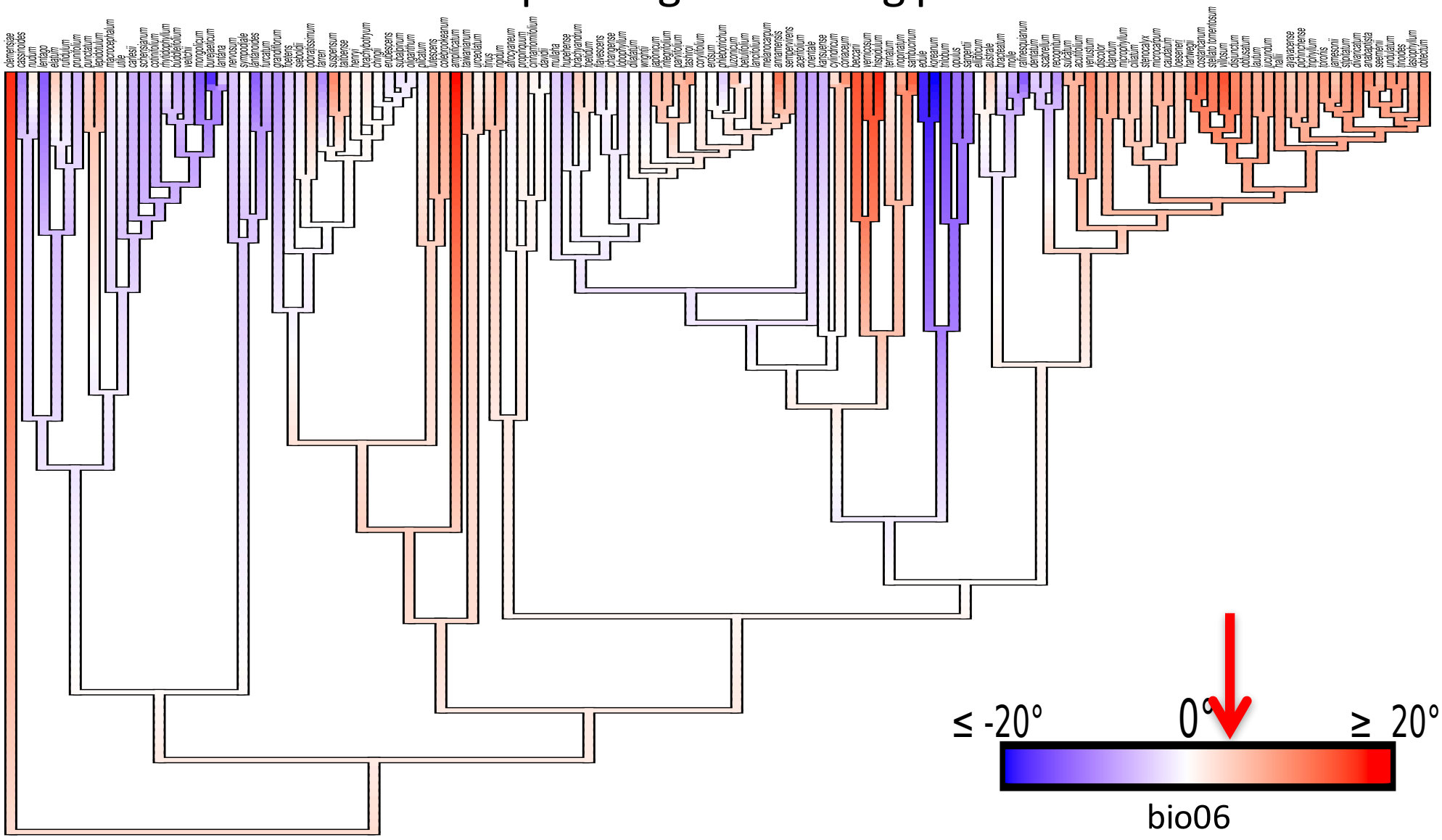
V. costaricanum



V. seemenii



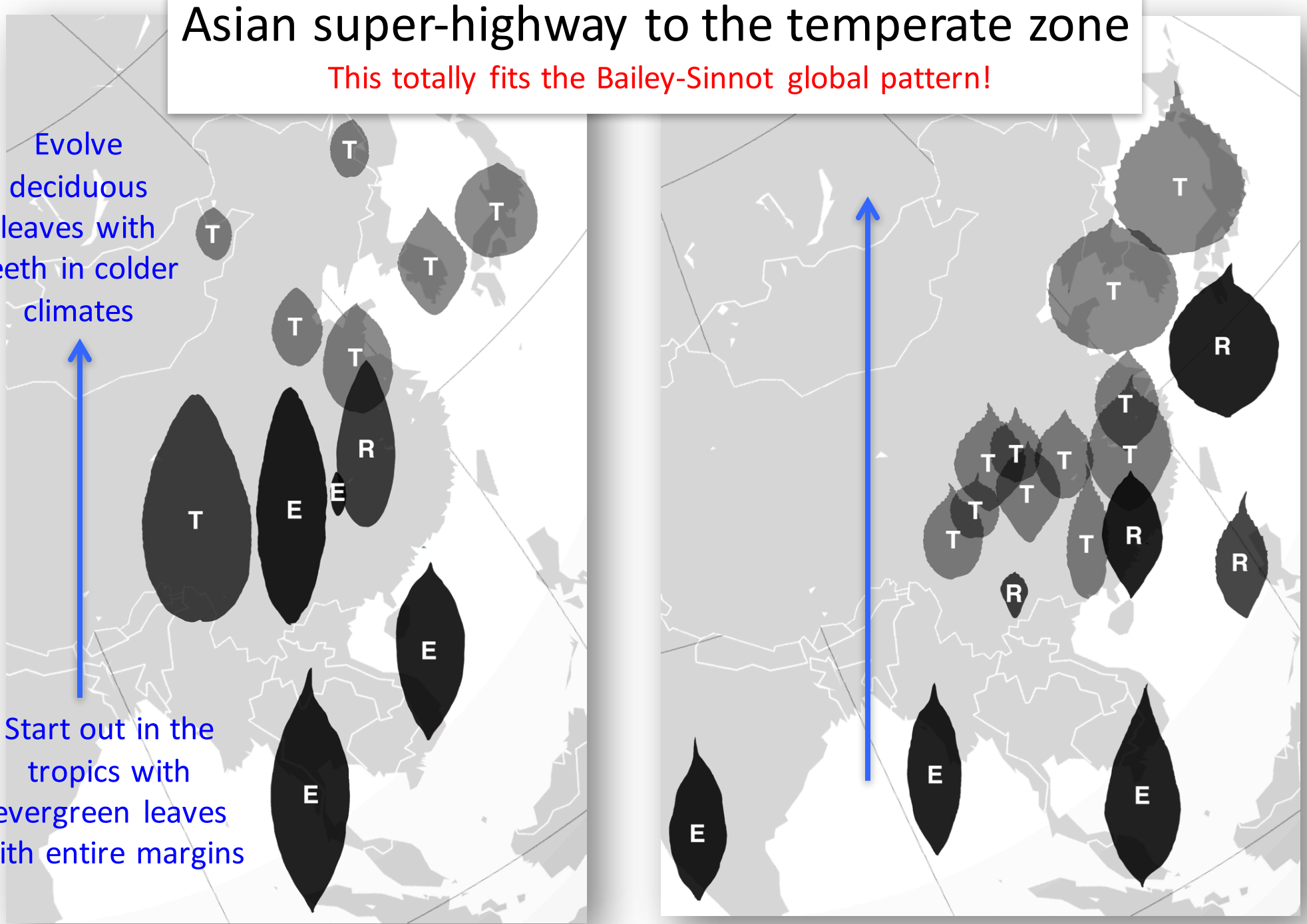
As many as 10 clades adapted independently to colder forests, with a prolonged freezing period



Viburnum probably started out in warmer forests, without prolonged freezing

Asian super-highway to the temperate zone

This totally fits the Bailey-Sinnot global pattern!

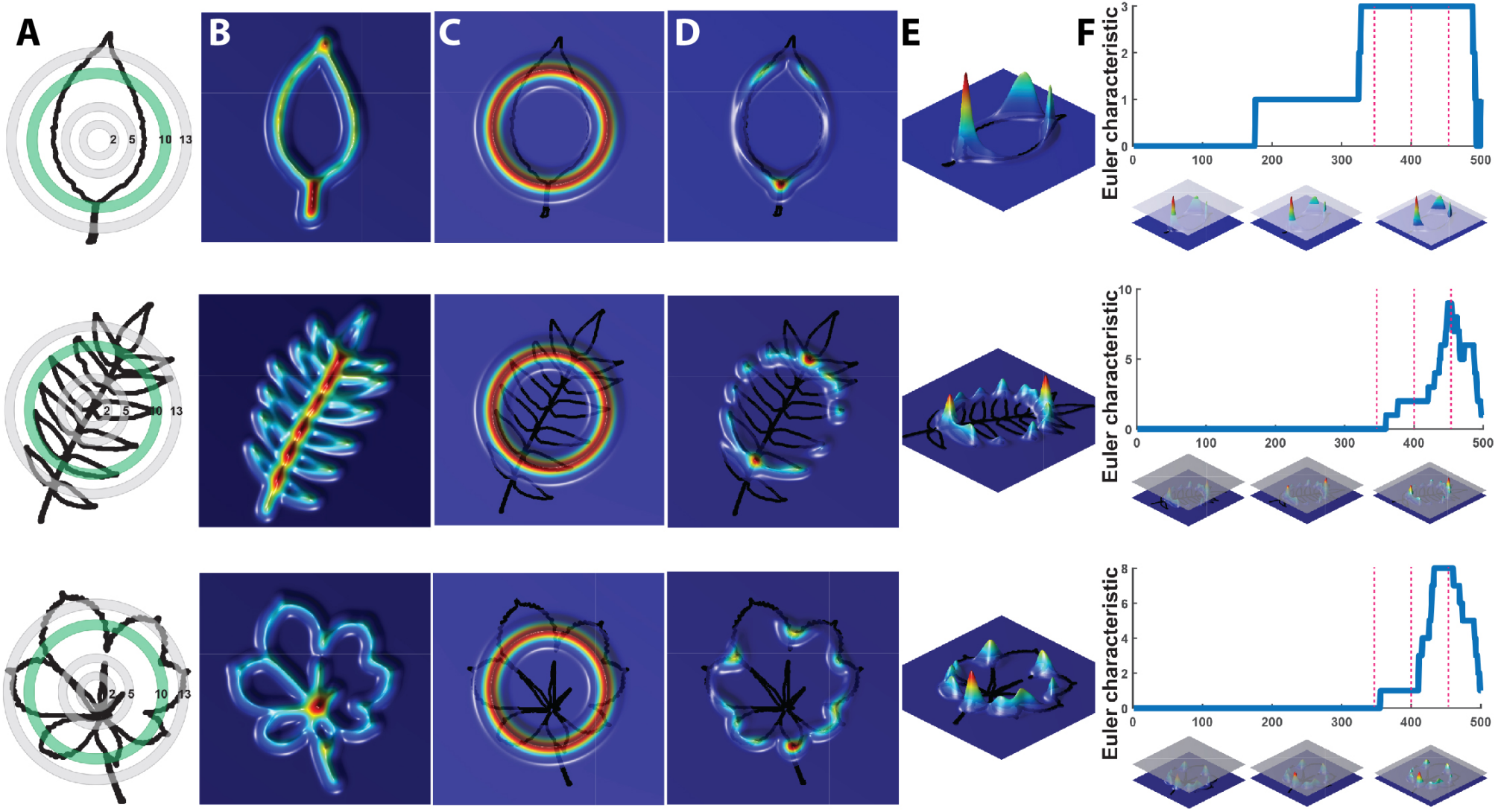


one *Viburnum* lineage

another *Viburnum* lineage

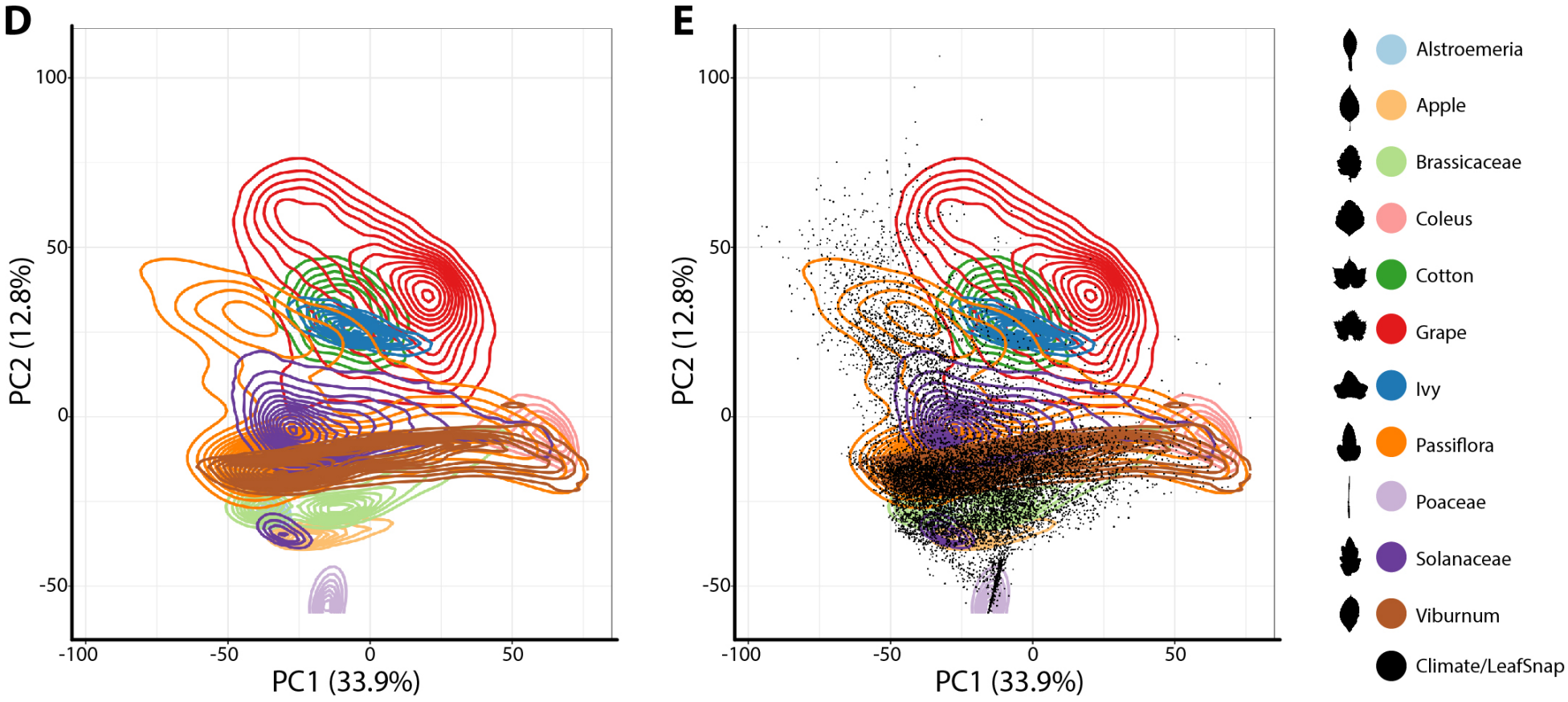
How about if we scale this whole procedure up, automate every step, and tackle this on a global scale? – sounds good, but leaves come in so many forms, and we don't have universal metrics!

But, people are working on this – e.g., here's a topological approach called "persistent homology"



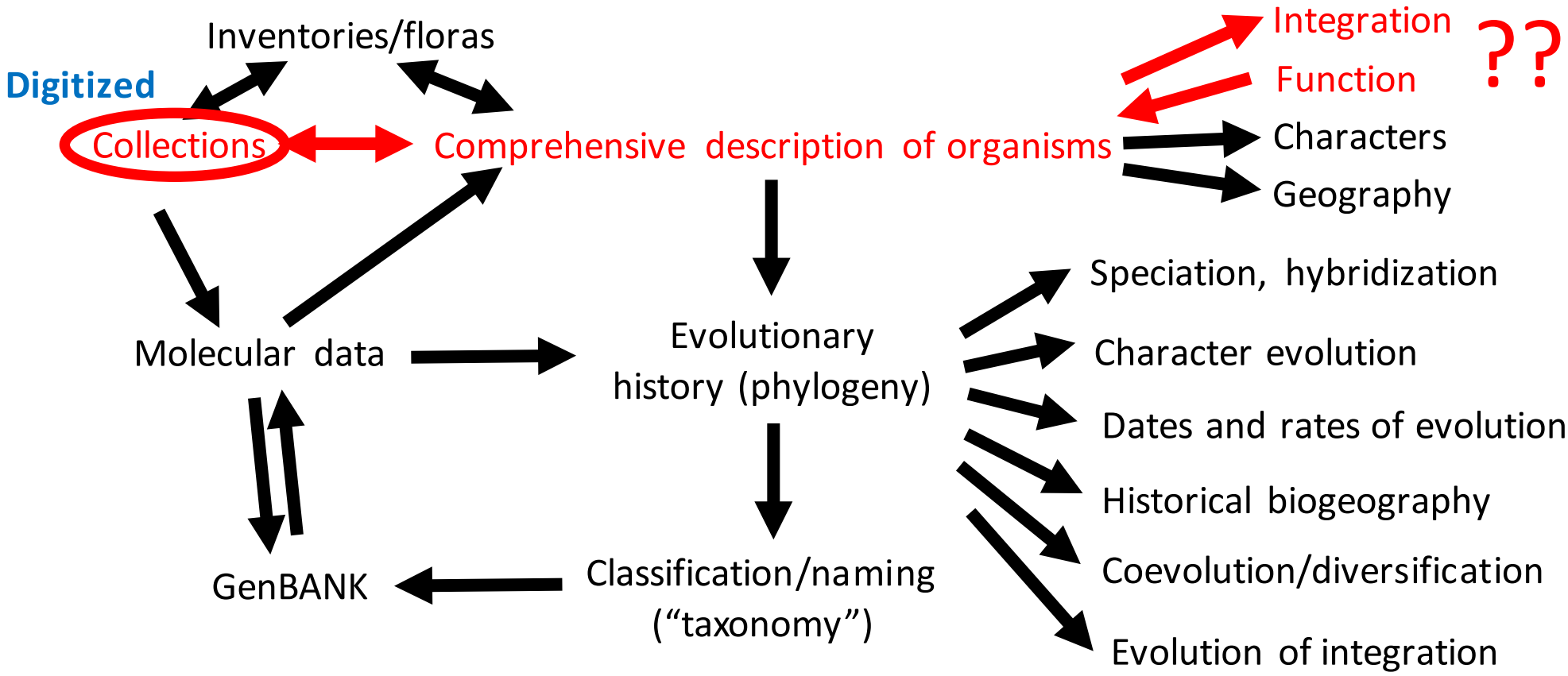
Mao Li, Dan Chitwood et al. A persistent homology leaf morphospace predicts plant family and region. MS in prep.

Using this approach we can display a wide variety of leaves in the same space – but we still have a long ways to go!!



Mao Li, Dan Chitwood et al. A persistent homology leaf morphospace predicts plant family and region. MS in prep.

"Functional traits"* and data on organismal integration are hard to extract from herbarium specimens, and virtually impossible using just standard images



But, there are some real possibilities with respect to "scaling law"!

* e.g., stomatal density, leaf mass per area (LMA), leaf palisade anatomy, wood density, vessel diameter, etc.

"Corner's rules" connect leaves to stems and to branching patterns

E. J. H. Corner, 1906-1996
University of Cambridge

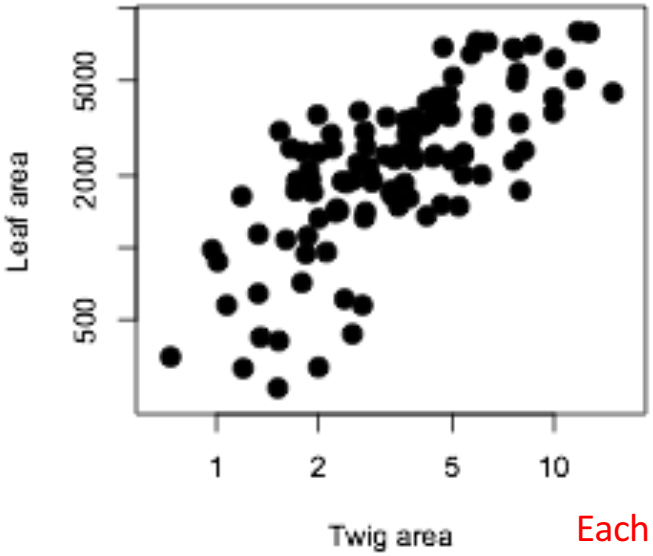
- Positive correlations between twig, leaf, and inflorescence size
- Negative correlation of leaf size with branching density
- These expectations relate to the efficient filling of spaces for light interception – e.g., avoiding self shading

"Corner's rules" at a glance in *Viburnum*

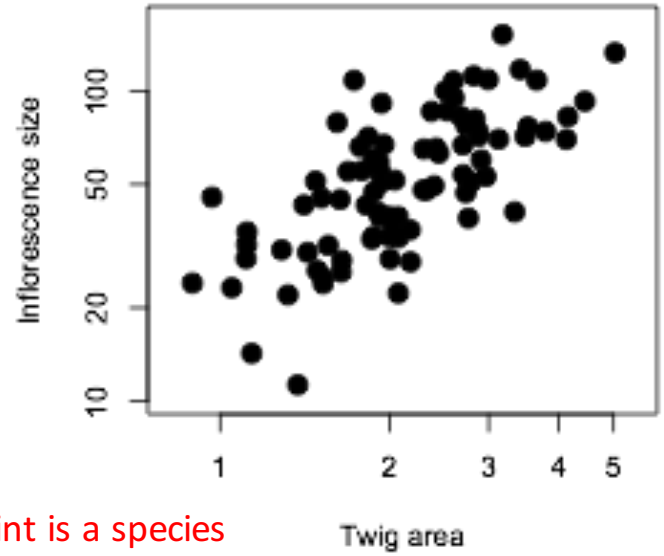
"Corner's rules" in *Viburnum* from images of herbarium specimens

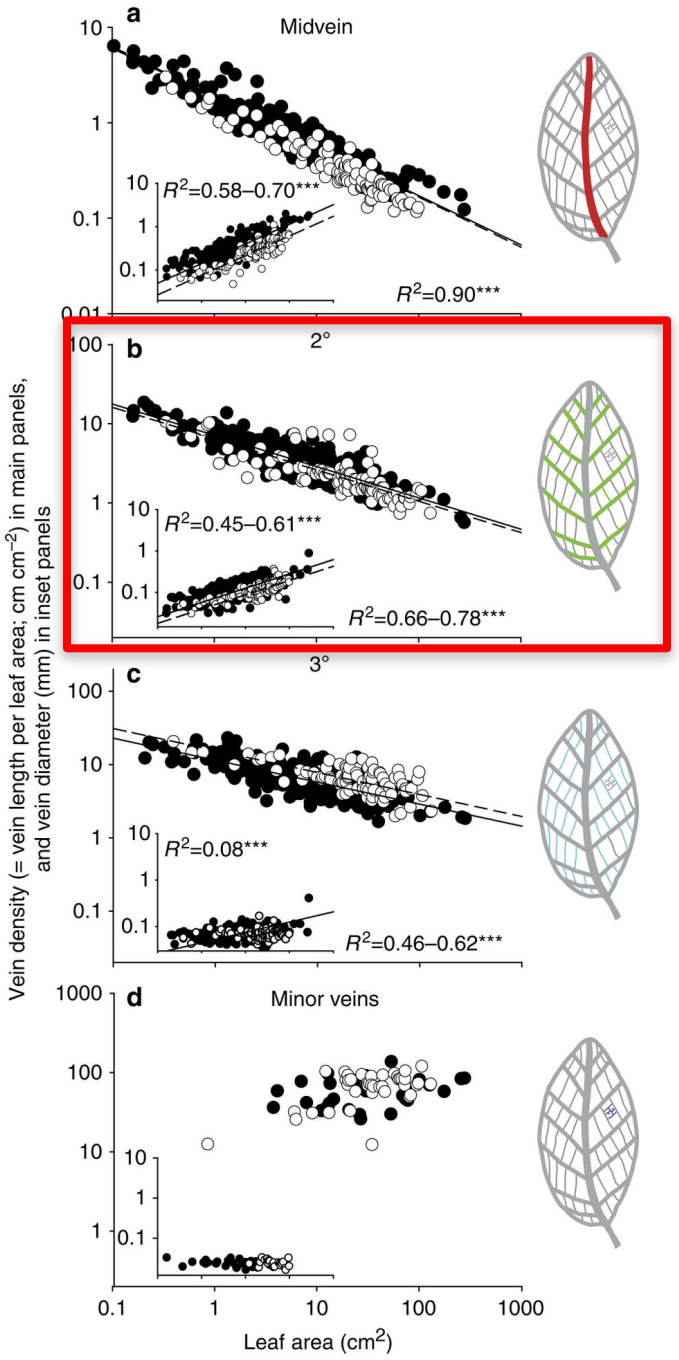
Matt Ogburn

Ogburn, Edwards, and Donoghue, in prep.

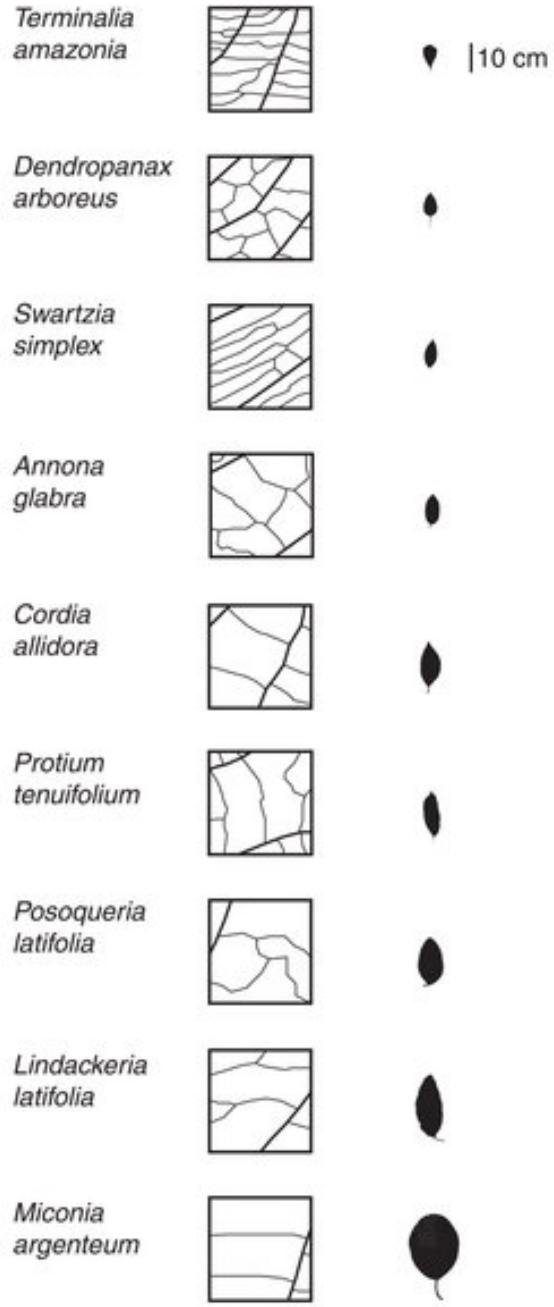


Each point is a species

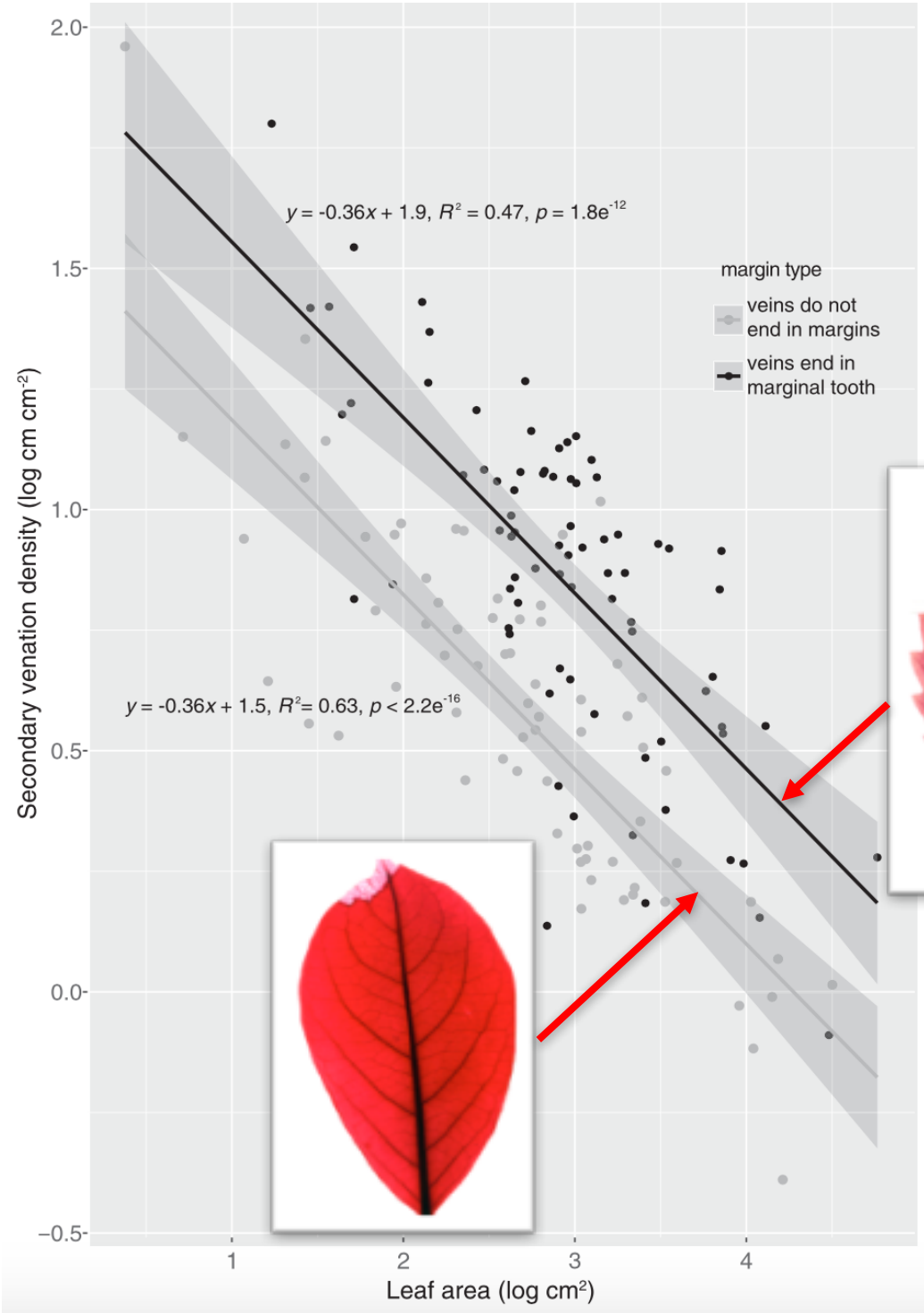




Another scaling law: in general, vein density is higher in small leaves, and lower in large leaves!



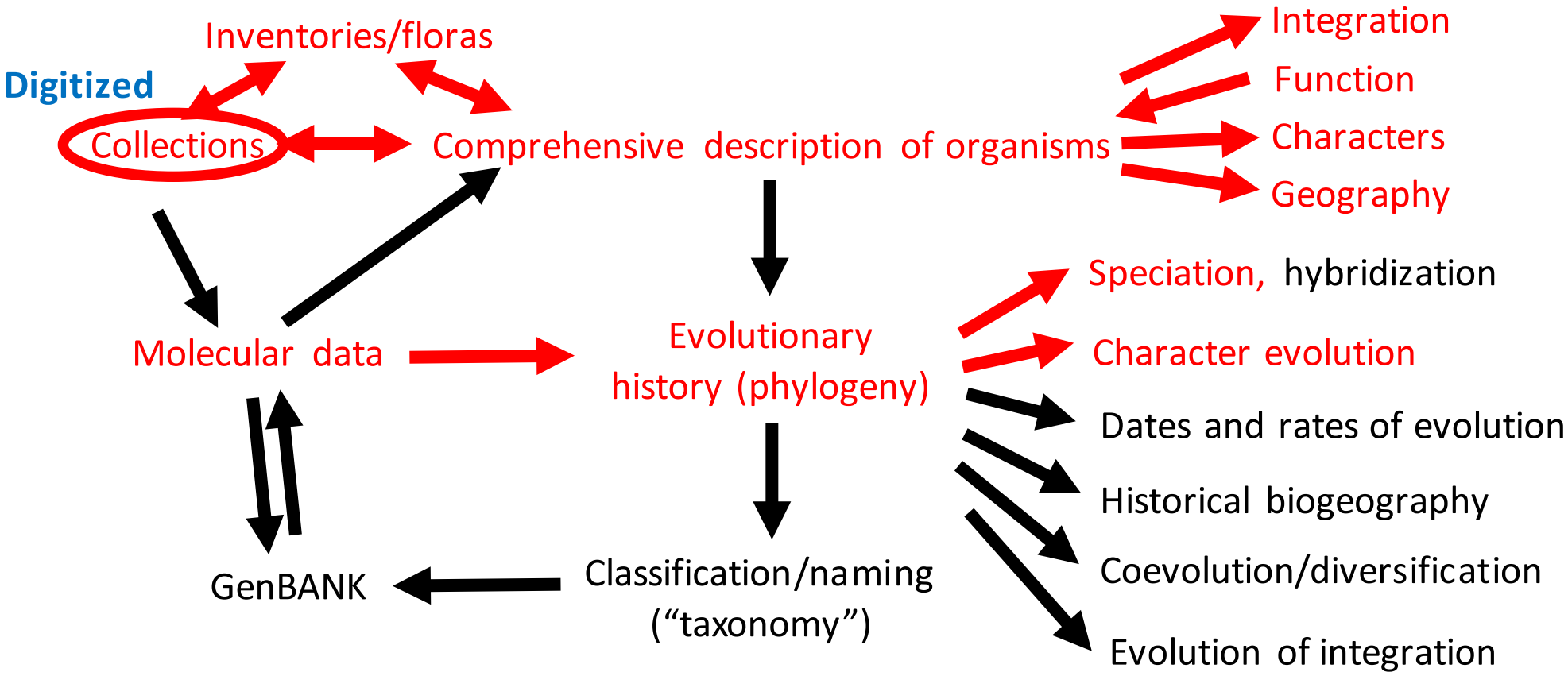
Sack et al., 2012, Developmentally based scaling of leaf venation architecture explains global ecological patterns. Nature Communications



Now, using *Viburnum* specimen images, we've discovered that for, leaves of a given size, secondary vein density is significant higher in leaves with marginal teeth!

Edwards et al. 2017. Amer. Jour. Bot.

I've given examples of some of the research uses of digitized specimens, but there are others that could cover ALL of the research areas below!



We need such examples to stimulate further use!!!!!!

Thank you, and the NSF !