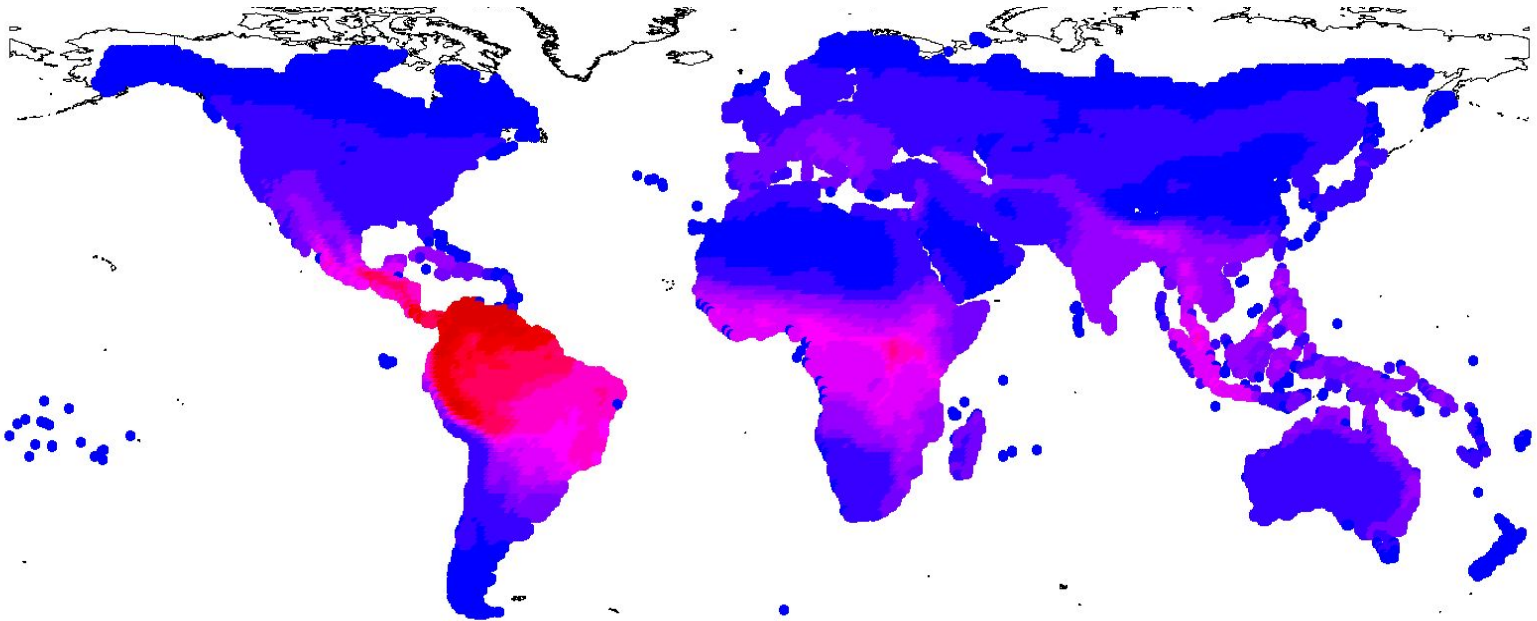


Digital Databases and the Multiplicity of Biodiversity Views



Terrestrial mammals numbers
Soberon & Ceballos, 2009

Jorge Soberon
Department of Ecology and Evolutionary Biology and Museum of Natural
History
University of Kansas

First, what is *biodiversity*?

- Simply put, the total manifestation of life on the planet. This is the “political” definition.
- The devil is in the details
- Biodiversity is not a simple concept, like, say temperature, that can be measured with thermometers.



AMNH, cdn.c.photoshelter.com

What is biodiversity? II

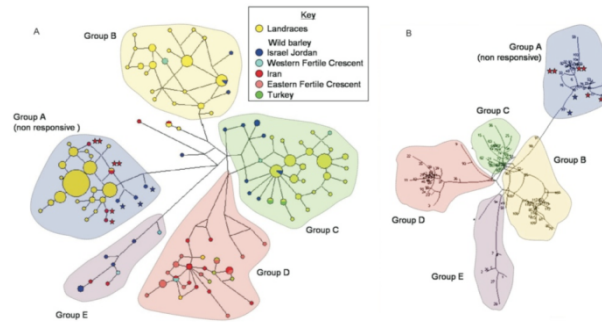
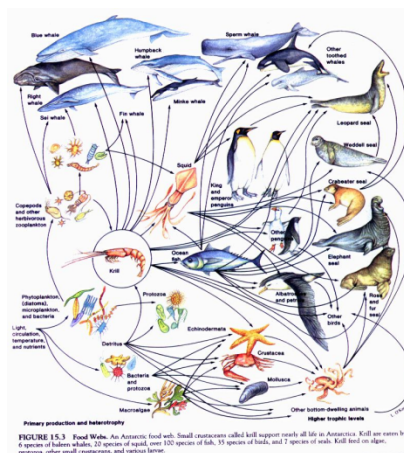
- Biodiversity is a complex and badly defined concept.
- It makes reference to other complex and poorly defined concepts
- And, alas, there are no *biodiversitometers*



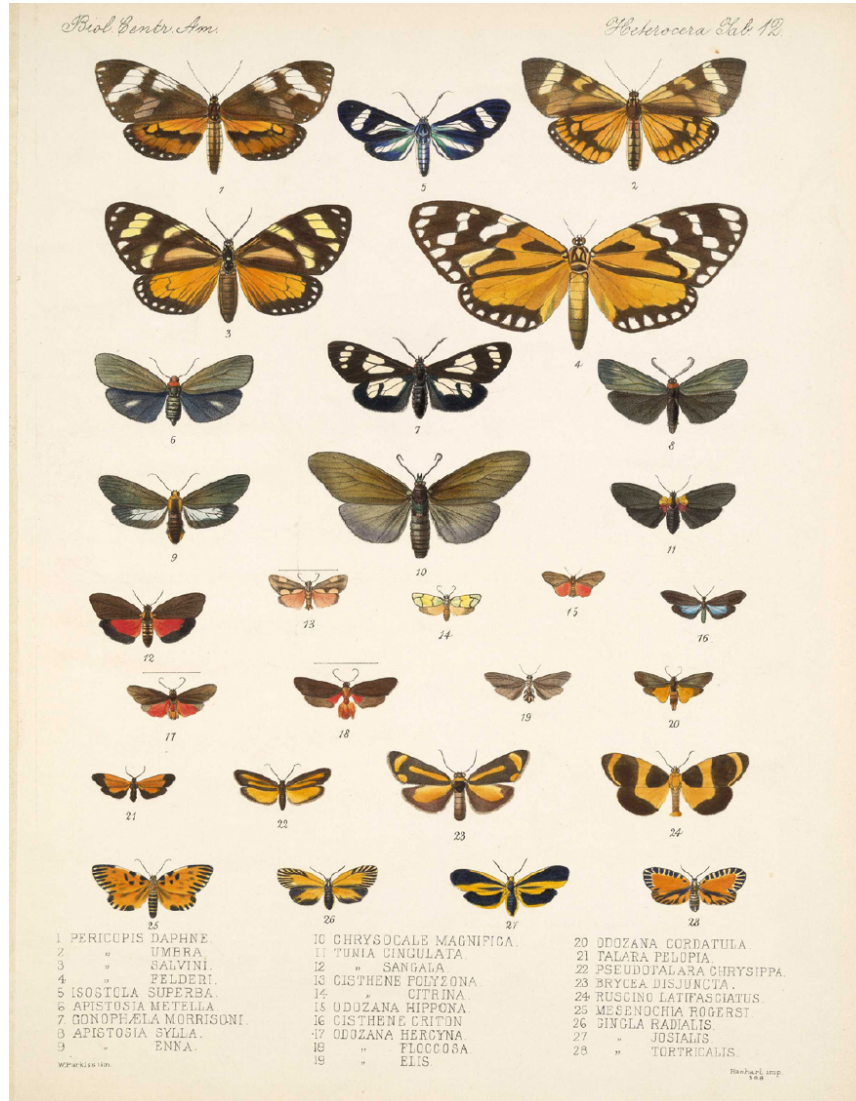
Somebody please invent
Mr. Spock's *Tricorder*

Surrogates of Biodiversity

- Morpho
- Eco
- Phylo
- Geno
- Proteo
- Ethno



Taxo View



Pericopidae.
Biologia Centrali Americana

MorphoView

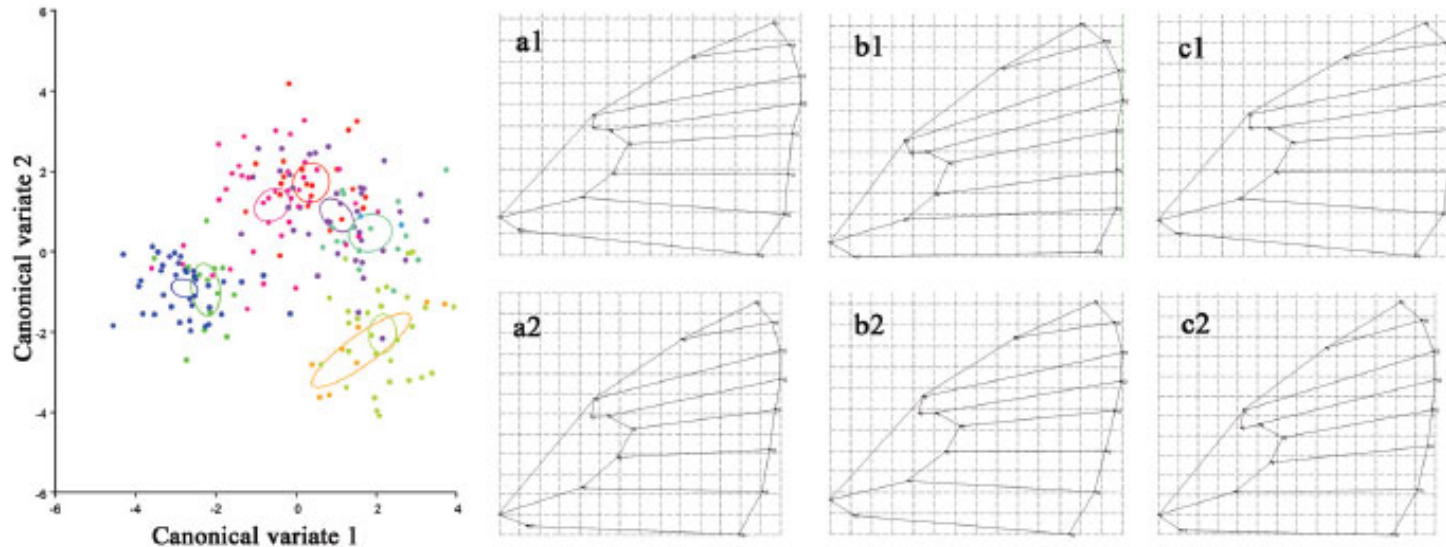
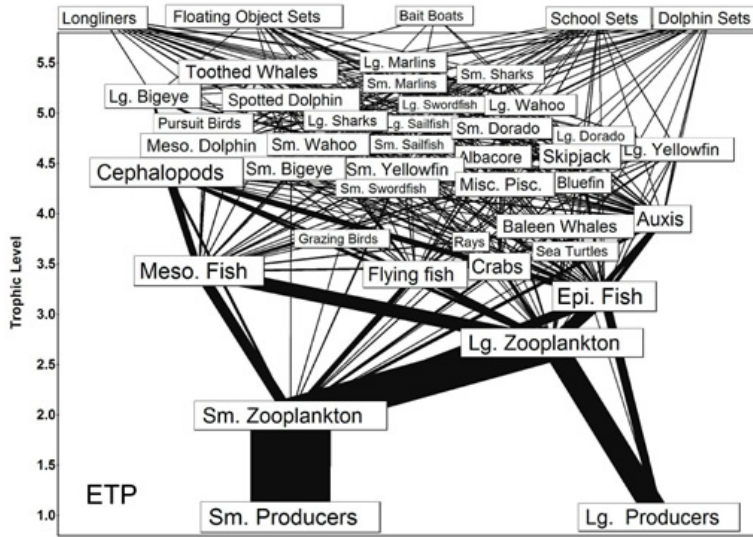
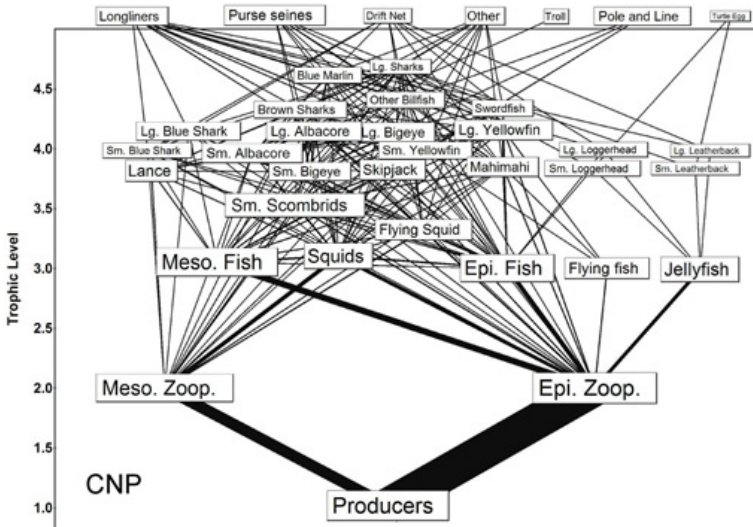


Fig. 3. Canonical variate analysis of the differentiation among sampling sites for *Auca coctel*. Color represents a shape variable which groups all individuals from the same site. *Dark red* points: La Serena (a); *dark green*: Collipulli; *dark blue*: Tumbes (c); *lilac*: Valle Elqui; *light green*: Dichato; *orange*: Concepción; *red*: Valparaíso; *calypso*: Los Lagos (b); *fluorescent green*: Loncoche. a1, b1, c1: female; a2, b2, c2: male.

Eco View

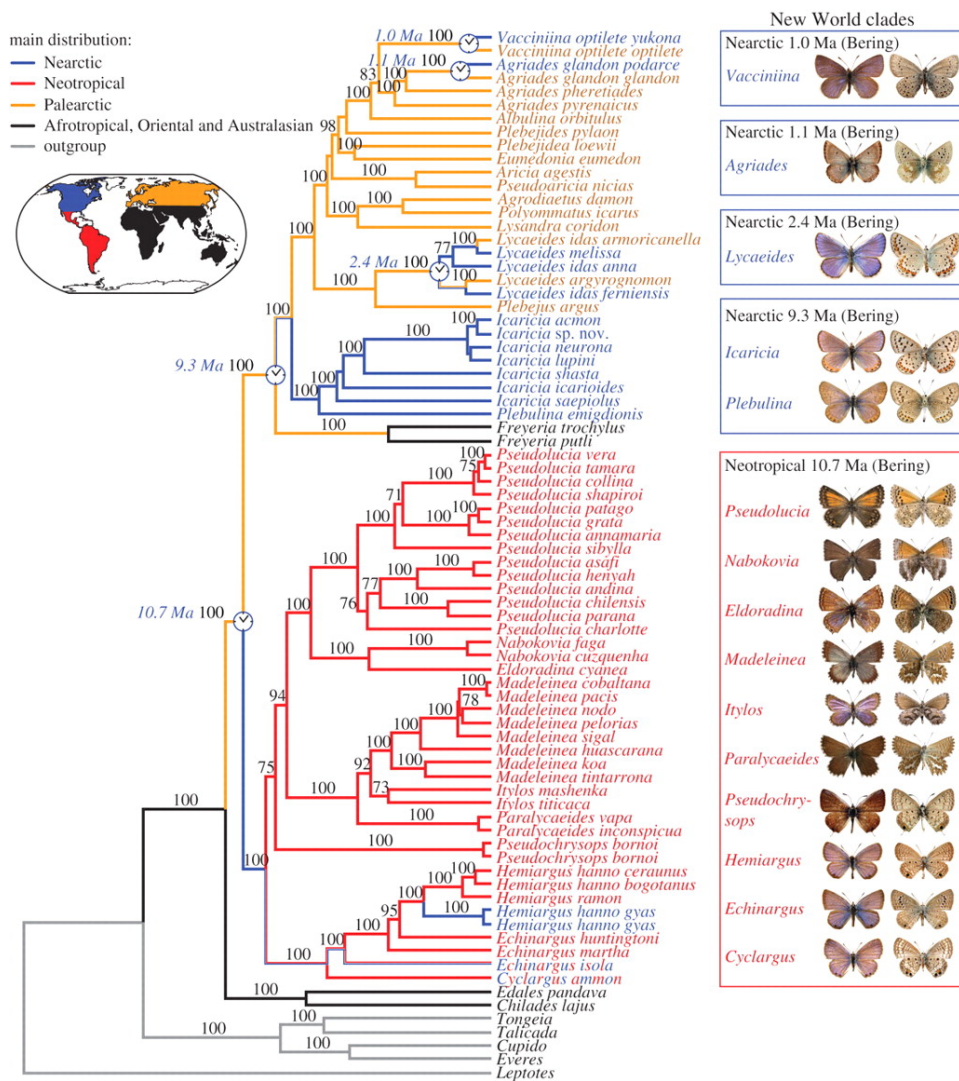


Full food webs of the Eastern Tropical Pacific (ETP)

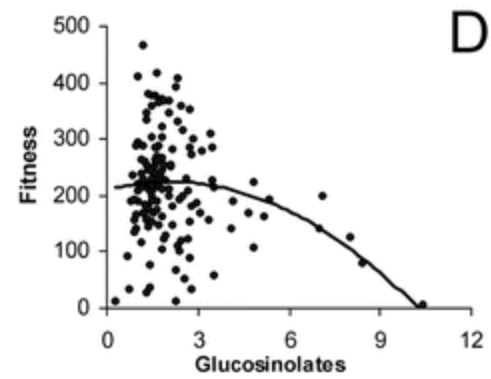
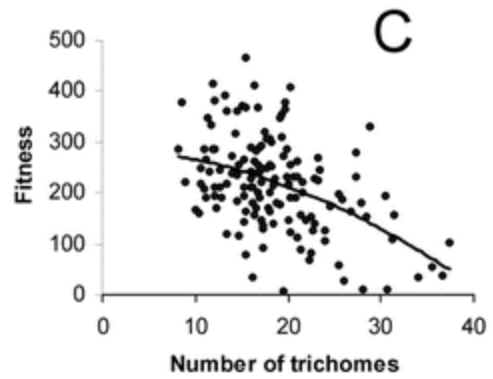
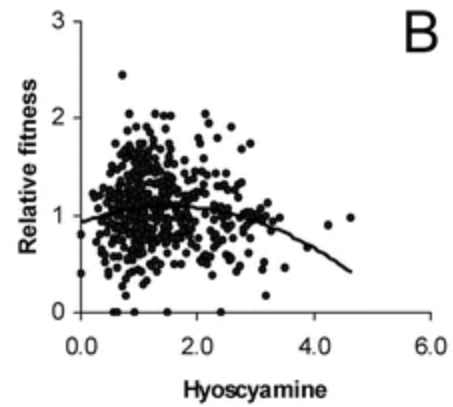
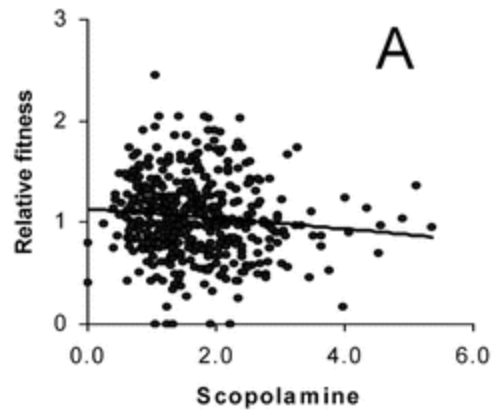


And the Central North Pacific (CNP)

Phylo View



ChemoView



Ethno View

Datura stramonium



Libelus , de la Cruz, 1552



Materia Medica Dioscorides, 512

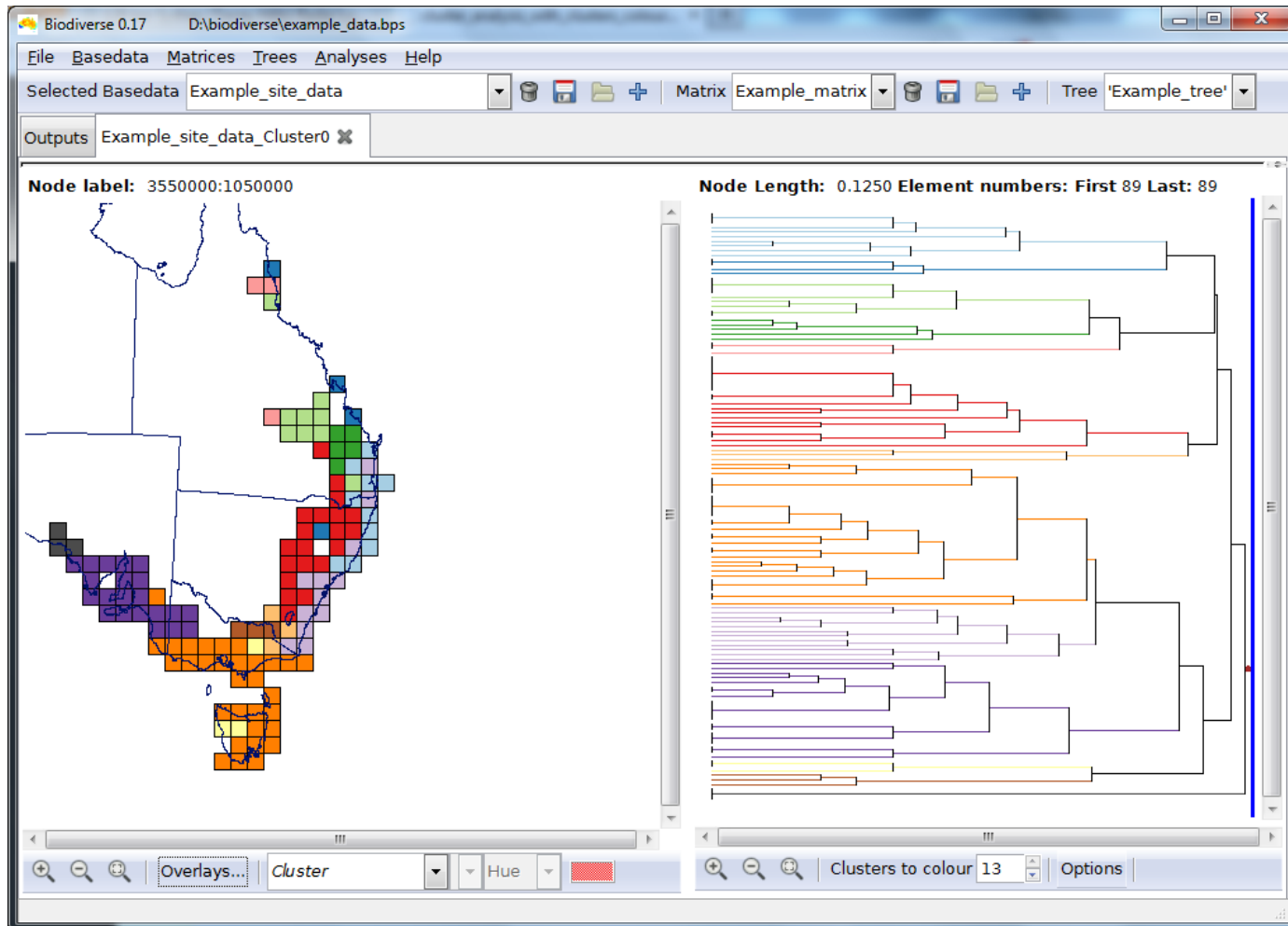


Quatro Libros,
Hernandez, 1652



Mexican specimen, Kew, 1874

Or, preferably, integrated views



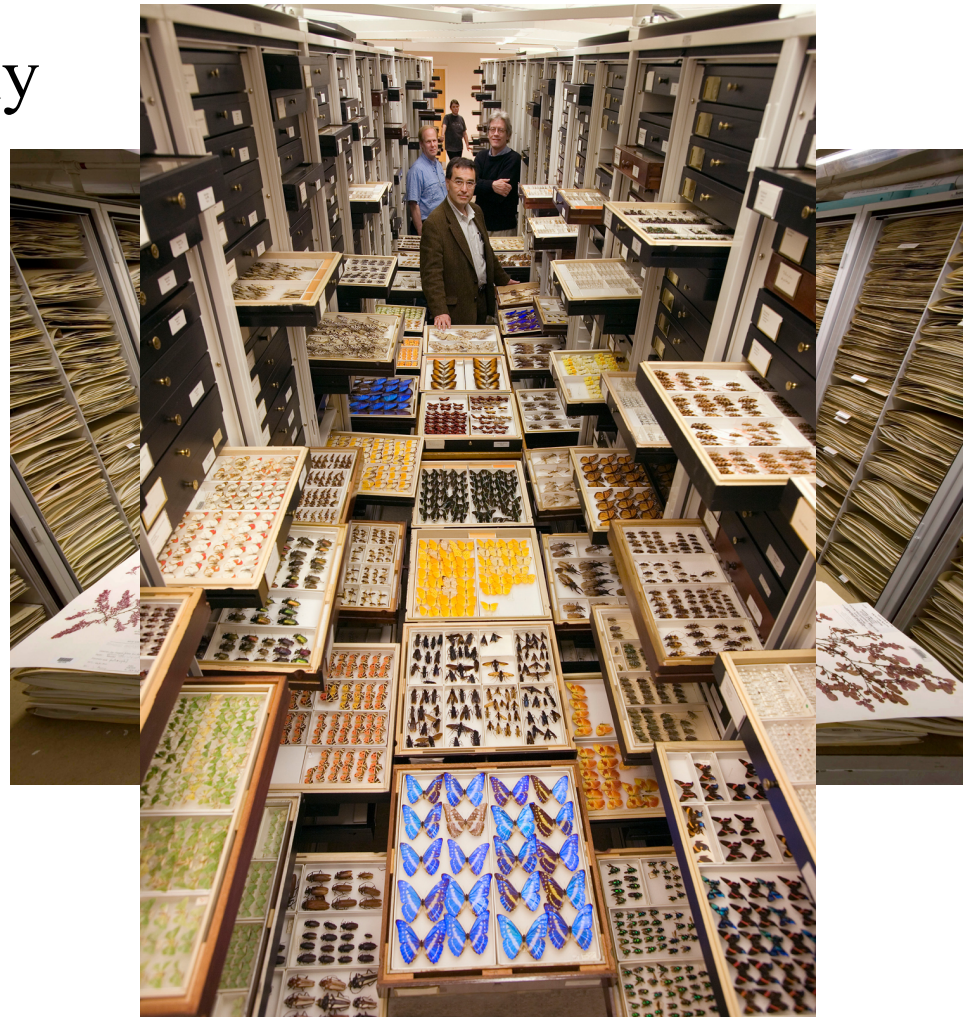
Biodiverse, by Shawn Laffan. <http://code.google.com/p/biodiverse/>

Integrating knowledge domains

- This should be the future. Integrating different points of view.
- What is *integrating*?
 - Simplest: overlaying
 - More difficult: statistical relationships
 - Truly difficult: theoretical
- Whatever the meaning, to integrate one needs *Digitally Available Knowledge* (DAK, Sousa-Baena et al. 2013 Diversity & Distributions 20:369-381)

So, what data is available?

- At a global level, mostly **occurrence data**
- The data about an observation, or a specimen of:
 - What (name)
 - Where (lat, long)
 - When (date)
 - Says who (collector)
- Darwin Core, ABCD...

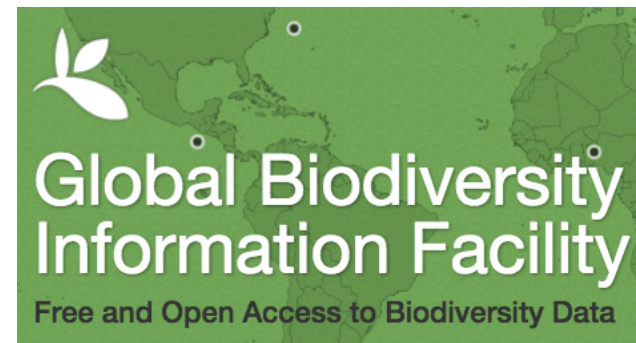
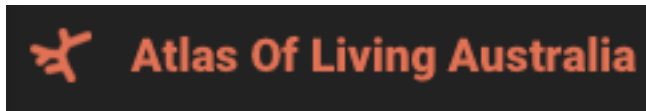


Images from the Smithsonian NMNH

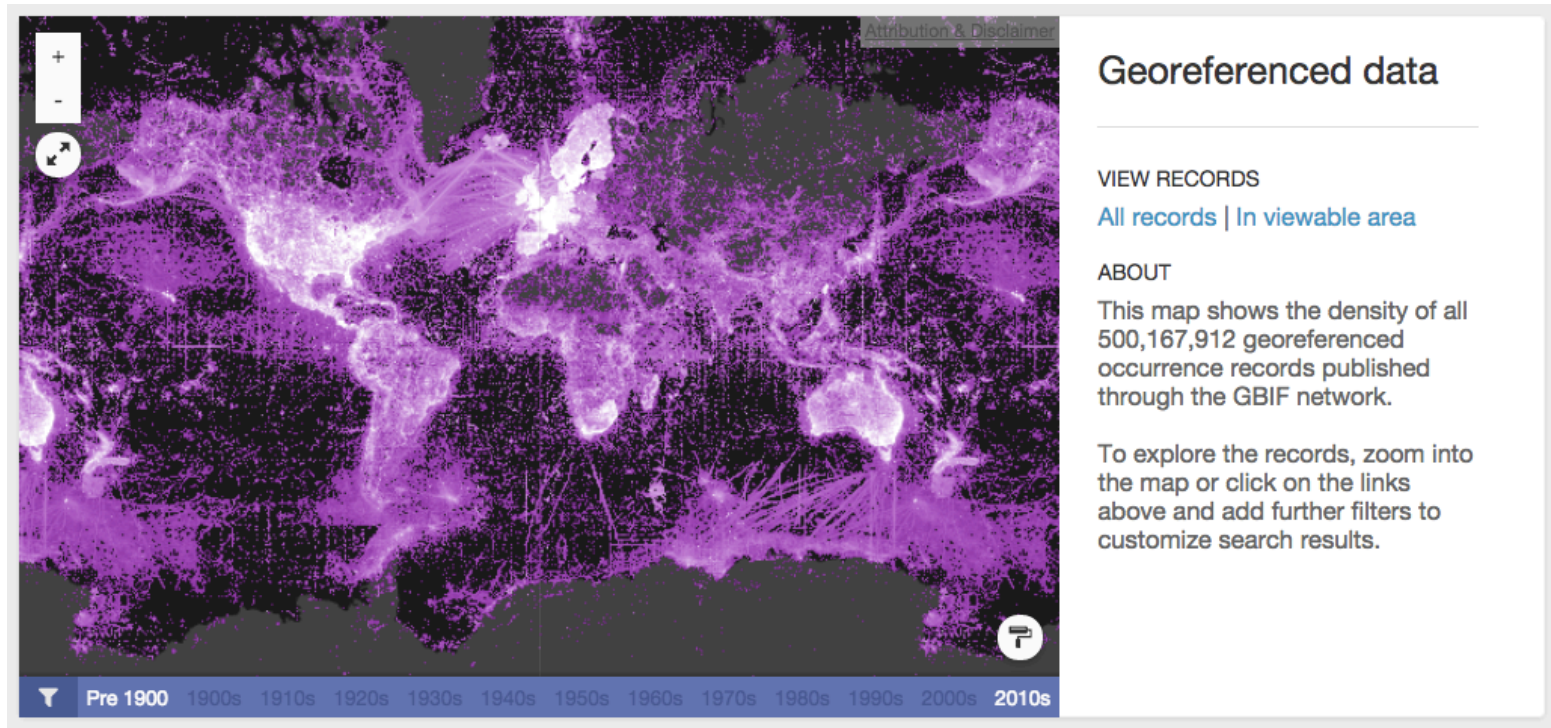


Instituto de Investigación de Recursos Biológicos Alexander von Humboldt

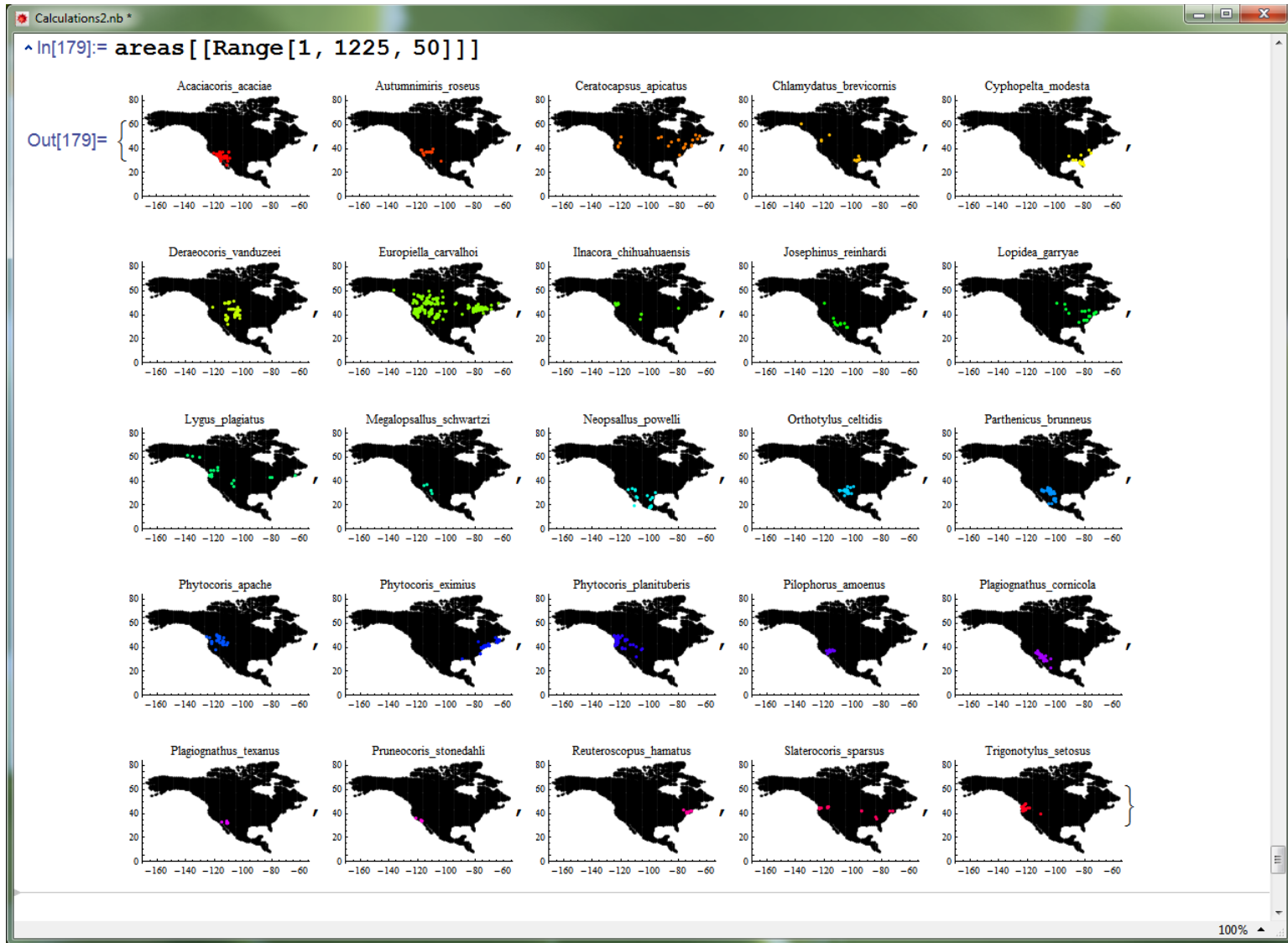
Investigación en biodiversidad y servicios ecosistémicos para la toma de decisiones

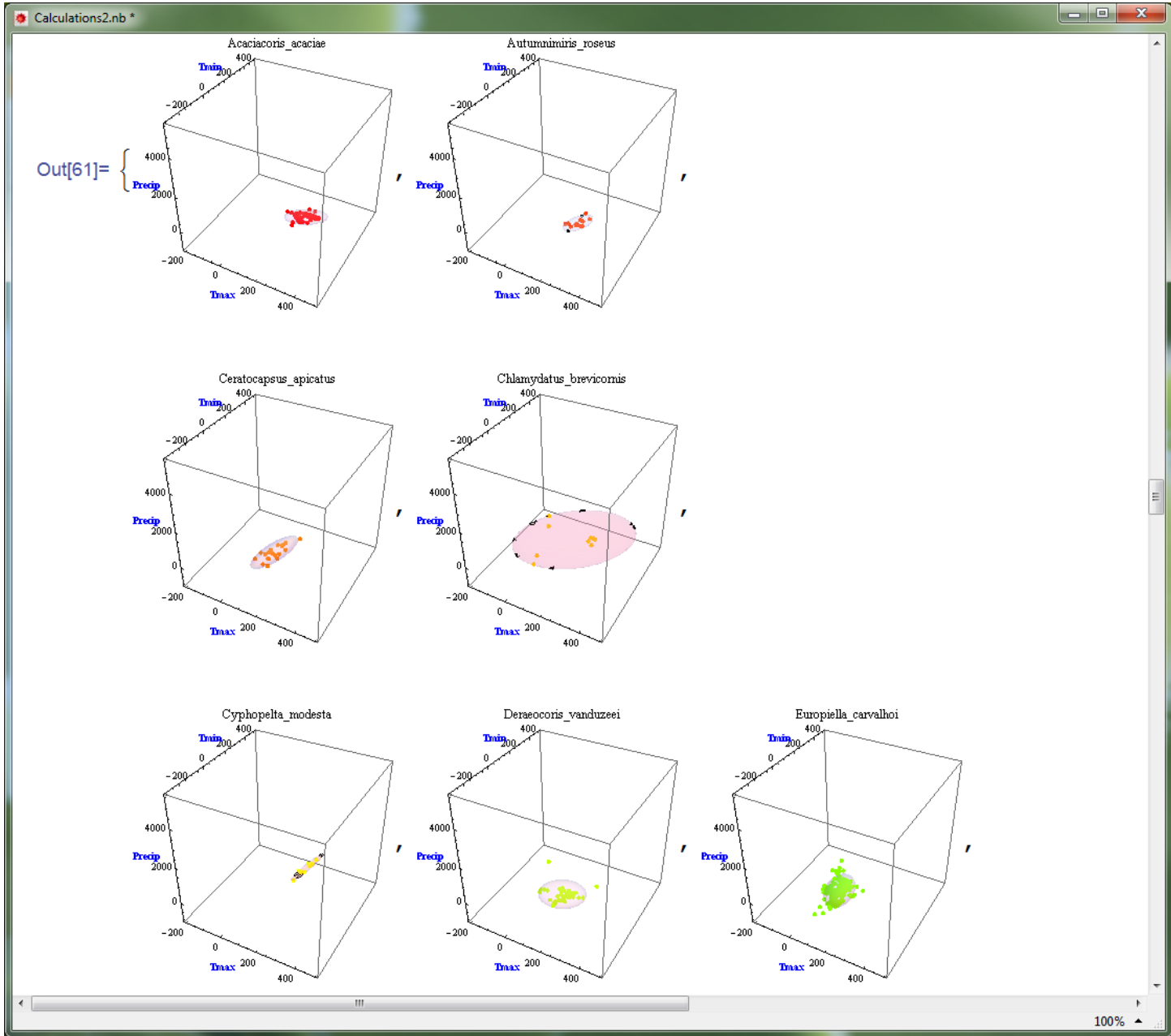


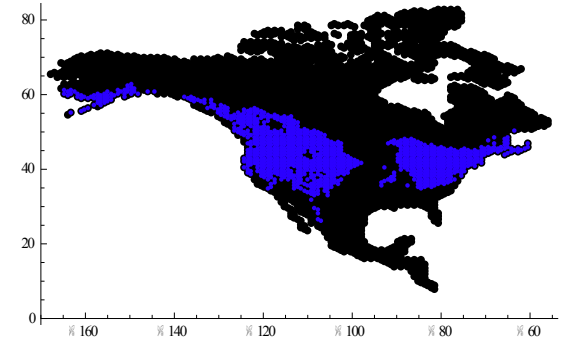
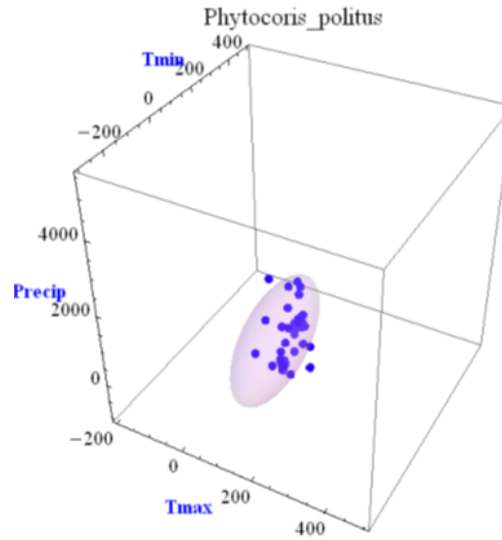
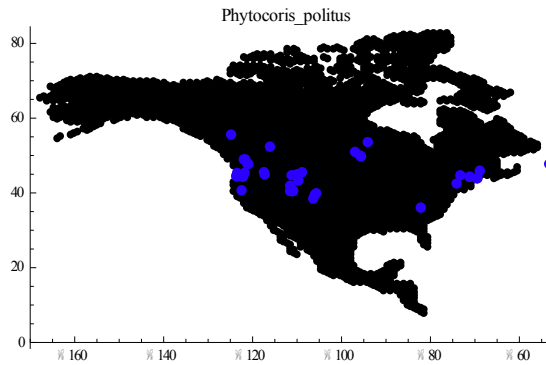
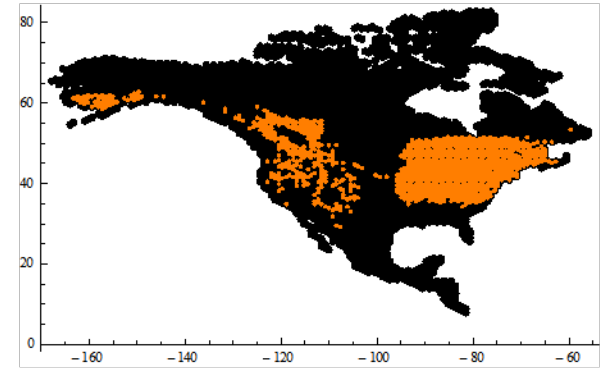
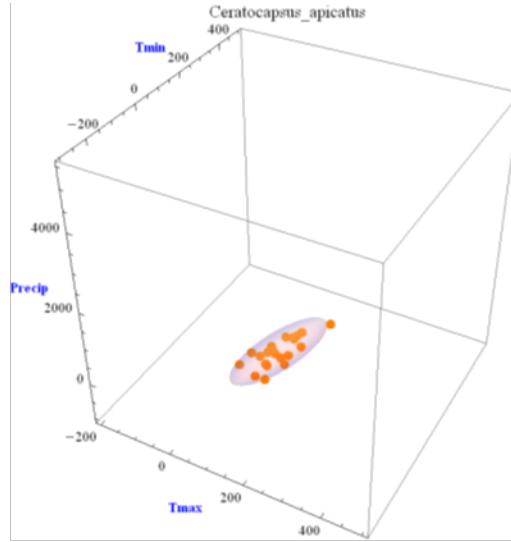
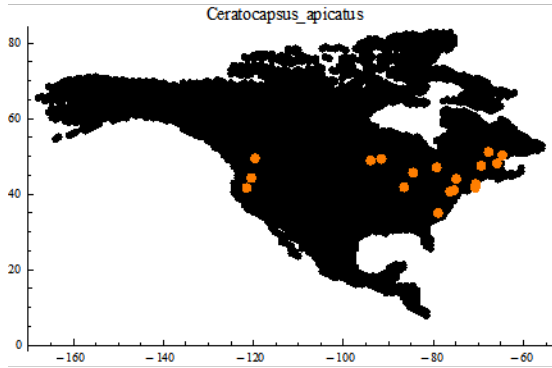
Presence only data is abundant and useful (although not for children)



The Miridae (data from Weirauch et al.)



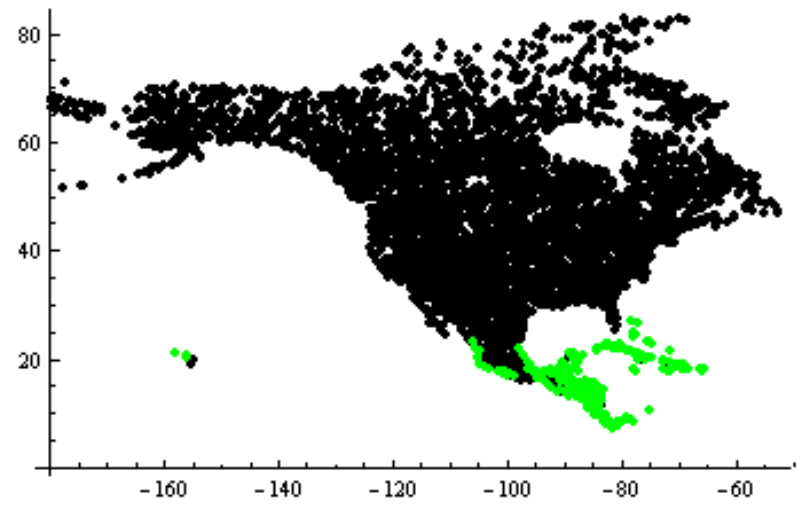
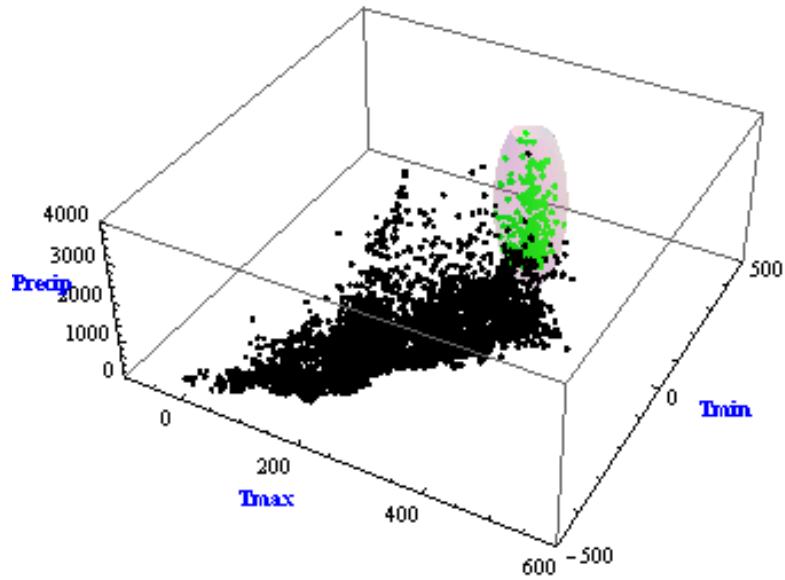


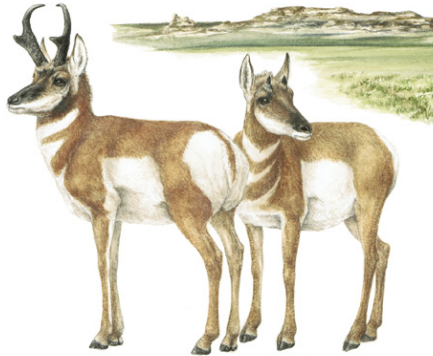




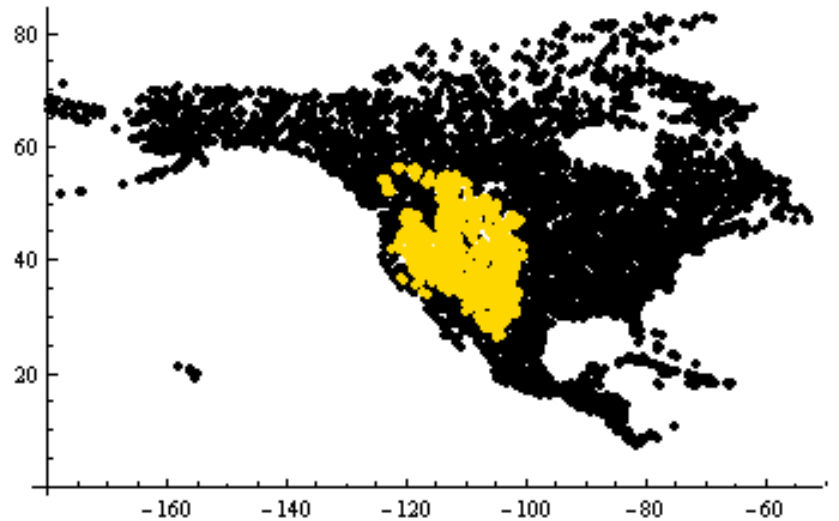
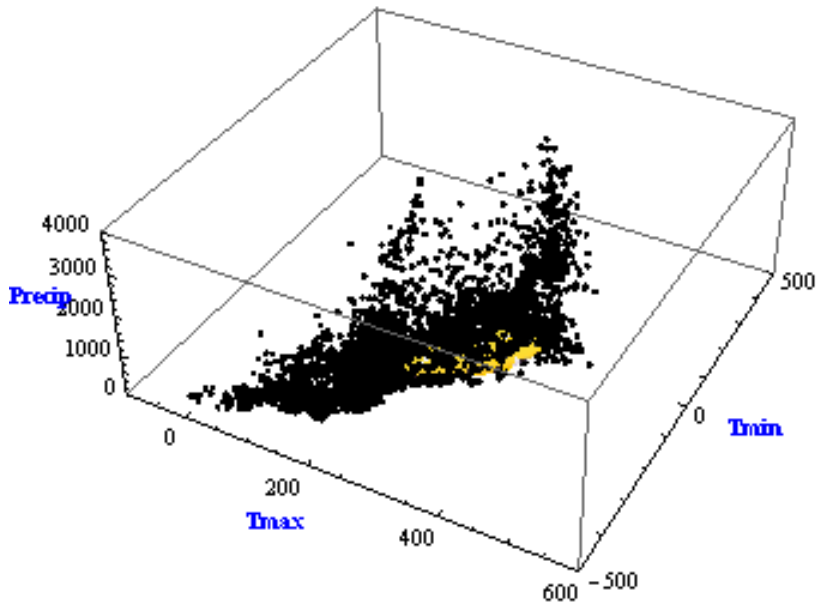
Alouatta palliata
INBio, Costa Rica

Fuente: INBio
© Derechos reservados



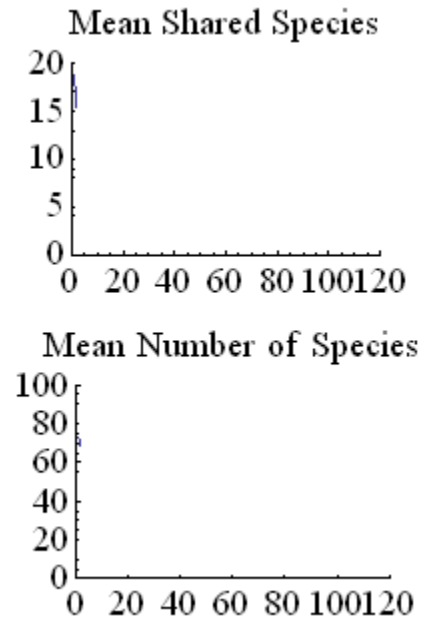
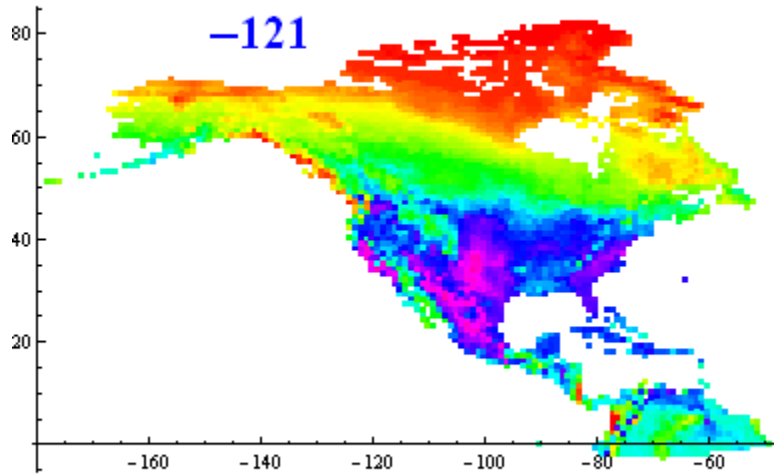


Antilocapra americana
Smithsonian NMNH



Mammal potential numbers since Interglacial (120,000 BP)

Purple ~ 80
Blue ~ 70
Green ~ 60
Yellow ~ 50
Red > 40



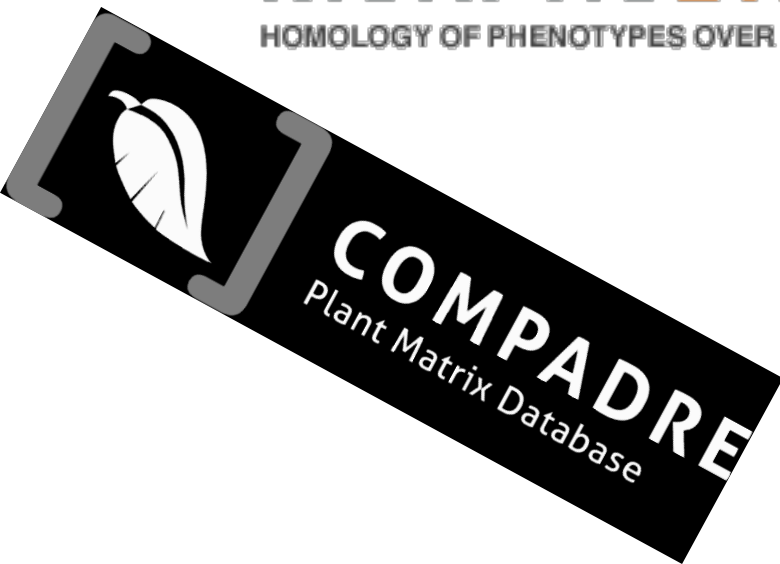
By *integrating*, the above can be...

- Dynamic maps of body size, ethnobotanic knowledge, secondary chemistry, interactions...
- Interacting, maybe dynamic perspectives on different knowledge domains
- The name is a link to attributes
- And the coordinates are a link to where
- These two are powerful links

There are already quite a few things to
integrate with



MORPHOBANK
HOMOLOGY OF PHENOTYPES OVER THE WEB



However...

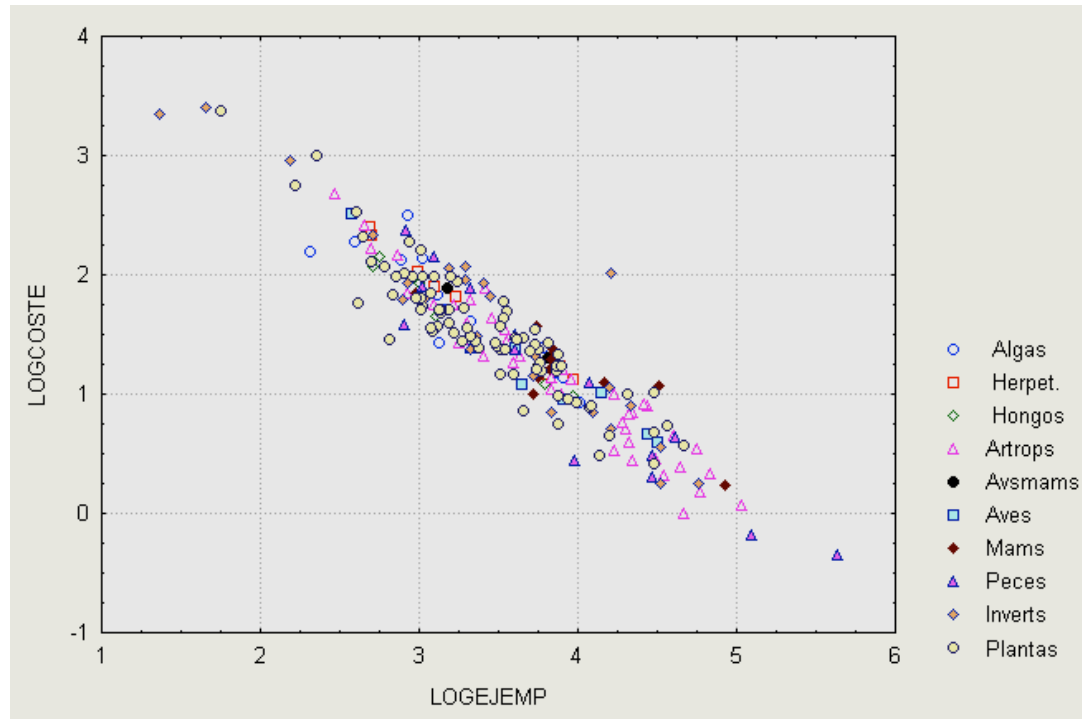
- Digitizing is expensive

Table I. Targeted Number of Objects to Be Digitized, Currently Realized Number of Digitized Objects, and Average Cost per Object per Digistreet as of April 1, 2014

Digistreet	Target	Realized	Average Cost per Object in Euros (including overhead)
Herbarium sheets—Leiden	3,800,000	450,000 (3,500,000 scanned)	1.29
Microscopic slides	900,000	350,000 (590,000 scanned)	1.57
2D objects (books, journals, etc.)	900,000*	600,000*	1.87
Entomology	850,000	600,000	1.51
Herbarium sheets—Wageningen	800,000	800,000	1.47
Mollusks	510,000	640,000	1.37
Vertebrates dry	325,000	180,000	2.37
Geological specimens	220,000	70,000	1.90
Wood samples	125,000	125,000	1.27
Specimens preserved on alcohol	100,000	70,000	4.65
Total	8,530,000	3,885,000	

*Number of pages.

Which means that it pays to digitize first the large collections



Log 10 of cost per specimens of digitizing collections, as a function of their size.
Data from CONABIO, Mexico. Cost per specimen \$1-\$10 USD

And not only expensive in \$\$\$

- It requires time
- Personnel
- Space
- A long term commitment (databases are never finished)
- It will never be done unless the primary providers (collections) find uses.
- This means large-scale digitizing beyond labels will not take place unless it is in the direct interest of curators and systematists

But the benefits will be huge: Overlay

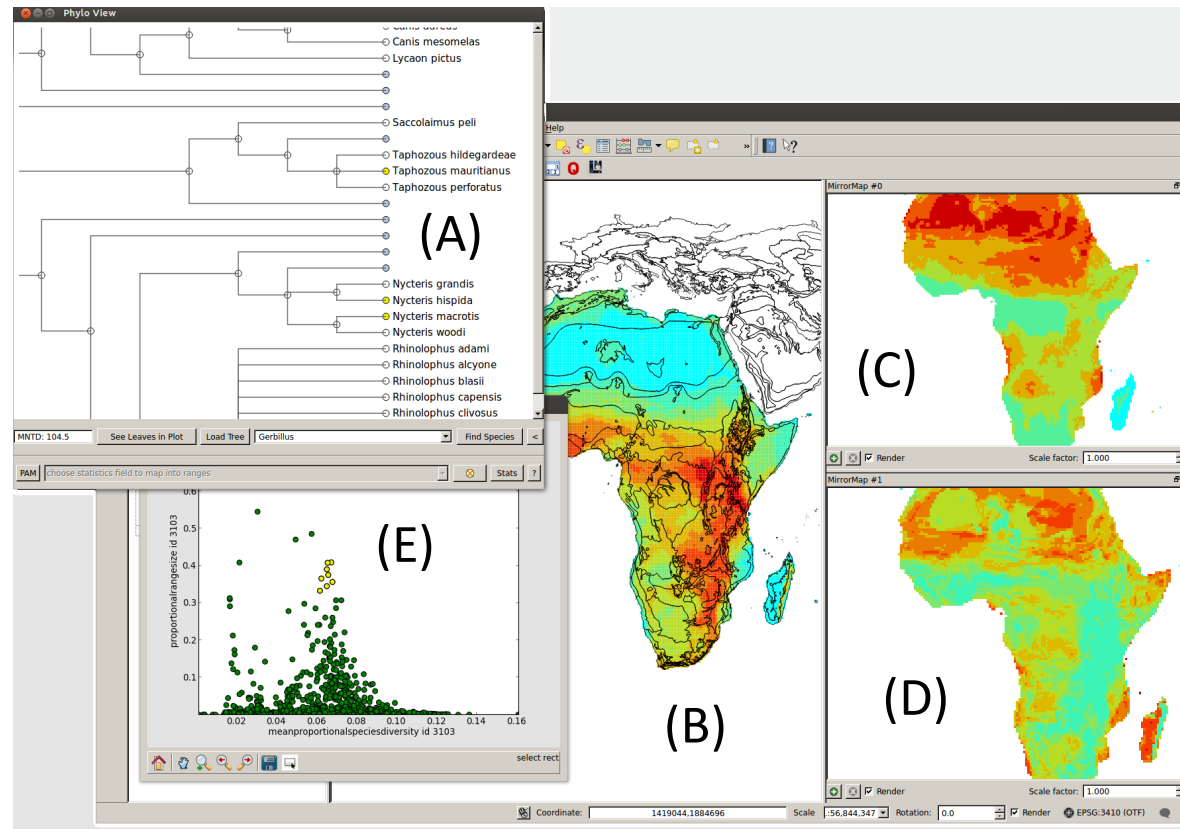


Figure 2. A species-based screen-capture of the QGIS plugin showing part of a mammal phylogeny (A) connected to a 800+ mammals of Africa PAM, with a map of species richness (B), mean proportional species-diversity (C), mean nearest taxon distance (D) and a scatterplot of range-size vs. mean phylogenetic distance (E) with the “brushed” species highlighted (yellow species in (A) and (E)).

Statistical

	Sites	Species' traits	Nodes
Species	I Incidence matrix	T Trait matrix	P Phylogeny matrix
Environmental variables	E Environment matrix	D '4 th corner' matrix	β_E Environment- phylogeny correlations
Biogeographic events	B Biogeography matrix		β_B Biogeography- phylogeny correlations

Figure 1 Schematic arrangement of data matrices **I**, **E**, **B**, **P** and **T** and the three association (**D**, β_E , β_B) matrices discussed in the text.

Theory

- A real challenge.
- Do we have a “theory of biodiversity”?
- Of course we have a Theory of Evolution, but this is too general.
- A theory of biodiversity may emerge from the integration of those different points of view.



Maybe this should be the future

- Creating DAK in different domains.
- Morpho
- Eco
- Phylo
- Creating the concepts and tools to integrate them
- Many doubt we can do it
- Will it be a case of “*ignoramus et ignorabimus*” (du Bois-Reymond)
- Or rather one of: “*Wir müssen wissen — wir werden wissen!*” (Hilbert)

Thanks to...

- Organizers of the symposium
- Jeff Cavner, A. Christen, H. Arita, P. Rodriguez, A. Lira, F. Villalobos, coworkers in the biodiversity perspectives stuff
- A. T. Peterson, for endless conversations on this.
- And the money folks

