PRAGMA VBE

Lifemapper

Lifemapper Evolves

Aimee Stewart

University of Kansas Biodiversity Institute

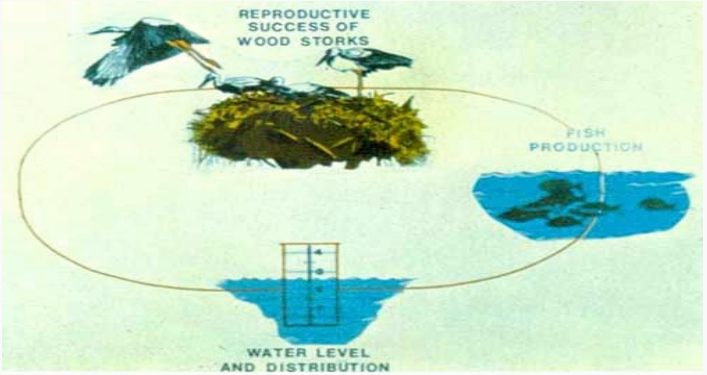
http://lifemapper.org

PRAGMA VBE

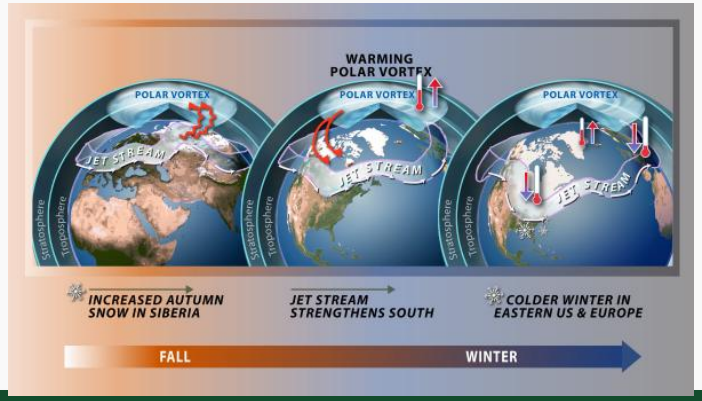
Lifemapper brings together …

Biodiversity Inventory

Macro-Ecological Modeling

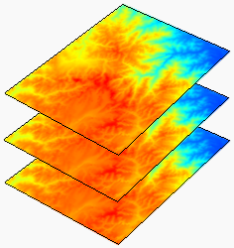


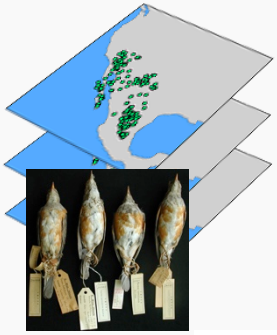
Global Climate Change



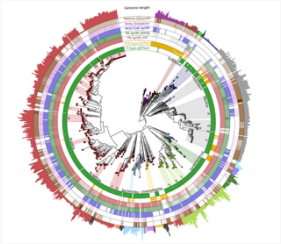
PRAGMA VBE

Lifemapper Analyses





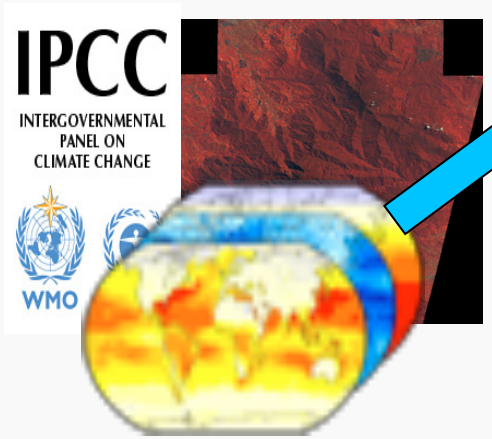
- Inputs
  - Environmental
  - Specimen occurrence points
  - Phylogenetic tree
- Tools
  - Species Distribution Modeling (SDM)
  - Macro-Ecological Analyses
  - Meta-Community Phylogenetic Analyses (MPCA)
- Results
  - Potential species habitat maps
  - Biodiversity / Phylogenetic Maps and Analyses



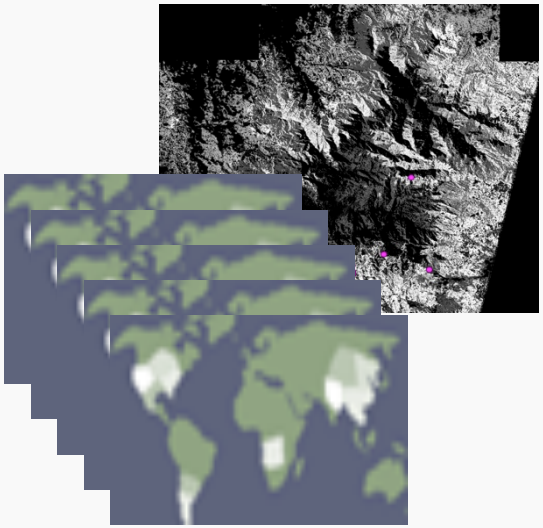
PRAGMA VBE

Species Distribution Modeling

Species Occurrence Data



Environmental Data



Potential Habitat

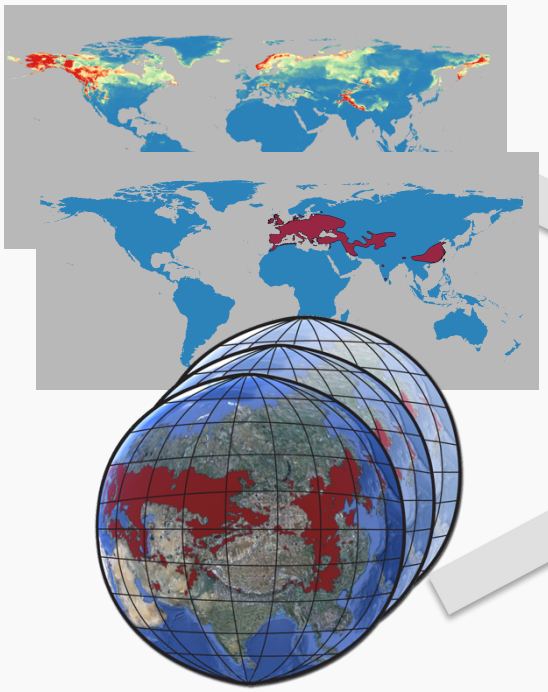
PRAGMA VBE

Lifemapper

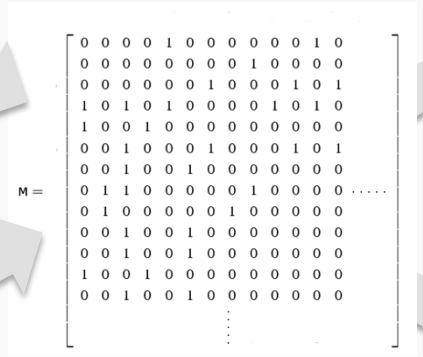
Demo

PRAGMA VBE

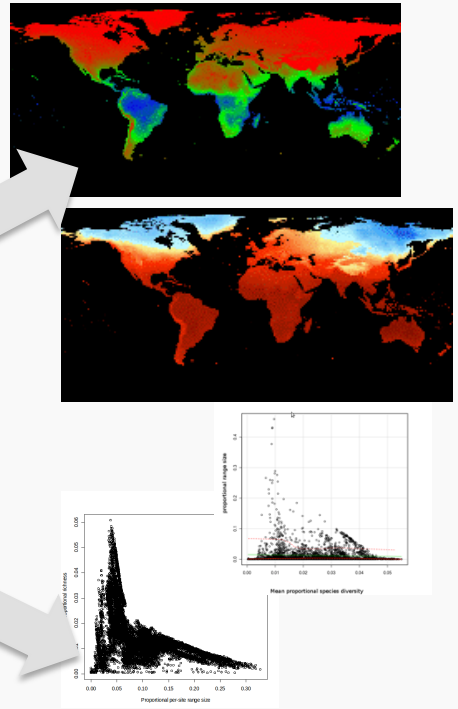
Macro-ecological Analyses



Species Habitat Data



Presence Absence Matrix (PAM)



Multi-species analyses

Range and Diversity Quantifications

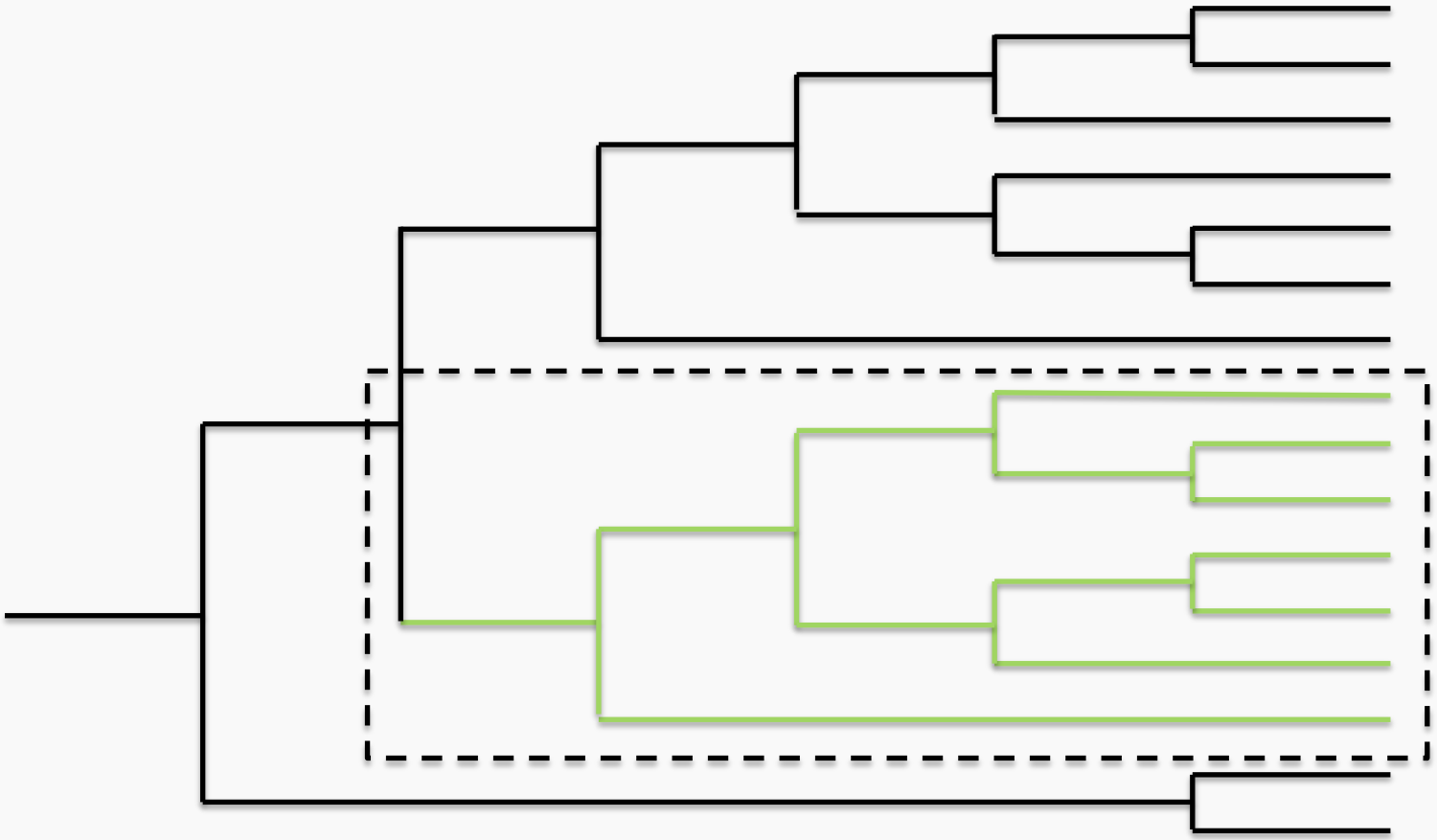
PRAGMA VBE

How to create a PAM

- What species group?
  - Subset by species in a study area
  - Subset by taxonomic or phylogenetic tree
- Define the area
  - Region bounding box
  - Site / cell size
- For every species, find presence/absence for every site
  - Lifemapper automates using input SDMs and thresholds
  - Calculate presence based on coverage and values in site

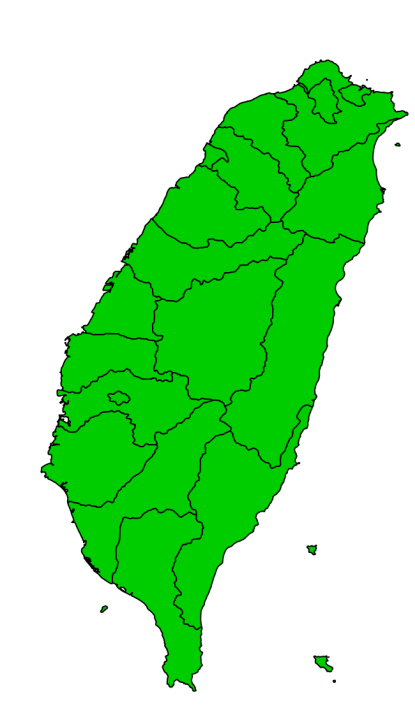
PRAGMA VBE

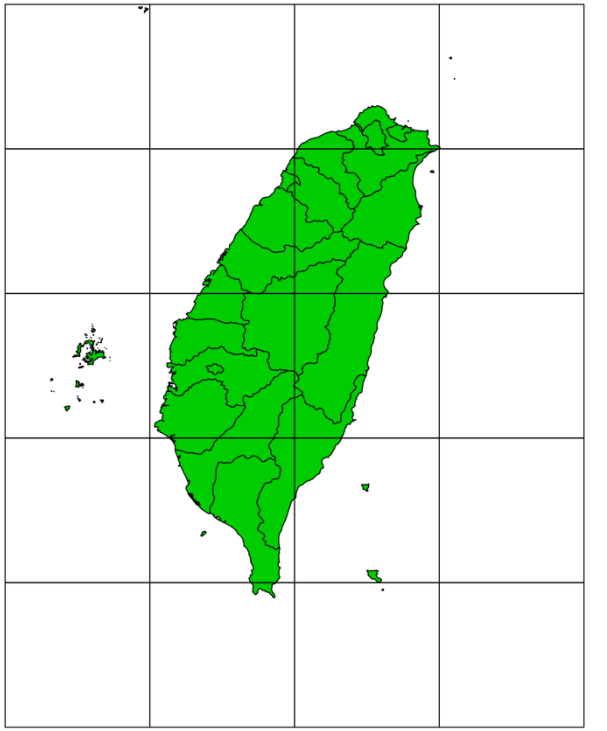
Select a species pool



PRAGMA VBE

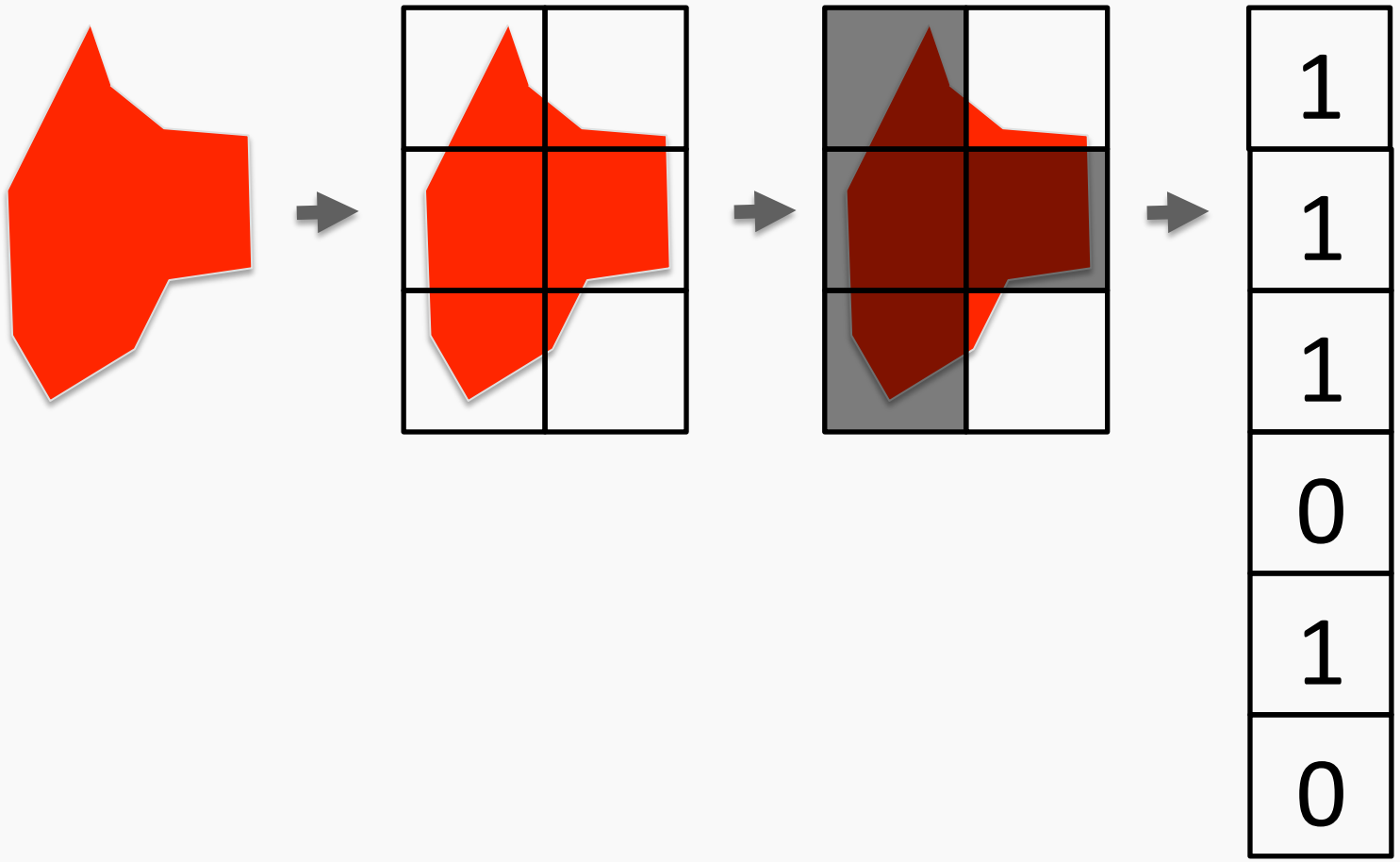
Create the grid





PRAGMA VBE

Determining presence



PRAGMA VBE

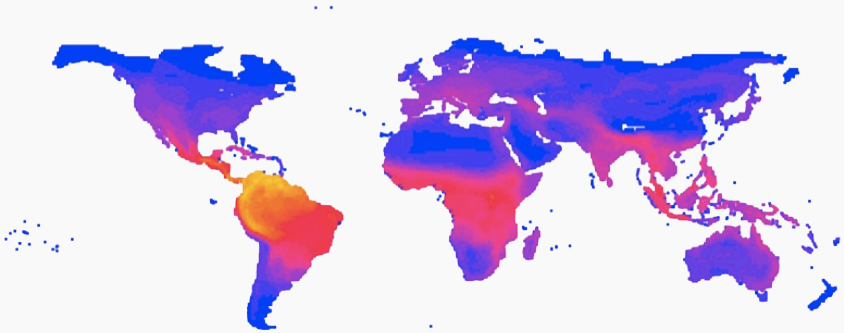
Site Statistics

- Species richness
  - The number of species present at each site
- Proportional species richness
  - The proportion of the number of species present at each site compared to the total number of species in the species pool
- Total range size
  - The sum of the range sizes of all of the species present at each site
- Proportional species range size
  - The proportion of the sum of the range sizes of all species at a site compared to the sum of the range sizes for all species in the species pool

PRAGMA VBE

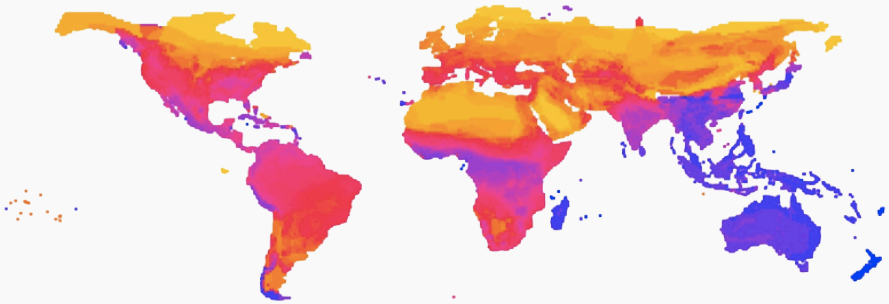
Lifemapper

Terrestrial Mammals



Proportional Species Richness

High	
Yellow	
Moderate	Red
Low	Blue



Per-site Range Size

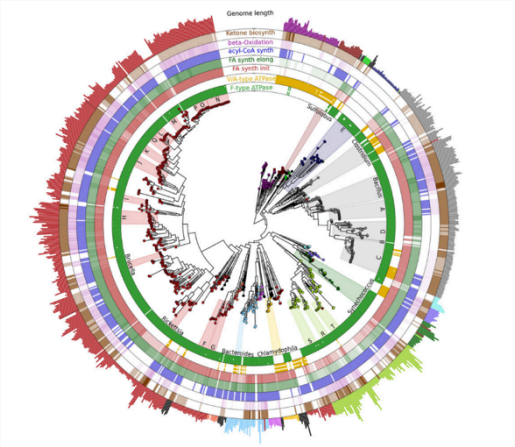
PRAGMA VBE

Global PAM and subsetting

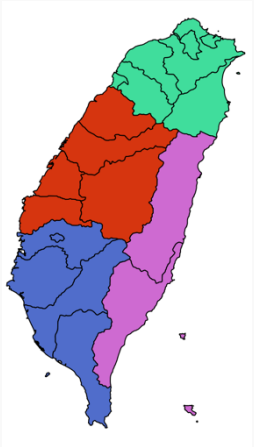
- Lifemapper creates a "**Global PAM"** for all species in a dataset
- Subsets of the Global PAM may be requested for analyses. Subset by:
  - Species
  - Geography
  - Time
- Study the effects of scale and time
  - Do patterns remain the same?

PRAGMA VBE

Meta-Community Phylogenetic Analyses (MCPA)



$$M = \begin{bmatrix} 0&0&0&0&1&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&1&0&1 \\ 1&0&1&0&1&0&0&0&0&1&0&1&0 \\ 1&0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&1&0&0&0&1&0&1 \\ 0&0&1&0&0&1&0&0&0&0&0&0&0 \\ 0&1&1&0&0&0&0&0&1&0&0&0&0 \\ 0&1&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&1&0&0&1&0&0&0&0&0&0&0 \\ 0&0&1&0&0&1&0&0&0&0&0&0&0 \\ 1&0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&1&0&0&0&0&0&0&0 \\ & & & & & & \vdots & & & & & & \end{bmatrix}$$



PRAGMA VBE

MCPA

Evaluates the interaction between

- phylogenetic structure
- historical biogeographic events
- environmental filtering

in driving species distributions in a community

Inputs

- PAM – inputs can be SDMs or range maps
- Phylogenetic tree
  - Must be binary
  - If it has branch lengths, the tree must be ultrametric
  - Encoded into a species X node matrix
- Environmental predictors
  - Intersected with sites to create a matrix matching PAM
- Biogeographic hypotheses (barriers or corridors for species distribution)
  - Encode all into a single matrix as (-1, 0, 1)

PRAGMA VBE

Lifemapper

Demo

PRAGMA VBE

Virtualization Technologies

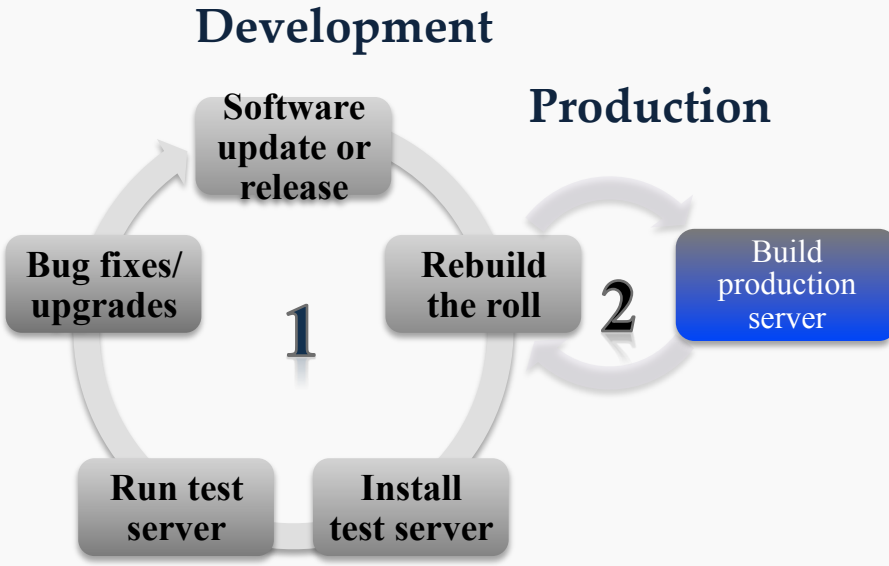
- Runs on Linux on x86 hardware
- Ideal to run on clusters
- Advantages in VBE
  - Can create bigger instance
  - Can have long running instance
  - Dynamic input data
  - Intended for use by multiple external clients

- Runs on Linux, Windows, OSX
- Ideal for laptops but
  - Must setup networking
  - Instance size/memory is limited by laptop hardware
- Advantages in VBE
  - Can special-purpose instance
  - Can have short-lived instance
  - Pre-defined stable input data
  - Intended for field work (with no network connection) and teaching tool

PRAGMA VBE

Virtualization Advantages

- **Increase availability and flexibility of Lifemapper**
  - Reduce cost of installing/configuring/ replicating
  - Ease burden of hardware and software integration
- **Reduce time from software update to server availability:**
  - Minimize time spent on build and configuration
  - Automate most hands-on tasks.
  - Implement tests components and configuration
- **Prepare for larger, high quality data and increased computational complexity:**
  - low resolution climate data → high resolution satellite imagery
  - single-species SDM experiments → multi-species macro-ecology experiments



Lifemapper

http://lifemapper.org

https://github.com/lifemapper/

Rocks

http://www.rocksclusters.org

PRAGMA

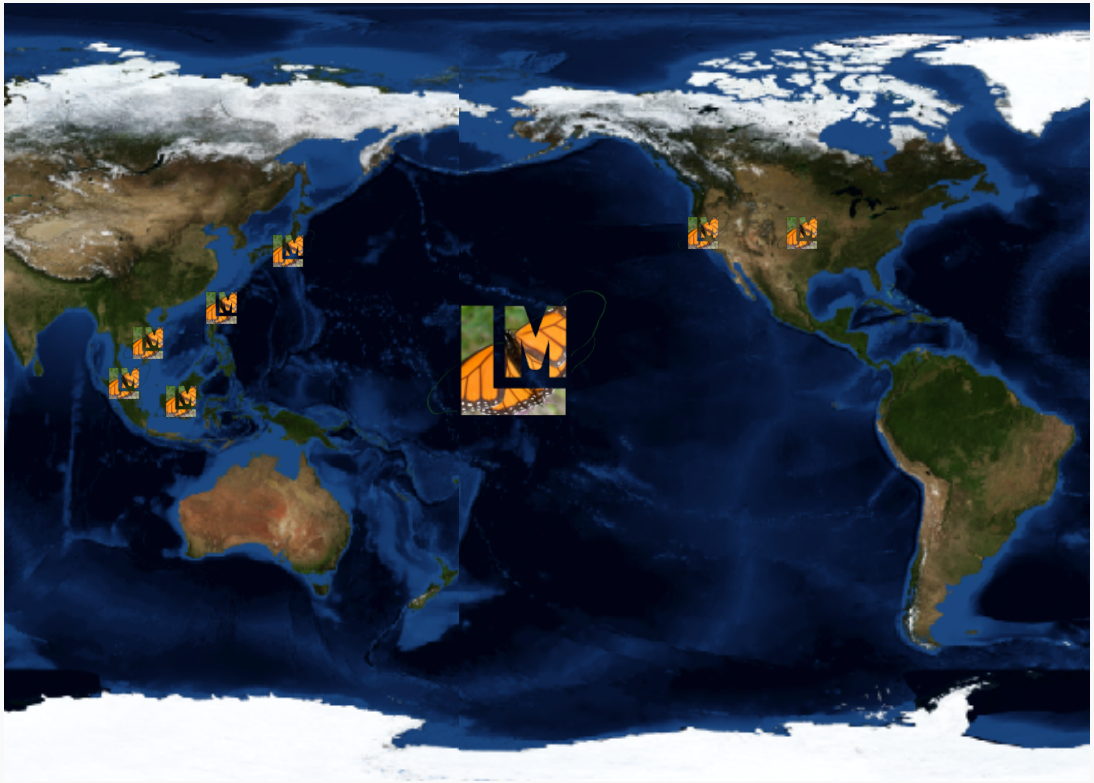
https://github.com/pragmagrid/lifemapper-compute

https://github.com/pragmagrid/lifemapper-server

PRAGMA VBE

Distributed Lifemapper

- Support focused and comparative studies
  - Taiwan at NCHC
  - Workshop for researcher introduction
  - Follow up with Taiwanese museums and researchers



PRAGMA VBE

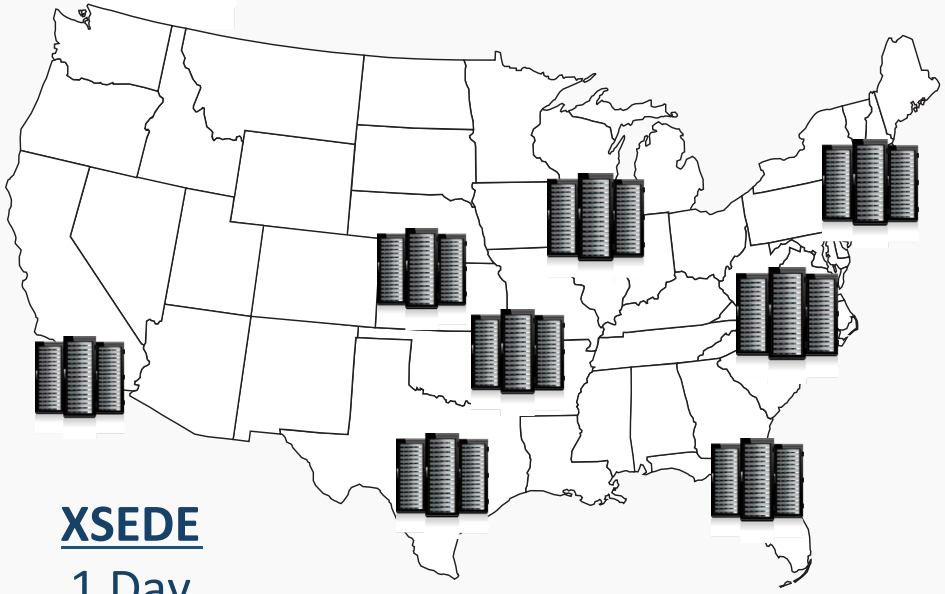
Scaling Computation from Desktop to National Grid

Inputs
500,000 species
4 climate scenarios
3 model algorithms
(= 6 million models)

Desktop
1.8 Years

KUBI Cluster
104 Days

KU ACF
5 Days



XSEDE
1 Day

PRAGMA VBE

Lifemapper funded by:

US NSF

OCI-1234983

BIO-1458422

BIO-1356732

USGS

BIP-G14AC00285

PRAGMA VBE

Biodiversity Expedition Next Steps

- Regional installations
- Move to HPC challenges – scaling
  - management / computational resources
  - Tune Notre Dame's workflow tools + compute resources
- UI changes allowing display through SAGE2
  - Replaced QGIS with generic browser-based UI
  - Works on dynamic web service calls OR local data
- Next steps:
  - Create/populate then move VC?
  - Containers? Virtual frontend + containers for computations?
  - Live connection b/w iDigBio and OpenTree (Biotaphy)
  - Marine models with NMMST (CENTRA)