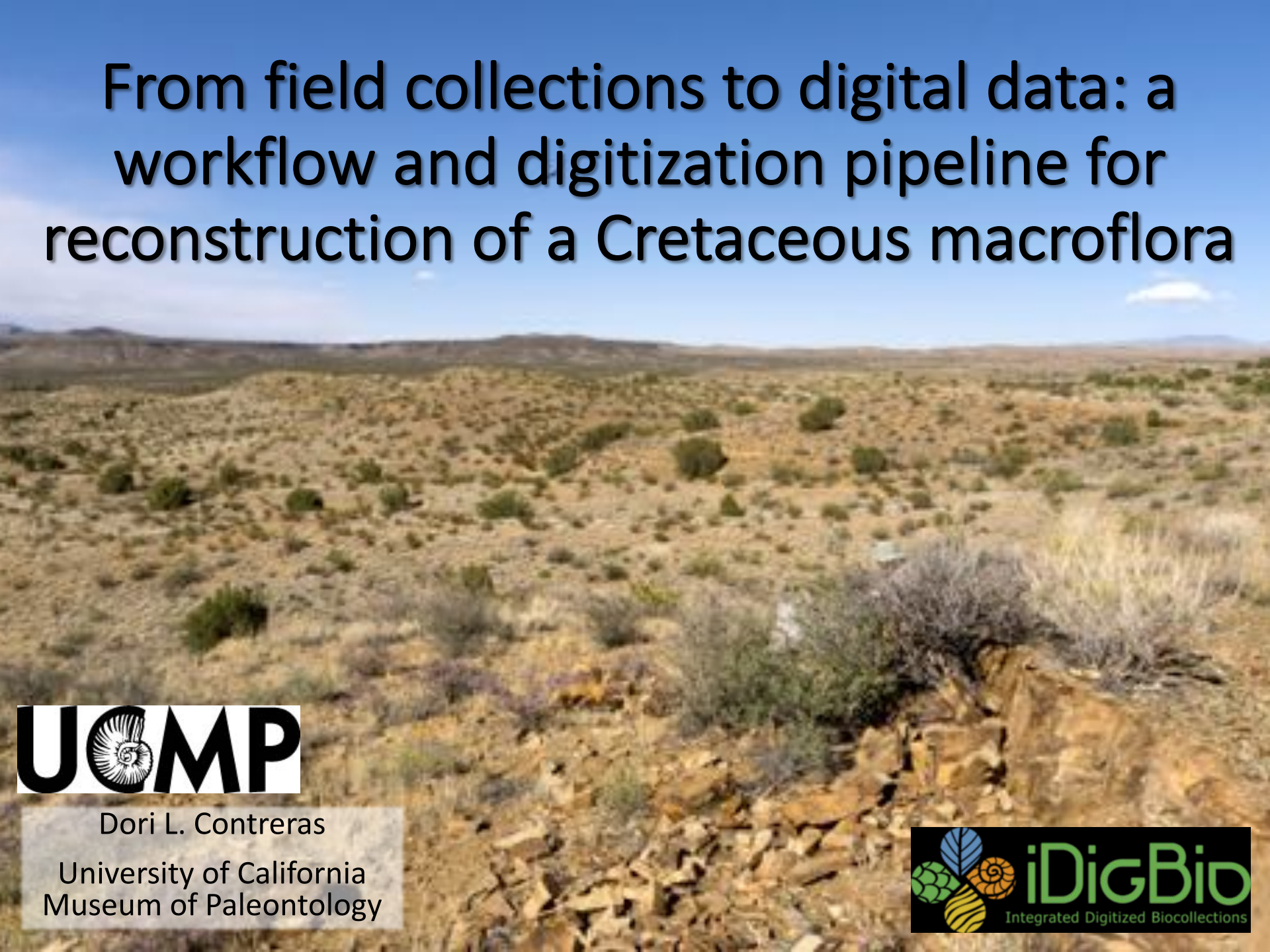
From field collections to digital data: a workflow and digitization pipeline for reconstruction of a Cretaceous macroflora



Dori L. Contreras

University of California
Museum of Paleontology

iDigBio
Integrated Digitized Biocollections

Meeting multiple needs

Research project goals:

- Collect new fossils
- Identify/describe
- Collect data/measure
- Analyze data
- Publish manuscript
  - (Wait! Need repository info and specimen numbers!)

Museum & archival goals:

- Locality info and supporting data
  - field notes, GPS, etc.
- Unique IDs for each specimen (catalog #s) tied to locality
  - georeferenced
- Care and house specimens
- Digitization – collection records and specimen photos
  - Online portals
  - Outreach
  - Accessibility to other researchers

Research based on new collections: integrating short and long-term needs

1) Collection
- Finding and collecting fossils
- Locality data and samples
- Field data (e.g. census counts)

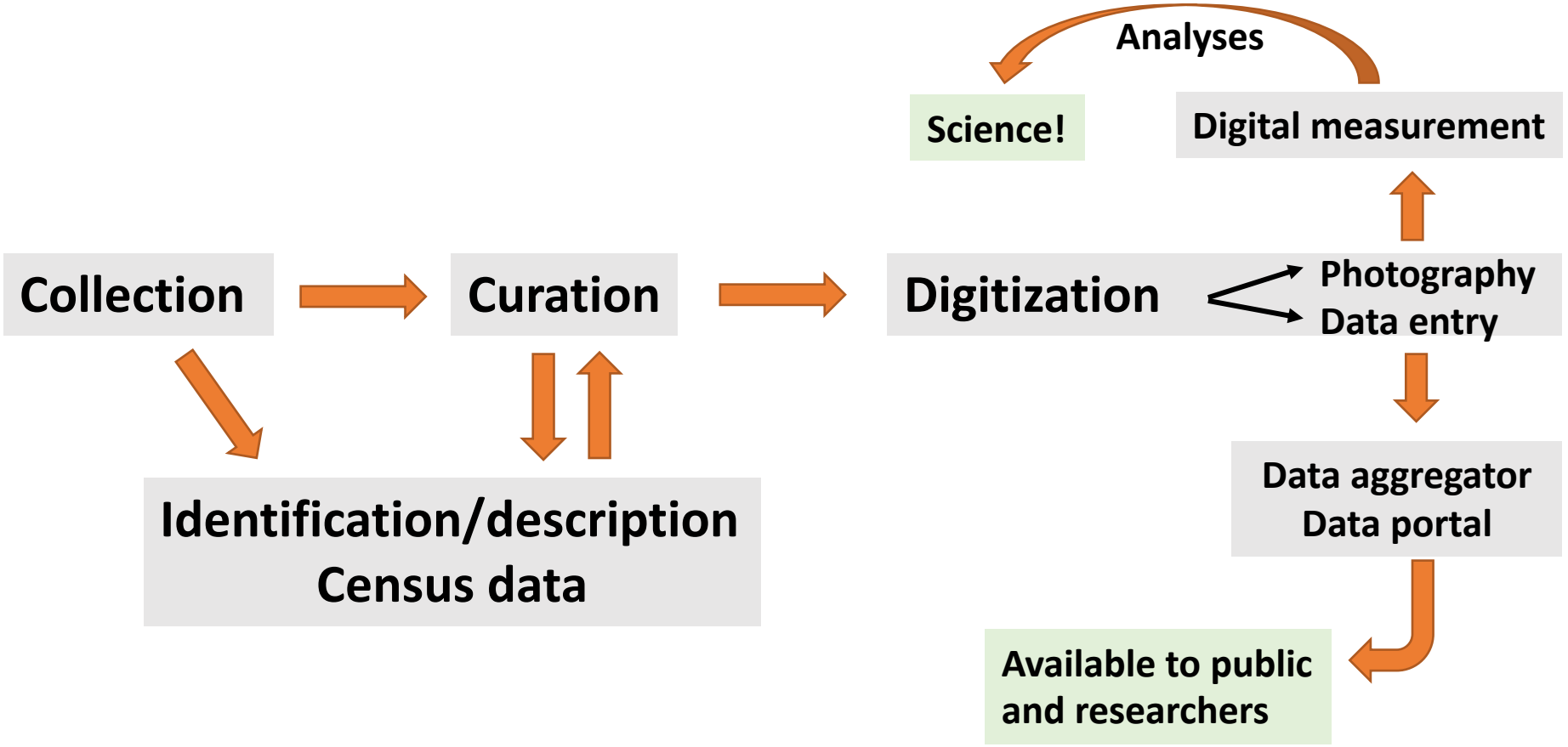
2) Identification/description; Census

3) Curation
- Specimen preparation
- Labels & specimen numbers
- Organize and storage

4) Digitization
- Photography
- Data entry

5) Digital data
- Measurements
- Data aggregators

Research based on new collections: integrating short and long-term needs



An example from ongoing project

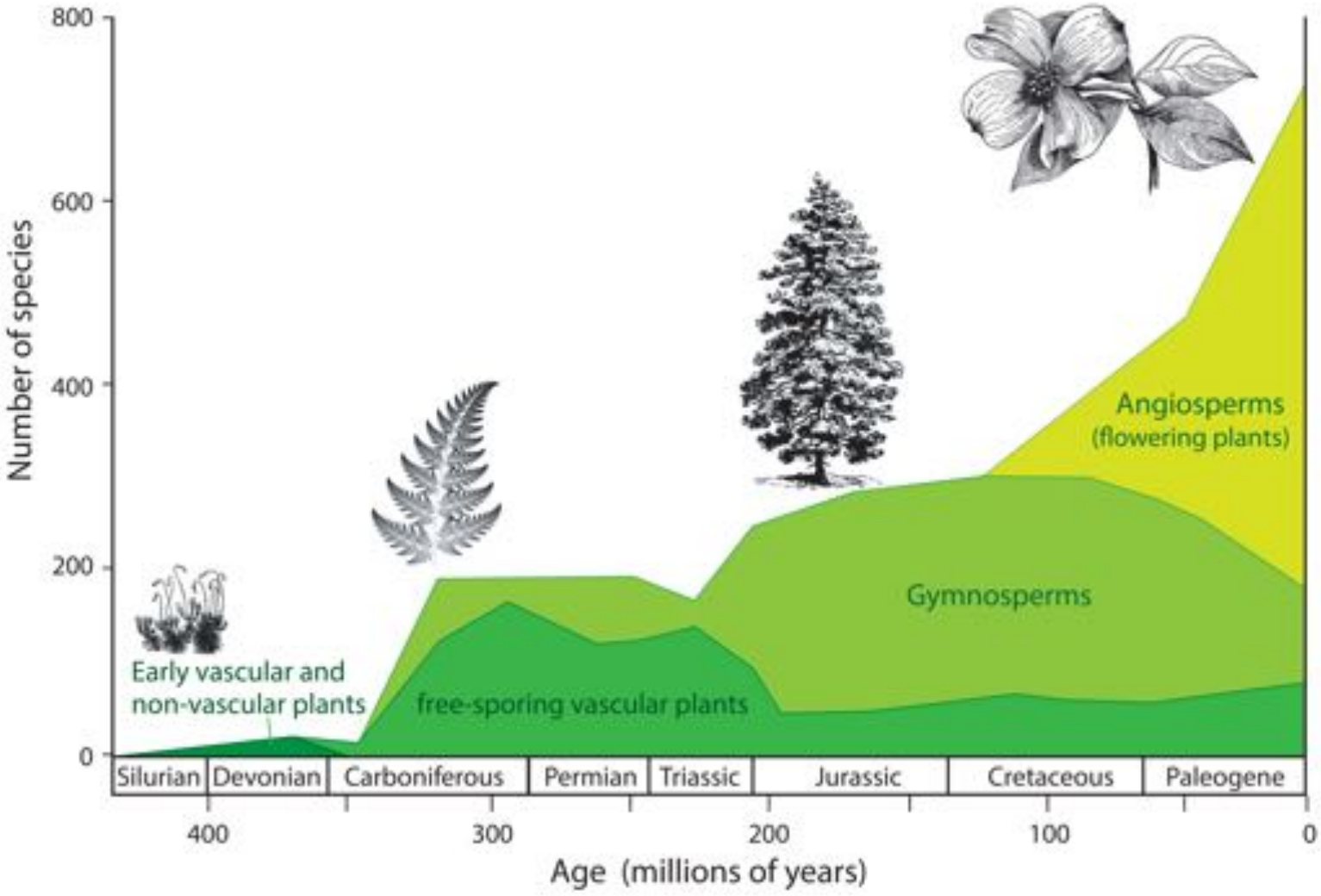
Reconstructing a Cretaceous forest

1) **Project description**

2) **Methods & Workflow**
   -roadblocks and solutions

3) **Results thus far and use of digital data**

AIM: to address fundamental questions about the structure, diversity, and functioning of forests during the Late Cretaceous ecological radiation of angiosperms



Redrawn from Niklas 1986

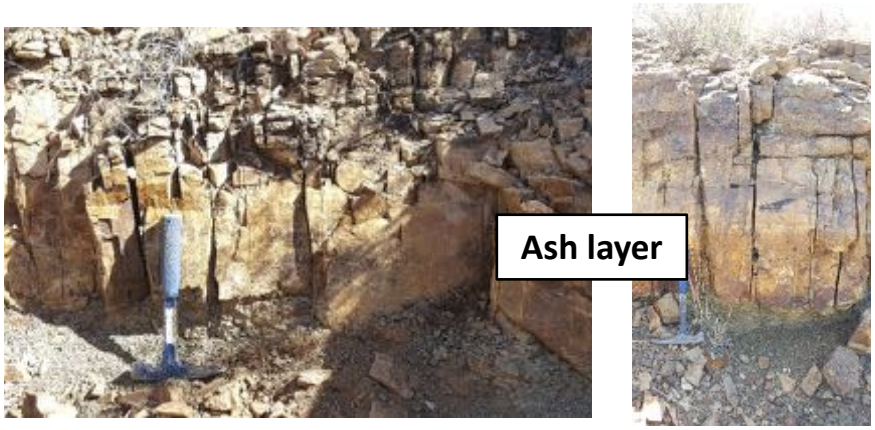
McRae Fm: Jose Creek Member

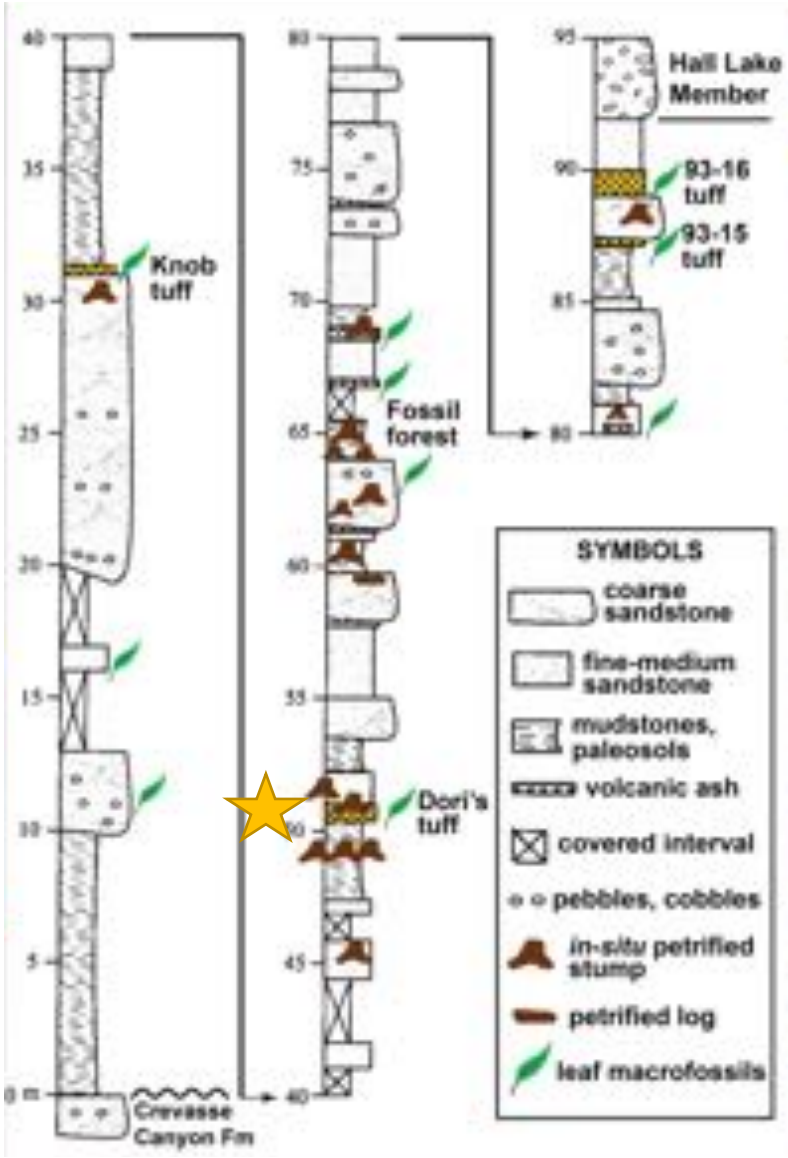
Late Campanian

South-Central New Mexico

“Dori’s tuff”

- Ashfall bed with little to no transport
- Single depositional event on stable floodplain
- 74.7 ± 0.6 Ma (U-Pb, Amato et al. accepted)





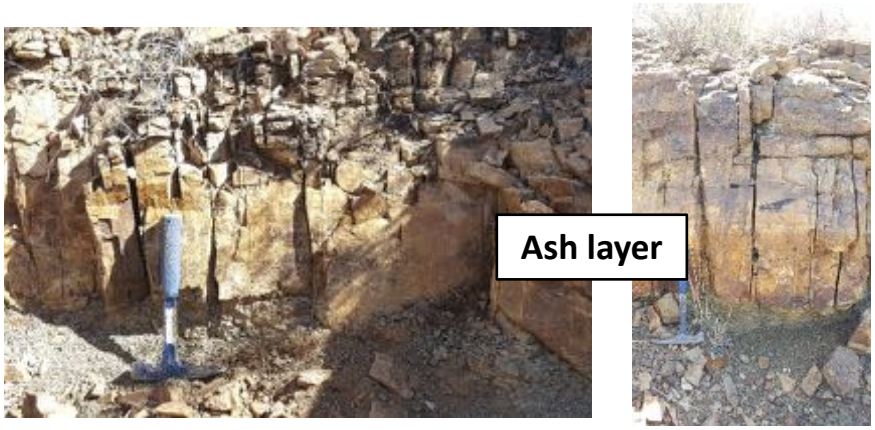
McRae Fm: Jose Creek Member

Late Campanian

South-Central New Mexico

“Dori’s tuff”

- Ashfall bed with little to no transport
- Single depositional event on stable floodplain
- 74.7 ± 0.6 Ma (U-Pb, Amato et al. accepted)



Project Overview:

Reconstructing structure & functional diversity of a Cretaceous forest

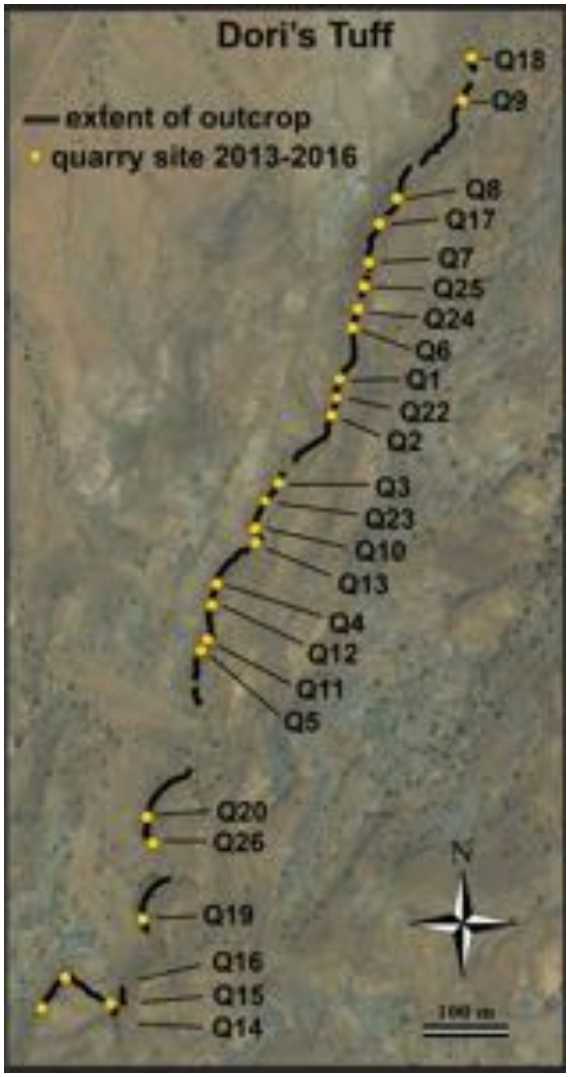
Phase 1. "Build" a forest

-describe taxonomic diversity
-relative abundance of taxa
-spatial structure of community

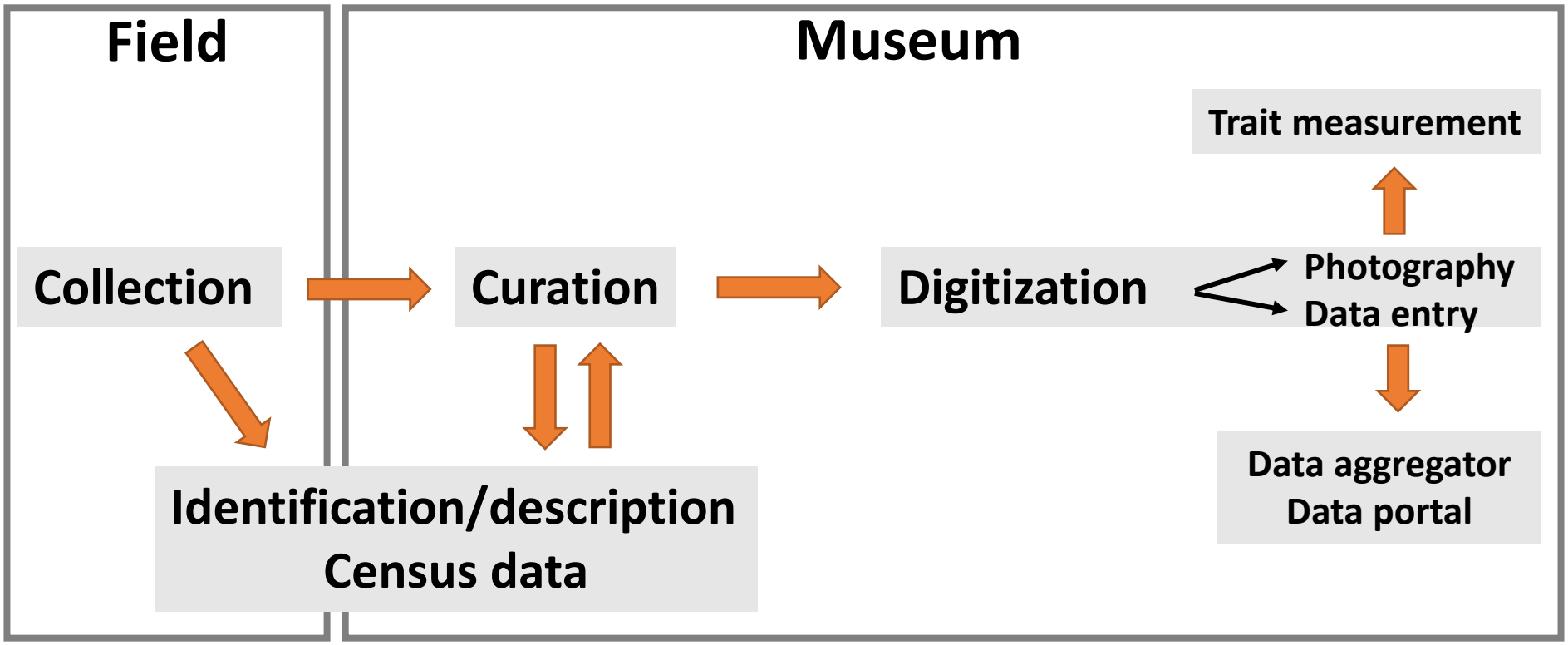
Phase 2. Measure the forest: quantify functional diversity

-measure functional leaf traits of all taxa
-reconstruct trait diversity across transect

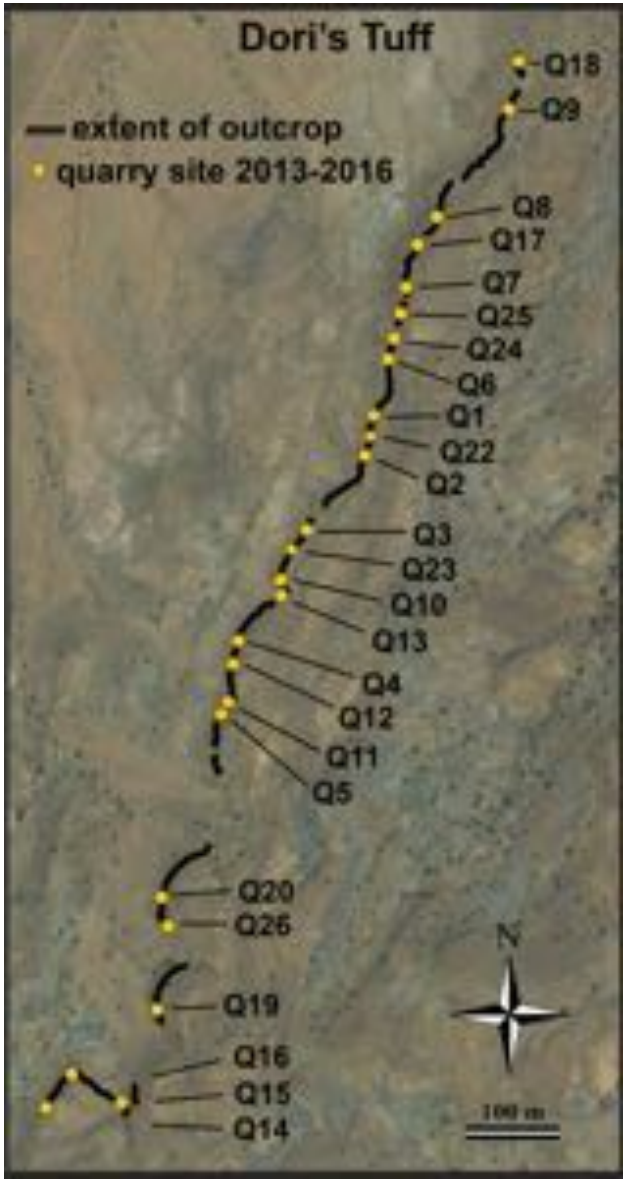
*quantitative and spatial explicit sampling scheme
*large sample sizes
*digital measurements of georeferenced specimens



A workflow from field to digital



Field Work – localities



<u>Established 26 quarry sites spanning deposit:</u>

- each treated as separate locality
- collected relevant site data for each:
  - GPS coordinates
  - sedimentological info and samples, quarry dimensions
  - photographs

Quarry site:

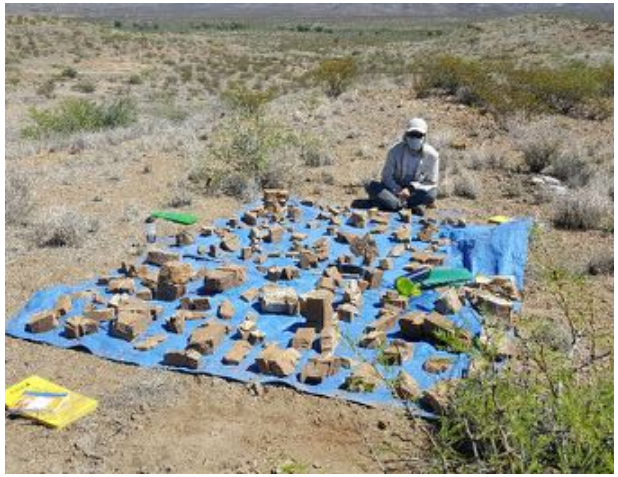
Field Work – collections and census

At each quarry:

- Bust out rock and expose fossils

  each rock gets a unique field ID number

- Census: Identify and record leaf morphotypes by:
  1) Relative abundance: number of leaf specimens
  2) percent cover: line-intercept method on rock surface in 2-cm increments (Wing *et al.* 1993, 2012)



Census: line increments

> 6,350 specimens
> 61,718 cm² of rock surface

"Keepers" are wrapped and labeled

> 1,945 rocks brought to UCMP

The spoils of field work

Spoils:
Locality info
Census databooks
Boxes of fossils

Collection → **Curation** → **Digitization** → **Photography** / **Data entry**

Trait measurement

Identification/description
Census data

Data aggregator
Data portal

The transition

Spoils:
Locality info
Census databooks
Boxes of fossils

Museum

Field work

Curation

Digitization

Photography
Data entry

Trait measurement

Data aggregator
Data portal

Identification/description
Census data

Museum and Lab: the transition

1) Data entry: census data

2) Unpacking fossils

- Matching parts and counterparts
- White striping
- Boxing and organizing into drawers
- Curatorial log
  *Rock ID

Museum and Lab: the transition

1) Data entry: census data

2) Unpacking fossils

***Reconcile for printing labels with species identifications**

Matching parts and counterparts

White striping

Boxing and organizing into drawers

Curatorial log
*Rock ID

101 Drawers to process….

Spoils:
Locality info
Census databooks
Boxes of fossils

Museum

The un-censused
-specimens that could not be identified in field
-new morphotypes

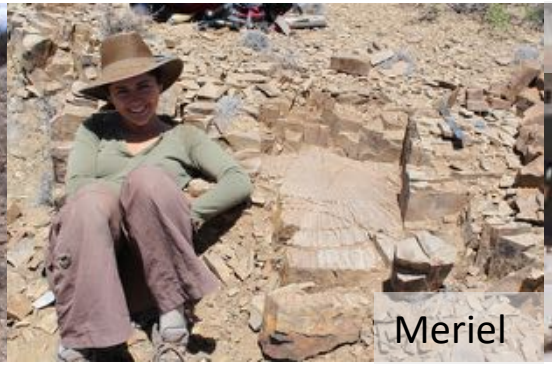
Collection → **Curation** → **Digitization** → **Photography** / **Data entry**

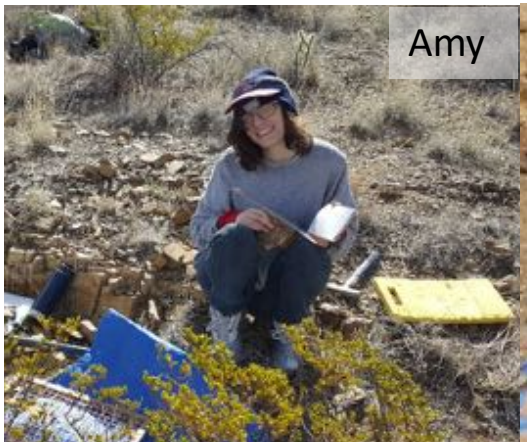
Trait measurement

Identification/description
Census data

The censused
-specimens ID'ed & well-preserved enough for trait measurements

Pipeline driven by people





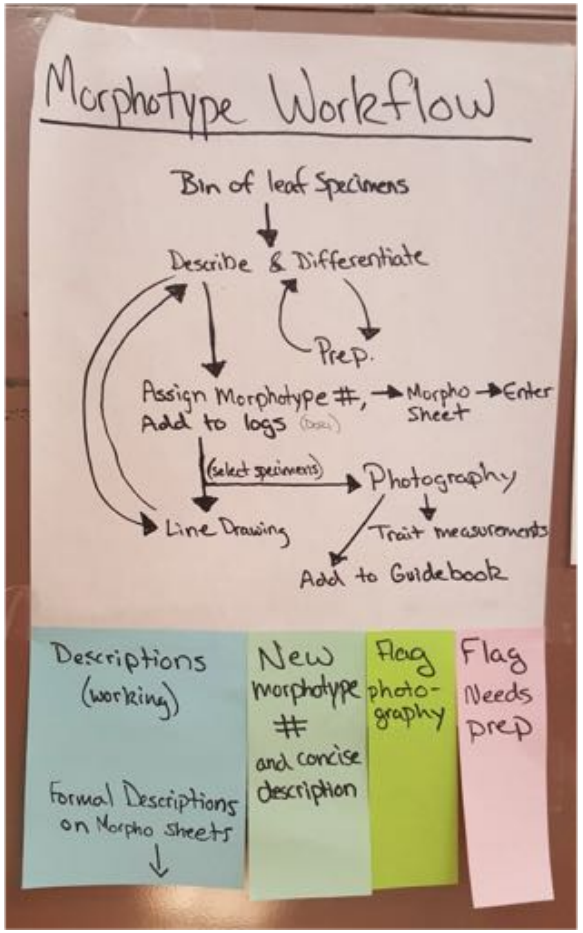
Processing: Two Methods

Batch processing

-one type of task at a time

-flag specimens for different processes

-let specimens build up until have a "batch" to process



Integration of tasks

-do everything at once!

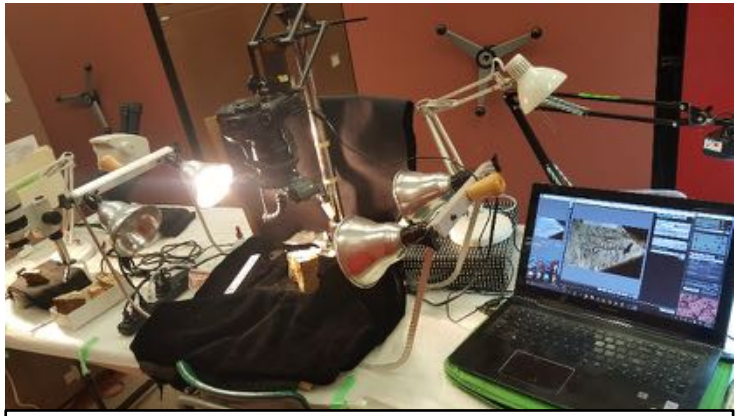
-drawer by drawer, processing each specimen completely

Re-organization and customization of workspace

Re-organization and customization of workspace

Prep station

White-stripe and census station



Photography station

Re-organization and customization of workspace

Prep station

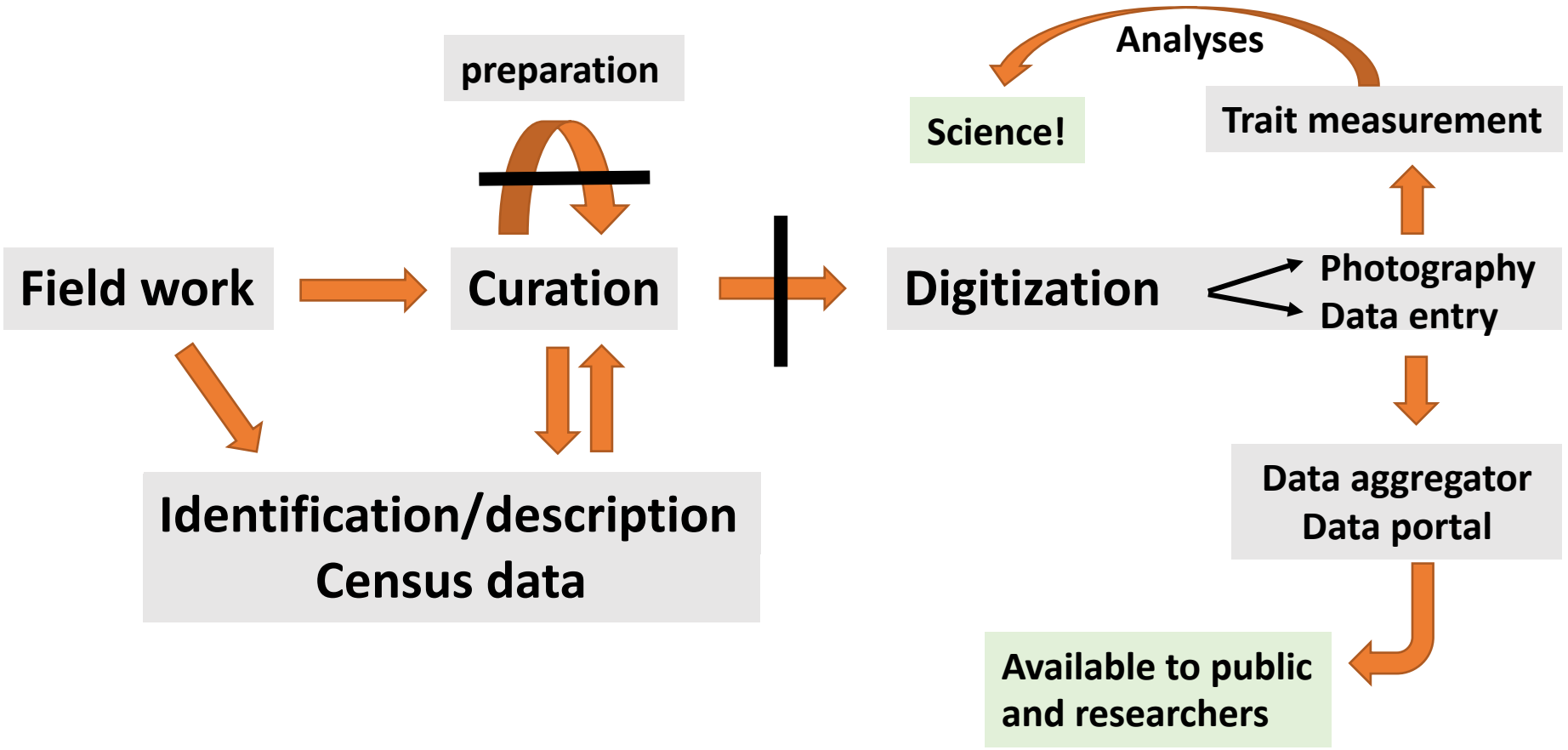
White-stripe and census station

Photograph

Re-organization and customization of workspace

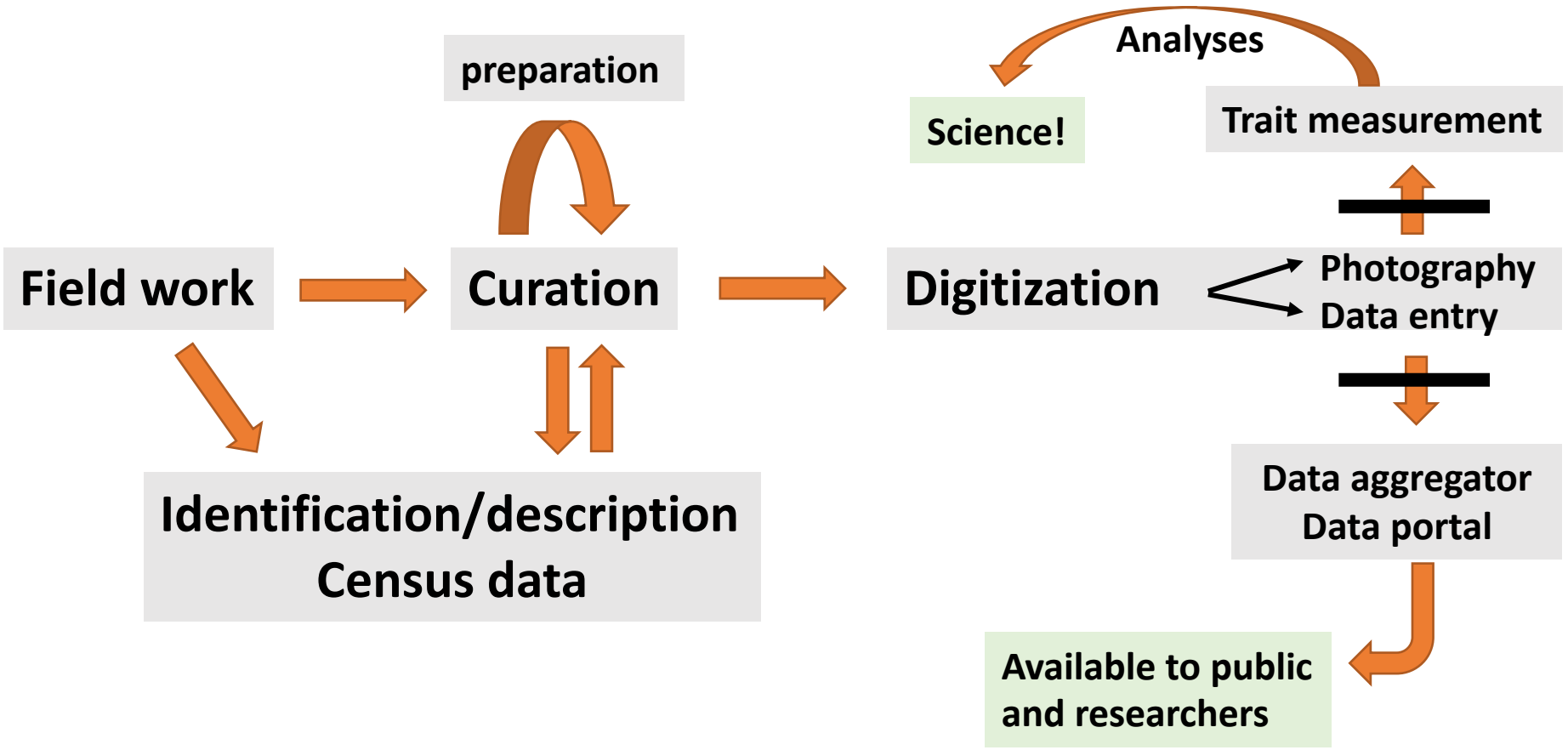
Shifting pinch-points in the workflow

Batch processing



Shifting pinch-points in the workflow

Integrated processing



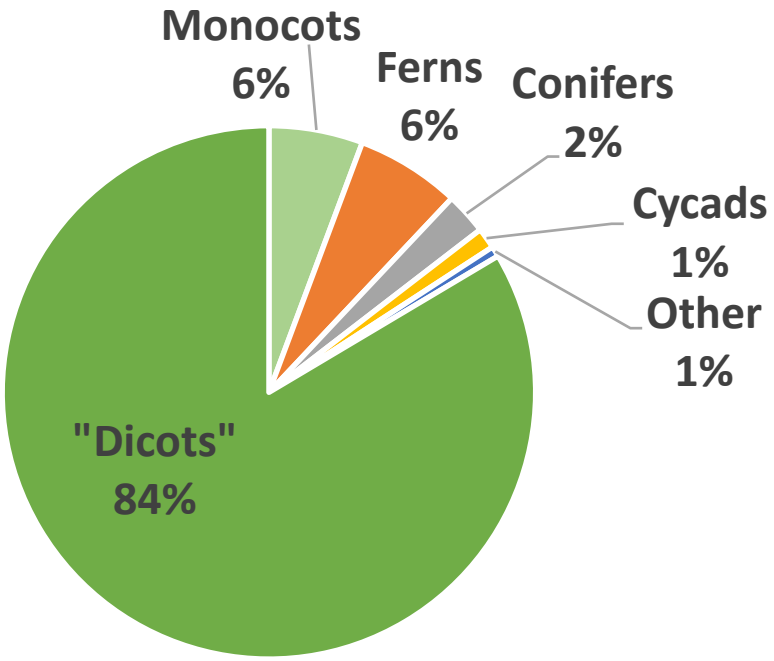
Putting all that data to use….!!

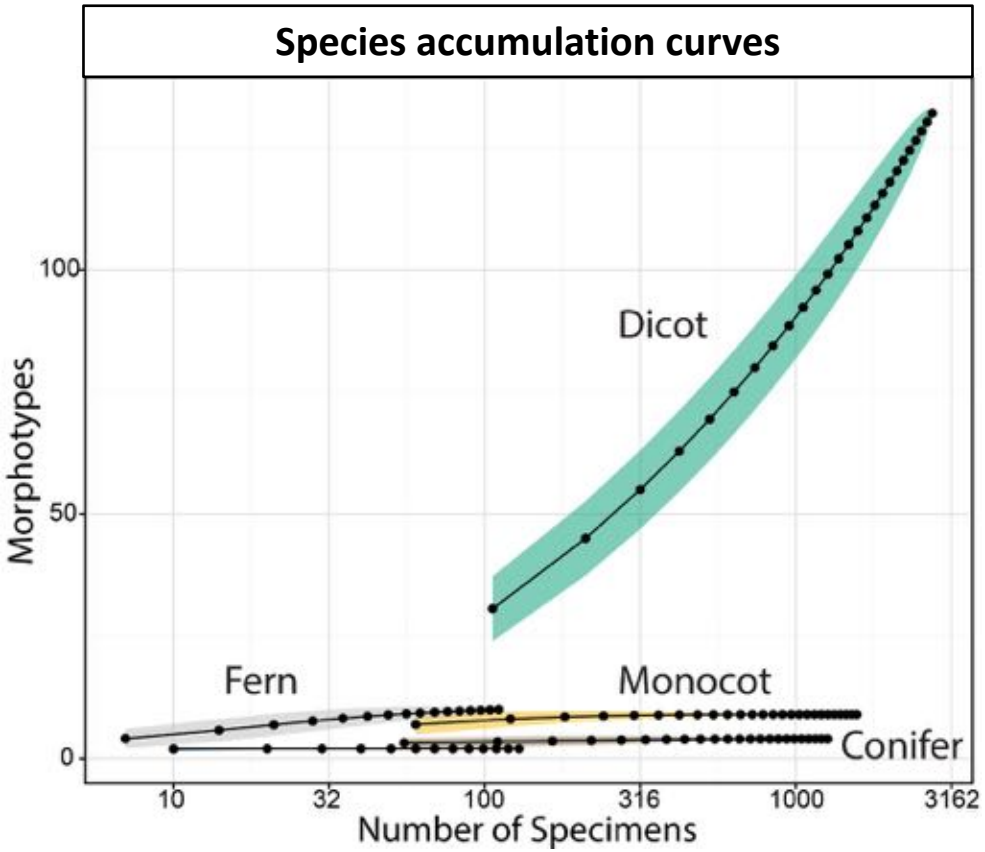
Phase 1 : Community structure
Results so far

Total Taxonomic Diversity (thus far)

Census:

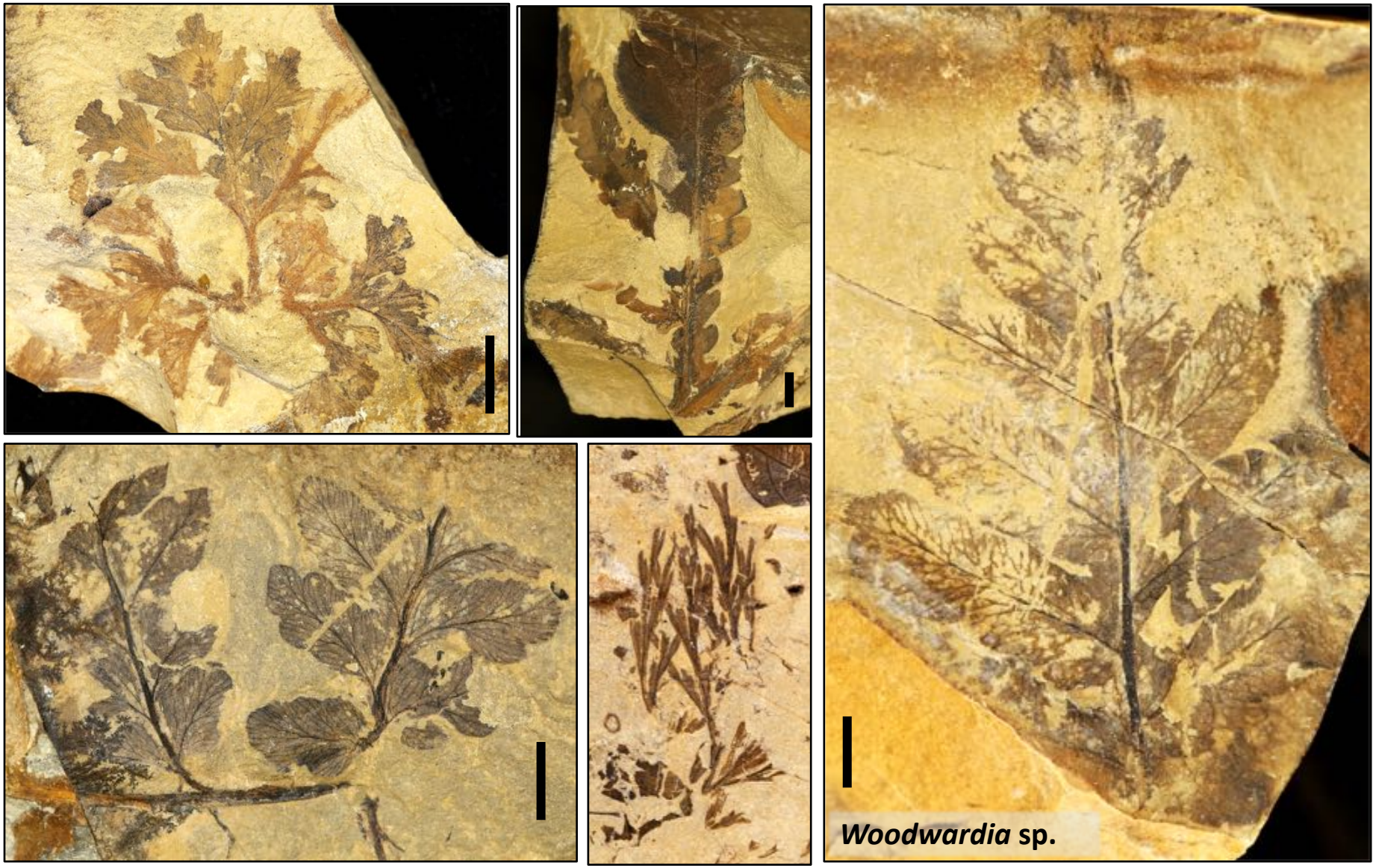
- **158 leaf morphotypes**
- Angiosperms: ~89% of diversity (141 morphotypes)



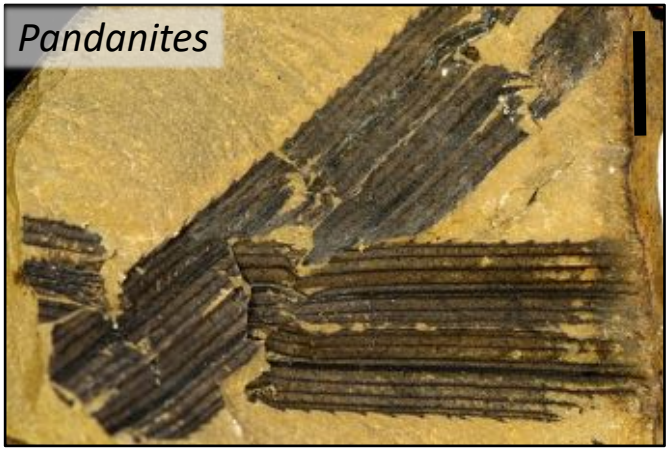


Conifers & Cycads

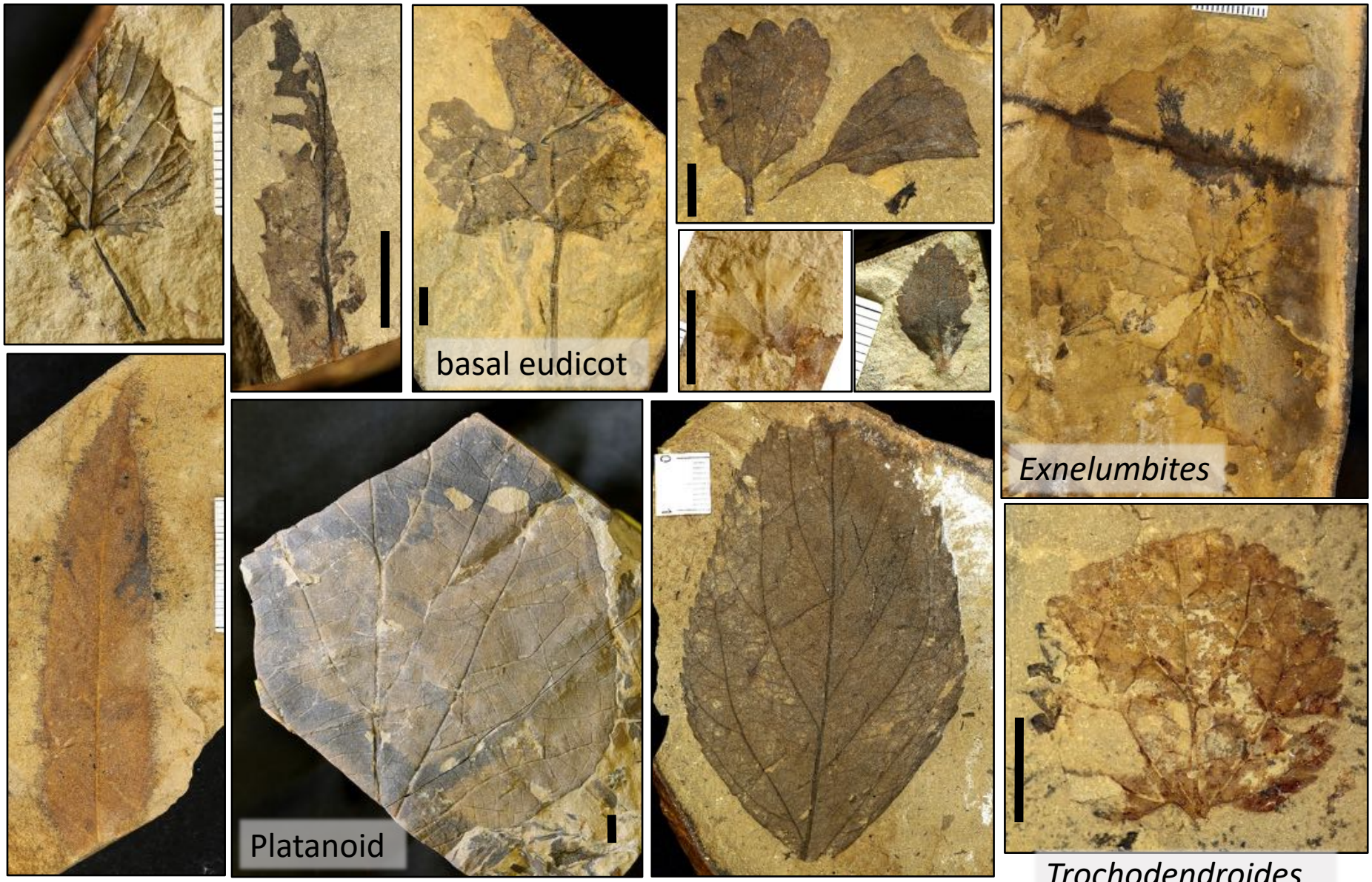
Ferns (rare, but distinct)



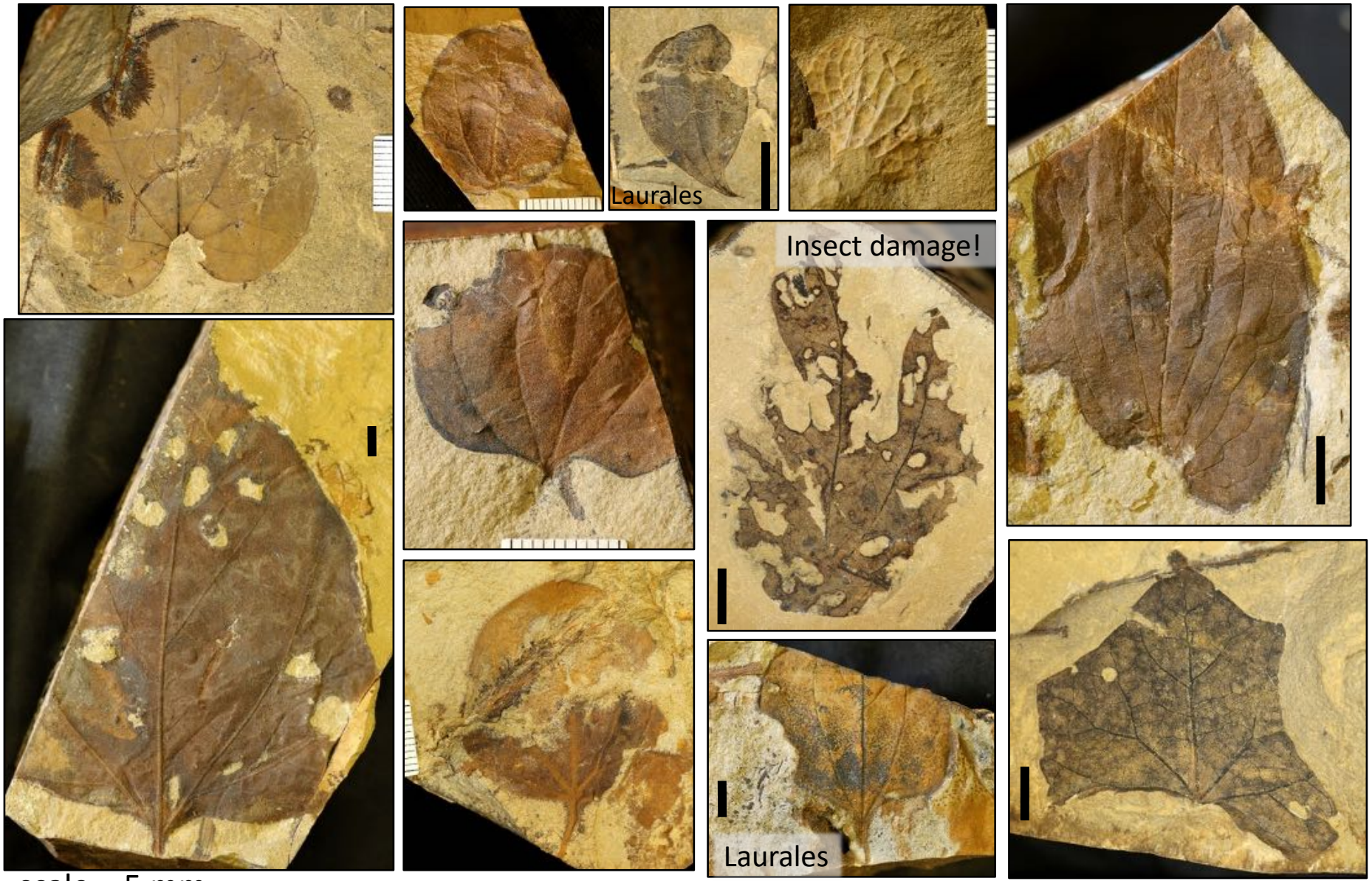
Angiosperms - monocots



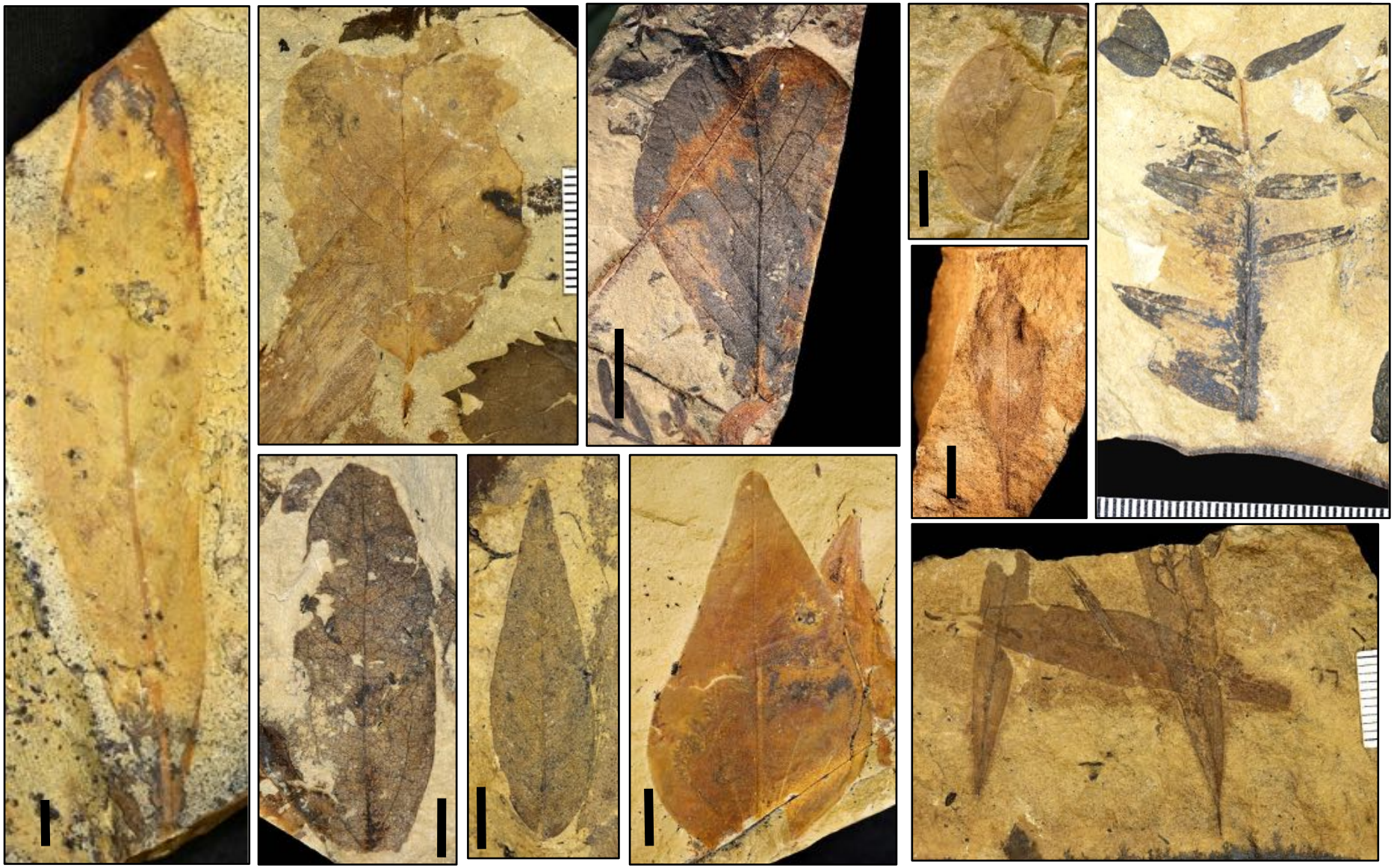
Angiosperms – toothed “dicots”



Angiosperms – palmate “dicots”



Angiosperms – pinnate “dicots”



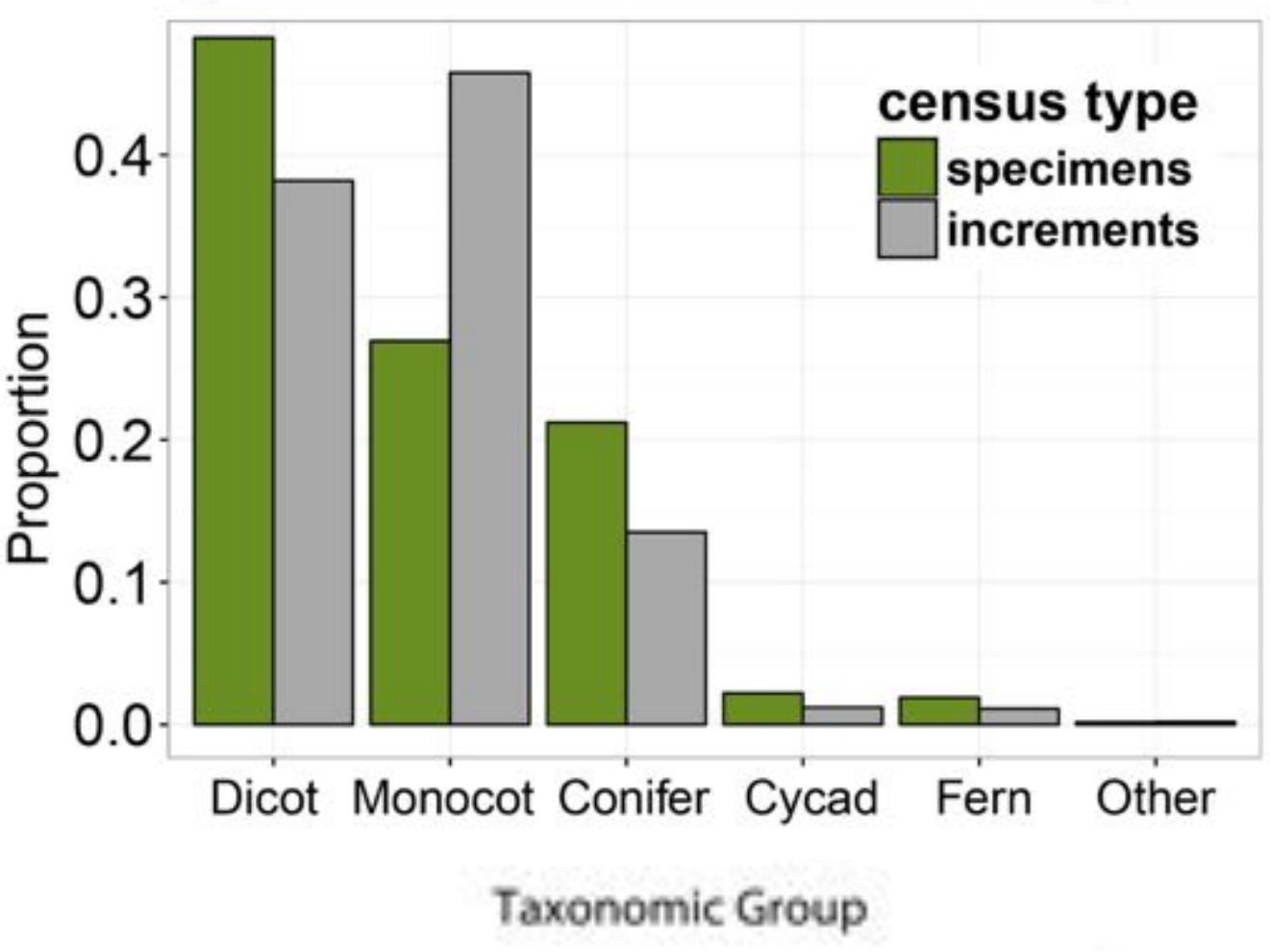
scale = 5 mm

Angiosperms – pinnate “dicots”

scale = 5 mm

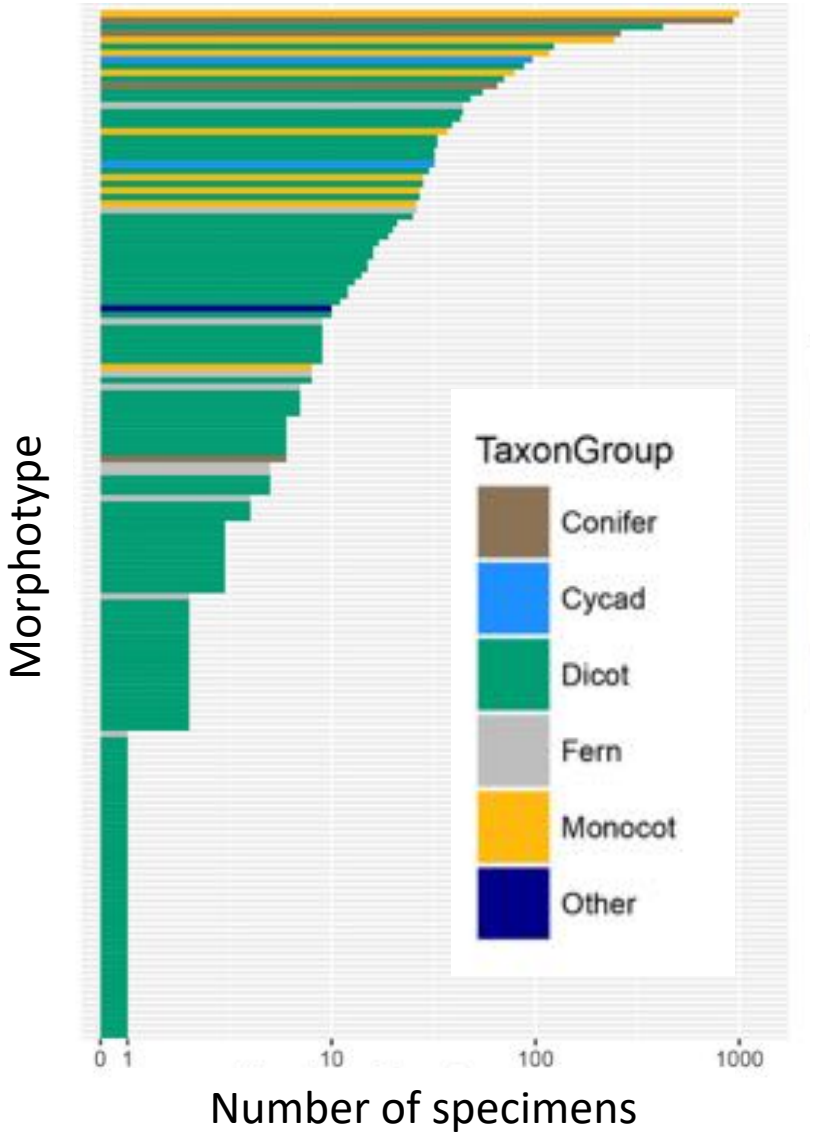
Community composition

Relative abundance and % cover



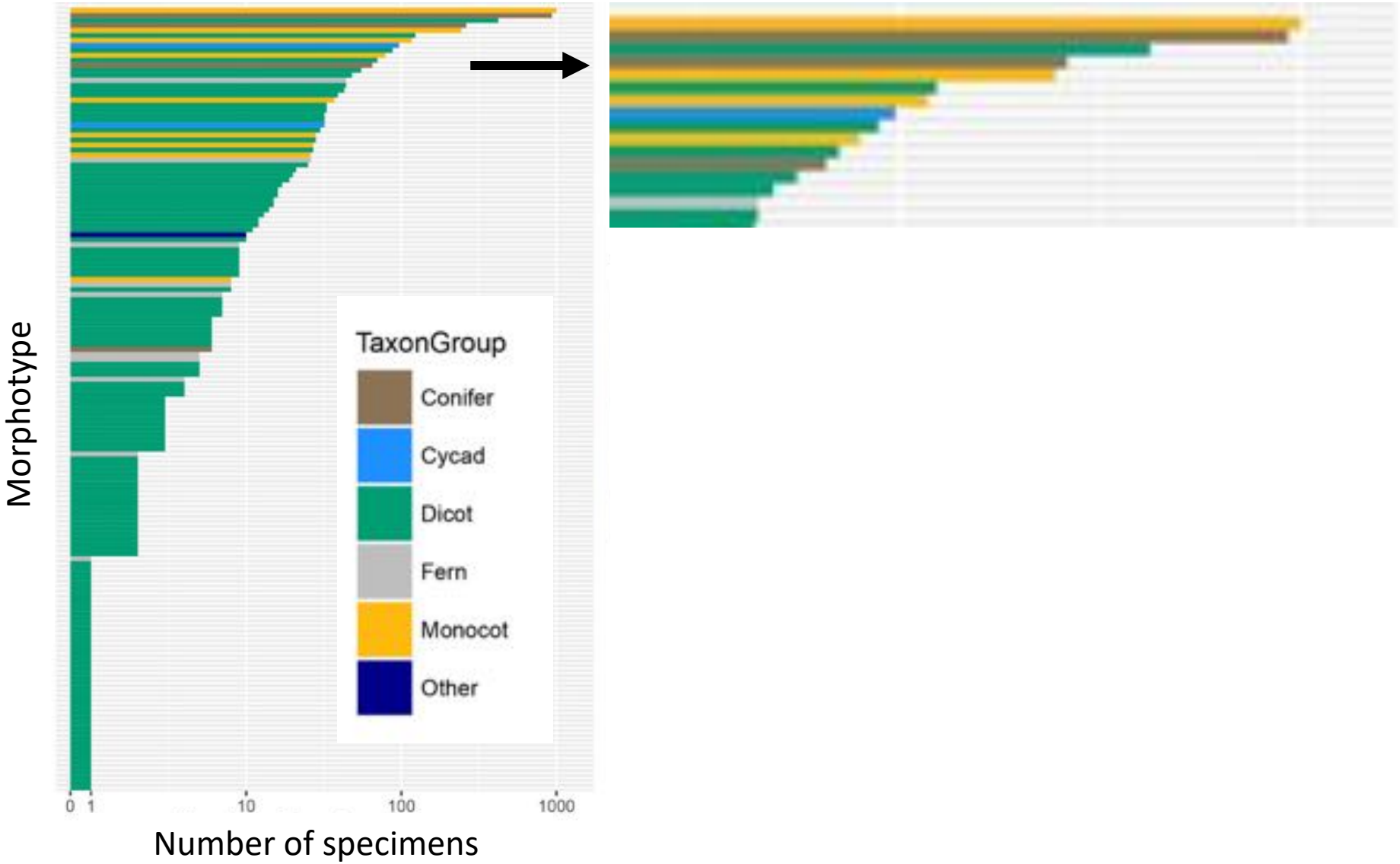
Community composition

Dominance and rank abundance



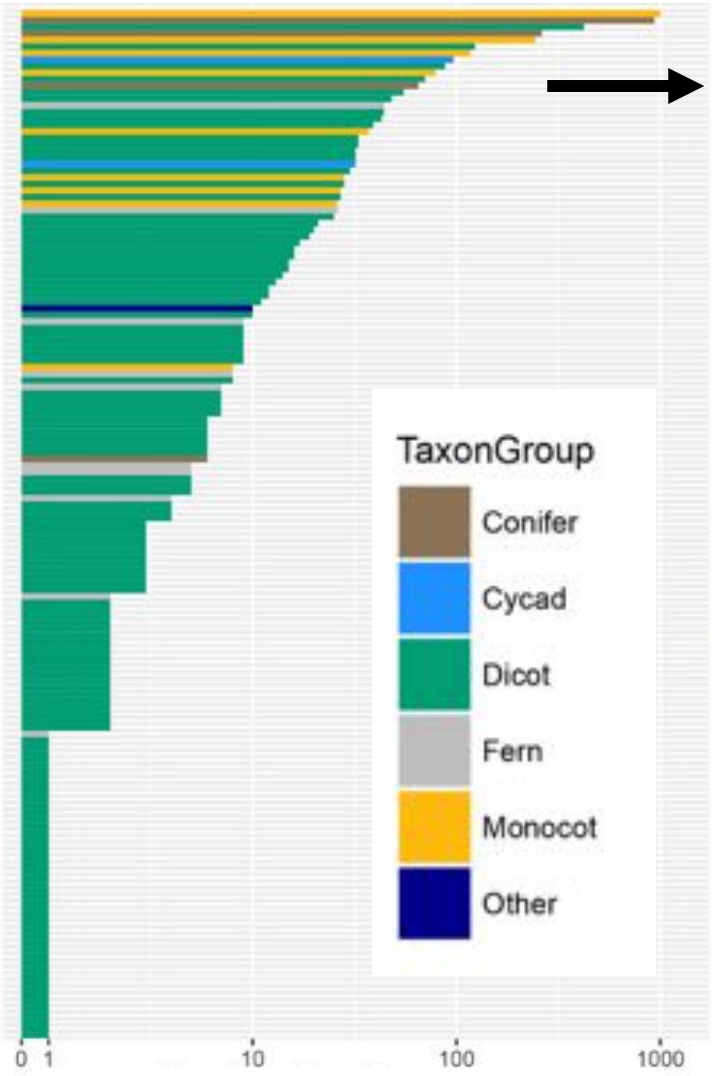
Community composition

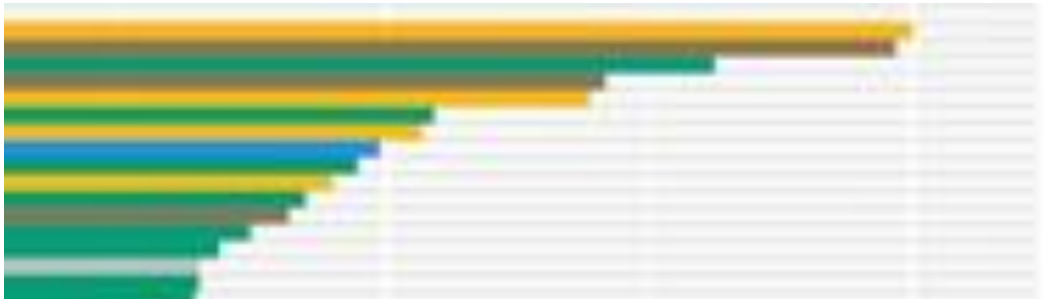
Dominance and rank abundance



Community composition

Dominance and rank abundance





Redwood relative

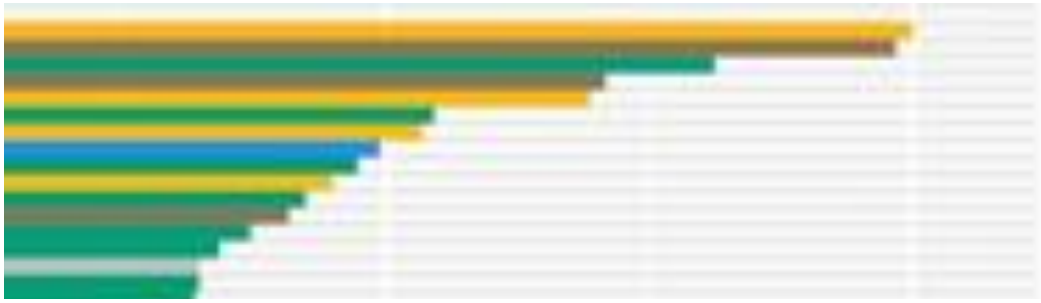
tree

Zingerberopsis

ground cover

Community composition

Dominance and rank abundance



Brachyphyllum sp.

Sabalites sp.

Dryophyllum sp.

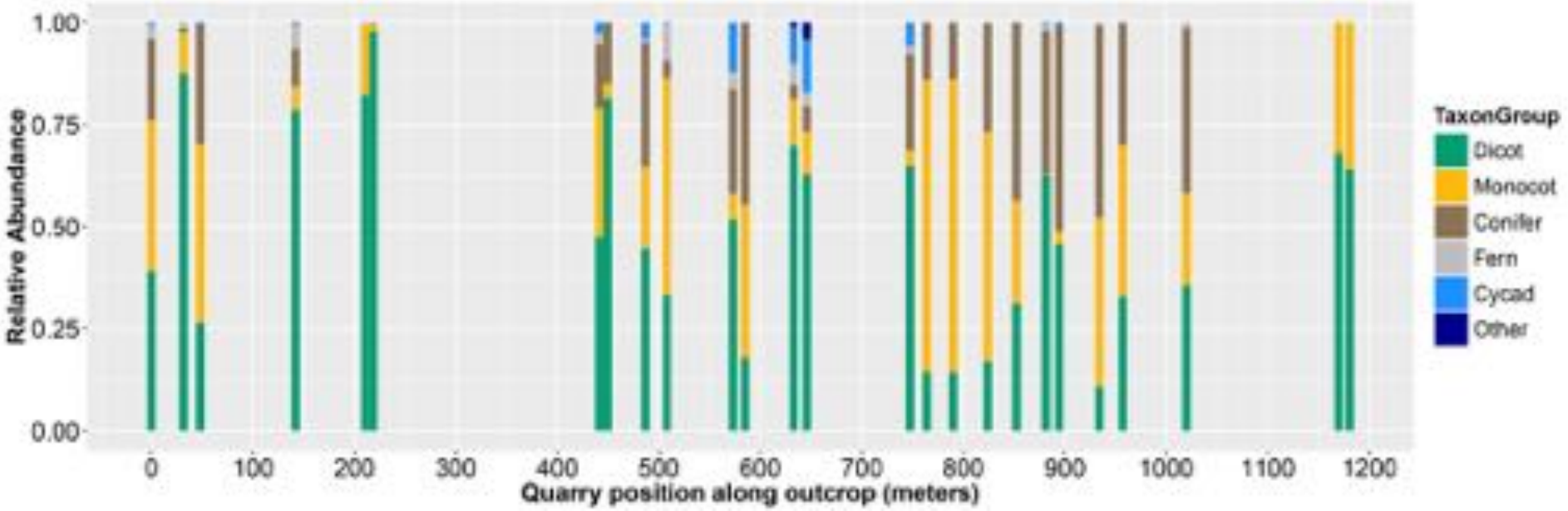
Redwood relative

tree

Zingerberopsis

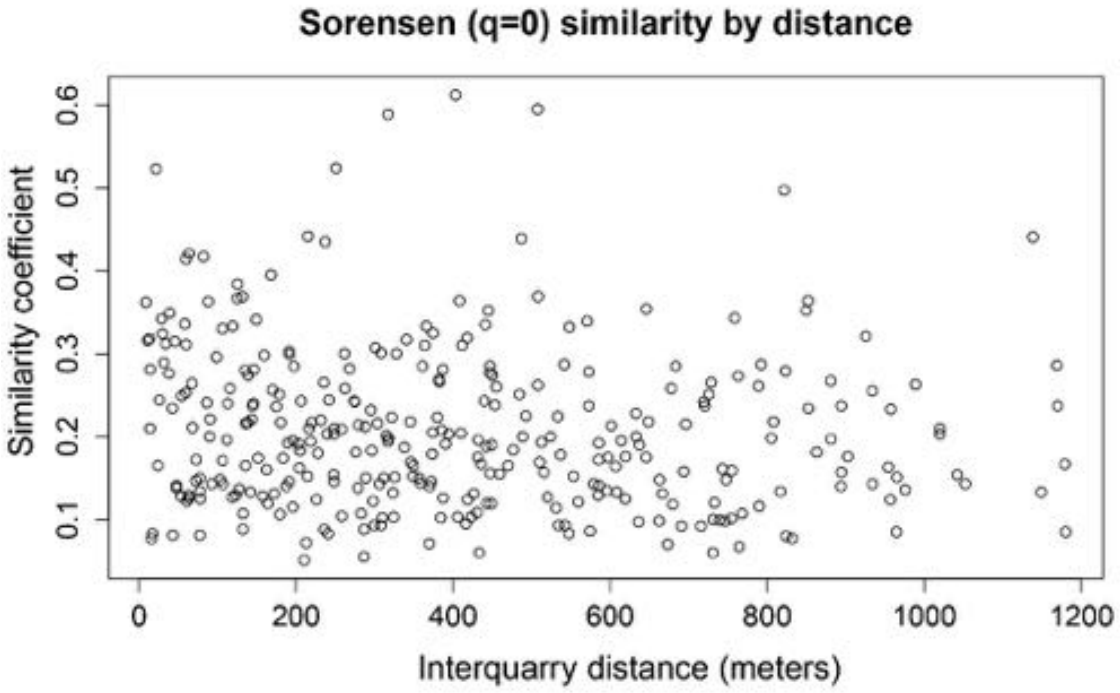
ground cover

Spatial heterogeneity



Spatial heterogeneity

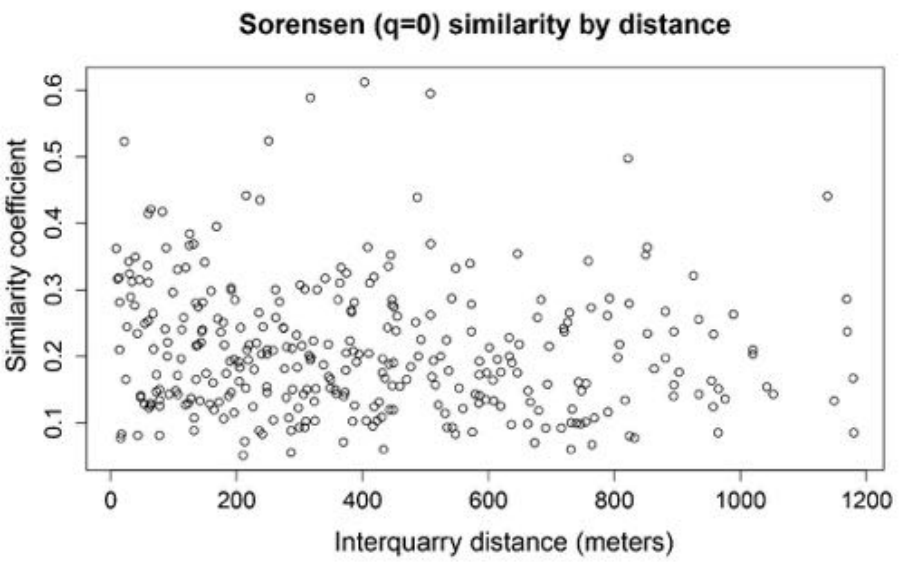
Species similarity between quarries

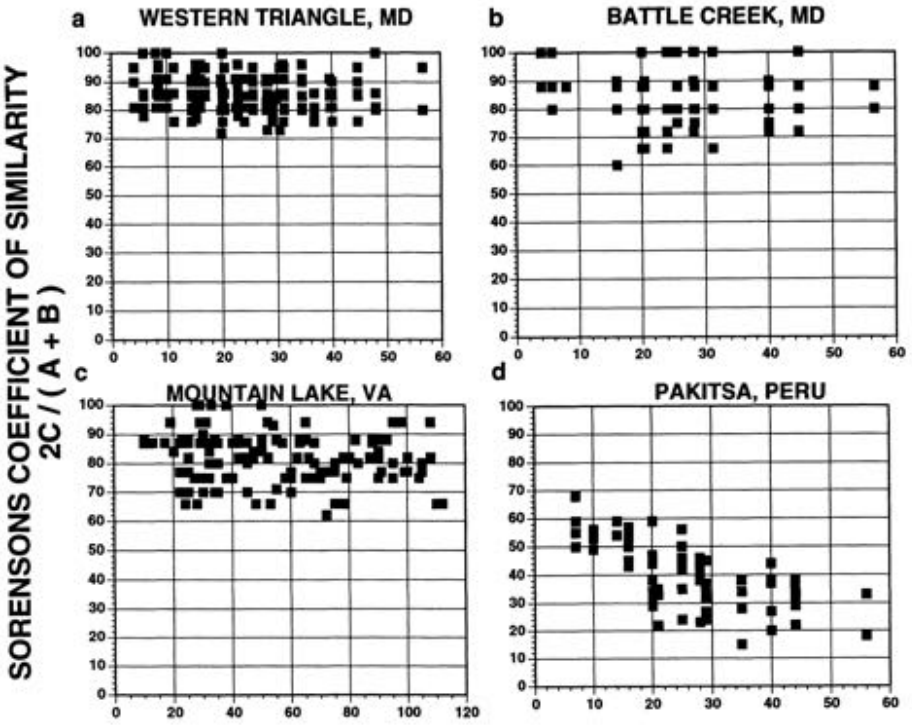


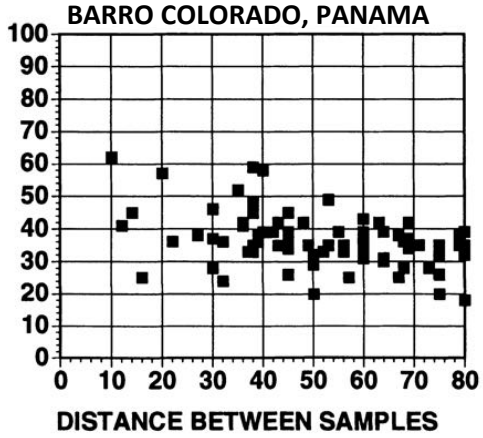
(calculated from Quarry GPS coordinates)

Spatial heterogeneity

Dori's Tuff:







From Burnham 1993

Phase 1 Interpretations:

1) The Jose Creek flora represents a hyper-diverse, spatially heterogeneous flora, with strong dominance by a redwood-like conifer, herbaceous monocot (*Zingerberopsis*), palm, & woody dicot (*Dryophyllum*).

2) Non-analog community – contains a mixture of plant groups that are no longer seen in association today (e.g., palms and redwoods)

Sabalites sp.

Dryophyllum

Redwood relative

tree

Zingerberopsis

ground cover

Project Overview:

Reconstructing functional diversity of a Cretaceous forest

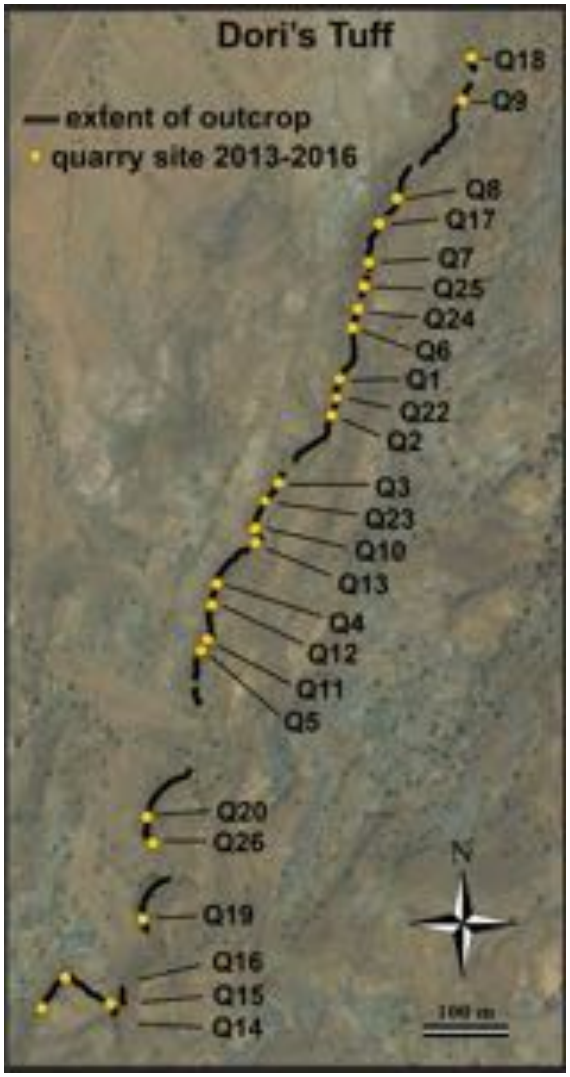
Phase 1. "Build" a forest

-describe taxonomic diversity
-relative abundance of taxa
-spatial structure of community

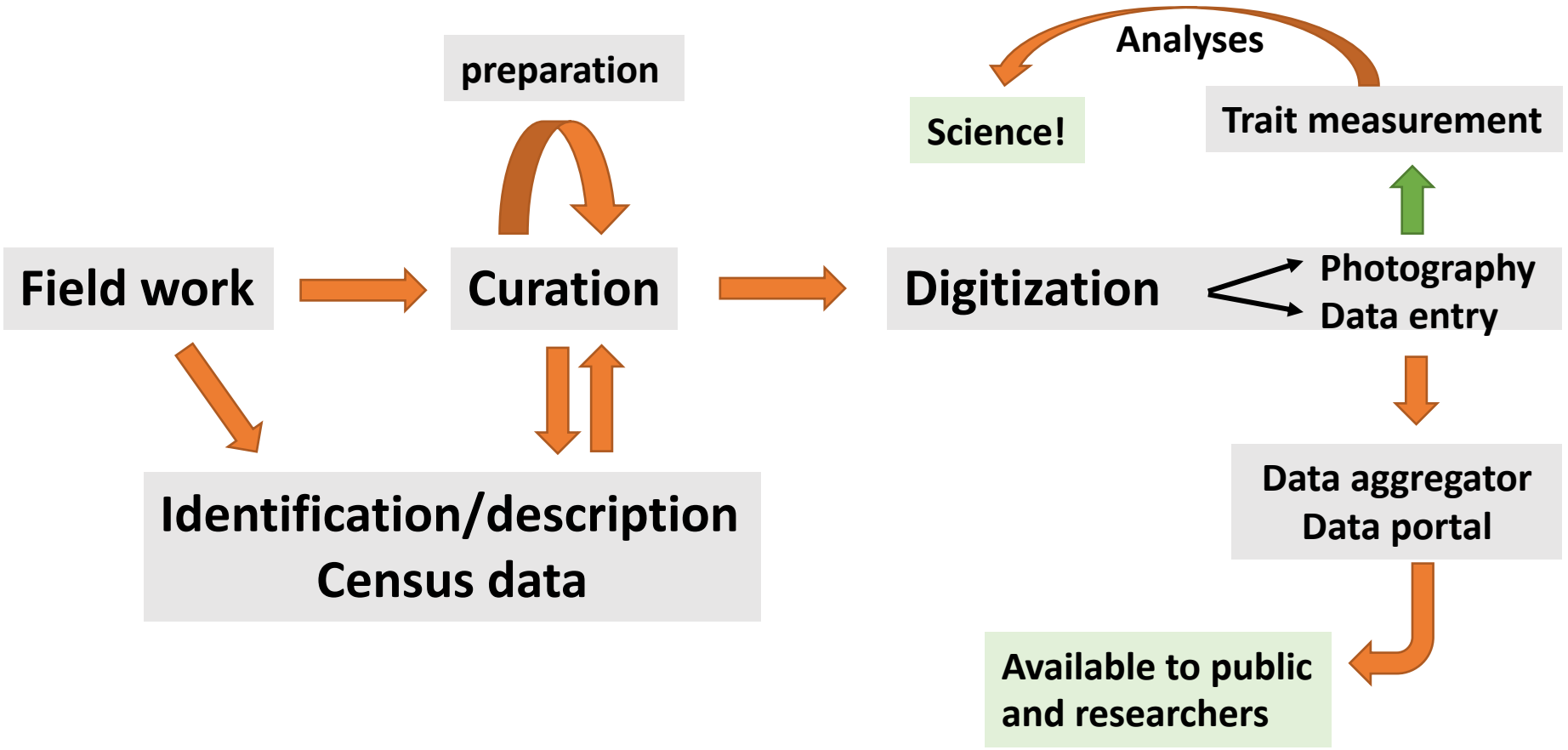
Phase 2. Measure the forest: quantify functional diversity

-measure functional leaf traits of all taxa
-reconstruct trait diversity across transect

*quantitative and spatial explicit sampling scheme
*large sample sizes
***digital measurements of georeferenced specimens**

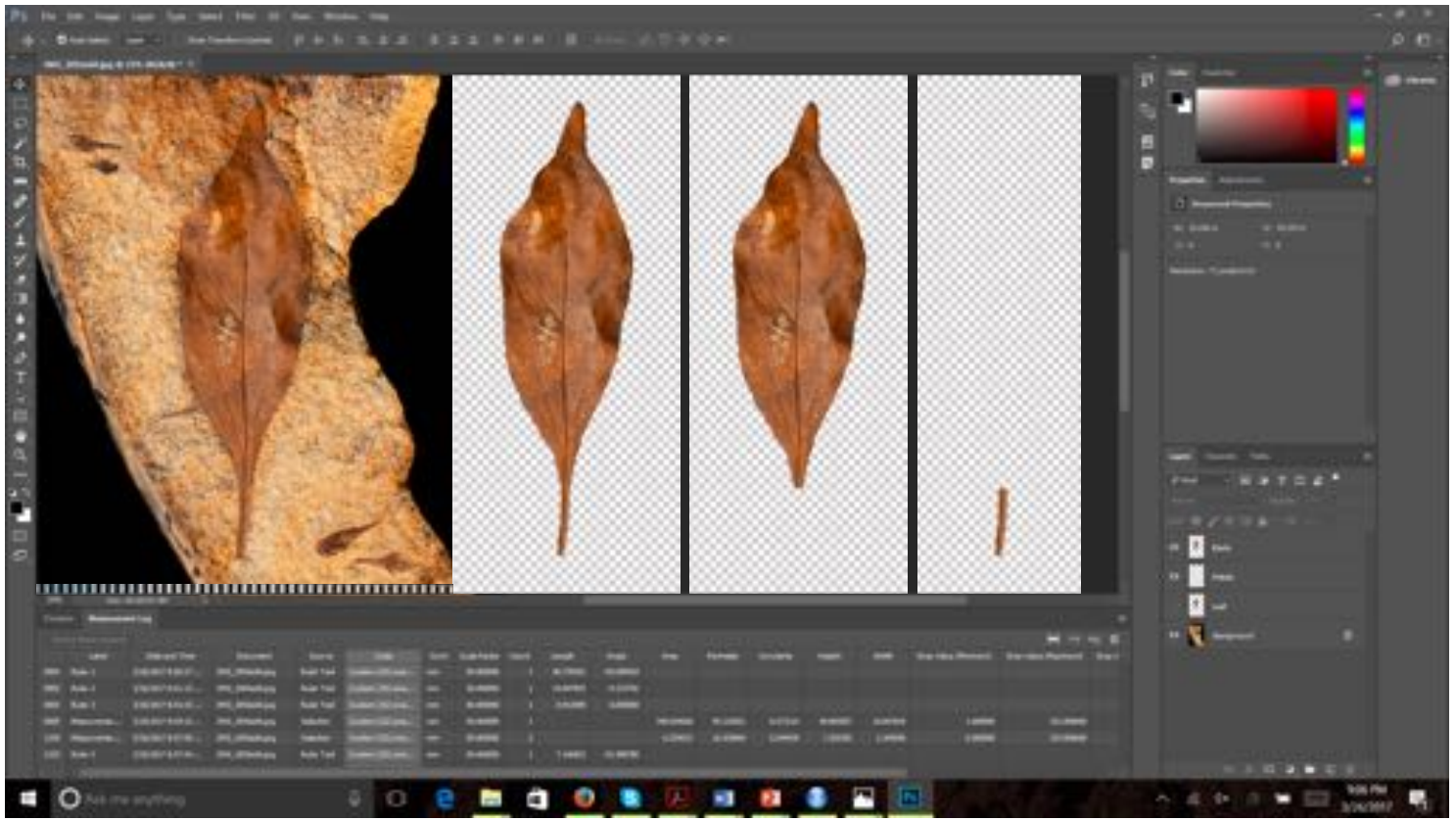


Using the digital data



Digital Trait Measurements

Digitized specimens piped into Adobe Photoshop for image analysis:



Leaf length
Leaf width
Tooth spacing

Leaf area
Leaf perimeter
Specific leaf area

Petiole length
Petiole width
Petiole area

Summary

- Workflow should simultaneously meet needs of research project and museum long term archival and digitization
- Assign rocks unique ID numbers in the field to bridge field data with museum data
- Increase efficiency with integration of tasks and customized workstation set-up
- Person-power! Importance of involving students and volunteers in research

Acknowledgements

LOOY LAB
Paleobotany, Palynology & Paleoecology
at University of California, Berkeley

UCMP

IDigBio and NSF

Amazing students and volunteers!

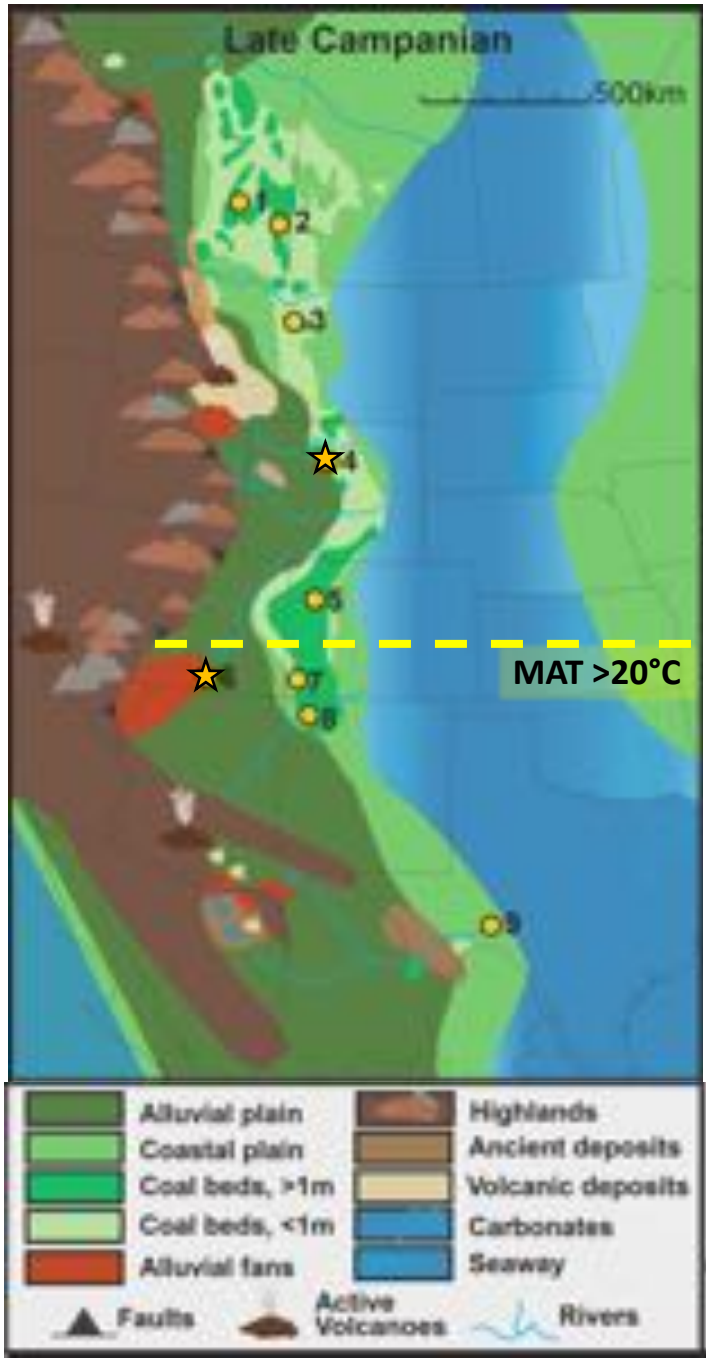
Looy Lab
UC Museum of Paleontology
Armendaris Ranch, Tom Waddell, Ted Turner

Diane Erwin, Cindy Looy, Pat Holroyd, Gary Upchurch, Greg Mack

Funding:

Lewis and Clark Fund for Exploration and Field Research (2015); UCMP Graduate Student Awards (2013, 2014, & 2015); Integrative Biology Graduate Research Fund: The Reshetko Family Scholarship (2015); Geological Society of America Graduate Student Research Grant (2014); Integrative Biology Graduate Research Fund (2014 & 2015); Sigma Xi Grants-in-Aid of Research UC-Berkeley Chapter (2014); Mid-American Paleontological Society (MAPS) Outstanding Student Research Award (2013); GRAC Research Funds, UC-Berkeley Integrative Biology Dept. (2013 & 2015); UCMP Graduate Student Fellowship; NSF Graduate Research Fellowship

TEXAS STATE BIOLOGY



Latitudinal gradient

Late Campanian Western Interior

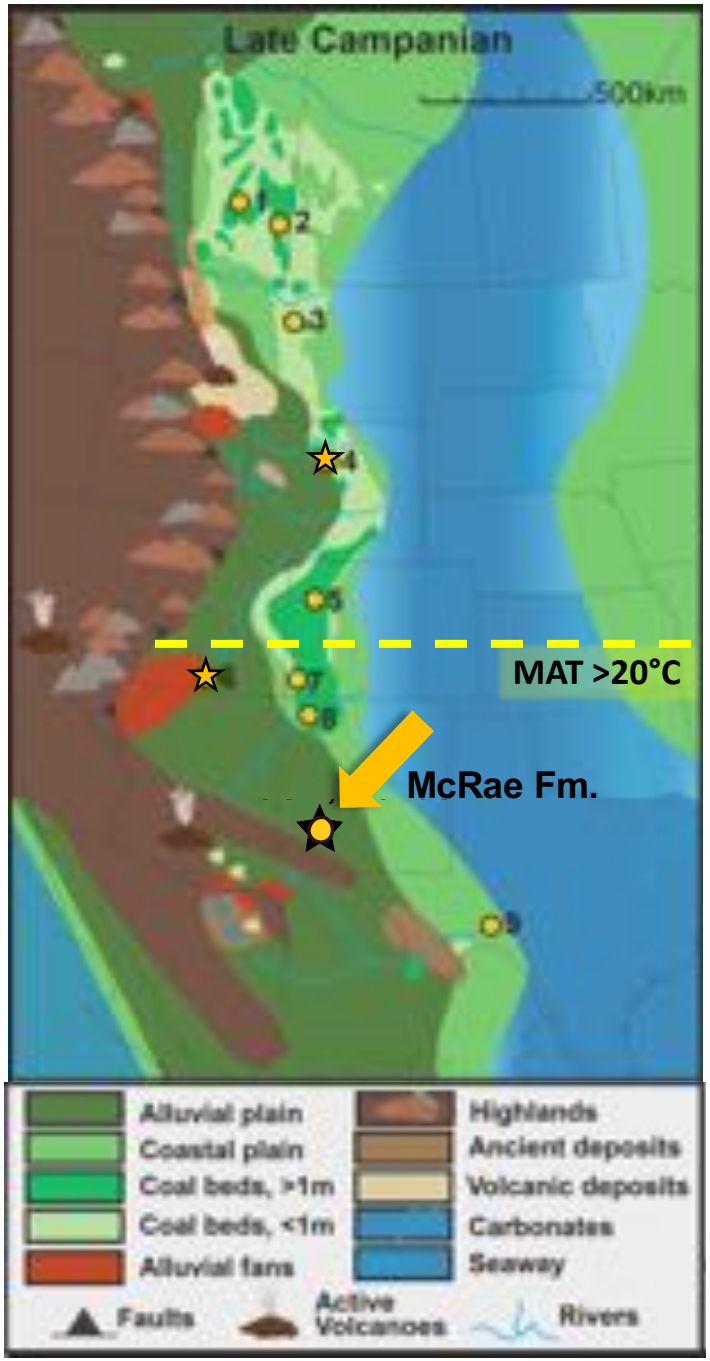
Gradient in angiosperm diversity and taxonomic composition of floras

- More prevalent at lower latitudes

Few quantitative landscape reconstructions > community structure

- Big Cedar Ridge (Wing et al. 2012)
- Kaiparowits (Miller et al. 2013)

Paleogeographic reconstruction redrawn from Roberts and Kirschbaum 1995;
Fossil floras plotted from Miller et al. 2013



Latitudinal gradient

Late Campanian Western Interior

Gradient in angiosperm diversity and taxonomic composition of floras

- More prevalent at lower latitudes

Few quantitative landscape reconstructions > community structure

- Big Cedar Ridge (Wing et al. 2012)
- Kaiparowits (Miller et al. 2013)

McRae Formation

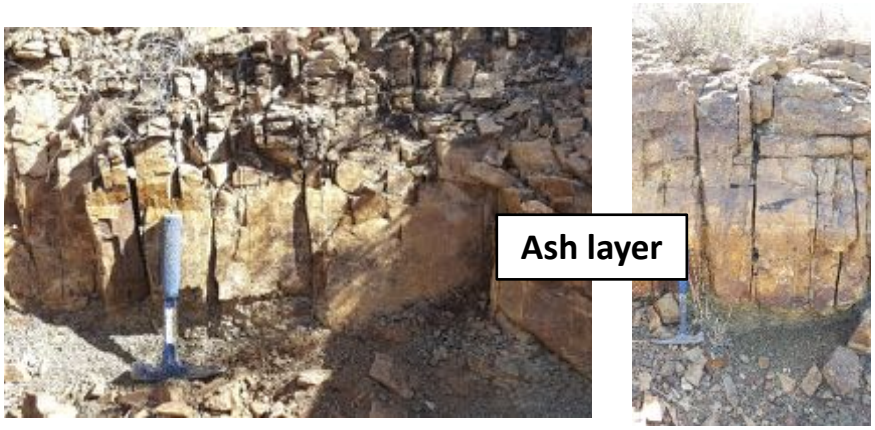
- **Alluvial Plain, terra firma**
- **>200 km inland, Southern NM**

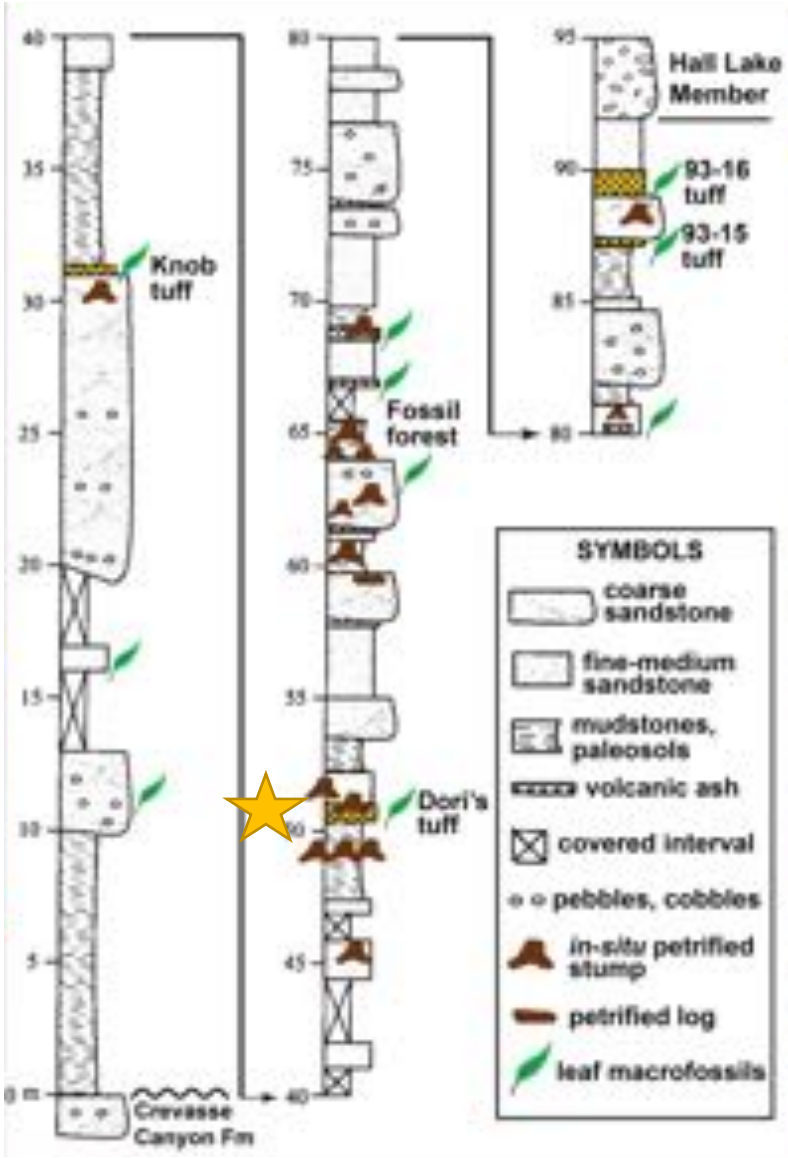
Paleogeographic reconstruction redrawn from Roberts and Kirschbaum 1995;
Fossil floras plotted from Miller et al. 2013

McRae Fm: Jose Creek Member

“Dori’s tuff”

- Ashfall bed with little to no transport
- Single depositional event on stable floodplain
- 74.7 ± 0.6 Ma (U-Pb, Amato et al. accepted)

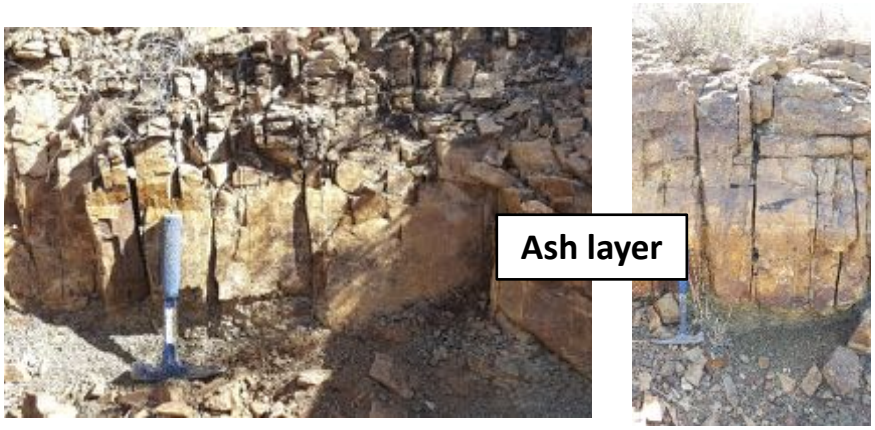




McRae Fm: Jose Creek Member

“Dori’s tuff”

- Ashfall bed with little to no transport
- Single depositional event on stable floodplain
- 74.7 ± 0.6 Ma (U-Pb, Amato et al. accepted)



Project Overview:

Reconstructing structure & functional diversity of a Cretaceous forest

<u>Phase 1.</u> "Build" a forest

-describe taxonomic diversity
-relative abundance of taxa
-spatial structure of community

<u>Phase 2.</u> Measure the forest: quantify functional diversity

-measure functional leaf traits of all taxa
-reconstruct trait diversity across transect

*quantitative and spatial explicit sampling scheme
*large sample sizes
*digital measurements of georeferenced specimens

Phase 1 Conclusions (thus far)

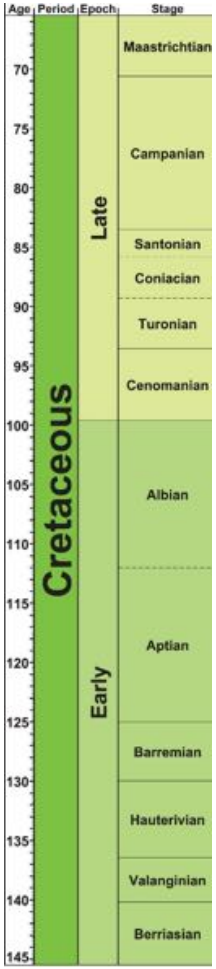
1) **It is among the earliest leaf macrofloras shown to be dominated by angiosperms by both species richness AND relative abundance across landscape (on terra firma!)**

2) **One of the most diverse leaf macrofloras known**

3) **Early glimpse into angio-dominant forest in warm-wet climate ("tropical"), with evergreen conifers still playing a prominent role in the canopy.**

4) **Suggests interior vegetation may show different trends than coastal/swamps.**

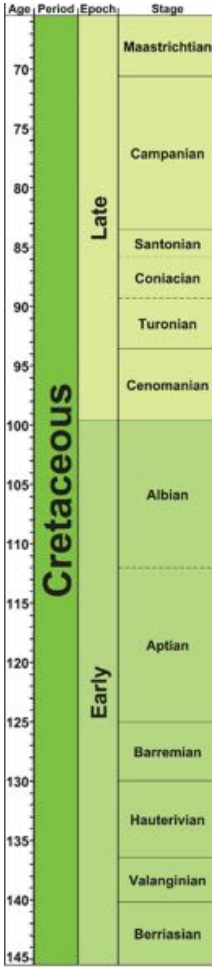
Timing of angiosperm ecological expansion and the restructuring of plant communities



high species richness in floras

appear in pollen record

Timing of angiosperm ecological expansion and the restructuring of plant communities

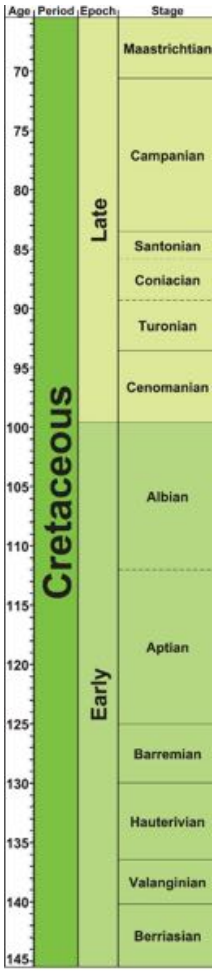


widespread dominance by abundance

high species richness in floras

appear in pollen record

Timing of angiosperm ecological expansion and the restructuring of plant communities



widespread dominance by abundance

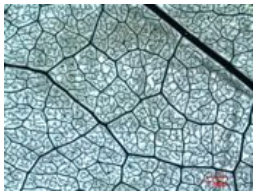
high species richness in floras

appear in pollen record

"ecological lag"?

presence in canopy and diversity of growth habits

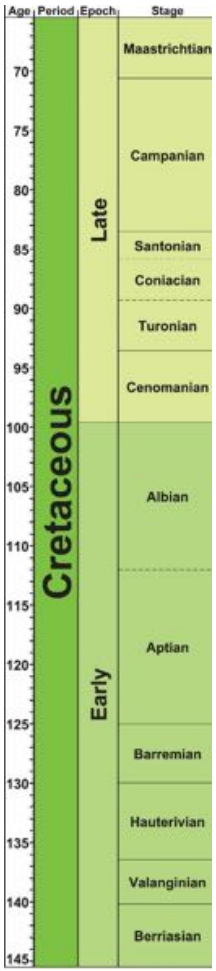
e.g. Crifo et al. 2014



ecophysiological traits

e.g. Field et al. 2011

Timing of angiosperm ecological expansion and the restructuring of plant communities



widespread dominance by abundance

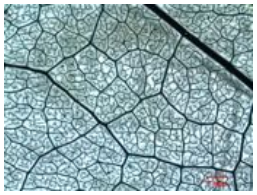
high species richness in floras

appear in pollen record

"ecological lag"?

presence in canopy and diversity of growth habits

e.g. Crifo et al. 2014



ecophysiological traits

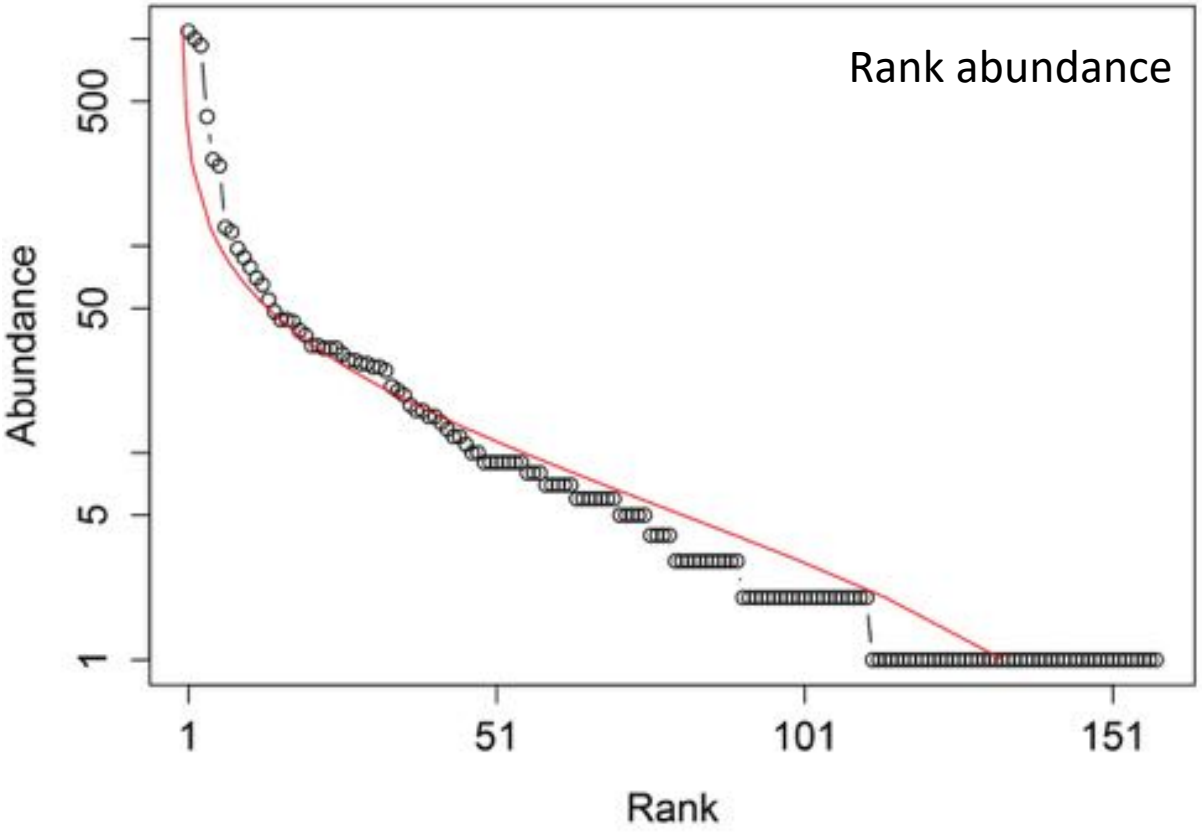
e.g. Field et al. 2011

<u>Unanswered questions:</u>

- Geographic and environmental patterns of angiosperm dominance by abundance
- Composition and structure of Late Cretaceous plant communities
- Functional diversity of angiosperms and communities

Community structure

Evenness



Most species richness is due to rare species

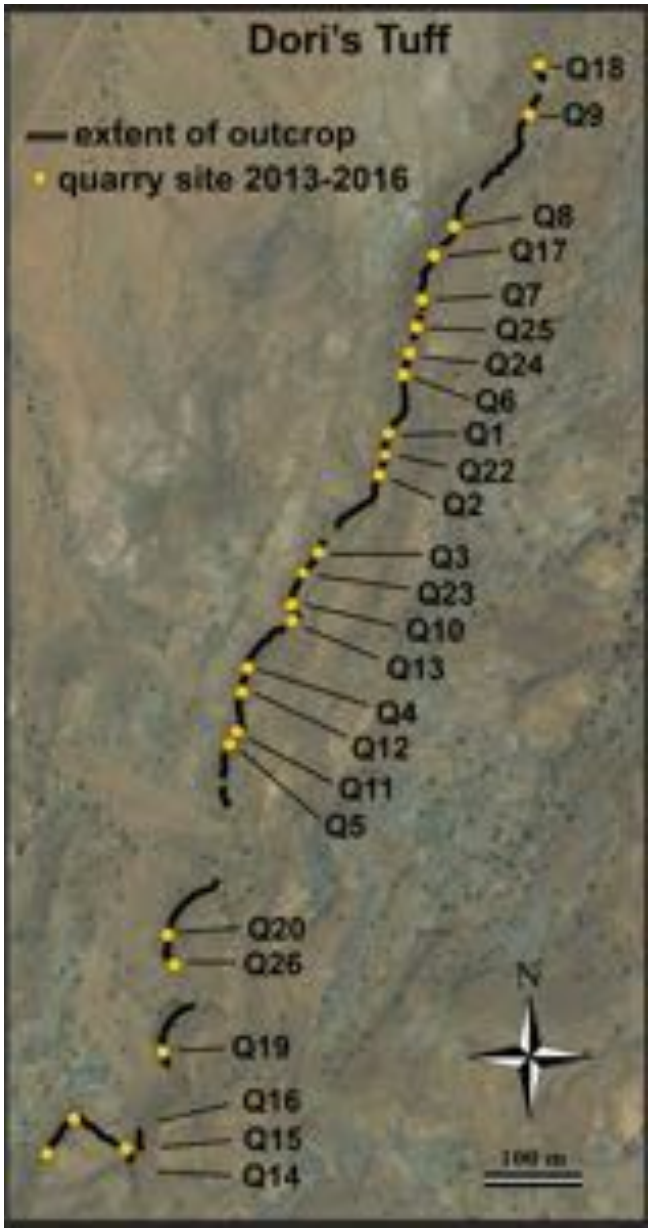
Spatial heterogeneity

Additive partitioning of diversity (richness):

(following Chao et al. 2012; Jost 2007; Lande 1996)

Total = within quarry + between quarry

$$\gamma = \alpha + \beta$$



Spatial heterogeneity

Additive partitioning of diversity (richness):

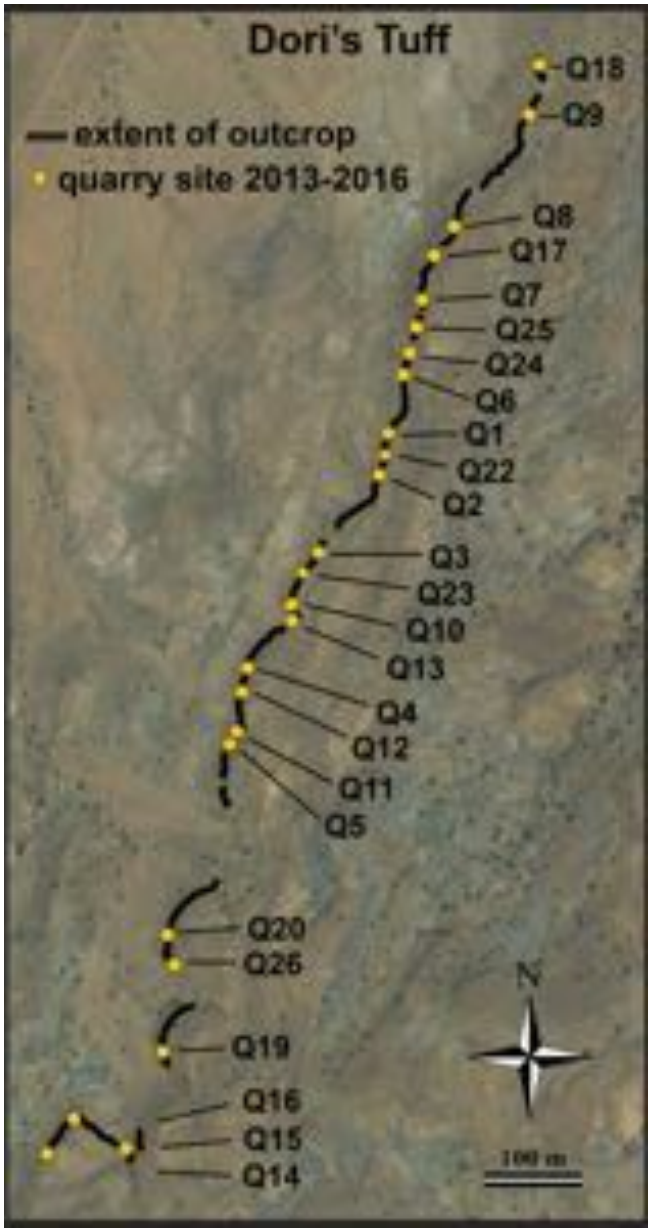
(following Chao et al. 2012; Jost 2007; Lande 1996)

Total = within quarry + between quarry

$$\gamma = \alpha + \beta$$

$$158 = 18 + 140$$

High species turnover across deposit



Spatial heterogeneity

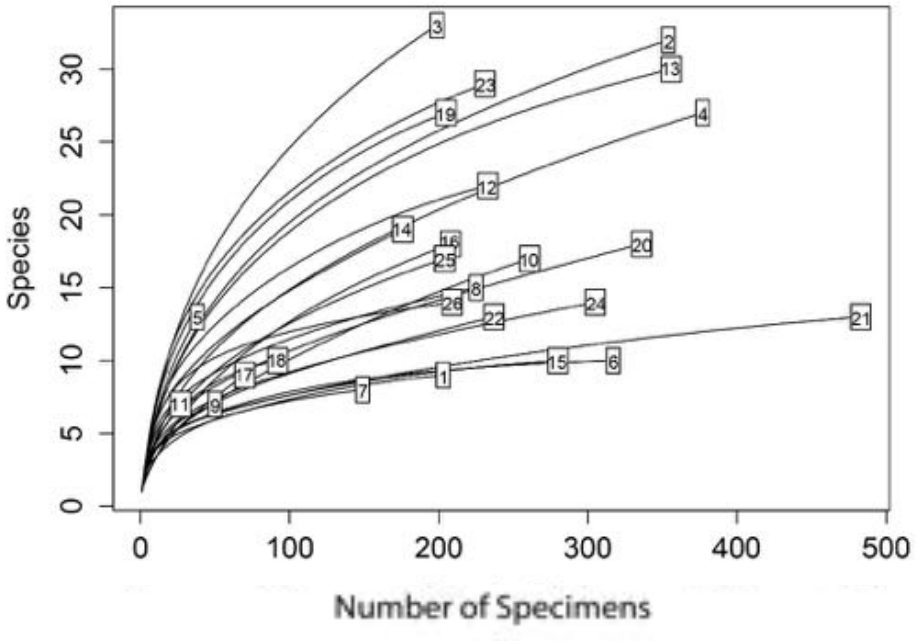
Additive partitioning of diversity (richness):

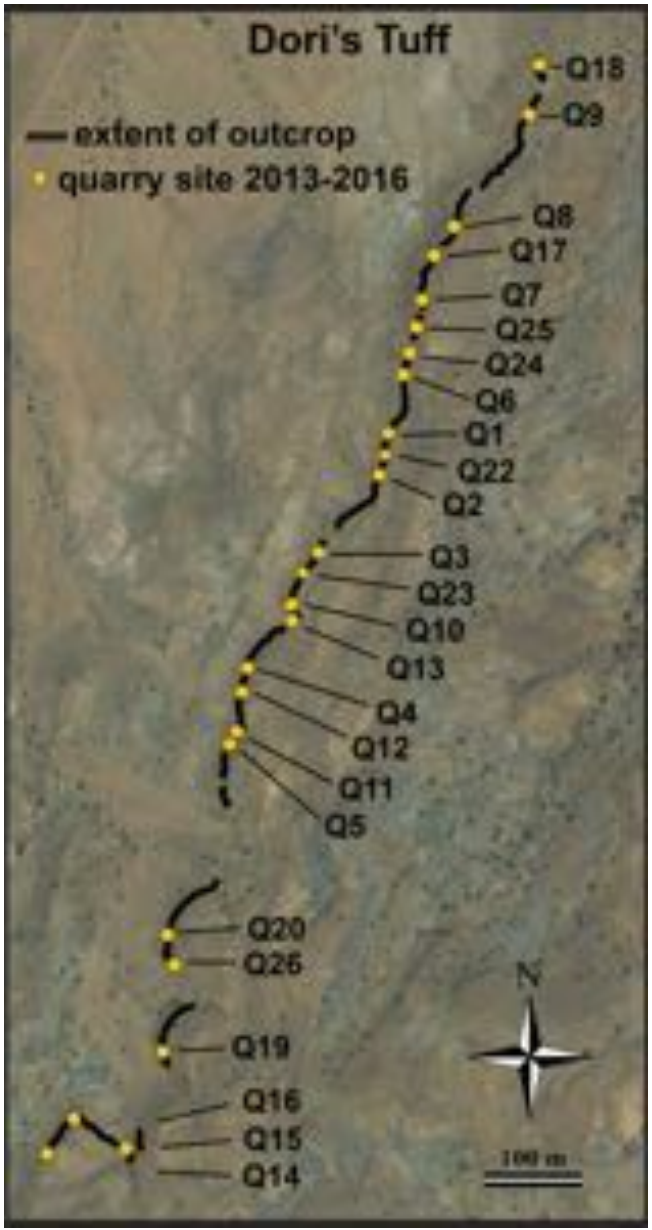
(following Chao et al. 2012; Jost 2007; Lande 1996)

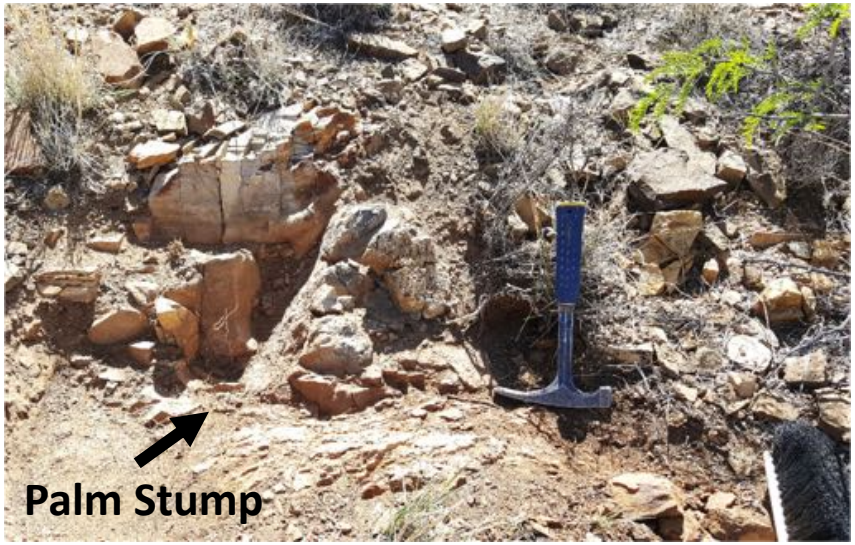
Total = within quarry + between quarry

$$\gamma = \alpha + \beta$$

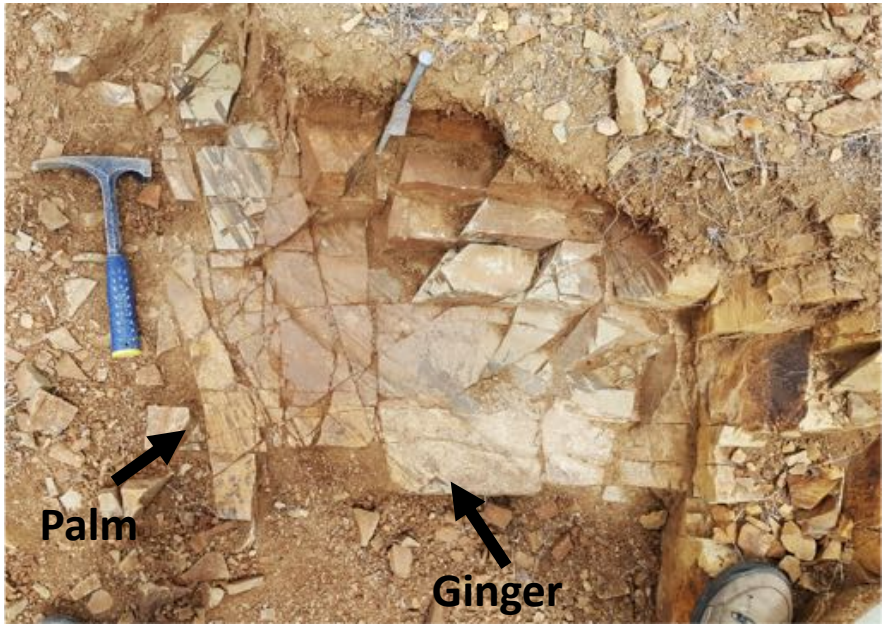
$$158 = 18 + 140$$





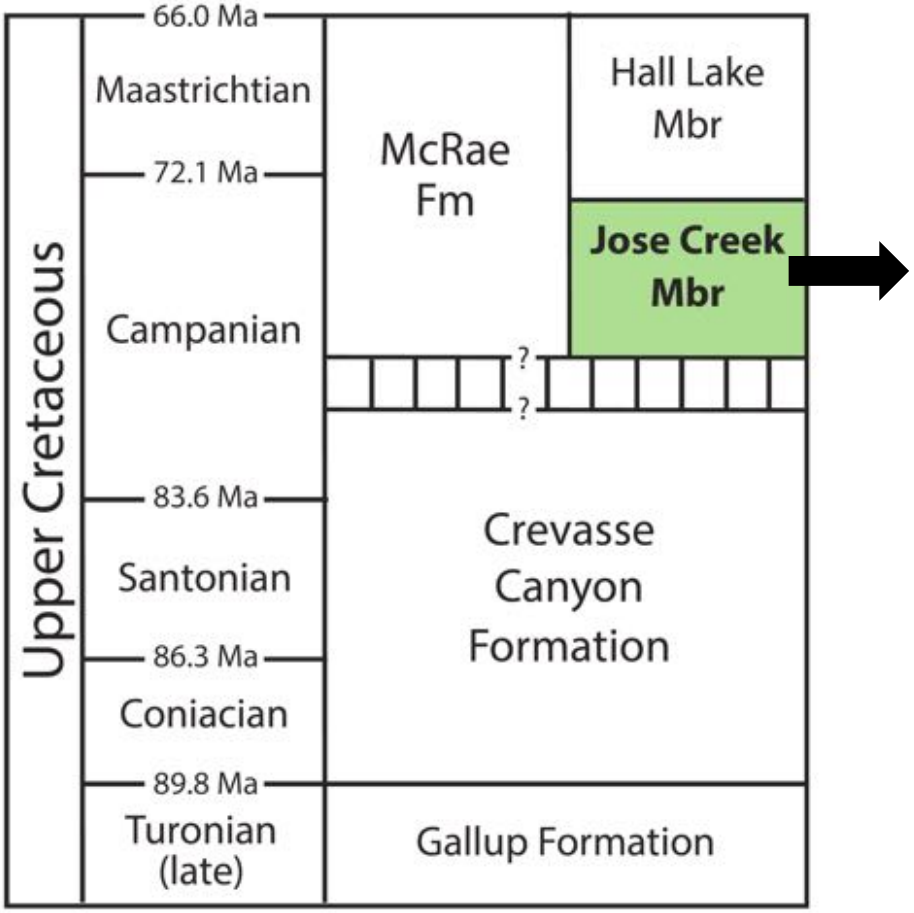

Palm Stump

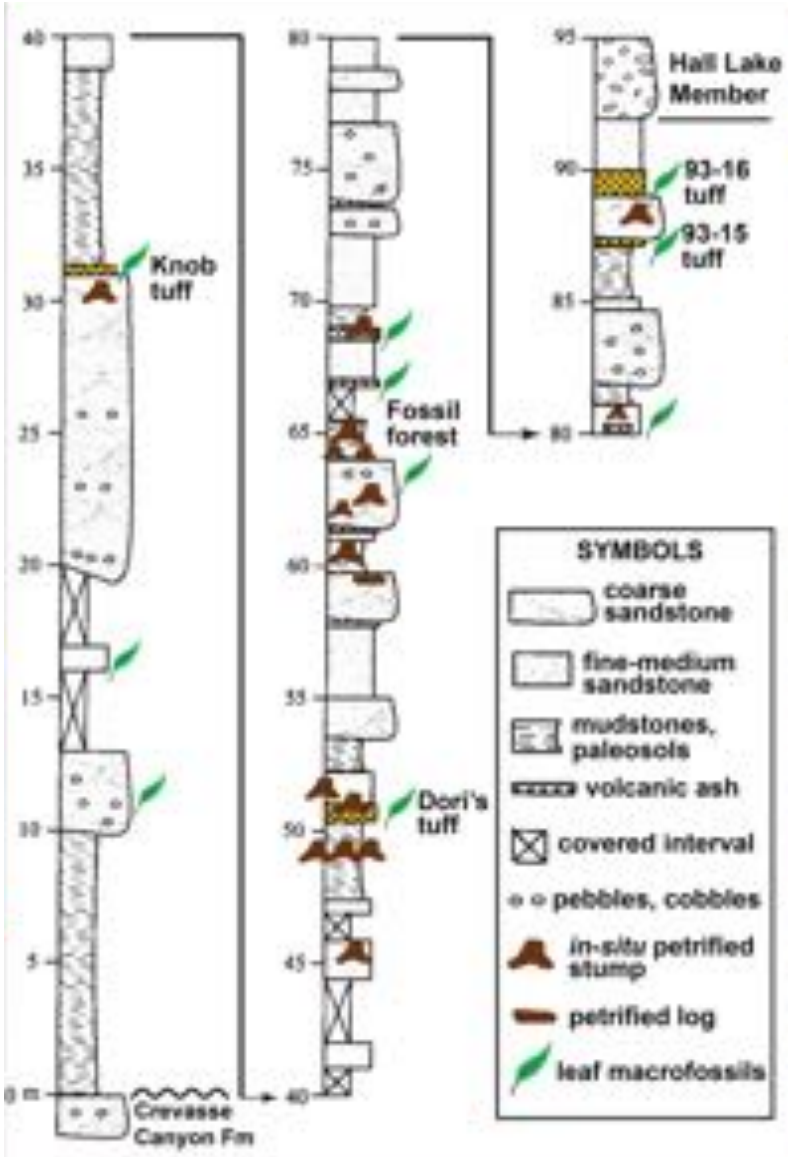

Ginger


Palm
Ginger

McRae Fm: Jose Creek Member

- Fluvial deposition on alluvial plain





Project Overview:

Reconstructing structure & functional diversity of a Cretaceous forest

Phase 1. "Build" a forest

-describe taxonomic diversity
-relative abundance of taxa
-spatial structure of community

Phase 2. Measure the forest: quantify functional diversity

-measure functional leaf traits of all taxa
-reconstruct trait diversity across transect

*quantitative and spatial explicit sampling scheme
*large sample sizes
*digital measurements of georeferenced specimens