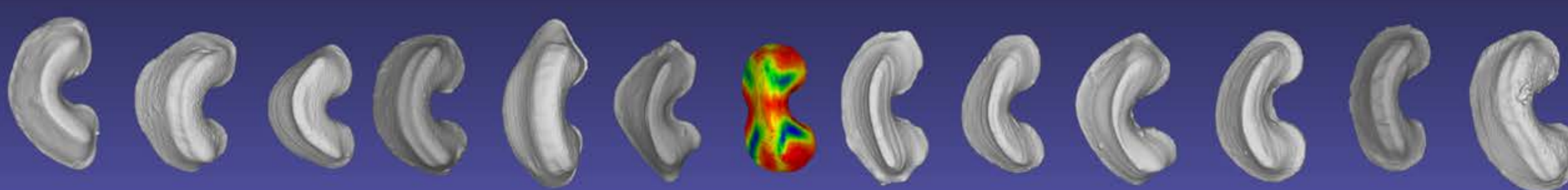
Shedding light on the evolution of grasses and grasslands through automated, quantitative imaging analyses of plant silica microfossils

Caroline A.E. Strömberg
Department of Biology & Burke Museum
University of Washington
Seattle, WA

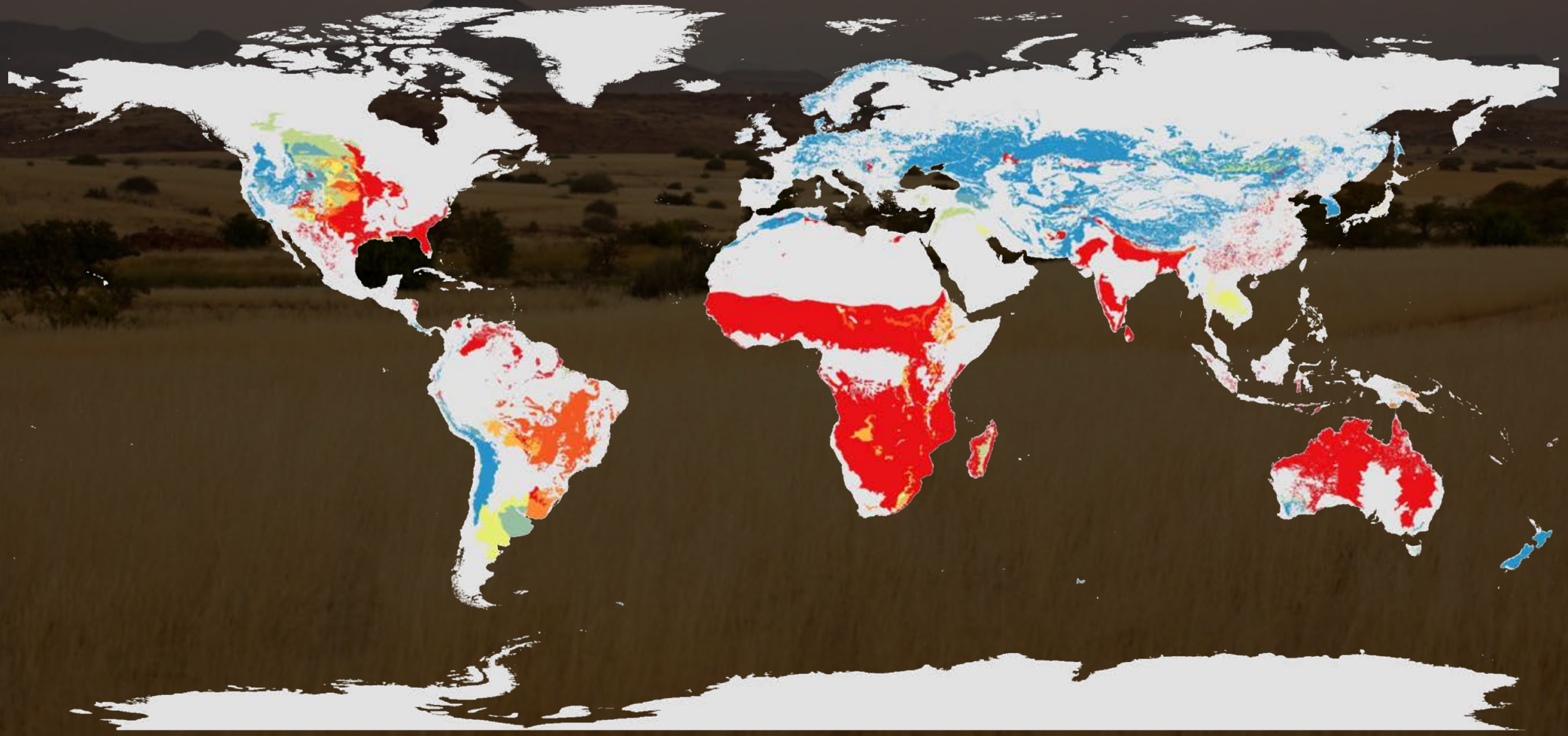


iDigBio Integrated Digitized Biocollections

Burke Museum University of Washington

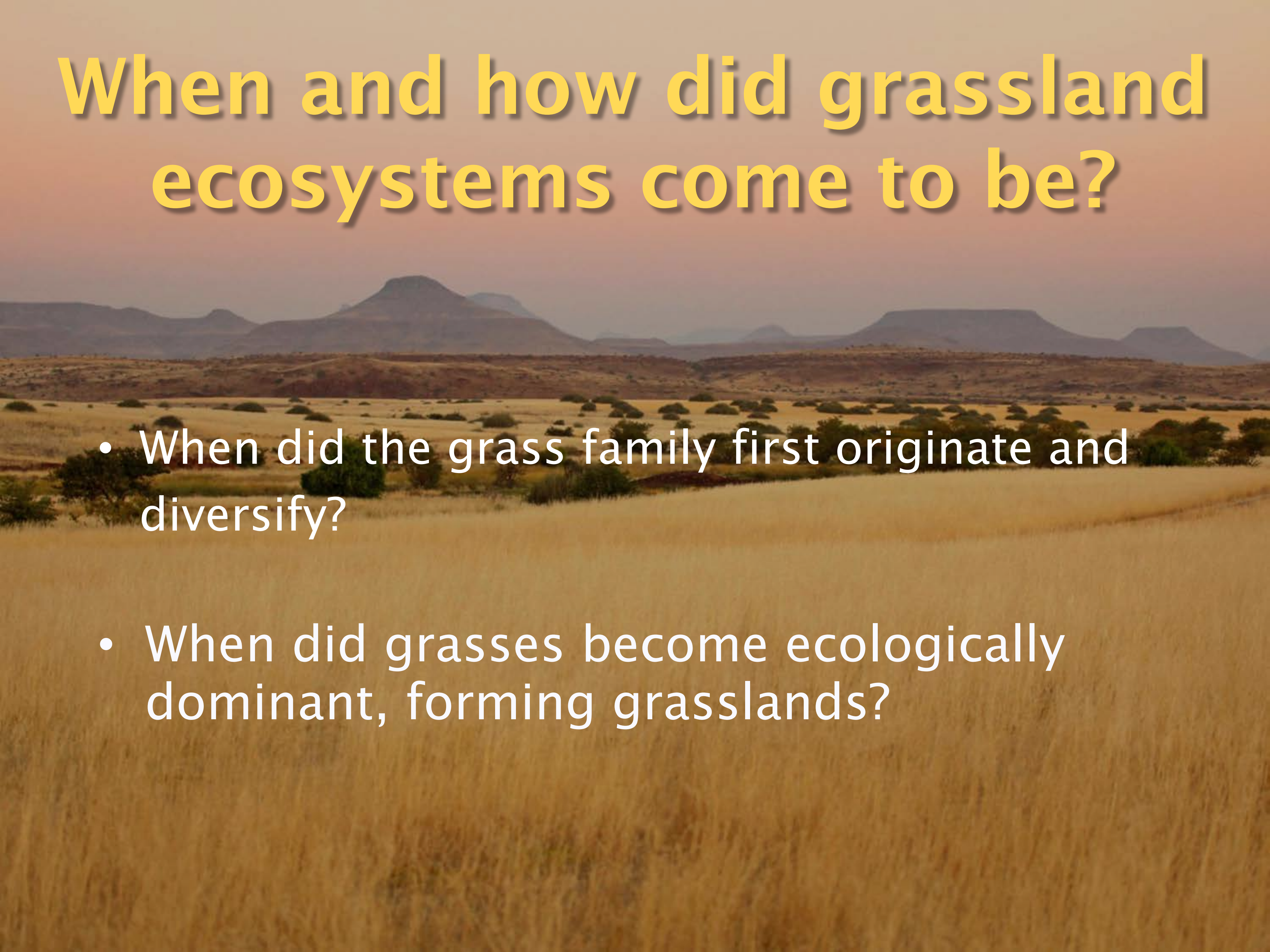
Grasslands are ecologically vital

- Grassy biomes make up >40% of Earth's land surface



Color = grass-dominated habitats

Lehmann et al. (in prep.)



When and how did grassland ecosystems come to be?

- When did the grass family first originate and diversify?
- When did grasses become ecologically dominant, forming grasslands?

Direct evidence for past grasslands

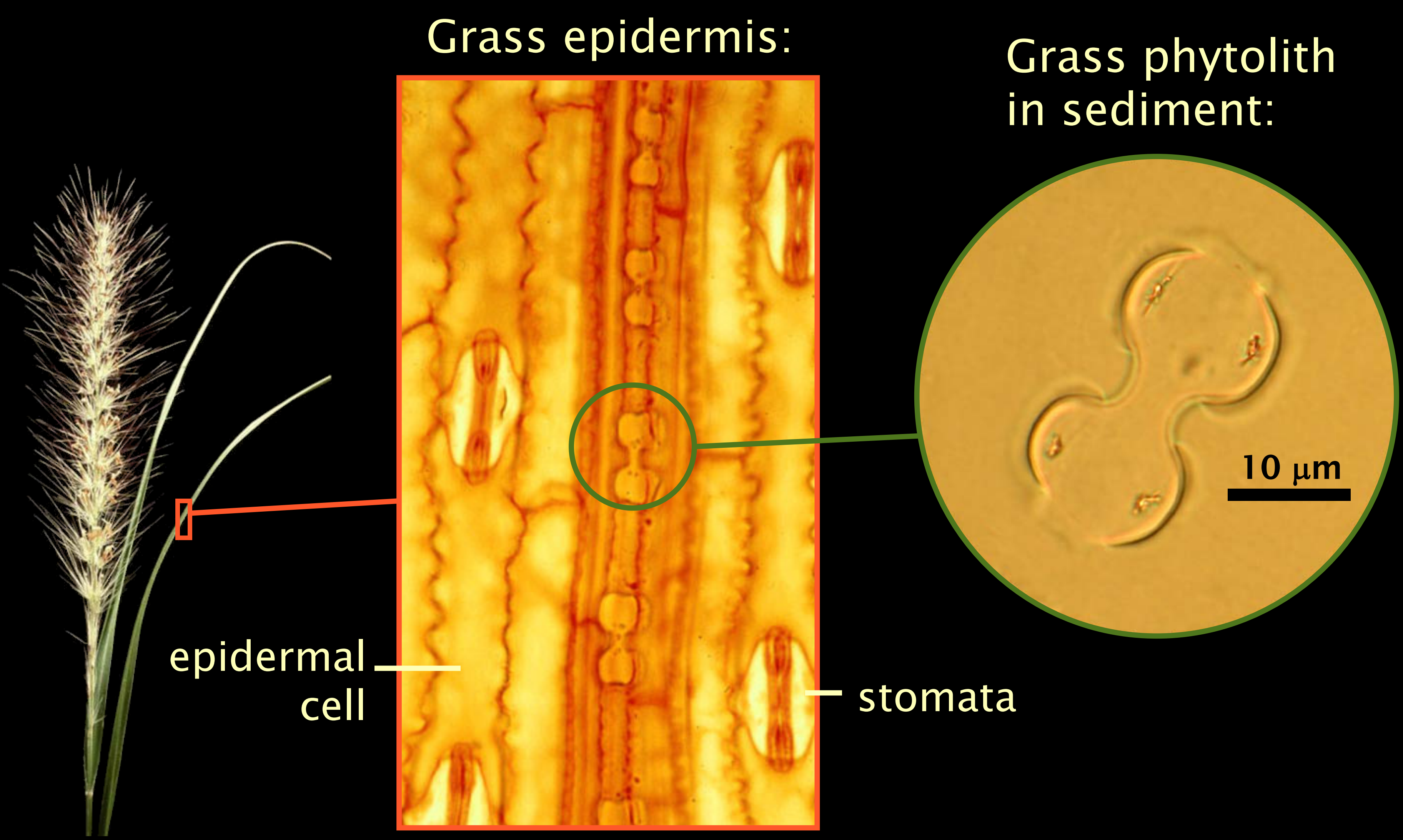
- Grass mesofossils and pollen are rare until the late Miocene—and often hard to interpret taxonomically

Pollen

Leaves

Fruit

Phytoliths *(plant silica)*



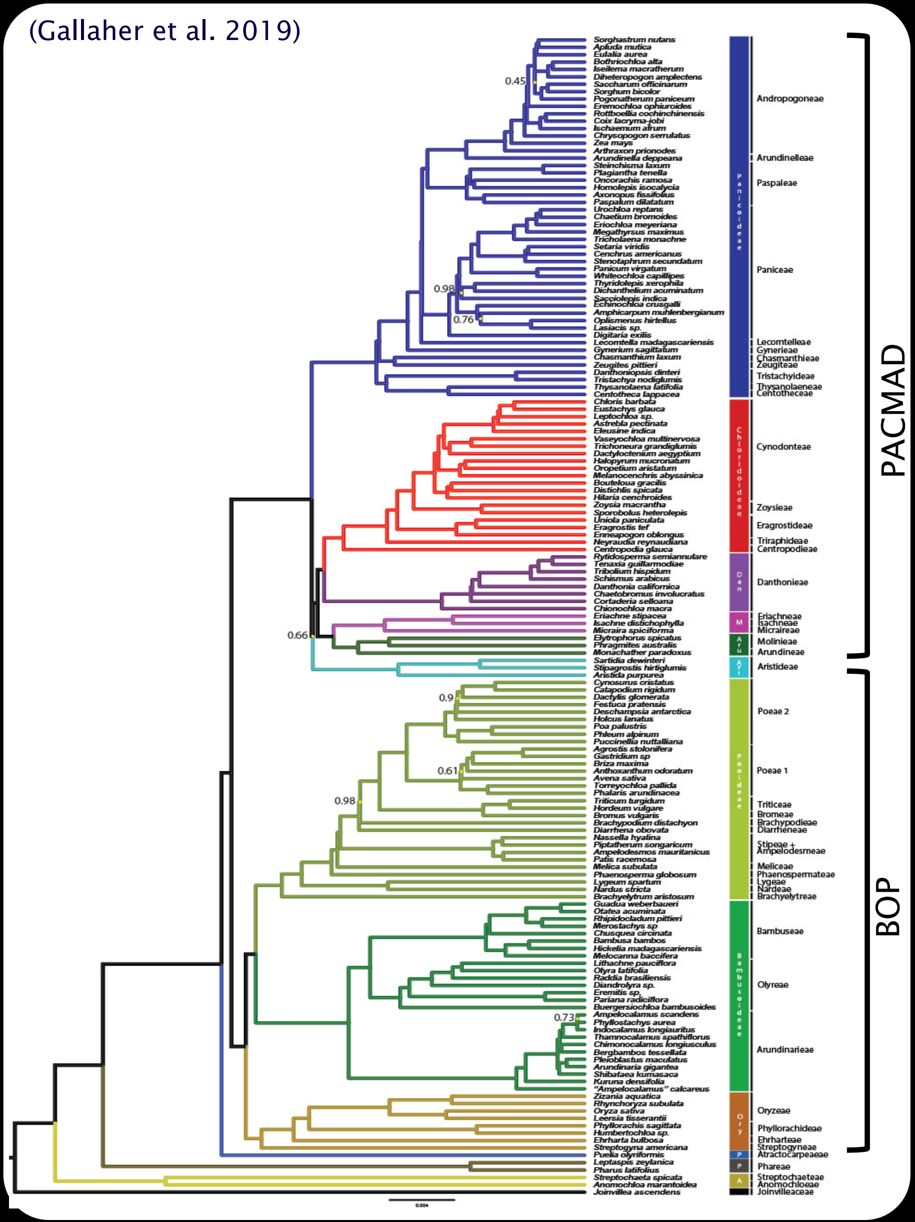
Phytoliths *(plant silica)*

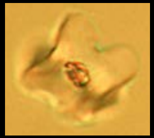
- Taxonomically useful variation within Poaceae (grass family)

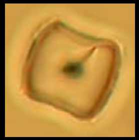
PACMAD:
Panicoideae
Arundinoideae
Chloridoideae
Micrairoideae
Aristidoideae
Danthonioideae

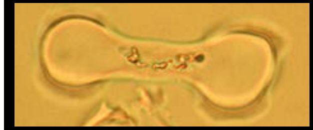
BOP:
Bambusoideae
Oryzoideae
Pooideae

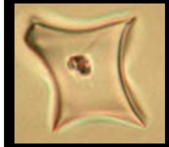
Early-diverging grasses





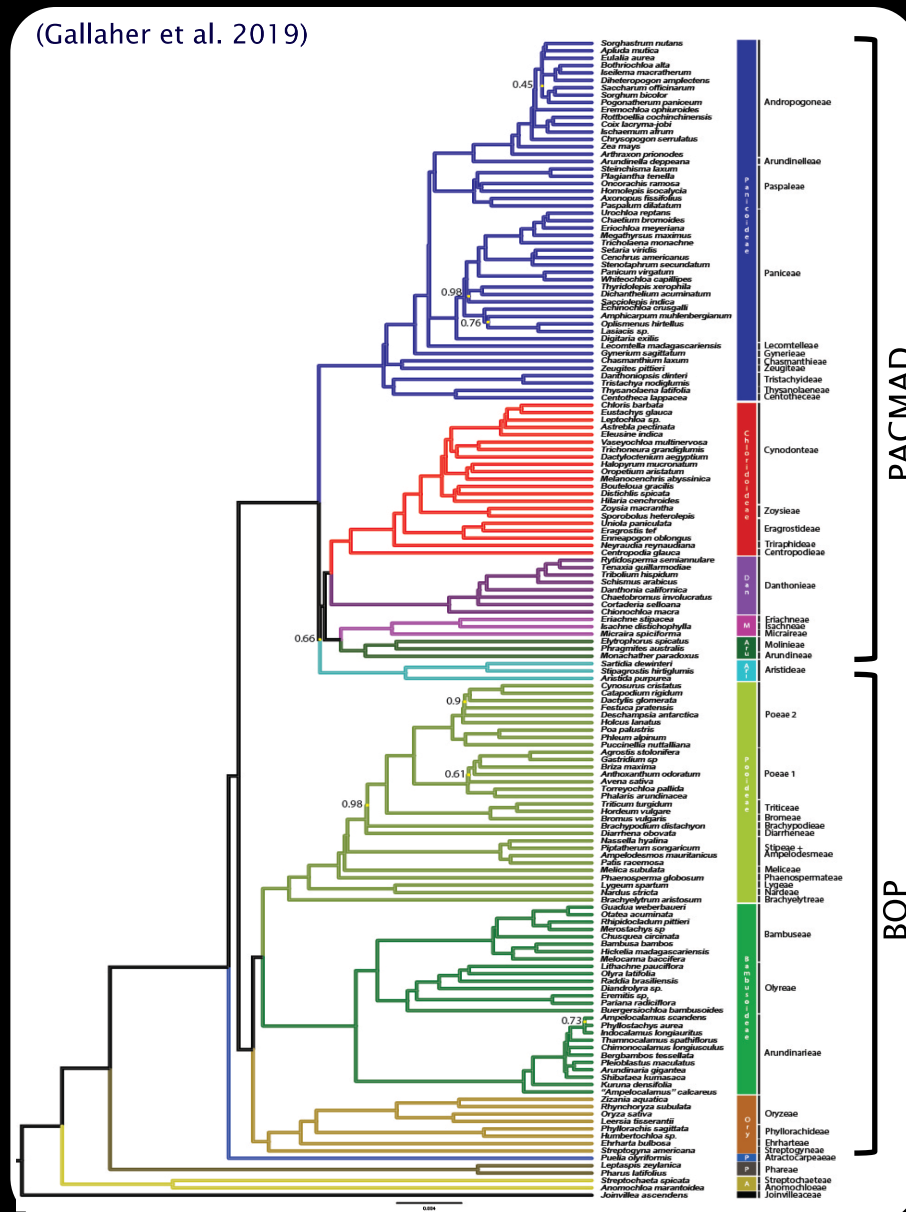






Phytoliths *(plant silica)*

- Taxonomically useful variation within Poaceae (grass family):
  - o Diversification of ancient grass lineages
  - o Ecology of past grass communities



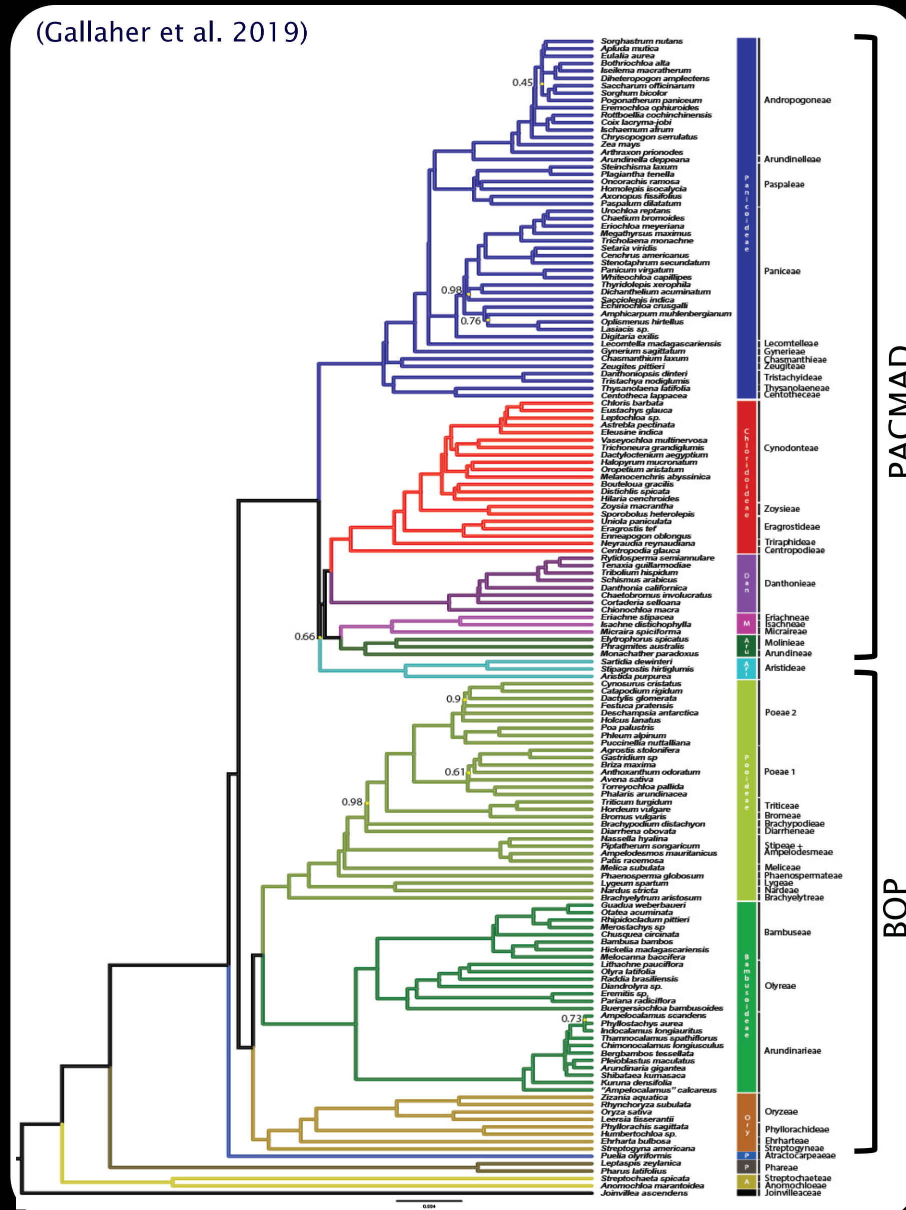
OPEN habitat (tropical)

OPEN habitat (temperate)

CLOSED habitat

Radiation of open-habitat grasses

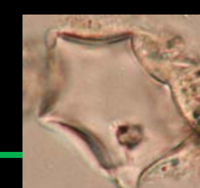
- Fossil phytolith morphotypes (Americas, Eurasia):
  - → Open-habitat grasses diversified by 40 Ma



40 Ma

19 Ma

40 Ma



35 Ma

(Strömberg 2004, 2005, Strömberg et al. 2007, 2013, Miller et al. 2012)

Grassland evolution in North America

- Earliest (early Miocene) grasslands were dominated by cool-temperate stipoid pooids

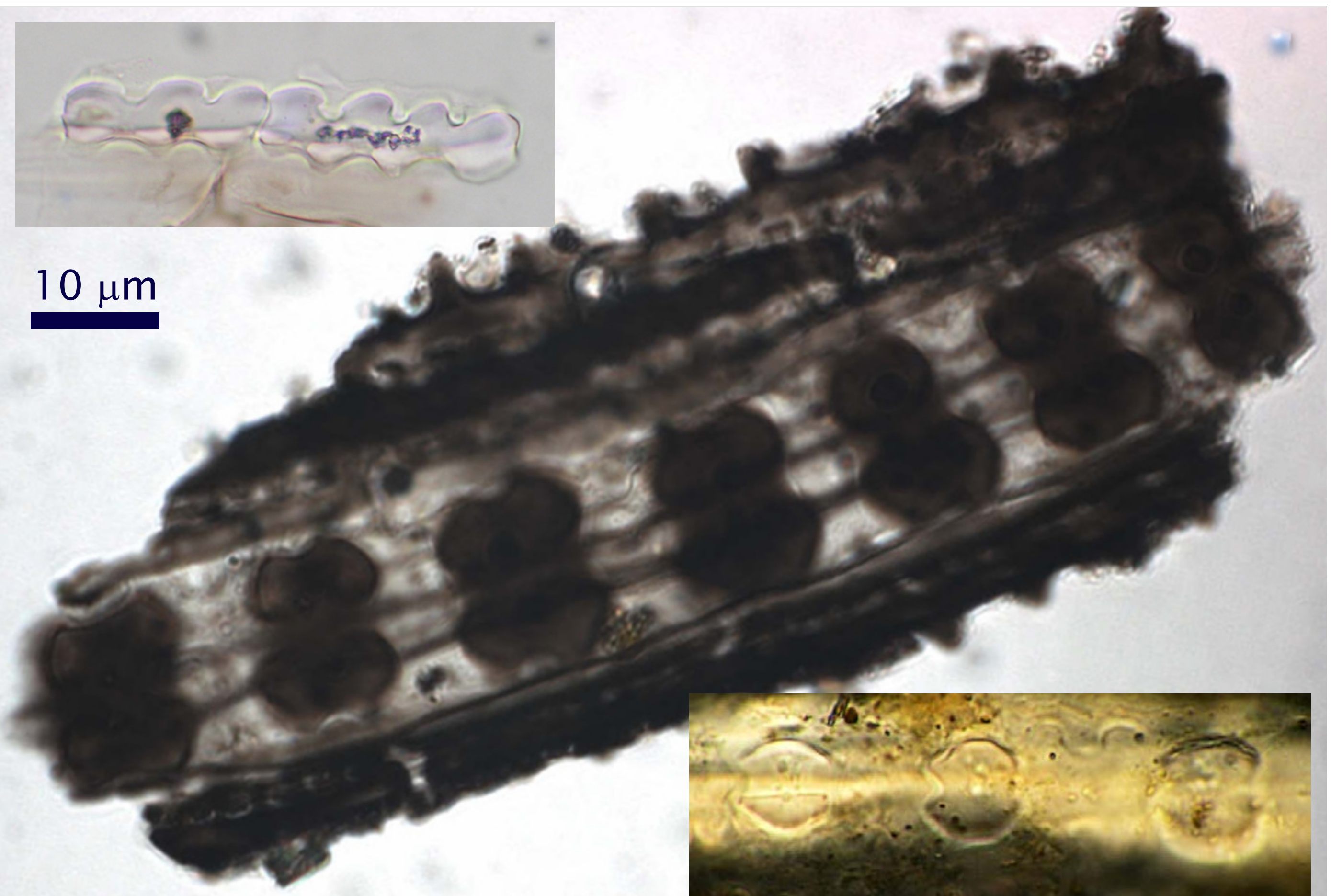
- Tropical, dry-adapted (C_4) chloridoids spread during the latest Miocene

(Strömberg 2004, 2005, 2011, Strömberg et al. 2007, 2013, Edwards et al. 2010)

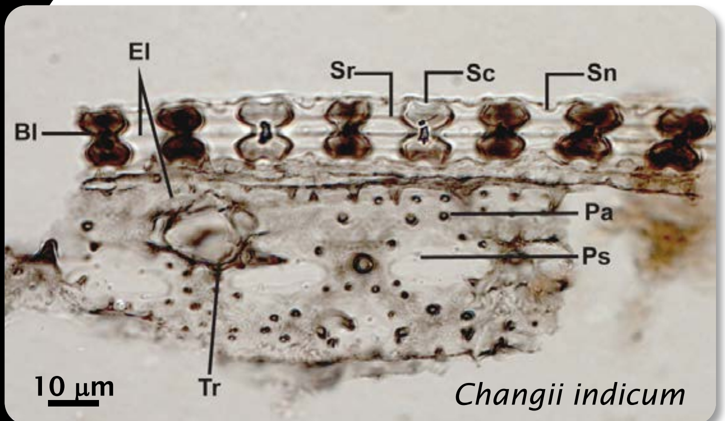
Early grass diversification

- Phytoliths (+cuticle) from Late Cretaceous dinosaur coprolites and sediment, central India

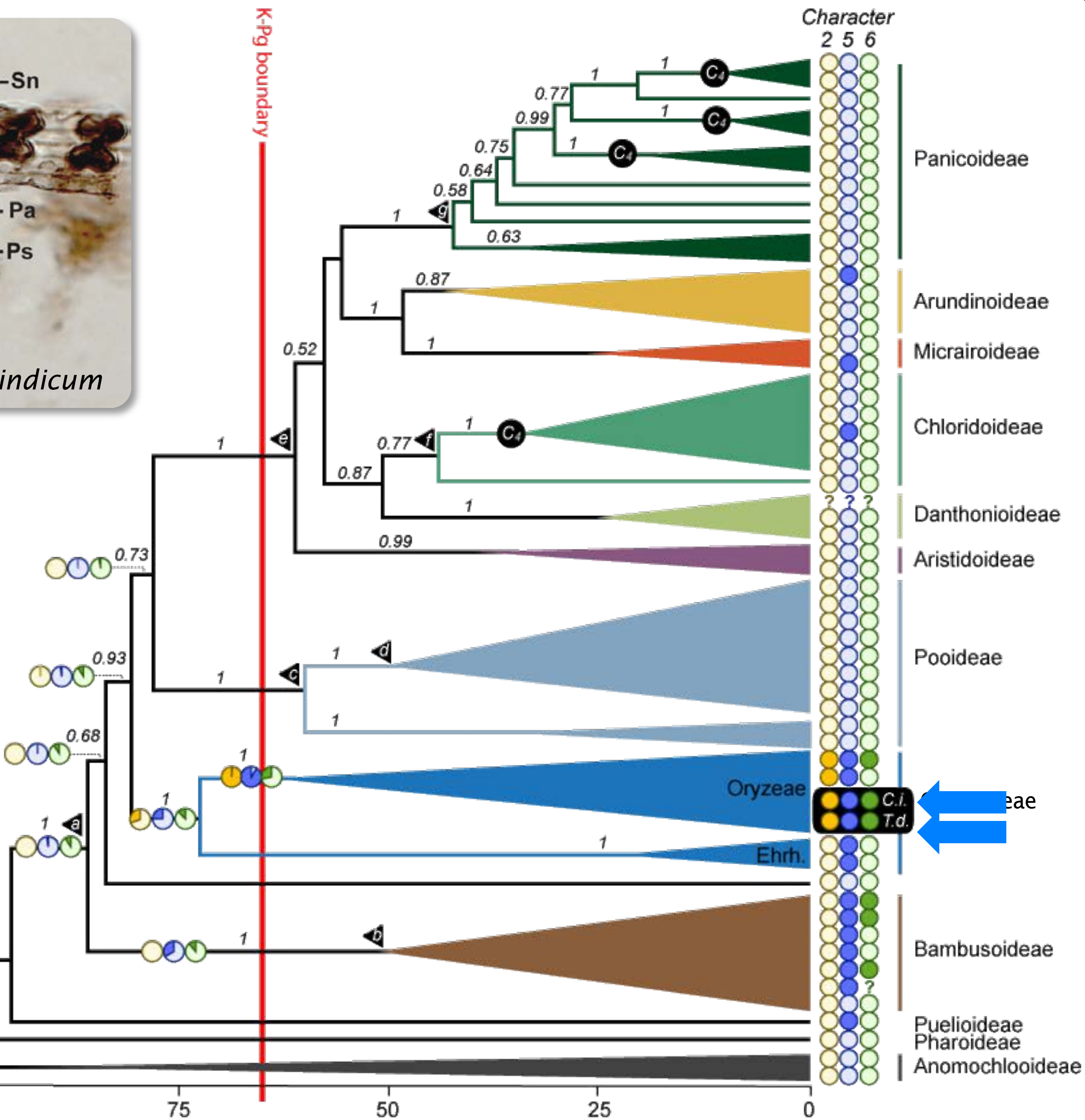
Prasad et al. (2005, 2011)



Early grass diversification



- Combined morphological + molecular phylogenetic analysis

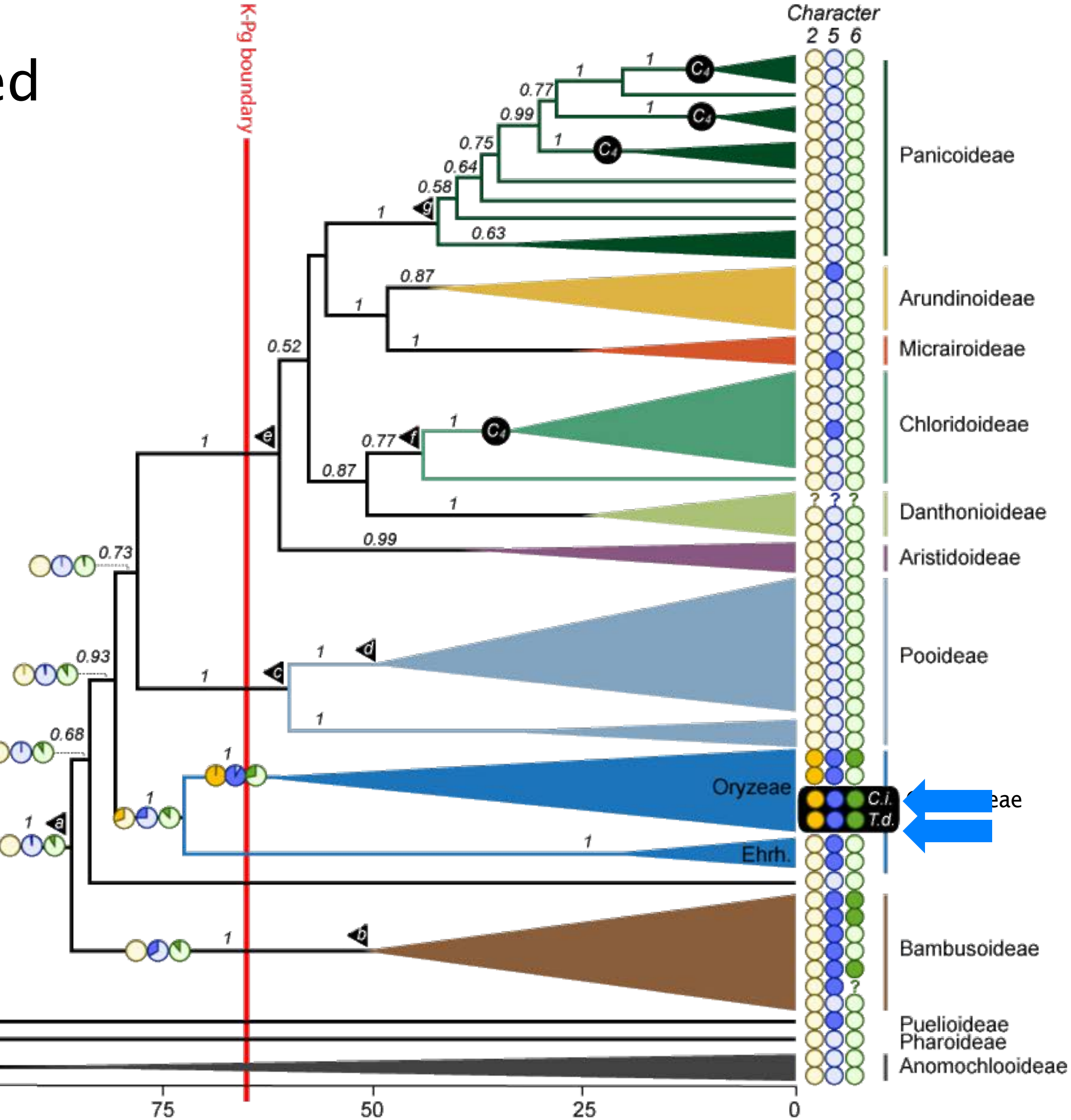


Prasad, Strömberg et al. (2011), *Nature Communications*

Early grass diversification

- Fossils are nested within Oryzeae

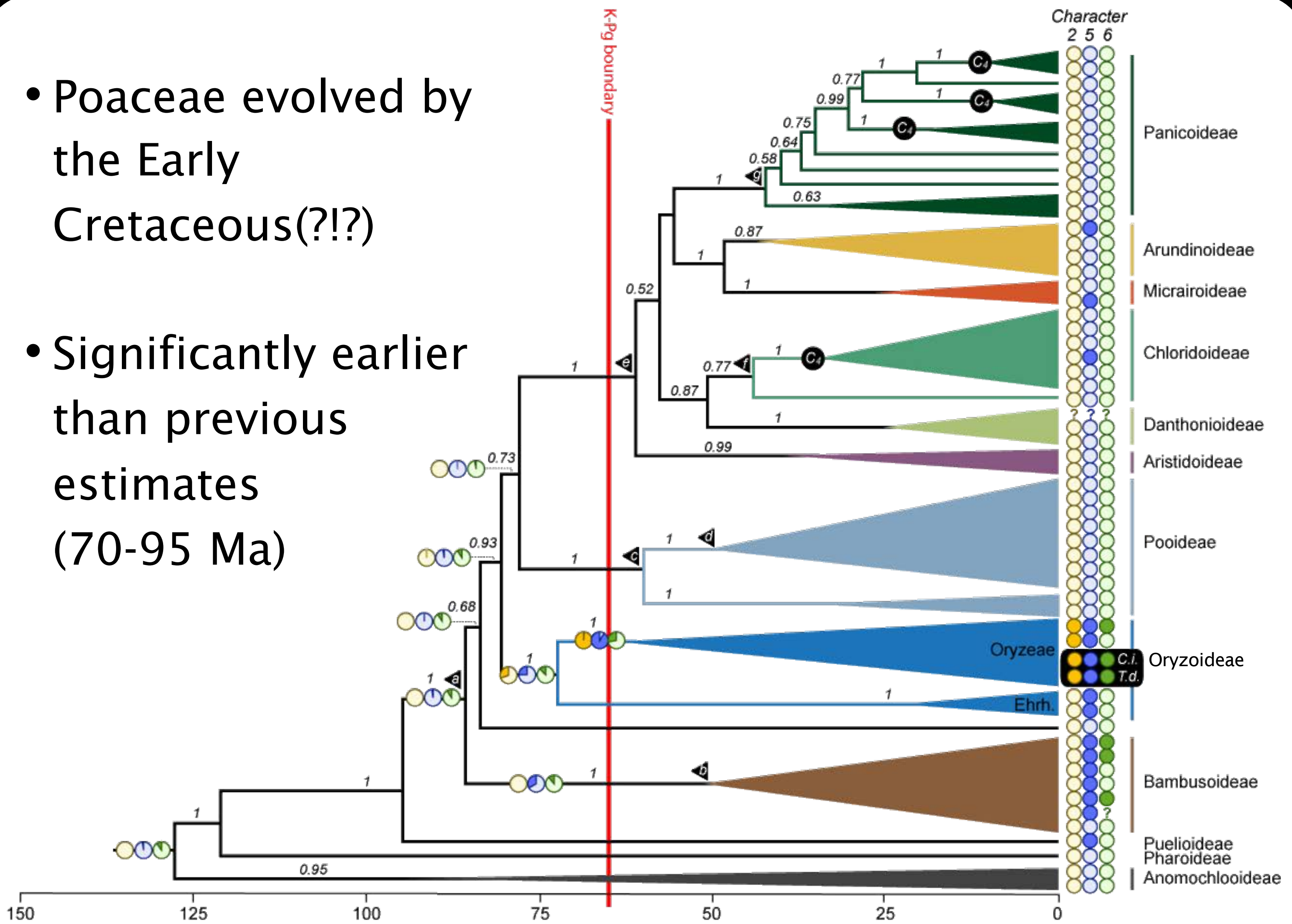
Characters:
2. Vertical bilobates
5. Papilla on long cells
6. Papilla on stomatal subsidiaries



Prasad, Strömberg et al. (2011), *Nature Communications*

Early grass diversification

- Poaceae evolved by the Early Cretaceous(?!?)
- Significantly earlier than previous estimates (70-95 Ma)

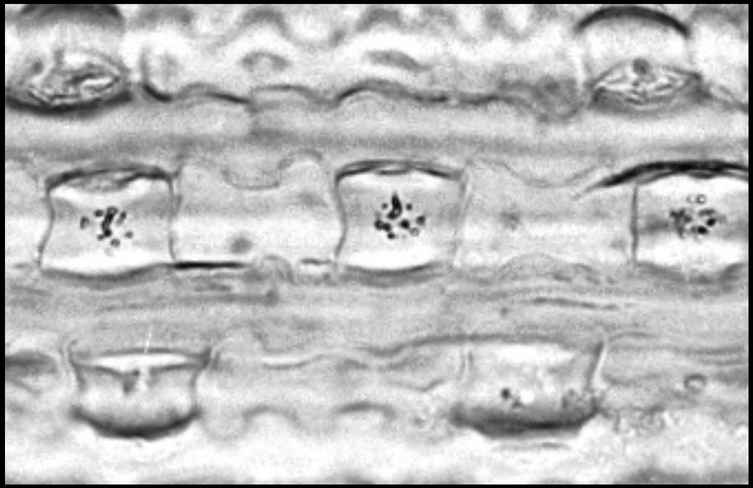


Prasad, Strömberg et al. (2011), *Nature Communications*

Problem with placing fossil phytoliths in Poaceae phylogeny

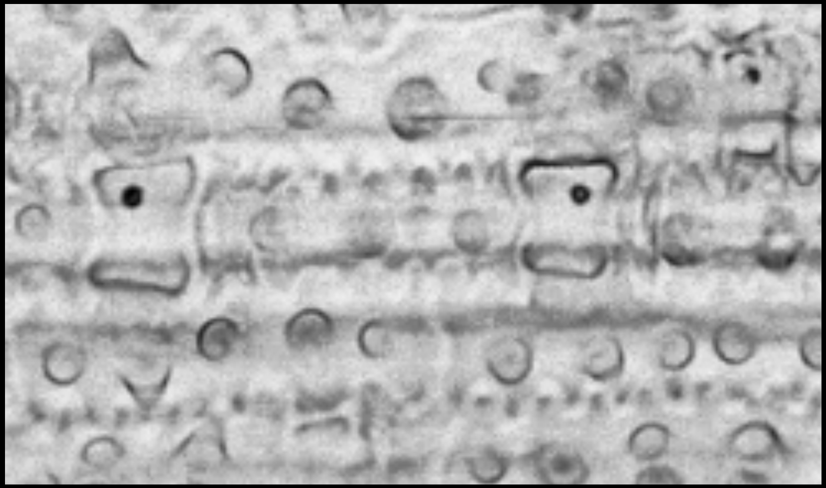
- Problem 1: Description and classification of grass phytoliths qualitative/semi-quantitative, 2-D, and subjective

“saddle-shaped”



Chloris

Dry-adapted, C_4 open-habitat grass



Sinobambusa

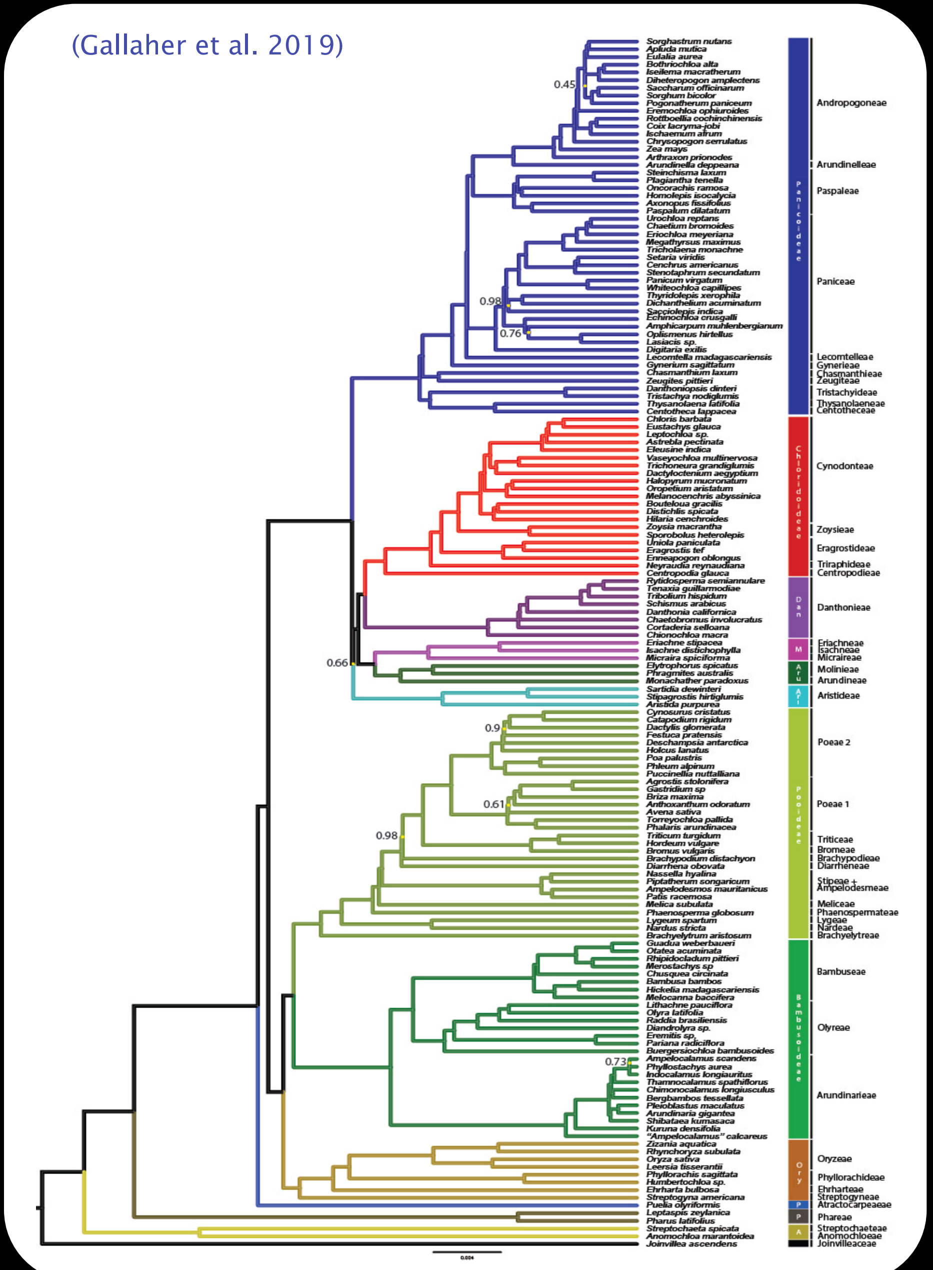
Mesophytic, tree-forming C_3 bamboo

(Watson and Dallwitz 1992-)

Problem with placing fossil phytoliths in Poaceae phylogeny

- Problem 2: Based on outdated grass taxonomy

- Recent phylogenies have dramatically changed our understanding of Poaceae relationships



Needed:

Precise, objective (and phylogenetically based) way to place grass phytoliths taxonomically—a "phylogenetic key"

Grass silica short cell phytoliths (GSSCP)

Creating a phylogenetic key to grass phytolith shape

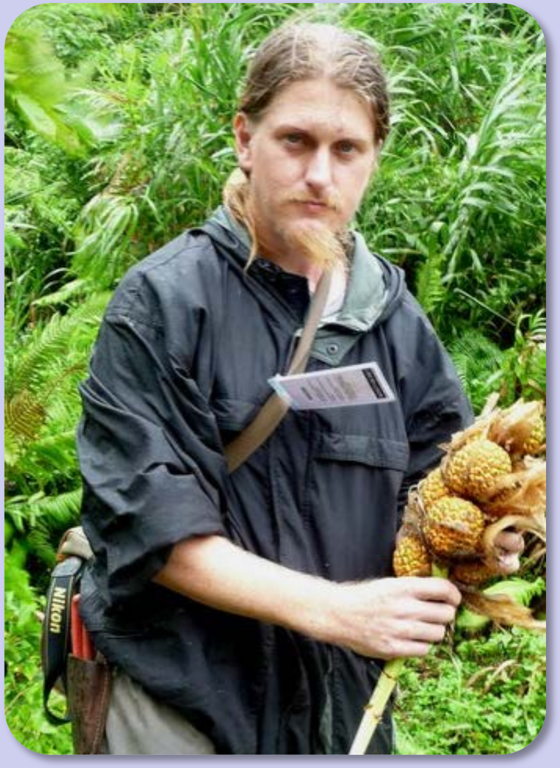
- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

Creating a phylogenetic key to grass phytolith shape

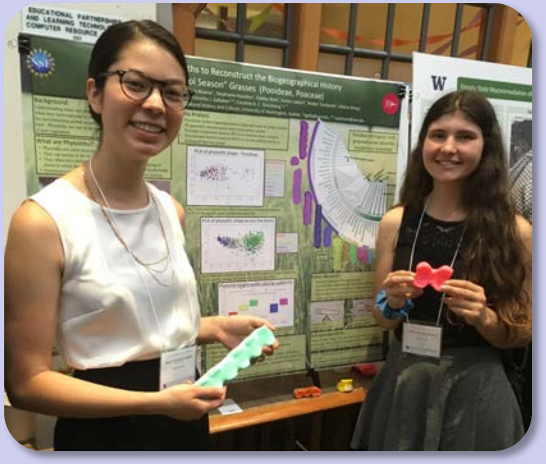
- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

Team GRASS

UW Undergraduates



Postdoc Tim Gallaher



Kari Jessett Callie Zender

Sultan Akbar

Grad student
Jessica Perry

Kevin Jamieson, UW

Anna Schorr, Nik Pershing, Elie Aboulafia, Brittany McManus, Casey O’Keefe, Claire Marvet, Brian Connor, Maura Kilpatrick, Yutaro Sakairi, Angel Alcaraz, Oakley Reid, Gwen Xiao, Claire Grant, Alex Arrendale, Sophia Druet, Tim Novak, Jessica O'Hanlon

Lynn Clark, Iowa State University

Ashly Senske Walter Tomberlin

Approaches to GSSCP identification/ placement

1. Landmark based 3-D surface geometric morphometrics
   + 3-D shape
   – landmarks, time consuming

2. Machine Learning and Computer Vision
   + regular brightfield images, no landmarks needed
   – 2-D shape, "black box"

Creating a (phylogenetic) key to GSSCP 3-D shape

Goals:

- Measure 3-D shape of GSSCPs using geometric morphometrics
- Map 3-D shape onto current phylogeny
- Correlate with ecological and physiological characters

→ Trace evolution of GSSCP shape and size across Poaceae

→ Establish GSSCP shape/size diagnostic of particular clades/ecologies/physiologies

Creating a phylogenetic key to grass phytolith shape

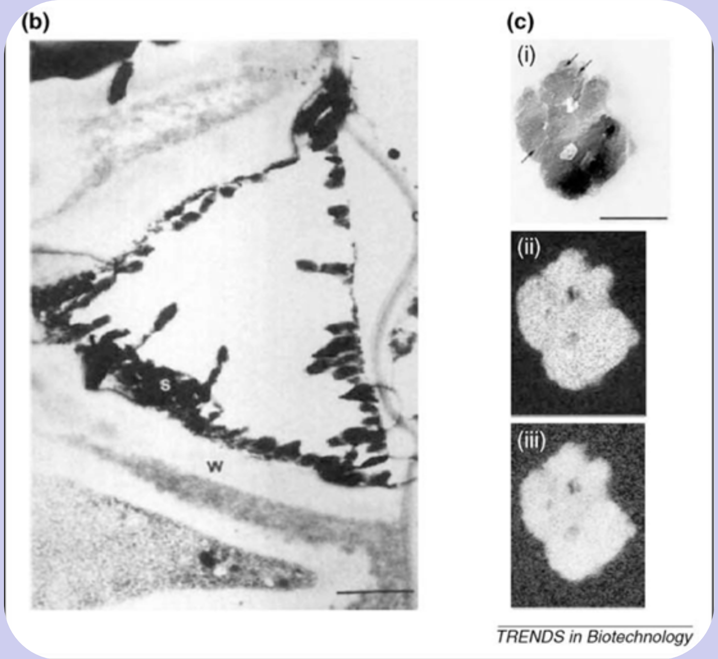
- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

Unique properties of phytoliths

1. Phytoliths consists mainly of SiO_2 (66–91%), organic carbon OC (1–6%), H_2O (0–11%), Al (0.01–4.55%), and Fe (0–2.1%)

→ Phytoliths do not auto-fluoresce and do not readily stain—or stain evenly

Developing phytolith

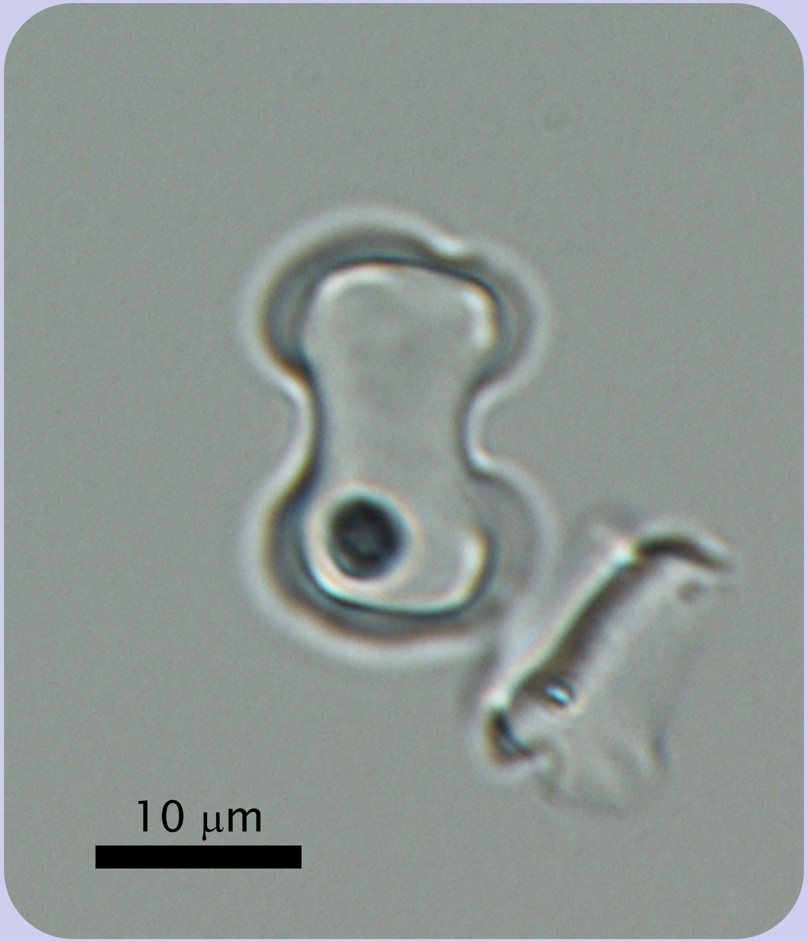


(Neethirajan et al. 2009)

Unique properties of phytoliths

2. GSSCP phytoliths are small (~7-40 micrometers)

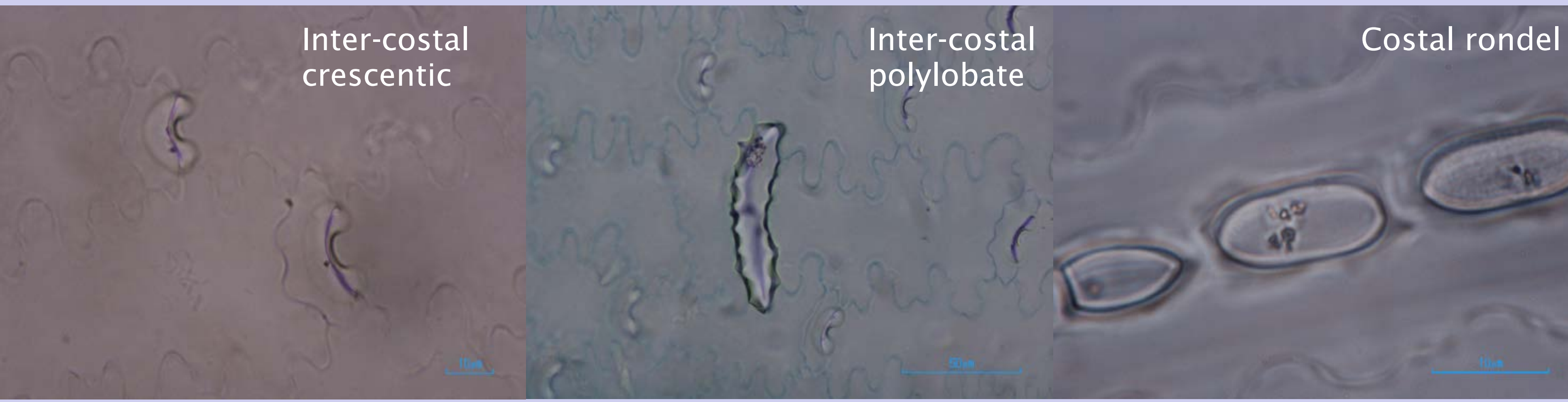
→ ***Resolution of e.g., micro-CT (100-200 micrometers) is not fine enough***



Unique properties of phytoliths

3. Most grass species make >>1 type of GSSCPs

Anomochloa marantoidea

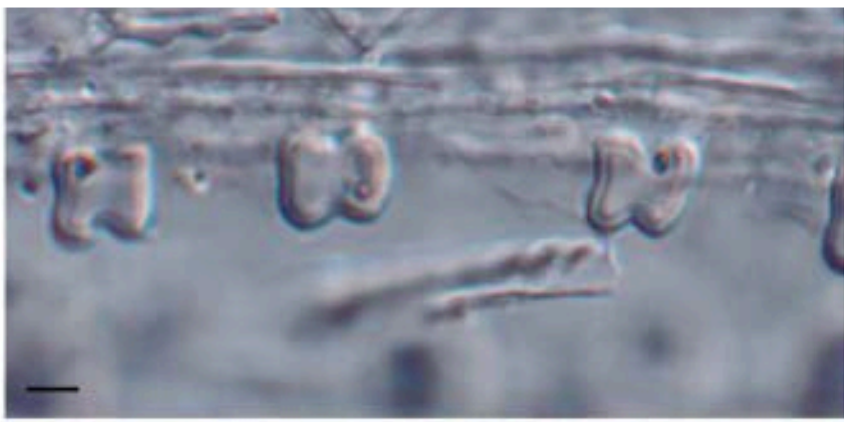


→ ***GSSCPs have to be studied like assemblages***

Unique properties of phytoliths

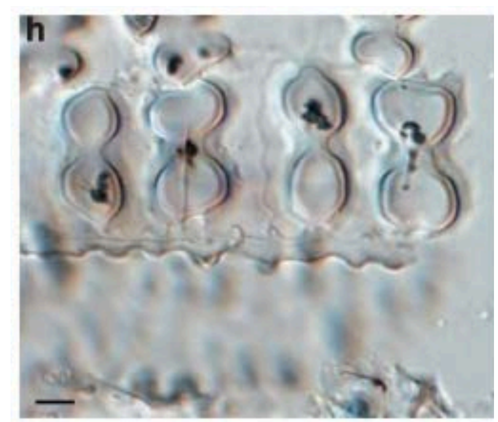
4. Similar GSSCP shapes can be oriented differently in the tissue in different species

Potamophila



Axial bilobate

Zizania

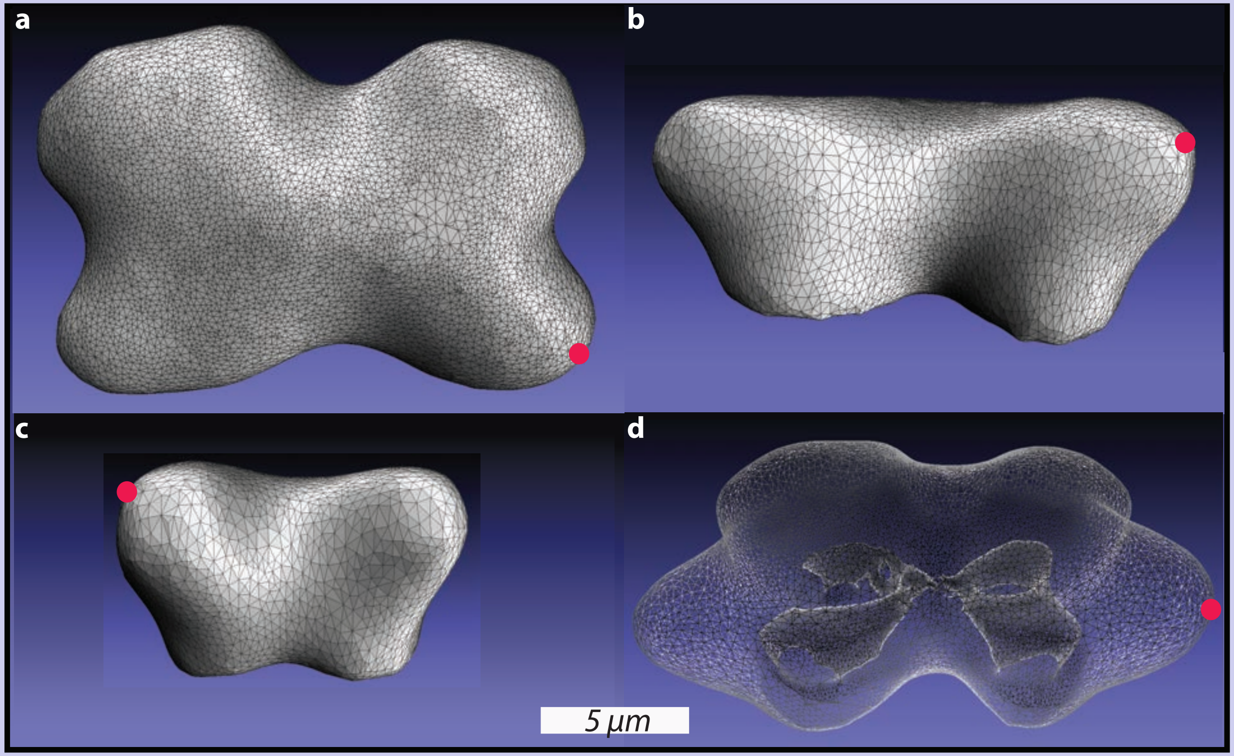


Transverse bilobate

→ ***GSSCPs have to be studied* in situ**

Unique properties of phytoliths

5. GSSCP have few good landmarks (homologous points)



→ Aligning GSSCPs relies on a few (type 3) landmarks only

Creating a phylogenetic key to grass phytolith shape

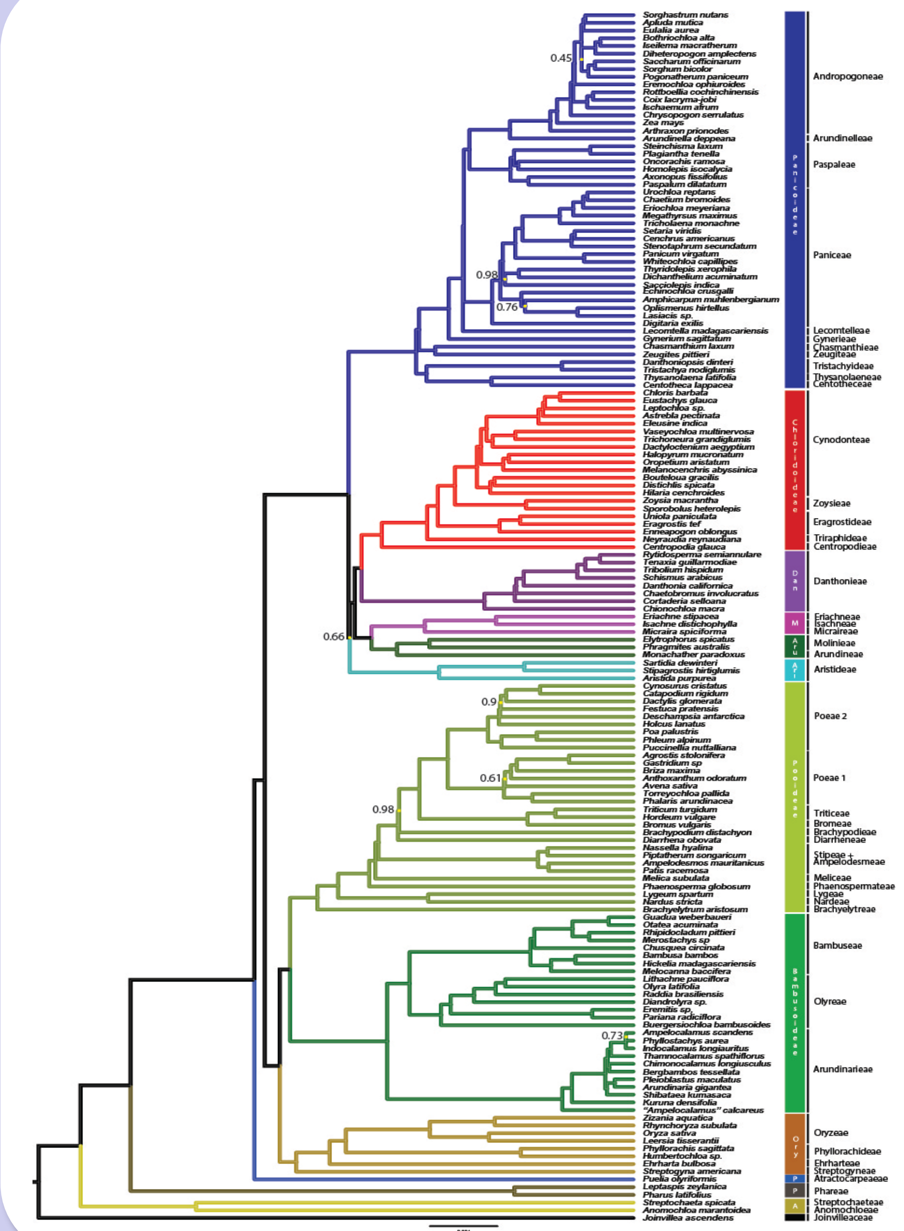
- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

Building a 3-D GSSCP shape key

Taxa sampled:

- >200 grass genera from all Poaceae subclades

- Leaf material

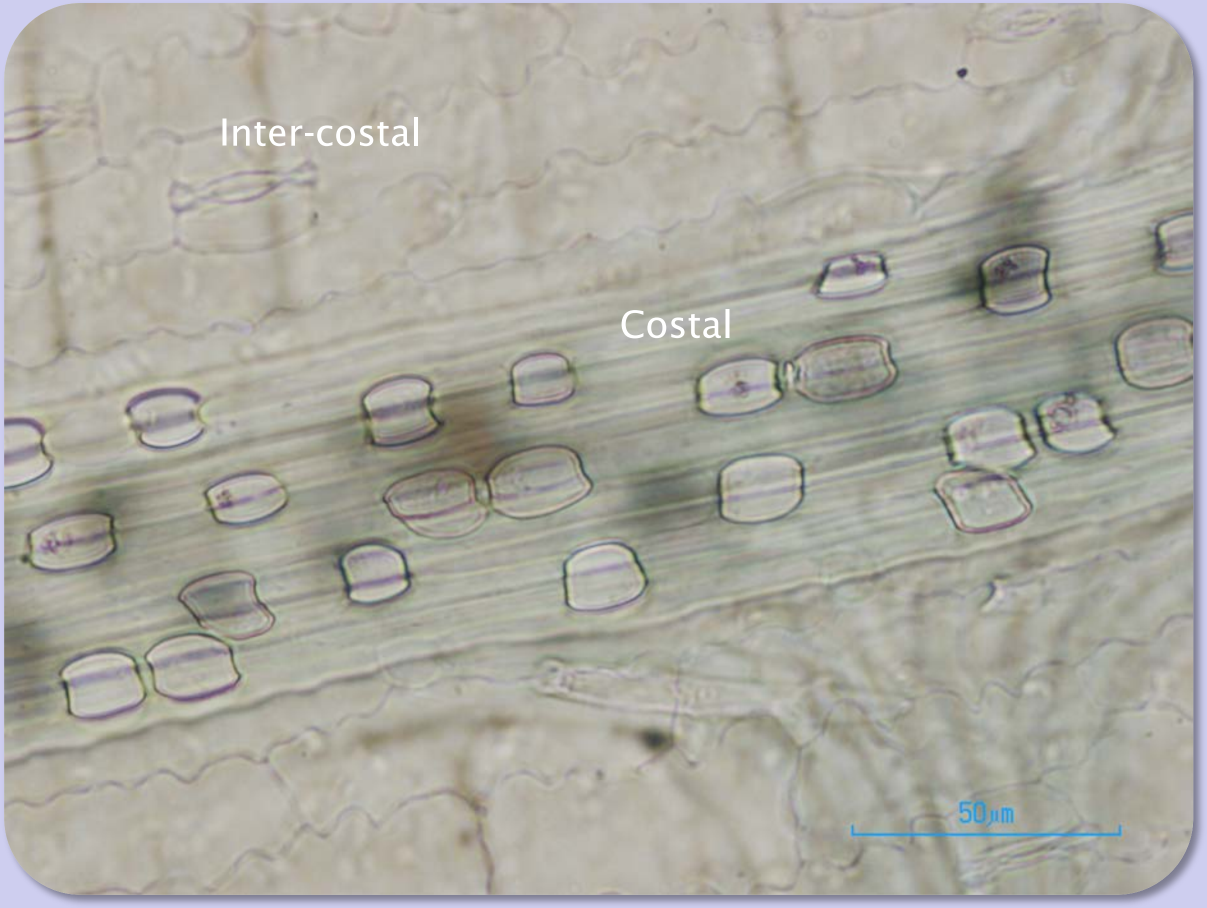


(Gallaher et al. 2019)

Building a 3-D GSSCP shape key

Data collected:

- Orientation and distribution of GSSCP shape

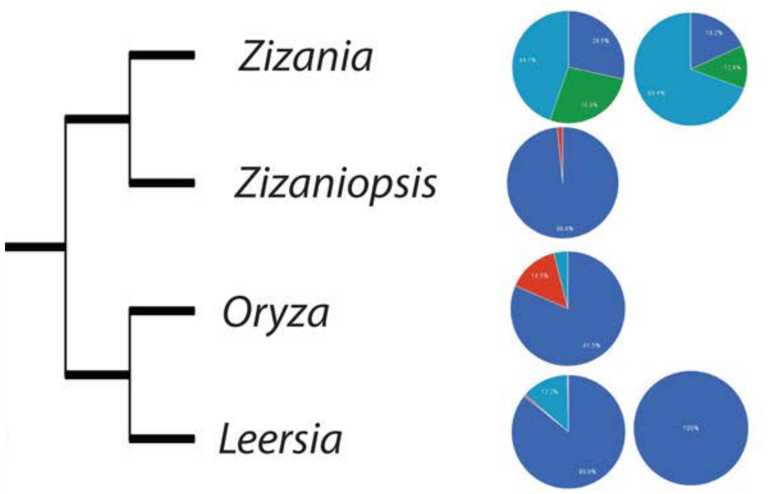


Building a 3-D GSSCP shape key

Data collected:

- Orientation and distribution of GSSCP shape

- Relative abundances of main morphotypes in GSSCP assemblages extracted from leaves

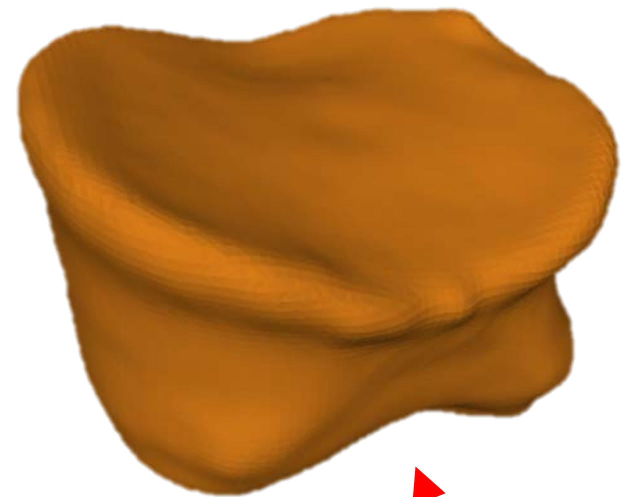


- 200 GSSCP /sample

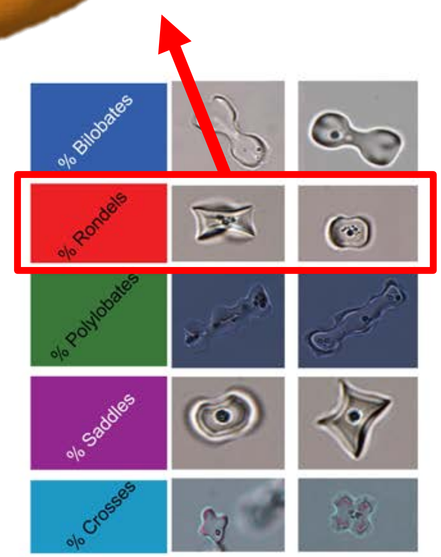
Building a 3-D GSSCP shape key

Data collected:

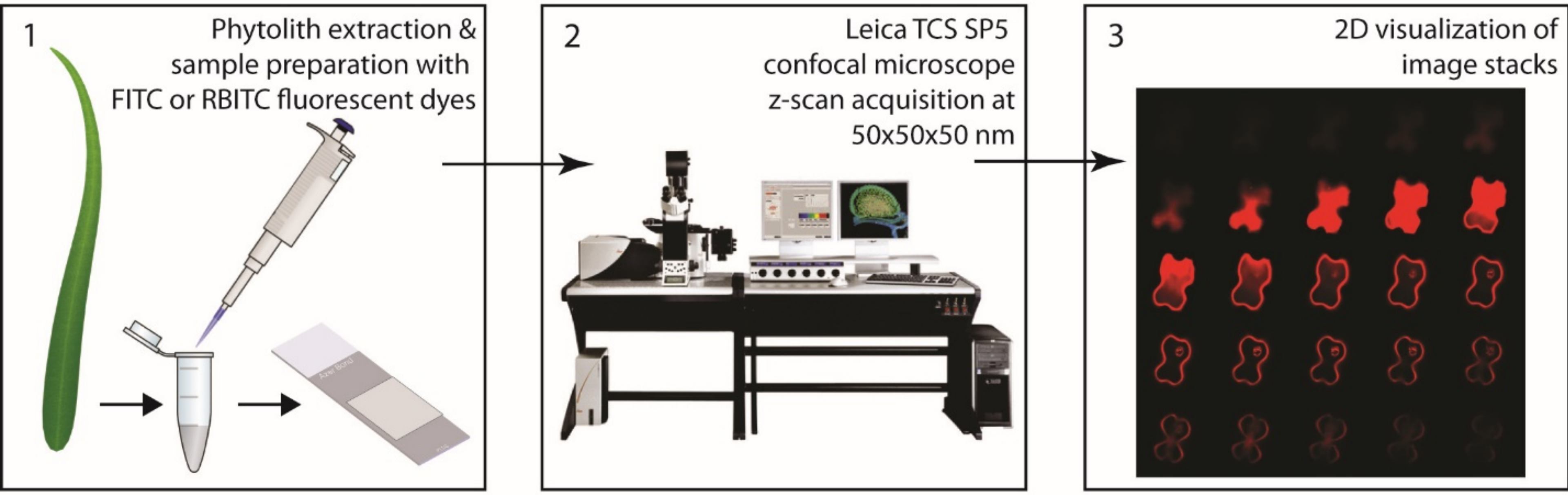
- Orientation and distribution of GSSCP shape
- Relative abundances of main morphotypes in GSSCP assemblages extracted from leaves
- 3-D shape within each main morphotype using confocal microscopy of extracted GSSCPs



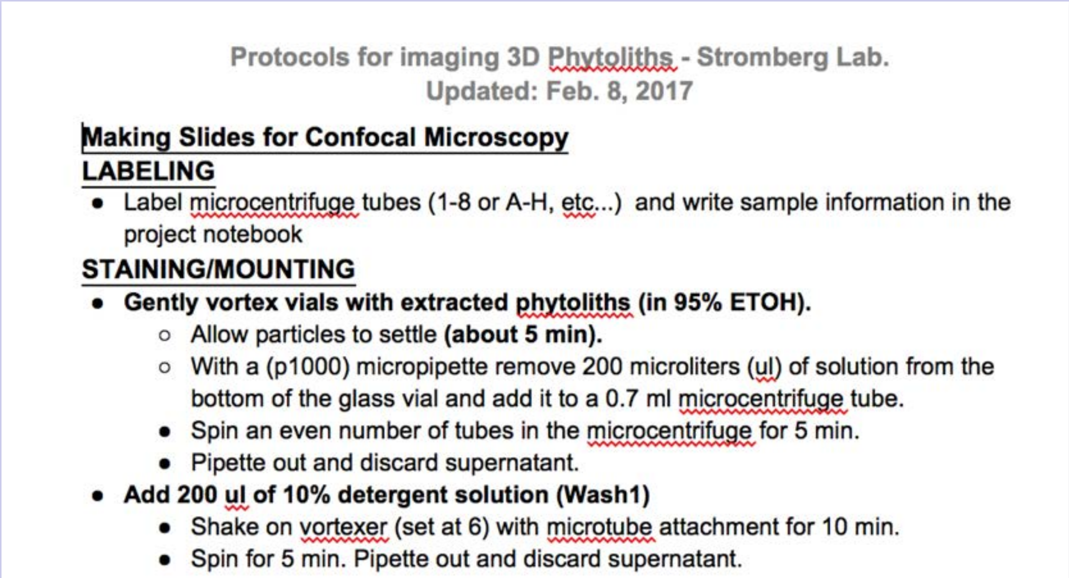
- >10 specimens /GSSCP morphotype



3-D data workflow: *Image acquisition*



- Detailed workflow protocol to ensure consistency



Protocols for imaging 3D Phytoliths - Stromberg Lab.
Updated: Feb. 8, 2017

Making Slides for Confocal Microscopy

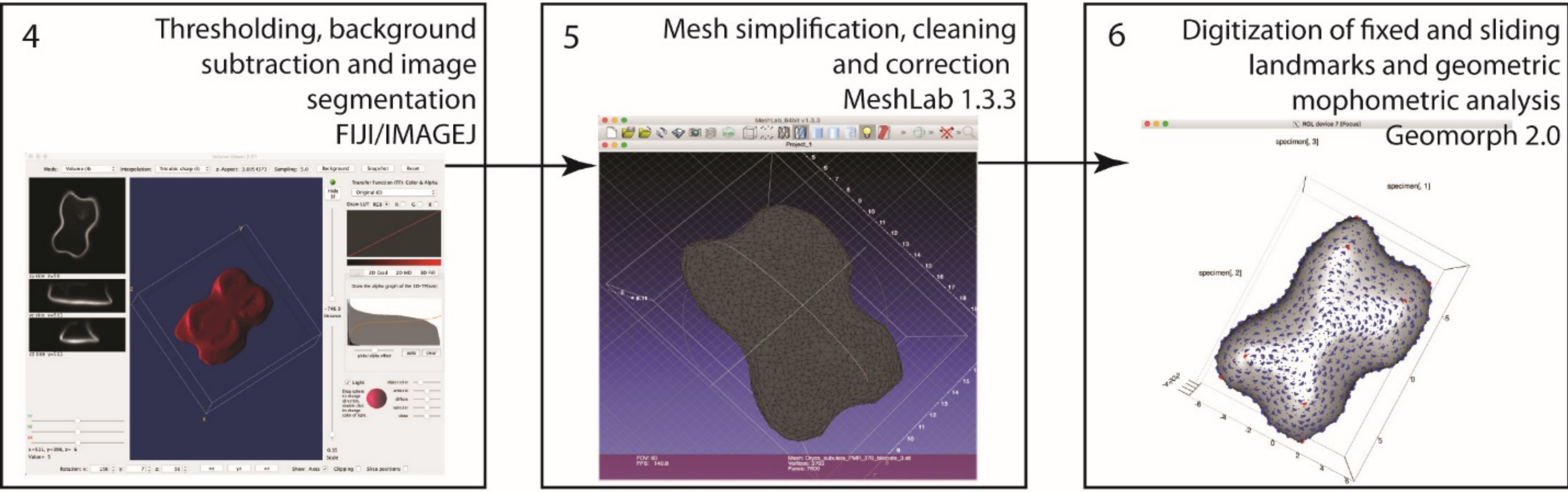
LABELING

- Label microcentrifuge tubes (1-8 or A-H, etc...) and write sample information in the project notebook

STAINING/MOUNTING

- **Gently vortex vials with extracted phytoliths (in 95% ETOH).**
    - Allow particles to settle **(about 5 min).**
    - With a (p1000) micropipette remove 200 microliters (ul) of solution from the bottom of the glass vial and add it to a 0.7 ml microcentrifuge tube.
    - Spin an even number of tubes in the microcentrifuge for 5 min.
    - Pipette out and discard supernatant.
- **Add 200 ul of 10% detergent solution (Wash1)**
    - Shake on vortexer (set at 6) with microtube attachment for 10 min.
    - Spin for 5 min. Pipette out and discard supernatant.

3-D data workflow: *Image processing and analysis*

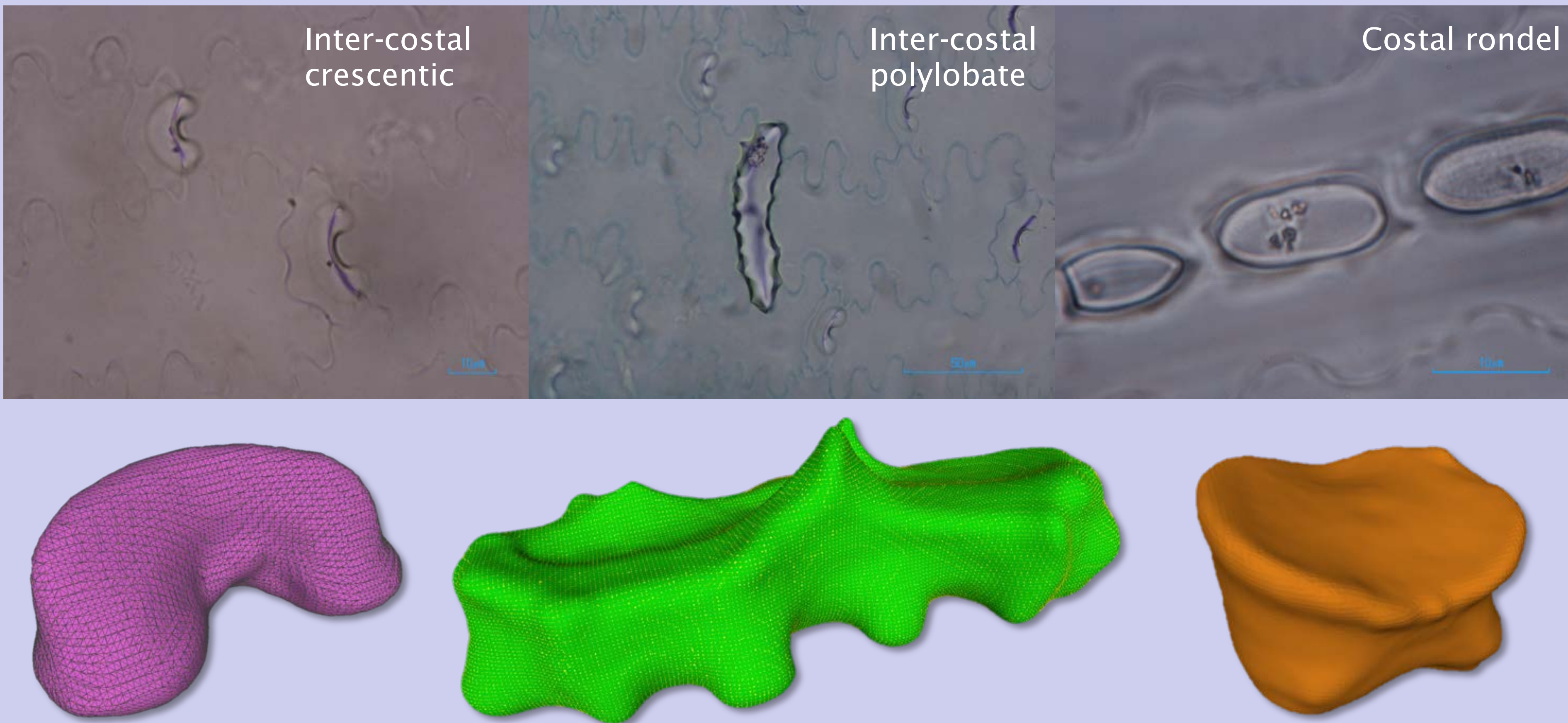


- Converting image stacks into 3-D models of GSSCPs
- Fit 6 (type 3) landmarks and 1,000 sliding surface semi-landmark points
- Transform and align meshes using Procrustes superimposition to remove size

3-D data workflow: *Outcomes*

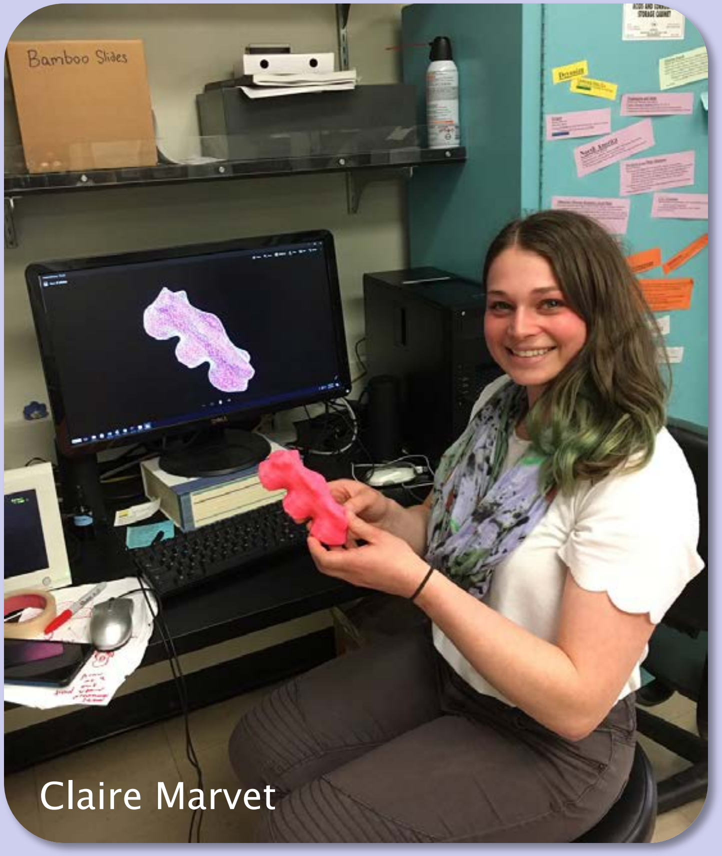
- 2,325 quantified 3-D GSSCP shapes from all 12 subfamilies for morphometric analysis and phylogenetic mapping

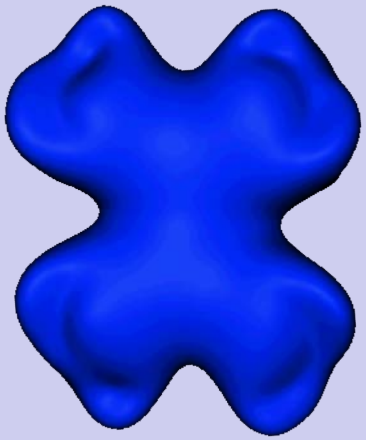
Anomochloa marantoidea



3-D data workflow: *Outcomes*

- 2,325 quantified 3-D GSSCP shapes from all 12 subfamilies for morphometric analysis and phylogenetic mapping
- Animations and 3-D printable objects

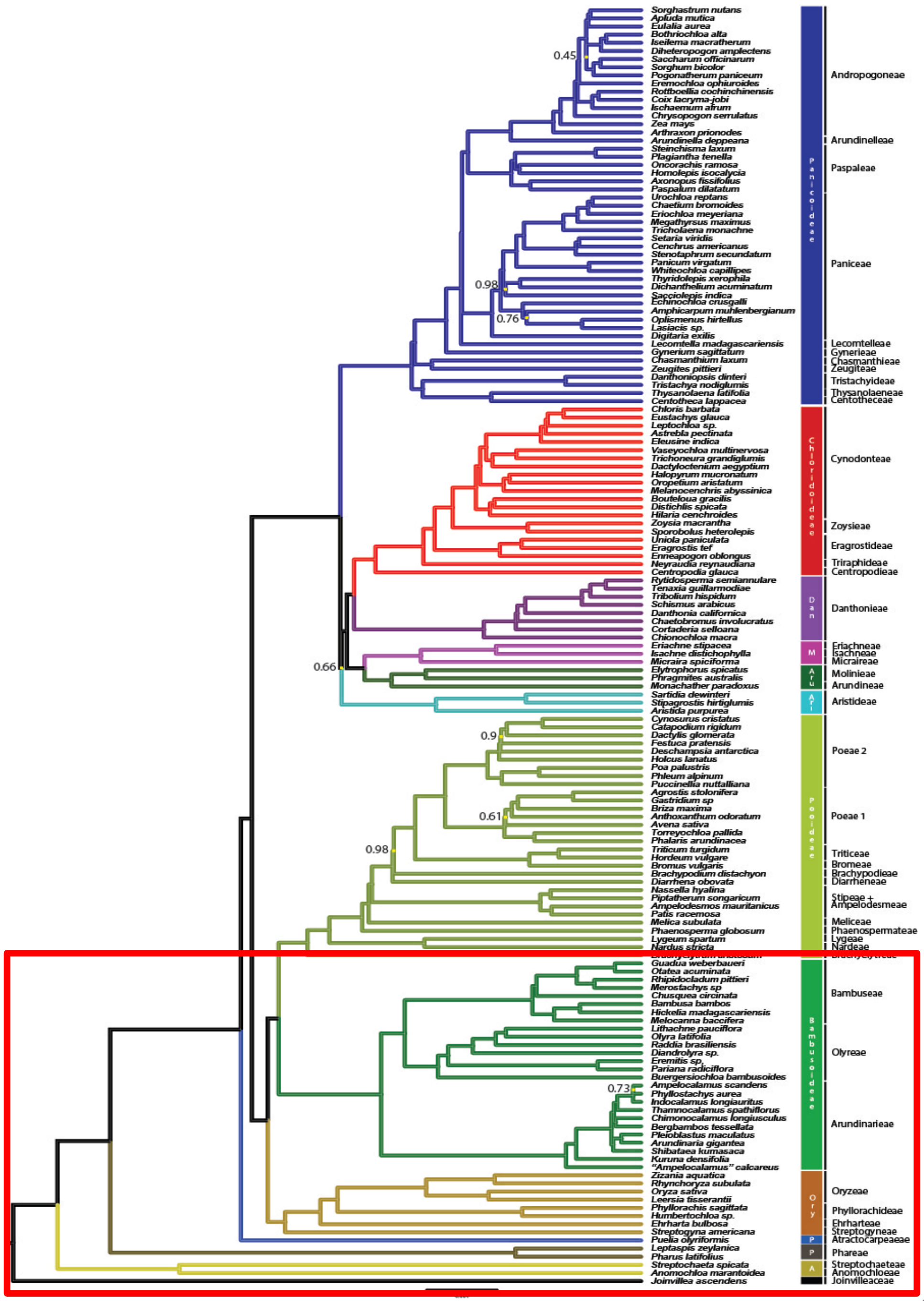




Results

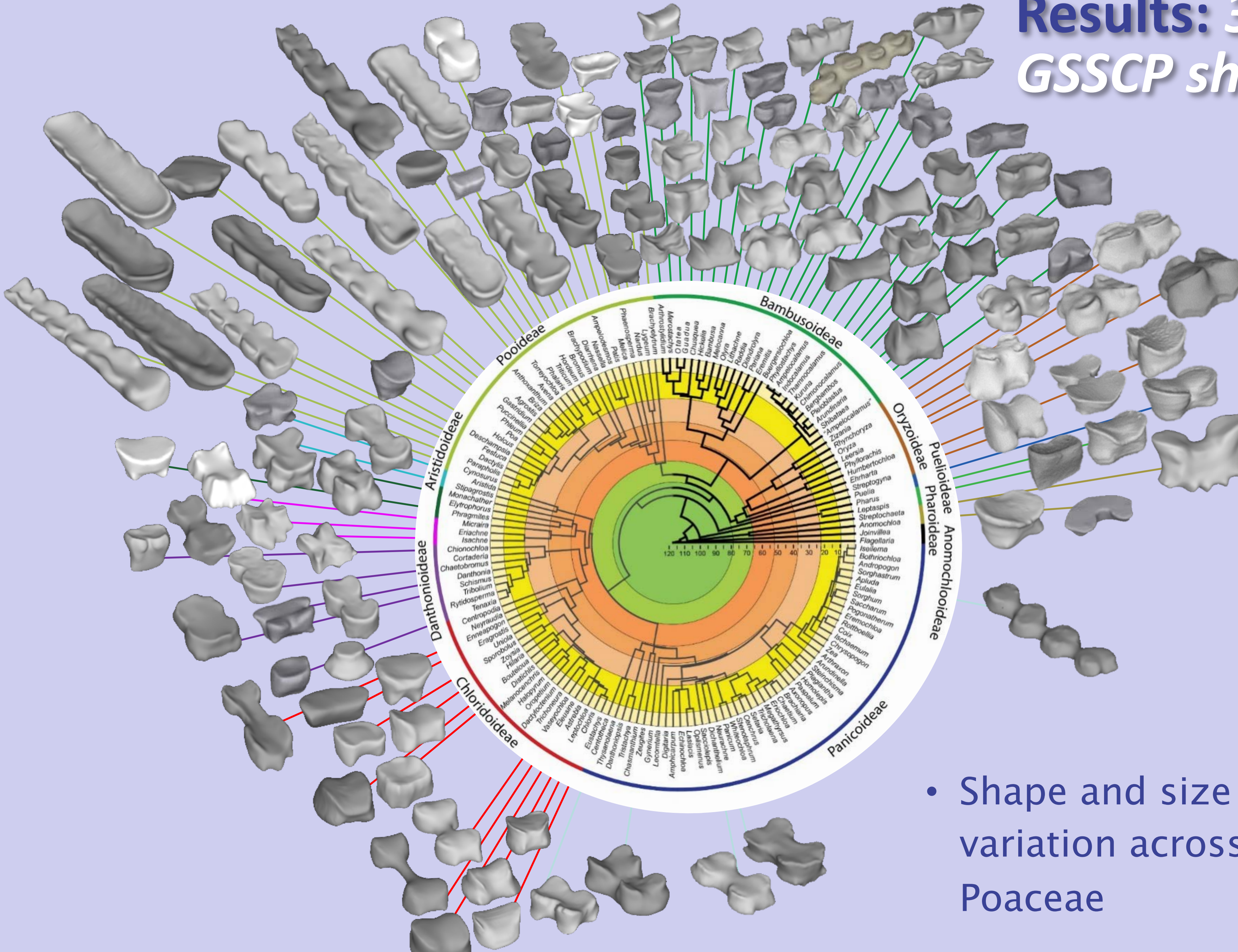
- Focus on early-diverging grasses (A+P+P), Bambusoideae, and Oryzoideae

(Gallaher et al. in prep.)



(Gallaher et al. 2019)

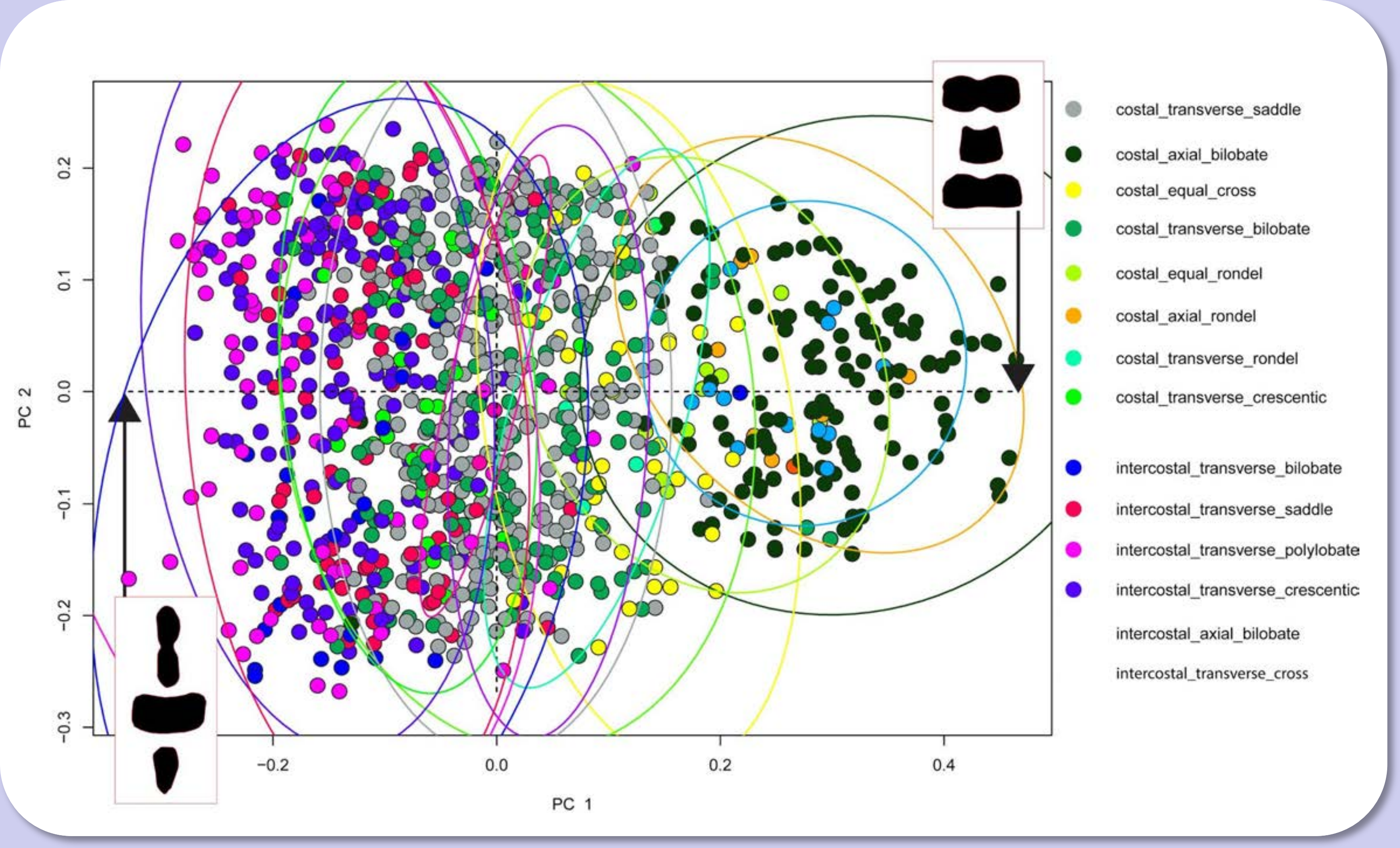
Results: *3-D GSSCP shape*



- Shape and size variation across Poaceae

Results: *3-D GSSCP shape classification*

- Principal Component visualization of GSSCP 3-D shape after Procrustes alignment (1,093 images)—there is signal but a lot of noise!

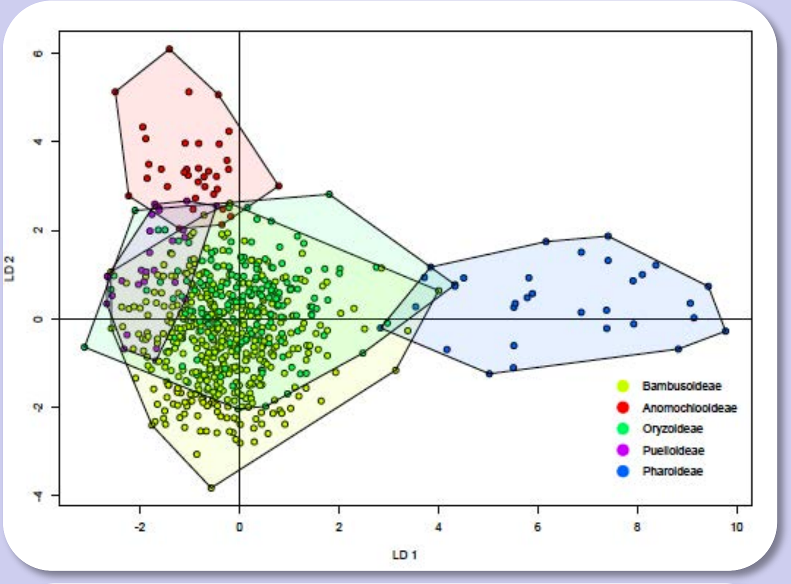


(Gallaher et al. in prep.)

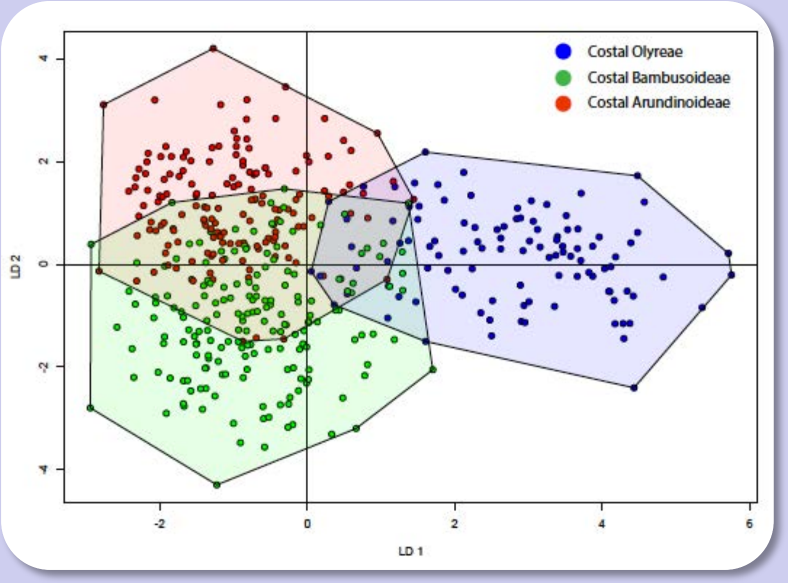
Results: *3-D GSSCP shape classification*

- Linear Discriminant Analysis (LDA) of GSSCP 3-D shape

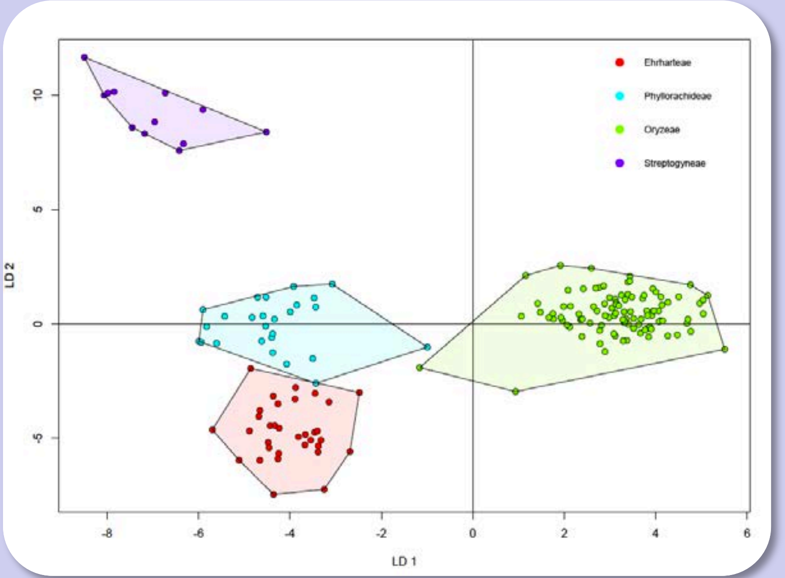
Subfamilies



Bambusoideae tribes



Oryzoideae tribes



- Data = first 66 PCs plus volume, surface area, centroid size, length, width, and height

(Gallaher et al. in prep.)

Results: *3-D GSSCP shape classification*

- LDA classification rates (Accuracy) assessed

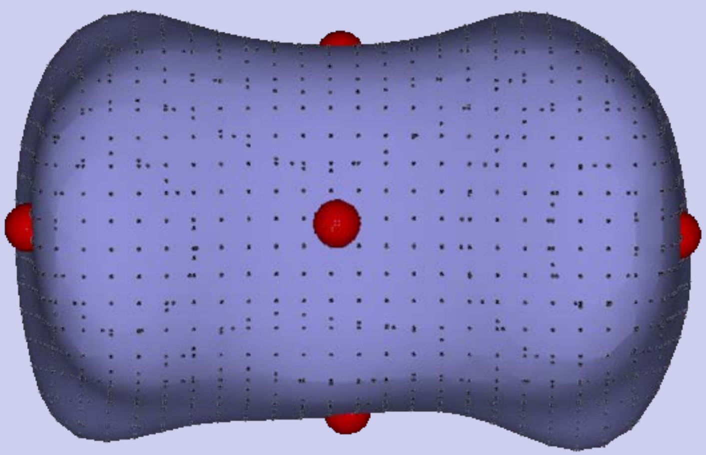
Taxon (# classes)	Accuracy LDA: Mean (min-max)	Kappa: Mean (min-max)	x random chance
Subfamily (5)	0.7821 (0.6585 - 0.8780)	0.5513 (0.3131 - 0.7509)	**2.8**
Tribe (12)	0.6516 (0.5116 - 0.7975)	0.5781 (0.4115 - 0.7511)	**6.9**
Subtribe (22)	0.5546 (0.3924 - 0.7381)	0.5134 (0.3279 - 0.7141)	**11.3**
Genus (46)	0.4514 (0.3415 - 0.6098)	0.4388 (0.3259 - 0.6005)	**20.2**

- Correct classification rates are high (3-20x random chance)

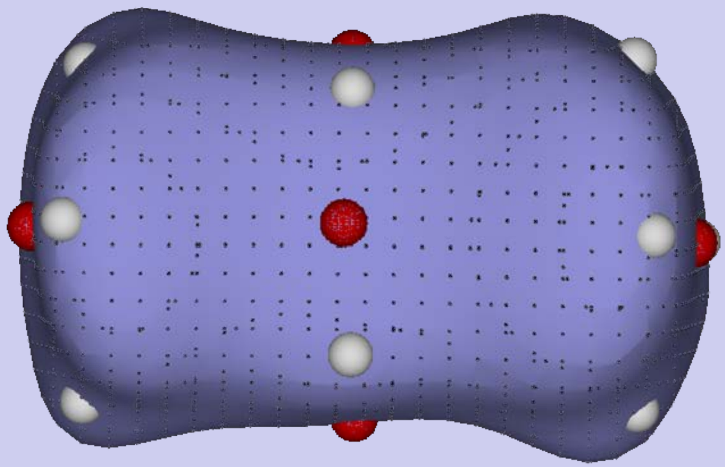
(Gallaher et al. in prep.)

Work in progress

- Switching from 6 to 16 (type 3) landmarks to better fit semi-landmarks to shape



6 (type 3) landmarks

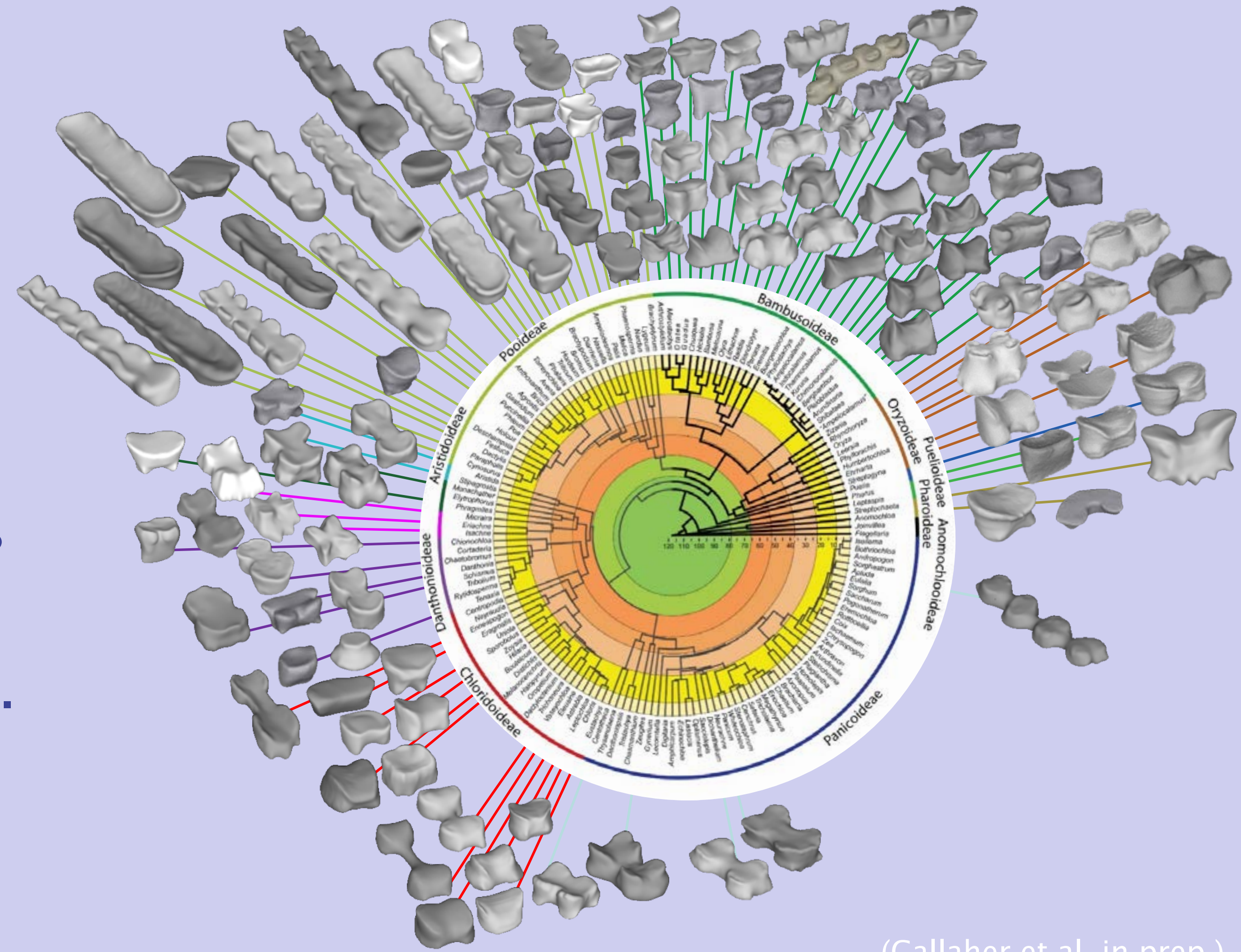


16 (type 3) landmarks

(Gallaher et al. in prep.)

Work in progress

- Testing for phylogenetic signal in 3-D shape and analyzing shape change across the Poaceae phylogeny



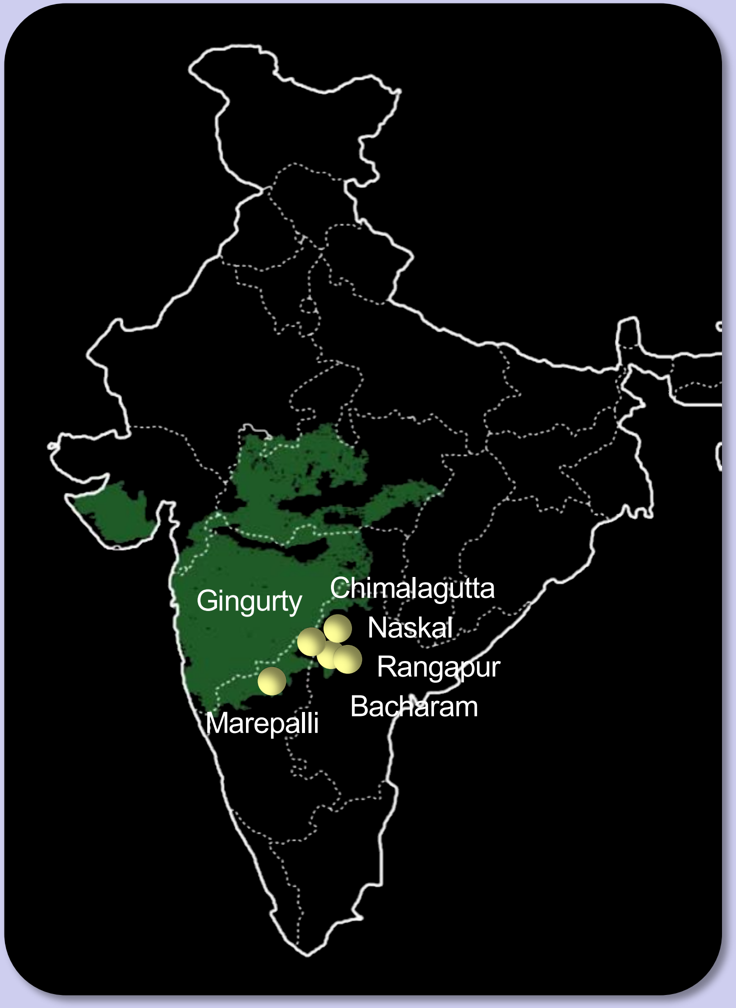
- Mapping grass ecology, physiology etc. and look for correlations

(Gallaher et al. in prep.)

Creating a phylogenetic key to grass phytolith shape

- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

GSSCPs from Late K-Paleocene sites in India

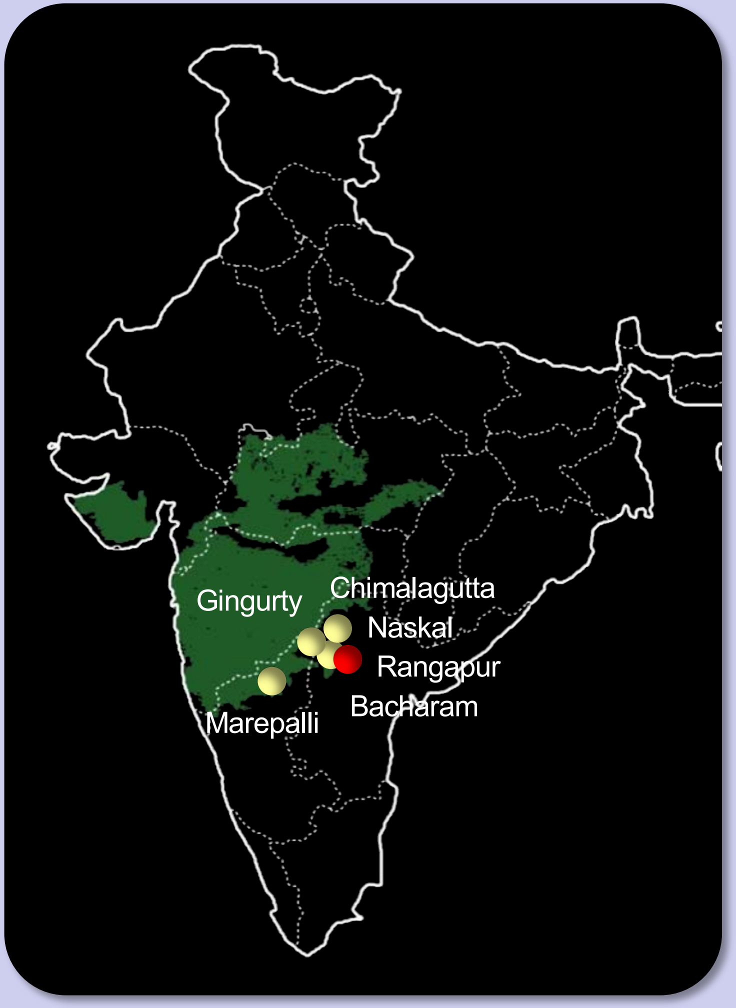


- Study of faunal localities in central India (sedimentology, fauna, flora)
- Late K and Paleocene localities studied so far all have grass phytoliths

Greg Wilson (University of Washington)

Jason Moore (University of New Mexico)

GSSCPs from Late K-Paleocene sites in India

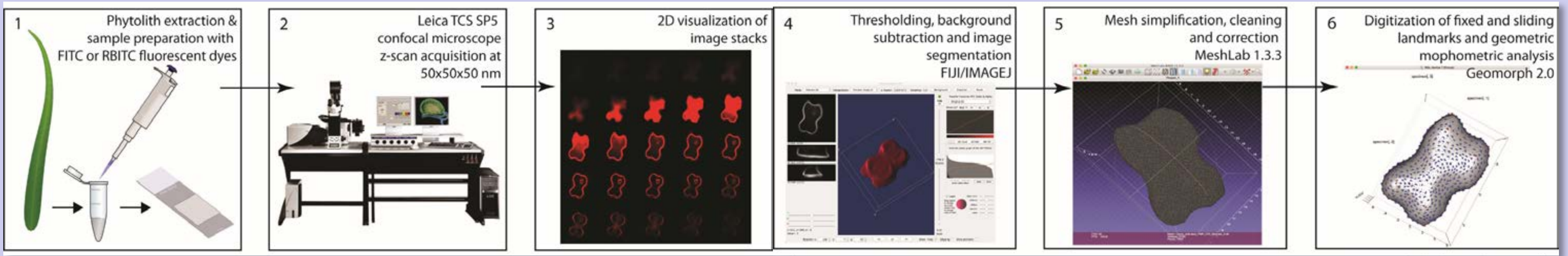


- Study of faunal localities in central India (sedimentology, fauna, flora)
- Late K and Paleocene localities studied so far all have grass phytoliths
- Preliminary results from the earliest Paleocene(?) Naskal site with abundant grasses

Greg Wilson (University of Washington)

Jason Moore (University of New Mexico)

GSSCPs from Naskal (early Paleocene?), India

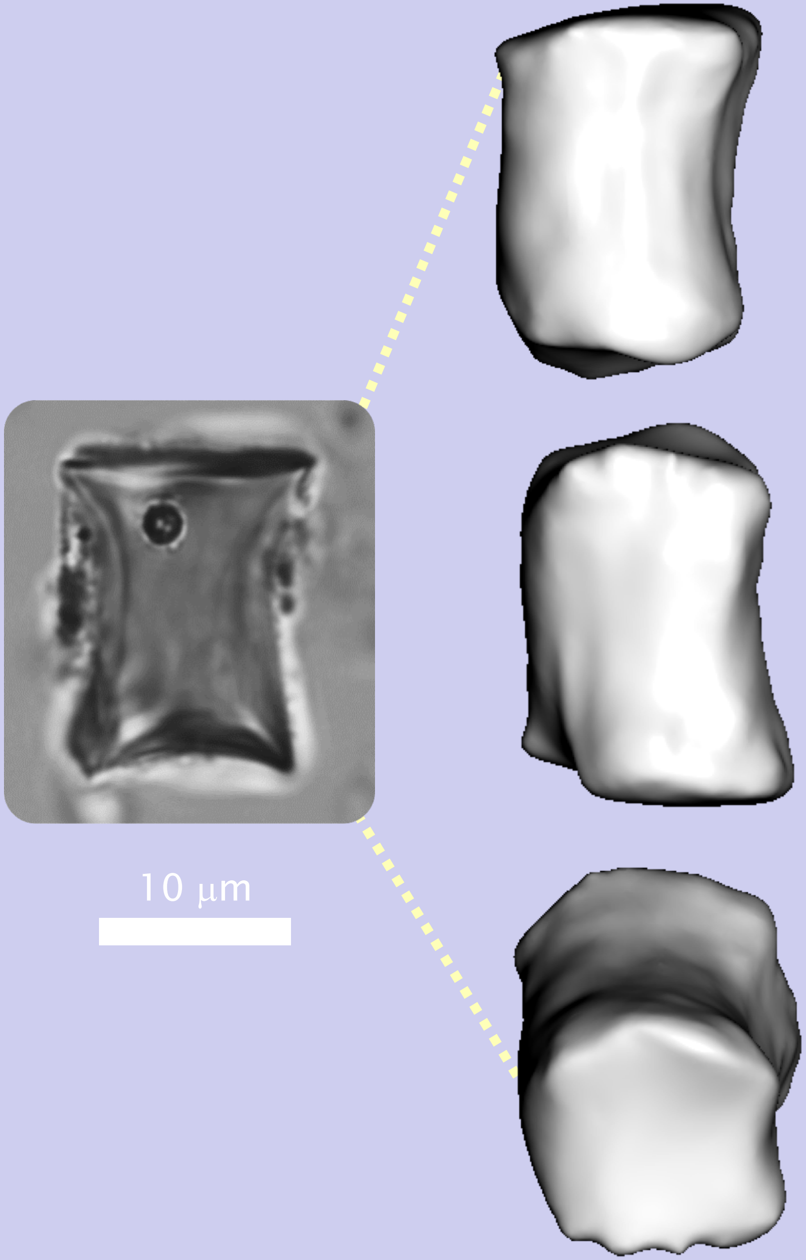


3-D data workflow (fossils):

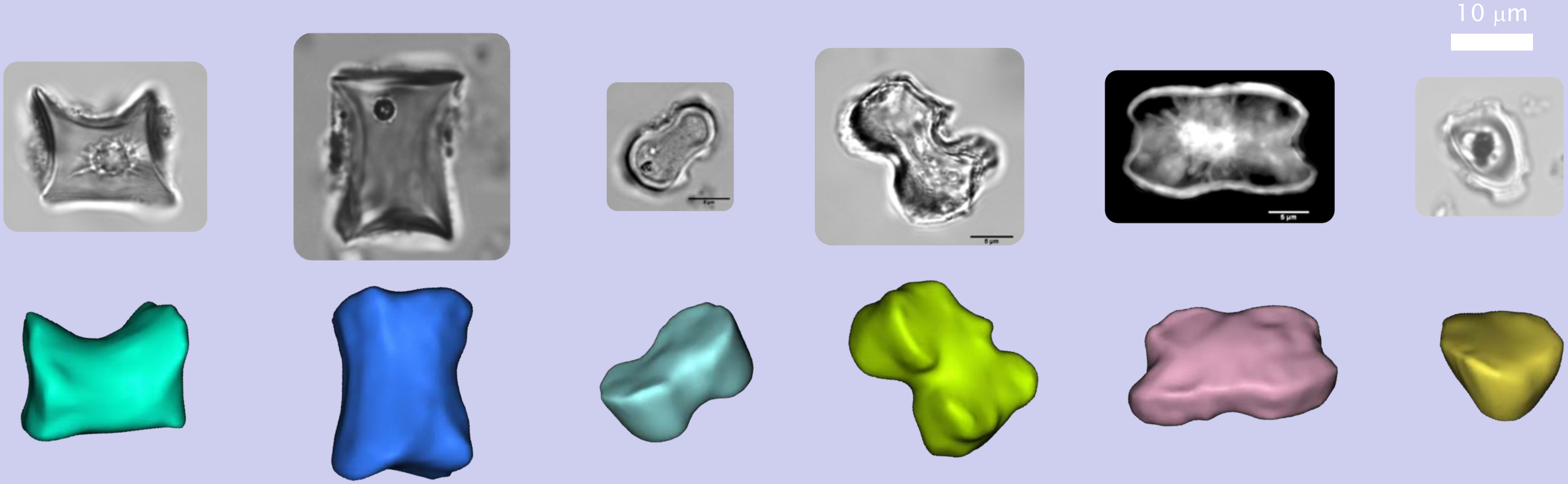
- Extraction, staining and imaging
- Converting image stacks into 3-D models of GSSCPs
- Fit 6 (type 3) landmarks and 1,000 sliding surface semi-landmark points
- Transform and align meshes using Procrustes superimposition to remove size

GSSCPs from Naskal (early Paleocene?), India

- Fossil GSSCPs are covered in debris and are often weathered obscuring fine details
- … but the the overall shape is visible
- Preliminary results for six GSSCPs



Taxonomic placements of Paleocene GSSCPs



- Linear Discriminant Analysis (LDA) using our modern dataset to determine taxonomic affinities

Taxonomic placements of Paleocene GSSCPs

Caveats with taxonomic assignment:

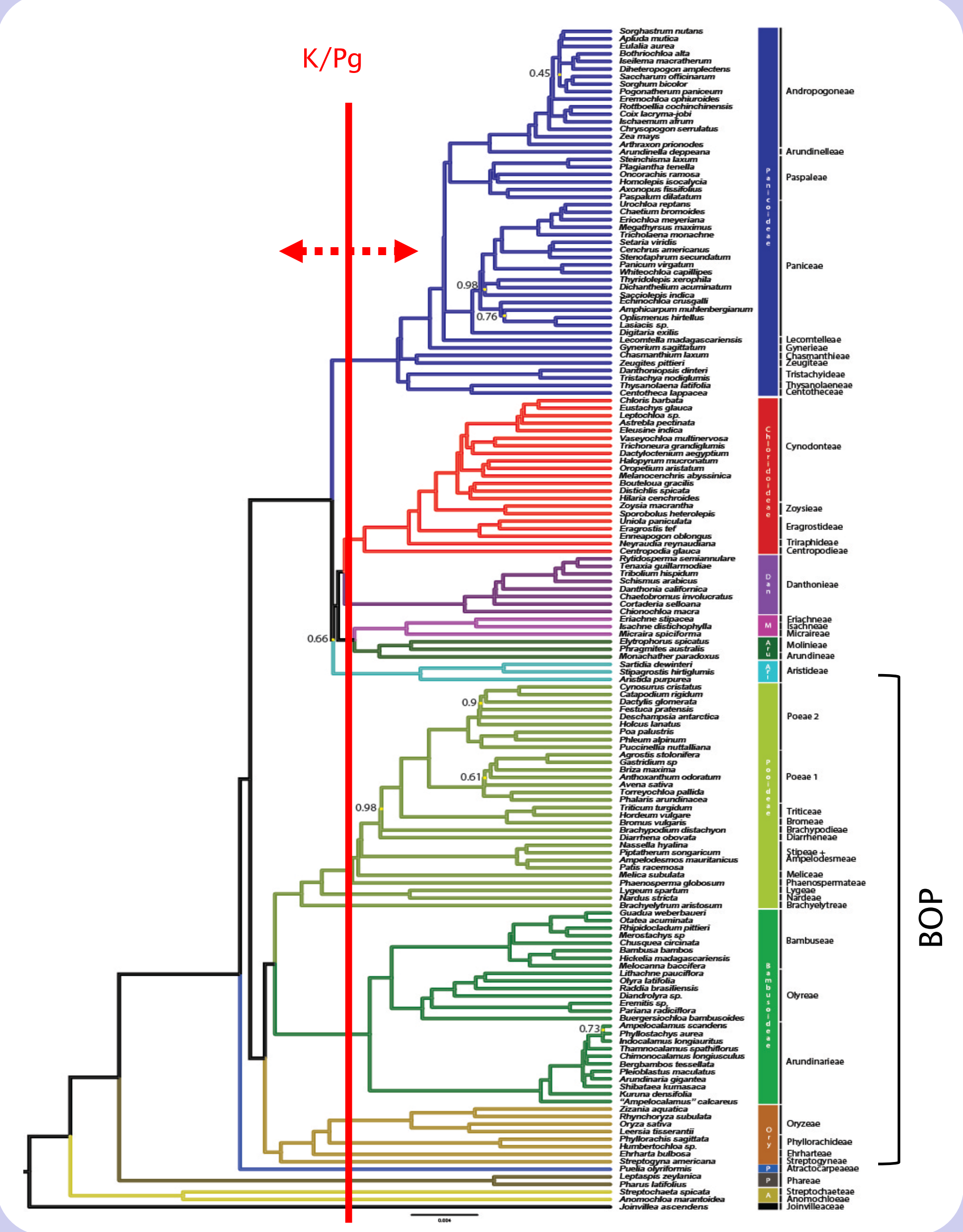
- So far very few GSSCPs have been analyzed
- Modern dataset have only 320 PACMAD specimens
- Phylogeny not explicitly included

Implications of analysis of Paleocene GSSCPs

- Representatives of BOP clades—similar to previous analyses

Prasad et al. (2005, 2011)

- Supports diversification of BOP well before 66 Ma



Creating a phylogenetic key to grass phytolith shape

- Approaches to GSSCP identification/placement
- Properties of GSSCP
- Building a 3-D shape key to GSSCP
- Preliminary application to the Late K-Paleocene of India
- GSSCP and Machine Learning & Computer Vision

Approaches to GSSCP identification/ placement

1. Landmark based 3-D surface morphometrics
   + 3-D shape
   – landmarks, time consuming

2. Machine Learning and Computer Vision
   + regular brightfield images, no landmarks needed
   – 2-D shape, "black box"

Tim Gallaher

Jessica Perry

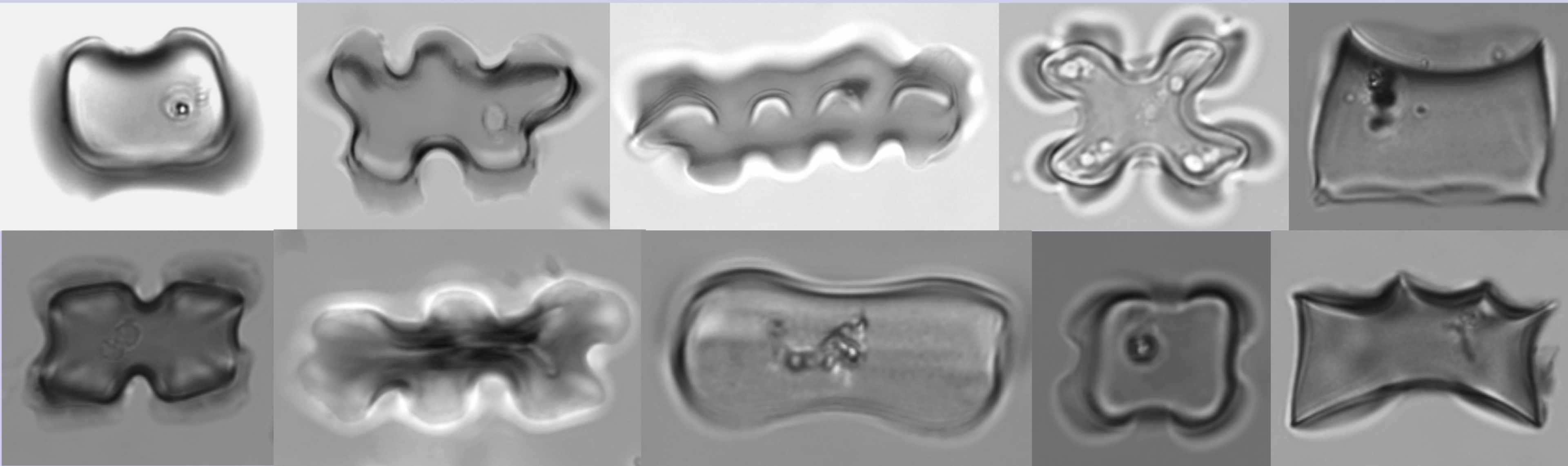
Kevin Jamieson

Other students:
Jifan Zhang, Xiangyun Meng, Claire Grant

RRF

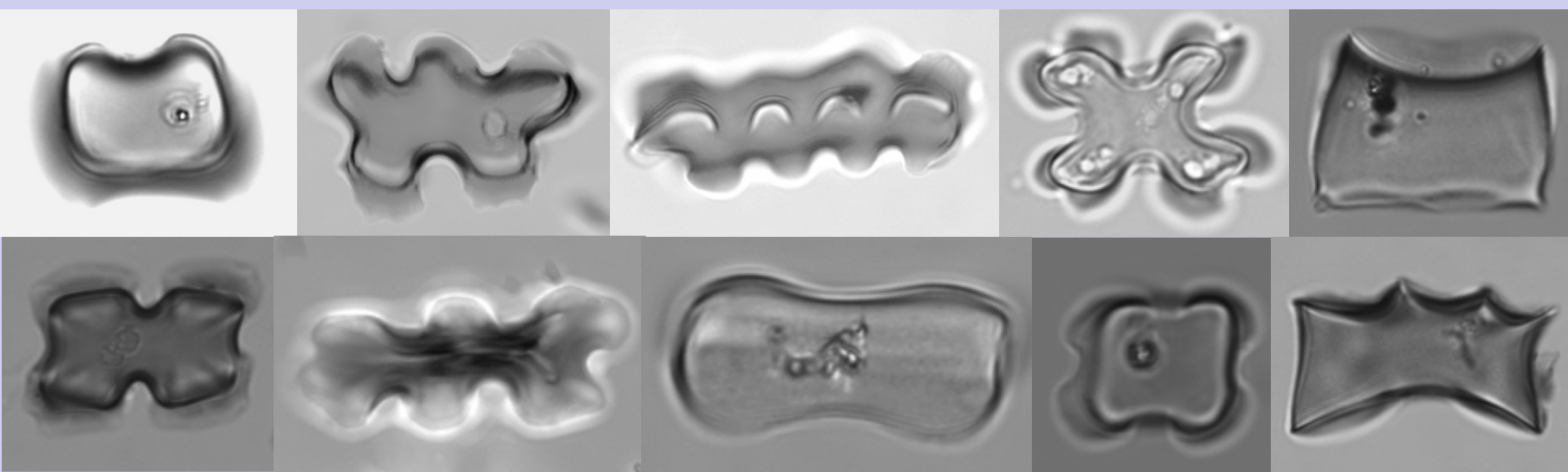
Machine Learning & Computer Vision (ML&CV)

- Uses training set of 2,480 2-D brightfield images of GSSCP from modern grasses
- Neural network trained to separate taxonomic groups using features of the images
- No alignment or landmarks needed



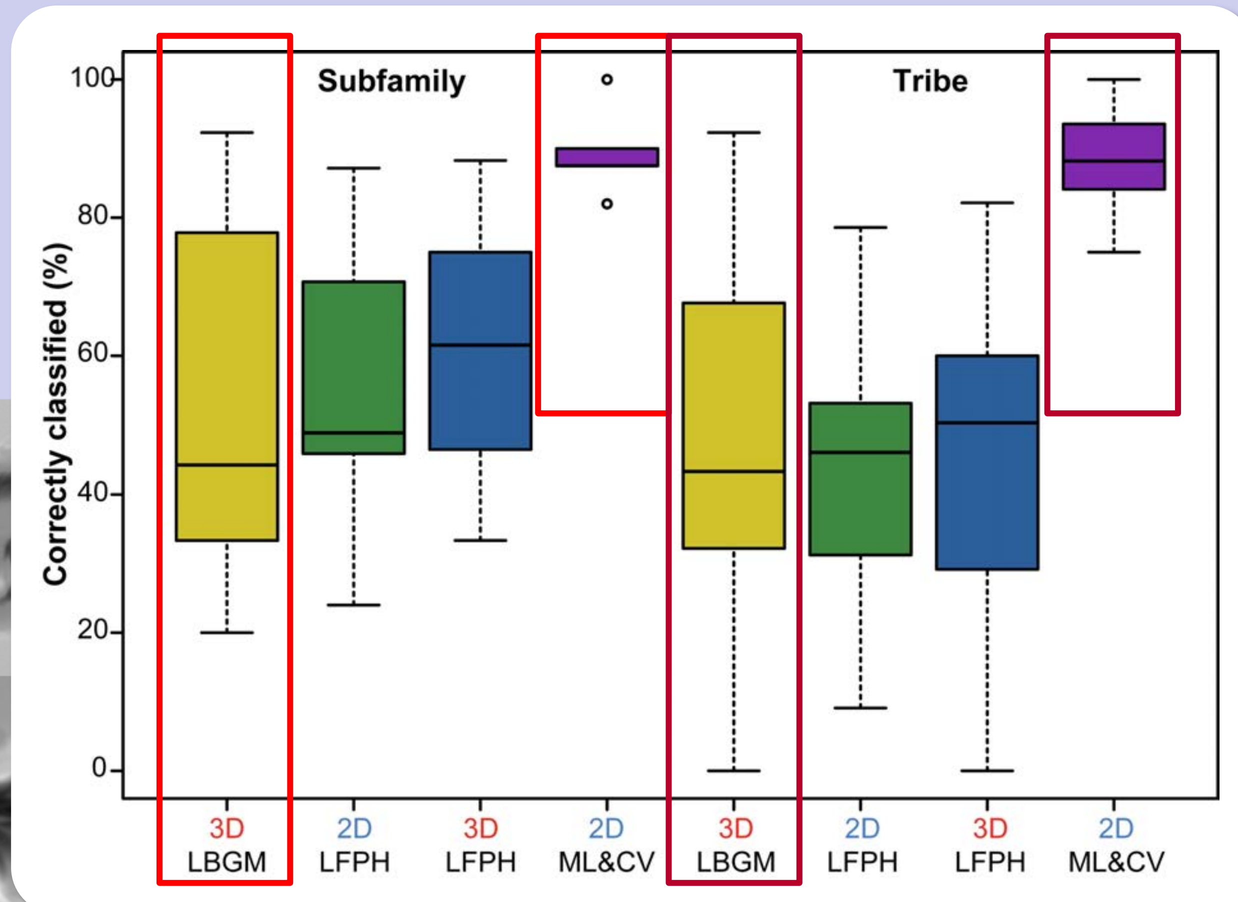
Machine Learning & Computer Vision (ML&CV)

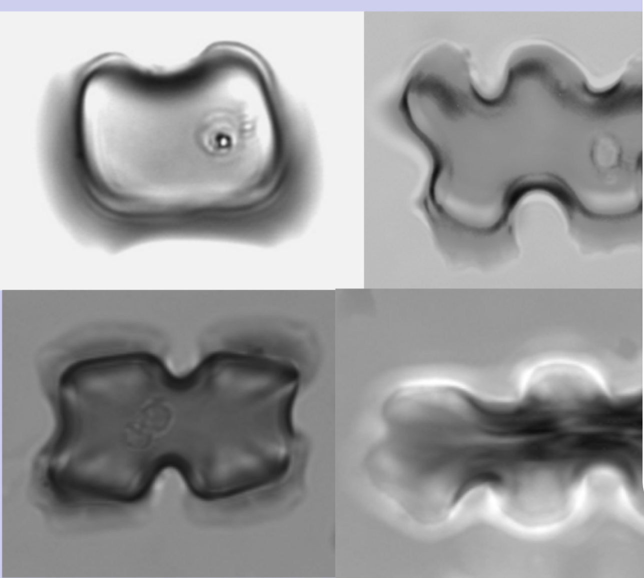
- Image traits cannot easily be extracted (“black box”)
- Not currently possible to incorporate phylogeny



Machine Learning & Computer Vision (ML&CV)

- Separation into subfamily and tribe superior to other (2-D and 3-D) methods
- Working to add statistic akin to confidence interval

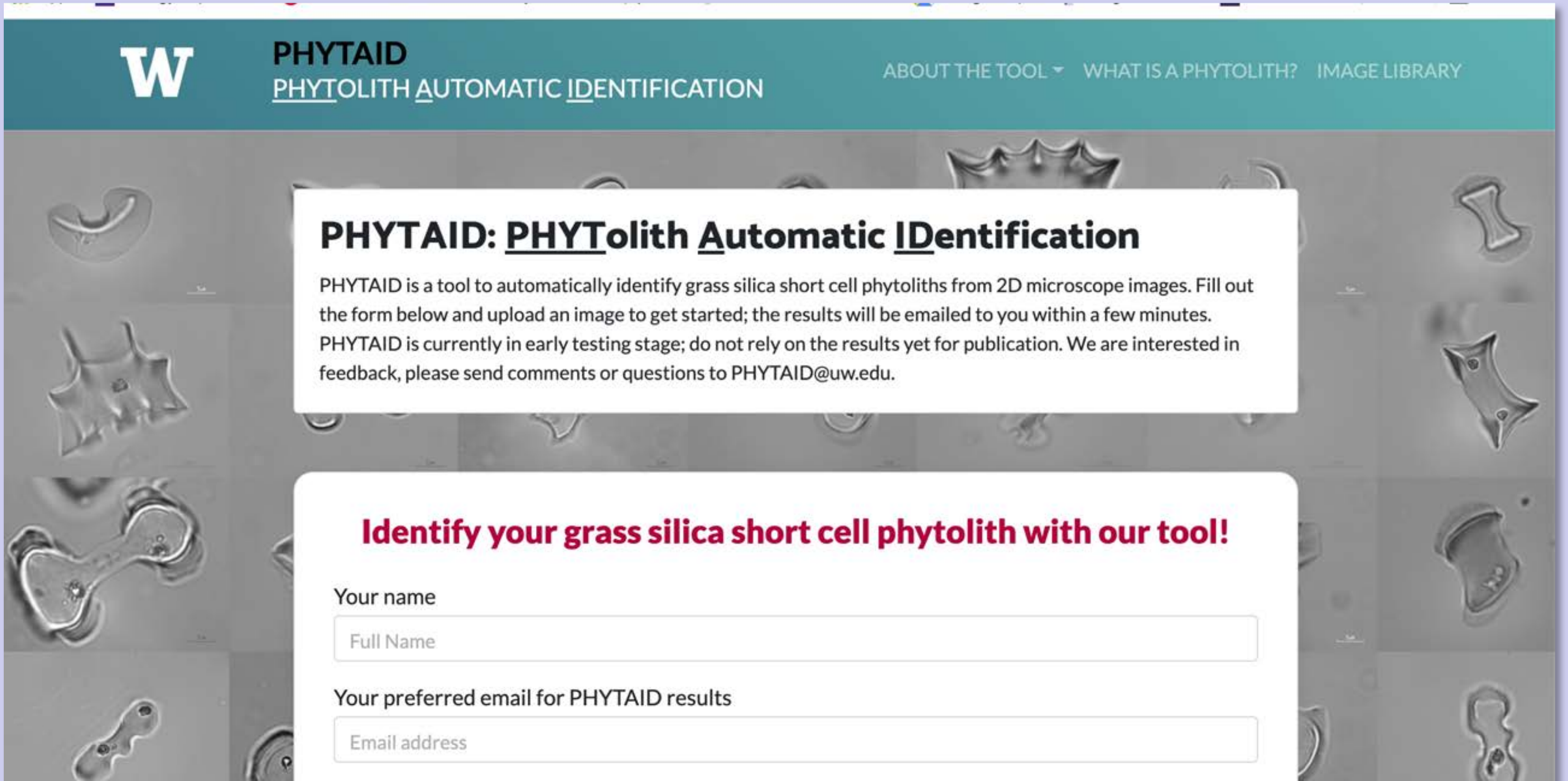




Sharing data

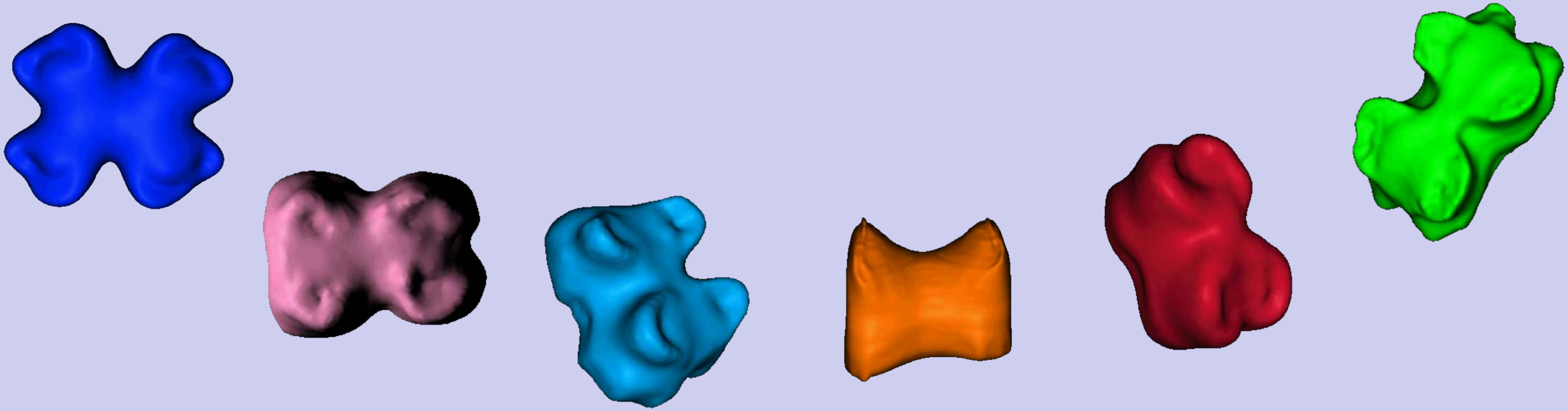
PHYTAID (*under construction*):

- Online platform where people can upload images and use ML&CV tool
- … and download (2-D, 3-D) images, videos, printable models etc.



Conclusions

- Quantitative methods (e.g., 3-D morphometrics, machine learning) have a great potential for making taxonomic placement of grass phytoliths more precise and robust
- Ultimately this will enhance research on grasses in paleobotany, paleoecology and archaeobotany



http://www.burkemuseum.org/phytaid

Acknowledgements

Strömberg Lab

Grad students:
Camilla Crifo
Will Brightly

Postdocs:
Georgina Erra
Tim Gallaher

Undergrads & volunteers:
Thien-Y Le
Kim Smith
Ryan Thummel
Nik Pershing
Anna Schorr
Elie Aboulafia
Brittany McManus
Casey O'Keefe
Ashly Senske
Erin Sofonowski
Claire Marvet
Sultan Akbar

Collaborators

Vandana Prasad
Dhananjay Mohabey
Ashok Sahni
Adam Leaché
Sue Hartley
Erika Edwards
Colin Osborne
Pascal-Antoine Christin
Elizabeth Kellogg
C4 Grasslands Working Group
Doris Barboni
Rosa Maria Albert
Katharina Neumann
Luc Vrydaghs
Rich Kay

Funding

Burke MUSEUM UNIVERSITY of WASHINGTON
UWBIOLOGY
NATIONAL GEOGRAPHIC SOCIETY INCORPORATED A.D. 1888
NSF
National Evolutionary Synthesis Center NESCent
UNIVERSITY OF WASHINGTON LVX SIT 1861
RRF

Thank you for your attention!