Reconstructing the extinction dynamics of *Picea critchfieldii*: Application of computer vision to fossil pollen analysis

Surangi W. Punyasena

Department of Plant Biology, University of Illinois

Shu Kong, Charless C. Fowlkes

Department of Computer Science, University of California, Irvine

and Stephen T. Jackson

Southwest Climate Science Center, US Geological Survey

BIO-DBI – ADVANCES IN BIOLOGICAL INFORMATICS

BIO-DBI – INNOVATIONS IN BIOLOGICAL IMAGING AND VISUALIZATION

Fossil pollen =
a microscopic census of
past vegetation,
preserved in
geologic sediments

POLLEN APPLICATIONS

Biostratigraphy

Paleoclimate

Paleoecology

Plant evolution

Forensics

DATA QUANTITY

How do we transform pollen analysis into a higher-throughput “big data” discipline?

DATA REPRODUCIBILITY

How do we improve the consistency and accuracy of pollen identifications?

TAXONOMIC RESOLUTION

How do we produce accurate, repeatable *species* identifications from pollen?

DATA QUANTITY

How do we transform pollen analysis into a higher-throughput “big data” discipline?

DATA REPRODUCIBILITY

How do we improve the consistency and accuracy of pollen identifications?

TAXONOMIC RESOLUTION

How do we produce accurate, repeatable *species* identifications from pollen?

DATA QUANTITY

How do we transform pollen analysis into a higher-throughput “big data” discipline?

DATA REPRODUCIBILITY

How do we improve the consistency and accuracy of pollen identifications?

TAXONOMIC RESOLUTION

How do we produce accurate, repeatable *species* identifications from pollen?

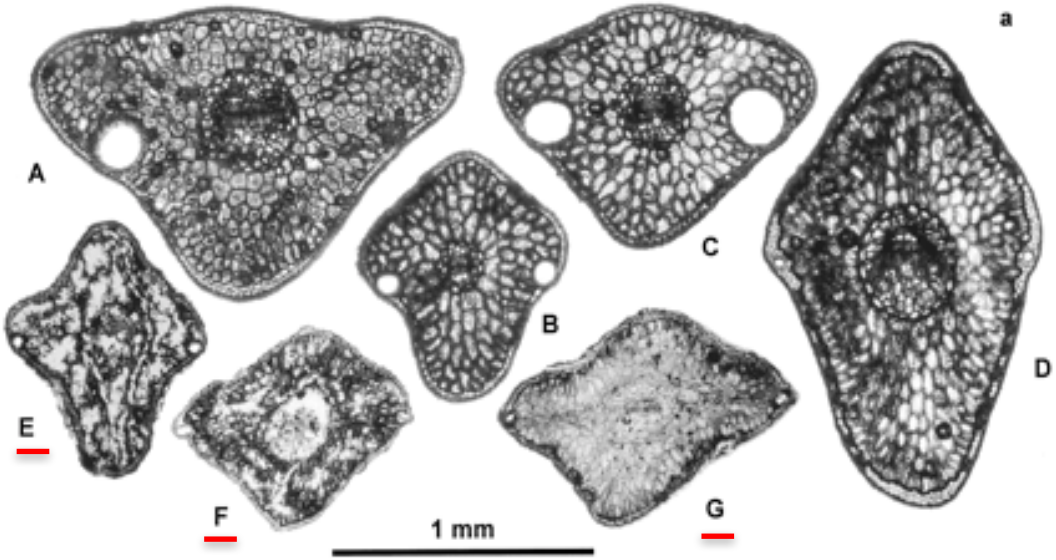
Late Quaternary extinction of a tree species in eastern North America

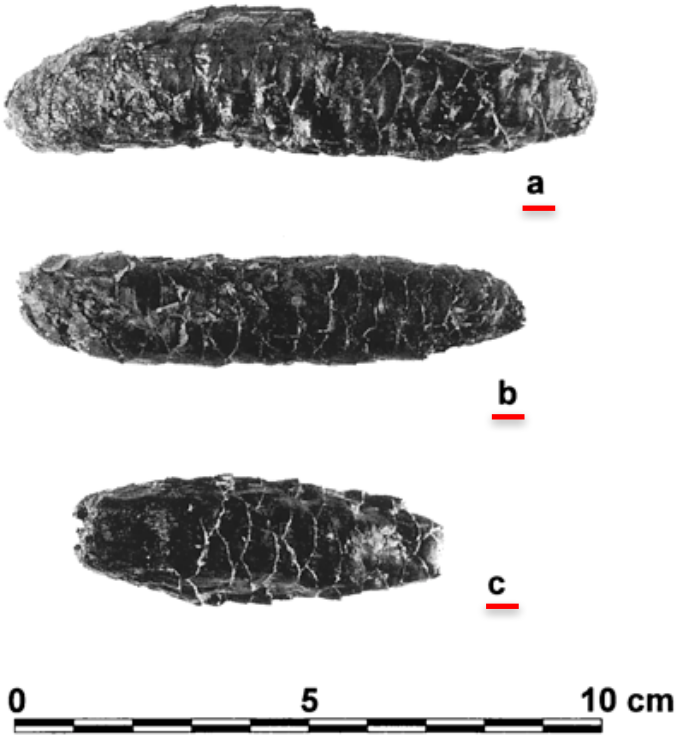
Stephen T. Jackson* and Chengyu Weng†

Department of Botany, University of Wyoming, Laramie, WY 82071

Edited by Margaret Bryan Davis, University of Minnesota, St. Paul, MN, and approved September 27, 1999 (received for review July 8, 1999)

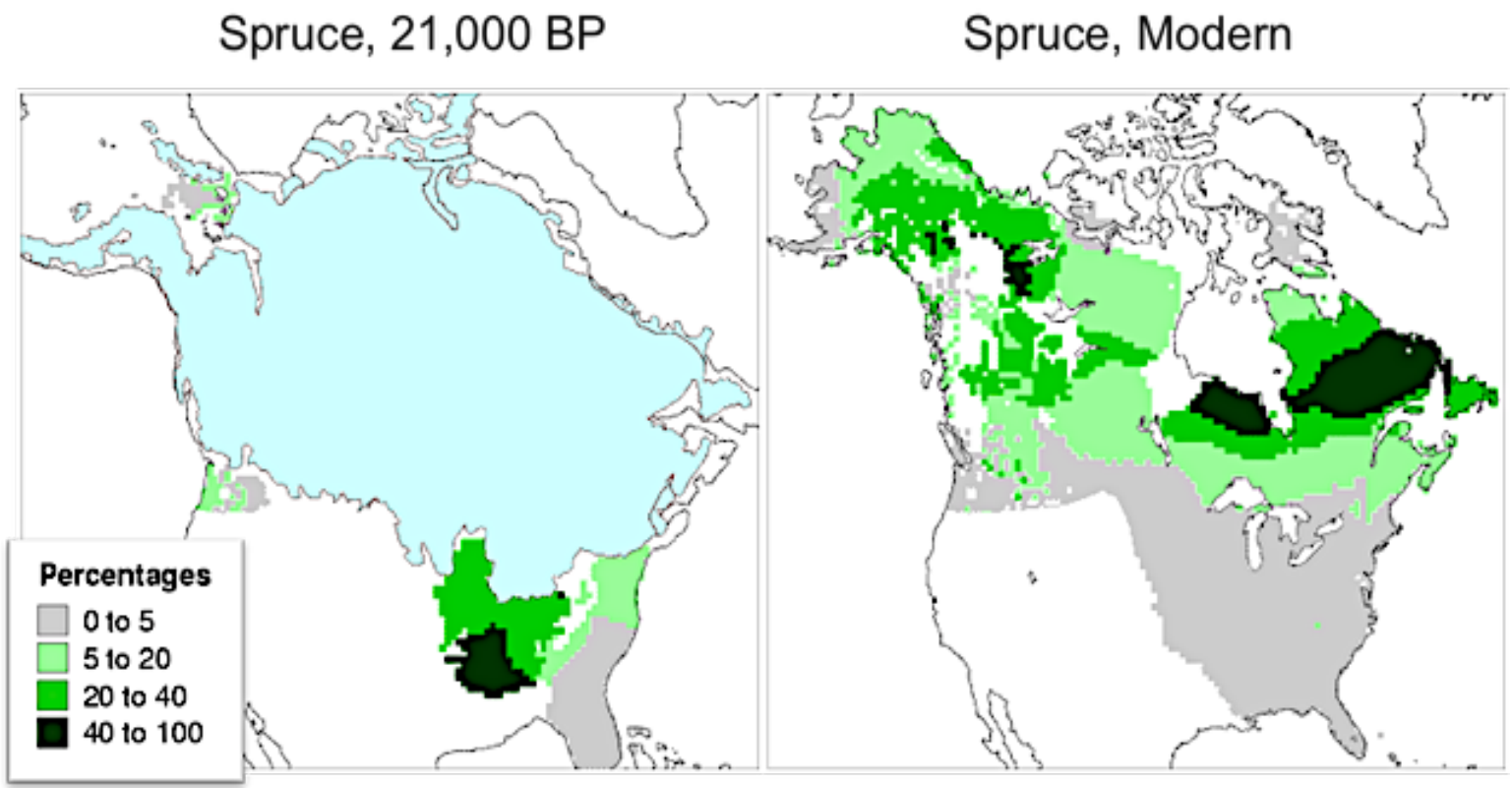
Proceedings of the National Academy of Sciences 1999





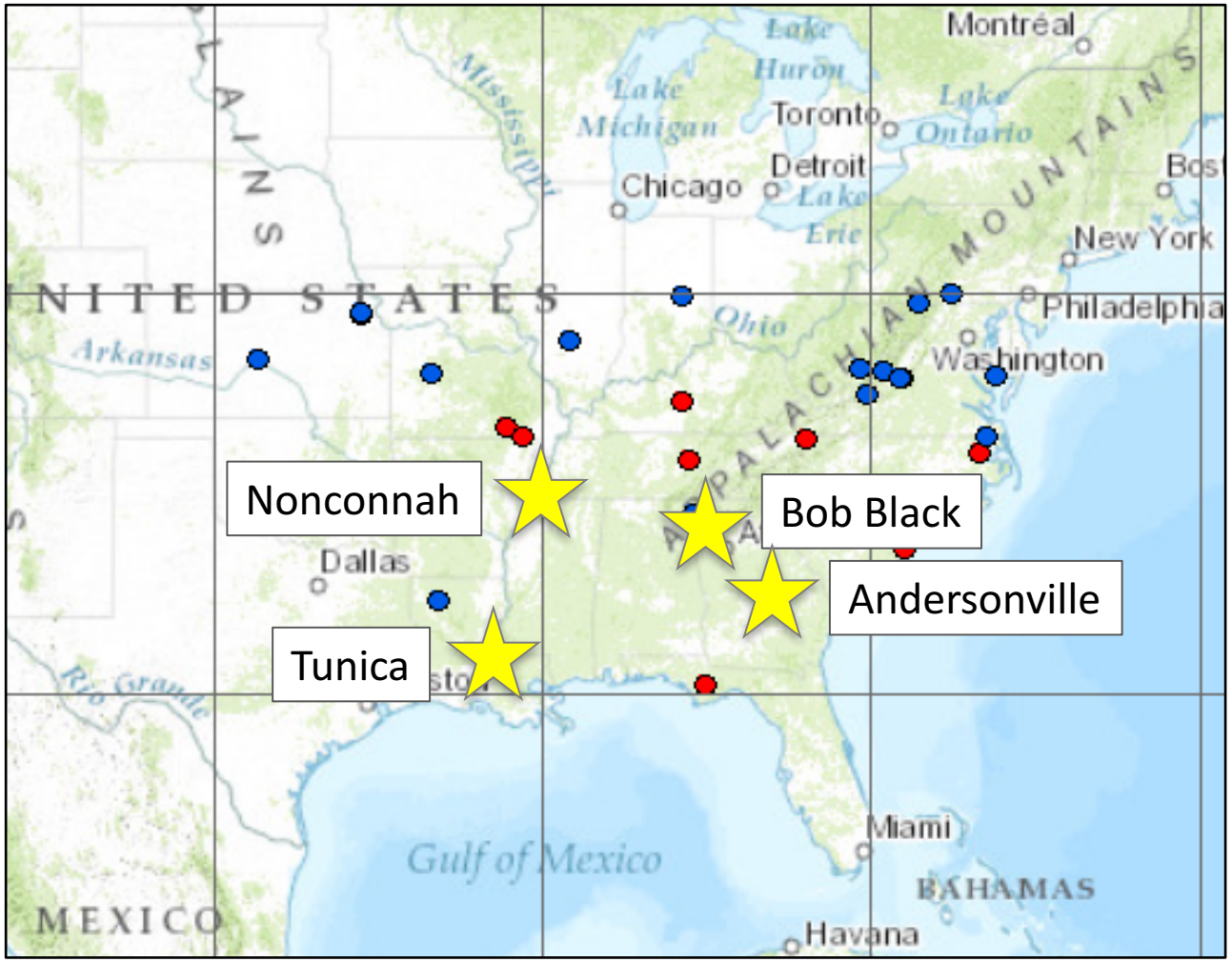
Mauricio Anton

SPRUCE PALEOBIOGEOGRAPHY



Pollen Viewer (Williams et al. 2004)

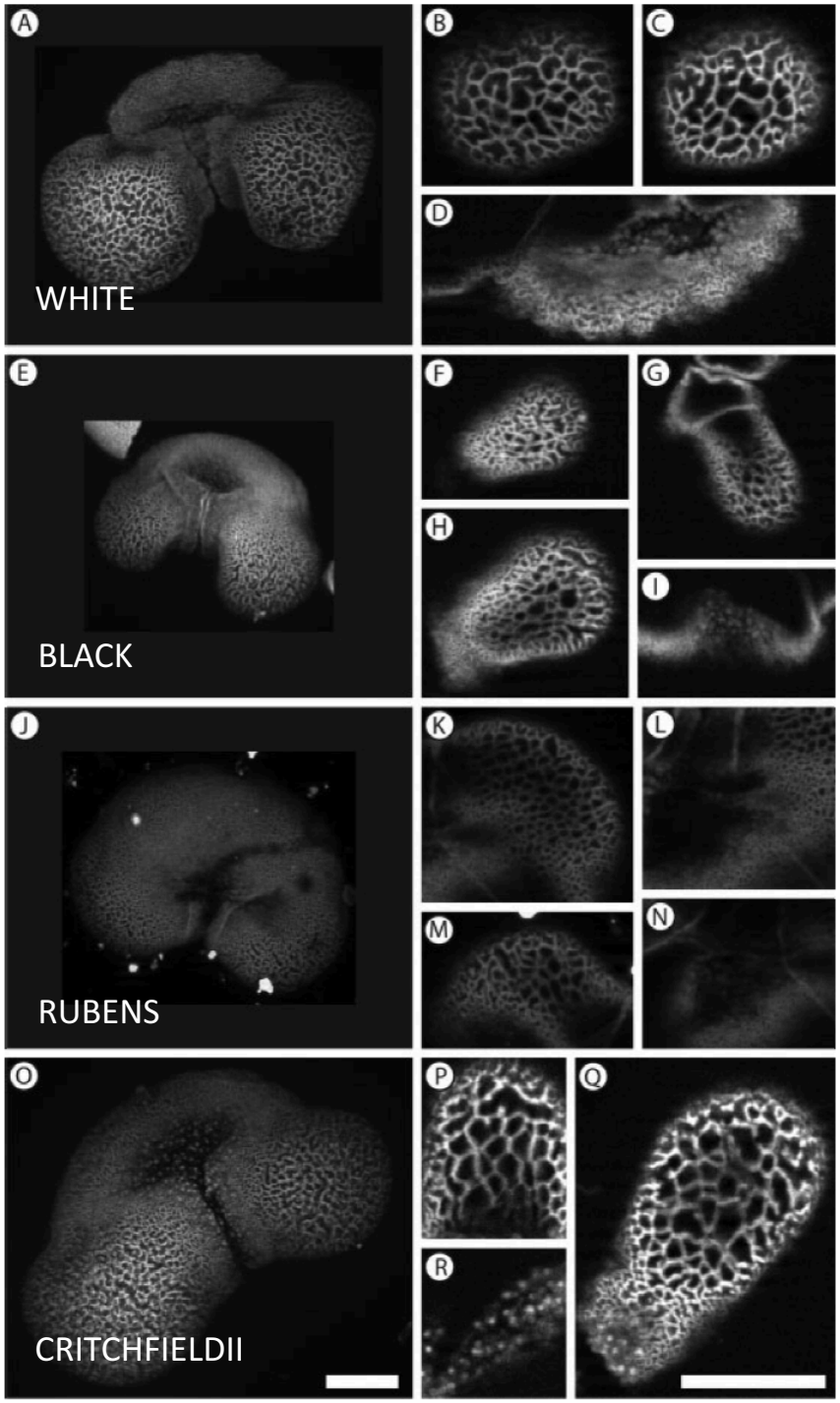
CRITCHFIELDII PALEOBIOGEOGRAPHY

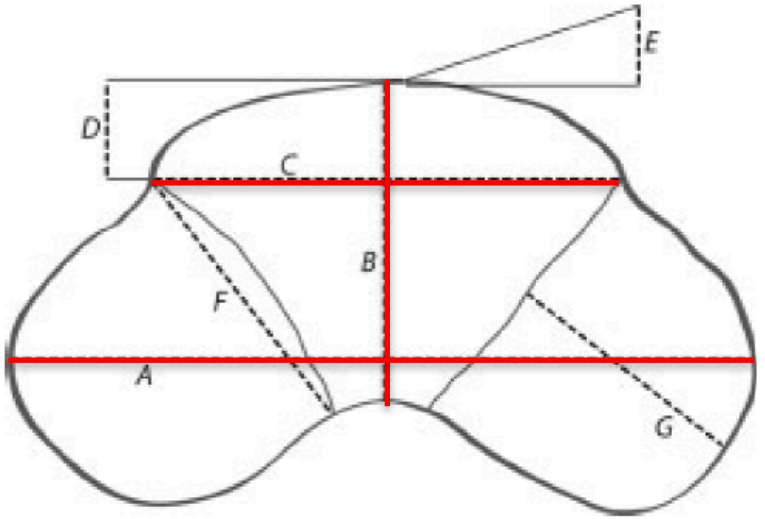


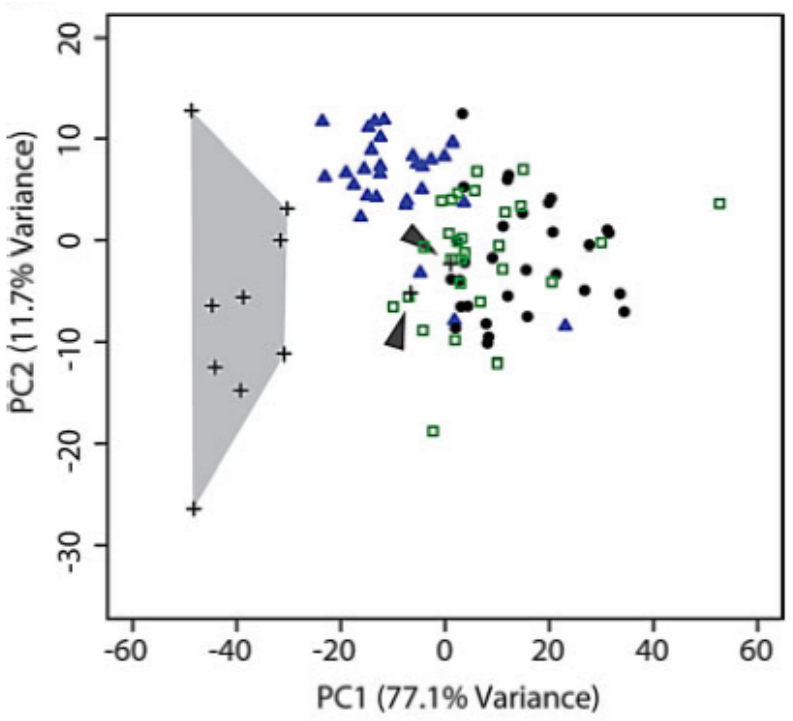
Critchfieldii presence established based on macrofossils

CRITCHFIELDII PALEOBIOGEOGRAPHY

Reconstruct pollen percentage abundances through time

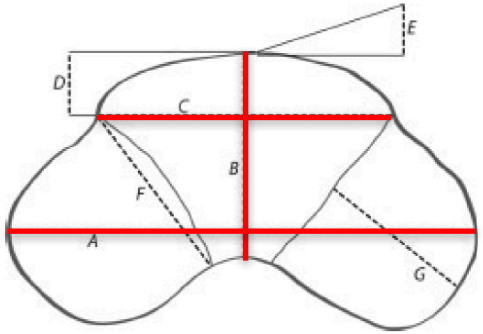






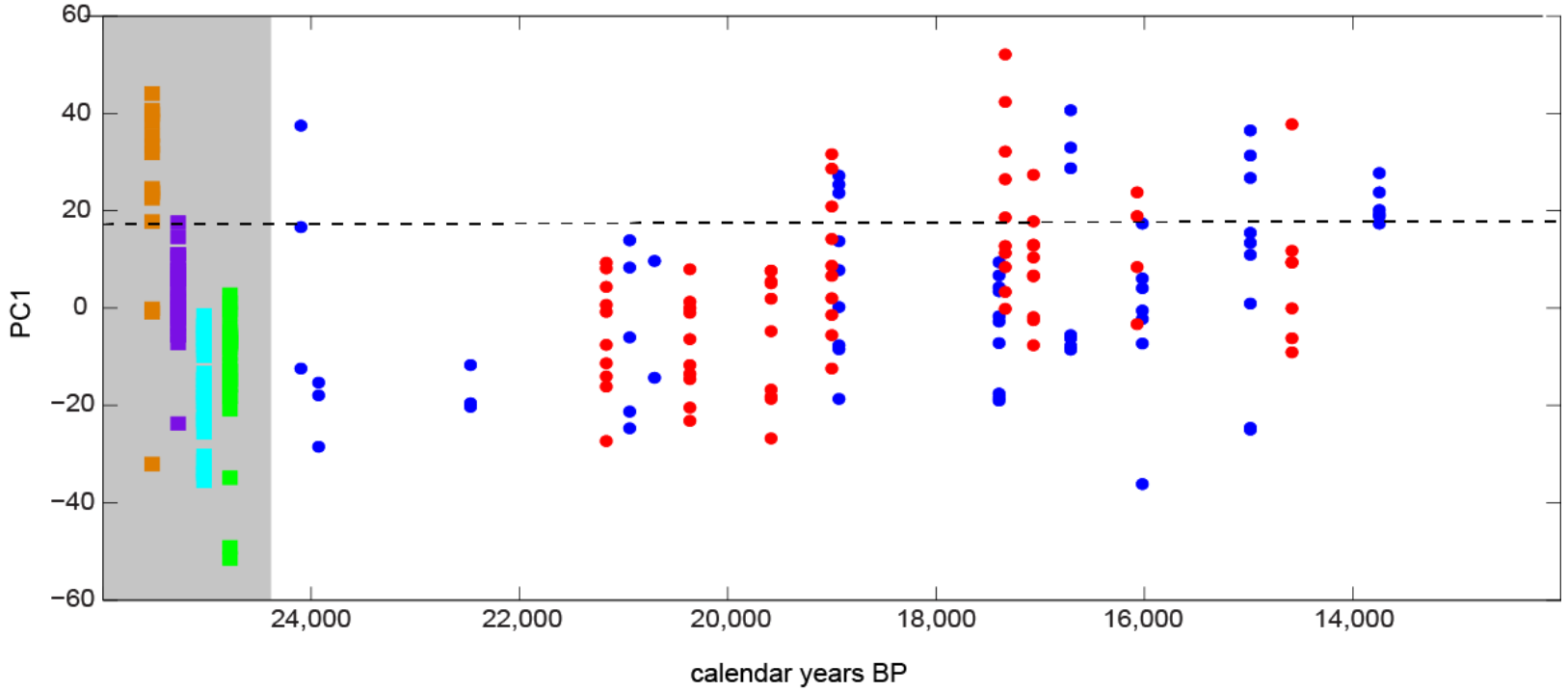
Mander, Rodriguez, Mueller, Jackson and Punyasena, 2014

RECONSTRUCTING PALEO-POPULATION DYNAMICS

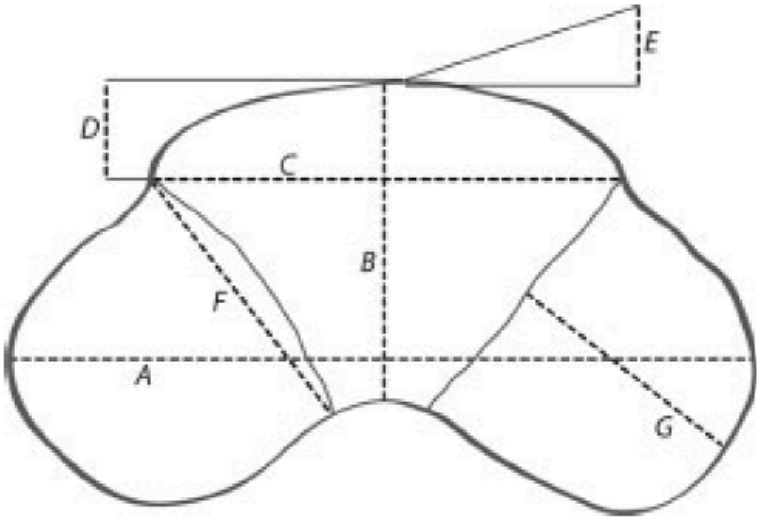


Anderson Pond, TN

Cupola Pond, MO



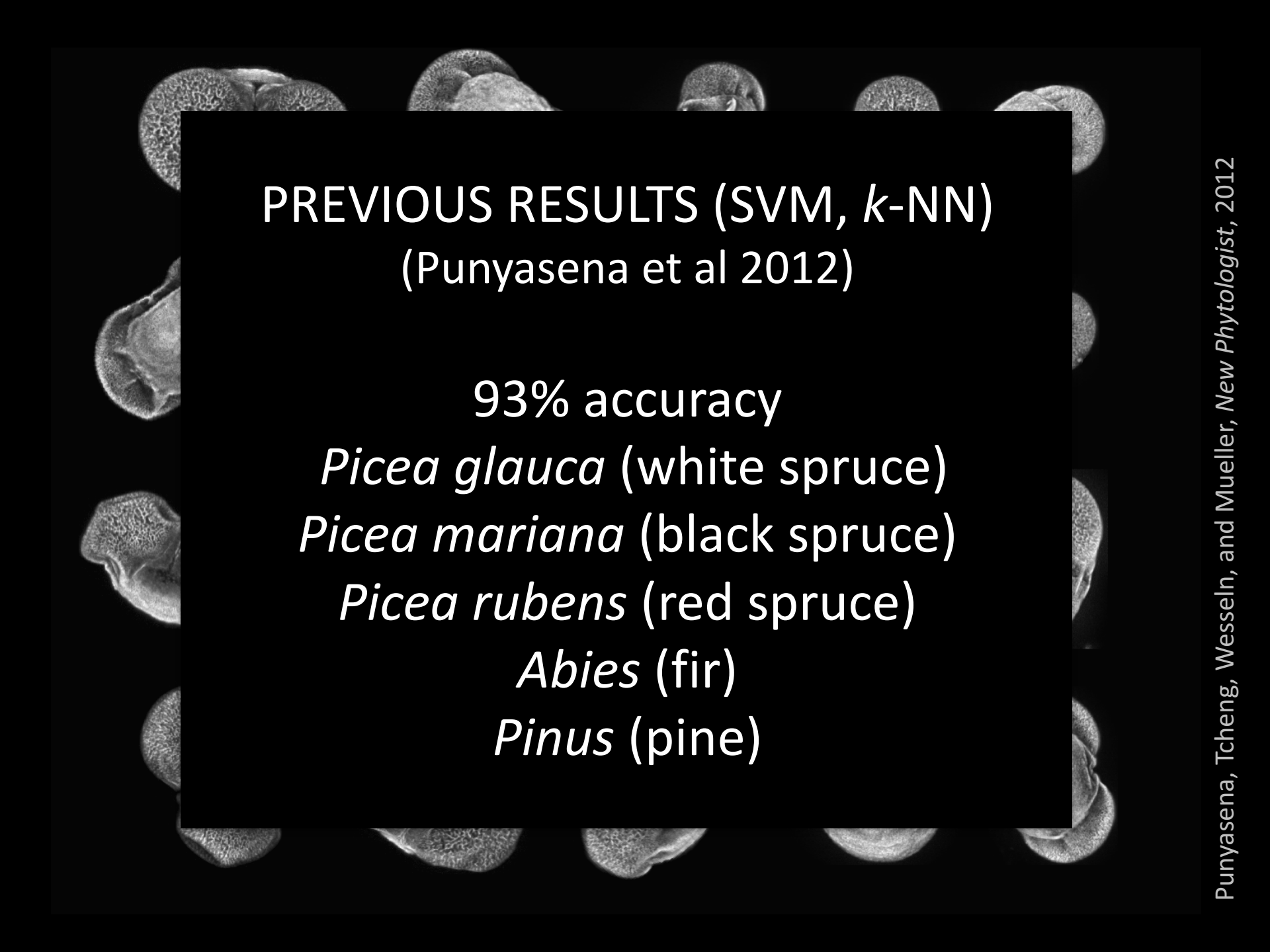
Punyasena, Rodriguez, and Jackson, unpublished



Morphometric approaches only allow us to classify grains in a specific orientation and ignores qualitative features (ornamentation, endoreticulation).

Collaborators at UC Irvine, Computer Science (Charless Fowlkes and Shu Kong) experimented with "deep learning" convolution neural nets as an alternative.

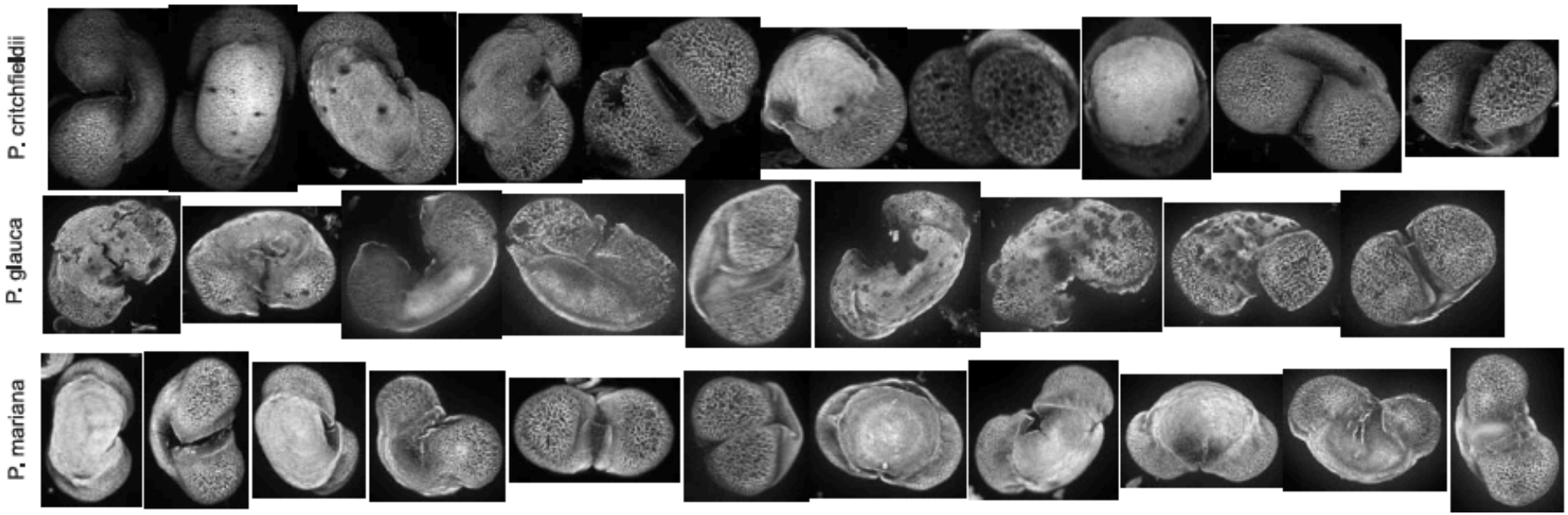
Punyasena, Tcheng, Wesseln, and Mueller, *New Phytologist*, 2012


PREVIOUS RESULTS (SVM, *k*-NN)
(Punyasena et al 2012)
93% accuracy
Picea glauca (white spruce)
Picea mariana (black spruce)
Picea rubens (red spruce)
Abies (fir)
Pinus (pine)
Punyasena, Tcheng, Wesseln, and Mueller, *New Phytologist*, 2012

COMPUTATIONAL GOAL

Develop a robust classification system for fossil specimens

- Tolerant of damage present in fossil material
- Train on modern reference material (white/black spruce), classify fossils
- Generalized solution/not overfit
- Critical for practical large-scale applications

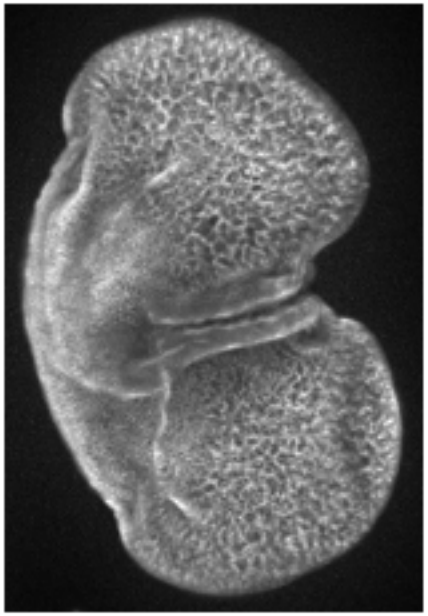


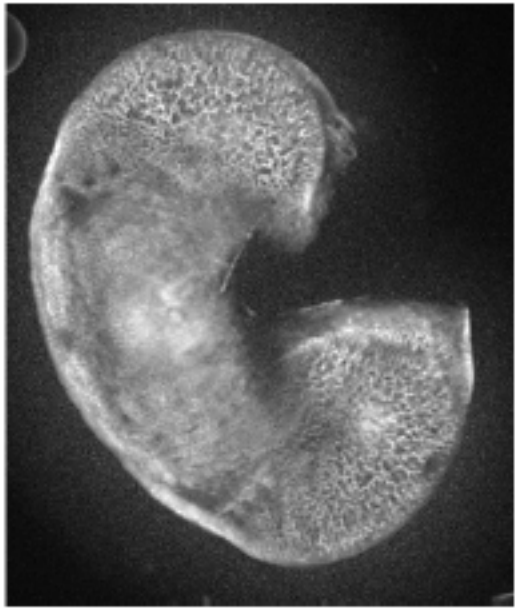
Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

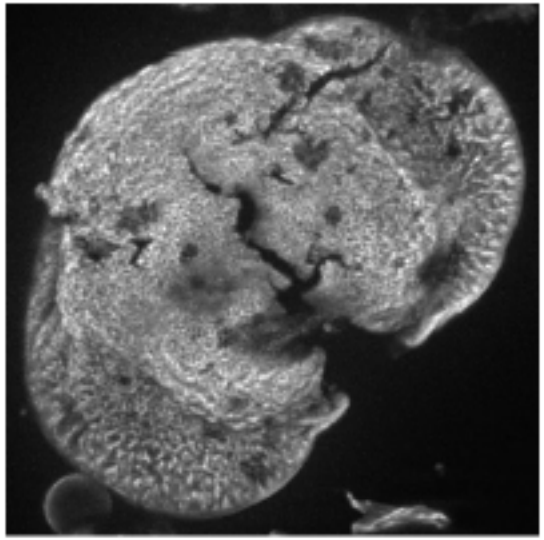
CLASSIFICATION PIPELINE

Cropped MIP Apotome fluorescence images

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016







MODERN REFERENCE
(TRAINING)

FOSSIL
(TESTING)

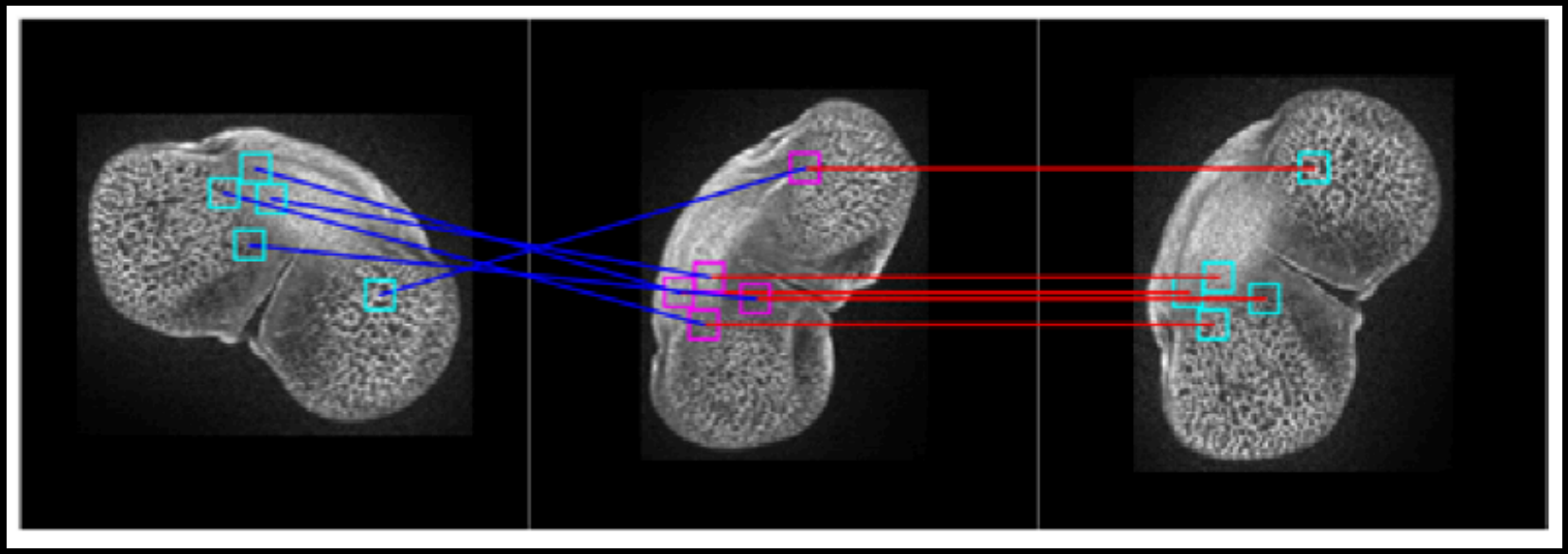
Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

CLASSIFICATION PIPELINE

Cropped MIP Apotome fluorescence images

Canonical rotations using
k-medoids clustering

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

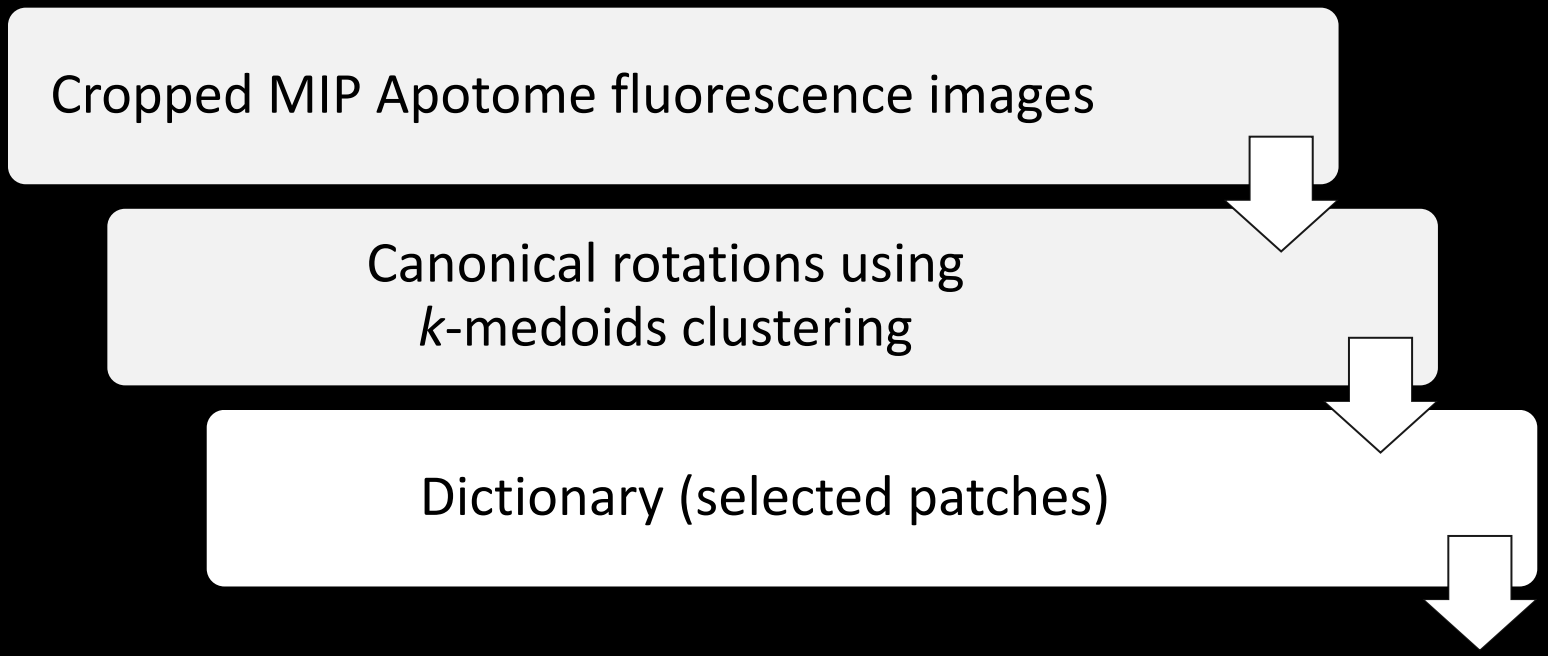


Only two representations were necessary:
equatorial and polar views.

Images were rotated to match canonical shapes.

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

CLASSIFICATION PIPELINE

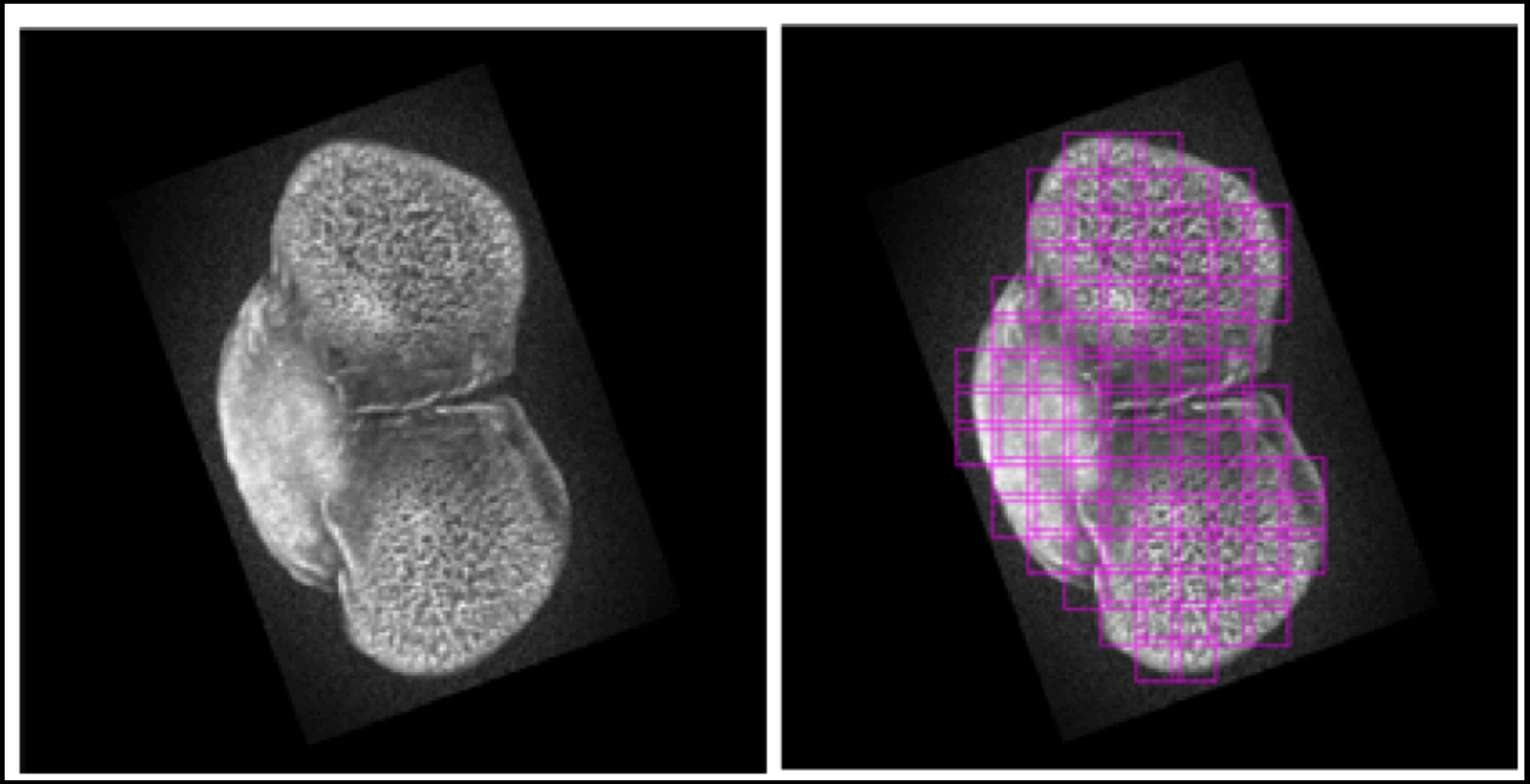


Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

Discriminative Patch Selection

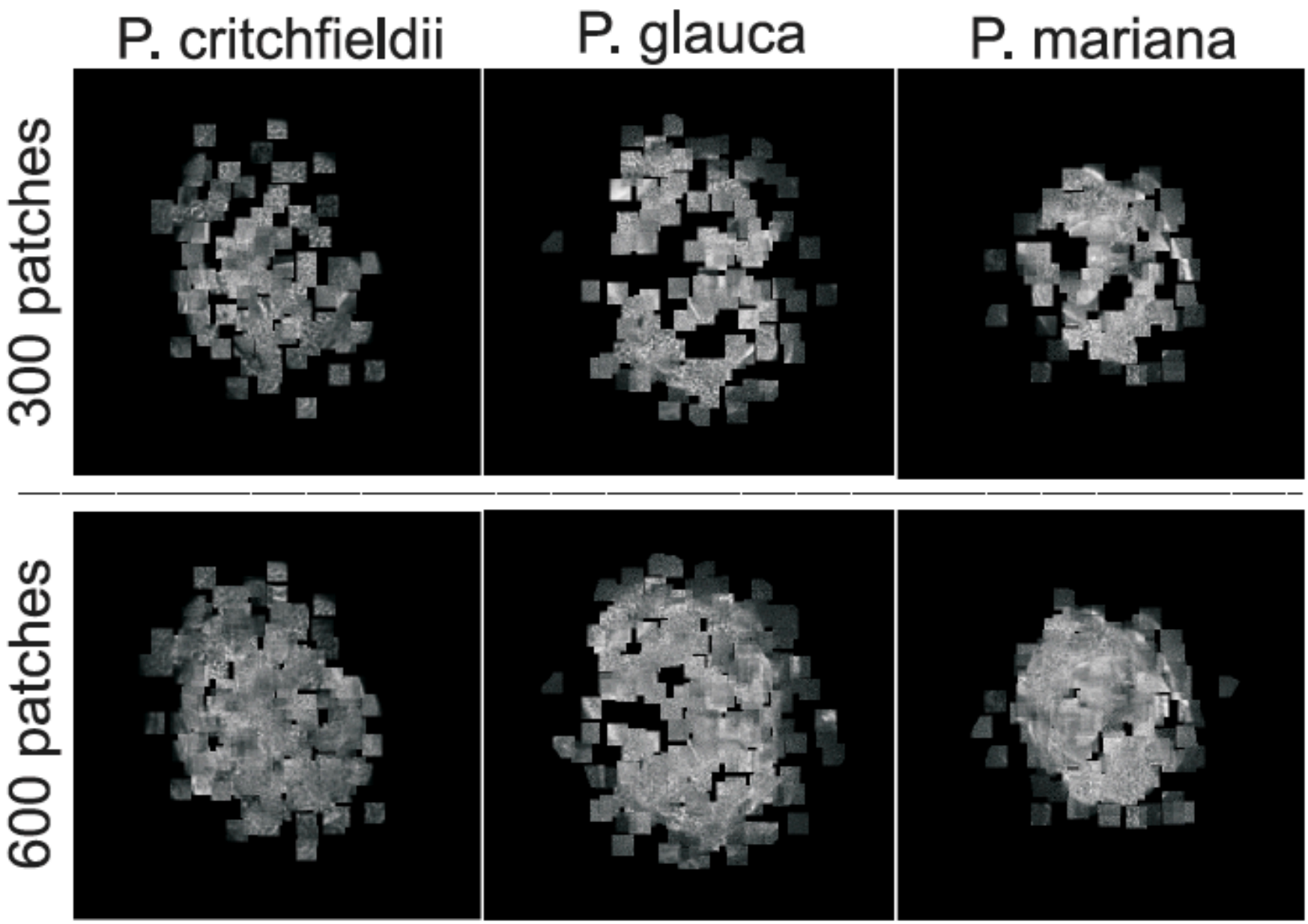
- Created a visual “dictionary”
  - Representation in feature space
  - Spatially distributed in input space
  - Discriminative power
  - Class balance
  - Cluster compactness
- Use greedy lazy forward selection

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016



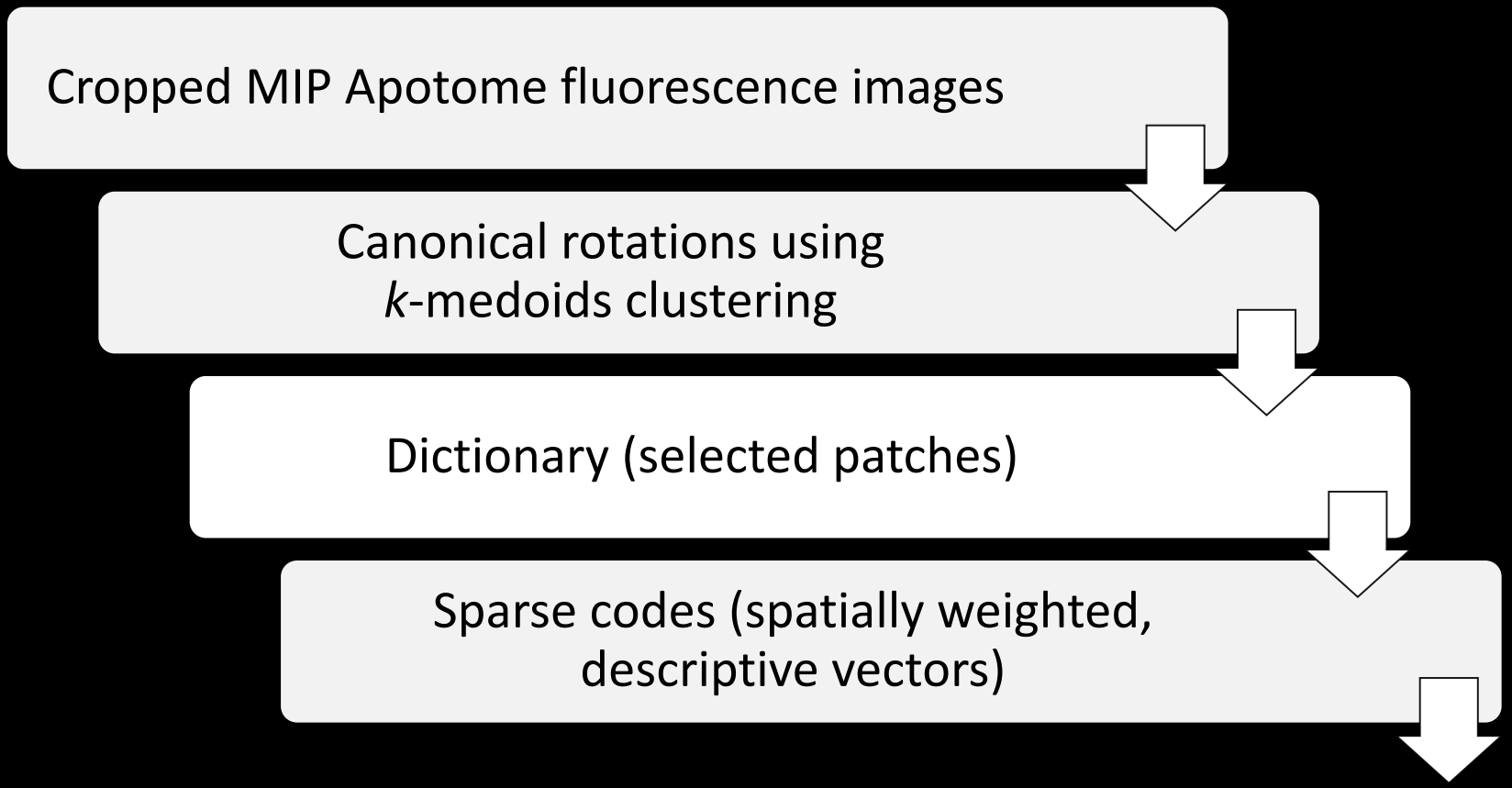
Overlapping patches (52 x 52 pixels)

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016



Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

CLASSIFICATION PIPELINE



Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

CLASSIFICATION PIPELINE

Sparse codes (spatially weighted, descriptive vectors)

Pre-trained CNN (VGGVeryDeep-19)

Extracted features at optimal layer (512-dimensional vectors)

Linear SVM classifier

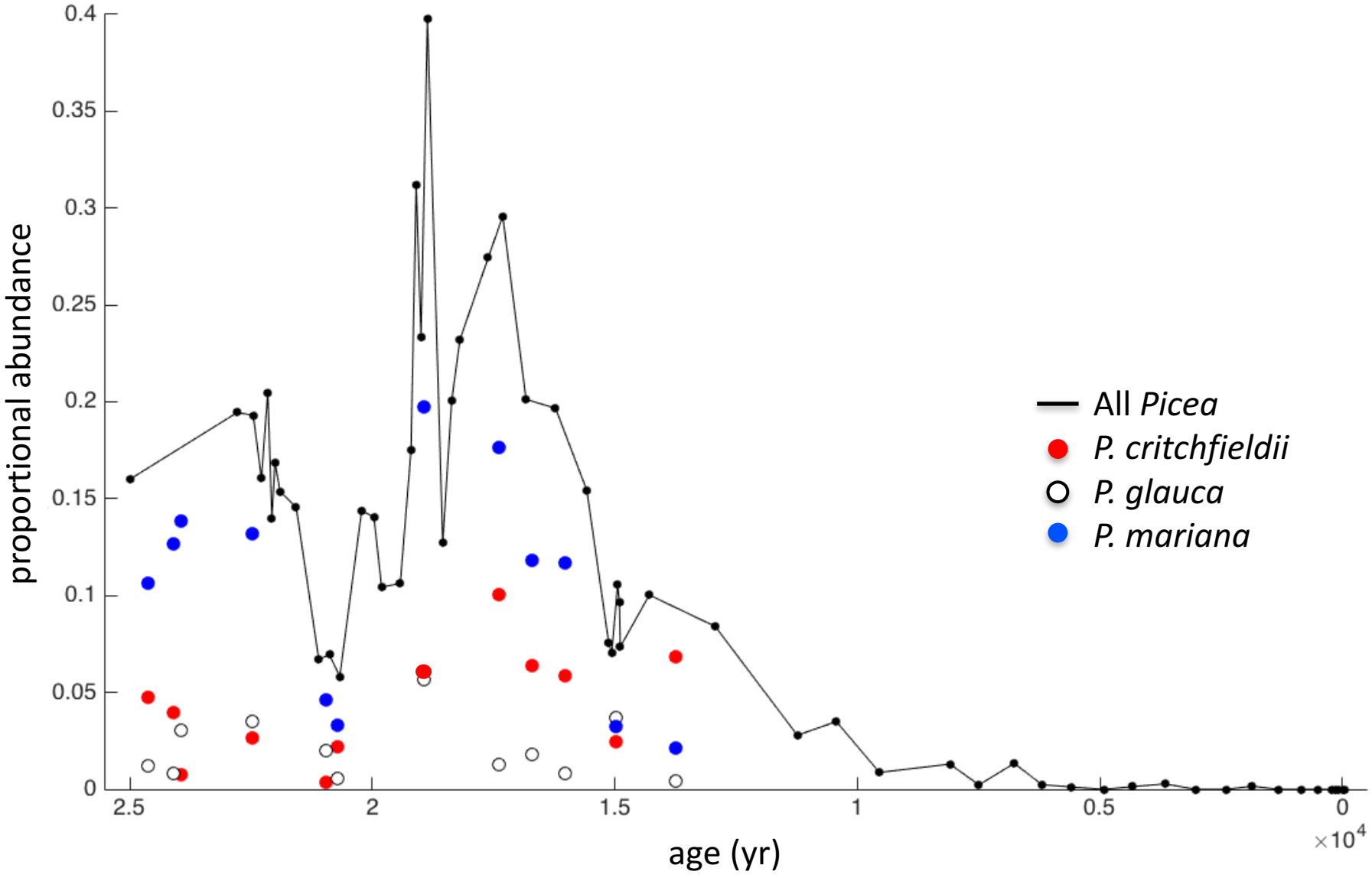
POLLEN CLASSIFICATIONS (86% ACCURACY)

Kong, Punyasena, and Fowlkes, IEEE Computer Vision and Pattern Recognition, 2016

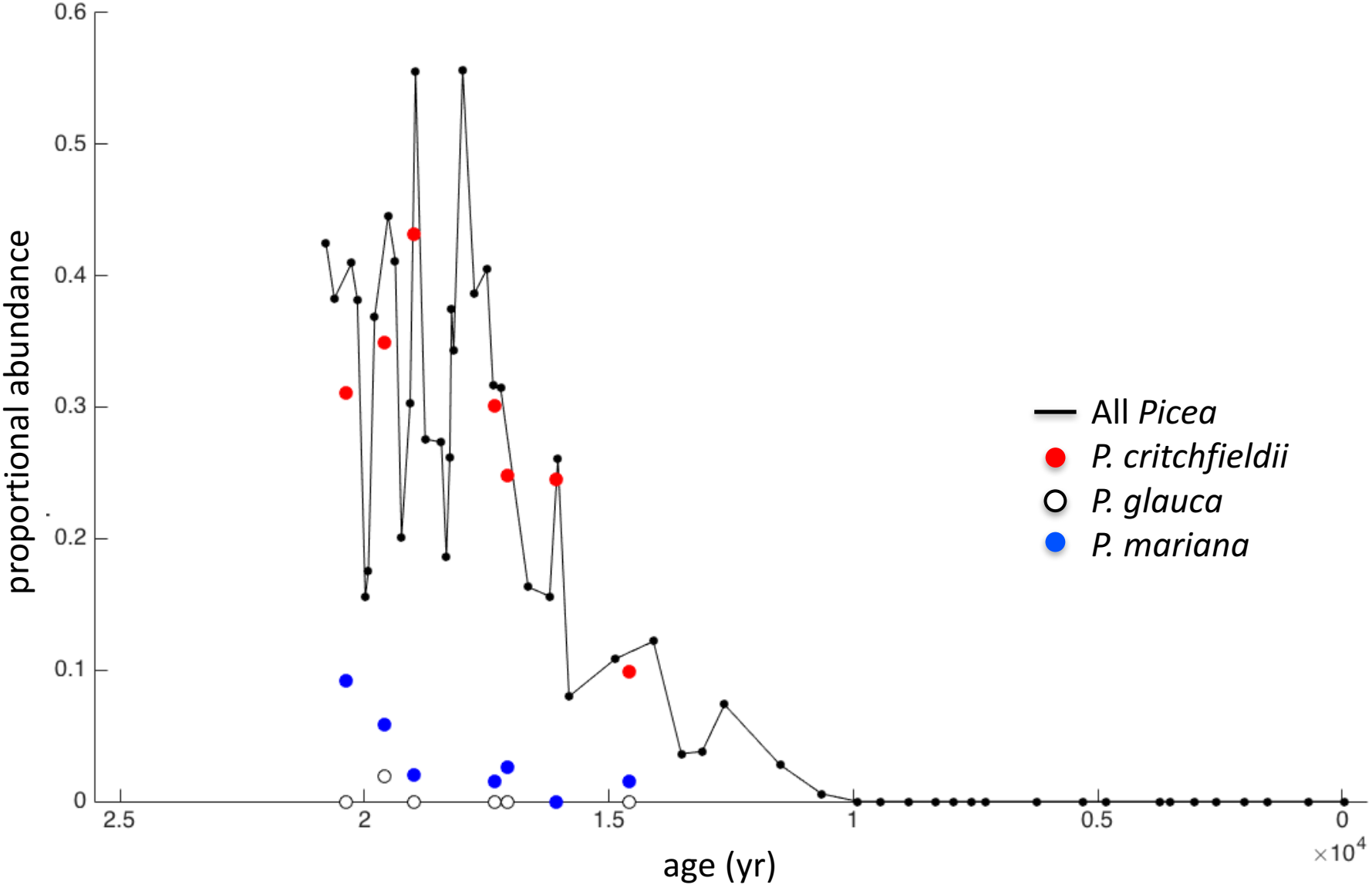
CRITCHFIELDII PALEOBIOGEOGRAPHY

Reconstruct pollen percentage abundances through time

ANDERSON POND, TN



CUPOLA POND, MO



SUMMARY

- *P. critchfieldii* was a significant component of the vegetation surrounding Anderson and Cupola
  - >40% of the pollen abundance of Cupola Pond, MO
- Expansion of spruce forests tied to an increase in *P. critchfieldii?*

NEXT STEPS

- Look at sites across southeastern and south-central US
  - Anecdotal reports of odd-looking spruce
- Better document the decline of *P. critchfieldii* and spruce forests (15 – 11 kyr)
- Relate differences in relative abundance of *P. critchfieldii* to climate

LESSONS LEARNED

- Analysis the result of a highly efficient, generalizable machine learning classification model
- Next-gen machine methods efficiently capture taxonomic (phylogenetic) information inherent in morphology
- Imaging more challenging than computational analyses
  - Need more images of vouchered, expertly identified specimens!
  - #ScanAllPollen

THANKS! Collaborators and Funders

Luke Mander

Open University, UK

Mayandi Sivaguru, Glenn Fried

Institute for Genomic Biology, UIUC

David Tcheng

Illinois Informatics Institute

Pietra Mueller

Illinois State Museum

Jackie Rodriguez

Cassie Wesseln

Ashwin Nayak

Ava Holz

UIUC IB 199 students

NSF-DBI – Advances in Biological Informatics

NSF-DBI/EF – Innovations in Biological Imaging & Visualization

National Center for Supercomputing Applications (NCSA) Institute for Advanced Computing Applications and Technologies (IACAT)

UIUC Campus Research Board

NSF XSEDE : Texas Advanced Computing Cluster (TACC); National Center for Supercomputing Applications (NCSA)