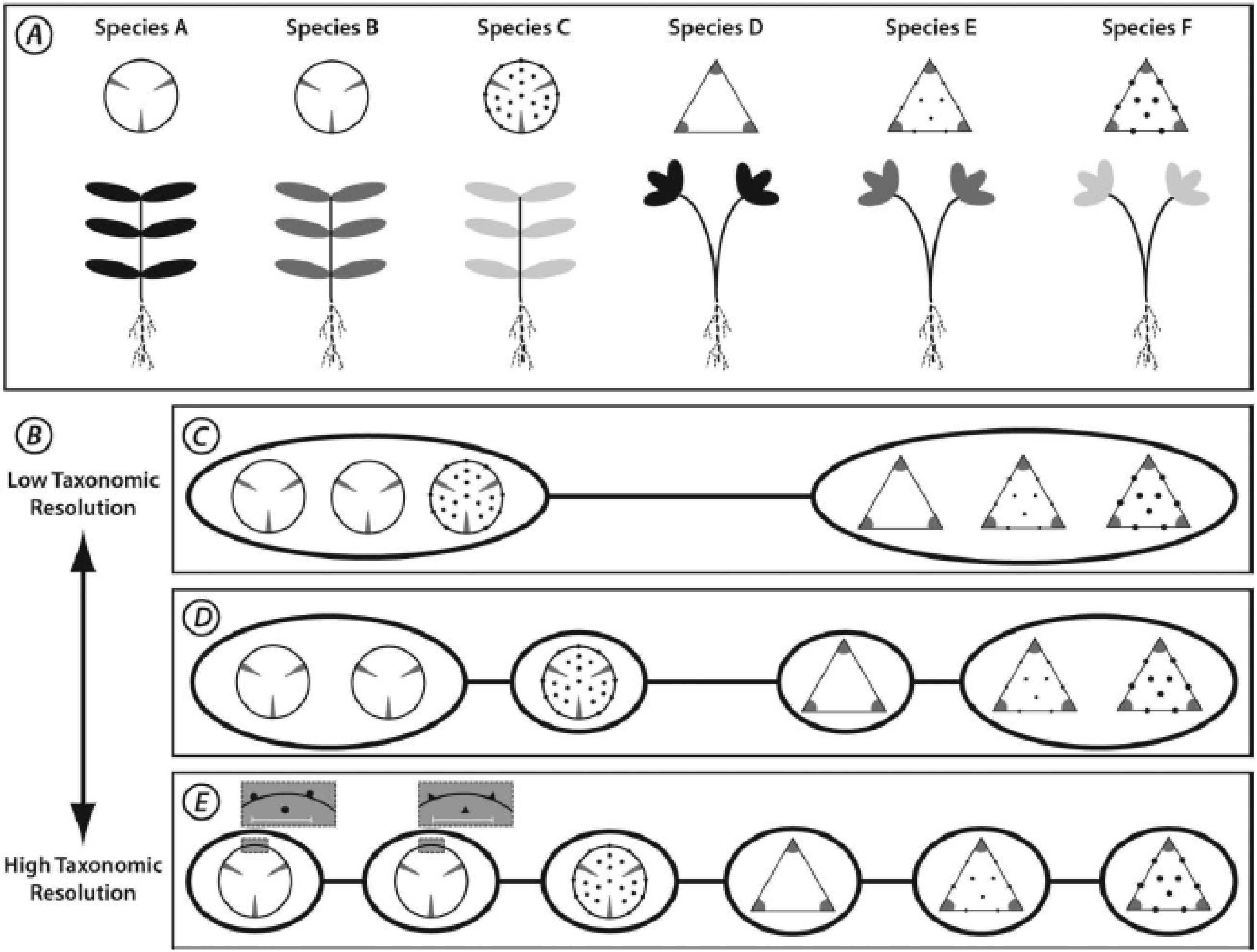
Identification of the Cenozoic pollen morphospecies *Striatopollis catatumbus* (Amherstieae, Fabaceae) using convolutional neural nets

Ingrid Romero[1], Shu Kong[2],
Charless C. Fowlkes[2], Surangi W. Punyasena[1]

[1]*Department of Plant Biology, University of Illinois at Urbana-Champaign*

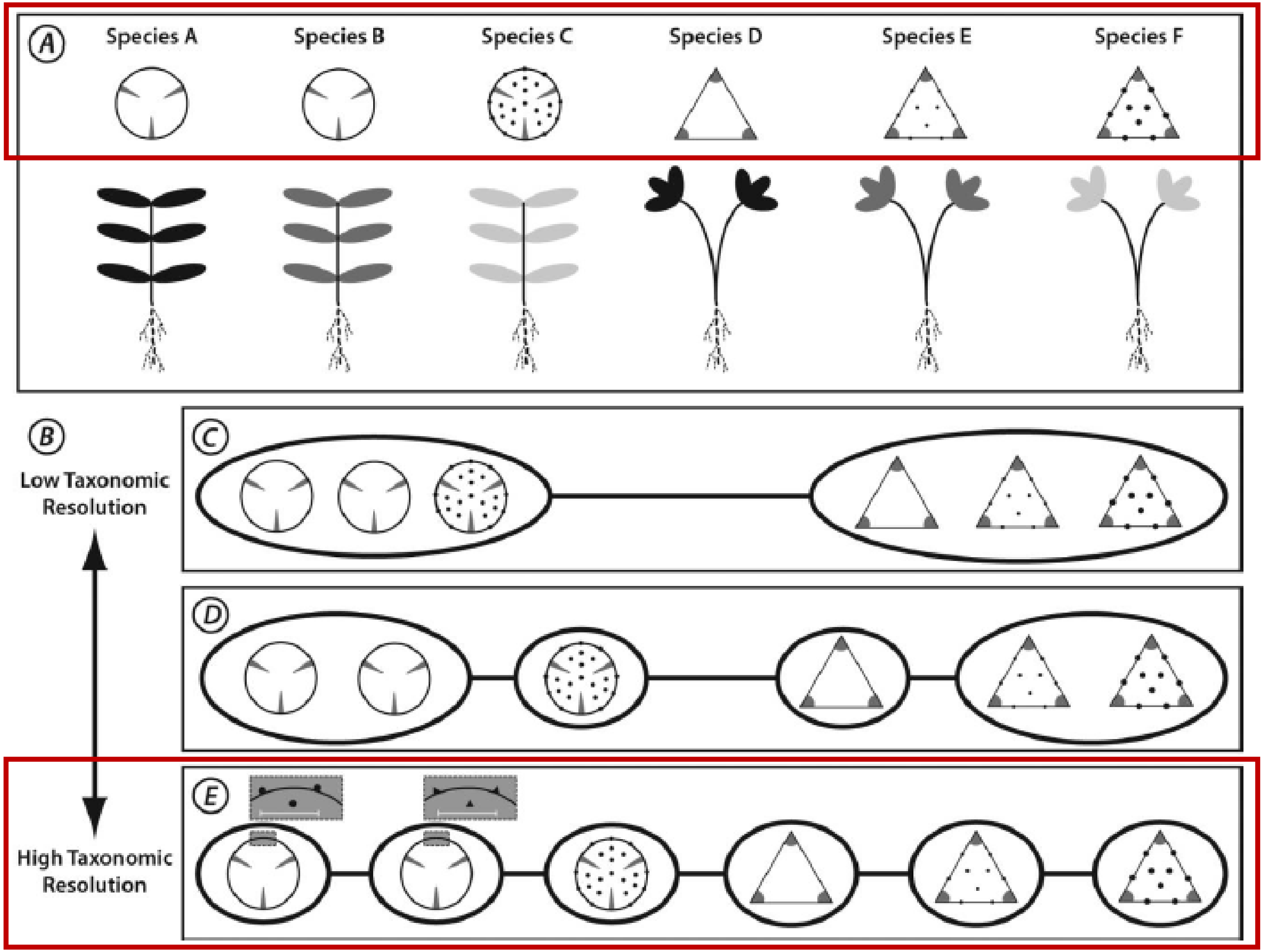
[2]*Department of Computer Science, University of California- Irvine*

Importance of increasing the taxonomic resolution



Mander and Punyasena (2016)

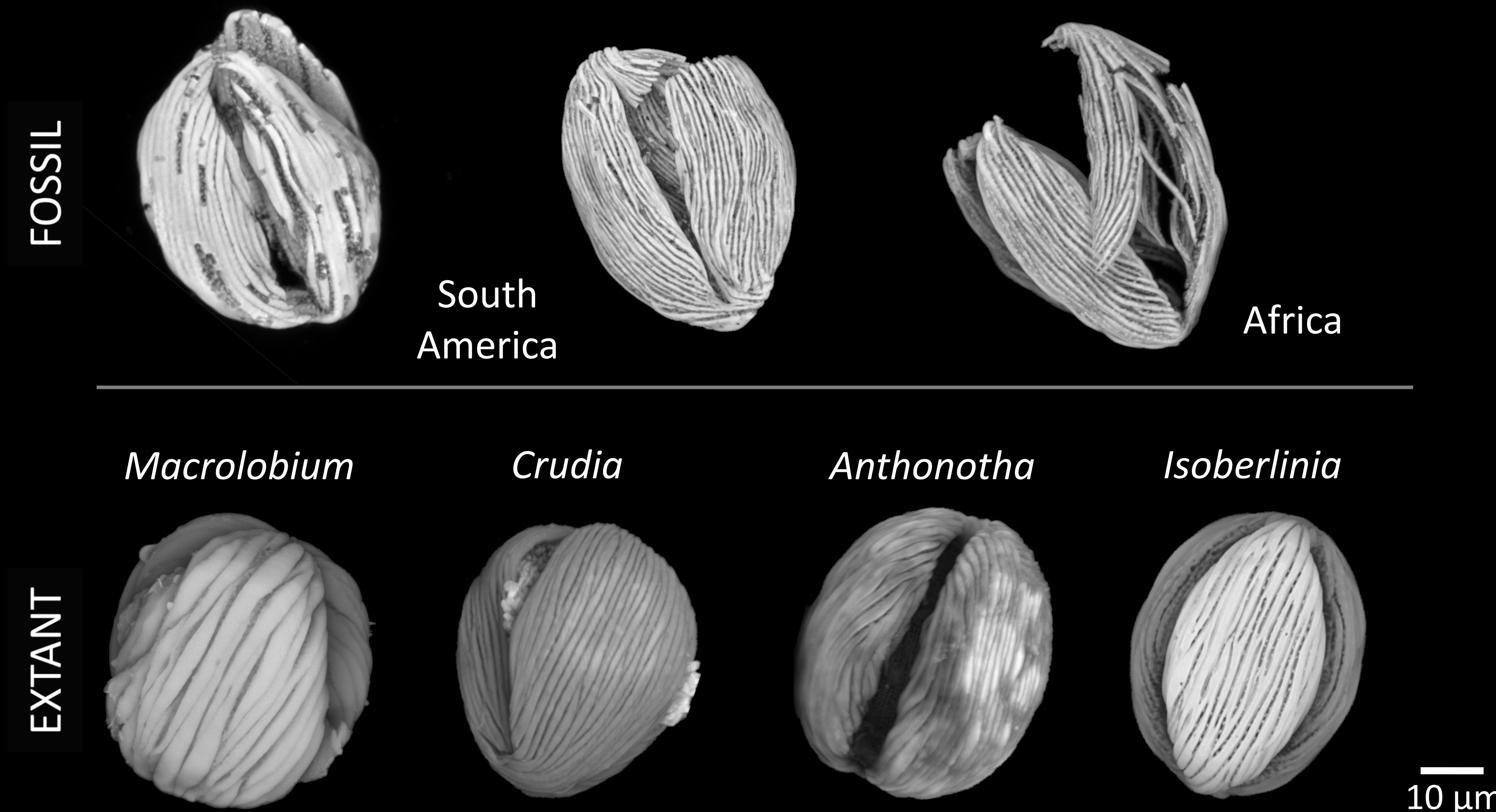
Importance of increasing the taxonomic resolution



Mander and Punyasena (2016)

Why *Striatopollis catatumbus*?

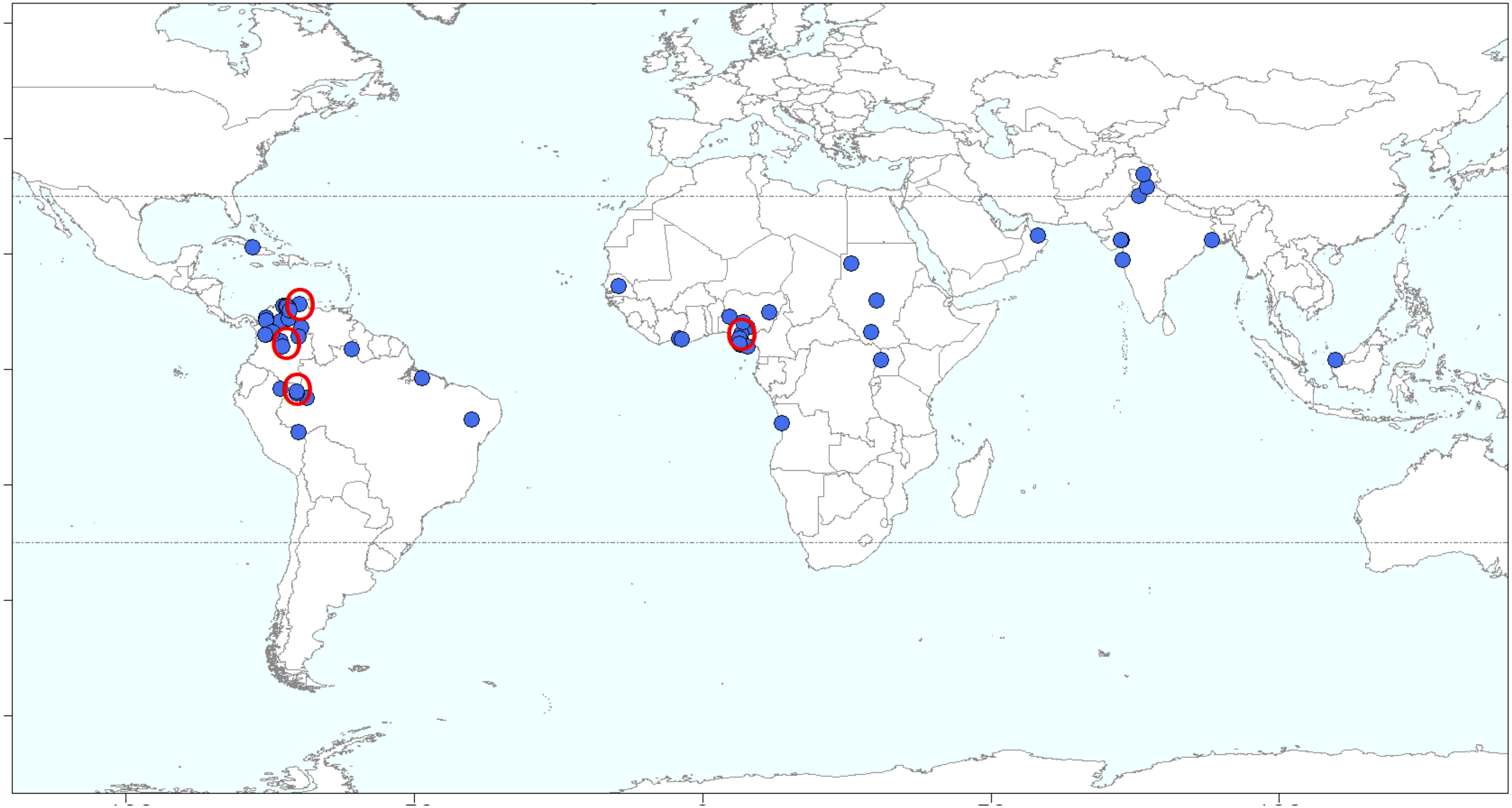
Affinity with the legume group Detarioideae/ Amherstia



Romero et al. (In preparation)

Why *Striatopollis catatumbus*?

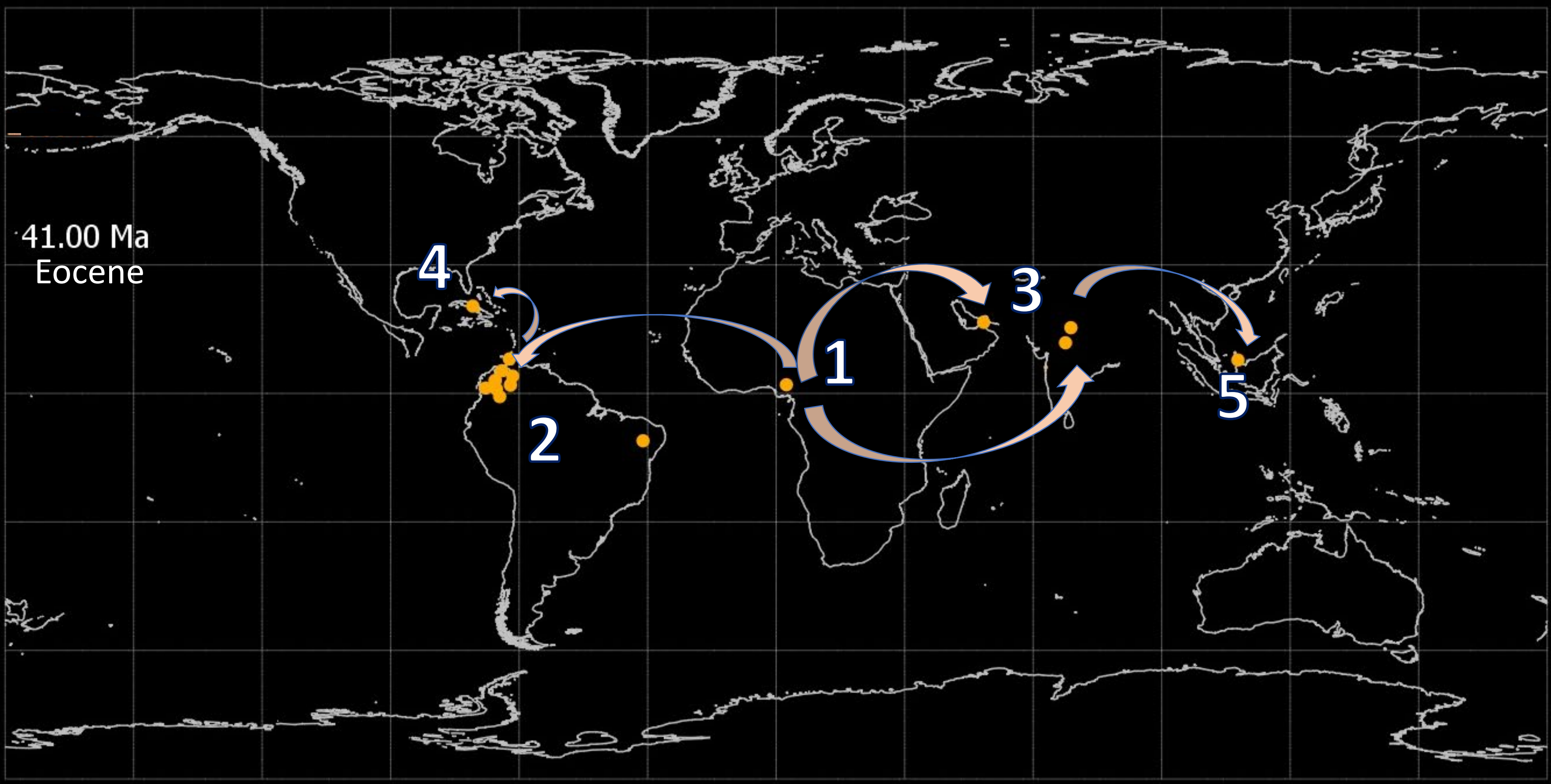
Pantropical distribution



Romero et al. (In preparation)

Why *Striatopollis catatumbus*?

Paleobiogeography



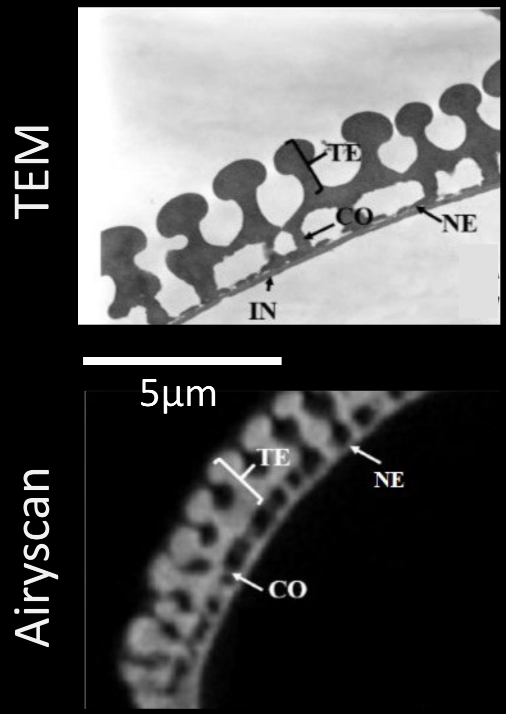
Africa: 65 Ma   South America: 58 Ma   Oman and India: 54 Ma

Central America: 49 Ma   Borneo: 40 Ma

Romero et al. (In preparation)

Airyscan confocal superresolution microscopy

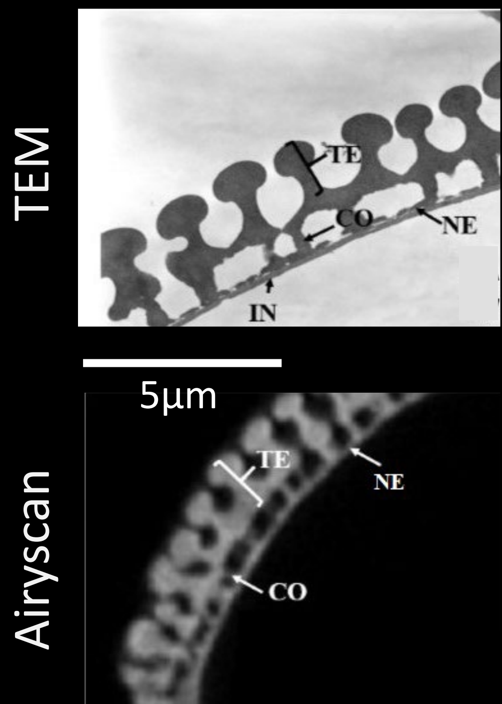
Comparison with EM



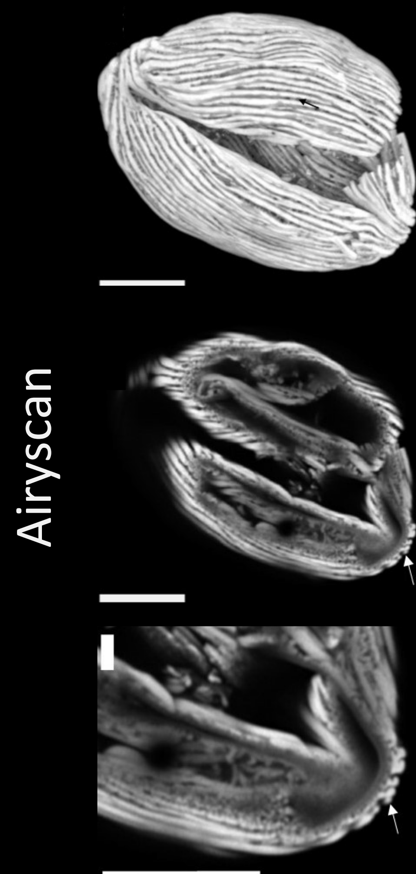
Romero et al. (In review)

Airyscan confocal superresolution microscopy

Comparison with EM



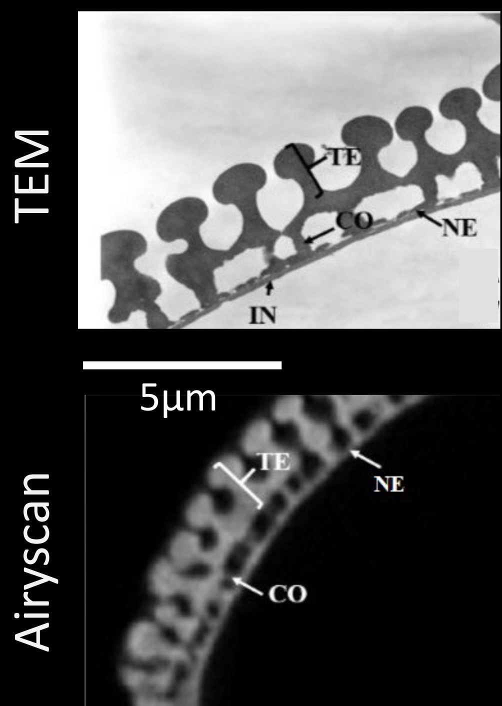
External and internal morphology



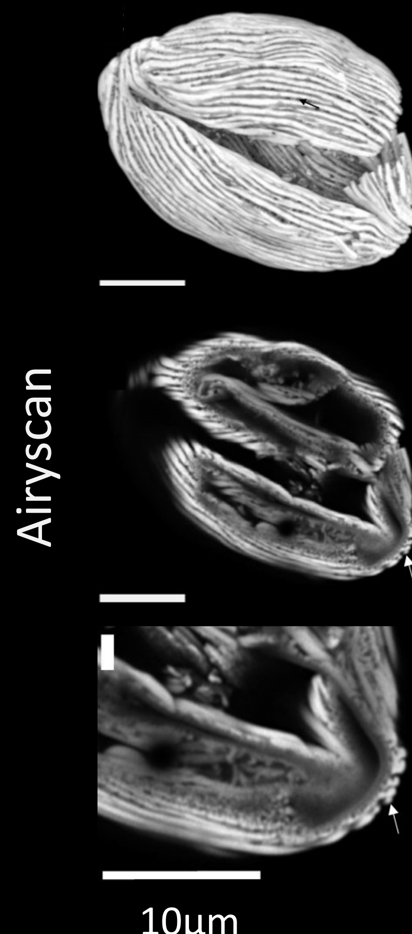
Romero et al. (In review)

Airyscan confocal superresolution microscopy

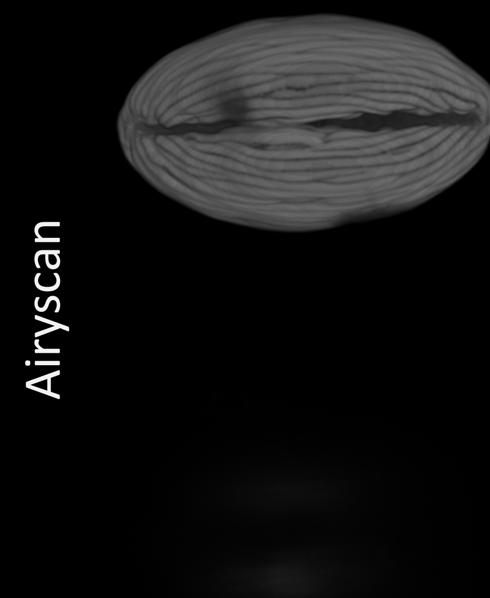
Comparison with EM



External and internal morphology

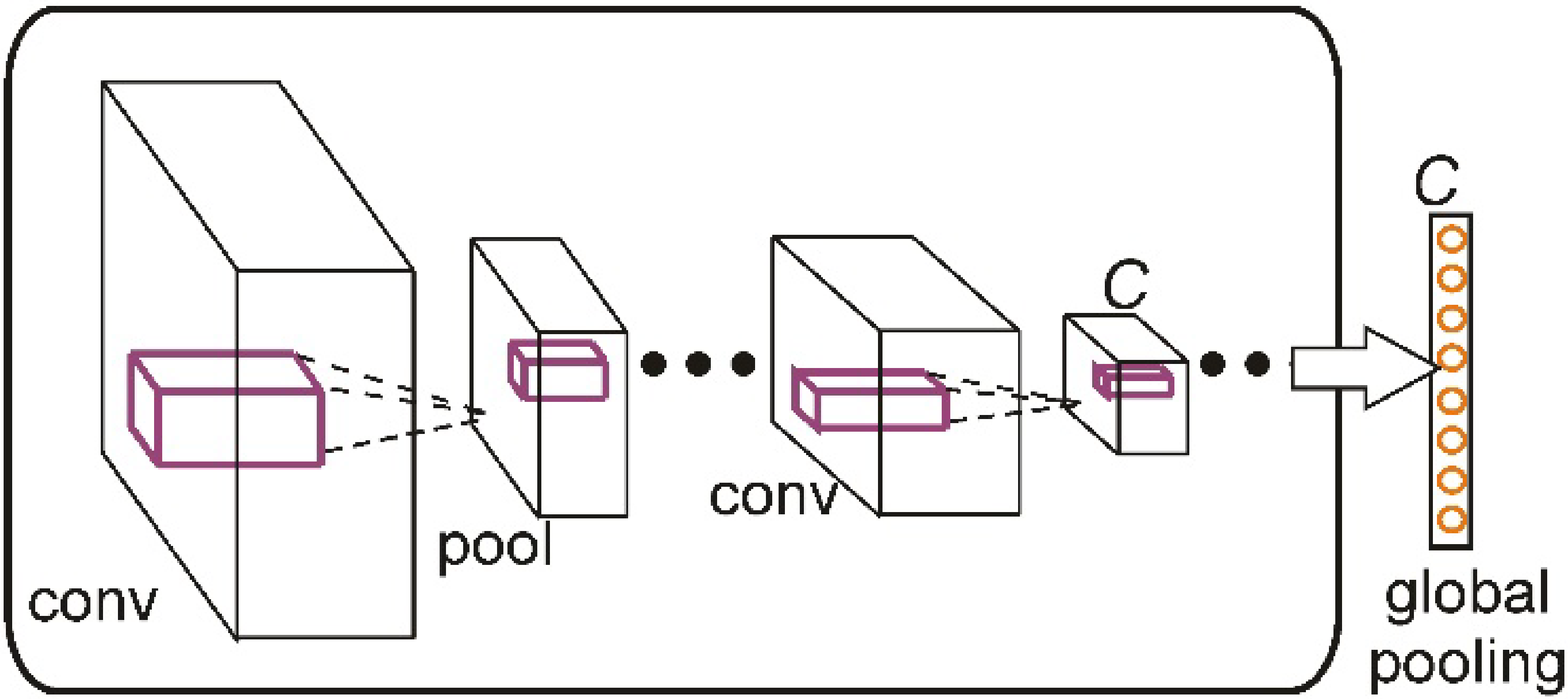


3D morphological data



Romero et al. (In review)

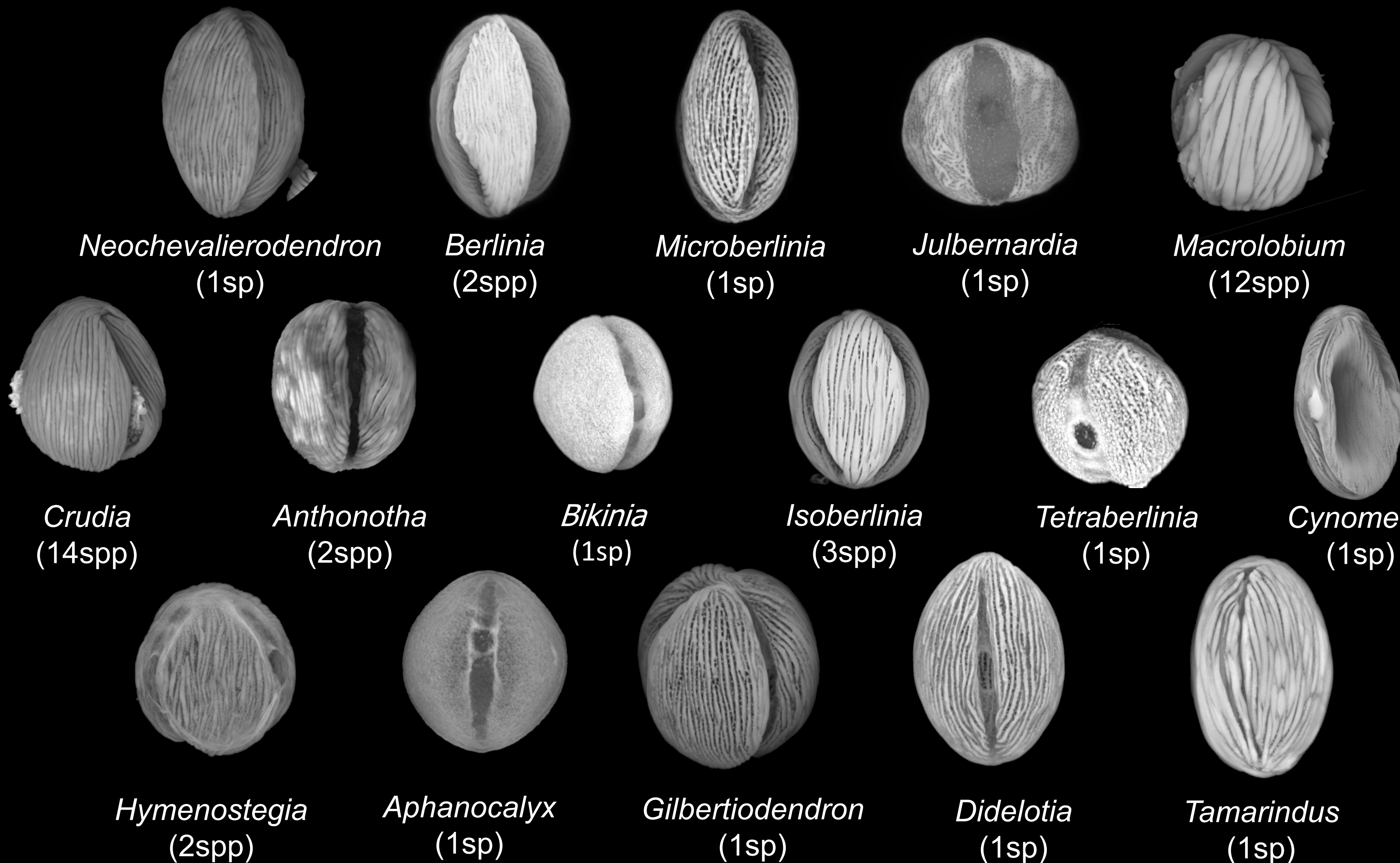
Convolutional neural networks - CNN



Romero et al. (In preparation)

Photos of modern pollen

A total of 459 pollen grains were imaged using Airyscan microscopy, 10 grains per species when possible.



Images are not scaled

Romero et al. (In preparation)

Modern pollen

Genus name	Images taken
Anthonotha	23
Aphanocalyx	9
Berlinia	20
Bikinia	10
Crudia	146
Cynometra	8
Didelotia	10
Gilbertodendron	10
Hymenostegia	21
Isoberlinia	31
Julbernardia	10
Macrolobium	121
Microberlinia	10
Neochevalierodendron	10
Tamarindus	10
Tetraberlinia	10

Romero et al. (In preparation)

Modern pollen

Genus name	Images taken
Anthonotha	23
Aphanocalyx	9
Berlinia	20
Bikinia	10
Crudia	146
Cynometra	8
Didelotia	10
Gilbertodendron	10
Hymenostegia	21
Isoberlinia	31
Julbernardia	10
Macrolobium	121
Microberlinia	10
Neochevalierodendron	10
Tamarindus	10
Tetraberlinia	10

→

70% training
30% testing

Romero et al. (In preparation)

Modern pollen

Genus name	Images taken
Anthonotha	23
Aphanocalyx	9
Berlinia	20
Bikinia	10
Crudia	146
Cynometra	8
Didelotia	10
Gilbertodendron	10
Hymenostegia	21
Isoberlinia	31
Julbernardia	10
Macrolobium	121
Microberlinia	10
Neochevalierodendron	10
Tamarindus	10
Tetraberlinia	10

70% training
30% testing

augment with
random rotations

Romero et al. (In preparation)

Modern pollen

Genus name	Images taken
Anthonotha	23
Aphanocalyx	9
Berlinia	20
Bikinia	10
Crudia	146
Cynometra	8
Didelotia	10
Gilbertodendron	10
Hymenostegia	21
Isoberlinia	31
Julbernardia	10
Macrolobium	121
Microberlinia	10
Neochevalierodendron	10
Tamarindus	10
Tetraberlinia	10

→ 70% training
30% testing

↓ augment with
random rotations

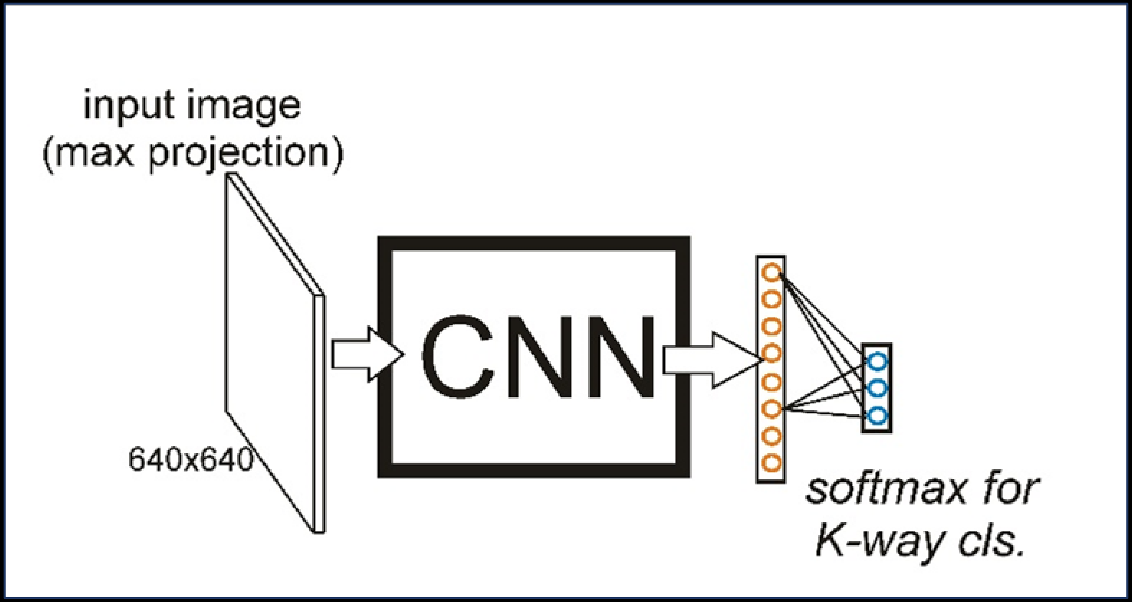
↓ ResNet
(50 layers,
cross-entropy
classifier)

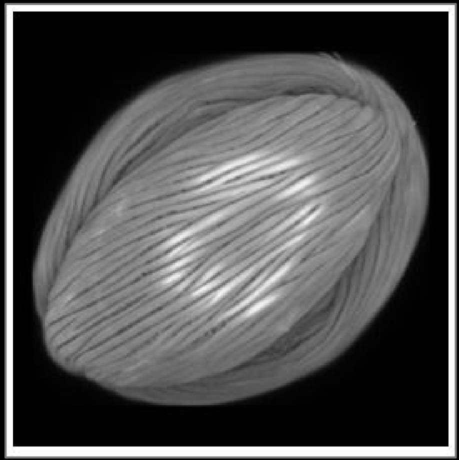
Romero et al. (In preparation)

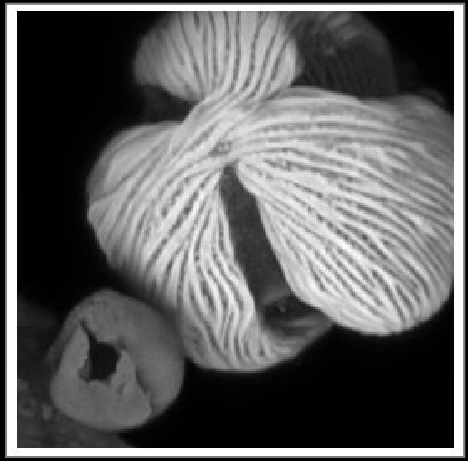
Modern pollen

Maximum projection model

Used the maximum intensity projection of the specimen image stacks as input for training and classification.





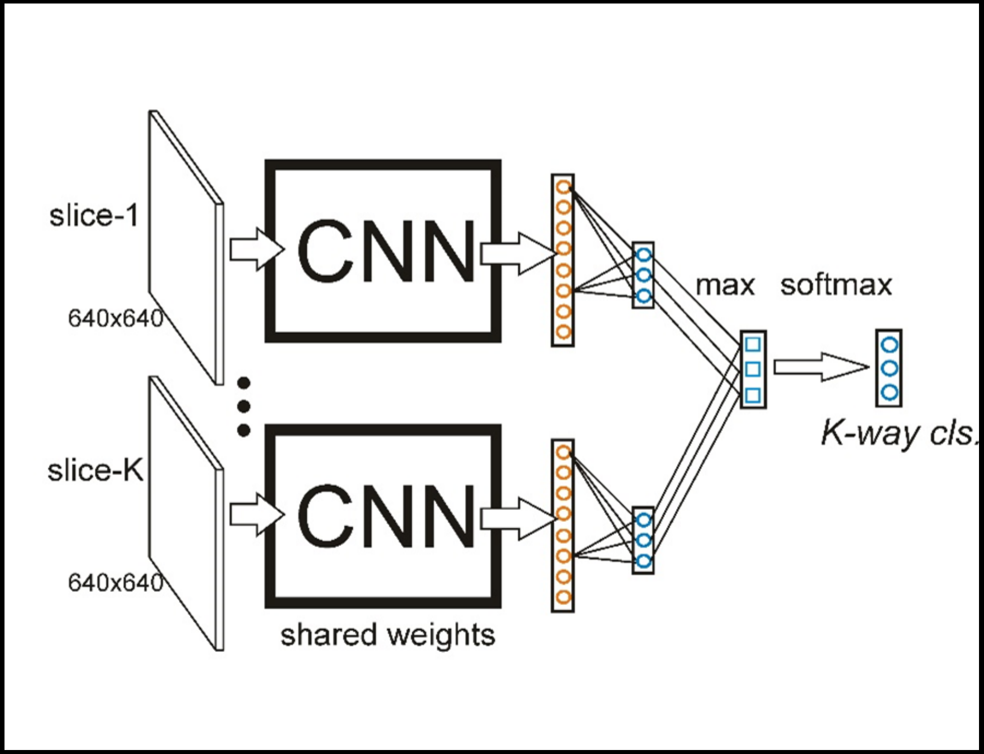


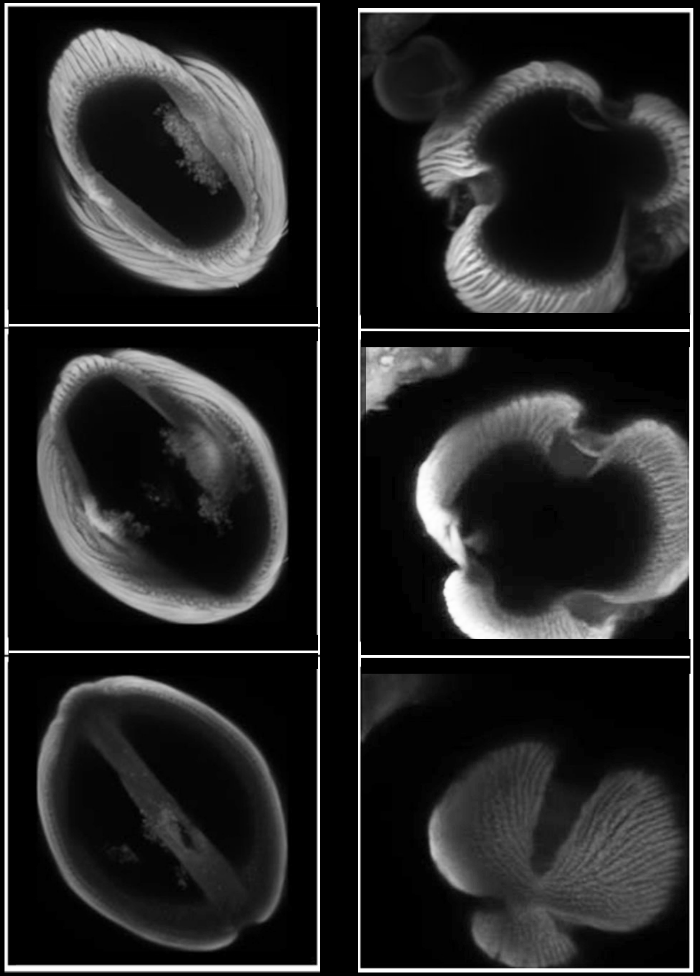
Romero et al. (In preparation)

Modern pollen

Multi-slice model

Used subsets of the image stack as input.



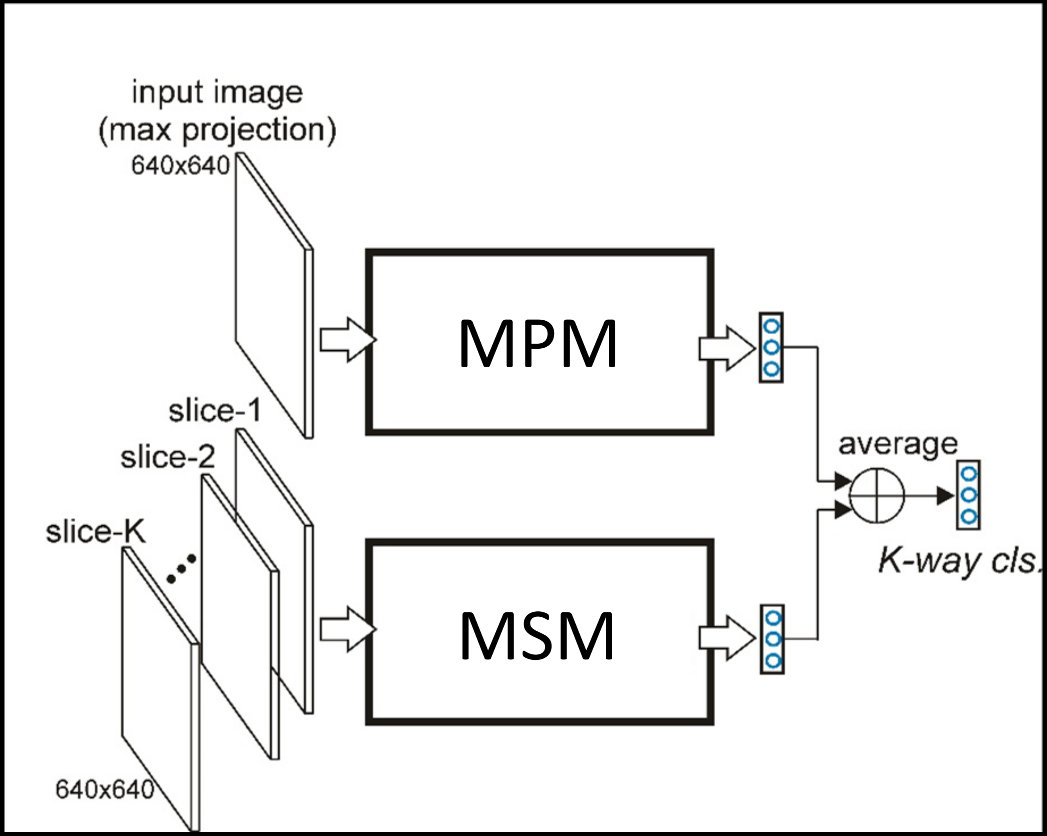


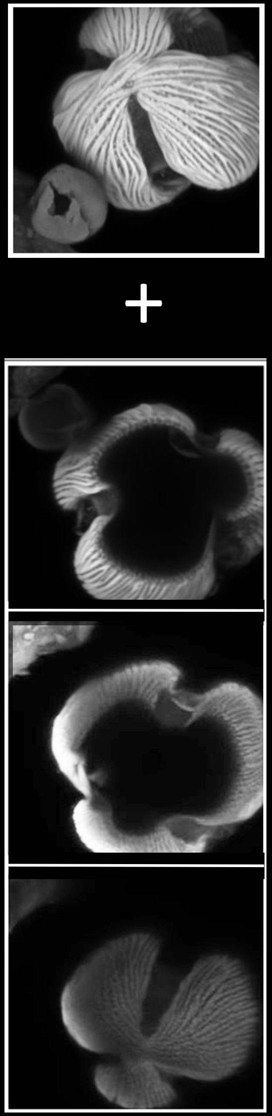
Romero et al. (In preparation)

Modern pollen

Fused model

Used the maximum projection model (MPM) and the multi-slice model (MSM) as input.

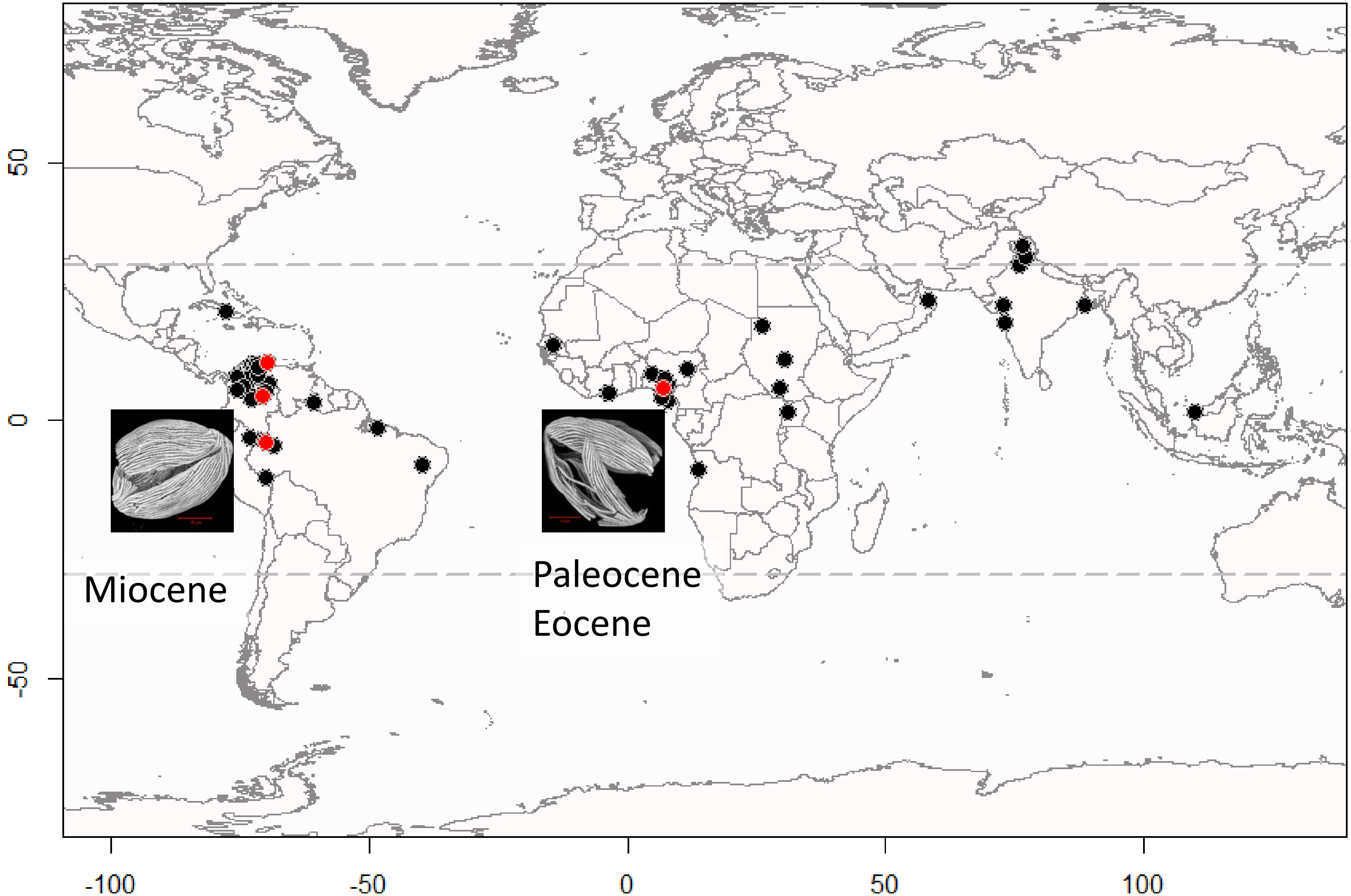




Romero et al. (In preparation)

Photos of fossil pollen

A total of 45 fossil pollen grains of *S. catatumbus* were imaged using Airyscan.

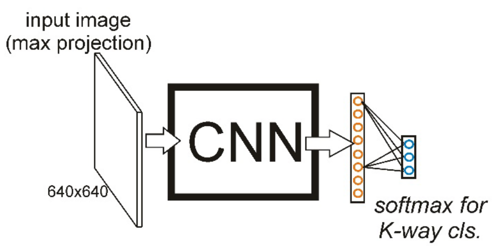


Romero et al. (In preparation)

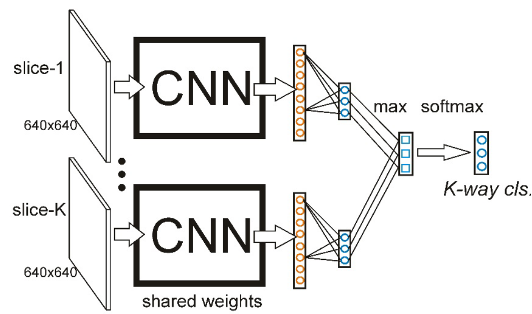
Fossil pollen

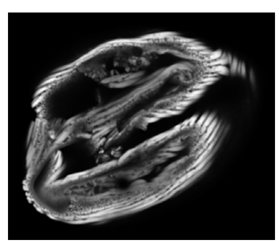
We applied the three models

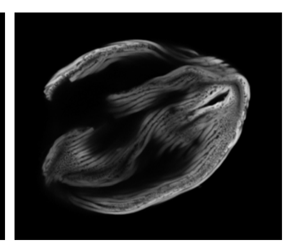
Maximum Projection model (MPM)



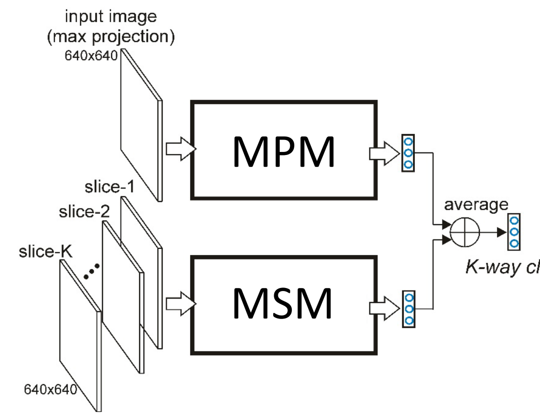
Multi-slice model (MSM)

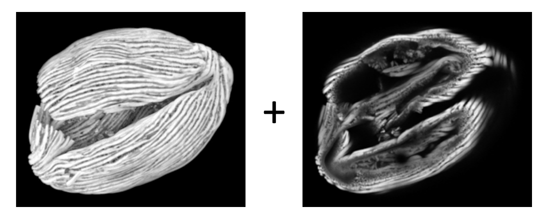






Fused model





Romero et al. (In preparation)

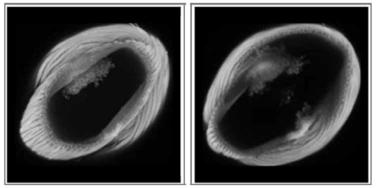
Results trained model – modern specimens

Maximum projection model – acc: 83.58% (3 of 16)

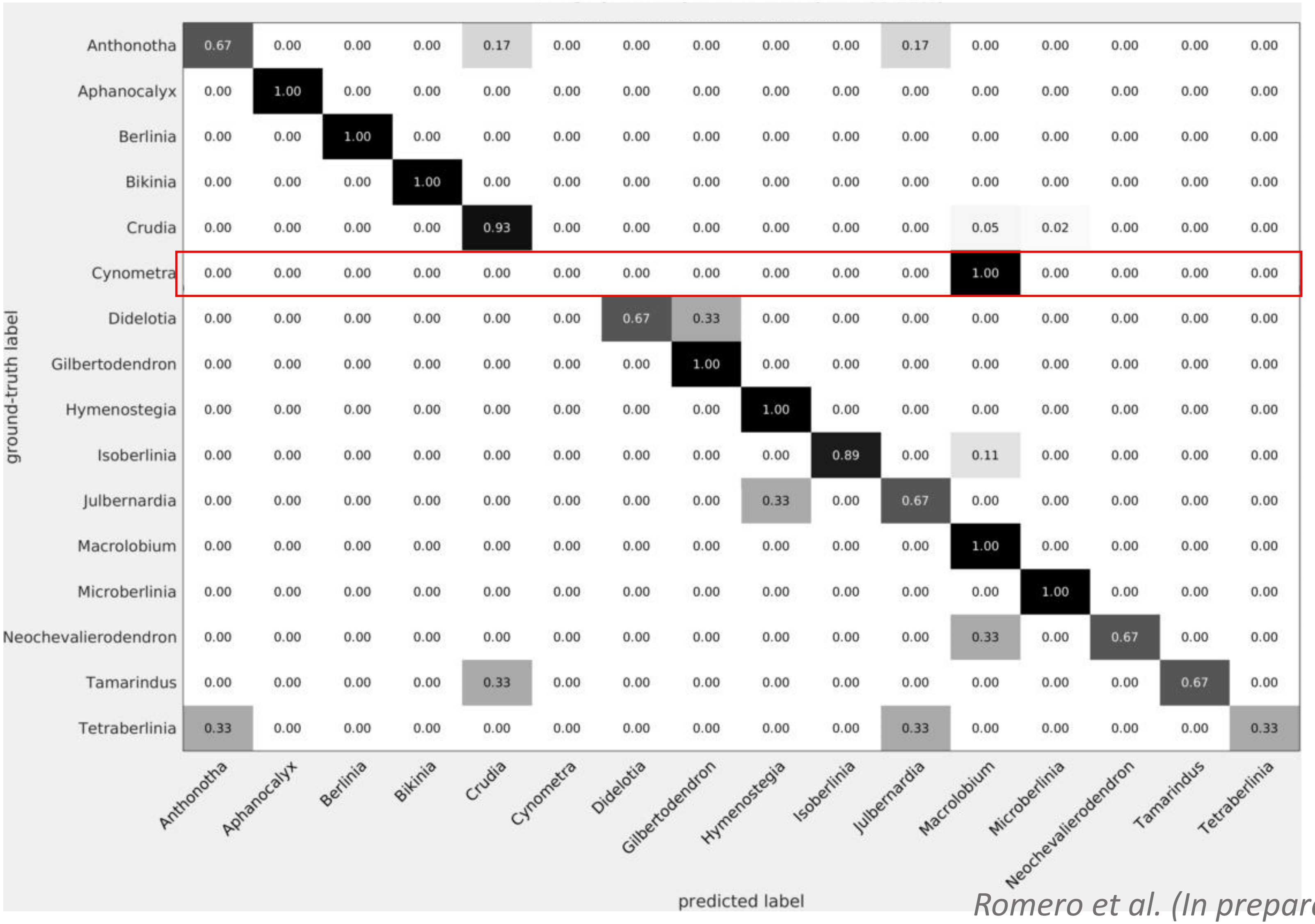
ground-truth label \ predicted label	Anthonotha	Aphanocalyx	Berlinia	Bikinia	Crudia	Cynometra	Didelotia	Gilbertodendron	Hymenostegia	Isoberlinia	Julbernardia	Macrolobium	Microberlinia	Neochevalierodendron	Tamarindus	Tetraberlinia
Anthonotha	0.67	0.00	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aphanocalyx	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Berlinia	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.00	0.00
Bikinia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crudia	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Cynometra	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Didelotia	0.00	0.00	0.00	0.00	0.33	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gilbertodendron	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hymenostegia	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Isoberlinia	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00	0.00	0.00	0.00
Julbernardia	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00
Macrolobium	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.03	0.00	0.00	0.92	0.00	0.00	0.00	0.00
Microberlinia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Neochevalierodendron	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.00	0.00
Tamarindus	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00
Tetraberlinia	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33

Romero et al. (In preparation)

Results trained model – modern specimens

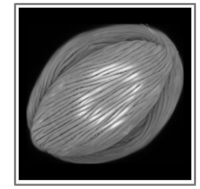


Multi-slice model – acc: 89.55% (9 of 16)

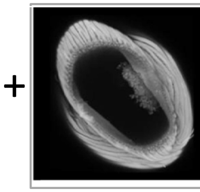


Romero et al. (In preparation)

Results trained model – modern specimens



+

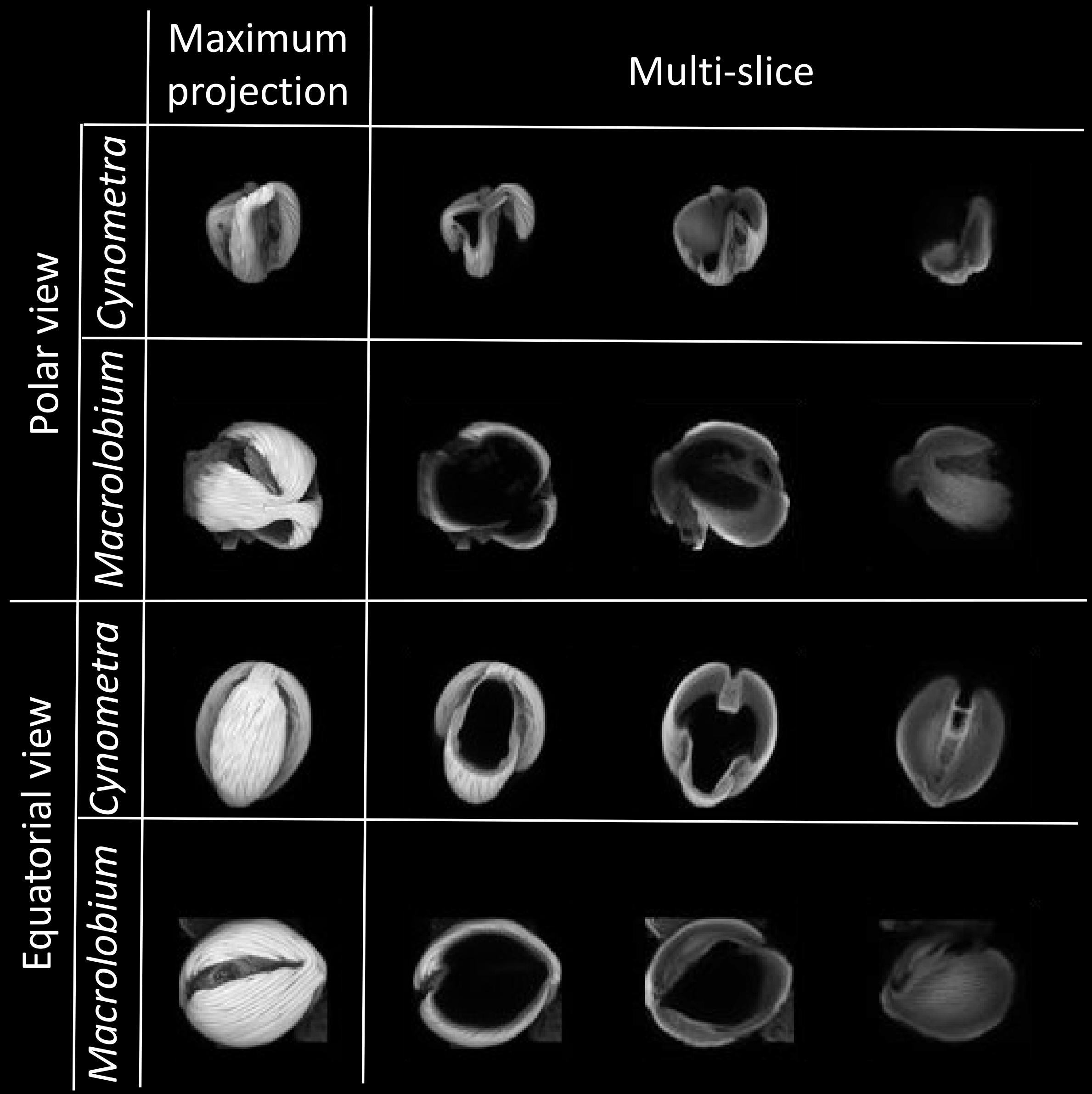


Fused model – acc: 90.30% (10 of 16)

ground-truth label \ predicted label	Anthonotha	Aphanocalyx	Berlinia	Bikinia	Crudia	Cynometra	Didelotia	Gilbertodendron	Hymenostegia	Isoberlinia	Julbernardia	Macrolobium	Microberlinia	Neochevalierodendron	Tamarindus	Tetraberlinia
Anthonotha	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aphanocalyx	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Berlinia	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bikinia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crudia	0.00	0.00	0.00	0.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.00	0.00	0.00
Cynometra	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Didelotia	0.00	0.00	0.00	0.00	0.33	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gilbertodendron	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hymenostegia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Isoberlinia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.00	0.11	0.00	0.00	0.00	0.00
Julbernardia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.67	0.00	0.00	0.00	0.00	0.00
Macrolobium	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.03	0.00	0.00
Microberlinia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Neochevalierodendron	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.00
Tamarindus	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00
Tetraberlinia	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67

Romero et al. (In preparation)

Results trained model – modern specimens



Romero et al. (In preparation)

Fossil specimens identification with CNN

Pred. label (probability)

Locality of collection	Sample ID	Maximum projection model	Multi-slice model	Fused model
Africa	01 S.catatumbus_Africa	Hymenostegia (0.284)	Julbernardia (0.998)	Julbernardia (0.496)
Africa	02 S.catatumbus_Africa	Anthonotha (0.745)	Cynometra (0.786)	Neochevalierodendron (0.479)
Africa	03 S.catatumbus_Africa	Crudia (0.548)	Crudia (0.971)	Crudia (0.727)
Africa	04 S.catatumbus_Africa	Cynometra (0.519)	Cynometra (0.961)	Neochevalierodendron (0.495)
Africa	05 S.catatumbus_Africa	Berlinia (0.483)	Berlinia (0.899)	Berlinia (0.546)
Africa	06 S.catatumbus_Africa	Isoberlinia (0.749)	Berlinia (0.980)	Berlinia (0.500)
Africa	07 S.catatumbus_Africa	Anthonotha (0.990)	Didelotia (0.840)	Anthonotha (0.997)
Africa	08 S.catatumbus_Africa	Cynometra (0.670)	Berlinia (0.707)	Neochevalierodendron (0.493)
Africa	09 S.catatumbus_Africa	Crudia (0.496)	Crudia (0.847)	Crudia (0.844)
Africa	10 S.catatumbus_Africa	Crudia (0.880)	Isoberlinia (0.811)	Anthonotha (0.596)
Africa	11 S.catatumbus_Africa	Berlinia (0.912)	Julbernardia (0.942)	Berlinia (0.998)
Africa	12 S.catatumbus_Africa	Anthonotha (0.746)	Didelotia (0.967)	Crudia (0.524)
Africa	13 S.catatumbus_Africa	Didelotia (0.453)	Crudia (0.984)	Crudia (0.787)
Africa	14 S.catatumbus_Africa	Anthonotha (0.768)	Anthonotha (0.920)	Anthonotha (0.984)
Africa	15 S.catatumbus_Africa	Crudia (0.523)	Crudia (0.543)	Crudia (0.527)
Africa	16 S.catatumbus_Africa	Anthonotha (0.854)	Microberlinia (0.952)	Microberlinia ([illegible])
Africa	17 S.catatumbus_Africa	Isoberlinia (0.438)	Berlinia (0.9[illegible]8)	Berlinia (0.[illegible]16)
Africa	18 S.catatumbus_Africa	Bikinia (0.998)	Julbernardia (0.876)	Bikinia (0.499)
Africa	19 S.catatumbus_Africa	Didelotia (0.457)	Crudia (0.941)	Crudia (0.516)
Africa	20 S.catatumbus_Africa	Crudia (0.382)	Crudia (0.909)	Crudia (0.563)
Africa	21 S.catatumbus_Africa	Crudia (0.528)	Crudia (0.905)	Crudia (0.741)
Africa	22 S.catatumbus_Africa	Didelotia (0.470)	Crudia (0.965)	Crudia (0.510)
Africa	23 S.catatumbus_Africa	Crudia (0.955)	Neochevalierodendron (0.826)	Crudia (0.925)
Africa	24 S.catatumbus_Africa	Crudia (0.456)	Cynometra (0.909)	Crudia (0.507)
Africa	25 S.catatumbus_Africa	Crudia (0.980)	Crudia (1.000)	Crudia (1.000)
Africa	26 S.catatumbus_Africa	Berlinia (0.417)	Berlinia (0.676)	Neochevalierodendron (0.501)
Africa	27 S.catatumbus_Africa	Crudia (0.746)	Crudia (0.940)	Crudia (0.722)
Africa	28 S.catatumbus_Africa	Crudia (0.936)	Crudia (0.997)	Crudia (0.994)
Africa	29 S.catatumbus_Africa	Crudia (0.969)	Crudia (0.933)	Crudia (0.998)
S. A	30 S.catatumbus_SA	Didelotia (0.906)	Anthonotha (0.738)	Anthonotha (0.672)
S. A	31 S_catatumbus_SA	Didelotia (0.598)	Crudia (0.784)	Crudia (0.491)
S. A	32 S_catatumbus_SA	Crudia (0.940)	Crudia (0.981)	Crudia (0.993)
S. A	33 S_catatumbus_SA	Crudia (0.926)	Crudia (0.797)	Crudia (0.986)
S. A	34 S_catatumbus__SA	Cynometra (0.945)	Berlinia (0.589)	Berlinia (0.500)
S. A	35 S_catatumbus__SA	Macrolobium (0.986)	Macrolobium (0.997)	Macrolobium (0.575)
S. A	36 S_catatumbus__SA	Cynometra (0.762)	Macrolobium (0.831)	Berlinia (0.609)
S. A	37 S_catatumbus__SA	Isoberlinia (0.741)	Berlinia (1.000)	Berlinia (0.848)
S. A	38 S_eudoscabratus_SA	Crudia (0.474)	Crudia (0.968)	Crudia (0.924)
S. A	39 S_Macrolobium_SA	Macrolobium (0.992)	Macrolobium (0.991)	Macrolobium (0.980)
S. A	40 S_Macrolobium_SA	Macrolobium (1.000)	Macrolobium (1.000)	Macrolobium (1.000)
S. A	41 S_Macrolobium_SA	Macrolobium (1.000)	Macrolobium (0.994)	Macrolobium (1.000)
S. A	42 S_Macrolobium_SA	Macrolobium (0.765)	Macrolobium (0.979)	Macrolobium (0.628)
S. A	43 S_Macrolobium_SA	Isoberlinia (0.308)	Cynometra (0.900)	Neochevalierodendron (0.494)
S. A	44 S_melenae_SA	Crudia (0.535)	Macrolobium (0.884)	Macrolobium (0.498)
S. A	45 S_tenuistriatus_SA	Crudia (0.838)	Crudia (0.769)	Crudia (0.841)

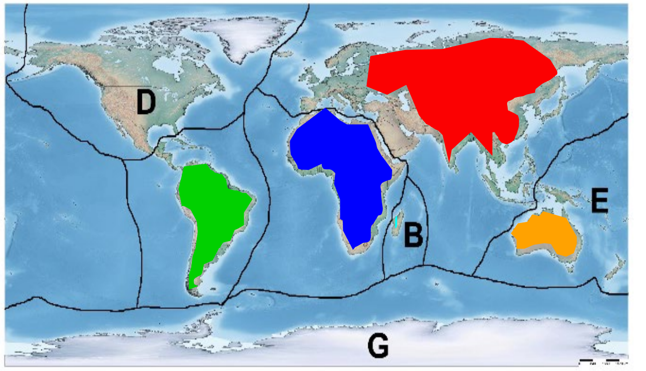
Fossil specimens identification with CNN

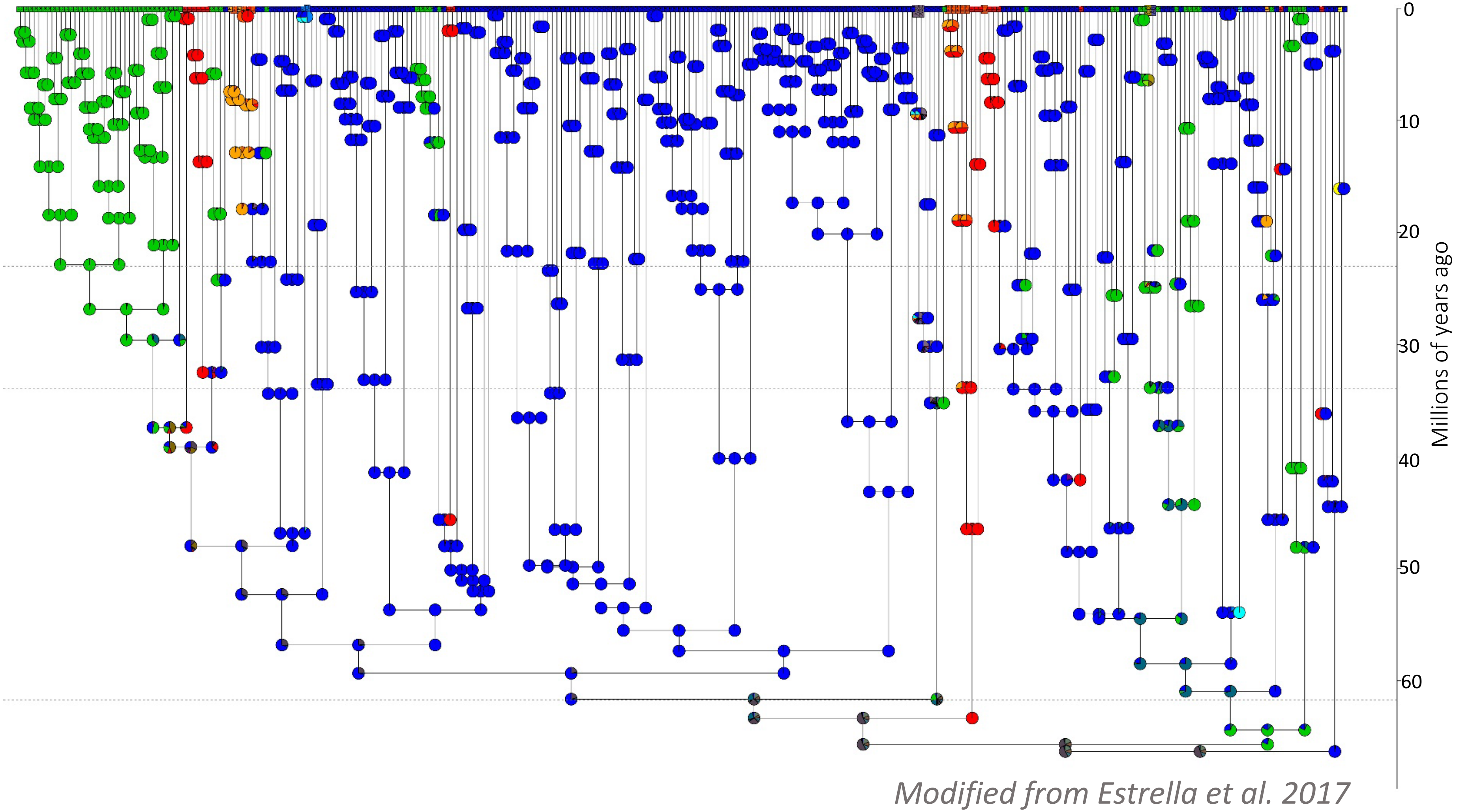
- 20 specimens with consistent identifications in all models
  11 African, 9 South American

- 19 specimens with consistent identifications in two models
  14 African, 5 South American

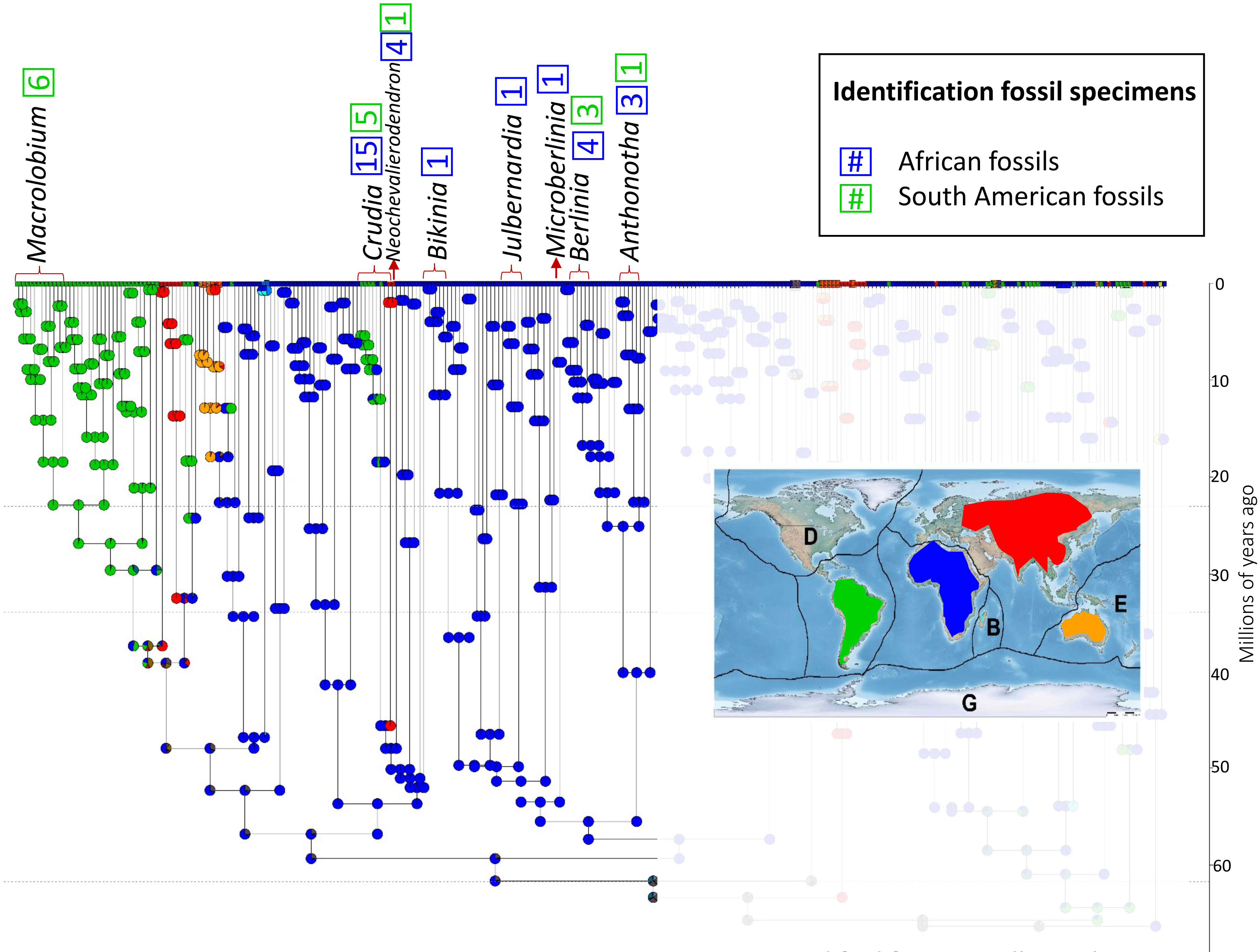
- 6 specimens with no consistent identifications
  4 African, 2 South American

Romero et al. (In preparation)





Modified from Estrella et al. 2017



Modified from Estrella et al. 2017

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Conclusions

- Airyscan microscopy captures high-resolution 3D images comparable to EM.

- CNNs efficiently discriminate subtle morphological differences in pollen.

- Results show high accuracy in classifying most of the genera, but *Cynometra* was never identified as *Cynometra.*

- Fossil classifications show biogeographic and phylogenetic consistency with extant genera. Affinity with more than one genus suggests intermediate morphologies.

- Potential to improve the accuracy and efficiency of fossil pollen identification.

- Opens the study of the evolution and ecology of plant clades to fundamentally novel interpretations of pollen data.

Acknowledgments

Michael A. Urban (PhD) from University of New Brunswick, Canada

Francisca Oboh-Ikuenobe (PhD) from Missouri University of Science and technology

Carlos D'Apolito (PhD) from Universidade Federal de Mato Grosso

David Seigler (PhD) from Herbarium Plant Biology, UIUC.

Mayandi Sivaguru (PhD) from Carl R. Woese Institute for Genomic Biology (IGB), UIUC.

James Solomon (PhD) from Missouri Botanical Garden.

Meghann Toner from the department of Botany at US National Herbarium.

Project funded by:

- Department of Plant Biology, UIUC
- National Science Foundation
  - NSF-DBI – Advances in Bioinformatics, NSF-DBI-1262561 to SWP and NSF-DBI-1262547 to CCF
  - NSF-IIS – Information and Intelligent systems, NSF-IIS-1618806 and NSF-IIS-1253538 to CCF