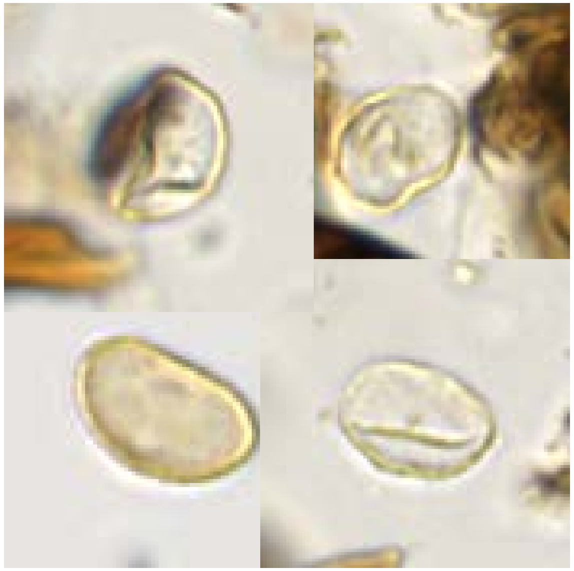
# Automating tropical pollen counts using convolutional neural nets: from image acquisition to identification

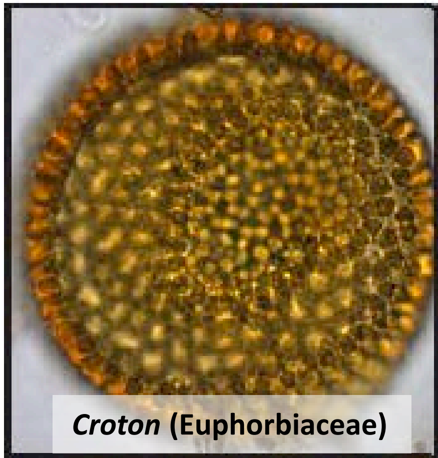
Derek S. Haselhorst[1], Shu Kong[2], Charless Fowlkes[2], J. Enrique Moreno[3], David K. Tcheng[4] and Surangi W. Punyasena[1]

[1]*University of Illinois at Urbana-Champaign*
[2]*University of California at Irvine*
[3]*Smithsonian Tropical Research Institute, Panama City, Republic of Panama*
[4]*National Center for Supercomputing Applications,*
*University of Illinois at Urbana-Champaign*

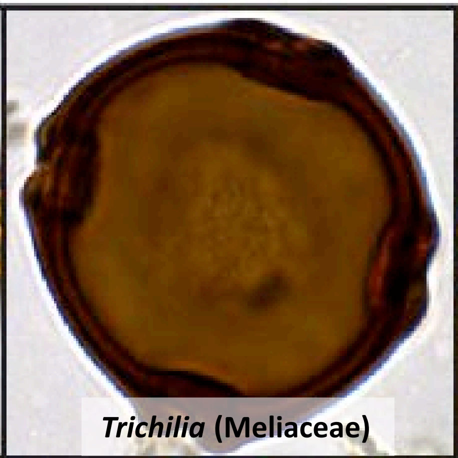
*Cecropia* (Urticaceae)



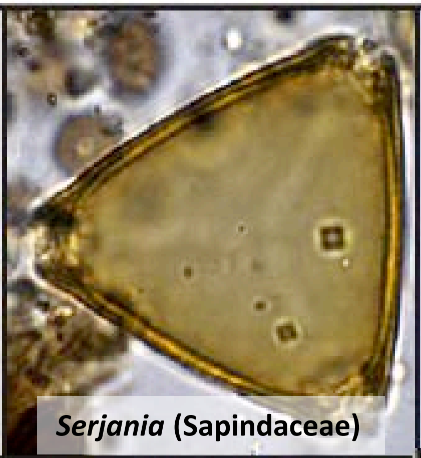
NSF Macrosystems Ecology
NSF Advances in Biological Informatics



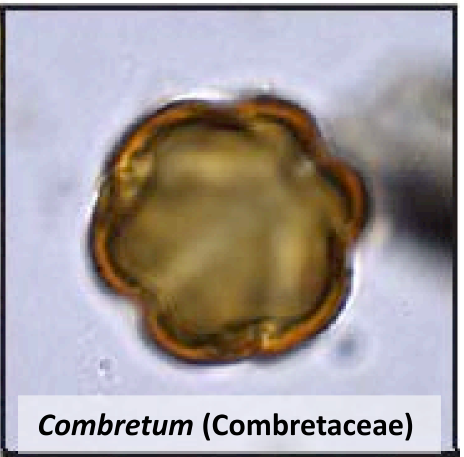
*Croton* (Euphorbiaceae)



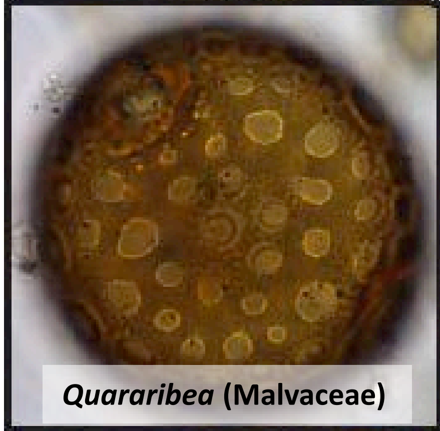
*Trichilia* (Meliaceae)



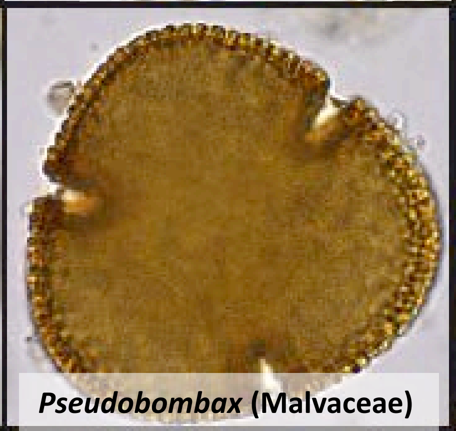
*Serjania* (Sapindaceae)



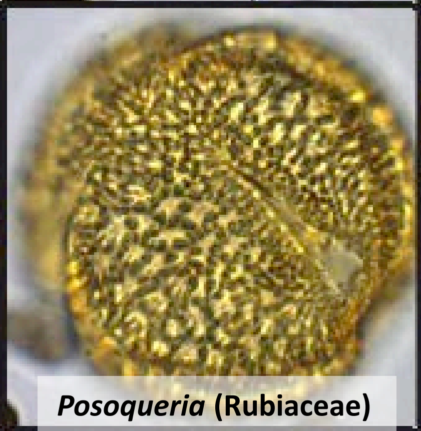
*Combretum* (Combretaceae)



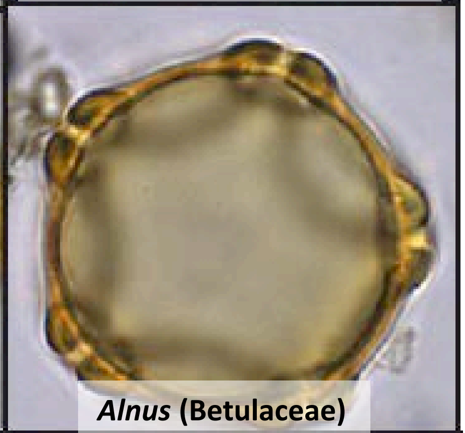
*Quararibea* (Malvaceae)



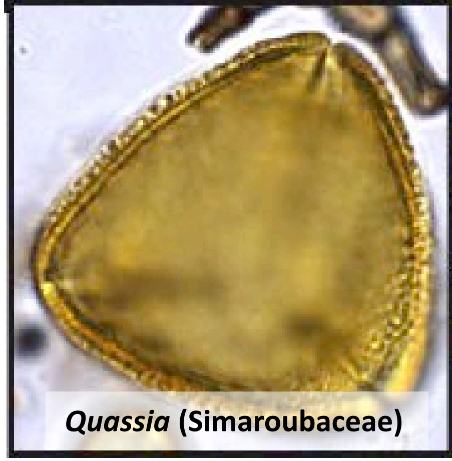
*Pseudobombax* (Malvaceae)



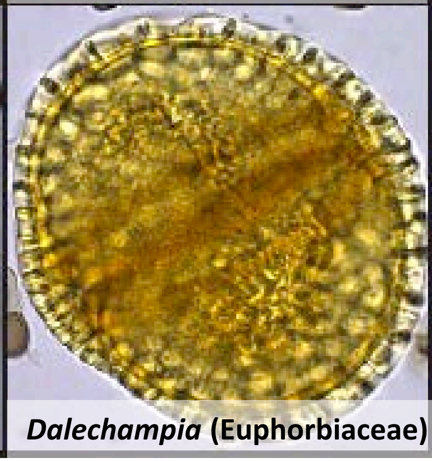
*Posoqueria* (Rubiaceae)



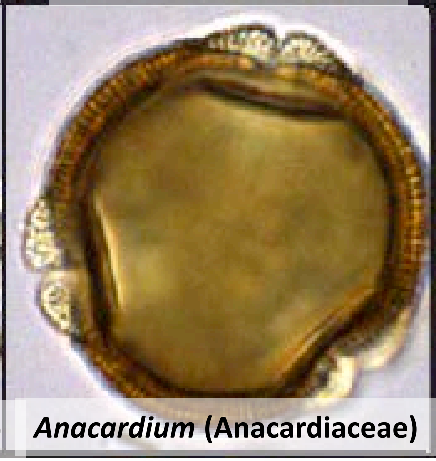
*Alnus* (Betulaceae)



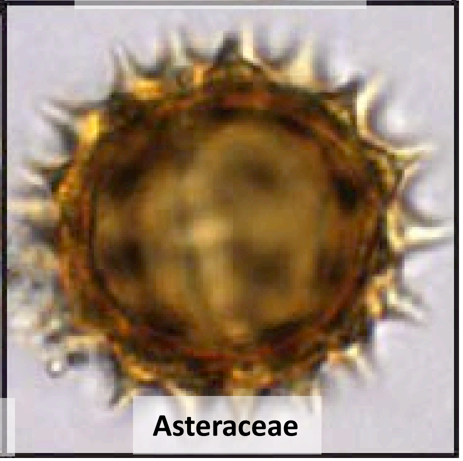
*Quassia* (Simaroubaceae)



*Dalechampia* (Euphorbiaceae)



*Anacardium* (Anacardiaceae)



Asteraceae

Haselhorst , Moreno and Punyasena, 2013, *PLoS ONE*

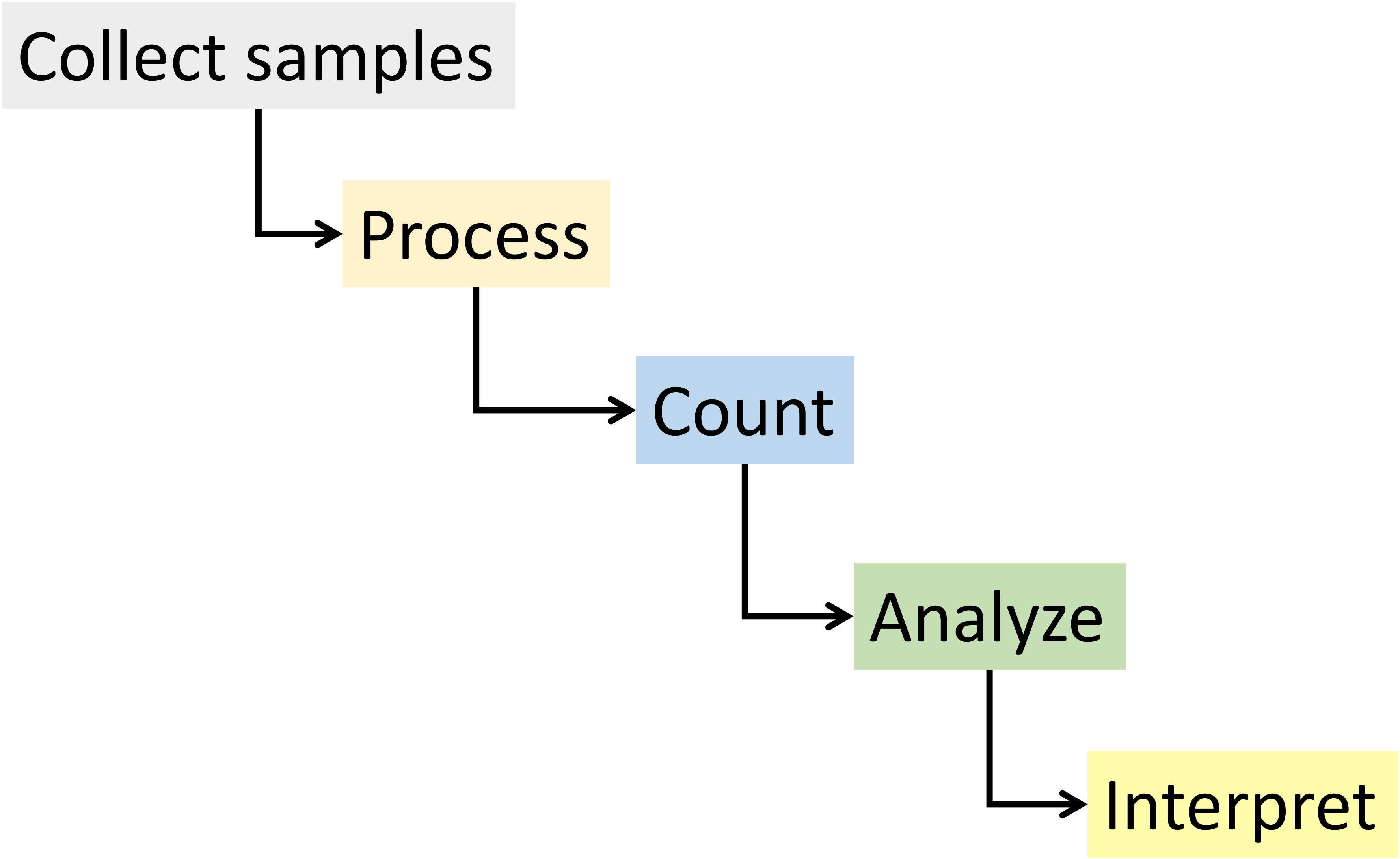
# POLLEN AS "BIG DATA"

~470 million years of plant history

Billions of potential specimens

~Continuous deposition across a range of environments

# REIMAGINING THE WORKFLOW



Collect samples

Process

Count

Analyze

Interpret

# IS AUTOMATION THE ANSWER?

**Quantity**: increase the throughput of pollen analysis

**Reproducibility**: improve the consistency and accuracy of pollen identifications

**Resolution**: produce repeatable recognition of ***species*** from pollen for more precise biome reconstructions

# TRAINING ON HYPERDIVERSE SAMPLES

- A 15-year pollen rain record from Barro Colorado Island, Panama
  - Obtained from a series of 20, evenly spaced pollen traps along two parallel transect in the 50 ha CTFS plot

- A 10-year pollen record from the Lutz weather tower
  - Images to be analyzed this summer

- ~ 130 pollen morphotypes

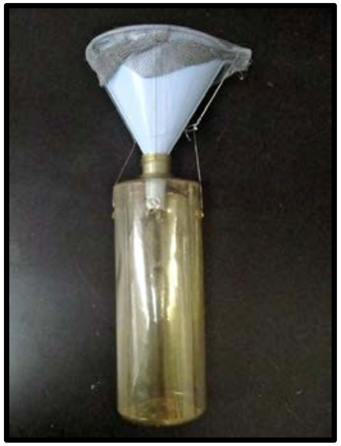
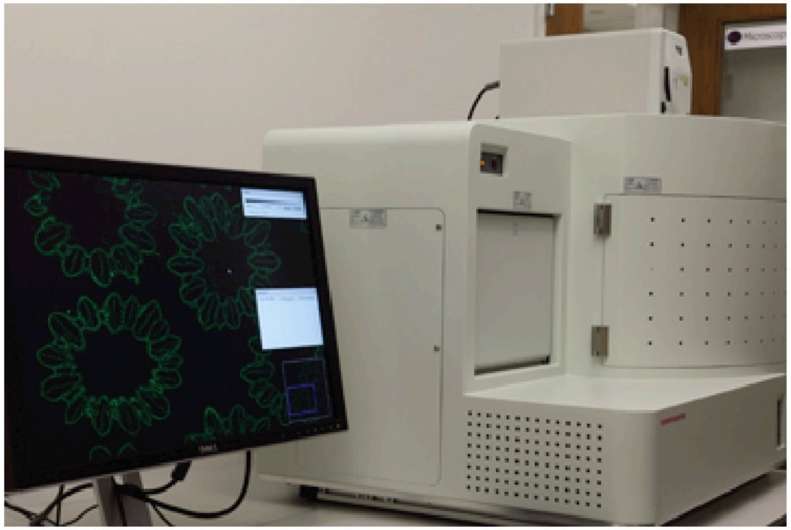
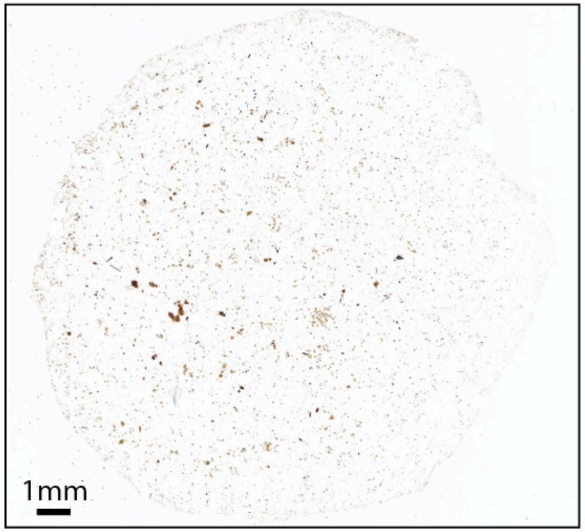
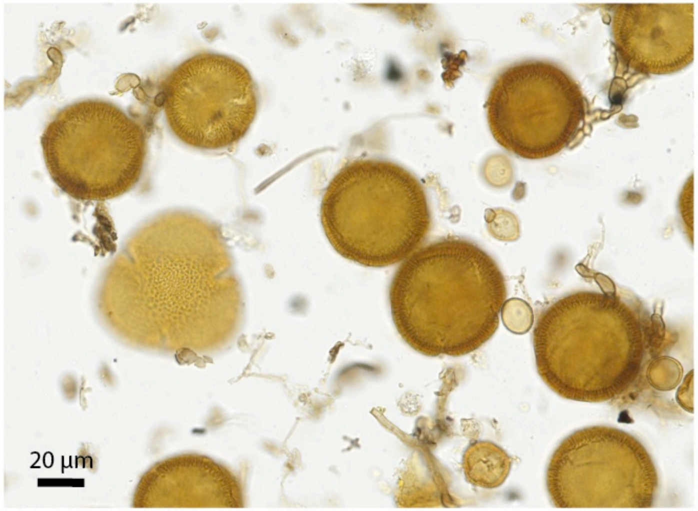


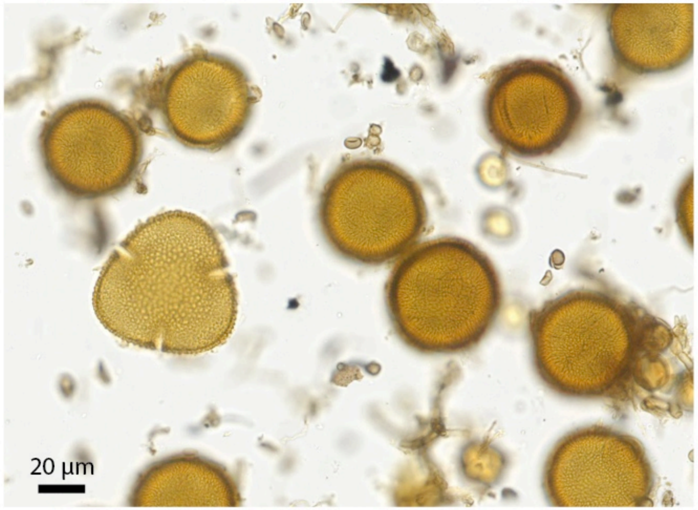
Photo: STRI

# IMAGING POLLEN SLIDES FOR COUNTING





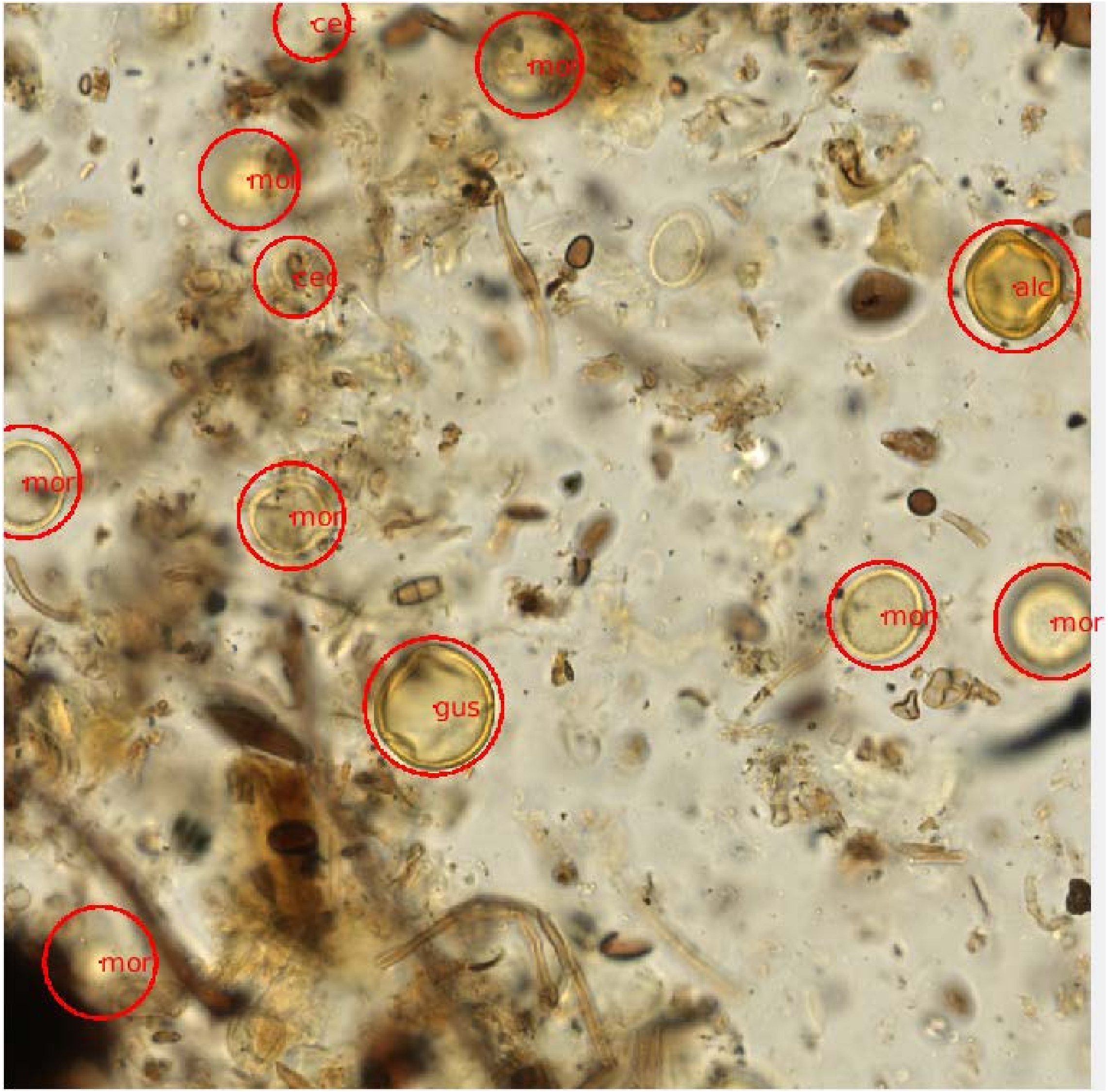




400x, 0.23 µm/pixel

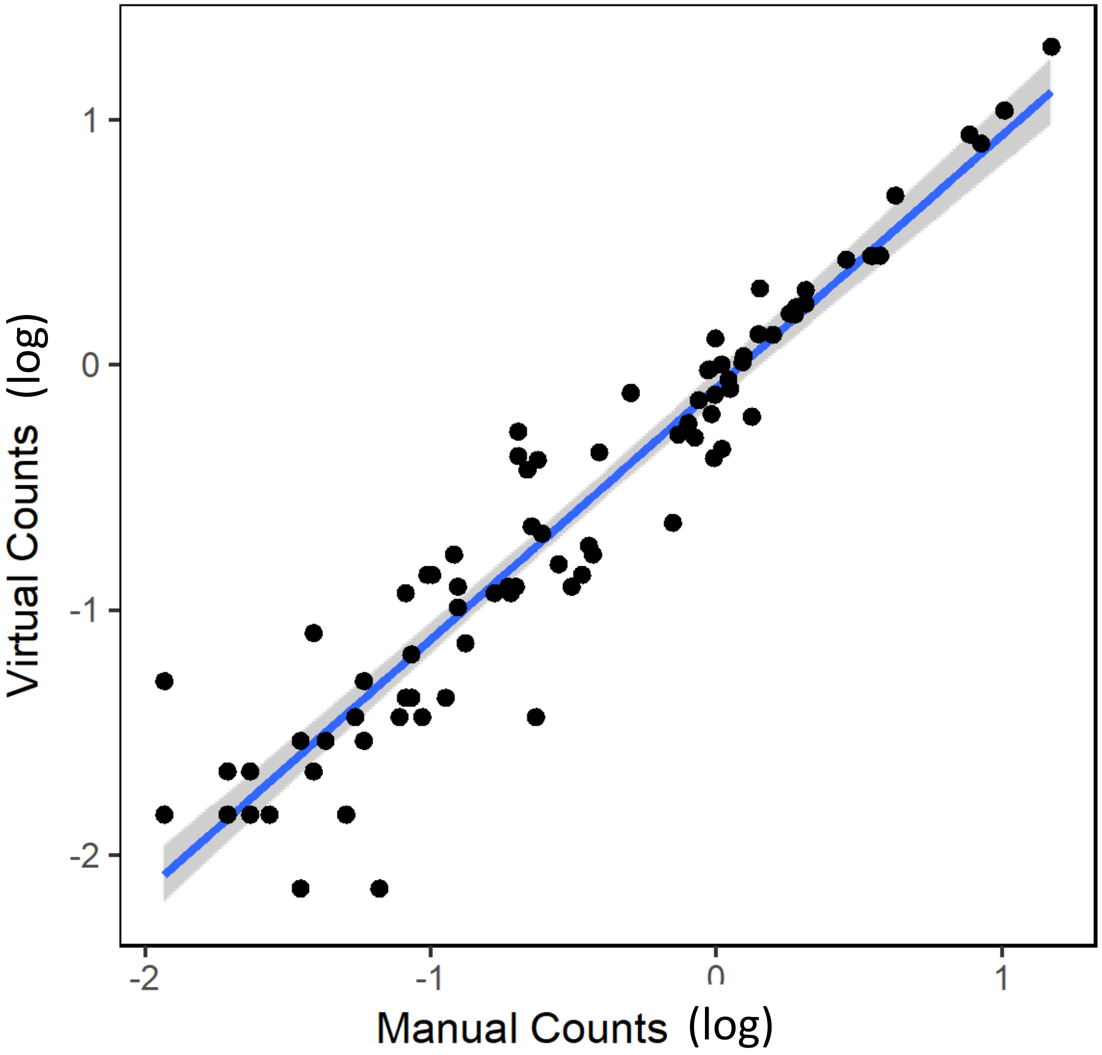
One sample (41 @ 1 µm axial planes) = ~400 GB

# VIRTUAL MICROSCOPE POLLEN IDENTIFICATIONS



- PNG images subsampled using a java script
- Slide images read from Matlab script
- Image metadata recorded for each individual pollen grain
  - Slide ID
  - Pollen coordinate
  - Pollen radius
  - 3-letter ID
  - Confidence level (0-9)
- Images and metadata then shared with UC-Irvine Computer Vision Collaborators

# COMPARING VIRTUAL AND LIGHT MICROSCOPY COUNTS



- Testing the fidelity of the virtual microscope using the 10-year Lutz tower record

- Can pollen can be identified at the same taxonomic resolution and frequency?

- R^2 = 0.97

- Observed differences are likely reflective of differences in counting strategy:

  Manual: slide transects

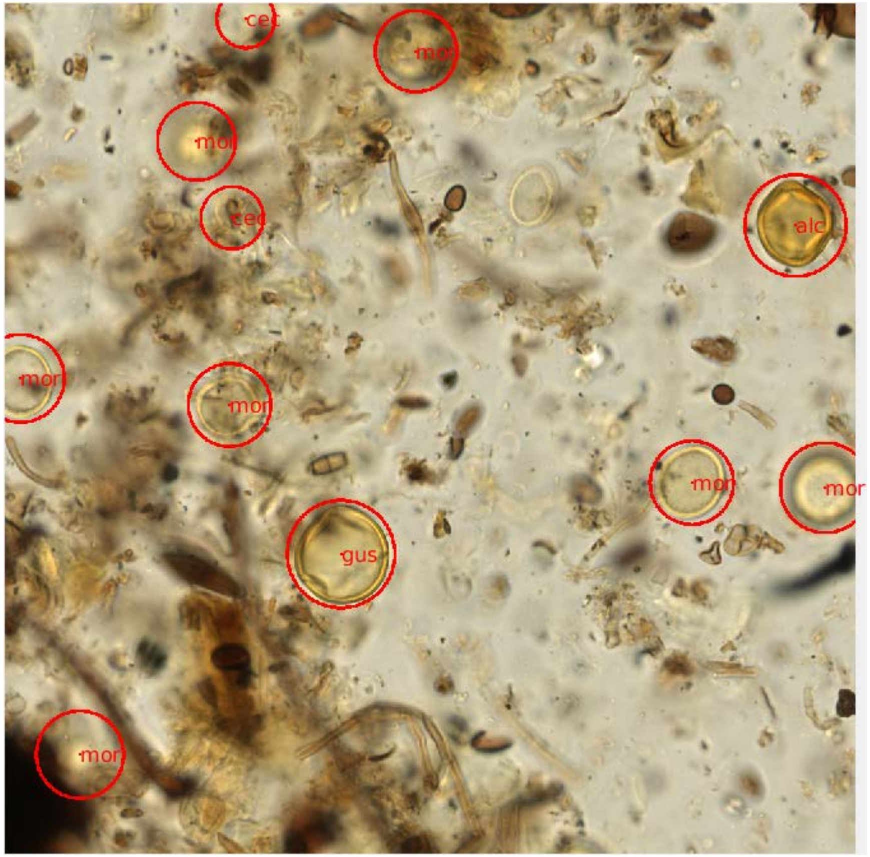
  Virtual: randomized images

(log)

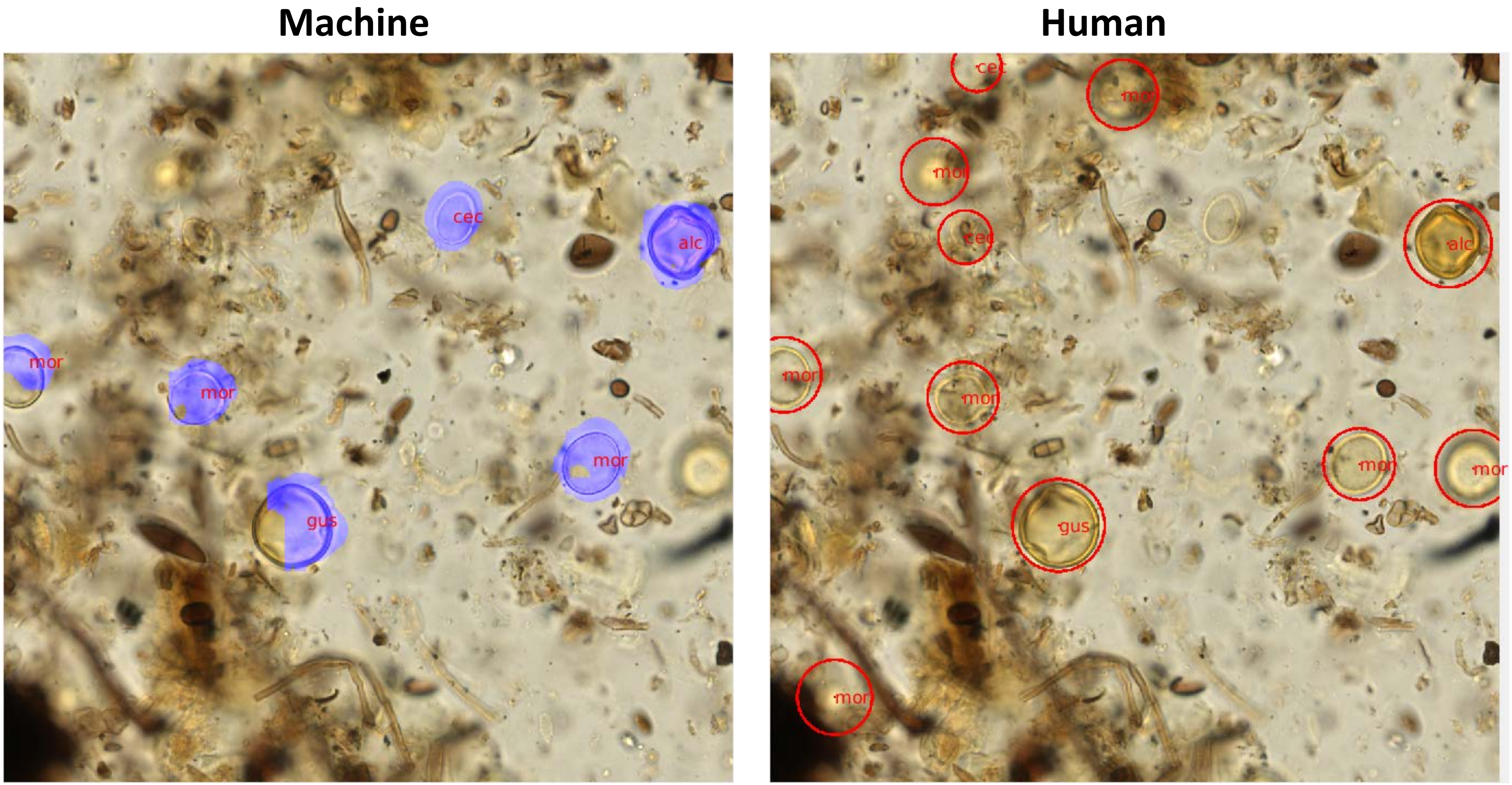
# TRAINING CONVOLUTIONAL NEURAL NETS (CNN) FOR POLLEN IDENTIFICATIONS FROM ANNOTATED IMAGE DATA

- Images were randomly split into training and testing sets
- Annotated training (ground truth) pollen image examples included the pollen id, location coordinate, and pollen grain radius
- CNN searches each image for patterns corresponding to each pollen id morphology
- Non-maximum suppression was used to identify pollen grains according to pollen ornamentation

**Human**



# SIMULTANEOUS POLLEN SEGMENTATION AND IDENTIFICATION



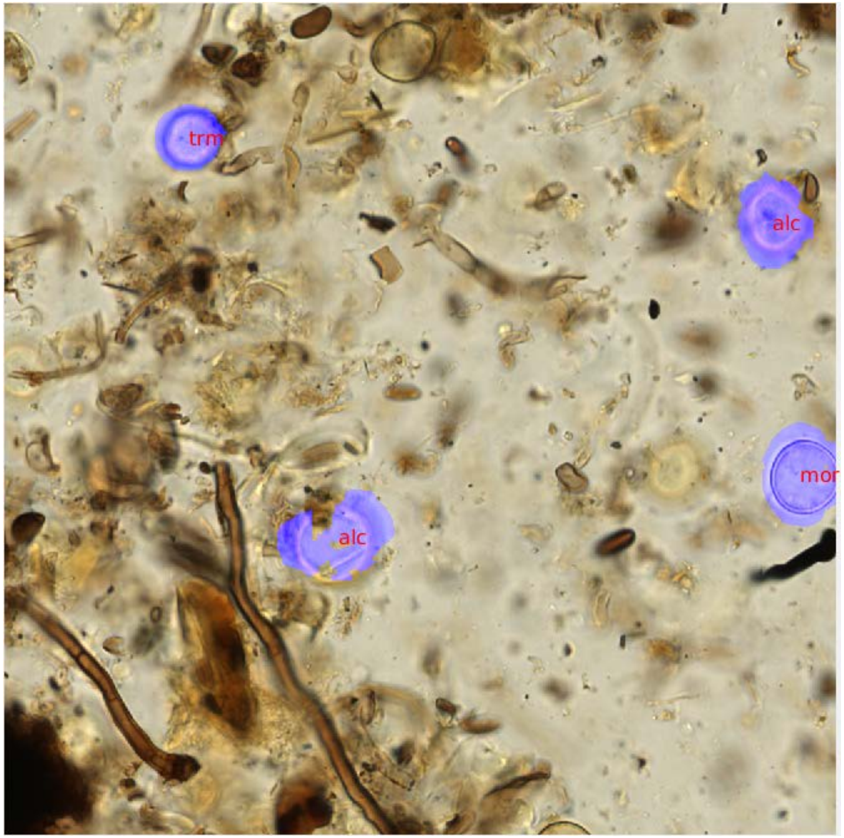
Machine

Human

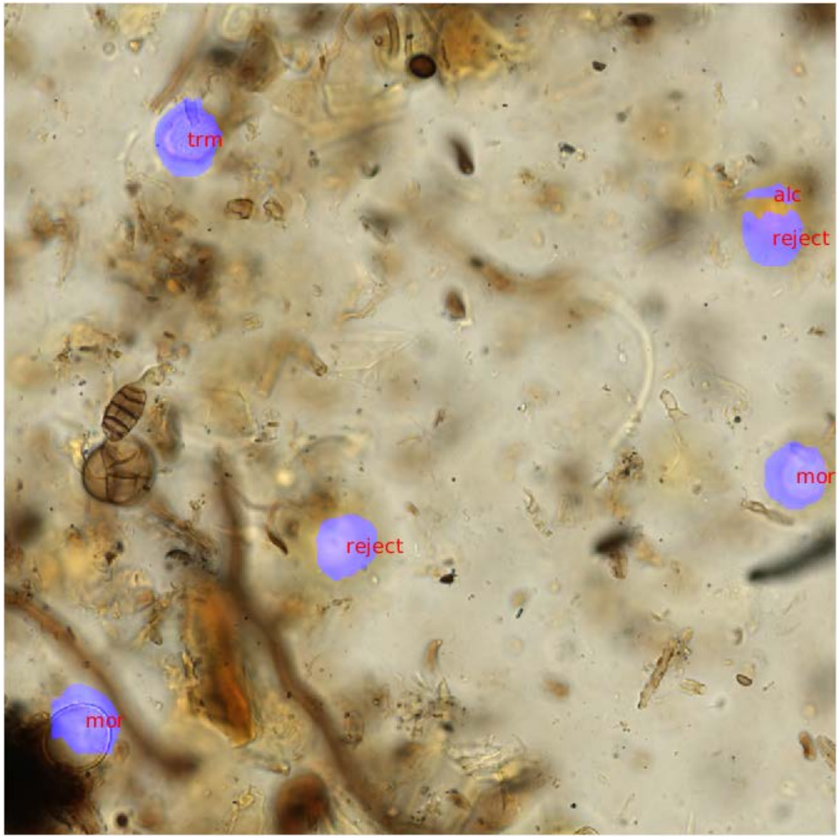
- 48-way classification matrices were constructed using the 47 most abundant pollen types and an additional category called “reject” comprised of pollen types not included in the 47 most abundant

# SIMULTANEOUS POLLEN SEGMENTATION AND IDENTIFICATION

**Machine**



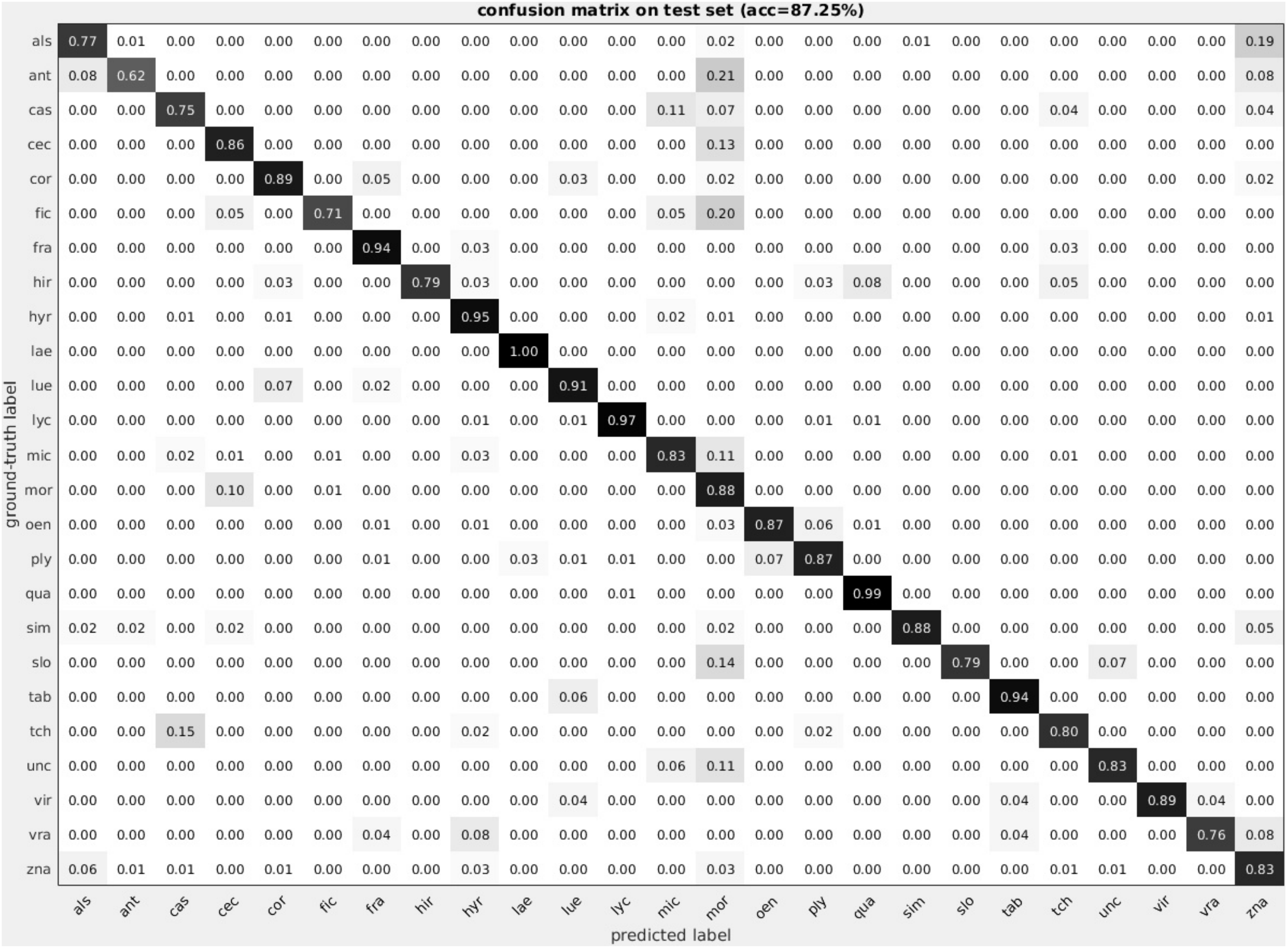
**Machine**



- 48-way classification matrices were constructed using the 47 most abundant pollen types and an additional category called “reject” comprised of pollen types not included in the 47 most abundant

# COMPLETE AUTOMATION: CONFUSION MATRICES

## ~70% accurate on full 47 pollen type training set, 87.25% on 25 most accurate types

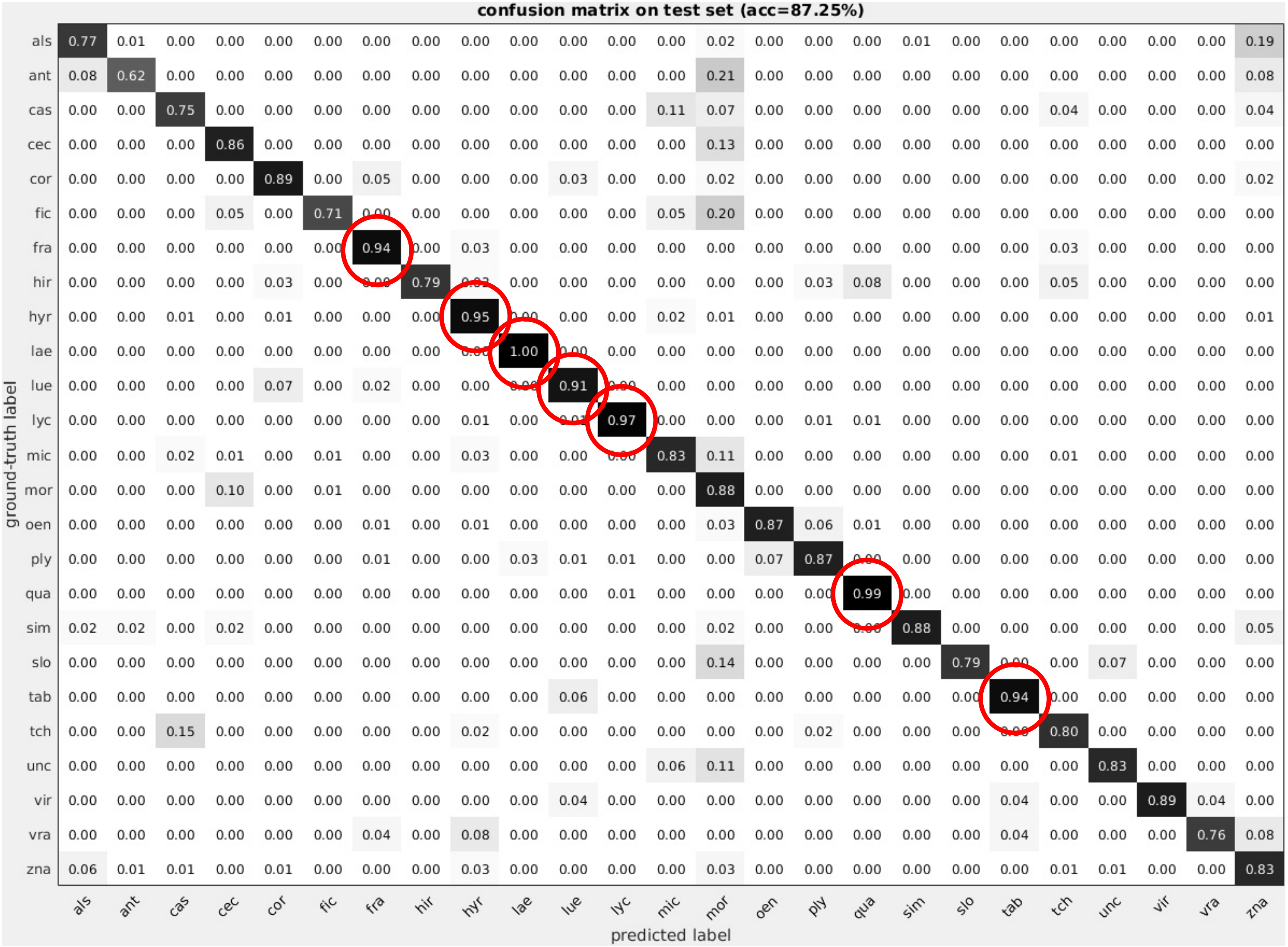

ground-truth label
predicted label

confusion matrix on test set (acc=87.25%)

| | als | ant | cas | cec | cor | fic | fra | hir | hyr | lae | lue | lyc | mic | mor | oen | ply | qua | sim | slo | tab | tch | unc | vir | vra | zna |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| als | 0.77 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.19 |
| ant | 0.08 | 0.62 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 |
| cas | 0.00 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.11 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.04 |
| cec | 0.00 | 0.00 | 0.00 | 0.86 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cor | 0.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| fic | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| fra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| hir | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.79 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.08 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| hyr | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| lae | 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 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| lue | 0.00 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.91 | 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 |
| lyc | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.97 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| mic | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.83 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| mor | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| oen | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.87 | 0.06 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| ply | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.01 | 0.01 | 0.00 | 0.00 | 0.07 | 0.87 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| qua | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| sim | 0.02 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| slo | 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 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.79 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 |
| tab | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tch | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 |
| unc | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.83 | 0.00 | 0.00 | 0.00 |
| vir | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.89 | 0.04 | 0.00 |
| vra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.76 | 0.08 |
| zna | 0.06 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.83 |

# COMPLETE AUTOMATION: CONFUSION MATRICES

>90% accuracy



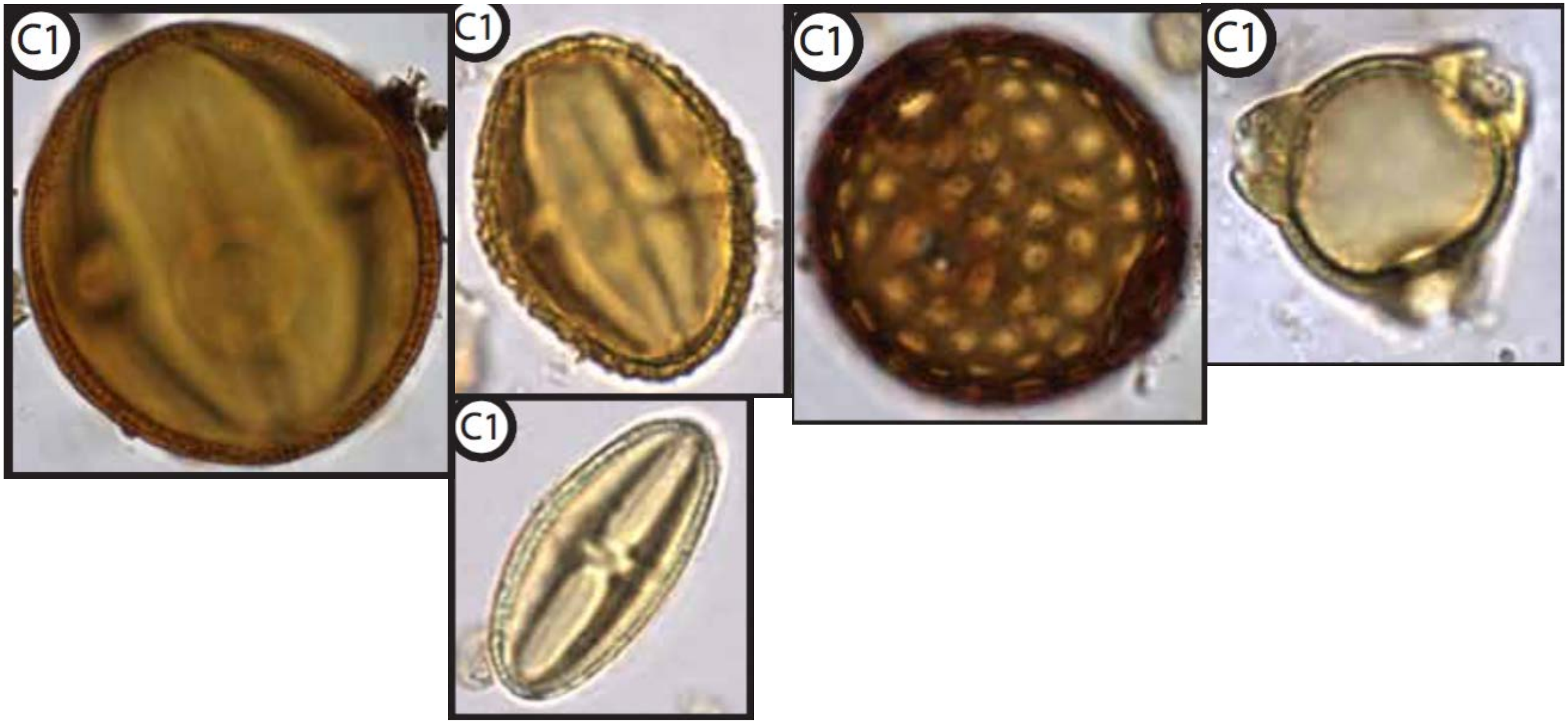
confusion matrix on test set (acc=87.25%)

ground-truth label / predicted label

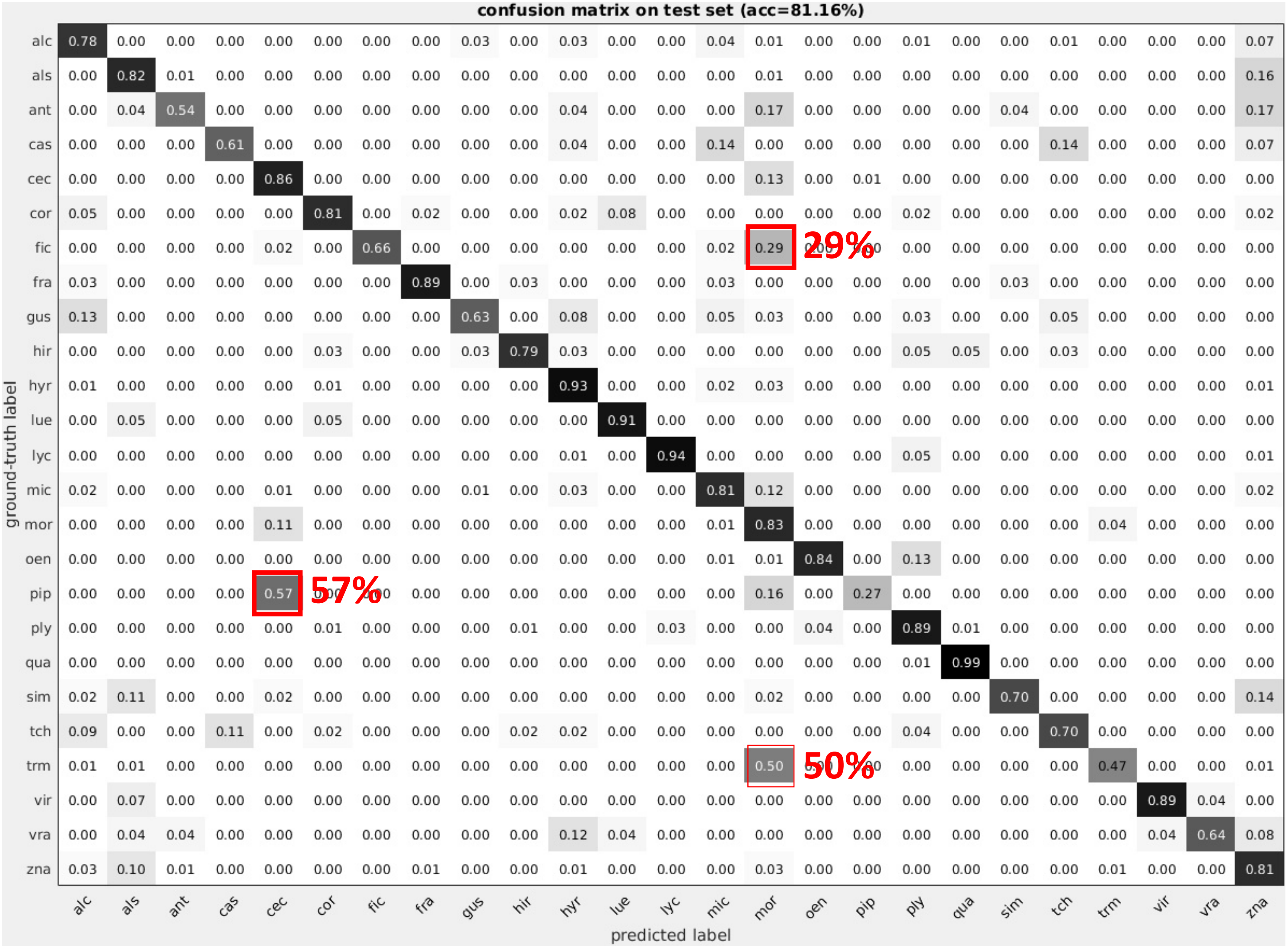
| | als | ant | cas | cec | cor | fic | fra | hir | hyr | lae | lue | lyc | mic | mor | oen | ply | qua | sim | slo | tab | tch | unc | vir | vra | zna |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| als | 0.77 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.19 |
| ant | 0.08 | 0.62 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 |
| cas | 0.00 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.11 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.04 |
| cec | 0.00 | 0.00 | 0.00 | 0.86 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cor | 0.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| fic | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| fra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| hir | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.79 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.08 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| hyr | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| lae | 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 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| lue | 0.00 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.91 | 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 |
| lyc | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.97 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| mic | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.83 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| mor | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| oen | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.87 | 0.06 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| ply | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.01 | 0.01 | 0.00 | 0.00 | 0.07 | 0.87 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| qua | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| sim | 0.02 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| slo | 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 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.79 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 |
| tab | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tch | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 |
| unc | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.83 | 0.00 | 0.00 | 0.00 |
| vir | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.89 | 0.04 | 0.00 |
| vra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.76 | 0.08 |
| zna | 0.06 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.83 |

# COMPLETE AUTOMATION: CONFUSION MATRICES

>90% accuracy



# COMPLETE AUTOMATION: CONFUSION MATRICES



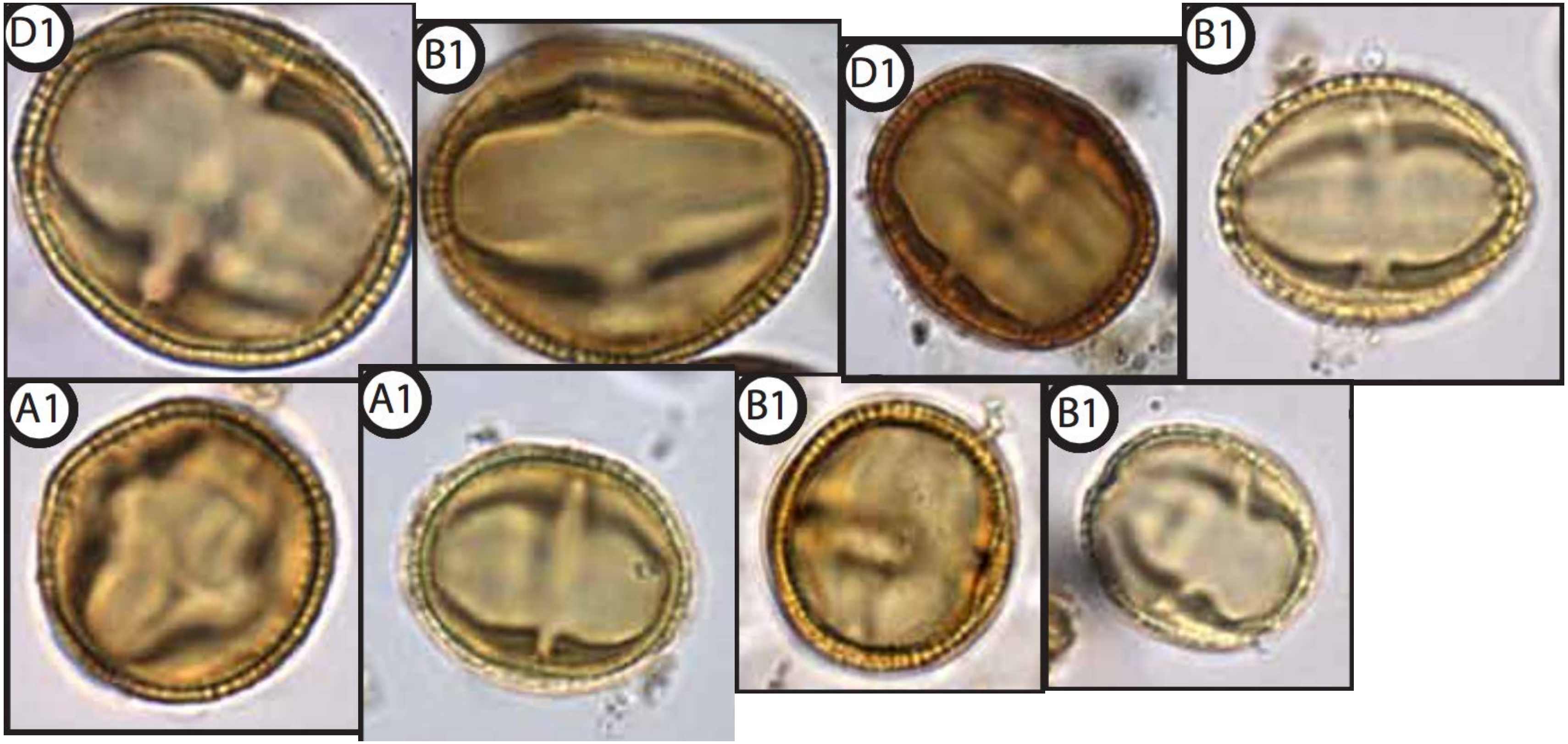
confusion matrix on test set (acc=81.16%)

ground-truth label / predicted label

| | alc | als | ant | cas | cec | cor | fic | fra | gus | hir | hyr | lue | lyc | mic | mor | oen | pip | ply | qua | sim | tch | trm | vir | vra | zna |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alc | 0.78 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.03 | 0.00 | 0.00 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.07 |
| als | 0.00 | 0.82 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.16 |
| ant | 0.00 | 0.04 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.17 |
| cas | 0.00 | 0.00 | 0.00 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.07 |
| cec | 0.00 | 0.00 | 0.00 | 0.00 | 0.86 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cor | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.81 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| fic | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.66 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.29 | [illegible] | [illegible] | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| fra | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| gus | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.63 | 0.00 | 0.08 | 0.00 | 0.00 | 0.05 | 0.03 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| hir | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.03 | 0.79 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.05 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| hyr | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.93 | 0.00 | 0.00 | 0.02 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| lue | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.91 | 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 |
| lyc | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| mic | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.03 | 0.00 | 0.00 | 0.81 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| mor | 0.00 | 0.00 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.83 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 |
| oen | 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 | 0.01 | 0.01 | 0.84 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| pip | 0.00 | 0.00 | 0.00 | 0.00 | 0.57 | [illegible] | [illegible] | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.16 | 0.00 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| ply | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.04 | 0.00 | 0.89 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| qua | 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 | 0.00 | 0.00 | 0.01 | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| sim | 0.02 | 0.11 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 |
| tch | 0.09 | 0.00 | 0.00 | 0.11 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 |
| trm | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | [illegible] | [illegible] | 0.00 | 0.00 | 0.00 | 0.00 | 0.47 | 0.00 | 0.00 | 0.01 |
| vir | 0.00 | 0.07 | 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 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.04 | 0.00 |
| vra | 0.00 | 0.04 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.64 | 0.08 |
| zna | 0.03 | 0.10 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.81 |

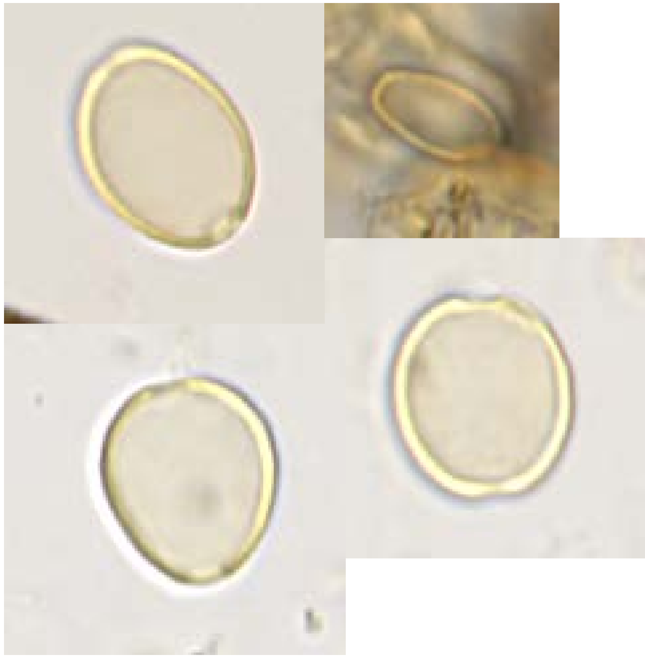
# COMPLETE AUTOMATION: CONFUSION MATRICES

<50% accuracy

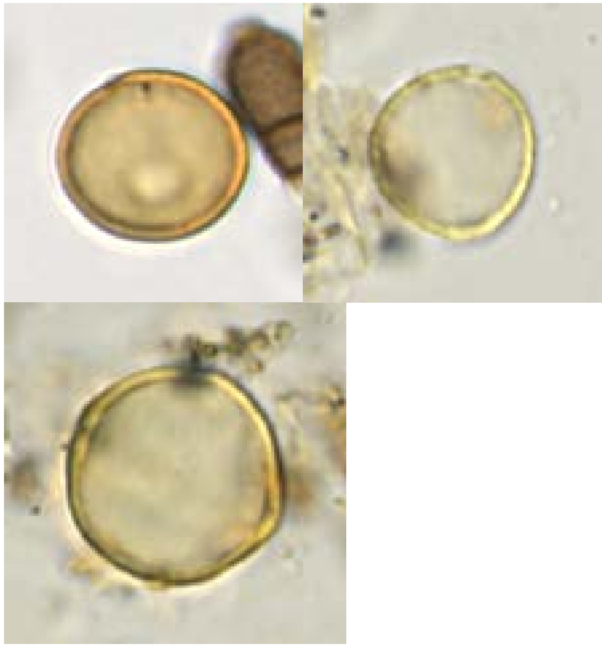


# THE PAIRWISE COMPARISONS IN THE CONFUSION MATRIX SHOWING THE MOST DISAGREEMENT ARE MORPHOLOGICALLY VERY SIMILAR

*Ficus* (**66% accuracy)** was misclassified **29%** of the time as *Brosinum*-type



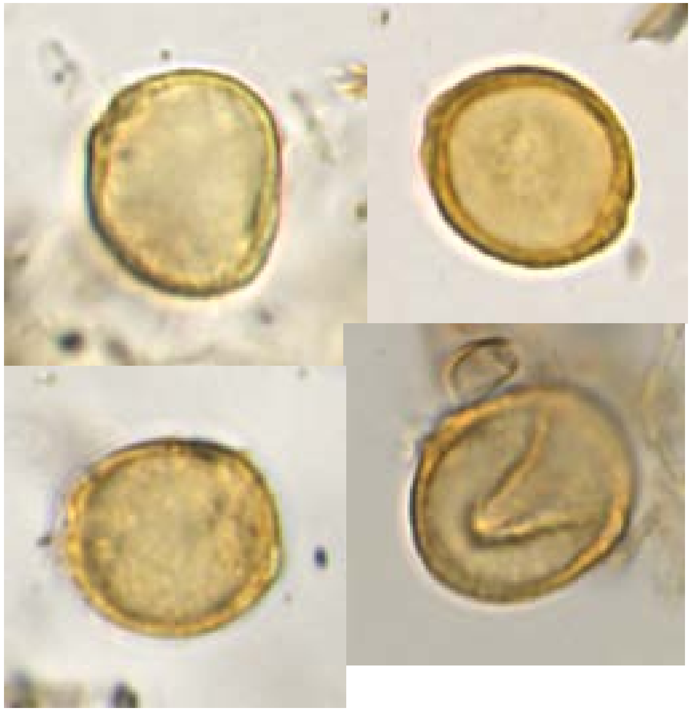
*Ficus* (Moraceae)



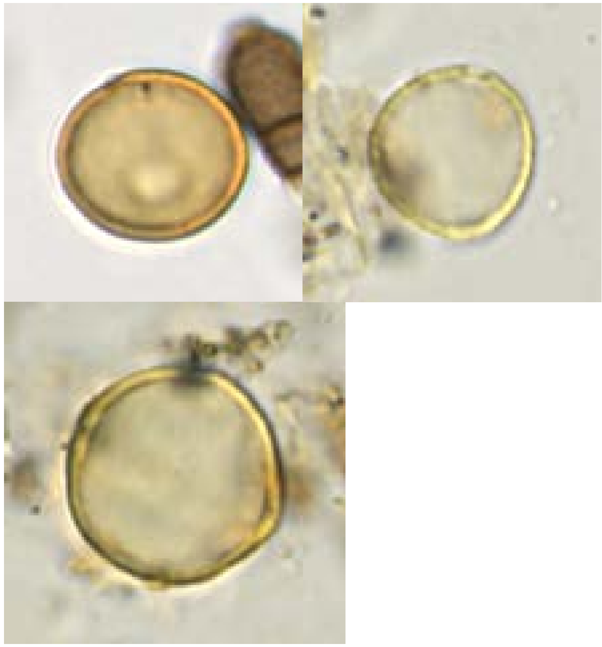
*Brosinum*-type(Moraceae)

# THE PAIRWISE COMPARISONS IN THE CONFUSION MATRIX SHOWING THE MOST DISAGREEMENT ARE MORPHOLOGICALLY VERY SIMILAR

Trema (**47% accuracy**) was misclassified **50%** of the times as *Brosinum*-type



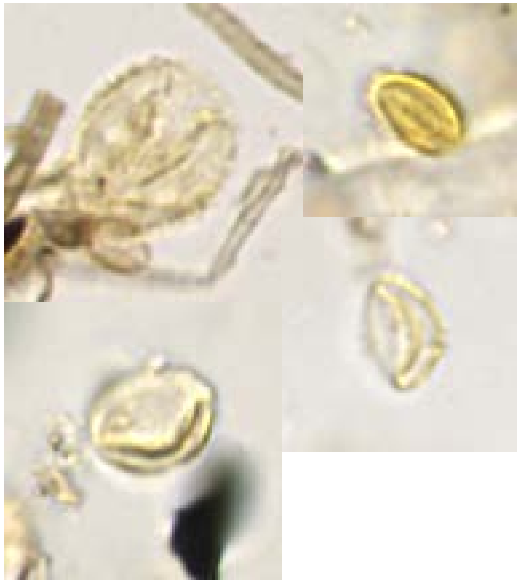
*Trema* (Ulmaceae)



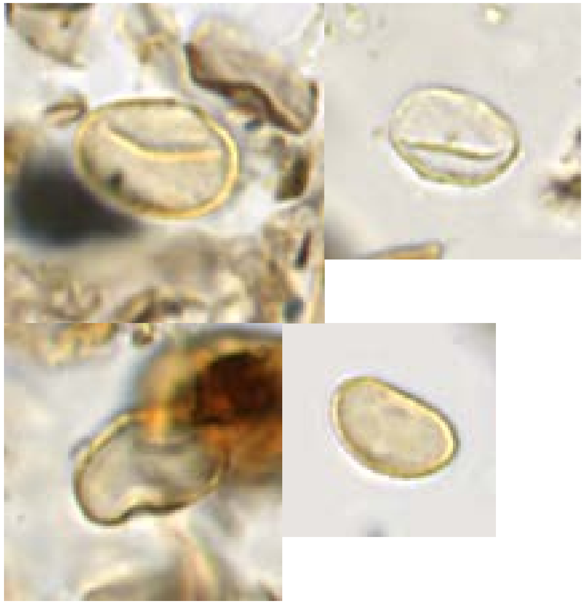
*Brosinum*-type(Moraceae)

# THE PAIRWISE COMPARISONS IN THE CONFUSION MATRIX SHOWING THE MOST DISAGREEMENT ARE MORPHOLOGICALLY VERY SIMILAR

Piperaceae (**27% accuracy**) was misclassified **57%** of the times as *Cecropia*



Piperaceae



*Cecropia* (Urticaceae)

# CONCLUSIONS

- Overall, our results are very promising given this difficult classification problem

- Our preliminary results show that automated segmentation and classification models can distinguish pollen types from hyper-diverse samples

- Model performance is poorest on pollen types that are morphologically very similar

- Predicted outputs should improve as the neural nets are trained on more tagged examples

# FUTURE DIRECTIONS

- The same system can be implemented using training data from herbarium and reference material collections

- Apply the methodology to fossil records

- Create a collaborative pollen identification database that harnesses the expertise of multiple palynologists

- Expand to other proxies
    - Diatoms
    - Phytoliths
    - Cuticles

# ACKNOWLEDGEMENTS

NSF Macrosystems Ecology
NSF Advances in Biological Informatics