Next-generation community ecology:

Exploring ecological and evolutionary drivers of planktonic foraminifera diversity using the Endless Forams database and a supervised machine learning classifier

Allison Hsiang[1,2] & Pincelli Hull[3]

[1] GeoBio-Center LMU, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 Munich, Germany
[2] Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 Munich, Germany.
[3] Department of Geology & Geophysics, Yale University, P.O. Box 208109, New Haven, CT 06520-8109 USA.

LUX ET VERITAS

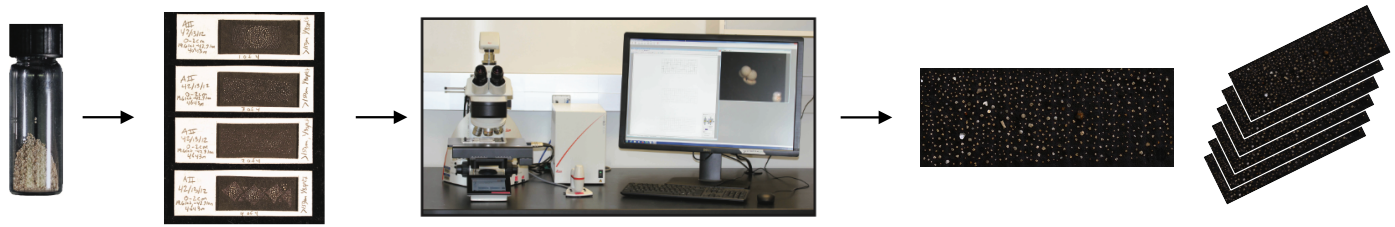
Naturhistoriska riksmuseet

GeoBio-Center LMU München

Paläontologie & Geobiologie LMU München

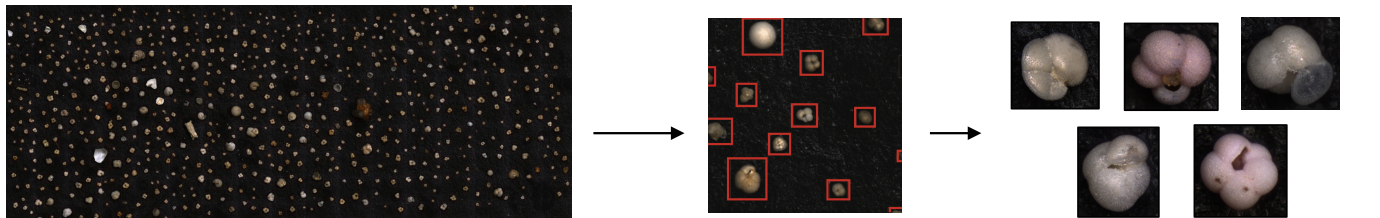
Building Big Data: Imaging pipeline

1. Sieve, arrange, and image samples



Sample | Slides | Leica DM6000 Imaging System | 2D EDF Image | Z-stack Images

2. Segment raw slide images using automatic object recognition



2D EDF Image | Example Zoom (After Object Identification) | Individual Object Images

AutoMorph

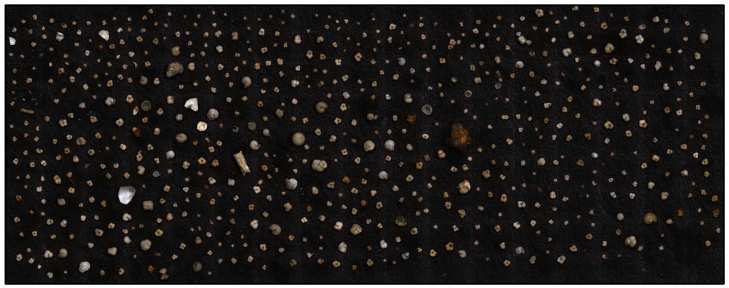
https://github.com/HullLab/AutoMorph

Modified from Hsiang et al. (2017) *AutoMorph:* Accelerating morphometrics with automated 2D and 3D image processing and shape extraction. *Methods in Ecology and Evolution*. 9(3):605-612.

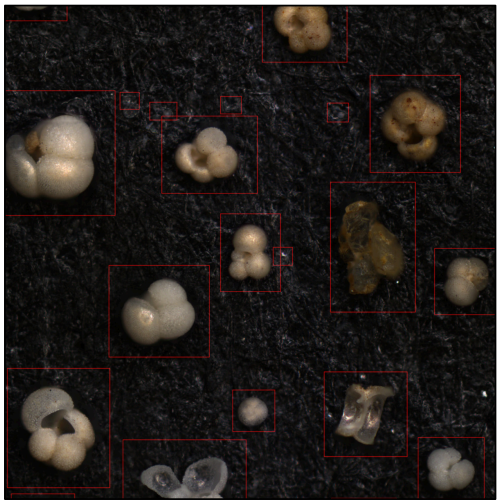
Specimens:

Yale Peabody Museum of Natural History

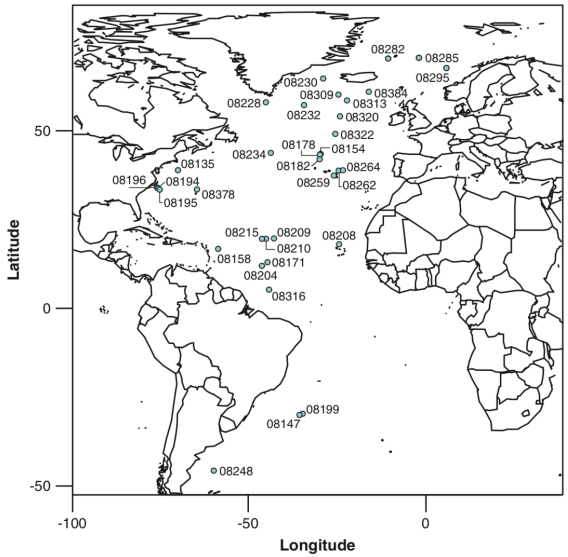
61,849 individuals



IP.307625



IP.307626

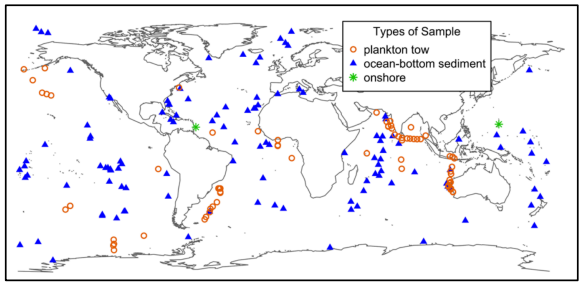


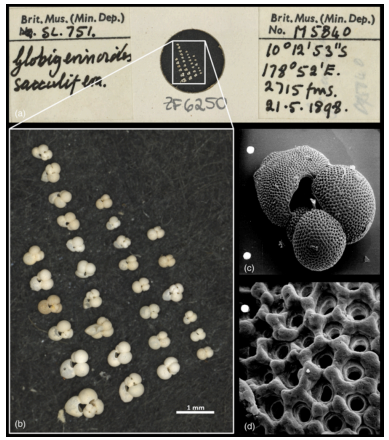
Elder et al. (2018) Sixty-one thousand recent planktonic foraminifera from the Atlantic Ocean. *Scientific Data*. 5:180109

Specimens:

Henry A. Buckley Collection
Natural History Museum, London

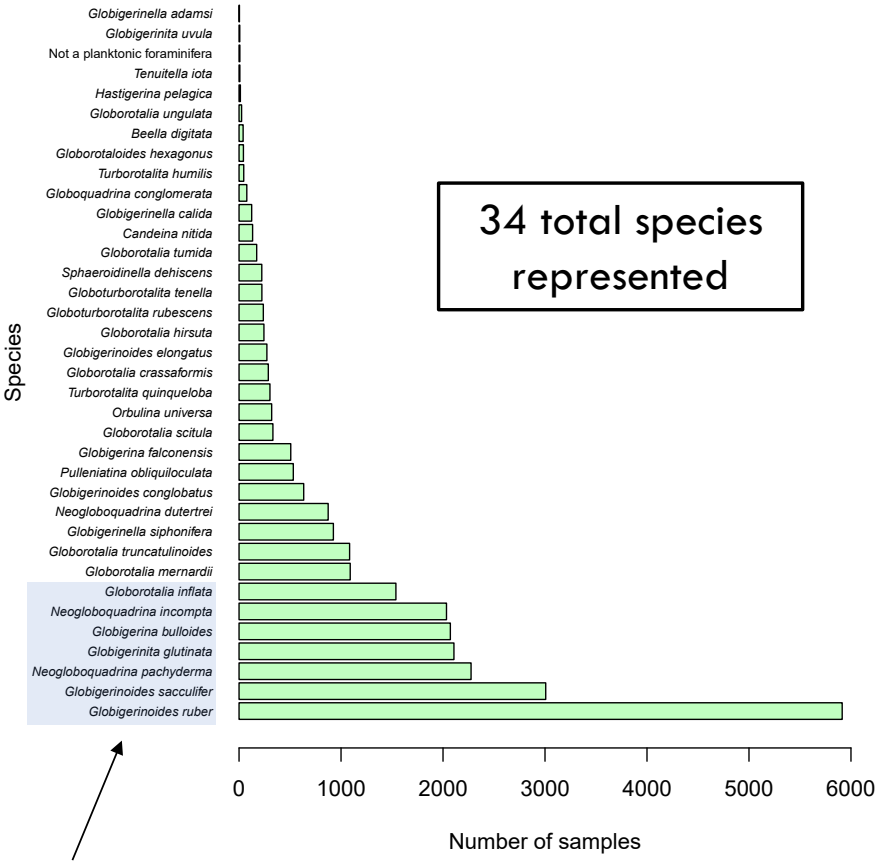
10,071 individuals





Rillo et al. (2016) The unknown planktonic foraminiferal pioneer Henry A. Buckley and his collection at The Natural History Museum, London *Journal of Micropalaeontology*. 36:191-194.

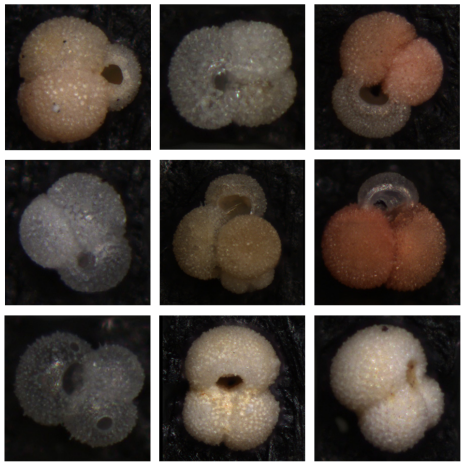
Taxonomic Resource: Diversity and variation



34 total species represented

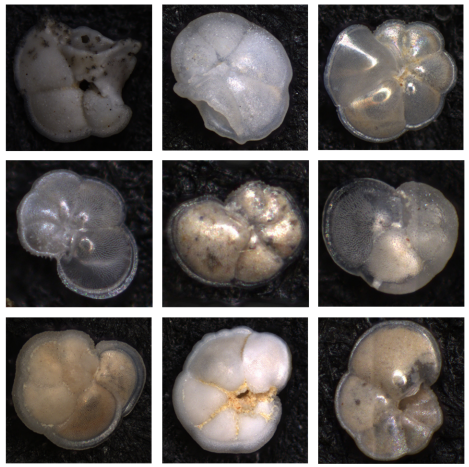
>1,000 images

Globigerinoides ruber



n = 6,442

Globorotalia menardii



n = 1,351

Globigerina bulloides

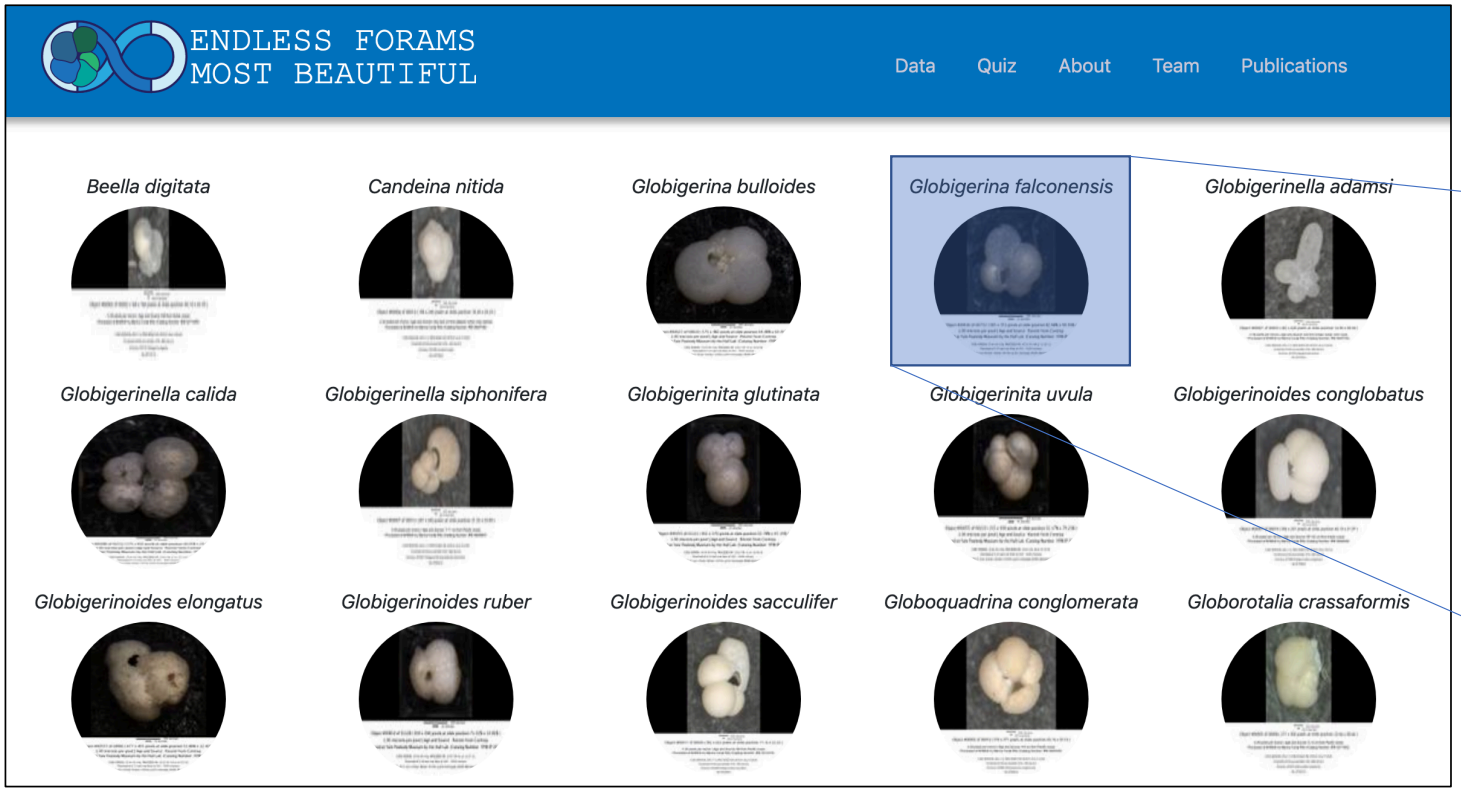
n = 2,608

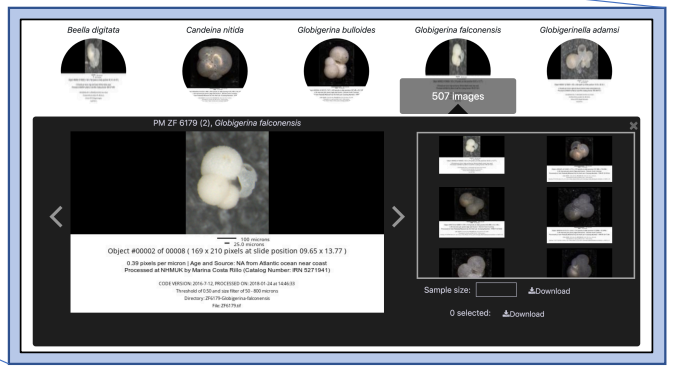
Modified from Hsiang et al. (2019) Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography & Paleoclimatology*. 34(7):1157-1177.

EndlessForams.org

Open-access image and taxonomy database

ENDLESS FORAMS MOST BEAUTIFUL

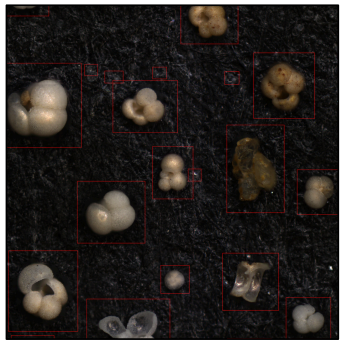




Supervised Machine Learning:
Training and validation sets

24,659 individuals
(high-quality labels)

YPM

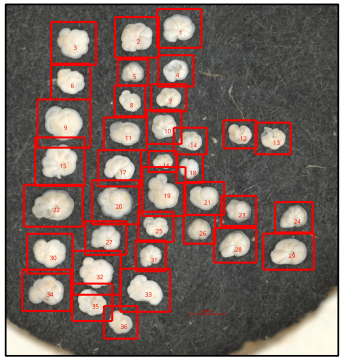


80% → Training: 19,680 individuals

20% → Validation: 4,889 individuals

10,071 individuals

NHM



80% → Training: 8,057 individuals

20% → Validation: 2,014 individuals

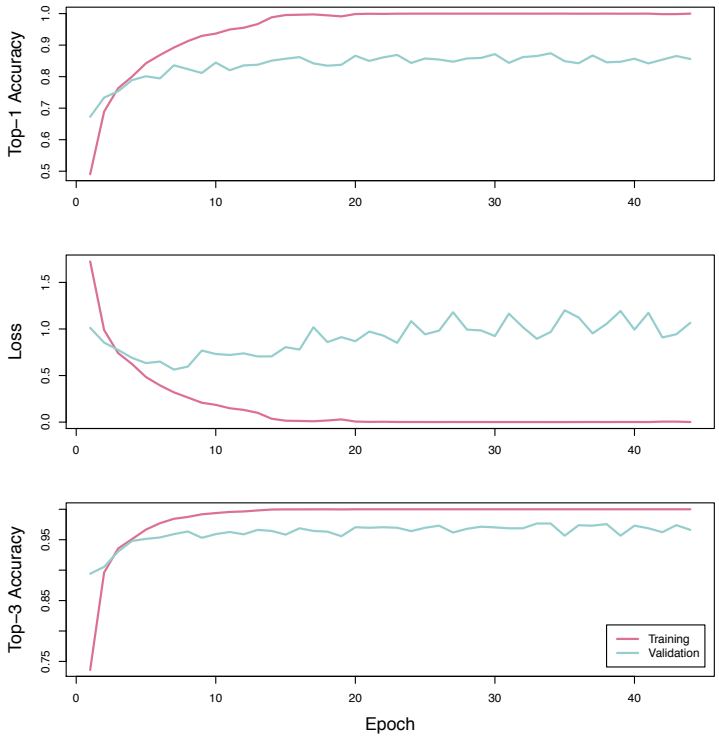
Final Datasets

Training: 27,737 individuals

Validation: 6,903 individuals

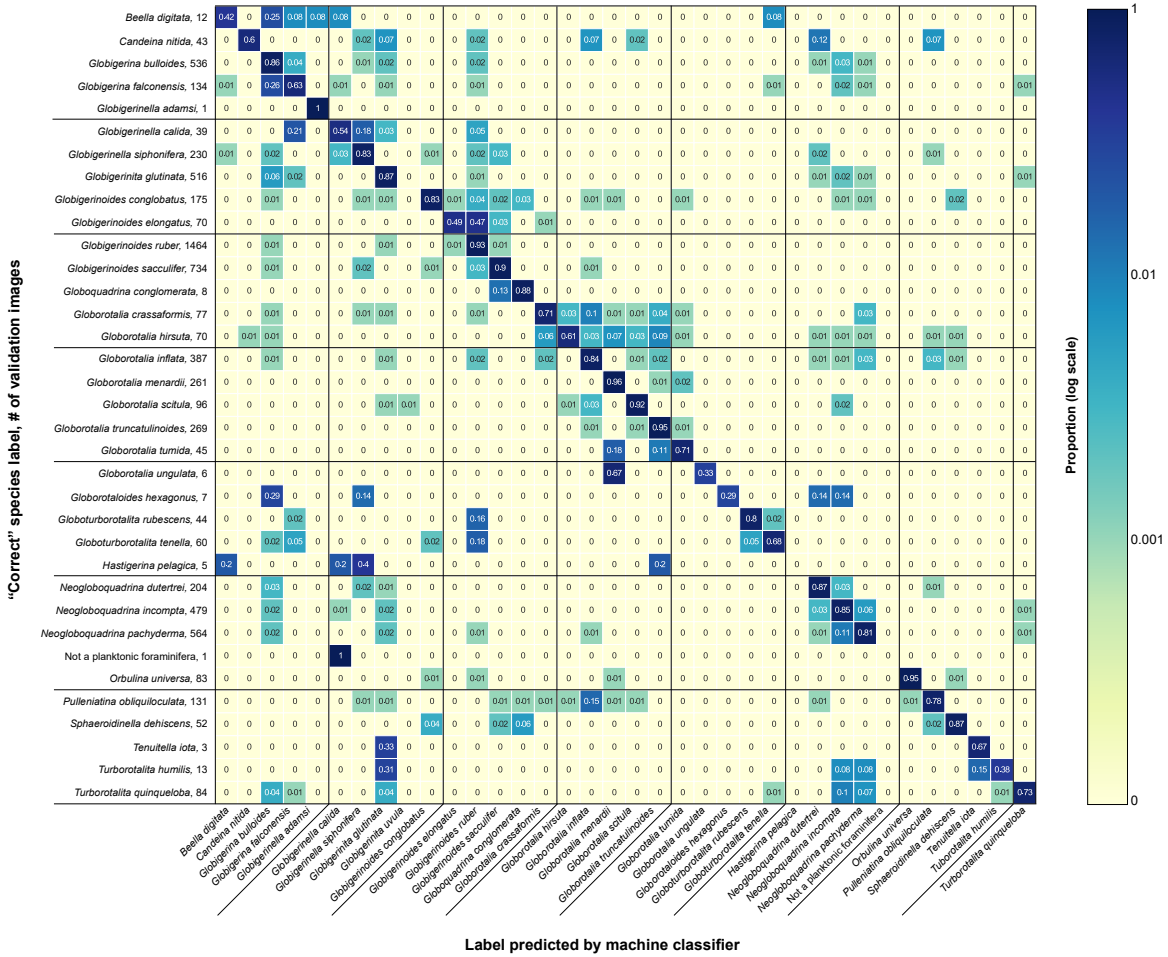
Supervised Machine Learning:

Machine accuracies



Maximum Top-1 Validation Accuracy: 87.41%
Maximum Top-3 Validation Accuracy: 97.66%

Average human accuracy:
71% (range: 64-85%)



Modified from Hsiang et al. (2019) Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography & Paleoclimatology*. 34(7):1157-1177.

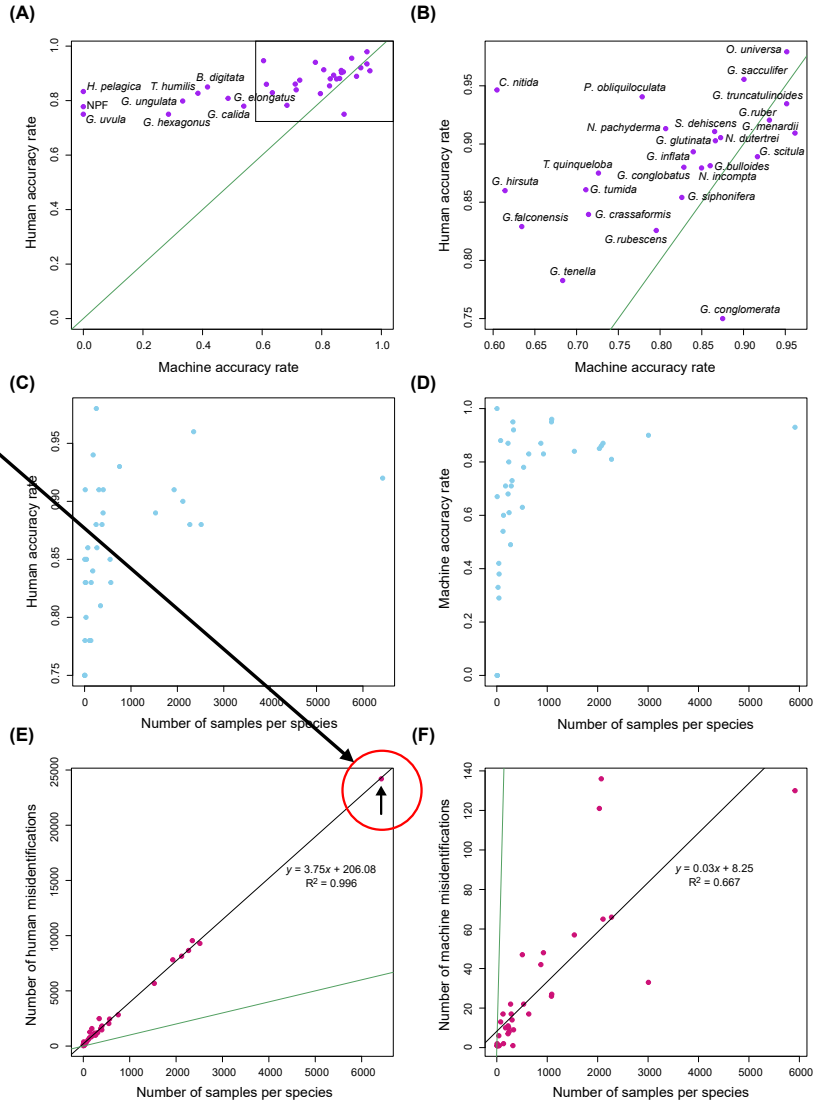
Supervised Machine Learning:
Human vs. machine performance

Globigerinoides ruber
(6,425 individuals)

24,202 identifications
mistaken for *G. ruber*

Human mistakes are more phylogenetically conservative

However, machine mistakes often repeat historical ambiguities (e.g., *Tenuitella iota* → *Globigerinita glutinata*)



Modified from Hsiang et al. (2019) Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography & Paleoclimatology*. 34(7):1157-1177.