☀ **Smithsonian**   **Menu**
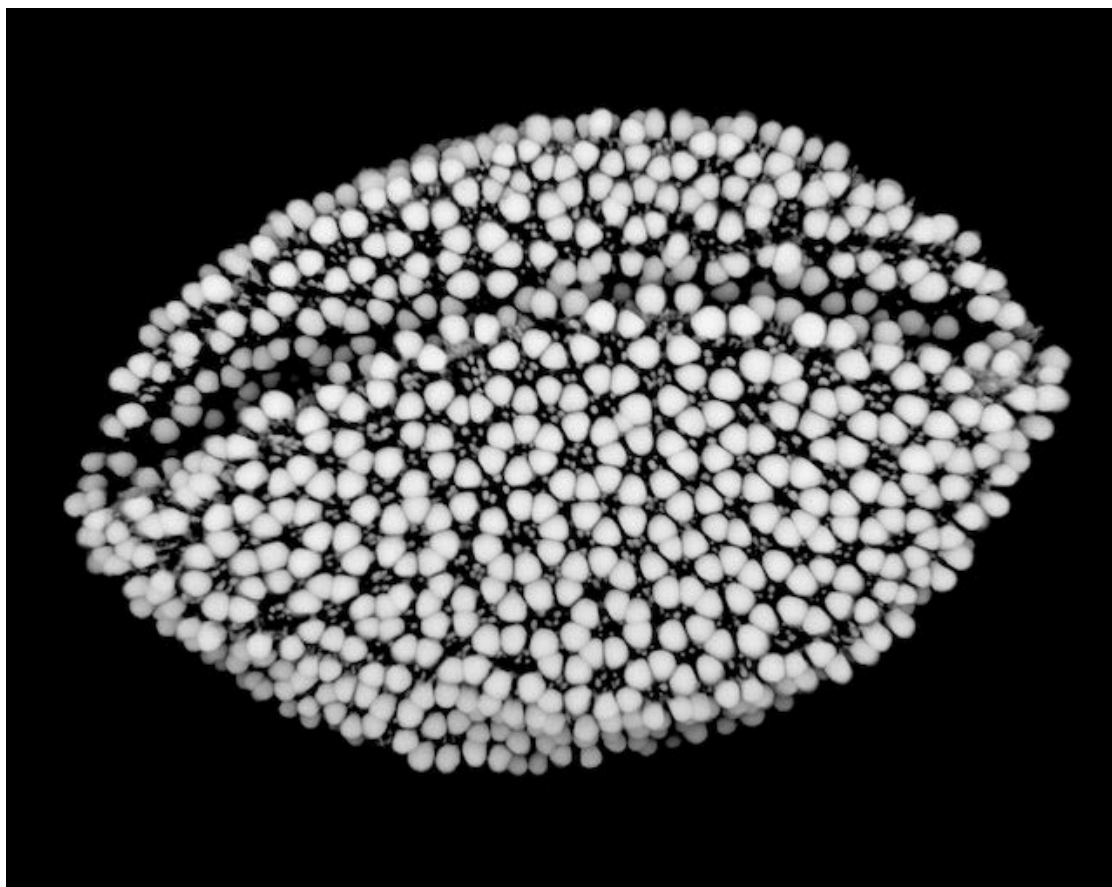
Smithsonian Digitizes Pollen From 18,000 Plant Species

Digitized Pollen Database for Paleontology Research, Allergy Medicine and More

News Release   •   August 1, 2025   •

Croton capitatus, one of more than 40 million recently digitalized images of pollen from the Smithsonian collections.

Media Contact

Beth King
> kingb@si.edu
> 202-633-4700 x 28216

Benjamin Marcus
> marcusba@si.edu
> 202-633-5194

A team of researchers from the Smithsonian Tropical Research Institute is digitizing images of pollen from more than 18,000 plant species from the tropics. These images are being used to train a machine-learning model to identify pollen grains, a job that usually takes hundreds of hours of microscopy work by pollen experts. The images also will make a wide range of new pollen analyses possible. The database, called PollenGEO, will be free online.

The Smithsonian pollen collection housed at the [Smithsonian Tropical Research Institute](#) (STRI) and the Smithsonian's [National Museum of Natural History](#) contains more than 18,000 species, making it one of the largest pollen collections in the world. Pollen reference database such as PollenGEO could potentially serve a myriad of functions in many areas of science and medicine. For example, quick and accurate pollen identification can help diagnose a pollen allergy, pinpoint where clothing at a crime scene came from, help investigate how ancient forests responded to climate change and date hydrocarbon deposits.

Pollen's value in paleontology derives from its durability—with some pollen grains lasting hundreds of millions of years, offering a window into Earth history that is precise in both time and space. Also, each plant species' pollen is unique and distinct from other species.

Previously, specialists identified pollen grains one by one under a microscope using illustrated handbooks as a reference. That process is very time consuming and can be particularly challenging in the tropics where there are thousands of plant species, many of them not yet identified. It is also challenging to identify pollen in ancient layers of rock because many of the plant species that produced the pollen are now extinct.

To resolve these challenges, more than 30 researchers and students at STRI led by staff palynologist, Carlos Jaramillo, are digitizing the entire Smithsonian palynological collection; they are uploading more than 40 million photos of pollen grains from known plant species to create a massive database. This dataset will be used to train AI models to aid pollen identifications.

Most samples derive from the Graham Palynological Collection, donated to STRI in 2008, which holds about 18,000 species of mostly tropical pollen on more than 23,000 microscope slides, each accompanied by an index card that describes the sample. About 100 volunteers working through the Smithsonian Transcription Center helped enter the information from the cards into the database. The collection also includes the Joan Nowicke collection, the Barro Colorado Island collection by Dave Roubik and Enrique Moreno, the Amazon collection made by Paul Collinvaux, and the Sian Ka'an collection, which contains 650 species from southeastern Mexico. In addition, approximately 1,000 fossil pollen samples have been scanned from museum collections at the National Museum of Natural History.

Training an AI model to use this massive database to identify samples required collaboration among experts in a range of fields, from botany to computer science. Associate professor Surangi Punyasena from the University of Illinois Urbana-Champaign is constructing the AI environment. Jaramillo's team is part of the Trans-Amazon Drilling project, a large-scale project using pollen in drilling cores from across the Amazon to understand the history of the forest. This project includes researchers from several institutions, such as the Universidade Federal de Mato Grosso and the Universidade Federal do Acre, both in Brazil, and the Open University.

The availability of PollenGEO and other online pollen databases will transform pollen identification from a solitary activity behind a microscope to a digital and universally accessible process.

# # #

Reference: Carlos Jaramillo et al. Digitizing collections to unlock the full potential of palynology: A case study with the Smithsonian palynology collection. Plants People Planet. doi:10.1002/ppp3.70073
SI-179-2025

Tags: [Research News](#)