Received: 6 November 2024 | Revised: 9 June 2025 | Accepted: 10 June 2025

DOI: 10.1002/ppp3.70073

Plants People Planet PPP Open Access

METHODS AND TECHNIQUES

# Digitizing collections to unlock the full potential of palynology: A case study with the Smithsonian palynology collection

Carlos Jaramillo[1] | Surangi W. Punyasena[2] | Daurys de Alba[1] | Roxana Alveo[1] | Angelica Arcila[1] | Jorge Bermudez[1] | Jonatan Bustos[1] | Dayenari Caballero-Rodriguez[1] | Karen Cardenas[1] | David Caro[1] | Francy Carvajal[1] | Ivonne Marcela Castañeda[1] | Shara Chaves[1] | Carlos D'Apolito[3] | Andres Diaz-Jaramillo[1] | Laura Diaz[1] | Luisa Gomez[1] | Mauricio León-Carreño[1] | Paula Lopera[1] | Maria Alejandra Lopez[1] | Priscila Lopez[1] | Luke Mander[4] | Jhonatan Martínez Murcia[1] | Carlos Moreno[1] | Enrique Moreno[1] | Enrique Neyra[1] | Brenda Orosco[1] | Jhon Ortiz[1] | Natalia Ossa[1] | Natalia Ovalle[1] | Carolina Ovalle[1] | Angelo Plata[1] | Ingrid Romero[1,5] | Bruno Scudeiro[1] | Silane A. F. da Silva Caminha[6] | Axel Tejada-Fajardo[1] | Vinicius Do Valle[1] | Thiago Wood[1]

[1]Smithsonian Tropical Research Institute, Panama City, Panamá

[2]University of Illinois, Urbana-Champaign, Urbana, Illinois, USA

[3]Universidade Federal do Acre, Rio Branco, Acre, Brazil

[4]The Open University, Milton Keynes, UK

[5]Smithsonian National Museum of Natural History, Washington, DC, USA

[6]Universidade Federal de Mato Grosso, Cuiaba, Mato Grosso, Brazil

**Correspondence**
Carlos Jaramillo, Smithsonian Tropical Research Institute, Balboa, Ancón, Panama City, Panamá.
Email: jaramilloc@si.edu

**Funding information**
Rubinoff Big Bet Endowment; Gregory D. and Jennifer Walston Johnson; 1923 Fund; Smithsonian Women's Committee, Grant/Award Number: SWC 13; The Anders Foundation; Smithsonian Tropical

**Societal Impact Statement**

Large palynological collections have been built over decades and contain vital information. However, they are often difficult to access and use effectively. What is the point of having such collections if they are not fully utilizable? To solve this problem, we digitized the Smithsonian palynological collection using both light and confocal microscopy. We digitized the pollen of 12,000 species and took 40 million photos. Our image library will support a wide array of applications, including environmental monitoring, public health, biodiversity studies, paleoclimate, and the analysis of landscape changes across spatial and temporal scales. It will also aid in geological correlations used in water exploration and in hydrocarbon storage/production.

**Summary**

- Palynology is a century-old practice, contributing data to various fields, from geology to medicine and forensics. Palynological analyses are highly time-consuming and involve visually finding, identifying, and counting thousands of palynomorph grains on microscope slides. These analyses are especially challenging in high-diversity tropical settings. Fortunately, the development of deep learning and the

Research Institute; Smithsonian Institution, Grant/Award Number: SILifeonaSustainablePlanetPathfinder

capability to digitize entire microscope slides are allowing palynology to enter a new era. Foundational to this transformation is building solid digital collections that can be achieved by digitizing botanical collections.

- We are digitizing the Smithsonian palynological collections, which contain ~18,000 species, most of which are Neotropical taxa in the *Graham Pollen Collection*. This digital product consists of high-resolution images of different types—transmitted light, differential interference contrast, and optical superresolution (Airyscan)—which will be freely available and lay the groundwork for training deep-learning models and applying novel image analysis to palynomorph morphology.
- Image quality matters, so we outline the best practices we have developed throughout the years of imaging and experimentation.
- High-resolution imaging of palynological collections holds the key to unraveling the full potential that the study of pollen and spores can offer.

**KEYWORDS**
digitization, paleobotany, paleoecology and paleontology, palynology, pollen and spores

Palynology is a century-old practice, contributing data to various fields, from geology to medicine and forensics. Palynological analyses are highly time-consuming and involve visually finding, identifying, and counting thousands of palynomorph grains on microscope slides. These analyses are especially challenging in high-diversity tropical settings. Fortunately, the development of deep learning and the capability to digitize entire microscope slides are allowing palynology to enter a new era. Foundational to this transformation is building solid digital collections that can be achieved by digitizing botanical collections. We are digitizing the Smithsonian palynological collections, which contain ~18,000 species, most of which are Neotropical taxa in the *Graham Pollen Collection*. This digital product consists of high-resolution images of different types—transmitted light, differential interference contrast, and optical superresolution (Airyscan)—which will be freely available and lay the groundwork for training deep-learning models and applying novel image analysis to palynomorph morphology. Image quality matters, so we outline the best practices we have developed throughout the years of imaging and experimentation. High-resolution imaging of palynological collections holds the key to unraveling the full potential that the study of pollen and spores can offer.

## 1 | INTRODUCTION

The discipline of palynology is over a century old (Edwards et al., 2017). It is used in various fields of research, from pollination ecology to vegetation paleoreconstructions and oil exploration. The core of any routine palynological analysis is classifying and counting pollen/spore grains. Thousands of isolated grains need to be detected and tallied and then identified through comparison to the pollen morphology of described plants, a task that is especially challenging in tropical settings because of the thousands of potential plant species that could serve as pollen sources. How do we accomplish this? A palynologist spends hundreds of hours scanning a microscope slide locating and then identifying pollen grains. Mastering palynology takes years of training, particularly in tropical settings. However, this training enables the recognition of only a fraction of taxa, often limited to a specific geographic area and a specific geological time. This problem is compounded when dealing with the fossil record as there are thousands of fossil pollen types that belong to extinct species. Another major caveat is that individual identifications are rarely confirmed by other experts, as the position of every grain on the microscope slide is rarely recorded. Counts are therefore not reproducible, and identifications are often trusted at face value. Furthermore, the identification of the isolated grains represents a major challenge. Across many plant clades, there is correlation between pollen morphological similarity and phylogenetic distance at species and genus level, but at higher hierarchical levels, this correlation is lost (Mander et al., 2021). Therefore, there are very few clades where a pollen grain trait is found as a synapomorphy at the family level. Classification is even more challenging for paleopalynologists because an isolated grain must be compared not only to modern taxa but also to the fossil taxa that have been formally described in the paleontological literature, that is, the roughly 4600 fossil taxa described for the tropics spanning the early Cretaceous to the Pleistocene (Jaramillo & Rueda, 2024). The analyst needs to review dichotomous keys in and reference published images to search published catalogs searching for the taxon that matches the mystery grain. It is a long and very inefficient process without global digital databases.

Over the past decade, there have been multiple efforts to produce digital palynological databases, most of them being open access (Table 1, Notes S1). For example, we have a database that

**TABLE 1** Digital pollen databases including extant and/or extinct taxa. None of them have confocal images. Photos taxa: number of photos per taxa. P/A, presence/absence of photos with multiple focal planes.

| Database | N°taxa | N°photos | Fossil/extant | Taxa | Dynamic photos (P/A) | Source |
|---|---|---|---|---|---|---|
| APSA | 16,099 | 45,000 | Both | 0–9 | 0 | https://apsa.anu.edu.au/samples/ |
| PalDat | 4533 | 36,207 | Extant | 2-8 | 0 | https://www.paldat.org/search/A |
| MUPA | 4408 | 27,645 | Extant | 2-8 | 0 | https://data.oreme.org/observation/pollen |
| GPP | 2340 | 2771 | Both | 1–3 | 1 | https://globalpollenproject.org |
| Palsys | 6800 | 25,000 | Fossil | 0-22 | 0 | https://palsys.org/genus |
| APD | 1165 | 5800 | Extant | 2-7 | 0 | https://africanpollendatabase.ipsl.fr/#/photos |
| HIP | 876 | 4300 | Extant | 2–10 | 0 | https://keyserver.lucidcentral.org/key-server/data/0f030b07-0200-4b0f-8,509-0a0808060703/media/Html/index.html |
| Discoverlife | 455 | 572 | Extant | 1–18 | 0 | https://www.discoverlife.org/mp/20p?res=120&see=I_POL/0005 |
| Pollen wiki | 2577 | 4061 | Extant | 3–24 | 0 | https://pollen.tstebler.ch/MediaWiki/index.php?title=Kategorie:Bild |
| NZFSP | 830 | 5000 | Fossil | 0–16 | 0 | https://pal.gns.cri.nz/catalog/index.htm |
| PCU | 315 | 2617 | Extant | 1–16 | 0 | https://science.uct.ac.za/plant-conservation/resources-databases/pcu-pollen-database |
| Pollen atlas | 173 | 1038 | Extant | 6 | 0 | https://pollenatlas.net/atlas/pollen-profiles |
| NPP ID | 1635 | 2169 | Fossil | 1–4 | 0 | https://non-pollen-palynomorphs.unigoettingen.de/NPPDatabase.html |
| MN-MP | 160 | 900 | Extant | 3–10 | 0 | https://pollen.jimdofree.com/galerie/pollen/ |
| MDC | 151 | 151 | Fossil | 1 | 0 | https://www.ucl.ac.uk/GeolSci/micropal/palynology.html |
| DC | 73 | 304 | Fossil | 1–9 | 1 | https://data.nhm.ac.uk/dataset/duxbury-collection-database |
| NPR | 6271 | 34,000 | Both | 5–6 | 1 | https://research.fit.edu/paleolab/pollen-database/ |
| STRI database | 5613 | 25,000 | Both | 5–6 | 0 | https://biogeodb.stri.si.edu/jaramillosdb/web/ |

contains the illustrations of 5613 taxa of fossil and extant pollen and spores from the Neotropics (4633 fossils, 980 extant) (Jaramillo & Rueda, 2024). Most of the illustrations are sourced from the publications themselves and typically depict only one or two planes of individual grains at most. However, there is still no database with many photos per taxon where the photos capture the complete depth of the grain (multiple focal planes) and the image appears as the grain would appear under a microscope. All current databases offer only a few focal planes (1–3) and a limited number of grains (Table 1).

Fortunately, microscopy and image analysis have gone through a revolution over the past decade to the point that the development of deep learning and the capability to digitize an entire microscope slide at multiple focal planes have made it possible for palynology to enter a new era. This paper outlines our experiences and recommendations for incorporating these approaches into the pollen analysis workflow. The manuscript is divided into two major sections. In Section 1, we describe best practice methods for digitizing a palynological collection at a massive scale to yield results that are useful for both long-term digital archival of a pollen specimen and for building image datasets useful for training neural networks. In Section 2, we explore some of the uses of digitized images that we believe are just the beginning of a revolution in the analysis of palynological images that will incorporate new methods and broaden the questions that can be addressed using palynological data.

## 2 | SECTION 1: DIGITIZING THE SMITHSONIAN PALYNOLOGICAL COLLECTION (EXTANT AND FOSSIL)

A growing number of studies demonstrate the feasibility of using artificial intelligence in palynological studies, including our own (Adaimé et al., 2023; Adaimé et al., 2024; Punyasena et al., 2022; Romero, Kong, et al., 2020). We are now scaling up our imaging and model development efforts to encompass the entire plant kingdom, with a focus on Neotropical clades. One of the most critical steps in using artificial intelligence for image analysis is building a robust training set. We are in the process of creating an ambitious dataset of hundreds of thousands of pollen images by digitizing the Smithsonian palynological collection and the Smithsonian collection of fossil pollen from the Neogene of tropical America.

We are digitizing 20 grains per taxon, using both brightfield (BF) and differential interference contrast (DIC) microscopy and five grains per taxon using confocal (CF) microscopy of each of the extant taxa in the collection, and 50 grains per taxon of the most abundant fossil taxa of the Neogene of the Neotropics in both BF and DIC (150 fossil species) (Figure 1).

We describe the process in detail in the headings that follow, as it requires several steps given the large file size and number of the images and the number of taxa to be digitized. We built a database network, named "PollenGeo," to manage the digitization process. The microscope equipment consists of four upright microscopes: a Zeiss

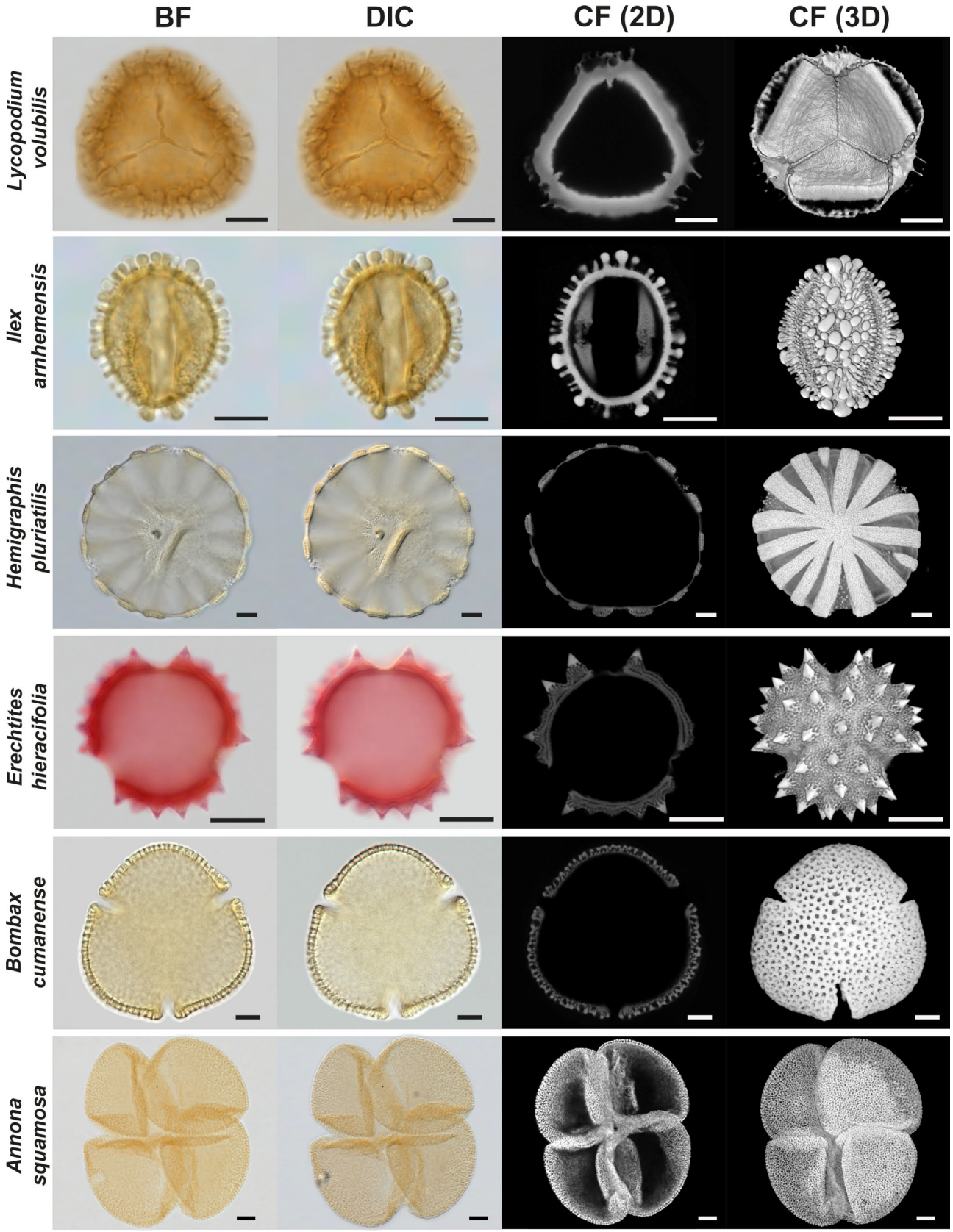


**FIGURE 1** Pollen and spore grains of six species are imaged using three microscope techniques: brightfield (BF), differential interference contrast (DIC), and confocal microscopy (CF). The confocal image includes one axial plane (2D) and a three-dimension convolution (3D). Scale bar is 10 μm.

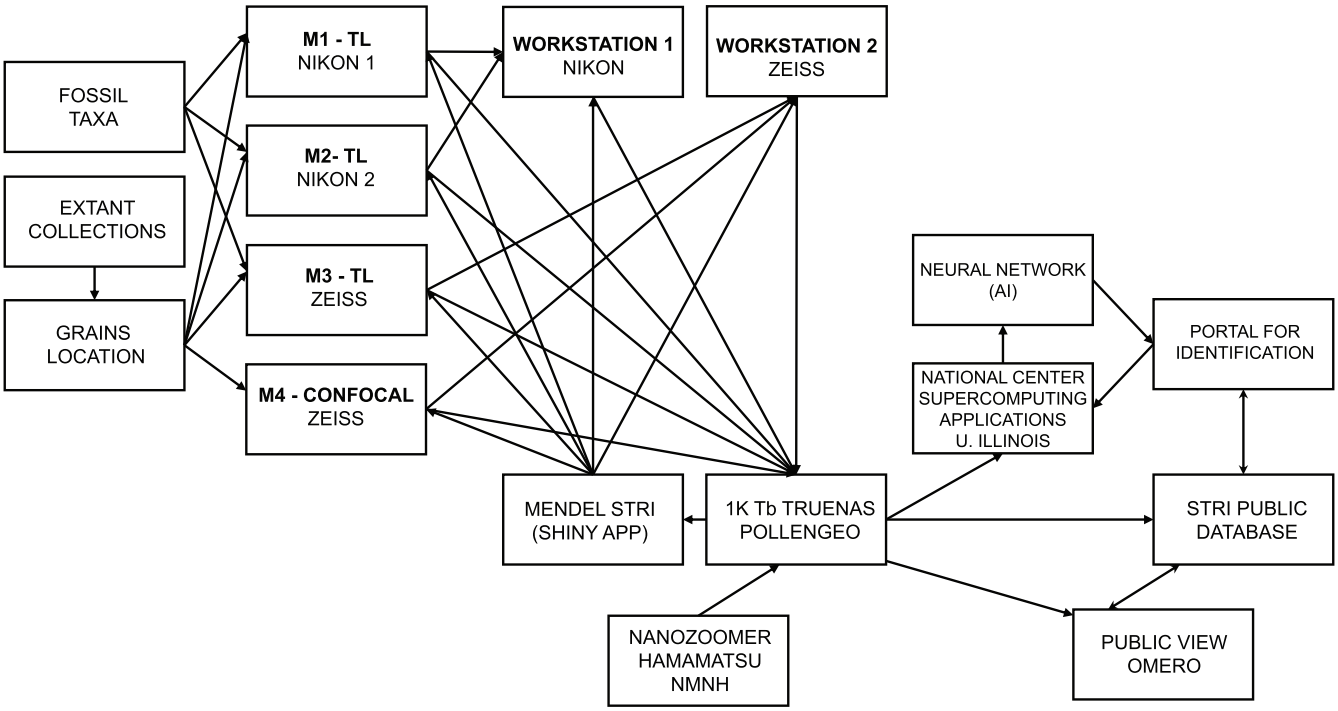


**FIGURE 2** "PollenGeo" database network illustrating all the components of the digitization process. This consists of four upright microscopes: a Zeiss confocal LSM 980, a Zeiss Axio Imager with a computer-controlled Piezo stage, and two Nikon Eclipse 80i with a computer-controlled Piezo stage. We scan full slides using a Hamamatsu Nanozoomer S20. Software used includes OMERO (hosted on a Smithsonian server), Nikon NIS-elements AR 5.42.04, Zeiss Zen 3.2 and 3.4, and a custom-built MySQL database on a shiny server. NMNH: Smithsonian Museum of Natural History, STRI: Smithsonian Tropical Research Institute.

confocal LSM 980, a Zeiss Axio Imager with a computer-controlled Piezo stage, and two Nikon Eclipse 80i with a computer-controlled-Piezo stage. Plan Apochromat with a numerical aperture (NA) of 1.4 and 60× to 63× objectives, and NA of 1.3 and 0.95 and 40× objectives. We also scan full slides using a Hamamatsu Nanozoomer S20 with 400× optical magnification and NA 0.95. Software used includes Open Microscopy Environment (OMERO) (hosted on a Smithsonian server), Nikon NIS-Elements AR 5.42.04, Zeiss Zen 3.2 and 3.4, and a custom-built MySQL database on a Shiny server (Figure 2).

## 2.1 | Step one: Collection curation

The acetolyzed Smithsonian palynological collection housed at the Smithsonian National Museum of Natural History (Dataset S1) contains 18,746 species, 5685 genera, and 548 families. It is composed of several collections (Moreno et al., 2014) including the Graham Palynological Collection, which was built by Alan Graham starting in 1954 with 15,202 species mostly from the Neotropics, the largest collection for this region (Stevens et al., 2014). This collection, donated by Dr. Graham to the Smithsonian in 2008, comprises 23,271 microscope slides corresponding to 15,202 taxa, and each one has a corresponding physical index card that contains data on the pollen that is on the slide (Figure 3). The second collection is the Joan Nowicke collection, with 5315 slides corresponding to 4141 taxa, mostly from temperate North America. Next is the Barro Colorado collection, which served as the basis for the widely used 1991 catalog, *Pollen and Spores from the Barro Colorado Island* (Roubik & Moreno, 1991), with 1305 taxa. The Amazon collection (Colinvaux et al., 1999), with 1038 taxa representing the most common taxa found in the Holocene palynological record of Amazonia. The last collection is the Sian Ka'an (México) collection, comprising 650 species from lowland tropical forest of southeastern Mexico (Palacios-Chavez & Ludlow-Wiechers, 1991). Sometimes the same species was in different collections. When that occurred, we chose the best slide (e.g., grains with the best preservation and large abundance) to digitize out of all collections.

Our first major challenge was digitizing the information on the Graham Collection's index cards. To that end, we used the Smithsonian Transcription Center, where an army of digital volunteers helped to transcribe the cards. We photographed each of the 23,271 index cards and uploaded them to the Transcription Center <https://transcription.si.edu> where ~100 volunteers helped to transcribe the cards, a process that took about 6 months. This effort allowed us to build a database that contained the metadata associated with each slide including the slide number, species, genus, tribe, family, author, herbarium, herbarium number, herbarium abbreviation, precise locality, province, country, collectors, state, territory, processing technique, processing data, collection date range, and a link to the index card image. A second step was updating the taxonomy, a task that required intensive work, given the constant evolution of botanical nomenclature as research

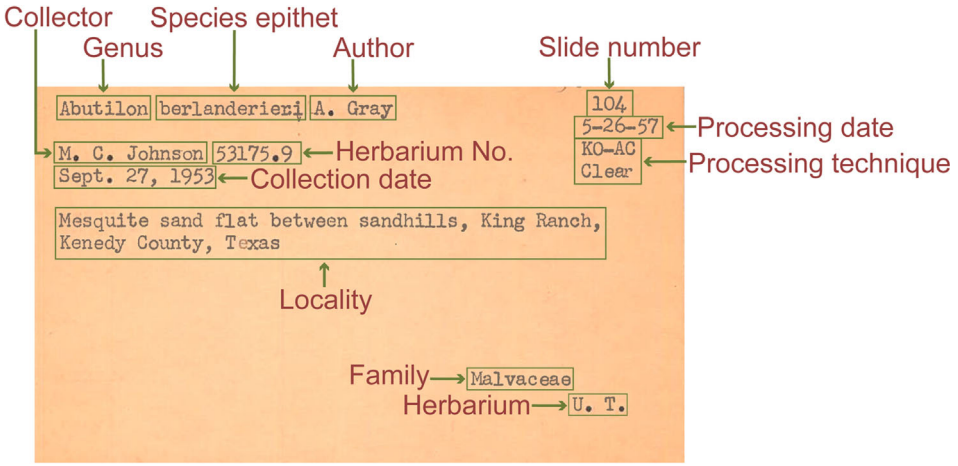


**FIGURE 3** Example index card from the Graham Palynological collection. The information in the cards was digitized by an army of volunteers at the Smithsonian transcription center following predefined categories as illustrated in this figure.

progresses. We compared the digitized names with the taxonomic lists of the World Flora Online (WFO, <www.worldfloraonline.org>), the taxonomic name resolution service (TNRS, <tnrs.biendata.org>), and the r-package WorldFlora 1.14-5 (Kindt, 2020). We selected the WFO opinion if discrepancies were found.

## 2.2 | Step two: Selecting what to digitize

We digitized 20 grains per taxon to capture the full variation in grain orientation (e.g., polar view vs. equatorial view) and grain size/morphology within each taxon. To do this, each slide was manually reviewed, and slide numbers were verified against the initial database derived from the Transcription Program using a query written in R. This step could not be automated due to the unpredictable condition of the labels, so manual inspection of each slide was essential. The quality of the slides also was variable, requiring microscopic examination to remove those slides not suitable for digitization. For example, the mounting medium used for permanent slide preparations may include Canada balsam, glycerin, glycerin jelly, silicone oil, and many types of industrial glues with differing densities. Each medium has unique properties, including light refractive indices and resistance to deterioration over the years (e.g., drying, lightening of grains, changes in shape and size of grains, formation of air bubbles, contamination, broken slides, and poor recovery), with almost 30% of slides already deteriorated and no longer useful for digitization (we established 13 criteria for discarding slides, see Notes S2). This is a common problem with old palynological collections (>30 years old) (Cushing, 2011), which makes digitizing them even more critical as existing pollen slides in herbaria and museums have a limited lifespan.

We selected the slide in the best condition when there were multiple slides of the same taxon, and only specimens identified at the species level were considered. We located 20 grains per slide using a Nikon Eclipse Ni BF microscope (Figure 2) and annotated the coordinates (X and Y values) of each of them into PollenGeo. These coordinates were then translated automatically by PollenGeo into the coordinates of any of the four microscopes used in the digitization.

## 2.3 | Step three: BF and DIC digitization

Once all the slides in a box (containing up to 100 slides each) were revised and grains were selected, the grains were imaged using one of the three light microscopes, either the two Nikon Eclipse 80i or the Zeiss Axioimager for the BF and DIC digitization.

PollenGeo provided the coordinates of all the grains that were previously chosen and the analyst proceeded to import the coordinates into the software that controls the microscope stage, either NIS in Nikon or Zen in Zeiss. Once 20 grains were located, the microscopist shifted to the objective with highest numerical aperture (NA 1.4, 63×, if the grain was larger than ~80 µm, we used 40× as the entire grain in 63× would be larger than the area captured by the camera), establishing the top and bottom axial plane of each grain, and choosing the microscopy technique (BF) or DIC. The software then calculated the number of Z-planes needed for the highest optical resolution (e.g., when using an objective with a numerical aperture of 1.4, Z-planes were spaced 0.2 µm apart). Once these parameters were established for all grains in the slide, imaging was then automated by the microscope software, a process that took 15 to 20 min. This process incorporated several software features to enhance the photographs, such as white balance and continuous autoexposure (see Notes S3 for a detailed step-by-step manual of each microscope). Once the imaging concluded, it was repeated using a second technique (either DIC or BF), to ensure that each grain is taken in both BF and DIC. BF and DIC produced different visualizations of pollen specimens, with DIC providing more morphological detail. However, BF is more widely used by the palynological community than DIC, which requires a more expensive setup. By including both imaging modalities, we produced a dataset that could be used by palynologists worldwide.

Fossil grains were imaged following the same protocol. The only difference was that the England Finder (that is a glass microscope slide with a uniquely indexed grid pattern that allows individual areas of interest to be referenced) coordinates of each grain were added to the file name. We also digitized 50 grains per species instead of 20, as

the variation of fossil taxa due to differences in preservation is often significant.

The microscope software, Zen or NIS, wrote the image stack (all Z-planes) of each grain into the hard disk of the computer that controls the microscope. The microscopist inputted the location of the photos into PollenGeo, and PollenGeo renamed them using a script that did not require the user's input. This file label contains the name of the taxon and the technique, and it also links the file to the PollenGeo database, which will then extract the metadata from the photograph itself. Lastly, the files are moved from the computer that runs the microscope to a 1000-Tb unit within the Smithsonian Network (TrueNas-PollenGeo). This was a critical step, as this unit has a large storage space and multiple automatic backups. We had two 8-h shifts per microscope per day, generating ~1000 files per day and ~2.5 Gb each. Therefore, it was essential for a project of this scale to have a large storage unit protected from accidental deletions.

As part of our protocol, the microscopist checked every single photograph taken from all the slides in a box before moving to a new one. If an error was detected, the image dataset was corrected right away. Errors occur due to two main reasons, either there was vibration in the floor when the photograph was taken or the stage shifted. We attempt to minimize these errors by placing the microscopes on an anti-vibration table and avoid walking, while the photographs are being taken. Small displacements (3–4 μm) were enough for a part of the grain to be out of the predefined Z-planes. We also added one or two grains to the imaging (photographing 21 or 22 grains to ensure that we had a minimum of 20 usable images).

We have also found that slides with a mounting medium that is too thick (>100 μm) were difficult to digitize due to the risk of accidentally fracturing the slides or scratching the objectives. We excluded these slides, as they involve considerably more time and risk. The same is true for preparations that do not have a homogeneous spread of material under the coverslip. Excess resin at the edges, for example, hinders the movement of the objectives, limiting the acquisition of images to only the center of the slide. Sometimes, liquid or viscous mounting media (e.g., glycerin) moved with the pressure of the oil immersion required by the 60× objective, and grains could not be digitized as grains were displaced while the picture was being taken. Another common problem during light microscopy digitization involves the light settings. There are many factors that can alter the light going through the slide, including bubbles, cracks, stains, type and thickness of mounting media, the age of the slide, excess organic matter, and degradation. These conditions can cause abrupt changes in light intensity in photographs even at different focal planes or cause photos to have blue or yellow colorations instead of white. Although it is possible to set a single light setting per technique and apply it to all selected grains, the microscopist must identify when individual adjustments are needed to correct these problems, even if the slide appears clean and homogeneous at first glance. This process may be automatized in the near future by using a properly trained neural network.

## 2.4 | Step four: Confocal superresolution

Zeiss Airyscan superresolution confocal microscopy (CF) has become an essential tool in palynological studies, offering high-resolution, three-dimensional optical light images of the internal and external structure of pollen grains and spores that rival electron microscopy and does not require additional sample preparation (Collevatti et al., 2024; Quamar et al., 2022; Romero, Kong, et al., 2020; Romero, Urban, & Punyasena, 2020).

Airyscan uses a 32-channel array detector that captures more light and improves the signal-to-noise ratio compared to traditional confocal systems that employ a single pinhole (Huff, 2015; Korobchevskaya et al., 2017; Sivaguru et al., 2018). With a resolution of up to 140 nm, Airyscan facilitates the observation of fine morphological details in palynomorphs (Huff, 2015; Romero, Urban, & Punyasena, 2020; Sivaguru et al., 2018). Furthermore, multiple detectors allow for improved image sharpness and clarity without the need to increase laser intensity or reduce image acquisition speed.

Once a slide box was digitized using light microscopy (DIC and BF), it was sent to the CF room. The microscopist uploaded the coordinates of all the grains that were taken in BF/DIC into the CF computer by querying PollenGeo and selected the five best to image. We imaged only five grains because the acquisition takes much longer than light microscopy, and we choose to image a large proportion of taxa rather than many grains from fewer taxa.

Laser tuning was a critical step in our digitization process. We employed the TRACK1-MPLX-CHA channel, which includes 405-, 488-, 561-, and 639-nm lasers. We used the 405-, 488-, and 561-nm lasers, applied individually or in combination depending on the specific characteristics of each grain. Laser selection and intensity, along with parameter settings such as Master Gain and Digital Offset, are based on the fluorescence response of each palynomorph and the mounting medium, whether glycerin gelatin, Canada balsam, or other resins. Each palynomorph has a unique fluorescence pattern that requires a customized configuration of the lasers, as several factors influence this pattern, including the thickness of the exine and the presence of stains such as safranin. Thicker exines, for example, may require higher laser intensity to adequately penetrate and reveal internal details without causing oversaturation. The 561-nm laser is particularly effective at intensifying exine fluorescence, but its improper use can lead to oversaturation that obscures fine morphological details. In contrast, the 405-nm laser tends to excite the mounting medium, which requires careful adjustment to prevent it from overshadowing the palynomorph. These adjustments are vital to obtain clear images that highlight the exine structure. In addition, the type of mounting medium can alter how the sample scatters or absorbs light, thus impacting fluorescence quality. This variability makes meticulous configuration of lasers essential to capture the key morphological features of each species. A common complication is the loss of information caused by pixel oversaturation or underexposure. Oversaturation can obscure fine pollen details, while underexposure can result in areas without useful data. To mitigate these problems, we used the Range

Indicator, a tool that identifies in real-time oversaturated (shown in red) and underexposed (shown in blue) areas.

In the image acquisition process, it is essential to correctly adjust sampling and averaging, two key parameters that are specifically tailored to the physical characteristics of the palynomorphs, such as size and volume. Sampling determines the type of data acquisition, whereas averaging significantly improves image quality by increasing the signal-to-noise ratio, albeit at the cost of longer acquisition time. There are three sampling modes in our system: SR-4Y (Super Resolution 4Y), SR-8Y (Super Resolution 8Y), and CO-8Y (Confocal 8Y). SR-4Y is the default mode, used for medium-sized grains (between 20 and 50 µm), as it offers the best resolution. For larger samples or those with significant depths, such as grains from the Malvaceae family, we opted for faster acquisition modes such as SR-8Y or CO-8Y, which reduce acquisition time in exchange for a slight decrease in resolution, but without compromising image quality. Averaging is available in various configurations: none, 2×, 4×, 8× and 16×. For smaller grains or those exhibiting extremely fine morphological details, we used a higher averaging level, such as 8× or 16×, to maximize resolution and ensure that all details were captured. In contrast, for larger grains, we used lower averaging, such as 2×, or combine 4× with SR-8Y sampling. This combination of sampling and averaging ensured that we obtained high-quality images with the appropriate resolution for each type of palynomorph.

After setting up the imaging, including crop area adjustment, lasers, sampling, averaging, and start and end planes for Z-stacks, we started the acquisition process (see Notes S4 for a detailed step-by-step manual). Images were captured sequentially from the foreground to the background, according to the previously defined settings. The acquisition time can vary considerably, from 1 to 60 min, depending on the species and the specific settings of each sample. Once the acquisition was completed, we processed the images using the Airyscan module. This step allowed us to integrate all stacks into a single file in .czi format. This process ensures that each palynomorph is captured integrally, without omitting segments, and ensures that it adequately occupies the area defined for its acquisition. Once all parameters have been verified, we selected the "Best Fit" option to save the final image. This step is crucial to ensure maximum visual quality of the morphological details.

Once the five imaging for each species were completed, we proceeded to rename and store the images using PollenGeo following the same protocol for DIC/BF as described above. We generated about 90 Gb of image data per day.

## 2.5 | Step five: Cropping

Microscope images were taken using the entire field of view offered by the camera chip (also called scene) as multiple grains could often be observed within the same scene (Figure 4), and several grains could be captured at once. The final step in digitization consisted of cropping out individual grains from each scene (Figure 4) to reduce the file size and exclude particles that are not pollen grains. This process was done manually using a fully dedicated workstation that was not connected to a microscope (Figure 2). This step also served as the last line of quality control before the images go into the digital repository. The individual responsible for image cropping opened the BF, DIC, and CF images taken as described in the steps above and checked if the image is of a good quality and the taxon name in the file corresponded to the taxon in the PollenGeo database. If the image failed either of those two parameters, the issue was noted in the PollenGeo database, and the microscopist who took the photo received a message indicating the photo needed to be retaken. If the photograph passes this assessment, every grain was manually cropped into a separate file. In addition to cropping pollen grains on the X–Y-axes, noninformative Z-planes (planes without in-focus sections of the grain) were removed (Figure 4). The photo quality control also included verifying several items, including resolution, the absence of digital noise, or oversaturation, taxonomy (erroneous identifications due to contamination), that the image was not a duplicate, and ensuring that entire grain was imaged (see Notes S5 for a step-by-step manual of the cropping process).

We used the NIS for the Nikon images (.nd2), and Zen for the Zeiss images (.czi). Once the individual grains were extracted, the original photograph was discarded. The files containing the individual grains were renamed by PollenGeo, including the slide code, taxa name, the original photo number, the technique, the acronym "SG" that designates it as a single grain, and the appropriated file extension (.nd2 or .czi). PollenGeo additionally collected the metadata from the original photo and the individual grains.

## 2.6 | Step six: User interaction

Users will be able to interact with the digitized collection in several ways once the work is completed. First, we will produce small-file. gif images of each photograph such that the user can see the animated grain across all its Z-planes. These images will avoid the need to upload/download the full-resolution image, thus facilitating user interaction. Nevertheless, we will always offer the option to download the full-resolution image in its native format. Second, we will upload the images from both extant and fossil taxa to our morphological electronic database (Jaramillo & Rueda, 2024) at <https://biogeodb.stri.si.edu/jaramillosdb/web/morphological/>. This is a public database, custom-built using MySQL 8.2 and PHP 8 that contains the morphological description of thousands of fossil taxa and most of the taxa in the extant collection (Table 1). Users are able to search for a taxon or a trait or combination of traits (e.g., aperture and ornamentation). We hope to offer in the near future the possibility of a fluid interaction with R for the user to perform queries and retrieve data. Lastly, by mid-2026, we will create a simple website where the user can access OMERO, search for a taxon, visualize the grains as if using a microscope, and see the images' associated metadata (e.g., where pollen/spore comes from, slide, position of the grain in the slide, technique, resolution, number of Z-planes, and file size).

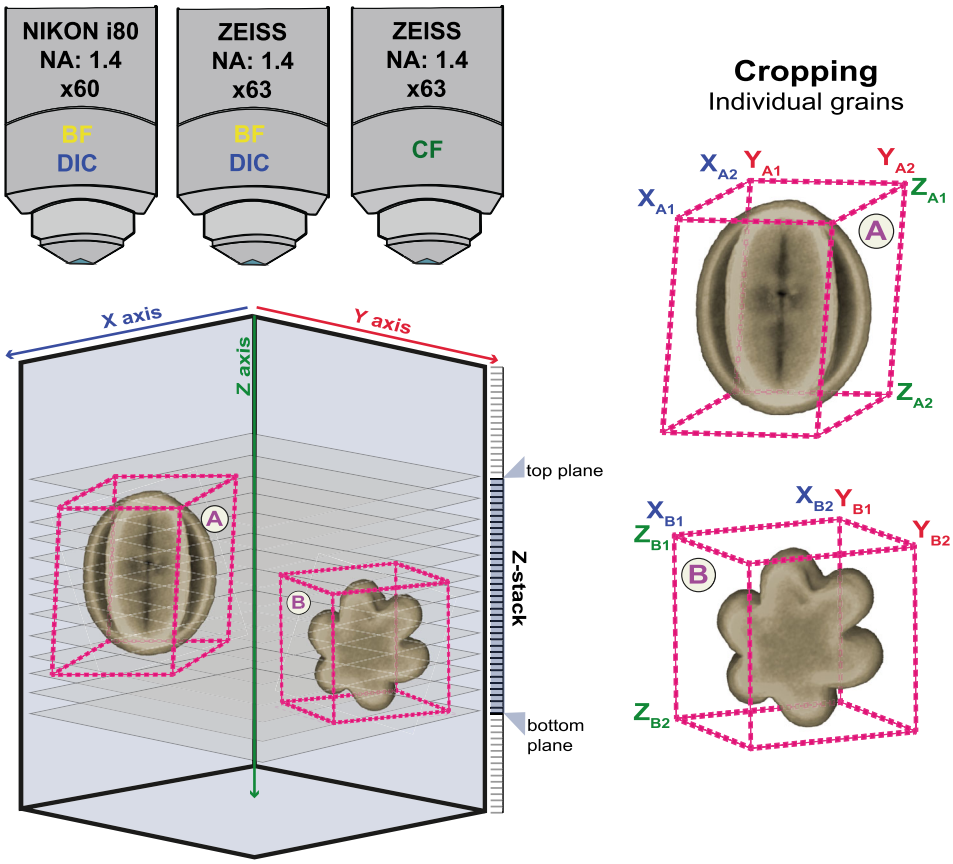


**FIGURE 4** Nikon and Zeiss objectives were used to digitize the grains in brightfield (BF), differential interference contrast (DIC), and confocal (CF) microscopy. Grains are digitized in multiple focal Z-planes and often there is more than one grain in a single scene. Once digitized, the grains are cropped in both x-y plane and the z-planes and saved as individual files. (A) Grain in equatorial view. (B) Grain in polar view. NA, numerical aperture.

## 3 | DIGITIZING ADDITIONAL COLLECTIONS (EXTANT AND FOSSIL)

The Smithsonian palynological collection is one of several large collections across the world, each one with different geographic coverage. For example, the ISEM collection at the Université de Montpellier contains ~30,000 taxa; the Royal Botanic Gardens, Kew, contains ~16,000 taxa; the Muséum National d'Histoire Naturelle in Paris has 3500 extant taxa, along with thousands of single-grain fossil specimens of the Boltenhagen Collection; the collection at the Swedish Museum of Natural History holds >10,000 extant taxa; and the fossil palynological collection of the Norwegian Offshore Directorate has >100,000 slides. These are just a few of the world's largest collections. There are also hundreds of smaller palynological collections in institutions across the globe that could be digitized.

Many institutions are actively engaged in digitization efforts, including the University of Montpellier, MNHN Paris, Amgueddfa Cymru—National Museum of Wales, the Natural History Museum in London, Cardiff University, and the University of York. While the initial cost in equipment and person effort can be high, the result is a permanent digital archive that is can outlast the estimated 30-year lifespan of the original palynological slide (Cushing, 2011) and is more accessible to the global community. Imaging and digitization of palynological collections support the research and education missions of the many museums and institutes in which these collections reside. These larger institutes can also serve as imaging hubs once their own collections are digitized and lead the effort in digitizing smaller collections housed in university research labs and regional museums.

We anticipate that an aggregator could soon be developed to allow a single taxon query to search all palynological digitization products simultaneously. Aggregating dispersed collections of pollen and spore images would allow researchers to use these images in new and novel ways, including training deep-learning models for region-specific analyses or universal foundation models, as described below.

## 4 | SECTION 2: APPLICATIONS

Digital libraries of pollen images provide the critical training data needed to develop and fine-tune computer-vision models for interpreting pollen morphology. This data-centered need motivated our large-scale digitization effort. We envision that our digital images will be used in a wide variety of applications, most of which still need to be developed and even imagined, including automated taxa

identification and counting, novel morphometrics, and large-scale analysis addressing multiple ecological and evolutionary questions. Our own research in developing machine-learning models in both low diversity (Johnsrud et al., 2013; Punyasena et al., 2012; Romero, Urban, & Punyasena, 2020) and higher diversity tropical systems (Punyasena et al., 2022) demonstrates how deep learning can be used to not only automate pollen identification but also improve upon expert classifications and ecological and evolutionary interpretations (Adaimé et al., 2023, 2024, 2025; Punyasena et al., 2012; Romero, Kong, et al., 2020).

Recent studies have used DNA metabarcoding to identify pollen by comparing the DNA of a sample against a reference library of DNA sequences of genetic markers (e.g., Evans & Kitson, 2020 {Lowe, 2022 #3981}). This technique is an excellent complement to the palynological analysis. However, DNA sequencing cannot be used if DNA has degraded as happens with most fossils with a few exceptions (Kjær et al., 2022) and many environmental samples. Furthermore, palynological analyses do not require expensive equipment, a simple microscope and the elements in a kitchen are enough (Traverse, 2007).

## 4.1 | Neural networks to aid identifications

Since the late 1960s, the palynological community has been interested in automating pollen-grain recognition (Flenley, 1968; France et al., 2000). A decade-old review anticipated that computer power and digital microscopy had reached a point where some degree of automation could finally be effectively applied to palynological studies (Holt & Bennett, 2014). However, pollen identification remains a highly specialized, laborious, and primarily visual skill, despite extensive published work on automated identification methods. Over the past decade, there has been a growing interest in attempts to automate pollen-grain recognition, mostly in the aerobiological medical fields (allergies) and honey quality control, including teams in Algeria (Menad et al., 2019); Australia (Lagerstrom et al., 2015); France (Bourel et al., 2020; Chung & Rodriguez, 2017); Brazil (Astolfi et al., 2020; da Silva et al., 2014; de Geus et al., 2019; Gonçalves et al., 2016; Rodrigues et al., 2015); China (Wang et al., 2021); Bulgaria (Nikolov & Tsankova, 2018); Czech Republic (Pospiech et al., 2021); Germany (Oteros et al., 2015; Ronneberger et al., 2002; Viertel & König, 2021; Viertel & König, 2022); Guadeloupe (Barnacin et al., 2023); Italy (Battiato et al., 2020, 2021); Japan (Miki & Kawashima, 2021); New Zealand (Allen et al., 2008; Hodgson et al., 2005; Holt et al., 2011; Lagerstrom et al., 2015; Li et al., 2004; Li & Flenley, 1999); Russia (Khanzhina et al., 2018; Khanzhina et al., 2023); Spain (Boldeanu et al., 2022; Bonton et al., 2002; De Saotero et al., 2004; Díaz-López et al., 2015; Gallardo-Caballero et al., 2019; García et al., 2012; Marcos et al., 2015; Pozo-Baños et al., 2012; Redondo et al., 2015; Sevillano et al., 2020; Ticay-Rivas et al., 2011; Valiente et al., 2023); Sweden (Midtvedt et al., 2021); the Netherlands (Polling et al., 2021); Turkey (Kaya et al., 2014); the United States (A. I. Daood et al., 2016b; A. Daood et al., 2016a; Filipovych et al., 2016; Nguyen et al., 2013; Ranzato et al., 2007); and Uzbekistan (Jumanov & Safarov, 2022).

*How will our effort make a difference?* The majority of this research is driven by experts in machine learning rather than by practicing palynologists. Machine methods have still not been adopted by the larger palynological community because of the mismatch between the expertise of many pollen experts and the skills needed for computer-vision analysis and the challenge of producing the large image dataset needed to effectively train machine models (Table 2).

Large datasets are a critical element for the development of automated pollen identification analyses. We have already digitized ~470,000 individual grains from 11,000 species and our goal by the end of 2025 is to end with ~700,000 grains from 18,000 species. An effort of this magnitude is unparalleled anywhere in the world (Table 1). We are confident that once our image dataset is complete and available to the public, other research teams will be able to use our images to train deep-learning models. We also hope that other researchers will be able to contribute similar images for additional species using the infrastructure we have already built.

Deep-learning models are typically constrained by the image data upon which they are trained. Building a larger shared image dataset will allow a global community of researchers to experiment with developing clade-specific or region-specific models tailored to their research questions, or, alternatively, use a growing pool of palynological images to develop foundation models (sensu Bommasani et al., 2022) that can be used universally in the study of all pollen and spores.

To help the broader palynological community take advantage of this digital image resource, we are also developing a web-accessible palynology image analysis platform for hosting published computer-vision tools for automated pollen identification, including our own work developing deep-learning models for phylogenetically meaningful characterizations of pollen morphology and high-throughput automated pollen analyses (e.g., Adaimé et al., 2023, 2024, 2025; Punyasena et al., 2012, 2022; Romero, Kong, et al., 2020). We anticipate that it will serve as a user-friendly gateway to machine-learning analysis for the diverse communities that employ pollen data and is accessible to researchers without experience in programming or machine learning. It will be the first of its kind for palynology, serving as a centralized system for sharing published machine-learning toolboxes and providing an incentive for the community to share pollen images and expert knowledge in pollen identification.

## 4.2 | Recording routine palynological counts

The process of counting pollen and spores under the microscope is performed daily by palynologists worldwide. This time-consuming process requires years of training and yet it yields counts that are not fully replicable, as recording the coordinates of and photographing each of the 300–400 grains counted per slide would add tens of additional hours to the process. However, there is a solution: directly

**TABLE 2** Training sets employed over the past two decades in algorithms for pollen identification.

| Number of species | Number of individual grains | Reference |
|---|---|---|
| 4 | 100 | (Hodgson et al., 2005) |
| 4 | 72 | (Li & Flenley, 1999) |
| 5 | 100 | (Filipovych et al., 2016) |
| 5 | 6472 | (Polling et al., 2021) |
| 6 | 300 | (Holt et al., 2011) |
| 6 | 93 | (Nikolov & Tsankova, 2018) |
| 7 | 2482 | (Pospiech et al., 2021) |
| 7 | 500 | (Oteros et al., 2015) |
| 7 | 210 | (da Silva et al., 2014) |
| 7 | 630 | (Duller et al., 1999) |
| 8 | 1102 | (Bourel et al., 2020) |
| 9 | 768 | (Nguyen et al., 2013) |
| 10 | 392 | (Daood et al., 2016b) |
| 10 | 300 | (Kaya et al., 2014) |
| 11 | 4235 | (Gallardo-Caballero et al., 2019) |
| 11 | 1774 | (Khanzhina et al., 2018) |
| 12 | 4061 | (Díaz-López et al., 2015) |
| 13 | 234 | (Li et al., 2004) |
| 15 | 1800 | (Marcos et al., 2015) |
| 15 | 1800 | (Redondo et al., 2015) |
| 15 | 2890 | (Lagerstrom et al., 2015) |
| 15 | 1800 | (Marcos et al., 2015) |
| 17 | 426 | (García et al., 2012) |
| 17 | 345 | (Pozo-Baños et al., 2012) |
| 17 | 426 | (Ticay-Rivas et al., 2011) |
| 19 | 19,000 | (Valiente et al., 2023) |
| 19 | 38,000 | (Boldeanu et al., 2022) |
| 20 | 7745 | (Khanzhina et al., 2023) |
| 23 | 805 | (Menad et al., 2019) |
| 23 | 575 | (Viertel & König, 2021) |
| 23 | 805 | (Gonçalves et al., 2016) |
| 26 | 1300 | (Ronneberger et al., 2002) |
| 27 | 3686 | (Ranzato et al., 2007) |
| 30 | 1060 | (Daood et al., 2016a) |
| 30 | 350 | (Bonton et al., 2002) |
| 33 | 22,759 | (Wang et al., 2021) |
| 46 | 19,000 | (Sevillano et al., 2020) |
| 73 | 2523 | (Astolfi et al., 2020) |
| 73 | 13,000 | (Battiato et al., 2020; Battiato et al., 2021) |
| 80 | 1505 | (Barnacin et al., 2023) |
| 134 | 3752 | (de Geus et al., 2019) |

recording counts and identifications on digitized slides. Slide-scanning microscopes like the Hamamatsu NanoZoomer, Zeiss Axioscan, and many others have been developed in the last decade for pathology research and are capable of scanning entire slides with multiple focal planes. Counts can then be produced on a screen rather than behind a microscope. We use a NanoZoomer N20 to scan our palynological slides that consist of two 23 × 23 mm coverslips mounted on standard 75 × 25 mm microscope slides. Twenty-one focal planes of the material in each coverslip produce an image that is ~22 Gb in size. The image is then uploaded into OMERO (Allan et al., 2012), a client–server environment for managing, viewing, annotating, and analyzing imaging data that is stored on a central server with a database for the metadata, and that can be accessed via a web-browser anywhere in the world.

Digitally recording counts not only preserves the metadata for each individual expert identification, it also produces training data that can be used to train future neural networks capable of further streamlining the pollen counting process (Punyasena et al., 2022). It also allows the interaction of multiple analysts across the world working on the same image, as all are accessing the image via OMERO, thus transforming the practice of palynology from a solitary endeavor to a communitarian science.

## 4.3 | Novel morphometrics

Morphological descriptions of pollen and spores have been carried out almost since the start of the discipline (Traverse, 2007), with a heavy emphasis on size (length and width of morphological elements) and categorical terms (e.g., type of sculpture, type of ornamentation, and type of shapes). A vast number of studies have been published on pollen morphology. A Google Scholar query for "pollen" "morphology" returns 979,000 results and even a more specialized query "pollen" "morphometrics" returns 6530 results (queries performed March 4, 2025). There have been notable synoptic studies of pollen morphology across angiosperms as a whole (e.g., Erdtman, 1952; Wortley et al., 2015), studies of angiosperm pollen macroevolution using discrete characters to quantify morphology (e.g., Lupia et al., 1999), and studies of pollen morphological evolution in large clades such as the Asterales in a phylogenetic context (Jardine et al., 2022). There have also been focused studies of single genera. For example, de Souza et al. (2020) studied the pollen morphology of 102 species in the genus *Croton* and used 10 morphological categorical traits (e.g., type of sexine elements in the lumen of rosettes 0 = absent, 1 = clava, 2 = pilum,3 = baculum, 4 = granulum, and 5 = gemma). This is an example of a classical and thorough pollen morphological study. However, in common with other studies that use categorical traits (also known as discrete characters) to describe morphology, there could be considerable uncaptured morphological variation within each category (Mander et al., 2021). An instance where such limitations occur is a

study of pollen morphology on Barro Colorado Island (BCI), Panama, in which pollen grains were assigned one of 22 different classes of primary surface ornamentation but subtle variations within each of these classes were not captured (Mander et al., 2021). In the case of Asteraceae pollen on BCI, for example, the pollen surface is covered by spines, but variation in the density of these spines between taxa was not captured (Mander et al., 2021).

However, we think that pollen morphology offers more information than can be described using terminology and encoded in systems of discrete characters, and that could be used to enhance the morphological descriptions. In particular, the use of digital image analysis will augment significantly the number of meaningful characters that can be used in analyses of pollen morphology. For example, we studied the pollen morphology of the palm Mauritiinae using light and confocal microscopy and developed several algorithms that generate quantitative data. This enabled us to measure a set of parameters not easily obtained by hand or using other microscopy methods and showed the potential to automate analysis (Collevatti et al., 2024). We measured nine quantitative parameters and one qualitative parameter, including volume, aperture, skewness, and spine shape and density. We then used a CART and Random Forest analysis that correctly identified both fossil and extant grains. This is only an example, as we are at the early stages of a revolution in the analysis of palynological morphospace, where several paths are being developed (e.g., Mander et al., 2013; Punyasena et al., 2012; Quamar et al., 2022; Romero, Kong, et al., 2020; Romero, Urban, & Punyasena, 2020; Trivedi et al., 2022), and there is still room for creating new ways to analyze pollen morphology. For example, creating metrics for measuring surface ornamentation such as the geometry of reticula, or the outline of projecting elements, or the spatial distribution of sculptural elements. We anticipate that this, coupled with neural networks as described above, will facilitate the study of variation in pollen morphology over evolutionary time scales. In particular, certain fossil groups, such as the marine microfossils (e.g., radiolaria, diatoms, and foraminifera), have a sufficiently complete fossil record that the types of speciation can be reconstructed from morphometric data, with the gradual speciation of the diatom *Rhizosolenia* a well-known example (Benton & Pearson, 2001). However, while the fossil record of pollen and spore rivals that of many such groups in terms of abundance and completeness, the nature of our current morphometric approaches means that is very difficult to observe transitions between morphotypes generated by systems of discrete morphological characters rooted in the language of descriptive palynology (Punt et al., 2007). There are some notable morphometric studies of single genera through time, such as a study of *Echitriporites* pollen across the Cretaceous–Paleogene (K-Pg) boundary in northern South America, which showed that only one species *E. trianguliformis* persisted after the K-Pg boundary in this region (Cárdenas et al., 2019), but developing metrics to quantify fine scale-variation in surface ornamentation over long timescales could allow modes of speciation among plants to be reconstructed in the manner shown for marine fossil groups (Benton & Pearson, 2001).

Deep learning also provides new approaches for quantifying and analyzing morphological characters. Notably, features identified by neural networks can be used directly in phylogenetic placement (Adaimé et al., 2023) or in ecological inference (Adaimé et al., 2024, 2025).

## 4.4 | Big data analysis

The morphological diversity of pollen grains is extremely high, especially in tropical rainforests (Roubik & Moreno, 1991). However, we still do not know whether this variation reflects ancestry, or ecological function, or is purely a result of biogeochemical drift. Pollen ornamentation forms by primexine phase separation coupled to membrane undulations, which reach equilibrium in some taxa but in most, it forms in kinetic arrest (Radja et al., 2019), and it is still uncertain what forces drive the kinetic arrest. Morphology is very conservative in some families (e.g., Poaceae) whereas it is highly variable in others (e.g., Acanthaceae) (Mander & Punyasena, 2014). The relation of function and morphology is also unclear (Mander et al., 2021), and perhaps pollen morphology is an all-general-purpose design not specifically fitted for a single task (Crane, 1986). A recent analysis comparing the morphological pollen distance of 700 taxa in BCI with its corresponding phylogenetic distance found a strong relationship at the species and genus level, but this relationship breaks down at higher levels (Mander et al., 2021). There was no relationship between morphology and biotic and abiotic pollination, and some pollen morphotypes or pollen traits were unique to certain pollination ecologies. These results, however, represent a small proportion of tropical diversity. We envision a similar set of analyses but on the large-scale dataset that our digitization program is producing. A big data approach would also facilitate the large-scale comparison of different biogeographical regions. For example, Jaramillo and Dilcher (2001) showed an 11% and 11.5% similarity in the taxonomic composition of vegetation in northern South America and tropical Africa during the Eocene and Paleocene, respectively, highlighting that biogeographical differences in the rainforests of the Neotropics and Old World tropics are geologically ancient. Recently, the Paleocene–Eocene pollen and spores systematics from tropical West Africa have been overhauled, with the description of two new spore species and one new genus and 18 new species of angiosperm pollen (Mander et al., 2023). The updated systematics of Mander et al. (2023) were harmonized with that of northern South America using type material inspected in museum collections and an online morphological database (Jaramillo & Rueda, 2024) to overcome barriers in taxonomic practice between the two biogeographic regions. However, while this will permit a detailed comparison of tropical vegetation in northern South America and West Africa during the Paleocene and Eocene that builds on the analysis of Jaramillo and Dilcher (Jaramillo & Dilcher, 2001), it will cover just two regions and at most two timepoints. Synthesizing a pantropical view of rainforest evolution using the empirical data of the fossil record, and spanning the birth of modern-aspect rainforests in the aftermath of the K-Pg mass extinction (Carvalho et al., 2021) to the present day, demands a big data approach

that leverages modern and fossil pollen and spore collections such as the Smithsonian Graham-Nowicke palynological collection together with neural networks and automation.

## 5 | CONCLUSIONS

We are at the threshold of an exciting new era in palynological research that will unlock the full power of palynological analysis, aided by a digital revolution in microscopy, the rapid development of neural network algorithms that aid pollen and spore identification, and image analysis that can add novel ways to quantify morphology. Digitizing massive collections like the Smithsonian Graham-Nowicke palynological collection plays a pivotal role in this revolution, transforming the practice of the entire field of palynology, with ripple effects across various fields, including paleontology, entomology, medicine, forensics, archeology, and paleoclimate.

### AUTHOR CONTRIBUTIONS

Carlos Jaramillo created and directed the whole project. Angelica Arcila, Roxana Alveo, Jorge Bermudez, J. Bustos, David Caro, Francy Carvajal, Mauricio León-Carreño, Shara Chaves, Karen Cardenas, Andres Diaz-Jaramillo, Laura Diaz, Luisa Gomez, Maria Alejandra Lopez., Paula Lopera, Priscila Lopez, Jhonatan Martínez Murcia, Enrique Moreno, Enrique Neyra, Brenda Orosco, Natalia Ossa, Natalia Ovalle, Carolina Ovalle, Angelo Plata, Bruno Scudeiro, Axel Tejada-Fajardo, Vinicius Do Valle, and Thiago Wood performed the digitization. Jhon Ortiz constructed PollenGeo. Daurys de Alba, David Caro, and Carlos Moreno wrote code supporting several aspects of the digitization. Silane A. F. da Silva Caminha and Carlos D'Apolito digitized fossil taxa. Ingrid Romero digitized whole slides and developed paths to analyze confocal images. Surangi W. Punyasena constructed the artificial intelligence environment and helped to design the general structure of the project. Luke Mander created multiple venues to analyzed palynomorphs. Carlos Jaramillo wrote the manuscript with the collaboration of all coauthors.

### ACKNOWLEDGMENTS

Funding provided by the Smithsonian Institution, the Anders Foundation, Gregory D. and Jennifer Walston Johnson, the 1923 Fund, the Rubinoff Big Bet Endowment, and the Smithsonian Women's Committee, and the Smithsonian Life on a Sustainable Planet Pathfinder. Thanks to Natasha Atkins and Camilo Jaramillo for editing. Thanks to *Plants, People, Planet* Editor Rhian Smith and three anonymous reviewers for reviewing the manuscript and offering suggestions that highly improved it.

### CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

### DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the Supporting information of this article.

### ORCID

*Carlos Jaramillo* https://orcid.org/0000-0002-2616-5079
*Surangi W. Punyasena* https://orcid.org/0000-0003-2110-5840
*Angelica Arcila* https://orcid.org/0009-0009-8756-5808
*Jorge Bermudez* https://orcid.org/0009-0006-6604-9503
*Jonatan Bustos* https://orcid.org/0000-0002-7998-4342
*Dayenari Caballero-Rodriguez* https://orcid.org/0000-0001-5537-9116
*Karen Cardenas* https://orcid.org/0000-0001-6446-0119
*David Caro* https://orcid.org/0009-0001-8244-3256
*Francy Carvajal* https://orcid.org/0000-0002-2836-993X
*Ivonne Marcela Castañeda* https://orcid.org/0000-0002-1353-6862
*Shara Chaves* https://orcid.org/0000-0002-4552-6140
*Carlos D'Apolito* https://orcid.org/0000-0003-1602-0201
*Andres Diaz-Jaramillo* https://orcid.org/0009-0001-8070-8089
*Luisa Gomez* https://orcid.org/0009-0007-3912-2063
*Mauricio León-Carreño* https://orcid.org/0000-0001-8657-5902
*Maria Alejandra Lopez* https://orcid.org/0009-0009-4498-7154
*Luke Mander* https://orcid.org/0000-0003-4347-2705
*Jhonatan Martínez Murcia* https://orcid.org/0000-0002-8909-2655
*Enrique Moreno* https://orcid.org/0000-0002-9767-7522
*Enrique Neyra* https://orcid.org/0000-0001-7570-7121
*Brenda Orosco* https://orcid.org/0009-0007-5449-1510
*Natalia Ossa* https://orcid.org/0000-0002-7785-7585
*Carolina Ovalle* https://orcid.org/0000-0003-0577-5019
*Angelo Plata* https://orcid.org/0000-0001-8509-0567
*Ingrid Romero* https://orcid.org/0000-0003-1376-5830
*Bruno Scudeiro* https://orcid.org/0000-0002-9494-6408
*Silane A. F. da Silva Caminha* https://orcid.org/0000-0003-4853-2789
*Axel Tejada-Fajardo* https://orcid.org/0000-0003-1098-0367
*Vinicius Do Valle* https://orcid.org/0009-0007-7723-7494
*Thiago Wood* https://orcid.org/0009-0001-2742-8774

### REFERENCES

Adaimé, M.-E., Kong, S., & Punyasena, S. W. (2023). Deep learning approaches to the phylogenetic placement of extinct pollen morphotypes. *PNAS Nexus*, *3*(1), pgad419. https://doi.org/10.1093/pnasnexus/pgad419

Adaimé, M.-E., Kong, S., Urban, M. A., Alayne, F., Street-Perrott, D., Verschuren, I., & Punyasena, S. W. (2024). Reconstructing the diversity dynamics of paleo-grasslands using deep learning on superresolution images of fossil *Poaceae* pollen. *bioRxiv*, 2024.09.23.612957. https://doi.org/10.1101/2024.09.23.612957

Adaimé, M.-E., Urban, M. A., Kong, S., Jaramillo, C., & Punyasena, S. W. (2025). Pollen morphology, deep learning, phylogenetics, and the evolution of environmental adaptations in *Podocarpus*. *New Phytologist*, *247*, 1460–1473. https://doi.org/10.1111/nph.70250

Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., MacDonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., … Swedlow, J. R. (2012). OMERO: Flexible, model-driven data management for experimental biology. *Nature Methods*, *9*(3), 245–253. https://doi.org/10.1038/nmeth.1896

Allen, G. P., Hodgson, R. M., Marsland, S. R., Flenley, J. R. (2008). Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic objects. *15th International conference on mechatronics and machine vision in practice (M2VIP08)*. Auckland, New Zealanda. 221-226. https://doi.org/10.1109/MMVIP.2008.4749537

Astolfi, G., Gonçalves, A. B., Menezes, G. V., Borges, F. S. B., Astolfi, A. C. M. N., Matsubara, E. T., Alvarez, M., & Pistori, H. (2020). POLLEN73S: An image dataset for pollen grains classification. *Ecological Informatics*, *60*, 101165. https://doi.org/10.1016/j.ecoinf.2020.101165

Barnacin, E., Henry, J.-L., Nagau, J., & Jack, M. (2023). Comparison of machine learning and deep learning algorithms for automatic classification of 80 different pollen species. *International Conference on Computational Vision*. Jun 2022, Venice, Italy. https://hal.science/hal-04114411v1

Battiato, S., Guarnera, F., Ortis, A., Trenta, F., Ascari, L., Siniscalco, C., De Gregorio, T., & Suarez, E. (2021). Pollen Grain Classification Challenge 2020 Challenge Report. In *ICPR 2020 workshops, LNCS 12668* (vol. 12668, pp. 469–479). https://doi.org/10.1007/978-3-030-68793-9_34

Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., & Siniscalco, C. (2020). Detection and classification of pollen grain microscope images. In *2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 4220–4227). https://doi.org/10.1109/CVPRW50498.2020.00498

Benton, M. J., & Pearson, P. N. (2001). Speciation in the fossil record. *TRENDS in Ecology and Evolution*, *16*, 405–411. https://doi.org/10.1016/S0169-5347(01)02149-8

Boldeanu, M., González-Alonso, M., Cucu, H., Burileanu, C., Maya-Manzano, J. M., & Buters, J. T. M. (2022). Automatic pollen classification and segmentation using U-nets and synthetic data. *IEEE Access*, *10*, 73675–73684. https://doi.org/10.1109/ACCESS.2022.3189012

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2022). On the opportunities and risks of foundation models. *arXiv*, 2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Bonton, P., Boucher, A., Thonnat, M., Tomczak, R., Hidaldo, P. J., Belmonte, J., & Galan, C. (2002). Colour image in 2D and 3D microscopy for the automation of pollen rate measurement. *Image Analysis & Stereology*, *21*, 25–30.

Bourel, B., Marchant, R., de Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., & Beaufort, L. (2020). Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. *Computers & Geosciences*, *140*, 104498. https://doi.org/10.1016/j.cageo.2020.104498

Cárdenas, D., De la Parra, F., & Espinosa-Campuzano, C. (2019). Morphologic variation of two key biostratigraphical proteaceous-like pollen taxa across the Cretaceous–Paleogene boundary in northern South America. *Grana*, *58*, 276–291.

Carvalho, M. R., Jaramillo, C., de la Parra, F., Caballero-Rodriguez, D., Herrera, F., Wing, S., Turner, B. L., D'Apolito, C., Romero-Baez, M., Narváez, P., Martínez, C., Gutierrez, M., Labandeira, C., Bayona, G., Rueda, M., Paez-Reyes, M., Cárdenas, D., Duque, Á., Crowley, J. L., … Silvestro, D. (2021). Extinction at the end-Cretaceous and the origin of modern Neotropical rainforests. *Science*, *372*, 63–68. https://doi.org/10.1126/science.abf1969

Chung, F., & Rodriguez, T. (2017). *Multi-focal image segmentation, classification and authentication: A general framework applied on microscope pollen images*. Éditions universitaires européennes.

Colinvaux, P., Oliveira, P. E., & Moreno, J. E. (1999). *Amazon pollen manual and atlas*. Hardwood Academic Publishers.

Collevatti, R. G., Castañeda, M., Silva-Caminha, S. A. F., & Jaramillo, C. (2024). Application of confocal laser microscopy for identification of modern and fossil pollen grains, an example in palm Mauritiinae. *Review of Palaeobotany and Palynology*, *327*, 105140. https://doi.org/10.1016/j.revpalbo.2024.105140

Crane, P. R. (1986). Form and function in wind dispersed pollen. In S. Blackmore & I. K. Ferguson (Eds.), *Pollen and spores: Form and function* (pp. 179–202). Academic Press.

Cushing, E. J. (2011). Longevity of reference slides of pollen mounted in silicone oil. *Review of Palaeobotany and Palynology*, *164*, 121–131. https://doi.org/10.1016/j.revpalbo.2010.12.001

da Silva, D. S., Quitan, L. N., Goncalves, A. B., Pistori, H., & Borth, M. R. (2014). Application of wavelet transform in the classification of pollen grains. *African Journal of Agricultural Research*, *9*, 908–913.

Daood, A., Ribeiro, E., & Bush, M. (2016a). Pollen recognition using multilayer feature decomposition. *Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Research Society Conference*, 26–31. https://api.semanticscholar.org/CorpusID:12567871

Daood, A. I., Ribeiro, E., & Bush, M. B. (2016b). Classifying pollen using robust sequence alignment of sparse Z-stack volumes. *International Symposium on Visual Computing*, *10072*, 331–340. https://doi.org/10.1007/978-3-319-50835-1_31

de Geus, A. R., Barcelos, C. A. Z., Batista, M. A., & da Silva, S. F. (2019). Large-scale pollen recognition with deep learning. In *2019 27th European signal processing conference (EUSIPCO)* (pp. 1–5). https://doi.org/10.23919/EUSIPCO.2019.8902735

De Sa-otero, M. P., Gonzalez, A., Rodriguez-Damian, M., & Cernadas, E. (2004). Computer-aided identification of allergenic species of Urticaceae pollen. *Grana*, *43*, 224–230. https://doi.org/10.1080/00173130410000749

de Souza, L. R., da Silva, O. L. M., dos Santos, F. A. R., & Carneiro-Torres, D. S. (2020). Pollen morphology of *Croton* of the New World and pollen evolution of tribe Crotoneae (Euphorbiaceae). *Review of Palaeobotany and Palynology*, *281*, 104279.

Díaz-López, E., Rincón, M., Rojo, J., Vaquero, C., Rapp, A., Salmeron-Majadas, S., & Pérez-Badia, R. (2015). Localisation of pollen grains in digitised real daily airborne samples. In J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, F. J. Toledo-Moreo, & H. Adeli (Eds.), *Artificial computation in biology and medicine* (pp. 348–357). Springer International Publishing.

Duller, A., Guller, G., France, I., & Lamb, H. F. (1999). Apollen image database for evaluation of automated identification systems. *Quaternary Newsletter*, *89*, 4–9.

Edwards, K. J., Fyfe, R. M., & Jackson, S. T. (2017). The first 100 years of pollen analysis. *Nature Plants*, *3*, 17001. https://doi.org/10.1038/nplants.2017.1

Erdtman, G. (1952). *Pollen morphology and plant taxonomy, angiosperms*. Almqvist & Wiksell.

Evans, D. M., & Kitson, J. J. (2020). Molecular ecology as a tool for understanding pollination and other plant–insect interactions. *Current Opinion in Insect Science*, *38*, 26–33. https://doi.org/10.1016/j.cois.2020.01.005

Filipovych, R., Daood, A., Ribeiro, E., & Bush, M. (2016). Pollen recognition in optical microscopy by matching multifocal image sequences. In *2016 23rd international conference on pattern recognition (ICPR)*. Cancun, Mexico (pp. 2127–2132). https://doi.org/10.1109/ICPR.2016.7899950

Flenley, J. R. (1968). *The problem of pollen recognition*. The problem of pollen recognition.

France, I., Duller, A. W. G., Duller, G. A. T., & Lamb, H. F. (2000). A new approach to automated pollen analysis. *Quaternary Science Reviews*, *19*, 537–546. https://doi.org/10.1016/S0277-3791(99)00021-9

Gallardo-Caballero, R., García-Orellana, C. J., García-Manso, A., González-Velasco, H. M., Tormo-Molina, R., & Macías-Macías, M. (2019). Precise pollen grain detection in bright field microscopy using deep learning techniques. *Sensors*, *19*(16), 3583. https://doi.org/10.3390/s19163583

García, N. M., Chaves, V. A. E., Briceño, J. C., & Travieso, C. M. (2012). Pollen Grains Contour Analysis on Verification Approach. In E. Corchado, V. Snášel, A. Abraham, M. Woźniak, M. Graña, & S.-B. Cho (Eds.), *Hybrid artificial intelligent systems* (pp. 521–532). Springer Berlin Heidelberg.

Gonçalves, A. B., Souza, J. S., Silva, G. G., Cereda, M. P., Pott, A., Naka, M. H., & Pistori, H. (2016). Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. *PLoS One*, *11*(6), e0157044. https://doi.org/10.1371/journal.pone.0157044

Hodgson, R. M., Holdaway, C. A., Zhang, Y., Fountain, D. W., & Flenley, J. R. (2005). Progress towards a system for the automatic recognition of pollen using light microscope images. In *ISPA 2005. Proceedings of the 4th international symposium on image and signal processing and analysis, 2005* (pp. 76–81). https://doi.org/10.1109/ISPA.2005.195387

Holt, K., Allen, G., Hodgson, R., Marsland, S., & Flenley, J. (2011). Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology*, *167*(3), 175–183. https://doi.org/10.1016/j.revpalbo.2011.08.006

Holt, K. A., & Bennett, K. D. (2014). Principles and methods for automated palynology. *New Phytologist*, *203*, 735–742. https://doi.org/10.1111/nph.12848

Huff, J. (2015). The Airyscan detector from ZEISS: Confocal imaging with improved signal-to-noise ratio and super-resolution. *Nature Methods*, *12*(12), i–ii. https://doi.org/10.1038/nmeth.f.388

Jaramillo, C., & Rueda, M. (2024). *A morphological electronic database of Cretaceous-Cenozoic fossil pollen and spores from northern South America and extant pollen and spores V24*. Smithsonian Tropical Research Institute. https://biogeodb.stri.si.edu/jaramillosdb

Jaramillo, C. A., & Dilcher, D. L. (2001). Middle paleogene palynology of Central Colombia, South America: A study of pollen and spores from tropical latitudes. *Palaeontographica Abteilung B*, *258*, 87–213.

Jardine, P. E., Palazzesi, L., Tellería, M. C., & Barreda, V. D. (2022). Why does pollen morphology vary? Evolutionary dynamics and morphospace occupation in the largest angiosperm order (Asterales). *New Phytologist*, *234*(3), 1075–1087. https://doi.org/10.1111/nph.18024

Johnsrud, S., Yang, H., Nayak, A., & Waduge Punyasena, S. (2013). Semi-automated segmentation of pollen grains in microscopic images: A tool for three imaging modes. *Grana*, *52*(3), 181–191. https://doi.org/10.1080/00173134.2013.768291

Jumanov, I., & Safarov, R. (2022). Optimization of the processing of images of pollen grains based on the use of their specific characteristics and geometric features. In *AIP conference proceedings* (vol. 2637, 040016). https://doi.org/10.1063/5.0118857

Kaya, Y., Mesut Pınar, S., Emre Erez, M., Fidan, M., & Riding, J. B. (2014). Identification of *Onopordum* pollen using the extreme learning machine, a type of artificial neural network. *Palynology*, *38*(1), 129–137. https://doi.org/10.1080/09500340.2013.868173

Khanzhina, N., Kashirin, M., & Filchenkov, A. (2023). New Bayesian focal loss targeting aleatoric uncertainty estimate: Pollen image recognition. In *2023 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 4253–4262). https://doi.org/10.1109/CVPRW59228.2023.00448

Khanzhina, N., Putin, E., Filchenkov, A., & Zamyatina, E. (2018). Pollen grain recognition using convolutional neural network. In *ESANN 2018 proceedings, European symposium on artificial neural networks, computational intelligence and machine learning* (pp. 409–414). Bruges.

Kindt, R. (2020). WorldFlora: An R package for exact and fuzzy matching of plant names against the world Flora online taxonomic backbone data. *Applications in Plant Sciences*, *8*, e11388. https://doi.org/10.1002/aps3.11388

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., Schmidt, A. M. A., & Kjeldsen, K. K. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, *612*(7939), 283–291. https://doi.org/10.1038/s41586-022-05453-y

Korobchevskaya, K., Lagerholm, B. C., Colin-York, H., & Fritzsche, M. (2017). Exploring the potential of Airyscan microscopy for live cell imaging. *Photonics*, *4*(3), 41. https://doi.org/10.3390/photonics4030041

Lagerstrom, R., Holt, K., Arzhaeva, Y., Bischof, L., Haberle, S., Hopf, F., & Lovell, D. (2015). Pollen image classification using the Classifynder system: Algorithm comparison and a case study on New Zealand honey. In C. Sun, T. Bednarz, T. D. Pham, P. Vallotton, & D. Wang (Eds.), *Signal and image analysis for biomedical and life sciences* (pp. 207–226). Springer International Publishing. https://doi.org/10.1007/978-3-319-10984-8_12

Li, P., & Flenley, J. R. (1999). Pollen texture identification using neural networks. *Grana*, *38*(1), 59–64. https://doi.org/10.1080/00173130750044717

Li, P., Treloar, W. J., Flenley, J. R., & Empson, L. (2004). Towards automation of palynology 2: The use of texture measures and neural network analysis for automated identification of optical images of pollen grains. *Journal of Quaternary Science*, *19*(8), 755–762. https://doi.org/10.1002/jqs.874

Lupia, R., Lidgard, S., & Crane, P. R. (1999). Comparing palynological abundance and diversity: Implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology*, *25*, 305–340. https://doi.org/10.1017/S009483730002131X

Mander, L., Jaramillo, C., & Oboh-Ikuenobe, F. (2023). Descriptive systematics of upper Palaeocene–lower Eocene pollen and spores from the northern Niger Delta, South-Eastern Nigeria. *Palynology*, *47*, 2200525. https://doi.org/10.1080/01916122.2023.2200525

Mander, L., Li, M., Mio, W., Fowlkes, C. C., & Punyasena, S. W. (2013). Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1770), 20131905. https://doi.org/10.1098/rspb.2013.1905

Mander, L., Paris-Fukuchi, C., Dick, C., Punyasena, S., & Jaramillo, C. (2021). Phylogenetic and ecological correlates of pollen morphological diversity in a Neotropical rainforest. *Biotropica*, *53*, 74–85.

Mander, L., & Punyasena, S. W. (2014). On the taxonomic resolution of pollen and spore records of Earth's vegetation. *International Journal of Plant Sciences*, *175*(8), 931–945. https://doi.org/10.1086/677680

Marcos, J. V., Nava, R., Cristobal, G., Redondo, R., Escalante-Ramirez, B., Bueno, G., Deniz, O., Gonzalez-Porto, A., Pardo, C., Chung, F., & Rodríguez, T. (2015). Automated pollen identification using microscopic imaging and texture analysis. *Micron*, *68*, 36–46. https://doi.org/10.1016/j.micron.2014.09.002

Menad, H., Ben-Naoum, F., & Amine, A. (2019). Deep convolutional neural network for pollen grains classification. *National Study Day on Research on Computer Sciences*, *2351*. https://api.semanticscholar.org/CorpusID:142503774

Midtvedt, B., Helgadottir, S., Argun, A., Pineda, J., Midtvedt, D., & Volpe, G. (2021). Quantitative digital microscopy with deep learning. *Applied Physics Reviews*, *8*, 011310. https://doi.org/10.1063/5.0034891

Miki, K., & Kawashima, S. (2021). Estimation of pollen counts from light scattering intensity when sampling multiple pollen taxa—Establishment of an automated multi-taxa pollen counting estimation system (AME system). *Atmos. Meas. Tech.*, *14*(1), 685–693. https://doi.org/10.5194/amt-14-685-2021

Moreno, E., Vergara, D., & Jaramillo, C. (2014). Las colecciones palinológicas del Instituto Smithsonian de Investigaciones Tropicales (STRI), Panamá. *Boletín de la Asociación LAtinoamericana de Paleobotánica y Palinología*, *14*, 207–222.

Nguyen, N. R., Donalson-Matasci, M., & Shin, M. C. (2013). Improving pollen classification with less training effort. In *2013 IEEE workshop on applications of computer vision (WACV)* (pp. 421–426). https://doi.org/10.1109/WACV.2013.6475049

Nikolov, D. N., & Tsankova, D. D. (2018). Features extraction for pollen recognition using Gabor filters. *Food Science and Applied Biotechnology*, *81*(2), 86–95.

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., Traidl-Hoffmann, C., Schmidt-Weber, C., & Buters, J. T. M. (2015).

Automatic and online pollen monitoring. *International Archives of Allergy and Immunology*, *167*(3), 158–166. https://doi.org/10.1159/000436968

Palacios-Chavez, R., & Ludlow-Wiechers, B. (1991). *Flora palinológica de la Reserva de la Biosfera de Sian Ka'an, Quintana Roo, México*. Centro de Investigaciones de Quintana Roo.

Polling, M., Li, C., Cao, L., Verbeek, F., de Weger, L. A., Belmonte, J., De Linares, C., Willemse, J., de Boer, H., & Gravendeel, B. (2021). Neural networks for increased accuracy of allergenic pollen monitoring. *Scientific Reports*, *11*(1), 11357. https://doi.org/10.1038/s41598-021-90433-x

Pospiech, M., Javurkova, Z., Hrabec, P., Pavel, S., Ljasovska, S., Bedna, J., & Tremlova, B. (2021). Identification of pollen taxa by different microscopy techniques. *PLoS One*, *16*, e0256808.

Pozo-Baños, M., Ticay-Rivas, J. R., Cabrera-Falcón, J., Arroyo, T.-G. C. M., Sánchez-Chavez, L., Santiago, T., Pérez, S. T., Alonso, J. B., & Ramírez-Bogantes, M. (2012). Image processing for pollen classification. In L. Gbolagade Akeem (Ed.), *Biodiversity enrichment in a diverse world*. IntechOpen. Ch. 19

Punt, W., Hoen, P. P., Blackmore, S., Nilsson, S., & Le Thomas, A. (2007). Glossary of pollen and spore terminology. *Review of Palaeobotany and Palynology*, *143*, 1–81.

Punyasena, S. W., Haselhorst, D. S., Kong, S., Fowlkes, C. C., & Moreno, J. E. (2022). Automated identification of diverse Neotropical pollen samples using convolutional neural networks. *Methods in Ecology and Evolution*, *13*, 2049–2064. https://doi.org/10.1111/2041-210X.13917

Punyasena, S. W., Tcheng, D. K., Wesseln, C., & Mueller, P. G. (2012). Classifying black and white spruce pollen using layered machine learning. *New Phytologist*, *196*(3), 937–944. https://doi.org/10.1111/j.1469-8137.2012.04291.x

Quamar, F., Singh, P., Garg, A., Tripathi, S., Farooqui, A., Shukla, A. N., & Prasad, N. (2022). Pollen characters and their evolutionary and taxonomic significance: Using light and confocal laser scanning microscope to study diverse plant pollen taxa from Central India. *Palynology*, *46*, 1–13. https://doi.org/10.1080/01916122.2022.2070294

Radja, A., Horsley, E. M., Lavrentovich, M. O., & Sweeney, A. M. (2019). Pollen cell wall patterns form from modulated phases. *Cell*, *176*(4), 856–868e810.

Ranzato, M., Taylor, P. E., House, J. M., Flagan, R. C., LeCun, Y., & Perona, P. (2007). Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, *28*(1), 31–39. https://doi.org/10.1016/j.patrec.2006.06.010

Redondo, R., Bueno, G., Chung, F., Nava, R., Víctor Marcos, J., Cristóbal, G., Rodríguez, T., Gonzalez-Porto, A., Pardo, C., Déniz, O., & Escalante-Ramírez, B. (2015). Pollen segmentation and feature evaluation for automatic classification in bright-field microscopy. *Computers and Electronics in Agriculture*, *110*, 56–69. https://doi.org/10.1016/j.compag.2014.09.020

Rodrigues, C. N. M., Goncalves, A. B., Silva, G. G., & Pistori, H. (2015). Evaluation of machine learning and bag of visual words techniques for pollen grains classification. *IEEE Latin America Transactions*, *13*, 3498–3504.

Romero, I. C., Kong, S., Fowlkes, C. C., Urban, M. A., Jaramillo, C., Oboh-Ikuenobe, F., D'Apolito, C., & Punyasena, S. W. (2020). Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 28496–28505.

Romero, I. C., Urban, M. A., & Punyasena, S. W. (2020). Airyscan superresolution microscopy: A high-throughput alternative to electron microscopy for the visualization and analysis of fossil pollen. *Review of Palaeobotany and Palynology*, *276*, 104192. https://doi.org/10.1016/j.revpalbo.2020.104192

Ronneberger, O., Schultz, E., & Burkhardt, H. (2002). Automated pollen recognition using 3D volume images from fluorescence microscopy. *Aerobiologia*, *18*(2), 107–115. https://doi.org/10.1023/A:1020623724584

Roubik, D. W., & Moreno, J. E. (1991). *Pollen and spores of Barro Colorado Island*. Missouri Botanical Garden.

Sevillano, V., Holt, K., & Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLoS ONE*, *15*(6), e0229751. https://doi.org/10.1371/journal.pone.0229751

Sivaguru, M., Urban, M. A., Fried, G., Wesseln, C. J., Mander, L., & Punyasena, S. W. (2018). Comparative performance of airyscan and structured illumination superresolution microscopy in the study of the surface texture and 3D shape of pollen. *Microscopy Research and Technique*, *81*(2), 101–114. https://doi.org/10.1002/jemt.22732

Stevens, W. D., Montiel, O. M., & Raven, P. H. (2014). *Paleobotany and biogeography: A festschrift for Alan Graham in his 80th year*. Missouri Botanical Gargen.

Ticay-Rivas, J. R., del Pozo-Baños, M., Travieso, C. M., Arroyo-Hernández, J., Pérez, S. T., Alonso, J. B., & Mora-Mora, F. (2011). Pollen classification based on geometrical, descriptors and colour features using decorrelation stretching method. In L. Iliadis, I. Maglogiannis, & H. Papadopoulos (Eds.), *Artificial intelligence applications and innovations* (pp. 342–349). Springer Berlin Heidelberg.

Traverse, A. (2007). *Paleopalynology* (Second ed.). Springer. https://doi.org/10.1007/978-1-4020-5610-9

Trivedi, A., Srivastava, A., Farooqui, A., Khan, S., Pokharia, A. K., Ferguson, D. K., & Singh, V. K. (2022). Pollen morphological study in subfamily Papilionoideae using confocal laser scanning microscopy. *Journal of Palaeosciences*, *71*, 123–142. https://doi.org/10.54991/jop.2022.538

Valiente, J. M., Juan-Borras, M., Lopez-Garcia, F., & Escriche, I. (2023). Automatic pollen recognition using convolutional neural networks: The case of the main pollens present in Spanish citrus and rosemary honey. *Journal of Food Composition and Analysis*, *123*, 105605. https://doi.org/10.1016/j.jfca.2023.105605

Viertel, P., & König, M. (2021). Deep learning in palynology. In *41. GIL-Jahrestagung, Informations- und Kommunikationstechnologie in kritischen Zeiten* (pp. 331–336). Gesellschaft für Informatik e.V.

Viertel, P., & König, M. (2022). Pattern recognition methodologies for pollen grain image classification: A survey. *Machine Vision and Applications*, *33*(1), 18. https://doi.org/10.1007/s00138-021-01271-w

Wang, Z., Wang, Z., & Wang, L. (2021). Automatic 3D pollen recognition based on convolutional neural network. *Scientific Programming*, *2021*, 5577307. https://doi.org/10.1155/2021/5577307

Wortley, A. H., Wang, H., Lu, L., Li, D.-z., & Blackmore, S. (2015). Evolution of angiosperm pollen. 1. Introduction. *Annals of the Missouri Botanical Garden*, *100*, 177–226.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Jaramillo, C., Punyasena, S. W., de Alba, D., Alveo, R., Arcila, A., Bermudez, J., Bustos, J., Caballero-Rodriguez, D., Cardenas, K., Caro, D., Carvajal, F., Castañeda, I. M., Chaves, S., D'Apolito, C., Diaz-Jaramillo, A., Diaz, L., Gomez, L., León-Carreño, M., Lopera, P., … Wood, T. (2025). Digitizing collections to unlock the full potential of palynology: A case study with the Smithsonian palynology collection. *Plants, People, Planet*, 1–16. https://doi.org/10.1002/ppp3.70073