

THE *Current*

March 23, 2026

[Harrison Tasoff](#)

The Big Bee Project brings natural history collections into the 21st century

Museum collections have underpinned scientific research for centuries. But physical specimens in boxes and drawers don't easily lend themselves to the research techniques of the new millennium. "How can we apply these techniques to natural history collections, especially when much of the intrinsic information a specimen has to offer is difficult to quantify?" asks [Katja Seltmann](#), director of UCSB's Cheadle Center for Biodiversity & Ecological Restoration.

Enter the Big Bee Project: a pioneering initiative to bring natural history collections into the century of AI, big data and networked databases.

Plan A is Plan Bee

Led by UC Santa Barbara, the project is a collaboration across 13 U.S. institutions to create over one million high-resolution 2D and 3D images of bee specimens and annotated datasets of bee traits. The campaign was given \$3 million in funding by the National Science Foundation to cover three years of work.

Scientists have studied animal traits since biology became a scientific discipline. Before the advent of genetics and genomics, scientists could only study an

organism's phenotype, or physical characteristics, to understand its position in the tree of life. Some traits — like length, weight and color — lend themselves to straightforward measurement. But others — like shape and pattern — defy traditional quantification.

Seltmann's group is the first to look at phenotypic variation in bees using museum specimens at a big-data scale. And they are doing it through machine learning, computer vision and crowdsourcing.

For instance, Seltmann's team uploaded detailed photos of bees to the Notes From Nature database, where more than 5,000 volunteers took body measurements to build out an annotated dataset. Researchers found that input from volunteers was comparable to trained scientists, emphasizing the efficiency of crowdsourcing these kinds of tasks.

In another part of the project, Seltmann's group used computer vision and machine learning to quantify how hairy different bees are, in terms of both density and color. They were able to characterize 611 bee species across 377 genera. The team published [a paper](#) on the correlations they found between hair coverage and how bees adapt to different climates and environmental changes.

Seltmann has also teamed up with engineering professor [B.S. Manjunath](#), an expert in computer vision at UCSB, to analyze the structure of bee wings. Wing venation can be a diagnostic characteristic between species, but taking advantage of this currently requires hours of expert analysis. Automating this process could enable rapid, non-invasive species identification from a photo out in the field.

Beyond the bee

UCSB's contribution to the [Big Bee Project](#) has already produced 36 publications, 63 posters and talks, and 23 shared datasets and reports. But the project's impact extends beyond entomology. The initiative enabled Seltman's group to pilot a number of innovative techniques and codify new standards for natural history collection research. It also enabled the group to carry out fundamental work, like deciding what metrics to use to quantify these complex traits and developing the methods to do so.

It's simple to count things, and measurements are straightforward to take. But it's challenging to describe the various forms that life can take. The shape of a leaf or the venation on a wing can be quite subjective. "So scientists have really been pushing to turn these into quantitative things, like numbers, matrices and graphs," Seltmann explained. At the same time, they're building large language models and neural networks to identify and organize this new kind of data.

This novel approach to natural history collections changes what is possible to do with them. "We can ask brand new questions because it is now far easier than it ever was to pull patterns from specimens and turn those into things that we can analyze with statistics," Seltmann said.

This is particularly true for image data, which has never been more ubiquitous in science as in daily life. "Storing images is a no-brainer," said Manjunath, whose lab developed [BisQue](#), a platform to store, visualize, organize and analyze images in the cloud. But his involvement is taking this to the next stage: "What can we do with those images? What can we learn with those images using computer vision and AI?"

In addition to working with Seltmann on bee wings, Manjunath's team has applied BisQue to topics as sundry as detecting neural pathologies from CT scans and identifying prairie dogs in aerial photos. The system accepts over 100 file types and can be augmented with custom software modules for each project. The group is currently working to integrate natural language processing into BisQue's interface to enable more organic interactions between scientists and software.

While the current Big Bee Project is sunsetting, Seltmann's work on natural history collections will continue. The next phase of her work will be to expand access to the wealth of data they've documented to researchers and professionals in other disciplines. She's particularly interested in collaborations with engineers and material scientists mining biology for solutions to human challenges.

Tags

[Artificial Intelligence](#)

[Data Science](#)

Media Contact

Harrison Tasoff

Science Writer

(805) 893-7220

harrisontasoff@ucsb.edu

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.