

Smarter tools for peering into the microscopic world

Breakthroughs in microbial analysis could enhance disease prediction, environmental protection and our understanding of the planet's smallest life forms

By Richard Harth, ASU News
December 11, 2025

The microscopic organisms that fill our bodies, soils, oceans and atmosphere play essential roles in human health and the planet's ecosystems. Yet even with modern DNA sequencing, figuring out what these microbes are and how they are related to one another remains extremely difficult.

In a pair of new studies, researchers at Arizona State University introduce powerful tools that make this work easier, more accurate and far more scalable. One tool improves how scientists build microbial family trees. The other provides a software foundation used worldwide to analyze biological data.

Together, these advances strengthen the scientific foundations of microbiome research, disease tracking, environmental monitoring and emerging fields like precision medicine.

"Our team builds open-source software tools because we believe that when everyone can access and extend scientific tools, the entire community benefits and discovery accelerates," said [Qiyun Zhu](#), lead author of the new studies.

Zhu is a researcher with the [Biodesign Center for Fundamental and Applied Microbiomics](#) and an assistant professor at ASU's [School of Life Sciences](#). He is joined by ASU colleagues and international collaborators.

The [first study](#), on improving marker genes, appears in the journal Nature Communications. The [second study](#), describing an open-source software library known as [scikit-bio](#), appears in Nature Methods.

Family affair

Building detailed and accurate evolutionary trees is essential for understanding how microbes evolve and influence the world. Better evolutionary trees improve disease tracking and help scientists follow how harmful microbes change over time. They also sharpen environmental research, showing how microbial communities respond to pollution or climate shifts. Clearer microbial identification also strengthens studies of the gut microbiome and its role in health.

Uncovering how microbes are related begins with choosing the right marker genes — the signposts in DNA that trace their evolutionary history.

For many years, scientists relied on the same small set of traditional marker genes. But in the growing field of metagenomics, researchers now work with millions of genomes, often directly from environmental samples. Metagenomics allows scientists to scoop up all the DNA in an environment and sequence it at once, revealing entire hidden communities of microbes.

These genomes are extremely valuable, but they're often incomplete or uneven in quality. That makes it hard to use a fixed set of marker genes and expect accurate evolutionary results.

To solve this, Zhu and colleagues helped develop TMarSel (short for Tree-based Marker Selection). Instead of choosing genes by hand, TMarSel automatically searches through thousands of possible gene families and selects the combination that builds the most reliable evolutionary tree. It evaluates each gene for how common it is, how informative it is and how much it contributes to a stable, meaningful picture of microbial relationships.

The result is a flexible, data-driven way to build microbial trees that work well even for large and diverse groups of organisms — and even when many genomes are only partly complete.

Scikit-bio: Ancestry.com for microbes

Zhu is also a lead developer of scikit-bio, a vast, open-source software library. Scikit-bio gives scientists the tools they need to analyze huge biological datasets. It is particularly useful for studying microbiomes — communities of microbes that live in a specific environment, such as the human gut.

Biological data sets are unlike any other kind of data: they are extremely large, very sparse and often include thousands of interconnected features. Standard data-analysis programs are not built for this level of fragmentation and complexity. Scikit-bio fills this gap by offering more than 500 functions for tasks such as:

- Comparing microbial communities.

- Calculating diversity.

- Transforming compositional data.

- Analyzing DNA, RNA and protein sequences.

Building and modifying phylogenetic trees.

Preparing data for machine learning.

The project is community-driven, supported by more than 80 contributors and maintained with rigorous testing and documentation. It has already been cited in tens of thousands of scientific papers across medicine, ecology, climate science and cancer biology. It has become an essential tool for researchers analyzing the microbiome and other large, data-rich areas of modern biology.

A new era in microbial research

As biological datasets grow, tools like scikit-bio and TMarSel make large-scale research more reliable and reproducible.

The studies reinforce ASU's expanding role at the intersection of biology and computation. Zhu's work shows how combining evolutionary insight with advanced software engineering can produce tools used by scientists around the world.

As DNA sequencing continues to become faster and cheaper, scientists will uncover even more of the microbial universe. Tools like TMarSel and scikit-bio ensure that this flood of data can be transformed into real scientific insight.

This story originally appeared on [ASU News](#).

Main image



A digital tree built from DNA-like connections represents how scientists trace the ancestry of microbes. New tools such as TMarSel and scikit-bio are helping researchers build clearer microbial family trees and make better sense of the growing flood of genetic data. Graphic courtesy of the Biodesign Institute

Text image(s)



