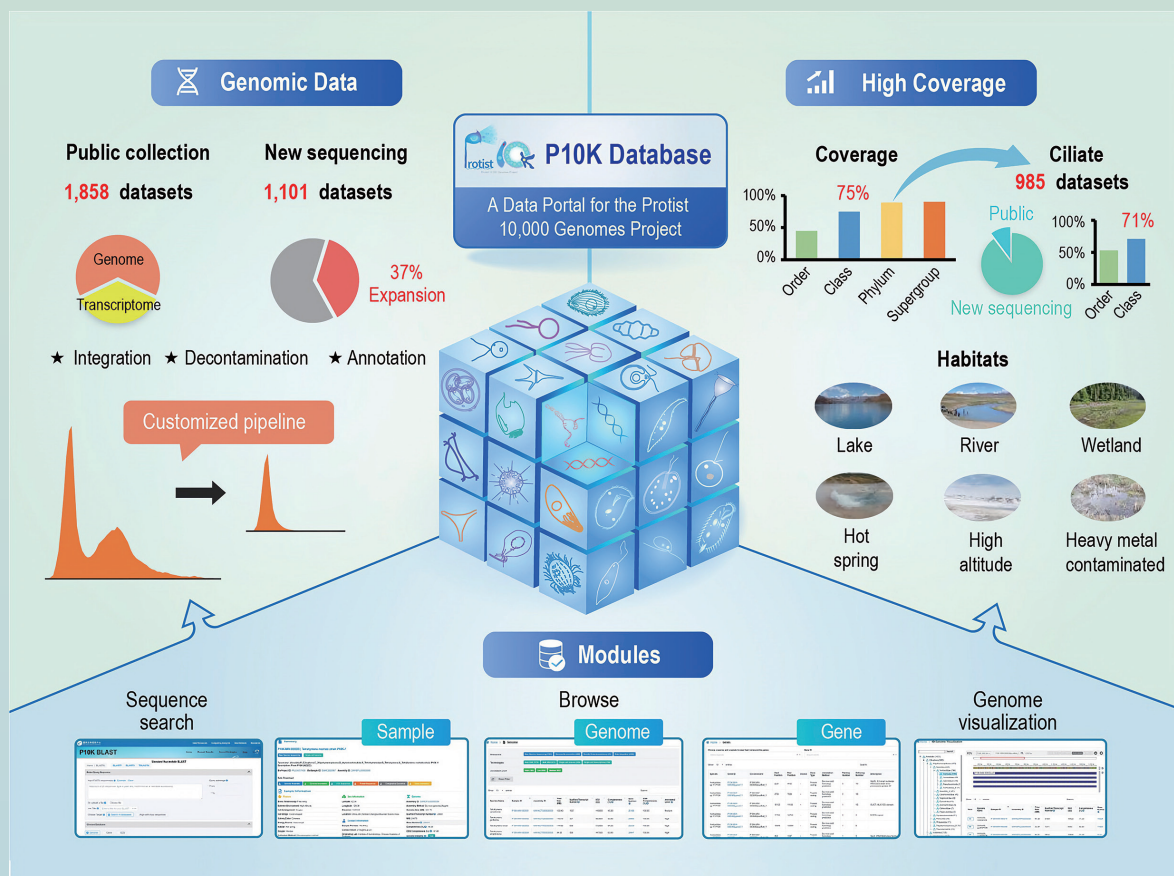


# Seeing the Unseen

By YAN Fusheng

The microscopic world is full of wonder, but peering into this realm often requires cutting-edge tools. Through large-scale genome sequencing, researchers have shed new light on the intricate inner workings of protists, single-cell organisms that play pivotal roles across ecosystems. Their latest findings unveil striking genetic intricacies underlying the astounding diversity of these minute yet mighty microbes.



The P10K project aims to create a comprehensive genetic resource database for microbes called protists. (Image by CAS)

Protists – a broad swath of single-celled eukaryotes distinct from animals, plants, and fungi – are tiny creatures with an outsized impact. Populating oceans, lakes, and soils across the globe, these single-cell organisms play indispensable roles in maintaining balance within ecosystems.

Yet despite their profound influence, protists remain enigmatic inhabitants of the microbial world. Compared to their better-studied bacterial counterparts, relatively little is known about their genetic makeup. But now, an ambitious initiative is pulling back the curtain on these overlooked microbes by sequencing thousands of protist genomes – a technological feat made possible through advanced single-cell analysis.

In December 2019, a group of scientists led by the Institute of Hydrobiology (IHB) of the Chinese Academy of Sciences (CAS) launched the Protist 10,000 Genomes Project (P10K). The main goal of this ambitious project is to construct an extensive genetic database for protists.

Up to now, this project has already amassed nearly 3,000 genomes and transcriptomes from protists representing a staggering diversity of environmental niches and taxonomic groups.

Its recent publication in *Nucleic Acids Research* on November 1, 2023 unveils a substantial catalog of protists' genetic data, providing a key resource for deciphering how protists evolve, adapt, and keep ecosystems in balance.

The study generated an invaluable collection of genetic information spanning the kaleidoscopic diversity of protists, which will tremendously benefit researchers across many fields.

The research team implemented sophisticated techniques to sequence single protist cells collected from habitats like oceans and lakes across China. These challenging samples traversed a vast diversity, from

hermits living in extreme environments to parasitic species interacting with animal hosts.

Advanced sequencing platforms decoded the genomes within individual cells, circumventing the need to culture finicky microbes in the laboratory. Powerful computational tools then cleaned up the raw genetic code, weeding out contamination from other microbes.

By leveraging cutting-edge single-cell sequencing coupled with specialized bioinformatic pipelines, they efficaciously extracted genetic insights from tiny amounts of starting material.

The project's analysis uncovered a cornucopia of genetic diversity, with distinct codes specifying varied traits that enable protists to thrive in various settings. For example, certain ciliates – which employ hair-like structures to move around – were found to use alternative codons – triplets of nucleotides in DNA or RNA that encode for specific amino acids – compared to many organisms. This likely provides advantages in their specialized environmental niches.

Intriguingly, the researchers discovered discrepancies between the genetic codes embedded in the sequenced genomes versus established databases. This suggests potential inaccuracies in the existing taxonomic classification system for protists. These findings could inform efforts to refine the protist family tree.

In addition to sequencing protist genomes, the team has built an intuitive online database (visit via <https://ngdc.cnca.ac.cn/p10k/>) to provide free access to the voluminous genetic repository.

Though the P10K initiative has greatly expanded protist genomic resources, much discovery still lies ahead. The ambitious project aims to eventually sequence over 10,000 genomes, ensuring an unparalleled understanding of Earth's most abundant eukaryotes.

## Reference

Gao, X., Chen, K., Xiong, J., Zou, D., Yang, F., Ma, Y., . . . Miao, W. (2024). The P10K database: a data portal for the protist 10 000 genomes project. *Nucleic Acids Research*, 52(D1), D747-d755. doi:10.1093/nar/gkad992