

This Ambitious Project Wants to Sequence the DNA of All Complex Life on Earth

The Earth BioGenome Project seeks to genetically profile over a million plants, animals, and fungi as they build an atlas of complex life on Earth.

By Shelly Fan

5 min. read · [View original](#)

“We’re only just beginning to understand the full majesty of life on Earth,” wrote the founding members of the [Earth BioGenome Project in 2018](#). The ambitious project raised eyebrows when first announced. It seeks to genetically profile over a million plants, animals, and fungi. Documenting these genomes is the first step to building an atlas of complex life on Earth.

Many living species remain mysterious to science. A database resulting from the project would be a precious resource for monitoring biodiversity. It could also shed light on the genetic “dark matter” of complex life to inspire new biomaterials, medicines, or spark ideas for synthetic biology. Other insights could tailor agricultural practices to ramp up food production and feed a growing global population.

In other words, digging into living creatures’ genetic data is set to unveil “unimaginable biological secrets,” wrote the team.

The problem? A hefty price tag. With an estimated cost of \$4.7 billion, even the founders of the project called it a moonshot. However, against all odds, the project has made progress, with 3,000 genomes already sequenced and 10,000 more species expected by 2026.

While lagging its original goal of sequencing roughly 1.7 million genomes in a decade, the project still hopes to hit this goal by 2032—later than the original goalpost, but with a much lower price tag

thanks to more efficient DNA sequencing technologies.

Meanwhile, the international team has also built infrastructure to share gene sequencing data, and machine learning methods are further helping the consortium analyze thousands of datasets—helping characterize new species and monitor DNA data for endangered ones.

Expanding the Scope

Genetic material is everywhere. It's an abundant resource to make sense of life of Earth. As genetic sequencing becomes faster, cheaper, and more reliable, recent studies have begun digging into information represented by DNA from species across the globe.

One method, dubbed metagenomics, captures and analyzes microbial DNA gathered in a variety of environments, from city sewers to boiling hot springs. The method captures and analyzes all DNA from a particular source to paint a broad genetic picture of bacteria from a given environment. Rather than bacteria, the [Earth BioGenome Project](#), or EBP, is

aiming to sequence the genomes of individual eukaryotic creatures— basically, those that keep most of their DNA in a nut-like structure, or nucleus, inside each cell.

Humans, plants, fungi, and other animals all fall into this group. In one estimate, there are roughly 10 to 15 million eukaryotic species on our planet. But just a little over two million have been documented.

Sequencing DNA from eukaryotic cells could vastly expand our knowledge of Earth's genetic diversity. Such a database could also be a treasure trove for synthetic biology. Scientists have already tinkered with the genetic blueprints of life in bacteria and yeast cells. Deciphering—and then reprogramming—their genes has led to advances such as coaxing bacteria cells to pump out biofuels, degradable materials, and medicines such as insulin.

Charting eukaryotes' genomes could further inspire new materials or medicines. For example, [cytarabine](#), a chemotherapy drug, was initially isolated from a sponge-like sea creature and

approved by the FDA to treat blood cancers that spread to the brain. Other plant-derived medications are already being used to tackle viral infections or to control pain. From nearly 400,000 different plant species, hundreds of medicines have already been approved and are on the market. Similarly, deciphering plant genetics have galvanized ideas for new biodegradable materials and biofuels.

Genetic sequences from complex organisms can “provide the raw materials for genome engineering and synthetic biology to produce valuable bioproducts at industrial scale,” wrote the team.

Medical and industrial uses aside, the effort also documents biodiversity. Creating a DNA digital library of all known eukaryotic life can pinpoint which species are most at risk—including species not yet fully characterized—providing data for earlier intervention.

“For the first time in history, it is possible to efficiently sequence the genomes of all known species and to use genomics to help discover the remaining 80 to 90

percent of species that are currently hidden from science,” wrote the team.

Soldiering On

The project [has three phases](#).

Phase one lays the groundwork. It establishes the species to be sequenced, builds digital infrastructure for data sharing, develops an analysis toolkit. The most important goal is to build a reference DNA sequence for species similar in genetic makeup—that is, those in a “family.”

Reference genomes are incredibly important for genetic studies. True to their name, scientists rely on them as a baseline when comparing genetic variants—for example, to track down genes related to inherited diseases in humans or sugar content in different variants of crops.

Phase two of the project will begin analyzing the sequencing data and form strategies to maintain biodiversity. The last phase integrates all previous work to potentially revise how different species fit into our evolutionary tree. Scientists will also integrate climate data into this

phase and tease out the impacts of climate change on biodiversity.

The international project began in 2018 and included the US, UK, Denmark, and China, with most DNA specimens sequenced at facilities in China and the UK. Today, 28 countries spanning six continents have signed on. Most DNA material isolated from individual species is directly sequenced on site, reducing the cost of transportation while increasing fidelity.

Not all participants have easy access to DNA sequencing facilities. One institution, Wellcome Sanger, developed a portable DNA sequencing lab that could help scientists working in rural areas to capture the genetic blueprints of exotic plants and animals. The device sequenced the DNA of a type of sunflower with potential medicinal properties in Africa, among other specimens from exotic locations.

EBP follows in the footsteps of other global projects aiming to sequence the Earth's microbes, such as the [National Microbiome Initiative](#) or the [Earth Microbiome Project](#). Once also

considered moonshots, these have secured funding from government agencies and private investments.

Despite the enthusiasm of its participants, EBP is still short billions of dollars to guide it to full completion. But the project's price tag—originally estimated in the billions of dollars—may be far less.

Thanks to more efficient and cheaper genetic sequencing methods, the current cost of phase one is expected to be half the original estimate—around \$265 million.

It's still a hefty sum, but for participants, the resulting database and methods are worth it. "We now have a common forum to learn together about how to produce genomes with the highest possible quality," Alexandre Aleixo at the Vale Institute of Technology, who participated in the project, [told Science](#).

Given the influence bacterial genetics has already had on biomedicine and biofuels, it's likely that deciphering eukaryote DNA can spur further inspiration. In the end, the project relies

on a global collaboration to benefit humanity.

“The far-reaching potential benefits of creating an open digital repository of genomic information for life on Earth can be realized only by a coordinated international effort,” [wrote](#) the team.

Image Credit: [M. Richter](#) on [Pixabay](#)