



THE EARTH BIOGENOME PROJECT: BUILDING THE ULTIMATE LIBRARY OF LIFE

Mark Blaxter^{1*} and Harris A. Lewin^{2,3}

¹Tree of Life, Wellcome Sanger Institute, Cambridge, United Kingdom

²Global Futures Laboratory, Walton Center for Planetary Health, Arizona State University, Tempe, AZ, United States

³Genome Center, Department of Evolution and Ecology, University of California, Davis, Davis, CA, United States

YOUNG REVIEWERS:



ALLISON

AGE: 10

Earth is full of amazing life forms, from giant whales to tiny microbes. Scientists want to understand and protect this biodiversity—but first, they need to know what is out there. That is where the Earth BioGenome Project (EBP) comes in. This worldwide effort aims to read the complete set of genetic instructions (genomes) for every known eukaryotic species on Earth. These genome sequences can reveal how species are related, how they evolved, and how they might help us in the future. EBP scientists must collect high-quality samples, use special tools to read long pieces of DNA, and stitch those pieces together using powerful computers. Then they must figure out what the genes do and share the data with the world. It is a massive task—but if they succeed, the result will be a global library of life that could help protect species, improve health, and teach us more about how life works.

EUKARYOTES

Organisms whose cells have a nucleus.

Eukaryotes include animals, plants, fungi, and many microbes. They are different from bacteria and archaea, which do not have a cell nucleus.

SPECIES

A group of living things that can mate and have offspring. Dogs, oak trees, and humans are all different species with their own unique traits and behaviors.

BIODIVERSITY

The variety of life on Earth, including all plants, animals, fungi, and microbes. Biodiversity keeps ecosystems healthy and helps support life—including human life—on the planet.

GENOMES

The complete set of genetic instructions found in every living thing's DNA. Genomes tell each organism how to grow, survive, and carry out all its functions.

DNA SEQUENCING

A method scientists use to read the exact order of DNA building blocks—A, T, C, and G—in an organism's genome, revealing important information about its biology.

A PLANET FULL OF LIFE

Earth is a living planet, home to more types of life than you can possibly imagine—ranging from tiny, single-celled organisms to 200-ton blue whales and towering redwood trees that can live for thousands of years. Life can be split into three domains: *Bacteria*, *Archaea*, and *Eukarya*. The **eukaryotes**, organisms whose cells have a nucleus, include plants, animals, fungi, and tiny microbes called protists. Scientists estimate that the number of eukaryotic **species** that currently live on our planet is around 10 million [1, 2]! Fewer than 2 million eukaryotic species have been discovered and named **so far**, which means that millions of fascinating organisms are yet to be found. These mind-blowing numbers illustrate Earth's enormous **biodiversity**, which plays a huge role in keeping ecosystems healthy and balanced, and which supports human life on this shared planet.

Life has evolved over the last 3 billion years from very primitive beginnings to the wonderful diversity we see today. Evolution has not always gone smoothly. Life on Earth has gone through five major extinctions in the past, caused by events like the asteroid impact that wiped out the dinosaurs or massive volcanic eruptions and other dramatic Earth changes. Today we are living through a sixth mass extinction, with species disappearing as fast as they did during the worst die-offs in Earth's history. It is depressingly likely that this sixth extinction is not being caused by outside forces, but by the actions of humans—such as releasing greenhouse gases into the atmosphere, destroying wilderness areas for farming and expanding our cities, polluting ecosystems, and overharvesting certain species. We are destroying biodiversity, and if we lose species before we understand them, we may never know how they function in Earth's ecosystems—or how they could help us.

One way scientists can study species—even those that are rare or hard to find—is by looking at their **genomes**, the DNA that contains the basic instructions (or “code”) for each living thing. Genomes can be examined by a technique called **DNA sequencing**, in which scientists “read” an organism's genetic code. Genome sequences can help researchers identify new species, understand how organisms are related to each other, and even discover traits that could be useful in medicine, agriculture, and conservation.

SEQUENCING LIFE FOR THE FUTURE OF LIFE

To unlock the information in genomes and preserve knowledge about Earth's species before it is lost, a large group of scientists from all over the world launched the **Earth BioGenome Project (EBP)**—a massive

effort to create a digital library of life by sequencing the genomes of all known eukaryote species on Earth [3, 4].

The EBP has an ambitious goal: to complete this enormous task in just 10 years. Scientists are currently wrapping up Phase I, which focused on sequencing one example genome from approximately 5,000 major branches (called families) of the tree of life, as well as another 5,000 species important for conservation or scientific interest. So far, EBP researchers have completed this work for over 4,000 species from more than 1,000 families (for recent numbers of sequences completed, see [here](#).) Now, as they move into Phase II, they plan to collect specimens for 300,000 species and sequence half of those (150,000 species!) in just 4 years. These 150,000 will include about half of all known genera—the groups that sit one level above species in the taxonomic hierarchy—and also include species that are especially important for conservation, human health, agriculture, or scientific research because of their unique traits. The ultimate goal of Phase III is to complete the sequencing of all known eukaryotic species by 2035.

The EBP has been called a “moonshot” project because, like sending humans to the moon the first time, this project is ambitious, challenging, and exciting—and all the technologies (and money) needed to complete the project are not yet in place. Phase II of EBP is no simple task—researchers will need to produce over 3,000 new genomes per month. This work requires a careful, step-by-step process, and each stage has its own unique challenges.

HOW DO SCIENTISTS SEQUENCE GENOMES?

The EBP aims to provide more than just a “rough draft” of each organism’s genome—the goal is to create reference genomes, which are high-quality genetic maps, carefully assembled and checked for accuracy and assigned to individual chromosomes. Scientists call them reference genomes because they serve as the “gold standard” for studying that species and its relatives, helping researchers identify genetic differences between individuals, track evolutionary relationships, and compare species across the tree of life. To build these high-quality genetic blueprints, scientists follow a multi-step process described in the sections that follow and summarized in [Figure 1](#).

Step 1: Collecting Specimens

The first step in genome sequencing is to find and collect a sample from the species scientists want to study. For some eukaryotic species, this is simple—scientists can easily collect a leaf from a common tree or a single specimen of a common insect like a ladybird (also called ladybug). But for others, it is much harder. Many species live deep in the ocean, in dense jungles, or in extreme environments like the Arctic.

Figure 1

Creating high-quality reference genomes. Step 1: researchers collect a sample from the species they want to sequence, along with data called “metadata” that track details of each sample (like where and when it was collected). Step 2: DNA is extracted from the specimen and broken into fragments, which are sequenced. Step 3: sequences are assembled in the right order to recreate the genome, like putting a puzzle together. Step 4: the genome is analyzed (annotated) to figure out which genes it contains. Step 5: data are organized, safely stored, and shared with researchers all over the world using databases connected to the internet.

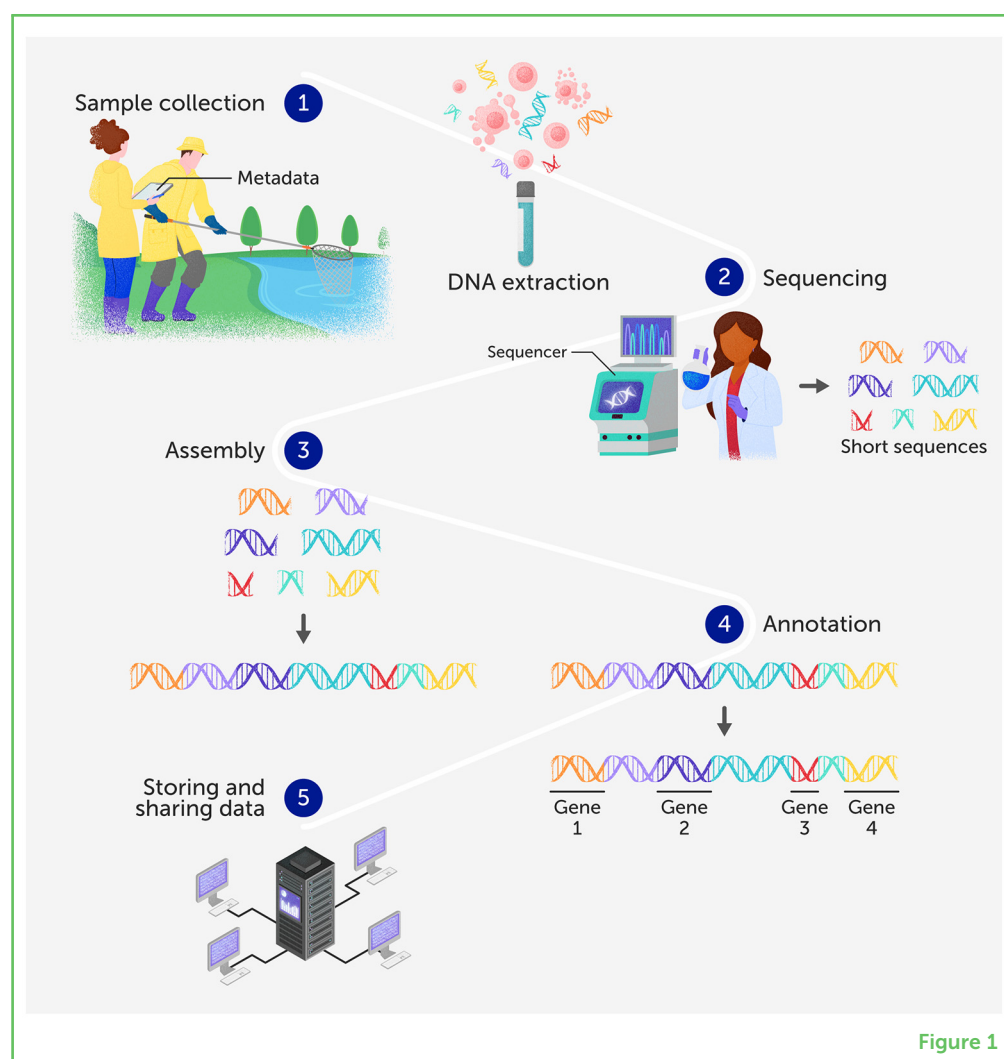


Figure 1

Others are microscopic or so rare that scientists may have only seen them a few times in the wild.

To collect specimens for the EBP, researchers must work with conservationists, biologists, Indigenous Peoples, and local communities—the people who know where to find certain species. Sometimes they gather samples from zoos, botanical gardens, museums, or research centers instead of collecting wild organisms. Collecting samples must be done carefully and ethically. Collectors follow strict rules to make sure they do not harm endangered species or take samples without first getting permission from local governments and Indigenous communities.

Genetic material is delicate, so samples must be carefully preserved so the DNA does not break down before it reaches the lab where sequencing will take place. Samples can be stored in chemicals that keep the DNA intact, and some are frozen in special **biobanks** until they can be sequenced. Keeping samples frozen or safely transporting them from remote locations can be expensive and difficult.

BIOBANK

A storage facility that keeps biological samples—like tissue or DNA—frozen or preserved for future research, including genome sequencing and studies on health, evolution, or conservation.

Step 2: Sequencing the Genome

Once scientists have a well-preserved specimen and the species is properly identified, the next step is to **sequence** its genome. Special chemicals are used to break open cells and pull out the DNA from the cell nucleus, which is then purified to remove any proteins or other molecules that might interfere with sequencing. To ensure the highest quality results, scientists need quite a lot of “clean”, undamaged DNA, and this can be a big challenge when working with rare or tiny specimens.

Because genomes are too long to read all at once, the DNA is first broken into smaller fragments, and then the “sequence” of each fragment is read. DNA is made up of four chemical building blocks, the letters of the **genetic code**, called A, G, C, and T. Specialized instruments read the sequence of the bases in each fragment. The longer the fragments are, the easier it is for scientists to put the pieces of the sequence together to create the full genome, so the EBP currently uses some very clever technologies that can read the sequences of single DNA molecules tens of thousands of DNA letters long. Modern sequencing technologies are incredibly fast, allowing scientists to read billions of DNA letters in a matter of hours.

To make it easier to sequence samples collected in remote areas, EBP scientists are developing a “sequencing lab in a box”—a portable system that contains everything needed to sequence DNA right where species are found (**Figure 2**). These mobile labs will reduce the need to preserve samples and transport them back to distant laboratories. When sequencing is done locally, researchers and communities living in biodiverse regions can play a more direct role in studying and benefiting from their own natural resources.

Step 3: Assembling the Genome

Sequencing a genome is not as simple as reading a book from start to finish. Because DNA must be chopped into pieces during sequencing, scientists need to put those pieces back together—a process called genome assembly. Assembly can be extremely difficult, especially for species with large or complex genomes. Computer scientists called bioinformaticians use powerful computers and computer programs that look for overlapping sequences between fragments, piecing them together like matching edges in a giant jigsaw puzzle. Even with the best sequencing methods, some parts of the genome can be tricky to assemble correctly, and special techniques are often needed to help place sequences in the right order.

Genome assembly is a big challenge for the EBP. The smallest eukaryote genome we know of has “only” 1.2 million DNA letters, but the largest has over 160 billion letters. The human genome has 3 billion letters. Many genomes contain repeated sections that are nearly identical, making it difficult to tell where one repeat ends

Figure 2

Genetic material is delicate and breaks down easily, so analyzing DNA right where species are found can be a big advantage. Since some samples must be collected from species that live in very remote areas, we are developing a portable “sequencing lab in a (big) box”, called the gBox, that contains all the equipment and computer power needed to sequence DNA anywhere in the world, including countries that cannot afford such equipment or do not have the expertise to run it. Training will be provided with the gBox.

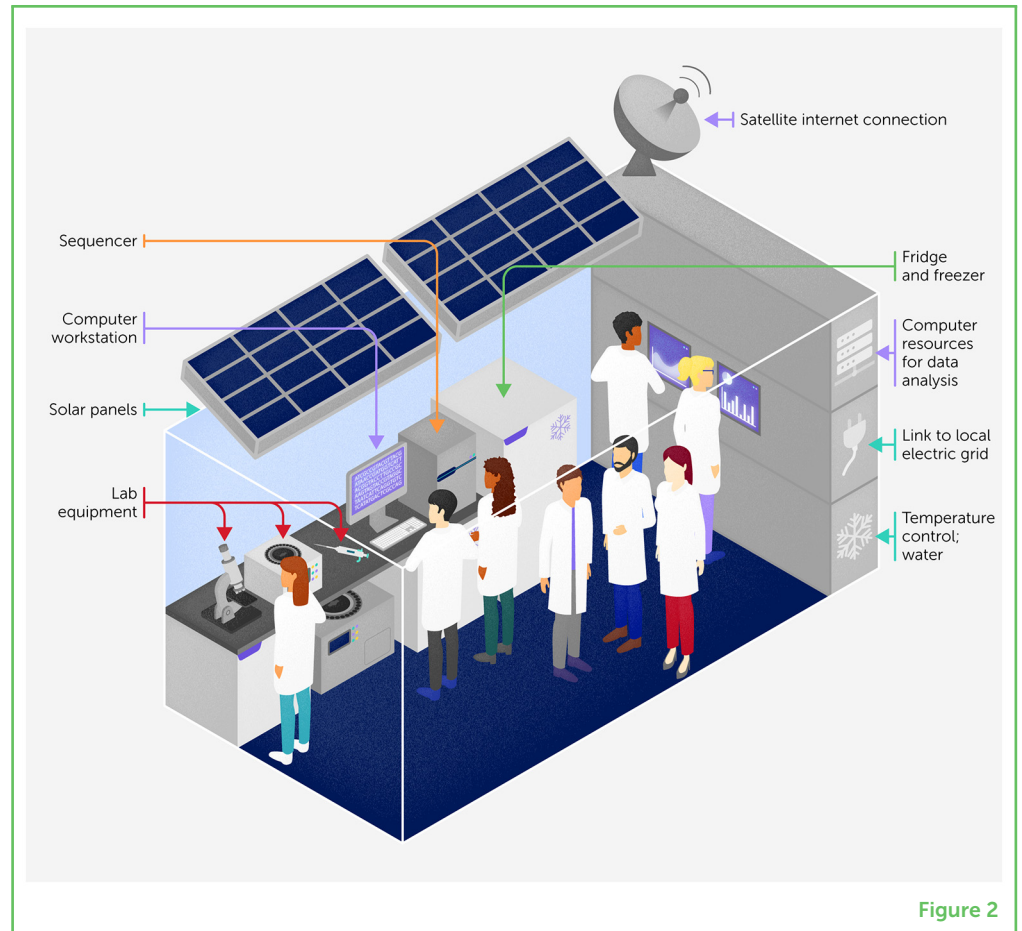


Figure 2

and another begins. Advances in computing power and sequencing technology will eventually help scientists piece together even the most complex genomes.

Step 4: Analyzing the Genome

Once a genome is assembled, scientists must figure out what the DNA sequences actually do—what the genetic code they have read means. This step, called **annotation**, is like labeling the parts of a blueprint to show what each section of a building is for. Without annotation, a genome is just a long string of letters that cannot be interpreted.

To annotate a genome, scientists search the genome for genes—sections of DNA that provide instructions for making proteins. They find the genes in three ways. The first is by using RNA data from the same species to see which genes are active in that species. RNA is a molecule made when a gene is turned on, so when the RNA sequences are aligned to the genomic DNA, it shows which parts of the genome are being used. The second way to find genes is by comparing the genome sequences of *related* species and identifying the parts that have been conserved by evolution—the conserved areas are more likely to be genes. The third way is by training computer programs to find regions that look like genes based on the evidence gathered by the

ANNOTATION

The process of labeling parts of a genome, like genes and control regions, to help scientists understand what different DNA sequences do and how they affect an organism.

first two approaches. Genes are only part of the story—annotation also helps identify other important genome regions, such as sequences that control when genes are turned on or off. Annotation is an ongoing process. As scientists learn more about a species and discover new genes, they can go back and improve previous annotations.

Step 5: Storing and Sharing Data

EBP is creating a massive genome library that will allow researchers everywhere to compare genomes, which will improve understanding of life on Earth. EBP makes its data publicly available, so *anyone* can access and use it for discoveries in conservation, medicine, agriculture, and more. But with thousands of genomes being sequenced, storing and organizing all this information is a huge challenge.

EBP uses a secure global network of online databases to store genome sequences, along with important details (called metadata) like where and when each specimen was collected. Details ensure that the information is properly cataloged and easy to find. Scientists also follow strict data-sharing rules to make sure the information they collect is shared fairly [5, 6]. Some countries, Indigenous Peoples, and local communities have special connections to certain species, so researchers must make sure the benefits of new discoveries are shared with the people and places that helped make them possible.

WHY DOES EBP MATTER?

In short, EBP's mission is to unlock the hidden code of life and use that knowledge to benefit science, species conservation, and society (Figure 3) [7]. For example, understanding genomes helps scientists track how species evolve, identify those at risk of extinction, and develop conservation strategies to protect Earth's biodiversity. In farming, decoding plant genomes could help scientists create crops that resist diseases and tolerate harsh conditions, improving food security as Earth's climate changes. In medicine, studying genomes of animals that can regrow lost limbs or resist cancer may inspire new treatments for human injuries or diseases. Many plants and animals are poisonous or venomous to humans. Surprisingly, studying the ways these poisons and venoms interact with human systems can lead to new ideas for how to treat pain or combat disease.

Reaching this goal is both a scientific and financial challenge. Scientifically, researchers must continue developing faster, more accurate methods for sequencing, assembling, and analyzing many genomes. Sequencing all eukaryotic species will cost billions of dollars and require cooperation between governments, researchers, and conservation groups all over the world. While major research institutions and governments have already invested in EBP, much more money is needed to keep the project moving forward. The estimated cost of \$3.9 billion may sound high, but so is the potential

Figure 3

The mission of the EBP is to unlock the hidden code of life and use it to benefit science, other species, and society. **(A)** The project can increase overall scientific knowledge by providing details about the evolution of species and biodiversity. **(B)** This knowledge could be used to identify species that are at risk of extinction and develop conservation strategies, preserving Earth's biodiversity. **(C)** Decoding plant genomes could lead to crops that can resist pests, diseases, or harsh weather. **(D)** Studying animal genomes could lead to medical advances like new treatments for diseases, ways to prevent cancer, or methods for treating serious injuries.

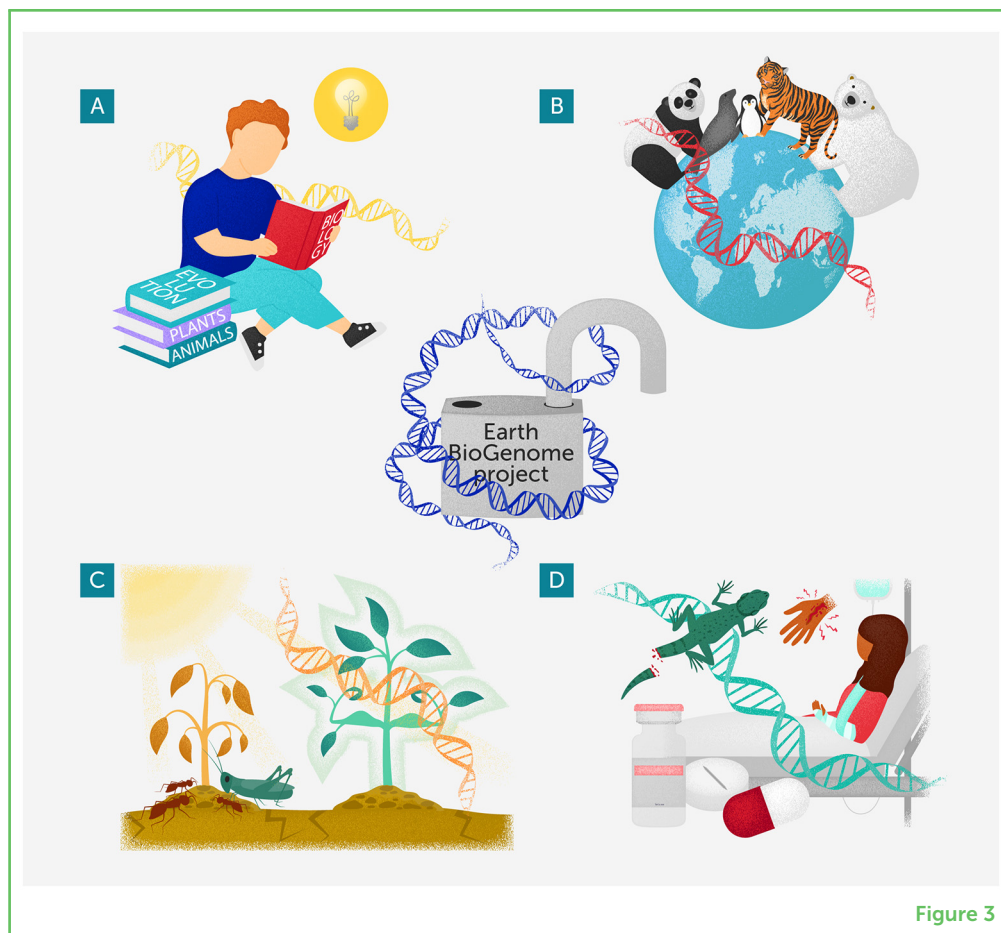


Figure 3

impact. By building a genetic library of life that is open to researchers everywhere, EBP is creating a resource that could drive discoveries for generations—helping scientists understand, protect, and use the world's biodiversity in ways we can only begin to imagine.

ACKNOWLEDGMENTS

MB: Wellcome Trust grants 206194 and 218328, Gordon and Betty Moore Foundation grant CBMF8897, Biodiversity Genomics Europe Project funded by Horizon Europe under the Biodiversity, Circular Economy and Environment sectors (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. HL: Arizona State University, University of California, Davis, Monell Foundation.

AI TOOL STATEMENT

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

ORIGINAL SOURCE ARTICLE

Blaxter, M., Lewin, H. A., DiPalma, F., Challis, R., da Silva, M., Durbin, R., et al. 2025. The Earth BioGenome Project Phase II: illuminating the eukaryotic tree of life. *Front. Sci.* 3:1514835. doi: 10.3389/fsci.2025.1514835

REFERENCES

1. Wiens, J. J. 2023. How many species are there on Earth? Progress and problems. *PLoS Biol.* 21:e3002388. doi: 10.1371/journal.pbio.3002388
2. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., and Worm, B. 2011. How many species are there on Earth and in the ocean? *PLoS Biol.* 9:e1001127. doi: 10.1371/journal.pbio.1001127
3. Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., et al. 2018. Earth BioGenome project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA* 115:4325–33. doi: 10.1073/pnas.1720115115
4. Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M, et al. 2022. The Earth BioGenome Project 2020: starting the clock. *Proc. Natl. Acad. Sci. USA* 119:e2115635118. doi: 10.1073/pnas.2115635118
5. Sherkow, J. S., Barker, K. B., Braverman, I., Cook-Deegan, R., Durbin, R., Easter, C. L., et al. 2022. Ethical, legal, and social issues in the Earth BioGenome Project. *Proc. Natl. Acad. Sci. USA*. 119:e2115859119. doi: 10.1073/pnas.2115859119
6. Mc Cartney, A. M., Anderson, J., Liggins, L., Hudson, M. L., Anderson, M. Z., TeAika, B., et al. 2022. Balancing openness with Indigenous data sovereignty: an opportunity to leave no one behind in the journey to sequence all of life. *Proc. Natl. Acad. Sci. USA* 119: e2115860119. doi: 10.1073/pnas.2115860119
7. Blaxter, M., Archibald, J. M., Childers, A. K., Coddington, J. A., Crandall, K. A., Di Palma, F., et al. 2022. Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. USA*. 119:e2115636118. doi: 10.1073/pnas.2115636118

SUBMITTED: 28 July 2025; **ACCEPTED:** 04 November 2025;

PUBLISHED ONLINE: 25 November 2025.

EDITOR: Robert T. Knight, University of California, Berkeley, United States

SCIENCE MENTORS: Natalie J. Nannas

CITATION: Blaxter M and Lewin HA (2025) The Earth Biogenome Project: Building the Ultimate Library of Life. *Front. Young Minds* 13:1674902. doi: 10.3389/frym.2025.1674902

CONFLICT OF INTEREST: The authors declare that the research was conducted in the absence of financial relationships that could be construed as a potential conflict of interest.

COPYRIGHT © 2025 Blaxter and Lewin. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

YOUNG REVIEWERS

ALLISON, AGE: 10

I am a 10 year-old 4th grader, my favorite subjects are math and science, and I love reading. I also love art and playing the piano. I love making friends and socializing with other people. I love animals and nature, and that is why I joined young reviewing in biodiversity. When I grow up I would love to be a teacher. I have lots of experience with explaining things to younger children because I have a little sister.

AUTHORS

MARK BLAXTER

Mark Blaxter is a biologist at the Wellcome Sanger Institute in Cambridge, United Kingdom, where he leads the Tree of Life programme. His research focuses on genomics, biodiversity, and evolution, with a particular interest in understanding how species are related and how their genomes can inform conservation. He has played a leading role in large-scale genome sequencing efforts, including the Darwin Tree of Life Project and the Earth BioGenome Project, which aim to read and share the complete genetic codes of all known eukaryotic species. [*mb35@sanger.ac.uk](mailto:mb35@sanger.ac.uk)

HARRIS A. LEWIN

Harris A. Lewin is a research professor at Arizona State University and Distinguished Professor Emeritus of Evolution and Ecology at the University of California, Davis. His research explores mammalian genome evolution, focusing on adaptation, speciation, and cancer. Before joining UC Davis, he spent nearly three decades at the University of Illinois at Urbana-Champaign, where he directed several major genomics centers. In 2017, he co-founded the Earth BioGenome Project and now serves as Chair of its Executive Council. Lewin is a member of the U.S. National Academy of Sciences and a recipient of the Wolf Prize in Agriculture.

