



OPEN ACCESS

EDITED BY
Frontiers in Science Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE

Paul Flicek

✉ paul.flicek@gmail.com

RECEIVED 21 August 2025

ACCEPTED 21 August 2025

PUBLISHED 04 September 2025

CITATION

Flicek P. Mapping the genomes of
Earth's interconnected biodiversity.
Front Sci (2025) 3:1690100.
doi: 10.3389/fsci.2025.1690100

COPYRIGHT

© 2025 Flicek. 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.

Mapping the genomes of Earth's interconnected biodiversity

Paul Flicek^{1,2*}

¹The Jackson Laboratory, Farmington, CT, United States, ²Department of Genetics, University of Cambridge, Cambridge, United Kingdom

KEYWORDS

DNA sequencing, reference genome, biodiversity, sequencing technology, evolution

An Editorial on the Frontiers in Science Lead Article

[The Earth BioGenome Project Phase II: illuminating the eukaryotic tree of life](#)

Key points

- The Earth BioGenome Project (EBP) has made significant progress toward its goal of sequencing the genomes of all eukaryotic species.
- The next phase of the project faces several key challenges, including developing strong regional sequencing hubs to build community relationships.
- The EBP will help drive technological breakthroughs in specimen storage, low-input genome sequencing, and computational methods that will benefit biodiversity research for decades to come.

Introduction

The Earth BioGenome Project (EBP) was launched in 2018 with a simple and compelling aim: to sequence the genomes of all eukaryotic species. The project scientists are attempting this despite not knowing the exact number of species on Earth, without a clear source of funding, and by relying on a combination of steady technological advances and the occasional dramatic breakthrough to achieve their goal. Now entering its second phase, the EBP is the type of heady scientific vision that can seem impossible until, suddenly, it is obvious that it will be done.

Earth in the Holocene Epoch is endowed with approximately 1.67 million eukaryotic species that have been named and described by taxonomists, although the actual number of species may be 10 million or more (1). A significant fraction of this biodiversity is currently threatened by habitat degradation and loss, climate change, exploitation, and other human activities. Meanwhile, biology has entered the genomic era: in the 25 years since the initial announcement of a complete draft of the human genome sequence, the ability to determine the genome sequence of any species has advanced at an almost unimaginable rate. DNA sequencing costs have dropped dramatically, from billions of United States dollars to only hundreds to sequence a single genome, while the total annual global sequencing capacity

has increased exponentially, to tens of millions of genomes. This low-cost capacity provides an effective and partly automated way to assess the vast collection of Earth's species, nearly all of which are uncharacterized at the genomic level.

In the 7 years since the EBP launch, much has changed to aid the feasibility of the project but, even with these advances, the list of challenges remains largely the same. The difference in 2025 is that the challenges are better understood and the updated plans for EBP Phase II are based on significant real-world experience. Described in the *Frontiers in Science* lead article by Blaxter et al. (2), these plans include adaptive sampling, increased genome quality, and expanded global partnerships.

Equitable global partnerships to power EBP Phase II

As more species are sequenced, scientific opportunities in biodiversity (3), conservation (4), and large-scale comparative genomics (5) are emerging rapidly. The newly revised Phase II of the EBP takes advantage of these experiences to refine its sequencing priorities while keeping a clear eye on the practicalities of making the project work. Indeed, the major challenges of the EBP do not involve the number of genomes for sequencing or the amount of data to be stored—the number of human genomes sequenced is far larger than the 1.67 million species that the EBP is attempting to corral. Instead, the issues lie with sample logistics. Key challenges remain in the collection, storage, DNA extraction, and management of biological assets, all while respecting the rights and interests of the communities where these species live. The EBP plans to address these challenges by collecting and biobanking 300,000 samples in Phase II, twice as many as they hope to sequence. They will leverage community-driven regional hubs and collect samples in opportunistic and efficient ways rather than adhering to a strict species list to achieve uniform phylogenetic coverage. The unsequenced Phase II samples will then be used to jump-start the future EBP Phase III.

At the heart of the new EBP plan is a globally distributed network of 25 hubs that will collect samples, sequence genomes, and produce assemblies. This is an ambitious and vitally important outcome of the project. Partnerships are critical to the success of the EBP yet pose specific challenges—unlike other aspects of the project, the effective parts of partnerships cannot be collectively automated and scaled. Cooperative relationships will need to be built and nurtured individually to fulfill their potential both within the project and in the global scientific community. Fortunately, the EBP is well placed to create this organization. Its “network of networks” governance structure is readily adaptable to increase the distribution of sites, capacity, and expertise across the world, and the Phase II call to establish a Foundational Impact Fund will help power organizational development.

Although necessary for the EBP's success, the partnership plan should be viewed through a broader lens as a concrete step toward building a durable global genomics capacity, primed in part by participation in this project. This is also the area where the EBP

appears to have the most work to do. By expanding beyond sample collection and data generation, scientists and institutes in the Global South can increasingly lead cutting-edge biodiversity research. As yet, however, it seems that relatively few of these partners have been defined, and the community of scientists authoring the EBP Phase II article mostly represents the Global North (2).

Nurturing breakthroughs

Ambitious projects often arise from and lead to technical breakthroughs in a virtuous cycle. While falling DNA sequencing costs inspired the initial EBP vision, it is the more recent advances in sequencing technology that mean the Phase II outputs will be both scientifically superior and less expensive than previously anticipated (2). Many, if not most, genomes will be sequenced to “reference quality”, with an increasing number now expected to be at or near telomere-to-telomere quality. This is a welcome advance compared with the previous plan that envisioned many more “draft” genomes. Having reference-quality genomes that are more complete with longer contiguous segments means that the genomes sequenced by the EBP will be valuable for a wider breadth of inquiry and reduce the need for re-sequencing with better technologies in the future. For example, research into genomic repeats (6), highly variable regions (7), gene family evolution (8), and population structures (9) are all dramatically improved with more complete and contiguous genome sequences.

Although many technical breakthroughs are serendipitous, others respond to obvious pressing needs. A potential advancement that would revolutionize the EBP and other environmental and conservation genomics projects is the development of ambient temperature specimen preservation techniques, alleviating the need for cold chain or live transport solutions. Efforts in this area are continuing at pace (10). A further technical challenge is the creation of reference genomes from unicellular and tiny species that contain far less DNA than required for current DNA extraction and sequencing protocols. Sequencing workflows with miniscule amounts of DNA will have dramatic benefits across many biological domains.

Finally, when the layers of the EBP are peeled back, the project is supported by a massive computational infrastructure that affects almost every part of the project and has its own set of ambitions, opportunities, and challenges. The major facets of informatics include (i) tracking and metadata management—essentially a large-scale distributed laboratory information management system (LIMS) project, (ii) production informatics for genome assembly and quality control, (iii) primary analysis associated with genome annotation and alignments, (iv) large-scale comparative analysis methods driving biological discovery, and (v) data management tools for access and sharing. Each of these areas has a powerful repertoire of tools adequate for Phase I. However, if the EBP is to meet its ultimate goals, advancement is required—either through significant development or complete replacement with tools not yet invented. Some parts of the computational stack will no doubt benefit from the headline-grabbing advances in artificial

intelligence, but changes in data standards and new algorithms (which are consistently finding more usable information in the evolutionary relationships between species) are likely to have the most impact. Importantly, the EBP has recognized that this computational power should be as green as possible (2).

Mapping the genomics of our Earth will give us insight into the DNA-based data system that interconnects life and the planet. Its value is immeasurable, and its secrets are truly everything we have.

Statements

Author contributions

PF: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author declared that no financial support was received for this work and/or its publication.

References

1. Scheffers BR, Joppa LN, Pimm SL, Laurance WF. What we know and don't know about Earth's missing biodiversity. *Trends Ecol Evol* (2012) 27(9):501–10. doi: 10.1016/j.tree.2012.05.008
2. Blaxter M, Lewin HA, Di Palma F, Challis R, da Silva M, Durbin R, et al. The Earth BioGenome Project Phase II: illuminating the eukaryotic tree of life. *Front Sci* (2025) 3:1514835. doi: 10.3389/fsci.2025.1514835
3. Iwaszkiewicz-Eggebrecht E, Goodsell RM, Bengsson B-Å, Mutanen M, Klinth M, van Dijk LJA, et al. High-throughput biodiversity surveying sheds new light on the brightest of insect taxa. *Proc R Soc B* (2025) 292(2046):20242974. doi: 10.1098/rspb.2024.2974
4. O'Connell LA, Rodríguez A, Kosch TA, Kwon TA, Bolsoni Lourenço L, Ortega-Andrade HM, et al. Genomics: using genomics approaches in amphibian conservation. In: Wren S, Borzée A, Marcec-Greaves R, Angulo A, editors. *Amphibian conservation action plan: a status review and roadmap for global amphibian conservation*. Gland: International Union for the Conservation of Nature (2024). doi: 10.2305/QWVH2717
5. Wright CJ, Stevens L, Mackintosh A, Lawniczak M, Blaxter M. Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* (2024) 8:777–90. doi: 10.1038/s41559-024-02329-4
6. Sproul JS, Hotaling S, Heckenhauer J, Powell A, Marshall D, Larracuent AM, et al. Analysis of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. *Genome Res* (2023) 33(10):1708–17. doi: 10.1101/gr.277387.122
7. Wold J, Koepfli K-P, Galla SJ, Eccles D, Hogg CJ, Le Lec MF, et al. Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol Ecol* (2021) 30:5949–65. doi: 10.1111/mec.16141
8. Chen W, Wang X, Sun J, Wang X, Zhu Z, Ayhan DH, et al. Two telomere-to-telomere gapless genomes reveal insights into *Capsicum* evolution and capsaicinoid biosynthesis. *Nat Commun* (2024) 15(1):4295. doi: 10.1038/s41467-024-48643-0
9. Secomandi S, Gallo GR, Rossi R, Fernandes CR, Jarvis ED, Bonisoli-Alquati A. Pangenome graphs and their applications in biodiversity genomics. *Nat Genet* (2025) 57:13–26. doi: 10.1038/s41588-024-02029-6
10. Teltscher F, Korlević P, Sheerin E, Makunin A, Lawniczak MKN. Breaking the cold chain: solutions for room temperature preservation of mosquitoes leading to high quality reference genomes. *bioRxiv* [preprint] (2025). doi: 10.1101/2025.07.03.662936

Conflict of interest

The author is a member of the steering committee of Rodent 2K, which is an Affiliated Project of the Earth BioGenome Project (see <https://www.earthbiogenome.org/affiliated-project-networks> for more information). In the past, the author was also part of other related projects, including the Darwin Tree of Life project.

Generative AI statement

The author declared that no generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.