


Forum

A global initiative for ecological and evolutionary hologenomics

Aoife Leonard¹,
Earth Hologenome Initiative
Consortium² and
Antton Alberdi ^{1,*}



The Earth Hologenome Initiative (EHI) is a global collaboration to generate and analyse hologenomic data from wild animals and associated microorganisms using standardised methodologies underpinned by open and inclusive research principles. Initially focused on vertebrates, it aims to re-examine ecological and evolutionary questions by studying host–microbiota interactions from a systemic perspective.

Towards a deeper understanding of hologenomes

Most animals on Earth live in intimate association with microorganisms, which provide their host with an array of functional capabilities that can influence physiological processes and individual fitness [1]. Hologenomics comprises the joint analysis of host genomes, **microbial metagenomes** (see [Glossary](#)), and their combined functional attributes, to unravel the underpinnings of such host–microbiota interactions [2]. To date, hologenomic information remains scarce, dominated by **amplicon sequencing data** derived from humans, model organisms, and farm/domestic animals, with limited representation of wild animals ([Figure 1A](#) and see also the supplemental information online). However, there is an ongoing rise in both the number of animal reference genomes and catalogues of functionally

annotated prokaryotic genomes [3,4]. In addition, powerful tools for **shotgun sequencing**-based data generation and analysis have also recently been developed [5–7]. These significant advancements now enable the comprehensive exploration of (meta)genome-wide information in wild host–microbiota systems or holobionts [8] ([Figure 1B](#)). Ecological niche shift, range expansion, invasion, domestication, disease transmission, and resistance development, extinction, and adaptation to environmental change are but a few of the eco-evolutionary processes that can be revisited from the integrative perspective of host–microbiota systems [9].

The complexity, spatiotemporal variability, and scale-dependent nature of eco-evolutionary processes require a large and coordinated effort to systematically generate comparable host–microbiota data from a taxonomically, ontogenically, functionally, environmentally, and geographically representative number of animal species worldwide. In response to this need, the EHI (www.earthhologenome.org) was established to promote, facilitate, coordinate, and standardise hologenomic research on wild organisms. Focused on terrestrial vertebrates in its initial phase, the EHI encompasses projects with diverse study designs and goals following standardised and open-access sample collection and preservation, data generation, and data management criteria. By establishing a robust consortium and implementing an open, transparent database, we strive to actively cultivate collaboration and knowledge transfer, improve data generation quality and efficiency, and yield beyond state-of-the-art scientific outputs.

Goals and anticipated outcomes

The EHI aims to achieve five strategic goals that will benefit the scientific community in terms of scientific outputs, methodological development, infrastructure availability, and networking, all grounded in solid ethical principles ([Figure 1C](#)).

Glossary

Amplicon sequencing data: genetic information obtained through sequencing of a selectively amplified DNA region.

CARE principles: research principles that consider the rights and interests of Indigenous peoples to ensure collective benefit, authority to control, responsibility, and ethics in working with research data.

FAIR principles: research principles that aim to facilitate data sharing by making research data findable, accessible, interoperable, and reusable.

Microbial metagenome: collective genetic material derived from a microbial community in a particular sample.

Reference genome: a representative version of a genome that serves as a template for comparing and analysing the genetic material of individuals within a species.

Shotgun sequencing data: genetic information obtained through sequencing total DNA.

Terabase: genomic data of a magnitude of one trillion (10^{12}) nucleotide bases.

Science

The main scientific objective of the EHI is to investigate eco-evolutionary processes through a host–microbiota lens, by jointly analysing whole-genome (re)sequencing data of wild animal hosts and genome-resolved metagenomic data of associated microbial communities, encompassing large spatiotemporal scales, including host genetic structure, host phenotypic traits, functional microbiome properties, and environmental features. Accounting for these factors increases our capacity to generate higher-order conclusions about animal–microbiota interactions, and their contributions to eco-evolutionary processes. Some examples include microbial contributions to local adaptation, adaptive radiation of host animals, and host contributions to functional diversification and spatial distribution of microorganisms. Collectively, these findings can provide valuable insights for policymakers and managers, empowering the development and implementation of microbiome-aware wildlife conservation and management strategies [10].

Methodology

The primary methodological objective of the EHI is to develop and implement

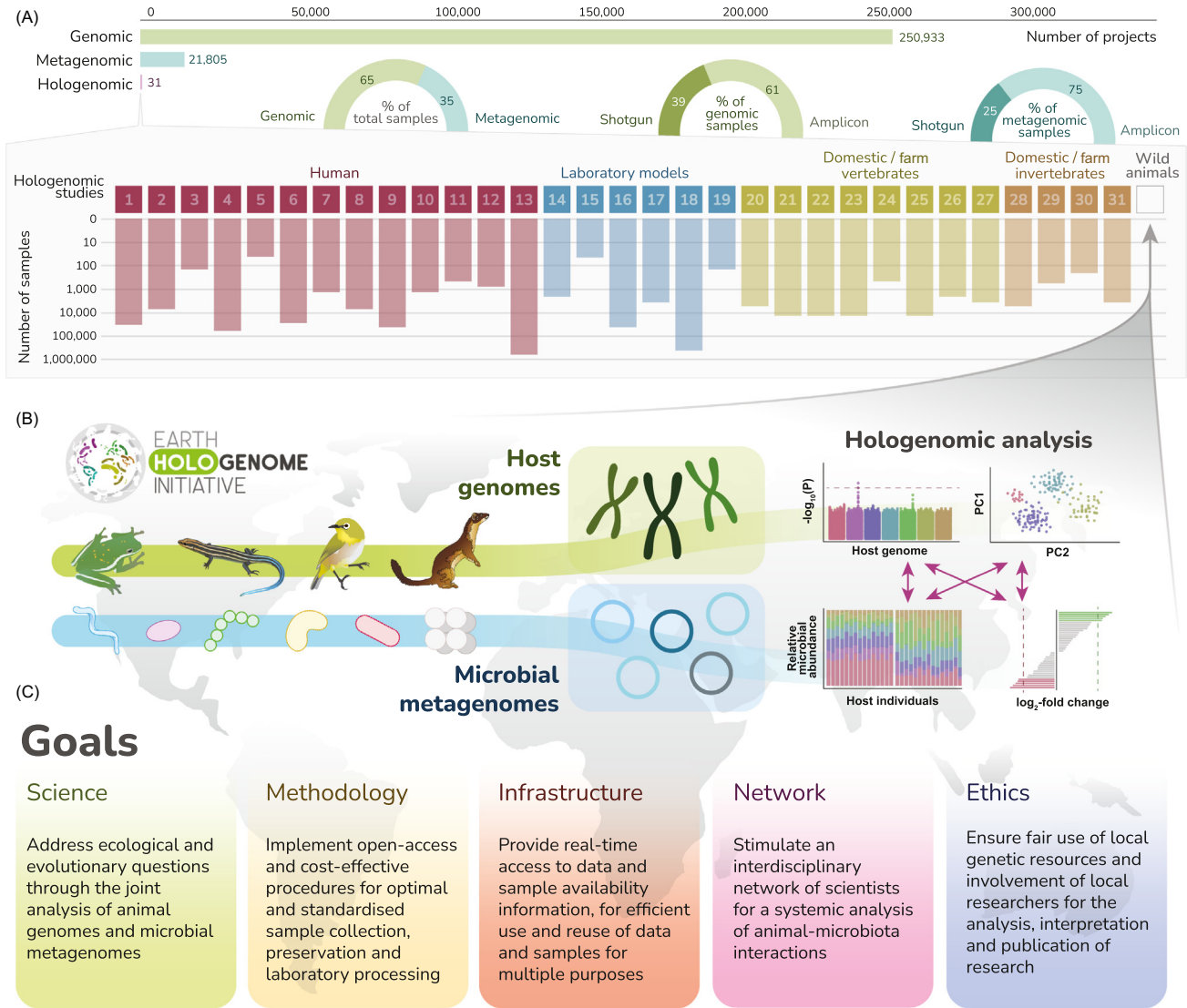


Figure 1. Overview of hologenomic datasets available at public INSDC databases and the goals of the Earth Hologenome Initiative (EHI). (A) A programmatic search (details in supplemental information online) for linked animal genomic and microbial metagenomic information in international research data archives yielded 31 research projects, none of them based on wild animals. (B) The EHI was conceived to fill this knowledge gap, by generating paired animal genomic and microbial metagenomic data in wild animals worldwide within an open and collaborative framework. (C) Overview of the five strategic goals of the EHI. INSDC stands for International Nucleotide Sequence Database Collaboration, and interconnects the DDBJ (DNA Data Bank of Japan), ENA (European Nucleotide Archive), and NCBI (National Center for Biotechnology Information, USA) databases.

standardised sampling, preservation, and laboratory methods based on open resources and knowledge. This effort aims to address the limited comparability and reproducibility of microbiome studies to date, resulting from the high sensitivity of microbiome analyses to cross-contamination and variations in sample

collection, preservation, and data generation methods [11]. All procedures are openly shared with the research community through the EHI website and developed so that they can be reproduced, automated, and deployed in different laboratories, with maximum cost-effectiveness.

Infrastructure

The main infrastructure goal of the EHI is to pioneer a novel strategy for transparent and reproducible research. This entails providing other researchers with direct access to real-time information on project progress, sample processing updates, and data generation status. In addition,

we aim to maximise outcomes from fieldwork efforts by making biological samples and DNA extracts available for other researchers, whenever allowed by legislation. This will aid multidisciplinary research investigations and help to decrease costly field sampling and disturbance of wild animal populations while fostering collaboration among researchers.

Network

Disentangling the complexity of host-microbiota interactions in wild animals requires joint contributions of researchers encompassing multiple disciplines,

geographical regions, and career development stages. This network is essential to make the most informed decisions regarding the scientific questions to be addressed, collect samples worldwide, and analyse and interpret the generated results. The EHI network provides an opportunity to build international bridges to foster multi-directional knowledge transfer among researchers interested in ecological and evolutionary hologenomics.

Ethics

Characterising the diversity, distribution, and structure of animal-microbiota interactions

across Earth is only achievable through the inclusive collaboration of diverse researchers and communities around the world, and responsible sharing of local genetic resources and knowledge. As such, the EHI is governed by open science, embracing **CARE** and **FAIR** data governance principles [12,13], and complying with all international, national, and regional regulations stemming from the United Nations' Convention on Biological Diversity (www.cbd.int).

Progress status and future steps

Conceived in 2020, the EHI is progressing through three phases (Figure 2A). The initial

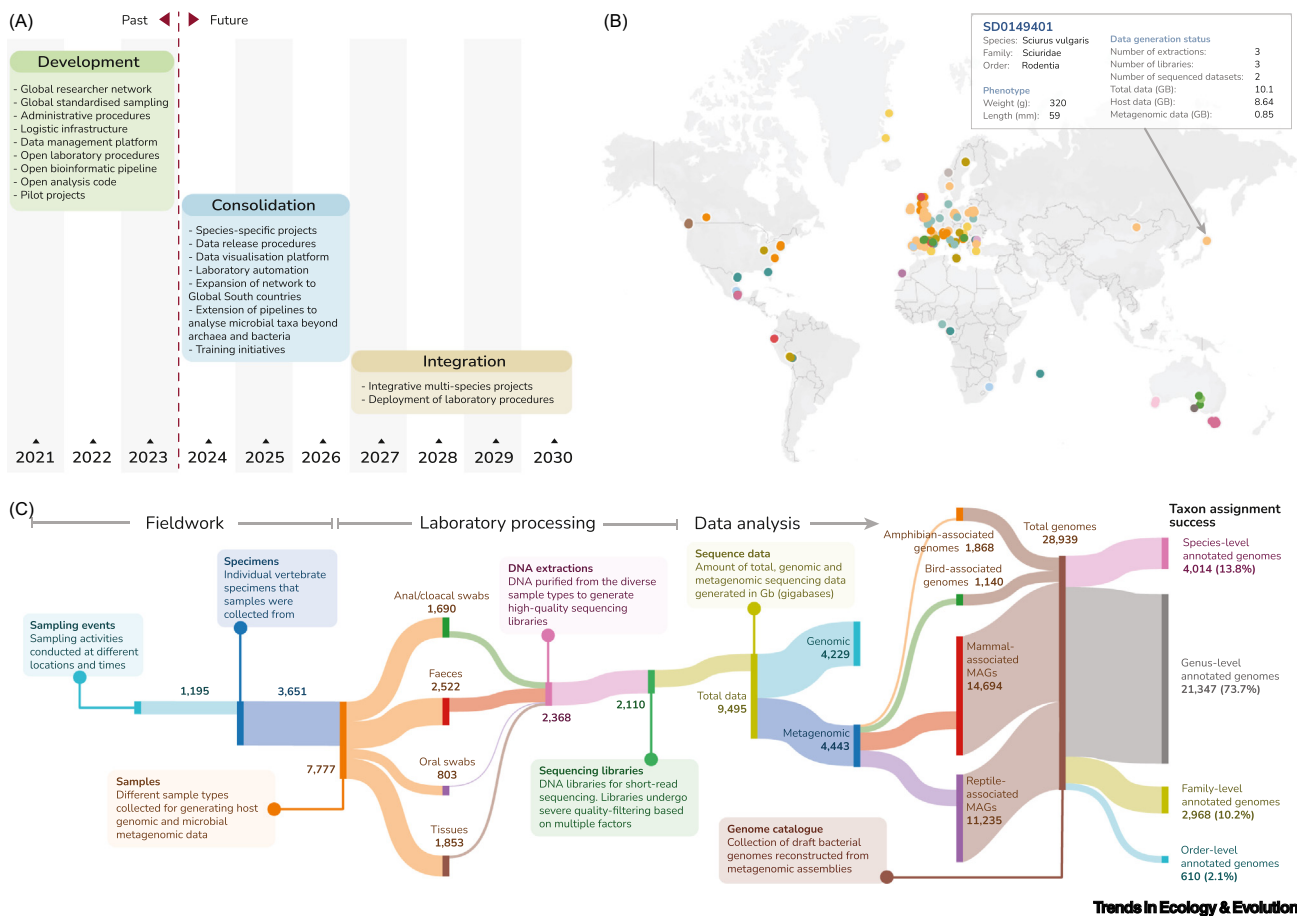


Figure 2. Overview of sample collection and data processing of the Earth Hologenome Initiative (EHI) in early 2024. (A) A 10-year roadmap indicating the main tasks associated with the three phases of the EHI. (B) World map with the geographical locations of sampled animals as shown on the www.earthhologenome.org website. Only samples stored in the EHI biobank and indexed in the database are displayed. The map reflects the geographical bias of the developmental phase of the EHI, which was primarily limited to Europe while administrative and logistic aspects of worldwide sampling were being addressed. (C) Sankey diagram displaying an overview of sample and data amounts throughout the EHI workflow. Abbreviation: MAG, metagenome-assembled genome.

developmental phase (2020–2023) focused on building the administrative, logistic, and technical framework, and generating preliminary data (see supplemental information online). This resulted in the distribution of 170 standardised EHI sampling kits among 86 participants, which yielded a cumulative submission of over 8500 samples from 244 distinct vertebrate species across the world (Figure 2B). The generation of hologenomic information from these samples has already produced >10 **terabases** of sequencing data, with an even split between host genomic and metagenomic data. **Reference genomes** with variable assembly levels were only available for 19% of the analysed species (see supplemental information online), which highlights the ongoing need to coordinate information exchange with reference genome-generating consortia [14]. The other half of the generated data, namely the metagenomic fraction, yielded nearly 30 000 draft bacterial genomes (Figure 2C). Fewer than 15% of these genomes were annotated at the species level, underscoring the fact that wild vertebrates harbour an enormous quantity of bacteria heretofore undescribed by genomic science. The knowledge generated in this preliminary phase has been crucial in shaping the design and guiding future steps. During the next consolidation phase (2024–2026) we are initiating targeted studies concentrating on one or a few focal taxa, following study-specific designs. The subsequent integration step (2027–2029) will entail the utilisation or reutilisation of hologenomic data derived from diverse individuals, populations, and species, with the goal of unravelling global patterns and addressing overarching biological questions at the forefront of science.

The EHI boldly steps into the realm of a largely unexplored scientific territory, with significant work ahead to achieve a comprehensive overview of host–microbiota systems or holobionts across taxa and biomes. With over 15 active projects and

more than 60 participating agreements in place, a rich scientific output is anticipated. Furthermore, the initiative's transparent procedures and collaborative ethos are likely to attract a wave of innovative researchers equipped with fresh perspectives and advanced analytical skills. This collective drive is poised to propel pioneering hologenomic research to deepen our insights into the intricate interconnectedness of Earth's life forms, while tackling challenges at the frontiers of the global research, biodiversity, and sustainability agendas [15].

Earth Hologenome Initiative Consortium

Javier Abalos, Titus Adhola, Windsor Aguirre, Ostaizka Aizpurua, Shahzad Ali, Franco Andreone, Fabien Aubret, Hefer D. Ávila-Palma, Lizbeth Fabiola Bautista Alcantara, Juan F. Beltrán, Rachel Berg, Thomas Bjørneboe Berg, Sandro Bertolino, Daniel T. Blumstein, Bazartseren Boldgiv, Zbigniew Borowski, Jean Philippe Boubli, Sven Büchner, Carlos Cabido, Carlos Camacho, Juan C. Chaparro, Anne Charmantier, Guillermo D'Elía, Luis P. da Silva, Bo Dalsgaard, Christophe de Franceschi, Ferran de la Cruz, Noé U. de la Sancha, Mathieu Denoël, Raphael Eisenhofer, Nathalie Feiner, Joana M. Fernandes, Jordi Figuerola, Leonida Fusani, Laura Gangoso, Roberto García-Roa, Stefania Gasperini, Nanna Gaun, M. Thomas, P. Gilbert, Ivan Gomez-Mestre, Gary R. Graves, Jim Groombridge, Emilie A. Hardouin, Mauricio Hernández, L. Gerardo Herrera M., Kathy Hodder, Peter A. Hosner, Natali Hurtado, Javier Juste, Sarah C. L. Knowles, Kevin D. Kohl, Carmi Korine, Yurii V. Kornilev, Stephanie Kramer-Schadt, Xavier Lambin, Ella Z. Lattenkamp, Jonas Lauritsen, Guoliang Li, Celeste María López, Adrià López Baucells, Tali Magory Cohen, Emiliano Manzo, Mélanie Marteau, Lynn B. Martin, Garazi Martin Bideguren, Maria Vittoria Mazzamuto, Ara Monadjem, Pirmin Nietlisbach, Daniel Bilyeli Øksnebjerg, Jasmin G. Packer, Michael L. Pepke, Juan

M. Peralta-Sánchez, Adrián Perdomo, Guillem Pérez i de Lanuza, Carlotta Pietroni, Mathilde Poyet, Carsten Rahbek, Thiruvarangan Ramaraj, Yuliaxis Ramayo-Caldas, Orly Razgour, Hugo Rebelo, Jiří Reif, Rebecca Rimbach, Ricardo Rocha, Rita G. Rocha, Carlos Rodríguez Fernandes, Claudia Romeo, Suvu Ruuskanen, Scott K. Sakaluk, Francesca Santicchia, Tom Sarraude, Rune Sørås, Martina Spada, Michael A. Steele, Mason R. Stothart, Emina Sunje, Alex O. Sutton, Marta Szulkin, Yu Takahata, Charles F. Thompson, Kasper Thorup, Geizecler Tomazetto, Laura Torrent, Nia Toshkova, Claudia Tranquillo, Manfredo Alejandro Turcios-Casco, Tobias Uller, Isolde van Riemsdijk, Guillermo Velo-Antón, Goedele Verbeylen, Elin Videvall, Christian C. Voigt, Lucas A. Wauters, Maren Wellenreuther, and Alexey Yanchukov.

Acknowledgments

The EHI could not be conceived without the trust and economic support provided by the Danish National Research Foundation through the grant DNRF143, and the Carlsberg Foundation through the grant CF20-0460. The authors also acknowledge the NERC (grant NE/M018660/1), and University of Exeter Europe Network Fund for the economic support provided to ongoing projects embedded within the EHI. M. Denoël is a Research Director at F.R.S.-FNRS (Fonds de la Recherche Scientifique). C. Rodríguez Fernandes thanks the support of cE3c through an assistant researcher contract (FCiência.ID contract #366) and FCT (Fundação para a Ciência e a Tecnologia) for Portuguese National Funds attributed to cE3c within the projects UIDB/00329/2020, UIDP/00329/2020, and LA/P/0121/2020. We also thank all the researchers who participated in the fieldwork, as well as the support people who facilitated the administrative paperwork, with a special mention for the Head of the Globe Institute Anders J. Hansen, who has signed dozens of participation agreements. Finally, we acknowledge that some of the elements included in the figures have been obtained and modified from the Integration and Application Network, University of Maryland—Center for Environmental Science (<https://ian.umces.edu/media-library/symbols>).

Declaration of interests

The authors have no interests to declare.

Supplemental information

Supplemental information to this article can be found online at <https://doi.org/10.1016/j.tree.2024.03.005>.

¹Center for Evolutionary Hologenomics (CEH), Globe Institute, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen, Denmark

²Earth Hologenome Initiative Consortium authors are listed in the back matter and their affiliations are available in the supplemental information online.

*Correspondence:

antton.alberdi@sund.ku.dk (A. Alberdi).

<https://doi.org/10.1016/j.tree.2024.03.005>

© 2024 Elsevier Ltd. All rights reserved.

References

- McFall-Ngai, M. *et al.* (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U. S. A.* 110, 3229–3236
- Alberdi, A. *et al.* (2022) Disentangling host-microbiota complexity through hologenomics. *Nat. Rev. Genet.* 23, 281–297
- Parks, D.H. *et al.* (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794
- Rhie, A. *et al.* (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746
- Shaffer, M. *et al.* (2020) DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* 48, 8883–8900
- Olm, M.R. *et al.* (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868
- Uritskiy, G.V. *et al.* (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, 158
- Bordenstein, S.R. and Theis, K.R. (2015) Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* 13, e1002226
- Hird, S.M. (2017) Evolutionary biology needs wild microbiomes. *Front. Microbiol.* 8, 725
- West, A.G. *et al.* (2019) The microbiome in threatened species conservation. *Biol. Conserv.* 229, 85–98
- Aizpurua, O. *et al.* (2023) Field and laboratory guidelines for reliable bioinformatic and statistical analysis of bacterial shotgun metagenomic data. *Crit. Rev. Biotechnol.* 1–19
- Wilkinson, M.D. *et al.* (2019) Addendum: the FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 6, 6
- Jennings, L. *et al.* (2023) Applying the “CARE Principles for Indigenous Data Governance” to ecology and biodiversity research. *Nat. Ecol. Evol.* 7, 1547–1551
- Lewin, H.A. *et al.* (2018) Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4325–4333
- Peixoto, R.S. *et al.* (2022) Harnessing the microbiome to prevent global biodiversity loss. *Nat. Microbiol.* 7, 1726–1735