Darwin Tree of Life Phase I

Impact and Outcomes

Darwin TREE of LIFE

Royal Botanic Garden Edinburgh
CABI
EMBL-EBI
wellcome sanger institute
Wytham WOODS
the natural place for science
UNIVERSITY OF CAMBRIDGE
Royal Botanic Gardens Kew
Earlham Institute
UNIVERSITY OF OXFORD
Natural History Museum
THE UNIVERSITY of EDINBURGH
Marine Biological Association
ZSL Institute of Zoology
UCD DUBLIN

Contents

This report was prepared by Sarah Hull of the Wellcome Sanger Institute for the Darwin Tree of Life Consortium partners to be shared and used by those institutions, their stakeholders and their funders. Unauthorized use and/or duplication of this material without express permission is prohibited.

Cover Image

"Thistle" the red deer, *Cervus elaphus,* and one of her calves. Samples from Thistle were used to produce the red deer reference genome for the Darwin Tree of Life Project in 2021.
Image credits: John Fletcher, Thistle's owner and caretaker.

Our Story So Far

The Darwin Tree of Life Project is established as the **global pioneer** in biodiversity genomics, and is generating high-quality reference genomes for Britain and Ireland's eukaryotic species at unprecedented scale. The field of biodiversity genomics has been revolutionized through adoption of the Darwin Tree of Life Project's methodological leadership and innovative tools.

As the largest contributor to the Earth BioGenome Project, the Project has released **genomes for over 1,700**[1] **species** representing all major taxonomic groups. At a **citation rate of 80%**, and with **over 20,000 genome downloads**, these genomes are a foundation for research worldwide.

The power of this work is greater than the genomic data resource alone; software tools and pipelines developed by the Project have been **adopted globally**, protocols have been downloaded **over 33,000 times**, and the Project has trained **more than 2,000 scientists from over 70 countries**.

The reference genomes are all shared freely with the global community, and have already been used to address critical challenges in biodiversity conservation, food security, and environmental protection. The Project's **"sequence locally, think globally"** approach ensures this British and Irish initiative delivers cascading scientific, economic and social benefits worldwide, providing for the first time a critical genomic foundation for understanding and preserving biodiversity during this period of rapid environmental change.

> ***[Darwin Tree of Life] genome data will be foundational for biological research for the century to come***[2]

---

[1] Species count and all other metrics are accurate as of 1st March 2025. For up to date figures, see https://portal.darwintreeoflife.org/ or contact@darwintreeoflife.org

[2] The Darwin Tree of Life Project Consortium. 2022. "Sequence locally, think globally: The Darwin Tree of Life Project." Proc. Natl. Acad. Sci. U.S.A. 119, no. 4 (November): e21156422118. https://doi.org/10.1073/pnas.2115642118

The Darwin Tree of Life Project

The vision of the Darwin Tree of Life Project is a future where our understanding of the planet and of ourselves is transformed for the better by our knowledge of genome sequences across the entire tree of life. Where conservation decisions are informed by the sensitivity of genomic technologies, where agricultural innovation is shaped by a molecular view of the entire farm ecosystem, and where drug development is guided by the power of nature's inherent variation.

The Project launched in 2019 to establish the critical partnerships, processes, tools, and techniques for sequencing reference genomes from all types of eukaryotic (non-bacterial) species, and to deliver at sufficient scale to unlock globally novel research and innovation. At the same time, to enrich the scientific community through open access tools, techniques, and data; and to train scientists in biodiversity genomics techniques.

Phase I of this Project was focused on sequencing a representative subset of all the species in Britain and Ireland and their seas, to demonstrate for the first time the power of large scale, place-based genomic sampling. The resulting rich dataset, protocols and tools is the foundation of the scale-up phase, Phase II, which will deliver all remaining findable species in the region, estimated at 30,000 species.

By the end of 2025, Phase I will have **collected samples from 10,000 species** and released **2,500 high-quality reference genomes**.

A Project Built on Partnership

The Darwin Tree of Life Project is delivered by a multi-organisational consortium made up of world-class institutions from across the scientific landscape in Britain and Ireland.

Each partner brings critical expertise in biology or technology, whether it is ecological understanding, taxonomic identification, sample collection best practice, species handling and lab culturing, specialised molecular biology techniques, genome sequencing, genome assembly, bioinformatics and data handling, or genome annotation expertise. All partners also support the Project by engaging with the wider research community through training and communications, and by using the genomes to address cutting edge scientific research questions in their varied fields.

Project Partners

CABI: specialising in culturing and processing of key fungal species.

The Earlham Institute: specialising in sample acquisition, taxonomic identification, culturing and processing of single celled organisms, specialised genome assembly and annotation, and management of documentation such as permits and manifests.

The EMBL European Bioinformatics Institute: specialising in annotation and analysis of the Project's reference genomes, development of novel bioinformatic tools, as well as hosting the collaborative portal and collaborating on innovations in genome browser development.

The Marine Biological Association: specialising in sample acquisition and processing, and taxonomic identification of marine species.

The Natural History Museum: specialising in sample acquisition, processing and taxonomic identification for terrestrial and marine animal species, voucher archiving, genetic barcoding support and engagement with taxonomic expert groups across the UK.

The Royal Botanic Garden, Edinburgh: specialising in sample acquisition and processing, taxonomic identification of vascular plants and bryophytes, sample archiving, and building the biodiversity genomics community in Scotland.

The Royal Botanic Gardens, Kew: specialising in sample acquisition and processing, taxonomic identification and culturing of fungal species, and sample archiving.

The University College Dublin: specialising in sample acquisition, taxonomic identification and processing of Irish species, and building the biodiversity genomics community in Ireland.

The University of Cambridge: specialising in biodiverse reference genome assembly and annotation.

The University of Edinburgh: specialising in sample acquisition and processing, sample archiving, and genome assembly.

The University of Oxford: specialising in intensive, geographically focused sample acquisition and processing, with a focus on insects; and the acquisition, processing and identification of protist organisms, as well as protist genome assembly and annotation

The Wellcome Sanger Institute: specialising in production level DNA extraction, sequencing, assembly, annotation and curation, as well as R&D on laboratory and assembly techniques.

ZSL Institute of Zoology: specialising in sample acquisition, processing and taxonomic identification of passerine birds.

The Success Story of Phase I

The Darwin Tree of Life Project is developing foundational genomic resources while training the next generation of biodiversity researchers in genomics techniques. The benefits of this will be five-fold:

- **Immediate Application:** Individual genome sequences will directly inform research, decision making, and solutions in the UK and worldwide
- **Methodological Adoption:** Research teams globally will adopt the Project's protocols, tools, and pipelines to accelerate research worldwide across multiple fields
- **AI-Enhanced Research:** Computational tools will be developed using the large dataset of diverse genomes to address major biological questions
- **Product Development:** Innovative products will be developed to address challenges in health, agriculture and food security
- **Earth BioGenome Project Completion:** The comprehensive digital blueprint of Earth's biodiversity will be captured, accelerated and transformed by this work

The first benefits are already evidenced in Phase I and will be further enhanced during Phase II.

Overcoming Fundamental Barriers

The most foundational achievement of the Project is dispelling doubts about the feasibility of sequencing all life on earth. Thanks to Darwin Tree of Life Phase I, the scientific community recognizes that genomic research faces no insurmountable barriers to application across the entire tree of life.

The development of a partnership network, operational protocols, and a software suite capable of addressing all complex life forms is an extraordinary accomplishment. Phase II promises even greater scientific and societal impact as the Project scales up to complete coverage in the UK and supports adoption of these tools globally.

The Darwin Tree of Life Project is at the forefront of global biodiversity genomics, generating reference genomes at an unprecedented scale and developing cutting-edge tools, techniques, and policies for an emerging field. Like its predecessor the Human Genome Project, this initiative's full scientific impact will unfold over decades, as it forms a foundation for transformative advances across many different fields.

Pioneering Biodiversity Genomics

The Earth BioGenome Project is a moonshot coordination effort to create a comprehensive digital blueprint of life on Earth: a dataset of high-quality reference genome sequences for all 1.8 million named eukaryotic species.

As of March 2025, the Darwin Tree of Life Project has contributed 30% of all Earth BioGenome Project reference genomes.

The dataset is accompanied by innovation and tools:

- The “Genomes on a Tree” database and platform, developed to coordinate sequencing across Project partners, now coordinates reference genome sequencing worldwide[3]
- The Project’s **Open Data Policy** has contributed to the global standard for biodiversity genomics[4]
- The **sample metadata approach** and **genome annotation approach** adopted and developed by the Project have been adopted by other major biodiversity genomics initiatives around the world[5]
- **8 software tools** developed by the Project are recommended by the Earth BioGenome Project for the most common genome sequencing and assembly approach[6]
- Software developed by the Project has been **incorporated into publicly available tools from other institutes**, for example the “Foreign Contamination Screen” provided by Genbank, NIH NCBI, USA.
- Rapid **annotation pipelines** for gene finding in non-human genomes developed by the Project have led to a four-fold increase in the number of genomes annotated annually by the Ensembl genome browser.
- The Ensembl **Rapid Release Genome Browser** and new website were developed to meet the needs of this Project, and benefitting multiple global biodiversity projects and other cutting-edge work.

The **genome note** publication format, developed by the Project with Wellcome Open Research, allows rapid publication of reference genome metadata and expert contributions without requiring a scientific hypothesis for each species. Without this innovation, the sample collectors, taxonomists, genomic researchers and other scientists involved in all stages of the process would not receive authorship recognition for their skilled contribution, and the scale of the project would be unsustainable.

---

[3] Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. Challis et al. 2023. Wellcome Open Res 2023, 8:24 doi:10.12688/wellcomeopenres.18658.1, https://goat.genomehubs.org/

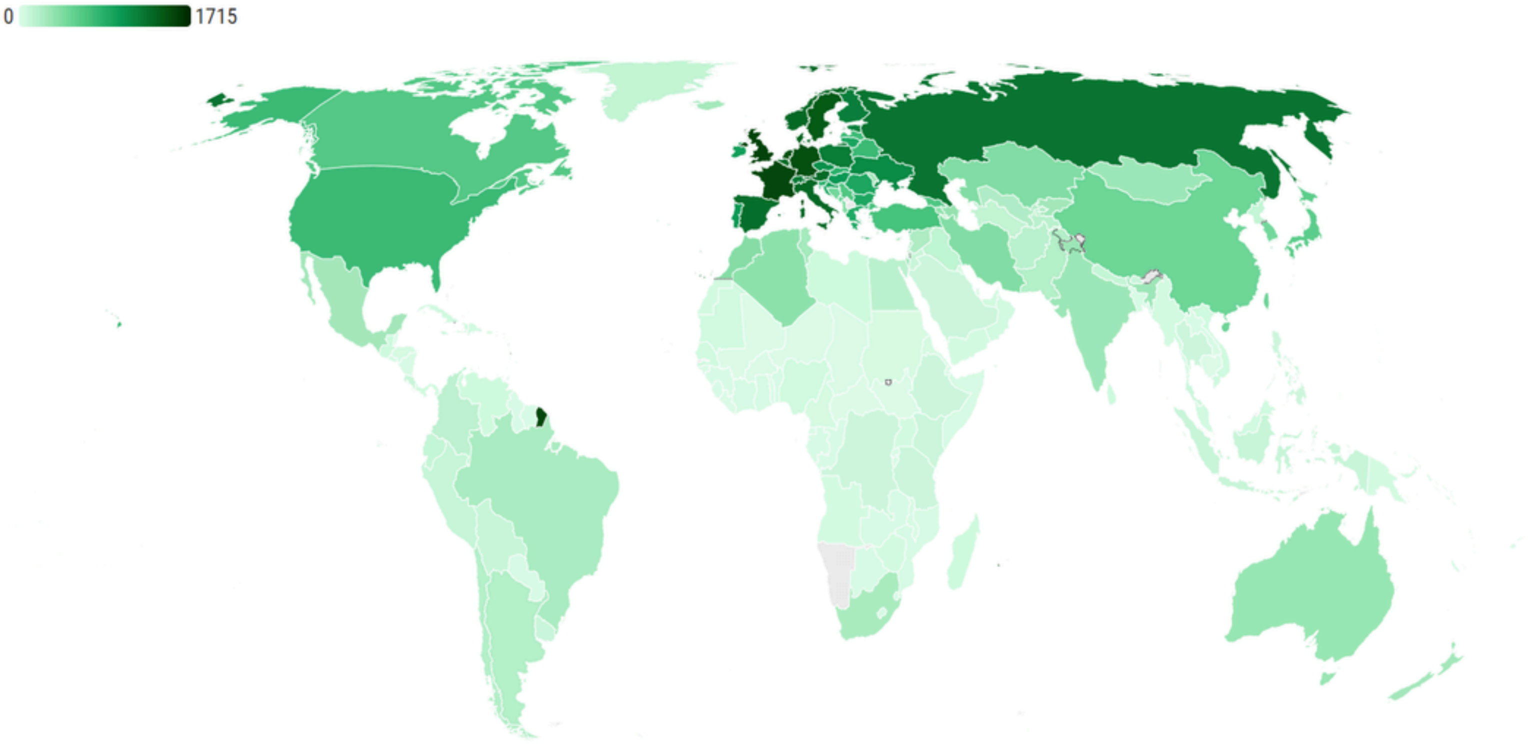
[4] Blaxter, Mark. 2024. “Darwin Tree of Life Open Data Release,” Darwin Tree of Life Data Release Policy version 1.0.4. Darwin Tree of Life Project Resources. https://www.darwintreeoflife.org/project-resources/

[5] Shaw F et al. COPO: a metadata platform for brokering FAIR data in the life sciences. F1000Research 2020, 9:495, https://doi.org/10.12688/f1000research.23889.1, https://f1000research.com/articles/9-495

[6] Earth BioGenome Project. 2024. “Earth BioGenome Project Report on Assembly Recommendations.” https://www.earthbiogenome.org/report-on-assembly-recommendations.

Sequence Locally, Think Globally

The ecosystems in Britain and Ireland and their seas are varied and complex, and most species in this region have ranges extending beyond these islands. While species are collected from Britain and Ireland and the surrounding seas, the Project's impact is global by design.



Sum of geographic ranges of species with genomes published by the Darwin Tree of Life Project

The Darwin Tree of Life genomes released prior to March 2025 can **benefit scientists in 191 countries** with high-quality reference genomes for at least one species native to their country. Reference genomes are also valuable tools for understanding related species, and Project genomes are frequently used by scientists outside Britain and Ireland as a scaffold for cost-effective generation of genomes for their local flora and fauna. This approach supports powerful comparative genomic studies in subjects such as climate change resilience and adaptability of vulnerable ecosystems.

The Project has adhered to and advocated for the principles of FAIR research throughout: that all scientific outputs should be findable, accessible, interoperable and reusable.[7] As a result, methods, raw data and analytical products are openly available and significant efforts are made to ensure they are accessible and useful to all researchers worldwide.

---

[7] Wilkinson, Mark, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific Data 3, no. 160018 (March). https://doi.org/10.1038/sdata.2016.18.

SUSTAINABLE DEVELOPMENT GOALS

Goal 14

Conserve and sustainably use the oceans, seas and marine resources for sustainable development.

Goal 15

Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss.

Scientists in The Faroe Islands, Guyana, Ecuador, and the Caribbean are adopting approaches developed by the Project to improve conservation decision making in local ecosystems under threat due to human-induced climate change. The Project shares protocols and tools as well as bespoke advice to support establishment of scientific infrastructure that will translate the Darwin Tree of Life Project achievements into advancement towards UN Sustainable Development Goals for marine and terrestrial ecosystem protection.[8]

The Project's "sequence locally, think globally" approach ensures that this British and Irish initiative delivers cascading economic and scientific benefits to global science and society, creating a lasting foundation for understanding and preserving all life on Earth.[9]

---

[8] UN DESA. 2024. The Sustainable Development Goals Report 2024 – June 2024. New York, USA: UN DESA. © UN DESA. https://unstats.un.org/sdgs/report/2024/

[9] The Darwin Tree of Life Project Consortium. 2022. "Sequence locally, think globally: The Darwin Tree of Life Project." Proc. Natl. Acad. Sci. U.S.A. 119, no. 4 (November): e21156422118. https://doi.org/10.1073/pnas.2115642118

Genomes for All

As of 1st March 2025, the Project had released **1,715 genomes** to the European Nucleotide Archive database, operating at an average release rate of 1 new genome every 9 hours. Collectively, these represent **all major taxonomic groups** in Britain and Ireland, including notoriously challenging species to handle such as protists, isopods and jellyfish; and species with challenging genomes to assemble, such as corals, birds, and certain plants.

Genome release rate is accelerating, having averaged one species genome released every 20 hours in 2024, or **37 species per month**, and projected to approach **90 per month in 2026 and beyond**.

Foundational Data

Release of genomes for over **1,715** species

Across **141** different taxonomic orders

At an average rate of **one genome every 9 hours**

Genomes collectively downloaded **over 20,000 times**

Researchers are making extensive use of these genomes. Comprehensive tracking across all databases is impossible, but Project genomes have been accessed by **20,000 genome downloads** through the one genome browser alone.

As of 1st March 2025:

- **498 publications** and preprints cite Project genomes
- **992 unique species** — representing 57% of all genomes released to date — have been cited at least once
- **80% of all genomes** that have been in the public domain for at least 1 year have been cited.

This excludes publications that omit citations for the genomes used, a relatively common flaw in large-scale comparative genomic research.[10]

---

[10] He, Chun, et al. 2024. "Large-scale genome analyses provide insights into Hymenoptera evolution." BioRxiv, (July). https://doi.org/10.1101/2024.07.01.601370, is an example which uses 131 species from the Darwin Tree of Life Project without citation.

Scientific Outcomes

Genomes collectively have already been cited **498 times**

80% of all genomes older than 1 year have already been **cited** in a research publication

The top 5 genomes alone have together been cited **134 times**

According to authorship affiliations reported on these publications, researchers from **75 countries** have directly benefited from Project genomes so far. As anticipated, the European scientific community is relatively enriched in the citing publications, but researchers are also publishing with authorship affiliations throughout Asia, Africa, and South America.

Authorship affiliation for publications citing Darwin Tree of Life genomes

Scientific Knowledge Exchange

The consortium's **75 published protocols** have collectively been **downloaded 33,150 times** by researchers worldwide, from the UK to Australia to Malaysia.

The Project has developed **11 software tools** and pipelines that have been **downloaded over 7,400 times** from open-source repositories.

An email-based query management system manned by Project scientists offers assistance to researchers seeking to use these protocols and tools, ensuring equitable access to help for everyone regardless of personal connections to the project. To date this enquiry tool has supported nearly 600 researcher enquiries.

Knowledge Exchange

Published **75 protocols** and **11 software tools**

Genome assembly pipeline downloaded **7,452 times**

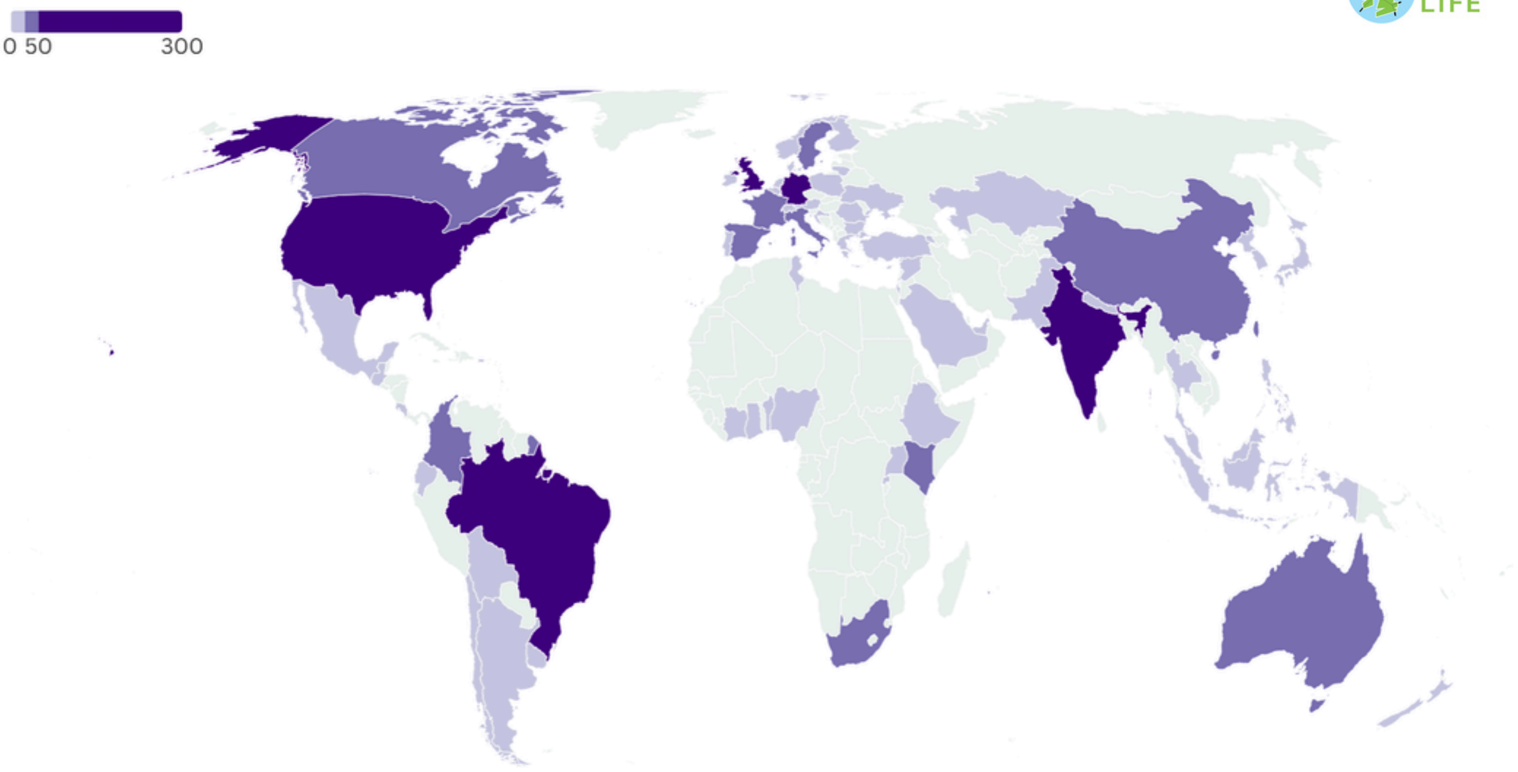
33,150 downloads across all published protocols

The Project delivers skills training to scientists worldwide, building on each partner's unique expertise, for example:

- Field-based sample collection and metadata best practices for genomics
- Lab processing of biodiverse samples in preparatino for sequencing
- Bioinformatic skills for genome assembly and curation in diverse species
- Implementation of Project-developed software tools and pipelines

to date, partners have conducted over **70 scientific training activities**, reaching more than **3,600 scientists**, including both group workshops and bespoke in-person training for individuals and organisations.

The Project's flagship training series, Biodiversity Genomics Academy, has provided free, interactive, online training workshops to over **2,000** scientists from **73 countries**. This bioinformatics training equips researchers worldwide with critical skills to apply biodiversity genomics in their local context.



Country of origin for attendees at the Biodiversity Genomics Academy in 2023-2024

Phase I Impact Case Studies

Genomes generated by the Project are already being used to address real-world challenges such as food security and biodiversity conservation. As the number of genomes released by the Project is increasing each year, the number and diversity of different ways the genomes are used will also increase.

Food Security: Sustainable Fisheries

Fishing quotas are essential for preventing overexploitation of marine resources, protecting biodiversity, and ensuring long-term food security.[11] Quotas based on accurate assessments of fish populations ensure fisheries maintain healthy stocks, ecosystems remain balanced and long-term economic stability is supported. The UK Centre for Environment, Fisheries and Aquaculture Science (Cefas) has adopted Darwin Tree of Life genome sequences to develop a novel approach for informing these quotas in the UK and **have updated the resulting policy advice to Defra.** [12]

Understanding Global Fish Populations

Using reference genomes for sole, *Solea solea,* and plaice, *Pleuronectes platessa*, Cefas scientists were able to identify separate populations within UK waters, enabling more accurate estimates of sustainable fishing rates in different regions around the UK. This work has been **shared with the International Council for the Exploration of the Sea (ICES) for potential global adoption** and is being prepared for publication.

The Darwin Tree of Life Project is prioritising collection and sequencing of key wild fish species to support sustainable fisheries management in the UK and worldwide.

Veterinary Vaccine Development: Black Scour Worm

The black scour worm, *Trichostrongylus colubriformis,* is a eukaryotic parasite responsible for billions of pounds of losses annually in the global livestock industry. The reference genome for this species was released by the Darwin Tree of Life Project in March 2024 following development of a specialized ultra-low input sequencing protocol.

Image credit: Sole meuniere by Jj Saezdeo licensed under CC BY-SA 4.0

---

[11] Marine Stewardship Council. 2025. “What is sustainable fishing?” Marine Stewardship Council. https://www.msc.org/what-we-are-doing/our-approach/what-is-sustainable-fishing.

[12] Department for the Environment, Food and Rural Affairs and Marine Management Organisation. 2024. “Quota Management Rules for England and the Crown Dependencies.” UK and England quota management rules. https://www.gov.uk/government/publications/uk-and-england-quota-management-rules.

This genome is now foundational to an international partnership between the Moredun Research Institute, the University of Glasgow, and the University of New England, Australia, to **develop a vaccine** that will combat rising rates of drug resistance in ruminant farms worldwide. This work is supported by £6 million funding from the Australian Federal Government. [13]

The protocols developed to grow, process and generate a genome sequence from this challenging organism have been adopted by genome sequencing projects worldwide, including the Earth BioGenome Project and European Reference Genome Atlas.

By developing and maintaining specialized expertise in handling even the most challenging taxa, the Darwin Tree of Life Project is enabling insights into economically significant pathogens, parasites and pests, and is paving the way for new solutions.

Species Reintroductions: Pine Hoverfly and Chequered Skipper Butterfly

The Royal Zoological Society for Scotland adopted the Darwin Tree of Life reference genome for Britain's rarest native insect, the critically endangered pine hoverfly, *Blera fallax*. This genome allowed the monitoring of genetic diversity in their captive breeding program and the assessment of genetic health in the reintroduced population.

Use of the genome sequence increased the efficiency of this reintroduction effort, supporting their successful work to establish a wild population of this important ecosystem pollinator.

The genome is also being used to determine the likely success of a potential cross-breeding program between the Scottish reintroduced population and wild Scandanavian populations.

Using a similar approach, the Project’s chequered skipper butterfly, *Carterocephalus palaemon,* genome has been used to evaluate a historic species reintroduction that took place as part of the UK Government’s Back from the Brink initiative in 2017-2021. [14] By analysing the butterfly population multiple generations after a reintroduction event, the authors were able to show that all surviving butterflies descended from just one of the two donor populations, each of which came from a different collection sites in Europe.

Image credit: Blera fallax2 by James Lindsey’s Ecology of Commanster Site licensed under CC BY-SA 2.5

[13] University of Glasgow. 2023. “A multinational £6M project to develop a novel parasite vaccine for sheep.” University of Glasgow University News. https://www.gla.ac.uk/news/archiveofnews/2023/august/headline_994489_en.html

[14] Back from the Brink. “Back from the Brink Summary Report.” Back from the Brink Resources. https://naturebftb.co.uk/downloads-category/summary-report/.

If the genome sequence had been available and used during the reintroduction effort, this genetic bias might have been detected sooner and introductions focused on the collection population that was most able to thrive in the new setting. As it stands, this evidence will be useful to improve future conservation efforts for this species.[15]

A Genomic Foundation for Conservation

Reference genome sequences are a critical tool for effective endangered species reintroductions. As multiple species genomes become available within a vulnerable ecosystem, conservationists will be able to simultaneously monitor the impact of species reintroductions on predator, prey, and competitor species – an increasingly essential capability as ecosystem-wide threats to biodiversity accumulate worldwide.

Species Conservation: Eurasian Otter

The Eurasian otter, *Lutra lutra,* genome, among the earliest published by the Project, is already improving conservation decision making for this protected British species.

This reference genome has been used by the Cardiff Otter Project to develop a bespoke approach to monitor the genetic health and history of UK otter populations. This uncovers significantly more information than traditional genetic methods that use only a small number of regions of the genome, and has uncovered previously undetected subpopulations as well as evidence of an interspecies breeding error in a major historic captive breeding program.[16]

The developers of this new approach also used the reference genome to support whole genome resequencing in otter populations around the UK, discovering differences in the impacts of hunting and pollution in the different regions.

They go on to conclude that the high quality reference genome is a **critical tool in all large-scale conservation work**, in order to avoid the significant confounding effects of ascertainment bias and genetic linkage that beset traditional genetic methods.

Image credit: European Otter (Lutra lutra) by Bernard Landgraf licensed under CC BY-SA 3.0

[15] Halford, Georgina, *et al.* 2025. “Genomic Monitoring of a Reintroduced Butterfly Uncovers Contrasting Founder Lineage Survival.” Evolutionary Applications 18, no. 2 (February). https://doi.org/10.1111/eva.70074.

[16] du Plessis, et al. 2023. “Genomics Reveals Complex Population History and Unexpected Diversity of Eurasian Otters (Lutra lutra) in Britain Relative to Genetic Methods.” Mol Biol Evol 40, no. 11 (September). https://doi.org/10.1093/molbev/msad207.

Species Conservation: Brown Trout

The brown trout, *Salmo trutta,* genome has become one of the Project's most cited resources, with **over 2,200 downloads** of its genome note. This key species is both an indicator of water quality and a source of income in local economies across Europe.

Researchers at the University of Girona in Spain used this genome to develop a testing approach that differentiates between native wild trout and hatchery-bred fish, which are frequently reared and released into the wild to boost local recreational fishing interests.

By applying this novel test to survey rivers across the Iberian Peninsula, the researchers showed that trout populations are facing significant genetic threat from these hatchery released fish, endangering long-term sustainability of these ecosystems.[17] This approach is now shared as a **monitoring tool for conservation programs** rebuilding native trout populations across Southern Europe.

Emerging Research

Some genomes released by the Darwin Tree of Life Project have triggered immediate research interest, either by revitalizing research communities, nucleating new research communities, or supporting the study of closely related organisms.

This organic interaction between released genomes and the scientific community demonstrates the value of comprehensive biodiversity sequencing beyond species with evident economic importance.

Fungal Innovation

Fungi comprise approximately 20% of the tree of life, yet 90% of all species are not known to science, and only a tiny fraction of the known species have a sequenced genome.[18] This is despite the clade being a uniquely rich source of biotechnological and medical innovation across drug molecules, antibiotics, antifungals, and fertilisers. The fungal genomes released by the Project are significantly accelerating developments in this major frontier of biological research.

---

[17] Casanova, et al. 2022. "Genomic Hatchery Introgression in Brown Trout (Salmo trutta L.): Development of a Diagnostic SNP Panel for Monitoring the Impacted Mediterranean Rivers." Genes 13, no. 2 (January): 255. https://doi.org/10.3390/genes13020255.

[18] Niskanen, Tuula, et al. 2023. "Pushing the Frontiers of Biodiversity Research: Unveiling the Global Diversity, Distribution, and Conservation of Fungi." Annual Reviews 48 (September): 149-176. https://doi.org/10.1146/annurev-environ-112621-090937.

Research currently underway at the Royal Botanic Gardens, Kew is using these genomes to identify potential novel medical compounds from fungal species with uses in Traditional Chinese Medicine (TCM) including the lion's mane, *Hericium coralloides,* and caterpillar fungus, *Cordyceps* spp. Related research on fungal families with known biotechnological applications, such as *Trichoderma* spp., will undoubtedly lead to further innovations in the agriculture and environmental biology sector.

Environmental DNA Monitoring

Environmental DNA (eDNA) approaches detect species through genetic material naturally shed into the environment. This non-invasive approach allows rapid, cost-effective biodiversity assessments across diverse ecosystems with minimal habitat disturbance. All eDNA approaches depend on knowledge of a biodiverse range of species genetic sequences. The foundational contribution of the Darwin Tree of Life Project to this emerging field, and the resulting anticipated advances in science, was recently recognised in a resource developed by **the Royal Society** in the UK.[19]

The Project's optimised genome sequencing and assembly pipeline is being deployed to address innovative research questions in eDNA and biodiversity monitoring. For example, it will be used in an upcoming major research project co-funded by the Novo Nordisk Foundation and Wellcome to analyse ancient eDNA from sediment cores. This will identify adaptations in plants and ecosystems that conferred resistance to historic changes in climate, and contribute to improvements in agricultural sustainability and crop resilience in the face of anthropogenic climate change.

Understanding Genome Packaging

In 2023, the Project released the genome sequence of the European Mistletoe, *Viscum album*. At 90 billion base pairs, this is the largest known genome from Britain and Ireland and 30 times larger than the human genome — despite being organised into only 10 chromosomes, whereas humans have 23.[20]

The technical challenges of assembling a genome of this scale led the Project to develop innovative approaches that are now widely used in the biodiversity genomics field.

Further, release of this genome catalyzed a multi-institutional partnership including the Universities of Edinburgh, Exeter, Sussex and Cambridge, as well as Queen Mary University of London and the Royal Botanic Gardens, Kew. This group is using mistletoe as a model organism to investigate fundamental questions about three-dimensional genome architecture.

---

[19] The Royal Society. 2025. "Environmental DNA: At A Glance." The Royal Society News and Resources. https://royalsociety.org/news-resources/projects/environmental-dna/

[20] Lythgoe, Luke. 2022. "2022: The year we built the biggest genome in Britain and Ireland." Darwin Tree of Life. https://www.darwintreeoflife.org/news_item/2022-the-year-we-built-the-biggest-genome-in-britain-and-ireland/.

Industrial Pest Management

The Webbing Clothes Moth, *Tineola bisselliella*, causes over $1 Billion in annual damage to the global textile industry and has destroyed countless irreplaceable historical artifacts in museums worldwide. Researchers at Boston University used the genome of a closely related species, *Tinea pellionella*, released by the Darwin Tree of Life Project in 2024, as a scaffold to develop a draft genome for the Webbing Clothes Moth.

Comparative analysis comparing this draft genome with multiple moth genomes from the Project enabled the researchers to study the molecular basis of this larvae's ability to digest cloth materials.[21] This demonstrates how Darwin Tree of Life genomes can invigorate research on pest species of all types, delivering global benefits.

Agricultural Parasite Research

The *Eimeria* family of protozoan parasites causes $10-12 billion annual losses in the global poultry industry due to enteric disease.[22] When the Darwin Tree of Life Project released the *Eimeria tenella* genome in 2022 it triggered a rapid flurry of citations in studies that are collectively shifting this research field from phenotypic to molecular assays.[23]

The recent publication of a sister species, *Eimeria* praecox, and the anticipated release of five additional *Eimeria* species genomes in 2026 will further accelerate this field of research, leading to improved treatment strategies and enhanced poultry farming efficiency worldwide.

Extinction Insights from Museum Specimens

The Project's genome for the Black-veined White butterfly, *Aporia crataegi*, a species that has been extinct in Britain since 1925, enabled cutting-edge "museomics" analysis of specimens collected between 1854 and 1924. This study revealed that there were extensive molecular indicators of the high extinction risk in the population before its disappearance.[24]

Monitoring for these indicators of genomic erosion in living butterfly patterns could proactively identify at-risk species for targeted conservation efforts in the UK. It is possible that taking a similar approach in other brances of the tree of life could reveal tell-tale genetic signs that could be monitored across the UK to improve targeting of conservation efforts and funding.

[21] Alqassar, Jasmine et al. 2024. "De Novo Genome Assembly and Annotation for the Synanthropic Webbing Clothes Moth (Tineola bisselliella): A Globally Distributed, Economically Important Pest." Genome Biology and Evolution 16, no. 12 (December): evae266. https://doi.org/10.1093/gbe/evae266.
[22] Lithgoe, Luke. 2021. "Eimeria: the chicken parasite costing farmers billions." Wellcome Sanger Institute Blog. https://sangerinstitute.blog/2021/09/23/eimeria-the-chicken-parasite-costing-farmers-billions/.
[23] Blake, Damer P. 2025. "Eimeria of chickens: the changing face of an old foe." Avian Pathology 54, no. 3 (January): 267-278. https://doi.org/10.1080/03079457.2024.2441180.
[24] Whitla, Rebecca, *et al*. 2024. "The last days of Aporia crataegi (L.) in Britain: Evaluating genomic erosion in an extirpated butterfly." Molecular Ecology 33, no. 9 (October): e17518.

The Economic Impact of the Darwin Tree of Life Project

The Darwin Tree of Life Project approach of delivering high-quality open access reference genomes at a high throughput dramatically **improves global research efficiency** and **speeds scientific discoveries**. In economic terms, a key value of the Darwin Tree of Life Project's reference genomes is their excellent reputation and their reusability.

Genomes generated on an as-needed basis by diverse research groups have a low likelihood of reuse by the scientific community, due to the risk that resulting genomes contain unknown errors. The Darwin Tree of Life Project has developed a reputation for high quality genomes with meaningful metadata and transparent protocols and processes, which underpins the Project genome's long-term reuse by the wider scientific community.

Savings to Global Science Funding

Considering only publications citing the Project's genomes released prior to March 2024, the collective saving to science funding across the citing research groups is estimated at **£42 million,** representing funding that would otherwise have been spent on generating genome assemblies for use in the citing studies. Thanks to Darwin Tree of Life, this funding was invested in addressing meaningful scientific questions rather than in reference genome generation with low reuse potential.

Support for Funding Applications

The availability of the genomes generated by the Darwin Tree of Life Project opens new opportunities for research in all countries where Project species are native. In a survey of scientists with an interest in the Project, **30% of respondents reported securing grant funding or fellowships** directly depending on the usage of Darwin Tree of Life outputs. These included early career researcher training positions and major research grants, with recipients in Britain and Ireland as well as France, Denmark, Sweden and Austria.

Further economic value is likely to be catalysed in the UK and European economies through novel research discoveries, AI models and product development building on the foundations of the Darwin Tree of Life genomes dataset.

The Societal Impacts of The Darwin Tree of Life Project

Biodiversity in local and global ecosystems is **critical for human physical health**[25], protection from infectious diseases,[26] mental health outcomes[27] and the global food chain;[28] as well as providing valuable recreation and learning opportunities at all stages of life.

The Office of National Statistics estimates UK natural ecosystem assets at £1.8 trillion, with the majority of this value derived from recreation health benefits at £489 billion, with UK broadleaf woodlands, a complex ecosystem with particularly high biodiversity, providing an estimated £1.1 billion in pollution removal value alone.[29] However, natural ecosystems in the UK and worldwide face significant threats through genetic diversity loss[30]. Recognising this, the UK government has established ambitious **National Targets to restore ecosystems and prevent species extinction,** however, current UK monitoring schemes lack long-term genetic diversity tracking[31], despite evidence that it is the most reliable early indicator of species decline prior to extinction.

The Darwin Tree of Life Project's genome sequences are critical resources that could transform how government agencies and conservation charities protect and rebuild wild ecosystems across Britain, Ireland, and Europe, particularly these critical high-value ecosystems with the most impact on our health, wellbeing, society and economy.

Conservation and Invasive Species

The Project's genomes significantly increase the number of genome sequences available for conservation and protection of species of significance in the UK.

As of March 2025, the Project has released genomes for **360 native species** designated as a priority for conservation in the UK according to the JNCC,[32] and **99 native species** under legislative protection by the UK government.

Invasive species have caused over £500 billion in economic losses to global agriculture, forestry, and fishery industries over the past 50 years.[33] Such species drive native species extinctions, reduce biodiversity, compete with beneficial native organisms, and fundamentally alter established ecosystem dynamics.

---

[25] Marselle, Melissa. 2021. "Pathways linking biodiversity to human health: A conceptual framework." Environment International 150 (May): 106420. https://doi.org/10.1016/j.envint.2021.106420.

[26] Keesing, Felicia, et al. 2010. "Impacts of biodiversity on the emergence and transmission of infectious diseases." Nature 468 (December): 647-652. https://doi.org/10.1038/nature09575.

[27] Hammoud, Ryan, et al. 2024. "Smartphone-based ecological momentary assessment reveals an incremental association between natural diversity and mental wellbeing." Scientific Reports 14, no. 7051 (April). DOI : 10.1038/s41598-024-55940-7.

[28] Renard, Delphine, and David Tilman. 2021. "Cultivate biodiversity to harvest food security and sustainability." Current Biology 31, no. 19 (October): R1154-R1158. https://doi.org/10.1016/j.cub.2021.06.082.

[29] Office for National Statistics (ONS), released 8 November 2024, ONS website, statistical bulletin, UK natural capital accounts: 2024

[30] Keck, François. 2025. "The global human impact on biodiversity." Nature 269 (March). https://doi.org/10.1038/s41586-025-08752-2.

[31] Department of Agriculture, Environment and Rural Affairs (DAERA), Scottish Government, Welsh Government, and UK Government. 2025. "Blueprint for Halting and Reversing Biodiversity Loss: the UK's National Biodiversity Strategy and Action Plan for 2030: Annex A." No. ISBN 978-1-86107-645-8. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://uk.chm-cbd.net/sites/gb/files/2025-02/NBSAP_Annex_A.pdf.

[32] Joint Nature Conservation Committee. 2024. "Terrestrial monitoring schemes." JNCC. https://jncc.gov.uk/our-work/terrestrial-monitoring-schemes/.

[33] Director-General for the Environment, European Commission. 2025. "Agriculture, forestry and fishery industries have lost hundreds of billions due to invasive alien species during the last 50 years, worldwide." "Science for Environment Policy": European Commission DG Environment News Alert Service, edited by SCU, The University of the West of England, Bristol

Reference genome sequences are critical tools for managing invasive species, and their absence is recognised as a limiting factor worldwide.[34] The Darwin Tree of Life Project focuses primarily on native British and Irish species, but a number of invasive UK species have also been sequenced, and many native UK species are invasive elsewhere in the world. For example, the Project has released genome sequences for English ivy, *Hedera helix* and Brown Trout, *Salman trutta*, which are both invasive in North America, Australia, and New Zealand.

With invasive species management increasingly turning to genomics, the Darwin Tree of Life project will increasingly **improve global invasive species management** with foundational genomic sequences. [35]

National Scientific Literacy

With approximately 1 in 120 people in the UK employed in the life sciences,[36] yet declining practical science education in schools,[37] the need for accessible, engaging cutting-edge science has never been greater. The Darwin Tree of Life Project uniquely enhances UK scientific literacy by:

- Enabling genomics teaching without introducing the subject of health disorders and disease such as genetic disorders and cancer
- Creating meaningful connections across UK 11-16 curriculum topics such as taxonomic classification, biodiversity, and genomic technologies
- Engaging non-scientists through local learned societies and conservation efforts
- Connecting activities like sample collection and DNA barcoding that can be conducted by citizen scientists, members of the public, and schoolchildren

A portfolio of outreach events have been delivered during Darwin Tree of Life Project Phase I, including:

- Citizen science sample collecting events
- DNA barcoding workshops in schools
- Online schools events with live chat to Darwin Tree of Life partner scientists
- The Royal Society Summer Science Exhibition 2022

---

[34] Matheson, P., McGaughran, A. Genomic data is missing for many highly invasive species, restricting our preparedness for escalating incursion rates. Sci Rep 12, 13987 (2022). https://doi.org/10.1038/s41598-022-17937-y

[35] McGaughran, Angela, et al. 2024. "Genomic Tools in Biological Invasions: Current State and Future Frontiers." Genome Biol Evol 16, no. 1 (January): evad230. doi: 10.1093/gbe/evad230.

[36] Gladstone, Emily, and Oxford Economics. 2024. "UK life sciences are set for growth, but challenges remain." Oxford Economics Consulting Report, (February). https://www.oxfordeconomics.com/resource/uk-life-sciences-are-set-for-growth-but-challenges-remain/.

[37] The Royal Society. 2024. "Students want to do more science experiments at school but have less access to practical work, says Royal Society and EngineeringUK survey." News from the Royal Society. https://royalsociety.org/news/2024/04/science-education-tracker-2023/.

Collectively these events reached over **5,000 non-scientists**, including **1,800 school-age children**. In an era of STEM skills gaps and declining practical science education, projects improving public scientific literacy and school STEM engagement are critical to the UK economy and the future of scientific innovation.[38]

Enriching National Collections

The UK has an unparalleled wealth of historic scientific materials going back centuries, having led the global scientific revolution of the 17th Century and the industrial revolution of the 18th Century. The resulting legacy is the UK's National Collections– registered organism collections that are maintained to preserve knowledge and further scientific discoveries.

The Darwin Tree of Life Project collaborates with all major UK National Collections, including the National Plant Collections at the Royal Botanic Garden Edinburgh and the Royal Botanic Gardens, Kew; and the National Collection of Fungus Cultures held at CABI, and the Culture Collection of Algae and Protozoa at CCAP. The Project enriches these historic scientific collections with species-level genomic sequence information, amplifying their value and enabling them to continue to drive scientific discovery now and in the future.

---

[38] Cambridge Industrial Innovation Policy. 2025. "Science and Engineering Workforce." https://www.ciip.group.cam.ac.uk/innovation/science-and-engineering-workforce-2025/.

Looking to the Future

During Phase I, the Darwin Tree of Life Project overcame all major hurdles in high-throughput, place-based reference genome sequence delivery. The insights, tools and genome outputs resulting from the outputs of Phase I are shaping this field and increasing the effectiveness of both local and global science. Delivery of Phase I will continue until all **10,000 species** collected have been processed, sequenced and published to the community.

To build on this success the project seeks to pivot into a **high-output Phase II**, focused on efficient sample collection and genome generation, with the target of releasing genomes from a further **20,000 species**. Collectively, Phase I and Phase II will **sequence all 30,000 findable species in Britain and Ireland, and generate the world's first ecosystem-level genomic dataset.**

Investments in open research, scientific skills training, convening the research community and scientific outreach that have characterised the Darwin Tree of Life Project will continue to be critical to the impact of this groundbreaking project.

Genomics is emerging as the next transformative technology for biodiversity and ecology, but this depends on the availability of reference genome sequences from a large number of diverse species. By providing reference sequences for all species found in Britain and Ireland, Phase II will unlock the application of genomics to the living world as a whole, and enable a great wealth of applications and societal benefits. Now is the time to do this – the technical means are in place, the Project consortium has proven itself, and the scientific community is actively adopting the genomes as a trusted high-quality resource.

The Darwin Tree of Life Project stands as a pioneer in this emerging field. Where Phase I has demonstrated the effectiveness of this approach to reference genome sequencing and laid the groundwork for even greater delivery, Phase II will cement Britain and Ireland's position as the global leaders in biodiversity genomics, while enabling scientific breakthroughs, driving economic growth and societal wellbeing, and providing essential tools for biodiversity conservation in this era of unprecedented environmental change.

Darwin TREE of LIFE

www.darwintreeoflife.org

This work was supported by the Wellcome Trust [108413/A/15/D],[220540/Z/20/A], [218328_Z-19_Z]; and Defra, as well as in-kind contributions from all partner organisations

Project delivered by:

Royal Botanic Garden Edinburgh
CABI
EMBL-EBI
wellcome sanger institute
Wytham Woods
University of Cambridge
Royal Botanic Gardens Kew
Earlham Institute
University of Oxford
Natural History Museum
The University of Edinburgh
Marine Biological Association
ZSL Institute of Zoology
UCD Dublin