Research Ideas and Outcomes 12: e187550
doi: 10.3897/rio.12.e187550

OPEN ACCESS

Grant Proposal

Biodiversity Genomics Europe (BGE) Project - Abridged Grant Proposal

Dimitris Koureas[‡,§], Pedro Beja[|,¶], Mark L. Blaxter[#], Astrid Böhne[¤], Sarah J. Bourlat[¤], Torbjørn Ekrem[«], Brent C. Emerson[»], Katharina F Heil[^], José Melo-Ferreira[|,˅,¶], Ben Price[¦], Rutger A. Vos[‡], Wouter Addink[‡], Tyler Alioto[ʔ,ʕ], Filippos A. Aravanopoulos[¢], Jonas J. Astrin[¤], Jean-Marc Aury[ℓ], Ian Barnes[¦], Claudia Bruschini[ϑ], Elena Buzan[P,Ⱥ], Guy Cochrane[ƨ], Tamás Cserkész[Ғ], Thanos Dailianis[₹], Elza J.E.M. Duijm[‡], Glenn Dunshea[«], Rosa Fernández[₦], Sónia Ferreira[|,¶], Giulio Formenti[ĸ], Sara Fratini[ϑ], Konstantinos Gkagkavouzis[¢,ʔ], Carole A. Goble[₩], Michal Grabowski[₸], Bjorn Grüning[♯♯], Ivo Gut[ʔ,ʕ], Marta Gut[ʔ,ʕ], Peter W. Harrison[ƨ,§§], Axel Hausmann[||], Jacob Höglund[¶¶], Laura Iacolina[P], Alessio Iannucci[ϑ], Kjetill S. Jakobsen[##], Urmas Kõljalg[¤¤], Henrik Lantz[««], Harris A. Lewin[»»], Jan N. Macher[‡,^^], Tereza Manousaki[₹], Fergal J. Martin[ƨ], Ximo Mengual[¤], Pedro H. Oliveira[ℓ], Rebekah Oomen[˅˅,ıı,ʔʔ,ʕʕ,¢¢], Michael J. Raupach[||], Ana Riesgo[ℓℓ], Hugues Roest Crollius[ϑϑ], Anna Somogyi[Ғ], Torsten H. Struck[ıı], Hannes Svardal[PP], Ave Tooming-Klunderud[##], Alexandros Triantafyllidis[¢,ʔ], Olga Vinnere Pettersson[ÅÅ,ƨƨ,ғғ], Maximilian Wagner[₹₹], Patrick Wincker[ℓ], Ni Yan[‡], Jose M. Alonso[‡], Ana Casino[₦₦], Claudio Ciofi[ϑ], Gabriela Daňková[‡], Peter M. Hollingsworth[ĸĸ], Mara K.N. Lawniczak[#], Camila J. Mazzoni[¢¢,ʔʔ], Robert M. Waterhouse[₩₩,₸₸]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
| CIBIO, Centro de Investigação Em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade Do Porto, Vairão, Portugal
¶ BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal
# Tree of Life, Wellcome Sanger Institute, Hinxton, United Kingdom
¤ Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Bonn, Germany
« Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway
» Instituto de Productos Naturales y Agrobiología (IPNA-CSIC), La Laguna, Tenerife, Islas Canarias, Spain
^ ELIXIR Hub, Hinxton, United Kingdom
˅ Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
¦ Natural History Museum, London, United Kingdom
ʔ Centro Nacional de Analisis Genomico (CNAG), Barcelona, Spain
ʕ Universitat de Barcelona (UB), Barcelona, Spain
¢ Faculty of Agriculture, Forest Science & Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece
ℓ Génomique Métabolique, Genoscope, Institut François Jacob, Commissariat à l'Energie Atomique (CEA), CNRS, Université Evry, Université Paris-Saclay, Evry, France
ϑ Department of Biology, University of Florence, Sesto Fiorentino (FI), Italy
P Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia
Ⱥ Faculty of Environmental Protection, Velenje, Slovenia
ƨ European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom
Ғ Hungarian National Museum Public Collection Centre - Hungarian Natural History Museum, Department of Zoology, Budapest, Hungary
₹ Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research (HCMR) Crete, Heraklion, Greece

© Koureas D et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

₦ Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain
₭ The Vertebrate Genome Laboratory, The Rockefeller University, New York, United States of America
₲ Department of Genetics, Development and Molecular Biology, School of Biology, Faculty of Science, Aristotle University of Thessaloniki, Thessaloniki, Greece
? Genomics and Epigenomics Translational Research (GENeTres), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Thessaloniki, Greece
₩ The Department of Computer Science, the University of Manchester, Manchester, United Kingdom
₸ Department of Invertebrate Zoology & Hydrobiology, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland
‡‡ Bioinformatics Group, Department of Computer Science, Albert-Ludwigs University Freiburg, Freiburg, Germany
§§ Health Data Research UK, London, United Kingdom
|| SNSB - Bavarian State Collection of Zoology, Munich, Germany
¶¶ Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway
¤¤ University of Tartu, Tartu, Estonia
«« Department of Medical Biochemistry and Microbiology, National Bioinformatics Infrastructure Sweden, SciLifeLab, Uppsala University, Uppsala, Sweden
»» Julie Ann Wrigley Global Futures Laboratory, Arizona State University, Tempe, United States of America
^^ Department of Environmental Biology, Institute of Environmental Sciences (CML), Leiden University, Leiden, Netherlands
˅˅ Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway
¦¦ Natural History Museum, University of Oslo, Oslo, Norway
?? University of New Brunswick Saint John, Saint John, Canada
ss Tjärnö Marine Laboratory, University of Gothenburg, Strømstad, Sweden
¢¢ Centre for Coastal Research, University of Agder, Kristiansand, Norway
ℓℓ National Museum of Natural Sciences, Madrid, Spain
ϑϑ Institut de Biologie de l'ENS (IBENS), CNRS UMR 8791, Université PSL, Paris, France
PP Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp, Belgium
AA SciLifeLab, Uppsala Biomedical Centre (BMC), Uppsala, Sweden
₴₴ National Genomics Infrastructure, Uppsala Genome Center, Uppsala, Sweden
₣₣ Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
₹₹ Faculty of Civil Engineering, University of Rijeka, Rijeka, Croatia
₦₦ Consortium of European Taxonomic Facilities (CETAF), Brussels, Belgium
₭₭ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom
₲₲ Berlin Center for Genomics in Biodiversity Research, Berlin, Germany
?? Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
₩₩ SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
₸₸ University of Lausanne, Lausanne, Switzerland

Corresponding author: Dimitris Koureas (dimitris.koureas@naturalis.nl)

Reviewable v 1

Received: 04 Feb 2026 | Published: 27 Feb 2026

Citation: Koureas D, Beja P, Blaxter ML, Böhne A, Bourlat SJ, Ekrem T, Emerson BC, Heil K, Melo-Ferreira J, Price B, Vos RA, Addink W, Alioto T, Aravanopoulos FA, Astrin JJ, Aury J-M, Barnes I, Bruschini C, Buzan E, Cochrane G, Cserkész T, Dailianis T, Duijm EJEM, Dunshea G, Fernández R, Ferreira S, Formenti G, Fratini S, Gkagkavouzis K, Goble CA, Grabowski M, Grüning B, Gut I, Gut M, Harrison PW, Hausmann A, Höglund J, Iacolina L, Iannucci A, Jakobsen KS, Kõljalg U, Lantz H, Lewin HA, Macher JN, Manousaki T, Martin FJ, Mengual X, Oliveira PH, Oomen R, Raupach MJ, Riesgo A, Roest Crollius H, Somogyi A, Struck TH, Svardal H, Tooming-Klunderud A, Triantafyllidis A, Vinnere Pettersson O, Wagner M, Wincker P, Yan N, Alonso JM, Casino A, Ciofi C, Daňková G, Hollingsworth PM, Lawniczak MKN, Mazzoni CJ, Waterhouse RM (2026) Biodiversity Genomics Europe (BGE) Project – Abridged Grant Proposal. Research Ideas and Outcomes 12: e187550. https://doi.org/10.3897/rio.12.e187550

Abstract

The Biodiversity Genomics Europe (BGE) Project has the overarching aim of accelerating the use of genomic science to enhance understanding of biodiversity, monitor biodiversity change, and guide interventions to address its decline. The BGE Project comprises activities focused on DNA Barcoding (Barcoding Stream) and Reference Genome Generation (Genomes Stream) for eukaryotic species across Europe, bringing together two European networks: the International Barcode of Life in Europe (iBOL Europe) and the European Reference Genome Atlas (ERGA). This publication is an abridged version of the successful grant proposal developed jointly by iBOL Europe and ERGA in response to the Horizon Europe call HORIZON-CL6-2021-BIODIV-01-01. Two key strands of genomic science form the basis of this proposal: DNA barcoding - sequencing short, standardised genomic regions to tell the world's species apart, transforming the speed of completion of the inventory of life on Earth and providing the foundations of a global bio-surveillance system for biodiversity; and genome sequencing - generating high-quality complete reference genomes for all species on Earth, transforming understanding of biodiversity at the genetic level, and delivering fundamental knowledge of how biological systems function and how species respond and adapt to environmental change. The BGE Project objectives are focused on (i) Capacity: To establish functioning biodiversity genomics networks at the European level to connect and grow community capacity to use genomic tools to tackle the biodiversity crisis; (ii) Production: To establish and implement large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation, and biomonitoring; and (iii) Application: To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines to improve the efficacy of management interventions and biomonitoring programmes.

Keywords

Biodiversity genomics, DNA barcodes, ERGA, Horizon Europe, iBOL Europe, reference genomes

List of participants

1. Stichting Naturalis Biodiversity Center† (Naturalis) Netherlands

2. Forschungsverbund Berlin e.V. (IZW) Germany

3. Università degli Studi di Firenze (UNIFI) Italy

3.1 Università degli Studi di Bari Aldo Moro (UNIBA) Italy

4. Staatliche Naturwissenschaftliche Sammlungen Bayerns (SNSB) Germany

5. Jyväskylän yliopisto (JYU) Finland

6. Associação BIOPOLIS (CIBIO) Portugal

7. Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) France

8. Leibniz-Institut zur Analyse des Biodiversitätswandels (LIB) Germany

9. Tartu Ülikool (UT) Estonia

10. Uppsala universitet (UU) Sweden

11. Fundacio Centre de Regulació Genòmica (CRG) Spain

12. ELIXIR/EBI - European Molecular Biology Laboratory (EMBL) Intergovernmental

12.1 Albert-Ludwigs-Universität Freiburg (ALU-FR) Germany

13. Norges teknisk-naturvitenskapelige universitet (NTNU) Norway

14. Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) Spain

15. Universitetet i Oslo (UiO) Norway

16. Consortium of European Taxonomic Facilities (CETAF) Belgium

17. Univerza na Primorskem Università del Litorale (UP) Slovenia

18. Uniwersytet Łódzki (UniLodz) Poland

19. Aristotelio Panepistimio Thessalonikis (AUTH) Greece

20. Panepistimio Kritis (NHMC) Greece

20.1 Hellenic Centre for Marine Research (HCMR) Greece

21. Faculty of Science University of Zagreb (UniZagreb) Croatia

22. Magyar Természettudományi Múzeum (HNHM) Hungary

23. V. N. Karazin Kharkiv National University (KKNU) Ukraine

24. Université de Lausanne (UNIL) Switzerland

25. International Barcode of Life Consortium (iBOL) Canada

26. The Rockefeller University (VGP) USA

27. The University of California Davis (EBP) USA

28. Royal Botanic Garden Edinburgh (RBGE) United Kingdom

29. Wellcome Sanger Institute - Genome Research Limited (Sanger) United Kingdom

30. Natural History Museum, London (NHM) United Kingdom

31. Earlham Institute (EI) United Kingdom

32. The University of Manchester (UMAN) United Kingdom

33. Consorcio para la Explotación del Centro Nacional de Análisis Genómico (CNAG) Spain

34. SIB Swiss Institute of Bioinformatics (SIB) Switzerland

† The Project coordinator.

Context

This paper is an abridged version of the original proposal developed to respond to the Horizon Europe call HORIZON-CL6-2021-BIODIV-01-01, which was submitted in October 2021. It contains the overarching scientific case for the Biodiversity Genomics Europe (BGE) project, alongside a description of our major ambitions and activities. Differences between this paper and the full "Description of Work" include redactions, e.g. of financial and personal information alongside our risk analysis; addition of formatted citations that were not included in the original proposal; minor edits to improve readability; and the inclusion of high-resolution versions of the figures. Other differences arise from post-award and in-project amendments, including changes in the BGE Consortium composition. The project started in September 2022, with a duration of 3.5 years. Note that since the submission of the proposal, BIOSCAN Europe has changed its name to iBOL Europe. The abridged proposal is published here to frame the publication of future outputs from the BGE project and Consortium members.

1. Excellence

1.1 Objectives and ambition

The Biodiversity Genomics Europe (BGE) Consortium has the overriding aim of accelerating the use of genomic science to enhance understanding of biodiversity, monitor biodiversity change, and guide interventions to address its decline.

Background and Project premise

Large-scale environmental change leads to massive biodiversity loss, with extinction rates at 10-100 fold above baseline and an estimated 25% of species threatened with extinction worldwide. This loss threatens to rapidly erode the provision of ecosystem goods and services that human society depends upon (IPBES et al. 2019). Addressing the global biodiversity crisis requires an understanding of the diversity of life on Earth,

how that diversity functions and interacts, and how biodiversity responds to different environmental pressures. Behind that simple statement lies a major challenge - the enormous complexity of Earth's biological systems. After centuries of research, around 2 million species have been scientifically described, but an estimated 80% of the world's multicellular species await scientific discovery and description (Mora et al. 2011). Even for described species, telling them apart is often difficult, and knowledge of their distributions, variation, properties, inter-dependencies, and conservation status remain patchy and incomplete.

Two key strands of genomic science are now addressing this complex challenge and form the basis of this proposal:

Large-scale DNA barcoding - sequencing short, standardised genomic regions to tell the world's species apart, transforming the speed of completion of the inventory of life on Earth and providing the foundations of a global bio-surveillance system for biodiversity.

Large-scale genome sequencing - generating high-quality complete reference genomes for all species on Earth, transforming understanding of biodiversity at the genetic level, and delivering fundamental knowledge of how biological systems function and how species respond and adapt to environmental change.

Building European capacity

The **International Barcode of Life (iBOL)** (Ratnasingham and Hebert 2007) and the **Earth BioGenome Project (EBP)** (Lewin et al. 2018) are two global initiatives, each providing a framework of governance and delivery for barcoding and reference genome generation, respectively. Within each of these initiatives, there is extensive and growing global participation. However, there remains a critical need to develop and strengthen functioning communities of practice, within and between these initiatives and at multiple scales.

At national levels for both iBOL and EBP, there is a need to build capacity, and establish mechanisms to democratise participation. Among countries, there is an urgency to ensure complementarity so that activities in individual countries achieve synergy rather than redundancy. More generally, there is an opportunity to better connect the currently separate programmes of iBOL and EBP to enable the skills and infrastructures of the DNA barcoding and genome sequencing communities to work together and capitalise on the common ground of DNA sequencing of biodiversity at scale. Overall, for the ambitious goal of massive scale sequencing of global biodiversity, there needs to be more effective international delivery networks, sharing of knowledge, working to common standards, and creating a flow of samples and data from the field and biorepositories to sequencing facilities and analytical pipelines. In turn, these initiatives in large-scale biodiversity genomics need to be fully integrated into the wider landscape of biological and environmental data infrastructures.

Coordinating efforts - Uniting research agendas

There is thus a clear need and opportunity for the BGE Consortium to unite the research agendas of DNA barcoding and reference genome generation and establish a step-change in biodiversity genomics research in Europe, aimed at underpinning biodiversity conservation and DNA-based biomonitoring. At a policy level, this aligns to the EU biodiversity strategy for 2030 (European Commission. Directorate General for Environment. 2020), and to its commitments, such as addressing the decline in pollinators, improving the conservation status of key species and habitats, promoting ecological restoration of forest systems and rivers, and minimising impacts from invasive species. Likewise, a step-change in the application of genomic science to biodiversity conservation aligns with the European Green Deal (European Commission. Directorate General for Communication. 2020), which stresses the importance of preserving and restoring ecosystems and biodiversity in terrestrial, freshwater, and marine systems.

The coordination and upscaling of biodiversity genomics in Europe will also capitalise on key assets and recent technological innovations. Developments in sequencing technologies are transforming the speed and cost-effectiveness of data generation, supporting an exponential increase in data availability. This growth in biodiversity genomics data is paralleled by developments in European and international data infrastructures (e.g. INSDC – International Nucleotide Sequence Database Collaboration, ELIXIR, DiSSCo – Distributed System of Scientific Collections, BOLD – Barcode of Life Data Systems, GBIF – Global Biodiversity Information Facility), which offer opportunities to enhance management, analysis, sharing and interoperability of multiple data sources. Underpinning these technological and data assets are the extensive and globally important natural history collections in European museums, herbaria, botanic gardens, zoos and culture collections. These biorepositories of millions of expertly verified samples represent an invaluable resource for biodiversity genomics projects, supporting the provision of materials to establish reference libraries of barcodes and genomes.

Collectively, these assets and policy drivers create a compelling case for the European biodiversity and genomics networks to join forces to deliver data and influence at scale in Europe and provide leadership in partnership working to shape the wider international landscape of biodiversity genomics science. The BGE consortium assembled here, addresses this challenge and brings together two newly formed networks, BIOSCAN Europe (which focuses on DNA barcoding) (iBOL Europe 2025) and the European Reference Genome Atlas (ERGA) (which focuses on genome sequencing) (Mazzoni et al. 2023), to support its delivery.

This Project will coordinate and upscale DNA barcoding and genome sequencing of biodiversity in Europe to enhance understanding of biodiversity and biodiversity declines and develop synergies by aligning the efforts and resources of the DNA barcoding and genome sequencing communities. The Project will tackle three fundamental objectives, directly aligned to the specified outcomes of the call (Table 1).

Table 1.

The BGE Project has three key objectives (Capacity, Production, Application) aligned to the desired outcomes of the Topic.

Our objectives		Specified outcomes of the Topic (HE-CL6-2021-BIODIV-01-01)
CAPACITY	Establish functioning biodiversity genomics networks at the European level to connect and grow community capacity to tackle the biodiversity crisis using genomic tools	Creation and management of the European node of the International Barcode of Life.
		Creation of a European hub (node) affiliated to the Earth BioGenome Project.
		Leverage active support and cooperation of citizen scientists and other non-professional taxonomists.
PRODUCTION	Establish and implement large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation and biomonitoring	Set-up the necessary networks, technologies, quality standards, reference atlas and taxonomic expertise through Europe to identify systematically, and comprehensively specific, intraspecific and ecosystem diversity through genomics techniques, such as full-genome sequencing, barcoding and metabarcoding.
APPLICATION	Apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines to improve the efficacy of management interventions and biomonitoring programmes	Advances in the assessment of pan-European biodiversity via genome sequencing and/or DNA barcoding of threatened/endangered species, ecological keystone species and economically important species.
		Pan-European barcoding of pollinators by completing the Barcode of Life for European bees, butterflies, moths and hoverflies.

Project ambition

To establish functioning biodiversity genomics networks at the European level we will build European nodes for iBOL and EBP and connect these nodes to the international scientific community, policy makers, practitioners, and wider society. Our ambition is to:

- Future-proof our networks on biodiversity genomics research - working together to develop mid- and long-term operational and scientific aspirations that act as roadmaps for future investments in the field of biodiversity genomics in Europe. The Project contributes by setting up organisational models and planning mechanisms for the European nodes of iBOL and EBP, and via their respective networks, building a set of collaborations encompassing all European countries.
- Lower access thresholds to biodiversity genomics research - improving the accessibility of skills, technologies, samples and data to increase participation in biodiversity genomics research. This includes addressing imbalances in access to technologies by opening up transnational access to sequencing facilities, as well as building in-country capacity. It includes training to enhance engagement with practices in the domain and also understanding of relevant data and outputs.

- Connect our biodiversity genomics networks - developing and establishing strong strategic, scientific, and operational links between the barcoding and genome research networks in Europe. The Project contributes by bringing together people and organisations to form joint research agendas and share resources, and demonstrate the added value derived from the joint development of best practices.
- Promote co-creation and citizen engagement - to establish a better dialogue between the scientific networks for biodiversity genomics research and society. The Project contributes by supporting citizen scientists and amateur taxonomists to directly contribute to the scientific process from inception to delivery of impact.

To establish and implement large-scale biodiversity genomic data generation pipelines for Europe we will align efforts across 20 countries connecting samples to sequencing facilities to data-management and data-processing workflows, supported by standardised operating procedures and infrastructure research and development. Our ambition is to:

- Establish distributed capacity - building a truly distributed and inclusive mechanism for delivering large-scale scientific outputs in Europe. The Project brings together research communities, museums, and sequencing centres across multiple countries to share the load of processing material. The material will come from an international sampling effort from field campaigns and multiple natural science collections.
- Build economies of scale - significantly reduce unit costs by benefiting from a strong specialisation strategy across partners. The Project contributes by understanding what these specialisations might look like for different community members and demonstrating how by reducing the extent of the scope of individual institutions, we can improve efficiency and lower costs.
- Connect previously disjoined resources to deliver knowledge - developing a European mechanism where information deriving from different sources of biodiversity genomics research (including European, national, and institutional projects) can be joined together. The Project contributes by rolling out technological solutions to transform or capture barcoding and genome data in a structured and standardised way. To this end we focus on further developing (meta)data standards and enhancing existing pipelines and systems.

To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines we will use DNA barcoding and genome sequencing to enable biodiversity characterisation, conservation, and biomonitoring in taxa and systems of high conservation, ecological, and socioeconomic importance. Our ambition is to:

- Better embed the use of biodiversity genomics data in science policy - demonstrating through strong case studies how biodiversity genomics data can provide powerful and reliable contributions to biodiversity knowledge gaps. The Project contributes through case studies that focus on threatened/endangered species, ecological keystone species, and economically important species, including pollinators.

- Establish European-wide large-scale biodiversity genomics research mechanisms - enabling large-scale, continuous, continent-wide systems for effective and efficient biomonitoring, species discovery, genome research, and biodiversity conservation. The Project demonstrates, through case studies, the feasibility of a pan-European approach and also establishes the Standard Operating Procedures (SOPs), as well as the technical and scientific roadmaps for scaling up.

Looking into the future - Our longer-term scientific ambitions

The importance of diversity at the species and genetic levels is central to the Convention on Biological Diversity (CBD 1992) and constitutes the foundation of both iBOL and EBP. The visions laid out by the iBOL and EBP communities to unveil the diversity of species and genomes of life on Earth through DNA sequencing cannot be completed by small groups working in isolation. Here, we will establish a connected European biodiversity genomics community. This community will be fundamental to continental and global DNA-based biomonitoring and genomic-informed biodiversity conservation and restoration actions over the decades to come.

DNA-based biomonitoring of species is most efficiently and effectively carried out by targeted approaches utilising minimal DNA barcodes to enable processing of large sample sizes. These barcode markers can be used to identify the species present in any kind of sample, whether it is a whole organism, parts of an organism, a mixture of taxa or even traces of DNA left in the environment (eDNA). The important link between these DNA barcodes and a species identity requires the investment in a reference index of DNA barcode sequences from expertly identified species. Toward this end, we will make a sizable contribution of 15,000 expertly identified species of the highest priority for European biomonitoring selected through a collaborative Europe-wide gap analysis. Together with targeted curation of the European species already present in barcode databases, we will set the foundations for a complete DNA barcode inventory of the European biota.

Reference genomes carry important information about intraspecific genetic diversity as well as being fundamental to building new tools to study intraspecific genetic diversity and how genes and other functional regions contribute to determining species phenotypes as well as the biology and function of communities of species. In the short-term, sequencing genomes of selected species will support the development of new methods for standardised monitoring of intraspecific genetic diversity of endangered and other key species, a critically important component of biodiversity which to date has not been adequately integrated into national and international biodiversity monitoring programmes. The Project will deliver reference genomes for species distributed across Europe under the categories of critical biodiversity and biodiversity hotspots and will establish an integrated network linked to the EBP. This network will provide the toolkits and processes necessary for subsequent generation of reference-quality, complete genome sequences for all species across Europe. In the longer-term, the broadening of

reference genome databases will enable the use of metagenomics in biodiversity monitoring and the evaluation of species community genetic function, to inform the success of biodiversity restoration programmes.

While the data we generate in the Project will be important, even more critical is the establishment of the connected European biodiversity genomics community. The BGE Consortium will share ways of working, through agreed sample collection standards, analytical tools and guidance, and co-created SOPs for the latest technical advances. This shared practice will bring down costs and ramp up the scale of the work we will need to do to realise our long-term goal of comprehensively monitoring and understanding biodiversity through genomics. Through this work we will position the biodiversity genomics community in Europe to quantify species and genetic diversity and track its change through space and time at the required scale. While the longer-term vision of generating DNA barcodes and reference genomes for all European species is undoubtedly ambitious, major advances in sequencing technologies now make this affordable, and a coordinated effort of pan-European and international research groups makes it feasible. The BGE Consortium established here is represented by some of the major, most impactful, and established organisations in the field of biodiversity science and genomics. Similar to the influence that the Human Genome Project had in the last two decades, the massive scale deployment of genomic approaches to biodiversity will fundamentally and dramatically change conservation biology and basic biological research.

1.2 Methodology

Project structure

Our Project is organised into two major Streams of work, focusing on DNA Barcoding (iBOL) and Genome Sequencing (EBP), led by BIOSCAN Europe and ERGA, respectively. The Work Package (WP) and management structure of the Project are shown in Fig. 1, and represents a matrix structure of the DNA Barcoding and the Genome Sequencing Streams, with Pillars representing steps of a value chain that addresses the needs of these two Streams.

Streams focus on creating a value chain connecting the delivery WPs. The WPs for the DNA Barcoding Stream are: WP2, WP4, WP6, WP8, WP10. The WPs for the Genome Sequencing Stream are WP3, WP5, WP7, WP9, WP11. A third Stream (WP12) is formed at the interface of the two networks and delivers work that represents the joint investments and shared resources needed for the Project implementation and the science at the interface of DNA barcoding and genome sequencing.

Pillars form the five areas of work of the Project (node development, sampling, sequencing, data, applications). Pillars include WPs that contribute to the steps in the delivery chain across the two networks. Connectivity in the vertical axis (Figure 1) is essential in terms of coordinating innovation and capacity building activities.

Delivery WPs address the three overarching Project Objectives (Table 1), as elaborated below. (1) CAPACITY will be primarily addressed by WP2 and WP3, which focus on building the European Nodes for iBOL and EBP respectively. (2) PRODUCTION will be primarily addressed by WPs 4-9, which include paired WPs for DNA barcoding and genome sequencing, focusing on samples (WPs 4,5), sequencing (WPs 6,7), and data processing and informatics (WPs 8,9). (3) APPLICATION will be primarily addressed by WP10 and WP11, focusing on applications of DNA barcoding and genome sequencing, respectively, to biodiversity characterisation, conservation and biomonitoring. Finally, WP12 delivers cross-cutting Tasks and activities at the interface of the DNA barcoding and genome sequencing Streams, and thus contributes to all three objectives.

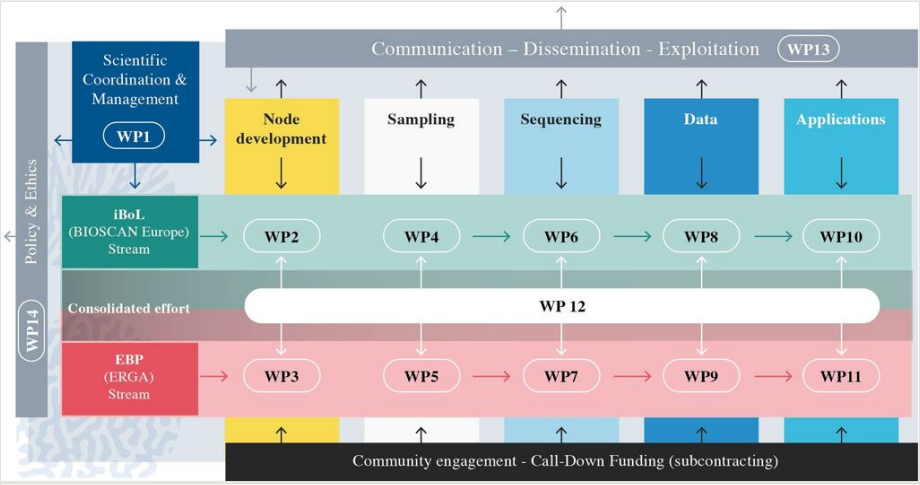

Figure 1. doi

Work package (WP) and management structure of the BGE Project.

Supporting WPs enable effective delivery and impact of the Project. Management of the Project is delivered through WP1. Communication, Dissemination, & Exploitation is addressed in WP13. Legal and Ethical compliance, as well as contributions to European regulatory frameworks, is delivered through WP14.

DNA Barcoding Stream

The BIOSCAN Europe work Stream focuses on building and curating DNA barcode reference libraries to characterise and identify European biodiversity and developing DNA-based biomonitoring programmes to track biodiversity change. Our approach is driven by the importance of establishing well-curated barcode reference libraries as the foundation step. Obtaining high-quality DNA barcode sequences from well-identified specimens is the critically important link between barcode sequences and species identity. Well-curated reference libraries enable large-scale, cost-effective biodiversity monitoring by high-throughput DNA barcoding of unknown samples, with assessments of species identity and diversity obtained from reference library queries. Our approach to

the DNA Barcoding Stream thus focuses heavily on the investment in reference library construction to enable biomonitoring of critically important taxa and ecosystems in Europe and supporting this with use cases to improve the protocols and harmonisation of DNA-based biomonitoring approaches. Another major focus of our work is building capacity and workflows to enable future large-scale delivery of DNA barcoding projects in Europe. The five WPs and key Tasks in the BIOSCAN Europe Stream are summarised in Fig. 2.

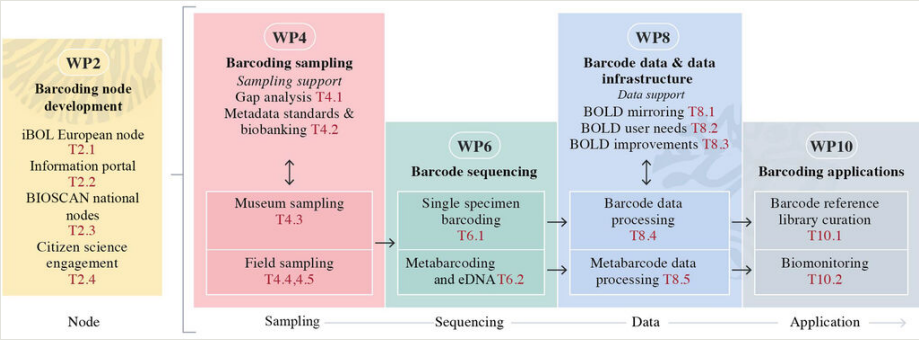

Figure 2. doi

Barcoding WPs and Task interrelationships.

(Obj 1) CAPACITY - To establish functioning biodiversity genomics networks at the European level to connect and grow community capacity to use genomic tools to tackle the biodiversity crisis

To establish functioning biodiversity genomics networks at the European level, we will build a European node for the iBOL and connect this node to the international scientific community, policy makers, practitioners, and wider society (WP2). The current phase of the global iBOL programme (BIOSCAN) involves more than 40 countries as formal partners, including strong national barcoding programmes in several European countries. However, there is no formal mechanism within Europe to coordinate efforts and achieve synergies, resulting in a fragmented landscape of activity. To address this, we will build on our recently established collective of 85 partner institutes that joined BIOSCAN Europe as a precursor to a European node of iBOL. We will establish a governance structure and formalise BIOSCAN Europe as the European node of iBOL [T2.1], evaluating optimal business models to maximise the sustainability of operations beyond the lifetime of the Project. BIOSCAN Europe will be supported by a Secretariat, who will coordinate the delivery of workshops, training events and information exchange, and the establishment of working groups to provide strategic direction for DNA barcoding in Europe. A central information portal will be established as a community nexus [T2.2] for regional, national, and transnational barcoding projects. A further key component of a functioning iBOL network in Europe is the establishment and strengthening of national barcoding programmes. To tackle this, we will develop model approaches to grow national-level capacity [T2.3], supporting the organisation of barcoding communities into

coherent national nodes. The initial focus will be pilot studies for Greece (with high biodiversity, but no national barcoding programme) and Poland (with a national barcoding framework at an early stage of development), establishing a transferable framework for community organisation, communication, expertise sharing, and connections to funders and policy makers. To provide wider support to national level barcoding activities in Europe, we will run a scheme enabling the wider community to access Project funds as subcontractors to generate barcodes to support national barcode reference library construction [T2.1].

To further grow the network of participants involved in DNA barcoding in Europe, we will capitalise on the distributed skills base in species identification and taxonomy among amateur natural historians. Working through established citizen science frameworks and natural history society networks, we will engage amateur experts to support the collection and identification of samples for DNA barcoding [T2.4] and then work together to interpret the findings of the barcode data. This engagement with specialist citizen scientists will be supported by wider public participation in BioBlitz sampling for DNA barcoding and eDNA biomonitoring of marine invasive species in ports and harbours [T2.4]. The critically important feedback loop from data gathering to sharing findings will be supported by a dedicated web-interface within the barcoding information portal.

(Obj 2) PRODUCTION - To establish and implement large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation, and biomonitoring

To establish and implement large-scale biodiversity genomic data generation pipelines for Europe, we will align efforts across 20 countries connecting samples to sequencing facilities to data-management and data-processing workflows, supported by SOPs and infrastructure research and development. This involves three major steps: Sampling (WP4), Sequencing (WP6), and Data Processing and Analysis (WP8).

Barcode Sampling (WP4): We will establish a prioritised sample supply chain delivering high-quality samples for DNA barcoding and metabarcoding to sequencing facilities. Our barcode reference library construction activities will primarily focus on completing taxonomic coverage of key groups that underpin knowledge of ecosystem health in Europe: with a major focus on insect pollinators, and taxa to support freshwater and marine biomonitoring (invertebrates, fish, and aquatic plants). The existing barcode databases hold records of thousands of European species in these groups, but there are numerous gaps, e.g. (Weigand et al. 2019), and the data are of varying levels of quality. We will undertake a gap analysis to prioritise our sampling [T4.1], identifying the species that are missing/underrepresented, and use this to develop a sampling strategy based on field and museum sampling. Sampling from the field has the benefit of providing fresh easy-to-sequence material, but fieldwork costs are expensive, and tracking down ‘missing’ species to fill reference library gaps can be time-consuming and difficult. There are also substantial costs involved in making voucher specimens and obtaining verified identifications of newly collected material. Sampling from museum collections has the benefit of building on past investments into collection, vouchering and verification, but

with the additional challenge that museum specimens frequently have more fragmented DNA, requiring more costly sequencing approaches. Our preliminary analyses and past experience suggest that the most time and cost-effective strategy will be a split of 75% museum and 25% field sampling. We will implement standardised metadata collection processes [T4.2], and sample a total of 45,000 samples of 15,000 species, with c. 33,000 specimens obtained from Europe's scientific collections [T4.3] and c. 12,000 from new field collections [T4.4; T4.5]. The museum sampling will focus on a network of 20 major and regional museums (many of which are participants in this Project), and will use our established minimally destructive sampling techniques, enabling the sampling, data entry and imaging of 150 samples per person per week. Our field sampling will be delivered via the extensive network of field biologists within the Project, and strongly supported through the citizen science programme outlined in Task T2.4.

To enable the biomonitoring of biological communities in a comparable and scalable way, we will undertake field sampling for DNA-based biomonitoring [T4.3]. We will deliver field campaigns to collect community samples of insect pollinators (450 sites), terrestrial arthropods and soil cores from restoration sites (50 sites), terrestrial arthropods from sites subject to agricultural intensification (30 sites), terrestrial arthropods, bryophytes, and lichens from mountain ranges (40 sites), and marine invasive species (55 sites). Sampling will be carried out across Europe, following standardised protocols, and documented effective sampling methods for the target organisms (e.g., pan traps, Malaise traps). Collectively this will involve a total of 3,500 sampling events. All samples collected in WP4 will comply with applicable legal and ethical requirements, and all samples and associated DNA extracts will be archived in biobanks [T4.2] to ensure access for long-term use.

Barcode Sequencing (WP6): We will establish and implement a functioning sequencing pipeline for single specimen barcoding and bulk/environmental DNA metabarcoding across Europe. Specimens collected in Tasks 4.3-4.5 will be sequenced to recover standard barcode markers using either amplicon sequencing or low coverage genome skimming approaches [T6.1]. Freshly collected specimens will be amplicon sequenced using low cost, high-throughput platforms (PacBio and ONT, e.g., (Srivathsan et al. 2021)). As museum specimens typically contain highly fragmented DNA, we will develop and refine scalable protocols based on two approaches to recover barcodes depending on DNA template quality, either multiple overlapping short-amplicon sequencing, e.g. (D'Ercole et al. 2021), or low coverage, highly multiplexed genome skimming with reduced volume library preparations, e.g. (Alsos et al. 2020). While the latter skimming approach is not barcoding sensu stricto, it can have higher success rates for barcode recovery from historical material. It is also simpler to implement at scale across diverse taxa as it does not require taxon-specific primers, and can be comparable in cost to amplicon sequencing for highly fragmented DNA, while providing additional sequence data beyond the barcode loci, e.g. (Bohmann et al. 2020).

Both bulk sample and environmental DNA (eDNA) metabarcoding will be used to generate local community composition profiles [T6.2] for pan-European biomonitoring. Standard DNA extraction and PCR amplification procedures will be used, and libraries

will be prepared using automated workstations for optimal quality control for individual libraries and comparability between libraries. Approximately 5000 community sample libraries will be sequenced to an average depth of 1 million 2x 250 paired end reads per sample, using a NovaSeq 6000 platform, compatible with the largest amplicon sizes and species richness for bulk sample and eDNA metabarcoding.

Barcode Data Processing and Analysis (WP8): We will design and implement both functional and performance enhancements in the existing DNA barcoding (meta)data infrastructure and support the processing and management of data generated during this Project. Our ability to scale up operations, as a community of practice, heavily relies on the availability of open, robust and sustainable data infrastructures. The Barcode of Life Datasystem (BOLD) infrastructure is the main focal point for the publication, annotation and subsequent curation of barcode data and metadata for iBOL. The BGE Project provides a unique opportunity to contribute to the further development of the BOLD data infrastructure from a European perspective. Specific issues to address include strengthening the long-term sustainability and resilience of BOLD, enhancing its functionality, and further integrating BOLD with existing European research infrastructures. Such developments will facilitate the data workflows within BGE production pipelines, and also enable the scaling up of those production pipelines beyond the duration of the BGE Project.

We will work closely with the development team of BOLD (in Canada) to establish a European mirror site for BOLD [T8.1], focusing on data publication in Europe, while ensuring complete and continued synchronisation with the Canadian instance. We will run a user community survey and workshop to synthesise user needs and use this to develop a technical roadmap of BOLD outlining and prioritising user requirements [T8.2]. During the Project we will then implement technical enhancements to BOLD [T8.3] to increase functionality and contribute to the delivery of the roadmap. Particular importance will be given to enhancements that provide maximum value to users and support integration with the existing landscape of data and data processing tools in Europe. In addition to our focus on strategic developments of BOLD, we will also facilitate the processing of barcode data as they become available from the partner sequencing centres. This includes establishing pipelines to support specimen-based data uploads to BOLD and the European Nucleotide Archive (ENA) [T8.4], and implementation of novel denoising procedures for quality control of metabarcoding data followed by subsequent data upload to mBRAVE and ENA [T8.5].

(Obj 3) APPLICATION - To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines to improve the efficacy of management interventions and biomonitoring programmes

To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines we will use DNA barcoding data to enable biodiversity characterisation, conservation, and biomonitoring in taxa and systems of high conservation, ecological, and socioeconomic importance (WP10). To deliver this for the DNA Barcoding Stream we have structured our effort into two major strands of work: (i)

the production of a curated European DNA barcode reference library for key taxa, and (ii) the exploitation of metabarcoding data to characterise, conserve, and monitor biodiversity.

To meet the challenge of large-scale reference library curation at such an unprecedented European scale, we have assembled partner institutes with (i) collections and curators representing the broad spectra of taxa and taxonomic expertise needed and (ii) capacity to deliver. These partners will work in a coordinated way to quality control, validate, and taxonomically assign 45,000 Project generated barcode sequences from 15,000 species, together with the curation of an additional 200,000 existing barcode sequences from 25,000 species [T10.1]. To do this, we will develop and implement a reference library curation pipeline involving automated clustering of sequences, followed by classification of matches against a community compiled rule set, e.g. BAGS: (Fontes et al. 2021), and then deploy highly targeted human interventions to reconcile species identity/specimens placement mismatches. To maximise the quality and efficiency of this step, we will draw upon the citizen science expertise of the European amateur natural history societies from Task T2.4. Collectively this will deliver both a scalable barcoding curatorial process and tightly curated European reference libraries for pollinators and key indicator species for freshwater and marine biomonitoring.

To meet the challenge of achieving harmonisation across national boundaries for pan-European biomonitoring, we will develop workflows using bulk eDNA and metabarcoding data and implement and synthesise important use cases. Using the production capacity of WPs 2, 4 & 6 for sampling, sequencing, and data processing, we will design and pilot a diverse set of biomonitoring use cases to evaluate patterns of species diversity and community composition in key systems [T10.2]. These include: (1) A pan-European assessment of pollinator diversity focusing on an easily deployable protocol allowing rapid assessment of pollinator community health. (2) Assessment of the biodiversity richness of sites undergoing ecological restoration by establishing and deploying a standardised sampling of arthropods and soil fungi. (3) An assessment of the impacts on arthropod diversity in sites subject to agricultural intensification. (4) Building a DNA barcoding climate change observatory network consisting of establishing baseline sampling along altitudinal gradients in key mountain ranges across Europe to track biodiversity shifts associated with climate change. (5) A citizen science based eDNA biomonitoring programme to enable early detection of invasive marine species in harbours (supported by Task 2.4). In all cases, we will use the metabarcoding data to characterise communities, detect changes, compute ecological quality indicators and Essential Biodiversity Variables. Drawing on our connections to existing major biomonitoring initiatives (e.g., LIFEPLAN, DNAqua-Net, EU-PoMs), we will synthesise approaches and develop a practical toolkit and guidance for the harmonisation of DNA-based biomonitoring to support convergence towards standardised approaches.

Genome Sequencing Stream

The ERGA work Stream will establish a European distributed infrastructure for generating reference-quality, complete genome sequences for species across the whole of European biodiversity. This process involves a complex, transdisciplinary workflow. Building reference genomes demands the use of a combination of state-of-the-art genomic technologies starting from high-quality samples containing well-preserved DNA molecules. The data generated is processed by various bioinformatics tools and often demands high-performance computing to complete genome assemblies. The functional part of the genome is annotated based on multi-tissue transcriptome data. The resulting annotated reference genome together with the large amounts of raw data that underlie it are made available through public repositories. Such a complex workflow can be long and expensive, and we view it as essential to our success that the European genomic community establishes tight connections to streamline the production of reference genomes through a distributed, harmonised, and expandable infrastructure. To contribute to as well as benefit from global genomic efforts, ERGA will be established as the European node of the EBP and a network of excellence in the field of genomics applied to biodiversity research and conservation, promoting capacity building across Europe.

An atlas of complete reference genome sequences of important European species for agriculture, fisheries, and key ecosystem processes, as well as endemic and threatened species, pests and disease vectors will become a resource to the scientific, industrial, and regulatory communities for the systematic exploration of the genomic landscape of the European biota, the conservation of biodiversity and ecosystem services. These goals are achieved through integrated and complementary activities across five WPs summarised in Fig. 3 and detailed below.

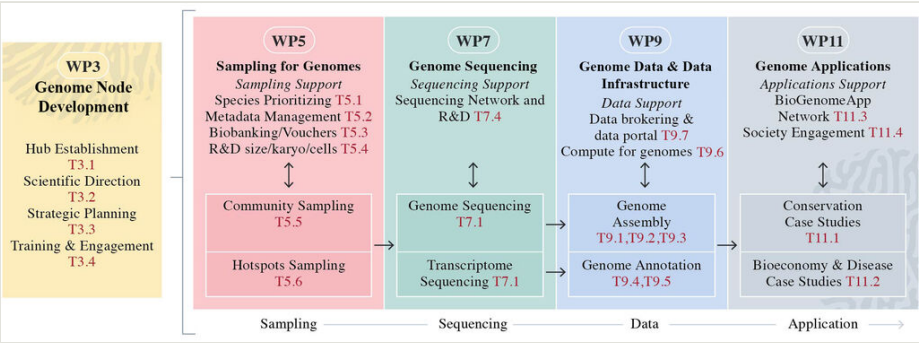

Figure 3. doi

Reference genomes WPs and Task interrelationships.

(Obj 1) CAPACITY - To establish functioning biodiversity genomics networks at the European level to connect and grow community capacity to use genomic tools to tackle the biodiversity crisis

To establish functioning biodiversity genomics networks at the European level we will build the European node for the EBP and connect this node to the international scientific community, policy makers, practitioners, and wider society (WP3). This will leverage the existing ERGA network and coordinate scientific activities, capacity building and training, communication, and public engagement, and planning future needs and networking strategies.

WP3 will support the transition of ERGA from an informal, mutual understanding among research institutions to a formal agreement for the establishment of a European node of the Earth BioGenome Project [T3.1]. This will provide legal and administrative support to further develop and formalise the existing ERGA executive bodies and committees at the heart of the network. Scientific direction for the ERGA community [T3.2] will leverage interactions across scientific committees established to mirror those of the EBP (Sample Collection & Processing, Sequencing & Assembly, Annotation, Data Analysis, IT & Informatics) and Science “plus” committees (Ethics, Social & Legal Issues, Citizen Science & Outreach, Media Communication & Public Affairs, and Training & Knowledge Transfer) while aligning ERGA activities with other worldwide genome sequencing efforts and connecting with the EBP. Working with international partners like the EBP and the Vertebrate Genomes Project (VGP), as well as profiling existing European capacities, networks, and infrastructures will help engagement with regional, and national initiatives and allow us to build a coherent strategic action plan for scaling up reference genome generation [T3.3]. Node development relies on supporting participation in the network through engaging members and the wider community. Emphasis will thus be given to the development, coordination, and support of training and engagement initiatives [T3.4] devised by the ERGA Training and Knowledge Transfer (TKT) and Citizen Science (CS) committees. We will identify scientific priorities and practicalities through extensive consultations with the genomics community to build an effective TKT/CS portfolio, and foster participation and involvement in other European training networks and knowledge transfer initiatives.

(Obj 2) PRODUCTION - To establish and implement large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation and biomonitoring

To establish and implement large-scale biodiversity genomic data generation pipelines for Europe we will align efforts across 20 countries connecting samples to sequencing facilities, to data-management and data-processing workflows, supported by SOPs and infrastructure research and development. Genome sequence production involves three major steps: Sampling (WP5), Sequencing (WP7), and genome informatics for Data Processing and Analysis (WP9).

Sampling for reference genome production (WP5): The main needs with respect to sampling efforts are centred on complementary activities of (i) the logistical support required for community-provided samples to be processed comprehensively, and (ii) developing robust processes for efficient field sampling of European biodiversity. Both require developing sampling protocols, coordination, performing and supporting sample

collection, vouchering and metadata management, which we have begun to explore within the ERGA network. Species selection will comprise an identification of taxonomic and geographic gaps in public repositories and establishment of prioritisation guidelines [T5.1]. Key criteria include maximising phylogenetic and geographic coverage and including species present in areas bearing particularly rich or unique biodiversity (hotspots) or species that are currently identified as important for societal, economic or ecosystem needs (critical biodiversity, pollinators, and ERGA case studies). Sampling of prioritised taxa will be performed in different tasks (T5.5, T5.6, T11.1, T11.2), two of which contain pre-selected species (T11.1, T11.2). The Species Prioritisation task (T5.1) will guide taxa selection for critical biodiversity (e.g. threatened species, vectors of diseases, species providing ecosystem services, wild species of economic importance, invasive species), biodiversity hotspots (areas particularly rich in species, rare species, and/or threatened species), and pollinators. Developing guidance on metadata recording, coordination and supervision of metadata deposition and data exchange [T5.2] is essential, here within T5.5, T5.6, and WP11, and for the metadata repository/portal developed in WP9. All samples will also need to be vouchered and stored as part of a biobanking network [T5.3], complemented by establishing and optimising protocols for genome size estimation, karyotyping, cell culturing and access to frozen viable material [T5.4]. These support activities of WP5 will directly contribute to sample delivery via the ERGA community for “European critical biodiversity” [T5.5], and via BGE Partners for “biodiversity hotspots” [T5.6]. The procedural components and the R&D (Research and Development) activities will together build a solid framework for scaling up sampling in future biodiversity genomics initiatives.

Genome sequencing (WP7): Because technologies and national/regional sequencing capacities are ever-changing it is important to consider not only how to deliver outputs but also how to shape the future landscape so that scaling up can be achieved in a distributed fashion across Europe. We will thus deliver reference-quality genomes from critical European biodiversity, biodiversity hotspots, pollinators, and ERGA case studies [T7.1] while also creating a European network of biodiversity sequencing centres [T7.2] to foster the exchange of expertise across the continent, including supporting core R&D in complex sequencing. High-volume sequencing methods for EBP-standard genome assembly will be implemented on biological samples from T5.5 & T5.6 to deliver a nominal 450 Gbase span encompassing between 350 and 500 species by distributing delivery efforts across BGE sequencing partners. In this way, we will produce the necessary data as well as building capacity through a network of sequencing centres with transfer of know-how through training and parallel activities in WP6, including improvement of existing procedures in DNA processing, long-read and long-range DNA sequencing, and sequencing of transcriptomes.

Genome informatics (WP9): The data analysis and management Tasks are designed to tackle challenges in scaling up and streamlining, standardisation and quality control, as well as community participation and access. For genome generation, this requires developing and maintaining robust computational workflows to produce high-quality genome assemblies and an integrated platform for the collection, aggregation, and

sharing of sample metadata seamlessly connected to the underlying genomic data resources. Development of packaged tools and workflows [T9.1] will deliver solutions to be deployed on heterogeneous compute environments for optimal genome assemblies using different datatypes, where support for the manual curation and quality control [T9.2] for finalisation to chromosome-level and approval of assemblies for public release is critical. A distributed approach ensures broad participation and engagement so processing of data produced by the sequencing centres (WP7) will be coordinated across Project partners and the wider genomics community to deliver high-quality genome assemblies [T9.3]. Similarly, genome annotation requires both community expertise and capacity for high-throughput, so workflows will be developed and deployed leveraging community annotation expertise [T9.4] while working closely with established large-scale data processing pipelines to produce annotations for all assemblies [T9.5] through Ensembl Rapid Release and the ENA, as well as coordinating inclusion of third party annotations into Ensembl. These key components of the value chain - from development, curation, assembly, to annotation will make use of compute resources [T9.6] provided by the European Galaxy server (Afgan et al. 2018) to generate high-quality annotated reference genomes. Acknowledging the complexity of these processes, having to link back to data sources (WP5), and future scaling up endeavours, it is clear that the Project must develop an integrated platform for the collection, aggregation, and sharing of sample metadata [T9.7] connected to the genomic data. This will include development of an ERGA Data Portal to ensure FAIR open access to genome data and a finer grained tracking infrastructure of the sampling and sequencing stages within the COPO data brokering platform (Shaw et al. 2020).

(Obj 3) APPLICATION - To apply genomic tools to enhance understanding of European biodiversity and biodiversity declines to improve the efficacy of management interventions and biomonitoring programmes

To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines we will develop the use of genome sequencing to enable biodiversity characterisation, conservation, and biomonitoring in taxa and systems of high conservation, ecological, and socioeconomic importance. While the use of genome data in biodiversity research and conservation has been steadily advancing, large-scale routine exploitation of reference genomes is still largely an ambition rather than a reality. In WP11 we therefore aim to establish a network of excellence that will drive this ambition forwards using concrete case study examples under two major subjects: (i) systems of conservation concern and endemism, tackling situations of steep population declines and the understanding of the adaptive dynamics of endemic species [T11.1], and (ii) systems of bioeconomic importance for which the provision of ecosystem services is threatened by disease [T11.2]. Integrating genomic approaches in biodiversity conservation and research, developing, and evaluating applications with cutting-edge standards, promoting communication and showcasing the use of reference genomes in biodiversity applications will be supported by the formation of a dedicated BioGenomeApp network of excellence [T11.3]. This also requires concerted efforts to build bridges between genomic research, conservation practitioners, other relevant

stakeholders, and the public via the case studies and citizen science actions to engage society with genomics approaches to understanding biodiversity [T11.4]. This programme addresses a recognised gap between the analysis of genomic data, predominantly addressed by basic research, and the full implementation of these tools by conservation practitioners and stakeholders.

Consolidated effort: the interface of the DNA Barcoding and Genome Sequencing Streams

The Tasks of this joint Stream Work Package (WP12) are designed to connect expertise from the BIOSCAN Europe and ERGA networks in combined actions to capitalise on the benefits of working together. They recognise the potential for economies of scale and the urgent need for standardising activities and harmonising data management to enable scaling up of biodiversity genomics. They represent concrete actions that bring the two communities together immediately to address imminent challenges and plan for an increasingly integrated future.

(Obj 1) CAPACITY - To establish functioning biodiversity genomics networks at the European level to connect and grow community capacity to use genomic tools to tackle the biodiversity crisis

Developing capacity across the BIOSCAN Europe and ERGA networks will be achieved by supporting interactions between networks to address current geographic and methodological disconnects. We will run joint network training activities [T12.1] with a focus on establishing European pollinators as a model for joint ecosystem sampling [T12.2] to connect and grow the biodiversity genomics community. Training events will disseminate the protocols, procedures, methods, and expertise developed within the Project to the wider community. These events are designed to provide training to record metadata and preserve specimens for future reference genome sequencing, while also carrying out onsite barcoding using the portable ONT sequencing technology. These events will bring parties together, including wider engagement by coordinating with citizen science Tasks [T2.4], especially around the common agro-ecological theme of coordinated assessments and monitoring of European pollinators.

(Obj 2) PRODUCTION - To establish and implement large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation, and biomonitoring

We will exploit the complementary skills of the iBOL and EBP communities in our Project to tackle scientific challenges at the interface of DNA barcoding and genome sequencing. This includes the use of reference genomes to design high-resolution multi-locus assays for species identity (focusing on plants where standard barcodes often fail), and for rapid and cost-effective assessment of intraspecific species variation (here focusing on insect pollinators) [T12.3]. We will also tackle the practicalities of securing sample supply chains for barcoding inventories and reference genome sequencing for the difficult-to-identify diminutive organisms in species-rich groups that represent a large proportion of the

unknown diversity of life on Earth. To address this, we will develop workflows that barcode and preserve 'dark taxa' [T12.4], in species-rich arthropod groups from underrepresented European regions. This includes research and development towards sampling protocols that preserve specimen morphology while releasing sufficient quality and quantity DNA for both barcoding and genomes. Barcoding dark taxa will use the high-throughput of the PacBio Sequel II platform to generate barcodes for 50,000 specimens. These barcodes will permit clustering of specimens into barcode sequence clusters, to enable downstream taxonomic annotation and ultimate integration into reference libraries. More generally, to scale up production and enable an integrated future landscape for biodiversity genomics in Europe there is a need for alignment and enhancement of (meta)data processing across the two networks. We will build on existing data management infrastructures to develop an integrated framework supporting both barcoding and genome generation workflows [T12.5; T12.6], including data brokering for processing, validation, and submission of sample and sequencing metadata, as well as informatics solutions to harmonise across data portals and facilitate change management: a FAIR biodiversity data ecosystem.

(Obj 3) APPLICATION - To apply genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines to improve the efficacy of management interventions and biomonitoring programmes

Characterising intraspecific genetic diversity and inter-specific divergence is essential for effective biomonitoring and advancing assessments of pan-European biodiversity. We will build on the multi-locus assays developed in T12.3 to operationalise high-resolution biodiversity assessment tools. This includes establishing and employing tools for capturing intraspecific genetic variation [T12.7] focused on pollinators co-occurring across European bioregions (sampled in T12.2), to identify neutral and potentially adaptive genetic diversity, and genetic markers that are informative for enhancing species discrimination. It also involves using reference genome derived multi-locus assays (from T12.3) to optimise efficient multi-locus barcodes [T12.8] to increase resolution in key groups, using plants as an exemplar to guide future developments. These application Tasks will thus exploit the added value of consolidating efforts at the interface of barcoding and genome sequencing for biomonitoring.

Supporting WPs

Management (WP1): The management of the BGE Consortium is delivered through WP1. Effective management of the Project will rely on (a) a clear governance structure with agreed responsibilities and authorities for each of the governance bodies, (b) clear lines of reporting, and (c) agreed Project SOPs that include quality control of all outputs. The Project will be led by a Project Director and two Project Deputy Directors, ensuring the connection of the Project leadership with the two networks (BIOSCAN Europe and ERGA). We approach the Project management in a way that ensures alignment to the structure of the Project (See Section Project structure above). As such, each Stream is led by a Consortium partner, who provides guidance and ensures the connectivity of WPs

across the Stream and is also responsible for the leadership of the corresponding Node Development WP (WP2 & WP3). Each Pillar is coordinated by a Consortium partner, who ensures the thematic coherence of the included WPs and represents the Project. A detailed management structure has been drafted and will be included in the Consortium Agreement. When possible, the Project will make use of virtual collaboration tools to minimise the need for onsite meetings. To maximise the overall impact of the Project, we will establish an external Science & Impact Advisory Board to provide guidance to the Project management team, and to also support the implementation of WP13. The Science & Impact Advisory Board (SIAB) is external to the project consortium and will be commissioned to execute the following functions: a) Provide expert advice on key topics related to the scientific direction of the project and especially around items where decisions must still be taken (for example in definition of taxonomic and geographic priorities for field sampling), b) provide advice on the topics linked to the sustainable development of tools and services, the expected scientific impact and adoption by the community, c) act in a peer review capacity, participating in the quality review of major project outputs and oversight of quality assurance processes, and d) provide guidance on topics related to policy implementation and compliance (including DSI). The terms of reference of the SIAB and its exact position in the governance and quality control of the BGE outputs will be defined in the BGE Consortium Agreement. The Project management team will manage and maintain oversight of subcontracting activities within the project.

Communications (WP13): Communications for the Project is delivered by WP13, which focuses on three main areas of work, namely (1) Communications and publicity [T13.1], (2) Dissemination of scientific outputs, and their subsequent translation into non-technical summaries [T13.2], and (3) Exploitation of knowledge produced during the Project and connecting this knowledge to key environmental stakeholders [T13.3] including policy makers and conservation organisations.

Policy and Ethics (WP14): The upscaling of biodiversity sequencing in Europe requires the establishment of a clear policy framework to guide the working practices of the BGE Consortium. This will involve developing the policy schema under which massive scale European sequencing will operate, while facilitating streamlining information on existing policies (EU and global) to understand their implementation in practical terms for flows of data and samples. Regulatory issues to address include compliance with the legislative frameworks involved in collection and holding biological samples, adherence to Access & Benefit Sharing (ABS) regulations, and appropriate consideration of digital sequence information (DSI) in light of ongoing negotiations under several multilateral environment agreements. Ethical issues to address include avoidance of harm, and animal welfare for sampling of sentient organisms, particularly vertebrates, inequalities of accessibility to biodiversity genomic technologies associated with wealth inequalities, and promotion of diversity and inclusivity for internal and external stakeholders. To address regulatory and ethical issues, we will create a web-based matrix with policy components and requirements towards the establishment of a Genomics Table of Policies (GeTAP) [T14.1]. A Social Innovation Team, running as a multidisciplinary advisory panel, will be set-up to identify methodological approaches and working practices that ensure legislative

compliance, and best ethical practice [T14.2]. Supported by the Project advisory board, Project risks will be monitored in the wider policy landscape (such as changes in DSI legislation) [T14.3].

Emphasis on value, not volume

In creating this proposal, we made a calculated decision to invest in connecting people and establishing best practices for biodiversity genomics. This comes at an initial cost of generating lower volumes of data, but it best positions us to successfully scale over the coming years because the foundational principles and networks will have been established through this application. This approach thus establishes the foundation for scaling the production of barcode reference libraries and reference genomes in Europe and a coherent deployment of genomic tools for biomonitoring and biodiversity conservation.

Interplay with national and international research and innovation activities

BGE Project collaboration with European Research Infrastructures (RIs)

Despite the highly aspirational nature of the BGE Project to design, establish and pilot, for the first time, a mechanism for the coordination of barcoding and genome sequencing efforts across Europe, we are not developing this in isolation. The BGE Project takes stock of the existing landscape of operating RIs. We do that in relation to national, European (e.g. ESFRIs and ENA) and international resources. By the end of the Project, we expect to have established a strong strategic and, to the extent possible, operational connection between BGE and relevant infrastructures. Specifically, BGE will work towards developing a strong interoperability framework between existing genomic and barcoding data systems. BOLD, ENA, ELIXIR, and DiSSCo will work together to ensure seamless meta(data) pipelines from capture to reuse in a FAIR and open manner. We will further develop capacity around BOLD and a bespoke reference genome metadata portal by connecting them to the wider European ecosystem of e-infrastructures, making use of public Cloud services for storage and computing. Finally, we will link up data generated by BGE to a multitude of other relevant taxonomic, geographic, image, and trait data by making use of the DiSSCo digital specimens' infrastructure. We will benefit from existing global resources like Catalogue of Life - CoL (for taxonomic information) and GBIF (for occurrence data).

Leveraging national projects and their results

In order to deliver impact at the scale envisioned it is crucial that we make best (re-)use of all available distributed resources and outputs from existing projects that run at national level. The BGE Project is geared towards delivering the necessary interconnections between existing investments in infrastructures, expertise, and data, rather than duplicating effort. As such, we have identified and connected to, also through our existing Consortium members, several national projects (Table 2) that (are expected to) deliver relevant outputs. Their connections and integration with BGE will be coordinated via WPs

2+3 for network-specific activities, and more generally, via WP1 for cross-cutting activities.

Table 2.

Examples of national funded projects or initiatives that are linked to the BGE Project.

Country	Linked project or initiative
Estonia	DiSSCo Estonia; UNITE
Finland	LifePlan
Germany	German Barcode of Life
Italy	Endemixit
Netherlands	ARISE Infrastructure
Norway	Norwegian Barcode of Life; EBP-Norway
Portugal	BioGenome Portugal
Spain	Catalan Initiative for the EBP (CBP)
Sweden	Swedish Earth BioGenome Initiative
United Kingdom	Darwin Tree of Life; BIOSCAN-UK; UK Barcode of Life

Data management plan and Open Science Practices

Data management principles: The BGE Project adheres to the principles of producing and mobilising open and FAIR data (Table 3). To coordinate data management within the Project, BGE will develop a guiding Data Management Plan (DMP). The DMP will specifically cover the handling of research data during and after the Project, data collection and processing; methodologies and standards; data sharing and open access; curation and preservation. The DMP will also provide the dataset metadata specification that will be used in the data registry, following an appropriate, relevant standard. It will specify the recommended licensing schemes, preferably using the Creative Commons Public Domain (CC0) and Attribution (CC-BY) licences. In the cases where the datasets cannot be publicly shared, the reasons will be mentioned in its metadata description (e.g., ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Where appropriate, we will make special provisions for data that fall under the regulatory frameworks linked to ABS regulation. Such provisions will be made in collaboration with experts involved in WP14. The BGE DMP will become part of the Project handbook and be used as the data management guiding document for all BGE Partners. At its core, our DMP will support the reproducibility of our scientific results, and act as clear guidance of how the Project beneficiaries can publish Project outputs using available tools and services.

Enabling access to other Project outputs: Software Code: Newly generated software code will be deposited in public repositories and documented in a way that makes it reusable

by others. The licence policy for software code will abide by the Open Source Initiative licences (e.g. Apache Licence 2.0, GNU General Public Licence (GPL) or MIT licence). Further specifications will be provided in the Project DMP. Open Access Publications: All publications produced by the beneficiaries of the Project and in the context of its work programme will be published as open access. Other Project outputs (including models, workflows etc.): All Project outputs will be deposited in open public repositories (e.g., Zenodo) under a collection of outputs that represent the Project.

Table 3.

Implementation of the FAIR data principles in the BGE Consortium.

FAIR produced data(sets)	**How the Project responds (further detailed in Project DMP)**
Findability	Datasets and data records produced will always be assigned actionable Persistent Identifiers (PIDs). Specific data types will be deposited to thematic repositories to ensure their optimal discoverability.
Accessibility	All data outputs and publications will be open by default. The DMP will further refine cases where embargo periods or access restrictions (e.g., location of protected species) are required.
Interoperability	We will make optimum use of available community (meta)data standards that refer to the use of data types. This can include standards such as DwC (occurrence metadata), MIxS (family of standards for genomic information), IIIF (image metadata). DMP will further specify the list of standards.
Reusability	Licence policy for datasets will include Creative Commons licences CC0 and CC-BY. More restrictive licences for other outputs can be applied. Criteria will be specified in the DMP.

Co-creation and citizen involvement: This Project will engage citizen scientists in the co-creation of knowledge in two main ways. Firstly, a major strand to our citizen science involvement will be engagement with amateur expert natural historians. To support the flow of expertly verified samples into the Project we will work with natural history societies across Europe to collect and identify samples for DNA barcoding, harnessing the taxonomic skills and knowledge of amateur natural historians [T2.4]. This sampling component of the work will be followed up with online workshops to enhance the curation of DNA barcode reference libraries [T2.4] working with amateur natural historians across Europe to update species' identities in light of DNA barcoding data. Collectively these activities aim to utilise the expertise of amateur natural historians to drive forward the completion of DNA barcode reference libraries in Europe, and to feedback data and findings to enhance the knowledge and understanding of participating citizen scientists.

The second major strand of citizen engagement by BGE is with the broader public. This capitalises on the extensive and growing public interest in biodiversity conservation and the effectiveness of DNA sequencing technologies as a connection point for citizen science participation. We will engage the public via a citizen science marine biomonitoring programme [T2.4], in which we will harness public participation in the collection of water samples for eDNA barcoding to support the detection and control of marine invasive species in and around ports and harbours. We will also run BioBlitzes

collecting and processing samples suitable for DNA barcoding and genome sequencing [T2.4], engaging citizen scientists with the biodiversity in their local area and introducing them to technological innovations in biomonitoring.

Collectively these activities aim to utilise the efforts of public participants in the Project in collecting and prioritising samples for DNA sequencing, and in turn to enrich participant connections with nature by feeding back insights into the natural world through the optic of DNA. To enhance the dynamic interactive nature of this co-creation, we will establish a feedback mechanism to directly solicit public views to guide future sampling of biodiversity for genomics research via a citizen science mobile app [T11.4]. We will follow the ten Principles of Citizen Science Engagement (Robinson et al. 2018), and all our activities will be supported by online resources to ensure accessible information flows and feedback on the Project's findings. We will also run workshops using hand-held DNA sequencers to demonstrate biodiversity sequence data generation [T12.1]. These workshops will be aligned to BioBlitzes and will bring together citizen scientists and early career researchers to explore the field of biodiversity genomics and support learning and exchange of different perspectives and experiences.

Provisions for gender balance

The Project Consortium will commit to promoting gender equality in research and innovation as set out in the European Commission (EC) Gender Equality Strategy for 2020-2025 (European Commission 2020), throughout all phases of the Project's implementation. Oversight of our performance, statistics, and strategy for addressing gender inclusivity and wider diversity and inclusivity will be undertaken by the Social Innovation Team, established in WP14.

Gender balance in scientific content: The only sex or gender dimension that is meaningful for the scientific content of the biodiversity genomics data we generate is that where possible and relevant, we will seek to sample the heterogametic sex for reference genome data generation (e.g., the XY or ZW sex) to provide assemblies for both sex chromosomes.

Gender balance in leadership and decision-making: The gender distribution regarding scientific management of the Project has resulted in 43% (13/30) female institutional leads. Likewise, our coordination committee overseeing the activities that have led to the Project proposal comprises three women (of 7). We will continue to monitor and ensure high representation of female leads as the Project progresses and new opportunities for leadership arise.

Gender equality in recruitment: We have solicited information from every Project partner on their institution's commitment to gender equality in hiring and pay. Some institutions have mandatory unconscious bias training that must be completed by interview panel members before shortlisting and interviews commence, some institutions have pay reviews identifying gender pay gaps and addressing them through increased pay. While it is not possible to standardise recruitment procedures and requirements across 30

institutions from 20 countries, we will share the measures being taken at various partner sites across the Project and encourage adoption where possible.

Gender balance in training provision: At a Project level, we will evaluate the composition of our researchers, ensuring that high diversity (including gender, age, ethnicity) and inclusivity is aimed for and alert where teams are skewed towards a lack of diversity. We commit to producing no multi-author (> 3) publications that are 100% male, ensuring engagement at an authorship level from women on all projects that will lead to a publication.

Gender awareness in community engagement: Our citizen science events will emphasise our interest in reaching underrepresented groups, and we will monitor representation through anonymised surveys prior to and after events to evaluate our reach and refine our approach should we observe a lack of diversity in those engaging with Project events.

2. Impact

2.1 Project's pathways towards impact

The BGE Project is a foundational joint step towards achieving shared capacity across Europe for delivering a scalable genomics system to deliver the production of knowledge for effective biomonitoring, biodiversity conservation, and species discovery (Fig. 4). We address this challenge as follows:

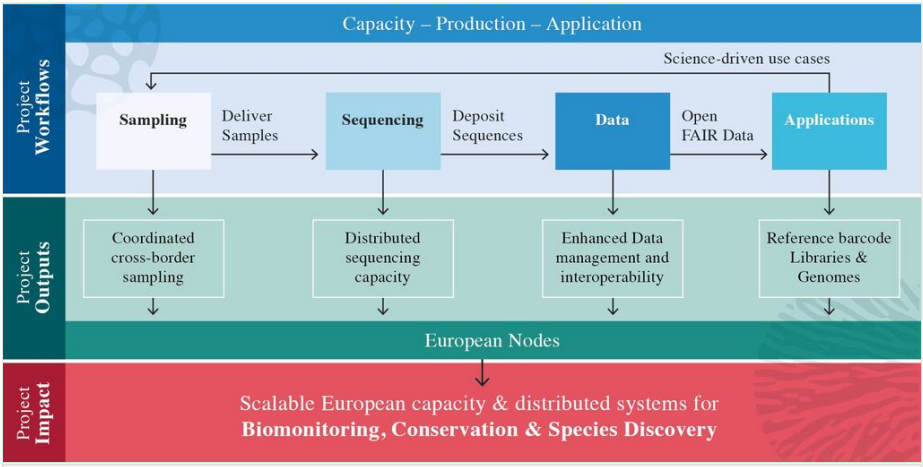

Figure 4. doi

Project value proposition, describing how Project workflows lead to outputs that generate longer-term impact.

- We establish the required capacity by connecting people, data, and infrastructures, bringing together currently fragmented activities, into a common

programme of work. This will be accomplished by building strong European nodes for the two networks, and taking the first steps to establish operational collaboration between the two communities of DNA barcoding and genome sequencing researchers;

- We demonstrate the feasibility of the cross-border and cross-community content production through establishing large-scale scientific content delivery through distributed facilities and teams using shared workflows, shared toolkits, and resources generating a pipeline from field → lab → data → application;
- We enhance knowledge of pan-European biodiversity, by the application of genomic tools to intra-specific diversity, species discovery, species identification, and monitoring biodiversity change and dynamics in taxa and systems of high conservation, ecological, and socioeconomic importance.

BGE response to Topic's objectives - Impact

The Project brings together BIOSCAN Europe / iBOL with ERGA / EBP for the first time at this scale, establishing an inclusive and aspirational partnership that will create a step-change in the European landscape of biodiversity genomics. As such, the Consortium will address all six objectives listed in the Topic description, as follows:

*EC Topic Objective 1: Creation and management of the European node of the International Barcode of Life*How we respond: We establish a European node for iBOL -- called BIOSCAN Europe (iBOL Europe 2025) -- as a focal point for coordinating and strengthening DNA barcoding activities in Europe. BIOSCAN Europe will develop an operational model designed for sustainable continuity beyond the lifetime of the Project. BIOSCAN Europe will have a gender-balanced governance structure and be supported by a Secretariat (with responsibility for coordination of activities, including organisation of training and engagement events), a barcode information portal (as a communication resource for the DNA barcoding community in Europe), and will work with national nodes to build in-country capacity, and with the international barcoding community to build global momentum.

EC Topic Objective 2: Creation of a European hub affiliated to the Earth BioGenome Project…How we respond: We establish a European hub (node) for the EBP -- called ERGA (ERGA 2025) -- as a focal point for coordinating and strengthening reference genome sequencing activities in Europe. ERGA will establish frameworks for communication, expertise sharing, and connections to funders and policymakers. ERGA will promote interactions among scientific, ethical, social, citizen science, media, public affairs, training, and knowledge transfer communities to align ERGA activities with other genome sequencing efforts around the world. We will engage with regional, national, and global initiatives to build an action plan for scaling up reference genome sequencing and continue to build global momentum through the EBP.

EC Topic Objective 3: Development of the necessary networks, technologies, quality standards, reference atlas and taxonomic expertise through Europe…How we respond: We enable the large-scale production of biodiversity genomic data across Europe by

establishing a common (meta)data framework, sharing vetted and robust SOPs, and creating a flow of samples and data, from the field and museums, through DNA sequencing facilities and data processing and management pipelines. Through these activities, we will serve data to a large community of researchers working at the interface of genomic science and biodiversity monitoring and conservation.

*EC Topic Objective 4: Advances in the assessment of pan-European biodiversity via genome sequencing and/or DNA barcoding of threatened/endangered species...*How we respond: We undertake assessments of pan-European biodiversity through (a) establishment of high-quality, well-curated reference libraries to underpin DNA-based biomonitoring, (b) delivering exemplar DNA-based biomonitoring of key habitats encompassing terrestrial and aquatic ecosystems, (c) delivering high-throughput DNA barcoding of 'dark taxa' to provide insights into the diversity and distribution of poorly known taxonomic groups, (d) developing high-quality reference genomes of key European taxa, and (e) delivering focal case studies building on the reference genomes to assess intraspecific diversity in species of conservation, ecological, and socioeconomic importance.

*EC Topic Objective 5: Pan-European barcoding of pollinators by completing the Barcode of Life for European bees, butterflies, moths and hoverflies.*How we respond: We perform a gap analysis and subsequently build a high-quality, well-curated reference library for European pollinators including European bees, butterflies, moths, and hoverflies, generating new barcode sequences from existing expertly verified museum specimens and freshly collected material, augmented with critical evaluation and curation of barcode records in existing public databases.

*EC Topic Objective 6: The active support and cooperation of citizen scientists and other non-professional taxonomists.*How we respond: We engage citizen scientists with biodiversity genomics research to support the co-creation of knowledge and enhance the speed and quality of data generation. This will involve engaging expert amateur natural historians with collecting and identifying samples for sequencing and incorporating their expertise into the interpretation of the resulting data. We will also engage the wider public with BioBlitzes and sampling for biomonitoring and incorporating their perspectives and views in the setting of future research priorities.

Measurable Project outputs - KPIs

The table below (Table 4) points to the major tangible and measurable results. These outputs will also act as Key Performance Indicators (KPIs) for the Project and be used to evaluate progress across the linked Tasks. Our Milestones are purposefully drafted to monitor progress across these KPIs.

Table 4.

Major tangible and measurable Project results.

Type of Project result	Measurable / Tangible key Project Results
Data	• 450 Gbp of annotated, curated reference genome assemblies encompassing 350-500 species; • 45,000 specimen barcodes and (meta)data; • 50,000 'dark taxa' specimens barcoded, with tissue stored for genome sequencing; • Curated European Barcode reference libraries (pollinators, freshwater & marine species); • 5,000 metabarcode libraries sequenced for biomonitoring; • Metadata of ca. 100,000 archived Project samples and DNA extracts.
Software / Services	• Enhanced European BOLD systems; • Reference genome public data portal; • Open repository for protocols and SOPs for biodiversity genomics.
Roadmaps	• Planning for ERGA and BIOSCAN Europe legal entities; • Design study and roadmaps (short-, mid- and long-term) for scaling up European level mechanisms for biodiversity genomics.
Policy Recommendations / Best Practices	• SOPs for sampling, sample processing and metadata collection; • Scientific guidelines for distributed biodiversity genomic practices; • Pipelines for the bulk data quality assessment & assurance; • Best practices for FAIR (meta)data management and data; • Workflows for intraspecific genetic diversity monitoring tools, and multi-locus barcoding; • Ethics compliance guidelines.
Skills & Knowledge	• Project Knowledge Platform; • Community training courses (online and on-site); • Establishing a network of sequencing centres and interfacility training and knowledge transfer.
(pre-)standards	• Material collection metadata standards; • Biobanking standards.
Publications	• Methodological publications sharing protocols for delivering biodiversity genomics at scale; • Data release papers describing the publication of reference barcodes and reference genomes; • Biomonitoring publications focusing on tracking biodiversity change using metabarcoding; • Showcase of the application of genomics in conservation; • Application of genomics in Bioeconomy and disease control.

Long-term & broader impact of the Project

Barcoding - Project impact to science & society

Biodiversity discovery: The delivery and curation of c 45K reference library specimens (culminating in T10.1), along with c.50K 'dark taxa' specimens [T12.4], and 3,500 metabarcoding samples [T10.2] will lead to the discovery of numerous new species by detecting lineages divergent from known taxa. Likewise, these data from across Europe will increase record density and add knowledge to species distributions. This discovery element of the Project is important as conservation programmes, and policy interventions aiming to address biodiversity loss are fundamentally based on knowledge of which species exist, and where they occur. In many species groups, particularly in arthropod orders like Diptera and Hymenoptera, previous DNA barcoding studies have detected substantial numbers of cryptic species. Indeed, large-scale DNA barcoding studies have led to the detection of species new to science across all the major clades of multicellular life, including fungi, plants, animals, and stramenopiles. This Project, and the subsequent legacy impact of a large-scale coordinated DNA barcoding network in Europe, will make a substantial contribution towards species discovery and distributional knowledge for the European biota.

Biological identification: By establishing robust well-curated reference libraries for ecologically important species guilds (i.e. pollinators) and systems (i.e. indicator species for freshwater and marine systems), we will make a substantial contribution to the species identifications required to underpin assessments of ecosystem health in key terrestrial, freshwater, and marine habitats. A commonly overlooked and under-resourced element of DNA barcoding is establishing high-quality reference libraries based on verified biological specimens. Without a good reference library, sequence data remains uncoupled from the existing taxonomic framework and associated biological knowledge. Thus, biomonitoring programmes lacking reference libraries are often disconnected from existing knowledge on species and species traits; an issue we will directly tackle in this Project.

An additional substantial outcome of this Project is the development of workflows for multi-locus barcoding methods and new approaches to achieve species-level resolution in organismal groups where standard DNA barcoding methods fail. We will deliver this by focusing on developing multi-locus barcodes for plants and use these to establish the conceptual framework for integrating multi-locus barcoding into the existing standardised barcoding workflows. This will make a significant contribution to our knowledge and ability to tackle biodiversity monitoring in Europe. The inability of current plant barcodes to distinguish species in many genera is a major barrier for DNA-based-identification and biomonitoring of the dominant biomass component of terrestrial ecosystems.

Improved biomonitoring: The standardised biomonitoring toolkit and workflows that we will establish in this Project will make an important contribution towards the harmonisation and standardisation of DNA-based biomonitoring in Europe. By drawing on the diverse systems used in this Project and the other major biomonitoring

programmes which we collaborate with (LIFEPLAN, DNAqua-Net), we will provide generalised guidance aiming to move DNA-based biomonitoring in Europe towards standardised approaches to track biodiversity change. This will represent a significant contribution towards biodiversity conservation and policy implementation.

Genome sequencing - Project impact to science & society

Biodiversity discovery: The delivery of a nominal 450 Gbase of genome span encompassing between 350 and 500 species will make an important contribution towards the discovery and characterisation of eukaryotic genetic diversity in Europe. The annotated reference genomes will provide the foundation for assessing diversity within and among taxonomic groups, with unprecedented detail, including the underpinnings of functional variation. Combining these with population resequencing data will showcase how reference genomes are key components of the discovery and detection of species or genetically distinct varieties, and support decision-making for habitat restoration and resource allocation. The outputs contribute to discovering and cataloguing intra- and inter-specific genetic diversity, which is essential to the development of applications that are designed to conserve biodiversity and ensure the sustainability of ecosystem services, while anchoring the development of new cutting-edge monitoring tools. The discovery of new, including functional, genomic variation will substantially contribute to the development of important assessment tools and help connect the research network to stakeholders.

Biodiversity dynamics: Reference genomes enable the high-resolution reconstruction of species evolutionary histories, including the recent demographic dynamics influenced by anthropogenic environmental change, new disease pressures or overexploitation. It also provides power to detect and understand local adaptation, which is fundamental to guide conservation and management actions. Each reference genome generated in the Project will provide the foundation to understand and characterise biodiversity dynamics across a variety of key taxonomic groups and keystone species. Our reference genomes will anchor the development of biodiversity applications aimed at understanding adaptations, characterising and quantifying sharp species declines (including those due to overexploitation and stock management), and the impacts of diseases affecting wildlife. This will inform the development of management and conservation actions to prevent species declines and promote restoration.

Biodiversity monitoring: Genomics is already widely applied to address ecological questions that are central to understanding the functioning of ecosystem communities and to tackle threats, often caused by human activities. The Project outputs will contribute substantially to the taxonomic coverage, data quality, and (meta)data interoperability for reference genome catalogues of European species. This will greatly improve the effectiveness of applying genomics techniques to biodiversity monitoring in space and time, including increasing the accuracy and resolution of metagenomic studies quantifying community composition, and measuring intraspecific genetic diversity and functional potential, both of which depend on the availability of high-quality annotated genomes with accurate and comprehensive (meta)data. This will create novel avenues

for the development of cutting-edge monitoring tools that in the future can be replicated across taxonomic groups and ecosystems.

Joining forces & establishing solid and sustainable connections

The impact of the Project's joint activities will be to bring together members of traditionally distinct research networks to develop solutions that align currently largely disconnected protocols, data management strategies, and applications in biodiversity and conservation. The DNA Barcoding and Genome Sequencing Streams address the key outputs of biodiversity discovery, identification, dynamics, and biomonitoring in a coordinated and complementary manner. Tasks at the interface contribute to these outputs, but with a strong focus on exploration and development aiming to set the community up for an increasingly integrated future. For example, fully consolidating sampling efforts across the two networks at this stage would be unworkable given different requirements for sampling, namely the time- and cost-intensive requirement for maintaining cold chain protocols to preserve high molecular weight DNA for reference genome sequencing, which are currently incompatible with the requirement for much larger sample sizes of individuals and species for building DNA barcode reference libraries. Nevertheless, our joint activities will establish sampling protocols for both genome sequencing and DNA barcoding to pave the way for a truly united future sampling strategy. With respect to data management, the Project recognises the urgent need to develop an integrated framework supporting both barcoding and genome generation workflows. Alignments and enhancements of (meta)data processing across the two networks will facilitate the scaling up of production and help to coerce the inputs (sampling) and outputs (downstream data use and reuse) towards greater standardisation and harmonisation.

Towards a distributed European infrastructure

The BGE Project will design and pilot the interoperation of organisations that now individually produce scientific content on biodiversity genomics. As such, it will essentially set the foundations for a pan-European distributed scientific system for continental scale delivery of genomic knowledge. We will pilot these connections against the needs of specific biodiversity applications (WP10, WP11), which are used as case studies for the organisation of the European shared production pipelines. At the end of the Project, we anticipate achieving solid proof-of-concept of how our European collaboration can deliver more than the sum of its parts, and a good understanding of the further steps that would enable a fully integrated systemic collaboration model and a European distributed infrastructure.

Impact on national & European policy frameworks

The longer-term aspiration of BGE is to enable a pan-European mechanism for scaled and high-resolution biodiversity data. Such capacity can produce a significant volume of high-quality genomic information that can plug into national metrics and increase the reliability of currently used biodiversity indicators. As such, BGE as a Project and as an

initiative can have a significant impact in the future development of national policy frameworks for biodiversity. Furthermore, we expect that the data generated by BGE will be annotated in such a way that can be included in the registry of resources used by the European Biodiversity Knowledge Centre for further informing biodiversity indicators at European level.

Opportunities for wider industrial involvement

The defragmentation of the European biodiversity genomics landscape that BGE works towards could have a positive effect on the collaboration frameworks with industrial partners. Sequencing technologies are primarily driven by industrial or human health sectors. The consolidation of the European biodiversity community will underscore to industry the urgent needs for innovation in the sequencing sector. As a result, this may lead to a more coherent market, where procurement of sequencing hardware or services is done in a coordinated, efficient way and to incentivisation for more industry-driven innovation and cost reduction driven by a larger market.

2.2 Measures to maximise impact - Dissemination, exploitation, and communication

We have structured the BGE Consortium to establish a knowledge exchange flow from individual WPs all the way through to stakeholders (Fig. 5), blurring the customary divide between internal and external communications. Our knowledge exchange flow encompasses data and knowledge generation from individual work packages (WP4-12); synthesis and knowledge exchange with the specialist communities of iBOL and EBP via their respective European nodes (WP2-3); and Consortium-wide communication and engagement with a diverse set of stakeholders, including the public, environmental agencies, policymakers, and conservation bodies via a dedicated communication, dissemination, and knowledge exploitation Work Package (WP13).

This structure is designed to accommodate the distinctive elements of iBOL and EBP, and the importance of building lasting knowledge exchange networks for their respective European nodes, while at the same time maximising synergies from coordinated approaches and harmonisation of efforts. WP13 fulfils this role of achieving synergy and acts as an amplifier and conduit for the European nodes of iBOL and EBP, and provides the space and mechanisms in which experts in media communications, science communications, and the science policy interface, will work with research scientists and the European node coordinators of iBOL and EBP to connect the work of the Consortium to stakeholders.

Communications: To publicise and communicate the work and findings of the BGE Consortium, a Communications Strategy will be developed at the outset of the Project [T13.1]. This will involve identifying and targeting the wide range of relevant audiences, including existing networks of biodiversity and genomics researchers, key stakeholders such as European governmental representatives and funding agencies, environmental agencies, and conservation organisations, mainstream press and media, citizen

scientists, including amateur taxonomists and natural historians, and the wider general public. The delivery of the Communication Strategy will include establishing a coherent brand and communication style for BGE to maximise visibility and promote understanding of its activities, while also accommodating the need to tailor communications to different audiences. The communication activities will use the BGE Knowledge Platform (developed in T13.2) as a baseline resource for content, and support this with a broader suite of communications tools to ensure effective provision of information (i.e., the main Project website, newsletters, press releases, social media, printed materials, videos, and podcasts, etc.). High priority will be given to accessibility and inclusivity in determining the mode and style of communications.

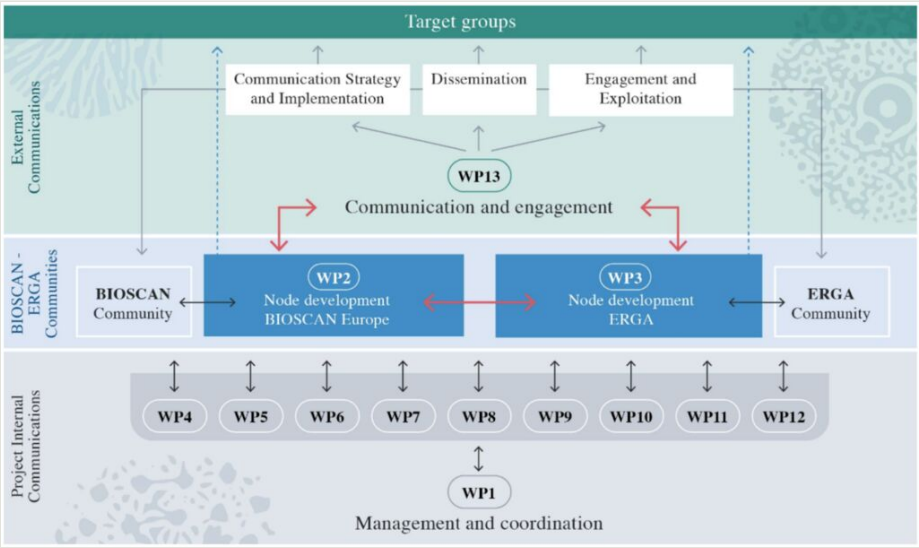

Figure 5. doi

Contribution of WPs to the different levels of communication (internal, community and external) of the Project. Arrows point to interrelationships and flows of information between WPs for effective coordination of the communication strategy.

Dissemination: To share the technical findings of the Project, and to translate key findings into non-technical summaries we will establish, populate, and maintain an online Knowledge Platform [T13.2] as part of the BGE website. The content for this Knowledge Platform will be co-created by specialists in science communication from WP13, who will work closely with WP2 (the European node for iBOL) and WP3 (the European node for EBP), to translate the technical findings from the Project into accessible summaries suitable for policy and practitioner audiences. The Knowledge Platform will also create a space for interactive scientific discussion via the BGE social media accounts.

Engagement and Exploitation: To connect the Consortium's scientists with key stakeholders and guide exploitation of the Project's findings we will develop and implement an Engagement and Exploitation Strategy [T13.3]. This will ensure that stakeholder needs are characterised, understood and serviced, and that the findings of

the Project are exploited to maximise impact. This will include mapping and understanding the stakeholder landscape and incorporating aspects such as their relative “knowledge distance” with the field of genomics research. It also involves identifying the most effective connection points with national and international bodies, and facilitation of the participation of BGE Partners in relevant fora at the science-policy interface. A key element to WP13 is the identification of exploitable results, target groups, and pathways tailored to each of them. Of particular importance is the exploitation of developments of methodologies, processes, and pipelines that support standardised characterisation and biomonitoring of biodiversity, along with specific applications directly linking case study examples to drivers of biodiversity change. The Science & Impact Advisory Group (established in WP1) will provide a mechanism for guiding the translation of the Consortium’s work to policy and practical applications.

To provide a major showcase for the Project, and a focal point for stakeholder engagement, we will organise two conferences targeted at connecting biodiversity genomics with biodiversity and environmental stakeholders. The first delivered after one year, will be an online conference organised to establish relationships and knowledge exchange partnerships, and communicate the conceptual area of the Project. The second will be a major in-person international conference near the end of the Project focusing on dissemination and exploitation of the research outputs and seeking further opportunities for joint funding.

The terms of Intellectual Property Rights (IPR) management will be specified in detail in the Consortium Agreement to be signed at the beginning of the Project. Beneficiaries will be required to take measures to implement the principles set out in the Code of Practice annexed to the Commission Recommendation on the management of intellectual property in knowledge transfer activities.

2.3 Summary

The Project dissemination, exploitation, and communication measures associated with the expected results and aligned with the specific needs are summarised in Table 5.

The impacts of Project outcomes and their target groups are summarised in Table 6.

3. Implementation

3.1 Work plan

The overall structure of the Project is introduced in Section 1.2. Graphical presentations of the WPs and their components are shown in Figures 1-3. In this section, the detailed work plan descriptions are presented for each WP.

Work package WP1 – Management & Coordination

Objectives: WP1 encompasses overall scientific coordination, financial, and administrative management of the Project, including governance, monitoring, quality assurance, risk management, and reporting. The WP will also provide the required framework for in-Project subcontracting tenders.

Table 5.

Summary of dissemination, exploitation, and communication measures.

SPECIFIC NEEDS	EXPECTED RESULTS	D & E & C MEASURES
Functioning biodiversity genomics networks at the European level	Establishment of the European Nodes of iBOL and EBP	*Dissemination*: Sharing of iBOL and EBP European activities and findings via node portals; *Exploitation*: Major stakeholder conference connecting BGE findings to policy makers and practitioners; *Communication*: Social media and press releases to increase public awareness
DNA barcode reference libraries to enable reliable DNA-based biomonitoring	Comprehensive barcode reference libraries for European pollinators and key indicator species for freshwater and marine biomonitoring	*Dissemination* Data release publication of curated reference libraries; non-technical summary for environmental policy and practitioner stakeholders; *Exploitation*: Operational protocols for open access use of reference libraries for stakeholders; *Communication.* Amateur natural historian and broader citizen science engagement to build and curate reference libraries. Social media, and press releases communicating discovery of species new to science
High-quality reference genomes as a foundation resource for biodiversity science	Reference genomes from 350-500 species assembled and annotated and publicly available	*Dissemination*: Data release scientific publication of curated reference genomes; *Exploitation*: Operational protocols for large scale reference genome production from biodiversity; *Communication*: Social media and press releases
Application of genomic tools to enhance understanding of pan-European biodiversity and biodiversity declines	Protocols established and implemented harmonising DNA biomonitoring and assessments of genome wide diversity in taxa and systems	*Dissemination*: Scientific publications freely available via BGE Knowledge Platform, non-technical summary briefs produced for environmental and conservation organisations; *Exploitation*: Harmonised methods and protocols guiding and supporting use of genomic data in biodiversity conservation; *Communication*: Social media, press releases, presentations, popular articles engaging public

Table 6.

Summary of expected impacts from Project outcomes.

TARGET GROUPS	OUTCOMES	IMPACTS
European DNA barcoding & genome sequencing communities, environmental policy makers and practitioners, citizen scientists	iBOL and EBP communities in Europe operational, connecting and growing community capacity to use genomic tools to tackle the biodiversity crisis	Upscaling genomic data integration into policy and practice to address biodiversity declines
Environment agencies, biodiversity research community, private companies engaged in biomonitoring	Highly curated European reference libraries for delivering DNA-based biomonitoring on pollinators, and for freshwater and marine habitats	Increased reliability of biomonitoring results and increased knowledge of the distribution of European biodiversity

TARGET GROUPS	OUTCOMES	IMPACTS
European scientific community, private companies	Availability of unprecedented knowledge of genome diversity and structure to underpin application of genome sequence data to biodiversity conservation	Reference genomes use by scientific community transforms the power and scale of population genomic studies to understand species biology, track change in genetic diversity, and guide conservation management
Environmental and conservation organisations, private companies	Results uptake by environmental and conservation organisation to inform biodiversity management	Increasingly routine use of biodiversity genomic data in environmental and conservation management

Description: T1.1 Project Management: Financial and administrative management, internal communication, liaison with the EC and stakeholders, day-to-day management, documentation, resource planning, monitoring and reporting. A dedicated programme manager and an administrator will be assigned. T1.2 Scientific Coordination: Overall scientific coordination of the Project, quality assurance and risk management. This Task is led by the Project Director in collaboration with Stream and pillar coordinators and WP leaders to ensure implementation and realisation of Project objectives. T1.3 Infrastructure Design Study: This Task brings together the key findings across the Project to produce a consolidated report, as a design study and implementation blueprints for the establishment of a distributed European Infrastructure for BGE. The report will position such activity within the current landscape of relevant RIs (incl. ELIXIR, DiSSCo, LifeWatch, GBIF, ENA and others). The report will be used to inform the next steps in scaling up our joint operations.

Work package WP2 – Barcoding Node Development

Objectives: WP2 focuses on establishing a European node for iBOL and delivering a programme of work to connect and grow the barcoding community in Europe.

Description: T2.1 iBOL European Node Establishment: We will establish a European node for iBOL, supported by a Secretariat, to coordinate and strengthen DNA barcoding activities in Europe. We will build a governance structure, develop an operational model designed for sustainable continuity beyond the Project, and coordination of network training and engagement events. T2.2 Barcode Information Portal: This Task will establish a European Barcode Information Portal, as a community nexus, and maintain the platform's contents in consultation with the wider community. The platform will register projects and resources and provide a map of national level barcoding and DNA-based biomonitoring programmes. T2.3 BIOSCAN National Nodes: This Task will develop model approaches to grow national level barcoding capacity in Europe, supporting the organisation of barcoding communities into coherent national nodes. The initial focus will be on running pilot studies for Greece and Poland to establish a transferable framework for community organisation, communication, expertise sharing, and connections to funders and policy makers. T2.4 Citizen Science Engagement: This Task will coordinate the participation of citizen scientists and amateur expert taxonomists in the co-creation of knowledge to enhance the construction of DNA barcode reference libraries and

biomonitoring programmes. This includes engaging amateur taxonomists in sample collection, preservation and identification, and the interpretation of DNA barcoding results. It also involves wider public engagement via participation in BioBlitzes and a citizen science marine biomonitoring programme.

Work package WP3 – Genome Node Development

Objectives: WP3 focuses on establishing a European node for EBP and delivering a programme of work to connect and grow the reference genome sequencing community in Europe.

Description: T3.1 EBP European Node Establishment: This Task will provide administrative, legal and logistical support to implement the transition from an informal, mutual understanding among scientists promoting ERGA joint research activities and knowledge transfer, to a formal agreement for the establishment of a European node of the Earth BioGenome Project. Development of a legal entity will entail network management, support, and further development of the ERGA committees, and building on current ethical, legal, and social issues linked to data protection and sample permitting. The Task will support the creation of transferable frameworks for communication, expertise sharing, and connections to funders and policy makers. T3.2 ERGA Scientific Direction: This Task will coordinate ERGA's science policy and vision for practical implementation across ERGA members, committees, the council of country representatives, scientific advisors, and stakeholders. It will provide scientific direction throughout ERGA's establishment phase, including aligning initiatives across committees (Sampling & Sample Processing, Sequencing & Assembly, Annotation, Data Analysis). It also entails guiding the integration of science+ committee efforts (including ELSI, citizen science, stakeholder engagement, training) into ERGA's scientific activities, as well as supporting regional and national nodes to strengthen the distributed reference genome delivery network. T3.3 ERGA Strategic Planning: This Task will focus on future scoping and strategic planning for scaling up reference genome establishment. It entails engagement with global, regional, and national initiatives to build a coherent action plan for effective alignments of the short- mid- and long-term goals; profiling existing European capacities, initiatives, networks, infrastructures across all steps including sampling (vouchering, biobanking, karyotyping, cell lines, etc.), sequencing (with a pan-European vision for distributed capacity), data management (efficiency, access, interoperability), and applications in driving enhanced assessments of European biodiversity; and synthesising experiences of in-Project partners into facts-based policy recommendations on requirements for ensuring efficient scaling up. T3.4 Training and Engagement: This Task will coordinate and support training and engagement initiatives proposed and executed by ERGA members through the Training and Knowledge Transfer (TKT, focused on informatics skills that allow consortium and community members to process and analyse the genome data) and Citizen Science (CS, e.g. delivering core practical skills in sampling and sample preservation as a route to citizen engagement) committees. It entails soliciting community input to identify priorities (scientific topics) and practicalities (existing expertise/resources, training modes/scales) to build an effective TKT/CS

portfolio; providing administrative and eligible costs support during development, set-up (logistics), and running of TKT/CS activities; building on and feeding into existing training networks (ELIXIR Training Platform), and knowledge transfer initiatives (European Knowledge Centre for Biodiversity); and building a Knowledge Platform for community and citizen engagement with ERGA.

Work package WP4 – Barcoding Sampling

Objectives: WP4 will establish the sample supply chain for DNA barcoding to support reference library construction and biomonitoring. This includes gap analysis, metadata management, and biobanking along with field and museum sampling.

Description: T4.1 Gap Analysis: T4.1 will provide priority lists of missing or underrepresented species in relevant public DNA sequence databases to complement the barcode reference library. A focus will be on terrestrial insect pollinators, freshwater macroinvertebrates, aquatic plants, and fish, and marine fish and invertebrates used in European biomonitoring. T4.1 will identify field sampling localities via GBIF/species distribution modelling and suitable museum collections through DiSSCo to retrieve needed specimens. The process of the Gap Analysis involving identification of species to sample, and the localities to sample them from will be ongoing and iterative through the project to M36, reflecting the large scale logistical challenges of targeting and sampling 45,000 individuals from 15,000 species. T4.2 Metadata Standards and Biobanking: T4.2 will define standards on the collection of specimen and sample metadata leading to data tables for use by Project partners. T4.2 will also coordinate long-term biobanking of collected/processed samples in BIOSCAN Europe, define biobanking SOPs and offer sample storage in a central biobank hub for all partners that do not possess their own facilities. T4.3 Museum Collection Sampling: T4.3 will tissue sample 33,000 expertly identified museum specimens from 11,000 species, providing tissues to sequencing centres to fill gaps in DNA barcode reference libraries and aid in the development of intraspecific genetic diversity monitoring tools for priority taxa. T4.3 will operate in close coordination with the field sampling Tasks T4.4 and T4.5. T4.4 Terrestrial Field Sampling: T4.4 will sample terrestrial habitats, aiming at (A) filling gaps in barcode libraries of prioritised groups [T4.1], and (B) developing metabarcoding biomonitoring workflows. Specimen sampling will focus on pollinators (6,000 specimens) and species of management significance (2,100 specimens) of 2,700 species. Community sampling for biomonitoring will focus on pollinator communities (450 sites), ecological restoration (50 sites), ecological intensification (30 sites), and climate change in mountains (40 sites). T4.5 Aquatic Field Sampling: T4.5 will sample 3,900 individual specimens of 1,300 species in the field to fill gaps in the barcode reference library of freshwater and marine organisms. T4.5 will also coordinate and perform sampling of eDNA from 55 sites for assessment of marine invasive species. A total of 550 marine eDNA sampling events across Europe will be performed, including from citizen science efforts in collaboration with T2.4.

Work package WP5 – Genome Sampling

Objectives: WP5 delivers quality biological material (focal areas 'European critical biodiversity' and 'biodiversity hotspots', with special attention to pollinators) for reference genome generation guaranteeing legal and ethical collection, documentation and vouchering of specimens, including biobanking of material. Metadata are documented in a central database connecting to genome generation progress and species prioritisation.

Description: T5.1 Gap Analysis and Species Prioritising: Species prioritisation for biodiversity hotspots [T5.6] and European critical biodiversity [T5.5] including pollinators [T12.2 & 12.4], developed in exchange with citizen scientists [T11.4, T3.3] and community consultation (two open calls, one already conducted), leading to identification of taxonomic and geographic gaps for reference genomes at the very start of the project [T5.1A], and establishment of prioritisation guidelines for the duration of the project and for other genomics consortia [T5.1B]. T5.2 Metadata and sample management: Coordination, guidance, and oversight of high-quality sample and metadata collection, storage, and transfer (within WP5; to WP9, WP11, WP12) under deployment and extension
of reproducible standards. Facilitating metadata exchange between local data/ institutional databases and central repository (WP9). Progress monitoring to continue prioritisation, species lists and their integration with WP9. Identification of experts for target taxonomic groups. T5.3 Biobanking and vouchering: T5.3 coordinates long-term biobanking and vouchering of samples; optional central processing (esp. cell cultures)/ storage by "a biobanking coordinator" at ZFMK. T5.3A coordinates distributed biobanking efforts (defining standards). T5.3B offers long-term storage of DNA extracts, fixed tissue and viable samples (active cell culture) as future-proof, best quality samples at biobank hub. Optional (guidance on) morphological vouchering at ZFMK [T5.3C]. T5.4 R&D SOPs for genome size estimation, karyotypes, and cell cultures: Protocol establishment for genome size estimation, karyotyping, and for difficult taxa cell culturing (routine cultures WP5.3). Genome size estimates by Feulgen staining inform genome sequencing efforts (particularly for rare species) from fixed specimens (e.g., ethanol/ museum) (WP7). Chromosome morphology guides genome assembly (WP9); Cell culturing and permanent frozen viable material (WP5.3) as source of genetic material for future efforts and for karyotyping. Task output informs WP5.1 for prioritisation. T5.5 Critical biodiversity community sampling: Coordination and support of community sampling for reference genomes of European critical biodiversity including pollinators [T12.2, T12.4] and ecologically/economically important species. Local experts will sample and undertake taxonomic identification [T5.5A]. T5.5B ensures snap freezing and submission of material for HMW-DNA extraction to WP7. T5.5C sends frozen/viable material to biobank (hub). T5.5D submits voucher material to suitable collections. T5.5E deposits collected metadata (exchange T5.2). T5.6 Biodiversity hotspots field sampling: Coordination and implementation of sampling for reference genomes from biodiversity hotspots including sampling pollinators (WP12). Partners/teams sample and undertake taxonomic identification [T5.6A]. T5.6B ensures snap freezing and submission of material for HMW-

DNA extraction to WP7. T5.6C ensures biobanking. T5.6D guarantees vouchering in suitable collections. T5.6E deposits collected metadata (exchange T5.2).

Work package WP6 – Barcoding Sequencing

Objectives: WP6 comprises single specimen barcoding and genome skimming to fill current gaps in barcode reference libraries; and environmental DNA and bulk sample metabarcoding to provide data for Project partners assessing biodiversity patterns across carefully selected habitats in Europe. The WP will also drive development of low cost, high-throughput DNA sequencing, enabling wider implementation of these technologies across partners in the Project.

Description: T6.1 Reference Barcode Sequencing: This Task will combine amplicon sequencing and genome skimming to sequence DNA barcodes from 15,000 priority species missing from reference barcode libraries. A total of 45,000 specimens will be sequenced, using both museum specimens [T4.3] and field collecting [T4.4 & T4.5]. T6.2 Metabarcoding for Biomonitoring: This Task will apply metabarcoding technologies to monitor (a) pollinator communities, (b) ecological restoration, (c) ecological intensification, (d) high mountain biodiversity, and (e) marine biodiversity. A total of 3,500 bulk and eDNA samples collected in Tasks 4.4 and 4.5 will be sequenced via 5,000 libraries with up to ~1M read pairs each.

Work package WP7 – Genome Sequencing

Objectives: WP7 will deliver reference genomes - to EBP standards - at scale for the Project. The Tasks include core research and development in complex sequencing, sequencing reference genomes from European critical biodiversity, biodiversity hotspots, and pollinators and case studies. A European network of biodiversity sequencing centres will be created and supported to ensure open knowledge distribution across the continent.

Description: T7.1 Reference Genome Sequencing: This Task will use high-volume genome sequencing to generate data for reference genome assembly of 350-500 species (~450 Gbase span) as supplied by collections made in Tasks 5.5 and 5.6, including extraction, sequencing on advanced platforms, and data quality control. Delivery will be distributed across sequencing partners. T7.2 Sequencing knowledge network: Development and coordination of a European network of sequencing centres with deep experience of genomics processes, developing new technologies and approaches to improve existing procedures and supporting T7.1 by sequencing 10-20 “difficult” genomes. Openly publishing SOPs and protocols for all to benefit.

Work package WP8 – Barcoding Data & Data Infrastructure

Objectives: WP8 is focusing on designing and delivering [T8.1 - T8.3] an enhanced, and connected to other data resources, European mirror of BOLD, improving long-term

sustainability and functionality. Furthermore, it facilitates data deposition and quality assurance from the Project production pipeline [T8.4, T8.5].

Description: T8.1 BOLD European mirror: This Task will establish a European instance for BOLD data publication, with emphasis on redundancy of the data in compliance with European requirements for data location within the Union while ensuring synchronisation between this and Canadian instances. The Task will also work on improving scalability and functionality of the data publication programming interfaces. The output of the Task will make use of an enhanced, centralised code base that will contribute to the development of additional functionality specific to the Project aims. T8.2 BOLD user needs: The Task will identify use cases and needs for iBOL public data access and will connect with stakeholders and users to develop a medium and long-term technical roadmap for iBOL's data architecture, e-infrastructure, and end user services. The output of the Task will make use of the European instance of iBOL data to inform technical decisions for its future, with emphasis on resource links, usability, and FAIR-by-design. T8.3 BOLD Enhancements: This Task will implement the prioritised roadmap items of Task 8.2. This will be achieved by add-ons to the European mirror and by updates to the central codebase, as relevant, including application programming interfaces (APIs) for interoperability with "Sample Data Brokering and Tracking" data services. The Task will also pilot and implement exchange infrastructure to support resource linking between sequence data generated from barcodes to genomes with voucher specimen infrastructure (DiSSCo) and other RIs. T8.4 Barcode data processing: This Task will quality control the new reference library barcode and skimming sequence data provided by the sequencing centres [T6.2] and upload the approved skim and barcode sequences to public repositories so that they can be incorporated into the reference library curation Task [T10.1] and used in DNA-based identification in this Project and beyond. T8.5 Metabarcode data processing: This Task will address the challenge of removing error and artefact from metabarcoding data that is needed for (i) accurate site-based species inventory and (ii) reliable estimation of haplotype variation within species. This Task will use metabarcode+barcode data generated in other Tasks to develop and refine pipelines to remove errors associated with (i) taxonomic inflation at the community level, and (ii) haplotype inflation within species.

Work package WP9 – Genome Data & Data Infrastructure

Objectives: WP9 will develop computational infrastructure to generate high-quality genome assemblies using different types of sequencing data. This includes support for manual curation to chromosome-level. These assemblies will be annotated automatically using Ensembl pipelines and delivered through Ensembl Rapid Release and the ENA. Third party annotations generated from the community will also be incorporated. Data services will provide an integrated platform for the collection, aggregation, and sharing of metadata. The ERGA Data Portal will ensure FAIR open access to genome data generated by the Project.

Description: T9.1 Assembly Development: This Task will develop robust computational workflows to establish high-quality genome assemblies using different types of sequencing data produced by the sequencing centres. It will leverage assembly workflow development experience gained from the Vertebrate Genomes Project and build an open analysis platform in the Galaxy Project. T9.2 Assembly curation: This Task provides expertise and support for the manual curation for the finalisation and approval for public release. It will include capacity building activities that enable community curation in a decentralised fashion. Software and training will be provided to in-Project partners and the wider genomics community. T9.3 Assembly Delivery: This Task will coordinate the use of the assembly workflows by in-Project partners and the genomics community to process the different data types produced by the sequencing centres and deliver high quality genome assemblies. It will also provide support and guidance to maintain common quality control standards and timely delivery of the outputs. T9.4 Community Annotation: This Task will develop workflows for community genome annotation and assessment of annotation. To accommodate taxon-specific issues, the Task will leverage the experience of partners in genome annotation to compile reusable annotation workflows as well as specific tools, maximising the potential of Galaxy environments provided by Compute for Genomes. T9.5 Annotation Delivery: This Task will provide the services to annotate genome sequences arising from the Assembly Delivery. Transcriptomic data alongside cross species protein data will be utilised to generate gene annotation and also include third party annotations into Ensembl. Annotation will be delivered through Ensembl Rapid Release and the European Nucleotide Archive. T9.6 Compute for Genomes: This Task will provide compute resources to be used by Assembly Development, Assembly Delivery, and Assembly Curation to produce reference genomes. The developed workflows will be deployed leveraging the open analysis platform experience of the Galaxy Project to enhance accessibility and scalability of assembly and annotation workflows. T9.7 Brokering and Portal: This Task will develop an integrated platform for the collection, aggregation, and sharing of the fundamental specimen and sample metadata within ERGA, seamlessly connected to the underlying genomic data resources. We will develop an ERGA Data Portal to ensure FAIR open access to the genome data generated by the Project.

Work package WP10 – Barcoding Applications

Objectives: WP10 focuses on generation of tightly curated barcode reference libraries for European pollinators, and freshwater and marine species, and then utilising this resource to deliver and harmonise cost-effective BIOSCAN monitoring workflows for terrestrial, marine and freshwater biodiversity.

Description: T10.1 Reference Library Curation: This Task will produce and curate barcode reference libraries for European pollinators, freshwater, and marine species. This includes: (A) quality and plausibility control of new sequences and metadata, and taxonomic validation; (B) revision, curation, and public release of existing sequences and metadata. This includes 45,000 new barcodes from the Project, plus curation of 200,000 existing barcodes. T10.2 Biomonitoring Interpretation and Application: This Task will

analyse metabarcoding biomonitoring use cases consisting of: (A) pollinator communities, (B) ecological restoration sites, (C) agricultural intensification sites, (D) high mountain biodiversity, and (E) marine invasive species, and will undertake a synthesis to provide a toolkit for harmonisation and standardisation of DNA-based biomonitoring approaches, as well as producing Essential Biodiversity Variables (EBVs) and Water Framework Directive (WFD) relevant ecological quality indicators.

Work package WP11 – Genome Applications

Objectives: WP11 will demonstrate the application of reference genomes. It will develop applied case studies showing that reference genomes are key components of biodiversity research, conservation and bioeconomy, which will be connected by networking and other activities that will maximise applicability and the link with stakeholders and public engagement.

Description: T11.1 Conservation Genomics Applications: This Task will develop case studies that demonstrate the application of reference genomes and genomics to the conservation of species affected by anthropogenic changes in climate, habitats, and ecosystems. These will tackle repeated adaptations of endemic species and strong declines of species impacted by harvesting and urbanisation. T11.2 Bioeconomy and Disease Applications: This Task will develop case studies that demonstrate the application of reference genomes and genomics to species of economic importance and disease control, thus contributing to the sustainability of ecosystem services. These will contribute to stock monitoring and management and characterise host susceptibility and vector capacity. T11.3 BioGenomeApp Network: This Task will connect, track, and boost the development of genome applications. It will identify difficulties of implementation of genomics in biodiversity applications toolkits; evaluate the application potential of community-suggested reference genomes; support case studies [T11.1-11.2] and applied activities from ERGA with cutting-edge research standards; and link communication among genome application developers. T11.4 Genomics for Society: This Task will build a bridge between genomic research, conservation practitioners, and the public, linking case studies [T11.1-11.2] to stakeholders and the public, providing support for their engagement while fostering ERGA multi-stakeholder dialogue. It will also coordinate the upgrading of (mobile) apps for citizen scientists and support community sampling in the citizen science and stakeholder perspective.

Work package WP12 – Joint Network Activities

Objectives: WP12 is focused on priority areas in biodiversity research, training, and conservation where current challenges are recognised and there are clear needs for well-aligned developments that will accelerate future work. The Tasks that make up this joint Stream WP are designed to connect expertise from the two networks in joint actions that will demonstrate the benefits of working together. They recognise the potential for economies of scale and the urgent need for standardising activities and harmonising

data management to enable future scaling up of biodiversity genomics. They represent concrete joint actions that bring the two communities together to address key challenges.

Description: T12.1 Joint Network Training: This Task will enhance capacity to deliver an inventory of barcodes for European taxa and to collect and store suitable material for reference genome sequencing through training in core practical skills that ensure citizen scientists can perform sampling and preservation (sampling protocols suitable for preserving HMW-DNA). This training will also be made available to citizen scientists participating in Task 2.4 to enrich their involvement with the Project. Training of early career scientists will go deeper into all stages from sampling to DNA extraction, sequencing, and data analysis targeted to enhance more specialist skills and expertise for those with existing scientific backgrounds. T12.2 Pollinator Population Sampling: This Task will coordinate population-level sampling of pollinators targeting individual specimens from selected populations and species for intraspecific genetic diversity assessment and developing monitoring applications (connecting to T12.3; T12.7). It will deliver 20-30 individual specimens per sampling locality across European biomes. Reference genomes will be generated via the genome establishment Tasks, and barcodes will be cross-checked and curated via the reference barcode library Tasks. T12.3 Multi-Locus Sequence Assays: This Task will develop protocols for multi-locus sequence assessment of species identity and variation focusing on high-resolution multi-locus DNA barcodes in plants and population genomic assays for insects, building on in-Project produced genomic resources. The work encompasses nucleic acid extraction, assay design, library preparation and sequencing to deliver data for activities at the interface of BIOSCAN Europe and ERGA goals. T12.4 Dark Biodiversity Genomics: This Task will develop cost-effective, non-destructive 96-well plate DNA extractions for small arthropod samples from malaise trapping in Greece, Poland, and Ukraine, and from in-Project BioBlitzes. The Task will deliver 50k barcodes for data taxa alongside protocols that preserve morphology and deliver DNA sufficient for genomes T12.5 FAIR Data Infrastructures: This Task will build on genomics (meta)data management systems of the Darwin Tree of Life, iBOL, ELIXIR, and other infrastructures, services, and interoperability resources to provide a comprehensive FAIR data foundation. It will support ERGA and BIOSCAN Europe workflows in an integrated framework using the COPO data brokering platform to support the processing, validation, and ingestion of standardised sample, barcode, and sequencing metadata into the biodiversity genomics data ecosystem. T12.6 Metadata Change Management: This Task will coordinate change management of Project documents for sample collection (e.g. changes to sample manifest specifications) and management of those changes to data brokering and downstream delivery into the public repositories, analytical workflows, and data sharing resources. T12.7 Intraspecific Genetic Diversity: This Task will develop protocols to design monitoring tools for capturing intraspecific genetic compositions, an Essential Biological Variable for biomonitoring. Focused on pollinators (c.f. T12.2; T12.3), it will combine analyses of full reference genomes and population-level whole genome resequencing data and develop and validate multi-locus assays capable of capturing genetic diversity across species. T12.8 Developing Genomic Barcodes: This Task will focus on developing strategies for achieving species level discrimination of plants (c.f. T12.3), making use of reference

genomes to develop and test standardised multi-locus DNA barcoding protocols to discriminate species, for which standard barcodes fail. This will guide future developments of multi-locus barcoding approaches in plants & other groups. More specifically, Tasks 12.7 and 12.8 (building on Tasks 12.2 and 12.3) will follow the general principle of using reference genomes augmented with resequencing to identify informative regions of the genome for further analysis. Once such regions have been identified in an initial sample set, these regions will be sequenced and compared in an expanded panel of samples using target-capture approaches for plants (for the development of multi-locus barcodes for species identification), and whole genome population resequencing for insects (for the development of multi-locus assays for screening intra-specific genetic diversity). The resulting data will be used to identify the most efficient subsets of loci that provide the maximum information content, which are then suitable for guiding widespread deployment of efficient assays for larger sample sets.

Work package WP13 – Communication and Engagement

Objectives: WP13 supports the communication, dissemination, and exploitation of knowledge from the BGE Consortium. This builds on the knowledge flows between individual work packages and the European nodes of iBOL and EBP in WPs 2 & 3, and adds to this by connecting, translating, and amplifying knowledge exchange with external stakeholders. The core elements of WP13 are the establishment and delivery of a Project-wide communication strategy to increase visibility, inclusivity, and coherence of BGEs messaging; the distillation and communication of technical findings via an accessible online Knowledge Platform; and targeted stakeholder engagement and advocacy to maximise policy traction and exploitation of the Project outputs.

Description: T13.1 Communication Strategy & Implementation: This Task will develop and implement a Communications Strategy to publicise and communicate the work of the BGE Consortium. Communication activities will use the BGE Knowledge Platform [T13.2] as a content resource and utilise a broad suite of communications tools to ensure effective provision of information including a main Consortium website, newsletters, press releases, social media, printed materials, videos, and podcasts. T13.2 Dissemination: This Task will share the technical findings of the Project, and translate them into non-technical summaries, and establish, populate, and maintain an online Knowledge Platform as part of the BGE website. The content for this Knowledge Platform will be co-created by specialists in science communication, who will work closely with WP2 and WP3 to translate the technical findings from the Project into accessible summaries suitable for policy and practitioner audiences. T13.3 Engagement and Exploitation: This Task will develop and implement an Engagement and Exploitation Strategy to connect the Consortium's scientists with key stakeholders and guide exploitation of the Project's findings. This will include mapping the stakeholder landscape, identifying the most effective connection points for policy engagement, and supporting the exploitation of methodologies, processes, and pipelines that enable standardised characterisation and biomonitoring of biodiversity. The Task will also lead the development of two conferences

(one online, one in-person) targeted at connecting biodiversity genomics with biodiversity and environmental stakeholders.

Work package WP14 – Policy & Ethics

Objectives: WP14 aims to identify the necessary policies, guidelines, best practices, and codes of conduct relevant to genomic research; co-create and implement a policy framework; set-up a "Social Innovation Team" to provide advice and guidance to Project partners to ensure alignment with synergetic initiatives and the implementation of best practices; and ensure compliance with the EU Ethics regulations through monitoring activities.

Description: T14.1 Policy contribution: An influential multidisciplinary "Social Innovation Team" will be set-up to identify methodologies and tools that will ensure ethical behaviour, social commitment, decision-making support, science-driven approach and international coherence and alignment of policies and regulatory frameworks (i.e., EU Taxonomy regulation). It will also help monitor the Project outputs regarding compliance with the six principles (EU Green Deal) of ‘do not harm’ (EU Taxonomy regulation) and how biodiversity policies could be mainstreamed. T14.2 Policy compliance: We will ensure coherence with existing policies, from sampling, data extraction, and registration and archiving. Consistency will lead to a set of policy elements ranked based on their enforcement, either informative (e.g. in a clearance registry), recommended (allocated in e.g. repository for Best Practices), or endorsed (community Codes of Conduct). T14.3 Ethics: The Social Innovation Team will assist in the monitoring, with the support of the external advisory board, of the Project risks associated with the production and use of digital sequence information and biological material. Also, it will co-assess the monitoring of gender equality plans, diversity and inclusion, for internal and external stakeholders to the Project (i.e. citizen scientists, taxonomists).

Project Milestones and Deliverables

The Project milestones and deliverables are listed with their due dates (in months with respect to Project duration) in Tables 7, 8, respectively.

Table 7.

Project milestones.

Milestone	Due date
M1.1 Project kick-off meeting	3
M1.2 Project mid-Term report	20
M1.3 All hands meeting 1	15
M1.4 All hands meeting 2	26
M2.1a Governance framework for BIOSCAN Europe	12

Milestone	Due date
M2.1b Reference library subcontracting report	36
M2.2 Barcode Information Portal architecture design	12
M2.3a National node workshops	24
M2.4a Citizen science events call 1	9
M2.4b Citizen science events call 2	24
M3.1 ERGA network management and formalisation	21
M3.2 ERGA network scientific direction implementations	21
M3.3 Strategic planning for future scaling up	21
M3.4 Development of training and engagement	21
M4.1a Initial gap analysis	3
M4.1b 15,000 specimens from gap analysis traced (museum, field)	20
M4.1c 30,000 specimens from gap analysis traced (museum, field)	30
M4.2a First call for samples to central Project biobank	15
M4.2b Second call for samples to central Project biobank	30
M4.3a First assignment of taxa to Task 4.3 partners	6
M4.3b Tissue sampling completed for 21,000 museum specimens	24
M4.4a 8100 terrestrial specimens identified and sampled for sequencing	29
M4.4b 80% of terrestrial metabarcoding samples collected	26
M4.5a 3000 aquatic specimens identified and sampled for sequencing	26
M4.5b 80% of marine eDNA samples collected	26
M5.1a Identification of gaps in available reference genomes for European biodiversity and development of species prioritisation guidelines for genome sequencing	11
M5.1b Call for a second round of suggested species and regions by the community closed	14
M5.2a SOPs for specimen sampling, and their metadata collection and transfer integrated into up to date species lists	9
M5.2b 90% of samples properly sampled and vouchered with correct metadata	32
M5.3a Definition of biobanking standards and guidelines for the Project, linked to Global Genome Biodiversity Network (GGBN)	9
M5.3b Last call for samples needing deposition in the central Project biobank at ZFMK/LIB	36
M5.4a Development of SOPs for karyotyping	16
M5.4b Genome size estimation for 100 to 200 species	16
M5.4c Development of cell culture protocols for non-model taxa	16
M5.5a Submission of material for DNA extraction to sequencing centre and of frozen voucher material to biobank for 50% of species	18
M5.5b Submission of material for DNA extraction to sequencing centre and of frozen voucher material to biobank for 90% of species	27
M5.6a For 50% of the targeted 150-200 species, submission of material for DNA extraction to sequencing centre and of frozen voucher material to biobank done	15

Milestone	Due date
M5.6b For 90% targeted 150-200 species, submission of material for DNA extraction to sequencing centre and of frozen voucher material to biobank	27
M6.1a Tranche 1 - 11,000 reference barcodes delivered	24
M6.1b Tranche 2 - 27,000 reference barcodes delivered	30
M7.1a Data sufficient for 100 Gbase of genome span generated	23
M7.1b Data sufficient for 450 Gbase of genome span generated	36
M7.2a Establishment and population of biodiversity genomics sequencing knowledgebase portal	42
M7.2b Delivery of comprehensive training across partners	36
M8.1 BOLD European portal (beta)	10
M8.2 BOLD user needs workshop	6
M8.3 European BOLD status report	24
M8.4 Deposit 22k reference barcodes in BOLD	30
M8.5 Benchmarking for denoising	20
M9.1 First assembly workflow deployed	12
M9.2a Provision of curation training material	12
M9.2b Curation training completed	12
M9.3a Curated assemblies for #90 Gbase genome span	26
M9.3b Curated assemblies for #230 Gbase genome span	30
M9.3c Curated assemblies for #450 Gbase genome span	38
M9.4 Community annotation workflow deployed	15
M9.5a Public release reference annotation	26
M9.5b ERGA third party annotation release1	34
M9.7a Release ERGA Data Portal V1	10
M9.7b Release COPO V1 metadata user interfaces	14
M10.1a Library curation automation implementation	11
M10.1b Library curation phase 1	24
M10.2a Draft BIOSCAN biomonitoring workflow	16
M10.2b BIOSCAN biomonitoring workflow	28
M11.1 Progress report on conservation case study development	24
M11.2 Progress report: bioeconomy & disease case studies	24
M11.3 Gap analysis on implementation of genomics in biodiversity applications	21
M11.4 Gap analysis on stakeholder and public engagement in genomic applications	21
M12.1 Joint network training activities	26
M12.2 Pollinator population sampling	13
M12.3 Multi-locus sequence assays and resequencing	24
M12.4 Dark taxa genomics	28

Milestone	Due date
M12.5 Data infrastructures for biodiversity genomics	26
M12.6 (Meta)data change management	26
M12.7 Pollinator intraspecific genome diversity	26
M12.8 Species resolution with multi-locus sequencing	32
M13.1 BGE communication strategy	10
M13.2 BGE dissemination plan	10
M13.3 BGE Engagement and Exploitation Strategy	15
M14.1a Workshop on BGE policy framework components	15
M14.1b Interface development and testing	29
M14.3 Panel structure and governance	27
M14.4 Mid-term review	25

Table 8.
Project deliverables.

Number	Deliverable Title	Type	Due date
D1.1	Project Handbook	R — Document, report	2
D1.2	Data Management Plan	R — Document, report	8
D1.3	Infrastructure design study	R — Document, report	42
D1.4	Data Management Plan Update	R — Document, report	36
D2.1	European Barcoding Node Business Plan	R — Document, report	42
D2.2	Online Barcode Information Portal	DEC —Websites, patent, filings, videos, etc	24
D2.3	National node operational models report	R — Document, report	40
D2.4	Citizen Science Engagement Website	DEC —Websites, patent, filings, videos, etc	42
D3.1	Formalisation of ERGA	R — Document, report	42
D3.2	Scientific direction for ERGA establishment	R — Document, report	42
D3.3	Coordinated planning for efficient scaling up	R — Document, report	39
D3.4	Training and engagement activities	R — Document, report	42
D4.1	Gap analysis report	R — Document, report	36
D4.2	Metadata collection standards	R — Document, report	4
D4.3	Biobanking standards	R — Document, report	15
D4.4	Museum specimen sampling	R — Document, report	34
D4.5	Terrestrial field sampling	R — Document, report	30
D4.6	Aquatic field sampling	R — Document, report	28

Number	Deliverable Title	Type	Due date
D5.1	Species prioritisation for genome sequencing	R — Document, report	12
D5.2	Metadata collection and transfer, samples coordination	R — Document, report	36
D5.3	Biobanking and sample storage	R — Document, report	36
D5.4	SOPs karyotyping, genome size, cell culturing	R — Document, report	16
D5.5	Material for genome sequencing critical European biodiversity	R — Document, report	30
D5.6	Material for genome sequencing European biodiversity hotspots	R — Document, report	30
D5.7	Prioritisation guidelines for the wider community	R — Document, report	12
D6.1	Reference barcode sequence report	R — Document, report	36
D6.2	Metabarcoding QC report	R — Document, report	41
D7.1	Data for 450 Gb of reference genomes	R — Document, report	36
D7.2	Open repository of genomics protocols	R — Document, report	42
D8.1	Instance of BOLD Europe	R — Document, report	12
D8.2	BOLD Europe design study	R — Document, report	12
D8.3	Deliver advanced BOLD Europe functionality	R — Document, report	42
D8.4	Deposit reference barcodes	R — Document, report	38
D8.5	Metabarcode denoising pipelines	R — Document, report	34
D9.1	Genome assembly/quality control tools	R — Document, report	30
D9.2	Reference genome assembly submission	R — Document, report	42
D9.3	Reference genome annotation	R — Document, report	42
D9.4	ERGA Public Data Portal	R — Document, report	39
D10.1	European barcode reference libraries	R — Document, report	42
D10.2	Biomonitoring harmonisation and standardisation	R — Document, report	42
D11.1	Conservation genomics case studies	R — Document, report	40
D11.2	Bioeconomy and disease control case studies	R — Document, report	40
D11.3	Developed genomics application toolkits	R — Document, report	42
D11.4	Engagement practices and results	R — Document, report	42
D12.1	Joint training activities	R — Document, report	36
D12.2	Population-level sample collection	R — Document, report	20
D12.3	Multi-locus sequence dataset	R — Document, report	38
D12.4	Future-proofing dark taxa samples	R — Document, report	38
D12.5	Developed FAIR data infrastructures	R — Document, report	39
D12.6	Metadata change management implementations	R — Document, report	39
D12.7	Developed intraspecific monitoring tools	R — Document, report	40
D12.8	Multi-locus barcoding protocol	R — Document, report	42

Number	Deliverable Title	Type	Due date
D13.1	Communication Impact Assessment Report	R — Document, report	42
D13.2	BGE Knowledge Platform	R — Document, report	40
D13.3	BGE Conference	R — Document, report	40
D14.1	Genomics Table of Policies (GeTAP)	R — Document, report	39
D14.2	European biodiversity genomics Policy Panel	R — Document, report	39
D14.3	Monitoring process report	R — Document, report	40

3.2 Capacity of participants and Consortium as a whole

The Biodiversity Genomics Europe (BGE) Project partnership was built by joining forces of two distinct, yet highly interlinked communities; the barcoding (BIOSCAN Europe) and the reference genome (ERGA) communities in Europe. As such, the Consortium achieves a very wide coverage of stakeholders across the European landscape of active organisations in biodiversity genomics and its applications to species discovery, species dynamics, biodiversity conservation, and biomonitoring. The BGE consortium is composed of representatives of 30 of the major, most impactful and established European organisations, across 20 countries, in the fields of biodiversity science, genomics, informatics, conservation and public engagement with science, along with three associated overseas partners representing the leaders in genetics and genomics applied to conservation and monitoring of biodiversity worldwide. The partners of the Consortium come together from different backgrounds and perspectives, inspired by collegiality in decision-making and benefits sharing. They are all strong supporters of open science, both in terms of reproducibility and open data policies. Our partnership is organically connected to the wider networks of BIOSCAN Europe and ERGA, which together represent more than 200 organisations across most European countries (Fig. 6).

The diversity of Tasks and research partners involved in DNA barcoding and production of reference genomes provide a unique opportunity for building interactional expertise, development of shared protocols, common laboratory techniques and bioinformatic pipelines. The Project's partners fully complement one another from sample collection through sequencing and bioinformatic work to the biodiversity application case studies, in line with proper ethical, legal, and social issues linked to collection permits, sample and data handling. Management and coordination of the Project are secured by the main experienced partners, which will also develop the European nodes of iBOL and the EBP in close partnership with iBOL and EBP. This endeavour is complemented by expertise from 18 partner organisations working on sample identification and provision for DNA barcode and sequencing, which in turn sees five main sequencing centres involved as leaders in the field of genomics. Data handling and management also sees high complementarity in the involvement of specialised bioinformatics expertise of individual partners and the main ELIXIR network. Similarly, the application of DNA barcodes and genomes to case studies involves dedicated leadership and south and eastern European countries housing a variety of biodiversity hotspots. Finally, additional partners were

selected for their extensive experience in citizen science and dissemination of science to stakeholders and the public.

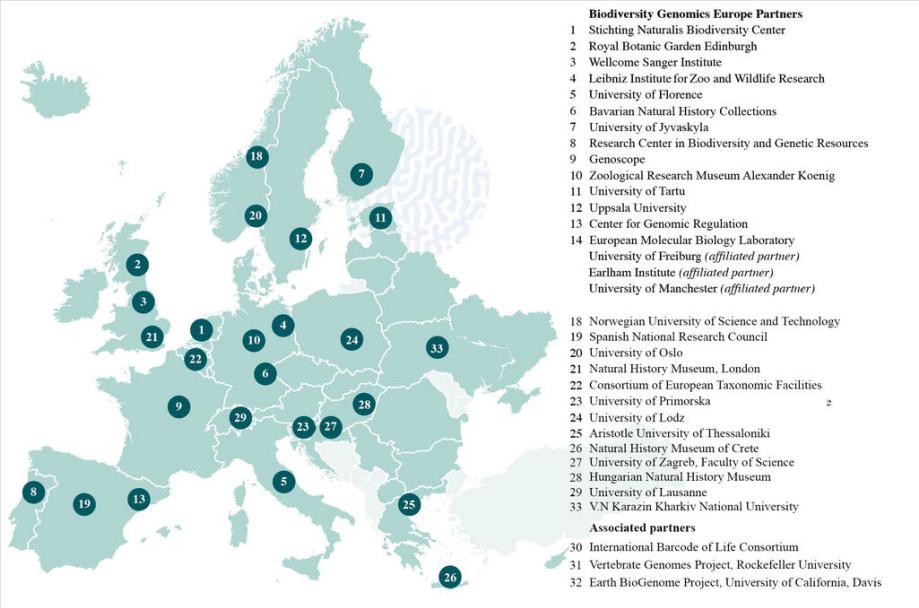

Figure 6. doi

Map displaying the countries (in darker shade) with existing members to the BIOSCAN Europe and ERGA networks as well as the European location of the BGE Project Consortium members (numbers). Affiliated and associated partners (30-32) are not shown in the map.

The Consortium sees benefits from the Naturalis Biodiversity Center as Project leader, with extensive experience in overall scientific coordination and management of European interdisciplinary projects. Naturalis will also fully contribute to the development of the European node of the iBOL, communication, dissemination, best practices as well as DNA barcoding and all joint network activities. This leadership is accompanied by 11 organisations with both disciplinary and inter-disciplinary knowledge in the scientific coordination and infrastructure of the project. The Royal Botanic Garden of Edinburgh (RBGE), The Natural History Museum of London (NHM) and the Zoological Research Museum Alexander Koenig (ZMFK), have extensive experience in curation of biological collections and biodiversity conservation programmes; the Wellcome Sanger Institute, a world leader in genome research; the Leibniz Institute for Zoo and Wildlife Research (IZW) performing basic and applied research on wildlife; the Portuguese Research Centre in Biodiversity and Genetic Resources (CBIO) and the Spanish National Research Council (CSIC), which develop world-class multidisciplinary research in biodiversity and evolutionary biology; three higher teaching and research institutions in life and environmental sciences: the University of Lausanne (UNIL), University of Florence (UNIFI - leading entity of the ERGA-associated Italian Joint Research Unit) and the Norwegian University of Science and Technology (NTNU); the ELIXIR/EBI

intergovernmental organisation which coordinates, integrates and sustains bioinformatics resources; and finally the Consortium of European Taxonomic Facilities (CETAF) with extensive expertise in policies, guidelines, best practices and ethics regulations.

Four additional institutions of higher education and research will join ten members of the coordination team (RBGE, NHM, ZFMK, Sanger, IZW, CBIO, UNIL, UNIFI, NTNU, CETAF) to help create the national and continental iBOL and EBP nodes. These are the University of Primorska, Slovenia (UP), University of Lodz, Poland (UNILODZ), Aristotle University of Thessaloniki (AUTh) and Karazin Kharkiv National University, Ukraine (KKNU).

Central to the BGE Project are the development of networks, technologies and expertise to characterise European biodiversity through genomic techniques. For this, RBGE will also contribute to DNA barcoding and related biodiversity applications, while ZMFK will mainly contribute to the coordination of sampling efforts for both DNA barcoding (together with NHM) and genome sequencing. Sanger will make an essential contribution to sampling and DNA extraction protocols and together with UNIFI, to genome sequencing and data analysis. UNIFI will also be involved in sampling and biodiversity applications for DNA barcoding. IZW will contribute to genome data analysis and with CIBIO and CSIC to sampling and the development of barcode, genome, and joint biodiversity applications. UNIL will be involved in joint network activities and NTNU will offer expertise in taxonomy and DNA (meta)barcoding. ELIXIR will make a major contribution to DNA barcode and genome data analysis and management. CETAF will contribute, along with UP, to the identification and development of necessary policies, guidelines, best practices and ethics regulations of the BGE Consortium.

UP has extensive expertise in citizen sciences and will also be involved in stakeholders and public engagement. AUTH, KKNU and UNILODZ (Polish national contact point of iBOL) will contribute with specimens and expertise on taxonomy and DNA barcoding; AUTH and UNILODZ will be involved particularly in model approaches to grow national level barcoding capacity in Europe. Moreover, the Consortium will benefit from the expertise in sampling from the field and biological collections, as well as training and education, from the Bavarian Natural History Collections (SNSB), the Hungarian Natural History Museum (HNHM) and the Natural History Museum of Crete (NHMC). The University of Zagreb (UNIZAGREB) will contribute with case studies that demonstrate the application of reference genomes to biodiversity conservation in south central eastern Europe. The University of Tartu Natural History Museum (UT) with expertise in metadata management, will focus on identifying use cases and needs for the iBOL data access roadmap and will be piloting the prioritised data services.

Three main DNA sequencing and data analysis centres: Genoscope (CEA), the Spanish National Center for Genomic Analysis (CNAG) and the Swedish SciLifeLab will contribute to the Project by generating large-scale whole genome sequencing data, supporting assembly manual curation, and coordinating the delivery of high-quality genome assemblies. The SciLifeLab will also contribute to the biodiversity applications of reference genome data. Additional bioinformatics and data management will be delivered by the Earlham Institute, an affiliated partner with excellent expertise in

computational science, and biotechnology applications, and three additional academic institutions, the Jyväskylän University (JYU) in Finland (with expertise in mathematical ecology / ecological modelling) and the University of Freiburg (ALU) and University of Manchester (UNIMAN), the latter two as affiliated partners. Finally, the University of Oslo (UiO), home to the Earth BioGenome Project Norway, will provide expertise in biodiversity sampling, DNA extraction and sequencing of challenging organisms.

This partnership will engage in joint research, training and community work by integrating biodiversity conservation expertise and social science and humanities skills to make the dissemination of research products accessible to all levels of society. Expertise in citizen science as well as ethical, social and legal issues linked to open science is widespread across Naturalis, RBGE, CETAF, IZW, UP, ZMFK, NHM and Sanger. Similarly, diverse expertise is in place (Naturalis, RBGE, UNIFI, CIBIO, UP, CSIC, UniLodz, UniZagreb, CETAF, AUTH) to publicise the work conducted by BGE and implement an Engagement and Exploitation Strategy to connect the Consortium's partners to key stakeholders. The partnership has a relatively balanced gender ratio (male:female) of 57% : 43% among principal investigators of the participating institutions.

All partners involved in sampling using either museum or field-based collections will have access to critical infrastructures for vouchering and taxonomic identification provided in-house or through the natural history museum partners. DNA extraction and sequencing will be completed by five sequencing centres involved in the Project, while the ELIXIR network and other partners involved in data analysis and management will provide necessary computing facilities and resources for DNA barcode and reference genome delivery.

Finally, one Canadian and two US organisations will provide guidance as associated partners for the development and maintenance of the iBOL and EBP European Nodes. The first is the International Barcode of Life itself, a research alliance aimed at building the DNA barcode reference libraries, the analytical protocols and the international collaboration required to inventory and assess biodiversity. iBOL is represented in the Project by its international lead organisation, the University of Guelph. The second is the Earth BioGenome Project, the confederated network whose goal is to characterise the genomes of all of Earth's eukaryotic biodiversity. EBP is represented by its co-founder and international lead organisation at University of California, Davis. The Vertebrate Genome Project (VGP) is the third associated partner with a goal of generating near error-free reference genome assemblies of all extant vertebrate species on Earth. VGP is represented by its international lead organisation at Rockefeller University Bringing on board the very same organisations that created the homonymous networks is key to guarantee the success of the BGE consortium which will firmly establish the EU as the world-leading union in genomics applied to biodiversity monitoring, conservation, and restoration.

Acknowledgements

The authors, comprising the BGE Consortium's Executive Board and contributors to the original drafting and finalisation of the writing of the proposal presented in this publication, acknowledge the valuable input, feedback, and constructive discussions from all other contributors. The Biodiversity Genomics Europe (BGE) Project was funded by Horizon Europe (Grant no. 101059492) under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme.

Funding program

Biodiversity, Circular Economy and Environment call (REA.B.3)

Grant title

Biodiversity Genomics Europe

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46 (W1). https://doi.org/10.1093/nar/gky379
- Alsos IG, Lavergne S, Merkel MKF, Boleda M, Lammers Y, Alberti A, Pouchon C, Denoeud F, Pitelkova I, Pușcaș M, Roquet C, Hurdu B, Thuiller W, Zimmermann N, Hollingsworth P, Coissac E (2020) The Treasure Vault Can be Opened: Large-Scale Genome Skimming Works Well Using Herbarium and Silica Gel Dried Material. Plants 9 (4). https://doi.org/10.3390/plants9040432
- Bohmann K, Mirarab S, Bafna V, Gilbert MTP (2020) Beyond DNA barcoding: The unrealized potential of genome skim data in sample identification. Molecular Ecology 29 (14): 2521-2534. https://doi.org/10.1111/mec.15507
- CBD (1992) Convention on Biological Diversity. URL: https://www.cbd.int/
- D'Ercole J, Prosser SJ, Hebert PN (2021) A SMRT approach for targeted amplicon sequencing of museum specimens (Lepidoptera)—patterns of nucleotide misincorporation. PeerJ 9 https://doi.org/10.7717/peerj.10420

- ERGA (2025) The European Reference Genome Atlas (ERGA). URL: https://www.erga-biodiversity.eu/
- European Commission (2020) A Union of Equality: Gender Equality Strategy 2020–2025. European Commission. Issue: COM(2020). URL: https://ec.europa.eu/newsroom/just/redirection/document/68222
- European Commission. Directorate General for Communication. (2020) The European Green Deal. Publications Office, LU. URL: https://data.europa.eu/doi/10.2775/19001
- European Commission. Directorate General for Environment. (2020) EU biodiversity strategy: bringing nature back into our lives. Publications Office, LU. URL: https://data.europa.eu/doi/10.2779/9896
- Fontes J, Vieira P, Ekrem T, Soares P, Costa F (2021) BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. Molecular Ecology Resources 21 (2): 573-583. https://doi.org/10.1111/1755-0998.13262
- iBOL Europe (2025) iBOL Europe – European node of the International Barcode of Life. URL: https://iboleurope.org/
- IPBES, Díaz S, Settele J, Brondízio ES, Ngo HT, Guèze M, Agard J, Arneth A, Balvanera P, Brauman KA, Butchart SHM, Chan KMA, Garibaldi LA, Liu KIJ, Subramanian SM, Midgley GF, Miloslavich P, Molnár Z, Obura D, Pfaff A, Polasky S, Purvis A, Razzaque J, Reyers B, Chowdhury RR, Shin YJ, Visseren-Hamakers IJ, Willis KJ, Zayas CN (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Zenodo. Version Number: summary for policy makers. https://doi.org/10.5281/ZENODO.3553579
- Lewin H, Robinson G, Kress WJ, Baker W, Coddington J, Crandall K, Durbin R, Edwards S, Forest F, Gilbert MTP, Goldstein M, Grigoriev I, Hackett K, Haussler D, Jarvis E, Johnson W, Patrinos A, Richards S, Castilla-Rubio JC, Van Sluys M, Soltis P, Xu X, Yang H, Zhang G (2018) Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences 115 (17): 4325-4333. https://doi.org/10.1073/pnas.1720115115
- Mazzoni C, Ciofi C, Waterhouse R (2023) Biodiversity: an atlas of European reference genomes. Nature 619 (7969): 252-252. https://doi.org/10.1038/d41586-023-02229-w
- Mora C, Tittensor D, Adl S, Simpson AB, Worm B (2011) How Many Species Are There on Earth and in the Ocean? PLoS Biology 9 (8). https://doi.org/10.1371/journal.pbio.1001127
- Ratnasingham S, Hebert PN (2007) BOLD : The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x
- Robinson L, Cawthray J, West S, Bonn A, Ansine J (2018) Ten principles of citizen science. In: Haklay M, Hecker S, Bowser A, Makuch Z, Vogel J, Bonn A (Eds) Citizen Science: Innovation in Open Science, Society and Policy. https://doi.org/10.2307/j.ctv550cf2.9
- Shaw F, Etuk A, Minotto A, Gonzalez-Beltran A, Johnson D, Rocca-Serra P, Laporte M, Arnaud E, Devare M, Kersey P, Sansone S, Davey R (2020) COPO: a metadata platform for brokering FAIR data in the life sciences. F1000Research 9 https://doi.org/10.12688/f1000research.23889.1
- Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R (2021) MinION barcodes: biodiversity discovery and identification by everyone, for everyone. Ecology. https://doi.org/10.1101/2021.03.09.434692

- Weigand H, Beermann A, Čiampor F, Costa F, Csabai Z, Duarte S, Geiger M, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand A, Willassen E, Wyler S, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra K, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, Van Der Hoorn B, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher J, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price B, Rinkevich B, Teixeira ML, Várbíró G, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of The Total Environment 678: 499-524. https://doi.org/10.1016/j.scitotenv.2019.04.247