BIOSCAN Strategic Plan 2019-2026

Vision

The International Barcode of Life Consortium (iBOL) is a partnership among nations represented by key institutions active in biodiversity genomics. iBOL's goal is to implement a cost-effective DNA-based identification system for all eukaryotic species and to advance the use of this system to benefit science and society

iBOL is currently working to deliver its second major program of activity, BIOSCAN, which serves as the central component in a three-part vision.

1. BARCODE 500K (2010-2015)

iBOL completed its first program, BARCODE 500K, in 2015, delivering a reference library of DNA barcodes for more than 500,000 species. This library, managed within the Barcode of Life Data Systems (BOLD) platform, plays a significant role in modern taxonomy and in support of biodiversity surveys and monitoring, particularly as the foundation for metabarcoding.

2. BIOSCAN (2019-2016)

The iBOL Consortium officially launched its current research program, BIOSCAN on June 16, 2019, with an estimated cost of $180 million over a seven-year period (2019-2026).

BIOSCAN builds on BARCODE 500K and aims to revolutionize our understanding of biodiversity and our capacity to manage it via three overarching research themes:

1. **Species discovery**

   Our planet is an island of life in the cosmos. DNA barcoding has been aiding species discovery for 15 years, but millions of species await analysis.

   BIOSCAN will use new protocols and sequencing platforms to increase the pace of analysis while decreasing cost. Ten million specimens from freshwater, marine, and terrestrial ecosystems will be analysed.

   Goal: Generate barcode coverage for 2 million species

2. **Species interactions**

No organism is an island; it is a complex ecosystem. Species interactions are central to the functioning of biological systems, but most remain unknown.

BIOSCAN will use taxonomically targeted primer sets on the DNA extract from single specimens to reveal their commensals, mutualists, parasites and parasitoids – the symbiome.

Goal: Reveal species interactions by targeting the symbiome

3. **Species dynamics**

   Biodiversity is in retreat, and current monitoring programs provide sparse insights into the shifting distributions and abundances of species.

   BIOSCAN will employ metabarcoding to lay the foundation for an earth observation system. It will examine biological communities from at least half the world's ecoregions to begin the task of compiling comprehensive biodiversity baselines.

   Goal: Scan biological communities at 2,000 sites

3. Planetary Biodiversity Mission (2026 onwards)

BIOSCAN is a significant program, promising both to deliver rich insights into global biodiversity and ecology and to trigger major advances in the effectiveness, efficiency and cost of barcoding and metabarcoding solutions. However, it is also intended to serve as a bridge to a future program, the Planetary Biodiversity Mission (PBM), scheduled to commence in 2026. THe PBM is intended to provide a global system for DNA-based monitoring of the world's species and ecosystems.

This strategic plan lays out the areas of focus for the work of the iBOL community to deliver the BIOSCAN vision and establish the systems required for the PBM.

Overview

BIOSCAN brings together a wide range of stakeholders with varying primary interests, including taxonomists and collection managers, ecologists, regulatory bodies, governments and intergovernmental bodies, biodiversity informatics platforms, local communities and citizen scientists, as well as researchers whose main focus is genomic studies, sequencing technology or bioinformatics. Each of these groups will play a significant role in delivering BIOSCAN and has perspectives and interests that will help to shape implementation of the program. BIOSCAN will succeed by enabling all of these groups to develop and own a shared vision and by encouraging cooperation to form a rich global community of interest.

The BIOSCAN strategy includes five interlocking objectives. Each of these is critical for delivering the BIOSCAN vision, and advances towards each objective reinforce efforts around the other four.

1. **Community** - Build and support a global community of practice with the skills and linkages needed to survey and monitor all species via barcoding and metabarcoding.
2. **Capabilities** - Develop and operate world-class infrastructure for low-cost sequencing and bioinformatics to support global-scale biodiversity monitoring.
3. **Coverage** - Develop and deploy efficient and reproducible approaches for DNA-based sampling and monitoring of all eukaryotic groups in all ecosystems.
4. **Impact** - Deliver data and knowledge products that maximise the ability of science and society to explore and interpret species richness and patterns of biodiversity.
5. **Sustainability** - Ensure sustainable and persistent operation of core infrastructure and preservation of samples, DNA extracts, and data from barcoding and metabarcoding activities.

Each objective is supported by a set of strategies. In some cases, these are already clearly defined, but further work is required across the iBOL community to refine planning for many others.

The objectives and strategies presented in this document serve several purposes:

- They communicate the nature and structure of the BIOSCAN program and its planned deliverables
- They establish the major areas of investment required to deliver BIOSCAN
- They provide a structure for the iBOL community to plan decentralised projects and investments in a way that contributes to the overall vision
- They will serve as a structure for discussion and communications within the iBOL community around BIOSCAN plans and progress
- They will provide the structure for annual work planning and progress reporting activities at the global and national scales

Objectives

Community

Build and support a global community of practice with the skills and linkages needed to survey and monitor all species via barcoding and metabarcoding.

BIOSCAN needs to engage stakeholders in all countries to understand the potential of a cheap DNA-based identification system for all the world's species, and to work together to resolve remaining challenges and to apply barcoding tools to support innovative research and improved management of natural and managed systems.

Achieving this objective will involve efforts to engage researchers, government representatives, NGOs, consultants, citizen scientists, and local communities from all countries as partners in BIOSCAN's activities. It will involve development of close alliances with organisations working in other areas of genomic research, taxonomy, ecology, environmental monitoring, and conservation. It will involve preparation and dissemination of clear information resources and training materials in forms and language appropriate to each of these groups.

During BARCODE 500K, DNA barcoding was also supported by the Consortium for the Barcode of Life (CBOL) which partnered with iBOL in capacity building. CBOL is no longer funded, so iBOL now has responsibility for supporting all aspects of community development.

Strategies

1. **Expand national membership and participation in the iBOL consortium.**

   Support barcoding initiatives in all regions and nurture strong national communities of researchers able to participate in all aspects of BIOSCAN.

2. **Develop online community to engage interested stakeholders in BIOSCAN discussions and planning.**

   Use web-based communications tools to build a strong community of practice and to facilitate collaborative activity throughout the network.

3. **Maintain a culture throughout iBOL activities that welcomes and nurtures involvement by all.**

   Ensure that iBOL functions as an inclusive and supportive community for researchers and stakeholders from all backgrounds.

4. **Deliver tools and infrastructure that support the needs of stakeholder communities.**

   Recognise the diverse perspectives and needs of collaborators in different regions and communities, and work with them to deliver value from BIOSCAN.

5. **Build capacity by producing and disseminating clear documentation and training materials to support rapid adoption of barcoding.**

   Work collaboratively to develop the needed range of support materials relating to all aspects of BIOSCAN to enable rapid adoption and contribution by new partners.

6. **Develop partnerships with other biodiversity genomics networks to create synergies across sampling, sequencing, bioinformatics and preservation.**

   Map common interests and opportunities to combine efforts and maximise benefits from sample collection, DNA extraction, sequencing and bioinformatics investments.

7. **Develop partnerships with networks and communities engaged in prokaryotic genomics and metagenomics.**

   Collaborate to ensure that metagenomics solutions and tools efficiently support exploration of all kingdoms of life.

8. **Develop partnerships with ecological and environmental monitoring networks.**

   Develop and adopt protocols that integrate efficiently with broader efforts to survey and monitor ecosystems.

Capabilities

Develop and operate world-class infrastructure for low-cost sequencing and bioinformatics to support global-scale biodiversity monitoring.

BIOSCAN requires investment in robust and large-scale infrastructure, adequate to support the development of a comprehensive barcode reference library for all eukaryotic species and to accommodate and analyse massive volumes of samples for metabarcoding. This infrastructure may be distributed around the world, but it must provide predictable and well documented best-practice handling and management of specimens and community samples, all the way from field collection through accession in well-managed collections and repositories, DNA extraction and sequencing, to scalable bioinformatics pipelines and trusted data management.

BARCODE 500K and subsequent barcoding activities, both in research and applied uses, has established standards and best practices for many aspects of iBOL's activity. investments by iBOL members, particularly those by the Canadian government and other funding agencies to the Centre for Biodiversity Genomics in Guelph, have built and funded pipelines that offer cost-effective processing for large volumes of specimens and samples. However, even with increasing efficiency from sequencing and computing platforms, much more capacity will be required for BIOSCAN to achieve its vision. Investment is needed to establish further core facilities around the world, to expand available IT capacity, particularly for mBRAVE, and to support the archival storage of specimens and DNA in secure and trusted collections.

Additionally, work is required for the community to resolve remaining issues and define standards for barcoding and metabarcoding those taxonomic groups that remain challenging with current markers, primers, and platforms.

Strategies

1. **Support task forces to develop and finalise best-practice approaches for barcoding and metabarcoding challenging taxonomic groups.**

   Evaluate options and develop standards for barcode markers and pipelines offering good taxonomic resolution for all eukaryotic groups.

2. **Expand investment in global sequencing capacity following iBOL standards and best practices and contributing to the BIOSCAN vision.**

   Develop a network of genomics centres offering high-quality and low-cost services for all aspects of barcoding and metabarcoding processes.

3. **Continue innovation to develop lower-cost solutions for processing and sequencing individual specimens and samples.**

Optimise all stages from sample collection through DNA extraction, sequencing, archival and bioinformatics to minimise costs for rapid high-quality barcoding and metabarcoding.

4. **Expand investment in data storage and computing capacity for maintenance and growth of BOLD and mBRAVE.**

   Ensure that informatics investment scales adequately to handle growing volumes of DNA and barcode data.

5. **Expand investment in national collections and repositories for persistent and secure preservation of specimens, samples and DNA from BIOSCAN activity.**

   Build on well-curated specimens and archival DNA to ensure that BIOSCAN supports sustainable science outcomes.

6. **Work with the CBD, national governments, and other stakeholders to ensure compatibility between BIOSCAN's mission and the implementation of the Nagoya Protocol.**

   Deliver BIOSCAN outcomes via models that address the concerns of international stakeholders regarding ownership and ethics.

Coverage

Develop and deploy efficient and reproducible approaches for DNA-based sampling and monitoring of all eukaryotic groups in all ecosystems.

The Global Malaise Program (GMP) has demonstrated the potential associated with a simple and replicable protocol for field collection of a broad class of organisms. The GMP will continue its activities and will be a core sub-program within BIOSCAN, focused on weekly collection and processing of Malaise samples from terrestrial ecosystems all over the world.

Significant effort has also been applied to use of barcoding and metabarcoding approaches to study biodiversity in other systems, especially in soils and in marine and freshwater environments. BIOSCAN requires development and operationalisation of protocols for sampling all components of biodiversity to offer the broadest perspective possible on local community composition.

In developing and adopting sampling protocols, and even in regard to Malaise trapping, BIOSCAN will benefit from close collaboration with ecologists, statisticians, and modelers to ensure good experimental design and to maximise the information value of each sample. The seven years of BIOSCAN offer the opportunity to explore the benefits arising from the capture of different sets of associated measurements and observations simultaneously with collecting organisms for barcoding or metabarcoding and to define best-practice models for use in the PBM. It is likely that some stakeholders and affiliated activities will be unable or prevented from following a more rigorous protocol, so there must be clear processes to document relevant aspects both of field collection and subsequent processing.

The historical materials held in natural history collections remain a foundational resource for biodiversity research. Many species held in collections are rarely or never detected through contemporary field research. Type specimens are particularly important as they anchor the definition for scientific names. BIOSCAN will continue to work with natural history collections to promote and support efforts to improve handling of ancient DNA and to sequence historical materials.

Strategies

1. **Coordinate efforts internationally to contribute to populating the barcode reference library with missing species and with representatives from unsampled regions.**

   Develop networks of collaborators that together contribute materials, DNA and sequences sampling all taxonomic groups and all ecosystems.

2. **Expand efforts to sequence historical materials from natural history collections, particularly type specimens.**

   Develop and promote efforts that deliver high-quality sequences that anchor barcodes and BINs in the context of taxonomic understanding of species boundaries and of existing species names.

3. **Engage ecologists, statisticians, and modelers to ensure that BIOSCAN sampling protocols as far as possible reflect best practice for field survey and monitoring programs.**

   Develop guidance and protocols that encapsulate good experimental design and maximises information secured from field sampling activities.

4. **Expand the Global Malaise Program to all regions and terrestrial ecosystems.**

   Use malaise trapping as a model approach for standardising methods and pipelines and for engaging international stakeholders.

5. **Develop and promote standard protocols and programs for surveying and monitoring soil communities.**

   Work with soil ecologists and existing soil metagenomics activities to standardise sampling and identification of soil biota.

6. **Develop and promote standard protocols and programs for surveying and monitoring marine communities.**

   Work with marine researchers to standardise sampling and identification of biota from marine habitats.

7. **Develop and promote standard protocols and programs for surveying and monitoring freshwater communities.**

   Work with freshwater ecologists to standardise sampling and identification of biota from freshwater habitats.

8. **Develop and promote standard protocols and programs for surveying and monitoring other terrestrial communities and ecosystems.**

   Work with other research communities to develop and promote standard approaches to sample and identify terrestrial biota using approaches that supplement coverage from malaise trapping and soil samples.

Impact

Deliver data and knowledge products that maximise the ability of science and society to explore and interpret species richness and patterns of biodiversity.

The Barcode of Life Data Systems (BOLD) platform, supported by the rich data associated with most barcode specimens, has contributed significantly to the success and growth of iBOL. The capabilities of BOLD and of the new Multiplex Barcode Research And Visualization Environment (mBRAVE) jointly serve as a powerful workbenches for thousands of researchers and hundreds of projects.

BIOSCAN will continue to enhance these platforms and to deliver data products and tools that increase benefits to researchers, governments, industry, and the public. This will be achieved by ensuring that BOLD, mBRAVE, and future tools serve as a best-practice digital research infrastructure, by seeking resources or partnerships to create channels for new audiences to use the data, and by engaging in dialog with current and potential stakeholder groups to understand their requirements and to optimize solutions.

Strategies

1. **Enhance BOLD and mBRAVE to support the functional needs of target audiences, including support to accelerate description of well-defined species represented by BINs and to facilitate the use of metabarcoding data in environmental management.**

   Work with current and potential user communities to ensure that BIOSCAN data are delivered in optimal forms to meet their needs.

2. **Ensure that the materials and sequences underlying any BIN and the algorithms used to organise BINs are clear and well documented with metadata.**

   Follow best practice for open data and open science so that the basis and validity of all aspects of barcoding and metabarcoding results can readily be understood and assessed.

3. **Ensure that BINs are persistent and that changes over time in the set of associated materials are well documented and discoverable.**

   Develop the BIN system as a stable contribution to digital exploration and description of biodiversity and to computable tracking of OTU concepts over time.

4. **Ensure that data in BOLD and mBRAVE are to the fullest extent possible, well documented and accessible under open licenses and following the FAIR**

principles.

Follow best practice internationally for accessibility and reuse of BIOSCAN data.

5. **Develop tools and engage with taxonomists and other experts to curate BINs and their associated scientific names and classification.**

   Facilitate expert review, curation and annotation of BINs and their associated OTUs to ensure correct use of scientific names.

6. **Work with INSDC for increased synergy and data exchange, improved curation, and good citation for data from the iBOL community.**

   Streamline processes for data submission and reuse to reduce burden on researchers and enable them easily and efficiently to contribute sequences for access through multiple infrastructures.

7. **Work with GBIF for increased discovery, reuse, and citation of data from the iBOL community as part of the global knowledge base for biodiversity and species distributions.**

   Ensure that spatial evidence from BIOSCAN feeds efficiently into global datasets for understanding and managing biodiversity.

8. **Raise the profile and understanding of DNA barcoding and metabarcoding in the context of intergovernmental initiatives and agreements, including CBD, IPBES, FAO, and CITES.**

   Address concerns and highlight benefits from broad international adoption of barcoding solutions to address global needs.

9. **Work with regulatory bodies and their officers to embed barcoding and metabarcoding solutions within operational processes for conservation, biosecurity, health, food security, and product certification.**

   Support the needs of governments for cost-effective approaches reliably to identify species and to assess biodiversity.

10. **Engage with citizen science groups, naturalists' societies, and local communities as partners in DNA-based survey and monitoring efforts.**

    Identify methods and practices that enable the widest possible range of stakeholders to contribute samples, DNA, sequences and other benefits to BIOSCAN.

Sustainability

Ensure sustainable and persistent operation of core infrastructure and preservation of samples, DNA and data from barcoding and metabarcoding activities.

Much of the funding required to meet BIOSCAN's estimated $180 million cost must still be raised. Funding will be required in many countries to support fieldwork to collect specimens and samples, in a smaller number of countries to establish high-throughput genomics facilities for barcoding and metabarcoding, around BOLD and mBRAVE to increase capacity and enrich function, and internationally to support research, including taxonomic activity, to respond to the flood of new data.

To date, the majority of funding for iBOL has been contributed by the Canadian government and other funders through CBG in Guelph. BIOSCAN needs to diversify the funding streams to support international expansion of its activities and to ensure sustainability of the infrastructure and services. Different funding sources are likely to be suitable for different aspects of the program. Clear fundable proposals will be developed for all of these, and the iBOL community will collaborate in seeking the resources required for each aspect.

Over time, as the barcode reference library approaches completion, and as processes and pipelines become more standardised and widely adopted, many of the BIOSCAN deliverables will no longer require ongoing investment. However, certain components, including the barcode reference library and tools for its curation and management (corresponding to BOLD today) and the engines for interpretation of metabarcoding data (corresponding to mBRAVE), will remain important tools for research and applied use. BIOSCAN must therefore develop a sustainability plan and long-term governance model for these critical deliverables and a lifecycle plan for all components, to ensure that capabilities and services are not interrupted.

Strategies

1. **Ensure that BIOSCAN has a governance model that distributes ownership and trust across the consortium.**

   Maximise participation and sustainability through shared ownership of the vision, products and outcomes.

2. **Plan the life cycle for all major components of the BIOSCAN infrastructure globally, ensuring that operationally important capabilities and services are maintained.**

   Avoid failure of trusted infrastructure by forward planning and sustainable planning for key components and services.

3. **Ensure that the barcode reference library and the linkage between barcode records and specimens is robust and supported by digital repositories with long-term funding.**

   Ensure persistence of the library of reference sequences as these serve as the trusted basis for species identifications.

4. **Develop clear value propositions for investments in all aspects of BIOSCAN, including field collection, sequencing, bioinformatics, and preservation of materials and DNA.**

   Work collaboratively across the iBOL consortium to expand and diversify the funding base for BIOSCAN.

5. **Collaborate internationally to develop and submit well-designed proposals that contribute to the implementation of this strategic plan.**

   Submit funding applications to suitable sources internationally and as a means to multiple core facilities to underpin BIOSCAN.

6. **Focus on establishing at least one genomics centre elsewhere in the world from CBG, to increase capacity and to add redundancy.**

   Avoid the risks associated with reliance on a single centre for delivering capacity for BIOSCAN.