White Paper

Establishing a national initiative for Earth BioGenome Norway

(EBP-Nor)

The EBP-Nor Working group

Professor Kjetill S. Jakobsen, University of Oslo

Professor Arild Johnsen, University of Oslo

Professor Carl Henrik Gørbitz, Director UiO:Life Science

Associate Professor Simen Rød Sandve, Norwegian University of Life sciences

Professor Øystein Lie, Norwegian University of Life Sciences

CEO Hanne Mette Dyrlie Kristensen, The Life Science Cluster

Professor Andreas Henjol, University of Bergen

Professor Anders Goksøyr, University of Bergen

Associate Professor Michael D. Martin, Norwegian University of Science and Technology

Professor Trygve Brautaset, Norwegian University of Science and Technology

Professor Kim Præbel, The Arctic University of Norway

Professor Trond Jørgensen, The Arctic University of Norway

Dr. Steinar Bergseth, The Research Council of Norway, has participated as an observer

Secretariat:

Dr. Jason Whittington

Dr. Thomas Marcussen

Oslo, Ås, Bergen, Trondheim, Tromsø 15 June 2020

Executive summary

Biodiversity is increasingly threatened by human-related activities. Understanding and preserving ecosystems and species are global imperatives for human survival and prosperity. Knowing the complete DNA sequences (genomes) of all species of life on Earth will provide the ultimate fundament for addressing biological questions of all kinds (fundamental and applied) and will represent an unprecedented gold-mine of scientific data for biotechnology, medicine and drug development, bioprospecting, new biomaterials, biofuels and bioproduction (including aquaculture, horticulture and livestock production). The Earth BioGenome Project (EBP) is a global non-profit initiative established for this purpose: to catalogue the genomes of the entirety of all of Earth's 1.5 million eukaryotic species over a period of ten years. EBP has member institutions across the world including in USA, Canada, Mexico, Brazil, Chile, Columbia, UK, Germany, Spain, Norway, Denmark, Sweden, Saudi Arabia, China, South Korea and Australia. The Nordic countries are represented by University of Oslo, SciLifeLab (University of Stockholm, Technical University, Stockholm (KTH), Uppsala University) and University of Copenhagen. Here we propose a Norwegian national network node of EBP, EBP-Nor, set up to sequence all Norwegian eukaryotic species (estimated to 45,000 in total) over the same period. EBP-Nor is a bottom-up initiative embraced by the seven major Norwegian universities by signing a Memorandum of Understanding (MOU). A Norwegian EBP will have a huge impact on the development of the future bioeconomy, preserving biodiversity and providing knowledge needed to handle future anthropogenic eco-evolutionary disease crises like SARS, Ebola and Covid-19. The initial aims, constitution, stakeholders, structure, scientific plans, and requirements for establishing the EBP-Nor project are outlined as a basis for wider discussion.

1. Background: Laying a foundation for a new and sustainable bio-economy

Genome DNA sequences are the “blueprints of life” and are products of evolution. All cellular life on Earth is related by descent from a common ancestor ~4 billion years ago, and there are currently around ~1.5 million known species on Earth (excluding prokaryotes and viruses).

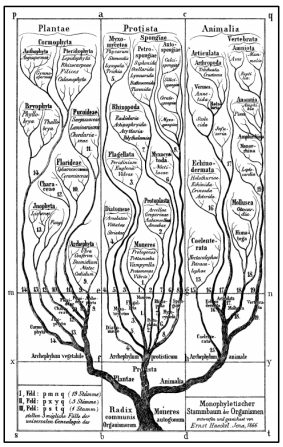
Biodiversity is increasingly threatened by human-related activities, directly by habitat destruction and species exploitation, and indirectly by climate change, spread of invasive species and pests as a result of globalization. We are in the midst of the sixth great extinction event of life on our planet, with **the risk of losing up to 50% of existing species by 2050**, which not only threatens wildlife species, but also **imperils the global food supply**[1].

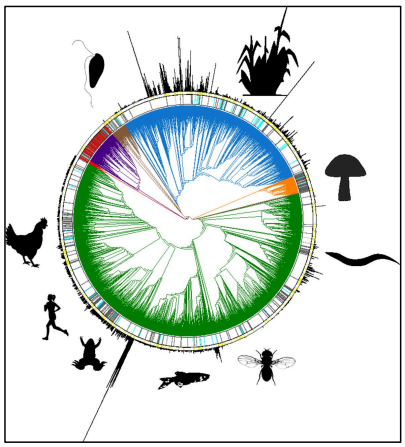
Therefore, conservation of species and ecosystems, and increasing our understanding of their ecological interactions, are global imperatives for human survival and prosperity.

As a "blueprint of life" an organism’s genome DNA sequence encodes ultimately everything the organism needs for maintaining itself and reproducing itself. However, sequencing the genome of an organism (i.e. making a reference genome) is a first step towards accessing and understanding this information. It will require additional projects for example generating population genome information (with a reference genome at hand this will be only 1/100 or even less compared to the reference genome costs), dedicated experiments based on the genome sequence and mining of genes with a potential of use in biotech, agriculture and

[1] Lewin et al. 2018. Earth BioGenome Project Sequencing life for the future of life. *PNAS* 115: 4325–4333.

medicine. Having an atlas of all DNA sequences (i.e. reference genomes) of all species of life on Earth will ultimately **provide fundamental insights into biology**, transform our **ability to tailor conservation of life**, and **deliver a comprehensive view of nature's toolkit for future biotechnology** to provide humanity with **food**, **medical treatment, drugs, vaccines, biofuels** and **biomaterials**. It will produce a complete catalogue of life on the planet and thus be a historic mission to acquire fundamental knowledge about the natural world - in the same spirit as particle physics, and exploration of the universe – and represent a gold-mine of data to be utilized **for the benefit of mankind** directly and in follow-up projects (e.g medicine, blue-green economy).





Box 1. The tree of life. Ernst Haeckel's tree (left) from 1866 showing three kingdoms: Plantae, Protista and Animalia. The modern view (right[2]) is more sophisticated, and we make a fundamental distinction between prokaryotes (lacking a cell nucleus), i.e. Archaea (red) and Bacteria (purple), and eukaryotes (which have a cell nucleus) and include fungi (yellow), plants (blue), protists (brown) and animals (green). The inner circle shows phylogenetic relationships (as a tree) and the middle circle genomic information is depicted: complete genomes (red), chromosome level (blue), scaffolds/contigs (dark/light grey). The outer circle shows existing transcriptome data and the genome sizes are shown on the outside.

Recent advances in DNA sequencing technology, notably the advent of long sequence reads, have made full-genome sequencing for all species of life feasible for the first time, both in terms of processing and cost. These improvements in technology have catalyzed the concerted and coordinated worldwide programme to sequence life on Earth – the **Earth BioGenome Project (EBP)**. EBP will accelerate the ongoing and initiate new initiatives to complete ALL species on Earth FAST in a global collaboration. This global collaboration represents the largest effort so far in the life sciences. It will transform the field and drive innovation, promote a sustainable bioeconomy, and also drive our understanding of ecosystems and human impact and thus aid in protecting biodiversity. Needless to say, such knowledge will substantially affect our ability to deal with and prevent eco-evolutionary crises such as the Covid-19 pandemic.

1.1. Sequencing life for the future of life: The Earth BioGenome Project (EBP)

The Earth BioGenome Project (EBP, https://www.earthbiogenome.org/) is a global non-profit initiative that aims to sequence and catalogue the genomes of all of Earth's 1.5 million currently described eukaryotic species (Box 1) over a period of ten years. The project was officially launched November 1, 2018 – already then embracing countries in North and South America, Europe, Asia and Oceania. A scientific paper presenting the vision for the project was published in *PNAS* in April 2018[2].

The grand vision of EBP is creating a new foundation for biology to drive solutions for

[2] Lewin et al. 2018. Earth BioGenome Project Sequencing life for the future of life. *PNAS* 115: 4325–4333.

1. **benefiting society and human welfare** (medicinal resources including new drugs, bioprospecting, cancer treatment, pandemics control, agriculture, food and feed productions, aquaculture, fisheries, biomaterials, biofuels, environment).
2. **protecting biodiversity** (climate change, human activities, conservation plans, ecosystem restoration), and
3. **understanding ecosystems** (new species, relationships, evolution and evolutionary drivers, ecosystems dynamics).

The EBP initiative complements the initiatives on the other two domains of life, Bacteria and Archaea (Box 1).

EBP is a "community of networks" and defines standard protocols for data generation,

Box 2. Institutional members of EBP (as of May 2020).

Australia Museum, Australia
BGI, Shenzhen, China
BioPlatforms, Australia
British Columbia Cancer Research Center, Canada
Center for Translational Biodiversity Genomics, Denmark
Dalhousie University, Canada
Duke University, USA
Earlham Institute, UK
FAPESP, Brazil
George Washington University, USA
Harvard University, USA
Institut D'Estudis Catalans, Spain
King Abdul Aziz University, Saudi Arabia
Korea Polar Research Institute, Korea
Korea University, Korea
Max Planck Institute of Molecular Cell Biology and Genetics, Germany
Natural History Museum, Denmark
Novim, USA
Rockefeller University, USA
Royal Botanic Gardens, Kew, UK
SciLifeLabs Genomics (Univ of Stockholm, KTH, Uppsala Univ), Sweden
Smithsonian Institution, USA
Spacetime Ventures, Brazil
Universidad de los Andes, Colombia
Universidad Nacional Autonoma de Mexico, Mexico
Universitat Pompeu Fabra, Barcelona
University of California, Davis, USA
University of California, Santa Cruz, USA
University of Cambridge, UK
University of Chile, Chile
University of Copenhagen, Denmark
University of Florida, USA
University of Illinois, Urbana-Champaign, USA
University of Oslo, Norway*
University of Sydney, Australia
USDA-ARS, USA
Wellcome Sanger Institute, UK

***UiO represents seven major Norwegian universities – UiO, NMBU, UiA, UiB, NTNU, Univ Nord and UiT – that are part of EBP through a common Norwegian Memorandum of Understanding (MOU).

analysis, and storage for the affiliated genome sequencing initiatives. As of May 2020, EBP comprises 30+ institutional members (Box 2) and a substantial number of affiliated project networks (Box 3).

The **data produced by EBP will be made publicly available** in a DNA database of biological information and will constitute a platform for future scientific research and support environmental and conservation initiatives.

Box 3. EBP affiliated projects and networks (as of May 2020).*

- Genome 10K (originally 10.000 vertebrate genomes – later largely expanded)
  - Vertebrate genome project (repr. of each family of vertebrates)
    - 10.000 bird genomes (B10K; part of the VGP)
    - Fish 10.000 genomes (Fish 10K; part of the VGP)
    - 1000 Bat genomes (part of VGP)
- 1000 Fungal genomes (1KFG)
- 10.000 plant genomes (10KP)
- 5.000 insect genomes (i5K)
- 1000 Chilean genomes
- 200 Mammals project
- Ag100 Pest (USDA)
- BRIDGE Columbia
- California Conservation Genomics Project (Cal GCP)
- Darwin Tree of Life -DToL (UK – run by Sanger Inst, funded by Wellcome Trust etc.)
- 150 Canadian Genomes
- Diversity initiative of Southern California Ocean (DISCO)
- DNA Zoo
- Global Ant Genomics Alliance
- Global Invertebrate Genome Alliance
- Open Green Genomes
- PhyloAlps
- **EBP-Nor**

* This list is not exhaustive. New initiatives arise frequently. Each project funded individually or by the research institutions being involved (by means of National (Research Councils)/EU/NSF/USDA/NIH funding). DToL funded from private and governmental bodies.

The cost of sequencing all of Earth's 1.5 million eukaryotic species has been estimated to US $4.7 billion, similar to the cost of sequencing the first human genome (The Human Genome project; HGP)[3]. **The return on investment for EBP is likely to be very large and may exceed that of HGP, estimated in 2011 at 141:1**[4]. By 2020, the return on investment of HGP has become substantially larger (approximately 10-fold).

2. A Norwegian node of the EBP

A Norwegian node of the global EBP – **EBP-Nor** – has been established and currently includes **seven major Norwegian universities – UiO, NMBU, UiA, UiB, NTNU, Uni Nord and UiT,** (Box 4). Between these institutions a Memorandum of Understanding (MOU) has been signed by the rectors, and at the scientist level a working group representing the MOU institutions has been established as an initial step. The next steps will be to formalize the organization, start the initial scientific work (sampling and initiate sequencing), along with a process of securing funding. For sampling of biodiversity, we will take advantage of the of the established biobanks such as MarBank, Svalbard Global Seed Vault, the Natural History Museums, Nature management institutions as well as the nation-wide community of researchers involved in EBP-Nor.

EBP-Nor embraces the three main visions of EBP (point 1.1) and has **four main goals**:

1. Sequencing all eukaryotic species in Norway, estimated at 45,000 species[5]. Since most Norwegian species occur also elsewhere and as EBP-Nor will be a **coordinated effort**

[3] Lewin et al. 2018. Earth BioGenome Project Sequencing life for the future of life. *PNAS* 115: 4325–4333.
[4] Tripp & Grueber 2011. Economic Impact of the Human Genome Project. Battelle Memorial Institute. 1–58.
[5] Elven & Sølie 2016. Kunnskapsstatus for artsmangfoldet i Norge 2015. Utredning for Artsdatabanken 1/2016. Artsdatabanken, Norge.

together with the Nordic countries, UK and other relevant EBP countries, the actual number of species to be sequenced by EBP-Nor will be far below 45,000, **likely between 9,000 and 22,500** depending on the number of participant EBP countries. EBP-Nor will be carried out in three phases that include a first stage, **Launch,** where high-quality genomes are generated for a diverse, but limited (20-50), set of species. In the next **Ascent** phase, we will scale up and do up to 500 species, and in the final **Orbit** phase we are at the production scale of 10,000 genomes or more (see below for details). For the Ascent and Orbit phases the contributions from the other participating institutions / countries will have a strong impact.

2. **Establish a Norwegian nation-wide network of collaborating nodes**, from universities to research institutes. This will be an inclusive process with the only demand that each lab adheres to the standards of EBP-Nor for sampling, sequencing approaches, and genome assembly.
3. **Actively collaborate with other countries to achieve global EBP goals**. EBP-Nor will participate in the EBP, play an active role in further developing the global strategy of the EBP and will collaborate with other countries in their BioGenome Projects, primarily neighbouring countries in Europe, the Arctic, and the Antarctic.
4. Through the EBP goals 1, 2, and 3 **EBP-Nor will generate added values through genome-based knowledge for utilization** in biotechnology, bioprospecting, aquaculture, marine resource management, protecting biodiversity, materials sciences, biofuels, drug discovery, and medical treatment, that will benefit Norway as a nation.

EBP-Nor will conceive, coordinate and deliver a concerted series of sub-projects which will explore different designs and approaches, take leadership in developing analytic methods and deliver a substantial contribution towards the global goal of the EBP of sequencing all eukaryotic species.

EBP-Nor is a bottom-up researcher-led initiative embraced by the universities of Norway. In Box 4 the institutions that have signed the MOU (in fall 2019) are listed along with their relevant expertise. Natural candidate institutions with relevant and in many cases complementary expertise is also included. A goal is to have as many as possible of the (and possibly additional research institutes) on board in 2020 (perhaps early 2021).

Box 4. Norwegian institutions embracing the EBP-Nor initiative. Institutions indicated with an asterisk signed a committing Memorandum of Understanding (MOU) in fall 2019.

Institutions	**Relevant expertise for EBP-Nor**
*University of Oslo (UiO)	Genome sequencing and genomics, bioinformatics, ecology and evolution, taxonomy, eukaryotic microbiology, mol. and cell biol.
*Norwegian University of Life Sciences (NMBU)	Agri- aquaculture genomics, genome sequencing, bioinformatics, genome evolution, mol. and cell biol.
*Norwegian University of Science and Technology (NTNU)	Microbiology, biotechnology, systems biology, genomics, marine bioprospecting, taxonomy, DNA barcoding.
*University of Agder (UiA)	Marine biology and ecology, marine genomics.
*University of Bergen (UiB)	Marine and aquaculture genomics, genome sequencing, bioinformatics, evolutionary and developmental biology, marine and terrestrial ecology, conservation biology.
*Nord University (Uni Nord)	Marine biology, fish genomics, population and comparative genomics.
*The Arctic University of Norway (UiT)	Marine bioprospecting, marine genomics, arctic biology, fishery sciences, aquaculture.
Candidate institutions	
Institute of Marine Sciences (IMR)	Marine biology, marine genetics and genomics, fishery management, marine bioprospecting.
Nofima	Aquaculture, functional genomics, food and feed biotechnology.
NIBIO	Forestry, fungal biology, plant disease.
NINA	Conservation biology, wildlife management.
NIVA	Aquatic biology and ecology, algal collection.
SINTEF	Biotechnology, microbiology, bioprospecting.
The University Centre in Svalbard (UNIS)	Arctic biology.

EBP-Nor will be an inclusive consortium constituted of **Norwegian scientists** (evolutionary biologists, genomicists, computational biologists, ecologists, taxonomic experts and others), **natural history museum** organisations, both national (Oslo, Bergen, Trondheim, Tromsø) and international (e.g. Royal Botanic Gardens, Kew; Royal Botanic Garden Edinburgh; Wildlife Trust; Smithsonian Institution). Furthermore, collaborations with several **international genome / bioinformatics centres** are already established (such as Sanger Institute, EMBL-EBI, Rockefeller University) and will be further extended in a near future (see 2.5). At a wider **stakeholder scale,** sample collections (zoos, others), government agencies, interested stakeholder communities (naturalists, university students, schoolchildren), public engagement entities, commercial entities (farming bodies, technology companies) and funding bodies will be engaged.

EBP-Nor will be open for private funding and sponsorships. Several processes have been initiated, and we are currently in an ongoing dialogue with several large private enterprises with interests in the marine environments, biodiversity research, bio economy/blue-green shift and bioprospecting (Box 5). We have experienced a huge interest in EBP-Nor from private companies (biotech, pharma, bioproducts, biomaterials, aquaculture, bioproduction, marine technology, energy etc.) and governmental ministries, directorates, resource

management as well as research and innovation funding bodies (Box 5). We seek to obtain funding from academic funding bodies (such as RCN, RFF, Nordforsk), other governmental funding and private funding from industry, finance and foundations (see Appendix for cost estimates).

EBP-Nor will develop an **academic EBP research community geared to process, analyse and present the data generated** that, as a whole, will be **a foundation for long-term utilisation of this unique genome and metadata resource**. Included in this would be the creation of training courses and degree / post-graduate qualifications. International collaborations will be developed and engaged with this community to facilitate the advancement of the global

Box 5: Potential stakeholders and interested parties*

Industry:

- **Pharma**: (e.g. Roche, Bayer, GlaxoSmithKline, AstraZeneca, Novo Nordisk, Lundbeck)
- **Biotech**: (Norw. companies e.g. ArcticZymes, PCI Biotech; large int. e.g. NovoZymes) Chemical/biomaterials industry: (e.g. Borregaard)
- **Bioproduction**: Forestry, pulp and paper (e.g. Norske Skog, Nordskog, Norges Skogeierforbund, Borregaard); Aquaculture (e.g. Marine Harvest, AquaGen, Pharmac, Skretting, EWOS); Fishery/Marine (e.g. Marine Harvest, Aker Biomarine)
- **Energy**: (e.g. Equinor, ABB, Statkraft)
- **Environment**: (e.g. REV Ocean)
- **Finance, shipping, insurance**: (e,g. DNB, Nordea, Storbrand etc)

Private foundations:

- e.g. FHF – Norwegian Seafood Research Fund, Olav Thon Stiftelsen, Kavli Trust, The Kavli Foundation, Trond Mohn Foundation (Bergen), The Nansen Foundation

Environmental organizations:

- e.g. Zero, WWF, Bellona, Naturvernsforbundet, Sabima, Natur og Ungdom

Governmental (ministries, directorates, resource management etc.):

- **Governmental ministries**: Climate and Environment, Education and Research, Agriculture and Food, Trade, Industry and Fisheries, Ministry of Health and Care Services, Local Government and Modernisation (ICT policy), Petroleum and Energy
- **Directorates**: Norwegian Environment Agency (Miljødirektoratet), Norwegian Directorate of Fisheries

Research and Innovation support, industry/academia contact points:

- The Research Council of Norway**, Innovation Norway, The Research Council of Norway**, Kreftforeningen, Oslo Cancer Cluster, The Life Science Cluster**, UiO:LifeScience**, RRF – Regional Research Fund, Technology Transfer Offices (TTOs), Nordforsk

Relevant projects:

- Digital Life Norway (DLN)**, Elixir Norway incl. BioMedData, The Nansen Legacy, EMBRC-Eric, Oslofjord-Skagerak initiative, Svalbard Global Seed Vault, Norwegian Barcode of Life (NorBOL), GBIF Norway

* This is by no means a comprehensive overview of all potential organizations, industry, stakeholders, collaborators and putative funding bodies. The above-mentioned parties are meant to serve as examples, and we have not yet interacted with all of them – but a dialogue has been established with some (both industry, funding and governmental bodies.

** Provided support for establishing/running the EBP-Nor Working group and writing of this document.

EBP nodes.

EPB-Nor is envisaged to run over a period of approximately 10 years, with the work structured in three phases (cost estimates for the three phases are given in the Appendix):

- **Phase 1: Launch** (proof of concept) **(1-2 years)**: Sequence the genomes of 20–50 species, with a focus on species of particular interest to Norway, such as "iconic" species selected in interaction with the interested public, such as industry, funding bodies, and school classes. This initial phase is intended for establishing expertise, pipelines, consolidate EBP-Nor to identify and distribute tasks according to expertise and infrastructure nationally, develop collaboration between EBP-Nor and EBP, and to serve as a demonstration to funding bodies of the joint capacity to deliver in the next two phases. This stage will include sampling of specimens, biobanking (including collection/registration of metadata), sequencing, analysis and curation of sequence data (annotation etc.), storage, and making the data publicly accessible.
- **Phase 2: Ascent** (scaling up) **(2 years)**: Sequence the genomes of 200–500 species making use of the pipelines established in Phase 1. This stage will include denser sampling of specimens, establishing biobanks and scaling up of sequencing, sequence analyses, storage and coupling sequence data and metadata. This large national effort will be established in close coordination with EBP initiatives in Nordic and other relevant countries, particularly with regard to species selection. A smaller set of species will be subject to population studies.
- **Phase 3: Orbit** (production) **(6-7 years)**: Sequence the genomes of the remaining thousands of species from all types of ecosystems, phylogenetically across the tree of life. Selected complete ecosystems will be studied comprehensively, such examples may be Svalbard, the Oslofjord / Skagerrak region, west coast ecosystems, Lofoten, mountain ecosystems (e.g. Hardangervidda, Jotunheimen). This phase will involve even tighter collaboration at the Nordic, UK, Canada, US and European level, as well as EBP at the global level. Data accessibility will be ensured by curated biodiversity biobanks and open access international sequence databases.

2.1 Special competitive advantages in Norway

In an international context, Norway has particular competitive advantages within the **marine sciences** (**marine ecosystems, fisheries and aquaculture**), **High-Arctic ecosystems** (including Svalbard), **mountain and fjord ecosystems**. We have a strong standing internationally in **marine genomics** (exemplified by the genomes of Atlantic cod and Atlantic salmon) and in comparative and evolutionary genomics. Importantly, Norway has initiated several huge and **important marine research initiatives** with potential strong synergies to EBP-Nor, including The Nansen Legacy, The Oslofjord-Skagerrak Initiative, The Solberg government's Ocean initiative, EMBRC-Eric and the privately funded REV Ocean (see Box 5). We also have **a strong Research Institute sector** within the marine and aquaculture sciences including Institute of Marine Research (IMR), Nofima, SINTEF and The Norwegian Institute for Water Research (NIVA) (see Box 4). In addition, the previously mentioned Digital Life Norway, Elixir Norway, the bioinformatics units Center for Bioinformatics (UiO) and Computational Biology Unit (UiB) and the highly competent biodiversity research environments at Norwegian Institute for Nature Research (NINA), Directorate for Nature Management (DN), Norwegian Institute for Bioeconomy Research (NIBIO) and Natural History Museums at our large universities (UiO, UiB, NTNU, UiT) all represent special competitive advantages.

2.2. Added value of the EBP-Nor initiative

The added value of the EBP-Nor initiative can be structured into three broad categories reflecting the three main visions of EBP.

Benefiting society and human welfare, through ecosystem services (i.e. ways by which ecosystems contribute to the benefit of society and human welfare) and biological assets. There is a huge untapped potential in mining genomes, using e.g. machine learning (i.e. extracting patterns from large datasets), for genetic resources and for synthetic biology (i.e. template for imitating nature's biological functions and processes), and for genomics-informed management of natural resources such as fisheries and other wildlife. EBP-Nor will enable **discovery of new medicinal resources** for human health, including **pandemics control** (vaccine development); identifying **new genetic variation**, or **new species**, for **improving agriculture and aquaculture** (e.g., yields, disease resistance) – discovering is already a prioritised goal of the Solberg government's Ocean initiative[6]; discovering **novel biomaterials**, **new energy sources**, and **biochemical functions and uses**; improving **environmental quality** (soil, air, and water). EBP-Nor will **attract biotech, pharma, genomics, aquaculture industry involvement** - either as collaborators (innovation partners), as funders, or both. The ongoing shift from oil-based to a sustainable bioeconomy will be ideal and timely for companies within the energy or transport sector to engage in sustainable and future-directed projects (see Box 5).

Protecting biodiversity. EBP-Nor will enable developing more efficient, **evidence-based plans for conservation, protection, and regeneration of biodiversity** on the national level. This will be achieved through determining the role of climate change on biodiversity, clarifying how human activities (pollution, habitat encroachment, etc.) and invasive species affect Norwegian biodiversity, developing evidence-based conservation plans for rare and endangered species, and creating genomic resources to restore damaged or depleted ecosystems. EBP-Nor will act synergistically with existing taxonomical efforts and boost collaboration between molecular biologists and taxonomists.

Scientific value. EBP-Nor will aid us **revise and reinvigorate our understanding of biology in general**, and in particular Norwegian biodiversity, including its adaptations, and ecosystem composition and functions. Norway features almost exclusively species that have immigrated since the last glacial period, during the past 10,000 years. This puts EBP-Nor in a unique position to decipher, by comparative genomic approaches, **the process of speciation itself, and how species respond to climatic and other environmental changes, particularly in light of the vulnerability of high-latitude regions to current climate changes**. This knowledge has direct relevance for (and parallels to) the dynamics **of spread and control of alien species,** including pests, diseases and pandemics. Furthermore, EBP-Nor will **drive development of new methods, techniques and expertise in nascent fields of research (**e.g. genomics, bioinformatics, machine learning, systems biology). A final legacy of EBP-Nor will be publicly available databases (sequences, biobank, metadata) that will fuel future research and innovation.

2.3. Return on investment

The coordinated EBP initiative will fundamentally transform our understanding of life on Earth that we ultimately depend on. For that reason, the return of investment in EBP-Nor, being an EBP affiliate, is expected to be very high. This number is notoriously difficult to quantify

[6] Nærings- og fiskeridepartementet 2019. Blå muligheter – Regjeringens oppdaterte havstrategi. 1–50.

because the outcomes will affect not only science, but also how we protect and make use of biodiversity for benefiting society and human welfare (cf. the three EBP visions).

A reference genome in itself is a result of billions of years of evolutionary fine-tuning and an unprecidented resourse societal and commercial innovations. However, the scope and impact of EBP, goes far beyond just providing the genomic DNA sequences: EBP (and EBP-Nor) will provide a fundament for further research projects utilising the genome information in experimental settings, and for a far more efficient genomic and biological exploration of individuals, populations and comparative (species) studies that will have deep impacts on science, innovation and society. Thus, EBP-Nor will have as a major obligation to facilitate accompanying projects that address specific questions of fundamental and applied nature utlizing the genome information generated by EBP-Nor (for example by far more efficient generation of genome data from individuals).

It is natural to compare EBP with the **Human Genome Project (HGP)**, the international collaboration to sequence the human genome, for which the **return on investment was estimated to 141:1, for USA alone**, for the first 7 years after the completion of HGP (2003-2010)[7]. In 2020, an estimate would be a 10-times increase from 2010 based on the development of the biotech, pharma and genomics industry[8]. Given that EBP will sequence 1.5 million species for a cost similar to or lower than that of sequencing "just" the human genome (by 2018)[9], a conservative estimate is **a return on investment for EBP-Nor ten times that of the Human Genome Project** (in 2010), i.e. 1400:1, within the first two decades of completion.

In practice, **it is expected that innovations and discoveries derived from just a few Norwegian species / genomes may alone generate returns to society far greater than the investment.**

A particularly relevant example is ciclosporin (alternative spelling cyclosporine), an immunosuppressant medication used especially in various organ transplants to reduce the risk of organ rejection. Ciclosporin was isolated in 1971 from the fungus *Tolypocladium inflatum* collected on Hardangervidda, Norway, and came into medical use in 1983. Ciclosporin drugs (marketed by Novartis, Abbot, Merck and more) had **a market size of 1,2 billion USD in 1997** and since then there has been at least a 10-times increase (also driven by treatment of psoriasis, rheumatoid arthritis, Chron's disease and more) – a more than 10 billion USD market! Ciclosporin was in 2017 the 248th most commonly prescribed medication in the US. The World Health Organization considers it among the safest and most effective medicines needed in a health system[9]. Ciclosporin is an established (and old) example and illustrates the time it takes for (some) drugs to reach their potential. Currently, several biotech companies in Norway have developed or are in the process of developing enzymes, compounds, drugs and biomaterials based on our country's biodiversity (some examples: cold adapted enzymes, redox enzymes for depolymerization of cellulose, polyunsaturated fatty acids, antimicrobial compounds and anti-cancer drugs). **EBP-Nor will enable Norway to boost innovation and also to have improved control of the economic exploitation of new drugs, compounds and inventions based on our country's biodiversity genome data.**

[7] Tripp & Grueber 2011. Economic Impact of the Human Genome Project. Battelle Memorial Institute. 1–58.
[8] Senior 2020. Europe's biotech renaissance. Nature Biotechnology 38, 408-415
[9] Lewin et al. 2018. Earth BioGenome Project Sequencing life for the future of life. PNAS 115: 4325–4333.
[10]World Health Organization 2019. World Health Organization model list of essential medicines.

2.4. EBP-Nor technology platform, expertise and national synergies

EBP-Nor is a bottom-up initiative. It will be open to all Norwegian research environments that would like to contribute. This implies that all contributing labs will comply to the same technology standards – decided by EBP and EBP-Nor.

Genomics: The strongest **genomics environments in Norway are associated with the large universities and university hospitals**. Today there is a fairly large sequencing capacity at **UiO and OUS** (counting high capacity short-read machines and long-read (PacBio) technology). In addition, there are sequencing core-facilities with considerable experience and sequencing capacity at **UiB, NTNU, NMBU, Uni Nord and UiT**. With some instrumental additions and increased automatization, the sequencing capacity of Norway will be able to handle the initial phases (1 and 2) of EBP-Nor. The aforementioned institutions together with additional institutions and research institutes (such as UiA, UNIS, IMR, Nofima, NINA, NIVA, NIBIO) will constitute the backbone of EBP-Nor research and technology platform.

Biodiversity biobanking: EBP-Nor will require the establishment of curated biodiversity biobanks for deposition of samples, vouchers (including type materials), and storage of various biomaterials (e.g. tissues, cultures, environmental samples, seeds) for future analysis as well as loans. Samples representing population material for selected species will also be included. Importantly, metadata describing the samples, specimens etc. will be crucial to link up to the genome data. These biobanks will be established and curated in collaboration with the **Natural History Museums in Oslo / Bergen / Trondheim / Tromsø, MarBank (Tromsø) and with the Svalbard Seed Vault**. Sampling and storage of biomaterial will represent a substantial cost (end effort). Biobanking will require appointment of dedicated (curator) positions, but we will also take advantage of the aforementioned existing collections and their personnel.

Bioinformatics, computational biology / statistics and storage: Competence and sufficient capacity within bioinformatics will be keys to the implementation of EBP-Nor. Tight collaborations with **Elixir-Norway** and other bioinformatics environments – e.g. **Center for Bioinformatics** (CBI), UiO; **Computational Biology Unit** (CBU), UiB; **UiT Machine Learning Group** (UiT) and national computational resources (supercomputing resources/ **UNINETT-Sigma2**, **Norstore** etc.) – will be crucial. Furthermore, we see it as strategically advantageous to establish a scientific and organisational association with **the national computational biotechnology initiative Digital Life Norway (DLN)**. DLN has already established a national platform for computation intensive biotech activities, it harbors strong innovation-facilitating support, and represents a national network with research groups and projects overlapping with researchers that will be involved in EBP-Nor. Finally, there is a need to build up storage capacities (and facilitate easy access) to the data that will be generated.

2.5 International collaboration and synergies

As an EBP-affiliated node, EBP-Nor will be an integrated part of the EBP network of communities. This implies that **EBP-Nor is one of the deciding bodies in the international network**, and can suggest and vote for guidelines, approaches and species to be included (when and by which country/institution). Most importantly, we will have a voice whose impact will depend on the strength of Norwegian funding and scientific contribution to the EBP community. EBP itself does not fund any projects (the individual countries, states, or institutions do), but provides the guidelines (a moving target) for genome standards, coordinates which species to be sequenced by which affiliate project, and is the platform for

international communication and co-operation. From a Nordic collaborative point of view, it is fundamental that EBP-Nor has already initiated collaborations with the Swedish and Danish EBP initiatives at Uppsala / Stockholm and Copenhagen / Natural History Museum, respectively. We seek to establish collaborations with also Finnish and Icelandic institutions. **The goal is to have the Nordic countries united in this effort**; this will also provide strength to the Nordic allies in an EBP-setting (i.e. reduce investment for each, but with the same return of investment). At a wider geographic scale, EBP-Nor is already a collaborator with UK through the Sanger Institute / Darwin Three of Life Project (UK) and US institutions through the Vertebrate Genome Project (Rockefeller University) and EBP-US (UCL-Davis). We will seek to obtain collaborations with Canadian Universities as well as other relevant European research (Germany a prime candidate). These coordinated actions with respect to the choice of species for sequencing will be absolutely needed for the feasibility of EBP-Nor.

2.6. Public engagement

As an EBP node, EBP-Nor will actively provide the interested public, including industry and funding bodies, with open access to the benefits of the sequenced eukaryotic genomes of Norway. Particularly, **we will establish user-friendly and pedagogical websites describing the diverse species, their habitats, ecosystems and their genome resources intended for the interested public**. This represents popularized science produced in close collaboration with the Natural History Museums and other partners. We plan to have **thematic events, covering oceanic/marine, freshwater, mountainous and Arctic ecosystems and fjord ecosystems (terrestrial and aquatic)**. Exhibitions at **Natural History Museums** and other relevant locations will also be produced. In order to reach out to the general research and innovation community, including industry, **development of easily accessible, efficiently searchable and continuously updated databases**, is needed. This will be instrumental for the success of EBP-Nor, and it will require allocated resources.

EBP-Nor: Appendices

Appendix 1. Principal steps in genome sequencing

Critical steps in the genome sequencing efforts planned in EBP-Nor include **1) Collection and storage of samples; 2) Genome sequencing; 3) Genome assembly; 4) Genome annotation.** All data regarding the specimens, from sampling/species/metadata to sequences generated to assemblies and annotations need to be stored in an accessible manner (for presentation and distribution).

Sample acquisition and storage are not trivial. For acquisition of a sample regulatory issues need to be resolved, adequate tissues need to be collected and stored in a way that will support the needed quality of DNA (and RNA) for the sequencing. **The metadata** concerning a particular sample are crucial for the project. **DNA sequencing** will be carried out according to EBP's standards. Currently this include long-read sequencing (PacBio), additional long-range protocols (based on Illumina reads) and if possible, optical mapping at the chromosome level. The standard protocols may change over time, and there may be species that we are unable to sequence with the standard EBP approaches. **Genome assembly** will at all times be carried out with the state-of-the-art methods. Methodologically this is a "moving target", improved/more accurate, faster, less computation intensive, are constantly being generated. The main challenge is to be able maintain sufficient bioinformatics competence in Norway to keep up with the development. This naturally has a cost. **Genome annotation** implies that the different elements in the genome (using the assembly) are identified and named. For example, annotation will involve identifying the protein-coding genes (open reading frames; ORFs), RNA coding genes (microRNAs etc.), known repetitive elements (including transposable elements), proviruses, telomers and centromers (and more). Protein-coding genes may be identified by homology to known proteins, from expression data or purely from ORFs. The annotated genomes – along with the primary sequencing, expression and meta-data need to be **stored** allowing for **data presentation and distribution** that will enable scientists in an efficient way to make use of the data.

Appendix 2. Organisation of EBP-Nor

EBP-Nor is a bottom-up initiative, made possible by the engagement of individual researchers and supported by the University rectors and leaders of research institutes. It is therefore crucial that we establish an organization that include all the involved institutions for example as a **scientific working group** (or several working groups) or a general assembly. It is also natural to establish geographically local collaborative teams. In terms of national coordinated governance, **a scientific steering group** representing the main organisations, will be established. It is important that this steering group is able to act with speed and efficiency. A **scientific leader** and a deputy leader will be appointed. Furthermore, **a Board**, representing

the initiative-taking institutions will be established. The leader of the Board should come from a different institution than the Scientific leader. Finally, EBP-Nor will appoint a **Scientific Advisory Committee (SAC)** from EBP-member countries in the Nordic countries, Europe and Overseas.

Appendix 3. Implementation of EBP-Nor

The three scientific phases that we envision are described in the main text. Implementation of EBP-Nor is highly dependent on the funding we are able to attract. However, our plans, following the previously initiated progress (and funding in UK) of our Nordic collaborators in Sweden and Denmark, imply **a start of EBP-Nor early in 2021** (with a strong and ongoing planning phase in 2020). Norway has already initiated a "pre-test" of EBP-standard genomes through the **new versions of the Atlantic cod and Atlantic salmon genomes**. Furthermore, ongoing EBP-standard genomes of haddock, coastal cod, polar cod, burbot, capelin, brown trout, Atlantic puffin and reed warbler (and more) will provide very important information about how implementation of the different phases of EBP-Nor should be planned (in detail). We will engage the public at an early stage – to select for some of the species to be sequenced initially. The iconic genomes, the project, knowledge generated will be communicated to the general public. In the next steps we will follow a phylogenetically guided approach where the main orders, families and finally at the genus level, will be sequenced. The detailed plans will have to be made as a common effort with our Nordic and Northern Europe and American collaborators.

We envision that, in addition to the sampling and sequencing capacity, sufficient **bioinformatics competence** and **computational capacity** will be crucial for the national implementation of EBP-Nor. Finally, successful implementation includes **an outreach program** that captures all the relevant groups.

Appendix 4. Estimations of costs for the three phases of EBP-Nor

The three phases as described in the main text consist of 1) **Launch** (*proof of concept*), 2) **Ascent** (*scaling up*) and 3) **Orbit** (production phase). This implies that we envision an initial phase (Launch) where we prove that the technology, infrastructure, collaborations, national share of labor and the organization of EBP-Nor is up and running. The second phase (Ascent) is intented to show that we can do large-scale reference genomes (and analyses, storage, curation and availability) using the pipelines established in the Launch phase. The third phase (Orbit) is the production phase where we will be up to "maximum speed" sequencing thousands of species, selected ecosystems as well as population genomics (on selected species). The strategy suggested here will naturally be subject to revisions along the line.

Cost estimates:

Launch (1-2 years; 20-50 species; establishing the platform and national collaboration ("dugnad"):

Total costs (sequencing, analyses, infrastructure, coll. ***15-20*** *million NOK*

Ascent (2 years; 200-500 species, biobank establishment, connecting genome data to metadata, data storage, "outreach"/communication, industry, Nordic collaborations)

Total costs (sampling, biobanking, sequencing, analyses +) ***100-200*** *mill NOK*

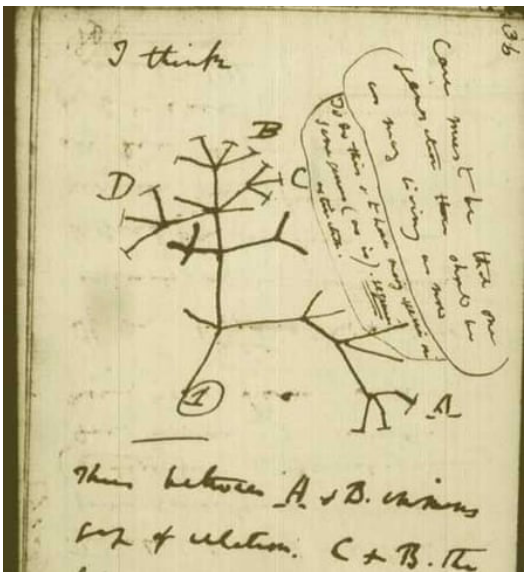
Orbit (6-7 years); thousands of species, ecosystem genomics, pop. genomics, Nordic and international coll., strong societal (and industry) involvement

Total costs (sampling, biobanking, sequencing, industry) ***1000–1700*** *mill NOK*

The numbers here given are "guestimates", but based on other EBP projects such as "The Darwin Tree of Life (UK) where the total cost of sequencing all (estimated) 65.000 species in Great Britain has been estimated to be around 8000 million NOK. The Californian EBP initiative have so far reached a current funding of 100 million NOK (for only the conserbation biology part; i.e. endangered species). It we time this right, establish an efficient communication and collaboration in the Nordic countries, with Europe and the North America, we will be able do this far more cost efficient than planned for UK. An assumption may be that we do 2000-5000 species (due to species overlap between the countries) plus ecosystem and population genomics on selected sites and species. This will reduce the total costs for the **Orbit phase to 1000 – 1700 milion NOK** (as depicted above).

We envision that that EBP-Nor **will attract substantial private funding** from the marine-oriented sector, biotechnology, bioproduction, bioprospecting, chemical industry, medicine, energy companies as well as the governmental resource management, biodiversity protection, environmental agencies and more.

The establishment and running of the national Working group, writing of this document has been made possible by the support of:



Darwin's sketch of the «Tree of Life» from 1837. This famous drawing came to symbolise the theory of evolution by natural selection as a tree with discreate evolutionary brances. Today we know that this is more complex; ancient (and current) hybridisations and horizontal gene transfer between fairly closely related and distantly related organisms have contributed substantially to the Life on Earth. For example antibiotic resistance is basically a result of ongoing horizontal gene transfer.

Although the Tree of Life may be more like a «Web of Life» we still see the outline of hieractical structure as Darwin descibed in his fairly young age, but with web-like features as well – which is crucial knowlede for the field of infectious disease as well as biology in general.