

The Catalan BioGenome Project
(El projecte de seqüenciació del genoma de totes les espècies eucariotes endèmiques dels Països Catalans)

Montserrat Corominas Guiu (UB, president SCB)

Roderic Guigó Serra, (CRG, UPF, Secció de Ciències Biològiques de l'IEC).

The EBP. Understanding Earth's biodiversity and responsibly administrating its resources is among the top scientific and social challenges of this century. The Earth BioGenome Project (EBP) aims to sequence, catalog and characterize the genomes of all of Earth's eukaryotic biodiversity over a period of 10 years (<http://www.pnas.org/content/115/17/4325>). The outcomes of the EBP will inform a broad range of major issues facing humankind, such as the impact of climate change on biodiversity, the conservation of endangered species and ecosystems, and the preservation and enhancement of ecosystem services. It will contribute to our understanding of biology, ecology and evolution, and will facilitate advances in agriculture, medicine and in the industries based on life: it will, among others, help to discover new medicinal resources for human health, enhance control of pandemics, to identify new genetic variants for improving agriculture, and to discover novel biomaterials and new energy sources, among others.

At a sequencing cost of about US \$1,000 per eukaryotic genome, the estimated cost of the EBP to sequence the approximate 1.5 million known eukaryotes is about \$4.5 billion, which includes, in addition, the costs of sample acquisition, data storage and analysis. The far-reaching potential benefits of the EBP can be realized only by a coordinated international effort. The EBP envisions an organization model based on self-organized projects, many of which already exist, either targeting specific taxa, such as the Initiative to Sequence 5000 Arthropod Genomes (i5K), Bird 10,000 Genomes Project (B10K, <https://b10k.genomics.cn/>), 1000 Fungal Genomes Project (<https://genome.jgi.doe.gov/programs/fungi/1000fungalgenomes.jsf>), 10,000 Plant Genomes Project (10KP, <https://db.cngb.org/10kp/>), or biogeographical regions. Indeed, streamlined access to biological samples will often require the participation of naturalists familiar with the regional ecosystems. In this regard, the UK has launched on November 1st, 2018, the "Darwin Tree of Life Project" (<https://www.sanger.ac.uk/news/view/genetic-code-66000-uk-species-be-sequenced>) to sequence the genomes of the 66,000 eukaryotic species that inhabit the British Islands. The project will be lead by the Wellcome Trust Sanger Institute with a cost of about £100 million over the first five years.

The CBP. Similarly, we propose here, as part of the EBP, the Catalan BioGenome Project (CBP) with the aim of sequencing all eukaryotic species that are endemic (or indigenous) of *els Països Catalans*. *Els Països Catalans* are part of the Mediterranean basin, one of the world's 35 biodiversity hotspots - Earth's most biologically rich, yet threatened, areas. Covering more than 2 million square kilometers, this hotspot is the second largest in the world, and ranked third- richest in terms of plant diversity. It hosts, 13,000 endemic animal species. In terms of aquatic diversity, though the Mediterranean only covers 0.7% of the world's ocean area it is one of the major reservoirs of marine and coastal biodiversity, with 28% of endemic species and

7.5% of the world's marine fauna and 18% of its marine flora. At the same time, human activities have profoundly shaped the ecosystems of the Mediterranean Basin throughout recent history. These ecosystems are therefore a point of reference for understanding the effects of global change in biodiversity. *Els Països Catalans* because of the great variety of habitats and biotopes within a rather small territory, capture proportionally a large fraction of Mediterranean biodiversity.

Opportunity. The combination of a rich naturalist tradition and of the recent development of powerful technological infrastructures in genome sequencing and computing, places Catalonia in a unique position to successfully implement the CBP, and to make a substantial contribution to the EBP. The CBP will, on the one hand, help our scientists to play a leadership role in within the EBP, and in research in biodiversity in general. On the other hand, within the country, it will contribute to reinvigorate Natural History and research in biodiversity. It will also help to locally streamline the translation of basic research in biodiversity to medicine, agriculture and biology-based industries.

Proposed Structure (see attached Figure). The CBP will be lead by the IEC (SCB and ICHN). The CBP fits particularly well within the mandate of the IEC, which includes the promotion and development of research in the different fields of science and technology, specifically within the framework of *els Països Catalans*. It also appeals to the broad spectrum of biological disciplines distributed across the many sections of the SCB: molecular biology, genomics, bioinformatics, evolutionary biology, microbiology, ecology, and others. The two main components off the CBP will be, on one hand, the naturalist institutions from *els Països Catalans*, possibly lead by the ICHN (and including among others: Museu de Ciències Naturals, Zoològic, the Jardins Botànics from Barcelona and Valencia, and many non governmental (NG) naturalist and conservation groups, such as the Grup Ornithological Balear, etc.), which will be in charge of cataloguing and prioritizing species, and of sample acquisition. The second component will be the leading infrastructures in genomics (CNAG-CRG) and computing (BSC) that exist in the country. The CBP will also benefit from the strong research in genomics (CRG), biodiversity (IRBio, CREA), Agriculture (CIRAD) and Institute of Marine Sciences (Ciències del Mar), as well as in the research in taxonomy and phylogeny in many departments at the universities from *els Països Catalans*.

Pilot Phase (2019). The pilot phase of the project will focus on :

1. Cataloguing all eukaryotic endemic and/or indigenous species inhabiting *els Països Catalans*.
2. Prioritizing species for the CBP, including those targeted in the pilot phase. This will help to define the scale of the project, which we estimate between 5,000 and 10,000 species; note that there are more than 3000 indigenous plant species in Catalonia)
3. Establishing the Standard Operation Procedures (SOPs) for sample acquisition and preservation
4. Defining the protocols to transfer samples and data between the different components of the project.
5. Establishing the project's budget and securing the needed funding.

6. The pilot phase should also include the sequencing of a number of species (from 100 to 1000) corresponding to a budget between 0.5 and 2M euros.

Timeline (Pilot Phase)

1. Initial meeting at the IEC (Nov 2018). Finalizing of the white paper.
2. Initiate contacts with possible sponsoring/funding organizations. Ideally during 2018.
3. Forming of the Steering Committee, including representatives from the IEC (Secció de Ciències Biològiques, SCB and ICHN), CNAG-CRG and BSC. Ideally during 2018.
4. Open meeting aimed at the local community to present and discuss the project and to collect expressions of interest. First quarter of 2019
5. B-Debate including world-wide leaders in genome initiatives related to the EBM, and the local state holders.
6. Creation of the CBP (before summer 2019). (Web page to promote the project and to open the crow funding).
7. Sequencing of the species of the pilot phase.

Funding. The sources of funding would/could include among others:

- IEC
- Universities from *els Països Catalans*.
- Governments of Catalonia, Balearic Islands, Valencia and Andorra
- Museums of science and natural history, Parc Zoològic, Jardí Botànic, etc.
- NG naturalistic and conservation organizations and related NGs.
- Private organizations: La Caixa, Banc Sabadell, CosmoCaixa, ...
- Crowdfunding .

Outlook. The infrastructures, the protocols and the expertise developed during the CBP could serve to catalyze larger and more ambitious internationally wide initiatives, lead from our country. This could include, for instance, the sequencing of the genomes of all species of the Iberian peninsula (The Iberian BioGenome Project) and the sequencing of all species of the Mediterranean Basin (The Mediterranean BioGenome Project).

Acknowledgements. The idea for the CBP was initially suggested by Antonio Martinez, during a late conversation on a summer night.

