

**Note:** This document contains the Catalan translation of the article “The Catalan initiative for the Earth BioGenome Project: contributing local data to global biodiversity genomics” by Corominas *et al.* NAR Genomics and Bioinformatics (2024), <https://doi.org/10.1093/nargab/lgae075>. Please use this citation when referring to this work. In the Catalan version, the Material and Methods section has moved to the end compared to the original article.

**Nota:** Aquest document conte la versió catalana de l'article “The Catalan initiative for the Earth BioGenome Project: contributing local data to global biodiversity genomics” per Corominas *et al.* NAR Genomics and Bioinformatics (2024), <https://doi.org/10.1093/nargab/lgae075>. Si us plau, utilitzeu aquesta citació quan feu referència a aquest treball. En comparació amb l'article original, en la versió en català, la secció de Materials i Mètodes s'ha mogut al final.

## La iniciativa catalana per a l'Earth BioGenome Project: contribuint amb dades locals a la genòmica de la biodiversitat global

Montserrat Corominas<sup>1,2,3,\*†</sup>, Tomàs Marquès-Bonet<sup>4,5,6,7,†</sup>, Miquel A. Arnedo<sup>8,9</sup>, Mònica Bayés<sup>6,10</sup>, Jordina Belmonte<sup>11,12</sup>, Héctor Escrivà<sup>13</sup>, Rosa Fernández<sup>4</sup>, Toni Gabaldón<sup>5,14,15,16</sup>, Teresa Garnatje<sup>17,18</sup>, Josep Germain<sup>19</sup>, Manel Niell<sup>20</sup>, Ferran Palero<sup>21</sup>, Joan Pons<sup>22</sup>, Pere Puigdomènech<sup>3,23</sup>, The Catalan Initiative for the Earth BioGenome Project, Vanesa Arroyo<sup>20</sup>, Cristian Cuevas-Caballé<sup>1,9</sup>, Joan Ferrer Obiol<sup>24</sup>, Ivo Gut<sup>6,10</sup>, Marta Gut<sup>6,10</sup>, Oriane Hidalgo<sup>16,25</sup>, Guillem Izquierdo-Arànega<sup>1,9</sup>, Laia Pérez-Sorribes<sup>17,26</sup>, Emilio Righi<sup>27</sup>, Marta Riutort<sup>1,9</sup>, Joan Vallès<sup>3,9,28</sup>, Julio Rozas<sup>1,9</sup>, Tyler Alioto<sup>6,10</sup> i Roderic Guigó<sup>3,27,29,\*</sup>

<sup>1</sup> Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), 08028 Barcelona, Catalunya, Espanya

<sup>2</sup> Institut de Biomedicina (IBUB), Universitat de Barcelona (UB), 08028 Barcelona, Catalunya, Espanya

<sup>3</sup> Institut d'Estudis Catalans (IEC), 08001 Barcelona, Catalunya, Espanya

<sup>4</sup> Institut de Biologia Evolutiva (IBE, UPF-CSIC), PRBB, 08003 Barcelona, Catalunya, Espanya

<sup>5</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalunya, Espanya

<sup>6</sup> Centre Nacional d'Anàlisi Genòmica (CNAG), 08028 Barcelona, Catalunya, Espanya

<sup>7</sup> Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalunya, Espanya

<sup>8</sup> Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), 08028 Barcelona, Catalunya, Espanya

<sup>9</sup> Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), 08028 Barcelona, Catalunya, Espanya

<sup>10</sup> Universitat de Barcelona (UB), 08028 Barcelona, Catalunya, Espanya

<sup>11</sup> Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Catalunya, Espanya

<sup>12</sup> Institut de Ciència i Tecnologia Ambientals (ICTA-UAB), Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Catalunya, Espanya

<sup>13</sup> Universitat Sorbona, CNRS, Biologie Intégrative des Organismes Marins, BIOM, F-66650, Banyuls de la Merenda, França

<sup>14</sup> Barcelona Supercomputing Center - Centre Nacional de Supercomputació (BSC-CNS), 08034 Barcelona, Catalunya, Espanya

<sup>15</sup> Institut de Recerca Biomèdica (IRB Barcelona), Institut de Ciència i Tecnologia de Barcelona (BIST), 08028 Barcelona, Catalunya, Espanya

- <sup>16</sup> CIBER de Enfermedades Infecciosas, Instituto de Salud Carlos III, 28029 Madrid, Espanya
- <sup>17</sup> Institut Botànic de Barcelona (IBB), CSIC-CMCNB, 08038 Barcelona, Catalunya, Espanya
- <sup>18</sup> Jardí Botànic Marimurtra - Fundació Carl Faust, 17300 Blanes, Catalunya, Espanya
- <sup>19</sup> Institució Catalana d'Història Natural, 08001 Barcelona, Catalunya, Espanya
- <sup>20</sup> Andorra Recerca + Innovació (ARI), AD600 Sant Julià de Lòria, Andorra
- <sup>21</sup> Institut Cavanilles de Biodiversitat i Biologia Evolutiva (ICBIBE), 46980 Paterna, València, Espanya
- <sup>22</sup> Departament de Biodiversitat Animal i Microbiana, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), 07190 Esporles, Illes Balears, Espanya
- <sup>23</sup> Centre de Recerca en Agrigenòmica, CSIC/IRTA/UAB/UB, 08193 Bellaterra, Catalunya, Espanya
- <sup>24</sup> Departament de Ciències Ambientals i Polítiques, Universitat de Milà, 20122 Milà, Itàlia
- <sup>25</sup> Royal Botanic Gardens, Kew, TW9 3DS Richmond, Regne Unit
- <sup>26</sup> Estación Biológica de Doñana, CSIC, 41092 Sevilla, Espanya
- <sup>27</sup> Centre de Regulació Genòmica (CRG), Institut de Ciència i Tecnologia de Barcelona (BIST), 08003 Barcelona, Catalunya, Espanya
- <sup>28</sup> Laboratori de Botànica (UB), Unitat Associada al CSIC, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, 08028 Barcelona, Catalunya, Espanya
- <sup>29</sup> Universitat Pompeu Fabra (UPF), 08003 Barcelona, Catalunya, Espanya

\* Autors a qui cal adreçar la correspondència. Montserrat Corominas, Telèfon: +34 934 037 003; correu electrònic: [mcorominas@ub.edu](mailto:mcorominas@ub.edu), i Roderic Guigó, correu electrònic: [roderic.guigo@crg.cat](mailto:roderic.guigo@crg.cat). La correspondència en català també es pot adreçar a Marta Riutort, correu electrònic: [mriutort@ub.edu](mailto:mriutort@ub.edu) i Julio Rozas, correu electrònic: [jrozas@ub.edu](mailto:jrozas@ub.edu).

† Contribució comparable

## Resum

La Iniciativa Catalana per a l'Earth BioGenome Project (CBP) és una xarxa de projectes afiliada a l'EBP que té com a objectiu seqüenciar el genoma de les més de 40.000 espècies eucariotes que s'estima que viuen als territoris de parla catalana, territoris que ocupen una àrea de gran biodiversitat. Tot i cobrir menys de l'1 % de la superfície d'Europa, són la llar d'aproximadament una quarta part de totes les espècies eucariotes conegudes d'aquest continent, que inclouen una alta proporció d'endemismes, molts dels quals estan amenaçats. Aquesta tendència probablement empitjorarà, ja que es preveu que els efectes del canvi global seran especialment greus a tota la conca mediterrània, particularment en els ecosistemes d'aigua dolça i les zones de muntanya. Seguint el model de l'EBP, la CBP és una organització en xarxa que ha estat capaç d'involucrar molts socis científics i no científics. En la fase pilot, s'estan seqüenciant els genomes de 52 espècies. Com a model d'estudi en la conservació de la biodiversitat, aquí destaquem el genoma de la baldriga balear (*Puffinus mauretanicus*), seqüenciat sota el paraigua de la CBP.

## Introducció

El Projecte del Biogenoma de la Terra o Earth BioGenome Project (EBP, <https://www.earthbiogenome.org>) és una xarxa internacional de xarxes que té com a objectiu seqüenciar els genomes de les aproximadament 1,8 milions d'espècies eucariotes conegudes que viuen a la Terra (1–3). Dins de l'EBP, la Iniciativa Catalana per a l'Earth BioGenome Project (CBP, <https://www.biogenoma.cat>) és una xarxa de projectes afiliada que té com a objectiu principal seqüenciar els genomes de les més de 40.000 espècies eucariotes que s'estima que viuen als territoris de parla catalana. Aquests territoris, que s'estenen per Espanya, França, el Principat d'Andorra i la ciutat sarda de l'Alguer (Itàlia), han compartit històricament una forta tradició cultural, principalment reflectida en l'ús de la llengua catalana. Es troben a la intersecció de les plaques europea i africana (4), i a la cruïlla entre les regions biogeogràfiques eurosiberiana i mediterrània. Aquests territoris, tot i cobrir menys de l'1 % del territori europeu (70.520 km<sup>2</sup>, amb 2.500 km de costa), alberguen aproximadament una quarta part de totes les espècies eucariotes conegudes d'Europa. També es caracteritzen per un elevat nombre d'endemismes. Moltes espècies endèmiques estan amenaçades, una tendència que empitjorarà en el futur, ja que l'impacte del canvi climàtic probablement serà intens a la conca mediterrània, particularment a les zones de muntanya (5-8).

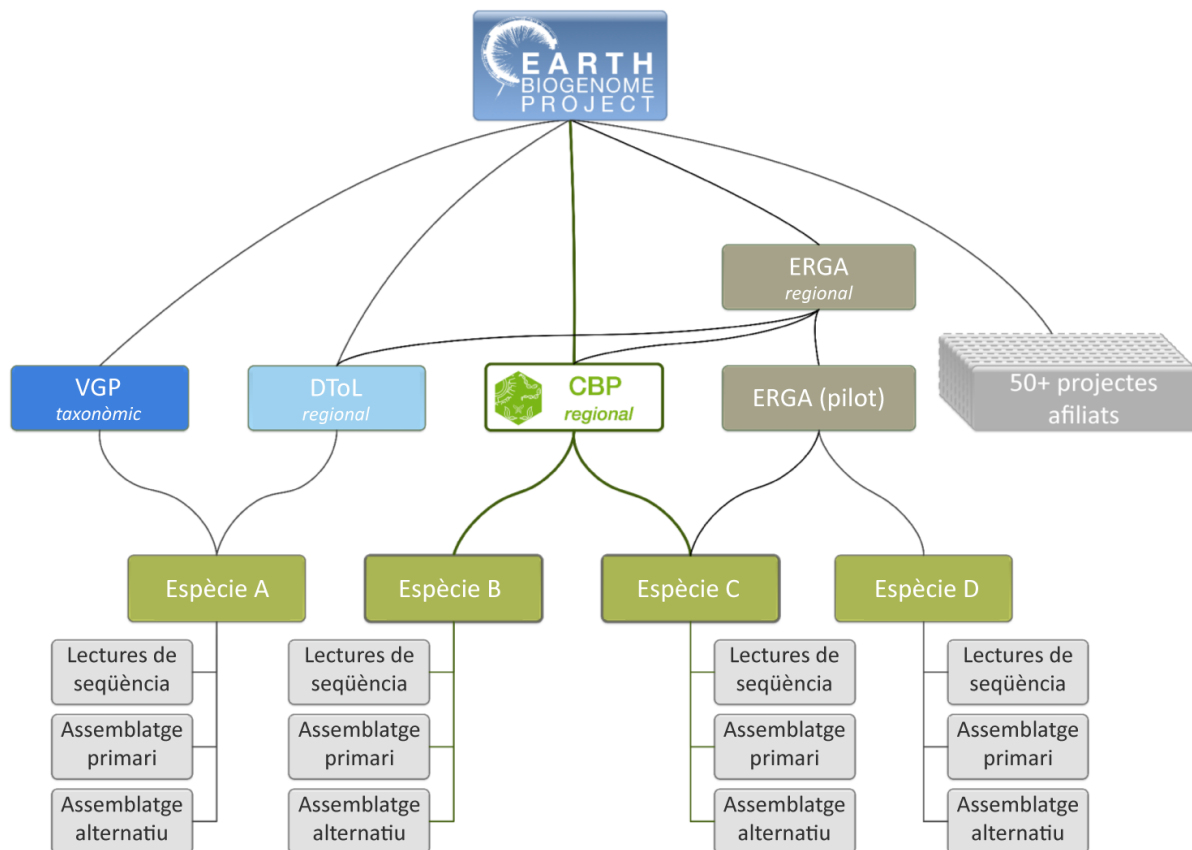
En aquest article, descrivim la CBP, incloent-hi la justificació del projecte, els seus orígens i història, la seva estructura i estat actuals, i la seva direcció futura. Per la seva naturalesa global, l'èxit de l'EBP depèn, en gran manera, de la participació de les comunitats científiques de tot el món i, ateses les seves implicacions més enllà de la ciència, representa una oportunitat única per a involucrar la societat en el seu conjunt en la promoció de la consciència sobre la rellevància de la biodiversitat per a la salut regional i global. La CBP ha estat capaç d'involucrar molts actors científics, acadèmics i socials, i, en aquest sentit, és un exemple d'iniciatives regionals i transnacionals de baix a dalt en el camp de la genòmica de la biodiversitat. En l'article, també informem sobre algunes característiques del genoma de la baldriga balear (*Puffinus mauretanicus*), un dels primers genomes seqüenciats per la CBP, com a exemple de la rellevància de la genòmica en la conservació de la biodiversitat.

## Resultats

### La Iniciativa Catalana per a l'Earth BioGenome Project

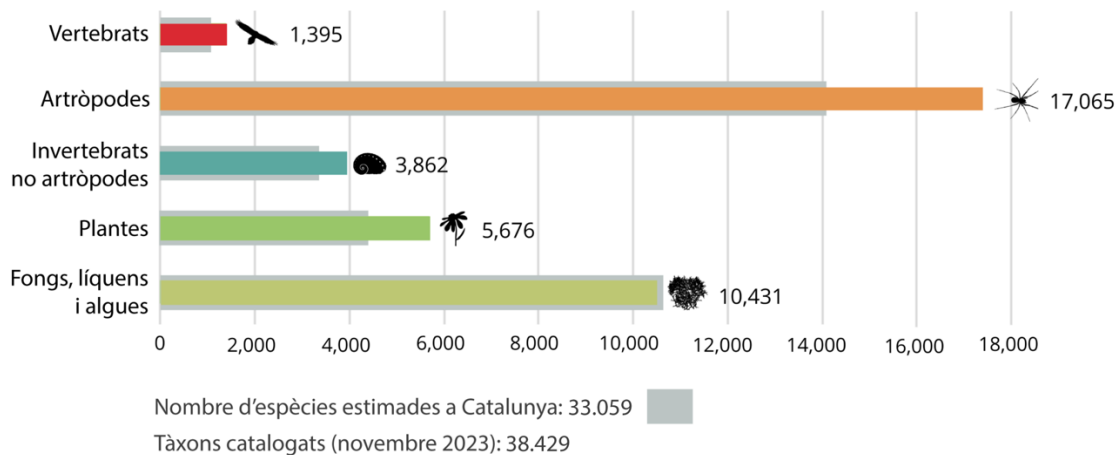
La CBP és una organització en xarxa impulsada el 2019 per la Societat Catalana de Biologia (SCB, <https://scb.iec.cat>) i la Institució Catalana d'Història Natural (ICHN, <https://blogs.iec.cat/ichn>), dues societats filials de l'Institut d'Estudis Catalans (IEC, <https://www.iec.cat>). L'IEC és l'acadèmia encarregada de la promoció de la ciència i la cultura als territoris catalans. L'octubre de 2018 la SCB va publicar un llibre blanc que descriu els antecedents, els objectius i un esbós proposat de la CBP (<https://www.biogenoma.cat/wp-content/uploads/2023/04/2018CatalanBioGenomeProjectNov12b.pdf>). El finançament inicial per a començar les activitats de la CBP va provenir de l'IEC i del llegat de Leandre Cervera, president de la SCB (1935-1963) en bona part durant el període en què les activitats de la Societat estaven prohibides (16). El finançament principal per a la primera fase (2023-2025) és proporcionat pel Departament de Recerca i Universitats de la Generalitat de Catalunya a l'IEC. La CBP també compta amb el suport d'Andorra Recerca + Innovació (ARI, <https://www.ari.ad/>).

La CBP és un membre afiliat de l'EBP i un projecte regional associat a l'European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu>). ERGA és el soci paneuropeu de l'EBP i té com a objectiu coordinar la producció de genomes d'alta qualitat representatius de la biodiversitat eucariota a Europa (17, 18). Orgànicament, dins de la International Nucleotide Sequence Database Collaboration (INSDC), el BioProject de la CBP (número d'accés PRJEB49670) penja directament dels BioProjects EBP (PRJNA533106) i ERGA (PRJEB43510) (Figura 1).



**Figura 1.** Les dades i els assemblatges dels genomes enviats a la International Nucleotide Sequence Database Collaboration (INSDC) s'agrupen en projectes paraigua quant a l'espècie, que, al seu torn, es poden vincular a un o més projectes afiliats a l'EBP. Els projectes regionals, com la CBP i el Darwin Tree of Life (DTOL) (3), es poden agrupar en iniciatives regionals més grans, com l'European Reference Genome Atlas (ERGA).

La fase pilot de la CBP es va llançar l'estiu de 2020, amb una convocatòria oberta per a projectes de seqüenciació, seguida d'una segona convocatòria oberta l'estiu de 2021. Durant aquesta fase, la ICHN va treballar en paral·lel per a crear un catàleg digitalitzat de les espècies eucariotes que viuen a Catalunya, basant-se en catàlegs anteriors (19). El novembre de 2023, aquest catàleg contenia 38.429 espècies, però probablement continuarà més de 40.000 espècies quan sigui complet (Figura 2). En qualsevol cas, aquest catàleg servirà de referència per a prioritzar els genomes a seqüenciar durant les diferents fases de la CBP. Els criteris de priorització inclouen la posició i novetat filogenètica, l'interès dels grups de recerca locals, el grau d'endemisme, l'estat de conservació i els interessos biomèdics, agrícoles o industrials, i l'adhesió als principis de l'ERGA (18, 20).



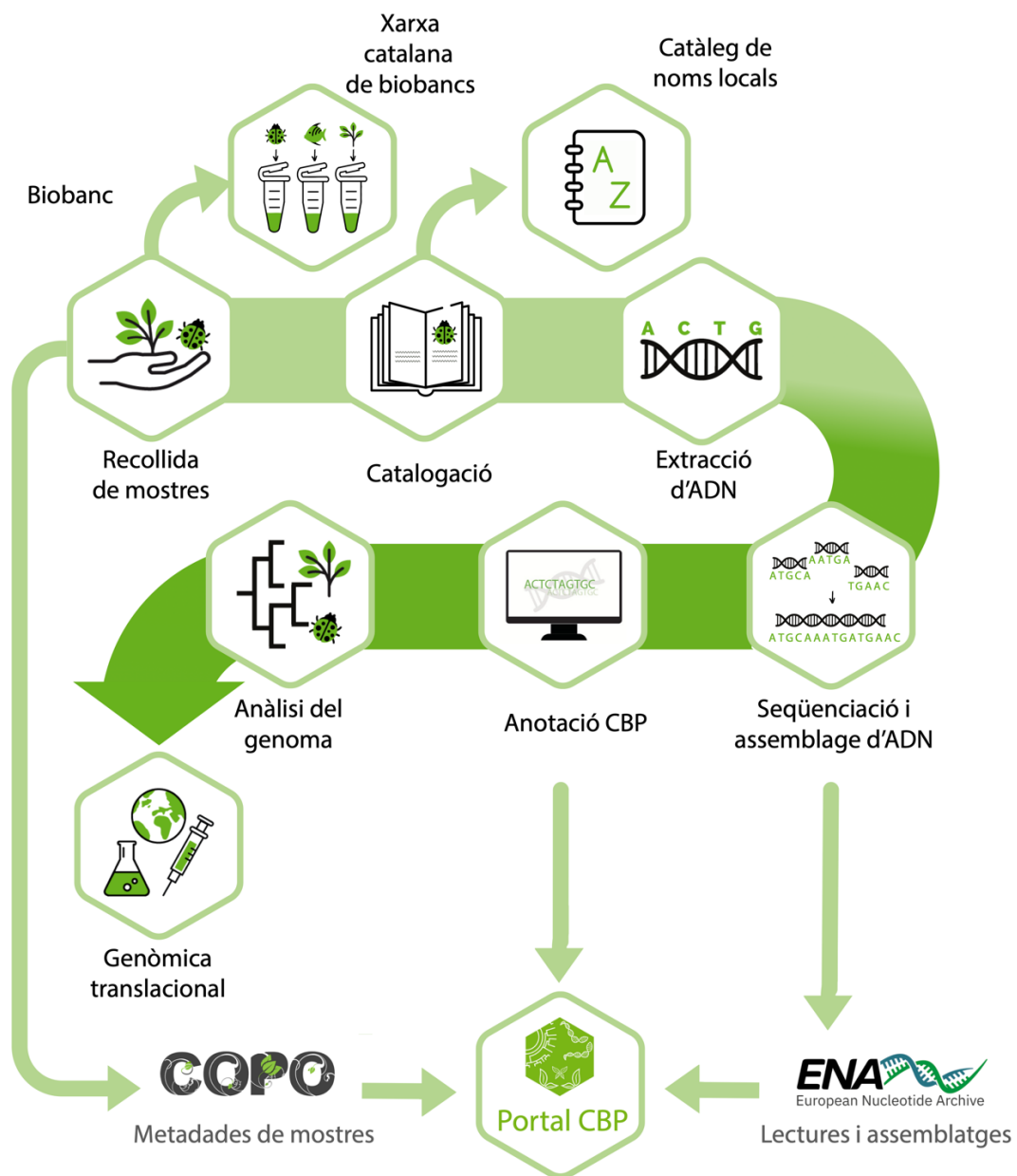
**Figura 2.** Catàleg de biodiversitat de Catalunya. Nombre de tàxons estimats inicialment (gris) i catalogats actualment de diferents grups taxonòmics. Les xifres fan referència exclusivament a Catalunya i no inclouen la resta de territoris de parla catalana.

L'objectiu general de la fase pilot ha estat, però, connectar tots els components existents per a crear la infraestructura necessària per a una producció eficient de seqüències de genomes. Seguint àmpliament el model del Darwin Tree of Life (2) i de l'ERGA (20) (Figura 3), en primer lloc, es recullen, identifiquen, anoten, codifiquen i guarden en biobancs mostres de les espècies prioritzades en cada fase. L'estructura de biobanc de la CBP encara s'està definint, però els principals centres inclouen el Museu de Ciències Naturals de Barcelona, el biobanc del Zoo de Barcelona i el Centre Balear de Biodiversitat per a les mostres d'animals, i l'Institut Botànic de Barcelona i el Centre de Recerca en Agrigenòmica per a les mostres de plantes. A continuació, les mostres se sotmeten a seqüenciació d'ADN i ARN. El centre de seqüenciació principal de la CBP és el Centre Nacional d'Anàlisi Genòmica (CNAG). Al CNAG, la seqüenciació es realitza generalment utilitzant tecnologies de lectures curtes i llargues, així com Hi-C (captura de conformació de cromosomes) amb l'objectiu de complir els estàndards de l'EBP (21), incloent-hi un N50 mínim d'1 Mb (un milió de bases) per als còntigs (és a dir, el 50% dels còntigs), un N50 d'escala cromosòmica pels esquelets, i una qualitat de consens superior a QV40 sempre que sigui possible. Després del control de qualitat de les dades, els genomes s'assemblen ([https://github.com/cnag-aat/assembly\\_pipeline](https://github.com/cnag-aat/assembly_pipeline)), es curen manualment (<https://gitlab.com/wtsi-grit/rapid-curation>) i s'anoten. La CBP produeix la seva pròpia anotació, utilitzant un conjunt de procediments (*pipelines*) del CNAG (<https://github.com/cnag-aat>), que es distribueix a través del portal de la CBP (vegeu més avall). Això s'ha de considerar complementari a l'anotació produïda dins del consorci ERGA, que també es pot distribuir a través del portal. Les dades inicials (lectures de seqüència) i els assemblatges es distribueixen obertament als repositoris públics designats.

Procediments addicionals per a l'anotació gènica (22), l'anàlisi funcional (23) i l'anàlisi filogenòmica (24) s'estan reimplementant com a fluxos de treball escalables i reproduïbles utilitzant NextFlow (25) i SnakeMake (26).

També s'està fent un esforç significatiu dins de la CBP cap a la divulgació i el compromís públic. Durant la fase pilot, el projecte s'ha presentat en diverses reunions i tallers internacionals. S'han realitzat nombroses xerrades i seminaris per al públic en general, incloent-hi dos tallers d'extracció d'ADN i una exposició destinada a estimular l'interès públic per la genòmica (vegeu [www.biogenoma.cat/en/outcomes](http://www.biogenoma.cat/en/outcomes)). S'han distribuït pancartes de

carrer amb informació sobre les activitats de la Iniciativa a Barcelona i altres ciutats (Figura suplementària S1). També es poden descarregar fullets informatius des del portal web ([www.biogenoma.cat](http://www.biogenoma.cat)).



**Figura 3.** Flux de treball de la CBP (vegeu el text principal per a una explicació).

### Estat actual

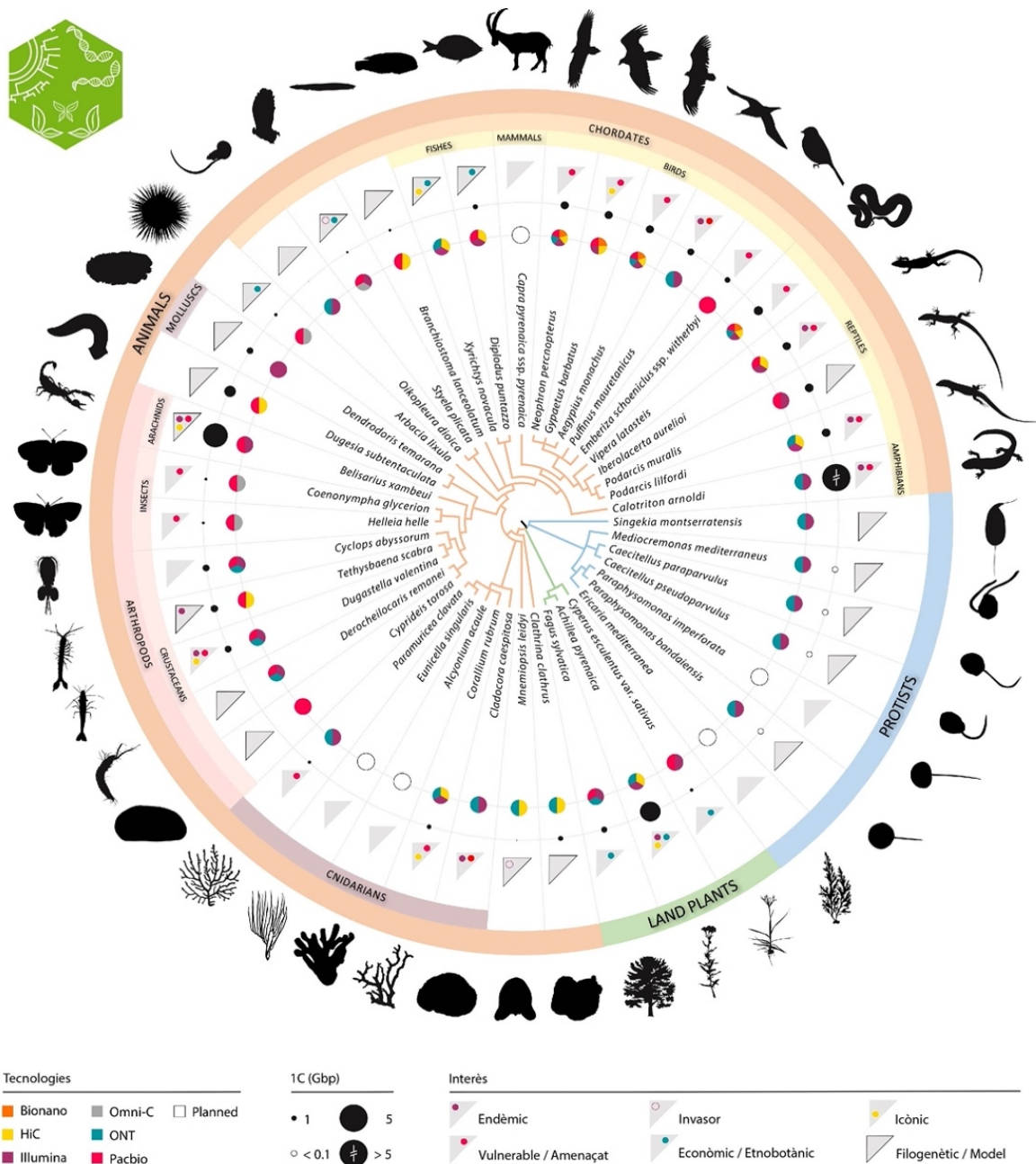
El novembre de 2023, 52 genomes, que representen espècies de diversos grups taxonòmics, hàbitats i ecosistemes, estan en procés sota el paraigua de la CBP (<https://goat.genomehubs.org/projects/CBP>), cinc dels quals ja han estat publicats (9,27-30) (Figura 4). La CBP ha seguit un enfocament de baix a dalt, responent als interessos de la comunitat investigadora local. Les espècies objectiu en la fase inicial inclouen aquelles pertanyents a tàxons poc explorats, com el flagel·lat d'aigua dolça (*Singekia montserratensis*),

un eucariota microbià (31); espècies rares, endèmiques o difícils de localitzar, com l'escorpi cec català (*Belisarius xambeui*), l'amfipode monotípic de les aigües subterrànies (*Sensonator valentiensis*), el cuc gegant català (*Norana najaformis*), la gamba cangur (*Dugastella valentina*), la gamba amb bigotis (*Derocheilocaris remanei*) i la baldriga balear (*Puffinus mauretanicus*) (vegeu més avall); espècies emergents com a models biològics, com la sargantana de paret (*Podarcis muralis*) i el sarg imperial (*Diplodus puntazzo*) que també estan en el punt de mira d'aquesta fase pilot. L'objectiu de la CBP és també proporcionar genomes per a espècies en perill d'extinció, com el tritó del Montseny (*Calotriton arnoldi*), l'amfibi més amenaçat d'Europa, o l'icònic coral vermell (*Corallium rubrum*); espècies tradicionalment usades en medicina, com la camamilla pirinenca (*Achillea ptarmica* subsp. *pyrenaica*); i espècies d'interès econòmic, com el rascló perlat (*Xyrichtys novacula*) o la xufa (*Cyperus esculentus*), entre d'altres.

Una prioritat dins de la CBP és el desenvolupament de mètodes bioinformàtics per a ajudar a afrontar els desafiaments que representen la gran quantitat de dades dels projectes de genòmica de la biodiversitat i el compromís amb l'accessibilitat ràpida de les dades. Així, els esforços s'han centrat específicament en el desenvolupament del BioGenome Portal (32) (<https://ebp.biogenoma.cat/>), una eina web personalitzable dissenyada per a millorar l'accessibilitat a les dades generades per la CBP o per qualsevol altre projecte de genòmica de la biodiversitat.

### **El genoma de la baldriga balear: un model d'estudi en la conservació de la biodiversitat**

El genoma de la baldriga balear (*Puffinus mauretanicus*) és un dels primers genomes publicats sota el paraigua de la CBP (9). Es pot accedir a la seqüència genòmica de la baldriga balear i a les seves dades associades a través de l'ENA (Arxiu Europeu de Nucleòtids a <https://www.ebi.ac.uk/ena/browser/view/PRJNA780920>) i del portal de la CBP a (<https://dades.biogenoma.cat/organisms/48681>). Aquest ocell marí pelàgic de mida mitjana, endèmic de les Illes Balears, està catalogat com a críticament amenaçat per la Unió Internacional per a la Conservació de la Natura (UICN) i les seves poblacions estan disminuint a una taxa anual alarmant del 7,4-14%. Les principals amenaces per a la supervivència de l'espècie provenen de les activitats humanes, especialment la captura accidental en pesqueries d'arts de palangre i la depredació a les colònies de cria per espècies invasores al·lòctones (33, 34). Reduir les taxes de captura accidental en pesqueries, assegurar hàbitats de cria lliures de depredadors i protegir l'hàbitat d'alimentació són mesures de conservació urgents per a evitar l'extinció de l'espècie (35-37). La disponibilitat d'un genoma de referència i de dades de genòmica de poblacions per a la baldriga balear millorarà la proposta de polítiques de conservació basades en la ciència que contribueixin a la implementació de les mesures esmentades.



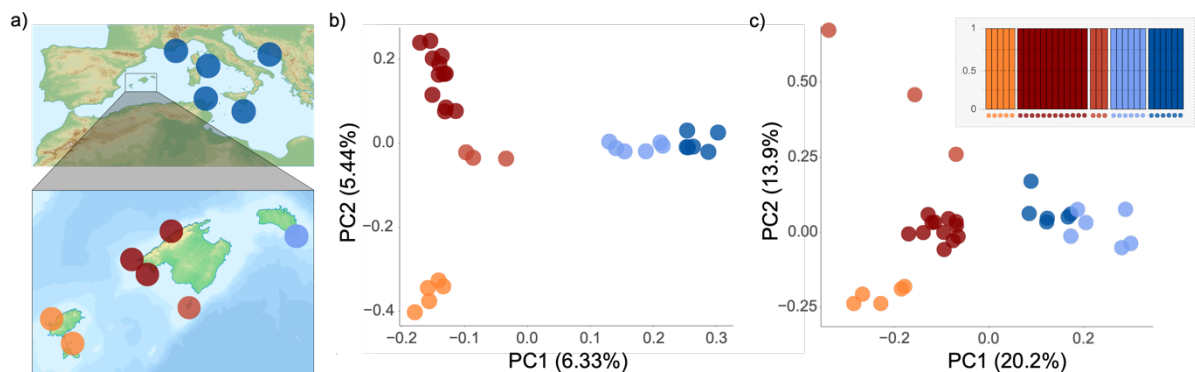
**Figura 4.** Projectes de genomes sota el paraigua de la Iniciativa Catalana per a l'Earth BioGenome Project (CBP), a juny de 2023. La figura recull la posició filogenètica de l'espècie, la mida del genoma quan es coneix, les tecnologies emprades per a la seqüenciació i la justificació de la selecció d'espècies. A la part superior esquerra, hi ha el logotip de la CBP.

Tot i ser limitada, la informació d'un sol genoma pot proporcionar coneixement sobre la història evolutiva d'una espècie, incloent-hi la inferència de potencials esdeveniments que han determinat la seva diversitat genètica actual. Mitjançant l'ús de mètodes basats en coalescència per a analitzar la diversitat genòmica, és possible determinar els canvis històrics de mida efectiva de la població. Fent ús d'aquests mètodes, s'ha inferit que la baldriga balear va patir una forta disminució de la població durant l'últim període interglacial (~120.000 anys) (9). S'esperaria que aquesta forta disminució de la població reduís la diversitat genètica de l'espècie, i en última instància augmentés el risc d'extinció a través de mecanismes genètics, els quals inclouen l'acumulació de mutacions perjudicials, la pèrdua de potencial adaptatiu i la depressió per endogàmia (38,39). No obstant això, els nivells actuals d'heterozigosi al llarg



del genoma de la baldriga balear són aproximadament equivalents als observats en altres ocells marins amb mides de població més grans, suggerint que: (i) aquesta forta disminució demogràfica no ha afectat fortament els nivells actuals d'heterozigosi, (ii) que aquests nivells eren molt alts abans de la disminució, o (iii) que la barreja entre poblacions en períodes interglacials podria haver mantingut alts nivells d'heterozigosi (9). No obstant això, la pressió que les activitats humanes ja estan exercint sobre les poblacions d'aquesta espècie podria conduir eventualment a una pèrdua dràstica de diversitat.

Els genomes de referència també són necessaris per a refinar les anàlisis de genòmica de poblacions, que permeten obtenir coneixements sobre aspectes crítics rellevants per a la conservació. En aquest context, l'anàlisi de reseqüenciació de genomes complets de diferents poblacions de la baldriga balear, juntament amb genomes complets de la baldriga mediterrània (*P. yelkouan*) estretament relacionada (Izquierdo-Arànega, Cuevas-Caballé, *et al.*, en preparació) (Figura 5A), mostra una caracterització detallada de l'estructura poblacional quan es comparen les diferents colònies d'aquests animals, permetent l'assignació genètica d'individus a la seva colònia natal (Figura 5B). Com s'ha esmentat abans, la captura accidental en pesqueries d'art de palangre és l'amenaça principal per a la supervivència de la baldriga balear i una que probablement conduirà l'espècie a l'extinció llevat que les taxes de captura accidental es redueixin significativament (33).



**Figura 5.** a) Mapa de distribució de les principals colònies de *Puffinus mauretanicus* i *P. yelkouan* a tota la Mediterrània occidental. b) Anàlisi de components principals (PCA), que resumeix la variació entre les dades de reseqüenciació de tot el genoma de 32 individus *Puffinus* basat en 1.304.832 SNP autosòmics. c) PCA dels mateixos individus utilitzant només els 61 SNP altament diferenciats. El requadre de la cantonada superior dreta mostra la probabilitat d'assignació a cada població (columnes) de cada individu (cercles) mitjançant el nostre panell d'SNP obtingut amb assignPOP (15). Aquestes xifres només tenen finalitats il·lustratives. La descripció del conjunt de dades i els mètodes estadístics utilitzats es descriuran detalladament en un manuscrit en preparació (Izquierdo-Arànega, Cuevas-Caballé *et al.*, en preparació).

L'assignació de baldrigues balears capturades accidentalment en els arts de pesca a la seva colònia d'origen permetria establir la connexió entre les amenaces al mar i les colònies de cria, aportant informació important per a la planificació d'estratègies de gestió per a mitigar-ne la captura accidental. Amb aquest objectiu s'està desenvolupant un panell d'SNP altament distintius (Figura 5C). Aquest panell facilitarà la determinació de la colònia d'origen de baldrigues capturades de forma accidental, ajudant així a identificar possibles punts crítics de captura accidental. A més, aquesta eina permetrà monitorar la possible disminució de l'heterozigosi al llarg del temps, la qual cosa podria comprometre la capacitat evolutiva de l'espècie.

En general, per a assegurar la persistència d'espècies rares i en perill d'extinció, necessitem més accions de conservació basades en la ciència. Projectes com l'EBP, i els seus afiliats, com la CBP, estan produint genomes de referència a un ritme sense precedents. Aquests projectes milloraran l'ús de recursos genètics destinats a proporcionar informació rellevant per a la conservació, i poden ser essencials per a implementar polítiques de conservació efectives.

## **Diversitat lingüística, cultural i biològica**

Tot i que les raons no s'entenen bé, hi ha una forta evidència de la col·localització geogràfica entre la diversitat biològica i la lingüística (40, 41). No és sorprenent, per tant, que les amenaces tant a les espècies vives com a les llengües tendeixin a produir-se a les mateixes regions geogràfiques. Totes dues podrien ser desencadenades pels mateixos processos socioeconòmics i polítics, que amenacen la integritat i la supervivència de les cultures locals i dels entorns en què viuen (40). L'extinció de les llengües és encara més dramàtica que l'extinció de la biodiversitat, amb la desaparició del 50-90 % de les llengües del món prevista a finals del segle XXI (42). Això tindrà un impacte enorme en la nostra capacitat de traduir el coneixement obtingut dels projectes de genòmica de la biodiversitat en aplicacions en salut humana i en altres àrees. S'ha demostrat que l'extinció d'una llengua contribueix a la pèrdua del coneixement mèdic tradicional perquè els usos de plantes medicinals estan descrits, majoritàriament, en una única llengua (43).

Més enllà de consideracions pràctiques, els noms comuns d'animals i plantes en una llengua determinada sovint reflecteixen la relació dels humans amb la biodiversitat inherent a la cultura més àmplia en què va sorgir la llengua. Anomenar objectes és una de les primeres activitats lingüístiques de qualsevol societat humana, i això també s'aplica als organismes vius. Al mateix temps, els noms perduren més temps que els usos i altres pràctiques, fins i tot quan el coneixement tradicional s'erosiona o es perd.

L'EBP i els projectes afiliats ofereixen una oportunitat única de recollir els noms comuns en els idiomes locals dels exemplars destinats a la seqüenciació en aquests projectes. En aquest sentit, s'han fet diversos esforços als nostres països per a recollir els noms comuns catalans, entre d'altres, d'ocells (44, 45), papallones (46), mamífers marins (47), peixos (48, 49), fongs (50, 51) i plantes, en particular en aquest últim cas, atesa la forta tradició etnobotànica als territoris de parla catalana (52). En el cas de les plantes, per exemple, hi ha 35.000 noms catalans corresponents a 6.500 tàxons (53), extrets de fonts tradicionals i acadèmiques, i actualment el web públic <https://etnobotanica.iec.cat> conté aproximadament 80.000 informes de noms locals i tradicionals de plantes d'uns 1.600 tàxons, que han estat recollits per prop de 2.000 persones. Tot i que la connexió entre la llengua, la identitat cultural i la terra és particularment important en les cultures indígenes (vegeu p. ex. (54)), argumentem que aquest vincle encara es pot rastrejar a les societats postindustrials (en particular, en aquelles en què la diversitat lingüística també està amenaçada). Per això, la CBP ha creat un grup de treball per a coordinar les diferents iniciatives de recopilació dels noms catalans de les espècies i d'elaboració de descripcions estandarditzades del seu significat, que podrien ser utilitzades per altres nodes de l'EBP.

## Discussió

Seqüenciar, catalogar i caracteritzar els genomes de totes les espècies de la Terra conduirà a una comprensió sense precedents de tots els fenòmens de la vida, ja que permetrà identificar els esdeveniments genètics subjacents a l'aparició dels trets biològics dels organismes vius (55). La finalització amb èxit de l'EBP i els projectes associats requereix, però, la participació de socis científics d'arreu del món i s'estén, més enllà de les organitzacions científiques, a la societat en general. Així, creiem que el paper d'estructures, com la CBP, construïda també per a servir d'interfície entre ciència i ciutadania, és crucial per a l'èxit d'aquests projectes. Alguns exemples d'iniciatives similars en àmbits geogràfics molt diferents són el Darwin Tree of Life (2), la Coalició Portuguesa per a la Genòmica de la Biodiversitat (56), el Projecte BioGenome d'Àfrica (57) o el Projecte Genòmica de Conservació de Califòrnia (58), entre d'altres.

De fet, la CBP ha desencadenat una col·laboració local sense precedents entre disciplines científiques diferents dins de la biologia (des de la història natural fins a la genòmica). En el futur, aquesta col·laboració pot ajudar a potenciar les institucions d'història natural amb infraestructures tecnològiques avançades i els recursos humans necessaris per a garantir la catalogació i la preservació d'exemplars, teixits i ADN per a les generacions futures, com ja està passant en moltes institucions d'arreu del món. Dinamitzaria, d'aquesta manera, la investigació taxonòmica en un moment en què es troba en el seu mínim històric (vegeu p. ex. (59)).

La CBP, d'altra banda, busca implicar tota la comunitat apel·lant a un entorn cultural fortament compartit. Per exemple, atesa l'heterogeneïtat i sovint la distància entre els diferents paisatges catalans, la CBP necessitarà la contribució d'una extensa xarxa d'associacions naturalistes familiaritzades amb els ecosistemes locals per a facilitar l'accés a les mostres biològiques. A part d'un esforç científic, concebem la CBP com a part d'un moviment transformador mundial que incrementa la consciència social sobre l'amenaça que suposa la pèrdua de biodiversitat per al benestar humà, i que implica globalment la societat cap a una relació diferent i més equilibrada amb la natura.

Més enllà de les fronteres dels territoris de parla catalana, i dins de la xarxa de xarxes EBP, la CBP pretén tenir un paper central en els projectes genòmics de biodiversitat a Europa, organitzats actualment sota ERGA, i els de la península Ibèrica i la conca mediterrània, les regions geogràfiques a les quals pertanyen els territoris catalans. Per tant, la CBP manté forts vincles amb ERGA-Espanya, ERGA-Andorra i ERGA-França. No obstant això, la CBP està oberta a la col·laboració i a compartir recursos i experiència amb socis d'arreu del món. Després de la fase pilot del projecte, la CBP té garantit un finançament continuat per iniciar la primera fase de producció durant el trienni 2023-2025. Contribuint tant amb dades com amb recursos, la CBP pretén implicar científics i ciutadans en el que potser és un dels projectes globals més importants de la història de la biologia.

## Materials i mètodes

Per a explorar els patrons de l'estructura poblacional en les baldrigues mediterrànies, vam utilitzar dades de reseqüenciament de genomes complets de 36 individus de baldriga balear ( $n = 30$ ) i baldriga mediterrània ( $n = 6$ ) (Izquierdo-Aránega, Cuevas-Caballé, *et al.*, en preparació). Les seqüències es van mapar al genoma de referència de *P. mauretanicus* (9) utilitzant BWA mem v0.7.17 (10) amb els paràmetres per defecte i es va realitzar la detecció

de variants genètiques combinant HaplotypeCaller de GATK4 v4.1.9 (11) i Freebayes v1.2 (12). Posteriorment, vam filtrar el conjunt de polimorfismes d'un únic nucleòtid (SNP) resultant amb VCFtools v.0.1.15 (13) utilitzant llindars de qualitat (--minQ 30), cobertura (--minDP 5 i --maxDP 200) i absència (--max-missing 0,75). Vam utilitzar el fitxer VCF resultant per a realitzar una anàlisi de components principals (PCA) implementada en PLINK v1.90 (14).

Per a construir un conjunt adient d'SNP per a l'assignació de baldrigues a la seva colònia d'origen, vam seleccionar els SNP més diferenciats a partir de comparacions d' $F_{ST}$  per parelles entre colònies utilitzant VCFtools. Vam seleccionar un conjunt final de 61 SNP, agafant els més diferenciats dins de cada comparació per parelles, assegurant-nos que només incloguéssim cada SNP una vegada. Per a validar el nostre panell d'SNP, vam realitzar una PCA implementada en PLINK i vam avaluar l'assignació genòmica de cada individu a les seves colònies d'origen mostrejades utilitzant Assignpop v1.2.4 (15) amb validació creuada K-fold utilitzant el model LDA.

## Disponibilitat de les dades

No hi ha dades associades a aquest manuscrit. Totes les dades produïdes per la CBP es poden consultar a través del portal de la CBP a <https://dades.biogenoma.cat>.

## Dades suplementàries

Les dades suplementàries estan disponibles a *NAR Genomics and Bioinformatics* en línia.

## Agraïments

Volem agrair especialment a Harris Lewin i a la direcció de l'EBP, i a Mark Blaxter del Darwin Tree of Life, el seu suport continu i incondicional. També volem donar les gràcies al Consell Directiu de la Societat Catalana de Biologia (SCB) pel seu compromís amb la iniciativa i a Carlos Camillero-Robles per la seva ajuda amb les figures en català. La Figura 4 compta amb les contribucions originals de Mercè Rocabosch, a qui agraiem el seu generós esforç. Finalment, voldríem posar de manifest que la idea del projecte va ser suggerida per l'Antonio Martínez Márquez durant la conversa d'una nit d'estiu.

## Finançament

La CBP agraeix el finançament del Departament de Recerca i Universitats de la Generalitat de Catalunya a través de l'Institut d'Estudis Catalans (IEC). La CBP està parcialment finançada per l'IEC, el Zoo de Barcelona, el Departament d'Acció Climàtica, Alimentació i Agenda Rural de la Generalitat de Catalunya, l'Euroregió Pirineus Mediterrània i la Universitat de Barcelona (UB). El suport institucional al Centre Nacional d'Anàlisi Genòmica (CNAG) i al Centre de Regulació Genòmica (CRG) prové del Ministeri de Ciència i Innovació d'Espanya a través de l'Institut de Salut Carlos III i de la Generalitat de Catalunya a través del Departament de Salut i el Departament de Recerca i Universitats.

## Declaració de conflicte d'interessos

## Referències

1. Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M.T.P., *et al.* (2018) Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 4325-4333.
2. Blaxter, M. L. (2022) Sequence locally, think globally: the Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2115642118.
3. Lewin, H. A., Richards, S., Aiden, E. L., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., *et al.* (2022) The Earth BioGenome Project 2020: starting the clock. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2115635118.
4. Casas-Sainz, A. M. and de Vicente, G. (2009) On the tectonic origin of Iberian topography. *Tectonophysics*, **474**, 214-235.
5. Tuel, A. i Eltahir, E. A. B. (2020) Why is the Mediterranean a climate change hot spot? *J. Clim.*, **33**, 5829-5843.
6. Hoegh-Guldberg, O., Jacob, D., Taylor, M., Guillén Bolaños, T., Bindi, M., Brown, S., Camilloni, I. A., Diedhiou, A., Djalante, R., Ebi, K., *et al.* (2019) The human imperative of stabilizing global climate change at 1.5°C. *Science*, **365**, eaaw6974.
7. Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.P., Iglesias, A., Lange, M. A., Lionello, P., Llasat, M. C., Paz, S., *et al.* (2018) Climate change and interconnected risks to sustainable development in the Mediterranean. *Nat. Clim. Chang.*, **8**, 972-80.
8. Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X. D., *et al.* (2015) Elevation-dependent warming in mountain regions of the world. *Nat. Clim. Chang.*, **5**, 424-430.
9. Cuevas-Caballé, C., Ferrer Obiol, J., Vizueta, J., Genovart, M., Gonzalez-Solís, J., Riutort, M. i Rozas, J. (2022) The first genome of the balearic shearwater (*Puffinus mauretanicus*) provides a valuable resource for conservation genomics and sheds light on adaptation to a pelagic lifestyle. *Genome Biol. Evol.*, **14**, evac067.
10. Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Doi: <https://arxiv.org/abs/1303.3997>, 26 de maig de 2013, preimpresió: no revisat.
11. Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., *et al.* (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv doi: <https://doi.org/10.1101/201178>, 24 de juliol de 2018, preimpresió: no revisat.
12. Garrison, E. i Marth, G. (2012) Haplotype-based variant detection from short-read sequencing — free bayes — variant calling — longranger. arXiv doi: <https://arxiv.org/abs/1207.3907>, 20 de juliol de 2012, preimpresió: no revisat.
13. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.
14. Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M. i Lee, J. J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
15. Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L. and Ludsins, S. A. (2018) assignPOP: an r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods Ecol. Evol.*, **9**, 439-446.
16. Casassas i Simó, O., Camarasa i Castillo, J. M. i Junyent i Rodríguez, C. (2012) Cent anys de la Societat Catalana de Biologia, la primera societat filial de l'Institut d'Estudis Catalans. *Treballs De La Societat Catalana De Biologia*, **63**, 299-324.
17. Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., *et al.* (2022) The era of reference genomes in conservation genomics. *Trends Ecol. Evol.*, **37**, 197-202.
18. Mazzoni, C. J., Ciofi, C. i Waterhouse, R. M. (2023) Biodiversity: an atlas of European reference genomes. *Nature*, **619**, 252.
19. Folch i Guillén, R. (2010) A: *Història natural dels Països Catalans Catalana*. Edició: Barcelona: Enciclopèdia Catalana, 1985-2010.
20. Böhne, A., Fernández, R., Leonard, J. A., McCartney, A. M., McTaggart, S., Melo-Ferreira, J., Monteiro, R., Oomen, R. A., Pettersson, O. V. i Struck, T. H. (2023) Contextualising samples:

- supporting reference genomes for European biodiversity through sample and associated metadata collection. bioRxiv doi: <https://doi.org/10.1101/2023.06.28.546652>, 30 de juny de 2023, preimpressió: no revisat.
21. Lawniczak, M. K. N., Durbin, R., Flicek, P., Lindblad-Toh, K., Wei, X., Archibald, J. M., Baker, W. J., Belov, K., Blaxter, M. L., Bonet, T. M., *et al.* (2022) Standards recommendations for the Earth BioGenome Project. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2115639118.
  22. Santesmasses, D., Mariotti, M. i Guigó, R. (2018) Selenoprofiles: a computational pipeline for annotation of selenoproteins. A: *Methods in Molecular Biology*. Vol. 1661.
  23. Vlasova, A., Pulido, T. H., Camara, F., Ponomarenko, J. i Guigó, R. (2021) FA-nf: a functional annotation pipeline for proteins from non-model organisms implemented in nextflow. *Genes (Basel)*, **12**, 1645.
  24. Fuentes, D., Molina, M., Chorostecki, U., Capella-Gutiérrez, S., Marcet-Houben, M. i Gabaldón, T. (2022) PhylomeDB V5: an expanding repository for genome-wide catalogues of annotated gene phylogenies. *Nucleic Acids Res.*, **50**, D1062-D1068.
  25. DI Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E. i Notredame, C. (2017) Nextflow enables reproducible computational workflows. *Nat. Biotechnol.*, **35**, 316-319.
  26. Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., *et al.* (2021) Sustainable data analysis with Snakemake [versió 2; avaluació d'experts: 2 aprovada]. *F1000Res*, **10**, 33.
  27. Ledoux, J. B., Cruz, F., Gómez-Garrido, J., Antoni, R., Blanc, J., Gómez-Gras, D., Kipson, S., López-Sendino, P., Antunes, A., Linares, C., *et al.* (2020) The genome sequence of the octocoral *Paramuricea clavata* - a key resource to study the impact of climate change in the mediterranean. *G3: Genes, Genomes, Genetics*, **10**, 2941-2952.
  28. Cruz, F., Gómez-Garrido, J., Gut, M., Alioto, T. S., Pons, J., Alós, J. i Barcelo-Serra, M. (2023) Chromosome-level assembly and annotation of the *Xyrichtys novacula* (Linnaeus, 1758) genome. *DNA Res.*, **30**, dsad021.
  29. Gomez-Garrido, J., Cruz, F., Alioto, T. S., Feiner, N., Uller, T., Gut, M., Sánchez Escudero, I., Tavecchia, G., Rotger, A., Otalora Acevedo, K. E., *et al.* (2023) Chromosome-level genome assembly of Lilford's wall lizard, *Podarcis lilfordi* (Günther, 1874) from the Balearic Islands (Spain). *DNA Res.*, **30**, dsad008.
  30. Marlétaz, F., Firbas, P. N., Maeso, I., Tena, J. J., Bogdanovic, O., Perry, M., Wyatt, C. D. R., de la Calle-Mustienes, E., Bertrand, S., Burguera, D., *et al.* (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature*, **564**, 64-70.
  31. Torruella, G., Galindo, L. J., Moreira, D., Ciobanu, M., Heiss, A. A., Yubuki, N., Kim, E. i López-García, P. (2023) Expanding the molecular and morphological diversity of Apusomonadida, a deep-branching group of gliding bacterivorous protists. *J. Eukaryot. Microbiol.*, **70**, e12956.
  32. Righi, E. i Guigó, R. (2023) The BioGenome Portal: a web-based platform for biodiversity genomics data management. bioRxiv doi: <https://doi.org/10.1101/2023.12.20.572408>, 22 de desembre de 2023, preimpressió: no avaluació d'experts.
  33. Genovart, M., Arcos, J. M., Álvarez, D., McMin, M., Meier, R., Wynn, B., R., G. i Oro, D. (2016) Demography of the critically endangered Balearic shearwater: the impact of fisheries and time to extinction. *J. Appl. Ecol.*, **53**, 1158-1168.
  34. Arcos, J. M. (compilador) (2011) International species action plan for the Balearic shearwater, *Puffinus mauretanicus*. A: *SEO/BirdLife & BirdLife International*.
  35. Cortés, V. i González-Solís, J. (2018) Seabird bycatch mitigation trials in artisanal demersal longliners of the Western Mediterranean. *PLoS One*, **13**, e0196731.
  36. Rodríguez, A., Arcos, J. M., Bretagnolle, V., Dias, M. P., Holmes, N. D., Louzao, M., Provencher, J., Raine, A. F., Ramírez, F., Rodríguez, B., *et al.* (2019) Future directions in conservation research on petrels and shearwaters. *Front. Mar. Sci.*, **6**, 00094.
  37. Louzao, M., Delord, K., García, D., Boué, A. i Weimerskirch, H. (2012) Protecting persistent dynamic oceanographic features: transboundary conservation efforts are needed for the critically endangered balearic shearwater. *PLoS One*, **7**, e35728.
  38. Frankham, R. (2005) Genetics and extinction. *Biol. Conserv.*, **126**, 131-140.
  39. Charlesworth, D. i Willis, J. H. (2009) The genetics of inbreeding depression. *Nat. Rev. Genet.*, **10**, 783-796.
  40. Maffi, L. (2005) Linguistic, cultural and biological diversity. *Annu. Rev. Anthropol.*, **34**, 599-617.
  41. Gorenflo, L. J., Romaine, S., Mittermeier, R. A. i Walker-Painemilla, K. (2012) Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8032-8037.

42. Nettle, D. i Romaine, S. (2000) A: *Vanishing voices: the extinction of the world's languages*. Oxford University Press.
43. Cámara-Leret, R. i Bascompte, J. (2021) Language extinction triggers the loss of unique medicinal knowledge. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2103683118.
44. Ortega i Gonzàlez, E. i (2017) A: *Diccionari etimològic dels noms científics dels ocells dels Països Catalans*. 1a ed. Institut d'Estudis Catalans, Barcelona.
45. Aguiló, C. i Mestre, A. (2017) *Atles ornitològic de les Illes Balears*. Barcelona/Palma, Institut d'Estudis Catalans / Institut d'Estudis Balearics.
46. Vila, R., Stefanescu, C. i Sesma, J. M. (2018) *Guia de les papallones diürnes de Catalunya*. Lynx Edicions, Bellaterra.
47. Leonart, J. (2012) *Els mamífers marins i els seus noms*. *Terminàlia*, **5**, 7-25.
48. Alegre i Urgell, M. (1992) A: *Espècies pesqueres d'interès comercial: Nomenclatura oficial catalana*. Generalitat de Catalunya. Departament de Cultura, Barcelona.
49. Mercader, L., Lloris, D. i Rucabado, J. (2001) A: *Tots els peixos del Mar Català: Diagnòs i claus d'identificació*. Institut d'Estudis Catalans, Barcelona.
50. Niell, M. i Girbal, J. M. (2006) *Els noms populars dels bolets a Andorra*. *Revista Catalana de Micologia*, **28**, 209-216.
51. Cuello, J. (2007) A: *Els noms dels bolets*. Lynx Edicions, Bellaterra.
52. Gras, A., Garnatje, T., Marín, J., Parada, M., Sala, E., Talavera, M. i Vallès, J. (2020) The power of wild plants in feeding humanity: a meta-analytic ethnobotanical approach in the Catalan linguistic area. *Foods*, **10**, 61.
53. Vallès, J., Veny, J., Vigo, J., Bonet, M. À., Julià, M. A. i Villalonga, J. C. (2014) A: *Noms de plantes. Corpus de fitonomia catalana*. Termcat - Centre de Terminologia i Universitat de Barcelona. Barcelona.
54. Blythe, J. i Brown, R. M. (2003) Maintaining the links: Language, identity and the land. A: *Proceedings of the 7th FEL Conference, Broom, Western Australia, 22-24 September 2003*. Foundation for Endangered Languages, Bath, Regne Unit.
55. Guigó, R. (2023) Genome annotation: from human genetics to biodiversity genomics. *Cell Genomics*, **3**, 100375.
56. Marques, J. P., Alves, P. C., Amorim, I. R., Lopes, R. J., Moura, M., Myers, E. W., Sim-Sim, M., Sousa-Santos, C., Judite Alves, M. V., Borges, P. A., et al. (2023) Building a Portuguese Coalition for Biodiversity Genomics. *EcoEvoRxiv*, ver 2.
57. Ebenezer, T. G. E., Muigai, A. W. T., Nouala, S., Badaoui, B., Blaxter, M., Buddie, A. G., Jarvis, E. D., Korch, J., Kujala, J. O., Lewin, H. A., et al. (2022) Africa: sequence 100,000 species to safeguard biodiversity. *Nature*, **603**, 388-392.
58. Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., et al. (2022) Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *J. Hered.*, **113**, 577-588.
59. Crisci, J. V., Katinas, L., Apodaca, M. J. i Hoch, P. C. (2020) The end of botany. *Trends Plant Sci.*, **25**, 1173-1176.

## Apèndix

Altres membres de la CBP que han participat en el projecte i no es mencionen a la llista d'autors: Laura Aguilera, Francisco Cámara, Jèssica Gómez-Garrido, Fernando Cruz, Ignacio Sánchez-Escudero (Centre Nacional d'Anàlisi Genòmica, Universitat de Barcelona, 08028 Barcelona, Catalunya, Espanya); Cristian R. Altaba (Conselleria d'Agricultura, Pesca i Medi Natural, Govern de les Illes Balears; Research Group on Human Evolution and Cognition, Universitat de les Illes Balears, 07122 Palma, Illes Balears, Espanya); Rui Alves (Departament de Ciències Mèdiques Bàsiques, Universitat de Lleida, 25002 Lleida, Catalunya, Espanya); Enrique Arboleda, Maria Capa, Laura Triginer (Centre Balear de Biodiversitat, Universitat de les Illes Balears, 07122 Palma, Illes Balears, Espanya); Vicent Arbona (Departament de Biologia, Bioquímica i Ciències Naturals, Universitat Jaume I, 12071 Castelló, País Valencià, Espanya); Conxita Avila, Laura Baldo, Alba Enguidanos, Andrea Prófumo, Owen S. Wangesteen (Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia; Institut de Recerca de la Biodiversitat, Universitat de Barcelona, 08028 Barcelona,

Catalunya, Espanya); Margarida Barceló-Serra (Institut Mediterrani d'Estudis Avançats, CSIC-UIB, 07190 Esporles, Illes Balears, Espanya); Laura Botigué, Amparo Monfort (Centre de Recerca en Agrigenòmica, CSIC-IRTA-UAB-UB, Campus UAB, 08193 Bellaterra, Catalunya, Espanya); Bernat Burriel-Carranza; Salvador Carranza, Javier del Campo, Gabriel Mochales-Riaño, Marc Palmada-Flores (Institut de Biologia Evolutiva, IBE, UPF-CSIC, PRBB, 08003 Barcelona, Catalunya, Espanya); Héctor Candela (Instituto de Bioingeniería, Universitat Miguel Hernández d'Elx, 03202, Elx, Espanya); Cristian Cañestro, Carlos Carreras, Marc Domènech, Carles Galià-Camps, Sara Guirao-Rico, Jesús Lozano-Fernández, Marta Pascual, Cinta Pegueroles, Alejandro Sánchez-Gracia (Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Institut de Recerca de la Biodiversitat, Universitat de Barcelona, 08028 Barcelona, Catalunya, Espanya); Miguel A. Carretero (CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universitat de Porto, Campus de Vairão. 4485-661 Vairão, Portugal); Oriol Grau (Observatori de Recerca del Parc Natural de l'Alt Pirineu, CSIC, Unitat d'Ecologia Global CREA-CSIC-UAB, 08913 Bellaterra, Catalunya, Espanya); Romina Garrido, Xavier Grau-Bové, Toni Hermoso, Iana Kim (Centre de Regulació Genòmica, Institut de Ciència i Tecnologia de Barcelona (BIST), 08003 Barcelona, Catalunya, Espanya); José A. Jurado-Rivera (Departament de Biologia, Genètica, Universitat de les Illes Balears, 07122 Palma de Mallorca, Illes Balears, Espanya); Carles Lalueza-Fox (Natural Museu de Ciències Naturals de Barcelona, 08019, Barcelona, Catalunya, Espanya); Jean-Baptiste Ledoux (CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universitat de Porto, 4099-002 Porto, Portugal); Ramiro Logares, Ramon Massana, Francesc Piferrer (Institut de Ciències del Mar, CSIC, 08003, Barcelona, Catalunya, Espanya); Juli Mauri (Fundació Barcelona Zoo, 08013, Barcelona, Catalunya, Espanya); Francesc Mesquita-Joanes, Juan S. Monrós (Institut Cavanilles de Biodiversitat i Biologia Evolutiva, ICBIBE, Universitat de València, 46980 Paterna, València, Espanya); Jaume Pellicer (Institut Botànic de Barcelona, CSIC-CMCNB, 08038 Barcelona, Catalunya, Espanya; Royal Botanic Gardens, Kew, Richmond TW9 3DS, Regne Unit); Josep Peñuelas (Unitat d'Ecologia Global CREA-CSIC-UAB, Campus UAB, 08913 Bellaterra, Catalunya, Espanya); Cristina Roquet (Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalunya, Espanya); Aurora Ruiz-Herrera (Departament de Biologia Cel·lular, Fisiologia i Immunologia, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08913 Bellaterra, Catalunya, Espanya); Guifré Torruella (Barcelona Supercomputing Center - Centre Nacional de Supercomputació (BSC-CNS), 08034 Barcelona, Catalunya, Espanya); Xavier Turon, Marc Ventura (Centre d'Estudis Avançats de Blanes, CEAB-CSIC, 17300 Blanes, Girona, Catalunya, Espanya).



## Figura suplementària S1



**Figura S1.** Activitats de difusió de la CBP. a) Banderoles al carrer amb informació sobre les activitats de la Iniciativa a la ciutat de Barcelona. Una mostra aleatòria d'aquestes activitats inclou: b) participació en la Jornada de Portes Obertes del PRBB, c) xerrada dins del cicle *Visions de la Ciència*, organitzat per les «Biblioteques de Barcelona», d) taller d'extracció d'ADN al Museu de Ciències Naturals de Granollers, e) participació en la «Fira Recerca en Directe».