The Earth BioGenome Project: Convergence Science at the Nexus of the Global Biodiversity and Climate Crises

Harris A. Lewin
University of California, Davis

www.earthbiogenome.org
Twitter: @EBPgenome

Our understanding of life is severely constrained

- Estimated 12-15 million eukaryotic species on Earth
- Only 1.8 million have been named

Much of eukaryotic life is under threat

More than 38,500 species
are threatened with extinction

That is still 28% of all assessed species.

IUCN 2021

"The sixth great extinction spasm of geological time is upon us."

E.O. Wilson, *The Diversity of Life*

Number of plant/animal species threatened with extinction

1,000,000

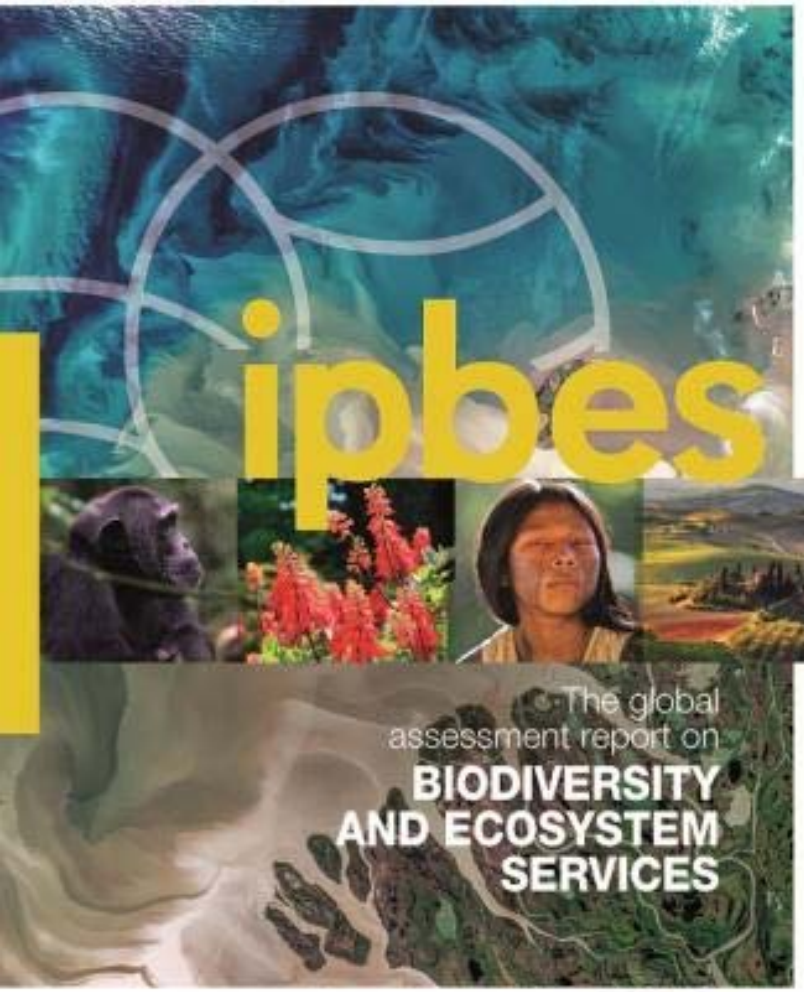
Connecting Global Priorities: Biodiversity and Human Health

A State of Knowledge Review

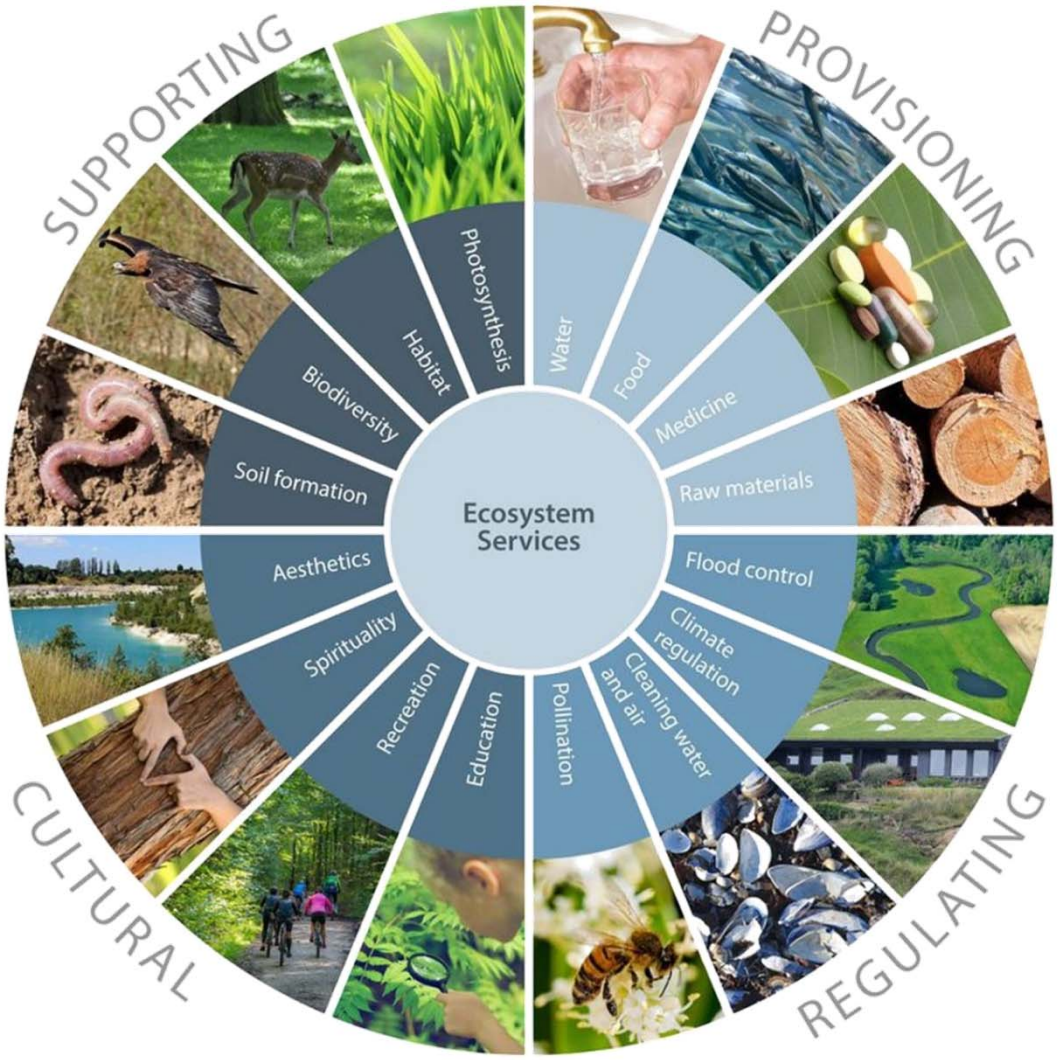
UNEP

Convention on Biological Diversity

World Health Organization


ipbes
The global assessment report on
BIODIVERSITY AND ECOSYSTEM SERVICES
ipbes

Biodiversity provides the basis for ecosystem services that support life on our planet



UN BIODIVERSITY CONFERENCE

11 - 15 October 2021
VIRTUAL

25 April - 8 May 2022
IN PERSON

2020 UN BIODIVERSITY CONFERENCE

COP 15 - CP/MOP10 - NP/MOP4

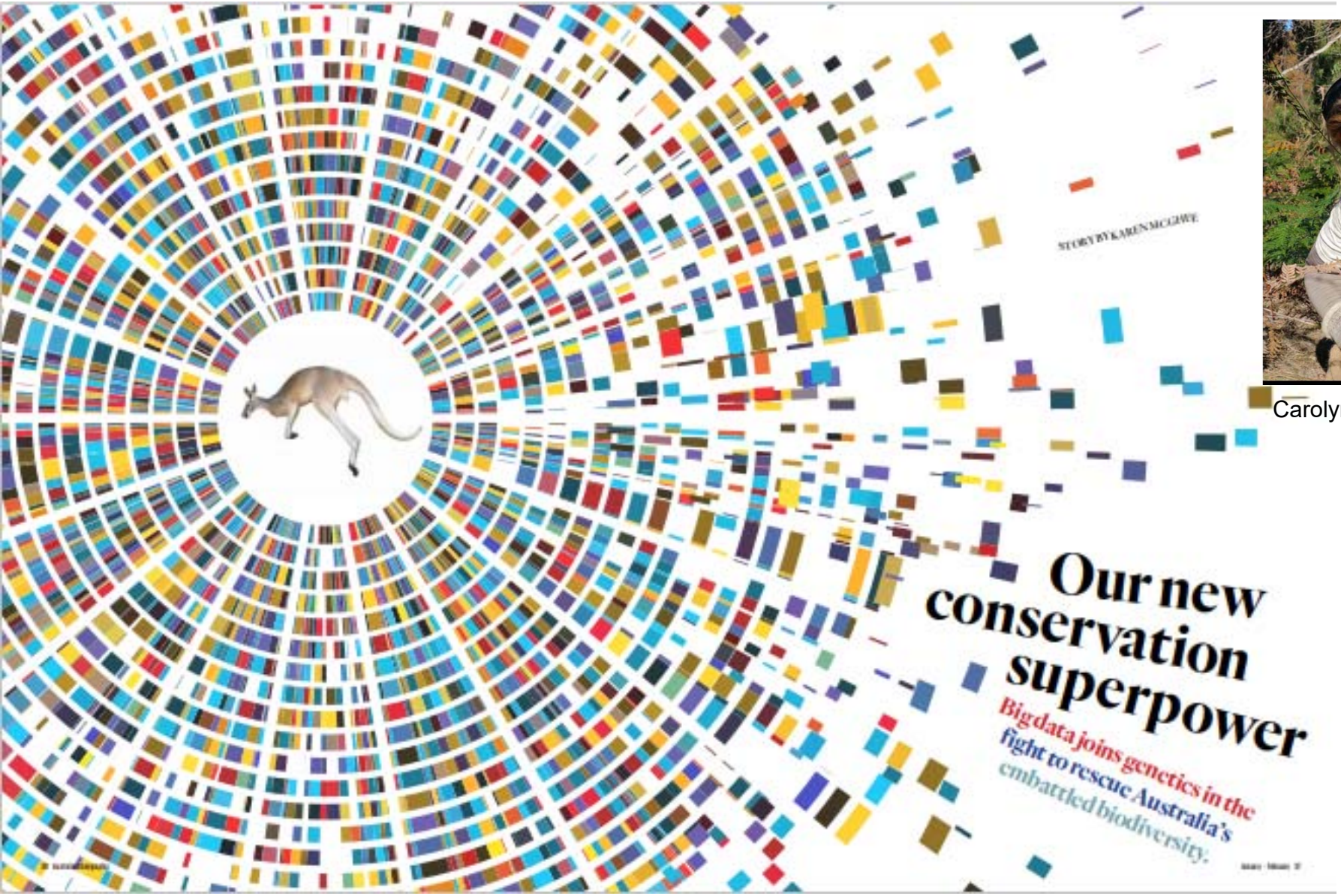
Ecological Civilization-Building a Shared Future for All Life on Earth

KUNMING · CHINA

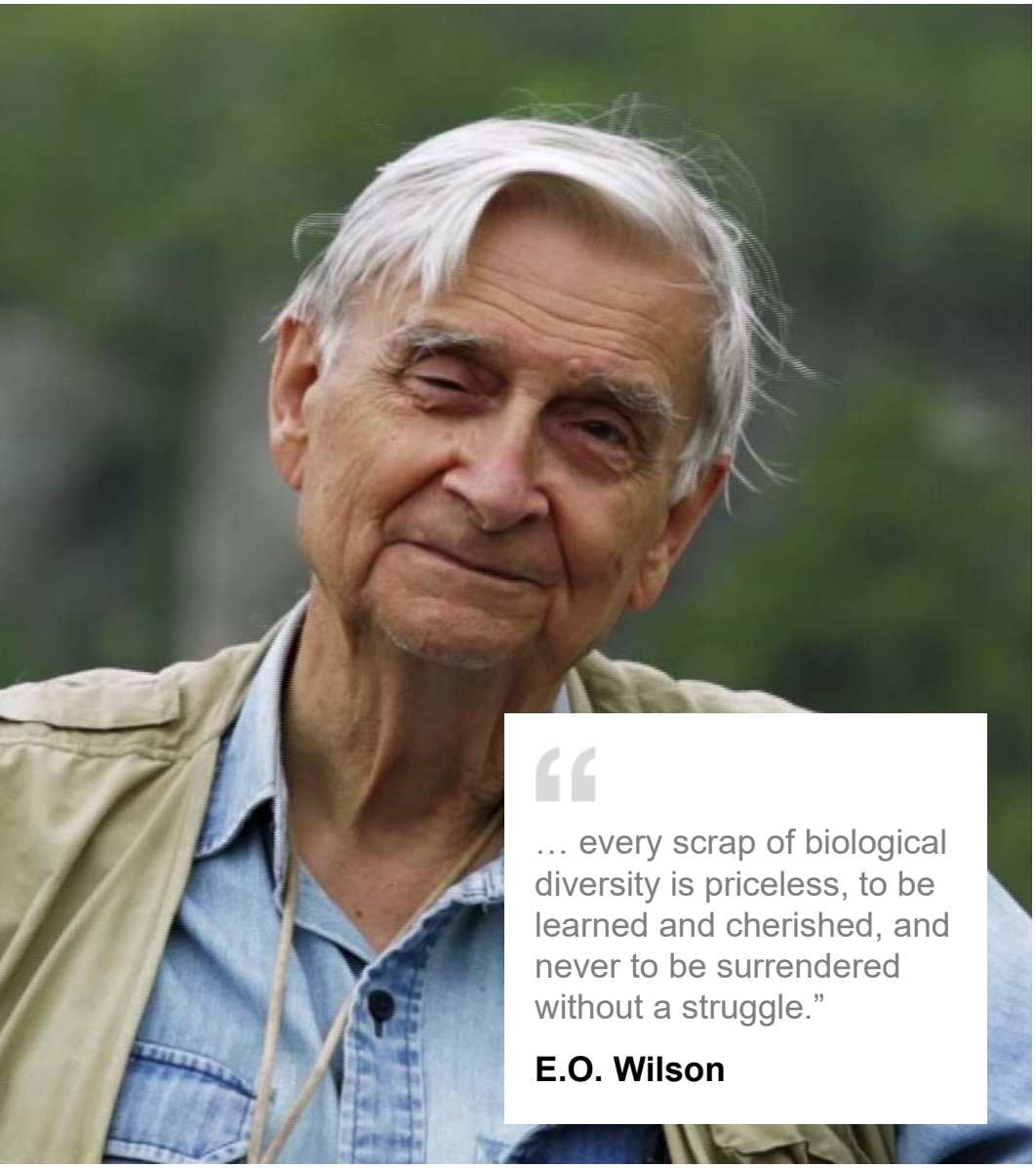
Convention on Biological Diversity

www.cbd.int/meetings/COP-15

#COP15



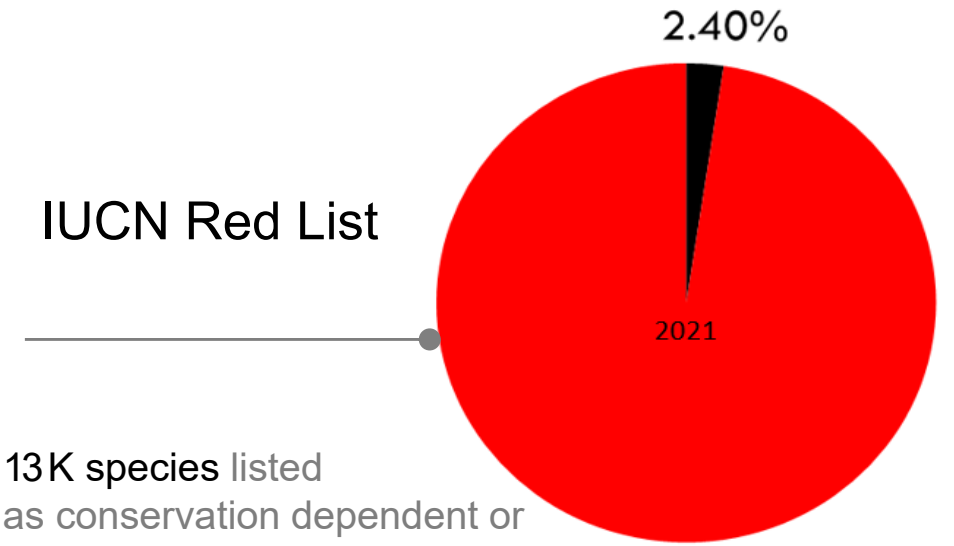
Carolyn Hogg, U. Sydney



> … every scrap of biological diversity is priceless, to be learned and cherished, and never to be surrendered without a struggle.”
>
> **E.O. Wilson**

Endangered species are without genomic tools

IUCN Red List



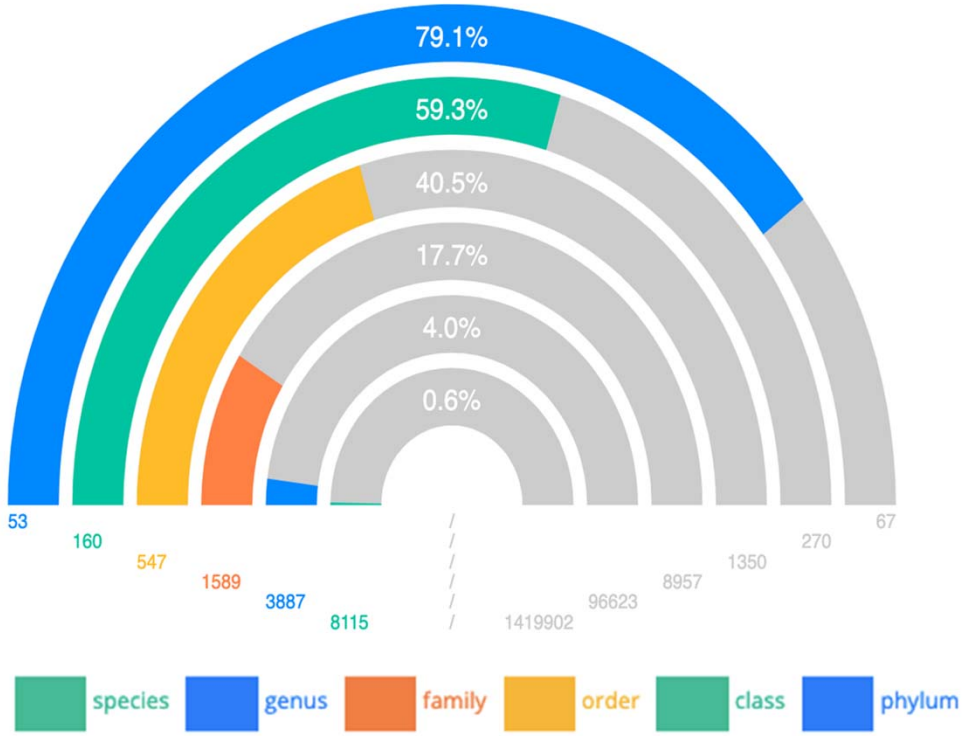
13K species listed as conservation dependent or worse

Hogg et al., *PNAS* in press

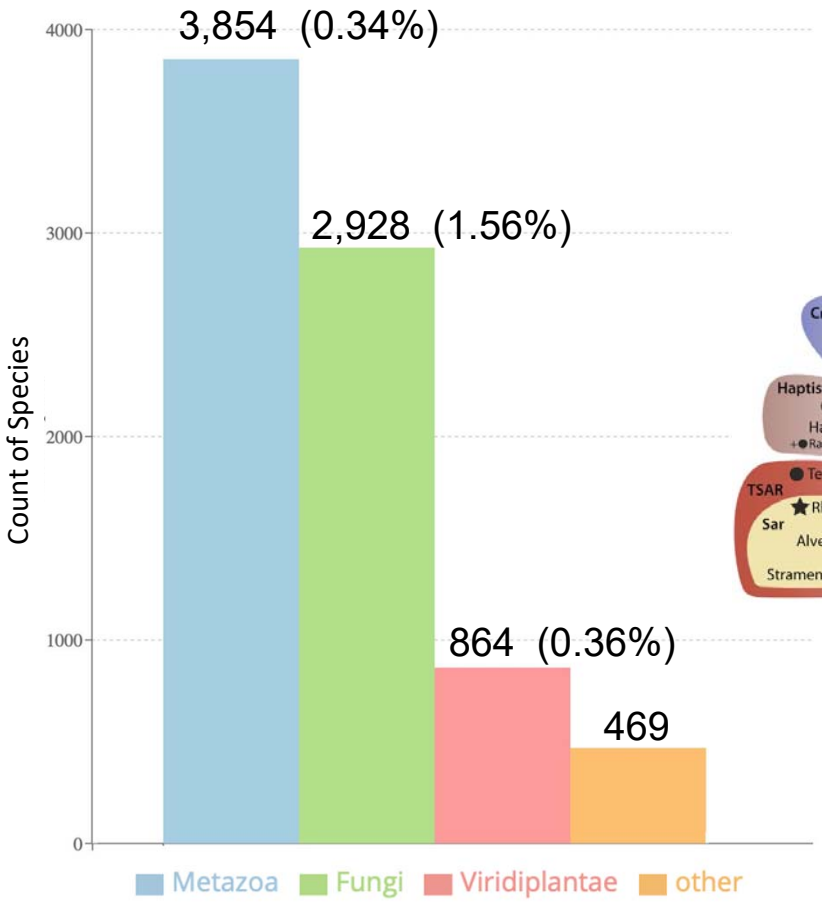
Global Status of Eukaryotic Genome Sequencing

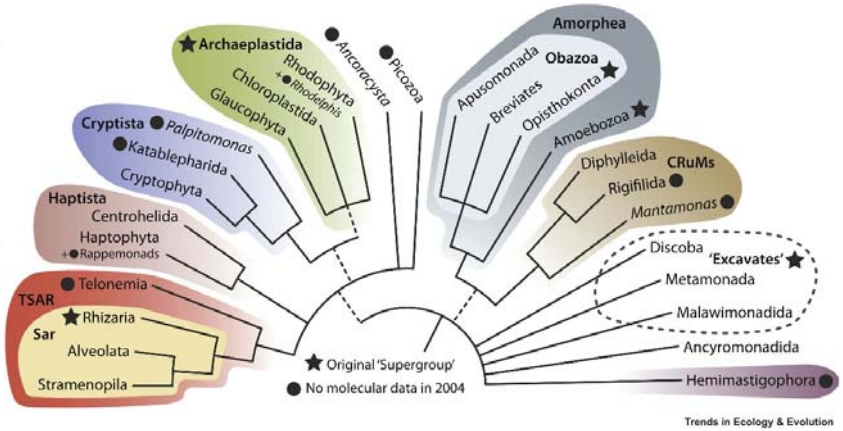
8,115 distinct eukaryotic taxa as of September 21, 2021

Only **0.6%** of known eukaryotic species have been sequenced



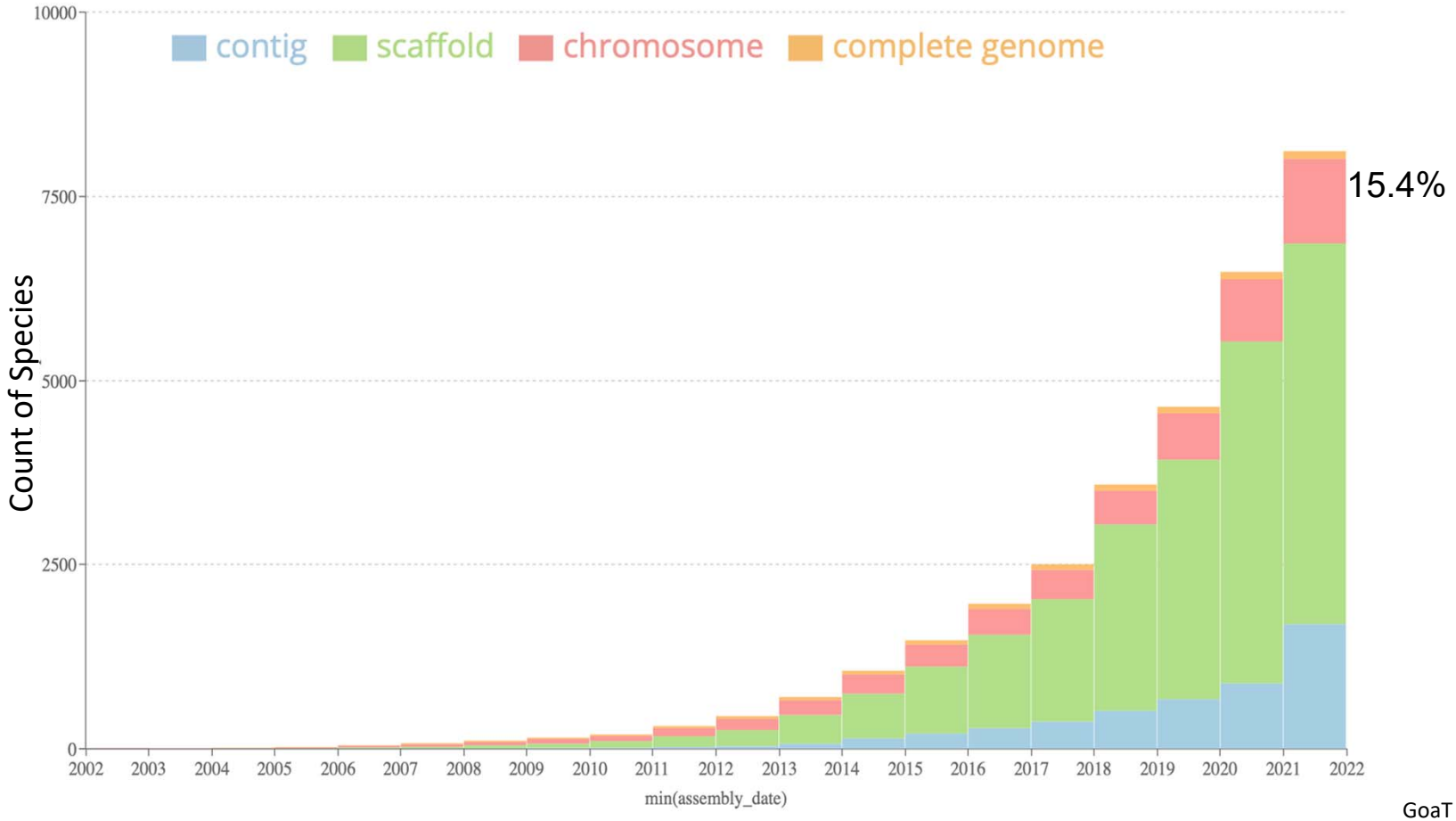
Sequencing Across Eukaryotic Kingdoms





GoaT data [release 2021.09.21]

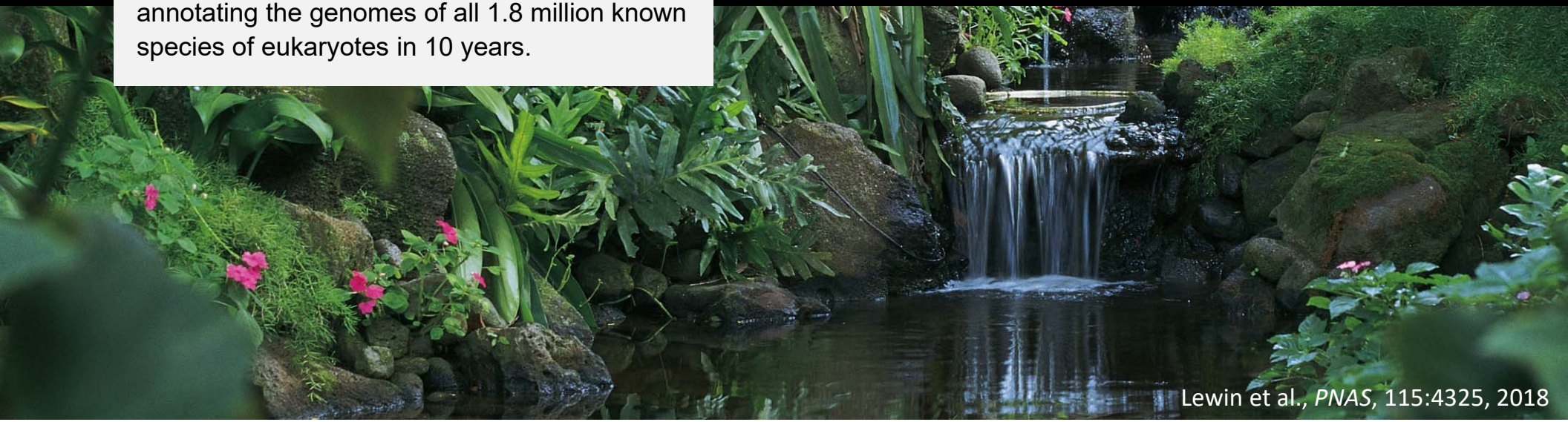
Whole Genome Assemblies of Eukaryotic Species Over Time



GoaT data release 2021.09.21

What is the Earth BioGenome Project?

The Earth BioGenome Project is a confederated international network-of-networks that has the common goal of sequencing and annotating the genomes of all 1.8 million known species of eukaryotes in 10 years.



Lewin et al., *PNAS*, 115:4325, 2018

EBP…More than a Moonshot

The EBP strives to be global model for large-scale scientific collaborations, providing equal opportunity and benefits for all people, everywhere

The Earth BioGenome Project:

A Confederated International **Network-of-Networks**

43
institutions members

49
Affiliated Projects

+5000
Participants

22
countries

Global hub and spokes model

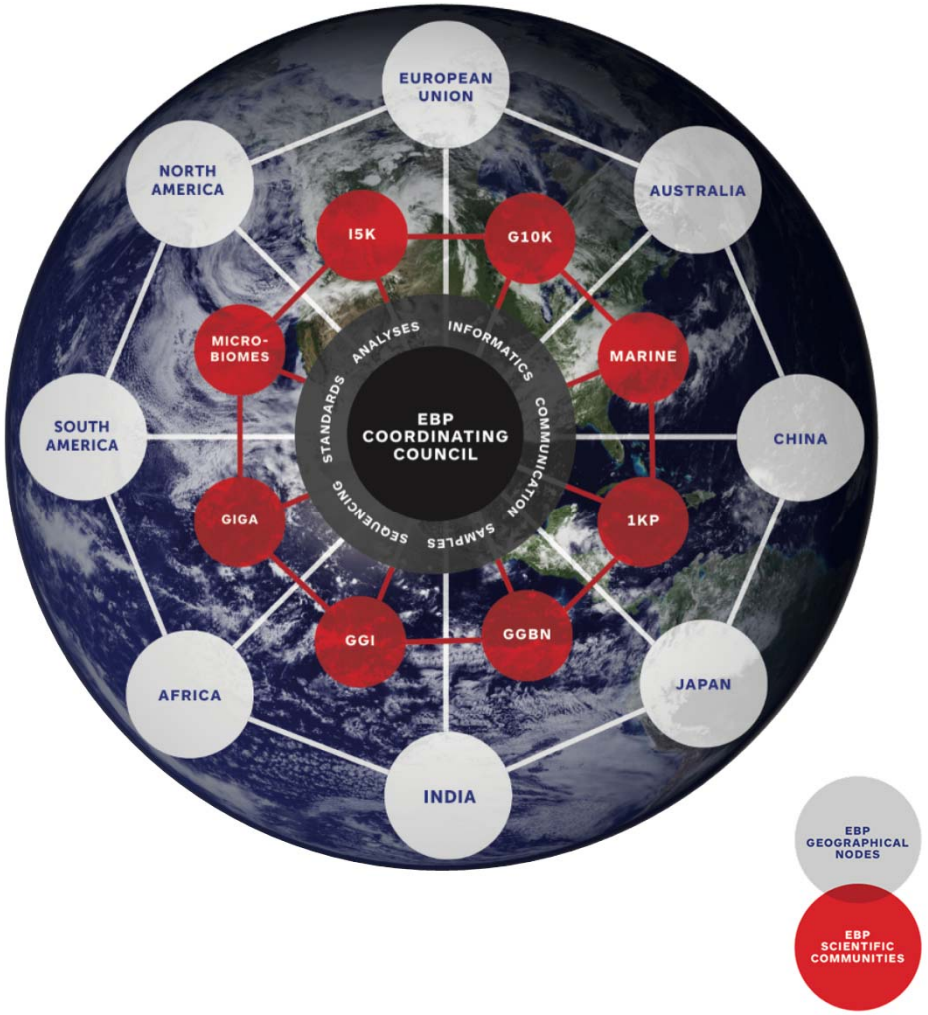
The Earth BioGenome Project:

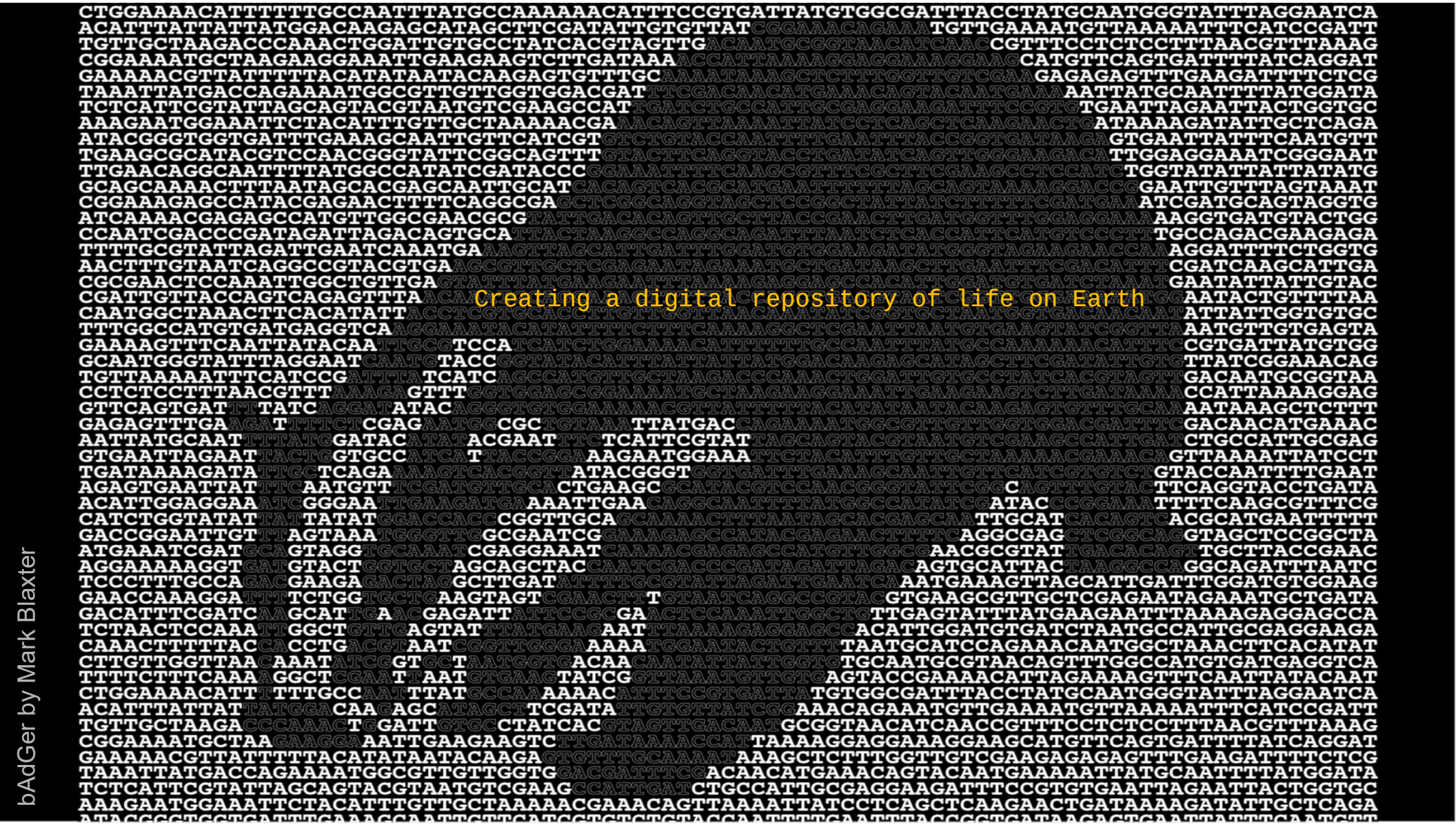
A Confederated International **Network-of-Networks**

Committed to open data access and compliance with the Convention on Biological Diversity and the Nagoya Protocol on Access and Benefit Sharing (ABS)

Committed to diversity, equity, inclusion and justice among the EBP community of scientists and peoples of the world.

Model of EBP Network-of-Networks





bAdGer by Mark Blaxter

Why Sequence all Eukaryotes?

Blaxter et al., *PNAS* in press

Building genomics-informed conservation

Understanding ecosystem function, stasis, and change

Revealing the deep logic of eukaryotic gene regulation

Tracking genomic changes in symbiosis

Decoding the genomics of complex traits

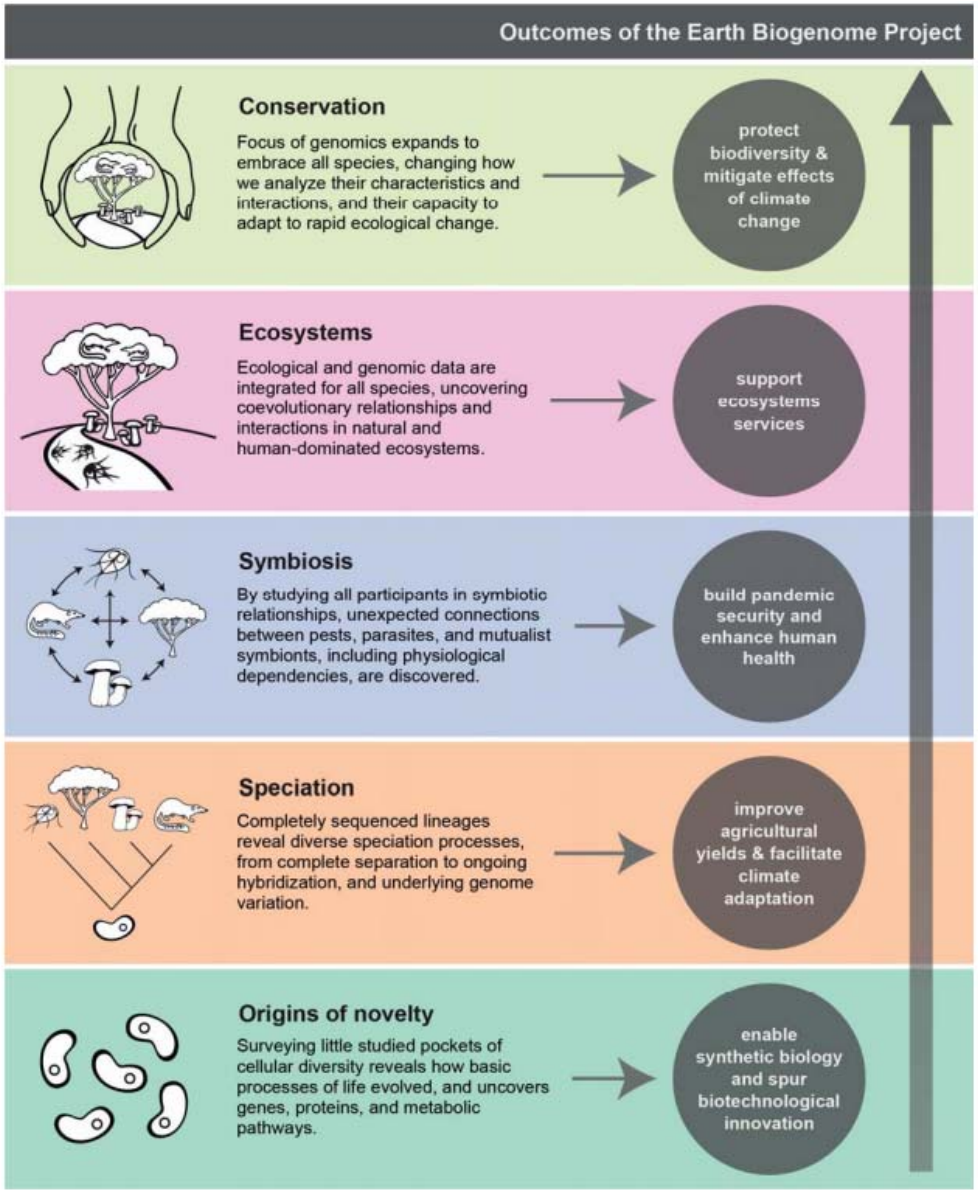
Probing the diversity of sexual systems

Decrypting chromosome evolution

Exploring diversity in the genomics of speciation

Defining the origin of eukaryotic cells

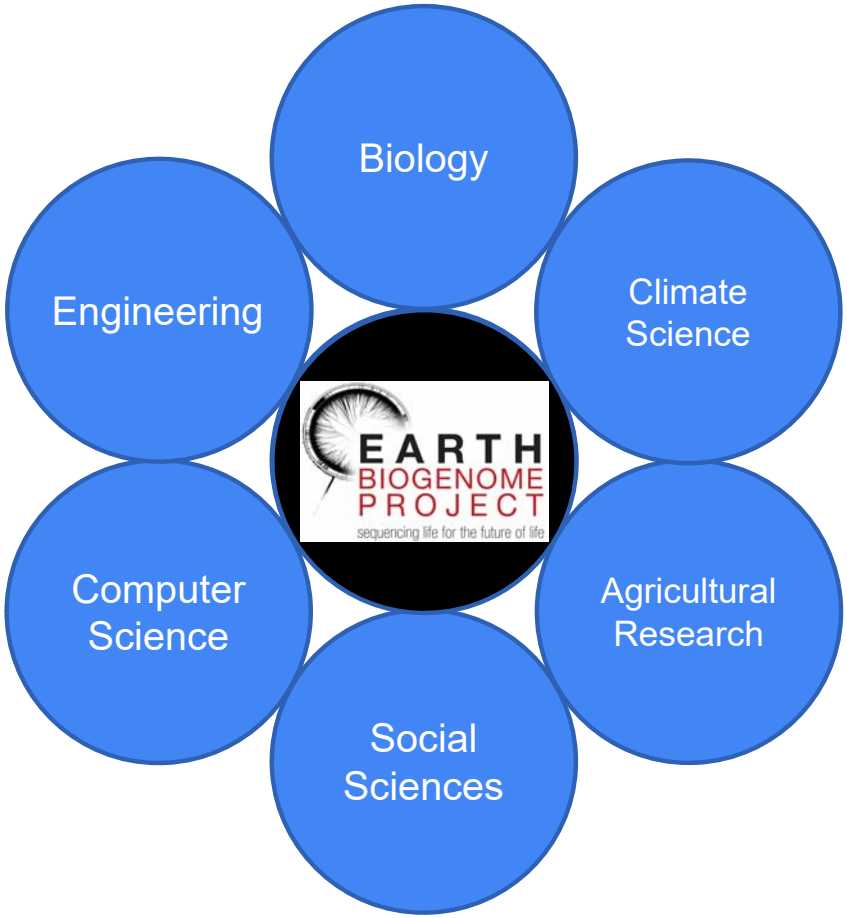
Discovering the trees of life



Biodiversity Genomics: Responding to Global Needs & Opportunities

- Mitigate Effects of Climate Change
  - Prevent loss of biodiversity
  - Adaptation of agriculturally important plants, animals and insects
  - Prevent spread of invasive species
  - Forest preservation and reconstitution
  - Protect, stabilize and restore Earth's ecosystems
- Build a Sustainable Bioeconomy
  - Nature's blueprints for engineering biology (synthetic biology)
  - Resources for the biotechnology industry (e.g., CRISPR, optogenetics, etc).
  - New pharmaceuticals from nature
  - Organisms and enzymes for bioprocessing
  - Clean energy
- Pandemic Preparedness & Biosecurity (human, animal and plant)
- Conservation of Endangered Species

Convergence Science at the Nexus of the Global Biodiversity and Climate Crises



EBP Strategy 1: The Phylogenomic Wave

EARTH BIOGENOME PROJECT
sequencing life for the future of life

- Domains: 3 (Eubacteria, Archaea, Eukarya)
- Eukaryotic Kingdoms: 5 (animal, plant, fungi, chromista and protozoa)
- Eukaryotic Phyla: 70 (33 animal; 8 plant; 5 fungi; 24 chromists+protozoa
- Eukaryotic Classes: 287
- Eukaryotic Orders: 1305
- Eukaryotic Families: 9302 (Phase I; reference quality, Y1-Y3)
- Eukaryotic Genera: 160,000-200,000 (Phase II, Y4-Y7)
- Eukaryotic Species: ~1.8 million known (Phase III, Y8-Y10)

Sources: Global Genome Initiative (Jonathan Coddington & Mike Trizna); Catalog of Life (Luisa Abucay & Yuri Roskov)
NCBI, September, 2018

EBP Strategy 2: Ecosystem Sequencing; "Google Life"

- Location Sampling (e.g. Ocean Sampling Day Consortium; Genomic Observatories Network; NEON; Critical Zone Observatory; CAL*e*DNA)
- Sequence all organisms in a particular geographical area (e.g., within biodiversity hotspots); soil, land, water and air (Wytham Woods, Darwin Tree of Life Project)
- Enables studies of the effect of environmental change on biodiversity (bio-surveillance & genomic ecology)

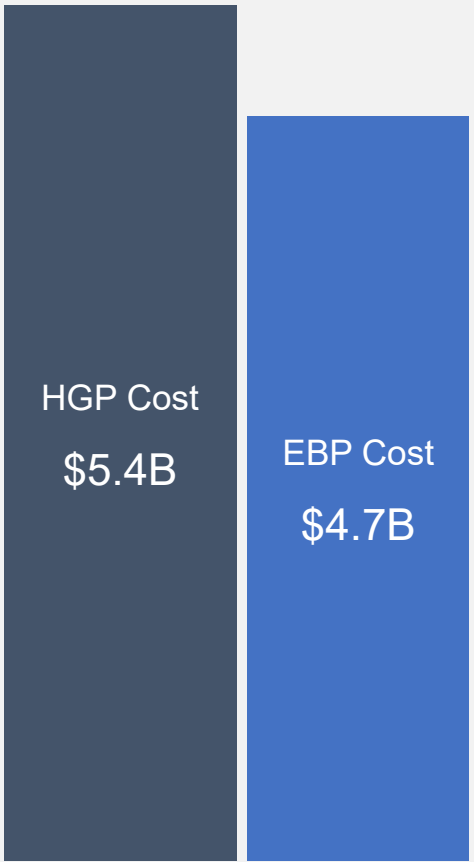
What will it cost and what are the expected returns?

The Human Genome Project (HGP)

Cost > $5.4 billion in 2017

+ other investments in genomics = $14.5 billion

ROI of the HGP estimated at ~$1 trillion, or 65:1 (Battelle Report, 2013) in the US alone



The Earth BioGenome Project (EBP)

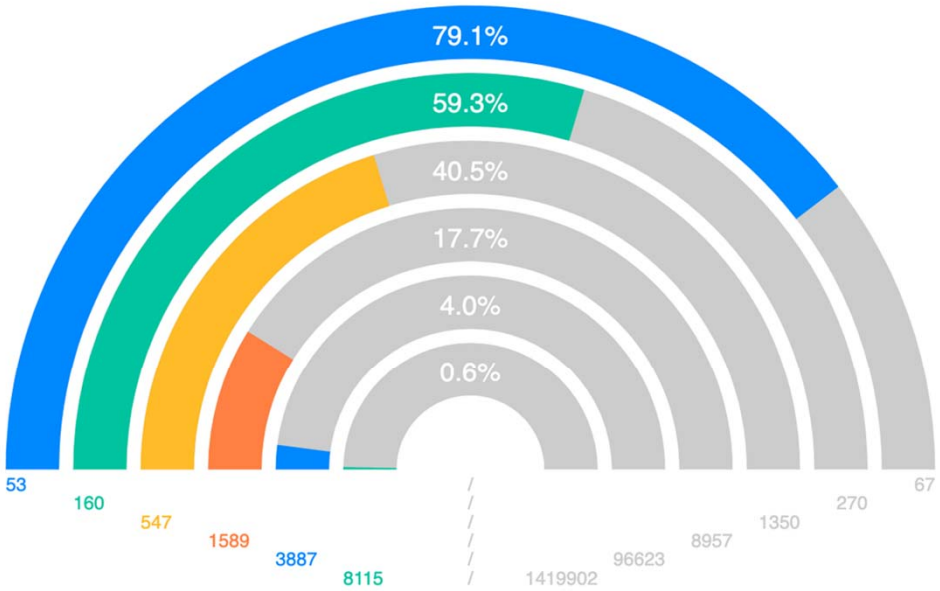
$4.7 billion over 10 years (globally)

ROI of the EBP will likely be many times more than the HGP

EBP Contribution to Eukaryotic Genome Sequencing

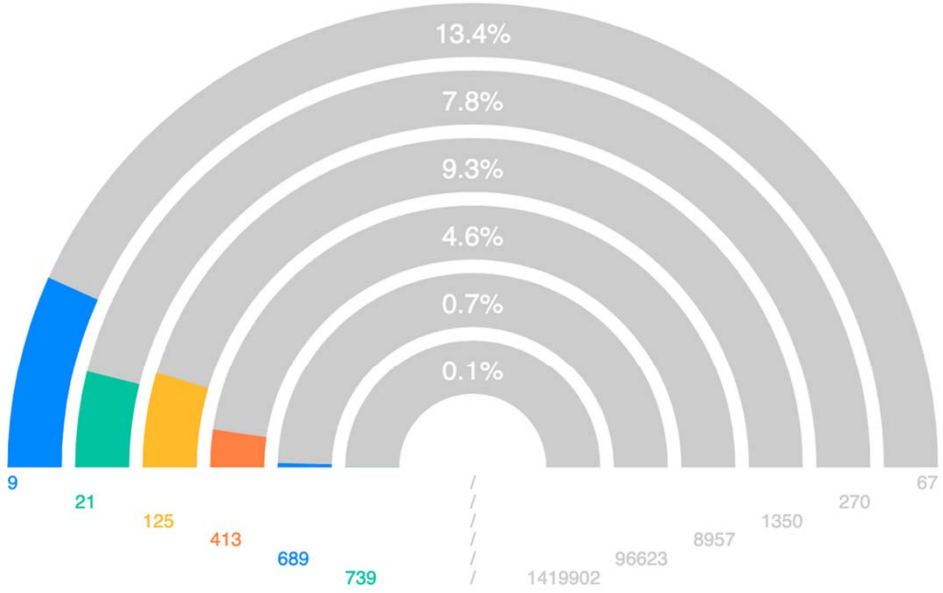
as of September 21, 2021

All NCBI taxa



Taxa with assemblies out of all Euk taxa in NCBI

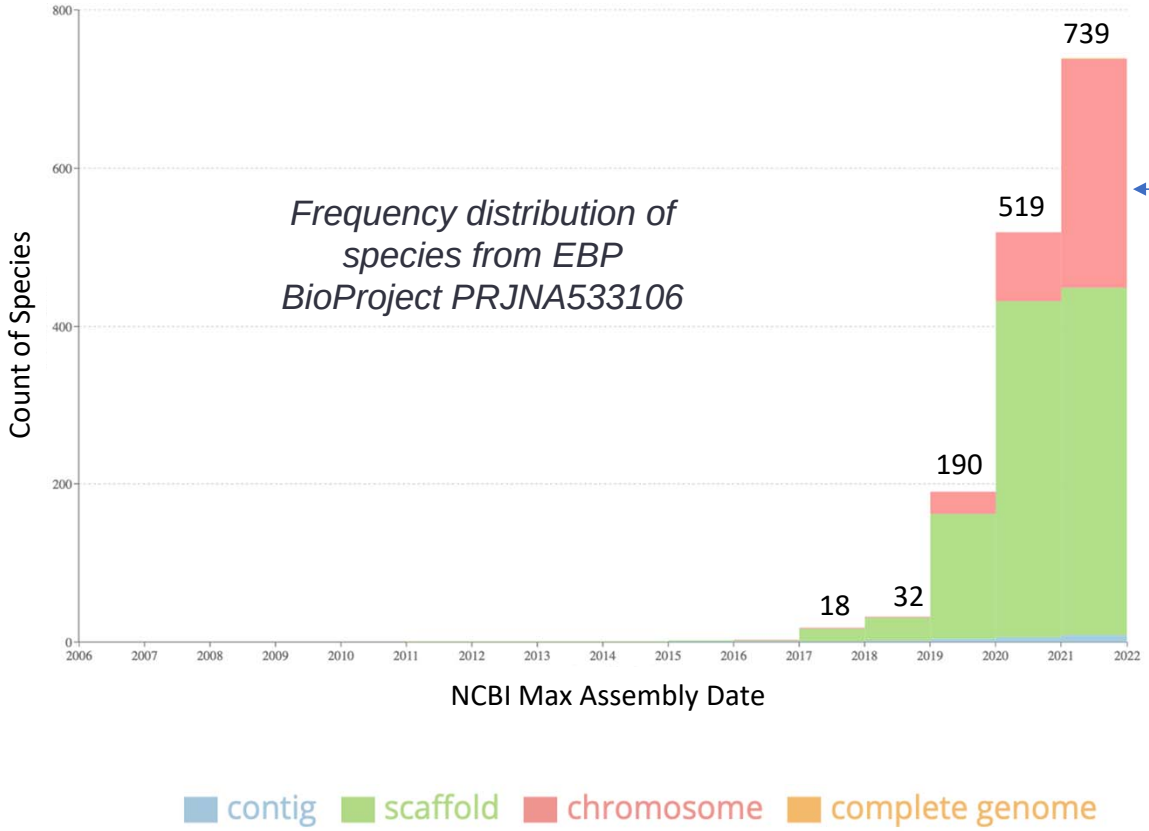
EBP Umbrella (BioProject PRJNA533106)



EBP taxa with assemblies out of all Euk taxa in NCBI

GoaT data release 2021.09.21

Progress in Whole Genome Sequencing of Eukaryotic Species
EBP Only



40% of all chromosome-level assemblies

GoaT data release 2021.09.21

Sequencing summary: All EBP affiliated projects: September, 2021

EBP Affiliated Projects (N=49)	No. species	No. reference quality	No. families w/reference quality	No. reference quality (in progress)	No. draft quality (in progress)
Total	1,719	412	316	3,021	2,156

The Technical Challenges Ahead: Opportunities for Convergence Research

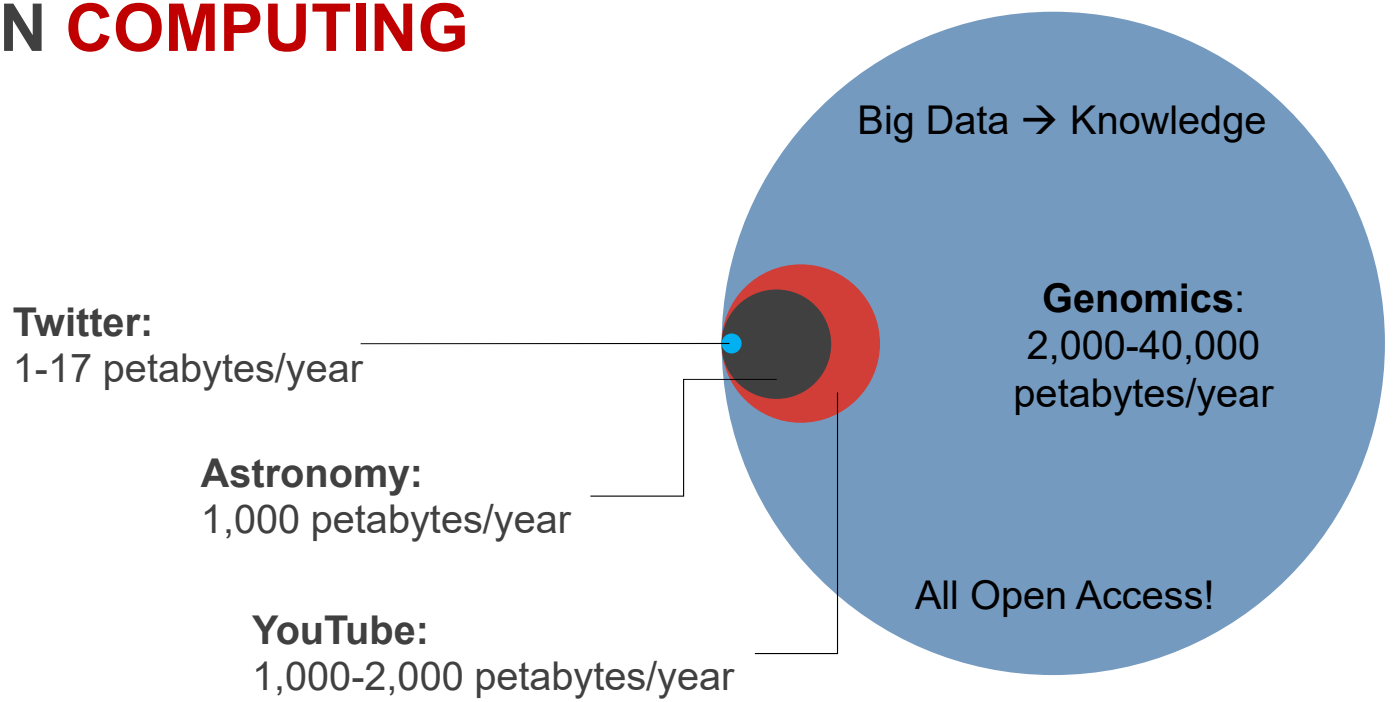
- **Sourcing, acquiring and vouchering thousands of specimens**
- HMW DNA and RNA isolation at scale
- Managing data workflows internationally
- Sequencing capacity and throughput
- **Assembly and curation at scale**
- **Annotation at scale**
- **Whole genome alignments at scale**
- Comparative genomic analysis and **data visualization at scale**

- Sequencing capacity & throughput
  - Phase I (9,300 in Y1-3)
    - 9 genomes/day √
  - Phase II (~180,000 in Y4-7)
    - 123 genomes/day (↑14x)
  - Phase III (1.32 M in Y8-10)
    - 1,205/day (↑9.8x)

EBP WILL SPUR INNOVATION IN TECHNOLOGY

- Sample collection
- Sample identification
- Sample sequencing

EBP WILL SPUR INNOVATION IN COMPUTING



Source: "Big Data: Astronomical or Genomical?" *PLoS Biology*, 7, 2015

1.8 million species x 2 Gb = 3.6×10^{15} bytes of data (finished sequence only)

Tree of life @ Sanger

Chromosomally complete genomes from across the tree of life

Professor Mark Blaxter

Tree of Life Programme Lead

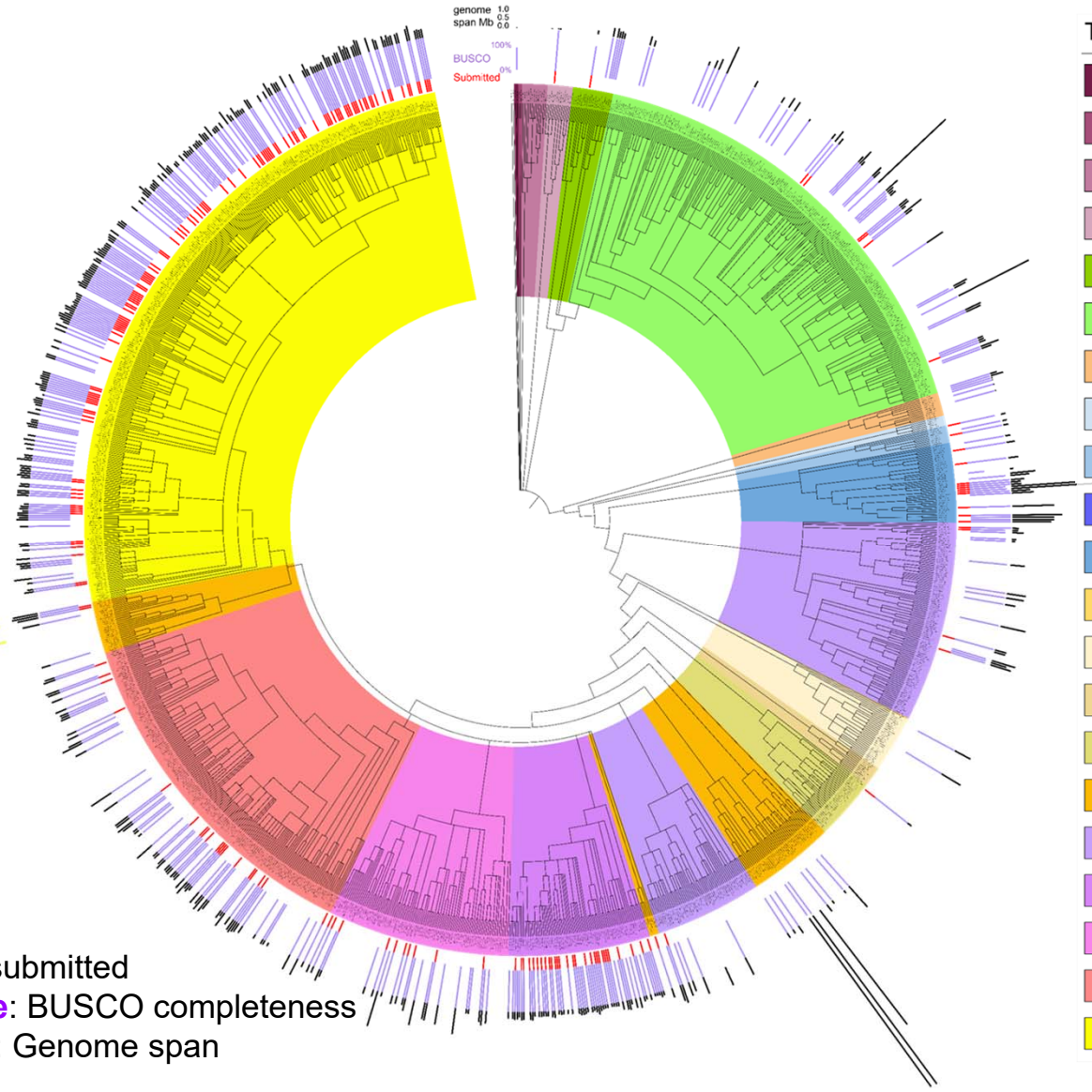
wellcome sanger institute

DToL Progress

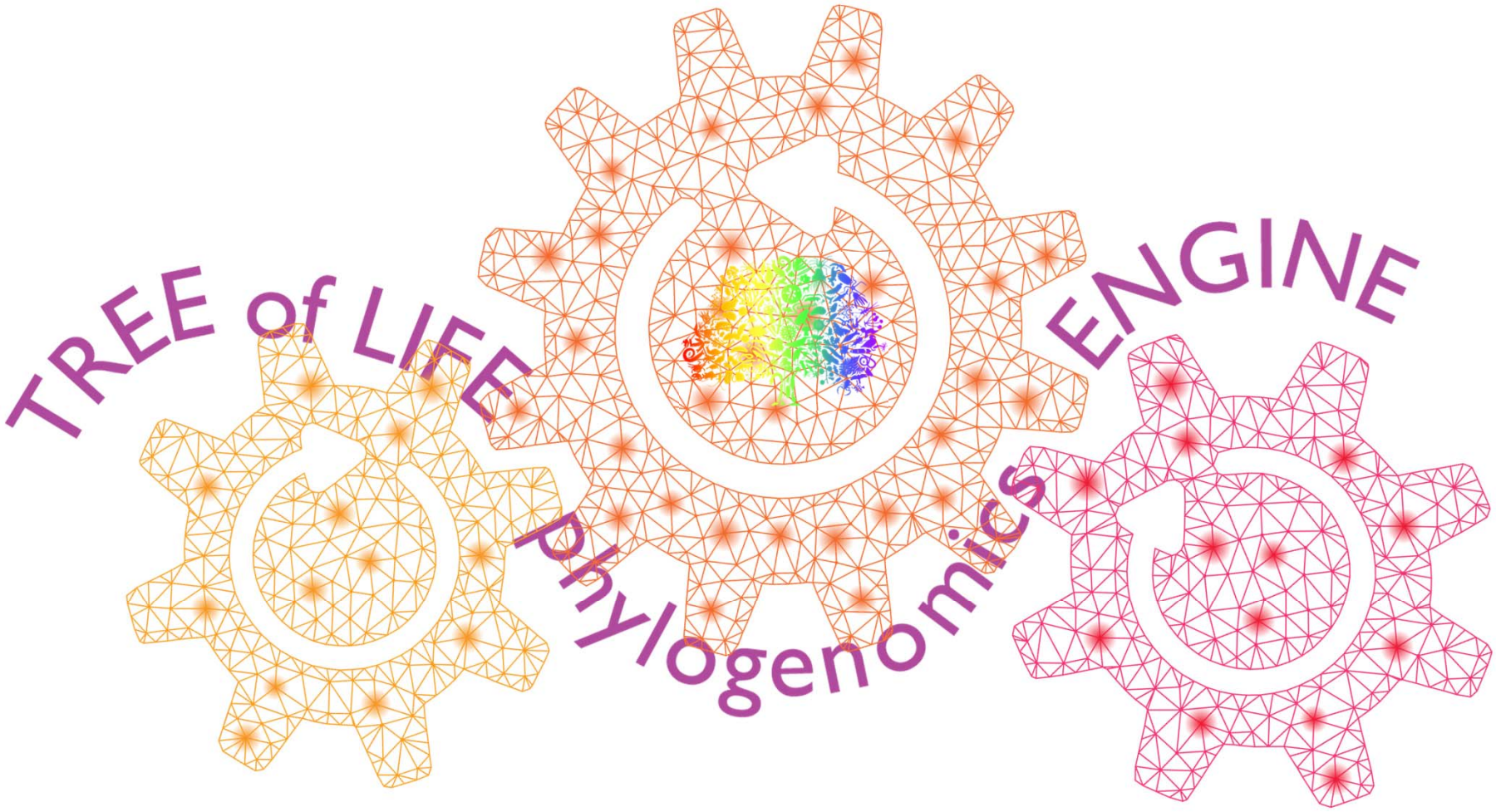
Nov 2021

210 species submitted

~1300 species in progress



Red: submitted
Purple: BUSCO completeness
Black: Genome span



Linked data management, wet lab and bioinformatics tools that can deliver genomes at scale

wellcome sanger institute

A taste of what's to come!

28 April 2021

Erich Jarvis, Chair

GENOME 10K

- Near-complete genomes for 16 vertebrate species
- Changed what's known about several species: discovered previously unknown chromosomes in platypus and zebra finch
- Discovered key differences between marmoset and human brain-related genes
- Discovered new insights into evolution of key neurochemicals oxytocin and vasotocin
- Gained new insights for conservation: Kākāpō, a critically endangered parrot from New Zealand; is able to purge deleterious mutations from its genome despite low genetic diversity

Conserve & regenerate biodiversity

Zoonomia Consortium, *Nature* 587:240-245, 2020

Genomic diversity and threatened species status

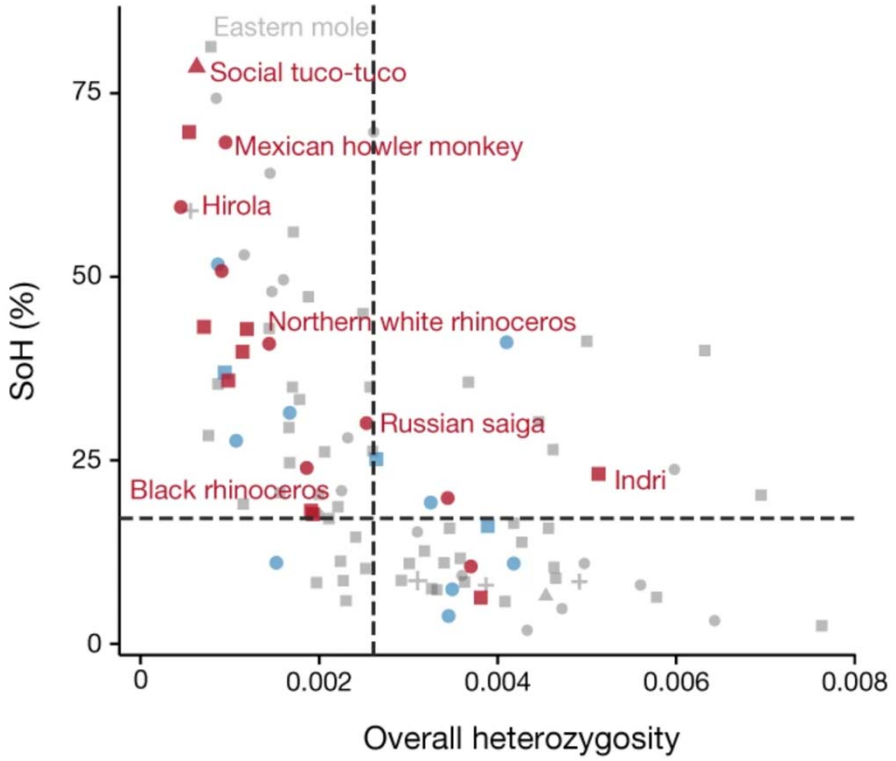
Fossa

Arctic fox

Hirola

Bumblebee bat

Snowshoe hare



Aye-aye

Geoffroy's spider monkey

Southern three-banded armadillo

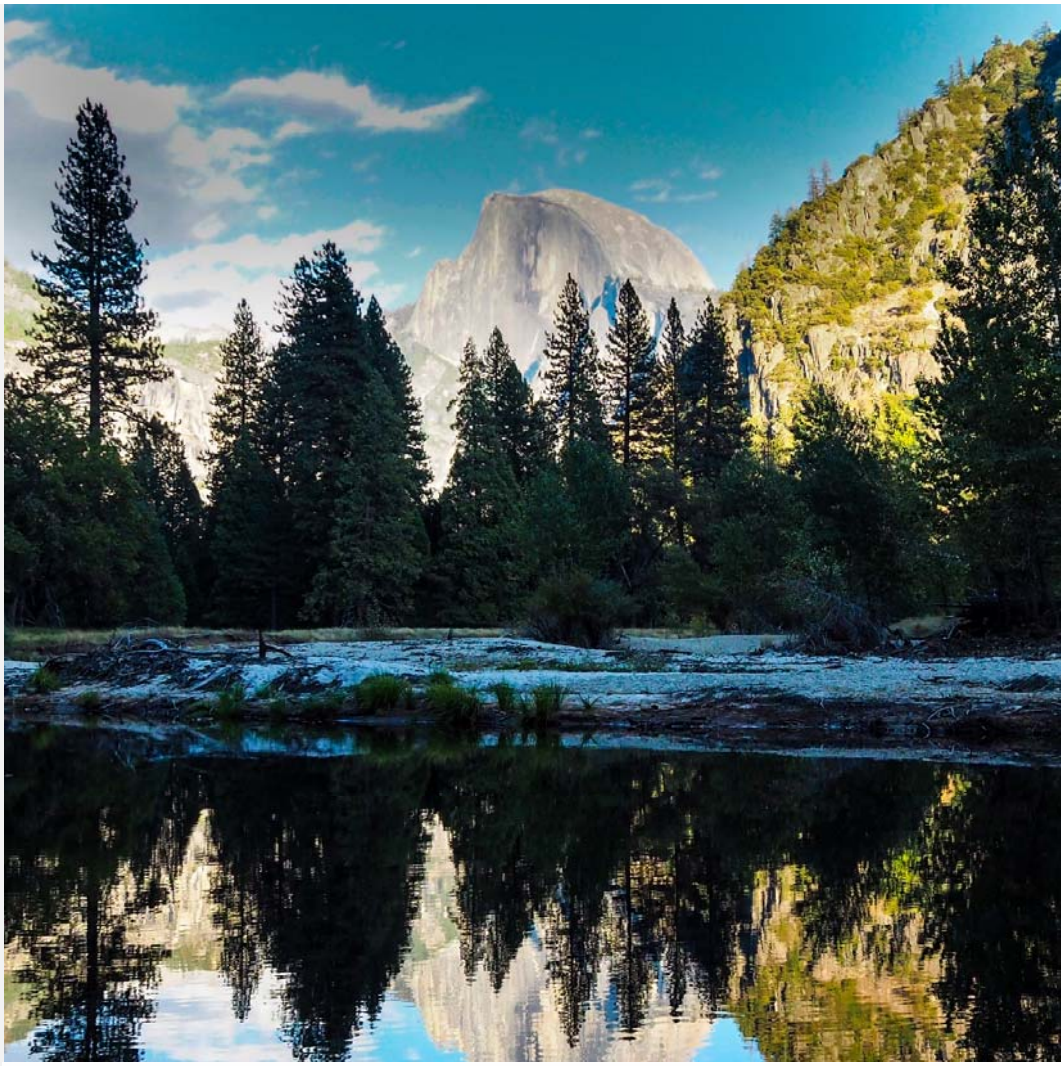
Giant anteater

Brown-throated sloth

California Conservation Genomics Project (CCGP)

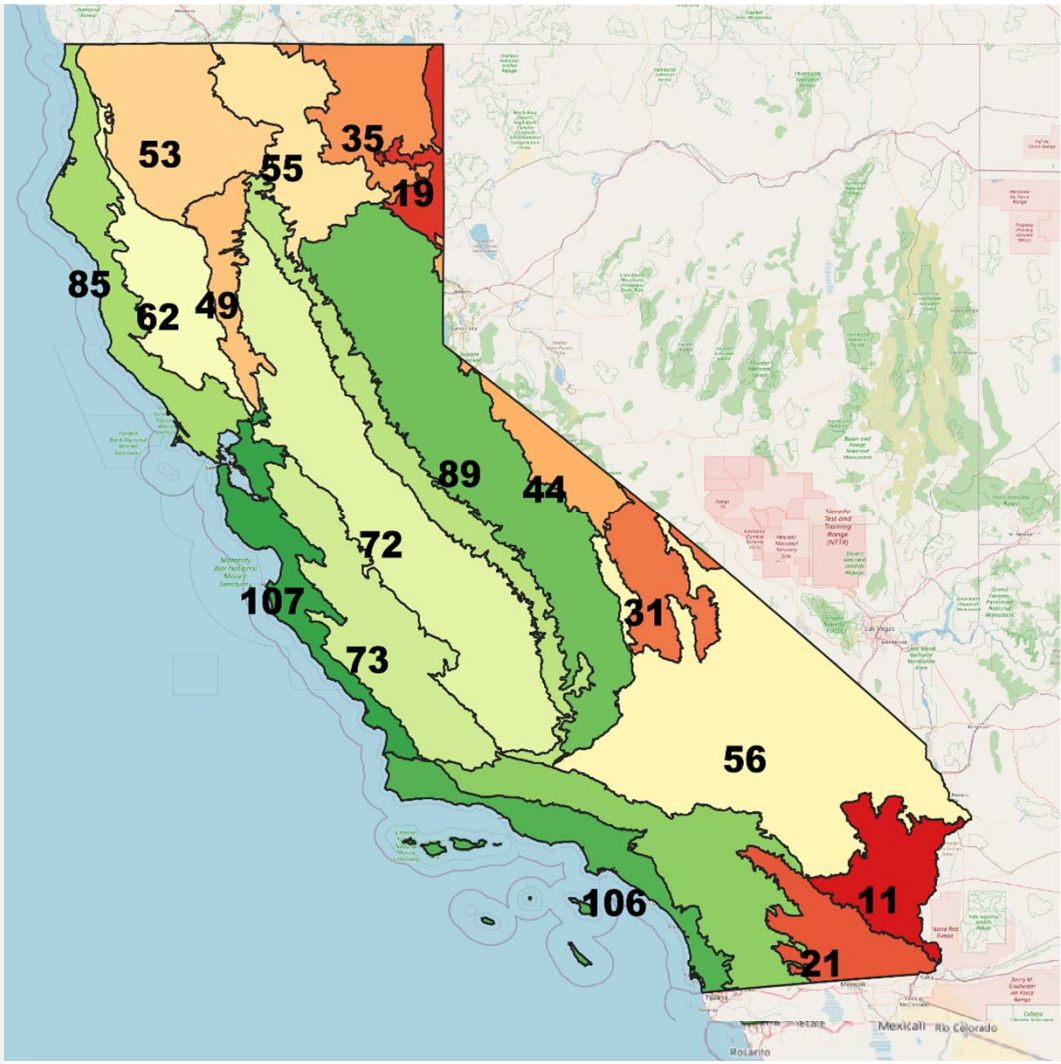
Brad Shaffer
Director, UCLA La Kretz Center for Conservation Science

A collaborative effort to conserve California flora and fauna using conservation & landscape genomics of threatened, endangered and commercially exploited species



California Conservation Genomics Project (CCGP)

- 225 species, ~130 projects
- 79 PIs
- 10 UC campuses
- ~150 reference genomes (HiFi + Omni-C)
  UC Davis Genome Center, UCSC & Dovetail
- Annotation (RNASeq)
  Illumina-EBP100 Program)
- 150 re-sequenced genomes/species
  (~20,0000 re-sequenced genomes)



ccgproject.org

cdfa

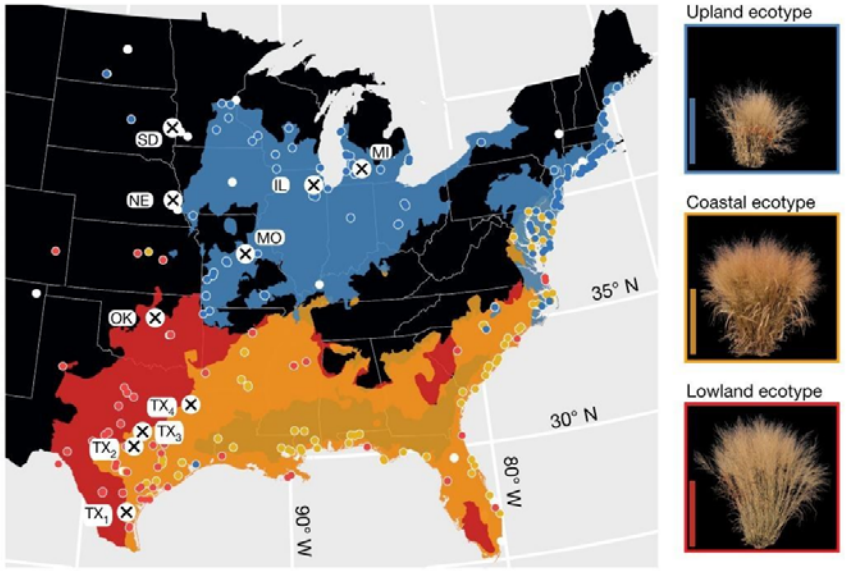
California Department of Fish and Wildlife

Adaptation and resilience to climate change

Article | Open Access

Published: 27 January 2021

Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass



Lovell et al., *Nature* 2021

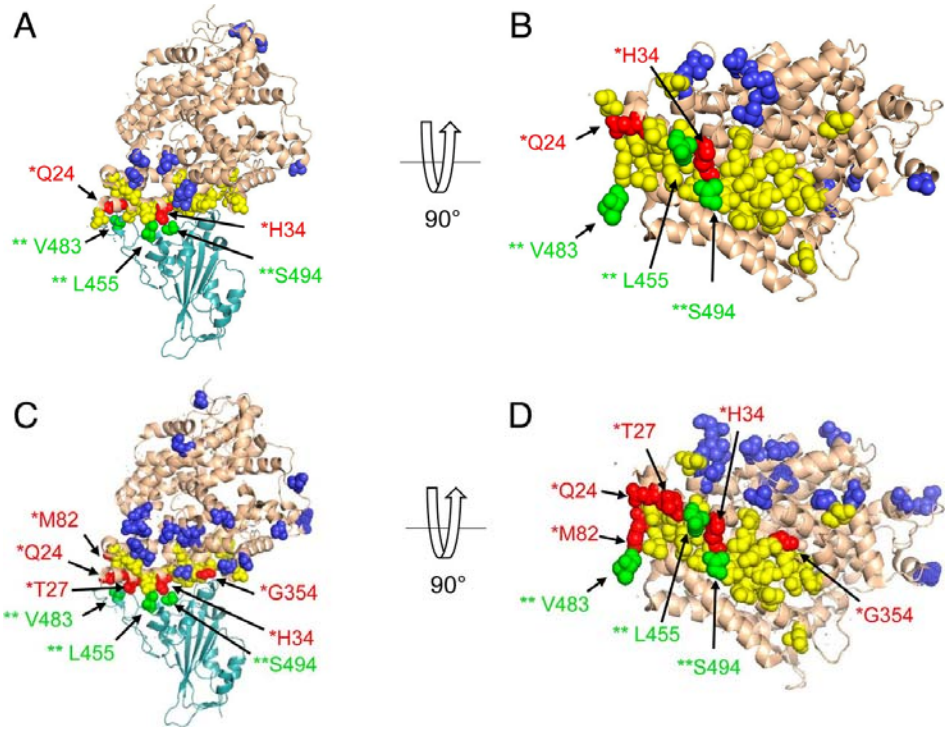
Improved pandemic prediction & responsiveness

PNAS 1808

Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates

Joana Damas, Graham M. Hughes, Kathleen C. Keough, Corrie A. Painter, Nicole S. Persky, Marco Corbo, Michael Hiller, Klaus-Peter Koepfli, Andreas R. Pfenning, Huabin Zhao, Diane P. Genereux, Ross Swofford, Katherine S. Pollard, Oliver A. Ryder, Martin T. Nweeia, Kerstin Lindblad-Toh, Emma C. Teeling, Elinor K. Karlsson, Harris A. Lewin

doi: https://doi.org/10.1101/2020.04.16.045302



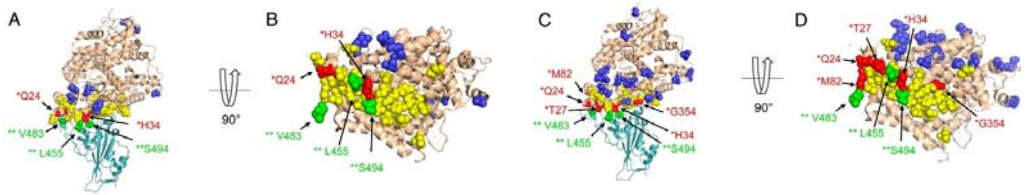
Improved pandemic prediction & responsiveness

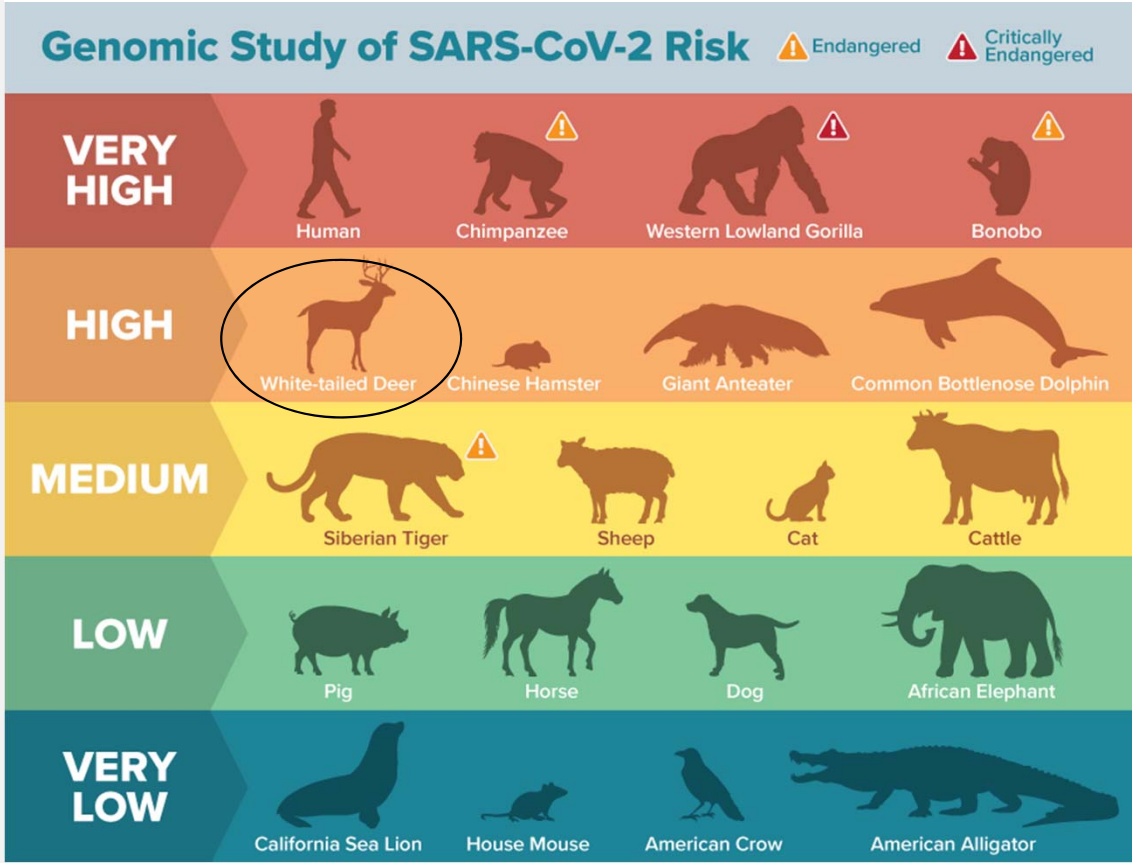
PNAS 2065

Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates

Joana Damas, Graham M. Hughes, Kathleen C. Keough, Corrie A. Painter, Nicole S. Persky, Marco Corbo, Michael Hiller, Klaus-Peter Koepfli, Andreas R. Pfenning, Huabin Zhao, Diane P. Genereux, Ross Swofford, Katherine S. Pollard, Oliver A. Ryder, Martin T. Nweeia, Kerstin Lindblad-Toh, Emma C. Teeling, Elinor K. Karlsson, Harris A. Lewin

doi: https://doi.org/10.1101/2020.04.16.045302

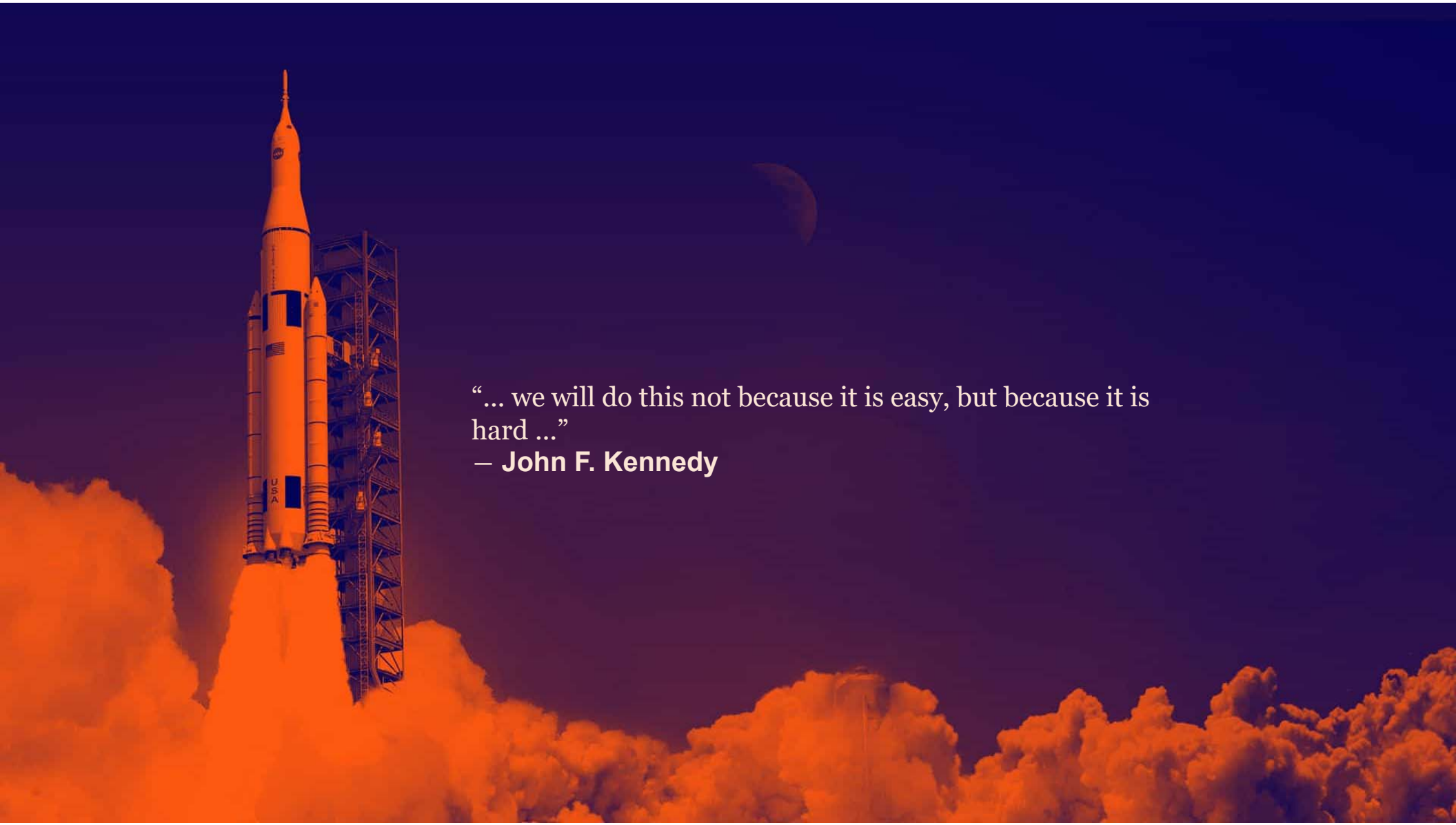




Improvements to human welfare

New technologies
to feed the world

New medicines and
biomaterials from nature


"... we will do this not because it is easy, but because it is hard ..."
— John F. Kennedy

Dr. Harris A. Lewin
Chair, Earth BioGenome
Project Working Group
University of California, Davis

Dr. W. John Kress
Co-Chair, Earth BioGenome
Project Working Group
Smithsonian Institution

Dr. Gene E. Robinson
Co-Chair, Earth BioGenome
Project Working Group
University of Illinois at Urbana-Champaign

Mark Blaxter
Stephen Richards
Olga Vinnere Pettersson
Kerstin Lindblad-Toh
Adam Phillippy
warren johnson
Arang Rhie
Rebecca Johnson
Katharina Hoff
Harris Lewin
Kerstin Howe
Vincent(Xiao feng) WEI
Kim Pruitt
erichjarvis
Andrew Torrance
John Kress
Richard Durbin
Jonathan Coddington
Paul Flicek
mara

EARTH
BIOGENOME
PROJECT
Sequencing life for the future of life
www.earthbiogenome.org
Twitter: @EBPgenome