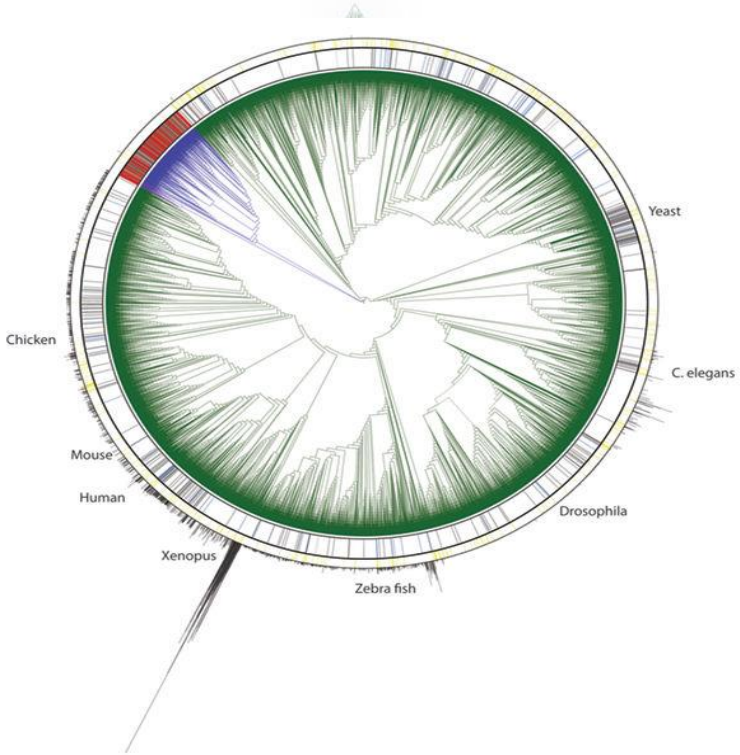
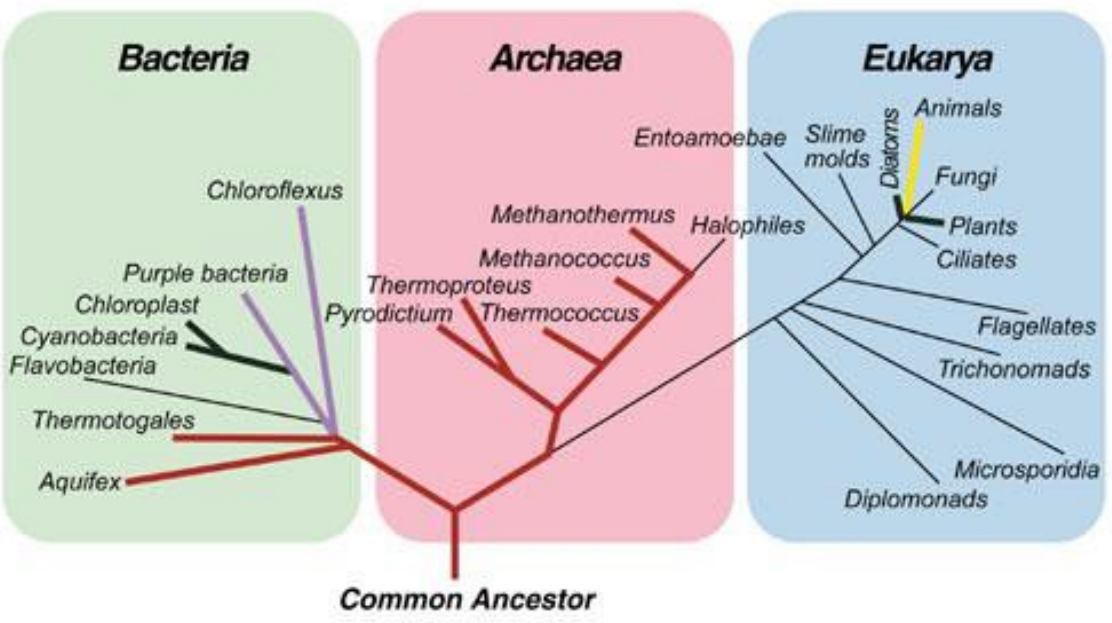


The Earth BioGenome Project:
Sequencing Life for the Future of Life

Harris Lewin, Gene Robinson, John Kress,
and the EBP Working Group

The Earth BioGenome Project: Grand Challenge



Mission: To sequence the DNA of every known species from the three life domains in 10 years

The Earth BioGenome Project: Grand Challenge

Vision: *To create a new foundation for science that drives solutions for preserving Earth's biodiversity and sustaining human societies*

Organizational Meetings

- **Establishing a National and Global Framework to Sustain Biodiversity: "The Earth BioGenome Project"**

  November 6, 2015

  Smithsonian Castle Commons, Washington, D.C.

- **The Earth BioGenome Project**

  August 8, 2016

  US National Academy of Sciences Council Meeting

  Woods Hole, MA

- **The Earth BioGenome Project**

  September 19, 2016

  OSTP

  Washington, DC

- **Earth BioGenome Workshop & Global Biodiversity Genomics Conference**

  February 20-23, 2017

  National Museum of Natural History, Washington, D.C.

Science
AAAS

Sequencing all life captivates biologists

“To sequence everything in the world—that is the reason we are here.”

Huanming Yang, quoted by Elizabeth Pennisi, *Science*

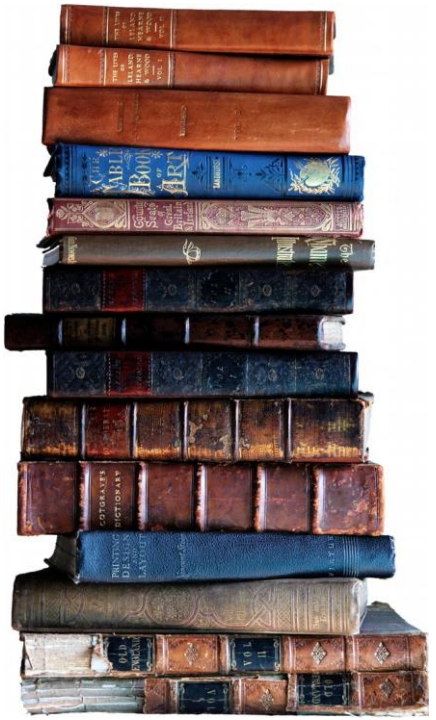
WHY SEQUENCE LIFE?

Maximize returns to society and human welfare

Conserve, protect, and restore biodiversity

Revolutionize biology

Genomes are books of life



UCDAVIS genome center

How many species are there?

Species	Earth			Ocean		
	Catalogued	Predicted	±SE	Catalogued	Predicted	±SE
Eukaryotes						
Animalia	953,434	7,770,000	958,000	171,082	2,150,000	145,000
Chromista	13,033	27,500	30,500	4,859	7,400	9,640
Fungi	43,271	611,000	297,000	1,097	5,320	11,100
Plantae	215,644	298,000	8,200	8,600	16,600	9,130
Protozoa	8,118	36,400	6,690	8,118	36,400	6,690
Total	1,233,500	8,740,000	1,300,000	193,756	2,210,000	182,000
Prokaryotes						
Archaea	[illegible]2	455	160	1	1	[illegible]
Bacteria	10,358	9,680	3,470	652	1,320	436
Total	10,860	10,100	3,630	653	1,32[illegible]	43[illegible]
Grand Total	**1,244,360**	**8,750,000**	**1,300,000**	**194,409**	**2,210,000**	**182,000**

Press Release 16-052

Researchers find that Earth may be home to 1 trillion species

Largest analysis of microbial data reveals that 99.999 percent of all species remain undiscovered

National Science Foundation
WHERE DISCOVERIES BEGIN

Mora et al., PLoS Biology 9:e1001127, 2011

UC DAVIS genome center

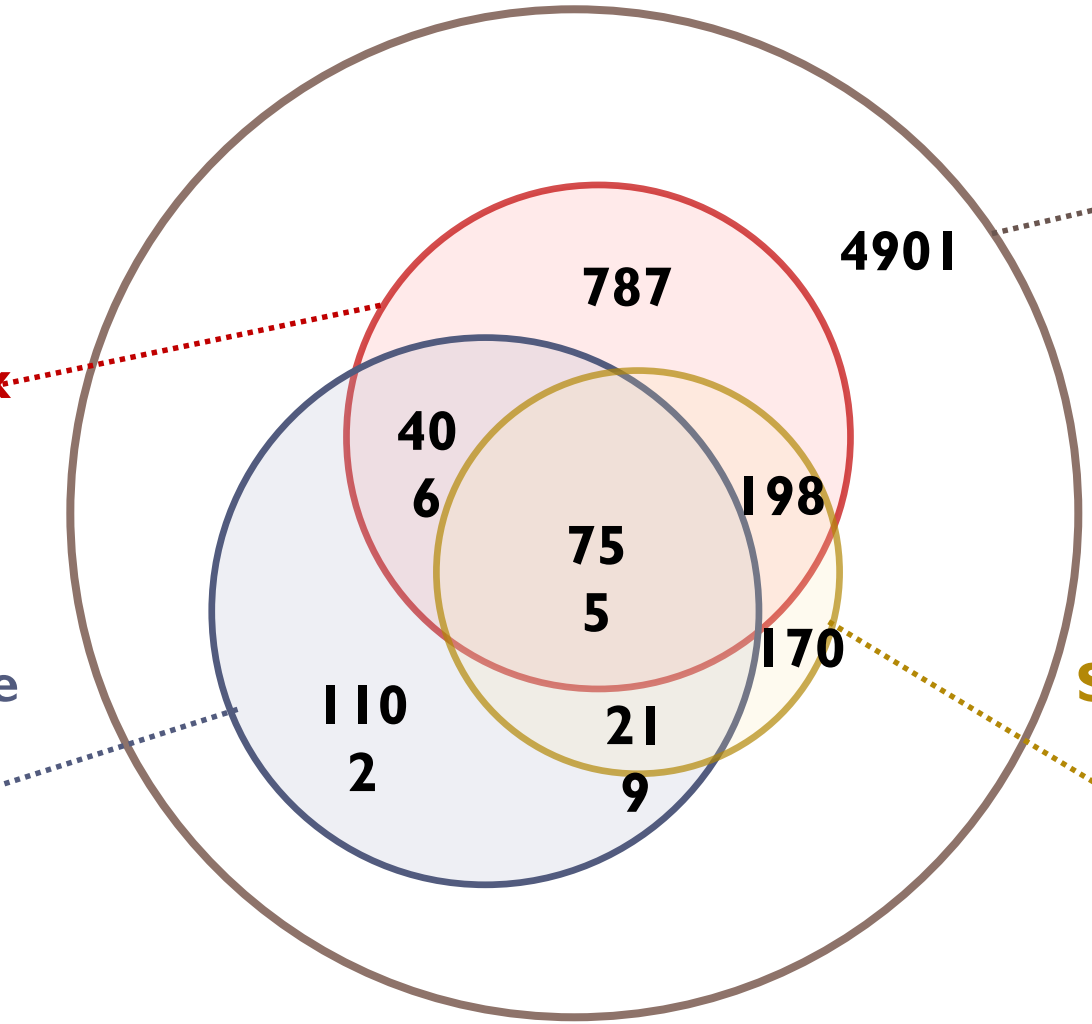
EBP Strategy 1: The Phylogenomic Wave

- Domains: 3 (Eubacteria, Archaea, Eukarya)
- Eukaryotic Kingdoms: 5 (animal, plant, fungi, chromista and protozoa)
- Eukaryotic Phyla: 61 (35 animal; 10 plant; 2 fungi; 14 chromists+protozoa)
- Eukaryotic Classes: 266
- Eukaryotic Orders: 1253
- Eukaryotic Families: 9330 (Phase I)
- Eukaryotic Genera: 140,000-200,000 (Phase II)
- Eukaryotic Species: ~1.5 million known (Phase III)

Sources: Global Genome Initiative Knowledge Portal (Jonathan Coddington); Catalog of Life (Luisa Abucay & Yuri Roskov)

The Sample Challenge

All Life Families



Jon Coddington et al., Smithsonian, February 2017

UCDAVIS genome center

EBP Strategy 2: "Google Life"

- Location Sampling (e.g. Ocean Sampling Day Consortium; Genomic Observatories Network; NEON; Critical Zone Observatory; CALeDNA)
- Sequence all organisms in a particular geographical area (e.g., within biodiversity hotspots); soil, land, water and air
- Enables studies of the effect of environmental change on biodiversity (genomic ecology)
- Produce a multidimensional and dynamic view of life on earth

Is it feasible to sequence 1.5 million Euks?

- YES – even with today's technology

- Next generation sequencers
  - 10,000 genomes/year
  - 15 Illumina XTen machines (or 30 NovaSeq 6000) can do 1.5 million genomes in 10 years
  - Assumes 3 Gbp average genome (range 2.9 Mbp – 130 Gbp) and 30-50x coverage

- Costs
  - Biorepositories, sample collection, processing and distribution
  - BioObservatories
  - Sequencers
  - Genome Sequencing and Annotation
  - Diversity and RNA sequencing
  - IT and Bioinformatics
  - **Total Estimated Cost: ~$4.3 billion**

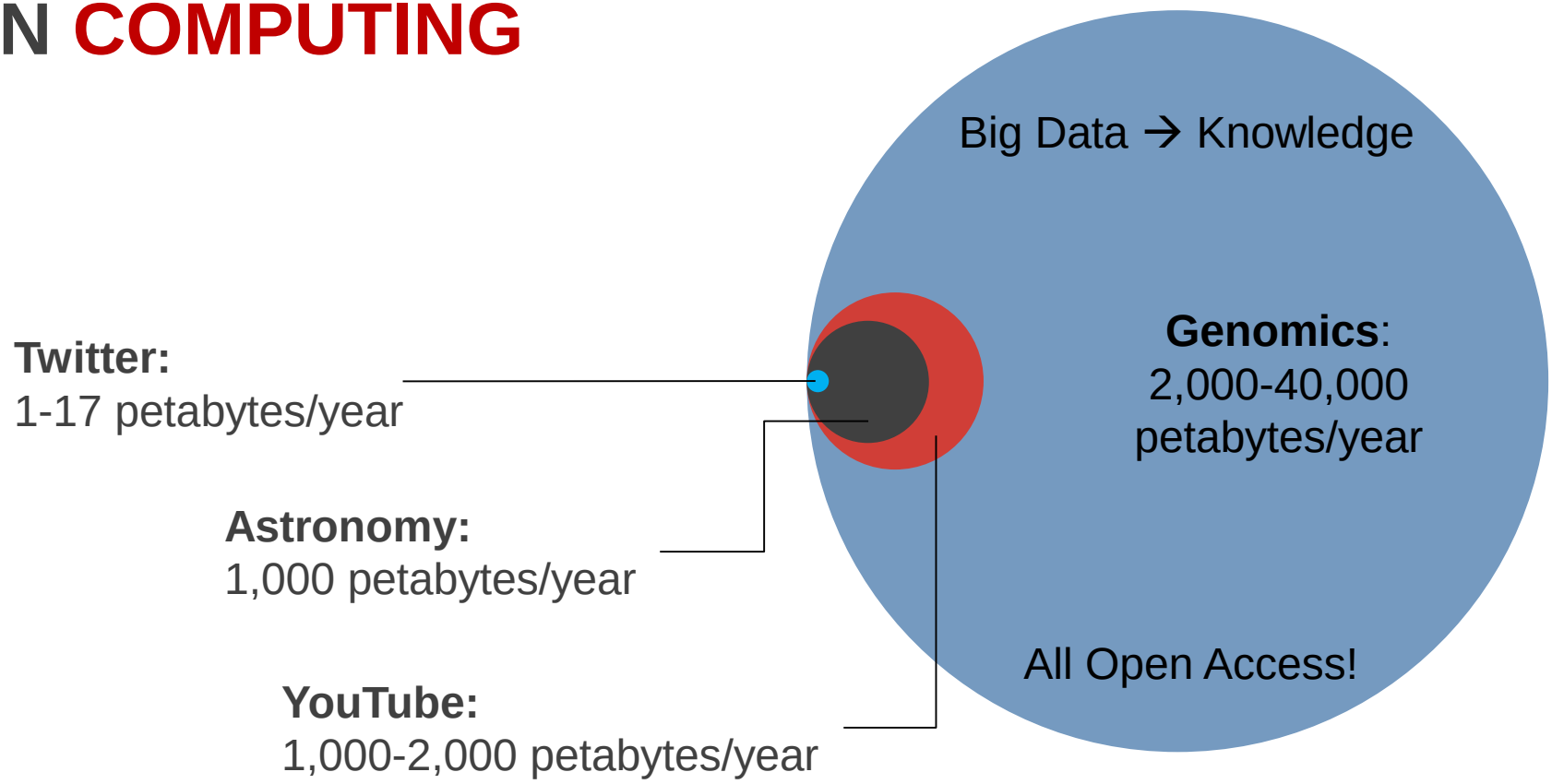
- Cost of Human Genome Project: $2.7 billion 1991 dollars ($4.814 billion in 2017 dollars)
  - Cost of B2 bomber program (through 2004) = $44.75 billion (1997 dollars)
    - $2.13 billion per aircraft ($3.1524 billion in today's dollars)

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

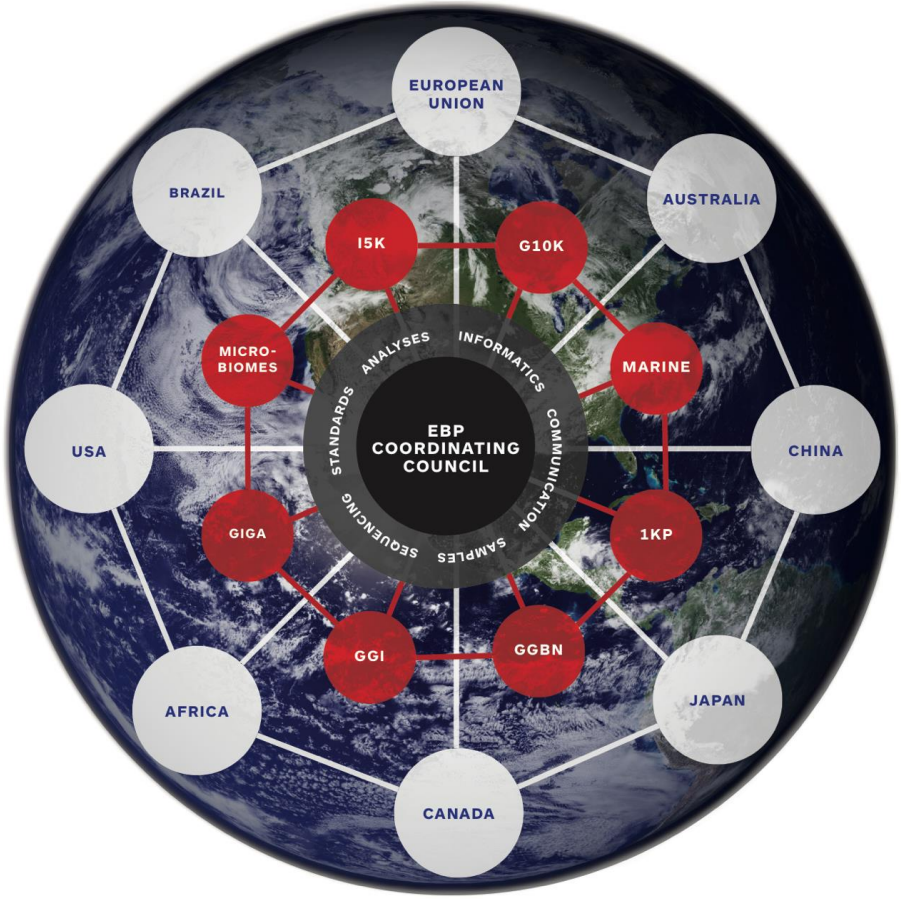
EBP WILL INSPIRE THE
NEXT GENERATION OF SCIENTISTS

UCDAVIS genome center

GLOBAL NETWORK OF COMMUNITIES

Open access

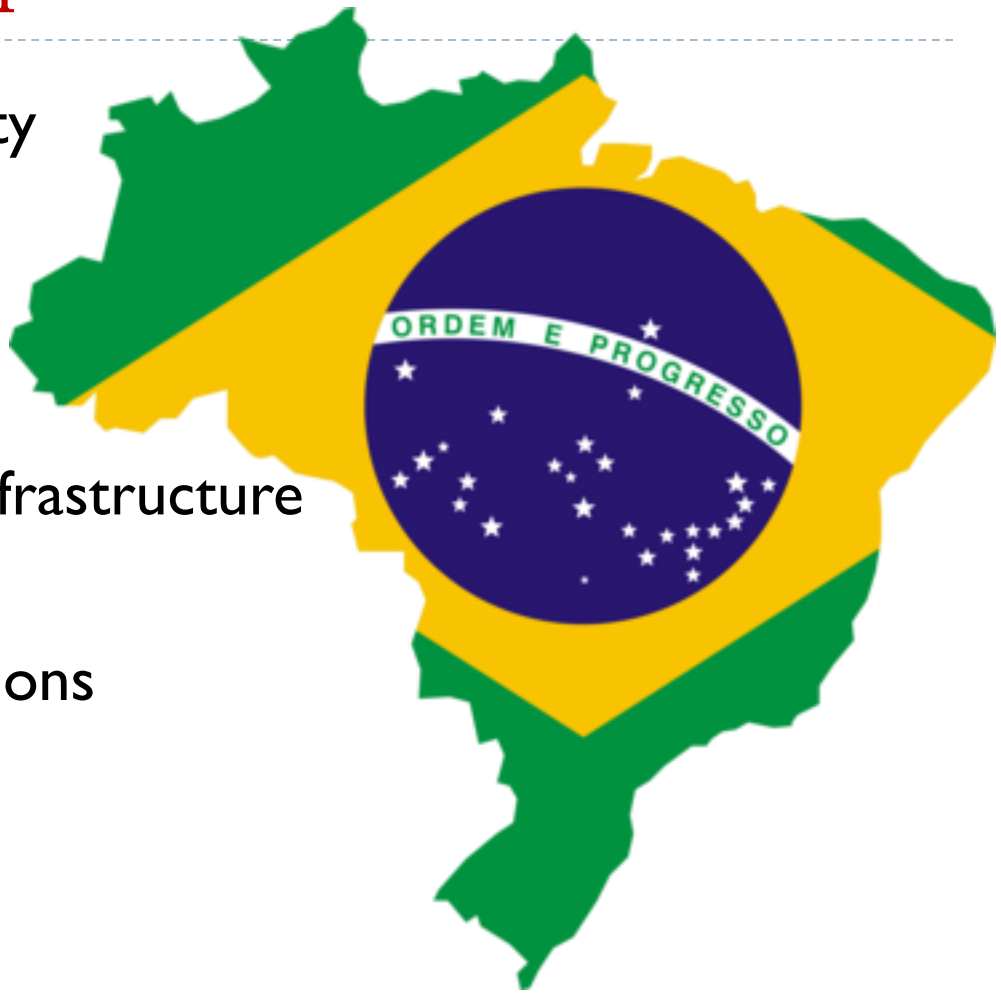
Compliance with the Convention on Biological Diversity and the Nagoya Protocol on Access and Benefit Sharing (ABS)



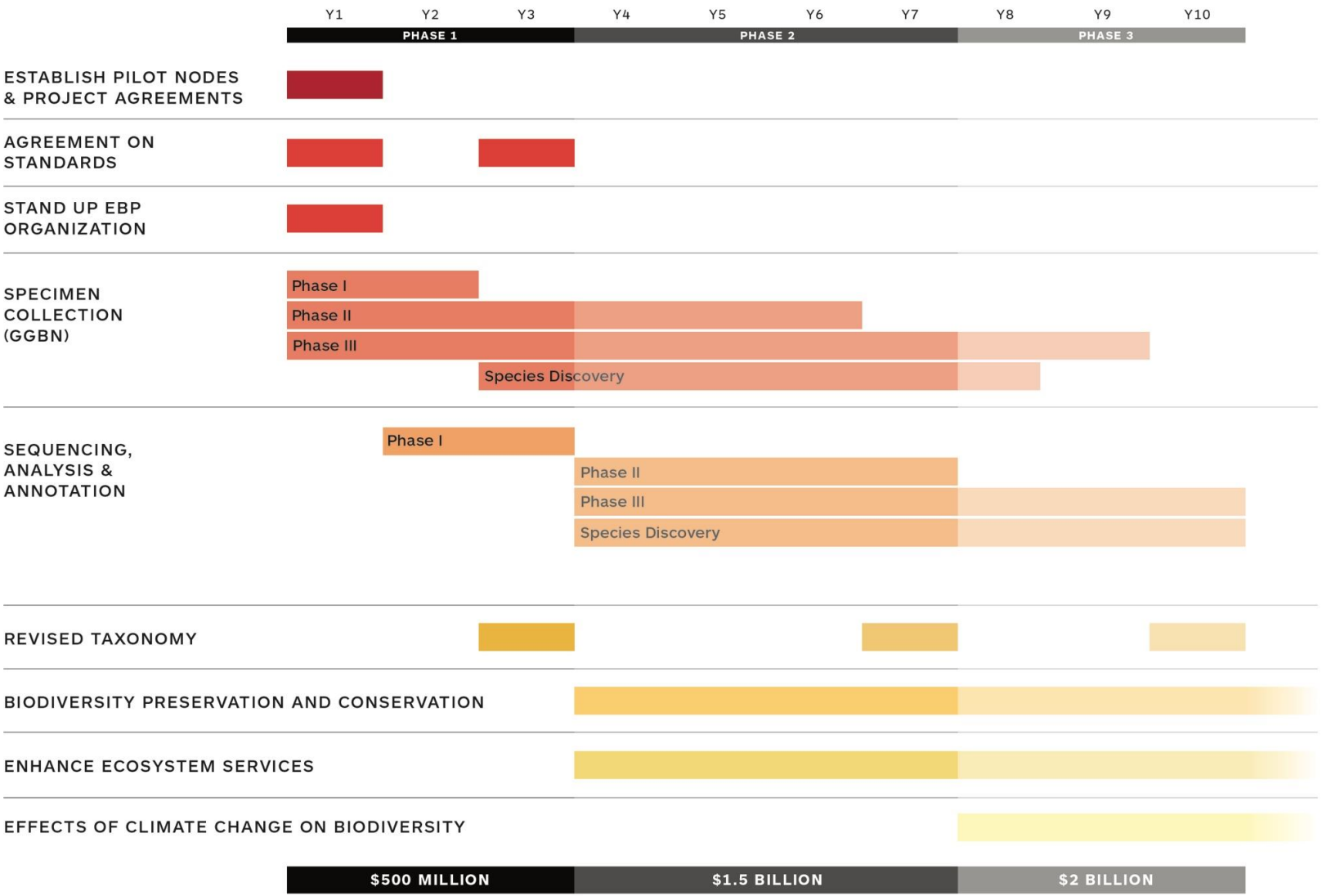
International EBP working group ***already established.***

Major Role For Brazil

- ~10% of world's biodiversity
- Curated collections
- Well-established science infrastructure
- Global research collaborations



PROJECT ROADMAP (2018 launch)

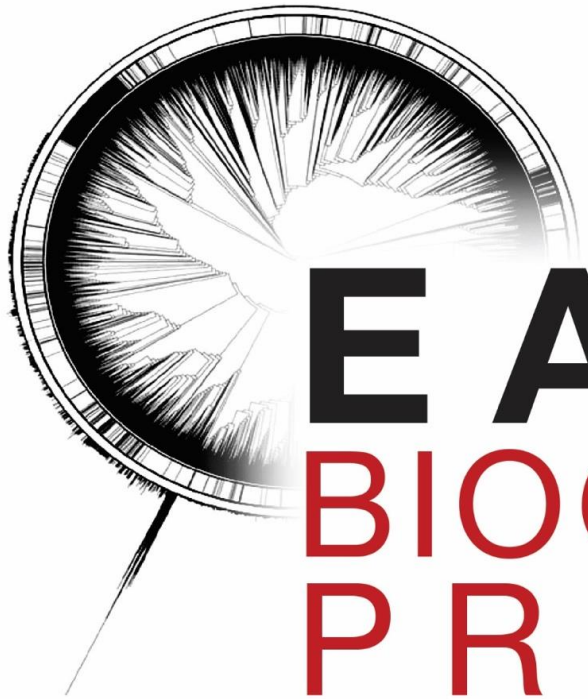


Funding Sources for the EBP

- National funding agencies (global)
  - Frame as a global grand challenge
  - Most US federal agencies would have some stake in this project
- Private Corporations
  - Technology providers have a stake
  - Ecosystem health has economic value
- Foundations
  - Attractive as a big idea whose time has come
  - Preserving biodiversity has great impact on culture and the environment
- Crowd sourcing and funding
  - Citizen science model (e.g., Kittybiome; CALeDNA)
  - Commemorative sponsorship

EBP: the most ambitious project in the history of biology

- First true global mega-genome sequencing project
- Involve thousands of scientists and millions of citizens around the world
- Very Big Data (exceeding 1 Exabyte) will drive new computational architectures, methods and models
- Revolutionize our understanding of biology
- Radically improve conservation efforts
- Create new resources for agriculture, medicine and ecosystem services



EARTH BIOGENOME PROJECT

sequencing life for the future of life