



CENTER OF
PLANT SCIENCES



Sant'Anna
Scuola Universitaria Superiore Pisa

Advanced Genomics

Genomics at work



A few lectures on applied genomics

- **Genomics and evolution**

- The tree of life – genomics of biodiversity
- Staging biological evolution in the lab: the long term evolution experiment (LTEE)
- Optical mapping and reconstruction of SVs
- Horizontal gene transfer

- **Genomics and climate change**

- Ancient DNA sequencing
- Environmental DNA (eDNA) sequencing
- High-throughput genotyping methods based on sequencing: arrays and Restriction site associated DNA (RAD) markers
- Combining genomics with climate models to capture evolution for local adaptation

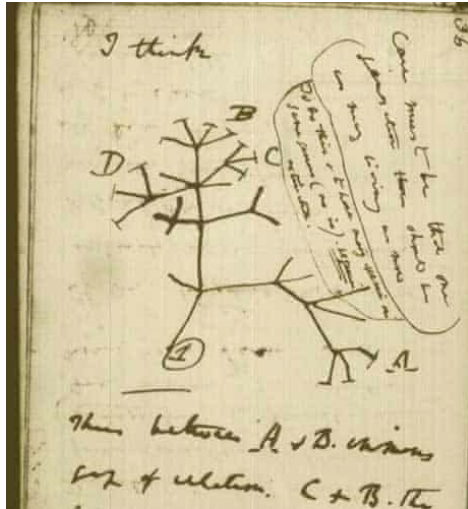
- **Genomics and humans**

- Mitochondrial DNA typing
- Population genetics of modern humans
- Ancestry kits and genomic prediction
- Concepts of quantitative genetics

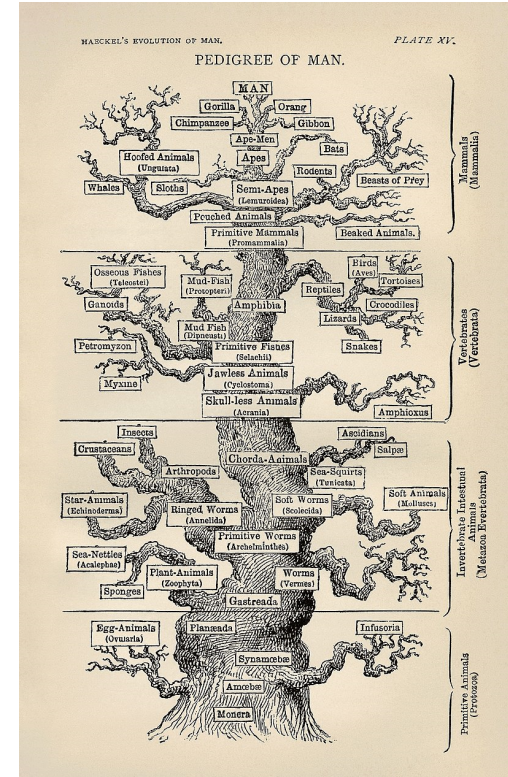


Genomics and evolution

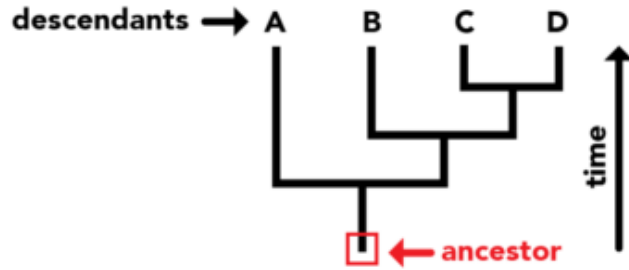
The tree of life of life is a metaphor, model and research tool used to explore the evolution of life and describe the relationships between organisms, both living and extinct



From Darwin notebook B



The root of the tree represents the ancestral lineage, and the tips of the branches represent the descendants of that ancestor



As you move from the roots to the tip, you move ahead in time



When a speciation event occurs, a single ancestral lineage gives rise to two or more daughter lineages

Early works used genotyping of selected markers, such as rRNA genes

- Easy to amplify
- Conserved across very long phylogenetic distances (everybody needs ribosomes!)

Clear separation between bacteria, archaea, eukarya

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4576–4579, June 1990
Evolution

Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya

(Euryarchaeota/Crenarchaeota/kingdom/evolution)

CARL R. WOESE*†, OTTO KANDLER‡, AND MARK L. WHEELIS§

*Department of Microbiology, University of Illinois, 131 Burrill Hall, Urbana, IL 61801; †Botanisches Institut der Universität München, Menzinger Strasse 67, 8000 Munich 19, Federal Republic of Germany; and ‡Department of Microbiology, University of California, Davis, CA 95616

Contributed by Carl R. Woese, March 26, 1990

4578 Evolution: Woese *et al.*

Proc. Natl. Acad. Sci. USA 87 (1990)

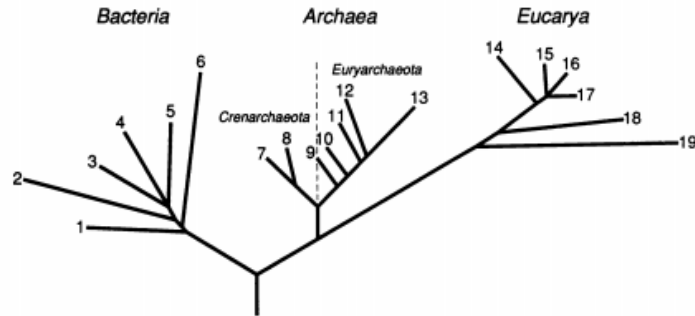


FIG. 1. Universal phylogenetic tree in rooted form, showing the three domains. Branching order and branch lengths are based upon rRNA sequence comparisons (and have been taken from figure 4 of ref. 2). The position of the root was determined by comparing (the few known) sequences of pairs of paralogous genes that diverged from each other before the three primary lineages emerged from their common ancestral condition (27). [This rooting strategy (28) in effect uses the one set of (aboriginally duplicated) genes as an outgroup for the other.] The numbers on the branch tips correspond to the following groups of organisms (2). Bacteria: 1, the Thermotogales; 2, the flavobacteria and relatives; 3, the cyanobacteria; 4, the purple bacteria; 5, the Gram-positive bacteria; and 6, the green nonsulfur bacteria. Archaea: the kingdom Crenarchaeota: 7, the genus *Pyrodicticum*; and 8, the genus *Thermoproteus*; and the kingdom Euryarchaeota: 9, the Thermococcales; 10, the Methanococcales; 11, the Methanobacteriales; 12, the Methanomicrobiales; and 13, the extreme halophiles. Eucarya: 14, the animals; 15, the ciliates; 16, the green plants; 17, the fungi; 18, the flagellates; and 19, the microsporidia.

Metagenomics

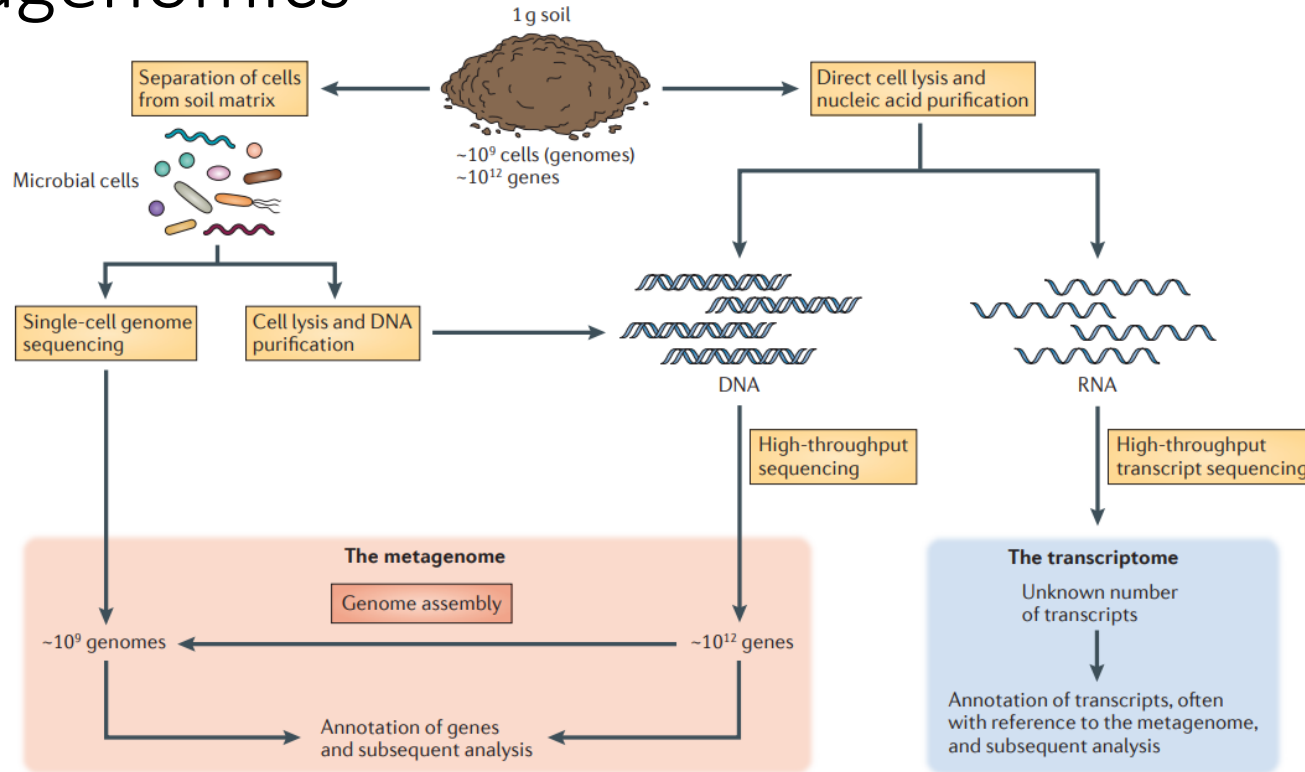
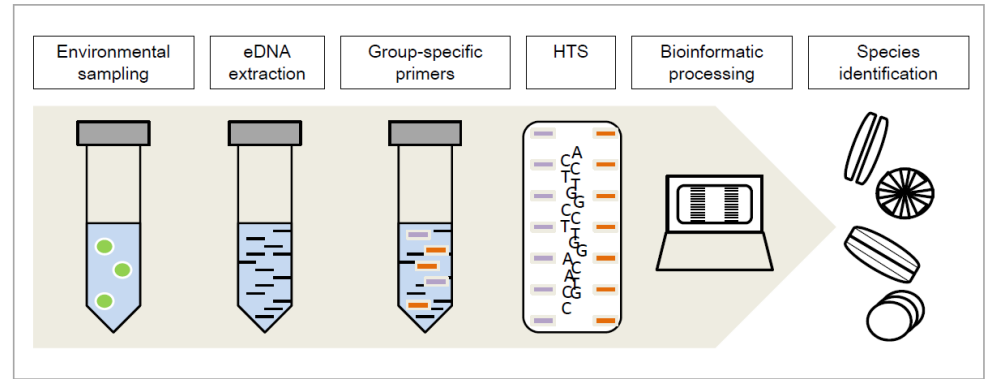
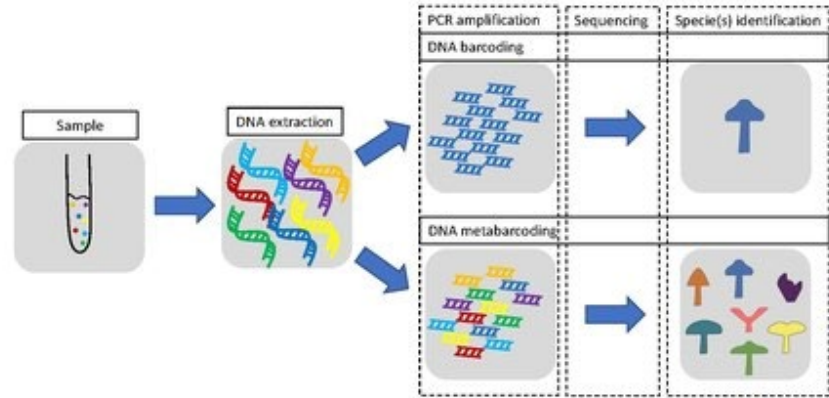


Figure 1 | **Metagenomic and metatranscriptomic analyses of soil samples.** Schematic representation of the main stages involved in generating metagenomic and metatranscriptomic libraries from 1g of soil. The abundances

of microbial cells and genes vary considerably in soil, but they are typically in the order of the amounts indicated here. Notably, the total number of transcripts in 1g of soil at any particular time is difficult to estimate.

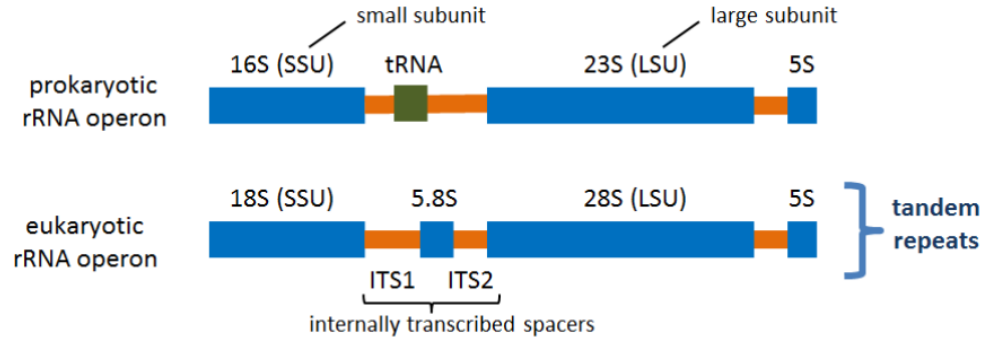
Metabarcoding

- The amplification step uses primers specific for different taxa (*e.g.* plants, fungi, animal)
- Primers are designed on very conserved sequences, for example mitochondrial genes (*e.g.* COI)
- The «full» composition of the samples can be reconstructed downstream



Metabarcoding markers

Different primers
target different taxa

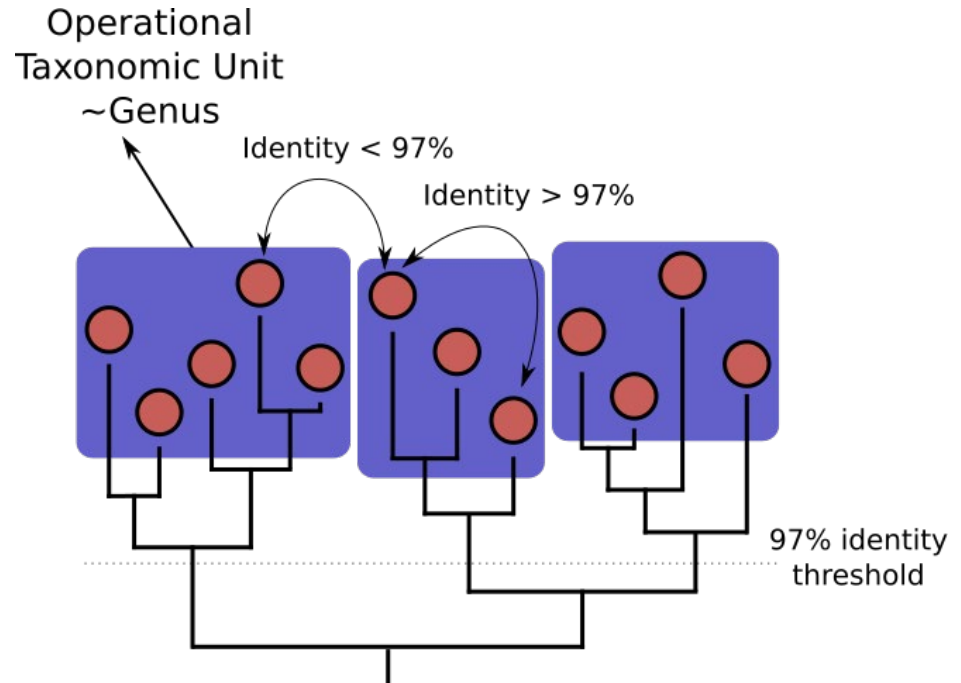


- *E.g.* for vascular plants, plastidial markers *matK* and *rbcl*
- Choice of marker also depends on mutation rate and subsequent taxonomic definition obtainable
- See <https://www.boldsystems.org/>

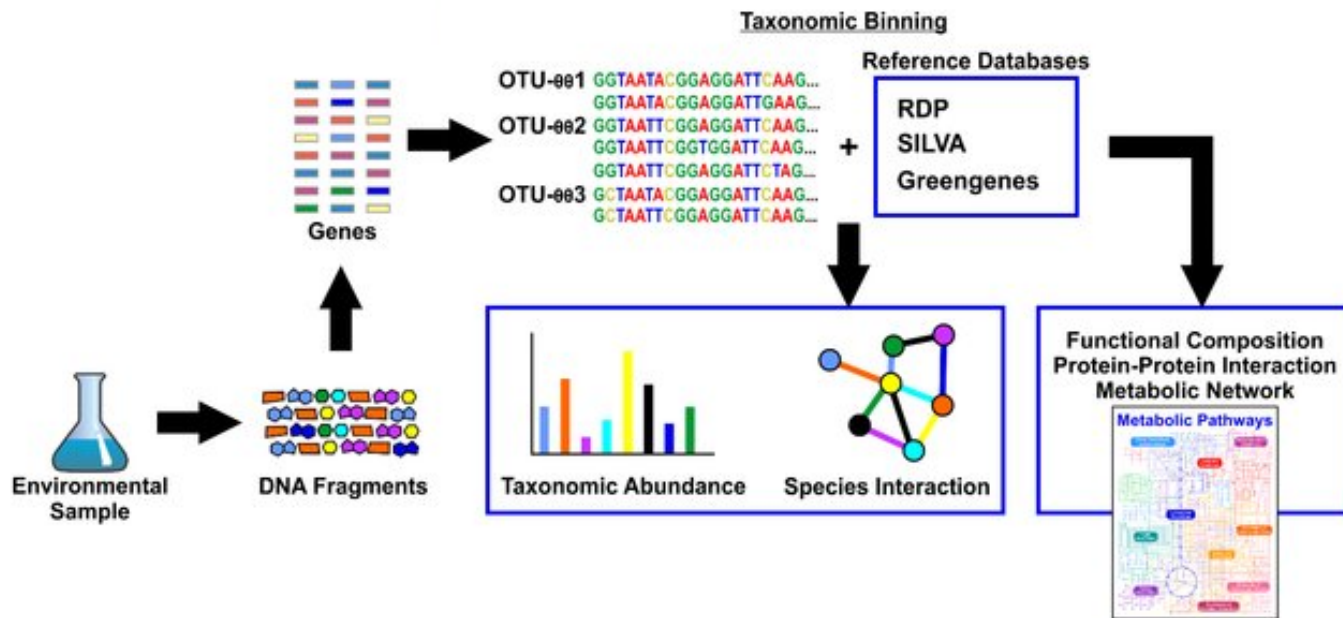
A hypothetical metagenomics analysis pipeline starts from amplification of sampled DNA

- Through bioinformatics, amplicons are assigned to operational taxonomic units (OTU). OTUs each represent one individual taxa, whatever that is

OTUs are devoid of taxonomic assignment, and are defined on the basis of sequence similarity



- Databases can be used to assign taxonomy to OTUs at different levels (species, genus, etc)



- Barcoding and metabarcoding are a step behind metagenomics
- The analogy is with amplification-based molecular markers VS full genome sequencing

(meta) barcoding-based methods

- Targeted amplification
- Cheaper data (less sequences produced)
- Less diversity, no function information
- Possible bias from target selection
- «Simple» dataset

(meta) genomic methods

- Naive method, no prior assumptions
- All sort of sequence in the genome produced
- More sequencing depth needed
- Highly complex dataset

A new view of the tree of life

Laura A. Hug^{1†}, Brett J. Baker², Karthik Anantharaman¹, Christopher T. Brown³, Alexander J. Probst¹, Cindy J. Castelle¹, Cristina N. Butterfield¹, Alex W. Hernsdorf³, Yuki Amano⁴, Kotaro Ise⁴, Yohey Suzuki⁵, Natasha Dudek⁶, David A. Relman^{7,8}, Kari M. Finstad⁹, Ronald Amundson⁹, Brian C. Thomas¹ and Jillian F. Banfield^{1,9*}

The tree of life is one of the most important organizing principles in biology¹. Gene surveys suggest the existence of an enormous number of branches², but even an approximation of the full scale of the tree has remained elusive. Recent depictions of the tree of life have focused either on the nature of deep evolutionary relationships^{3–5} or on the known, well-classified diversity of life with an emphasis on eukaryotes⁶. These approaches overlook the dramatic change in our understanding of life's diversity resulting from genomic sampling of previously unexamined environments. New methods to generate genome sequences illuminate the identity of organisms and their metabolic capacities, placing them in community and ecosystem contexts^{7,8}. Here, we use new genomic data from over 1,000 uncultivated and little known organisms, together with published sequences, to infer a dramatically expanded version of the tree of life, with Bacteria, Archaea and Eukarya included. The depiction is both a global overview and a snapshot of the diversity within each major lineage. The results reveal the dominance of bacterial diversification and underline the importance of organisms lacking isolated representatives, with substantial evolution concentrated in a major radiation of such organisms. This tree highlights major lineages currently underrepresented in biogeochemical models and identifies radiations that are probably important for future evolutionary analyses.

Metagenomics unlocks new perspectives on life on Earth

- Newly sequenced 1000+ organisms, mostly uncultured bacteria sampled in shallow aquifer systems, a deep subsurface research site in Japan, a salt crust in the Atacama Desert, grassland meadow soil in northern California, a CO₂-rich geyser system, and two dolphin mouths
- 3,083 organisms in total
- Extraction of rDNA sequences and comparison across genomes

A more accurate tree of life

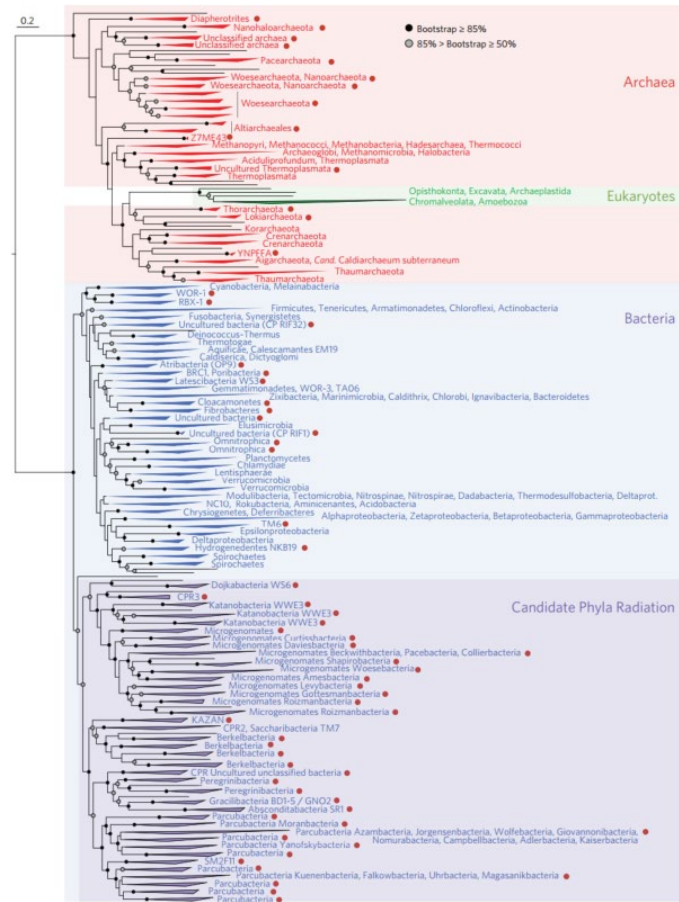
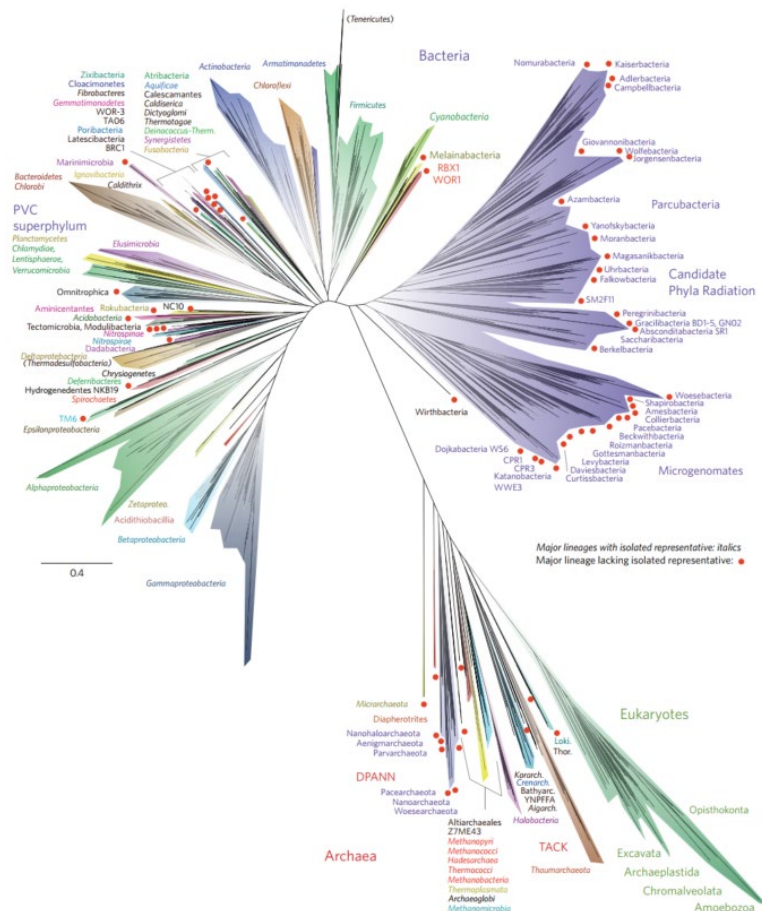
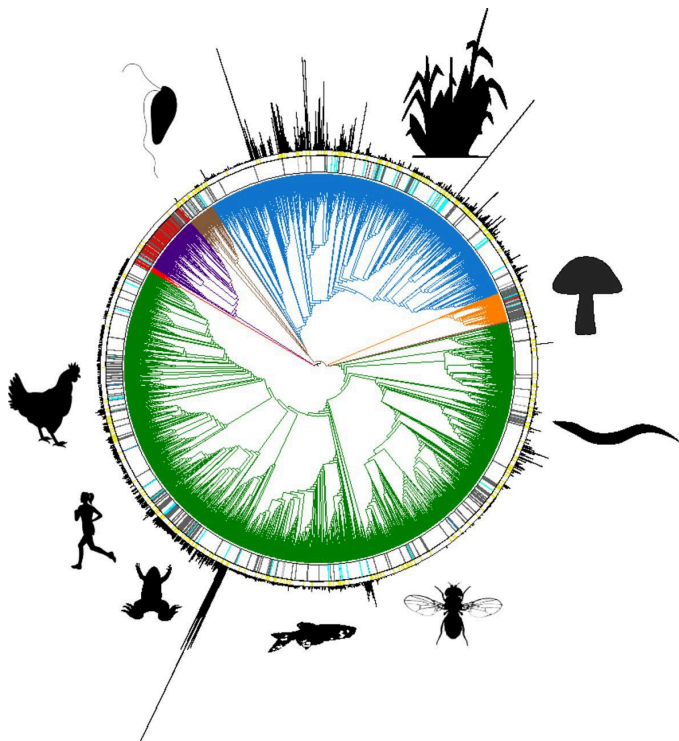




Figure 2 | A reformatted view of the tree in Fig. 1 in which each major lineage represents the same amount of evolutionary distance. The threshold for groups (coloured wedges) was an average branch length of <0.65 substitutions per site. Notably, some well-accepted phyla become single groups and others are split into multiple distinct groups. We undertook this analysis to provide perspective on the structure of the tree, and do not propose the resulting groups to have special taxonomic status. The massive scale of diversity in the CPR and the large fraction of major lineages that lack isolated representatives (red dots) are apparent from this analysis. Bootstrap support values are indicated by circles on nodes—black for support of 85% and above, grey for support from 50 to 84% . The complete ribosomal protein tree is available in rectangular format with full bootstrap values as Supplementary Fig. 1 and in Newick format in Supplementary Dataset 2.

Beyond barcoding



Earth BioGenome Project: Sequencing life for the future of life

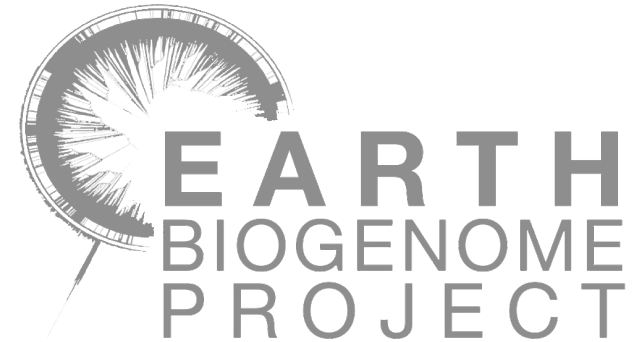
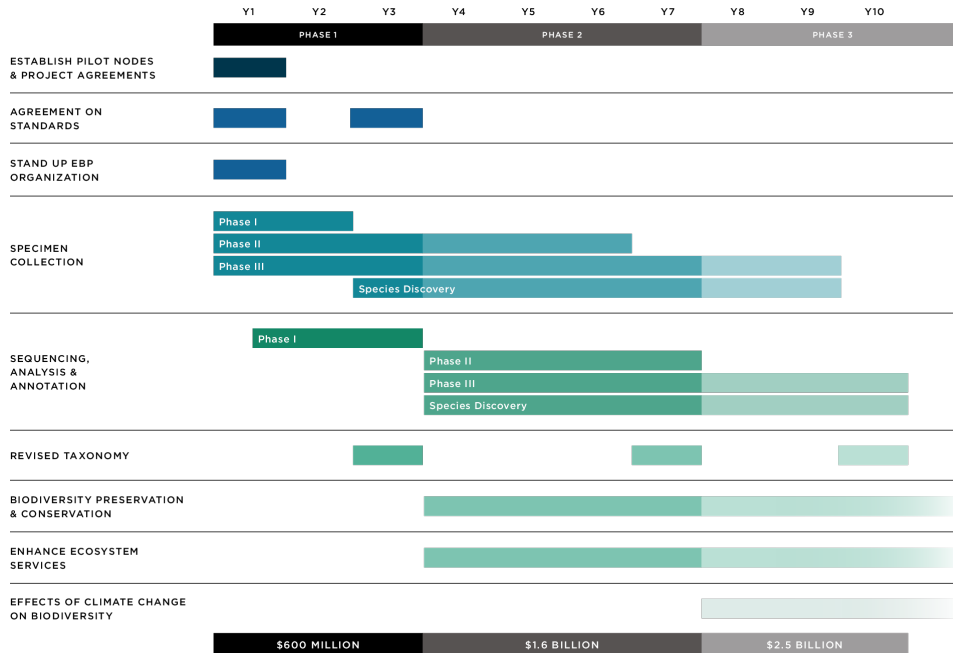
Harris A. Lewin, Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington,  Keith A. Crandall, Richard Durbin,  Scott V. Edwards, Félix Forest, M. Thomas P. Gilbert, Melissa M. Goldstein, Igor V. Grigoriev, Kevin J. Hackett, David Haussler, Erich D. Jarvis, Warren E. Johnson, Aristides Patrinos, Stephen Richards, Juan Carlos Castilla-Rubio, Marie-Anne van Sluys, Pamela S. Soltis, Xun Xu, Huanming Yang, and Guojie Zhang

PNAS April 24, 2018 115 (17) 4325–4333; first published April 23, 2018; <https://doi.org/10.1073/pnas.1720115115>

Edited by John C. Avise, University of California, Irvine, CA, and approved March 15, 2018 (received for review January 6, 2018)

- Started 2018
- Sequence EVERYTHING!
- Create annotated chromosome-scale reference assemblies for at least one representative species of each of the ~9,000 eukaryotic taxonomic families

Sequencing and annotating ~1.5 million known eukaryotic species in three phases over a 10-year period using a phylogenomic approach.



<https://www.earthbiogenome.org/>

Nothing in Biology Makes Sense Except in the Light of Evolution

- We owe everything, including the formation of our own species, to evolution
- History of life is driven by evolution
- All biodiversity, as made of living beings, comes from the past and is projected into the future through evolutionary change
- Biodiversity as a whole is object and propeller of evolution
- Contemporary evolution is a key to understand reality and guide applied research including conservation, breeding, and medicine



Theodosius Grygorovych Dobzhansky

The Origin of Species - 1859

"As more individuals ...are born than can possibly survive, and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in a manner profitable to itself, will have a better chance of surviving, and thus be naturally selected"

ON
THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

By CHARLES DARWIN, M.A.,
FELLOW OF THE ROYAL, GEOLOGICAL, LINNEAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE
ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

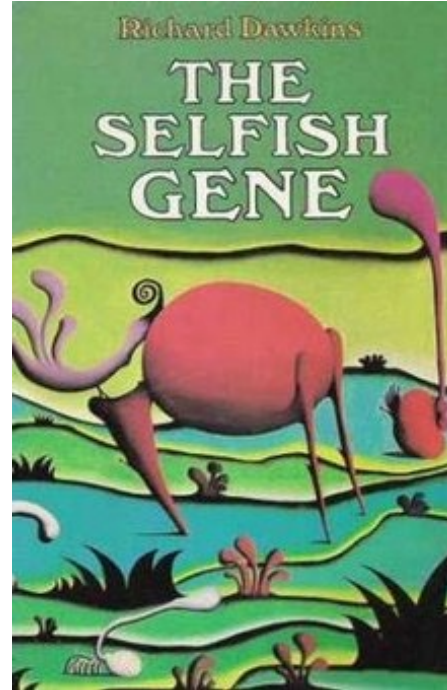
Key ideas of evolution by natural selection

- Many more individuals are born than survive (**COMPETITION**)
- Individuals vary in traits directly related to their ability to survive and reproduce (**VARIATION**)
- Advantageous traits are passed on to offspring (**HERITABILITY**)
- This process is repeated generation after generation over long periods of time (**ITERATION**).

Evolution – a modern view

Richard Dawkins

- Self-replicating hereditary mechanisms are at the centre of evolution
- Meme, shortening of *mimeme* (from Ancient Greek "to imitate")
- meme as a “cultural gene”



Richard Dawkins when he created the word "meme" in 1976



Chicken is an Egg's way of making another egg

- You are a vehicle for your genes to be transferred from one generation to another
- Variants in the same gene differ from their efficiency in running the vehicle, resulting in different likelihood to be inherited
- Variants compete for ability to programme the vehicle to help them to be transferred in as many copies as possible
 - Darwin: how does this trait increase fitness of an individual?
 - Dawkins: how does this trait increase a number of copies of variant of gene that is responsible for it?

Biological evolution pivots on DNA

- In biological evolution “the change” is not a sufficient condition for evolution
- The reason of change is not important
- **Change should be inherited**

When we talk about biological evolution:

- We refer to populations
- We observe changes of frequencies of features from generation to generation (e.g. resistance to antibiotics, color etc)
- Feature should be inherited from generation to generation

