INSTITUTE OF PLANT SCIENCES

Sant'Anna
School of Advanced Studies – Pisa

Environmental DNA typing

Breedtech workshop

September 2025

Sara Verni

(Capa and Hutcings, 2021)

Me: from the deep-sea … to the deep-field

But always in good company!

environmental DNA (eDNA)

Outline

- **Biodiversity**: what we know and why does it matter
- **Agrobiodiversity** and why it is essential for ecosystems services
- Barcoding
- environmental DNA
- Metabarcoding
- Metagenomics
- To the future and beyond: **portable** genomics in the field!

- «The variety and variability of living organisms and the ecological systems that include them»

(Office of Technology Assessment, US Congress, 1987).

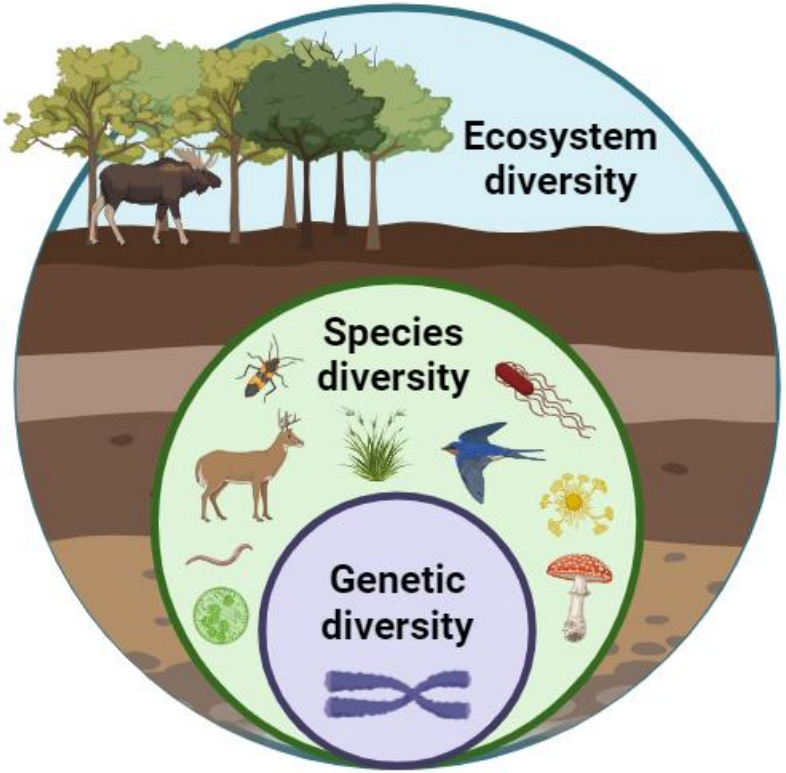
- «Variability of all kinds among living organisms»

(IUCN, The World Conservation Union, 1994).

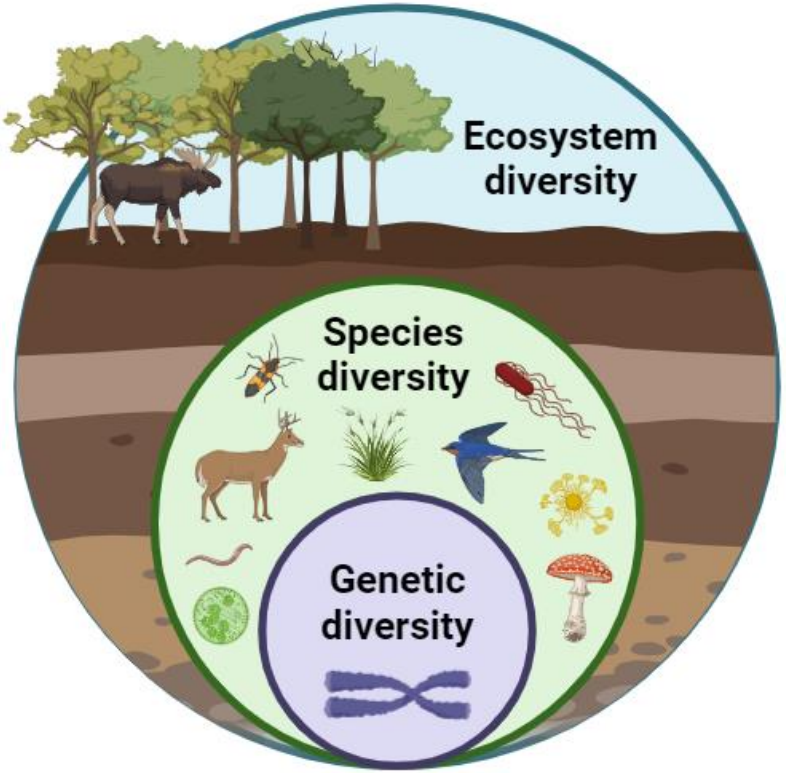
- «The total variety of life on Earth»

(Takacs, 1996).

- **Biodiversity** refers to the variety of all life on Earth.
- Exists at different levels, from ecosystems to species to genes.



- **Biodiversity** refers to the variety of all life on Earth.
- Exists at different levels, from ecosystems to species to genes ⟹ *functional diversity*
- Genetic diversity provides the potential for a species to adapt to environmental changes.



Why is the study of biodiversity **important**?

Why is the study of biodiversity **important**?

Intrinsic value

Human beings are an integral part of nature and the environment.

Anthropocentric value

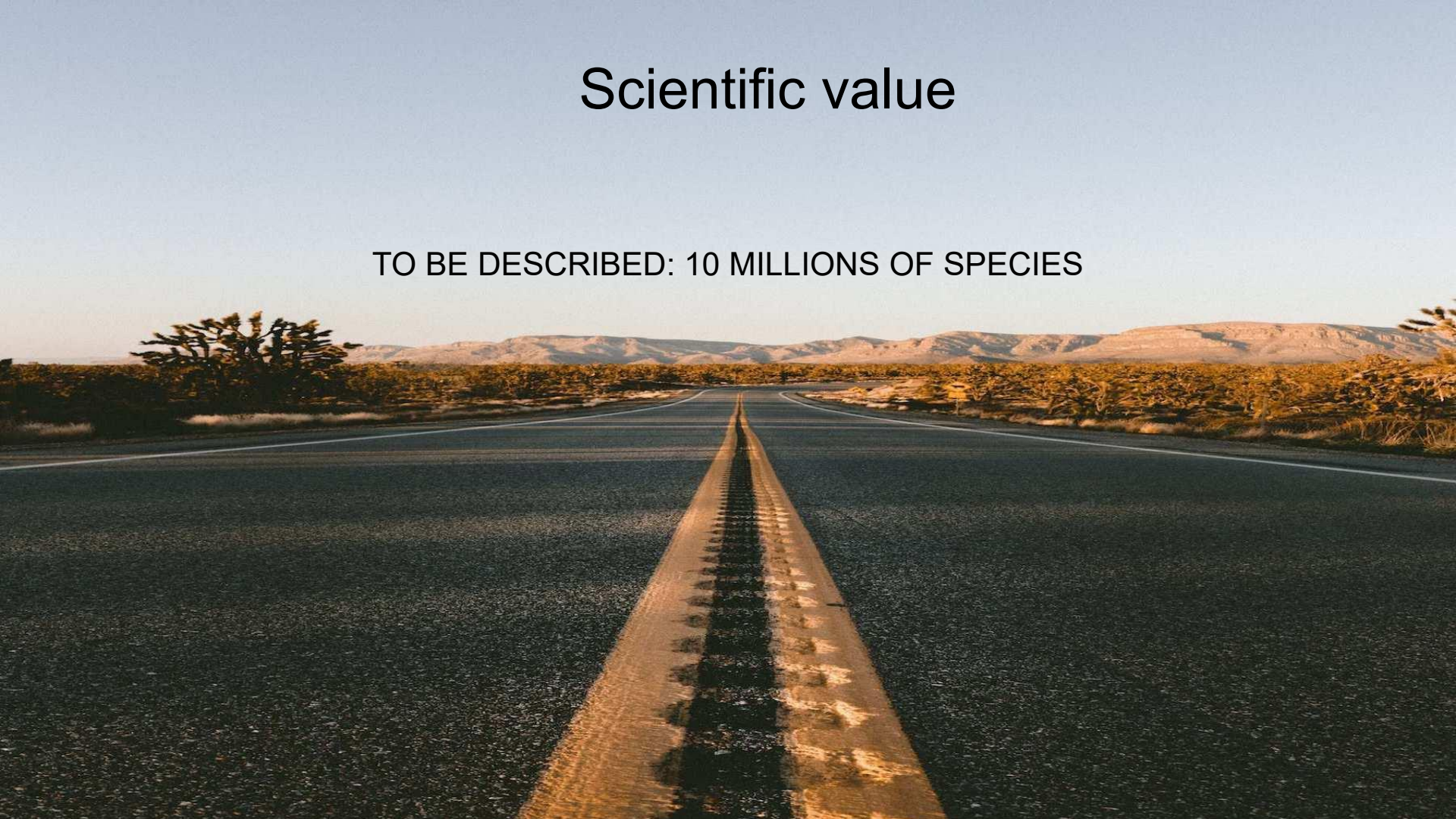
The direct and indirect economic benefits derived from it.

Aesthetic and **recreational** value

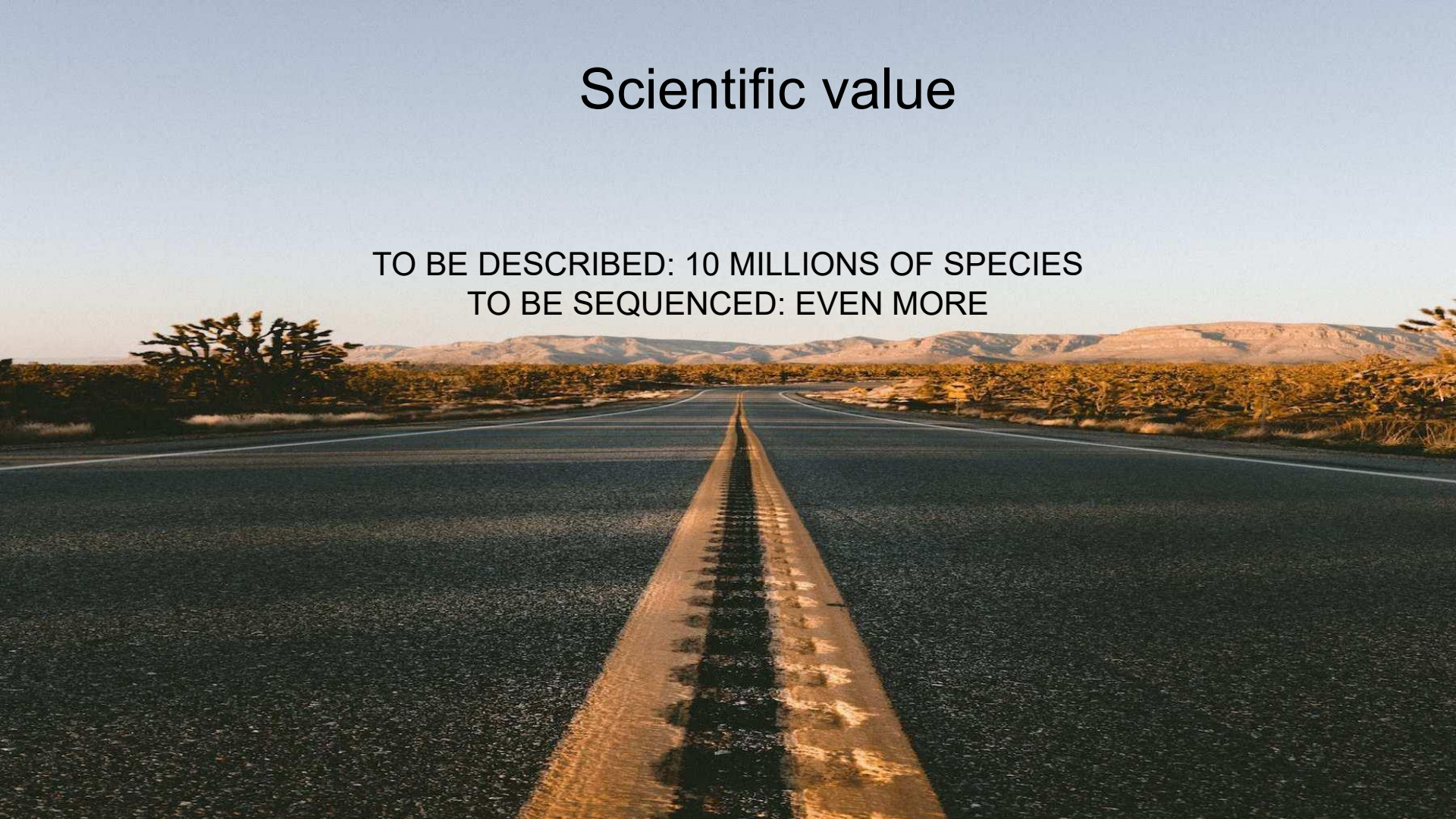
Tourism, sport, contemplation.

Scientific value

TO BE DESCRIBED: 10 MILLIONS OF SPECIES



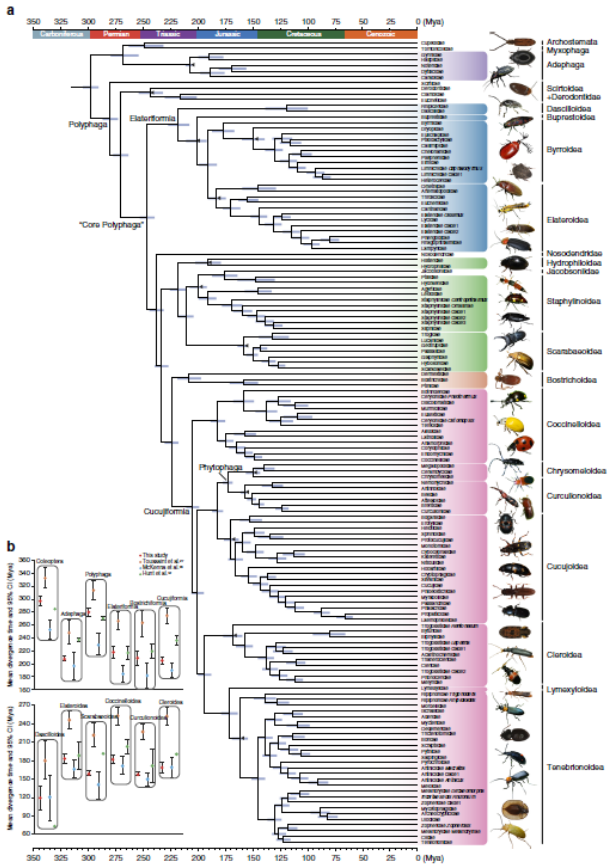
Scientific value



TO BE DESCRIBED: 10 MILLIONS OF SPECIES
TO BE SEQUENCED: EVEN MORE

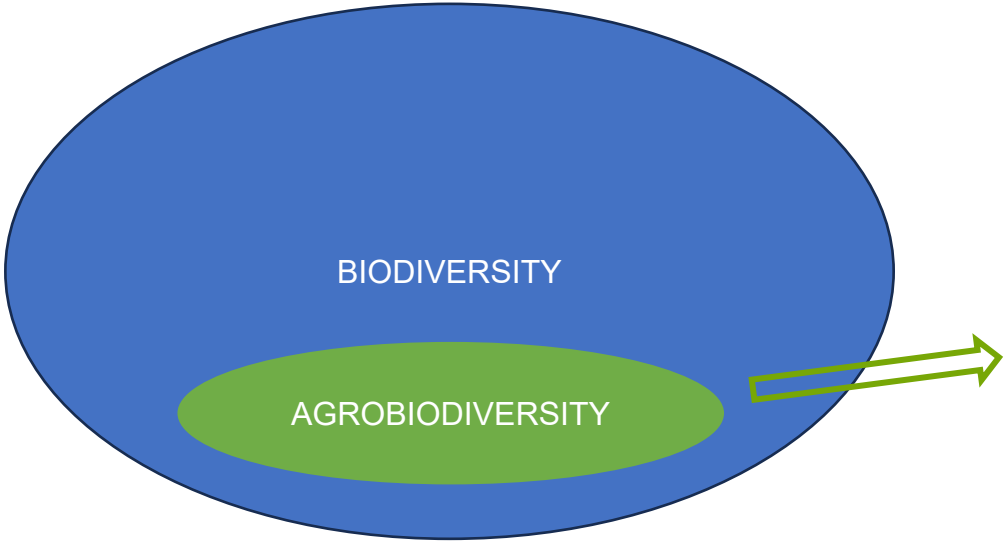
“The Creator must have an inordinate
fondness for beetles”

J.B.S. Haldane



Environmental DNA typing

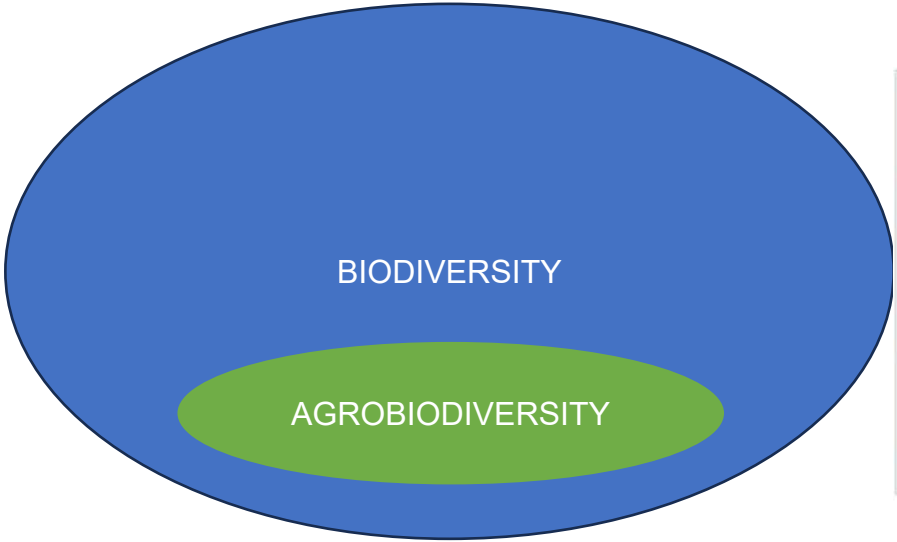
AGROBIODIVERSITY IS CENTRAL TO OVERALL BIODIVERSITY

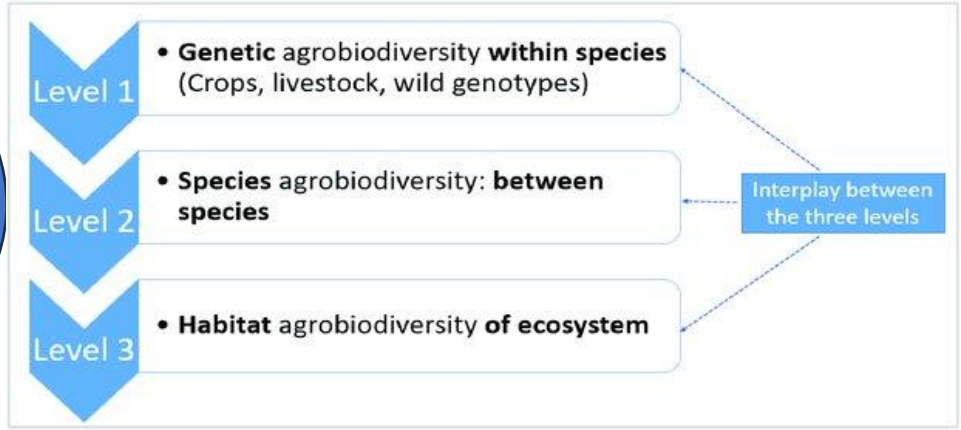


- Mixed agro-ecosystems
- Crop species/varieties
- Livestock and fish species
- Plant/animal germoplasm
- Soil organisms in cultivated areas
- Biocontrol agents for pests
- Wild species
- **Cultural and local knowledge**

Agrobiodiversity is the result of natural selection processes and the careful selection and inventive developments of farmers, herders and fishers over millennia.

LEVEL OF AGROBIODIVERSITY





Agrobiodiversity is the result of natural selection processes and the careful selection and inventive developments of farmers, herders and fishers over millennia.

Who lives in a field?

Weeds / spontaneous herbs
Perennial plants at field margins

Taraxacum officinale

Papaver rhoeas

Chenopodium album

Who lives in a field?

Pollinators
Other insects (enemies of pests and pests)
Birds
Small mammals
Amphibians and reptiles

Coccinella septempunctata

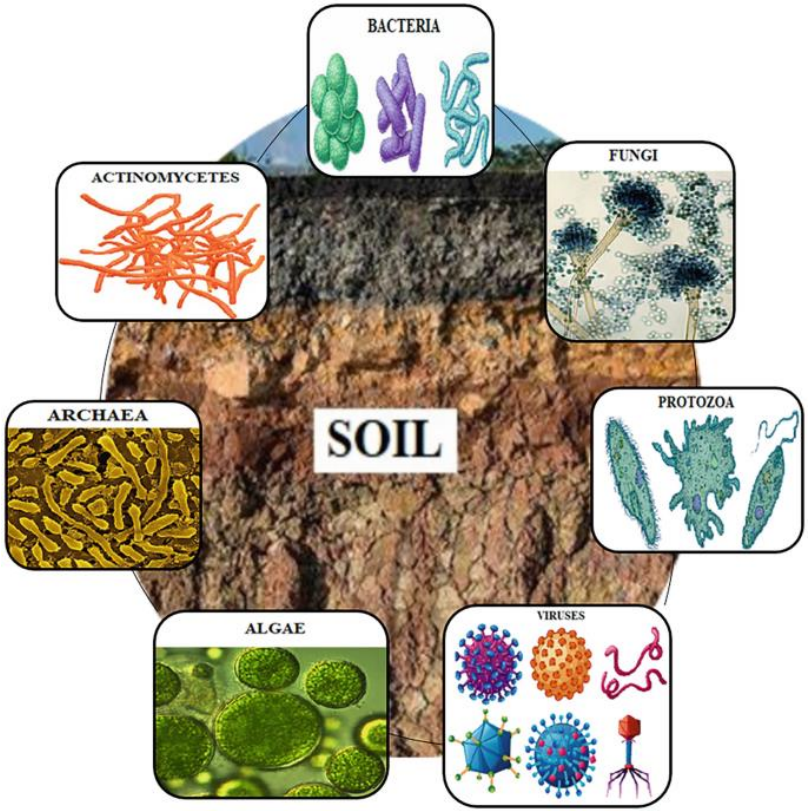
Bombus spp.

Podarcis muralis

Sturnus vulgaris

Microtus arvalis

Who lives in a field?



Ecosystem services of agrobiodiversity

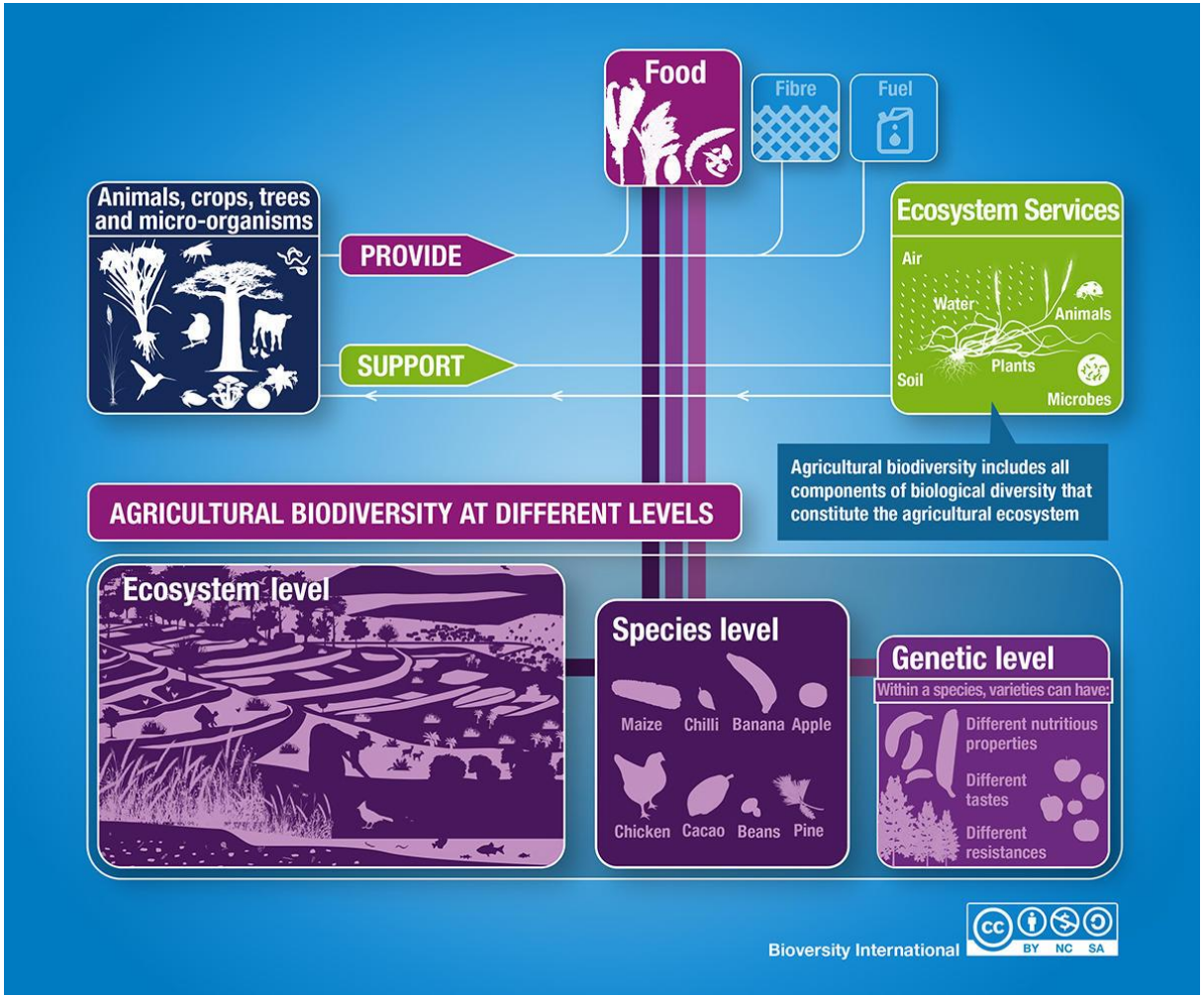
a property or process in an ecosystem that confers either direct or indirect benefits to humans.

These renewal processes and ecosystem services are largely biological

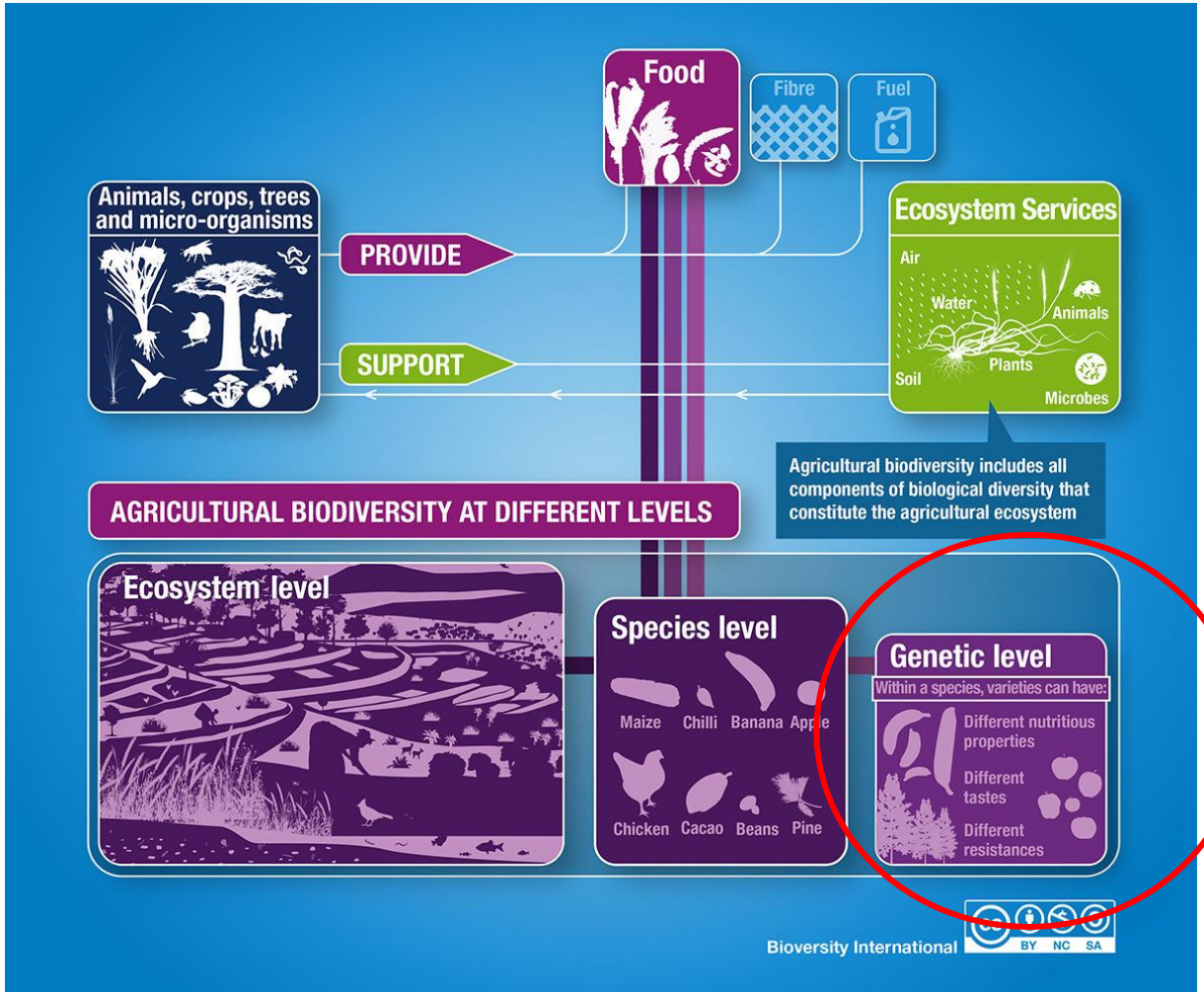
their persistence depends upon maintenance of biological diversity (Altieri, 1994)

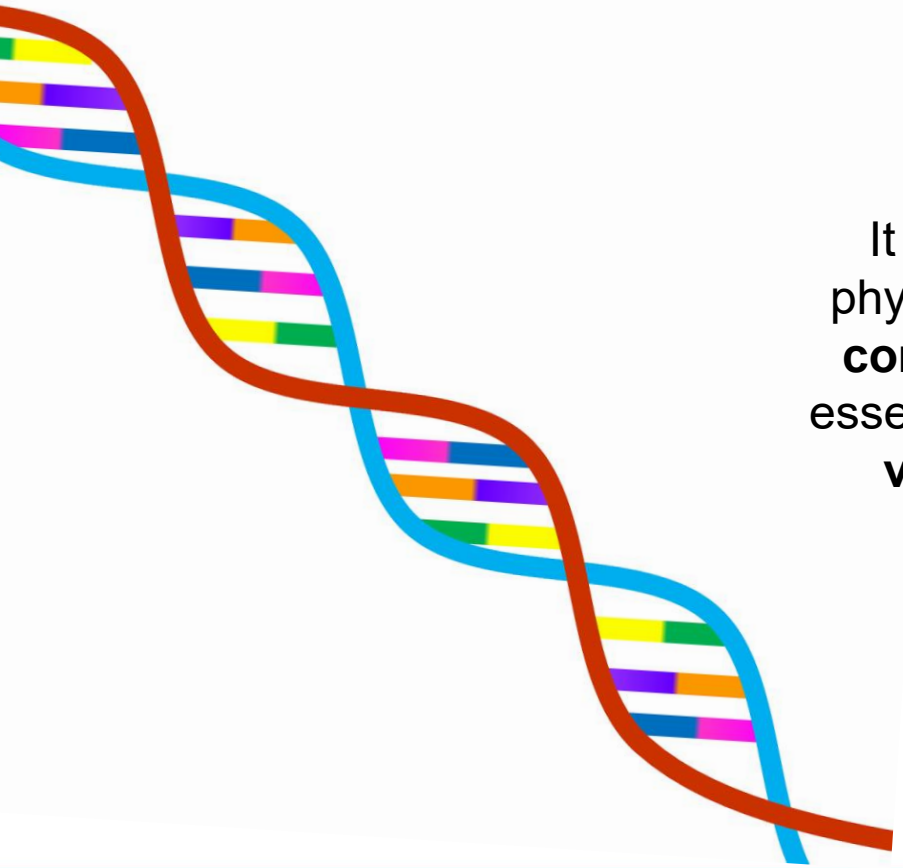
When these natural services are lost due to biological simplification, the economic and environmental costs can be quite significant

To summarize



To summarize

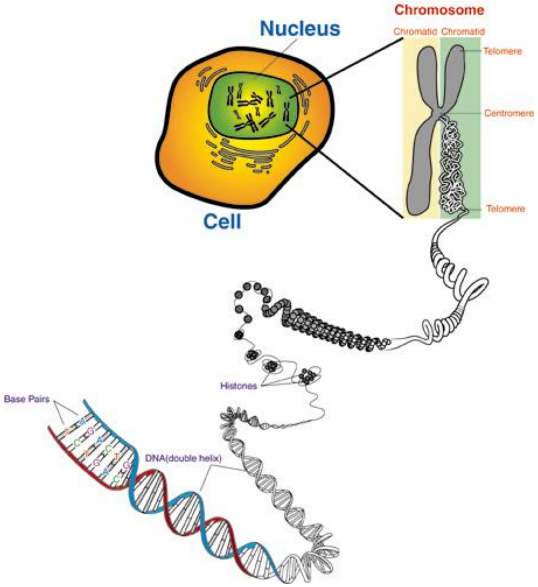




Genetic diversity

It enables an individual to interact with their physical and biological environment, as well as **compete** both within and between species. In essence, it enables **tolerance of environmental variations and survival through natural selection.**

The key to heredity: DNA



Genome: the whole GENETIC MATERIAL - DNA - In a cell, in an organism, or in a species

It is the **POTENTIAL** In terms of **functionality** of organisms

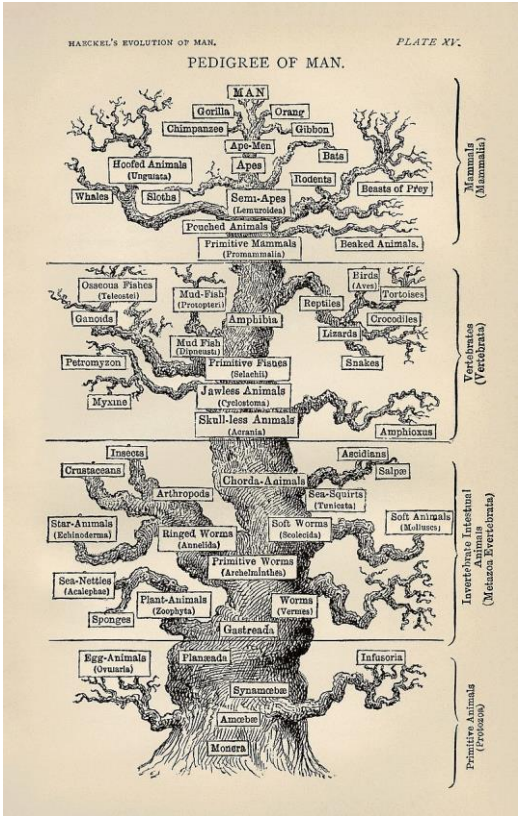
How do we measure genetic biodiversity?

DNA-based phylogenies: genetic diversity is assessed by reconstructing evolutionary trees from DNA or genomic data.

Why it matters: phylogenetic approaches capture not only the number of species, but also the **depth of evolutionary history** they represent.

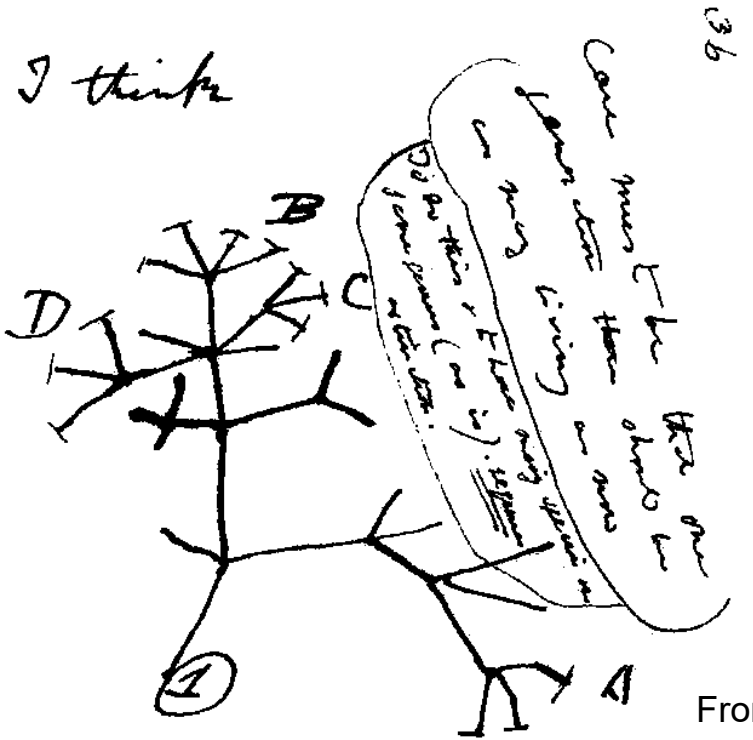
Genomics and evolution

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



Phylogenetic reconstruction

Reconstructing relationships between taxa in the form of an evolutionary tree



From Darwin notebook

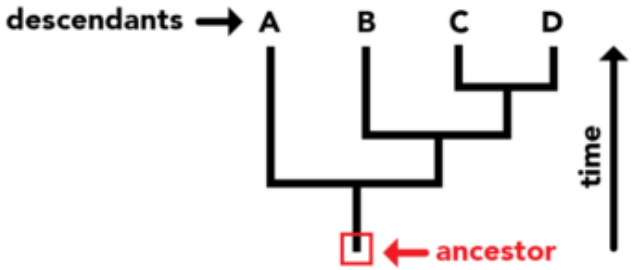
Study of characters

Any characteristic used by to highlight variations within and between species.

Character status

The various ways in which a character can manifest itself ⇨ it allows the characters to be coded and analysed

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

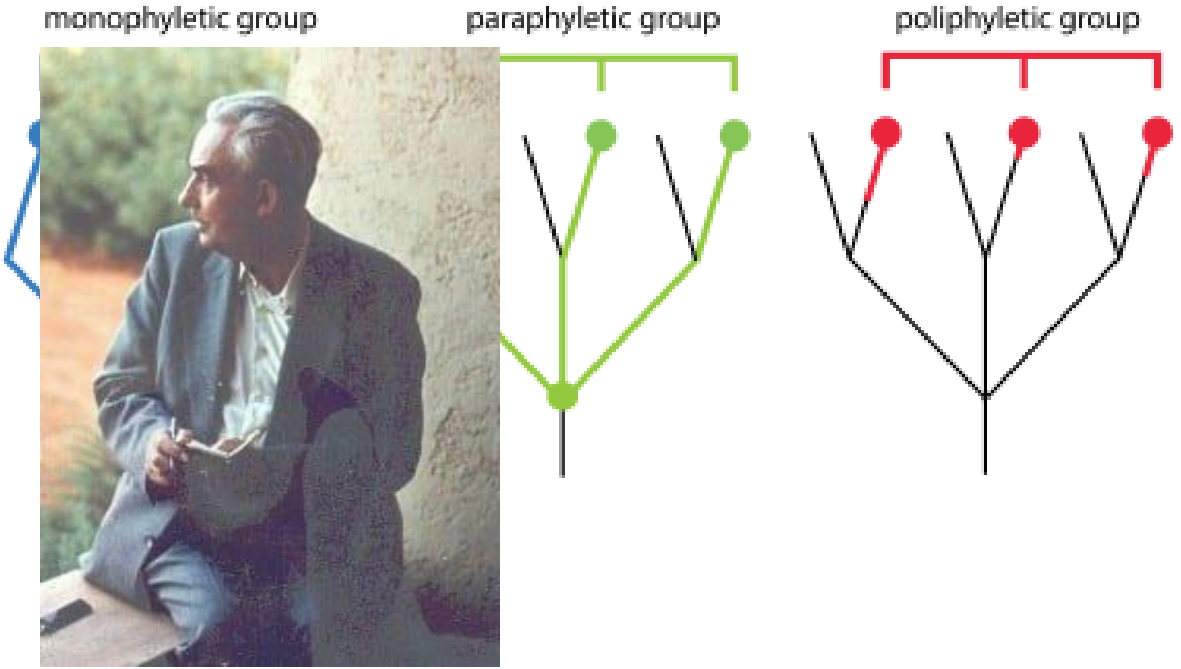
When a speciation event occurs, a single ancestral lineage gives rise to two or more daughter lineages

Species concept

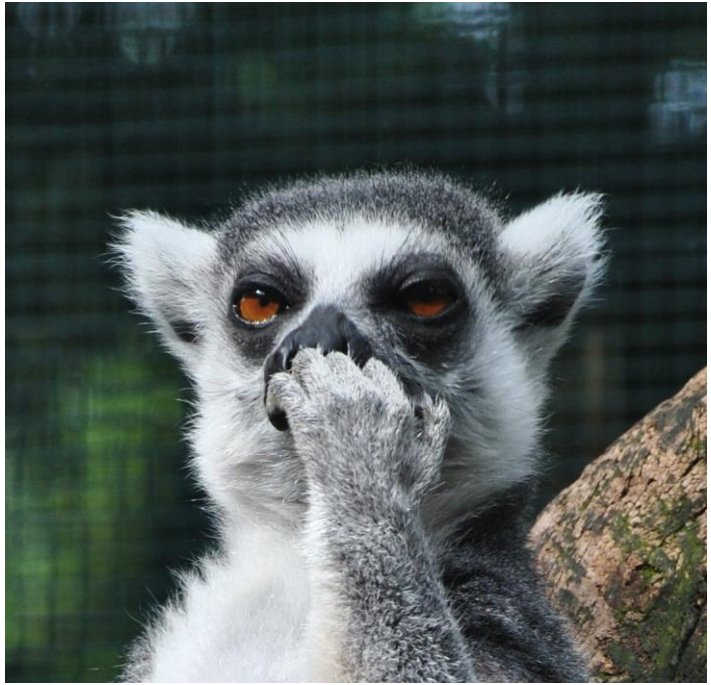
Phylogenetic concept

An irreducible (fundamental) grouping of organisms, diagnostically distinct from other similar groups and within which there is an ancestor-descendant parental model.

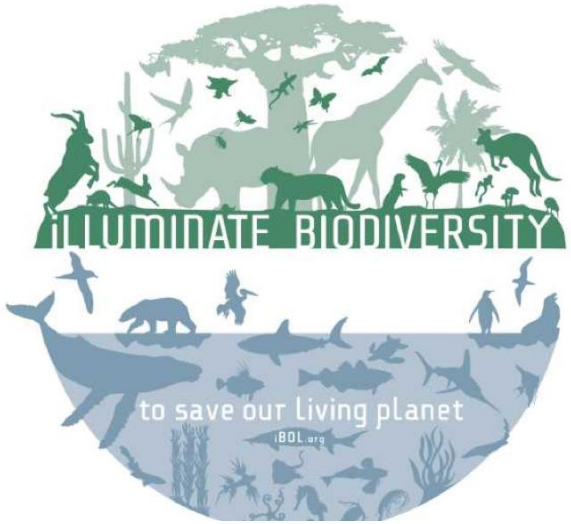
Willi Henning
(1913-1976)
Zoologo
Leipzig University



HOW DO WE RECONSTRUCT PHYLOGENY AND THEREFORE BIODIVERSITY?



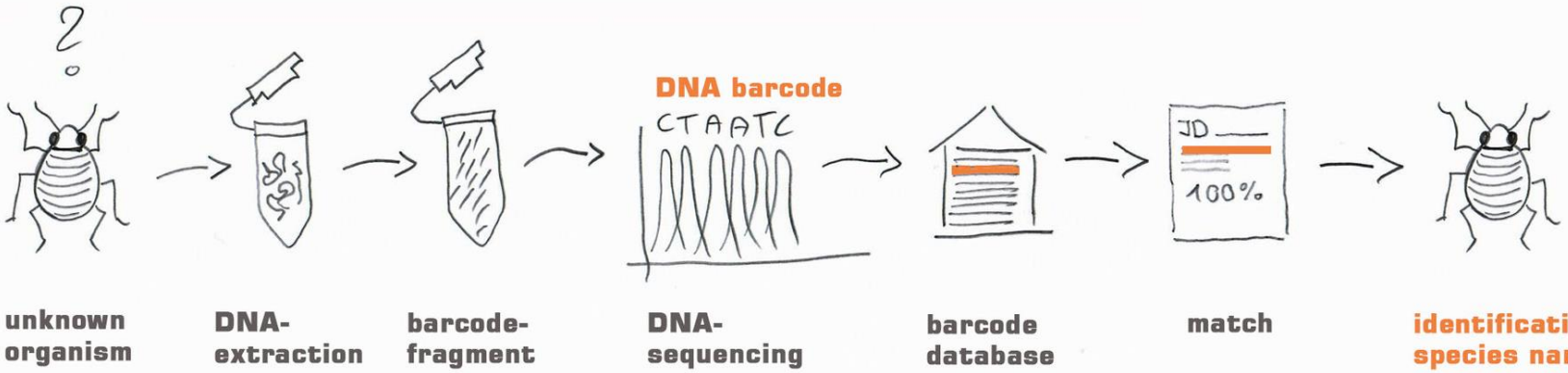
Barcoding

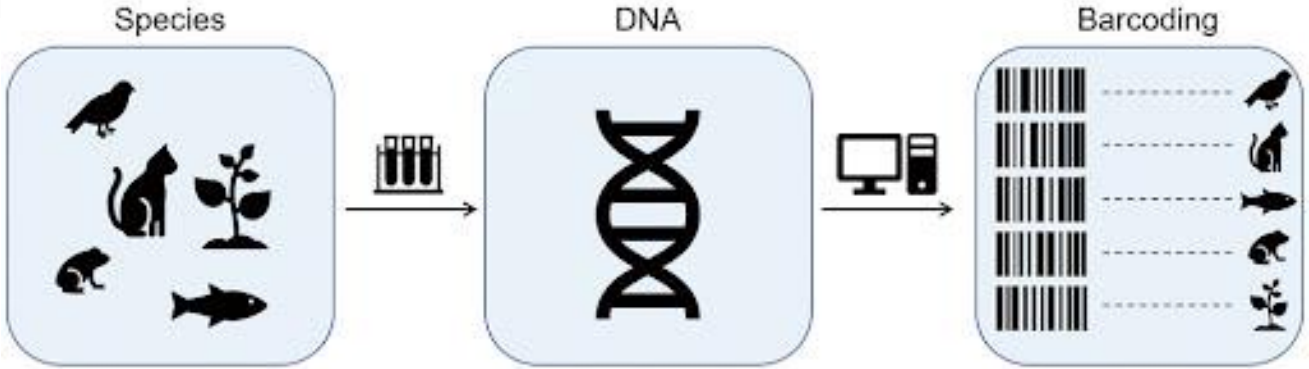


A technique, proposed in 2003 by Canadian researcher Paul D. N. Herbert:

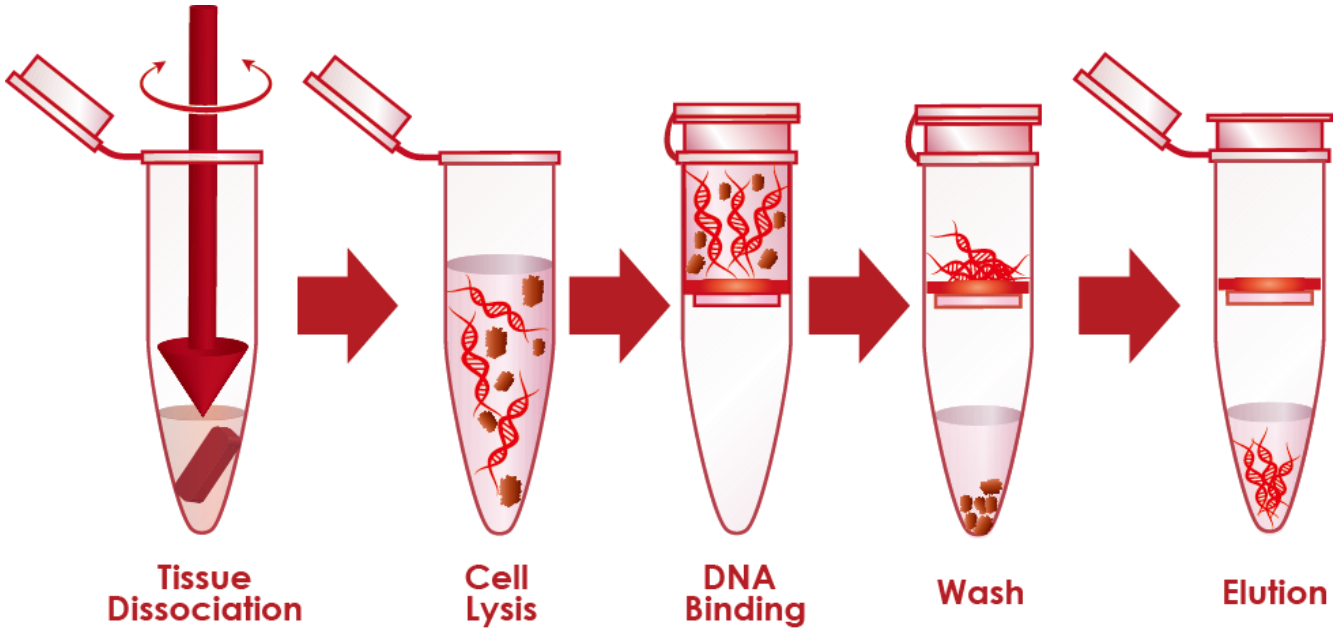
a specific DNA sequence becomes the key to uniquely identifying the species to which an organism belongs

How does DNA Barcoding work ?

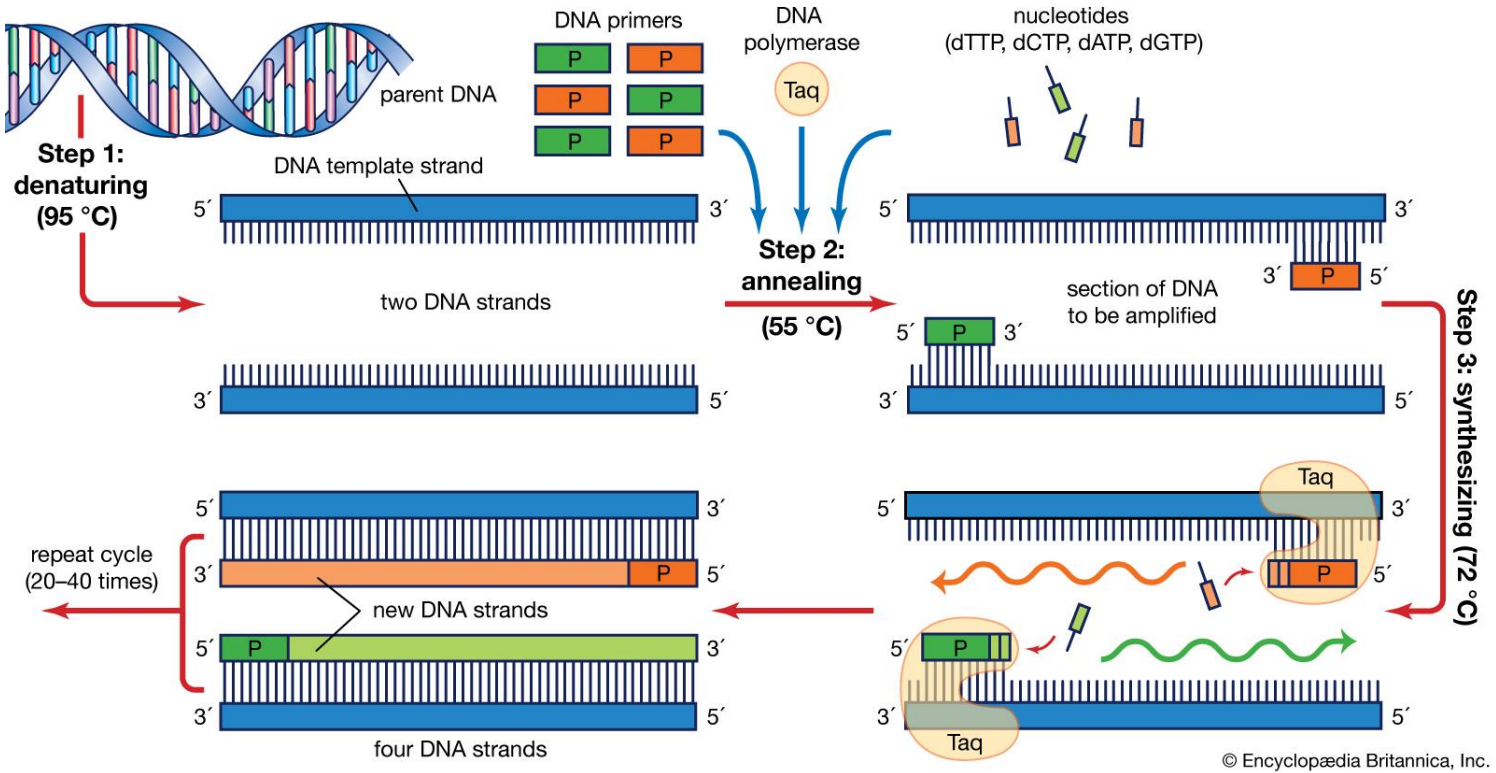


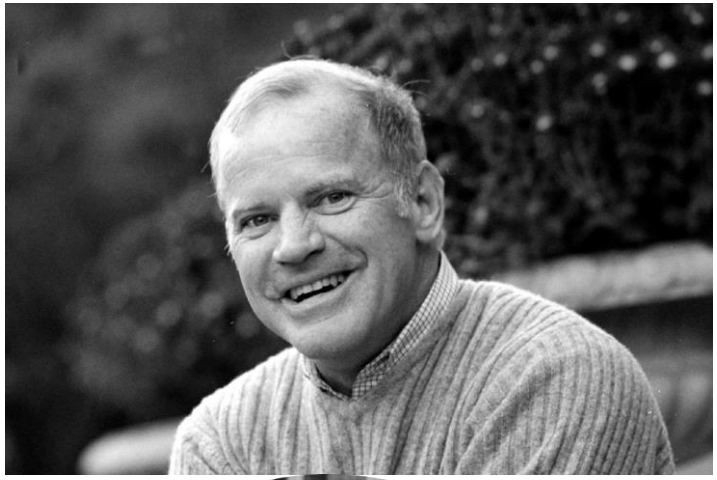


DNA extractions



PCR: POLIMERASE CHAIN REACTION





Kary Mullis (1944-2019)

Inventor of PCR
Nobel Prize in Chemistry 1993

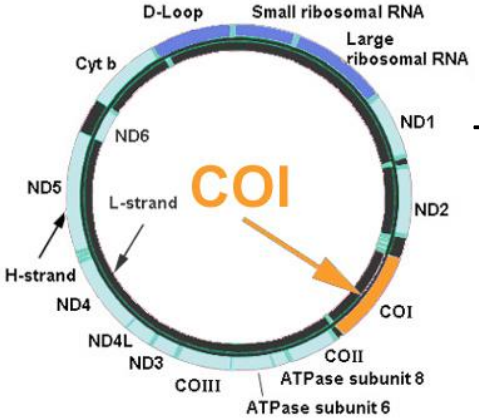
Micheal Smith (1932-2000)

Inventor of PCR
Nobel Prize in Chemistry 1993

The number of mutations (i.e., genetic distance) can be used as an indicator of the time elapsed since the separation of two generations.

Barcoding markers

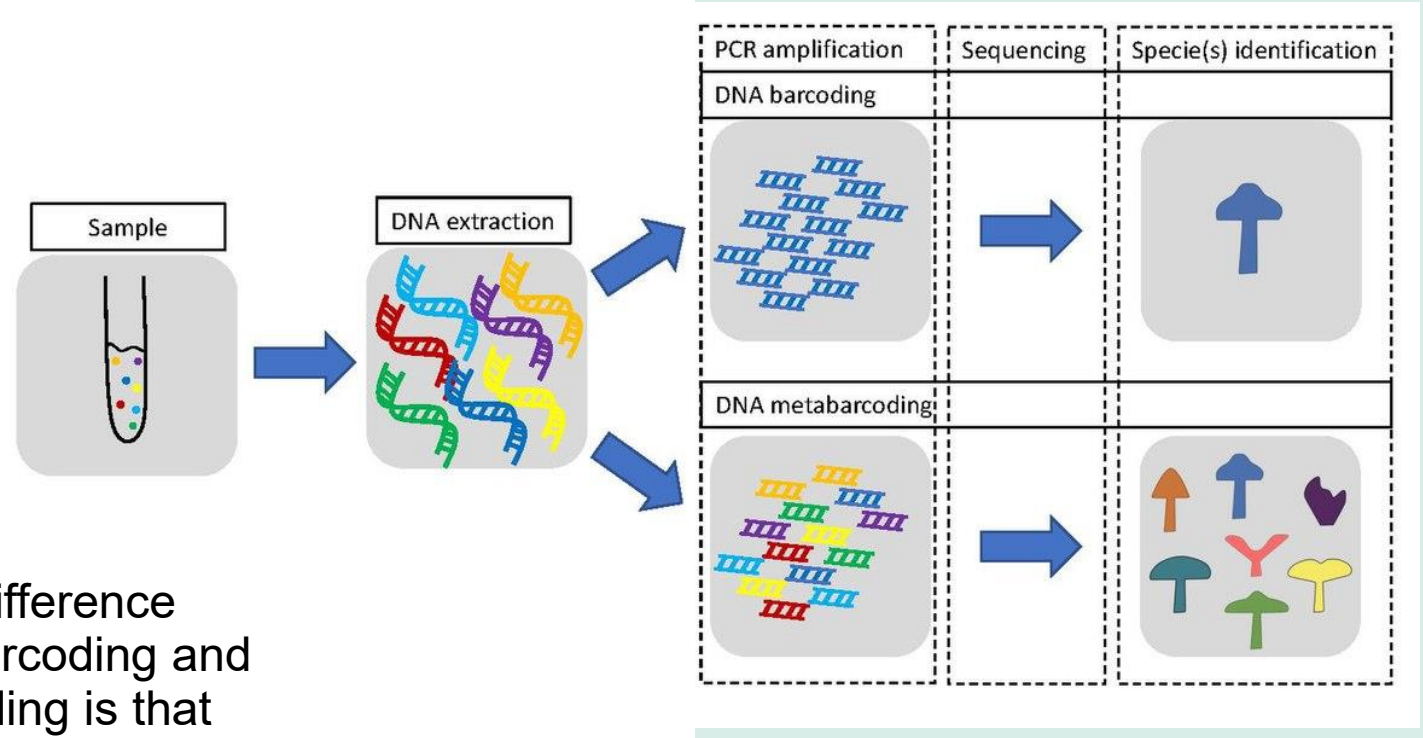
precise **interspecific** variability, but **low variability within individuals** of the same species.



The mitochondrial gene CO1 (cytochrome c oxidase 1)

648 bp

Hebert et al. (2003) suggested a threshold of 3% for COI to separate different species.

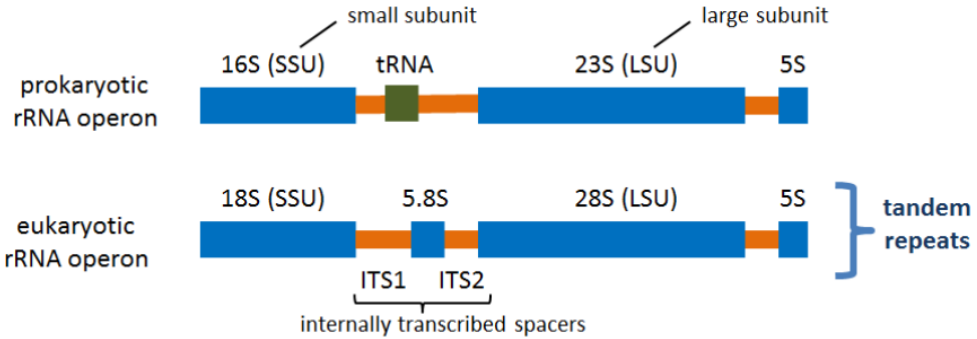


The main difference between barcoding and metabarcoding is that metabarcoding does not focus on a specific organism, but **aims to determine the species composition within a sample.**

Metabarcoding

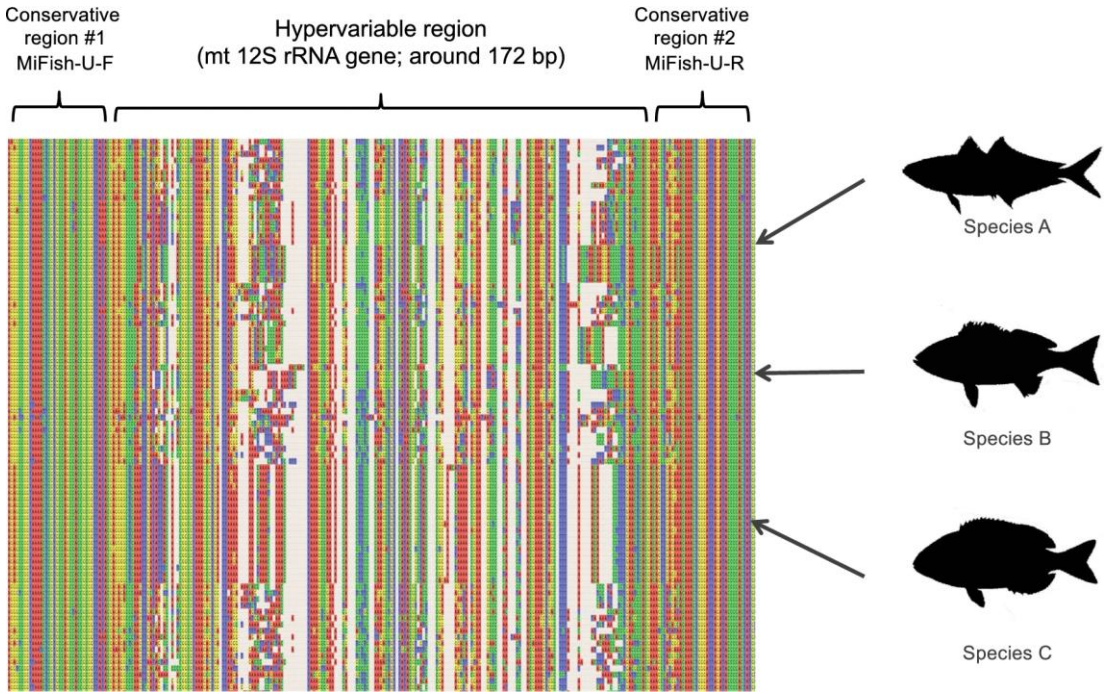
Metabarcoding markers

Different primers target different taxa

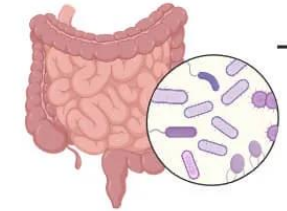


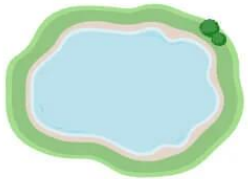
- *E.g.* for vascular plants, plastidial markers matK and rbcL
- Choice of marker also depends on mutation rate and subsequent taxonomic definition obtainable
- See https://www.boldsystems.org/

Metabarcoding markers

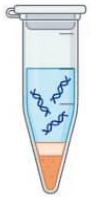


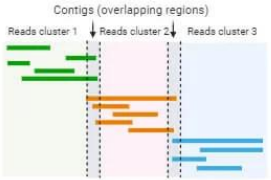
Metagenomics





Sample





DNA → DNA library → DNA sequencing → Sequence analysis and reconstruction

Tools, Softwares, Analysis ← Assembled genome sequence

Metagenomics

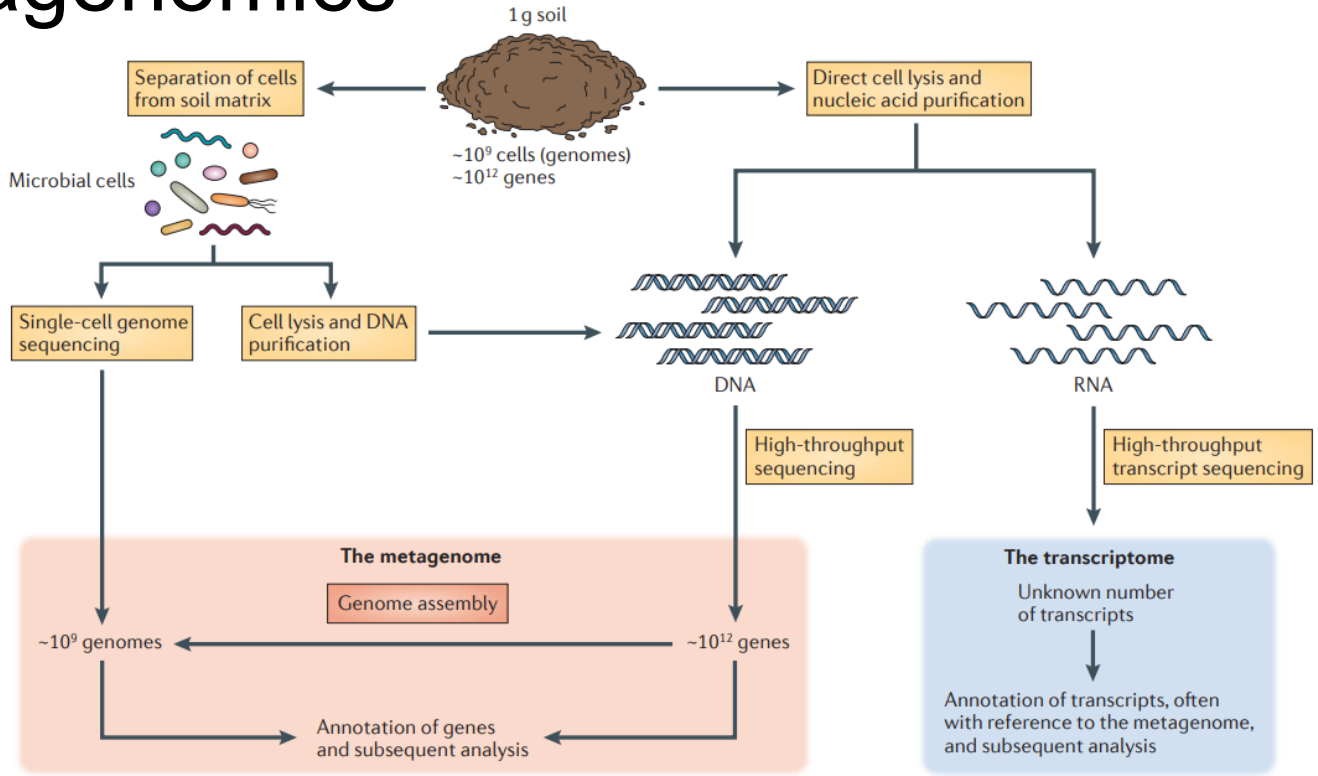


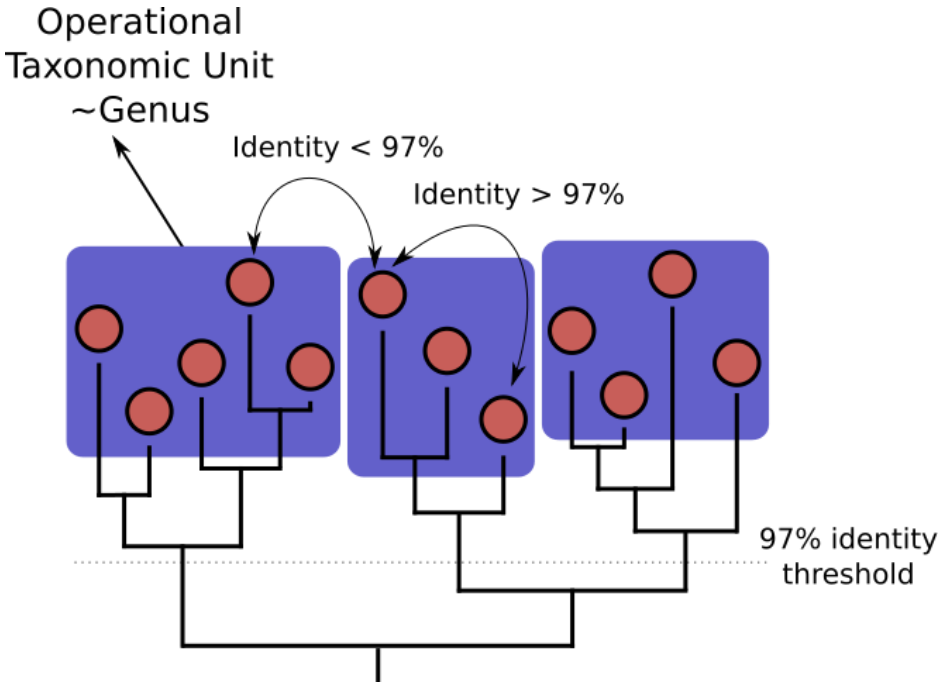
Figure 1 | **Metagenomic and metatranscriptomic analyses of soil samples.** Schematic representation of the main stages involved in generating metagenomic and metatranscriptomic libraries from 1 g of soil. The abundances of microbial cells and genes vary considerably in soil, but they are typically in the order of the amounts indicated here. Notably, the total number of transcripts in 1 g of soil at any particular time is difficult to estimate.

[Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology (agrilife.org)]

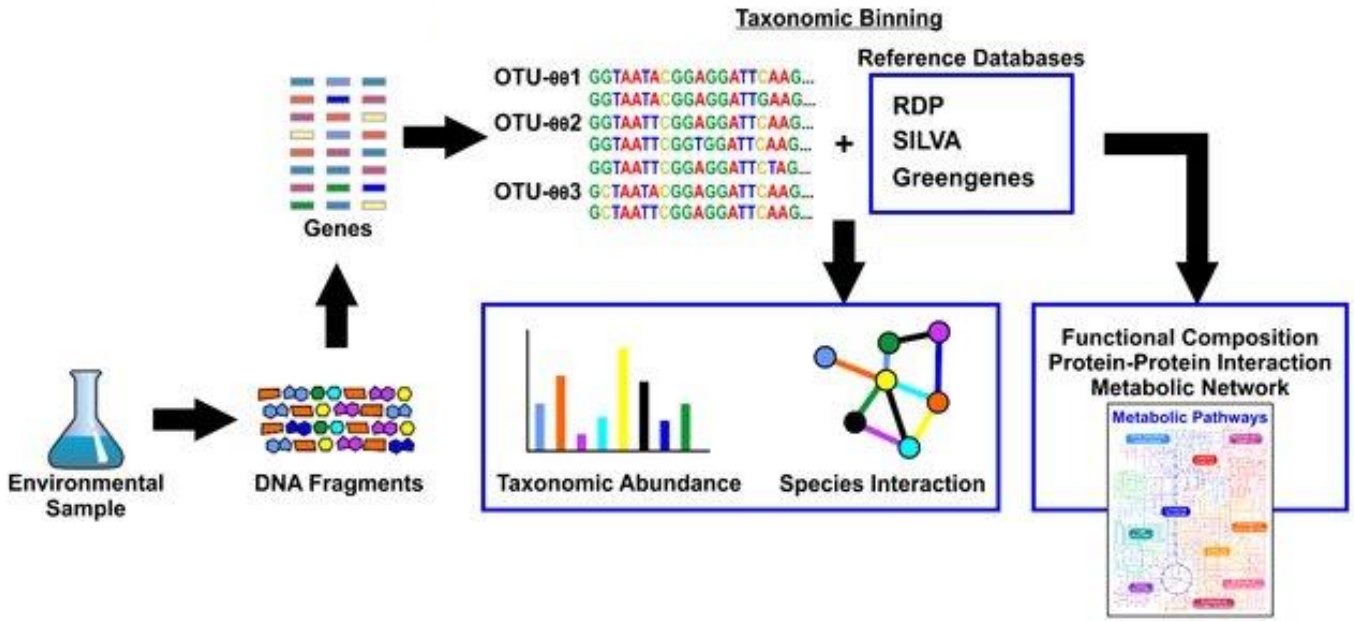
A hypothetical metagenomics analysis pipeline starts from amplification of sampled DNA

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

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



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



Kucharzyk et al 2018

- 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

Early approaches in molecular phylogenetics

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, archea, 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

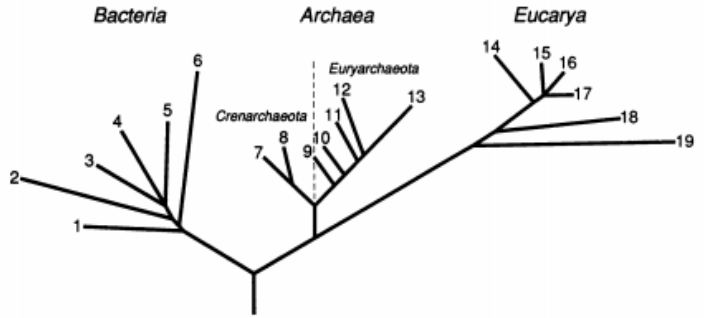


FIG. 1. Universal phylogenetic tree in rooted form, showing the three domains. Branching order and branch lengths are based upon rRNA sequence comparisons (and have been taken from figure 4 of ref. 2). The position of the root was determined by comparing (the few known) sequences of pairs of paralogous 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. Archae: the kingdom Crenarchaeota: 7, the genus *Pyrodictium*; 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.

nature microbiology

LETTERS

PUBLISHED: 11 APRIL 2016 | ARTICLE NUMBER: 16048 | DOI: 10.1038/NMICROBIOL.2016.48

OPEN

A new view of the tree of life

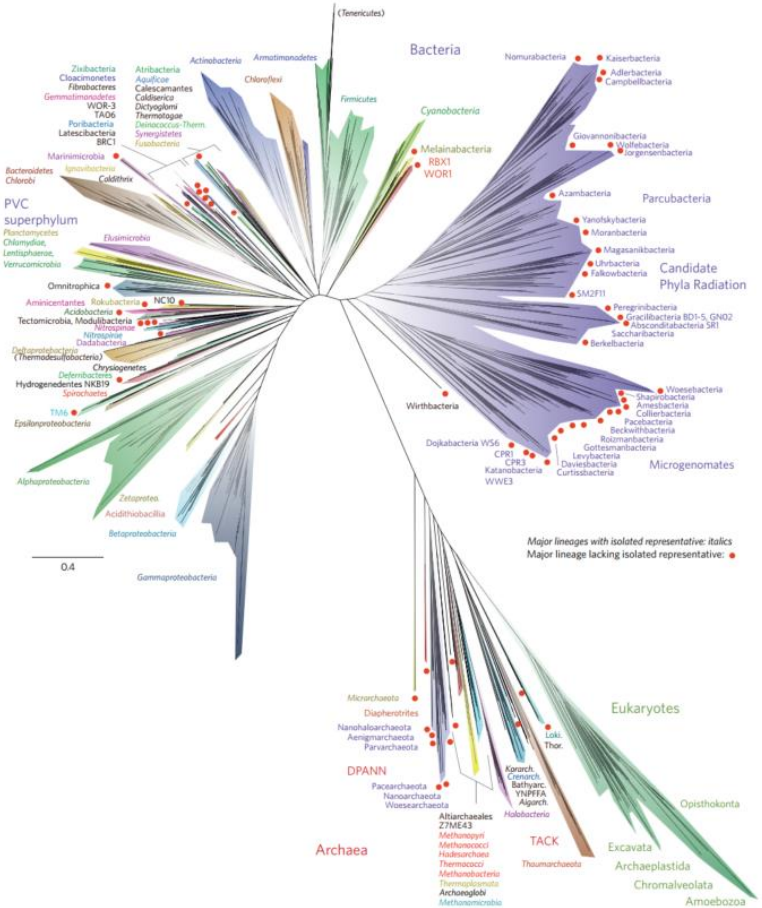
Laura A. Hug[1†], Brett J. Baker[2], Karthik Anantharaman[1], Christopher T. Brown[3], Alexander J. Probst[1], Cindy J. Castelle[1], Cristina N. Butterfield[1], Alex W. Hernsdorf[3], Yuki Amano[4], Kotaro Ise[4], Yohey Suzuki[5], Natasha Dudek[6], David A. Relman[7,8], Kari M. Finstad[9], Ronald Amundson[9], Brian C. Thomas[1] and Jillian F. Banfield[1,9*]

The tree of life is one of the most important organizing principles in biology[1]. Gene surveys suggest the existence of an enormous number of branches[2], 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 eukaryotes[6]. 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



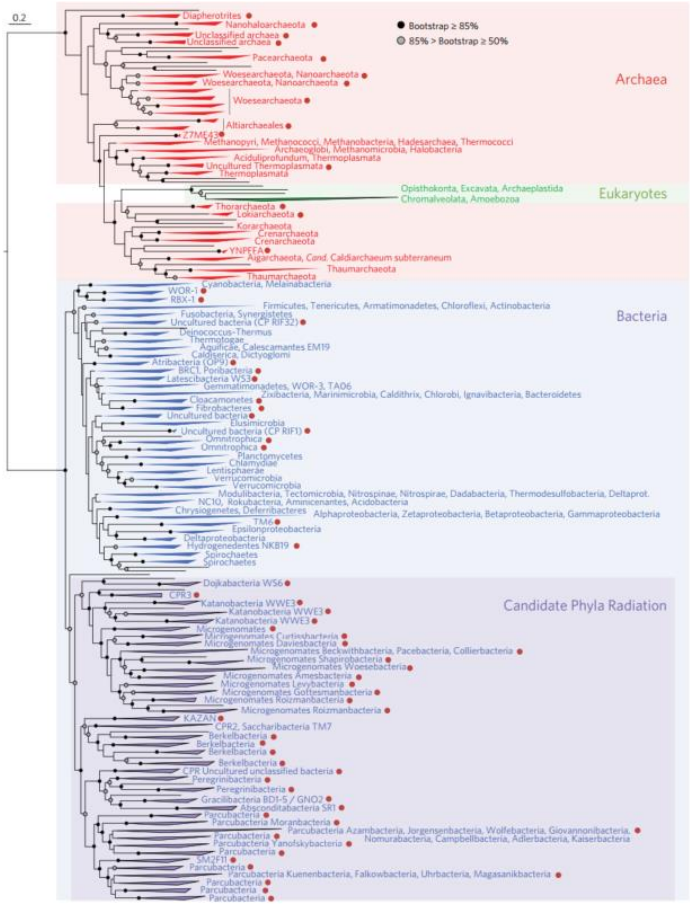


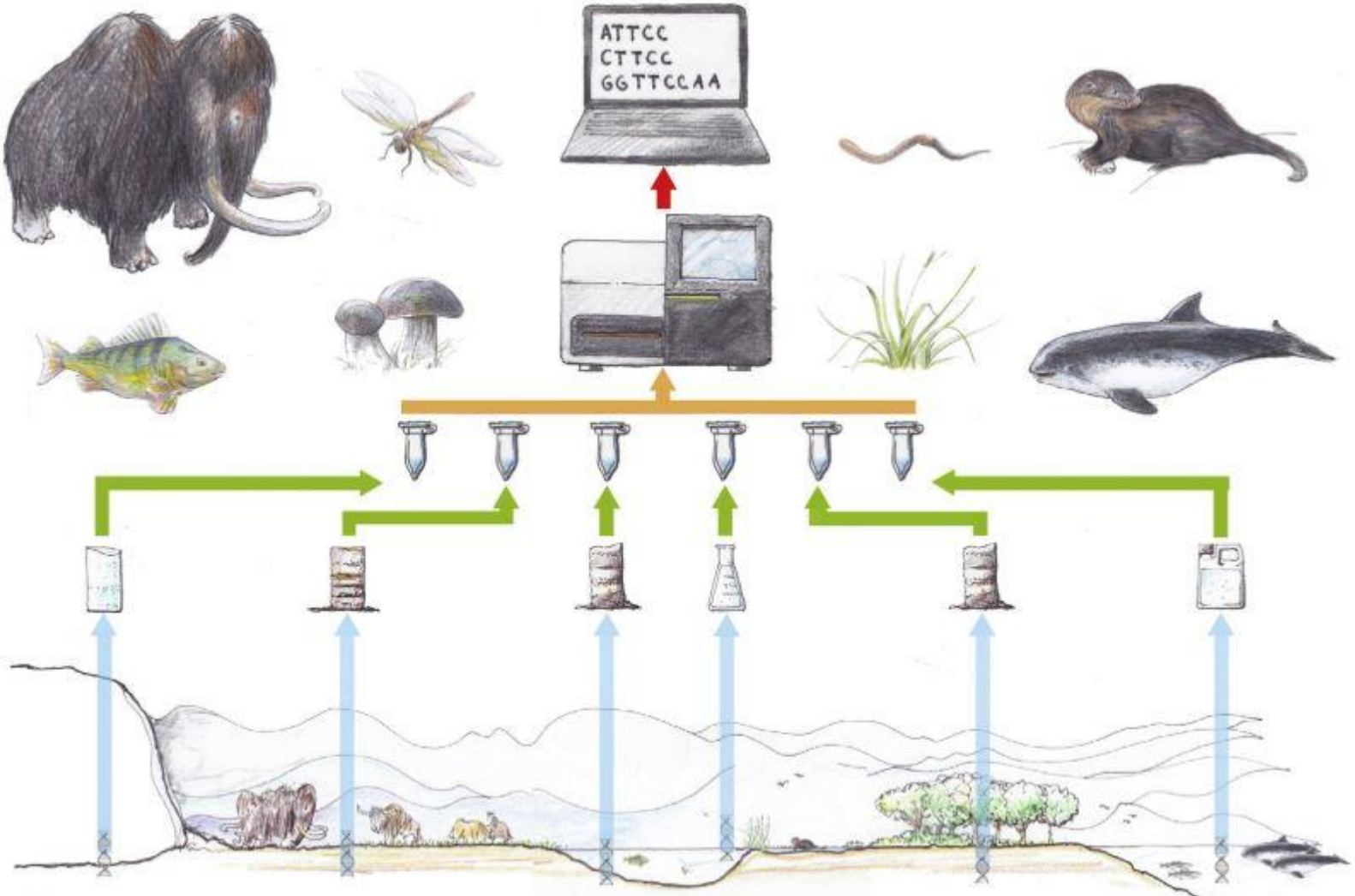
Figure 2 | A reformatted view of the tree in Fig. 1 in which each major lineage represents the same amount of evolutionary distance. The threshold for groups (coloured wedges) was an average branch length of <0.65 substitutions per site. Notably, some well-accepted phyla become single groups and others 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.

A step forward: environmental DNA (eDNA)

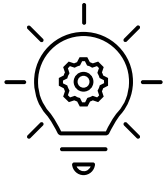
Some definitions:

DNA captured from an environmental sample **without first isolating any target organisms** (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012).

Traces of DNA can be from faeces, mucus, skin cells, organelles, gametes or even extracellular DNA. Environmental DNA can be sampled from **modern environments or ancient environments** (Thomsen & Willerslev, 2015).

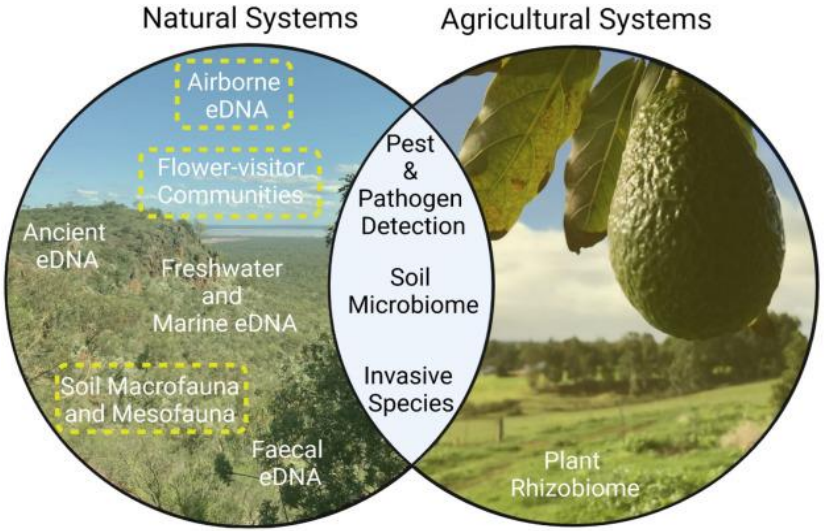

ATTCC
CTTCC
GGTTCCAA

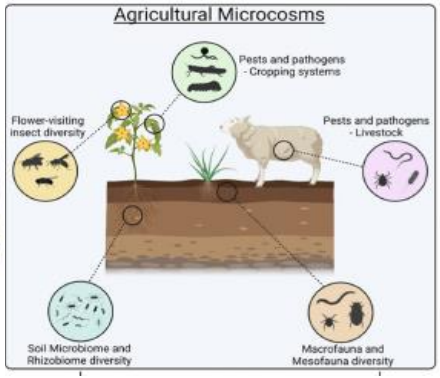
A step forward: environmental DNA (eDNA)

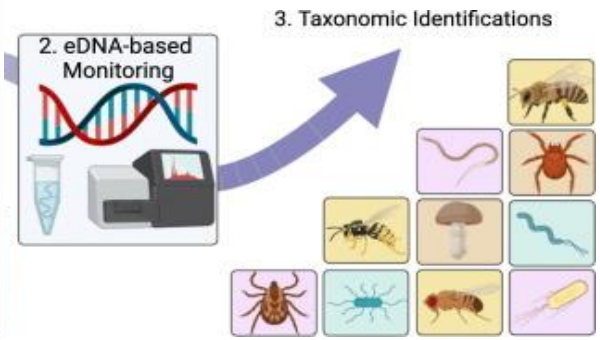


We could use it to identify the agrobiodiversity within an agroecosystem

eDNA in agricultural systems



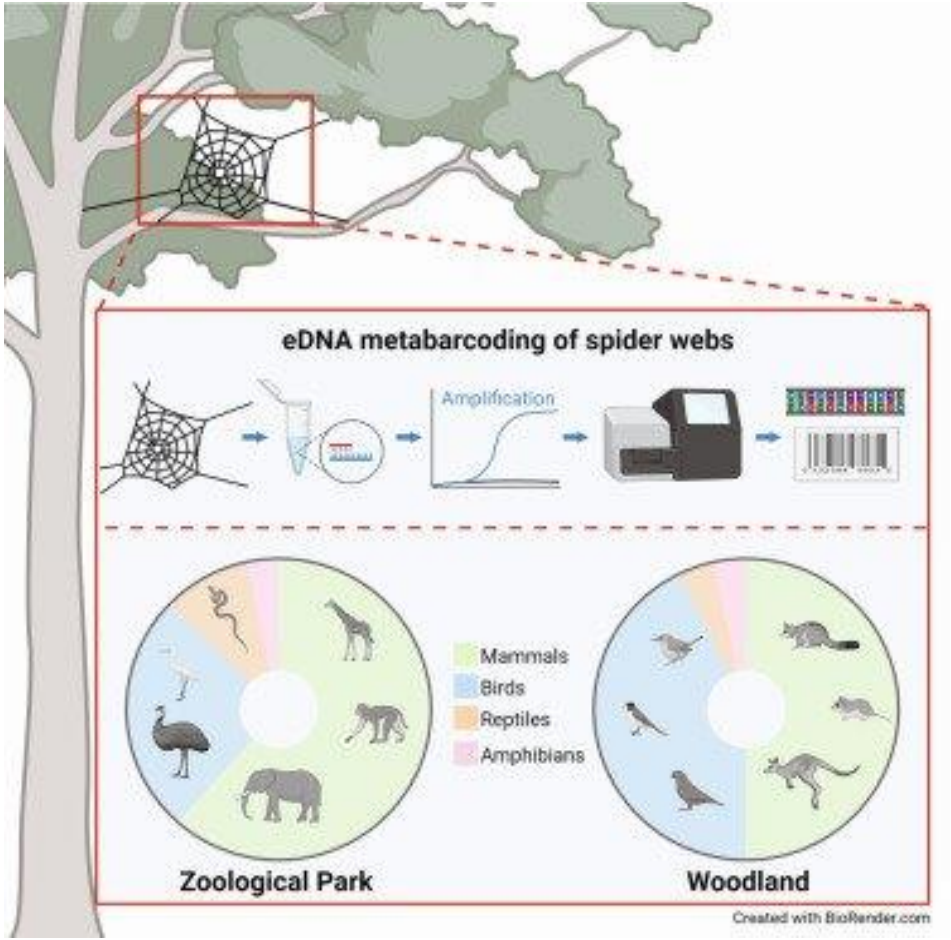




Yellow boxes designate applications of eDNA which are used in natural systems and are emerging in food production systems.
(Kestell et al., 2022)

Accurate identifications to capture diversity and rapid scalable monitoring methods are necessary to identify emerging threats, inform on ecosystem health, and provide **evidence for new management practices**

eDNA from spider webs

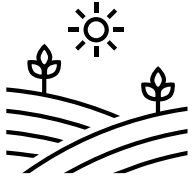


Experimental approach to biotraps for eDNA sampling

- Spider are key **predators** present all year
- Very **diversified** group in agrosystems
- Spiderwebs filter a **large volume of air** and capture (among other things) eDNA

(Newton et al., 2023)

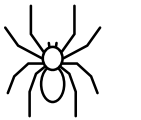
The B-REAL project: characterization of **below-ground and above-ground biodiversity** in multiple sites to monitor biodiversity as associated with **different agricultural practices** conducted in B-REAL

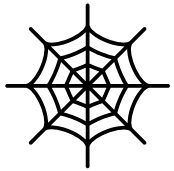


300 samples

3 substrates for eDNA

aereal part of plants





spiderwebs

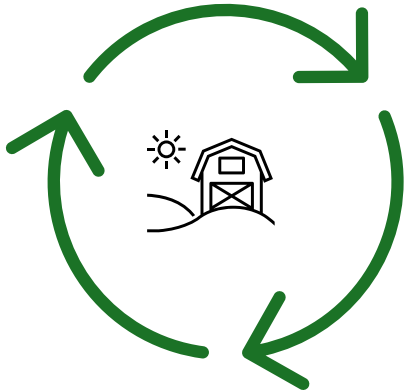
soil

Colombia/Perù

Kenya

Expected results

1. A definition of the **potential of biodiversity characterization** by eDNA for each substrate



2. The production of **eDNA sequencing data** from all samples involved in the study

3. The characterization of biodiversity associated **in each target substrate in each sampling area**

WORKFLOW

Study design

Basic science or applied? (e.g., environmental biomonitoring)

What is your study goal?
- **presence/absence**
- **diversity assessment**
- **absolute quantification**

What taxa will you target?

Is the scale of inference for your sample type appropriate to your question?

Can you compare complementary data types? (e.g. traditional vs. eDNA)

Does your sampling/ replication scheme provide good statistical power?

In the field

What type of sample is needed? (water, soil, air)

What metadata should you collect?

How many replicates will you collect?

Does your sampling protocol minimize/ control for :
- **contamination** (e.g., positive and negative controls)
- **any known biases** (e.g., inhibitors, sample volume)

In the laboratory

Sample Handling Phase

What extraction method? (physical vs. chemical)

How much sample?

What locus and primers?

Do you need to generate reference sequence data?

Are technical replicates needed?

What library preparation method will you use?

How many samples will you index and pool?

What sequence depth is needed per sample ?

What read length will you use?

DNA Processing Phase

What sequencing platform will you use?

Do you need paired-end sequencing?

Have you included appropriate quality assurances? (e.g., mock community, qPCR, bioanalyser traces)

Does your laboratory protocol minimize/ control for:
- **contamination** (e.g., positive and negative controls)
- **any known biases** (e.g., primer bias, coverage, taxonomic resolution)

At the keyboard

How complete is the reference database?

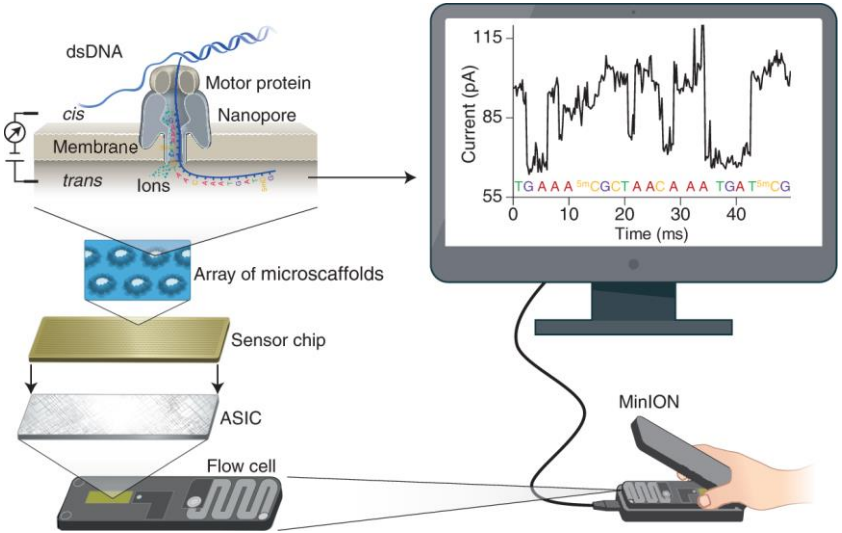
Do you have adequate sequencing coverage across samples?

Are you using appropriate choices for software tools, parameters?

Are your biological conclusions upheld using alternative parameters and workflows?

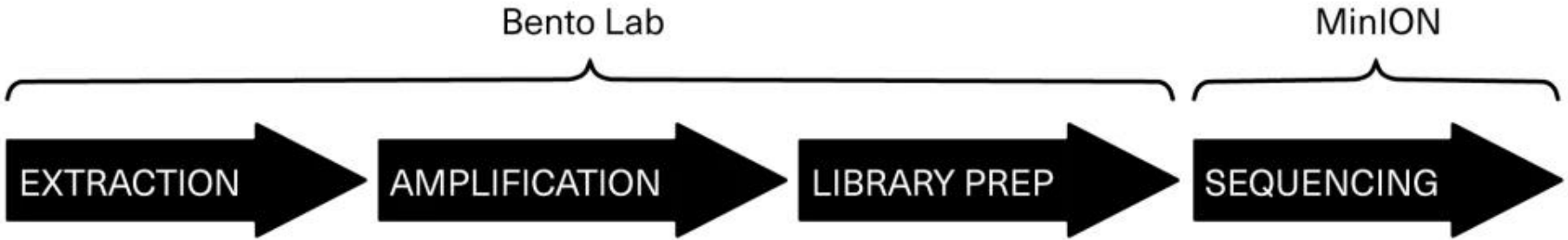
Are you including appropriate quality filtering of your data? (see Box 2)

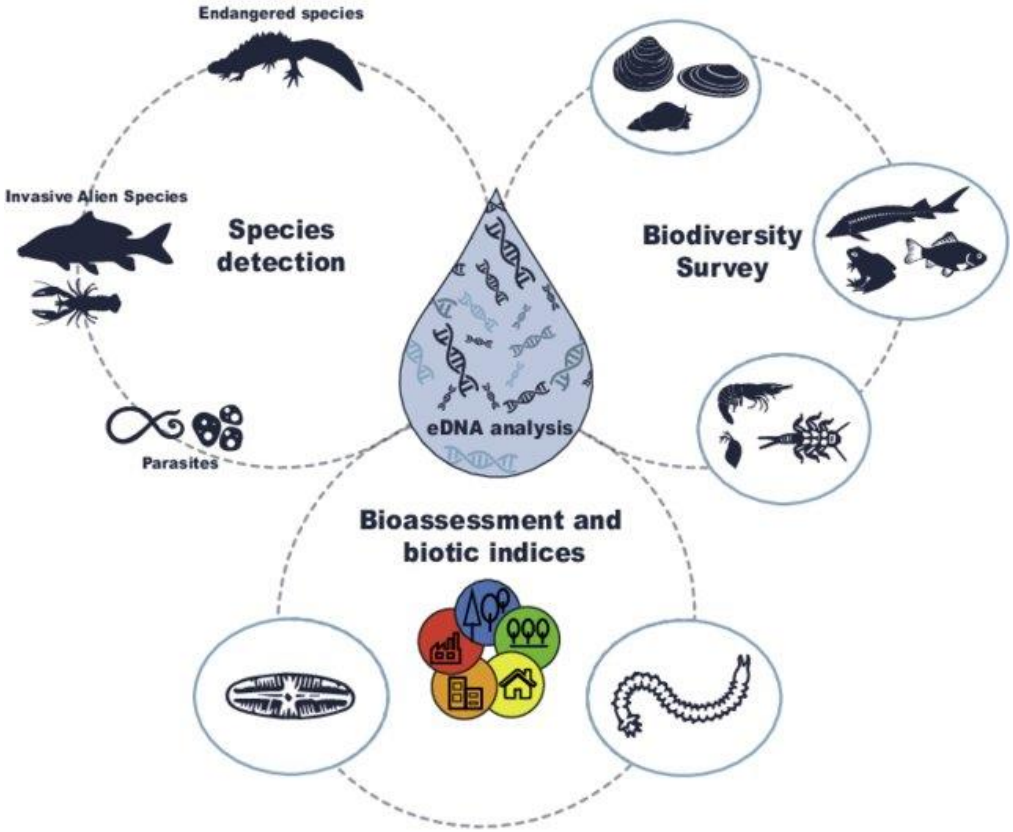
Real time sequencing with pocket size devices



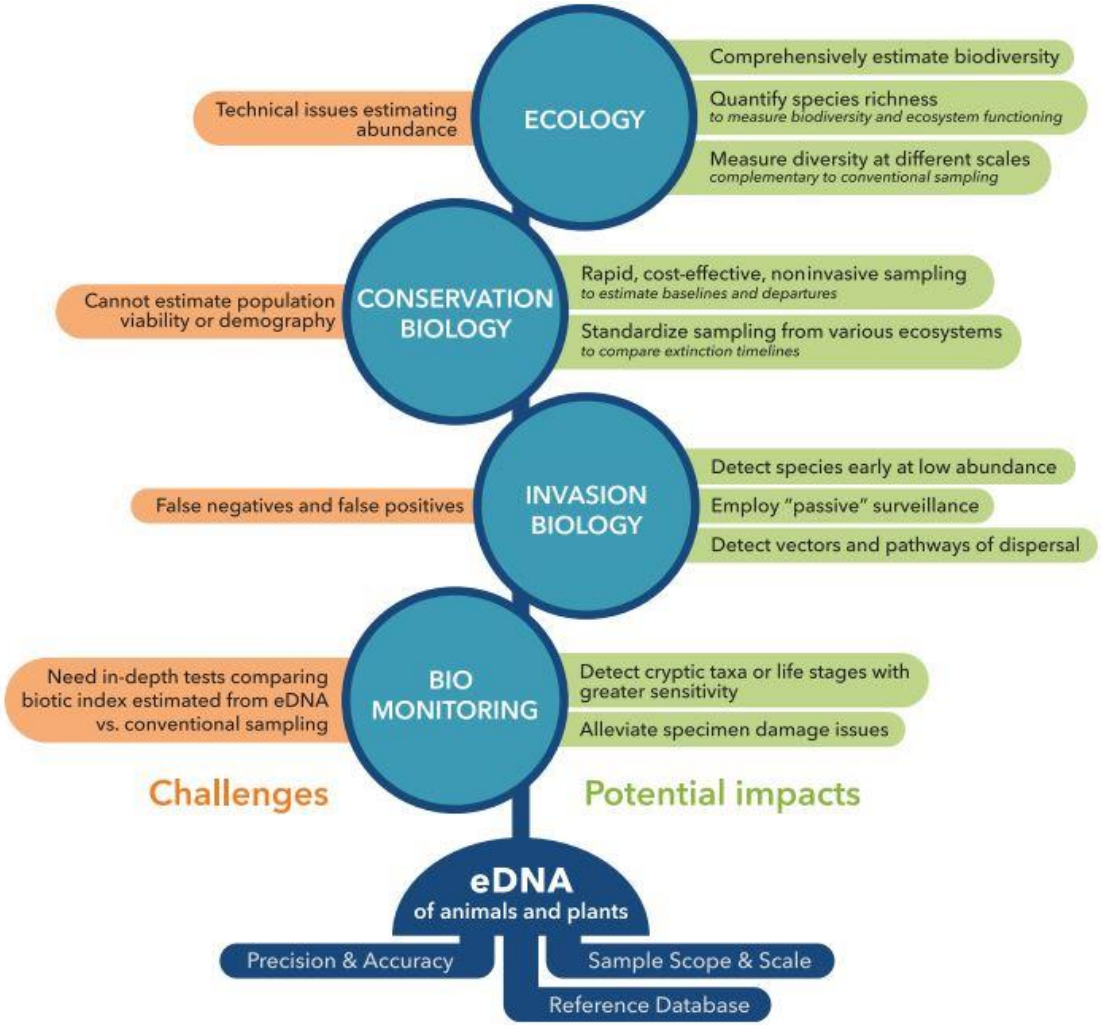
Readout of ion current changes occurring when single- stranded DNA passes through a protein pore

Real time sequencing with pocket size devices





Environmental DNA typing


ECOLOGY
Technical issues estimating abundance
Comprehensively estimate biodiversity
Quantify species richness
to measure biodiversity and ecosystem functioning
Measure diversity at different scales
complementary to conventional sampling
CONSERVATION BIOLOGY
Cannot estimate population viability or demography
Rapid, cost-effective, noninvasive sampling
to estimate baselines and departures
Standardize sampling from various ecosystems
to compare extinction timelines
INVASION BIOLOGY
False negatives and false positives
Detect species early at low abundance
Employ "passive" surveillance
Detect vectors and pathways of dispersal
BIO MONITORING
Need in-depth tests comparing biotic index estimated from eDNA vs. conventional sampling
Detect cryptic taxa or life stages with greater sensitivity
Alleviate specimen damage issues
Challenges
Potential impacts
eDNA
of animals and plants
Precision & Accuracy
Sample Scope & Scale
Reference Database

eDNA challenges

- Optimizing and standardizing procedures
- The problem of contamination
- Primer design and reference database
- Understanding spatial and temporal dynamics: How long does it persist? How does it expand in between?
- At sea: large volumes of water, additional abiotic factors such as salinity, currents
- Population parameters that cannot be obtained by eDNA metabarcoding alone

Time for questions!

?

EDNA: ADVANTAGES

- No need for experts for taxonomic recognition
- No identification errors
- No impact on the environment and biological communities (applications in marine protected areas)
- Useful for elusive or small species
- Low costs

Edna advantages

- No need for experts for taxonomic recognition
- No identification errors
- No impact on the environment and biological communities (applications in marine protected areas)
- Useful for elusive or small species
- Low costs