General introduction to barcoding and metabarcoding

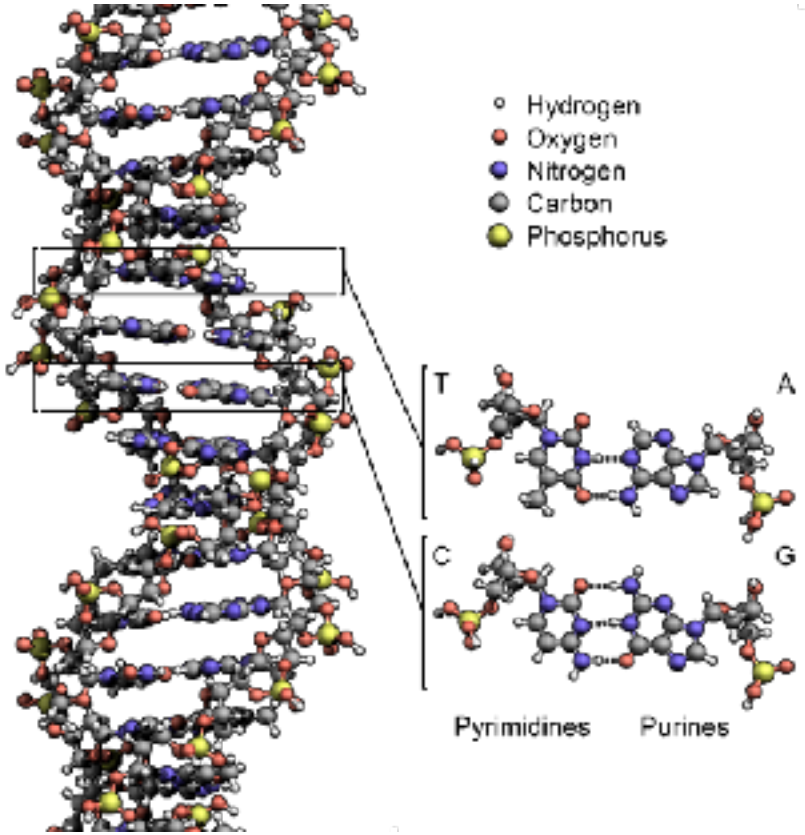
Agnès Bouchez

BIOLAWEB

Summary

- **DNA - some basics**
- DNA and barcoding
- DNA and metabarcoding

Funded by European Union

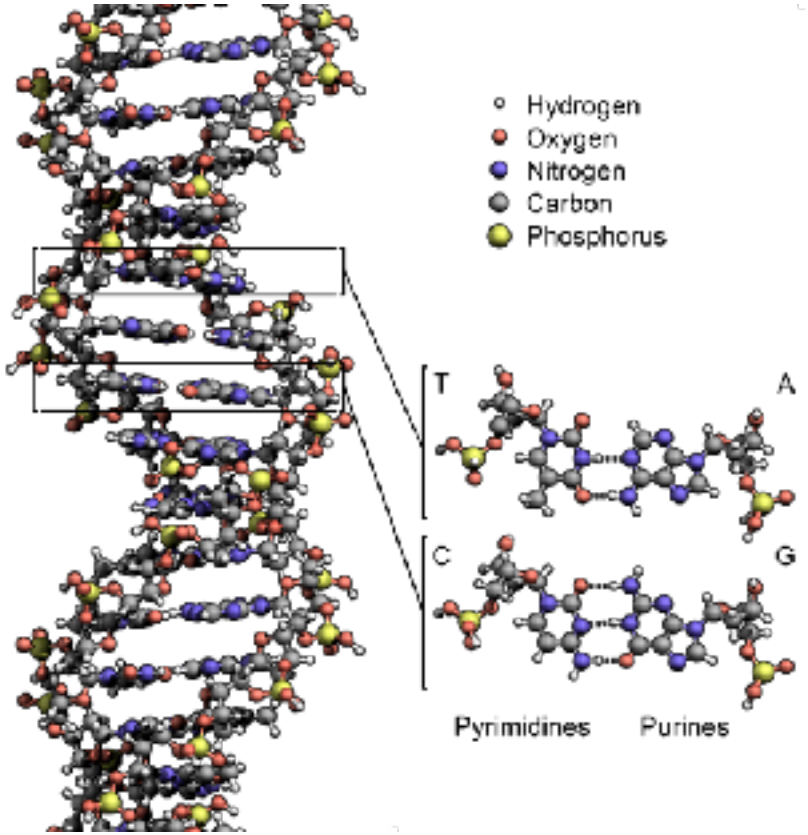


DNA = DesoxyriboNucleic Acid

Watson and Crick 1953

DNA molecule → a 3-dimensional double helix

DNA in all living organisms

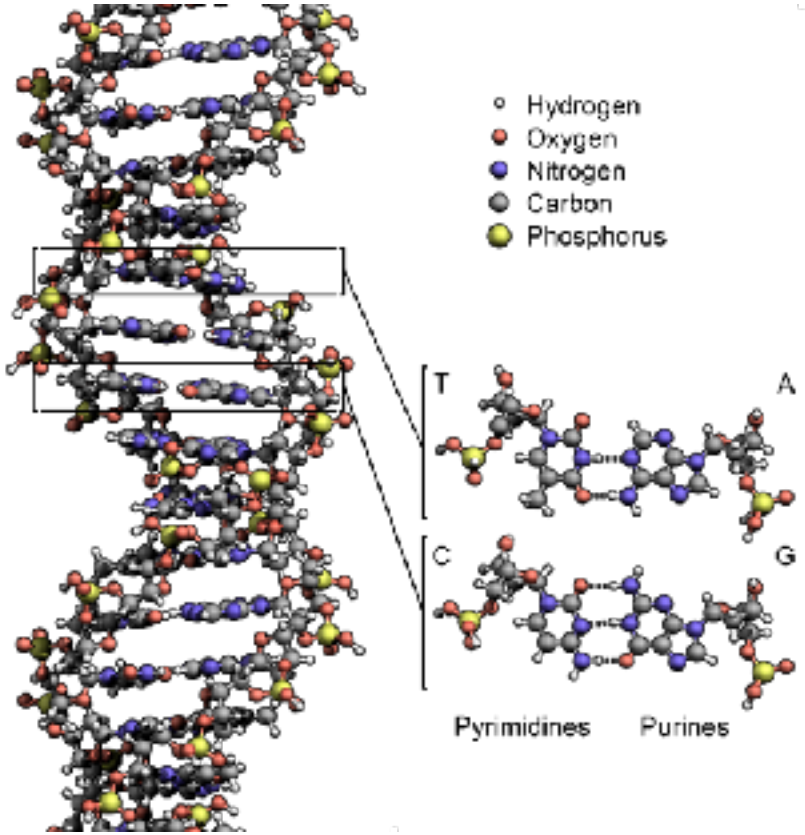


DNA = DesoxyriboNucleic Acid

- All living organisms store genetic information using DNA

Funded by European Union

DNA in all living organisms

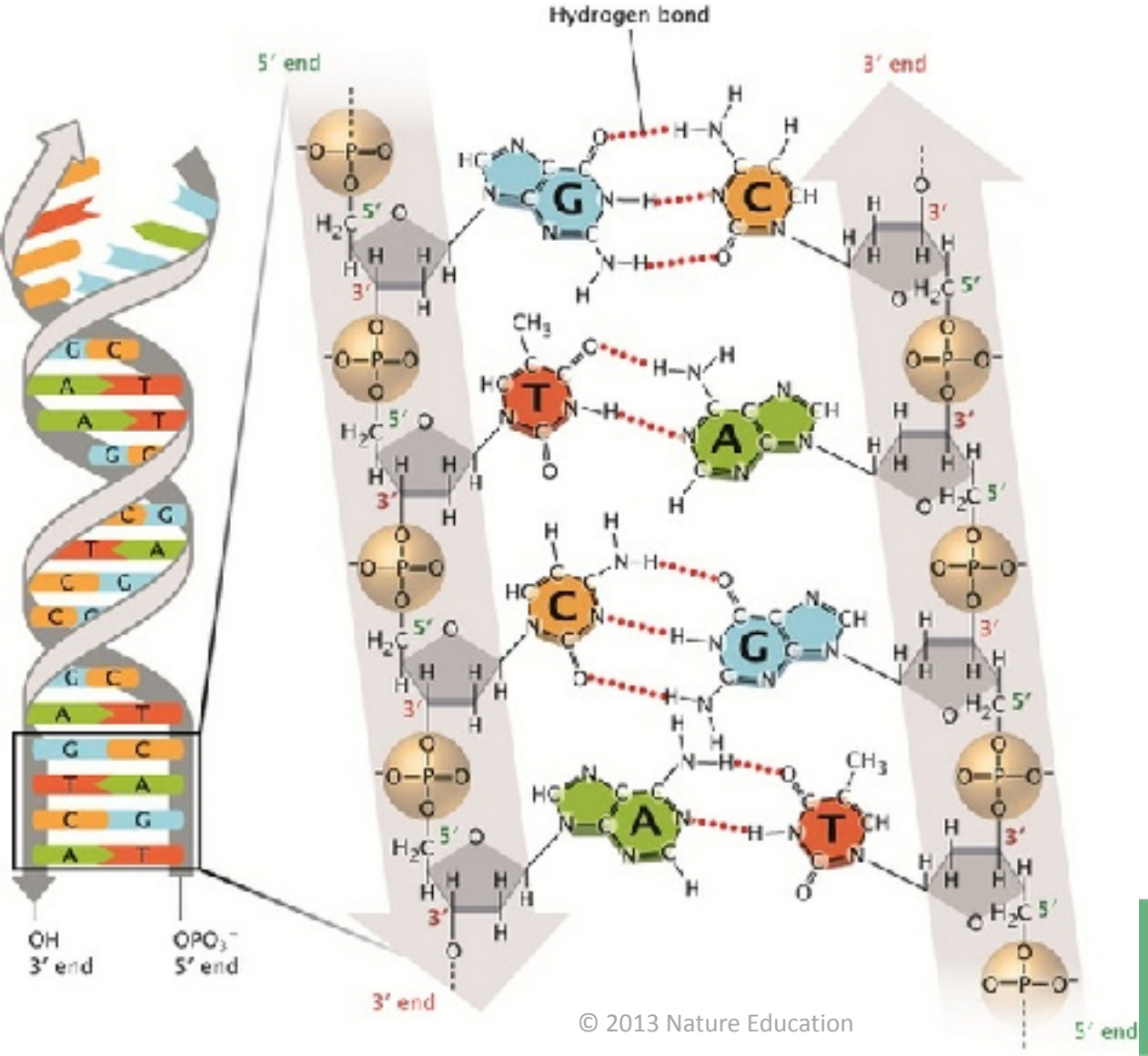


DNA = DesoxyriboNucleic Acid

- All living organisms store genetic information using DNA
- The genetic code is based on an assembly of 4 nucleic bases:
  - ✓ A = Adenine ✓ T = Thymine
  - ✓ G = Guanine ✓ C = Cytosine
- Bases go by pair A-T G-C linked by hydrogen bonds

Funded by European Union

DNA in all living organisms

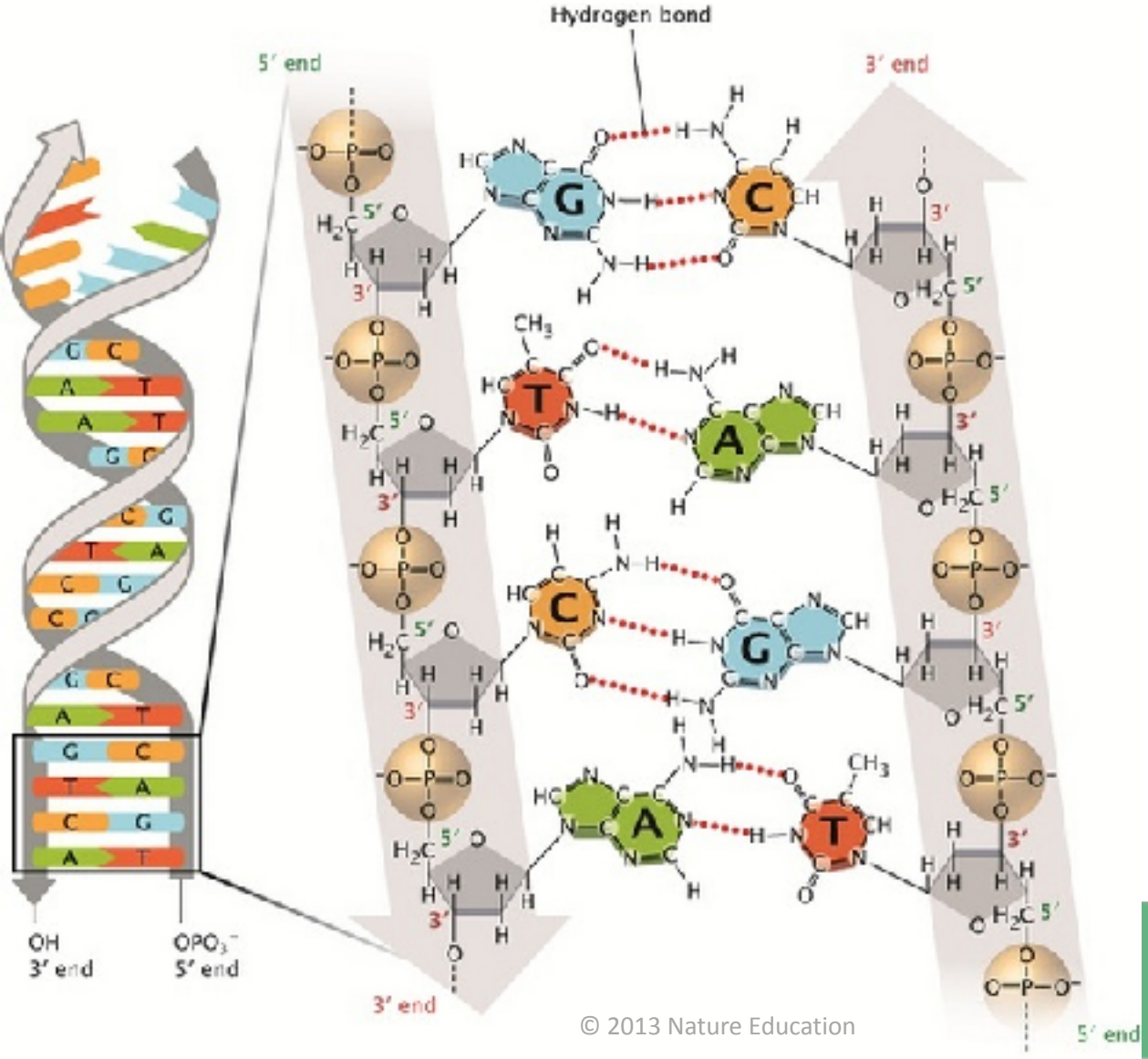


© 2013 Nature Education

DNA = DesoxyriboNucleic Acid

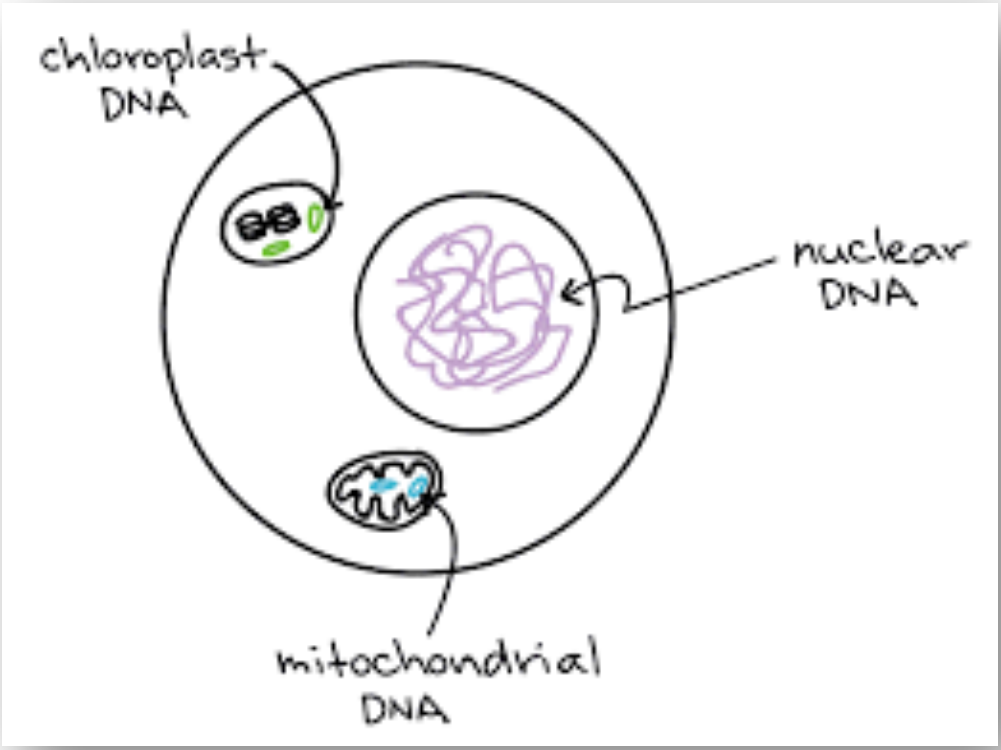
- All living organisms store genetic information using DNA
- The genetic code is based on an assembly of 4 nucleic bases:
  - ✓ A = Adénine
  - ✓ T = Thymine
  - ✓ G = Guanine
  - ✓ C = Cytosine
- Bases go by pair A-T & G-C linked by 2 or 3 hydrogen bonds, respectively
- Base/sugar/phosphate: Nucleotide
- Nucleotides assemble in a double helix

DNA in all living organisms



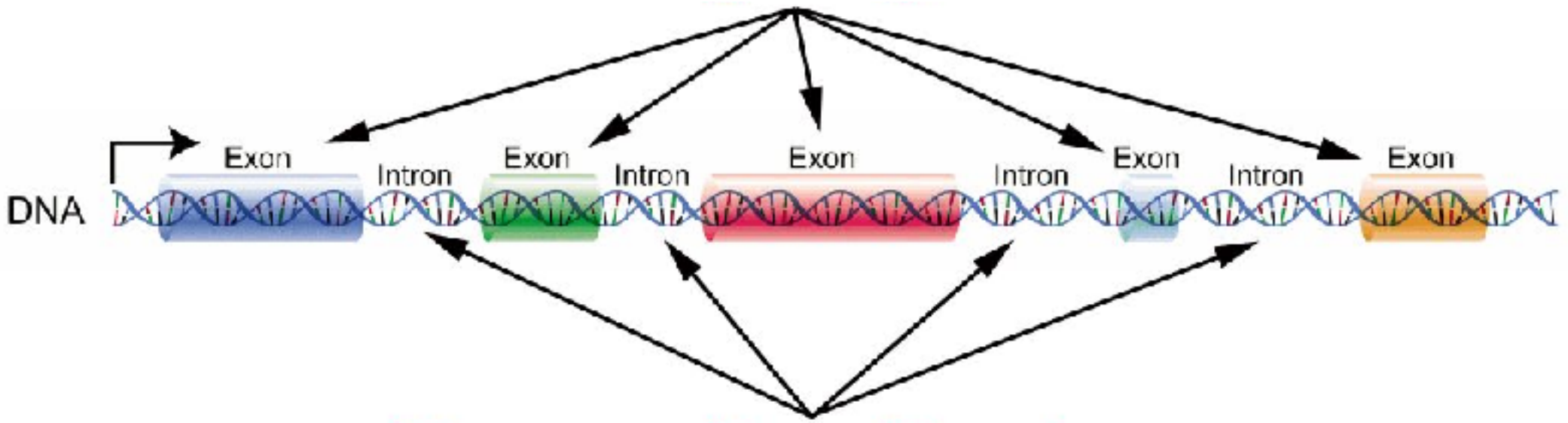
© 2013 Nature Education

DNA In the nucleus and organelles



DNA in all living organisms

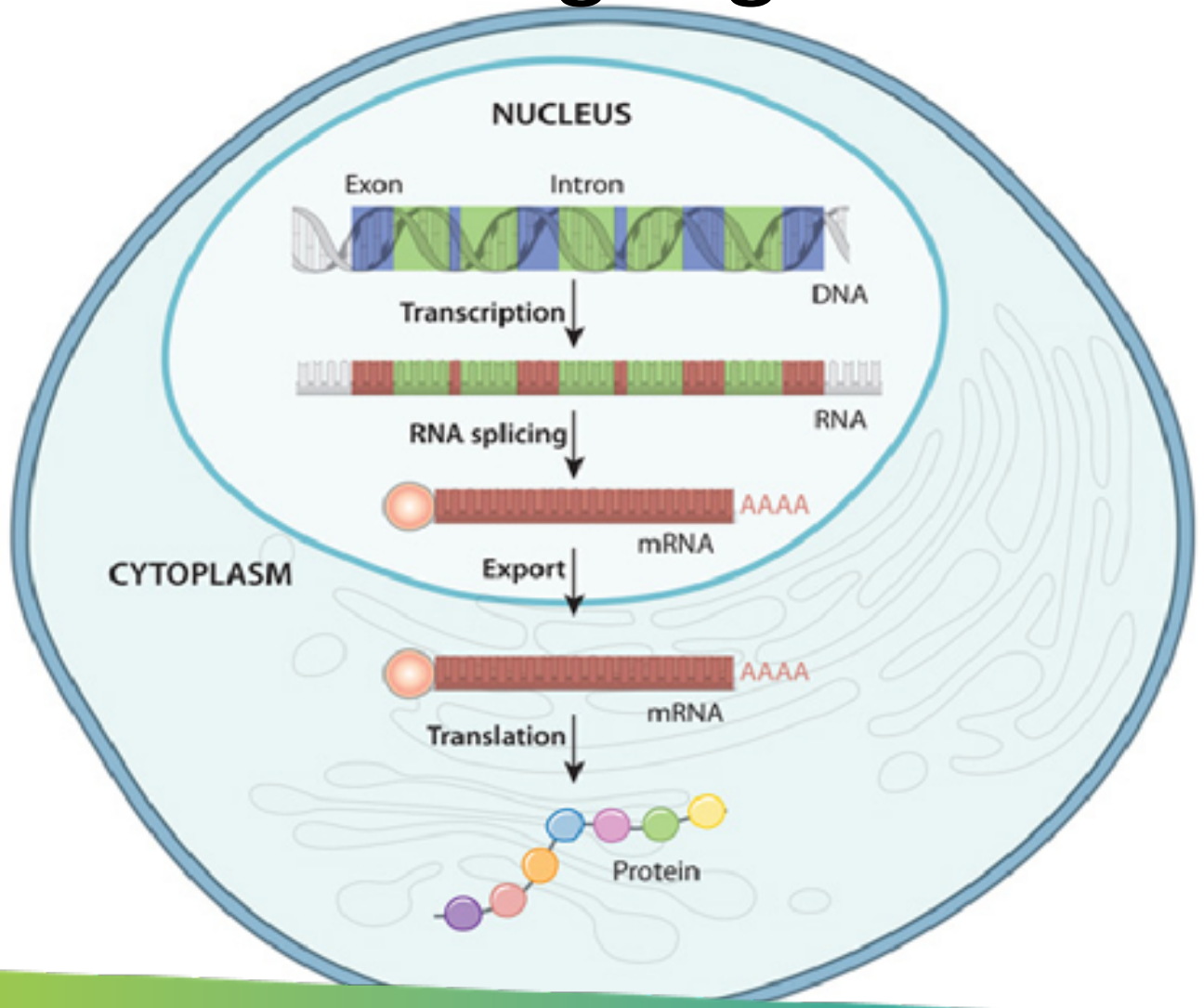
Coding Regions (2%)



Non-coding Regions (98%)

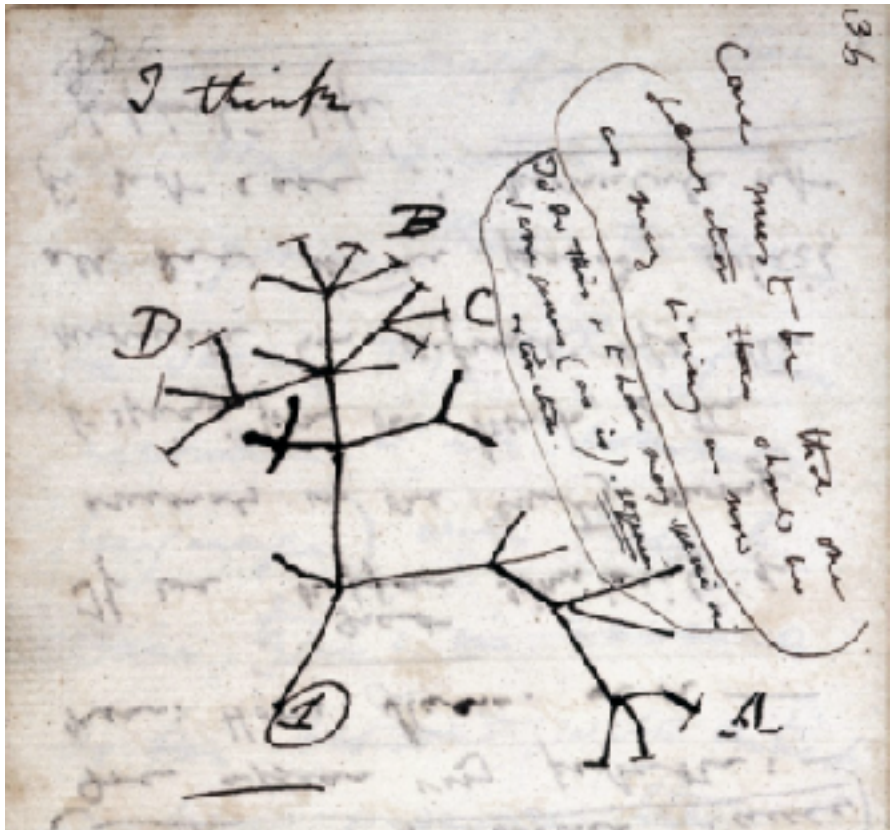
DNA in all living organisms

From DNA
to proteins



DNA in all living organisms

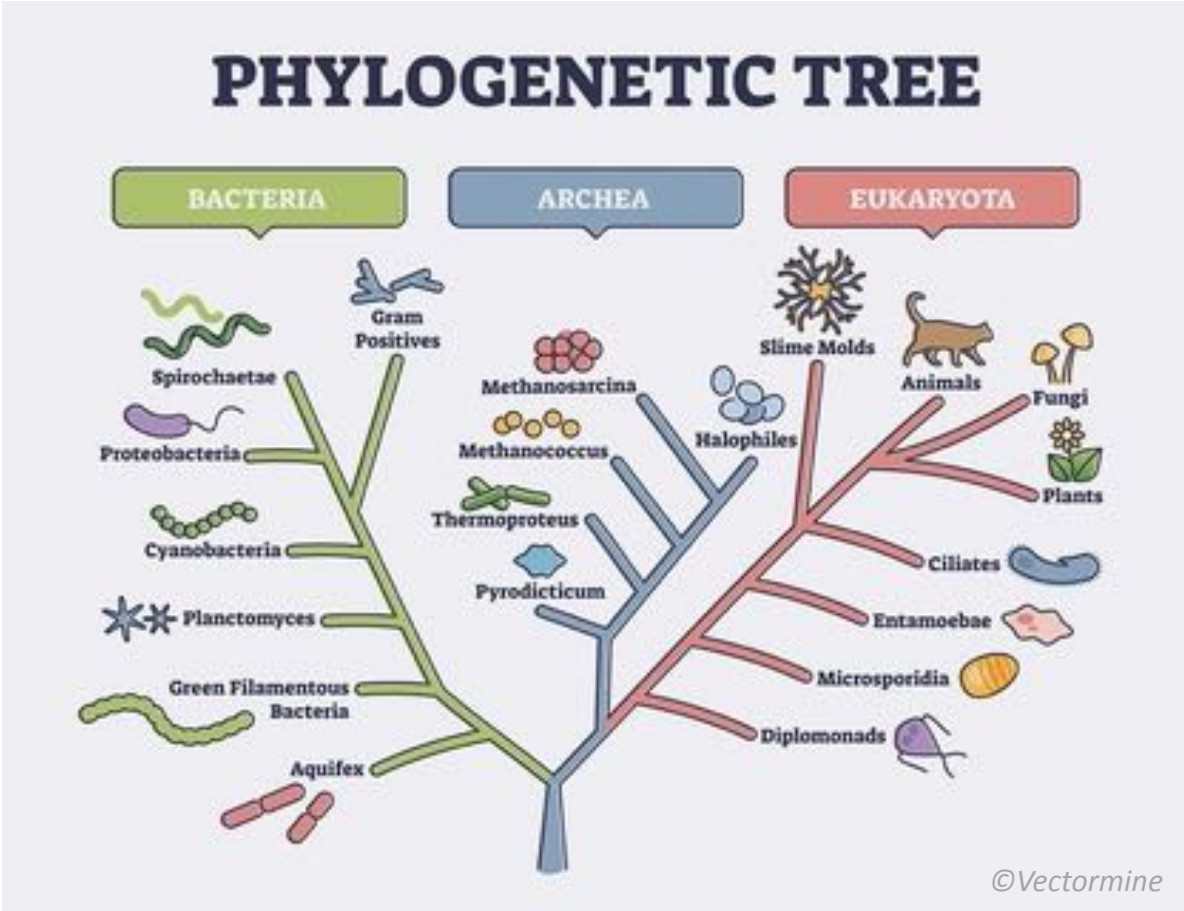
- Similar genes ... inherited from a common ancestor



Darwin 1837

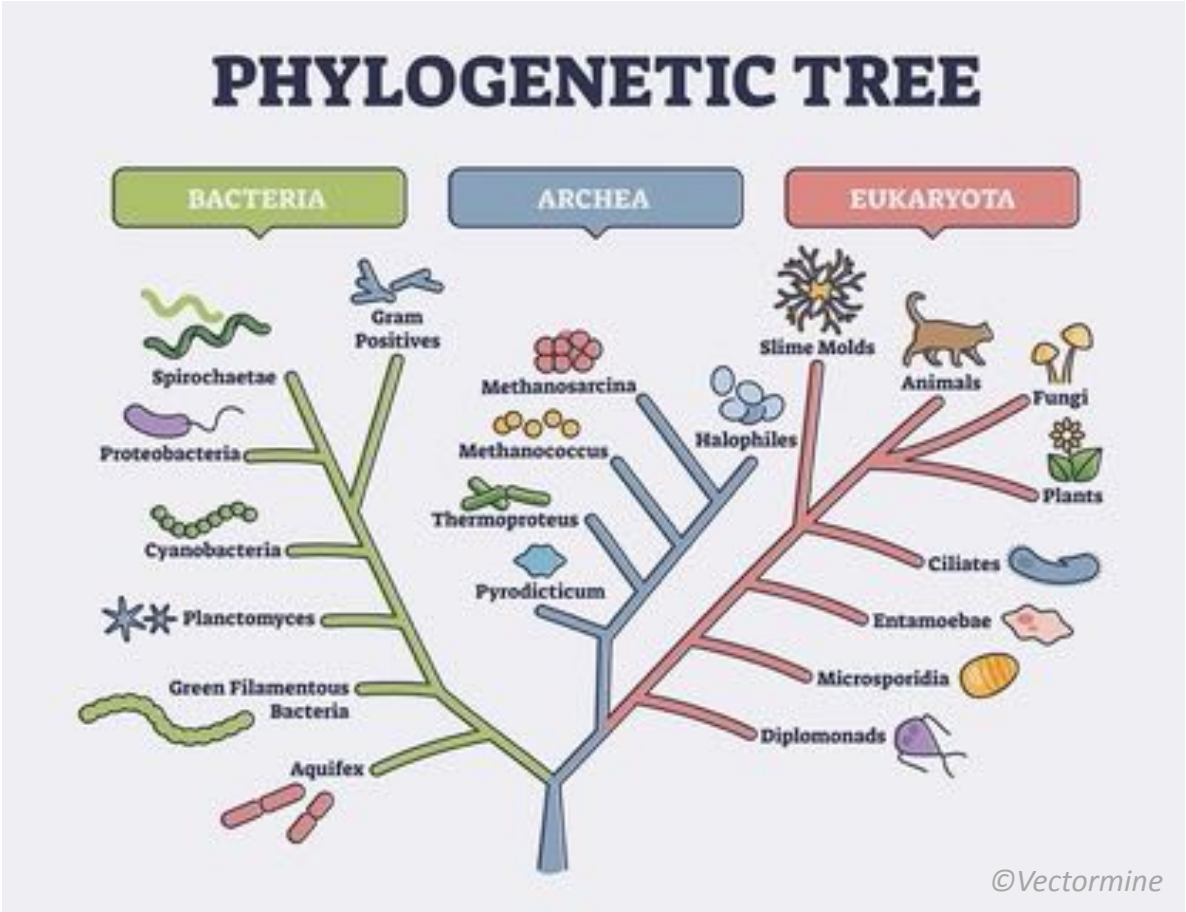
DNA in all living organisms

- Similar genes ... inherited from a common ancestor



DNA in all living organisms

- Similar genes ... inherited from a common ancestor
- But each living organism has a different genetic code



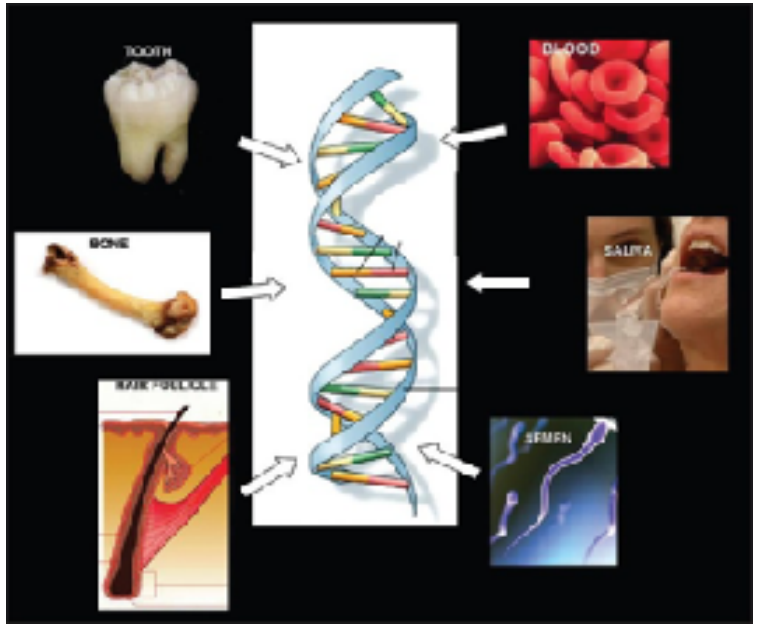
BIOLAWEB

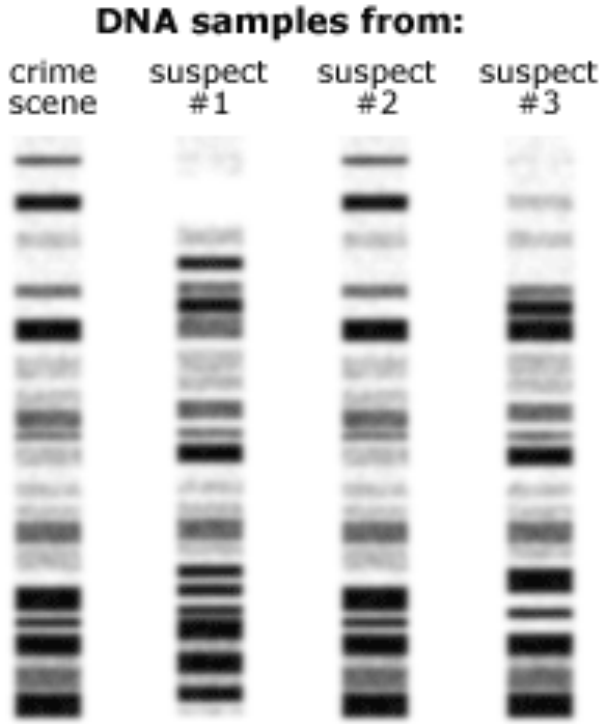
Summary

- DNA - some basics
- **DNA and barcoding**
- DNA and metabarcoding

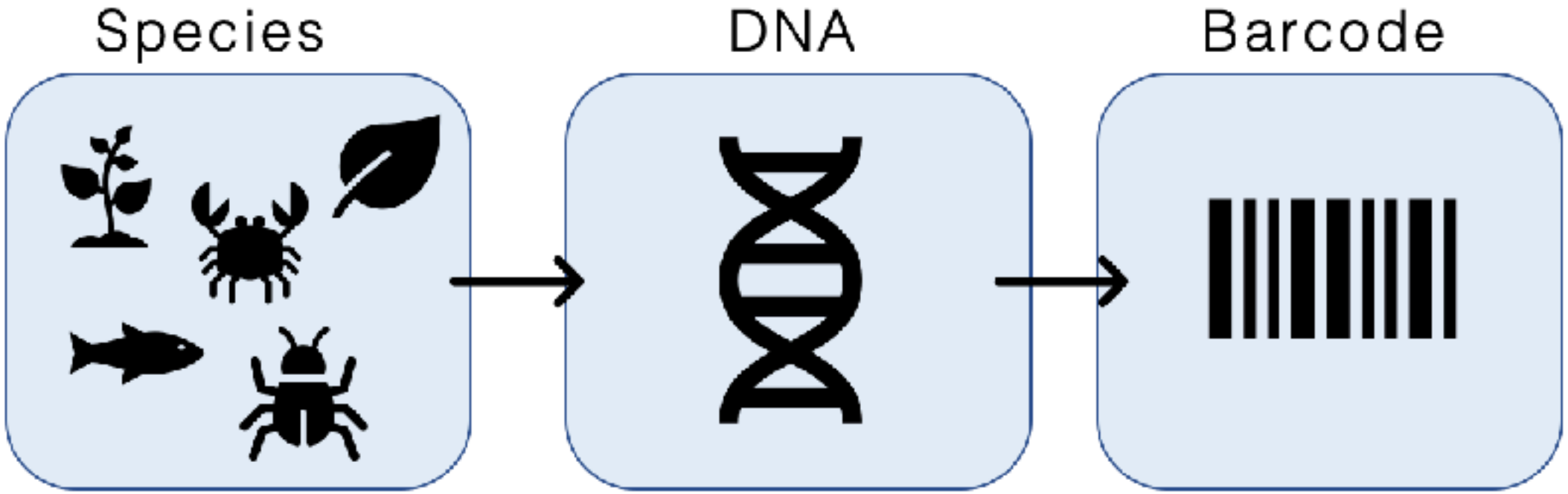
DNA: an identification tool

- In all living organisms
- In all their cells… and cell-remains





ID tool à DNA Barcoding



ID tool à DNA Barcoding

THE ROYAL SOCIETY

Received 29 July 2002
Accepted 30 September 2002
Published online 8 January 2003

Biological identifications through DNA barcodes

Paul D. N. Hebert*, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard

Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Hebert et al. 2003

DNA Barcoding

- Mitochondrial gene: cytochrome c oxidase subunit 1
- Barcode: a 658 pb region of the gene COX1 (or COI)

« can serve as the core of a global bioidentification system for animals »

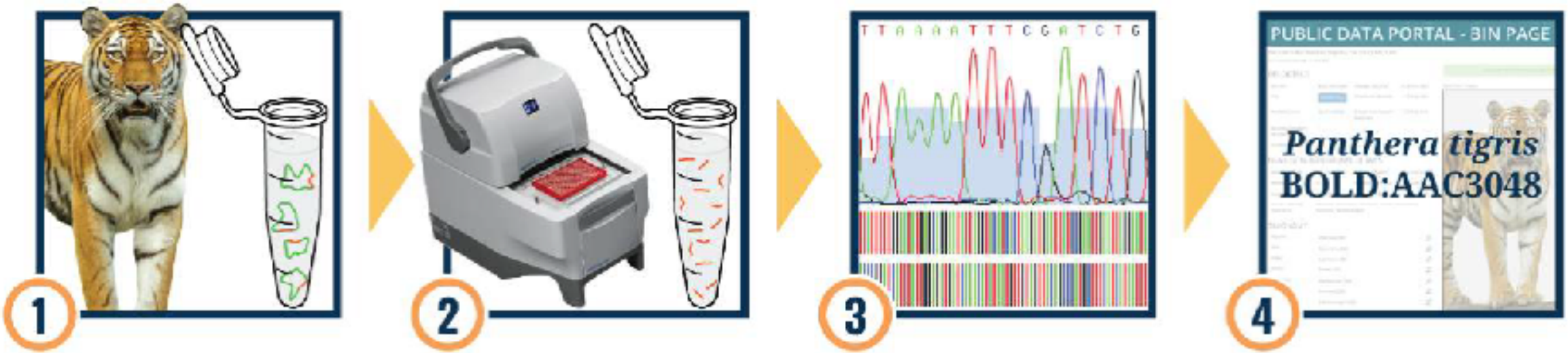
DNA Barcoding

- Mitochondrial gene: cytochrome c oxidase subunit 1
- Barcode: a 658 pb region of the gene COX1 (or COI)

« can serve as the core of a global bioidentification system for animals »

- ❑ A DNA-based taxonomic ID tool for species
- ❑ A tool for sharing taxonomic expertise, overcoming the lack of experts
- ❑ A standardised method using specific information present:
  - in all species
  - at all their life stages (including early life stages)
  - In all their tissues

DNA Barcoding



DNA extracted from a sample of the specimen to be identified → amplified barcode → sequenced barcode (read) → sequence compared to previously identified sequences (reference base)

ID

Funded by European Union

DNA Barcoding

- ✓ Insects : 100 families from 8 orders
- ✓ **1 specimen per family (previously identified – morpho-taxonomy)**
- ✓ Sequencing of the COX1 barcode
  - Reference DNA/taxonomy
  - Taxonomic classification is consistent: morpho / DNA

- ✓ **50 specimens to identify**
- ✓ Sequencing of the COX1 barcode and comparison to references

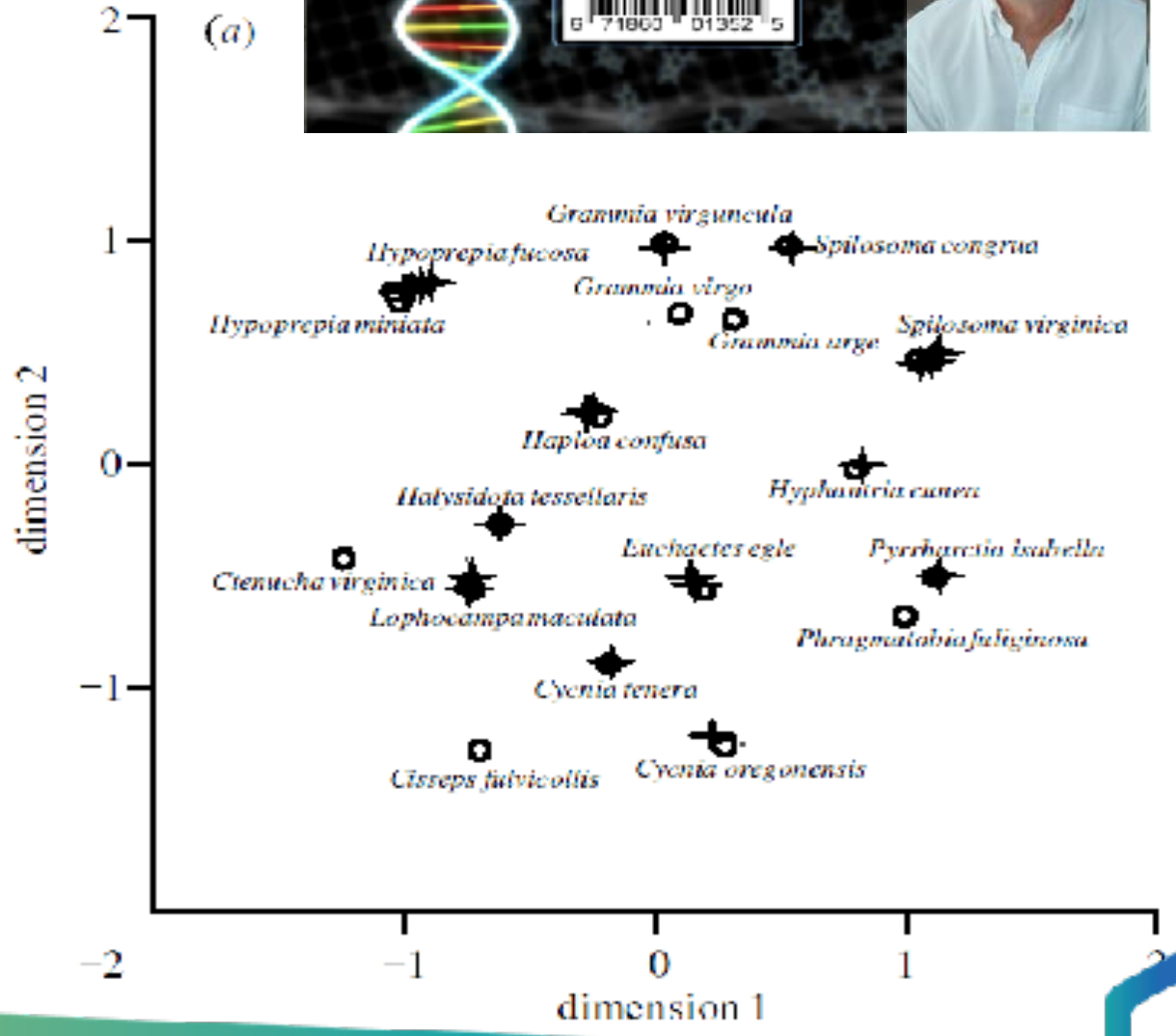
→ 100% correctly identified at the order level

DNA Barcoding

Test on **Lepidoptera**: a challenging group as there is little polymorphism in the COI barco

- ✓ 200 species with ref DNA/taxonomy
- ✓ **Test on 150 specimens to identify**

- ✓ 100% correct ID at species level by comparing sequenced barcodes to reference barcodes
- ✓ ex: 18 sp articideae (+ tests; o refs)

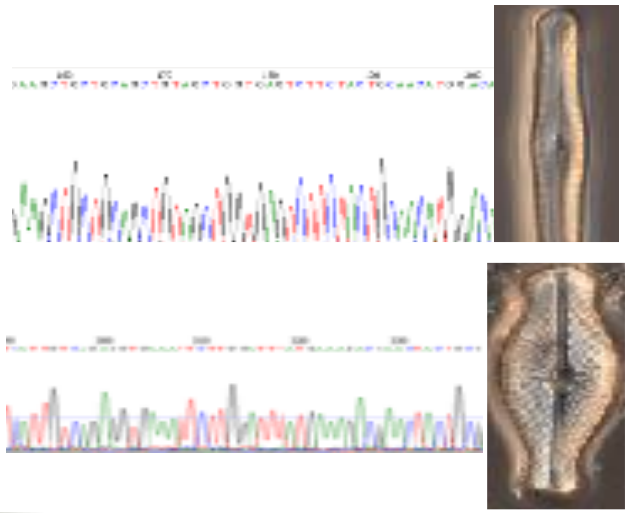


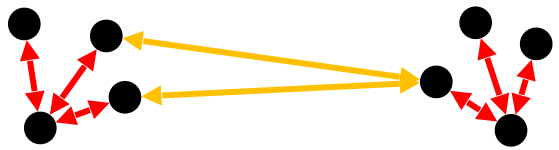
Funded by European Union

DNA Barcoding

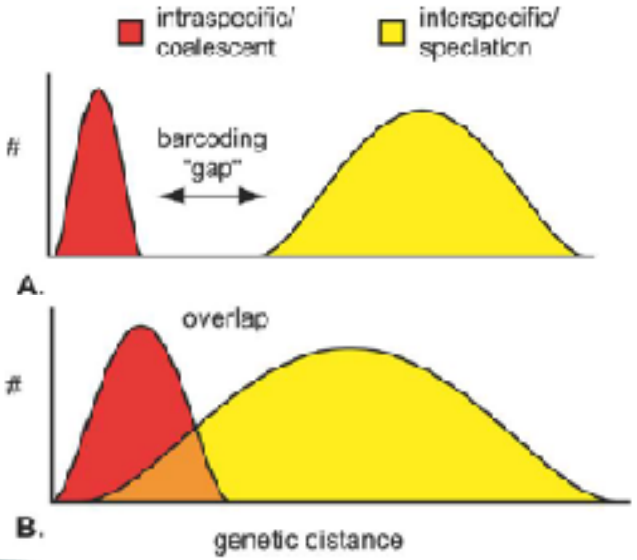
Hebert et al. 2003

- Standardised approach to taxonomic ID
- Barcode: small DNA fragment specific to a taxa
- Identification is based on the existence of a “barcoding gap”





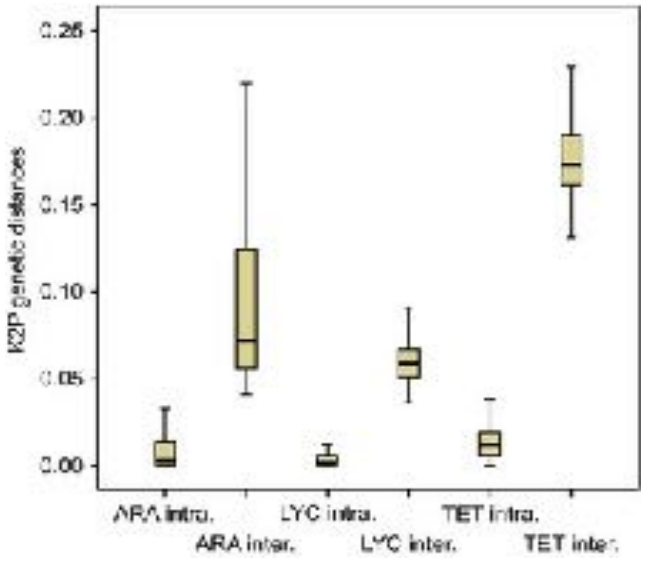
species1 species2

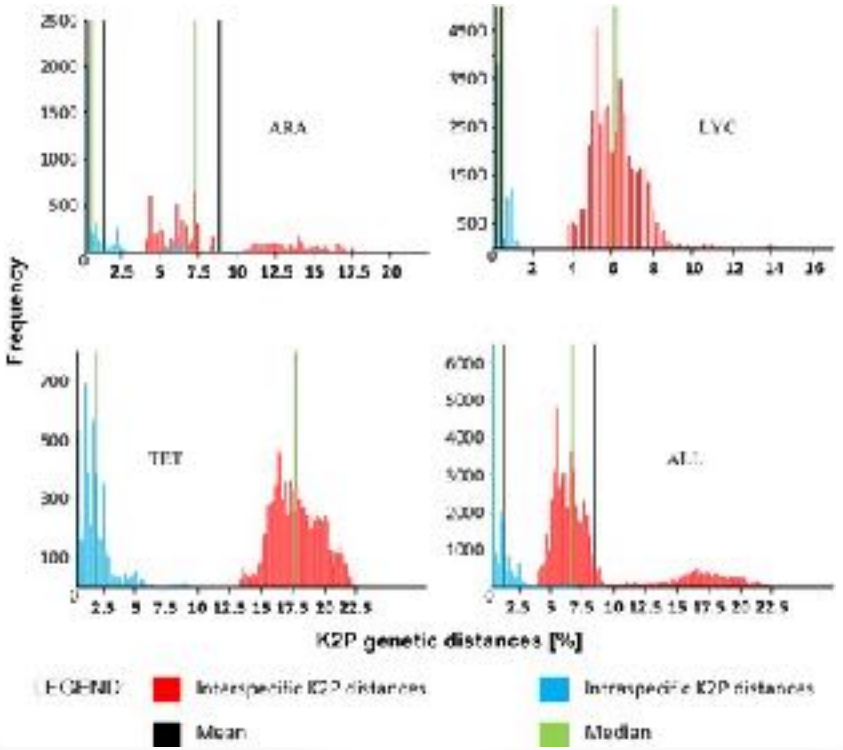


DNA Barcoding: example

- ✓ 1203 individual spider barcodes
- ✓ 162 sp of 3 families
- ✓ (*Araneidae, Tetragnathidae, Lycosidae*)

- ➢ Good efficiency of the DNA barcode for the identification of spider species.
- ➢ The size of the barcoding-gap depends on the taxonomic group.





Candek and Kuntner 2015

DNA Barcoding

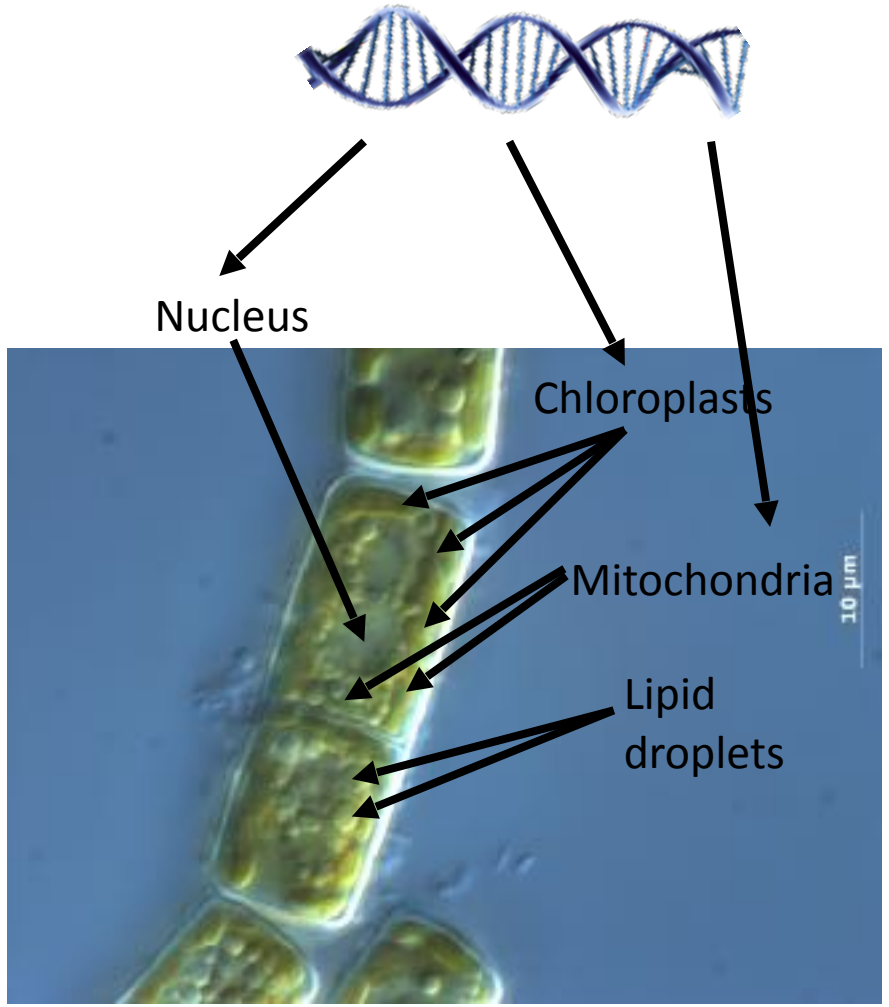
- Identification to species level using a single "universal" molecular marker s barcode... Not an achievable goal!

- Group-specific barcodes
  - Metazoans: COI
    - Fish: 12S
  - Fungi: ITS
  - Plants : *rbc*L et MATK
    - Diatoms : *rbc*L (18S)

DNA Barcoding: diatoms

Barcodes

- ***rbc*L** chloroplastic gene coding for the large sub-unit of the RuBisCo (photosynthesis)
- **CO1** mitochondrial gene coding for the cytochrome c oxydase I
- **rDNA 18S** nuclear gene coding for ribosomal RNA18S



DNA Barcoding: diatoms

Gomphonema parvulum :

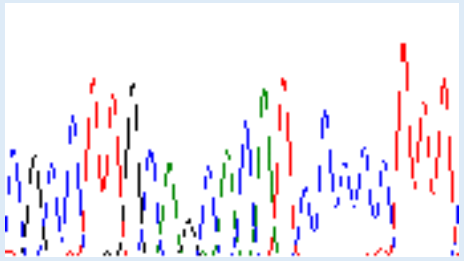
25 strains from tropical streams

11 strains from temperate streams

+ 1 *G. exilissimum* (tropical stream)

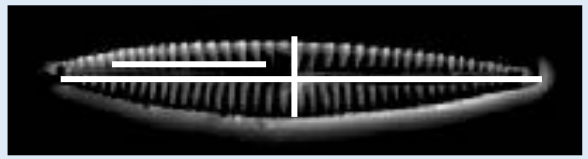
+ 1 *G.* cf. *lagenula* (tropical stream)

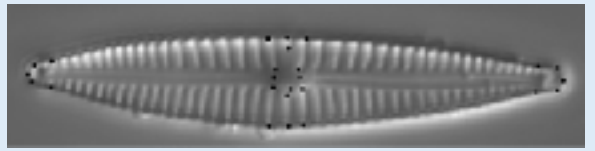
Sanger sequencing



Kermarrec et al., 2013

Morphotaxonomy





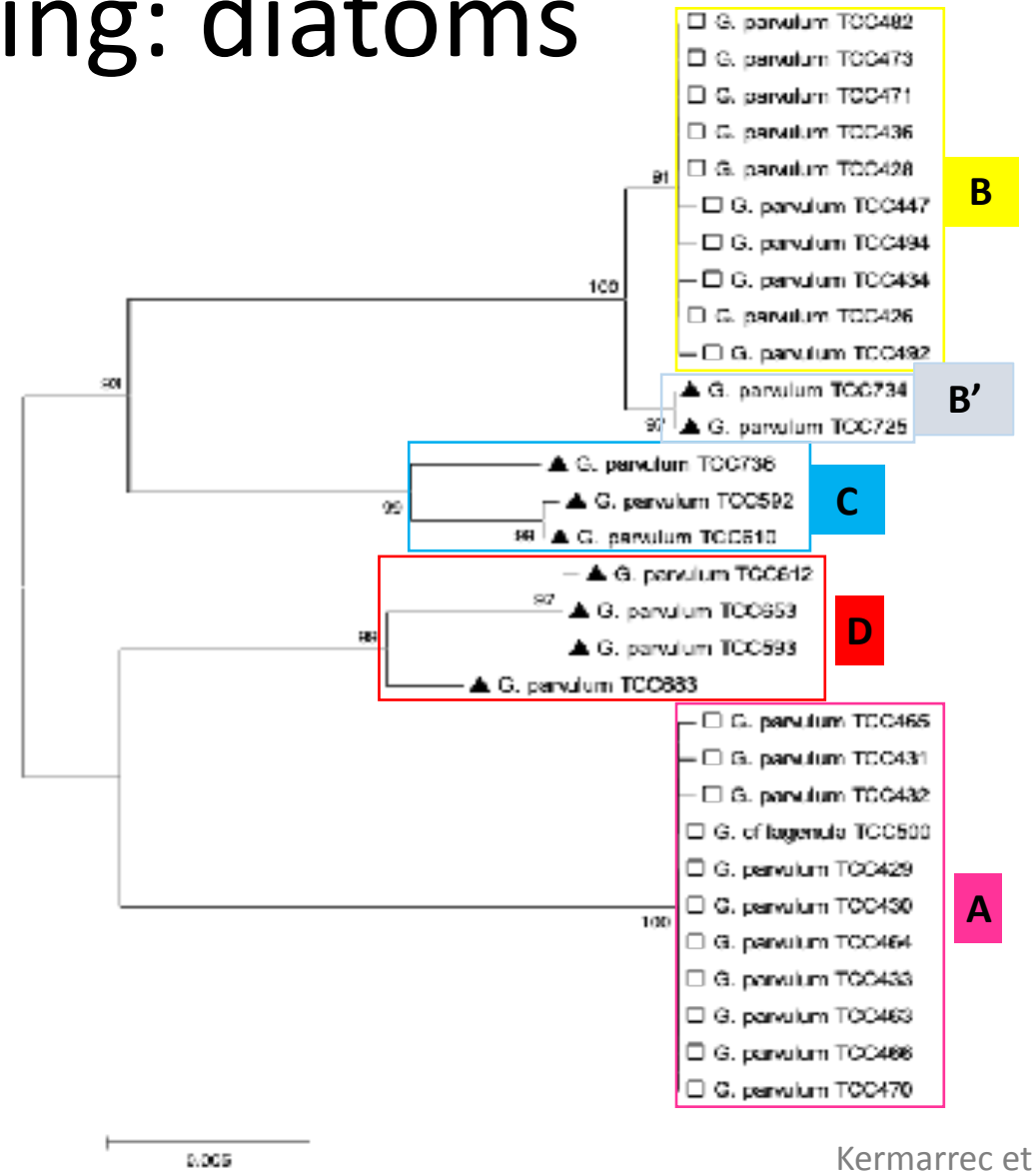
DNA Barcoding: diatoms

18S: allows defining the group of *G. parvulum* in the *Gomphonema* genera

cox1 / *rbc*L:

- 4 main clades: A, B/B', C et D

→ highlighting cryptic diversity in *G. parvulum*

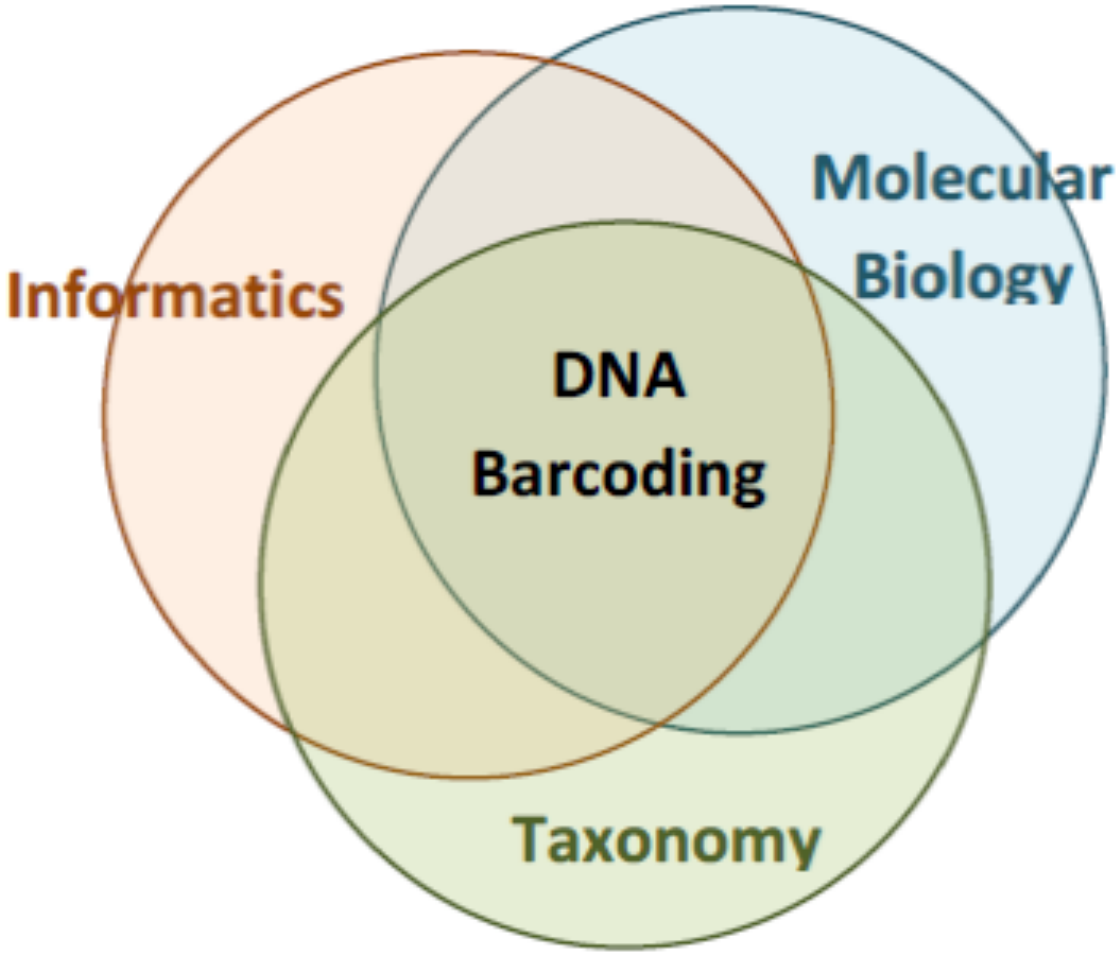


Kermarrec et al., 2013

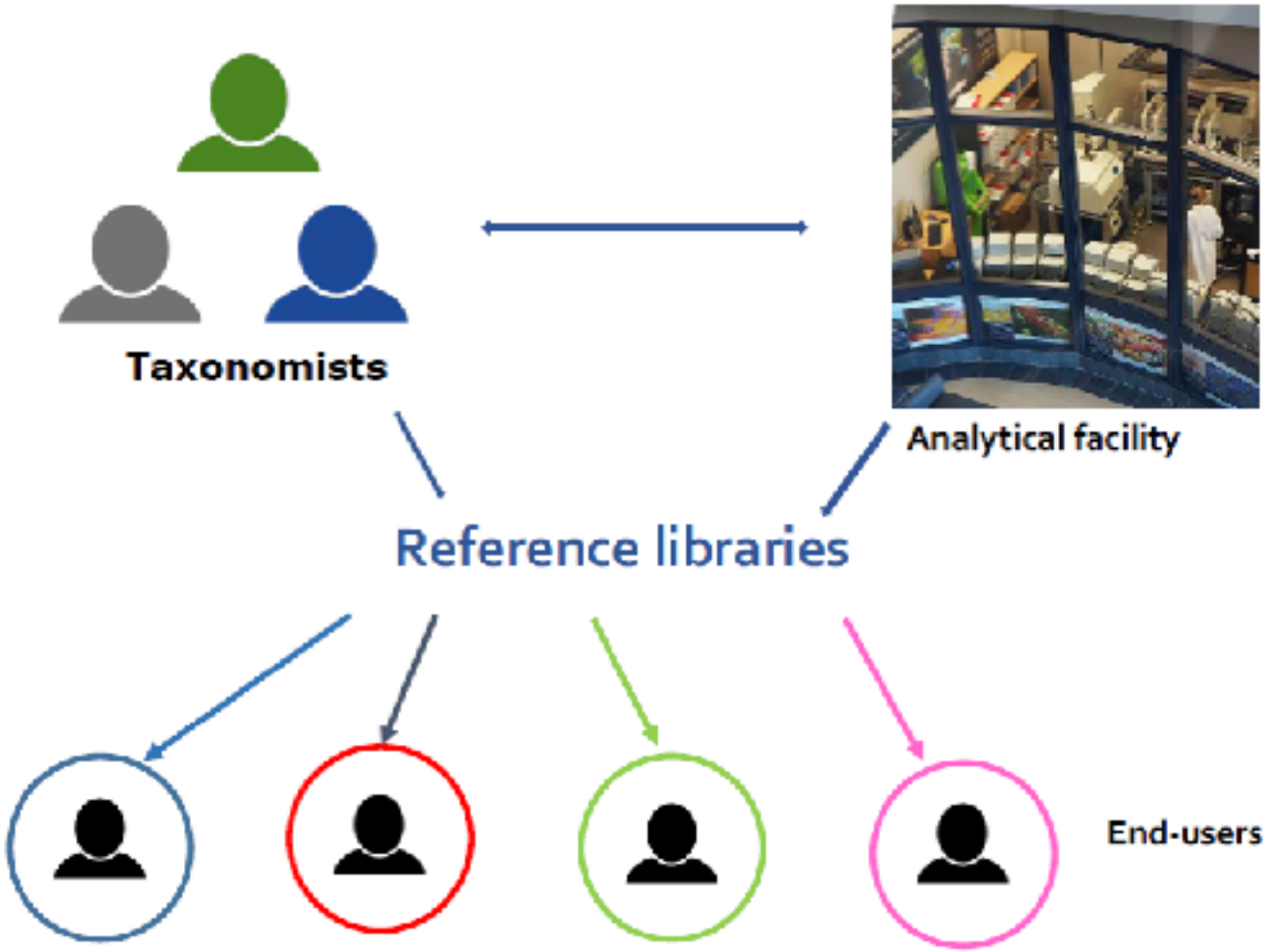
Funded by European Union

DNA Barcoding

A tool for taxonomic ID at the crossroads of several disciplines

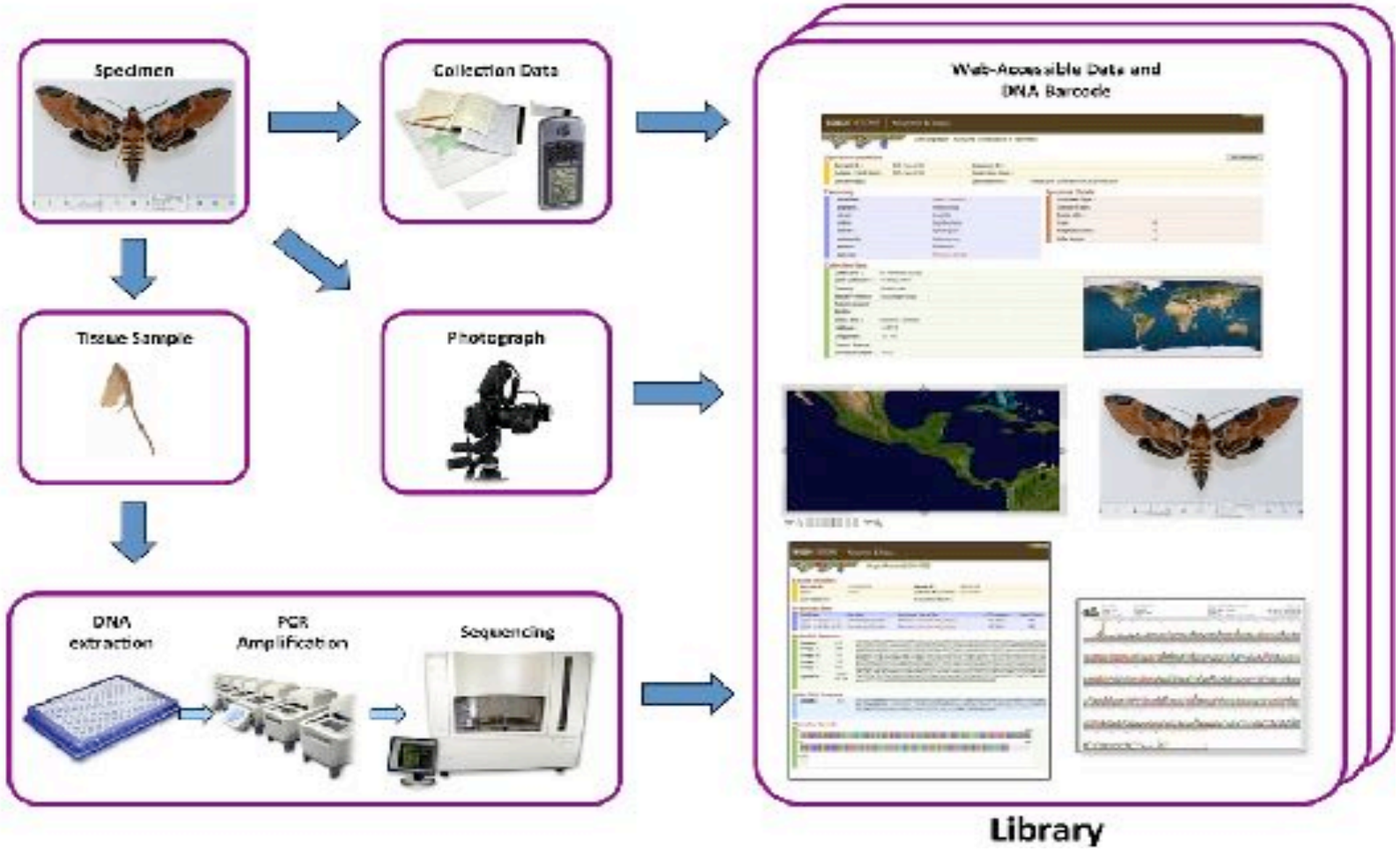


DNA Barcoding

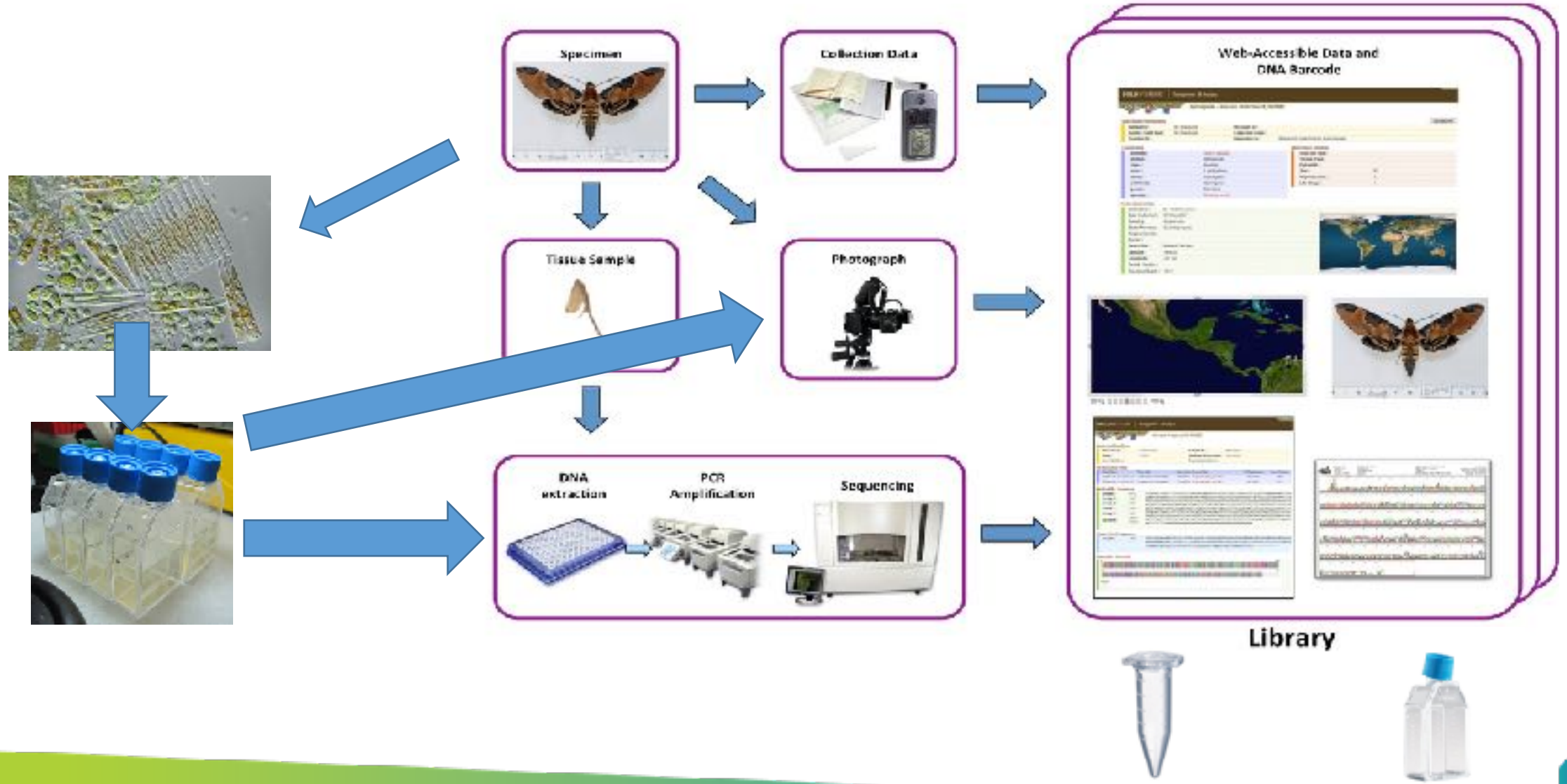


DNA Barcoding

Key point:
The Reference Library!



DNA Barcoding



Funded by European Union

DNA Barcoding: uses

Endangered species

Invasive alien species

Environmental monitoring

Food fraud

Wildlife crime

Pathogens and vectors

Identify:

- Invasive alien species
- Endangered species
- «*sensitive*» species *(commercial, biomedical, pathogenic…)*
- Indicator species or key species

For the purposes of:

- Quarantine inspection
- Fraud control
- Environmental assessment
- Biodiversity conservation

Funded by European Union

DNA Barcoding: uses

Detection of morphologically 'unidentifiable' organisms

→ early detection

- Incomplete specimens (tissue, fragment…) *e.g. paleo samples*
- Degraded specimens (feces, food…) *e.g. gut content*
- Young stages (larva) *e.g. macroinvertebrate inventories*

But also:

- **Access to undescribed diversity**
- **Access to cryptic diversity**

Funded by European Union

BIOLAWEB

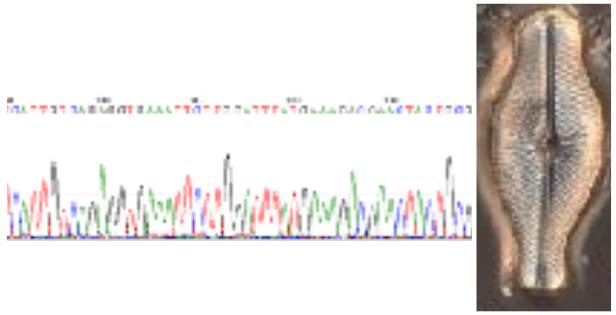
Summary

- DNA - some basics
- DNA and barcoding
- **DNA and metabarcoding**

DNA Barcoding → DNA Metabarcoding

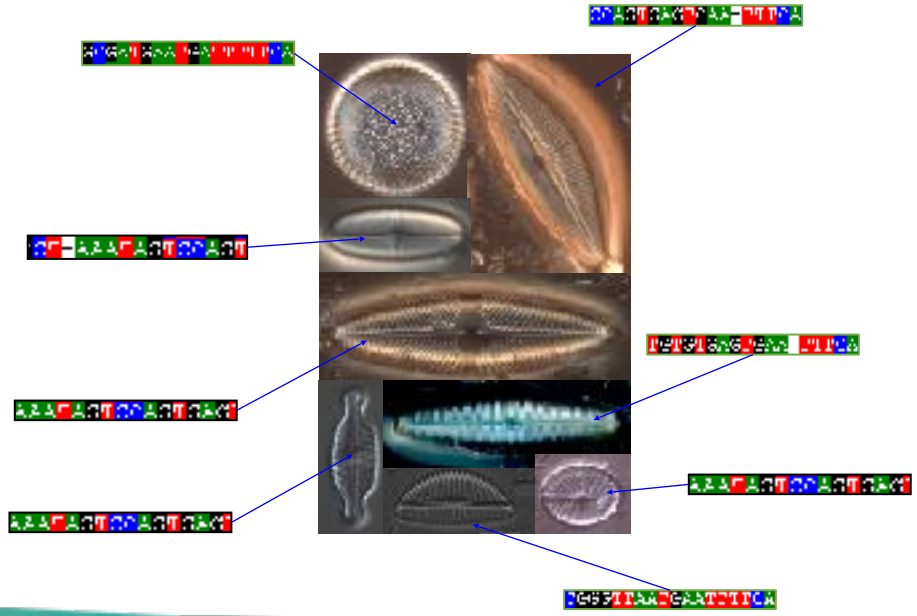
- **2003: DNA Barcoding** (Hebert et al. 2003)

Standardised DNA-based ID tool
Barcode: short DNA fragment
taxa-specific
easy to sequence



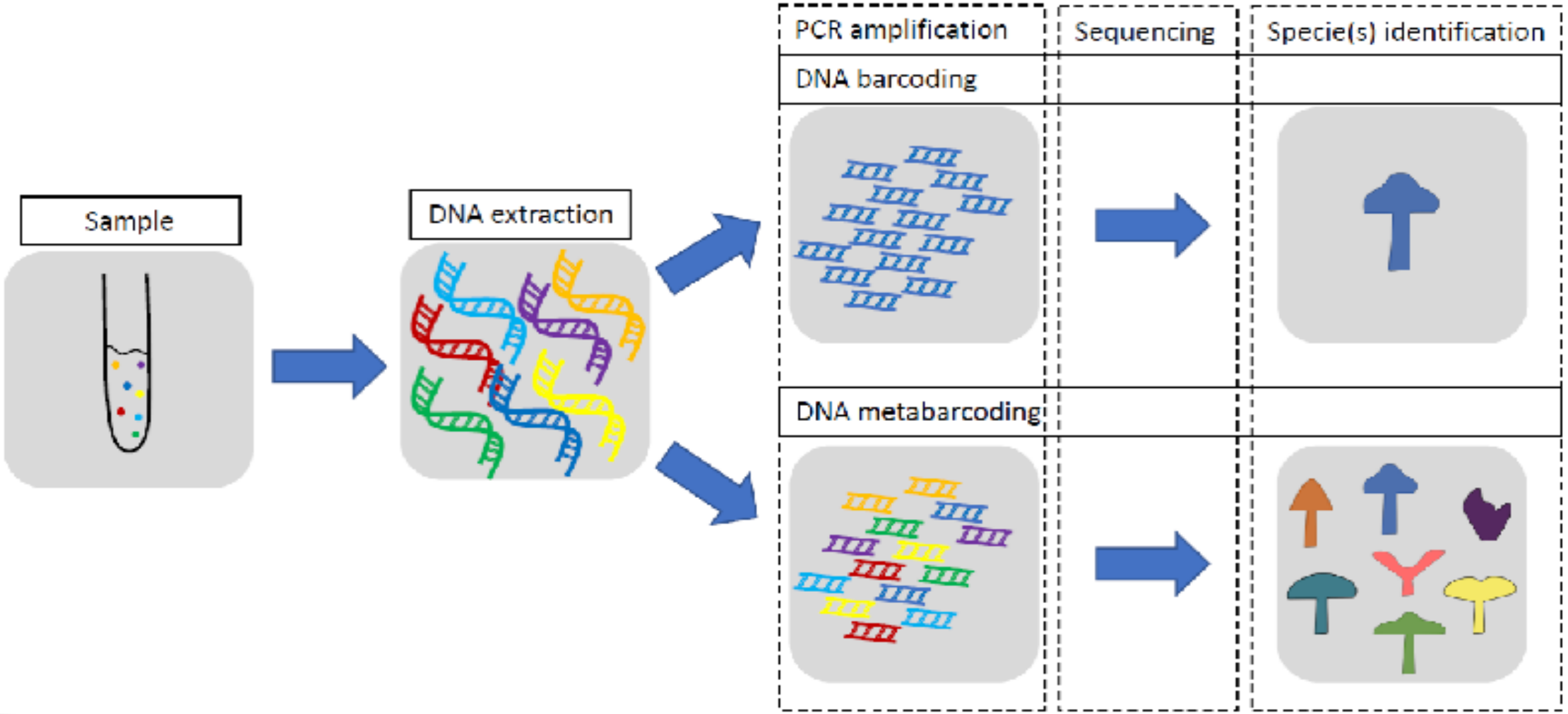
- **2012: DNA Metabarcoding** (Taberlet et al. 2012)

- Extending the concept to the study of natural samples
- Identification of organisms within a community
- Use of "next-generation", "high-throughput" sequencing



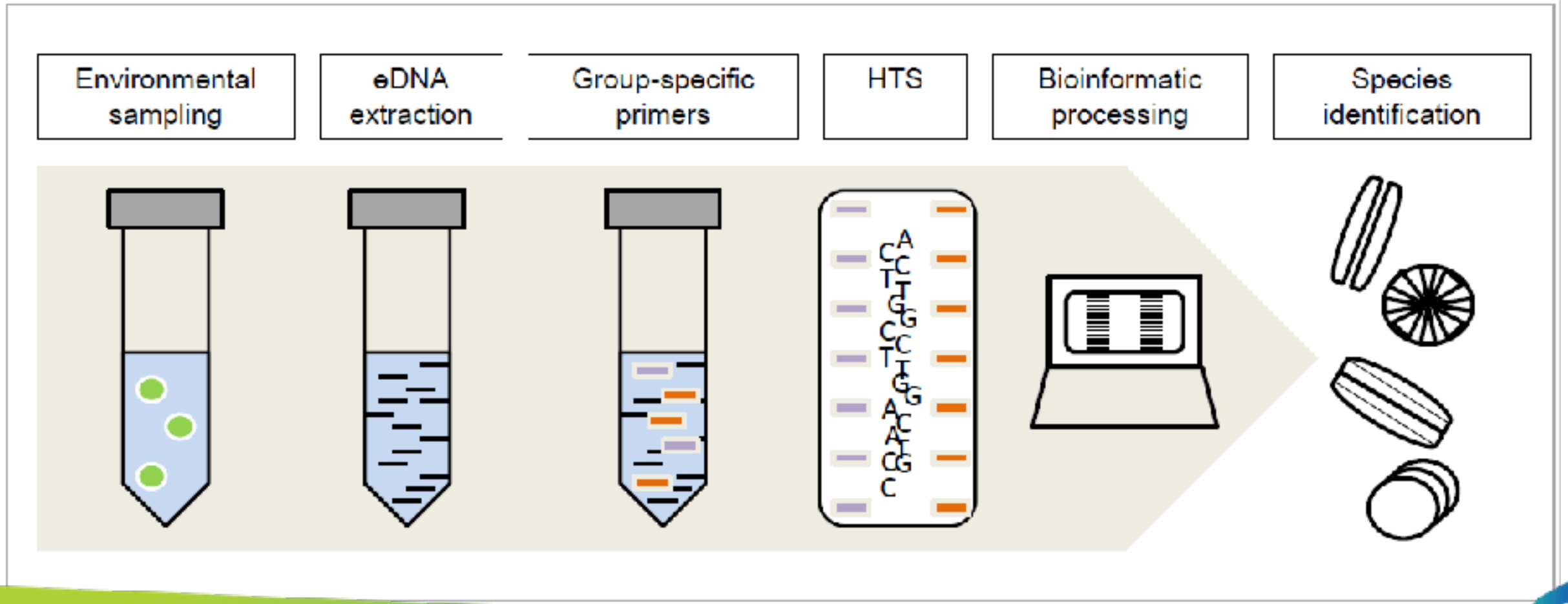
>> Benefits: Reduce analysis time, reduce costs, facilitate intercalibration

Barcoding / Metabarcoding

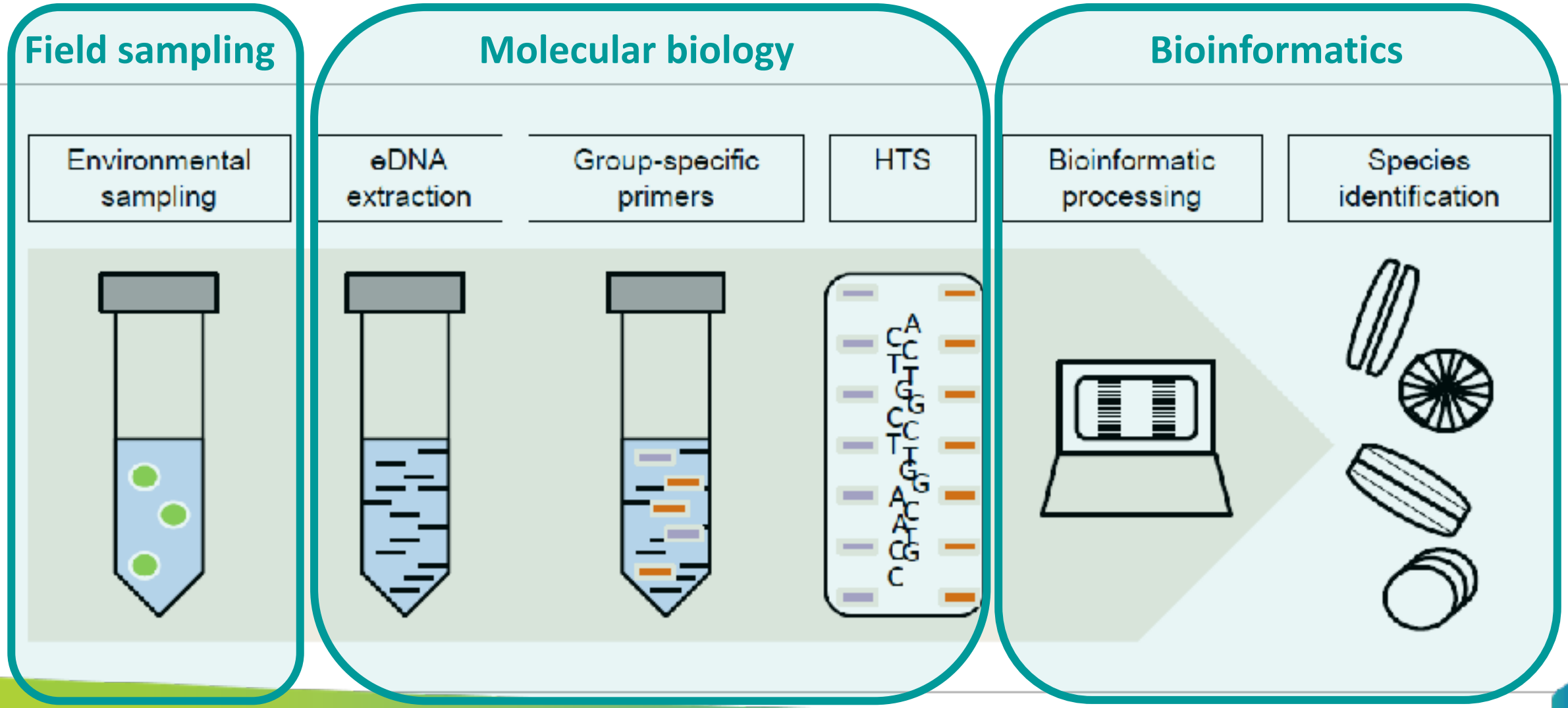


Funded by European Union

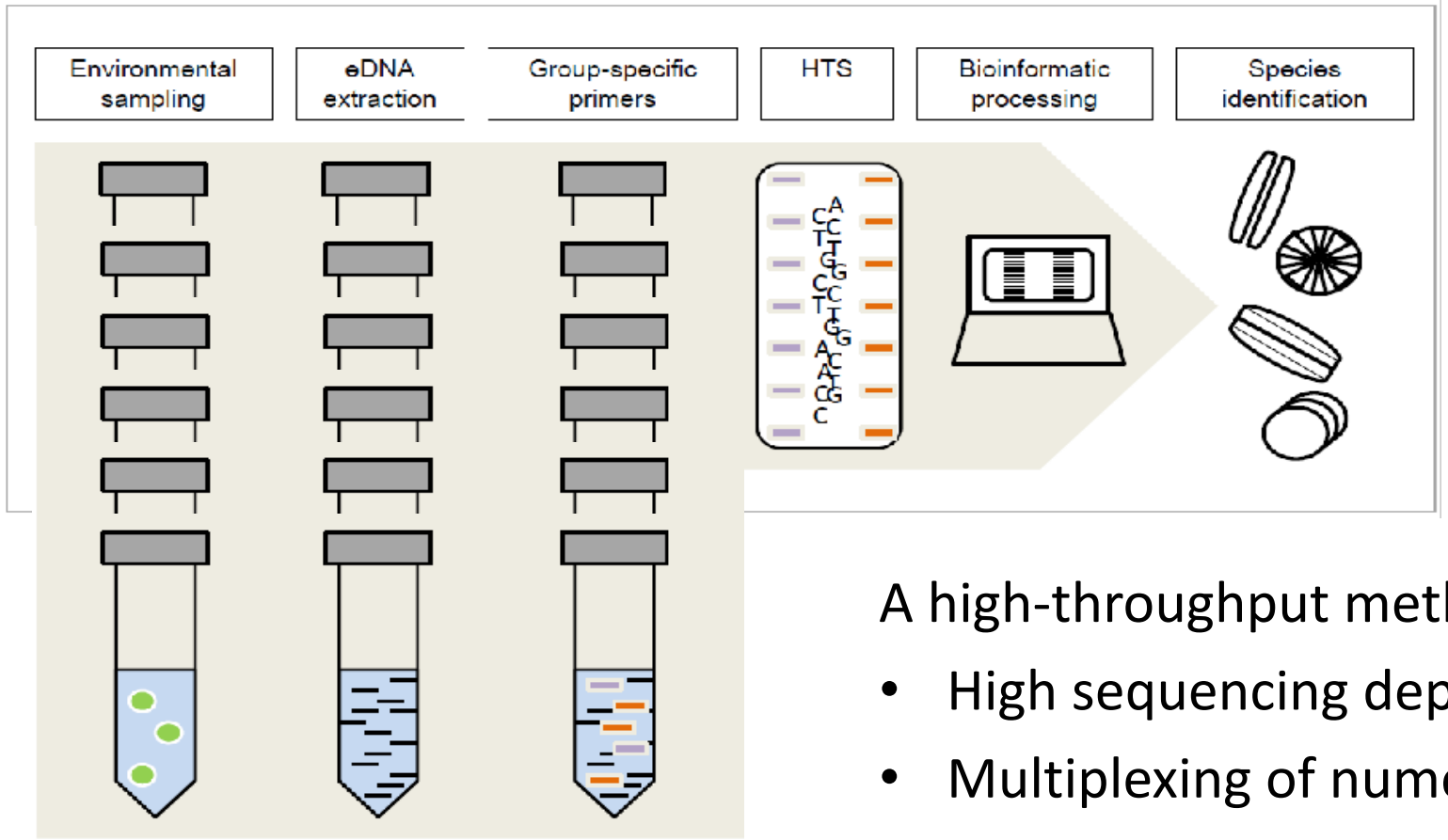
DNA Metabarcoding



DNA Metabarcoding



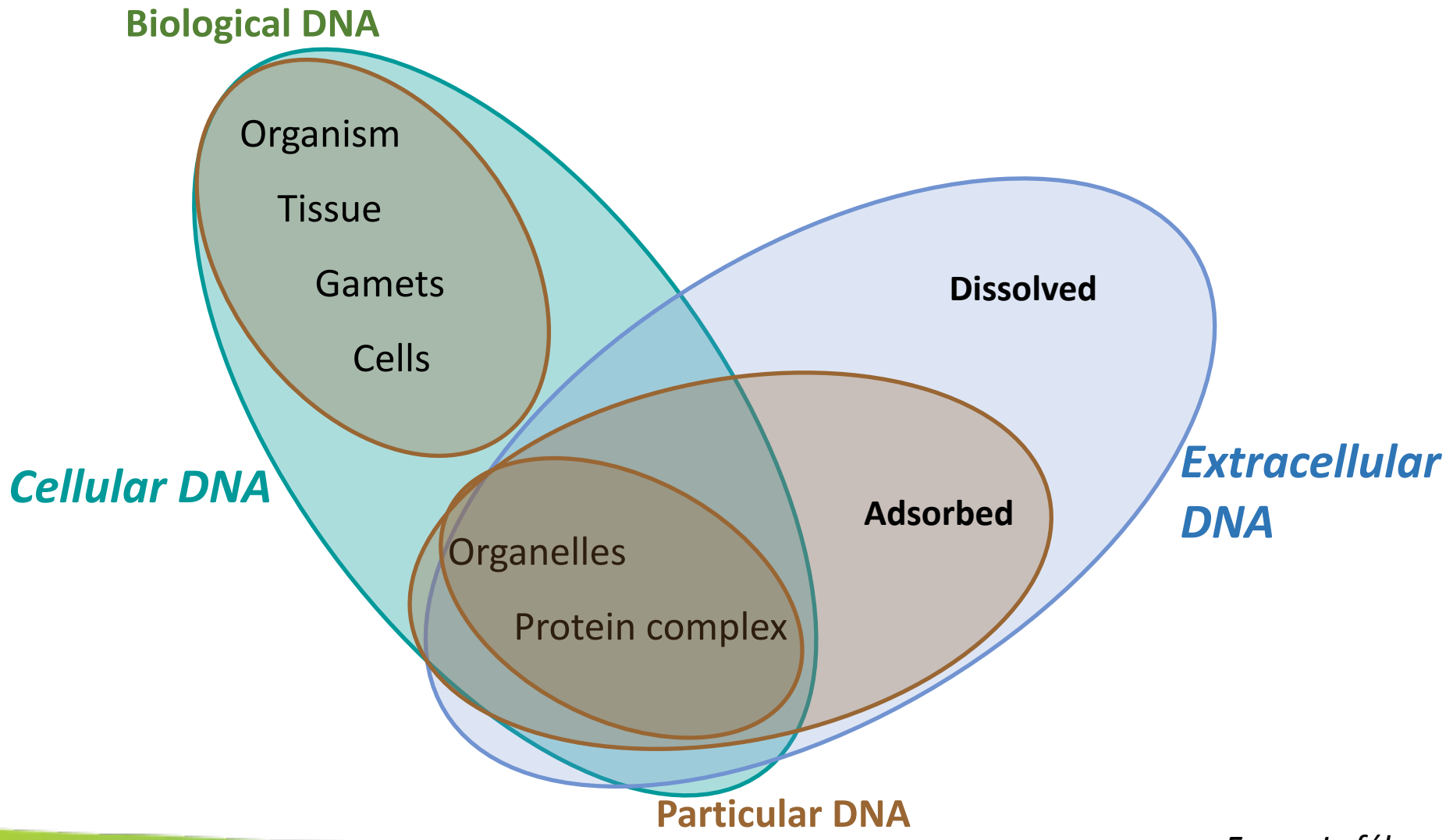
DNA Metabarcoding



A high-throughput method

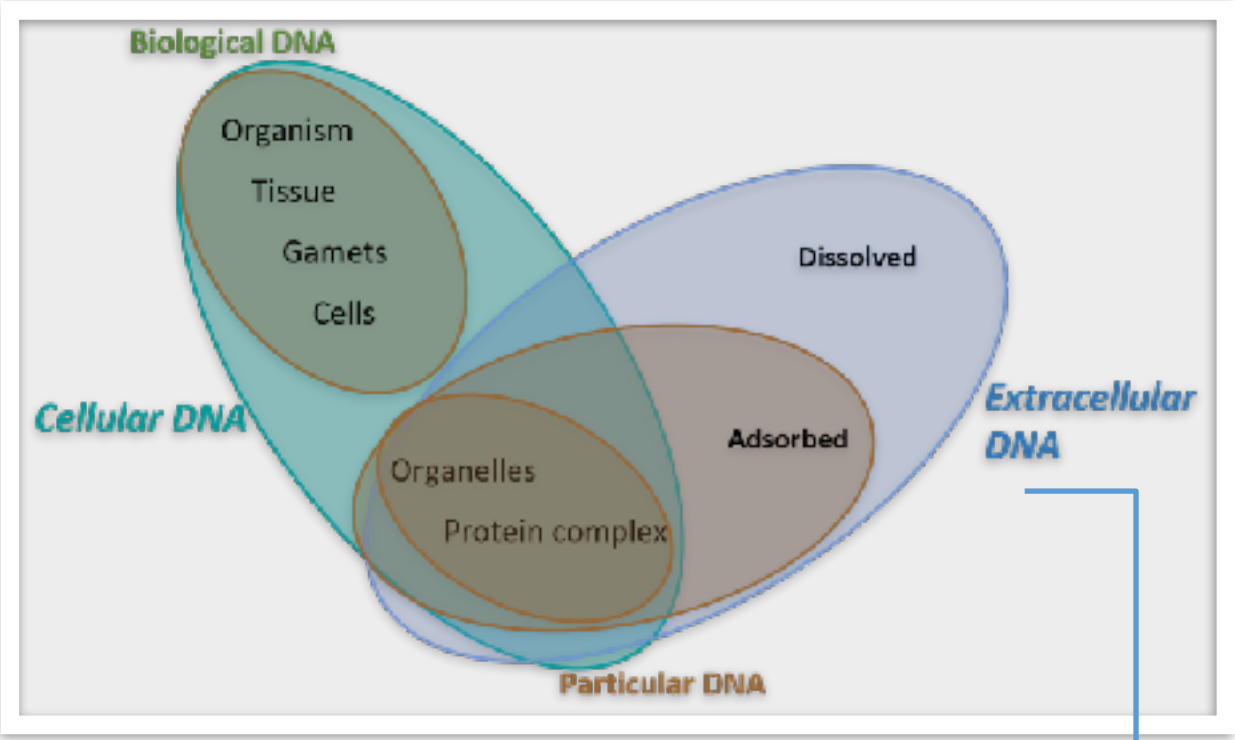
- High sequencing depth per sample
- Multiplexing of numerous samples

Metabarcoding: DNA material



From Lefébure - LEHNA

Metabarcoding: DNA material



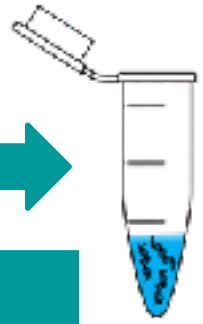
Environmental DNA / eDNA ?

in the broad sense: all DNA in environmental samples

in the strict sense: extracellular DNA in environmental samples

DNA Metabarcoding: uses

sampling
DNA extraction



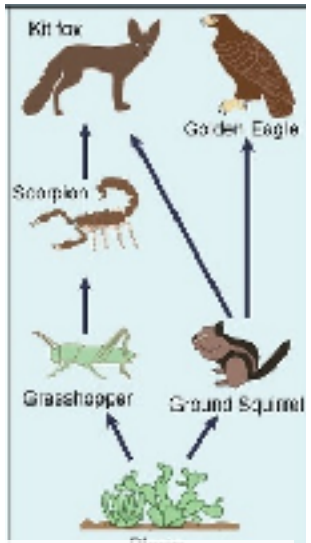
amplification
sequencing
bioinformatics

plant diversity

pollution responses

biomonitoring

invasive species

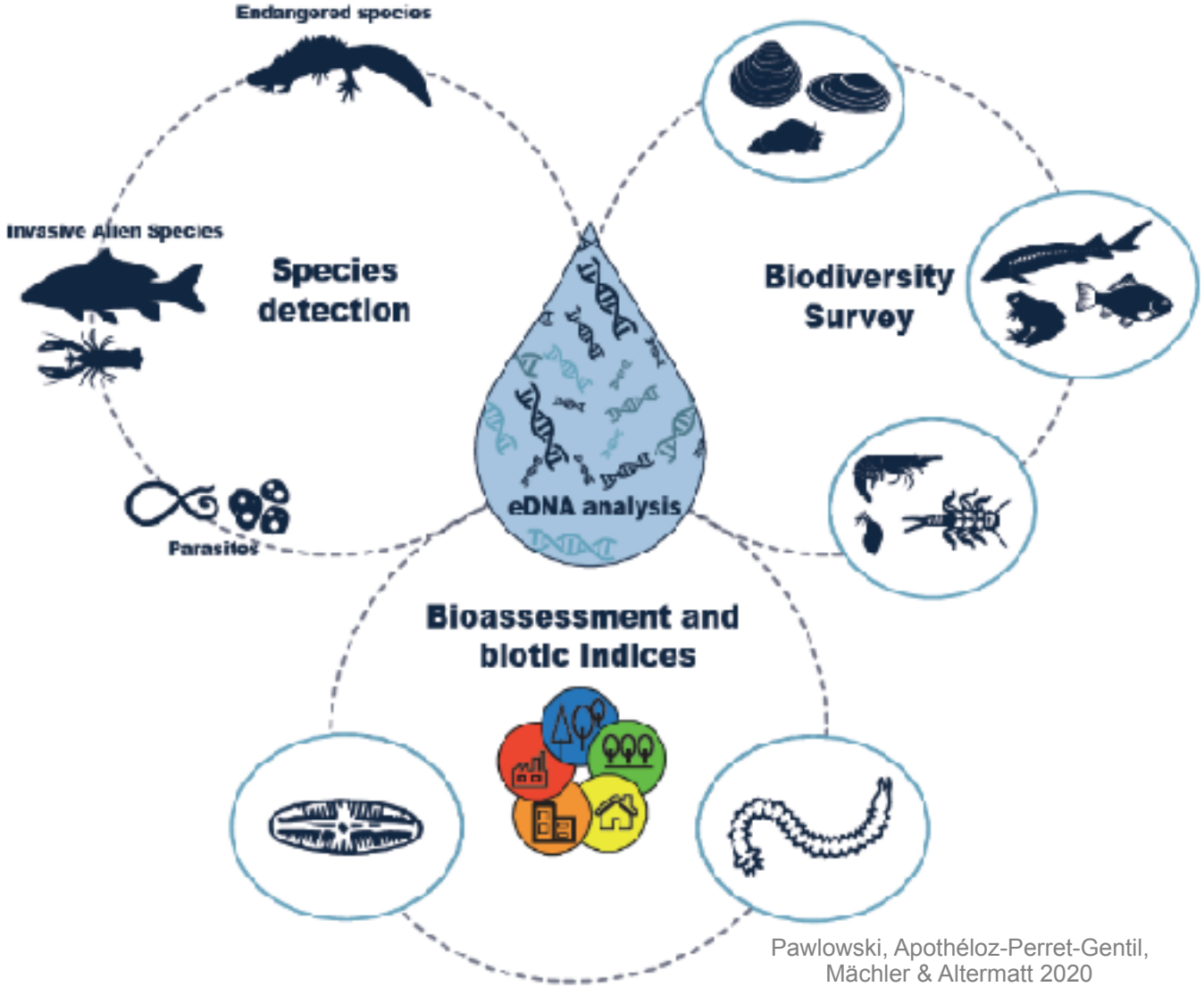


diet analysis

ancient ecosystem

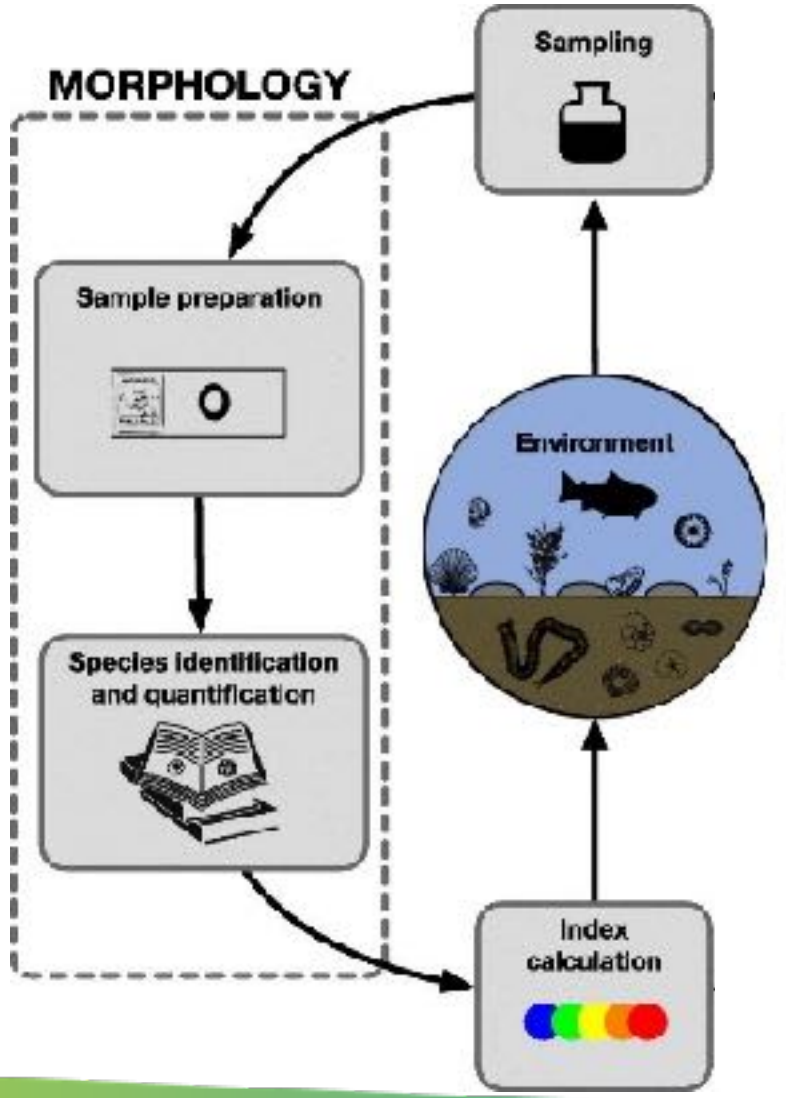
DNA Metabarcoding: uses

→ Many uses for eDNA in aquatic environments

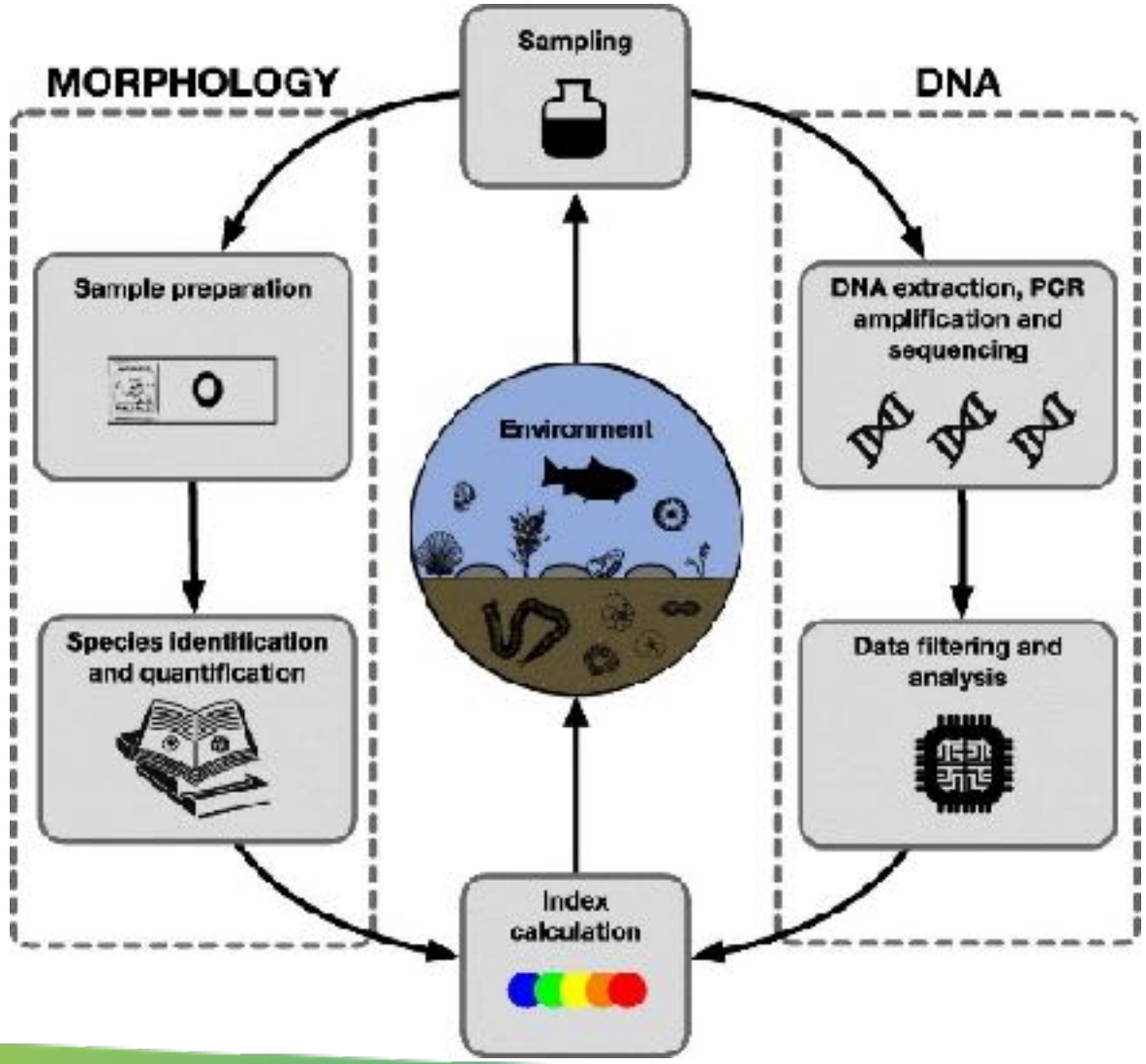


Pawlowski, Apothéloz-Perret-Gentil, Mächler & Altermatt 2020
Swiss Federal Office for the Environment

DNA Metabarcoding: uses



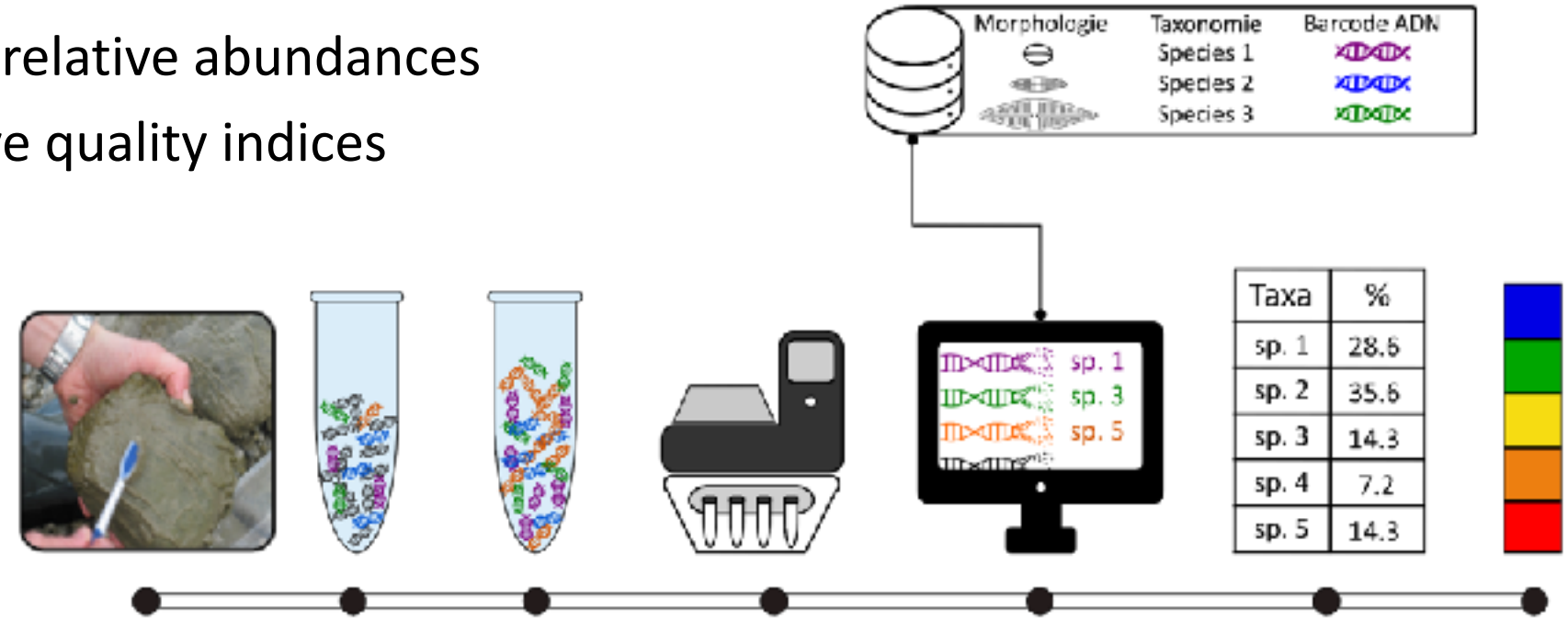
DNA Metabarcoding: uses



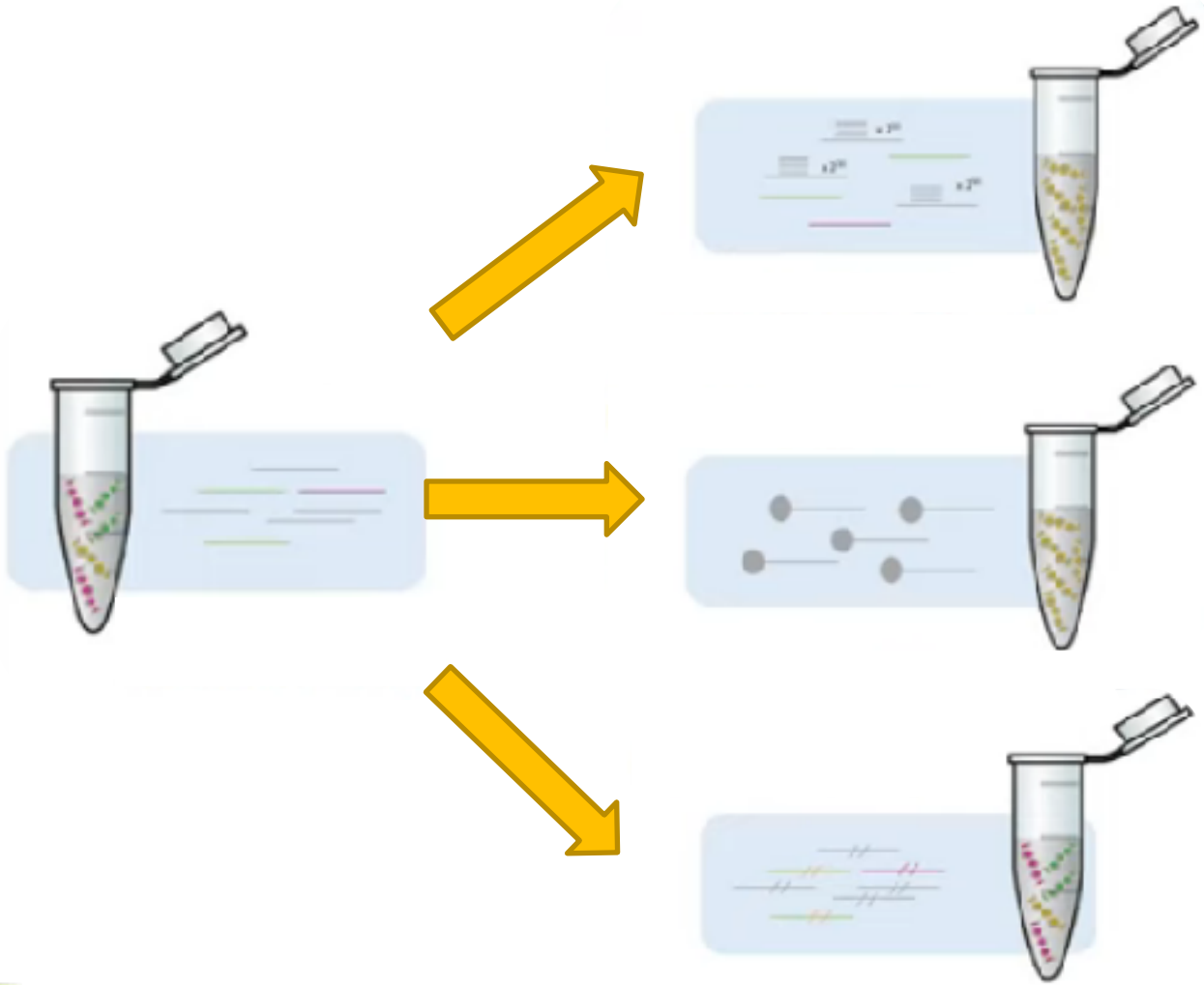
DNA Metabarcoding: uses

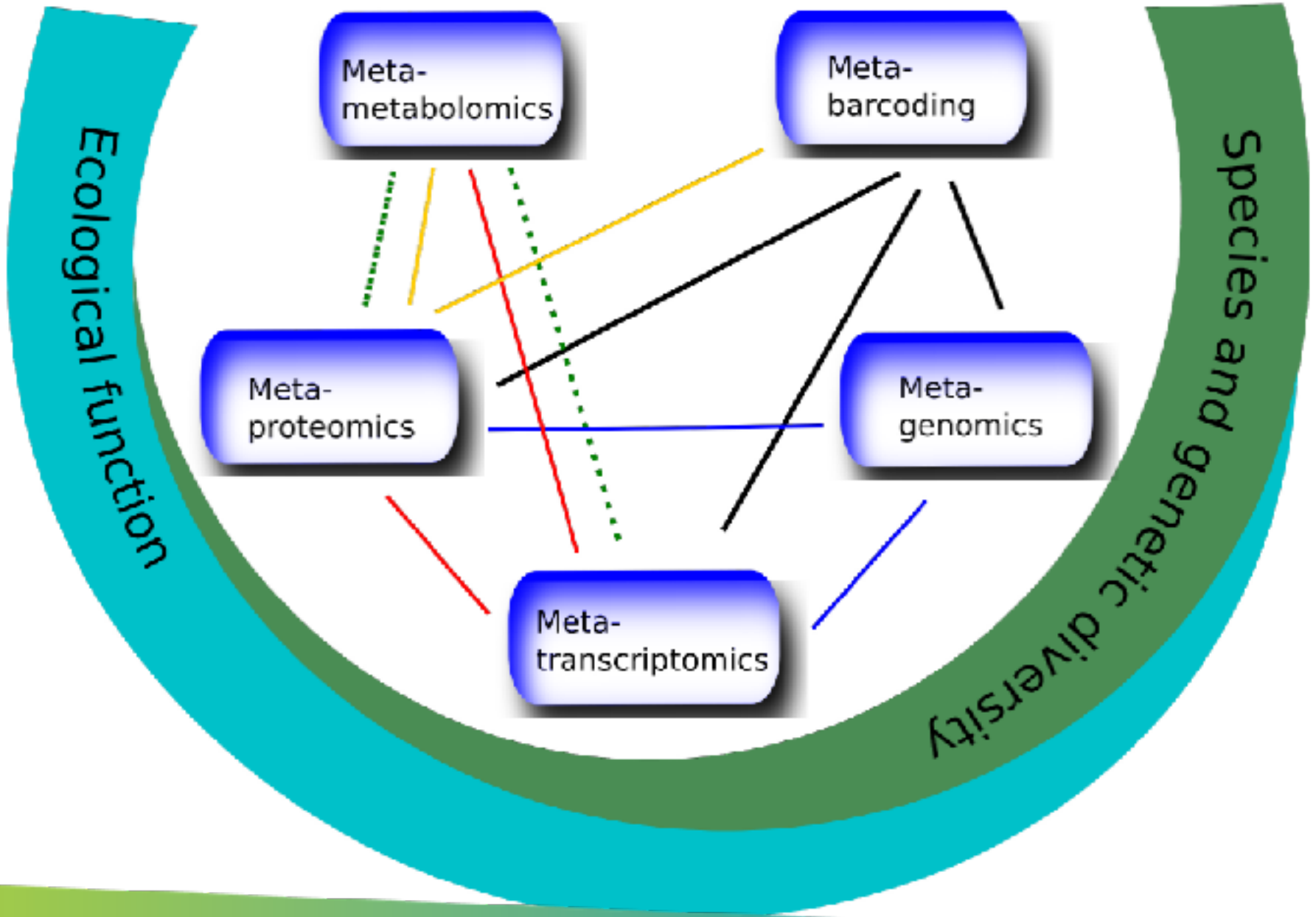
Diatoms: biological quality element (BQE)

- taxonomic inventories
- with relative abundances
- derive quality indices



DNA Metabarcoding: uses




Ecological function
Species and genetic diversity
Meta-metabolomics
Meta-barcoding
Meta-proteomics
Meta-genomics
Meta-transcriptomics

Questions ?