University of Belgrade
Institute of Chemistry, Technology and Metallurgy
National Institute of the Republic of Serbia

www.biolaweb.com

BIOLAWEB

Acronym: BIOLAWEB Boosting Institute of Chemistry, Technology and Metallurgy in Water Biomonitoring

Grant No: 101079234

Type of action: HORIZON Coordination and Support Actions (HORIZON - CSA)

Duration: 36 months

eDNA Workshop

Barcoding/Metabarcoding

Belgrade, October 2023

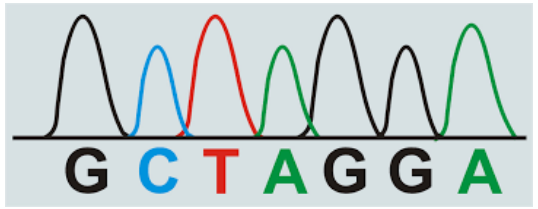
BIOLAWEB
presentation

Funded by
the European Union

www.biolaweb.com

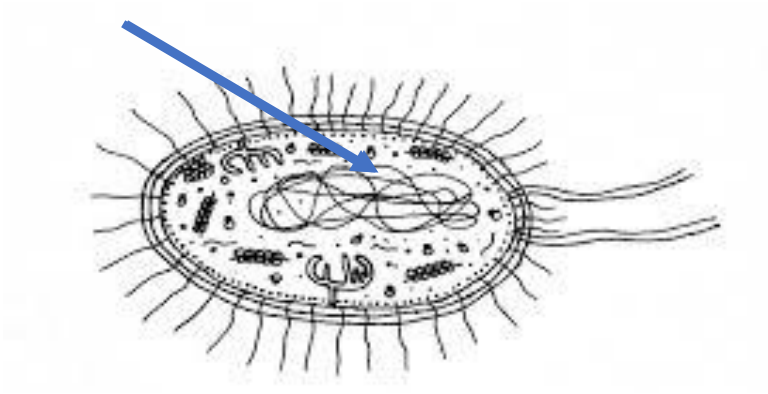
What is a barcode?

a machine-readable code: in the form of numbers and a pattern of parallel lines to identify a certain item,

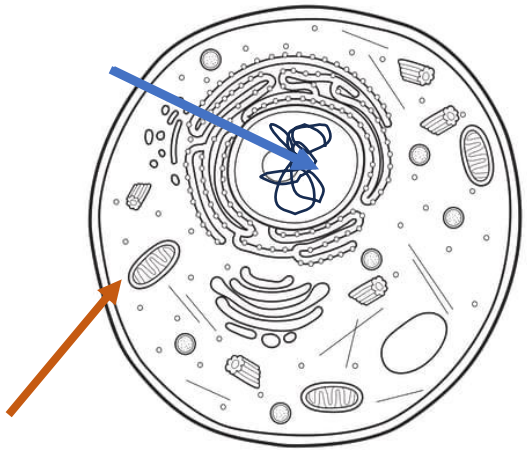


to identify a certain gene/organism

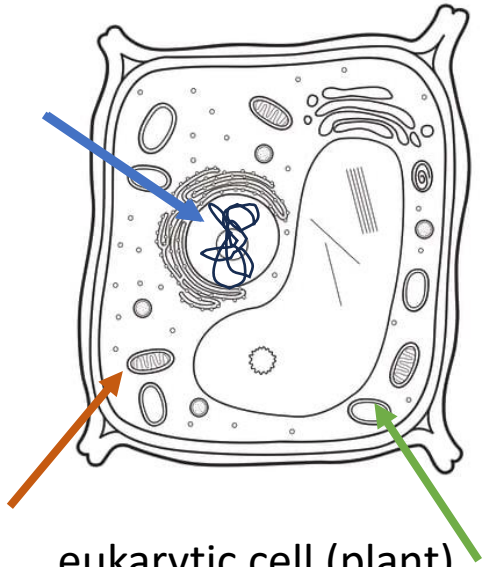
Barcode



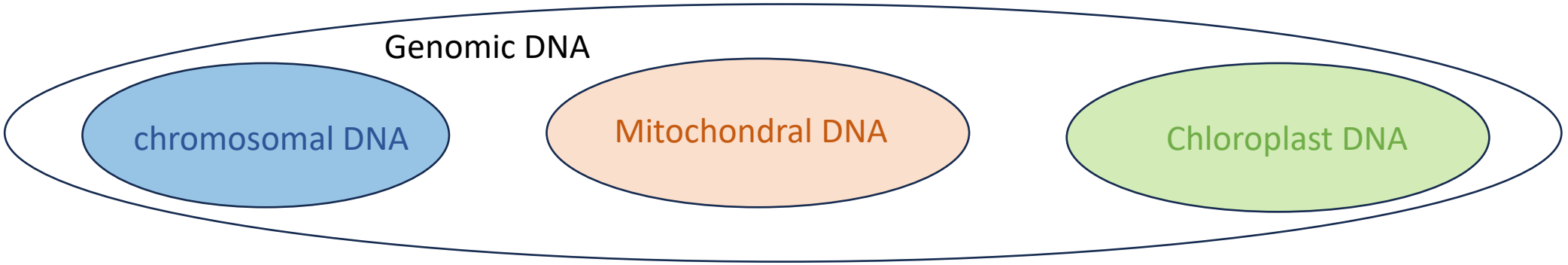
prokaryotic cell



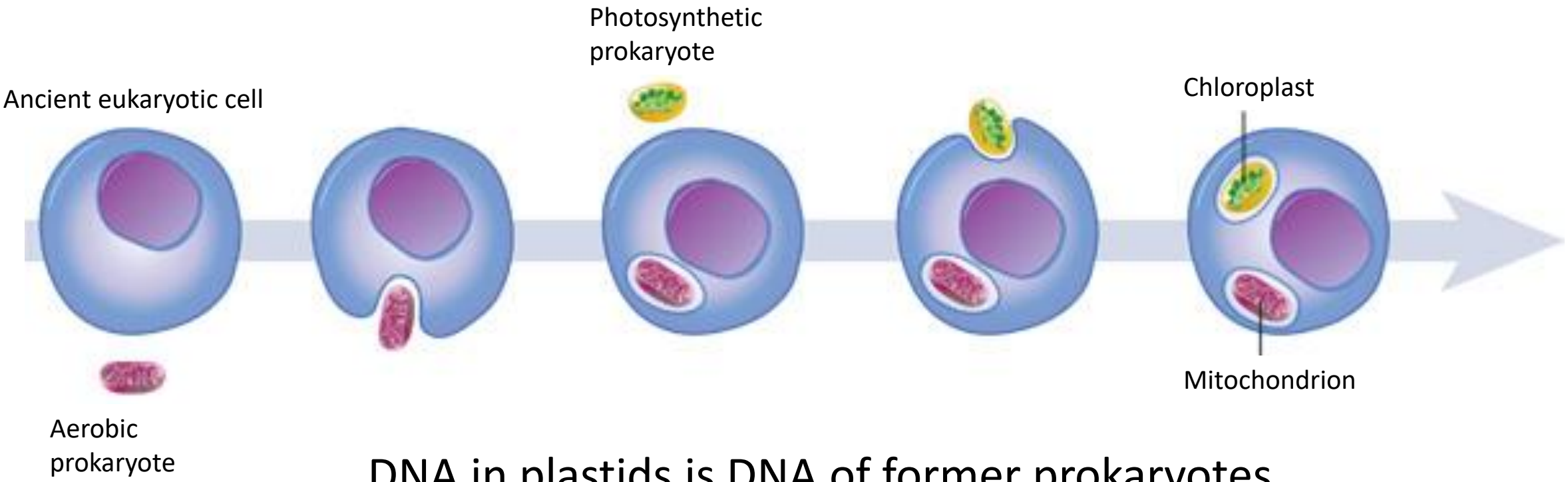
eukarytic cell (animal)



eukarytic cell (plant)



Barcode



DNA in plastids is DNA of former prokaryotes

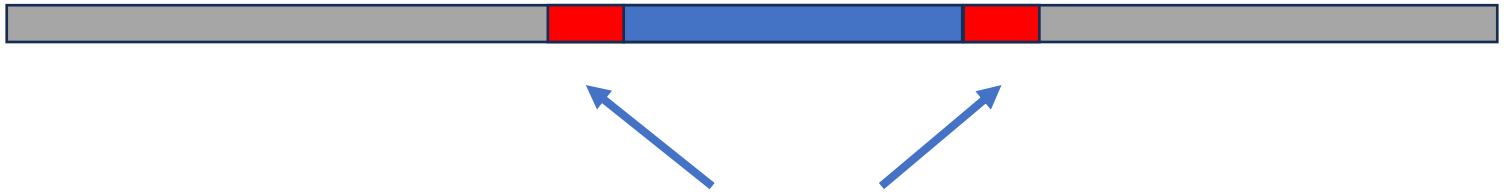
BIOLAWEB

Barcode

Group of organisms	Marker gene/locus used for barcoding
Animals	COI, *Cytb*, 12S, 16S
Plants	*matK*, *rbcL*, *psbA-trnH*, ITS
Bacteria	COI, *rpoB*,16S, *cpn60*, *tuf*, RIF, *gnd*
Fungi	ITS,TEF1α, RPB1 (LSU), RPB2 (LSU), 18S (SSU)
Protists	ITS, COI, *rbcL*, 18S, 28S

Barcode selection

marker gene
= barcode
(variable region)

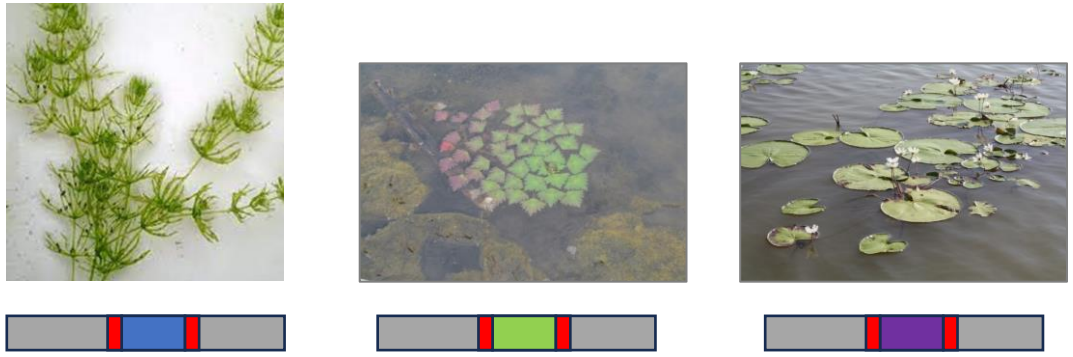


Primer
(conserved region)

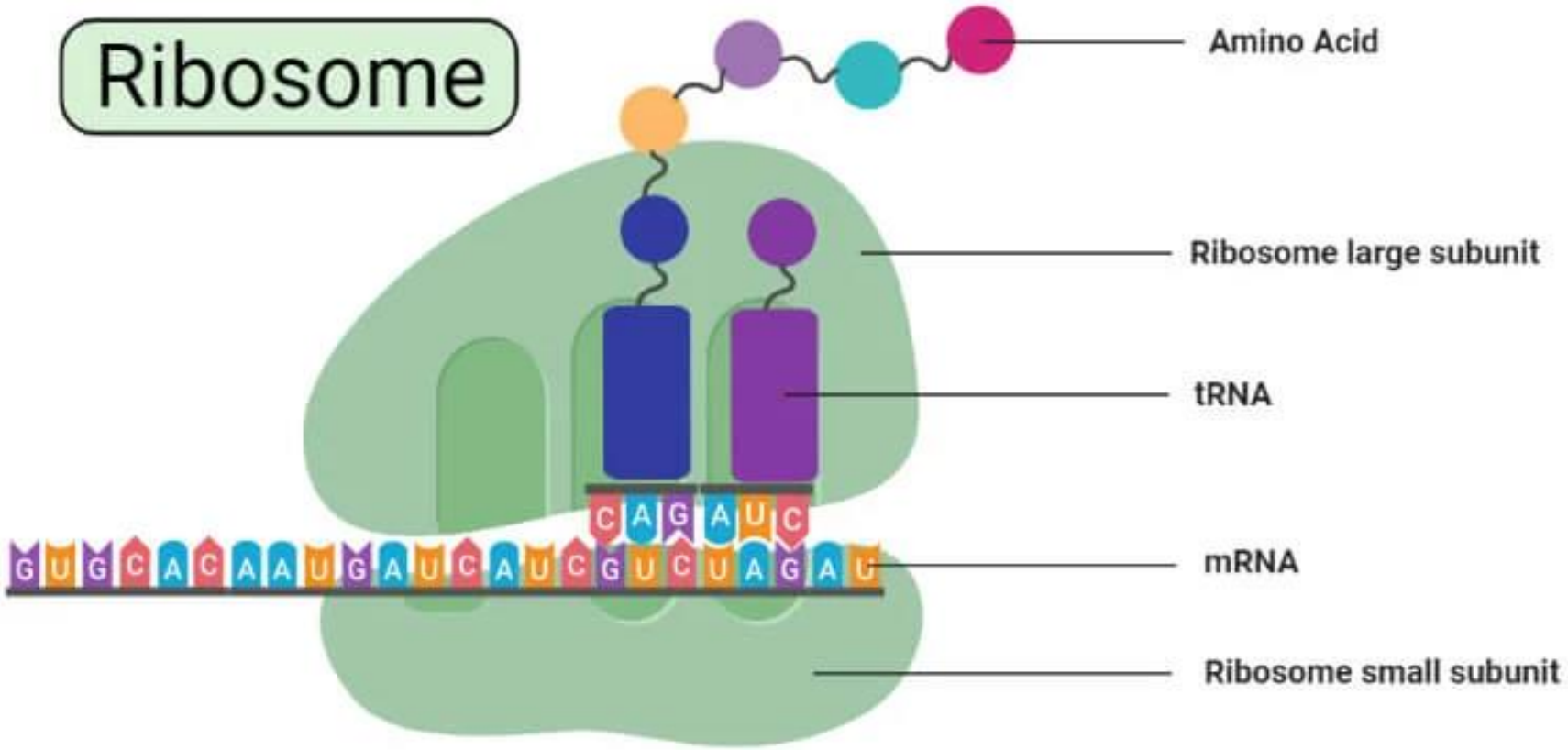
(i) significant interspecific genetic variability

(ii) conserved flanking sites for developing universal PCR primers

(iii) rel. short sequence length depending on sequence technology

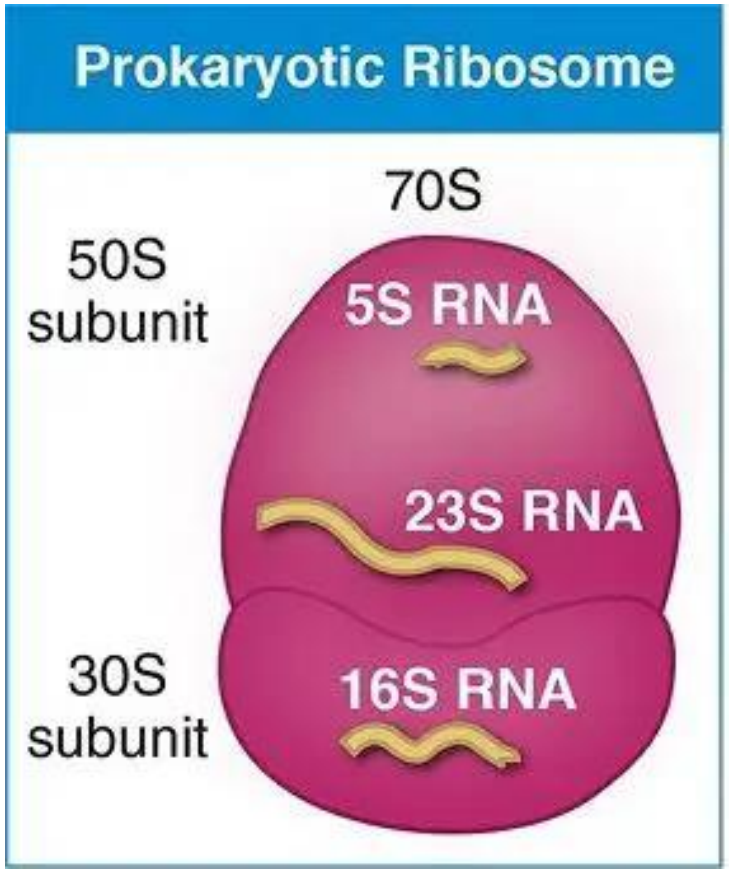


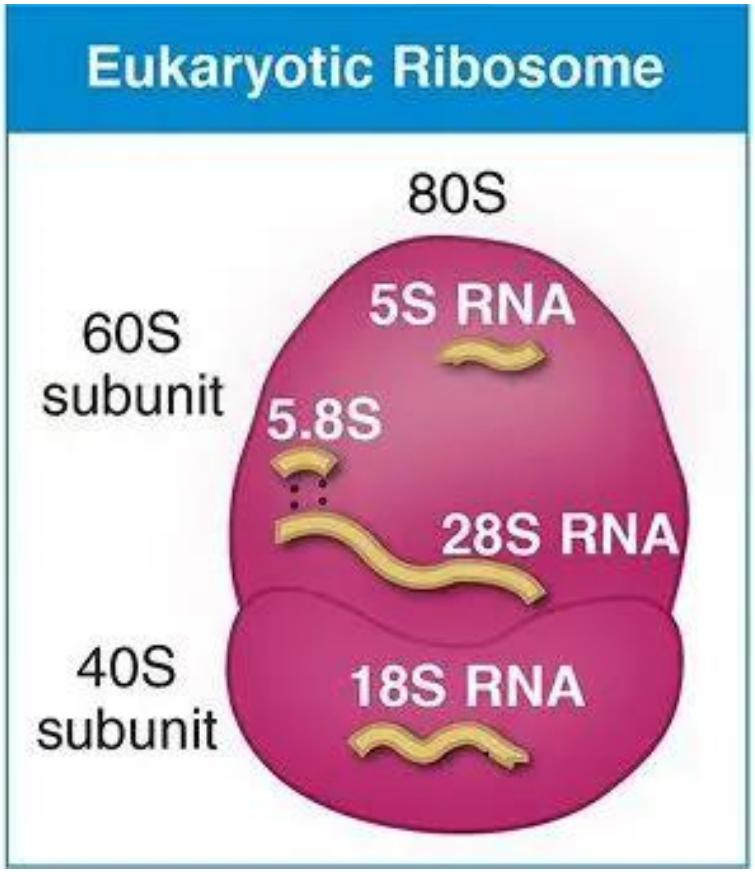
Example: Marker genes like 16S can be found in all organisms or like 18S in eukaryotic cells



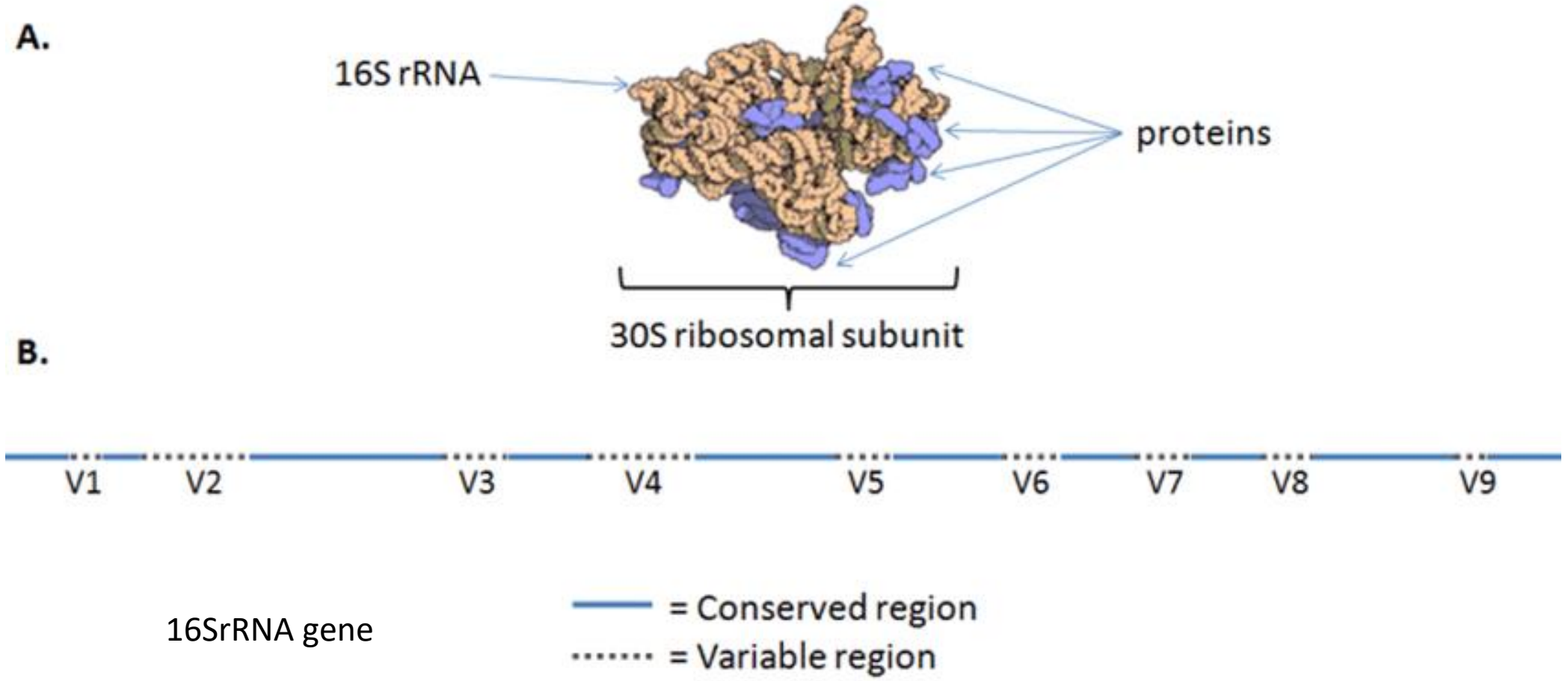
Barcode

16S rRNA gene is coding for 16S rRNA not for proteins!



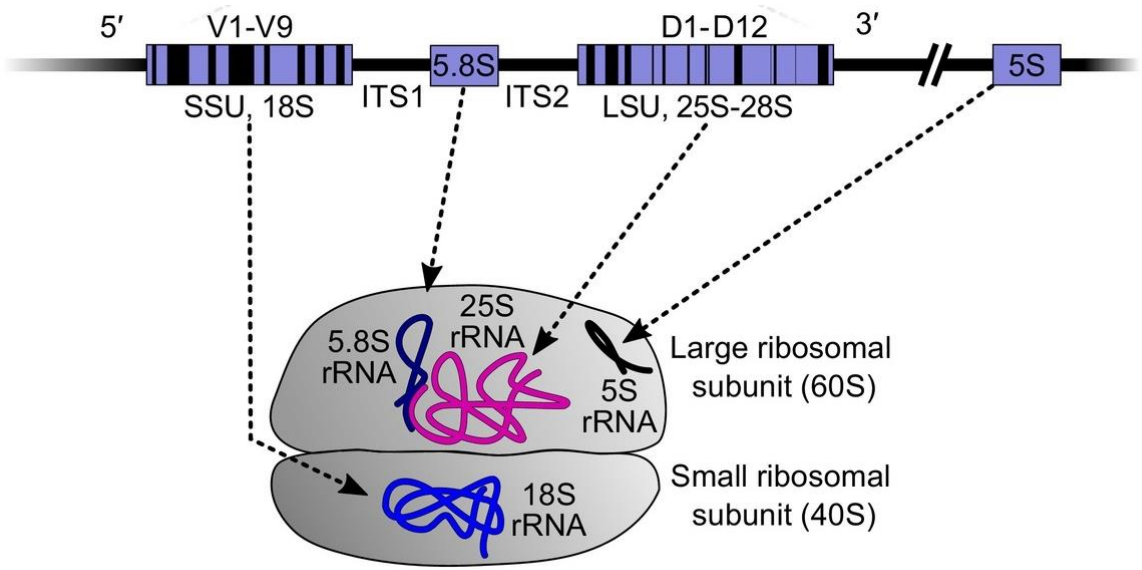


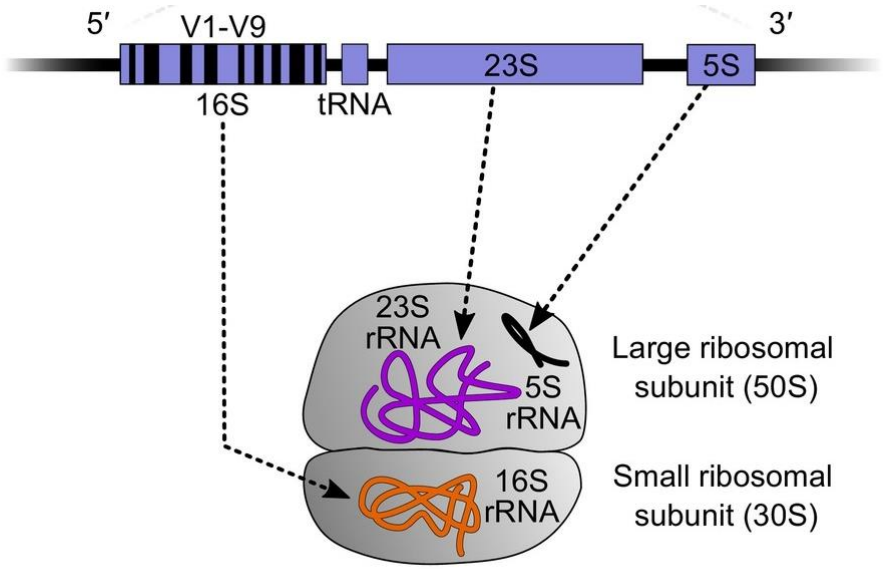
Barcode

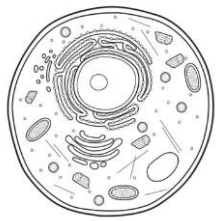

A.
16S rRNA
proteins
30S ribosomal subunit
B.
V1
V2
V3
V4
V5
V6
V7
V8
V9
16SrRNA gene
= Conserved region
= Variable region

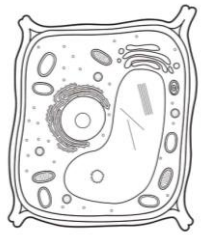
Barcode

Ribosomal genes
- organized in an operon

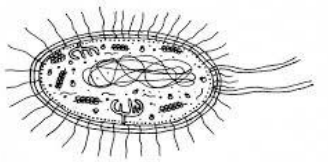








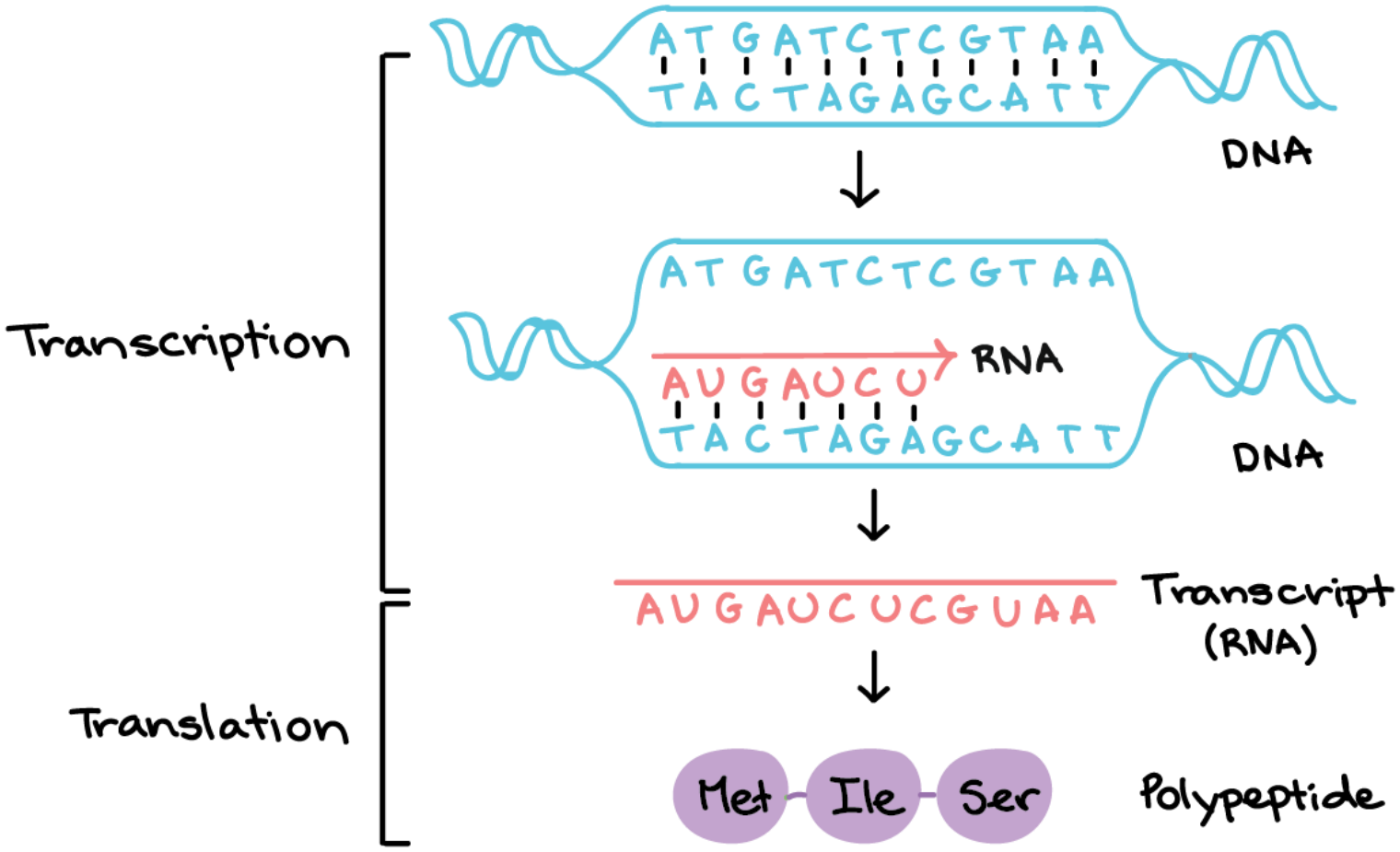
eukarytic cell (animals, plants)

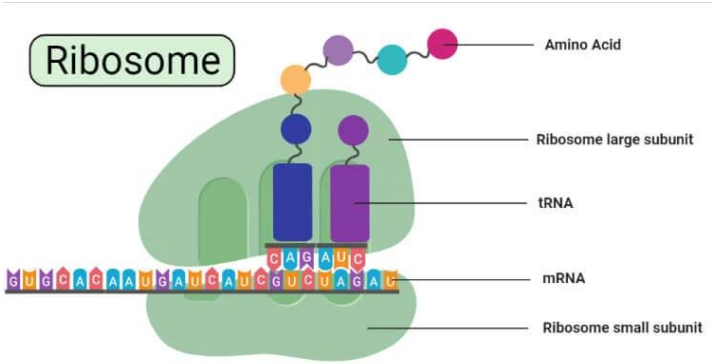


prokaryotic cell

Barcode

DNA is important in several processes





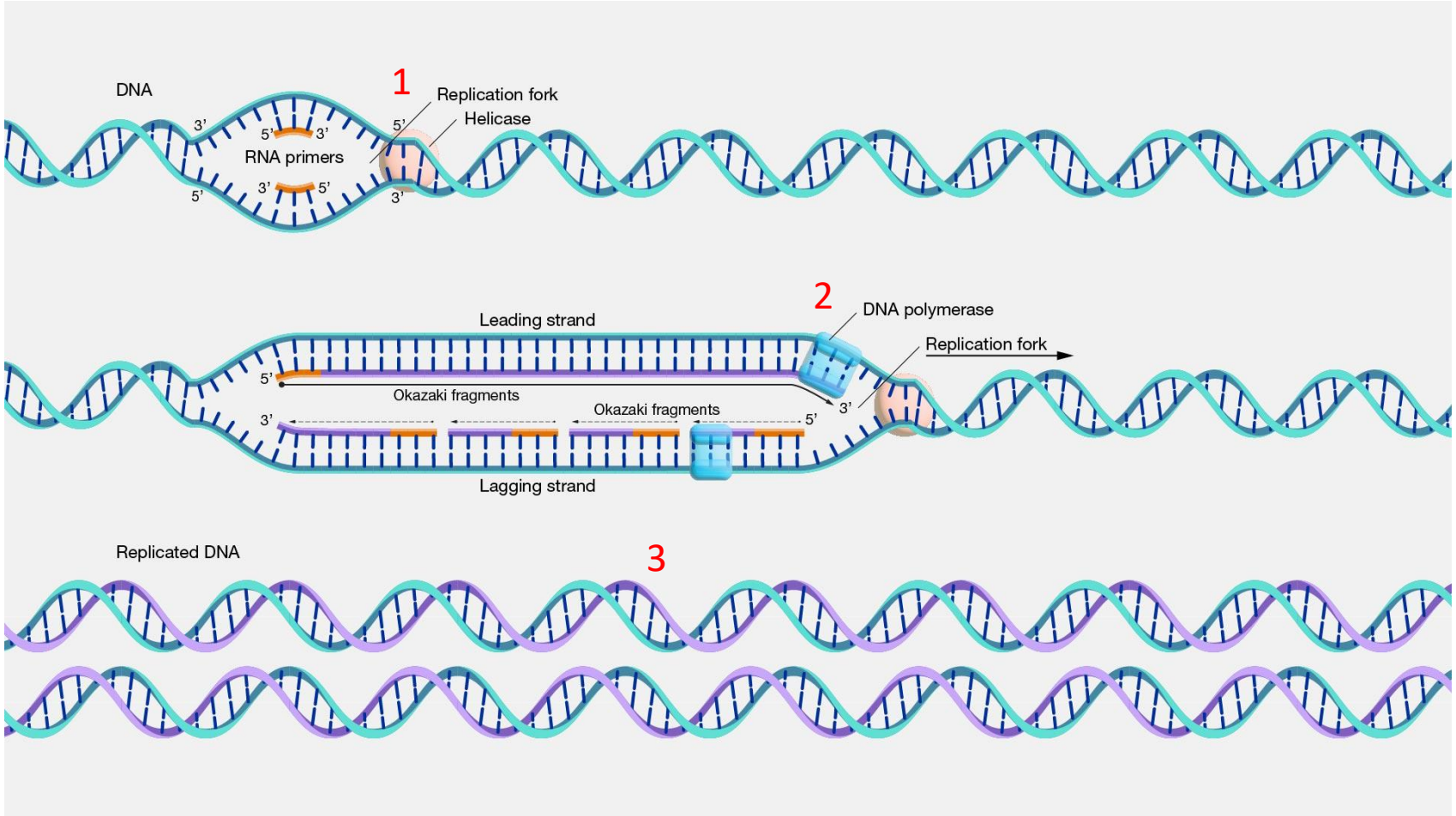
Barcode

DNA replication

1 initiation

2 elongation

3 termination

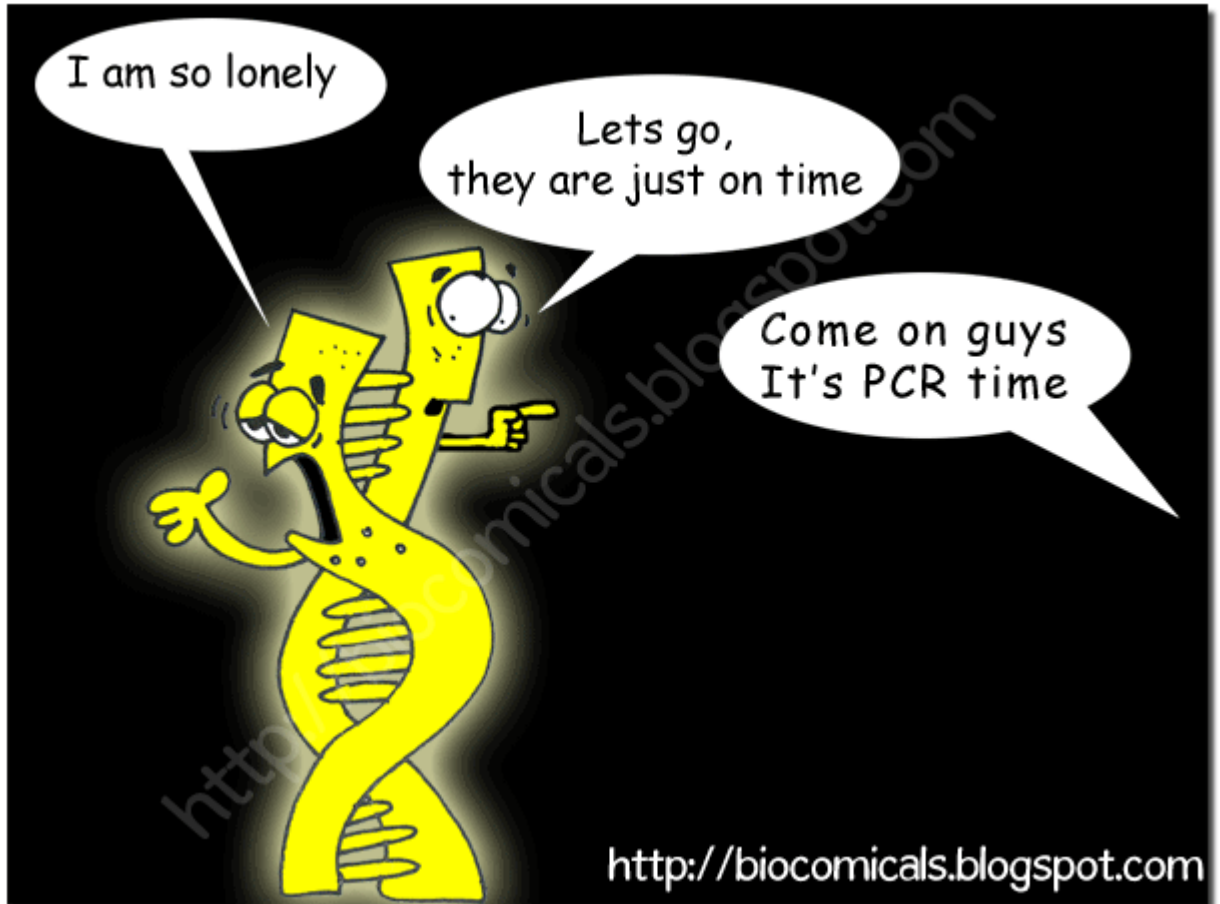


- DNA replication takes place in the cytoplasm in prokaryotes and in the nucleus in eukaryotes.
- a natural process, at body temperature

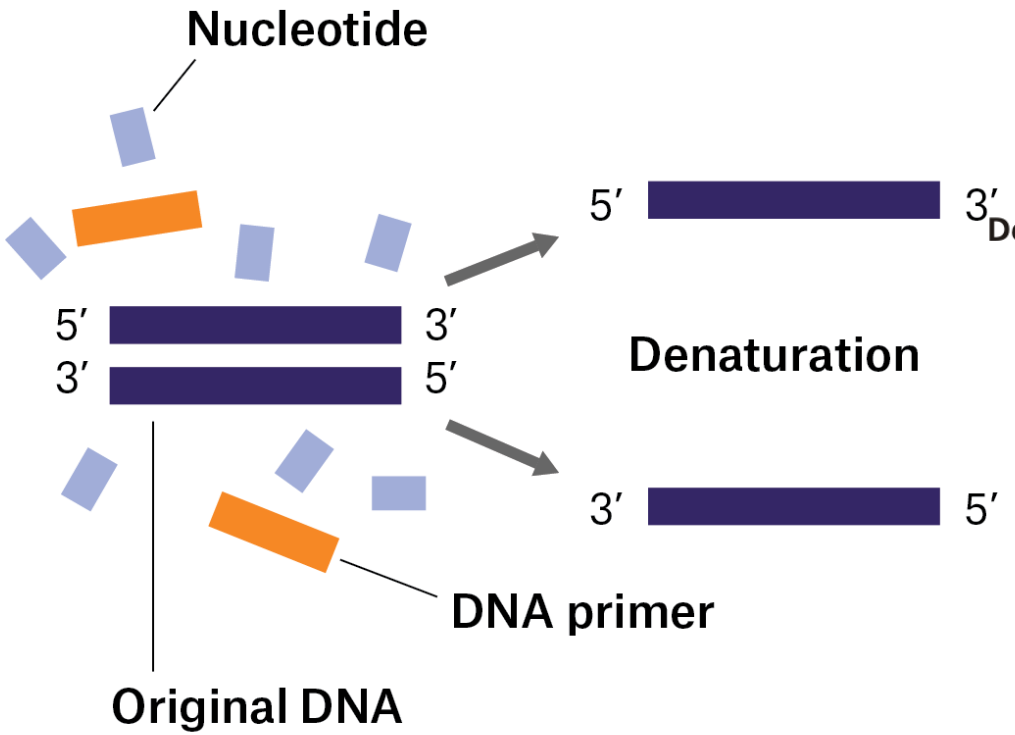
Barcoding

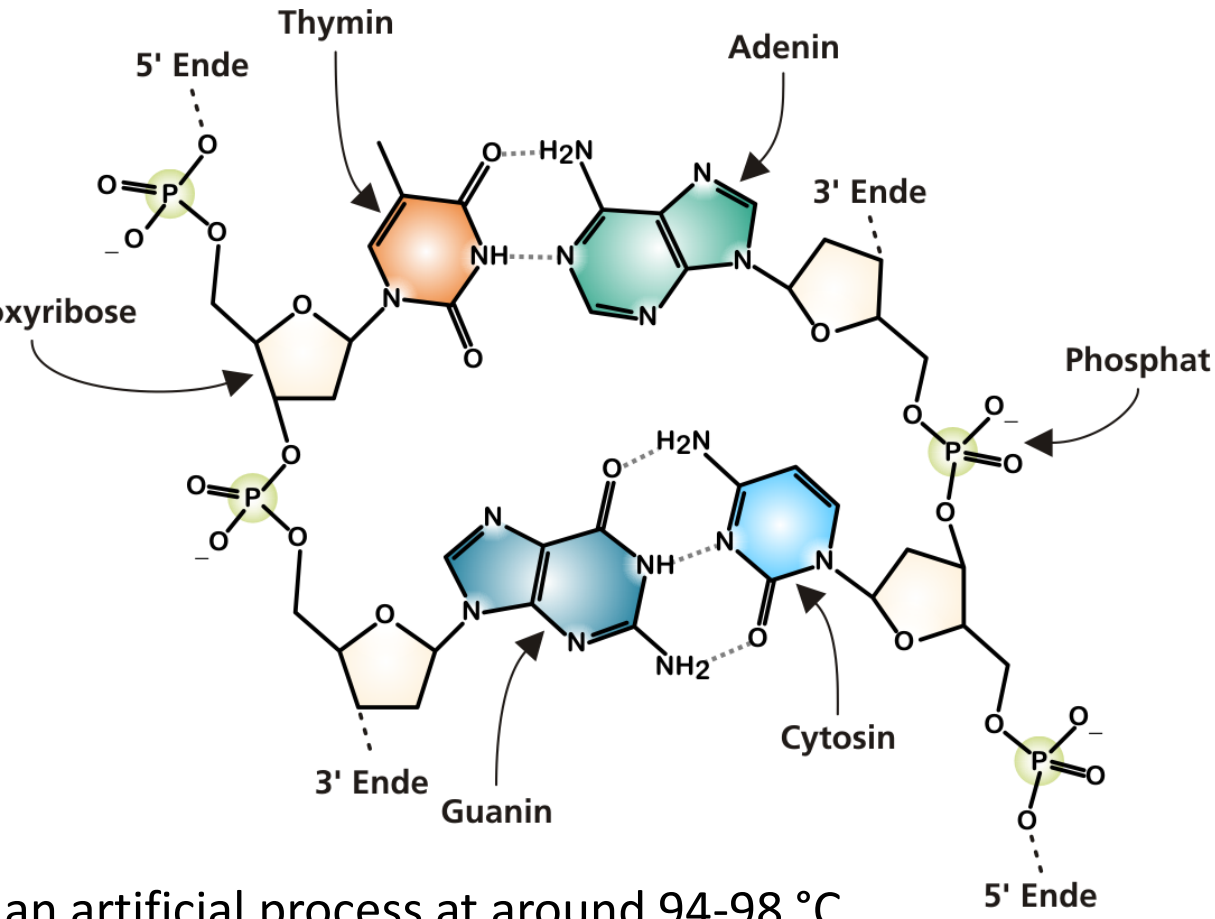
PCR = Polymerase chain reaction
A method to multiply DNA

Genomic DNA



Barcoding

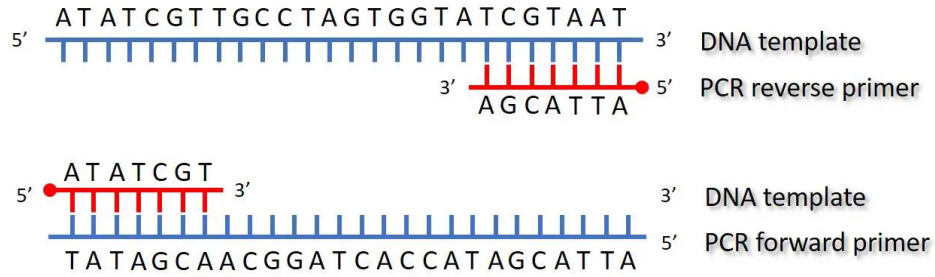


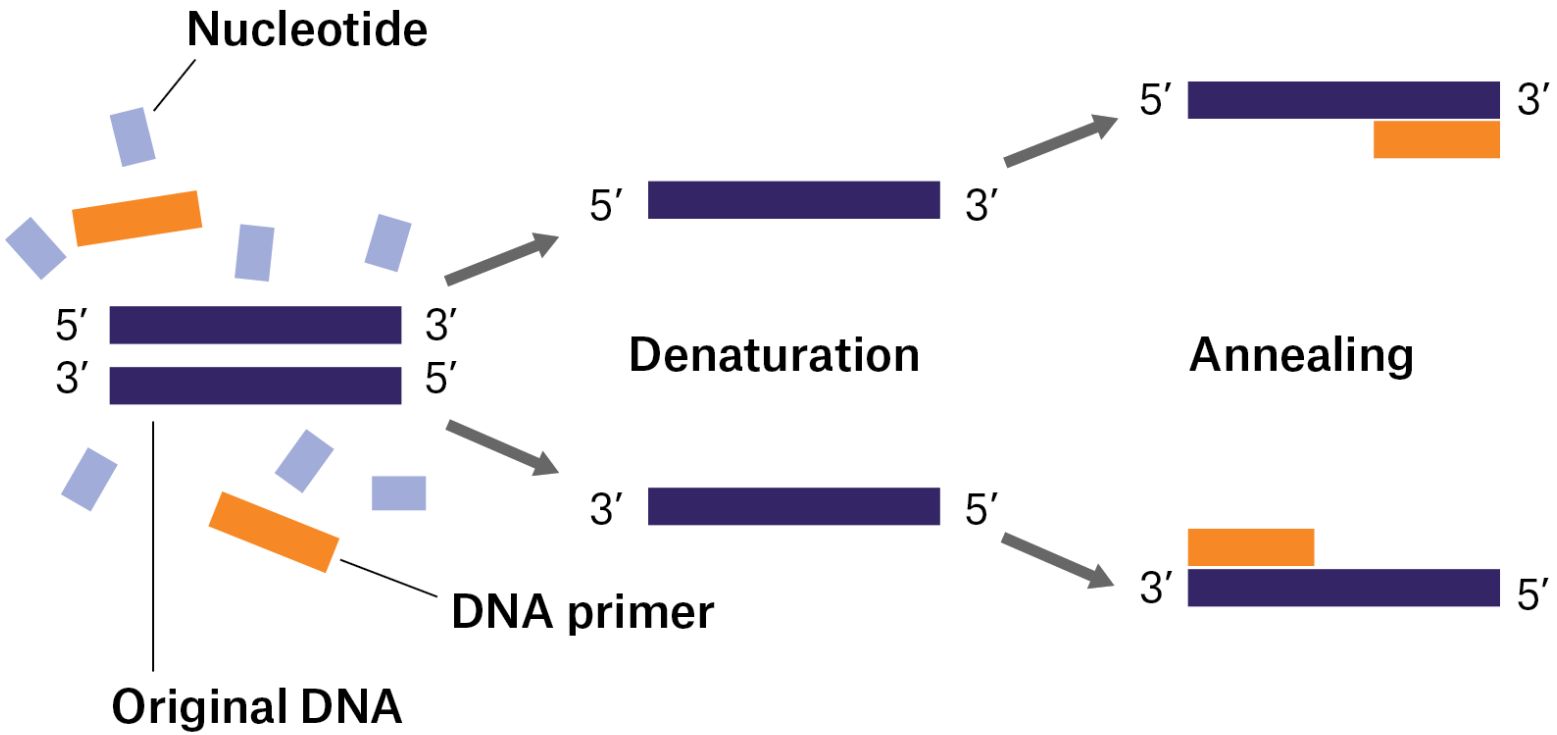


Denaturation: an artificial process at around 94-98 °C

BIOLAWEB

Barcoding

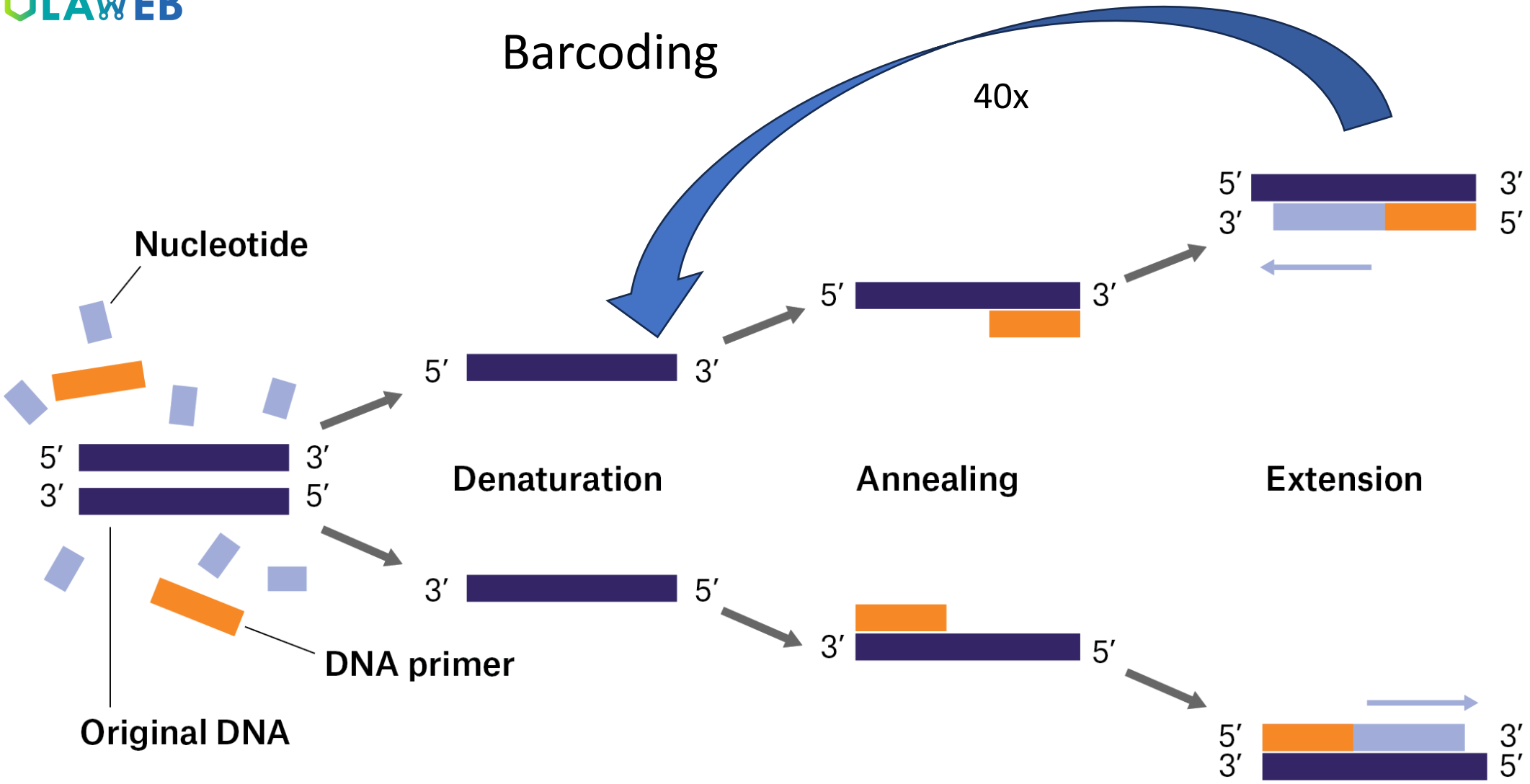




Primer annealing: an artificial process at around 45-62 °C

BIOLAWEB

Barcoding



Extension: an artificial process at around 70-72 °C

Needs: Enzyme DNA polymerase + nucleotides

Barcoding/Metabarcoding

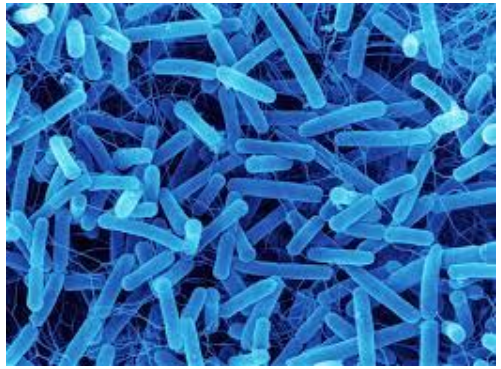
Must match the PCR technology

Up to 3.4 kb is possible
Often 200 -1000 b

PCR machine

BIOLAWEB

The perfect barcode does not exist to target diverse organism groups



Barcode selection

How do you select?

From other studies (publications)

- tested

- known bias

- check them against Ref. libraries

Group of organisms	Marker gene/locus used for barcoding
Animals	COI, *Cytb*, 12S, 16S
Plants	*matK*, *rbcL*, *psbA-trnH*, ITS
Bacteria	COI, *rpoB*,16S, *cpn60*, *tuf*, RIF, *gnd*
Fungi	ITS,TEF1α, RPB1 (LSU), RPB2 (LSU), 18S (SSU)
Protists	ITS, COI, *rbcL*, 18S, 28S

https://en.wikipedia.org/wiki/DNA_barcoding#cite_note-88.

May be they are not suitable to your study

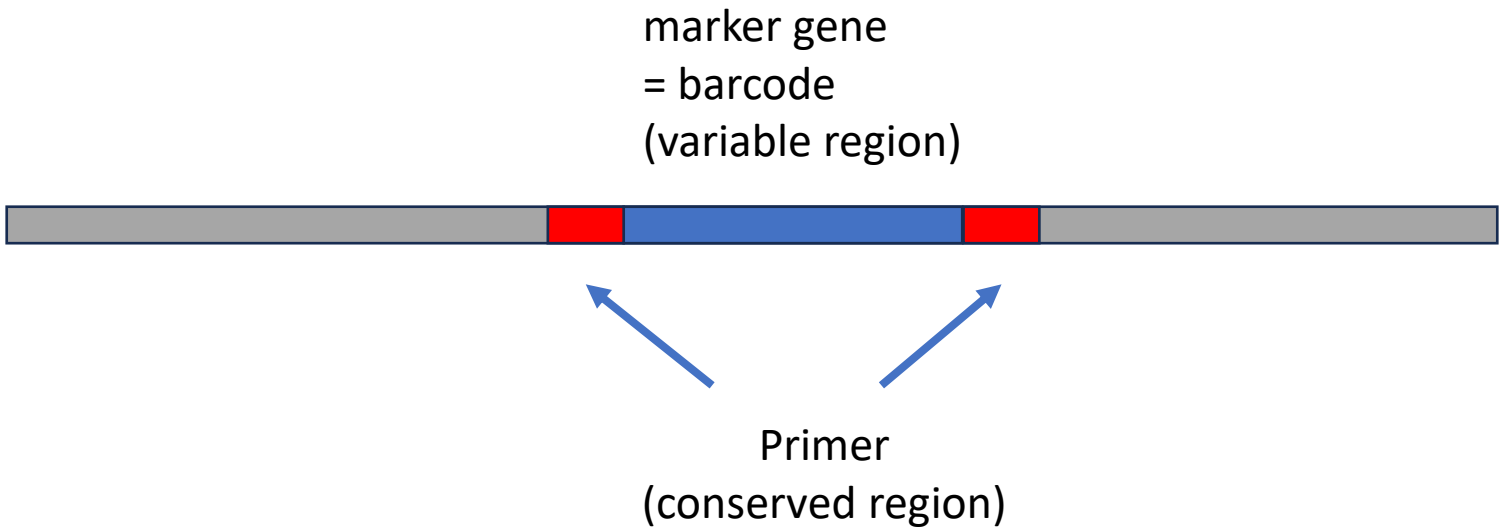
Design your own primer

- using reference sequences closest to your unknown organism

- needs optimalization and time

Barcode selection: Primer

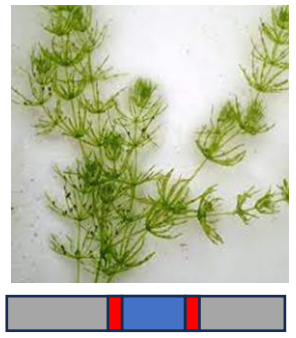
To target a group of organisms

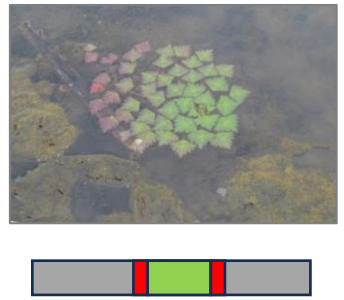


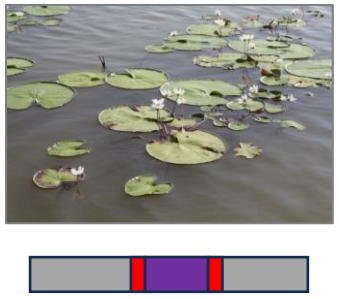
(i) significant interspecific genetic variability

(ii) conserved flanking sites for developing universal PCR primers

(iii) rel. short sequence length depending on sequencing technology

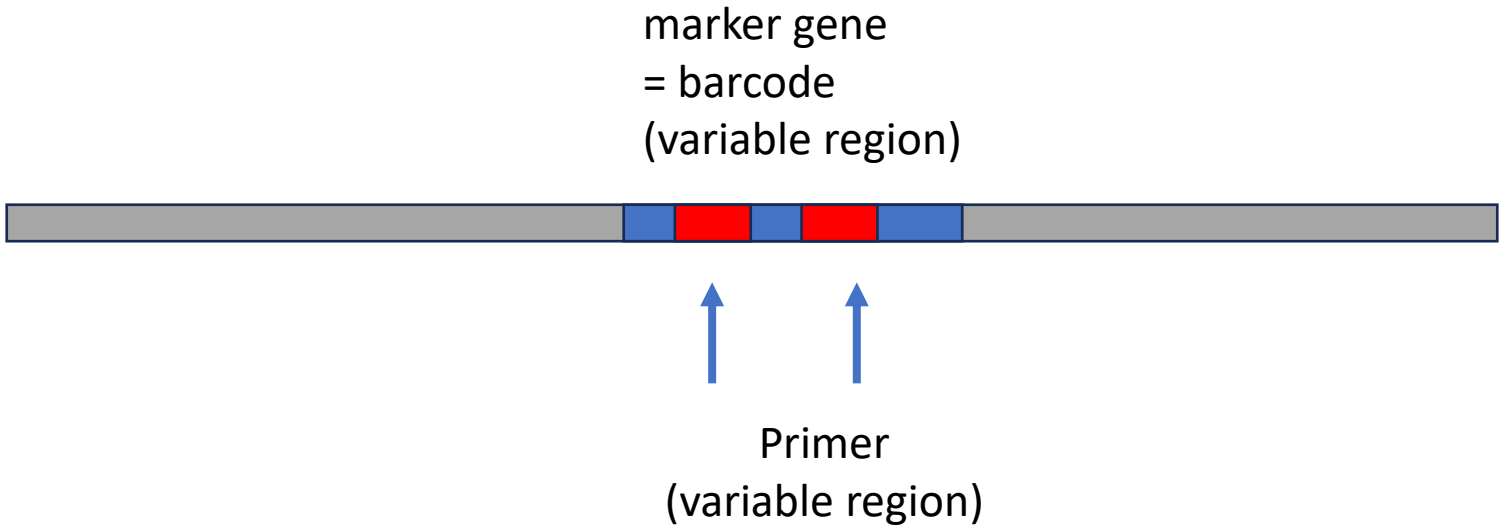






Barcode selection: Primer

To target a single species

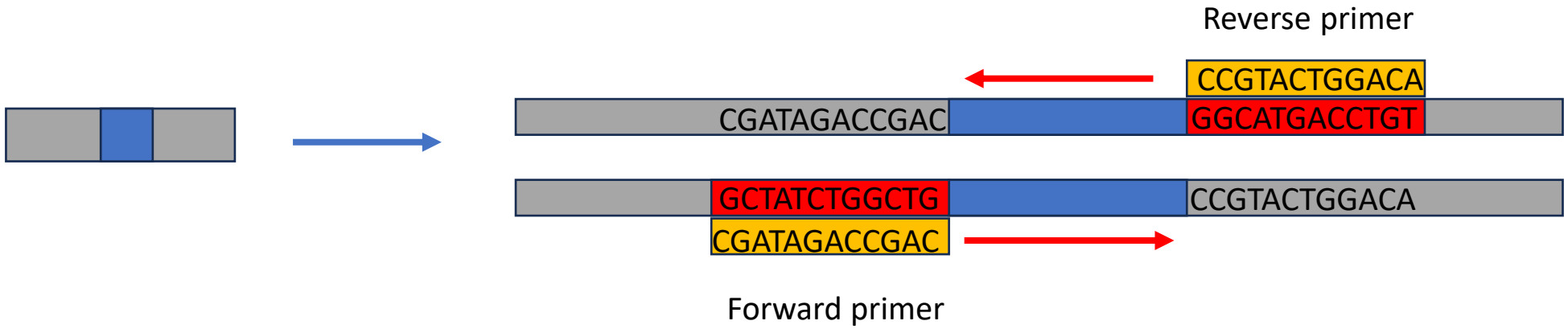


(i) significant intraspecific genetic variability

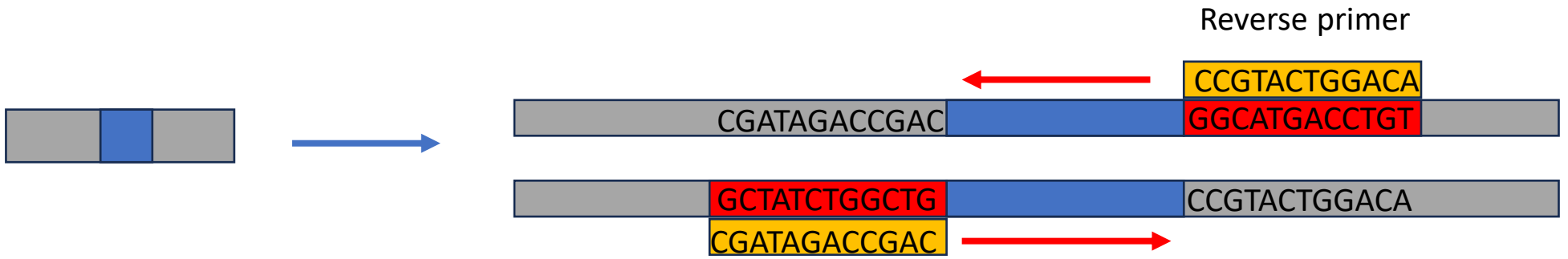
(ii) rel. short sequence length depending on sequencing technology or qPCR

Barcode selection: Primer

Primer



Barcode selection: Primer



Degenerate primers

CGATAGACCGAC

CGATACACCGAC

CGATASACCGAC

Code	Description
M	AC
R	AG
W	AT
S	CG
Y	CT
K	GT
V	ACG
H	ACT
D	AGT
B	CGT
N	ACGT

Barcode selection: Primer

How to make your primers?

BIOLAWEB

eurofins | Genomics

Quick Order

Markets | RUO Products | Company | Contact | EN

Integrated DNA Technologies acquires Archer™ next generation sequen

IDT INTEGRATED DNA TECHNOLOGIES

Integrated DNA Technologies

Search

Order by stock par

PRODUCTS & SERVICES | APPLICATIONS & SOLUTIONS | SUPPORT

PrimerQuest™ Tool

Design primers or assays for PCR, qPCR, or sequencing (any species).

- Customization of ~45 parameters, allowing qPCR assay designs:
  - With specific primer, probe, or amplicon criteria
  - Across a specified location
- Design algorithm includes multiple checks to reduce primer-dimer formation
- Provides flexible sequence entry and batch entries (up to 50 sequences)

Integrated DNA Technologies acquires Archer™ next generation sequencing research assay

IDT INTEGRATED DNA TECHNOLOGIES

Search

Order by stock part number »

PRODUCTS & SERVICES | APPLICATIONS & SOLUTIONS | SUPPORT & EDUCATION

OligoAnalyzer™ Tool

Understand the expected properties of your oligos ***before*** you order them.

- Calculator for GC content, melting temperature (T_m), molecular weight, extinction coefficient, µg/Ol nmol/OD, and more
- Identify secondary structure potential
- Minimize dimerization
- Use NCBI BLAST™

BIOLAWEB

Self-Dimer

```
4 bp, delta G = -6.6 kc/m (bad!)  (worst= -36.6)
        5' GGGAAAATTCCAGGATCTAT 3'
              ||||   ||||
3' TATCTAGGACCTTAAAAGGG 5'
```

Cross dimer

```
5'-CGGAAACAAGGAGGA
     ||||   ||||
3'-TATGAAGGACCTTACTTCCC-5'
```

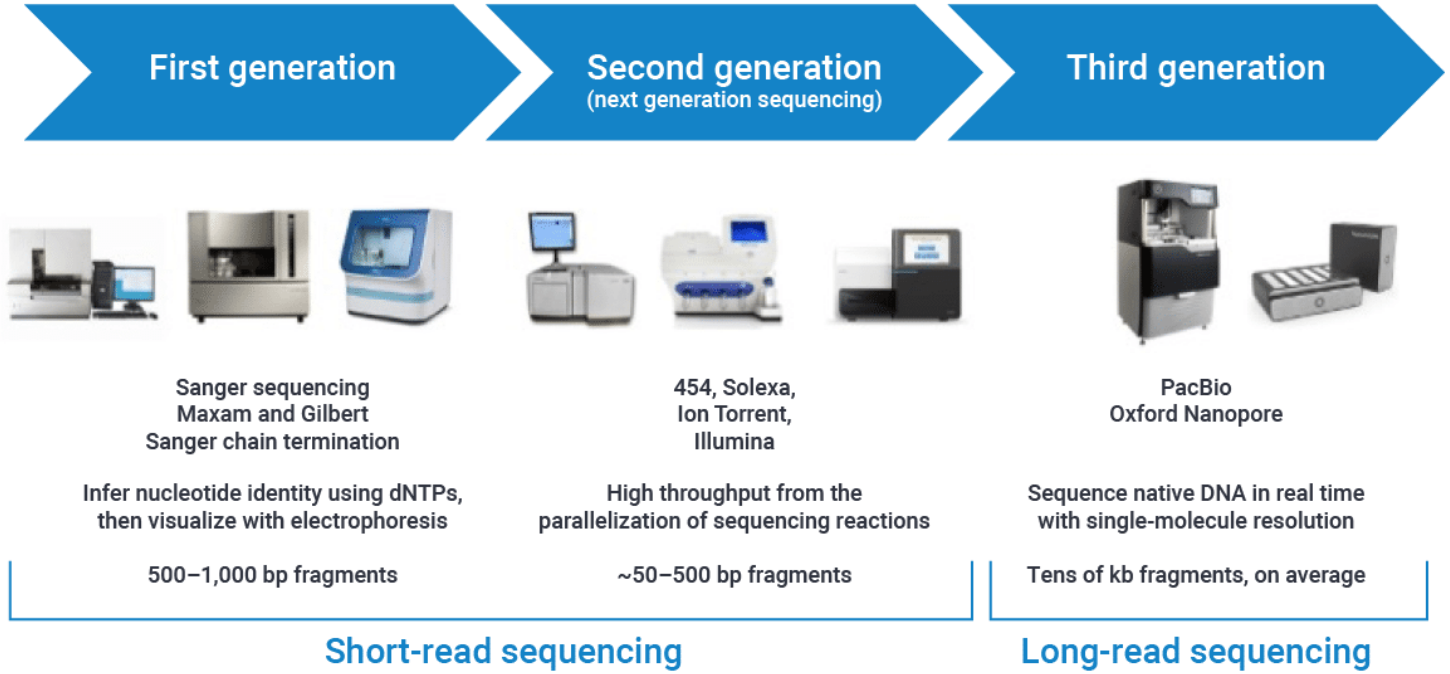
Hairpin

```
5'-GTCAGGATC
       |||
3'-CTATGTACGCCTTA
```

Barcode selection

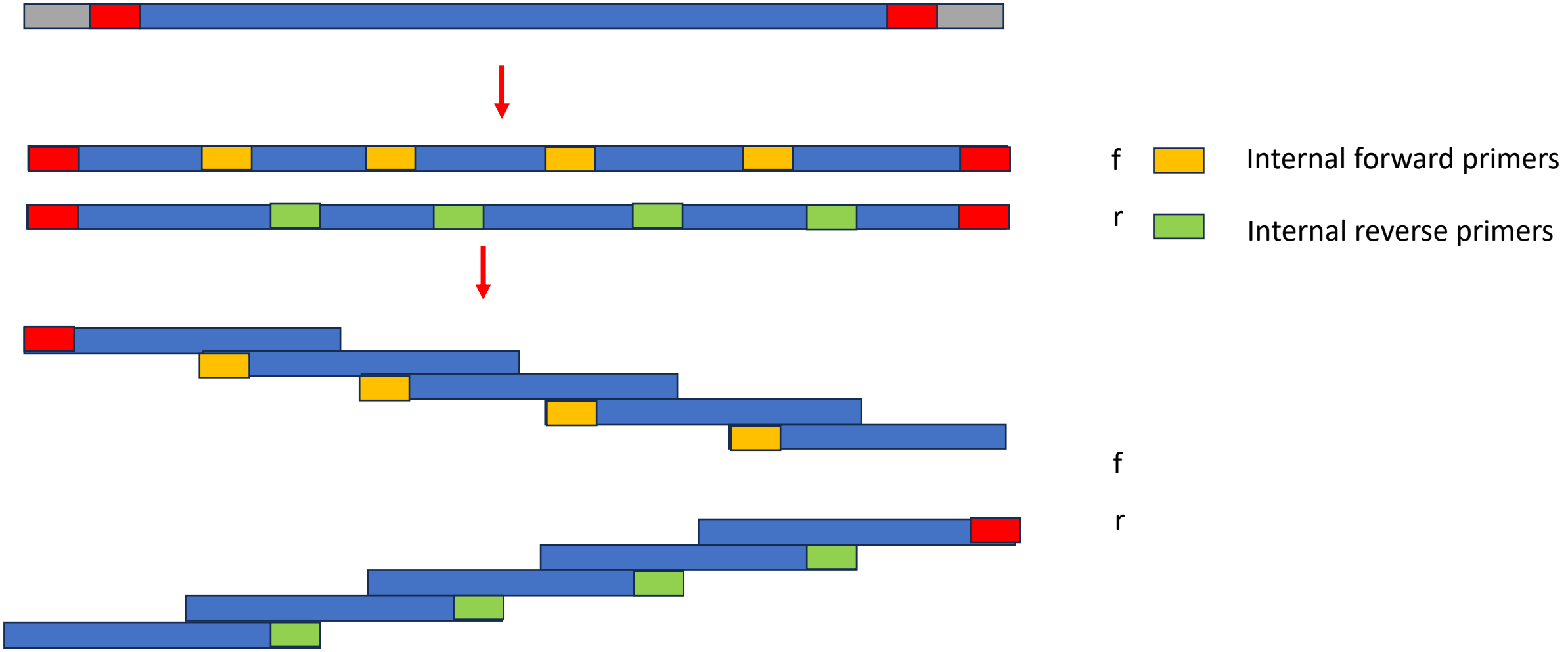
Must match the sequencing technology

16S	ca. 1500 bp
matK	ca. 1600 bp
rbcl	ca. 1500 bp



BIOLAWEB

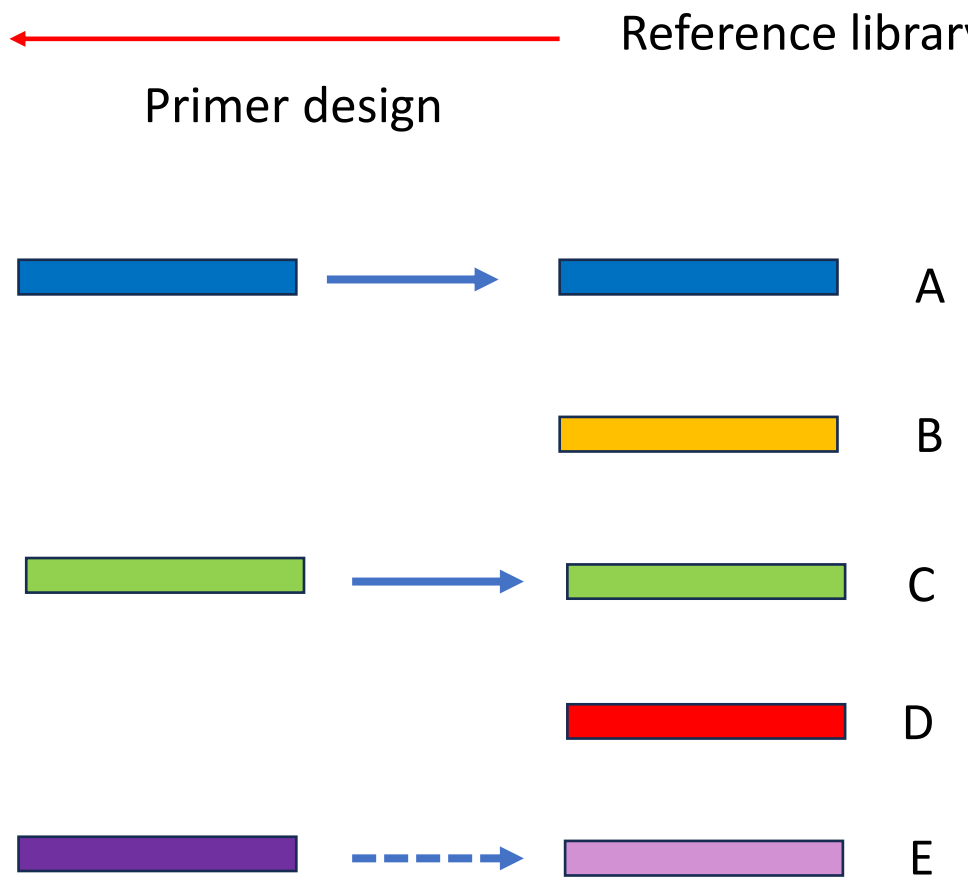
PCR > 1000 bp



Barcoding/Metabarcoding: Reference libraries

A prerequisite to assign your sequence(s) and identify your taxon

sample



Reliability increases with the number of confirmed sequences and taxa

Barcoding/Metabarcoding: Reference libraries

A

B

C

D

E

- as complete as possible

- as best curated as possible

BIOLAWEB

Barcoding/Metabarcoding: Reference libraries curated

BOLD SYSTEMS DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN

BOLD = Barcode of Life Database

PUBLIC DATA PORTAL - RECORD LIST

characeae | PUBLIC DATA | SEARCH

Specimens: DWC XML TSV
Sequences: FASTA TRACE
Combined: XML TSV
Map: Generate from

Records 1 to 100 **Page** 1 2 3 4 5 next>> **Records Per Page** 100

BBYUK2012-12 - Nitella sp. [rbcLa:552]
Taxonomy: Charophyta, Charophyceae, Charales, Characeae, Nitella
Identifiers: CCDB-18344-H3[sampleid], 05-074[fieldid], BABY-6276[museumid]
Depository: Research Collection of B. A. Bennett
Collected in: Canada, Yukon Territory

CYTC5629-12 - Chara vulgaris [COI-5P:1000,COII:748,COXIII:798,atp6:756]

Results Summary

Found **458** published records,
with **458** records with sequences,
forming **0** BINs (clusters),
with specimens from **19** countries,

Barcoding/Metabarcoding: Reference libraries not curated

National Center for Biotechnology Information
NCBI

ENA
European Nucleotide Archive

DDBJ
DNA Data Bank of Japan

250 million sequences
ca. 160 000 taxa

Barcoding/Metabarcoding:
Chara spp. (matK)

matK primer and rbcl primer from former studies to be tested for metabarcoding

F-matK-Chara AGAATGAGCTTAAACAAGGAT
R-matK-Chara ACGATTTGAACATCCACTATAATA

Chara-matK-BT2F DATATGGCAACAYCAAAAGAC
Chara-matk-BT2R ATACAGACCATGCAGCYTT

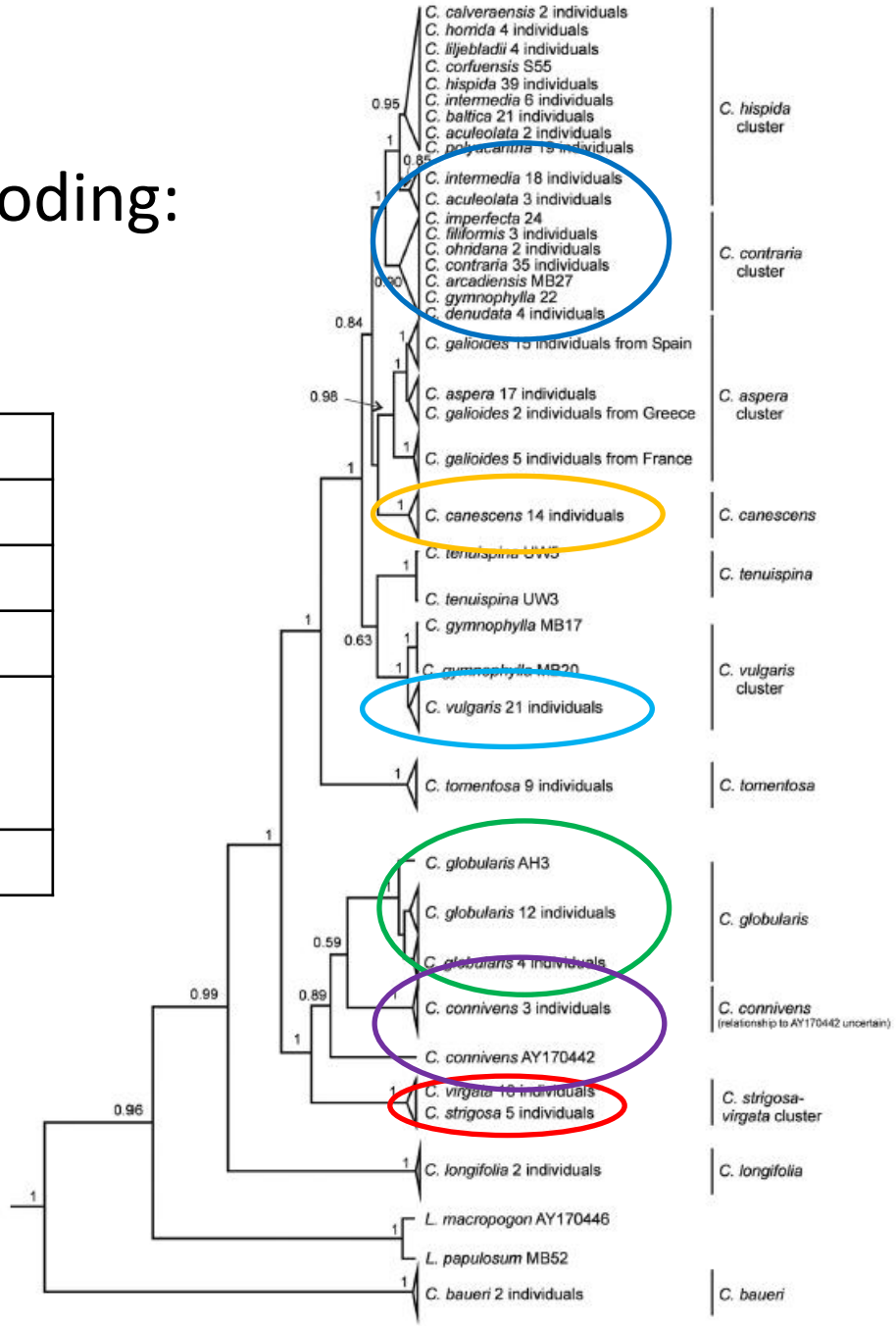
matKF2 AATGAGCTTAAACAAGGATTC
matKR1a CGTCCATGTAGATCTAATACTAG

Chara_matKF2 GAACGAATCCGTGATAAAAGC
Chara_matKR2 CTTCGGCCTTTCAAAAAGAA

rbcLa-F ATGTCACCACAAACAGAGACTAAAGC Levin et al. 2003
rbcLa-R GTAAAATCAAGTCCACCRCG

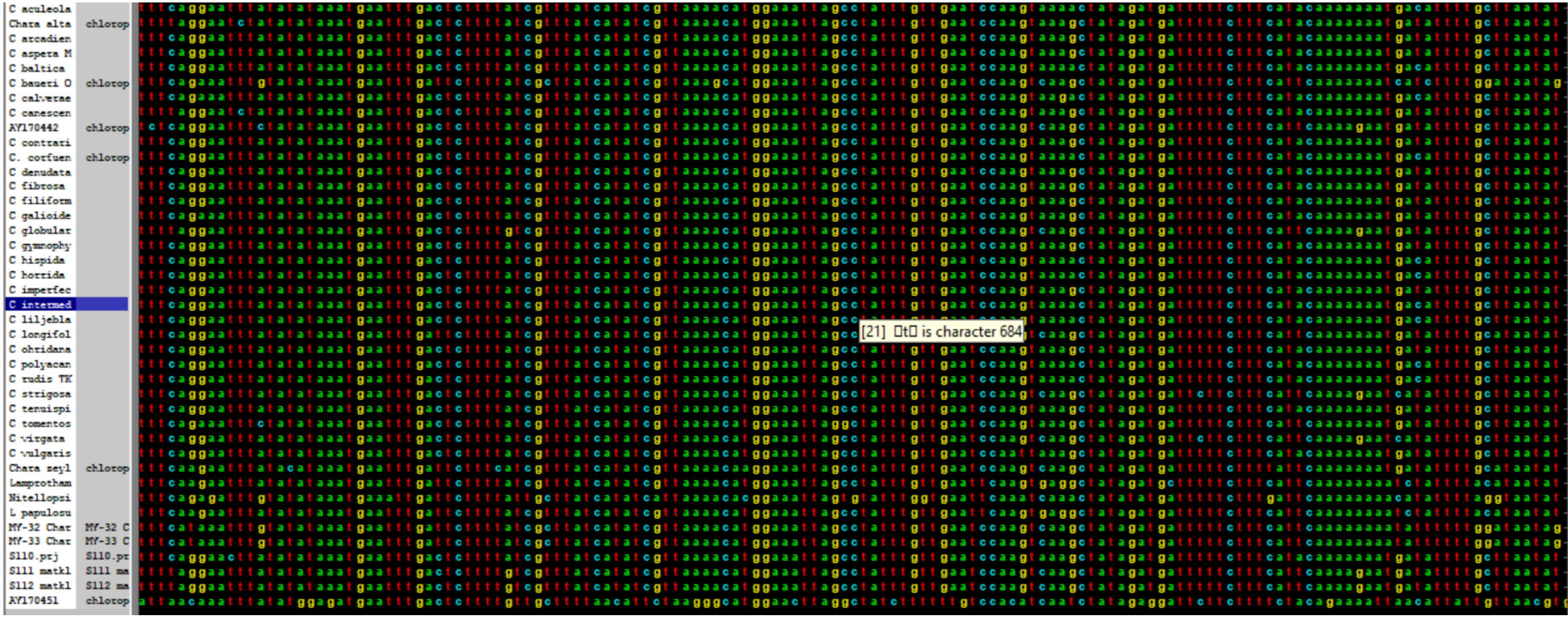
Barcoding/Metabarcoding: Chara spp. (matK)

Chara virgata	Sava lake
Chara contraria	Sava lake
Chara globularis	Sava lake
Chara connivens	Sava lake
Chara canescens	Pečena Slatina; Plava banja
Chara vulgaris	Markovačko Lake



Schneider et al. 2016:
345 Chara matK sequences

Barcoding/Metabarcoding:
Chara spp. (matK)

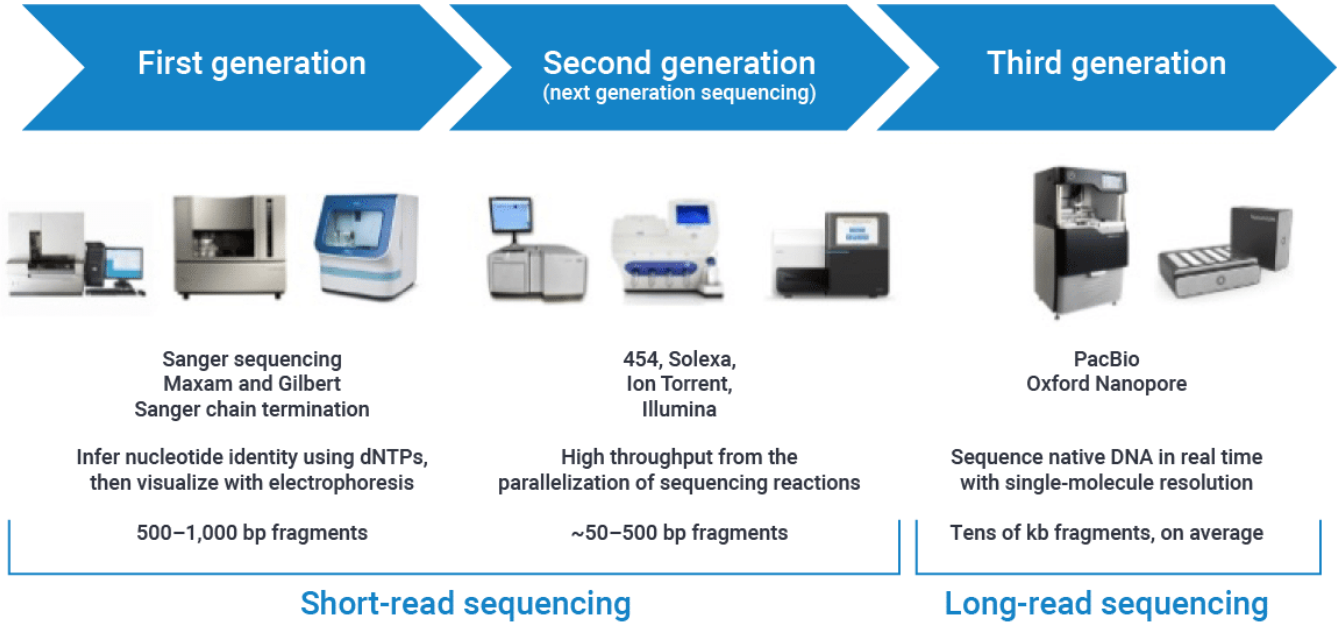


Barcoding/Metabarcoding: Chara spp.

- For several Characeaen taxa in Serbia metabarcoding is expected to work fine

- for some Characeaen taxa (Chara contraria) new markers have to be found

Chara virgata	Sava lake
Chara contraria	Sava lake
Chara globularis	Sava lake
Chara connivens	Sava lake
Chara canescens	Pečena Slatina; Plava banja
Chara vulgaris	Markovačko Lake



BIOLAWEB

Macrophyte (Characeae) barcoding/metabarcoding

First generation

Sanger sequencing
Maxam and Gilbert
Sanger chain termination

Infer nucleotide identity using dNTPs,
then visualize with electrophoresis

500–1,000 bp fragments

short DNA fragments

Reference library Characeae (matK, rbcl, ITS)

BOLD = Barcode of Life Database

Third generation

PacBio
Oxford Nanopore

Sequence native DNA in real time
with single-molecule resolution

Tens of kb fragments, on average

short to ultra long DNA or RNA fragments

Take home message

DNA related methods are cheaper and faster in the long run

However, a cooperation between taxonomists and molecular biologist is also a must in the future even with a good reference library

If not the risk of misinterpretations of data is high

Acknowledgement

This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No. 101079234

Funded by the European Union

Thank you for your attention!

www.biolaweb.com