



Metabarcoding : the main steps - part 2

Clarisse Lemonnier

The INRAE logo is positioned at the bottom left of the slide. It consists of the letters "INRAE" in a bold, teal, sans-serif font. The "E" is stylized with a circular element at its bottom right corner. The logo is partially overlaid by a large, abstract graphic of overlapping hexagons in various shades of green and teal that occupies the left side of the slide.

INRAE



Summary

Sampling

DNA extraction

PCR

Sequencing

Bioinformatic analysis

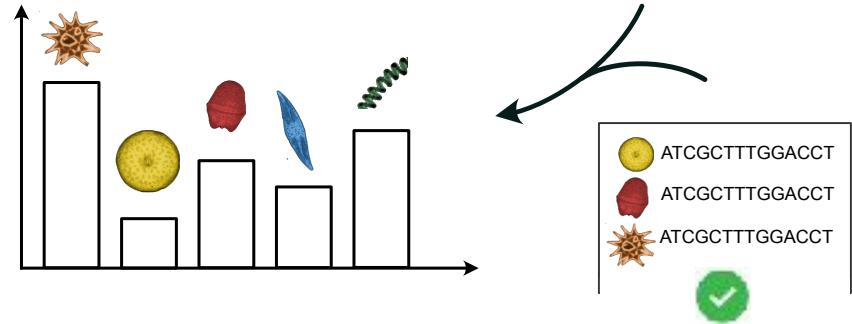
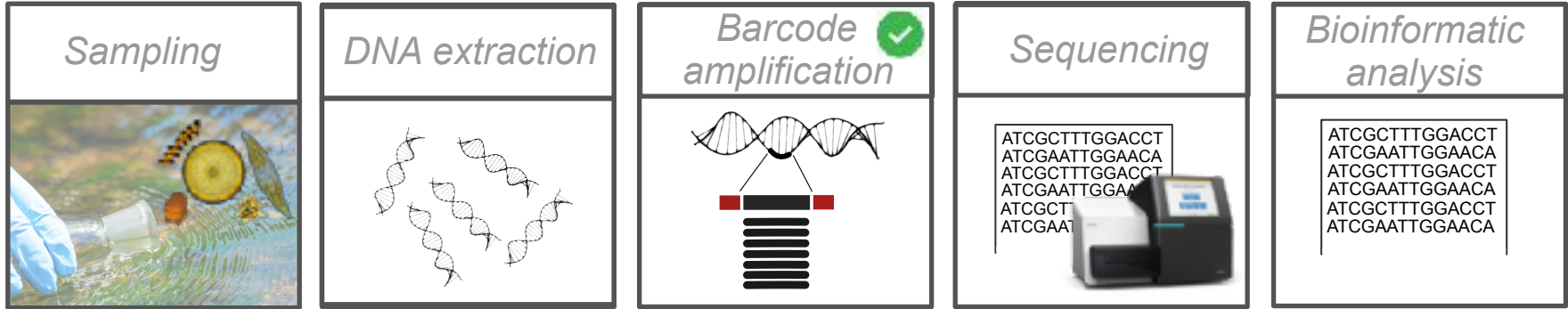
Taxonomic assignation

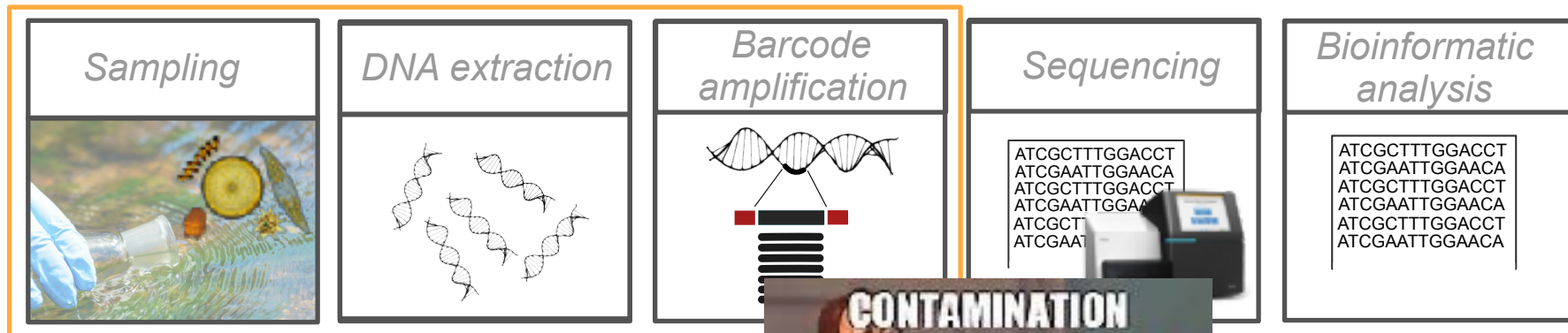


Funded by European Union

www.biolaweb.com

Metabarcoding steps






« Clean » conditions

Avoid contaminations



-  ATCGCTTTGGACCT
-  ATCGCTTTGGACCT
-  ATCGCTTTGGACCT

Metabarcoding steps

*Diatom biofilm :
thousands of cells/cm²*

*Ancient DNA:
A very few/m²*

Sampling



DN



de
ation






ormatic
lysis

TGGACCT
TGAACA
TGGACCT
TGAACA
TGGACCT
TGAACA

« Clean » conditions

Avoid contaminations



 ATCGCTTTGGACCT
 ATCGCTTTGGACCT
 ATCGCTTTGGACCT

« Clean » conditions

DNA extraction



Clean material



« Sterile » environment



Negative controls



Sampling

Clean dishes.

Use of hydrogen peroxide, HCl, bleach, autoclave, sterile devices ...

Add negative controls





Sampling

Clean dishes.

Use of hydrogen peroxide, HCl, bleach, autoclave, sterile devices ...

Add negative controls

Sample preservation : preservation of DNA

-80°C or -20°C that preserve DNA (ideal for long-term preservation)

Other solutions, such as buffers or ethanol



Metabarcoding steps

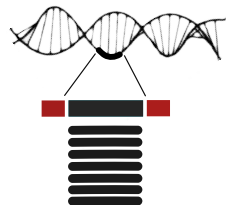
Sampling



DNA extraction



Barcode amplification



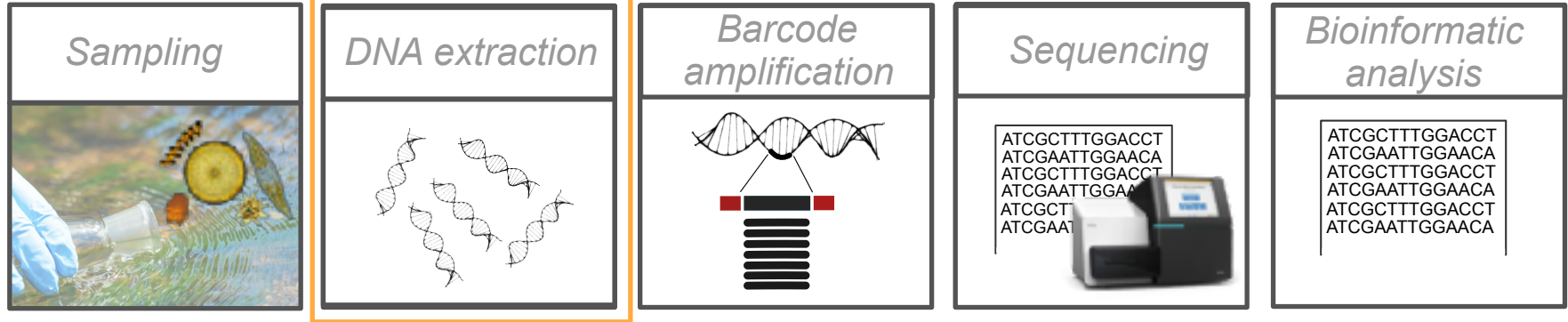
Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```



Bioinformatic analysis

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```



Objectives

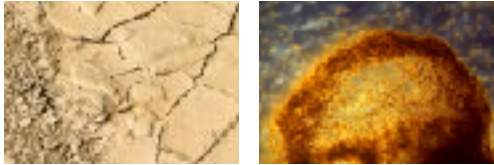
Extract DNA of good quality and quantity from the targeted organisms

In a given sample, the DNA molecule is associated with different molecules :



Of the matrix

Mineral and organic molecules



Within the cell

Intracellular organelles

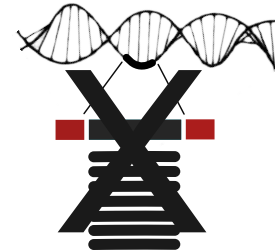


Within DNA molecule

Histones



In a given sample, there can be a lot of inhibitors for PCR



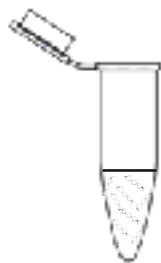
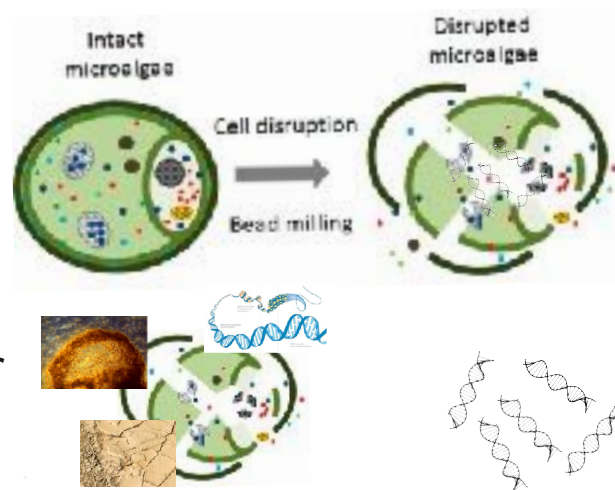
Type of inhibitor	Molecule or ion	Source	Mechanism/s ¹	Reference
Polymerase inhibitors	Al^{3+}	Sampling using aluminum-ant-shafted swabs	Alters ion composition	[29]
	Alginate	Sampling with calcium alginate swabs	Adsorption of Mg^{2+} or entrapment of polymerase	[29]
	Bile salts (choleic and deoxycholic acid)	Feces	Direct effect on polymerase	[15, 51]
	Calcium ions	Milk	Competition with the polymerase cofactor Mg^{2+}	[24]
	Collagen	Bone	Alteration of ion composition by binding cations	[62]
	EDTA	Anti-coagulant	Chelation of Mg^{2+}	[63]
	$FeCl_3$		Release of iron ions	[51]
	Free radicals	UV treatment of PCR tubes	Reaction with polymerase	[64-66]
	Falvic acid	Soil	Binding to polymerase	[51]

Main steps of DNA extraction

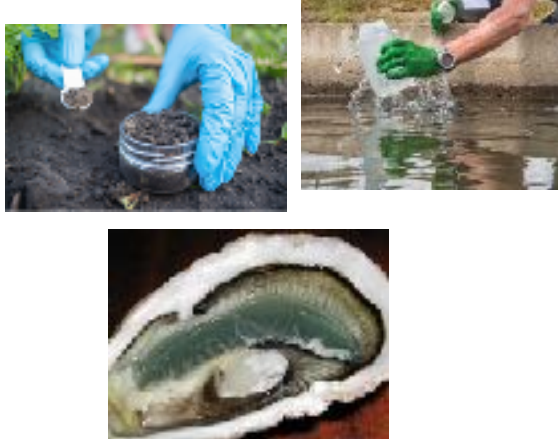
1/ Lysis of cellular membrane

2/ Separation of DNA from cellular debris, proteins and inhibitors

3/ Elution of DNA



Needs to be adapted
to the matrix



DNA extraction

Needs to be adapted
to the organisms



DNA extraction



Needs to be adapted
to the question

DNA extraction

Main steps of DNA extraction

1/ Lysis of cellular membrane

2/ Separation of DNA from cellular debris, proteins and inhibitors

3/ Elution of DNA

Phenol-chloroform protocol

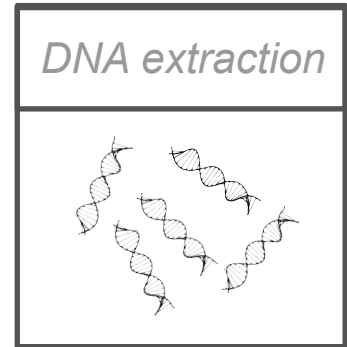


Commercial kits



DNA extraction





1/ Lysis of cellular membrane

Physical

Break the cells physically

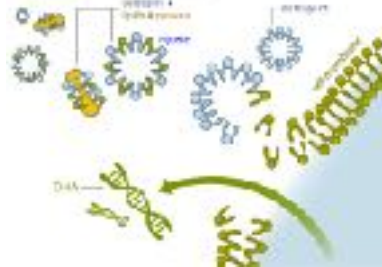
- Bead-beating
- freeze/thaw cycles
- Ultrasonication
- ...



Chemicals

chemicals that will destabilize the lipid membrane :

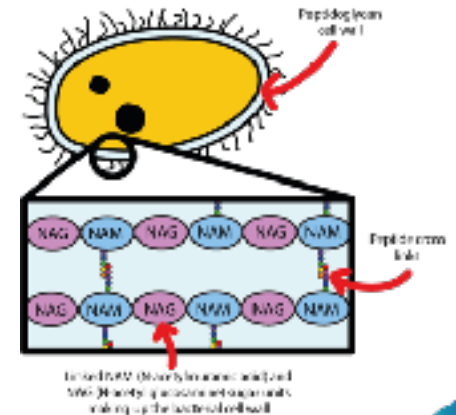
- Detergents
- Solvents
- Acids
- ...



Enzymatic

Use of enzymes that will disrupt the cell membrane:

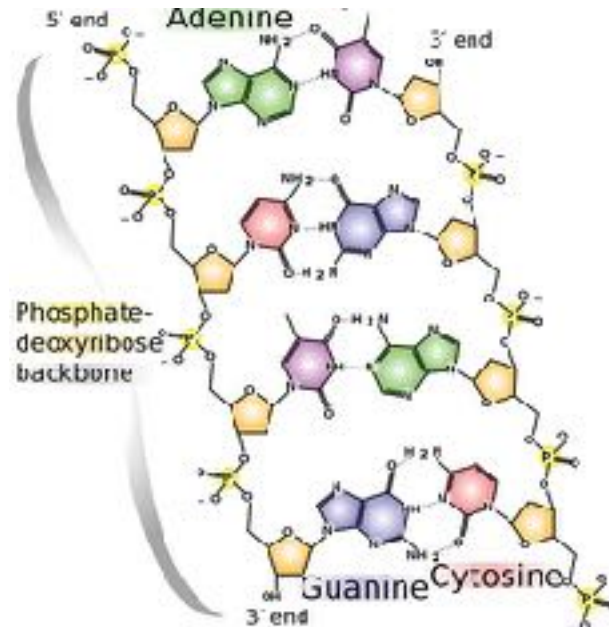
- Cellulase
- Lysozyme



2/ Separation of DNA from membranes debris / proteins

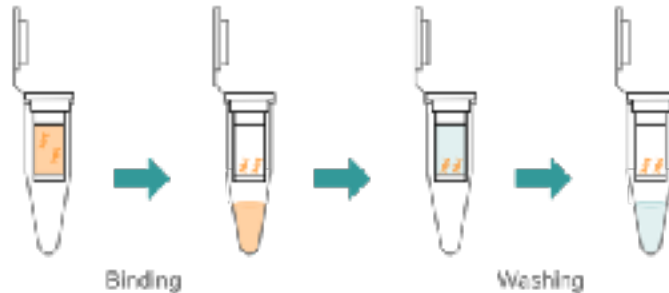
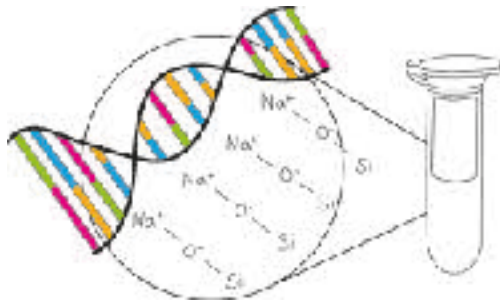
Use the natural properties of DNA :

- hydrosoluble
- Negatively charged



2/ Separation of DNA from membranes debris / proteins

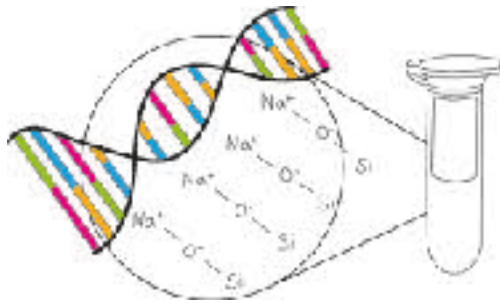
Use of silice column (kits)



Low pH and high salts
(chaotropic) : DNA binds to the
silicate membrane

2/ Separation of DNA from membranes debris / proteins

Use of silice column (kits)



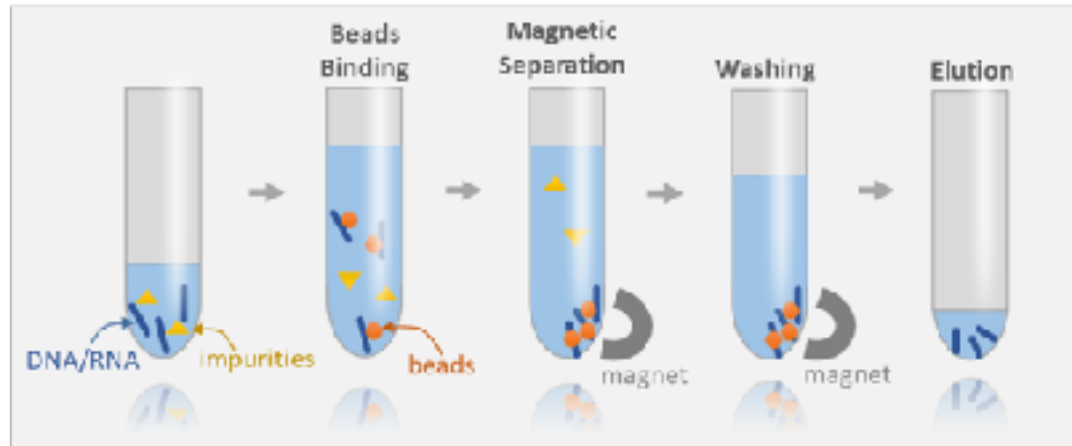
Low pH and high salts
(chaotropic) : DNA binds to the
silicate membrane



Higher pH - low salt : silicate
membrane become negative
and DNA is eluted

2/ Separation of DNA from membranes debris / proteins

Use of magnetic beads (kits)



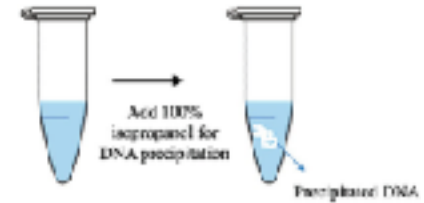
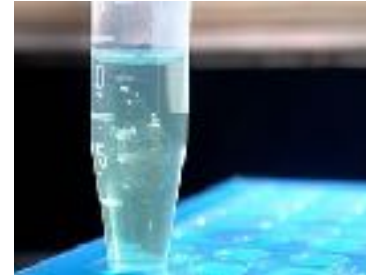
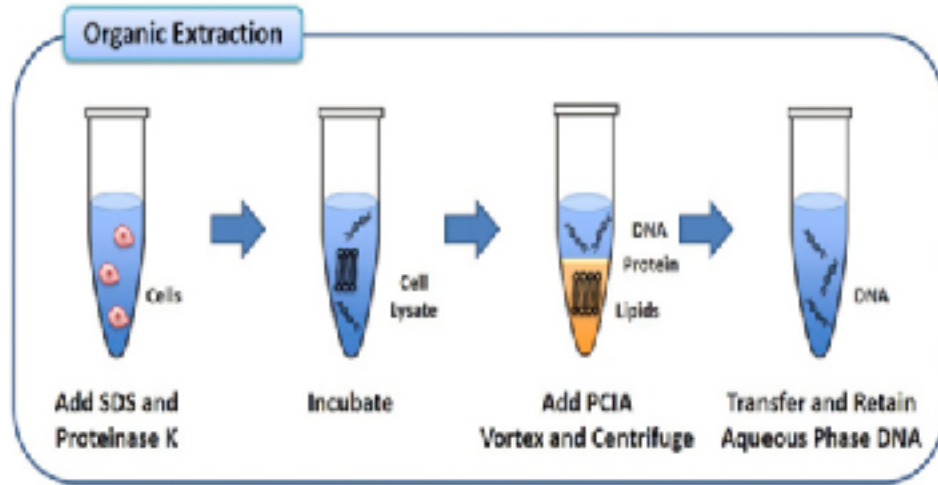
Magnetic beads are small (20-30nm) iron oxide particles

They are coated with silica to bind dna

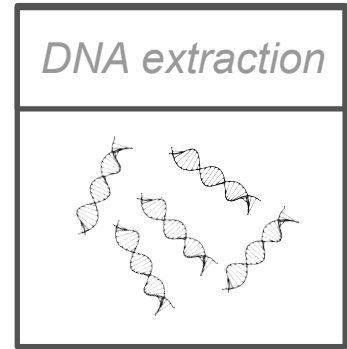
Works also with salts and solution with different pH

2/ Separation of DNA from membranes debris / proteins

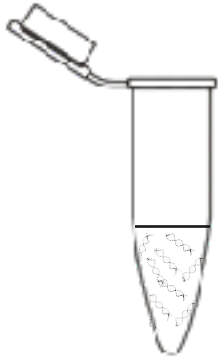
Use of organic solvents (phenol-chloroform extraction)



DNA is precipitated with salts and isopropanol



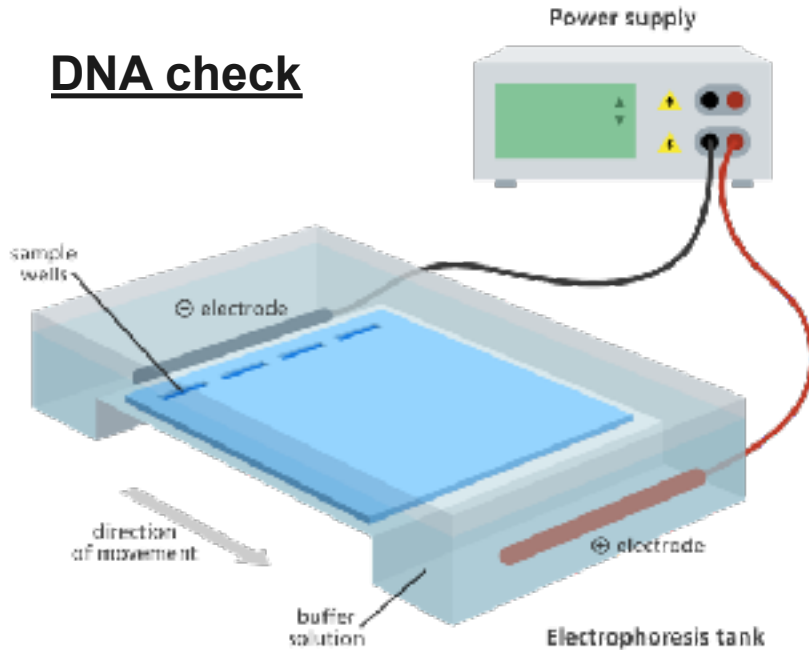
DNA extraction



DNA extraction



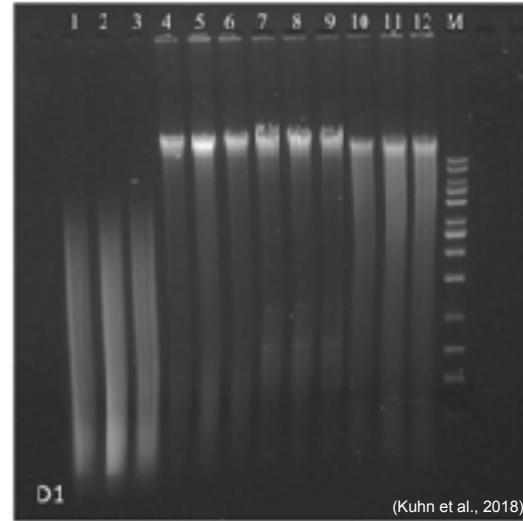
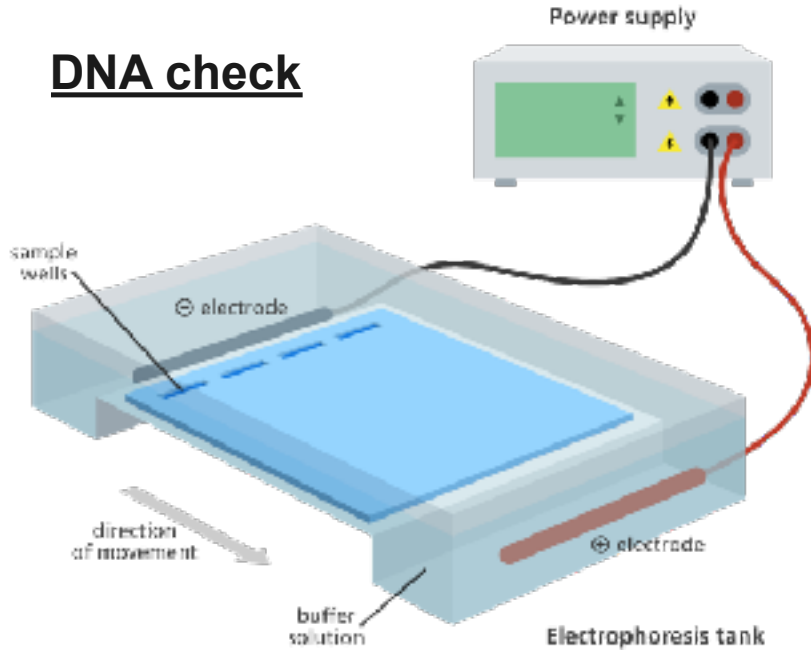
DNA check



DNA extraction

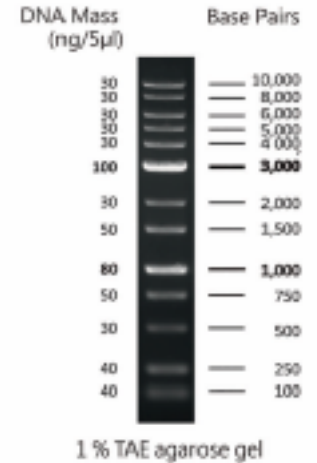
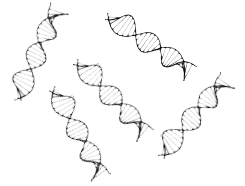


DNA check



DNA intercalant that fluoresce under UV light

DNA extraction





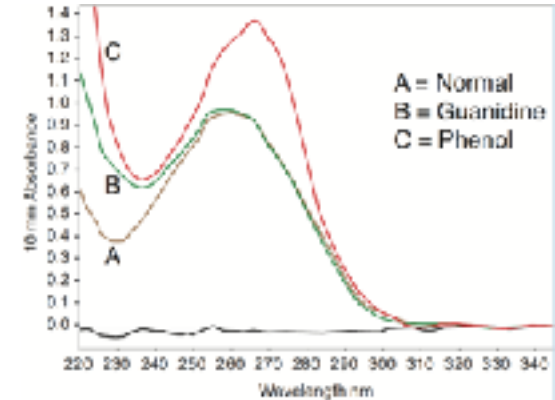
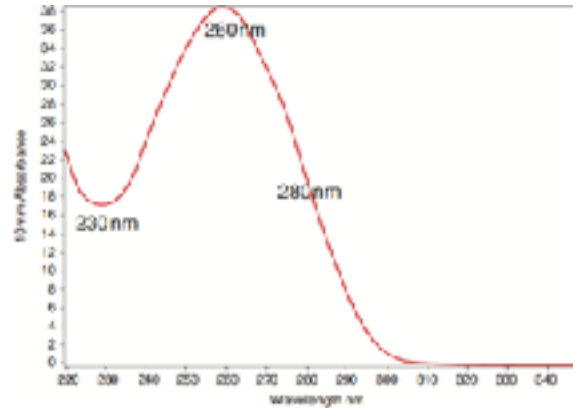
DNA quantity and quality check

Spectrophotometer



Nanodrop

Absorbance at 260nm :DNA quantity



Protein and organic contaminations:
260/230 or 260/280

Metabarcoding steps

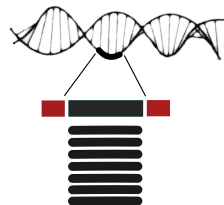
Sampling



DNA extraction



*Barcode
amplification*



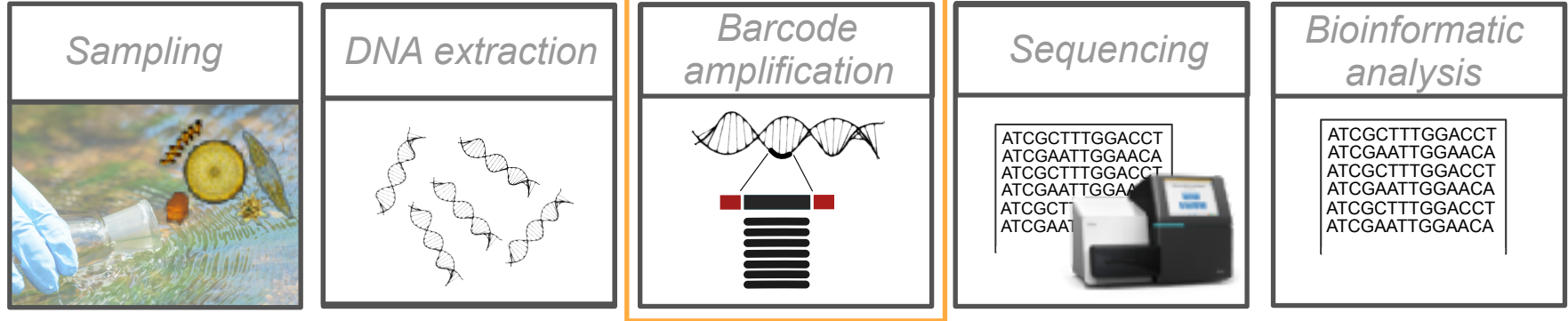
Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```



*Bioinformatic
analysis*

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```

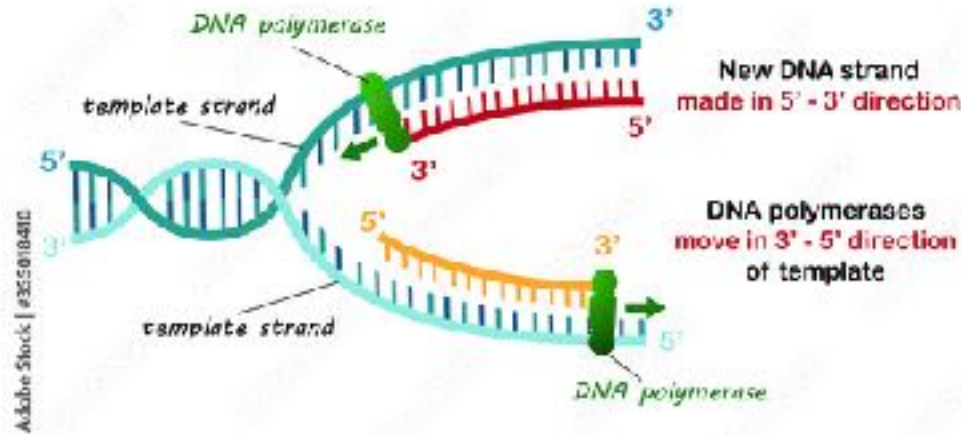
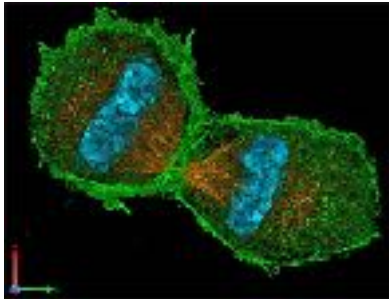


Objectives

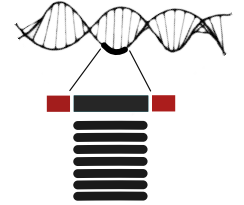
Amplify the barcode in thousands of identical copies

PCR steps

In the cell :

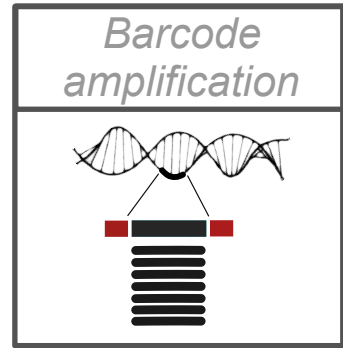


Barcode amplification



The DNA polymerase cannot synthesize a new DNA strand de novo. She can only add new nucleotides to a previous existing strand

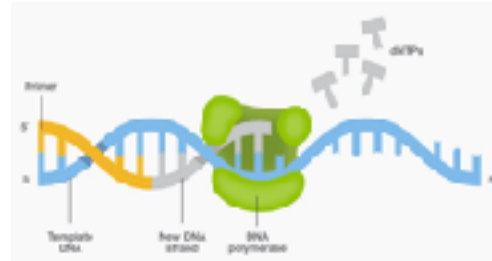
PCR steps



In the lab :

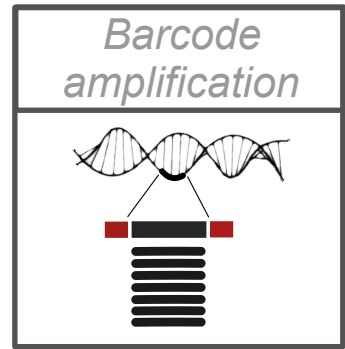
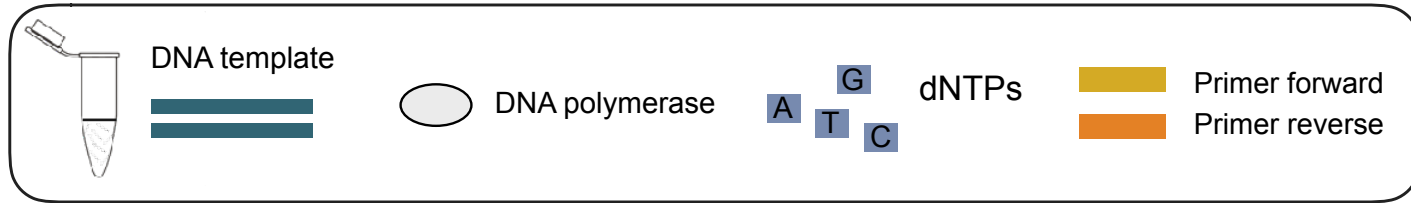
Artificially separate
the 2 DNA strands

Use primers for the DNA
polymerase to start the synthesis
of a new strand

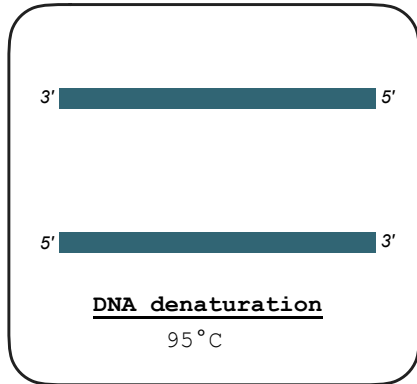
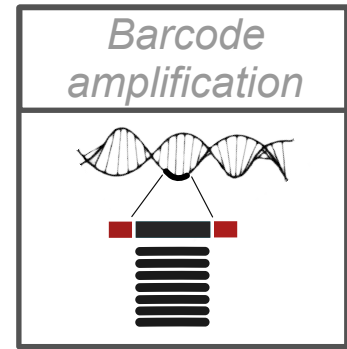
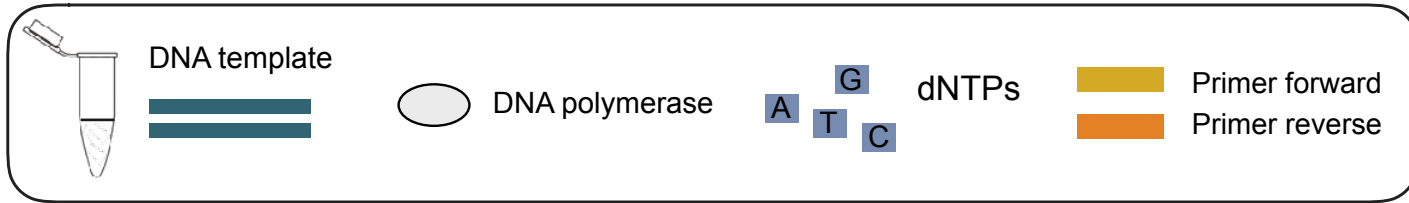


Use of a high-fidelity
DNA polymerase

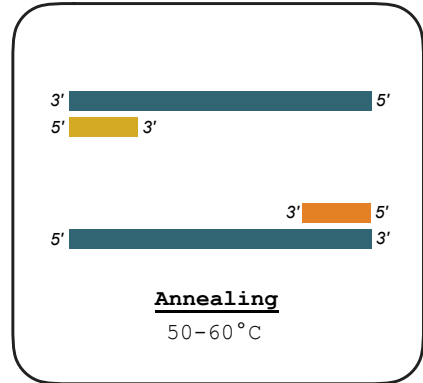
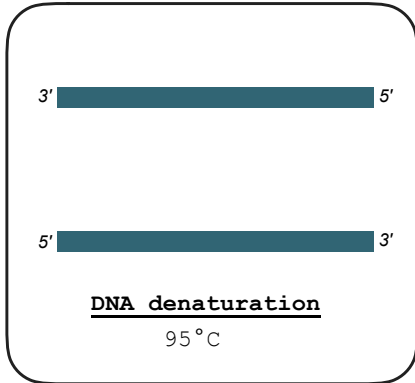
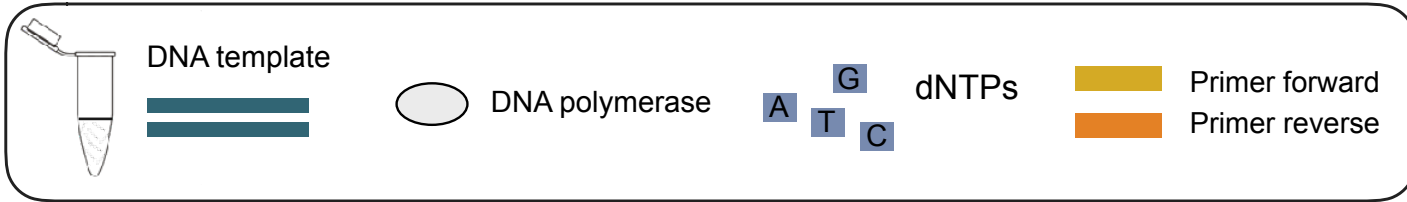
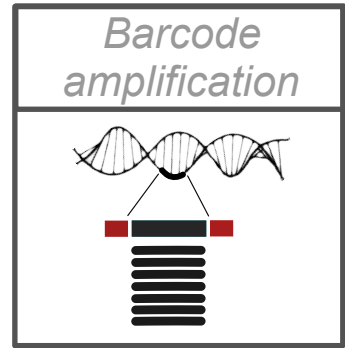
Different steps of PCR



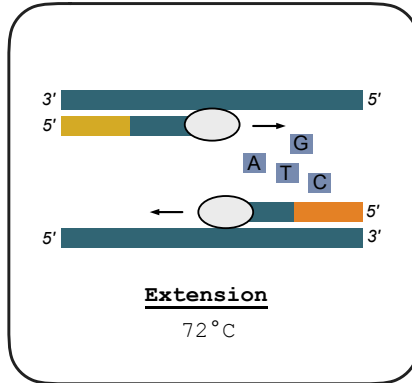
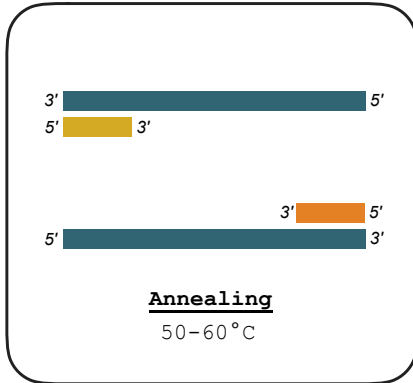
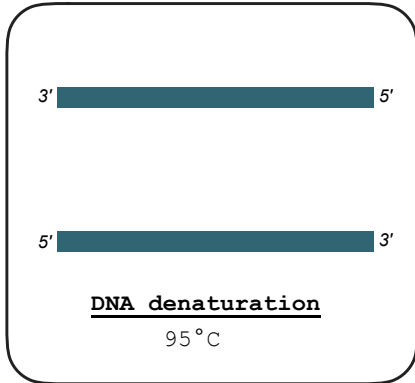
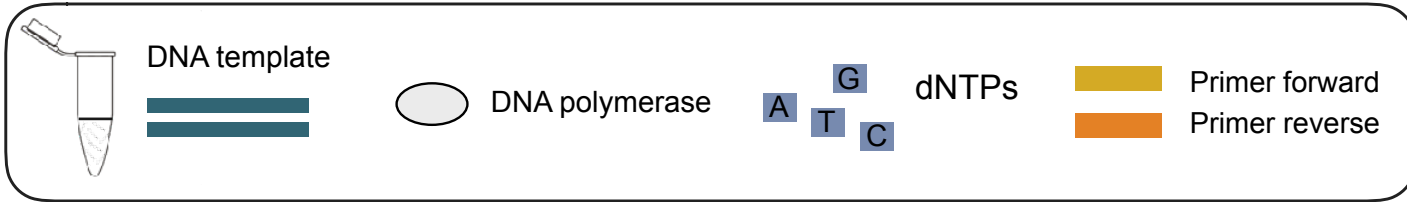
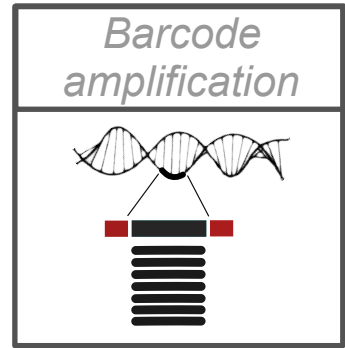
Different steps of PCR



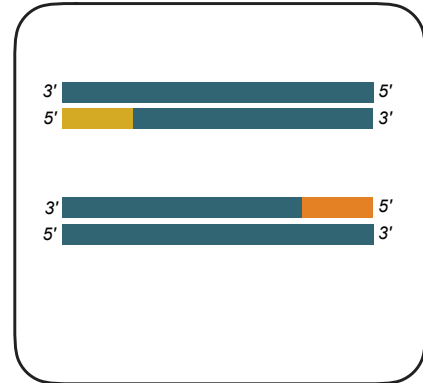
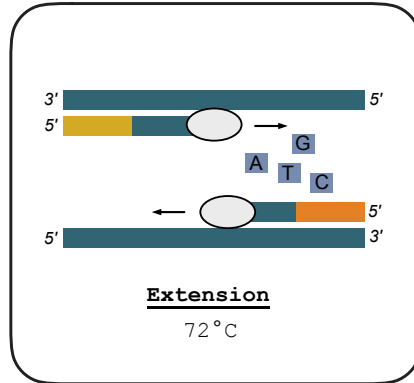
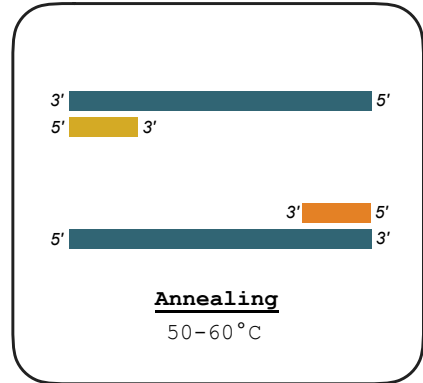
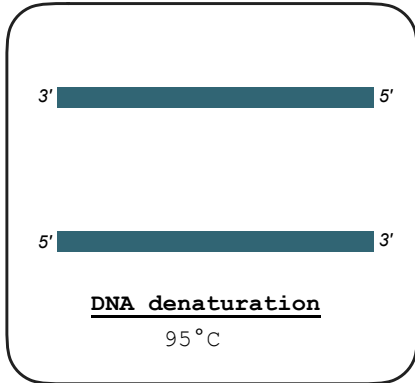
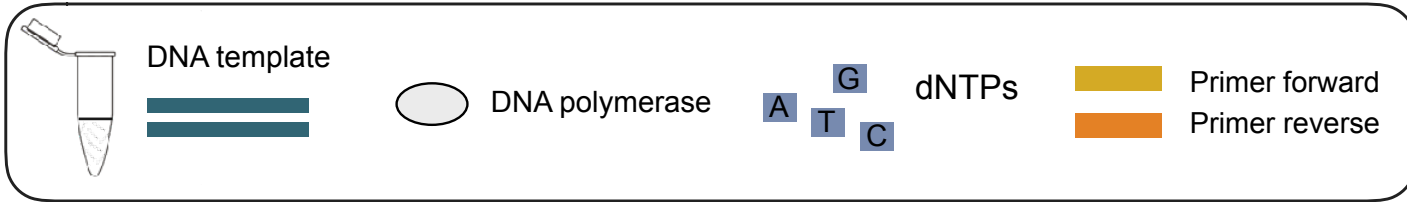
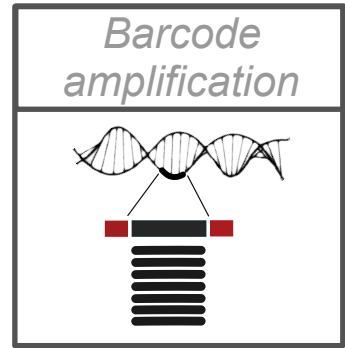
Different steps of PCR



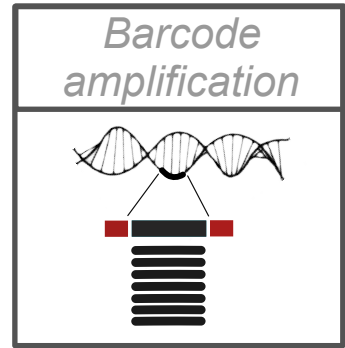
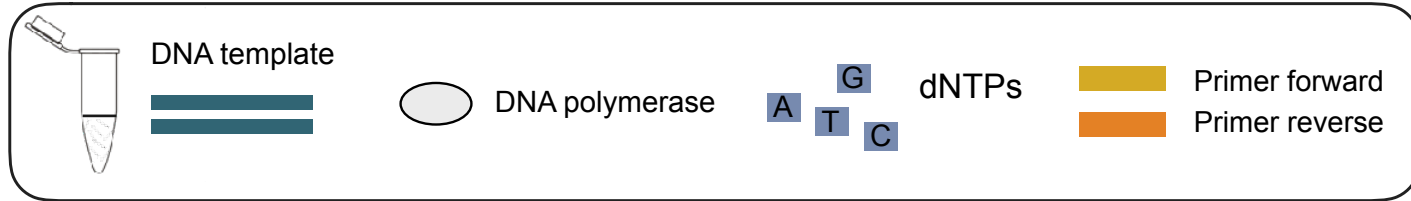
Different steps of PCR



Different steps of PCR



Different steps of PCR

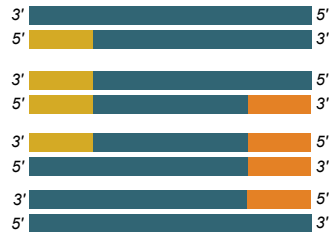


1st cycle



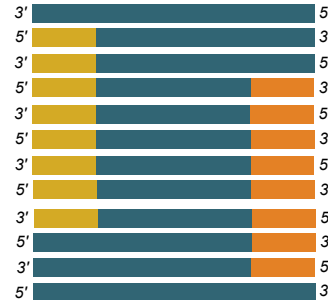
4 amplicons

2nd cycle

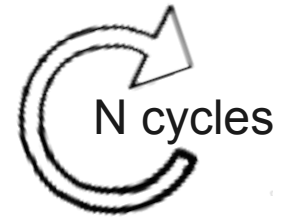


8 amplicons

3rd cycle



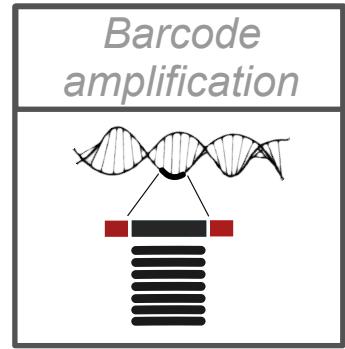
16 amplicons



At the end : *in theory* you have 2^n copies of your original molecule

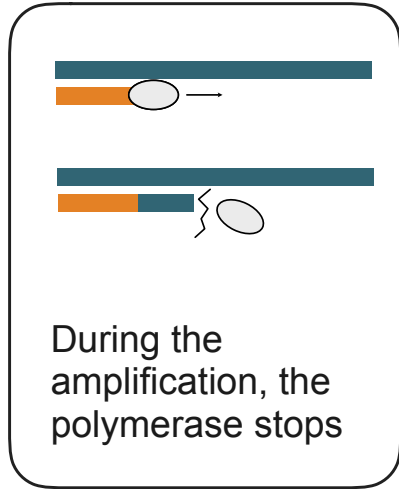
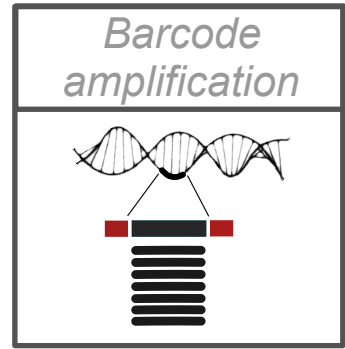


During the PCR...Creation of chimera



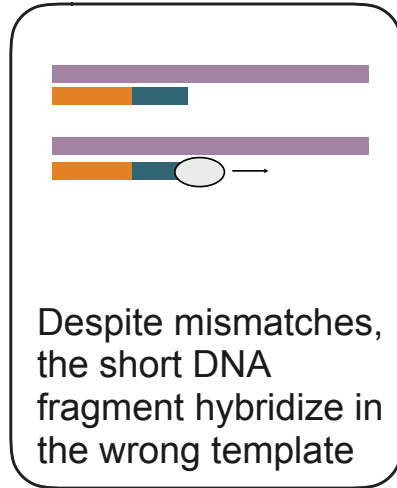
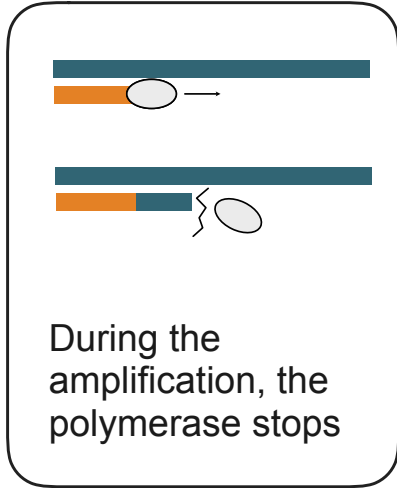
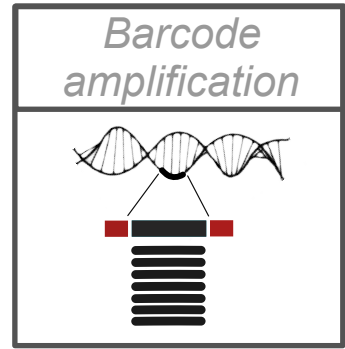


During the PCR...Creation of chimera



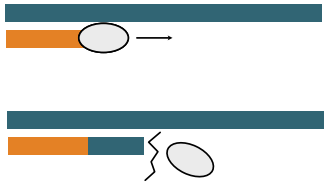
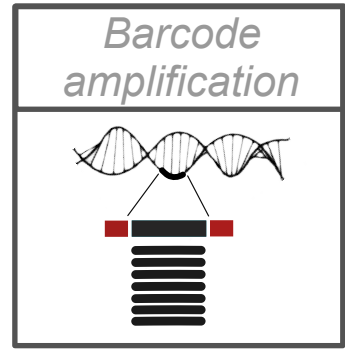


During the PCR...Creation of chimera

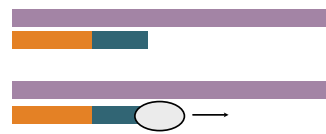




During the PCR... Creation of chimera



During the amplification, the polymerase stops



Despite mismatches, the short DNA fragment hybridize in the wrong template

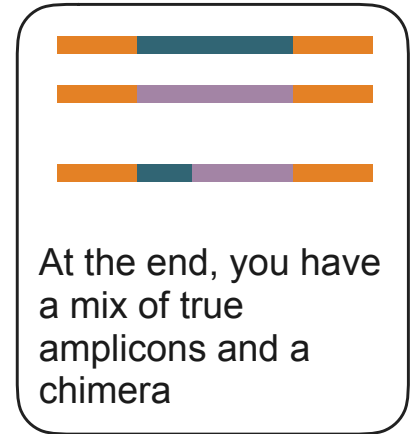
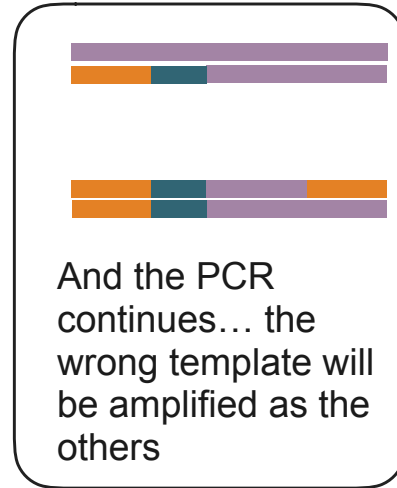
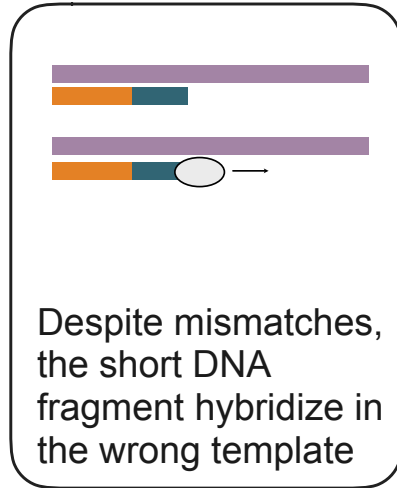
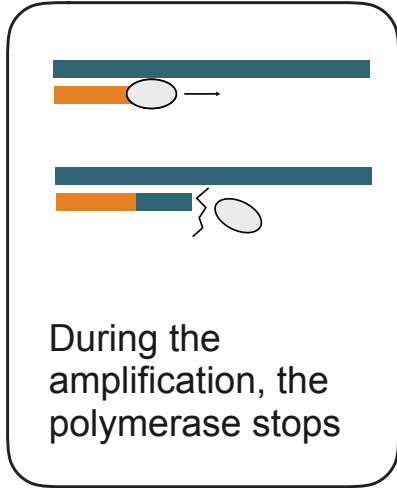
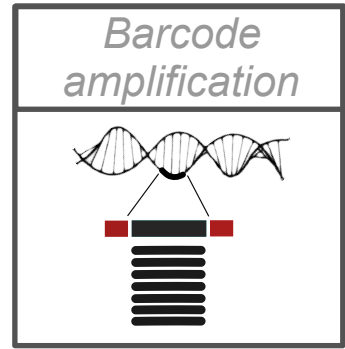


And the PCR continues... the wrong template will be amplified as the others

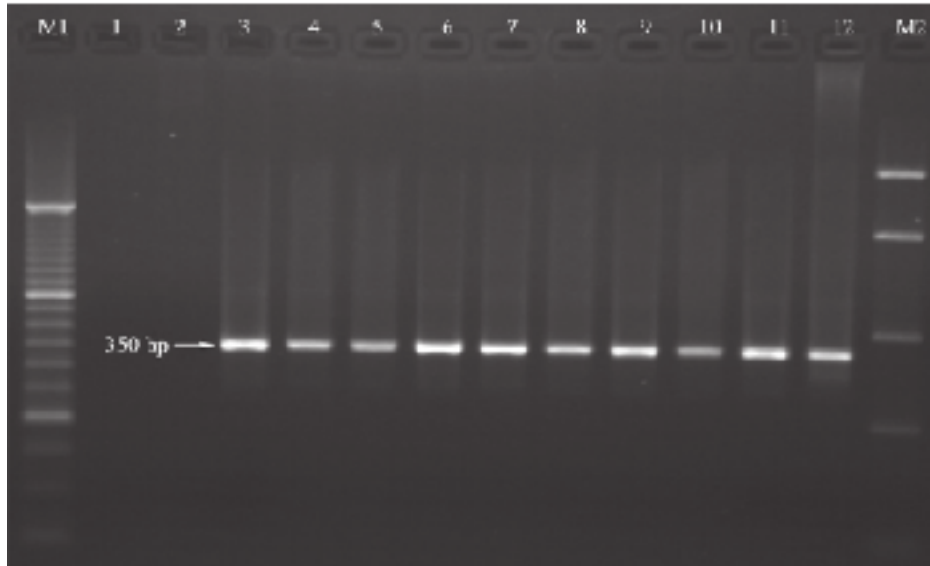




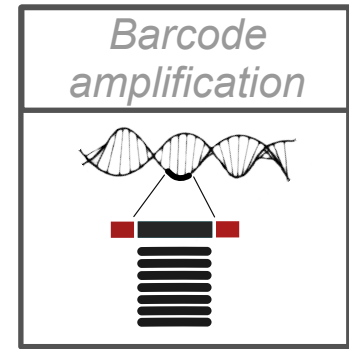
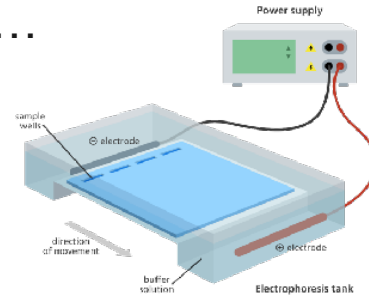
During the PCR... Creation of chimera

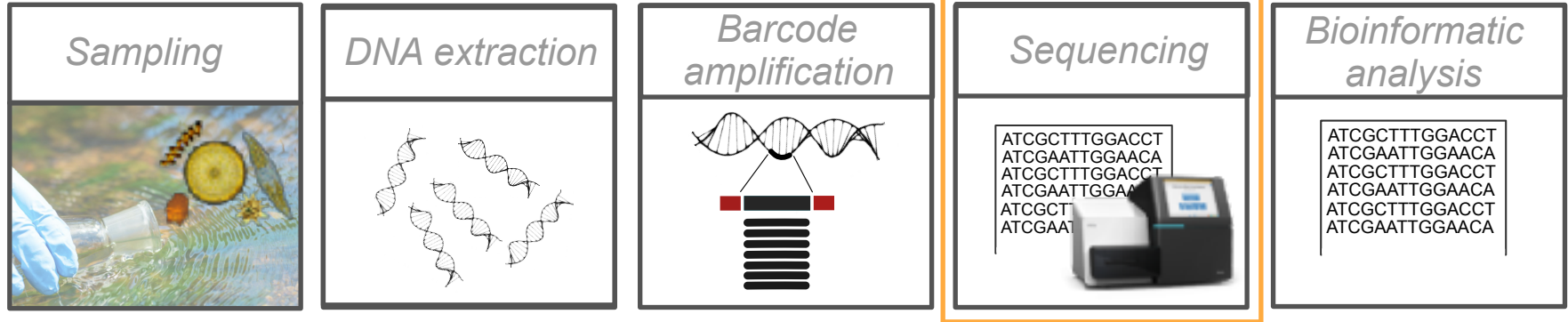


After the PCR...



Check if we have well amplified the barcodes with electrophoresis





Objectives

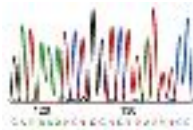
Get the DNA sequence of all amplified barcodes

Sequencing technologies

1st generation (1990s)



Sanger



2nd generation (2005 -)



Roche 454, Illumina, Ion torrent

3rd generation (2011 -)



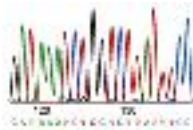
PacBio, Oxford Nanopore

Sequencing technologies

1st generation (1990s)



Sanger



96 seq/run

400-900pb

0.3% error rate

2nd generation (2005 -)



Roche 454, Illumina, Ion
torrent

3rd generation (2011 -)



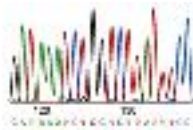
PacBio, Oxford Nanopore

Sequencing technologies

1st generation (1990s)



Sanger



96 seq/run

400-900pb

0.3% error rate

2nd generation (2005 -)



Roche 454, Illumina, Ion torrent

Millions seq/run

150-300 pb

0.1 - 1% error rate

3rd generation (2011 -)



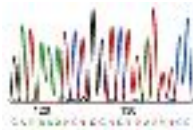
PacBio, Oxford Nanopore

Sequencing technologies

1st generation (1990s)



Sanger



96 seq/run

400-900pb

0.3% error rate

2nd generation (2005 -)



Roche 454, Illumina, Ion torrent

Millions seq/run

150-300 pb

0.1 - 1% error rate

3rd generation (2011 -)



PacBio, Oxford Nanopore

~1 millions seq/run

100 000 pb

12% but decreasing

Sequencing technologies

1st generation (1990s)



Not adapted to
metabarcoding

400-900 pb

0.3% error rate

2nd generation (2005 -)



Roche 454, Illumina, Ion
torrent

Millions seq/run

150-300 pb

0.1 - 1% error rate

3rd generation (2011 -)



Still in development
(error rates)

100 000 pb

12% but decreasing

Illumina MiSeq

Sequencing length of 400-500 bp max

Paired-end sequencing



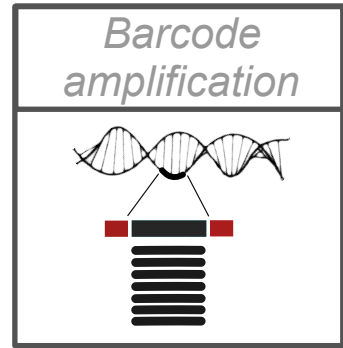
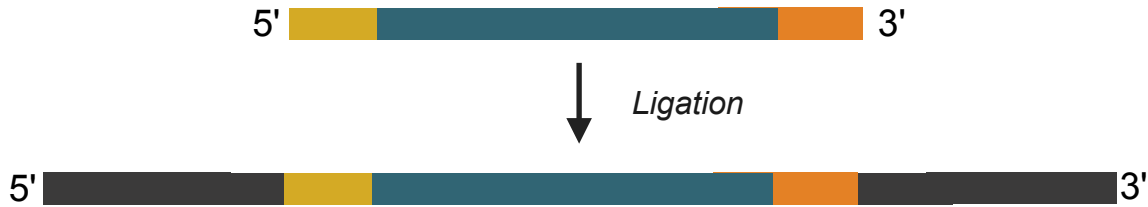
Up to 50 millions paired-end sequences in one run

Run last around 1day



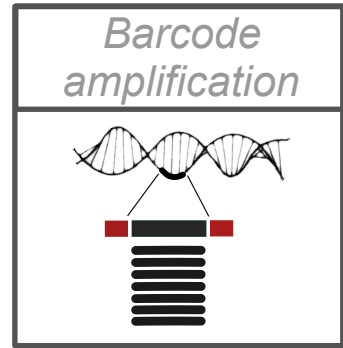
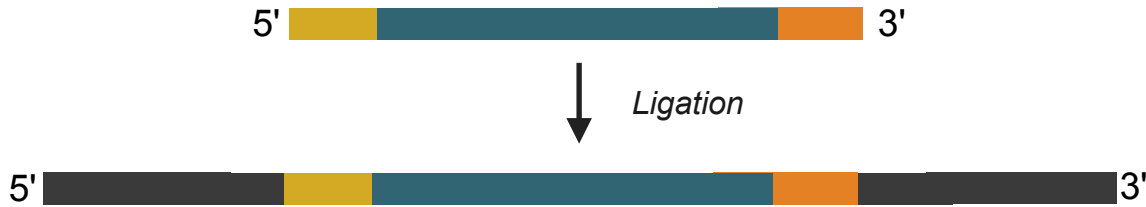
Multiplexed PCR

Amplicons need some additional elements for Illumina technology



Multiplexed PCR

Amplicons need some additional elements for Illumina technology

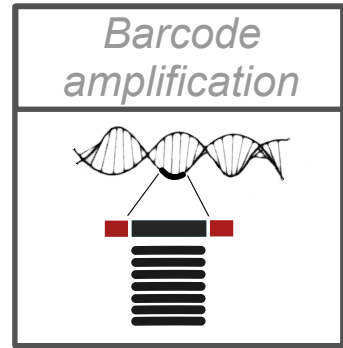


1 run Illumina : 50 000 000 sequences

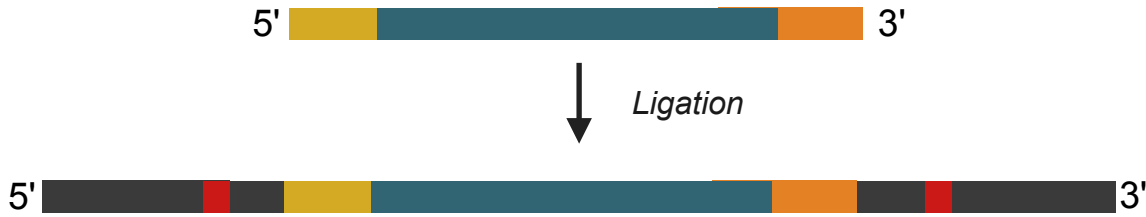
1 sample : 50 000 -150 000 sequences

—> Put multiple sample in one run

Multiplexed PCR



Amplicons need some additional elements for Illumina technology



■ Index

1 run Illumina : 50 000 000 sequences

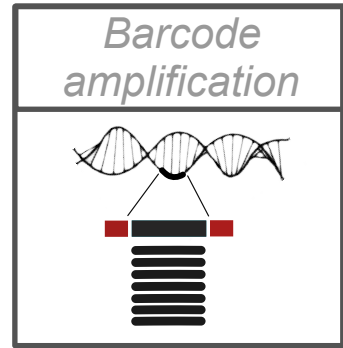
1 sample : 50 000 -150 000 sequences

—> Put multiple sample in one run

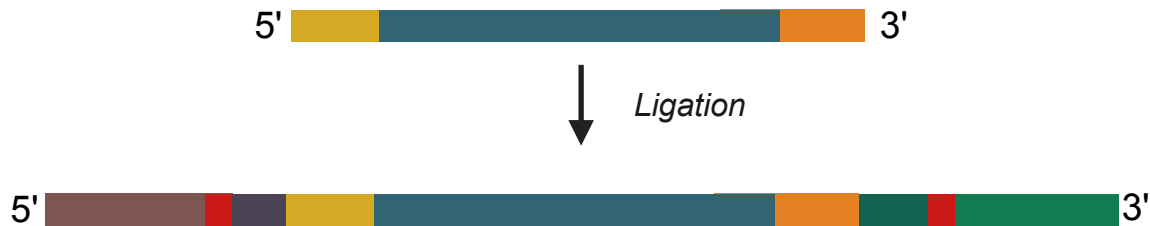
Add a combination of index that will be specific for each sample.

Index : short sequence ~5bp

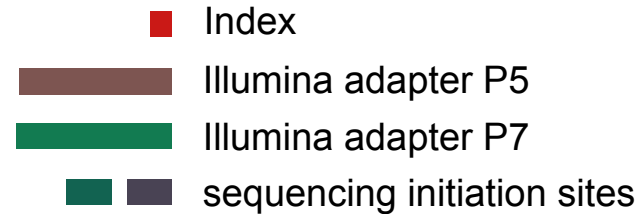
Multiplexed PCR



Amplicons need some additional elements for Illumina technology



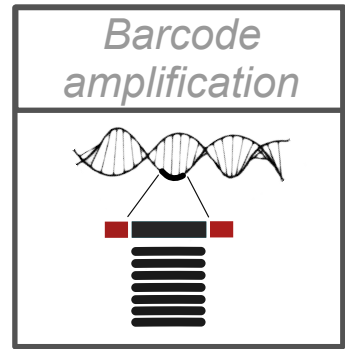
- + Add specific adapters that our amplicon will bind to the Illumina flowcell
- + Sequencing initiation sites



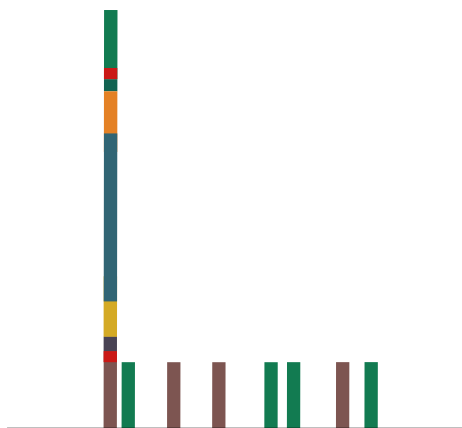
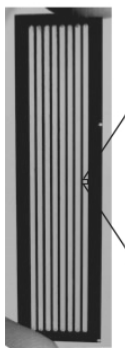
Illumina technology



Ready for sequencing !



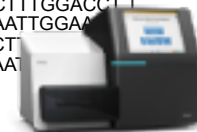
Illumina technology



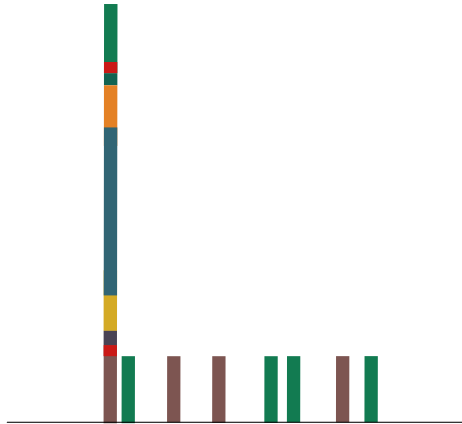
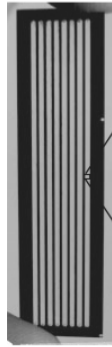
Fixation in the flow
cell

Sequencing

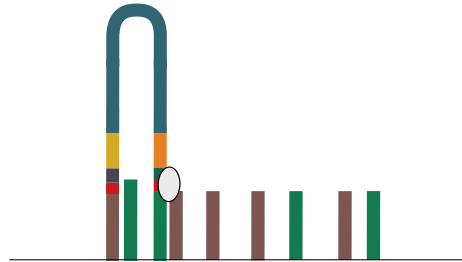
```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
```



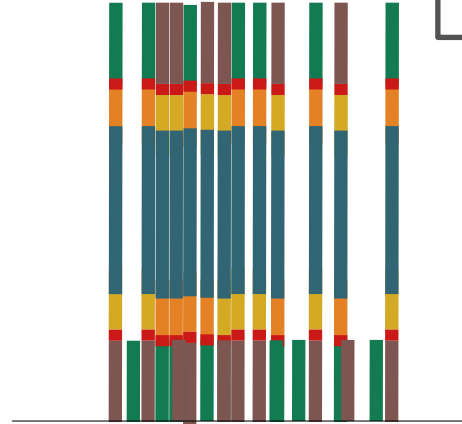
Illumina technology



Fixation in the flow cell



Bridge amplification

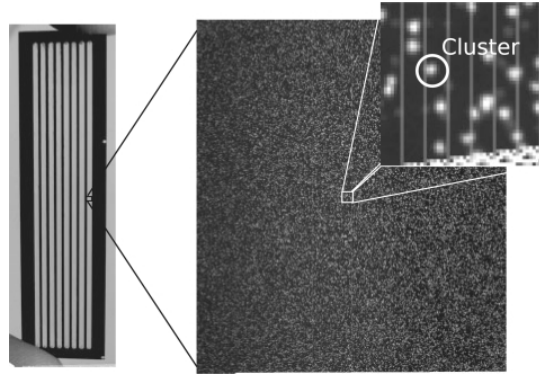


Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
```



Illumina technology

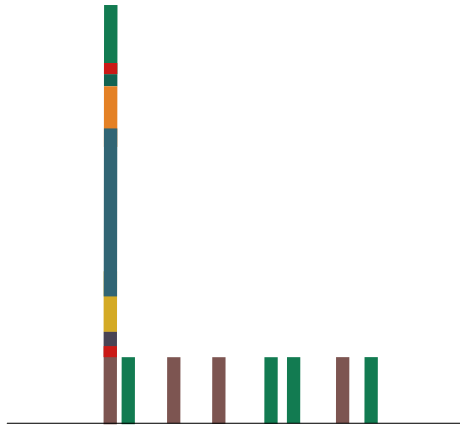


Clusters are visible with a high resolution camera

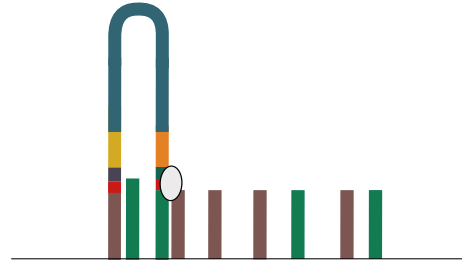
Sequencing

```

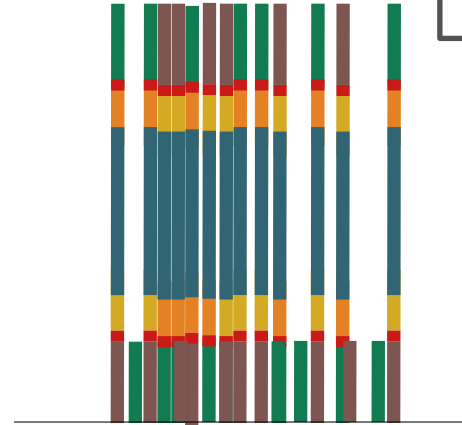
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
    
```



Fixation in the flow cell



Bridge amplification



Illumina technology



Fluorescent nucleotides with reversible terminator

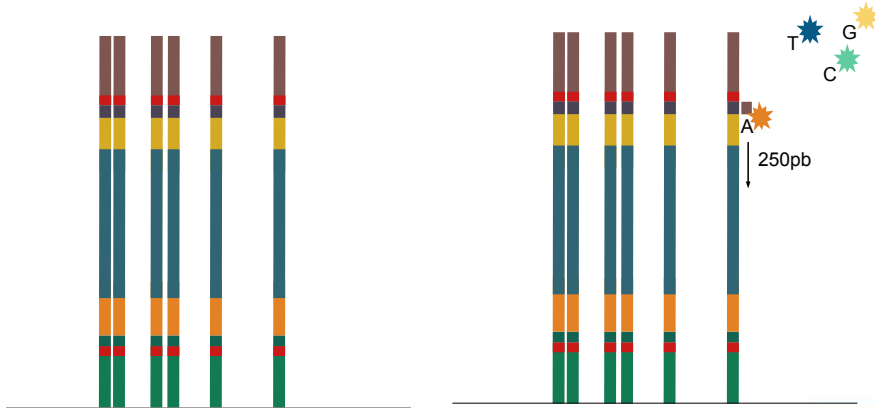
Sequencing

```

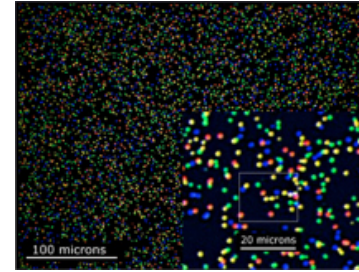
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
    
```



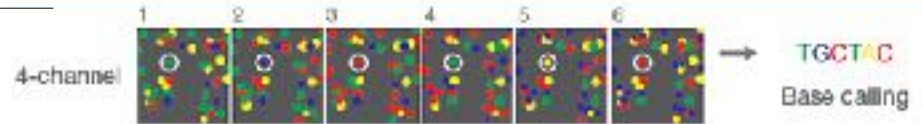
Forward strand



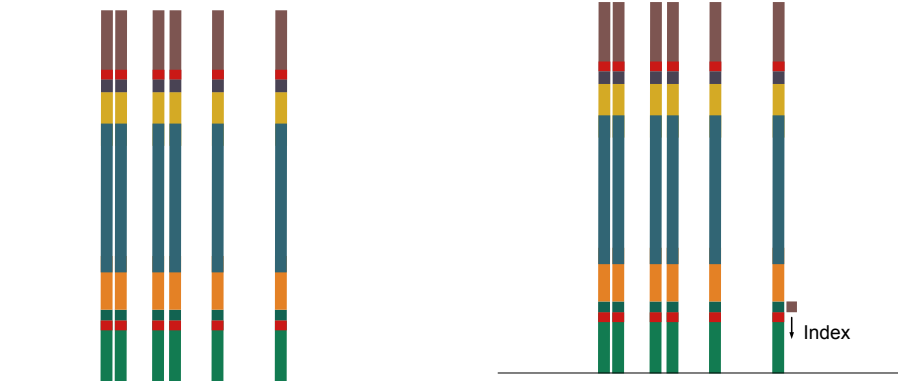
Sequencing of read 1



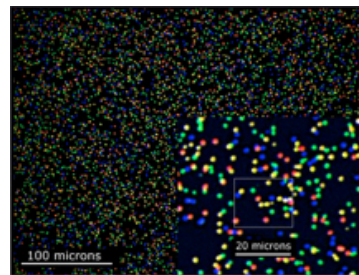
Read 1 



Illumina technology



Sequencing of index 1



Read 1 +

Sequencing

```

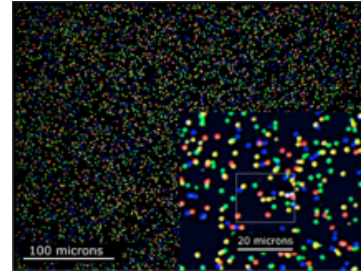
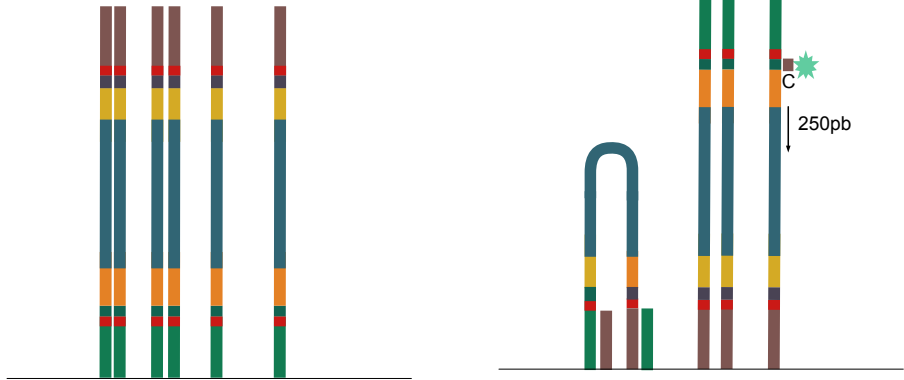
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
            
```



Illumina technology

Forward strand

Reverse strand



Sequencing of read 2 and index 2

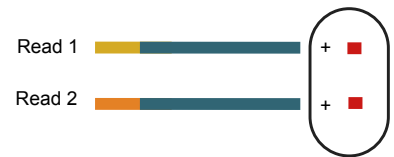
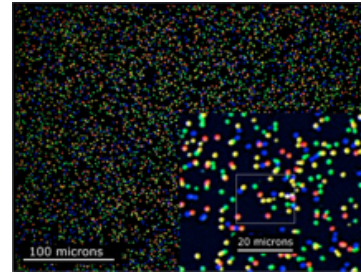
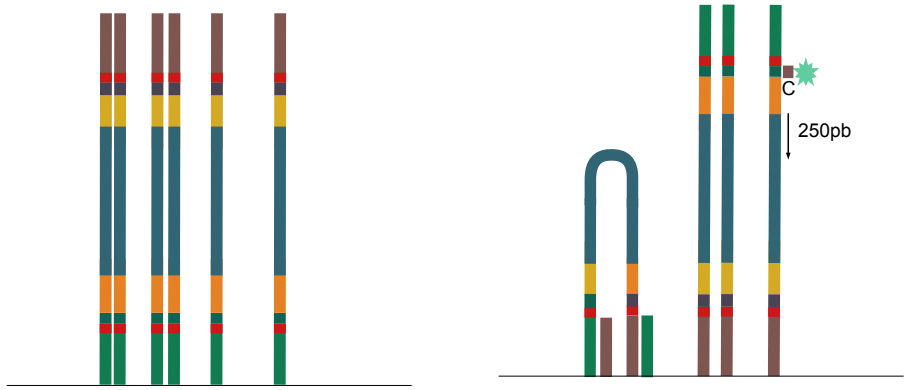
Sequencing

```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
    
```



Illumina technology



Sequencing

```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
            
```



Demultiplexing



Output file



©2011, Illumina Inc. All rights reserved.



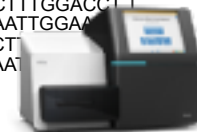
```
@seq_1_ID  
ATCCTGTCGTCTGCT....  
+ 1>AAAFAPAFFF  
@seq_2_ID  
ATCCTGTCGTCCGG...
```

12 Go

```
@seq_1_ID  
ATCCTGTCGTCTGCT....  
+ 1>AAAFAPAFFF  
@seq_2_ID  
ATCCTGTCGTCCGG...
```

Sequencing

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGA  
ATCGCT  
ATCGAAT
```



Output file : fastq file

Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
```



Define

@M01028:94:00000000-AHT10:1:1101:15604:1824 1:N:0:10

ATGCTCCAGCAGCTGCGGTAATACGGAGGATGCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTCCGCAGGCGGATCAATAAGT
CAGTGGTGAAAGCCATAGCTCAACTATGGAAGTCCATTGAAACTGTTGATCTTGAGTCTAGTTGACGTAGGCGGAATGTGACTT
GTAGCGGTGAAATGCATAGATATGTCACAGAACACCAATTGCGAAGGCAGCTTACGTAAGTACGACTGACGCTCTTGGAC

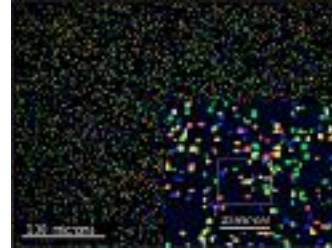
+

3AABAFFFFFFFFGGGGGEGGGHHHGEGG2FFB3BDEEGG2F5FECBEEHDFGGFFFG3AFFG?GCCE0>EE?FG/
>EE3FF3DG44FGHEHGBGE3GGHGF2?GGFF?GFCGFBHHFHFBFB@2F2
FHGBHHD1GHGGH1CHFF11GFBD1CFCAD1<@CCC0CC0;0<CGF0C-@GA.0;;0C0;;FFCCFB9090;/;9/
BE.;FB09-9-:....FF;B.;//9.9.//9.9..9B///

@M01028:94:00000000-AHT10:1:1101:12447:1845 1:N:0:10

Sequence

Quality score



```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
    
```



Define

@M01028:94:00000000-AHT10:1:1101:15604:1824 1:N:0:10
 ATGCTCCAGCAGCTGCGGTAATACGGAGGATGCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTCCGCAGGCCGGATCAATAAGT
 CAGTGGTGAAAGCCATAGCTCAACTATGGAAGTCCATTGAAACTGTTGATCTTGAGTCTAGTTGACGTAGGCCGAATGTGACTT
 GTAGCGGTGAAATGCATAGATATGTCACAGAACACCAATTGCGAAGGCAGCTTACGTAAGTCTGACTGACGCTCTTGGAC
 +
 3AABAFFFFFFFFGGGGGEGGGHHHGEGG2FFB3BDEEGG2F5FECBEEHDFGGFFFEFG3AFFG?GCCE0>EE?FG/
 >EE3FF3DG44FGHEHGBGE3GGHGF2?GGFF?GFCGFBHHFHFBF@2F2
 FHGBHHD1GHGGH1CHFF11GFBD1CFCAD1<@CCC0CC0;0<CGF0C-@GA.0;;0C0;;FFCCFB9090;/;/9/
 BE.;FB09-9-:.....FF;B.;;/9.9.//9.9..9B///
@M01028:94:00000000-AHT10:1:1101:12447:1845 1:N:0:10





Sequence

Quality score







Get a fastq file for each sample

Read 1

@ Seq 1

+

@ Seq 2

+

@ Seq 3

Read 2

@ Seq 1

+

@ Seq 2

+

@ Seq 3

Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGA
ATCGCT
ATCGAAT
```



~ 100 000
sequences

Metabarcoding steps

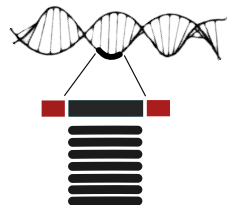
Sampling



DNA extraction



*Barcode
amplification*



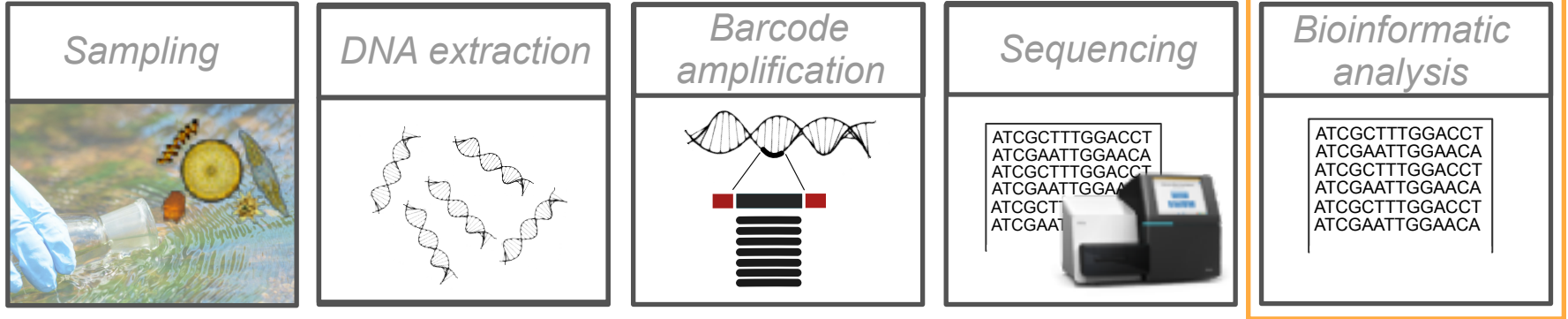
Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```



*Bioinformatic
analysis*

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGAATTGGAACA
```



Objectives

Correct the reads from PCR and sequencing errors, and try to reconstruct meaningful « groups » of sequences that can be equivalent as species

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```

- 1 Get complete barcode sequences of high quality**
- 2 « Reconstruct » species**
- 3 Taxonomic assignation**

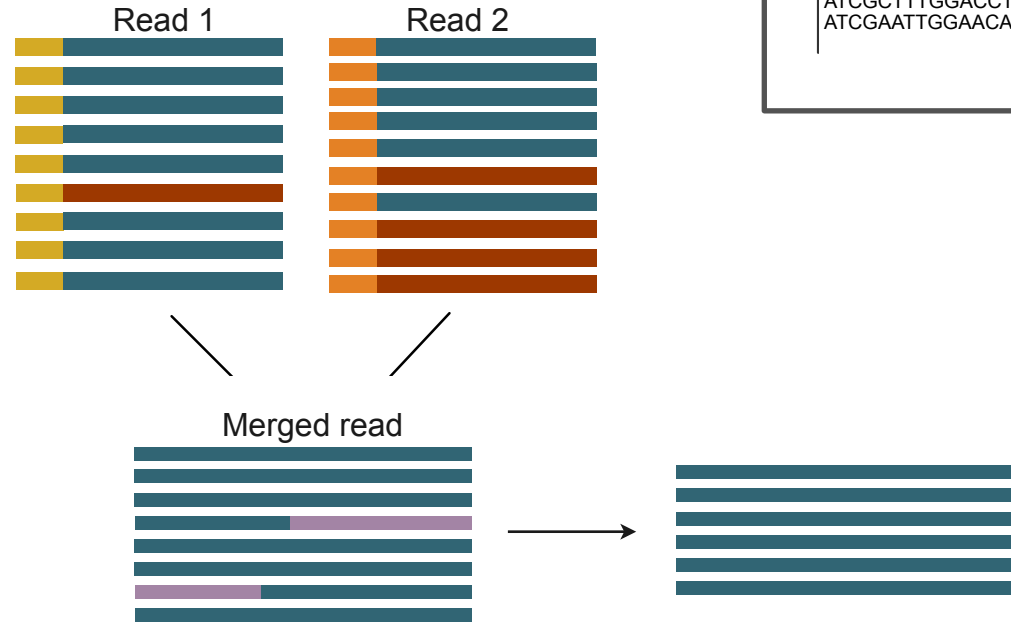
1 Get complete barcode sequences of high quality

Filter reads based on their quality profile

Remove primers

Merge the read 1 and read 2

Remove chimera



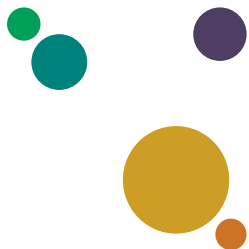
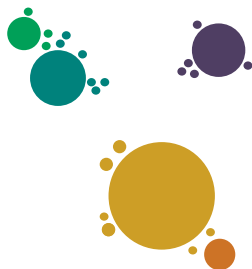
```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```

2 « Reconstruct » species

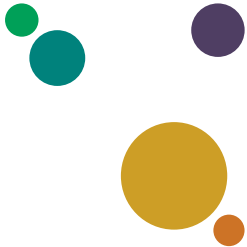
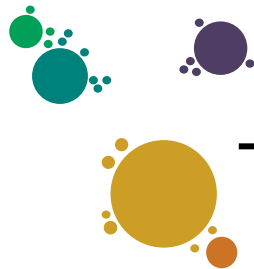
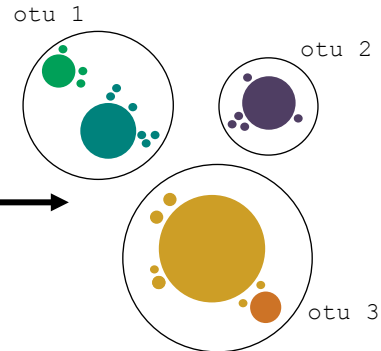
```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```

Why? Because we have to deal with 2 things : natural intraspecific variability and sequencing errors.

At the end you don't have a true species as it is artificially reconstructed sequence

2 « Reconstruct » speciesSample sequencesAmplicon reads

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```

2 « Reconstruct » speciesSample sequencesAmplicon readsClustering (OTUs)

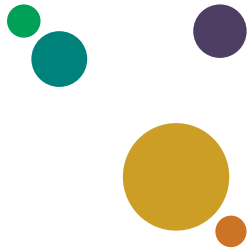
```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```

2 « Reconstruct » species

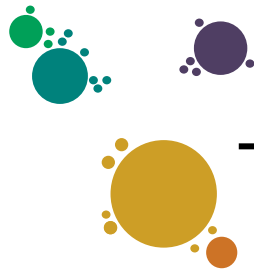
```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```

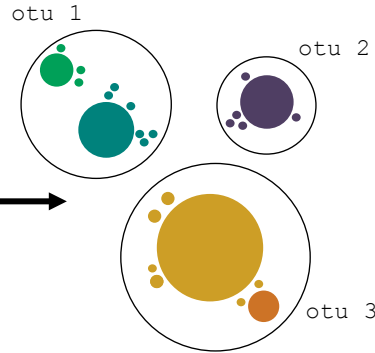
Sample sequences



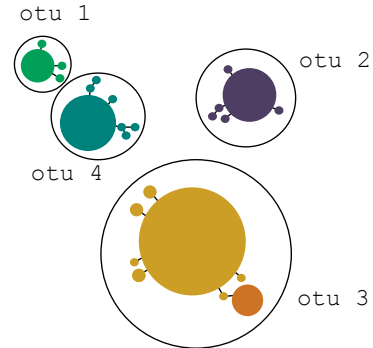
Amplicon reads



Clustering (OTUs)



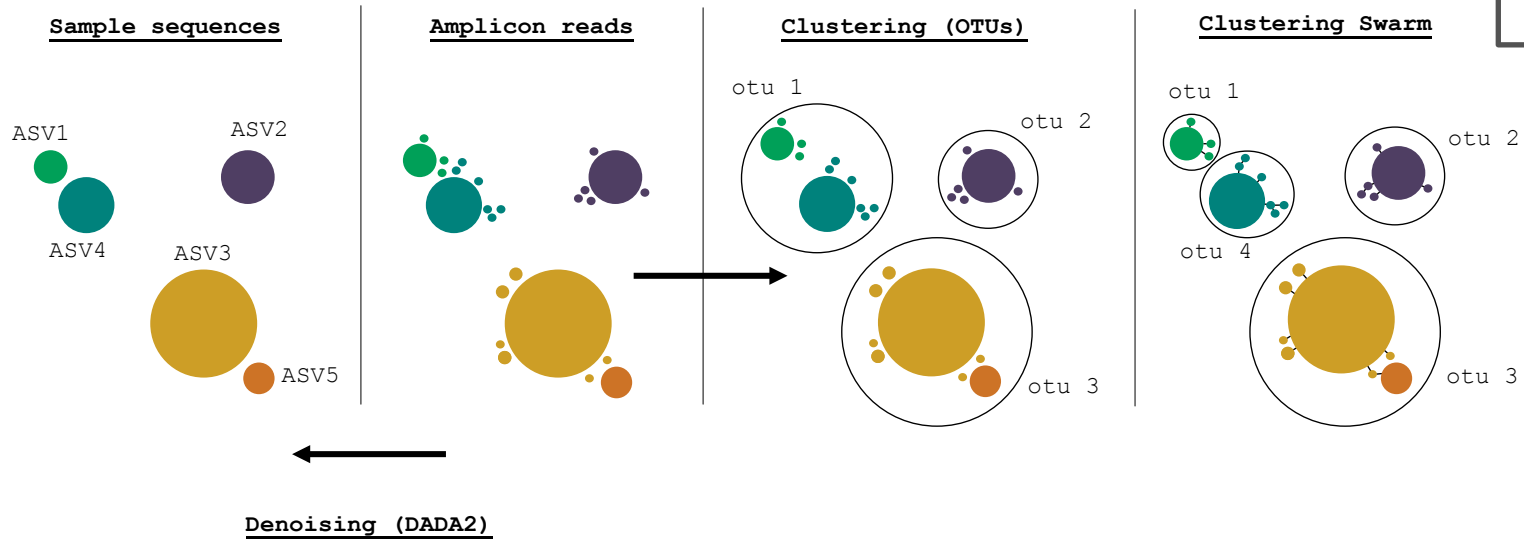
Clustering Swarm



2 « Reconstruct » species

```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```



3 Taxonomic assignment

OTUs/ASV
sequences



Sequences are
compared to
reference
databases

Reference database



*Bioinformatic
analysis*

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```

Sequences are not
aligned as it would
be too time-
consuming and
resource expensive

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```

- 1 Get complete barcode sequences of high quality
- 2 « Reconstruct » species
- 3 Taxonomic assignation



**Bioinformatic
workflow**

Different workflows

*Bioinformatic
analysis*

@seq_1_ID

ATCCTGTCGTCTGCT....

+ 1>AAAFAFAFFF

@seq_2_ID

ATCCTGTCGTCCGG...

@seq_1_ID

ATCCTGTCGTCTGCT....

+ 1>AAAFAFAFFF

@seq_2_ID

ATCCTGTCGTCCGG...

12 Go

= millions of reads

= thousands of OTUs/ASVs



ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

**Use of bioinformatic
(it is not an option)**

Different workflows

*Bioinformatic
analysis*

```
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA  
ATCGCTTTGGACCT  
ATCGAATTGGAACA
```



There are different softwares available to do all the bioinformatic analysis :



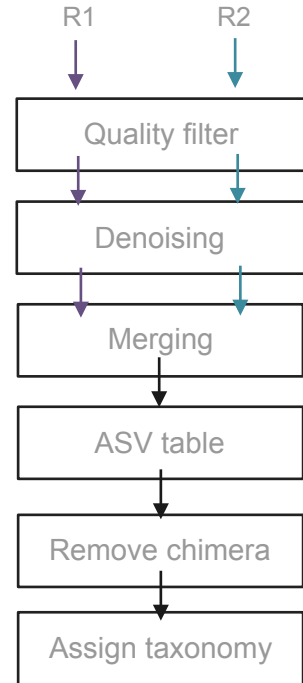
Mothur : open
source software
Own language



QIIME: open
source software
Python



DADA2 : open
source software
R language



DADA2 pipeline

Online tutorial



DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution

DADA2
Amplicon Sequencing. Exactly.

The DADA2 R package is the software you need for DADA2 analysis. For the R package, see the [DADA2 R package](#).

Installation

```

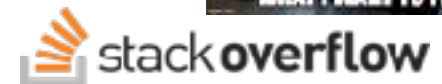
install.packages("dada2")
library(dada2)
  
```

Tutorials

Learn the DADA2 pipeline from the ground up with the DADA2 tutorial. The tutorial is available in [HTML](#) or [PDF](#) format. The tutorial is available in [HTML](#) or [PDF](#) format.



Widely used language within biologists



Metabarcoding steps

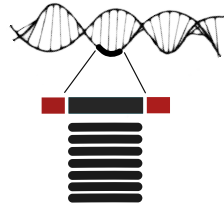
Sampling



DNA extraction



Barcode amplification



Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```

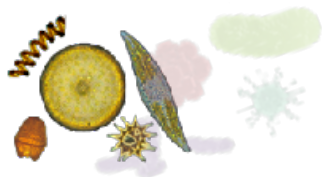
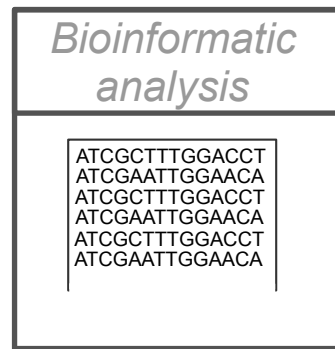
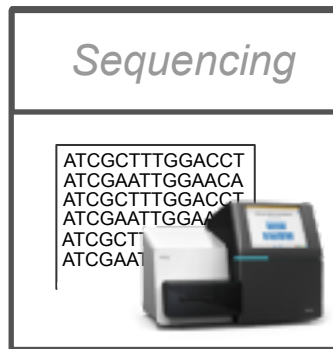
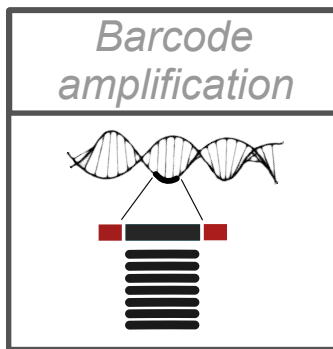
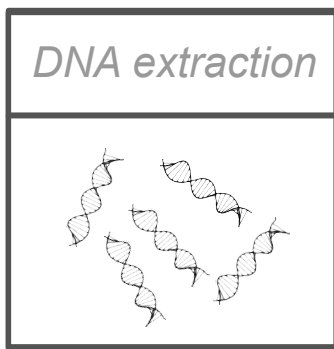


Bioinformatic analysis

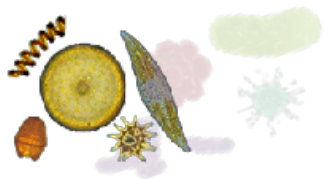
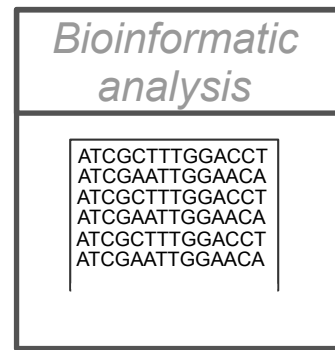
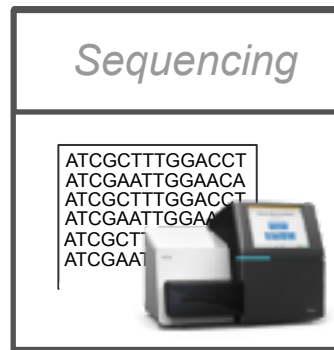
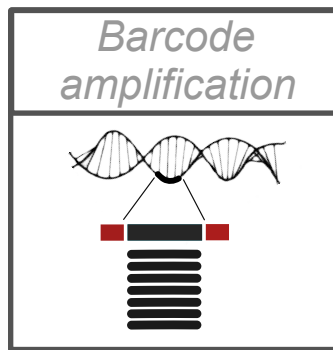
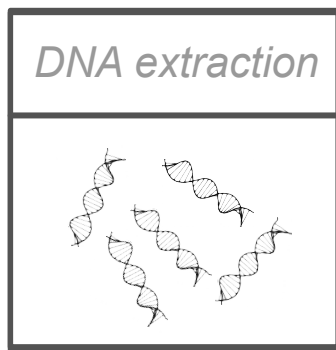
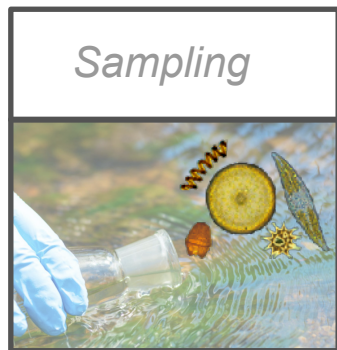
```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```



Metabarcoding steps - without optimization




Metabarcoding steps - without optimization

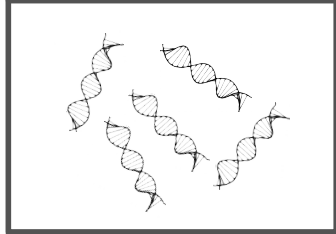


Metabarcoding steps - without optimization

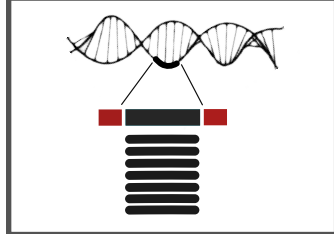
Sampling



DNA extraction



Barcode amplification



Sequencing



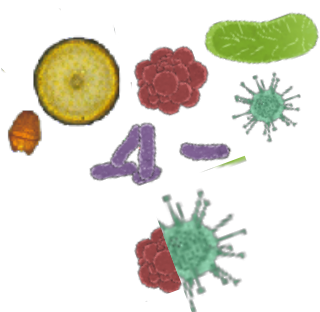
```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```

Bioinformatic analysis

```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```

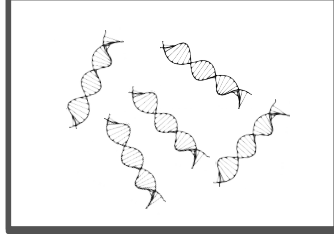


Metabarcoding steps - without optimization

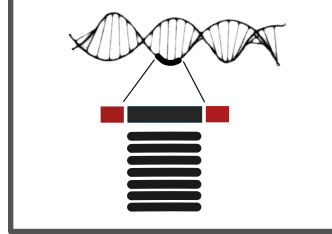
Sampling



DNA extraction



Barcode amplification



Sequencing

```

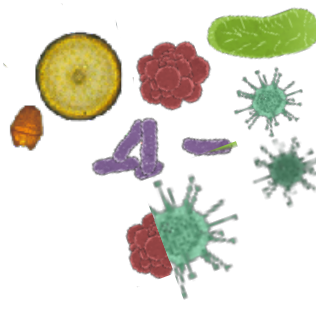
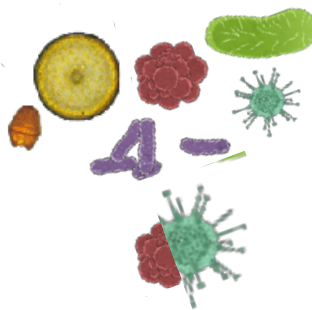
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```



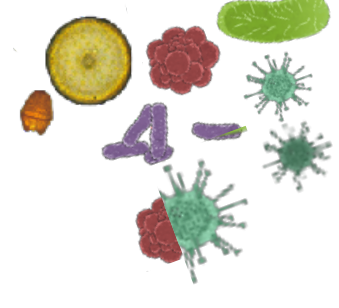
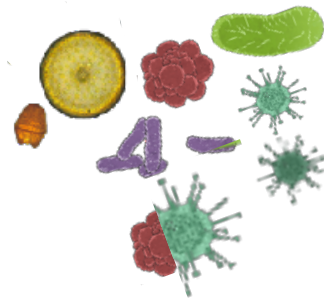
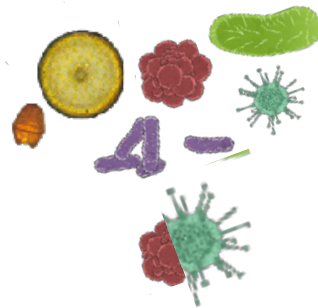
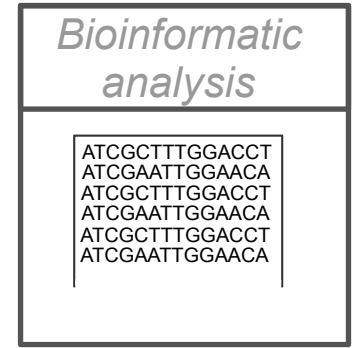
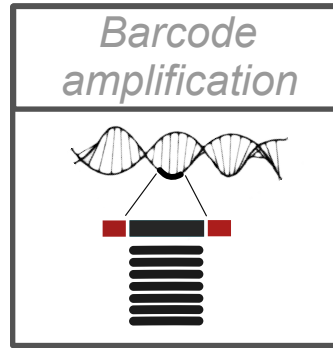
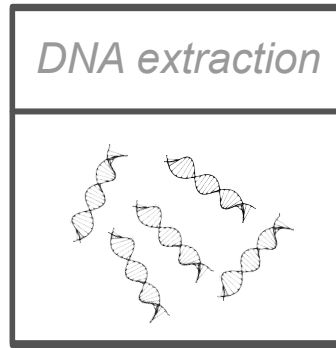
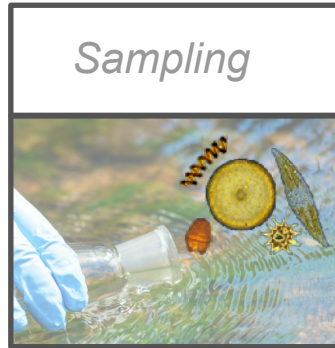
Bioinformatic analysis

```

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```




Metabarcoding steps - without optimization

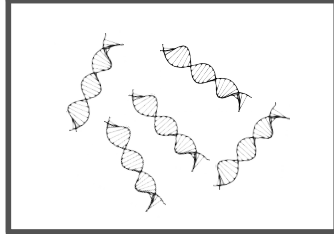


Metabarcoding steps - without optimization

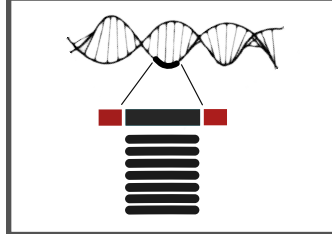
Sampling



DNA extraction




Barcode amplification



Sequencing

```

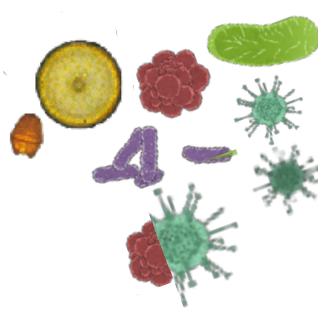
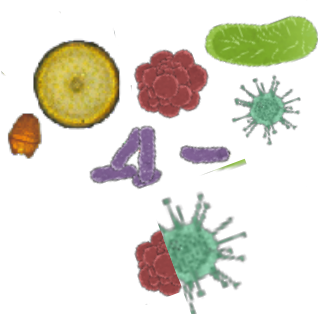
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```



Bioinformatic analysis

```


ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
    
```



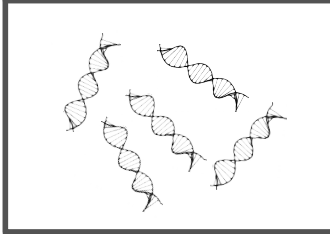
NOT APPROVED

Metabarcoding steps - with optimization

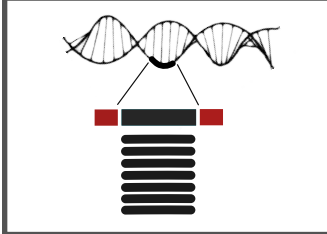
Sampling



DNA extraction




Barcode amplification



Sequencing

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```

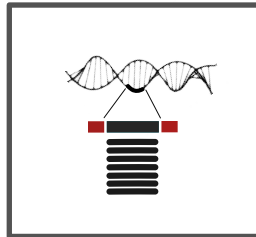
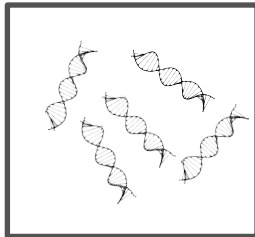


Bioinformatic analysis

```
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
```



You'll see in the next presentations how this workflow was designed, optimized, tested and verified to do biomonitoring of diatoms



How we are currently doing this for the biomonitoring of phytoplankton...



Any questions?

