The benefits of long read HiFi sequencing for metabarcoding analysis

Adrien Castinel, Jean Mainguy, Sylvie Combes, Carole Iampietro, Christine Gaspin, Denis Milan, Cécile Donnadieu, Claire Hoede, Géraldine Pascal, Olivier Bouchez

HAL Id: hal-03740586
https://hal.inrae.fr/hal-03740586
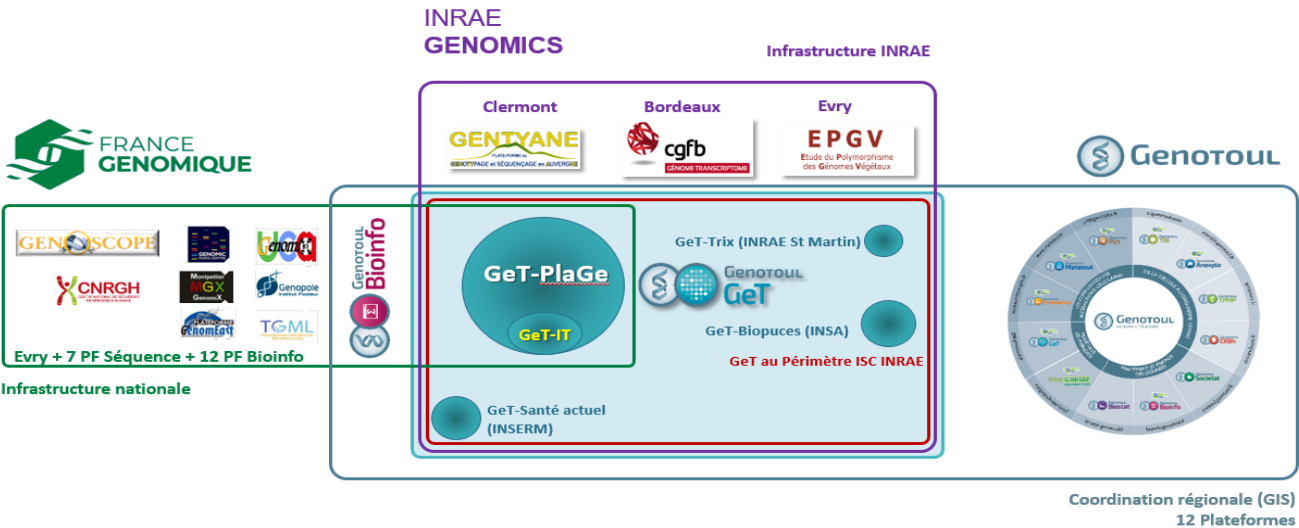
Submitted on 29 Jul 2022

GeT-PlaGe: sequencing core facility

- **GeT (Genome and Transcriptome):**
  Multi-site facility of Genotoul close to their scientific community

- **Member of national and local networks**
  Genotoul, INRAE genomics, France Genomique…
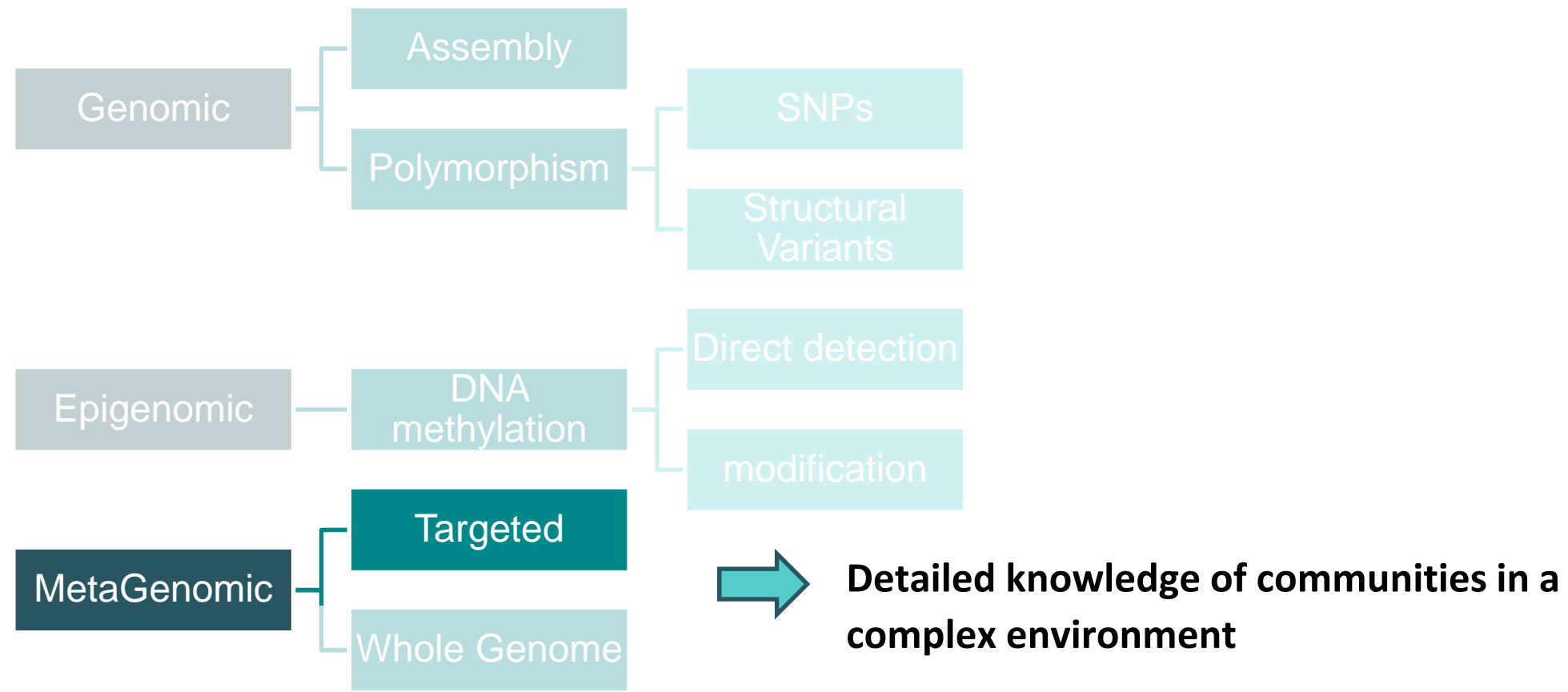
- **GeT-PlaGe Equipment:**

2 Sequel II

SeqOccIn

SeqOccIn Project: carried by GeT-PlaGe and Bioinfo Genotoul

Aim of the project:
Acquire expertise in the use of long read sequencing technology in 3 domains:

Genomic
- Assembly
- Polymorphism
  - SNPs
  - Structural Variants

Epigenomic
- DNA methylation
  - Direct detection
  - modification

MetaGenomic
- Targeted
- Whole Genome

→ **Detailed knowledge of communities in a complex environment**
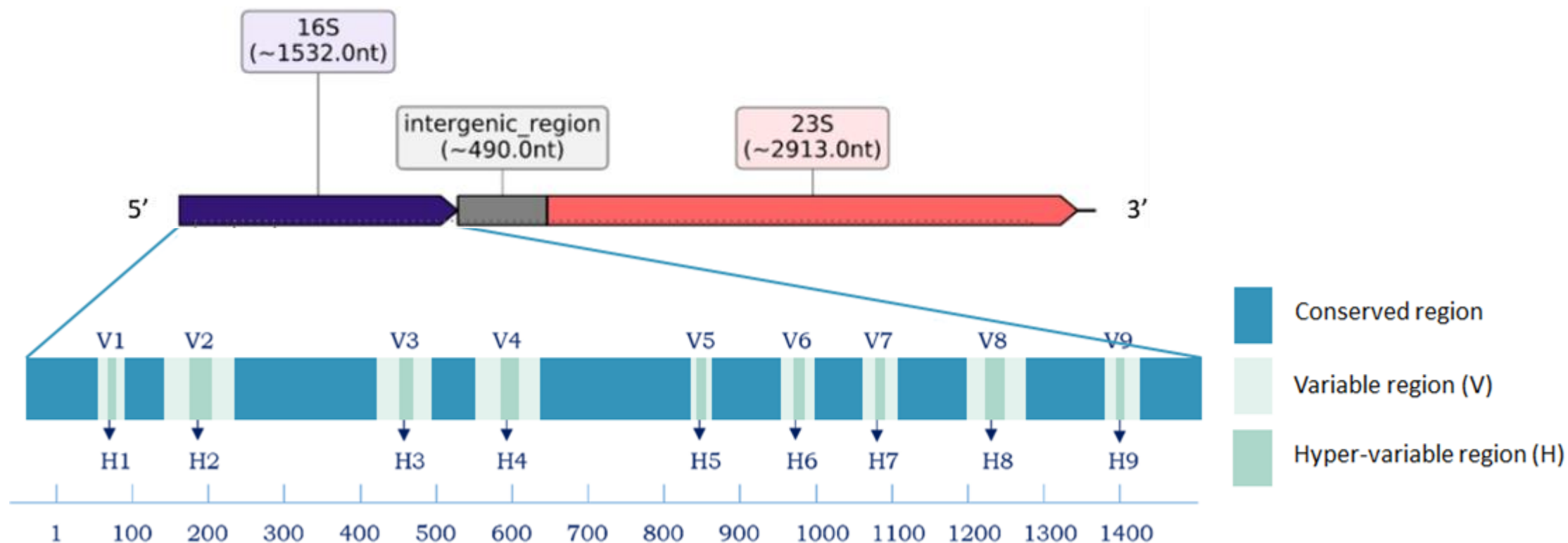
METABARCODING

What is metabarcoding?

- Some DNA fragments are highly conserved within a species and variable between species. These are the genetic markers or barcodes.

- Metabarcoding, by identifying barcodes through sequencing, allows blind identification of all species present in a sample at once.
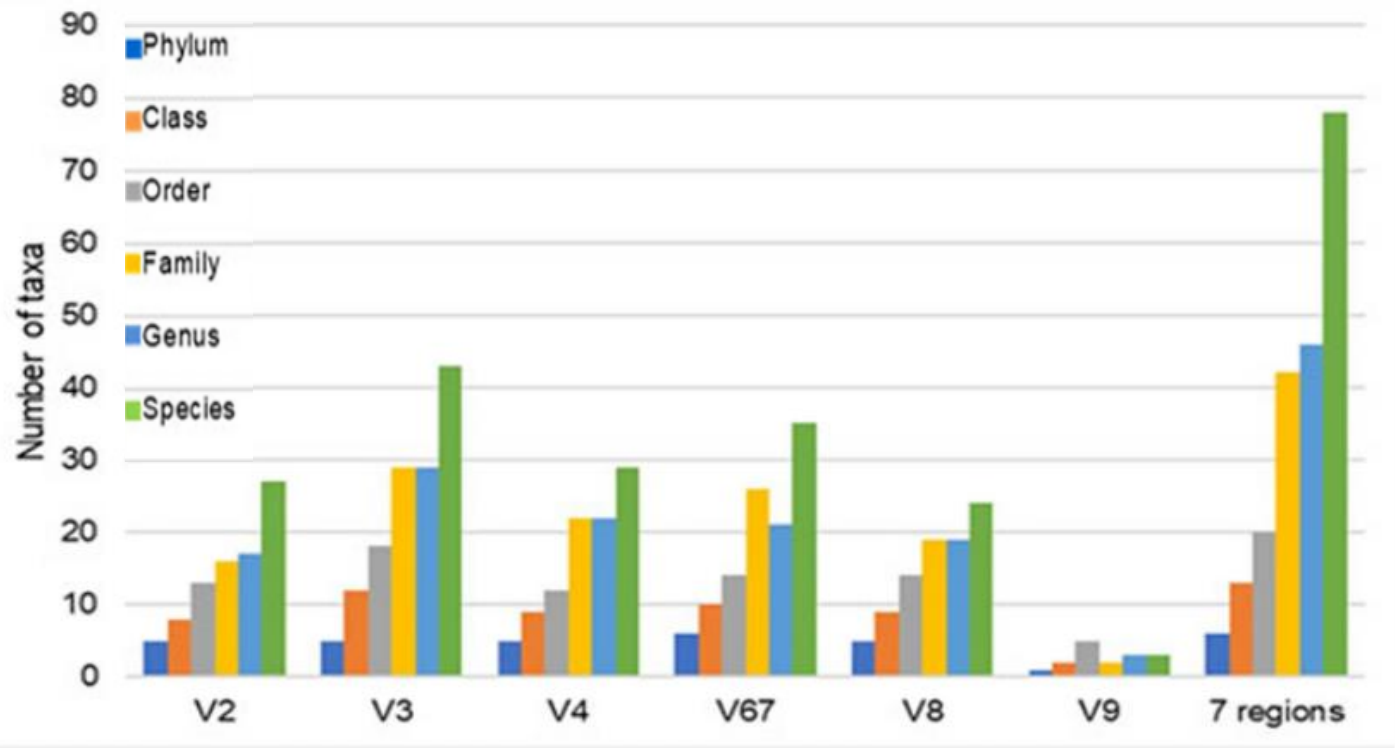

Community analysis → DNA extraction → Barcode amplification → Amplicon sequencing → Bioinformatics process → Diversity characterisation
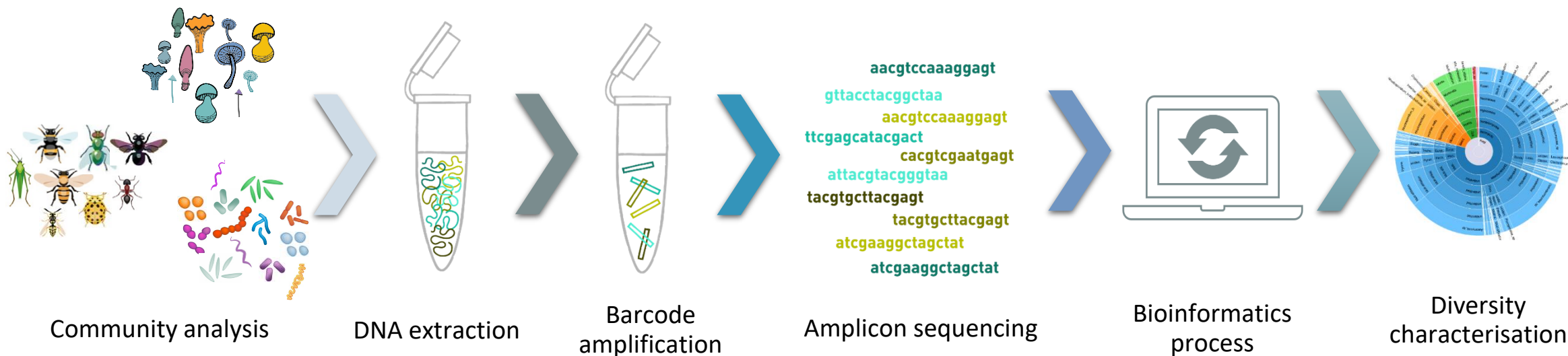
16S: a bacterial marker

- Gene encoding the 16S subunit of ribosomal RNA
- Highly conserved gene in all bacteria
- Highly conserved regions (for primers) interleaved with variable regions (bacteria identification)
- Choice of the regions depends on the community to be analyzed

Number of taxa identified at each taxonomical using individual and concatenated hypervariable regions of the 16S rRNA gene.

- Analysis of a combination of sequences from V2 to V9 regions identified more taxa.
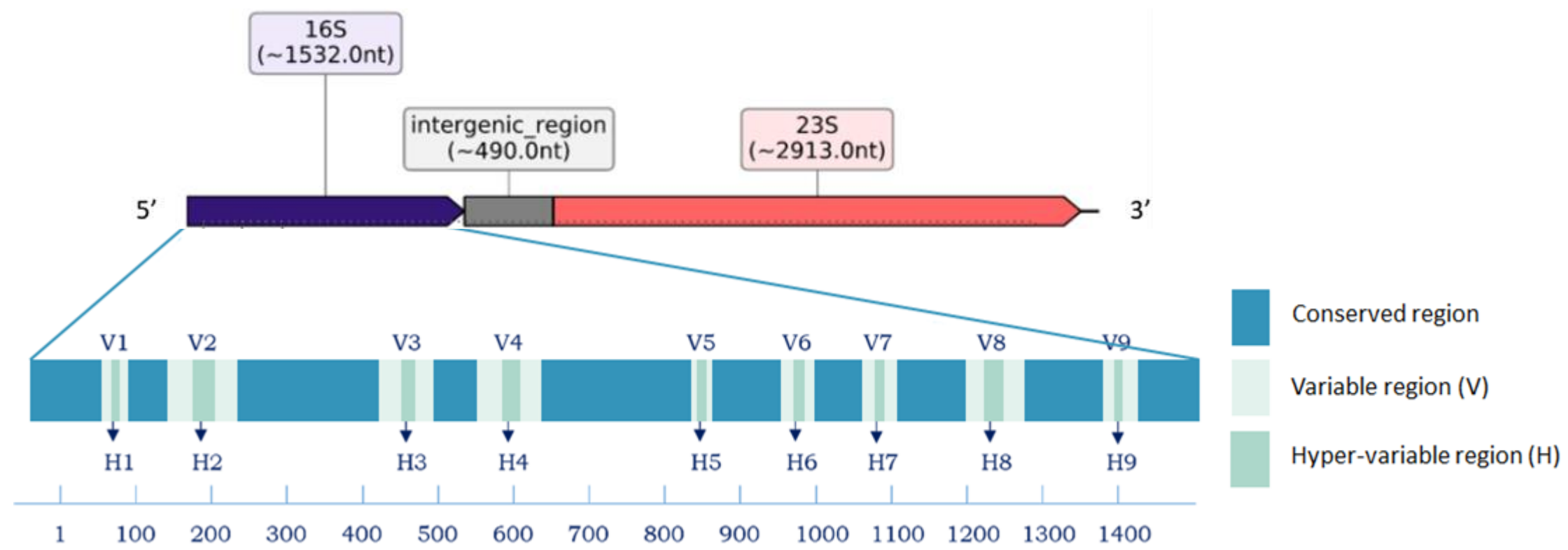
Original Paper | Published: 22 November 2020

Comparison of different hypervariable regions of 16S rRNA for taxonomic profiling of vaginal microbiota using next-generation sequencing
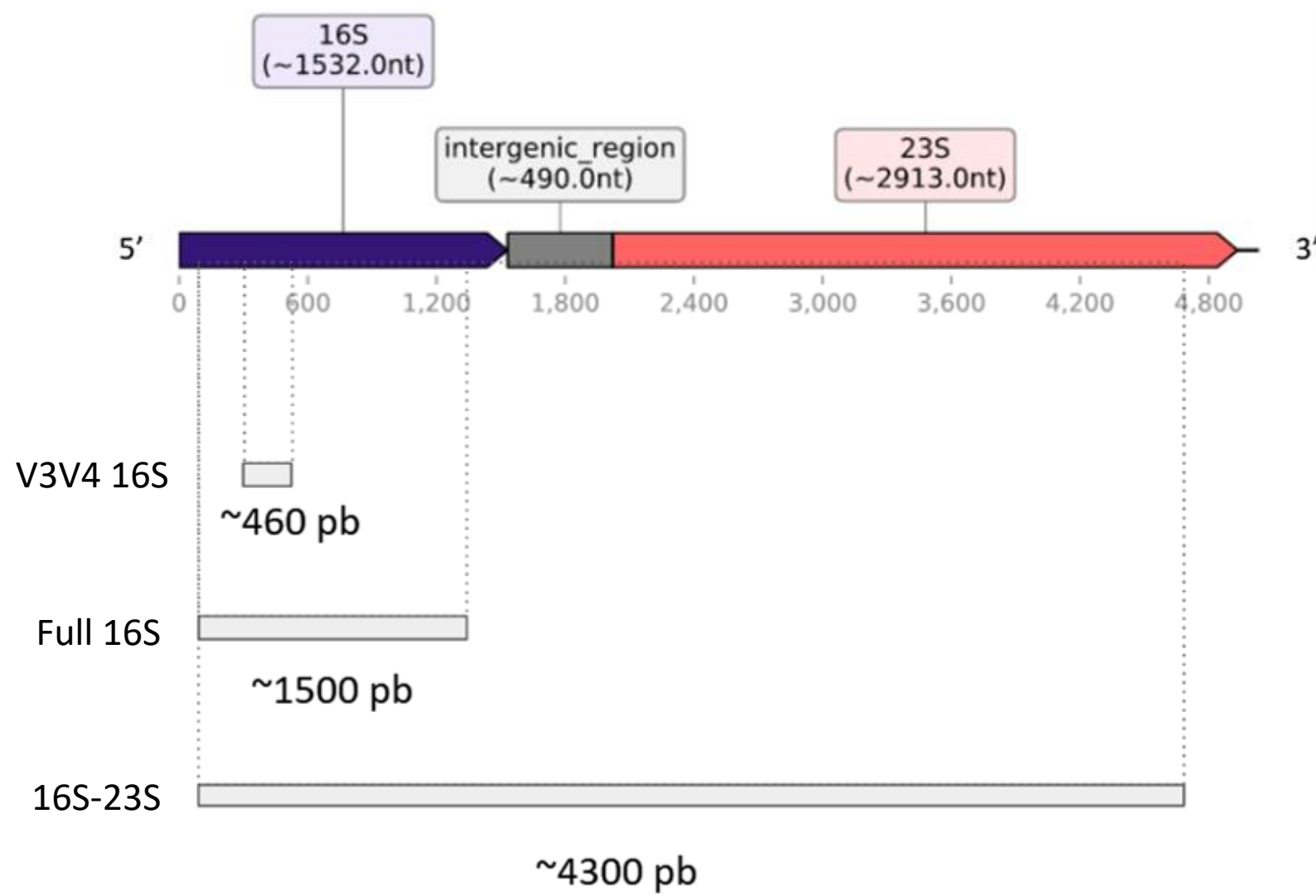
Auttawit Sirichoat, Nipaporn Sankuntaw, Chulapan Engchanil, Pranom Buppasiri, Kiatichai Faksri, Wises Namwat, Wasun Chantratita & Viraphong Lulitanond ✉

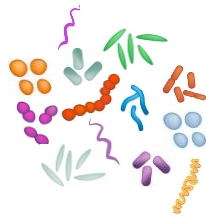# rRNA genes - metabarcoding long read

What would be the contribution of "long reads" in metabarcoding?

Focus on 16S / 16S23S markers

Known biaises in metabarcoding

SeqOccIn

Community analysis

DNA extraction

Barcode amplification

Amplicon sequencing

```
aacgtccaaaggagt
gttacctacggctaa
   aacgtccaaaggagt
ttcgagcatacgact
      cacgtcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
      tacgtgcttacgagt
atcgaaggctagctat
   atcgaaggctagctat
```

Bioinformatics process

✓ 16S copy number

✓ Horizontal gene transferts

✓ DNA extraction kits/protocols

PCR
✓ Polymerase efficiency
✓ Polymerase contaminations
✓ non homogenous amplification

✓ Sequencing errors
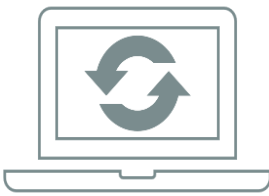
✓ Analysis softwares

✓ Databases exhaustivity

Metabarcoding process for long read

Community analysis

DNA extraction

Barcode amplification

Amplicon sequencing

aacgtccaaaggagt
gttacctacggctaa
aacgtccaaaggagt
ttcgagcatacgact
cacgtcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
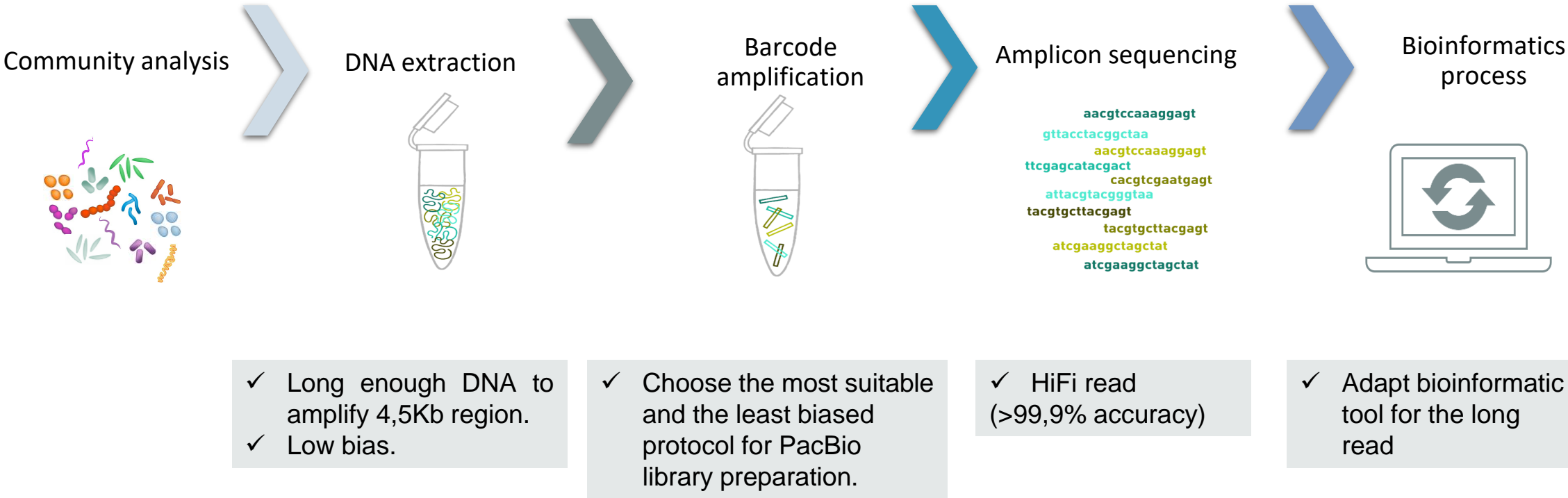tacgtgcttacgagt
atcgaaggctagctat
atcgaaggctagctat

Bioinformatics process

✓ Long enough DNA to amplify 4,5Kb region.
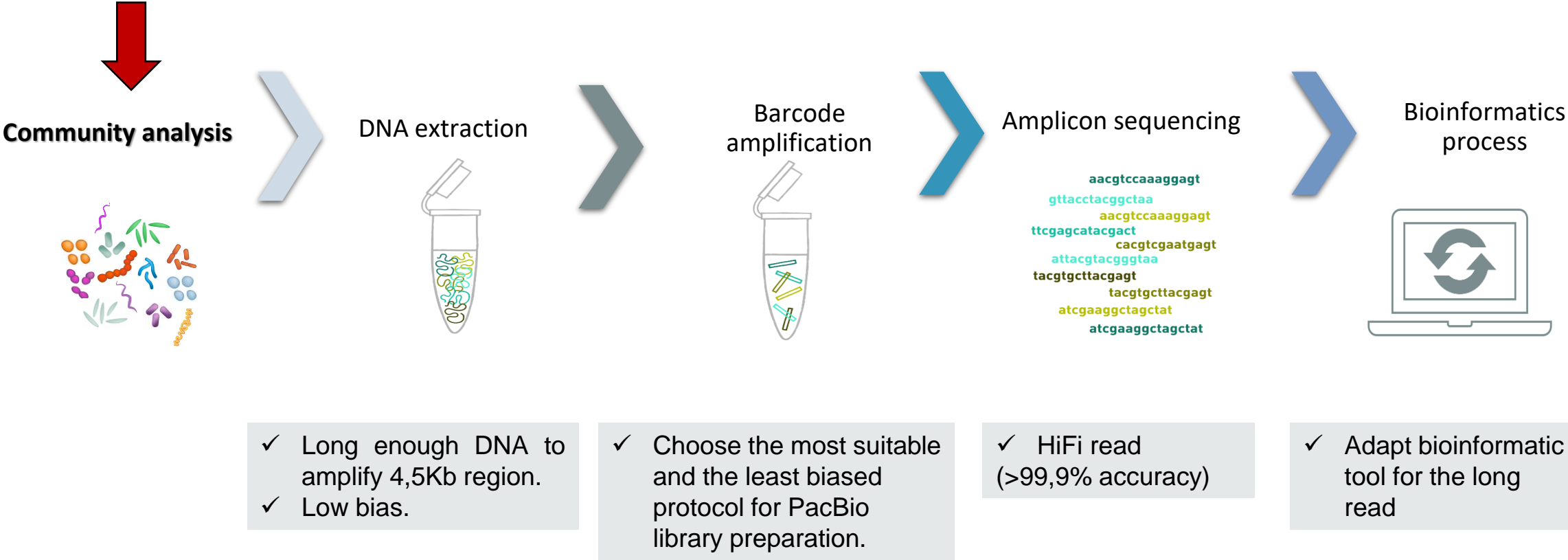✓ Low bias.

✓ Choose the most suitable and the least biased protocol for PacBio library preparation.

✓ HiFi read (>99,9% accuracy)

✓ Adapt bioinformatic tool for the long read

Metabarcoding process for long read

Community analysis → DNA extraction → Barcode amplification → Amplicon sequencing → Bioinformatics process

Amplicon sequencing sequences:
aacgtccaaaggagt
gttacctacggctaa
aacgtccaaaggagt
ttcgagcatacgact
cacgtcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
tacgtgcttacgagt
atcgaaggctagctat
atcgaaggctagctat

DNA extraction
- ✓ Long enough DNA to amplify 4,5Kb region.
- ✓ Low bias.

Barcode amplification
- ✓ Choose the most suitable and the least biased protocol for PacBio library preparation.

Amplicon sequencing
- ✓ HiFi read (>99,9% accuracy)

Bioinformatics process
- ✓ Adapt bioinformatic tool for the long read

Simple community

- **Bacterial mock** *(ZymoBIOMICS)*

8 bacterias (3 Gram- and 5 Gram+)
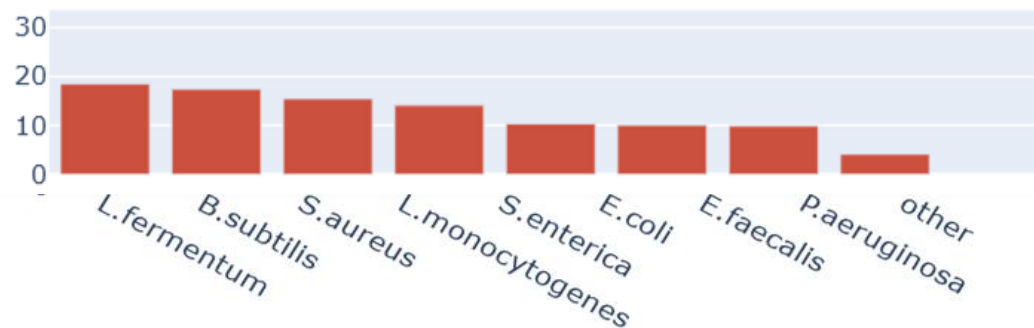
- **DNA mock** *(ZymoBIOMICS):* artificial mix of DNA from individualy extracted strain

Complex community

- **Pig faeces** (hard to extract, inhibitors)

32 faeces samples from 16 individuals (GenPhySE) exposed or not to mycotoxine Fumonisin B1
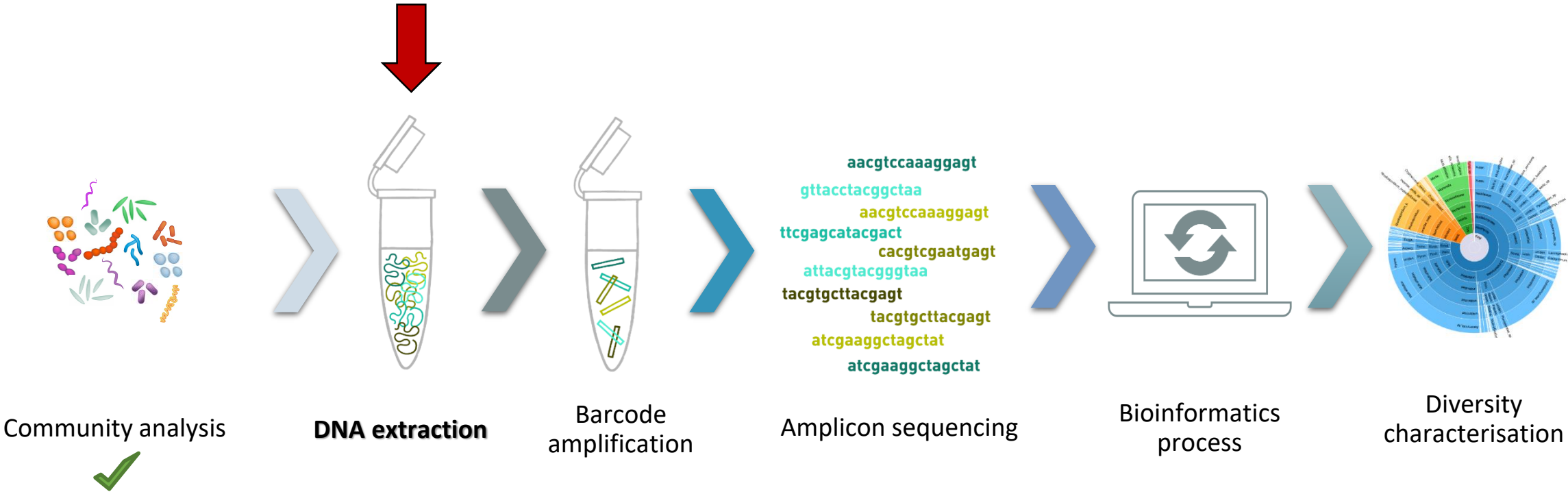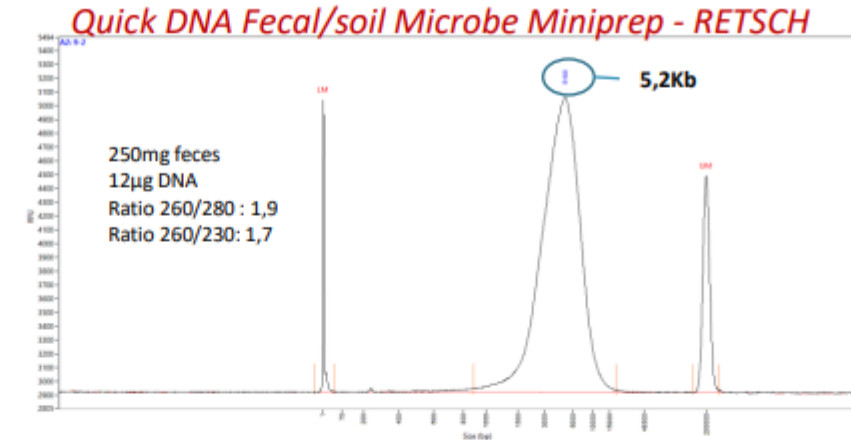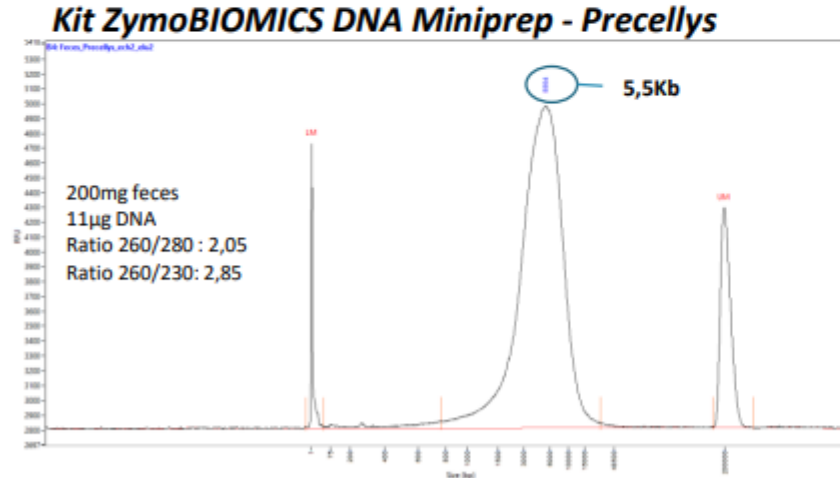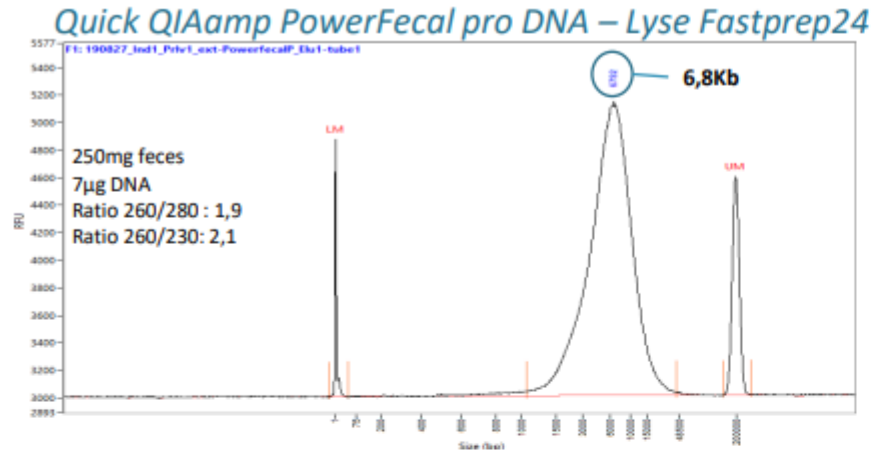
- Short read (V3-V4) data available

toxins   MDPI

Article

Fumonisin-Exposure Impairs Age-Related Ecological Succession of Bacterial Species in Weaned Pig Gut Microbiota

Ivan Mateos [1,2], Sylvie Combes [1,*], Géraldine Pascal [1], Laurent Cauquil [1], Céline Barilly [1], Anne-Marie Cossalter [3], Joëlle Laffitte [3], Sara Botti [1], Philippe Pinton [3] and Isabelle P. Oswald [3,*]

Metabarcoding process: DNA extraction

Community analysis   DNA extraction   Barcode amplification   Amplicon sequencing   Bioinformatics process   Diversity characterisation

DNA extraction kit evaluation

3 kits (among 5 tested) gave good results in term of size & purity

✔ **Qiagen QIAmp PowerFecal Pro DNA**
✔ **Zimobiomics DNA miniprep**
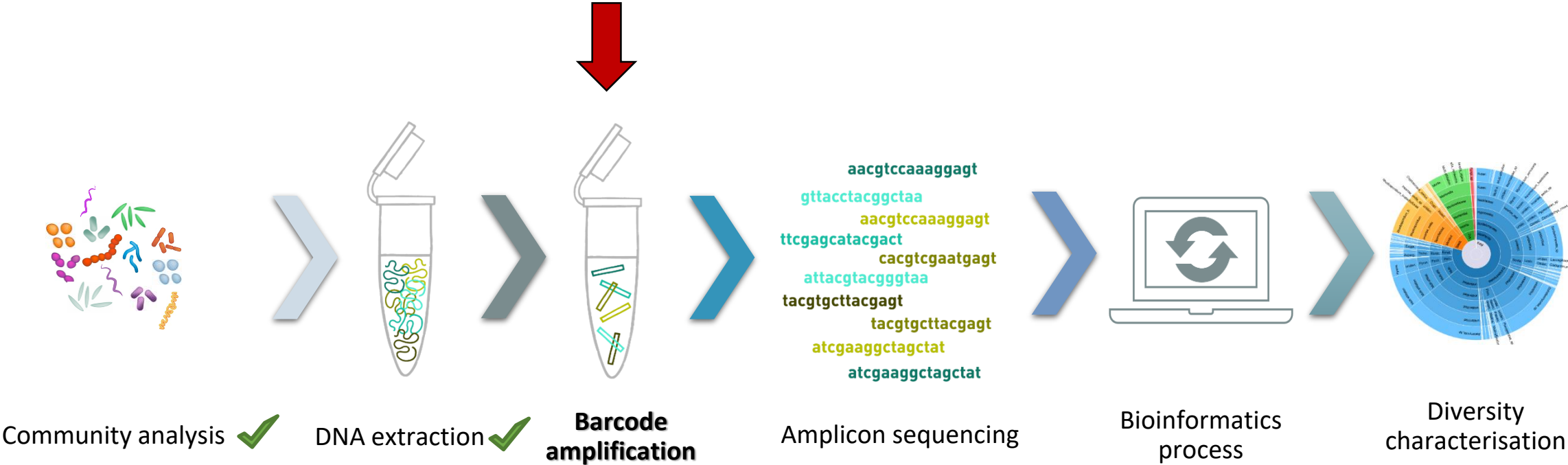✔ **Zimobiomics Quick-DNA Fecal/Soil (reference kit)**

➡ **Efficacity evaluation by sequencing on Illumina MiSeq (reference method)**

DNA extraction: validation on short read (V3-V4 region)
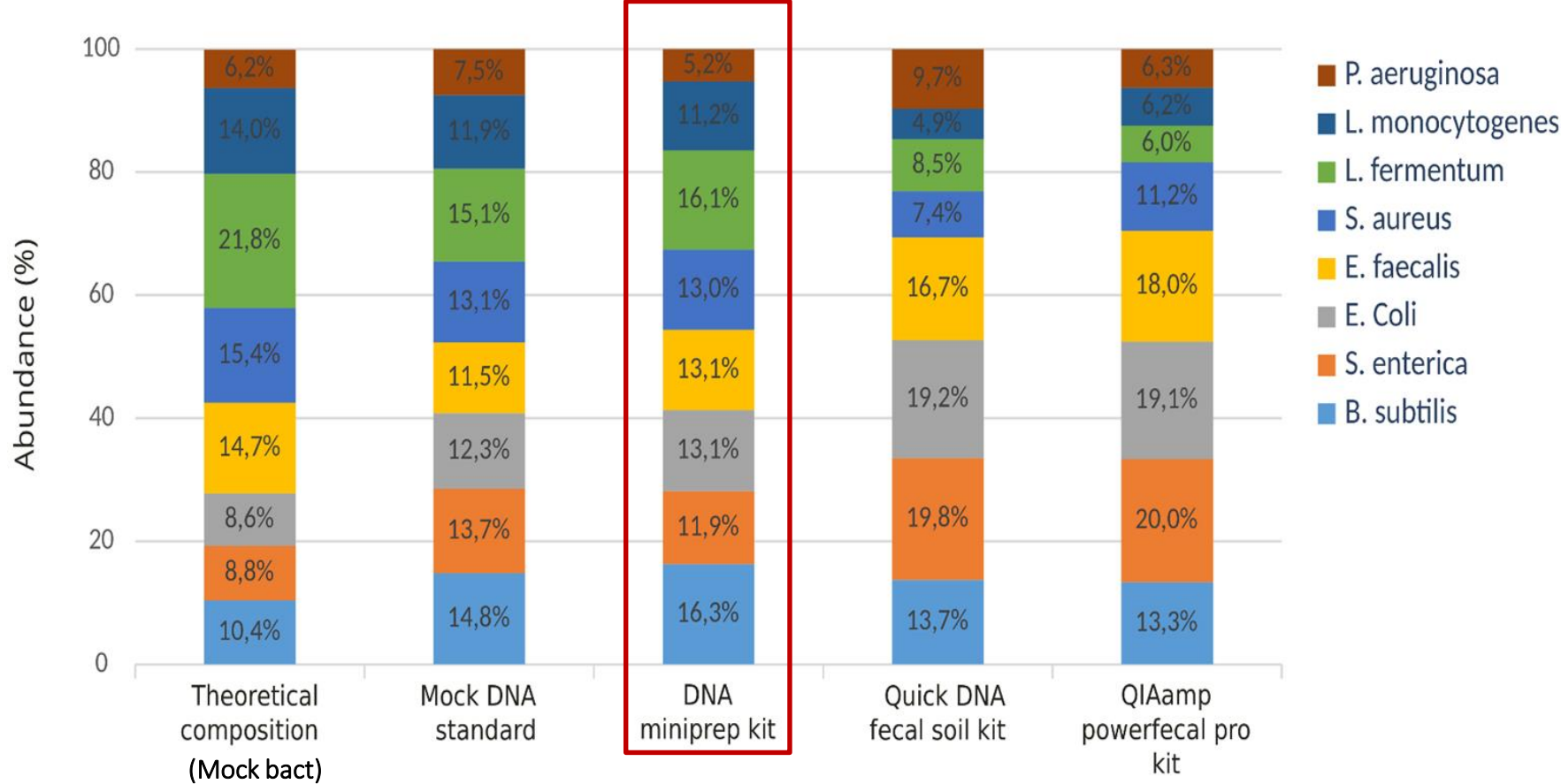
➢ **Slight biases when sequencing mock DNA**

➢ **Zymo DNA miniprep kit gave the best results, choosen for metabarcoding**

Metabarcoding process: barcode amplification
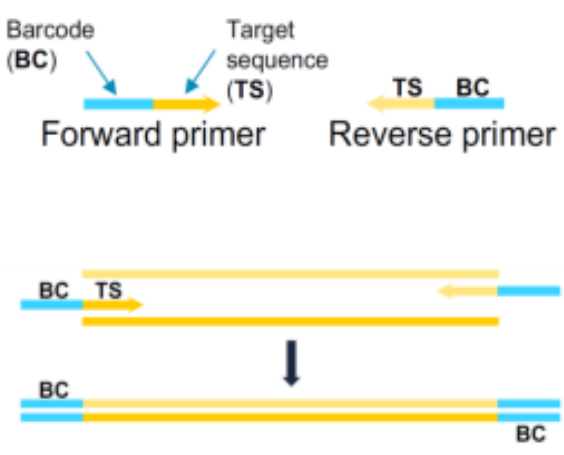
> Protocols have been tested with 16S and 16S-23S barcodes on mock and pig faeces samples
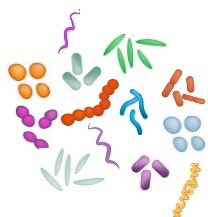
Metabarcoding process: amplicon sequencing

Community analysis ✓   DNA extraction ✓   Barcode amplification ✓   **Amplicon sequencing**   Bioinformatics process   Diversity characterisation
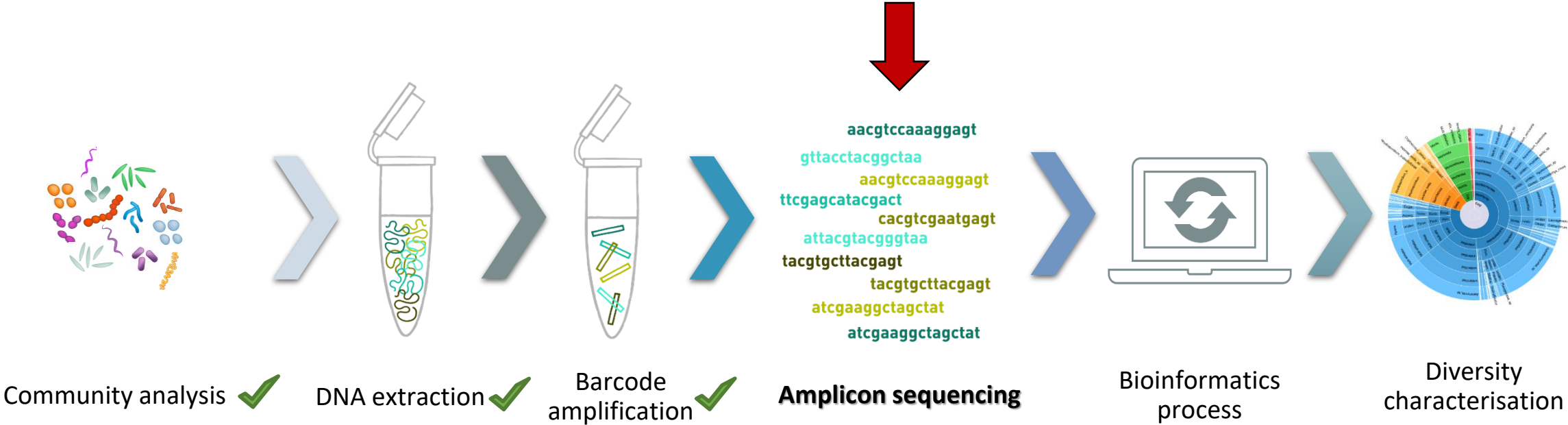
Sequencing results for 16S and 16S-23S barcodes

Protocol	Subreads N50	HiFi Reads	Gb CCS	Quality
16S				
BUP	1,602	2 813 935	4,4	Q40
BTSP	1,569	3 139 687	4,8	Q40
BOA	1,568	2 834 523	4,4	Q42
16S-23S				
BUP	4,2	2 962 319	11,7	Q46
BTSP	4,204	3 173 610	12,7	Q45
BOA	4,325	2 534 890	10,1	Q43

➢ **Sequencing Ok!**
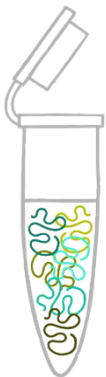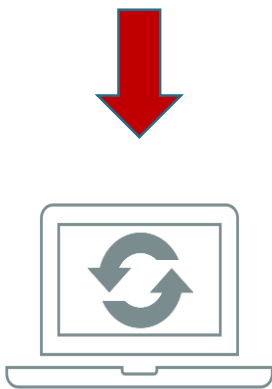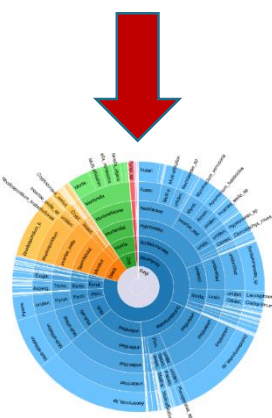
Community analysis ✓ DNA extraction ✓ Barcode amplification ✓ Amplicon sequencing ✓ **Bioinformatics process** **Diversity characterisation**

Bioinformatics process for metabarcoding
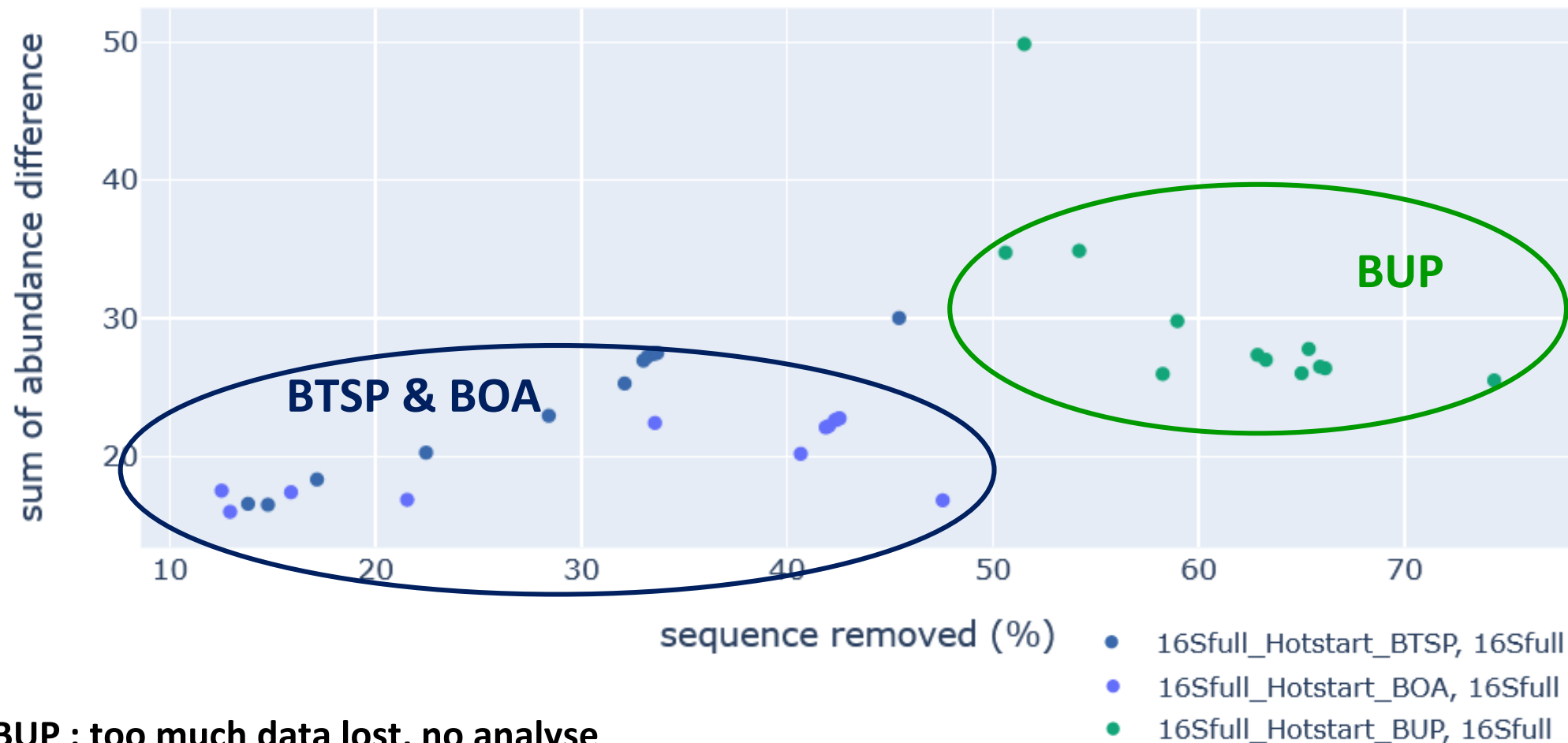

Sequencing

Demultiplexing
Each index is assignes to its original sample

Denoising and clustering
Sequences are groupes by similarity. Amplification and sequencing errores are masked.

Taxonomic affiliation
Each group is affiliated with a taxon through a reference database.

Abundance table
Detect and count the taxa present in each sample.

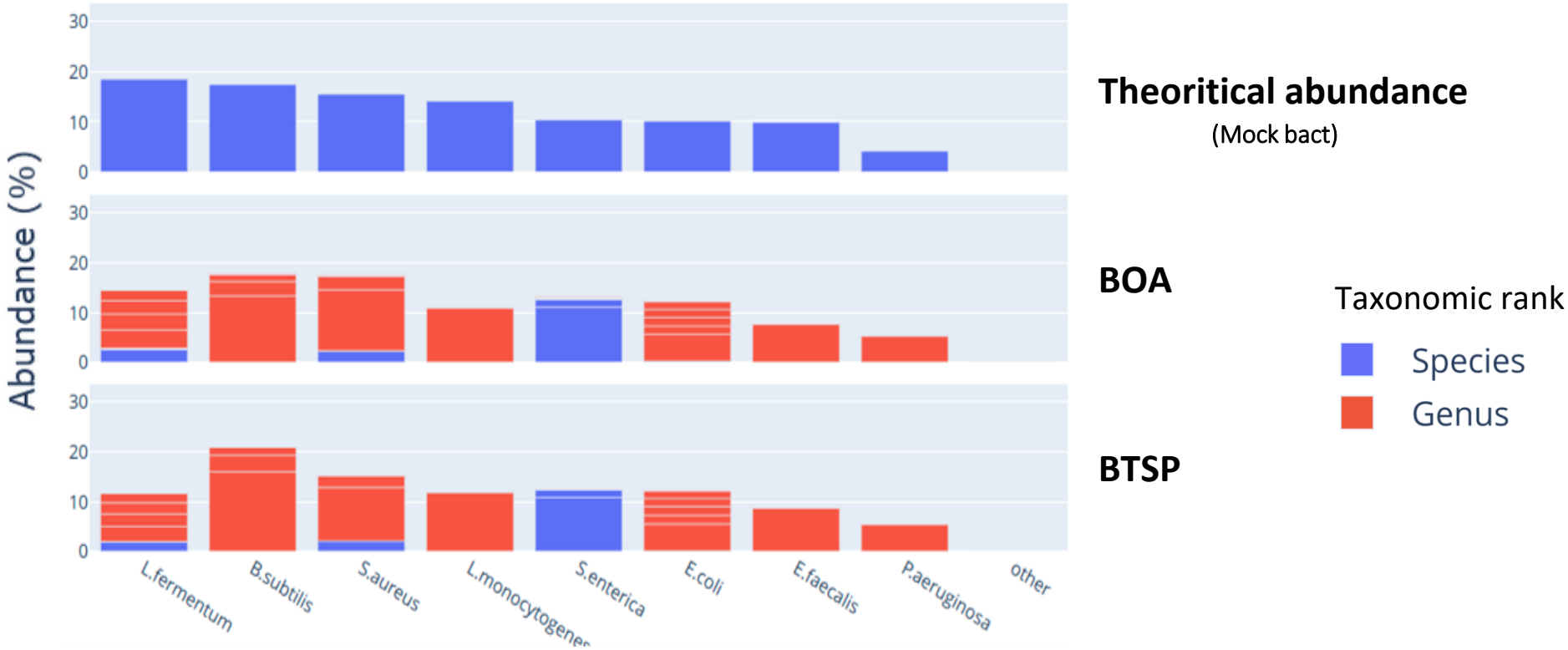*OTU : Operational Taxonomic Unit*

25

% of sequences lost during analysis

- ➤ **BUP : too much data lost, no analyse**

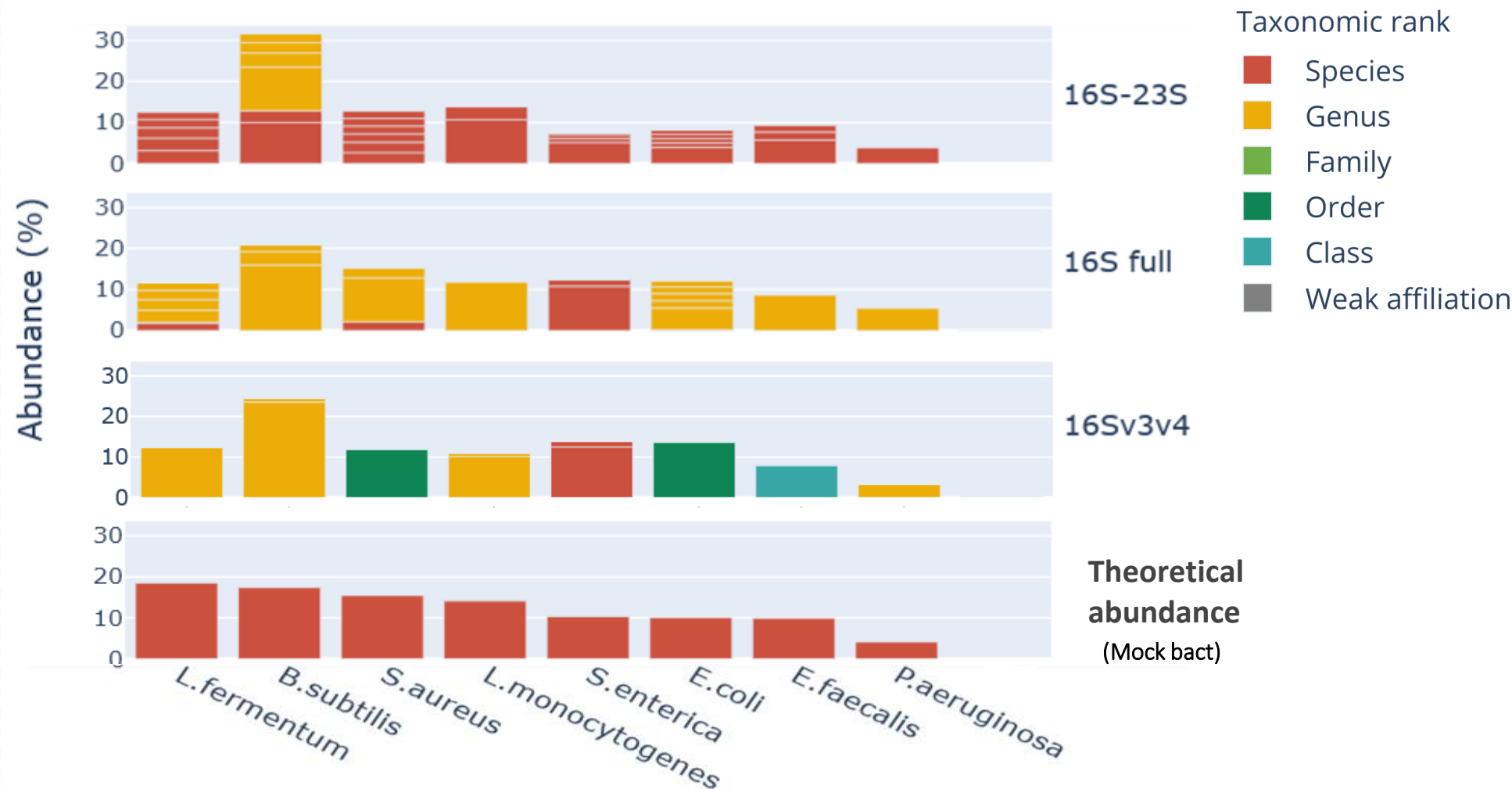- ➤ **BTSP & BOA protocols: same behaviour whatever the software used**

Mock DNA 16S rRNA sequencing on PacBio Sequel II, HiFi

➢ **Similar results obtained for BTSP & BOA sequencing**

> A longer barcode allows to describe more precisely a **simple community**.

31
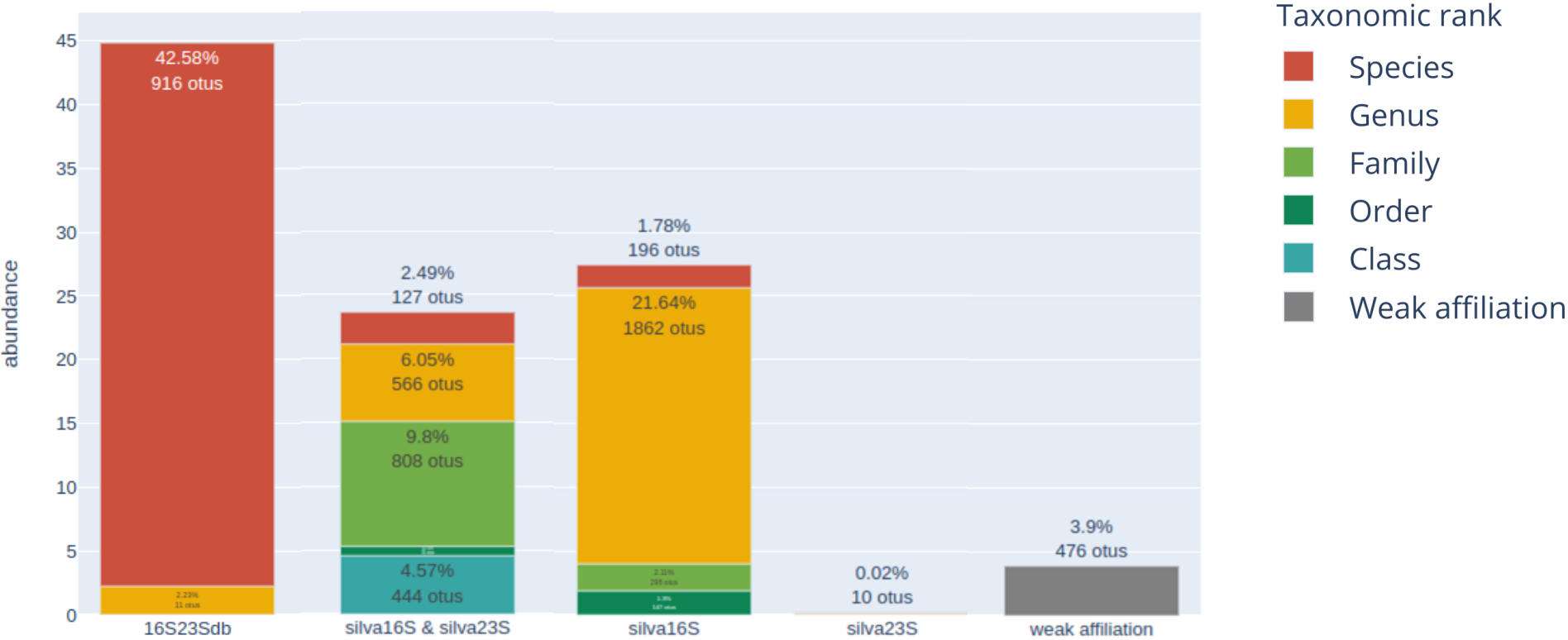
➢ **A longer barcode allows to describe more precisely a complex community.**

Conclusions

- Implemented an **optimized process to performed long read metabarcoding** on 16S and 16S-23S with PacBio HiFi technology :

Community analysis	>	DNA extraction	>	Barcode amplification	>	Amplicon sequencing	>	Bioinformatics process

Amplicon sequencing sequences:
```
aacgtccaaaggagt
gttacctacggctaa
       aacgtccaaaggagt
ttcgagcatacgact
           cacgtcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
           tacgtgcttacgagt
atcgaaggctagctat
       atcgaaggctagctat
```

| Pig faeces	Zymo DNA miniprep kit	BOA and BTSP protocols on 16S and 16S-23S barcodes	PacBio HiFi technology	FROGS Adapted to long read

- **16S-23S region is more informative** than full length 16S

- Problem remains with the database exhaustivity that require time to be completed

Perspectives

- Further development of a taxonomic affiliation method by combining the 16S-23S, 16S and 23S databases.

- Next step is to **deploy** all **SeqoccIn developments** to the **scientific community** => **Search for pilot projects.**

get-plage.contact@genotoul.fr
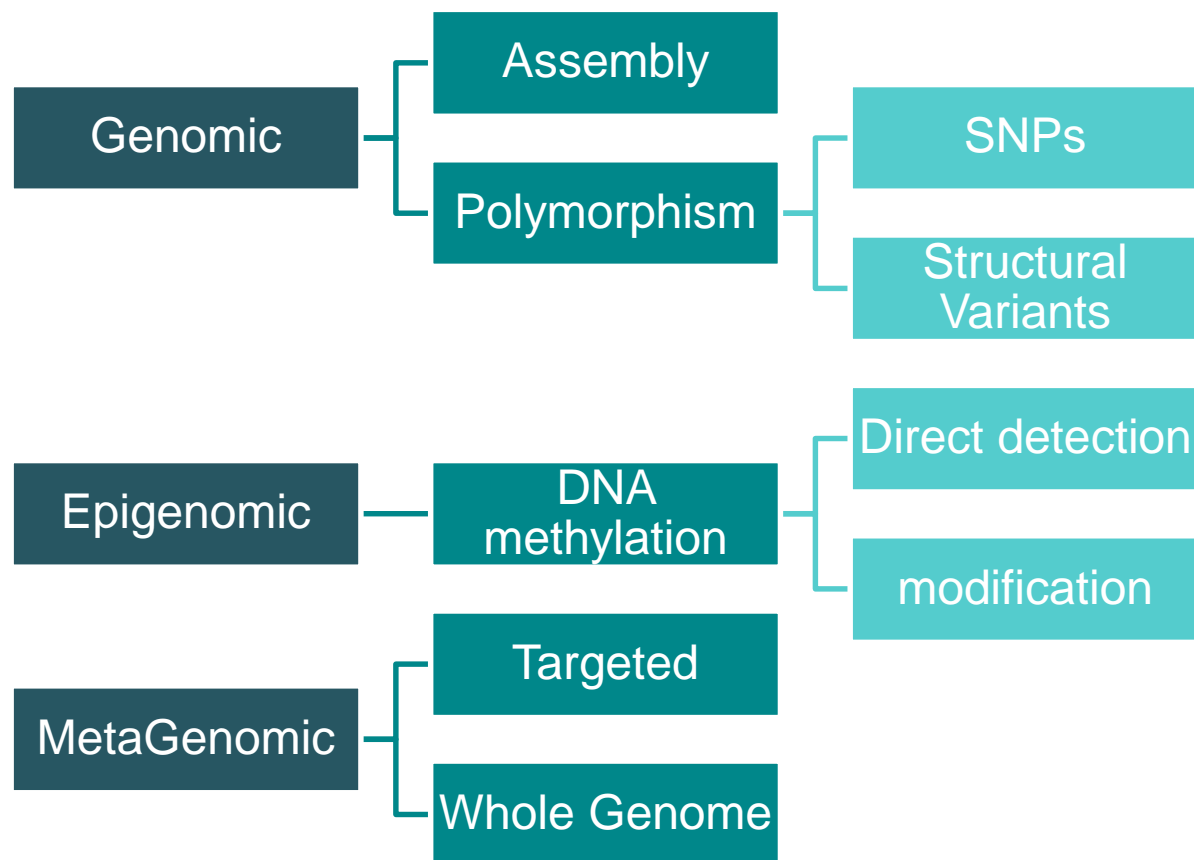
Thanks to

PacBi●

SeqOccIn

Coordination

Cécile Donnadieu
Christine Gaspin
Carole Iampietro
Denis Milan

Jean Mainguy
Olivier Bouchez
Sylvie Combes
Claire Hoede
Géraldine Pascal