



HAL
open science

The benefits of long read HiFi sequencing for metabarcoding analysis

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

► To cite this version:

Adrien Castinel, Jean Mainguy, Sylvie Combes, Carole Iampietro, Christine Gaspin, et al.. The benefits of long read HiFi sequencing for metabarcoding analysis. PACBIO DISCOVERIES ROADSHOW PARIS, PACBIO, May 2022, Paris, France. hal-03740586

HAL Id: hal-03740586

<https://hal.inrae.fr/hal-03740586>

Submitted on 29 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The benefits of long read HiFi sequencing for metabarcoding analysis

Adrien Castinel

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...

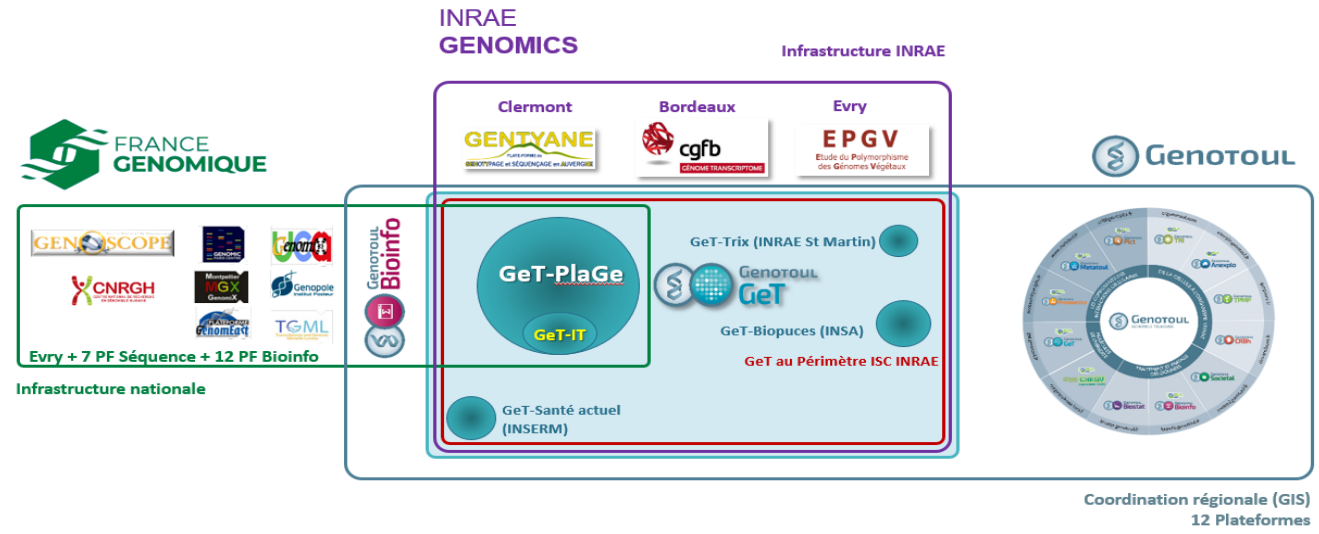


GeT-PlaGe Equipment:

2 Sequel II



PacBio

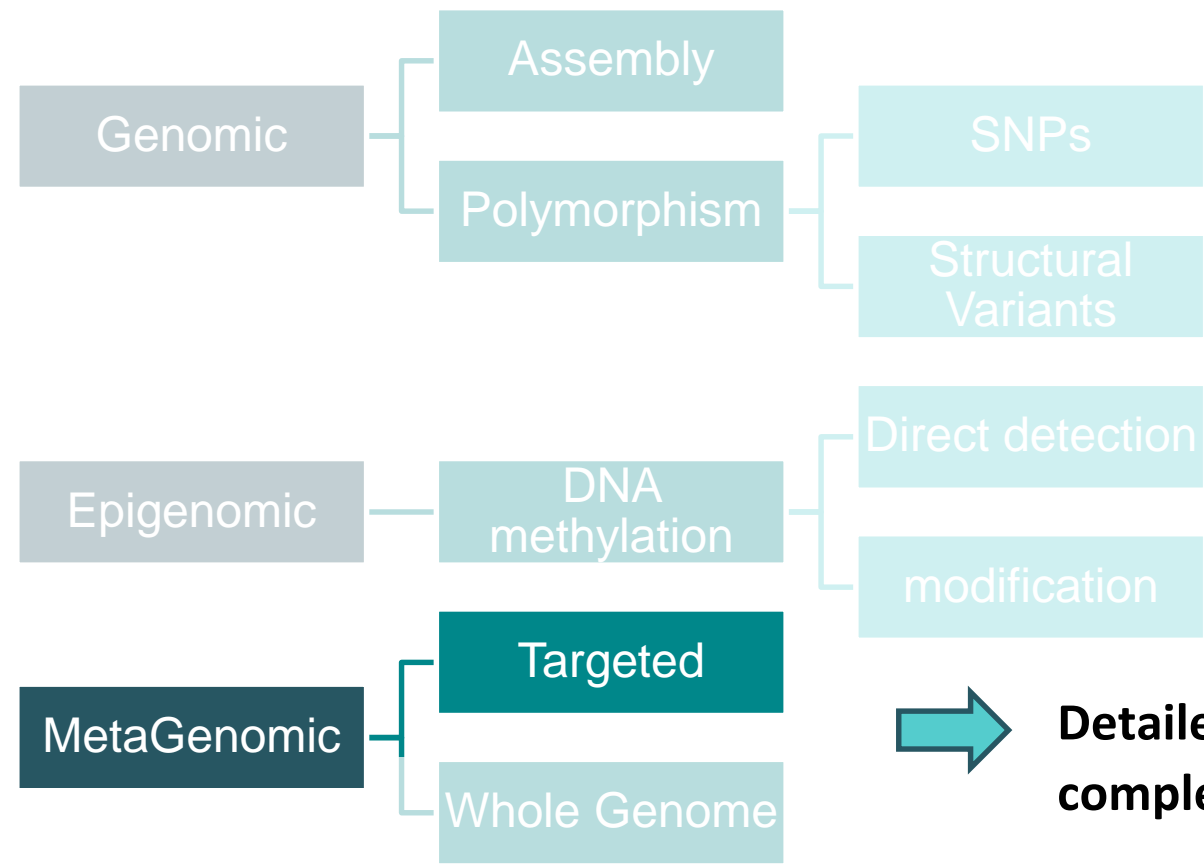


SeqOccln

SeqOccln 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:

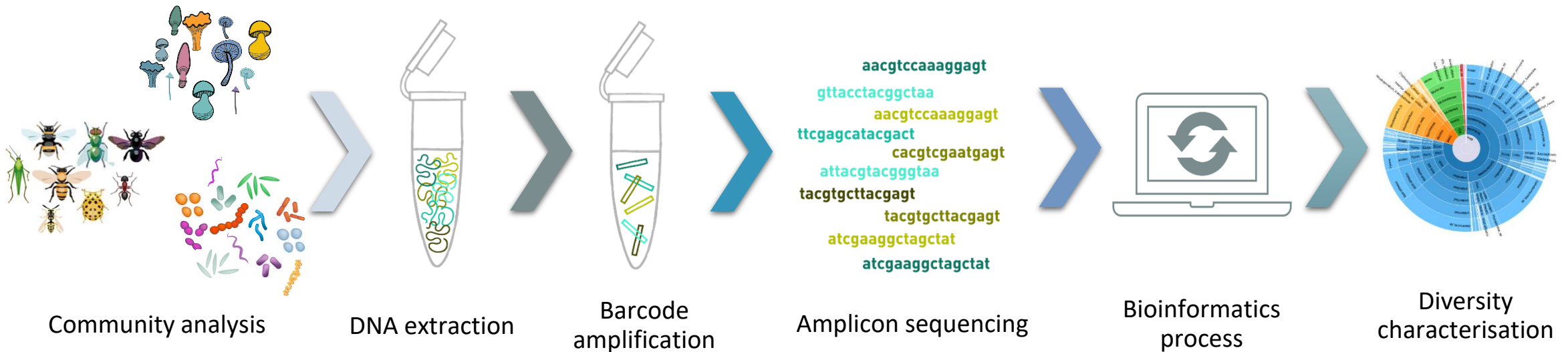


➔ Detailed knowledge of communities in a complex environment

METABARCODING

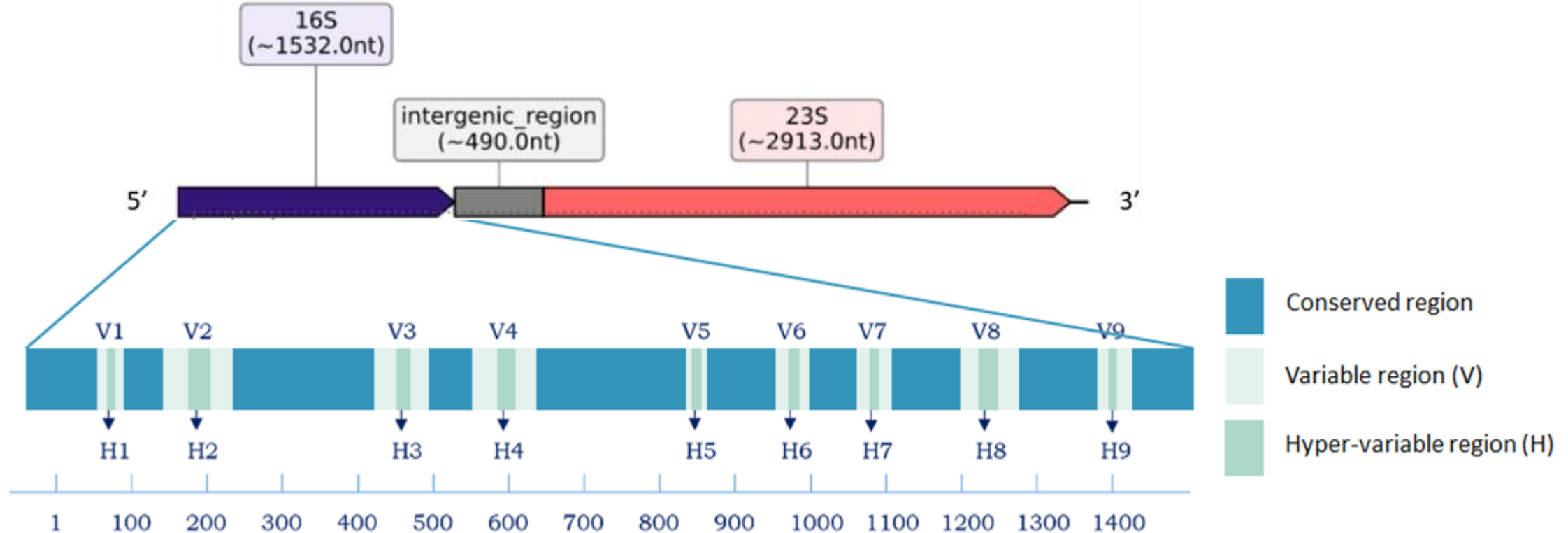
What is metabarcoding?

- 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.



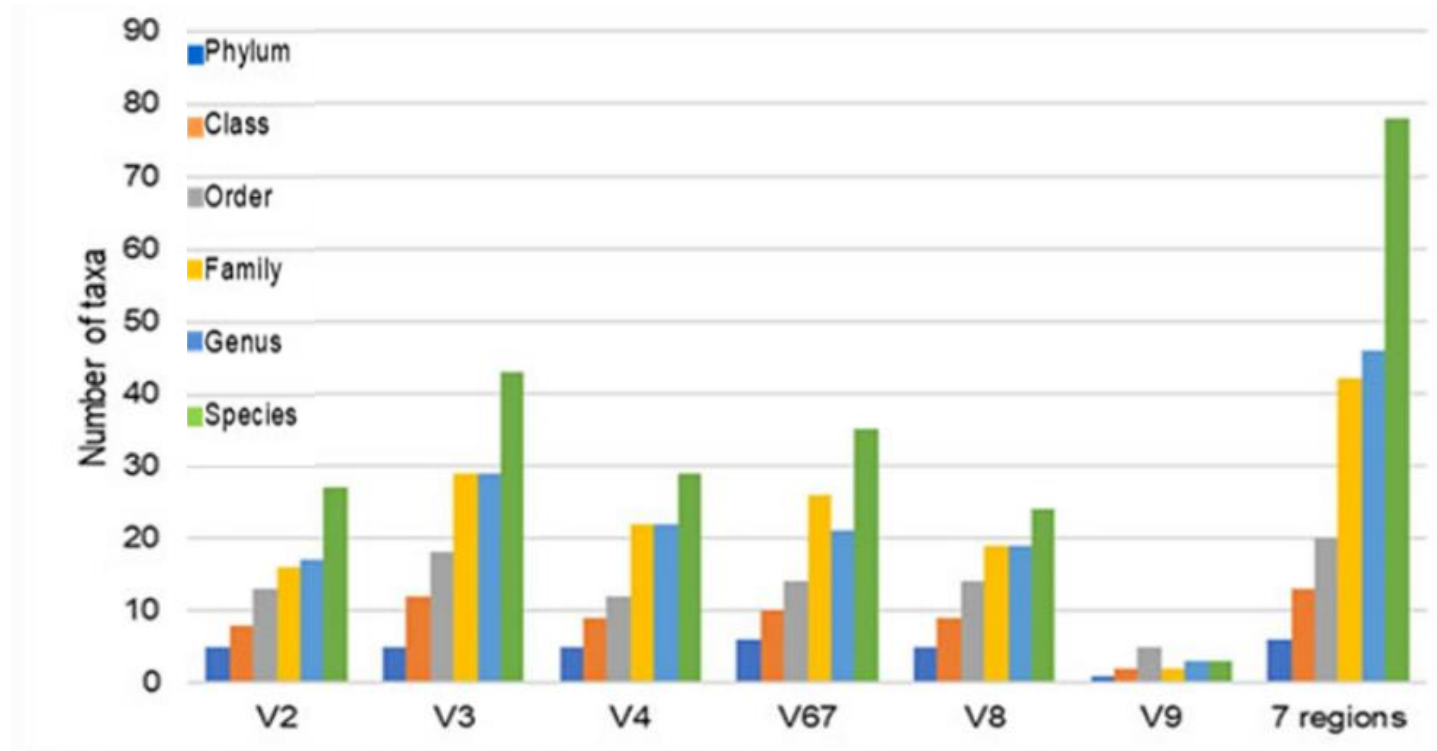
16S: a bacterial marker

Ribosomal RNA: bacterial markers



- 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

Comparison of different hypervariable regions of 16S rRNA



- Analysis of a combination of sequences from V2 to V9 regions identified more taxa.

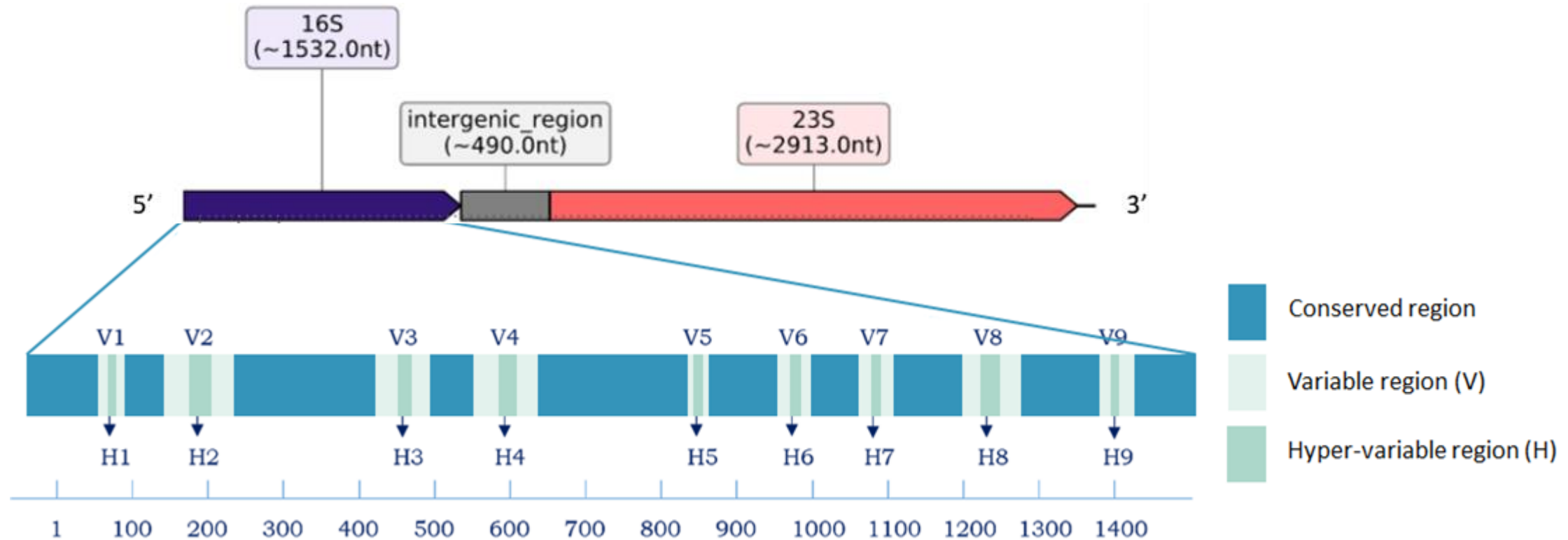
Original Paper | Published: 22 November 2020

Comparison of different hypervariable regions of 16S rRNA for taxonomic profiling of vaginal microbiota using next-generation sequencing

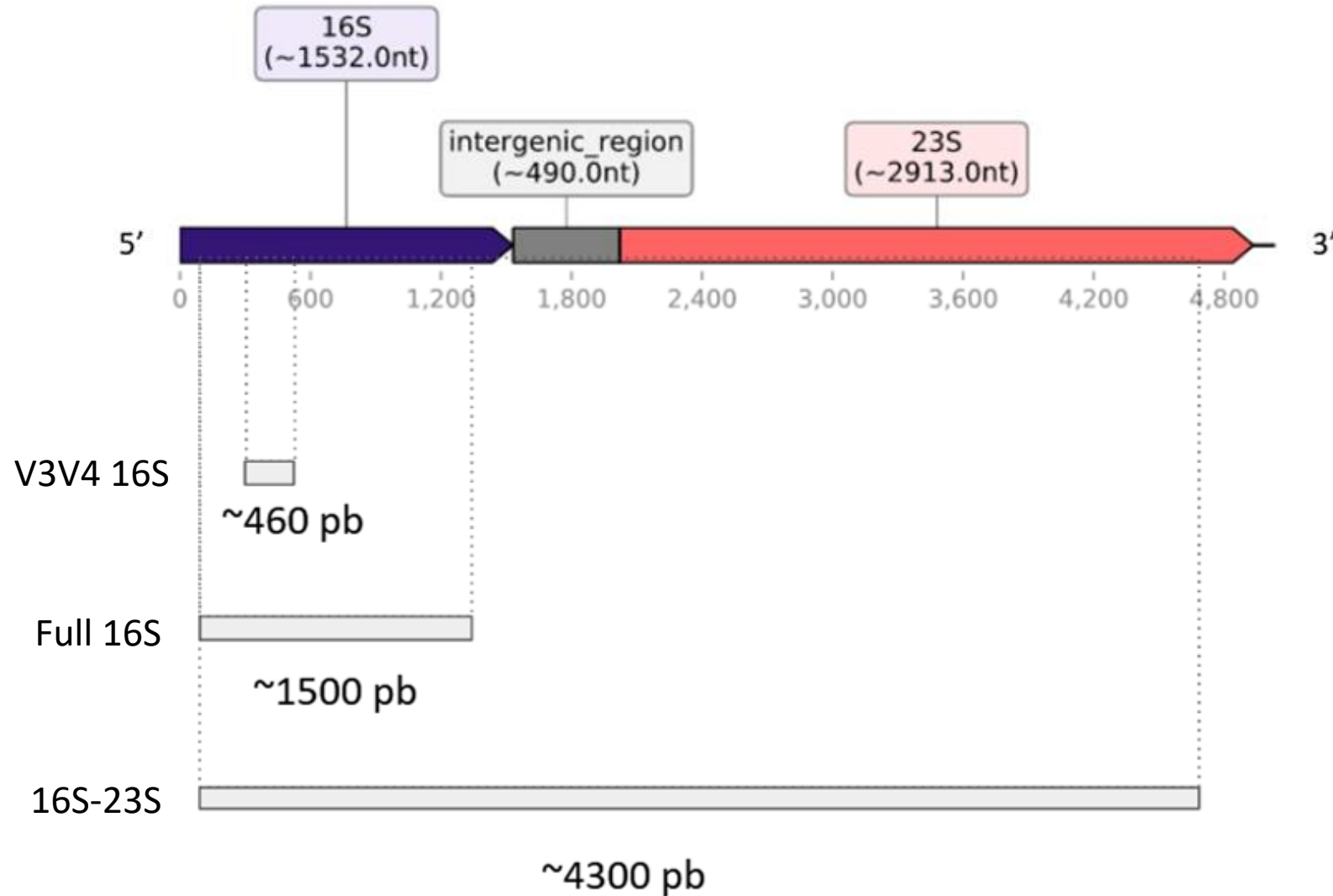
[Auttawit Sirichoat](#), [Nipaporn Sankuntaw](#), [Chulapan Engchanil](#), [Pranom Buppasiri](#), [Kiatchai Faksri](#), [Wises Namwat](#), [Wasun Chantratita](#) & [Viraphong Lulitanond](#)

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

rRNA genes - metabarcoding long read



rRNA genes - metabarcoding long read

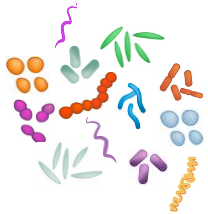


What would be the contribution of "long reads" in metabarcoding?

Focus on 16S / 16S23S markers

Known biases in metabarcoding

Community analysis



- ✓ 16S copy number
- ✓ Horizontal gene transfers

DNA extraction



- ✓ DNA extraction kits/protocols

Barcode amplification



- PCR
- ✓ Polymerase efficiency
 - ✓ Polymerase contaminations
 - ✓ non homogenous amplification

Amplicon sequencing

```

aacgtcaaaggagt
gttacctacggctaa
aacgtcaaaggagt
ttcgagcatagact
cagctcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
tacgtgcttacgagt
atcgaaggctagctat
atcgaaggctagctat
  
```

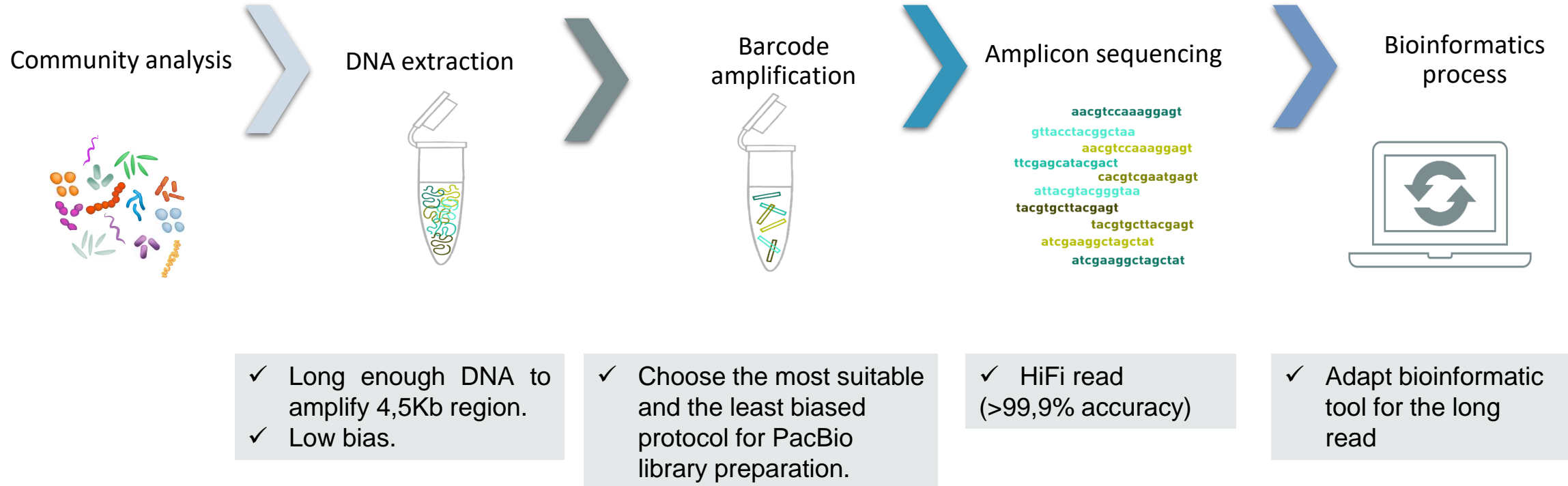
- ✓ Sequencing errors

Bioinformatics process

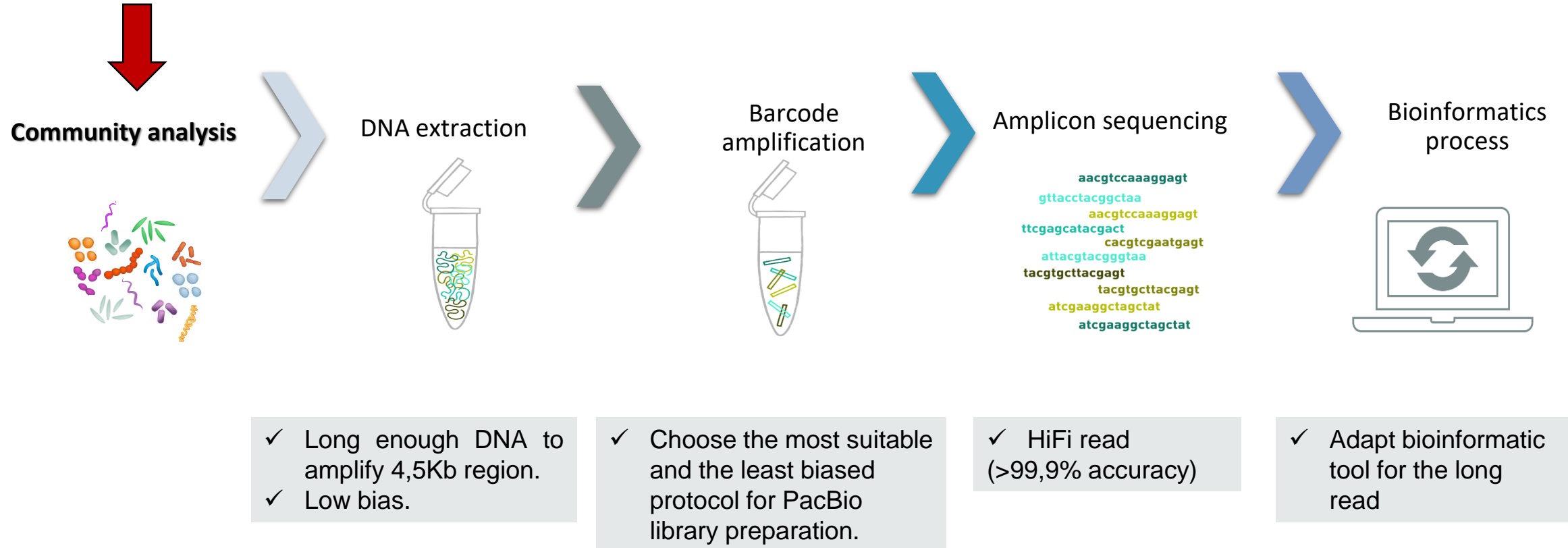


- ✓ Analysis softwares
- ✓ Databases exhaustivity

Metabarcoding process for long read



Metabarcoding process for long read

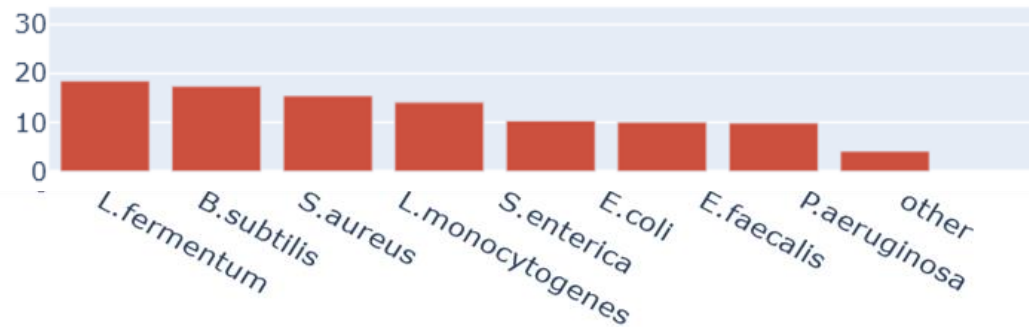


Biological matrix used in the project

Simple community

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

8 bacteria (3 Gram- and 5 Gram+)



- **DNA mock** (*ZymoBIOMICS*): artificial mix of DNA from individually 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

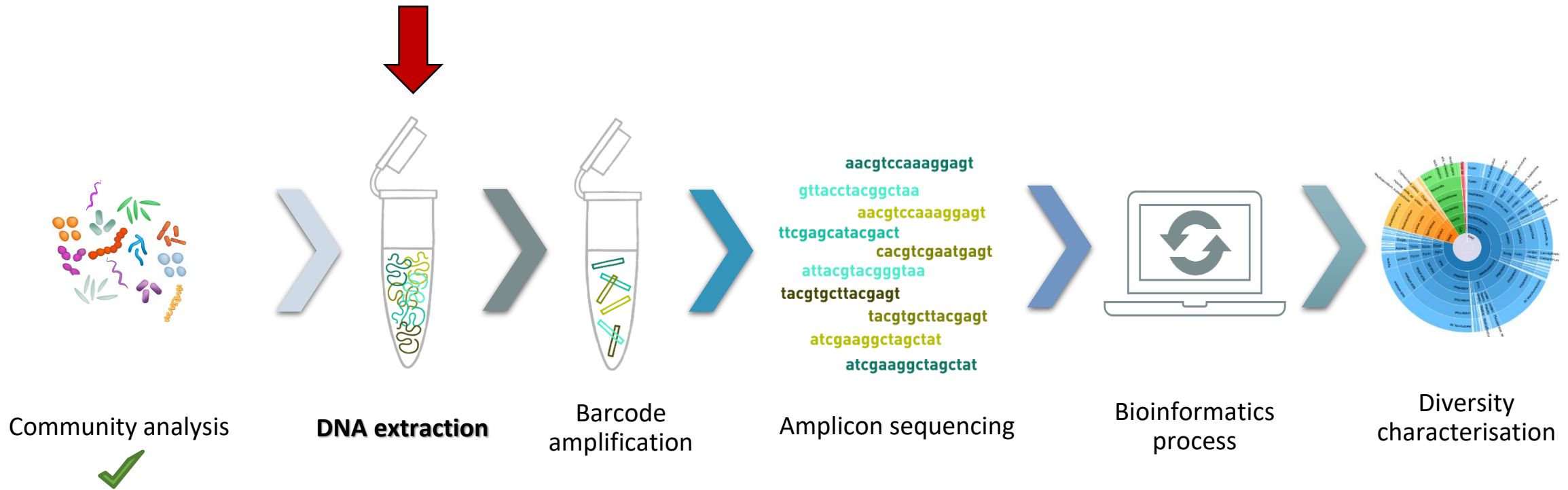


Article

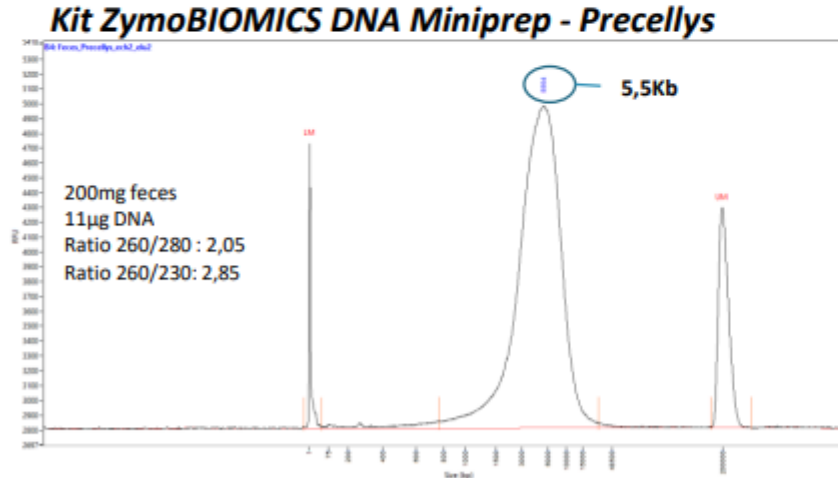
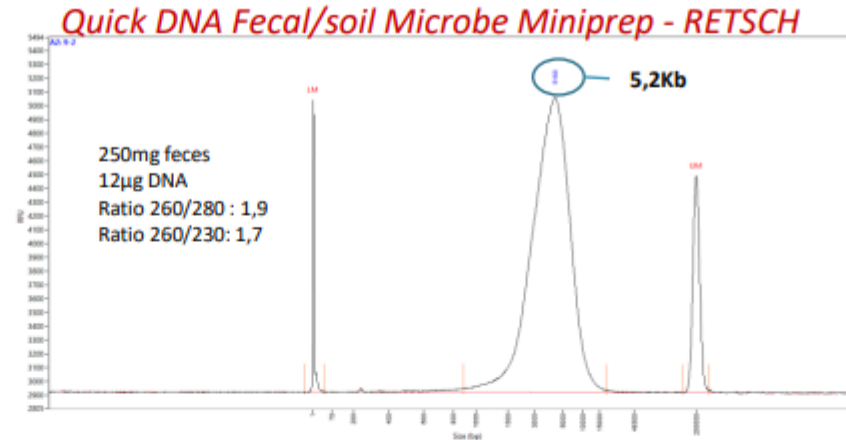
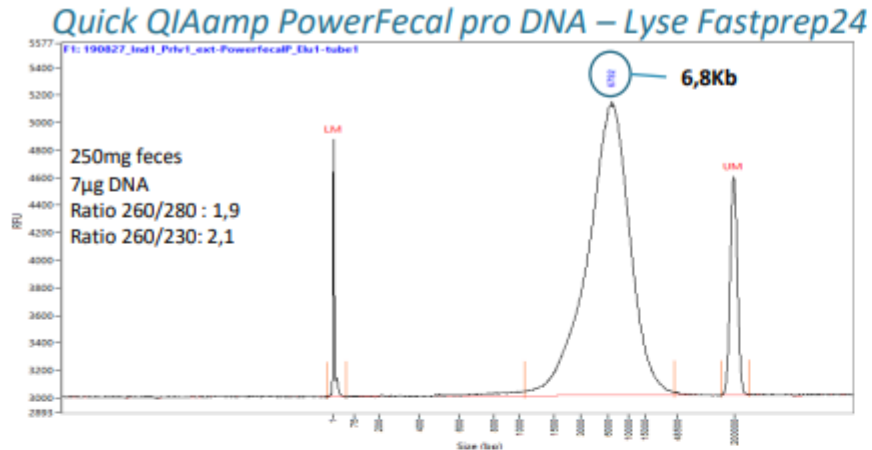
Fumonisin-Exposure Impairs Age-Related Ecological Succession of Bacterial Species in Weaned Pig Gut Microbiota

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 ⁴, Philippe Pinton ³ and Isabelle P. Oswald ^{3,*}

Metabarcoding process: DNA extraction



DNA extraction kit evaluation



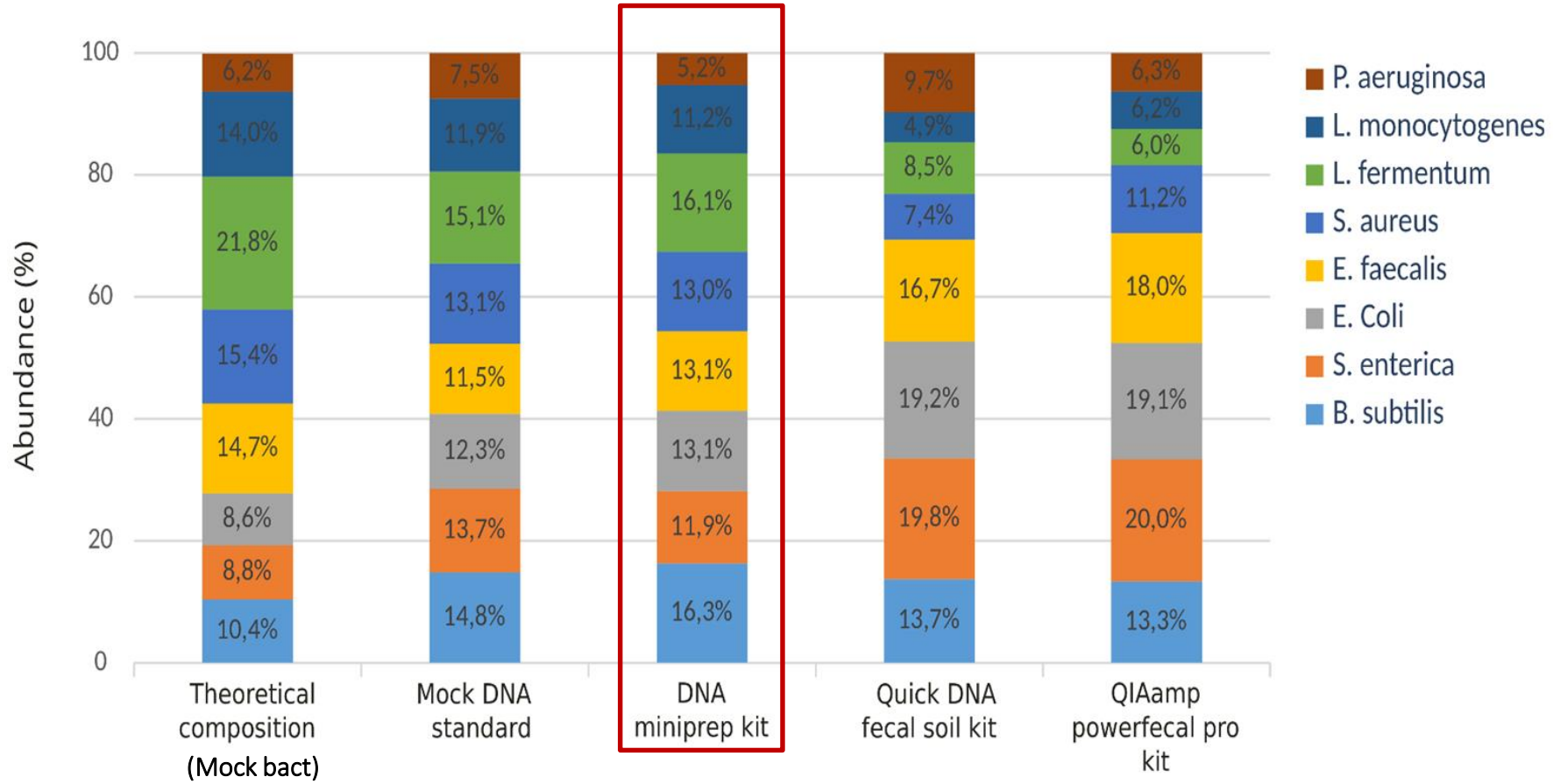
3 kits (among 5 tested) gave good results in term of size & purity

- ✓ Qiagen QIAamp PowerFecal Pro DNA
- ✓ Zimobiomics DNA miniprep
- ✓ Zimobiomics Quick-DNA Fecal/Soil (reference kit)



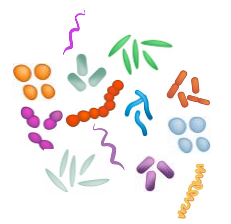
Efficacy evaluation by sequencing on Illumina MiSeq (reference method)

SeqOclin DNA extraction: validation on short read (V3-V4 region)

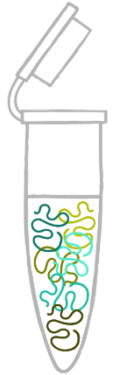


- Slight biases when sequencing mock DNA
- Zymo DNA miniprep kit gave the best results, chosen for metabarcoding

Metabarcoding process: barcode amplification



Community analysis ✓



DNA extraction ✓



Barcode amplification



```

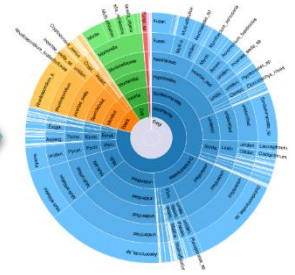
aacgtccaaggagt
gttacctacggctaa
aacgtccaaggagt
ttcgagcatagcact
cacgtcgaatgagt
attacgtacgggtaa
tacgtgcttacgagt
tacgtgcttacgagt
atcgaaggctagctat
atcgaaggctagctat

```

Amplicon sequencing



Bioinformatics process



Diversity characterisation

Comparison of PacBio protocols

Barcoded Universal Primers (BUP)

PCR 1



PCR 2



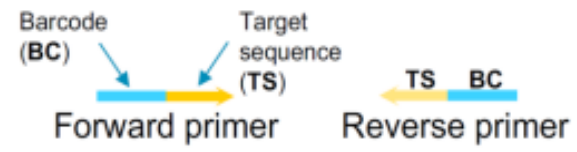
2 PCR

Complexity primer design : ● ● ○ ○

PCR optimisation : ● ● ○ ○

Cost : ● ○ ○ ○

Barcoded Target-Specific Primers (BTSP)



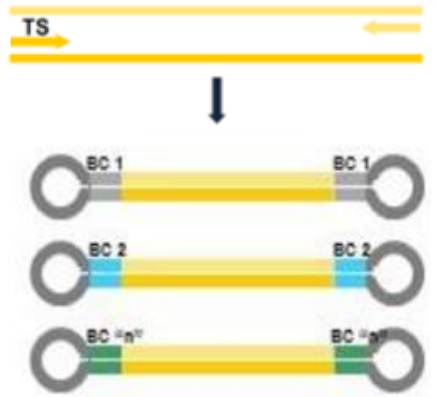
1 PCR

Complexity primer design : ● ● ● ●

PCR optimisation : ● ● ○ ○

Cost : ● ○ ○ ○

Barcoded Overhang adapters (BOA)



1 PCR

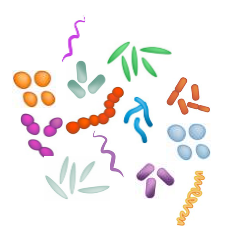
Primer design : ● ○ ○ ○

PCR optimisation : ● ○ ○ ○

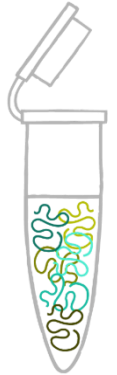
Cost : ● ● ○ ○

➤ 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 ✓

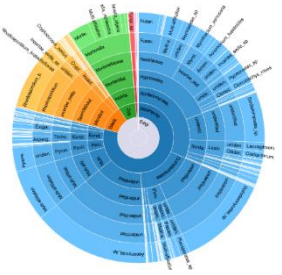


aacgtccaaggagt
 gttacctacggctaa
 aacgtccaaggagt
 ttcgagcatacgact
 cacgtcgaatgagt
 attacgtacgggtaa
 tacgtgcttacgagt
 tacgtgcttacgagt
 atcgaaggctagctat
 atcgaaggctagctat

Amplicon sequencing



Bioinformatics process



Diversity characterisation

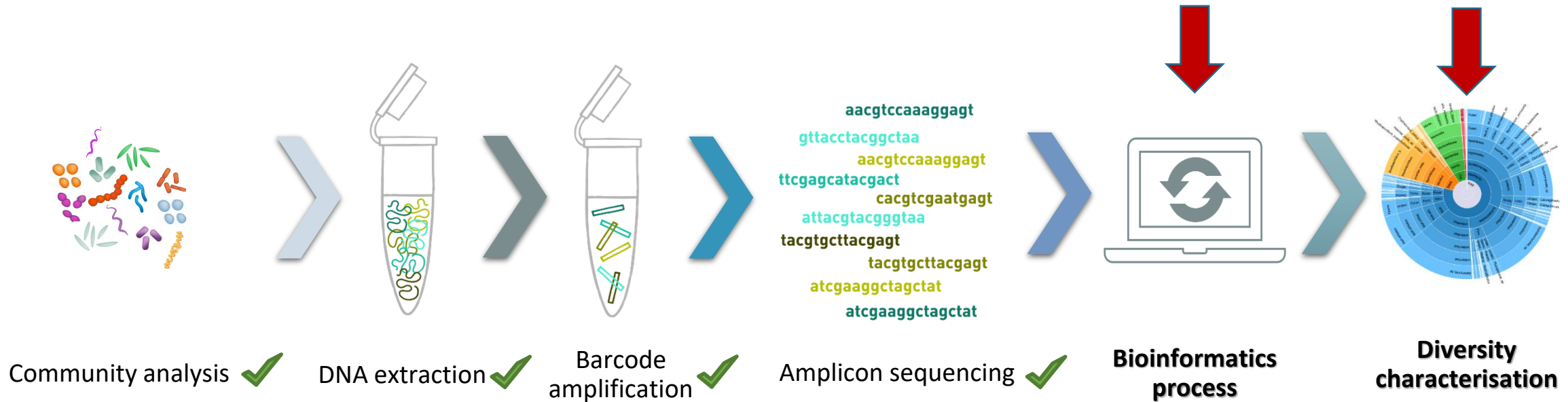
Sequencing results for 16S and 16S-23S barcodes



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!

Metabarcoding process: bioinformatics process

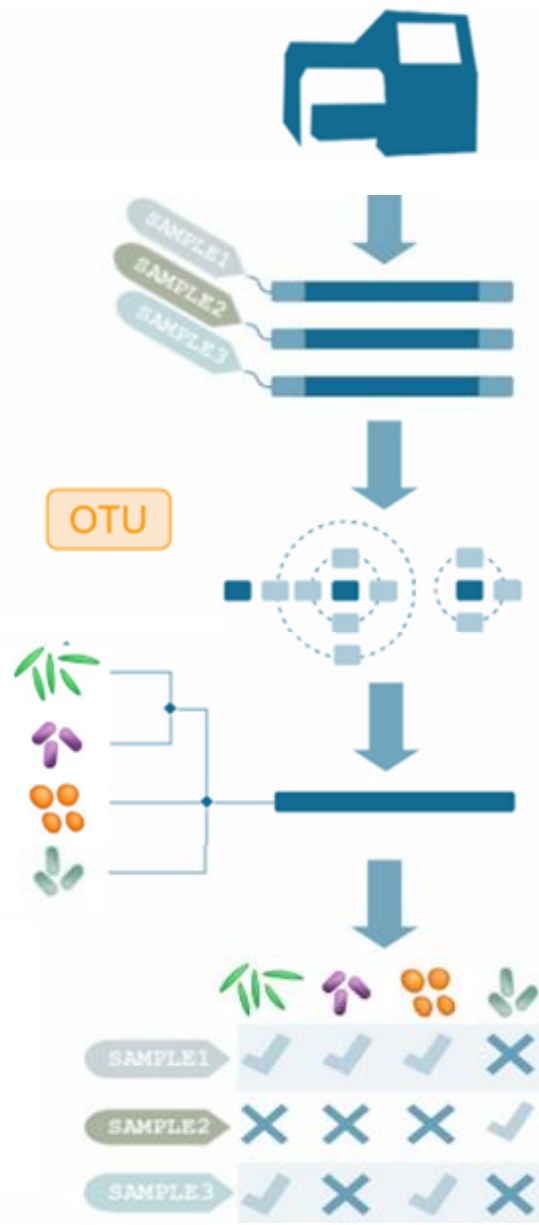


Bioinformatics process for metabarcoding



FROGS: Find, Rapidly, OTUs with Galaxy Solution FREE
 Frédéric Escudié, Lucas Auer, Maria Bernard, Mahendra Mariadassou, Laurent Cauquil, Katia Vidal, Sarah Maman, Guillermina Hernandez-Raquet, Sylvie Combes, Géraldine Pascal ✉ Author Notes
 Bioinformatics, Volume 34, Issue 8, 15 April 2018, Pages 1287–1294, <https://doi.org/10.1093/bioinformatics/btx791>
 Published: 07 December 2017 [Article history](#) ▾

*OTU : Operational Taxonomic Unit



Sequencing

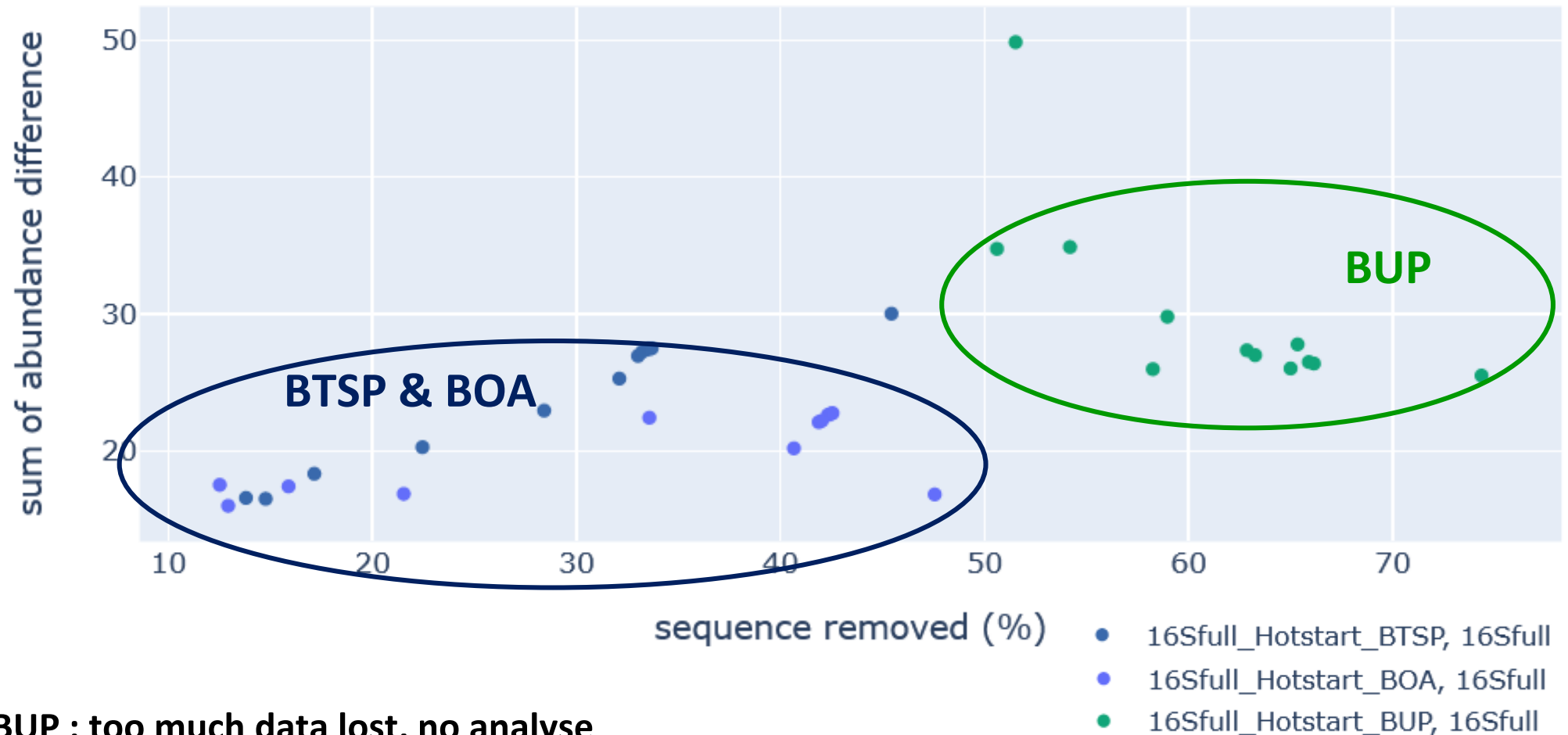
Demultiplexing
Each index is assigned to its original sample

Denoising and clustering
Sequences are grouped by similarity. Amplification and sequencing errors 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.

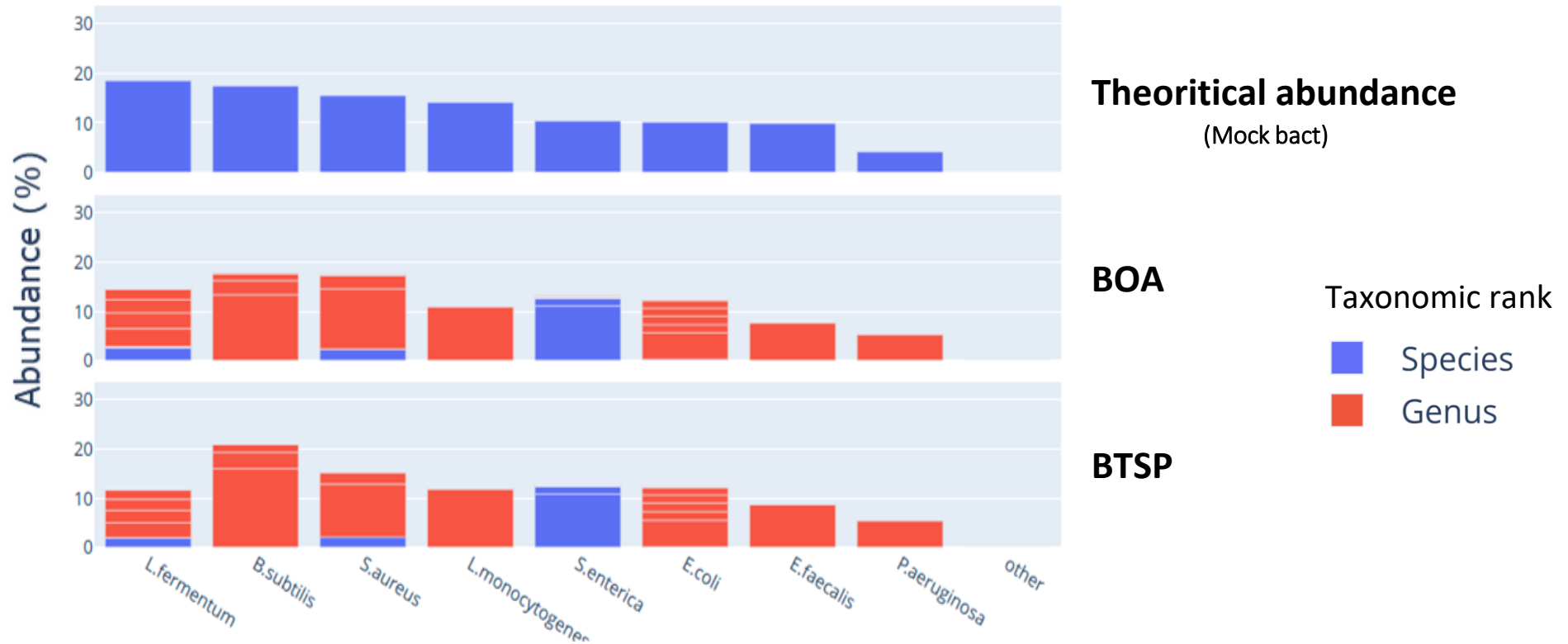
% of sequences lost during analysis



- **BUP** : too much data lost, no analyse
- **BTSP & BOA** protocols: same behaviour whatever the software used

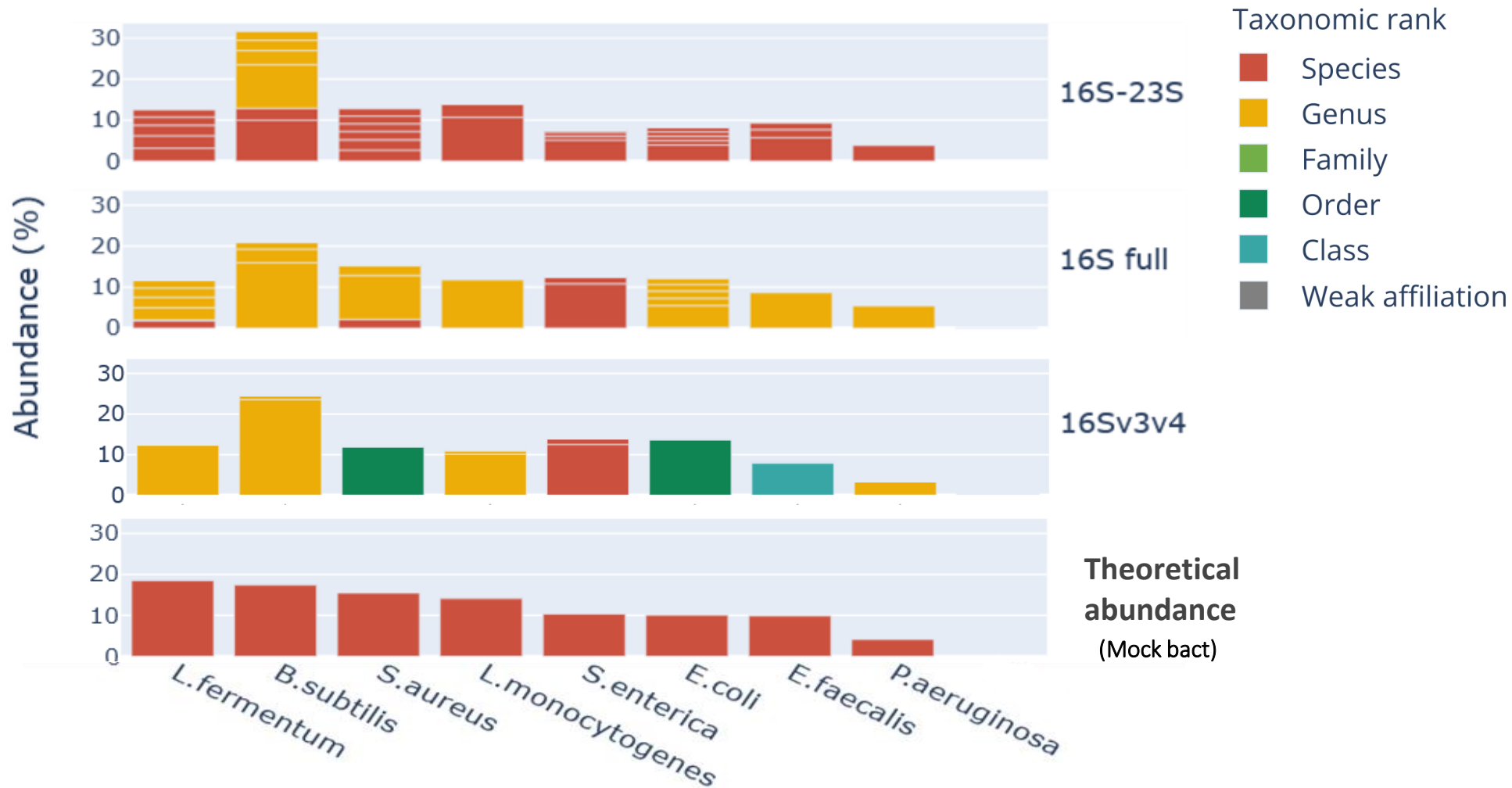
BTSP & BOA comparison: 16S rRNA sequencing

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



➤ Similar results obtained for BTSP & BOA sequencing

Results: affiliation of OTUs from metabarcodings with different targets - mock community



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

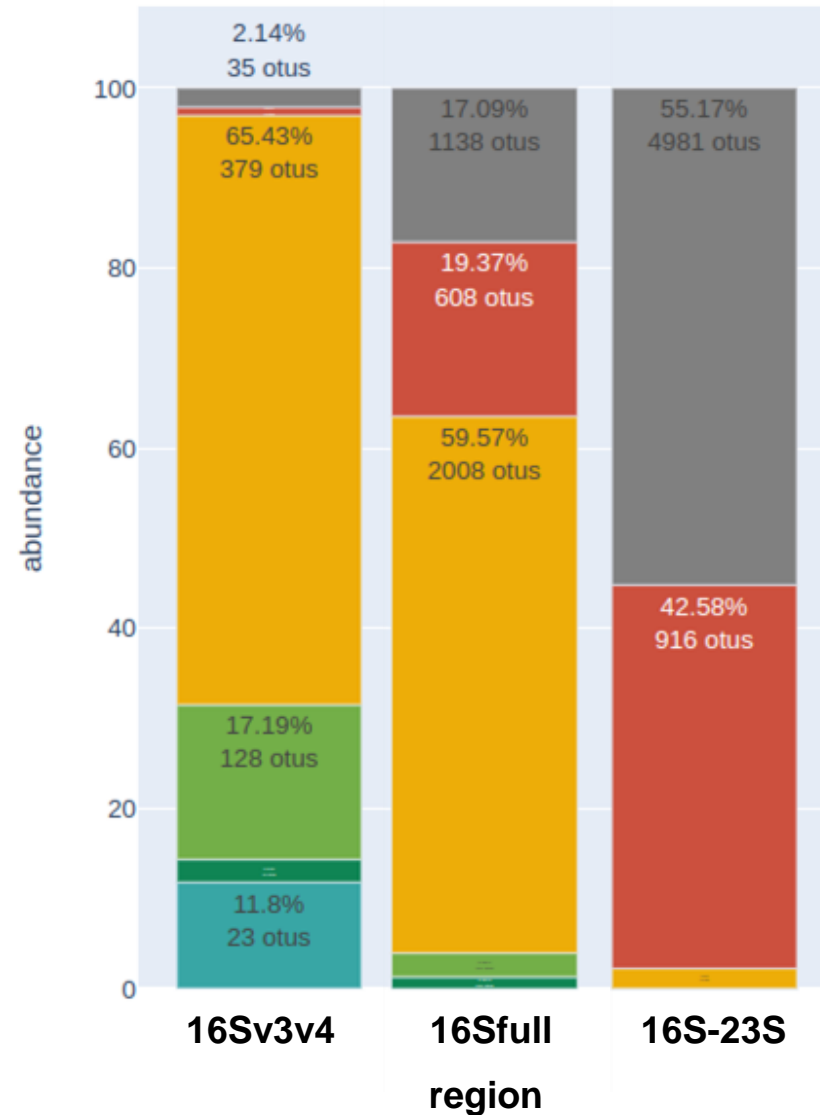
Results: Affiliation of OTUs from metabarcodings with different targets - pig faeces sample



Taxonomic rank

- Species
- Genus
- Family
- Order
- Class
- Weak affiliation

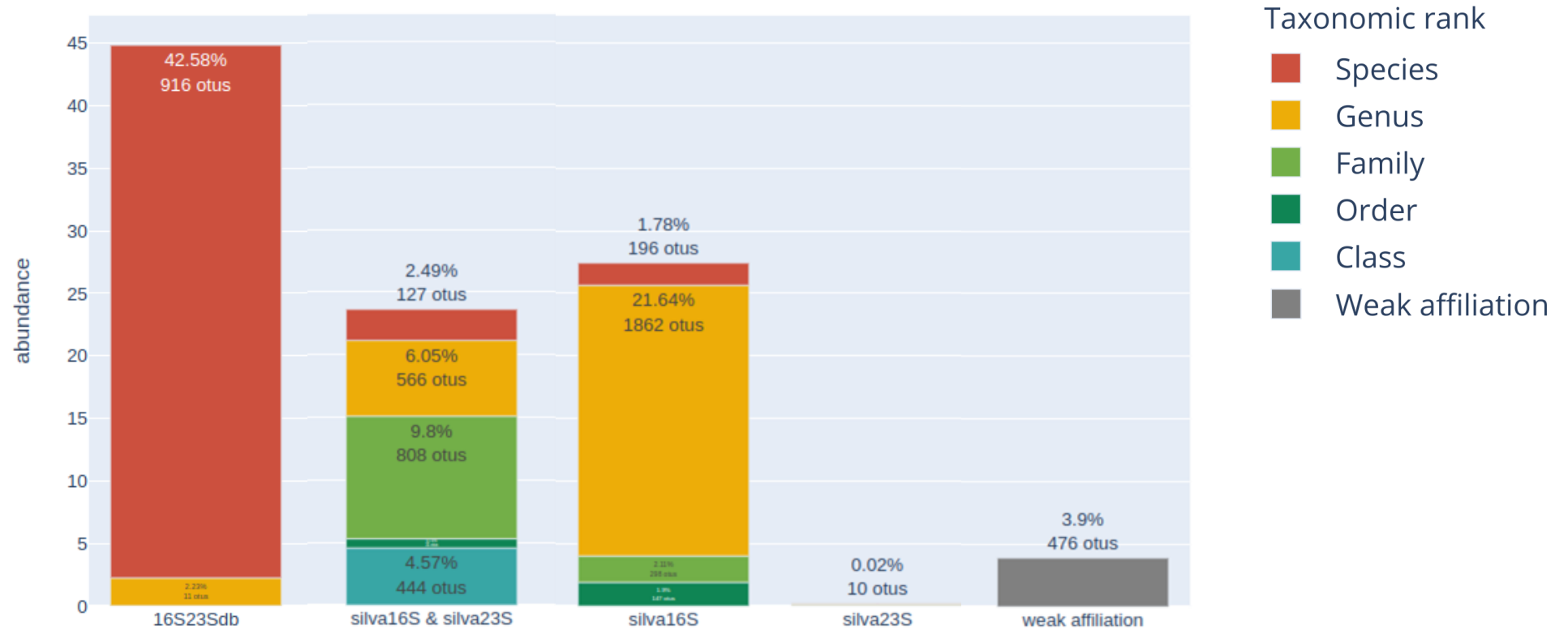
98% identity
99% coverage



vs other databases



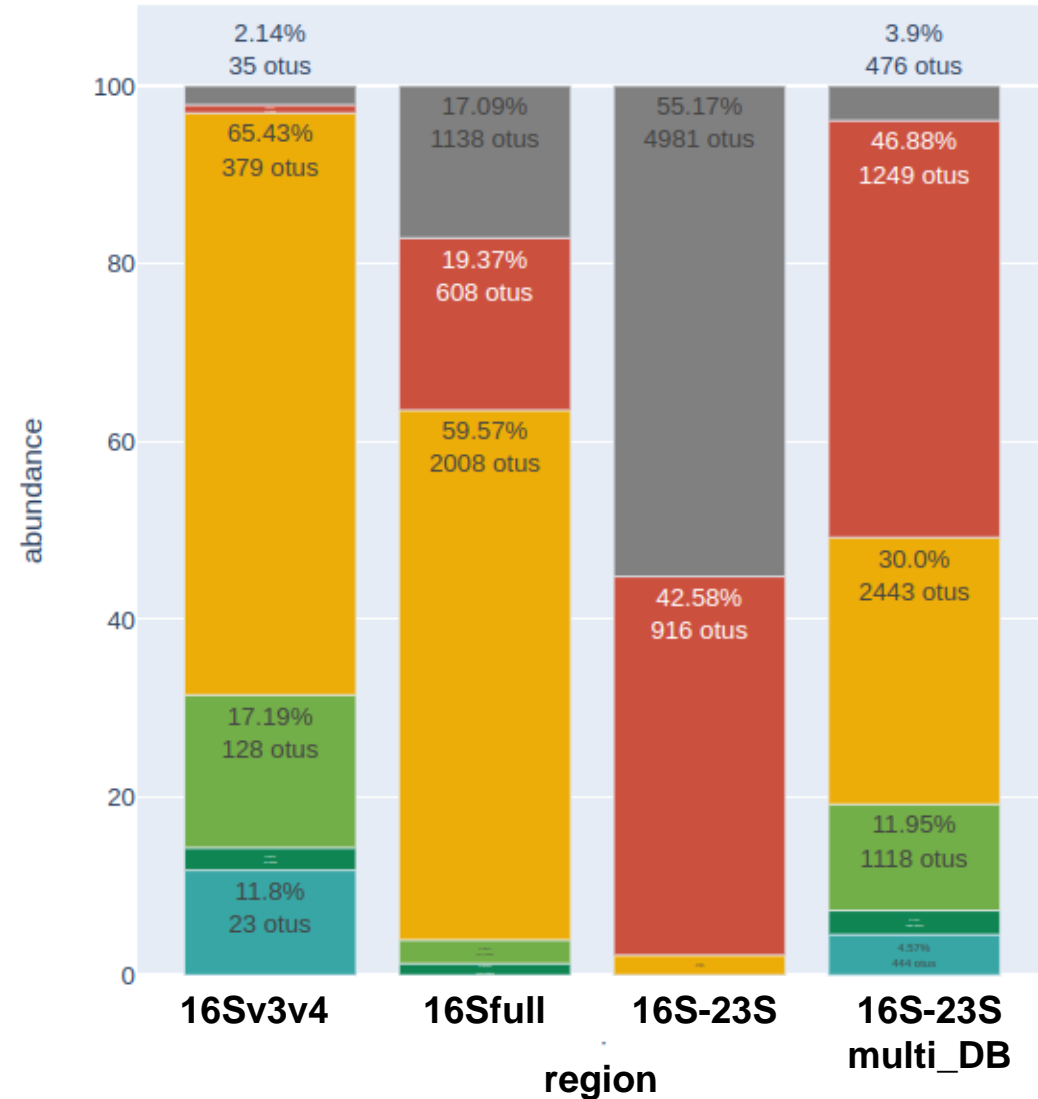
Results: Affiliation of OTUs from 16S23S metabarcoding by different databases – pig faeces sample



Results: Affiliation of OTUs from metabarcodings with different targets - pig faeces sample

Taxonomic rank

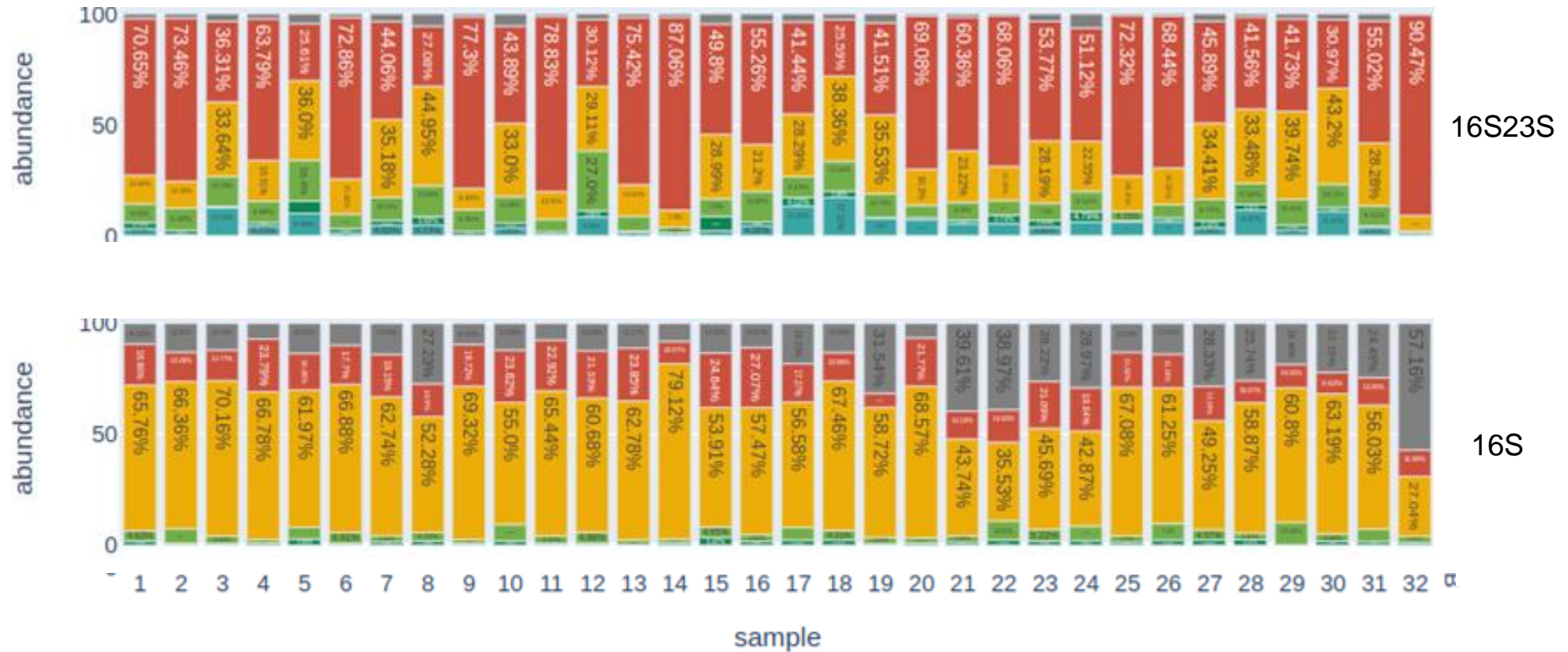
- Species
- Genus
- Family
- Order
- Class
- Weak affiliation



Results: Affiliation of OTUs from metabarcodings with different targets - 32 pig faeces samples

Taxonomic rank

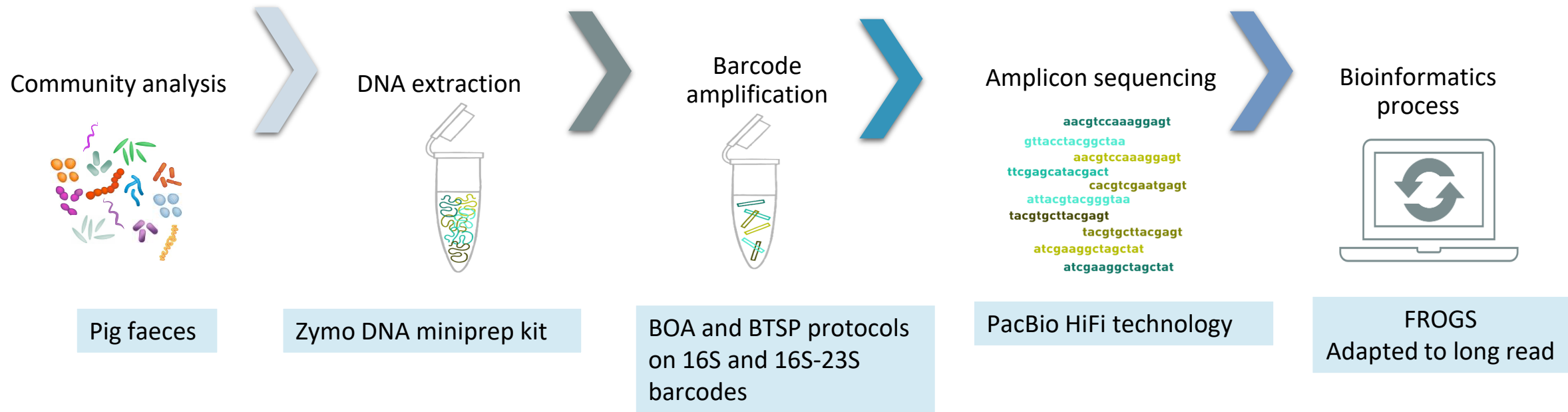
- Species
- Genus
- Family
- Order
- Class
- Weak affiliation



➤ 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 :

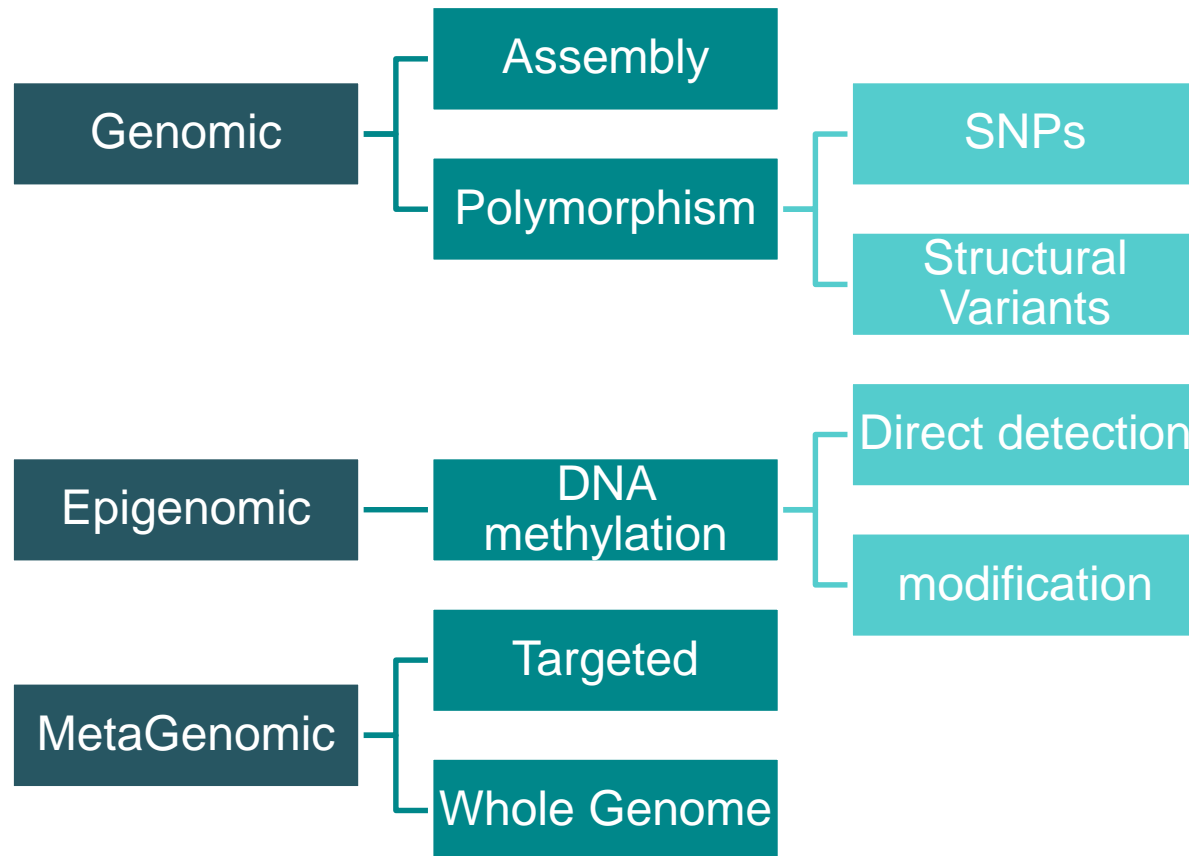


- 16S-23S region is more informative** than full length 16S
- Problem remains with the database exhaustivity that require time to be completed



- Further development of a taxonomic affiliation method by combining the 16S-23S, 16S and 23S databases.
- Next step is to **deploy all Seqoccln developments to the scientific community => Search for pilot projects.**

get-plage.contact@genotoul.fr



Thanks to

PacBio

SeqOccln

Coordination

Cécile Donnadieu
Christine Gaspin
Carole Iampietro
Denis Milan

Jean Mainguy
Olivier Bouchez
Sylvie Combes
Claire Hoede
Géraldine Pascal

