Analyse de données métagénomiques 16S

Module 20

Olivier Rué
Migale

Mahendra Mariadassou
MaIAGE

Cédric Midoux
PROSE & MaIAGE

September 11, 2023

Introduction

Practical informations

- 9h00 - 17h00
- 2 breaks morning and afternoon
- Lunch at INRAE restaurant (not mandatory)
- Questions are strongly encouraged
- Everyone has something to learn from each other

Better knwow you

Who are you?

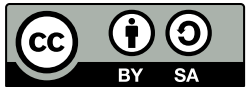
- Institution / Laboratory / position

What is your scientific topic?

- Studied ecosystem
- Scientific question
- Experimental design

What is your background?

- Already treated shotgun data?
- Background in bioinformatics?

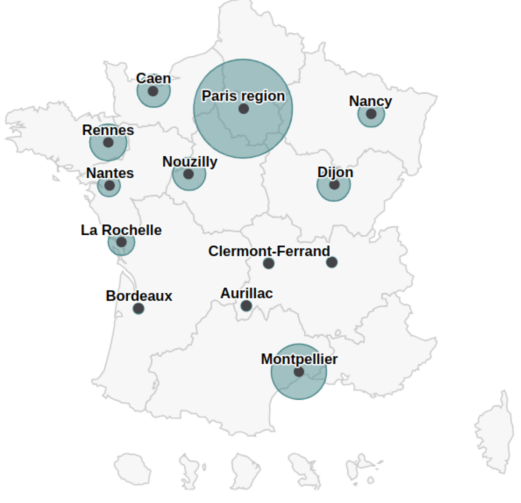


Better know us

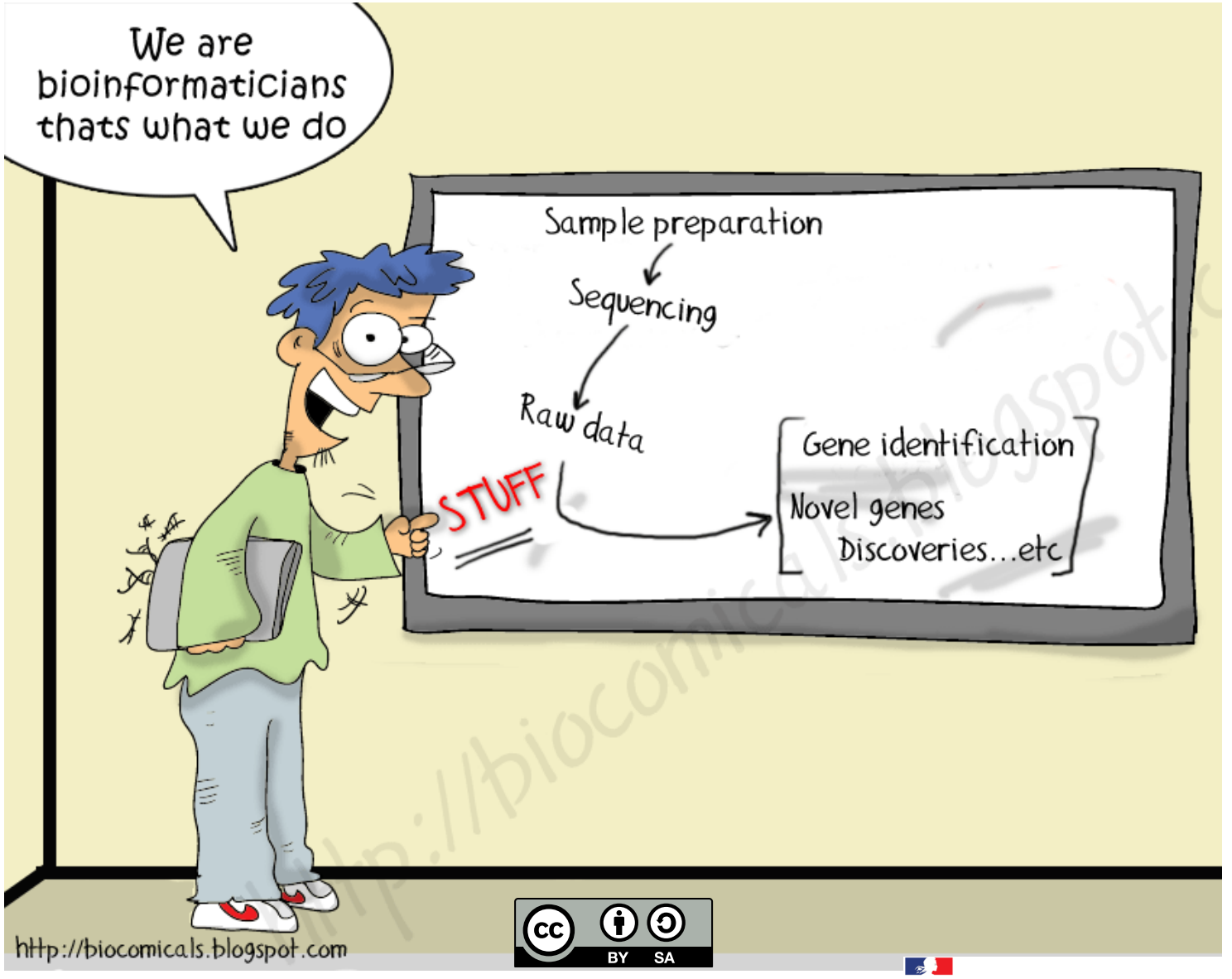
- Open infrastructure dedicated to life sciences
  - Computing resources, tools, databanks…
- Dissemination of expertise in bioinformatics
- Design and development of applications
- Data analysis

Data analysis service

- We are specialized in genomics/metagenomics
- 3 **Bioinformaticians** and 2 **Statisticians**
- More than 140 projects since 2016
- 2 types of partnership
  - Classical collaboration (we perfom the analyses) 🛠
  - Accompaniment (we help you do the analysis yourself) 🤝



Our expectations

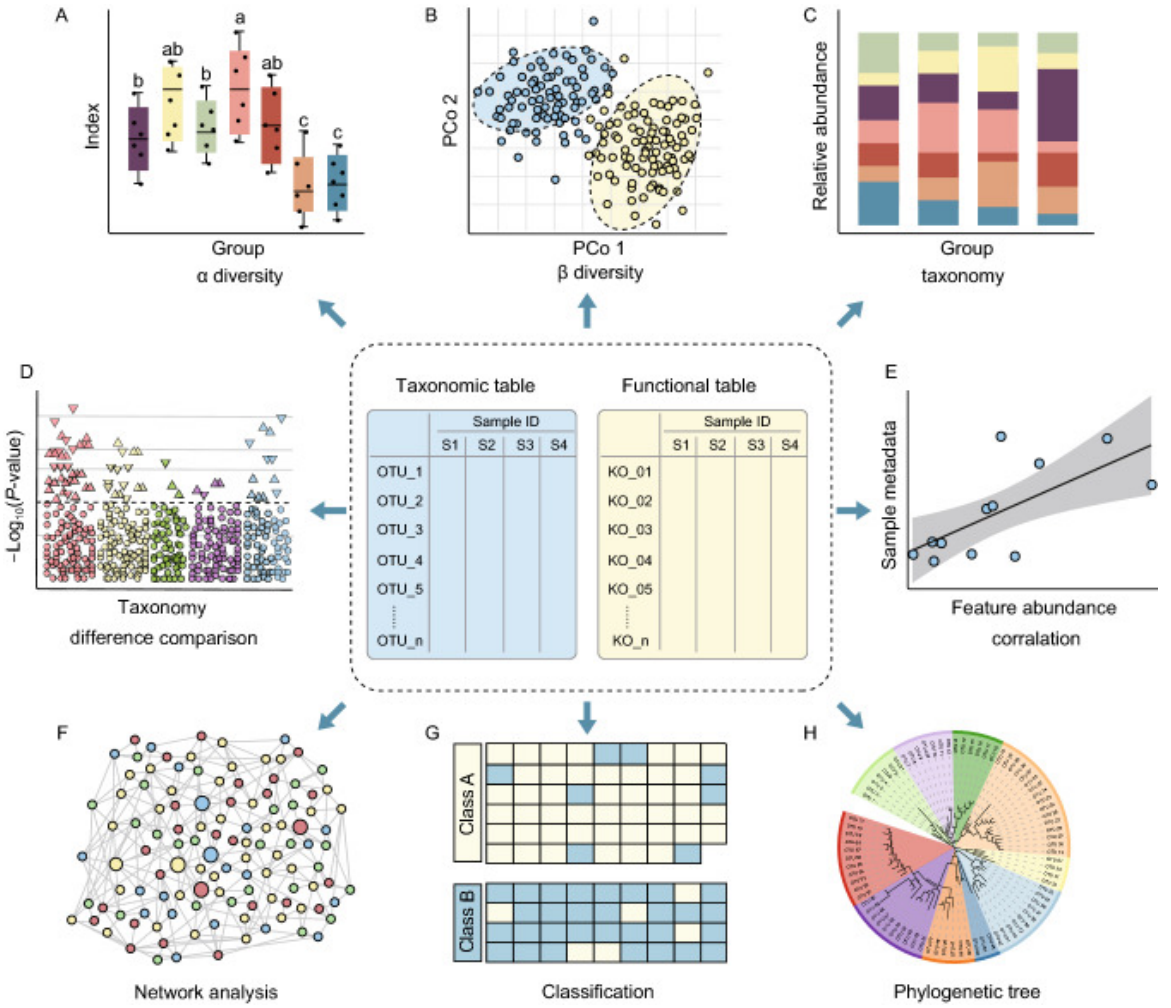


Aim of this training

After this 4 days training, you will:

- Know the outlines, **advantages** and **limits** of amplicon sequencing data analysis
- Be able to use **FROGS** (through Galaxy) and **phyloseq** (through **easy16S**) tools on the training data set
- Be able to identify tools and parameters adapted to your own analyses

Aim of this training



Liu et al., 2020: A practical guide to amplicon and metagenomic analysis of microbiome data [1]

Program

DAY 1

- Introduction
- Data import on Easy16S
- α and β diversities
- Ordination

DAY 2

- PERMANOVA and hypothesis tests
- Differential abundance
- Analysis of *Ravel* and *Mach* data
- Introduction to amplicon analysis (1)

DAY 3

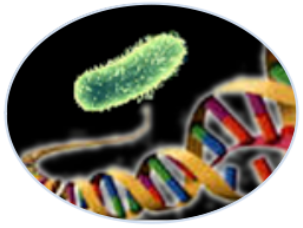
- Introduction to amplicon analysis (2)
- Introduction to Galaxy
- Quality control
- FROGS (1)

DAY 4

- FROGS (2)
- FROGSfunc
- Analysis of your data

Introduction to amplicon analyses

Meta-omics using next-generation sequencing (NGS)



DNA

RNA

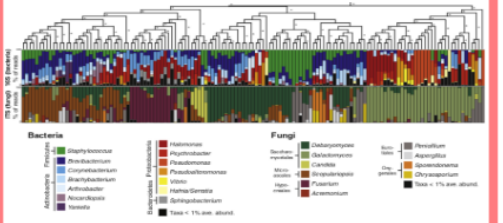
Metagenomics

Metatranscriptomics

Amplicon sequencing

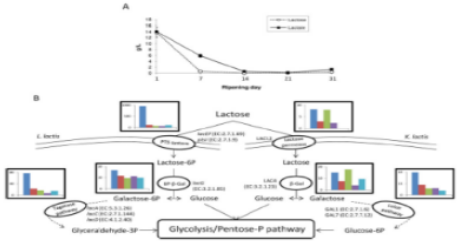
Shotgun sequencing

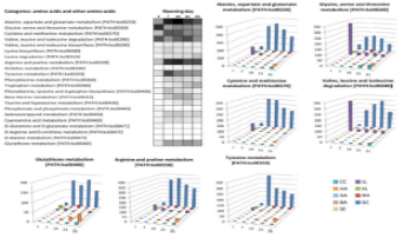
RNA sequencing



Wolfe *et al.*, 2014

Almeida *et al.*, 2014





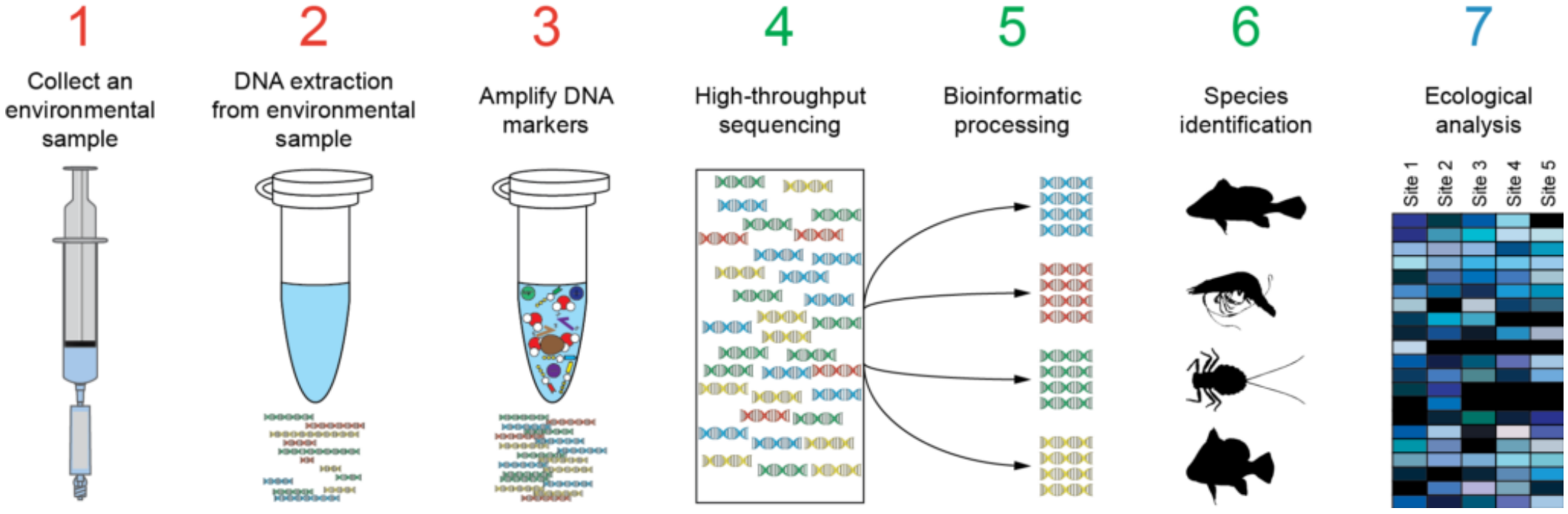
Dugat-Bony *et al.*, 2015

Who is here?

What can they do?

What are they doing?

Meta-omics using next-genertation sequencing (NGS)



Strengths and weaknesses of amplicon analyses?

http://scrumblr.ca/strengths_weaknesses

Strengths

- Detect subdominant microorganisms present in complex samples → microbial inventories
- Get (approximate) relative abondances of different taxa in samples
- Analyze and compare many taxa (hundreds) at the same time
- Taxonomic profiles of the communities (usually up to genus level, and sometimes up to species or strain)
- Low cost

Weaknesses

- Compositional data, many biases -> no absolute quantification
- Exact identification of the organisms difficult
- Hard to distinguish live and dead fractions of the communities
- No functional view of the ecosystem

The gene marker power

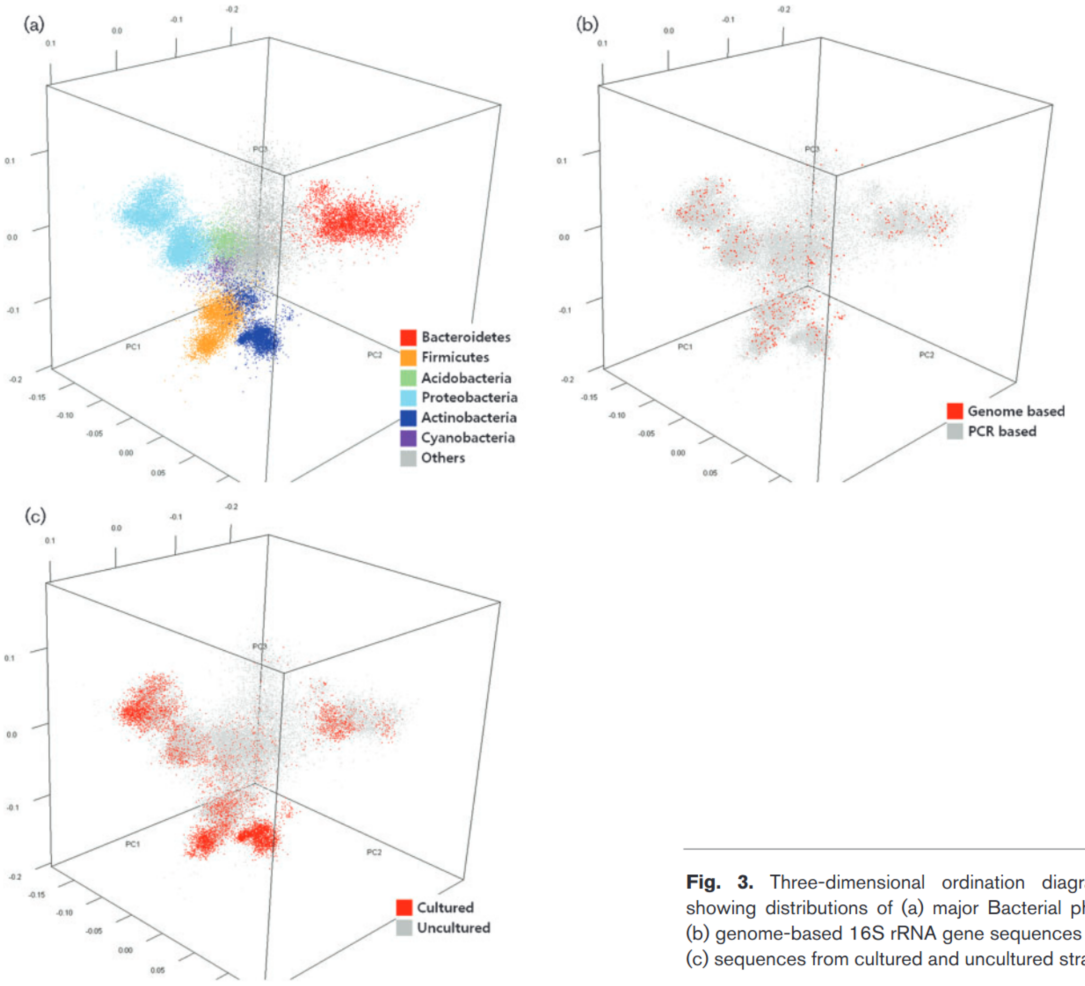
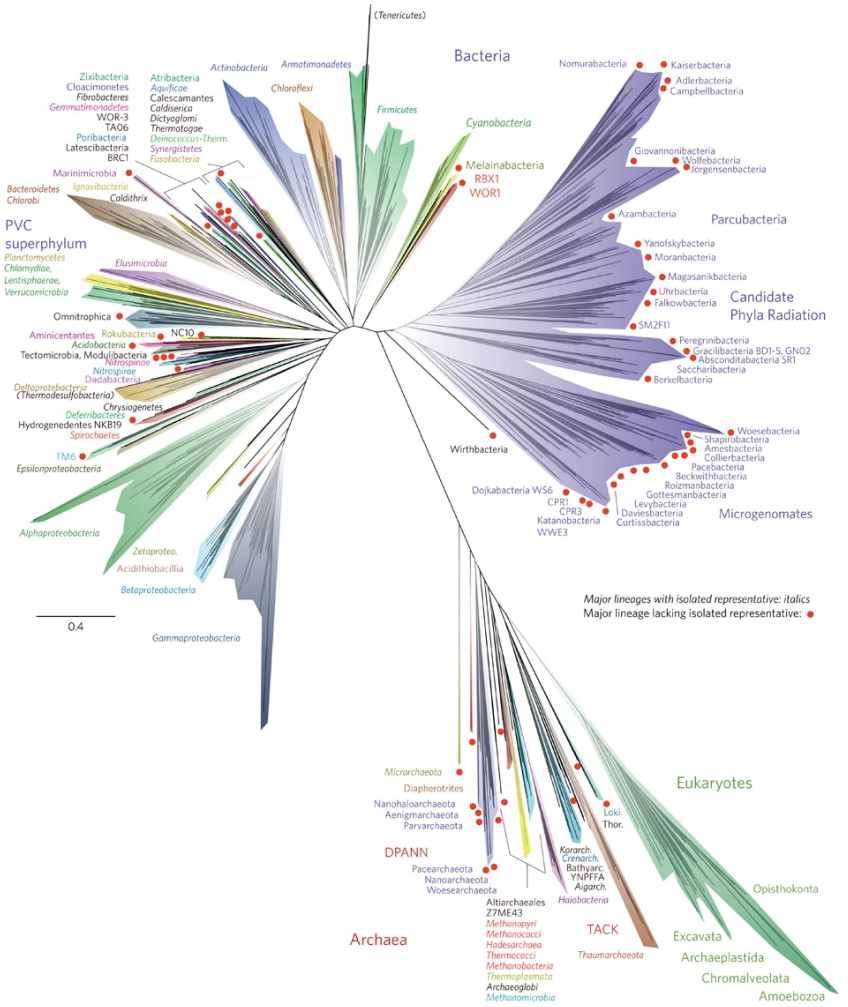


Fig. 3. Three-dimensional ordination diagrams showing distributions of (a) major Bacterial phyla, (b) genome-based 16S rRNA gene sequences and (c) sequences from cultured and uncultured strains.

Kim et al., 2012: Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species [2]

Microbial tree of life



Hug et al., 2016: A new view of the tree of life [3]

Story of barcoding

- Early 2000's: beginning of barcoding
- 1st DNA barcode: 65 bases of the mitochondrial gene of Cytochrome Oxidase I (COI) dedicated to the identification of vertebrates
- 2007: 1st international published database (BOLD)
- 2009: chloroplastic markers - RBCL (Ribulose Biphosphate Carboxylase; 553 pairs of bases) and MATK (MATurase K; 879 pairs of bases) → standard markers for plants
- 2012: ITS, standard marker of fungi (length between 361–1475 bases in UNITE 7.1)
- 16S marker, mainly used for bacteria but no designated standard.

Choice of a marker gene

The perfect / ideal gene marker:

- is ubiquist
- is conserved among taxa
- is enough divergent to distinguish stains
- is not submitted to lateral transfer
- has only one copy in genome
- has conserved regions to design specific primers
- is enough characterized to be present in databases for taxonomic affiliation

Bacterial targets

The genes that have been proposed for this task include those encoding :

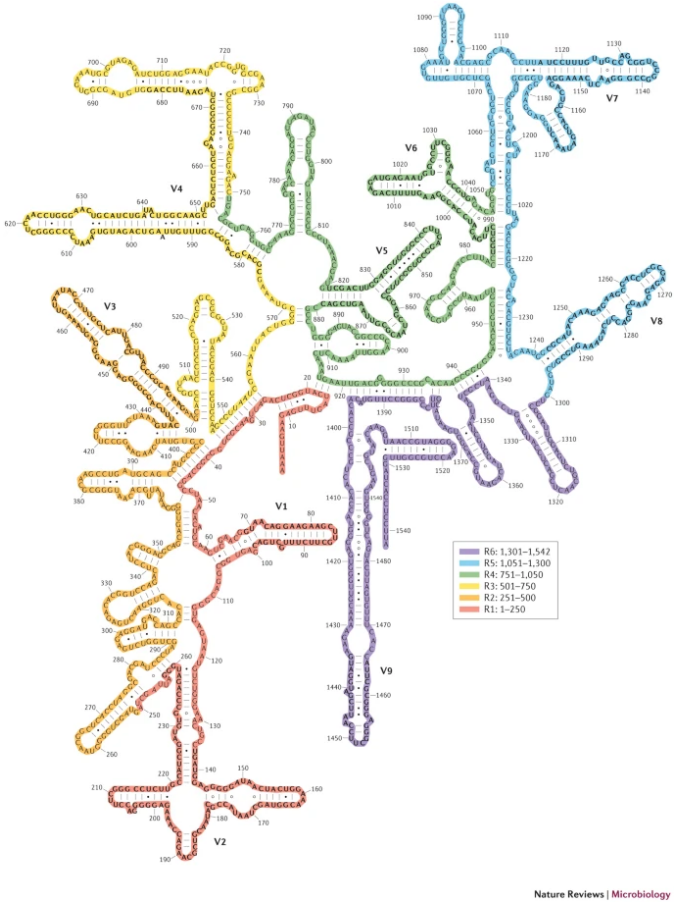
- 16S / 23S rRNA
- DNA gyrase subunit B (*gyrB*)
- RNA polymerase subunit B (*rpoB*)
- TU elongation factor (*tuf*)
- DNA recombinase protein (*recA*)
- protein synthesis elongation factor-G (*fusA*)
- dinitrogenase protein subunit D (*nifD*) …

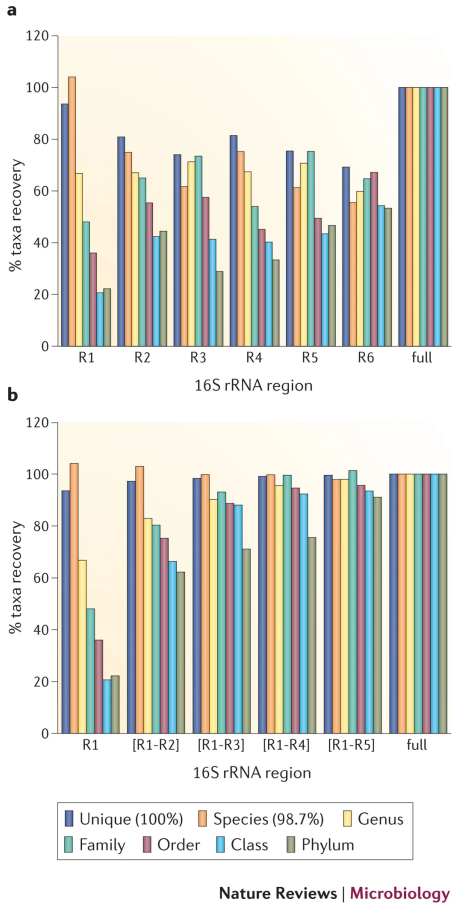
Bacterial lineages vary in their genomic contents, which suggests that different genes might be needed to resolve the diversity within certain taxonomic groups.

The gene encoding the small subunit of the ribosomal RNA

- The most widely used gene in molecular phylogenetic studies
- Ubiquist gene: 16S rDNA in prokayotes ; 18S rDNA in eukaryotes
- Gene encoding a ribosomal RNA : non-coding RNA (not translated), part of the small subunit of the ribosome which is responsible for the translation of mRNA in proteins
- Not submitted to lateral gene transfer
- Availability of databases facilitating comparison
  - Silva v138.1 - 2021: available SSU/LSU sequences to over 10,700,000

The 16S resolution

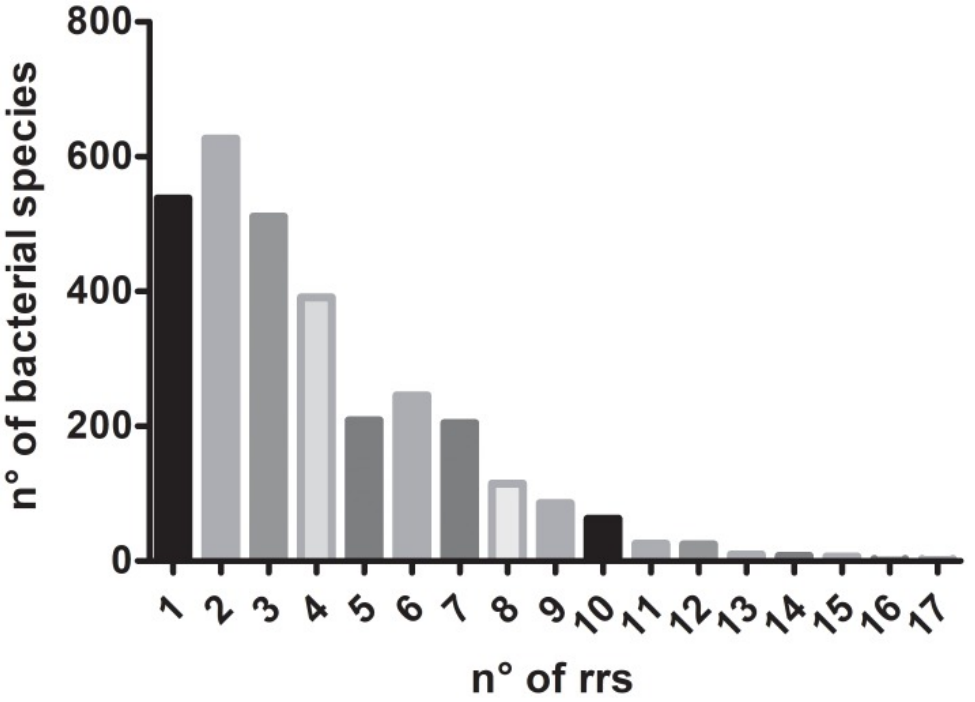




Analysis of the taxa recovery rate indicates a great underestimation of ta… partial sequences are used. Although the situation tends to ameliorate as longer segments are considered, near full-length 16S rRNA genes sequences are required for accurate rich… taxa.

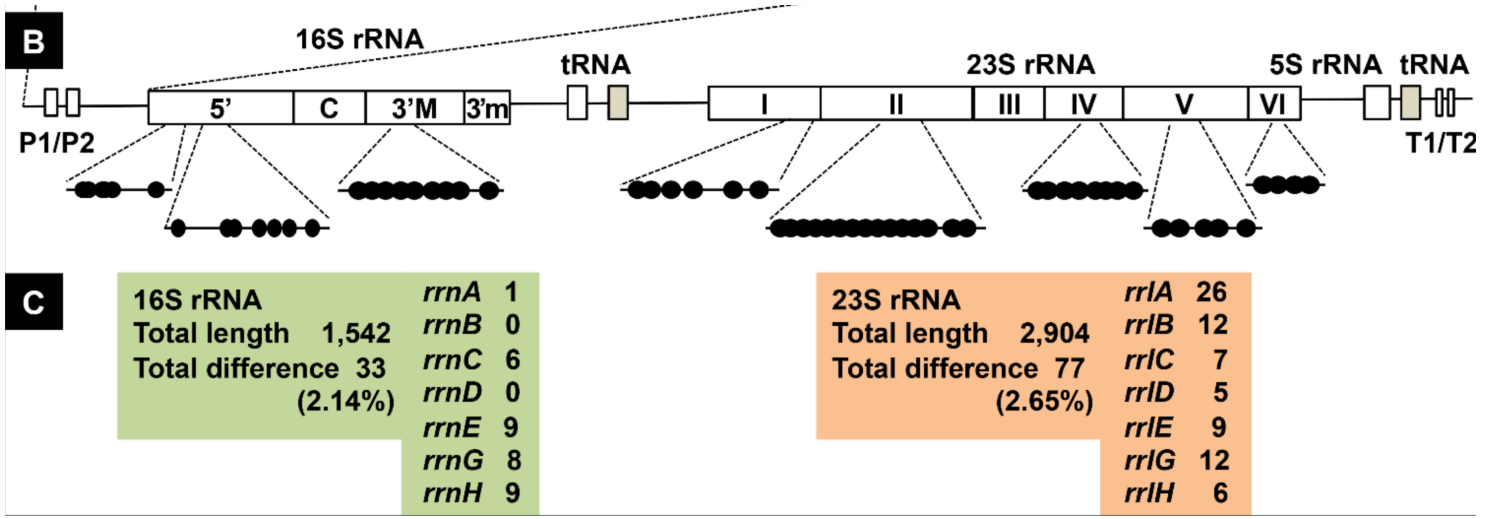
16S rRNA copy number

Median of the number of 16S rRNA copies in 3,070 bacterial species according to data reported in rrnDB database – 2018



Espejo and Plaza, 2018: Multiple ribosomal RNA operons in bacteria; their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA [5]

16S rRNA copy variation



[B] The positions of sequence variation within 16S and 23S rRNA are shown along the gene organization of rrn operons. A total of 33 and 77 differences were identified in 16S rRNA and 23S rRNA, respectively.

[C] The number of bases that are different from the conserved sequence are shown for 16S and 23S rRNA for each rrn operon

Maeda et al., 2015: Strength and regulation of seven rRNA promoters in Escherichia coli [6]

16S rRNA copy variation

- Only a minority of bacterial genomes harbors identical 16S rRNA gene copies
- Sequence diversity increases with increasing copy numbers
- While certain taxa harbor dissimilar 16S rRNA genes, others contain sequences common to multiple species

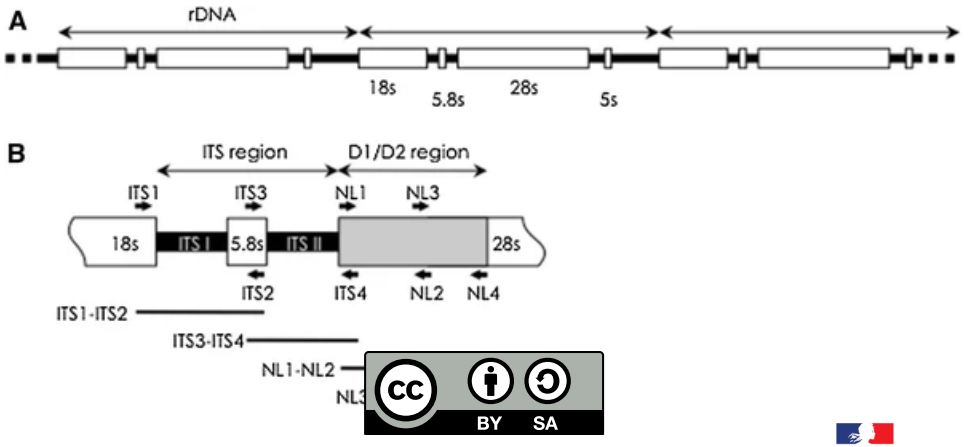
gyrB: an alternative of 16S

- A single-copy housekeeping gene that encodes the subunit B of DNA gyrase, a type II DNA topoisomerase, and therefore plays an essential role in DNA replication.
- Essential and ubiquitous in bacteria
- Higher rate of base substitution than 16S rDNA does
- Sufficiently large in size for use in analysis of microbial communities.
- Also present in Eukarya and sometimes in Archaea but it shows enough sequence dissimilarity between the three domains of life to be used selectively for Bacteria.

Poirier, Rué et al., 2019: Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing [7]

Fungal ITS

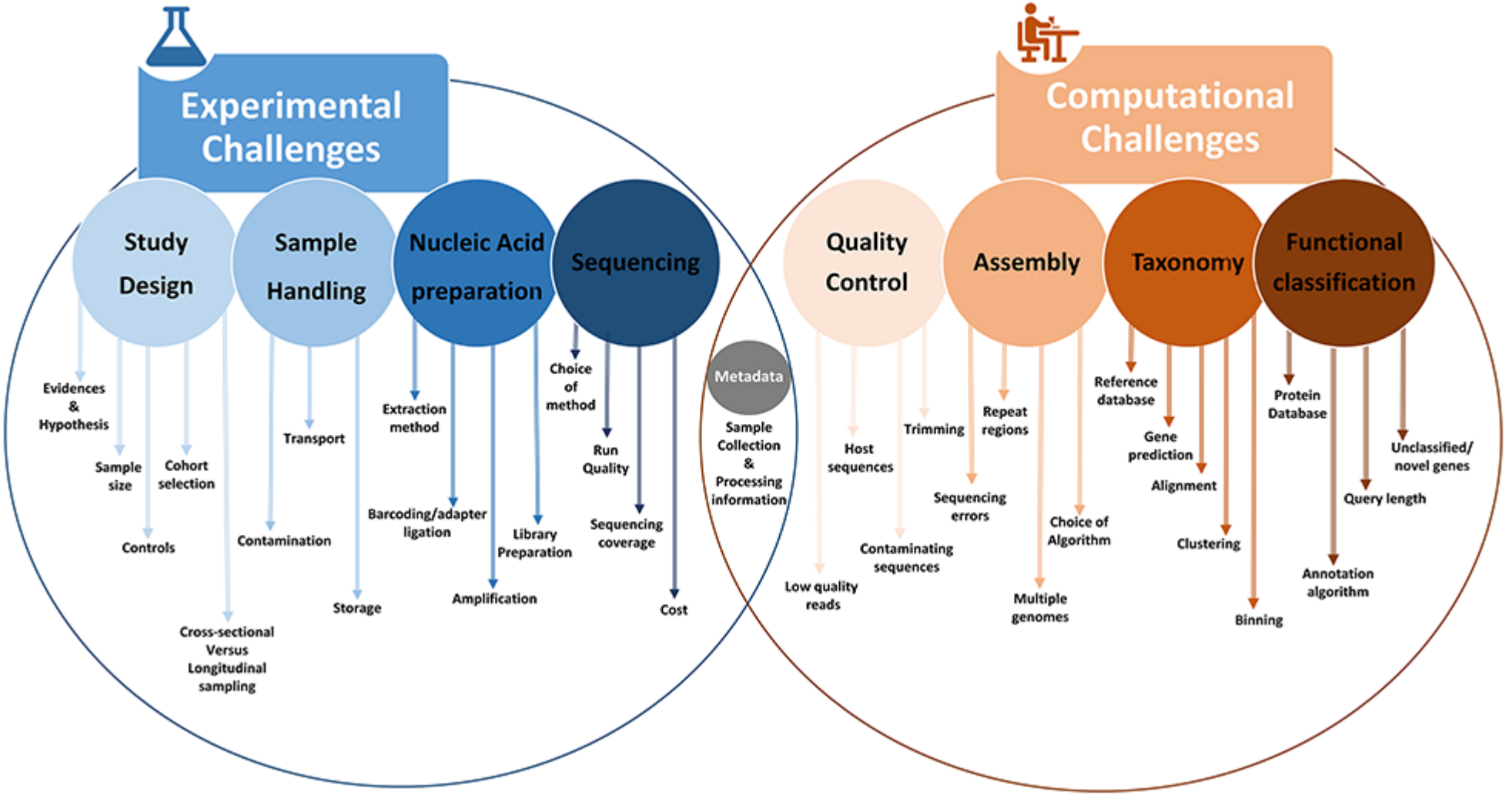
- ITS: Internal Transcribed Spacer
- Size polymorphism of ITS (from 361 to 1475 bases in UNITE 7.1)
- Highly conserved regions of the neighboring of ITS1 and ITS2
- Lack of a generalist and abundant ITS databank (several small specialized databanks)
- Multiple copies (14 to 1400 copies (mean at 113, median at 80))
- FROGS deals very good with ITS [8]
  - small and long fragments contrary to many tools



Lofgren et al., 2019: Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles [9]

Planning an experiment

Challenges



Bharti and Grimm, 2019: Current challenges and best-practice protocols for microbiome analysis [10]

Experimental design: challenges and solutions

- In general, any hypothesis should primarily be supported by meticulous literature driven evidence and preliminary testing using small-scale/pilot studies to avoid uncertainty in biological signals, trials and failures
  - **Number of samples**: variability between similar samples / choosing appropriate sample sizes based on statistical principles can certainly help to avoid biases and spurious interpretations
  - **Controls**: needed to identify whether a signal is real and not just a stochastic or spurious result
  - **Cross-sectional or longitudinal studies**: it is equally important to cautiously plan identical sample collection times for each replicate to avoid biases
  - **Metadata**: help to avoid false interpretation of results and highlights the effective size of individual factors

Sample collection and handling

- **Contamination**: changes in temperature, humidity, or other factors could alter or contaminate samples. Minimizing the time of sample collection and using aseptic laboratory resources, including gloves, masks and head covers, help to reduce contamination
- **Transportation**: Transit conditions and duration can influence the quality and quantity of extracted nucleic acids
- **Storage and safety**: Several studies have assessed the effect of storage conditions on compositional changes in microbial samples

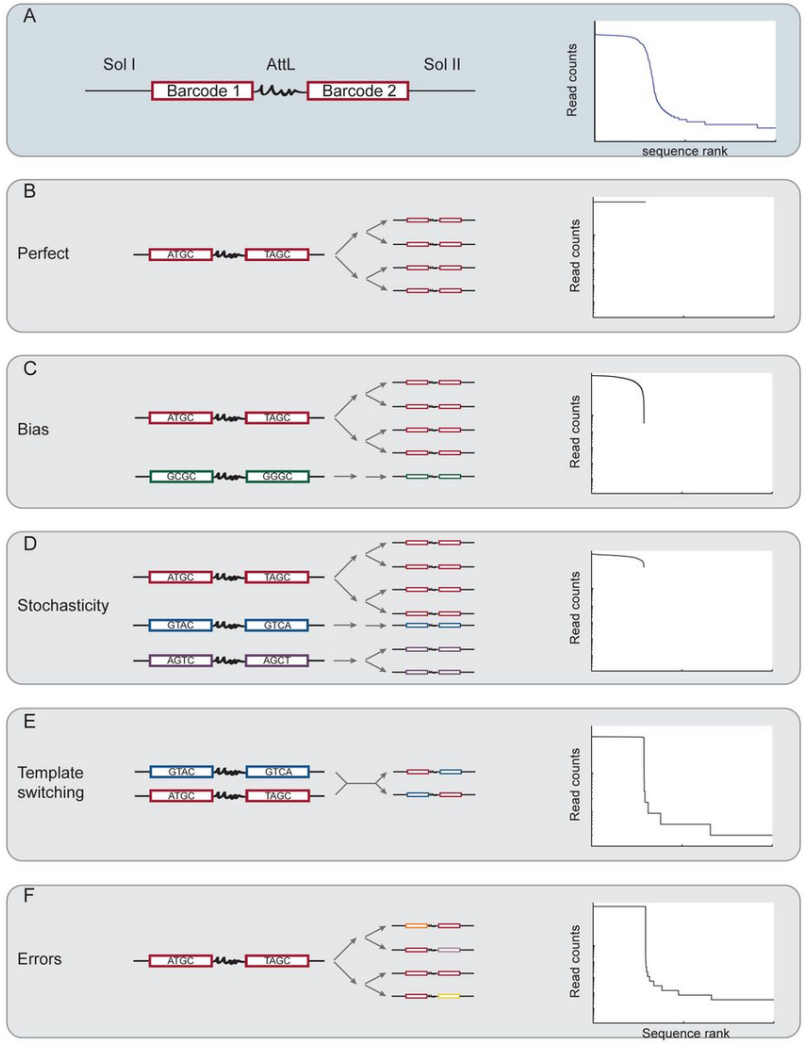
DNA extraction and preparation

- mechanical lysis/bead beating or chemical lysis
- **amplification using barcode primer pairs**, purification, and preparation of purified DNA libraries are done before sequencing
  - universal primers are not so universal [11]
  - amplification bias

Amplification bias

- Amplification by PCR has sequence-dependence efficiency, especially the sequence that binds to primers.
- If one sequence is amplified 10% more than another in one round, it will be 1.130 = 17.4 x more abundant after 30 rounds.
- This effect is most important when the sequence has one or more mismatches with the primer.
- With one mismatch, amplification efficiency is usually significantly less, and with two or more mismatches the sequence may not be amplified to detectable levels.

Amplification bias



- C and D impact the abundance without adding new sequences
- E and F add new sequences

Sequencing technologies

Sequencing technologies

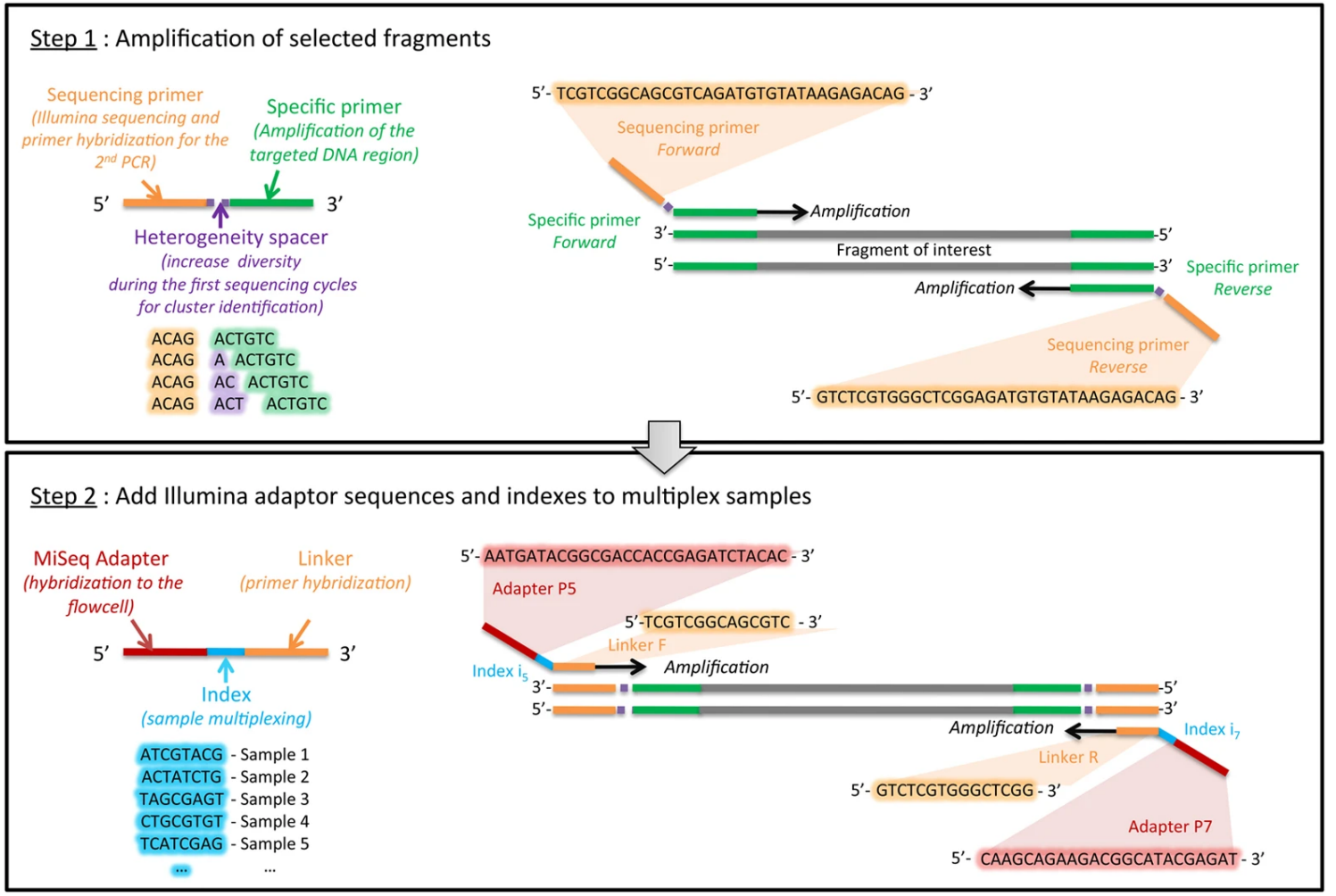
Sequencing technologies

Table 2 Comparison between next-generation sequencing technologies

Method	Illumina	Pacific Bio	Nanopore	Pyrosequencing (454)	SOLiD
Read length per run	50–300 base pair	10–25 kilo base pair	500–2.3 mega base pair	Approximate 800 base pair	50 base pair
Time taken per run	1 to 10 days	Up to 30 h	1 min–72 h	24 h	1 to 2 weeks
Cost	$148 per Gb	$2000 Gb	$60–80 per sample	$7000 per sample	$15,000 per 100 Gb
Accuracy	98%	99.9%	98.9–99.6%	99.9%	99.9%
Advantages	Cost-effective, high-yield sequence reads	Fast, long read lengths	Real-time analysis, long read lengths	Fast, long read lengths	High accuracy
Disadvantages	Instrument cost, high maintenance of instrument, read length	Low high throughput	Error prone	Homopolymer error	Long run time, low read length

Tan et al., 2022: Bioinformatics approaches and applications in plant biotechnology [12]

Illumina technology



Cruaud et al., 2017: High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy [13]

Illumina technology

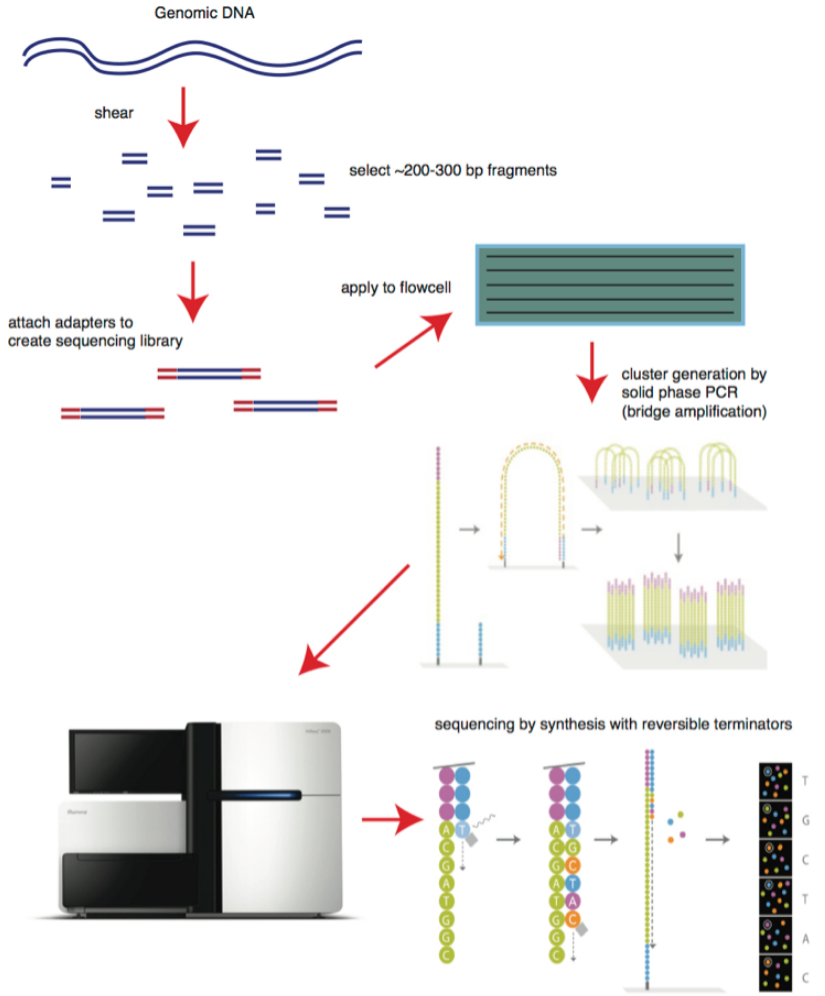
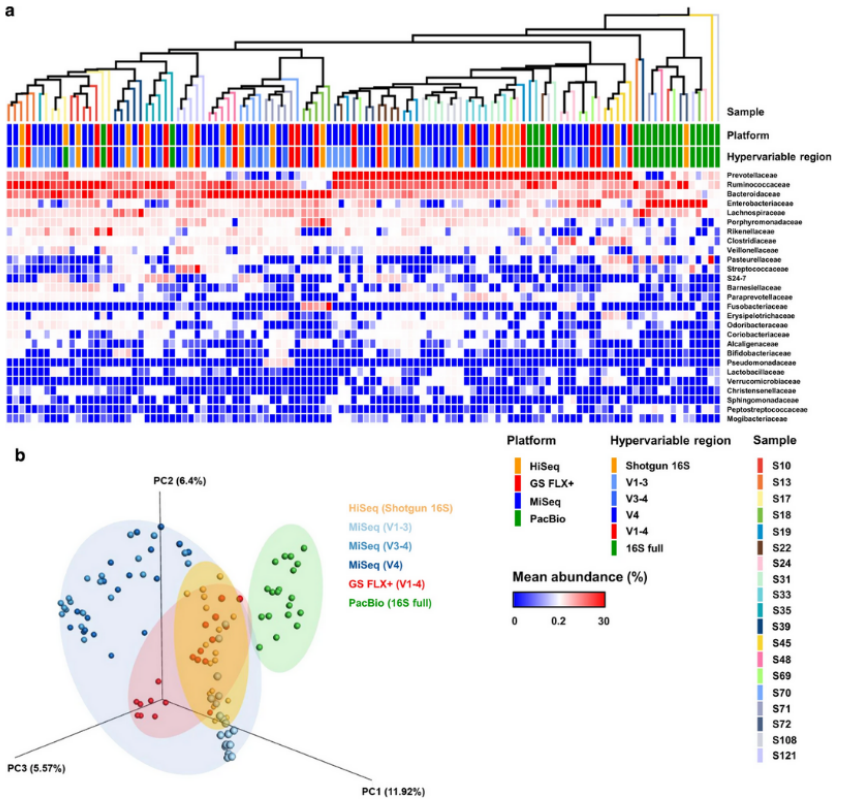


Image credit: BiteSizeBio, 2012

Illumina technology

Effect of sequencing technology



Fecal samples collected from 19 human subjects were sequenced using the indicated platforms: GS FLX+ (V1–4, red), Illumina MiSeq (V1–3, light blue; V3–4, blue; V4, dark blue), and PacBio CCS (V1–9, green). Whole-genome shotgun sequences generated by Illumina HiSeq (Shotgun 16 S, orange) were included as a reference for community structure without amplification bias. (**a**) The sequence data were clustered using a UPGMA dendrogram based on the Bray-Curtis dissimilarity matrix, and samples from the same individual are shown in the same color. The relative abundances of bacterial taxa are displayed as a heatmap over 27 families (>1% relative abundance). (**b**) The sequence data were clustered by principal component analysis.

Whon et al., 2018: The effects of sequencing platforms on phylogenetic resolution in 16 S rRNA gene profiling of human feces [14]

Sequencing biases

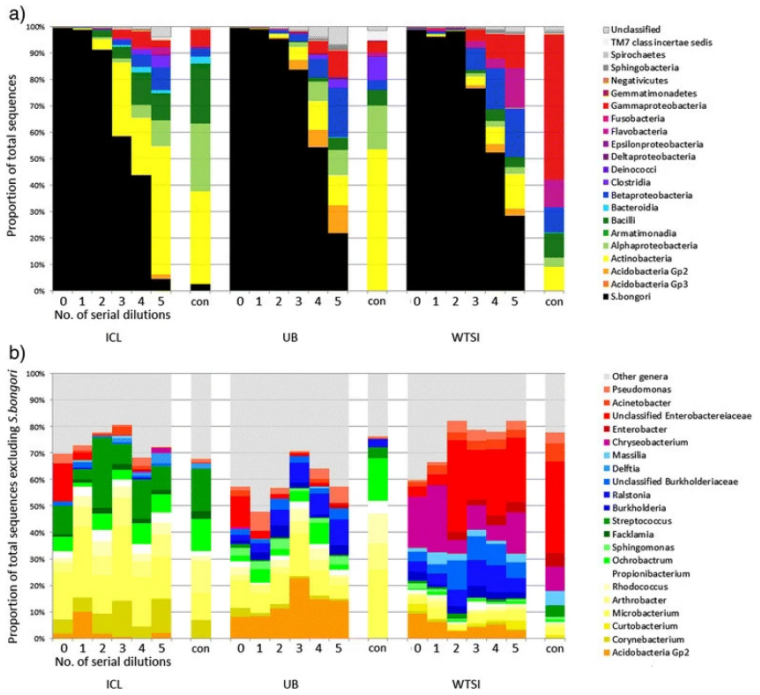
- Contamination between samples during the same run
- Contamination during successive runs (residual contaminants)
- Variability between runs: take into account for experimental plan
- Variability inside run: add some controls

Salter et al., 2014: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses [15]

Interest of controls

Figure 1

From: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses



Summary of 16S rRNA gene sequencing taxonomic assignment from ten-fold diluted pure cultures and controls. Undiluted DNA extractions contained approximately 10^8 cells, and controls (annotated in the Figure with 'con') were template-free PCRs. DNA was extracted at ICL, UB and WTSI laboratories and amplified with 40 PCR cycles. Each column represents a single sample; sections **(a)** and **(b)** describe the same samples at different taxonomic levels. **a)** Proportion of *S. bongori* sequence reads in black. The proportional abundance of non-*Salmonella* reads at the Class level is indicated by other colours. As the sample becomes more dilute, the proportion of the sequenced bacterial amplicons from the cultured microorganism decreases and contaminants become more dominant. **b)** Abundance of genera which make up >0.5% of the results from at least one laboratory, excluding *S. bongori*. The profiles of the non-*Salmonella* reads within each laboratory/kit batch are consistent but differ between sites.

Salter et al., 2014: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses [15]

Interest of controls

Table 1 List of contaminant genera detected in sequenced negative `blank' controls

From: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses

Phylum	List of constituent contaminant genera
Proteobacteria	Alpha-proteobacteria:
	Afipia, Aquabacterium[e], *Asticcacaulis, Aurantimonas, Beijerinckia, Bosea, Bradyrhizobium*[d], *Brevundimonas*[c], *Caulobacter, Craurococcus, Devosia, Hoeflea*[e], *Mesorhizobium, Methylobacterium*[c], *Novosphingobium, Ochrobactrum, Paracoccus, Pedomicrobium, Phyllobacterium*[e], *Rhizobium*[c,d], *Roseomonas, Sphingobium, Sphingomonas*[c,d,e], *Sphingopyxis*
	Beta-proteobacteria:
	Acidovorax[c,e], *Azoarcus*[e], *Azospira, Burkholderia*[d], *Comamonas*[c], *Cupriavidus*[c], *Curvibacter, Delftia*[e], *Duganella*[a], *Herbaspirillum*[a,c], *Janthinobacterium*[e], *Kingella, Leptothrix*[a], *Limnobacter*[e], *Massilia*[c], *Methylophilus, Methyloversatilis*[e], *Oxalobacter, Pelomonas, Polaromonas*[e], *Ralstonia*[b,c,d,e], *Schlegelella, Sulfuritalea, Undibacterium*[e], *Variovorax*
	Gamma-proteobacteria:
	Acinetobacter[a,d,c], *Enhydrobacter, Enterobacter, Escherichia*[a,c,d,e], *Nevskia*[e], *Pseudomonas*[b,d,e], *Pseudoxanthomonas, Psychrobacter, Stenotrophomonas*[a,b,c,d,e], *Xanthomonas*[b]
Actinobacteria	*Aeromicrobium, Arthrobacter, Beutenbergia, Brevibacterium, Corynebacterium, Curtobacterium, Dietzia, Geodermatophilus, Janibacter, Kocuria, Microbacterium, Micrococcus, Microlunatus, Patulibacter, Propionibacterium*[e], *Rhodococcus, Tsukamurella*
Firmicutes	*Abiotrophia, Bacillus*[b], *Brevibacillus, Brochothrix, Facklamia, Paenibacillus, Streptococcus*
Bacteroidetes	*Chryseobacterium, Dyadobacter, Flavobacterium*[d], *Hydrotalea, Niastella, Olivibacter, Pedobacter, Wautersiella*
Deinococcus-Thermus	*Deinococcus*
Acidobacteria	Predominantly unclassified Acidobacteria Gp2 organisms

The listed genera were all detected in sequenced negative controls that were processed alongside human-derived samples in our laboratories (WTSI, ICL and UB) over a period of four years. A variety of DNA extraction and PCR kits were used over this period, although DNA was primarily extracted using the FastDNA SPIN Kit for Soil. Genus names followed by a superscript letter indicate those that have also been independently reported as contaminants previously. [a]also reported by Tanner *et al.* [12]; [b]also reported by Grahn *et al.* [14]; [c]also reported by Barton *et al.* [17]; [d]also reported by Laurence *et al.* [18]; [e]also detected as contaminants of multiple displacement amplification kits (information provided by Paul Scott, Wellcome Trust Sanger Institute). ICL, Imperial College London; UB, University of Birmingham; WTSI, Wellcome Trust Sanger Institute.

Salter et al., 2014: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses [15]

Illustration

Here, we showed that contaminant OTUs from extraction and amplification steps can represent more than half the total sequence yield in sequencing runs, and lead to unreliable results when characterizing tick microbial communities. We thus strongly advise the routine use of negative controls in tick microbiota studies, and more generally in studies involving low biomass samples

ORIGINAL RESEARCH ARTICLE
Front. Microbiol., 09 June 2020 | https://doi.org/10.3389/fmicb.2020.01093

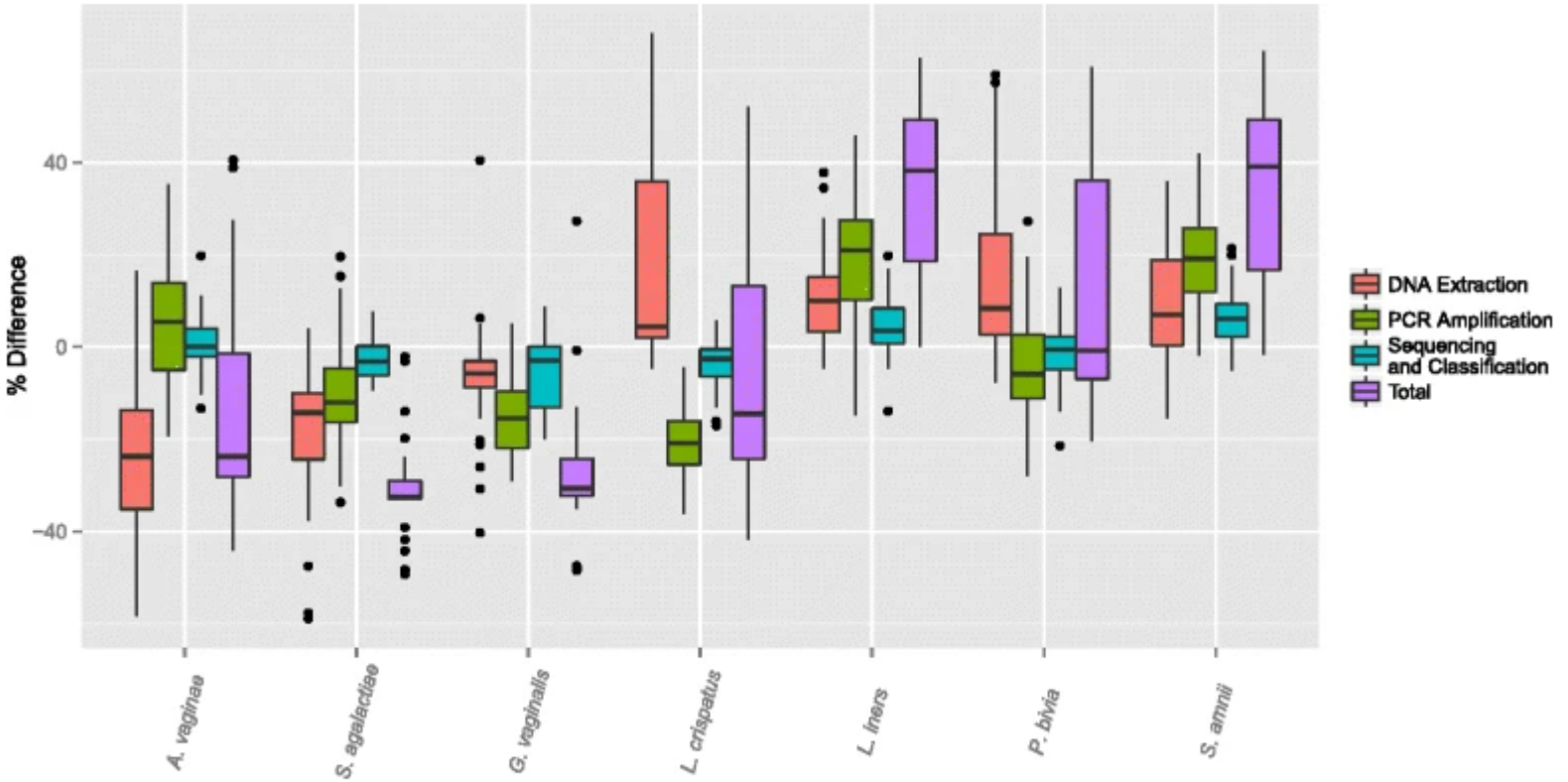
Taxon Appearance From Extraction and Amplification Steps Demonstrates the Value of Multiple Controls in Tick Microbiota Analysis

Emilie Lejal[1], Agustín Estrada-Peña[2], Maud Marsot[3], Jean-François Cosson[1], Olivier Rué[4,5], Mahendra Mariadassou[4,5], Cédric Midoux[4,5,6], Muriel Vayssier-Taussat[7] and Thomas Pollet[1,8*]

[1]UMR BIPAR, Animal Health Laboratory, INRAE, ANSES, Ecole Nationale Vétérinaire d'Alfort, Université Paris-Est, Maisons-Alfort, France
[2]Faculty of Veterinary Medicine, University of Zaragoza, Zaragoza, Spain
[3]Laboratory for Animal Health, Epidemiology Unit, ANSES, University Paris-Est, Maisons-Alfort, France
[4]INRAE, MaIAGE, Université Paris-Saclay, Jouy-en-Josas, France
[5]INRAE, Bioinfomics, MIGALE Bbioinformatics Facility, Université Paris-Saclay, Jouy-en-Josas, France
[6]INRAE, PROSE, Université Paris-Saclay, Antony, France
[7]Animal Health Department, INRAE, Nouzilly, France
[8]UMR ASTRE, CIRAD, INRAE, Montpellier, France

Lejal et al., 2020: Taxon appearance from extraction and amplification steps demonstrates the value of multiple controls in tick microbiota analysis [16]

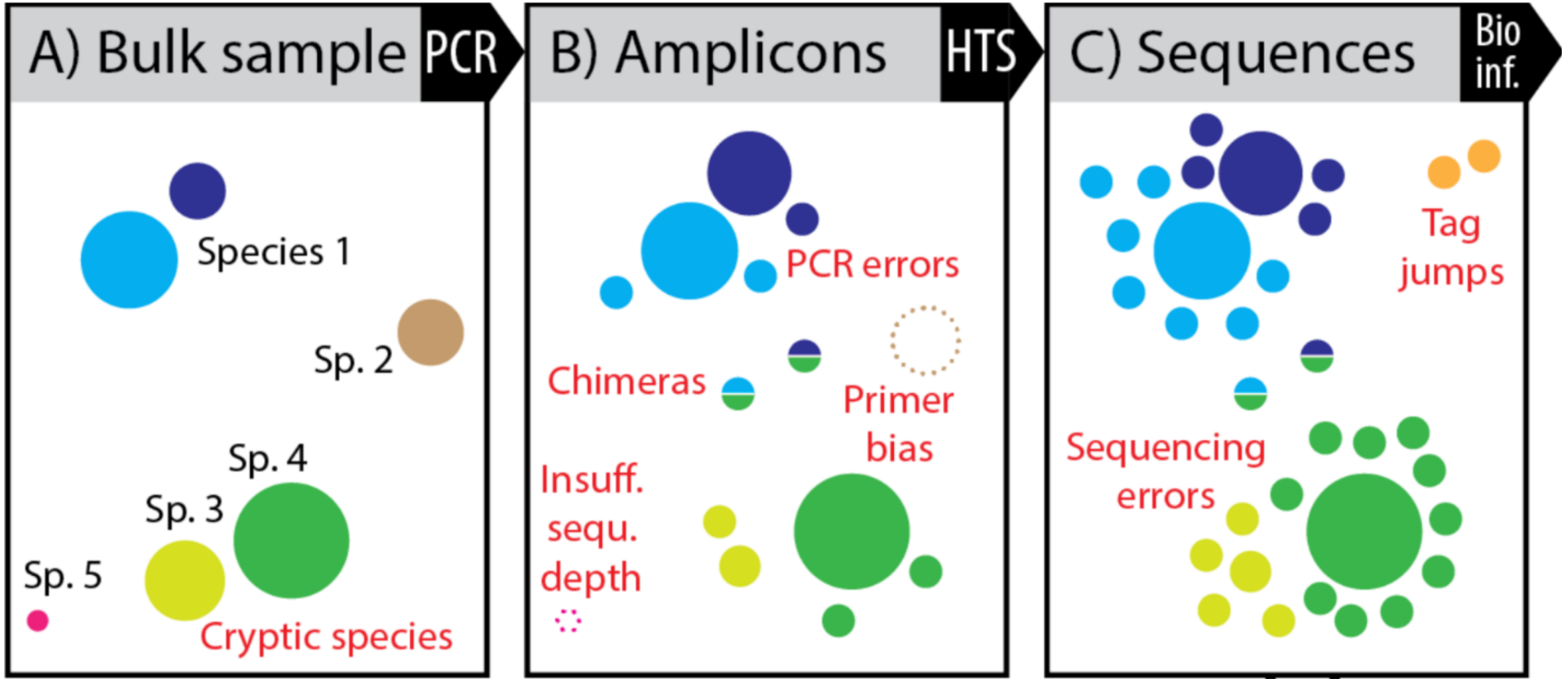
Synthetis of biases



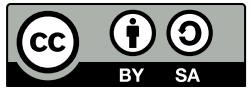
Observed bias by bacterium. The observed bias (the observed minus the actual proportions) for each bacterium in the experimental design due to the different effects of our DNA Extraction, PCR amplification, and sequencing and taxonomic classification protocols. The total bias is also plotted for each bacterium. For each box and whisker plot, only the samples including the bacterium were included.

Brooks et al., 2015: The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies [17]

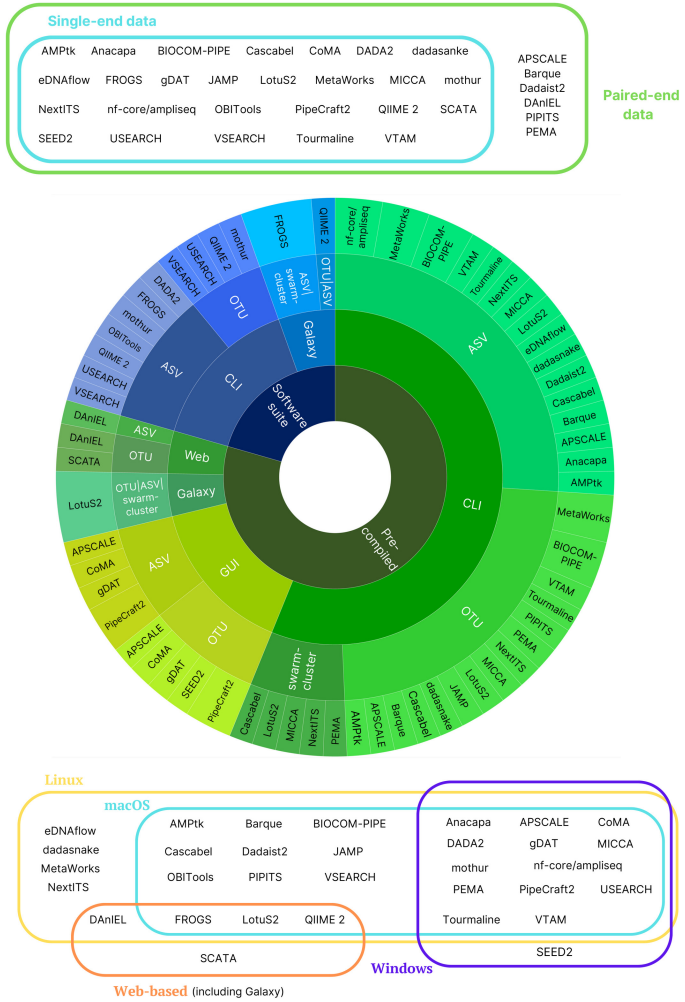
Synthesis



Bioinformatics



A pile of pipelines



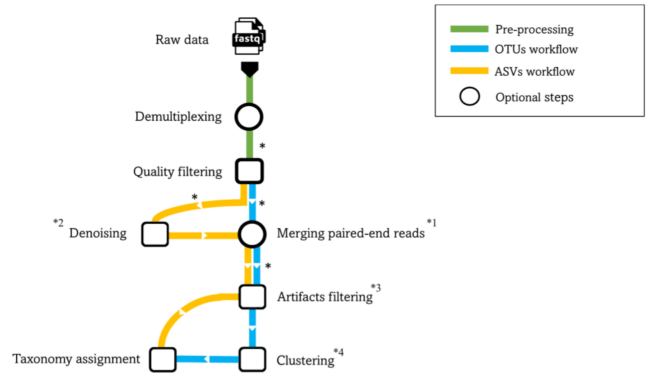
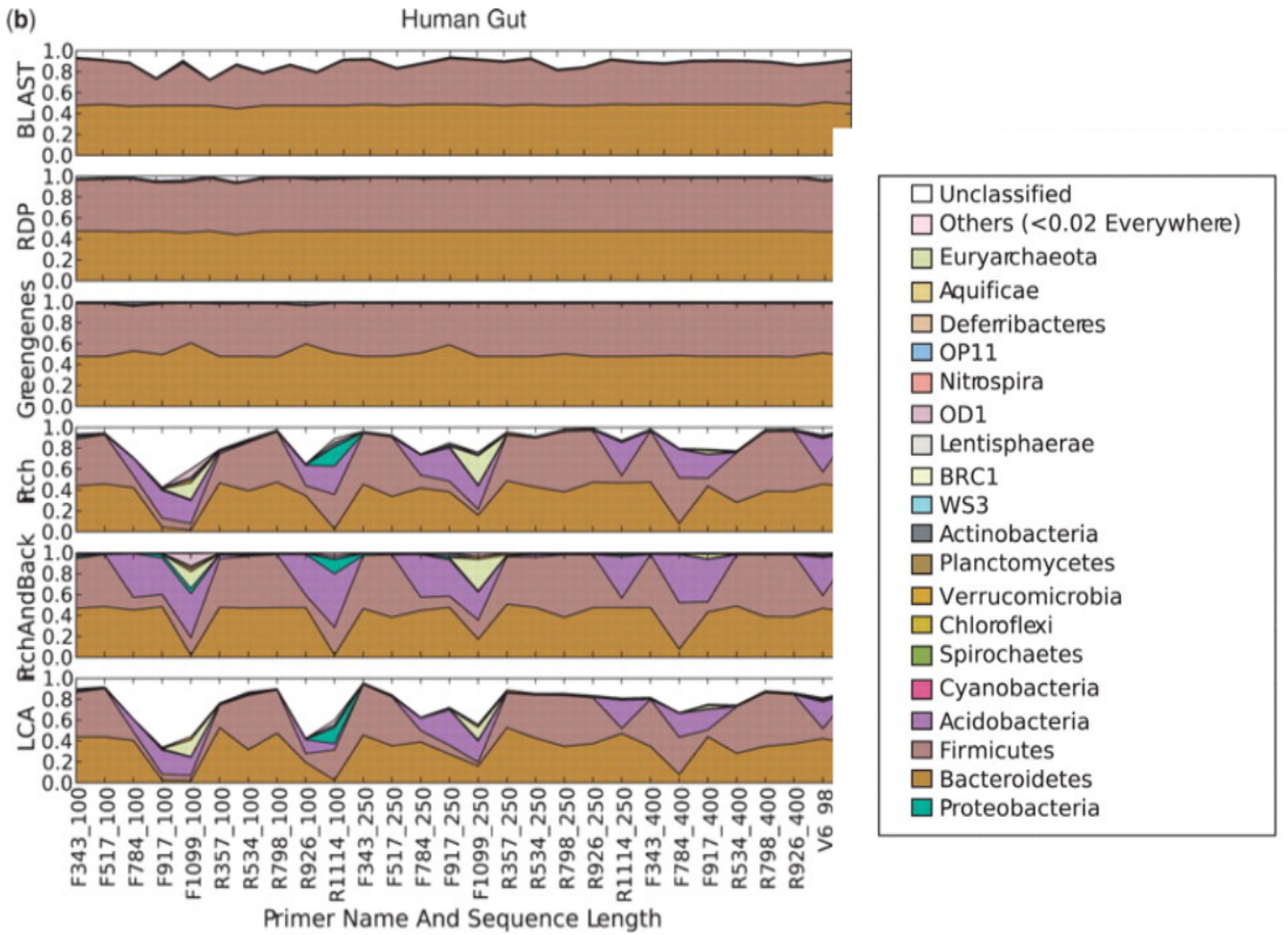


FIGURE 1 Examples of basic bioinformatics workflows for metabarcoding data. The workflow begins with demultiplexing, assigning reads to respective samples based on unique molecular identifiers. Next, quality filtering removes low-quality reads to reduce errors and improve reliability. Denoising algorithms identify and correct sequencing errors while preserving biological variation. For paired-end reads, merging combines forward and reverse reads into single-end sequences. Artifacts filtering aims to remove artifacts such as chimeras and NUMTs. Clustering groups of sequences into features. Finally, taxonomic assignment of the features against a reference database. * Primer trimming between any of these steps can be applied. *1 Only for paired-end data (may be performed before or after quality filtering). *2 Error correction; formation of ASVs. *3 Including chimera filtering, off-target gene removal (pseudogene removal, ITS extraction). *4 Formation of OTUs/swarm-clusters.

Hakimzadeh et al., 2023: A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses [18]

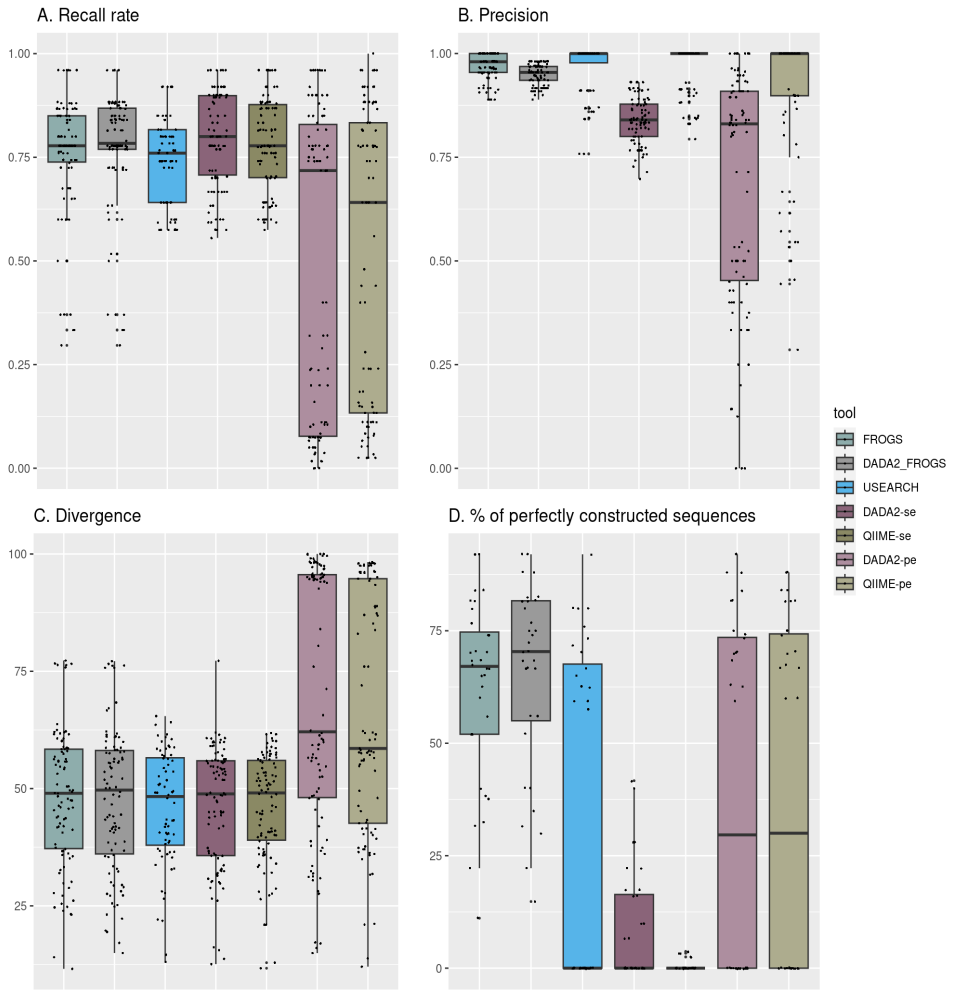
Benchmarking



Compositions at the phylum level for Human gut and, using a range of different methods (separate subpanels within each group).

Liu et al., 2008: Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers [19]

Benchmarking

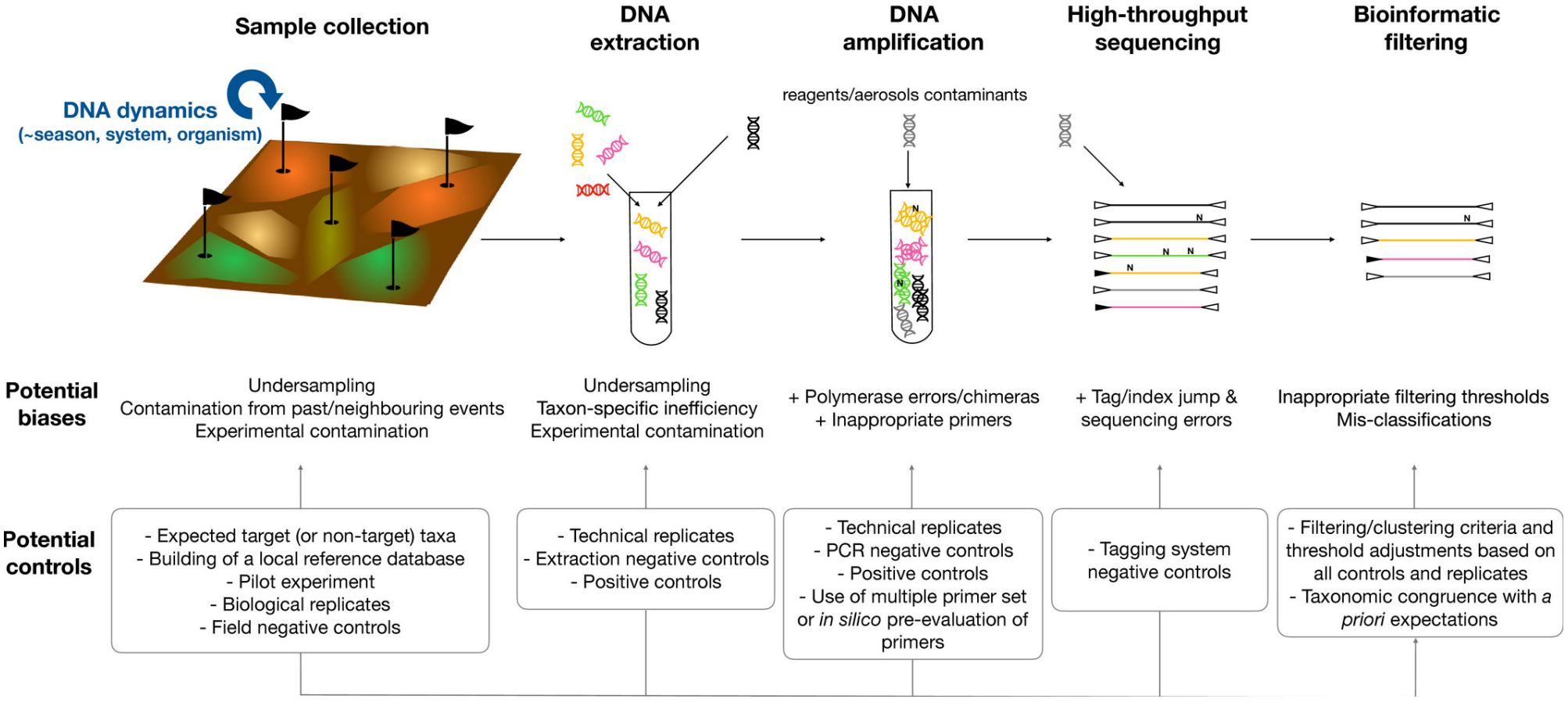


Quality parameters obtained with the seven bioinformatics pipelines. A) Recall rate (TP/(TP+FN)) reflects the capacity of the tools to detect expected species. B) Precision (TP/(TP+FP)) shows the fraction of relevant species among the retrieved species. C) Divergence rate is the Bray-Curtis distance between expected and observed species abundance. D. Percentage of perfectly reconstructed sequences is the fraction of predicted sequences with 100% of identity with the expected ones.

Rué et al., 2023: Comparison of metabarcoding taxonomic markers to describe fungal communities in fermented foods [20]

Conclusion 1: sequencing data do not contain exactly what you sampled…

Summary



Conclusion 2: … but you now know how to deal with

Key advices

- Discuss with all partners (bioinformaticians & statisticians) involved in the project
  - scientific aspects
  - financial aspects
- Use controls!
- If possible, perform a preliminary analysis

References

1. Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein & Cell. 2020;12:315–30. doi:10.1007/s13238-020-00724-8.

2. Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, et al. Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. International Journal of Systematic and Evolutionary Microbiology. 2012;62:716–21. doi:https://doi.org/10.1099/ijs.0.038075-0.

3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nature microbiology. 2016;1:1–6.

4. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Reviews Microbiology. 2014;12:635–45.

5. Espejo RT, Plaza N. Multiple ribosomal RNA operons in bacteria; their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA. Frontiers in microbiology. 2018;9:1232.

6. Maeda M, Shimada T, Ishihama A. Strength and regulation of seven rRNA promoters in escherichia coli. PLoS One. 2015;10:e0144697.

7. Poirier OAP Simon AND Rué. Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing. PLOS ONE. 2018;13:1–26. doi:10.1371/journal.pon

8. Bernard M, Rué O, Mariadassou M, Pascal G. FROGS: a powerful tool to an special management of internal transcribed spacers. Briefings in Bioinform

doi:10.1093/bib/bbab318.
9. Lofgren LA, Uehling JK, Branco S, Bruns TD, Martin F, Kennedy PG. Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Molecular Ecology. 2019;28:721–30. doi:https://doi.org/10.1111/mec.14995.
10. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Briefings in Bioinformatics. 2019;22:178–93. doi:10.1093/bib/bbz155.
11. Alard J, Lehrter V, Rhimi M, Mangin I, Peucelle V, Abraham A-L, et al. Beneficial metabolic effects of selected probiotics on diet-induced obesity and insulin resistance in mice are associated with improvement of dysbiotic gut microbiota. Environmental Microbiology. 2016;18:1484–97. doi:https://doi.org/10.1111/1462-2920.13181.
12. Tan YC, Kumar AU, Wong YP, Ling APK. Bioinformatics approaches and applications in plant biotechnology. Journal of Genetic Engineering and Biotechnology. 2022;20:1–13.
13. Cruaud P, Rasplus J-Y, Rodriguez LJ, Cruaud A. High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. Scientific reports. 2017;7:41948.
14. Whon TW, Chung W-H, Lim MY, Song E-J, Kim PS, Hyun D-W, et al. The effects of sequencing platforms on phylogenetic resolution in 16 s rRNA gene profiling of human feces. Scientific data. 2018;5:1–15.
15. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology. 2014;12. doi:10.1186/s12915-014-0087-z.
16. Lejal E, Estrada-Peña A, Marsot M, Cosson J-F, Rué O, Mariadassou M, et al. Taxon appearance from extraction and amplification steps demonstrates the value of multiple controls in tick microbiota analysis. Frontiers in Microbiology. 2020;11:1093.

17. Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. BMC microbiology. 2015;15:1–14.
18. Hakimzadeh A, Abdala Asbun A, Albanese D, Bernard M, Buchner D, Callahan B, et al. A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. Molecular Ecology Resources. 2023.
19. Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic acids research. 2008;36:e120–0.
20. Rué O, Coton M, Dugat-Bony E, Howell K, Irlinger F, Legras J-L, et al. Comparison of metabarcoding taxonomic markers to describe fungal communities in fermented foods. bioRxiv. 2023. doi:10.1101/2023.01.13.523754.