THE 4th VIETNAM SCHOOL OF BIOLOGY (VSOB-4)

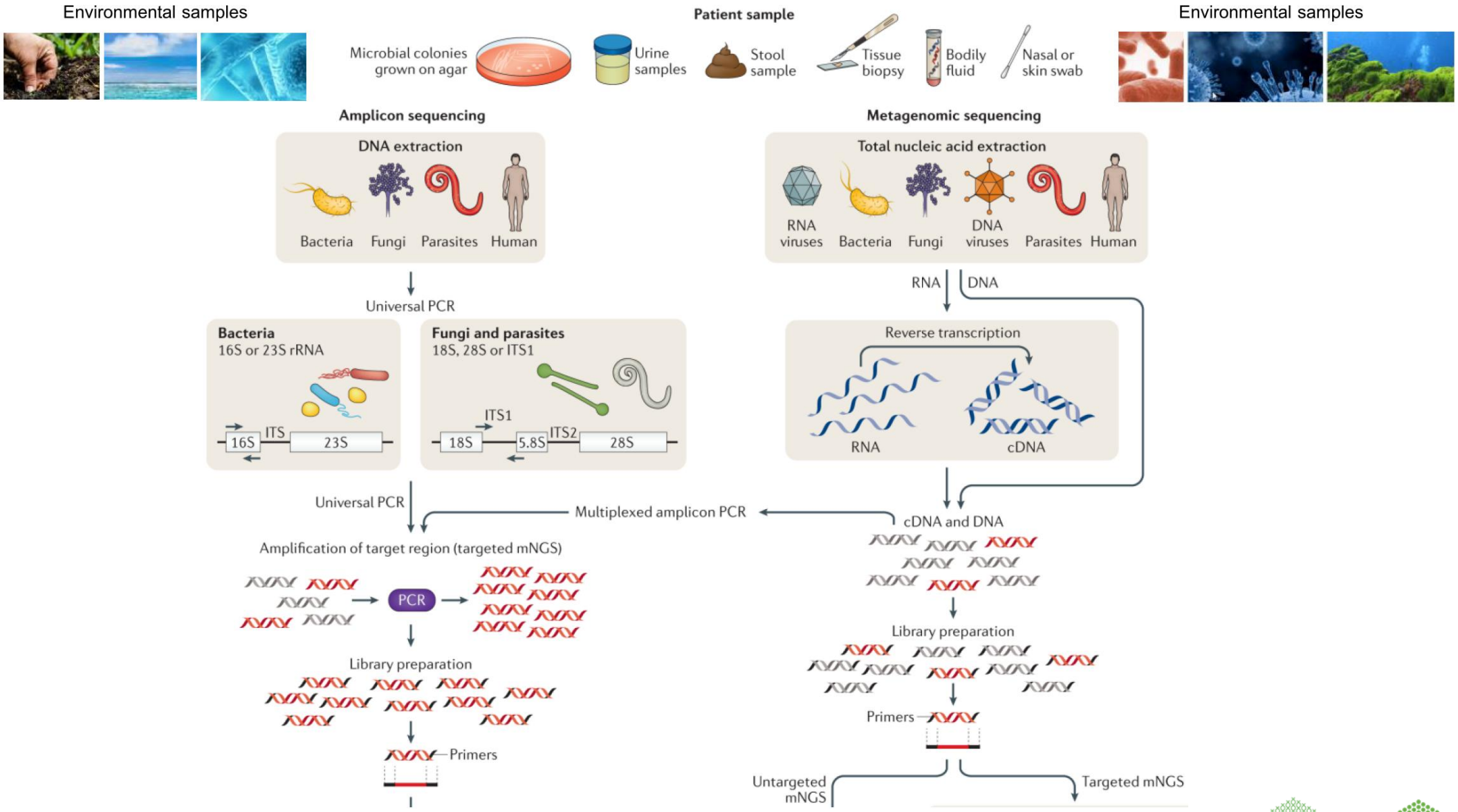
Metagenome Analysis in One Health Practice – From 16S to Shotgun

September 03rd-06th, 2025, ICISE, Quy Nhon, Vietnam

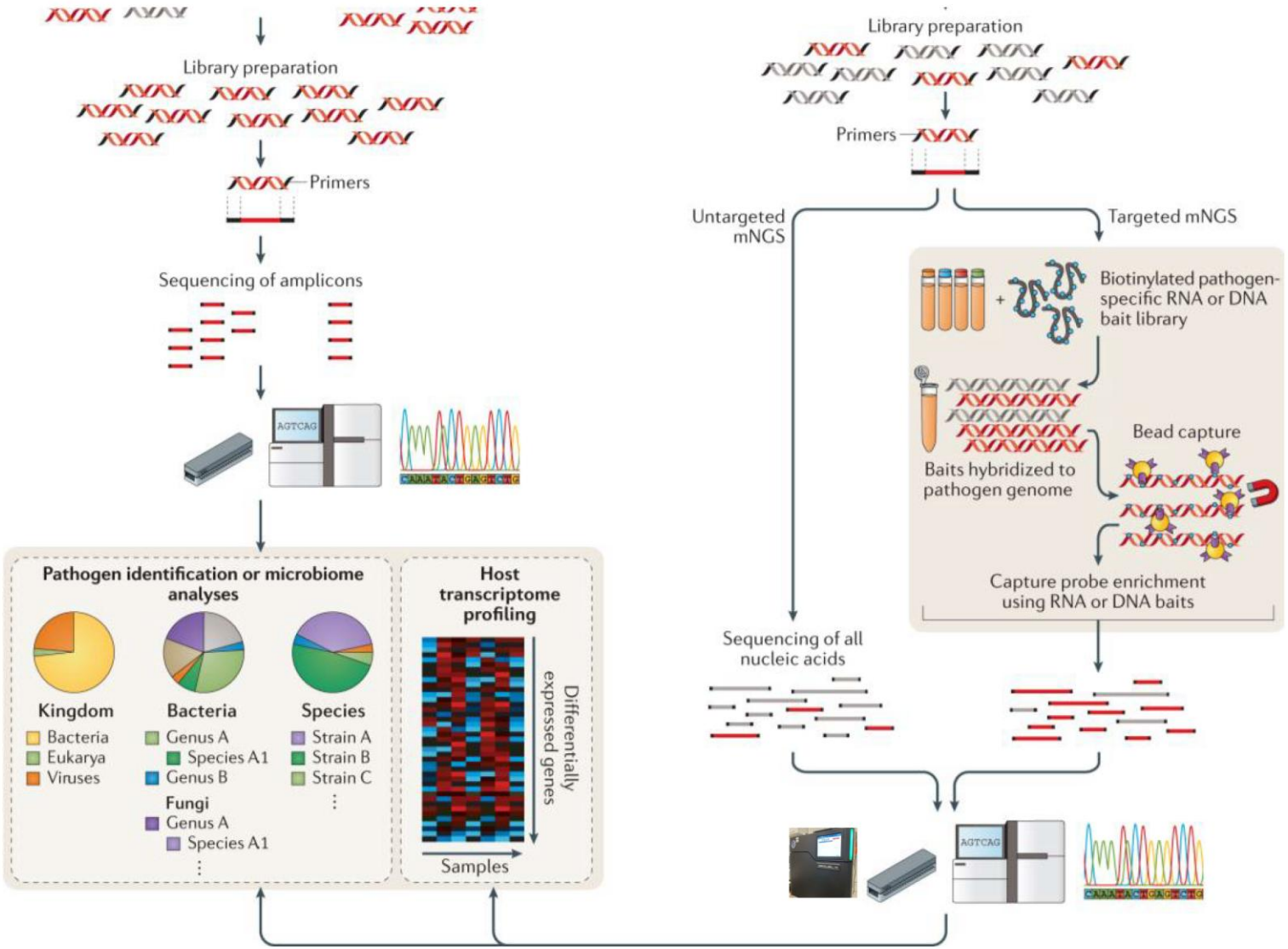
Lecture 1
Metagenomics library preparation

Lecturer: Thuy Vy Nguyen, PhD

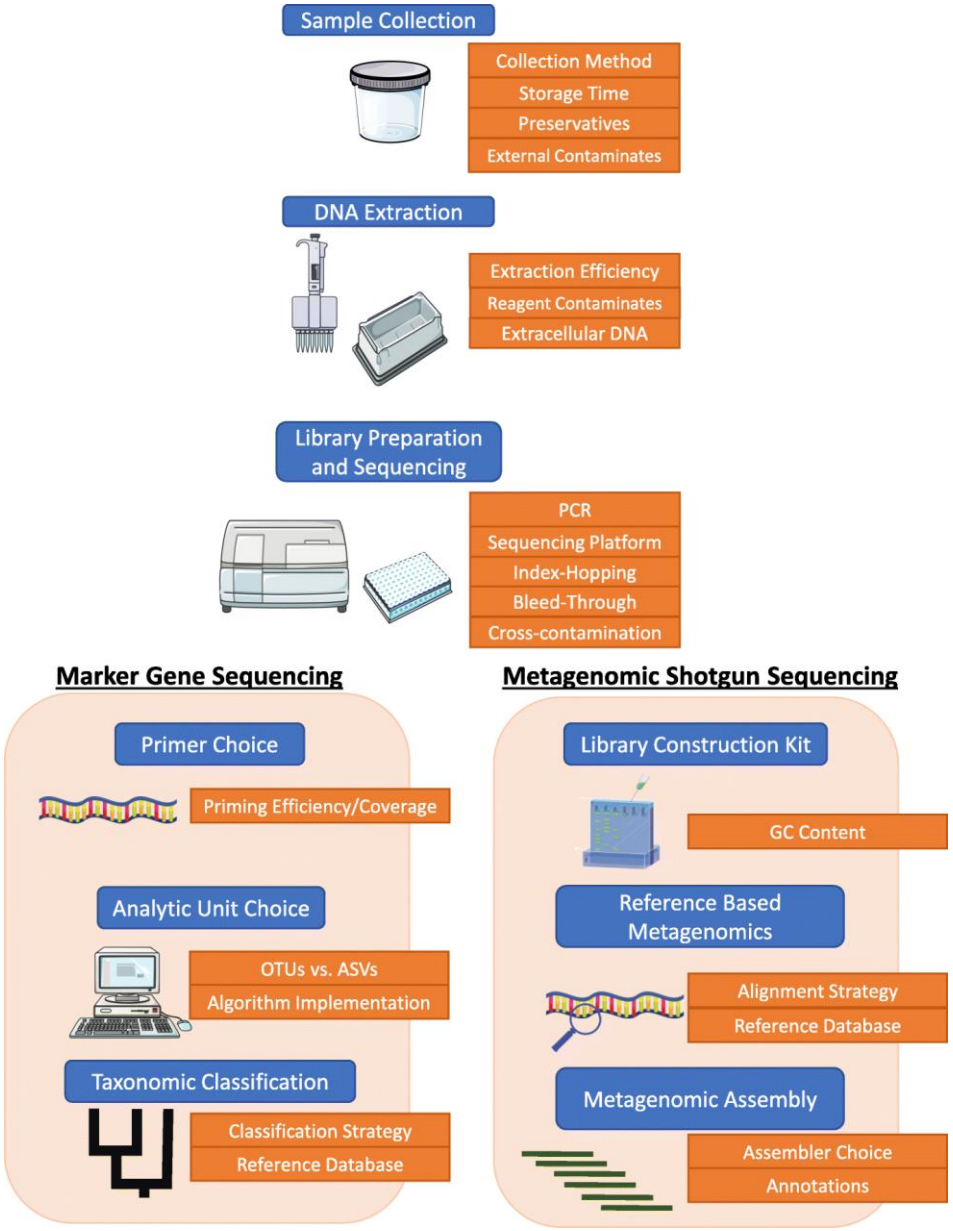
METAGENOMICS WORKFLOWS



METAGENOMICS WORKFLOWS



BUT BIAS IS BEHIND EVERY STEP…



Nearing at al., Microbiome (2021)
doi: 10.1186/s40168-021-01059-0

EXPERIMENTAL DESIGN, DECISIONS…

Points to consider

- Background information on the studied system (*e.g.*, genome size)
- Sample choice and its homogeneity, contamination
- Complexity of a “sample”, its repertoire, incl. base composition
- Proper sampling
- Biological replication, randomization, controls
- Sample harvesting, DNA/RNA isolation, library preparation
- Read-type, read-length, read-depth/coverage
- Data QC, analysis, corroboration & validation
- *What comes next?*

FROM SAMPLE TO GOOD DATA: HOW TO GET THERE

- Preparation of sequencing library is still required
- *A sequencing library should fully encompass complexity of the 'sample'*
- There are many factors influencing/compromising it
- And people don't always realize this
- An ideal: no library preparation at all
- Reality: direct sequencing of nucleic-acid not yet possible, for some applications and nucleic-acid molecules of interest it may never be…

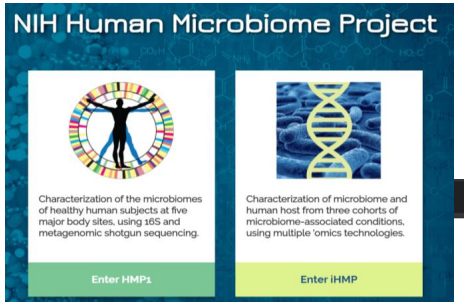
EXPERIMENTAL DESIGN

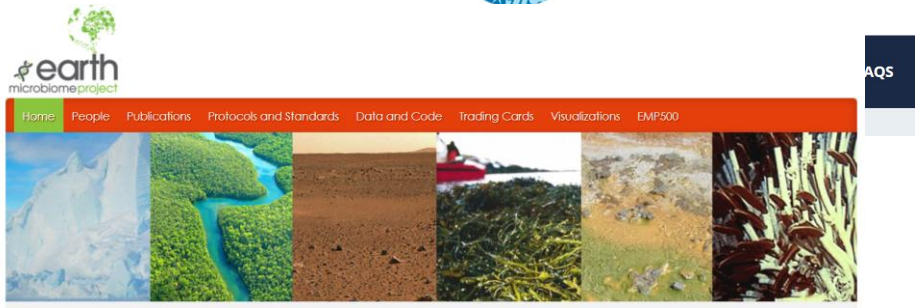
- Previous studies
- Pilot studies allow for estimating the number of taxa, the relative abundance of each taxa -> sample size is calculated
- 2-step studies: 16S screening -> shotgun metagenomics.

(Quince et al., 2017; Goodrich et al., 2014; Tickle et al., 2013; La Rosa et al., 2012)

SAMPLE COLLECTION & PROCESSING

- Previous studies
- Refer to some international metagenomics projects





SAMPLE COLLECTION, PRESERVATION, AND TRANSPORTATION

347 Table 1: Average dissimilarities and dispersion (Standard error) of stored samples relative to samples that
348 were not stored (i.e. DNA isolation immediately upon sample collection (0h)). The entries for 0h represented
349 the variation among replicates.

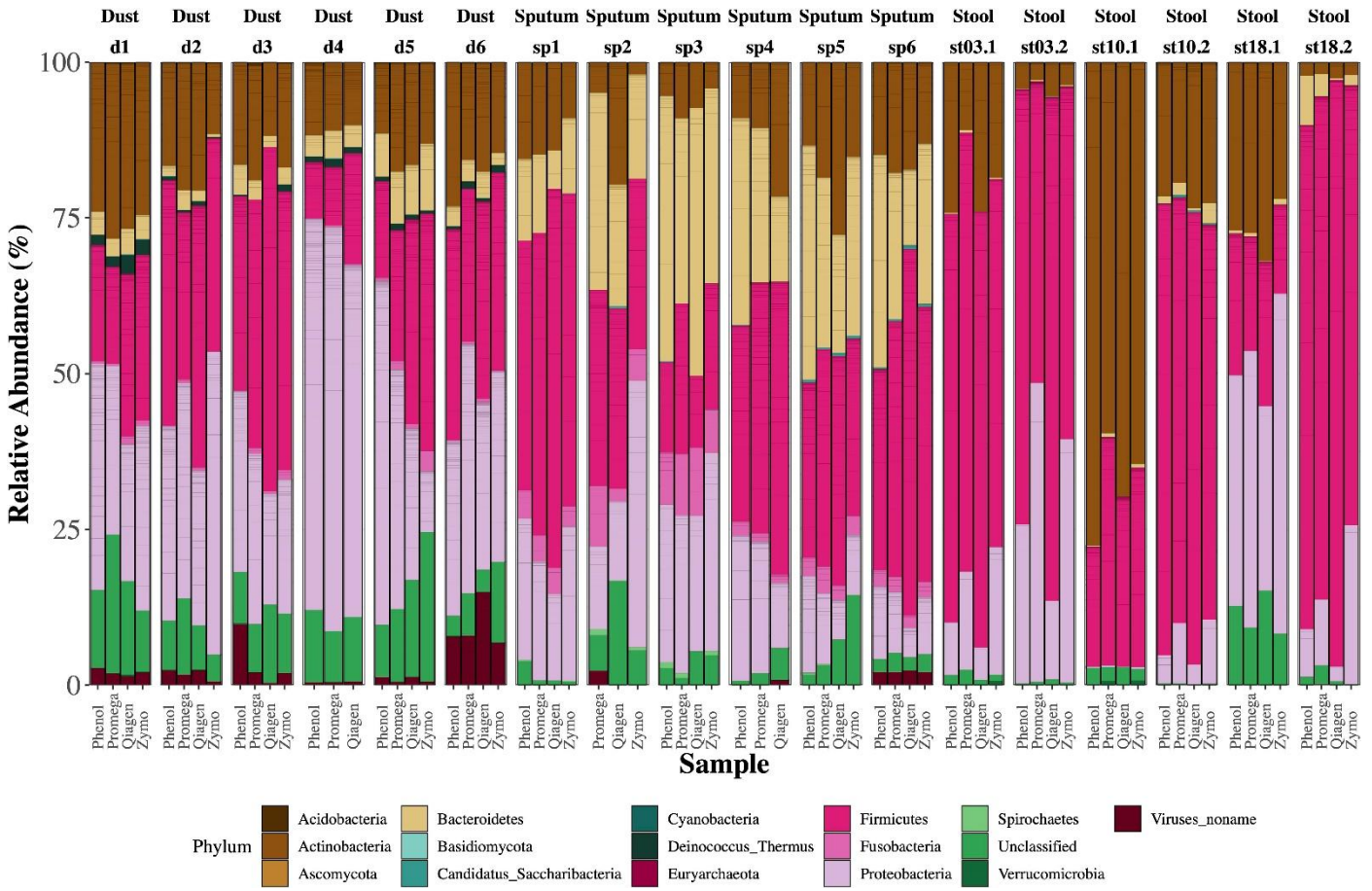
	Pig feces 1	Pig feces 2	Sewage 1	Sewage 2
0h	0,066 (0,003)	0,081 (0,003)	0,073 (0,003)	0,057 (0,003)
16h -80 °C	**0,159 (0,005)**	**0,116 (0,003)**	**0,103 (0,002)**	0,077 (0,002)
64h -80 °C	**0,152 (0,005)**	**0,113 (0,002)**	**0,121 (0,004)**	0,076 (0,001)
16h -20 °C	**0,141 (0,004)**	**0,124 (0,003)**	**0,109 (0,002)**	0,082 (0,003)
64h -20 °C	**0,139 (0,004)**	**0,126 (0,002)**	**0,109 (0,002)**	0,088 (0,002)
16h 5 °C	0,073 (0,003)	0,098 (0,004)	0,089 (0,008)	0,061 (0,003)
64h 5 °C	0,098 (0,002)	**0,137 (0,004)**	0,091 (0,003)	0,073 (0,001)
16h 22 °C	0,096 (0,002)	**0,130 (0,004)**	0,087 (0,002)	0,082 (0,002)
64h 22 °C	**0,142 (0,002)**	**0,180 (0,003)**	**0,127 (0,002)**	**0,126 (0,001)**

350

(From PhD thesis of Casper Sahl Poulsen, DTU, Denmark)

DNA EXTRACTION

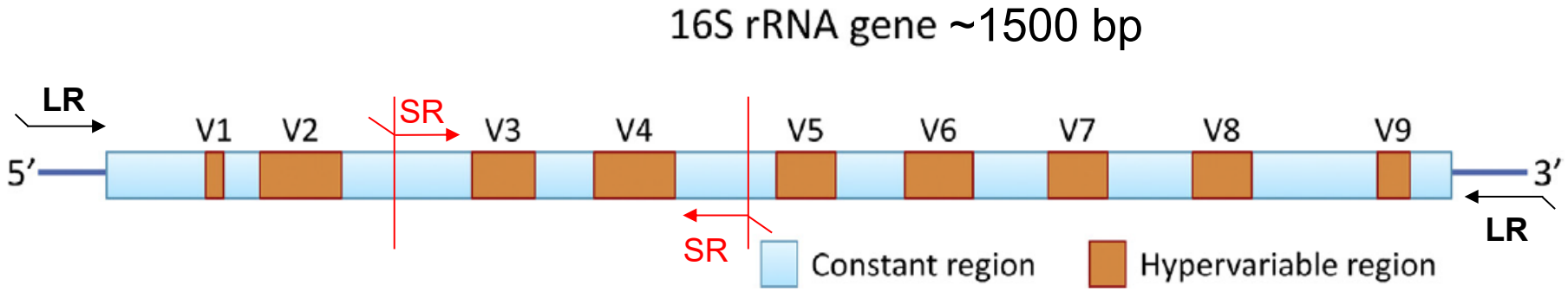
Lysation and DNA extraction, many methods available, different biases



Use the same extraction kit consistently throughout the project

Sui Hui-yu et al,, 2020, Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing

16S AMPLICON SEQUENCING



OPEN ACCESS Freely available online
PLoS one

Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing

Purnima S. Kumar[1]*, Michael R. Brooker[1], Scot E. Dowd[2], Terry Camerlengo[3]

- V1-V3 favours *Prevotella, Fusobacterium, Streptococcus, Granulicatella, Bacteroides, Porphyromonas* and *Treponema*
- V4-V6 favours *Streptococcus, Treponema, Prevotella, Eubacterium, Porphyromonas, Campylobacter* and *Enterococcus*
  - failed to detect *Selenomonads*, and *Mycoplasma*
- V7-V9 favours *Veillonella, Streptococcus, Eubacterium, Enterococcus, Treponema, Catonella* and *Selenomonas*
  - failed to detect *Fusobacteria*

16S AMPLICON SEQUENCING

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE

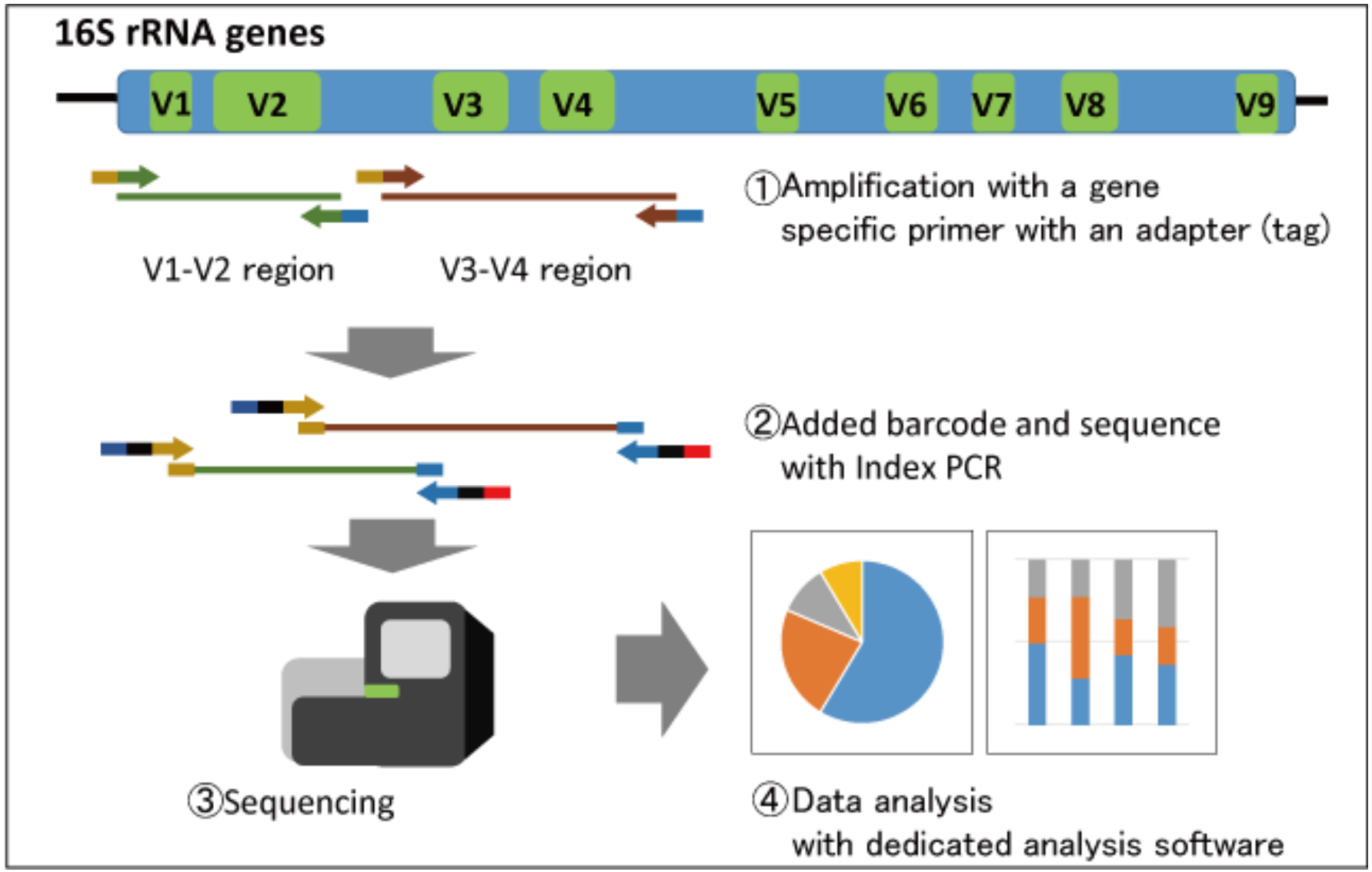
https://doi.org/10.1128/mSphere.01202-20

Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing

Isabel Abellan-Schneyder,[a] Monica S. Matchado,[b] Sandra Reitmeier,[a] Alina Sommer,[a] Zeno Sewald,[a] Jan Baumbach,[b,c,d] Markus List,[b] Klaus Neuhaus[a]

The primer choice has a significant influence on the resulting microbial composition; we show that microbial profiles generated using different primer pairs need independent validation of performance.

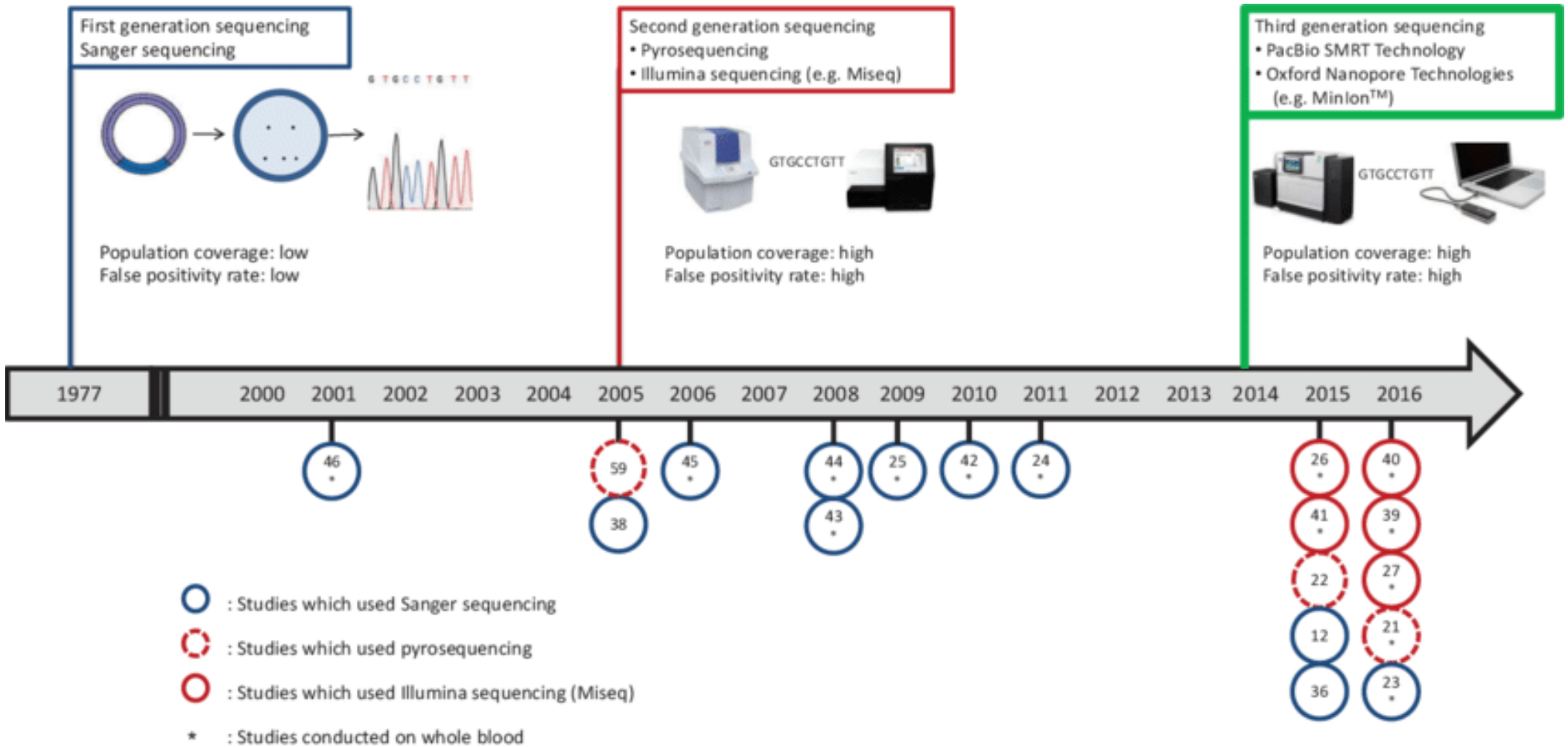
16S AMPLICON SEQUENCING



https://sourdoughstartersfall20.wordpress.ncsu.edu/2020/10/09/so-lets-talk-about-sequencing-our-starters/

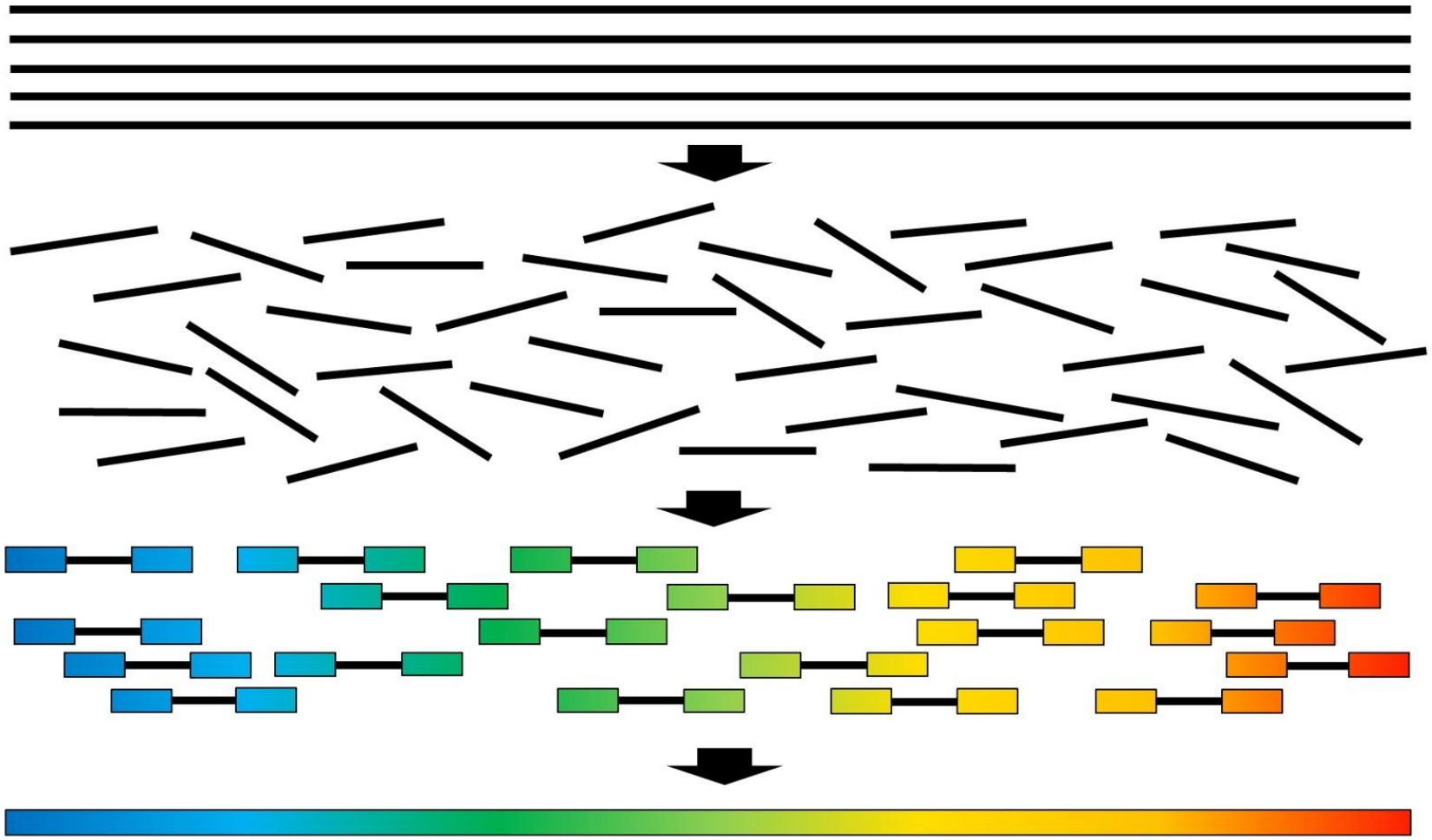
METAGENOMICS - NUCLEIC-ACID SEQUENCING

SEQ GENERATION **GEN I** **GEN II** **GEN III**



Rutanga JP et al. 16S metagenomics for diagnosis of bloodstream infections: opportunities and pitfalls. 2018

SHOTGUN METAGENOMICS



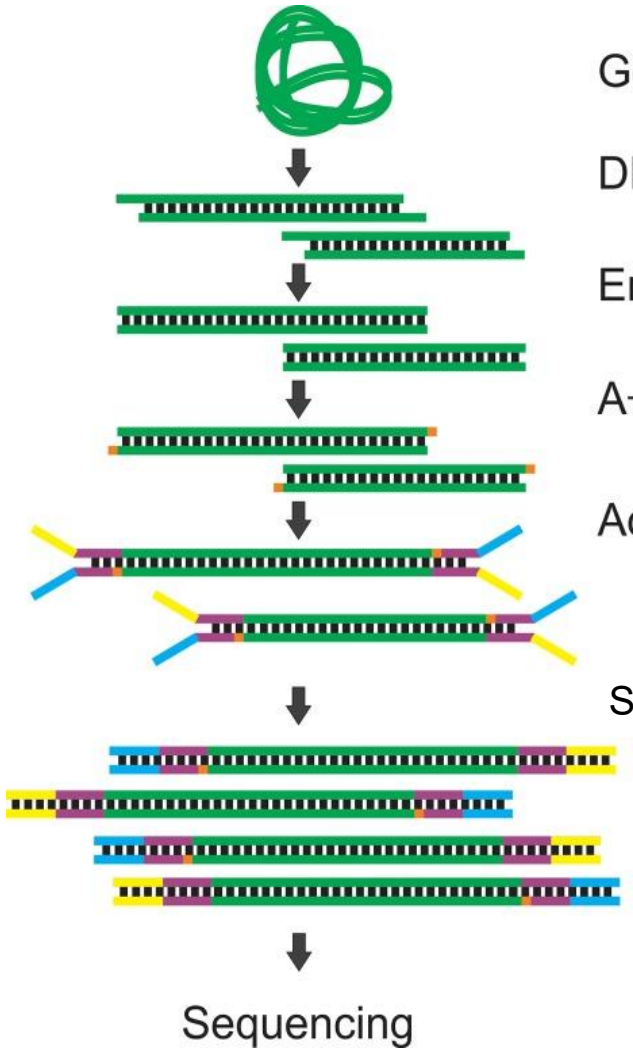
Multiple copies of un-sequenced genome

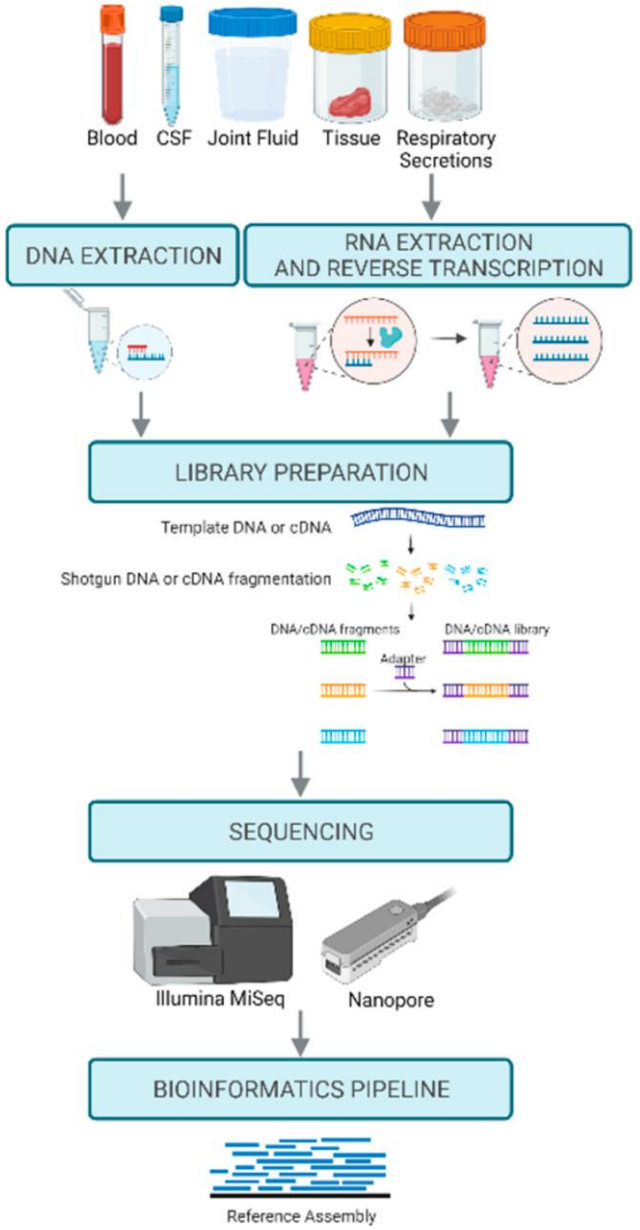
"shotgunned" un-sequenced fragments

Fragment ends sequenced and aligned

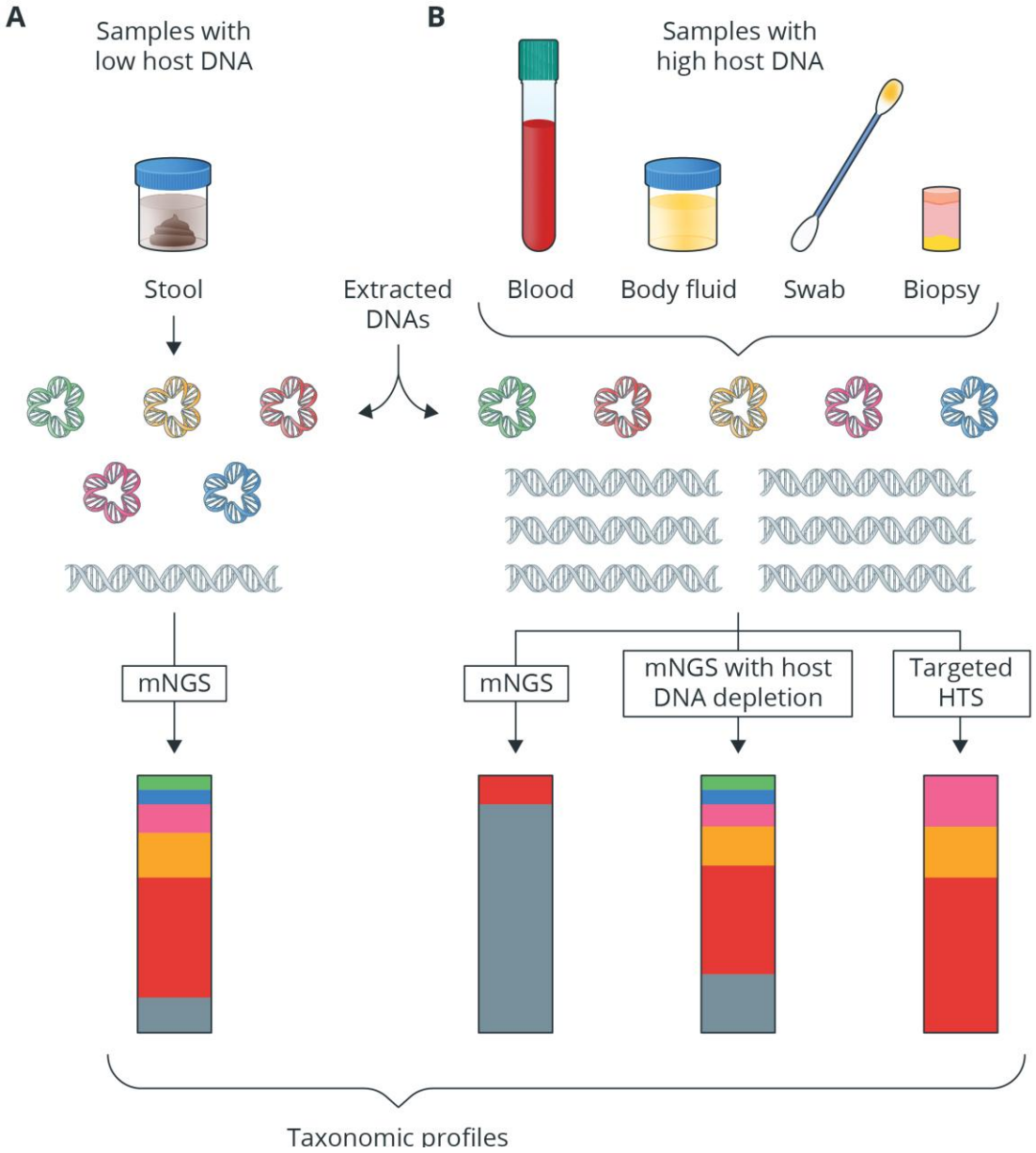
Reassembled sequenced genome

SHOTGUN METAGENOMICS



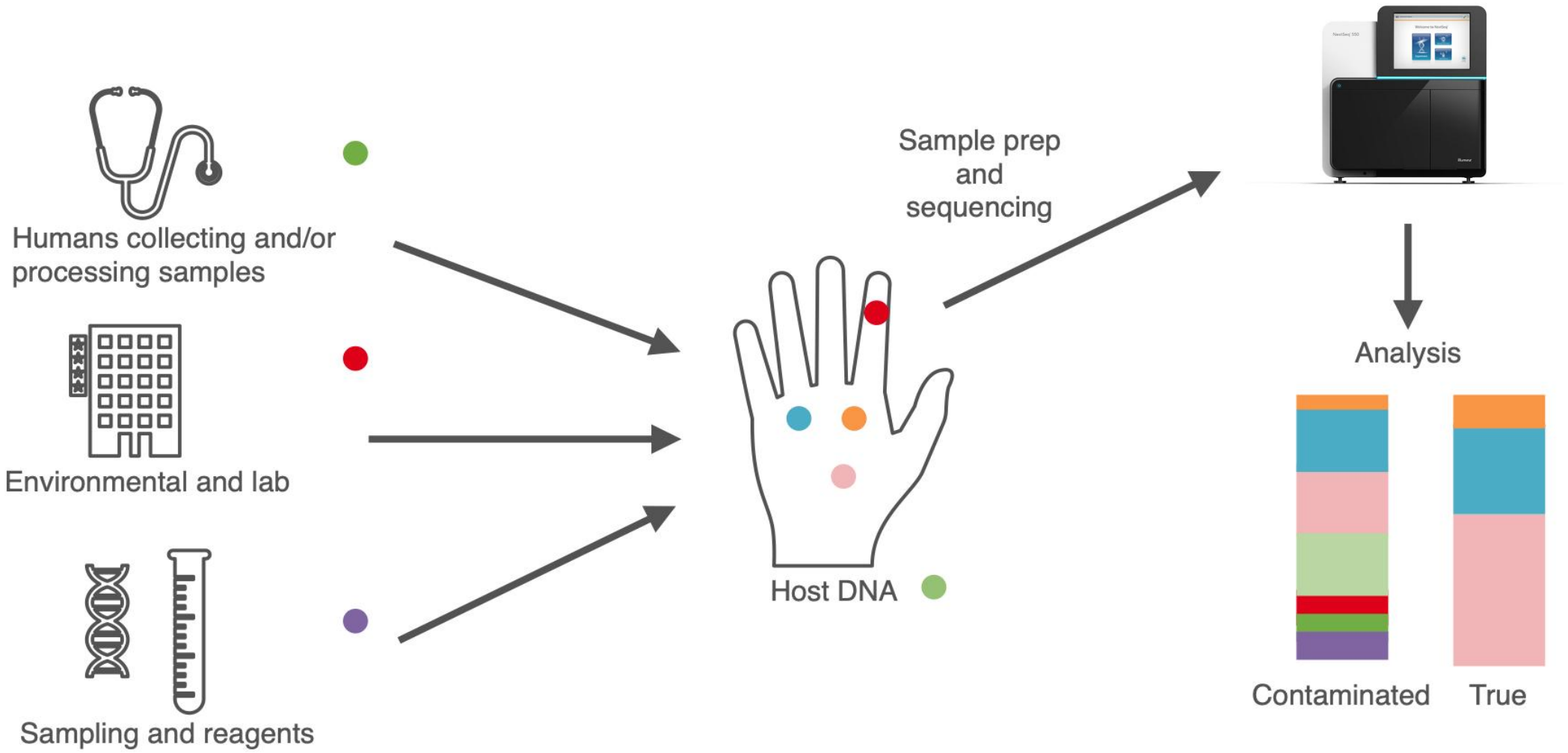


SHOTGUN METAGENOMICS

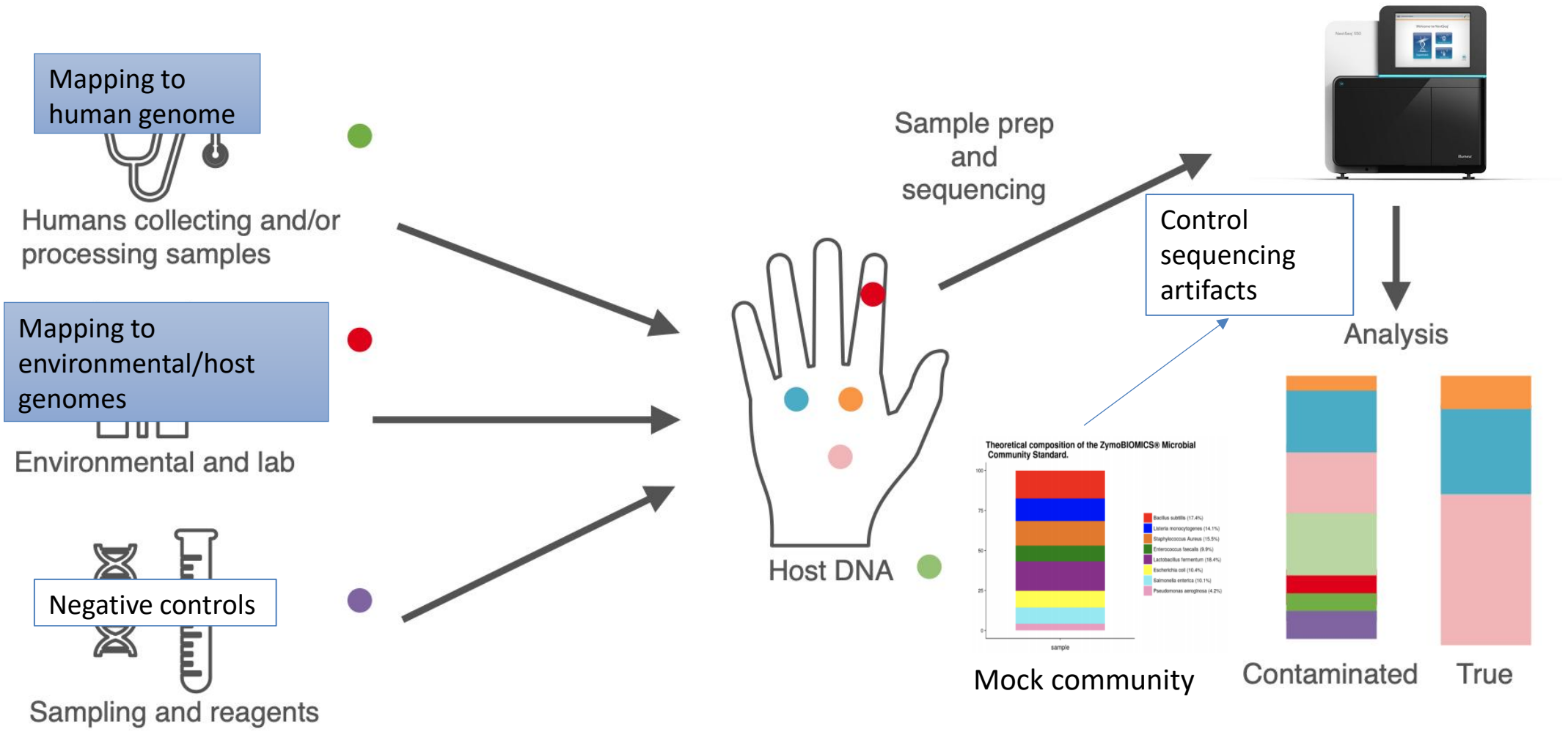


https://www.novogene.com/amea-en/resources/blog/metagenomics-host-dna-removal-or-not/

SOURCES OF CONTAMINATION

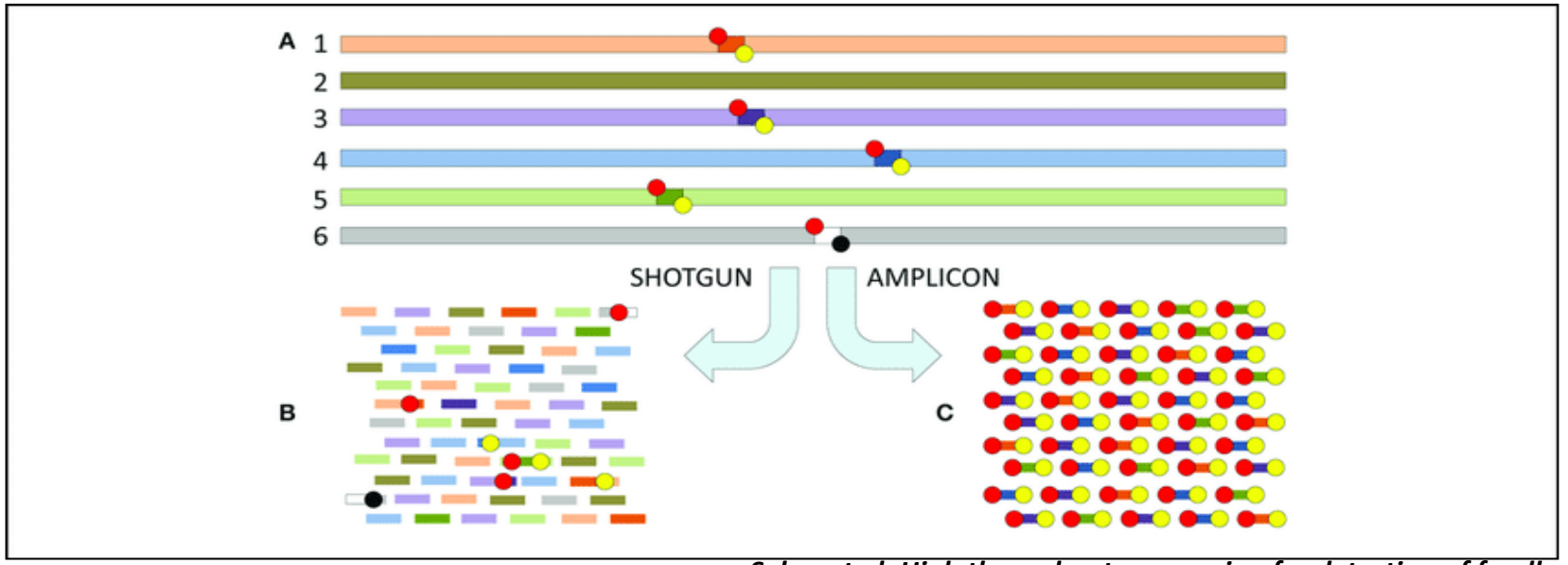


SOURCES OF CONTAMINATION



METAGENOMICS

2 APPROACHES



Sekse et al. High throughput sequencing for detection of foodborne pathogens. 2017

SHOTGUN METAGENOMICS

TAXONOMIC AND FUNCTIONAL STUDIES

- Species- and strain-level identification
- Complete genome assembly of community members
- Novel function discovery; draft genomes of uncultivable species
- Functional studies

METABARCODING

TAXONOMIC STUDIES

- 16S rDNA (bacteria), 18S rDNA (eukaryotes), ITS (fungi)
- Not applicable to viruses
- PCR bias may affect results
- Resolution mostly at genus level
- Suitable for low-biomass, host DNA–contaminated samples (skin, lung, …)
- Low cost, fast, simple

CONCLUSION

- Metagenomics research is really, incredibly exciting… but….
- ***Every*** step counts
- Be very careful & consistent, at all stages
- 16S – cheap, biased but effective
- WGS –information rich, less biased
- Beware contamination, include controls

Limitations of metagenomics sequencing. EBI course "Metagenomics_Bioinformatics" Jun 2019.