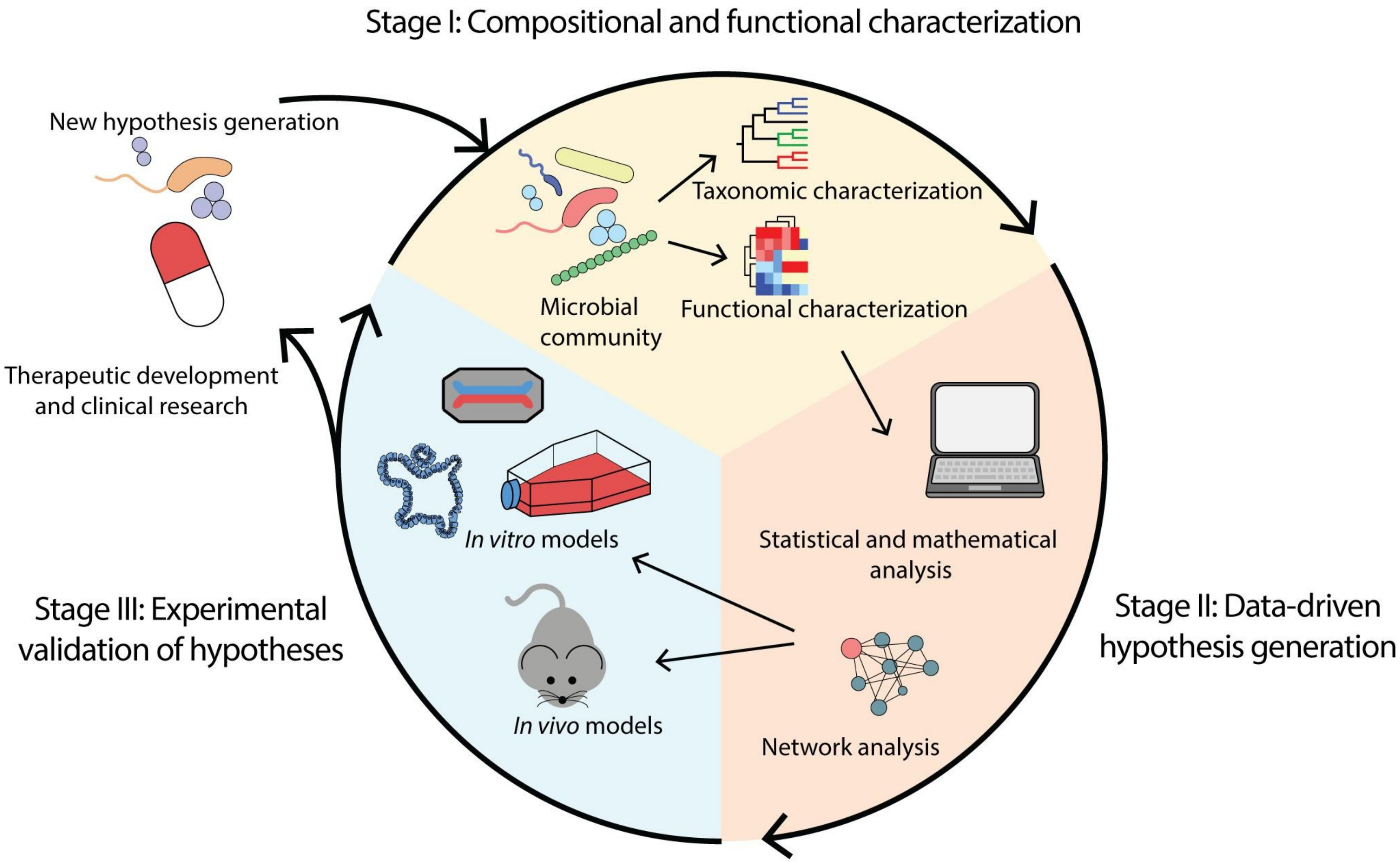
Microbiome experiment design

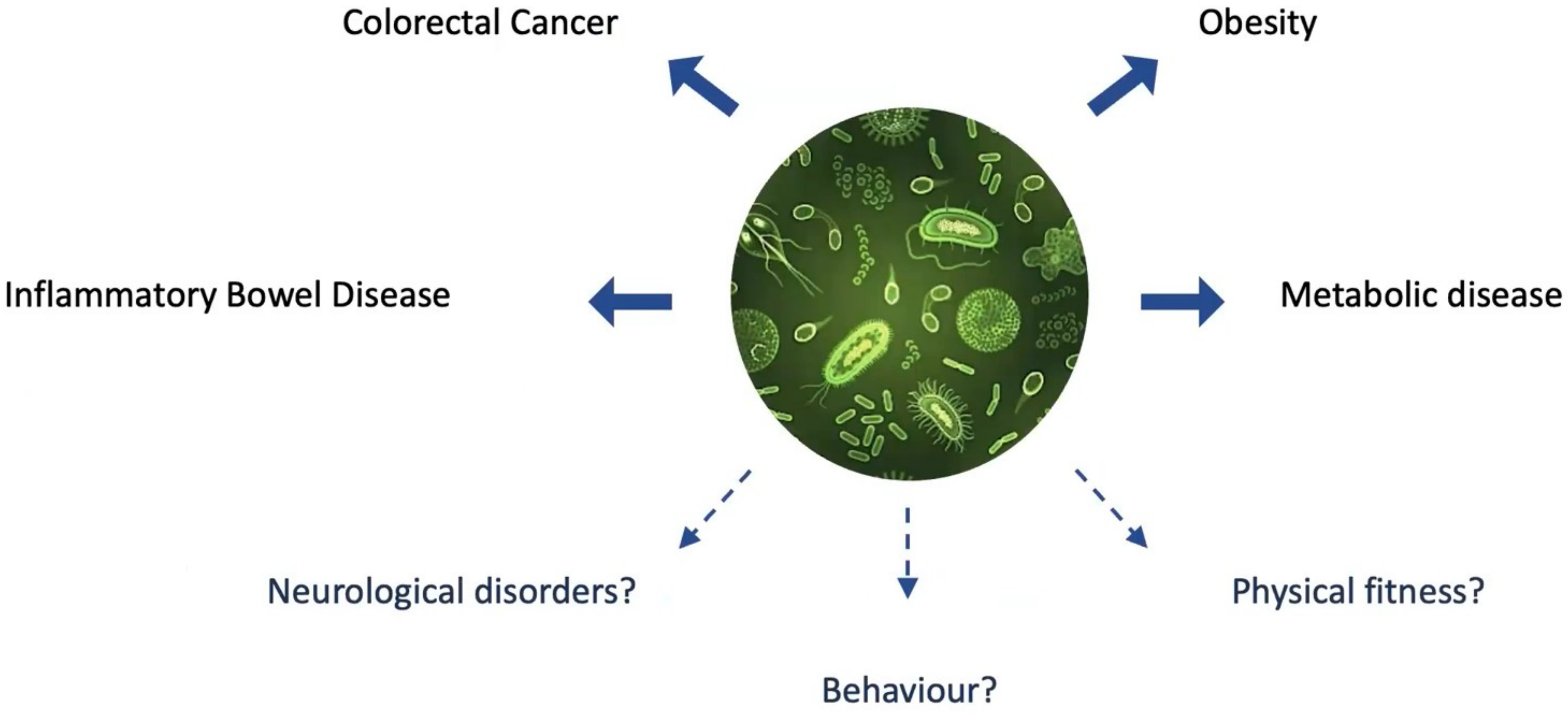
Content

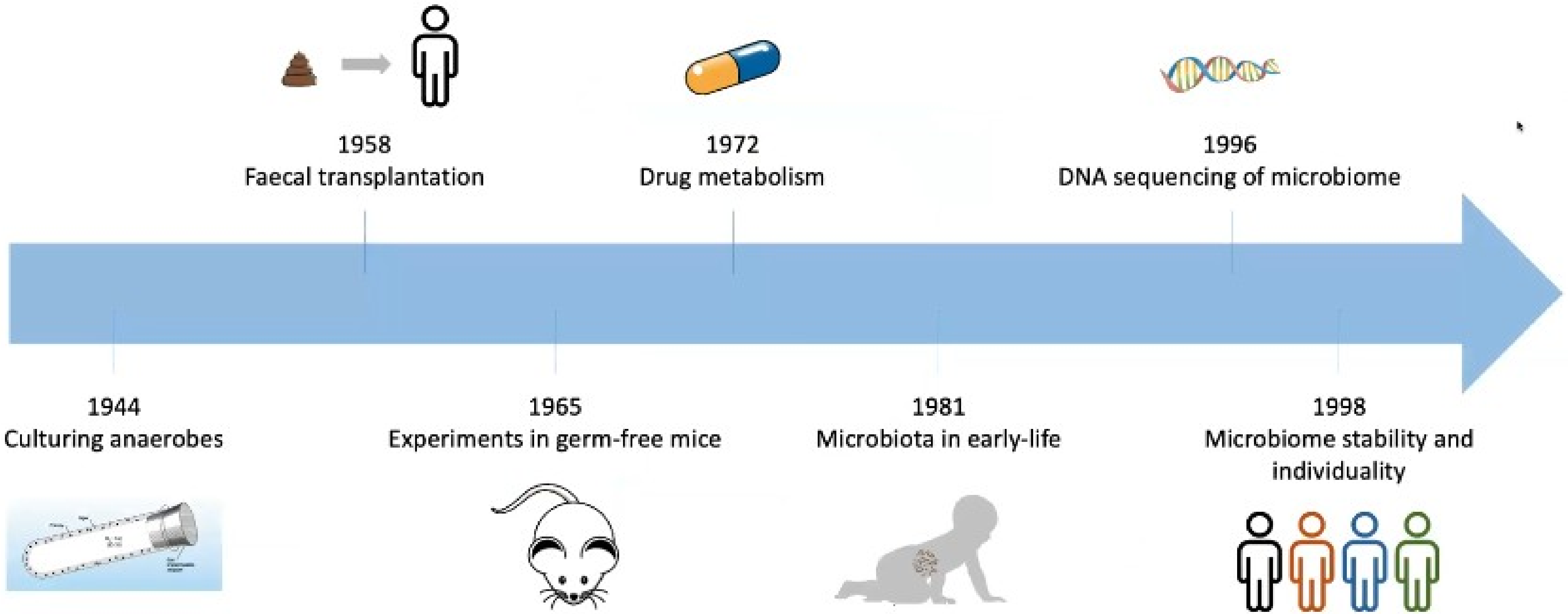
- Introduction
- the microbiome in translational medicine
- Making informed decisions: study design
- Measuring the microbiome: taxonomy and function
- Statistical considerations in microbiome analysis: a roadmap
- Sample study – Freshwater monitoring by nanopore sequencing

Introduction



the microbiome in translational medicine

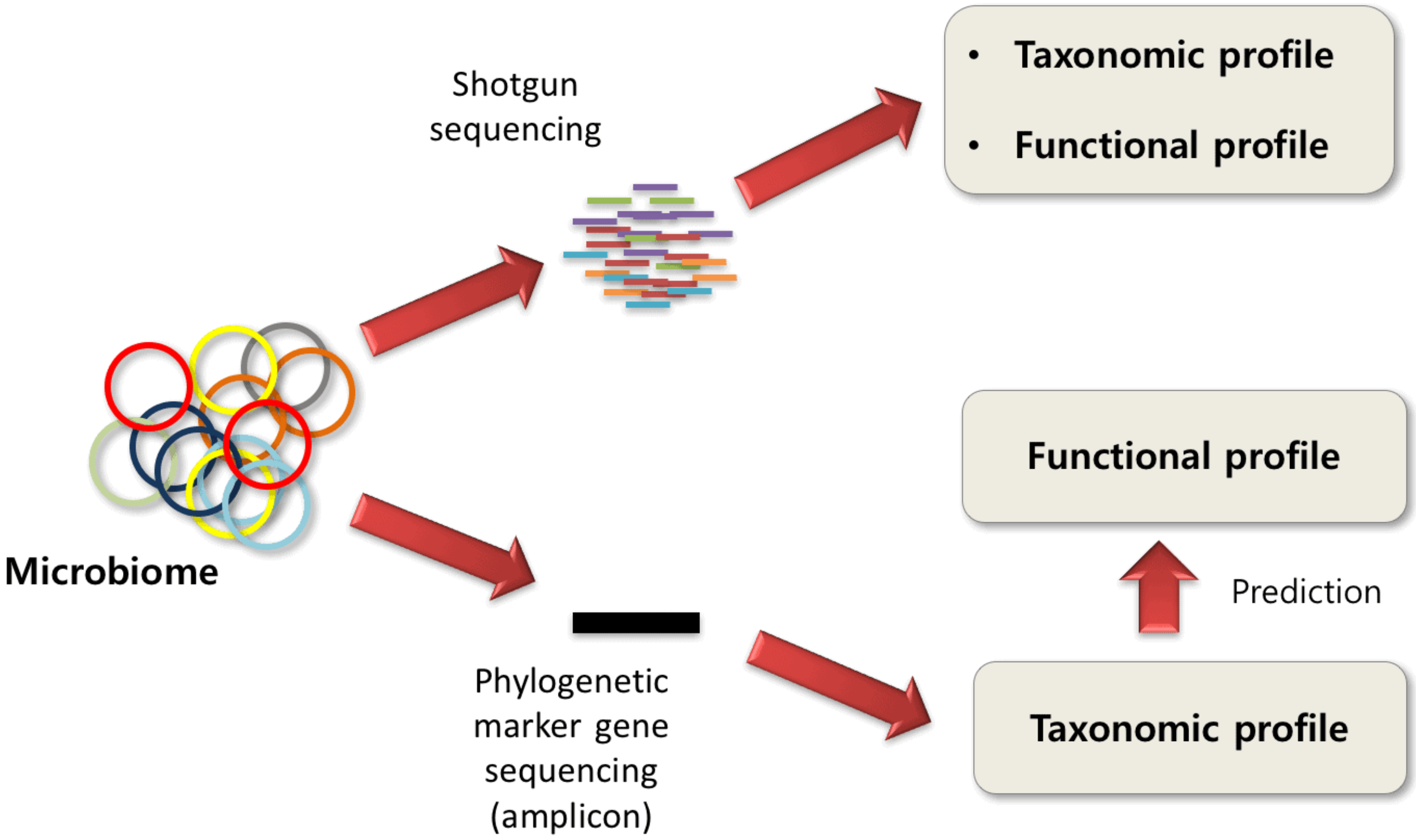



1958
Faecal transplantation
1972
Drug metabolism
1996
DNA sequencing of microbiome
1944
Culturing anaerobes
1965
Experiments in germ-free mice
1981
Microbiota in early-life
1998
Microbiome stability and individuality

the microbiome in translational medicine

- The microbiome's potential roles in medical practice include:
  - serving as a diagnostic tool
  - determining treatment responses
  - monitoring disease progression
  - guiding therapeutic interventions.
  - offers insights into antibiotic side effects, immune processes, and small-molecule drug identification.

Measuring the microbiome: taxonomy and function

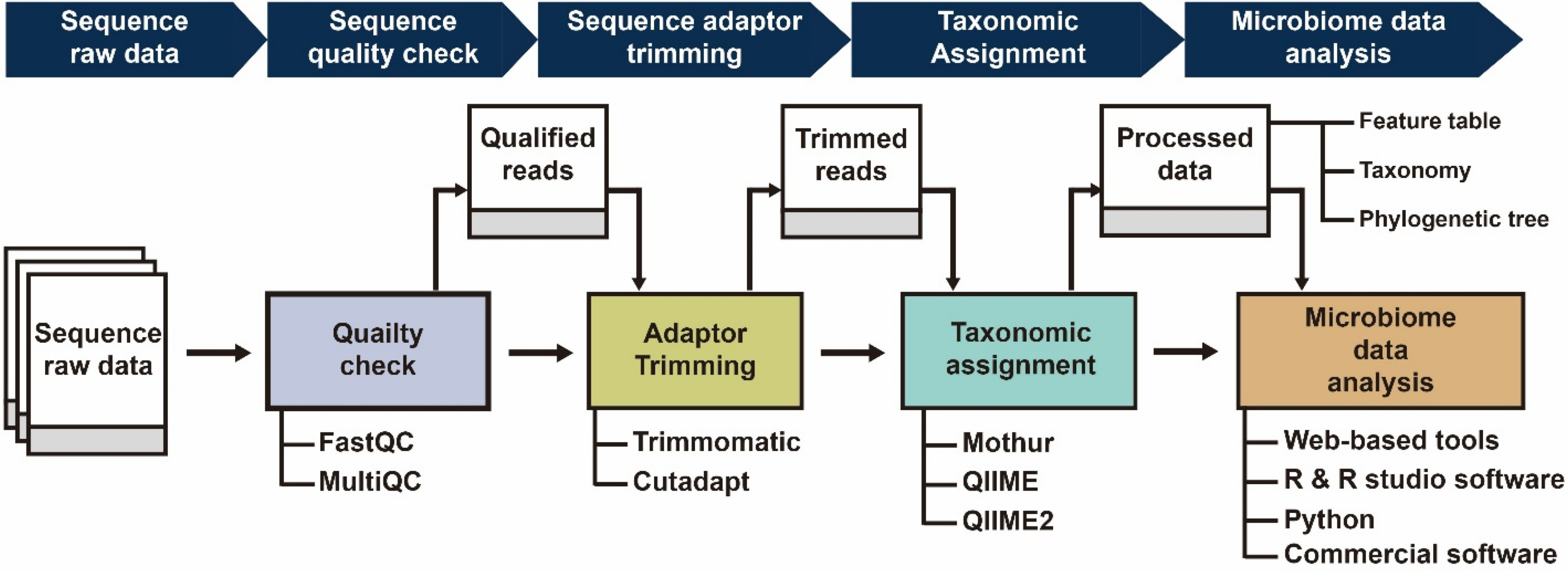


Measuring the microbiome: taxonomy and function

Bacterial and Fungal Microbiomes	Metagenomic Sequencing
Methodology: Typically studied via sequencing of conserved genes like 16S rRNA (bacteria) and ITS (fungi).	**Approach:** Sequencing entire microbial genomes (shotgun sequencing) to achieve species and strain-level identification and functional profiling
Advantages: Cost-effective for large-scale studies but limited in providing detailed strain-level information crucial for understanding microbial interactions in diseases	**Advantages:** Offers detailed insights into microbial diversity, functional attributes, and genetic variations.
Limitations: These methods provide taxonomic information but do not indicate whether organisms are live, their metabolic states, or detailed functional profiles	**Challenges:** High cost and computational complexity, especially for deeply sequencing multiple samples.

Statistical considerations in microbiome analysis: a roadmap

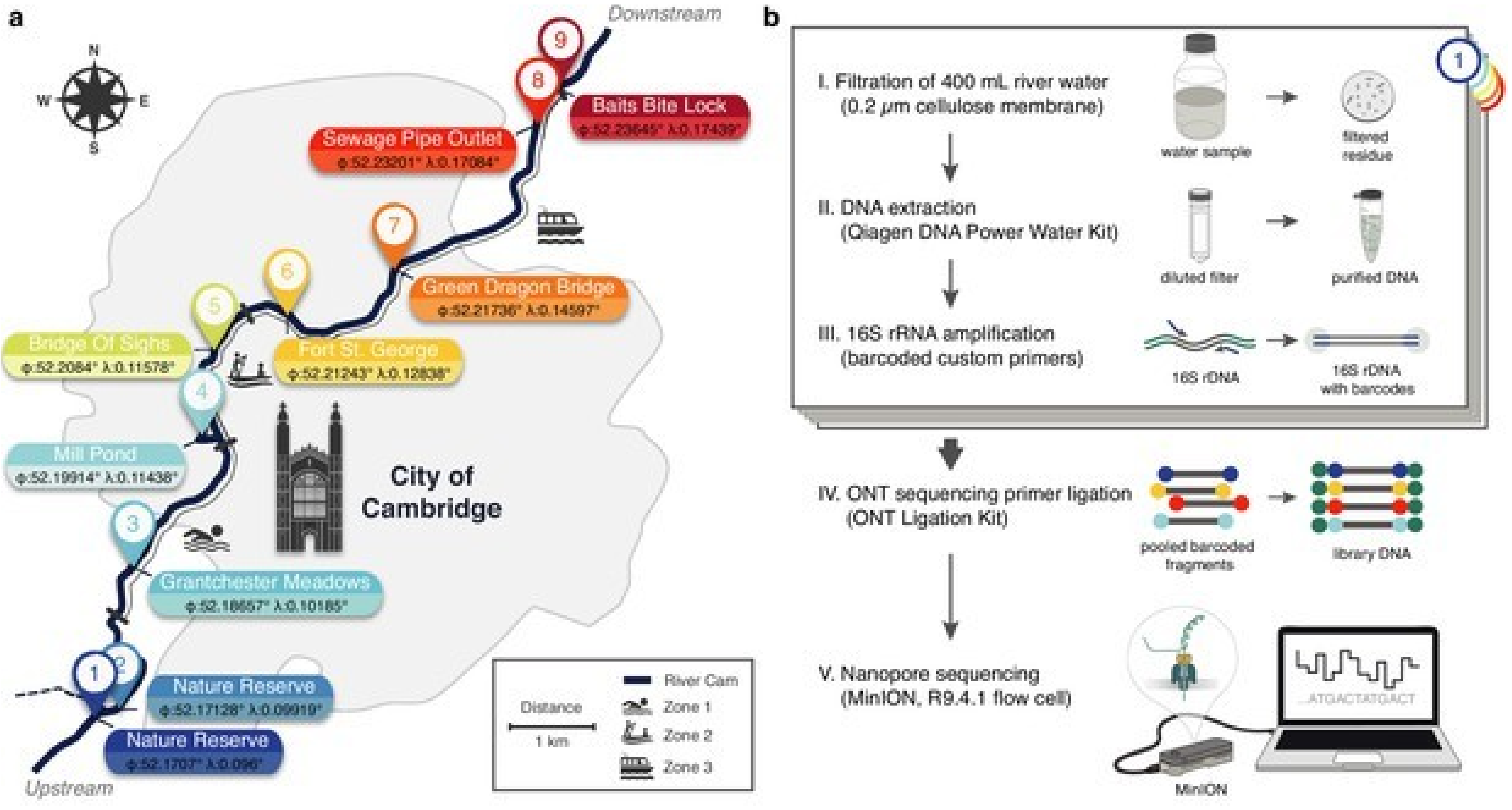
- A variety of downstream statistical analyses are necessary to analyze taxonomic and functional characteristics of microbiome communities that were measured by either 16S or metagenomic profiling. The design and choice of these statistical analyses is closely connected with the research objectives of the study.


Sequence raw data
Sequence quality check
Sequence adaptor trimming
Taxonomic Assignment
Microbiome data analysis
Qualified reads
Trimmed reads
Processed data
Feature table
Taxonomy
Phylogenetic tree
Sequence raw data
Quailty check
Adaptor Trimming
Taxonomic assignment
Microbiome data analysis
FastQC
MultiQC
Trimmomatic
Cutadapt
Mothur
QIIME
QIIME2
Web-based tools
R & R studio software
Python
Commercial software

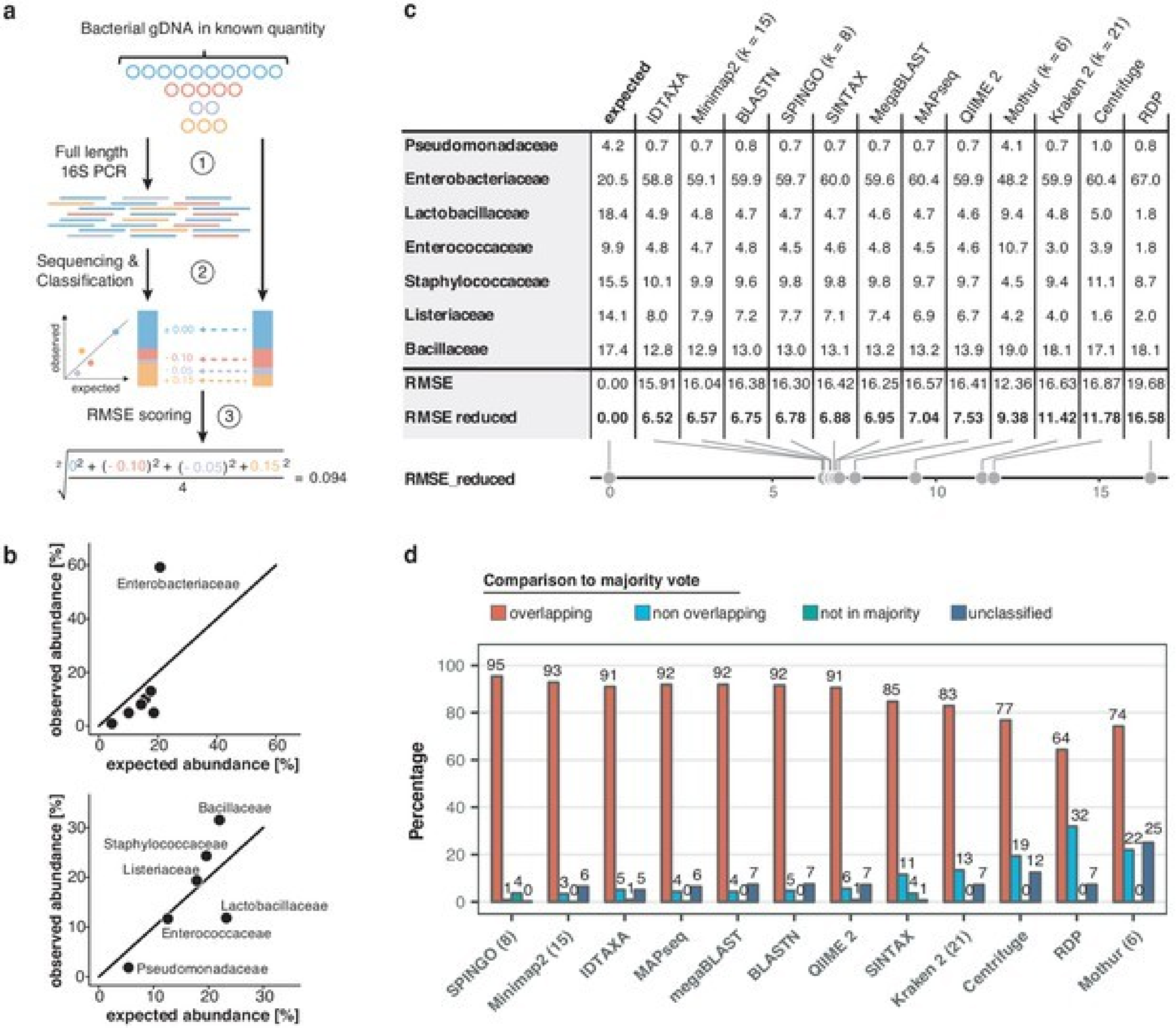
Freshwater monitoring by nanopore sequencing

Abstract

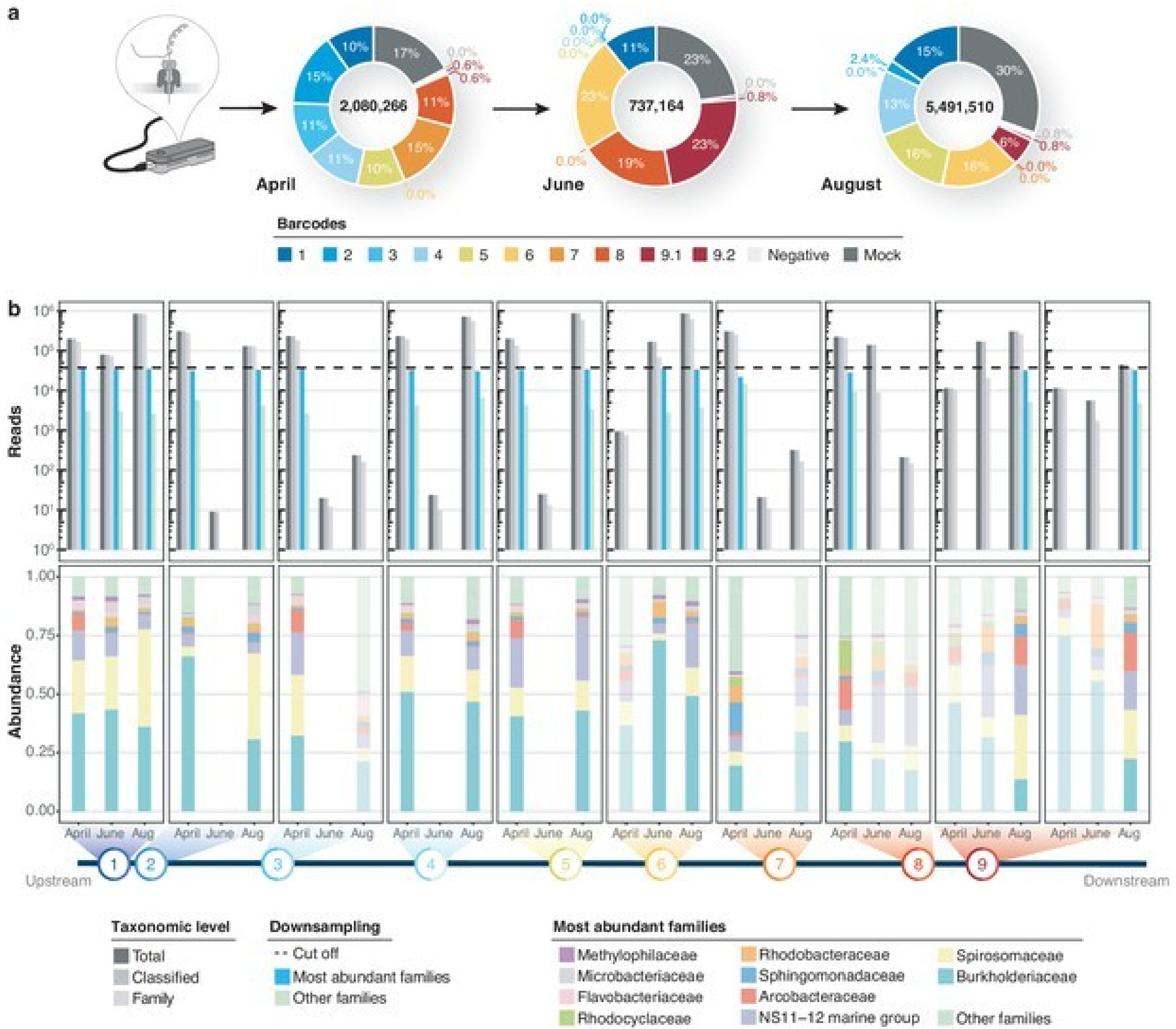
1. **Methodology:** Utilizing nanopore sequencing to trace all aquatic DNA, providing detailed insights into microbiota compositions and spatiotemporal variations in freshwater samples.
2. **Study Details:** The study presents optimized experimental and bioinformatics protocols based on surface water samples from a river in Cambridge, UK. It benchmarks twelve taxonomic classification tools for nanopore sequences.
3. **Findings:** Nanopore metagenomics effectively identifies core microbiomes in hydrological contexts and detects fine temporal gradients, aligning well with traditional physicochemical measurements.
4. **Public Health Applications:** The data reveal relevant signals from sewage contamination and pathogen presence at species-level resolution, indicating potential for enhanced environmental monitoring initiatives using portable sequencing devices.

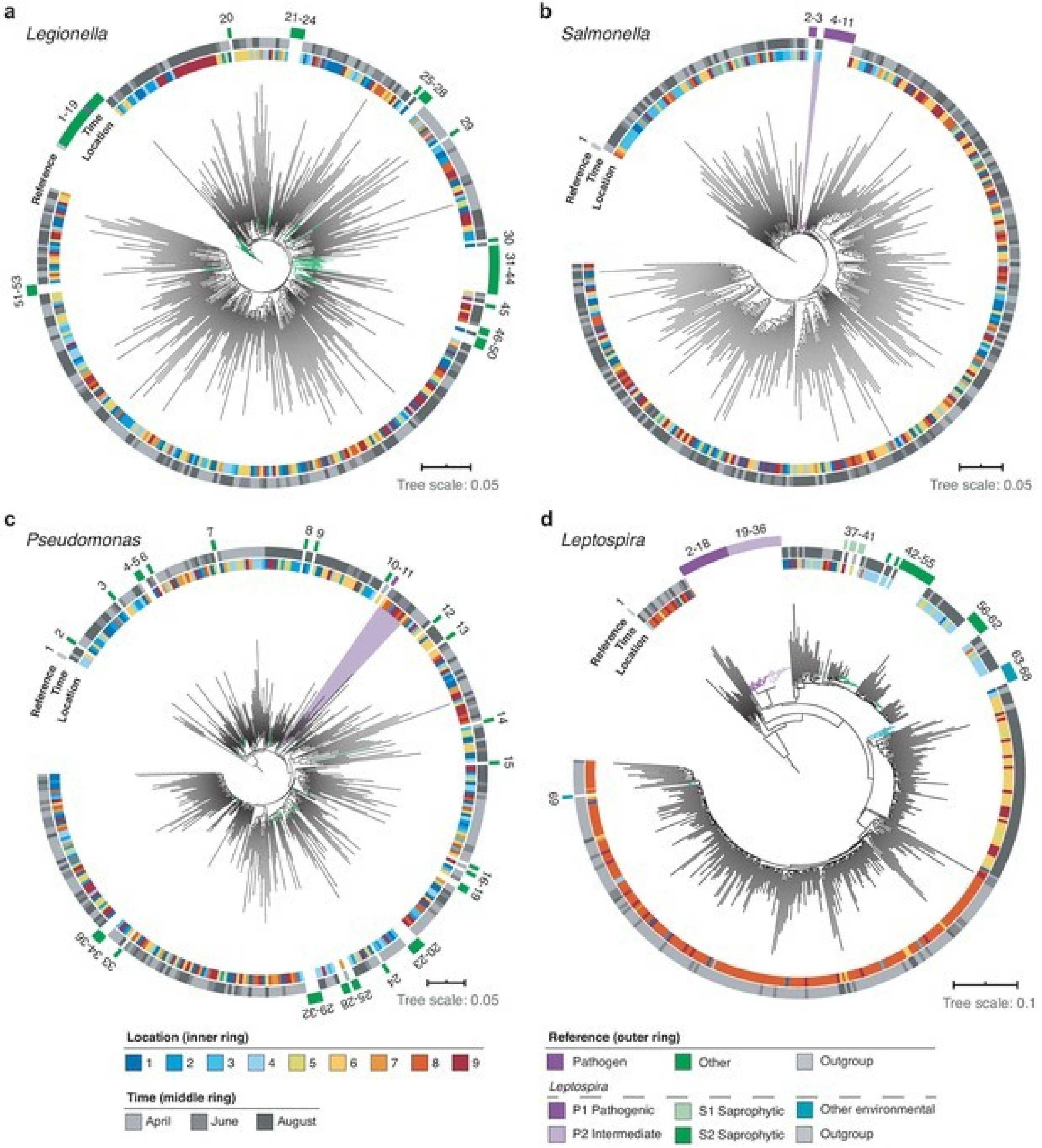


a
Downstream
Upstream
N
S
W
E
Baits Bite Lock
ϕ:52.23645° λ:0.17439°
Sewage Pipe Outlet
ϕ:52.23201° λ:0.17084°
Green Dragon Bridge
ϕ:52.21736° λ:0.14597°
Fort St. George
ϕ:52.21243° λ:0.12838°
Bridge Of Sighs
ϕ:52.2084° λ:0.11578°
Mill Pond
ϕ:52.19914° λ:0.11438°
Grantchester Meadows
ϕ:52.18657° λ:0.10185°
Nature Reserve
ϕ:52.17128° λ:0.09919°
Nature Reserve
ϕ:52.1707° λ:0.096°
City of Cambridge
River Cam
Distance
1 km
Zone 1
Zone 2
Zone 3
b
I. Filtration of 400 mL river water (0.2 μm cellulose membrane)
water sample
filtered residue
II. DNA extraction (Qiagen DNA Power Water Kit)
diluted filter
purified DNA
III. 16S rRNA amplification (barcoded custom primers)
16S rDNA
16S rDNA with barcodes
IV. ONT sequencing primer ligation (ONT Ligation Kit)
pooled barcoded fragments
library DNA
V. Nanopore sequencing (MinION, R9.4.1 flow cell)
...ATGACTATGACT
MinION



	expected	IDTAXA	Minimap2 (k = 15)	BLASTN	SPINGO (k = 8)	SINTAX	MegaBLAST	MAPseq	QIIME 2	Mothur (k = 6)	Kraken 2 (k = 21)	Centrifuge	RDP
Pseudomonadaceae	4.2	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	4.1	0.7	1.0	0.8
Enterobacteriaceae	20.5	58.8	59.1	59.9	59.7	60.0	59.6	60.4	59.9	48.2	59.9	60.4	67.0
Lactobacillaceae	18.4	4.9	4.8	4.7	4.7	4.7	4.6	4.7	4.6	9.4	4.8	5.0	1.8
Enterococcaceae	9.9	4.8	4.7	4.8	4.5	4.6	4.8	4.5	4.6	10.7	3.0	3.9	1.8
Staphylococcaceae	15.5	10.1	9.9	9.6	9.8	9.8	9.8	9.7	9.7	4.5	9.4	11.1	8.7
Listeriaceae	14.1	8.0	7.9	7.2	7.7	7.1	7.4	6.9	6.7	4.2	4.0	1.6	2.0
Bacillaceae	17.4	12.8	12.9	13.0	13.0	13.1	13.2	13.2	13.9	19.0	18.1	17.1	18.1
RMSE	0.00	15.91	16.04	16.38	16.30	16.42	16.25	16.57	16.41	12.36	16.63	16.87	19.68
RMSE reduced	**0.00**	**6.52**	**6.57**	**6.75**	**6.78**	**6.88**	**6.95**	**7.04**	**7.53**	**9.38**	**11.42**	**11.78**	**16.58**


a
2,080,266
737,164
5,491,510
April
June
August
Barcodes
1 2 3 4 5 6 7 8 9.1 9.2 Negative Mock
b
Reads
Abundance
April June Aug
Upstream
Downstream
Taxonomic level
Total
Classified
Family
Downsampling
Cut off
Most abundant families
Other families
Most abundant families
Methylophilaceae
Microbacteriaceae
Flavobacteriaceae
Rhodocyclaceae
Rhodobacteraceae
Sphingomonadaceae
Arcobacteraceae
NS11–12 marine group
Spirosomaceae
Burkholderiaceae
Other families



- **High-resolution phylogenetic clustering of candidate pathogenic genera in the River Cam.**
- Phylogenetic trees illustrating multiple sequence alignments of exemplary River Cam nanopore reads (black branches) classified as
  - (a) *Legionella*,
  - (b) *Salmonella*,
  - (c) *Pseudomonas*,
  - (d) *Leptospira*,
  - together with known reference species sequences ranging from pathogenic to saprophytic taxa within the same genus (colored branches). Reference species sequences are numbered in clockwise orientation around the tree.

Discussion

- Technology and Accuracy: Nanopore sequencing achieves **approximately 92% sequencing accuracy**, enabling the identification of significant shifts in human pathogen communities across rural-to-urban river transitions.
- Bioinformatics Challenges: Evaluates twelve taxonomic classification tools for nanopore sequences, highlighting varying rates of read misclassification and performance metrics. **Minimap2 emerges as more accurate but slightly slower compared to faster tools like Kraken 2 and Centrifuge**.
- Microbial Diversity and Composition: Nanopore amplicon sequencing identifies core microbiome structures and temporal variations in a freshwater environment, **primarily classifying taxa at genus and family levels.**
- Public Health and Environmental Monitoring: Demonstrates the potential for nanopore sequencing to **detect pathogenic bacteria like Legionella and Salmonella in freshwater samples**, supporting comprehensive environmental monitoring beyond traditional methods.
- Practical Considerations: Addresses experimental challenges such as DNA extraction yields, PCR biases, and sequencing throughput fluctuations, suggesting improvements in workflow optimizations and nanopore technology advancements.

Reference

- Insights into study design and statistical analyses in translational microbiome studies, https://atm.amegroups.org/article/view/13582/html
- Study Design Considerations, Simon R Carding, Falk Hildebrand & George M Savva, https://quadram.ac.uk/wp-content/uploads/2021/06/1.-Study-Design-Considerations.pdf
- Freshwater monitoring by nanopore sequencing, https://elifesciences.org/articles/61504#s2

Thank you for your listening