

Metataxonomic Analysis Experimental Design

Swamy Adapa

Bioinformaticians need to be really good at:

- Data cleaning and reformatting
- Programming
- Domain expertise (biology)
- Data visualization
- Statistical modeling
- Problem solving and some sanity check.
- Data/project (multi-task) management.
- Code sharing via GitHub - reproducibility
- Scientific Communication skills to a group of audience from different fields.
- Figure editing for pub
- Response to reviewers on why we don't need to do analysis X
- Installing tools with usually bad docs.
- Converting files again and again
- Good eyesight to see people from far down the corridor and duck aside before they ask you when "it" will be finished.
- Not going crazy when someone says "that's real quick cuz the computer does everything for you" 😅

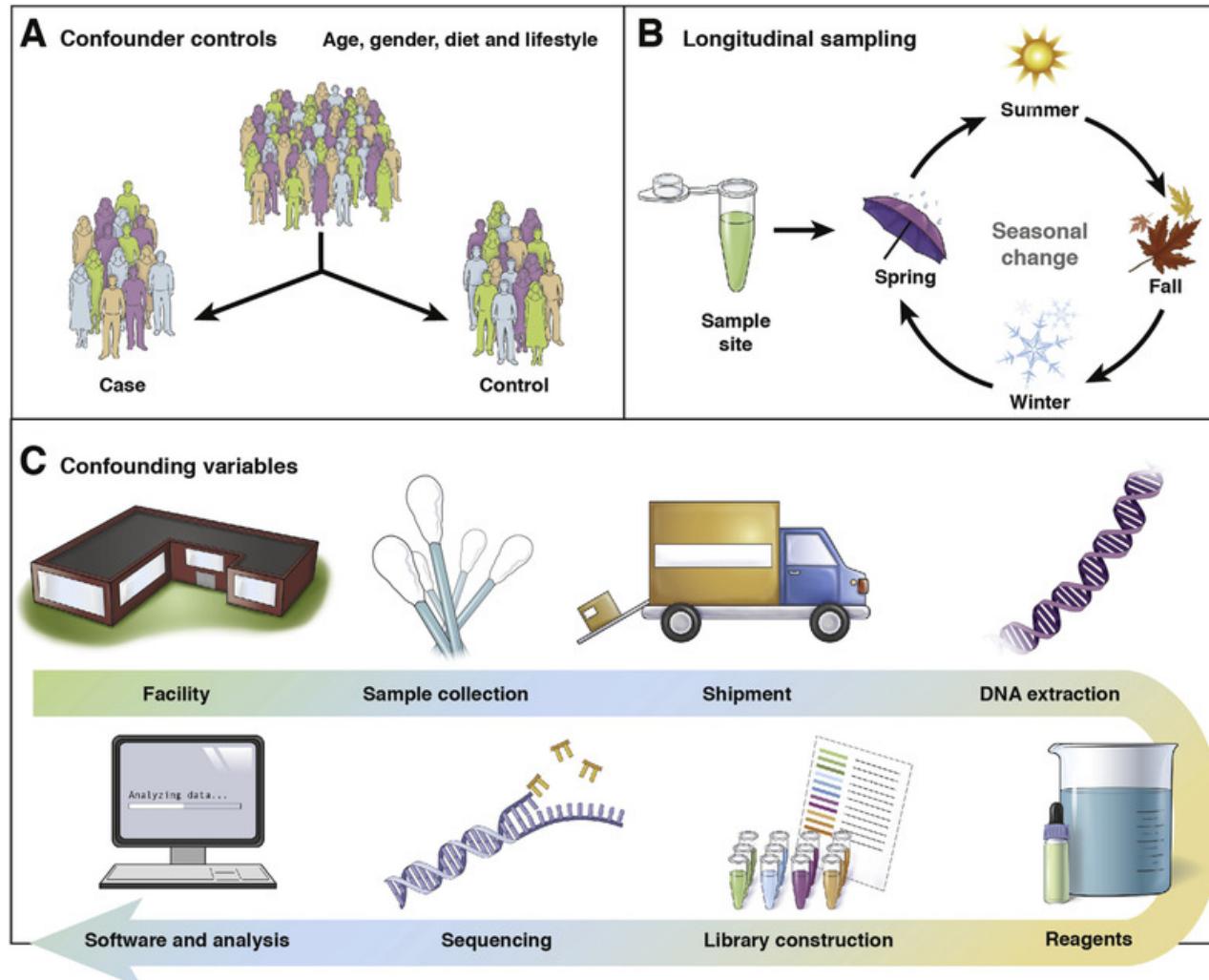
Study Design

- Clearly define the objective(s) of the study before deciding on the study design and statistical analysis.
- You should know your final statistical model and comparisons before beginning your experiment.

Effect size and target population:

- Effect Size - Pilot study or from scientific literature
- Study size - anticipated effect size and similar previous studies (power analysis)
- Well-defined inclusion and exclusion criteria - reduce heterogeneity in the cohort

Numerous factors influence clinical microbiome experiment



Challenges



(Work by Uwe Kils) <http://www.ecoscope.com/iceberg>

Control Group(s)

- Clinical phenotype distinct from the one under study while matching other relevant criteria to avoid confounding factor
- Longitudinal studies,
 - Collecting baseline samples before and after treatment to establish participant's own control.
- Multiple control groups selected on various criteria can provide more insights

Collect as many samples as possible.

As the increased statistical power of larger samples sizes cannot be overlooked when the difficulties of gaining approval and recruiting subjects represent a substantial barrier.

Collection Parameters

Donor Criteria

Gender

- Male, Female
- Specific cohorts

Age

- Adult, Geriatric, Pediatric

Race/Ethnicity

- White, Black, Hispanic, Asian
- Specific cohorts by Race/Ethnicity

BMI

- >30, <20, Specific Range
- Specific cohorts by BMI range/value

Dietary Restrictions

- Vegan, Vegetarian, Paleo, Etc
- Specific cohorts by dietary type

Dx Criteria

Diagnosis

- Physician confirmed
- Screened
- Test results
- Disease Severity/Flares/etc

Treatment specifics

- Naïve
- Specific Treatment
- Specific washout
- Specific exclusion (such as Antibiotics)
- Current medications

Timing and frequency of sample collection:

Level of invasiveness of the sampling procedure should always be minimized.

- Cross-sectional sampling
 - Diagnostic microbiome signatures
 - Temporally stable microbiome types.
- Repeated sampling (same subject)
 - To get insight into temporal dynamics and a comprehensive view of microbial community changes.
 - Monitoring disease severity or response to a treatment.
 - Discriminate between dysbiosis associated with disease and preclinical states and short-term changes followed by resilience.
 - Sampling frequency should be similar between subjects

Sample Criteria

Sample Types

Gut (Feces)	Urine
Nostril	Mouth
Skin	Hair
Earwax	

Sample consistency/viscosity/origin

- Loose, firm, etc
- Location of skin swab/hair collection

Volume required

- 3 grams
- Fill a specific container

Time points

- During an active flare
- Pre/post antibiotics
- 4 weeks apart

Possible confounding factors:

Animal Studies: maternal, environment, facility, cage and vendor effects.

- A more stable microbiome can be generated by selective breeding of siblings over several generations
- Use germ-free mice gavaged or transplanted with the same inoculum from mixed-genotype litters
- Order mice from the same vendor
- Standardize housing conditions and minimize exposure of animals to stressors
- Randomize and cohoused treatment-control mice across litters and/or cages
- Distribute littermates over multiple cages and maximize the number of cages for each study group if treated and untreated animals cannot be cohoused
- Introduce systematic variation of parameters within one research institution to avoid extreme local standardization and reduced reproducibility between different sites

Possible confounding factors:

Human studies: lifestyle and clinical factors

Apply exclusion criteria (e.g., list of exclusion criteria from the NIH Human Microbiome Project)

Balance factors across the experimental groups

Carefully select control groups

Consistency

Procedures, protocols & collected information

- Implement consistent procedures and minimize extraneous variables across groups
- Collect maximum information about samples and experimental procedures to allow reproducibility

Collection - Shipment

Time to preservation / shipment / receipt

- 4 hrs, 24 hrs, 1-7 days, none
- Initial store at -20C, transfer to -80 within specific time frame

Collection device

- Fisherbrand™ Commode Specimen Collection System
- DNA Genotek Inc
- NorgenBiotekSaliva
- BD BBL™ CultureSwab™ EZ collection and transport systems

Storage container

- OMNIgene®•GUT
- NorgenBiotekSaliva, Urine, Stool
- DNA Genotek Inc

Storage solution

- Container specific
- Custom

What is the optimal protocol for collecting a microbiome sample for analysis?

There is still an ongoing debate on the best way to collect and store a sample for analysis of the microbiome.

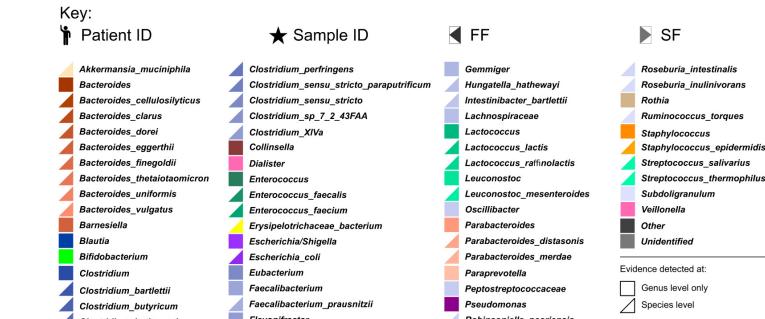
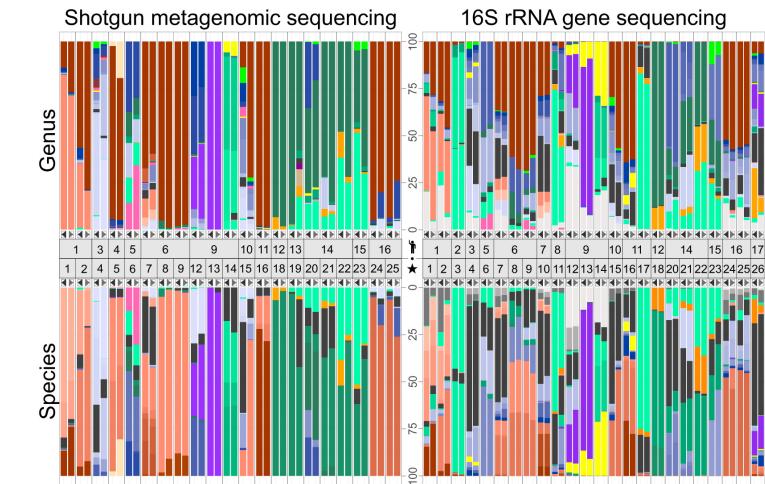
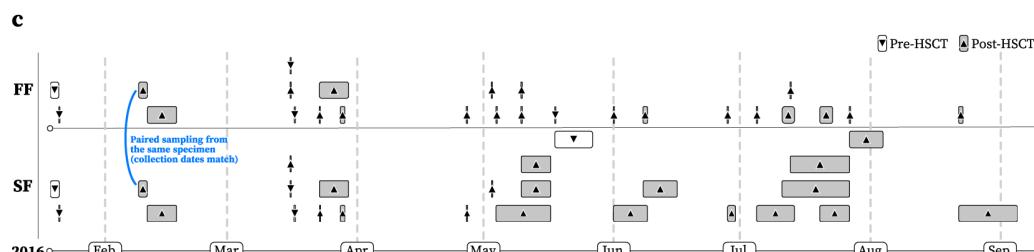
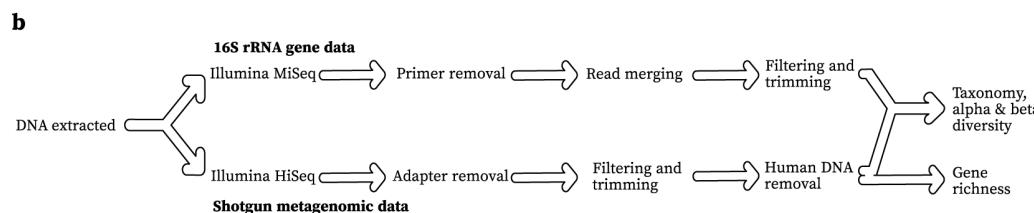
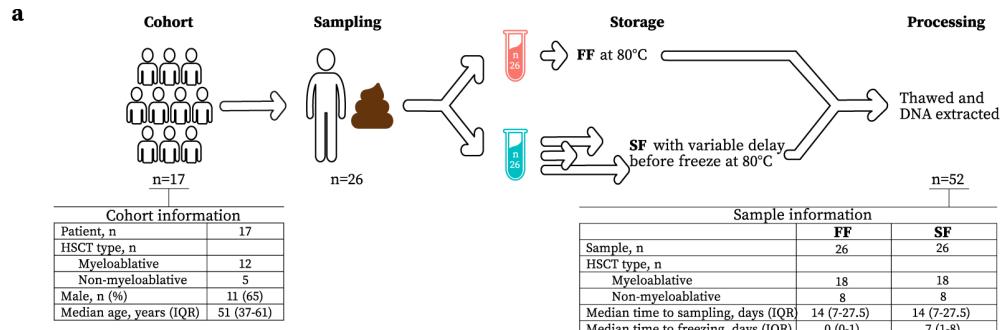
In short, there is no perfect method because the choice will depend on feasibility, cost, patient acceptance, and which methods will be used to read the microbiome downstream.

Even the most aggressive homogenizing procedure to break bacterial cell walls still may miss important organisms.

Sample collection

Fresh Frozen vs Stabilized Frozen

- This is not true in all cases – A huge divide in the community.
- Study shows that SF is comparable to FF protocols in a diseased and hospitalized cohort using both 16S rRNA gene and shotgun metagenomic sequencing data



Sample preparation

What may seem like small technical details introduced during sample extraction/preparation can lead to large changes, or technical bias, in the data.

- Prepare more samples than you are going to need, i.e. expect some will be of poor quality, or fail
- Preparation stages should occur across all samples at the same time (or as close as possible) and by the same person
- Spend time practicing a new technique to produce the highest quality product you can, rely on.
- Quality should be established using Fragment analysis traces (pseudo-gel images, RNA RIN > 7.0)
 - DNA/RNA should not be degraded
 - 260/280 ratios for RNA should be approximately 2.0 and 260/230 should be between 2.0 and 2.2. Values over 1.8 are acceptable
- Quantity should be determined with a Fluorometer, such as a Qubit.

Sequencing errors

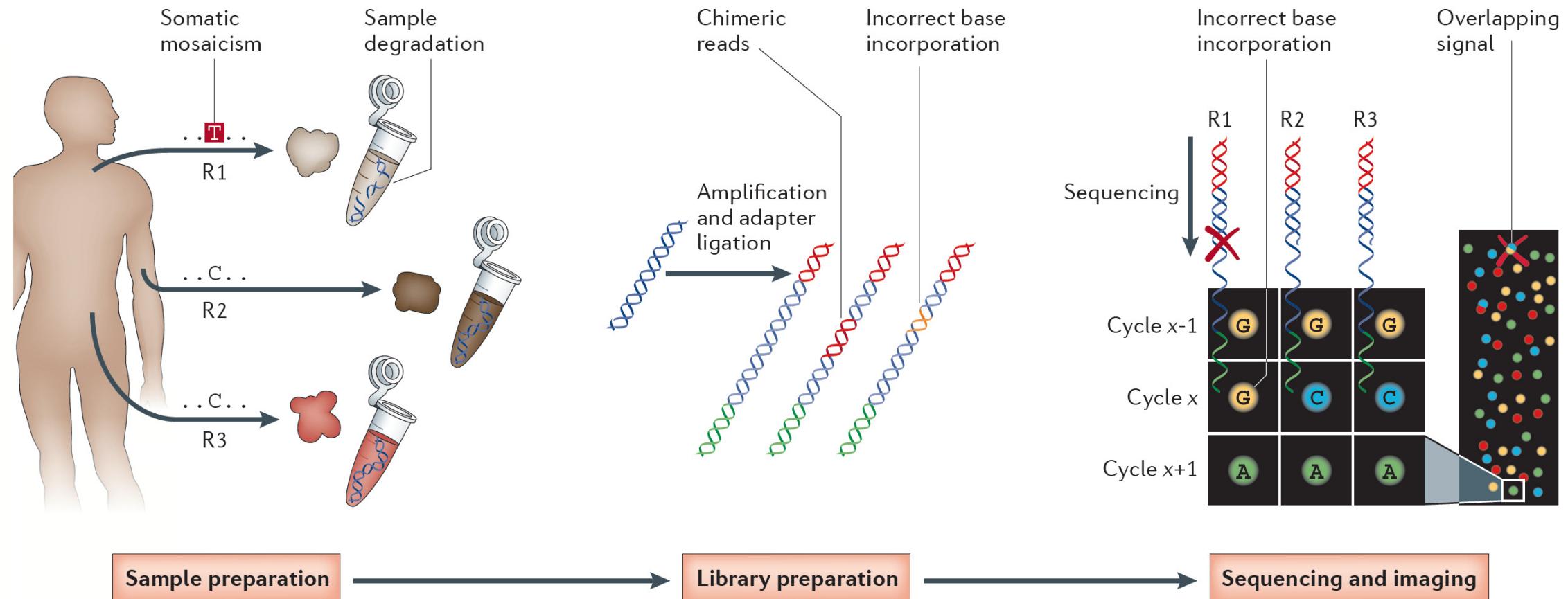
- Sequencing errors or mis-called bases occur when a sequencing method calls one or more bases incorrectly leading to an incorrect read.
- The chance of a sequencing error is generally known and quantifiable, thanks to extensive testing and calibration of the sequencing machines
- Each base in a read is assigned a quality score, indicating confidence that the base has been called correctly.

Sources of Sequencing Errors

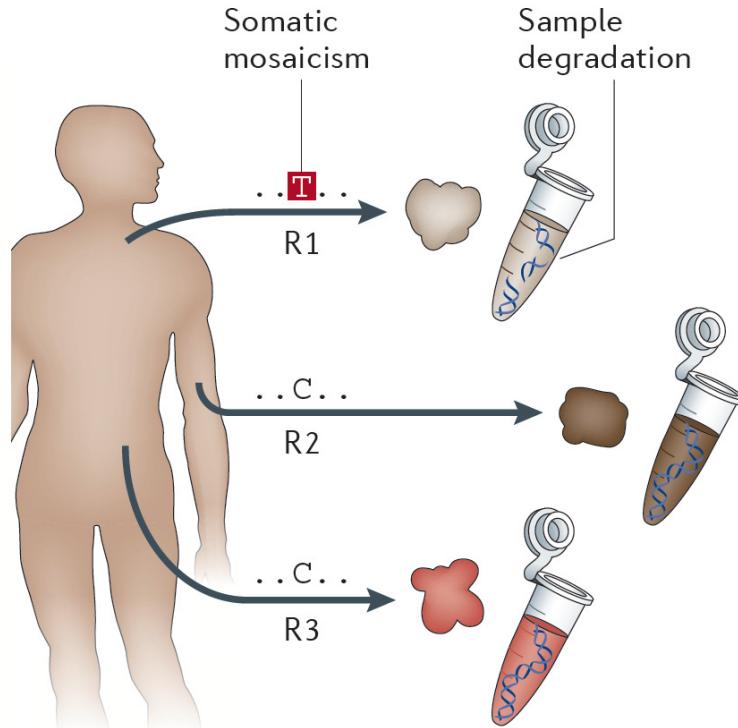
- The importance and the relative effect of each error source on downstream applications depend on many factors, such as:
 - sample acquisition
 - reagents
 - tissue type
 - protocol
 - instrumentation
 - experimental conditions
 - analytical application
 - the ultimate goal of the study.

Sources of Sequencing Errors

- Sequencing errors can stem from any time point throughout the experimental workflow, including initial sequence preparation, library preparation and sequencing.

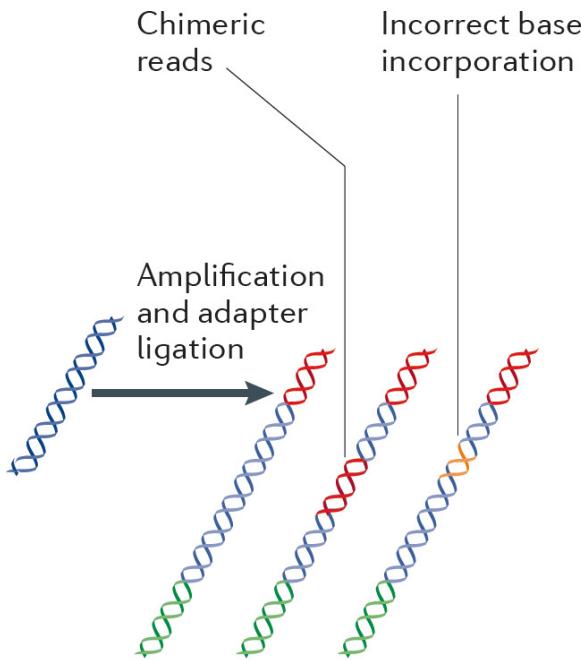


Sample Preparation



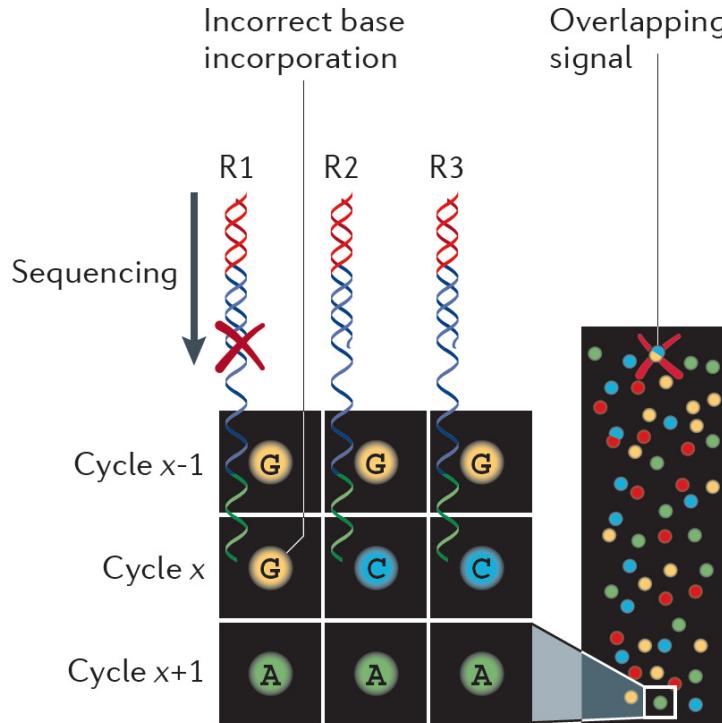
- User errors; for example, mislabelling
- Degradation of DNA and/or RNA from preservation methods; for example, tissue autolysis, nucleic acid degradation and crosslinking during the preparation of formalin-fixed, paraffin-embedded (FFPE) tissues
- Alien sequence contamination; for example, those of mycoplasma and xenograft hosts
- Low DNA input

Library Preparation



- User errors; for example, carry-over of DNA from one sample to the next and contamination from previous reactions
- PCR amplification errors
- Primer biases; for example, binding bias, methylation bias, biases that result from mispriming, nonspecific binding and the formation of primer dimers, hairpins and interfering pairs, and biases that are introduced by having a melting temperature that is too high or too low
- 3'-end capture bias that is introduced during poly(A) enrichment in high-throughput RNA sequencing
- Private mutations; for example, those introduced by repeat regions and mispriming over private variation
- Machine failure; for example, incorrect PCR cycling temperatures
- Chimeric reads
- Barcode and/or adaptor errors; for example, adaptor contamination, lack of barcode diversity and incompatible barcodes

Sequencing and Imaging



- User errors; for example, cluster crosstalk caused by overloading the flow cell
- Dephasing; for example, incomplete extension and addition of multiple nucleotides instead of a single nucleotide
- ‘Dead’ fluorophores, damaged nucleotides and overlapping signals
- Sequence context; for example, GC or AT richness, homologous and low-complexity regions, and homopolymers
- Machine failure; for example, failure of laser, hard drive, software and fluidics
- Strand biases

Marker sequencing

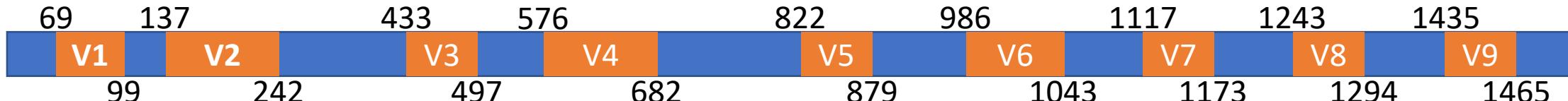
Depending on the sample type you are interested in sequencing, we can use a variety of amplicons :

- Prokaryotic communities (16S V4, V1-V3, V3-V5, V6-V9)
- Eukaryotic communities (18S)
- Fungal communities (ITS2)
- Archaeal communities (16S V4-V5)

No published primer pool covers all known organisms.

Rational primer pool design can match known variants in databases.

16S rRNA Gene Structure



Target Region Selection – Primer Pairs

*September 2016

16S rRNA

- Present in all known prokaryotic organisms
- No or less lateral gene transfer
- Extremely variable and conserved regions

Experimentally proved primer pairs

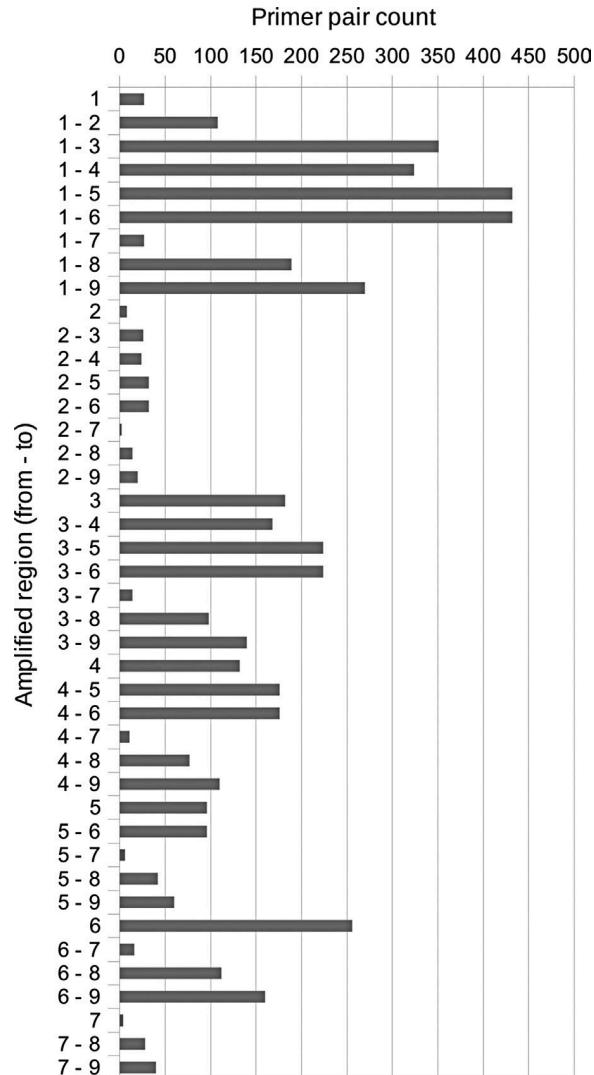
- 20 years of extensive literature
- probeBase repository

Commonly used primer pairs

- V13 - 27F and 534R
- V35 - 357F and 926R
- V69 - U968f and 1492r-MP

V4 region is about 250 bp

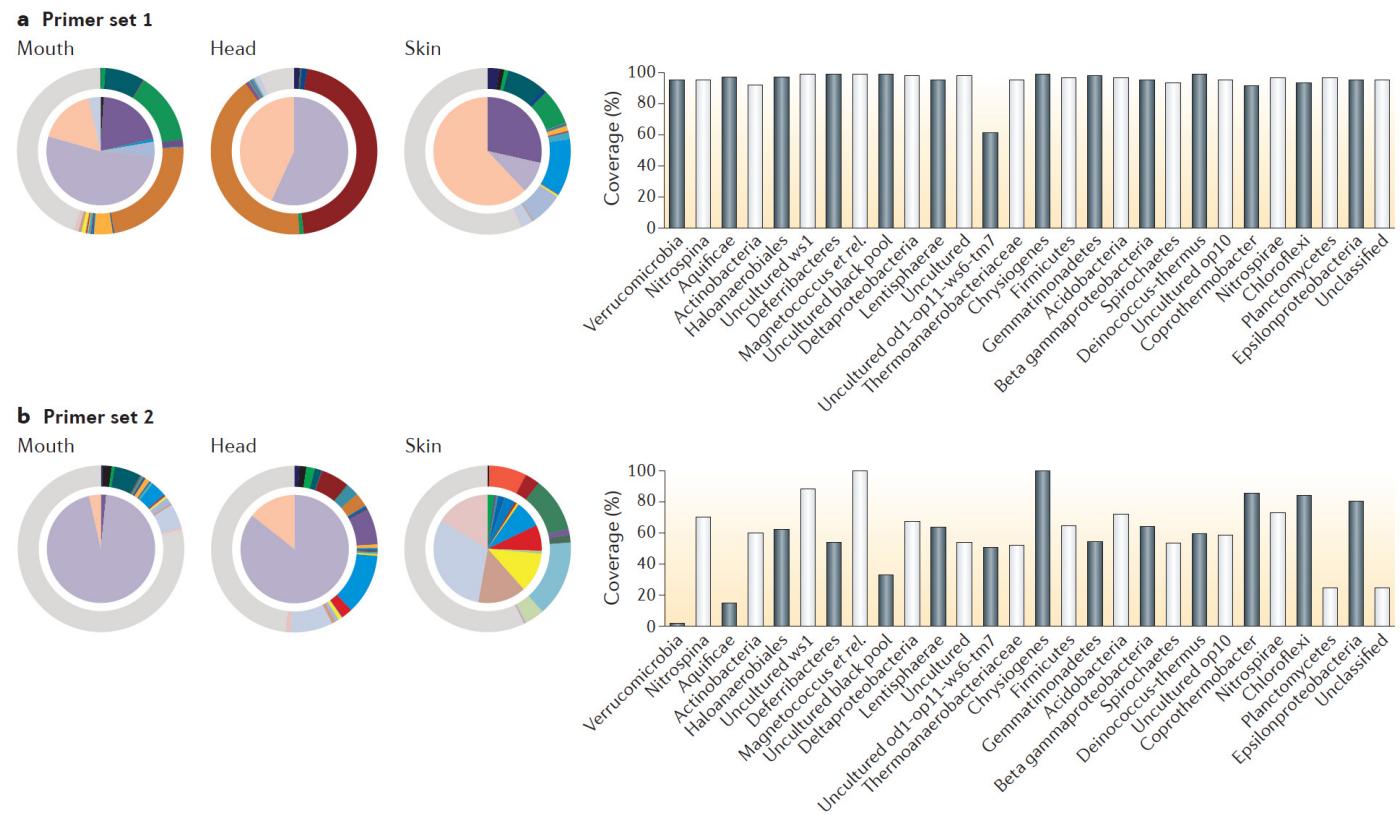
- Good phylogenetic coverage and precision
- Preferred region – Recent studies



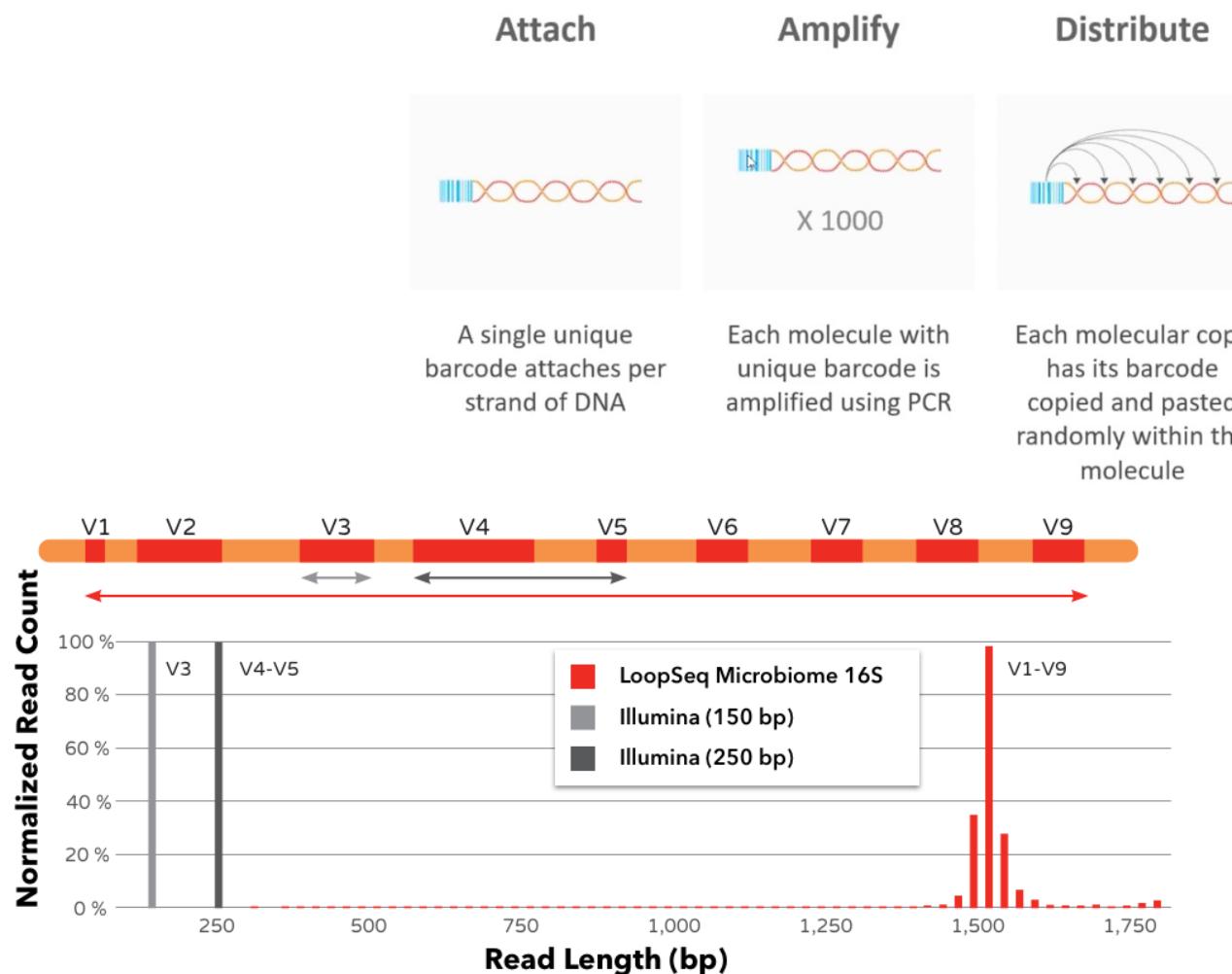
Exp	V1	V2	V3	V4	V5	V6	V7	V8	V9	Biome
133	991	1250	1265	969	463	340	247	214	26	Soil
37	207	245	1415	3906	169	44	20	13	15	Fecal
24	318	339	253	210	122	42	76	76	71	Agricultural
19	60	68	66	57	76	58	20	6	11	Forest soil
14	210	224	278	72	17	17	0	0	0	Dairy products
10	32	33	12	93	9	64	64	56	1	Grasslands
10	4	4	6	2	45	97	106	129	2	Rhizosphere
8	4	3	1	38	19	0	0	0	0	Contaminated
7	142	142	154	178	787	787	781	0	0	Digestive system
7	12	16	12	4	1	14	14	13	0	Permafrost
7	145	244	77	40	0	0	0	0	0	Rumen
6	87	88	90	9	0	14	14	14	0	Desert
6	1181	1614	1369	317	0	223	0	0	0	Vagina
4	5	3	2	1	0	0	0	0	0	Sand
3	0	0	14	2	1	0	0	0	0	Agricultural land
3	0	0	0	8	1	1	0	0	0	Boreal forest
3	8	8	12	5	0	0	0	0	0	Oil-contaminated
3	30	28	25	4	0	0	0	0	0	Tropical rainforest
3	0	0	0	22	4	0	0	0	0	Wetlands
2	575	575	0	2	0	0	0	0	0	Activated Sludge
2	0	0	11	11	0	0	0	0	0	Bioreactor
2	1	4	1	0	0	0	0	0	0	Environmental
2	0	0	62	10	0	0	0	0	0	Host-associated
2	1	1	16	45	0	0	0	0	0	Plants
2	930	963	1083	13	0	0	0	0	1	Skin
1	0	0	0	0	385	385	0	0	0	Abyssal plane
1	0	0	0	338	0	0	0	0	0	Agricultural wastewater
1	34	34	0	0	0	0	0	0	0	Amphibia
1	1	1	21	0	0	0	0	0	0	Animal
1	31	29	7	0	0	0	0	0	0	Bioremediation
1	0	0	0	0	0	1	1	0	0	Birds
1	10	10	10	0	0	0	0	0	0	Circulatory system
1	9	9	8	1	0	0	0	0	0	Coastal
1	0	0	3	0	0	0	0	0	0	Control
1	1	1	1	0	0	0	0	0	0	Crop
1	1	1	0	0	0	0	0	0	0	Freshwater
1	0	0	0	0	0	0	0	0	0	Fungi
1	0	0	0	19	0	0	0	0	0	Groundwater
1	0	0	14	18	0	0	0	0	0	Gut
1	0	0	0	996	0	0	0	0	0	Human
1	1	1	4	0	0	0	0	0	0	Intestine
1	1	1	0	112	0	0	0	0	0	Lab enrichment
1	0	0	0	1001	0	0	0	0	0	Large intestine
1	0	0	0	0	0	10	10	10	1	Marine
1	0	0	0	1	1	1	0	0	0	Mine drainage
1	144	144	144	0	0	0	0	0	0	Oral
1	12	12	5	0	0	0	0	0	0	Salt marsh
1	4	4	3	0	0	0	0	0	0	Sediment
1	9	9	8	0	0	0	0	0	0	Silt
1	0	0	0	10	0	0	0	0	0	Uranium contaminated
1	0	0	2	2	0	0	0	0	0	Volcanic
1	0	0	2	2	1	0	0	0	0	Wastewater

Effects of primer choice in targeted amplicon sequencing.

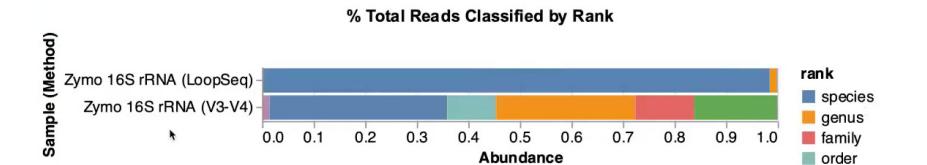
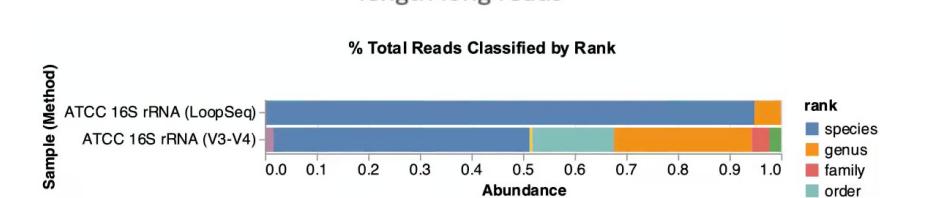
- Optimal primer set should amplify a region that is informative both taxonomically and phylogenetically, depending on the desired analyses
- 16S rDNA hypervariable region V6 poorly replicates full-length taxonomic assignments compared with the V4 or V2 region



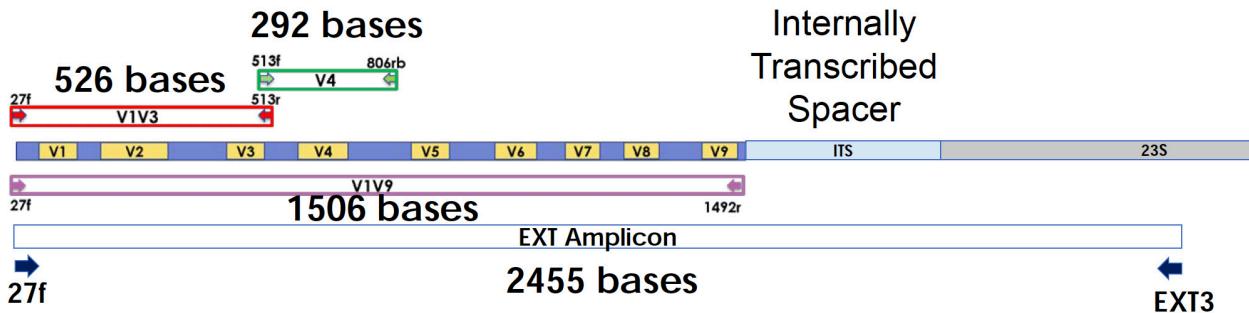
Synthetic Long Read 16S - Illumina



38\$ per sample for 200 samples

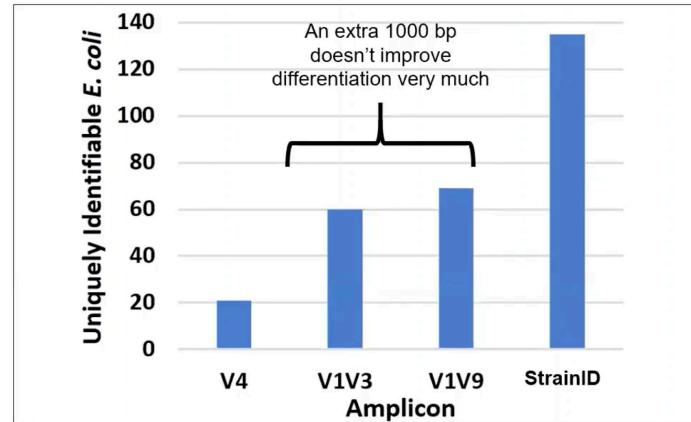


Long Read 16S - PacBio



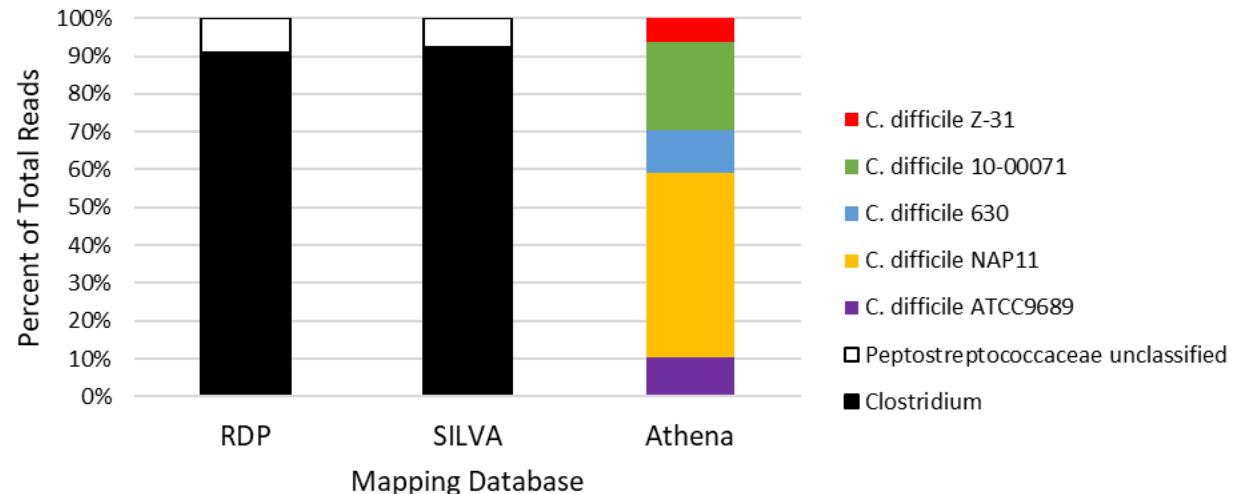
- | | |
|------|----------------|
| V4 | Family/Genus |
| V1V3 | Genus/Species |
| V1V9 | Species |
| EXT | Species/Strain |

StrainID differentiates strains better than V1V9



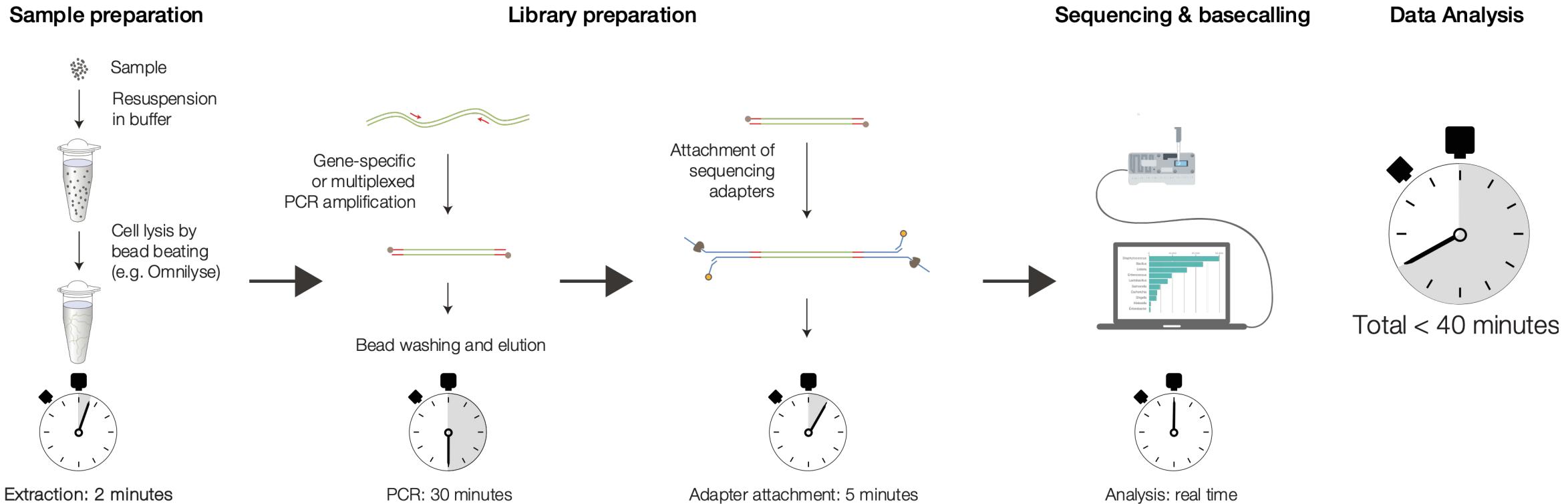
- V1-V9: 16S only
- StrainID:
 - 16S
 - ITS (internal spacer region)
 - partial 23S
 - ~2500 bp fragment

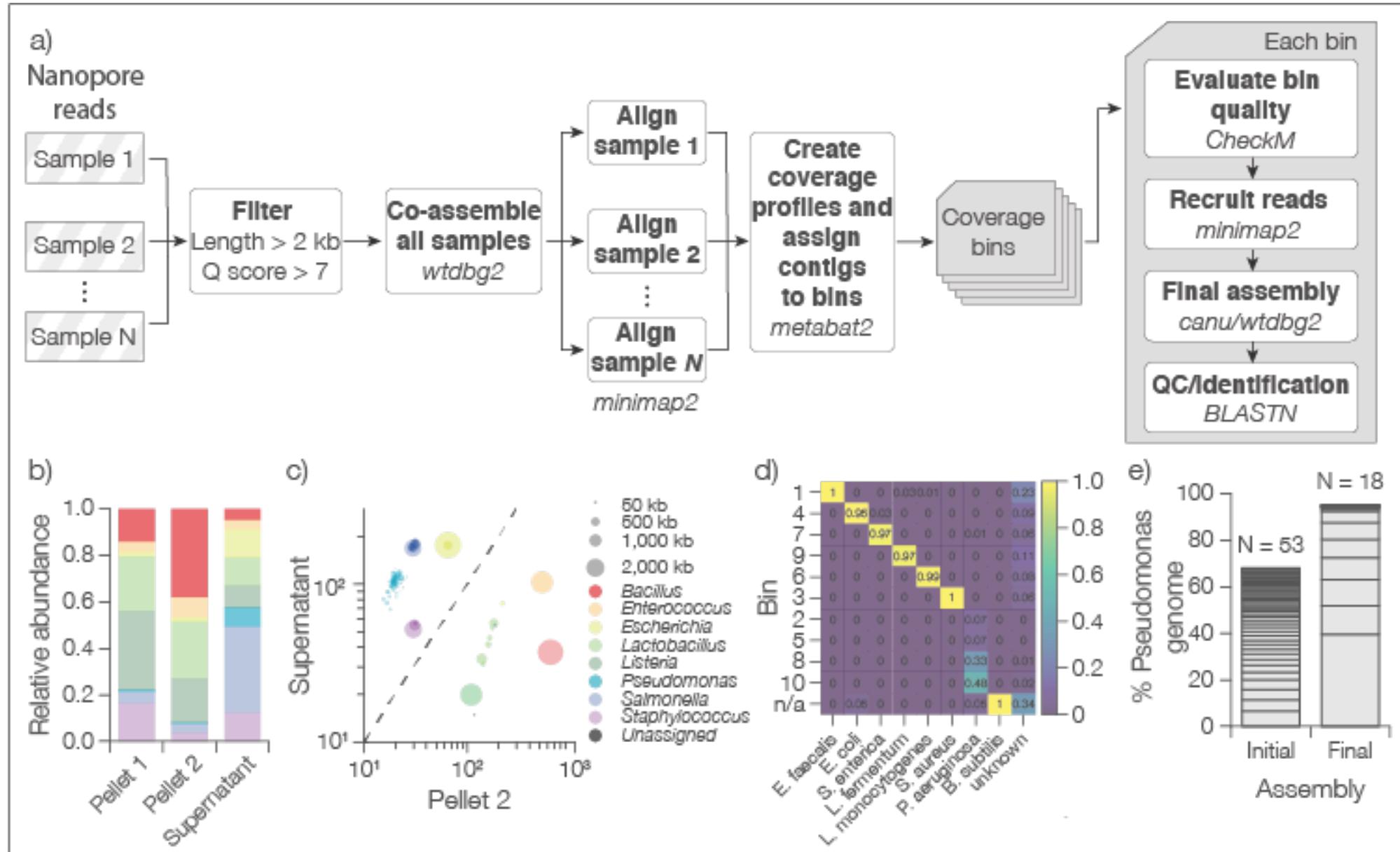
StrainID increases sequence diversity and identifies all 137 *E. coli* strains



Long-read Nanopore Sequencing

The 16S Barcoding Kits and EPI2ME 16S analysis workflow allows users to perform genus-level identification from single reads; with access to basecalled files for detailed investigations at the species and sub-species level.





Biases

- Universality of the primers
 - Reduce the effective richness of the sample
 - Not necessarily match all the taxa contained in our sample
 - Consider existing Primer sets (literature)
- Amplification issues
 - Alter the relative abundance profiles - Shannon entropy
 - Different combinations of PCR primers
- Region selection
 - Varying phylogenetic performances
 - Choose wisely
- Contaminant DNA
 - Kitome - Samples containing a low microbial biomass
 - Template-free “blanks” - correct abundance profiles
- Non-uniqueness of 16S rRNA genes
 - Ribosomal operon – sequence dissimilarity even in closely related species.
 - Use RNA operon copy number database (rrnDb) to adjust abundance profiles

16S rRNA Databanks

Annotated reference sequence – very important
Ribosomal Database Project (RDP)

- First released in 1992 and contained 447 entries of full length 16S sequences.
- Current release 11.04 accumulates 3,224,600 aligned and annotated 16S rRNA framed in a revised taxonomic tree

GreenGenes

- Started in 2002 as a semimanually maintained core alignment of nearly full-length, high-quality, chimera-checked 16S rRNA sequences extracted from other existing databases.
- Current release 13.5 includes 1,262,986 sequences coherently compared with a secondary structure-aware aligner and encoded in a 7682-character multiple sequence alignment

16S rRNA Databanks

SILVA

- Started in 1992
- Currently at release 128 and is the official database of the software package ARB, a long-term project aimed at providing a tool for handling, analyzing and visualizing sequences and related information.
- A core alignment is maintained - termed SSU Ref NR and containing 597,607 nearly full-length sequences carefully checked, representative and nonredundant) to grant accuracy in phylogenetic analysis.

Reproducible reanalysis

- Microbiome Quality Control project
 - Showed that differences in the computational pipeline, even on the same data, could lead to large differences in the inferred outcomes at levels from the species to the phylum.
 - Many different laboratories could independently reproduce similar results on the same samples by following a consistent written protocol.

Version Control

- The version and the parameters used when invoking a software during data analysis should be recorded.
 - Different versions of the same software could have different default parameters, and uncorrected bugs that could impact the analysis
- Versions of packages used by specific software such as R or python, should be documented as well

Reproducible reanalysis

- Dedicated workflow management software (e.g., Snakemake, bpipe or Galaxy)
 - To ensure that data analysis can be repeated by using analysis software used in the same order and with the same parameters
- Use containers (i.e., lightweight virtual computers)
 - To ensure the workflows are executed in the same computing environment when repeating data analysis (e.g., Docker and Docker-based solutions such as biocontainers or bioboxes, singularity)
- Leverage crowdsourcing to validation of results by external researchers using their own analysis methods
 - Starting from the raw data or restricted to specific steps of the data analysis
 - Can also be applied to evaluate reproducibility of experimental methods
 - To verify results and biological conclusions of a metagenomics study

Power Analysis

- Microbiome studies often compare groups of microbial communities with different environmental exposures, or to which different interventions have been applied.
- Pairwise distance metrics facilitate standardized comparison of community membership between individual study subjects.
- The design of microbiome studies demands consideration of statistical power
 - An adequate number of subjects must be recruited to ensure that the effect expected from the exposure or intervention of interest can be detected

Tools

- Micropower
 - Kelly, Brendan J., et al. "Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA." *Bioinformatics* 31.15 (2015): 2461-2468.
- HMP tool
 - La Rosa PS, Brooks JP, Deych E, Boone EL, Edwards DJ, et al. (2012) Hypothesis Testing and Power Calculations for Taxonomic-Based Human Microbiome Data. PLOS ONE 7(12): e52078.
- ShinyMB
 - Mattiello, Federico, et al. "A web application for sample size and power calculation in case-control microbiome studies." *Bioinformatics* 32.13 (2016): 2038-2040.

Genomic Standards Consortium (GSC) - MIMARKS

- The Genomic Standards Consortium (GSC) is an open-membership working body formed in September 2005.
- The aim of the GSC is making genomic data discoverable.
- The GSC enables genomic data integration, discovery and comparison through international community-driven standards.
- GSC proposed standards for describing genomic sequences—the “minimum information about a genome sequence” (MIGS)—and metagenomic sequences—the “minimum information about a metagenome sequence” (MIMS)

Specification projects	MIGS					MIMS	MIMARKS			New checklists
Checklists	EU	BA	PY	JI	ORG	metagenomes	survey	specimen	e.g., pan-genomes	
Shared descriptors	collection date, environmental package, environment (biome), environment (feature), environment (material), geographic location (country and/or sea, region), geographic location (latitude and longitude), investigation type, project name, sequencing method, submitted to INSDC									
Checklist-specific descriptors	assembly, estimated size, finishing strategy, isolation and growth condition, number of replicons, ploidy, propagation, reference for biomaterial			target gene						
Applicable environmental packages (measurements and observations)	Air Host-associated Human-associated Human-oral Human-gut Human-skin Human-vaginal					Microbial mat/biofilm Miscellaneous natural or artificial environment Plant-associated Sediment Soil Wastewater/sludge Water				

The core MIxS team developed the following environmental packages
These packages are available as separate spreadsheets:

Air ([download MIxS-air only](#))

Built-environment ([download MIxS-built environment only](#))

Host-associated ([download MIxS-host associated only](#))

Human-associated ([download MIxS-human associated only](#))

Human-gut ([download MIxS-human gut only](#))

Human-oral ([download MIxS-human oral only](#))

Human-skin ([download MIxS-human skin only](#))

Human-vaginal ([download MIxS-human vaginal only](#))

Microbial mat/biofilm ([download MIxS-microbial mat/biofilm only](#))

Miscellaneous natural or artificial environment ([download MIxS-misc. natural or artificial environment only](#))

Plant-associated ([download MIxS-plant associated only](#))

Sediment ([download MIxS-sediment only](#))

Soil ([download MIxS-soil only](#))

Wastewater/sludge ([download MIxS-wastewater/sludge only](#))

Water ([download MIxS-water only](#))

Hydrocarbon resources-cores ([download from GitHub](#))

Hydrocarbon resources-fluids/swabs ([download from GitHub](#))

Newly Published MIxS Packages:

Minimum Information About a Single Amplified Genome (MISAG) ([download from GitHub](#))

Minimum Information About a Metagenome-Assembled Genome (MIMAG) ([download from GitHub](#))

Minimum Information About an Uncultivated Virus Genome (MIUViG) ([download from GitHub](#))

In-Progress MIxS Standards:

Parasite Microbiome MIxS (MIxS-PMP)

Agricultural Microbiome MIxS (MIxS-Ag)

GitHub: <https://github.com/GenomicsStandardsConsortium/mixs>

MIxS

"Minimum Information about any (X) Sequence" (MixS) specification

Purpose

Store the Spreadsheet files for current MixS releases. This repo currently houses versions 4 and 5, and may house future versions. Older versions of the MixS standard are available at <https://github.com/GenomicsStandardsConsortium/mixs-legacy>.

To request changes to the MixS standards, please use the issue tracker in this repo.

The old content of this repository has moved to <https://github.com/GenomicsStandardsConsortium/mixs-ng>

We are working to serve MixS as RDF and JSON. The repo for that work is
<https://github.com/GenomicsStandardsConsortium/mixs-rdf>.

Reuse and citation of content on this repository

The MixS standards and the content of this repo are freely available under the Creative Commons 0 (open source) <https://creativecommons.org/share-your-work/public-domain/cc0/> agreement.

Cite the standard

If you use any of the MixS standards, please cite [this paper] (<https://www.nature.com/articles/nbt.1823>):

DOI: <https://doi.org/10.1038/nbt.1823>

PLoS Pathog. 2019 Oct; 15(10): e1008028.

Published online 2019 Oct 10. doi: [10.1371/journal.ppat.1008028](https://doi.org/10.1371/journal.ppat.1008028)

PMCID: PMC6786532

PMID: [31600339](https://pubmed.ncbi.nlm.nih.gov/31600339/)

Parasite microbiome project: Grand challenges

Nolwenn M. Dheilly,^{1,*} Joaquín Martínez Martínez,^{2,*} Karyna Rosario,³ Paul J. Brindley,^{4,5} Raina N. Fichorova,⁶ Jonathan Z. Kaye,⁷ Kevin D. Kohl,⁸ Laura J. Knoll,⁹ Julius Lukeš,¹⁰ Susan L. Perkins,¹¹ Robert Poulin,¹² Lynn Schriml,¹³ and Luke R. Thompson^{14,15}

June L. Round, Editor

► Author information ► Copyright and License information Disclaimer

This article has been cited by other articles in PMC.

The first Parasite Microbiome Project (PMP) Workshop (January 9–14, 2019, Clearwater, Florida, United States) hosted researchers from across continents and disciplines to lay the foundation of the PMP consortium. The PMP vision is to catalyze scientific discourse and explorations through a systems approach, toward an integrated understanding of the microbiota of parasites and their impact on health and disease. The participants identified knowledge gaps and grand challenges in the field of host-parasite-microbe interactions summarized here. The PMP will provide an interactive centralized platform and resource for transdisciplinary collaboration to propel the field of parasitology forward by disentangling complex interactions between parasites and hosts, their respective microbiota, and microbial communities in the parasite's direct environment (Fig 1).

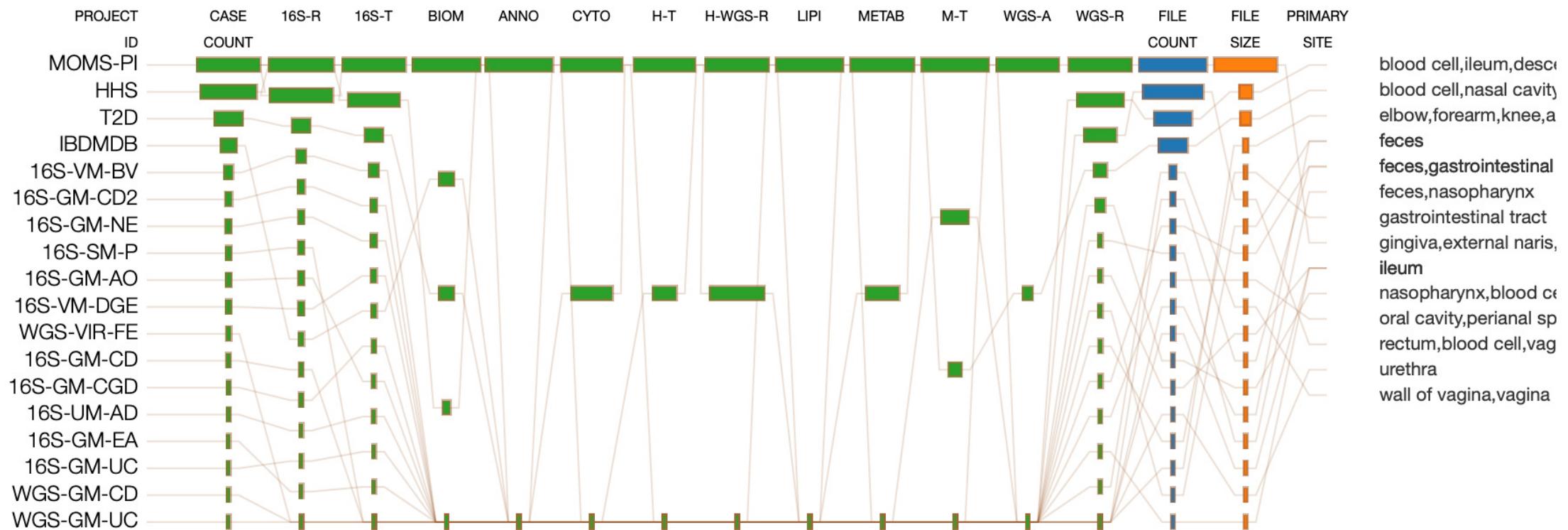


Human Microbiome Project

Data Analysis and Coordination Portal

<https://portal.hmpdacc.org>

Sample count per Data Category



Mgnify (EBI): Metagenome database with assemblies

<https://www.ebi.ac.uk/metagenomics>



Human
(97725)



Digestive
system
(65625)



Aquatic
(41710)



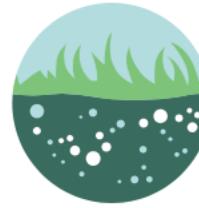
Marine
(30841)



Plants
(22217)



Digestive
system
(19016)



Soil
(17232)



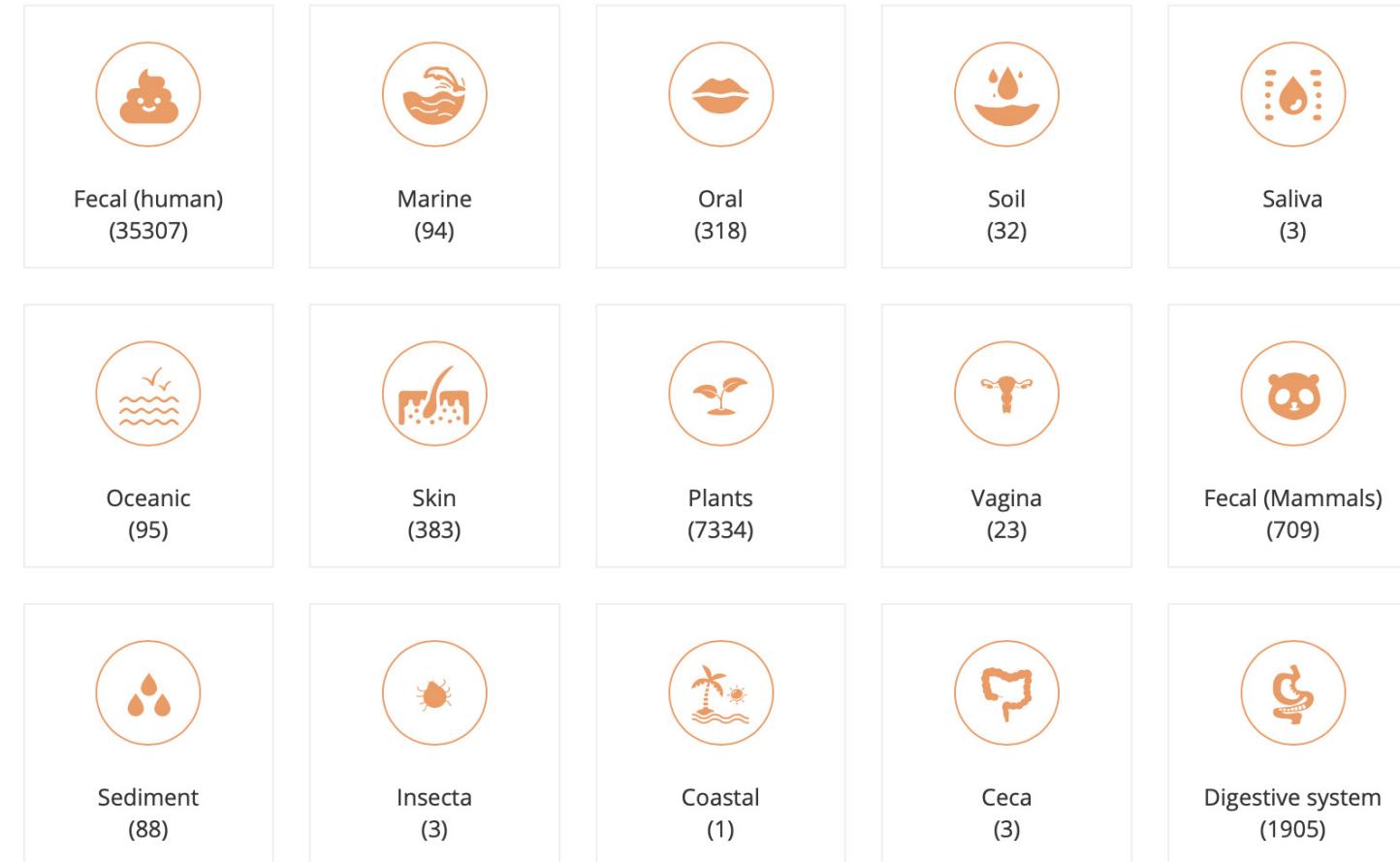
Skin
(7780)



Wastewater
(3477) Food
production
(1801)

MDB: A microbiome database hosted by China National GeneBank(CNGB)

<https://db.cngb.org/microbiome>



QIITA: rapid, web-enabled microbiome meta-analysis

<https://qiita.ucsd.edu>

Generated on: 02-11-20 00:57

Studies

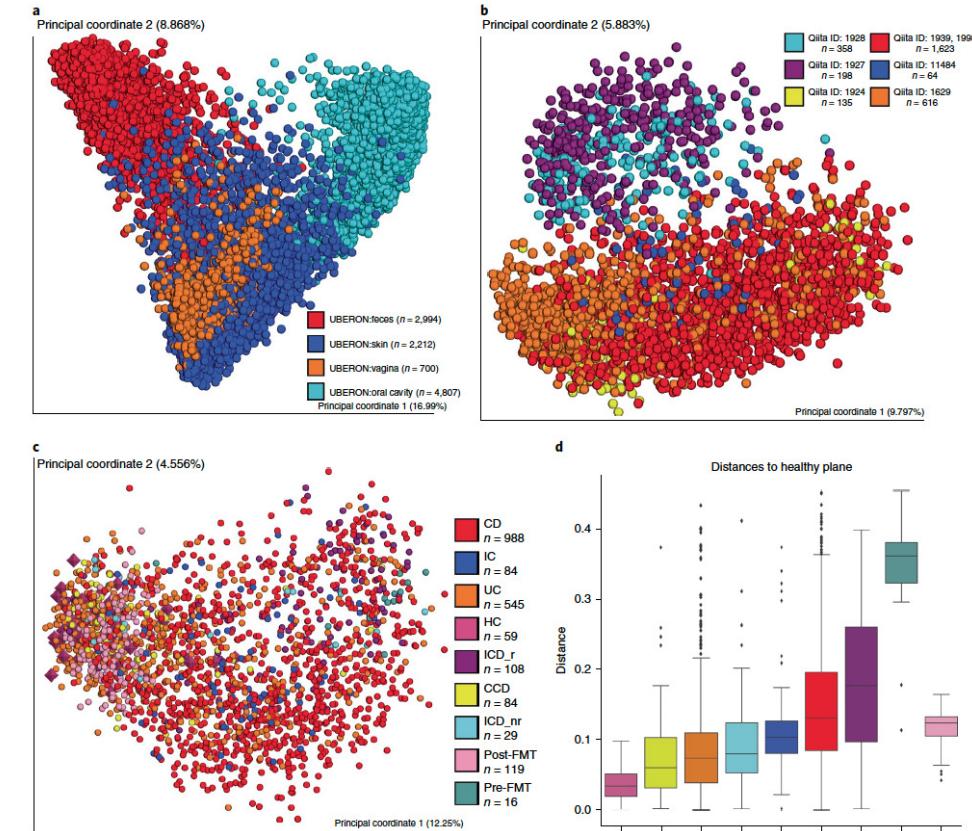
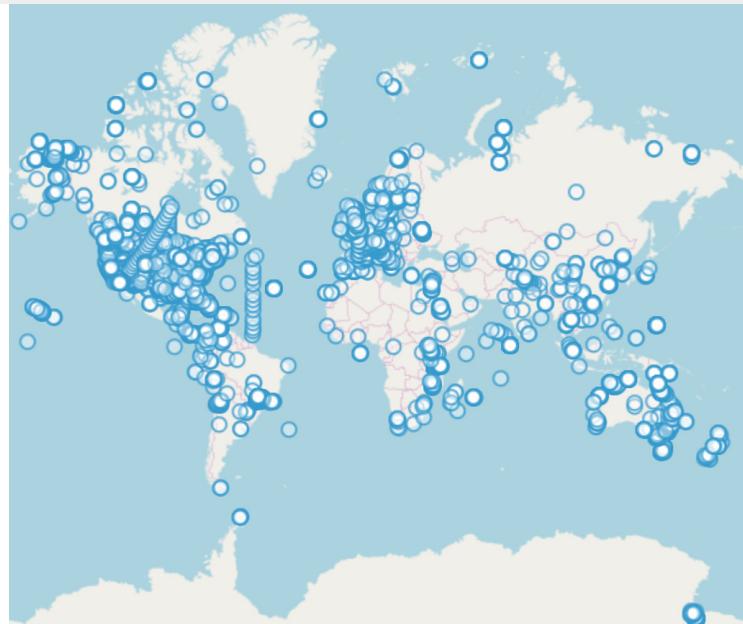
public: 521
private: 245
sandbox: 1,241
submitted to EBI: 346

Samples

public: 235,726
private: 133,675
sandbox: 305,419
submitted to EBI: 149,948
submitted to EBI (prep): 167,264

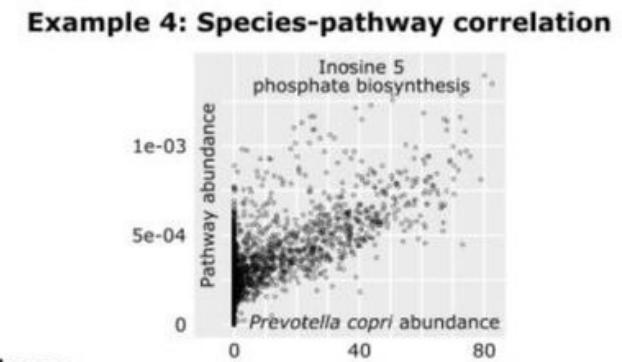
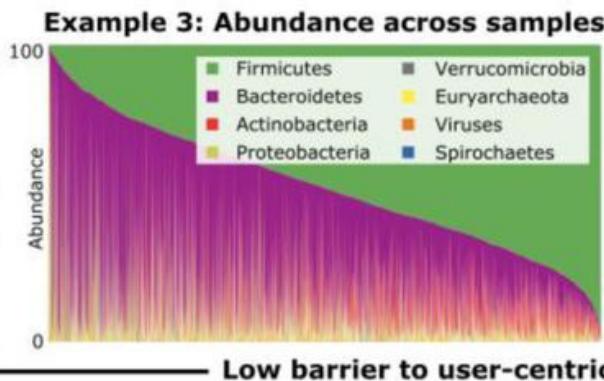
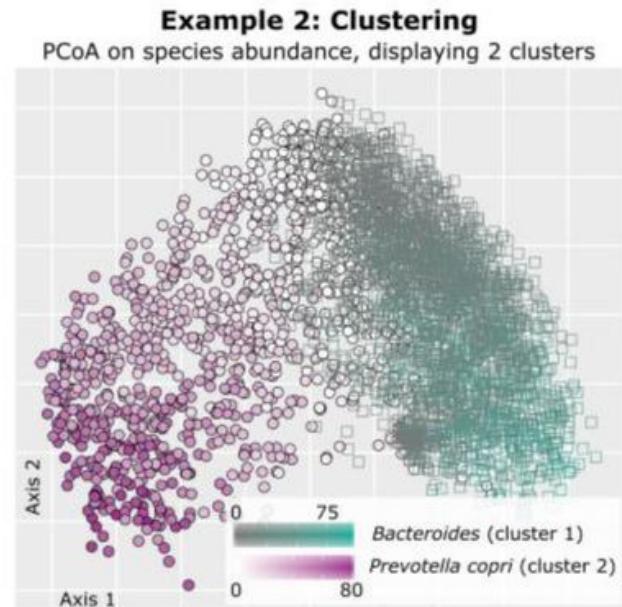
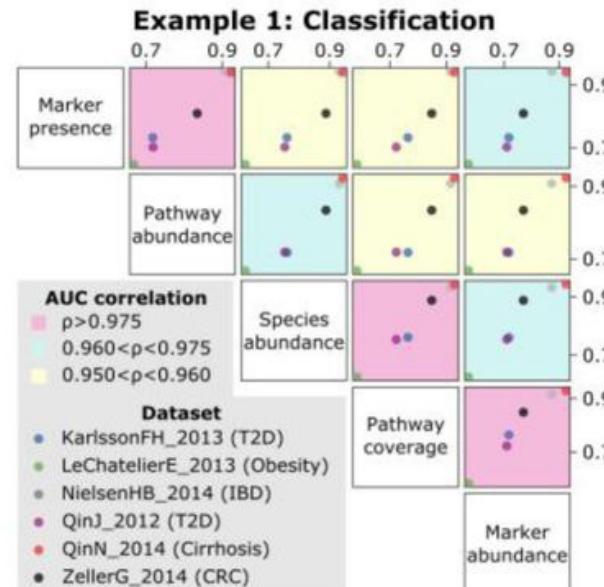
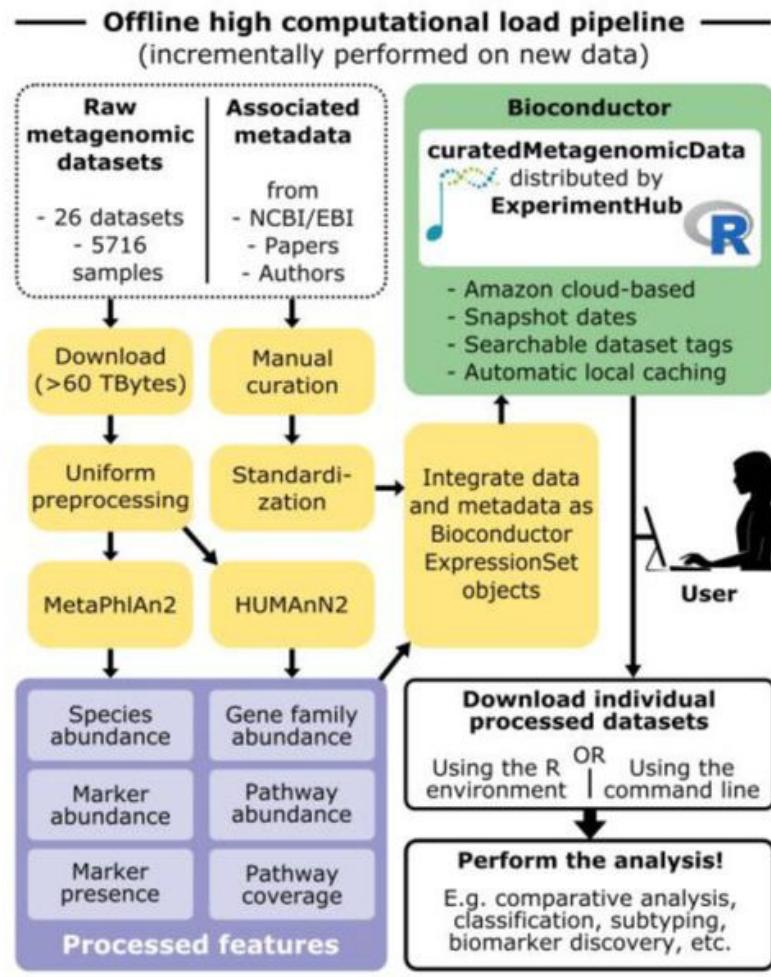
Samples per data type

16S: 236,801
18S: 7,425
ITS: 9,569
Metagenomic: 7,555
Metabolomic: 407



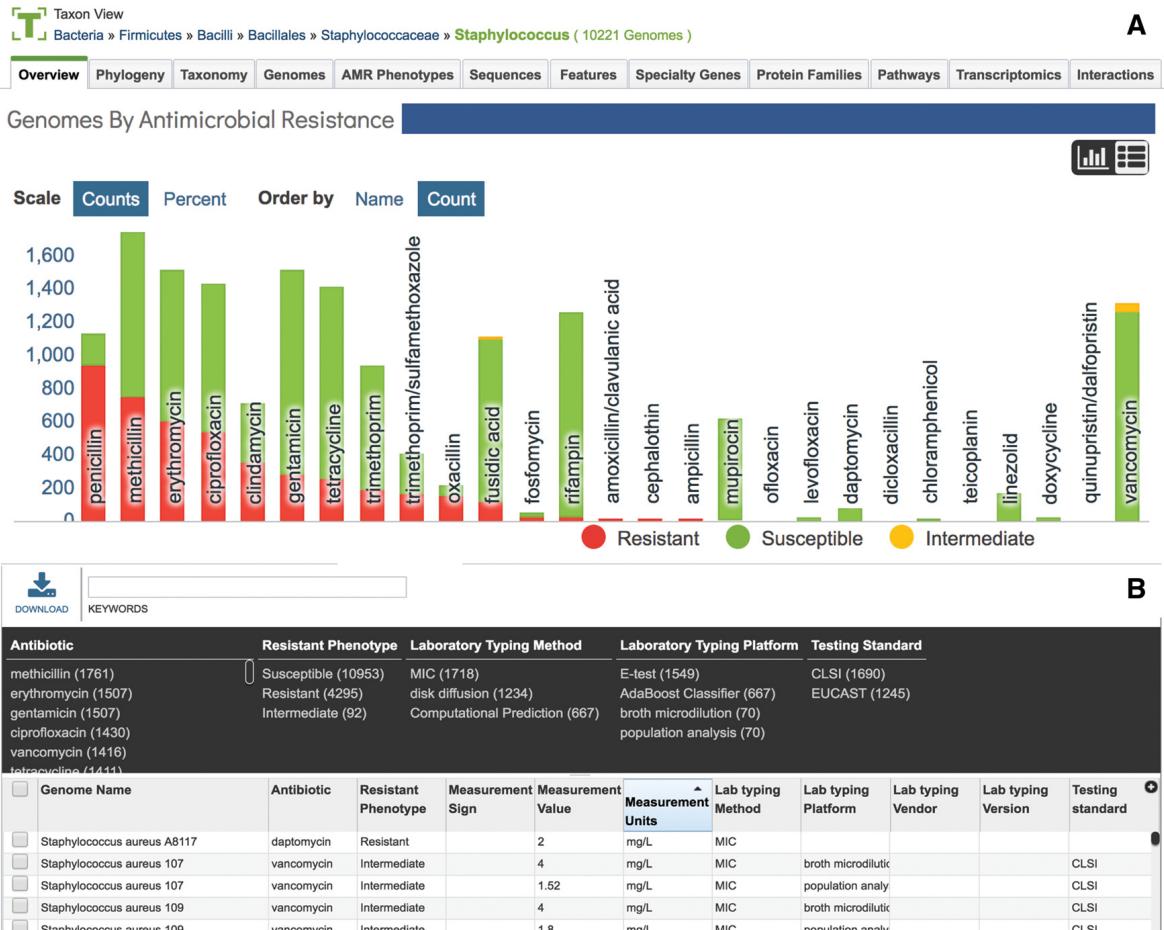
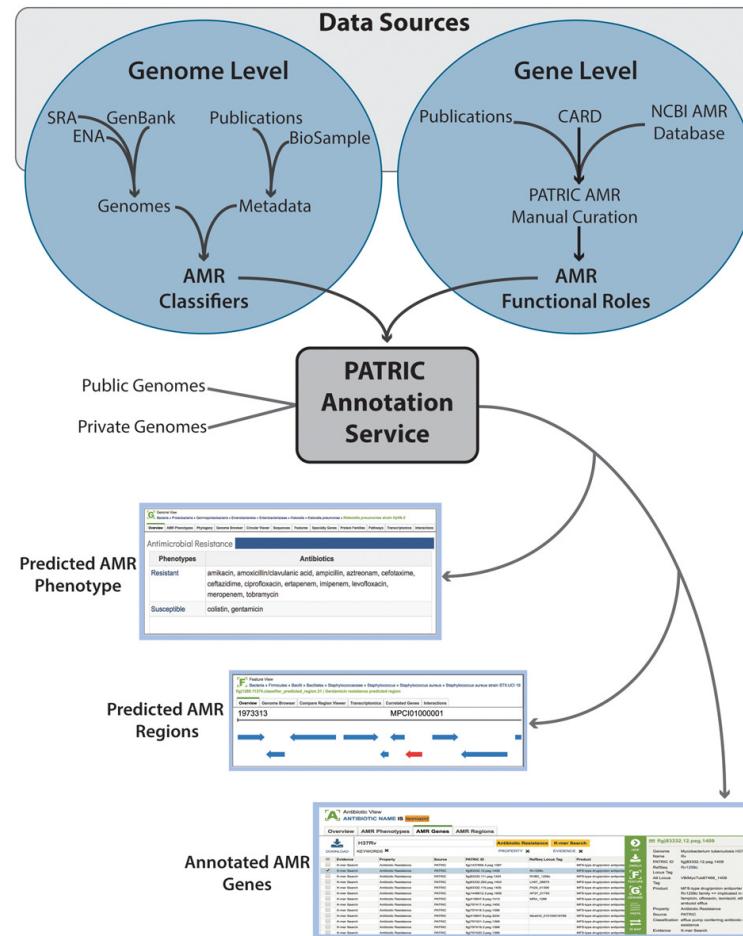
Gonzalez, A., Navas-Molina, J. A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., ... & Sanders, J. G. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. *Nature methods*, 15(10), 796-798.

curatedMetagenomicData : Curated and processed metagenomic data through ExperimentHub



PATRIC: Antibiotic resistance data

<https://patricbrc.org>



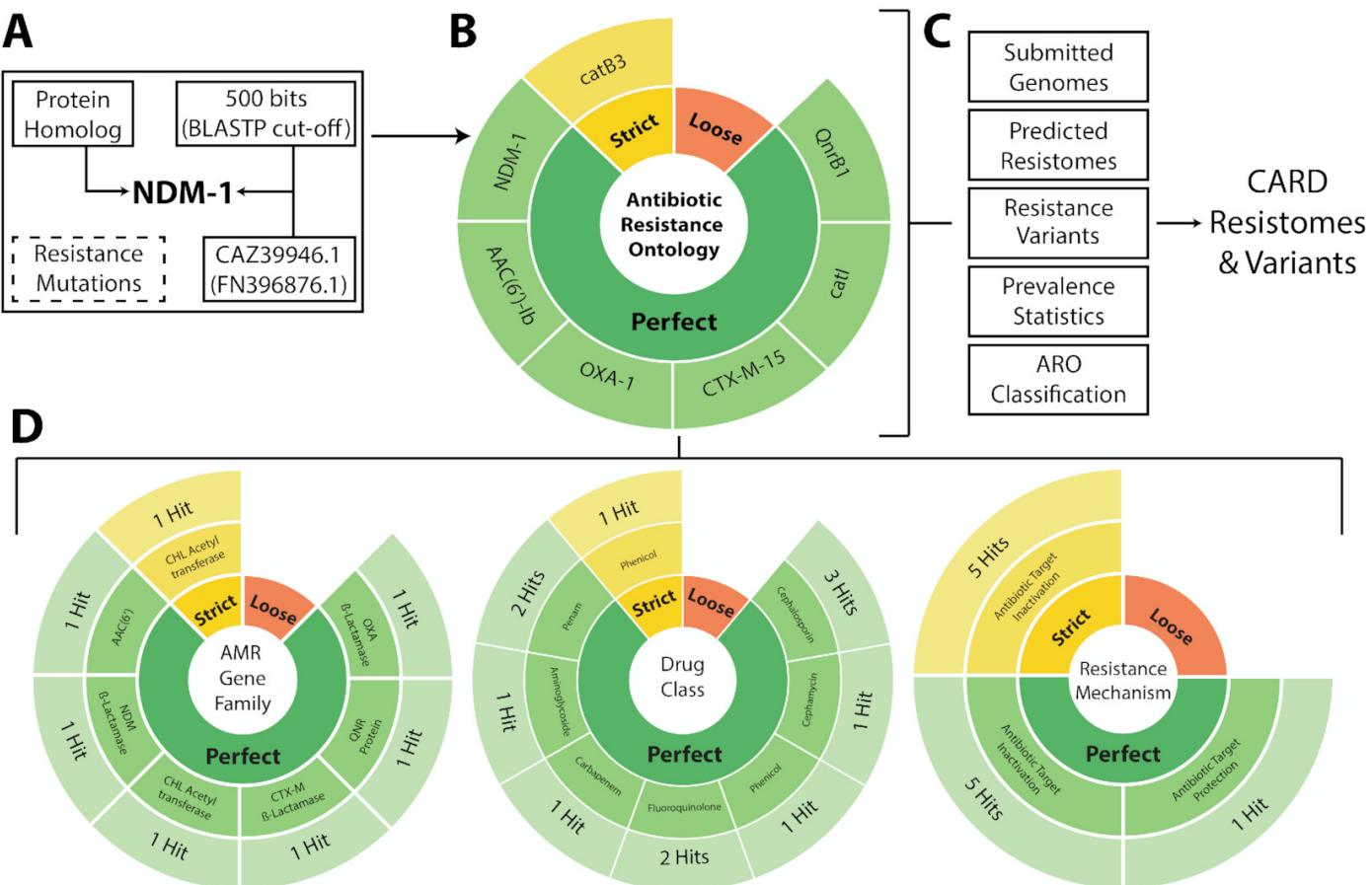
National Database of Antibiotic Resistant Organisms (NDARO)

Antonopoulos, D. A., Assaf, R., Aziz, R. K., Brettin, T., Bun, C., Conrad, N., ... & Kenyon, R. W. (2019). PATRIC as a unique resource for studying antimicrobial resistance. *Briefings in bioinformatics*, 20(4), 1094-1102.

CARD: antibiotic resistance gene database

RGI - Resistance Gene Identifier (cli)

<https://card.mcmaster.ca>



Microbiome molecular epidemiology.

