Assessing Performance of Metagenomic Profiling Using Microbial Genomic DNA Reference Material Mixtures

Jason G. Kralj, Dieter M. Tourlousse, Stephanie L. Servetas, Samuel P. Forry, Scott Jackson

Complex Microbial Systems Group, Biosystems and Biomaterials Division, Materials Measurements Laboratory
National Institute of Standards and Technology, 100 Bureau Dr. MS 8313; (301) 975-4130; jason.kralj@nist.gov

Introduction

Pathogen DNA Reference Materials to Benchmark Analysis

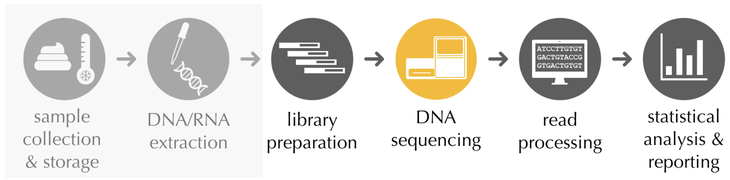
Metagenomics enable simultaneous analysis for (nearly) unlimited numbers of potential pathogens

- Multiplexed, *by design*--unlike PCR- and culture-based techniques
- Unbiased – all DNA subjected to same procedures

Transitioning technologies from the bench to the bedside/backyard stymied by lack of reproducibility across the analysis pipelines

- Regulatory bodies have established performance metrics for evaluating clinical and environmental decision making
- Developers are eager to benchmark their methods with these defined criteria
  - ❑ Translating a method to real-world application, and
  - ❑ Instill confidence in the analyses

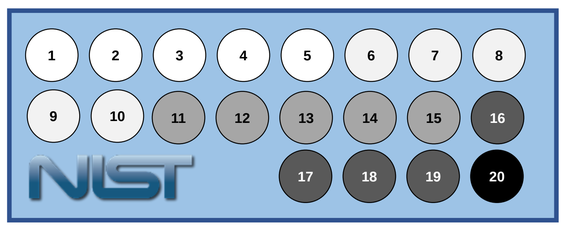
The materials and methods needed to evaluate these new tools are lacking because the sample analysis workflow is complex, with multiple opportunities for bias and error to propagate.

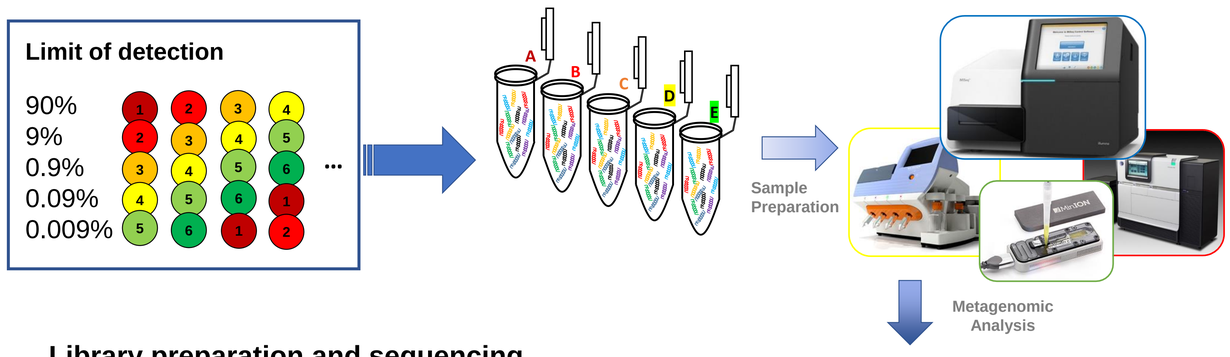


Schematic of the sample processing workflow depicts how each processing step skews information (through error and bias) to yield a result that may appear different from ground truth.

Experimental Methods*

Make mixtures to suit *YOUR* application





Library preparation and sequencing

Nextera XT DNA Library Prep Kit
MiSeq, V3 chemistry 2x301 bp

Read quality control

fastp v0.20.0

Read simulations

BBTools v38.26 (randomreads.sh)

Quantification genome relative abundances

kallisto v0.46.0

Species-level taxonomic profiling

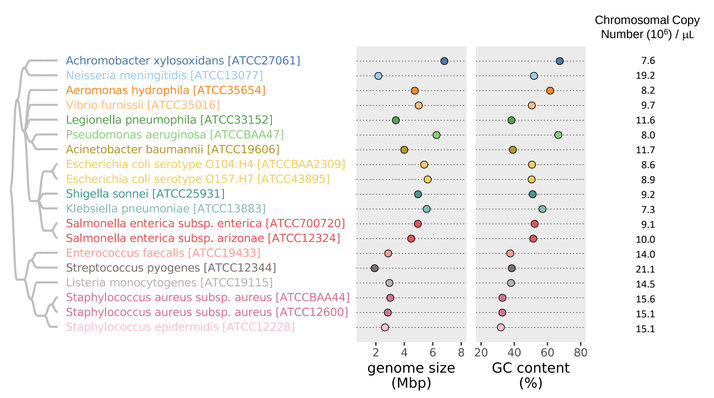
Centrifuge v1.0.4_beta
Metaphlan2 v2.7.8
Gottcha v1.0c

*Disclaimer: Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

NIST Candidate Reference Material 8376

- 20 constituents
- **DNA** from isolate bacteria + PGP Human Cell Line
- Modular
- Assembled genomes

- Near neighbors
- High/Low GC content
- Gram +/-
- Genome sizes
- AMR genes
- Disease sites



Mixture Design for Examining LOD, Informatics

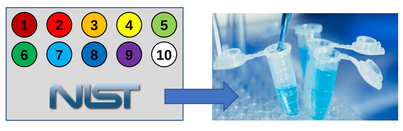
Wide range of concentrations w/ Latin square-type design

- Pools similar (gram +/-, G/C)

6 test samples

- Equigenomic, 5 $\log_{10}$ dilutions
- Subsampled *in silico-generated and experimental*

1. How does detection change vs. concentration?
2. How does taxonomic classifier affect sequencing reads interpretation?

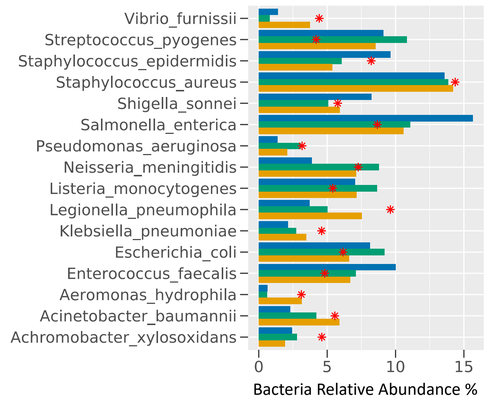


Pools

Annapolis			
ATCC 43895	*Escherichia*	*coli*	o157:h7
ATCC BAA 44	*Staphyloccus*	*aureus*	USA 300
ATCC BAA 47	*Pseudomonas*	*aeruginosa*	
ATCC 13077	*Neisseria*	*meningitidis*	
Baltimore			
ATCC BAA 2309	*Escherichia*	*coli*	o104:h4
ATCC 12600	*Staphyloccus*	*aureus*	
ATCC 27061	*Achromobacter*	*xylosoxidans*	
ATCC 35016	*Vibrio*	*furnissii*	
Chesapeake			
ATCC 700720	*Salmonella*	*enterica*	enterica
ATCC 12228	*Staphylococcus*	*epidermidis*	
ATCC 35654	*Aeromonas*	*hydrophila*	
ATCC 19115	*Listeria*	*monocytogenes*	
District			
ATCC 12324	*Salmonella*	*enterica*	arizonae
ATCC 12344	*Streptococcus*	*pyogenes*	
ATCC 19606	*Acinetobacter*	*baumannii*	
ATCC 19433	*Enterococcus*	*faecalis*	
Ellicott			
ATCC 13883	*Klebsiella*	*pneumoniae*	
ATCC 25931	*Shigella*	*sonnei*	
ATCC 33152	*Legionella*	*pneumophila*	

Results

Evaluating Database Effect – Same Data, Different Results

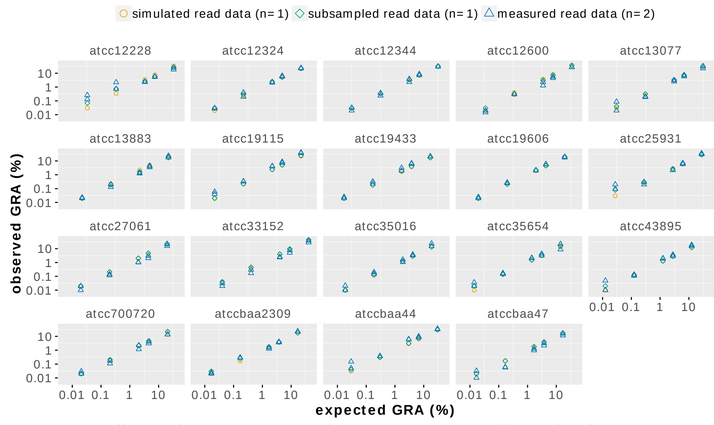


Limited vs. pan-genome database usage demonstrates how interpretation of the same data can be biased, despite all species having database representation. For applications using abundance criteria, e*x post facto* corrections, database curation, and/or multiple tools may be required.

`Centrifuge` v1.0.4_beta
Input *
p_compressed+h+v
p+h+v
Custom DB19 database

"Equigenomic" mixture of 19 strains (16 sp)

Simulated v. Subsampled v. Actual Mixtures



Kallisto with custom DB19 Database. GRA = Genome Relative Abundance

Analysis using full *in silico* (simulated) mixtures, subsampled isolate + *in silico* mixtures, and mixed DNA samples are in good agreement, and can all be used to evaluate performance.

Assembled genomes (19 components)

- sequencing read simulation
- rapid experimental space screening

Adding wet-lab experimental results

- verify the simulated sample results (and vice versa),
- Improve confidence analysis protocols performing properly

Effect of algorithm – Different Tools for Different Applications

centrifuge v 1.0.4_beta with p+h+v database
gottcha v1.0c with GOTTCHA_BACTERIA_c4937_k24_u30_xHUMAN3x.species database
metaphlan2 v2.7.8 with mpa_v20_m200 database

The measured vs. expected relative abundance for the 6 sample mixtures. With "default" databases, large differences observed how each taxonomic classification tool interprets the **same data**. Data represented <0.003 should have either been not measured or expected. Filled symbols = sample mixtures species, w/ open symbols = non-RM species.

Raw results from each taxonomic classifier tested show some of the biases of each tool. These include incorrectly identifying and excluding species, and incorrectly estimating relative abundances.

Discussion & Conclusions

Caveats

- Tool+database linked, confounding effects
- Latin-square design groups taxa, may mask correlations
- New evidence (McLaren, et al. *bioRxiv* (2019)) suggesting using taxon proportions to correct biases in relAb give superior sample composition estimates

RM facilitates evaluation of workflow biases

- *In silico and* subsampled reads mimic physical materials → **use both**
  - Develop analysis methodologies
  - Benchmark system behavior
- Sample composition and workflow (e.g. classifier) effects can be probed simultaneously to identify biases and errors

Significant work remains to develop rigorous benchmarking protocols for specific applications

Acknowledgements: Erica Romsos, Pete Vallone, Heike Sichtig, David Catoe, Justin Zook, Vanya Paralanov, Jayan Rammohan, and Steve Choquette (NIST); Patrick Chain (LANL); Jonathan Jacobs (Qiagen)
Funding: FDA CDRH