





LEMMI: a continuous benchmarking platform for metagenomics classifiers

LEMMI.ezlab.org

Evgeny.Zdobnov @ unige.ch

ICCMg 2021

The pipeline



MYCOBACTERIUM TUBERCULOSIS
GENOME SEQUENCING REPORT
NOT FOR DIAGNOSTIC USE

Patient Name	JOHN DOE	Barcode	
Birth Date	2000-01-01	Patient ID	12345678910
Location	SOMEPLACE	Sample Type	SPUTUM
Sample Source	PULMONARY	Sample Date	2016-12-25
Sample ID	A12345678	Sequenced From	MGIT CULTURED ISOLATE
Reporting Lab	LAB NAME	Report Date/Time	2017-01-01, 15:36
Requested By	REQUESTER NAME	Requester Contact	REQUESTER@EMAIL.COM

Summary

The specimen was positive for Mycobacterium tuberculosis. It is resistant to isoniaizd and rifampin. It belongs to a cluster, suggesting recent transmission.

Organism

The specimen was positive for Mycobacterium tuberculosis, lineage 2.2.1 (East-Asian Beijing).

Drug Susceptibility

binty of resista	nce.		Extensive drug resistance predicted
Drug class	Interpretation	Drug	Resistance Gene (Amino Acid Mutation)
	Susceptible	Ethambutol	No mutation detected
First Line	ousception	Pyrazinimide	No mutation detected
r in se carre	Desistant	Isoniazid	katG (S315T)
	Resistant	Rifampin	rpoB (S531L)
		Streptomycin	No mutation detected
		Ciprofloxacin	No mutation detected
		Ofloxacin	No mutation detected
Second Line	Susceptible	Moxifloxacin	No mutation detected
		Amikacin	No mutation detected
		Kanamycin	No mutation detected
		Capreomycin	No mutation detected

Scaling-up and mixing the puzzles

Who is there?(what they can do?)How many?





Is the bioinformatics solved?



Which one to pick? Why?

Selected abstract: Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples
 Jutte De Vries
 Netherlands
 DETAILS ABOUT THE EVENT

Why would you trust the authors without a critical evaluation?



Constructive reviews help a bit



(1) Independent benchmarks



Can you afford to wait? Did they do what **you** need?

(2) Community building

MDPI

Open Access Published: 02 October 2017

Critical Assessment of Metagenome Interpretation –a benchmark of metagenomics software

Alexander Sczyrba 🖂, Peter Hofmann, [...] Alice C McHardy 🖂





Article Viral Metagenomics in the Clinical Realm: Lessons Learned from a Swiss-Wide Ring Trial

Thomas Junier ^{1,*}, Michael Huber ^{2,*}, Stefan Schmutz ², Verena Kufner ^{2,†}, Osvaldo Zagordi ^{2,†}, Stefan Neuenschwander ^{3,†}, Alban Ramette ^{3,†}, Jakub Kubacki ^{4,†}, Claudia Bachofen ^{4,†}, Weihong Qi ^{5,†}, Florian Laubscher ^{6,†}, Samuel Cordey ^{6,†}, Laurent Kaiser ^{6,†}, Christian Beuret ^{7,†}, Valérie Barbié ⁸, Jacques Fellay ^{1,9} and Aitana Lebrand ^{8,*}



Could **you** reuse/reproduce? Did you act on the results?

What is to benchmark?



Relative abundance of known species?

Diagnostic of low abundance microbes in human host context?

(3) Continuous benchmarking



Live Evaluation of computational Methods for Metagenome Investigation

Go to LEMMI >> ☑

Continuous integration

is *key*, as snapshot evaluations are dead on arrival.

Multi-objective ranking

according to certain objectives or technical limitations.

Effective distribution

using containers (i.e. isolated software packages) for plug and play and reproducibility.

Seppey, Manni, Zdobnov 2019. https://doi.org/10.1101/507731

Bridging developers & users



(A) Solves the problem of the objectivity

- Highly automated:
 - Evaluation in a matter of weeks
 - Less arbitrary / human interpretation

- Use public data in an isolated/controlled environment
 - Redefine on the fly what is known or unknown
 - Faster generation of new samples than sequencing

(B) Solves the problem of the time gap

- Novel tools are released frequently
- A software update can have a major impact
 - On resource usage
 - Debugging can change results.
- Companion or third-party scripts can be game changers

Expect softwares to evolve <u>on new releases</u>, not only on new publications, <u>re-evaluate</u> as often as necessary on comparable problems.

(C) Solves the problem of reproducibility

- Software containers
 - For portability / distribution
 - Multi-step pipelines as one benchmarking item

(D) Solves the problem of flexibility

LEMMI	Taxonomic Cla	ssifiers A L	_iv	e Evaluation of Computational Methods for Metagenome Investigation		UNIVERSITÉ DE GENÈVE FACULTY OF MEDICINE
Home I	Rankings I	Details by datasets	1	Evaluated methods: description, sources, and containers II FAQ II Documentation II Discussion board II	Ì.	About/Contact

Rank	Method	Parameters	Reference	Score	Rank	Method	Parameters	Reference	Scor
1	Ganon	k=19+fwd reads only	BUILT I RefSeq/08.2018 1rep.	0.577	1	Kraken 2.0.7 + Bracken-2.0	k=35+l=150	BUILT I RefSeq/08.2018 All	0.692
2	Kraken 2.0.7 + Bracken-2.0	k=35+l=150	BUILT I RefSeq/08.2018 All	0.575	2	Kraken 2.0.7 + Bracken-2.0	k=35+l=150	BUILT I RefSeq/08.2018 1rep.	0.65
3	Metacache v0.5.0	k=16	BUILT RefSeq/08.2018 All	0.567	3	Clark-I	light	BUILT I RefSeq/08.2018 All	0.64

Rank	Method	Parameters	Reference	Score
1	Ganon	k=19+fwd reads only	BUILT I RefSeq/08.2018 1rep.	0.767
2	Metacache v0.5.0	k=22	BUILT I RefSeq/08.2018 1rep.	0.731
3	Metacache v0.5.0	k=16	BUILT RefSeq/08.2018 All	0.729

ank	Method	Parameters	Reference	Score
1	Ganon	k=19+fwd reads only	BUILT I RefSeq/08.2018 1rep.	0.993
2	Metacache v0.5.0	k=16	BUILT I RefSeq/08.2018 All	0.991
3	Kraken 2.0.7 + Bracken-2.0	k=35+l=150	BUILT I RefSeq/08.2018 All	0.99

Choose custom criteria

SD Species detect	tion, true and false predictions, i	gnoring < 100 reads
RA Relative abundar	nce of organisms, e.g. to study micr	obial imbalance
LA Low abundance :	score, e.g. for pathogens identificat	ion
RB Reads taxonomic	c binning, how reads are clustered ir	nto taxonomic groups
Predictions to consider:		
1) Taxonomic rank		
 1) Taxonomic rank Species only 	O Both	O Genus only
 1) Taxonomic rank Species only 2) Low coverage (count 	\bigcirc Both of reads < 100) ^(?)	O Genus only
 1) Taxonomic rank Species only 2) Low coverage (count Include for all metrics 	O Both of reads < 100) ^(?) s ⊙ Ignore	 Genus only Add the low abundance score (?)
 1) Taxonomic rank Species only 2) Low coverage (count Include for all metrics Choice and weight of the operation 	O Both of reads < 100) ^(?) s O Ignore different metrics:	 Genus only Add the low abundance score (?)
 1) Taxonomic rank Species only 2) Low coverage (count Include for all metrics Choice and weight of the only 1) Taxa detection recall 	O Both of reads < 100) ^(?) s O Ignore different metrics:	 Genus only Add the low abundance score (?)
 1) Taxonomic rank Species only 2) Low coverage (count Include for all metrics Choice and weight of the only 1) Taxa detection recall Important 	 ○ Both of reads < 100)^(?) s ○ Ignore different metrics: ○ Somewhat 	 Genus only Add the low abundance score (?) Not at all
 1) Taxonomic rank Species only 2) Low coverage (count Include for all metrics Choice and weight of the only 1) Taxa detection recall Important 2) Taxa detection precision 	 Both of reads < 100)^(?) Ignore different metrics: Somewhat 	 Genus only Add the low abundance score (?) Not at all

Get the corresponding ranking and scoring of tools

LEMMI fingerprint (?): SD.ALGORITHM.beta01.20200615

Click on each score to display the values underlying it

Rank	Method	Parameters	Reference (?)	Score (?
1	ganon 0.1.0	k=19+fwd reads only	BUILT+GE I RefSeq/08.2018 1rep.	0.31
2	ganon 0.2.3	k=19 paired	BUILT+GE I RefSeq/08.2018 1rep.	0.295
3	ganon 0.2.3	k=19 single	BUILT+GE I RefSeq/08.2018 1rep.	0.279
4	MetaCache 0.5.0	k=22	BUILT+GE I RefSeq/08.2018 1rep.	0.268
5	MetaCache 0.5.0	k=16	BUILT+GE I RefSeq/08.2018 1rep.	0.242
6	Kraken 2.0.7 Protein	k=15	BUILT+GE I RefSeq/08.2018 1rep.	0.207
7	Kraken 2.0.7 + Bracken 2.0	k=35+l=150	BUILT+GE I RefSeq/08.2018 1rep.	0.118
8	Centrifuge 1.0.3	Default	BUILT+GE I RefSeq/08.2018 1rep.	0.113
9	Kaiju 1.6.0	Greedy	BUILT+GE I RefSeq/08.2018 1rep.	0.099
10	CLARK-I	light	BUILT+GE I RefSeq/08.2018 1rep.	0.075
11	CCMetagen	k=16+prefix=TG	BUILT+GE I RefSeq/08.2018 1rep.	0.069

Get the corresponding ranking and scoring of tools

Ci



LEMMI fingerprint (?): SD.ALGORITHM.beta01.20200615

Click on each score to display the values underlying it

Rank	Method	Parameters	Reference (?)	Score (?
1	ganon 0.1.0	k=19+fwd reads only	BUILT+GE I RefSeq/08.2018 1rep.	0.31
2	ganon 0.2.3	k=19 paired	BUILT+GE I RefSeq/08.2018 1rep.	0.295
3	ganon 0.2.3	k=19 single	BUILT+GE I RefSeq/08.2018 1rep.	0.279
4	MetaCache 0.5.0	k=22	BUILT+GE I RefSeq/08.2018 1rep.	0.268
5	MetaCache 0.5.0	k=16	Common	0.242
6	Kraken 2.0.7 Protein	k=15	reference used	0.207
7	Kraken 2.0.7 + Bracken 2.0	k=35+l=150	assessment	0.118
8	Centrifuge 1.0.3	Default		0.113
9	Kaiju 1.6.0	Greedy	BUILT+GE I RefSeq/08.2018 1rep.	0.099
10	CLARK-I	light	BUILT+GE I RefSeq/08.2018 1rep.	0.075
11	CCMetagen	k=16+prefix=TG	BUILT+GE I RefSeq/08.2018 1rep.	0.069

Get the corresponding ranking and scoring of tools

LEMMI fingerprint (?): SD.DEFAULT.beta01.20200615

Click on each score to display the values underlying it

Rank	Method	Parameters	Reference (?)	Score (?)
1	Kraken 2.0.7 + Bracken 2.0	k=35+l=150	Assessing	0.604
2	ganon 0.1.0	k=19+fwd reads only	different	0.589
3	MetaCache 0.5.0	k=16	also an option	0.58
4	ganon 0.2.3	k=19 paired		0.568
5	ganon 0.2.3	k=19 single	BUILT I RefSec/08.2018 1rep.	0.542
6	Kraken 2.0.7 Protein	k=15	BUILT I RefSeq/08.2018 All	0.535
7	MetaCache 0.5.0	k=22	BUILT I RefSeq/08.2018 1rep.	0.494
8	Kraken 1.1 + Bracken 2.0	k=31	BUNDLED Minikraken/8G/2017	0.482
9	MetaPhIAn 2.7.7	Default	BUNDLED mpa_v20_m200	0.465
10	Kraken 2.0.7 + Bracken 2.0	k=35+l=150	BUNDLED Minikraken2/8G/2019	0.459



Including computational resources needs



LEMMI.v1 growth

Our team selects tools, makes them compatible with LEMMI and runs the benchmark

SD	Species detection	Category METHOD ALGOR	ITHMS	
Rank	Method	Parameters	Reference	Score
1	ganon 0.1.0	k=19+fwd reads only	BUILT+GE I RefSeq/08.2018 1rep.	0.31
2	ganon 0.2.3	k=19 paired	BUILT+GE I RefSeq/08.2018 1rep.	0.295
3	ganon 0.2.3	k=19 single	BUILT+GE I RefSeq/08.2018 1rep.	0.279
	to complete ranki	ings >		

C BB | emmi-data.ezlab.org/beta01.20190417/

< >

Index of /beta01.20190417/

/	
bracken1.tar.gz	29-Apr-2019
centrifuge.tar.gz	29-Apr-2019
ganon.tar.gz	29-Apr-2019
kaiju.tar.gz	29-Apr-2019
kraken2.tar.gz	29-Apr-2019
metacache05.16.tar.gz	29-Apr-2019
metacache05.31.tar.gz	29-Apr-2019
metaphlan2.tar.gz	29-Apr-2019

The community can access results and containers and contact us with requests

LEMMI.v2-standalone should facilitate developers entry

Our team keeps integrating tools and pipelines in a public instance of LEMMI.v2

The community of users and developers becomes involved, preparing and sharing compatible pipelines, while benefiting from personal benchmarking

... to complete rankings > lemmi-data.ezlab.org/beta01.20190417/

Index of /beta01.20190417/

../ bracken1.tar.gz centrifuge.tar.gz ganon.tar.gz kaiju.tar.gz kraken2.tar.gz metacache05.16.tar.gz metacache05.31.tar.gz metaphlan2.tar.gz

 $\langle \rangle$

C 88

> 29-Apr-2019 29-Apr-2019 29-Apr-2019 29-Apr-2019 29-Apr-2019 29-Apr-2019 29-Apr-2019 29-Apr-2019

The benchmarking "triangle"

LEMMI.v2 gives you control over these three aspects.

The new LEMMI.v2 allows multiple «scenarios»

Create samples that represent your biological problem Choose a reference that scales with your computational environment

- Taxonomy: GTDB or NCBI
- Host organism
- Contaminant organisms
- Target organisms
- Number of reads
- Read length (for now illumina)
- Number of samples

2021_9_PROK_GIDB	
Bacterial target organ Inspired by dental san	iisms, nples
valuated under the G axonomy	TDB
tools at the moment	
calibration and 4 eva	luation

LEMMI.v2: select and compare a subset of tools

Detailed results for the benchmark 2021_9_PROK_GTDB Taxonomic rank: phylum class order family genus species centrifuge metacache ganon Precision 0.890 0.838 0.612 0.625 0.651 Recall 0.470 **F1** 0.734 0.601 0.631 Runtime (min) 10.563 59.404 59.096 Memory usage (GB) 58.323 39.724 219.684 centrifuge metacache ganon Precision Recall **Comparative Memory Usage** F1 Score **Comparative Runtime Efficiency**

Gweon *et al. Environmental Microbiome* (2019) 14:7 https://doi.org/10.1186/s40793-019-0347-1

RESEARCH ARTICLE

The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples

H. Soon Gweon^{1,2*}, Liam P. Shaw³, Jeremy Swann³, Nicola De Maio³, Manal AbuOun⁴, Rene Niehus⁵, Alasdair T. M. Hubbard³, Mike J. Bowes², Mark J. Bailey², Tim E. A. Peto^{3,6}, Sarah J. Hoosdally³, A. Sarah Walker^{3,6}, Robert P. Sebra⁷, Derrick W. Crook^{3,6}, Muna F. Anjum⁴, Daniel S. Read², Nicole Stoesser^{3*} and on behalf of the

REHAB consortium

Abstract

Background: Shotgun metac the investigation of antimicro method, and reference database used. As the field changes so rapidly and newer methods become available, we strongly recommend that researchers with shotgun metagenomic data review excellent benchmarking efforts such as CAMI [21] and LEMMI [22] and assess the tools using a particular quantitative metric rather than making a (perhaps arbitrary) choice for their analysis. Investigating the robustness of conclusions to choice of method is also a recommended step [23, 24].

Open Access

LEMMI creates a link between method developers and users

Help us to make it happen - use and recommend LEMMI !

Try and give us your feedback

https://lemmi-v2-beta.netlify.app

Chat with the lead developer: Mathieu.Seppey@unige.ch

ezlab.org

C

ĥ ó

Tools

instead of CheckM Biome & AI

- LEMMI 🏆 nove Jenchmarking
- BUSCO scoring core genes
- OrthoDB 🙀 "equivalent" genes
- mirMap 🎡 miRNA targets
 ..more

We focus on metagenomics and machine learning. Genomics is a Big Data field now, and we develop tools to automate analyses. Focusing on: Viruses/phages, bacteria AMR/virulence, Taxonomy.

Genomics

- Dipluran Campodea augens
- Damselfly Calopteryx splendens
- Ice crawler Galloisiana yuasai
- Bristletail Lepismachilis y-signata
- Firebrat Thermobia domestica