CAMI II: identifying best practices and issues for metagenomics software

By providing challenges to the metagenomics community based on complex and realistic metagenome benchmark datasets, CAMI — the community-driven initiative for the Critical Assessment of Metagenome Interpretation — has created a comprehensive assessment of the performance of metagenomics software for common analyses. As part of its second contest, CAMI II, it evaluates ~5,000 submissions from 76 software programs and their different versions.

This is a summary of:
Meyer, F. et al. Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nat. Methods* https://doi.org/10.1038/s41592-022-01431-4 (2022).

Published online:
8 April 2022

The problem

Metagenomics — the study of the collective genomes of microorganisms from environmental samples — provides fundamental insights into the structures and functions of microbial communities. Computational processing is key to metagenomics, and assessing the performance of metagenomics software thoroughly, comprehensively and without bias is thus crucial. Lack of knowledge about the most suitable metagenomics software for a specific application can lead to issues such as lower quality metagenome assemblies and recovered genomes, or a reduced ability to detect the microorganisms present in the sampled community.

The solution

CAMI is an initiative that aims to assess metagenomics analysis software by offering challenges to the community based on comprehensive and realistic benchmarking datasets. These datasets cover typical experimental setups and data-generating techniques for frequently studied environments.

For the CAMI II contest, we offered metagenome datasets representing a marine environment, a plant-associated environment — including fungal and host plant genome sequences — and a high-strain-diversity environment that we called 'strain madness'. Datasets were generated by sampling short and long reads from ~1,700 genomes and ~600 circular elements. More than half of the genomes were newly sequenced and the others were high-quality public genomes. Specific challenges were provided for metagenome assembly software; genome and taxonomic binners, which cluster sequences of the same genome (or taxon); and taxonomic profilers, which infer the identities and relative abundances of microbial taxa present[1]. Further, we offered a clinical pathogen detection challenge based on a short-read metagenome dataset from a patient's blood sample and case report. To enable reproducibility, challenge participants were encouraged to provide both results and the exact software version used, together with all parameter settings and reference databases.

Overall, participants from around the world submitted ~5,000 sets of results from 76 programs and their versions. Analysis of the results identified computationally efficient and well-performing software with regards to key measures (Fig. 1), with a substantial performance improvement compared to software assessed in the CAMI challenge[2,3]. For metagenome assembly, long-read sequences proved particularly valuable for difficult-to-assemble regions such as 16S rRNA genes. Overall assembly quality was shown to depend on preprocessing, genome coverage and the presence of closely related strains. Most metagenome assembly software did not resolve individual strains, in some cases intentionally. The presence of related strains and assembly quality also affected genome binners, which showed variable performances across metrics and data types. For taxonomic profilers, several methods consistently ranked highly across performance measures. Furthermore, several submissions identified the causal pathogen in the clinical pathogen challenge.

The implications

CAMI II has delivered detailed insight into the performances of common types of metagenome analysis software across different datasets and application scenarios. This study will provide guidance for researchers choosing appropriate analysis software and suggests relevant research areas for method developers, such as the taxonomic assignment and profiling of Archaea, viruses and taxa at low bacterial ranks, as well as achieving reproducibility for causal pathogen detection from clinical samples. Metagenomics has great potential for clinical pathogen diagnostics and treatment[4]; however, further assessments are needed, and there are more hurdles to overcome before its clinical application.

Evaluating software performance is a moving target, as progress in methods and data generation is rapid. It will be essential to regularly reassess the state of the art in future challenges and to include further meta-omics data modalities in these efforts.

Alice C. McHardy and Fernando Meyer, Helmholtz Centre for Infection Research, Brunswick, Germany.

EXPERT OPINION

"The authors should be commended for this excellent work, which I am confident will become an indispensable guideline for the community and should substantially enhance the quality of metagenomic analyses." **Nikos Kyrpides, DOE Joint Genome Institute, Berkeley, CA, USA**

FIGURE

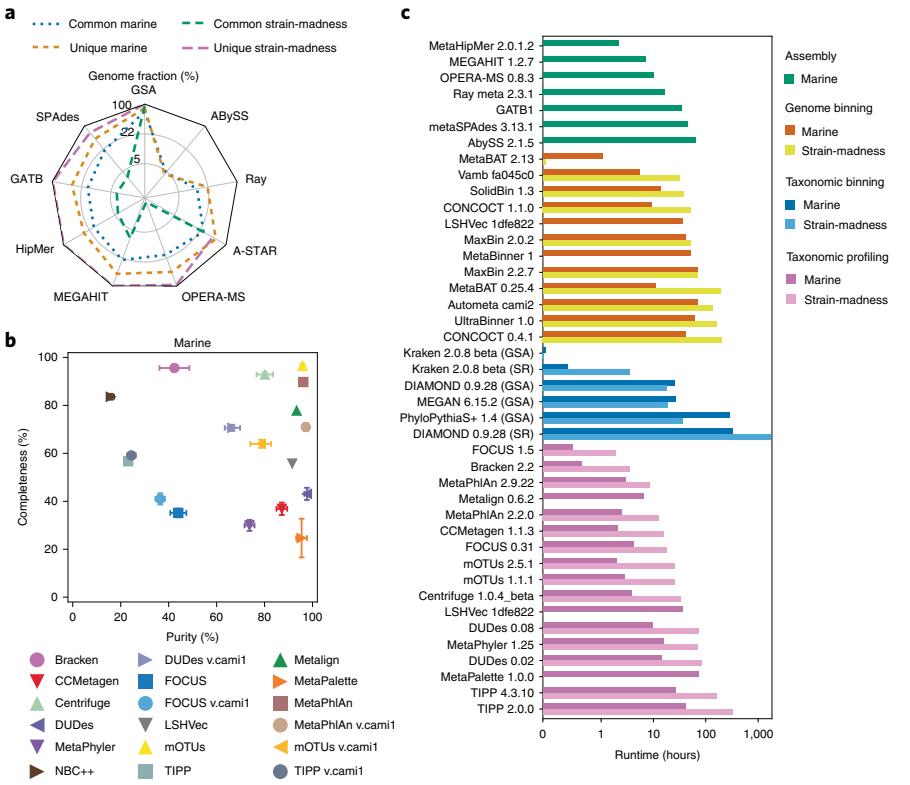

Fig. 1 | CAMI II results for two of five assessed method categories and method runtimes. a, Fraction of genomes covered by assembled contigs, for unique genomes (no other genome with ≥95% average nucleotide identity in the dataset) and common genomes. **b**, Purity versus completeness of taxonomic profiles for the marine dataset at genus rank. **c**, Method runtimes in log scale for all assessed categories.

BEHIND THE PAPER

CAMI was officially founded in 2014 by Thomas Rattei, Alex Sczyrba and myself during a metagenomics program at the Isaac Newton Institute for Mathematical Sciences in Cambridge. From then on, CAMI rapidly grew into a widely supported initiative in the microbiome research community, with many open-to-all events and workshops held for gathering inputs. We thank everyone who has supported and contributed to CAMI and made it to what it is: all scientists who contributed creative ideas, data, thoughts on evaluation measures, inputs on required standards, efforts in running metagenomics software, and code and analyses for interpreting the results; John Toland, the former head of INI, who funded several CAMI follow-up workshops; Mihai Pop, for doing the same in Maryland; and the ISMB conference, which hosts yearly CAMI sessions. **A.C.M.**

REFERENCES

1. Breitwieser, F. P., Lu, J. & Salzberg, S. L. A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* **20**, 1125–1136 (2019). **A Review article outlining common computational methods in metagenomics.**

2. Sczyrba, A. et al. Critical assessment of metagenome interpretation — a benchmark of metagenomics software. *Nat. Methods* **14**, 1063–1071 (2017). **This article reports the results of the first CAMI challenge.**

3. Meyer, F. et al. Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit. *Nat. Protoc.* **16**, 1785–1801 (2021). **A tutorial article demonstrating how to assess metagenomics software in the CAMI framework.**

4. Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat. Rev. Genet.* **20**, 341–355 (2019). **A Review article that outlines the potential and challenges of using metagenomics for clinical applications.**

FROM THE EDITOR

"With an ever increasing number of computational tools for metagenomic analysis being developed, objective and comprehensive benchmarking is needed but requires tremendous efforts. The results reported by the second round of the Critical Assessment of Metagenome Interpretation challenges (CAMI II) benefit the community by assessing the strength and weakness of state-of-the-art methods and guide the direction of future method development." **Editorial Team, *Nature Methods***