Critical Assessment of Metagenome Interpretation Enters the Second Round

Andreas Bremges,[a,b] Alice C. McHardy[a]

aComputational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany

bGerman Centre for Infection Research (DZIF), Braunschweig, Germany

Bioinformatic methods are key components in the analysis of large omics data sets now routinely generated in microbiome research. Methods are evaluated using a large variety of benchmark data sets and evaluation metrics in original research articles, and self-assessments may be biased, which complicates method comparisons from the literature (1). CAMI, the initiative for the Critical Assessment of Metagenome Interpretation, aims to develop community-wide standards and facilitates benchmarking by providing realistic standard data sets and organizing community challenges on a regular basis. It engages the microbiome research community in public workshops and hackathons to design future CAMI challenges and identify the most relevant evaluation metrics. The first CAMI challenge opened in 2015 and ran for 3 months. Developers could assess metagenome assembly, genome and taxonomic binning, as well as taxonomic profiling methods on metagenome benchmark data sets of different complexities, derived entirely from organisms not present in public genome databases (2). All data are available for further benchmarking (https://doi.org/10.5524/100344), together with open-source software to facilitate performing and reproducing evaluations (http://microbiome-cosi.org/cami/resources).

Currently, CAMI is preparing for a second round of challenges, tentatively planned to open later this year. CAMI will provide data sets representing different environments and again offer assembly, taxonomic and genomic binning, as well as taxonomic profiling challenges (Fig. 1). Two multisample "toy" data sets representing microbial communities from different human body sites and from mouse gut are already provided to allow participants to prepare for the challenges (https://data.cami-challenge.org/participate). These practice data sets are generated from known genomes, and therefore reference-based methods (e.g., using genome databases for their analysis) might perform better here than for real shotgun metagenomic data, where a substantial portion of microbial community members have not been sequenced (3, 4). The second CAMI challenge data sets will therefore again include new genomes from taxa (at different evolutionary distances) not found in public databases. Furthermore, a new focus will be on establishing the value of long sequencing reads for microbiome research, with data sets providing both long- and short-read data. Lastly, a clinical pathogen discovery challenge will be offered, mimicking an emergency diagnostic situation in the clinic.

We invite everyone to join the CAMI effort and participate in the upcoming challenges!

Address correspondence to Alice C. McHardy, Alice.McHardy@helmholtz-hzi.de.

The views expressed in this Editorial do not necessarily reflect the views of this journal or of ASM.
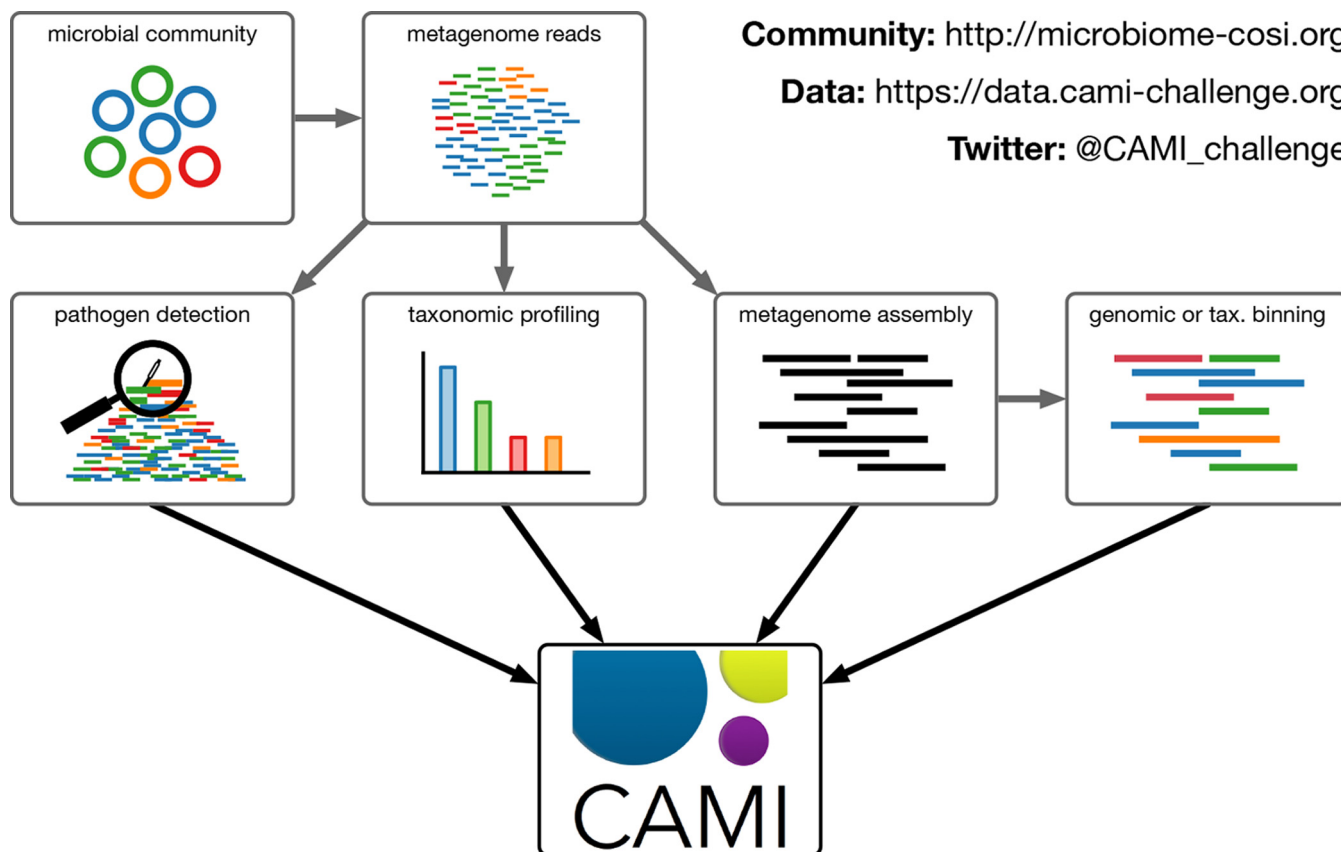
FIG 1 CAMI will provide data sets representing different environments and again offer metagenome assembly, profiling, and binning challenges, as well as a new pathogen detection challenge, which mimics an emergency diagnostic situation in the clinic.

REFERENCES

1. Norel R, Rice JJ, Stolovitzky G. 2011. The self-assessment trap: can we all be better than average? Mol Syst Biol 7:537. https://doi.org/10.1038/msb.2011.70.

2. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, DeMaere MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaitre C, Peterlongo P, Rizk G, Lavenier D, Wu YW, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin HH, Liao YC, et al. 2017. Critical assessment of metagenome interpretation—a benchmark of metagenomics software. Nat Methods 14:1063–1071. https://doi.org/10.1038/nmeth.4458.

3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.

4. Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP, Göker M, Coates RC, Hadjithomas M, Pavlopoulos GA, Paez-Espino D, Yoshikuni Y, Visel A, Whitman WB, Garrity GM, Eisen JA, Hugenholtz P, Pati A, Ivanova NN, Woyke T, Klenk HP, Kyrpides NC. 2017. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. Nat Biotechnol 35:676–683. https://doi.org/10.1038/nbt.3886.