



Critical Assessment of Metagenome Interpretation (CAMI)

Towards a comprehensive, independent and unbiased evaluation of computational metagenome analysis methods

Alice C. McHardy
HHU Düsseldorf, HZI Braunschweig

Why CAMI?

Tool development for shotgun metagenome data sets is a very active area: Assembly, (tax.) binning, taxonomic profiling

- Method papers present evaluations using many different metrics, simulated data sets (snapshots) and are difficult to compare
- It is unclear to everyone which tools are most suitable for a particular task and for particular data sets
- Comparative benchmarking requires extensive resources and there are pitfalls



Critical Assessment of Metagenome Interpretation (CAMI)

- Initial targets: Assembly, (tax.) binning and profiling
- Extensive simulated data sets will be provided
- Competition scheduled to open in **late 2014**
- Standards, overview of toolscape and different use cases, facilitate future benchmarking, indicate promising directions for development, suggestions for experimental design
- Publication with participants and data contributors

www.cami-challenge.org
[CAMI Google+-Group](#)



CAMI Events

- **March, Cambridge:** Meeting & discussion at MTG meeting at Newton Institute
- **August 25th, Seoul:** Roundtable at **ISME 2014** to decide on evaluation details and dissemination of results
- **September, Cambridge:** Hackathon at Newton Institute
- **End of 2014:** Tentative start of competition
- **Early 2015:** Evaluation meeting

Important contest principles

- Data sets need to be as realistic as possible
- Evaluation measures should be informative to developers and understandable also by applied community
- Reproducibility (provide scripts after the competition, describe data generation procedures)
- Participants should not see any of the data before



CAMI contributors

Contributors: Stephan Majda, Eddy Rubin, Nikos Kyrpides, Paul Schulze-Lefert, Nicole Shapiro, Tanja Woyke, Hans-Peter Klenk, Johannes Droege, Ivan Gregor, Peter Hofmann, Eik Dahms, Jessika Fiedler, Ruben Garrido-Oter, Julia Vorholt, Yang Bai, Girish Srinivas, Phil Blood, Mihai Pop, Aaron Darling, Matthew DeMaere, Dmitri Turaev, Chris Hill, Peter Belmann, Andreas Bremges

And you?

Steering committee: Alice McHardy, Alexander Sczyrba, Thomas Rattei



www.cami-challenge.org

CAMI Workpackages

WP1: Benchmark data set generation (McHardy, Rattei)

WP2: Assembly evaluation (Sczyrba, Pop)

WP3: (Taxonomic) binning evaluation (McHardy)

WP4: Taxonomic profiling evaluation (Koslicki, McHardy)

WP5: Runtime benchmarking framework (Blood,..)

WP6: Up- and download sites (Rattei, Blood, Sczyrba)

WP1 - Benchmark data sets

- Challenging and as realistic as possible
- Samples of high complexity, medium complexity communities (differential abundance, time series, single sample)
- Strain level variation
- Different taxonomic distances to sequenced genomes (deep branchers included)
- Simulate Illumina and PacBio reads from unpublished assembled genomes
- Distribute simulated metagenome samples (unassembled, later then the perfect assembly), along with one NCBI taxonomy version and reference collection

Simulated Metagenome Sample Generation

