CAMI

Critical Assessment of Metagenome Interpretation (CAMI)

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

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

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

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 compeitition, describe data generation procedures)
- Participants should not see any of the data before

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

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,.)

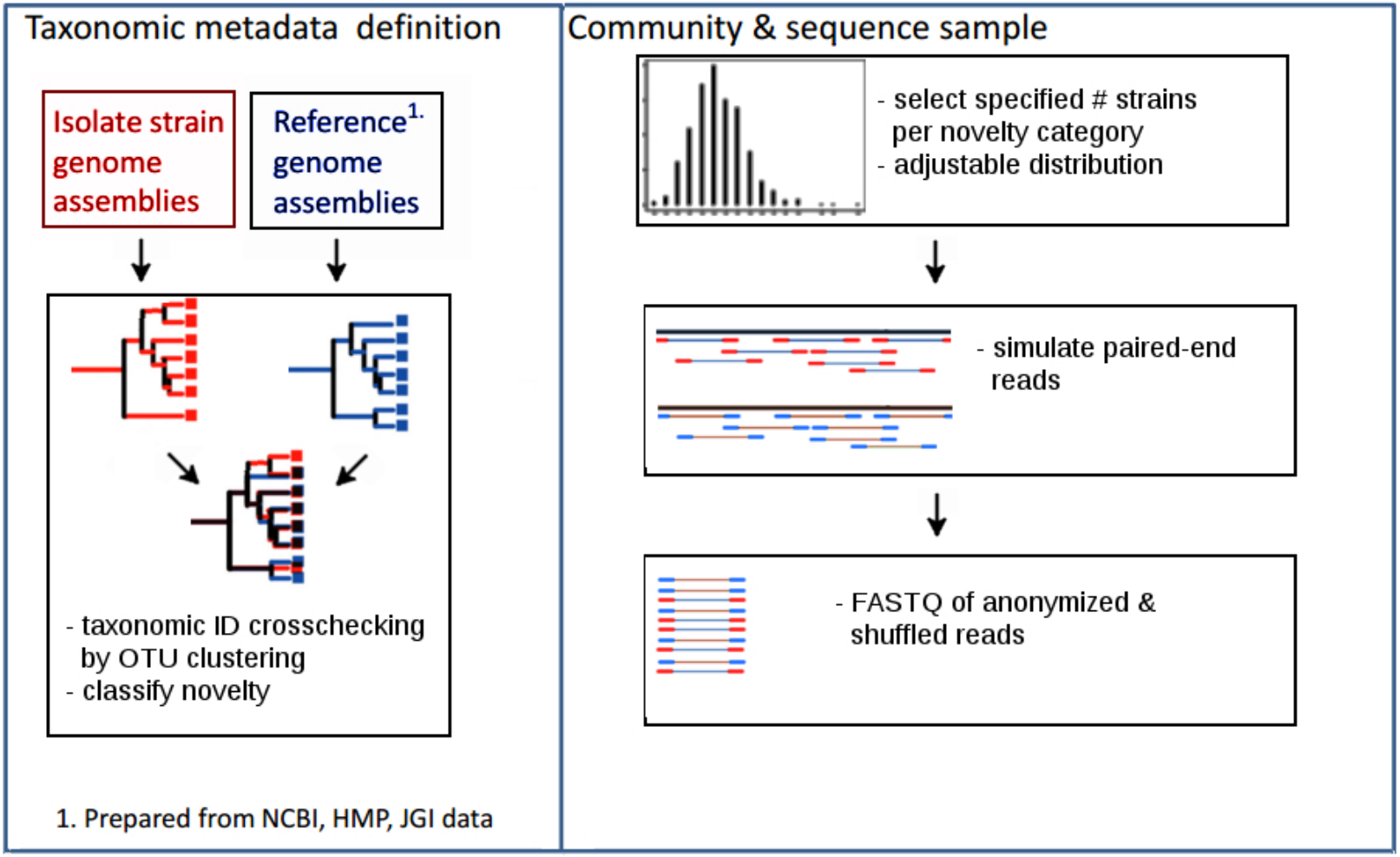
WP6: Up- and download sites (Rattei, Blood, Sczyrba)

WP1 - Benchmark data sets

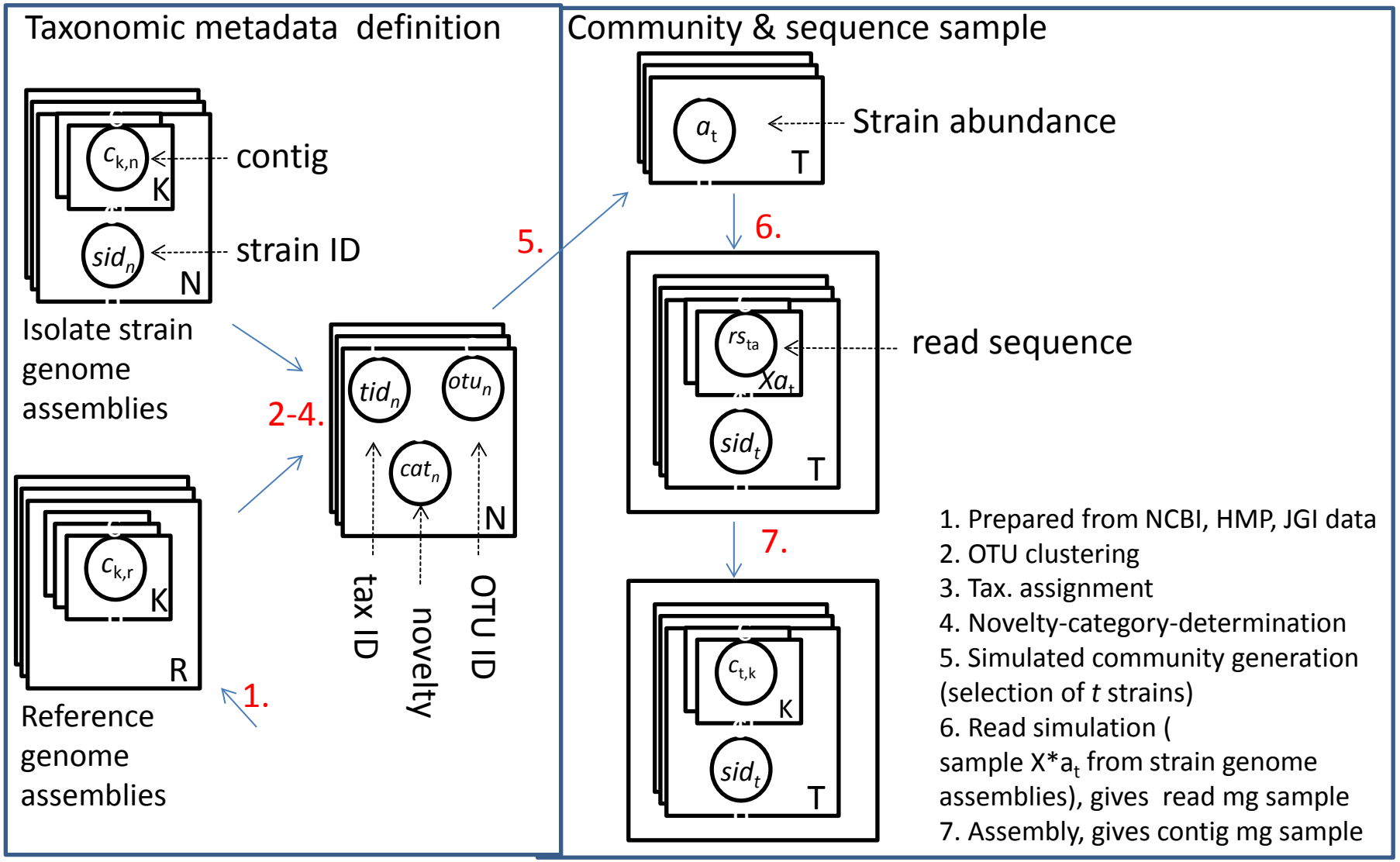
- Challenging and as realistic as possible
- Samples of high complexity, medium complexity, low complexity communities (differential abundance, time series, single sample – with diff. Insert sizes?)
- Strain level variation
- Different taxonomic distances to sequenced genomes (deep branchers included)
- Simulate Illumina **HiSeq**/ MiSeq /PacBio reads from unpublished assembled genomes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

WP1

Simulated Metagenome Sample Generation



Simulated Metagenome Sample Generation



WP1: Status

- Generated first simulated sample data sets (MiSeq Illumina, paired-end)
- Automatic checking of taxonomic assignments of genomes being implemented
- Time series simulations being in development
- Genome data: In process of being generated

WP2: Assembly measures

- Reference-independent measures (M.Pop)
- Reference-dependent measures (A.Sczyrba, P.Belmann)
- Provide assemblies for binners after assembly part has ended (A. Sczyrba will create gold standard assemblies with 1.1. mapping of unassembled reads to it)

WP2: Status (AS, PB)

- Reference-dependent pipeline:
  - Align contigs to references to calculate specific metrics
  - Basic statistics (#contigs, N50, total length, longest contig, …)
  - Misassemblies (relocations, translocations. Inversions, …)
  - Genome statistics (fraction assembled, #genes, …)
- *Currently testing pipelines:*
  - *QUAST [1]*
  - *Feature Response Curves [2]*
- *Run-time issues*
  - *Computational costs very high*
  - *Working on scaling up pipelines*
- *Open question: how to present results? (dynamic vs. static plots, webpage, etc)*

[1] Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29(8), 1072–1075. doi:10.1093/bioinformatics/btt086
[2] Vezzi, F., Narzisi, G., & Mishra, B. (2012). Feature-by-Feature – Evaluating De Novo Sequence Assembly. PloS One, 7(2), e31002. doi:10.1371/journal.pone.0031002

WP2: Status (MP, CH)

- Reference-independent pipeline:
  - Align reads to assemblies to calculate specific contig coverages.
  - Compare **assembly likelihoods** using LAP [1,2].
  - Find regions of potential misassemblies within contigs:
    - Large variances in **depth of coverage**.
    - **Incorrect mate-pair fragment size** using REAPR [3].
      - REAPR is run independently on bins of contigs with similar coverage.
    - Align portions of singleton (unaligned) reads to find potential **breakpoints.**
  - Regions are outputted in GFF format.
- *Currently testing pipeline.*

[1] Ghodsi et al. "De novo likelihood-based measures for comparing genome assemblies." BMC research notes 6.1 (2013): 334.

[2] Hill et al. "De novo likelihood-based measures for comparing metagenomic assemblies." BIBM 2013.

[3] Hunt et al. "REAPR: a universal tool for genome assembly evaluation." Genome Biol 14 (2013): R47.

WP3 - (Taxonomic) binning measures

- CAMI output format, version 1 with specification (initial version up for discussion in google+ group)
- Taxonomy used will be specified for contest
- Collect evaluation scripts realizing different measures
  - (macro-) precision and –recall, accuracy for different ranks, taxonomy-based measures (earth movers distance, i.e. UniFrac, and similar), measures of bin consistency (taxonomy-aware, or not)

WP4 - Taxonomic profiling measures

- Allow assessment of tools such as QUIKR, Metaphyler
- Return a vector of taxon abundances as results
- Collect evaluation scripts from developers (precision /recall, rank correlation, others..)
- CAMI profiling format defined

Comments

- Many tools also use different taxonomies from NCBI taxonomy, must be mapped to provided ref. Taxonomy for contest.
- Taxonomy is in output format

WP5: Runtime benchmarks (optional)

- Run at the PSC, already installing a list of tools, also host data there (Phil Blood)
- Integrate evaluation measures and tools into joint framework, run tools within VMs (dockers)
- ANL offered to host runs (Folker Meyer)

Initial list of tools to benchmark / installed at PSC

- Assembly: MetaVelvet, IDBA-DU, SOAP-denovo, Velvet, TitusBrowns pipeline before assembly, RayMeta,….

- Binning: Megan (blastx/RAPSEARCH/PAUDA vrs nr), CARMA, CONOCT, PPS+, taxator-tk (vs. nr/nr_genomes), GroopM, Phymm(BL),…

WP5: Status (PB)

- Many of the assembly codes already available at PSC
- Need to install binning codes
- Use representative data set to benchmark complete pipeline(s) and assess computational requirements

WP6: Up-/Download site

- Website www.cami-challenge.org
- Download site (ready until mid-October)
  - AMultiple download sites (host data at PBC/Bielefeld
  - user needs to sign agreement to not distribute data before getting access, participants remain anonymous, download data with key (A. Sczyrba)
- Upload site (until mid-october)
  - No feedback on results during competition, but repeated submission possible
  - Parse check of results (report format errors), i.e. how many predictions were submitted (validator scripts)