Critical Assessment of Metagenome Interpretation –
the second round of community-driven challenges

Alice C. McHardy
Computational Biology of Infection Research
Helmholtz Centre for Infection Research

and the CAMI Initiative

Computational Metagenomics

Sequencing

GAGAGATATTTTAGACCCACGATGATCC
AGCGATTTGACACGATGATAG
AGCGATTTGACACGGAGAGTATAGAA
AGCGATTTGTAGTATATCGAGGGG

Assembly

[...]sdATGACGATTCCGAAACACGAGCCGCGCGAGGTCTTCGATCGGGCGATCGAGCATACGCGGGCGCTTTCCC
CGGCCACTTTCGATCAGTGGTTCGGGGGAGTTCAGTTCGATGATCTGACCGACGGCGTCCTGACGCTGCG
CGAGCAGCAGCGCGCCATAGCGCGCCAGCACGTCGCGGGCCCGCGCCATGCGCCTCGCCGCGTCGTCCTC
GGATCCCGTGAAGATGAGCACGAGCGCCGTGACCTCGCGCCGCTCGC

CCTCGACGGGGGTGCGCTCGTTGGCCCCGTTCGGCTCCTCTTGGAGCACGGCGCTCGCCTGGATCTCGGC

GCTCTTCTCCTCGTGGAACGGCTCCAGGAAATCGGCGAGGTCGTTCGCCCCGAACCGCTCGGCCGTCGCG
TAGAAGAAGCCGAGGAGCGCCTCGTGCAGCCGCCCGGCGTCGGGGAACCGATCCTCCGGGTTCTTCGCGA
GCATCGTGCTCACGACGTCGGCGAGCGACCGGGGCACGTCGGGGCGCACGAGCGCGAGCGGCGGATACTC

GCTCGCCTGCAGCCGCCGCATGATCTCGAACGCCGTCGGCGCCGCGAACGGGTTCGACCCGGCGAGCATC
GATGGTCGTGTTCCGCGGGATGAGCTTGAAGCTCACGCCGCCGCCGGTCTCGACGCCGAGCGACAGCGGC

GTCACGTCGAGGAGCAGCACCTCGTCCACCTGCCCGGAGAGCGCGGCGCCCTGCAGCGCAGCGCCGACCG
CGACGACCTCGTCGGGGTTGACGCCCTTGTTCGGCTCCCGGCCGAAGAACTCGCGCACCGCCGCCTGCAC
CGCGGGCATCCGGGTCATGCCGCCGACGAGGACCACCGTGTTGACGGCGGAGACGGGGAGCTTGGCGTCC
CCGAGCGTCGCGCGACAGACGTCGATGGTCCGCCGGATGAGGCCCTCGCAGAGCATCTCGAGCTCGTTGC
GCCGCATCGTCCGCTCGAGGTGGAGCGGCCCTCCGCCCGGGCCGACGGCGATGAAGGGGATGTTGATCTC

GGTCTCGAGCGACGACGAGAGCTCGTGCTTCGCCTTCTCGGCCGCCTCCTTGAGCCGCTGCAGCGCCATG
CGATCCCGGCGCAGATCGATGCGGCTCTTCCCCTCGAACTCGTCGGCGAGCAGGTCGATGATCCGCTGGT
CGAAGTCCTCGCCGCCGAGGTGCGTGTCGCCGCCGGTCGCCTTGACGCTGAAGACCCCGCTCGCGATCTC
GAGGATCGAGATATCGAACGTGCCTCCGCCGAGATCGTAGACCGCGATCGTCTCGGCCTTCACCTTGTCG [...]

Genome & Taxonomic 'binning'

[...]sdATGACGATTCCGAAACACGAGCCGCGCGAGGTCTTCGATCGGGCGATCGAGCATACGCGGGCGCTTTC
CC CGGCCACTTTCGATCAGTGGTTCGGGGGAGTTCAGTTCGATGATCTGACCGACGGCGTCCTGACGCTGCG
CGAGCAGCAGCGCGCCATAGCGCGCCAGCACGTCGCGGGCCCGCGCCATGCGCCTCGCCGCGTCGTCCTC
GGATCCCGTGAAGATGAGCACGAGCGCCGTGACCTCGCGCCGCTCGC

CCTCGACGGGGGTGCGCTCGTTGGCCCCGTTCGGCTCCTCTTGGAGCACGGCGCTCGCCTGGATCTCGGC

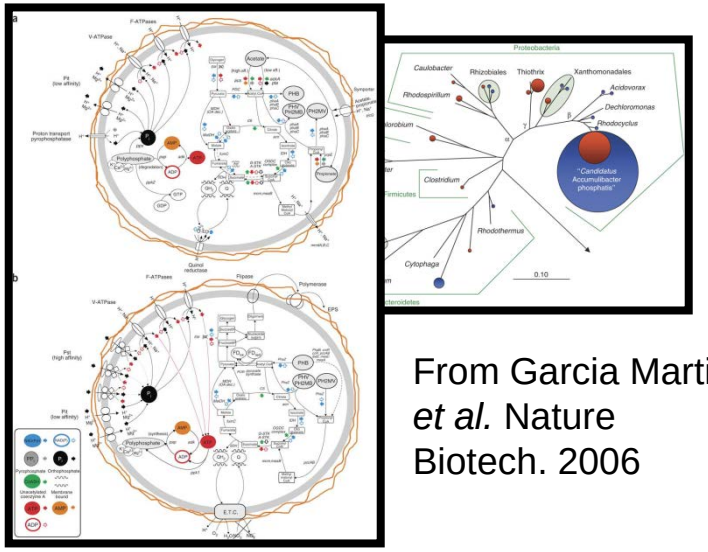
GCTCTTCTCCTCGTGGAACGGCTCCAGGAAATCGGCGAGGTCGTTCGCCCCGAACCGCTCGGCCGTCGCG
TAGAAGAAGCCGAGGAGCGCCTCGTGCAGCCGCCCGGCGTCGGGGAACCGATCCTCCGGGTTCTTCGCGA
GCATCGTGCTCACGACGTCGGCGAGCGACCGGGGCACGTCGGGGCGCACGAGCGCGAGCGGCGGATACTC

GCTCGCCTGCAGCCGCCGCATGATCTCGAACGCCGTCGGCGCCGCGAACGGGTTCGACCCGGCGAGCATC
GATGGTCGTGTTCCGCGGGATGAGCTTGAAGCTCACGCCGCCGCCGGTCTCGACGCCGAGCGACAGCGGC

GTCACGTCGAGGAGCAGCACCTCGTCCACCTGCCCGGAGAGCGCGGCGCCCTGCAGCGCAGCGCCGACCG
CGACGACCTCGTCGGGGTTGACGCCCTTGTTCGGCTCCCGGCCGAAGAACTCGCGCACCGCCGCCTGCAC
CGCGGGCATCCGGGTCATGCCGCCGACGAGGACCACCGTGTTGACGGCGGAGACGGGGAGCTTGGCGTCC
CCGAGCGTCGCGCGACAGACGTCGATGGTCCGCCGGATGAGGCCCTCGCAGAGCATCTCGAGCTCGTTGC
GCCGCATCGTCCGCTCGAGGTGGAGCGGCCCTCCGCCCGGGCCGACGGCGATGAAGGGGATGTTGATCTC

GGTCTCGAGCGACGACGAGAGCTCGTGCTTCGCCTTCTCGGCCGCCTCCTTGAGCCGCTGCAGCGCCATG
CGATCCCGGCGCAGATCGATGCGGCTCTTCCCCTCGAACTCGTCGGCGAGCAGGTCGATGATCCGCTGGT
CGAAGTCCTCGCCGCCGAGGTGCGTGTCGCCGCCGGTCGCCTTGACGCTGAAGACCCCGCTCGCGATCTC
GAGGATCGAGATATCGAACGTGCCTCCGCCGAGATCGTAGACCGCGATCGTCTCGGCCTTCACCTTGTCG [...]

Annotation



From Garcia Martin *et al.* Nature Biotech. 2006

Bacteria 0.7

Archaea 0.3

Proteobacteria 0.2

Firmicutes 0.05

…

Profiling

NATURE METHODS | **METHAGORA**

The Critical Assessment of Metagenome Interpretation (CAMI) competition

27 Jun 2014 | 6:36 PM | Posted by Tal Nawy | Category: Bioinformatics, Computational, Guest Post, Metagenomics

Alice McHardy, Alex Sczyrba and Thomas Rattei announce a new initiative for assessing metagenomics methods in this guest post.

Alice McHardy
FOLKER MEYER

Alex Sczyrba
A. SCZYRBA

Thomas Rattei
ANJA VENIER

In just over a decade, metagenomics has developed into a powerful and productive method in microbiology and microbial ecology. The ability to retrieve and organize bits and pieces of genomic DNA from any natural context has opened a window into the vast universe of uncultivated microbes. Tremendous progress has been made in computational approaches to interpret this sequence data but none can completely recover the complex information encoded in metagenomes.

A number of challenges stand in the way. Simplifying

Towards a comprehensive and objective evaluation of metagenomics software

HELMHOLTZ CENTRE FOR INFECTION RESEARCH

CASP 1 (1994)
Critical Assessment of Techniques for Protein Structure Prediction

PROTEINS: Structure, Function, and Genetics 23:301–317 (1995)

A Critical Assessment of Comparative Molecular Modeling of Tertiary Structures of Proteins*

Steven Mosimann, Ron Meleshko, and Michael N.G. James
Medical Research Council of Canada, Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

ABSTRACT **In spite of the tremendous increase in the rate at which protein structures are being determined, there is still an enormous gap between the numbers of known DNA-derived sequences and the numbers of three-dimensional structures. In order to shed light on the biological functions of the molecules, researchers often resort to comparative molecular modeling. Earlier work has shown that when the sequence alignment is in error, then the comparative model is guaranteed to be wrong. In addition, loops, the sites of insertions and deletions in families of homologous proteins, are exceedingly difficult to model. Thus, many of the current problems in comparative molecular modeling are minor versions of the global pro-**

there are several commercial and public domain computer programs that have been developed for modeling; these programs remove much of the tedium from the process. There are numerous reasons for constructing comparative molecular models of proteins. The molecular model may explain the structural basis of existing experimental results and can provide one with structural information on which further experiments can be planned, executed, and evaluated. Site-specific mutations of the gene coding for the specific protein can provide important data regarding the protein's function. Perhaps, some of the most revealing experiments are those designed to predict and to probe the molecular reasons for an enzyme's specificity.[3] On a more practica molecular model can sometimes be used successfully

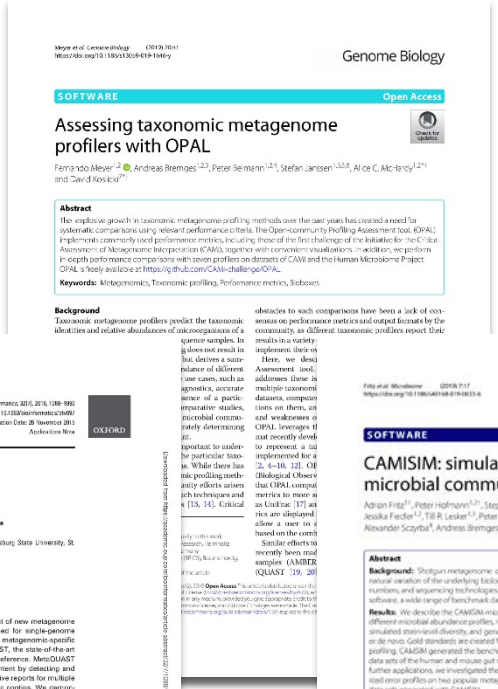
HELMHOLTZ CENTRE FOR INFECTION RESEARCH

CAMI Principles

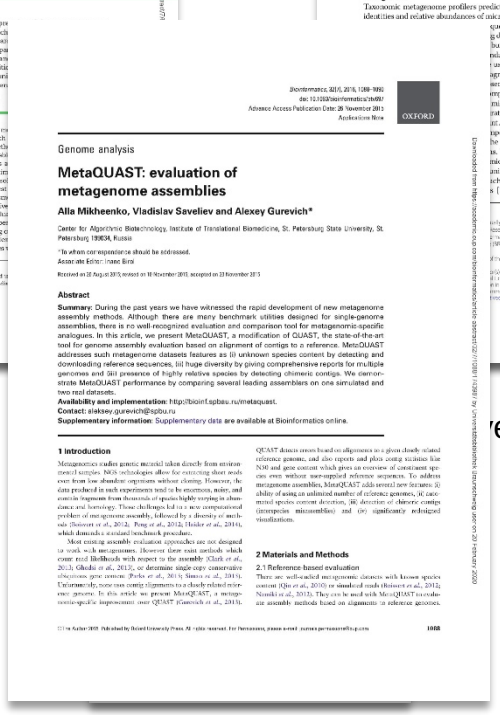
- Design decisions made by the community (datasets, evaluation measures, principles)
- Extensive, high-quality benchmark datasets from unpublished data
- Evaluation measures: informative to developers and applied researchers
- Reproducibility: data generation, programs, evaluation
- Benchmark common method categories

CAMI Benchmarking Software

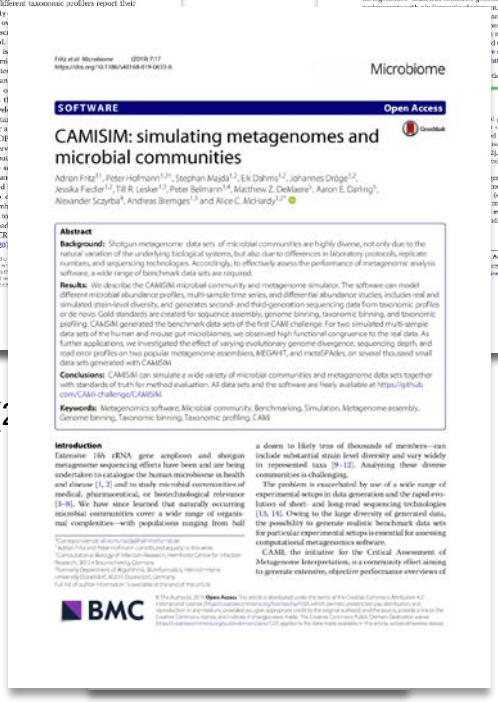
Meyer *et al.* (2018)



Bremges *et al.* (2020)

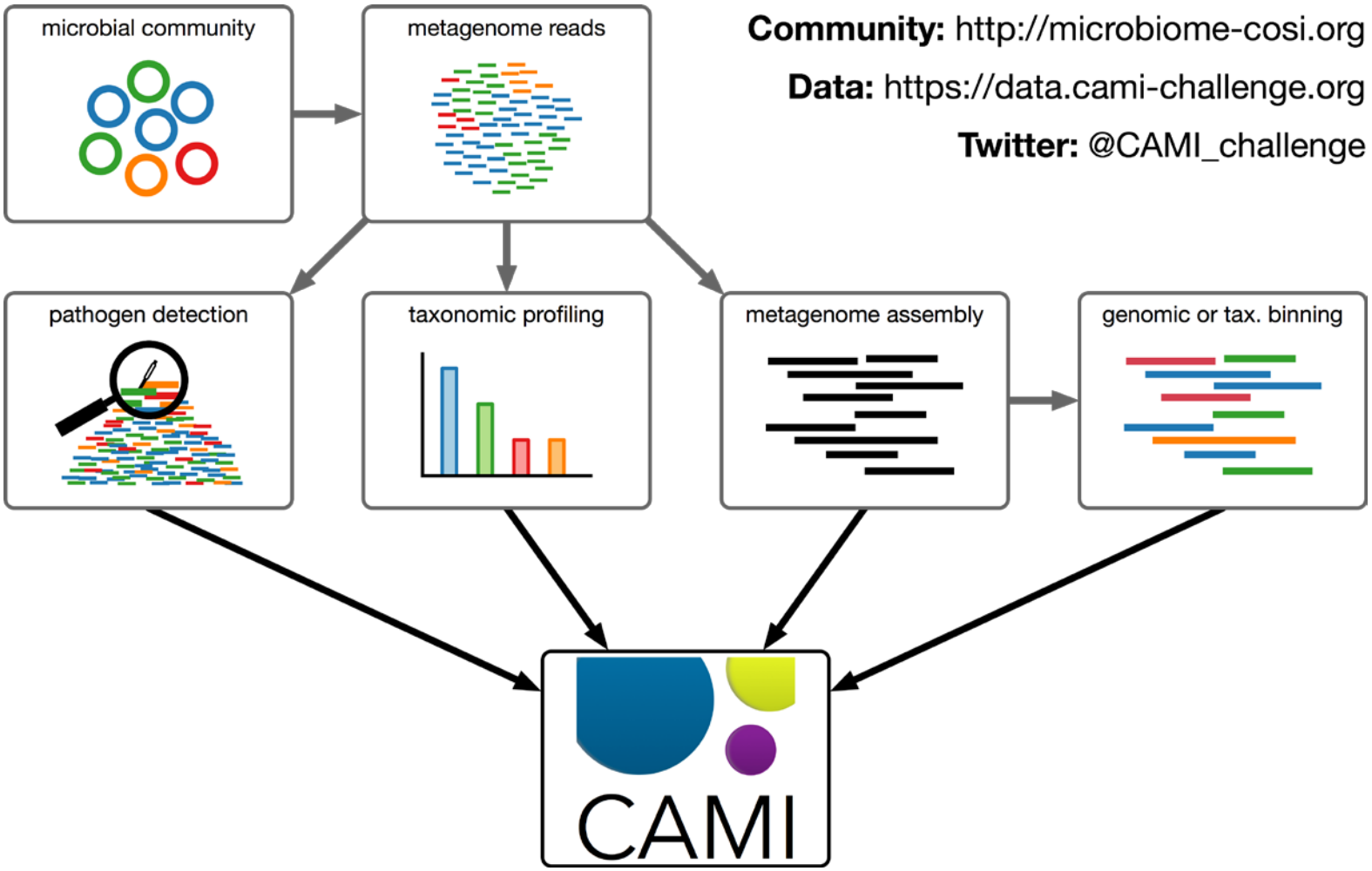


Mikheenko *et al.* (2016)



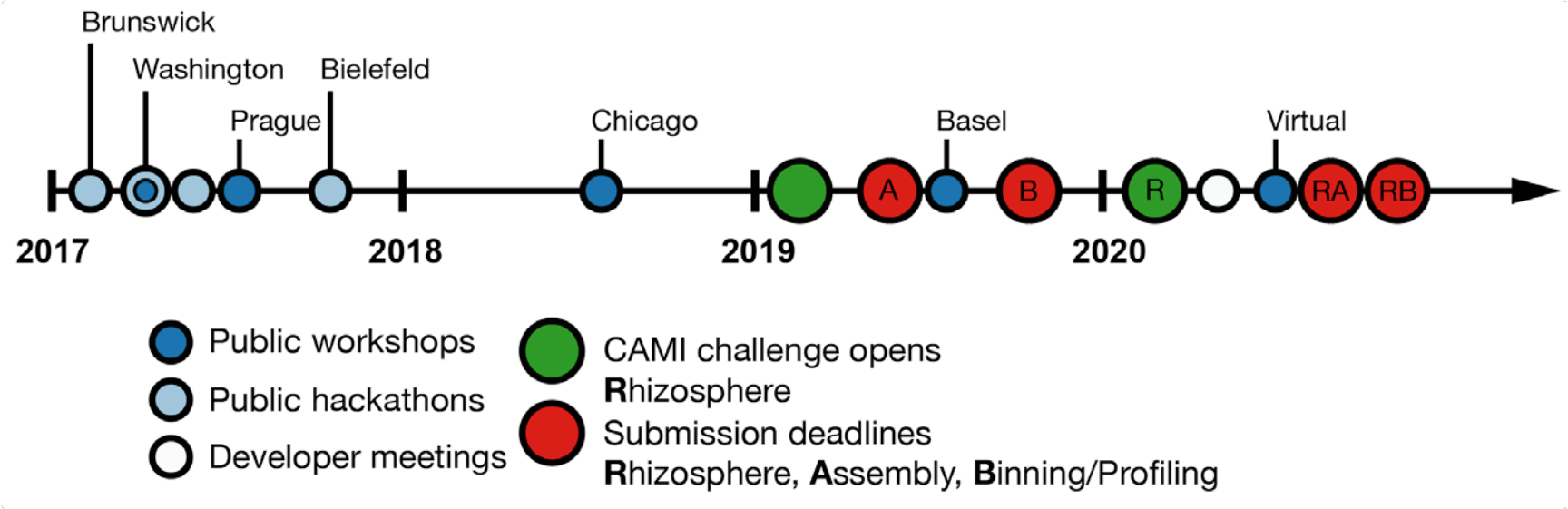
Fritz *et al.* (2019)

CAMI II Challenges



Meyer, Fritz *et al.,* Nature Methods, 2022

Timeline: CAMI II Challenge



https://data.cami-challenge.org/

http://microbiome-cosi.org/cami/participate/schedule

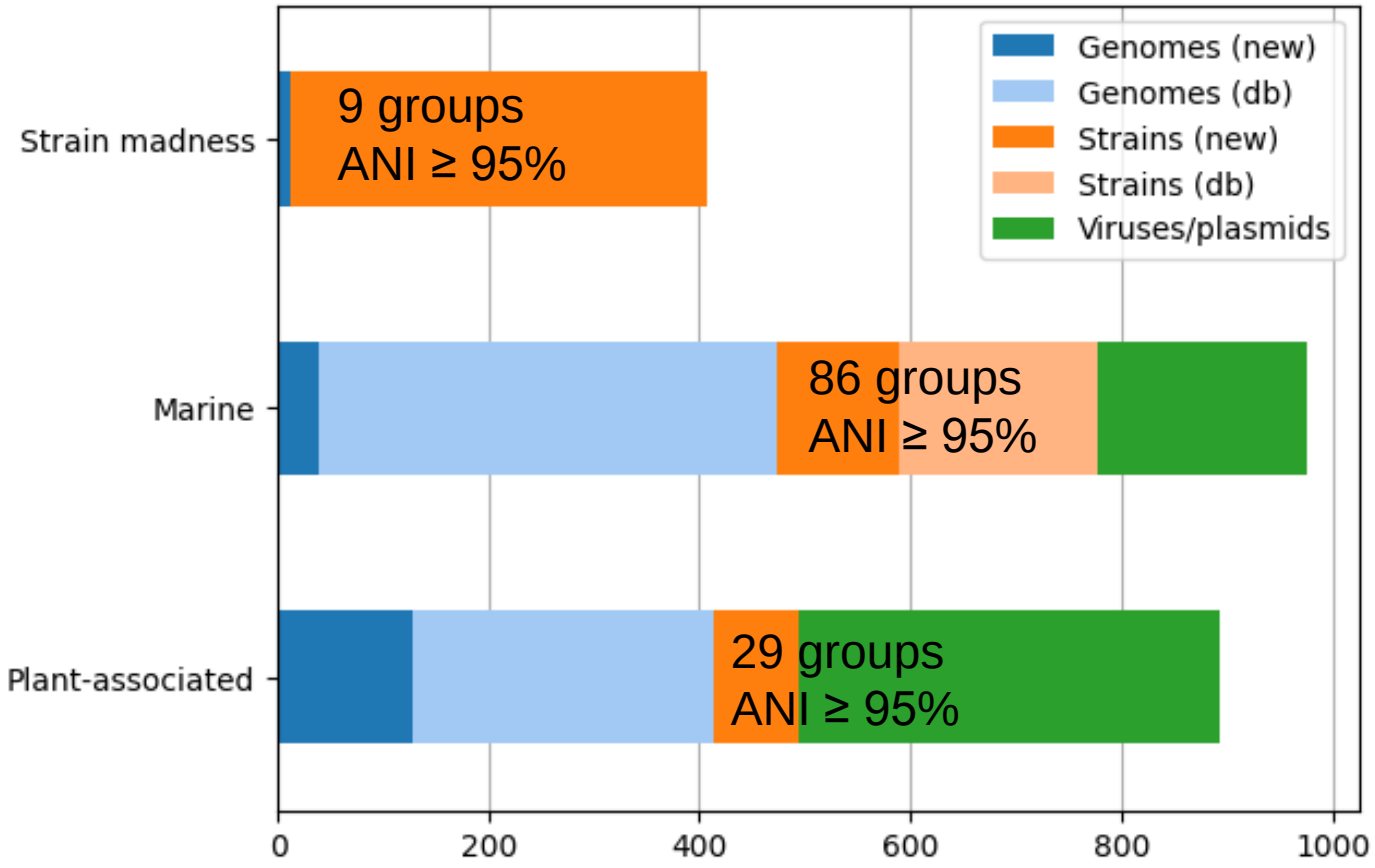
CAMI II Challenge Datasets

- Unpublished and public genomes
- Multi-sample (10-100), long and short reads
- Environment specific (Marine, Plant-associated)
- „Strain madness“, Clinical pathogen detection

Strain madness	Marine environment	Plant-associated
2 x 100 samples	2 x 10 samples	3 x 21 samples
2 x 200 Gb	2 x 50 Gb	3 x 105 Gb
Read lengths 2 x150 bp (Illumina) 7400 bp (PacBio)	Read lengths 2 x150 bp (Illumina) 7400 bp (PacBio)	Read lengths 2 x150 bp (Illumina) 7400 bp (PacBio) 3000 bp (ONP)

Simulated from ~1,680 microbial genomes (incl. fungi), plant host, 599 novel viruses, plasmids and other circular elements

CAMI II Challenge Datasets



Challenge Participants

Early 2015: >40 registered participants;
Early 2020: >350 registrations

https://data.cami-challenge.org

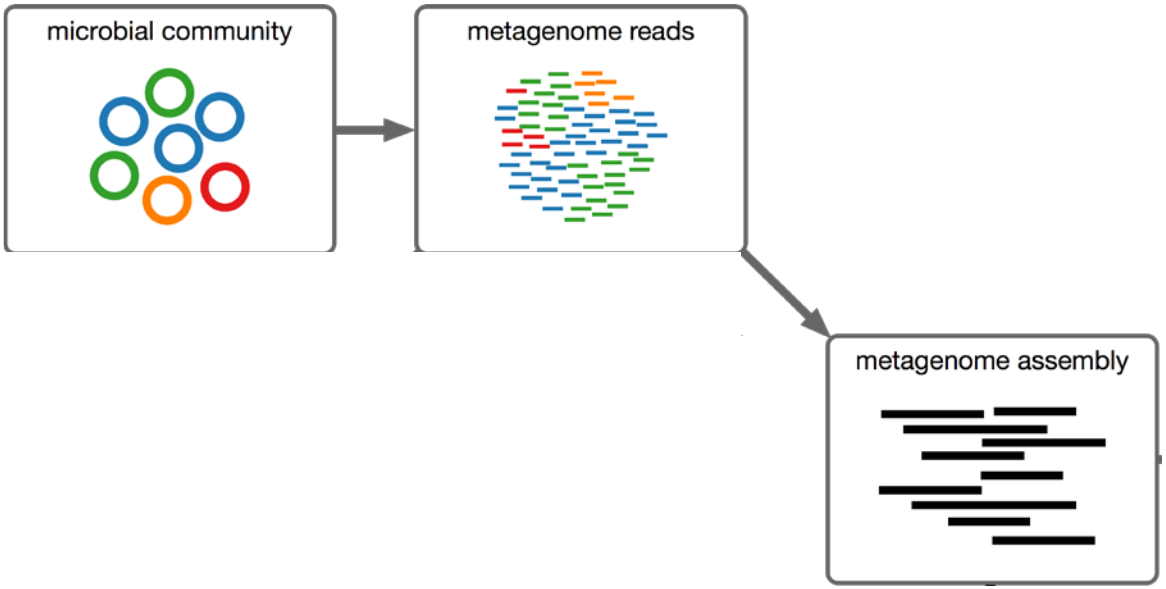
HELMHOLTZ CENTRE FOR INFECTION RESEARCH

CAMI II Challenge Submissions

Assembly, binning, taxonomic binning, profiling, pathogen detection

30 teams; 5,002 results; 76 program versions

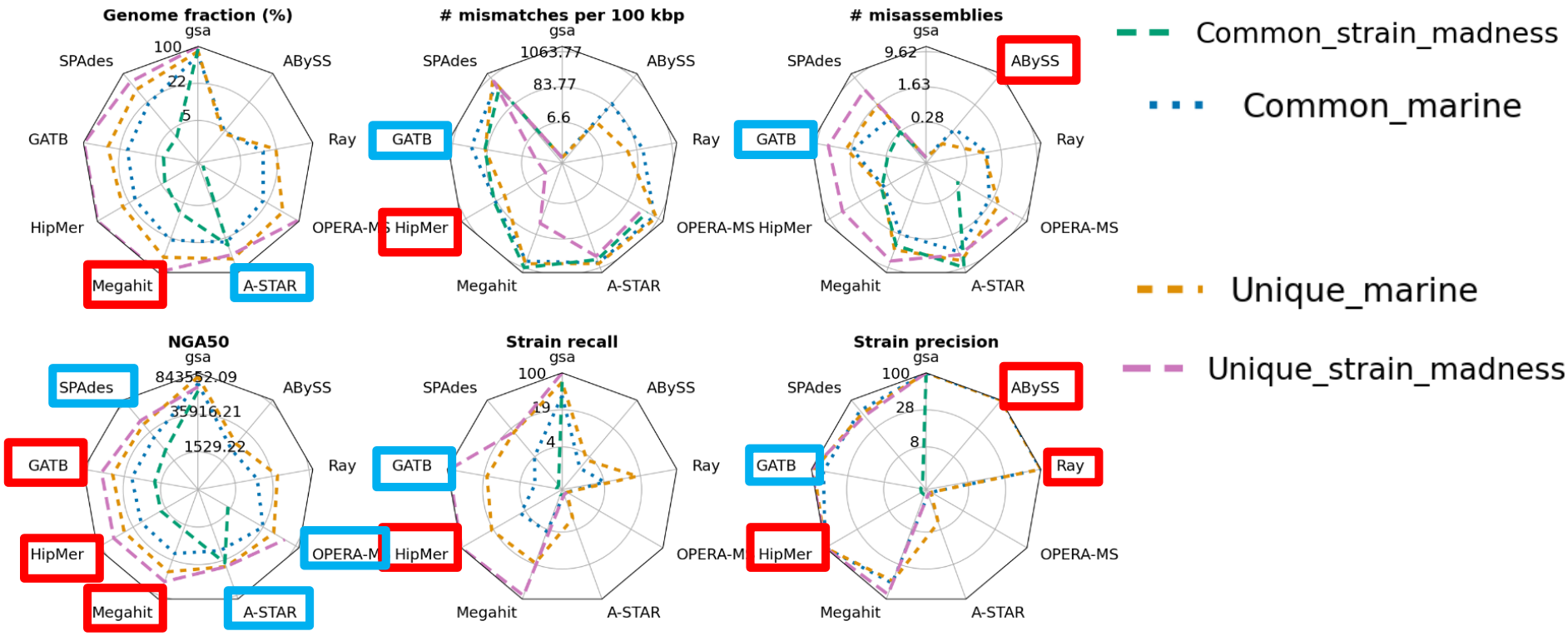
	Submissions	Reference	Reproducibility
Kraken v2.0.8-beta + Bracken v2.2/2.5	2	Lu J et al. PeerJ Computer Science 2017	Requires manual search
LSHVec	1	Shi L, Chen B. bioRxiv 2019	
MetaPhlAn v2.2.0/2.9.14	2	Segata N et al. Nature Methods 2012	Requires manual search
MetaPhyler v1.25	1	Liu B et al. BMC Genomics 2011	Requires manual search
NSSAC	2	Porter W	
Pathoscope v2.0.7	1	Owen E et al. Genome research 2013	Requires manual search
CCmetagen v 1.1.3	1	Vanessa R. M. et al. Genome biology 2020	Requires manual search



ASSEMBLY CHALLENGE

HELMHOLTZ CENTRE FOR INFECTION RESEARCH

Assembly: common and unique strains



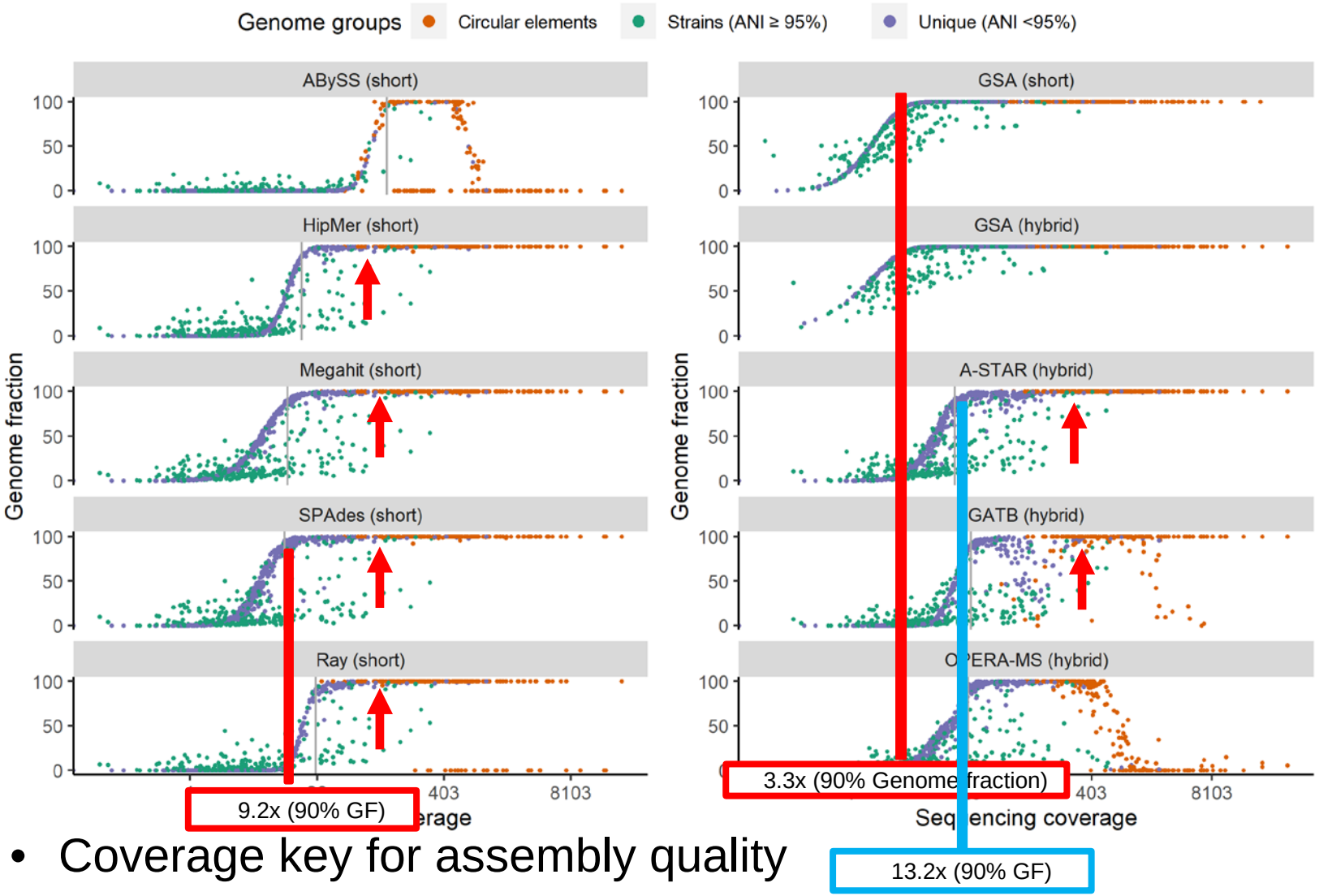
Assemblies better for **unique** than **common** genomes

Different methods have different strengths

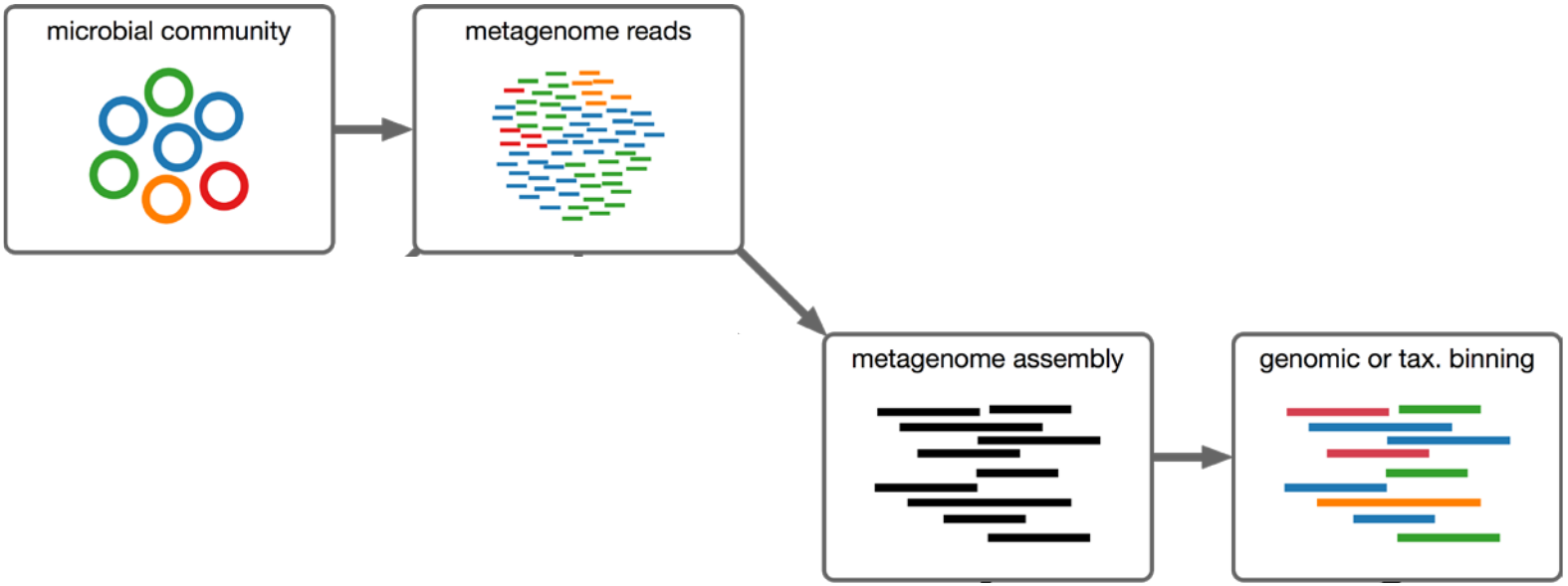
HipMer best across metrics: marine, common marine genomes

GATB for strain madness, common strain madness genomes

Assembly: coverage



- Coverage key for assembly quality
- Many methods assemble high-copy circular elements well



GENOME BINNING CHALLENGE

High quality genome bins

Genome binner	Contamination	>90% completeness			
		Marine		Strain madness	
		gsa	ma	gsa	ma
Gold standard		780	779	408	408
MetaWRAP	<10%	254	**92**	43	9
UltraBinner	<10%	**293**	77	48	**10**
MetaBinner	<10%	232	77	51	8
Vamb	<10%	266	72	21	9
MaxBin	<10%	198	59	29	8
CONCOCT	<10%	84	48	**82**	**10**
Autometa	<10%	197	35	3	2
MetaBAT	<10%				6
SolidBin	<10%			34	
LSHVec	<10%			0	

37%

12%

2%

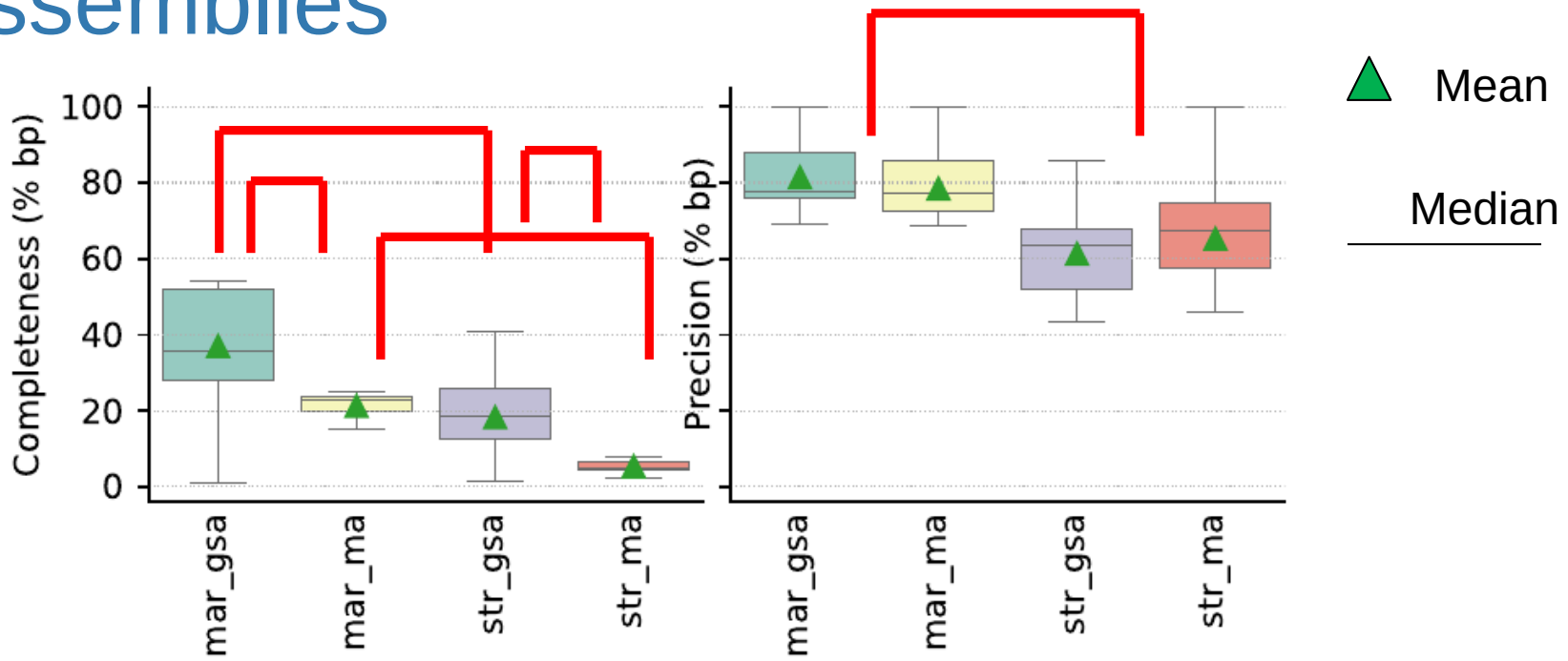
20%

gsa: gold standard assembly
ma: Megahit assembly

#recovered genomes:	100%	50%	0%

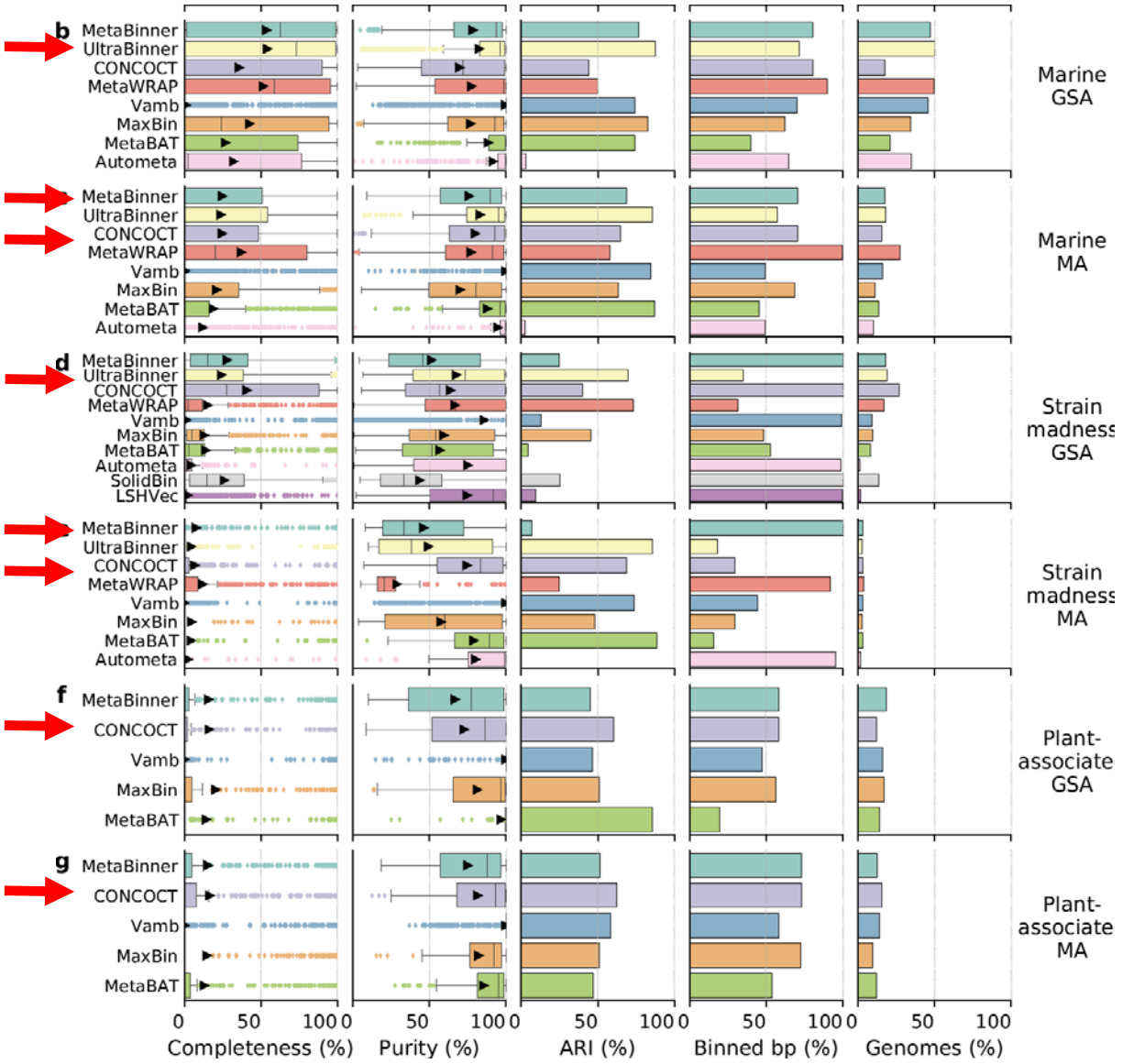
- related to assembly quality and strain diversity
- **18.5x** more recovered from marine gsa than strain madness ma

Genome Binning: gold standard and real assemblies

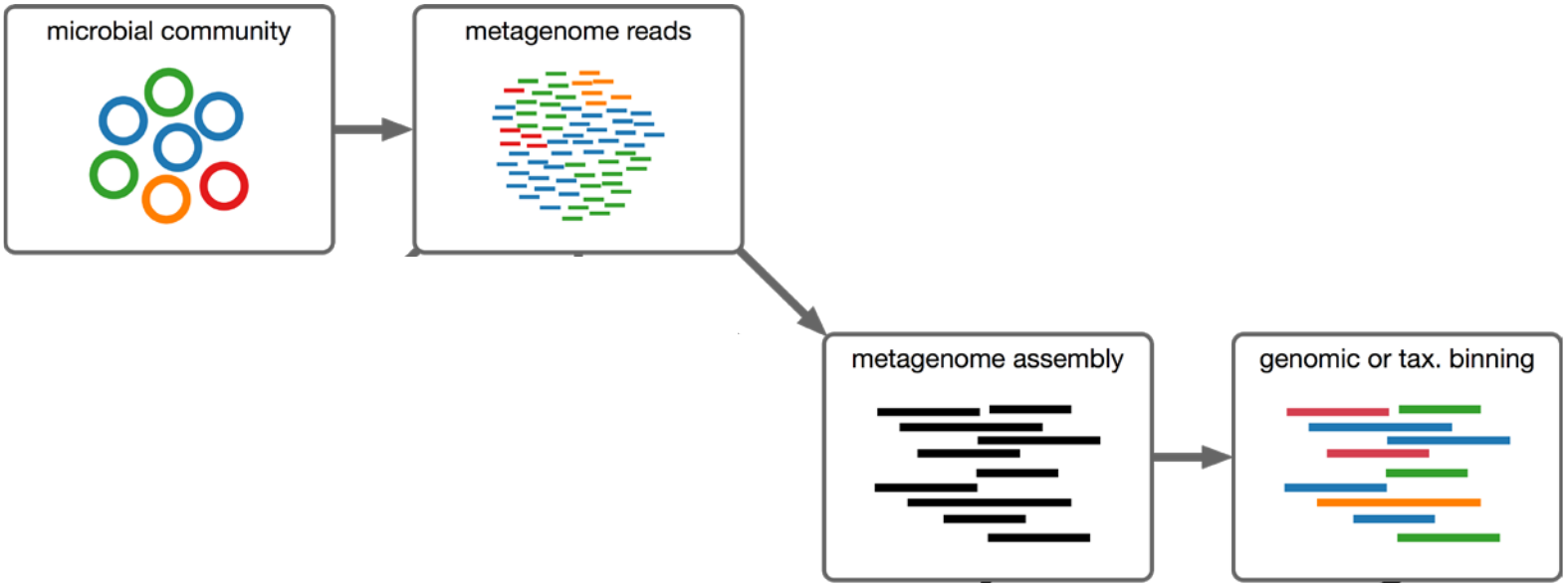


- Av. completeness (recall) of methods decreases with assembly quality and inc. strain diversity (marine: 40 vs. 20%; strain madness 20 vs. 5%)
- Av. purity (precision) of methods decreases mostly with inc. strain diversity (80% marine vs. 60% strain madness)

Genome Binning: gsa and ma



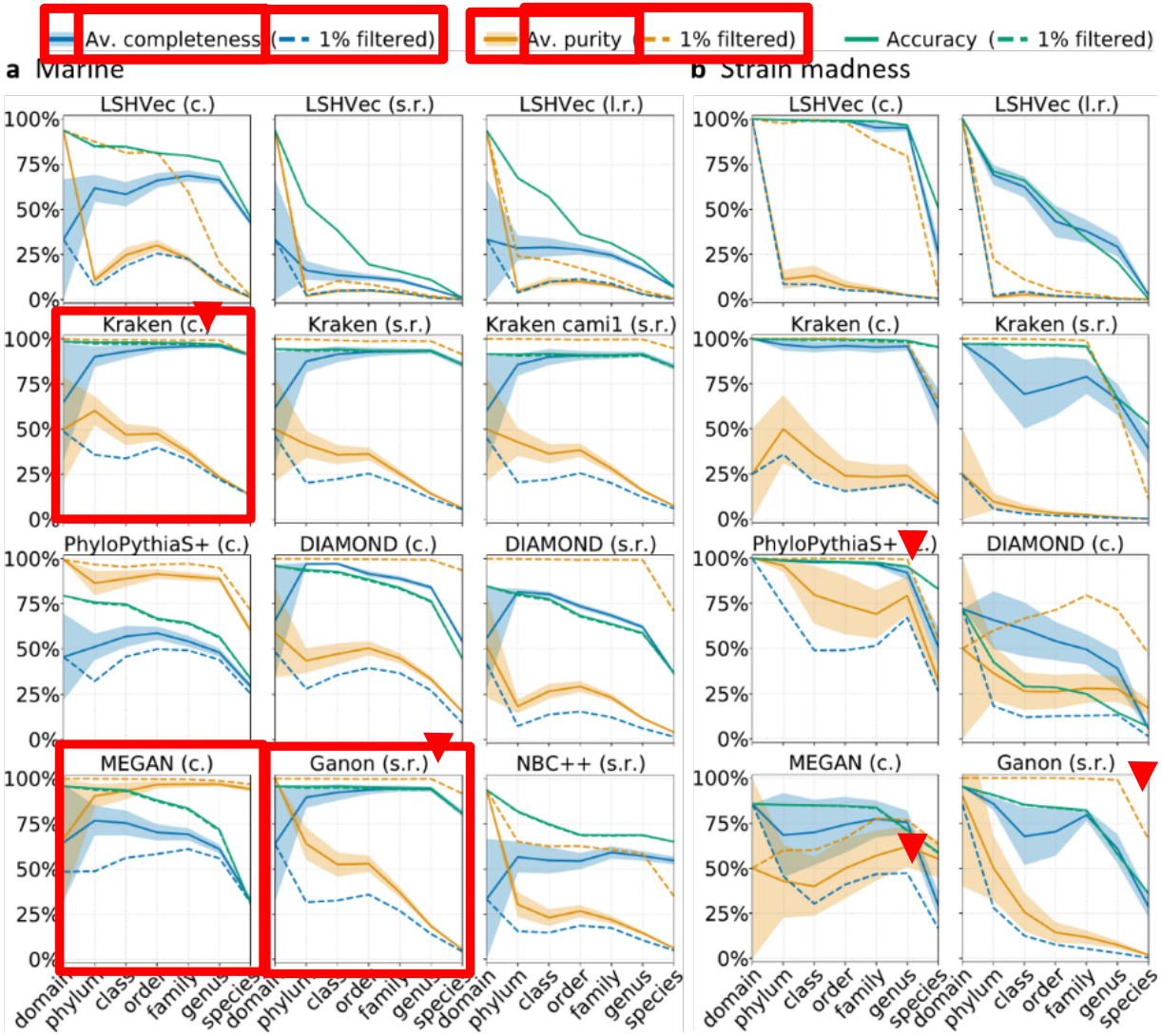
- Different methods excelled for individual metrics and data partitions
- Best trade-off performances across metrics by **UltraBinner** for gsas
- .. **MetaBinner, CONCOCT** for mas
- **.. CONCOCT** on plant-associated assemblies



TAXONOMIC BINNING CHALLENGE

HELMHOLTZ CENTRE FOR INFECTION RESEARCH

Taxon Binning



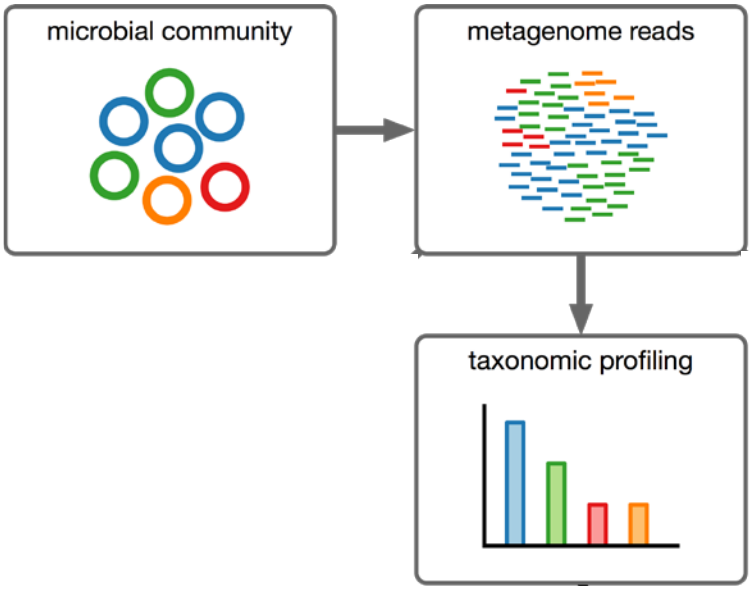
Some methods perform well until genus level (family in CAMI 1), depends on data

Across all data sets, metrics, ranks MEGAN, Kraken on contigs, Ganon s.r. best

Filtering (↑purity, ↓completeness) helps some methods

No method did well on viral sequences

HELMHOLTZ CENTRE FOR INFECTION RESEARCH



TAXONOMIC PROFILING CHALLENGE

HELMHOLTZ CENTRE FOR INFECTION RESEARCH

Metrics for Taxonomic Profilers

Metric	Best method (score)	Second best method (score)	Third best method (score)
Recall			
Precision			
L1 Norm			
Unifrac			

Presence/absence of taxa

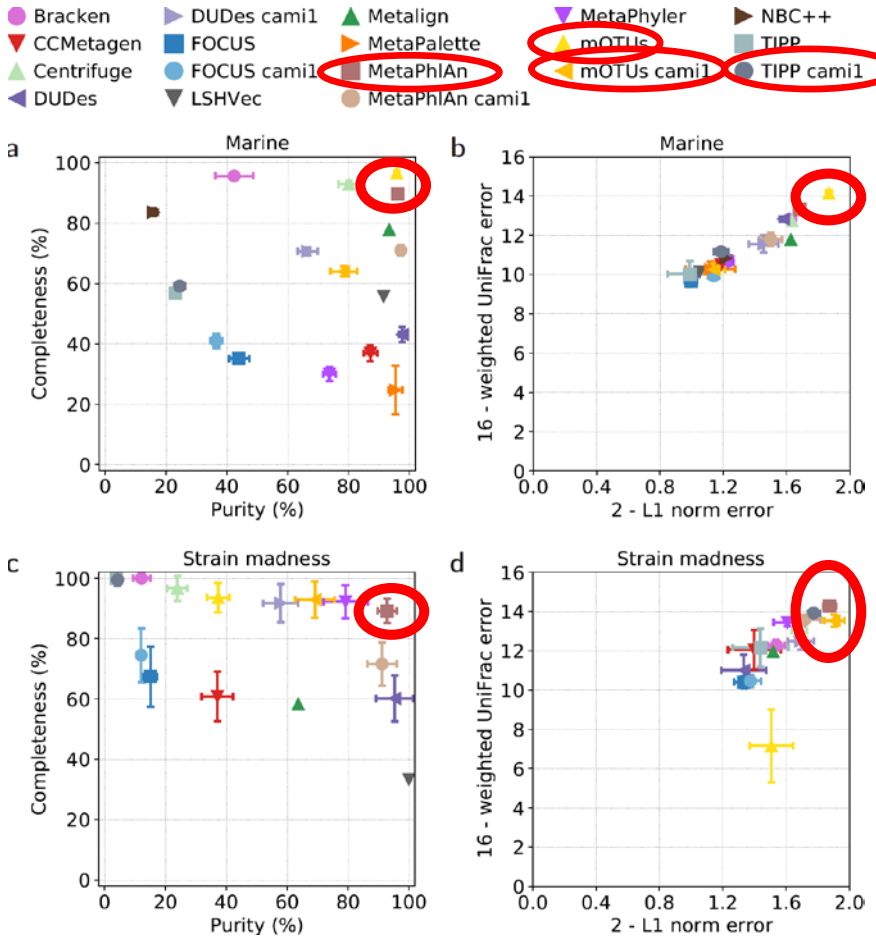
Recall (Completeness)

Precision (Purity)

Abundance estimates

L1-Norm & weighted Unifrac

Taxon Profiling

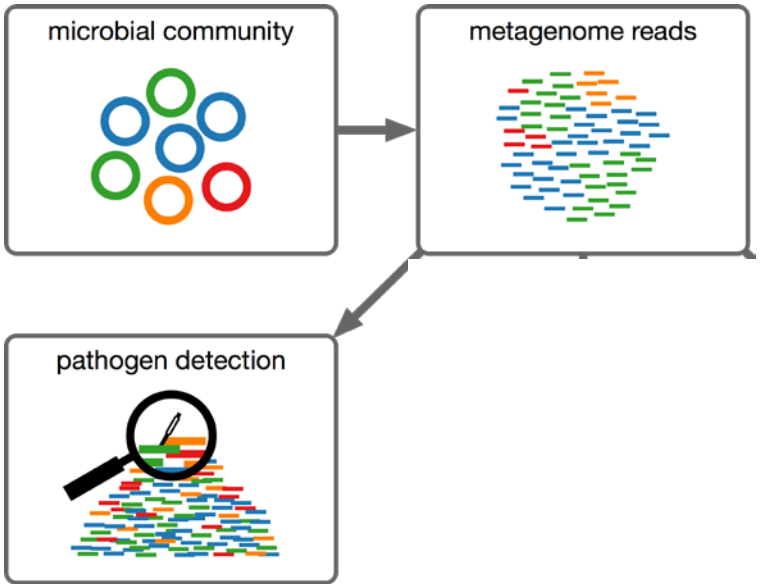


Genus rank

Performances good until genus rank

Several methods excel across data sets and metrics

- Taxon prediction: mOTUs, MetaPhlAn
- Abundance estimates: mOTUs, MetaPhlAn, TIPP cami1 and mOTUs cami1
- Alpha-diversity (Shannon equitability): MetaphlAn and DUDes
- Viruses, Archaea not well detected

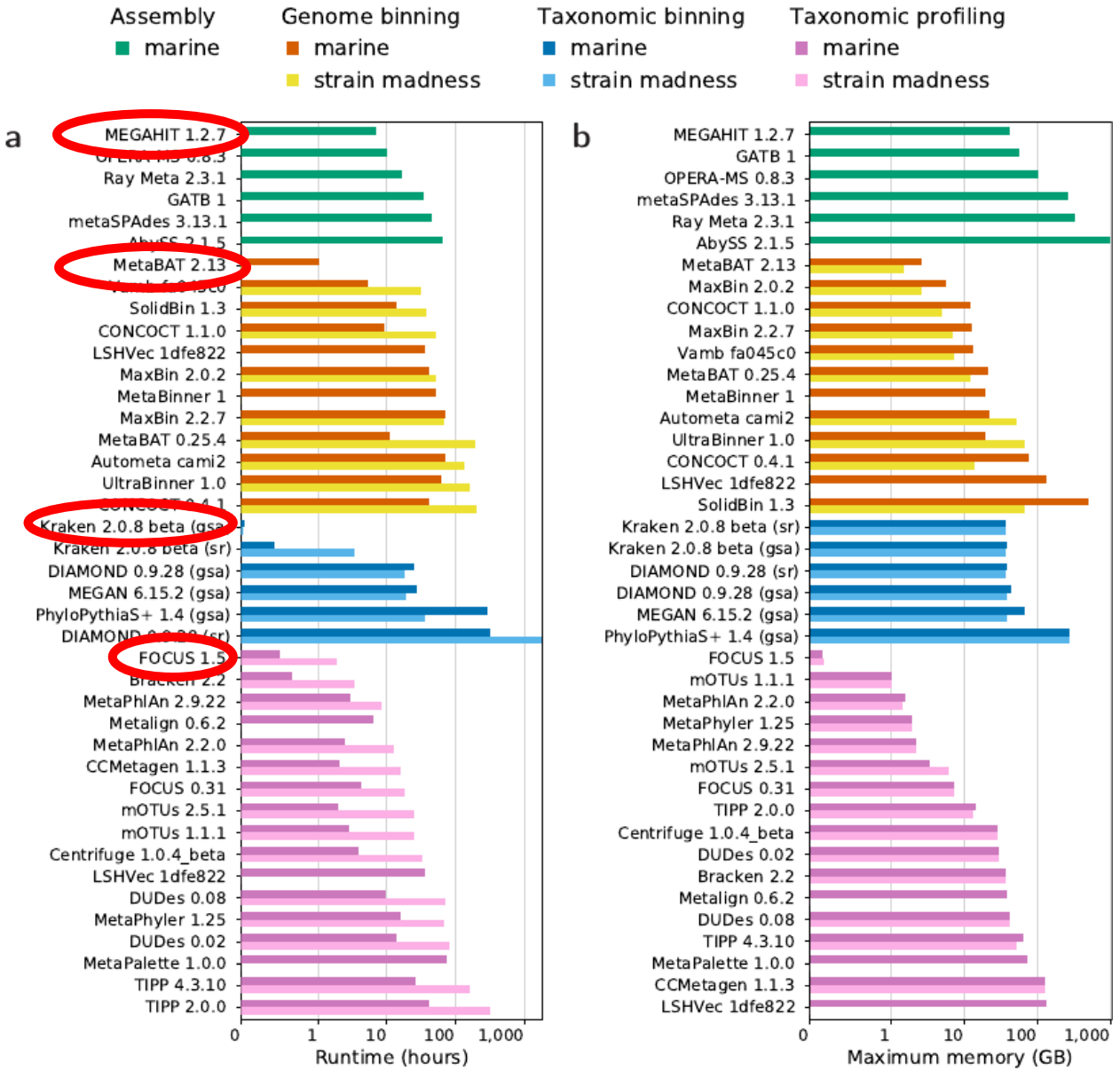


CLINICAL PATHOGEN DETECTION CHALLENGE

Clinical Pathogen Detection Challenge

- Short-read sample from fatally ill patient with unknown infection and case description
- 3 of 10 submissions identified causal pathogen (Crimean-Congo hemorrhagic fever orthonairovirus)
- None fully reproducible, as all manually curated

Runtime and memory requirements



- Efficient methods in all categories
- Large differences within categories: fastest 10-10^3x faster than last placed method

Key results

Assembly

- Substantial (up 30%) improvements across metrics since CAMI 1
- Hybrid assemblers are better for difficult regions (16S rRNA)
- Strain diversity and low coverage is a challenge (short read and hybrid)

Genome binning

- Highly variable across metrics and data types
- Substantial improvement by ensembl binners
- Assembly quality AND strain diversity is a challenge for genome binning

Key results

Taxon binners and profilers

- Performant and computationally efficient software across a range of conditions and relevant metrics
- Profilers matured, good in taxon identification, abundance and diversity estimates (less variance across methods than in CAMI 1)
- Good until genus rank, low taxonomic bacterial ranks, viruses and Archaea challenging

Clinical pathogen detection

- Successful, but not reproducible

Further details

CAMI II: Meyer, Fritz *et al.,* Nat. Methods 2022

Tutorial: Meyer *et al.,* Nat. Protocols 2021

CAMI I: Sczyrba *et al.,* Nat. Methods 2017

New CAMI benchmarking portal

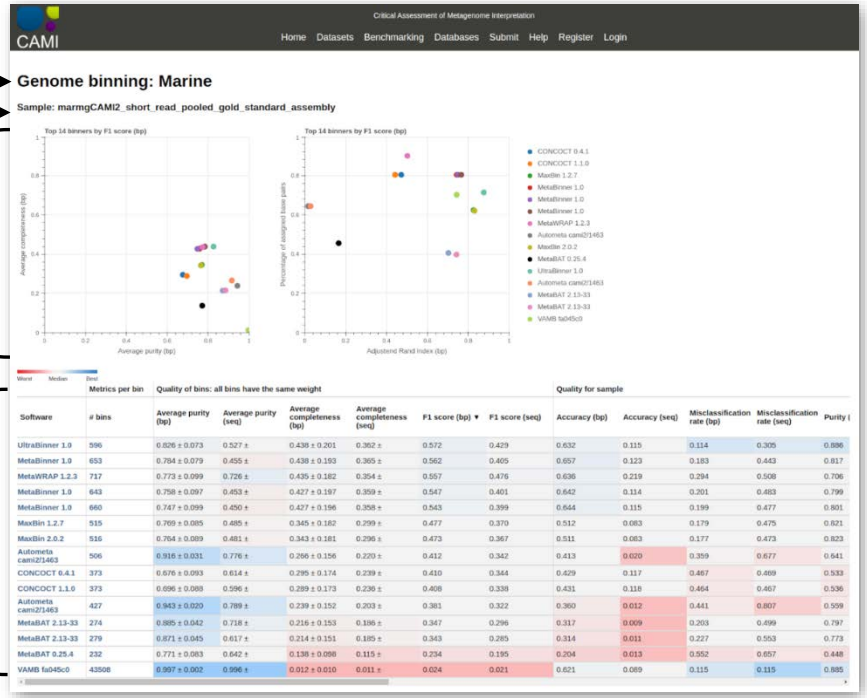
Online, automatic benchmarking
- Assembly
- Binning
- Taxonomic profiling

CAMI standards

New users' results and from CAMI I and II Challenges

Easy, drag-and-drop upload

Rich visualizations



https://cami.bifo.helmholtz-hzi.de

Outlook

CAMI III

- clinical metagenomics challenges
  - pathogens
  - AMR phenotyping

- CAMI II best practice workflows

CAMI II Contributors

F. Meyer, A. Fritz, Z.-L. Deng, D. Koslicki, A. Gurevich, G. Robertson, T.-R. Lesker, M. Alser, D. Antipov, F. Beghini, D. Bertrand, J.J. Brito, C.T. Brown, J. Buchmann, A. Buluç, B. Chen, R. Chikhi, P.T.L.C. Clausen, A. Cristian, P.W. Dabrowski, A.E. Darling, R. Egan, E. Eskin, E. Georganas, E. Goltsman, M.A. Gray, L.H. Hansen, S. Hofmeyr, P. Huang, L. Irber, H. Jia, T.S. Jørgensen, S.D. Kieser, T. Klemetsen, A. Kola, M. Kolmogorov, A. Korobeynikov, J. Kwan, N. LaPierre, C. Lemaitre, C. Li, A. Limasset, F. Malcher-Miranda, S. Mangul, V.R. Marcelino, C. Marchet, P. Marijon, D. Meleshko, D.R. Mende, A. Milanese, N. Nagarajan, J. Nissen, S. Nurk, L. Oliker, L. Paoli, P. Peterlongo, V.C. Piro, J.S. Porter, S. Rasmussen, E.R. Rees, K. Reinert, B. Renard, E.M. Robertsen, G.L. Rosen, H.-J. Ruscheweyh, V. Sarwal, N. Segata, E. Seiler, L. Shi, F. Sun, S. Sunagawa, S.J. Sørensen, A. Thomas, C. Tong, M. Trajkovski, J. Tremblay, G. Uritskiy, R. Vicedomini, Zi. Wang , Zhe. Wang, Zho. Wang, A. Warren, N.P. Willassen, K. Yelick, R. You, G. Zeller, Z. Zhao, S. Zhu, J. Zhu, R. Garrido-Oter, P. Gastmeier, S. Hacquard, S. Häußler, A. Khaledi, F. Maechler, F. Mesny, S. Radutoiu, P. Schulze-Lefert, N. Smit, T. Strowig, A. Bremges, A. Sczyrba, A.C. McHardy

>100 contributors from 77 institutions

Get involved!

Follow @cami_challenge

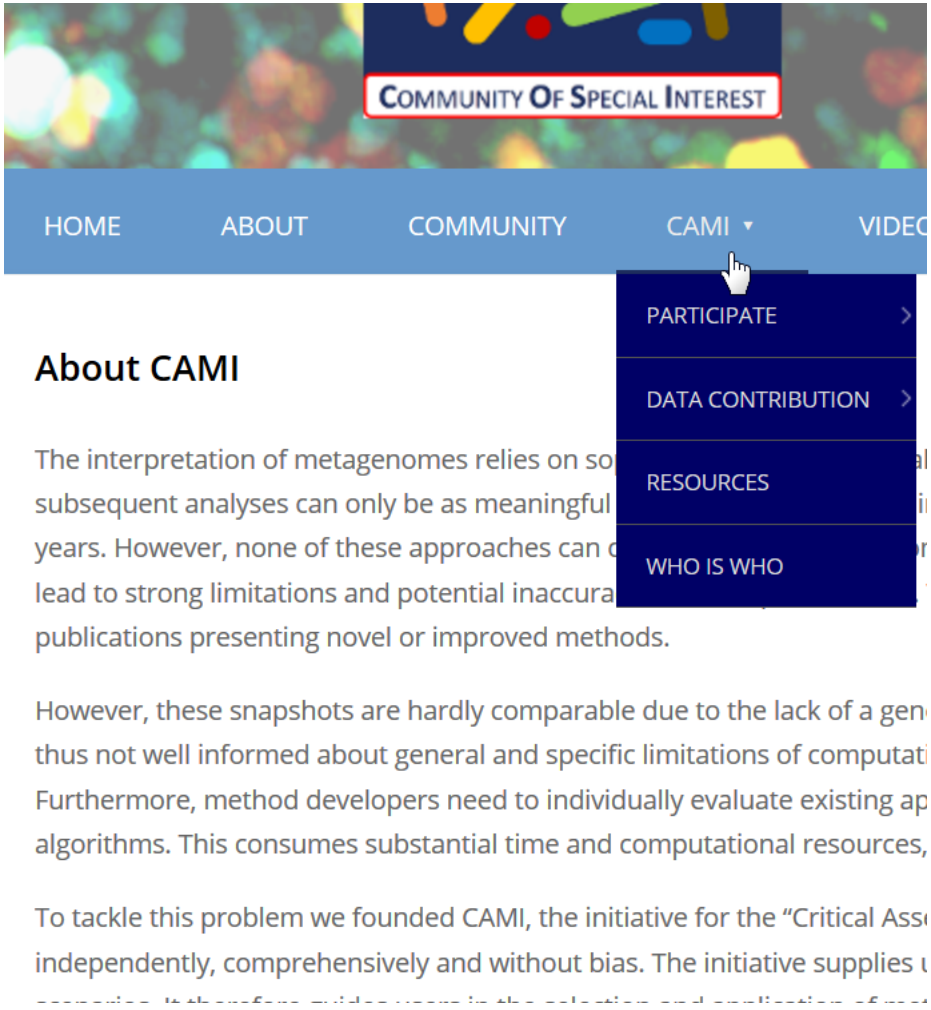
Methods to benchmark?

New challenge wishlist?

Data to contribute?

Contact@CAMI-challenge.org

Alice.McHardy@helmholtz-hzi.de



https://www.microbiome-cosi.org/