



# Metagenome assembly and post-assembly analysis of complex clinical samples

**Alexey Gurevich**, Department of Clinical Bioinformatics  
Geneva, 20.10.2022

# Research Background



Alexey Gurevich

FOLLOW

Jun. Prof. of Bioinformatics, Helmholtz Inst. for Pharm. Res. Saarland (HIPS) & Saarland University

☐ TITLE

CITED BY YEAR

## Metagenome assembly

<a href="#">Assembling single-cell genomes and mini-metagenomes from chimeric MDA products</a> S Nurk, A Bankevich, D Antipov, AA Gurevich, A Korobeynikov, A Lapidus, ... Journal of Computational Biology 20 (10), 714-737	977	2013
<a href="#">metaFlye: scalable long-read metagenome assembly using repeat graphs</a> M Kolmogorov, DM Bickhart, B Behsaz, A Gurevich, M Rayko, SB Shin, ... Nature Methods 17 (11), 1103-1110	189	2020

## (Meta)genome assembly quality control

<a href="#">QUAST: quality assessment tool for genome assemblies</a> A Gurevich, V Saveliev, N Vyahhi, G Tesler Bioinformatics 29 (8), 1072-1075	5293	2013
<a href="#">MetaQUAST: evaluation of metagenome assemblies</a> A Mikheenko, V Saveliev, A Gurevich Bioinformatics 32 (7), 1088-1090	371	2016

## Metagenome interpretation quality control

<a href="#">Critical assessment of metagenome interpretation—a benchmark of metagenomics software</a> A Sczyrba, P Hofmann, P Belmann, D Koslicki, S Janssen, J Dröge, ... Nature methods 14 (11), 1063-1071	565	2017
<a href="#">Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit</a> F Meyer, TR Lesker, D Koslicki, A Fritz, A Gurevich, AE Darling, A Sczyrba, ... Nature protocols 16 (4), 1785-1801	22	2021
<a href="#">Critical Assessment of Metagenome Interpretation: the second round of challenges</a> F Meyer, A Fritz, ZL Deng, D Koslicki, TR Lesker, A Gurevich, G Robertson, ... Nature methods 19 (4), 429-440	39	2022

# Research Background



Alexey Gurevich

FOLLOW

Jun. Prof. of Bioinformatics, Helmholtz Inst. for Pharm. Res. Saarland (HIPS) & Saarland University

<input type="checkbox"/>	TITLE			CITED BY	YEAR
--------------------------	-------	--	--	----------	------

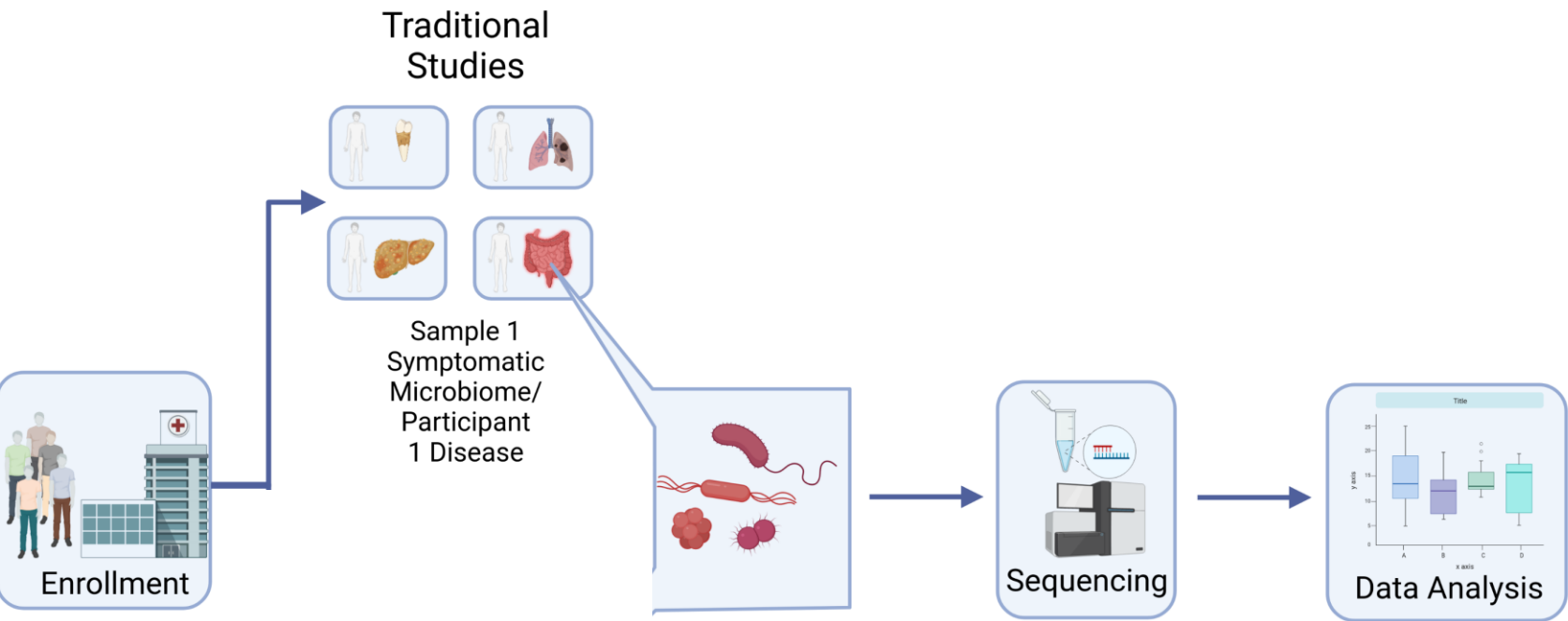
## Metagenomics for finding natural products, e.g., novel medicines

<a href="#">Metabolic fingerprints from the human oral microbiome reveal a vast knowledge gap of secreted small peptidic molecules</a>	22	2017
A Edlund, N Garg, H Mohimani, A Gurevich, X He, W Shi, PC Dorrestein, ... Msystems 2 (4), e00058-17		
<a href="#">Dereplication of microbial metabolites through database search of mass spectra</a>	162	2018
H Mohimani, A Gurevich, A Shlemov, A Mikheenko, A Korobeynikov, ... Nature communications 9 (1), 1-12		
<a href="#">MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities</a>	43	2019
L Cao, A Gurevich, KL Alexander, CB Naman, T Leão, E Glukhov, ... Cell systems 9 (6), 600-608. e4		
<a href="#">Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery</a>	15	2021
B Behsaz, E Bode, A Gurevich, YN Shi, F Grundmann, D Acharya, ... Nature communications 12 (1), 1-17		
<a href="#">Nerpa: A tool for discovering biosynthetic gene clusters of bacterial nonribosomal peptides</a>	4	2021
O Kunyavskaya, AM Tagirdzhanov, AM Caraballo-Rodríguez, LF Nothias, ... Metabolites 11 (10), 693		

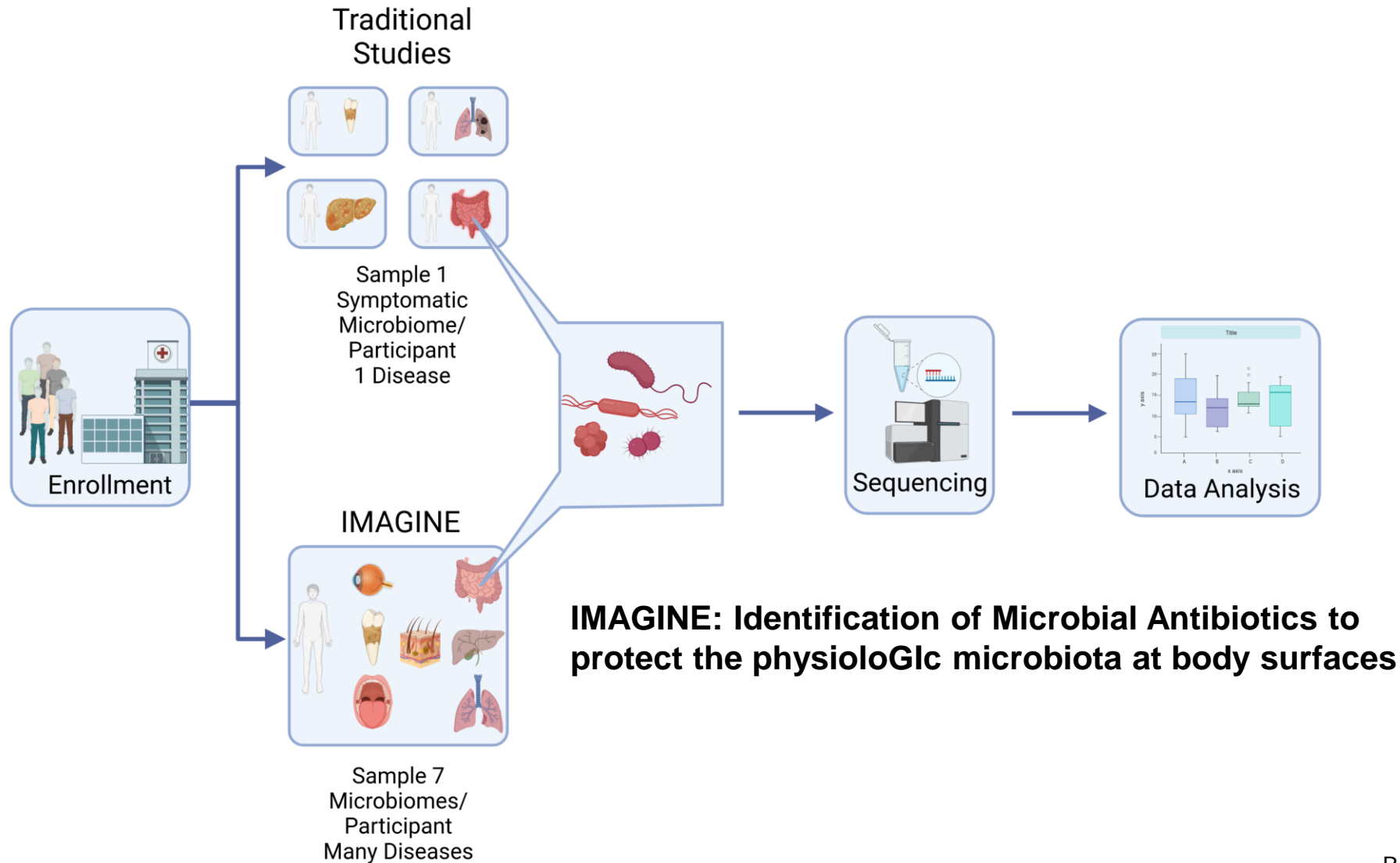
## Metagenomics for clinical samples

Since July 2022

# Clinical samples

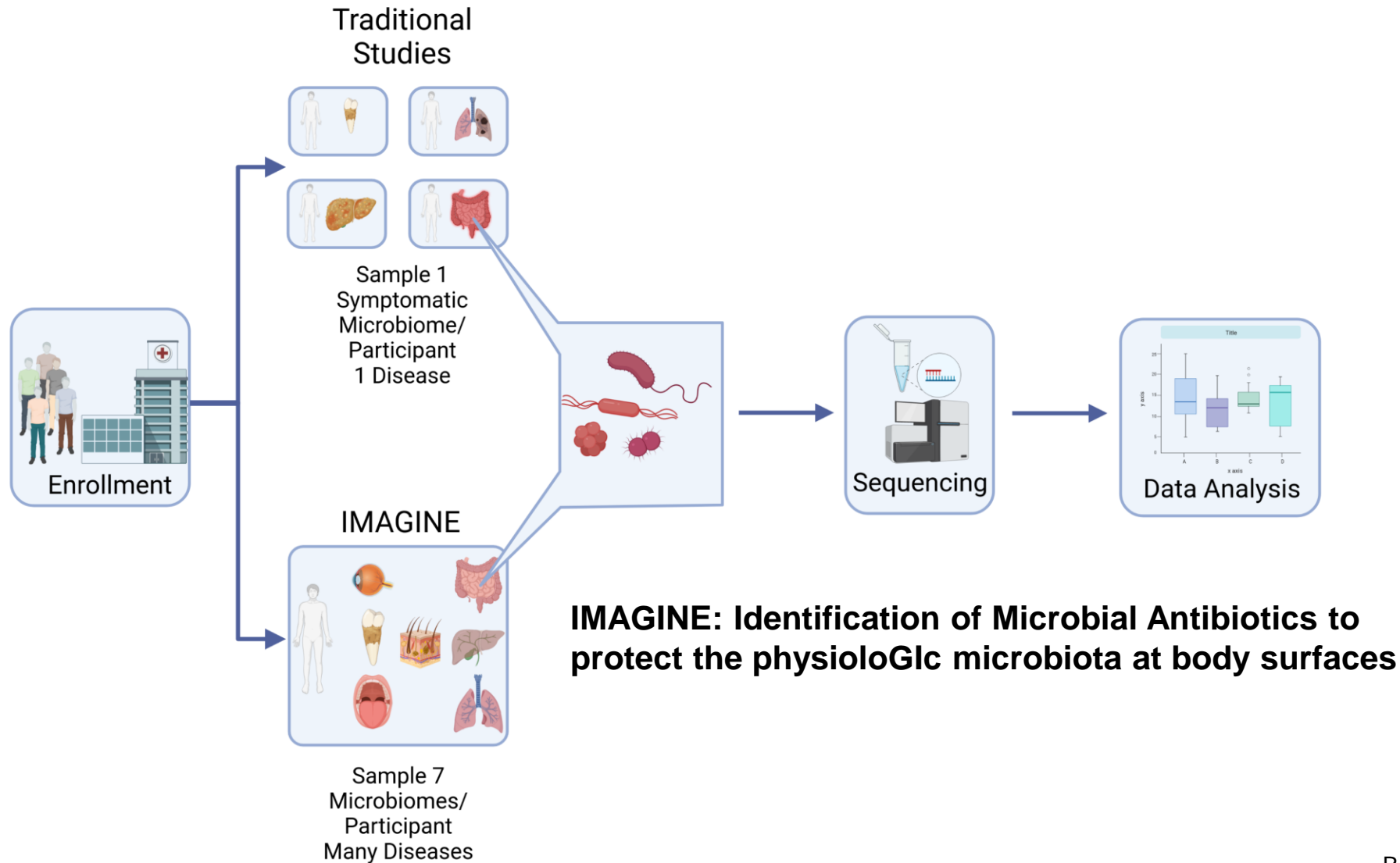


# Clinical samples



- 10 clinics involved
- Extensive documentation of clinical data
- Whole microbiome sequencing

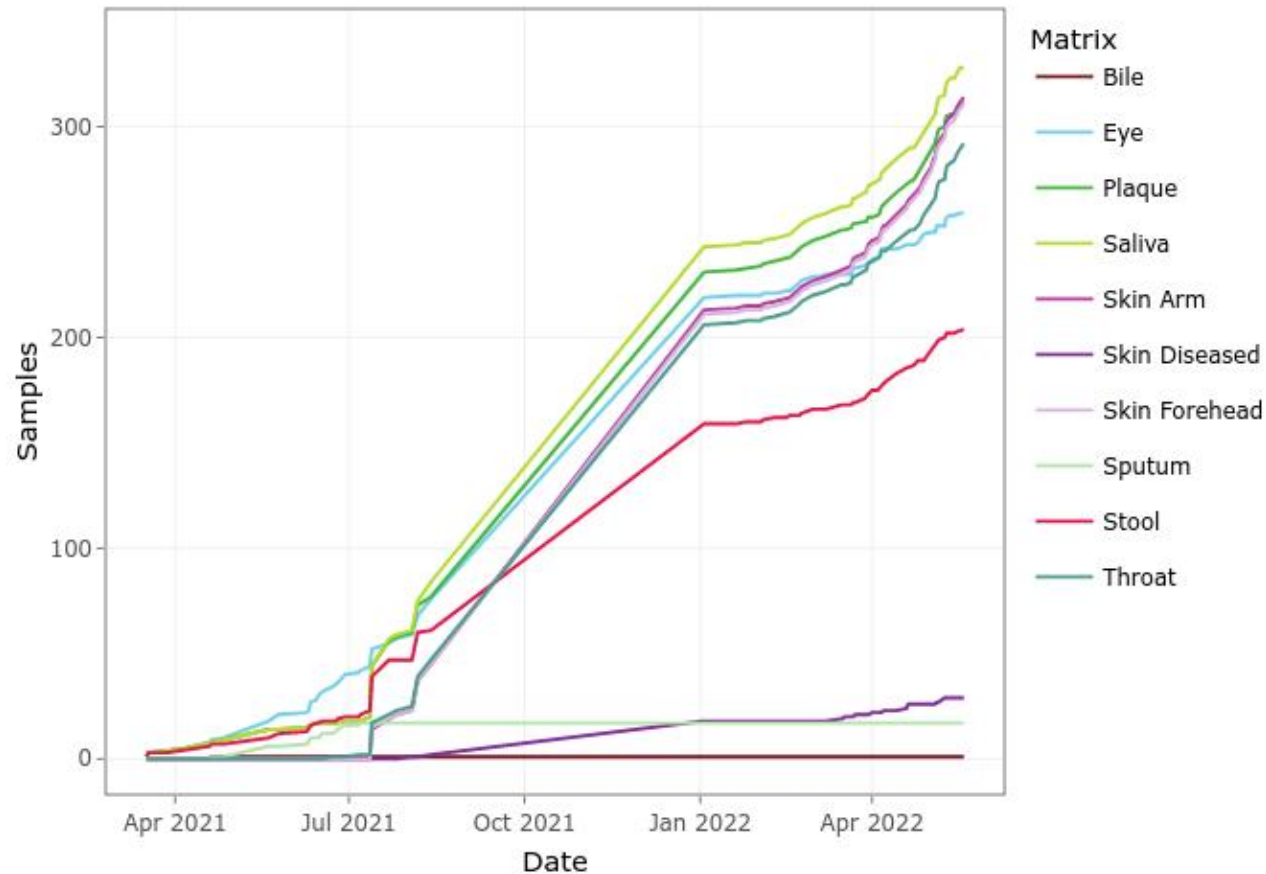
# Clinical samples



- 10 clinics involved
  - Extensive documentation of clinical data
  - Whole microbiome sequencing
- +**
- Long reads (ONT)
  - Transcriptome
  - Metabolome (MS/MS)



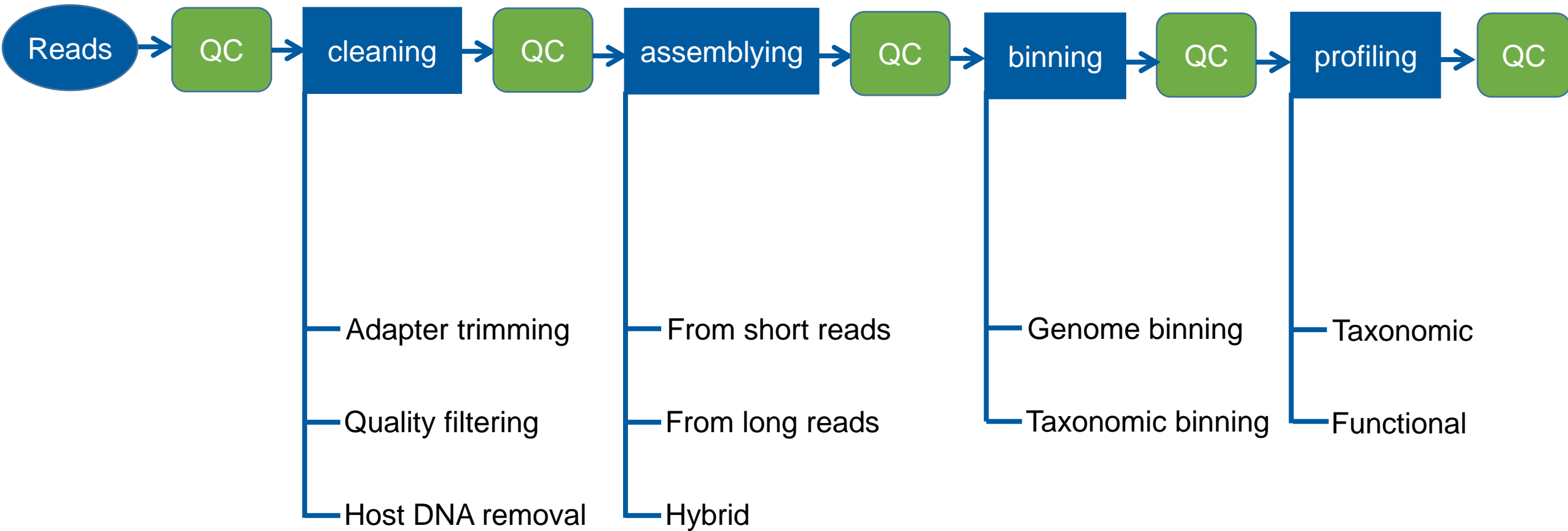
# Clinical samples – the IMAGINE project



## The latest numbers

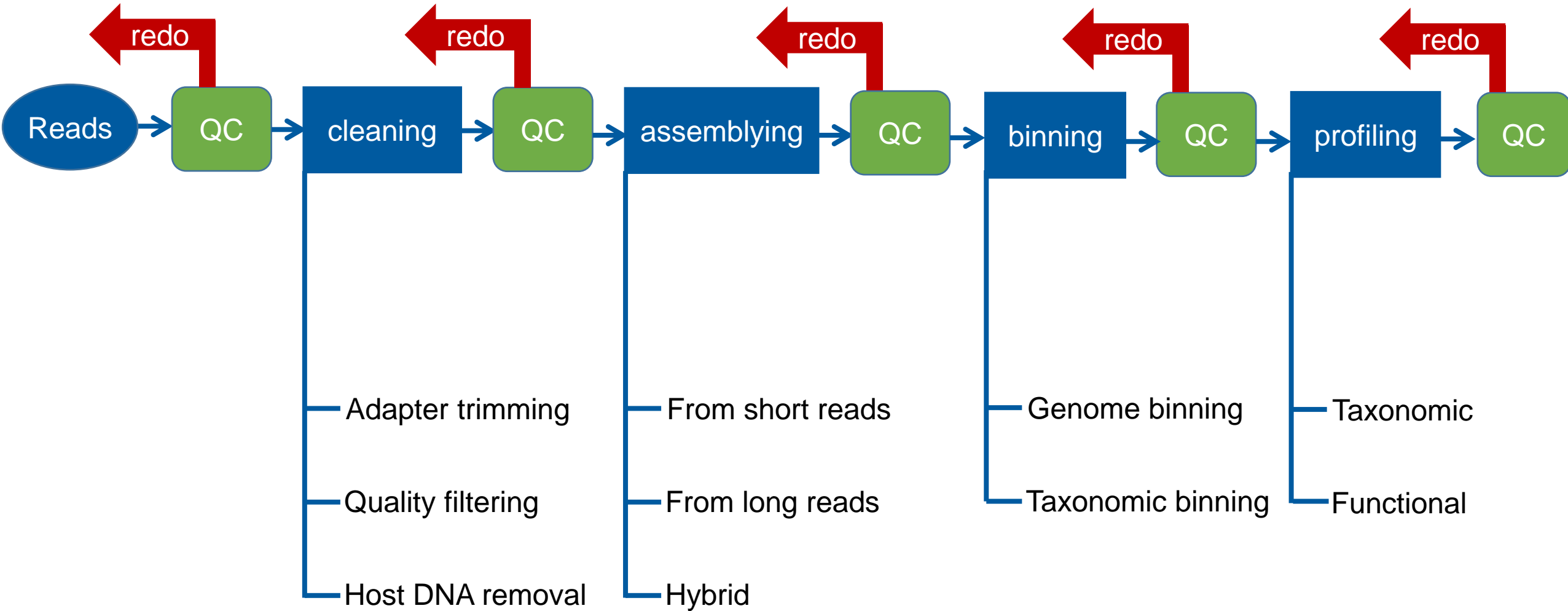
- > 2,500 samples collected (5-8 samples per patient)
- > 2,000 samples sequenced (~20 Gbp per sample)
- 877 samples from 259 patients processed computationally

# Metagenome assembly and post-assembly analysis

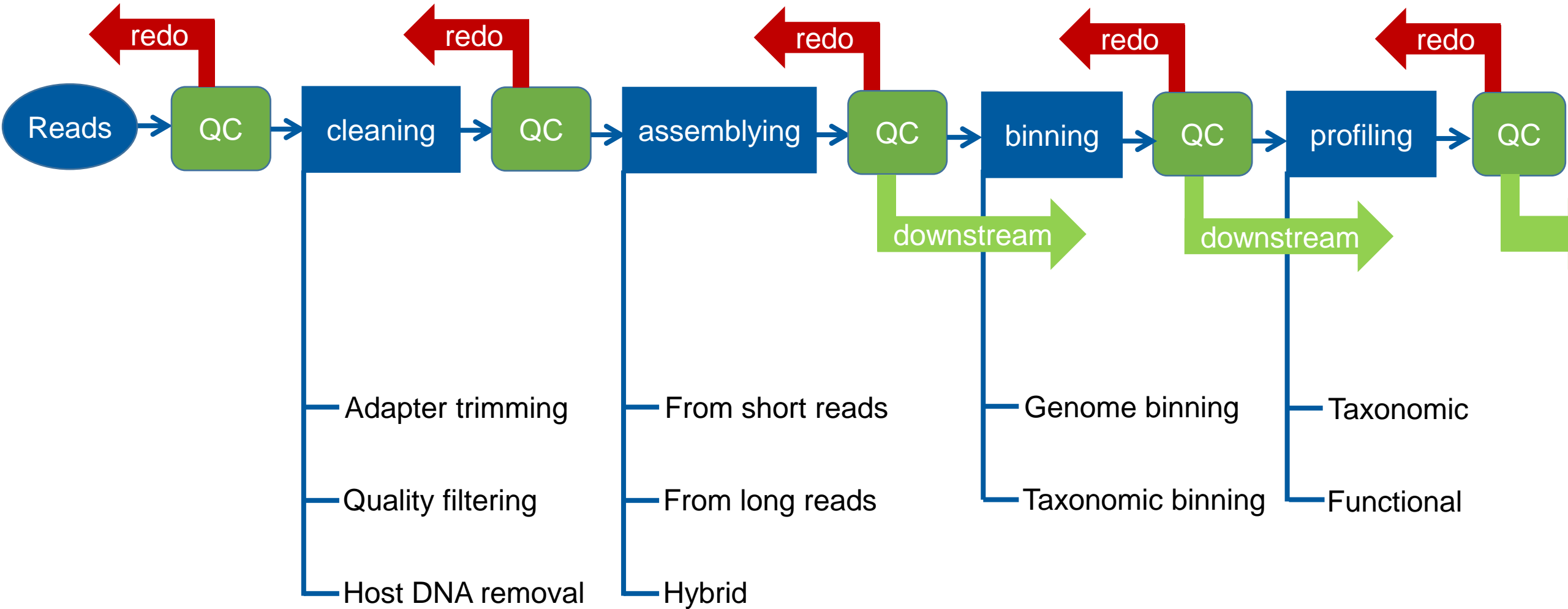




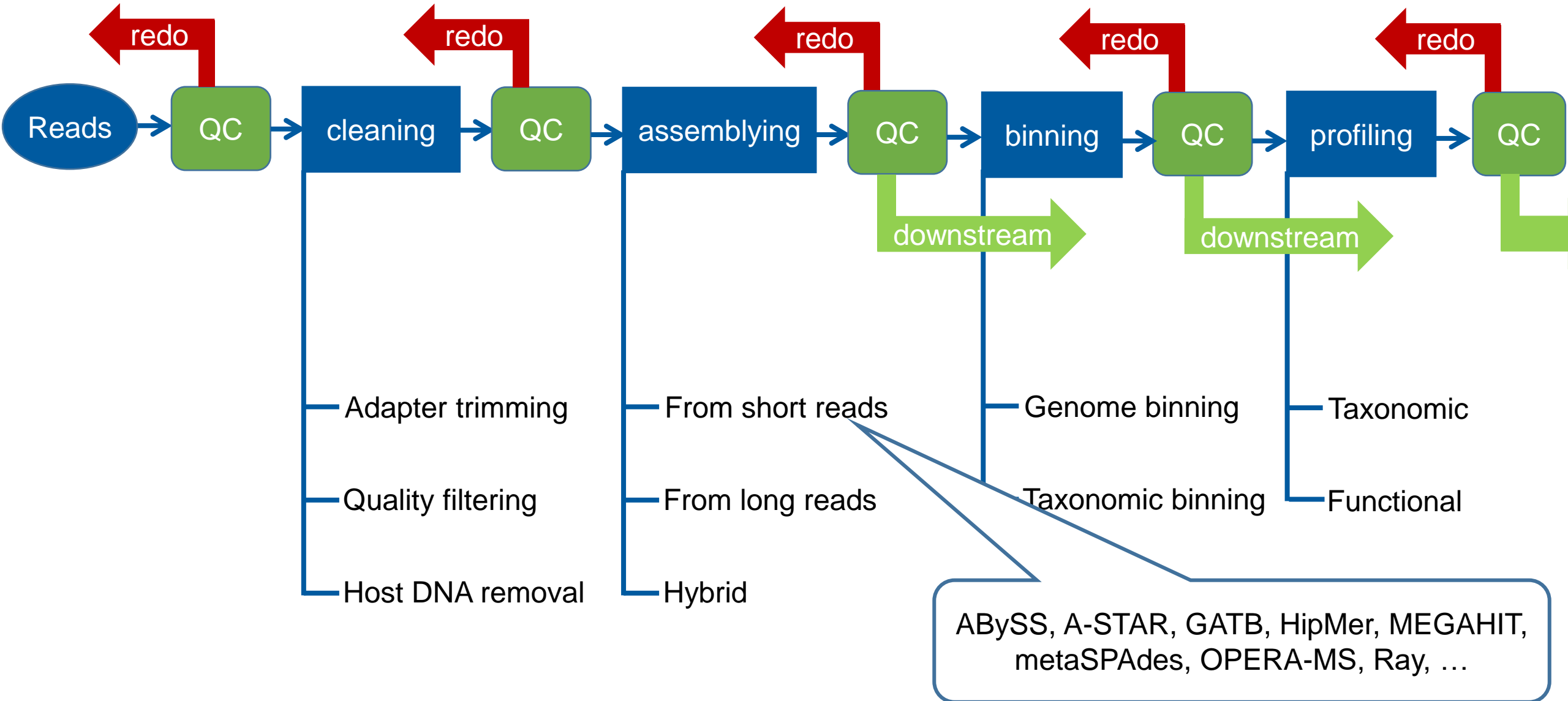
# Metagenome assembly and post-assembly analysis



# Metagenome assembly and post-assembly analysis



# Metagenome assembly and post-assembly analysis



# CAMI challenges



Open Access | Published: 02 October 2017

## Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software

[Alexander Sczyrba](#) ✉, [Peter Hofmann](#), [Peter Belmann](#), [David Koslicki](#), [Stefan Janssen](#), [Johannes Dröge](#), [Ivan Gregor](#), [Stephan Majda](#), [Jessika Fiedler](#), [Eik Dahms](#), [Andreas Bremges](#), [Adrian Fritz](#), [Ruben Garrido-Oter](#), [Tue Sparholt Jørgensen](#), [Nicole Shapiro](#), [Philip D Blood](#), [Alexey Gurevich](#), [Yang Bai](#), [Dmitrij Turaev](#), [Matthew Z DeMaere](#), [Rayan Chikhi](#), [Niranjan Nagarajan](#), [Christopher Quince](#), [Fernando Meyer](#), ... [Alice C McHardy](#) ✉

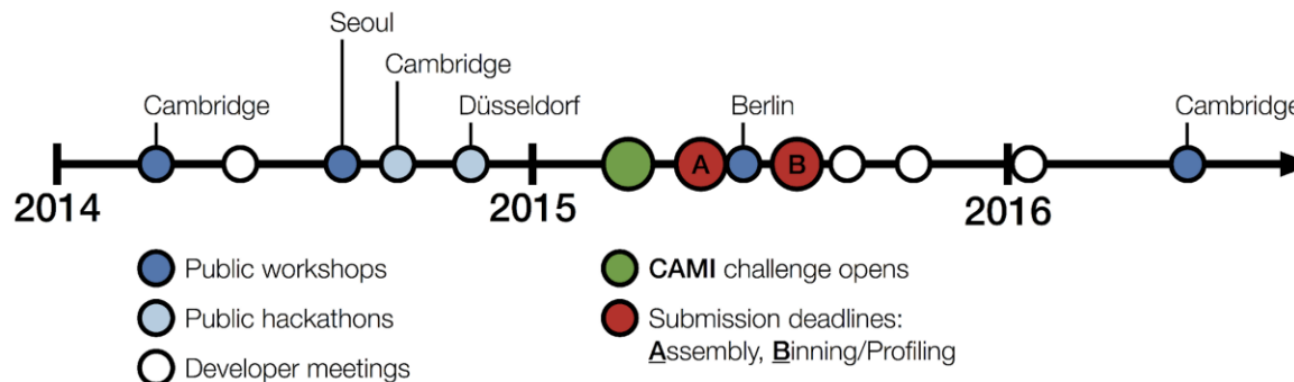
+ Show authors

*Nature Methods* **14**, 1063–1071 (2017) | [Cite this article](#)

### Key numbers


- ~700 newly sequenced microorganisms and ~600 novel viruses and plasmids were used for simulating metagenome benchmark datasets
- 215 submissions, representing 25 programs, were received from 16 teams worldwide
- 6 genome assemblers
- 5 genome binners
- 4 taxonomic binners
- 10 taxonomic profilers

### Timeline



Analysis | [Open Access](#) | [Published: 08 April 2022](#)

## Critical Assessment of Metagenome Interpretation: the second round of challenges

[Fernando Meyer](#), [Adrian Fritz](#), [Zhi-Luo Deng](#), [David Koslicki](#), [Till Robin Lesker](#), [Alexey Gurevich](#), [Gary Robertson](#), [Mohammed Alser](#), [Dmitry Antipov](#), [Francesco Beghini](#), [Denis Bertrand](#), [Jaqueline J. Brito](#), [C. Titus Brown](#), [Jan Buchmann](#), [Aydin Buluç](#), [Bo Chen](#), [Rayan Chikhi](#), [Philip T. L. C. Clausen](#), [Alexandru Cristian](#), [Piotr Wojciech Dabrowski](#), [Aaron E. Darling](#), [Rob Egan](#), [Eleazar Eskin](#), [Evangelos Georganas](#), ...  
[Alice Carolyn McHardy](#)  [+ Show authors](#)

[Nature Methods](#) **19**, 429–440 (2022) | [Cite this article](#)

### Timeline

#### Part a

16.01.2019 Challenge opens

17.05.2019 Assembly challenge closes

03.06.2019 Clinical pathogen detection  
challenge starts

25.10.2019 Challenge closes

#### Part b (Rhizosphere)

14.02.2020 Dataset provided

29.09.2020 Assembly challenge closes

30.09.2020 “Correct” assemblies  
provided

31.01.2021 Challenge closes

### Key numbers

- ~1,700 new and known genomes and ~600 new viruses and plasmids were used for simulating metagenome benchmark datasets
- 5,002 submissions, representing 36 (76) programs, were received from 30 teams worldwide
- 8 (20) genomes assemblers
- 8 (18) genome binners
- 7 (9) taxonomic binners
- 13 (22) taxonomic profilers

# Do your own benchmark!



Review Article | [Published: 01 March 2021](#)

## **Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit**

[Fernando Meyer](#), [Till-Robin Lesker](#), [David Koslicki](#), [Adrian Fritz](#), [Alexey Gurevich](#), [Aaron E. Darling](#),  
[Alexander Sczyrba](#), [Andreas Bremges](#) & [Alice C. McHardy](#) ✉

[Nature Protocols](#) **16**, 1785–1801 (2021) | [Cite this article](#)

**CAMISIM** for simulating metagenomes and microbial communities

<https://github.com/CAMI-challenge/CAMISIM/>

[Fritz\\*, Hofmann\\*, et al. CAMISIM: Simulating metagenomes and microbial communities. Microbiome, 7:17. \(2019\)](#)

**MetaQUAST** for assessing metagenome assemblies

<http://quast.sourceforge.net/metaquast.html>

[Mikheenko, Saveliev, and Gurevich. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics, 32, 7, 1088 \(2016\)](#)

**AMBER** for assessing genome binning methods

<https://github.com/CAMI-challenge/AMBER>

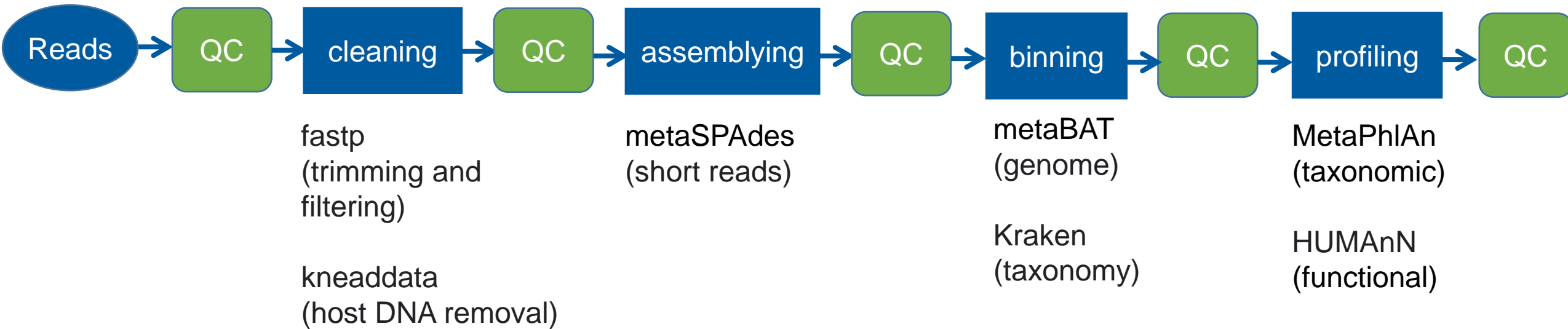
[Meyer, et al. AMBER: Assessment of Metagenome BinnERs, GigaScience, 7, 6 \(2018\)](#)

**OPAL** for assessing taxonomic profilers

<https://github.com/CAMI-challenge/OPAL>

[Meyer, et al. Assessing taxonomic metagenome profilers with OPAL, Genome Biology, 20:51 \(2019\)](#)

# The IMAGINE data processing



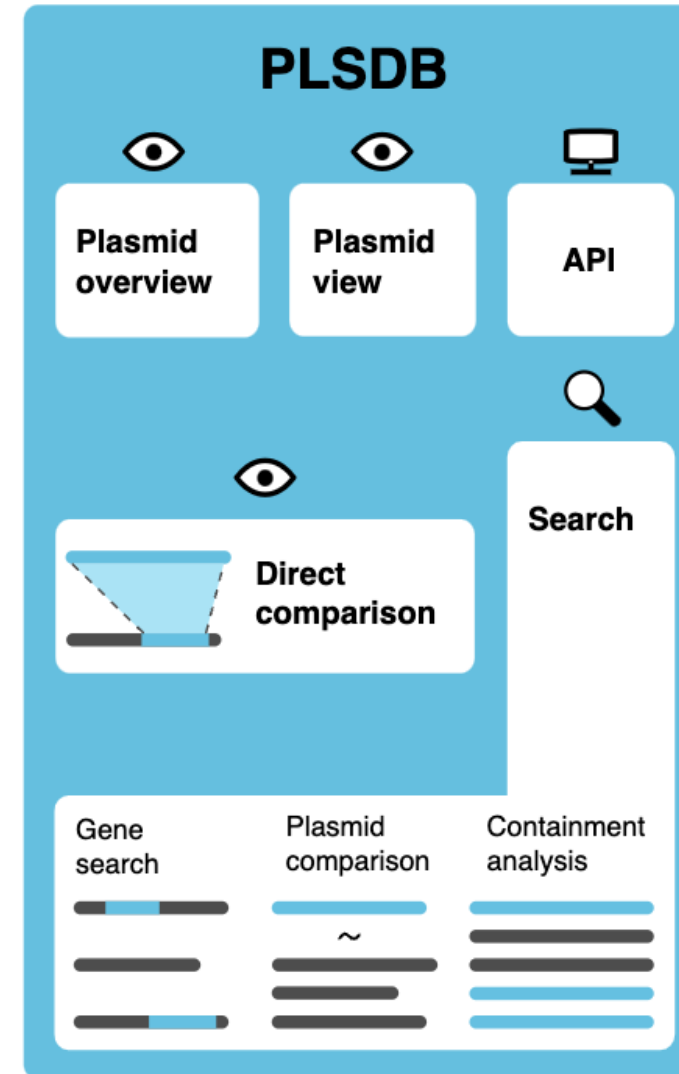


# HIPS as an Institute of the HZI in Braunschweig



# Plasmids in IMAGINE

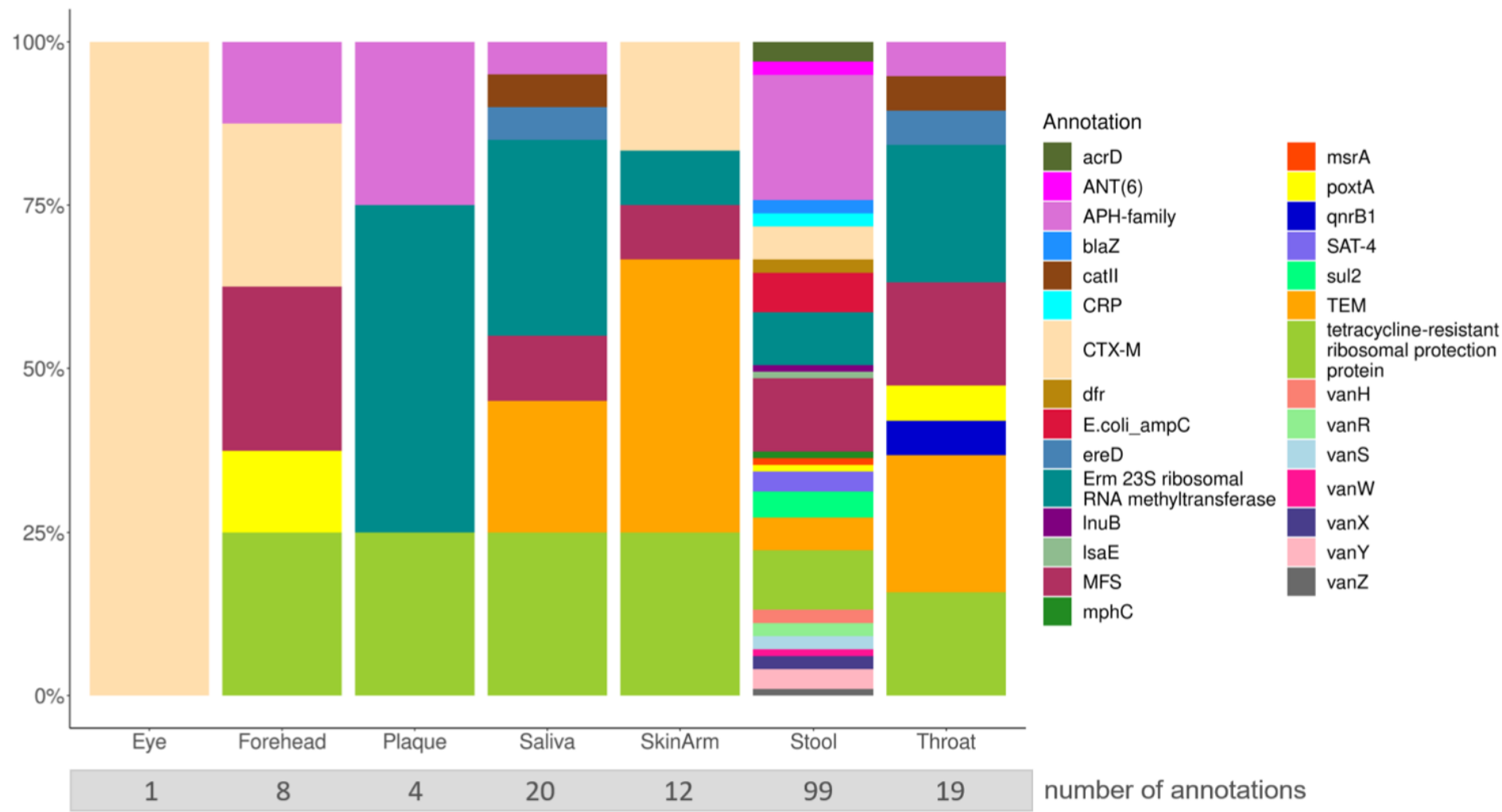
Bacterial plasmids play a key role in the horizontal transfer of antimicrobial resistance genes



> 34,000 records

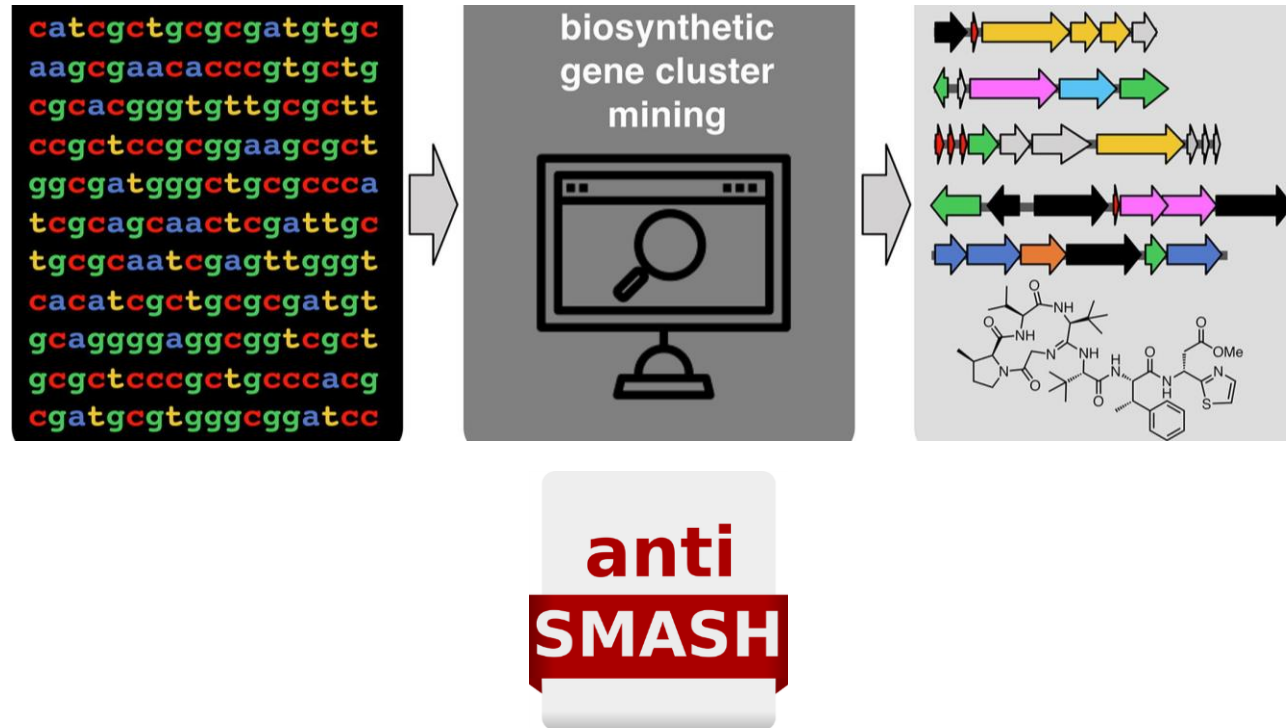
# Plasmids in IMAGINE\*

Resistance factors in PLSDb plasmids mapped to metagenome assemblies



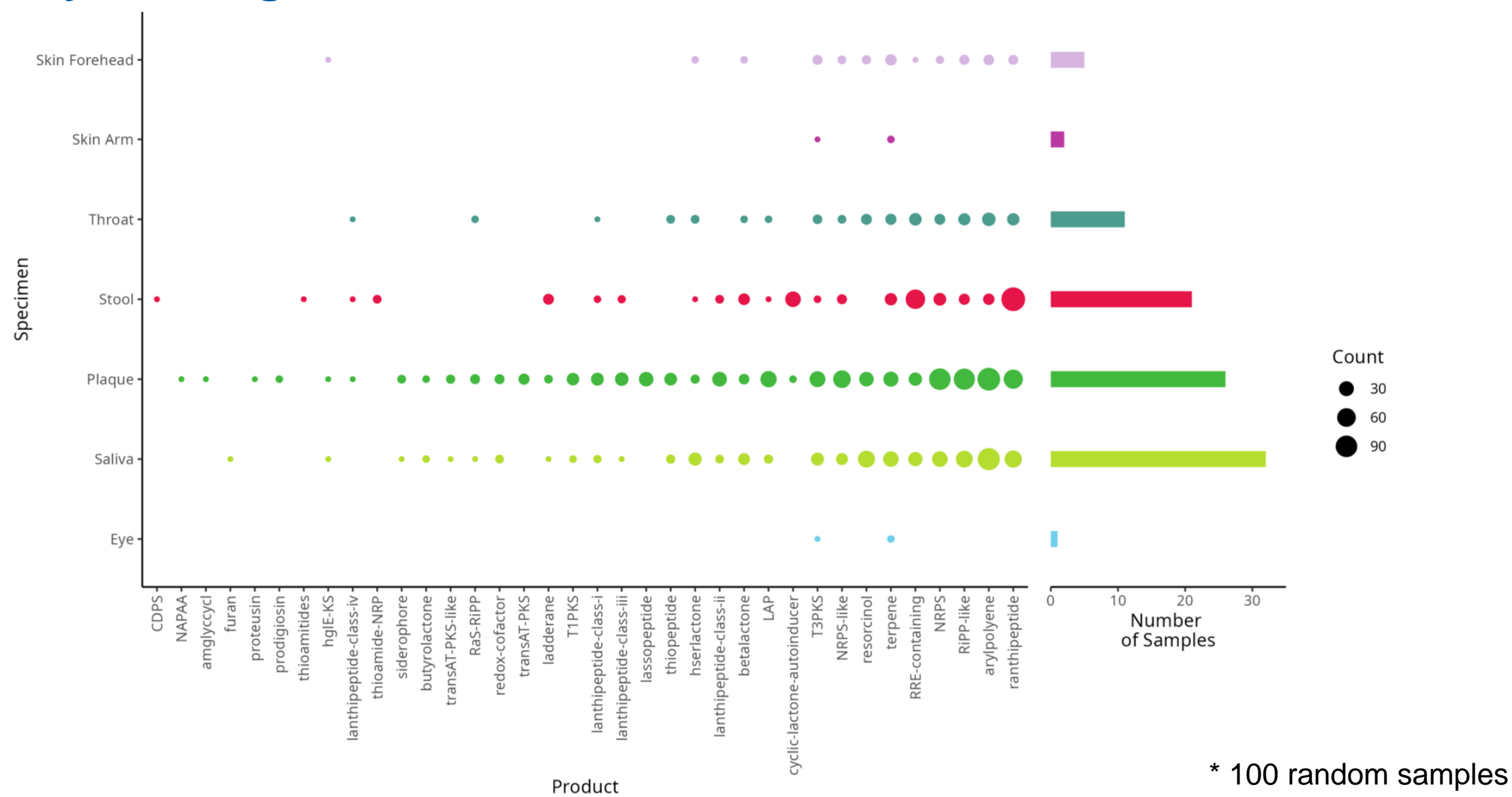
# Biosynthetic gene clusters in IMAGINE

Many natural products are encoded in BGCs



<https://antismash.secondarymetabolites.org/>

# Biosynthetic gene clusters in IMAGINE\*



**More information will follow..**

# Acknowledgements

## HIPS / UoS

- Prof. Dr. Rolf Müller
- Prof. Dr. Andreas Keller
- Georges Schmartz
- Pascal Hirsch

## The CAMI initiative

- Prof. Dr. Alice McHardy (HZI Braunschweig)
- Prof. Dr. Alexander Sczyrba (Bielefeld University)

## Funding

**HIPS** **HELMHOLTZ**  
Institut für Pharmazeutische Forschung Saarland



 International Conference  
on Clinical Metagenomics

# Thank you!

✉ [alexey.gurevich@helmholtz-hips.de](mailto:alexey.gurevich@helmholtz-hips.de)     [@Alexey\\_Gurevich](https://twitter.com/Alexey_Gurevich)

<https://www.helmholtz-hips.de/en/research/teams/team/human-microbe-systems-bioinformatics/>