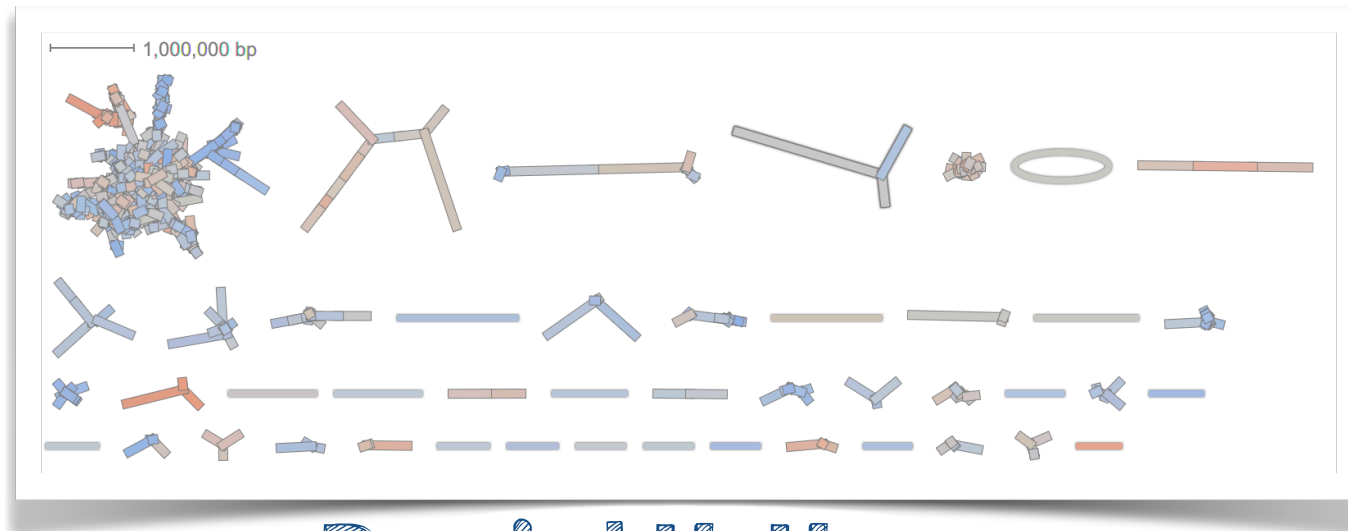


Analysis of long-read microbiome sequencing data



Daniel H. Huson

Outline

- Long-read metagenomics
- Assembly+DIAMOND+MEGAN-LR
- Mobile sequence analysis using MAIRA
- Bioreactor monitoring using MMonitor
- Analysis of long-read assemblies using ALORA

Outline

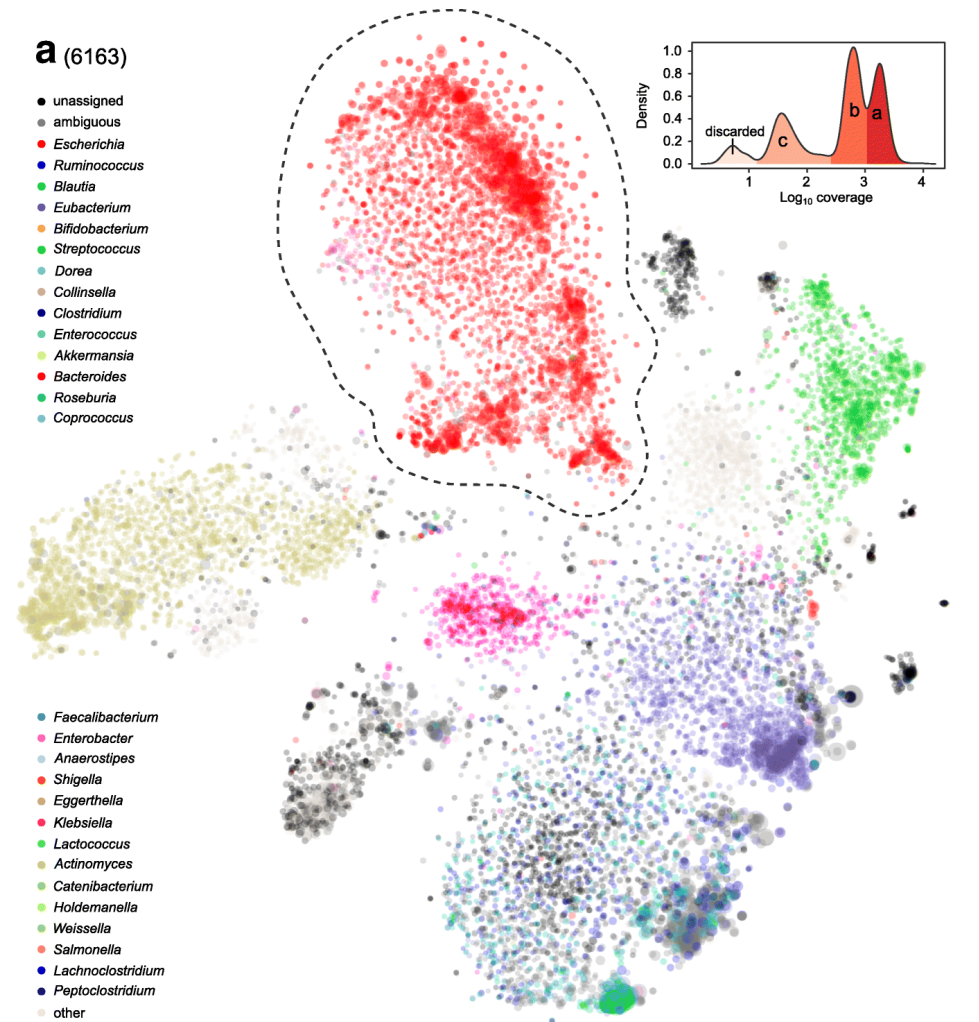
- Long-read metagenomics
- Assembly+DIAMOND+MEGAN-LR
- Mobile sequence analysis using MAIRA
- Bioreactor monitoring using MMonitor
- Analysis of long-read assemblies using ALORA

Microbiome Read-Length Paradox

- Short reads are short and plentiful...
 - So short-read datasets appear to require assembly
 - *But* resulting sequences are disappointingly short...
 - Long reads are long...
 - So usually longer than average assembled short reads
 - *But* assembly results in *very* long sequences
- ➔ Assembly of short reads is optional, but long reads should always be assembled...

Limitation of Short-Read Metagenomics

- *Assembly* of metagenomic short reads produces large numbers of tiny contigs - never complete chromosomes

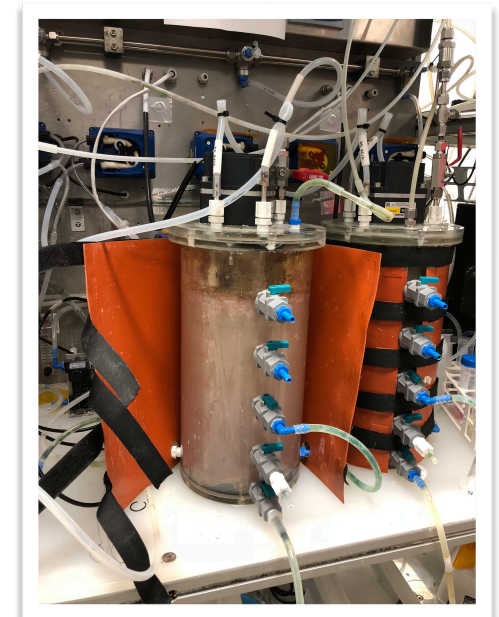


Long-Read Metagenomics

EBPR waste-water bio-reactor

- MinION sequencing 2018
 - Reads: ~695,000 (~ 6 Gb)
 - Length: ~9 kb mean (2 bp - 66 kb)
- Assembly (Unicycler): (Li 2016, Vaser *et al* 2017, Wick *et al*, 2017)
 - ~1,700 contigs
 - length 1.3 kb - 5.2 Mb (mean 61 kb)

Joint work with: Rohan Williams, Krithika Arumgam, Irina Bessarab and others at NUS and SCELSE



Krithika Arumgam



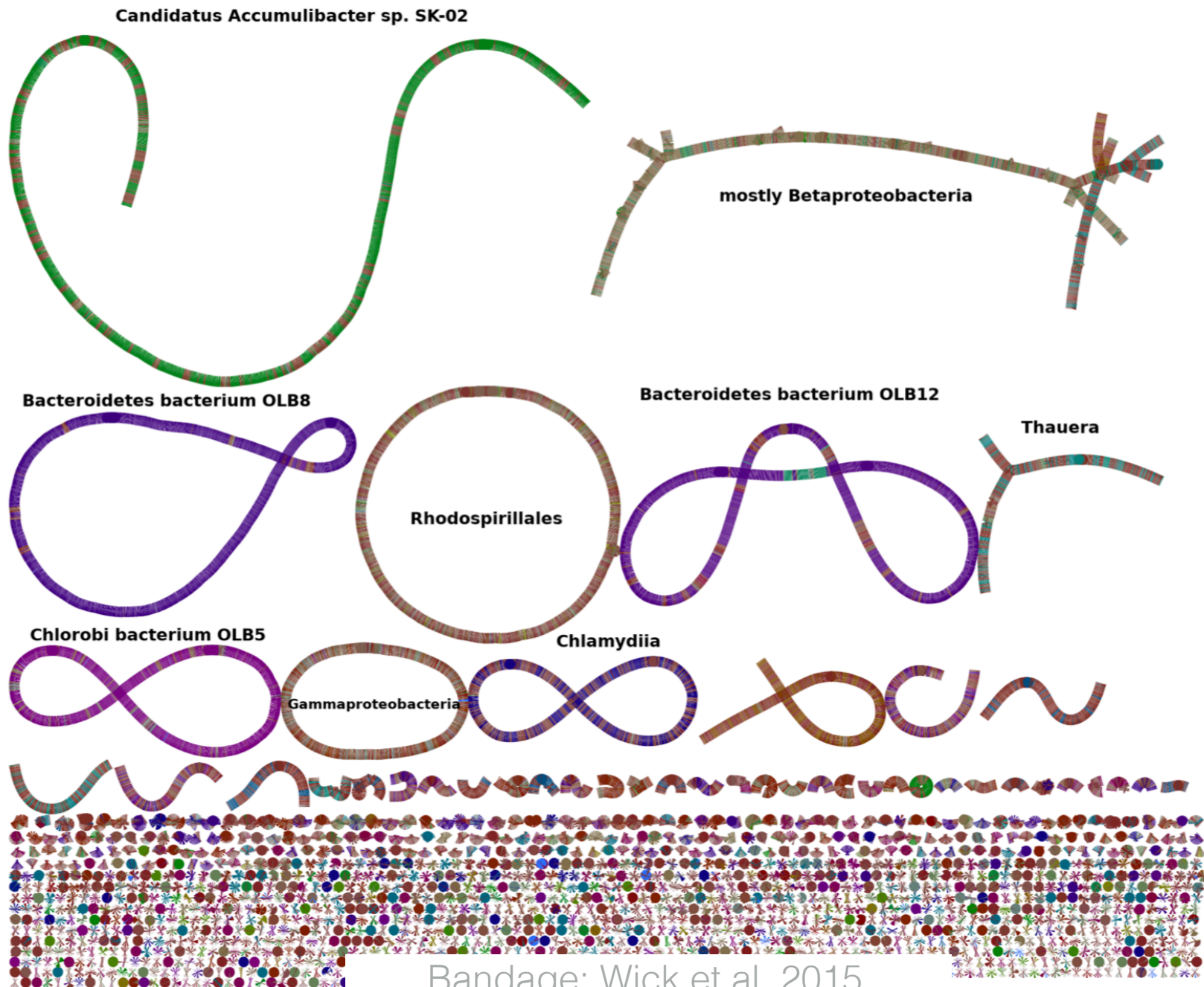
Short report | [Open Access](#) | Published: 16 April 2019

Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data

[Krithika Arumgam](#), [Caner Bağcı](#), [Irina Bessarab](#), [Sina Beier](#), [Benjamin Buchfink](#), [Anna Górska](#), [Guanglei Qiu](#), [Daniel H. Huson](#) & [Rohan B. H. Williams](#) ✉

Microbiome 7, Article number: 61 (2019) | [Cite this article](#)

Bandage Visualization of Assembly Graph



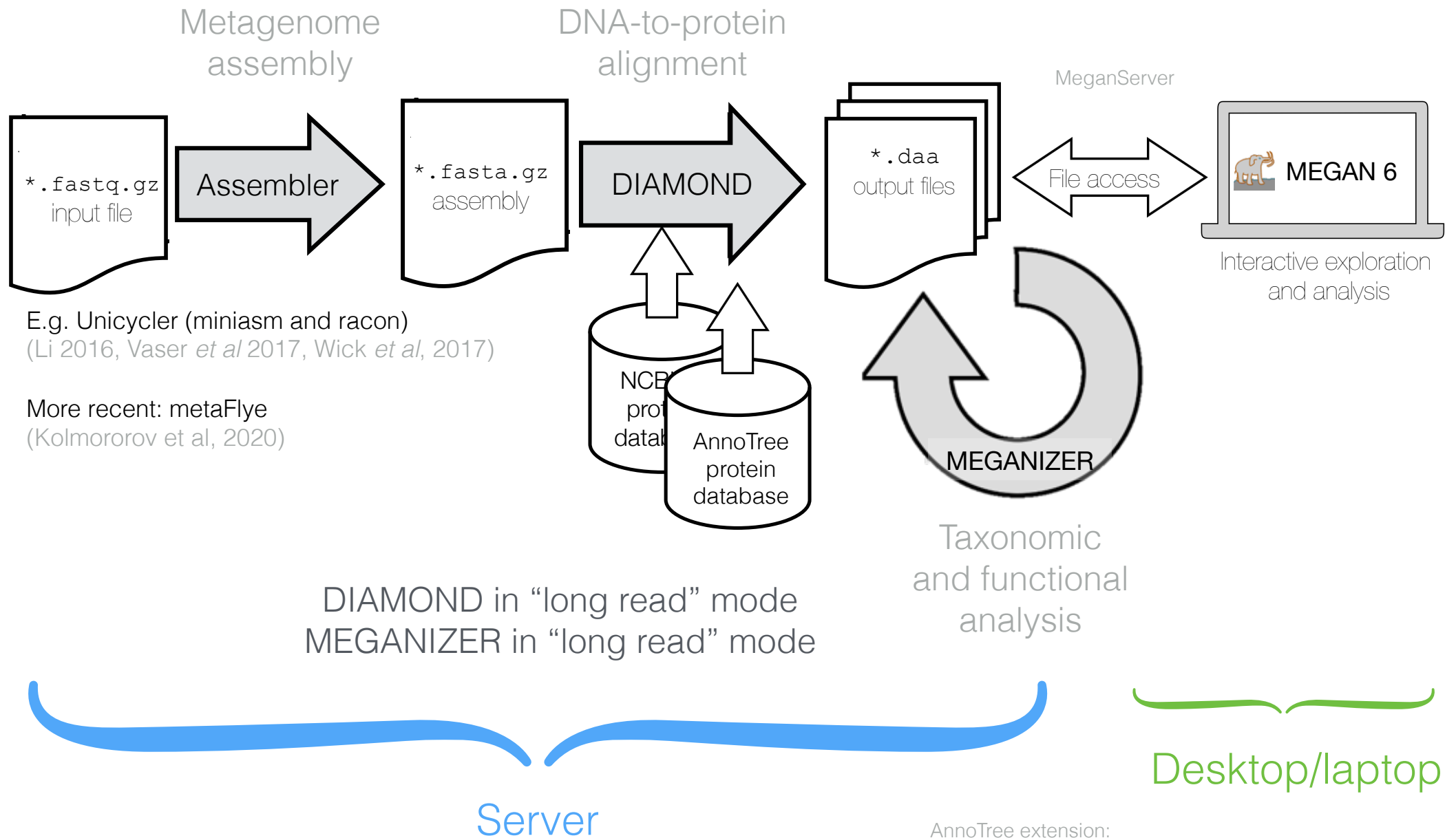
Bandage: Wick et al, 2015

Layout: Hachul S., Jünger M., 2007

Outline

- Long-read metagenomics
- **Assembly+DIAMOND+MEGAN-LR**
- Mobile sequence analysis using MAIRA
- Bioreactor monitoring using MMonitor
- Analysis of long-read assemblies using ALORA

Long-Read Analysis Pipeline



AnnoTree extension:
 Anupam Gautam
 Caner Bagci
 Hendrik Felderhoff

Taxonomic Bins $\geq 50\%$ Complete

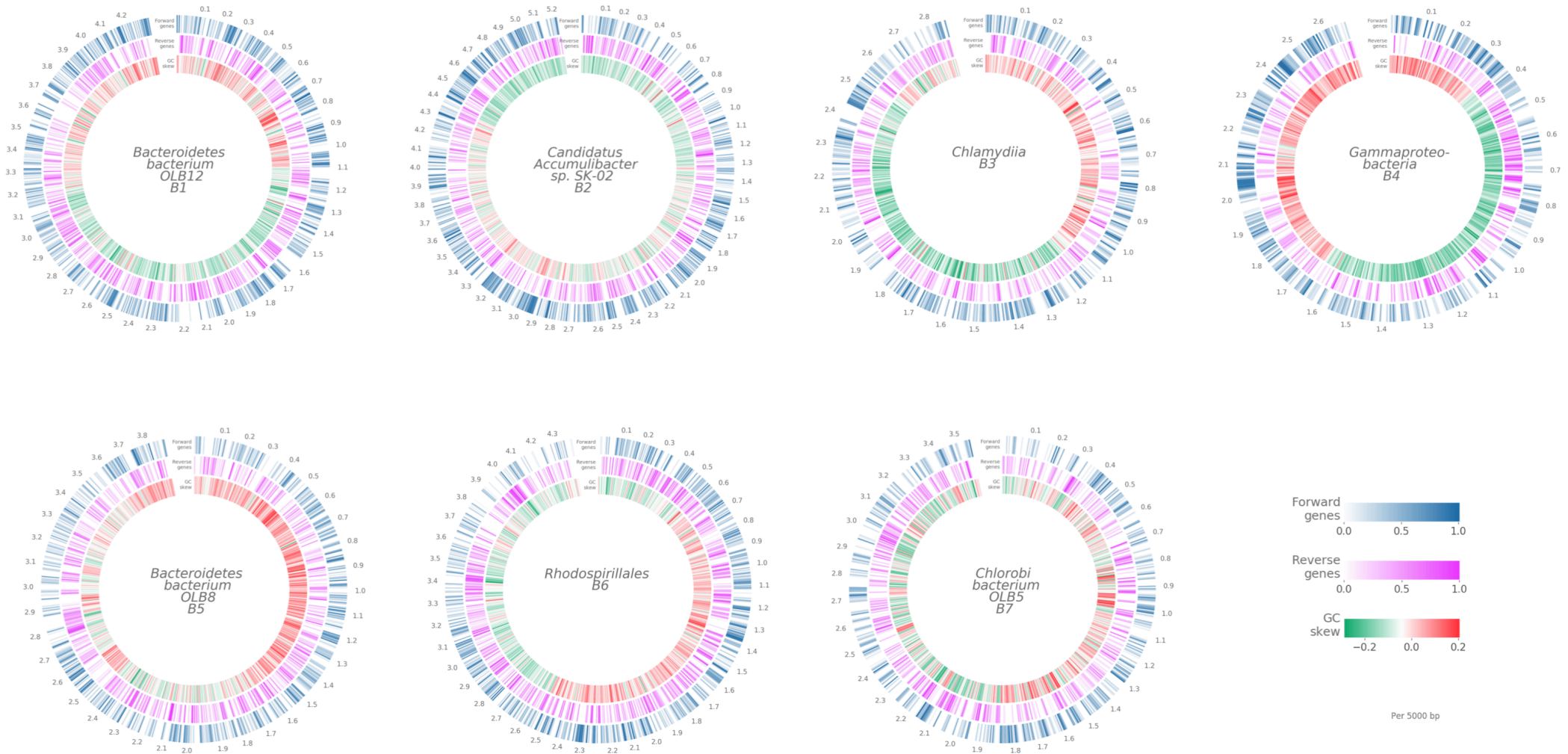
DIAMOND+MEGAN taxonomic bin		Unicycler contigs	Total (Mb)	Aligned (Mb)	Average coverage	CheckM		rRNA	Prokka tRNA	CDS
						Complete.	Contam.			
High quality draft genomes:										
B1	<i>Bacteroidetes bacterium</i> OLB12	1	4.2	3.5	57.3	95%	0.1%	6	39	4,163
B2	<i>Candidatus Accumulibacter</i> SK-02	1	5.2	4.1	384.2	94%	0.6%	4	53	4,915
B3	<i>Chlamydia</i> (class)	1	2.8	1.8	48.8	94%	2%	6	39	3,387
B4	<i>Gammaproteobacteria</i> (class)	43	4.7	3.0		93%	2%	6	52	4,833
	-longest contig		2.7	1.6	25.1	93%	0.2%	3	40	3,359
B5	<i>Bacteroidetes bacterium</i> OLB8	1	3.8	3.0	52.1	93%	1%	6	37	3,394
B6	<i>Rhodospirillales</i> (order)	1	4.4	3.0	29.5	92%	0.5%	3	47	4,015
B7	<i>Chlorobi bacterium</i> OLB5	1	3.5	2.5	38.7	88%	1%	3	41	4,131
Medium quality draft genomes:										
B8	<i>Thauera</i> (genus)	25	4.6	4.0		89%	4%	12	64	4,040
	-longest contig		0.8	0.7	32.7	14%	0%	0	5	672
B9	<i>Sphingobacteriales bacterium</i> 44-15	59	3.2	2.8		76%	1%	2	17	2,953
	-longest contig		0.2	0.1	10.2	0%	0%	0	0	172
B10	<i>Bacteroidetes</i> (phylum)	43	3.9	2.6		72%	7%	1	12	1,997
	-longest contig		1.2	0.8	14.1	32%	0%	0	3	807
B11	<i>Candidatus Contendobacter</i> B J11	39	2.5	2.0		59%	9%	2	37	2,668
	-longest contig		0.3	0.3	15.4	19%	0%	0	7	295
Low quality draft genomes:										
B12	<i>Betaproteobacteria</i> (class)	111	6.6	5.5		89%	79%	6	71	4,655
	-longest contig		0.4	0.3	37.1	10%	0%	0	1	372
B13	<i>Nitrospira</i> (genus)	34	4.2	3.7		83%	13%	0	6	563
	-longest contig		1.1	0.9	17.6	27%	0%	0	2	99
B14	<i>Chloroflexi</i> (phylum)	151	5.4	4.3		71%	29%	0	11	3,565
	-longest contig		0.2	0.2	13.3	8%	0%	0	1	86

Arumugam et al, 2019

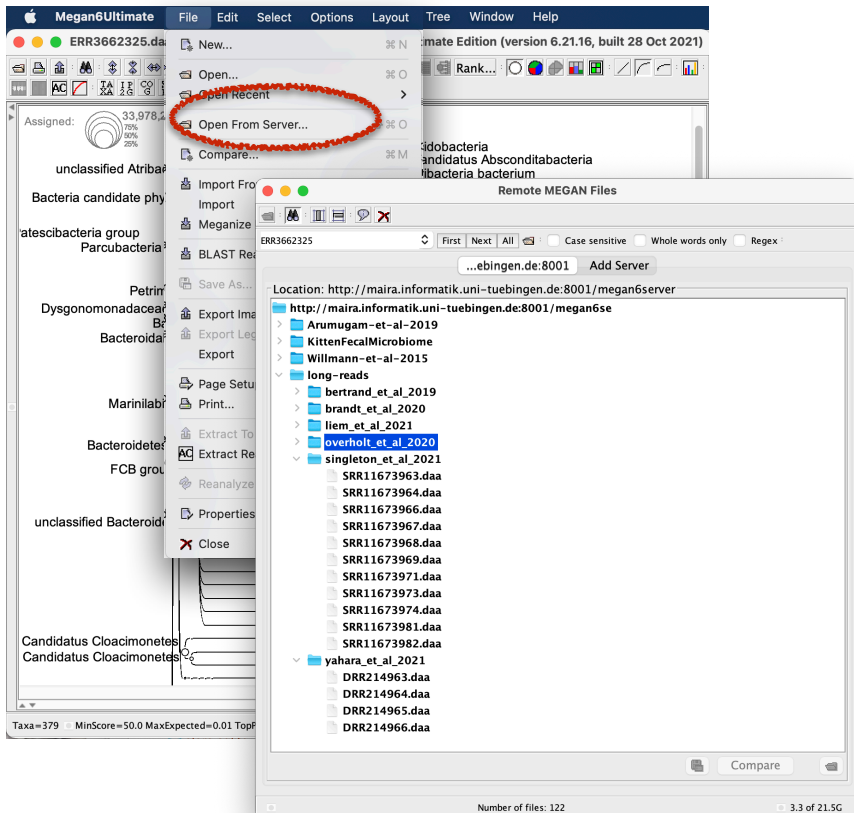
CheckM (Parks et al. 2014)

Prokka (Seemann, 2014)

Assembled Chromosomes



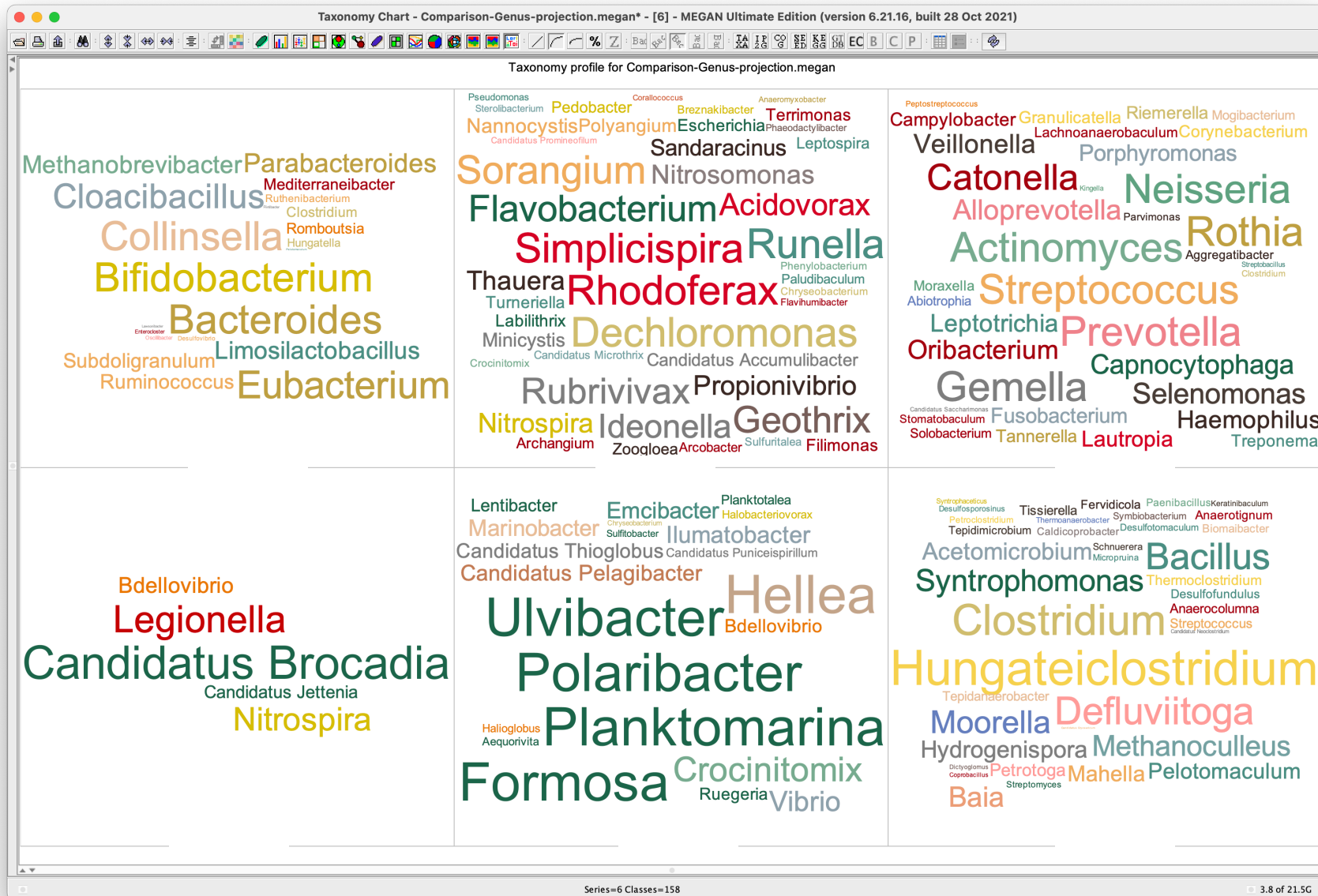
Long-Read Microbiomes



- Bertrand et al (2019)
 - 22 human gut samples, ONT
- Brandt et al (2020)
 - 20 biogas/wastewater samples, ONT
- Liem et al (2021)
 - 2 seawater samples, ONT
- Overholt et al (2020)
 - 1 groundwater samples, ONT
- Singleton et al (2021)
 - 11 wastewater samples, ONT
- Yahara et al (2021)
 - 4 oral samples, ONT

...via MeganServer

One Sample from Each Study



One Sample from Each Study

Quiz!

seawater

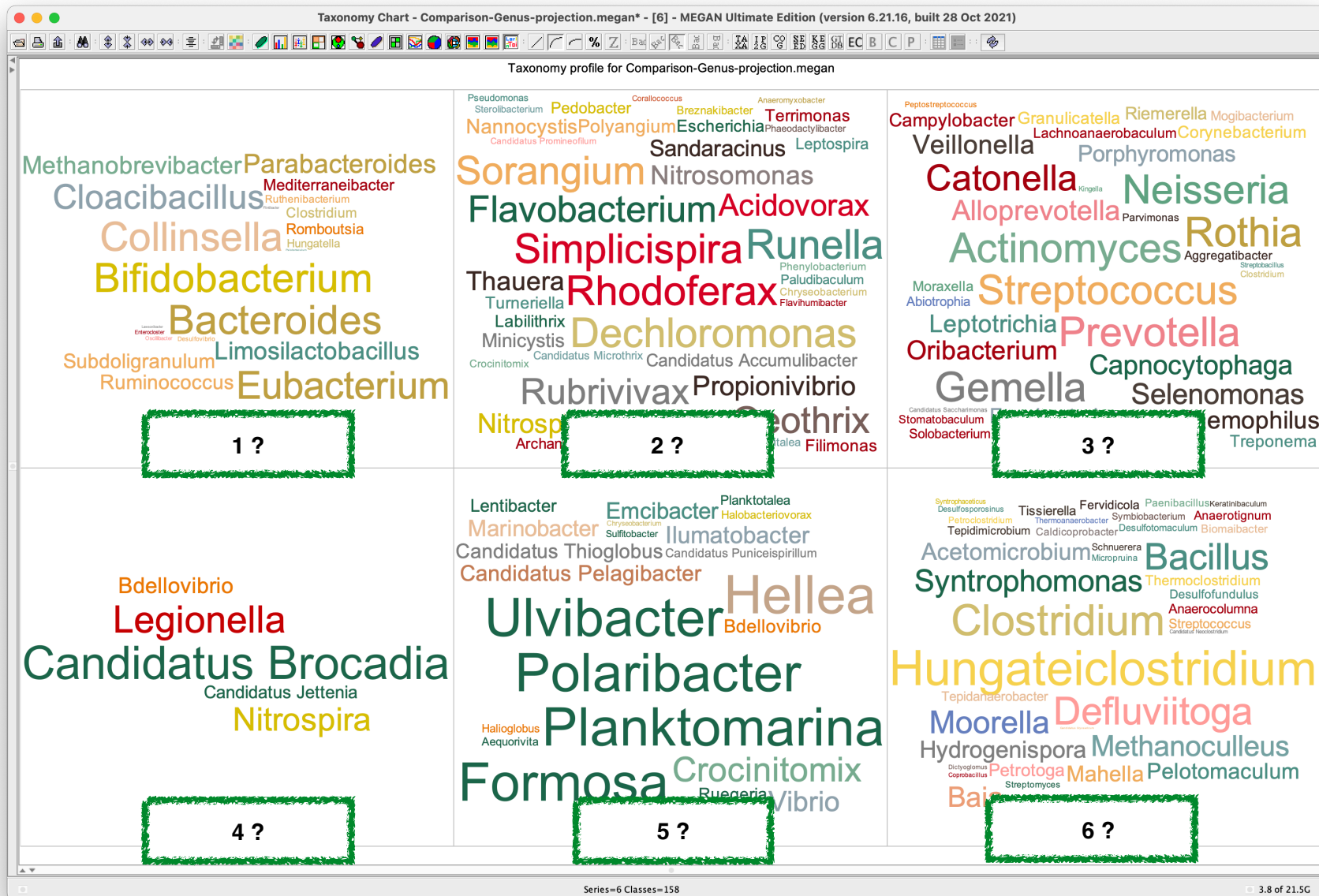
groundwater

human gut

human oral

wastewater

biogas



How to Use DIAMOND+MEGAN



PROTOCOL |  Open Access |   

DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences

Caner Bağcı, Sascha Patz, Daniel H. Huson 

First published: 03 March 2021 | <https://doi.org/10.1002/cpz1.59>

<https://doi.org/10.1002/cpz1.59>



Outline

- Long-read metagenomics
- Assembly+DIAMOND+MEGAN-LR
- **Mobile sequence analysis using MAIRA**
- Bioreactor monitoring using MMonitor
- Analysis of long-read assemblies using ALORA

Real-Time Sequence Analysis

- Laptop software for analysis of Nanopore reads
- Analyse sequences as they are produced
- Goal is to determine
 - strains,
 - virulence- pathogenicity- & antibiotic resistance factorsin the field

Benjamin Albrecht and Caner Bagci (funded by DFG)

MAIRA - Mobile Analysis of Long Reads



MAIRA software
by Benjamin Albrecht
and Caner Bagci
Funded by DFG



java

Untitled - MAIRA

NCBI Taxonomy Virulence factors Antibiotic resistance

Enter to filter table

Genus	Probability	Run Species
No content in table		

Enter to filter table

Species	Completeness	Coverage
No content in table		

MAIRA (version 0.1.6-beta, built 30 Aug 2019)

MAIRA

Mobile analysis of long reads
Benjamin Albrecht, Caner Bagcı & Daniel H. Huson
University of Tübingen, 2019

Lineage Info Composition

MAIRA - Mobile Analysis of Long Reads

The screenshot displays the MAIRA software interface. On the left, there are two tables. The top table lists various bacterial genera with their probabilities and run species status. The bottom table lists species with their completeness and coverage. The main area shows a taxonomic tree starting from the Bacteria superkingdom, branching into Proteobacteria and Firmicutes, and further into specific families and species. A progress bar at the bottom indicates that 14,000 reads have been processed.

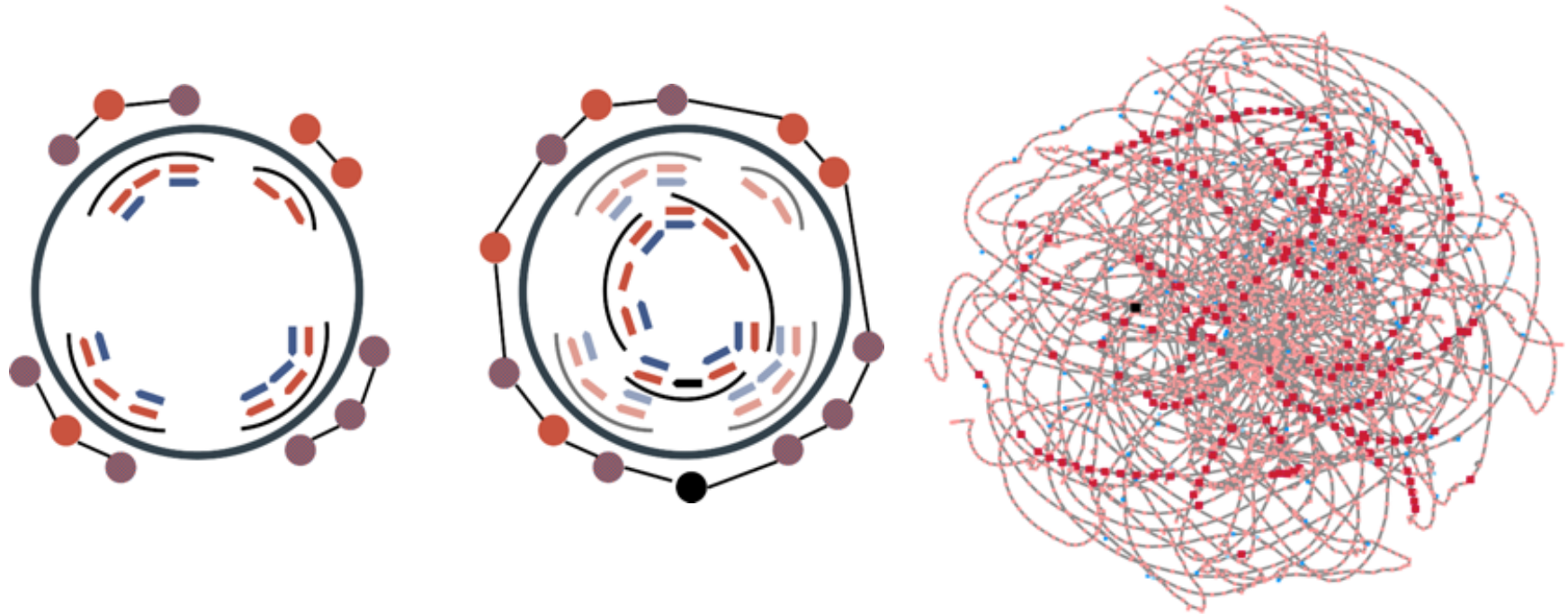
Genus	Probability	Run Species
547 Enteroba...	1.0	●●●●●
1578 Lactoba...	1.0	●●●●●
▼ Escherichia,S...	1.0	●●●●●
561 Escheri...	1.0	●●●●●
620 Shigella	1.0	●●●●●
1350 Enteroc...	1.0	●●●●●
1386 Bacillus	1.0	●●●●●
1279 Staphylo...	1.0	●●●●●
1637 Listeria	0.99	●●●●●
286 Pseudom...	0.74	●●●●●
590 Salmonella	0.33	●●●●●
1301 Streptoc...	0.16	●●●●●
570 Klebsiella	0.13	●●●●●
2737 Vagococ...	0.11	●●●●●
544 Citrobacter	0.08	●●●●●

Species	Completeness	Coverage
1280 Staphylococ...	0.91	2.82
562 Escherichia coli	0.74	2.91
1351 Enterococcu...	0.78	3.21
1639 Listeria mon...	1.0	2.98
1613 Lactobacillus...	0.77	3.29
1390 Bacillus amy...	0.39	2.62
1423 Bacillus subt...	0.76	2.92

Reads processed: 2k... 3k... 4k... 6k... 12k... 14k...

Protein Synteny Graph

- Taxonomic and functional analysis uses a graph that contains all proteins and models genomes:



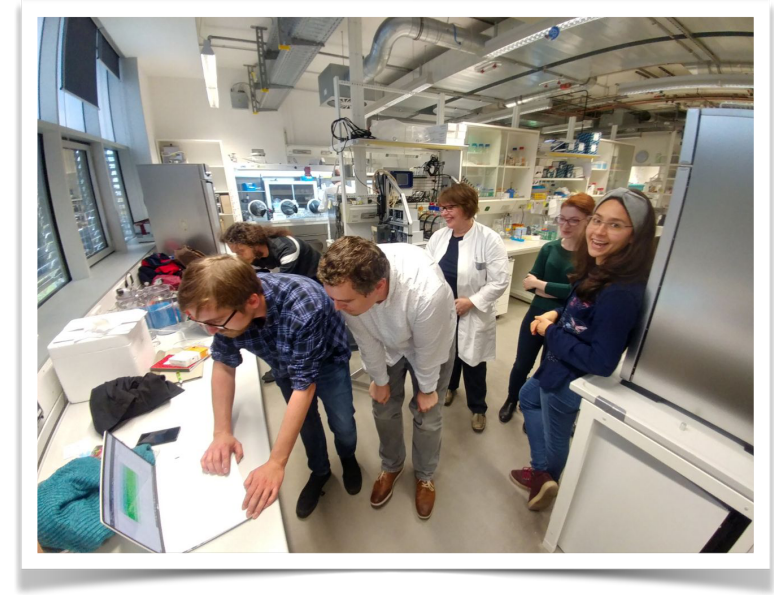
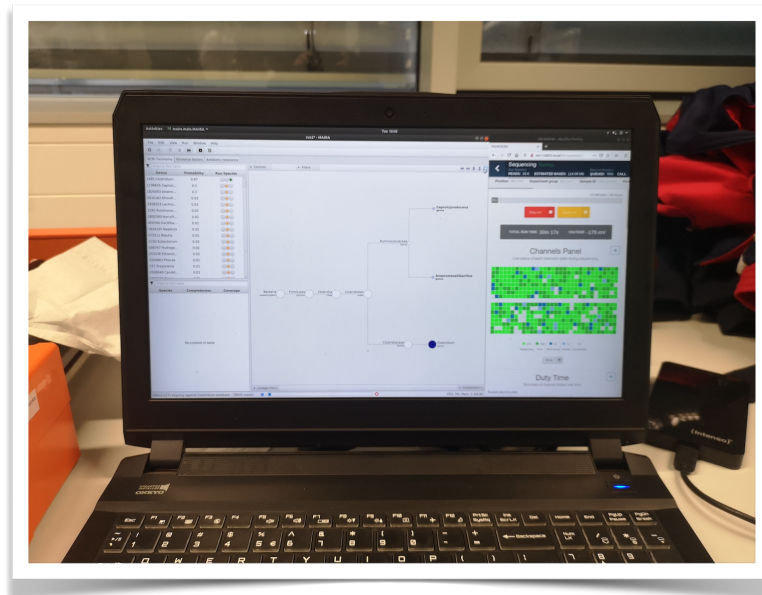
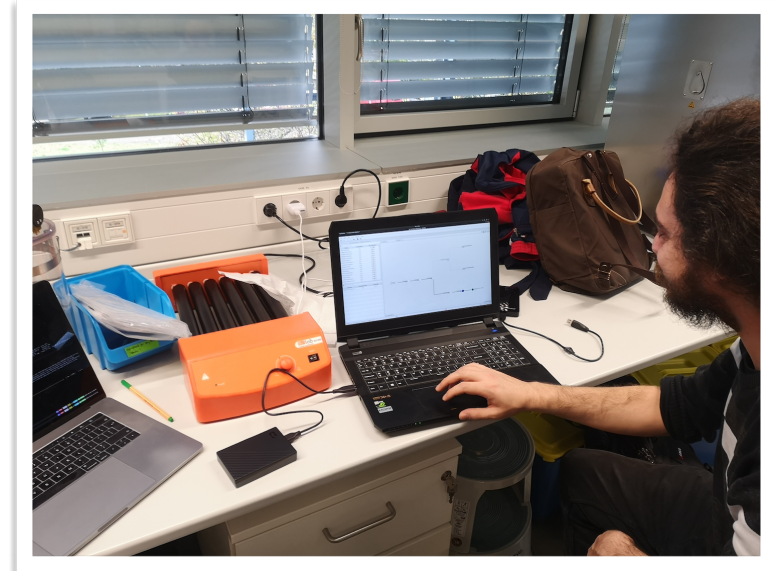
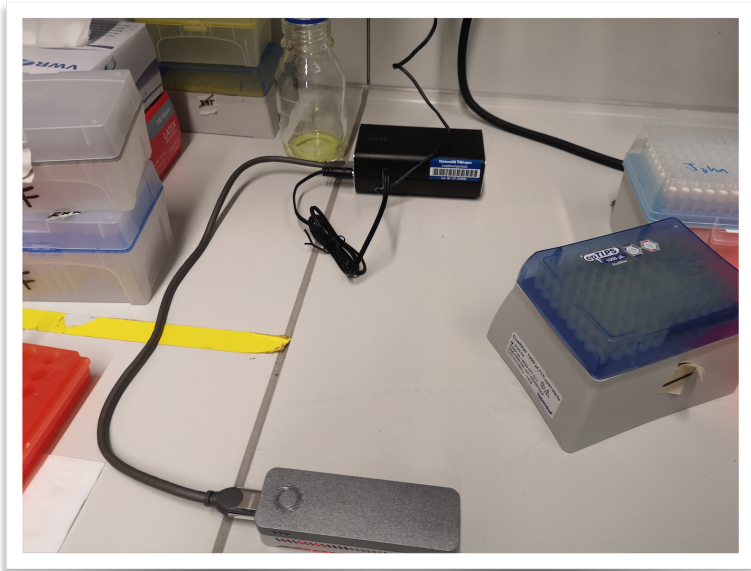
Research | [Open Access](#) | [Published: 17 September 2020](#)

MAIRA- real-time taxonomic and functional analysis of long reads on a laptop

[Benjamin Albrecht](#), [Caner Bağcı](#) & [Daniel H. Huson](#) 

[BMC Bioinformatics](#) **21**, Article number: 390 (2020) | [Cite this article](#)

Running MinION, MinIT and MAIRA



Outline

- Long-read metagenomics
- Assembly+DIAMOND+MEGAN-LR
- Mobile sequence analysis using MAIRA
- **Bioreactor monitoring using MMonitor**
- Analysis of long-read assemblies using ALORA

Bioreactor Monitoring

Bioreactor experiments:

- Grow microbes under controlled conditions
 - Understand and/or optimize specific processes
 - Monitor experimental parameters
 - Collect biomass samples and sequence *later*
- ★ Use MinION to monitor community *during* experiment...

CMFI project with Lars Angenent (Biotechnology) and Ruth Ley (Microbiome). PhD students Ulrike Biehain (Bioreactors) and Timo Lucas (Bioinformatics)

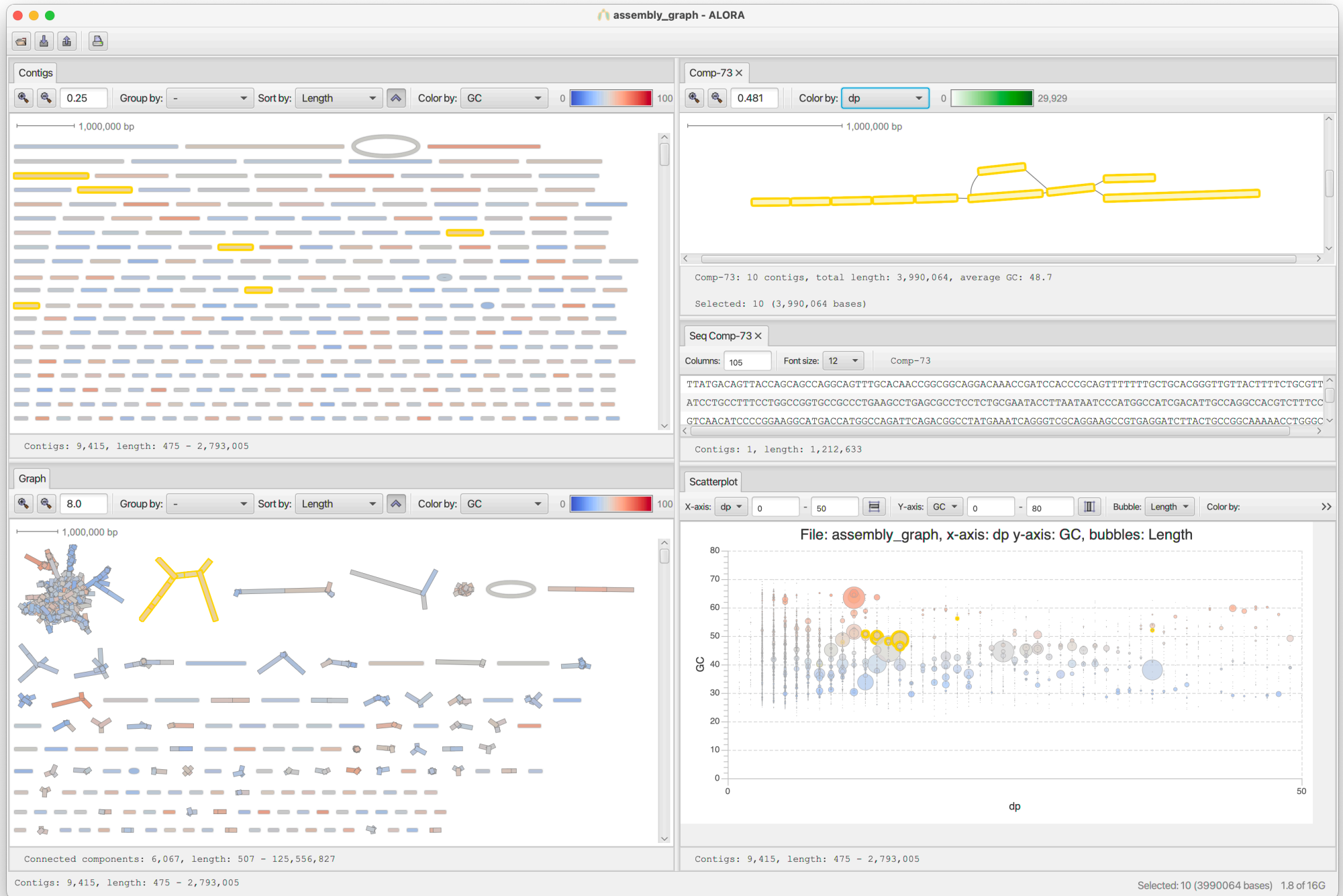
MMonitor



Outline

- Long-read metagenomics
- Assembly+DIAMOND+MEGAN-LR
- Mobile sequence analysis using MAIRA
- Bioreactor monitoring using MMonitor
- Analysis of long-read assemblies using ALORA

ALORA - Analysis of Long Read Assemblies



Summary

- Long-read metagenomics is much more fun...
- Try: Assembly+DIAMOND+MEGAN+MeganServer
- **M**A**I**R**A** for laptop-based analysis while-u-wait
- MMonitor for microbiome monitoring
- **ALORA** for interactive analysis of long-read assemblies

Thank You!

Joint work with:

- Benjamin Albrecht, Caner Bagcii, Anupam Gautam, Timo Lucas & Sascha Patz Tübingen
- Irina Bessarab, Krithika Arumugam and Rohan Williams SCELSE Singapore
- Lars Angenent and Ulrike Biehn Biotechnology, Tübingen
- Ruth Ley MPI Microbiome Science, Tübingen

Funding:

- Life Sciences Institute at NUS, NRF/MOE and NRF-EW, Singapore (MEGAN-LR)
- Deutsche Forschungsgemeinschaft (MAIRA & BinAC)
- DFG EXC2124 Controlling Microbes to Fight Infection (MMonitor)