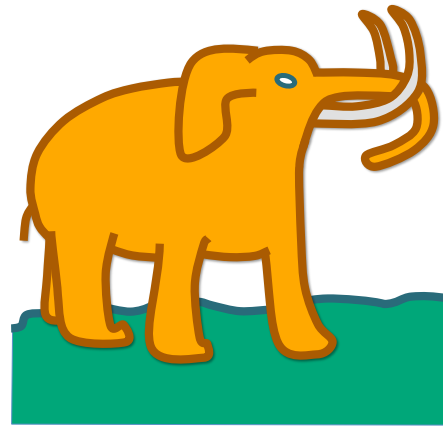


Introduction to Microbiome Analysis using **DIAMOND+MEGAN**



Daniel H. Huson



Outline

- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Outline

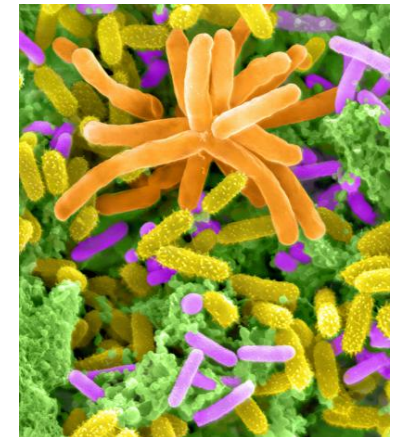
- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Microbiome

- Traditionally, microbes are studied in pure culture
- **Genome:**
 - Entire DNA sequence of a single organism
- *But:* most microbes don't live in isolation and many can't be cultured
- **Microbiome:**
 - Collection of microbes in a specific theatre of activity
- **Metagenome:**
 - Entire DNA sequence of a microbiome



www.innovations-report.de



www.physorg.com

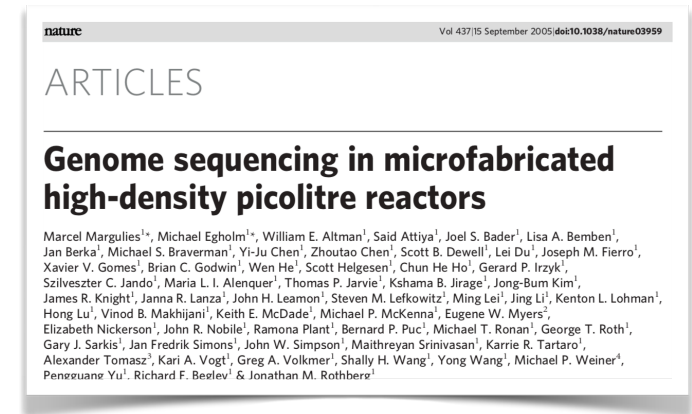
Sources of Studied Microbiomes

- Soil samples
- Water samples
- Seabed samples
- Air samples
- Ancient bones
- Host-associated samples
- Human microbiome
- ...





- First NGS technique 454 released
- Intended for genome sequencing...



★ Use NGS to sequence ancient DNA?

★ Use NGS to sequence metagenomic DNA?

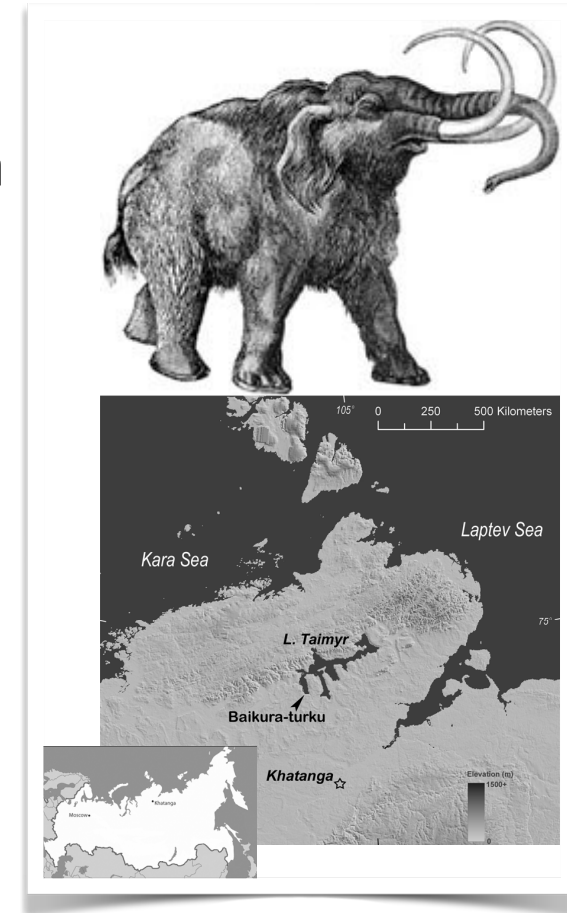


NGS = next generation sequencing

Mammoth DNA & Metagenome (2006)

- DNA collected from permafrost mammoth (28,000 years old)
- DNA extracted from 1g bone
- DNA sheared to 500-700 bp
- Sequenced using 454
- ~302,000 reads, length ~95 bp

- ★ Can use NGS for ancient DNA
- ★ First NGS metagenomics paper



REPORTS

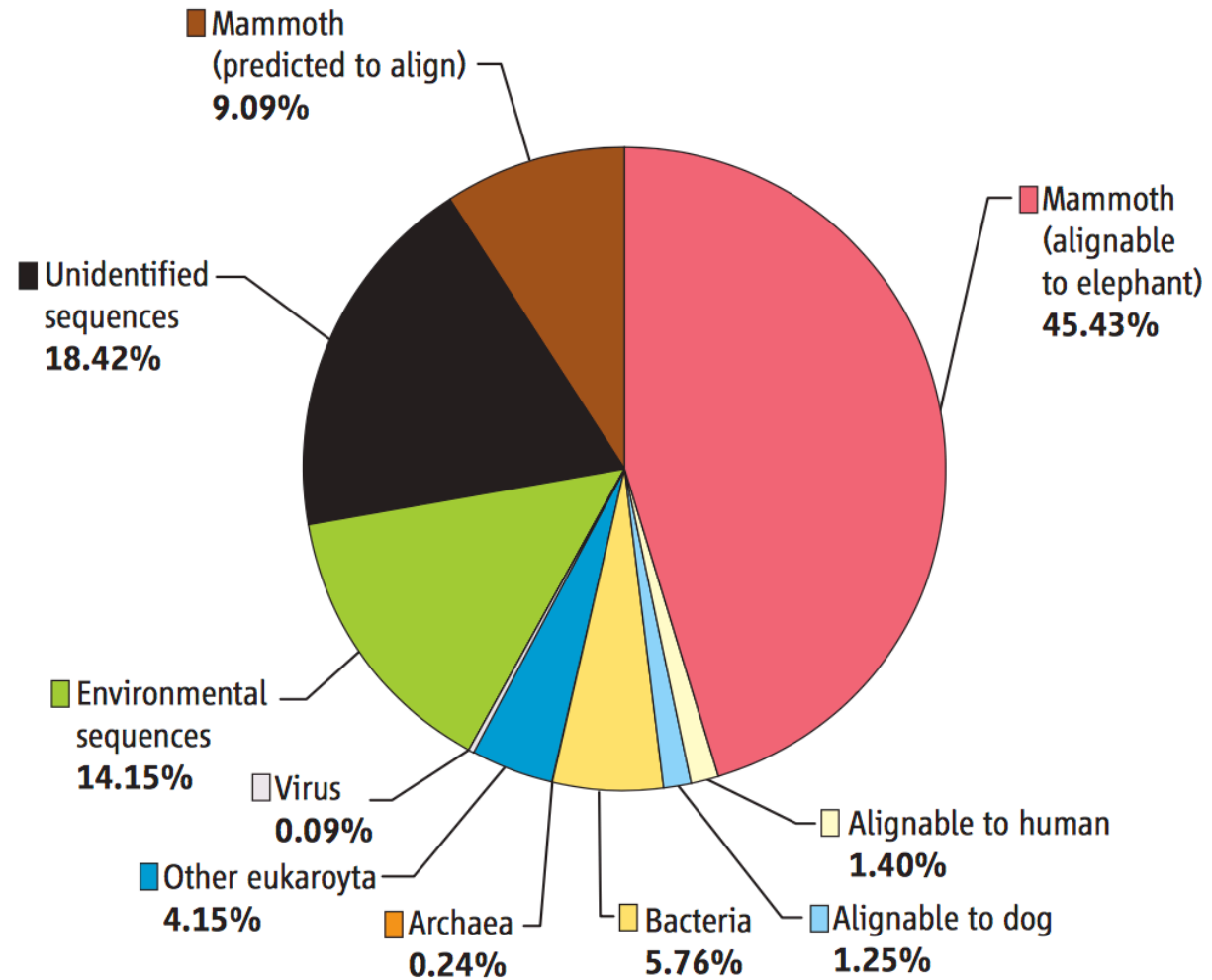
Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA

Hendrik N. Poinar,^{1,2,3*} Carsten Schwarz,^{1,2} Ji Qi,⁴ Beth Shapiro,⁵ Ross D. E. MacPhee,⁶ Bernard Buigues,⁷ Alexei Tikhonov,⁸ Daniel H. Huson,⁹ Lynn P. Tomsho,⁴ Alexander Auch,⁹ Markus Rampp,¹⁰ Webb Miller,⁴ Stephan C. Schuster^{4**}

Science, 2006

Mammoth Bone Metagenome (2006)

Fig. 1. Characterization of the mammoth metagenomic library, including percentage of read distributions to various taxa. Host organism prediction based on BLASTZ comparison against GenBank and environmental sequences database.

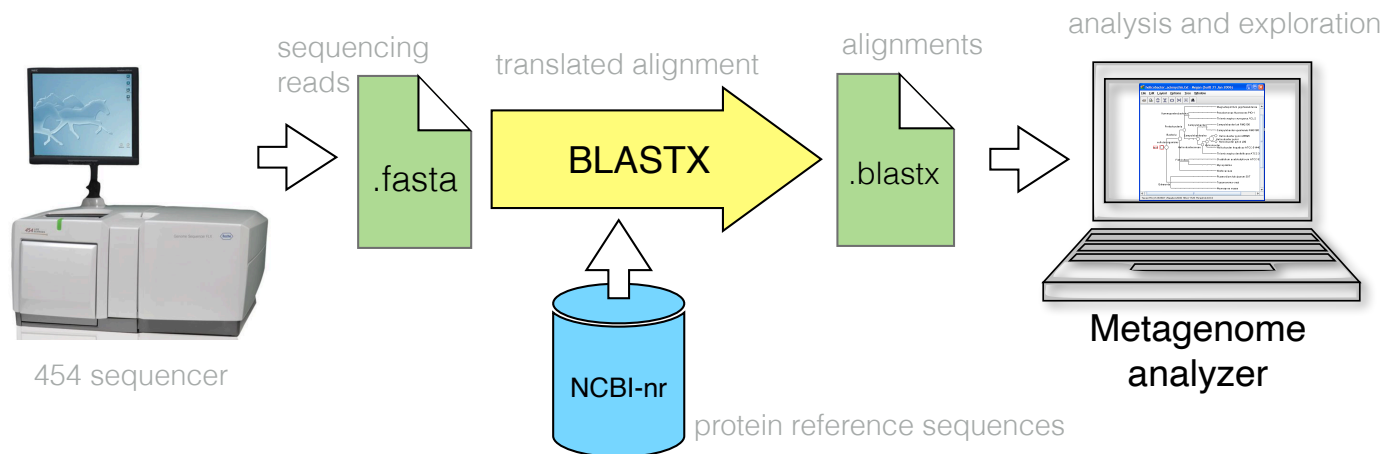


Poinar et al, Science 2006

How to Analyze Metagenomic Reads? (2006)

Basic idea (with Stephan Schuster at Penn State):

- BLASTX non-host reads against NCBI-nr
- Assign reads to NCBI taxonomy using naive LCA (lowest common ancestor) approach
- Develop GUI to explore assignments and alignments

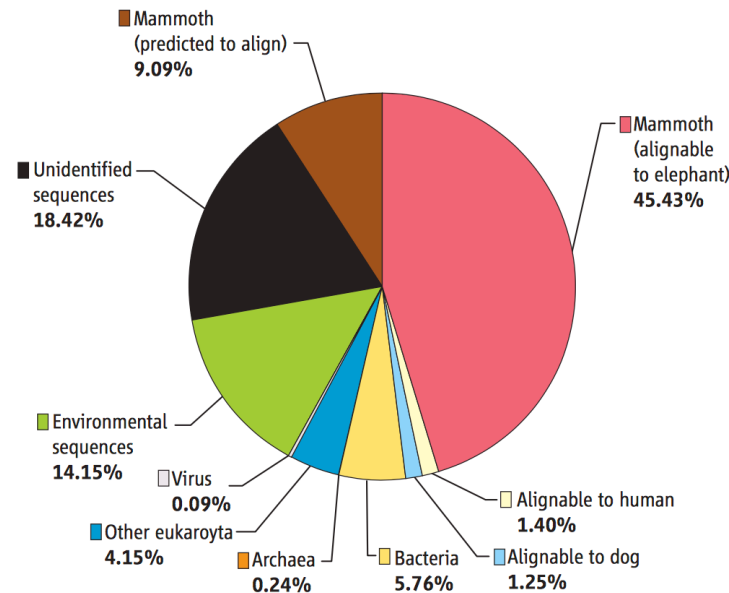


2006 MEGAN analysis pipeline

How to Analyze Metagenomic Reads? (2006)

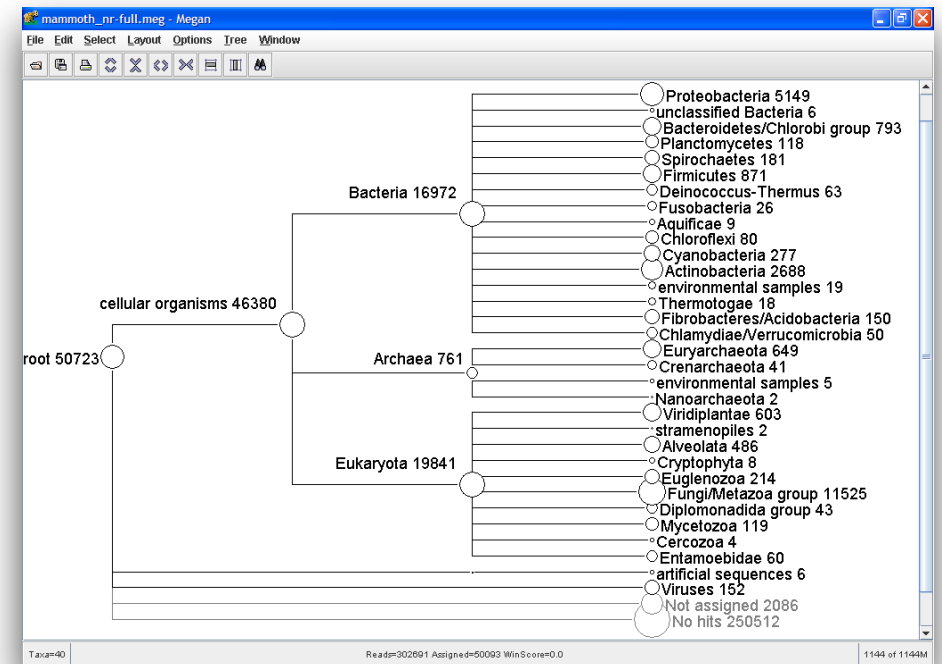
- MEGAN (MEtagenome ANalyzer 1.0)

Fig. 1. Characterization of the mammoth metagenomic library, including percentage of read distributions to various taxa. Host organism prediction based on BLASTZ comparison against GenBank and environmental sequences database.



Poinar et al, Science 2006

MEGAN 1.0



H. et al, Genome Research, 2007

Computational Bottleneck (2006)

- Compare all reads against the NCBI-nr protein database
- Year 2006:
 - 300,000 reads of length ~100bp
 - NCBI-nr: 3 million entries, ~1 billion letters
- ★ **BLASTX** took a couple of weeks on a small cluster

(NCBI-nr today: ~ 250 million entries)

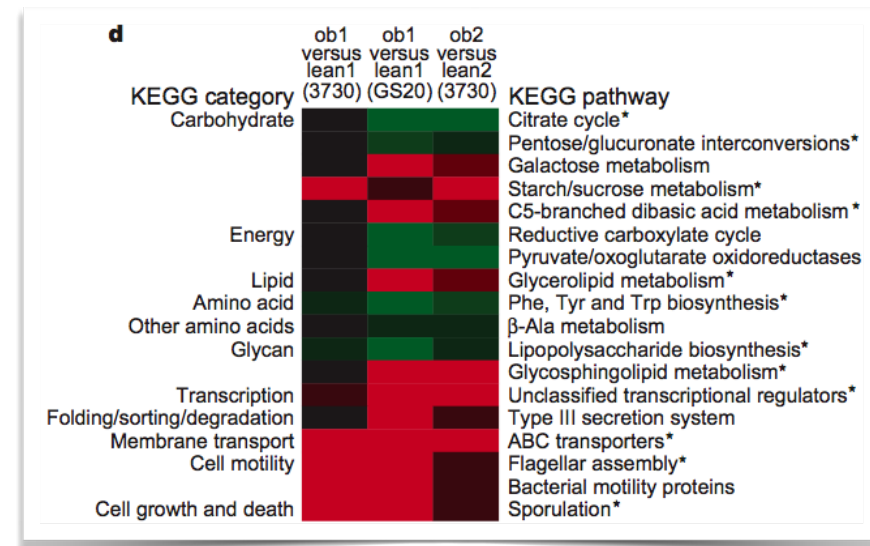
Obesity-Associated Gut Microbiome

Turnbaugh *et al* (2006):

- Caecal microbial DNA of ob/ob, ob/+, +/+ mice
- Sanger sequencing:
 - 39.5 Mb
 - read length 750 bp
- 454 sequencing:
 - 160 Mb
 - read length 93 bp



- Change in relative abundance of Bacteroidetes and Firmicutes
- Change in functional capacity (toward energy harvesting)



Vol 464 | 4 March 2010 | doi:10.1038/nature08821

nature

MetaHIT 2010

ARTICLES

A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin^{1*}, Ruiqiang Li^{1*}, Jeroen Raes^{2,3}, Manimozhiyan Arumugam², Kristoffer Solvsten Burgdorf⁴, Chaysavanh Manichanh⁵, Trine Nielsen⁴, Nicolas Pons⁶, Florence Levenez⁶, Takuji Yamada², Daniel R. Mende², Junhua Li^{1,7}, Junming Xu¹, Shaochuan Li¹, Dongfang Li^{1,8}, Jianjun Cao¹, Bo Wang¹, Huiqing Liang¹, Huisong Zheng¹, Yinlong Xie^{1,7}, Julien Tap⁶, Patricia Lepage⁶, Marcelo Bertalan⁹, Jean-Michel Batto⁶, Torben Hansen⁴, Denis Le Paslier¹⁰, Allan Linneberg¹¹, H. Bjørn Nielsen⁹, Eric Pelletier¹⁰, Pierre Renault⁶, Thomas Sicheritz-Ponten⁹, Keith Turner¹², Hongmei Zhu¹, Chang Yu¹, Shengting Li¹, Min Jian¹, Yan Zhou¹, Yingrui Li¹, Xiuqing Zhang¹, Songgang Li¹, Nan Qin¹, Huanming Yang¹, Jian Wang¹, Søren Brunak⁹, Joel Doré⁶, Francisco Guarner⁵, Karsten Kristiansen¹³, Oluf Pedersen^{4,14}, Julian Parkhill¹², Jean Weissenbach¹⁰, MetaHIT Consortium†, Peer Bork², S. Dusko Ehrlich⁶ & Jun Wang^{1,15}

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

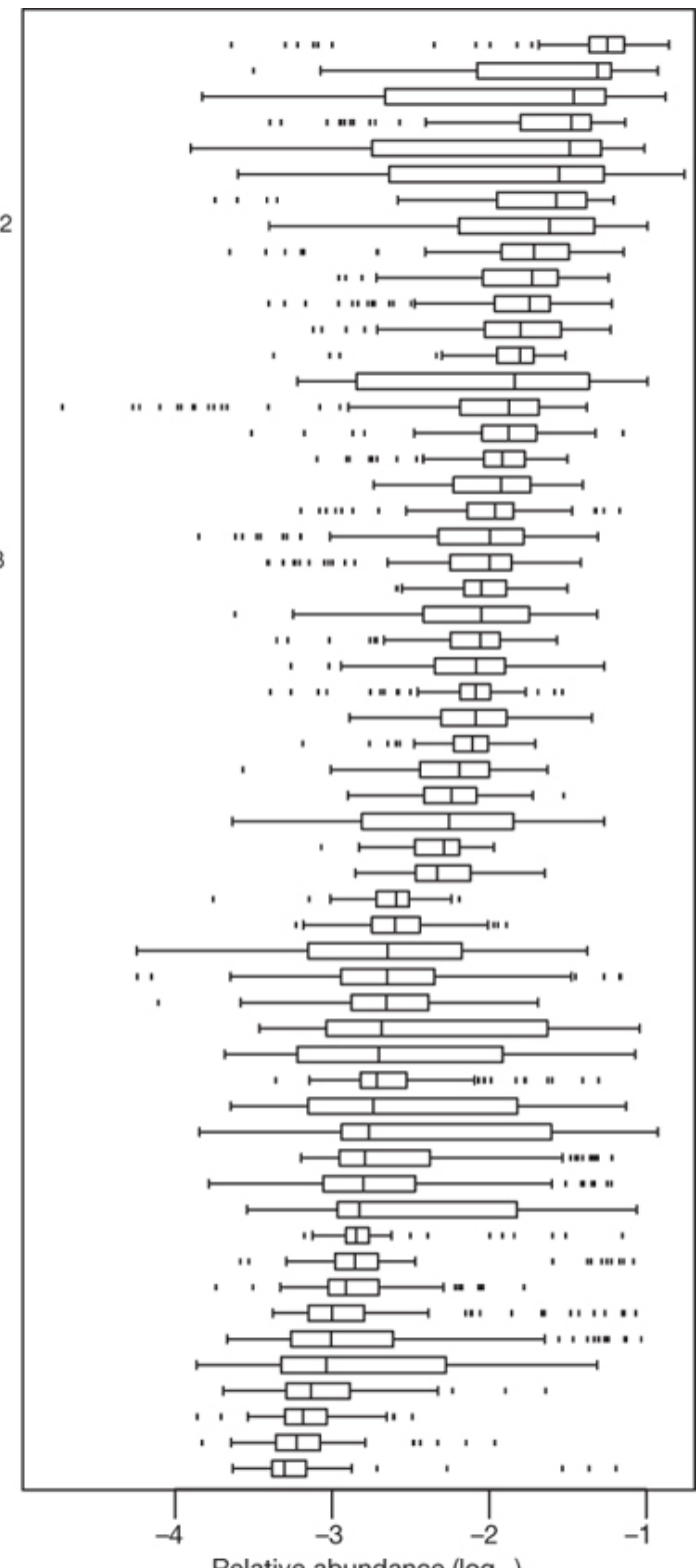
- 576Gb of sequence from 124 individuals

Core of Human Gut Microbiome

- 57 species present in $\geq 90\%$ of individuals with coverage $> 1\%$
- High variability
- Bacteroidetes and Firmicutes most abundant

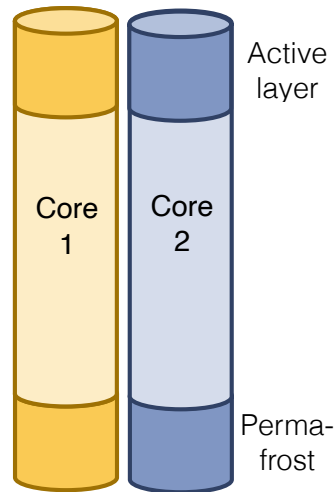
BLASTX at Super Computer
 Center in Barcelona, then
 MEGAN analysis

Bacteroides uniformis
Alistipes putredinis
Parabacteroides merdae
Dorea longicatena
Ruminococcus bromii L2-63
Bacteroides caccae
Clostridium sp. SS2-1
Bacteroides thetaiotaomicron VPI-5482
Eubacterium hallii
Ruminococcus torques L2-14
 Unknown sp. SS3 4
Ruminococcus sp. SR1 5
Faecalibacterium prausnitzii SL3 3
Ruminococcus lactaris
Collinsella aerofaciens
Dorea formicigenerans
Bacteroides vulgatus ATCC 8482
Roseburia intestinalis M50 1
Bacteroides sp. 2_1_7
Eubacterium siraeum 70 3
Parabacteroides distasonis ATCC 8503
Bacteroides sp. 9_1_42FAA
Bacteroides ovatus
Bacteroides sp. 4_3_47FAA
Bacteroides sp. 2_2_4
Eubacterium rectale M104 1
Bacteriodes xylanisolvens XB1A
Coprococcus comes SL7 1
Bacteroides sp. D1
Bacteroides sp. D4
Eubacterium ventriosum
Bacteroides dorei
Ruminococcus obeum A2-162
Subdoligranulum variabile
Bacteroides capillosus
Streptococcus thermophilus LMD-9
Clostridium leptum
Holdemania filiformis
Bacteroides stercoris
Coprococcus eutactus
Clostridium sp. M62 1
Bacteroides eggerthii
Butyrivibrio crossotus
Bacteroides finegoldii
Parabacteroides johnsonii
Clostridium sp. L2-50
Clostridium nexile
Bacteroides pectinophilus
Anaerotruncus colihominis
Coprococcus gnauvus
Bacteroides intestinalis
Bacteroides fragilis 3_1_12
Clostridium asparagiforme
Enterococcus faecalis TX0104
Clostridium scindens
Blautia hansenii



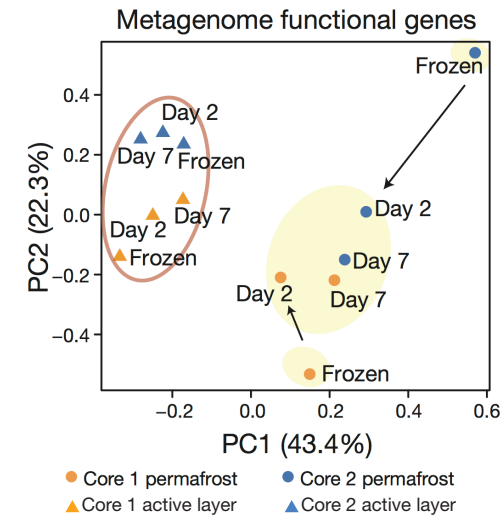
Permafrost Study (2011)

(Mackelprang *et al*, Science 2011)



Their question:
Functional changes
during thawing?

Frozen, day 2, day 7



- Align ~250 million Illumina reads against KEGG
- 800,000 CPU hours at Super Computer Center in Berkeley



on 100 cores

Outline

- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Translated Alignment

- Read:

```
>HISEQ:457:C5366ACXX:2:1101:5937:60460 (101 bases)
TTATATTAATTAGAAAACCAATTAAAATACGAACGTTATGAAGAAGTACATTTGC...
```

- Translation (frame +3):

```
..I L I R K P I K N T N V M K K Y I C ...
```

- Translated alignment:

```
>EEC52678.1 Length = 65
```

```
Score = 56 bits (135), Expect = 1e-05
```

```
Identities = 22/33 (67%), Positives = 27/33 (82%), Gaps = 0/33 (0%)
```

```
Frame = +3
```

```
Query:          3  ILIRKPIKNTNVMKKYICTVCEYIYDPEQGDPE 101
                  +L +K  K   VM+KYICT+CEY+YDPEQGDPE
Sbjct:          1  MLSKKKFKQKRVMEKYICTICEYVYDPEQGDPE 33
```

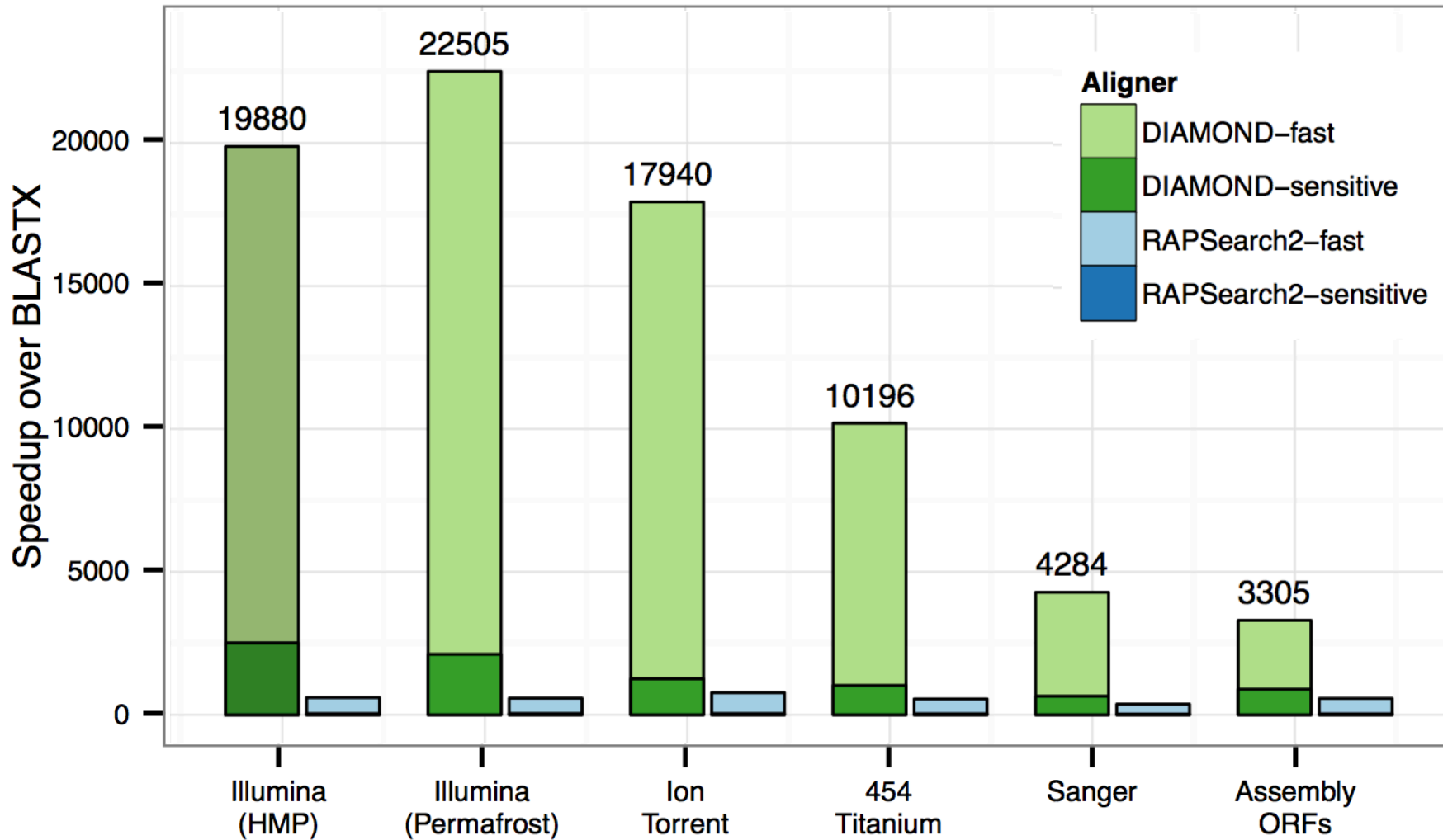
DIAMOND BLAST!

- Translated alignment tool DIAMOND
- DIAMOND replaces BLASTX on microbiome sequencing reads
- Very similar sensitivity to BLASTX on short reads
- Much, much faster...

**Fast and sensitive protein
alignment using DIAMOND**

Benjamin Buchfink¹, Chao Xie^{2,3} &
Daniel H Huson^{1,2} **NATURE METHODS** 2015

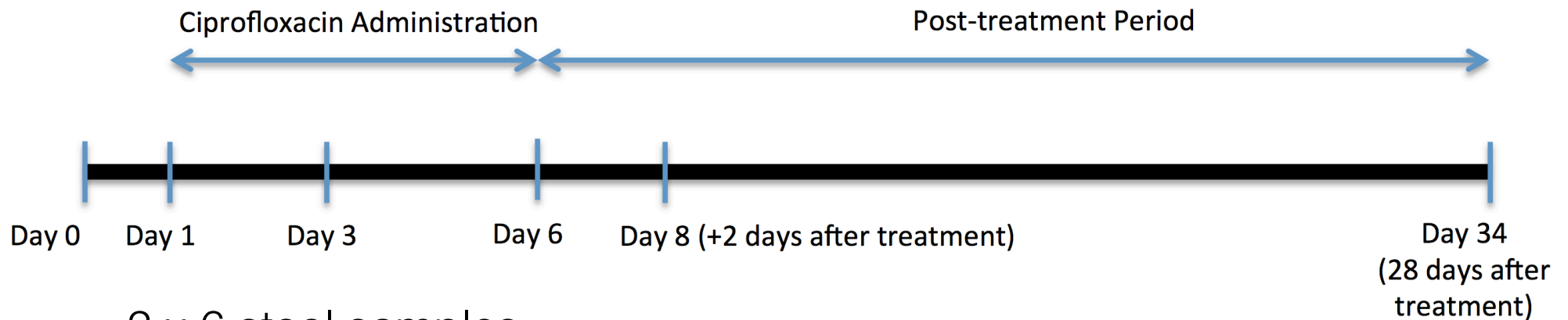
DIAMOND Performance



ASARI- Antibiotic Resistance Pilot Study



- Two volunteers, subject 1 and subject 2



- 2 x 6 stool samples
- Shotgun sequencing
 - ~60 million reads per sample (101 bp per read)
 - ~800 million reads in total
- Initial analysis: compare against NCBI-nr protein database

Performance of DIAMOND+MEGAN

- 12 human gut samples, total 816 million HiSeq reads

Sample	Reads	DIAMOND (s)	Alignments	Aligned reads	Meganizer (s)
Alice 0	66 393 401	19 062	627 405 772	44 900 227	9 299
Alice 1	64 923 975	15 771	595 715 349	43 498 105	11 338
Alice 3	55 092 349	13 435	515 249 349	37 675 494	8 621
Alice 6	66 289 376	16 801	910 892 059	52 627 776	11 771
Alice 8	57 957 661	14 134	790 946 244	45 358 448	13 911
Alice 34	64 380 386	15 615	608 114 143	44 741 897	11 962
Bob 0	61 232 588	14 573	825 213 917	48 882 884	12 058
Bob 1	65 763 766	16 203	841 038 616	51 408 892	12 270
Bob 3	89 034 641	34 598	1 233 571 041	72 017 720	15 789
Bob 6	89 339 172	27 333	1 138 796 522	70 344 161	15 507
Bob 8	78 001 118	19 734	1 049 831 855	63 336 241	13 423
Bob 34	57 627 119	15 406	780 844 319	455 681 58	11 433
Total	816 035 552	222 665	9 917 619 186	620 360 003	Max: 15 789
Time		≈ 62 h			≈ 5 h

doi:10.1371/journal.pcbi.1004957.t001

- Complete analysis in 62+5 hours on a single server



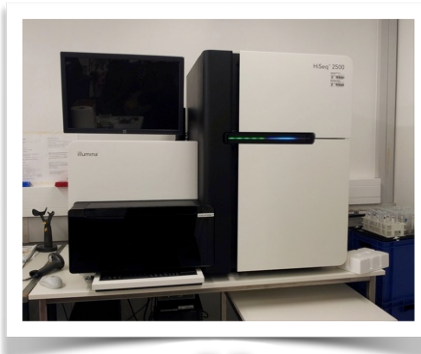
Outline

- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Three Computational Questions



Hundreds of Samples



High-throughput
DNA sequencing

Billions of sequences

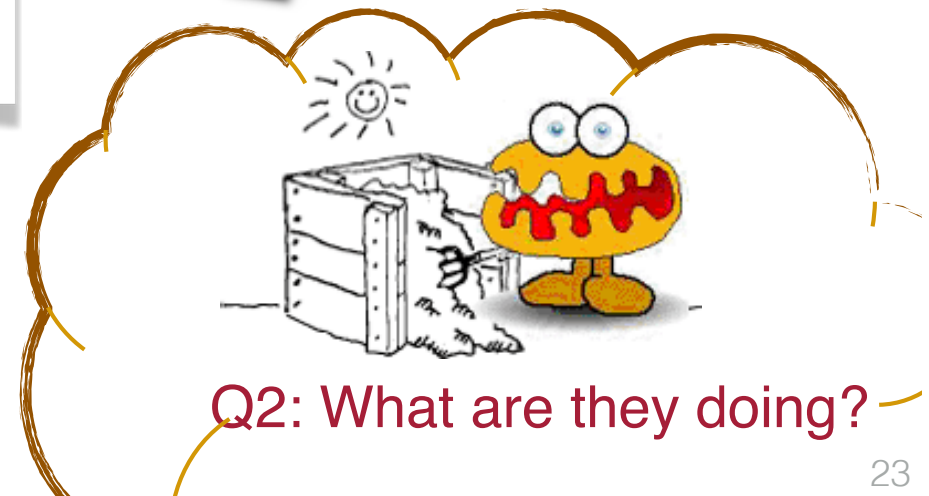


Basic computational
analysis

Many
CPU hours

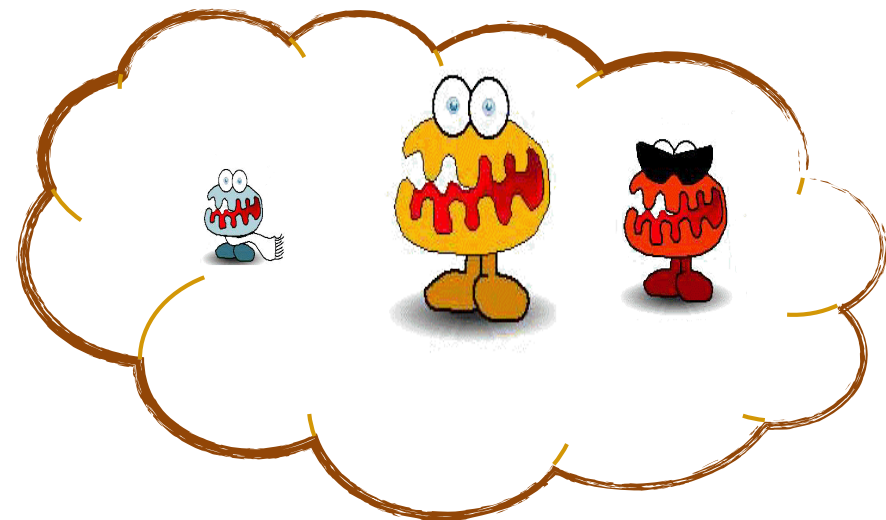


Q1: Who is out there?



Q2: What are they doing?

Q3: How do they compare?

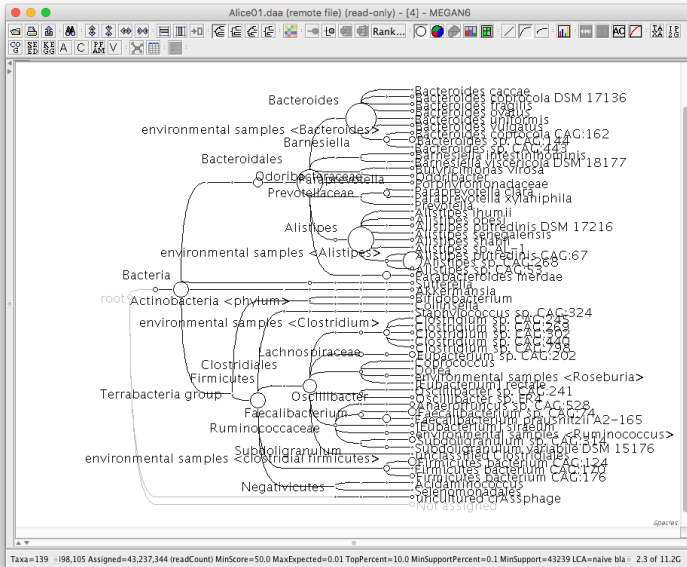


Outline

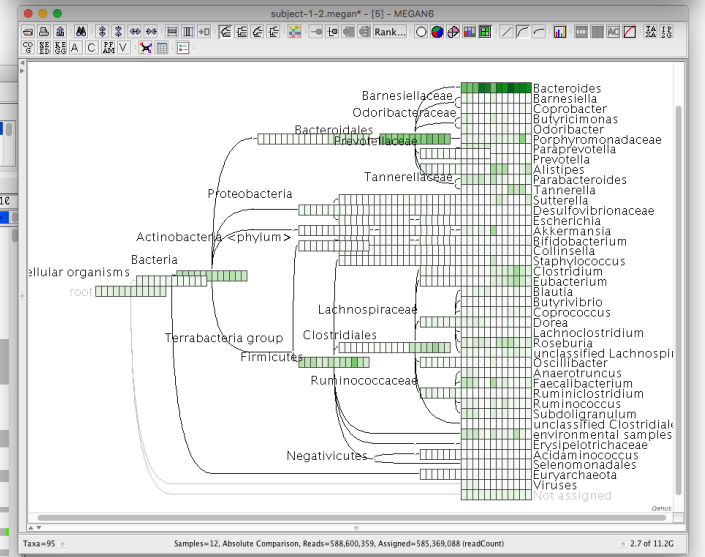
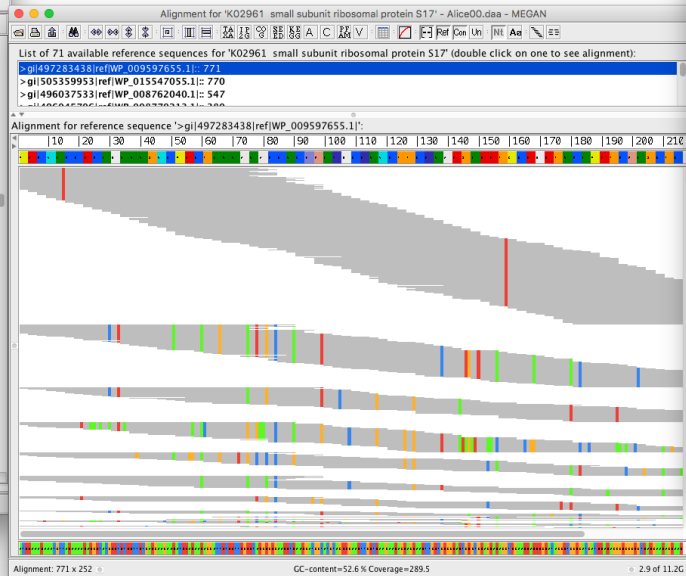
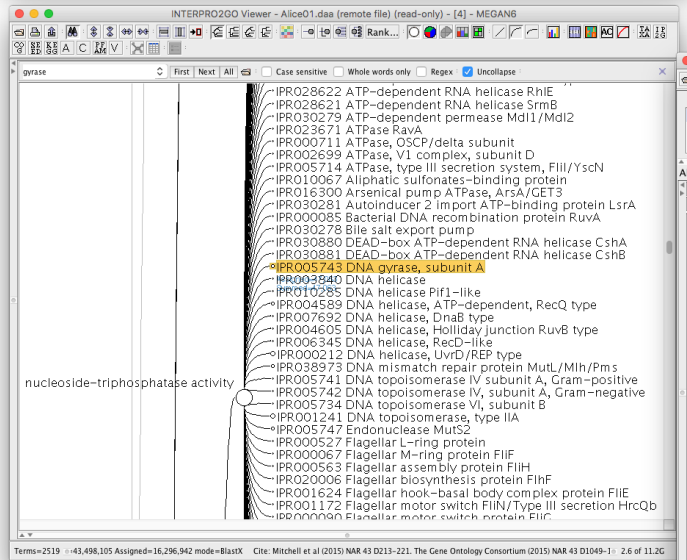
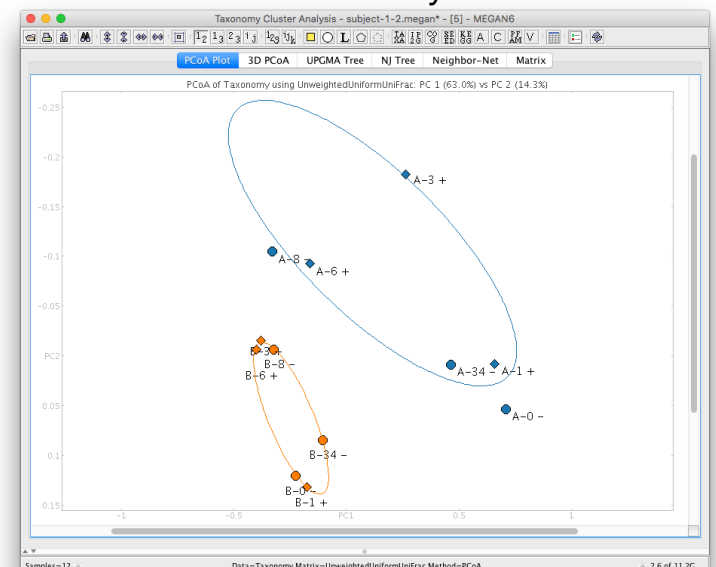
- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- **MEGAN taxonomic and functional binning**
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Interactive MEGAN Analysis

Taxonomic content



PCoA analysis



Functional content

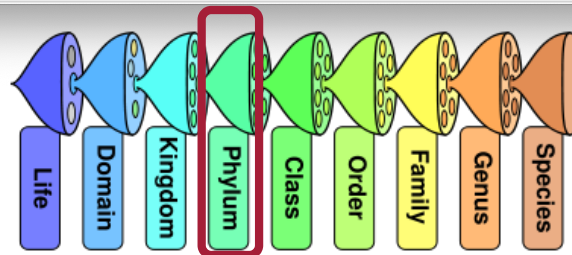
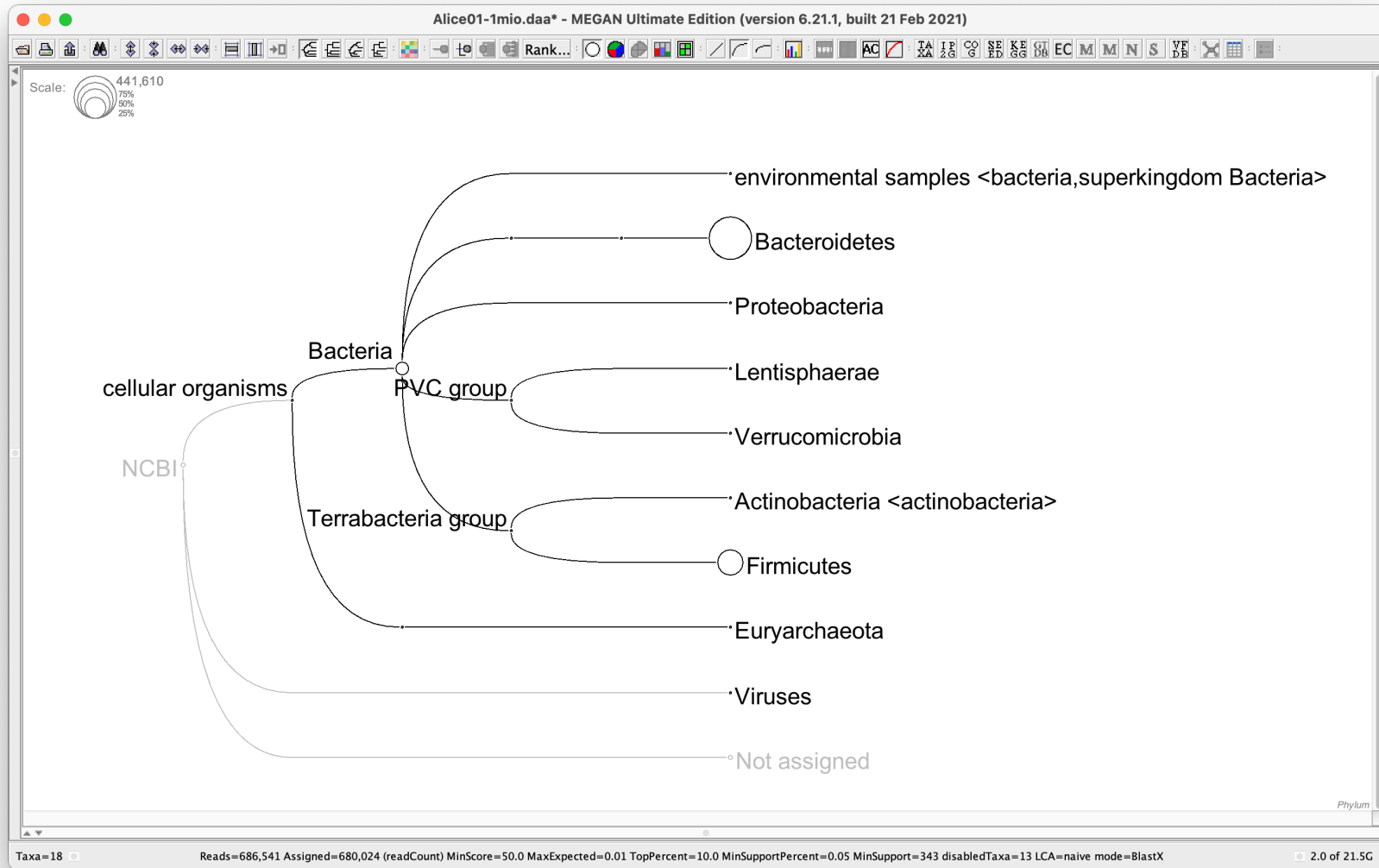
Gene-centric alignment and assembly

Comparative analysis

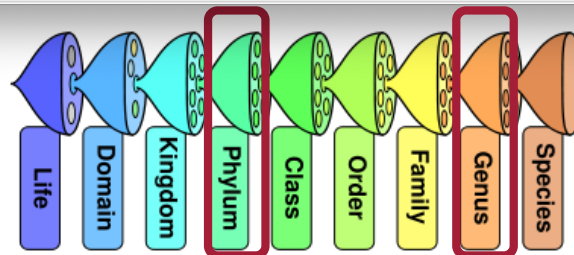
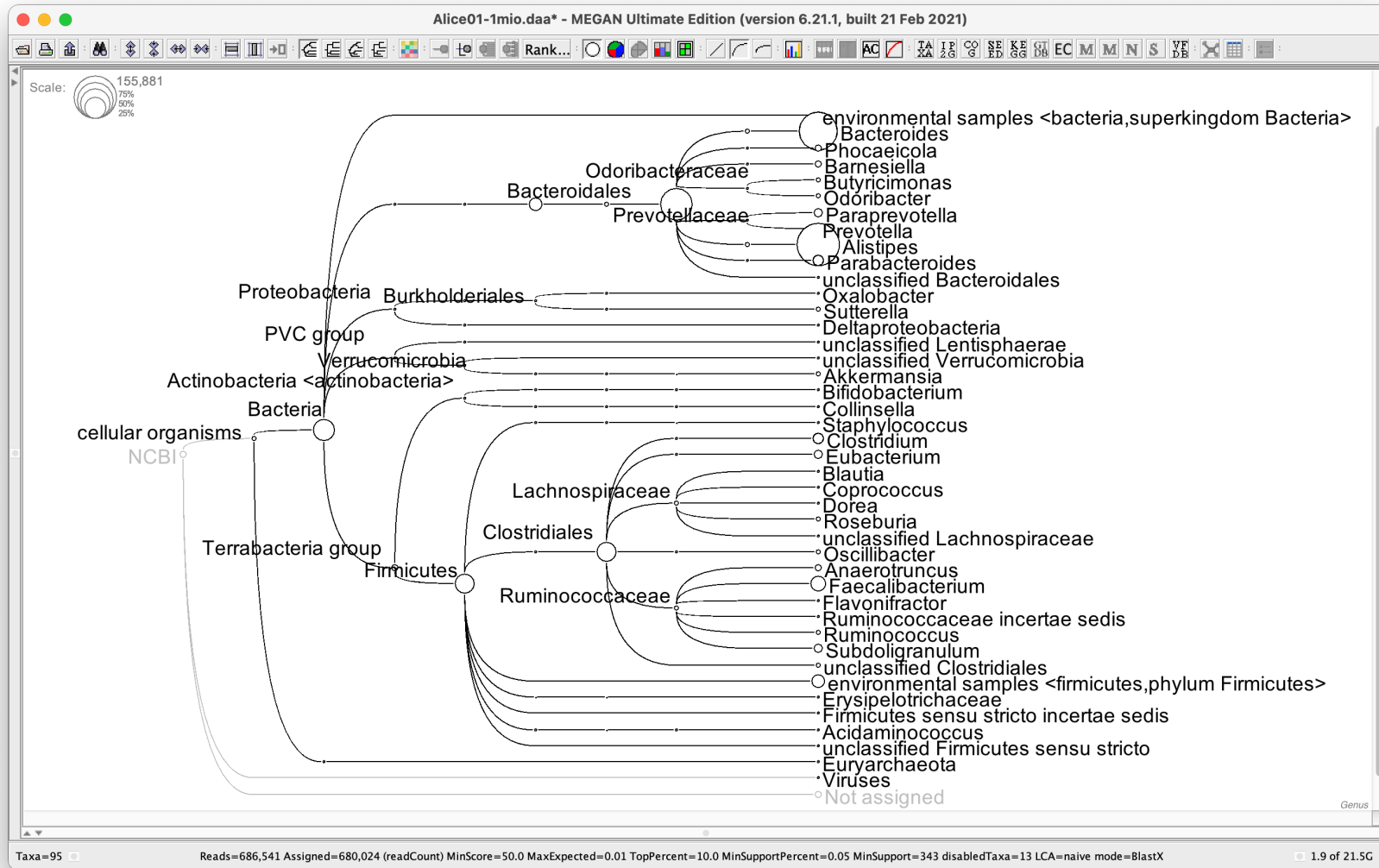


Taxonomic Content

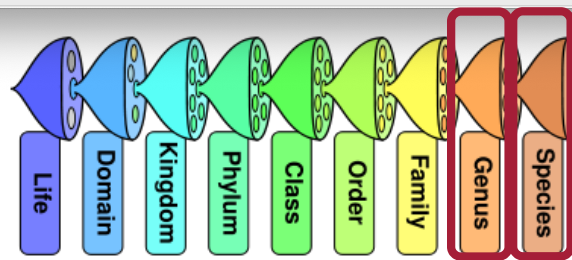
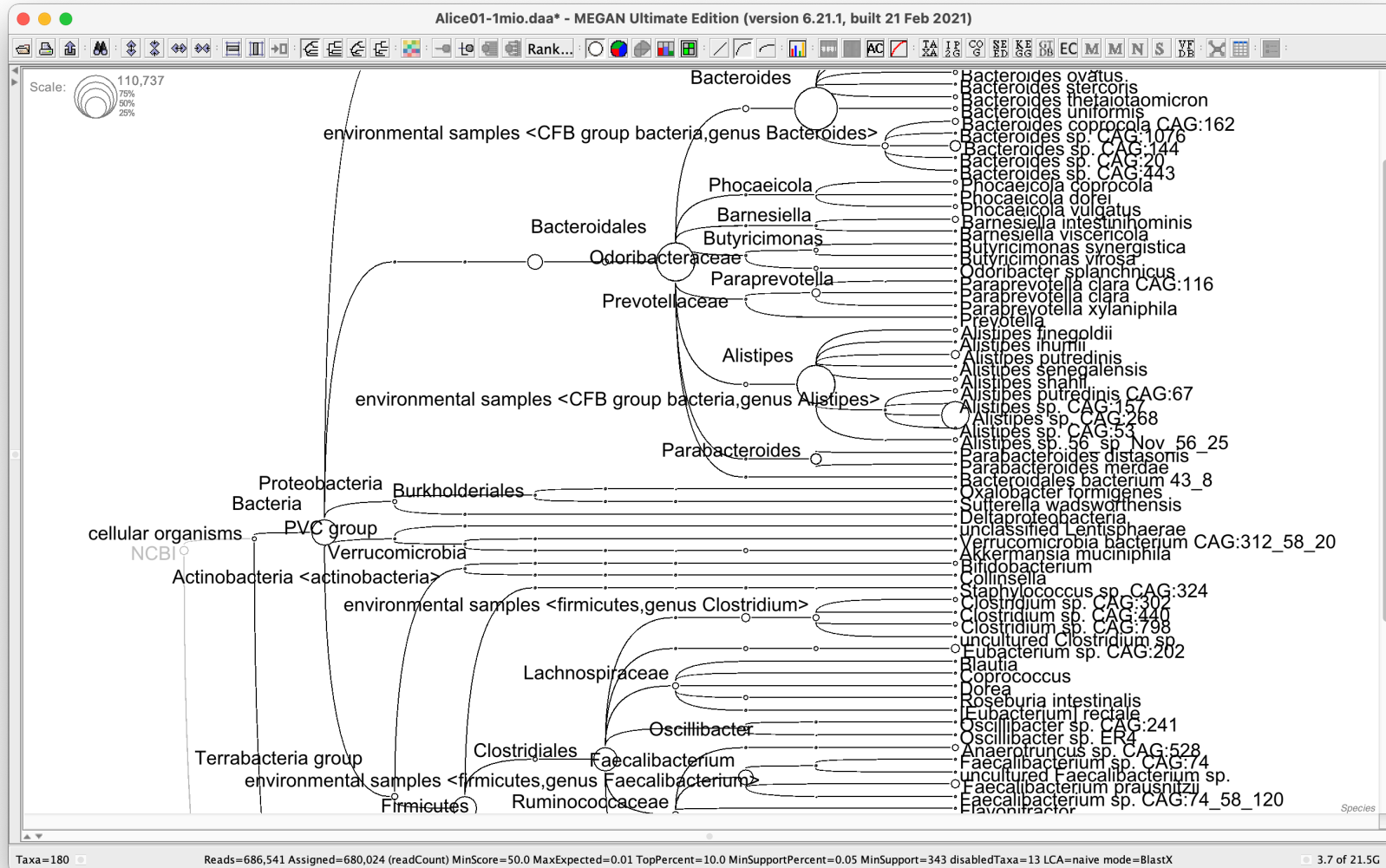
ASARI human gut microbiome



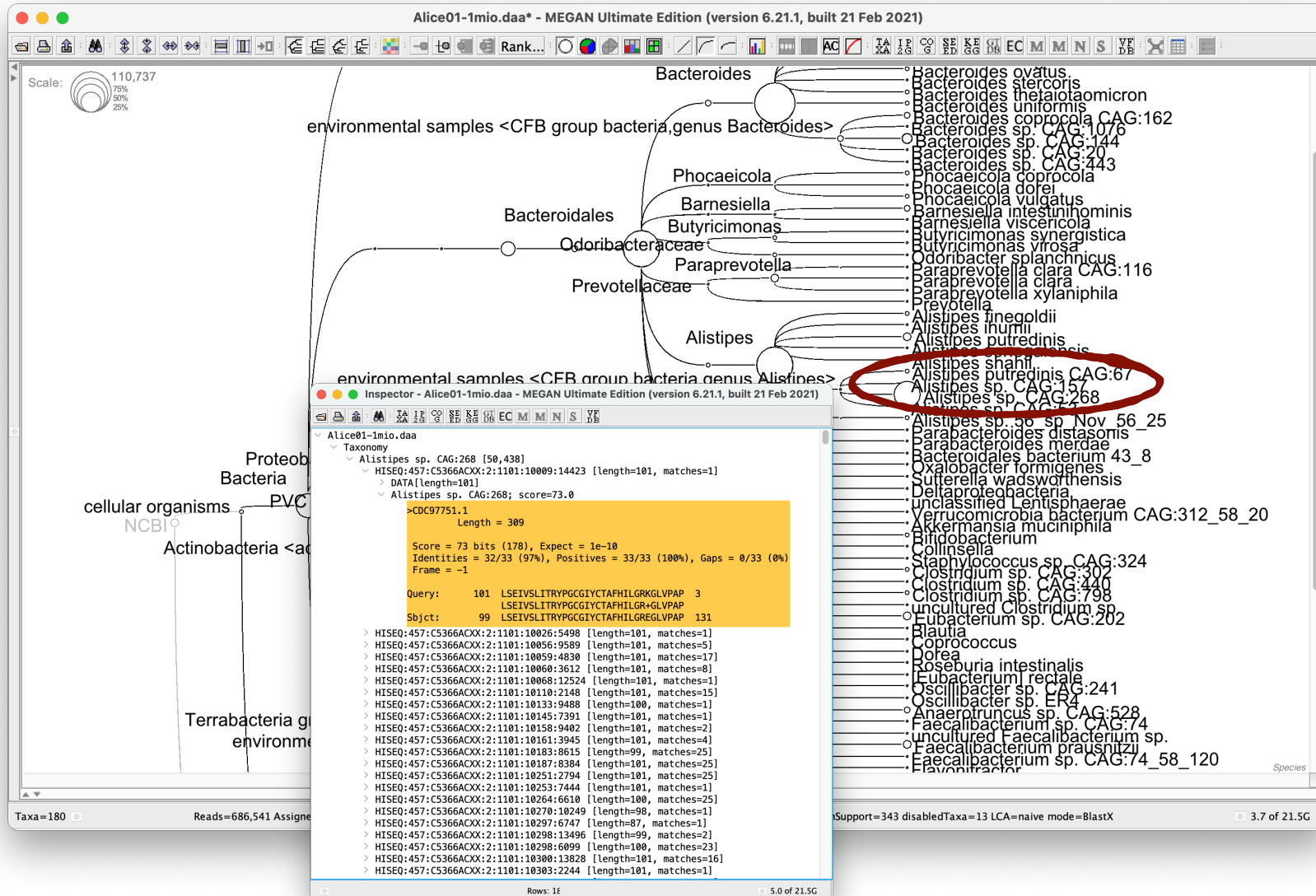
Taxonomic Content



Taxonomic Content



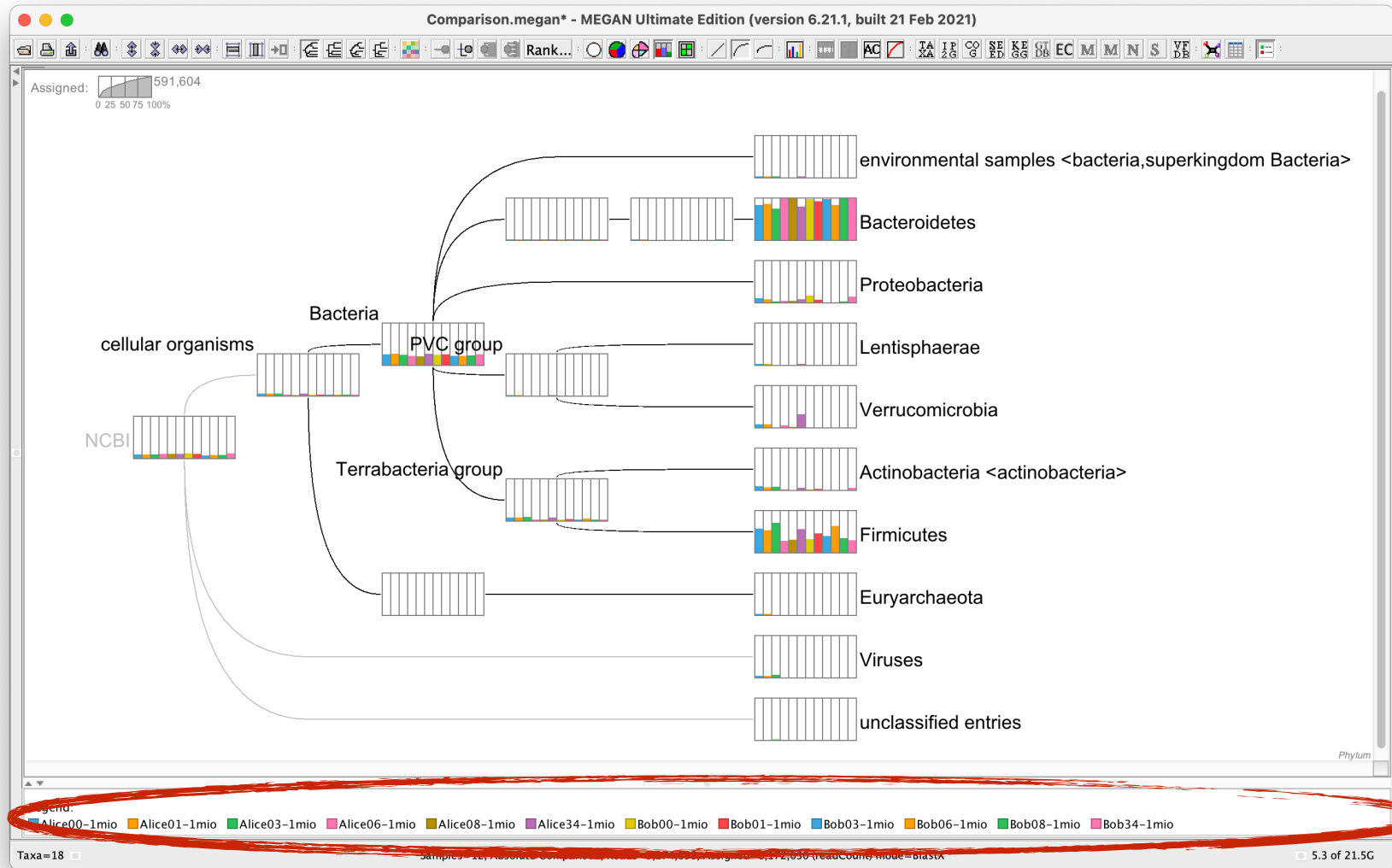
Drill Down to Details...



Comparison

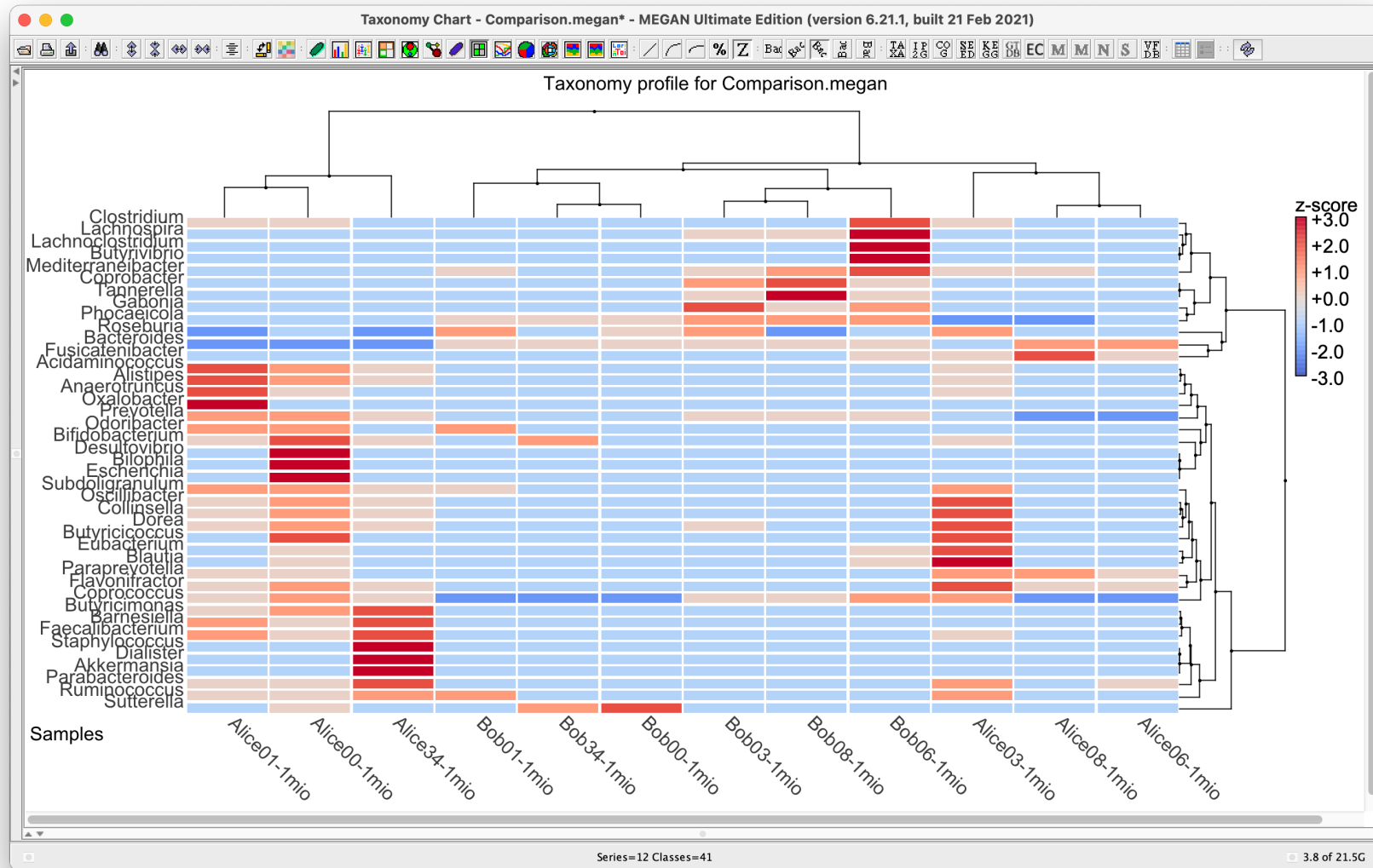


Q3: How do they compare?



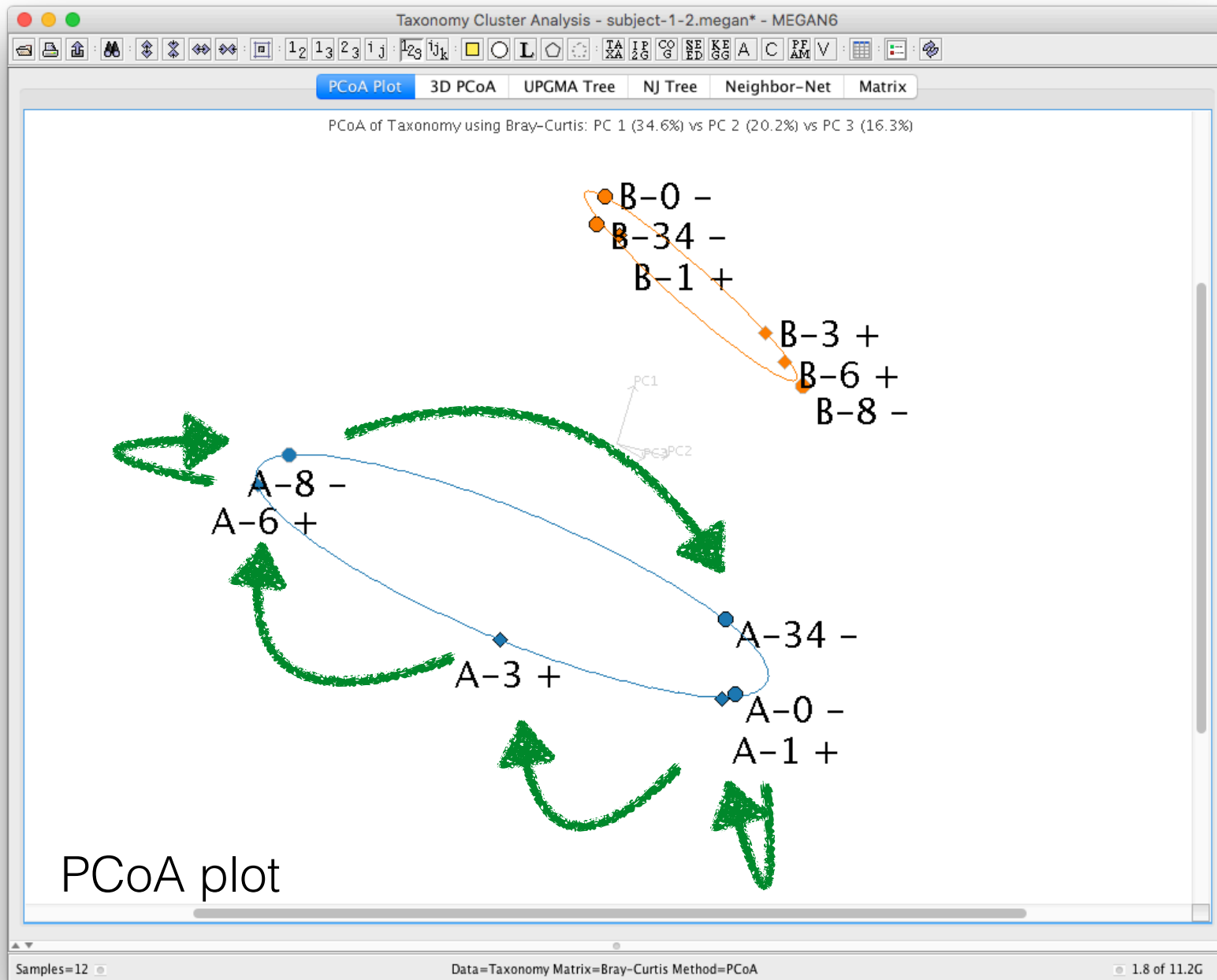
All 12 ASARI human gut samples together

Comparison

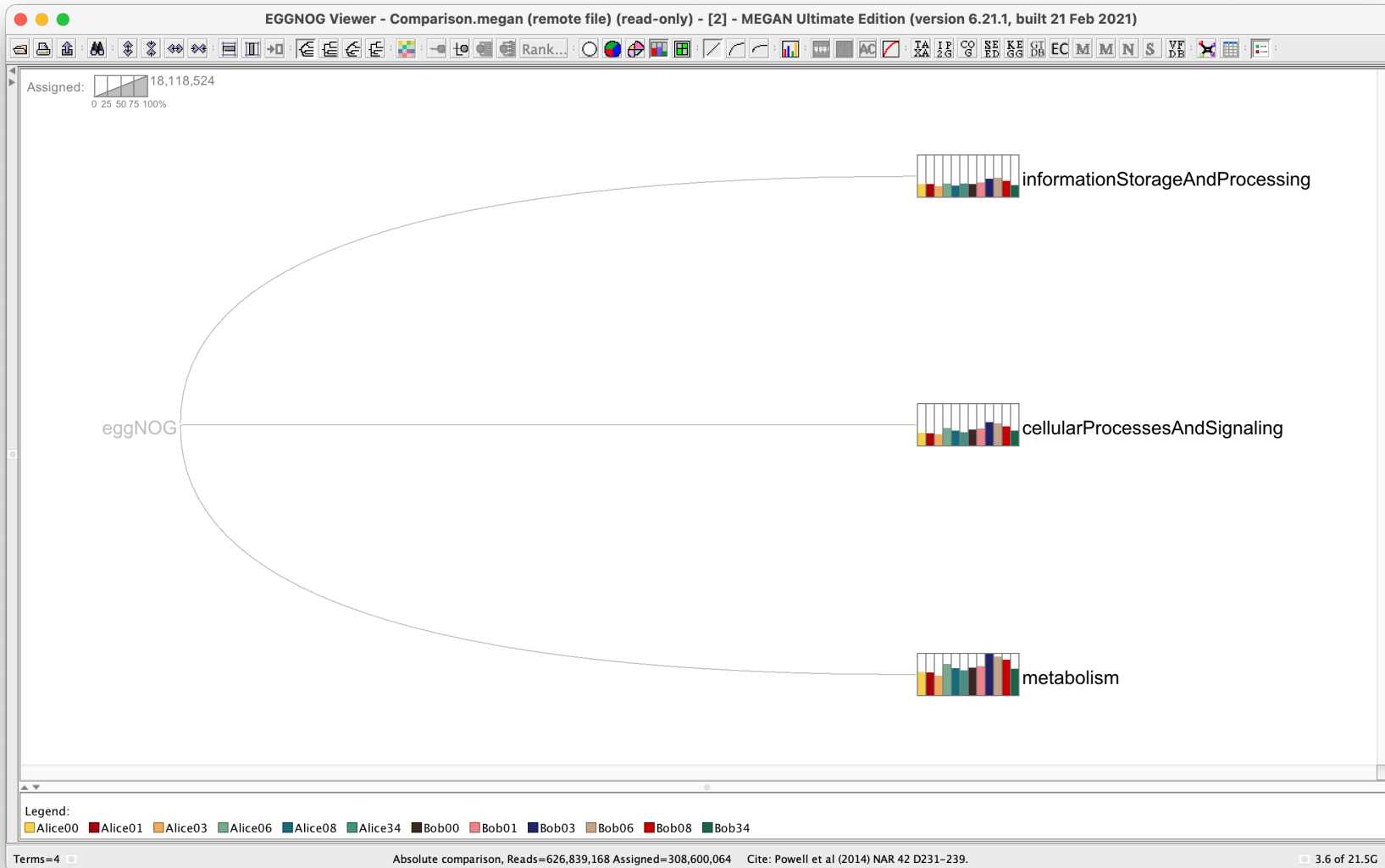


All 12 ASARI human gut samples together

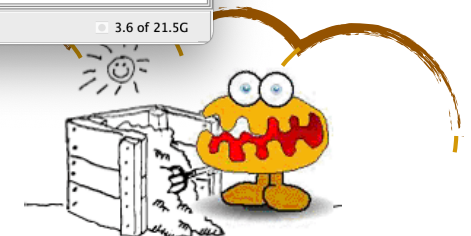
E.g.: Does the Microbiome Rebound?



Functional Content

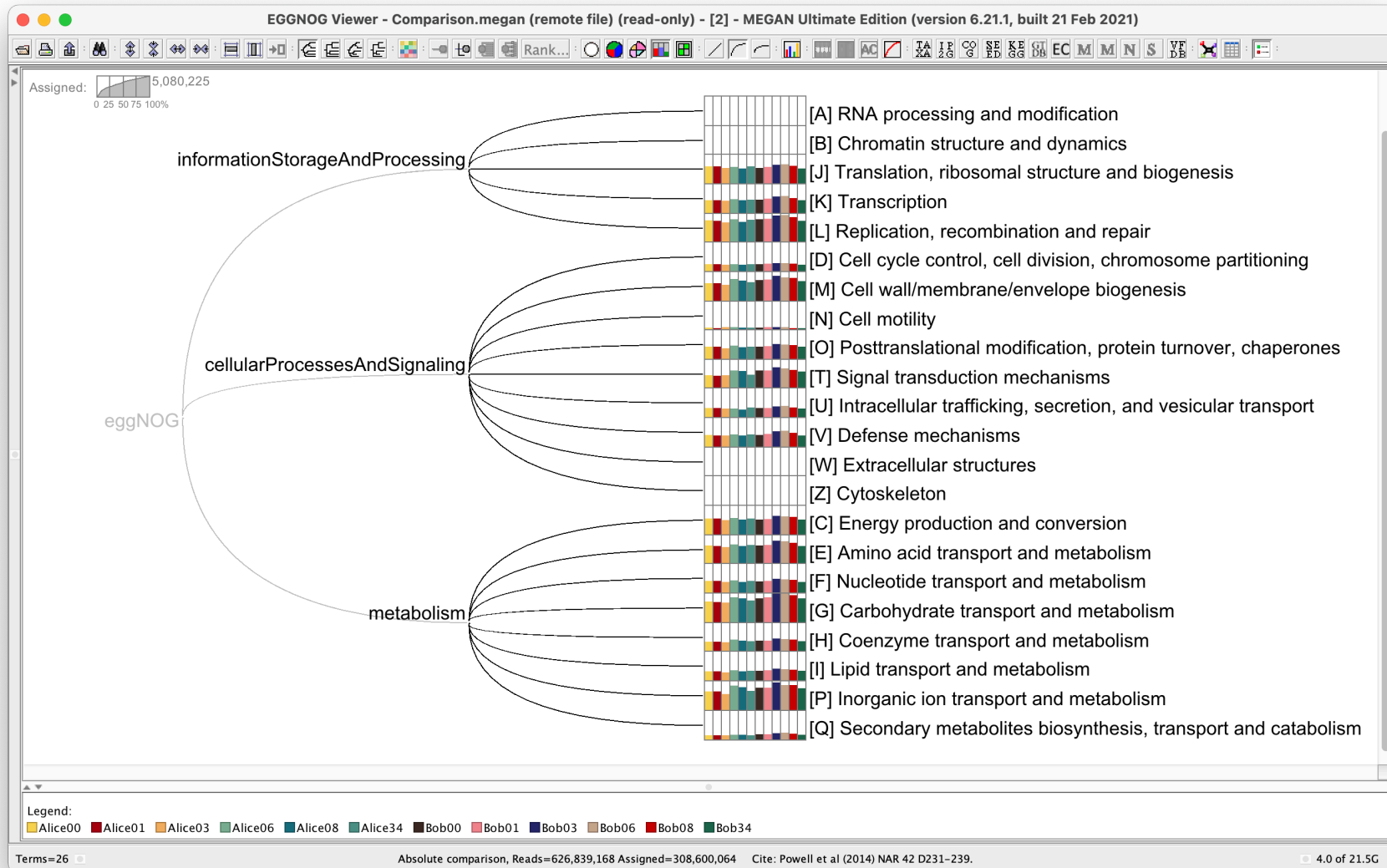


eggNOG classification
(Powell et al, 2014)

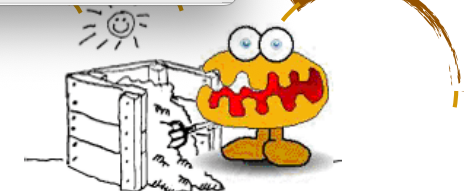


Q2: What are they doing?

Functional Content

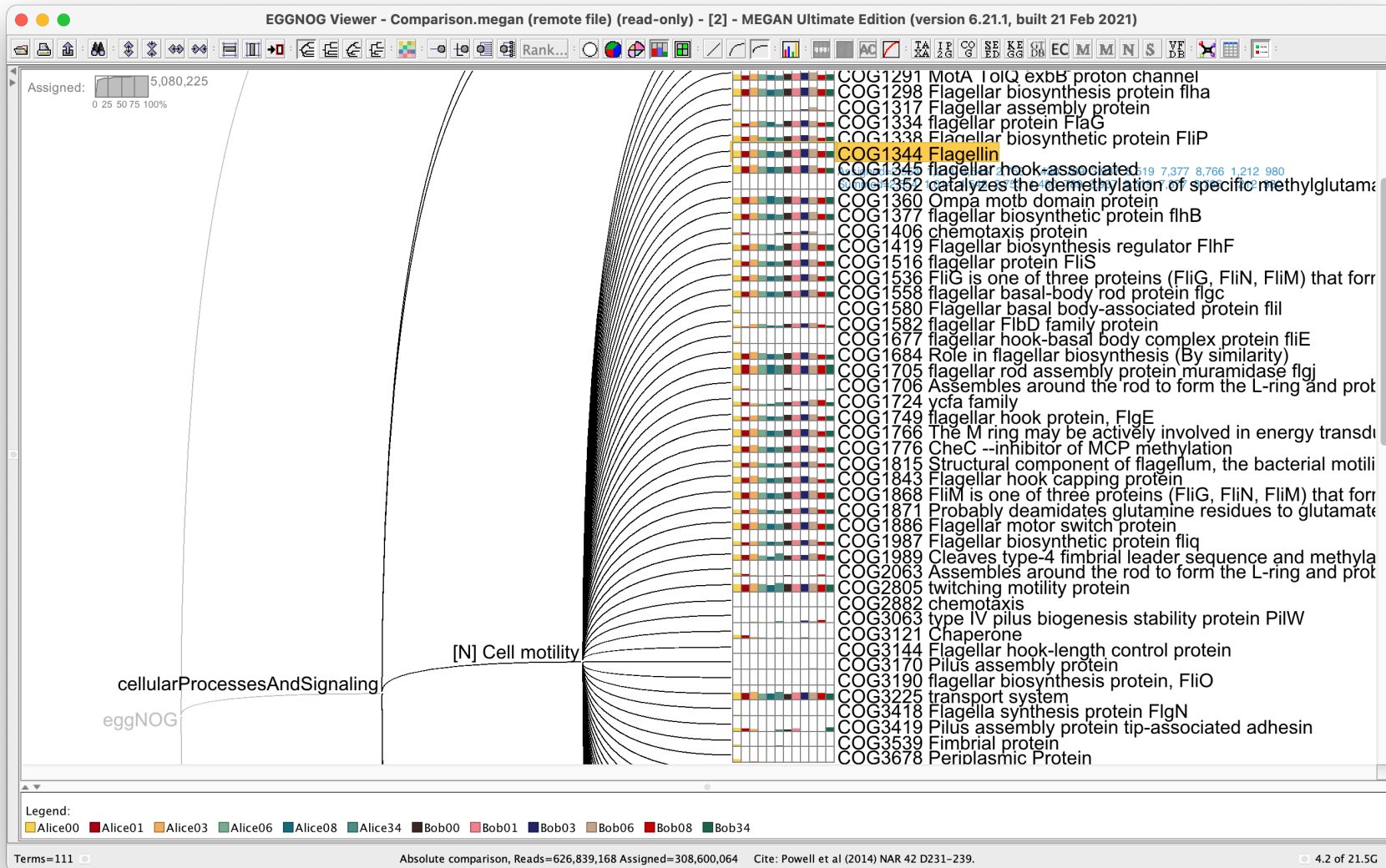


eggNOG classification
(Powell et al, 2014)

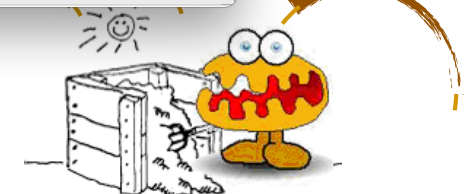


Q2: What are they doing?

Functional Content



eggNOG classification
(Powell et al, 2014)



Q2: What are they doing?

MEGAN Binning

- Taxonomic binning using: NCBI taxonomy or GTDB
- Functional binning using:
 - InterPro families (Mitchell et al, 2015)
 - eggNOG/COG (Powell et al, 2014)
 - SEED (Overbeek et al, 2014)
 - KEGG (license required) (Kanehisa M & Goto S, 2000)
 - EC numbers

Outline

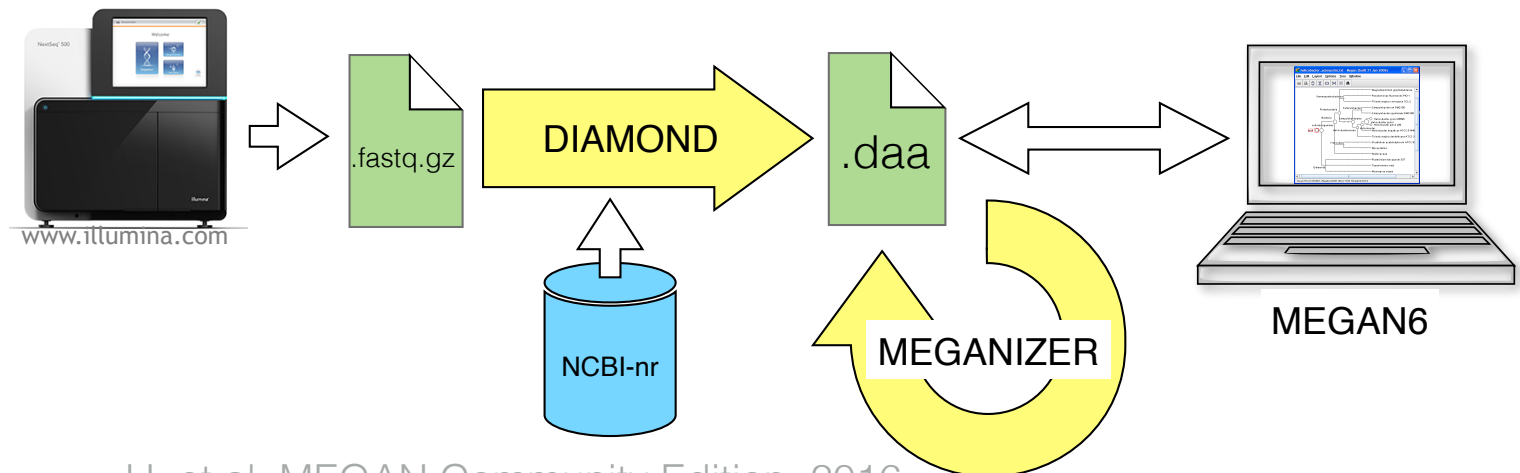
- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

DIAMOND+MEGAN

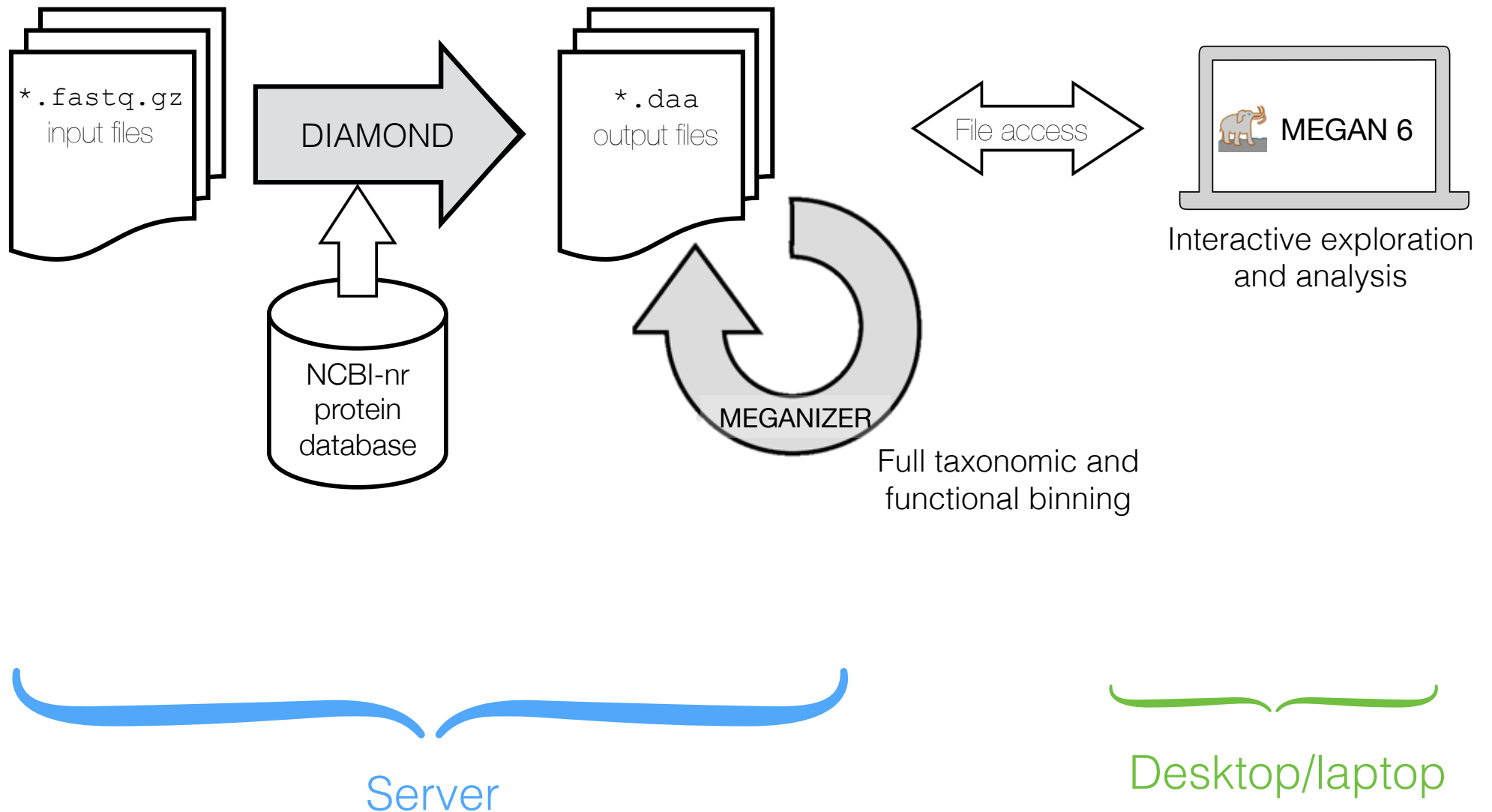
Meganizer program available with MEGAN

- Performs taxonomic and functional binning of reads
- Indexes all data
- Appends results to the DIAMOND output file
- Reduces the total number of files generated in a metagenome analysis to **2**

- Basic Pipeline:



DIAMOND+MEGAN Pipeline



Running DIAMOND

1. Download and install DIAMOND on a server:

```
wget http://github.com/bbuchfink/diamond/releases/  
download/v2.0.9/diamond-linux64.tar.gz
```

```
tar -xzf diamond-linux64.tar.gz
```

```
or: conda install -c bioconda diamond
```

2. Obtain the latest NCBI-nr database:

```
wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz
```

3. Build the DIAMOND index:

```
diamond makedb --in nr.gz -d nr
```

4. Run DIAMOND on a fasta or fastq file:

```
diamond blastx -d nr -q reads.fastq.gz -o reads.daa -f 100
```

Running Meganizer

1. Download MEGAN

- installer `MEGAN_Community_unix_6_21_10.sh` and
- mapping file `megan-map-Jan2021.db.zip` from:

<https://software-ab.informatik.uni-tuebingen.de/download/megan6>

2. Run the installer in console mode:

```
./MEGAN_Community_unix_6_21_10.sh -c
```

3. Unzip the mapping file:

```
unzip megan-map-Jan2021.db.zip
```

4. 6. Run meganizer on each DIAMOND output file:

```
MEGAN/tools/daa-meganizer -i reads.daa -mdb megan-map-Jan2021.db
```


Running MEGAN

1. Download MEGAN installer e.g.
MEGAN_Community_macos_6_21_9.sh from:
<https://software-ab.informatik.uni-tuebingen.de/download/megan6>
2. Double-click to install in interactive mode
3. Download all meganized DAA files
4. Launch MEGAN and then use File→Open

Alternatively, run the Megan-Server program on your server and then access files directly within MEGAN

Outline

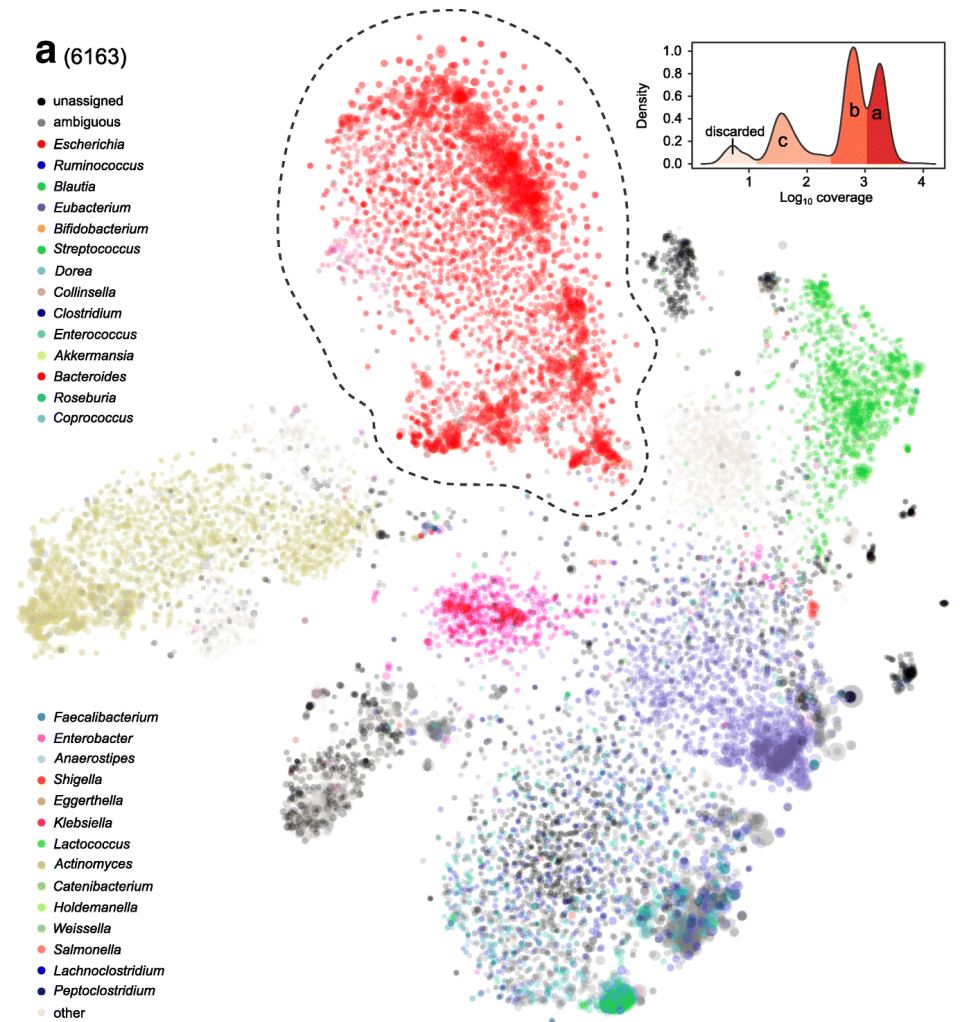
- Introduction to microbiome analysis
- Protein alignment against the NCBI-nr database
- Who is out there, what are they doing, how do they compare?
- MEGAN taxonomic and functional binning
- The DIAMOND+MEGAN pipeline
- Long-read metagenomics

Microbiome Read-Length Paradox

- Short reads are short and plentiful...
 - **So:** short read microbiome datasets should benefit from assembly
 - **But:** the resulting sequences are usually disappointingly short...
 - Usually far from chromosomal length....
- Long reads are long...
 - **So:** usually longer than average assembled short reads
 - **But:** assembly results in *very* long sequences
 - Complete chromosomes can be obtained...
- Assembly of short reads is optional, but long reads should always be assembled...

Limitation of Short-Read Metagenomics

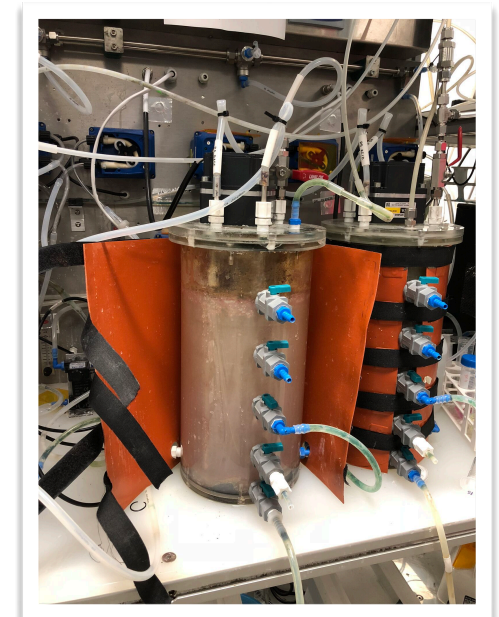
- *Assembly* of metagenomic short reads produces large numbers of tiny contigs - never complete chromosomes



Long-Read Metagenomics

- EBPR waste-water bio-reactor
- MinION sequencing 2018
 - Reads: ~695,000 (~ 6 Gb)
 - Length: ~9 kb mean (2 bp - 66 kb)
 - Short Read Archive SRX5120474

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](#) ✉

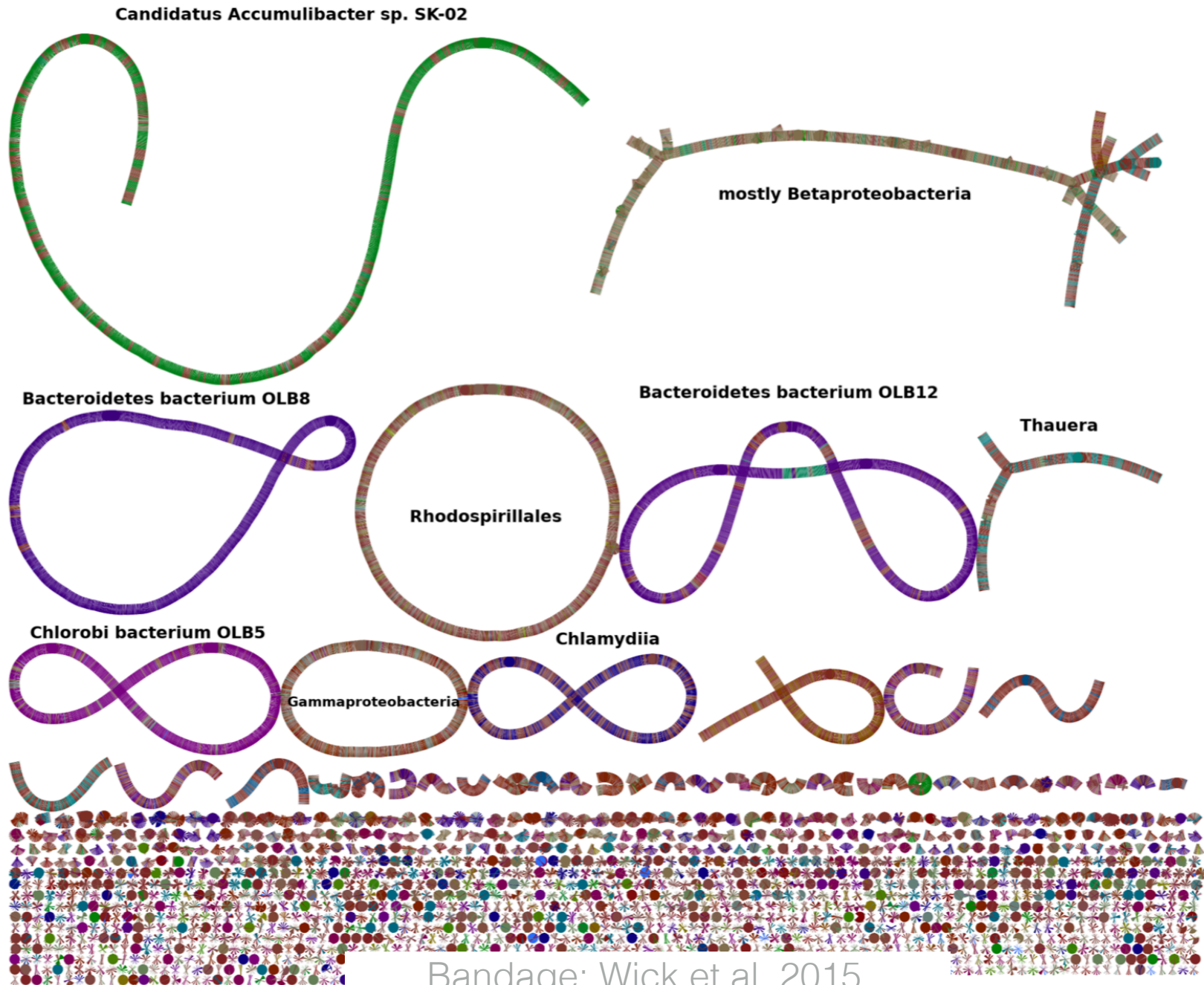
[Microbiome](#) 7, Article number: 61 (2019) | [Cite this article](#)



Long-Read Metagenome Assembly

- Input:
 - Reads: ~695,000 (~ 6 Gb)
 - Length: ~9 kb mean (2 bp - 66 kb)
- Assembly using Unicycler (miniasm and racon)
(Li 2016, Vaser *et al* 2017, Wick *et al*, 2017)
- Output:
 - Contigs: ~1,700 (~ 104 Mb)
 - Length: ~ 61 kb mean (1.3 kb - 5.2 Mb)

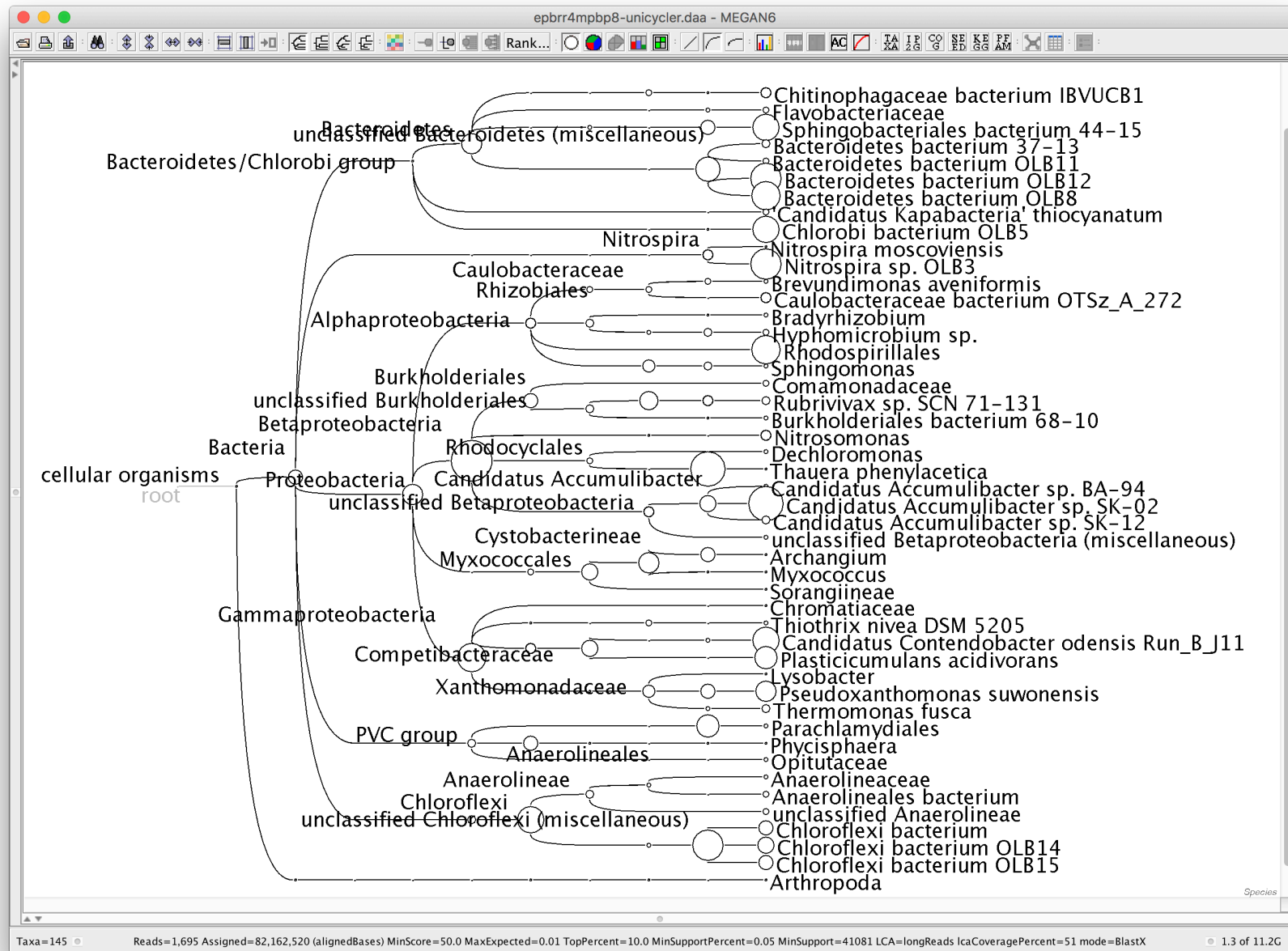
Bandage Visualization of Assembly Graph



Bandage: Wick et al, 2015

Layout: Hachul S., Jünger M., 2007

Taxonomic Binning of Contigs



DIAMOND+MEGAN

Taxonomic Bins $\geq 50\%$ Complete

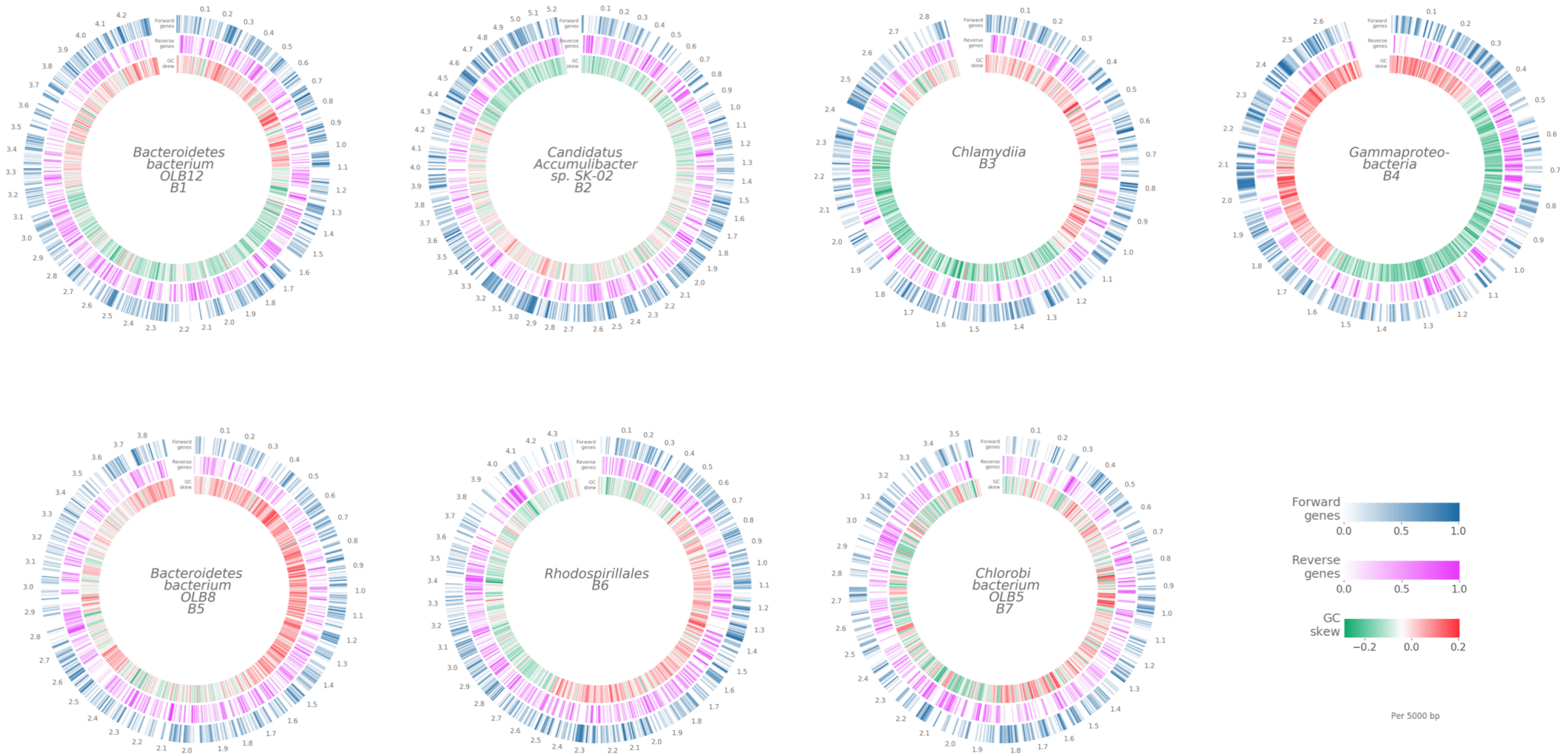
DIAMOND+MEGAN taxonomic bin		Unicycler contigs	Total (Mb)	Aligned (Mb)	Average coverage	CheckM		rRNA	Prokka tRNA	CDS
						Complete.	Contam.			
High quality draft genomes:										
B1	<i>Bacteroidetes bacterium</i> OLB12	1	4.2	3.5	57.3	95%	0.1%	6	39	4,163
B2	<i>Candidatus Accumulibacter</i> SK-02	1	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

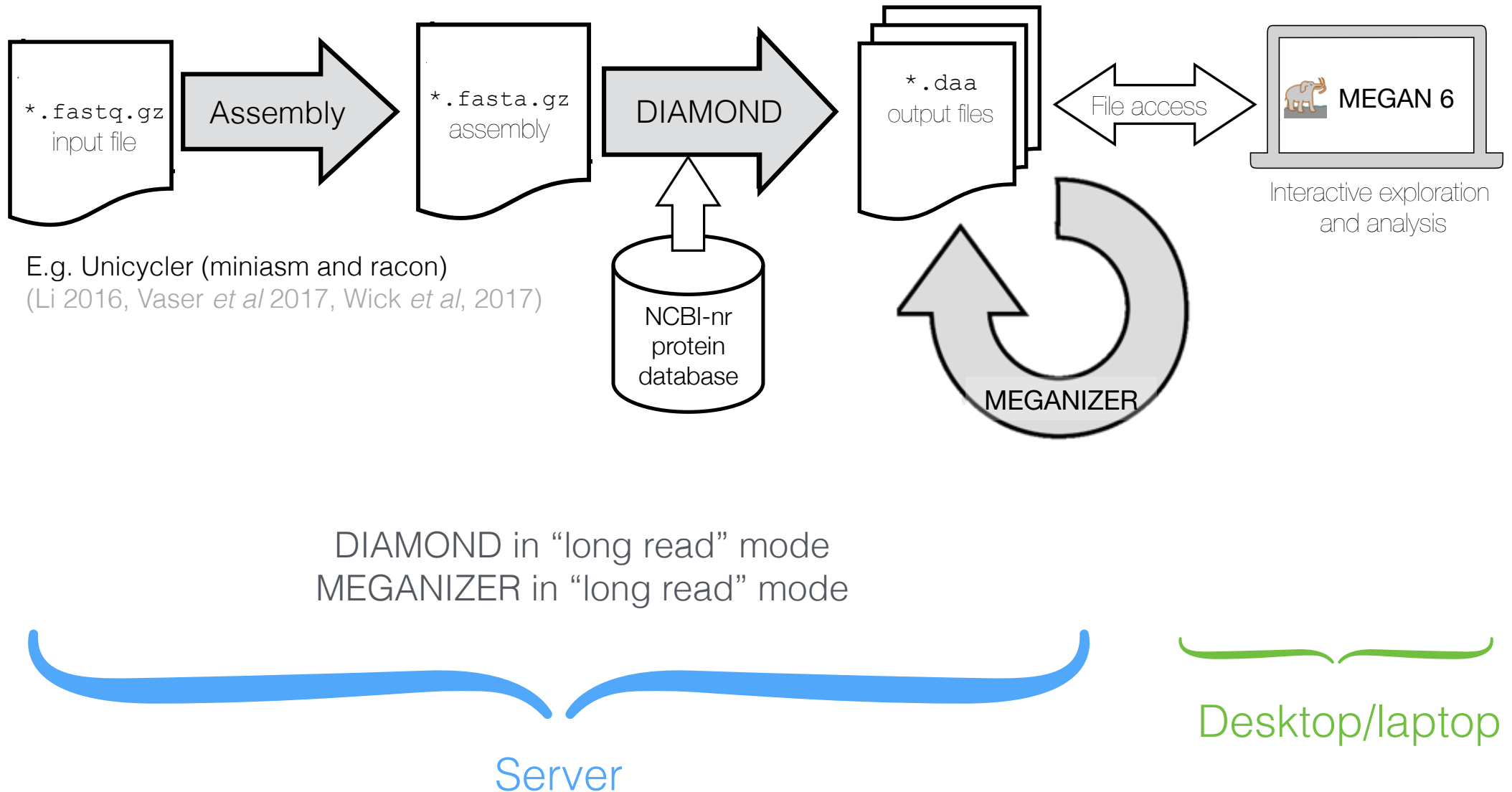
CheckM (Parks et al. 2014)

Prokka (Seemann, 2014)

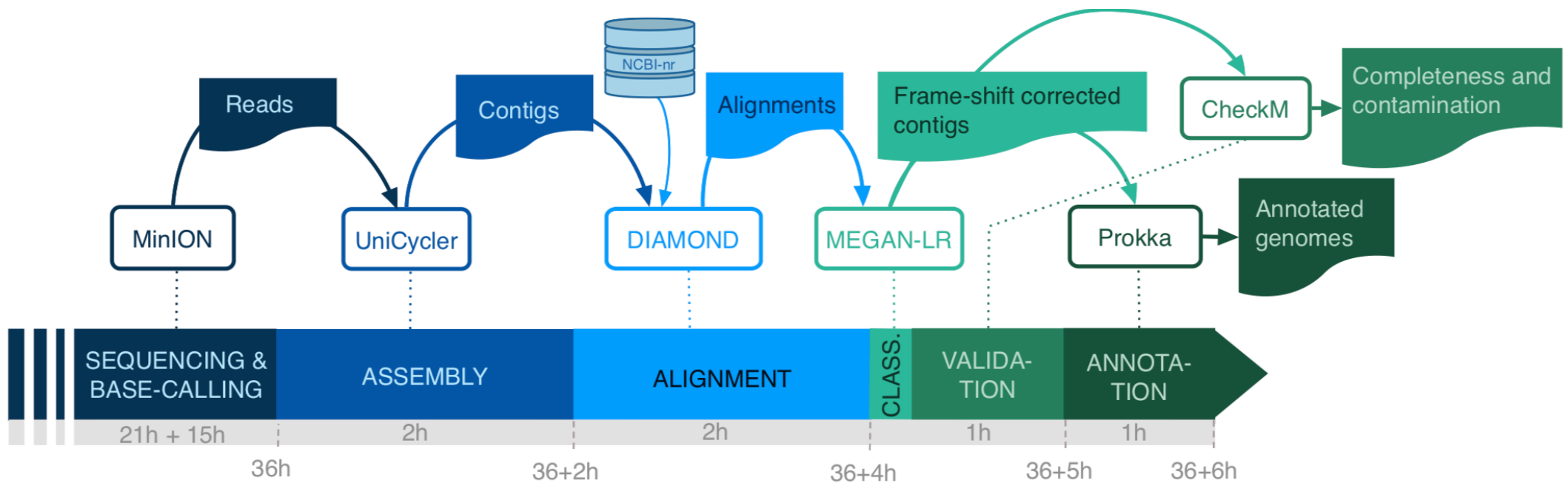
Assembled Chromosomes



Long-Read Analysis Pipeline



Long-Read Analysis Pipeline



Running Assembly

There is much active research into long-read assembly. Unicycler is one of many tools.

- Install the *Unicycler* assembler as follows:

```
git clone https://github.com/rrwick/Unicycler.git
```

```
cd Unicycler
```

```
make
```

- Or: `conda install -c bioconda unicycler`

(Li 2016, Vaser *et al* 2017, Wick *et al*, 2017)

Running Assembly

- Run Unicycler as follows:

```
unicycler -l reads.fq.gz -o reads_asm --keep 3 -t 16
```

- Option `-l reads.fq.gz` to specify file of long reads
- Option `-o reads_asm` to specify output directory
- Option `--keep 3` to keep intermediate files
- Option `-t 16` to specify the number of threads
- Output: `reads_asm/assembly.fasta`

(Li 2016, Vaser *et al* 2017, Wick *et al*, 2017)

Running DIAMOND on Assemblies

- Run as for short reads, but with additional options:

```
diamond blastx -d nr -q assembly.fasta -o assembly.daa  
-f 100 -F 15 --range-culling --top 10
```

- Option **-F 15** to activate frame-shift alignment
- Options **--range-culling --top 10** to ensure that alignments along the whole sequence are reported

Running Meganizer on Assemblies

- Run as for short reads, but with an additional option:

```
MEGAN/tools/daa-meganizer -i assembly.daa  
-mdb megan-map-Jan2021.db -1
```

- Option **-1** to specify long-read mode

Detailed Protocols



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>



Hands-on Session

- <https://software-ab.informatik.uni-tuebingen.de/download/public/tutorial-aug2021/welcome.html>

DIAMOND+MEGAN Tutorial ISME 2-August-2021 Download Page

1. Presentation:

- Video: [Introduction-to-DIAMOND+MEGAN-August2021.mp4](#)
- Slides: [Introduction-to-DIAMOND+MEGAN-August2021.pdf](#)
- Both: [04-Presentation.zip](#)

2. Tutorial outline:

- [DIAMOND+MEGAN-Tutorial-August-2021.pdf](#).

3. MEGAN program installers:

- Mac OS: [MEGAN_Community_macos_6_21_10.dmg](#).
- Linux: [MEGAN_Community_unix_6_21_10.sh](#).
- Windows: [MEGAN_Community_windows-x64_6_21_10.exe](#).
- Other: [MEGAN download page](#)

4. Short read datasets:

- One million reads each: [01-Short-Read-Data-1mio.zip](#) (2.5 GB)
- Summary only: [01-Short-Read-Data-summary.zip](#) (1 MB)

5. Long-read dataset:

- Full dataset: [02-Long-Read-Data-full.zip](#) (234 MB)
- Summary only: [02-Long-Read-Data-summary.zip](#) (0.1 MB)

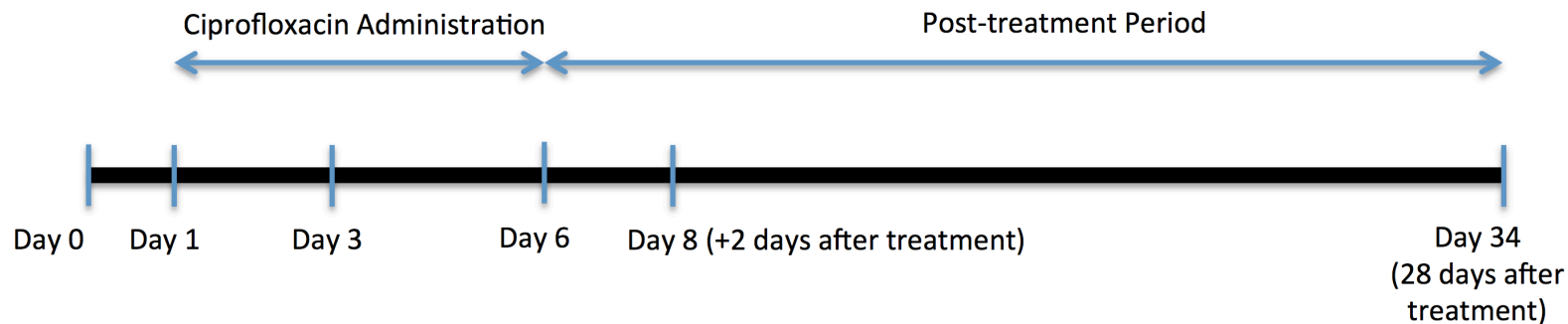
6. Papers:

- [03-Papers.zip](#).



Short-Read Data

Alice and Bob, 6 time points each



- Each subsampled to 1 mio reads:
 - [01-Short-Read-Data-1mio.zip](#) (2.5 GB)
- Summary only:
 - [01-Short-Read-Data-summary.zip](#) (1 MB)

Long-Read Data

- Nanopore reads from enrichment reactor:

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

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

[Krithika Arumugam](#), [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](#)



Krithika Arumugam

- Full dataset:
 - [02-Long-Read-Data-full.zip](#) (234 MB)
- Summary only:
 - [02-Long-Read-Data-summary.zip](#) (0.1 MB)

Thank You!

Joint work with:

- Benjamin Albrecht, Caner Bagci, Xi Chen, Timo Lucas, Sascha Patz
& Lars Angenent Tübingen
- Irina Bessarab, Krithika Arumugam and Rohan Williams SCELSSE Singapore

Funding:

- Deutsche Forschungsgemeinschaft (MAIRA & BinAC)
- Life Sciences Institute at NUS
- NRF/MOE and NRF-EW, Singapore