Complete and validated genomes from metagenomes

Daniel J. Giguere[1], Alexander T. Bahcheli[1], Benjamin R. Joris[1],
Julie M. Paulssen[2], Lisa M. Gieg[2], Martin W. Flatley[3], Gregory B. Gloor[1]

1 Department of Biochemistry, Western University, London, Canada
2 Department of Biological Sciences, University of Calgary, Calgary, Canada
3 Senior Staff Engineer, Suncor Energy, Sarnia, Canada

Introduction

• Metagenome assembly typically results in fragmented, **incomplete assemblies**
• **Long read technology** enables resolving repetitive regions
• Standards for a complete genome are not achieved for most metagenomic-assembled genomes (MAGs)
• We investigated an algal-bacterial community that is known to **degrade 1-adamantanecarboxylic acid**[1]

Methods

1. **High molecular weight** DNA extraction
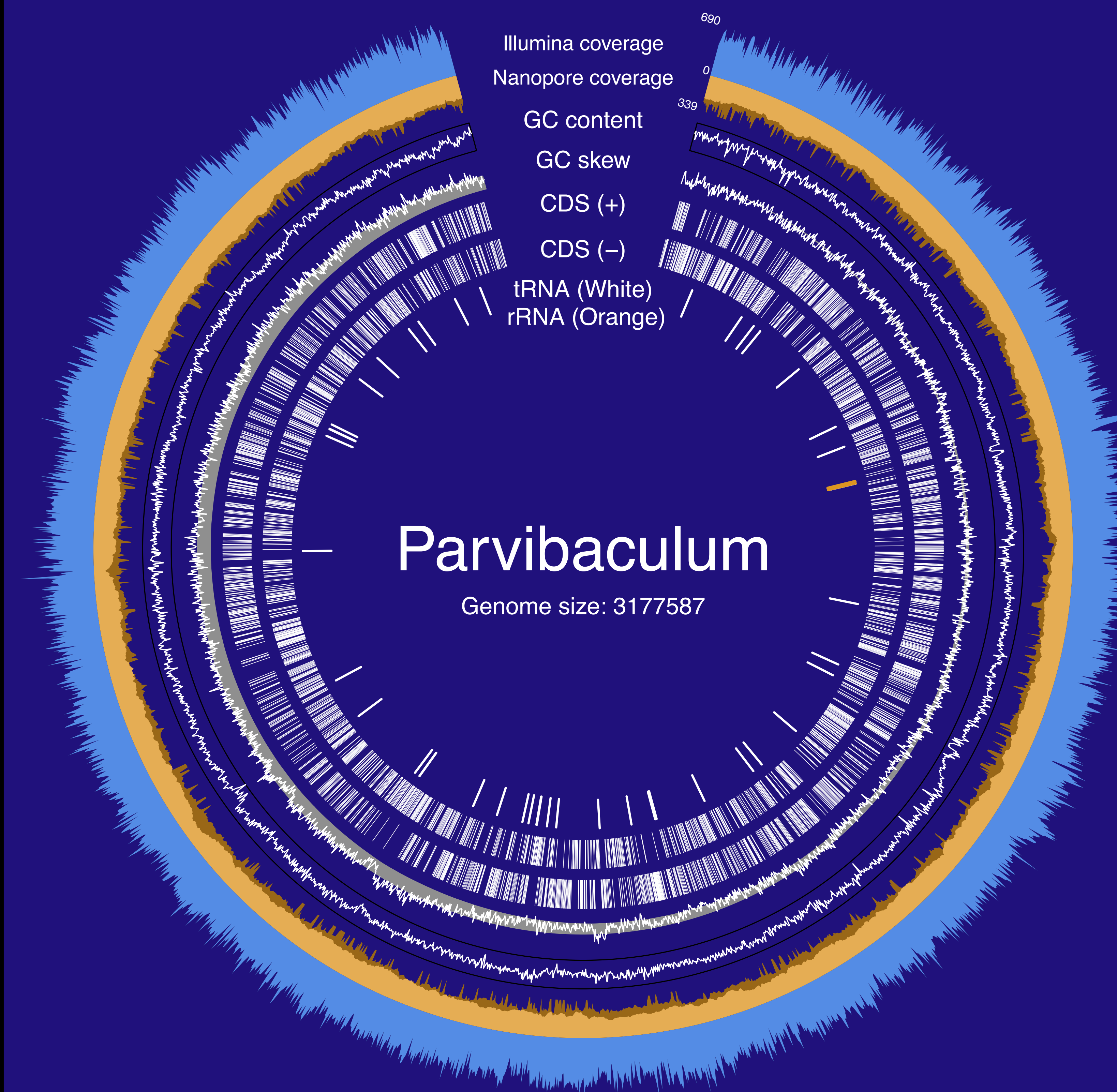2. Oxford Nanopore minION and Illumina NextSeq

Long read assembly

Read N50 > 25 kb

Filter reads and correct contigs
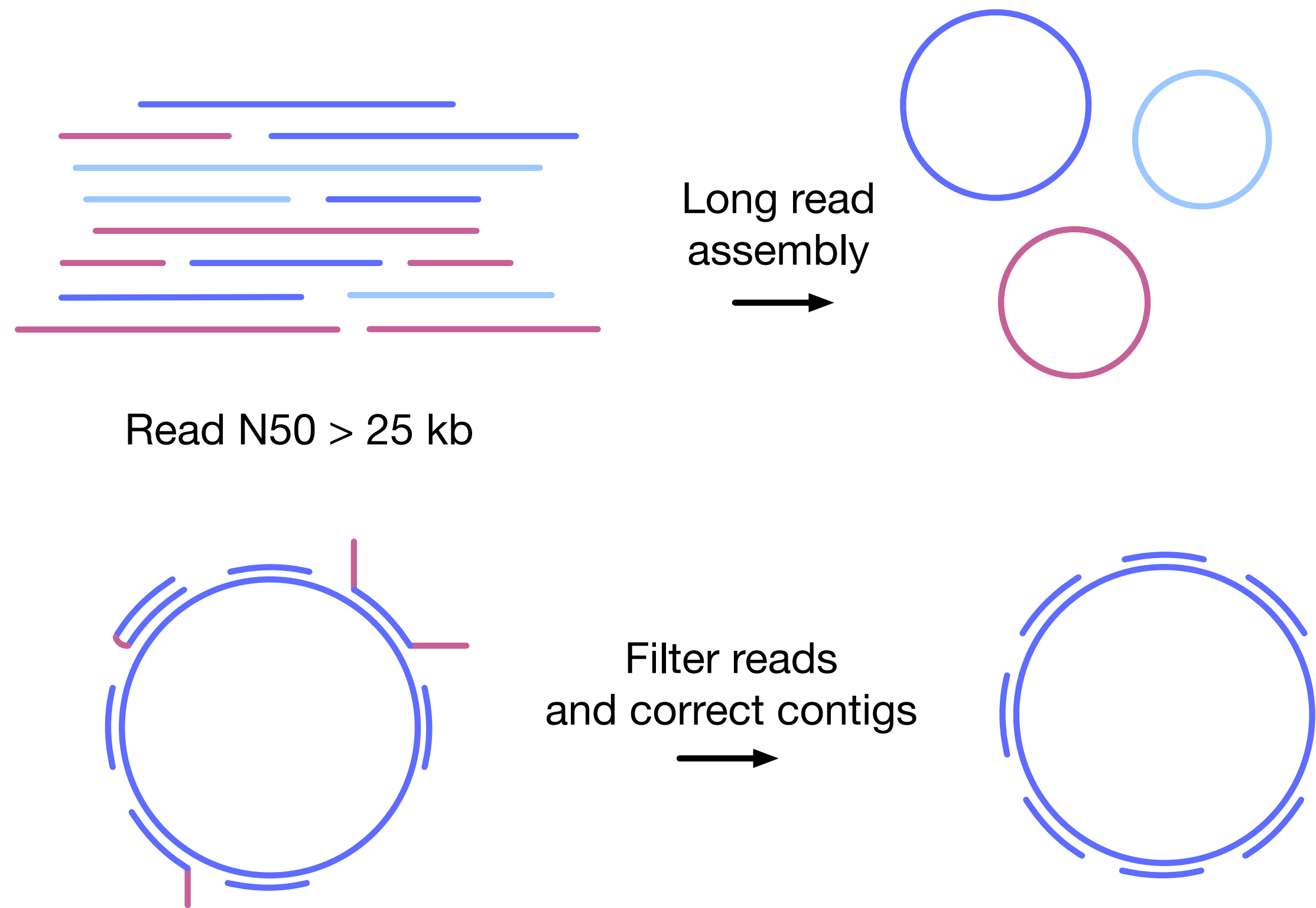
Results

• **13 MAGs** representing more than 60% of total DNA sequences[2]

Length	%GC	%C	%R	Illumina coverage	Nanopore coverage	rRNA genes	tRNA genes	Predicted taxonomy
3.14	62.8	98.59	0.00	392	193	2	48	Parvibaculum
4.47	64.4	100.00	2.82	402	94	2	49	Rhizobiaceae
3.73	63.9	100.00	0.00	68	32	2	46	Unknown
3.84	63.5	98.59	1.41	184	67	4	45	Blastomonas
3.74	72.4	100.00	0.00	121	39	4	63	UBA2363
5.16	42.6	98.59	0.00	180	97	6	40	Algoriphagus
3.98	66.9	100.00	1.41	41	16	4	51	Tabrizicola
3.96	71.9	98.59	2.82	125	57	2	53	UBA4742
4.68	66.1	100.00	2.82	57	17	4	52	Rhodobacteraceae
5.79	66.3	100.00	2.82	35	14	4	55	Aquimonas
0.79	51.0	85.92	4.23	88	36	2	40	UBA1547
3.06	65.9	100.00	0.00	1107	355	2	48	Brevundimonas
2.88	64.4	100.00	0.00	94	35	2	44	Oceanicaulis

We generated complete genomes directly from bacterial communities

Illumina coverage
Nanopore coverage
GC content
GC skew
CDS (+)
CDS (−)
tRNA (White)
rRNA (Orange)

690
0
339

Parvibaculum

Genome size: 3177587

Discussion

• **Species level variants** with possible mobile genetic elements within a single species can be detected

Overlapping reads
17
0
Coverage
400
200
2.97 mb   3.03 mb   3.09 mb

Add 35 kb insertion found by manual inspection

Overlapping reads
25
0
Coverage
400
200
0
2.97 mb   3.03 mb   3.09 mb
Chromosome position
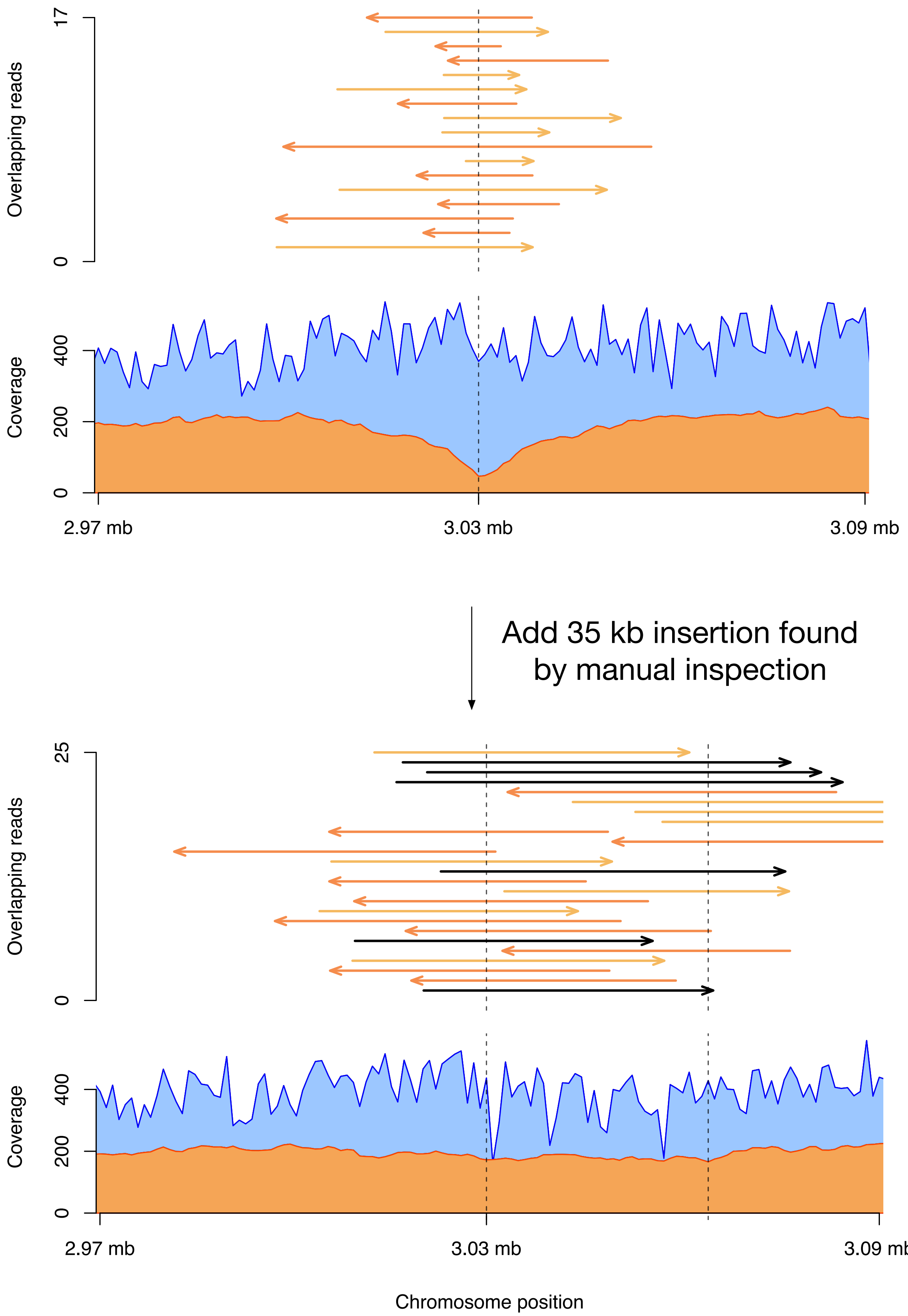
• Genomes from **majority of population** recovered 10 of 13 MAGs appear to be complete genomes
• Illumina-only assembly suffers from **PCR bias** in communities with high GC variability, would result in fragmented assembly

Future work

• What pathways are present that contribute to 1-adamantanecarboxylic acid degradation?
• Can we identify functional mobile genetic elements?
• Can we apply this method to more diverse communities?

References

[1]Paulssen JM, Gieg LM. Biodegradation of 1-adamantanecarboxylic acid by algal-bacterial microbial communities derived from oil sands tailings ponds. Algal Research, 2019.
[2]Giguere DJ, et al. Complete and validated genomes from a metagenome. bioRxiv, 2020.

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

github.com/dgiguer   @danieljgiguere   dgiguer@uwo.ca

SUNCOR ENERGY   NSERC CRSNG   Ontario Graduate Scholarship