In-Silico enrichment of bacterial plasmids using nanopore adaptive sampling

Jens-Uwe Ulrich[1,2], Lennard Epping[3], Tanja Pilz[3], Birgit Walther[4], Kerstin Stingl[5], Torsten Semmler[3], Bernhard Y. Renard[1]

[1] Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, [2] Department of Mathematics and Computer Science, Free University of Berlin, [3] Genome Sequencing and Genomic Epidemiology, Robert Koch Institute, Berlin, [4] Advanced Light and Electron Microscopy, Robert Koch Institute, Berlin, [5] National Reference Laboratory for Campylobacter, Department of Biological Safety, German Federal Institute for Risk Assessment (BfR), Berlin; jens-uwe.ulrich@hpi.de

Introduction

Horizontal gene transfer via plasmids plays a major role regarding the spread of antibiotic resistance genes (ARGs). However, characterization of ARGs located on low abundant plasmids by whole genome sequencing is challenging. Although sample preparation methods can enrich the proportion of plasmid DNA before sequencing, these methods are expensive and laborious, and they might introduce a bias by enriching only for specific plasmid DNA sequences.

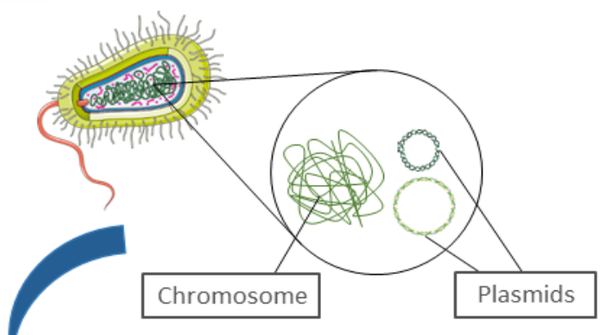
Aim

- Investigate adaptive sampling for the enrichment of low abundant plasmids using MinKNOW or ReadBouncer
- Inspect potential cost savings by utilizing adaptive sampling on expired flow cells

Methods

sampling

- 5 bacterial strains with different plasmid abundance
  - *Campylobacter jejuni*
  - *Campylobacter coli*
  - *Klebsiella pneumoniae*
  - *Enterobacter hormaechei*
  - *Salmonella enterica*



cultivation

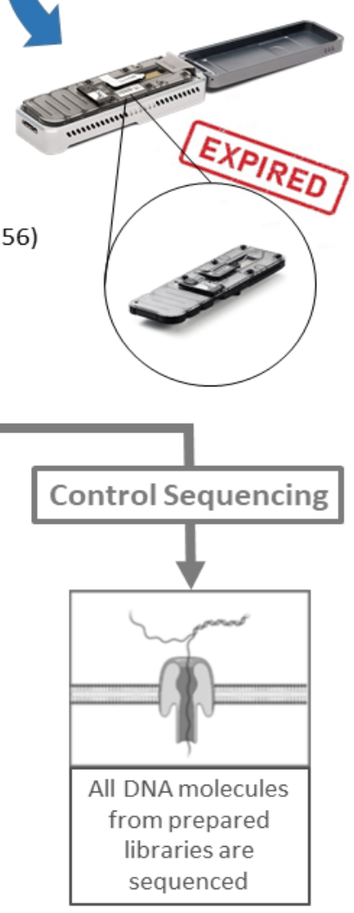
- *Campylobacter* strains streaked on Columbia blood agar
- Others were streaked on Lauria Bertani plates

extraction & preparation

- *Campylobacter jejuni* DNA extracted with MagAttract HMW Genomic Extraction Kit (Qiagen)
- DNA of other four strains extracted with QIAamp DNA Mini Kit
- Library preparation using Rapid Barcoding Kit

sequencing & adaptive sampling

- Using expired FLO-MIN106D flowcells
- Divide flowcells into *adaptive sampling* (channels 1-256) and *control* region (channels 257-512)
- Adaptive sampling either done using *MinKNOW* (deplete mode) or *ReadBouncer*
- Trying to reject all bacterial chromosome reads to enrich for low-abundant plasmids



Results

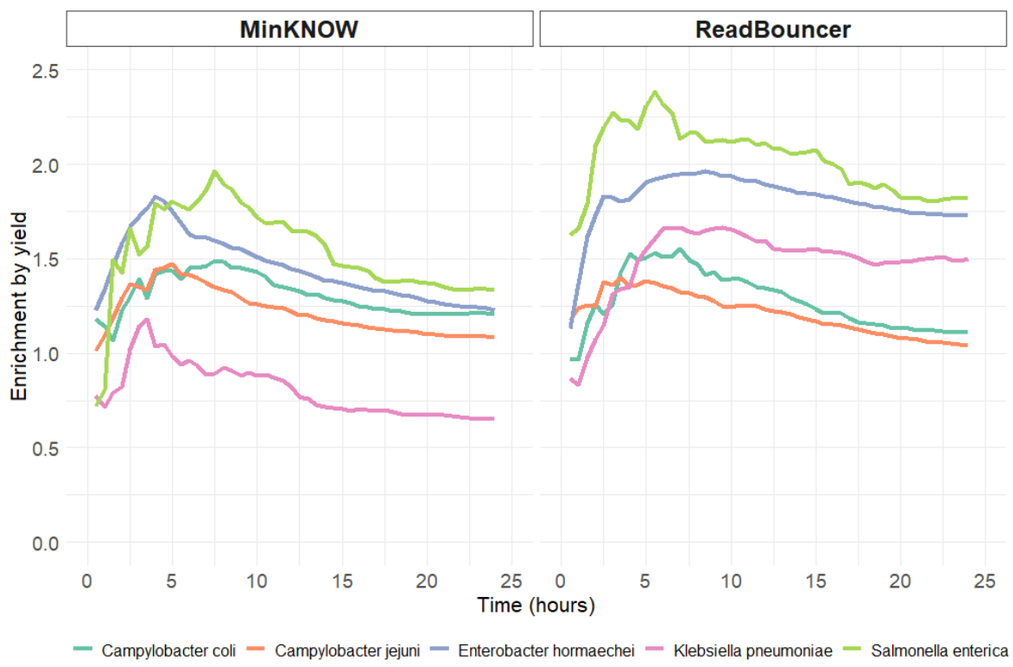


Fig 1: Comparison of enrichment by yield in five bacterial samples between *MinKNOW* and *ReadBouncer*

	Salmonella enterica			*Enterobacter hormaechei*			*Klebsiella pneumoniae*		
Adaptive Sampling (channels 1-256)									
Time (hours)	1	2	3	1	2	3	1	2	3
Plasmid reads	73	190	321	853	2,404	4,219	77	187	318
NGA50	93,838	93,830	93,828	292,652	292,567	292,570	102,925	154,202	154,190
Ref. avg. coverage depth	9	23	40	12	37	65	3	8	16
Ref. coverage ≥ 10x(%)	50.39	100	100	36.98	99.81	99.84	0.0	34.10	89.10
Mismatches per 100kb	188	132	132	91	25	24	577	29	5
Indels per 100kb	676	658	655	130	25	24	722	78	21
Control (channels 257-512)									
Time (hours)	1	2	3	1	2	3	1	2	3
Plasmid reads	43	98	150	606	1,429	2,357	124	289	492
NGA50	91,692	93,851	93,828	109,532	292,998	292,565	103,578	153,446	154,178
Ref. avg. coverage depth	5	11	17	9	22	36	4	8	14
Ref. coverage ≥ 10x(%)	6.19	75.92	100	17.54	77.40	98.73	3.17	35.73	78.62
Mismatches per 100kb	263	145	129	190	35	25	161	177	45
Indels per 100kb	876	685	656	276	54	26	290	198	38

Table 1: Plasmid Assembly statistics of adaptive sampling and control region at three different time points of sequencing for three species when using *ReadBouncer*. Reads were quality filtered and separately assembled using metaFlye assembler. Assembly statistics were provided by Quast.

Conclusions

- Adaptive sampling results in plasmid enrichment by yield up to 1.8x after 24 hours of sequencing
- No negative impact on the enrichment of target sequences when using expired flowcells
- *ReadBouncer* performs better if regions of high sequence similarity are located on the chromosome and one of the plasmids
- Assemblies of plasmids after two hours of sequencing with adaptive sampling show higher accuracy than plasmid assemblies after three hours of standard sequencing

- Kav, A.B. Et al. A method for purifying high quality and high yield plasmid dna for metagenomic and deep sequencing approaches. Journal of microbiological methods, 95(2):272–279, 2013.
- Lynch, M. et al. Ecology and exploration of the rare biosphere. Nature Reviews Microbiology, 13(4):217–229, 2015.
- Ulrich, J. U. et al. ReadBouncer: precise and scalable adaptive sampling for nanopore sequencing. Bioinformatics, 38 (Supplement_1):i153–i160, 06 2022.
- Kolmogorov, M. et al. metaflye: scalable long-read metagenome assembly using repeat graphs. Nature Methods, 17(11):1103–1110, 2020.
- Gurevich, A. et al. Quast: quality assessment tool for genome assemblies. Bioinformatics, 29(8):1072–1075, 2013
- Ulrich, J. U. et al. (2022). In-silico enrichment of bacterial plasmids by chromosomal depletion using nanopore adaptive sampling. *bioRxiv*.

DZIF German Center for Infection Research
Federal Ministry of Education and Research
DFG Deutsche Forschungsgemeinschaft German Research Foundation
ROBERT KOCH INSTITUT
HPI Hasso Plattner Institut Digital Engineering · Universität Potsdam