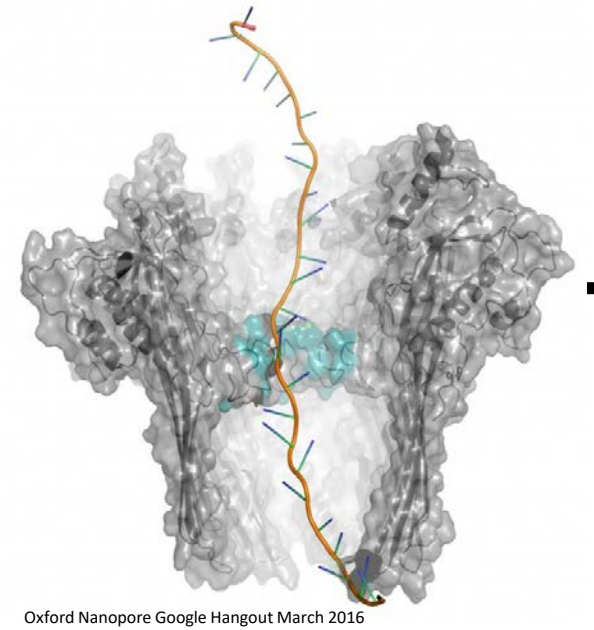
JHU Microbiome Seminar
October 16, 2017

JOHNS HOPKINS
BIOMEDICAL ENGINEERING

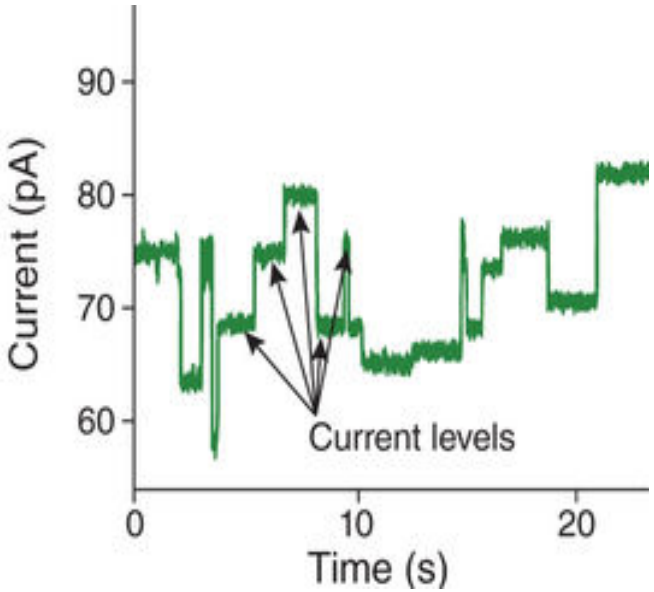
Nanopore Sequencing for exploration of Infectious Disease

Winston Timp
Department of Biomedical Engineering
Johns Hopkins University

Nanopore: Single Molecule Sequencing



Oxford Nanopore Google Hangout March 2016



Deamer et al 2016, Nature Biotech

ATCGATCGATAGTA
TTAGATACGACTAG
CGATCAG

- Oxford Nanopore Technologies, CsgG biological pore
- No theoretical upper limit to sequencing read length, practical limit only in delivering DNA to the pore intact
- Palm sized sequencer
- Predicted sequencing output 3-6Gb

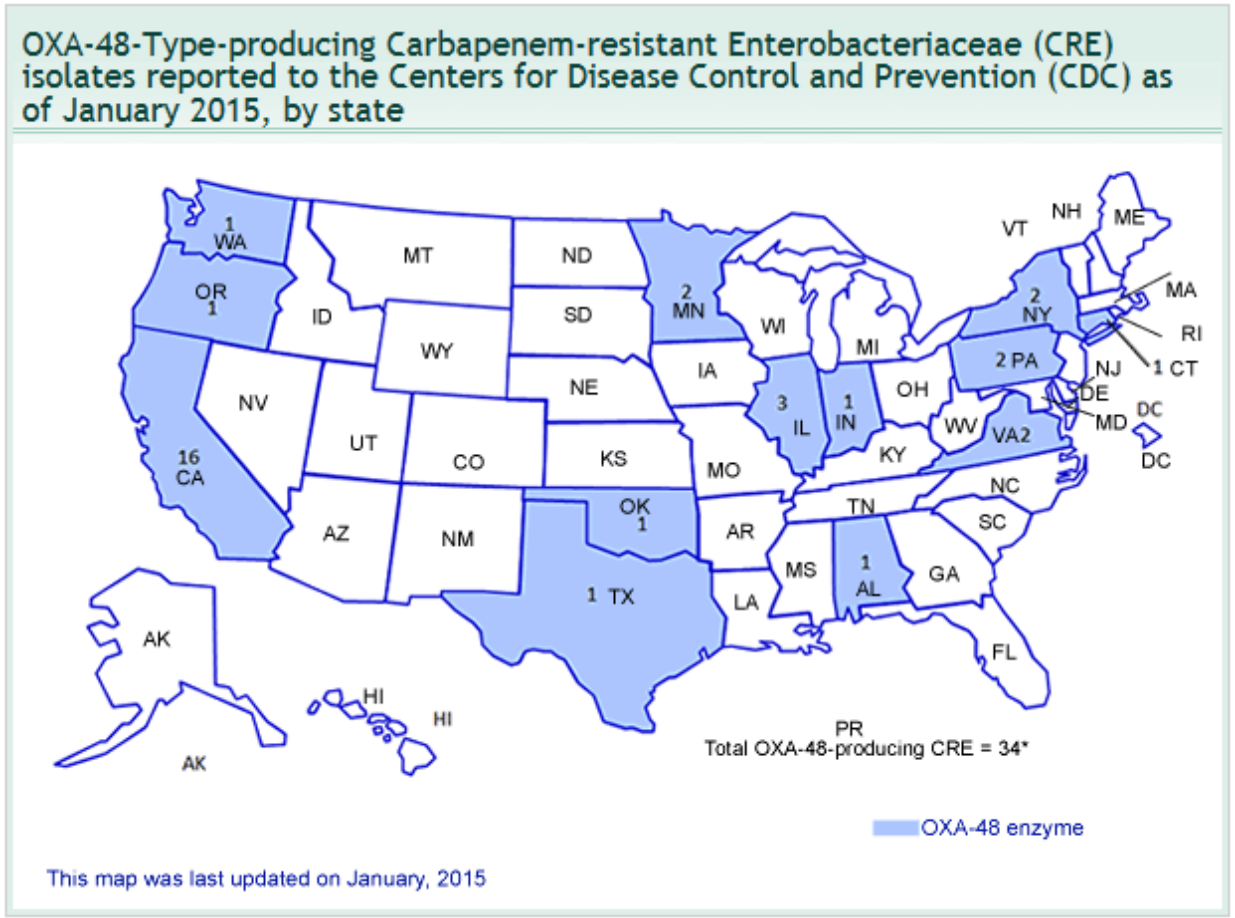
Disclosure: Timp has two patents (US 2011/0226623 A1; US2012/0040343 A1) licensed to ONT

Hypervirulent (hypermucoviscous) *K. pneumoniae*

New variant of *Klebsiella pneumoniae*
First described in the Asian Pacific Rim 1980s
Now increasingly recognized in other countries

Defining clinical features:

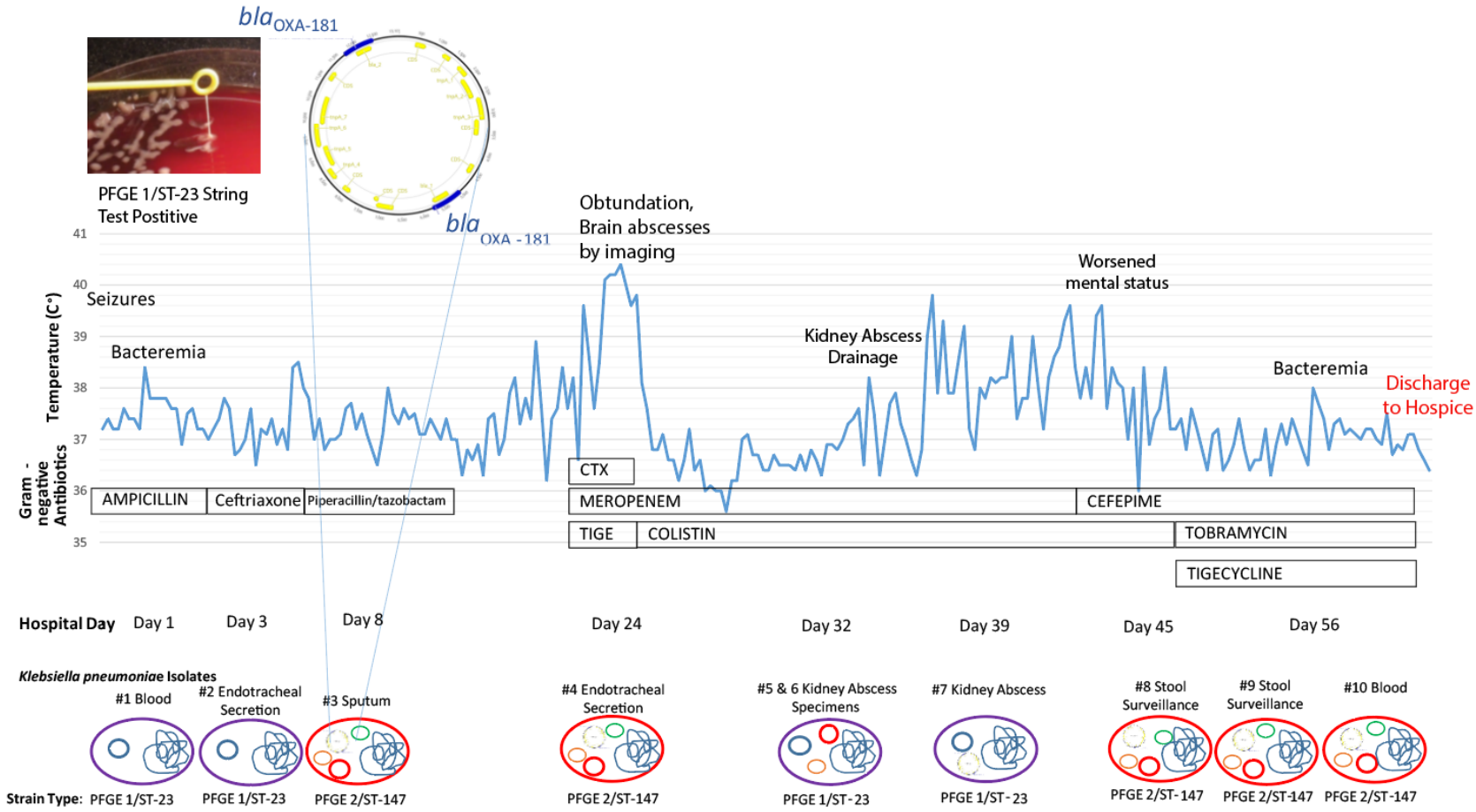
- Serious, life-threatening community-acquired infection in younger healthy hosts
- Liver abscess, pneumonia, meningitis and endophthalmitis
- Metastatic spread



Source: Shon, Rajinda, Russo 2013
CDC
http://www.cdc.gov/hai/organisms/cre/TrackingCRE.html#CREmapOXA

Isolate Sequencing: Patient

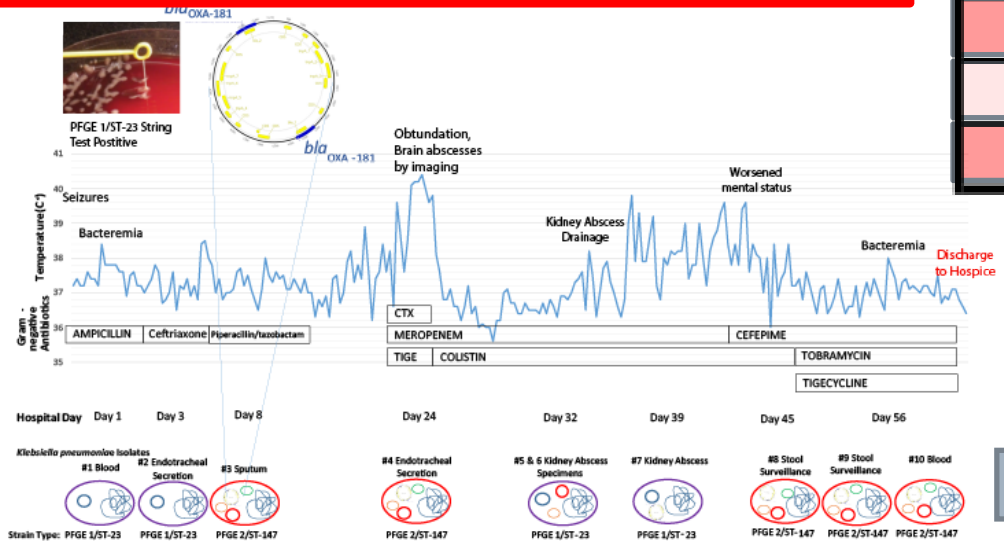
- 44yo m with cirrhosis. Transferred from Holy Cross Hospital on 7/26/15 with seizure, sepsis and *K. pneumoniae* pneumonia. Found to have kidney and brain abscesses.
- Travel history: Social visits to India in April/May2015
- Medical care abroad: Hospital stay in India after a fall
- Previous admission 6/21 – 6/28 (MICU) -> Bacteremia and pneumonia
- Admitted to MICU 7/27; Contact isolation started 8/7 (CRE from sputum collected 8/4)
  - *K. pneumoniae* (varying resistance from sputum, renal abscess, blood)
  - Identified as multi-drug resistant (8/7)
  - Identified as hypermucoviscous (8/7)
  - Positive for *bla*$_{OXA-48}$ enzyme (9/15)
  - Verified to be *bla*$_{OXA-48}$ by DHMH/CDC



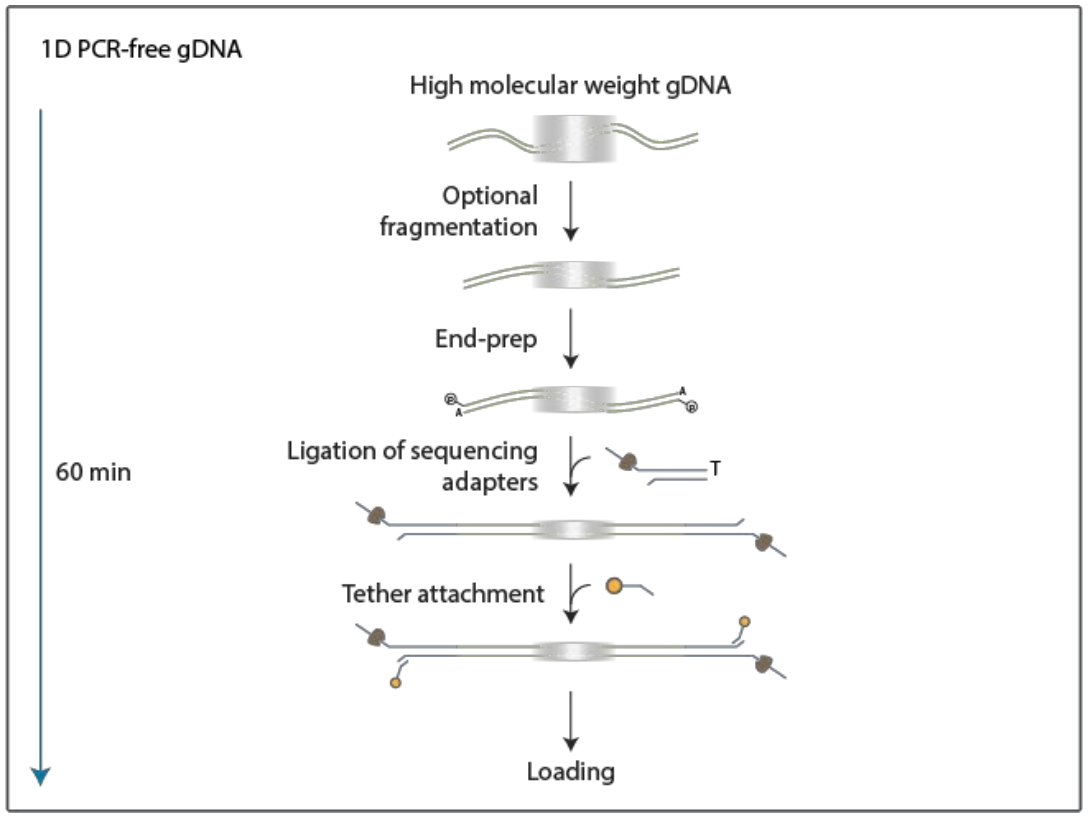
Isolates Sequenced

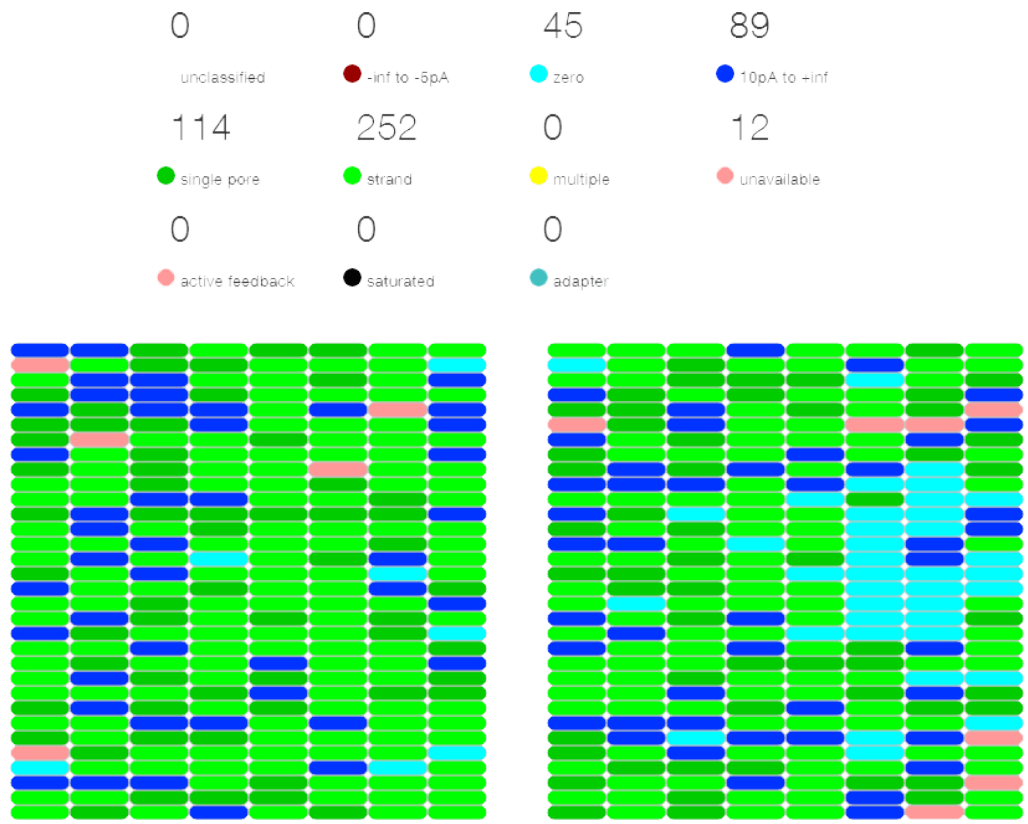
- Isolates grown up and extracted with MoBio Power Biofilm kit
- All isolates sequenced on Illumina
- 9 isolates sequenced on nanopore

Isolate	Hospital Day	Source	Resistance	Sequencer
1	1	Blood	No	Illumina and Nanopore
2	3	Endo/Nasal	No	Illumina and Nanopore
3	8	Sputum	Yes	Illumina Only
4	24	Endo/Nasal	Yes	Illumina and Nanopore
5	32	Kidney Abcess	No	Illumina Only
6	32	Kidney Abcess	No	Illumina and Nanopore
7	39	Kidney Abcess	No	Illumina and Nanopore
8	45	Stool	Yes	Illumina and Nanopore
9	45	Stool	Yes	Illumina and Nanopore
10	56	Blood	Yes	Illumina and Nanopore
11	50	Room	Yes	Illumina Only
12	50	Room	Yes	Illumina and Nanopore



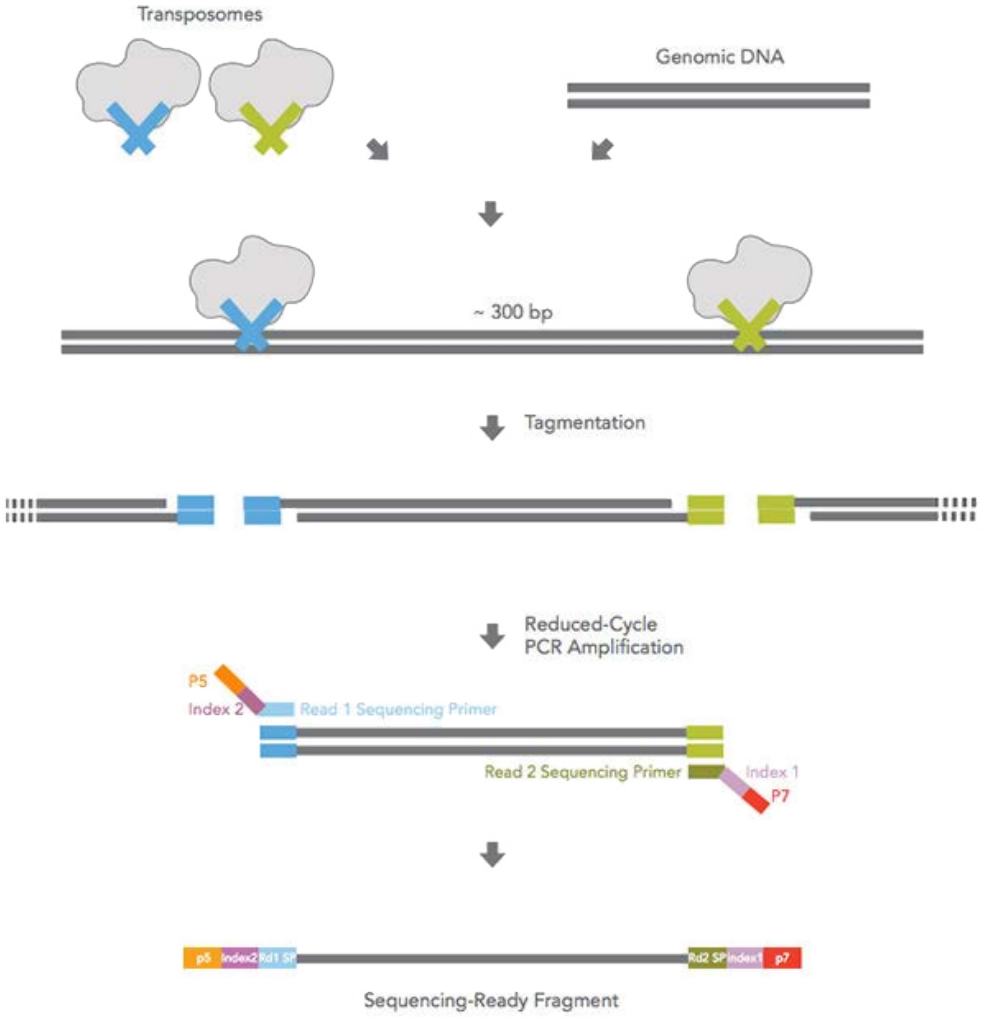
Nanopore Library Prep

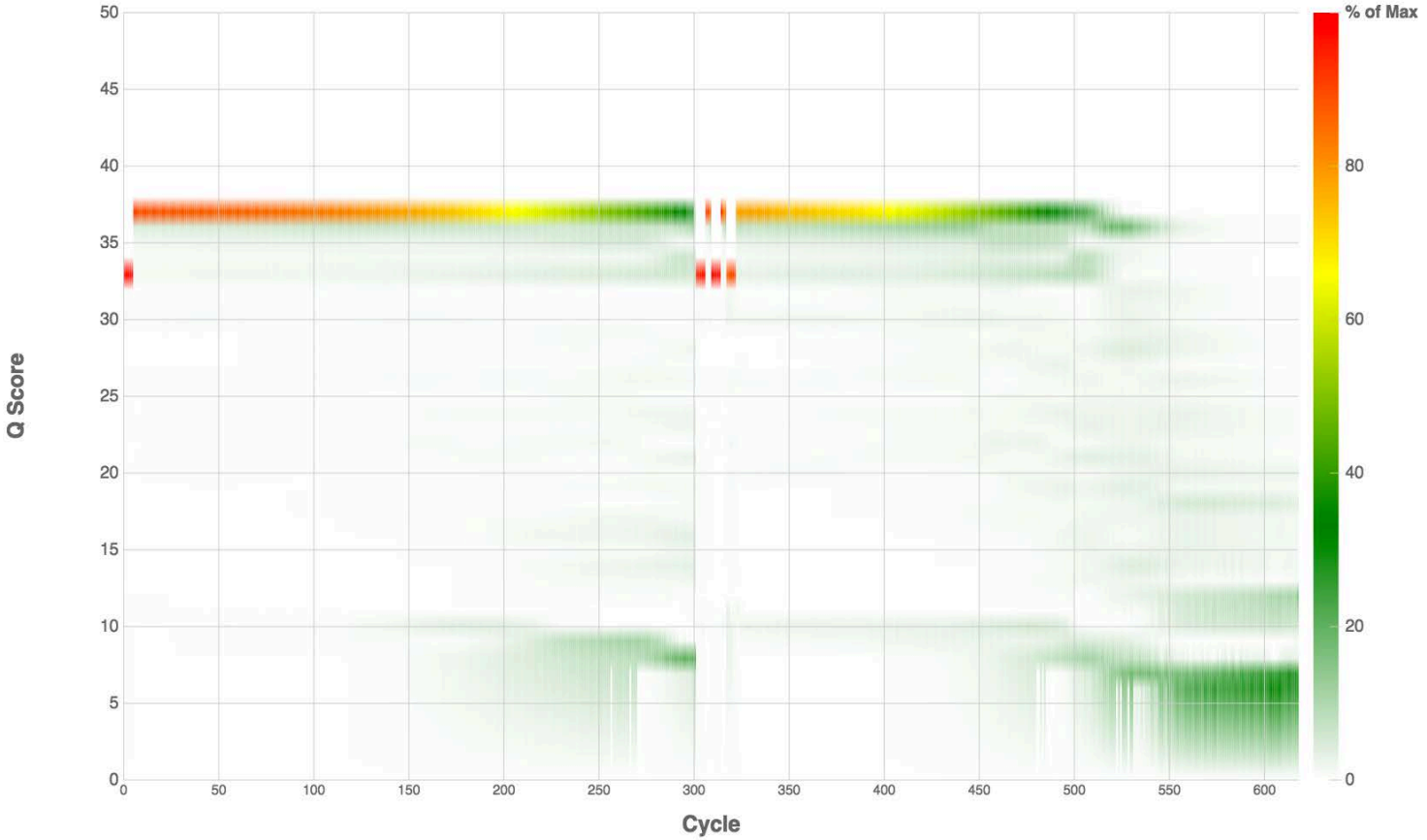




- Library prep is very similar to methods for short-read sequencing
- For DNA shearing we use Covaris gTubes or Diagenode Megaruptor
- After end-repair and A-tailing, leader adapter with motor protein is ligated
- MinION arrays 512 channels (with 4 pores possible per channel) (shown right from running software); dark green pores are sequencing, light green available, other colors inactive.

Illumina Nextera





- Performed Illumina Nextera (transposase based) library prep, needs only 1ng
- Tagmentation (transposase) based prep is straightforward, then sequenced on the MiSeq
- Used 600v3 kit, ~60 hrs runtime
- Multiplexed 3 samples per run.

Run data

Sample	Source	Illumina Reads (M)	Illumina Yield (Gb)	Illumina Chemistry	Nanopore Reads (k)	Nanopore Yield (Mb)	Read Length	Nanopore Chemistry
1	blood	9.2	4.0	600v3	361.0	2158.0	5977	R9
2	endo/nasal	4.6	2.6	600v3	103.6	535.4	5169	R7
3	sputum	6.6	3.0	600v3				
4	endo/nasal	5.9	3.3	600v3	75.7	362.7	4789	R7
5	kidney abcess	10.6	5.1	600v3				
6	kidney abcess	7.5	3.5	600v3	189.4	912.6	4819	R9
7	kidney abcess	6.3	3.5	600v3	100.6	427.5	4250	R7
8	stool	6.4	3.6	600v3	41.7	188.5	4522	R7
9	stool	3.6	2.0	600v3	97.5	467.0	4789	R7
10	blood	9.8	4.7	600v3	359.7	1906.6	5300	R9
11	room	6.1	3.0	600v3				
12	room	10.1	4.9	600v3	446.5	2489.9	5577	R9

Nanopore sequencing yield improved dramatically R7->R9 – from 100s of Mb to typical yield of ~1-5 Gb in our hands – 1-2.5 Gb shown here.

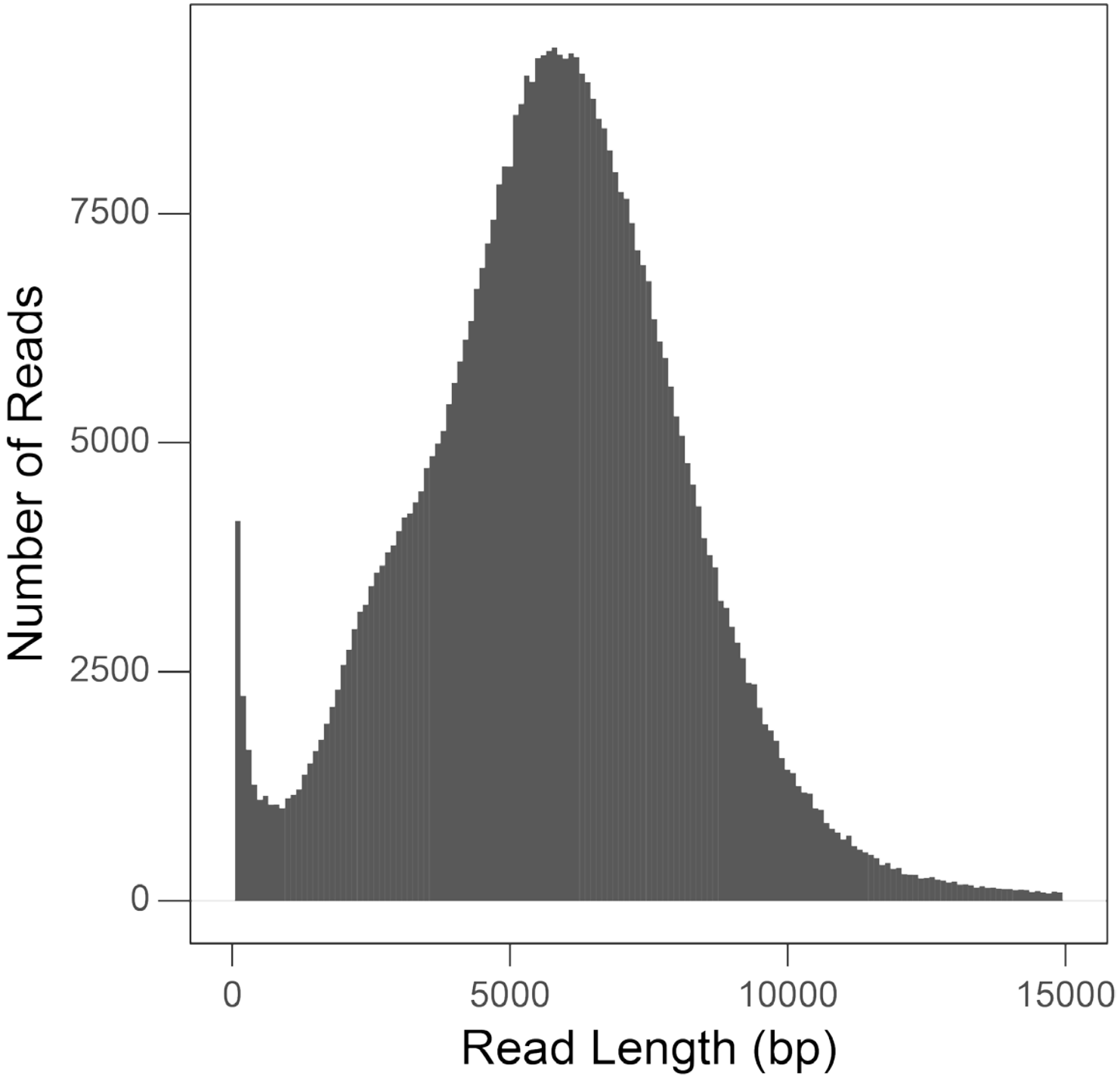
Run data

- Nanopore sequencing of bacterial isolates from the Simner Lab achieves base pair yields often >10Gb per flowcell.
- For reference (in bulk) costs per flowcell are $600 per run – these costs are now competitive with Miseq costs
- For a 2x300 v3 run on the MiSeq (generating 15Gb) the cost is ~$1400, or $93.80 per Gb
- For a nanopore flowcell with average yield of recent bacterial runs (8.9Gb) the cost is $500 or $56.18 per Gb.

Sample	Number of Reads (k)	Median Read Length (bp)	Yield (Gb)
139	1668	8051	13.66
134	959	9697	10.03
140	918	3672	4.93
42	840	8872	7.7
70	620	7409	4.678
25	1361	7292	11.7
162	1121	7551	9.91
29	508	7960	4.7
59	828	6671	6.64
71	1245	7343	10.72
84	1105	8308	10.41
154	721	7402	6.14
94	1546	5615	12.87
33	1014	6685	9.02
133	880	8823	8.75
B1	1897	4889	11.39
19	900	5627	8.75

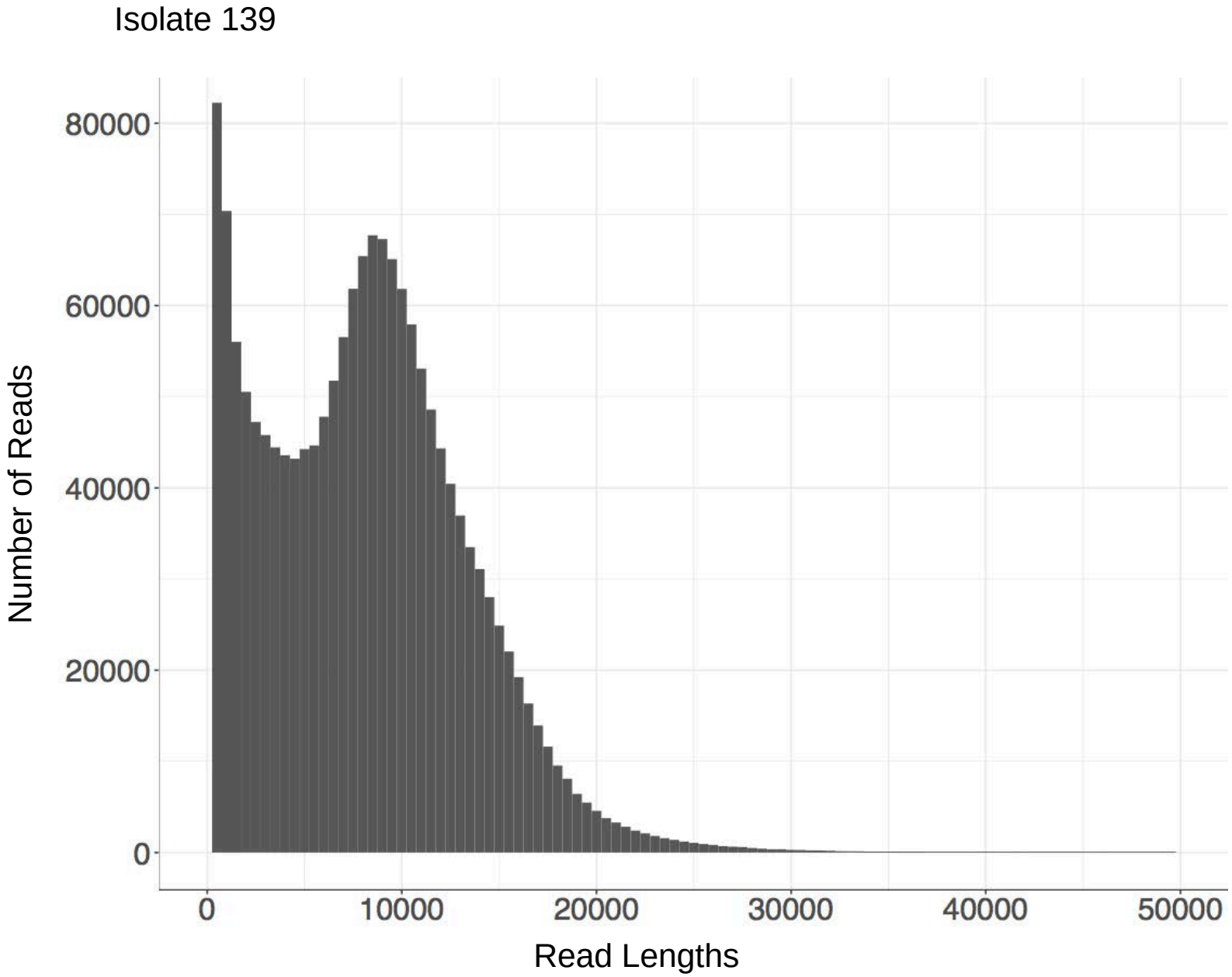
Read length histogram

- Read length for nanopore sequencing was relatively short
- Due to prep methods:
  - HMV was hard to extract effectively, bead beating likely broke DNA
  - We also sheared (Covaris gTubes)
  - To get plasmid sequences, either tagmentation or shearing is likely required to get smaller circular sequences



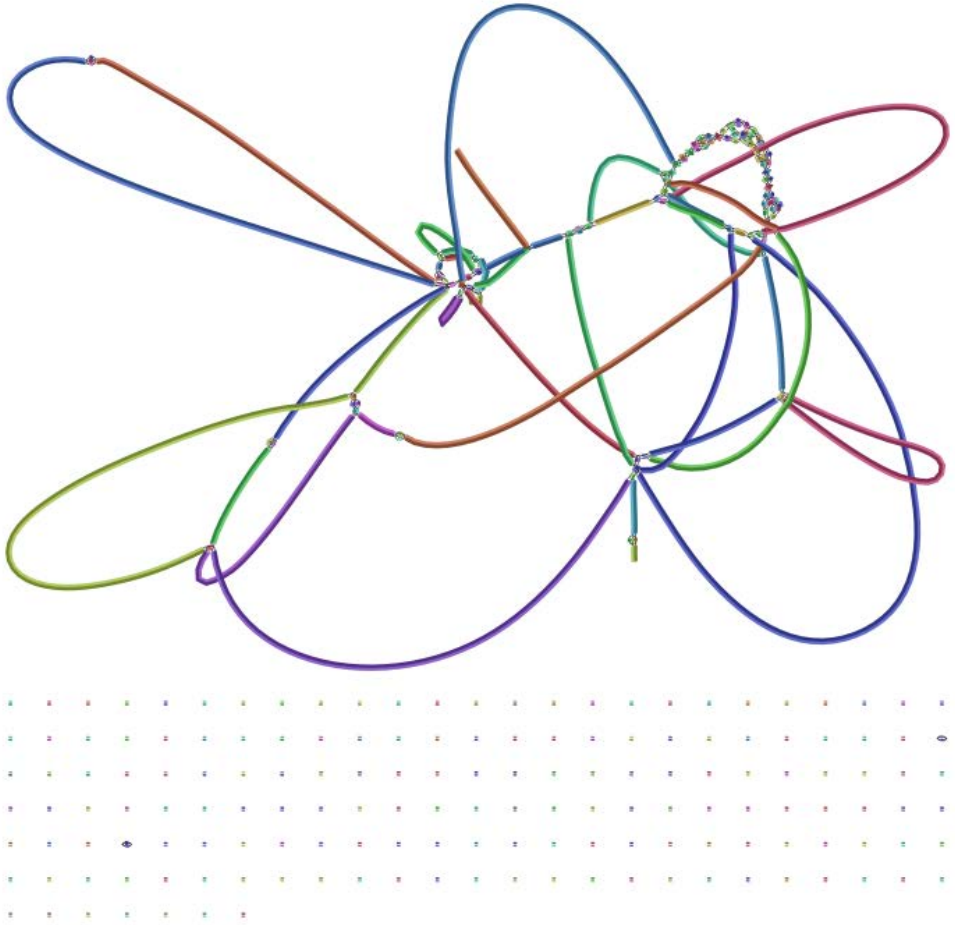
Read length histogram

- Read length for nanopore sequencing was relatively short
- Due to prep methods:
  - HMV was hard to extract effectively, bead beating likely broke DNA
  - We also sheared (Covaris gTubes)
  - To get plasmid sequences, either tagmentation or shearing is likely required to get smaller circular sequences

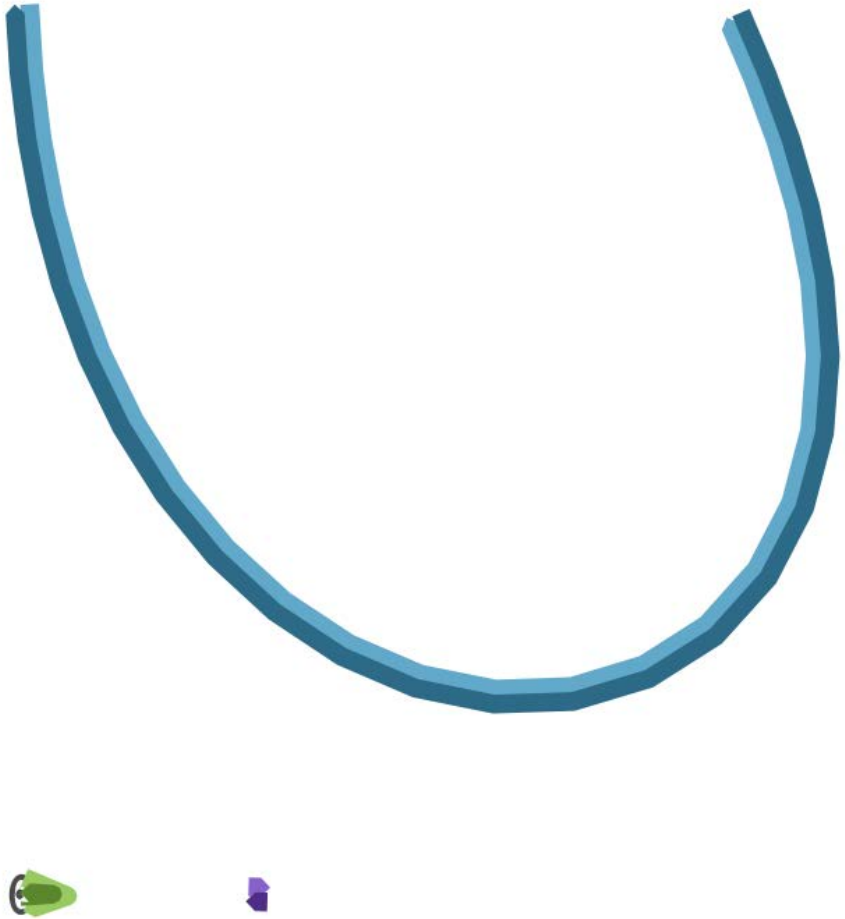


Assemblies

SPAdes (Illumina only)



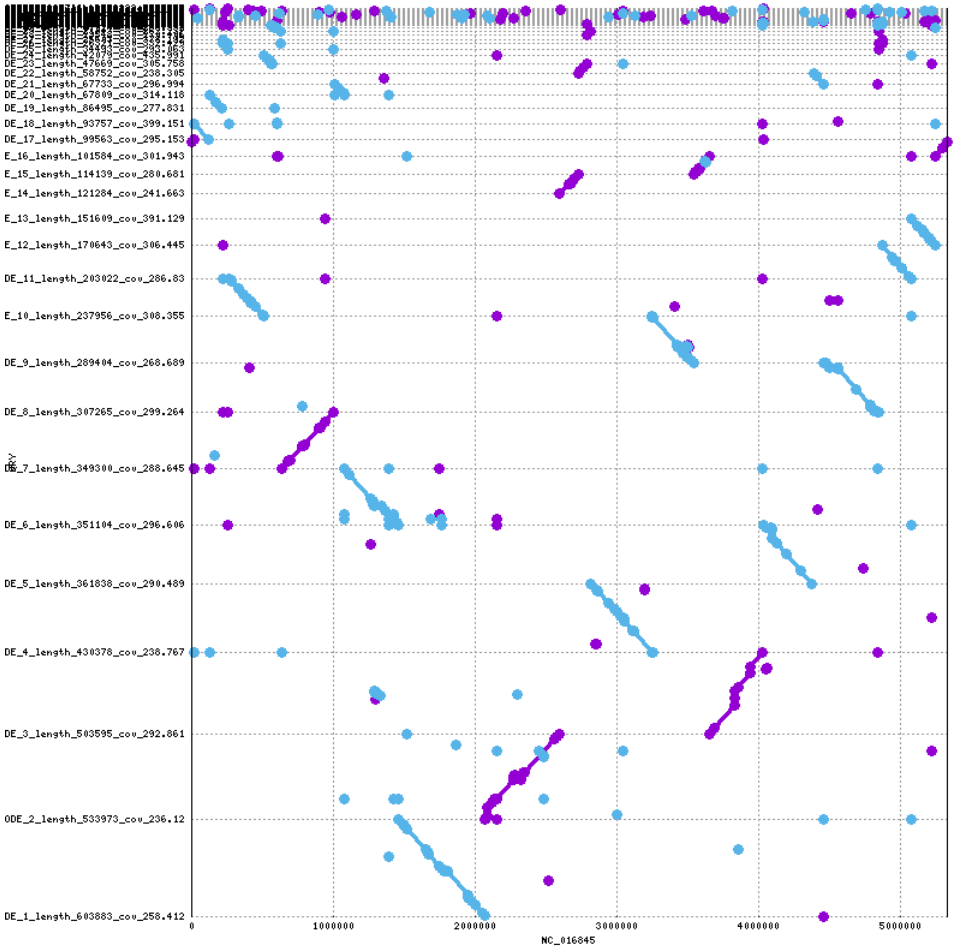
Canu (Nanopore only)



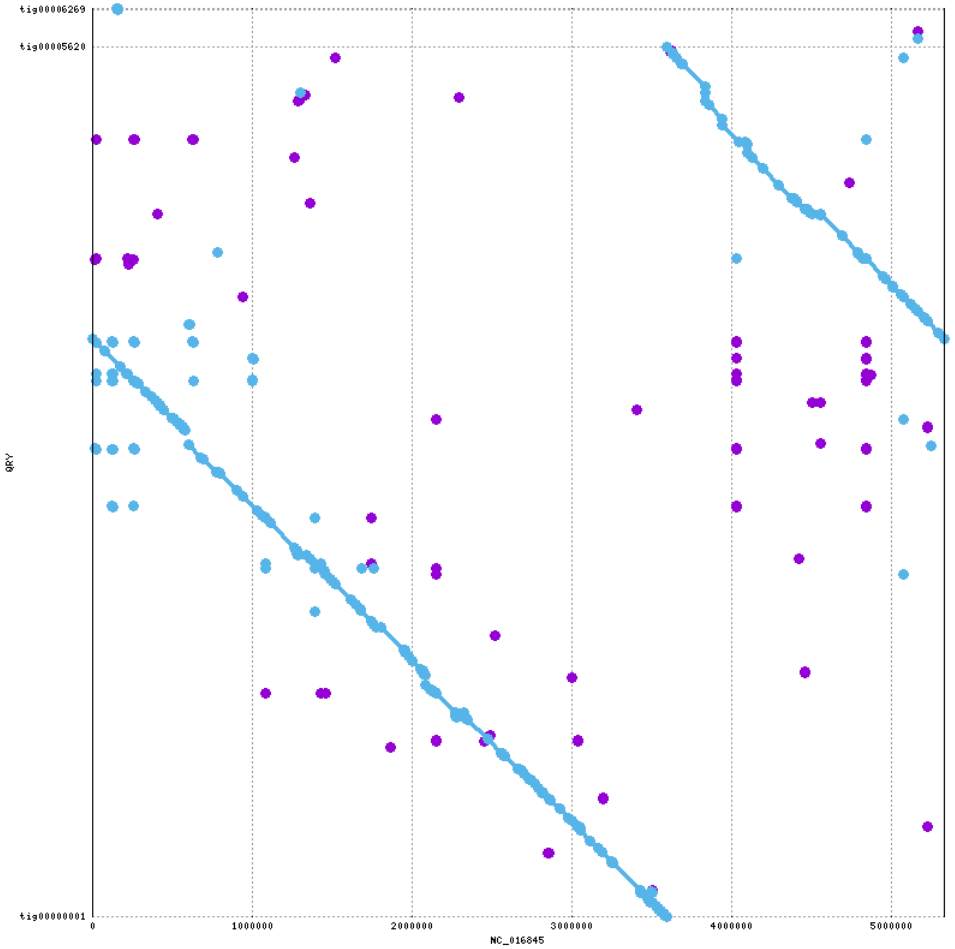
Long reads **really** help in getting complete assemblies – using 2X the yield of Illumina reads, but the N50 is still ~0.3 Mb for no effort Illumina only, while it's full length chromosome from nanopore only

Assemblies

SPAdes (Illumina only)



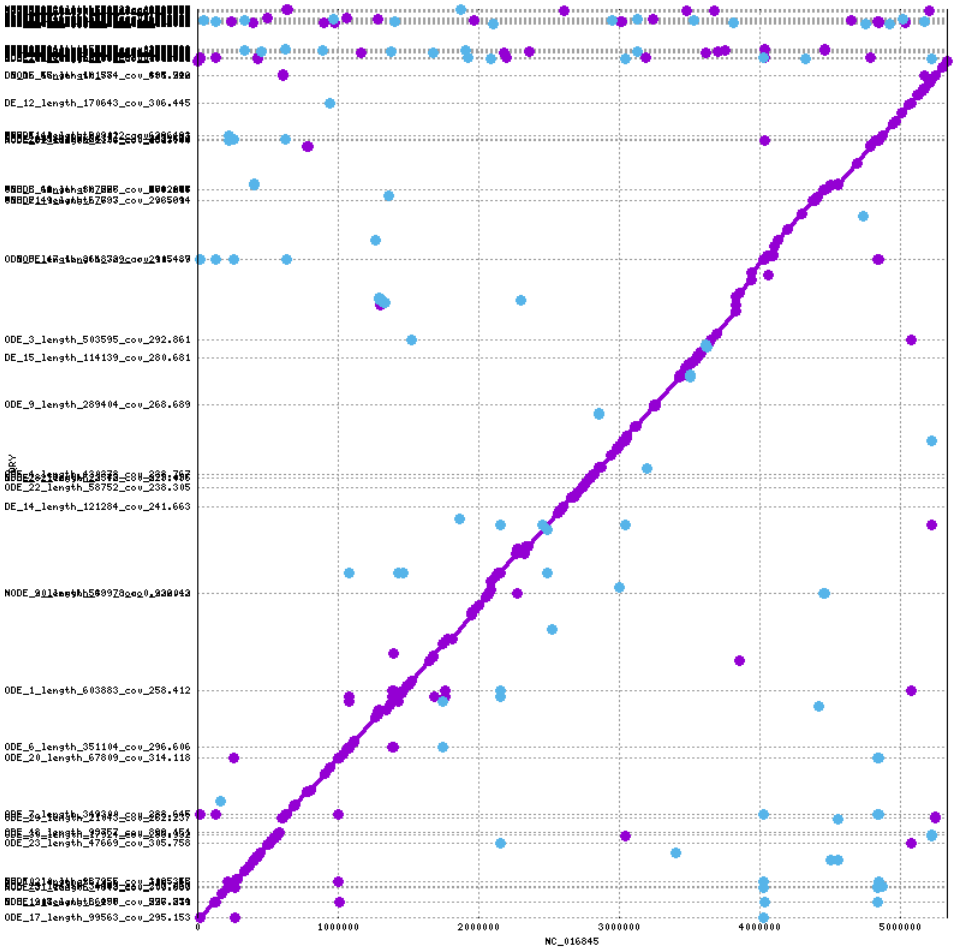
Canu (Nanopore only)



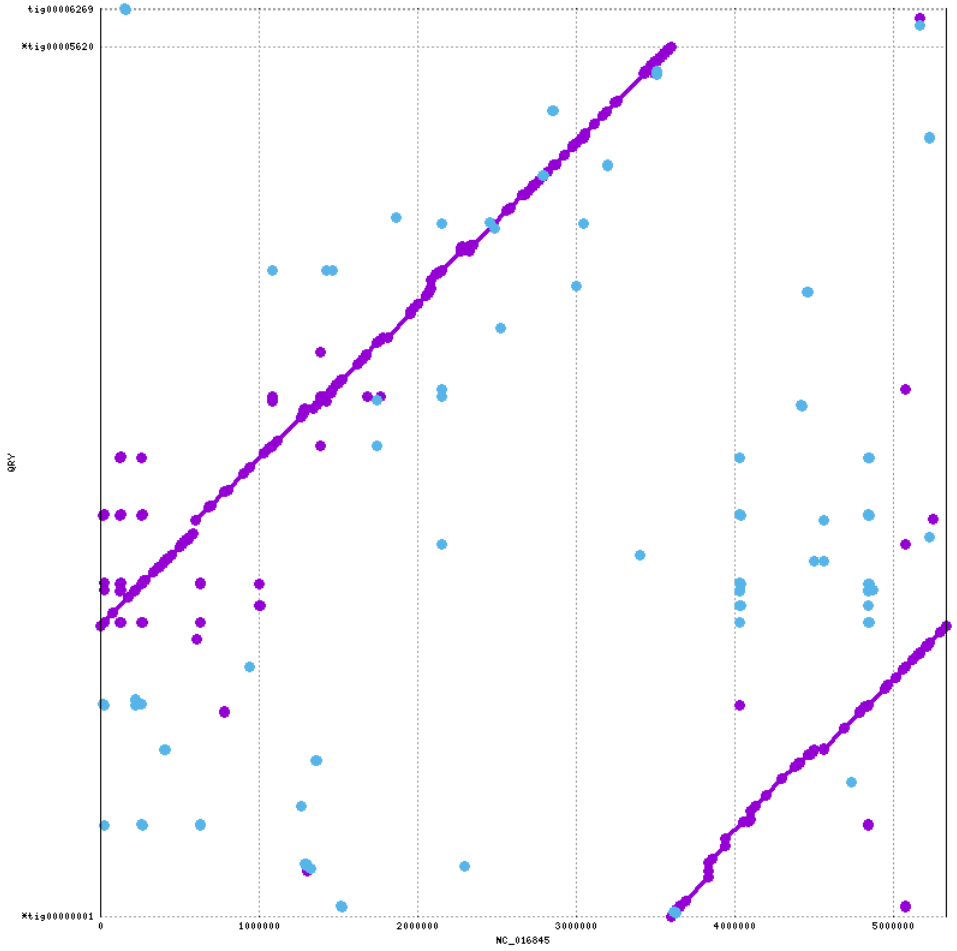
Aligning the assemblies with nucmer – both assemblies capture the chromosome of our isolates

Assemblies

SPAdes (Illumina only)



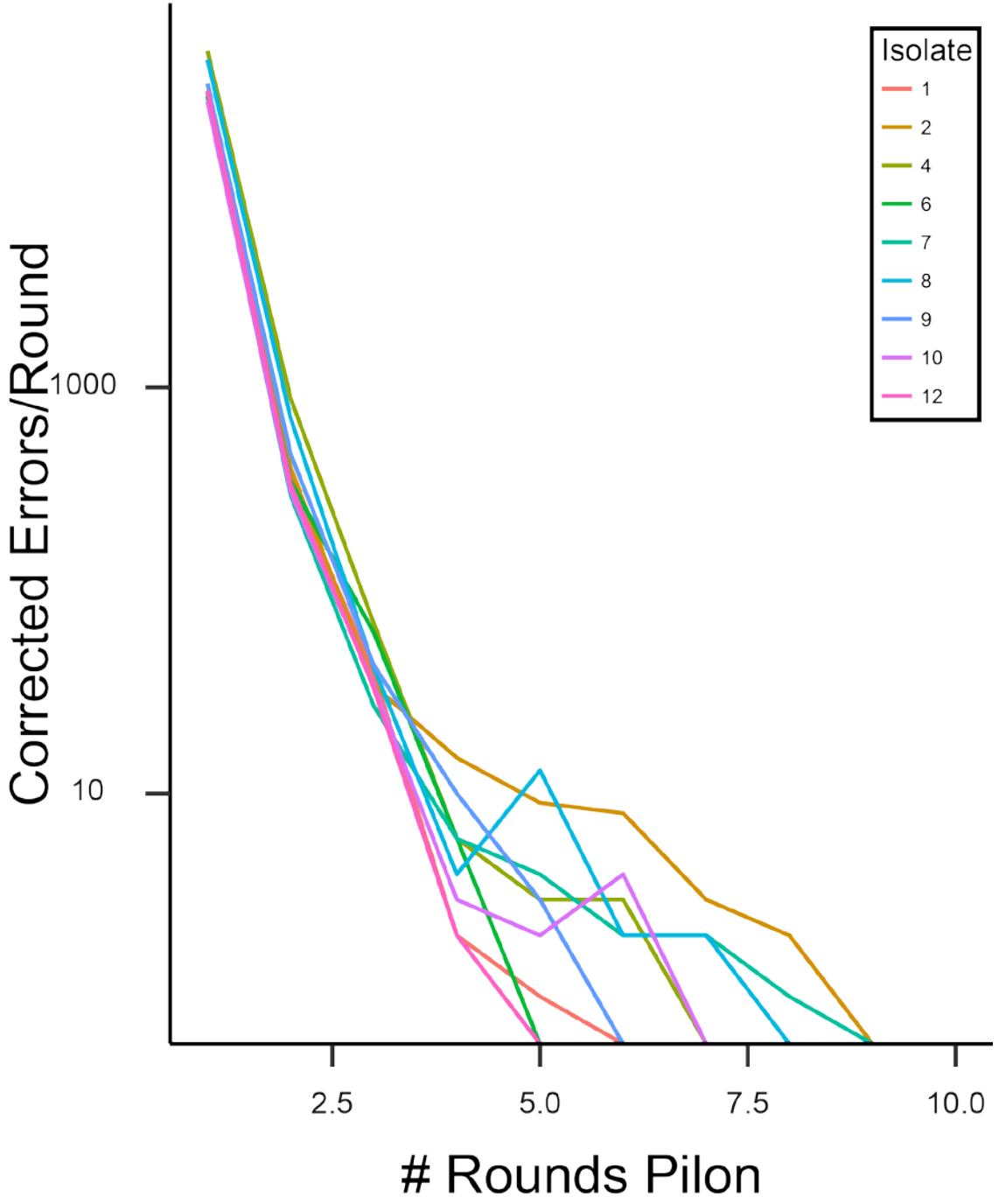
Canu (Nanopore only)



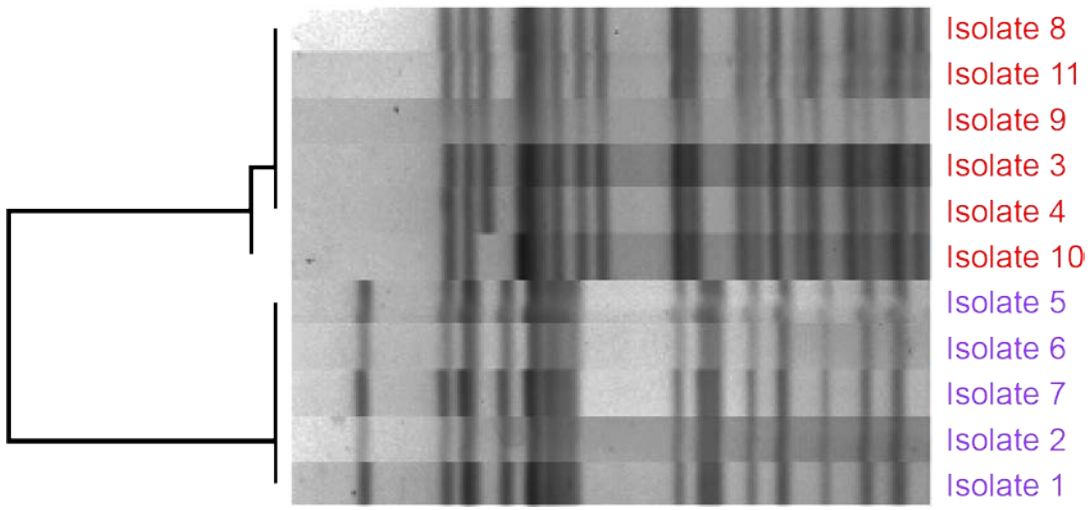
Aligning the assemblies with nucmer – both assemblies capture the chromosome of our isolates

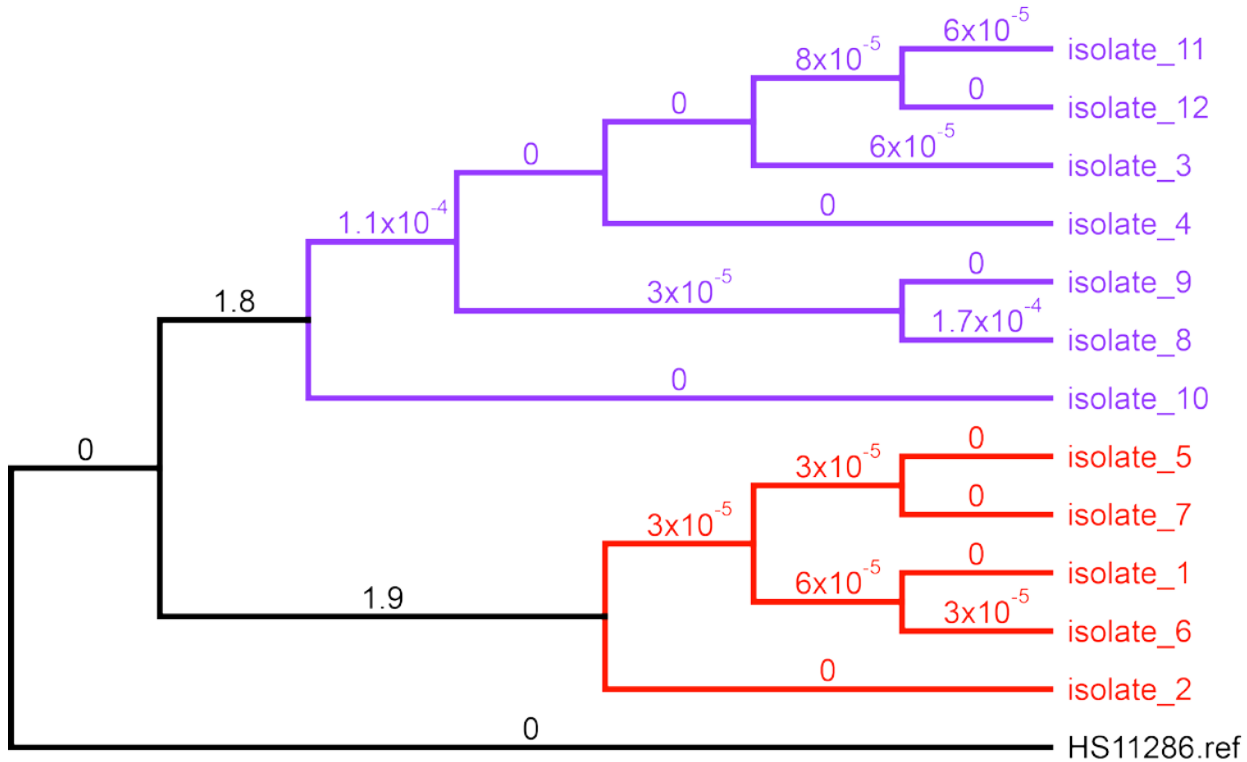
Pilon Correction

- Though ONT error rate is improving, still not sufficient from raw basecalled for SNPs
- Application of Illumina data via pilon allows for correction
- Multiple rounds of correction needed – initially Illumina reads won't align correctly to error prone scaffold
- Next we are trying to apply *nanopolish* which returns to the electrical signal to correct nanopore only assemblies



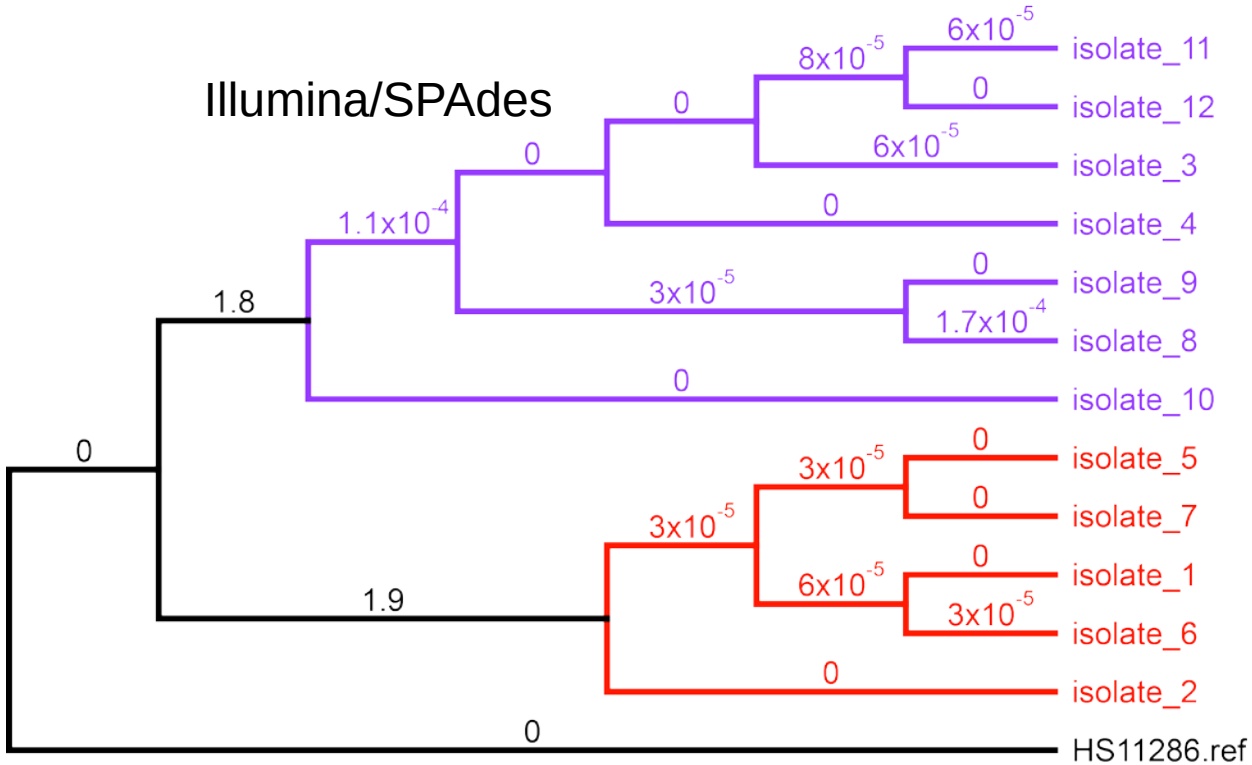
Tree comparing isolates

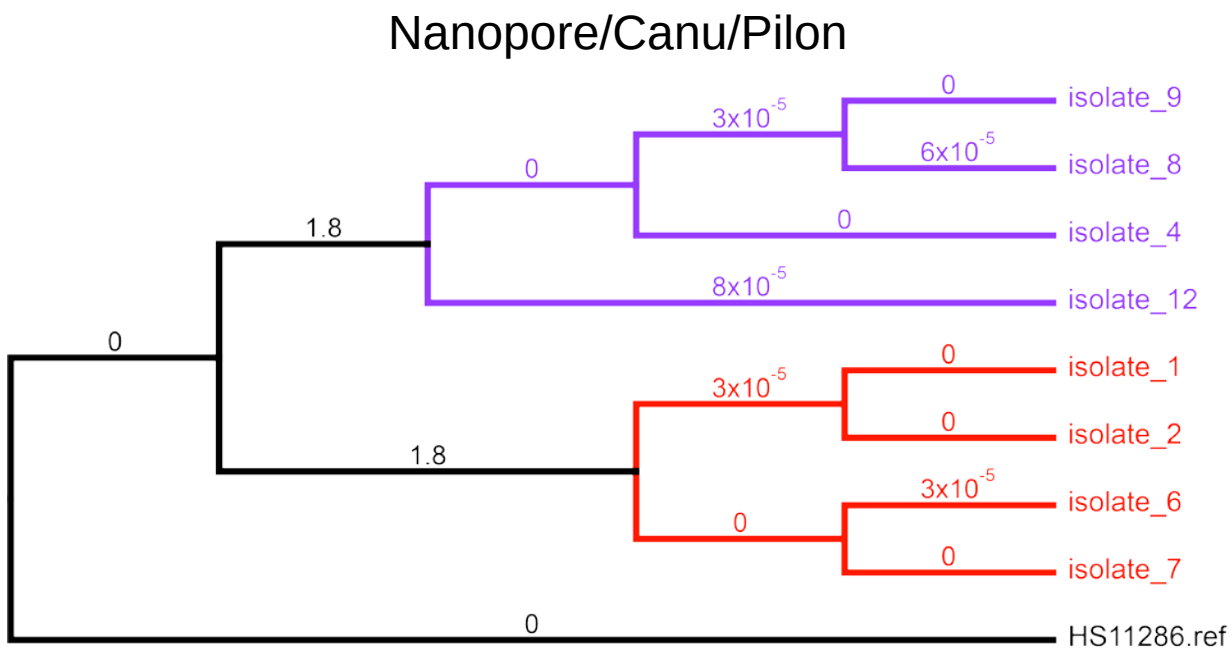




We performed a simple comparison of our isolates using HarvestTools (Parsnp) and found that it largely agreed with the PFGE results`

Illumina Nanopore



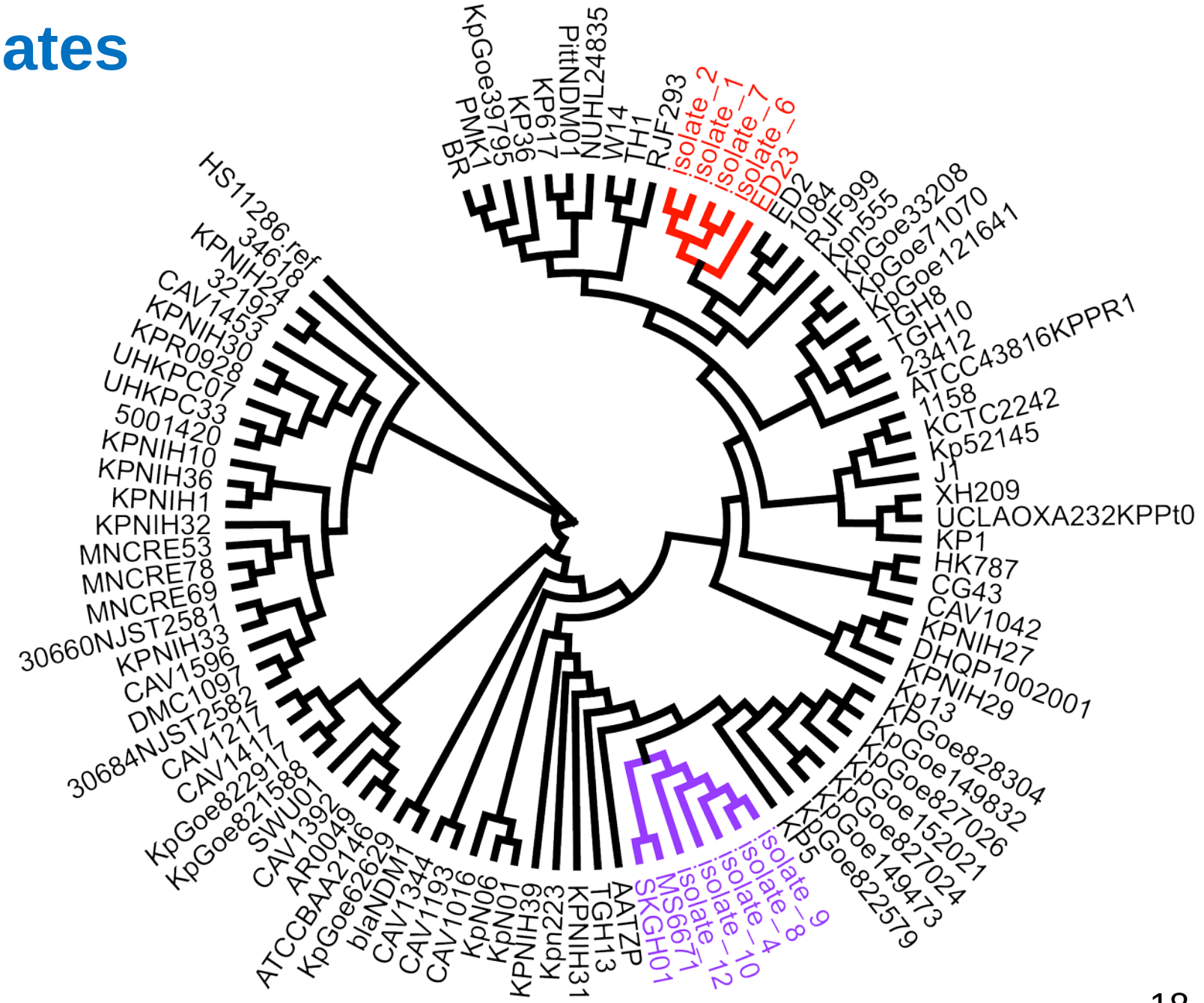


We generated phylogenetic trees from both the SPAdes assemblies and the nanopore assemblies (after 1 round of polishing with pilon from Illumina reads) The nanopore results still showed a larger number of differences, likely due to the higher error rate, especially from some of our older nanopore data.

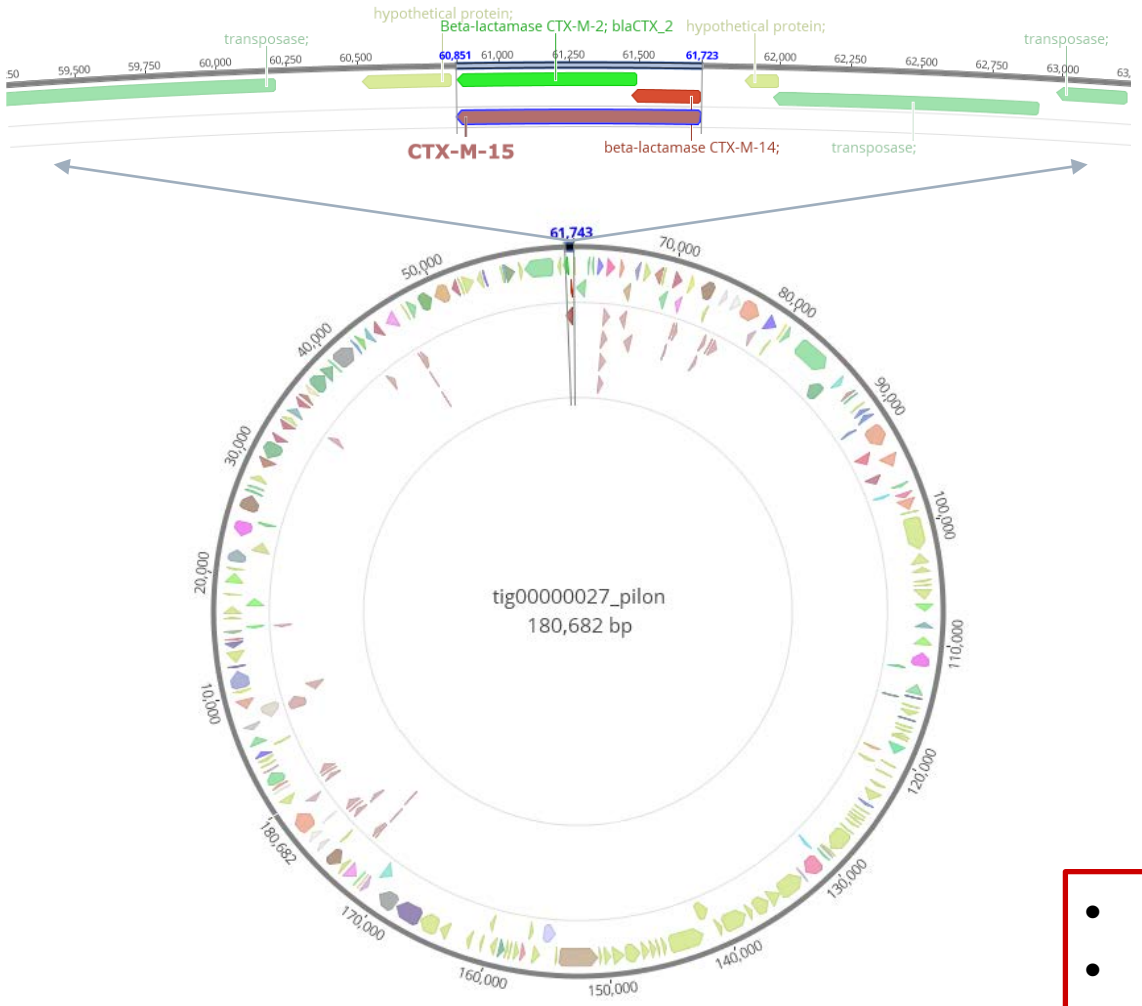
Versus Reference Isolates

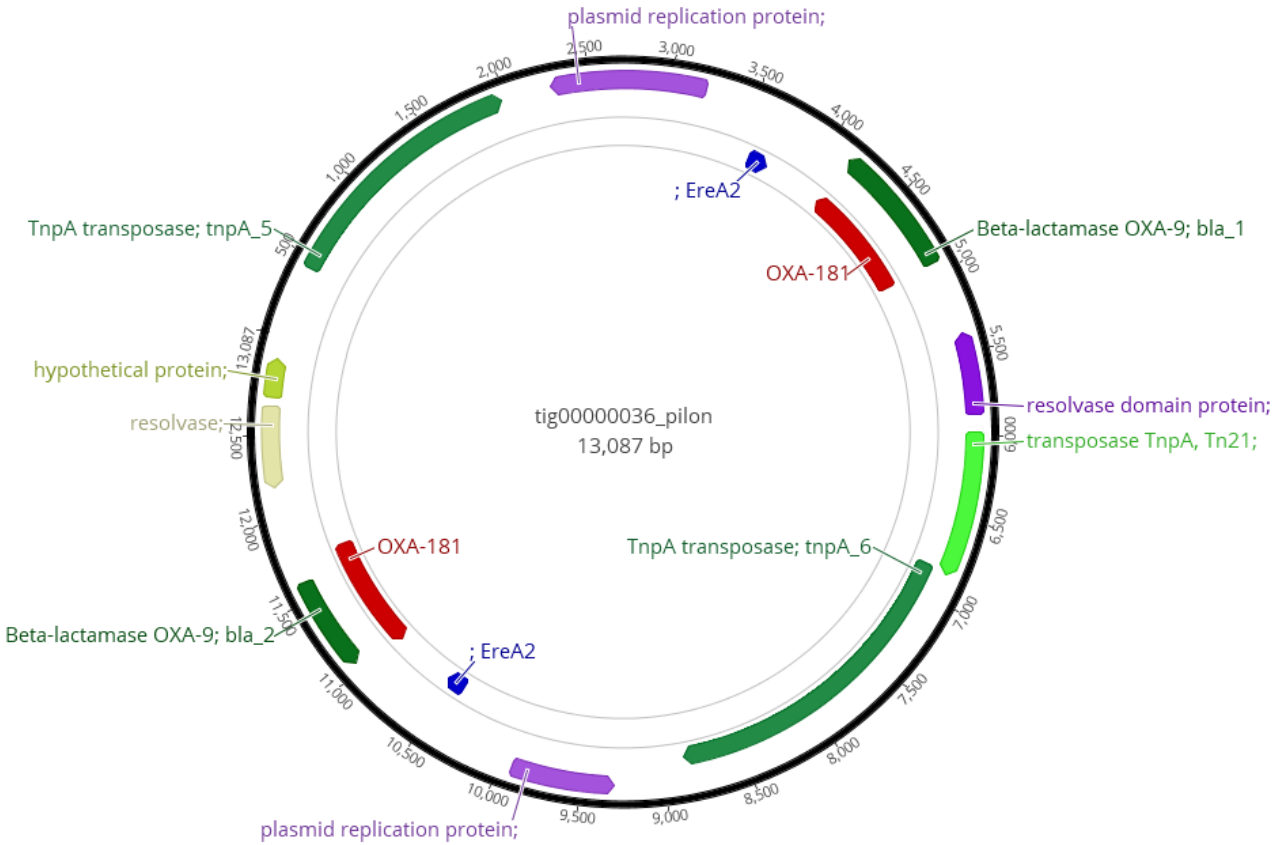
A comparison of our isolate assemblies to NCBI reference genomes for K. pneumo gave clear clustering with specific strains

- HMV strain (MLST ST-23) is closest to MS6671 which is a strain first isolated in the United Arab Emirates (UAE)
- XDR strain (MLST ST-147) is closest to ED23 which is a strain first isolated in Taiwan



AMR genes

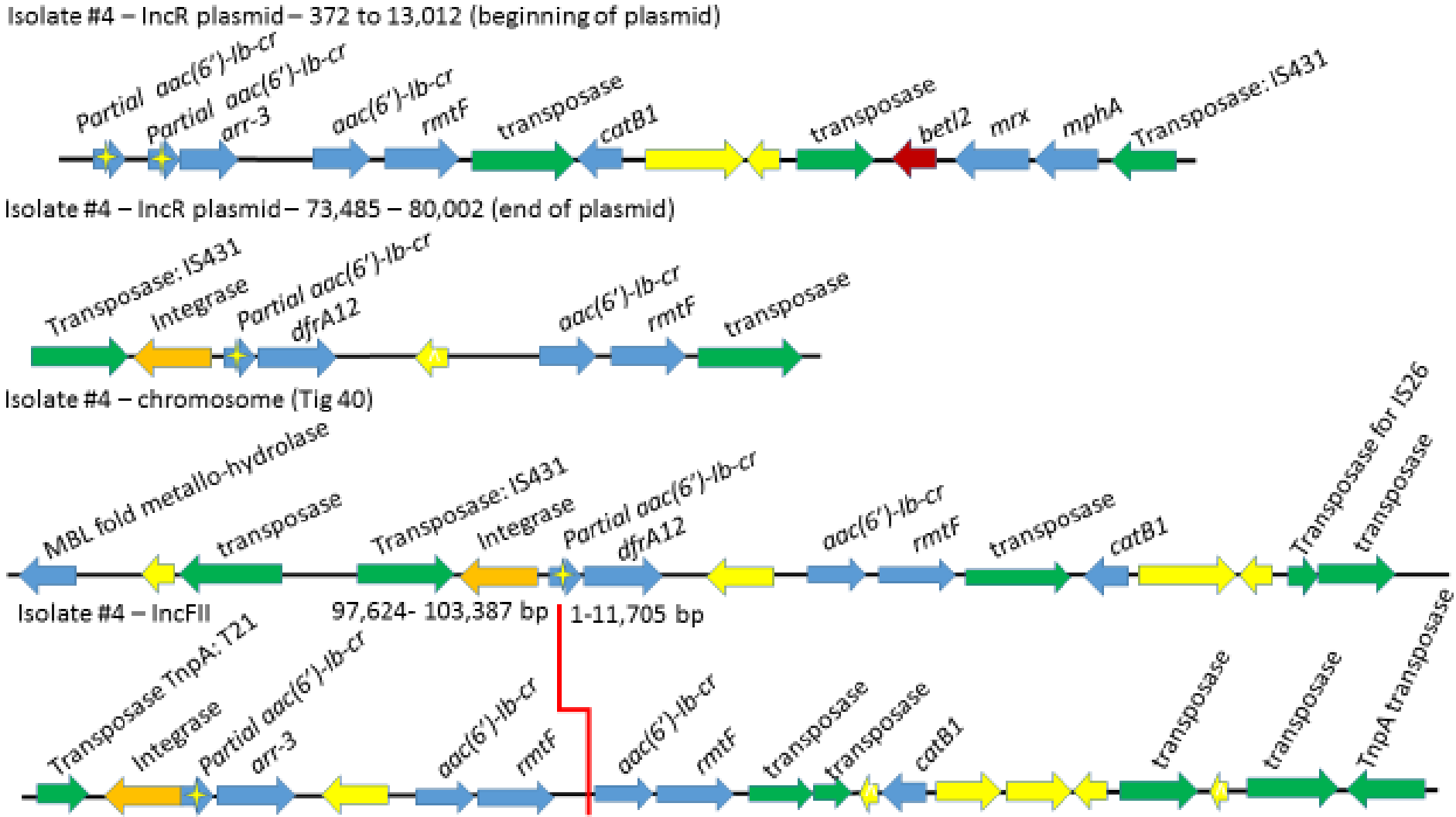




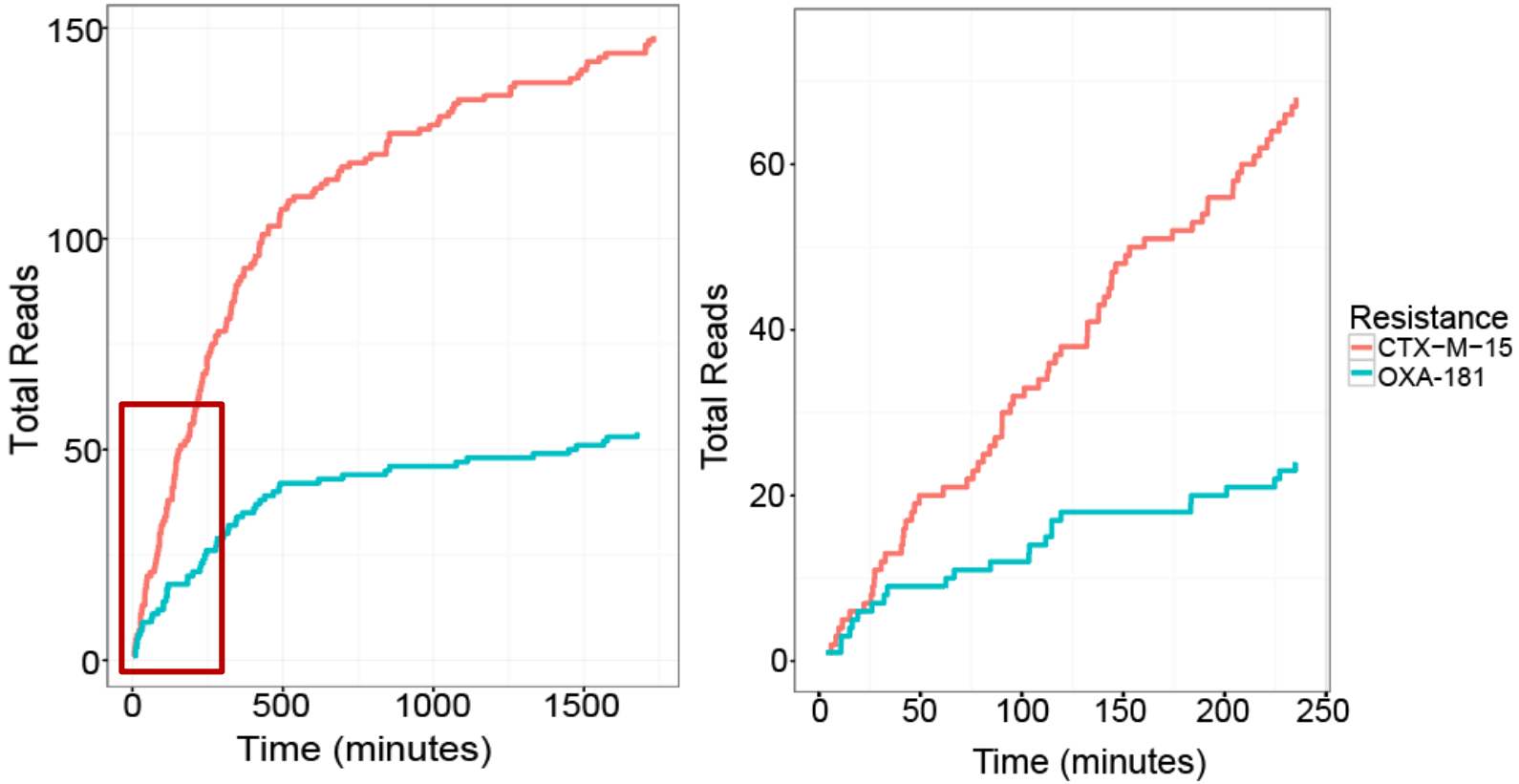
- Annotated assemblies with prokka and CARD databse
- Two specific AMR genes (CTX-M-15 and OXA-181) were discovered and placed within their genetic context (plasmids)

Mobile Elements

- With our complete assemblies we can observe transfer between the plasmids and the chromosome
- Isolate #4, the larger class 1 integron containing *dfrA12*, *aac(6')Ib-cr, rmtF*, *catB7, mrx* and *mphA* was integrated in the chromosome.



Real-time Detection



Sample	OXA-181 detection time	CTX-M-15 detection time
4	4.8 min	2.8 min
8	10.3 min	3.6 min
9	10.73 min	9.37 min
10	0.99 min	1.00 min
12	13.07 min	6.04 min

Resistance identified within 15 minutes!

As reads can be identified as they come off the sequencer, we can identify AMR rapidly
Our retrospective analysis showed the resistance was identified for all of our isolates within 15 minutes

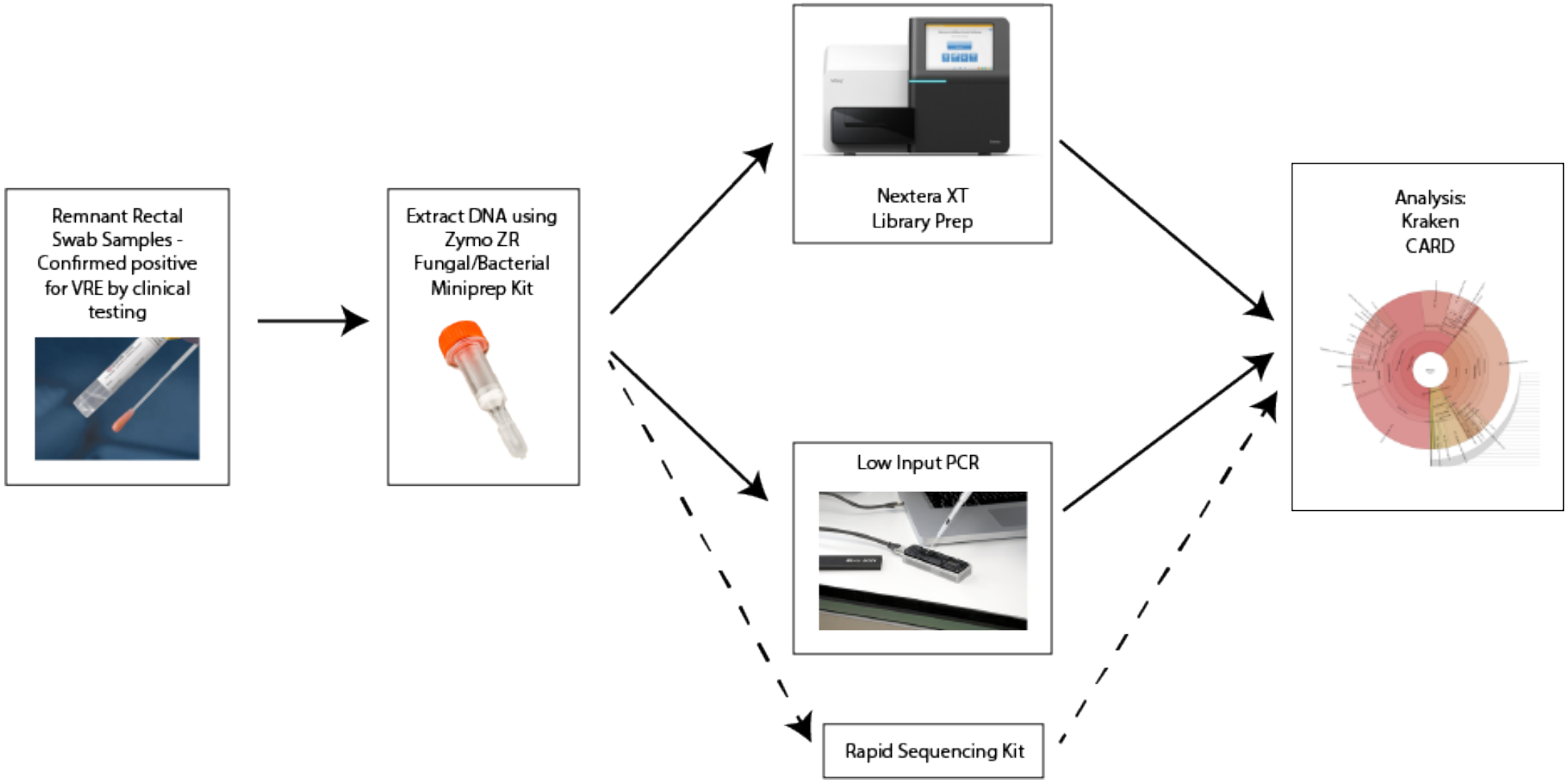
Metagenomic Sequencing: VRE resistance

- In the United States, VRE is commonly acquired in a healthcare setting
- 20000 rectal swabs tested for VRE per year at JHHS
- Major Organisms of Interest:
  - *E. faecalis* (vanB resistance)
  - *E. faecium* (vanA resistance)

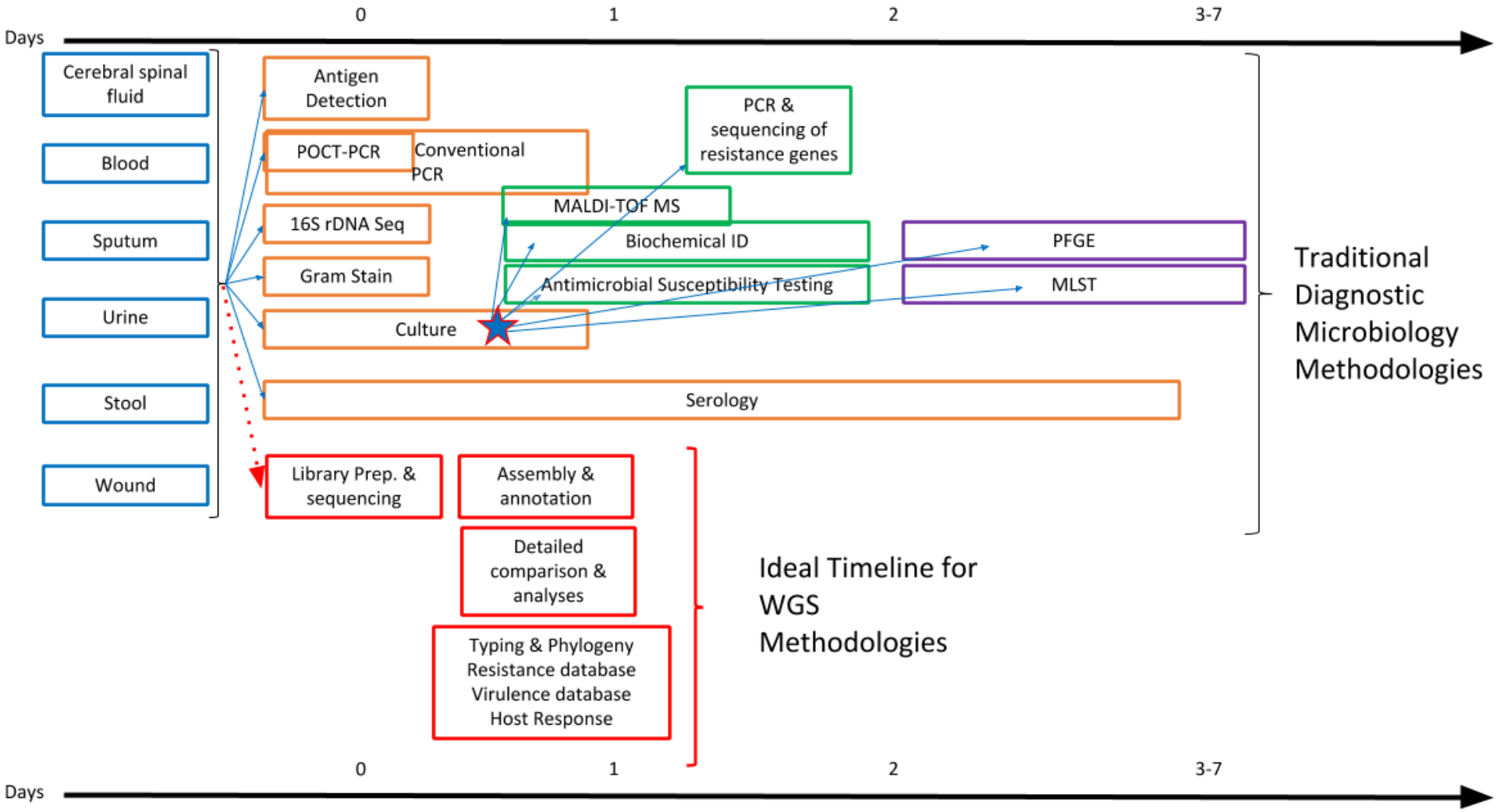
% Enterococci that are Vancomycin Resistant

Species	Europe	US	Canada	Asia-Pacific	Latin-America
E. faecium	**8.8**	**79.4**	**22.4**	**14.1**	**48.1**
E. faecalis	**1.0**	**8.5**	**0.1**	**0.01**	**3.1**
All enterococci	**4.0**	**35.5**	**6.0**	**11.9**	**12.9**

Assay Pipeline



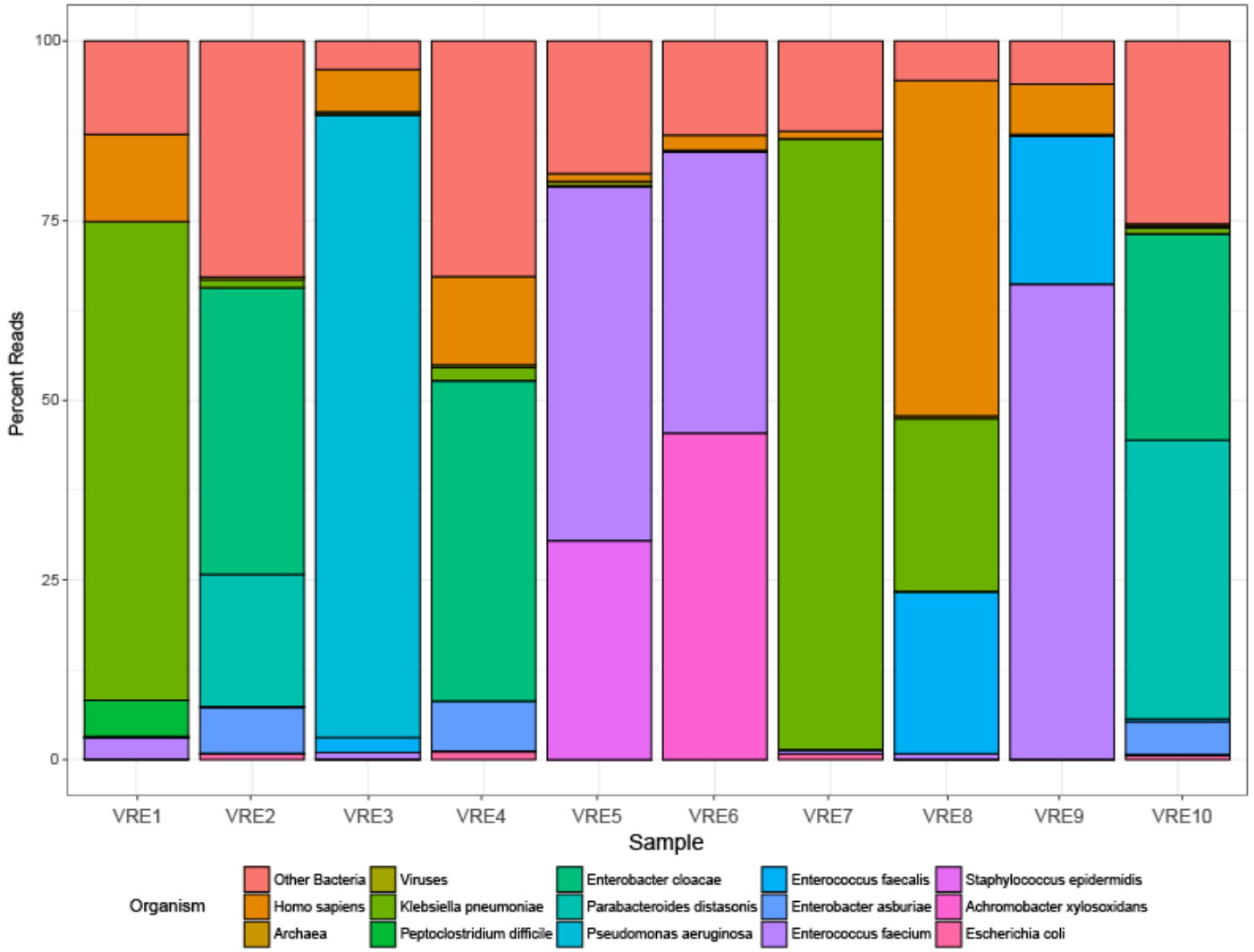
Identification of Pathogenic Organisms



Simner et al forthcoming.

Organism Distribution

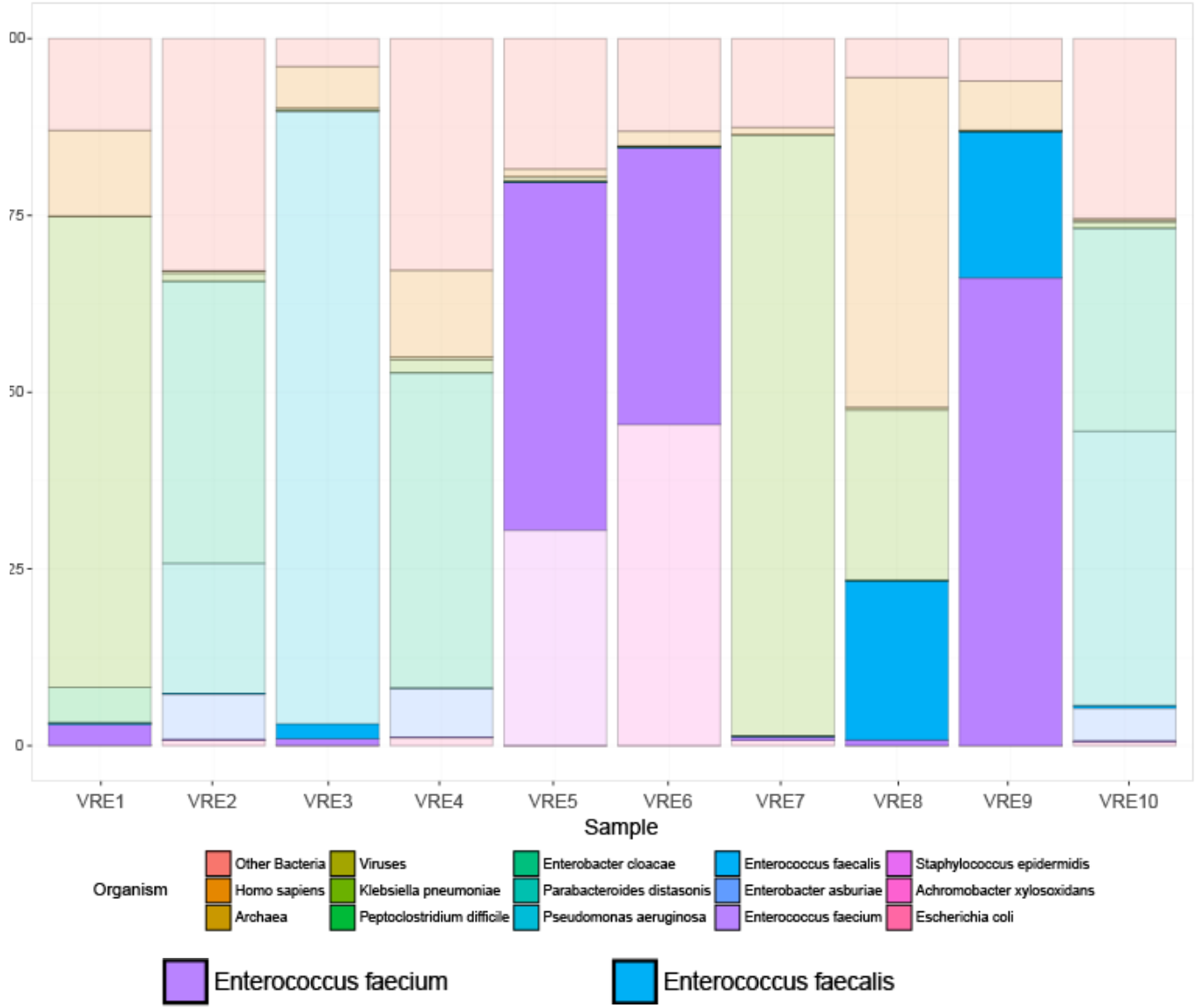
	Results via Culture		Results via Sequencing	
Sample	VRE Organisms	CRO	% VRE	% Human
1	*E. faecium*	No	0.06	8.33
2	*E. faecalis*	No	0.06	0.02
3	*E. faecium*	*P. aeruoginosa*	0.89	5.63
4	*E. faecium*	*E. cloacae*	0.016	7.13
5	*E. faecium*	No	23.6	0.53
6	*E. faecium*	No	13.5	0.75
7	*E. faecium*	*K. pneumoniae*	0.23	0.5
8	*E. faecium*	No	0.7	41.19
9	*E. faecium*	No	23.5	2.52
10	*E. faecalis*	No	0.18	0.14



Organism Distribution - VRE

	Results via Culture		Results via Sequencing	
Sample	VRE Organisms	CRO	% VRE	% Human
1	*E. faecium*	No	0.06	8.33
2	*E. faecalis*	No	0.06	0.02
3	*E. faecium*	*P. aeruoginosa*	0.89	5.63
4	*E. faecium*	*E. cloacae*	0.016	7.13
5	*E. faecium*	No	23.6	0.53
6	*E. faecium*	No	13.5	0.75
7	*E. faecium*	*K. pneumoniae*	0.23	0.5
8	*E. faecium*	No	0.7	41.19
9	*E. faecium*	No	23.5	2.52
10	*E. faecalis*	No	0.18	0.14

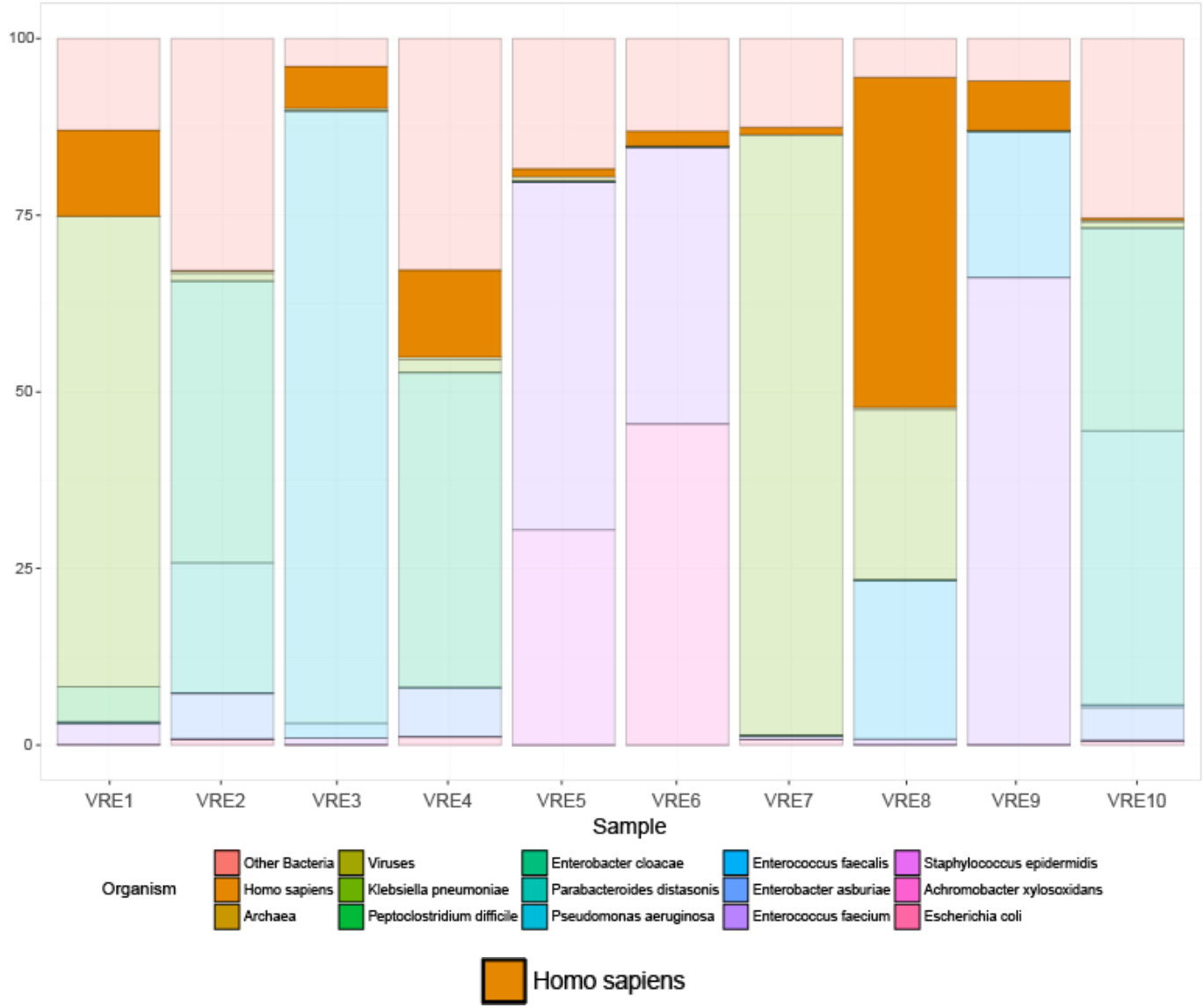
VRE do not generally make up a large part of the microbiome



Organism Distribution - Human

	Results via Culture		Results via Sequencing	
Sample	VRE Organisms	CRO	% VRE	% Human
1	*E. faecium*	No	0.06	8.33
2	*E. faecalis*	No	0.06	0.02
3	*E. faecium*	*P. aeruoginosa*	0.89	5.63
4	*E. faecium*	*E. cloacae*	0.016	7.13
5	*E. faecium*	No	23.6	0.53
6	*E. faecium*	No	13.5	0.75
7	*E. faecium*	*K. pneumoniae*	0.23	0.5
8	*E. faecium*	No	0.7	41.19
9	*E. faecium*	No	23.5	2.52
10	*E. faecalis*	No	0.18	0.14

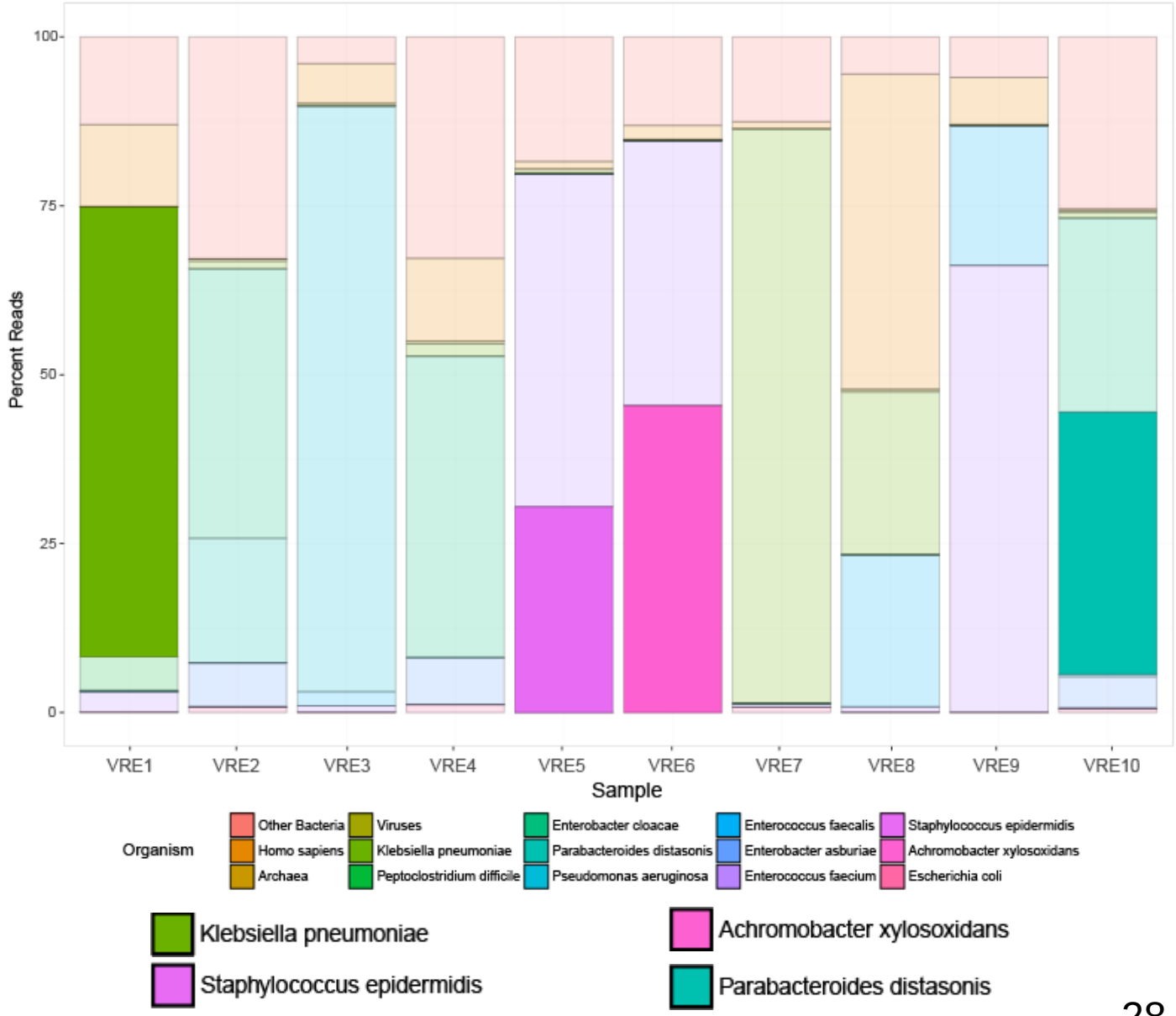
Human contamination levels vary, but were low in these samples.



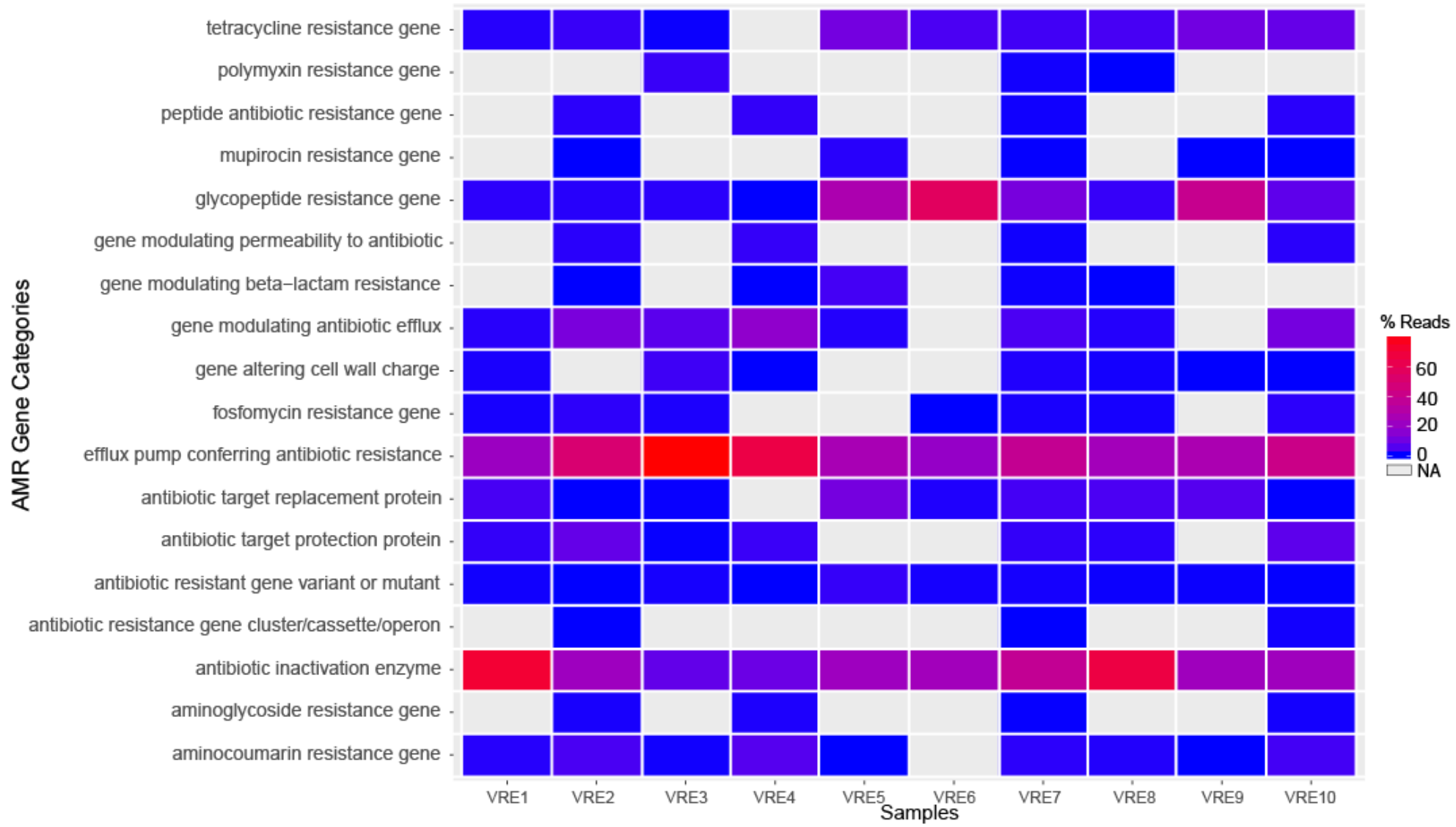
Organism Distribution – Other

	Results via Culture		Results via Sequencing	
Sample	VRE Organisms	CRO	% VRE	% Human
1	*E. faecium*	No	0.06	8.33
2	*E. faecalis*	No	0.06	0.02
3	*E. faecium*	*P. aeruoginosa*	0.89	5.63
4	*E. faecium*	*E. cloacae*	0.016	7.13
5	*E. faecium*	No	23.6	0.53
6	*E. faecium*	No	13.5	0.75
7	*E. faecium*	*K. pneumoniae*	0.23	0.5
8	*E. faecium*	No	0.7	41.19
9	*E. faecium*	No	23.5	2.52
10	*E. faecalis*	No	0.18	0.14

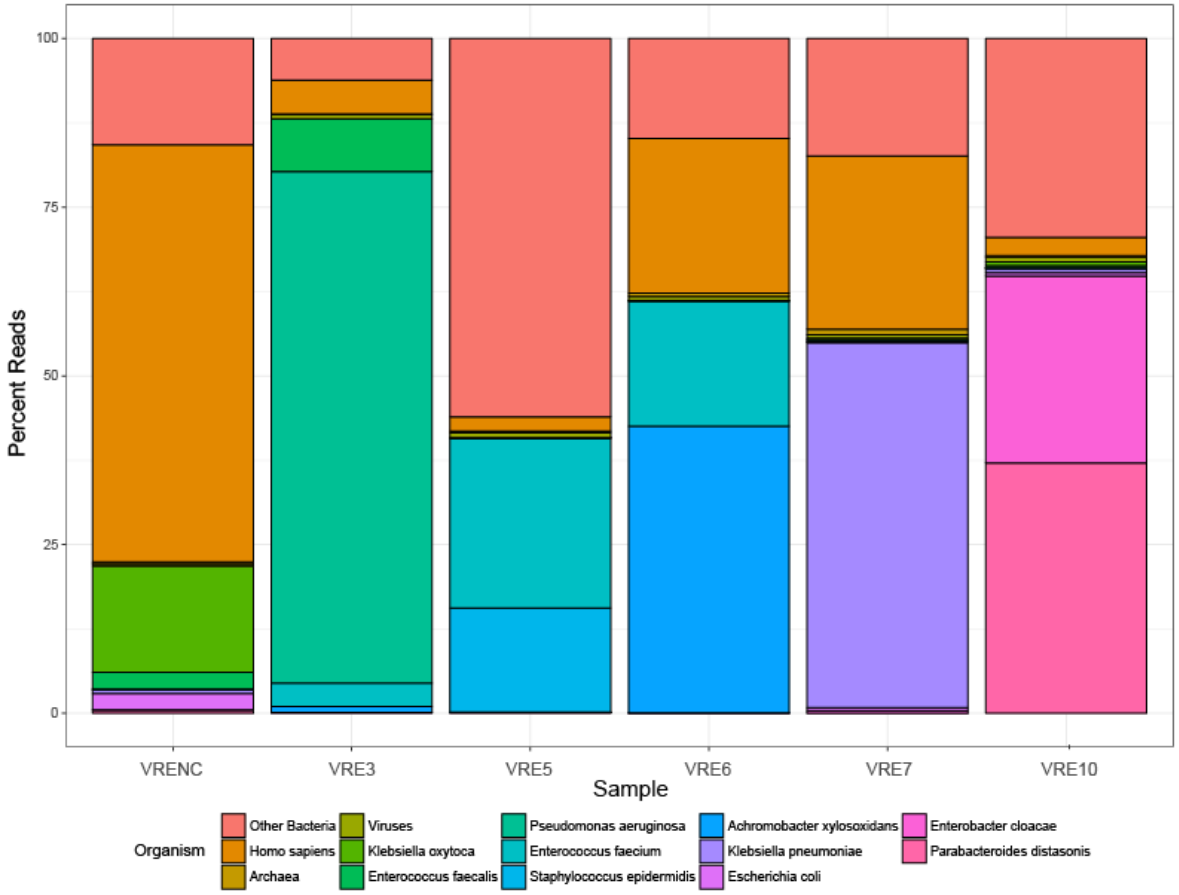
We can get more information using shotgun metagenomics sequencing than just from culture alone

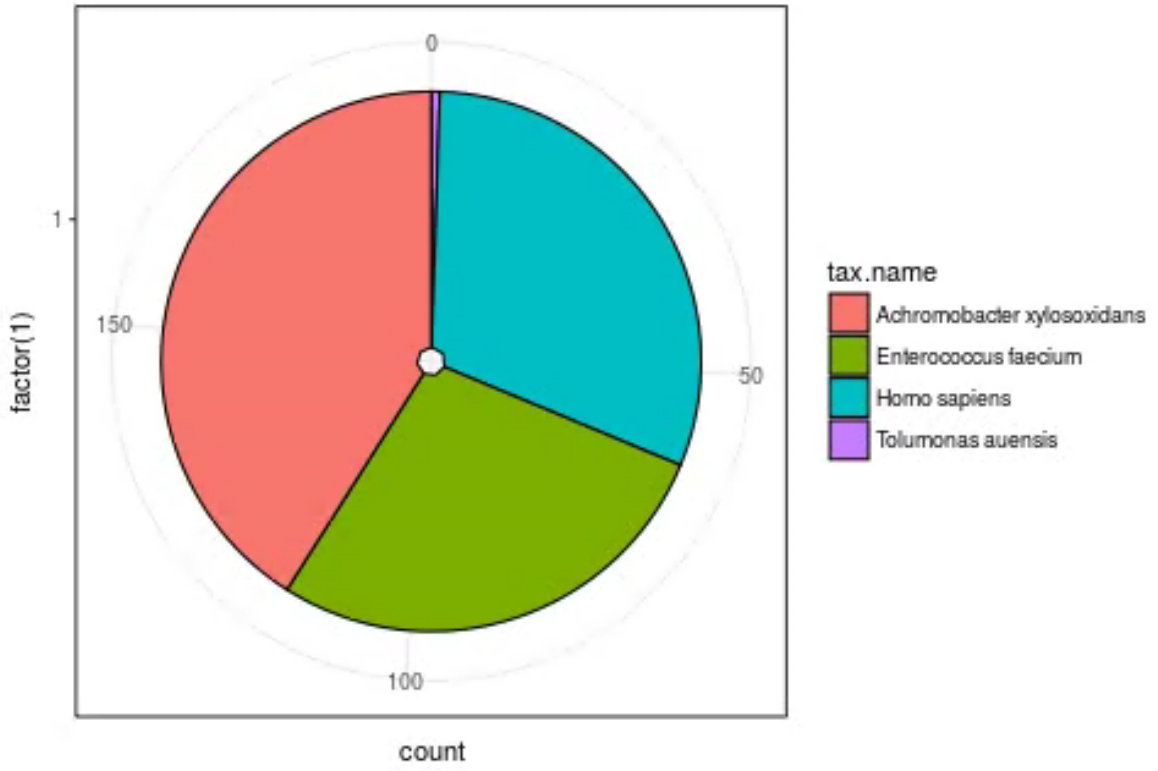


Kraken with customized CARD database



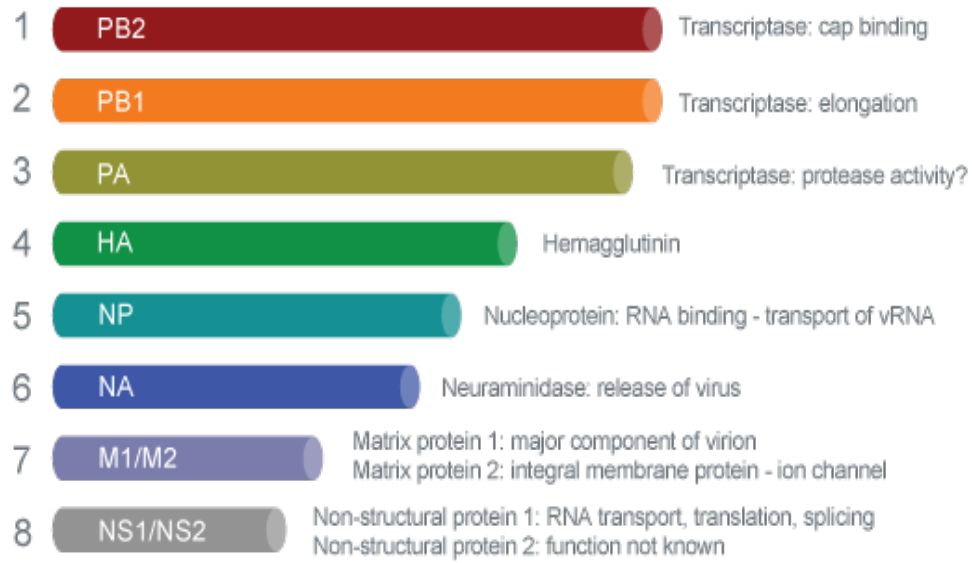
Samples – Nanopore





Influenza Genome

- Influenza – single stranded, helically shaped RNA virus
- Influenza is made of 8 different genomic RNA segments.
- These segments can be amplified by primers which are specific to each segment, but conserved between influenza strains
- Fragments range in size between 900-2.4kb; nanopore would allow for full length sequencing.



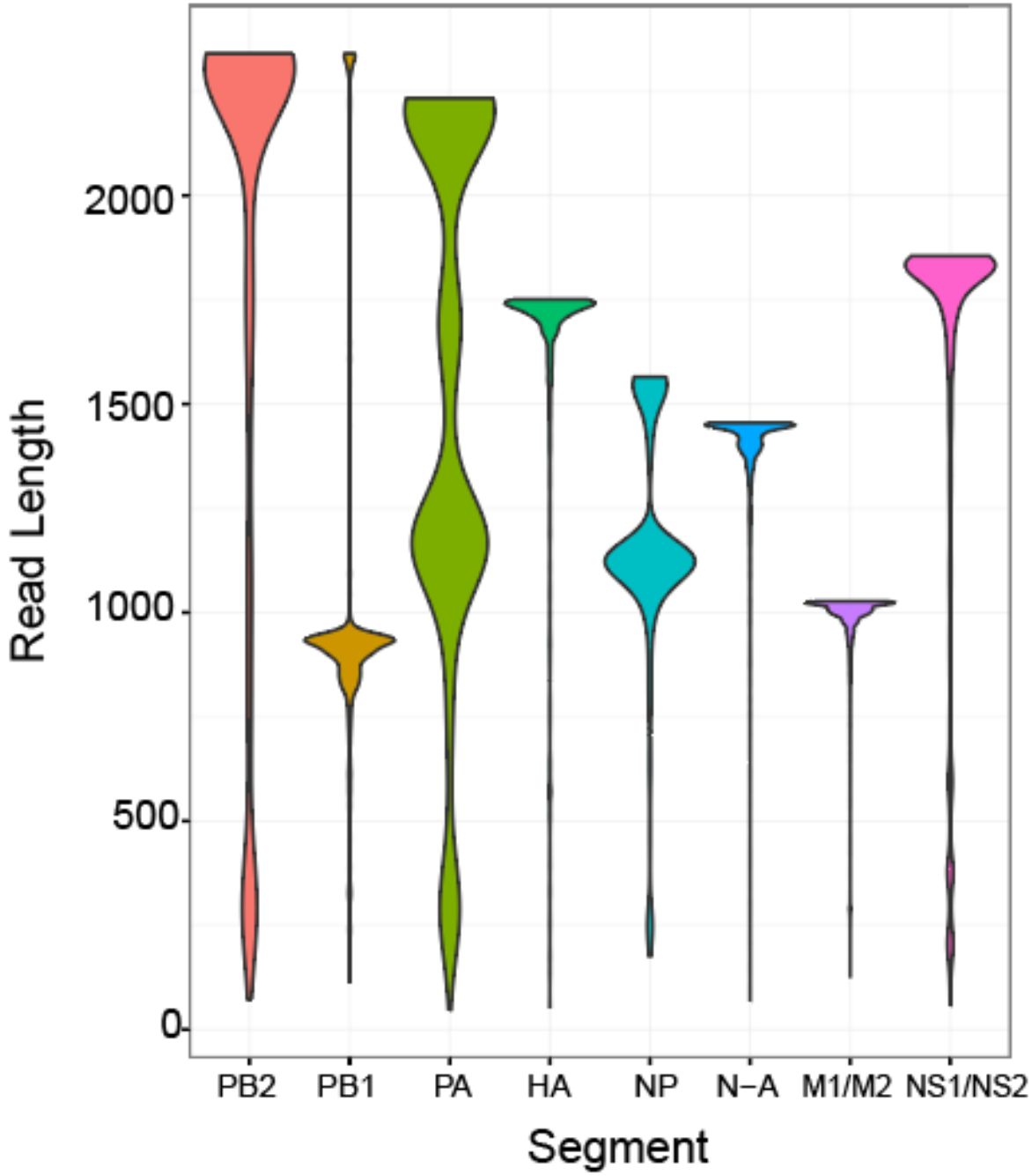
Influenza A genomic RNAs

Flu Sequencing

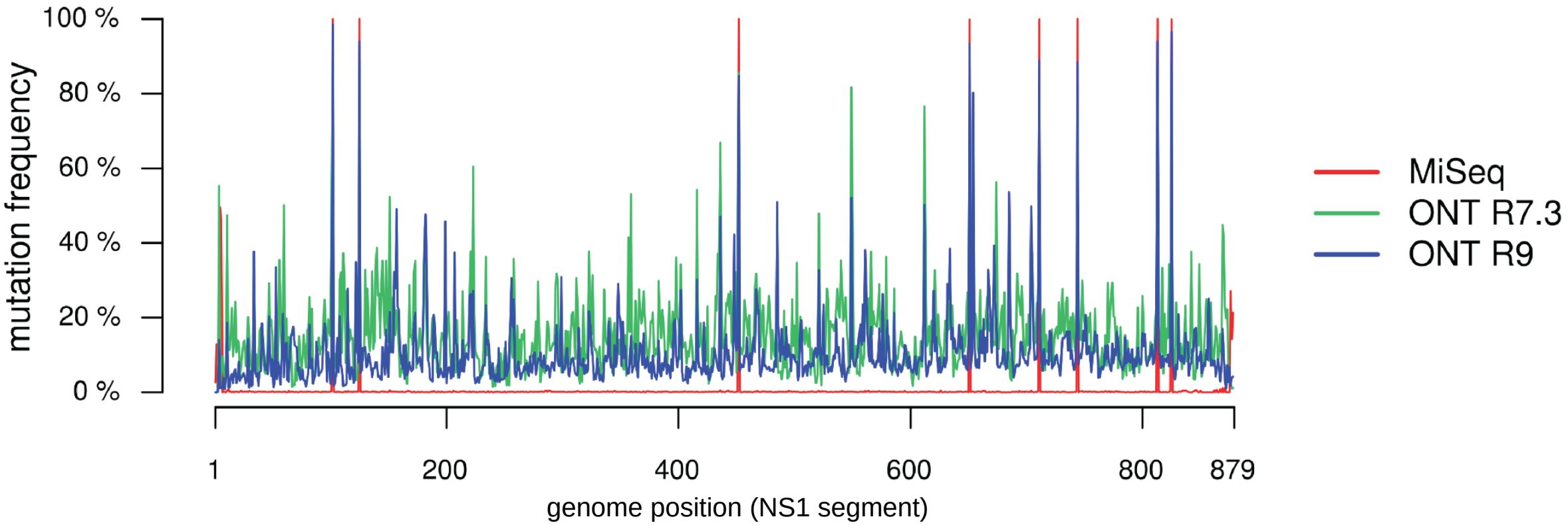
Segment	# of Reads
PB2	622
PB1	1147
PA	4267
HA	2196
NP	1291
N-A	4354
M1/M2	10905
NS1/NS2	1478

Majority of reads are full length

Full length reads don't occur in some segments due to defective interference particles



Mutation Analysis



Mutation frequency: high-depth Illumina data vs. R7 and R9 Oxford chemistry

Older nanopore chemistry (green) was quite error rich, but newer chemistry shows better error rate.

We are going to add nanopolish to see if this can be improved further.

Conclusions/Future Directions

- Sequencing can provide more insight into environmental context of organisms than just culture alone
- Nanopore sequencing could be useful as an aid in providing proper treatment for infectious diseases
  - Get full coverage of pathogenic organisms with one flowcell
  - Rapid time of detection of genes of interest
- Application of nanopolish to improve nanopore-only assemblies
- Application of nanopolish to call methylation in assemblies

Acknowledgments

- **Timp Lab – Johns Hopkins University**
- Winston Timp
- Rachael Workman
- Stephanie Hao
- Yunfan Fan

National Institute of Allergy and Infectious Diseases

1R21AI130608-01 (Simner)

- **Schatz Lab – Johns Hopkins University / CSHL**
- Michael Schatz
- James Gurtowski

- **Simner Lab – Johns Hopkins School of Medicine**
- Patricia (Trish) Simner, PhD
- Belita Opene
- Annie Antar
- Yehudit Bergman

- **CEIRS (Influenza)**
- Rich Rothman
- Lauren Sauer
- Peter Thielen (APL)
- Thomas Mehoke (APL)

RGI-CARD Kraken

Number of reads at each classification level