CAPHRI

CAPHRI Care and Public Health Research Institute

Maastricht University | Maastricht UMC+

Nanopore Sequencing for Rapid VRE Outbreak Detection: Evaluating Data Quality Thresholds

Rebekka Koeck[1], Sébastien Matamoros[2], Erik Beuken[1], Casper Jamin[3], Aldert Bart[4], Paul Savelkoul[1], Lieke van Alphen[1]

1. Department of Medical Microbiology, Infectious diseases and Infection Prevention, Maastricht University Medical Center+, Maastricht, The Netherlands.
2. Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Center, Amsterdam, The Netherlands
3. Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
4. Centraal Bacteriologisch en Serologisch Laboratorium, Tergooi MC Hilversum, The Netherlands

BACKGROUND: Why consider Oxford Nanopore (ONT) sequencing for VRE?

Aim of VRE sequencing:

VRE = Vancomycin-resistant Enterococcus

Detect transmission events

→ implement containment strategies

→ prevent further transmission

Sequencing requirements:

- Short turnaround time (TAT)
- Accurate transmission detection

Current gold standard: Illumina

TAT:

Library preparation: 1.5 days

Sequencing: 72 hours

Minimum: 5 days

Throughput:

12 samples per run

ONT sequencing

TAT:

Library preparation: 1 days

Sequencing: 4-72 hours

Potential minimum: 1.5 days

Throughput:

Flexible sample pooling

METHODS: isolates

ONT sequencing data:

14 VRE isolates:

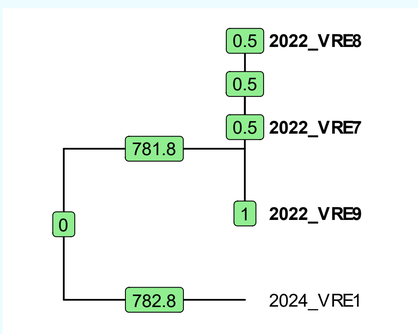
- from 3 Dutch Centres (Maastricht, Amsterdam, Hilversum)
- Clonally related & unrelated isolates

Illumina sequencing data:

4 isolates from Maastricht (MUMC+)

Used to make hybrid assemblies (Unicycler)

MUMC+ isolates - Illumina



3 Clonally-related isolates (Sequence type 787)

1 unrelated isolate (Sequence type 1299)

ONT data processing

Basecalling models: Dorado SUP 4.3.0 / 5.0.0

Minimum read length: 1 Kb / 2 Kb

Minimum average read quality: 10 / 12 / 15

Percentage highest quality base pairs: 80 / 90 / 100

Genome assembly: Flye / Unicycler

+ / - Assembly polishing: Medaka

RESULTS: ONT data processing choices affect assembly similarity

2022_VRE7 assembly similarity

Data pre-processing					Genome assembly					SNP distance vs. hybrid assembly
Basecalling model	Minimum read length	Minimum read quality	Percentage data used	Number of input reads	Assembler	Polishing	Number of contigs	Total genome length	Average depth of coverage	
SUP 5.0.0	2 Kb	10	90%	**51,517**	Unicycler	-	**6**	**3,031,405**	**99 x**	**17**
SUP 5.0.0	1 Kb	10	80%	**57,733**	Unicycler	-	**7**	**3,070,483**	**106 x**	**1**
SUP 5.0.0	1 Kb	10	80%	**57,733**	Unicycler	+	**7**	**3,070,918**	**106 x**	**4**
SUP 5.0.0	1 Kb	15	80%	**51,129**	Unicycler	+	**7**	**3,035,435**	**96 x**	**41**
SUP 5.0.0	1 Kb	15	90%	**72,231**	Unicycler	+	**6**	**3,031,287**	**105 x**	**57**

ONT-only assemblies of **2022_VRE7** differ from the corresponding hybrid assembly by 1 – 57 SNPs

Representative assemblies shown

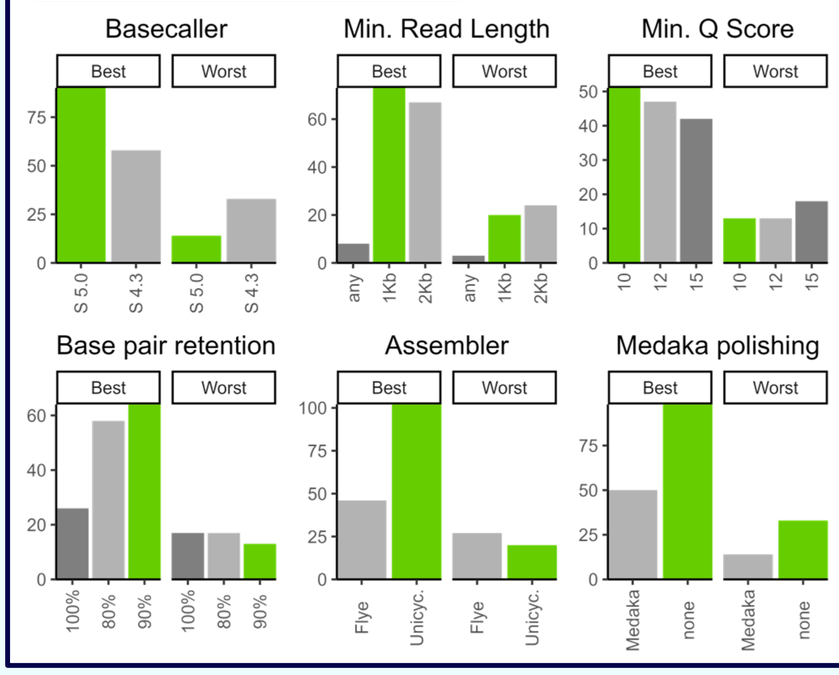
2022_VRE8: 0 – 42 SNPs

2022_VRE9: 0 – 22 SNPs

2024_VRE1: 0 – 17 SNPs

(split K-mer analysis – SKA)

Selecting appropriate parameters



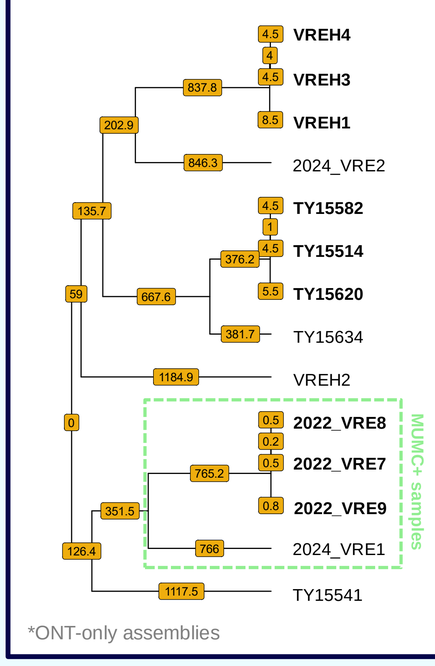
"Best" assemblies (n = 888) differ by <5 SNPs from their corresponding hybrid assemblies

"Worst" assemblies (n = 282) differ by >20 SNPs from their corresponding hybrid assemblies

(or worst 10 assemblies with >10 SNP difference)

Green = selected parameter (most common in "best" assemblies and least common in "worst" assemblies

Applying the selected parameters



*ONT-only assemblies

The assemblies created with the parameters most commonly seen in the best assemblies, accurately reflect the SNP differences and clusters obtained with Illumina sequencing.

Validation with a larger dataset and repeatability experiments are needed.

CONCLUSION: ONT data can replicate Illumina sequencing defined cluster assignment but data processing choices matter!

Correspondence to:

Rebekka Koeck | Email: rebekka.koeck@mumc.nl | www.linkedin.com/in/rebekka-koeck