

BTO 2019.057 | November 2019

BTO rapport

Verkenning van de
toepassingsmogelijkheid
en van MinION
sequencing

BTO

Verkenning van de toepassingsmogelijkheden van MinION sequencing

BTO 2019.057 | November 2019

Opdrachtnummer

402045/041

Projectmanager

Michiel Hootsmans

Opdrachtgever

BTO - Verkennend onderzoek

Kwaliteitsborger(s)

Peer Timmers (gedelegeerd door Paul van der Wielen)

Auteur(s)

Leo Heijnen

Verzonden naar

Dit rapport is verspreid onder BTO-participanten.
Een jaar na publicatie is het openbaar.

Jaar van publicatie
2019

Meer informatie

Leo Heijnen
T +316069743
E leo.heijnen@kwrwater.nl

Keywords: NGS, Next Generation Sequencing, DNA, Microbiology

Postbus 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2019.057 | November 2019 © KWR

Alle rechten voorbehouden.

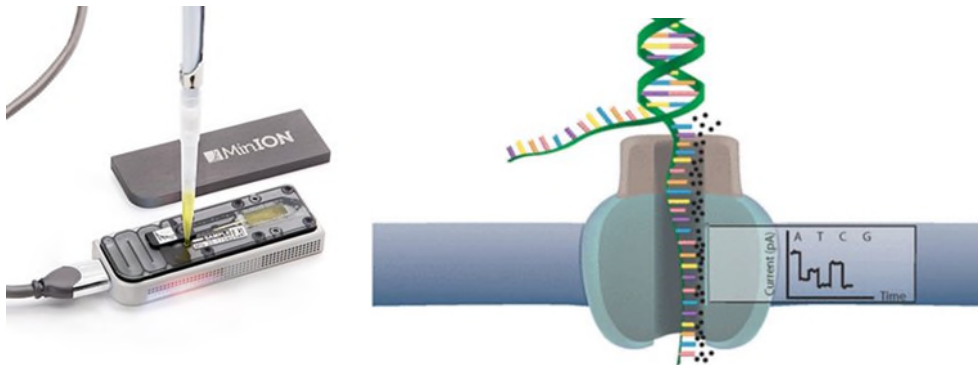
Niets uit deze uitgave mag worden vervoelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

BTO Managementsamenvatting

Resultaten Next Generation Sequencing met MinION nog onvoldoende betrouwbaar; ook toepassen op locatie nog niet mogelijk

Auteur(s) Leo Heijnen

Next Generation Sequencing biedt mogelijkheden om de efficiëntie van zuiveringsprocessen en de kwaliteit van het geproduceerde en gedistribueerde drinkwater te monitoren op microbiologische waterkwaliteit. Een nieuwe, commercieel verkrijgbare technologie die gericht is op toepassing op locatie is getest voor drinkwatertoepassingen. Deze geminiaturiseerde MinION DNA-sequencer, met het formaat van een USB-stick, biedt echter nog te weinig betrouwbare resultaten (hoge *error rate*). Bovendien is het nog niet mogelijk om de benodigde voorbehandeling (filtratie, isolatie en PCR) op locatie uit te voeren. Verdere stappen in de ontwikkeling van de MinION sequencer en de analysesoftware zijn nodig om de interessante mogelijkheden voor toepassing dichterbij te krijgen. Gezien de snelle ontwikkelingen in de technologie biedt de MinION een interessante belofte voor de niet al te verre toekomst.



De MinION sequencer en het principe van het gebruik van nanoporiën voor het bepalen van DNA-sequenties

Belang: snel een beeld vormen van (verandering in) de microbiologische waterkwaliteit

Een snel beeld van microbiologische waterkwaliteitsveranderingen is waardevol bij het aansturen en optimaliseren van zuiveringsprocessen. Next Generation Sequencing (NGS) van (drink)water maakt identificatie van vrijwel de gehele microbiologische bacteriegemeenschap (*microbial profiling*) en van de aanwezige genen (*metagenomics*) mogelijk. Voor NGS is een laboratorium nodig. Er is nu echter ook commercieel een geminiaturiseerde MinION DNA-sequencer verkrijgbaar, met het formaat van een USB-stick, een nieuwe stap in de toepassingsmogelijkheden van NGS

technologieën. Deze DNA-sequencer werkt op basis van nanoporiën, is met de juiste hardware eenvoudig en mogelijk op locatie te gebruiken en maakt via internet gebruik van databases van de leverancier.

Als deze MinION-methode betrouwbare en bruikbare resultaten levert, zou inzet ervan mogelijk sneller een beeld van de microbiologische waterkwaliteit kunnen opleveren.

Aanpak: analyses van praktijk- en kunstmatig samengestelde monsters

NGS is uitgevoerd op praktijkmonsters en op kunstmatig samengestelde mengsels met zowel de MinION als met de standaard

laboratoriumapparatuur (*Illumina*). De resultaten van deze analyses zijn met elkaar vergeleken om de betrouwbaarheid en inzetbaarheid van de MinION te testen.

Resultaten: nog onvoldoende betrouwbaar, niet toepasbaar op locatie

De analyseresultaten maken duidelijk dat het mogelijk is om met MinION informatie te verkrijgen over de samenstelling van de bacteriële gemeenschap en de aanwezigheid van antibioticumresistentiegenen. De kwaliteit van de verkregen data is echter nog onvoldoende betrouwbaar, er is een hoge *error rate*. Bovendien is het niet mogelijk de MinION-analyses op locatie uit te voeren, omdat er voor het uitvoeren van het benodigde voortraject (filtratie, isolatie en PCR) voor watermonsters nog geen goede methoden beschikbaar zijn die snel en *on site* kunnen worden uitgevoerd. Er is meer ontwikkeling van de techniek nodig om

MinION mogelijk in de toekomst wel praktisch te kunnen inzetten.

Toepassing: MinION nog niet geschikt voor de praktijk, maar belofte voor de toekomst

NGS analyses met MinION zijn, vanwege de beperkte betrouwbaarheid (hoge *error rate*), nog niet direct toepasbaar in de praktijk. De ontwikkelingen in de *nanopore sequencing* technologie en de analysesoftware verlopen snel, zodat toepassingsmogelijkheden wel binnen enkele jaren te verwachten zijn.

Rapport

Dit Verkennend (kraamkamer-)onderzoek is beschreven in het rapport *Verkenning van de toepassingsmogelijkheden van MinION sequencing* (BTO 2019.057).

Meer informatie

Leo Heijnen
T 030 60 69 743
E leo.heijnen@kwrwater.nl

KWR

PO Box 1072
3430 BB Nieuwegein
The Netherlands



Inhoud

Inhoud	1
1 Achtergrond	2
2 Materiaal en methoden	4
2.1 Praktijkmonsters	4
2.2 Isolatie van DNA	4
2.3 Next generation sequentieanalyse (NGS)	4
3 Resultaten	13
3.1 Gebruik van MinION: algemeen	13
3.2 Kwaliteit MinION en Illumina MiSeq sequenties	14
3.3 Microbial profiling door sequentieanalyse van 16S RNA amplicons op MinION en Illumina	15
3.4 MinION voor metagenomics analyses	21
3.5 MinION voor het identificeren van antibioticum resistentie genen	31
4 Conclusies en discussie	32
5 Literatuur	36
Bijlage I ABR genen	39

1 Achtergrond

De uitdagingen op het gebied van microbiologische waterkwaliteit liggen (i) op het verbeteren van deze waterkwaliteit bij een deel van de productielocaties en (ii) op het ontwikkelen van efficiënte strategieën voor het monitoren van zuiveringsprocessen en de kwaliteit van het geproduceerde en gedistribueerde drinkwater. De huidige parameters geven vaak slechts beperkt inzicht in het optreden van eventuele kwaliteitsveranderingen. Verbeteren en optimalisatie van meetstrategieën geeft de mogelijkheid om beter en efficiënter inzicht te krijgen in (veranderingen van) de microbiologische waterkwaliteit.

Een belangrijke technologische ontwikkeling op het gebied van nieuwe microbiologische methoden is de toepassing van Next Generation Sequencing (NGS). Met deze technologie is de identificatie van vrijwel de gehele microbiologische bacteriegemeenschap ("Microbial profiling") in het (drink)water mogelijk door sequencing van het 16S rRNA marker gen. Met deze informatie kan inzicht worden verkregen in (i) de microbiologische processen in het drinkwatersysteem en de eventuele (ongewenste) veranderingen die daarbij optreden, van bron tot tap en (ii) veranderingen van waterkwaliteit door externe microbiologische besmettingen (bv fecaal- of bodemmateriaal). Met NGS komen mogelijkheden beschikbaar om beter inzicht te verkrijgen in de zuiveringsprocessen en de drinkwaterkwaliteit. Zoals ook is vastgesteld in de impactstudie "opkomende DNA technieken" (<https://www.btonet.nl/2611/impactstudie-opkomende-dna-technieken.html?qvId=3905474&qvCrc=WWaCjavJvoinbkdb>), zijn veel toepassingen van NGS-technieken te verwachten.

De ontwikkeling van een geminiaturiseerde DNA-sequencer door de firma Oxford-Nanopore (MinION) lijkt te kunnen gaan zorgen voor een nieuwe stap in de toepassingsmogelijkheden van NGS technologieën. Deze DNA-sequencer, met het formaat van een USB-stick, is enige jaren beschikbaar maar toepassingsmogelijkheden zijn nog beperkt. De MinION heeft een sterke ontwikkeling doorgemaakt op het gebied van kwaliteit, "throughput" en leeslengte waardoor mogelijkheden voor sequentie analyse van lange fragmenten en toepassing "on-site" zich lijken aan te dienen. Sequencing van langere DNA-fragmenten maakt het in potentie mogelijk om met "Microbial profiling" de bacteriën in een populatie te identificeren tot op soortniveau. Identificatie tot op soortniveau geeft betrouwbaarder informatie over de potentiële eigenschappen van de aanwezige bacteriën en de biochemische processen waarbij deze bacteriën betrokken zijn. Ook maakt sequencing van langere fragmenten het eenvoudiger om primers voor specifieke detectie van die bacteriesoort met bijvoorbeeld qPCR te ontwikkelen. Daarnaast maakt de hoge throughput en lange leeslengte het wellicht mogelijk om sequencing met MinION te gebruiken om de analyses niet te beperken tot sequencing van het 16S rRNA marker gen maar voor sequencing van willekeurige DNA strengen (Metagenomics). Met deze metagenomics analyses beperkt de analyse zich niet tot microbial profiling, maar worden ook de aanwezige genen (waaronder ook genen betrokken bij antibioticumresistentie) en daarmee de biochemische processen die kunnen plaats vinden geïdentificeerd.

Het doel van dit verkennend onderzoek is:

- Inzicht krijgen in de toepassingsmogelijkheden van MinION sequencing voor drinkwateronderzoek:

- Voor “Microbial Profiling” van bacteriële populaties.
- Voor “Metagenomics” om inzicht te krijgen in:
 - De samenstelling van de aanwezige microbiologische populaties en hun genen in water.
 - Voor het onderzoek naar het voorkomen van antibiotica resistentiegenen (AMR genen).

Om inzicht te krijgen in de toepassingsmogelijkheden van MinION sequencing is er gekozen om dit onderzoek te beperken tot het gebruik van de laboratoriumprotocollen van Oxford-Nanopore en het gebruik van de bioinformatica software die Oxford-Nanopore voor deze analyses aanbiedt.

2 Materiaal en methoden

2.1 Praktijkmonsters

In dit onderzoek zijn analyses uitgevoerd op verschillende monsters uit de waterketen, de monsters zijn weergegeven in Tabel 1. De monsters afkomstig van RWZI "Nieuwe Wetering" zijn verzameld in het kader van TKI onderzoek "DNA fingerprinting in afvalwater- en oppervlaktewaterbeheer"

(<https://www.tkiwatertechnologie.nl/project/dna-fingerprinting-in-afvalwater-en-oppervlaktewaterbeheer/>).

Tabel 1. Beschrijving van monsters gebruikt in deze studie

	Monster	Datum monstername	Volume Voor DNA isolatie	Concentratie DNA (ng/μl)
1	Drinkwater Nieuwegein	23-05-2018	1000 ml	1,95
3	Oppervlaktewater (Lekkanaal Nieuwegein)	06-03-2018	300 ml	0,97
4	Rioolinfluent (RWZI "Nieuwe Wetering")	06-03-2018	50 ml	9,08
5	Riooleffluent (RWZI "Nieuwe Wetering")	06-03-2018	200 ml	59,2
6	Mock community (Bacteriemengsel)			<0,01

2.2 Isolatie van DNA

Voor het isoleren van DNA uit de watermonsters is een volume (zie Tabel 1) water gefiltreerd over polycarbonaat membraanfilters met een poriegrootte van 0,2 μm. DNA is geïsoleerd door gebruik te maken van de DNeasy Power Biofilm Kit van de firma Qiagen. Bij het gebruik van deze kit wordt het DNA eerst uit de cellen vrijgemaakt door een combinatie van mechanische (bead-beating) en chemische lysis van de cellen. Vervolgens wordt het DNA gezuiverd m.b.v. een affiniteitszuivering in spinfilters resulterend in een DNA-suspensie met een volume van 100 μl. De concentratie van het gezuiverde DNA is vervolgens gemeten met de Qubit fluorometer na fluorescente kleuring met Qubit dsDNA HS Assay Kit (beide Fisher Scientific). De gemeten concentraties zijn weergegeven in Tabel 1.

2.3 Next generation sequentieanalyse (NGS)

In dit onderzoek zijn NGS analyses uitgevoerd met de MinION sequencer van Oxford Nanopore en de MiSeq van Illumina. Met beide systemen is het mogelijk om de samenstelling (volgorde van de DNA-bouwstenen: A, C, G en T) van een enorm aantal DNA-moleculen in korte tijd te bepalen. De sequencers van Illumina zijn momenteel de meest gebruikte sequencers waarbij de MiSeq het meest wordt toegepast door onderzoekslaboratoria. Bedrijven die sequencing services op commerciële basis aanbieden (zoals Baseclear in Leiden) maakt gebruik van sequencers met een grotere capaciteit (zoals HiSeq of NovaSeq). KWR heeft de beschikking over de Illumina MiSeq. De MinION sequencer van Oxford Nanopore is een relatief nieuwe, veelbelovende en relatief goedkope (€1000 tegen ca. € 70.000 voor een Illumina MiSeq) sequencer ter grootte van een USB stick. In Figuur 1 zijn de beide systemen weergegeven.

Figuur 1. MinION en Illumina MiSeq sequencers



De technologie waarmee beide sequencers de samenstelling van DNA moleculen bepalen ("sequencen") is compleet verschillend en wordt in de paragrafen 2.3.1 en 2.3.2 kort uitgelegd.

Tabel 2. Eigenschappen van MinION en Illumina MiSeq (informatie van de leveranciers)

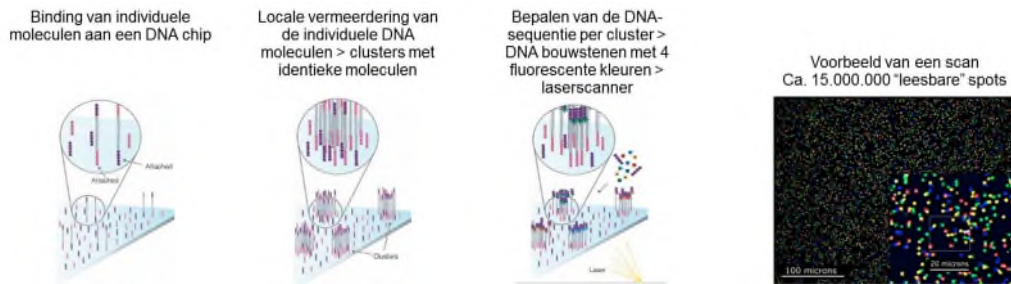
	MinION	Illumina MiSeq
Kosten sequencer (k€)	1	70
Kosten per run (per flowcell, €)	900	1200
Leeslengte	> 1X10 ⁶ bp*	2X300 bp
Maximale hoeveelheid sequentiedata (per run)	30 GB	15GB
Data quality	ca. 10% "errors"	<0,01% errors

* De leeslengte bij MinION sequencing is afhankelijk van de lengte van de aangeboden DNA-moleculen. Omdat elke isolatiemethode zorgt voor het, door mechanisch krachten, aanbrengen van breuken in de DNA-moleculen (shearing) is deze lengte sterke mate afhankelijk van de toegepaste methode voor isolatie van DNA.

2.3.1 Illumina sequencing

Bij de Illumina MiSeq worden de individuele DNA fragmenten uit een te analyseren monster eerst voorzien van synthetische DNA-moleculen aan het uiteinde van de fragmenten (adapters). Vervolgens worden deze adaptersequenties gebruikt om de individuele moleculen te binden aan synthetische DNA-moleculen welke zijn gekoppeld aan een vaste drager in een flowcell (DNA chip). Aansluitend worden reagentia door de flowcell geleid voor het uitvoeren van een PCR (Polymerase Chain Reaction) waarbij de flowcell wordt onderworpen aan wisselingen van temperatuur. Deze PCR-reactie zorgt lokaal voor vermeerdering van de individuele, aan de flowcell gebonden, DNA-moleculen waardoor clusters met identieke DNA-moleculen ontstaan (Figuur 2).

Figuur 2. Principe van Illumina sequencing



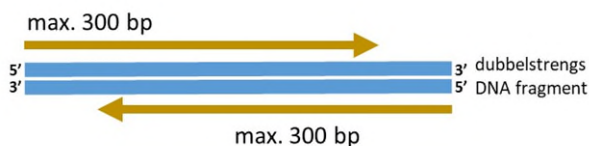
Het bepalen van de DNA-sequenties van de geclusterde DNA-moleculen vindt plaats door het uitvoeren van cyclische reacties waarbij tijdens elke cyclus een mengsel van nucleotiden wordt aangeboden waarbij elk nucleotide (A, C, G of T) een fluorescente kleurstof bevat. Tijdens elke cyclus wordt 1 nucleotide van elk uniek cluster van DNA moleculen gekopieerd door toevoeging van de fluorescent nucleotides (A, C, G of T). Aan het einde van iedere cyclus wordt een foto gemaakt van alle clusters van DNA moleculen op de flowcell, zodat zichtbaar is welk nucleotide is ingebouwd. Voor Illumina MiSeq zijn dat ongeveer 15.000.000 locaties.

Deze cycli worden herhaald totdat de sequentie van max. 300 nucleotiden van de ene streng is bepaald. Vervolgens worden max. 300 bp van de andere streng van het dubbelstrengs DNA molecuul bepaald (Figuur 3).

Op deze manier wordt, door dit cyclische proces maximaal 600X te herhalen, de sequentie van maximaal 2X300 nucleotiden van de individueel gebonden DNA-moleculen bepaald.

N.B. Vanwege het relatief grote aantal fouten met Illumina kits waarmee de sequentie van 2X300 nucleotiden wordt bepaald (<http://blog.mothur.org/2014/09/11/Why-such-a-large-distance-matrix/>) is er bij KWR voor gekozen om kits toe te passen waarmee 2X250 nucleotiden worden bepaald.

Figuur 3. Schematische weergave van sequentiale analyse van 2X300 nt m.b.v. Illumina sequencing.



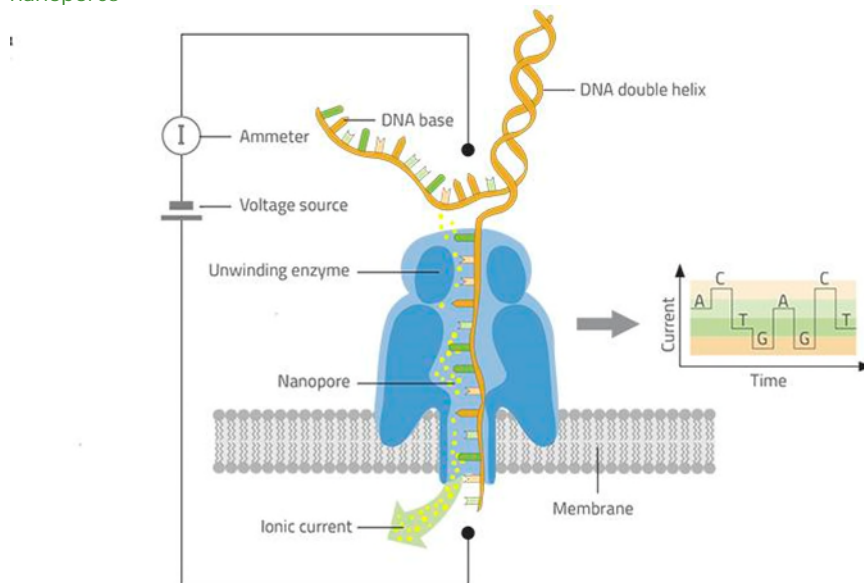
2.3.2 Nanopore sequencing

Bij nanopore sequencing wordt gebruik gemaakt van kunstmatig gesynthetiseerde eiwitporiën in een polymeermembraan. Tijdens het proces van DNA-sequencing wordt één streng van een individueel dubbelstrengs DNA molecuul, onder invloed van een eiwit ("Motor Protein") met een constante snelheid door één van de poriën geleid (Figuur 4). Op het moment dat de individuele nucleotiden (A, C, G of T) van het DNA molecuul de porie passeren, zorgt deze passage voor een verstoring van de ionenstroom (die onder invloed van elektrische spanning door de poriën plaats vindt). De mate van verstoring van de ionenstroom is afhankelijk van de nucleotiden die zich in de porie bevindt. Door meting van de verstoringen van de ionenstroom kan de sequentie van het passerende DNA-molecuul worden bepaald (animatie:

<https://vimeo.com/337258910>). De aanwezigheid van 2048 poriën in de membraan maakt het mogelijk om de sequentie van 2048 moleculen gelijktijdig te bepalen met een snelheid van ca. 400 bp/sec. Zodra een molecuul de porie volledig is gepasseerd en de sequentie ervan is bepaald kan de porie worden gebruikt voor een volgend DNA-

molecuul. De lengte van de te sequencen DNA moleculen kent bij het toepassen van deze technologie vooralsnog geen beperking. Deze lengte wordt alleen beperkt door de mogelijkheid voor het isoleren van zeer lange DNA-moleculen aangezien alle methoden voor het isoleren van DNA in meerdere of mindere mate (afhankelijk van de toegepaste methode) zorgen voor de introductie van breuken in de DNA strengen t.g.v. mechanische krachten (shearing). De verbeteringen van de kwaliteit van Nanopore zijn vooral gelegen in de ontwikkeling van verbeterde membranen met verbeterde en stabielere poriën. In dit onderzoek zijn versie 9 flowcells gebruikt.

Figuur 4. Schematische weergave van het proces voor het bepalen de DNA-sequenties m.b.v. nanopores



Oxford nanopore heeft een aantal verschillende sequencers op "de markt" waarbij deze vooral verschillen in het aantal flowcells dat gelijktijdig kan worden gebruikt voor een analyse. Door het gelijktijdig gebruik van meerdere flowcells is het mogelijk om meer data per monster te genereren of een groter aantal monsters gelijktijdig te analyseren. Met het MinION systeem is het mogelijk om één flowcell, met GridION vijf en met PromethION maximaal 48 flowcells gelijktijdig te gebruiken (Figuur 5).

Figuur 5. MinION, GridION en PromethION van Oxford Nanopore

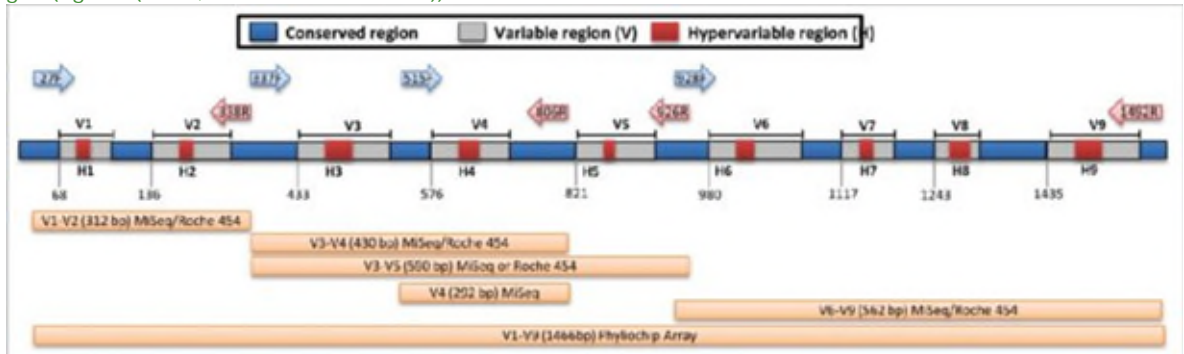


2.3.3 Microbial profiling door sequentie analyse van 16S rRNA genen

Het 16S rRNA gen is aanwezig in alle prokaryoten (bacteriën en archaea) waarbij delen van de sequentie van dit gen bij alle prokaryote groepen (vrijwel) identiek (geconserveerde regio's) zijn, terwijl de sequentie van andere delen van dit gen juist sterk varieert tussen verschillende groepen (variabele regio's). De DNA-sequentie van deze variabele regio's wordt gebruikt voor het identificeren van de aanwezige bacteriën

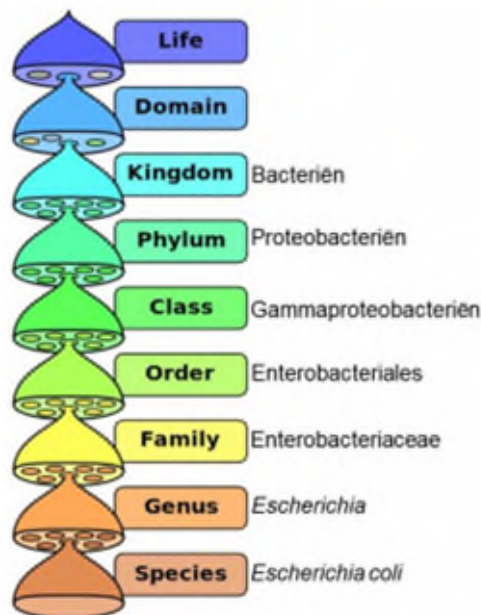
en archaea. Het gebruik van PCR maakt het mogelijk om een specifiek deel van het 16S rRNA gen te vermeerderen. Voor de PCR wordt gebruik gemaakt van synthetische DNA-moleculen (primers) die specifiek binden aan geconserveerde sequenties van het 16S rRNA maar flankeren aan een variabele regio. Er zijn veel verschillende primercombinaties in omloop waarmee amplificatie van fragmenten met verschillende lengte van verschillende regio's van het 16S rRNA gen mogelijk is (Figuur 6).

Figuur 6. Veel gebruikte primers voor vermenigvuldiging van variabele regio's van het 16S rRNA gen (figuur: (Shahi, Freedman et al. 2017)).



Met NGS is het vervolgens mogelijk om de DNA-sequentie te bepalen van een mengsel van een groot aantal (max. ca. 15.000.000 in één MiSeq run) individuele PCR-fragmenten (amplicons) van het 16S rRNA gen. De DNA sequenties van de amplicons wordt m.b.v. software vergeleken met sequenties van 16S rRNA genen uit databases en op basis van deze vergelijking ingedeeld in groepen op verschillende taxonomische niveaus. Een voorbeeld van de taxonomische indeling voor de bacteriesoort *Escherichia coli* is weergegeven in Figuur 7. Doordat met Illumina sequencing slechts een deel van de DNA-sequentie van het 16S rRNA gen wordt bepaald (ca. 300 bp), is het vaak niet mogelijk om de sequenties tot op het niveau van "species" (soort) te identificeren. Identificatie van een groot deel van de sequenties is mogelijk tot op het niveau van familie en een kleiner deel tot op het niveau van genus (geslacht). Wanneer men het gehele 16S rRNA gen (ca. 1400 bp) sequenced is het mogelijk om veel van de sequenties tot op speciesniveau te identificeren.

Figuur 7. Taxonomische rangen met als voorbeeld de taxonomische indeling van *E. coli*



2.3.3.1 Gedeeltelijke sequencing van 16S rRNA genen met Illumina MiSeq

Voor sequentieanalyse van een deel van het 16S rRNA gen met Illumina sequencing is gebruik gemaakt van primers (515F en 806R, zie Figuur 6) waarmee de variabele regio 'v4' wordt geamplificeerd (Caporaso, Lauber et al. 2011). Voor deze amplificatie worden langere primers gebruikt welke zijn voorzien adapters voor binding aan de Illumina flowcell (paragraaf 2.3.3.1). In een tweede PCR stap worden de amplicons voorzien van monster-specifieke index sequenties (barcodes). Deze barcodes zijn specifiek voor elk monster waardoor het mogelijk wordt om in één analyse meerdere monsters te sequencen. Scheiding van monsterdata vindt plaats op basis van de barcode. De stappen voor het verkrijgen van de amplicons en het genereren van de sequenties zijn uitgevoerd zoals voorgeschreven door Illumina (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf).

Processing van ruwe data

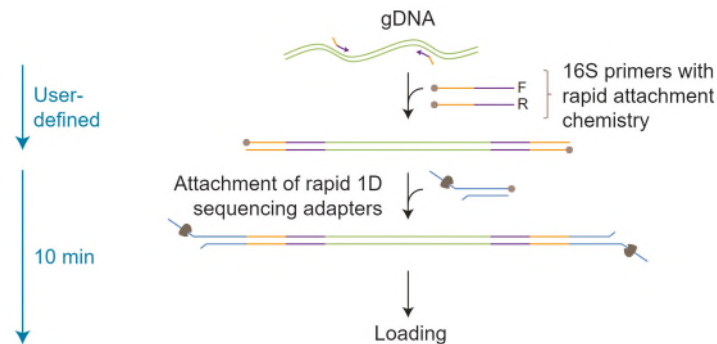
De software van Illumina is gebruikt voor het identificeren scheiding van de verschillende monsters op basis van deze barcodes. Het programma MOTHUR (Schloss, Westcott et al. 2009) is gebruikt voor het assembleren van de paired end sequenties (Figuur 3), het verwijderen van artefact sequenties (zoals chimere sequenties) en het clusteren van sequenties op basis van de mate van sequentieovereenkomst tot "Operational Taxonomic Units" (OTU's op basis van 97% sequentieovereenkomst). Het taxonomisch classificeren van de OTU's is gedaan door de sequenties te vergelijken met de 16S rRNA database van Silva (Quast, Pruesse et al. 2013, Yilmaz, Parfrey et al. 2014).

2.3.3.2 Sequencing van "full length" 16S rRNA genen met MinION

Voor het analyseren van de gehele 16S rRNA genen met de MinION sequencer is gebruik gemaakt van de 16S rapid barcoding kit van Oxford Nanopore. Alle stappen zijn uitgevoerd zoals beschreven door Oxford Nanopore. Bij het gebruik van deze kit worden de primers 27F (AGAGTTTGATCMTGGCTCAG) en 1492R

(CGGTTACCTTGTTACGACTT) gebruikt voor het amplificeren van vrijwel de volledige sequentie van het 16S rRNA gen m.b.v. PCR (Figuur 6). Om gelijktijdige analyse van meerdere monsters mogelijk te maken wordt ook bij Nanopore sequencing gebruik gemaakt van barcode sequenties. Een schematische weergave van deze PCR is weergegeven in Figuur 8.

Figuur 8. Schematische weergave van 16S sequencing met de 16S rapid barcoding kit van Oxford Nanopore.



Na het opbrengen van het monster op de flowcel is een runtijd van ca. 11 uur toegepast.

Processing van ruwe data

De online software van Oxford Nanopore is gebruikt voor het scheiden van de verschillende monsters op basis van de barcode sequenties en het taxonomisch identificeren van de sequenties. Voor de taxonomische identificatie worden door de software van Oxford Nanopore de sequenties vergeleken met een online 16S rRNA database van het National Centre for Biotechnology Information (NCBI): <https://www.ncbi.nlm.nih.gov/bioproject?db=bioproject&Cmd=DetailsSearch&Term=3175%5Buid%5D> m.b.v. het programma Blast (Altschul, Gish et al. 1990) Deze database bevat 18,927 16S rRNA gensequences van verschillende organismen. Elke MinION read wordt geklassificeerd op basis van de mate van overeenkomst met referentiesequenties.

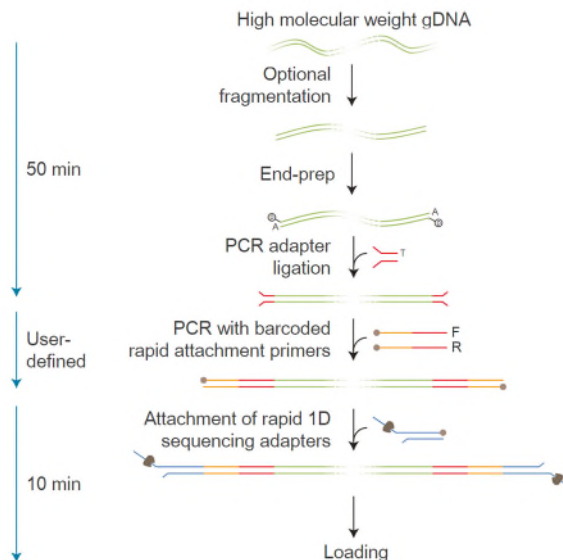
2.3.4 Sequencing van metagenomes met MinION

Bij het sequencen van metagenomes wordt het bepalen van de DNA-sequenties niet beperkt tot het 16S rRNA gen maar vindt de sequencing plaats op willekeurige DNA-fragmenten uit een monster. Hierdoor is het mogelijk om meer sequentie-informatie over de, in een monster aanwezige organismen, te verkrijgen. Dit maakt het mogelijk om, naast taxonomische informatie over de organismen, ook informatie te verkrijgen over de aanwezige genen. Maar, doordat sequentieanalyse zich niet beperkt tot één marker-gen (zoals 16S rRNA), geven metagenomics analyses alleen informatie over de meest dominante organismen uit de populatie.

In principe kunnen metagenomics analyses direct uitgevoerd worden op het DNA uit een monster en is er geen PCR noodzakelijk om daarmee een marker-gen te vermenigvuldigen. Voor het uitvoeren van metagenomics analyses zonder PCR is de beschikbaarheid van voldoende DNA (> 1 µg) uit een monster van belang. Bij het opwerken van volumes zoals zijn weergegeven in Tabel 1 was het niet mogelijk om voldoende DNA te isoleren voor het uitvoeren van metagenomics analyses zonder toepassing van PCR. Daarom is gekozen voor het toepassen van PCR Barcoding Kit

(SQK-PBK004) van Oxford Nanopore waarbij gebruik gemaakt is van het door de leverancier geleverde protocol. In Figuur 9 zijn de stappen uit dit protocol schematische weergegeven.

Figuur 9. Schematische weergave van de stappen uitgevoerd met de PCR Barcoding Kit



Tijdens het isoleren van DNA ontstaan er breuken op willekeurige plaatsen in de lange DNA-strengen en raken de uiteinden beschadigd t.g.v. mechanische krachten tijdens het isolatieproces (shearing), vooral tijdens bead-beating. De uiteinden van het geïsoleerde DNA worden eerst gerepareerd ("End-prep") en vervolgens worden er synthetische dubbelstrengs DNA fragmenten (adapters) aan deze uiteinden gekoppeld via een ligase reactie ("PCR adapter ligation"). Aansluitend worden primers gebruikt die binden aan de adaptersequenties voor het uitvoeren van een PCR waarbij alle fragmenten worden vermeerderd. Om ervoor te zorgen dat de PCR zo min mogelijk invloed heeft op de samenstelling van het DNA ("PCR-bias") worden er bij deze PCR weinig vermeerderingscycli (15 cycli) uitgevoerd. Voor elk monster wordt een aparte set PCR-primers gebruikt waarbij de primers zijn voorzien van barcodes. Voor elk monster worden andere barcodes gebruikt. Door het toepassen van barcodes is het mogelijk om meerdere monsters gelijktijdig te sequencen. Na het sequencing wordt namelijk bepaald welke sequenties bij elk monster horen door gebruik te maken van de barcodes. Ten slotte worden de fragmenten voorzien van sequencing adapters (1D sequencing adapters) welke noodzakelijk zijn voor de Nanopore sequencing reactie. Daarna wordt het monster op de MinION sequencer gebracht. Gedurende een runtijd van ca. 11 uur is sequencing data gegenereerd.

Processing van ruwe data

De online software van Oxford Nanopore is gebruikt voor het scheiden van de verschillende monsters op basis van de index sequenties. Voor taxonomische identificatie van de verkregen sequenties is gebruik gemaakt van het programma WIMP (What's in my Pot, (Juul, Izquierdo et al. 2015)). Dit programma wordt online, via de EPI2ME service van Oxford Nanopore, aangeboden en is gebaseerd op het programma "Centrifuge" (Kim, Song et al. 2016) waarbij de sequenties worden vergeleken met een genomische sequentie database van een groot aantal bacteriën, virussen en eukaryote organismen (zie Tabel 3). Voor het identificeren van antibioticum resistentiegenen in de dataset is gebruik gemaakt van de "FASTQ antimicrobial resistance" module uit de EPI2ME service van Oxford Nanopore. Deze module maakt gebruik van een

sequentievergelijking met sequenties uit de CARD database. Deze CARD database is een veel gebruikte database voor antibioticum resistentiegenen (Jia, Raphenya et al. 2017).

Tabel 3. Inhoud van de database welke door WIMP gebruikt wordt voor identificatie van sequenties

RefSeq Database	Number of different taxa in the database
Bacteria	5083
Viruses	7482
Fungi	249
Archaea	230
Total	13044

3 Resultaten

Om een indruk te krijgen van de mogelijkheden van MinION sequencing zijn in dit onderzoek een aantal analyses hiermee uitgevoerd. Er zijn analyses uitgevoerd om inzicht te krijgen in de toepasbaarheid van MinION voor "Microbial Profiling" (gerichte sequencing van sequenties van het 16S rRNA gen) en "Metagenomics" (sequencing van willekeurige DNA moleculen uit een populatie). Hiervoor zijn analyses uitgevoerd op de volgende monsters:

- Molecular profiling door analyse van 16S rRNA sequenties:
 - Analyse van een samengesteld mengsel van bacteriën (mock community) met MinION en vergelijking van de resultaten met Illumina MiSeq sequencing.
- Metagenomics analyses met MinION:
 - Analyse van een mock community
 - Analyse van DNA uit drinkwater, RWZI-Influent en RWZI effluent

3.1 Gebruik van MinION: algemeen

In 2017 is verkennend onderzoek uitgevoerd naar de toepassingsmogelijkheden van metatranscriptomics analyses voor drinkwateronderzoek (Heijnen 2018). Binnen dit project is ook zeer kort gekeken naar de mogelijkheden van nanopore sequencing voor full length 16S rRNA sequencing. Dit onderzoek maakte toen duidelijk dat het mogelijk was om sequenties te bepalen van het volledige 16S rRNA gen met MinION. De kwaliteit van de sequenties was toen niet voldoende om de bacteriën uit de gemeenschap tot op het soortniveau te bepalen. Bij Oxford Nanopore (de ontwikkelaar van MinION) zijn sindsdien stappen gezet voor verbetering van de membranen die de nanopores bevatten en de software voor de "vertaling" van de elektrische signalen naar sequentie-informatie. Om zo goed mogelijk op de hoogte gebracht te worden van de meest recente technische aspecten is eerst een bezoek gebracht aan Oxford Nanopore voor het volgen van een workshop, vervolgens is gestart met experimenten.

Voor het uitvoeren van analyses met de MinION sequencer zijn er speciale criteria waaraan de computer die de MinION aanstuurt en de verbinding tussen de computer en internet moeten voldoen. Vanwege de grote hoeveelheid data en de snelheid waarmee deze wordt gegenereerd is een snelle computer met een grote solid state disk (SSD) noodzakelijk. Aangezien gegenereerde data via internet naar een server wordt gestuurd en deze server vervolgens bewerkte data weer terugstuurt was het bij KWR tevens noodzakelijk om diverse beveiligingsinstellingen aan te passen. Het optimaliseren van deze instellingen heeft tijdens het uitvoeren van experimenten geleid tot vertragingen en diverse mislukte of afgebroken experimenten t.g.v. ICT problemen.

Gedurende de sequencing wordt door de MinION software (Epi2Me) "live" weergegeven welke nanoporiën ("Channels") van de flowcell actief gebruikt worden voor het bepalen van de sequenties van DNA-moleculen (groene channels in Figuur 10). Gedurende het proces raken de poriën beschadigd of verstopt waardoor het aantal "inactieve" poriën toeneemt.

Figuur 10. Weergave van de actieve poriën in de membraan van de flowcell



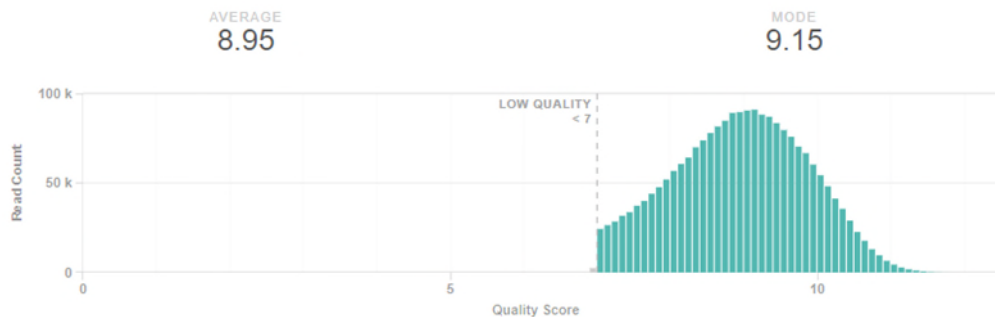
3.2 Kwaliteit MinION en Illumina MiSeq sequenties

Gedurende het proces van het bepalen van de sequenties door sequencers (zoals MinION of MiSeq) wordt een maat voor de kwaliteit van elke gelezen bouwsteen (nucleotide) weergegeven, deze maat voor de kwaliteit wordt PHRED score genoemd (Ewing and Green 1998). Met deze PHRED score wordt de kans dat een nucleotide verkeerd is geïdentificeerd weergegeven waarbij elke nucleotide een eigen PHRED score heeft. De volgende formule wordt toegepast om deze kans te berekenen:

$$P = 10^{-\frac{Q}{10}}$$

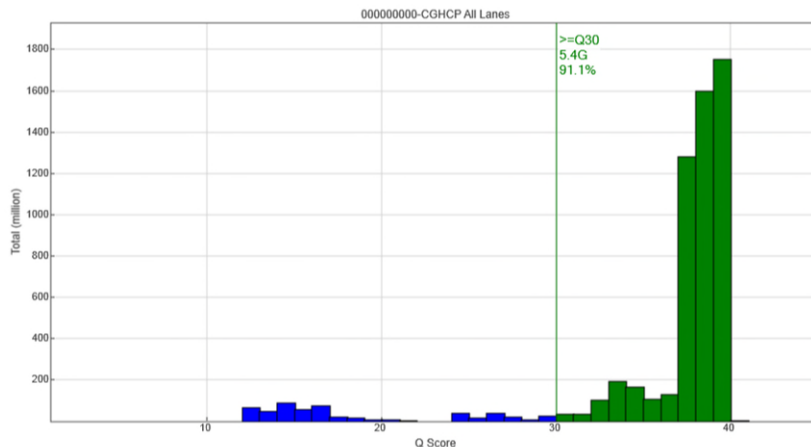
Hierin is P de kans op een fout en Q is de door de sequencer gegenereerde PHRED score. Dit betekent dat er een kans is van 10% op een fout bij een PHRED score van 10 en een kans is van 0,01% op een fout bij een PHRED score van 40.

Figuur 11. Kwaliteitscores MinION



In Figuur 11 is de verdeling van PHRED scores weergegeven van 16S rRNA fragmenten bij sequentieanalyse op de MinION en in Figuur 12 de PHRED scores bij sequentieanalyse op de Illumina MiSeq. De gemiddelde PHRED score van alle nucleotiden op de MinION sequencer is 9,15. Dit wil zeggen dat er voor elke nucleotide een kans van ca. 12% dat deze verkeerd bepaald is. Bij de Illumina MiSeq heeft een groot deel van de nucleotiden een PHRED score dichtbij 40 en heeft 91,1% van de nucleotiden een PHRED score hoger dan 30. Een score van 30 wil zeggen dat er een kans van 0,1% is dat een nucleotide verkeerd is bepaald.

Figuur 12. Kwaliteitsscores Illumina MiSeq



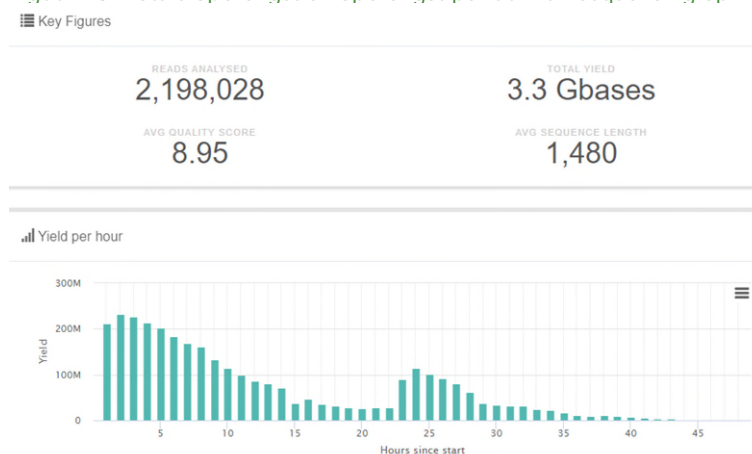
Bovenstaande voorbeelden zijn representatief voor meerdere experimenten en geven aan dat sequenties bepaald op de Illumina Miseq van hogere kwaliteit zijn dan sequenties bepaald op de MinION.

3.3 Microbial profiling door sequentieanalyse van 16S RNA amplicons op MinION en Illumina

Voor het vergelijken van de analyseresultaten van microbial profiling analyses uitgevoerd op MinION en Illumina zijn er analyses met beide systemen uitgevoerd op een mengsel van bekende bacteriesoorten (mock community).

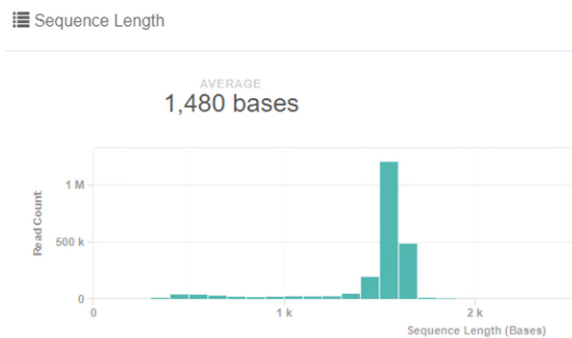
De analyses op de MinION sequencer zijn gericht op het volledige 16S rRNA gen (ca. 1450 bp) en de analyses op de Illumina Miseq op een fragment van ca 300 bp van de v4 regio van het 16S rRNA gen. Bij sequentieanalyse op MinION is de sequentie bepaald van 2.198.028 fragmenten waarmee ca. $3,3 \times 10^9$ nucleotiden aan sequentiedata is gegenereerd (Figuur 13). De hoeveelheid sequentiedata die per uur wordt bepaald neemt gedurende de run sterk af t.g.v. de toename van inactieve poriën in de membraan (Figuur 10).

Figuur 13. Totale opbrengst en opbrengst per uur van sequencing op MinION



Na de sequencing op MinION wordt de lengte van de DNA-strengen waarvan de sequentie is bepaald weergegeven (Figuur 14). Deze lengte van de fragmenten is gemiddeld 1480 bp (bases), dit is in overeenstemming met de gemiddelde lengte van de geamplificeerde 16S rRNA fragmenten (gemiddeld 1465bp) verlengd met de monster-specifieke barcodes.

Figuur 14. Fragmentlengte MinION sequencing



3.3.1 Analyse van de mock community

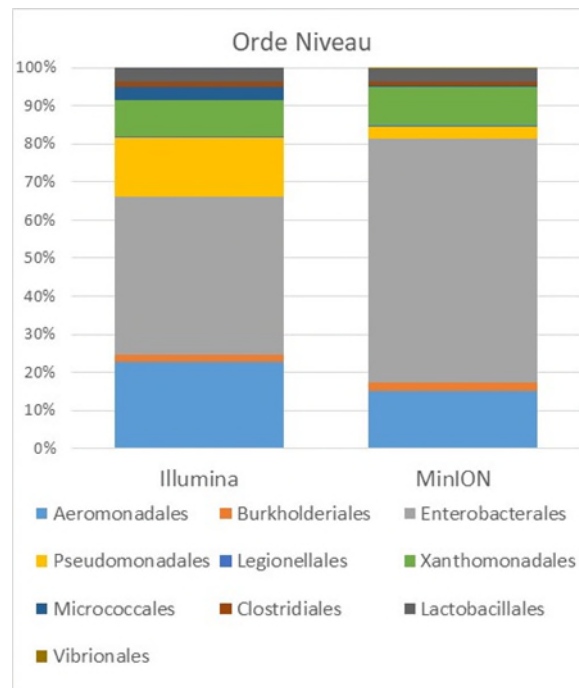
De analyse van de mock community is een herhaling van het experiment uit 2017 (Heijnen 2018). Deze mock community bestaat uit een mengsel van bacteriecellen afkomstig van referentiestammen uit de KWR bacteriedatabase. De bacteriestammen die aanwezig zijn in deze mock community zijn weergegeven in Tabel 4 waar op hetzelfde mengsel van cellen een analyse is uitgevoerd met MinION en Miseq.

Tabel 4. Samenstelling van de mock community

Samenstelling mock community (naamgeving volgens KWR database)
Aeromonas salmonicida
Aeromonas veronii
Burkholderia multivorans
Hermiinimonas glaciei
Citrobacter freundii
Klebsiella pneumoniae
Enterobacter aerogenes
Escherichia coli
Rahnella aquatilis
Acinetobacter radioresistens
Legionella pneumophila
Pseudomonas fluorescens
Stenotrophomonas maltophilia
Staphylococcus warneri
Clostridium sp.
Enterococcus faecium

In 2017 werd vastgesteld dat de resultaten van taxonomische identificatie op orde-niveau na sequencing van 16S rRNA amplicons met Miseq in grote lijnen vergelijkbaar was met taxonomische identificatie na sequencing met MinION. De mogelijkheid om 16S rRNA sequenties tot op soortniveau te kunnen identificeren zou een groot voordeel van sequentieanalyse van het volledige 16S rRNA zijn (zoals dat met MinION mogelijk is). De data van 2017 liet echter zien dat, door de beperkte kwaliteit van MinION data, identificatie tot op soortniveau niet mogelijk was. In dit project is dit experiment herhaald met als doel om te onderzoeken of er tussen 2017 en eind 2018 in de ontwikkeling van MinION stappen zijn gezet waardoor betrouwbare identificatie op soortniveau wel mogelijk is geworden. De vergelijking van de resultaten met MinION en Illumina zijn niet zo uitgebreid uitgevoerd als in het onderzoek van 2018 (Heijnen 2018). In Figuur 15 is een schematische weergave van de samenstelling van deze mock communities op orde niveau weergegeven. De vergelijking van beide analyses laat zien dat beide sequentieanalyse methoden op klasse niveau een redelijk vergelijkbaar resultaat laten zien. De klasse van *Vibrionales* (0,2%) wordt alleen aangetoond bij het gebruik van MinION terwijl er in de mock community geen bacteriën behorende bij deze klasse aanwezig zijn. Alle andere klassen worden met beide methoden waargenomen. Wel zijn er relatieve verschillen in de dominantie van de klassen waarneembaar. Deze verschillen kunnen het gevolg zijn van het vermeerderen van verschillende 16S rRNA fragmenten waarvoor primers worden gebruikt die mogelijk zorgen voor de introductie van bias in de resultaten.

Figuur 15. Resultaten van de analyse van de mock community op "orde niveau" met MinION en Illumina sequencing



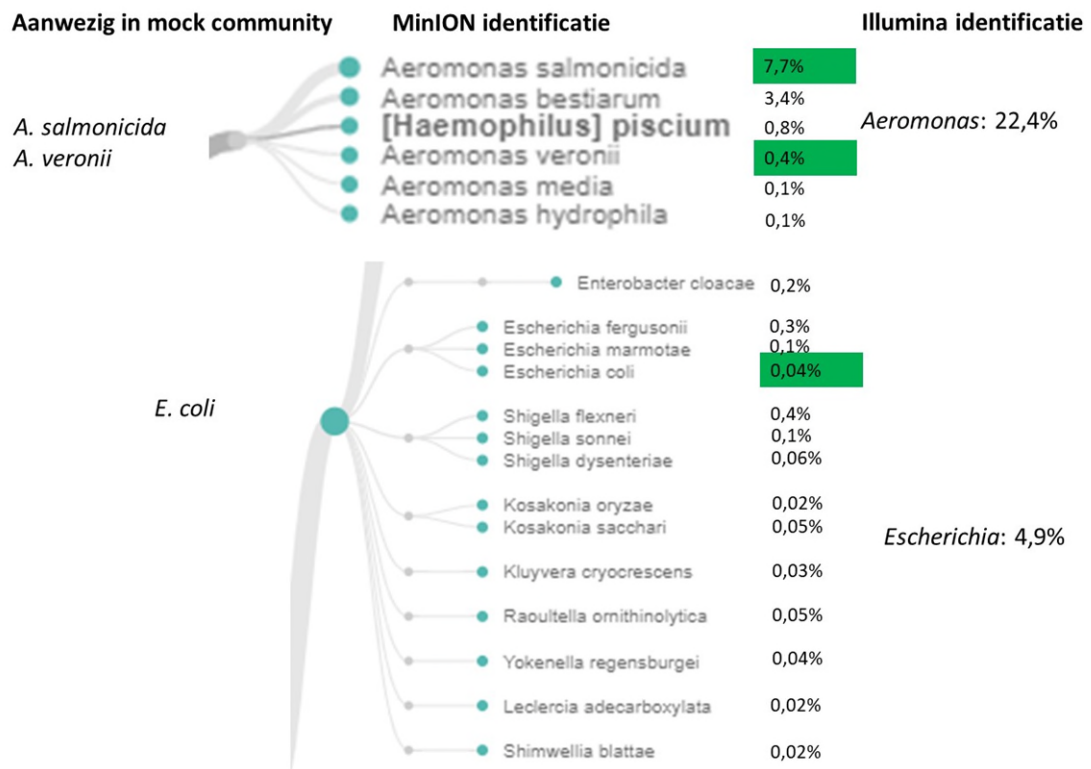
Bij het identificeren van de sequenties tot op genusniveau worden veel van de met Illumina sequencing geïdentificeerde genera ook aangetoond bij het gebruik van MinION. Met Illumina sequencing kan geen onderscheid gemaakt worden tussen een aantal genera welke behoren tot de familie van de *Enterobacteriaceae* (*Citrobacter*, *Klebsiella* en *Enterobacter*) terwijl het wel mogelijk is om deze te onderscheiden met MinION. Met Illumina sequencing wordt 99,4% van de sequenties ook inderdaad geïdentificeerd als sequenties afkomstig van de genera zoals die aanwezig zijn in de mock community. Bij het gebruik van MinION is dit slechts 50,5% van de sequenties en wordt 49,5% van de sequenties geïdentificeerd als sequenties van genera die niet aanwezig zijn in de mock community. Het detecteren van bacteriegenera in de microbial profiles die niet aanwezig zijn in de mock communities geeft aan dat een groot deel van de sequenties (49,5%) niet juist wordt geïdentificeerd bij het gebruik van MinION sequencing. Waarschijnlijk is dit het gevolg van de aanwezigheid van veel sequentiefouten in MinION data (hoge "error rate").

Tabel 5. Identificatie 16S rRNA sequenties verkregen met Illumina en MinION sequencing tot op genusniveau

Soort in mock community (naamgeving volgens KWR database)	Illumina sequencing		MinION sequencing	
	Identificatie	Aantal sequenties		Identificatie
		141984	223193	
		% sequenties		
Aeromonas salmonicida	Aeromonas	22,7	14,7	Aeromonas
Aeromonas veronii				
Burkholderia multivorans	Burkholderia	1,1	0,9	Burkholderia
Herminiimonas glaciei	Herminiimonas	0,7	0,7	Herminiimonas
Citrobacter freundii	Enterobacteriaceae	27,6	9,8	Citrobacter
Klebsiella pneumoniae			6,3	Klebsiella
Enterobacter aerogenes	Enterobacteriaceae	3,7	0,2	Enterobacter
Escherichia coli	Escherichia	4,9	0,5	Escherichia
			0,6	Shigella
Rahnella aquatilis	Yersinia	5,1	0,6	Yersinia
Acinetobacter radioresistens	Acinetobacter	8,5	1,6	Acinetobacter
Legionella pneumophila	Legionella	0,3	0,1	Legionella
Pseudomonas fluorescens	Pseudomonas	7,0	1,7	Pseudomonas
Stenotrophomonas maltophilia	Stenotrophomonas	9,6	6,0	Stenotrophomonas
			2,9	Xanthomonas
Staphylococcus warneri	Micrococcus	3,5	0,3	Micrococcus
Clostridium sp.	Clostridium	1,0	0,7	Paraclostridium
Enterococcus faecium	Enterococcus	3,8	2,9	Enterococcus
	Other Genera	0,6	49,5	Other genera

Alleen genera die minimaal 0,1% van de gemeenschap uitmaken zijn weergegeven

Bij het identificeren van MinION sequenties tot op soortniveau worden veel meer soorten in de populaties waargenomen dan er soorten in de mock community aanwezig zijn. Als voorbeeld is in Figuur 16 de identificatie van de in de mock community aanwezige twee *Aeromonas* soorten (*A. salmonicida* en *A. veronii*) en *Escherichia coli* weergegeven.

Figuur 16. Voorbeeld van identificatie van *Aeromonas* en *E. coli* met MinION sequencing

Bij het identificeren van *Aeromonas* worden de twee soorten uit de mock community inderdaad teruggevonden in de gemeenschap. Maar, daarnaast worden ook nog vier andere soorten gevonden die niet aanwezig zijn in de mock community. In de populatie is *A. salmonicida* de meest dominante (7,7%) *Aeromonas* soort en vormt *A. bestiarum* volgens de identificatie met MinION 3,4% van de populatie terwijl deze soort niet aanwezig is in de mock community. Bij het identificeren van *E. coli* uit de mock community wordt met de MinION analyse een heel scala van verschillende, aan *E. coli* verwante bacteriën uit de *Enterobacteriaceae* familie, ten onrechte in de gemeenschap aangetoond waaronder een aantal pathogene *Shigella* soorten.

Deze resultaten laten zien dat identificatie van 16S rRNA sequenties op soortniveau niet betrouwbaar is bij het gebruik van MinION sequencing en dat technische aanpassingen aan het protocol noodzakelijk zijn waarbij de sequentie van elk individueel molecuul meerdere keren wordt bepaald resulterend in betrouwbaarder sequencingdata (Calus, Ijaz et al. 2018). Het ten onrechte aantonen van pathogenen, zoals *Shigella flexneri* en *Shigella sonnei* laat bovendien zien dat de 16S rRNA sequencing met MinION kan leiden tot onjuiste conclusies. Vergelijkbare resultaten werden verkregen in eerder onderzoek bij KWR (Heijnen 2018) en elders (Benitez-Paez, Portune et al. 2016). De recente verbeteringen aan de MinION flowcell, de daarin aanwezige nanoporiën en de analysesoftware hebben nog niet geresulteerd in een methode waarmee nanopore sequencing kan worden toegepast voor betrouwbare microbial profiling analyses. Mogelijk dat MinION sequencing van 16S rRNA sequenties wel kan worden toegepast om daarmee een globaal beeld te krijgen van de samenstelling van een bacteriële gemeenschap en de veranderingen die daarin optreden wanneer alleen informatie tot op genus niveau wordt gebruikt. Het gebruik van de volledige 16S sequentie kan wel degelijk leiden tot het identificeren van bacteriën in populaties (in afvalwater) tot soortniveau (Numberger, Ganzert et al. 2019). Bij deze studie is echter gebruik

gemaakt van het PacBio platform van de firma Pacific Biosciences (Menlo Park, Californië, USA). Met het PacBio systeem is het mogelijk om de sequentie van lange DNA-fragmenten ("full length" 16S rRNA) met een lage "error rate" (99,99% nauwkeurig) te bepalen, echter wel tegen aanmerkelijk hogere kosten. Ook moet voor dit systeem de kwaliteit van het DNA zeer hoog zijn.

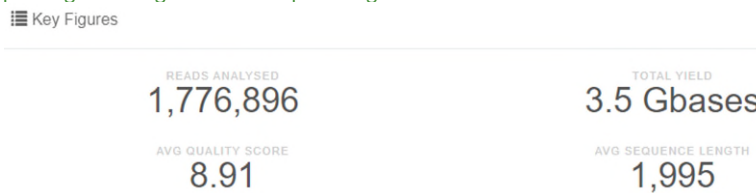
3.4 MinION voor metagenomics analyses

Om een eerste indruk te krijgen van de mogelijkheden om MinION sequencing toe te passen voor metagenomics analyses zijn er analyses uitgevoerd op DNA afkomstig van Drinkwater (in duplo), RWZI-Influent en Effluent van een RWZI

In eerste instantie was het ook de bedoeling om metagenomics analyses uit te voeren op een kunstmatig samengesteld monster van een mengsel van bekende bacteriën (mock community). Hiervoor is gebruik gemaakt van een commercieel verkrijgbaar product van Zymo Research (ZymoBIOMICS Microbial Community Standard). Dit product bevat een mengsel van acht bekende bacterie- en twee bekende schimmelsoorten. Het bleek in dit experiment echter niet mogelijk om sequenties van deze mock community te bepalen. Waarschijnlijk is dit het gevolg van de lage concentratie cellen resulterend in een te lage concentratie DNA.

Bij het sequencen van het metageenoom van bovenstaande vier monsters is de sequentie bepaald van 1.776.896 fragmenten met een gemiddelde lengte van 1995 nucleotiden waarmee ca. $3,5 \times 10^9$ nucleotiden aan sequentiedata is gegenereerd (Figuur 17).

Figuur 17. Opbrengst metagenomics sequencing



Sequencing van bovenstaande 1.776.896 reads is verdeeld over de vier geanalyseerd monsters. In Tabel 6 zijn het aantal fragmenten waarvan voor elk monster de sequentie is bepaald (totaal) en het aantal reads dat kon worden geïdentificeerd (door het vinden van een overeenkomstige sequentie in de database) en het percentage geïdentificeerde reads weergegeven.

Tabel 6. Totaal aantal reads, het geïdentificeerde aantal reads en het percentage geïdentificeerde reads per monster

	Reads		
	Totaal	Geïdentificeerde	% Geïdentificeerde
Drinkwater 1	225559	31547	14,0
Drinkwater 2	418552	54640	13,1
Rioolinfluent	803883	317889	39,5
Riooleffluent	222673	45530	20,4

De resultaten laten zien dat er relatief weinig sequenties afkomstig van drinkwater kunnen worden geïdentificeerd (14,0 en 13,1%). Voor RWZI-Influent en -RWZI-effluent was dit aandeel groter. Dit is waarschijnlijk het gevolg van de aanwezigheid van relatief veel onbekende sequenties in drinkwater.

Voor de analyse van drinkwater is het DNA van twee identieke drinkwatermonsters geanalyseerd om daarmee te zien of de analyse van twee identieke monsters ook

vergelijkbare resultaten oplevert. Het resultaat van deze vergelijking is weergegeven in Figuur 18.

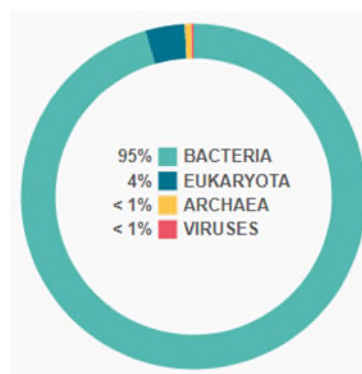
Figuur 18. Metagenomics analyses op twee identieke drinkwatermonsters

	Drinkwater	
	1	2
	% van geïdentificeerde reads	
Rhizobiales	9,8	9,5
Rhodobacterales	1,8	1,7
Rhodospirillales	2,0	2,0
Sphingomonadales	3,8	3,7
Methylococcales	1,8	1,9
Pseudomonadales	3,4	3,3
Xanthomonadales	1,7	1,8
Enterobacterales	2,7	2,7
Burkholderiales	12,4	12,4
Nitrosomonadales	1,4	1,4
Myxococcales	3,6	3,6
Nitrospirales	3,9	4,3
Corynebacterales	1,8	2,1
Streptomyetales	1,8	1,7
Micrococcales	1,5	1,3
Bacillales	1,4	1,3
Primates	3,1	3,1

De vergelijking van de analyseresultaten van twee identieke monsters laat zien dat het resultaat van deze analyses zeer vergelijkbaar is.

Van alle vier geanalyseerde monsters is het grootste deel van de sequenties afkomstig van bacteriën (95%), een kleiner deel is afkomstig van mensen (4%) en minder dan 1% is afkomstig van Archaea of Virussen (Figuur 19).

Figuur 19. Aandeel van eukaryoten, virussen, archaea en virussen in geïdentificeerde sequenties van alle monsters



Tabel 7. Aandeel van eukaryoten, virussen, archaea en virussen in geïdentificeerde sequenties van per monster

Geïdentificeerde Reads	Bacteria		Eukarya		Archea		Virussen		
	Aantal Reads	%	Aantal Reads	%	Aantal Reads	%	Aantal Reads	%	
Drinkwater	86187	76187	88,40	8315	9,65	589	0,68	107	0,12
Rioolinfluent	317889	309040	97,22	5471	1,72	2072	0,65	561	0,18
Riooleffluent	45530	42580	93,52	2297	5,05	180	0,40	90	0,20

3.4.1 Eukaryote organismen en virussen

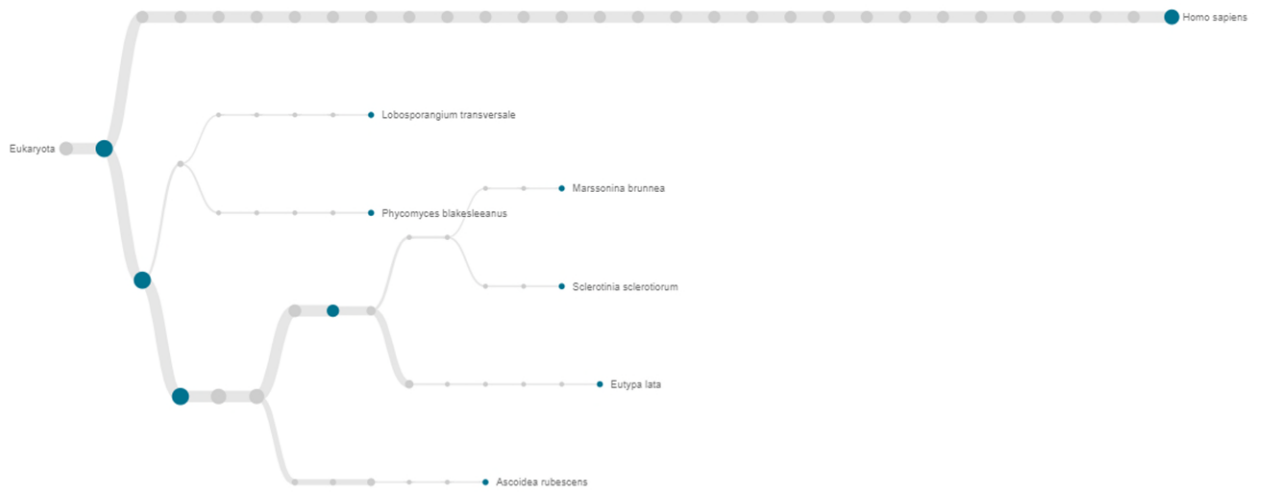
In Figuur 20 is grafisch weergegeven van welke eukaryote organismen er DNA is aangetoond met de metagenomics analyses op MinION.

Figuur 20. Sequenties van dominante eukaryote organismen in drinkwater (A), rioolinfluent (B) en riooleffluent (C)

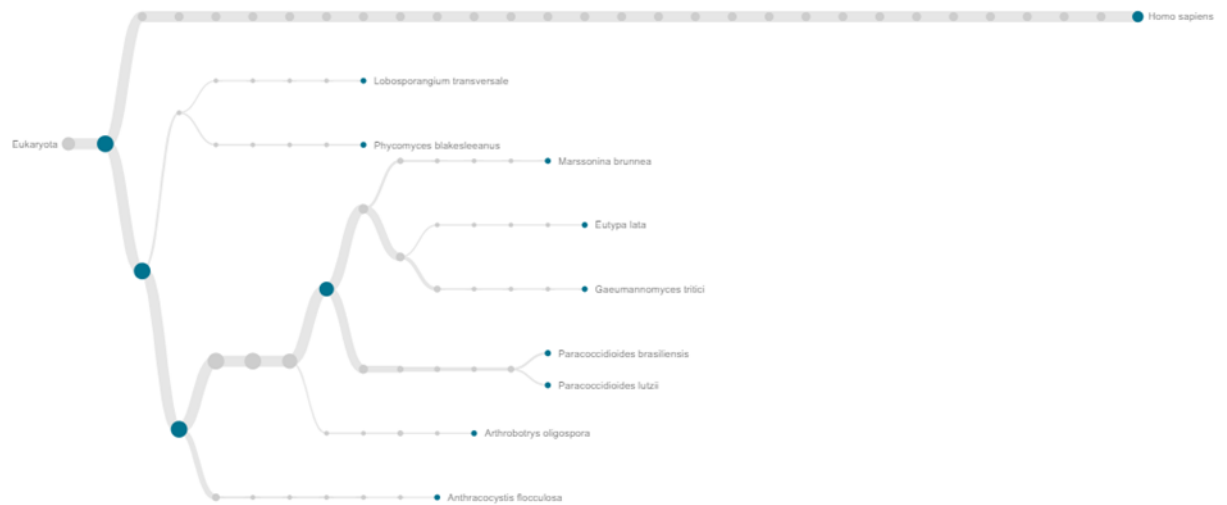
A.



B.



C.

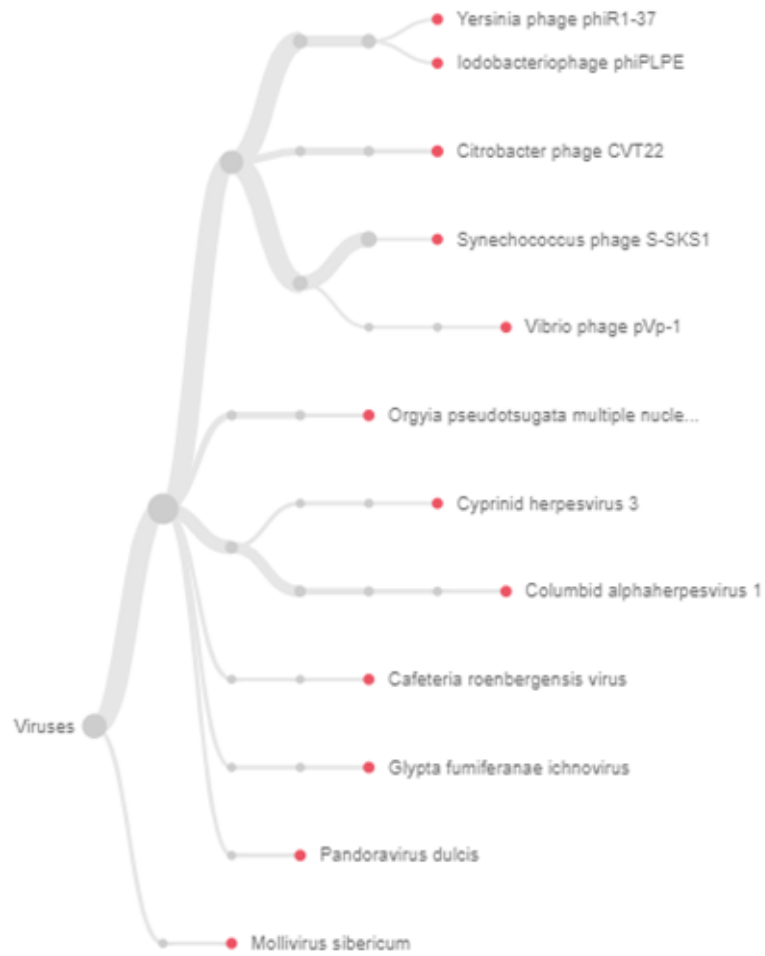


Het grootste deel van de DNA-sequenties van eukaryote organismen is geïdentificeerd als DNA van de mens (0,3% van alle sequenties in RWZI-Influent en 0,4% in drinkwater). In RWZI-Influent is dit waarschijnlijk het gevolg van de aanwezigheid van humane cellen, die via de huishoudens (t.g.v. contact met water zoals wassen en feces) in RWZI-Influent terecht zijn gekomen. De aanwezigheid van humaan DNA in drinkwater kan verklaard worden door de verspreiding van humane cellen via de drinkwaterzuivering en de introductie van humane cellen tijdens de laboratoriumwerkzaamheden. Een klein deel van de eukaryote sequenties in de monsters is afkomstig van verschillende schimmelsoorten. DNA van de schimmelsoorten *Lobosporangium transversale* en *Phycomyces blakesleeianus* wordt in drink- en rioolwater en RWZI-effluent aangetoond. Een aantal andere schimmels wordt alleen in RWZI-influent en/of -effluent aangetoond.

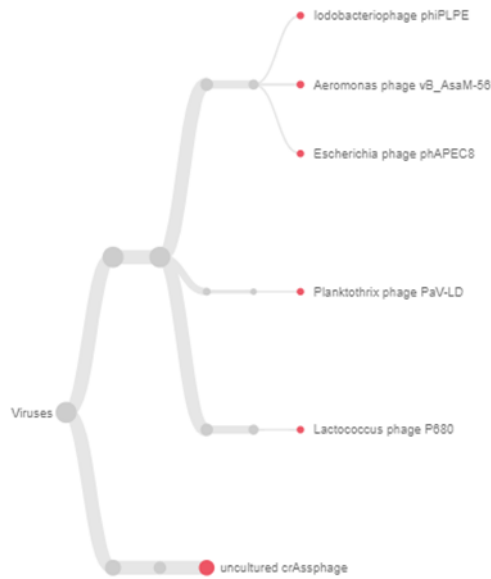
In de monsters worden ook sequenties afkomstig van virussen aangetoond. In het drinkwatermonster is 0,005% van alle sequenties geïdentificeerd als virussequenties, in RWZI-influent is dat 0,07% van alle sequenties en in -effluent 0,04% van alle sequenties. Een overzicht van de meest voorkomende virussen is weergegeven in Figuur 21.

Figuur 21. Sequenties van dominante virussen in drinkwater (A), rioolinfluent (B) en riooleffluent (C)

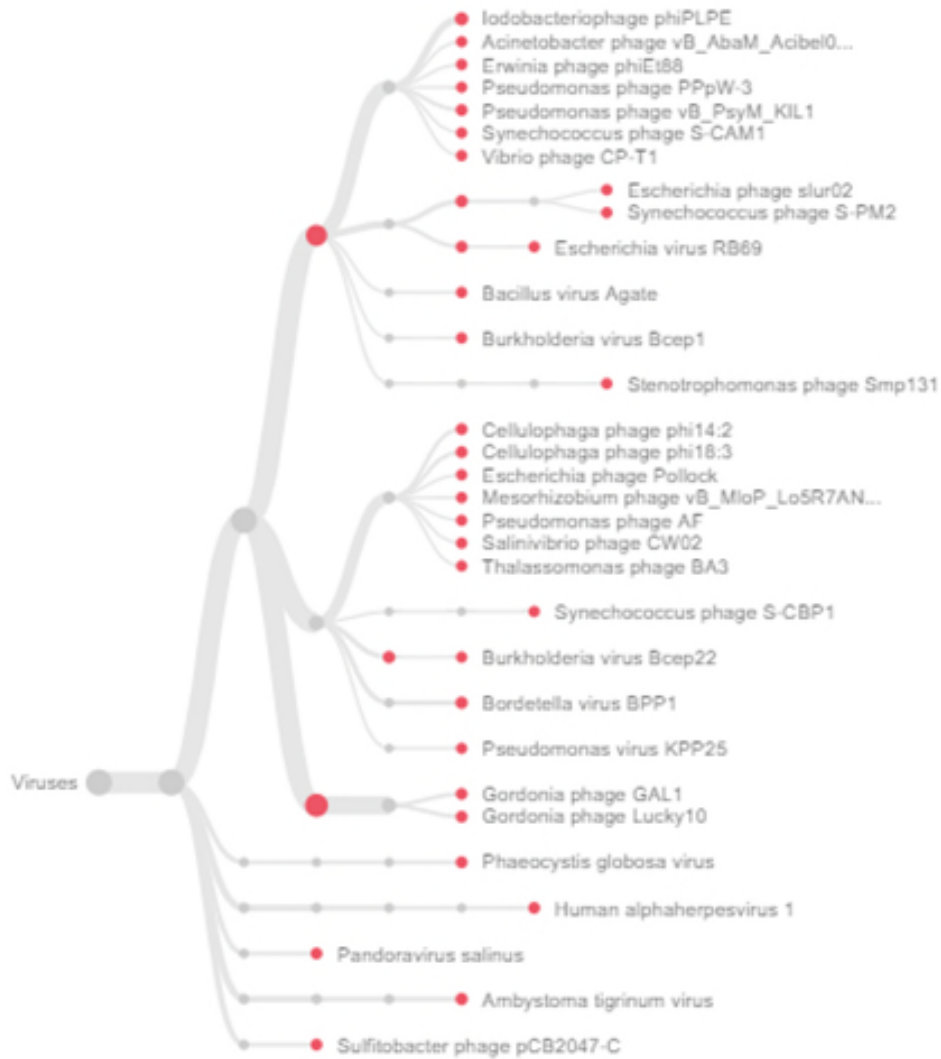
A.



B.



C.



De aanwezigheid van virale DNA-sequenties lijkt verrassend aangezien er membranen met een poriëgrootte van 0,2 µm zijn gebruikt voor filtratie van het water. Aangezien de afmeting van vrijwel alle virussoorten kleiner dan 0,2 µm is, valt te verwachten dat deze deeltjes de membraan zullen passeren. Veel van de in de monsters aangetoonde virussen zijn virussen die bacteriën als gastheer hebben (fagen) en het is te verwachten dat deze aangetoonde fagen in bacteriecellen aanwezig zijn en daardoor worden geconcentreerd met membraanfiltratie.

In het onderzochte drinkwatermonster is het interessant om te zien dat er, naast DNA-sequenties van een aantal fagen, ook sequenties worden aangetoond met overeenkomsten met de virussoorten waarvan bekend is dat de virusdeeltjes zeer groot zijn en waarschijnlijk zullen worden tegengehouden op een membraanfilter:

Mollivirus sibericum: een recent ontdekt (geïsoleerd uit permafrost) zeer groot virus met afmeting tussen 0,6 en 1.5 µm (Legendre, Lartigue et al. 2015).

Pandoravirus dulcia: een recent ontdekt zeer groot virus met afmeting van ca. 1 µm met amoeben als gastheer (Philippe, Legendre et al. 2013).

Cafeteria roenbergensis: een recent ontdekt zeer groot virus, geïsoleerd uit zeewater (Fischer, Allen et al. 2010)

Columbid alphaherpesvirus 1: een zeer groot herpesvirus, oorspronkelijk geïsoleerd uit duiven (Guo, Li et al. 2017)

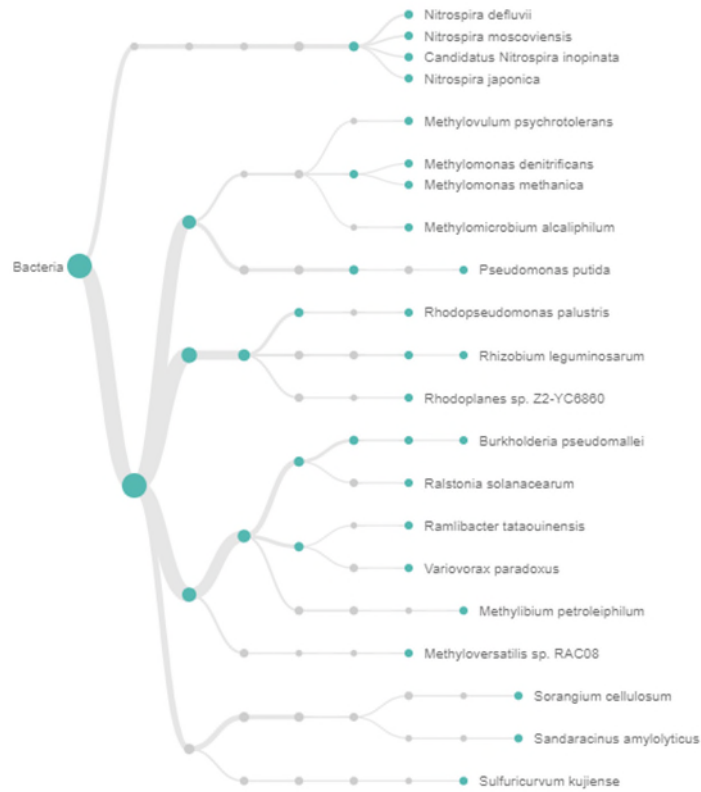
Cyprinid herpesvirus 3: een groot herpesvirus, veroorzaakt ziekte bij karpers (Michel, Fournier et al. 2010).

3.4.2 Bacteriele populaties

In Figuur 22 is een overzicht gegeven van de identificatie van het DNA van de dominante bacteriën uit de populaties in drinkwater, RWZI-influent en -effluent.

Figuur 22. Sequenties van dominante (minimaal 0,1% van de populatie) bacteriën in drinkwater (A), riolinfluent (B) en rioleffluent (C)

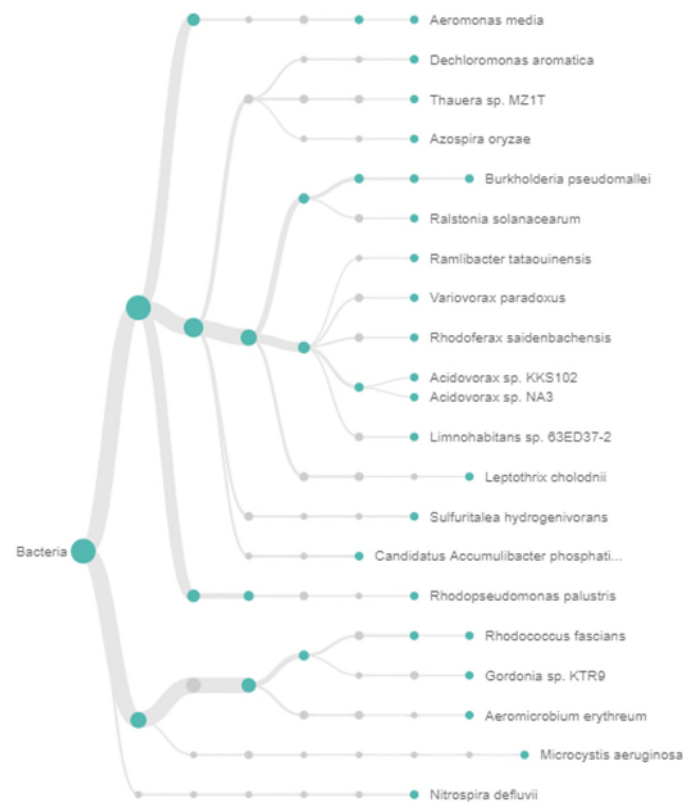
A.



B.



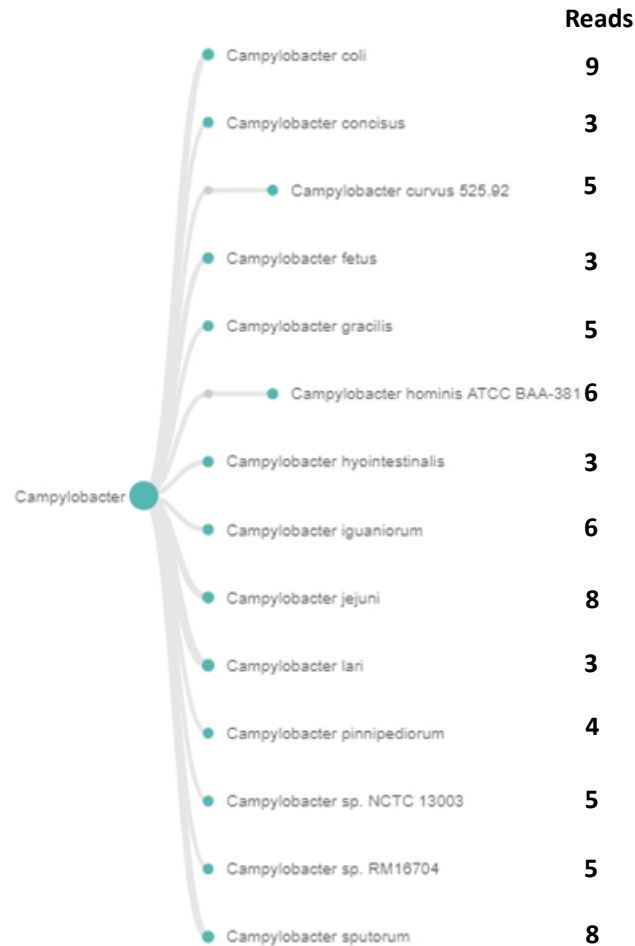
C.



Deze analyses laten zien dat er, zoals te verwachten, er grote verschillen zijn tussen de bacteriepopulaties in drinkwater, RWZI-influent en -effluent. In het drinkwater zijn, zoals ook in andere studies is aangetoond (Martiny, Albrechtsen et al. 2005, Van Assche, Crauwels et al. 2019), verschillende soorten van het genus *Nitrospira* dominant. Ook soorten behorende tot de *Pseudomonas*, *Rhizobium* en *Burkholderia* genera zijn eerder in drinkwater aangetoond (Perrin, Bouchon et al. 2019). De bacteriesoorten welke worden aangetoond in RWZI-influent komen goed overeen met de bacteriën waarvan al eerder beschreven is dat deze dominant zijn in RWZI-influent ((McLellan, Fisher et al. 2015, Numberger, Ganzert et al. 2019), <http://midasfieldguide.org/>). Veel van deze bacteriesoorten zijn van fecale herkomst, zoals verschillende *Bacteroides* soorten, *Faecalibacterium prausnitzii* en *Escherichia coli*. Daarnaast zijn verschillende *Aeromonas* soorten zeer dominant, een bacteriegroep waarvan bekend is dat deze in hoge concentraties voorkomt in RWZI-Influent (<http://www.waterpathogens.org/book/aeromonas>). In het RWZI-effluent worden bij deze analyses bacteriegroepen aangetoond waarvan bekend is dat deze dominant voorkomen in rioolwater-slib en -effluent (<http://midasfieldguide.org/>). Zo zijn bacteriën die een belangrijke rol spelen bij fosfaat verwijdering zeer dominant aanwezig ("*Candidatus Accumilibacter phosphatatis*). Ook van *Rhodoferax* is bekend dat deze dominant voorkomen in de rioolwaterzuivering (McLellan, Fisher et al. 2015). Hoewel meer onderzoek om de betrouwbaarheid van MinION voor het uitvoeren van metagenomics analyses analyses goed vast te stellen geven bovenstaande resultaten de eerste indicatie dat met metagenomics analyses op MinION een globaal beeld kan worden verkregen van de samenstelling van de dominante bacteriën in de populatie. Om een beter inzicht te krijgen van de mogelijkheden van MinION sequencing is in meer detail gekeken naar de zeldzamere sequenties in de populaties. Hiervoor is vooral

gekeken of er sequenties voorkomen van pathogene bacteriesoorten waarvan bekend is dat deze niet voorkomen in drinkwater in Nederland. In Figuur 23 is een voorbeeld gegeven van het aantal reads dat verkregen wordt met MinION sequencing en door de WIMP (What's In My Pot) software wordt geclassificeerd tot het genus van *Campylobacter*.

Figuur 23. Aantal als *Campylobacter* geklassificeerde MinION reads



Ook voor andere pathogene bacteriesoorten is onderzocht in hoeverre er in de, met MinION sequencing verkregen dataset, sequenties uit drinkwater worden geïdentificeerd als sequenties afkomstig van pathogenen.

Tabel 8. Sequenties van pathogenen in drinkwater na sequencing met MinION en identificatie met WIMP

Geclassificeerde pathoogeen met MinION en WIMP	Aantal reads
<i>Legionella pneumophila</i>	31
<i>Mycobacterium tuberculosis</i>	22
<i>Salmonella enterica</i>	16
<i>Vibrio cholera</i>	14
<i>Campylobacter jejuni</i>	8
<i>Yersinia pestis</i>	6
<i>Shigella sonnei</i>	1

Deze resultaten laten zien dat metagenoom analyse na sequencing met MinION en classificatie van de verkregen sequenties zorgt voor de identificatie van verschillende

pathogene micro-organismen waarvan een deel kan leiden tot ernstige ziekten. Het is uitgesloten dat deze pathogenen ook werkelijk in het drinkwatermonster aanwezig zijn. Het is eerder te verwachten dat de identificatie van deze bacteriesoorten het gevolg is van een combinatie van 1) het hoge aantal sequentiefouten in MinION data en 2) problemen met de identificatie van sequenties door "WIMP".

Ook de metagenomics analyses laten zien dat sequencing met MinION en identificatie van de sequenties met WIMP gebruikt kan worden voor het verkrijgen van globaal inzicht van de aanwezige bacteriën in een gemeenschap en de verschillen tussen gemeenschappen in verschillende typen monsters (drinkwater, rioolinfluent, -effluent). Deze aanpak is momenteel echter niet bruikbaar om inzicht te krijgen in het voorkomen van specifieke bacteriesoorten in een gemeenschap zoals het voorkomen van pathogenen.

3.5 MinION voor het identificeren van antibioticum resistentie genen

De metagenomics data is ook gebruikt voor het identificeren van DNA-sequenties die een rol hebben bij resistentie tegen antibiotica (ABR genen). In Tabel 9 is gegeven hoeveel reads er per monster zijn geanalyseerd, hoeveel van deze reads sequenties bevatten die significante homologie vertonen met ABR genen en het aantal ABR reads per miljoen reads. Het aantal ABR reads per miljoen reads wordt gebruikt als een maat voor de hoeveelheid ABR genen in een milieu. In het RWZI-influent is dit duidelijk hoger dan in RWZI-effluent en drinkwater. Dit is in overeenstemming met en ligt in dezelfde range als eerder beschreven bij onderzoek van afvalwater in Europa (Hendriksen, Munk et al. 2019).

Tabel 9. Het aantal sequentie reads met een relatie met genen betrokken bij antibiotacumresistentie

Monster	Reads	ABR reads	ABR reads Per 1.000.000 reads
Drinkwater	532427	36	68
Rioolinfluent	478573	308	644
Riooleffluent	170235	12	71

In bijlage I is een overzicht gegeven van de genen die betrokken zijn bij antibioticumresistentie en door de software zijn geïdentificeerd in de MinION metagenome data. Deze resultaten laten zien dat MinION sequencing in principe toegepast kan worden voor het identificeren van ABR genen in metagenoom data. Er zijn nog maar zeer weinig studies waarmee de verkregen data vergeleken kan worden om daarmee de waarde van MinION te kunnen beoordelen.

4 Conclusies en discussie

In dit verkennend onderzoek is onderzocht welke toepassingsmogelijkheden er zijn voor het gebruik van de MinION sequencer voor het microbiologisch onderzoek in de drinkwatersector. De MinION sequencer is een DNA-sequencer met het formaat van een USB stick en een kostprijs van \$1000 waarmee het mogelijk is om de samenstelling van individuele DNA moleculen te bepalen. Het systeem wordt door de leverancier (Oxford Nanopore) gebracht als een systeem waarmee DNA-sequenties en analyse van de data eenvoudig on-site plaats zou kunnen vinden. De mogelijkheid om de sequentie te bepalen van lange DNA-fragmenten is een potentieel voordeel van MinION t.o.v. het populaire Illumina systeem.

Het gebruik van de MinION sequencer

Na het volgen van een training is de MinION relatief eenvoudig te gebruiken. Er zijn op de website van Oxford Nanopore handige online instructie video's beschikbaar waarin de werking en bediening van het systeem duidelijk wordt uitgelegd. Het is wel van belang om rekening te houden met de voorwaarden die er aan ICT gesteld worden. Het is van belang dat de computer de beschikking heeft over voldoende en snelle (SSD) opslagcapaciteit en dat er onbeperkte communicatie tussen het systeem en een Oxford Nanopore server mogelijk is. Alle andere activiteiten (zoals een virusscanner) kunnen de MinION laten stoppen. Vanwege diverse ICT problemen bij gebruikers verkoopt Oxford Nanopore een klein "ICT kastje" die als vervanger van de computer dienst kan doen (MiniT: <https://nanoporetech.com/products/minit>). Voor de analyse van watermonsters is het noodzakelijk om te starten met voldoende biomassa waarvoor een filtratiestap noodzakelijk is. Ook is het van belang dat voldoende DNA wordt geïsoleerd en dat dit DNA voldoende zuiver is en ontdaan van stoffen die de sequencing kunnen verstoren. Hierdoor is het noodzakelijk goede en robuuste methoden te gebruiken voor het isoleren van DNA uit watermonsters. Vervolgens is een PCR-stap noodzakelijk voor het vermeerderen van het 16S rRNA gen (microbial profiling) of voor het vermeerderen van al het DNA (metagenomics). Het uitvoeren van filtratie, isolatie en PCR is momenteel nog arbeidsintensief en er zijn voor watermonsters nog geen goede methoden beschikbaar die snel en "on-site" kunnen worden uitgevoerd. Dit maakt "on-site" sequencing van de microbiologische gemeenschap in watermonsters momenteel nog niet mogelijk. Om dit wel mogelijk te maken zijn er vooral ontwikkelingsstappen nodig in het traject vóór de sequencer (monsterconcentratie, DNA-isolatie en voorbereiding van het DNA (PCR) voor sequencing).

Bij het lezen van de DNA-sequenties van individuele moleculen maakt de MinION relatie veel fouten (hoge "error-rate"). In de uitgevoerde experimenten was de error rate bij sequencing met MinION ca. 12%. De gemiddelde error-rate van de Illumina sequencer is <0,1%.

Het gebruik van MinION voor microbial profiling

Met microbial profiling wordt gerichte sequentie analyse van alleen het 16S rRNA gen gebruikt om inzicht te krijgen in de samenstelling van een bacteriële populatie. De mogelijkheid om de sequentie van lange DNA-fragmenten te bepalen is in potentie een groot voordeel van Nanopore sequencing. Dit maakt het mogelijk om de sequentie van het gehele 16S rRNA gen te bepalen waarmee identificatie tot op soortniveau mogelijk zou moeten zijn. Bij de analyse met MinION van een kunstmatig samengesteld monster

met hierin bekende bacteriesoorten ("Mock community") was de identificatie van dit mengsel tot op het taxonomisch niveau van "orde" in grote lijnen vergelijkbaar met resultaten verkregen met de Illumina MiSeq sequencer. Bij de identificatie van MinION sequenties tot op genusniveau wordt een groot deel (49,5%) van de sequenties geïdentificeerd tot genera waarvan geen bacteriën aanwezig zijn in de mock community. Bij identificatie tot op soortniveau werd, naast de soorten uit de mock community, ook een scala aan andere bacteriesoorten waargenomen. Dit geeft aan dat de er met MinION en de daaraan gekoppelde software onjuiste informatie wordt verkregen over de soortensamenstelling van de mock community. Het is te verwachten dat dit voor een groot deel het gevolg is van de hoge error rate van MinION sequencing. Mogelijk speelt ook de toegepaste software en de daarbij behorende database een rol.

Vooralsnog lijkt MinION sequencing, bij het toepassen van de door Oxford Nanopore aangeboden laboratorium- en bioinformatica-workflow, dus alleen geschikt voor het genereren van een globaal overzicht van de samenstelling van bacteriële populaties. Identificatie tot op soort- of genusniveau levert met deze workflow vooralsnog geen betrouwbare data op. Vergelijkbare resultaten zijn eerder ook in ander onderzoek waargenomen (Shin, Lee et al. 2016, Mitsuhashi, Kryukov et al. 2017). Vandaar dat een alternatieve workflow is ontwikkeld (NanoAmpli-Seq) waarmee het mogelijk is de sequentie van meerdere kopieën van elk individueel 16S rRNA molecuul door een nanopore te laten bepalen en vervolgens software toe te passen om hieruit een sequentie met hoge kwaliteit te genereren (Calus, Ijaz et al. 2018). Deze aanpak zorgt voor 16S rRNA sequenties met hoge kwaliteit ($\approx 99,5\%$ nauwkeurigheid) waarmee betrouwbare identificatie tot op soortniveau mogelijk is. Deze aanpak voor meer en gecompliceerder laboratoriumhandelingen waardoor "on site" sequencing vooralsnog niet haalbaar lijkt.

Het gebruik van MinION voor metagenomics analyses

Bij het gebruik van een metagenomics aanpak is sequencing niet beperkt tot het 16S rRNA gen maar wordt sequencing uitgevoerd op willekeurige, uit een monster afkomstige, DNA-fragmenten. Dit geeft de mogelijkheid om inzicht te krijgen in de samenstelling de dominante organismen in een monster. De analyse van een mock community geeft de beste mogelijkheid om de waarde van metagenomics analyses met MinION te kunnen beoordelen. Binnen dit project lukte het echter niet om metagenomics analyses uit te voeren op een commercieel verkrijgbare mock community van de firma Zymo Research. Waarschijnlijk was dit het gevolg van de lage concentratie cellen in deze mock community. Als alternatief zijn er metagenomics analyses uitgevoerd op DNA uit drinkwater en RWZI-Influent -effluent). Met deze metagenomics analyses wordt informatie verkregen over de aanwezigheid van eukaryote organismen, virussen, bacteriën en Archaea. Van de eukaryote organismen wordt vooral het DNA van mensen en ook van verschillende schimmels in de geanalyseerde monsters aangetoond. Vrijwel alle aangetoonde DNA-sequenties van virussen zijn afkomstig van virussen die zich in bacteriën vermeerderen (fagen) of van zeer grote virussen. Aangezien membraanfiltratie (poriegrootte: 0,2 μm) is gebruikt voor het concentreren van de organismen uit de monsters komt het detecteren van DNA van deze virussen overeen met de verwachting en is te verwachten dat andere (kleinere) virussen de membraan zijn gepasseerd. De analyse van de bacteriële populaties laat zien dat de dominante sequenties vooral afkomstig zijn van bacteriesoorten die, op basis van literatuurgegevens, ook te verwachten zijn in drink- en RWZI-Influent. Als er in detail wordt gekeken naar de geïdentificeerde zeldzamere sequenties in de dataset van drinkwater dan valt op dat er DNA van pathogene bacteriesoorten wordt waargenomen. Het zal niet zo zijn dat deze bacteriesoorten, of

het DNA ervan, in drinkwater aanwezig zal zijn, het is eerder te verwachten dat het aantonen van deze sequenties het gevolg is van de hoge error rate van MinION sequencing. Daarnaast zal de gebruikte software (WIMP) en de beperkte omvang van de door WIMP gebruikte bacteriële database (genomische sequenties van 5083 bacteriesoorten) een rol spelen bij de verkeerde identificatie van een beperkt deel van de sequenties. Deze verkeerde identificatie zal waarschijnlijk beperkte invloed hebben op de analyseresultaten waarbij het verkrijgen van een globaal beeld van de samenstelling van een populatie het doel is van de analyse. Als het bestuderen van zeldzamere sequenties, zoals sequenties van lage concentraties pathogene organismen, in een populatie het doel is dan kan dit leiden tot verkeerde conclusies bij het gebruik van MinION en WIMP. Waarschijnlijk is het mogelijk om betrouwbaarder metagenomics analyses uit te voeren bij het gebruik van andere software. Bij het toepassen van WIMP wordt een identificatie uitgevoerd op de sequentie van elke individueel molecuul. Er zijn echter ook mogelijkheden om eerst de sequentie te bepalen van een grote hoeveelheid willekeurige DNA-fragmenten en vervolgens software te gebruiken om deze sequenties te assembleren en aansluitend te identificeren. Deze aanpak zal waarschijnlijk zorgen voor betrouwbaarder metagenomics data waarmee ook de identificatie van functionele genen mogelijk is door gebruik te maken van alternatieve software en databases zoals MG-RAST (Keegan, Glass et al. 2016). Het onderzoeken van het gebruik alternatieve software en databases viel buiten de scope van dit onderzoek maar is zeker interessant voor vervolgonderzoek.

Het gebruik van MinION voor het identificeren van ABR genen

De, met MinION verkregen, metagenomics data maakt het ook mogelijk om DNA-sequenties te identificeren die een rol hebben bij resistentie tegen antibiotica (ABR genen). De resultaten van analyses van rioolwater laten zien dat het aantal gedetecteerde ABR genen in dezelfde range ligt als in een andere studie (Hendriksen, Munk et al. 2019). Uitgebreider onderzoek is nodig om de waarde de waarde en kwaliteit van, met MinION verkregen, ABR data te kunnen beoordelen.

Conclusies:

- Het uitgevoerde onderzoek heeft een eerste indruk gegeven over de toepassingsmogelijkheden van MinION sequencing en de bijbehorende analysesoftware.
- De analyse van watermonsters met MinION sequencing is nu alleen nog mogelijk op een specialistisch laboratorium (niet "on site").
- De kwaliteit van de verkregen MinION sequenties is aanmerkelijk lager (hoge "error rate" dan de kwaliteit van andere DNA-sequencing methoden (zoals Illumina).
 - Inmiddels zijn verbeterde flowcells beschikbaar waarmee lagere "error rates" worden verkregen.
 - Ook zijn alternatieve protocollen ontwikkeld waarmee de sequentie van individuele moleculen meerdere keren wordt bepaald waardoor lagere "error rates" worden verkregen.
- Het gebruik van MinION voor microbial profiling maakt het mogelijk om een globale indruk te krijgen van de samenstelling van de bacteriële gemeenschap in watermonsters.
 - Verder onderzoek is nodig om te bepalen in hoeverre data op hogere taxonomische niveaus (orde, genus) voldoende betrouwbaar en toepasbaar is.

- Identificatie op soortniveau levert geen betrouwbare informatie en kan lijden tot onjuiste conclusies.
- Het gebruik van MinION voor metagenomics maakt het mogelijk om een globale indruk te krijgen van de samenstelling van de dominante organismen (eukaryoten, virussen, bacteriën en archaea) en de aanwezigheid van antibioticumresistentiegenen in watermonsters.
 - De analyses geven een beeld van de dominant aanwezige virussen
 - De analyses geven een beeld van de dominant aanwezige eukarya
 - De analyses geven een beeld van de dominant aanwezige archaea
 - De analyses geven een beeld van de dominant aanwezige genen die betrokken zijn bij antibioticumresistentie
 - Identificatie van bacteriën op soortniveau levert geen betrouwbare informatie op en kan lijden tot onjuiste conclusies (aanwezigheid van pathogenen).
 - Verder onderzoek is nodig om te bepalen in hoeverre data over de meest dominante organismen in de populaties voldoende betrouwbaar en toepasbaar is.
- Voor routinematige toepassing van MinION is verdere verbetering van de kwaliteit van sequencing (lagere error rates) en software noodzakelijk.

5 Literatuur

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-410.
- Benitez-Paez, A., K. J. Portune and Y. Sanz (2016). "Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer." *Gigascience* **5**: 4.
- Calus, S. T., U. Z. Ijaz and A. J. Pinto (2018). "NanoAmpli-Seq: A workflow for amplicon sequencing from mixed microbial communities on the nanopore sequencing platform." [bioRxiv](#).
- Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer and R. Knight (2011). "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample." *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4516-4522.
- Ewing, B. and P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." *Genome Res* **8**(3): 186-194.
- Fischer, M. G., M. J. Allen, W. H. Wilson and C. A. Suttle (2010). "Giant virus with a remarkable complement of genes infects marine zooplankton." *Proceedings of the National Academy of Sciences* **107**(45): 19508-19513.
- Guo, Y., S. Li, X. Sun, Y. He, H. Zhao, Y. Wang, P. Zhao and M. Xing (2017). "Complete genome sequence and evolution analysis of a columbid herpesvirus type 1 from feral pigeon in China." *Arch Virol* **162**(7): 2131-2133.
- Heijnen, L. (2018). "Toepassingsmogelijkheden van: Metatranscriptomics, Microbial profiling en MinION sequencing." [BTO 2018.044](#).
- Hendriksen, R. S., P. Munk, P. Njage, B. van Bunnik, L. McNally, O. Lukjancenka, T. Röder, D. Nieuwenhuijse, S. K. Pedersen, J. Kjeldgaard, R. S. Kaas, P. T. L. C. Clausen, J. K. Vogt, P. Leekitcharoenphon, M. G. M. van de Schans, T. Zuidema, A. M. de Roda Husman, S. Rasmussen, B. Petersen, A. Bego, C. Rees, S. Cassar, K. Coventry, P. Collignon, F. Allerberger, T. O. Rahube, G. Oliveira, I. Ivanov, Y. Vuthy, T. Sopheak, C. K. Yost, C. Ke, H. Zheng, L. Baisheng, X. Jiao, P. Donado-Godoy, K. J. Coulbaly, M. L. Malania, A. Nitsche, A. Brinkmann, C. K. S. Saba, B. Kocsis, N. Solymosi, T. R. Thorsteinsdottir, A. M. Hatha, M. Alebouyeh, D. Morris, M. Cormican, L. O'Connor, J. Moran-Gilad, P. Alba, A. Battisti, Z. Shakenova, C. Kiiyukia, E. Ng'eno, L. Raka, J. S. J. gYb_cz5"6 fn] yzJ "6Ufh_Yj Jvz 7 "DYbbnz < "FU'UbXUgz G" DUFja UbbUbzA "J "<UVYfz P. Pal, G.-J. Jeunen, N. Gemmell, K. Fashae, R. Holmstad, R. Hasan, S. Shakoore, M. L. Z. Rojas, D. Wasyl, G. Bosevska, M. Kochubovski, C. Radu, A. Gassama, V. Radosavljevic, S. K i YfmzF "Ni b][U. A cbhUbYnz A "M: "HUnz 8"; Uj U cz z z? "DUgh Wcj Uz D" Hfi g_UzA " Trkov, K. Esterhuysen, K. Keddy, M. Cerdà-Cuellar, S. Pathirage, L. Norrgren, S. Örn, D. G. J. Larsson, T. V. d. Heijden, H. H. Kumburu, B. Sanneh, P. Bidjada, B.-M. Njanpop-Lafourcade, S. C. Nikiema-Pessinaba, B. Levent, J. S. Meschke, N. K. Beck, C. D. Van, N. D. Phuc, D. M. N. Tran, G. Kwenda, D.-a. Tabo, A. L. Wester, S. Cuadros-Orellana, C. Amid, G. Cochrane, T. Sicheritz-Ponten, H. Schmitt, J. R. M. Alvarez, A. Aidara-Kane, S. J. Pamp, O. Lund, T. Hald, M. Woolhouse, M. P. Koopmans, H. Vigre, T. N. Petersen, F. M. Aarestrup and c. The Global Sewage Surveillance project (2019). "Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage." *Nature Communications* **10**(1): 1124.
- Jia, B., A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. Brinkman, G. D. Wright and A. G. McArthur (2017). "CARD 2017: expansion and model-centric curation

of the comprehensive antibiotic resistance database." *Nucleic Acids Res* **45**(D1): D566-d573.

Juul, S., F. Izquierdo, A. Hurst, X. Dai, A. Wright, E. Kulesha, R. Pettett and D. J. Turner (2015). "What's in my pot? Real-time species identification on the MinION™." *bioRxiv*: 030742.

Keegan, K. P., E. M. Glass and F. Meyer (2016). MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. *Microbial Environmental Genomics (MEG)*. F. Martin and S. Uroz. New York, NY, Springer New York: 207-233.

Kim, D., L. Song, F. P. Breitwieser and S. L. Salzberg (2016). "Centrifuge: rapid and sensitive classification of metagenomic sequences." *Genome Res* **26**(12): 1721-1729.

Legendre, M., A. Lartigue, L. Bertaux, S. Jeudy, J. Bartoli, M. Lescot, J.-M. Alempic, C. Ramus, C. Bruley, K. Labadie, L. Shmakova, E. Rivkina, Y. Couté, C. Abergel and J.-M. Claverie (2015). "In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba." *Proceedings of the National Academy of Sciences of the United States of America* **112**(38): E5327-E5335.

Martiny, A. C., H. J. Albrechtsen, E. Arvin and S. Molin (2005). "Identification of bacteria in biofilm and bulk water samples from a nonchlorinated model drinking water distribution system: detection of a large nitrite-oxidizing population associated with *Nitrospira* spp." *Appl Environ Microbiol* **71**(12): 8611-8617.

McLellan, S. L., J. C. Fisher and R. J. Newton (2015). "The microbiome of urban waters." *Int Microbiol* **18**(3): 141-149.

Michel, B., G. Fournier, F. Lieffrig, B. Costes and A. Vanderplasschen (2010). "Cyprinid herpesvirus 3." *Emerging infectious diseases* **16**(12): 1835-1843.

Mitsuhashi, S., K. Kryukov, S. Nakagawa, J. S. Takeuchi, Y. Shiraishi, K. Asano and T. Imanishi (2017). "A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer." *Scientific Reports* **7**(1): 5657.

Numberger, D., L. Ganzert, L. Zoccarato, K. Mühldorfer, S. Sauer, H.-P. Grossart and A. D. Greenwood (2019). "Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing." *Scientific Reports* **9**(1): 9673.

Perrin, Y., D. Bouchon, V. Delafont, L. Moulin and Y. Héchard (2019). "Microbiome of drinking water: A full-scale spatio-temporal study to monitor water quality in the Paris distribution system." *Water Research* **149**: 375-385.

Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013).

"Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." *Science* **341**(6143): 281-286.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glockner (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." *Nucleic Acids Res* **41**(Database issue): D590-596.

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn and C. F. Weber (2009). "Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities." *Appl Environ Microbiol* **75**(23): 7537-7541.

Shahi, S. K., S. N. Freedman and A. K. Mangalam (2017). "Gut microbiome in multiple sclerosis: The players involved and the roles they play." *Gut Microbes* **8**(6): 607-615.

Shin, J., S. Lee, M.-J. Go, S. Y. Lee, S. C. Kim, C.-H. Lee and B.-K. Cho (2016). "Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing." *Scientific Reports* **6**: 29681.

Van Assche, A., S. Crauwels, J. De Brabanter, K. A. Willems and B. Lievens (2019). "Characterization of the bacterial community composition in water of drinking water production and distribution systems in Flanders, Belgium." *MicrobiologyOpen* **8**(5): e00726-e00726.

Yilmaz, P., L. W. Parfrey, P. Yarza, J. Gerken, E. Priesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig and F. O. Glockner (2014). "The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks." Nucleic Acids Res **42**(Database issue): D643-648.

Bijlage I ABR genen

Drinkwater				
	identity	name	URL	Reads
1	82.76	Neisseria gonorrhoeae 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40097	5
2	86.73	Chlamydomonas reinhardtii 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40087	4
3	89.46	Neisseria meningitidis 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40099	4
4	83.92	Helicobacter pylori 16S rRNA mutation conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/40112	4
5	86.18	Salmonella enterica serovar Typhimurium 16S rRNA mutation in the rrsD gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40114	4
6	79.69	Streptomyces cinnamomeus EF-Tu mutants conferring resistance to elfamycin	https://card.mcmaster.ca/ontology/39943	3
7	83.07	Planobispora rosea EF-Tu mutants conferring resistance to inhibitor GE2270A	https://card.mcmaster.ca/ontology/39945	2
8	82.32	Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to kanamycin	https://card.mcmaster.ca/ontology/40020	2
9	83.69	Escherichia coli 16S rRNA mutation conferring resistance to edeine	https://card.mcmaster.ca/ontology/39807	1
10	82.87	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/39960	1
11	83.97	Mycobacterium tuberculosis rpsL mutations conferring resistance to Streptomycin	https://card.mcmaster.ca/ontology/39979	1
12	83.86	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to gentamicin C	https://card.mcmaster.ca/ontology/39980	1
13	89.01	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	https://card.mcmaster.ca/ontology/39986	1
14	83.99	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to paromomycin	https://card.mcmaster.ca/ontology/39987	1
15	81.16	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to streptomycin	https://card.mcmaster.ca/ontology/39989	1
16	79.88	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	https://card.mcmaster.ca/ontology/40147	1
17	76.92	Salmonella enterica parC conferring resistance to fluoroquinolones	https://card.mcmaster.ca/ontology/40662	1

Rioolinfluent				
	identity	name	URL	Reads
1	87.11	Helicobacter pylori 16S rRNA mutation conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/40112	43
2	94.16	Neisseria meningitidis 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40099	32
3	82.53	Salmonella enterica serovar Typhimurium 16S rRNA mutation in the rrsD gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40114	25
4	93.01	tet36	https://card.mcmaster.ca/ontology/36336	8
5	79.66	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to streptomycin	https://card.mcmaster.ca/ontology/39989	7
6	90.06	Neisseria gonorrhoeae 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40097	7
7	96.04	CfxA6	https://card.mcmaster.ca/ontology/39650	6
8	87.89	Escherichia coli rpoB mutants conferring resistance to rifampicin	https://card.mcmaster.ca/ontology/39872	6
9	90.6	Escherichia coli 16S rRNA mutation in the rrsH gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/39956	6
10	84.15	Escherichia coli EF-Tu mutants conferring resistance to Pulvomycin	https://card.mcmaster.ca/ontology/39953	5
11	90.72	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/39960	5
12	89.2	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to tobramycin	https://card.mcmaster.ca/ontology/39992	5
13	81.61	Enterococcus faecium EF-Tu mutants conferring resistance to GE2270A	https://card.mcmaster.ca/ontology/40022	5
14	85.69	Mycobacterium smegmatis 16S rRNA mutation in the rrsA gene conferring resistance to neomycin	https://card.mcmaster.ca/ontology/40146	5
15	80.58	Escherichia coli 16S rRNA mutation in the rrsC gene conferring resistance to kasugamicin	https://card.mcmaster.ca/ontology/39917	4
16	83.88	Escherichia coli 16S rRNA mutation in the rrnB gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/39961	4
17	96.76	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	https://card.mcmaster.ca/ontology/39986	4
18	87.13	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/39994	4
19	91.5	Pasteurella multocida 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40095	4
20	95.17	sul1	https://card.mcmaster.ca/ontology/36549	3
21	99.18	OXA-119	https://card.mcmaster.ca/ontology/38175	3
22	95.54	CblA-1	https://card.mcmaster.ca/ontology/39433	3
23	82.68	Mycobacterium abscessus 16S rRNA mutation conferring resistance to neomycin	https://card.mcmaster.ca/ontology/39822	3
24	78.89	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to gentamicin C	https://card.mcmaster.ca/ontology/39980	3
25	90.79	Escherichia coli 16S rRNA mutation in the rrnB gene conferring resistance to streptomycin	https://card.mcmaster.ca/ontology/39990	3
26	83.43	Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to streptomycin	https://card.mcmaster.ca/ontology/40076	3
27	82.53	Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to amikacin	https://card.mcmaster.ca/ontology/40077	3
28	82.01	OpmH	https://card.mcmaster.ca/ontology/40317	3
29	95.82	tetO	https://card.mcmaster.ca/ontology/36329	2
30	97.6	tetW	https://card.mcmaster.ca/ontology/36333	2
31	83.86	acrB	https://card.mcmaster.ca/ontology/36355	2
32	91.02	mel	https://card.mcmaster.ca/ontology/36910	2
33	80.37	mexF	https://card.mcmaster.ca/ontology/37184	2
34	97.11	OXA-129	https://card.mcmaster.ca/ontology/38211	2
35	98.76	aadA	https://card.mcmaster.ca/ontology/39001	2
36	98.11	catQ	https://card.mcmaster.ca/ontology/39121	2
37	83.53	mexW	https://card.mcmaster.ca/ontology/39465	2
38	94.99	Escherichia coli 16S rRNA mutation conferring resistance to edeine	https://card.mcmaster.ca/ontology/39807	2

39	82.66	Clostridium difficile EF-Tu mutants conferring resistance to elfamycin	https://card.mcmaster.ca/ontology/39941	2
40	88.52	Escherichia coli 16S rRNA mutation in the rrnB gene conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/39995	2
41	84.24	Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to viomycin	https://card.mcmaster.ca/ontology/40021	2
42	83.05	Chlamydomophila psittaci 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40087	2
43	84.66	Borrelia burgdorferi 16S rRNA mutation conferring resistance to gentamicin	https://card.mcmaster.ca/ontology/40106	2
44	81.72	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to kanamycin A	https://card.mcmaster.ca/ontology/40144	2
45	88.11	Mycobacterium smegmatis 16S rRNA mutation in the rrsA gene conferring resistance to kanamycin A	https://card.mcmaster.ca/ontology/40145	2
46	78.9	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	https://card.mcmaster.ca/ontology/40147	2
47	84.8	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to viomycin	https://card.mcmaster.ca/ontology/40149	2
48	86.66	Pseudomonas aeruginosa gyrA conferring resistance to fluoroquinolones	https://card.mcmaster.ca/ontology/40319	2
49	84.13	Pseudomonas aeruginosa parE conferring resistance to fluoroquinolones	https://card.mcmaster.ca/ontology/40320	2
50	95.16	tetE	https://card.mcmaster.ca/ontology/36312	1
51	83.15	mexB	https://card.mcmaster.ca/ontology/36517	1
52	96.94	ErmF	https://card.mcmaster.ca/ontology/36637	1
53	79.08	macB	https://card.mcmaster.ca/ontology/36674	1
54	79.77	adeJ	https://card.mcmaster.ca/ontology/37161	1
55	95.6	mdtB	https://card.mcmaster.ca/ontology/37173	1
56	95.51	baeS	https://card.mcmaster.ca/ontology/37209	1
57	86.16	cpxR	https://card.mcmaster.ca/ontology/37211	1
58	98.57	evgS	https://card.mcmaster.ca/ontology/37213	1
59	96.12	mdtG	https://card.mcmaster.ca/ontology/37728	1
60	96.89	OXA-1	https://card.mcmaster.ca/ontology/37796	1
61	97.26	OXA-5	https://card.mcmaster.ca/ontology/37800	1
62	85.01	OXA-12	https://card.mcmaster.ca/ontology/37807	1
63	98.2	OXA-368	https://card.mcmaster.ca/ontology/37955	1
64	94.93	OXA-212	https://card.mcmaster.ca/ontology/38111	1
65	97.64	MOX-2	https://card.mcmaster.ca/ontology/38583	1
66	94.46	AER-1	https://card.mcmaster.ca/ontology/38881	1
67	97.04	aadA22	https://card.mcmaster.ca/ontology/39019	1
68	97.84	APH(3'')-Ib	https://card.mcmaster.ca/ontology/39039	1
69	98.97	APH(6)-Id	https://card.mcmaster.ca/ontology/39060	1
70	94.18	APH(6)-Id	https://card.mcmaster.ca/ontology/39060	1
71	83.33	dfrE	https://card.mcmaster.ca/ontology/39309	1
72	84.85	ImrD	https://card.mcmaster.ca/ontology/39316	1

73	78.2	ceoB	https://card.mcmaster.ca/ontology/39444	1
74	91.24	dfxB3	https://card.mcmaster.ca/ontology/39456	1
75	80.45	smeB	https://card.mcmaster.ca/ontology/39486	1
76	97.11	msrE	https://card.mcmaster.ca/ontology/39685	1
77	83.31	Mycobacterium abscessus 16S rRNA mutation conferring resistance to kanamycin	https://card.mcmaster.ca/ontology/39820	1
78	75.38	Morganella morganii gyrB conferring resistance to fluoroquinolone	https://card.mcmaster.ca/ontology/39890	1
79	87.03	Salmonella serovars gyrB conferring resistance to fluoroquinolone	https://card.mcmaster.ca/ontology/39891	1
80	90.76	Escherichia coli parE conferring resistance to fluoroquinolones	https://card.mcmaster.ca/ontology/39900	1
81	82.35	Streptomyces rishiriensis parY mutant conferring resistance to aminocoumarin	https://card.mcmaster.ca/ontology/39902	1
82	81.5	Escherichia coli EF-Tu mutants conferring resistance to kirromycin	https://card.mcmaster.ca/ontology/39952	1
83	85.36	Escherichia coli EF-Tu mutants conferring resistance to Enacyloxin IIa	https://card.mcmaster.ca/ontology/39954	1
84	97.87	Escherichia coli soxR mutants	https://card.mcmaster.ca/ontology/39965	1
85	89.3	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to G418	https://card.mcmaster.ca/ontology/39981	1
86	90.94	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to kanamycin A	https://card.mcmaster.ca/ontology/39983	1
87	90.15	Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to paromomycin	https://card.mcmaster.ca/ontology/39987	1
88	84.93	Propionibacterium acnes 16S rRNA mutation conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/40101	1
89	83.15	Borrelia burgdorferi 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40104	1
90	87.42	Mycobacterium chelonae 16S rRNA mutation conferring resistance to amikacin	https://card.mcmaster.ca/ontology/40116	1
91	84.47	Mycobacterium chelonae 16S rRNA mutation conferring resistance to kanamycin A	https://card.mcmaster.ca/ontology/40117	1
92	89.24	Mycobacterium chelonae 16S rRNA mutation conferring resistance to tobramycin	https://card.mcmaster.ca/ontology/40118	1
93	82.41	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to hygromycin B	https://card.mcmaster.ca/ontology/40142	1
94	83.46	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to streptomycin	https://card.mcmaster.ca/ontology/40143	1
95	97.14	PmrC	https://card.mcmaster.ca/ontology/40186	1
96	88.52	Pseudomonas aeruginosa parC conferring resistance to fluoroquinolone	https://card.mcmaster.ca/ontology/40355	1
97	79.14	nfsA	https://card.mcmaster.ca/ontology/40407	1
98	80.89	aminocoumarin resistant alaS	https://card.mcmaster.ca/ontology/40521	1
99	93.96	kdpE	https://card.mcmaster.ca/ontology/40534	1
100	95.55	Escherichia coli mutant UhpT conferring resistance to fosfomycin	https://card.mcmaster.ca/ontology/40592	1
101	81.69	Neisseria gonorrhoeae parC conferring resistance to fluoroquinolone	https://card.mcmaster.ca/ontology/40639	1
102	77.81	Salmonella enterica parC conferring resistance to fluoroquinolones	https://card.mcmaster.ca/ontology/40662	1
103	82.83	msbA	https://card.mcmaster.ca/ontology/40720	1

Rioleffluent					
	identity	model	name	URL	Reads
1	88.73	rRNA mutation model	Neisseria meningitidis 16S rRNA mutation conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40099	2
2	80.39	protein homolog model	qepA	https://card.mcmaster.ca/ontology/36587	1
3	97.91	protein homolog model	mdtA	https://card.mcmaster.ca/ontology/37172	1
4	93.5	protein homolog model	mdtF	https://card.mcmaster.ca/ontology/37176	1
5	80.43	protein homolog model	Streptomyces rishiriensis parY mutant conferring resistance to aminocoumarin	https://card.mcmaster.ca/ontology/39902	1
6	81.36	protein homolog model	Streptomyces cinnamoneus EF-Tu mutants conferring resistance to elfamycin	https://card.mcmaster.ca/ontology/39943	1
7	86.32	protein variant model	Mycobacterium tuberculosis rpsL mutations conferring resistance to Streptomycin	https://card.mcmaster.ca/ontology/39979	1
8	86.13	rRNA mutation model	Escherichia coli 16S rRNA mutation in the rrnB gene conferring resistance to tetracycline	https://card.mcmaster.ca/ontology/39995	1
1	88.86	rRNA mutation model	Salmonella enterica serovar Typhimurium 16S rRNA mutation in the rrsD gene conferring resistance to spectinomycin	https://card.mcmaster.ca/ontology/40114	1
2	91.63	rRNA mutation model	Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to kanamycin A	https://card.mcmaster.ca/ontology/40144	1
3	82.48	protein homolog model	mexK	https://card.mcmaster.ca/ontology/40341	1