

A network diagram consisting of various sized light blue circles connected by thin white lines, set against a solid blue background. The circles vary in size, with some being significantly larger than others, and they are interconnected in a complex, non-linear fashion.

Bedrijfstakonderzoek
BTO 2021.054 | Mei 2022

Microbial profiling met NGS voor drinkwateronderzoek

Bedrijfstakonderzoek

KWR

Bridging Science to Practice

Rapport

Microbial profiling met NGS voor drinkwateronderzoek

BTO 2021.054 | Mei 2022

Dit onderzoek is onderdeel van het collectieve Bedrijfstakingonderzoek van KWR, de waterbedrijven en Vewin.

Opdrachtnummer

402045/036 en 402045/100

Projectmanager

Michiel Hootsmans

Opdrachtgever

BTO - Thematisch onderzoek - Biologische veiligheid

Auteurs

Leo Heijnen en Peer Timmers

Kwaliteitsborger

Paul van der Wielen

Verzonden naar

Dit rapport is verspreid onder BTO-participanten.

Een jaar na publicatie is het openbaar.

Keywords

NGS, Microbiologie, Sequencing, DNA, microbiële gemeenschap, microbiële populatie, moleculaire technieken

Jaar van publicatie
2022

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

PO Box 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



Mei 2022 ©

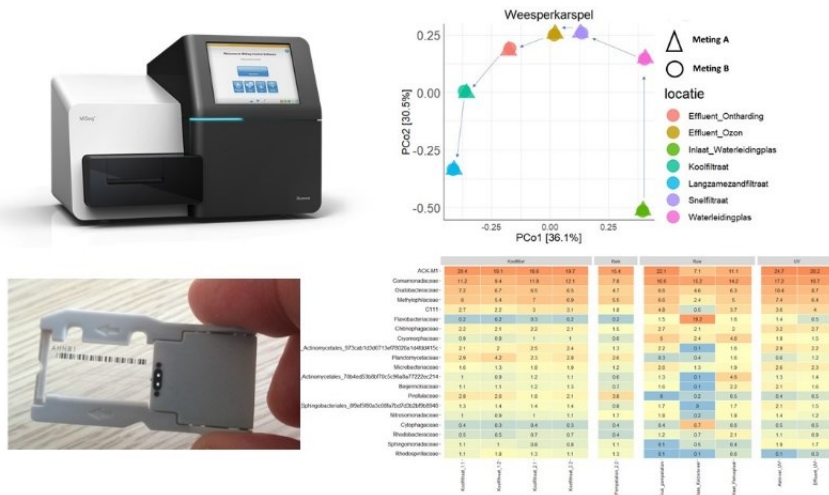
Alle rechten voorbehouden aan KWR. Niets uit deze uitgave mag - zonder voorafgaande schriftelijke toestemming van KWR - worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier.

Managementsamenvatting

Met standaardisering vormt microbial profiling met NGS een waardevolle toevoeging voor drinkwateronderzoek

Auteurs: Leo Heijnen en Peer Timmers

Microbial profiling maakt het mogelijk met één analyse een overzicht te krijgen van de aanwezige micro-organismen in een watermonster. Voor vrijwel elke stap binnen de microbial-profiling-analyse kunnen verschillende keuzes worden gemaakt en elk van die keuzes kan de uiteindelijke analyseresultaten sterk beïnvloeden. Het standaardiseren van de procedures is daarom van groot belang. Er is een gestandaardiseerd protocol ontwikkeld voor de uitvoering van de analyses in het laboratorium en er zijn stappen gezet voor de ontwikkeling van controles waarmee de kwaliteit van de analyses kan worden gewaarborgd. De stappen die zijn gezet laten zien dat reproduceerbare analyses met microbial profiling kunnen worden uitgevoerd, waarmee al relatief kleine veranderingen van de microbiologische gemeenschap zichtbaar gemaakt kunnen worden.



Microbial profiling met NGS: de apparatuur voor het genereren van data (links) en visualisatie van analyseresultaten (rechts)

Belang: microbial profiling ontwikkelen voor monitoring waterkwaliteit en zuiveringsprocessen

Voor de levering van microbiologisch onberispelijk drinkwater is het van belang om de kwaliteit van het water in het traject van winning tot distributie te kunnen beoordelen en de processen die plaatsvinden tijdens de zuivering goed te begrijpen en efficiënt te kunnen beheren. Daarvoor is de samenstelling van de microbiële gemeenschap in het water ook

waardevolle informatie. Deze is vast te stellen met microbial profiling: daarbij wordt met Next Generation Sequencing (NGS) in één analyse de samenstelling en volgorde (sequentie) van grote aantallen DNA-moleculen van het 16S rRNA gen bepaald. Deze methode is snel en kost weinig en geeft met een analyse een overzicht van de samenstelling van vrijwel de complete microbiologische populatie in watermonsters.

Mogelijk kan deze technologie in de toekomst worden gebruikt als “brede microbiologische screening” van de microbiologische waterkwaliteit, de werking van microbiële zuiveringsprocessen en het optreden van ongewenste nagroei tijdens zuivering of distributie. De technologie van microbial profiling is zowel op het gebied van laboratoriumtechnieken als op het onderdeel data-interpretatie nog sterk in ontwikkeling. Voor vrijwel elke stap binnen de microbial-profiling-analyse kunnen verschillende keuzes worden gemaakt, en deze keuzes kunnen veel invloed hebben op de uiteindelijke analyseresultaten. Het standaardiseren van de procedures is daarom van groot belang om microbial profiling te ontwikkelen tot een krachtig middel voor efficiënte monitoring van de waterkwaliteit tijdens zuivering en distributie.

Aanpak: literatuur- en experimenteel-onderzoek ingezet voor gestandaardiseerd protocol

Aan de hand van literatuur en experimenten is voor elk onderdeel van de analyseprocedure voor microbial profiling geïnventariseerd welke keuzes gemaakt kunnen worden en welke effecten de gemaakte keuzes kunnen hebben op de resultaten. Op basis van deze inventarisatie is een gestandaardiseerd protocol ontwikkeld voor de uitvoering van het werk op het laboratorium en is onderzocht welke “bioinformatica-tools” kunnen worden toegepast voor het interpreteren van (grote hoeveelheid) DNA-sequentiedata. De ontwikkelde procedure is vervolgens toegepast in de praktijk om de veranderingen van de microbiologie tijdens zuiveringsstappen te onderzoeken.

Resultaten: standaardisering microbial profiling belangrijk voor ontwikkeling veelbelovend middel

De resultaten van het onderzoek laten zien dat microbial profiling reproduceerbare resultaten kan opleveren met drinkwatermonsters. Daarmee kan microbial profiling worden toegepast voor het identificeren van minimale veranderingen in de microbiële gemeenschap.

Het standaardiseren van analyses blijkt van groot belang om data uit verschillende onderzoeken te kunnen vergelijken. Voor het waarborgen van de kwaliteit van de analyses is ook het toepassen en ontwikkelen van controlemonsters (“mock communities”) van groot belang.

Toepassing: microbial profiling geeft waardevolle kennis over drinkwaterzuivering en -distributie

Deze studie laat zien dat microbial profiling kan worden toegepast om een betrouwbaar en reproduceerbaar beeld te krijgen van de samenstelling van de microbiologie in watermonsters. Daarmee komt het gebruik van microbial profiling voor routinematige kwaliteitsmetingen steeds dichterbij. Vooral de ontwikkeling en optimalisatie van eenvoudig toepasbare apparatuur gaat erg snel, vooral de snelle kwaliteitsverbetering van draagbare sequencers, zoals MinION van Oxford Nanopore, is daarbij veelbelovend.

Rapport

Dit onderzoek is beschreven in het rapport *Microbial profiling met NGS voor drinkwateronderzoek* (BTO 2021.054). Lees ook meer in:

- (Timmers, Slootweg et al. 2022)
- (Vavourakis, Heijnen et al. 2020)
- (Roeselers, Coolen et al. 2015)

Jaar van publicatie
2022

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

PO Box 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

KWR

Mei 2022 ©

Alle rechten voorbehouden aan KWR. Niets uit deze uitgave mag - zonder voorafgaande schriftelijke toestemming van KWR - worden veeleenvoudigd, 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.

Inhoud

Rapport	2
<i>Managementsamenvatting</i>	3
Inhoud	5
1 Verklarende lijst met termen	8
2 Microbial profiling met NGS voor drinkwateronderzoek	9
3 Microbial profiling: de technologie	11
3.1 DNA sequencing	11
3.1.1 Illumina sequencing	13
3.2 Microbial profiling	15
4 Verschillende stappen voor microbial profiling	18
4.1 Monstername, conservering en concentratie	18
4.1.1 Monstername	18
4.1.2 Concentreren van monsters	19
4.1.3 Conserveren van monsters	19
4.2 Isolatie van DNA van de microbiële gemeenschap	19
4.2.1 Algemeen	19
4.2.2 Effect van DNA-isolatiemethoden op microbiële profielen	21
4.2.3 Effect van DNA concentratie op microbiële profielen	22
4.3 Het genereren van een “fragmentbibliotheek” met 16S rRNA fragmenten	23
4.3.1 Primerkeuze en PCR 16S rRNA gen	23
4.3.2 Genereren van PCR-fragmenten voor Illumina sequencing	23
4.4 Bioinformatica voor dataverwerking en interpretatie	24
4.4.1 Platforms, software en databases	24
4.4.2 Dataverwerking met Qiime2	25
4.4.3 Demultiplexing	25
4.4.4 Joining pairs	26
4.4.5 Ruisonderdrukking/clustering	26
4.4.6 Taxonomische classificatie en fylogenie	27
4.4.7 Diversiteitsanalyses alpha- en beta-diversiteit	28
4.4.8 Differentiële abundantie	31
4.4.9 Staafdiagrammen/heatmaps en meer statistiek	31
4.4.10 Qiime viewer	32

5	Kwaliteitscontroles	33
5.1	Blanco monsters als negatieve controle	33
5.2	Mock communities als positieve controle	35
6	Bepalen van minimale DNA-concentratie voor microbial profiling	38
6.1	Celaantallen, DNA-concentratie en sequencereads	38
6.2	Alpha-diversiteit	39
6.3	Beta-diversiteit	41
6.4	Conclusie	43
7	Praktijktoepassingen microbial profiling	44
7.1	Microbial profiling voor detectie van verontreinigingen in drinkwater	44
7.1.1	Alpha-diversiteit	44
7.1.2	Beta-diversiteit	46
7.1.3	Differentiële abundantie en indicatorsoorten	47
	Analysis of composition of microbiomes (ANCOM)	51
7.1.4	Source Tracker	55
7.1.5	Conclusie	57
7.2	Veranderingen in microbiële gemeenschappen tijdens zuivering	57
7.2.1	Zuivering productielocatie Berenplaat (Evides)	58
7.2.2	Zuivering Weesperkarspel (Waternet)	62
7.2.3	Oppervlaktewaterzuivering “De Punt” (Waterbedrijf Groningen)	65
7.2.4	Oppervlaktewaterzuivering Leiduin (Waternet)	68
7.2.5	Drinkwaterzuivering Scheveningen (Dunea)	70
7.2.6	Drinkwaterzuivering Loosbroek (Brabant Water)	73
7.2.7	Drinkwaterzuivering Zele (De Watergroep)	76
7.2.8	Monsters van PWN	78
8	Conclusies	83
8.1	De technologie: de sequencingsystemen	83
8.2	Standaardisatie van microbial profiling	84
8.2.1	Monsternamen en conservering	84
8.2.2	Isolatie van DNA	84
8.2.3	Effect van DNA concentratie	84
8.2.4	Keuze van primers	84
8.2.5	Bioinformatica voor interpretatie van data	84
8.2.6	Kwaliteitscontroles	85
8.3	Toepassing van microbial profiling voor drinkwateronderzoek	85
8.3.1	Reproduceerbaarheid	85
8.3.2	Benodigde DNA concentratie	85
8.3.3	Detectie van verontreinigingen	86
8.3.4	Praktijktoepassingen (casestudies)	86

9	Aanbevelingen	87
10	Literatuur	88
I	Workshop “Toepassing van NGS in de drinkwatersector”	91
II	Protocol 16S v4 sequencing op Illumina MiSeq	92
III	Rekenvoorbeeld Bray-Curtis	98

1 Verklarende lijst met termen

ASV: Amplicon sequence variant. De term ASV wordt gebruikt voor individuele unieke sequenties afkomstig van NGS analyses waarbij de sequentie van PCR-amplicons van merkergenen wordt bepaald (zoals 16S rRNA). ASV is een alternatief voor Operational Taxonomic Unit (OTU).

Beide ASVs en Otus worden in Qiime2 **features** genoemd

Baseparen (bp): het aantal DNA-bouwstenen van de dubbele helix van DNA

DNA: Desoxy-Nucleïc-Acid, biochemisch macromolecuul dat fungeert als belangrijkste drager van erfelijke informatie. Het bestaat uit 4 bouwstenen: A(denine), G(uanine), C(itosine) en T(hymine).

DNA-sequentie: samenstelling van de DNA streng

16S-rRNA gen: gen coderend voor een deel van de 30S subunit van ribosomen bij bacteriën en archaea. de sequentie van dit gen wordt algemeen toegepast voor taxonomische identificatie.

Microbioom: de complete collectie (gemeenschap) van micro-organismen in een bepaald milieu bepaald met whole genome sequencing..

Microbiota: de complete collectie (gemeenschap) van micro-organismen in een bepaald milieu bepaald met sequencing van het 16S rRNA gen.

NGS: Next Generation Sequencing, ook "high throughput sequencing" (HTS) genoemd. Het betreft een technologie voor het bepalen van de sequentie van een zeer grote aantal DNA-moleculen

Nucleinezuren: DNA en RNA

OTU: Operationele Taxonomische Unit, taxonomische unit voor clustering van sequenties waarbij de mate van sequentieovereenkomst de basis is voor clustering van de sequenties. Voor 16S rRNA gen-sequenties wordt 97% sequentieovereenkomst aangehouden voor taxonomische clustering op soortniveau.

RNA: Deoxy-Ribonucleïc-Acid, biologisch macromolecuul dat essentieel is voor de cellulaire processen in alle bekende levensvormen. Het bevat de code voor de vorming van eiwitten en bestaat uit 4 bouwstenen: A(denine), G(uanine), C(itosine) en U(racil).

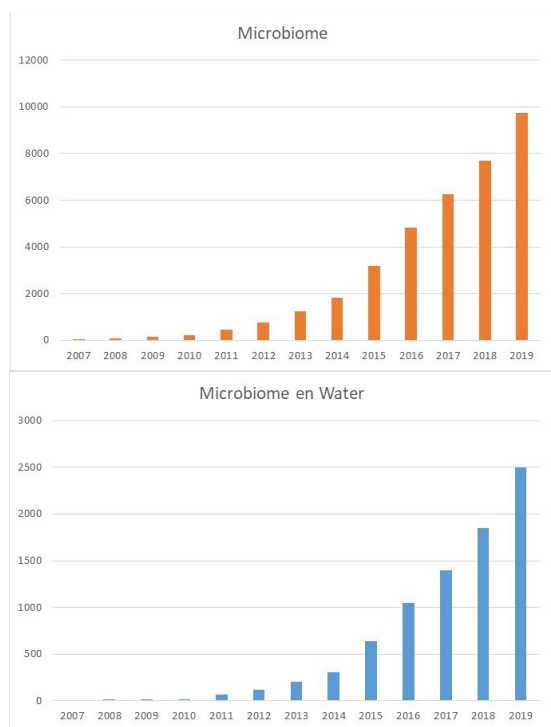
Sequentie: =DNA-sequentie

2 Microbial profiling met NGS voor drinkwateronderzoek

Voor de levering van microbiologisch goed en veilig drinkwater is het van belang om de kwaliteit van het water in het traject van winning tot distributie te kunnen beoordelen en de processen die plaats vinden tijdens de zuivering goed te begrijpen en efficiënt te kunnen beheren. Met het bepalen van de samenstelling van de microbiologische gemeenschap is het mogelijk om veranderingen van de microbiologie waar te nemen en de gevolgen hiervan te beoordelen. Deze veranderingen kunnen het gevolg zijn van:

- Veranderingen in de kwaliteit van het water vóór de zuivering;
- het optreden van veranderingen/problemen in de zuivering;
- het optreden van ongewenste situaties tijdens distributie zoals nagroei of hygiënische problemen na ingrepen in het leidingnet.

Recente technologische ontwikkelingen maken het mogelijk om, tegen lage kosten, de samenstelling en volgorde van de DNA-bouwstenen (DNA-sequentie) van enorme hoeveelheden DNA fragmenten te bepalen (Next Generation Sequencing: NGS). Door toepassingen van deze technologie voor sequencing van een deel van het 16S ribosomaal RNA (rRNA) gen kan in één analyse een vrijwel compleet beeld van de soortensamenstelling van de microbiologische gemeenschap (microbiota) gegenereerd worden ("microbial profiling"). Het is gebleken dat studies naar de samenstelling van het menselijk microbiota en het microbiota in verschillende milieus zorgen voor veel nieuwe toepasbare kennis waardoor het aantal microbiota studies, sinds de introductie van NGS, een enorme vlucht heeft genomen (Figuur 1). Ook het onderzoek naar het microbiota van water maakt een sterke groei door en de eerste toepassingen van microbial profiling in de Nederlandse en Vlaamse drinkwatersector zijn veelbelovend (Roeselers, Coolen et al. 2015, Heijnen 2016, Brouwer, Wielen et al. 2017, Liu, Zhang et al. 2018, Vavourakis, Heijnen et al. 2020, Learbuch, Smidt et al. 2021). Het valt te verwachten dat de toepassingsmogelijkheden van microbial profiling in de komende jaren sterk zullen toenemen. Als startpunt van dit project werd in september 2018 over dit onderwerp een workshop gehouden met sprekers uit de drinkwatersector en de academische onderzoekswereld (korte samenvatting in Bijlage I). Tijdens deze workshop werd duidelijk dat er, naast de vele mogelijkheden van microbial profiling, ook beperkingen zijn aan de methoden. Waardevolle toepassing van microbial profiling is alleen mogelijk als de voordelen, maar ook de beperkingen, van deze technologie duidelijk zijn. Daarnaast is het van belang om een gestandaardiseerde procedure te ontwikkelen waarmee betrouwbare, reproduceerbare en controleerbare resultaten worden verkregen. Alleen dan kan microbial profiling betrouwbaar worden ingezet voor toepassing in het microbiologisch drinkwateronderzoek.



Figuur 1. Aantal wetenschappelijke artikelen met "Microbiome" of "Microbiome en Water" in de titel (bron PubMed)

Om de mogelijkheden van microbial profiling met NGS voor en met de drinkwatersector in kaart te brengen is het noodzakelijk om:

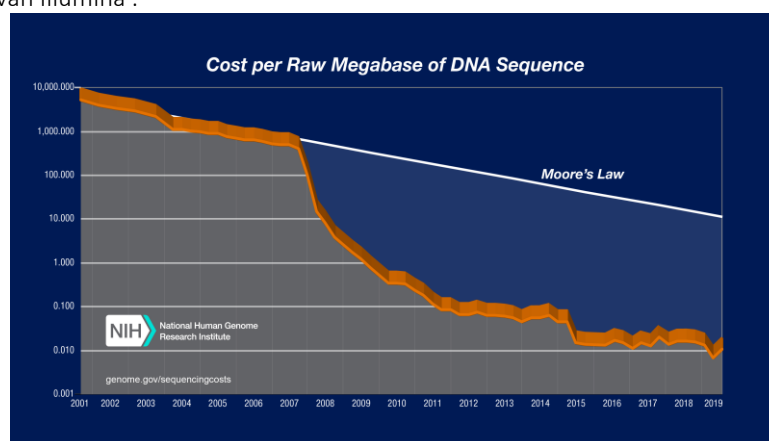
- De mogelijkheden en beperkingen van microbial profiling te kennen
- De prestatiekenmerken van deze analyses te kennen
- De juiste controles te ontwikkelen waarmee de kwaliteit van de analyses kan worden geborgd.

Dit onderzoek heeft tot doel om inzicht te krijgen in deze onderdelen. Dit inzicht is verkregen door literatuuronderzoek uit te voeren en dit te combineren met experimenteel onderzoek. Aansluitend is een beperkte (praktische) casestudie uitgevoerd waarbij is onderzocht in hoeverre de gestandaardiseerde methode kan worden ingezet om een beeld te krijgen van de veranderingen die optreden in verschillende drinkwaterzuiveringen. Deze casestudies zijn aangevuld met een literatuuroverzicht van de toepassingen van microbial profiling bij verschillende drinkwateronderzoeken.

3 Microbial profiling: de technologie

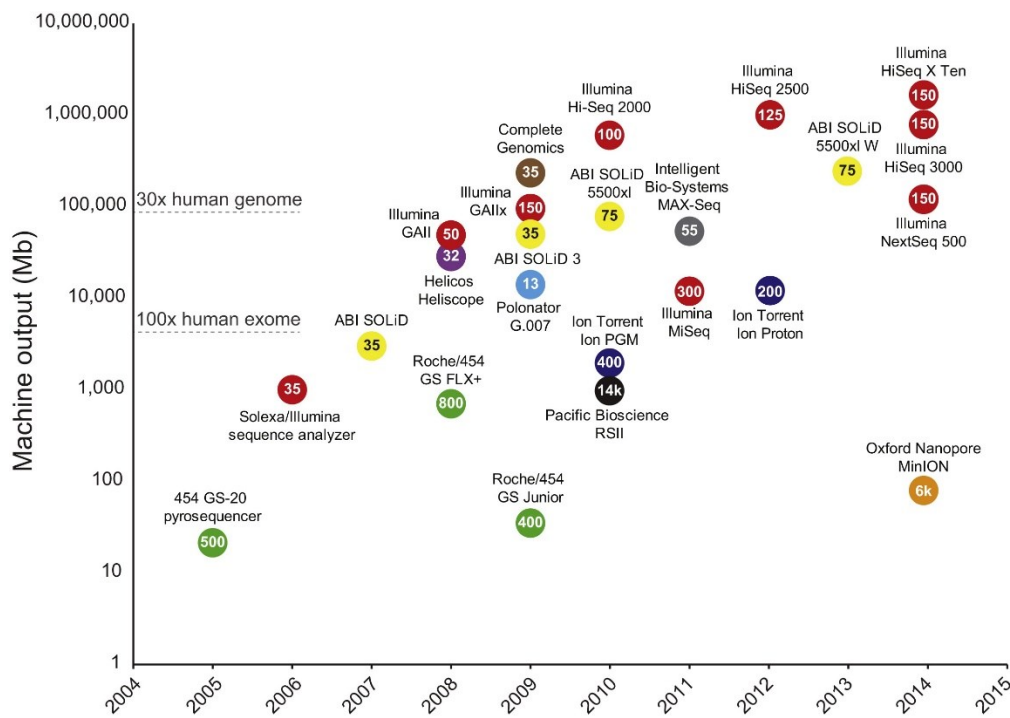
3.1 DNA sequencing

Een DNA-molecuul bestaat uit vier bouwstenen: Adenine (A), Guanine (G), Cytosine (C) en Thymine (T). Met sequencing wordt de samenstelling en volgorde van deze bouwstenen bepaald. De eerste methode voor het bepalen van de sequentie (samenstelling en volgorde van de vier bouwstenen van DNA-moleculen: A, C, G en T) van DNA-moleculen is ontwikkeld in 1977 (Sanger, Nicklen et al. 1977) en is bekend geworden onder de naam "Sanger sequencing". Met deze methode is het mogelijk om de sequentie van één individueel DNA-fragment te bepalen. Deze methode is lange tijd gebruikt voor het bepalen van DNA sequenties en was ook de methode voor het bepalen van de sequentie van het humane genoom (HUGO), een project dat in 1988 is gestart en in 2003 afgerond waarmee 99% van de sequentie van één individu is bepaald (Collins, Morgan et al. 2003). Toepassing van Sanger sequencing was erg arbeidsintensief en daardoor zeer kostbaar. De kosten voor uitvoering van Sanger sequencing werden enigszins beperkt door automatisering van een deel van de laboratoriumhandelingen. Voor het bepalen van de DNA-sequentie van 1 miljoen baseparen m.b.v. Sanger sequencing waren de kosten in het jaar 2000 ca. \$10.000 en door vergaande automatisering werden deze kosten beperkt tot ca. \$700 in 2007 (Figuur 2). Onder invloed van het HUGO project werden revolutionaire nieuwe sequencing technieken ontwikkeld, in 2005 werd het eerste "Next generation sequencing" systeem (454 sequencing) ontwikkeld waarmee het mogelijk werd om de DNA sequentie van grote hoeveelheden DNA-fragmenten geautomatiseerd te bepalen. Na het commercialiseren van dit systeem en de implementatie ervan bij verschillende laboratoria, namen de kosten voor sequencing verder af. Door de ontwikkeling van verbeterde NGS apparatuur die meer data per run genereren, zijn de kosten voor uitvoeren van sequencing nog verder terug gelopen tot rond de \$0,01 per miljoen baseparen (Figuur 2). Door de ontwikkeling van deze verbeterde systemen is de verkoop van 454 sequencing (Roche diagnostics) gestopt en sinds 2006 zijn er door diverse fabrikanten verschillende systemen ontwikkeld met elk, op het vlak van "throughput" hun eigen specifieke kenmerken (Figuur 3 en Tabel 1) waarbij de markt momenteel vooral wordt gedomineerd door sequenceapparaten van Illumina.



Figuur 2. Ontwikkeling van de kosten voor sequencing van DNA

De Illumina systemen zijn geschikt voor het sequencen van enorme aantallen korte DNA-fragmenten ("short reads"); van 25 miljoen fragmenten met het MiSeq systeem tot 400 miljard fragmenten met het NovaSeq systeem. Hierbij wordt van elk fragment maximaal ca. 600 bp nauwkeurige sequentie-informatie verkregen. De kosten van deze apparatuur zijn vrij hoog: ca. € 80.000 voor een MiSeq tot ca. € 1000.000 voor een HiSeq systeem.



Figuur 3. Ontwikkeling van de output (Mbs) voor verschillende NGS sequencing technologieën van 2004 t/m 2015.

Inmiddels zijn ook systemen in opkomst waarmee het mogelijk is om de DNA-sequentie van veel langere fragmenten ("long reads": tot meer dan 1 miljoen bp) te bepalen zoals het MinION-systeem van Oxford Nanopore en het PacBio RSII van Pacific Biosciences. De investeringskosten voor de MinION zijn bovendien erg laag (ca. € 1000), het systeem is eenvoudig en eventueel ook op locatie te gebruiken.

Tabel 1. High throughput sequencing technologieën met de karakteristieken, overgenomen uit (da Fonseca, Albrechtsen et al. 2016)

Sequencing technology/Platform	Detection method	Library types	Maximum read length (bp)	Reqs per run (maximum)	Error rate (approximate)	Pros	Cons
454/GS FLX titanium XL+	Pyrophosphate detection (M., 1998)	Single end/	1000	1,000,000	0.2% (Shao et al., 2013)	Medium size reads. Errors are well characterized (Shao et al., 2013).	Will be discontinued in 2016. Inaccurate homopolymer detection (M., 1998). Emulsion PCR ^a .
454/GS Junior Systems		Paired end	800	100,000			
Illumina-Solexa/Hiseq 4000	Fluorescence, reversible terminators (Bentley et al., 2008)	Single end/	2 × 150	5,000,000,000	0.2–0.8% (Quail et al., 2012)	Widely used. Flexible library preparation methods. High-throughput well suited for resequencing projects. Good characterization of biases (Schirmer et al., 2015).	Short reads. Not optimal for de novo assembly.
Illumina-Solexa/MySeq		Paired end	2 × 300	25,000,000			
Life Technologies/SOLiD	Fluorescence di-base probes (McKernan et al., 2009)	Single end/ Paired end	1 × 75/2 × 50	1,400,000,000	0.01% (Buermans and den Dunnen, 1842)	Second most used (van Dijk et al., 2014). Second highest throughput. Each base is read twice, thus decreasing the error rate. Short running times.	Color space not supported by many mappers. Short reads. Emulsion PCR ^a .
Life Technologies/Ion Torrent	Hydrogen ion (pH) sensor (Merriman et al., 2012)	Single end/ Paired end	400	5,500,000	1.8% (Quail et al., 2012)	Longest reads. Good for improving de novo assemblies. Single molecule sequencing. Long reads. No GC bias (Cherf et al., 2012). Portable. Scalable. Real-time data. Single molecule. It is possible to read both strands of the DNA sequence.	Bias against AT-rich regions (Quail et al., 2012). Inaccurate homopolymer detection (Buermans and den Dunnen, 1842). Emulsion PCR ^a . Low throughput. High cost-throughput ratio. High error rates (Quail et al., 2012). High error rates. Quality scores only defined by the quality of the alignment to a reference sequence. Existing mapping and assembly software do not deal with long and high error reads.
PacBio RS II/SMRT	Fluorescence phospho-linked nucleotides (Eid et al., 2009)	Single end	20,000	55,000	13% (Quail et al., 2012)		
Oxford Nanopore/MinION	Electrical sensing (Cherf et al., 2012)	Single end	2000	60,000	30% (Madoui et al., 2015)		

^a The emulsion PCR step adds an extra bias. Only 1/3 of the molecules will contain exactly one molecule, and thus be useful during the sequencing process.

3.1.1 Illumina sequencing

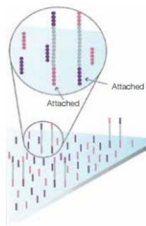
Voor NGS analyse van 16S rRNA gensequenties van microbiële gemeenschappen worden momenteel met name sequenceapparaten van de firma Illumina toegepast. Zoals eerder beschreven is het met dit systeem mogelijk om betrouwbare sequenties te genereren van fragmenten met een maximale lengte van ca. 500bp. Eerst wordt, op het DNA dat uit een monster is geïsoleerd, een PCR uitgevoerd waarbij een fragment van het 16S rRNA gen van bacteriën en/of archaea wordt geamplificeerd (16S-amplicon). Vervolgens wordt op het 16S-amplicon een tweede PCR stap uitgevoerd. Hiermee wordt het geamplificeerde 16S rRNA genfragment van elk monster binnen één NGS experiment aan één of beide uiteinden voorzien van unieke “indexsequenties” van ieder zeven nucleotiden lang. Na deze tweede PCR worden amplicons (met indexsequenties) van verschillende monsters in equimolaire hoeveelheden gemengd en dit mengsel wordt door het sequenceapparaat geanalyseerd. Binnen het sequentieapparaat vindt de sequentieanalyse plaats in een DNA chip (Figuur 4).



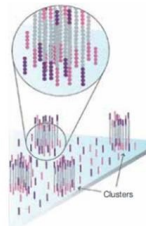
Figuur 4: DNA chip voor sequentieanalyse met Illumina MiSeq

Deze DNA chip bevat een kanaaltje waar de reagentia doorheen gepompt worden (doorstroomcel of “flowcell”). Tijdens de analyse worden individuele moleculen uit het ampliconmengsel eerst gebonden aan de doorstroomcel en vervolgens wordt er een cyclisch proces gestart (Figuur 5). Tijdens dit proces worden PCR-reagentia door de doorstroomcel geleid waarbij deze reagentia zorgen voor lokale vermeerdering van de individuele ampliconmoleculen waardoor clusters met identieke DNA-moleculen ontstaan. Ten slotte vinden cyclische sequencereacties plaats waarbij tijdens elke cyclus een mengsel van nucleotiden wordt aangeboden waarbij elk nucleotide (A, C, G of T) een fluorescente kleurstof bevat met voor elk nucleotide een eigen kleur. Tijdens elke cyclus wordt, afhankelijk van de sequentie van het gebonden ampliconmolecuul, een nucleotide ingebouwd waarbij de kleur van de fluorescente stof aangeeft welk nucleotide is ingebouwd (A, C, G of T). Aan het einde van elke cyclus wordt een foto gemaakt van de doorstroomcel en kan bepaald worden welke fluorescente kleur (welk nucleotide) is ingebouwd op alle locaties (ca. 15 miljoen locaties). Door herhaling van dit cyclische proces (maximaal 600 keer), wordt de sequentie van max. 600 baseparen van het 16S rRNA amplicon bepaald.

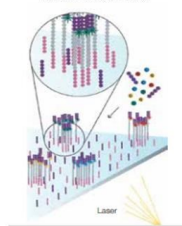
Binding van individuele moleculen aan een DNA chip



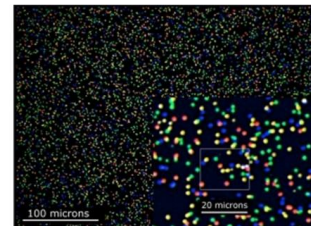
Locale vermeerdering van de individuele DNA moleculen > clusters met identieke moleculen



Bepalen van de DNA-sequentie per cluster > DNA bouwstenen met 4 fluorescente kleuren > laserscanner



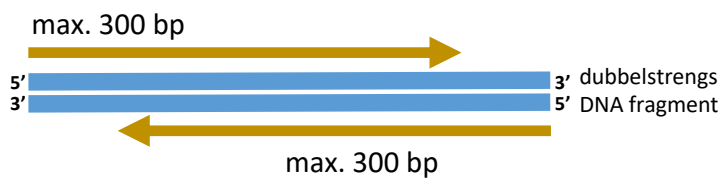
Voorbeeld van een scan
Ca. 15.000.000 “leesbare” spots



Figuur 5: basisprincipe van NGS analyses op de miseq

Na afloop van de sequencerun bepaalt de software in welke volgorde de nucleotiden op de verschillende locaties zijn ingebouwd en welke DNA sequenties er in de monsters aanwezig zijn. Tijdens dit proces wordt ook duidelijk welke indexsequenties er aan de uiteinden van de individuele moleculen aanwezig zijn. Deze informatie wordt gebruikt om vast te stellen uit welk monster een sequentie afkomstig is.

Met systemen van Illumina (zoals Illumina MiSeq) is het mogelijk om de sequentie van korte DNA-fragmenten nauwkeurig te bepalen met een foutenpercentage (error rate) van slechts ca. 0,1%. Bij het gebruik van de Illumina-systemen is het verder mogelijk om van elke streng van een dubbelstrengs DNA fragment een lengte van maximaal 300 bp aan sequentie-informatie te bepalen. Hierbij is het mogelijk om van elk fragment twee keer 300 bp sequentie-informatie te verkrijgen waarbij de keuze gemaakt kan worden om de sequentie-informatie te laten overlappen waardoor de beide runs elkaar bevestigen. Door die stap neemt de kwaliteit van de sequentiedata toe (Figuur 6). Een andere mogelijkheid is om te kiezen voor een gedeeltelijke overlap waardoor sequentieanalyse van langere fragmenten mogelijk is (met lagere kwaliteit).

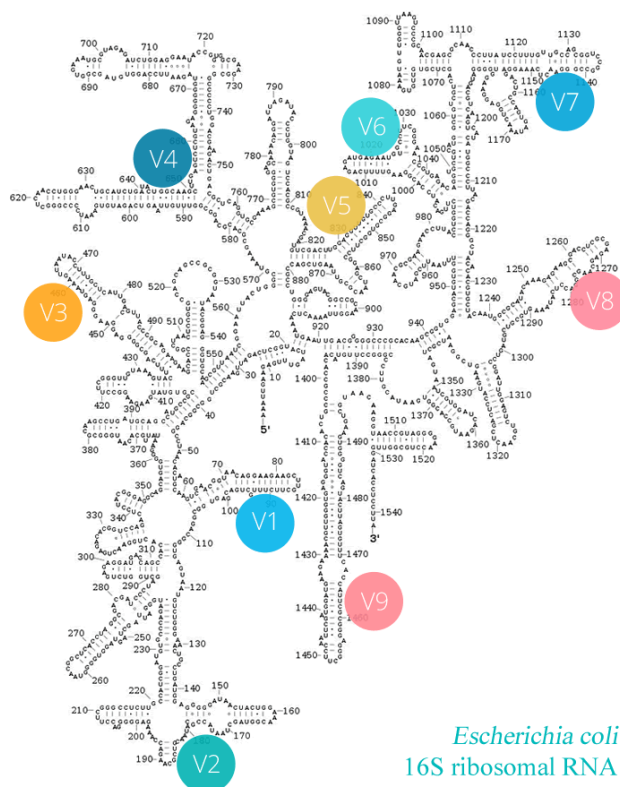


Figuur 6. Schematische weergave van sequencing van korte fragmenten (b.v. Illumina sequencing)

Als de lengte van de fragmenten zo gekozen wordt dat overlappende runs ontstaan (zoals in Figuur 6) dan wordt het mogelijk om met de resultaten van de ene run de resultaten van de andere run te verifiëren. Deze verificatie zorgt voor een kwaliteitsverbetering en resulteert, in combinatie met enkele softwarematige kwaliteitsverbeteringen, in een uiteindelijke sequentie met een foutenpercentage ("error rate") van slechts 0,007% (https://www.mothur.org/wiki/MiSeq_SOP).

3.2 Microbial profiling

Bij het toepassen van NGS analyses voor het uitvoeren van microbial profiling wordt de sequencing specifiek uitgevoerd op een deel van het 16S rRNA gen. Dit gen (schematisch weergegeven in Figuur 7), met een lengte van



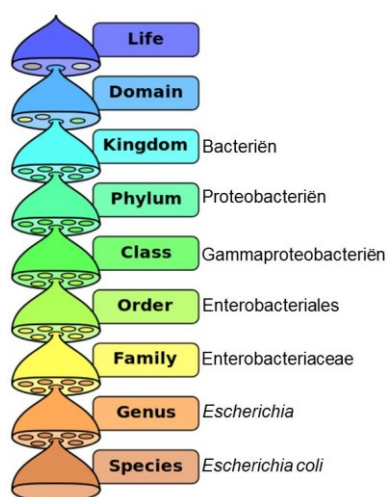
ca. 1600 baseparen (bp), is aanwezig in alle bacteriën en archaea. De samenstelling van dit gen (DNA-sequentie) bevat delen die voor alle bacteriën vrijwel gelijk zijn (geconserveerde delen) en delen met veel

Figuur 7. Schematische weergave 16S rRNA sequentie van *E. coli* met hierin de hypervariabele gebieden weergegeven (V1-V9).

Overgenomen van: <https://era7bioinformatics.com/en/page.cfm?id=2716&title=full-length-16s-taxonomic-profiling-with-pacbio> (Era7Bioinformatics)

variatie tussen de verschillende soorten (variabele delen). De sequentie van dit gen wordt daarom algemeen gebruikt als moleculaire marker voor taxonomische identificatie van de aanwezige bacterie- en/of archaeasoorten.

Voor de identificatie van de verkregen 16S rRNA gensequenties zijn databases beschikbaar met hierin de 16S rRNA gensequenties van een groot aantal bekende bacterie- en archaeasoorten. Vergelijking van de 16S rRNA gensequenties van de bacteriën uit een monster met 16S rRNA gensequenties van bekende bacteriën uit de databases, maakt het mogelijk om de bacteriën uit de te onderzoeken populatie (taxonomisch) te identificeren. Bij Illumina sequencing is alleen sequentieanalyse van relatief korte DNA-fragmenten (300-550 bp) mogelijk. Hierdoor is het vooralsnog vaak niet mogelijk om de 16S rRNA gensequenties te identificeren tot op soortniveau maar blijft identificatie beperkt tot familie of genusniveau (Figuur 8).



Figuur 8. Taxonomische indeling van organismen met, ter illustratie, een voorbeeld van de indeling van de bacteriesoort *Escherichia coli*.

Het analyseren van langere 16S rRNA fragmenten maakt het in principe wel mogelijk om sequenties tot op soortniveau te identificeren. Alternatieve sequencing methoden waarmee dit mogelijk is, zijn in opkomst, zoals MinION van Oxford Nanopore of PacBio sequencing. Echter, naast voordelen voor sequencing van langere fragmenten, staan bij andere systemen ook nadelen tegenover zoals de hoge kosten van bijvoorbeeld PacBio sequencing (Tabel 2) en vanwege de hoge foutenpercentage van Oxford Nanopore sequencing (Tabel 2) is betrouwbare identificatie van 16S sequenties met dit systeem momenteel nog niet eenvoudig (Calus, Ijaz et al. 2018, Heijnen 2019). Intussen zijn er ook methoden in opkomst waarbij het, door toepassing van “synthetische” lange reads (Callahan, Grinevich et al. 2021) ook met Illumina sequencing van korte fragmenten mogelijk lijkt om een groter deel van de sequenties tot op soortniveau te identificeren. Daarnaast zijn er nieuwe machinaal leren (machine learning) en alignment gebaseerde methoden voor taxonomische classificatie ontwikkeld waarmee identificatie van korte 16S rRNA gensequenties tot op soortniveau verbeterd is (Bokulich, Kaehler et al. 2018). De toepassingsmogelijkheden van deze nieuwe technieken zal in het BTO onderzoek van 2021 en 2022 worden onderzocht.

Tabel 2. Vergelijking van eigenschappen van verschillende sequencing systemen (overgenomen van (Donkor 2013))

Method	Single-Molecule Real-Time Sequencing (Pacific Bio)	Ion Semiconductor (Ion Torrent Sequencing)	Sequencing by Synthesis (Illumina)	Chain Termination (Sanger Sequencing)
Read length	2,900 bp	200 bp	50 to 250 bp	400 to 900 bp
Accuracy	99%	98%	98%	99.9%
Reads per run	35–75 thousand	up to 5 million	up to 3 billion	N/A
Time per run	30 min to 2 h	2 h	1 to 10 days,	20 min to 3 h
Cost per 1 million bases (in US\$)	\$2	\$1	\$0.05 to \$0.15	\$2,400
Advantages	Rapid and has longest read length.	Equipment is relatively less expensive and Fast.	Sequence yield could be very high depending upon equipment model	Long individual reads. Wide application
Disadvantages	Yield tends to be low at high accuracy. Equipment is very expensive.	Prone to homopolymer errors.	Equipment can be very expensive.	Equipment is expensive and not suitable for larger sequencing projects.
References	[46,48].	[46,47]	[43,44,46]	[39]

4 Verschillende stappen voor microbial profiling

Voor het uitvoeren van microbial profiling is het noodzakelijk om een groot aantal verschillende stappen uit te voeren waarbij verschillende keuzes gemaakt kunnen worden voor iedere stap. Veel van deze keuzes hebben invloed op de uiteindelijke analyseresultaten. Hierdoor kan het relatieve aandeel van bepaalde microbiële groepen veranderen door het maken van keuzes in elke stap van de procedure (zowel het laboratoriumdeel als het analyse- of bioinformaticadeel). Deze keuzes zullen (enige) bias introduceren in de resultaten en kunnen daarmee ook invloed hebben op de uiteindelijke conclusies die getrokken worden. De verschillen die bij het maken van keuzes kunnen optreden, hoeven waardevolle toepassing van microbial profiling zeker niet in de weg te staan. Het is van belang om bewust te zijn van de gevolgen van het maken van de keuzes op het eindresultaat en op basis daarvan weloverwogen te besluiten welke keuzes te gebruiken voor elke stap in het proces. Door de introductie van bias is het uiteindelijk van belang om een gestandaardiseerde methode toe te passen die reproduceerbare resultaten oplevert. Binnen dit hoofdstuk worden de verschillende stappen behandeld die nodig zijn om tot microbiële profielen voor drinkwatermonster te komen en worden de verschillende keuzemogelijkheden per stap besproken.

Voor het genereren van microbiële profielen uit watermonsters worden de volgende stappen uitgevoerd (zie hoofdstuk 1 voor toelichting van de verschillende gebruikte termen) en deze zullen per subparagraaf worden behandeld in dit hoofdstuk:

- Bemonstering, concentrering en conservering
 - Monstername
 - Concentreren monster
 - Conservering monster
- Isolatie van DNA van de microbiële gemeenschap
 - Algemeen
 - DNA-isolatiemethoden
 - DNA-concentratie
- Het genereren van een “fragmentbibliotheek” met 16S rRNA fragmenten (“bibliotheekbereiding”)
 - Primerkeuze en PCR 16S rRNA gen
 - Genereren van PCR-fragmenten voor Illumina sequencing
- Bioinformatica voor dataverwerking en interpretatie:
 - Dataverwerking met QIIME2
 - Demultiplexing
 - Joining pairs
 - Ruisonderdrukking (“denoising”)/clustering
 - Taxonomische classificatie en fylogenie
 - Diversiteitsanalyses alpha- en betadiversiteit
 - Differentiële abundantie
 - Barplots/heatmaps en meer statistiek
 - QIIME2 viewer

4.1 Monstername, conservering en concentratie

4.1.1 Monstername

De voorwaarden voor het nemen van monsters voor het uitvoeren van microbial profiling verschillen weinig van de voorwaarden voor het nemen van monsters voor veel andere moleculair-microbiologische analyses (o.a. qPCR). De belangrijkste criteria zijn:

- a) De microbiologie in het genomen monster dient representatief te zijn voor de situatie.
 - Afhankelijk van de situatie kan het tijdstip, de locatie of de manier van monstername (vooral voor de homogeniteit) kritisch zijn.
- b) Er is speciale aandacht nodig om de introductie van microbiële contaminaties te voorkomen.
 - De introductie van dode en levende micro-organismen alsook het DNA van micro-organismen moet zoveel als mogelijk worden voorkomen.
- c) Om inzicht te krijgen in eventuele variabiliteit van de monstername is het van belang om analyses op meerdere identieke monsters uit te voeren.
 - Minimaal drie biologische (monster-)replica's wordt sterk geadviseerd.

Omdat met het gebruik van microbial profiling vrijwel de volledige microbiële populatie in een watermonster zichtbaar gemaakt wordt, is er speciale aandacht nodig om te voorkomen dat andere micro-organismen (ook van geïnactiveerde/dode) of het DNA van andere micro-organismen worden geïntroduceerd. Vooral bij de analyse van monsters met weinig biomassa (zoals drinkwater) kan de invloed van contaminaties relatief groot zijn en moet daarom zoveel mogelijk worden voorkomen. Dit betekent dat speciale aandacht nodig is voor de materialen waarmee de monstername plaatsvindt (zoals o.a. monsterflessen). Het heeft de voorkeur om de gebruikte materialen bij de monstername zodanig te behandelen dat deze geen organismen of DNA bevatten. Het gebruik van steriel en DNA-vrij wegwerpmateriaal heeft de voorkeur. Toevoegingen aan monsterflessen, zoals met kweekmethoden gebruikelijk is (zoals thiosulfaat voor het neutraliseren van chloor), dienen te worden vermeden. Deze toevoegingen kunnen invloed hebben op de microbiologische populaties, zoals aangetoond voor thiosulfaat in onderzoek uitgevoerd in DPWE verband en vastgelegd in een memo (Heijnen 2019).

4.1.2 Concentreren van monsters

Vanwege de beperkte hoeveelheid biomassa in drinkwatermonsters is het van belang om de micro-organismen te concentreren uit een voldoende groot volume water. Voor het concentreren van micro-organismen uit watermonsters zijn er verschillende methoden beschikbaar. De algemeen toegepaste membraanfiltratiemethode is ook voor microbial profiling een veel toegepaste methode. Voor monsters met weinig biomassa kan het waardevol zijn om grotere volumes te concentreren met ultrafiltratie (Hemoflow) waarmee voldoende biomassa in het concentraat wordt verzameld. Bij de keuze voor de concentratiemethode is het van belang om een methode te selecteren waarmee het mogelijk is om voldoende DNA te isoleren. In de paragrafen 4.2.3 en 5.2 is experimenteel uitgezocht welke minimale DNA concentratie nodig is voor het genereren van betrouwbare resultaten.

4.1.3 Conserveren van monsters

Er zijn maatregelen nodig om te garanderen dat de microbiota die bij de analyses wordt onderzocht, representatief is voor de microbiota die in de praktijk in het water aanwezig is. Deze maatregelen zijn vooral gericht op het stabiliseren van de aanwezige microbiële gemeenschap door het voorkomen van groei en afsterving in de periode tussen monstername, monsterfiltratie en DNA-isolatie.

4.2 Isolatie van DNA van de microbiële gemeenschap

4.2.1 Algemeen

Voor het uitvoeren van betrouwbare microbial profiling analyses is voldoende DNA nodig en moet de samenstelling van het geïsoleerde DNA representatief zijn voor de bacteriële populatie in het monster. Vanwege de beperkte hoeveelheid biomassa in drinkwatermonsters is het van belang om een isolatiemethode te gebruiken met voldoende opbrengst. Naast voldoende opbrengst en concentratie is het belangrijk dat het DNA van verschillende soorten micro-organismen met dezelfde efficiëntie wordt geïsoleerd waardoor zo weinig mogelijk bias wordt geïntroduceerd t.g.v. de toegepaste isolatiemethode. Vanwege verschillen in samenstelling van de celwand van bacteriesoorten heeft vooral de stap waarbij het DNA wordt vrijgemaakt uit de bacteriële cellen (lysis) een belangrijke invloed op verschillen die kunnen optreden in de efficiëntie van DNA-isolatie van verschillende bacteriesoorten. Voor het verkrijgen van een goed beeld van de samenstelling van de microbiota en de relatieve

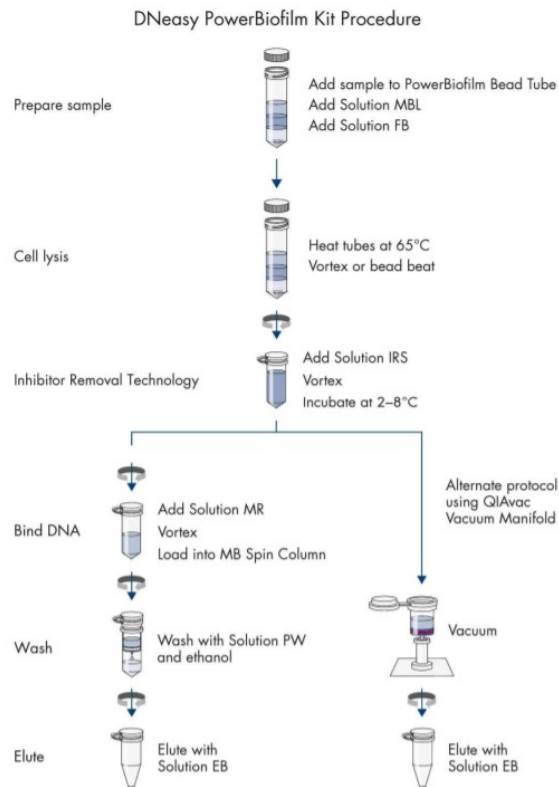
kwantitatieve verschillen tussen groepen van micro-organismen is het daarom van belang om een isolatiemethode toe te passen waarmee er weinig isolatieverschillen optreden tussen groepen van micro-organismen. Voor lysis van de microbiële cellen worden deze vaak behandeld met chemische stoffen (detergentia zoals triton-X100 of SDS of guanidine-iso-thiocyanaat) of enzymen (proteases zoals Proteinase-K of Lysozym). Met deze behandelingen worden eiwitten in de celwand aangetast en kunnen de nucleïnezuren uit de gelyseerde cellen vrijkomen. Daarnaast worden ook mechanische methoden gebruikt waarbij glasbeads worden toegevoegd aan het monster waarna, door zeer krachtig schudden met een speciaal voor dit doel ontwikkelde machine (bead-beater, Figuur 9), de celwanden mechanisch (door optredende botsingen tussen beads en cellen) worden aangetast (bead-beaten) en het DNA vrijkomt. Het gebruik van bead-beating blijkt vooral van belang voor het vrijmaken van DNA uit “grampositieve” bacteriën (Albertsen, Karst et al. 2015). Grampositieve bacteriesoorten bezitten een fysiek sterke celwand, samengesteld uit peptidoglycaan, waardoor bead-beating nodig is voor het efficiënter vrijmaken van DNA uit deze bacteriecellen.



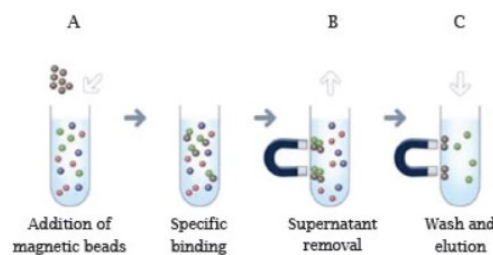
Figuur 9. Voorbeeld van een bead-beater

Na het vrijkomen van het DNA zijn er verschillende methoden beschikbaar waarmee, door toepassing van affiniteitszuivering, het DNA wordt geïsoleerd, geconcentreerd en wordt ontdaan van verontreinigingen die storend kunnen werken op het verdere proces (zoals inhibitors die kunnen zorgen voor remming van PCR reacties). De basis van deze isolatiemethoden wordt vooral gevormd door een affiniteitszuivering, waarbij silica wordt toegepast voor binding van DNA. Dit type zuivering kan plaats vinden in kolomformaat (Figuur 10a) of door gebruik te maken van silicabeads met een magnetische kern (Figuur 10b). Het gebruik van zuiveringen in kolomformaat is vooral interessant voor het onderzoeken van kleinere monsterseries, het gebruik van zuivering met magnetische silicabeads is relatief eenvoudig te automatiseren en daardoor beter bruikbaar voor “high throughput” toepassingen.

A.



B.



Figuur 10. Affiniteitszuivering van DNA met kolom- (A) of magnetische-beads (B), bron voor de figuur: Qiagen

4.2.2 Effect van DNA-isolatiemethoden op microbiële profielen

Er is vooral veel onderzoek gedaan naar het effect van DNA-isolatiemethoden op de analyseresultaten van de microbiota in de darm (feces). Een vergelijking van vijf DNA-isolatiemethoden op de microbial profiling resultaten van darmbacteriën liet een effect zien van de gebruikte isolatiemethode (Rintala, Pietilä et al. 2017). Deze studie liet echter ook zien dat dit effect kleiner was dan het effect van verschillende primerkeuzes op de microbiële profielen (zie ook paragraaf 4.3.1). Een andere studie (Mallott, Malhi et al. 2019) liet zien dat grampositieve bacteriën minder dominant worden aangetroffen in de profielen bij het gebruik van een DNA-isolatiemethode waarbij geen bead-beat stap wordt gebruikt (Qiaamp DNA stool kit van Qiagen) dan bij een methode waarbij wel bead-beating wordt toegepast (PowerSoil Kit van MoBio). Ook de studie van Lim et al (Lim, Song et al. 2018) maakt duidelijk dat de toepassing van bead-beating resulteert in een vollediger beeld van de microbiële samenstelling van de darmmicrobiota.

Het aantal studies naar het effect van DNA-isolatie op de microbiota in water is beperkt. In recent onderzoek werd het gebruik van de FastDNA kit for soil (MP Biomedicals) vergeleken met de PowerWater DNA isolation Kit (Qiagen, voorheen MoBio). Dit onderzoek liet zien dat meer DNA werd geïsoleerd met de PowerWater kit dan met de FastDNA kit (Brandt and Albertsen 2018). Met beide methoden bleek het wel mogelijk om microbiële profielen te genereren waarbij significante verschillen in de samenstelling van het microbiota werden waargenomen tussen de beide DNA-isolatiemethoden. De profielen verkregen na isolatie met de FastDNA kit blijken minder goed

reproduceerbaar en levert een lagere diversiteit op dan met de PowerWater kit, waardoor te verwachten is dat het beeld over de samenstelling van het microbiota met de FastDNA kit minder volledig is (Brandt and Albertsen 2018). Eerder onderzoek uitgevoerd bij KWR liet ook zien dat de DNA-opbrengst met de FastDNA kit aanmerkelijk lager is dan met de PowerWater en PowerBiofilm kits van MoBIO (Italiaander, Heijnen et al. 2011).

Optimale DNA-isolatiemethode

Bovenstaande inventarisatie laat zien dat voor microbiotaonderzoek robuuste DNA-isolatiemethoden moeten worden toegepast om een representatief beeld te krijgen van de samenstelling van de microbiële gemeenschap, zodat de introductie van bias in de analysesresultaten wordt geminimaliseerd. Uit eerdere onderzoeken blijkt dat een combinatie van chemische, eventueel aangevuld met een enzymatische behandeling, én bead-beaten het meest representatieve beeld geeft van de samenstelling van de microbiota. Vooral het gebruik van bead-beaten is van belang om ook het DNA uit grampositieve bacteriesoorten vrij te maken en daarmee te voorkomen dat deze bacteriesoorten, ten onrechte, minder goed worden gerepresenteerd in de profielen. Voor het onderzoek naar de microbiota in (drink)water is de keuze gemaakt om de DNeasy Power Biofilm Kit van Qiagen te gebruiken voor DNA-isolatie. Bij deze methode wordt een combinatie van bead-beating en chemische lysis gebruikt voor het vrijmaken van het DNA uit de cellen. Het vrijgemaakte DNA wordt vervolgens met hoge opbrengst gezuiverd door toepassing van kolomzuivering (Italiaander, Heijnen et al. 2011).

4.2.3 Effect van DNA concentratie op microbiële profielen

Voor het onderzoek naar de samenstelling van de microbiota in drinkwater heeft men onder reguliere praktijkcondities te maken met water waarin relatief weinig biomassa aanwezig is. Dergelijke lage biomassaconcentraties in de te onderzoeken monsters zorgt ervoor dat de introductie van contaminaties (via laboratoriumreagentia, -artikelen of -handelingen of kruiscontaminaties met DNA uit andere monsters) relatief veel invloed kan hebben op de uiteindelijke microbiële profielen, waardoor verkeerde conclusies getrokken kunnen worden over de aanwezige micro-organismen in het monster (Eisenhofer, Minich et al. 2019). Dit maakt de analyse van controles, te weten blancomonsters (DNA-vrij water; inclusief DNA-extractie) en monsters zonder DNA-extractie (“no-template controles”) van groot belang (Eisenhofer, Minich et al. 2019). In het artikel van Eisenhofer en Minich wordt ook een overzicht gegeven van de microbiële taxa die bij verschillende studies gevonden zijn in blancomonsters en no-template controles. In paragraaf 5.1 wordt beschreven welke taxa zijn aangetroffen in het onderzoek van blancomonsters bij KWR.

Naast de introductie van DNA-contaminatie is er bij de analyse van monsters met weinig biomassa ook het risico dat de concentratie van de minder dominante organismen te laag is om nog zichtbaar te kunnen maken in de profielen waardoor een onvolledig beeld ontstaat van de samenstelling van de microbiële populatie. Onderzoek naar het effect van de concentratie van DNA voor het genereren van betrouwbare microbiële profielen uit humane feces laat zien dat slechts weinig feces nodig is (2 mg) voor het genereren van reproduceerbare profielen en dat, bij de onderzochte concentraties, de toegepaste isolatiemethode meer invloed heeft op het profiel dan de concentratie DNA (Velásquez-Mejía, de la Cuesta-Zuluaga et al. 2018). Het is echter niet duidelijk of dit ook het geval is voor monsters met weinig biomassa zoals drinkwatermonsters. Om uit monsters met weinig biomassa toch microbiële profielen te genereren, is het mogelijk om meer cycli toe te passen voor het uitvoeren van de PCR-reacties om daarmee meer PCR-fragmenten te genereren. Bij het toepassen van meer cycli neemt de kwaliteit van de sequenties echter af (hogere foutenpercentage) en wordt de invloed van eventuele contaminaties groter, waardoor extra aandacht nodig is voor de vergelijking met negatieve controles (blanco monsters en “no template controles”) (Witzke, Gullic et al. 2020). Om de invloed van contaminaties bij de analyse van monsters met weinig biomassa te minimaliseren zijn er ook mogelijkheden om reagentia te behandelen met enzymen die DNA afbreken (DNase), dit blijkt effectief om 99% van de contaminerende sequenties te verwijderen (Stinson, Keelan et al. 2019).

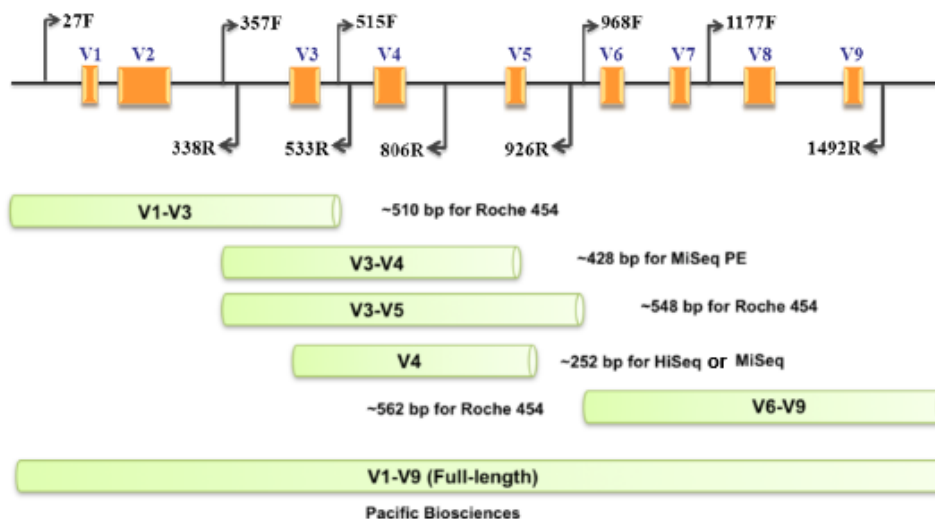
Er kan algemeen gesteld worden dat de analyse van monsters met weinig biomassa uitdagingen met zich meebrengt. Om deze problemen te beperken heeft het de voorkeur om te streven naar het uitvoeren van analyses op monsters met voldoende biomassa. In hoofdstuk 6 wordt een studie beschreven die bij KWR is uitgevoerd en waarbij is onderzocht bij welke DNA-concentraties van het drinkwater van productielocatie Tull en 't Waal (bemonsterd aan de tap van het KWR-laboratorium) de microbiële profielen nog steeds van goede kwaliteit zijn.

4.3 Het genereren van een “fragmentbibliotheek” met 16S rRNA fragmenten

4.3.1 Primerkeuze en PCR 16S rRNA gen

Het 16S rRNA gen bestaat uit delen waar de sequentie tussen verschillende soorten micro-organismen zeer weinig varieert (geconserveerde gebieden) en delen die zeer veel sequentieverschillen bevatten (hypervariabele gebieden, zie Figuur 7).

Om voldoende onderscheid tussen genera of soorten van micro-organismen mogelijk te maken wordt er bij microbial profiling voor gekozen om de sequentie te bepalen van een hypervariabel gebied van het 16S rRNA gen. Om te zorgen dat zoveel mogelijk bacteriesoorten uit de populatie kunnen worden gedetecteerd, wordt tevens gekozen om de PCR-primers te laten binden aan geconserveerde sequenties van het 16S rRNA gen. De sequentie van de geconserveerde gebieden zijn echter nooit volledig identiek tussen alle bacteriesoorten en de primers welke worden gebruikt om te binden aan het geconserveerde deel van het 16S rRNA gen (overzicht van veel gebruikte primerparen: Figuur 11) zullen daardoor suboptimaal binden aan het DNA van een beperkt aantal bacteriesoorten. Met de keuze van primers wordt dus altijd bias in de resultaten van de microbiële profielen geïntroduceerd. Onderzoek heeft laten zien dat de invloed van de keuze van de primerset een grotere bias introduceert in de resultaten dan de keuze van de DNA extractiekit (Albertsen, Karst et al. 2015, Wasimuddin, Schlaeppli et al. 2019). De bias veroorzaakt door primerkeuze hoeft echter geen probleem te zijn voor de toepassingsmogelijkheden van microbial profiling voor het drinkwateronderzoek. Het is dan wel van belang om te kiezen voor primers waarmee de, voor de onderzoeksvraag, relevante genera, efficiënt worden gedetecteerd en bewust te zijn van de door primerkeuze geïntroduceerde bias en kennis te hebben van de gevolgen van deze bias op detectie van bepaalde microbiële groepen. Daarnaast zorgt deze bias ervoor dat men zeer voorzichtig moet zijn met het combineren van data uit studies waarbij verschillende primerparen zijn gebruikt voor het genereren van data.



Figuur 11. Overzicht van veelgebruikte primerparen en de variabel delen van het 16S rRNA gen die met de primerparen worden geamplificeerd t.b.v. microbial profiling (bron: Pacific Biosciences of California, Inc.)

4.3.2 Genereren van PCR-fragmenten voor Illumina sequencing

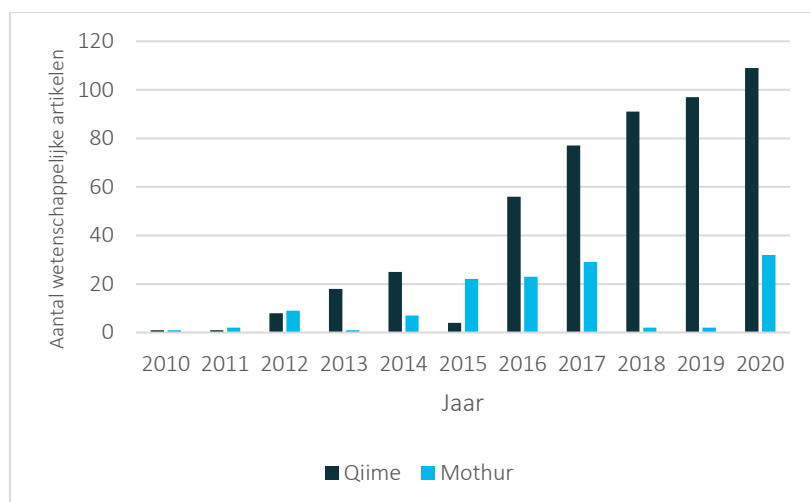
Bij het gebruik van het Illumina-systeem worden de PCR-fragmenten aan de uiteinden voorzien van sequenties waarmee de fragmenten aan de chip binden (Figuur 4) en die sequenties dienen als startpunt voor de sequencereactie. Daarnaast worden de fragmenten voorzien van indexsequenties op basis waarvan de

verschillende gemengde monsters in een sequencerun kunnen worden onderscheiden (paragraaf 3.1.1). Om praktische redenen is binnen het onderzoek gekozen om het genereren van PCR-fragmenten voor sequencing in twee stappen uit te voeren: eerst een PCR op het 16S rRNA gen en vervolgens een PCR waarmee de uiteinden worden verlengd. In bijlage II is het protocol weergegeven dat bij KWR in gebruik is voor het uitvoeren van de laboratoriumhandelingen.

4.4 Bioinformatica voor dataverwerking en interpretatie

4.4.1 Platforms, software en databases

Na het daadwerkelijke sequencen begint het meest arbeidsintensieve werk: de bioinformatica bestaande uit dataverwerking, validatie en analyse van de verkregen sequenties. Deze analyses worden meestal gedaan in Linux-gebaseerde omgevingen waarbij binnen een pijplijn ('pipeline') meerdere scripts worden toegepast. Deze pijplijn houdt dus in dat opeenvolgende softwarescripts alle dataverwerking, validatie en analyse van de sequenties uitvoeren. Er zijn twee veelgebruikte pijplijnsoftwareprogramma's in omloop hiervoor: MOTHUR (Schloss, Westcott et al. 2009) en Qiime2 (Caporaso, Kuczynski et al. 2010, Bolyen, Rideout et al. 2018). Eerder onderzoek heeft laten zien dat de invloed van de gebruikte pijplijnen database op de uiteindelijke verkregen microbiële profielen groot kan zijn (Siegwald, Caboche et al. 2019). Hierdoor is het belangrijk om een goed onderbouwde keuze te maken voor de te gebruiken pijplijnsoftware. Bij KWR was in eerste instantie gekozen voor het gebruik van het programma 'MOTHUR', omdat het destijds werd gezien als een betrouwbaardere pijplijnsoftware dan Qiime1 (de voorloper van Qiime2). Tijdens dit project is echter besloten om over te stappen naar het ondertussen ontwikkelde en sterk verbeterde Qiime2 (als opvolger van Qiime1). Er waren meerdere redenen om de overstap van MOTHUR naar Qiime2 te maken. De doorontwikkeling van Qiime2 gaat momenteel vrij snel terwijl MOTHUR weinig wordt doorontwikkeld. Bovendien wordt het beheer en de doorontwikkeling van MOTHUR gedaan door een kleine onderzoeksgroep, waardoor dit pakket kwetsbaar lijkt om op de lange termijn beschikbaar te blijven. In tegenstelling hiermee wordt de basis van Qiime2 gevormd door ontwikkelaars van meerdere onderzoeksgroepen die gesteund worden door een grote gemeenschap van gebruikers. Deze gebruikers en ontwikkelaars ontwikkelen continu nieuwe tools en software die geïntegreerd worden in het Qiime2 platform via zogenaamde "plugins". Tevens heeft Qiime2 een gebruikers/ontwikkelaarsforum waar probleemoplossing en vragen snel en efficiënt beantwoord worden. Verder heeft Qiime2 in tegenstelling tot Mothur veel handige visualisatiemogelijkheden (online en tussen gebruikers eenvoudig uitwisselbaar) en statistische analyses geïntegreerd. Deze eigenschappen maken Qiime2 momenteel tot de meest gebruikte analysesoftware voor 16S rRNA genanalyses (Figuur 12) en de taxonomische identificatie van 16S rRNA gensequenties blijkt met Qiime2 nauwkeuriger dan met MOTHUR (Almeida, Mitchell et al. 2018). Tot slot is Qiime2 software ook bruikbaar voor de analyse van andere taxonomische genen voor identificatie van eukaryote organismen zoals schimmels en protozoën (op basis van het 18S rRNA gen en/of het ITS gen). Voor het runnen van Qiime2 bij KWR wordt gebruik gemaakt van een Linux server. Het gebruik van Qiime2 verloopt via commando's via een "command line".



Figuur 12. Aantal wetenschappelijke artikelen waarin Qiime en MOTHUR worden genoemd. In 2018 is Qiime 2 als opvolger van Qiime 1 beschikbaar gekomen.

Aanvullend wordt bij KWR voor visualisatie van data en het uitvoeren van additionele statistische analyses gebruik gemaakt van pakketten die ontwikkeld zijn om te gebruiken binnen het statistische programma “R” (R-packages). Deze pakketten zijn ontwikkeld door verschillende onderzoeksgroepen en zijn relatief eenvoudig toe te passen. Het gebruik van “R-packages” verloopt via “R-studio”, een omgeving waarbinnen de R-scripts relatief eenvoudig kunnen worden gerund. R en R-studio werken zowel binnen een Windows- en Linuxomgeving.

4.4.2 Dataverwerking met Qiime2

De dataverwerking en interpretatie van de sequenties bestaat uit een aantal opeenvolgende stappen, waarbij bij elke stap weloverwogen keuzes gemaakt moeten worden om tot een betrouwbaar resultaat te komen. Elke keuze brengt daarbij nieuwe bias met zich mee waarvan de voor- en nadelen daarom goed afgewogen moeten worden. Het is van belang om te weten wat deze stappen precies inhouden zodat bij datainterpretatie rekening wordt gehouden met de gekozen bias. De volgende stappen vinden plaats in de Qiime2 pijplijn die bij KWR gebruikt wordt:

1. Demultiplexing
2. Joining pairs
3. Ruisonderdrukking (denoising)/ Clustering
4. Taxonomische classificatie en fylogenie
5. Diversiteitsanalyse (alpha en beta)
6. Differentiële abundantie, barplots/heatmaps en statistiek

In de volgende paragrafen worden deze stappen binnen het Qiime2 platform kort uitgelegd. Gedetailleerde informatie kan teruggelezen worden op de Qiime2 website: <https://docs.qiime2.org/2020.8/tutorials/>

4.4.3 Demultiplexing

Demultiplexing is het proces waarbij de sequenties uit de verschillende monsters, die in één Miseq sequencing run opgebracht waren, gescheiden worden. Normaliter worden in één sequencing run een zeer groot aantal verschillende monsters gemengd om kosten en tijd te besparen. Zoals beschreven in 3.1.1 zijn de uiteinden van alle verkregen PCR fragmenten van elk uniek monster voorzien van specifieke “indexsequenties” voordat de PCR-fragmenten van verschillende monsters gemengd (multiplexed) worden. Hierdoor heeft elk verkregen sequentie een indexsequentie gekregen die direct gekoppeld is aan het monster waar deze sequentie uit is gekomen. Deze indexsequenties worden vervolgens bij het demultiplexing gebruikt om de sequenties te scheiden naar ieder uniek monster waaruit ze afkomstig waren. De demultiplexing softwarestap van Qiime2 levert voor elk monster twee computerbestanden op: een forward reads en een reverse reads bestand. Daarnaast wordt ook een overzicht

gegeven van het aantal reads per monster en de kwaliteitscore voor elke base die gesequencet is (zogenaamde phred scores).

4.4.4 Joining pairs

Na het demultiplexen kunnen per monster de forward en reverse reads aan elkaar geplakt worden tot 1 read. Deze stap heet 'joining pairs'. Met de sequenceapparaten van Illumina (zoals de Illumina MiSeq) is het mogelijk om de sequentie van korte DNA-fragmenten zeer nauwkeurig te bepalen met een foutenpercentage (error rate) van slechts ca. 0,1%. Hierdoor is het mogelijk om bij het gebruik van de sequenceapparaten van Illumina voor elke streng van een dubbelstrengs DNA fragment een lengte van maximaal 300 bp aan sequentie-informatie te bepalen. **6**).

Als de lengte van de fragmenten zo gekozen wordt dat de forward en reverse read vrijwel compleet overlappen (zoals in Figuur 6), dan wordt het mogelijk om met de resultaten van de forward streng de resultaten van de reverse streng te verifiëren. Deze verificatie zorgt voor een extra kwaliteitsverbetering en resulteert, in combinatie met enkele softwarematige kwaliteitsverbeteringen, in een uiteindelijke sequentie met een foutenpercentage van slechts 0,007% (https://www.mothur.org/wiki/MiSeq_SOP). Dit lage foutenpercentage is vooral van belang bij het gebruik van clusteringmethoden op basis van het percentage sequentieovereenkomst. Deze methode wordt toegepast om operational taxonomic units (OTU's) (soorten) te definiëren. Bij OTU clustering zullen eventuele sequentiefouten daarom resulteren in (onterechte) toename van het aantal OTU's en daarmee overschatting geven van de microbiële diversiteit. Om deze reden is besloten om primers te gebruiken die vrijwel compleet overlappende runs genereren waardoor de kwaliteit van deze "dubbel gelezen" sequentie zeer goed is. Bij het gebruik van ASV's voor het clusteren van sequenties is deze hoge kwaliteit van de sequenties van minder belang (zie paragraaf 4.4.5). In Qiime 2 wordt deze stap uitgevoerd door beide sequenceruns eerst te combineren (via de "joining pairs" stap) en aansluitend een kwaliteitscontrole uit te voeren met de tool 'vsearch'. Het resultaat is dat sequenties worden verwijderd die niet genoeg overlap vertonen of die te lage kwaliteitscores (phred scores, (Ewing and Green 1998)) hebben. De waarde van kwaliteitscore waarbij sequenties verwijderd worden kan men zelf bepalen.

4.4.5 Ruisonderdrukking/clustering

Ruisonderdrukking (Denoising) is een stap waarbij, zoals de term zegt, de 'ruis' uit de data gehaald wordt. Dit is de stap waar de grootste slag wordt gemaakt in kwaliteitsfiltering van de sequentiereads. Er zijn verschillende opties mogelijk voor ruisonderdrukking in Qiime2, zoals de DADA2 software (Callahan, McMurdie et al. 2016) of de Deblur software (Amir, McDonald et al. 2017). Beide tools verwijderen slechte sequentiereads met veel fouten, chimeersequenties (sequenties die tijdens de PCR-reacties zijn ontstaan uit 16S rRNA fragmenten van twee verschillende bacteriesoorten), singletons (16S rRNA genfragment waarvan de DNA-sequentie maar één keer is aangetroffen in het monster), en passen tevens dereplicatie toe (het identificeren van unieke sequenties waarbij alle kopieën daarvan worden verwijderd). Wanneer gekozen wordt voor de Deblur software, moet er eerst een kwaliteitsfilteringstap ondergaan worden (met q scores), bij de DADA2 software is dit niet nodig. In het algemeen geldt dat als je meerdere monsters die in verschillende sequencing runs geanalyseerd zijn wilt vergelijken, het beter is om Deblur te gebruiken. Daarnaast is het sneller dan DADA2. DADA2 heeft de voorkeur wanneer alle monsters in 1 run gesequencet zijn. Omdat bij KWR monsters vaak over verschillende runs gesequencet worden, gebruiken wij standaard Deblur.

Clustering is een methode waarbij sequentiereads gegroepeerd worden in (OTU's) die op elkaar lijken (waarbij $\geq 97\%$ overeenkomst wordt gebruikt als arbitraire grens voor het onderscheiden van sequenties op bacterieel soortniveau). Dit proces heet OTU picking en werd bij KWR gebruikt bij het verwerken van sequentiedata met het MOTHUR platform. Momenteel neemt het gebruik van OTU clustering door onderzoekers sterk af en wordt deze methodiek sinds 2020 bij KWR niet meer gebruikt. Tegenwoordig wordt alleen ruisonderdrukking uitgevoerd op de afzonderlijke sequentiereads met Deblur. Hierdoor worden dus geen OTU's meer verkregen, maar zogenaamde amplicon sequence variants (ASV's).

ASV's worden dus niet via clustering op basis van 97% sequentieovereenkomst verkregen. Deze ASVs worden verkregen door eerst een (softwarematige) kwaliteitsverbetering van de sequenties uit te voeren en door daarna alleen unieke sequenties te clusteren (Alle ASV's met 100% sequentieovereenkomst) en wordt elke unieke ASV gezien als een aparte cluster. Dit hele proces wordt uitgevoerd tijdens ruisonderdrukking van de sequentiereads met behulp van de Deblur software. Een meer gedetailleerde toelichting over het verschil tussen het gebruik van OTU's en ASV's wordt hier weergegeven: <https://www.zymoresearch.com/blogs/blog/microbiome-informatics-otu-vs-asv>.

Het gebruik van ASV's is minder gevoelig voor de aanwezigheid van sequentiefouten (die altijd, maar bij Illumina sequencing wel beperkt, aanwezig zijn). Doordat geen clustering plaats vindt op basis van 97% sequentieovereenkomst (zoals bij OTU clustering het geval is) kunnen ook kleine, maar soms relevante, sequentieverschillen resulteren in verschillende taxonomische identificaties waardoor, met het gebruik van ASV's, een groter deel van de sequenties kan worden geïdentificeerd tot op genus- of soortniveau (species). De voordelen die clustering op basis van ASV's heeft t.o.v. clustering van OTU's heeft er dus voor gezorgd dat er een verschuiving is opgetreden van software voor OTU clustering naar software voor ASV clustering.

Het resultaat van de ruisonderdrukkingsoftware binnen Qiime2 zijn twee computerbestanden, een bestand waar voor elke ASV de frequentie per monster wordt gegeven (feature table) en een ander bestand waar de sequentie voor elke ASV wordt weergegeven (feature data).

4.4.6 Taxonomische classificatie en fylogenie

Door vergelijking van de ASV-sequenties met referentiesequenties van het 16S rRNA gen van verschillende micro-organismen die zijn opgenomen in een internationale database, worden de ASV's vervolgens taxonomisch geïdentificeerd. Dit wordt gedaan met "machine-learning based classification". Ook hier is de keuze voor machine-learning based classification methoden weer bepalend voor het resultaat. Het gebruik van de juiste machine learning methode kan resulteren in betere classificatie, namelijk meer sequenties tot op soortniveau. Bij KWR gebruiken we machine-learning based classification methoden die getest zijn op zoetwatermonsters (Bokulich, Kaehler et al. 2018). In de toekomst willen we zelf een eigen machine learning methode ontwikkelen die getest is op drinkwatermonsters, zodat we nog meer sequenties tot op soortniveau kunnen bepalen.

De keuze voor de database met referentiesequenties van het 16S rRNA gen van bacteriën bepaalt hoe de ASV's geïdentificeerd worden. Er zijn verschillende databases beschikbaar, zoals Greengenes (<https://greengenes.secondgenome.com/>), The ribosomal database project (RDP, <http://rdp.cme.msu.edu/>) en SILVA (<https://www.arb-silva.de/download/arb-files/>). Ervaring van het onderzoeksproject beschreven in deze rapportage heeft aangetoond dat bij gebruik van de Greengenes database met ongetrainde classifier, geen enkele readsequentie toegewezen werd aan het genus *Escherichia*, maar slechts maximaal tot de familie *Enterobacteriaceae*. Dit was onder andere het geval voor rioolwater- en RWZI-effluentmonsters, waar met volledige zekerheid *Escherichia coli* in aanwezig is. Ook wanneer we handmatig de readsequenties bekeken met BLAST searches, werden deze met 100% sequentieovereenkomst toegewezen aan *E. coli*. Wanneer alleen de readsequenties verkregen met de forward primer getrimd werd (en je dus langere sequentiereads van >250 bp overhoudt), was de classificatie met de Greengenes database wel voldoende.

De meest recente SILVA 16S rRNA database (tijdens het schrijven van dit rapport is dat versie 138.1 uit augustus 2020)(Yilmaz, Parfrey et al. 2014)) geeft een betere resolutie dan de Greengenes database. Daarbij leidde de SILVA database vaker tot classificatie op genus- en zelfs soortniveau: 70% van de sequentiereads wordt met deze combinatie tot soortniveau geïdentificeerd. Daarom wordt door KWR de SILVA database gebruikt, ook omdat continu nieuwe taxons worden beschreven (ook uit drinkwater) en omdat taxons continu opnieuw worden geïdentificeerd en aan de SILVA database worden toegevoegd, terwijl de andere databases daar niet up-to-date in zijn. Bij elke NGS analyse wordt de meest recente SILVA 16S rRNA database gebruikt.

Om de relatie tussen de taxons van de ASV's te beschrijven, kan de fylogenie bepaald worden. De fylogenie is de afstammingsgeschiedenis van een groep (micro-)organismen en wordt bepaald door een stamboom te maken waarbij de afstand tussen de takken de evolutionaire afstand weergeeft. Dit kan later gebruikt worden in analyses waarbij de fylogenie van belang is. Om de fylogenetische afstand tussen ASV's en referentiesequenties te bepalen, moeten de 16S rRNA gensequenties 'aligned' (uitgelijnd) worden. Dit uitlijnen houdt als het ware in dat alle 16S rRNA gensequenties met elkaar vergeleken worden, waarbij de overeenkomsten en de verschillen van nucleotiden tussen de sequenties vastgelegd worden. Deze overeenkomsten en verschillen zijn beide bepalend voor de evolutionaire afstand die aan de hand hiervan berekend wordt.

4.4.7 Diversiteitsanalyses alpha- en beta-diversiteit

Alpha-diversiteit meet de mate van diversiteit *binnen* individuele monsters. Beta-diversiteit meet de mate van diversiteit of verschil *tussen* monsters. Voor beide analyses is het belangrijk dat de data eerst genormaliseerd wordt, anders is een vergelijking niet mogelijk. Elk monster bevat namelijk, door technische factoren, een verschillend aantal totale sequentiereads. Wanneer een monster meer sequentiereads heeft dan de ander, dan is de kans groot dat de diversiteit hoger is in het monster met meer sequentiereads. Door te normaliseren kan hiervoor gecorrigeerd worden. Deze normalisatie kan door middel van het omzetten van de absolute abundantiedata naar relatieve abundantie of door 'rarefaction' uit te voeren. In het algemeen heeft rarefaction de voorkeur. Rarefaction houdt in dat uit elk monster een vast aantal keer een willekeurig submonster genomen wordt en hierop de analyses gedaan worden. Hierbij wordt voor elk monster dezelfde minimum sampling diepte gekozen, waardoor genormaliseerd wordt voor de verschillen in aantal sequencing reads tussen monsters. Deze keuze hangt af van het minimum aantal sequentiereads van de monsters die worden vergeleken. Omdat het subbemonsteringsproces willekeurig is, zal elke analyse op dezelfde data net wat andere resultaten geven.

Alpha-diversiteit

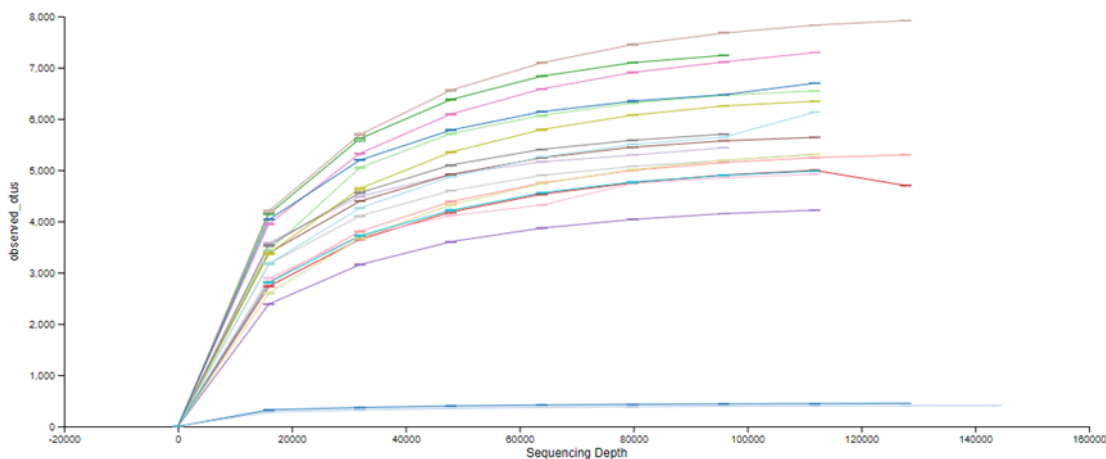
De alpha-diversiteit is de diversiteit binnen een monster. Voor het berekenen van de alpha-diversiteit kan gekozen worden uit een scala aan methoden (indices), bijv. de Shannon index of Faith's phylogenetic diversity index. Over de verschillende indices is meer terug te lezen op de Qiime2 website: <https://forum.qiime2.org/t/alpha-and-beta-diversity-explanations-and-commands/2282>

De kracht van deze indices is dat één getal (de index) de diversiteit binnen het gehele monster uitdrukt. Dit maakt het gemakkelijk om deze indices tussen monsters te vergelijken en statistische toetsen uit te voeren om te achterhalen of er significante verschillen in diversiteit zijn tussen monsters. De statistische toetsen die meestal toegepast kunnen worden zijn non-parametrische testen zoals de Kruskal Wallis tests en Wilcoxon rank test omdat NGS data in het algemeen niet normaal verdeeld is. Het is van belang dat de gebruiker goed begrijpt wat de index uitdrukt, zodat de interpretatie van deze index en vergelijking tussen monsters goed gedaan wordt. Voor verschillende onderzoeksvragen kunnen namelijk verschillende indices toepasselijk zijn. Voor diversiteitsindexen wordt gebruik gemaakt van species 'richness' (het aantal soorten) en de 'evenness' (distributie van abundantie van de verschillende soorten) en mogelijk de taxonomische/fylogenetische diversiteit (de genetische relatie tussen verschillende soorten).

Bij KWR wordt nog niet specifiek gebruik gemaakt van één diversiteitsindex, maar worden vaak meerdere van de veelgebruikte indices berekend en wordt per project bekeken welke index het beste gebruikt kan worden die past bij de onderzoeksvraag. Een voorbeeld (de faith PD index) wordt gegeven in 6.2 en 7.1.1. In de praktijkvoorbeelden in 7.2 wordt ook gebruik gemaakt van de faith PD index, zodat vergelijking van data mogelijk is binnen dit rapport.

Naast rarefaction voor normalisatie, kunnen ook Alpha- rarefactionplots gemaakt worden. Deze plots helpen om te beoordelen of monsters diep genoeg gesequencet zijn om de gehele microbiële diversiteit van een monster te bepalen. Als een monster bijvoorbeeld maar heel weinig sequentiereads heeft, dan kunnen alleen de meest dominante soorten aangetroffen worden in een monster. Wanneer dieper wordt gesequencet (meer

sequentiereads per monster worden gegenereerd), worden ook de minder dominante soorten gedetecteerd. Voor drinkwater zonder een desinfectieresidu is dit extra van belang, omdat in die monsters de diversiteit erg hoog is: er zijn veel verschillende soorten aanwezig en de dominante soorten komen ook in relatief lage aantallen voor. Dit betekent dat veel dieper gesequencet moet worden met drinkwatermonsters dan met andere monsters waar de diversiteit lager is. In Figuur 13 is een voorbeeld gegeven van een Alpha-rarefactionplot. Op de x-as wordt de sequencingdiepte gegeven (in aantal sequentiereads) en op de y-as het aantal geobserveerde ASV's. Elke lijn is een apart monster en elk punt is een moment dat de rarefaction uitgevoerd is, dus een willekeurig submonster uit de sequentiereads is genomen. Bij elk submonster komen er nieuwe sequentiereads bij en dat resulteert ook in het aantreffen van extra ASV's, maar als het goed is wordt per extra submonster het aantal nieuwe ASV's steeds minder. Wanneer de lijn een asymptoot bereikt, is het monster voldoende diep gesequencet om alle ASV's in het monster te detecteren. Aan de hand van een aantal analyses op verschillende drinkwatermonsters, hebben we geobserveerd dat drinkwatermonsters minimaal ca. 100.000 (maar bij minder diversiteit soms ook minder) reads moeten bevatten om deze asymptoot te bereiken. Dit heeft gevolgen voor het aantal monsters dat per run geanalyseerd kan worden. Ervaring bij KWR heeft geleerd dat er ca. 50-60 monsters per run meegenomen kunnen worden wanneer men drinkwatermonsters analyseert en geïnteresseerd is in de gehele microbiële populatie.



Figuur 13: Een voorbeeld van een Alpha-rarefactionplot waar op de x-as de sequencingdiepte weergegeven is (in aantal sequentiereads) en op de y-as het aantal geobserveerde ASV's. Elke lijn stelt één onderzocht monster voor.

Beta-diversiteit

De alpha-diversiteit is de diversiteit tussen verschillende monsters. De Beta-diversiteit kan ook met verschillende indices geanalyseerd worden en ook hier geldt dat elke index andere resultaten geeft. Het is daarom van belang dat de juiste keuze in indices gemaakt wordt voor een betrouwbare vergelijking tussen monsters. Voor Beta-diversiteitanalyse wordt meestal gebruik gemaakt van de mate van overeenkomst of verschil in ASV's tussen monsters. Deze overeenkomsten/verschillen worden weergegeven in een afstandswaarde tussen dezelfde ASV's in twee monsters. Hieronder een aantal manieren waarop deze afstandswaarde berekend kunnen worden:

- Jaccard distance: een **kwalitatieve** maat voor verschillen in ASV-populatie tussen monsters. Deze maat berekent hoeveel overeenkomstige ASV's er tussen twee monsters zijn gedeeld door het totaal aantal ASV's in beide monsters:

$$Jaccard = \frac{a_{i,j}}{a_{i,j} + b_i + c_j}$$

waarbij i het ene monster en j het andere monster is; $a_{i,j}$, het aantal overeenkomstige ASV's is tussen monster i en monster j; b_i , het aantal ASV's die in monster i aanwezig zijn, maar niet in monster j en c_j , het

aantal ASV's die in monster j aanwezig zijn maar niet in monster i. Het getal dat hier uitkomt (tussen 0 en 1) wordt ook wel de afstand (distance) tussen twee monsters genoemd.

- Bray-Curtis dissimilarity: een **kwantitatieve** maat voor verschillen in de ASV-populatie tussen monsters. Deze maat vergelijkt hoeveel verschillende ASV's er zijn tussen monsters, maar ook hoeveel van elke ASV voorkomt in ieder monster en hoe de aantallen ASV's verschillen tussen monsters.

$$\text{Bray-Curtis} = \frac{2A_{ij}}{S_i + S_j}$$

waarbij i het ene monster en j het andere monster is; A_{ij} , het laagste aantal sequenties van iedere ASV die in allebei de monsters aanwezig zijn, S_i , het totaal aantal sequenties in monster i en S_j , het totaal aantal sequenties in monster j. Het getal dat hier uitkomt (tussen 0 en 1) wordt ook wel de afstand (distance) tussen twee monsters genoemd (rekenvoorbeeld is opgenomen in Bijlage I).

- Unweighted UniFrac distance: een **kwalitatieve** maat voor verschillen in ASV-populatie tussen monsters waarbij de fylogenetische relaties meegenomen worden. De index vergelijkt de fylogenetische afstand tussen ASV's in de verschillende monsters. Het getal dat hier uitkomt (tussen 0 en 1) wordt ook wel de afstand (distance) tussen twee monsters genoemd.
- Weighted UniFrac distance: een **kwantitatieve** maat voor verschillen in ASV-populatie tussen monsters gebaseerd op de fylogenetische relaties zoals in de Unweighted Unifrac distance, maar waarbij ook de hoeveelheid van elke ASV's meegewogen wordt.

Tabel 3: Een samenvatting van de verschillen tussen de verschillende Beta-diversiteitsindices

	ASVs	Phylogenetische boom
Unweighted	Jaccard index	Unweighted Unifrac
Weighted	Bray- Curtis	Weighted Unifrac

Het verschil van de beta-diversiteit met de alpha-diversiteit is dat de beta-diversiteit verschillen *tussen* monsters bekijkt, terwijl de alpha-diversiteit verschillen *binnen* een monster bekijkt. De beta-diversiteit wordt daarom niet door één getal weergegeven, maar er wordt een kruistabel gegeneerd tussen de monsters. In deze tabel wordt voor ieder monsterpaar een afstandsindexwaarde gegeven. Doordat met meerdere monsters het heel lastig is om de onderlinge afstand van ieder monster ten opzicht van alle andere monsters te doorzien met een kruistabel, worden zogenaamde ordinaatiemethoden toegepast op de matrixtabel zijn. Andere termen voor ordinatie zijn (multivariate) gradiëntanalyse of multidimensionale scaling. Ordinatie is gedefinieerd als het rangschikken van objecten langs gradiënten op grond van waarnemingen aan meer afhankelijke variabelen. In andere woorden, de kruistabel wordt teruggebracht naar een aantal ongedefinieerde variabelen, waarbij elke variabele een deel van de variantie van de kruistabel beschrijven. De twee of drie variabelen met de hoogste verklarende variantie worden vervolgens op een x en y of een x, y en z as geplote (de ordinaatplot). Ordinatie kan ook met zeer veel verschillende methoden worden toegepast en ook hier geldt dat elke methode de data anders weergeeft en dus de interpretatie sterk beïnvloed. Het is daarom zaak om de verschillende ordinaatiemethoden enigszins te begrijpen. De meest gebruikte vormen van ordinatie voor NGS data zijn: principal component analysis (PCA), principal coordinates analyses (PCoA), Nonmetric multidimensional scaling (NMDS), Redundantieanalyse (RDA), en Canonische correspondentieanalyse (CCA), maar er zijn er velen meer. Uitleg over deze verschillende ordinaatietechnieken kan

online teruggevonden worden op de qiime2 website (<https://forum.qiime2.org/t/alpha-and-beta-diversity-explanations-and-commands/2282>), aangezien het teveel is om dit onder te brengen in deze rapportage.

Ook kan op de beta-diversiteitsindices statistiek uitgevoerd worden om te bepalen of verschillen significant zijn tussen monsters. Deze methoden zijn meestal gebaseerd op gepaarde of ongepaarde varianten van de non-parametrische multivariate analysis of variance (PERMANOVA) testen.

Bij KWR wordt meestal gebruik gemaakt van Bray-Curtis dissimilarity, dat in PCA (met Hellinger transformatie), of PCoA wordt weergegeven. Er wordt gekozen voor Bray-Curtis dissimilarity omdat vaak een kwantitatieve maat nuttig is om niet alleen verschillen te zien in welke soorten voorkomen, maar ook in welke mate. Dit is vooral van belang wanneer bijvoorbeeld een zuivering wordt doorgemeten; daar verandert de microbiële populatie, maar omdat het een doorlopend systeem is met hetzelfde water dat gezuiverd wordt, zijn veranderingen in kwantiteit van de verschillende micro-organismen ook interessant. De reden waarom PCA en PCoA vaak gekozen wordt is omdat deze ordinatiemethode heel handig is om een eerste indruk te krijgen van wat de grote verschillen zijn tussen monsters, zonder bepaalde aannames van de data te maken. Andere methoden zoals RDA en CCA berusten meer op het effect van bepaalde parameters op microbiële populaties. Deze methoden zijn ook heel handig voor bepaalde onderzoeksvragen, maar worden niet altijd toegepast. In dit rapport zijn verschillende indices gebruikt als voorbeeld en toegelicht in 6.3 en 7.1.2. In de praktijkvoorbeelden in 7.2 wordt voor alleen gebruik gemaakt van Bray-Curtis dissimilarity in PCoA plots om analyses te reduceren en praktijkvoorbeelden vergelijkbaar te houden.

4.4.8 Differentiële abundantie

Differentiële abundantieanalyses worden gedaan om te achterhalen of de ASV-populatiesamenstellingen binnen een bepaalde experimentele conditie verschillen van de ASV-populatiesamenstellingen waargenomen binnen een andere experimentele conditie (bijvoorbeeld een controle). Er heerst sterke discussie tussen bioinformaticaxperts welke methoden het beste zijn om deze verschillen aan te tonen. Anderen beweren dat er überhaupt geen geschikte methode is om dergelijke verschillen voor NGS data te berekenen. Dit berust in het probleem dat data genormaliseerd moet worden voor deze analyses. Zoals eerder aangegeven bevat elk monster door technische aspecten een ander aantal sequentiereads en dat kan de oorzaak zijn dat bepaalde ASVs vaker voorkomen in het ene dan in het andere monster. Door de data te normaliseren kan hiervoor gecorrigeerd worden. Voor het testen van differentiële abundantie is normaliseren naar relatieve abundantie een optie. Het nadeel van relatieve abundantie is dat als een bepaalde ASV is toegenomen onder invloed van een experimentele conditie die aanwezig is in het ene monster maar niet in het andere monster, alle andere ASV's automatisch afnemen, en vice versa. Het gevolg is dat de positieve en negatieve correlaties die hierdoor worden aangetroffen tussen ASV's en experimentele condities daarom vaak vals positief zijn (Weiss, Xu et al. 2017). Enkele differentiële abundantiemethoden proberen daarom op een andere manier te normaliseren, zoals bijvoorbeeld Analysis of Composition of Microbiomes (ANCOM) (Mandal, Van Treuren et al. 2015), DESeq2 (Love, Huber et al. 2014) en Gneiss. Uitleg over deze methoden is terug te vinden in de referenties en op de Qiime2 website: <https://docs.qiime2.org/2020.8/tutorials/gneiss/> en <https://docs.qiime2.org/2017.8/tutorials/moving-pictures/#differential-abundance-testing-with-ancom>.

Omdat ANCOM en Gneiss zijn opgenomen in de Qiime2 workflow als plugins, worden deze bij KWR gebruikt wanneer differentiële abundantie tests benodigd zijn. Deze methoden zijn ook gebruikt in dit rapport als voorbeeld, en worden verder toegelicht in 7.1.3.

4.4.9 Staafdiagrammen/heatmaps en meer statistiek

Vaak is het gewenst om data op veel verschillende manieren te visualiseren, bijvoorbeeld met heatmaps en staafdiagrammen. Ook maken we gebruik van aanvullende statistiek wanneer dit nodig is in een project. In Qiime2 kan een groot deel van deze analyses gedaan worden, maar soms is er de voorkeur voor andere tools of kan een bepaalde analyse alleen door een andere tool worden uitgevoerd. Bij KWR wordt daarom ook gebruik gemaakt van het statistische softwarepakket 'R', waarvoor pakketten gemaakt zijn speciaal voor microbial profiling analyse, zoals

ampvis2 (<https://madsalbertsen.github.io/ampvis2/articles/ampvis2.html>) en phyloseq (<https://joey711.github.io/phyloseq/>). Uitleg over het gebruik van deze packages is te vinden op de websites.

4.4.10 Qiime viewer

In de rest van dit rapport gaan we vooral in op Qiime2 analyses en laten we daarom alleen visualisaties zien die direct uit Qiime2 komen. Het doel daarbij is dat de lezer van dit rapport in een later stadium de Qiime2 resultaten ook kan interpreteren. Een groot voordeel van Qiime2 is namelijk dat resultaten gemakkelijk met anderen gedeeld kunnen worden in een interactieve vorm met zogenaamde Qiime2 viewer bestanden (.qzv files). Deze bestanden kunnen met 'drag and drop' in de Qiime viewer geladen worden: <https://view.qiime2.org/>.

5 Kwaliteitscontroles

De kwaliteit van de microbial profiling analyses kan worden vastgesteld en geborgd met een geoptimaliseerde methode die, voor zover mogelijk, gestandaardiseerd is en is voorzien van ingebouwde controles. Deze optimalisatie en standaardisatie heeft niet alleen betrekking op het deel van de analyse dat in het laboratorium plaatsvindt. Ook het onderdeel dataverwerking is daarbij van groot belang, waarbij de keuze voor geschikte software, databases en visualisatiemogelijkheden een goede en efficiënte interpretatie van data mogelijk moet gaan maken.

5.1 Blanco monsters als negatieve controle

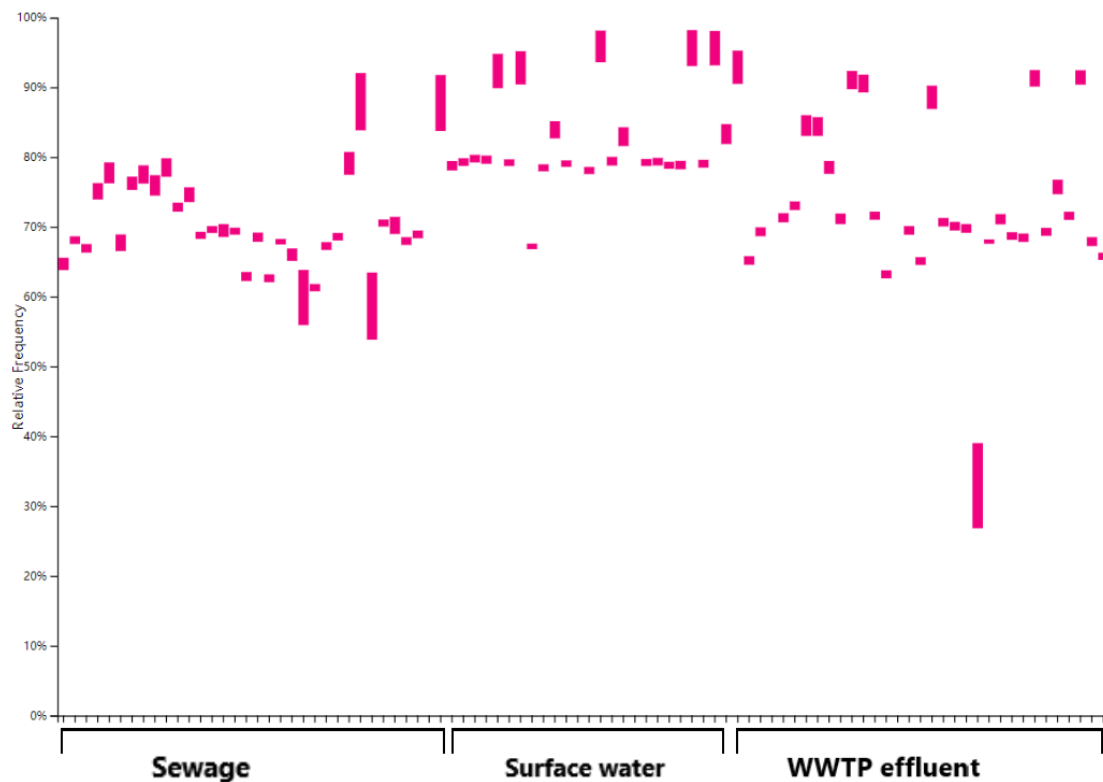
Blanco monsters zijn negatieve controles die in elke sequencing run meegenomen dienen te worden. Het is aan te raden om negatieve controles voor DNA isolatie en voor de PCR stap mee te nemen. De negatieve controle voor DNA isolatie uit water en biofilm is DNA-vrij water, dat ook wordt gefiltreerd over een filter zoals het normale watermonster of biofilmsuspensiemonster ook wordt gefiltreerd. De negatieve controle voor de PCR uitvoering is ook DNA-vrij water, waar dezelfde PCR op wordt uitgevoerd als op het reguliere monster. Wanneer deze negatieve controles sequencing resultaten geven, dan komt dit door besmetting tijdens de DNA isolatie of tijdens de PCR stap. Deze besmettingen kunnen het gevolg zijn van het optreden van kruiscontaminaties tussen monsters maar ook door de aanwezigheid van contaminerend microbiële DNA in de gebruikte extractie- of PCR-reagentia (Stinson, Keelan et al. 2019). Kruiscontaminaties tussen monsters lijken daarbij de grootste bijdrage te hebben (Minich, Sanders et al. 2019). Door het uitvoeren van analyses op blanco monsters is het mogelijk om hiermee rekening te houden en voor te corrigeren tijdens interpretatie van de resultaten. Wanneer monsters te veel lijken op blanco's zullen deze mogelijk verwijderd moeten worden uit verdere analyse.

Om te onderzoeken welke sequenties aangetoond worden in blanco monsters zijn de hierboven genoemde negatieve controles tijdens elke MiSeq run bij KWR meegenomen. In dit onderzoek zijn op die manier 63 blanco monsters geanalyseerd die tussen 2016 en 2019 gesequencet zijn, om te bepalen wat er precies in de blanco monsters aanwezig is. Hiervoor hebben we onderzocht welke 'core features' aanwezig zijn in blanco monsters. Core features zijn de kern (=core) ASV's (ASV's worden door Qiime2 features genoemd) die voorkomen in een bepaald type monster; dus de micro-organismen die vrijwel altijd in dat type monster voorkomen. Wanneer de micro-organismen die in de negatieve controles aanwezig zijn, teruggevonden worden in drinkwatermonsters, is het aannemelijk dat deze aanwezig zijn door (kruis)contaminatie van het monster, en dus niet in het drinkwatermonster zelf aanwezig waren. De analyse van de negatieve controles liet zien dat de blanco monsters onderling erg veel verschillen in microbiële profiel, wat verwacht is wanneer er kruiscontaminatie optreedt van verschillende monsters. Er werden wel enkele taxa gevonden die veel voorkomen in de meeste blanco monsters:

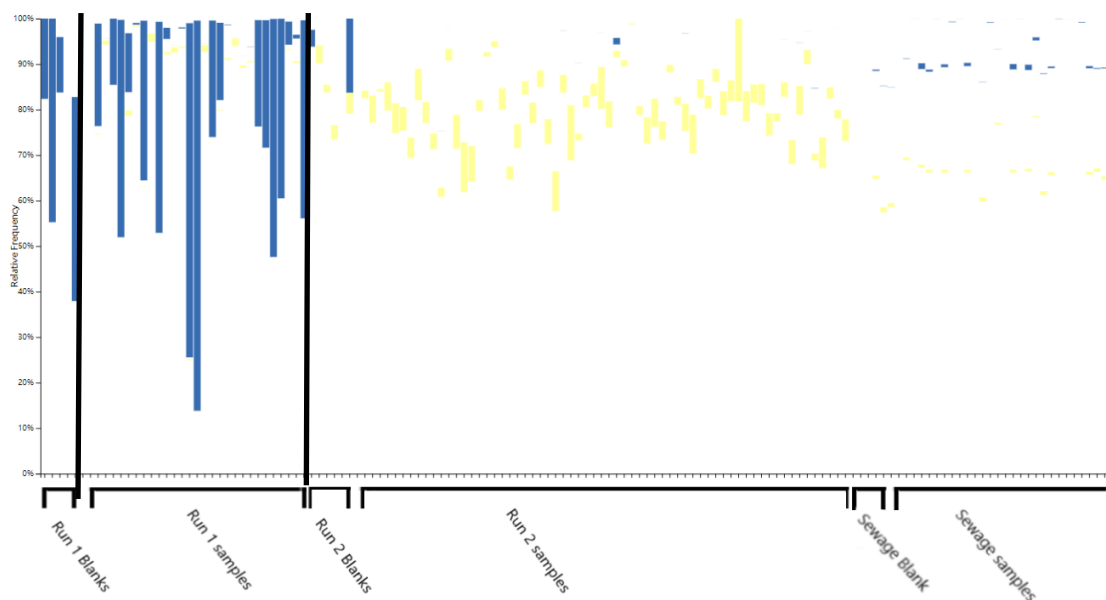
1. Een onbekende soort van het genus *Delftia* (uit de familie *Comamonadaceae*) werd gevonden in 73% van alle 63 blanco monsters.
2. Het genus *Propionibacterium* (uit de *Propionibacteriaceae* familie) kwam voor in 37% van de blanco monsters.
3. Het genus *Acinetobacter* (uit de *Moraxellaceae* familie) kwam voor in 28% van de blanco monsters.
4. Het genus *Staphylococcus* (uit de *Staphylococcaceae* familie) kwam ook voor in 28% van de blanco monsters.

Een verklaring voor de aanwezigheid van deze taxa kan komen door contaminatie van DNA extractiekits en het ultrapure water (voor de negatieve controles) dat gebruikt wordt. Het is daarom van belang om na elke run de sequentiereads van het blanco monsters te vergelijken met de sequentiereads van de daadwerkelijke monsters en de positieve controles (zie paragraaf 4.5.2). Wanneer bepaalde taxa voorkomen in de blanco monsters die niet dominant zijn in de andere monsters, is er eerder sprake van contaminatie door DNA extractiekits en het ultrapure water. In dat geval is het belangrijk om die contaminaties te elimineren uit de monsters. Dat kan handmatig, maar er zijn ook tools voor ontwikkeld die de mate van contaminatie identificeren, zoals Source Tracker (Knights, Kuczynski et al. 2011). Daarnaast is het mogelijk om in Qiime2 bepaalde sequenties te verwijderen uit de dataset (exclude-seqs). In onze ervaring is bij analyse van de 63 blanco monsters in alle runs alleen (nog) niet waargenomen dat er significante contaminatie van kits optreedt.

Een andere verklaring dat de blanco monsters sequenties kunnen bevatten naast contaminatie is door verkeerde demultiplexing, wat leidt tot bovengenoemde kruiscontaminatie. Tijdens demultiplexing kunnen fouten optreden, waardoor de verkeerde sequencing reads aan de verkeerde monsters toegewezen worden. Dit laatste is zeer waarschijnlijk het geval, sinds *Comamonadaceae* in bijna alle watermonsters gevonden is (Figuur 14) en wanneer demultiplexing verkeerd gaat, zou een *Comamonadaceae* sequentie uit een watermonster abusievelijk kunnen worden toegekend aan een blanco monster. Dit impliceert dus dat sequentiereads aan verkeerde monsters toegekend worden. Blanco monsters bevatten weinig DNA en dus weinig reads, waardoor de foute toewijzing van reads van andere monsters tijdens demultiplexing grote gevolgen heeft op het totale aantal reads. Dit effect is te zien in Figuur 15, waar de aanwezigheid van 2 taxa weergegeven is in 3 verschillende runs. In run 1 komt *Sulfuricurvum* (paars) veel voor, waardoor deze waarschijnlijk door verkeerde demultiplexing ook in de blanco monsters van die run voorkomt. In run 2 komt *Pirellulaceae* familie (geel) veel voor, waardoor deze waarschijnlijk door verkeerde demultiplexing ook in de blanco monsters van die run voorkomt. In de 3^e run (sewage) komen ze beide voor, en dat is ook terug te zien in de 'blanco monsters van die run. Deze kruiscontaminatie die door demultiplexing wordt veroorzaakt wordt ook wel 'index-hopping' genoemd. Het demultiplexen van monsters wordt namelijk gedaan aan de hand van index-sequenties die bij bepaalde monsters horen. Het is om deze reden van belang dat er niet teveel verschillende typen monsters aan één run toegevoegd worden, zoals bijvoorbeeld rioolwater- en drinkwatermonsters. Een deel van de dominante soorten uit het rioolwater kunnen door dit technische effect de drinkwatermonsters contamineren, waardoor de verkeerde conclusies worden getrokken over de bacteriële populatiesamenstelling van het drinkwater.



Figuur 14: Taxa-barplot van alle runs van de meng-experimenten waarbij rioolwater (sewage), oppervlaktewater (surface) en gezuiverd afvalwater (WWTP effluent) gemengd was met drinkwater. Alleen de relatieve aanwezigheid van de *Comamonadaceae* familie is weergegeven.



Figuur 15: Taxa-barplot van 3 runs (run 1, run 2 en sewage) waarbij alleen de relatieve aanwezigheid van *Sulfuricurvum kujiense* (blauw) en *Pirellulaceae* family (geel) gegeven is voor de monsters van de run (samples) en de blanco's van de runs (blanks).

5.2 Mock communities als positieve controle

Mock communities zijn gedefinieerde mengsels van microbiële cellen, virussen en/of DNA. Deze worden in vitro gemaakt om de microbiële compositie (of het DNA ervan) van de microbiota na te bootsen. De term "mock"

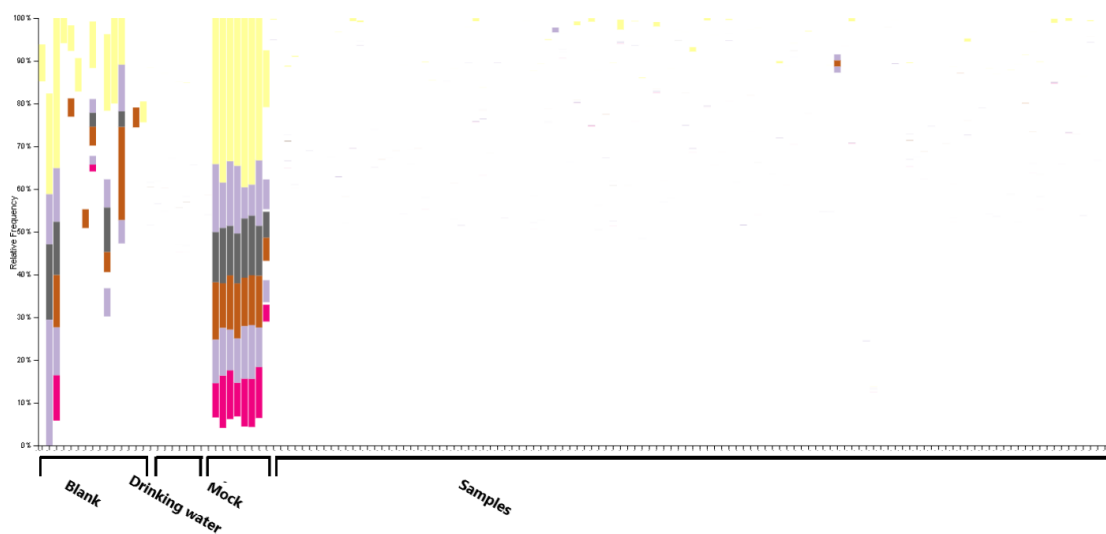
betekent dan ook imitatie. Deze mock communities worden geanalyseerd als positieve controles in sequencing runs. Doordat de precieze samenstelling van de mock community (welke micro-organismen en in welke verhouding) bekend is, kan men kwaliteitscontrole op de sequencingmethode doen door de sequencingresultaten te vergelijken met de daadwerkelijke compositie van de mock community. Ter controle van het gehele proces (lab-analyse en bioinformatica) is het van groot belang om bij elk experiment ook de analyse uit te voeren op een mock community (Yeh, Needham et al. 2018). In eerste instantie is bij KWR gestart met het ontwikkelen van een eigen mock community met hierin mengsels van bekende bacteriesoorten (Tabel 4). Bij de analyse van deze mock community bleek het onduidelijk of de toegevoegde *Klebsiella*- en *Rahnella*-stammen ook daadwerkelijk tot deze bacteriegenera behoorden en was aanvullend onderzoek nodig om dit grondig uit te zoeken. Echter, op dat moment kwam de firma Zymo met een commercieel verkrijgbare en eenvoudig te gebruiken mock community (<https://www.zymoresearch.com/collections/zymbiomics-microbial-community-standards/products/zymbiomics-microbial-community-standard>) en is besloten om in eerste instantie deze mock community in gebruik te nemen (samenstelling: *Listeria monocytogenes* - 12%, *Pseudomonas aeruginosa* - 12%, *Bacillus subtilis* - 12%, *Escherichia coli* - 12%, *Salmonella enterica* - 12%, *Lactobacillus fermentum* - 12%, *Enterococcus faecalis* - 12%, *Staphylococcus aureus* - 12%, *Saccharomyces cerevisiae* - 2%, and *Cryptococcus neoformans* - 2%).

Tabel 4. Samengestelde mock community met KWR laboratoriumstammen

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

In Qiime2 zijn meerdere opties aanwezig om een kwaliteitscontrole uit te voeren op basis van de mock community en om daaraan gekoppeld het foutenpercentage te bepalen. Dit is terug te lezen op deze pagina: <https://docs.qiime2.org/2017.12/tutorials/quality-control/>. In het kort, in Qiime2 kan de verwachte compositie van de mock community vergeleken worden met de geobserveerde mock community na de sequencerun. Het resultaat zijn waargenomen/verwachte ratio's, valspositieven en valsnegatieven. Daarnaast kan de kwaliteit van de sequencingrun geëvalueerd worden door de sequenties van bepaalde soorten in de mock communities te vergelijken met de sequenties van die soorten in de database om zo de frequentie van mismatches te identificeren.

Het technische effect van het onjuist toewijzen van indexsequenties (index-hopping) tijdens demultiplexing (zie paragraaf 4.5.1) is ook te zien in Figuur 16 waar alle taxa van de mock community in alle monsters van één run weergegeven zijn. Het is duidelijk te zien dat de blanco monsters vooral de taxa van de mock communities bevatten (omdat deze zelf weinig sequentiereads genereren en dus het aandeel verkeerde toekenning groter wordt). De daadwerkelijke monsters bevatten ook, maar voor een veel kleiner deel, deze taxa. Het is van belang om de juiste samenstelling van de mock communities te kiezen gebaseerd op het soort monsters dat geanalyseerd wordt. In het geval van drinkwatermonsters is het van belang om mock communities samen te stellen van soorten die normaliter in drinkwater voorkomen. Wanneer een mock community gekozen wordt zoals de Zymo mock community (die voornamelijk darmbacteriën bevat), dan komen deze soms in de monsters terecht (Figuur 16) en wekt dit de verkeerde indruk dat darmbacteriën in het drinkwater aanwezig is.



Figuur 16. Taxa-barplot van verschillende runs, waarbij alleen de relatieve aanwezigheid van de mock communities weergegeven is in de blanco monsters (blanks), drinkwater monsters (drinking water) en monsters van alle 5 runs (samples).

6 Bepalen van minimale DNA-concentratie voor microbial profiling

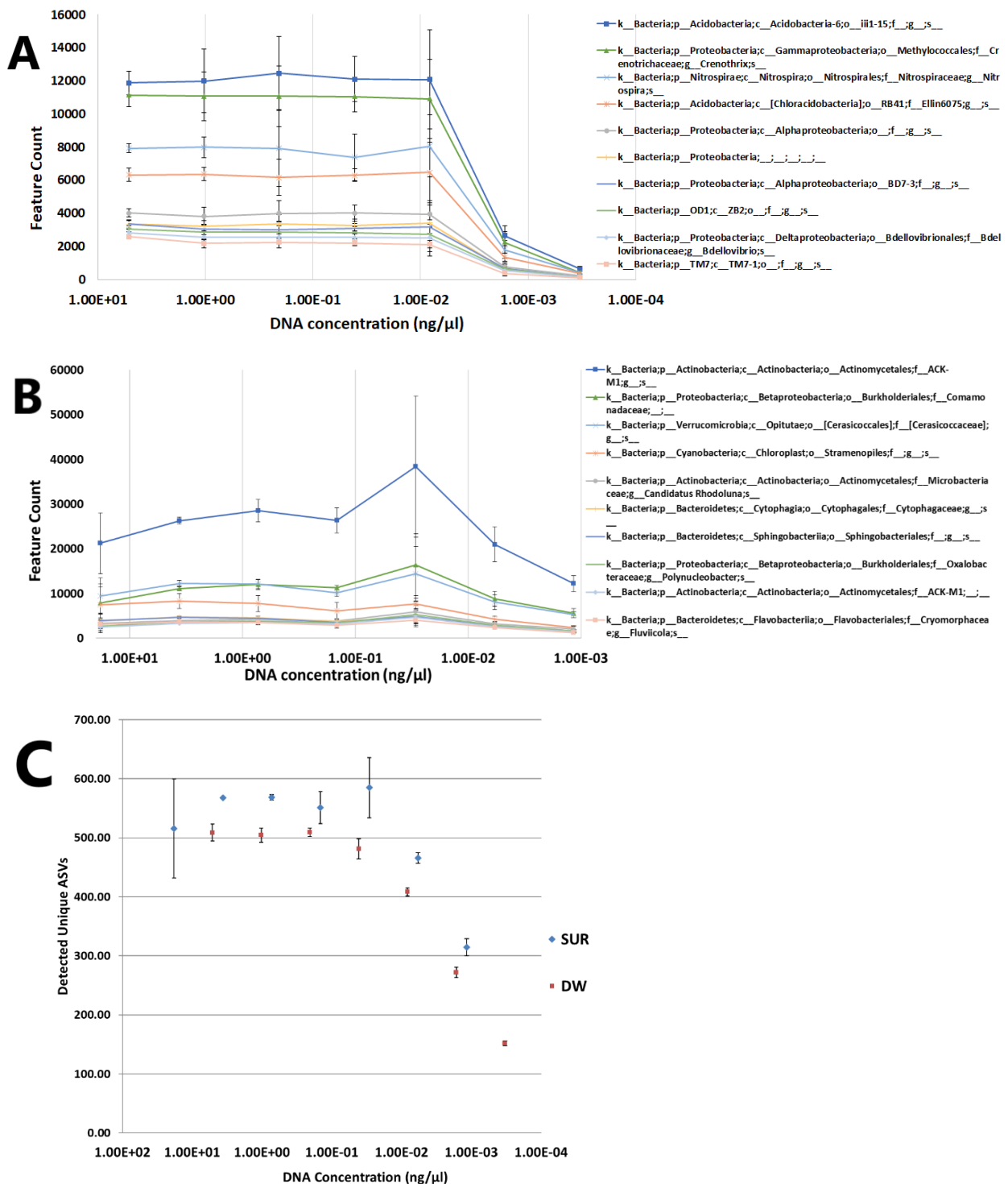
6.1 Celaantallen, DNA-concentratie en sequencereads

Omdat bij KWR vooral monsters gesequencet worden voor de drinkwaterbedrijven, is het van belang om de NGS analyses vooral op deze matrix te optimaliseren. Monsters die vanuit de drinkwaterbedrijven komen bevatten relatief weinig micro-organismen, waardoor weinig DNA te isoleren is uit deze monsters. Vooral grondwatermonsters bevatten zeer lage DNA-concentraties, waardoor sequenceanalyses problematisch kunnen zijn. Het is daarom van belang om te weten bij welke concentratie DNA een NGS analyse nog representatieve resultaten geeft. Om deze minimale DNA-concentratie te achterhalen, hebben we daarom verdunningsreeksen van DNA van drinkwatermonsters en oppervlaktewatermonsters gesequencet. Hiervoor is eerst het DNA geïsoleerd uit drinkwater (3L drinkwater van het KWR-lab in Nieuwegein) en oppervlaktewater (0,4L uit het Lekkanaal ter hoogte van de inlaat van de WRK) en vervolgens zijn analyses uitgevoerd (in drievoud) op verdunningsreeksen met zes verdunningsstappen van dit DNA waarbij bij elke verdunningsstap het DNA een factor 5 is verdund (onverdund tot 15.625 keer verdund). In de gebruikte (onverdunde) monsters is de concentratie ATP, het aantal dode en levende cellen (met flowcytometrie) en de concentratie geïsoleerd DNA bepaald. De DNA-concentratie is met een fluorometrische meting op de QuBit fluorometer bepaald. De resultaten van deze metingen zijn weergegeven in Tabel 5.

Tabel 5. Concentraties ATP, aantallen dode en levende cellen (flowcytometrie) en de concentratie geïsoleerd DNA in de gebruikte watermonsters.

Monster	Filtratie Volume (ml)	Elutie Volume (µl)	ATP (pg/ml)		FCM (cells/ml)			Concentratie DNA (ng/µl)						
			ATP	stdv	levende cellen	dode cellen	totaal cellen	Onverdund	5X	25X	125X	625X	3125X	15625X
Drinkwater	3000	100						5,44E+00	1,09E+00	2,18E-01	4,35E-02	8,70E-03	1,74E-03	3,48E-04
Drinkwater	3000	100	2,8	0,3	5,40E+04	2,90E+03	5,70E+04	4,88E+00	9,76E-01	1,95E-01	3,90E-02	7,81E-03	1,56E-03	3,12E-04
Drinkwater	3000	100						5,12E+00	1,02E+00	2,05E-01	4,10E-02	8,19E-03	1,64E-03	3,28E-04
Drinkwater							Gem:	5,15E+00	1,03E+00	2,06E-01	4,12E-02	8,23E-03	1,65E-03	3,29E-04
Oppervlaktewater	400	100						1,77E+01	3,54E+00	7,08E-01	1,42E-01	2,83E-02	5,66E-03	1,13E-03
Oppervlaktewater	400	100	250	55	2,70E+06	<1,0E5	2,70E+06	1,80E+01	3,60E+00	7,20E-01	1,44E-01	2,88E-02	5,76E-03	1,15E-03
Oppervlaktewater	400	100						1,89E+01	3,78E+00	7,56E-01	1,51E-01	3,02E-02	6,05E-03	1,21E-03
Oppervlaktewater							Gem:	1,82E+01	3,64E+00	7,28E-01	1,46E-01	2,91E-02	5,82E-03	1,16E-03

Om te weten wanneer in een monster nog voldoende DNA aanwezig is voor het genereren van een microbiel profiel die representatief is voor het totale drinkwater- of oppervlaktewatermicrobiota, is het belangrijk om te weten bij welke DNA-concentratie de microbiële gemeenschap gaat afwijken. Ten eerste zijn de absolute feature counts (oftewel hoe vaak unieke ASVs voorkomen) van belang, dus hoe diep gesequencet kan worden bij een bepaalde DNA concentratie. In Figuur 17 zijn deze absolute feature counts te zien voor de top 10 meest abundante ASVs in drinkwater (A) en oppervlaktewater (B). Drinkwater heeft in het algemeen minder feature counts dan oppervlaktewater. Te zien is verder dat de absolute feature counts drastisch afneemt bij een DNA concentratie rond de 10^{-2} ng/µl. De unieke ASVs verdwijnen echter al bij een DNA concentratie rond de 10^{-1} ng/ul voor drinkwater en rond de 10^{-2} ng/ul voor oppervlaktewater (Figuur 17c). Omdat de absolute feature counts van iedere ASV evenveel lager wordt bij afnemende DNA-concentratie, blijft de relatieve abundantie van de top 10 ASVs wel gelijk gedurende verdunning.



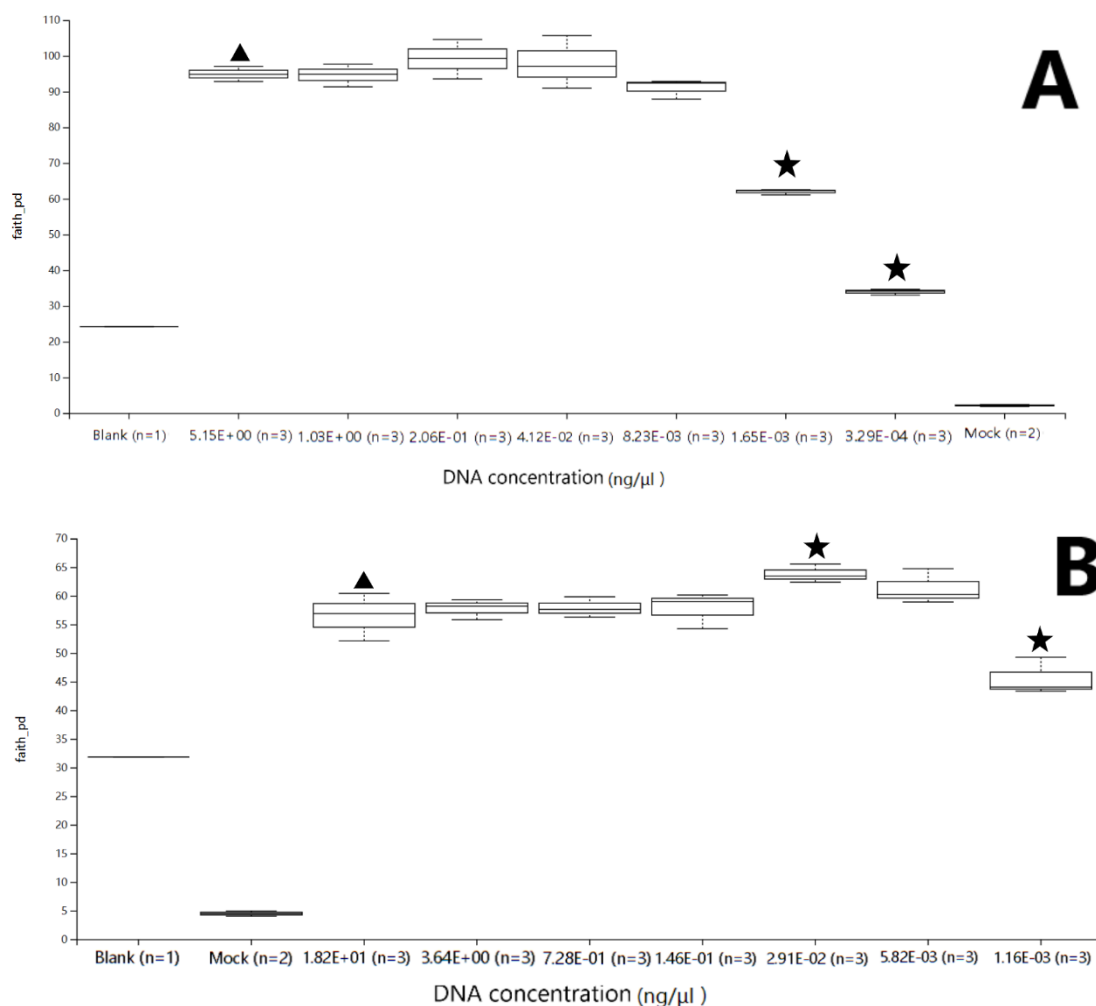
Figuur 17: Het totaal aantal features van de 10 meest abundante taxa in drinkwater (A), oppervlaktewater (B) en het totaal aantal unieke ASVs (C) van de DNA verdunningsreeks van drinkwater (DW) en oppervlaktewater (SUR). Standaarddeviaties tonen variaties tussen biologisch triplo metingen.

6.2 Alpha-diversiteit

De vraag is wanneer het verliezen van feature counts van de meest abundante ASVs en het verlies van unieke ASVs een effect heeft op de diversiteit en de microbiële compositie. Dit aspect kan worden bepaald met Faith's phylogenetic diversity (faith's PD) index, wat een diversiteitsindex is die de phylogenetische soortenrijkdom van de ASVs in acht neemt. Dat wil zeggen, dat een lagere diversiteit uitgedrukt in de faith's PD index betekent dat de

fylogenetische rijkdom lager is. De faith's PD index kan daarom 'tipping points' aangeven waar de soortenrijkdom drastisch verlaagd wordt door het verdwijnen van belangrijke fylogenetische soorten. Hierdoor is de faith's PD index zeer geschikt om te bepalen wanneer de DNA concentratie een effect heeft op de diversiteit (Faith 1992).

De diversiteit van de DNA-verdunningsreeks uit drinkwater, uitgedrukt in faith's PD index (Figuur 18A), is significant lager vergeleken met het onverdunde monster vanaf een verdunning van 56 keer (Kruskal-Wallis H test, $p < 0,05$), wat overeenkomt met een theoretische DNA concentratie van $1,65 \times 10^{-3}$ ng/ul. De diversiteit van de DNA-verdunningsreeks uit oppervlaktewater (Figuur 18B) is verrassend genoeg significant hoger vergeleken met het onverdunde monster bij een verdunning van 56 keer (Kruskal-Wallis H test, $p < 0,05$), overeenkomstig een theoretische DNA concentratie van $2,91 \times 10^{-2}$ ng/ul. Een significant lagere diversiteit werd geobserveerd bij een DNA concentratie van $1,16 \times 10^{-3}$ ng/ul (15.625 keer verdunning) (Kruskal-Wallis H test, $p < 0,05$).

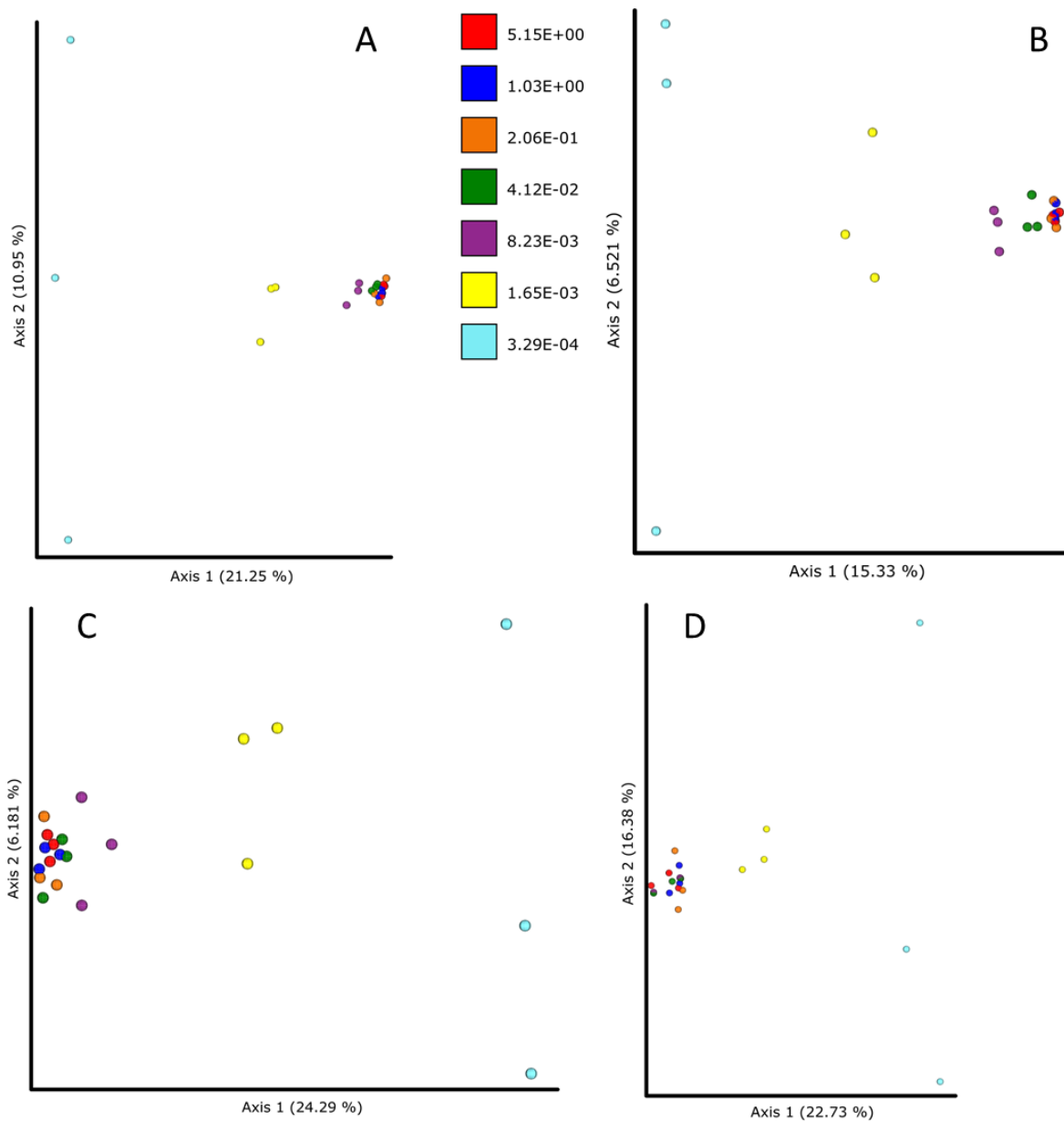


Figuur 18: De faith-pd index (y-axis) uitgezet tegen de DNA verdunningsreeksen voor drinkwater (A) en oppervlaktewater (B). Box plots geven de spreading voor de triplo ($n=3$) monsters aan. De boxplots met ster zijn significant verschillend van de onverdunde monsters met driehoek (Kruskal-Wallis H test, $p < 0.05$).

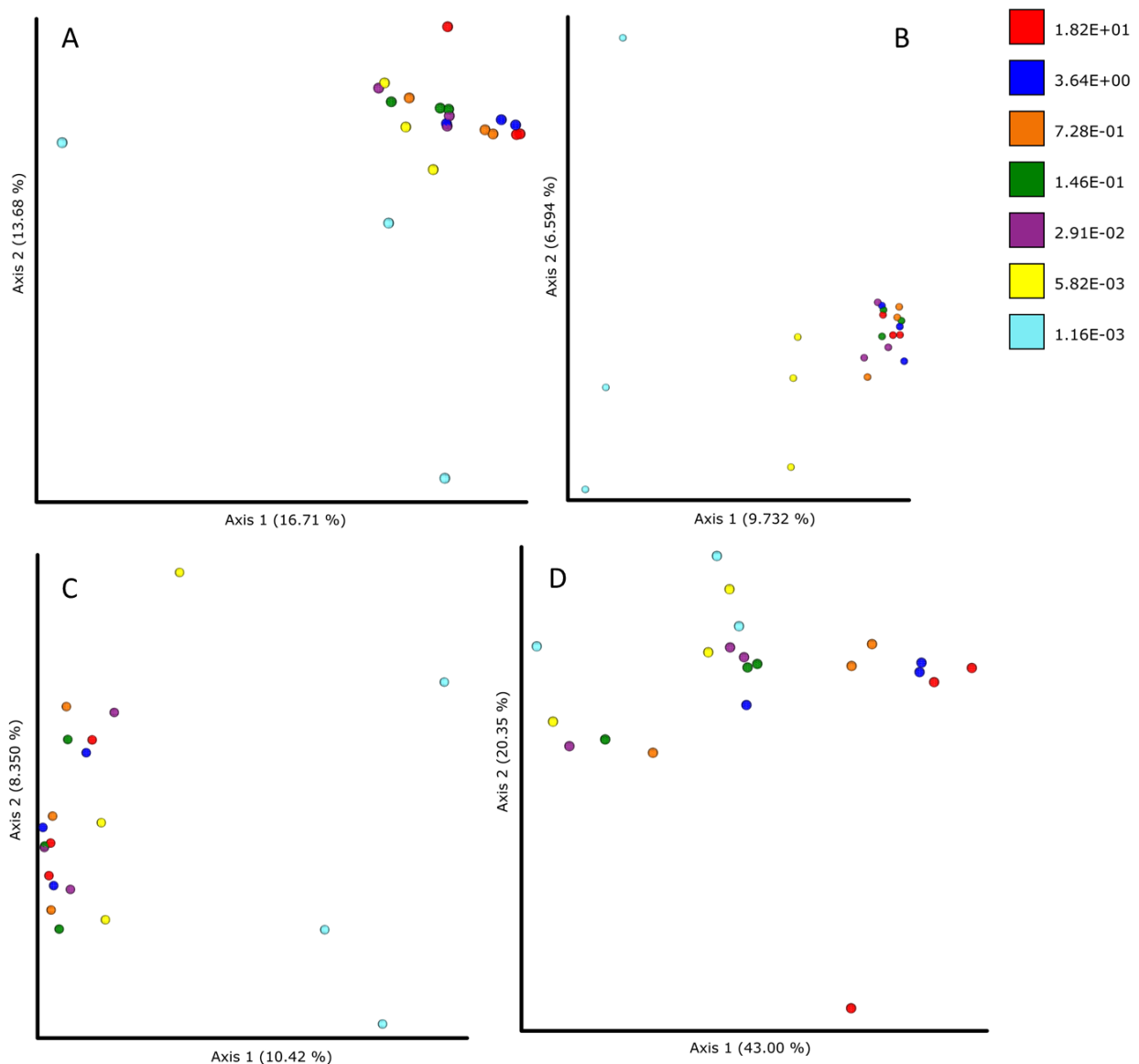
Terwijl het aantal unieke ASVs al afnam bij 10^{-1} ng/ul voor drinkwater en 10^{-2} ng/ul voor oppervlaktewater, heeft dit dus pas een effect op de absolute feature counts van dominante ASVs bij 10^{-2} ng/ul en op de totale diversiteit bij 10^{-3} ng/ul.

6.3 Beta-diversiteit

Nu heerst nog de vraag bij welke DNA concentratie de microbiële compositie verschillend is *tussen* de monsters (beta-diversiteit). Hiervoor zijn vier verschillende beta-diversiteitsindices berekend voor de DNA verdunningsreeksen van beide watertypen. Dit is ook gedaan ter illustratie voor de verschillende resultaten die deze indices opleveren.



Figuur 19: De beta-diversiteit van de drinkwater DNA verdunningsreeks met Bray Curtis (A), Jaccard distance (B), Un-weighted unifrac (C) en Weighted unifrac (D) index, weergegeven in een principal coordinates analyses (PCoA). Weergegeven zijn de monsters (circels), gekleurd aan de hand van de DNA concentratie (zie legenda), blanco monsters (ring) en Mock communities (ster). Er is geen significant verschil gevonden in compositie tussen de verschillende monsters (Permanova significance test, $p < 0.05$).



Figuur 20: De beta-diversiteit van de oppervlaktewater DNA verdunningsreeks met Bray Curtis (A), Jaccard distance (B), Un-weighted unifrac (C) en Weighted unifrac (D) index, weergegeven in een principal component analyses (PCA) met Hellinger transformatie. Weergegeven zijn de monsters (cirkels), gekleurd aan de hand van de DNA concentratie (zie legenda), blanco monsters (rode ring) en Mock communities (rode ster). Er is geen significant verschil gevonden in compositie tussen de verschillende monsters (Permanova significance test, $p < 0.05$).

In Figuur 19 en Figuur 20 zijn PCoA plots weergegeven van de DNA verdunningsreeksen van drinkwater en oppervlaktewater. Het is te zien dat verdunningen van het DNA van drinkwater met de vier hoogste concentraties (5,15; 1,03; 0,206 en 0,0412 ng/ μ l) in de PCoA plot dicht bij elkaar clusteren en de samenstelling van deze populaties verschillen dus weinig van elkaar. Bij de lagere concentraties (0,00823; 0,00165 en 0,000329 ng/ μ l) worden de verschillen in samenstelling van de populaties duidelijk zichtbaar in de plot waarbij elke volgende verdunningsstap resulteert in een duidelijke verschuiving in de plot. Dit verschil is met de vier beta-diversiteitsindices te zien.

Voor het oppervlaktewater is dit anders: de verdunningsreeksen van oppervlaktewater scheiden zich niet met alle beta-diversiteitsindexen van de onverdunde monsters. Bij de kwantitatieve methoden (Bray Curtis en Weighted Unifrac) is er minder verschil tussen de verdunningen terwijl dit meer het geval is voor de kwalitatieve methoden

(Jaccard en Unweighted Unifrac). Dit komt omdat de kwantitatieve methoden het aantal *verschillende* ASVs en de *hoeveelheid* van alle afzonderlijke ASVs meenemen, terwijl de kwalitatieve methoden alleen het aantal *verschillende* ASVs in de analyse meenemen. Hierdoor heeft een afname in ASVs die niet dominant voorkomen in de populatie een zelfde effect als een afname van ASVs die wel dominant voorkomen op kwalitatieve methoden. Wanneer er afname is van ASVs die zeldzaam zijn in een populatie, heeft dit dus weinig effect op de kwantitatieve methoden, en een groot effect op kwalitatieve methoden. Hierdoor is de scheiding in de kwantitatieve methoden dus minder goed. Bij de kwalitatieve methoden vindt er scheiding plaats van de monsters met de laagste twee concentraties ($5,82 \times 10^{-3}$ en $1,16 \times 10^{-3}$ ng/ μ l) met de andere monsters. Het aantal unieke ASVs en absolute feature counts voor oppervlaktewater nam af bij 10^{-2} ng/ μ l en deze DNA concentratie scheidt zich dan ook van andere monsters bij de kwalitatieve methoden. Dit geeft dus aan dat deze afname vooral plaatsvindt bij de zeldzame ASVs.

De drinkwatermonsters vertonen bij zowel kwalitatieve (Jaccard en unweighted unifrac) en kwantitatieve (Bray-Curtis en weighted unifrac) beta-diversiteitsindices goede scheiding. In tegenstelling tot oppervlaktewater lijken zowel de kwalitatieve als kwantitatieve indexen een duidelijke scheiding te geven. Dat ook de kwantitatieve indexen een duidelijke scheiding geeft bij drinkwater komt waarschijnlijk doordat er een sterke afname was van zeldzame en unieke ASVs. Het aantal unieke ASVs nam voor drinkwater namelijk al af bij 10^{-1} ng/ μ l en voor oppervlaktewater pas bij 10^{-2} ng/ μ l DNA verdunning. Daarnaast vertoont verdunning van drinkwater een eerder effect op de diversiteit dan bij het oppervlaktewater. Drinkwater heeft een hogere diversiteit dan oppervlaktewater, wat betekent dat het meer verschillende ASVs bevat. Dit betekent dus dat verdunning van DNA van drinkwater een groter effect heeft op deze zeldzame ASVs en de totale soortenrijkdom dan verdunning van oppervlaktewater.

De reden dat er geen significante verschillen gevonden zijn tussen de monsters (PERMANOVA, $p < 0,05$), kan komen door de grote variatie in microbiële compositie binnen de groepen die vergeleken zijn. Hierdoor is het niet mogelijk voor de statistische toets om grotere verschillen tussen t.o.v. binnen de groepen te zien.

6.4 Conclusie

Op basis van bovenstaande resultaten kan geconcludeerd worden dat de DNA concentratie van monsters hoger moet zijn dan 10^{-3} ng/ μ l, wat correspondeert met een geschatte totaal aantal cellen van 10-20 cellen/ml. Lagere DNA concentraties zorgen voor een verlies van de diversiteit (alpha-diversiteit) en een verschil in microbiële compositie (beta-diversiteit). Een DNA-concentratie van 10^{-3} ng/ μ l is goed haalbaar met filtratie van 500 tot 1000 ml watermonsters en eventueel de toepassing van hemoflow. Zelfs uit grondwater wordt met die filtratie- of concentreermethode DNA concentraties gekregen van $>0,1$ ng/ μ l.

7 Praktijktoeepassingen microbial profiling

In dit hoofdstuk zijn de praktische toepassingen van microbial profiling voor het drinkwateronderzoek in meer detail uitgezocht. Er is experimenteel onderzocht in hoeverre minimale verontreinigingen in drinkwater detecteerbaar zijn middels microbial profiling. Daarnaast zijn casestudies met drinkwatermonsters uit de praktijk uitgevoerd om een indruk te krijgen van de samenstelling van de bacteriepopulaties na de zuiveringsstappen van verschillende drinkwaterzuiveringen, waarmee de mogelijkheden van microbial profiling voor de drinkwatersector verder worden verkend.

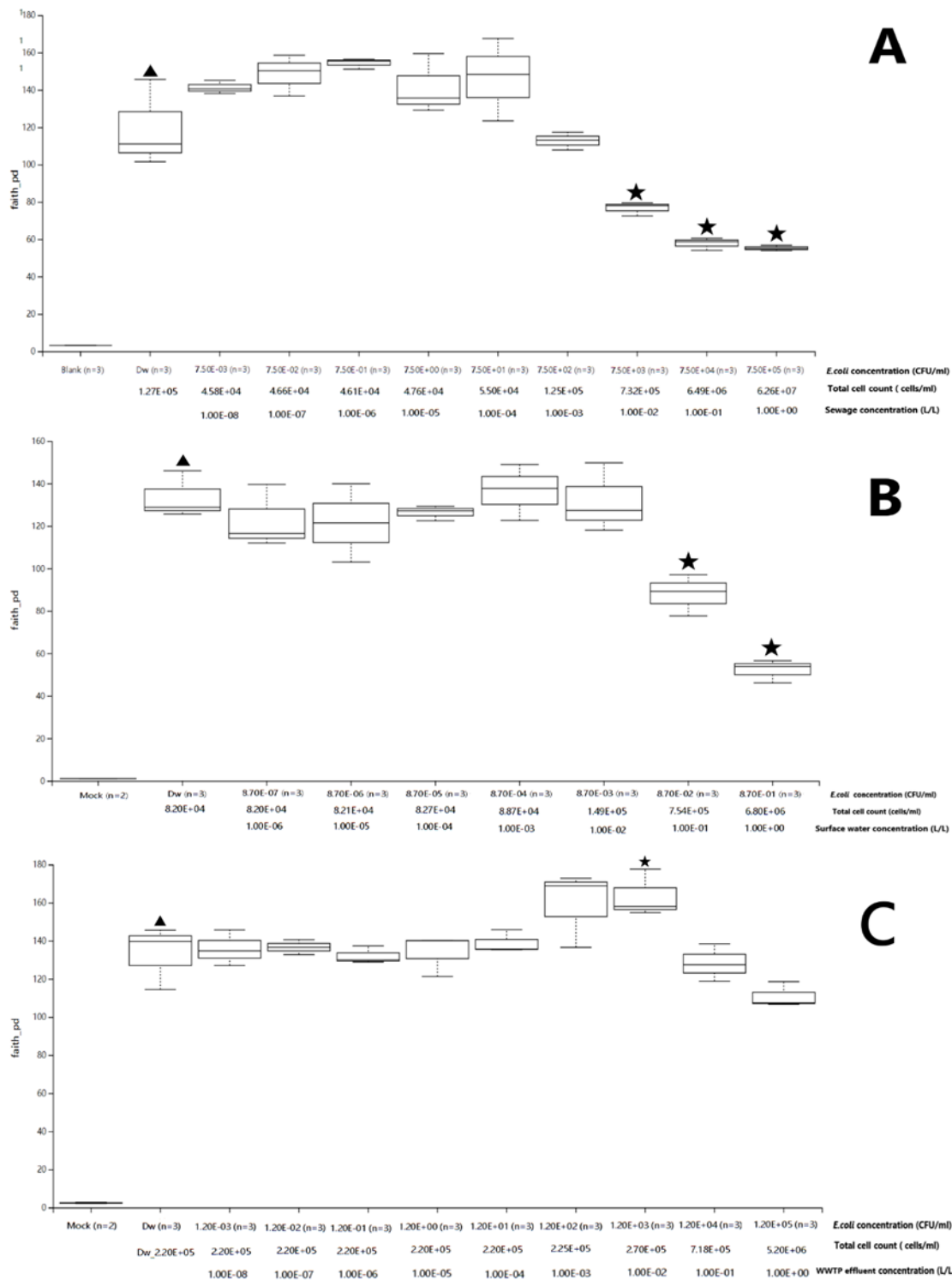
7.1 Microbial profiling voor detectie van verontreinigingen in drinkwater

Om te achterhalen of NGS kan dienen als gevoelige methode voor detectie van verontreinigingen in drinkwater en om een algemene indruk te krijgen van de mogelijkheden van NGS analyses voor het aantonen van zeldzame bacteriesoorten in de populaties (door detectie van hun 16S rRNA gensequenties), zijn er analyses uitgevoerd op drinkwatermonsters waaraan kunstmatig verontreinigingen zijn toegevoegd. Dit is gedaan door drinkwater te verontreinigen met verdunningsreeksen van verschillende potentiële verontreinigingsbronnen: rioolwater, oppervlaktewater en RWZI effluent. Voor het detecteren van verontreinigingen in drinkwater met behulp van microbial profiling, hebben we de verschillende methoden gebruikt die geïntroduceerd zijn in hoofdstuk 4, zoals alpha- en beta-diversiteit en differentiële abundantie. Daarnaast is ook een source tracking tool gebruikt die in dit hoofdstuk wordt toegelicht.

7.1.1 Alpha-diversiteit

In Figuur 21 zijn de resultaten van de Faith PD index (als maat voor alpha-diversiteit) te zien wanneer in drinkwater een toenemende mate van contaminatie plaatsvindt met rioolwater (A), oppervlaktewater (B) en RWZI effluent (C). De verontreiniging aangegeven met 10^0 bestaat volledig uit het verontreinigingsmonster terwijl de verontreinigingsverdunning weergegeven met 10^{-1} , voor 10% bestaat uit verontreiniging en 90% uit drinkwater; elke volgende verdunningsstap bevat vervolgens een factor 10 minder verontreinigingswater. De alpha-diversiteit fluctueert t.o.v. drinkwater bij toenemende mate van contaminatie. Door deze fluctuaties is er pas een significant verschil ($p < 0,05$) in diversiteit te zien bij een verdunning van drinkwater tot 10^{-2} met rioolwater (99% drinkwater en 1% rioolwater) en tot 10^{-1} met oppervlaktewater (90% drinkwater en 10% oppervlaktewater), terwijl voor het RWZI effluent geen significant verschil tijdens contaminatie in verschillende verdunningen wordt gezien.

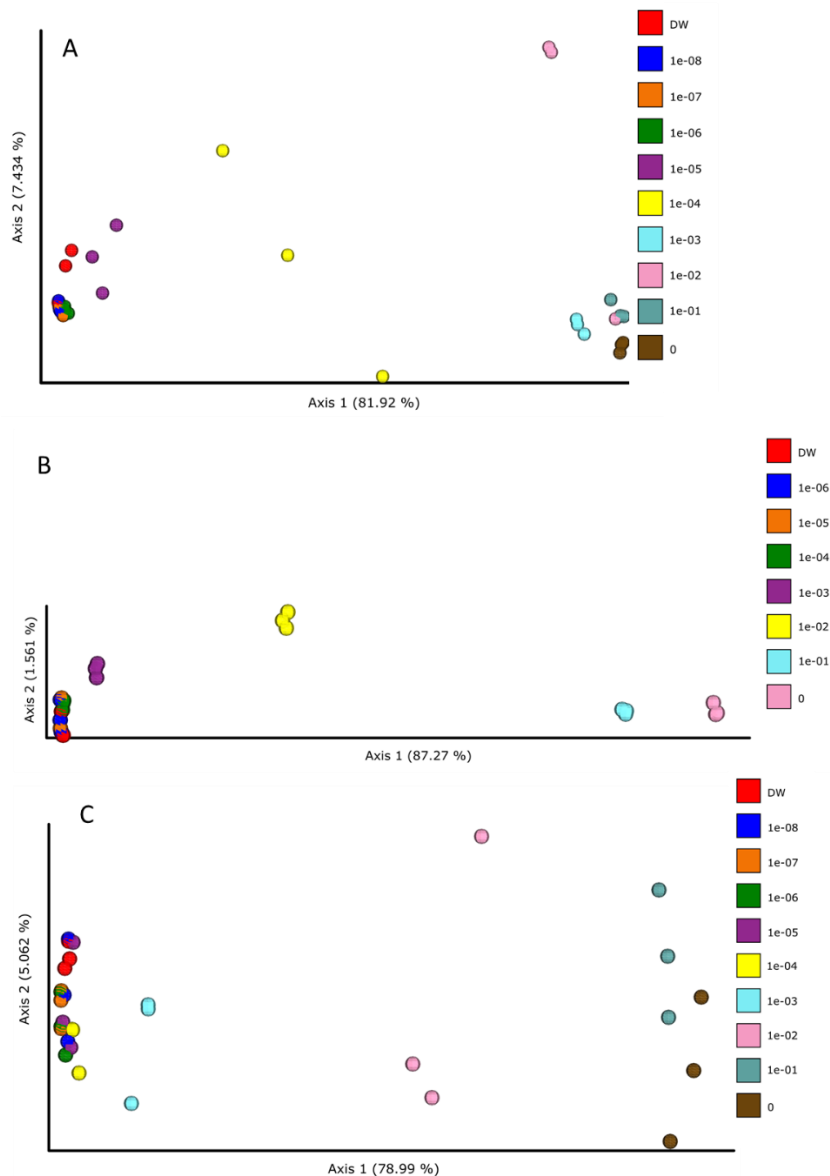
Dit verschil in detectie van drinkwater met verontreiniging van verschillende contaminatiebronnen is te verklaren door de hoogte van diversiteit van de verschillende contaminatiebronnen. De diversiteit van rioolwater en oppervlaktewater is significant lager dan die van drinkwater, waardoor bij een bepaalde lage verdunning de afname van diversiteit significant wordt. De diversiteit van het RWZI-effluent is echter niet significant verschillend van drinkwater, waardoor het mengen van rioleffluent met drinkwater in verschillende verhoudingen nooit zal leiden tot een significante afname of toename van de diversiteit. Het waarnemen van dezelfde mate van diversiteit tussen monsters (alpha-diversiteit) zegt echter niets over eventuele verschillen in de microbiële compositie (beta-diversiteit); het RWZI-effluent en drinkwater kunnen dus een geheel andere compositie hebben, maar wel een soortgelijke diversiteit. Dit toont aan dat de alpha-diversiteit geen goede maat is voor het detecteren van contaminatie van drinkwater door verschillende bronnen.



Figuur 21: Alpha-diversiteit uitgedrukt in Faith's PD index van riolwater (A), oppervlaktewater (B) en RWZI effluent (C) gemengd met drinkwater (DW), met sampling diepte van 66988, 535342 en 37746 reads, respectievelijk. De corresponderende *E. coli* koloniegallen (CFU/ml/ml) en totale celtellingen (cellen/ml) zijn naast de contaminatie concentratie (L/L, 1.00E+00 = 100% contaminatie) op de x-as weergegeven. Monsters die significant verschiden van het drinkwater (met driehoek erboven) zijn weergegeven met een ster (Kruskall Wallis, $p < 0,05$).

7.1.2 Beta-diversiteit

Een beta-diversiteitsanalyse is uitgevoerd met dezelfde vier beta-diversiteitsindices (Jaccard, Bray-Curtis, unweighted unifrac en weighted unifrac) als voor het DNA concentratie-experiment beschreven in hoofdstuk 6. Alle vier de verschillende indices vertoonden soortgelijke resultaten, en om die reden worden alleen de resultaten van één van de vier indices (de Bray Curtis index) weergegeven in Figuur 22.



Figuur 22. PCoA plots gebaseerd op Bray Curtis index van drinkwater verdund met rioolwater (A), oppervlaktewater (B) en RWZI effluent (C), met monsterdiepte van 66.988, 535.342 en 37.746 sequencereads, respectievelijk. Monsters (cirkels) zijn gekleurd aan de hand van de concentratie verdunding (L/L, zie legenda), waarbij 100% contaminatie de 0 L/L verdunding is en 0% contaminatie alleen drinkwater (DW) is. Er is geen significant verschil aangetroffen in microbiële compositie tussen de verschillende monstergroepen (PERMANOVA, $p < 0,05$).

Voor alle contaminatiebronnen is duidelijk een scheiding te zien in de verdunningen van de drinkwater gecontamineerde monsters. De monsters die het minst gecontamineerd zijn clusteren samen met het drinkwater en naarmate het percentage contaminatie van rioolwater, oppervlaktewater of RWZI effluent hoger wordt, wordt

de scheiding met het drinkwatermonster groter. De meest gecontamineerde monsters (10^{-1} , 90% drinkwater, 10% contaminatiebron) clusteren dichtbij de contaminatiebronnen (100% contaminatie).

De mate waarin de monsters bij elkaar clusteren aan de hand van de contaminatie verschilt per contaminatiebron. Dit verschil komt waarschijnlijk omdat in dit contaminatie-experiment naast verlies van ASVs die in drinkwater voorkomen (uitgaande van steeds meer verontreiniging met de contaminatiebron) een toename ontstaat van ASVs die in de contaminatiebron aanwezig zijn.

De monsters in dit contaminatie-experiment vertonen grote variatie binnen de monsters (grote spreiding tussen de triplo monsters). Dit kan komen doordat het mengen van de contaminaties met drinkwater in het laboratorium leidt tot het verliezen van verschillende zeldzame ASVs in ieder van de triplo monsters. Hierdoor zullen de triplos andere, unieke ASVs bevatten met grote spreiding. Deze spreiding *binnen* de clusters is waarschijnlijk de reden waarom er geen significante verschillen gevonden worden *tussen* de clusters (PERMANOVA, $p < 0,05$). Aangezien de beta-diversiteit door deze complicerende factoren geen geschikte maat lijkt voor het beoordelen van de mate van contaminatie, hebben we verdere statistische analyses achterwege gelaten. De verschillen tussen de microbiële profielen van de verschillende verdunningen kan ook met andere methoden worden bepaald, deze worden in de volgende paragrafen beschreven.

7.1.3 Differentiële abundantie en indicatorsoorten

In bovenstaande paragrafen is aangetoond dat indices voor de alpha- en beta-diversiteit van de gehele microbiële populatie niet volstaan om contaminatie van verschillende bronnen aan te tonen. De microbiële verschillen tussen de drinkwatermonsters, verdunningen en contaminatiebronnen zou mogelijk ook kunnen worden bepaald door specifiek naar bepaalde indicatorsoorten te kijken die wel voorkomen in de contaminatiebron, maar niet in het drinkwater. Er zijn verschillende mogelijkheden om uit een gehele microbiële populatie specifieke soorten te detecteren die alleen of meer aanwezig zijn in de contaminatiebron en het gecontamineerd water dan in het niet-gecontamineerde drinkwater. Veel van deze mogelijkheden berusten op het vinden van verschillen in ASVs tussen twee of meerdere type monsters en worden differentiële abundantiemethoden genoemd. Er zijn in dit project verschillende van dergelijke methoden getest op bruikbaarheid om contaminatiespecifieke differentiële abundantie tussen monsters te bepalen, waaronder de methoden genaamd “Gneiss” of “ANCOM”. Daarnaast hebben we met een andere methode de kerntaxa (core features) van de contaminatiebronnen bepaald en onderzocht of deze kerntaxa ook in de verdunningen worden waargenomen.

Kerntaxa

Kerntaxa zijn ASVs die voorkomen in verschillende monsters van hetzelfde ecosysteemtype (bijvoorbeeld drinkwatermonsters van verschillende productielocaties of influent/effluentmonsters van verschillende RWZIs) en daarmee de kern voor de microbiologie van dat ecosysteemtype vormen. Het bepalen van deze kerntaxa, kan helpen om aan te wijzen welke bacteriesoorten, -genera en/of -familie altijd voorkomen in bepaalde contaminatiebronnen. Dit kan helpen met het vaststellen van specifieke indicatorsoorten.

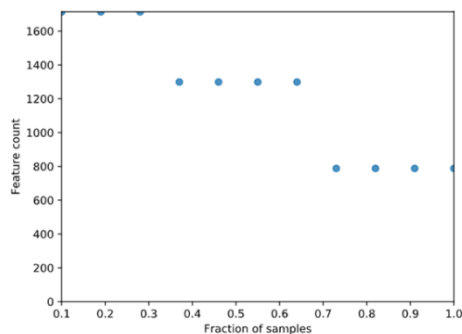
In eerste instantie hebben we de kerntaxa bepaald van de 100% contaminatiemonsters, dus van onverdund rioolwater, RWZI effluent en oppervlaktewater. De resultaten toonden dat in alle triplo monsters van rioolwater, RWZI effluent en oppervlaktewater respectievelijk 1496, 1939, en 788 unieke ASVs aanwezig zijn die in alle triplo's voorkomen.

De data wordt weergegeven zoals in Figuur 23 te zien is voor oppervlaktewater. Er wordt per fractie van een vooraf geselecteerd aantal monsters weergegeven hoeveel kern-ASVs deze bevatten en wat het deel van deze kern-ASVs

is ten opzichte van alle ASVs die in de monsters zijn aangetroffen. Omdat wij van rioolwater, oppervlaktewater en RWZI-effluent triplo monsters geanalyseerd hebben en per type matrix deze drie triplo monsters hebben geselecteerd voor de kerntaxa-analyse, kijken we alleen naar fractie 1 oftewel ASVs die in 100% van de triplo monsters voorkomen (de zogenaamde kern-ASVs). Vervolgens kunnen we in het bestand dat met deze analyse in Qiime2 wordt gegenereerd de abundantie van iedere kern-ASV achterhalen, zodat de dominante kern-ASVs kunnen worden geselecteerd. Abundantie is namelijk een belangrijk selectiecriteria om te bepalen of bepaalde kern-ASVs als indicator voor contaminatie kunnen dienen. Wanneer ASVs namelijk een zeer lage abundantie hebben in de contaminatiebron, is de kans groter dat deze ASVs niet gedetecteerd worden wanneer de contaminatie zeer gering aanwezig is in drinkwater en zijn deze ASVs dus geen goede indicatorsoorten ondanks dat ze wel specifiek zijn voor de contaminatiebron. Daarnaast bestaat het risico dat ASVs met lage abundantie afkomstig kunnen zijn uit andere monsters in dezelfde sequentierun door het onjuist toekennen van deze ASVs tijdens demultiplexing, of door contaminatie tijdens labwerkzaamheden (zie hoofdstuk 4).

Feature table summary

Sample count: 3
Feature count: 1714



Terminology key

Fraction of samples: The fraction of the total number of samples that a feature must be observed in for that feature to be considered "core".

Feature count: The number of "core" features.

Fraction of features: The fraction of the total number of features that are "core" features.

Feature list: The list of "core" feature IDs and their total percentile frequencies.

Fraction of samples	Feature count	Fraction of features	Feature list
0.10	1714	1.000000	TSV
0.19	1714	1.000000	TSV
0.28	1714	1.000000	TSV
0.37	1299	0.757876	TSV
0.46	1299	0.757876	TSV
0.55	1299	0.757876	TSV
0.64	1299	0.757876	TSV
0.73	788	0.459743	TSV
0.82	788	0.459743	TSV
0.91	788	0.459743	TSV
1.00	788	0.459743	TSV

Figuur 23: De output van qiime core features voor de triplo monsters van oppervlaktewater. Gegeven is het aantal kern-ASVs (feature count = 788) dat is gevonden in 100% van de monsters (fraction of samples = 1) en de fractie van de kern-ASVs ten opzichte van alle ASVs in de triplo monsters (fraction of features = 0,46, oftewel 46% van de totale ASVs zijn de 788 kern-ASVs).

De vijf kern-ASVs die het meest voorkomen in alle triplo monsters per verschillende watermatrix behoren tot:

Voor rioolwater:

- De genera *Arcobacter* en *Acinetobacter*, leden van de familie *Aeromonadaceae* en de orde *Fusobacteriales*.

Voor RWZI effluent:

- De genera *Collinsella*, *Thauera*, en "*Candidatus Accumulibacter*", leden van de familie *Conexibacteraceae*, en de klasse TM7-3 die behoren tot de orde xxx.

Voor oppervlaktewater:

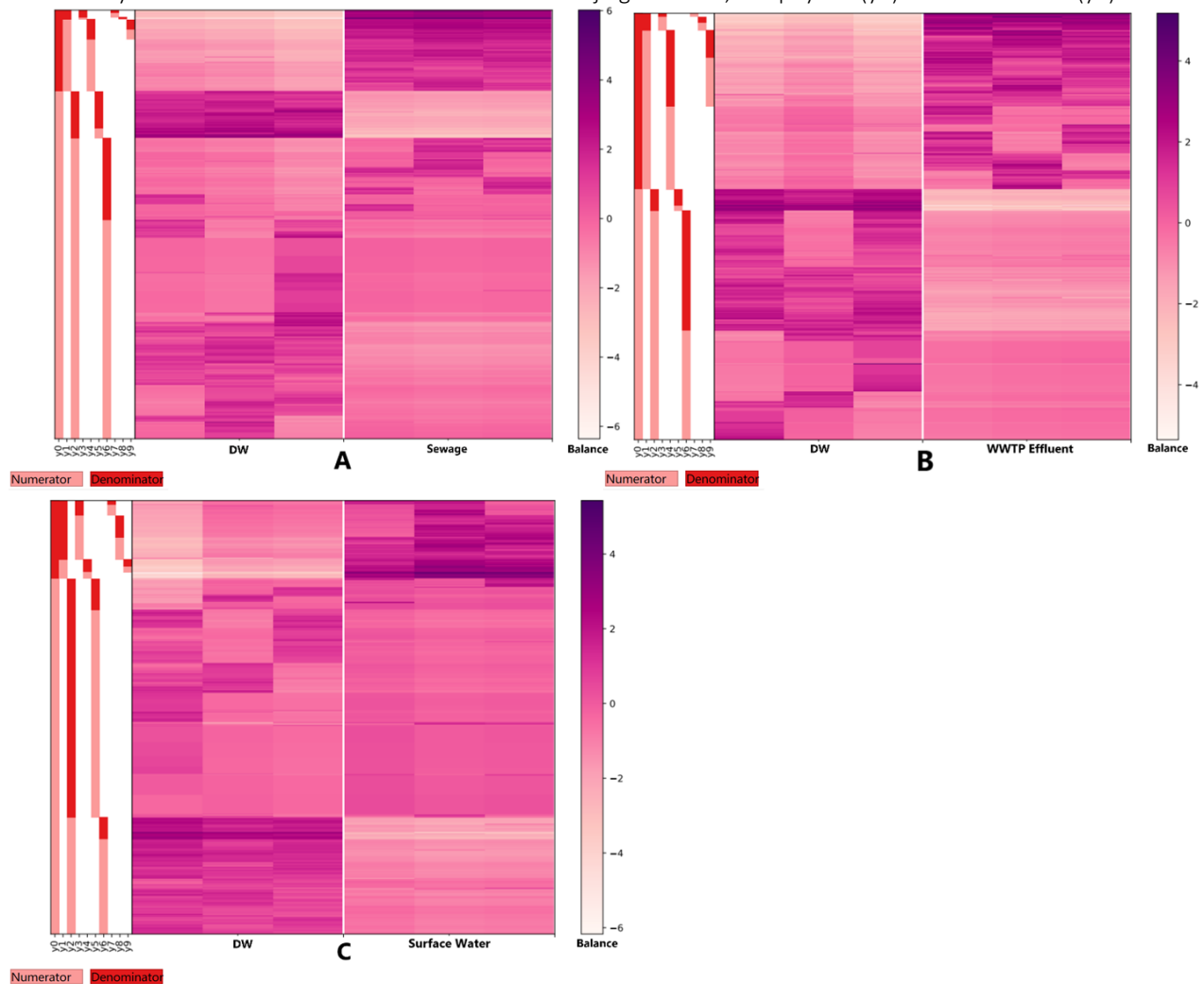
- Genera behorende bij de ordes *Actinomycetales* en *Methylophilales*.

De kern-ASVs geven dus op een goede manier weer welke ASVs het meest voorkomen in alle monsters van een bepaald monstertype. Het geeft alleen niet weer wat de onderscheidende ASVs zijn; dus de ASVs die differentieel voorkomen in bepaald monstertype. Die informatie hebben we juist nodig om te bepalen welke ASVs indicatief zijn voor contaminatie en er moet daarvoor gebruik gemaakt worden van een differentiële abundantietestmethode, zoals Gneiss of ANCOM.

Gneiss

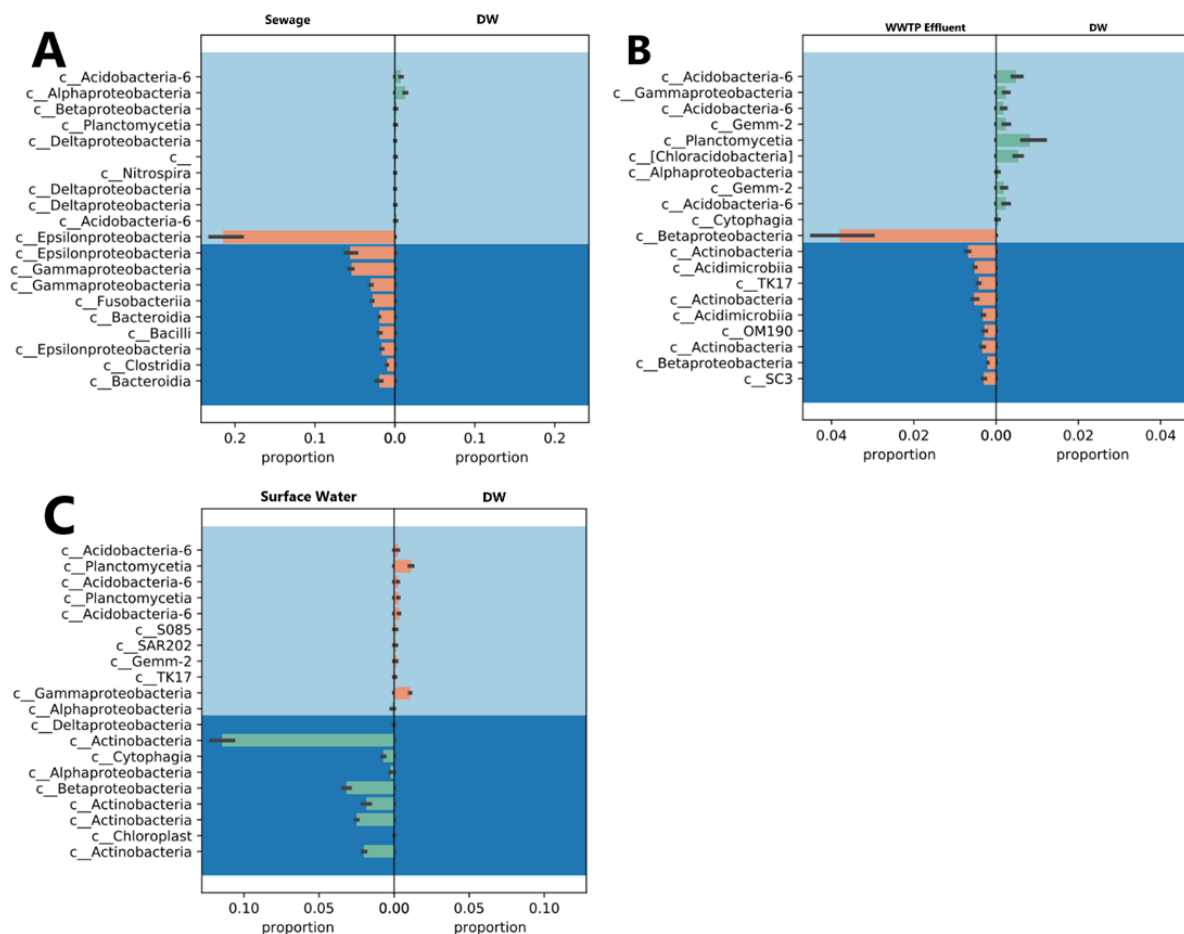
Gneiss is een Qiime2 plugin softwaretool dat gebruik maakt van gradiëntclustering om vervolgens als resultaat de abundantie van ASVs die specifieke zijn voor bepaalde monstertypes weer te geven in zogenaamde heatmaps (zie figuur 24). De gehele uitleg over Gneiss kan teruggelezen worden op de Qiime2 webpagina:

<https://docs.qiime2.org/2020.2/tutorials/gneiss/>. In het kort, Gneiss analyseert welke taxa meer voorkomen in een type monster t.o.v. een andere type monster. Binnen het onderzoek laten we drie vergelijkingen zien: de triplo drinkwatermonsters ten opzichte van de triplo monsters' van 100% rioolwater, oppervlaktewater en RWZI effluent, respectievelijk. De verdunningen van drinkwater met ieder van de drie contaminatiebronnen worden getoond, omdat de resultaten lieten zien dat Gneiss dan minder goed presteert. In Figuur 24 zijn de resultaten van de drie eerdergenoemde vergelijkingen te zien in de heatmaps. In deze heatmaps staan op de 'x-as' de triplo monsters van beide monsters en op de y-as de ASVs. De kleurcodering van de heatmap geeft aan in welk van de monsters de ASV meer voorkomt; hoe donkerder de kleur, hoe relatief meer deze ASV voorkomt in dat monster. Links van de heatmaps is een zogenaamde numerator en denominator te zien met een 'y'. Dit is simpel gezegd de samenvatting van de monsters van elk type. Wanneer ASVs in het algemeen meer voorkomen in de drinkwatermonsters, kleurt de numerator roze, wanneer ze meer voorkomen in de contaminatiebron kleurt de denominator rood. Het getal achter de 'y' staat voor het taxaniveau waartoe de ASVs zijn geclusterd; dus phylum (y0) tot en met soort (y9).



Figuur 24: Gneiss heatmap van 100% rioolwater (A), WWTP effluent (B), oppervlaktewater (C) vergeleken met drinkwater (DW). De kleurcodering van de heatmap staat voor de balans (balance, zie legenda), wat betekent dat bij een hogere balans, dat taxon relatief meer aanwezig is in het ene type monster t.o.v. het andere type monster. De numerator (drinkwater, roze) en denominator (contaminatiebron, rood) geven aan in welk type monster die taxa relatief meer voorkomen.

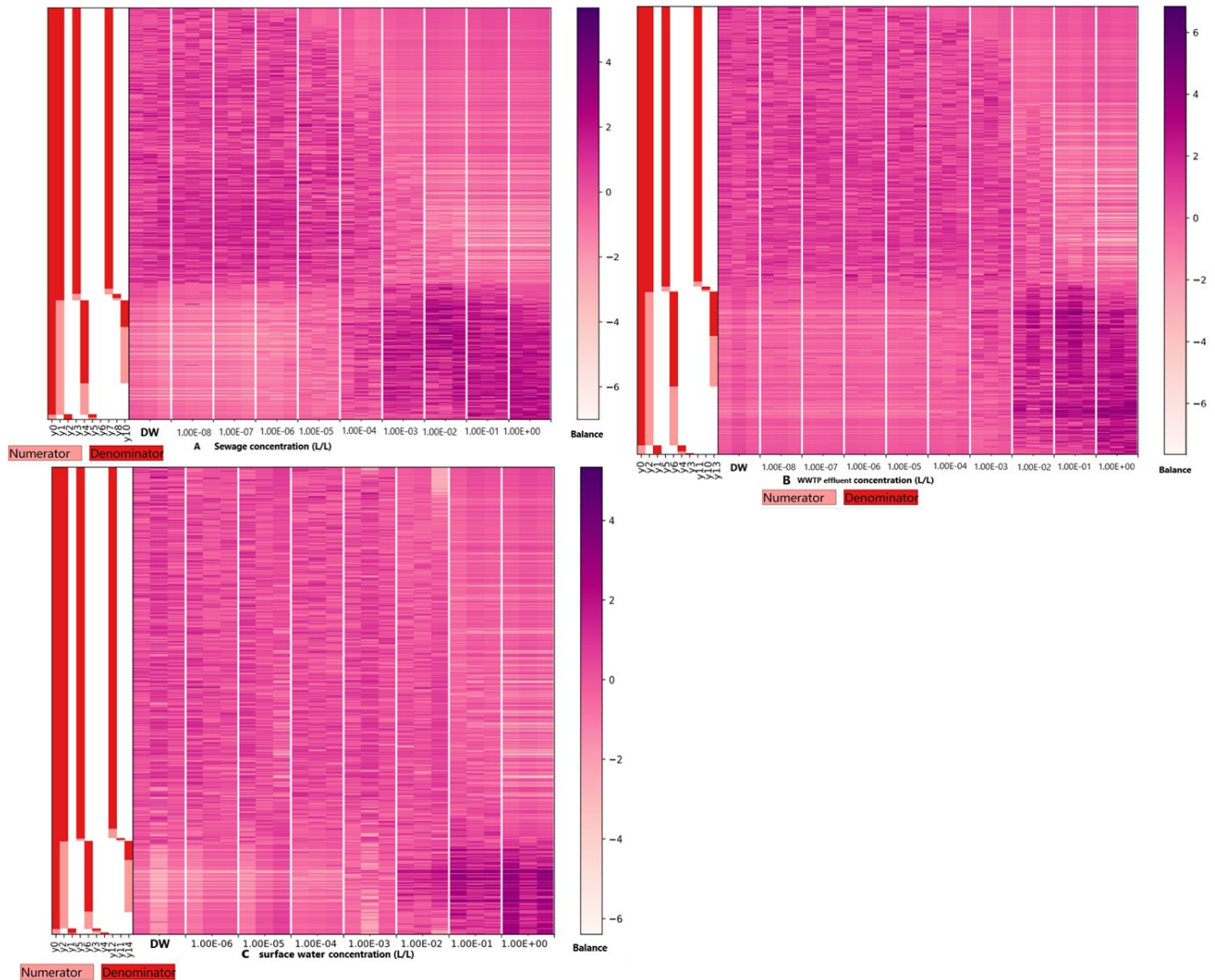
Om exact te achterhalen welke taxa-geclusterde ASVs meer voorkomen in een bepaald type monster, kan gebruikt gemaakt worden van “Qime gneiss balance-taxonomy”. De resultaten hiervan zijn gegeven in Figuur 25. In deze figuur zijn de taxa die meer voorkomen in de contaminatiebron (denominator) en welke in drinkwater (numerator) weergegeven tot klassenniveau. Het is duidelijk te zien dat in drinkwater ASVs die meestal behoren tot dezelfde klassen vaker voorkomen t.o.v. ASVs in de drie verschillende contaminatiebronnen, zoals bijvoorbeeld ASVs behorende tot de *Acidobacteria*, *Planctomycetia*, *Nitrospira*, *Gemmata* (Gemm-2), etc. In rioolwater komen daarentegen bijvoorbeeld bepaalde ASVs meer voor die behoren tot *Epsilonproteobacteria* en *Gammaproteobacteria*; bepaalde ASVs behorende tot de *Betaproteobacteria* zijn specifiek abundant in RWZI effluent en een ASV behorende tot de *Actinobacteria* was specifiek abundant in oppervlaktewater. Een onpraktisch onderdeel van de Gneiss softwaretool is dat je maar één taxonomisch niveau tegelijk kan analyseren en het daardoor arbeidsintensief is om de ASVs te bepalen die specifiek zijn voor een bepaald watertype. Verder waren er wat ‘bugs’ tijdens het exporteren van data in deze tool die mogelijk later verholpen worden in de ontwikkeling. Daarom wordt uit deze studie geconcludeerd dat de Gneiss softwaretool minder geschikt is voor het bepalen van differentiële abundantie van contaminatiebronnen in drinkwater, maar lijkt de tool wel geschikt voor het observeren van verschillen in taxa die meer of minder voorkomen tussen type monsters.



Figuur 25.: de Proportion plots van rioolwater (A), RWZI effluent (B), en oppervlaktewater water (C). De proportion plots geven ASVs weer die specifiek meer voorkomen in de denominator (contaminatiebron) of in de numerator (drinkwater).

Met Gneiss kan bijvoorbeeld wel gekeken worden welke taxa vooral veranderen tijdens contaminatie van drinkwater. Dit is een erg nuttige eigenschap, omdat we dan weten welke taxa vooral indicatief zijn voor contaminatie. Dit zijn zoals gezegd hoogstwaarschijnlijk de dominante taxa en met Gneiss kan dit in detail bekeken worden.

In Figuur 26 zijn deze resultaten te zien voor contaminatiereeksen van rioolwater (A), RWZI effluent (B) en oppervlaktewater (C). Hier is duidelijk te zien dat ASVs uit rioolwater bij een lagere verdunning nog steeds duidelijk aanwezig zijn in het drinkwaterprofiel, terwijl voor RWZI effluent en oppervlaktewater dit minder het geval is. Dit komt overeen met andere analyses waarbij rioolwaterinvloed gevoeliger te detecteren is dan de andere twee contaminatiebronnen, waarschijnlijk omdat het aantal bacteriën in rioolwater hoger is dan het aantal bacteriën in oppervlaktewater of RWZI-effluent.



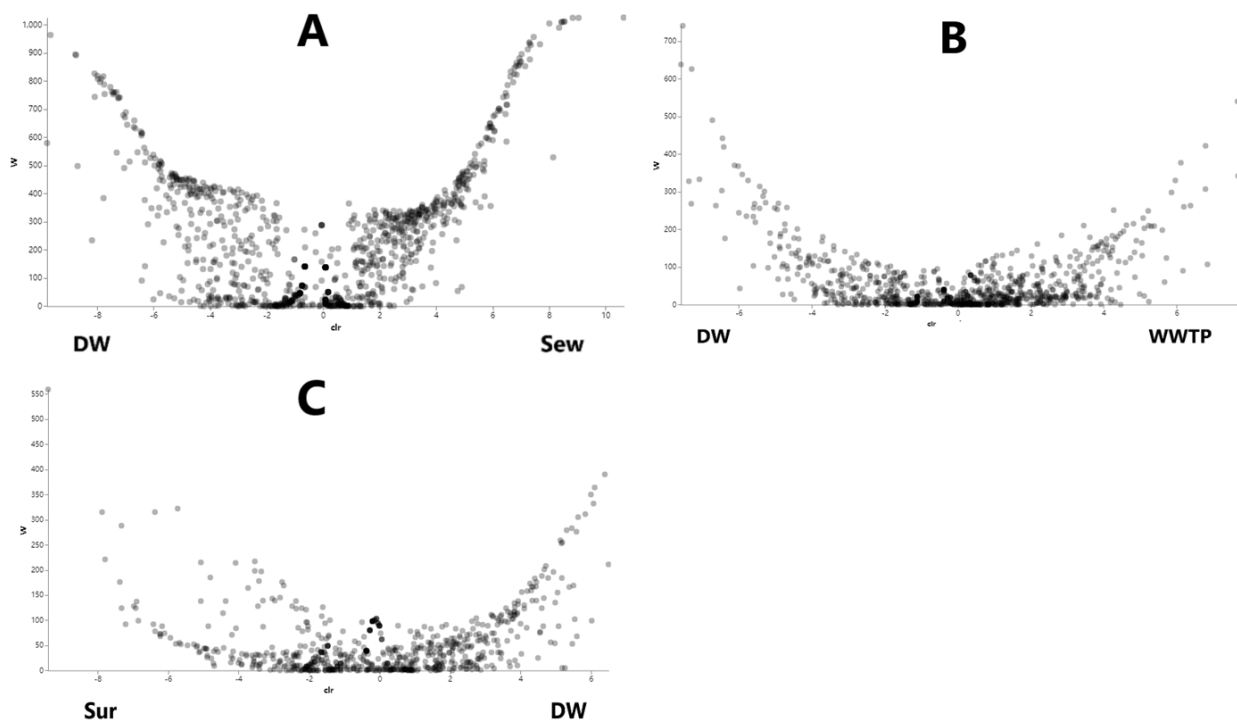
Figuur 26.: Gneiss heatmap van de verdunningsreeksen van contaminatie met rioolwater (A), WWTP effluent (B), oppervlaktewater (C) vergeleken met drinkwater (DW). De kleurcodering van de heatmap staat voor de balans (balance, zie legenda), wat betekent dat bij een hogere balans, dat taxon relatief meer aanwezig is in het type monster t.o.v. het andere type monster. De numerator (contaminatiebron, roze) en denominator (drinkwater, rood) geven aan in welk type monster die taxa relatief meer voorkomen.

Analysis of composition of microbiomes (ANCOM)

De 'analysis of composition of microbiomes (ANCOM)' is een andere softwaretool die gebruikt kan worden voor differentiële abundantie analyse binnen Qiime2 (Mandal, Van Treuren et al. 2015).

Deze tool gebruikt een andere manier om data te normaliseren dan met rarefying of relatieve abundantie. ANCOM identificeert taxa die differentiële abundantie vertonen tussen monsters door 1-voor-1 de log ratio van de abundantie van elk taxon te vergelijken met de abundantie van alle andere taxa in een monster (de zogenaamde compositionele logratio). Dit is een vorm van normaliseren die minder vals positieven zal identificeren. Na

vergelijking van alle monsters binnen een matrix die getest wordt, wordt het aantal keer dat de hypothese wordt afgewezen vastgelegd in een zogenaamde W-waarde. Hoe hoger deze W-waarde, hoe groter de kans dat de differentiële abundantie van die ASV statistisch significant is en dus vooral voorkomt in één matrix. Daarnaast wordt de compositionele logratio gegeven om aan te geven in welke monstertype die ASV differentieel abundant is. De waarden worden gevisualiseerd in een volcano plot met op de x-as de compositionele log-ratio en op de y-as de W-waarde, waarbij elk punt een specifieke ASV is (Figuur 27). Daarnaast wordt in een tabel gegeven welke ASVs statistisch significant meer voorkomen in een monstertype (Tabel 6). De significantie is bepaald door een grens van de W-waarde te stellen, waarbij alles boven die grens significant verschillend is. De grens van deze W-waarde wordt bij elke analyse opnieuw gesteld en is afhankelijk van de compositie van de data.



Figuur 27: Volcano plots van de ANCOM differentiële abundantie resultaten analyse op triplo metingen van rioolwater (A), RWZI effluent (B) en oppervlaktewater (C) vergeleken met drinkwater (DW). De x-as toont de compositionele log-ratio (clr) voor de verschillende monstertypes. De y-as toont de W-waarde die staat voor het aantal keer dat de nul hypothese wordt afgewezen.

Tabel 6: Taxa met de hoogste W-waarde in ANCOM analyse van rioolwater (sewage) vergeleken met drinkwater (DW). In de kolommen wordt de taxonomie van de ASVs gegeven, de percentiel van het aantal reads voor beide monstergroepen voor elke ASV, de W-waarde en of de nul hypothese afgewezen is. Een read van 1 betekent 0, omdat ANCOM logaritmisch is en pseudocounts gebruikt. Taxa met een W-waarde lager dan de gedefinieerde grens worden geannoteerd als 'FALSE' voor het afwijzen van de nulhypothese en zijn dus niet significant verschillend.

Group	DW					Sewage					W	Reject null hypothesis
	0	25	50	75	100	0	25	50	75	100		
Percentiel	0	25	50	75	100	0	25	50	75	100		
k__Bacteria;p__Proteobacteria;o__Epsilonproteobacteria;o__Campylobacteriales;f__Campylobacteraceae;g__Arcobacter;s__cryaeophilus	1	1	1	1	1	46774	53015	59256	60458.5	61661	1026	TRUE
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__	1	1	1	1	1	7246	8397	9548	10271	10994	1025	TRUE
k__Bacteria;p__Proteobacteria;o__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__	1	1	1	1	1	8968	10540.5	12113	12594	13075	1025	TRUE
k__Bacteria;p__Proteobacteria;o__Epsilonproteobacteria;o__Campylobacteriales;f__Campylobacteraceae;g__Arcobacter;s__	1	1	1	1	1	5123	6029.5	6936	7958	8980	1012	TRUE
k__Bacteria;p__Proteobacteria;o__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;s__	1	1	1	1	1	5498	5974.5	6451	7393	8335	1011	TRUE
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;s__prausnitzii	1	1	1	1	1	5006	5408	5810	6843.5	7877	1010	TRUE
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__	1	1	1	1	1	4908	5592	6276	7077	7878	1010	TRUE
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Carnobacterium;s__	1	1	1	1	1	3297	3518	3739	4445.5	5152	1005	TRUE
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__copri	1	1	1	1	1	4504	5143.5	5783	6341	6899	990	TRUE
k__Bacteria;p__Nitrospirae;c__Nitrospiria;o__Nitrospirales;f__Nitrospiraceae;g__Nitrospira;s__	8194	10860.5	13527	14044	14561	1	1	1	1	1	964	TRUE
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia;s__faecis	1	1	1	1	1	1912	2080	2248	2525.5	2803	957	TRUE
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Macellibacteroides;s__fermentans	1	1	1	1	1	1596	1765	1934	2218.5	2503	938	TRUE
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;s__	1	1	1	1	1	1588	1729	1870	2240.5	2611	934	TRUE
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella;s__aerofaciens	1	1	1	1	1	2341	2475.5	2610	3232	3854	931	TRUE
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__	1	1	1	1	1	1697	1907.5	2118	2283.5	2449	928	TRUE
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;s__	1	1	1	1	1	1454	1617.5	1781	2135	2489	913	FALSE
k__Bacteria;p__Proteobacteria;o__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Tolomonas;s__	1	1	1	1	1	1197	1332.5	1468	1687.5	1907	896	FALSE
k__Bacteria;p__Proteobacteria;o__	2835	4566.5	6298	6378	6458	1	1	1	1	1	895	FALSE
k__Bacteria;p__Planctomycetes;c__Planctomycetia;o__Planctomycetales;f__Planctomycetaceae;g__Planctomycetes;s__	2828	4438.5	6049	6141.5	6234	1	1	1	1	1	892	FALSE

Tabel 7: Taxa met de hoogste W-waarde in ANCOM analyse van RWZI effluent vergeleken met drinkwater (DW). In de kolommen wordt de taxonomie van de ASVs gegeven, de percentielen van het aantal reads voor beide monstergroepen voor elke ASV, de W-waarde en of de nul hypothese afgewezen is. Een read van 1 betekent 0, omdat ANCOM logaritisch is en pseudocounts gebruikt. Taxa met een W-waarde lager dan de gedefinieerde grens worden geannoteerd als 'FALSE' voor het afwijzen van de nulhypothese en zijn dus niet significant verschillend.

Group	DW					WWTP effluent					W-value	Reject null hypothesis
	0	25	50	75	100	0	25	50	75	100		
Percentiel	0	25	50	75	100	0	25	50	75	100		
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Hyphomicrobium;s__	363	494	625	708.5	792	90	96.5	103	158	213	741	TRUE
k__Bacteria;p__Proteobacteria;o__	1169	2282	3395	3565.5	3736	39	54.5	70	113	156	638	FALSE
k__Bacteria;p__Proteobacteria;o__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobacteraceae;g__Desulfobacterium;s__	4	4	4	4.5	5	1	1	1	1	1	626	FALSE
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Sulfuritalea;s__	8	20	32	35	38	471	528	585	814.5	1044	540	FALSE
k__Bacteria;p__Acidobacteria;c__DA052;o__Ellin6513f;g__	5	18	31	32	33	1	1	1	1	1	490	FALSE
k__Bacteria;p__Planctomycetes;c__Phycisphaera;o__mle1-8f;g__	7	15	23	30.5	38	12	18.5	25	27.5	30	442	FALSE
k__Bacteria;p__Actinobacteria;c__Actinobacteriia;o__Actinomycetales;f__Intrasporangiaceae;s__	1	1	1	1	1	226	245	264	280.5	297	422	FALSE
k__Bacteria;p__OP3;c__o__f;g__	28	72	116	116	116	1	1	1	1	1	419	FALSE
k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__koll13;g__	1	1	1	2	3	1	1	1	1	1	377	FALSE
k__Bacteria;p__Planctomycetes;o__	3	7.5	12	15	18	1	1	1	1	1	370	FALSE
k__Bacteria;p__Nitrospirae;c__Nitrospiria;o__Nitrospirales;f__Thermodesulfobivibrionaceae;g__LCP-6p__	1	1	1	2	3	1	1	1	1	1	368	FALSE
k__Bacteria;p__AC1;c__o__f;g__	1	2	3	3.5	4	1	1	1	1	1	346	FALSE
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella;s__	1	1	1	1	1	24	50.5	77	81	85	342	FALSE
k__Bacteria;p__OP3;c__koll11;o__f;g__	269	553	837	947	1057	148	211	274	307.5	341	333	FALSE
k__Bacteria;p__Proteobacteria;o__Gammaproteobacteria;o__34P16f;g__	30	48	66	73	80	1	1	1	1	1	330	FALSE
k__Bacteria;p__Verrucomicrobia;c__Verruco-5;o__WCHB1-41f;WCHB1-25;g__	1	1	1	1	1	1	1	1	2	3	330	FALSE
k__Bacteria;p__Gemmatimonadetes;c__Gemm-1;o__f;g__	75	112	149	155	161	404	444	484	543	602	328	FALSE
k__Bacteria;p__Chloroflexi;c__Dehalococcoidetes;o__Dehalococcoidales;f__Dehalococcoidaceae;g__Dehalococcoides;s__	1	2.5	4	4.5	5	1	1	1	1	1	314	FALSE
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__C39;s__	1	1	1	1	1	93	114.5	136	144	152	307	FALSE
k__Bacteria;p__Proteobacteria;o__Gammaproteobacteria;o__Methylocoales;f__Crenotrichaceae;g__Crenotrichum;s__	161	320	479	619.5	760	1	2	3	3	3	303	FALSE

Tabel 8: Taxa met de hoogste W-waarde in ANCOM analyse van oppervlaktewater vergeleken met drinkwater (DW). In de kolommen wordt de taxonomie van de ASVs gegeven, de percentielen van het aantal reads voor beide monstergroepen voor elke ASV, de W-waarde en of de nul hypothese afgewezen is. Een read van 1 betekent 0, omdat ANCOM logaritisch is en pseudocounts gebruikt. Taxa met een W-waarde lager dan de gedefinieerde grens worden geannoteerd als 'FALSE' voor het afwijzen van de nulhypothese en zijn dus niet significant verschillend.

Group	DW					Surface water					W-value	Reject null hypothesis
	0	25	50	75	100	0	25	50	75	100		
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Gordoniaceae;g__Gordonia;s__	1	1	1	1	1	1	2	3	3	3	559	TRUE
k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes;o__f__g__s__	3	5.5	8	24.5	41	1	2	3	4.5	6	390	FALSE
k__Bacteria;p__Acidobacteria;c__Holophagae;o__Holophagales;f__Holophagaceae;g__s__	1	3	5	13.5	22	1	1	1	3	5	364	FALSE
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiales;g__Pedomicrobium_s__australicum	1	1	1	3	5	1	1	1	1	1	350	FALSE
k__Bacteria;p__OP3;c__koll1;o__GIF10;f__g__s__	357	421	485	1796	3107	1	1	1	2	3	332	FALSE
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__Methylosinus;s__sp__o__rflum	1	1	1	1	1	2	2	2	8.5	15	322	FALSE
k__Bacteria;p__Verrucomicrobia;c__Opitatae;o__Opitales;f__Opitaceae;g__s__	1	3	5	19.5	34	8	13	18	22	26	315	FALSE
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__ACK-M1;g__s__	1	1	1	3.5	6	5474	9571	13668	14669	15670	315	FALSE
k__Bacteria;p__Planctomycetes;c__C6;o__d113;f__g__s__	1	5	9	17	25	1	1	1	1	1	311	FALSE
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Myxococaceae;g__s__	8	14	20	28	36	1	1	1	2.5	4	305	FALSE
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__ACK-M1;g__s__	1	1	1	1	1	645	1127	1609	1752.5	1896	288	FALSE
k__Bacteria;p__Acidobacteria;c__Chloracidobacteria;o__RB41;f__g__s__	66	75	84	225.5	367	1	1	1	1	1	283	FALSE
k__Bacteria;p__GN02;c__GN07;o__f__g__s__	1	8	15	44.5	74	1	1	1	1	1	279	FALSE
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__f__g__s__	802	1034	1266	3905	6544	3	5.5	8	8	8	276	FALSE
k__Bacteria;p__Chloroflexi;c__SHA-26;o__f__g__s__	10	13.5	17	55.5	94	1	1	1	1	1	259	FALSE
k__Bacteria;p__BRC1;c__NPL-UPA2;o__f__g__s__	1	1	1	5	9	1	1	1	1	1	255	FALSE
k__Bacteria;p__WS2;c__Kazan-3B-09;o__f__g__s__	7	9	11	43.5	76	1	1	1	2	3	253	FALSE
k__Bacteria;p__Cyanobacteria;c__Chloroplasto;f__g__s__	1	1	1	2	3	56	104.5	153	153.5	154	221	FALSE
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Sediminibacter;s__furfu__rosus	1	1	1	1	1	1	4.5	8	8.5	9	217	FALSE
k__Bacteria;p__Cyanobacteria;c__ML6351-21;o__f__g__s__	31	35	39	118	197	1	2.5	4	5	6	215	FALSE

In Figuur 27 is te zien dat alle volcano-plots een duidelijk verschil laten zien in differentiële abundantie van een aantal ASVs voor beide vergeleken matrices (drinkwater versus rioolwater, oppervlaktewater of RWZI-effluent). Wanneer deze grafieken geopend worden in Qiime2 view, kan ook bekeken worden welke ASV elk punt in de grafiek is. In de tabellen worden de ASVs gegeven met de hoogste W-waarde en hoeveel reads elk monstertype (100% contaminaties en drinkwater) dat getest is bevat. Omdat we de matrices die we vergelijken in triplo per matrix geanalyseerd hebben, worden hier het aantal reads per percentiel (0, 25, 50, 75, 100) gegeven. Daarnaast wordt de W-waarde die bij elke ASVs hoort gegeven, met het resultaat of de nulhypothese afgewezen wordt (False) of niet (True). Wanneer de nulhypothese wordt afgewezen, betekent dit dat de ASV significant verschilt tussen de twee matrices.

Voor rioolwater is duidelijk te zien dat de ASVs behorende bij de genera *Arcobacter*, *Acinetobacter*, *Faecalibacterium*, etc significant meer voorkomen in rioolwater dan in drinkwater (Tabel 6). Dit bevestigt de bovengenoemde Gneiss resultaten waarbij duidelijk werd dat ASVs behorende tot de klassen *Epsilonbacteria* (waartoe *Arcobacter* behoort) en *Clostridia* (waartoe *Faecalibacterium* behoort) meer voorkomen in rioolwater. De kern-ASVs voor rioolwater bevatten ook ASVs die behoren tot deze taxa. Ook wordt bevestigd dat *Nitrospira* en *Planctomycetaceae* meer voorkomen in drinkwater.

In oppervlaktewater t.o.v. drinkwater komen ASVs behorende tot *Actinomycetales* familie ACK-M1, *Verrucomicrobia* en Chloroplasten (van plantenresten in oppervlaktewater) meer voor (Tabel 8). In RWZI effluent t.o.v. drinkwater komen ASVs behorende tot *Actinomycetales*, *Rhodocyclales*, *Coriobacteriales*, en *Gemmatimonadetes* Gemm-1 meer voor (Tabel 7). Voor RWZI effluent en oppervlaktewater zijn er veel minder ASVs significant verschillend volgens ANCOM vergeleken met rioolwater. Een reden hiervan is dat ANCOM een zeer conservatieve methode is om differentiële abundantie te bepalen en de verschillen tussen RWZI effluent en oppervlaktewater t.o.v. drinkwater blijkbaar kleiner zijn dan met rioolwater. Hierdoor ontstaan minder valspositieven in de resultaten, maar dus ook minder werkelijke positieven. Daarnaast moet de data aan een aantal voorwaarden voldoen, de microbiële populaties die vergeleken worden mogen bijvoorbeeld niet te verschillend zijn (<25% van de ASVs mogen anders zijn tussen de groepen) en er kunnen maar twee condities vergeleken worden. Wanneer dit niet het geval is zullen de resultaten niet naar behoren zijn. Het lijkt desondanks dat deze resultaten valide zijn, omdat ook de kern-ASVs van het RWZI-effluent behoren tot de *Rhodocyclales* (genus *Thauera* en “*Candidatus Accumilibacter*”) en *Coriobacteriales* (genus *Collinsella*). De kern-ASVs van oppervlaktewater behoren

ook vooral tot de *Actinomycetales* familie ACK-M1. Daarnaast vertonen de Gneiss resultaten ook dat het RWZI-effluent en oppervlaktewater meer ASVs bevatten behorende tot de *Actinobacteria* (waar *Actinomycetales* onder valt) en dat oppervlaktewater meer Chloroplasten bevat. De overeenkomst tussen de kerntaxamethode, Gneiss en Ancom geeft aan dat ANCOM zeer goed bruikbaar is om differentiële abundantie van ASVs tussen monsters aan te tonen. Ondanks dat ANCOM leidt tot de detectie van minder significante differentiële ASV-abundantie, is ANCOM robuuster, accurater en objectiever dan een handmatige analyse om ASVs te identificeren die als indicatorsoort zou kunnen dienen.

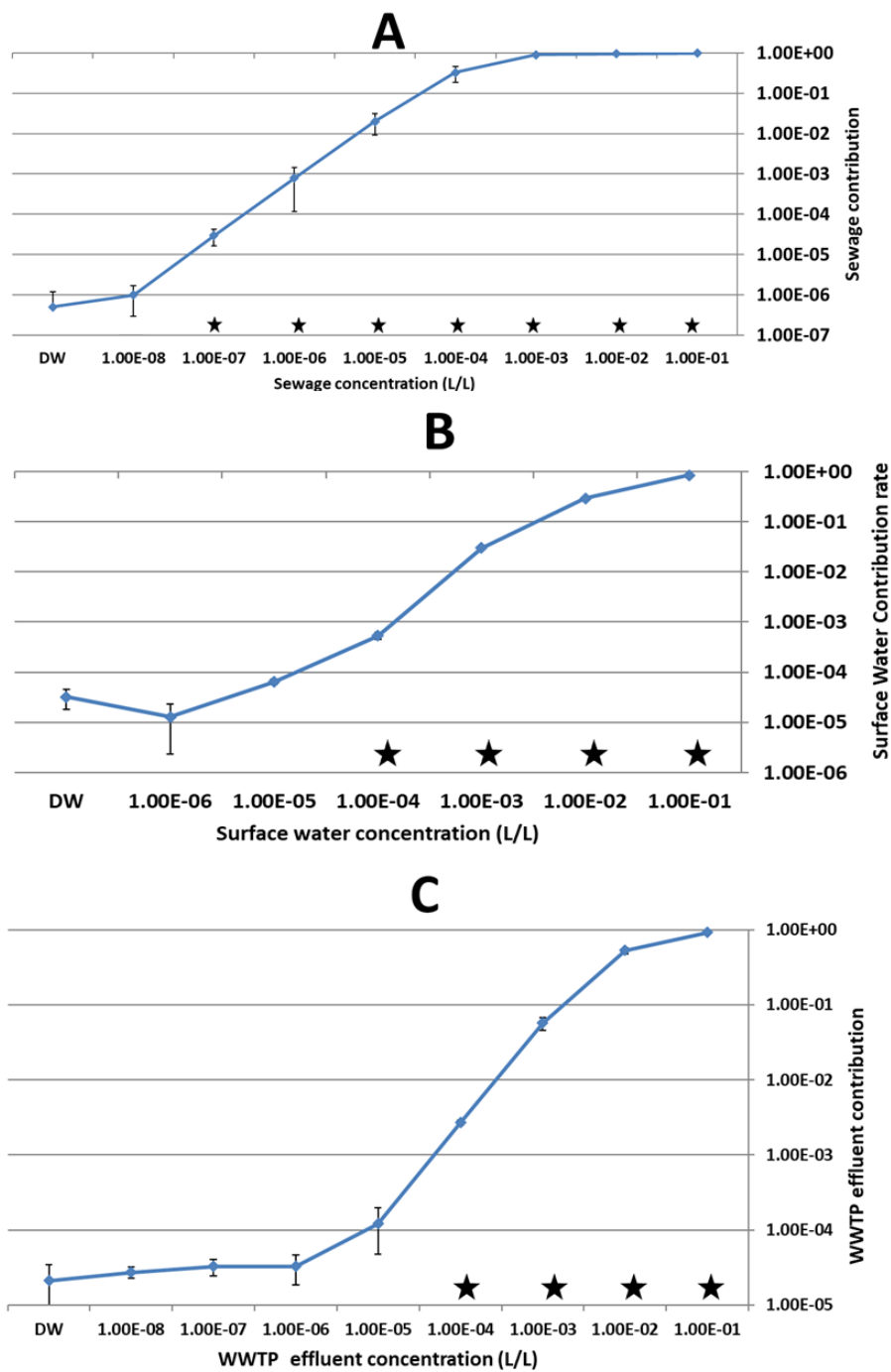
Bovenstaande technieken dienen dus om indicatorsoorten aan te wijzen uit verschillende contaminaties. De geïdentificeerde indicatorsoorten zouden mogelijk gebruikt kunnen worden in te ontwikkelen (RT)-qPCR analyses, waarbij de contaminatie gekwantificeerd kan worden, of erg gevoelig gedetecteerd kan worden.

Wanneer men geïnteresseerd is om contaminatie op te sporen door in eerste instantie te kijken naar de gehele microbiële gemeenschap biedt source tracker uitkomst, wat hieronder uitgelegd wordt.

7.1.4 Source Tracker

Er is in dit project ook een Qiime2 plugin softwaretool getest, genaamd Source Tracker (Knights, Kuczynski et al. 2011, Staley, Kaiser et al. 2018). Deze tool is al vaker toegepast door KWR om te achterhalen wat de bron(nen) zijn van verschillende watersoorten. Zo is deze tool toegepast op NGS-data van drinkwater en op NGS-data van, rioolwater en RWZI effluent om te achterhalen of verontreinigingen van deze twee bronnen te traceren zijn in oppervlaktewater en grondwater (Timmers 2020). In het in dit rapport beschreven experiment is onderzocht hoe gevoelig rioolwater, oppervlaktewater en RWZI effluent te traceren zijn in drinkwater. Source tracker maakt in tegenstelling tot de eerder beschreven methoden gebruik van de gehele microbiële populatie als vingerafdruk/indicator voor contaminatie. Dit maakt dat Source Tracker mogelijk gevoeliger is dan het gebruik van één of enkele indicatorsoorten en maakt het, door alle informatie uit de "fingerprint" te gebruiken, wellicht mogelijk om ook de verontreinigingsbron te achterhalen.

De tool Source Tracker gebruikt het microbiële profiel verkregen uit de contaminatiebronnen en vergelijkt deze met het microbiële profiel van de monsters die mogelijk gecontamineerd zijn. De resultaten worden weergegeven als percentages van de populatie die van de contaminatiebron afkomen. In Figuur 28 zijn de resultaten weergegeven voor de experimenten waar het rioolwater (A), oppervlaktewater (B) en RWZI effluent (C) als contaminatiebron in verdunningsexperimenten gebruikt zijn. Resultaten tonen dat voor verontreiniging van drinkwater met rioolwater al een significant verschil in de microbiële populatie van drinkwater ontstaat bij 10^{-7} verdunning van rioolwater in drinkwater. Voor verontreiniging van drinkwater met oppervlaktewater en RWZI effluent is dit effect bij 10^{-4} verdunning significant.



Figuur 28: Contributie (% van de microbiële populatie) in verdunningsreeksen van riool (A), oppervlaktewater (B) en RWZI effluent (C) (L/L) in drinkwater (DW), bepaald met de tool Source Tracker. Punten in de grafiek aangegeven met een ster zijn significant verschillend van drinkwater (t-test, $p < 0.05$).

Met de eerdergenoemde beta-diversiteit waren er geen verschillen te zien in de gehele microbiële compositie bij 10^{-4} verdunning van drinkwater met een contaminatiebron. Dit verschil tussen de beta-diversiteitsanalyse en Source Tracker kan komen door het algoritme dat Source Tracker gebruikt. Deze algoritme maakt namelijk gebruik van de kans dat bepaalde ASVs uit een bepaalde omgeving komen. Daarnaast zou de Bray Curtis index een heel ander resultaat opleveren als de mock communities en de blanco monsters niet meegenomen worden in de analyse. De monsters zouden dan waarschijnlijk verder uit elkaar liggen. Normaliter wordt de data nogmaals

geanalyseerd zonder de negatieve en positieve controles wanneer er gecontroleerd is dat deze naar behoren zijn. We hebben deze met opzet in de dataset gelaten om te dienen als voorbeeld analyse in dit rapport.

Eerder onderzoek heeft aangetoond dat Source Tracker nog steeds rioolwaterbacteriën kon detecteren bij een rioolwaterverduunning van 10^{-7} , terwijl *E. coli* niet meer aangetroffen werd in een rioolwater verduunning van 10^{-5} (Heijnen 2016). Dit gold ook voor RWZI effluent waar Source Tracker RWZI-effluentbacteriën detecteerde bij een 10^{-4} verduunning, terwijl *E. coli* niet meer gedetecteerd werd bij een 10^{-2} verduunning. Deze resultaten tonen dat van de verschillende methoden die in de in dit rapport beschreven studie zijn onderzocht, Source Tracker momenteel de gevoeligste tool is om contaminatie van een andere bron in drinkwater te detecteren. Het enige nadeel van deze tool is dat het microbiële profiel van de besmettingsbron bepaald moet zijn, terwijl met andere methoden voorkennis hierin uitkomst kan bieden.

7.1.5 Conclusie

We hebben met verschillende bioinformaticatools aangetoond hoe een NGS-dataset van drinkwater, rioolwater, oppervlaktewater, RWZI-effluent en verduunningen van deze laatste drie bronnen in drinkwater geanalyseerd kan worden. Verder hebben we laten zien welke methoden wel of niet geschikt zijn voor het aantonen van contaminatie van een andere bron in drinkwater. We hebben onderzocht met welke methoden de karakteristieke ASVs voor een bepaald monstertype kan worden achterhaald. We hebben ook laten zien welke methoden bruikbaar zijn om differentiële abundantie van ASVs in bepaalde monstertypen te bepalen; dus welke ASVs komen meer voor in een bepaald monstertype dan in een ander. Elke methode heeft zijn voor- en nadelen en het is daarom belangrijk om voor elke onderzoeksvraag goed na te gaan welke methode het beste gebruikt kan worden.

In het uitgevoerde contaminatie-experiment is duidelijk geworden dat Source Tracker de beste en meest gevoelige methode is om te achterhalen of contaminatie van drinkwater vanuit andere bronnen heeft opgetreden. Daarnaast kan met differentiële abundantieanalyse (Gneiss en ANCOM) en kern-ASVs-analyse bepaald worden welke ASVs vooral indicatief kunnen zijn voor contaminatie van drinkwater met een andere bron en welke ASVs gebruikt zouden kunnen worden als indicatorsoorten voor die contaminatie.

7.2 Veranderingen in microbiële gemeenschappen tijdens zuivering

Om inzicht te krijgen in de potentiële toepassingsmogelijkheden van microbial profiling zijn er door de leden van de BTO themagroep “Biologische veiligheid” voor de deelnemende drinkwaterbedrijven monsters aangeleverd voor het uitvoeren van beperkte praktijkstudies. De gekozen praktijkstudies worden vooral gebruikt om inzicht te krijgen in de verandering die optreden in de microbiologie in het water tijdens verschillende zuiveringsstappen. Hiervoor zijn er monsters geanalyseerd afkomstig van verschillende locaties in acht verschillende zuiveringen. Omdat uitgebreide monitoring buiten de scope en budget van dit onderzoek viel, zijn deze analyses vooral uitgevoerd op éénmalige monsters waarvan, in het kader van andere projecten, al DNA verzameld was. Door de analyse van enkelvoudige monsters is de uitvoering van statistische tests, om de significantie van waargenomen verschillen te beoordelen, niet mogelijk. De uitgevoerde analyses hebben zich beperkt tot het vergelijken van de microbiologie op het niveau van de gehele gemeenschap. De basis voor deze vergelijkingen wordt gevormd door een beschrijving van de microbiële gemeenschap waarvoor de aanwezigheid en relatieve aandeel van de ASV's de basis vormt. Van de in de monsters aangetoonde ASV's is ook een taxonomische identificatie uitgevoerd tot op genusniveau. Deze analyses zijn dus niet bedoeld om de samenstelling van de microbiële populaties in de onderzochte zuiveringen in detail te beschrijven maar geven een eerste indruk van potentiële mogelijkheden van microbial profiling voor het verkrijgen van inzicht in de samenstelling van de microbiologie tijdens zuiveringsprocessen en de veranderingen die daarin optreden t.g.v. de toegepaste zuiveringsstappen. Vervolgonderzoek zal nodig zijn om inzicht te krijgen in de robuustheid van deze analyses, microbiële variaties gedurende de bedrijfsvoering en het optreden van eventuele

seizoensinvloeden. Voor visualisatie van data is, vanwege beperkingen in het budget, in dit hoofdstuk vooral gebruik gemaakt van de, snelle en eenvoudige, mogelijkheden die door Qiime2 worden geboden. Deze visualisatiemogelijkheden zijn vooral bedoeld om een snel beeld te krijgen van de resultaten maar deze zijn minder geschikt in rapportages. Voor betere visualisaties zijn er van een aantal experimenten analyses uitgevoerd in de statistische software RStudio waarmee betere visualisaties mogelijk zijn.

7.2.1 Zuivering productielocatie Berenplaat (Evides)

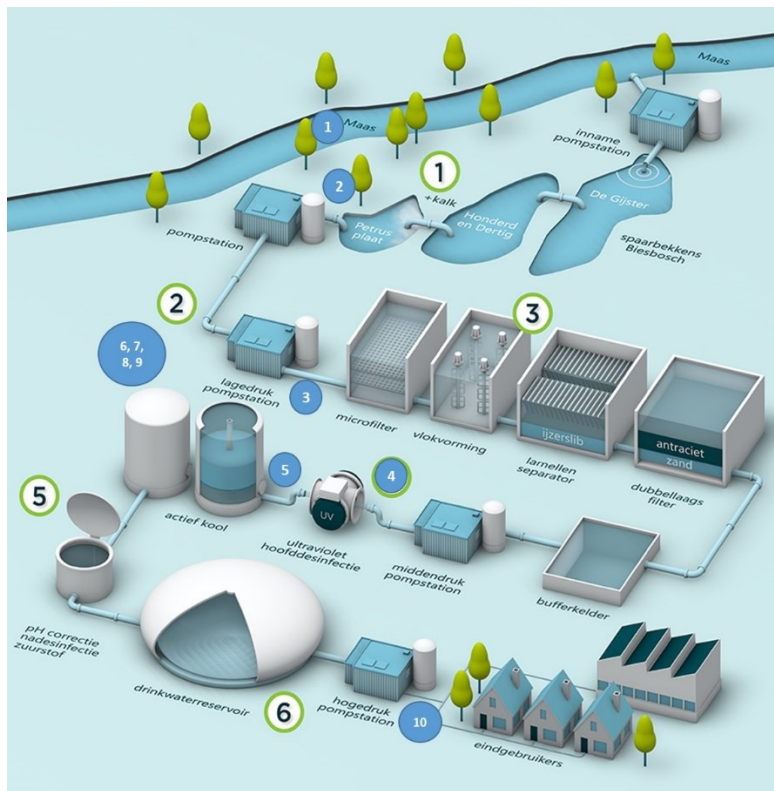
Bij oppervlaktewaterzuiveringslocatie Berenplaat is water bemonsterd na verschillende processen tijdens de zuivering van oppervlaktewater uit de Maas tot het geproduceerde drinkwater (Tabel 9, Figuur 29). De monsters zijn genomen op 7 februari in het winterseizoen bij een lage watertemperatuur van 4-5°C. Vervolgens is DNA geïsoleerd van deze monsters en is microbial profiling toegepast volgens het opgestelde standaardprotocol (Bijlage II).

Tabel 9. Geanalyseerde monsters van productiebedrijf Berenplaat van Evides

Monsternr.	Datum	Locatiecode	Locatie	Nr. in figuur 29*	Aantal reads**
70792	7-2-2019	RKEI00MEET	Maas (Keizersveer)	1	101.693
68979	10-1-2019	PPPT9M071VC	Uitlaat Petrusplaat	2	91.341
70793	7-2-2019	PPPT9M071VC	Uitlaat Petrusplaat	2	37.558
70794	7-2-2019	PBPL19LDXX	Lagedruk pompstation	3	118.624
70795	7-2-2019	PBPL47INFUV1	Aanvoer UV	4	108.259
70796	7-2-2019	PBPL50EFFUV1	Effluent UV	5	113.976
70797	7-2-2019	PBPL56KF11	Koolfiltraat_1.1	6	122.138
70798	7-2-2019	PBPL56KF12	Koolfiltraat_1.2	7	128.592
70799	7-2-2019	PBPL56KF21	Koolfiltraat_2.1	8	146.103
70800	7-2-2019	PBPL56KF22	Koolfiltraat_2.2	9	119.498
70801	7-2-2019	PBPL80HD2-2	Hogedruk Pompstation 2.2	10	108.364

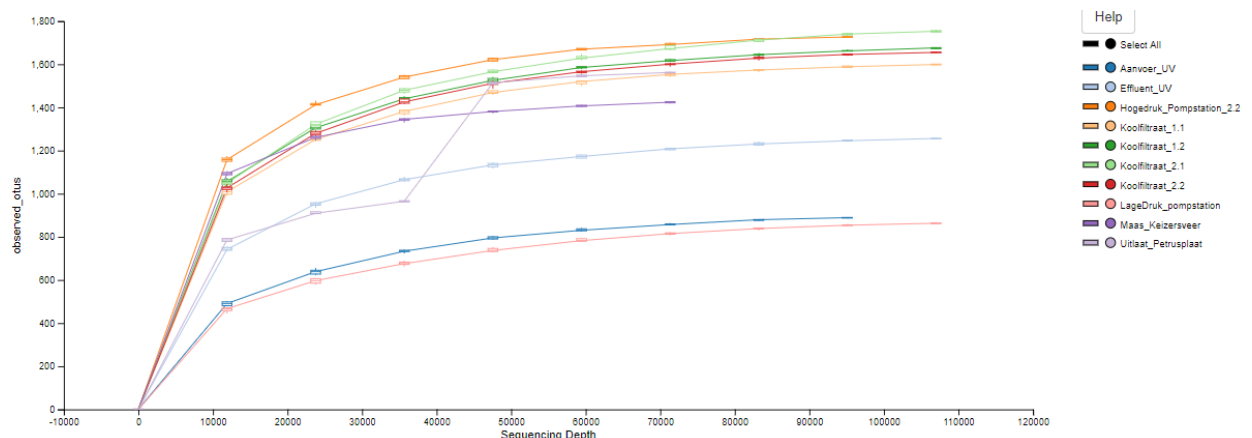
* Locaties zijn met blauwe cirkels weergegeven in Figuur 29

** Het totaal aantal kwalitatief goede sequentiereads (na het filteren van de sequentiereads met "Deblur")



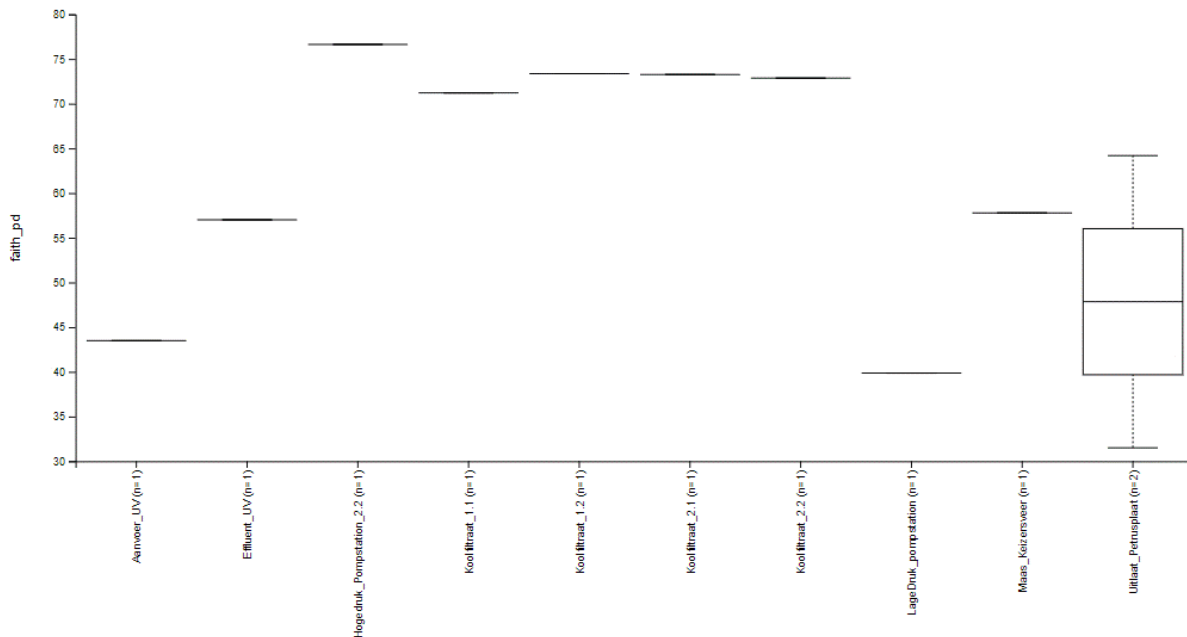
Figuur 29. Monsterlocaties (blauwe cirkels) bij productiebedrijf Berenplaat van Evides

Om te bepalen in hoeverre het aantal gegenereerde sequentiereads voldoende is geweest om een zo volledig mogelijk beeld te krijgen van de microbiële samenstelling van de populaties in de monsters zijn er rarefactionanalyses (zie paragraaf 5.3.1) uitgevoerd op de geanalyseerde monsters (Figuur 30). De rarefactionanalyses laten een afbuigende curve zien die aan het einde van de curve vrijwel horizontaal verloopt. Dit geeft aan dat de het vergroten van de sequencing diepte (van een groter aantal fragmenten de sequenties bepalen) niet zal leiden tot significant meer ASVs, waardoor geconcludeerd kan worden dat de verkregen sequencingdiepte een vrijwel volledig beeld geeft van de aanwezige ASVs in de monsters.



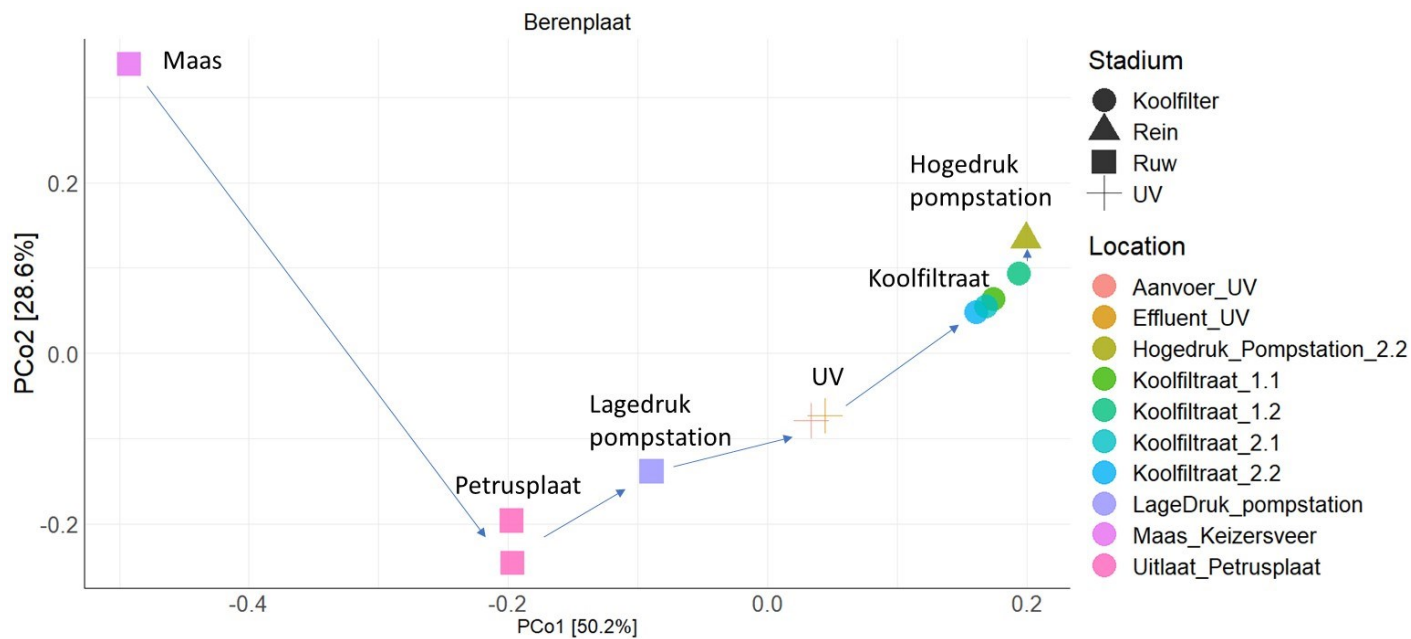
Figuur 30. Rarefactioncurves (figuur uit Qiime2) van de monsters van Berenplaat met de sequencingdiepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

Om een indruk te krijgen van de mate van diversiteit van de microbiële populaties, is de alpha-diversiteit geanalyseerd door de Faith's PD index te bepalen. De resultaten van deze analyse is weergegeven in Figuur 31. De analyses laten zien dat de laagste bacteriële diversiteit wordt aangetroffen in het ruwe oppervlaktewater en aan het begin van de zuivering. De diversiteit neemt toe in het traject na de UV-desinfectie door met name het biologische proces in de koolfilters en is het hoogst in het afgeleverde water na filtraatdesinfectie met chloordioxide en verblijf in de reinwaterkelder. Het analyseren van monsterreplica's is echter nodig om een betrouwbaarder inzicht te krijgen in de microbiële diversiteit in het water van de geanalyseerde locaties.



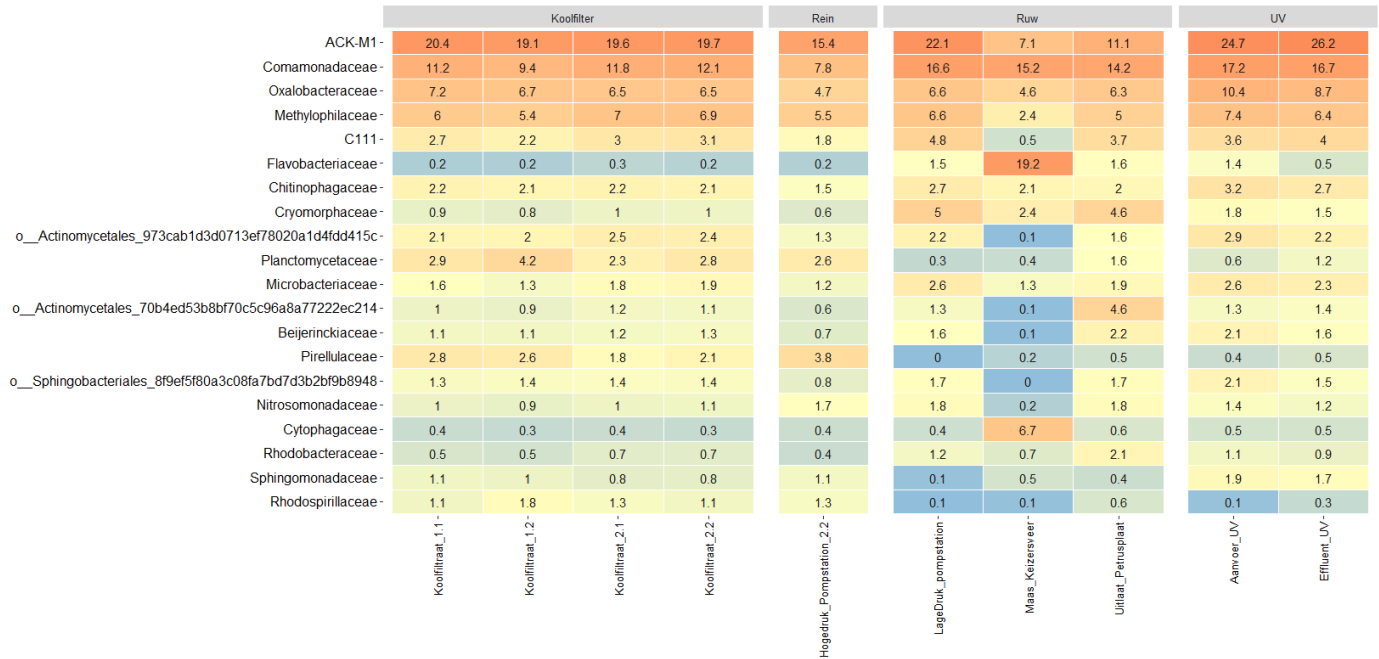
Figuur 31. Alpha-diversiteit (figuur uit Qiime2), uitgedrukt in Faith's PD index, van de bacteriële populaties in monsters van drinkwaterzuivering van Berenplaat

Om een beeld te krijgen van verschillen in de samenstelling van de microbiële populaties in het water bemonsterd na de verschillende zuiveringsprocessen zijn beta-diversiteitsanalyses uitgevoerd op basis van Bray-Curtis afstandsmatrix en vervolgens gevisualiseerd middels PCoA analyse (zie paragraaf 4.4.7). In Figuur 32 is een PCoA plot weergegeven waarmee verschillen tussen de populaties worden zichtbaar gemaakt. De PCoA-plot laat zien dat de bacteriële populatie in het water van de Maas bij Keizersveer zeer sterk verschilt van de andere monsters. Het verblijf in het spaarbekken van Petrusplaat zorgt voor een grote verandering in de microbiologie. Vervolgens vinden in de zuivering stapsgewijze veranderingen plaats in de microbiële gemeenschap tot het afgeleverde drinkwater (Hogedruk Pompstation). Het is daarbij opvallend dat de populatie in het water vóór UV vrijwel hetzelfde is als de populatie in het water na UV. Waarschijnlijk zal de UV behandeling wel effect hebben op de levensvatbaarheid van de micro-organismen in de populatie maar wordt deze niet zichtbaar in deze microbial profiles doordat het effect van UV op het DNA te beperkt is om met NGS-analyses zichtbaar te maken. In het traject van UV naar de koolfiltraten treden wel veranderingen op, de verschillen tussen de koolfilters zijn beperkt. Van koolfiltraat tot drinkwater uit de reinwaterkelder treden ook nog veranderingen op, die waarschijnlijk worden veroorzaakt door de nadesinfectie met een lage concentratie chloordioxide.



Figuur 32. De beta-diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van Berenplaat.

Een globale indruk van de samenstelling van de bacteriële populaties op familieniveau in de onderzochte monsters is weergegeven in een zogenaamde heatmap (Figuur 33). In deze figuur is de bacteriepopulatie weergegeven van de meest dominante ASVs in de populaties. Uit deze figuur volgt wederom dat een sterk afwijkende bacteriële populatie aanwezig is in het ruwe Maaswater, gedomineerd door de familie van *Flavobacteriaceae*. De relatieve aanwezigheid van deze familie neemt sterk af tijdens het verblijf in de spaarbekken en Petrusplaat waardoor het aandeel *Flavobacteriaceae* in de zuivering laag is. De familie van *Comamonadaceae* zijn dominant in alle onderzochte watermonsters. Bij het verder “inzoomen” tot op genusniveau blijkt dat er wel grote verschillen zijn tussen de leden van de *Comamonadaceae* familie in het ruwe oppervlaktewater en in de zuivering. Vanwege de grote diversiteit op genusniveau is het echter lastig om dit in figuren duidelijk te maken. In het ruwe Maaswater, behoren deze *Comamonadaceae* voor een groot deel tot het genus *Rhodoferrax*. De relatieve aanwezigheid van dit genus neemt echter sterk af tijdens het verblijf in de spaarbekken en Petrusplaat waardoor het aandeel *Rhodoferrax* bacteriën in de zuivering laag is. Het genus *Simplicispira* wordt juist weinig in het Maaswater aangetroffen, maar nemen toe tijdens de zuiveringsstappen van Maaswater tot UV en blijven dan aanwezig tot in het reinwater. Bacteriën behorende tot dit genus spelen mogelijk een rol bij denitrificatie (Siddiqi, Sok et al. 2020).



Figuur 33. Heatmap (figuur met AmpVis2 in RStudio) van de 20 meest dominante taxa op familieniveau in de onderzochte microbiële populaties in drinkwaterzuivering Berenplaat. De kleur geeft de mate van dominantie aan, de getallen geven het percentage sequenties behorende tot een taxon weer.

7.2.2 Zuivering Weesperkarspel (Waternet)

Om een eerste indruk te krijgen van de veranderingen die optreden van de microbiologie in het water van oppervlaktewaterzuivering Weesperkarspel zijn er analyses uitgevoerd op water afkomstig van verschillende stadia van deze zuivering. De monsters zijn afkomstig van enkele punten in het traject van oppervlaktewater wat de waterleidingplas ingaat tot het geproduceerde drinkwater (Tabel 10, Figuur 34). Om inzicht te krijgen in de reproduceerbaarheid van het protocol dat gebruikt wordt voor het genereren van de profielen zijn de analyses van de monsters in duplo uitgevoerd.

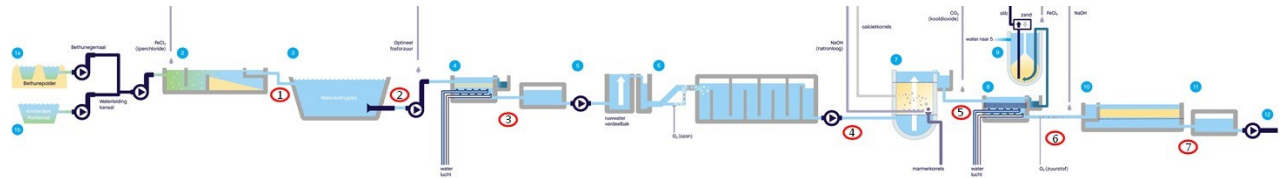
Tabel 10. Geanalyseerde monsters van productiebedrijf Weesperkarspel van Waternet

Monsternr.	Datum	Locatiecode	Locatie	Nr. in figuur*	Aantal reads**
84790	15-10-2019	WWL-VV-003	Inlaat Waterleidingplas	1	95817
84791	15-10-2019	WWL-VV-003	Inlaat Waterleidingplas	1	97477
84794	15-10-2019	PLV-SB-001	Waterleidingplas	2	107993
84795	15-10-2019	PLV-SB-001	Waterleidingplas	2	91616
84784	15-10-2019	PLV-VSF-001	Verzameld_Snelfiltraat_Loenderveen	3	107975
84785	15-10-2019	PLV-VSF-001	Verzameld_Snelfiltraat_Loenderveen	3	121984
84802	15-10-2019	PWK-KF-Inf001	Effluent_Ontharding	4	96550
84803	15-10-2019	PWK-KF-Inf001	Effluent_Ontharding	4	87903
84806	15-10-2019	PWK-OH-INF6313	Effluent_Ozonstraat	5	99018
84807	15-10-2019	PWK-OH-INF6313	Effluent_Ozonstraat	5	101382
84788	15-10-2019	PWK-VKF-003	Verzameld_Koofiltraat	6	90371
84789	15-10-2019	PWK-VKF-003	Verzameld_Koofiltraat	6	12720

84800	15-10-2019	PWK-VLZF-001	Verzameld_LangzameZandfiltraat	7	86336
84801	15-10-2019	PWK-VLZF-001	Verzameld_LangzameZandfiltraat	7	91421

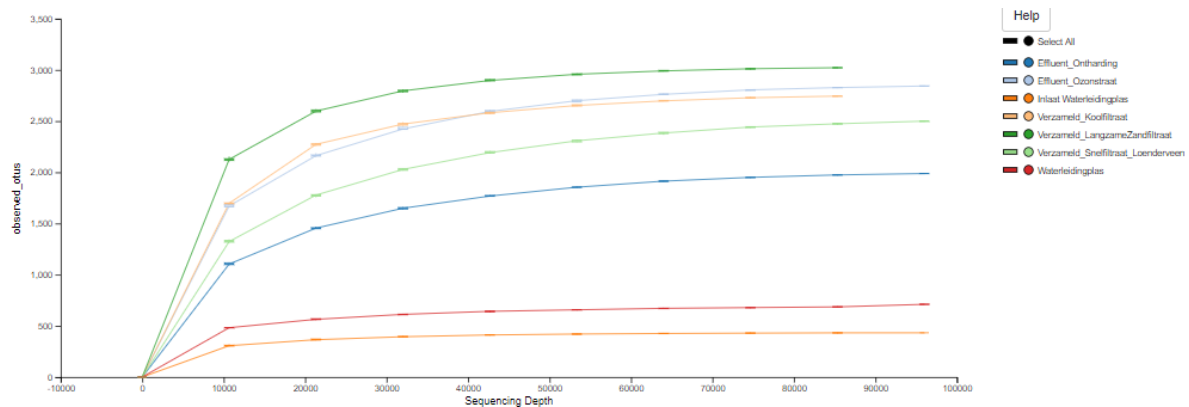
* Locaties zijn in Figuur 34 weergegeven in cirkels met rode rand

** Het totaal aantal kwalitatief goede reads (na het filteren van de reads met “Deblur”)



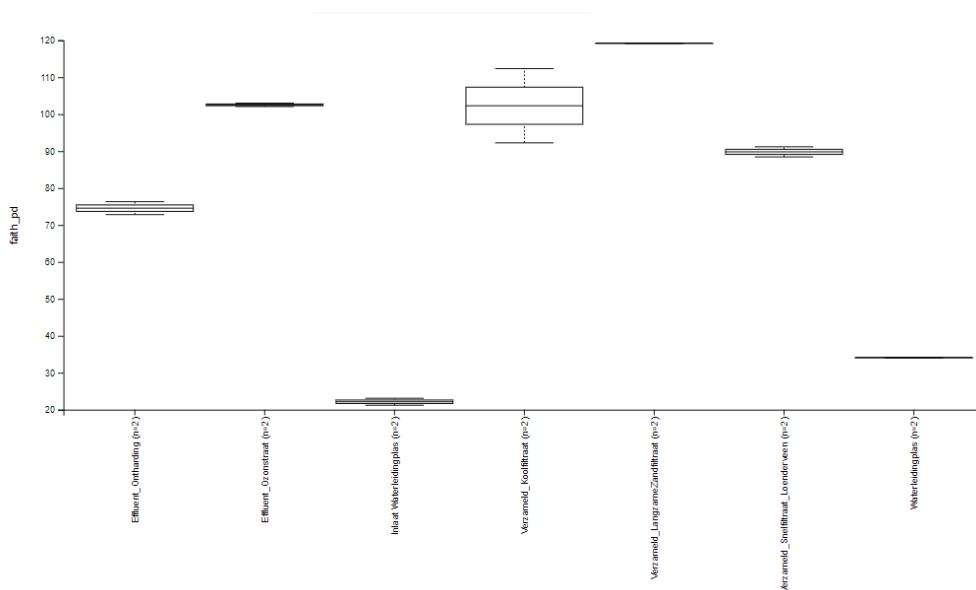
Figuur 34. Monsterlocaties (rood omcirkeld) bij productiebedrijf Weesperkarspel

De rarefactionanalyse (Figuur 35) laat ook voor deze monsters zien dat het aantal gegenereerde reads (Tabel 10) ruim voldoende is om een vrijwel volledig beeld te krijgen van de aanwezige ASV's in de monsters.



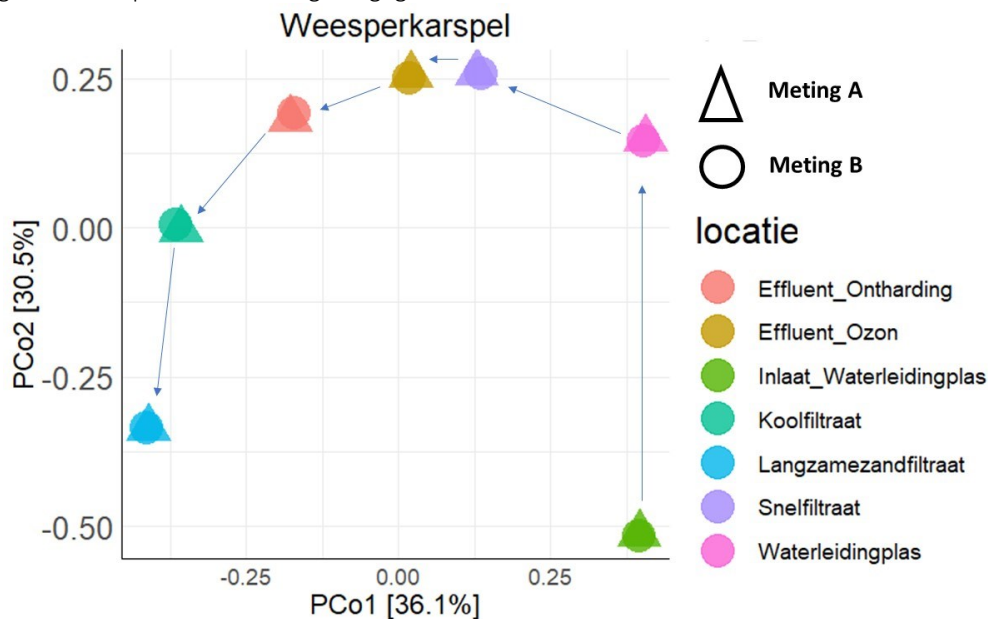
Figuur 35. Rarefaction curves van de monsters van Weesperkarspel met de sequencing diepte (x-as) en het aantal geobserveerde ASV's (y-as) in elk monster

De boxplots waarmee de alpha-diversiteit wordt weergegeven (Figuur 36) laat zien dat er weinig variatie is tussen duplo-analyses. En, net als bij de zuivering van Berenplaat, is de microbiële diversiteit het laagst in het oppervlaktewater dat het spaarbekken (Waterleidingglas) ingaat en neemt de diversiteit in de zuivering toe met de hoogste diversiteit in het water wat klaar is voor distributie (Langzame zandfiltraat). Deze waarnemingen worden bevestigd met de rarefactioncurves waarbij de monsters met minder microbiële diversiteit al tot een plateaufase komen bij een lagere sequencingdiepte.



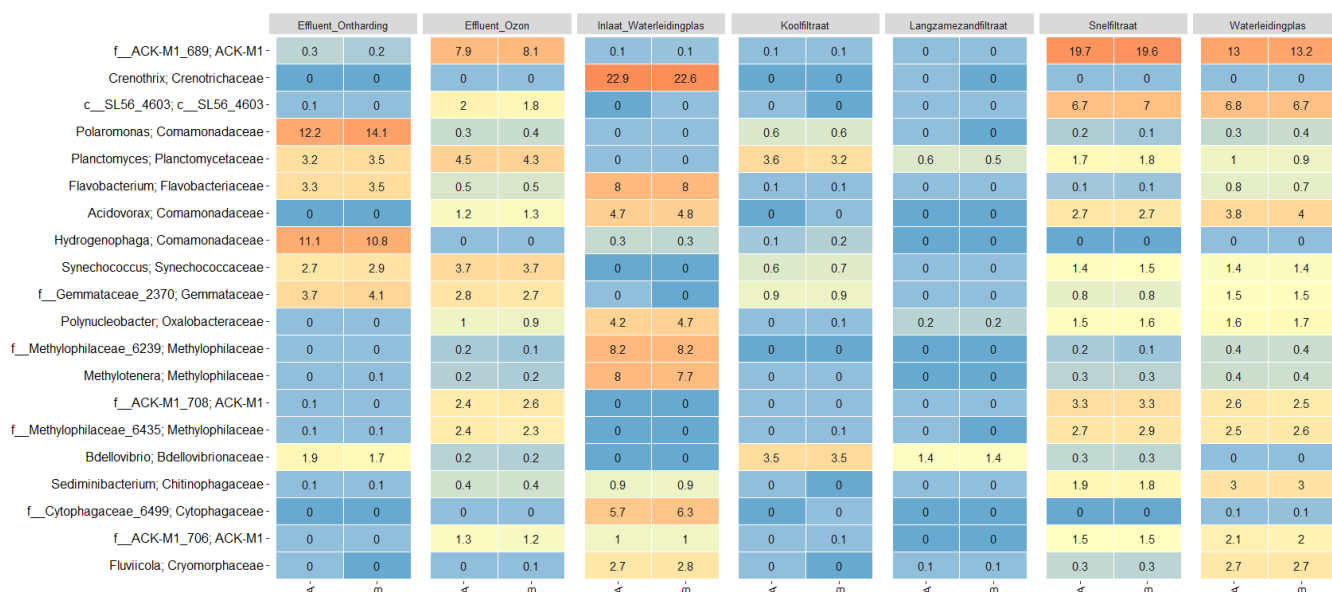
Figuur 36. Alpha-diversiteit, uitgedrukt in Faith's PD index, van de microbiële populaties in monsters van drinkwaterzuivering van Weesperkarspel

De PCoA plot waarin de beta-diversiteit wordt weergegeven (Figuur 37) laat zien dat er zeer weinig variatie is tussen duplomonsters (merk op dat elke kleur is samengesteld uit twee verschillende symbolen) waarbij de gehele procedure (watermonster/DNA-isolatie/PCR-reacties/sequencing/bioinformatica analyse) in duplo is uitgevoerd. De microbiële populaties in de monsters van Weesperkarspel ontwikkelen zich op een vergelijkbare manier als bij de zuivering van Berenplaat. Ook in Weesperkarspel is de microbiologie in het oppervlaktewater dat de waterleidingplas ingaat (inlaat Waterleidingplas) zeer afwijkend en deze microbiologie verandert sterk tijdens het verblijf in de Waterleidingplas (Waterleidingplas). In de zuivering verandert de samenstelling van de microbiologie, onder invloed van de processen in de zuivering, stapsgewijs tot een gemeenschap die sterk verschilt met de gemeenschap die de zuivering is ingegaan.



Figuur 37. De beta diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray-Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van Weesperkarspel.

Een meer gedetailleerd beeld van de samenstelling van de microbiële populaties in de onderzochte monsters van Weesperkarspel is weergegeven in de heatmap van Figuur 38.



Figuur 38. Weergave (figuur met AmpVis2 in RStudio) van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering Weesperkarspel. De taxa zijn gerangschikt op familie niveau, het genusniveau van de taxa is ook weergegeven

De data in de heatmap bevestigen de grote mate van overeenkomst tussen duploanalyses en de stapsgewijze veranderingen die ook in de PCoA plots zichtbaar zijn. Op genusniveau is duidelijk dat *Crenothrix* zeer dominant is in de inlaat van de waterleidingplas en, na verblijf in de waterleidingplas niet meer detecteerbaar is. Waarschijnlijk zijn deze methaanoxiderende bacteriën (Oswald, Graf et al. 2017) zeer dominant in het ruwe water van de Bethunepolder en het Amsterdam-Rijnkanaal. De onthardingsstap zorgt voor vermeerdering van bacteriën uit de familie van de *Comamonadaceae* (*Hydrogenophaga*).

7.2.3 Oppervlaktewaterzuivering “De Punt” (Waterbedrijf Groningen)

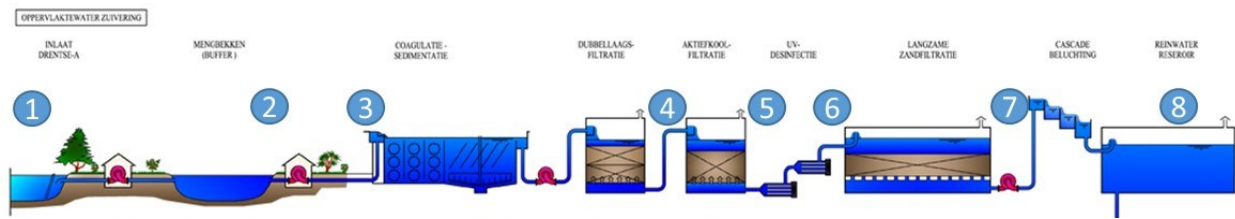
Bij drinkwaterzuivering “De Punt” wordt er drinkwater geproduceerd uit oppervlaktewater (uit de Drentse Aa) en grondwater en gemengd gedistribueerd. De watermonsters uit dit onderzoek zijn afkomstig van oppervlaktewaterzuivering. In Tabel 11 en Figuur 39 zijn de locaties aangegeven waarvan water is onderzocht.

Tabel 11. Geanalyseerde monsters van productiebedrijf “De Punt” van waterbedrijf Groningen

Monsternr.	Datum	Locatiecode	Nr. in figuur*	Aantal reads**
78116	26-6-2019	Rivierwater Drentse Aa	1	81474
78117	26-6-2019	Rivierwater Drentse Aa	1	84780
78118	26-6-2019	Effluent mengbekken	2	95556
78119	26-6-2019	Coagulatie Aanvoer FD	3	104310
78120	26-6-2019	Voeding actief kool filters	4	107230
78121	26-6-2019	Aanvoer UV - desinfectie (AKF 1-6)	5	112563
78122	26-6-2019	Aanvoer langzame zand filters	6	101650
78123	26-6-2019	Aanvoer cascade (verzameld filtaat)	7	73482
78124	26-6-2019	Reinwater	8	86991

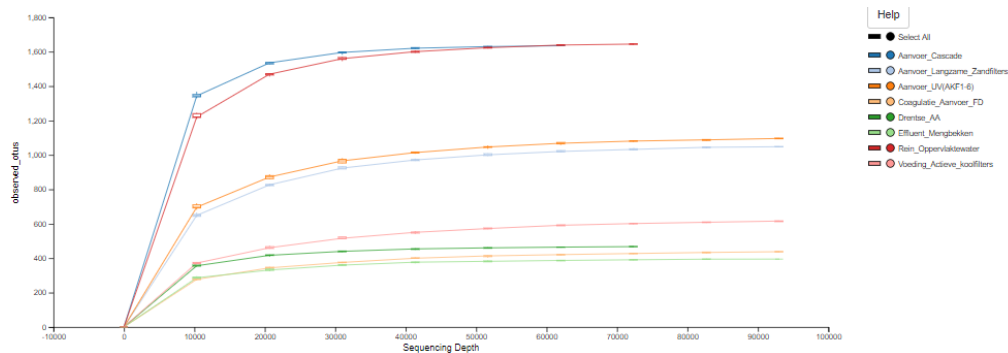
* Locaties zijn in **Error! Reference source not found.** weergegeven in blauwe cirkels

** Het totaal aantal kwalitatief goede reads (na het filteren van de reads met “Deblur”)



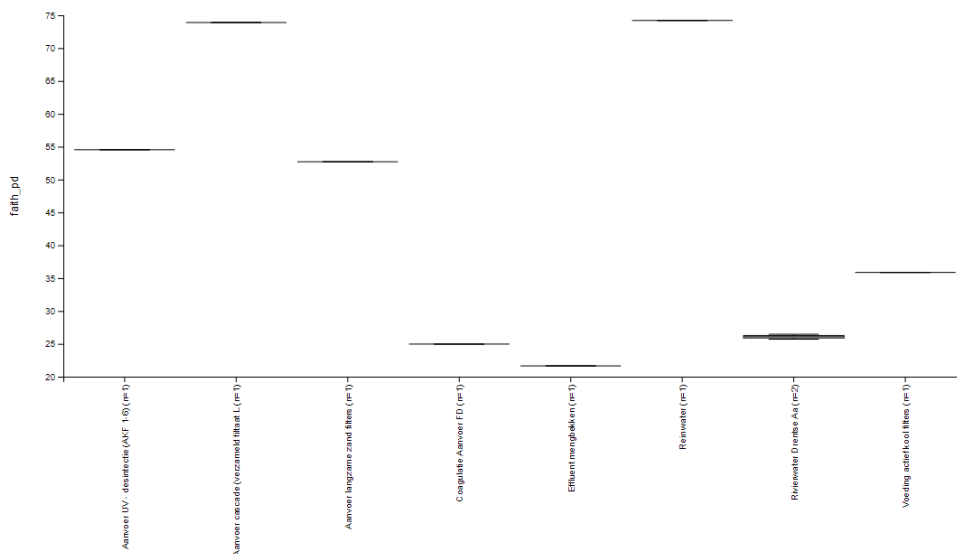
Figuur 39. Monsterlocaties (blauwe cirkels) bij drinkwaterzuivering “De Punt”

De rarefactionanalyses (Figuur 40) laat ook voor deze monsters zien dat het aantal gegenereerde reads een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in de monsters.



Figuur 40. Rarefaction curves van de monsters van “De Punt” met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

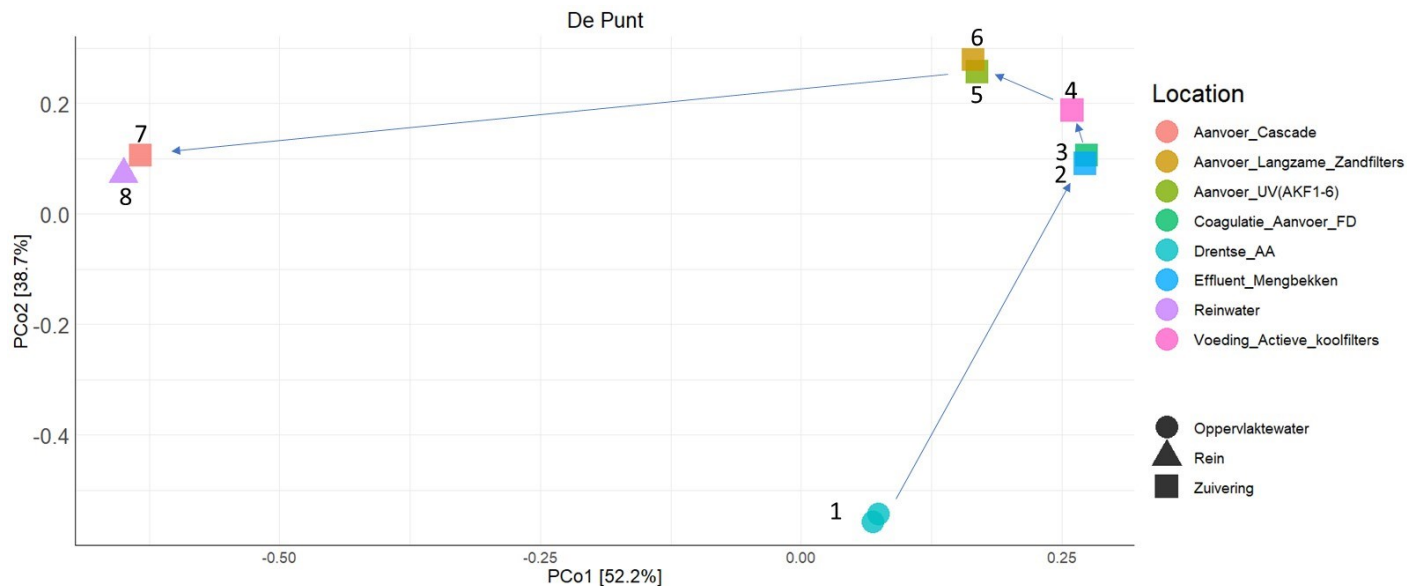
De alpha-diversiteit van de microbiële samenstelling in de monsters (Figuur 41) laat ook voor deze zuivering zien dat de diversiteit in het ruwe oppervlaktewater en het water aan het begin van de zuivering het kleinst is en vervolgens toeneemt tot de grootste diversiteit in het reine water.



Figuur 41. Alpha diversiteit, uitgedrukt in Faith's PD index, van de microbiële populaties in monsters van drinkwaterzuivering van “De Punt”

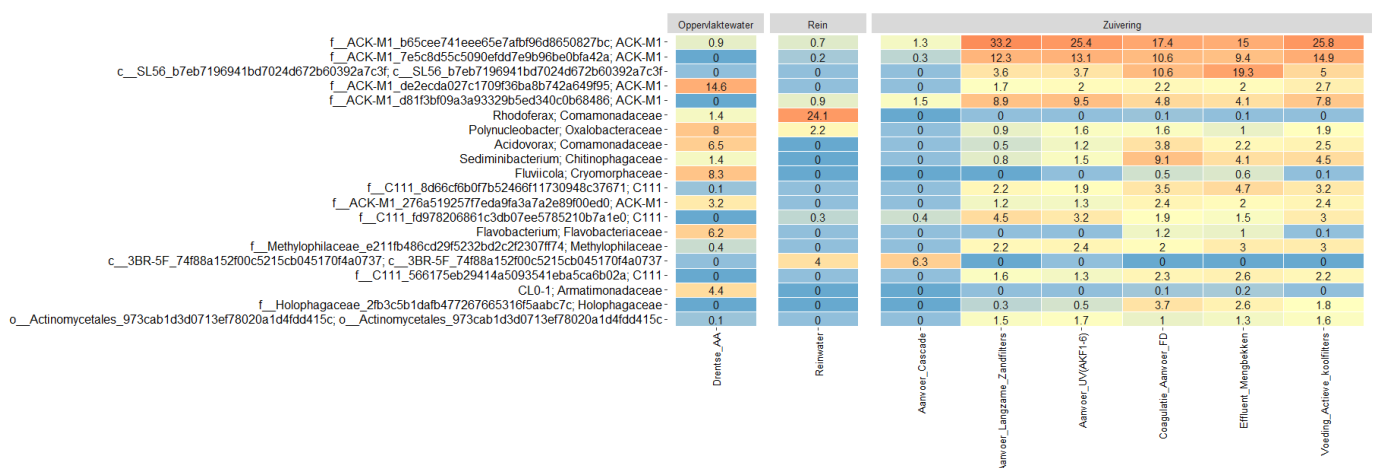
De PCoA plot (Figuur 42) laat zien dat er weinig variatie is tussen duplomonsters (van oppervlaktewater uit de Drentse Aa). De microbiële populaties in deze monsters laten ook voor deze zuivering zien dat de microbiologie in

het ruwe oppervlaktewater van de Drentse Aa sterk afwijkt van de microbiologie in het water na verblijf in het mengbekken. De relatieve veranderingen die vervolgens in zuivering, tijdens coagulatie/sedimentatie, dubbellaags filtratie en UV desinfectie plaatsvinden zijn beperkt. Aansluitend vindt er een grote verandering plaats door het langzame zandfiltratieproces. Waarschijnlijk is dit het gevolg van de UV desinfectiestap waarbij een groot deel van de aanwezige populatie wordt geïnactiveerd en wordt vervangen door een populatie die zich vermeerderd in het langzame zandfilter.



Figuur 42. De beta diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van De Punt.

Een meer getailleerd beeld van de samenstelling van de microbiële populaties in de onderzochte monsters van “De Punt” is weergegeven in een heatmap (Figuur 43)



Figuur 43. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering “De Punt”. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

De heatmap bevestigt dat in de zuivering grote veranderingen plaatsvinden tussen de aanvoer van het langzame zandfilter en de aanvoer van de cascadebeluchting. Dit is waarschijnlijk het gevolg van de UV desinfectiestap. Deze

stap zal waarschijnlijk zorgen voor afdoden van de aanwezige populatie, deze UV behandelde populatie zal “op DNA niveau” nog wel zichtbaar gemaakt worden in het water wat aangevoerd wordt in het langzame zandfilter maar zal zich daar niet verder ontwikkelen maar worden vervangen door een “langzame-zandfilterpopulatie” die veel lijkt op de populatie die ook in het reine water wordt waargenomen. De populatie in de zuivering wordt vooral gedomineerd door een aantal onbekende bacteriële taxa. In het reine water zijn *Rhodofera* bacteriën dominant terwijl deze nog niet zichtbaar zijn in de aanvoer van de cascadebeluchting.

7.2.4 Oppervlaktewaterzuivering Leiduin (Waternet)

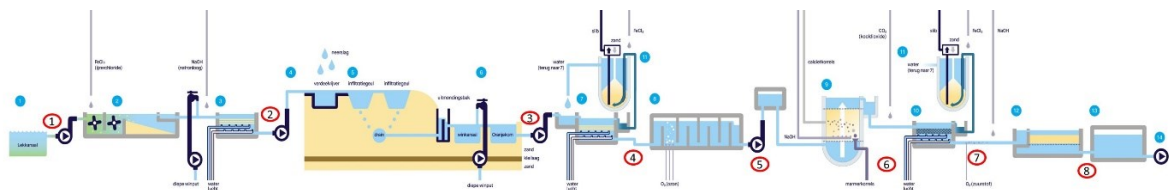
Bij de waterzuivering Leiduin wordt er drinkwater geproduceerd uit oppervlaktewater. Het water wordt ingenomen vanuit het lekkanaal in Nieuwegein waar ook een voorzuivering plaatsvindt bij de WRK. Vervolgens wordt het water naar de Amsterdamse Waterleidingduinen bij Vogelenzang getransporteerd. Na duinpassage vindt een nazuivering op de vestiging Leiduin plaats. In Tabel 12 en Figuur 44 zijn de locaties aangegeven waarvan water is onderzocht. Van elke monsterlocaties zijn monsters in duplo geanalyseerd.

Tabel 12. Geanalyseerde monsters van productiebedrijf Leiduin van Waternet

Monsternr.	Datum	Locatiecode	Locatie	Nr. in figuur*	Aantal reads**
84792	15-10-2019	PNG-OW-02	Inname_Lek	1	175703
84793	15-10-2019	PNG-OW-02	Inname_Lek	1	171352
84796	15-10-2019	PNG-VSF-001	Na_Snelfilters_WRK-I/II	2	153100
84797	15-10-2019	PNG-VSF-001	Na_Snelfilters_WRK-I/II	2	125111
84810	15-10-2019	PLD-SF-INF002	Oranjekom	3	185098
84811	15-10-2019	PLD-SF-INF002	Oranjekom	3	172333
84798	15-10-2019	PLD-SF-EFF001	Na_SnelleZandfilters	4	144062
84799	15-10-2019	PLD-SF-EFF001	Na_SnelleZandfilters	4	155272
84804	15-10-2019	PLD-OH-INF001	Na_Ozon_Leiduin	5	167036
84805	15-10-2019	PLD-OH-INF001	Na_Ozon_Leiduin	5	188246
84812	15-10-2019	PLD-OH-EFF016	Effluent_Ontharding	6	66695
84813	15-10-2019	PLD-OH-EFF016	Effluent_Ontharding	6	172407
84808	15-10-2019	PLD-KF-EFF001	Koolfiltraat_Zuidzijde	7	164596
84809	15-10-2019	PLD-KF-EFF001	Koolfiltraat_Zuidzijde	7	168861
84786	15-10-2019	PLD-LZF-17	Verzameld_LZF_Leiduin-2	8	207338
84787	15-10-2019	PLD-LZF-17	Verzameld_LZF_Leiduin-2	8	181887

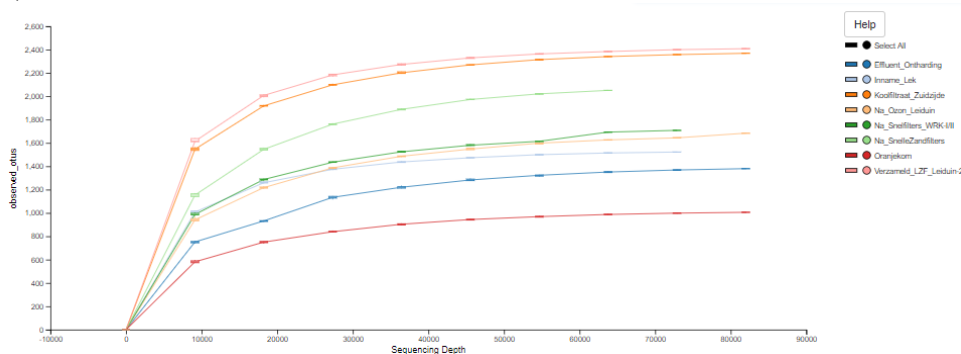
* Locaties zijn in **Error! Reference source not found.** weergegeven in rode cirkels

** Het totaal aantal kwalitatief goede reads (na het filteren van de reads met “Deblur”)



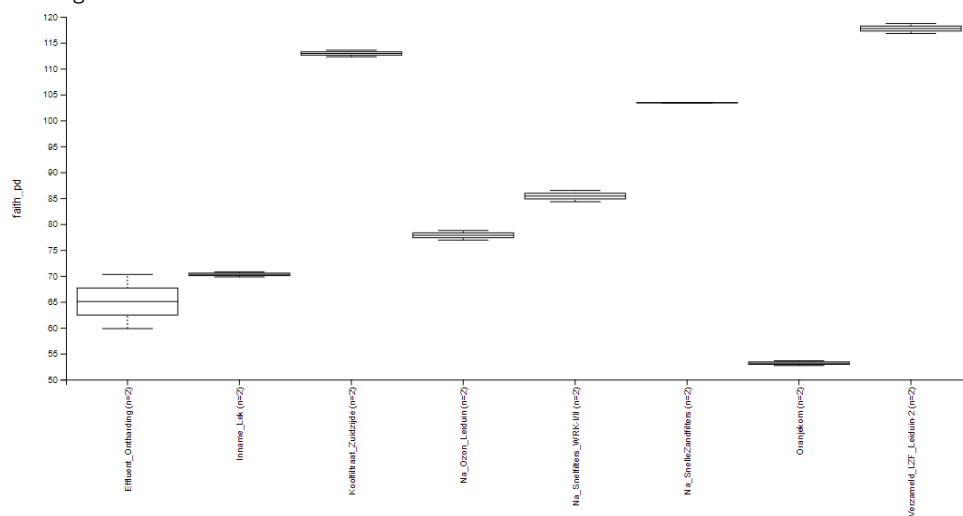
Figuur 44. Monsterlocaties (rood omcirkeld) bij drinkwaterzuivering “Leiduin”

De rarefaction analyses (Figuur 45) laat ook voor deze monsters zien dat het aantal gegenereerde reads een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in de monsters.



Figuur 45. Rarefaction curves van de monsters van productiebedrijf Leiduin van Waternet met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

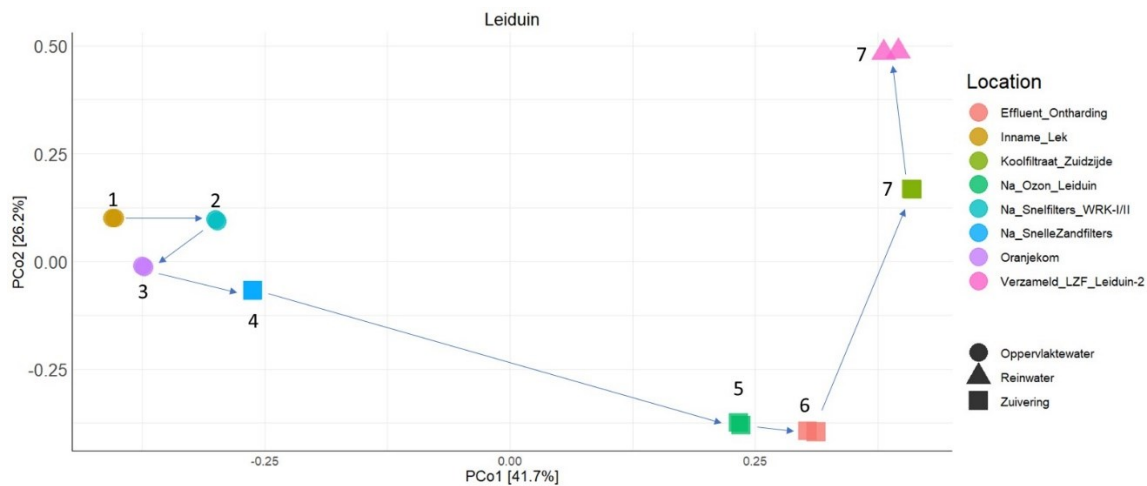
De alpha-diversiteit van de microbiële samenstelling in de monsters (Figuur 46) laat ook voor deze zuivering zien dat de diversiteit in het ruwe oppervlaktewater en het water aan het begin van de zuivering het kleinst is. Ten opzichte van het ingenomen water uit het lekkanaal neemt de diversiteit door het verblijf in het bekken van Oranjekom verder af. De diversiteit neemt vervolgens in de zuivering toe tot de grootste diversiteit in het water na de langzame zandfilters.



Figuur 46. Alpha diversiteit, uitgedrukt in Faith's PD index, van de microbiële populaties in monsters van drinkwaterzuivering van "Leiduin"

In de PCoA plot (Figuur 47) met de vergelijking van monsters van Leiduin is ook voor deze monsters duidelijk dat er zeer weinig variatie is tussen geanalyseerde duplomonsters (merk opnieuw op dat elke kleur in de plot is samengesteld uit symbolen). De microbiële populaties in deze monsters laten zien dat het effect van de voorzuivering van de WRK op de microbiologie in het ingenomen water relatief klein is. Vervolgens is ook het effect van transport naar en passage door de duinen relatief klein. De ozonbehandeling op het gepasseerde duinwater zorgt voor een grote verschuiving in de microbiële populatie waarna ook de relatieve bijdrage van de koolfiltratie (mogelijk i.c.m. ontharding) op de microbiële samenstelling groot is. Mogelijk wordt een groot deel van de aanwezige microbiële populatie door desinfectie met ozon vervangen door een populatie die zich in de stappen na de ozonbehandeling ontwikkelt. Waarschijnlijk is het effect van deze desinfectiestap slecht zichtbaar op DNA-niveau waardoor de verschillen tussen monsters "Effluent ontharding" en "Na Ozon Leiduin" maar weinig van elkaar verschillen, het effect van ozon wordt, door vermeerdering van andere soorten, pas zichtbaar in de volgende

stappen. Uiteindelijk zorgt het langzame-zandfilter aan het einde van de zuivering van Leiduin nog voor een flinke verandering in de microbiële populatie.



Figuur 47. De beta-diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van Leiduin.

De heatmap met taxonomische informatie van de dominante micro-organismen tot op familie- en genusniveau (Figuur 48) laat zien dat in het ruwe oppervlaktewater en het water na snelfilters verschillende bacteriesoorten uit de familie van *Comamonadaceae* dominant zijn. Na de desinfectiestap met ozon worden de genera *Sphaerotilus* en *Polaromonas* (familie *Comamonadaceae*) en een genus uit de familie van *Isosphaeraceae* dominant.

	Oppervlaktewater			Reinwater	Zuivering			
f__ACK-M1_b65cee741eee65e7afbf96d8650827bc; ACK-M1-	11.6	21	1.9	0	0	0	0	5.1
Polaromonas; Comamonadaceae	0.5	0.6	0	0	4.6	0.4	23.3	0
Acidovorax; Comamonadaceae	3.5	4.4	6.8	0	0	0	0	8.5
Flavobacterium; Flavobacteriaceae	1.2	0.5	8.9	0	7.8	0	0.4	3.2
Polynucleobacter; Oxalobacteraceae	2.7	4.1	3.5	0	0	0	0	7.7
Hydrogenophaga; Comamonadaceae	0.7	1	0.1	0	11.7	0.5	1	0
f__Cytophagaceae_289caaa4d4c1eecefd2c8185d8c0c; Cytophagaceae-	1.1	0.9	6.6	0	0	0	0	4.2
f__ACK-M1_276a519257f7eda9fa3a7a2e89f00ed0; ACK-M1-	2.3	1.9	3.5	0	0	0	0	4.9
Fluviicola; Cryomorphaeae	1.1	0.3	7.3	0	0.4	0	0	3.1
f__Methylophilaceae_e211fb486cd29f5232bd2c2f2307f74; Methylophilaceae-	3.8	5	0.9	0	0	0	0	1.9
Planctomyces; Planctomycetaceae	1.3	3.1	0.1	0.7	1	1.7	2.4	1
f__Isosphaeraceae_297c001b71204aa86f577f8f67e77a34; Isosphaeraceae-	0.1	0.1	0.6	0	2.2	0	6.6	1
Sphaerotilus; Comamonadaceae	0	0	0	0	0.5	0	9.6	0
o__Stramenopiles_9dd8611dcfec4a8a9c3461bf6401086; o__Stramenopiles_9dd8611dcfec4a8a9c3461bf6401086-	0	0	9.6	0	0.1	0	0.3	0
Sedimibacterium; Chitinophagaceae	2.1	1.6	3.4	0	0	0	0	2.7
f__Comamonadaceae_785682628a0269c3c999e0e5a3a9c9a5; Comamonadaceae-	1	0.9	3.2	0	0	0	0	3
Nitrosopumilus; Cenarchaeaceae	0	0.1	0	3.4	0	4	0	0
o__Cryptophyta_8043b2b8804509d26c91f97681803d8f; o__Cryptophyta_8043b2b8804509d26c91f97681803d8f-	0.1	0.1	1	0	0.4	0	4.2	1.5
Planctophila; ACK-M1-	2.6	2.8	0.6	0	0	0	0	1
o__Actinomycetales_973cab1d3d0713ef78020a1d4fd415c; o__Actinomycetales_973cab1d3d0713ef78020a1d4fd415c-	2.2	3.4	0.2	0	0	0	0	0.5

Figuur 48. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering Leiduin. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

7.2.5 Drinkwaterzuivering Scheveningen (Dunea)

Van de drinkwaterzuivering van Scheveningen zijn er monsters geanalyseerd van het oppervlaktewater wat wordt geïnfiltrerd in de duinen, het onttrokken water na duinpassage, het water na de snelfilters, het afgeleverde drinkwater en het water uit de distributie (Tabel 13).

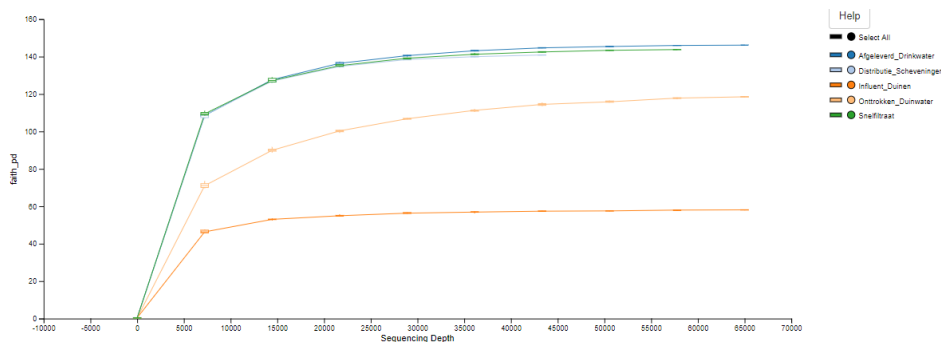
Tabel 13. Geanalyseerde monsters van productiebedrijf Scheveningen van Dunea

Monsternr.	Datum	Locatiecode	Locatie	Aantal reads*
84949	22-10-2019	DRZ_ADH_DN_VIJZEL	Distributie Scheveningen	223276

84958	22-10-2019	PSC_DW	Afgeleverd drinkwater	208410
84959	22-10-2019	WME_INF	Influent duinen	189272
84960	22-10-2019	PMO_VOW	Onttrokken duinwater	181542
84961	22-10-2019	PMO_SF_VEFF	Snelfiltraat	140090

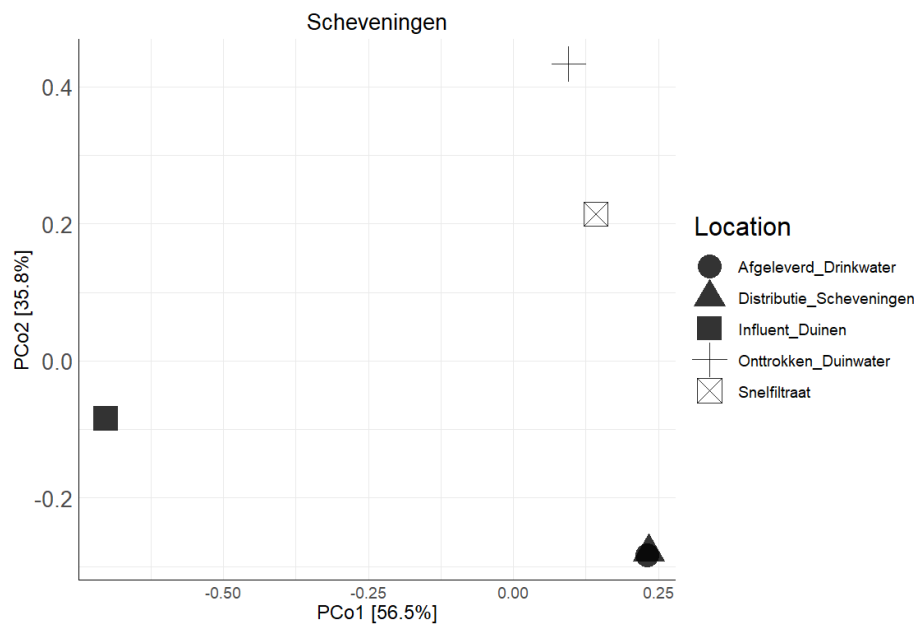
* Het totaal aantal kwalitatief goede reads (na het filteren van de reads met "Deblur")

De rarefaction analyses (Figuur 49) laat ook voor deze monsters zien dat het aantal gegenereerde reads een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in de monsters. Door het ontbreken van meerdere analyses van elke locatie is het voor Qiime2 niet mogelijk om de alpha-diversiteit weer te geven in boxplots. De rarefaction analyses maken wel duidelijk dat de microbiële diversiteit in het influent van de duinen het kleinst is, duinpassage lijkt te zorgen voor een toename van de diversiteit waarna de microbiële diversiteit in de nazuivering resulterend in de hoogste diversiteit in het afgeleverde en gedistribueerde drinkwater.



Figuur 49. Rarefaction curves van de monsters van drinkwaterzuivering Scheveningen van Dunea met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

De PCoA plot (Figuur 50) met de vergelijking van monsters van Scheveningen laat zien dat er grote verschillen zijn tussen de microbiologie in het water dat wordt geïnfiltreerd in de duinen en het water na duinpassage. De duinpassage heeft dus veel invloed op de microbiologie. De processen in het traject tussen het onttrekken van duinwater en snelfiltratie (ontharding, kooldosering, cascadebeluchting) zorgen ook voor een duidelijke microbiologische verandering waarna ook de langzame zandfilters een bijdrage leveren in de uiteindelijke samenstelling van de microbiologie in het afgeleverde water. Het water op de locatie waarop het monster uit de distributie is genomen bezit een microbiële populatie die vergelijkbaar is met het afgeleverde water.



Figuur 50. De beta diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van Scheveningen.

De heatmap uit Figuur 51 bevestigt dat de microbiologie in het water dat de duinen ingaat veel verschilt met de microbiologie in het water dat uit de duinen wordt onttrokken. In het water na de sneifilters, wat gedeeltelijke nazuivering heeft ondergaan (na onharding/kooldosering/cascadebeluchting en sneifiltratie), bestaat het dominante deel van de aanwezige microbiologie al voor een groot deel uit de taxa die ook aanwezig zijn in het afgeleverde water. In het influent van de duinen is een onbekende bacteriegenus zeer dominant, deze bacteriegenus is na duinpassage niet meer detecteerbaar. Na duinpassage zijn vooral *Methylococcaceae* en *Crenothrix* zeer dominant, deze bacteriën verdwijnen vervolgens in de nazuivering en worden vrijwel niet aangetroffen in het drinkwater. De samenstelling van de microbiologie in het water uit de distributie is zeer vergelijkbaar met de microbiologie in het afgeleverde water.

	Drinkwater_Distributie	Drinkwater_Pomptstation	Na-Duinen	Sneifiltraat	Voor_Duinen
f__Methylococcaceae_5c0d8abd40e05d5d426d281f4e3ac9d4; Methylococcaceae	0	0	21.4	1.7	0
Crenothrix; Crenothricaceae	0.1	0.1	16.9	1.7	0.1
f__ACK-M1_b65cee741eee65e7afb96d8650827bc; ACK-M1-	0	0	0	0	12.9
c__PBS-25_fbc91355f092201c9dd1f6e747cda9af; c__PBS-25_fbc91355f092201c9dd1f6e747cda9af	1.5	1.4	1	2.3	0
Fluviicola; Cryomorphaceae	0	0	0.3	1.2	4.7
f__Methylococcaceae_395991fcbcbda4726b3206edae56b3f; Methylococcaceae	0	0	5.5	0.5	0
Acidovorax; Comamonadaceae	0.1	0.7	0	0	5.1
f__Methylococcaceae_96f8db4b1df857a4b182fac2b352881; Methylococcaceae	0	0	5.2	0.5	0
Thiothrix; Thiotrichaceae	0	0	4.9	0.2	0
Planctomyces; Planctomycetaceae	0.2	0.3	0	0.4	4.1
Polynucleobacter; Oxalobacteraceae	0	0	0	0	4.4
Sulfuricum; Helicobacteraceae	0	0	2.1	1.7	0
f__Methylophilaceae_e211fb486cd29f5232bd2c2f2307f174; Methylophilaceae	0	0	0	0	3.5
cetales_973cab1d3d0713ef78020a1d4fd415c; o__Actinomycetales_973cab1d3d0713ef78020a1d4fd415c	0	0	0	0	3.2
Candidatus Methyloirabialis; Methyloirabiaceae	0.2	0.1	0.8	1.8	0
f__Helicobacteraceae_95ac04a85ab6f8fa049858244e4d2cd4; Helicobacteraceae	0	0	2.7	0.2	0
f__0319-6A21_db15b0f02d94dfb6d39f702ced2f4c4c5; 0319-6A21-	0.9	0.7	0.3	0.8	0
Planctophila; ACK-M1-	0	0	0	0	2.7
f__0319-6A21_46d08fe0abec27dc85101fe1c5f07350; 0319-6A21-	0.9	0.9	0.2	0.7	0
teriales_8f9ef5f80a3c08fa7bd7d3b2b9b8948; o__Sphingobacteriales_8f9ef5f80a3c08fa7bd7d3b2b9b8948	0	0	0	0	2.6

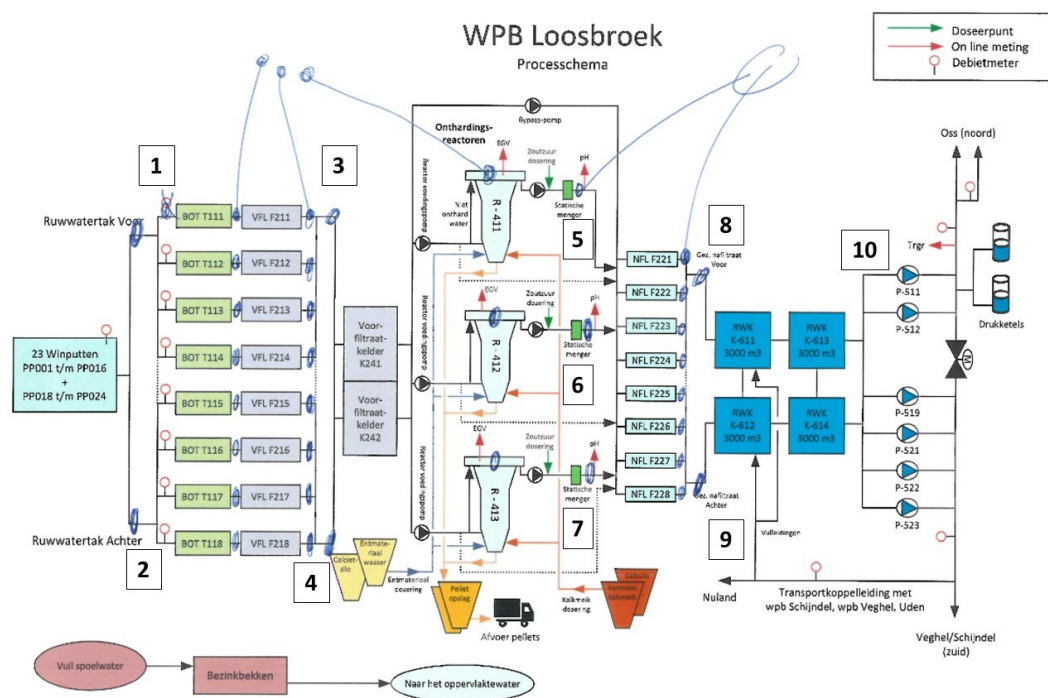
Figuur 51. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering Scheveningen. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

7.2.6 Drinkwaterzuivering Loosbroek (Brabant Water)

Bij drinkwaterzuivering Loosbroek van Brabant Water zijn er monsters, afkomstig van verschillende stadia van de zuivering, geanalyseerd. In Tabel 14 is weergegeven op welke locaties in de zuivering monsters zijn genomen en in Figuur 52 zijn deze locaties in het schema van de zuivering weergegeven..

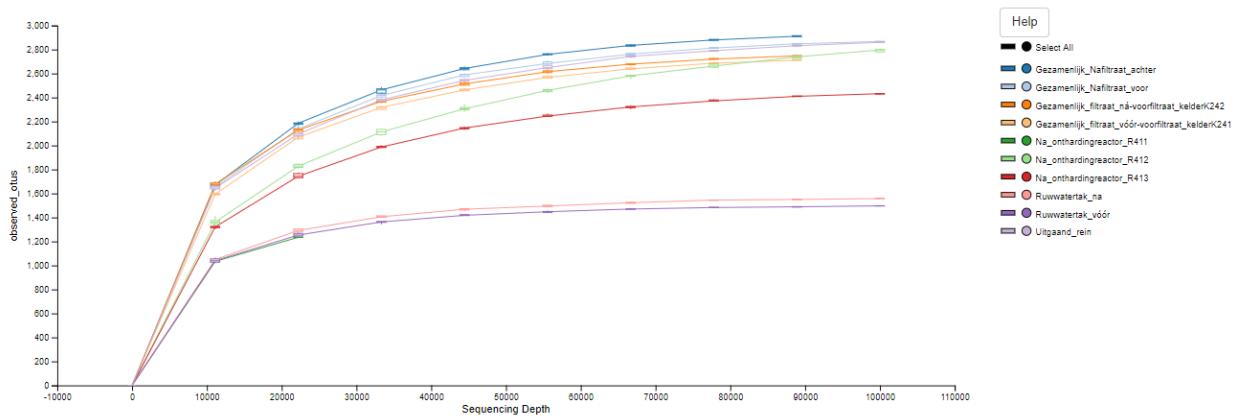
Tabel 14. Geanalyseerde monsters van productiebedrijf Loosbroek van Brabant Water

Monsternr.	Nr. (in Figuur 52)	Datum	Exacte locatie	Aantal reads*
85548	1	06-11-2019	Ruwwatertak_vóór	189404
85549	2	06-11-2019	Ruwwatertak_na	221277
85550	3	06-11-2019	Gezamenlijk voorfiltraat vóór	171797
85551	4	06-11-2019	Gezamenlijk voorfiltraat ná	180059
85552	5	06-11-2019	Na_onthardingreactor_R411	53293
85553	6	06-11-2019	Na_onthardingreactor_R412	340293
85554	7	06-11-2019	Na_onthardingreactor_R413	192309
85555	8	06-11-2019	Gezamenlijk_Nafiltraat_voor	188963
85556	9	06-11-2019	Gezamenlijk_Nafiltraat_achter	179085
85557	10	06-11-2019	Uitgaand_rein	204567



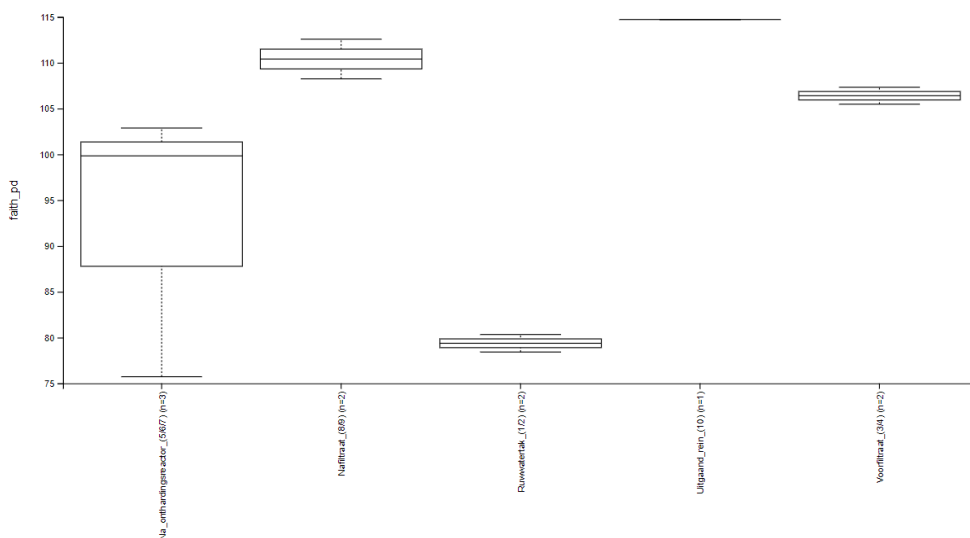
Figuur 52. Monsterlocaties (in genummerde vierkantjes) bij drinkwaterzuivering “Loosbroek”

De rarefactionanalyses (Figuur 53) laten zien dat het aantal gegenereerde reads voor elk monster een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in deze monsters.



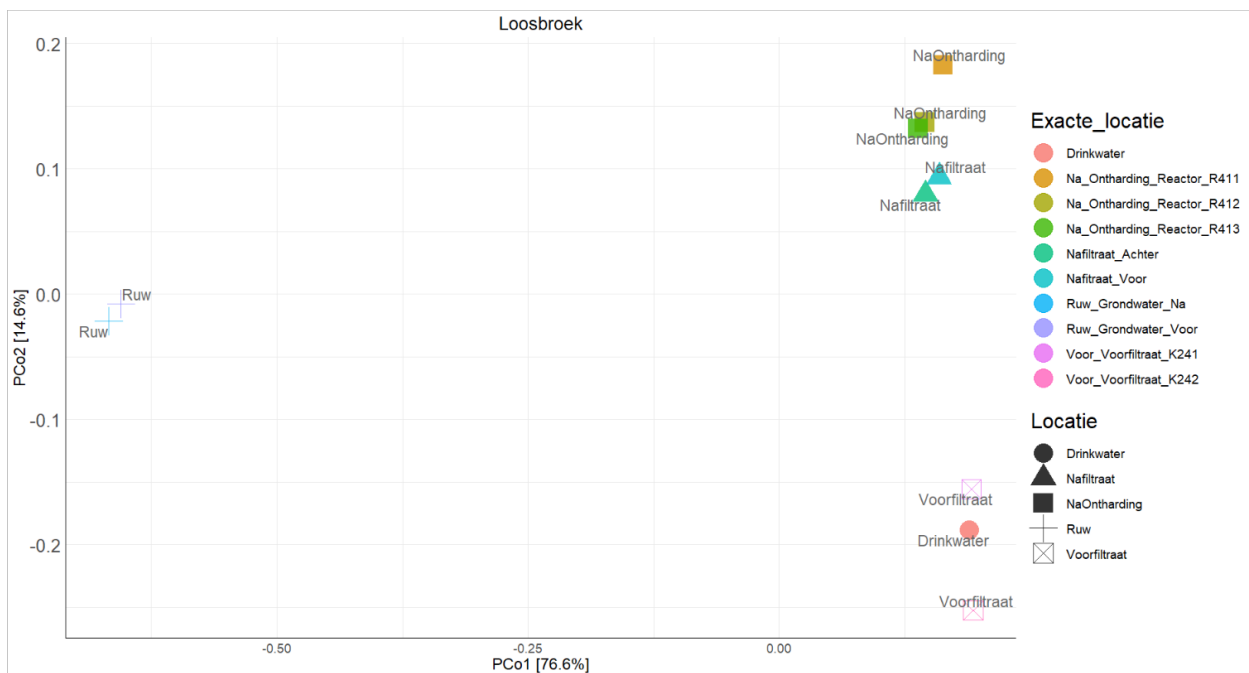
Figuur 53. Rarefaction curves van de monsters van drinkwaterzuivering Loosbroek van Brabant Water met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

De alpha-diversiteit van de microbiologie in de monsters (Figuur 54) laat zien dat de diversiteit in het ruwe water relatief laag is, de diversiteit neemt vervolgens toe in het voorfiltraat en daarna weer af in de onthardingsreactor. Na een toename van de diversiteit na de nafiltraatie is de diversiteit het hoogst in het uitgaande reine water. Er zijn vrij grote verschillen in de diversiteit tussen de bemonsterde onthardingsreactoren, vooral in reactor R412 lijkt de microbiële diversiteit groter. Echter vanwege éénmalige monsternamen kan niet worden gezegd of dit verschil ook statistisch significant is, uitgebreider onderzoek is nodig om beter inzicht te krijgen in eventuele verschillen tussen de reactoren.



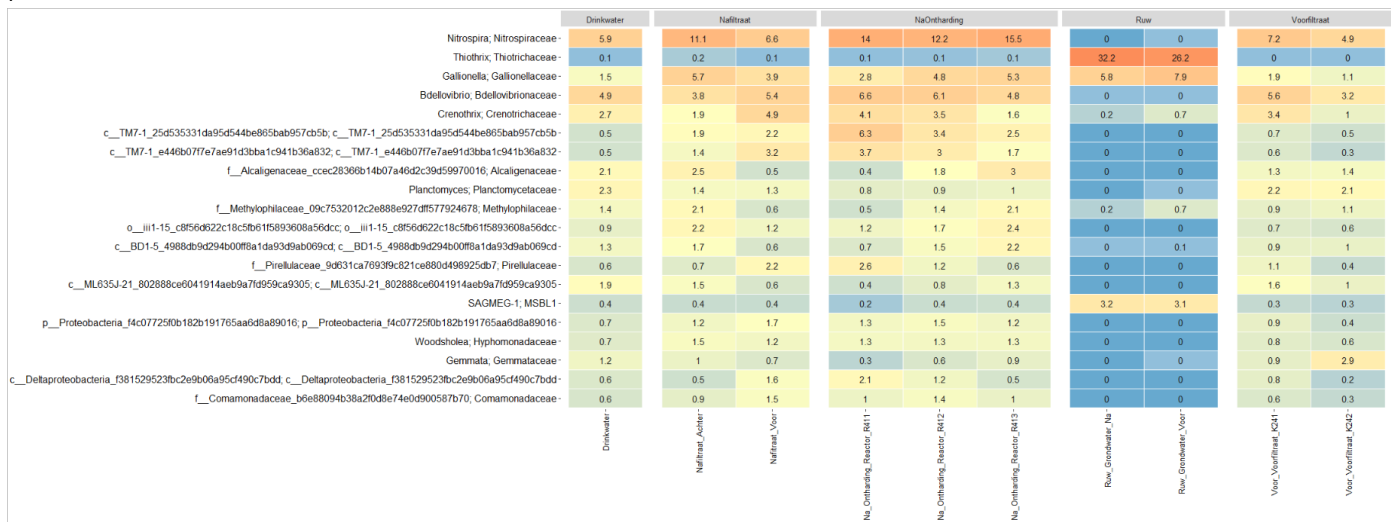
Figuur 54. Alpha diversiteit, uitgedrukt in Faith's PD index, van de microbiële populaties in monsters van drinkwaterzuivering van "Loosbroek"

De PCoA plot (Figuur 55) waarmee de samenstelling van de aanwezige microbiële populaties wordt vergeleken (beta-diversiteit) laat zien dat: de microbiologie in beide ruwwatertakken zeer vergelijkbaar is, daarna treedt er een grote verandering op waardoor de microbiologie in het voorfiltraat sterk afwijkt van de microbiologie in het ruwe water. De clustering van de monsters van het voorfiltraat dichtbij het drinkwatermonster en de monsters van het nafiltraat op grotere afstand van het drinkwater lijkt onlogisch. De verwachting is immers dat de microbiologie in het water na de zuiveringsstappen steeds meer op het drinkwater zal gaan lijken. Momenteel is onduidelijk of de situatie inderdaad is zoals dit experiment laat zien of dat er hier sprake is van verwisseling van nafiltraatmonsters met voorfiltraatmonsters.



Figuur 55. De beta-diversiteit (figuur met AmpVis2 in RStudio) op basis van Bray Curtis index, weergegeven in een principal coordinates analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering van drinkwaterzuivering Loosbroek.

De heatmap uit Figuur 56 bevestigen dat de grootste microbiologische veranderingen optreden tussen het ruwe water en het voorfiltraat, de microbiologische veranderingen in de stappen na de voorfiltratie zijn beperkt. In alle monsters zijn vooral bacteriën uit het phylum van proteobacteriën dominant. Het *Acidobacteria* phylum wordt vrijwel niet aangetroffen in het ruwe grondwater, vermeerderen waarschijnlijk tijdens de voorfiltratie en blijven vervolgens stabiel tijdens de volgende zuiveringsstappen. De heatmap (Figuur 56) laat ook zien dat, op genus niveau, de genera *Thiothrix* en *Gallionella* dominant zijn in het ruwe grondwater. *Thiothrix* behoort in zuivering niet tot de dominante soorten, van *Thiothrix* is bekend dat deze bacteriesoort betrokken kan zijn bij problemen met biofouling en de oxidatie van sulfide (Brigmon, Martin et al. 1997). *Gallionella* is een bodembacterie die vaak in grondwater wordt aangetoond en waarvan bekend is dat deze betrokken kan zijn bij oxidatie van ijzer (de Vet, Dinkla et al. 2011), *Gallionella* is dominant in het ruwe grondwater en blijft aanwezig in de zuivering. In de zuivering en in het afgeleverde drinkwater zijn vooral *Nitrospira* bacteriën dominant, die waarschijnlijk betrokken zijn bij de omzetting van ammonium.



Figuur 56. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering Loosbroek. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

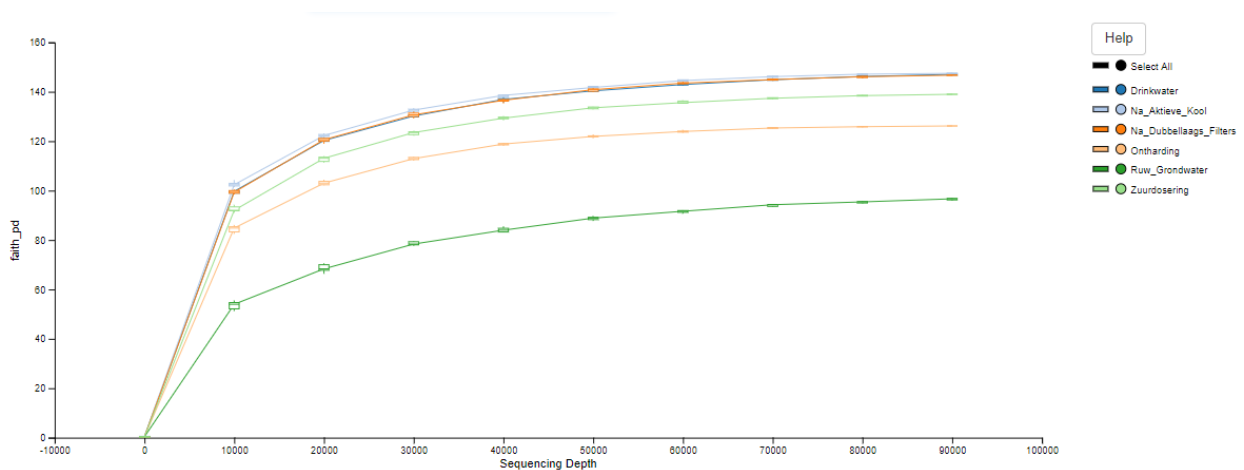
7.2.7 Drinkwaterzuivering Zele (De Watergroep)

Van drinkwaterzuivering Zele zijn er eenmalig monsters geanalyseerd welke afkomstig zijn van verschillende stappen in de zuivering. In Tabel 15 is een overzicht weergegeven van de geanalyseerde monsters en het aantal gegenereerde sequence reads per monster.

Tabel 15. Geanalyseerde monsters van productiebedrijf Zele van de Watergroep

Monsternr.	Datum	Locatiecode	Locatie	Aantal reads*
85512	05-11-2019	Ruw grondwater	Ruw_Grondwater	228508
85513	05-11-2019	Zuurdosering	Zuurdosering	211597
85514	05-11-2019	Ontharding	Ontharding	199071
85515	05-11-2019	DLF Totaal	Na_Dubbellaags_Filters	210836
85516	05-11-2019	AKF Totaal	Na_Aktieve_Kool	202255
85517	05-11-2019	Rein water	Drinkwater	231539

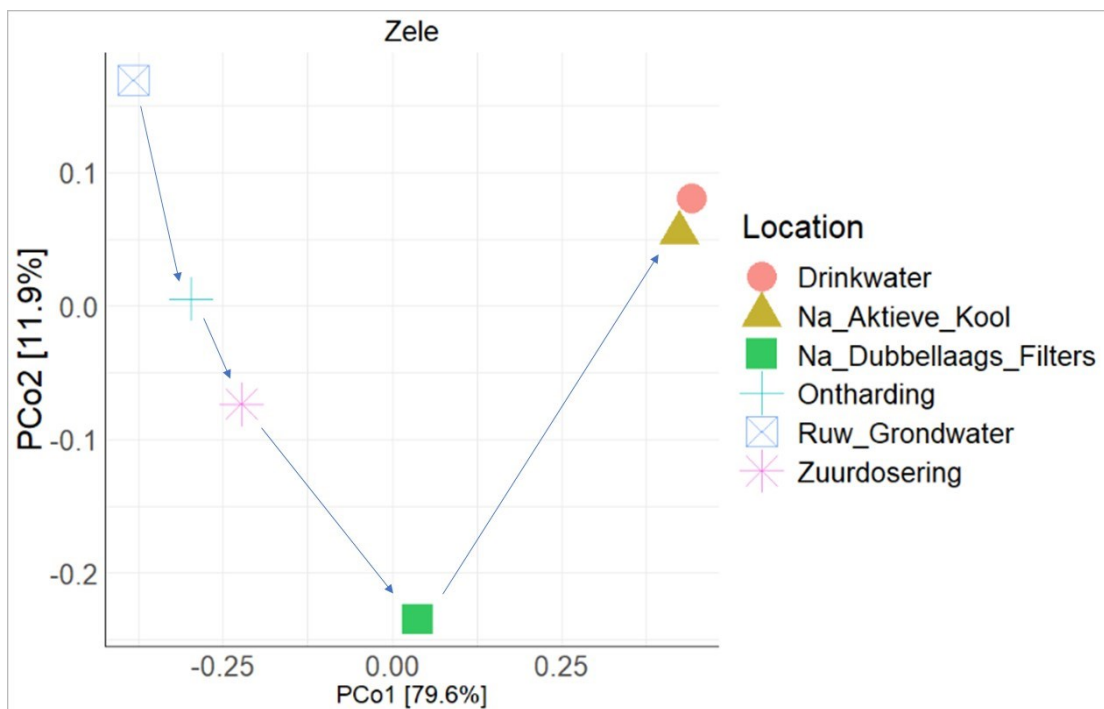
Ook voor de monsters van deze zuivering laten de rarefactionanalyses (Figuur 57) zien dat het aantal gegenereerde reads voor elk monster een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in deze monsters.



Figuur 57. Rarefaction curves van de monsters van drinkwaterzuivering Zele van De Watergroep met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

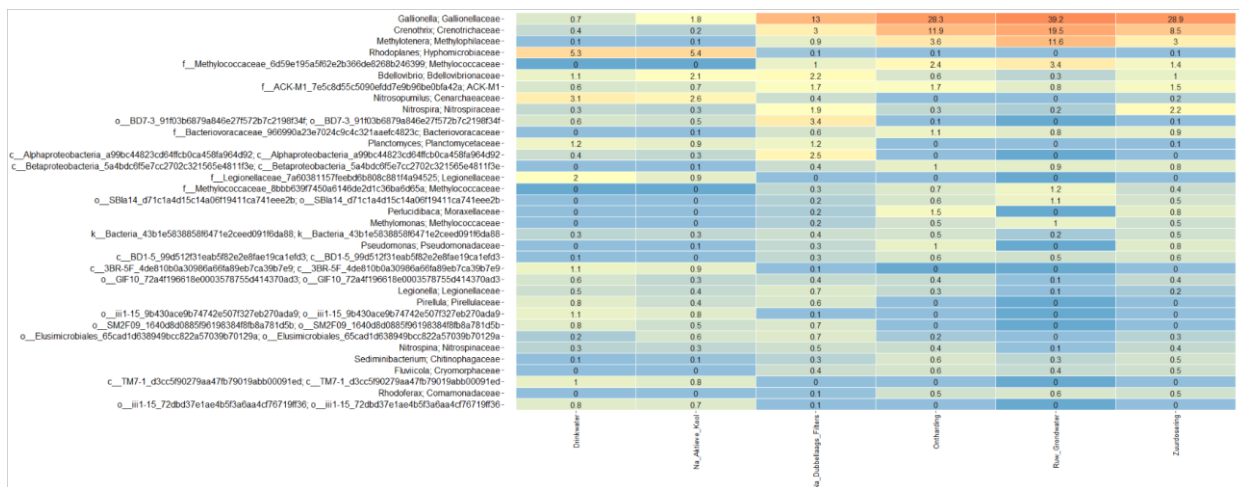
Door het ontbreken van meerdere analyses van elke locatie is het voor Qiime2 ook voor deze monsterset niet mogelijk om de alpha-diversiteit weer te geven in boxplots. De rarefactionanalyses maken wel duidelijk dat de microbiële diversiteit in ruwe grondwater het kleinst is waarna de diversiteit in de zuivering toeneemt en het hoogst is in het gedistribueerde drinkwater.

De PCoA plot (Figuur 58), waarmee de samenstelling van de aanwezige microbiële populaties wordt vergeleken (beta-diversiteit), laat zien dat de microbiologie in het ruwe grondwater sterk afwijkt van de microbiologie in de andere monsters. In de zuivering verandert de samenstelling van de microbiologie stapsgewijs waarbij de grootste veranderingen optreden tussen de zuurdosering en na passage door dubbellaagsfilters en tussen de dubbellaagsfilters en na actieve-koolfiltratie. In het traject tussen de actieve-koolfiltratie en het afgeleverde drinkwater blijft de microbiologie in het water relatief stabiel.



Figuur 58. De beta-diversiteit op basis van Bray Curtis index, weergegeven in een principal component analyses (PCoA) met Hellinger transformatie van de microbiële populaties in monsters van drinkwaterzuivering Zele.

De veranderingen welke worden zichtbaar gemaakt in de PCA plot (Figuur 59) worden bevestigd in de heatmap waarin de samenstelling van de microbiële gemeenschap wordt weergegeven. In de heatmap is te zien dat ook bij deze grondwaterwinning *Gallionella* zeer dominant is in het ruwe grondwater en tijdens de eerste stappen van de zuivering. *Gallionella* blijft, net als in de grondwaterzuivering van Loosbroek, aanwezig na de verschillende zuiveringsstappen. Net als *Gallionella* zijn ook de genera *Methylothera* en *Crenothrix* dominant in het ruwe grondwater, de dominantie van deze genera neemt af na actieve-koolfiltratie. De dominantie van de bacteriesoort *Rhodoplanes* en het archaeon *Nitrosopulimus* nemen na actieve-koolfiltratie toe en zijn dominant in het drinkwater. Vergelijkbare analyse zijn uitgevoerd door De Watergroep, een uitgebreide vergelijking van resultaten valt echter buiten de scope van dit onderzoek. Een zeer globale vergelijking van de resultaten laat vergelijkbare trends zien (persoonlijke communicatie, Jerom Vranken). Een uitgebreide vergelijking van data zal enig inzicht geven in de vergelijkbaarheid van data gegenereerd bij verschillende laboratoria.



Figuur 59. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in drinkwaterzuivering Zele. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

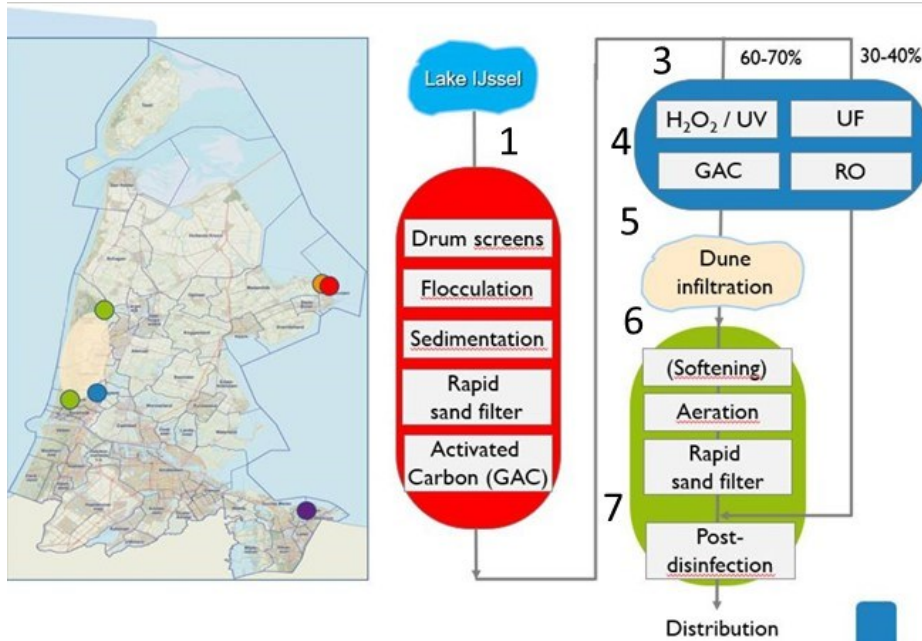
7.2.8 Monsters van PWN

Van PWN zijn er acht monsters geanalyseerd waarvan de locatiecodes en de locatie van herkomst in Tabel 16 en Figuur 60 is weergegeven.

Tabel 16. Geanalyseerde monsters van PWN

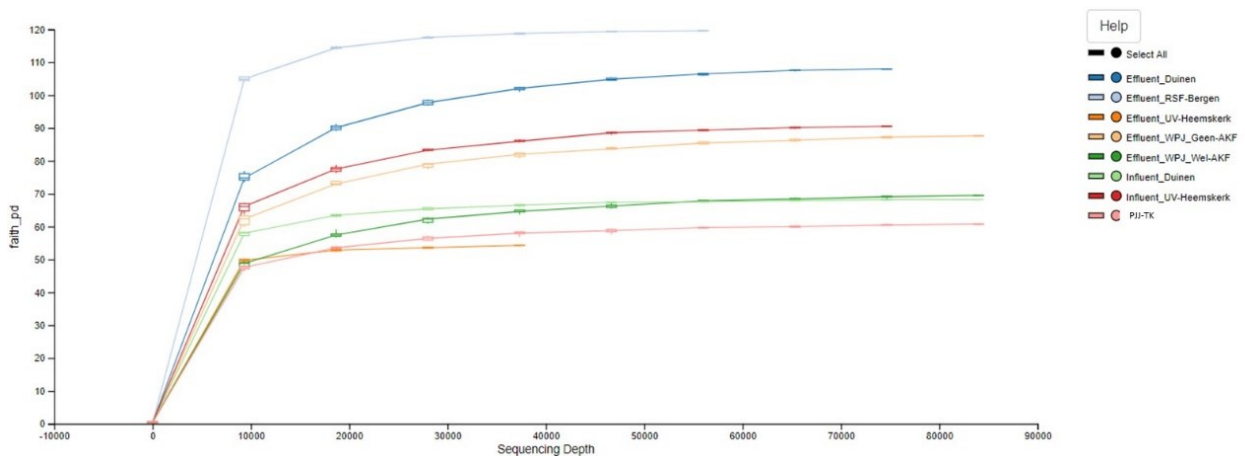
Monsternr.	Nr.	Datum	Locatiecode	Locatie	Aantal reads**
84956	5	22-10-2019	PLG-TL-EFF	Influent_Duinen	146605
84950	6	22-10-2019	PBG-PO-INF	Effluent Duinen	170360
84951	7	22-10-2019	PBG-UV1-INF	Effluent RSF-Bergen	161683
84955	1	22-10-2019	PPJ-TK	Influent WPJ	191282
84952	2a	22-10-2019	PPJ-TL1-INF	Effluent WPJ_Geen-AKF	187425
84954	2b	22-10-2019	PPJ-TL2-INF2	Effluent WPJ_Wel-AKF	159201
84957	3	22-10-2019	PLG-UF-INF	Influent_UV-Heemskerk	156564
84953	4	22-10-2019	PLG-UF-EFF	Effluent_UV-Heemskerk	70374

** Het totaal aantal kwalitatief goede reads (na het filteren van de reads met "Deblur")
Het nummer (Nr.) van de locatie correspondeert met het nr. weergegeven in Figuur 60



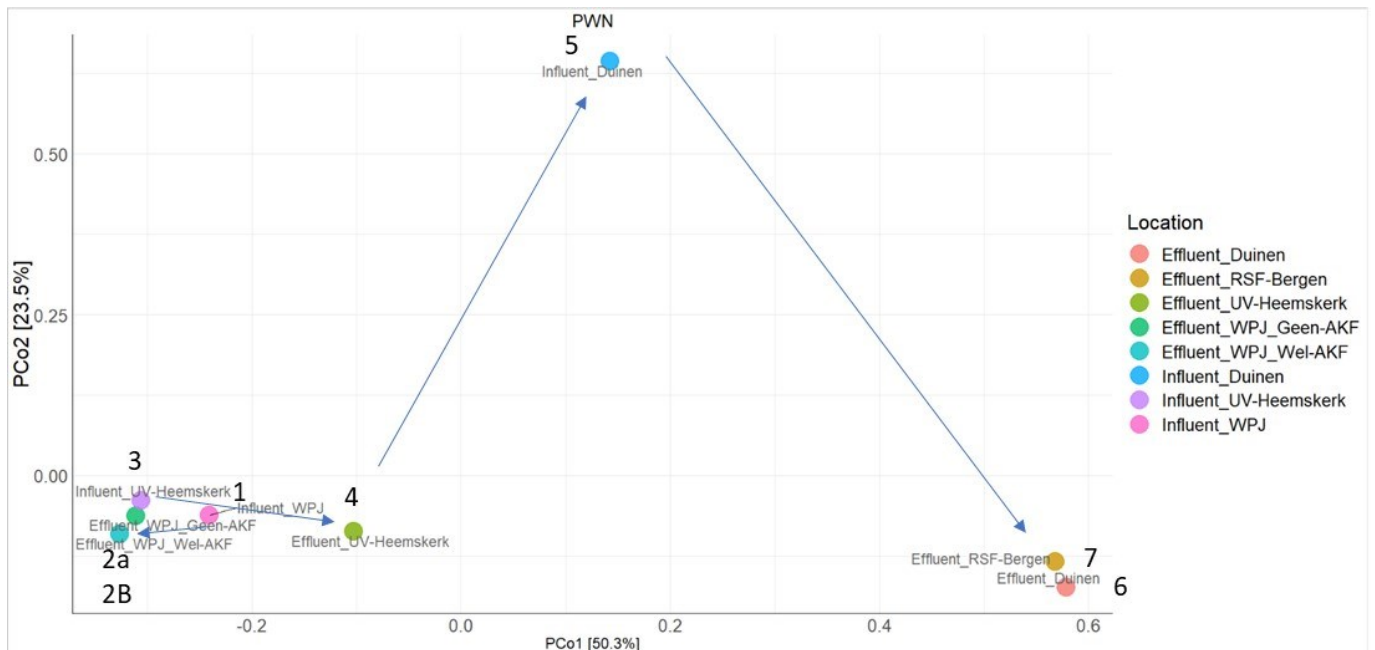
Figuur 60. Schematische weergave van bemonsterde zuiveringen. De WPJ zuivering is weergegeven in rood, Heemskerk in blauw en Bergen in groen

De rarefactionanalyses (Figuur 61) laat zien dat het aantal gegenereerde reads een overmaat vormen voor het verkrijgen van een vrijwel volledig beeld van de samenstelling van de microbiële populaties in de monsters. Ook voor het monster met relatief weinig sequentiereads (Effluent_UV-Heemskerk) is de sequencediepte voldoende om in het horizontale deel van de rarefactioncurve te komen waardoor aangenomen kan worden dat ook van dit monster een vrijwel volledig beeld van de microbiële samenstelling wordt verkregen. De rarefactioncurves laten ook zien dat de laagste diversiteit aanwezig is in het influent WPJ (PJJ-TK) en de hoogste diversiteit in het effluentmonster RSF-Bergen.



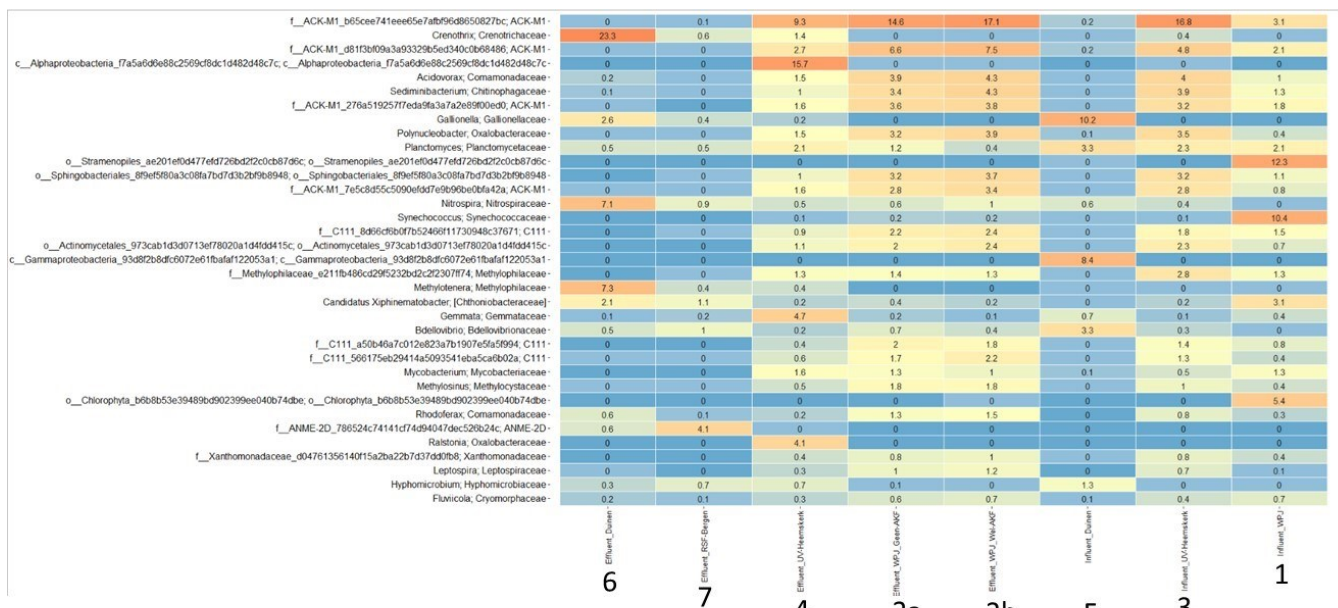
Figuur 61. Rarefaction curves van de monsters van de monsters van PWN met de sequencing diepte (x-as) en het aantal geobserveerde ASVs (y-as) in elk monster

De PCoA plot (Figuur 62) met de vergelijking van de monsters maakt duidelijk dat de veranderingen in van de microbiële populaties in de beginstappen van de zuivering (1-4) gering zijn waarbij ook de invloed van actieve koolfiltratie klein is. Ook de microbiologie in het effluent van de UV in Heemskerk lijkt nog sterk op de microbiologie in de WPJ zuivering. Waarschijnlijk worden in dit monster nog steeds de bacteriën uit WPJ aangetoond maar zijn deze door de UV/H₂O₂ stap afgedood. De microbiologie in het water wat wordt geïnfiltreerd in de duinen (5) is compleet anders, waarschijnlijk is de microbiële van WPJ afgedood door de UV/H₂O₂ en vervangen door een nieuwe populatie in de zuivering van Heemskerk. Aansluitend zorgt duinpassage opnieuw voor grote microbiële veranderingen waarna het effect van de zuiveringsstappen in de zuivering van Bergen beperkt zijn.



Figuur 62. De beta-diversiteit op basis van Bray Curtis index, weergegeven in een principal component analyses (PCoA) met Hellinger transformatie van de microbiële populaties in de monsters van PWN.

De veranderingen welke worden zichtbaar gemaakt in de PCA plot van Figuur 62 worden bevestigd in de heatmap van Figuur 63. waarin de samenstelling van de microbiële gemeenschap op familieniveau wordt weergegeven. Ook in deze monsters worden veel onbekende bacteriën aangetoond en is zichtbaar dat *Crenothrix* zeer dominant is na duinpassage.



Figuur 63. Heatmap met een weergave van de samenstelling van de 20 meest dominante taxa in de microbiële populaties in monsters van PWN. De taxa zijn gerangschikt op familieniveau, het genusniveau van de taxa is ook weergegeven

8 Conclusies

Voor het uitvoeren van microbial profiling analyses dient een groot aantal stappen op het laboratorium te worden uitgevoerd en aansluitend zijn bioinformatica-analyses noodzakelijk om interpretatie van de ruwe sequentiedata mogelijk te maken. Bij zowel de laboratorium- als de bioinformatica-analyses kunnen verschillende keuzes worden gemaakt, waarbij veel van deze keuzes invloed kunnen hebben op het eindresultaat. Dit maakt het onmogelijk om resultaten verkregen met verschillende protocollen op detailniveau te vergelijken en vraagt om uniformering van de aanpak en het toepassen van de juiste controles. In dit onderzoek zijn de verschillende stappen besproken die dienen te worden toegepast om tot betrouwbare microbiële profielen te komen en interpretatie daarvan en wordt het mogelijke gevolg van gemaakte keuzes samengevat. Voor het isoleren van DNA uit watermonsters en het genereren van 16S rRNA-gensequentiedata met de Illumina MiSeq sequencer is een gestandaardiseerd protocol ontwikkeld. Voor de interpretatie van 16S rRNA-gensequentie data is nog geen standaard workflow ontwikkeld, maar zijn de mogelijkheden van het programma Qiime2 onderzocht en gecombineerd met statistische R-scripts voor populatieonderzoek.

Door het grote aantal keuzes dat gemaakt kan worden en het grote effect van deze keuzes op het analyseresultaat is het van belang om de uitvoering van deze analyses zoveel mogelijk te standaardiseren en, bij het combineren van analyseresultaten van meerdere laboratoria, de workflow en het gebruik van de juiste controles, tussen deze laboratoria te uniformeren. Echter, vanwege de snelle ontwikkelingen op het vlak van technologie en bioinformatica, zijn verbeteringen te verwachten zodat gemaakte keuzes relatief korte tijd zullen worden toegepast waardoor het aannemelijk is dat gestandaardiseerde procedures regelmatig dienen te worden aangepast.

8.1 De technologie: de sequencingsystemen

Voor het uitvoeren van NGS analyses zijn er verschillende systemen van verschillende fabrikanten beschikbaar. Deze systemen verschillen in 1) de kosten van de systemen, 2) het aantal sequenties wat kan worden bepaald (throughput), 3) de lengte van de fragmenten waarvan de sequentie kan worden bepaald en 4) de nauwkeurigheid waarmee de sequentie van de fragmenten wordt bepaald (foutenpercentage of error rate). Momenteel worden vooral de sequencingsystemen van leverancier Illumina toegepast voor molecular profiling. Het foutenpercentage is bij Illumina sequencing zeer laag, de systemen zijn relatief eenvoudig en tegen lage kosten te bedienen. De mogelijkheid om alleen korte DNA-fragmenten (maximaal 300-400bp) te sequencen vormt een nadeel van Illumina sequencing. Hierdoor is het met de standaardmethoden niet mogelijk om 16S rRNA gensequenties te identificeren tot op soortniveau en is genusniveau tot nu toe het maximaal haalbare. Inmiddels zijn wel alternatieve procedures ontwikkeld waarmee het ook mogelijk is geworden om ook met Illumina sequencing de sequenties tot op soortniveau te identificeren. De mogelijkheden van deze alternatieve aanpak worden in het BTO programma van 2021-2022 verkend. Nieuwe sequencingsystemen, zoals bijvoorbeeld de MinION van Oxford Nanopore, zijn goedkoper, eenvoudiger te gebruiken en richten zich op sequencing van langere fragmenten. Vanwege het eenvoudiger gebruik en potentieel betere identificatie van deze langere fragmenten zijn deze systemen zeer aantrekkelijk, maar wordt de toepassing momenteel nog belemmerd door de hoge foutenpercentage. De kwaliteit van de MinION analyses verbetert echter snel waardoor te verwachten is dat deze systemen steeds meer toegepast zullen gaan worden.

8.2 Standaardisatie van microbial profiling

8.2.1 Monsternamen en conservering

De manier van monsternamen voor microbial profiling analyses wijkt weinig af van andere microbiologische of moleculairbiologische analyses. Een aantal zaken is aanvullend van belang:

- Gebruik van steriele, DNA-vrije monsterflessen
- Geen toevoegingen, zoals thiosulfaat, aan de monsterflessen. Deze toevoegingen kunnen, vanwege groeibevordering, resulteren in veranderingen in de populaties.
- De microbiële gemeenschap zoveel mogelijk stabiliseren door koelen (ijs of koelpatronen) tijdens transport.
- Isolatie van DNA binnen 24 uur na monsternamen om te waarborgen dat de samenstelling van de geanalyseerde microbiële populatie overeenkomt met de populatie tijdens monsternamen.
- Er wordt geadviseerd om van elk monster meerdere replica's te analyseren.

8.2.2 Isolatie van DNA

Om voldoende DNA te isoleren uit (water)monsters met weinig biomassa en bias in de analyseresultaten te voorkomen is het van belang om een DNA-isolatiemethode te gebruiken waarmee het DNA uit alle microbiële cellen met vergelijkbare en hoge efficiëntie verloopt. Vooral bij het proces van het vrijmaken van DNA uit de cellen kunnen efficiëntieverschillen bestaan, waardoor isolatie van DNA uit micro-organismen met een sterke celwand minder efficiënt verloopt. Literatuuronderzoek laat zien dat het toepassen van methoden waarbij een combinatie van fysische behandeling (bead-beating) en chemische lysis van de cellen wordt gebruikt om de celwanden open te "breken" de voorkeur heeft. Bij KWR wordt hiervoor de Qiagen Power Biofilm Kit (voorheen van MoBio) succesvol toegepast.

8.2.3 Effect van DNA concentratie

De analyse van monsters met weinig biomassa, zoals drink- of grondwater brengt extra uitdagingen met zich mee. Bij deze monsters kan het zijn dat de concentratie DNA te laag is om representatieve analyseresultaten te genereren en heeft de introductie van contaminaties (vanuit reagentia, laboratoriumhandelingen of gebruiksartikelen) relatief veel invloed op de analyseresultaten.

8.2.4 Keuze van primers

Literatuuronderzoek laat zien dat er veel verschillende primerparen bruikbaar zijn voor het uitvoeren van microbial profiling analyses (paragraaf 4.3.1). Vanwege de optimale fragmentlengte, de hoge dekking (coverage) en het grote onderscheidend vermogen wordt de V4 regio door veel onderzoeksgroepen gebruikt bij het toepassen van Illumina sequencing. Hierdoor zijn er veel V4 sequenties beschikbaar in de databases, waardoor identificatie van de sequenties beter mogelijk is. Bij KWR is daarom gekozen om de sequencing uit te voeren op de V4 regio. De primerkeuze is van grote invloed op de uiteindelijke analyseresultaten. Het combineren van onderzoeksprojecten waarbij data is verkregen met verschillende primerparen is daarom niet goed mogelijk en wordt afgeraden.

8.2.5 Bioinformatica voor interpretatie van data

Voor het interpreteren van data zijn er verschillende softwarepakketten beschikbaar waarbij de volgende twee de belangrijkste zijn: Qiime2 en MOTHUR. De vernieuwing en ontwikkeling van nieuwe tools verloopt bij Qiime2 sneller (ook door gebruikers) en Qiime2 heeft als belangrijk voordeel dat data snel en eenvoudig kan worden gevisualiseerd via een interactief platform (Qiime Viewer). Deze ontwikkelingen hebben ervoor gezorgd dat Qiime2 het meest gebruikte analyse pipeline is geworden. Een groot verschil tussen beide pijplijnen is ook dat bij Qiime2 wordt afgestapt van clustering van sequenties op basis van de mate van sequentieovereenkomst (OTU's met 97% sequentieovereenkomst) maar gebruik wordt gemaakt naar het gebruik van ASV's waarbij elk sequentieverschil wordt gezien als potentieel relevant. Het gebruik van ASV's wordt inmiddels gezien als de beste methode voor het clusteren van 16S rRNA-sequenties (Callahan, McMurdie et al. 2017). Bij KWR is in het verleden gestart met het

gebruik van MOTHUR, maar is vervolgens gekozen om over te stappen naar Qiime2. Bij MOTHUR is men voor visualisaties geheel aangewezen op het gebruik van R-scripts. Voor een beperkt aantal specifieke statistische analyses wordt Qiime2 nog aangevuld met R-scripts. In deze rapportage is de toepassing van verschillende R-scripts en Qiime2 tools gedemonstreerd.

Database

Voor het identificeren van de 16S rRNA gensequenties zijn er verschillende databases met referentiesequenties beschikbaar. De SILVA database (Quast, Pruesse et al. 2013) bevat kwalitatief goede sequenties en wordt het meest frequent voorzien van updates. Bij KWR wordt daarom gekozen voor het gebruik van deze database. Qiime2 bevat bovendien mogelijkheden om deze database te voorzien van nieuwe sequenties en de classifier, om sequenties met de database te vergelijken, te trainen.

8.2.6 Kwaliteitscontroles

Blanco monsters

Om eventuele contaminatiesequenties te kunnen onderscheiden van de sequenties uit de monsters is de analyse van blanco monsters, vooral bij de analyse van monsters met weinig biomassa, van groot belang. De analyse van een groot aantal blanco monsters (paragraaf 4.5.1) maakt duidelijk dat in veel blanco monsters sequenties worden aangetoond die ook in watermonsters worden aangetoond. Dit is waarschijnlijk het gevolg van contaminaties in reagentia of van sporadische leesfouten van de indexsequenties waarmee sequenties uit watermonsters ten onrechte worden geïdentificeerd als “blanco-sequenties”. Om eventuele fouten bij de interpretatie van de analyseresultaten te voorkomen is het van belang om bij elk experiment blanco monsters (blanco filtratiemonsters en blanco PCR-monsters) te analyseren en de resultaten hiervan te vergelijken met de resultaten van de te onderzoeken monsters.

Mock communities

De analyse van een kunstmatig samengestelde microbiële gemeenschap met bekende samenstelling (mock community) is van belang als controle van de gehele workflow (DNA-siolutie/16S-PCR/sequencing/bioinformatica) en voor het bepalen van het aandeel sequentiefouten (error rate) in de data. In dit onderzoek zijn twee verschillende mock communities toegepast, één samengesteld uit verschillende controlestammen en een commercieel verkrijgbare (Zymo Research) community. Het onderzoek maakt duidelijk dat een klein deel van de sequenties uit de mock communities, waarschijnlijk t.g.v. sequentiefouten in de index sequenties (“cross talking”), ook kan worden aangetoond in geanalyseerde watermonsters. Dit is vooral heel vervelend als in de mock communities micro-organismen aanwezig zijn waarvan de het aantonen van deze organismen in watermonsters tot verkeerde conclusies kan leiden (pathogenen of fecale indicatoren zoals *E. coli* of intestinale enterococchen). Om dit type problemen te voorkomen kan een andere mock community worden ontwikkeld waarin alleen micro-organismen worden gebruikt die normaal al in water aanwezig zijn. Ook kan onderzocht worden of het mogelijk is om “cross talking” sterk te beperken door lagere concentraties 16S amplicons van de mock communities tijdens de sequencing toe te voegen.

8.3 Toepassing van microbial profiling voor drinkwateronderzoek

8.3.1 Reproduceerbaarheid

Veel van de in dit onderzoek uitgevoerde analyses zijn in duplo of triplo uitgevoerd. Deze meervoudige analyses laten zien dat zeer reproduceerbare analyseresultaten worden verkregen met het toegepaste protocol.

8.3.2 Benodigde DNA concentratie

Experimenteel is vastgesteld (paragraaf 5.2) dat reproduceerbare analyses kunnen worden uitgevoerd op drink- of oppervlaktewatermonsters met een minimale concentratie DNA van 1 pg/μl (corresponderend met ca. 20 cellen

per ml). Lagere concentraties resulteren in verlies van alpha-diversiteit en verandering van de aangetoonde verschillen tussen populaties (de beta-diversiteit). Voor het detecteren van zeldzame sequenties in de populaties zijn hogere DNA-concentraties noodzakelijk (afhankelijk van de concentratie van deze zeldzame micro-organismen).

8.3.3 Detectie van verontreinigingen

Door analyses uit te voeren op drinkwater gemengd met verdunningsreeksen van riool- en oppervlaktewater is onderzocht in hoeverre microbial profiling toegepast kan worden voor het detecteren van verontreinigingen in drinkwater. Deze experimenten laten zien dat de verontreinigingen met hoge gevoeligheid kunnen worden aangetoond. De software tool “Sourcetracker” (Knights, Kuczynski et al. 2011) kan daarbij gebruikt worden om betrouwbaar het aandeel verontreiniging in de microbiële populaties te bepalen.

8.3.4 Praktijktoeepassingen (casestudies)

De toepassing van microbial profiling op monsters van praktijksituaties afkomstig van acht drinkwaterzuiveringen laat zien dat microbial profiling inzicht geeft in de samenstelling van de microbiologie in water en wat het effect is van verschillende zuiveringsstappen op de microbiologie. Deze casestudies laten zien welke veranderingen er optreden in de microbiologie en geeft, in potentie de mogelijkheid om veranderingen te monitoren en op basis daarvan zuiveringsprocessen te monitoren.

9 Aanbevelingen

Het in dit rapport beschreven onderzoek heeft geresulteerd in de volgende aanbevelingen

- Alle stappen in de gehele workflow hebben invloed op het eindresultaat, daarom wordt aanbevolen om alle stappen (laboratorium en bioinformatica) te standaardiseren.
- Het belang van het gebruik van goede controles is groot.
 - Blanco monsters moeten bij elke analyse worden meegenomen om de invloed van de aanwezigheid van contaminerende sequenties op de analyses zichtbaar te kunnen maken. Er wordt aanbevolen om twee verschillende blanco monsters mee te nemen: één monster die het gehele traject ondergaat (filtratie/DNA isolatie/PCR en sequencing) en één monster waarmee alleen het laatste deel van de procedure (PCR+sequencing) wordt onderzocht. Met deze verschillende blanco's wordt inzicht verkregen in het optreden van eventuele contaminaties tijdens de verschillende onderdelen van het gehele proces.
 - Om de integriteit van het gehele proces (laboratorium en bioinformatica) te kunnen monitoren is de analyse van een kunstmatig samengestelde gemeenschap met hierin bekende bacteriesoorten (mock community) van belang. Voor vervolgonderzoek wordt de ontwikkeling van een geoptimaliseerde mock community aanbevolen.
- Vanwege de korte lengte van de fragmenten waarop NGS analyses worden uitgevoerd is identificatie van de sequenties momenteel slechts mogelijk tot op familie of genus. De ontwikkeling van alternatieve methoden wordt aanbevolen en is reeds gestart in BTO 2021-2022 waarmee sequentie analyse van het gehele 16S rRNA gen plaats vindt, waardoor identificatie tot op soortniveau mogelijk wordt.
- Een korte inventariserende studie laat zien dat microbial profiling reproduceerbare profielen laat zien die goed inzicht geven in de samenstelling van de microbiologie in watermonsters tijdens zuiveringsprocessen en waarmee inzicht wordt verkregen in veranderingen die t.g.v. zuiveringsstappen optreden. Er wordt aanbevolen om dergelijke studies uitgebreider uit te voeren door met intensievere monstercampagnes metingen uit te voeren gedurende het jaar. Deze metingen zullen inzicht kunnen geven in de samenstelling en stabiliteit van de microbiologie na de verschillende zuiveringsstappen en de rol van de microbiologie bij zuiveringsprocessen.

10 Literatuur

- Albertsen, M., S. M. Karst, A. S. Ziegler, R. H. Kirkegaard and P. H. Nielsen (2015). "Back to Basics--The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities." *PLoS One* **10**(7): e0132783.
- Almeida, A., A. L. Mitchell, A. Tarkowska and R. D. Finn (2018). "Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments." *Gigascience* **7**(5).
- Amir, A., D. McDonald, J. A. Navas-Molina, E. Kopylova, J. T. Morton, Z. Zech Xu, E. P. Kightley, L. R. Thompson, E. R. Hyde, A. Gonzalez and R. Knight (2017). "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns." *mSystems* **2**(2): e00191-00116.
- Bokulich, N. A., B. D. Kaehler, J. R. Rideout, M. Dillon, E. Bolyen, R. Knight, G. A. Huttley and J. Gregory Caporaso (2018). "Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin." *Microbiome* **6**(1): 90.
- Bolyen, E., J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodríguez, J. Chase, E. Cope, R. Da Silva, P. C. Dorrestein, G. M. Douglas, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. Kaehler, K. B. Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolk, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Lottfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, I. I. M. S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K. C. Weber, C. H. D. Williamson, A. D. Willis, Z. Z. Xu, J. R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight and J. G. Caporaso (2018). "QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science." *PeerJ Preprints* **6**: e27295v27292.
- Brandt, J. and M. Albertsen (2018). "Investigation of Detection Limits and the Influence of DNA Extraction and Primer Choice on the Observed Microbial Communities in Drinking Water Samples Using 16S rRNA Gene Amplicon Sequencing." *Frontiers in microbiology* **9**: 2140-2140.
- Brigmon, R. L., H. W. Martin and H. C. Aldrich (1997). "Biofouling of groundwater systems by *Thiothrix* spp." *Curr Microbiol* **35**(3): 169-174.
- Brouwer, S., v. d. Wielen, Paul, M. Schriks, M. Claassen and L. Kosrs (2017). "Citizen science project toont waarde van burgers als wetenschappers." *H2O, Water Matters* 32-35.
- Callahan, B. J., D. Grinevich, S. Thakur, M. A. Balamotis and T. B. Yehezkel (2021). "Ultra-accurate microbial amplicon sequencing with synthetic long reads." *Microbiome* **9**(1): 130.
- Callahan, B. J., P. J. McMurdie and S. P. Holmes (2017). "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis." *The ISME Journal* **11**(12): 2639-2643.
- Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature Methods* **13**(7): 581-583.
- 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., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld and R. Knight (2010). "QIIME allows analysis of high-throughput community sequencing data." *Nature Methods* **7**: 335.
- Collins, F. S., M. Morgan and A. Patrinos (2003). "The Human Genome Project: Lessons from Large-Scale Biology." *Science* **300**(5617): 286.
- da Fonseca, R. R., A. Albrechtsen, G. E. Themudo, J. Ramos-Madrugal, J. A. Sibbesen, L. Maretty, M. L. Zepeda-Mendoza, P. F. Campos, R. Heller and R. J. Pereira (2016). "Next-generation biology: Sequencing and data analysis approaches for non-model organisms." *Mar Genomics* **30**: 3-13.

- de Vet, W. W., I. J. Dinkla, L. C. Rietveld and M. C. van Loosdrecht (2011). "Biological iron oxidation by *Gallionella* spp. in drinking water production under fully aerated conditions." *Water Res* **45**(17): 5389-5398.
- Donkor, E. S. (2013). "Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics." *Genes (Basel)* **4**(4): 556-572.
- Eisenhofer, R., J. J. Minich, C. Marotz, A. Cooper, R. Knight and L. S. Weyrich (2019). "Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations." *Trends Microbiol* **27**(2): 105-117.
- Ewing, B. and P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." *Genome Res* **8**(3): 186-194.
- Faith, D. P. (1992). "Conservation evaluation and phylogenetic diversity." *Biological Conservation* **61**(1): 1-10.
- Heijnen, L. (2016). "NGS voor detectie van fecale besmettingen." *BTO Rapport 2016.026*.
- Heijnen, L. (2019). "Verkenning van de toepassingsmogelijkheden van MinION sequencing." *BTO rapport 2019.057*.
- Heijnen, L. K.-M., A (2019). "NGS om de hygiënische betekenis van indicatororganismen in duinfiltraat te bepalen." *Memo DPWE onderzoek Project nr. 402635-009*.
- Italiaander, R., L. Heijnen and B. Wullings (2011). "Validatie van methoden voor DNA-extractie uit water." *BTO rapport 2011.033(s)*.
- Knights, D., J. Kuczynski, E. S. Charlson, J. Zaneveld, M. C. Mozer, R. G. Collman, F. D. Bushman, R. Knight and S. T. Kelley (2011). "Bayesian community-wide culture-independent microbial source tracking." *Nature Methods* **8**: 761.
- Learbuch, K. L. G., H. Smidt and P. van der Wielen (2021). "Influence of pipe materials on the microbial community in unchlorinated drinking water and biofilm." *Water Res* **194**: 116922.
- Lim, M. Y., E. J. Song, S. H. Kim, J. Lee and Y. D. Nam (2018). "Comparison of DNA extraction methods for human gut microbial community profiling." *Syst Appl Microbiol* **41**(2): 151-157.
- Liu, G., Y. Zhang, E. van der Mark, A. Magic-Knezev, A. Pinto, B. van den Bogert, W. Liu, W. van der Meer and G. Medema (2018). "Assessing the origin of bacteria in tap water and distribution system in an unchlorinated drinking water system by SourceTracker using microbial community fingerprints." *Water Research* **138**: 86-96.
- Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* **15**(12): 550.
- Mallott, E. K., R. S. Malhi and K. R. Amato (2019). "Assessing the comparability of different DNA extraction and amplification methods in gut microbial community profiling." *Access Microbiol* **1**(7): e000060.
- Mandal, S., W. Van Treuren, R. A. White, M. Eggesbø, R. Knight and S. D. Peddada (2015). "Analysis of composition of microbiomes: a novel method for studying microbial composition." *Microb Ecol Health Dis* **26**: 27663.
- Minich, J. J., J. G. Sanders, A. Amir, G. Humphrey, J. A. Gilbert and R. Knight (2019). "Quantifying and Understanding Well-to-Well Contamination in Microbiome Research." *mSystems* **4**(4).
- Oswald, K., J. S. Graf, S. Littmann, D. Tienken, A. Brand, B. Wehrli, M. Albertsen, H. Daims, M. Wagner, M. M. Kuypers, C. J. Schubert and J. Milucka (2017). "Crenothrix are major methane consumers in stratified lakes." *Isme j* **11**(9): 2124-2140.
- 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.
- Rintala, A., S. Pietilä, E. Munukka, E. Eerola, J.-P. Pursiheimo, A. Laiho, S. Pekkala and P. Huovinen (2017). "Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor." *Journal of Biomolecular Techniques : JBT* **28**(1): 19-30.
- Roeselers, G., J. Coolen, P. W. van der Wielen, M. C. Jaspers, A. Atsma, B. de Graaf and F. Schuren (2015). "Microbial biogeography of drinking water: patterns in phylogenetic diversity across space and time." *Environ Microbiol* **12**(10): 1462-2920.
- Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A* **74**(12): 5463-5467.
- 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.
- Siddiqi, M. Z., W. Sok, G. Choi, S. Y. Kim, J.-H. Wee and W. T. Im (2020). "Simplicispira hankyongi sp. nov., a novel denitrifying bacterium isolated from sludge." *Antonie van Leeuwenhoek* **113**(3): 331-338.
- Siegwald, L., S. Caboche, G. Even, E. Viscogliosi, C. Audebert and M. Chabé (2019). "The Impact of Bioinformatics Pipelines on Microbiota Studies: Does the Analytical "Microscope" Affect the Biological Interpretation?" *Microorganisms* **7**(10).

- Staley, C., T. Kaiser, A. Lobos, W. Ahmed, V. J. Harwood, C. M. Brown and M. J. Sadowsky (2018). "Application of SourceTracker for Accurate Identification of Fecal Pollution in Recreational Freshwater: A Double-Blinded Study." *Environ Sci Technol* **52**(7): 4207-4217.
- Stinson, L. F., J. A. Keelan and M. S. Payne (2019). "Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses." *Lett Appl Microbiol* **68**(1): 2-8.
- Timmers, P. B. v. d., Joost; Heijnen, Leo; Kardinaal, Edwin; Sollie, Susan; Medema, Gertjan (2020). "Microbiële vingerafdruk voor bronopsporing in oppervlaktewaterbeheer." *H2O Water Matters* **4** (Juni 2020).
- Timmers, P. H. A., T. Slootweg, A. Knezev, M. van der Schans, L. Zandvliet, A. Reus, D. Vughs, L. Heijnen, T. Knol, J. El Majjaoui, P. van der Wielen, P. J. Stuyfzand and K. Lekkerkerker-Teunissen (2022). "Improved drinking water quality after adding advanced oxidation for organic micropollutant removal to pretreatment of river water undergoing dune infiltration near The Hague, Netherlands." *J Hazard Mater* **429**: 128346.
- Vavourakis, C. D., L. Heijnen, M. Peters, L. Marang, H. A. M. Ketelaars and W. A. M. Hijnen (2020). "Spatial and Temporal Dynamics in Attached and Suspended Bacterial Communities in Three Drinking Water Distribution Systems with Variable Biological Stability." *Environ Sci Technol* **54**(22): 14535-14546.
- Velásquez-Mejía, E. P., J. de la Cuesta-Zuluaga and J. S. Escobar (2018). "Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces." *Appl Microbiol Biotechnol* **102**(1): 403-411.
- Wasimuddin, K. Schlaeppli, F. Ronchi, S. L. Leib, M. Erb and A. Ramette (2019). "Evaluation of primer pairs for microbiome profiling across a food chain from soils to humans within the One Health framework." *bioRxiv*: 843144.
- Weiss, S., Z. Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, C. Lozupone, J. R. Zaneveld, Y. Vázquez-Baeza, A. Birmingham, E. R. Hyde and R. Knight (2017). "Normalization and microbial differential abundance strategies depend upon data characteristics." *Microbiome* **5**(1): 27.
- Witzke, M. C., A. Gullic, P. Yang, N. J. Bivens, P. R. F. Adkins and A. C. Ericsson (2020). "Influence of PCR cycle number on 16S rRNA gene amplicon sequencing of low biomass samples." *Journal of Microbiological Methods* **176**: 106033.
- Yeh, Y.-C., D. M. Needham, E. T. Sieradzki and J. A. Fuhrman (2018). "Taxon Disappearance from Microbiome Analysis Reinforces the Value of Mock Communities as a Standard in Every Sequencing Run." *mSystems* **3**(3): e00023-00018.
- Yilmaz, P., L. W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, 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.

I Workshop “Toepassing van NGS in de drinkwatersector”

Kort verslag van de workshop

Traditioneel worden microbiologische analyses uitgevoerd door organismen te kweken. Bij moleculaire methoden wordt echter het genetische materiaal, DNA en RNA, in het water onderzocht. PCR wordt al veelvuldig toegepast bij onderzoek en bedrijfsvoering van drinkwater. Bij PCR wordt gericht naar één specifieke DNA of RNA sequentie gezocht. Next Generation Sequencing (NGS) maakt het mogelijk om, tegen lage kosten, de samenstelling en volgorde van de DNA-bouwstenen (DNA-sequentie) van enorme hoeveelheden DNA fragmenten te bepalen. Op 12 september 2018 werd bij KWR een workshop gehouden over de toepassingsmogelijkheden van Next Generation Sequencing (NGS) technologieën in de drinkwatersector. Deze BTO workshop werd bezocht door ca. 25 deelnemers afkomstig van de drinkwaterbedrijven hun laboratoria. De vragen die tijdens deze workshop centraal stonden waren: welke toepassingsmogelijkheden van DNA-sequencing zijn er nu, welke toepassingen worden in de toekomst verwacht en welke stappen moeten worden gezet om deze technologieën beter toepasbaar te krijgen. Om de huidige stand van de technologie en de mogelijkheden voor toepassing helder te krijgen werd de workshop gestart met presentaties van Gertjan Medema (KWR), Leo Heijnen (KWR), Louise Vanysacker (De Watergroep), Maarten Lut (Oasen, onderzoek i.s.m. Gang Liu), Nora Sutton (WUR) en Gerben Hermes (WUR). Deze presentaties gaven inzicht in de technologische mogelijkheden en beperkingen van NGS en zorgden voor een overzicht van interessante toepassingen. De middag werd gevuld met discussies waarbij werd geïnventariseerd voor welke onderzoeksvragen NGS van meerwaarde kan zijn en welke ontwikkelingen nodig zijn om NGS breder toe te passen. De workshop maakte duidelijk dat het met NGS mogelijk is om een vrijwel compleet beeld te genereren van de samenstelling van microbiële populatie in water. Hierdoor is NGS momenteel al van waarde om: a) de kwaliteit van het water in het traject van winning tot distributie te kunnen beoordelen en b) de processen die plaats vinden tijdens de zuivering te leren kennen en te kunnen beheeren. Voor een bredere toepassing moet de methode wel verder ontwikkeld en gestandaardiseerd worden. Een deel van de bacteriegroepen die in drinkwater worden aangetroffen is nog onbekend, uitbreiding van de huidige referentiedatabases zal zorgen voor beter inzicht in de betekenis van deze, nu nog, onbekende bacteriegroepen. Daarnaast is er momenteel nog geen gestandaardiseerde methode waardoor resultaten verkregen door verschillende onderzoeksgroepen moeilijk vergelijkbaar zijn. Voor de ‘natte’ analyse is het van belang om de juiste experimentele controles toe te passen en de prestatiekenmerken van een toegepaste methode goed te kennen. Voor de ‘droge’ analyse, het vergelijken van meetresultaten met databases, is ook standaardisatie nodig om resultaten onderling te vergelijken en veranderingen in de tijd te volgen. Deze onderdelen maken deel uit van het BTO project “Moleculair-biologische methoden voor microbiologie”.

II Protocol 16S v4 sequencing op Illumina MiSeq

Protocol voor sequencing van de v4 regio van 16S rRNA amplicons

Isolatie van DNA

Filtratie van watermonsters

Voor het filtreren van watermonsters wordt gebruik gemaakt van het standaard membraanfiltratieprotocol van LMB voor DNA toepassingen. Bij het filtreren van watermonsters met weinig biomassa (o.a. drinkwater) wordt een zo groot mogelijk volume gefiltreerd, het streven is minimaal 1 liter.

Isolatie van DNA

Voor het isoleren van DNA wordt gebruik gemaakt van het standaard LMB protocol voor het isoleren van DNA waarbij gebruik gemaakt wordt van de de Power Biofilm Kit van Qjagen (voorheen MoBio). Er wordt geen interne controle toegevoegd.

Meting van concentratie

De concentratie van het geïsoleerde DNA wordt bepaald met de QuBit waarbij gebruik wordt gemaakt van de High Sensitivity dsDNA kit.

Criteria moeten worden opgesteld.

Maken van amplicons

Amplicon primers

Voor het genereren van amplicons van de v4 regio van 16S rRNA genen worden de volgende primers gebruikt:

515F_IIIu: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTGCCAGCMGCCGCGTAA-3'**

806R_IIIu: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GGACTACHVGGGTWTCTAAT-3'**

Amplificatie 16S-V4

Met deze stap wordt het V4 fragment van 16S rRNA moleculen vermenigvuldigd

Mix per monster:

12,5 µl Kappa HiFi Mix

6,5 µl PCR H₂O

0,5 µl 515F_IIIu

0,5 µl 806R_IIIu

5 µl DNA

PCR profiel:

95°C 3 min

30 cycli:

98 °C 20 sec.

60 °C 15 sec.

72 °C 30 sec

5 min 72 °C

Hold at 4°C

PCR cleanup 1

Met deze stap worden de amplicons gezuiverd van de nog aanwezige primers, nucleotiden en reagentia.

Procedure

1. Centrifuge the Amplicon PCR plate at 1,000 × g at 20°C for 1 minute to collect condensation, carefully remove seal.
2. Using a multichannel pipette set to 25 µl, transfer the entire Amplicon PCR product from the PCR plate to the MIDI plate. Change tips between samples.
3. Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough depending on the number of samples processing.
4. Using a multichannel pipette, add 17 µl of AMPure XP beads to each well of the Amplicon PCR plate. Change tips between columns.
5. Seal plate and shake at 1800 rpm for 2 minutes if using a MIDI plate.
6. Incubate at room temperature without shaking for 5 minutes.
7. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.
8. With the Amplicon PCR plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant. Change tips between samples.
9. With the Amplicon PCR plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
 10. Using a multichannel pipette, add 180 µl of freshly prepared 80% ethanol to each sample well.
 11. Incubate the plate on the magnetic stand for 30 seconds.
 12. Carefully remove and discard the supernatant (set the pipet to 220 µl).
 13. With the Amplicon PCR plate on the magnetic stand, perform a second ethanol wash as follows:
 14. Using a multichannel pipette, add 180 µl of freshly prepared 80% ethanol to each sample well.
 15. Incubate the plate on the magnetic stand for 30 seconds.
 16. Carefully remove and discard the supernatant (set the pipet to 220 µl).
 17. Use a P20 multichannel pipette with fine pipette tips to remove excess ethanol.
 18. With the Amplicon PCR plate still on the magnetic stand, allow the beads to air-dry for 10 minutes.
 19. Remove the Amplicon PCR plate from the magnetic stand. Using a multichannel pipette, add 52.5 µl of 10 mM Tris pH 8.5 to each well of the Amplicon PCR plate.
 20. Gently pipette mix up and down 10 times, changing tips after each column (or seal plate and shake at 1800 rpm for 2 minutes). Make sure that beads are fully resuspended.
 21. Incubate at room temperature for 2 minutes.
 22. Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.

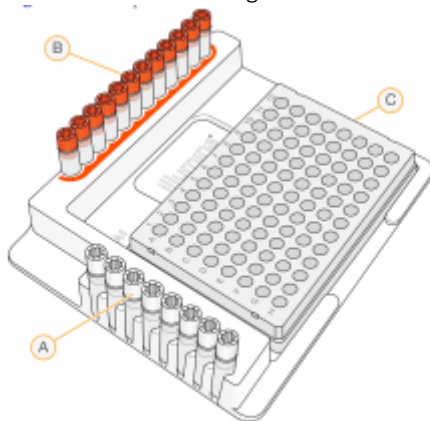
23. Using a multichannel pipette, carefully transfer 40 μl of the supernatant from the Amplicon PCR plate to a new 96-well PCR plate. Change tips between samples to avoid cross-contamination.

Index PCR

Met deze stap worden de fragmenten voorzien van index sequenties waarmee het mogelijk wordt om fragmenten die afkomstig zijn van verschillende monsters te onderscheiden. Dit wordt mogelijk gemaakt door elk monster te voorzien van verschillende index sequenties.

Procedure

1. Using a multichannel pipette, transfer 5 μl from each well to a new 96-well plate. The remaining 45 μl is not used in the protocol and can be stored for other uses.
2. Arrange the Index 1 and 2 primers in a rack (i.e. the TruSeq Index Plate Fixture) using the following arrangements as needed:
 - Arrange Index 2 primer tubes (white caps, clear solution) vertically, aligned with rows A through H.
 - Arrange Index 1 primer tubes (orange caps, yellow solution) horizontally, aligned with columns 1 through 12.



- A Index 2 primers (white caps)
 B Index 1 primers (orange caps)
 C 96-well plate

3. Place the 96-well PCR plate with the 5 μl of resuspended PCR product DNA in the TruSeq Index Plate Fixture.
4. Set up the following reaction of DNA
 - KAPAHifi Hotstart: 25 μl
 - H₂O: 10 μl
 - DNA: 5 μl
 - S-primer: 5 μl
 - N-primer: 5 μl
5. Gently pipette up and down 10 times to mix.
6. Cover the plate with Microseal 'A'.
7. Centrifuge the plate at 1,000 \times g at 20°C for 1 minute.
8. Perform PCR on a thermal cycler using the following program:
 - 95 °C: 3 min
 - 8 cycli met
 - 95 °C: 30 sec.
 - 55 °C: 30 sec.
 - 72 °C: 30 sec.
 - 72 °C: 5 min

- Hold at 4°C

PCR cleanup 2

Met deze stap worden de amplicons gezuiverd van de nog aanwezige primers, nucleotiden en reagentia.

Procedure

1. Centrifuge the Index PCR plate at 280 × g at 20°C for 1 minute to collect condensation, carefully remove seal.
2. Using a multichannel pipette set to 50 µl, transfer the entire Index PCR product from the PCR plate to the MIDI plate. Change tips between samples.
3. Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough depending on the number of samples processing.
4. Using a multichannel pipette, add 50 µl of AMPure XP beads to each well of the Index PCR plate. Change tips between columns.
5. Seal plate and shake at 1800 rpm for 2 minutes if using a MIDI plate. Incubate at room temperature without shaking for 5 minutes.
6. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.
7. With the Index PCR plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant. Change tips between samples.
8. With the Index PCR plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
 - a. Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to each sample well.
 - b. Incubate the plate on the magnetic stand for 30 seconds.
 - c. Carefully remove and discard the supernatant.
9. With the Index PCR plate on the magnetic stand, perform a second ethanol wash as follows:
 - a. Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to each sample well.
 - b. Incubate the plate on the magnetic stand for 30 seconds.
 - c. Carefully remove and discard the supernatant.
 - d. Use a P20 multichannel pipette with fine pipette tips to remove excess ethanol.
10. With the Index PCR plate still on the magnetic stand, allow the beads to air-dry for 10 minutes.
11. Remove the Index PCR plate from the magnetic stand. Using a multichannel pipette, add 27.5 µl of 10 mM Tris pH 8.5 to each well of the Index PCR plate.
12. Gently pipette mix up and down 10 times, changing tips after each column. Make sure that beads are fully resuspended.
13. Incubate at room temperature for 2 minutes.
14. Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.
15. Using a multichannel pipette, carefully transfer 25 µl of the supernatant from the Index PCR plate to a new 96-well PCR plate. Change tips between samples to avoid cross-contamination.

Controle van de library

Controleer de opbrengst en fragmentlengte van de library d.m.v. agarosegel electroforese (2% agarosegel) of m.b.v. de Bioanalyser.

Library Quantification, Normalization and Pooling

Voor het kwantificeren van de libraries wordt een fluorometrische meting met QuBit gebruikt waarbij een specifieke fluorescente kleurstof voor het kwantificeren van dubbelstrengs DNA wordt toegepast.

Bereken de concentratie in nM op basis van de gemeten concentratie in ng/µl. Gebruik hiervoor de volgende omrekening:

$$\frac{\text{Gemeten concentratie in ng/}\mu\text{l}}{(660 \text{ g/mol} \times 457^*)} \times 10^6 = \text{concentratie in nM}$$

* is de fragmentlengte, is 457 in het geval van de toegepaste V4 specifieke amplicon PCR

Library pool

Verdun elke library met Resuspension Buffer (RSB) of 10 mM Tris pH 8,5 tot een concentratie van 4 nM

Pool de libraries door van elk monster 5 µl (met een concentratie van 4 nM) te poolen.

Denaturatie van de Library en sample loading op de MiSeq

Voor het runnen op de MiSeq wordt de gepoolde library:

- Gedenatureerd met NaOH
- Verdund in hybridisatiebuffer
- Hitte gedenatureerd voor de start van de run

Aan elke run wordt PhiX toegevoegd tot% als interne controle en voor het aanbrengen van sequentievariatie in het monster (van belang voor de sequencing).

Voorbereiding

- Hitteblok of waterbad op 96 °C
- MiSeq reagent cartridge ontdooien: bij voorkeur in de koelkast de dag ervoor, eventueel in een waterbad bij kamertemperatuur.
- Maak een ijsbad met een mengsel van ijs en water
- Ontdooi buffer HT1 en bewaar op ijs

Denaturatie Library

- Meng:
 - 5 µl gepoolde library (4 nM)
 - 5 µl 0,2 N NaOH
- Vortex kort
- Spin down
- Incubeer: 5 min. bij kamertemp.
- Voeg 990 µl ijskoude buffer HT1 toe aan het 10 µl library/NaOH mengsel
- Toevoeging van HT1 resulteert in een amplicon library concentratie van 20 pM
- Bewaar gedenatureerde library op ijs

Verdunnen Library

De amplicon library kan volgens onderstaand schema verdund worden tot de gewenste concentratie.

Final Concentration	2 pM	4 pM	6 pM	8 pM	10 pM
20 pM denatured library	60 µl	120 µl	180 µl	240 µl	300 µl
Pre-chilled HT1	540 µl	480 µl	420 µl	360 µl	300 µl

Bij KWR wordt standaard een concentratie van 6 pM gebruikt.

- Deze wordt gemaakt door het mengen van:
 - 180 µl gedenatureerde library
 - 420 µl ijskoude HT1

Denaturatie PhiX controle

- Meng om hiermee een 4nM PhiX te krijgen:
 - 2 µl 10 nM PhiX library
 - 3 µl Tris HCl pH 8,5 (10 mM)
- Meng:
 - 5 µl 4 nM PhiX library
 - 5 µl 0,2 N NaOH
- Vortex kort

- Spin down
- Incubeer: 5 min. bij kamertemp.
- Voeg 990 µl ijskoude buffer HT1 toe aan het 10 µl PhiX library/NaOH mengsel om hiermee een 20 pM PhiX library te krijgen.
- Bewaar gedenatureerde library op ijs
- De 20 pM PhiX library kan volgens onderstaand schema verdund worden tot de gewenste concentratie:

Final Concentration	2 pM	4 pM	6 pM	8 pM	10 pM
20 pM denatured library	60 µl	120 µl	180 µl	240 µl	300 µl
Pre-chilled HT1	540 µl	480 µl	420 µl	360 µl	300 µl

- Verdun de 20 pM PhiX library tot een concentratie van 6 pM door te mengen:
 - 180 µl 20 pM gedenatureerde PhiX library
 - 420 µl ijskoude HT1
- Meng
- Bewaar op ijs

Mengsel amplicon library met PhiX library

Bij KWR wordt 20% PhiX toegevoegd bij de library, hiervoor wordt de PhiX library gemengd met de amplicon library.

- Meng de volgende volumes gedenatureerde PhiX en gedenatureerde amplicon library:
 - 120 µl gedenatureerde en verdunde PhiX
 - 480 µl gedenatureerde en verdunde amplicon library

Hitte denaturatie library

- Incubeer ampliconlibrary/PhiX library mengsel:
 - 2 min. 96°C
 - Meng door cupje rustig 2x te laten kantelen
 - Zet cupje op ijs

III Rekenvoorbeeld Bray-Curtis

De formule om de Bray-Curtis index te berekenen wordt beschreven door:

$$\text{Bray-Curtis} = \frac{2A_{i,j}}{S_i + S_j}$$

waarbij i het ene monster en j het andere monster is; $A_{i,j}$, het laagste aantal sequenties van iedere ASV die in allebei de monsters aanwezig zijn, S_i , het totaal aantal sequenties in monster i en S_j , het totaal aantal sequenties in monster j .

Stel dat in monster i verschillende aantallen sequenties worden gevonden die behoren tot de vier ASVs:

ASV 1: 10 sequenties

ASV 2: 5 sequenties

ASV 3: 20 sequenties

ASV 4: 3 sequenties

En in monster j werden twee dezelfde ASVs aangetroffen en twee nieuwe ASVs met de volgende sequentieverdeling:

ASV2: 8 sequenties

ASV3: 3 sequenties

ASV 5: 15 sequenties

ASV 6: 12 sequenties

Dan wordt de $A_{i,j}$ het minst aantal sequenties van ASV2 in de twee monsters (is 5 sequenties van monster i) + het minst aantal sequenties van ASV3 in de twee monsters (is 3 sequenties van monster j), oftewel $5+3=8$. S_i is het aantal sequenties van monster i (38) en S_j het aantal sequenties van monster j (38). Dat geeft vervolgens de volgende berekening:

$$\text{Bray-Curtis} = \frac{2 \times (5+3)}{(38+38)} = 0,21$$

