Norwegian University
of Life Sciences

Master's Thesis 2022 60 ECTS
Faculty of Biosciences

Phenotypic and Genotypic Characterization of Atlantic Salmon (*Salmo salar)* Gut Microbiota

Sander Johan Mangseth Johnsen
MSc Biology

Acknowledgments

This master thesis was performed at the faculty of biosciences at the Norwegian University of Life Sciences (NMBU), with Phillip Pope and Sabina Leanti La Rosa as my supervisors. This project was a part of the Improvafish project and would express my gratitude for allowing me to be a part of a project with goals that I admire.

I would like to personally thank Sabina, who supervised me from start to finish from everything to lab work to writing. She walked me through every step in the process while answering every question I had, never berating my many mistakes. Even when I got covid and had to stay in, she took time out of her day to continue my lab work to keep everything on schedule. Without her, this paper certainly wouldn't exist. I would also give my thanks to Phillip, for reading through my draft and giving very needed guidance in the final steps of the writing process.

I am also incredibly grateful to Arturo Vera_Ponce_De_Leon, who assisted me in all the programming and data analysis steps. He always walked me through whenever I had blocks regarding coding or errors, and he was always available to give me quick responses.

Abstract

Knowledge of the Atlantic Salmon (*Salmo salar*) gut microbiota falls behind the research done about the human gut microbiome. Creating greater understanding of the bacteria present and their capabilities helps to create a more sustainable aquaculture industry by increasing the possibilities of creating better fish health and farming efficiency. This thesis aims to make a characterization of the gut bacteria using a culturomic based approach using 16S rRNA analysis with Sanger sequencing and Oxford Nanopore sequencing for whole genome analysis in the aim of building upon the Salmon Microbial Genome Atlas (SMGA).

Samples was taken from the distal tract of the fish was plated onto different media and the best two media with the best growth was chosen for further work. 56 (34 BHI/22TSA) isolates were chosen for Sanger sequencing, while 45 isolates were processed further for ONT sequencing using the Promethion sequencer.

The gut microbiota was dominated by the *Pseudomonadota* phylum, with smaller quantities of *Bacillota* and *Actinomycetota*, while *Lelliotta* showed the most frequent hits on Genus level. These isolates showed predicted functionality for basic polysaccharide degradation like chitin and acetate for SCFA synthesis, with multiple hits for beta-mannan degrading enzymes.

Table of Contents

1. INTRODUCTION

The research of the unseeable but vastly important gut microbiome has undergone massive scrutiny from the scientific community over the last decade (Marchesi et al., 2016). The growing use and evolution of technologies like high throughput sequencing have expanded our knowledge about the functions and applications of the gut microbiome. These studies have been mainly focusing on the inner workings of the human gut microbiota, however the gut of livestock animals like pigs have also been in the spotlight due to their similar physiological properties that make them a valuable model animal (Heinritz et al., 2013). Although the research of these gut microbiomes has been relatively expansive, the study of teleost gut microbiome is still far behind (Llewellyn et at., 2014). A teleost in particular; the Atlantic Salmon (*Salmo salar)*, which has immense value as a research topic due to its prevalence in the aquaculture industry, has become a model organism for the study of marine gut microbiomes.

A microbiome is a collection of every group of microorganisms that reside in a habitat, and there exists one inside the digestive system of every vertebrate species (Ciric et al., 2018). The intricate interactions between microbiomes and their hosts are vast, and the notion that we are dependent on our microbiota has been studied through the comparison of fitness in animals with normal and deprived gut microbiota (Moran & Hammer, 2019). Evolution has turned these organisms that were most likely parasitic in nature into a mutualistic relationship that the host is dependent on (Moran et al., 2019), like larval mosquitoes that require aerobic microorganisms to deplete oxygen to fully develop (Coon et al., 2014). Studies have shown that the microbiota can have a significant effect on their hosts through metabolism, gut function and protection against gut related diseases (Egerton et al., 2018; Semova et al., 2012; Dehler et al., 2017). The most abundant carbon complexes are plant polysaccharides, but the energy locked in their bonds is mostly unobtainable without the gut microbiota that contains the enzymatic capabilities for their breakdown (Moran & Hammer, 2019) (Nishida & Ochman, 2018). The protection gut microorganisms bring against pathogenic intruders is one of the most documented areas of study (Moran & Hammer, 2019). While the mechanism behind the protection is less clear than its effects, studies have shown that they are diverse, ranging from modulation of the mucus layer to the production of antibacterial substances (Moran & Hammer, 2019).

As mentioned, the area of the fish microbiome has been explored, however it lags substantially behind compared to the knowledge of the mammalian gut microbiota (Heinritz et al., 2013). The microbial colonization in the fish starts when the sterile fish larvae are exposed to the surrounding water after hatching (Egerton et al., 2018). The consumed bacteria colonize in the newly forming gastrointestinal tract and form the initial gut microbiota, which further diversifies over time through feeding (Egerton et al., 2018). Interestingly, studies show that the microbial composition varies heavily in different life stages of the salmon, hinting that the environment the fish lives in can have significant effect on its microbial content (Llewellyn et al., 2016)

Although the Salmon gut microbiota contains bacteria, viruses, fungi, Protoctista, yeast and archaea, bacteria is by far the most dominant domain (Gajardo et al., 2016). Of the bacteria, they are primarily represented by *Pseudomonadota* (Proteobacteria), *Fusobacteriota*, *Bacillotas* (Firmicutes), *Bacteroidetes* and *Actinomycetota* (Actinobacteria) (Egerton et al., 2018) (Gajardo et al., 2016).

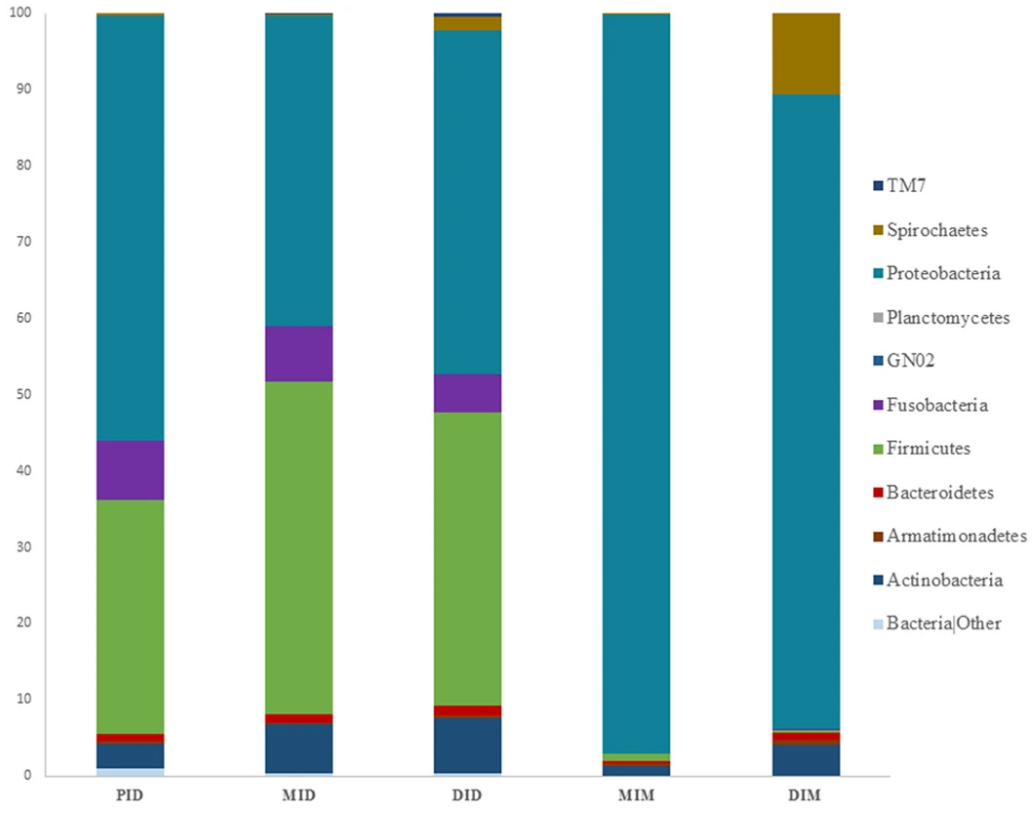


Figure 1: (Gajardo et al., 2016) Taxonomic classification of Salmon gut bacteria in the different gut compartments on phylum level using 16S rRNA sequencing of the V1 and V2 region of the 16S gene. Note that this figure uses the old nomenclature for

some phyla. Composition is shown from the Proximal Intestinal Digesta (PID), Mid Intestinal Digesta (MID), Distal Intestinal Digesta (DID), Mid Intestinal Mucosa and Distal Intestinal Mucosa.

1.1 Atlantic salmon

Atlantic Salmon is a ray-finned fish in the Salmonidae family and shares its genus together with trout (Artsdatabanken, s.a.). It is an anadromous species, which entails that its life cycle revolves around birth in freshwater before migrating to the ocean after a series of physiological changes known as smoltification. After feeding and maturing in saltwater, they end their life cycle by migrating back to freshwater for spawning (Vollestad, s.a.). In salmon farming, this process is replicated by following these three phases (Misund, s.a.). The first phase is the fertilization of the female roes when they are mixed with the milt from the males (SeafoodFromNorway, s.a.). After hatching, the fish feed on their yolk sac until it is empty, and they receive food in the form of pellets. The fish will then mature until it reaches smoltification where it is released into the ocean. The time it takes for the salmon to smoltify varies greatly (SeafoodFromNorway, s.a.). Due to selective breeding and controlled environmental variables such as feed formulation and water temperature, the salmon normally reaches smoltification in about 16 months (Khaw et al., 2021). However, earlier smoltification can be produced with the use of photoperiod and temperature manipulation to promote out of season smolts 7-8 months after hatching, although this has shown to have the possibility of a negative effect on survival and growth during the saltwater life cycle (Sigholt et al., 1998). The fish is then kept in the farms to mature until it reaches desired size.

In the wild, Atlantic salmon are carnivorous feeders that eat other fish species, fish larvae and planktonic crustaceans (Rikardsen & Dempson, 2010). Farmed salmon however eats mainly dry pellets that very often contains a mixture of raw vegetable material, like vegetable protein and carbohydrates and raw marine material, like fish oil and fish meal (SeafoodFromNorway, s.a.). The theme of fish feed is a huge part of the aquaculture industry. Creating the mix between sustainable, cheap and efficient feed while making sure high fish health and growth is maintained is a consistent issue for the industry. One area of study is the ability to promote beneficial changes in the farmed salmon through feed modulation. An example of this are prebiotics, that are non-digestible nutrients that exist to promote the growth of beneficial bacteria in the gut (Ringo et al., 2010). This idea behind feed affecting the bacterial composition in hosts has existed since the early 1900s, long before the concept had been named (Cheplin & Rettger,

1920). It was in 1995 when Glenn Gibson and Marcel Roberfroid released their study: "Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics". Since then, the term prebiotics has become a widely prevalent term in the scientific community, regardless of the controversies behind its naming (Hutkins et al., 2015). However, even though modulating microbiomes is useful, it is only when you understand the communities and how they affect their hosts that feed modulation becomes a valuable tool, especially in the field of aquaculture.

Norway is an ideal country for aquaculture due to its large coastline and stable cold temperatures with its position to the Atlantic Ocean (Andreassen & Martinussen, 2011), which allows it to be one of the largest producers of farmed Atlantic salmon (FAO, 2021). This industry has had a massive growth and reached finally reached a worth of 100 billion NOK in 2018, around 70% of this coming from salmonoids (i.e., Salmon and Rainbow trout) (Raudstein, 2020) (Richardsen et al., 2019). This massive industry has been a great boon for food production; however, it is often plagued by a great deal of issues. One of these issues are diseases caused by pathogens (e.g., bacteria, viruses and fungi), which have negative effects on fish welfare. As such, there is a lot of value in learning in what ways we can affect the microbial composition through feed modulation, an example being Soybean protein concentrate (SCP), that has been shown to change the microbial composition (Green et al., 2013).

Therefore, the more research is done into this area, the more valuable knowledge we can learn which increases the chances of creating a better, more sustainable food production industry.

1.2 Sequencing technologies – the culturomics approach

The evolution of DNA sequencing has had a massive growth in the 21st century as the technologies has been improved upon as the knowledge of the microbial scientific community grows. This is especially prominent in the field of metagenomics which has brought incredible insight into the huge microbiomes previously untouchable by old culture-based techniques (Lagier et al., 2016). Whilst culture-dependent methods are the bedrock of microbiology, this approach often fails to identify most of the microbial organisms present in the microbiota as most of the sequences come from uncultured bacteria (Lagier et al., 2012). This means that there exists "microbial dark matter" in every study performed, causing most of the information present to us to be un-identifiable (Rinke et al., 2013). For the study of the gut microbiome, metagenomics has revolutionized what we know about its relations to diseases and host health

(Lagier et al., 2016). However, with metagenomics it is difficult to ascertain detailed phenotypic knowledge of microorganisms, hence new culture-based approaches have gained increased prominence in the last decade to complement metagenomics and its inherent limitations (Diakite et al., 2020). Culturomics aims at identifying the bacterial species that were thought to once be unculturable by older -omics based techniques (Rinke et al., 2013). This approach is usually performed through two methods: MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometry and 16s rRNA amplification sequencing (Dubourg et al., 2013). The latter method will be the focus of this thesis by combining the identification of 16s amplicons with Sanger sequencing and long read Oxford Nanopore Sequencing.

The history of sequencing technologies very often starts with Sanger sequencing in 1977 which was developed by Frederick Sanger and his colleagues (Sanger et al., 1977). The main principle of the technology was the chain termination method, which allowed for different sized reads to be created randomly with the introduction of dideoxynucleoside trisphosphates (ddNTPs), which terminated the DNA replication process due to a lack of the 3' hydroxyl group. By running the reactions in 4 wells with radiolabeled ddNTPs, the DNA could be sequenced (Sanger et al., 1977). This was the start, and over decades this technology was improved to increase its efficiency, most notably the introduction of fluorescent labeling (Heather & Chain, 2016), however other technologies were also being developed alongside it. This ended up with the technology getting for the most part overtaken in efficiency and therefore popularity. On the NCBI RefSeq (Reference Sequence Database) sequences produced by Sanger sequencing constitutes barely 1% of the collected sequences while Illumina dominates with 82% (Segerman, 2020). Illumina is the state-of-the-art sequencing technology for short reads due to its low cost and high-quality output. This technology utilizes clustering on flow cells to create amplification on the DNA input which is then sequenced using SBS (Sequencing by Synthesis), as it tracks the addition of dNTPs containing fluorescently marked terminators that gets cleaved off allowing for further synthesis (Illumina, 2017). This parallel sequencing method is what allows Illumina technologies to generate high yield with low error rates (Illumina, 2017).

Although Illumina usually trumps over Sanger sequencing in terms of both effectiveness and popularity, the use of Sanger is still a viable option to use for bacterial classification. Sanger sequencing specializes in targeting specific genes with many samples as input due to its ~800

nucleotides length output with high quality (CD Genomics, s.a.). This makes it a great technology for identification of bacteria through the 16S rRNA gene.

The 16S rRNA gene has become one of the hallmark markers for phylogenetic identification due to its conserved features (Janda & Abbot, 2007). The gene is present in all prokaryotic organisms, making it a universal target for bacterial identification. Secondly, the function of the gene has remained constant over time, meaning that mutations in the sequence are more likely to reflect random changes than selective changes. Lastly, the length of the gene at around 1 500 base pairs is significant as it is long enough to provide statistically relevant information. In the length of the gene exists around fifty functional groups, and therefore selective changes in one group will not be enough to affect phylogenetic relationships (Patel, 2001). These factors are some reasons as to why 16S rRNA is so powerful. However, the approach does also have disadvantages, to note its low resolution making it less unfit technology for species level identification, and its inability to detect viruses and fungi due to their lack of the marker gene (Janda & Abbot, 2007).

To couple with the 16S rRNA sequencing, this thesis will also incorporate Oxford Nanopore Technology (ONT) for whole genome assembly. ONT has garnered massive spotlight as a sequencing tool due to its long reads and ease of use ever since 2014 with the commercial release of the MinION sequencer (Lu et al., 2016). The core concept behind ONT of pushing molecules through a nanopore has been around since 1995 (Akeson et al., 1995), and its iterations has led to the technology we know today. ONT is based on the migration of single strained DNA or RNA across a bi-lipid membrane through a membrane protein. As the DNA migrates through the protein, the ionic current is blocked leading to a measurable change in voltage. The change in voltage will also depend on the base of the nucleosides as they travel through (Jain et al., 2016). This process allows for incredibly long reads relative to any other major sequencing technology at this current time, being for the most part dependent on the length of the input. The price for this efficiency comes in the form of quality, which cannot compete to the short read technologies (Branton et al., 2008). However, short reads have an inherent disadvantage in the assembly of whole genomes due to the difficulty in creating accurate contigs with repeating elements (Moss et al., 2020). Long reads circumvent this issue due to spanning multiple repeat elements, leading to more accurate genome assembly. This ability to recover genomes makes it a great tool for

bacterial identification, especially when combined with short read technologies like Illumina that produce more accurate base calling (Branton et al., 2008).

1.3 Thesis aim

As shown, there is a big room for further culturomics based identification of the microbiome in the gut of Atlantic salmon. As more research is performed and more uncultured bacteria are identified, metagenomics and culturomics build into each other's strengths and weaknesses, allowing for exponential growth in the study of gut microbiomes. This thesis is a work in collaboration with Improvafish, a joint European project which aims to improve the sustainability of aquaculture through by modulating feed-microbiome-host axis in fish. My study focuses on the prediction of the salmon gut microbiota composition and its function to assist in the creation of a Salmon Microbial Genome Atlas (SMGA). The study aims to make use of both sanger 16S rRNA amplification sequencing and long read Oxford Nanopore sequencing to survey the microbial composition of cultured bacterial isolates and to gain insight into some of their genetic functions.

2. METHOD

2.1 Media

The experimental samples collected for this study were initially cultivated on 5 different media: BHI (Brain-Heart Infusion), TSA (Tryptic Soy Agar), LB (Luria Bertani), MRS (De Man, Rogosa and Sharpe agar) and Vibrio. The table containing the ingredients for the media can be found in table 1:

Table 1: The 5 different media that was used for initial samplings, and its ingredients. An extra 2.5% NaCl was added in addition to the NaCl already apart of the recipe since the isolation is of bacteria from a marine environment.

Media	Ingridients per 1L agar plates
BHI	37g BHI powder + 10 g NaCl + 15 g agar powder +2.5 %NaCl
TSA	30 g TSA powder + 50 g glucose + 10 g NaCl + 15g agar powder + 2.5 % NaCl
LB	10g Tryptone + 10g NaCl + 5 g yeast extract + 2.5 % NaCl
MRS	55g MRS powder + 10g NaCl + 15g agar powder + 2.5% NaCl
Vibrium	90 g Vibrio powder + 15 g agar powder + 2.5 % NaCl

2.2 Sampling

Samples were collected from the Center for Sustainable Aquaculture at NMBU. Eight random fish were collected from freshwater tanks and consequently euthanized. The fish were sprayed with 70% ethanol prior to incision, and then cut along the midline of the fish to expose the gut of the fish. The entire intestine was aseptically removed from the fish and cut at the intersection between the distal and proximal tract using a sterile scalpel. The gut contents of the distal tract were collected using sterile loops and sampled onto six Petri dishes of each medium. Two of the petri dishes were sampled without dilution of the gut samples, while four petri dishes were sampled with 100 micrograms of samples with serial dilutions from $1x10^{-1}$ to $1x10^{-4}$. The cultures were grown for around two weeks until sufficient cultures could be seen.

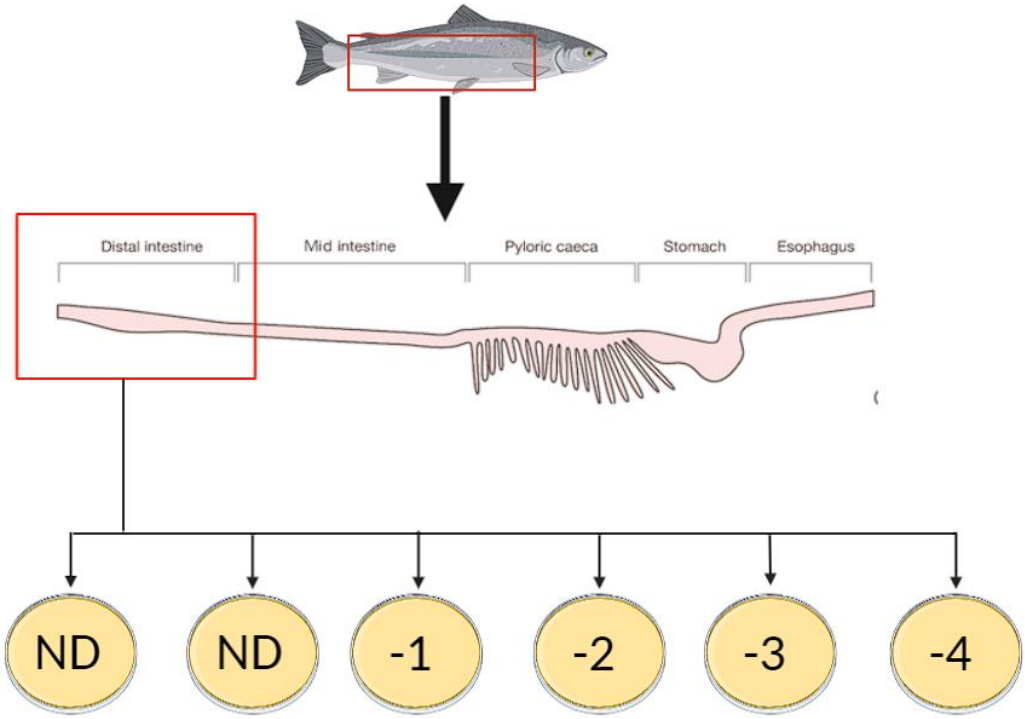


Figure 2: Flowchart showing the steps for sampling of the distal tract content. This process was performed for each of the 5 media. The serial dilutions ranged from $1x10^{-1}$ to $1x10^{-4}$. ND = No Dilution

2.3 Isolating cultures

After incubation, plates showed sufficient colonies on the BHI and TSA media plates, and as such decided to proceed with lab work on these two mediums. Single colonies from the $1x10^{-1}$ and $1x10^{-2}$ dilution were re-streaked onto new plates, as these plates showed separated cultures with limited fungi growth. These new plates were then incubated at 15 °C for 1 day, and the re-streaking and incubation steps were repeated.

From this process, 34 isolates from BHI and 22 isolates from TSA media were selected for DNA extraction and amplification of the 16S rRNA gene using polymerase chain reaction (PCR).

2.4 Preparation of bacterial collection frozen stock

Freeze stocks of bacterial isolates were prepared from liquid cultures. Tubes with 4 ml BHI or TSA were inoculated with a single colony from the respective plate, and cultures were incubated at 15 °C for a day. 765µl of the overnight cultures were added to cryotubes together with 235µl of sterile 85% glycerol. tubes stored at -80 °C.

2.5 16S rDNA amplification by colony polymerase chain reaction (PCR)

Amplification of the 16S rDNA gene on each isolate colony was performed using the 27F (5'-AGAGTTTGATCCTGGCTCAG-3) and 1492R (5'-GGTTACCTTGTTACGACTT-3') primers. Colonies were sampled in 20 µl distilled water.

12.5 µl Q5 Hot Start High-Fidelity 2X Master mix (New England BioLabs$_{inc}$), 1.25 µl Forward- and reverse primer, 8 µl nuclease-free water, and 2 µl from the template DNA was added to a 1.5 ml Eppendorf tube to a total volume of 25 µl. The Tubes were placed in an Eppendorf Mastercycler Gradient Thermal Cycler/Senseoquest Labcycler. The PCR reaction started with the initial denaturation at 98 °C for 2 minutes, followed by 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 55 °C for 30 seconds and then extension at 72 °C for 50 seconds. The final extension step was set at 72 °C for 2 minutes, and the tubes were then held in the machine at 4 °C overnight. The expected size of the PCR product was ~3 kb since the size of the 16S gene is ~ 1.5 kb and the samples contains both forward and reverse reads.

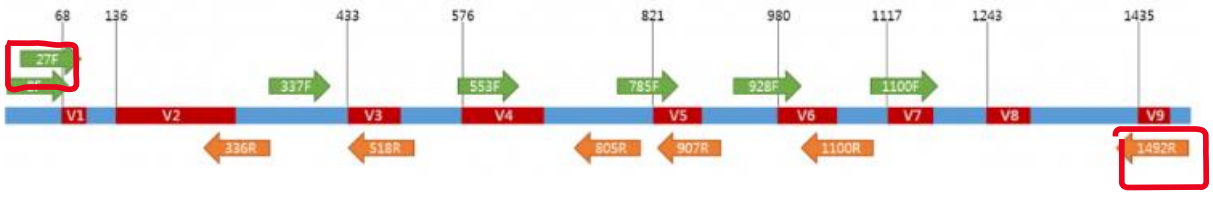


Figure 3: Description of the 16S rRNA gene. Consists of 10 conserved (blue) and 9 variable (red) regions. Also shows the range of 27F and 1492R primers.

2.6 Gel electrophoresis

Gel electrophoresis was used to confirm the presence of the PCR products. Gels were made by mixing 50 ml 1x tris-acetate-EDTA (TAE) and 1g agarose powder in an Erlenmeyer flask. The mixture was heated in a microwave until becoming transparent. When sufficiently cooled, 2.5 µl Peq-green was added for DNA detection using UV illumination. 4 µl PCR product was loaded in the gel with 6 µl distilled water and 2 µl #R0611 6X DNA loading dye, and then ran at 100 volts until sample had traveled for around 3 cm. The gel was then visualized using BIO-RAD Gel doc EZ Imager.

The samples that showed clear presence of PCR product were stored, while negative results were repeated through a new PCR. With the information gathered from the gels, 34 BHI and 22 TSA isolates were chosen for 16S rRNA sequencing.

2.7 NucleoSpin Gel and PCR Clean-up Kit

Preparation of the samples for sanger sequencing was done through the Nucleospin gel and PCR clean-up kit (MACHEREY-NAGEL), to remove primers, buffers, enzymes and any residual dNTPs from the PCR reaction. 42 µl PCR product was mixed with 84 µl NTI buffer. The mixture was added to the Nucleospin Clean-up column, which was placed into a 2 ml collection tube. The tube was then centrifuged for 1 minute at 11 000 x G, and flow through was discarded. 700 µl of buffer NT3 was added to the column and centrifuged again at same levels, and flow through was again discarded. This step was performed twice. The tube was then centrifuged at 11 000 x G for 2 minutes for complete removal of buffer NT3. The column was removed and placed into a 1.5 ml microcentrifuge tube, and 25 µl elution EB was added before incubating at room temperature for 1 minute. After incubation, the tube was centrifuged at 11 000 x G for 1 minute. The flow through containing the clean DNA was then kept and stored in a cooler ready to be sent to Eurofins for Sanger sequencing using the ABI 3730XL sequencing machine.

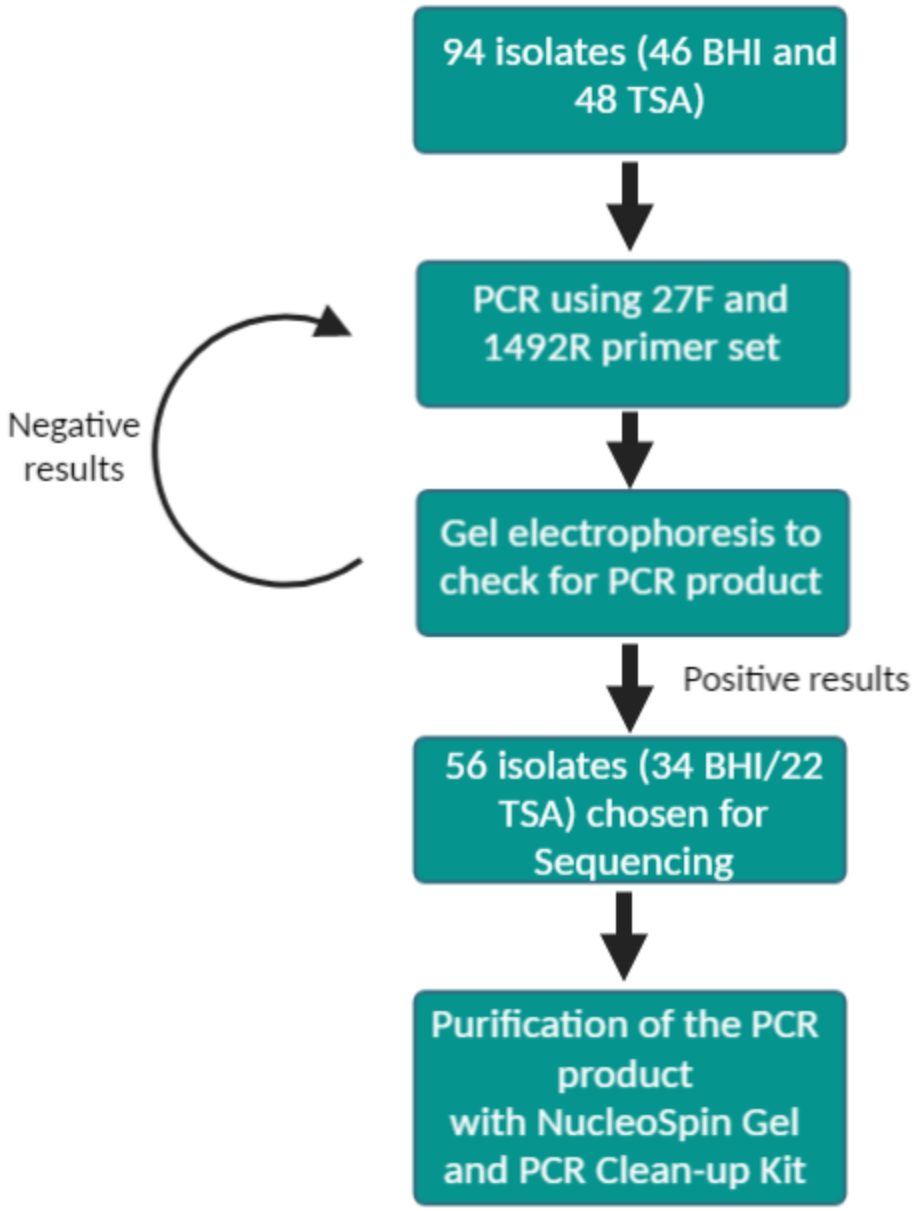


Figure 4: Culture isolation and processing flowchart. Positive and negative results refer to the presence of PCR product in the gel electrophoresis performed.

2.8 Sanger sequencing result processing

The DNA sequence of 16S DNA genes were analyzed and processed using BioEdit v7.2, a program for sequence alignment and analysis. Contigs were assembled using the forward and reverse strand with CAP v1.0 (Contig Assembly Program), an accessory program within BioEdit. The resulting assembly was then compared to the NCBI database using BLASTN with an E-value cutoff at 0.05, and the hit with the lowest E-value was chosen as the closest relative for the isolate. Phylogenetic trees were built using Clustal Omega v1.2.4, a web based multiple sequence alignment tool. The trees produced were processed with FigTree v1.4.4, a program for phylogenetic tree visualization.

Using the Sanger result as basis, the 56 samples were polished by discarding 9 of the repeating samples and the Nanopore sequencing was to proceed with the remaining 45 samples.

2.9 Extraction of high molecular weight DNA for ONT sequencing

To extract high molecular weight DNA from liquid cultures, a Nanobind CBB Big DNA kit (CIRCULOMICS) was performed using the gram-positive bacteria protocol. The samples were centrifuged at 16000 x G at 4 °C for 1 minute, and the supernatant was removed. 20 µl of 1x PBS was added and mixed with a P200 pipette to resuspend the cells. 100 µl of STET Buffer with added lysozyme was added and vortexed for 1 second 10 times, before incubated at 37 °C for 30 minutes. After incubation, 20 µl of proteinase K was added and the sample was vortexed before incubated at 55 °C and 900 rpm for 10 minutes. Optional RNA removal was performed by adding 20 µl RNase A and incubation for at room temperature for 3 minutes.
After RNA removal, 100 µl of Buffer BL3 is added to the sample and then pulse vortexed before another incubation at 55 °C at 900 rpm for 10 min. At this point the nanobind disks are added to the cell lysate together with 400 µl isopropanol, inversion mix 5 times then mix on a tube rotator for 10 minutes at 9 rpm. The tubes were then placed on the magnetic rack, which allows for removal of the supernatant. The tubes were mixed after adding 700 µl of buffer CW1 and then re-placed on the magnetic rack for removal of the supernatant. This same procedure is repeated twice using buffer CW2. The tubes are then centrifuged, and any residual liquid is removed. 50 µl of buffer EB is added to the samples and incubated at room temperature for 10 mins to

resuspend the DNA on the disk to the liquid. After incubation, the eluate is transferred to a new 1.5 ml microcentrifuge tube, making sure to centrifuge the nanobind disk tube to transfer all the liquid. The tubes containing the extracted DNA is then mixed and stored overnight in a cooler. The DNA concentration of the samples were measured using Nanodrop, and the values can be found in table S2 in appendix A. Gel electrophoresis was performed with the same procedure as the PCR product gels, with the exception that they ran for longer to make sure the HMW DNA was able to travel sufficiently.

2.10 Short read elimination

The XS SRE protocol was performed on 15 of the samples to remove reads shorter than 10 kb in length to increase the sequencing quality, as the gel electrophoresis performed after DNA extraction showed an abundance of shorter length DNA strands in the samples. The concentration of all samples was checked using Qubit dsDNA Broad Range Assay. 60 µl sample was added together with 60 µl buffer SRE XS in a 1.5 ml Eppendorf DNA LoBind tube and mixed carefully. The tube was centrifuged at 10 000 x G for 30 minutes at room temperature, and the supernatant was carefully removed while not disturbing the invisible DNA pellet. 700 µl 70% EtOH was added to the sample and centrifuged for 2 minutes, and the wash solution was removed. 50 µl buffer EB was then added to the tube and the sample was stored in a cooler. The final concentration was then checked with Qubit, and gel electrophoresis was performed to check if the SRE kit has any effect on the samples. The measured DNA concentration of the 15 isolates can be found in table S3. in appendix A.

2.11 Nanopore Ligation and library preparation

The Extracted DNA samples for each isolate were prepared for Nanopore sequencing following the Ligation Sequencing Kit (SQK-LSK109) with the Native Barcoding Expansion protocol (EXP-NBD196).

The first step was DNA repair and end-prep. The concentration of the 45 DNA samples was measured using qubit and adjusted to 40 µg/ml. 400 ng DNA was then allocated to a 96-well

plate and then added nuclease-free water to a total of 12 µl together with 3 control samples for a total of 48 samples. Each of the wells were then added 0.875 µl of NEBNext FFPE DNA Repair Buffer and Ultra II End-prep reaction buffer, 0.75 µl Ultra II End-prep enzyme mix and 0.50 µl NEBNext FFPE DNA repair mix. The plate was then incubated at 20 °C for 5 minutes and 65°C for 5 minutes. The samples were saved for the native barcode ligation.

A single barcode was chosen for each sample. In a new 96-well plate, the following reagents were added to each sample: 3 µl nuclease-free water, 0.75 µl end-prepped DNA, 1.25 µl Native barcode and 5 µl BLUNT/TA Ligase Master Mix. After mixing by pipetting, the samples were incubated at 20 °C for 20 minutes in a thermal cycler. After incubation, 1 µl of 0.5 M EDTA was added to each well and mixed thoroughly. After this, all libraries were pooled together and the total of 480 µl were carried forward. AMPure XP beads were resuspended by vortexing, and 384 µl of these beads were added to the library and mixed on a hula mixer for 10 minutes at room temperature. The beads were then pelleted using a magnetic rack for 5 minutes and the supernatant was removed. The beads are then washed using 700 µl 70% ethanol without disturbing the pellets on the rack, and the ethanol was then removed. This step was performed twice. The tube was then centrifuged and replaced on the magnetic rack to remove any residual ethanol, allowing the pellets to dry for around 30 seconds. After removing from the rack. the pellets were re-suspended in 35 µl nuclease free water and incubated for 10 minutes at 37 °C. By placing the tube back on the magnetic rack to pellet the beads, 35 µl of the eluate were transferred to a new 1.5 ml Eppendorf DNA LoBinding tube.

Following barcode ligation, adapter ligation and cleanup was performed on the eluate. In a fresh 1.5 ml Eppendorf DNA LoBinding tube the following reagents were added: 30 µl pooled barcoded sample, 5µl Adapter Mix II (AmII), 10 µl NEBNext Quick Ligation Reaction Buffer (5X) and 5µl Quick T4 DNA Ligase. After mixing and spinning down, the tube was incubated at room temperatures for 10 minutes. From this point the process is the same as the barcode ligation process from the addition of the AMPure XP beads, with the exception that the beads are washed using 125 µl long fragment buffer instead of the ethanol. After drying, the beads were suspended in 15 µl Elution buffer and incubated at 37 °C for 10 minutes. After pelleting the beads on the magnetic rack, 15 µl of the eluate can be transferred to a 1.5 ml Eppendorf DNA LoBinding

tube. The library is now prepared and ready to be primed and loaded on the Promethion flow cell following the manufacturer's instructions.

2.12 Promethion data processing

The data recovered from the Promethion sequencer had already been base called using Guppy, which converts the voltage data from the Promethion in FAST5 format to FASTQ format. The quality of the reads was then analyzed using Nanoplot v1.33.0 (parameters: --fastq –N50 –loglength) (De Coster et al., 2018) which is a plotting tool for sequence data. Nanoplot was also performed on filtered data, which was created using Filtlong v0.2.1 (parameters: --min_length 1000 –keep_percent 90) (Wick & Menzel, 2017), which filters out shorter reads by quality. With the current filtered reads, we assembled the genomes using Flye v2.9 (parameters: --nano-raw) (Kolmogorov et al., 2019).

The completeness of the assembled genomes was evaluated with BUSCO v1.0 (Benchmarking Universal Single-Copy-Orthologs) (parameters: -m geno --auto-lineage-prok) (Simão et al., 2015). BUSCO provides a measure of the assembly completeness through gene content. When running BUSCO, you have the option to choose the database to compare your assembled genomes to, or let BUSCO automatically detect the most appropriate database for your genome. This means that BUSCO will assign its genomic input to a taxonomic group according to its predicted gene contents. Although not specifically performed for taxonomic classification, the database assigned by the program will be noted to compare to the other tools used for taxonomic classification of the isolates.

One of the major issues with Long read technologies like ONT, is the accompanying high error rate (>5%) (Vaser et al., 2017). As such, the genomes are usually modified through post-assembly polishing. The two tools used in this thesis will be Racon (Rapid Consensus) (Vaser et al., 2016) and Medaka (Rescheneder et al., 2016). Racon is a read-to-assembly polishing method, as it compares the draft assembly to an input of mapped reads (Huang et al., 2021). Minimap2 v2.24 (Li, 2018) was used as a mapping tool with default parameters, which finds overlap of the ONT reads. Medaka is another polishing tool and is developed by ONT. This is a neural network tool, which tries to calculate the correct bases in an assembly through multiple algorithms that

recognize relationships (Huang et al., 2021). Processing with these tools aims to increase the quality of the Flye assembled genome.

Polishing of the genome assembly was performed with 2 rounds of Racon v1.4.3 with default parameters, and then further polishing with Medaka v1.6.0 (parameters: -g r941_prom_high_g4011). BUSCO was then performed again on the polished assemblies, to evaluate is the polishing influenced genome completeness.

With the polished medaka assemblies, taxonomic classification was recovered through MiGA (Microbial Genome Atlas) (Rodriguez et al., 2018). This web-based tool is a data management and processing of microbial genomes and metagenomes. It provides taxonomic classification based of AAI (Average Amino acid Identity) and ANI (Average Nucleotide Identity) values that compare the assembled genome to reference genomes present in the NCBI prokaryote genome database.

To complete the analysis, DRAM v1.3 (parameters: --trans_table 11) (Shaffer et al., 2020) was performed to annotate genes and claim prediction of CAZymes (Carbohydrate active enzymes), which are enzymes with the ability to degrade saccharides. DRAM (Distilled and Refined Annotation of metabolism) is a computational framework tool for decoding the metabolic functions of microbiomes. The genomic or metagenomic input has its genes predicted with Prodigal, and annotation is performed through searching the predicted genes with multiple genome databases like for example DBcan2 and KEGG that provide the metabolic functions of the microbiota.

All custom scripts and the commands to run them can be found in Appendix A

3. RESULTS

3.1 Preliminary lab work results

This section combines the results obtained during the isolate processing. Figure 6 and 7 shows the colony growth in the BHI and TSA plates after ~1 week of incubation. The pictures indicate the increased colony growth in the less diluted samples, while also showing the emergence of fungi growth in some TSA plates. The first gel picture shows the result of the SRE kit on TSA isolate 35 and 36 and confirms its effectiveness in removing shorter DNA fragments, allowing it to be executed on the rest of the isolates. Figure 8 shows gel electrophoresis performed after PCR. The gels confirm the presence of ~3 kb sized PCR products in the positive results. Figure 9 shows the gel electrophoresis results performed after DNA extraction using the Nanobind protocol. The gels show that there is a lot of high molecular weight DNA in the samples, with the DNA being mostly above 10 kb. However, there are also a lot of shorter strains that contaminate the samples.

Figure 5: Example Gel electrophoresis of SRE product of samples TSA 35 and 36. These two samples were performed separately to evaluate if the SRE kit would be effective, and the rest of the samples were performed due to positive results.

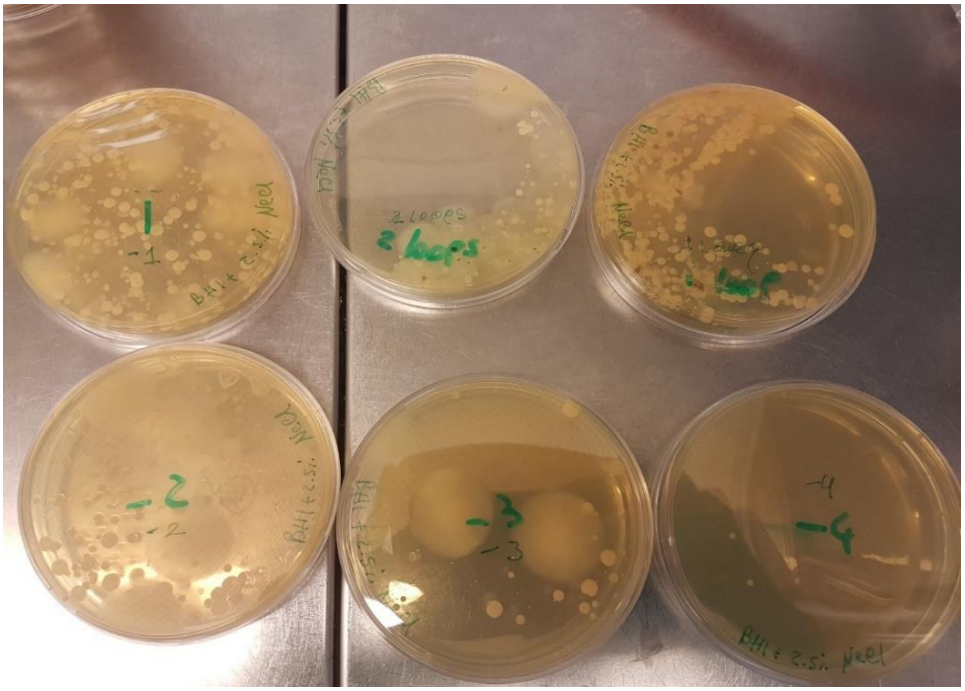


Figure 6: BHI culture growth after 1-2 weeks of incubation. The plates show in order from top left: -1 dilution, plate 2 with no dilution, plate 1 with no dilution, -2 dilution, -3 dilution and -4 dilution.

Figure 7: TSA culture growth after 1-2 weeks of incubation. The plates show in order from top left: -4 dilution, plate 2 with no dilution, plate 1 with no dilution, -1 dilution, -2 dilution, -3 dilution

Figure 8: Gel electrophoresis of PCR results for BHI and TSA isolates. 1st well contains 1kb HMW DNA ladder N3232S. K represent the control sample. ***A)*** *Isolate BHI 1-13.* ***B)*** *Isolate BHI 15-28.* ***C)*** *Isolate BHI 29 – 46.* ***J)*** *Isolate TSA 1-12.* ***K)*** *Isolate TSA 17-28 + Repeat for 1 and 2.* ***L)*** *Isolate TSA 35 – 46 + repeat for sample 17 and 18. Gel D - I are repeats of the negative BHI results in the initial gels or repeats due to sufficient contamination in control well. All gels have a corresponding control; however, some were conducted in pairs. Control A correspond to gel B, C to D and G to H.*

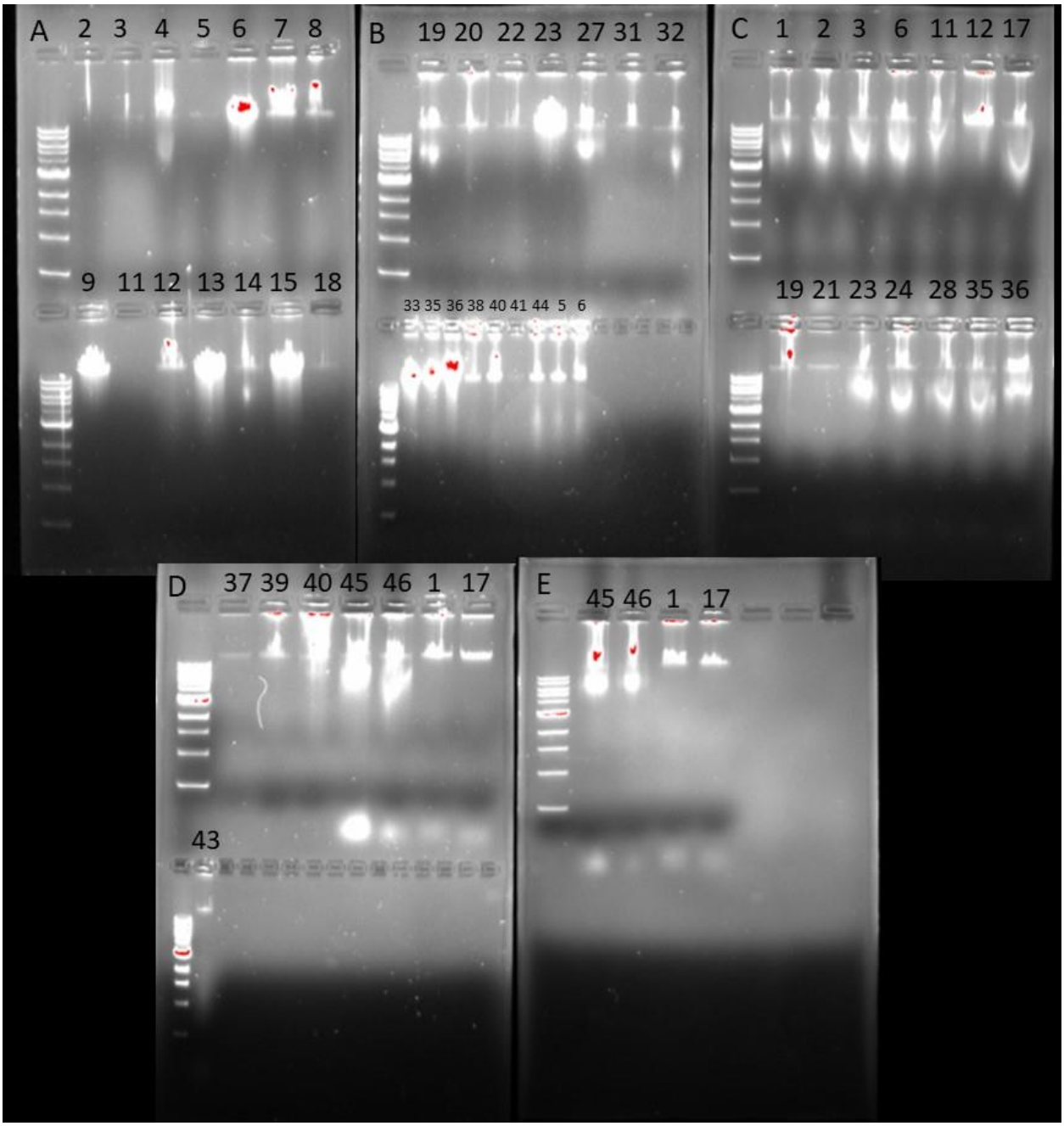


Figure 9: Gel electrophoresis of HMW DNA extraction for BHI and TSA isolates. 1st well contains 1kb HMW DNA ladder N3232S. ***A)*** *BHI isolate 2 – 18.* ***B)*** *BHI isolate 19 – 44 + TSA 5 and 6.* ***C)*** *TSA isolate 1 – 36.* ***D)*** *TSA isolate 37 – 46 + BHI 1, 17 and 43.* ***E)*** *Repeat run of TSA isolates 45 and 46 + BHI isolates 1 and 17 due to signs of high RNA contamination*

3.2 16S rRNA Sanger sequencing data

With the sequenced reads provided by Eurofins, the assembled contigs of the 16S rRNA genes were blasted to provide taxonomic classification with an average homology of ~96%. The full table containing all the samples query length and homology value can be found in table S3 in appendix B.

Using the assemblies built in BioEdit, a phylogenetic tree was constructed with Figtree which can be viewed in figure 10. This tree contains our 56 isolates, but also additional isolates provided by the currently incomplete SMGA database. These isolates, provided by Arturo Vera_Ponce_De_Leon, were isolated from Atlantic salmon taken from saltwater, and so these were added to the tree to compare the relatedness of the freshwater and saltwater salmon.

The taxonomic classification collected from Sanger sequencing was assigned using BLASTN on the NCBI database. Figure 15 shows the phylum and genus level classification, but species and strain level information can be found in table 2 from the main text and table S3 in Appendix B. Most of the hits show only one unique result, however strains like *Lelliottia_amnigena_strain_pp3* (17%), *Enterobacter_sp_KAR3* (15%) and *Lelliottia_amnigena_strain_JM121* (10%) was identified in multiple samples.

On phylum level, the BLASTN results assigned sequenced reads into the three phylums *Pseudomonadota* (69%), *Actinomycetota* (8%), *Bacillota* (4%) and assigned nine uncultured isolates (19%).

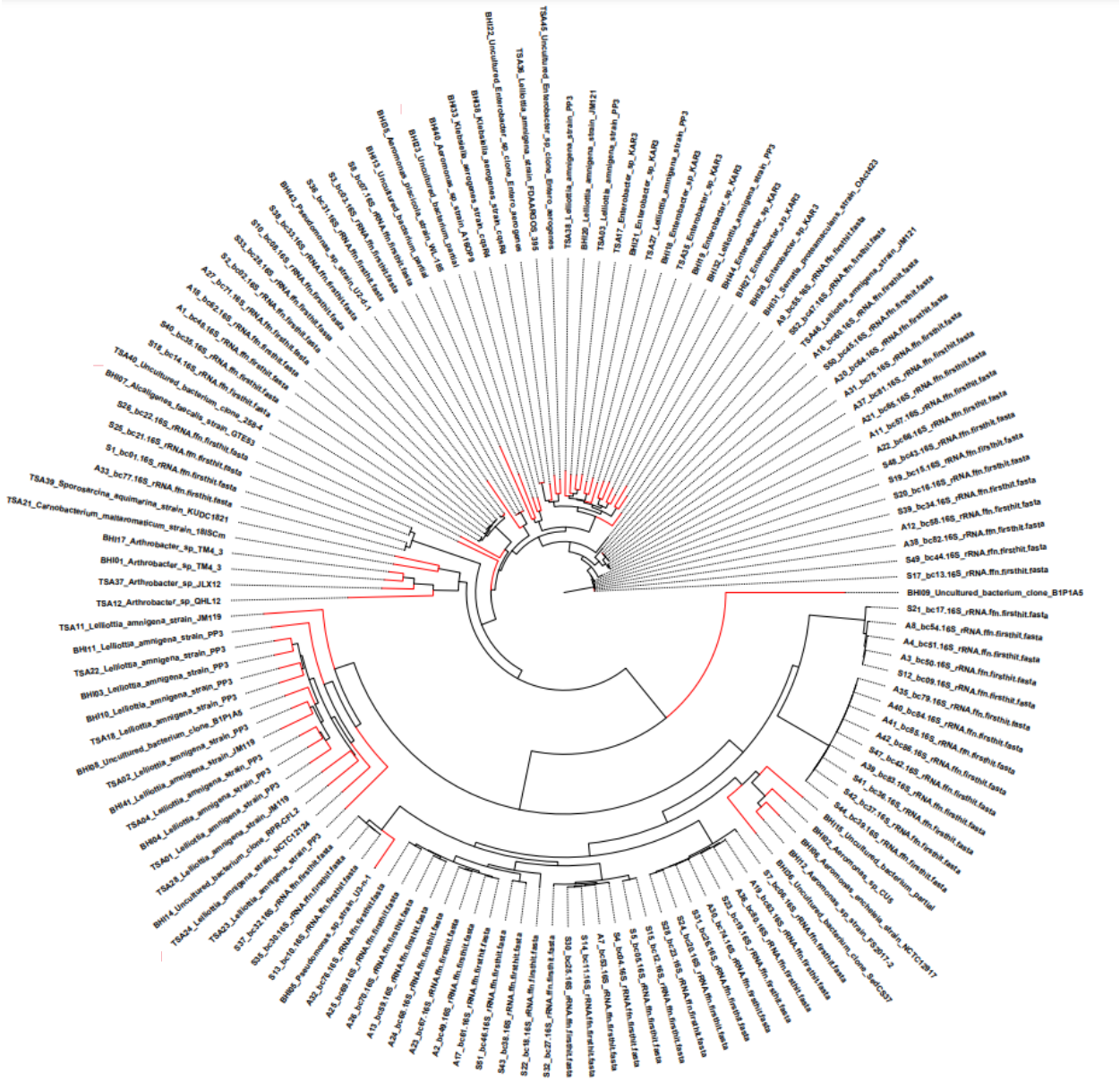


Figure 10: Phylogenetic tree constructed with Clustal Omega using 16S rRNA sequences obtained with Sanger sequencing. Contains both samples collected from salmon in freshwater tanks and samples collected from saltwater tanks, retrieved from the Salmon Microbial Genome Atlas v1 (in preparation). The freshwater samples are marked in red. A PDF file of this tree can be found in the appendix.

3.3 Promethion Sequencing Data

The Promethion sequencer produced a total of 3.01 million reads with a total of 26.33 Gb passed bases. However, only half of the assigned barcodes were for the isolates sampled for this thesis. Our 48 samples resulted in 1.89 million reads. The average Phred quality score was given as 12.4, which represents an approximate 5.8% error rate. The reads were filtered using Filtlong, which provided polishing of the reads. After Filtlong, 1.5 million reads remained with a new average quality of 12.9 or 5.1 % error rate. All plots showing the Nanoplots results can be found in the appendix. The samples are named using both their sample and barcode number, and the table showing their respective isolate is found in table S4 in the appendix.

For easier categorization of the different samples, they were divided into two groups depending on their read length: H20K and L20K. H20K refers to the isolate that produced >20 000 reads for the ONT sequence run. L20K refers to isolates with <20 000 reads.

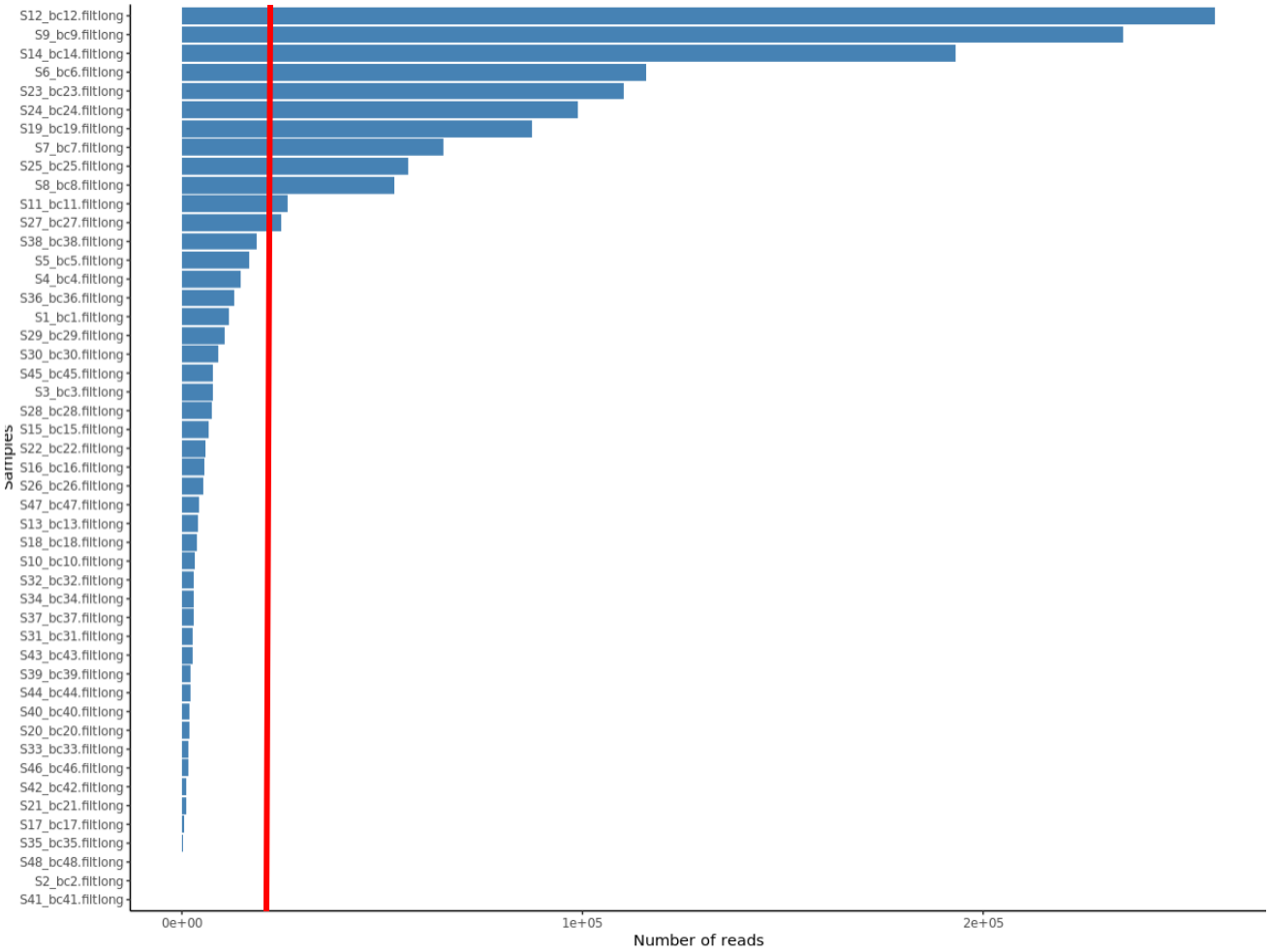


Figure 11: Table of sequenced read length. The red line shows the cutoff for higher quality isolates. The samples showing a length higher than 20 000 bases were S12_bc12 to S27_bc27.

Contig Assembly with Flye:

Flye was used to assemble contigs using the filtered Nanopore reads. The table containing the resulting genome size, N50 and contigs assembled can be found in table S4 in the appendix. Sample S35_bc35 did not provide any results due to an unknown error, and so this sample is therefore discarded for the rest of the processing.

Gene completeness with BUSCO:

BUSCO produced the result in table S4 in the appendix. To note is that sample S17_bc17 failed to find any matching due to an unknown error and was as such discarded, leaving a total of 43 samples. BUSCO assigns two values for genome completeness, one being the assessment with the domain-level dataset (generic) and the other being on the lowest taxa database assigned automatically by BUSCO (specified). BUSCO calculated an average completeness of 74% with the generic database and 69% completeness for the specified database. These values included the contigs assembled using both the high and short read length, so only counting the 12 higher quality isolates showed an average of 81% for the generic and 80% for the specified.

Polishing with Racon and Medaka:

Polishing with both Racon and Medaka produced the result shown in table 1.4 in the appendix. There are slight changes in the genome size and N50 values, but most notable are the improvements to the BUSCO completeness.. Both the generic and specified values see an improvement in genome completeness after polishing. The generic value increases from 74% to 83%, while the specified value sees an increase from 69% to 77%. Only counting the H20K samples, the completeness reaches an average of 98.7% in both categories.

Gene annotation with DRAM:

To complete the analysis of the Promethion sequence data, DRAM was performed to obtain information about the functional properties of the recovered bacteria. DRAM automatically produces a heatmap in the output that contains the metabolic functions of the annotated isolates. The resulting annotations can be found in figure 12, 13 and 14.

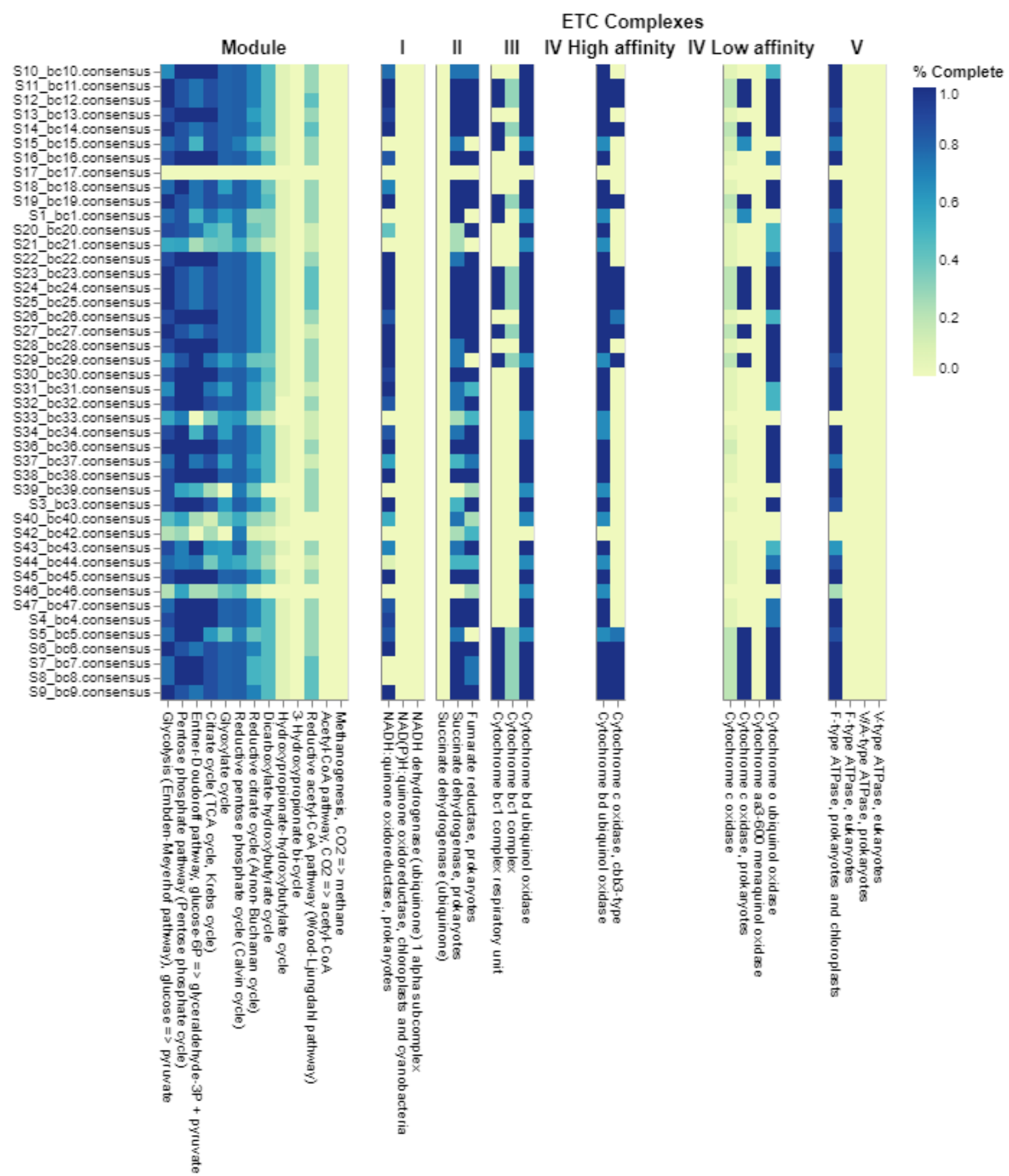


Figure 12: Heatmap of Modules and the ETC (Electron Transport Chain) Complexes provided using DRAM.

Figure 13: DRAM produced figure showing the presence of genetic functions detected for annotation. Contains the map over the CAZy (Carbohydrate Active Enzymes), Nitrogen metabolism, sulfur metabolism and other reductases.

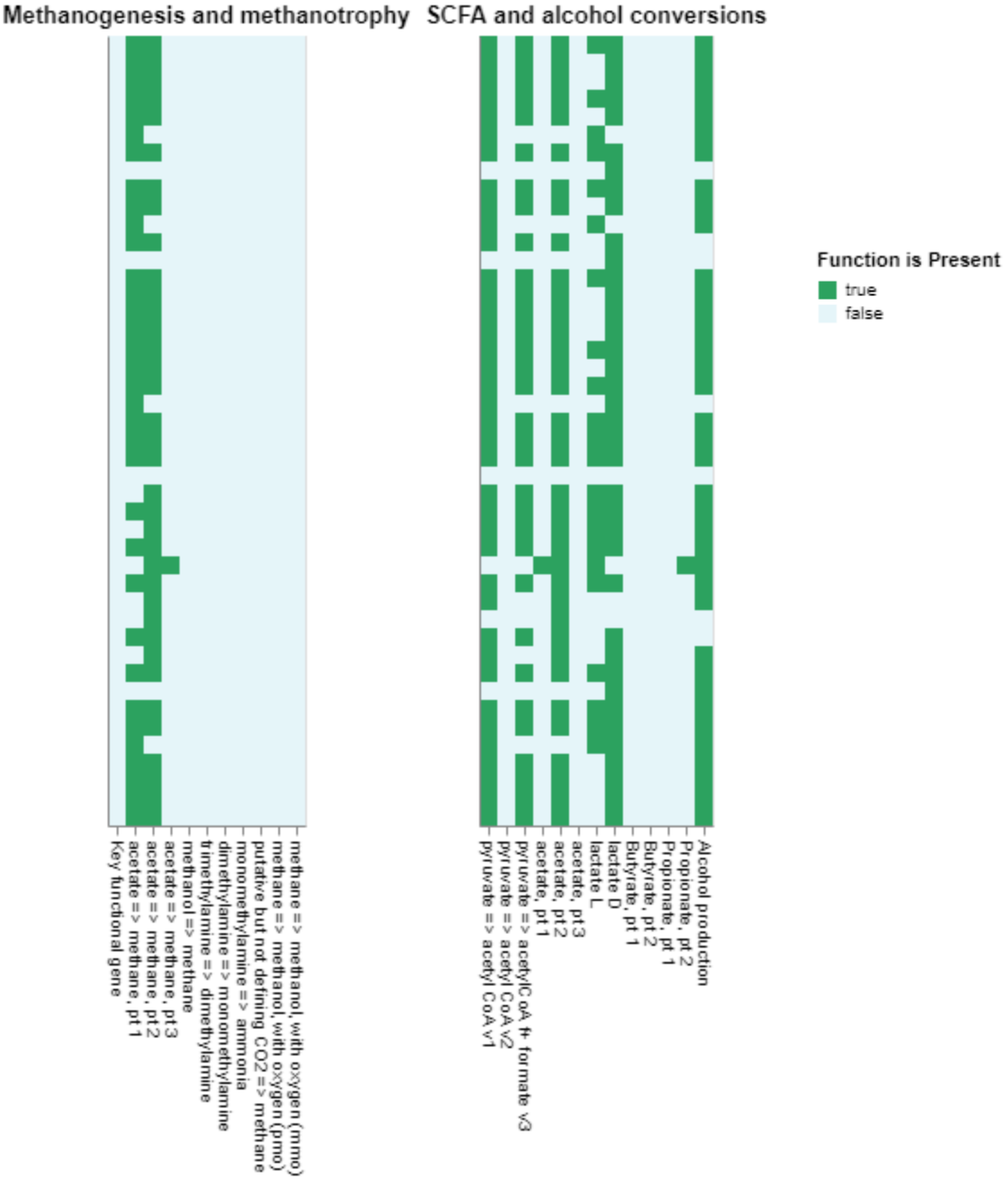


Figure 14: Continuation of figure 13. Shows the genetic presence of methanogenesis and methanotrophy + the short chain fatty acids and alcohol conversions.

Figure 12 shows the completion analysis of electron transport chain complexes and pathways, and a summary of the carbon metabolic pathways. DRAM uses functional marker genes to predicts the pathway coverage using the KEGG database. The darker the blue color, the more pathways are predicted for the respective modules. DRAM confirms that fermentation through glycolysis seems like the main source of energy production, with pathways like glycolysis, TCA cycle and the pentose phosphate pathway being prevalent in all the isolates. Isolates that show deviations from this with visibly lower completeness, like S46_bc46 and S42_bc42, is most likely due to low quality genomes (i.e., >20% completeness).

For the carbohydrate active enzymes, there are some notably predicted results. There is consistent predicted presence of genes that encode gene functions associated with polyphenol, pectin and chitin degradation. Genes encoding for predicted starch degrading enzymes can be found in four samples, and the presence of alpha- and beta-mannan degrading functionality can also be detected in multiple isolates. Nitrogen metabolism was predicted mainly in the form of genes encoding elements of the nitric oxide pathway, with the presence of nitrogen reductases and oxidases. Sulfur metabolism was also predicted for 7 isolates, with all enzymes functions being connected to the ion Thiosulfate and the metabolites it transforms to. Interesting results in the other reductases show the presence of TMAO (Trimethylamine N-oxide) reductase in S45_bc45, S38_bc38 and S36_bc36.

Figure 14 shows the remaining DRAM results for methanogenesis and methanotophy, and short chain fatty acid and alcohol conversions. Methane function was unlikely present in any of the isolates we recovered as our observed annotations linked to this category were limited to Acetate metabolism using acetyl-CoA synthetase and acetate kinase, whereas only S39_bc39 shows the presence of phosphate acetyltransferase. Lastly, Short Chain Fatty Acid (SCFA) conversions are relatively consistent in all high-quality isolates. Genes coding for Pyruvate to acetyl-CoA, Lactate and parts of Acetate metabolism are predicted.

3.4 Taxonomic classification of bacterial gut isolates

For the Nanopore data, taxonomic classification was assigned using MiGA, however the specific database assigned when running BUSCO also assigns a classification through comparison of universal genes when it calculates the genome completeness and so this is also included for comparison. BUSCO assigned 34 (85%) *Pseudomonadota*, 1 (2.5%) *Bacillota*, 2 (5%) *Actinomycetota* and 3 (7.5%) *Mycoplasmatota*. Collection of all the taxonomic data for Sanger, BUSCO and MiGA can be found in figure 15.

MiGA classification returned a total of 35 positive results, as 9 samples returned inactive due to low quality. Of the 35 samples, MiGA identified 32 (91%) *Pseudomonadotas*, 2 (6%) *Actinomycetota* and 1 (3%) *Bacillota* on phylum level.

MiGA also assigns genus level classification with a p-value less than 0.05 for all samples except S8_bc8. When comparing the genome classification to the 16S rRNA classification, there is a dominating presence of the genus *Lelliotta* making up ~30% of the genera, with other notable hits with *Aeromonas* and *Enterobacter*.

Table 2 provides information of the novelty of the isolates according to the ANI and AAI values assigned by MiGA and the closest relative assigned by BLASTN on the 16S rRNA sequenced reads.

3.5 Alternate taxonomic classification using GTDB-k

When analyzing the taxonomic classification of isolate S8_bc8 assigned by MiGA, it seemed necessary to try one more classification tool for comparison to assess whether the results were reliable. This was performed using GTDB-k v 2.0 (Chaumeil et al., 2020). GTDB-k uses the criteria of Relative Evolutionary Divergence (RED) and ANI for identification using the Genome Taxonomy Database (GTDB). The output provided this result for sample S8_bc8:

```
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Shewanellaceae;g__Shewanella;s__Shewanella sp002966515	97.49
```

This shows the taxonomy of the closest relative of the isolate and the ANI value of 97.49%. The text file containing the rest of the classification can be found in Appendix B. Only this isolate showed any difference in classification between the two classification tools.

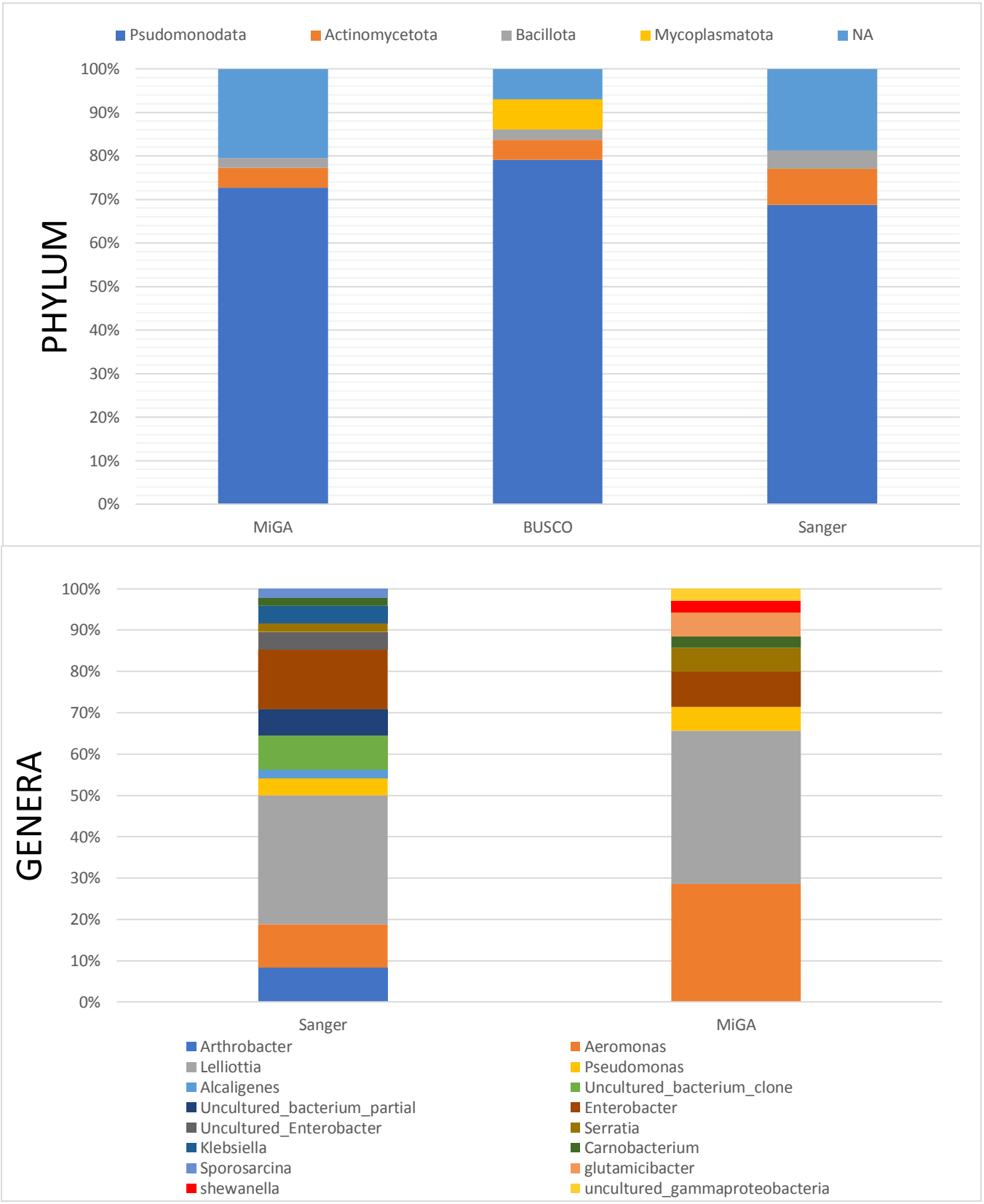


Figure 15: ***Taxonomic classification on phylum level (a) and Genus level (b). for figure a, classification was assigned by Sanger, BUSCO genome evaluation and MiGA. NA results refer to isolates with too low quality for classification. For the Sanger results, NA refer to results that identified uncultured bacteria.***
Figure b indicates results for Sanger and MiGA since BUSCO only assigns to higher taxa due to its comparison of universal gene content when assigning a database for comparison. The plots only include results with quality high enough for classification.

Table 2: Closest relative of isolate using the ANI (Average Nucleotide Identity) and AII (Average Amino acid Identity) for ONT sequenced isolates provided by MiGA, and percentage sequence identity of Sanger sequenced reads compared to closest relative assigned by BLASTN.

	ANI	AAI	Sanger classification
S1_bc1	Glutamicibacter nicotianae NZ CP033081 **(88.46%)**	Glutamicibacter arilaitensis Re117 NC 014550T **(95%)**	Arthrobacter_sp_TM4_3 **(98%)**
S3_bc3	Lelliottia amnigena NZ CP023529 **(98.78%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Lelliottia_amnigena_strain_PP3 **(97%)**
S4_bc4	Enterobacter sp. RHB15 C17 NZ CP057782 **(86.02)**	Enterobacter sp. RHB15 C17 NZ CP057782 **(95%)**	Lelliottia_amnigena_strain_PP3 **(95%)**
S5_bc5	Pseudomonas sp. BIGb0427 NZ CP062498 **(86.17%)**	Pseudomonas sp. BIGb0427 NZ CP062498 **(82.75%)**	Pseudomonas_sp_strain_U3-n-1 **(97%)**
S6_bc6	Aeromonas encheleia NZ LR134376T **(88.81%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Aeromonas_encheleia_strain_NCTC12917 **(96%)**
S7_bc7	Shewanella sp. WE21 NZ CP023019 **(97.53%)**	Shewanella baltica OS678 NC 016901 **(95%)**	Alcaligenes_faecalis_strain_GTE53 **(98%)**
S8_bc8	NA	Aeromonas sp. 2692 1 NZ CP038513 **(55.6%)**	Uncultured_bacterium_clone_B1P1A5 **(94%)**
S9_bc9	Aeromonas encheleia NZ LR134376T **(88.9%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Uncultured_bacterium_clone_B1P1A5 **(86%)**
S10_bc10	Lelliottia amnigena NZ CP023529 **(98.74%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Lelliottia_amnigena_strain_PP3 **(93%)**
S11_bc11	Aeromonas encheleia NZ LR134376T **(88.8%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Aeromonas_sp_strain_FS2017-2 **(99%)**
S12_bc12	Aeromonas media NZ CP070623 **(89.09%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Uncultured_bacterium_partial **(87%)**
S13_bc13	Lelliottia amnigena NZ CP023529 **(98.76%)**	Lelliottia amnigena NZ CP023529 **(95%)**	Uncultured_bacterium_clone_RPR-CFL2 **(95%)**
S14_bc14	Aeromonas media NZ CP070623 **(89%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Uncultured_bacterium_partial **(91%)**
S15_bc15	Glutamicibacter sp. ZJUTW NZ CP043624 **(88.42%)**	Glutamicibacter arilaitensis Re117 NC 014550T **(88.32%)**	Arthrobacter_sp_TM4_3 **(98%)**
S16_bc16	Lelliottia amnigena NZ CP023529 **(98.76%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Enterobacter_sp_KAR3 **(97%)**
S18_bc18	Lelliottia amnigena NZ CP023529 **(98.71%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Uncultured_Enterobacter_sp_clone_Entero_aerogenes **(98%)**
S19_bc19	Aeromonas encheleia NZ LR134376T **(88.78%)**	Aeromonas encheleia NZ LR134376T **(89.54%)**	Uncultured_bacterium_partial **(97%)**
S20_bc20	Lelliottia amnigena NZ CP023529 **(98.27%)**	Lelliottia amnigena NZ CP023529 **(80.93%)**	Enterobacter_sp_KAR3 **(96%)**
S22_bc22	Enterobacter sp. RHB15 C17 NZ CP057782 **(88.99%)**	Enterobacter sp. RHB15 C17 NZ CP057782 **(95%)**	Lelliottia_amnigena_strain_PP3 **(98%)**
S23_bc23	Aeromonas encheleia NZ LR134376T **(88.91%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Klebsiella_aerogenes_strain_cqsR4 **(98%)**
S24_bc24	Aeromonas media NZ CP070623 **(89.04%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Aeromonas_piscicola_strain_WL-185 **(93%)**
S25_bc25	Aeromonas encheleia NZ LR134376T **(88.73%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Uncultured_bacterium_clone_SedCS37 **(90%)**
S26_bc26	Lelliottia amnigena NZ CP023529 **(98.75%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Klebsiella_aerogenes_strain_cqsR4 **(98%)**
S27_bc27	Aeromonas encheleia NZ LR134376T **(88.79%)**	Aeromonas encheleia NZ LR134376T **(95%)**	Aeromonas_sp_strain_A16OP9 **(98%)**
S28_bc28	Enterobacter sp. RHB15 C17 NZ CP057782 **(85.89%)**	Enterobacter sp. RHB15 C17 NZ CP057782 **(95%)**	Lelliottia_amnigena_strain_JM119 **(95%)**
S29_bc29	Pseudomonas sp. R32 NZ CP019396 **(86.18%)**	Pseudomonas sp. BIGb0427 NZ CP062498 **(83.54%)**	Pseudomonas_sp_strain_U2-d-1 **(94%)**
S30_bc30	Lelliottia amnigena NZ CP023529 **(98.82%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Enterobacter_sp_KAR3 **(97%)**
S31_bc31	Lelliottia amnigena NZ CP023529 **(98.41%)**	Lelliottia amnigena NZ CP023529 **(86.79%)**	Lelliottia_amnigena_strain_PP3 **(96%)**
S32_bc32	Lelliottia amnigena NZ CP023529 **(98.72%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Lelliottia_amnigena_strain_PP3 **(95%)**
S34_bc34	Lelliottia amnigena NZ CP023529 **(98.62%)**	Enterobacter sp. 638 NC 009436 **(95%)**	NA
S37_bc37	Lelliottia amnigena NZ CP023529 **(98.58%)**	Lelliottia amnigena NZ CP023529 **(85.91%)**	Enterobacter_sp_KAR3 **(97%)**
S38_bc38	Serratia liquefaciens NZ CP014017 **(98.86%)**	Serratia grimesii NZ LT883155 **(95%)**	NA
S39_bc39	Carnobacterium maltaromaticum LMA28 NC 019425 **(98.68%)**	Carnobacterium maltaromaticum LMA28 NC 019425 **(95%)**	Carnobacterium_maltaromaticum_strain_18ISCm **(96%)**
S45_bc45	Serratia liquefaciens NZ CP014017 **(98.85%)**	Serratia grimesii NZ LT883155 **(95%)**	Sporosarcina_aquimarina_strain_KUDC1821 **(97%)**
S47_bc47	Lelliottia amnigena NZ CP023529 **(98.77%)**	Enterobacter sp. 638 NC 009436 **(95%)**	Uncultured_Enterobacter_sp_clone_Entero_aerogenes **(98%)**

4. DISCUSSION

4.1 16s rRNA analysis compared to whole genome analysis

The results obtained in this study are consistent with previous studies on the salmonoid microbial gut microbiota. In all three methods of classification, the phylum *Pseudomonadota* shows the highest number of hits in our isolates. Not counting the unclassified isolates, the *Pseudomonadotas* make up 91%, 85% and 84% of the MiGA, BUSCO and Sanger results respectively. This is quite consistent with contemporary research, which usually show the *Pseudomonadota* phylum as one of the biggest phylum's present in the salmon gut. Especially in the distal parts of the gut, as Gajardo, K. et al. found that the *Pseudomonodata* bacteria dominates in all compartments of the gut, and especially in the mid intestinal mucosa and the distal intestinal mucosa. The secondary phylums like *Bacillota* and *Actinomycetota* are much less abundant in this study, but the *Bacillota* phylum together with *Pseudomonadota* are known to generally show the highest abundance in studies of the gut (Egerton et al., 2018; Llewellyn et al., 2016; Gajardo et al., 2016; Zarkasi et al., 2014; Hovda et al., 2012). Overall, there is not much variation between the ONT and Sanger sequencing on the phylum level. BUSCO assigns two *Mycoplasmatota,* however the quality and completeness of these genomes are low (<30% after polishing) and is therefore not reliable as results.

It is at genus level that we see more of a divergence between the two technologies. Sanger analysis of the PCR amplified 16S rRNA genes for our isolates showed a higher variation of different genera with hits of *Arthrobacter*, *Alcaligenes*, *Klebsiella*, *Sporosarcina* and multiple uncultured hits. ONT analysis of the corresponding genomes predicted 10 isolates of the genus Aeromonas, which is double the hits from our 16S rRNA analysis, while not finding any hits of the Arthrobacter genus. ONT also classifies *Shewanella* and *Glutamicibacter*. These results indicate that there are differences on the isolates with smaller prevalence, although they find similar results on the most abundant ones.

4.2 Gene annotation

Insight into the metabolic and enzymatic capabilities of the bugs was obtained with DRAM, which produced the heatmaps shown in figure 12-14. Firstly, CAZyme annotation showed a wide array of consistent results. The presence of chitin, pectin and polyphenolic degrading enzymes are predicted. Chitin is a polymer that's widely abundant in nature (Beier & Bertilsson, 2013) and is commonly used in fish diets as many papers has studied its positive effects on growth and feed conversion (Elserafy, 2021). Pectin and polyphenols are polysaccharides commonly associated with fruits, and it is less probable that these are common in fish feed. Another interesting result was the prediction of beta-mannan degradation function in 13 isolates. Studies of Beta-mannans in the human gut has found that the presence of the polysaccharide promotes growth of SCFAs producing bacteria while reducing the abundance of mucus-degraders, leading to increased protection against inflammation and carcinogenic effects (Leanti La Rosa et al., 2019). Interestingly, of the 11 isolates that were assigned this CAZyme, 9 were classified as *Lelliotta amnigena NZ CP023529*. The other two isolates, S4_bc4 and S28_bc28, was assigned with an *Enterobacter* strain as their closest relative with ANI, however they only beat the *Lelliotta* strain *WB101 NZ CP028520* with 0.12 and 0.08 % for S4_bc4 and S28_bc4 respectively. These results can predict our isolates affiliated with the Lelliotta genus possibly degrade beta-mannan, however considering that not all the Lelliotta strain isolates showed the presence of beta mannan CAZymes, there would have to be more evidence to conclude this.

Figure 14 indicates the predicted presence of genes coding for acetate and lactate conversion in most of the high-quality isolates, while there seems to be no presence of either butyrate or propionate being produced by the isolates in this study. Older studies on SCFA contents in marine fish show that Acetate is the most abundantly produced SCFA, and that most of the production occurs in the distal tract (Clements, 1995). This might suggest reliable indication of gene coding functions for acetate in our bacterial isolates.

Other notable results from DRAM are the three isolates showing the presence of TMAO reductase in S36_bc36, S38_bc38 and S45_bc45. The TMAO amine oxide has been shown to protect against the protein destabilizing effect of urea in marine animals (Velasquez et al., 2016) and serves as an electron acceptor in the metabolism of multiple gut bacteria (Barret & Kwan,

1985). However, higher levels have known detrimental effects like cardiovascular and kidney diseases in humans (Velasquez et al., 2016).

4.3 Novelty

Using the ANI and AAI values provided by MiGA, we can comment on the novelty of our acquired isolates. These values were introduced as a concept for defining species in 2005 by Konstantinos, T et al. as an alternate method from DNA-DNA hybridization which had been the staple before then (Konstantinos et al., 2005; Kim et al., 2021). The ANI and AAI values compare the sequence identities of shared orthologous genes by observing the nucleotide identity and average amino acid identity respectively (Konstantinos et al., 2005), and the accepted threshold for species characterization has been set at ~95% (Jain et al., 2018; Yarza et al., 2014; Richter & Rosselló-Móra, 2009). My data set shows multiple hits >95% on both ANI and AAI values on the same isolates, however they seem to indicate different classifications. Most of the ANI hits over 95% assigns the isolate with *Lelliotta amnigena NZ CP023529* as the closest relative, however the AAI often assigns these as *Enterobacter*. As such, it might indicate that the variation that differs *Lelliotta* from *Enterobacter* lays in the DNA sequence and becomes more hidden during translation. Another possible scenario is that the taxonomic classification of the organism has been incorrectly assigned and has found its way onto the database.

Some isolates are assigned a species not present in the GenBank database for both. These are S5_bc5, S8_bc8, S15_bc15, S19_bc19 and S29_bc29 as they show ANI/AAI values under the 95% threshold. This could indicate the novelty of these sample, however the assembled genomes of S5_bc5, S15_bc15 and S29_bc29 show a genome completeness of <90% both on MiGA and assigned by BUSCO (table S4, appendix B). Therefore, the quality of these genomes might not be enough to conclude the novelty of these isolates. Furthering work on these samples through an increase in sequencing depth that might increase the reliability of these isolates. This can be achieved through a repeat of sequencing and choosing longer and higher quality reads for assembly or combining ONT with Illumina reads for polishing. However, S19_bc19 shows very high genome completeness of ~99 % after polishing and a quality of 90 % assigned by MiGA.

These values might show indication that the isolate belongs to a species not present in the database.

One sample which shows interesting properties is S8_bc8, with an AAI value of 55.6% (table 2) and no assigned ANI value. Considering that the completeness of this genome was good (99.2% and 100% for BUSCO and MiGA respectively) and its quality was relatively high (86% MiGA and over 20 000 assigned reads for sequencing), this might be a candidate for a novel genus. It is even possible that this might be a novel family, however the p-value is too high to conclude this ($p = 0.4$). To check the reliability of these findings, a second taxonomic classification tool was used on the genomes. GTDB-k results contradicts with the classification assigned by MiGA as it returns an ANI value of 97.49% with the closest relative being *Shewanella sp002966515*. Considering that MiGA and GTDB-k provided different results analyzing this sample, the rest of the isolates was compared with the two tools. On all the isolate that MiGA managed to classify, the two tools showed agreement on the classification, with no deviation in the higher quality isolates. I hypothesized that the deviation of the S8_bc8 sample had something to do with the *Shewanellaceae* family not being similarly represented in the databases respectively used by MiGA and GTDB-k, however MiGA correctly managed to classify the S7_bc7 isolate which was identified as *Shewanella sp. WE21*. This result would initially indicate a fault with the MiGA database, however in the classification by GTDB-k you can see that it assigns the isolate in the *Enterobacterales* order. This is unexpected considering that the *Shewanellaceae* family is a part of the *Alteromonadales* order (Satomi, 2014). As both tools seem to struggle in the classification of the isolate, it is possible to conclude that it is the sample that is the problem. Ideally I would test more classification methods to validate this conclusion, however due to time constraints this was not feasible.

4.4 Relatedness of fresh- and saltwater Salmon

In the analysis of my microbial isolates, it is important to note that the samples were all taken from Salmon originating from freshwater tanks. Like previously stated, the anadromous nature of the Atlantic salmon has a noticeable effect on its gut content (Lllewellyn et al., 2014). This is a substantial factor in the classification of the microbial isolates from my samples. As such, I was provided samples taken from saltwater Salmon for comparison of the two environments. From the phylogenetic tree in figure 10 there does not seem to be any clear evidence that the bacteria living in the gut are fully distantly related. The freshwater and saltwater samples are generally separated into closely related sister taxa; however, the overarching pattern shows that the samples are mixed when it comes to their evolutionary distance.

4.5 Media

Notable differences between the isolates of the two mediums are quite expected. There is a prevalence of gammaproteobacterial in the BHI plates, while no hits are found in the TSA plates. This seems to indicate that the bacteria in the class are too fastidious for the TSA medium. However, the TSA plates show a high composition of *Enterobacterales* which the medium is quite efficient at for culturing (Tankeshwar, 2022). The quality difference is also quite apparent between the two mediums, with BHI showing an average of 91.5% genome completeness after polishing. TSA isolates show an average of 65.3%. This low quality probably indicates something wrong with the making of the TSA media, or just a result of the intrinsic error rate of the ONT sequencing method.

4.6 Technical considerations

Only explores distal tract

The overarching goal of this study is to do an exploration into the composition and functions of Salmon gut microbiota to further the work needed to create a salmon microbial genome atlas. However, it is important to note that it only analyses samples taken from the distal tract of the gut. As such, this cannot be taken as a general characterization of the entire Salmon gut.

Lack of information of sampled Salmon

Due to the small scope of this master thesis, no consideration was taken for the conditions of the Salmon, except if it was taken from fresh- or saltwater tanks. As such, conditions like feed content, medical treatment and other factors that can influence the results are not known.

Small sample size

The end sample size of this study is relatively small, ending up with 56 isolates sequenced with sanger and 45 for ONT. When filtered down due to inactive and low-quality genomes, the total sample size becomes too small to make any major final conclusions about the salmon gut microbiota functionality.

7. CONCLUDING REMARKS

Using contemporary long read and 16S rRNA sequencing, this thesis aimed to perform a cultormics based study of the taxonomic composition and phenotypic characteristics of the Atlantic Salmon distal tract microbiota. Results confirm the dominance of the *Pseudomonadota* phylum with smaller abundances of *Bacillotas* and *Actinomycetota* which corresponds with other studies in the same field. Functional analysis using DRAM show prevalent ability of polysaccharide degradation like polyphenols, chitin and pectin, with rarer cases of Starch and beta-mannan CAZymes. SCFA conversions were dominated by acetate and lactate.

This study is small in scope, and further research is needed to confirm the reliability of notable results, as they were usually connected to less complete genomes. Therefore, combining the sequenced reads with higher quality Illumina reads might be a great step in providing more accurate conclusions. However, the building of a genome database requires years and thousands of genomes to become a valuable tool for further research. As such, the results in this paper might hold value as one part of a greater whole in the aim to build the salmon microbial genome atlas.

8. REFERENCES

1. Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quraishi, M. N., Kinross, J., Smidt, H., Tuohy, K. M., et al. (2016). The gut microbiota and host health: a new clinical frontier. *Gut*, 65 (2): 330-339. doi: 10.1136/gutjnl-2015-309990.
2. Heinritz, S. N., Mosenthin, R. & Weiss, E. (2013) "Use of pigs as a potential model for research into dietary modulation of the human gut microbiota," Nutrition Research Reviews. Cambridge University Press, 26(2), pp. 191–209. doi: 10.1017/S0954422413000152.
3. Llewellyn, M. S., Boutin, S., Hoseinifar, S. H. & Derome, N. (2014). Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. *Front Microbiol*, 5: 207-207. doi: 10.3389/fmicb.2014.00207.
4. Ciric M, Waite D, Draper J, Jones JB. Characterisation of gut microbiota of farmed Chinook salmon using metabarcoding. (2018) bioRxiv. Available from: https://doi.org/10.1101/288761
5. Moran, N. A., Ochman, H. & Hammer, T. J. (2019). Evolutionary and Ecological Consequences of Gut Microbial Communities. *Annual review of ecology, evolution, and systematics*, 50 (1): 451-475. doi: 10.1146/annurev-ecolsys-110617-062453.
6. Coon, K. L., Vogel, K. J., Brown, M. R. & Strand, M. R. (2014). Mosquitoes rely on their gut microbiota for development. *Mol Ecol*, 23 (11): 2727-2739. doi: 10.1111/mec.12771.
7. Egerton, S., Culloty, S., Whooley, J., Stanton, C. & Ross, R. P. (2018). The Gut Microbiota of Marine Fish. *Frontiers in Microbiology*, 9. doi: 10.3389/fmicb.2018.00873
8. Semova, I., Carten, Juliana D., Stombaugh, J., Mackey, Lantz C., Knight, R., Farber, Steven A. & Rawls, John F. (2012). Microbiota Regulate Intestinal Absorption and Metabolism of Fatty Acids in the Zebrafish. *Cell Host Microbe*, 12 (3): 277-288. doi: 10.1016/j.chom.2012.08.003.
9. Dehler, C. E., Secombes, C. J. & Martin, S. A. M. (2017). Environmental and physiological factors shape the gut microbiota of Atlantic salmon parr (Salmo salar L.). *Aquaculture*, 467: 149-157. doi: 10.1016/j.aquaculture.2016.07.017.
10. Nishida, A. H. & Ochman, H. (2018). Rates of gut microbiome divergence in mammals. *Mol Ecol*, 27 (8): 1884-1897. doi: 10.1111/mec.14473.
11. Egerton, S., Culloty, S., Whooley, J., Stanton, C. & Ross, R. P. (2018). The Gut Microbiota of Marine Fish. *Frontiers in Microbiology*, 9. doi: 10.3389/fmicb.2018.00873.
12. Llewellyn, M., McGinnity, P., Dionne, M. *et al.* (2016) The biogeography of the atlantic salmon (*Salmo salar*) gut microbiome. *ISME J* **10,** 1280–1284. https://doi.org/10.1038/ismej.2015.189
13. Gajardo, K., Rodiles, A., Kortner, T. M., Krogdahl, Å., Bakke, A. M., Merrifield, D. L. & Sørum, H. (2016). A high-resolution map of the gut microbiota in Atlantic salmon

(Salmo salar): A basis for comparative gut microbial research. *Sci Rep*, 6 (1): 30893-30893. doi: 10.1038/srep30893.
14. Artsdatabanken. Laksefamilien Salmonidae (s.a.) Available from: https://artsdatabanken.no/Taxon/Salmonidae/42660. (Accsessed: April 2022)
15. Vøllestad A. laks. (s.a.) Oslo: Store norske leksikon; [updated 2022]. Available from: https://snl.no/laks. (Accsessed: April 2022)
16. Misund B. (s.a.) Fiskeoppdrett. Stavanger: Store norske leksikon; [updated 2021]. Available from: https://snl.no/fiskeoppdrett. (Accsessed: April 2022)
17. SeafoodFromNorway. (s.a.) The Salmon life cycle. Available from: https://salmon.fromnorway.com/sustainable-aquaculture/the-salmon-lifecycle/. (Accsessed: April 2022)
18. Khaw, H. L., Gjerde, B., Boison, S. A., Hjelle, E. & Difford, G. F. (2021). Quantitative Genetics of Smoltification Status at the Time of Seawater Transfer in Atlantic Salmon (Salmo Salar). *Frontiers in genetics*, 12: 696893-696893. doi: 10.3389/fgene.2021.696893.
19. Sigholt, T., Åsgård, T. & Staurnes, M. (1998). Timing of parr-smolt transformation in Atlantic salmon (Salmo salar): effects of changes in temperature and photoperiod. *Aquaculture*, 160 (1): 129-144. doi: 10.1016/S0044-8486(97)00220-2.
20. Rikardsen AH, Dempson JB. (2010) Dietary Life-support: The Food and Feeding of Atlantic Salmon at Sea. In: Aas Ø, Einum S, Klemetsen A, Skurdal J, editors. Atlantic Salmon Ecology. United Kingdom: Wiley-Blackwell;. p. 115-43. doi: 10.1002/9781444327755.ch5
21. Ringo, E., Olsen, R. E., Gifstad, T. O., Dalmo, R. A., Amlund, H., Hemre, G. I. & Bakke, A. M. (2010). Prebiotics in aquaculture: a review. *Aquaculture nutrition*, 16 (2): 117-136. doi: 10.1111/j.1365-2095.2009.00731.x.
22. Cheplin, H. A. & Rettger, L. F. (1920). Studies on the Transformation of the Intestinal Flora, with Special Reference to the Implantation of Bacillus Acidophilus. II. Feeding Experiments on Man. *Proc Natl Acad Sci U S A*, 6 (12): 704-705. doi: 10.1073/pnas.6.12.704.
23. Gibson GR, Roberfroid MB. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr. 125(6):1401-12. doi: 10.1093/jn/125.6.1401.
24. Hutkins, R. W., Krumbeck, J. A., Bindels, L. B., Cani, P. D., Fahey, G., Goh, Y. J., Hamaker, B., Martens, E. C., Mills, D. A., Rastal, R. A., et al. (2015). Prebiotics: why definitions matter. *Curr Opin Biotechnol*, 37: 1-7. doi: 10.1016/j.copbio.2015.09.001.
25. Andreassen G, Martinussen T. (2011) Aquaculture in Norway. Oslo, Norway: Norwegian Seafood Federation. Available from: https://sjomatnorge.no/app/uploads/importedfiles/Aquaculture%2520in%2520Norway%25202011.pdf. (Accsessed: February 2022)

26. FAO. 2021. GLOBEFISH Highlights - A quarterly update on world seafood markets 1st issue 2021 January–September 2020 Statistics. Globefish Highlights No. 1–2021. Rome. https://doi.org/10.4060/cb4129en
27. Raudstein, M. (2020). Investigating prevalence and geographical distribution of Mycoplasma sp. in the gut of Atlantic salmon (Salmo salar L.): M.Sc. thesis. Ås, Norwegian University of Life Sciences.
28. Richardsen R, Myhre MS, Tyholt IL, Johansen U. (2019) Nasjonal betydning av sjømatnæringen. Trondheim, Norway: SINTEF. Available from: Nasjonal betydning av sjømatnæringen (sintef.no). (Accsessed: April 2022)
29. Green, T. J., Smullen, R. & Barnes, A. C. (2013). Dietary soybean protein concentrate-induced intestinal disorder in marine farmed Atlantic salmon, Salmo salar is associated with alterations in gut microbiota. *Vet Microbiol*, 166 (1-2): 286-292. doi: 10.1016/j.vetmic.2013.05.009.
30. Lagier, J.-C., Khelaifia, S., Alou, M. T., Ndongo, S., Dione, N., Hugon, P., Caputo, A., Cadoret, F., Traore, S. I., Seck, E. H., et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol*, 1 (12): 16203-16203. doi: 10.1038/nmicrobiol.2016.203.
31. Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J. F., Koonin, E. V., La Scola, B., Raoult, D. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study, Clinical Microbiology and Infection, Volume 18, Issue 12, Pages 1185-1193, ISSN 1198-743X, https://doi.org/10.1111/1469-0691.12023.
32. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499 (7459): 431-437. doi: 10.1038/nature12352.
33. Diakite, A., Dubourg, G., Dione, N., Afouda, P., Bellali, S., Ngom, I. I., Valles, C., Tall, M. L., Lagier, J.-C. & Raoult, D. (2020). Optimization and standardization of the culturomics technique for human microbiome exploration. *Sci Rep*, 10 (1): 9674-9674. doi: 10.1038/s41598-020-66738-8.
34. Dubourg, G., Lagier, J. C., Armougom, F., Robert, C., Hamad, I., Brouqui, P. & Raoult, D. (2013). The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis*, 32 (5): 637-645. doi: 10.1007/s10096-012-1787-3.
35. Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proc Natl Acad Sci U S A*, 74 (12): 5463-5467. doi: 10.1073/pnas.74.12.5463.
36. Heather, J. M. & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107 (1): 1-8. doi: 10.1016/j.ygeno.2015.11.003.

37. Segerman, B. (2020). The Most Frequently Used Sequencing Technologies and Assembly Methods in Different Time Segments of the Bacterial Surveillance and RefSeq Genome Databases. *Front Cell Infect Microbiol*, 10: 527102-527102. doi: 10.3389/fcimb.2020.527102.
38. Illumina®. (2017). An introduction to Next-Generation Sequencing Technology. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
39. CD Genomics. (s.a.) Sanger Sequencing: introduction, Principle and Protocol. Available from: https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/. (Accsessed: April 2022)
40. Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J Clin Microbiol*, 45 (9): 2761-2764. doi: 10.1128/JCM.01228-07.
41. Patel JB. 16S rRNA Gene Sequencing for Bacterial Pathogen Identification in the Clinical Laboratory. Mol. Diagn. 2001 Dec;6(4):313-21. doi: 10.1054/modi.2001.29158
42. Lu, H., Giordano, F. & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, proteomics & bioinformatics*, 14 (5): 265-279. doi: 10.1016/j.gpb.2016.05.004.
43. Akeson, M., Branton, D., Church, G Deamer, D. W. (1995). Characterization of individual polymer molecules based on monomer-interface interactions. Google patent, URL: https://patents.google.com/patent/US20120160687#nplCitations
44. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, 17 (1): 239-239. doi: 10.1186/s13059-016-1103-0.
45. Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., et al. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26 (10): 1146-1153. doi: 10.1038/nbt.1495.
46. Moss, E. L., Maghini, D. G. & Bhatt, A. S. (2020). Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*, 38 (6): 701-707. doi: 10.1038/s41587-020-0422-6.
47. De Coster, W., D'Hert, S., Schultz D. T., Cruts, M., Broeckhoven, C. (2018) NanoPack: visualizing and processing long-read sequencing data, *Bioinformatics*, Volume 34, Issue 15, Pages 2666–2669, doi: 10.1093/bioinformatics/bty149
48. Wick, R., Menzel, P. (2017). *Filtlong*. Available from: https://github.com/rrwick/Filtlong (Accessed: 10.03.2022).
49. Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P. (2019) Assembly of Long Error-Prone Reads Using Repeat Graphs, Nature Biotechnology, doi: 10.1038/s41587-019-0072-8
50. Simão, F. A., Waterhouse, W. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. doi: 10.1093/bioinformatics/btv351

51. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*, 27 (5): 737-746. doi: 10.1101/gr.214270.116.
52. Vaser, R. et al (2016) Racon. GitHub repository. Available at: https://github.com/lbcb-sci/racon (Accessed: 04.04.2022)
53. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34:3094-3100. doi:10.1093/bioinformatics/bty191
54. Huang, N., Nie, F., Ni, P., Luo, F., Gao, X. & Wang, J. (2021). NeuralPolish: a novel Nanopore polishing method based on alignment matrix construction and orthogonal Bi-GRU Networks. *Bioinformatics*, 37 (19): 3120-3127. doi: 10.1093/bioinformatics/btab354.
55. Rescheneder, P. et al (2016) Medaka. GitHub repository. Available at: https://github.com/nanoporetech/medaka (Accessed: 04.04.2022)
56. Rodriguez-R *et al*. (2018). The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research* 46(W1): W282-W288. doi:10.1093/nar/gky467
57. Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*, 48 (16): 8883-8900. doi: 10.1093/nar/gkaa621.
58. Zarkasi, K. Z., Abell, G. C. J., Taylor, R. S., Neuman, C., Hatje, E., Tamplin, M. L., Katouli, M. & Bowman, J. P. (2014). Pyrosequencing- based characterization of gastrointestinal bacteria of Atlantic salmon (Salmo salar L.) within a commercial mariculture system. *J Appl Microbiol*, 117 (1): 18-27. doi: 10.1111/jam.12514.
59. Hovda, M. B., Fontanillas, R., McGurk, C., Obach, A. & Rosnes, J. T. (2012). Seasonal variations in the intestinal microbiota of farmed Atlantic salmon (Salmo salar L.): Seasonal variations in the intestinal microbiota of Salmo salar L. *Aquaculture research*, 43 (1): 154-159. doi: 10.1111/j.1365-2109.2011.02805.x.
60. Beier, S. & Bertilsson, S. (2013). Bacterial chitin degradation-mechanisms and ecophysiological strategies. *Front Microbiol*, 4: 149-149. doi: 10.3389/fmicb.2013.00149.
61. Elserafy, S. S., Abdel-Salam, H. A., Dakrouni, A. M et al. (2021). Effect of shrimp waste extracted chitin on growth and some biochemical parameters of the Nile tilapia. *Egyptian Journal of Aquatic Biology and Fisheries*, 25(1): 313 – 329. Available from: https://ejabf.journals.ekb.eg/article_143244_799f5c16a73226580a2660309512da7c.pdf
62. McGivern, B. B., Tfaily, M. M., Borton, M. A., Kosina, S. M., Daly, R. A., Nicora, C. D., Purvine, S. O., Wong, A. R., Lipton, M. S., Hoyt, D. W., et al. (2021). Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat Commun*, 12 (1): 2466-2466. doi: 10.1038/s41467-021-22765-1.
63. Wade Abbott, D. & Boraston, A. B. (2008). Structural Biology of Pectin Degradation by Enterobacteriaceae. *Microbiol Mol Biol Rev*, 72 (2): 301-316. doi: 10.1128/MMBR.00038-07.

64. Leanti La Rosa, S., Leth, M. L., Michalak, L., Hansen, M. E., Pudlo, N., Glowacki, R., Pereira, G., Workman, C., Arntzen, M. Ø., Pope, P., et al. (2019). The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. doi: https://doi.org/10.1038/s41467-019-08812-y.
65. Clements, K. D. & Choat, J. H. (1995). Fermentation in Tropical Marine Herbivorous Fishes. *Physiological zoology*, 68 (3): 355-378. doi: 10.1086/physzool.68.3.30163774.
66. Velasquez, M. T., Ramezani, A., Manal, A. & Raj, D. S. (2016). Trimethylamine N-Oxide: The Good, the Bad and the Unknown. *Toxins (Basel)*, 8 (11): 326. doi: 10.3390/toxins8110326.
67. Barret, E. L., Kwan, H. S. (1985). BACTERIAL REDUCTION OF TRIMETHYLAMINE OXIDE. doi: 10.1146/annurev.mi.39.100185.001023.
68. Konstantinidis, K. T. & Tiedje, J. M. (2005). Genomic Insights That Advance the Species Definition for Prokaryotes. *Proc Natl Acad Sci U S A*, 102 (7): 2567-2572. doi: 10.1073/pnas.0409727102.
69. Kim, D., Park, S. & Chun, J. (2021). Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. *J Microbiol*, 59 (5): 476-480. doi: 10.1007/s12275-021-1154-0.
70. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*, 9 (1): 5114-8. doi: 10.1038/s41467-018-07641-9.
71. Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R. & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*, 12 (9): 635-645. doi: 10.1038/nrmicro3330.
72. Richter, M. & Rosselló-Móra, R. (2009). Shifting the Genomic Gold Standard for the Prokaryotic Species Definition. *Proc Natl Acad Sci U S A*, 106 (45): 19126-19131. doi: 10.1073/pnas.0906412106.
73. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36 (6): 1925-1927. doi: 10.1093/bioinformatics/btz848.
74. Satomi, M. (2014). The Family Shewanellaceae. I, s. 597-625. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-38922-1_226
75. Tankeshwar, A. (2022). *Tryptic Soy Agar (TSA): Composition, Preparation, Uses*. Available from: https://microbeonline.com/tryptic-soy-agar-tsa-composition-preparation-uses/ (Accessed: 22.04.2022).

9. APPENDIX

9.1 Appendix A: method

Extracted HMW DNA Nanodrop values:

Table S1: Nanodrop values of the extracted DNA obtained using the Nanobind HMW DNA extraction kit.

Sample	ng/µl	A260/280	A260/230
BHI01	526.2	1.86	1.74
BHI02	206	1.50	0.71
BHI03	182.6	1.36	0.74
BHI04	405.7	1.82	1.55
BHI05	661.4	1.92	1.83
BHI06	559.5	1.88	1.78
BHI07	421.3	1.85	1.89
BHI08	440.9	1.85	1.90
BHI09	439.2	1.79	1.31
BHI11	307.6	1.84	1.61
BHI12	345.9	1.93	1.87
BHI13	510.1	1.79	1.28
BHI14	733.9	1.85	1.56
BHI15	415.3	1.79	1.31
BHI17	225.0	1.97	1.79
BHI18	133.6	1.60	0.76
BHI19	278.3	1.91	1.10
BHI20	169.2	1.63	0.96
BHI22	150.4	1.58	0.92
BHI23	491.3	1.87	1.78
BHI27	451.2	1.87	1.43
BHI31	104.2	1.64	0.80
BHI32	575.7	1.86	1.44
BHI33	308.4	1.81	1.43
BHI35	338.8	1.85	1.73
BHI36	749.8	1.71	1.11
BHI38	36.0	0.96	0.24
BHI40	444.4	1.73	1.25
BHI41	59.9	1.33	0.48
BHI43	1029.5	2.02	2.12
BHI44	254.4	1.60	0.83

TSA01	799.2	1.87	1.38
TSA02	517.8	1.73	1.20
TSA03	735.3	1.83	1.51
TSA06	629.4	1.81	1.27
TSA11	681.3	1.82	1.29
TSA12	529.0	1.83	1.51
TSA17	849.3	1.75	1.31
TSA19	171.2	1.49	0.61
TSA21	283.2	1.46	0.63
TSA23	788.2	1.88	1.52
TSA24	627.5	1.82	1.25
TSA28	649.9	1.78	1.18
TSA35	607.8	1.77	1.05
TSA36	813.6	1.84	1.40
TSA37	144.2	1.38	0.55
TSA39	121.0	1.43	0.72
TSA40	1026.7	1.51	1.25
TSA45	907.2	1.83	1.35
TSA46	636.6	1.84	1.25

Qbit measured DNA concentration of the 15 isolates chosen for SRE:

Table S2: DNA concentration of isolates chosen for SRE protocol. Measured by Qbit.

	Isolate	Concentration (ng/µl)
BHI	19	13.7
	27	25.4
TSA	1	164
	2	81.3
	3	126
	6	322
	11	193
	17	26.5
	23	80
	24	291
	28	320
	35	280
	36	331

	45	279
	46	655

Document containing all bash scripts and parameters used for data processing:

C:\Users\turne\OneDrive\Dokumenter\MASTER OPPGAVE\SCRIPT.COLLECTION.pdf

9.2 Appendix B: Results

Sanger sequencing results:

*Table S3: 16s rRNA Sanger sequencing results. The forward and reverse reads were assembled using the CAP program on BioEdit and then blasted using BLASTN on the NCBI website. Hit with lowest E-value was chosen. *Samples with poor quality forward or reverse read caused problems in the assembly process. For these samples, only the better read was used instead.*

Sample	Name	query length	Identity
BHI01	*Arthrobacter_sp_TM4_3*	1436	98%
BHI02	*Aeromonas_sp_CU5*	1466	97%
BHI03	*Lelliottia_amnigena_strain_PP3*	1466	97%
BHI04	*Lelliottia_amnigena_strain_PP3*	1465	95%
BHI05	*Pseudomonas_sp_strain_U3-n-1*	1476	97%
BHI06	*Aeromonas_encheleia_strain_NCTC12917*	1466	96%
BHI07*	*Alcaligenes_faecalis_strain_GTE53*	894	98%
BHI08	*Uncultured_bacterium_clone_B1P1A5*	1498	94%
BHI09	*Uncultured_bacterium_clone_B1P1A5*	1625	86%
BHI10	*Lelliottia_amnigena_strain_PP3*	1471	97%
BHI11	*Lelliottia_amnigena_strain_PP3*	1499	93%
BHI12	*Aeromonas_sp_strain_FS2017-2*	1457	99%
BHI13	*Uncultured_bacterium_partial*	1751	87%
BHI14	*Uncultured_bacterium_clone_RPR-CFL2*	1519	95%
BHI15	*Uncultured_bacterium_partial*	1513	91%
BHI17	*Arthrobacter_sp_TM4_3*	1587	98%

BHI18	*Enterobacter_sp_KAR3*	1642	97%
BHI19	*Enterobacter_sp_KAR3*	1588	98%
BHI20	*Lelliottia_amnigena_strain_JM121*	1532	97%
BHI21	*Enterobacter_sp_KAR3*	1535	97%
BHI22*	*Uncultured_Enterobacter_sp_clone_Entero_aerogenes*	1471	98%
BHI23	*Uncultured_bacterium_partial*	1525	97%
BHI27	*Enterobacter_sp_KAR3*	1516	96%
BHI28	*Enterobacter_sp_KAR3*	1538	97%
BHI31*	*Serratia_proteamaculans_strain_OAct423*	1409	98%
BHI32	*Lelliottia_amnigena_strain_PP3*	1602	98%
BHI33*	*Klebsiella_aerogenes_strain_cqsR4*	1439	98%
BHI35	*Aeromonas_piscicola_strain_WL-185*	1548	93%
BHI36	*Uncultured_bacterium_clone_SedCS37*	1536	90%
BHI38*	*Klebsiella_aerogenes_strain_cqsR4*	1446	98%
BHI40*	*Aeromonas_sp_strain_A16OP9*	1437	98%
BHI41	*Lelliottia_amnigena_strain_JM119*	1519	95%
BHI43	*Pseudomonas_sp_strain_U2-d-1*	1512	94%
BHI44	*Enterobacter_sp_KAR3*	1487	97%

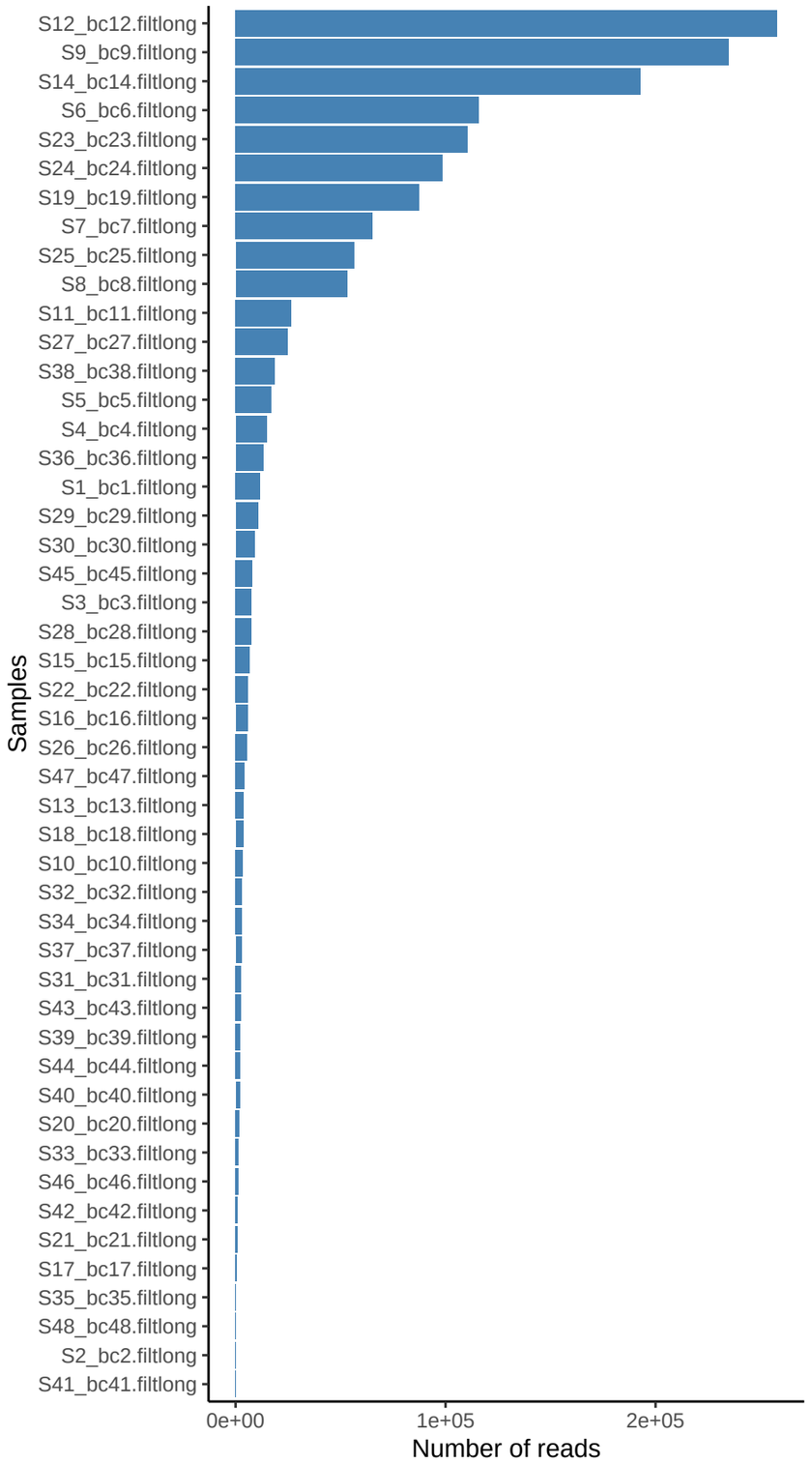
Sample	Name	query length	Identity
TSA01	*Lelliottia_amnigena_strain_PP3*	1524	96%
TSA02	*Lelliottia_amnigena_strain_PP3*	1471	95%
TSA03	*Lelliottia_amnigena_strain_PP3*	1559	97%
TSA04	*Lelliottia_amnigena_strain_PP3*	1510	96%
TSA11	*Lelliottia_amnigena_strain_JM119*	1609	88%
TSA12	*Arthrobacter_sp_QHL12*	1829	86%
TSA17	*Enterobacter_sp_KAR3*	1525	97%
TSA18	*Lelliottia_amnigena_strain_PP3*	1469	96%
TSA21	*Carnobacterium_maltaromaticum_strain_18ISCm*	1497	96%
TSA22	*Lelliottia_amnigena_strain_PP3*	1455	98%
TSA23	*Lelliottia_amnigena_strain_PP3*	1546	87%
TSA24	*Lelliottia_amnigena_strain_NCTC12124*	1484	92%
TSA27	*Lelliottia_amnigena_strain_PP3*	1637	98%
TSA28	*Lelliottia_amnigena_strain_JM119*	1634	96%
TSA35	*Enterobacter_sp_KAR3*	1581	97%
TSA35	*Arthrobacter_sp_JLX12*	1536	94%
TSA36*	*Lelliottia_amnigena_strain_FDAARGOS_395*	1477	98%
TSA37	*Arthrobacter_sp_JLX12*	1609	97%

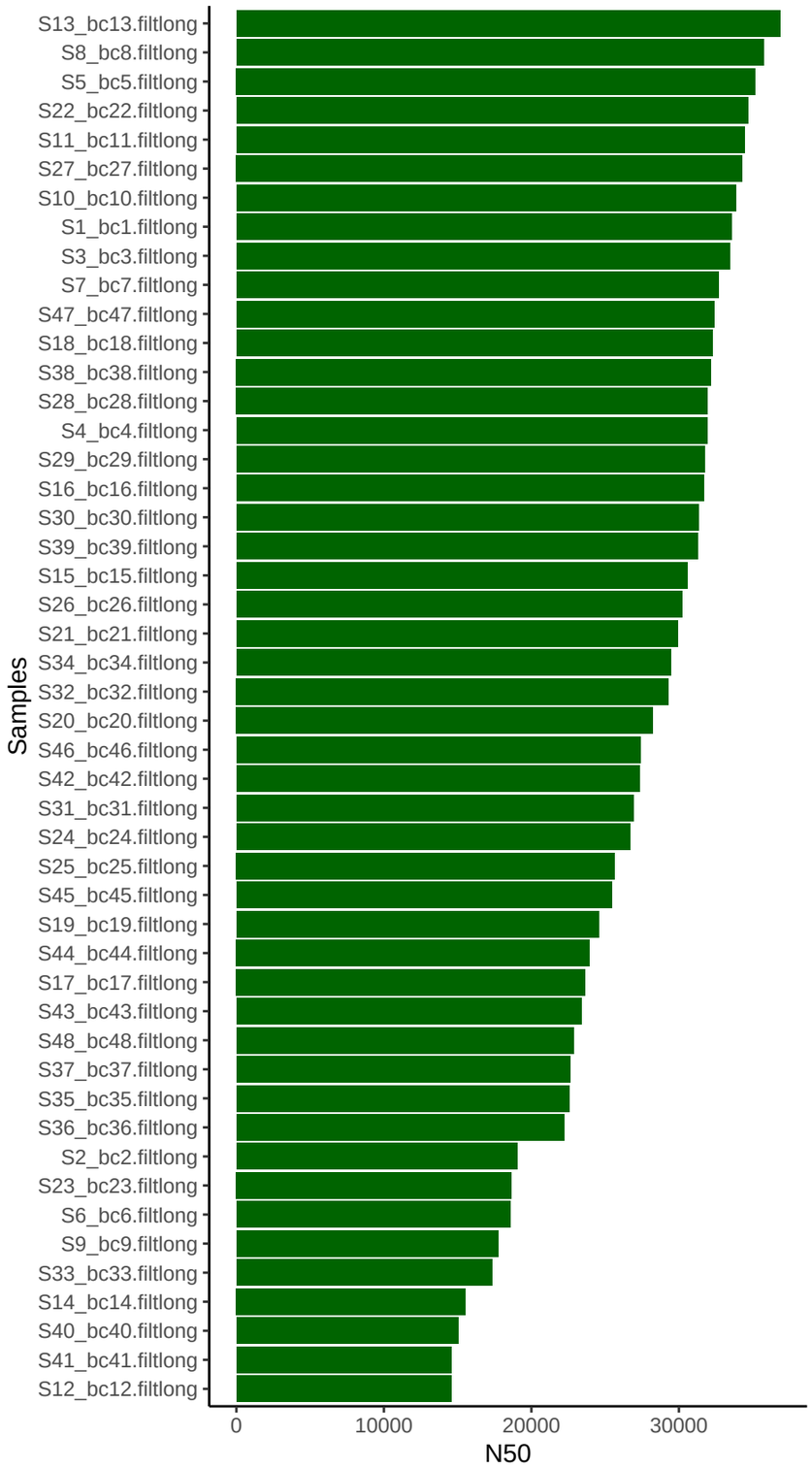
TSA38	*Lelliottia_amnigena_strain_PP3*	1543	98%
TSA39	*Sporosarcina_aquimarina_strain_KUDC1821*	1609	97%
TSA40	*Uncultured_bacterium_clone_258-4*	1543	98%
TSA45*	*Uncultured_Enterobacter_sp_clone_Entero_aerogenes*	1503	98%
TSA46	*Lelliottia_amnigena_strain_JM121*	1541	90%

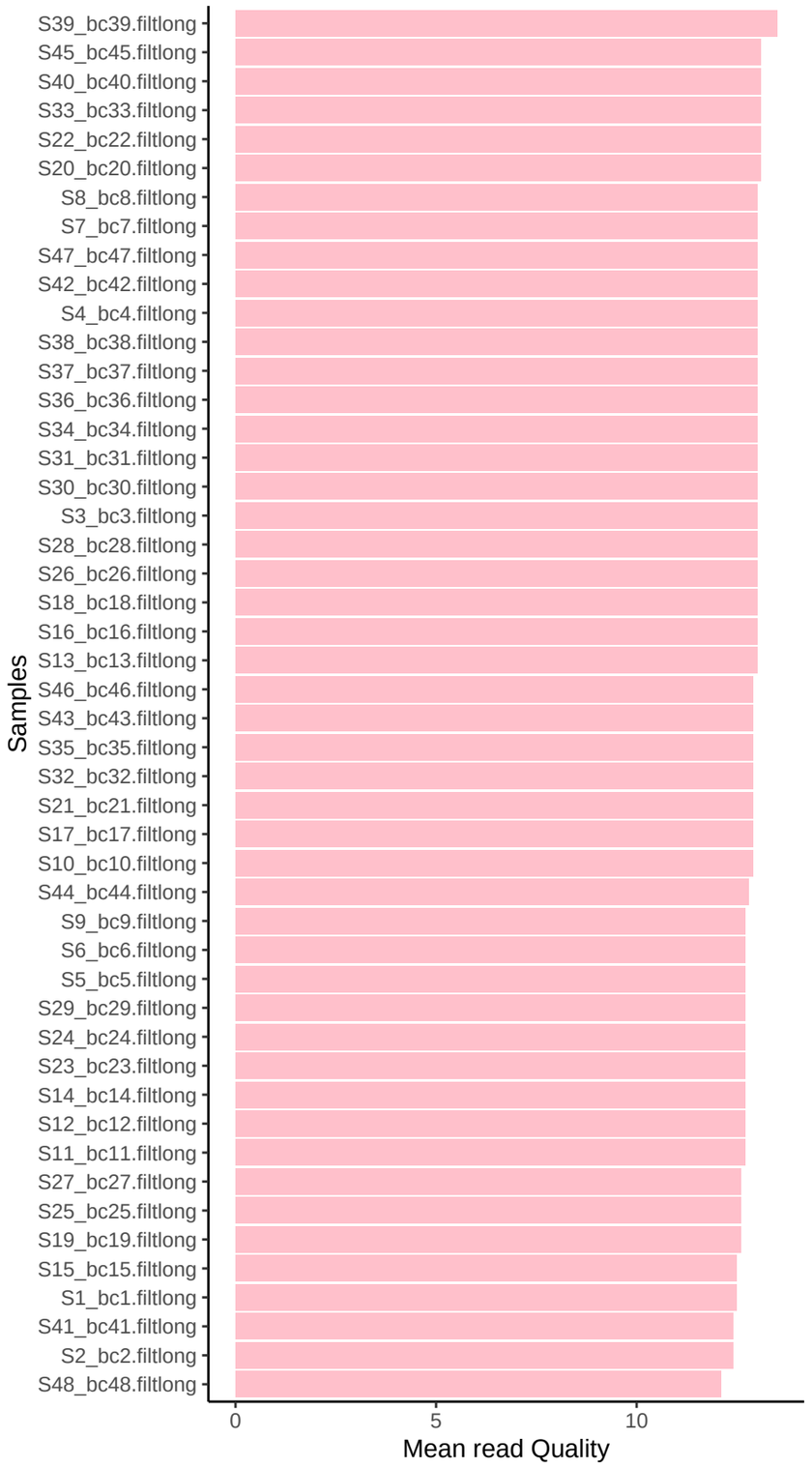
Phylogenetic tree with 16s rRNA results + Samples provided:

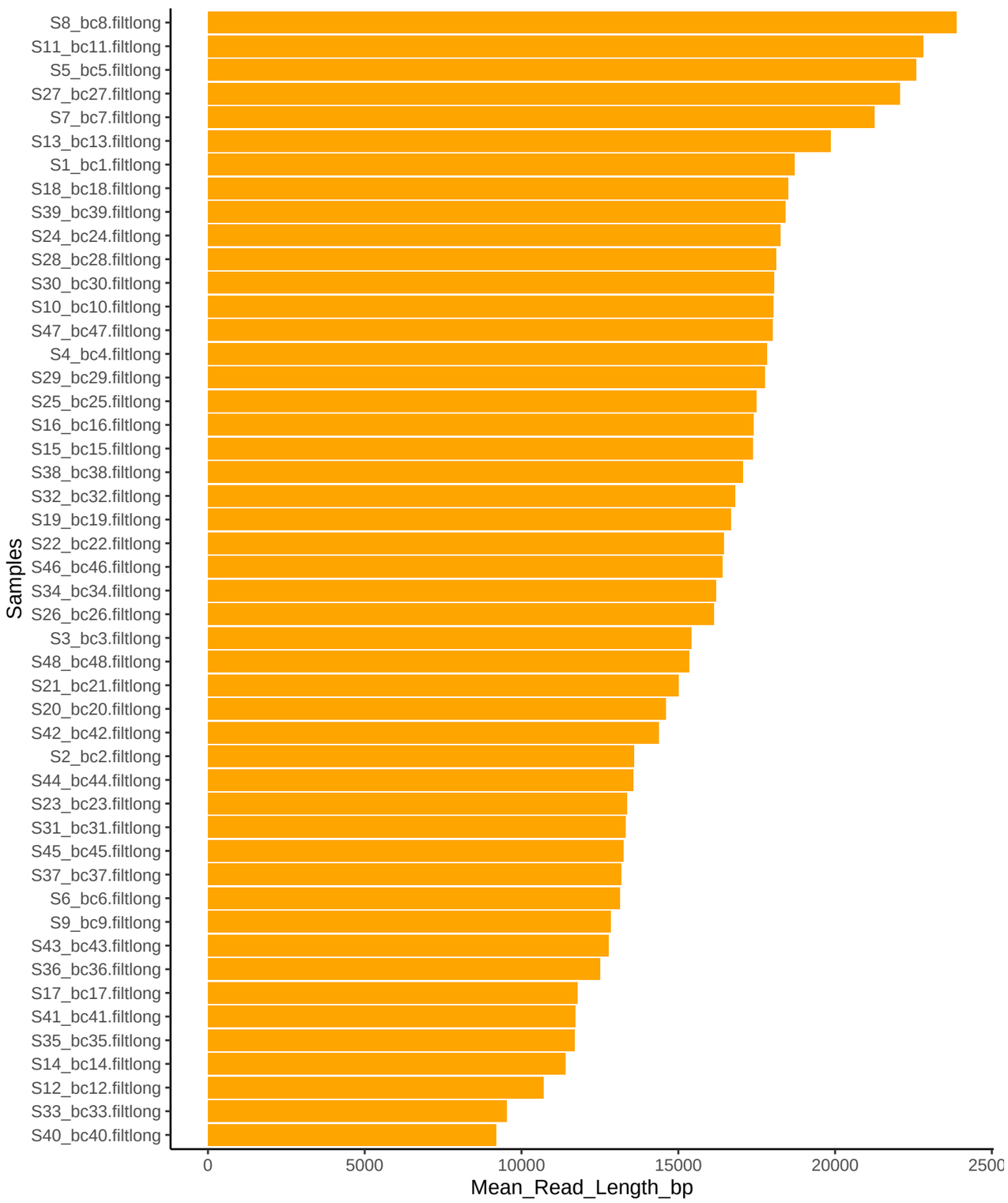
https://d.docs.live.net/8a11757707ea179e/Dokumenter/MASTER%20OPPGAVE/PDFs/cladogram1_full.pdf

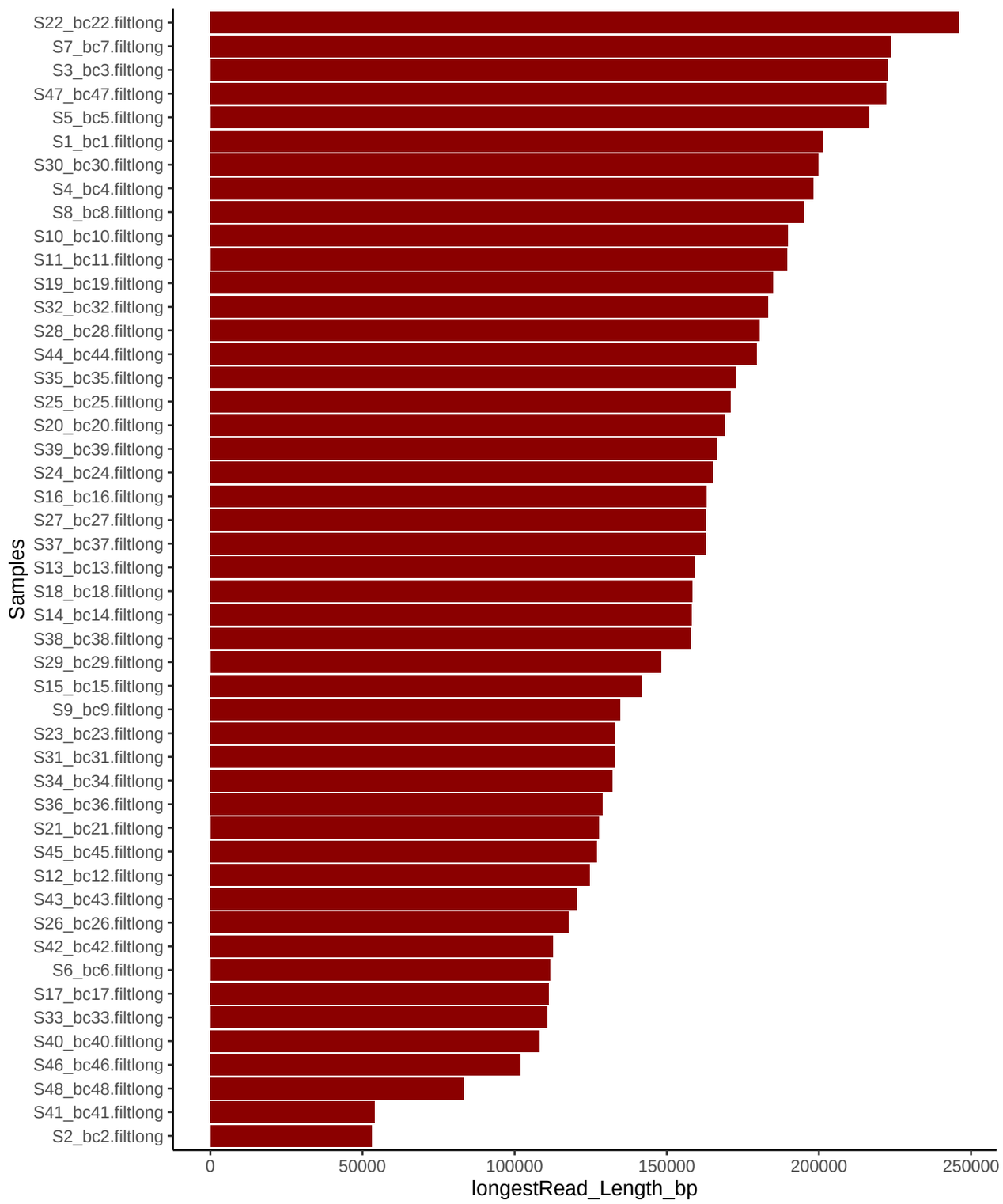
Barplots produced by Nanoplot:


S12_bc12.filtlong
S9_bc9.filtlong
S14_bc14.filtlong
S6_bc6.filtlong
S23_bc23.filtlong
S24_bc24.filtlong
S19_bc19.filtlong
S7_bc7.filtlong
S25_bc25.filtlong
S8_bc8.filtlong
S11_bc11.filtlong
S27_bc27.filtlong
S38_bc38.filtlong
S5_bc5.filtlong
S4_bc4.filtlong
S36_bc36.filtlong
S1_bc1.filtlong
S29_bc29.filtlong
S30_bc30.filtlong
S45_bc45.filtlong
S3_bc3.filtlong
S28_bc28.filtlong
S15_bc15.filtlong
S22_bc22.filtlong
S16_bc16.filtlong
S26_bc26.filtlong
S47_bc47.filtlong
S13_bc13.filtlong
S18_bc18.filtlong
S10_bc10.filtlong
S32_bc32.filtlong
S34_bc34.filtlong
S37_bc37.filtlong
S31_bc31.filtlong
S43_bc43.filtlong
S39_bc39.filtlong
S44_bc44.filtlong
S40_bc40.filtlong
S20_bc20.filtlong
S33_bc33.filtlong
S46_bc46.filtlong
S42_bc42.filtlong
S21_bc21.filtlong
S17_bc17.filtlong
S35_bc35.filtlong
S48_bc48.filtlong
S2_bc2.filtlong
S41_bc41.filtlong
Samples
0e+00
1e+05
2e+05
Number of reads


S13_bc13.filtlong
S8_bc8.filtlong
S5_bc5.filtlong
S22_bc22.filtlong
S11_bc11.filtlong
S27_bc27.filtlong
S10_bc10.filtlong
S1_bc1.filtlong
S3_bc3.filtlong
S7_bc7.filtlong
S47_bc47.filtlong
S18_bc18.filtlong
S38_bc38.filtlong
S28_bc28.filtlong
S4_bc4.filtlong
S29_bc29.filtlong
S16_bc16.filtlong
S30_bc30.filtlong
S39_bc39.filtlong
S15_bc15.filtlong
S26_bc26.filtlong
S21_bc21.filtlong
S34_bc34.filtlong
S32_bc32.filtlong
S20_bc20.filtlong
S46_bc46.filtlong
S42_bc42.filtlong
S31_bc31.filtlong
S24_bc24.filtlong
S25_bc25.filtlong
S45_bc45.filtlong
S19_bc19.filtlong
S44_bc44.filtlong
S17_bc17.filtlong
S43_bc43.filtlong
S48_bc48.filtlong
S37_bc37.filtlong
S35_bc35.filtlong
S36_bc36.filtlong
S2_bc2.filtlong
S23_bc23.filtlong
S6_bc6.filtlong
S9_bc9.filtlong
S33_bc33.filtlong
S14_bc14.filtlong
S40_bc40.filtlong
S41_bc41.filtlong
S12_bc12.filtlong
Samples
0
10000
20000
30000
N50


S39_bc39.filtlong
S45_bc45.filtlong
S40_bc40.filtlong
S33_bc33.filtlong
S22_bc22.filtlong
S20_bc20.filtlong
S8_bc8.filtlong
S7_bc7.filtlong
S47_bc47.filtlong
S42_bc42.filtlong
S4_bc4.filtlong
S38_bc38.filtlong
S37_bc37.filtlong
S36_bc36.filtlong
S34_bc34.filtlong
S31_bc31.filtlong
S30_bc30.filtlong
S3_bc3.filtlong
S28_bc28.filtlong
S26_bc26.filtlong
S18_bc18.filtlong
S16_bc16.filtlong
S13_bc13.filtlong
S46_bc46.filtlong
S43_bc43.filtlong
S35_bc35.filtlong
S32_bc32.filtlong
S21_bc21.filtlong
S17_bc17.filtlong
S10_bc10.filtlong
S44_bc44.filtlong
S9_bc9.filtlong
S6_bc6.filtlong
S5_bc5.filtlong
S29_bc29.filtlong
S24_bc24.filtlong
S23_bc23.filtlong
S14_bc14.filtlong
S12_bc12.filtlong
S11_bc11.filtlong
S27_bc27.filtlong
S25_bc25.filtlong
S19_bc19.filtlong
S15_bc15.filtlong
S1_bc1.filtlong
S41_bc41.filtlong
S2_bc2.filtlong
S48_bc48.filtlong
Samples
0
5
10
Mean read Quality


S8_bc8.filtlong
S11_bc11.filtlong
S5_bc5.filtlong
S27_bc27.filtlong
S7_bc7.filtlong
S13_bc13.filtlong
S1_bc1.filtlong
S18_bc18.filtlong
S39_bc39.filtlong
S24_bc24.filtlong
S28_bc28.filtlong
S30_bc30.filtlong
S10_bc10.filtlong
S47_bc47.filtlong
S4_bc4.filtlong
S29_bc29.filtlong
S25_bc25.filtlong
S16_bc16.filtlong
S15_bc15.filtlong
S38_bc38.filtlong
S32_bc32.filtlong
S19_bc19.filtlong
S22_bc22.filtlong
S46_bc46.filtlong
S34_bc34.filtlong
S26_bc26.filtlong
S3_bc3.filtlong
S48_bc48.filtlong
S21_bc21.filtlong
S20_bc20.filtlong
S42_bc42.filtlong
S2_bc2.filtlong
S44_bc44.filtlong
S23_bc23.filtlong
S31_bc31.filtlong
S45_bc45.filtlong
S37_bc37.filtlong
S6_bc6.filtlong
S9_bc9.filtlong
S43_bc43.filtlong
S36_bc36.filtlong
S17_bc17.filtlong
S41_bc41.filtlong
S35_bc35.filtlong
S14_bc14.filtlong
S12_bc12.filtlong
S33_bc33.filtlong
S40_bc40.filtlong
Samples
0
5000
10000
15000
20000
2500
Mean_Read_Length_bp


S22_bc22.filtlong
S7_bc7.filtlong
S3_bc3.filtlong
S47_bc47.filtlong
S5_bc5.filtlong
S1_bc1.filtlong
S30_bc30.filtlong
S4_bc4.filtlong
S8_bc8.filtlong
S10_bc10.filtlong
S11_bc11.filtlong
S19_bc19.filtlong
S32_bc32.filtlong
S28_bc28.filtlong
S44_bc44.filtlong
S35_bc35.filtlong
S25_bc25.filtlong
S20_bc20.filtlong
S39_bc39.filtlong
S24_bc24.filtlong
S16_bc16.filtlong
S27_bc27.filtlong
S37_bc37.filtlong
S13_bc13.filtlong
S18_bc18.filtlong
S14_bc14.filtlong
S38_bc38.filtlong
S29_bc29.filtlong
S15_bc15.filtlong
S9_bc9.filtlong
S23_bc23.filtlong
S31_bc31.filtlong
S34_bc34.filtlong
S36_bc36.filtlong
S21_bc21.filtlong
S45_bc45.filtlong
S12_bc12.filtlong
S43_bc43.filtlong
S26_bc26.filtlong
S42_bc42.filtlong
S6_bc6.filtlong
S17_bc17.filtlong
S33_bc33.filtlong
S40_bc40.filtlong
S46_bc46.filtlong
S48_bc48.filtlong
S41_bc41.filtlong
S2_bc2.filtlong
Samples
0
50000
100000
150000
200000
250000
longestRead_Length_bp

Flye assembly and BUSCO evaluation results:

Filename	Sample	Generic(%)	predicted(%)	Database	quality	Genome size	N50	Contigs
S1_bc1	BHI-01	88.7	87.2	micrococcales	L20K	3870650	3813288	2
S2_bc2	negative control	NA	NA	NA	L20K	NA	NA	NA
S3_bc3	BHI-03	91.1	87.5	enterobacterales	L20K	4542900	4542900	1
S4_bc4	BHI-04	91.9	91.1	enterobacterales	L20K	4542866	4542866	1
S5_bc5	BHI-05	75.8	73.4	pseudomonadales	L20K	5627038	5627038	1
S6_bc6	BHI-06	81.4	75.7	gammaproteobacteria	H20K	4967230	4967230	1
S7_bc7	BHI-07	91.1	91.7	alteromonadales	H20K	5029678	5029678	1
S8_bc8	BHI-08	90.3	92.6	alteromonadales	H20K	5029704	5029704	1
S9_bc9	BHI-09	74.2	74.3	gammaproteobacteria	H20K	4902195	4836759	3
S10_bc10	BHI-11	83.9	78.9	enterobacterales	L20K	4542798	4542798	1
S11_bc11	BHI-12	83.9	81.6	gammaproteobacteria	H20K	4776409	4776409	1
S12_bc12	BHI-13	76.6	72.9	gammaproteobacteria	H20K	4891931	4836829	3
S13_bc13	BHI-14	76.6	79.1	enterobacterales	L20K	4559574	4559574	1
S14_bc14	BHI-15	75	75.4	gammaproteobacteria	H20K	4886073	4837088	2
S15_bc15	BHI-17	85.5	81.8	micrococcales	L20K	3870621	3813287	2
S16_bc16	BHI-18	85.5	83.9	enterobacterales	L20K	4551154	4551154	1
S17_bc17	BHI-20	NA	NA	NA	L20K	NA	NA	NA
S18_bc18	BHI-22	85.5	81.3	enterobacterales	L20K	4606202	4606202	1
S19_bc19	BHI-23	78.2	75.7	gammaproteobacteria	H20K	4967260	4967260	1
S20_bc20	BHI-27	43.5	39.3	enterobacterales	L20K	4437385	1072863	7
S21_bc21	BHI-31	13.7	14.8	enterobacterales	L20K	1972411	243237	12
S22_bc22	BHI-32	88.7	86.6	enterobacterales	L20K	4542884	4542884	1
S23_bc23	BHI-33	78.2	77.5	gammaproteobacteria	H20K	4886363	4837372	2
S24_bc24	BHI-35	76.6	76.2	gammaproteobacteria	H20K	4886189	4837188	2
S25_bc25	BHI-36	83	77.3	gammaproteobacteria	H20K	4967840	4967840	1
S26_bc26	BHI-38	87.9	85.9	enterobacterales	L20K	5166555	4557862	8
S27_bc27	BHI-40	84.7	84.4	gammaproteobacteria	H20K	4968674	4968674	1
S28_bc28	BHI-41	92.7	88.2	enterobacterales	L20K	4542994	4542994	1
S29_bc29	BHI-43	78.2	72.8	pseudomonadales	L20K	6066977	5627290	2
S30_bc30	BHI-44	87.1	86.8	enterobacterales	L20K	4559515	4559515	1
S31_bc31	TSA-01	63.7	52	enterobacterales	L20K	4694543	2391451	7
S32_bc32	TSA-02	68.5	72.3	enterobacterales	L20K	4543212	4543212	1
S33_bc33	TSA-03	24.2	17.2	mollicutes	L20K	2414467	103698	30
S34_bc34	TSA-06	83.1	74.5	enterobacterales	L20K	4535827	3013583	2
S35_bc35	TSA-11	NA	NA	NA	L20K	NA	NA	NA
S36_bc36	TSA-12	NA	97.6	bacteria	L20K	13572177	2626093	17
S37_bc37	TSA-17	58.9	56.1	enterobacterales	L20K	4542373	1914482	3
S38_bc38	TSA-19	96	95.9	enterobacterales	L20K	5291749	5291749	1

S39_bc39	TSA-21	71.8	65.6	lactobacillales	L20K	3684904	3514817	2
S41_bc41	negative control	NA	NA	NA	L20K	NA	NA	NA
S40_bc40	TSA-28	NA	12.1	bacteria	L20K	2693277	116004	37
S42_bc42	TSA-35	12.9	7.3	mollicutes	L20K	1358272	155123	11
S43_bc43	TSA-36	39.5	38.9	enterobacterales	L20K	4498443	898867	7
S44_bc44	TSA-37	22.6	11.3	mollicutes	L20K	5092871	271946	35
S45_bc45	TSA-39	91.9	92.2	enterobacterales	L20K	5338531	5291378	2
S46_bc46	TSA-40	NA	20.2	bacteria	L20K	3249713	223268	22
S47_bc47	TSA-45	83.9	80	enterobacterales	L20K	4559334	4559334	1
S48_bc48	negative control	NA	NA	NA	L20K	NA	NA	NA

Table S4: Nanopore results, containing information about genome assembly by Flye and BUSCO evaluation

Flye assembly and BUSCO evaluation after polishing with Racon and Medaka:

Filename	Sample	Generic(%)	predicted(%)	Database	quality	Genome size	N50	Contigs
S1_bc1	BHI-01	93.5	95.9	micrococcales	L20K	3871060	3813704	2
S2_bc2	negative control	NA	NA	NA	L20K	NA	NA	NA
S3_bc3	BHI-03	92.7	90.4	enterobacterales	L20K	4544135	4544135	1
S4_bc4	BHI-04	95.2	96.6	enterobacterales	L20K	4543898	4543898	1
S5_bc5	BHI-05	74.2	72.9	pseudomonadales	L20K	5630237	5630237	1
S6_bc6	BHI-06	99.2	98.6	gammaproteobacteria	H20K	4971990	4971990	1
S7_bc7	BHI-07	100	99.7	alteromonadales	H20K	5030842	5030842	1
S8_bc8	BHI-08	99.2	99.7	alteromonadales	H20K	5030853	5030853	1
S9_bc9	BHI-09	99.2	98.9	gammaproteobacteria	H20K	4905116	4840567	3
S10_bc10	BHI-11	85.5	82.9	enterobacterales	L20K	4544642	4544642	1
S11_bc11	BHI-12	98.4	98.3	gammaproteobacteria	H20K	4779882	4779882	1
S12_bc12	BHI-13	98.4	98.9	gammaproteobacteria	H20K	4897180	4842040	3
S13_bc13	BHI-14	84.7	83.4	enterobacterales	L20K	4560883	4560883	1
S14_bc14	BHI-15	97.6	98.1	gammaproteobacteria	H20K	4891077	4842031	2
S15_bc15	BHI-17	84.7	89.2	micrococcales	L20K	3870946	3813608	2
S16_bc16	BHI-18	91.9	90.9	enterobacterales	L20K	4552488	4552488	1
S17_bc17	BHI-20	NA	NA	NA	L20K	NA	NA	NA
S18_bc18	BHI-22	87.9	85.7	enterobacterales	L20K	4606674	4606674	1
S19_bc19	BHI-23	99.2	98.6	gammaproteobacteria	H20K	4971992	4971992	1
S20_bc20	BHI-27	47.6	44.3	enterobacterales	L20K	4422469	1063719	7
S21_bc21	BHI-31	NA	16.9	bacteria	L20K	1962668	242215	12
S22_bc22	BHI-32	91.9	90.4	enterobacterales	L20K	4544282	4544282	1
S23_bc23	BHI-33	98.4	98.7	gammaproteobacteria	H20K	4891063	4842028	2
S24_bc24	BHI-35	96.8	98.1	gammaproteobacteria	H20K	4891068	4842036	2
S25_bc25	BHI-36	99.2	98.9	gammaproteobacteria	H20K	4971965	4971965	1
S26_bc26	BHI-38	92.7	90.4	enterobacterales	L20K	5160818	4559473	8
S27_bc27	BHI-40	98.4	98.9	gammaproteobacteria	H20K	4971981	4971981	1
S28_bc28	BHI-41	96	94.7	enterobacterales	L20K	4544010	4544010	1
S29_bc29	BHI-43	74.2	75	pseudomonadales	L20K	6069261	5629439	2
S30_bc30	BHI-44	92.7	92.5	enterobacterales	L20K	4560546	4560546	1
S31_bc31	TSA-01	60.5	56.4	enterobacterales	L20K	4677668	2390109	11
S32_bc32	TSA-02	79	77.2	enterobacterales	L20K	4544942	4544942	1
S33_bc33	TSA-03	29	17.7	mollicutes	L20K	2403309	103579	30
S34_bc34	TSA-06	90.3	75.9	enterobacterales	L20K	4536792	3014733	2
S35_bc35	TSA-11	NA	NA	NA	L20K	NA	NA	NA
S36_bc36	TSA-12	NA	97.6	bacteria	L20K	13555304	2627155	16
S37_bc37	TSA-17	71.8	59.8	enterobacterales	L20K	4540850	1914425	3
S38_bc38	TSA-19	99.2	98.8	enterobacterales	L20K	5293132	5293132	1

S39_bc39	TSA-21	72.6	69.6	bacilli	L20K	3687948	3517710	2
S41_bc41	negative control	NA	NA	NA	L20K	NA	NA	NA
S40_bc40	TSA-28	NA	11.3	bacteria	L20K	2682459	115791	37
S42_bc42	TSA-35	9.7	5.3	mollicutes	L20K	1349019	153998	11
S43_bc43	TSA-36	46	44.3	enterobacterales	L20K	4492183	898017	7
S44_bc44	TSA-37	28.2	13.9	mollicutes	L20K	5054383	270554	36
S45_bc45	TSA-39	97.6	95.7	enterobacterales	L20K	5340480	5293410	2
S46_bc46	TSA-40	NA	16.1	bacteria	L20K	3234167	222004	22
S47_bc47	TSA-45	87.1	86.1	enterobacterales	L20K	4560968	4560968	1
S48_bc48	negative control	NA	NA	NA	L20K	NA	NA	NA

Table S5: Nanopore results of Flye assembly and BUSCO evaluation after polishing with two rounds of Racon and one round of Medaka.

Table of MiGA assigned classification of isolates on phylum level:

Table S6: Assignment of samples based on phylum level classification given by MiGA.l

Psudomonodata	Actinomycetota	Bacillota	Inactive
S3_bc3	S1_bc1	S39_bc39	S17_bc17
S4_bc4	S15_bc15		S21_bc21
S5_bc5			S33_bc33
S6_bc6			S36_bc36
S7_bc7			S40_bc40
S8_bc8			S42_bc42
S9_bc9			S43_bc43
S10_bc10			S44_bc44
S11_bc11			S46_bc46
S12_bc12			
S13_bc13			
S14_bc14			
S16_bc16			
S18_bc18			
S19_bc19			
S20_bc20			
S22_bc22			
S23_bc23			

S24_bc24			
S25_bc25			
S26_bc26			
S27_bc27			
S28_bc28			
S29_bc29			
S30_bc30			
S31_bc31			
S32_bc32			
S34_bc34			
S37_bc37			
S38_bc38			
S45_bc45			
S47_bc47			

GTDB-k result text file

C:\Users\turne\OneDrive\Skrivebord\MASTER FILES\gtdb-k.result.txt

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway