NTNU

Norwegian University of
Science and Technology

# Intragenomic Variation Among 16S rRNA Copies in *Vibrio*

Significance of Lifestyle

Line Strand Karlsholm

Biotechnology (5 year)
Submission date: November 2017
Supervisor: Ingrid Bakke, IBT
Co-supervisor: Ragnhild Inderberg Vestrum, IBT

Norwegian University of Science and Technology
Department of Biotechnology and Food Science

## Aknowledgements

This master project was performed at the Department of Biotechnology and Food Science at the Norwegian University of Science and Technology (NTNU) in Trondheim.

My main supervisor for this thesis was Ingrid Bakke, who I would like to thank for all her invaluable feedback and support throughout my work on this thesis. Thank you for helping me improve my thesis and maintain my motivation throughout my time working on this project, your help has gone beyond what can be expected. I would also like to thank my co-supervisor Ragnhild Inderberg Vestrum for all the great feedback and the help I got learning new methods in the lab.

I would like to thank my friends, lab- and classmates for all their support and companionship through my time working on this project. I would also like to thank my family, whose support enabled me to complete my work on this thesis.

**Trondheim, November 2017** **Line Strand Karlsholm**

## Abstract

Intragenomic heterogeneity among 16S rRNA gene copies within a genome with several rRNA operons is a phenomenon observed in several bacteria. It has been suggested that these different 16S rRNA gene variants might have functional differences, and therefore expressed differently when the bacteria are in different environments. A recent gnotobiotic fish experiment yielded results indicating a difference in the abundance of the 16S rRNA gene variants at the DNA level for a *Vibrio* strain when living with different lifestyles. This subject, which has not been subjected to much research previously, was examined in this study.

In this thesis, the presence and abundance of intragenomic 16S rRNA gene copy variants of the *Vibrio anguillarum* strain HI610, the *Vibrio campbellii* strain BB120 and the *Vibrio* sp. strain RD5-30 were investigated, both at the RNA and DNA level. The 16S rRNA gene sequence variants for the three *Vibrio* strains were mapped, and their abundances in different lifestyles compared. A gnotobiotic cod larvae experiment and a growth experiment in M65 medium were conducted to compare *Vibrio* cells colonizing Atlantic cod (*Gadus morhua*) larvae, living planktonic in the rearing water of the fish, as well as growing in exponential and stationary phases of growth in M65 medium. DNA and RNA was extracted from samples taken from these experiments, and cDNA synthesized from RNA. Denaturing gradient gel electrophoresis (DGGE) and Illumina sequencing of 16S rRNA amplicons were the methods used to investigate potential 16S rRNA variants and their relative abundances for bacteria growing in different environments.

DGGE analysis proved to be unsuited for analyzing intragenomic sequence heterogeneity, as many bands migrating differently in the denaturing gradient turned out to represent identical nucleotide sequences. Illumina sequencing of 16S rRNA amplicons yielded only one sequence variant of the 16S rRNA gene in HI610 and BB120. For RD5-30, 17 sequence variants were found, involving combinations of 6 V1-region and 10 V3-region sequence variants. Some of these sequences variants might represent chimeras, and could either be the result of PCR errors or a biological recombination process. RD5-30 exhibited significant differences in relative abundance of the 16S rRNA gene sequence variants between sample groups, both at the DNA and RNA level. This indicated a regulation of the abundance of 16S rRNA gene variants at both

levels. Future research should focus on identifying potential functions of intragenomic hypervariable 16S rRNA heterogeneity and the mechanism responsible for the differences in abundances of these variants at the RNA and DNA level between different lifestyles.

## Sammendrag

Intragenomisk heterogenitet blant 16S rRNA-genkopier innenfor et genom med flere rRNA operon, er et fenomen som har blitt observert hos flere bakterier. Det har blitt foreslått at disse ulike 16S rRNA-genvariantene kan ha funksjonelle forskjeller, og at de derfor blir uttrykt ulikt når bakterier befinner seg i ulike miljø. Et nylig gnotobiotisk fiskeforsøk ga resultater som indikerte en forskjell i mengden av 16S rRNA-genvarianter på DNA-nivå for en *Vibrio*-stamme når den levde med ulike livsstiler. Dette emnet, som det ikke har blitt forsket mye på før, ble nærmere undersøkt i denne oppgaven.

I denne oppgaven var tilstedeværelsen og mengden av intragenomiske 16S rRNA-gen-kopivarianter for *V. anguillarum* stammen HI610, *V. campbellii* stammen BB120 og *Vibrio*-stammen RD5-30 undersøkt, på både RNA- og DNA-nivå. 16S rRNA-sekvensvarianter i de tre *Vibrio* stammene ble kartlagt, og mengden av disse variantene for ulike livsstiler ble sammenlignet. Et gnotobiotisk eksperiment med torskelarver og et veksteksperiment i M65-medium ble utført for å sammenligne *Vibrio*-celler som koloniserte torskelarver av atlanterhavstorsk (*G. morhua*), som levde planktonisk i oppdrettsvannet til fisken, samt voksende i eksponentiell- og stasjonærfase i M65-medium. DNA og RNA ble ekstrahert fra prøver tatt fra disse eksperimentene, og cDNA ble syntetisert fra RNA. DGGE og Illumina-sekvensering av 16S rRNA-amplikon var metodene som ble brukt for å undersøke potensielle 16S rRNA varianter og deres relative mengder for bakterier som vokste i ulike miljø.

DGGE analyse viste seg å være en uegnet metode for analyse av intragenomisk sekvensheterogenitet ettersom mange av båndene som migrerte annerledes i denatureringsgradienten viste seg å representere identiske sekvenser. Illumina-sekvensering av 16S rRNA amplikon ga kun en sekvensvariant av 16S rRNA-genet for HI610 og BB120. For RD5-30 ble det funnet 17 sekvensvarianter som alle hadde kombinasjoner av 6 V1-region varianter og 10 V3-region sekvensvarianter. Noen av disse sekvensvariantene representerer kanskje kimære sekvenser, og kan enten være et resultat av PCR-feil eller en biologisk rekombinasjonsprosess. RD5-30 viste signifikante forskjeller i relativ menge av 16S rRNA-sekvensvarianter mellom prøvegrupper, både på DNA- og RNA-nivå. Dette indikerte en regulering av mengden av 16S rRNA-genvarianter på begge nivåer. Fremtidig forskning bør

fokusere på å identifisere potensielle funksjoner av intragenomisk hypervariabel 16S rRNA heterogenitet og mekanismen som står bak forskjellene i mengdene av disse variantene på både RNA- og DNA-nivå mellom ulike livsstiler.

# Table of Contents

# 1 INTRODUCTION

## 1.1 Organization and function of rRNA operons in bacteria

Ribosomes are the site of protein synthesis in cells, and translate mRNAs into a sequence of amino acids (Stoddard et al., 2015). The ribosomes of prokaryotes consist of a small (30S) subunit and a large (50 S) subunit (Fukuda et al., 2016). The small 30S ribosomal subunit in bacteria consist of about 22 different ribosomal proteins and 16S ribosomal RNA (rRNA) (Maeda et al., 2015). The large subunit contains 31 different ribosomal proteins and two rRNAs (5S and 23S). Transfer RNAs (tRNAs) also interact with the ribosomes during protein synthesis (Snustad and Simmons, 2012). The function of the rRNAs in the ribosome includes playing a role in important catalytic activities such as the formation of peptide bonds and various interactions with tRNAs, mRNAs, ribosomal proteins, and translational factors (Maeda et al., 2015). The conserved secondary and tertiary structure of the rRNAs such as for 16S rRNA (Figure 1.1) have an important structural and functional role in the ribosomes (Snustad and Simmons, 2012, Wimberly et al., 2000, Shajani et al., 2011).

In bacteria, the rRNA genes are organized in operons that contain the 16S, 23S and 5S rRNA genes, as well as at least one tRNA in their spacer regions (Condon et al., 1995, Klappenbach et al., 2000, Acinas et al., 2004). These rRNA operons also contain at least one promoter region (Maeda et al., 2015). An example of the composition of the rRNA operon in bacteria, from one of the seven operons in *Escherichia coli*, can be seen in Figure 1.2. The entire operon is transcribed from its DNA template in one piece that goes through posttranscriptional cleavage into the individual rRNAs and tRNAs (Snustad and Simmons, 2012).

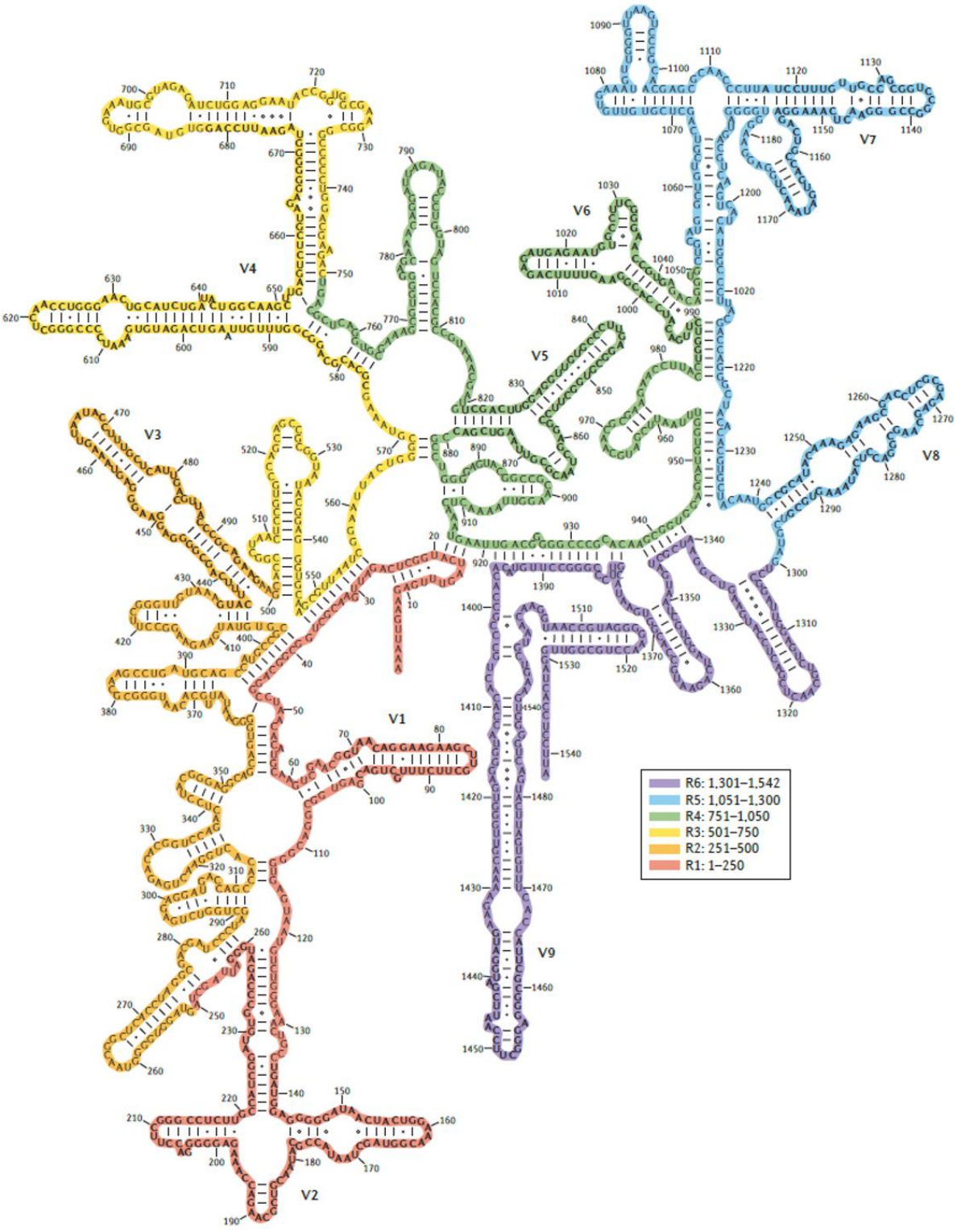


**Figure 1.1:** The secondary structure and nucleotide sequence of the 16S rRNA gene in *Escherichia coli*. The variable regions are marked V1 to V9 (Yarza et al., 2014).

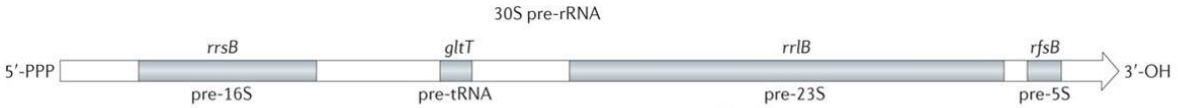


**Figure 1.2:** The *rrnB* operon from *E. coli*, which consists of the 16S rRNA gene, a tRNA, the 23S rRNA gene and the 5S rRNA gene, in that order. The figure is adapted from Mackie (2013).

The mechanism for the activation of the genes in the rRNA operon involves the binding of an RNA polymerase, and the number of possible transcripts of the genes from one operon is limited by the transcriptional rate of the polymerase. There are also limitations related to how many transcripts can initiated at one operon promotor. Transcription of the rRNA operon when the bacteria is rapidly growing can make up 70% of all transcription activity in the cell (Klappenbach et al., 2000). The number of ribosomes in a cell is proportional to the growth rate of the bacteria (Maeda et al., 2015). Translational power has been found to be higher for bacteria with a high number of rRNA operons (Dethlefsen and Schmidt, 2007).

Bacteria contain between 1 and 15 copies of the rRNA operon in their genome (Harth et al., 2007). Modern sequencing technology and genomics has led to the mapping of the copy number of rRNA operons in many bacterial species, and this information has been stored in a database found at https://rrndb.umms.med.umich.edu (Stoddard et al., 2015, Regents of the University of Michigan, 2017). Recent analyses have estimated the average copy number of the operon per bacteria to be about 3.61 (Sun et al., 2013). Variations in copy number have been observed for different strains of the same species (Acinas et al., 2004). A higher copy number of the operon correlates with a higher number of ribosomes and faster growth rates (Jensen et al., 2009, Maeda et al., 2015). A high copy number of rRNA operons reflects an ecological strategy of rapidly responding to favorable conditions – a copiotrophic or r-strategic lifestyle. Conversely, a low rRNA copy number is believed to correspond to an oligotrophic or K-strategic lifestyle, where the bacteria live in environments with low nutrient availability where it is important to allocate resources efficiently (Klappenbach et al., 2000, Koch, 2001, Stoddard et al., 2015, Verschuere et al., 1997). In *E. coli*, only five of its seven rRNA operons are essential for optimal growth, but it is believed that having all seven rRNA operons gives the bacteria an advantage in situations with changes in nutrient availability or temperature where rapid adaption to new environments are needed (Condon et al., 1995).

## 1.2 The 16S rRNA gene as a marker for microbial diversity and phylogeny

Ever since the late 19th century, microbiologists have been trying to find a way to determine the evolutionary relationships among bacteria. The morphology of the bacteria was initially used to decide bacterial phylogeny, but it could not be used to effectively distinguish phyla as bacteria have a much simpler morphology than eukaryotes (Woese, 1987). In the 1980's, researchers started moving away from using morphology to identify relationships between bacteria, and towards comparing stable parts of the bacterial genomes instead (Woese et al., 1990). 5S rRNA and tRNA gene sequences were initially used to create phylogenetic trees (Gray et al., 1984), before it was eventually decided that the part of bacterial genomes best suited for this purpose was the 16S rRNA gene, as originally suggested by Woese and Fox (1977). An overview of the secondary structure of the 16S rRNA gene as it appears in *E. coli* can be seen in Figure 1.1.

There are many reasons why the 16S rRNA gene was chosen as the primary target for bacterial phylogeny and taxonomic classification. Firstly, the 16S rRNA gene consists of both variable and conserved regions and has a size of about 1550 nucleotides (1542 nt in *E. coli*) (Clarridge, 2004, Yarza et al., 2014). There are ten conserved regions of the 16S rRNA gene, which have been highly conserved throughout the evolution of bacteria. The conserved regions serve as sites for so-called universal primers that target most bacteria, making it easy to amplify 16S rRNA regions for the bacteria present in a sample (Fukuda et al., 2016). There are nine hypervariable regions, named V1 to V9 (Mizrahi-Man et al., 2013). The variable regions of the 16S rRNA gene can be used to separate bacterial taxa and differentiate bacteria at the genus, or to some extent, even at the species level (Glaeser and Kampfer, 2015). It is possible to design primers that are specific to a certain taxon (e.g. genus or species) by targeting more specific gene sequences (Fukuda et al., 2016). Originally, a threshold of >3% 16S rRNA gene sequence dissimilarity was used to classify bacteria at the species level, but this threshold has been lowered to a 1-1.3% sequence dissimilarity (Stackebrandt and Goebel, 1994, Pei et al., 2010).

The 16S rRNA gene has been thought to not be greatly affected by horizontal gene transfer (HGT; transferring of genes between organisms that is not from parent to offspring) (Acinas et al., 2004, Alberts et al., 2015). The 16S rRNA gene has been researched and sequenced

extensively, and there are large databases of gene sequences such as GenBank and the Ribosomal Database Project (RDP), which contains over 3 million 16S rRNA sequences from different bacterial strains representing both cultivated microbes and environmental samples (Benson et al., 2005, Maidak et al., 2001, Michigan State University, 2016). All these different features of the 16S rRNA gene contribute to making it a suitable target to map changes throughout evolution and subsequent phylogeny mapping of bacteria (Acinas et al., 2004). The amplification of parts of the gene is a very useful tool in analyses of microbial communities, e.g. for assessing diversity and taxonomic assignments (Amato, 2017). Using 16S rRNA to classify bacteria at a species level is also very useful in identifying pathogens in clinical microbiology (Sul et al., 2011, Clarridge, 2004).

## 1.3 History of bacterial analysis – from cultivation to high throughput sequencing

In early days of microbiology research, cultivation of bacteria along with different analyses of their traits was required to be able to identify and classify them (Amann et al., 1995). Cultivation of environmental samples on plates were used to find the number of living cells in a sample, but when the same samples were investigated through microscopy the number of living cells were orders of magnitude higher (Amann et al., 1995). This phenomenon was named the "great plate count anomaly" by Staley and Konopka (1985). The inability to cultivate more than a fraction of the microorganisms in a sample was a large obstacle for research within the field of microbial ecology (Head et al., 1998). With new suggestions for microbial phylogeny, including using the 16S rRNA gene to classify bacteria, new methods were developed to help identify uncultured bacteria (Head et al., 1998, Woese and Fox, 1977, Woese et al., 1990).

The new methods developed to analyze microorganisms using the 16S rRNA gene included fluorescent *in-situ* hybridization (FISH), terminal restriction fragment length polymorphism (T-RFLP) and denaturing gradient gel electrophoresis (DGGE) (Head et al., 1998, Fukuda et al., 2016). The latter method has been used a lot in microbial ecology to analyze environmental samples (Head et al., 1998). Through casting a gel with a denaturing gradient during DGGE, PCR amplified 16S rRNA (DNA) amplicons can be separated based on their nucleotide sequence, as differently composed amplicons will denature at different denaturing percentages

when wandering through the gel (Fukuda et al., 2016, Head et al., 1998). Small variations in sequence lead to distinct bands as DNA molecules melt at different denaturing percentages (Fukuda et al., 2016).

DNA sequencing methods have largely taken over for the other methods, especially in areas of research where a large number of bacterial need to be identified, such as analyzing environmental samples in microbial ecology (Acinas et al., 2004). One much used first generation sequencing method that was also used to sequence the first whole bacterial genome was developed by Sanger et al. (1977) (Fleischmann et al., 1995). In this approach, fluorescent dideoxynucleotides (ddNTPs) terminated sequences are added randomly instead of dNTPs in a polymerase chain reaction (PCR) that contains both, and with sufficient amount of the template the sequence will be terminated with fluorescent ddNTP nucleotides at every possible nucleotide position in the sequence. The resulting terminated sequences are then separated by size through capillary electrophoresis, and the fluorescence of the ddNTPs are detected by laser excitation (Goodwin et al., 2016, Heather and Chain, 2016). The first generation sequencing machines had the advantage of long sequence reads of about one kilobase (kb) (Heather and Chain, 2016).

The next generation of DNA sequencing introduced high-throughput methods capable of massively parallel sequencing, allowing large amounts of DNA to be sequenced in one run (Heather and Chain, 2016). Among the first of these methods was 454 pyrosequencing, much used in analyses of microbial communities by sequencing 16S rRNA gene amplicons (Goodwin et al., 2016, Fukuda et al., 2016). The 454 pyrosequencing method utilizes sequencing by synthesis, and relies on the detection of released pyrophosphates upon the addition of nucleotides (Heather and Chain, 2016). It was the preferred method because of its long sequencing reads of up to 500 base pairs (bp). In recent years, 454 pyrosequencing has been replaced by Illumina sequencing for use in studies of microbial communities, as Illumina sequencing read lengths have increased and has a much higher output (Mizrahi-Man et al., 2013). Illumina sequencing is now the main method of choice for microbial ecology, and is used for metagenomics and transcriptomics (Amato, 2017, Wu et al., 2015). Illumina sequencing of 16S rRNA amplicons is currently the most widely applied approach to study

microbial diversity and phylogeny (Amato, 2017). It is also a sequencing technology now commonly used for whole genome sequencing of any species (Illumina, 2016).

Illumina sequencing allows for millions of sequence reads per run (Goodwin et al., 2016). The first step of Illumina sequencing is PCR with primers that contain adapter sequences that will allow the template DNA to hybridize to complimentary adapter sequences on the flow cell on which the Illumina sequencing takes place. The hybridized templates are amplified into clonal clusters through bridge amplification, and their sequences are found by incorporating fluorescently labeled nucleotides (Illumina, 2016). This happens much in the same way as in Sanger sequencing, only the terminating fluorescently nucleotides are only reversibly bound during Illumina sequencing (Heather and Chain, 2016, Illumina, 2016). After the fluorescence of the nucleotide added to a cluster is recorded, the fluorescent nucleotide is cleaved and so the next one in the sequence can be added (Goodwin et al., 2016). Even though sequence reads of Illumina sequencing can only be a maximum of 300 bp long with the newest Illumina MiSeq machines, the sequencing method can sequence longer fragments because it enables paired end reads (Goodwin et al., 2016, Heather and Chain, 2016). Paired end reads are achieved by first sequencing one flow cell bound strand, before it is washed away and the remaining complimentary strand is sequenced (Heather and Chain, 2016). Provided the forward and reverse reads have some overlap, they can then easily be aligned into a longer sequence read (Goodwin et al., 2016, Illumina, 2016).

## 1.4 Intragenomic heterogeneity of the 16S rRNA gene

The sequence of the rRNA genes has been found to vary between the rRNA operon copies within one genome. In particular, variation between 16S rRNA genes within the same genome has been subject to research (Mylvaganam and Dennis, 1992, Cilia et al., 1996, Nübel et al., 1996). One of the first discoveries of heterogeneity among the 16S rRNA operons of a species was found while sequencing the seven rRNA operons in a strain of *E. coli* (Cilia et al., 1996). Later, Dahllöf et al. (2000) used DGGE as a method to examine 16S rRNA heterogeneity within bacteria, and found several bands for the 16S rRNA genes of single bacterial strains in 12 out of 14 isolates from a marine rock, with each band supposedly corresponding to a discrete 16S rRNA gene variant.

A bacterium that has more than one rRNA operon in their genome is only rarely found to have identical 16S rRNA genes, though the variation between gene copies is usually very small (Vetrovsky and Baldrian, 2013). A variation of more than 1% between 16S rRNA genes in a bacterium was only found in 24 genomes after the analysis of the genomes of 883 prokaryotes by Pei et al. (2010). Vetrovsky and Baldrian (2013) found that the 16S rRNA genes of only 2.4% of bacterial genomes contained more than 1% sequence dissimilarity After analyzing 2143 genomes, Tian et al. (2015) found that 28 of these contained more than 2% 16S rRNA gene heterogeneity. However, Tian et al. (2015) also argues that intragenomic heterogeneity might be underestimated because high throughput sequencing methods might fill sequencing gaps with common 16S sequences and in that way "ignore" heterogeneity. Some bacteria do exhibit a larger variation between 16S rRNA copies, caused by extensive intragenomic variation in hypervariable regions of their 16S rRNA genes (Sun et al., 2013). One such bacteria with a large degree of variation is *Thermoanaerobacter tengcongensis* with 11.6% divergence between its 16S rRNA copies (Acinas et al., 2004). Generally, most of the intragenomic heterogeneity is found in the V1- and V6-regions and least heterogeneity is found in the V4- and V5-regions of the 16S rRNA gene (Sun et al., 2013).

According to the complexity hypothesis, the 16S rRNA gene should only to a small degree be affected by HGT. The complexity hypothesis was introduced by Jain et al. (1999). This hypothesis suggests that informational genes, meaning genes involved in transcription and translation, are more rarely subject to HGT than housekeeping genes, which are genes with conserved functions that are also referred to as operational genes (Glaeser and Kampfer, 2015). However, Wang and Zhang (2000) elaborated on this hypothesis in their simplified complexity hypothesis. They suggest that short segments of these informational genes, such as the 16S rRNA gene, might be subject to HGT between species of bacteria. They found evidence to support that the segment of the rRNA gene that was horizontally transferred were part of hairpin loops in the secondary structure of the gene, more specifically the stems of these hairpins. This transfer of DNA segments usually seems to conserve the secondary structure of the gene and presumably also its function, as the conservation of the stem-loop structure implies that it is important to the function of the gene (Moreno et al., 2002, Harth et al., 2007). This horizontal transfer of genes might take place through bacteriophages, which are abundant in marine environments, and bacteriophages that transduce genes to the bacteria *Vibrio parahaemolyticus*

have been isolated (Harth et al., 2007). Tian et al. (2015) found that HGT between 16S rRNA genes only occurs at an intraspecies or intragenus level.

A factor that should prevent heterogeneity of the 16S rRNA gene is gene conversion (Gonzalez-Escalona et al., 2005). Gene conversion in the rRNA genes of prokaryotes, as described by Liao (2000), involves a short segment of DNA being copied between genes that have been duplicated at some point but since diverged in function (paralogs). The end result of gene conversion is the homogenization of the gene as a type of concerted evolution (Gonzalez-Escalona et al., 2005). It has been speculated that these heterogeneous 16S rRNA genes within the same genome have some difference in function that bestows some sort of advantage on the host (Pei et al., 2010). The fact that the different 16S rRNA genes of a genome have not been entirely homogenized as a result of conversion and concerted evolution supports this theory (Jensen et al., 2009).

Looking more into the differential expression of heterogeneous 16S rRNA variants, indications have been found that the different variants are expressed differentially throughout the life cycle of bacteria, in the bacterium *Streptomyces coelicolor* (Kim et al., 2007). Duan et al. (2014) also found evidence supporting this in the bacterium *Pseudomonas* sp. UW4, as they found that different rRNA promotors in this bacterium are expressed preferentially at different life stages, temperatures and different levels of nutrient availability. It was suggested that by Nübel et al. (1996) that different growth conditions might utilize different ribosome compositions. When researching the expression of the 7 rRNA operons of *E. coli* in four different growth media, Maeda et al. (2015) found that the rrnE operon promotor had the highest activity in all growth conditions, while the activity of the six other rRNA operon promoters varied with culturing conditions.

Even when the secondary structure of the 16S rRNA gene appears intact, small variations in nucleotide sequence are not necessarily biologically insignificant (Jensen et al., 2009). Some research has been made to attempt to discover how the different variants of the 16S rRNA gene might differ functionally. Temperature has been shown to significantly affect the expression of the different 16S variants in *Haloarcula marismortui*, and their secondary structures might be

stable at different temperatures (López-López et al., 2007). Elongated helixes in the gene's secondary structure might be an adaption to the increased pressure that is found in for instance deep-sea environments (Lauro et al., 2007). In their investigation into characteristics of the 16S rRNA genes in a strain of *Vibrio splendidus*, Jensen et al. (2009) suggested that the existence of heterogeneous and numerous 16S rRNA operons in *V. splendidus* could be important adaptive qualities for the structural and functional dynamics of a bacterium with pathogenic, symbiotic and free-living lifestyles. Small sequence variations in the rRNA operons, down to the single nucleotide substitution level, might radically alter the function of a ribosome, as shown for *E. coli* (Kolmsee et al., 2011). Different 16S rRNA sequence variants may affect ribosome assembly and function, e.g. through interactions with ribosomal proteins (Mayerle and Woodson, 2013, Sapag et al., 1990). More research is needed on what selective factors might affect 16S rRNA gene composition within bacteria (López-Pérez et al., 2013).

In summary, there are several findings that indicate some bacteria differentially express their heterogeneous 16S rRNA gene variants in different life stages or in different environments. This, along with the knowledge that certain parts of the 16S rRNA gene interact with ribosomal proteins, is a strong indication that the different 16S rRNA gene variants can make functionally different ribosomes. It has been speculated that intragenomic variation between 16S rRNA copies can be related to complex life strategies in bacteria, however, there has not yet been much research done in this area (Kim et al., 2007).

**Implications of intragenomic 16S rRNA gene heterogeneity in diversity and phylogeny studies**

The presence of several copies of the rRNA operon and the intragenomic heterogeneity among the 16S rRNA copies in many bacterial species, may lead to an overestimation of diversity in microbial community analyses (e.g. DGGE and high throughput sequencing) (Vetrovsky and Baldrian, 2013, Sun et al., 2013). Taxa with high rRNA operon copy numbers, such as Gammaproteobacteria, have their abundance overestimated and the taxa with a low copy number tend to have their abundance underestimated (Vetrovsky and Baldrian, 2013). Acinas et al. (2004) analyzed bacterial genomes, and estimated the overestimation of diversity to be at least 2.5 times the actual value. Nine years later, Sun et al. (2013) presented a new estimate of

overestimation when using full-length 16S rRNA genes of 123.7%, in a study of more than 2000 complete genomes from 1212 different species.

The presence of heterogeneity among intragenomic 16S rRNA copies can also make it problematic to use the 16S rRNA gene to classify bacteria at a species and subspecies level (López-Pérez et al., 2013). Multilocus sequence analysis (MLSA), which involves the sequencing of multiple conserved housekeeping genes, has been suggested as a method to use in addition to 16S rRNA gene sequencing for a higher resolution species determination (Glaeser and Kampfer, 2015). MLSA is however limited by the need to sequence several loci (Machado and Gram, 2015). Use of alternative phylogenetic markers such as *rpoB* has also been suggested as an alternative to 16S rRNA for phylogenetic species determination. The *rpoB* gene encodes the RNA polymerase beta subunit and only exists as a single copy bacteria and has not shown any signs of heterogeneity (Dahllöf et al., 2000).

## 1.5 Features of the *Vibrio* genome and their rRNA operons

*Vibrio* is a genus of the Gammaproteabacteria class that usually colonize human or marine hosts where they often act as pathogens (Dorsch et al., 1992, Thompson and Klose, 2006). They are ubiquitous in marine environments, and require 1-2% NaCl for optimal growth (Thompson and Klose, 2006, Frans et al., 2011). Trucksis et al. (1998) discovered that the bacterium *Vibrio cholerae* has two chromosomes. Subsequent research on other species of *Vibrio* has found that all of them have genomes that are organized into two circular chromosomes of different sizes – one large chromosome of about three million base pairs in size and one smaller chromosome of about one million base pairs (Val et al., 2014, Kirkup et al., 2010, Heidelberg et al., 2000). The smaller chromosome is assumed to have its origin as a megaplasmid taken up by an ancestral species of *Vibrio* (Heidelberg et al., 2000). The two chromosomes are believed to a confer an evolutionary advantage to *Vibrio*, as they allow for quicker replication of the genome than if they were to have one large genome, thus shortening the generation time of the bacteria (Thompson et al., 2010). For *V. parahaemolyticus*, 10 of the 11 rRNA operons were found in the large chromosome and the remaining copy on the smaller chromosome (Makino et al., 2003).

*Vibrio* are known to have high genome plasticity, making them difficult to define at a species level (Thompson et al., 2010). According to https://rrndb.umms.med.umich.edu (Regents of the University of Michigan, 2017), species of *Vibrio* have between 6 and 14 rRNA operons in their genome. Within the *Vibrio* genus, the 16S rRNA gene sequence varies from <1% to 6% (Moreno et al., 2002). Most *Vibrio* strains have some intragenomic heterogeneity among their 16S rRNA operons (Gonzalez-Escalona et al., 2005). The hypervariable regions V1 and V3 of the 16S rRNA were first described as areas of potential intragenomic heterogeneity in *Vibrio* bacteria by Dorsch et al. (1992). This has since been found to be the case in e.g. *V. splendidus* 3d, where there was a 2% difference between 16S rRNA copies with most of the variability located in the V1- and V3-regions (Jensen et al., 2009). Harth et al. (2007) suggest that the V3-region of the 16S rRNA gene in *V. parahaemolyticus* is subject to recombination with other *Vibrio* species.

Horizontal gene transfer is common in *Vibrio* (Kirkup et al., 2010). As mentioned in Section 1.4, bacteriophages are common in marine environments and HGT through phages are likely to be common between species of *Vibrio* (Thompson and Klose, 2006, Harth et al., 2007). It appears that growth of *Vibrio* in their natural hosts might promote the occurrence of HGT. Evidence supporting this was reported in a study of the bacterium *Vibrio fischeri* and its host, the squid *Eupyrmna scolopes. V. fischeri* upregulated a chitin utilization pathway upon colonization of the host, indicating that *E. scolopes* provides *V. fischeri* with chitin (Thompson and Klose, 2006). Meibom et al. (2005) found that natural competence is induced in *V. cholerae* upon growth on chitin, making the bacteria more susceptible to HGT. Harth et al. (2007) suggested that high intergenomic recombination of rRNA operons between *Vibrio* species has been important for the evolution of the genus.

## 1.6 Fish microbiota

The term “microbiota” is used to refer to all the microorganisms that coexist with a host without harming it. In vertebrates, the gut is the most colonized organ, and the gut microbiota consists mostly of bacteria, and is a symbiotic relationship beneficial to both host and microbes

(Maynard et al., 2012, Sekirov et al., 2010). The microbiota of vertebrates has been found to be important for nutrient metabolism, enzyme and vitamin synthesis and immunity (Sekirov et al., 2010, Parekh et al., 2015, Fraune and Bosch, 2010). Disturbance of the human microbiota, dysbiosis, has a role in e.g. obesity and inflammatory bowel disease (IBD) (Maynard et al., 2012, Parekh et al., 2015, Konig et al., 2016).

For fish, the microbiota has been shown to share many of the same host responses to the gut microbiota with mammals (Bakke et al., 2015). The microorganisms that colonize fish are mainly bacteria (Nayak, 2010). Fish microbiota develops throughout the life cycle of the fish. Bacteria from the environment and those on the egg rapidly colonize fish once it has hatched, and they colonize the gut once the fish starts consuming the surrounding water (Llewellyn et al., 2014). The establishment and composition of the microbiota is affected by different factors, both biotic and abiotic (Sullam et al., 2012); including the feed, genetics, developmental stage, stress, and environmental factors such as ambient water microbes, temperature, salinity, as well as chemicals like antibiotics or pollutants (Llewellyn et al., 2014, Nayak, 2010, Sullam et al., 2012). A stable gut microbiota is established after the first feeding stages of the fish (Llewellyn et al., 2014).

Fish gut microbiota has been shown to upregulate various genes involved in innate immunity (Rawls et al., 2007, Rawls et al., 2004). The GI microbiota is believed to be the first line of defense for the fish against opportunistic pathogens and particularly important during early developmental stages (Llewellyn et al., 2014, Boutin et al., 2012). Furthermore, it is also considered important for development and maturation of the mucosal immune system. Other factors affected by the fish microbiota includes digestion through the production of digestive enzymes and nutrition absorption and metabolism, particularly cholesterol metabolism, but also production of vitamins like B12, amino acids, and essential growth factors (Llewellyn et al., 2014, Nayak, 2010).

As microbiology has become an important aspect of aquaculture, control of microbiology in conjunction with fish rearing is considered important to ensure the increased profitability of the aquaculture industry (Verschuere et al., 1997). For example, Atlantic cod (*Gadus morhua*) has

been considered as a candidate for aquaculture, but attempts to use it as such have been thwarted by early larval death (Rosenlund and Halldórsson, 2007, Forberg et al., 2011b). Opportunistic pathogens have been assigned a large part of the blame for this problem (Vadstein et al., 1993). Research is now centering on controlling the microbiology of the fish so that an advantageous microbiota can be achieved and sustained, instead of attempts to keep bacterial numbers in the rearing water low (Forberg et al., 2011b). Probiotics are microorganisms added to food or rearing water to increase host viability, and have been suggested as a potential tool for achieving this (Llewellyn et al., 2014). Another potential strategy is to control the microbes in the environment of the fish, meaning their rearing water, by using matured water which is treated in a way that promotes the growth of non-opportunistic bacteria (Skjermo et al., 1997). To be able to use these techniques effectively, more information is needed about the formation of the fish microbiota and on host-microbe interactions in early larval development and interactions between the fish and the bacteria in its environment (Forberg et al., 2011b, Bakke et al., 2015).

A valuable tool to study host-microbe interactions is gnotobiotic fish experiments (Llewellyn et al., 2014). In a gnotobiotic experiment, the fish are cultured without any microorganisms – in axenic conditions – or with the known microbiota composition. This method has been used to do research on the role of the microbiota in fish immunity and nutrition, and to study microbial interactions in the gut (Nayak, 2010). The most used model organism for gnotobiotic fish experiments is the zebrafish (*Danio rerio*) (Rawls et al., 2004). An axenic model for cod (*G. morhua*) larvae has also been developed and applied for studying host responses to bacteria (Forberg et al., 2011a).

## 1.7 Previous experiment as basis for this thesis

A gnotobiotic fish experiment with Atlantic cod (*G. morhua*) larvae was performed at NTNU during spring 2014 (R. Vestrum, unpublished results). Germ-free cod larvae were colonized with a *Vibrio* strain (RD5-30) previously isolated from cod larvae (Fjellheim et al., 2010). DNA was extracted from water and fish samples at 3, 8, 13 and 16 days post hatching. The V3-region was amplified by PCR for all samples and the PCR products were analyzed on a DGGE gel (Figure 1.3).

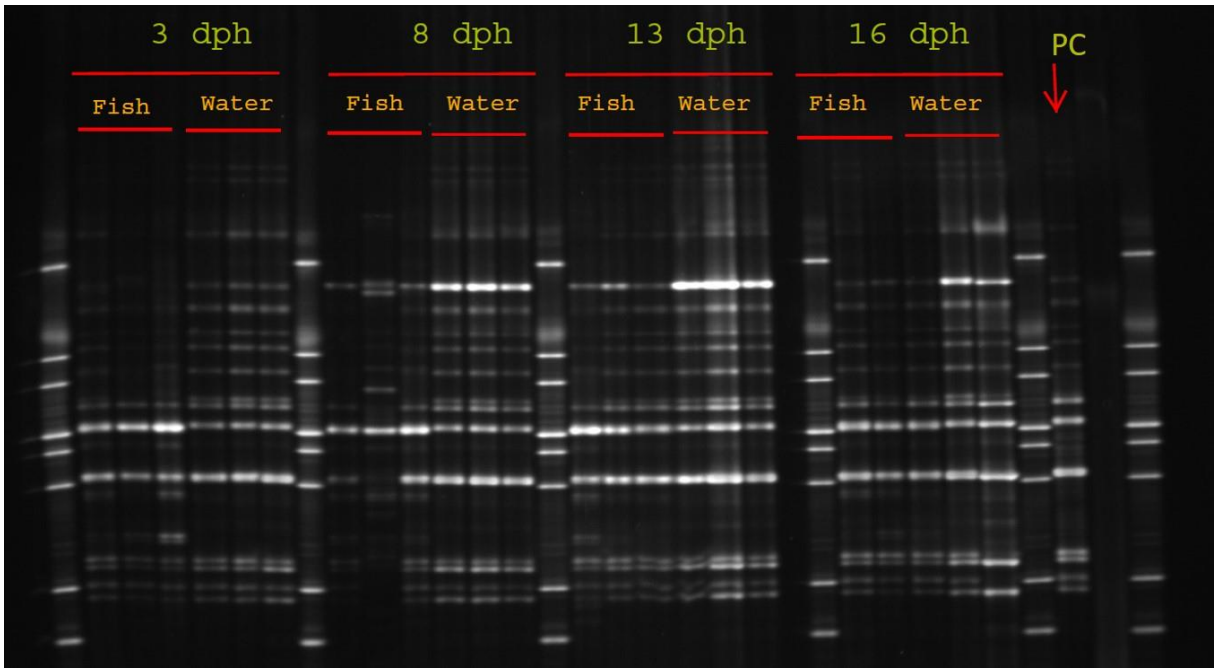


**Figure 1.3:** DGGE gel (8% acrylamide, 35-55% denaturing gradient) with PCR products representing the 16S rRNA V3-region obtained from DNA from cod (*G. morhua*) larvae that had been colonized by *Vibrio* sp. RD5-30 and rearing water at 3, 8, 13 and 16 days post-hatch. Lane marked PC contains DNA sample from pure culture of RD5-30 (R. Vestrum, unpublished results).

Surprisingly, even though the PCR products represented only one bacterial strain, as much as 14 bands were observed for each sample. Moreover, they were positioned throughout the denaturing gradient, indicating a large sequence variation between DGGE bands. There also seemed to be a difference in band pattern regarding the relative abundance of bands between the rearing water and fish DGGE profiles. This trend seems especially prevalent in the samples taken 13 and 16 days post-hatch. A possible interpretation of these observations is that this *Vibrio* strain contains extremely large intragenomic sequence variation in the 16S rRNA V3-region, and that as much as 14 different sequence variants exist. Moreover, the DGGE analysis indicates that the abundance of the sequence variants differs in the RD5-30 genome when the bacteria live planktonic in the water and when it is colonizing the fish larvae. This would have some very interesting implications, namely that different life strategies select for different abundances of the sequence variants at the DNA level.

## 1.8 Hypothesis and objectives

The hypothesis of this thesis is that different intragenomic 16S rRNA gene variants have functional differences, and that the relative expression of these gene variants in bacteria will vary according to their environment or lifestyle. To investigate this hypothesis, three different species of *Vibrio* – the *Vibrio* sp. isolate RD5-30, the *V. Anguillarum* strain HI610 and the *V. campbellii* strain BB120 – will be examined. The first objective is to map the different 16S rRNA gene variants for each strain, looking specifically at the V1- and V3-regions by investigating sequence variation of 16S rRNA amplicons by DGGE and Illumina sequencing. The second objective is to investigate the relative abundances of these gene variants under different environmental conditions and lifestyles, at both the DNA and RNA level. The different lifestyles investigated in this thesis are colonizing a fish host (*G. morhua*) and planktonic living in fish rearing water, as well as exponential and stationary phases of growth in the liquid growth medium M65.

# 2 MATERIALS AND METHODS

Two main experiments were conducted for this thesis to investigate intragenomic 16S rRNA gene sequence heterogeneity. The first experiment was a gnotobiotic fish experiment to investigate intragenomic heterogeneity in three *Vibrio* strains colonizing the fish (*G. morhua*) or its rearing water. The other experiment involved the sampling of the two *Vibrio* strains RD5-30 and BB120 in exponential and stationary phases of growth in liquid M65 culturing medium. The samples from the gnotobiotic fish experiment were subjected to DGGE analysis, while the samples from both experiments were analyzed by Illumina amplicon sequencing of the V1- and V3-regions of the 16S rRNA gene.

## 2.1 Biological materials

Species in the *Vibrio* genus are known to have high rRNA operon copy numbers (Regents of the University of Michigan, 2017). Species in this genus have also shown intragenomic heterogeneity and hypervariable regions in their 16S rRNA genes (Jensen et al., 2009). Based on these characteristics, three species of *Vibrio* were used in this thesis: *Vibrio anguillarum* strain HI610, *Vibrio campbellii* strain BB120 and *Vibrio* sp. isolate RD5-30 were chosen to further investigate intragenomic 16S rRNA heterogeneity. These strains have all been found to colonize *G. morhua* (Fjellheim et al., 2010). HI610 and BB120 are well known pathogens colonizing marine fish (Ronneseth et al., 2017, Yang et al., 2017). RD5-30, on the other hand, is a probiotic candidate isolated from *G. morhua* larvae (Fjellheim et al., 2010). The three *Vibrio* strains used will be referred to by their strain names – RD5-30, HI610 and BB120 – for the remainder of the thesis.

## 2.2 Gnotobiotic fish experiment

### 2.2.1 Experimental design

A gnotobiotic fish experiment was conducted with Atlantic cod (*G. morhua*) larvae and the three strains of *Vibrio* – HI610, RD5-30 and BB120. The *G. morhua* eggs were hatched under sterile conditions, allowing for gnotobiotic rearing of the cod larvae. There were five groups of flasks (Figure 2.1). Three of these groups consisted of cod larvae paired with one *Vibrio* strain and one group was an axenic control group containing cod larvae but no added bacteria. These four groups all had four replicates each. The remaining group of flasks contained only one strain of *Vibrio* but no cod larvae, with one replicate per strain.

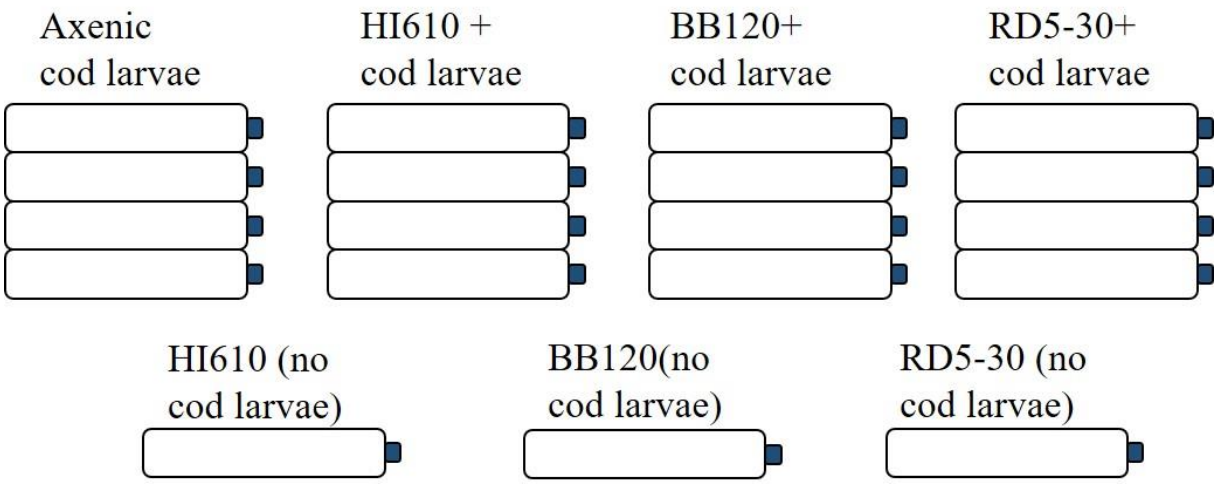


**Figure 2.1:** The setup of the flasks used in the gnotobiotic cod larvae experiment.

The protocol for disinfection of eggs of Atlantic cod larvae (*G. morhua*) was adapted from and Forberg et al. (2011a) and Salvesen et al. (1997). The Atlantic cod larvae were received from Havbruksstasjonen i Tromsø AS, a commercial hatchery, at a temperature of 5°C. The eggs were poured gently into a sieve and while in the sieve transferred to 2 liters filtered autoclaved seawater (FASW, Appendix A) containing 400 ppm glutaraldehyde. The sieve containing the eggs was gently stirred for 10 minutes. The eggs were then rinsed in 3 x 1 L fresh FASW. The eggs were stored at approximately 5°C for an hour before the same procedure was repeated under completely sterile conditions. The eggs were then transferred gently to petri dishes containing FASW with 10 ppm each of rifampicin and ampicillin. The petri dishes with eggs

were incubated in the dark at 6-6.5°C, and gently stirred daily to keep an even distribution of the eggs in the petri dish until hatching at approximately 90 day degrees. After hatching, the larvae were transferred to petri dishes containing fresh FASW and then sterile tissue culture flasks (NUNC™) containing 50 mL FASW. These flasks were stocked with 30 larvae each. Bacteria were added to the flasks at a final density of $10^6$ CFU/mL. All strains were cultured in Marine Broth (MB) medium (Appendix A, Table A.1). Measurements of bacterial concentrations were needed for both the gnotobiotic fish experiment and the growth experiment. To measure the amount of bacteria in a culture, optical density (OD) was measured by a spectrophotometer at 660 nm. The correct concentration of bacteria in a liquid culture was determined by converting the OD measurement of each strain cultivated in liquid MB medium into CFU/mL by the formula F1 (Vestrum, 2009).

$$\frac{CFU}{mL} = 1200 \cdot 10^6 \cdot OD_{660} \qquad \text{(F1)}$$

Fish mortality, air and water temperature was monitored daily throughout the experiment. At hatching, the air temperature was 9°C. After stocking, the temperature was increased by 1°C per day until 12°C was reached. The daily temperature recordings are listed in Appendix B. Continuous light was used from day 3 post-hatch. When approximately 30% mortality was reached, larvae and water was sampled from each of the replicate flasks. For HI610 and BB120, this was 3 days post-hatch (dph), while for RD5-30 it was 6 dph.

### 2.2.2 Sampling

Cod larvae were sacrificed by an overdose of Ethyl 3-aminobenzoate (MS222) (98% Sigma Aldrich) in FASW (1 g/L) and then rinsed in Milli-Q® water. For sampling for DNA and RNA extraction, 10 cod larvae were collected and spotted on a piece of plankton net (100 µm) and immediately put into a Nunc™ CryoTube™ Vial (Thermo Scientific) and snap frozen in liquid nitrogen and stored at -80°C until DNA or RNA extraction. Water was sampled from all flasks for flow cytometry analysis. A volume of 1.8 mL water was collected in a Nunc™ CryoTube™ Vial (Thermo Scientific), fixated with 0.5% glutaraldehyde, snap frozen in liquid nitrogen and stored at -80°C until analysis by flow cytometry. For DNA and RNA extraction, approximately

25 mL water was filtered through sterile 0.2 µm hollow fiber syringe filters (DynaGard®, Microgon Inc., California) and the filters were stored at -20°C until DNA or RNA extraction. To verify that the axenic flasks contained no bacteria, approximately 100 µL water from each flask was plated on Marine Agar (MA) (Appendix A, Table A.1) to check for growth. The plates were incubated at 20°C for two weeks and inspected daily.

## 2.3 Cultivation of *Vibrio* strains in liquid M65 growth medium to exponential and stationary phase

Growth curves were made for *Vibrio* strains BB120 and HI610 in M65 medium (Appendix A, Table A.2) to identify when the exponential and stationary phases of growth occurred. Overnight cultures were made by inoculating bacteria from glycerol stocks in Erlenmeyer flasks filled up 10% of its total volume by M65 medium, incubated at 28°C and 120 rotations per minute (rpm). The overnight cultures were inoculated to make new 1% cultures, 1 replicate per strain, also incubated at 28°C and 120 rpm. From each flask, 1 mL culture was sampled under sterile conditions every hour, and transferred to 1.5 mL polystyrene semi-micro cuvettes (VWR). For each sample, OD was measured by a spectrophotometer at 660 nm three times. The first OD measurements were done immediately after inoculation, at time = 0, and samples were taken at the specified time intervals until the stationary phases of the growth curves were reached.

RD5-30 and BB120 were then cultivated in M65 to generate samples from exponential and stationary phases of growth, with four replicate flasks per strain. The experiment was conducted as described for the growth curves, except the overnight cultures were inoculated with single bacterial colonies from M65 agar inoculated by glycerol stocks. Samples for DNA and RNA sequence analysis were taken in the exponential and stationary phases, as determined by the previously established growth curves. At each determined time point, 7 mL culture was sampled. 1 mL culture was transferred to 1.5 mL polystyrene semi-micro cuvettes (VWR) for OD measurements at 660 nm, and 2 x 1.5 mL culture was transferred to 2 mL microcentrifuge tubes and centrifuged at 13 000 rpm for 3 minutes. The supernatant was discarded and the tubes snap frozen in liquid nitrogen and stored at -20°C until DNA and RNA extraction.

## 2.4 Analytical methods

### 2.4.1 Flow cytometry

Flow cytometry is a method that can be used to determine many different cell characteristics, among them cell size, cell count and RNA or DNA content (BD Biosciences, 2000). This is achieved by having cells suspended in a fluid pass by a laser and measuring the scattering of the laser light and emitted fluorescence by detectors, and turning this into electronic information (BD Biosciences, 2000). A BD Accuri™ C6 Flow Cytometer was used for flow cytometry analysis of water samples from all flasks in the gnotobiotic fish experiment, and one FASW sample for comparison. The samples were all diluted with 1:10 filtered TE buffer (Appendix A, Table A.3) to an either 1:10 or 1:100 dilution of the sample - adapted to have a cell count as close to 1000 events per µL as possible - and vortexed. The samples were added 1% 1:50 diluted SYBR® Green II RNA gel stain from Invitrogen, a highly sensitive RNA stain that is detected by fluorescence (Molecular Probes Inc., 2001) and vortexed. The samples were then incubated in darkness for 15 minutes, vortexed once more and analyzed by flow cytometry. From the results, graphs plotting forward-scattered light (FCS) against FL1 were made for each sample. Forward scattered light values are proportional to cell size (BD Biosciences, 2000). FL1, also known as FITC, is a fluorochrome (fluorescein isothiocyanate) with a peak emission wavelength detection detected at 530 nm, while the SYBR™ Green II RNA stain has a fluorescence emission at 520 nm (BD Biosciences, 2000, Molecular Probes Inc., 2001).

### 2.4.2 DNA and RNA extraction and cDNA synthesis

DNA and RNA was extracted from cod larvae and water samples from the gnotobiotic fish experiment. DNA and RNA was also extracted from the growth experiment samples of RD5-30 and BB120 in their stationary and exponential phases of growth. For the cod larvae samples, the larvae were transferred into the initial lysis tubes used in the DNA or RNA isolation protocols. For the water samples, the fibers of the DynaGard® filters were removed and used as sample in the initial tubes. For the growth experiment samples, the pellets were transferred to the initial lysis tube for DNA and RNA extraction.

For DNA extraction, the Mo Bio Laboratories Inc. PowerSoil® DNA Isolation Kit was used, following the accompanying protocol (Appendix C). For RNA extraction the PowerMicrobiome™ RNA Isolation Kit from Mo Bio Laboratories Inc. was used, and the accompanying protocol followed (Appendix D). Both the DNA and RNA isolation kits used bead beating lysis protocols.

The RNA extracted was used as template for cDNA synthesis using a Prime Script 1st Strand cDNA Synthesis Kit (TaKaRa) (Appendix E). The DNA and cDNA was used in PCR amplifications. All DNA and cDNA was stored at -20°C until analysis.

### 2.4.3 Polymerase chain reaction (PCR)

Parts of the 16S rRNA gene was amplified by polymerase chain reaction (PCR) from the DNA and cDNA templates. A nested PCR protocol with an external and internal amplification was used to avoid co-amplification of eukaryotic DNA when amplifying the V3-region of samples from the gnotobiotic fish experiment. For external amplification, forward primer EUB8F and reverse primer 984yR were used (Table 2.1). For the internal amplification of the V3-region of the 16S rRNA gene, primers 338F-GC and 518R were used (Table 2.1). For PCR amplification of the V1-region, forward primer EUB8F-GC and reverse primer 179rev (Table 2.1) were used.

**Table 2.1:** Primer names and sequences for all primers used in PCR reactions in this thesis.

| Primer name | Sequence(5'-3') |
|---|---|
| **EUB8F** | AGAGTTTGATCMTGGCTCAG |
| **984yR** | GTAAGGTTCYTCGCGT |
| **338F-GC** | CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG ACTCCTACGGGAGGCAGCAG |
| **338F-GC-M13** | CAGGAAACAGCTATGACCCGCCCGCCGCGCGCGGCGGGCGG GGCGGGGGCACGGGGGG ACTCCTACGGGAGGCAGCAG |
| **518R** | ATTACCGCGGCTGCTGG |
| **EUB8F-GC** | CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG AGAGTTTGATCMTGGCTCAG |
| **179rev** | TGCGGTATTAGCCATCGTTTCC |
| **Vibrio48F-Ill** | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCTAAC ACATGCAAGTCGAGC |
| **Vibrio510R-Ill** | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNTTAG CCGGTGCTTCTTCTG |
| **M13R** | CAGGAAACAGCTATGACC |

PCR reactions were performed with Taq polymerase and reaction buffer from Qiagen, a total $MgCl_2$ concentration of 2 mM, 0.2 mM of each dNTP and 0.3 µM of each primer (Table 2.2). In the external PCR reactions, 0.3 µg/µL purified Bovine Serum Albumin (BSA) 100X (New England BioLabs Inc.) was also added.

**Table 2.2:** PCR components used in the PCR protocol for amplification of the 16S rRNA gene.

| PCR Reagent | Amount per reaction (µL) | Final concentration |
|---|---|---|
| **PCR buffer 10x** | 2.5 | 1x (Tris-Cl, KCl, $(NH_4)_2SO_4$ and 1.5 mM $MgCl_2$) |
| **10 mM dNTPs** | 0.5 | 200 µM each dNTP |
| **$MgCl_2$** | 0.5 | 2 Mm* |
| **Forward primer** | 0.75 | 0.3 µM |
| **Reverse primer** | 0.75 | 0.3 µM |
| **Taq DNA Polymerase** | 0.125 | 0.025 U/µL |
| **Template** | 1 | |
| **Filtered $H_2O$ (**Milli-Q®**)** | ~18.875 | |

*Total $MgCl_2$ concentration – 1.5 mM $MgCl_2$ from the buffer with an additional 0.5 mM $MgCl_2$ added

The temperature cycles used for the external PCR amplification of the V3-region is presented in Table 2.3. The program for the internal PCR was identical, except the annealing temperature which was increased to 53°C. Furthermore, the number of cycles were increased up to 24 cycles for samples with a low yield after 20+20 cycles. The PCR program used for amplification of the V1-region was similar to the program presented in Table 2.3, except the annealing temperature, which was 53 °C, and the number of cycles, which was 35.

**Table 2.3:** Temperature program used for PCR amplification of the V3-region of the 16S rRNA gene.

| Step | Length | Temperature (°C) | |
|---|---|---|---|
| Denaturation | 3 minutes | 95 | |
| Denaturation | 30 seconds | 95 | 20 cycles |
| Annealing | 30 seconds | 50 | |
| Elongation | 1 minute | 72 | |
| Final elongation | 10 minutes | 72 | |
| Cooling | 5 minutes | 4 | |
| Storage | ∞ | 10 | |

### 2.4.4 Agarose gel electrophoresis

The yield and quality of all PCR amplicons was inspected by agarose gel electrophoresis. Agarose gels (1%) were made by adding SeaKem® LE Agarose (Lonza) to 1 x TAE buffer (Appendix A, Table A.4) and the solution was homogenized by boiling. The solution was subsequently cooled to a temperature of approximately 65 °C and 5 µL GelRed (Qiagen) per 100mL 1% agarose solution was added to stain the DNA in the gel. The gel was then poured into a gel chamber and combs were inserted to make wells for application of samples. Once the gel had polymerized, the gel chamber was filled with 1 x TAE buffer. GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific) was applied to the gel. Prior to application to the gel, 5µL of PCR product was mixed with 1 µL of 6X DNA Loading Dye (Thermo Scientific). Small 1% agarose gels were run at 95 Volts for 1 hour and larger gels were run at 140V for 45 minutes. The agarose gels were then transferred to a SynGene G:BOX GelDoc for visualization of the DNA, and pictures of the gel were taken using GeneSnap from SynGene software.

### 2.4.5 Denaturing gradient gel electrophoresis (DGGE)

DGGE is a method used in microbial ecology, where how far a DNA molecule migrates down the linear denaturing gradient of the gel is dependent on its sequence (Muyzer, 1999, Fukuda et al., 2016). Forward primers with long Guanine-Cytosine rich sequences ("GC-clamps") attached were used for the PCR of samples to be analyzed by DGGE, to avoid complete denaturation of the two strands of the DNA molecules (Fukuda et al., 2016, Nübel et al., 1996). In this project the "GC-clamp" containing forward primers 338F-GC and EUB8F-GC (Table 2.1) were used; 338F-GC for amplification of the V3-region, and EUB8F-GC for the V1-region.

To cast the gel, two 8% acrylamide solutions with 0% and 80% denaturing reagents were prepared where 100% corresponds to 7 M urea and 40% formamide (Appendix A, Table A.5 and Table A.6). Acrylamide solutions with variable denaturing percentages were made by mixing these two solutions (Table 2.4), corresponding to the relevant denaturing gradient of the gel to be casted. The DGGE gels analyzing the V3-region of the 16S rRNA gene had a denaturing gradient of 35% to 55%, while the gels for the analyses of the V1-region had a 35%

to 50% denaturing gradient. The 80% denatured solution was filtered through a 0.45 µm pore size filter (VWR) prior to mixing with the 0% denatured solution. TEMED and 10% ammonium persulfate (APS) were added to the two solutions to initiate the polymerization process directly prior to casting of the gel (Table 2.4).

**Table 2.4:** Composition of solutions used for casting DGGE gels with different denaturing gradients.

| **Denaturing %** | **0% denatured solution (mL)** | **80% denatured solution (mL)** | **TEMED (µL)** | **10% APS (µL)** | **Total volume (mL)** |
|---|---|---|---|---|---|
| **35** | 13.5 | 10.5 | 16 | 87 | 24 |
| **50** | 9 | 15 | 16 | 87 | 24 |
| **55** | 7.5 | 16.5 | 16 | 87 | 24 |
| **0 "stacking gel"** | 8 | 0 | 40 | 10 | 8 |

A gradient mixer was used for casting the gel between two glass plates to generate a denaturing gradient. A 0% denatured "stacking gel" solution was casted on the top of the gel (Table 2.4). A 48-toothed comb was inserted on top of the gel. The gel was set to polymerize for one hour. It was then placed in an INGENY phorU heating tank containing approximately 18 L of 0.5 x TAE buffer preheated to 60°C. The PCR products (3 to 15µL) were mixed with 5µL loading dye and applied to the wells. The amount of the PCR product added was determined by the brightness of their respective bands in the 1% agarose gels, in an attempt to get relatively similar amounts of DNA in each sample applied. A marker was also applied to about every tenth well of the gel. This marker consisted of V3-region 16S rRNA PCR products pooled together. In this marker, the templates used were pure cultures of *Staphylococcus aureus*, *Ruminococcus obeum*, *Eubacterium formicigenerans*, *Ruminococcus productus*, *Fusobacterium prauznitzii*, *Clostridium celerescans*, *Eubacterium plutii*, *Eubacterium halii*, and *Bifidobacterium longum*.

The gel was then run for a minimum of 22 hours. After running, the gel was transferred from the glass plates onto a plastic sheet and stained with a solution containing 30 mL Milli-Q® water, 600 µL 50 x TAE buffer and 3 µL SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen). The gel was incubated in darkness with the staining solution for an hour, before it was rinsed

with Milli-Q® water and placed in the SynGene G:BOX GelDoc to visualize the band patterns and GeneSnap from SynGene was used to take pictures.

### 2.4.6 Reamplification and Sanger sequencing of bands from DGGE gels

DGGE bands were excised by using pipette tips, and the gel material was transferred to Eppendorf tubes containing 20µL of filtered Milli-Q® water. These tubes were then vortexed and used as template in PCR reactions. The protocol was as described in Section 2.4.3 (Table 2.2 and Table 2.3), but with forward primer 338F-GC-M13 and reverse primer 518R (Table 2.1), 53°C annealing temperature and 38-40 cycles. The PCR products were examined on 1% agarose gels and then purified with QIAquick® PCR Purification Kit (Qiagen). The procedure was followed as described in the manufacturer's protocol (Appendix F). Finally, 5 µL of each purified sample along with 5 µL of 5 µM primer M13R (Table 2.1) were added to 1.5 mL Eppendorf tubes marked with bar codes and sent to GATC Biotech for Sanger sequencing.

### 2.4.7 Design of primers spanning the V1- and V3-regions for Illumina sequencing

For Illumina sequencing, we aimed to generate amplicons spanning the entire V1- and V3-regions of the 16S rRNA genes of the *Vibrio* strains. The amplified fragments should be as small as possible while still spanning the entirety of both the V1- and V3-regions, to keep them within the maximum read length for 300 bp paired end MiSeq Illumina sequencing. For primer design, 16s rRNA sequences for different *Vibrio* species, including *V. campbellii*, *V. anguillarum* and *V. harveyi*, were downloaded from the RDP database (Michigan State University, 2016). The GenBank accession numbers of the *Vibrio* strains used in the alignment were: X74711, AJ421444, X74690, AJ002566, X74724, X74706, X74692, AJ293802, AJ316181, AJ312382, AJ316205, Z21731, X71818, AJ514917, AY662308, DQ980029, AY257971 and AY257974.

An alignment of all these sequences as well as the sequences acquired from the Sanger sequencing of DGGE gel bands (Section 2.4.6) was made with CloneManager (Scientific and Educational Software). Primers were designed to target regions conserved in all the *Vibrio*

species. GC content of the primers was checked to ensure suitable annealing temperatures. The designed primers covered an area of the 16S rRNA gene spanning from nucleotide (nt) 48 to 510, according to *E. coli* numbering (Figure 1.1). Adapter sequences necessary for their use in Illumina sequencing were added to the primers. Sequences for the resulting primers (Vibrio48F-Ill and Vibrio510R-Ill) are given in Table 2.1.

### 2.4.8 Preparation of amplicon library for Illumina sequencing

Samples from the gnotobiotic fish experiment performed for this thesis (referred to as the 2016 experiment) and the gnotobiotic experiment described in Section 1.7 (the 2014 experiment) as well as samples from the growth experiment were sequenced by MiSeq Illumina sequencing by the Norwegian High Throughput Sequencing Centre (NSC). For a full overview of the 76 samples in the library submitted for Illumina sequencing, see Appendix G. Before sequencing, an amplicon library was made according to the description below.

#### Amplification of the 16S rRNA V1-V3 region

The target region was first amplified for all samples using primers spanning the V1- and V3-regions; Vibrio48F-Ill and Vibrio510R-Ill (Table 2.1). A high fidelity DNA polymerase with a low error rate, Phusion Hot Start DNA Polymerase (Thermo Scientific) and a 5X Phusion HF Buffer™ from the same producer was used in the reaction mixture. The full composition of the PCR reactions is given in Table 2.5. The thermocycling program used for the PCR is presented in Table 2.6. The number of cycles was increased to 35 for samples showing weak bands on 1% agarose gels after 33 cycles.

**Table 2.5:** PCR components for amplification of the V1-V3 region of the 16S rRNA gene for Illumina sequencing.

| PCR Reagent | Amount per reaction (µL) | Final concentration |
|---|---|---|
| **5X Phusion HF Buffer™** | 4.0 | 1x (1.5 mM $MgCl_2$) |
| **10 mM dNTPs** | 0.5 | 200 µM each dNTP |
| **$MgCl_2$** | 0.4 | 2.5 mM* |
| **BSAx100 (10 mg/mL)** | 0.8 | 0.4 mg/mL |
| **Forward primer** | 0.6 | 0.3 µM |
| **Reverse primer** | 0.6 | 0.3 µM |
| **Phusion Hot Start DNA Polymerase** | 0.15 | 0.015 U/µL |
| **Template** | 1 | |
| **Filtered $H_2O$ (**Milli-Q®**)** | 11.95 | |

*Total $MgCl_2$ concentration – 1.5 mM $MgCl_2$ from the buffer with an additional 1 mM $MgCl_2$ added

**Table 2.6:** Temperature cycles used for PCR amplification of the V1-V3 region of the 16S rRNA gene for the Illumina amplicons.

| Step | Length | Temperature (°C) | |
|---|---|---|---|
| Denaturation | 1 minute | 98 | |
| Denaturation | 15 seconds | 98 | |
| Annealing | 20 seconds | 53 | 33 cycles |
| Elongation | 20 seconds | 72 | |
| Final elongation | 5 minutes | 72 | |
| Cooling | 1 minute | 4 | |
| Storage | ∞ | 10 | |

After the PCR amplification of the V1-V3 region of the 16S rRNA gene, the PCR products were purified and normalized by use of SequalPrep™ Normalization Plate (96) kit (Invitrogen) following the accompanying protocol (Appendix H).

**Indexing PCR**

After normalization, the samples were tagged with sequencing indexes that were incorporated into the forward and reverse primers, so each sample had a unique combination of index sequences in the 5'- and 3'-end of their PCR product. The indexes used were from the Nextera® XT Index Kit v2 Set D and the Nextera® XT Index Kit (96 indexes), both from Illumina. The PCR reaction mixture is given in Table 2.7.

**Table 2.7:** PCR components used in the indexing PCR reaction for amplification of the V1-V3 region of the 16S rRNA gene.

| **PCR Reagent** | **Amount per reaction (µL)** | **Final concentration** |
|---|---|---|
| **5X Phusion HF Buffer™** | 5.0 | 1x (1.5 mM $MgCl_2$) |
| **10 mM dNTPs** | 0.625 | 250 µM each dNTP |
| **$MgCl_2$** | 0.5 | 2.5 mM* |
| **Index 1** | 2.5 | |
| **Index 2** | 2.5 | |
| **Phusion Hot Start DNA Polymerase** | 0.188 | 0.015 U/µL |
| **Template** | 2.5 | |
| **Filtered $H_2O$ (**Milli-Q®**)** | 11.187 | |

*Total $MgCl_2$ concentration – 1.5 mM $MgCl_2$ from the buffer with an additional 1 mM $MgCl_2$ added

The PCR program used for the amplification of Illumina amplicons with indexes was as described in Table 2.6, but with only 12 cycles and a 50°C annealing temperature. After the indexing PCR, the PCR products were purified and normalized through the SequalPrep™ Normalization Plate (96) kit (Invitrogen) protocol (Appendix H).

**Pooling and concentration of amplicons**

After normalization, equal volumes of the PCR products were pooled into one amplicon library and concentrated through AmiconUltra 0.5 centrifugal filter devices (30K membrane, Merck Millipore) following the manufacturer's protocol (Appendix I). The purity and concentration were measured by NanoDrop and Qubit, respectively. The Illumina sequencing requirements set by The Norwegian High Throughput Sequencing Centre was minimum 50 ng (5 µL) DNA with a A260/280 ratio in the range 1.8-2.0 and a 260/230 ration within the 1.8-2.4 range (The Norwegian High Throughput Sequencing Centre, 2016). The purity was found to be 2.32 for the A160/280 ratio and 0.42 for the 260/230 ratio through NanoDrop measurements. These low measurements were due to the use of TE buffer as an elution buffer, as this buffer contains EDTA which absorbs at 230 nm. Qubit measured the concentration at 13.2 ng/µL. Finally, 30 µL of the amplicon library was sent to The Norwegian High Throughput Sequencing Centre for Illumina sequencing.

### 2.4.9 Analysis of Illumina amplicon sequencing data

The sequencing data were processed by Ingrid Bakke using the Usearch pipeline (version 8.1.1825; http://drive5.com/usearch/features.html). The major steps in the pipeline included demultiplexing, quality trimming, chimera removal, and clustering to obtain Operational taxonomic units (OTUs) at 99% similarity level. The subsequent taxonomy affiliation was based on the Utax script implemented in the Usearch pipeline with a confidence value threshold of 0.8 and the RDP reference data set version 15.

For the statistical analyses of the resulting OTU table, the program package PAST version 2.17 was used (Hammer et al., 2001). Principal coordinate analysis (PCoA) was based on Bray-Curtis similarities. In PCoA analysis, a distance matrix is generated based on the OTU profiles for the samples, and a plot is generated to visualize the distances between samples (Di Bella et al., 2013). In this two-dimensional plot, one axis represent the first coordinate, which is the direction that most separates the samples while the other axis represent the second coordinate, which is the direction separating the samples second best (Di Bella et al., 2013).

One-way permutational multivariate analysis of variance (PERMANOVA) tests were done in PAST to investigate whether groups of samples were different from each other in a statistically significant way. These tests were based on Bray-Curtis dissimilarities. The test assumes a null hypothesis of there being no differences between groups. If the test yields a p-value below the 0.05 threshold, the null hypothesis is void and the two groups are found to be significantly different (Anderson, 2001, Hammer et al., 2001).

Similarity percentage (SIMPER) tests, once again based on Bray-Curtis dissimilarities, were used to identify which OTUs were contributing the most to the differences found between sample groups (Poretsky et al., 2014).

CloneManager (Scientific and Educational Software) was used to generate sequence alignments for comparing representative OTU sequences.

Secondary structures for the various V1- and V3-region sequences identified by the Illumina sequencing was predicted using an online tool provided by Mathews group (2017) found at https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html. The sequence stretch used in these analyses correspond to the defined secondary structures including H6 (V1) and H17 (V3) as determined for *E.coli* (Kitahara et al., 2012). The most likely secondary structure was identified based on the lowest free energy.

# 3 RESULTS

## 3.1 Gnotobiotic cod larval experiment: distribution of *Vibrio* 16S rRNA intragenomic sequence variants in cod larvae and rearing water

To investigate whether the distribution of 16S rRNA intragenomic variants for the three strains of *Vibrio* varied when the bacteria lived planktonic in rearing water versus when they were colonizing fish, a gnotobiotic fish experiment was conducted with bacteria free larvae of Atlantic cod (*G. morhua*). These fish larvae were exposed to either *V. anguillarum* strain HI610, *V. Campbellii* strain BB120, *Vibrio* sp. isolate RD5-30 or no bacteria as a control treatment. Flasks containing only one of the *Vibrio* strains but no cod larvae were also used in the experiment.

### 3.1.1 Survival and microbial status in gnotobiotic cod larvae experiment

Fish mortality was monitored every day during the experiment, as sampling at least twenty living larvae per flask was desirable. Each flask started with approximately 30 fish at day 0 post-hatch. The survival curves of the fish reared in the flasks containing HI610, BB120, RD5-30 or axenic FASW as a control treatment, are shown in Figure 3.1. In the flasks containing a pathogenic strain (HI610 or BB120) survival dropped already at 3 days post-hatch (dph) (Figure 3.1). These flasks were therefore sampled at 3 dph to ensure enough live fish for sampling. At 6 dph, the flasks with the axenic cod larvae control treatment and those exposed to RD5-30, had an average survival of about 26 and 27 respectively, and all these flasks were sampled at this point in the experiment.

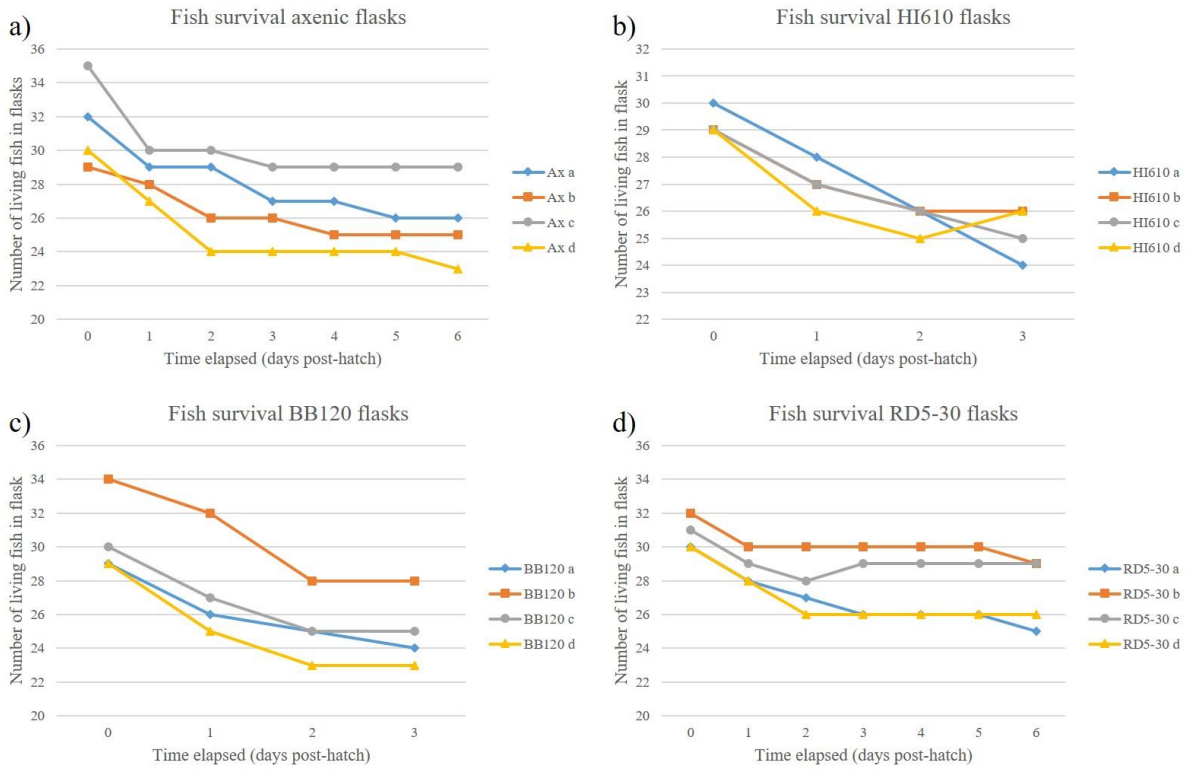


**Figure 3.1:** Curves tracking the survival of Atlantic cod larvae as a function of days post-hatch in the flasks in the gnotobiotic fish experiment in a) the axenic (Ax) control treatment flasks, b) the flasks containing HI610, c) the flasks containing BB120 and d) the flasks containing RD5-30. For each treatment, four parallel flasks were used (a-d).

Flow cytometry analysis of the rearing water was applied to evaluate the microbial status in the flasks, and indicated all axenic control flasks seemed to be bacteria free (example of plot in Appendix J, Figure J.1) and that the flasks exposed to the *Vibrio* strains appeared to contain mono-populations of bacteria (Representative plots shown in Appendix J, Figure J.2). The water from the axenic fish experiment flasks that was plated on MA also showed no signs of growth after two weeks incubation.

### 3.1.2 DGGE analysis and Sanger sequencing of V1- and V3-regions of 16S rRNA amplicons from gnotobiotic cod larvae experiment samples

To map how many different 16S rRNA sequence variants in variable regions V1 and V3 were present for each of the *Vibrio* strains in the gnotobiotic fish experiment samples, a DGGE analysis was used. DGGE was also used to investigate whether the distribution of these variants varied between the rearing water and fish samples, as well as the water samples from the flasks containing one of the bacterial strains but no cod larvae.

The DGGE profiles of all samples were investigated from extracts on both the DNA and RNA level. All PCR reactions resulted in amplification products of the expected size, with a few exceptions: PCR amplification of the DNA and RNA extracted from the BB120 cod larvae samples gave little or no PCR products. This trend was the same for both the V1- and V3-region of the 16S rRNA gene, indicating low amounts of the BB120 strain in the fish samples. As there was a lot of amplification product for the BB120 water samples, for both the DNA and RNA extracts, these results indicate that BB120 had not managed to colonize the cod larvae during the experiment.

The V1-region of the 16S rRNA was first analyzed through DGGE analysis of PCR products amplified from DNA isolated from fish and water samples (Figure 3.2).

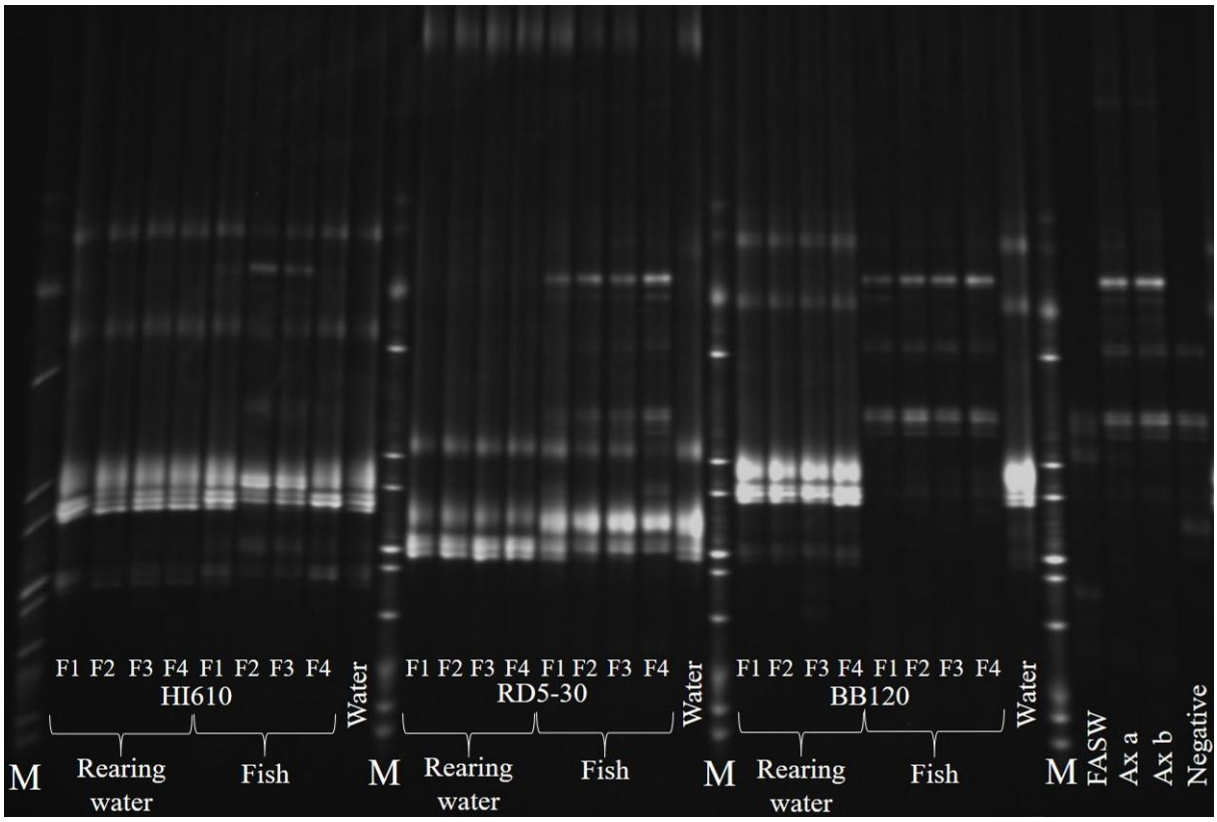


**Figure 3.2:** DGGE gel with 16S rRNA V1-region PCR products amplified from DNA isolated from cod larvae and water samples in the gnotobiotic fish experiment. Bacterial strain and sample type (water or fish) is indicated for each well. F1 through F4 indicates the replicate flask that the samples were obtained from, whereas lanes named just "water" indicates flasks that were added a *Vibrio* strain but no cod larvae. The four rightmost samples contain negative controls from FASW, fish samples from axenic flasks (Ax a and b) and a negative PCR amplification control. M indicates wells with DGGE markers.

All three *Vibrio* strains exhibited band profiles with more than a single band for all samples. Many of these bands were, however, closely positioned on the gel, indicating relatively small variations in sequence between the different variants the bands represent. HI610 samples had three to four bands, positioned very closely on the gel. Water and fish samples for the RD5-30 strain seemed to have slightly different band profiles, which indicates that the fish and water samples have different abundances of the V1-region sequence variants at the DNA level. BB120 fish and water samples had very different band profiles, but upon further inspection it seemed that the band profiles of the BB120 fish samples were identical to the axenic control fish samples. Along with the weak PCR amplification results, this further supports the theory that

BB120 did not manage to properly colonize the fish, and that these band profiles do not accurately represent BB120 in fish.

A DGGE analysis was also conducted for 16S rRNA V1-region amplified from RNA isolated from water and fish samples from the same gnotobiotic cod larvae experiment (Figure 3.3). This was done to investigate whether intragenomic V1-region sequence variants were expressed differently in the water and fish samples.

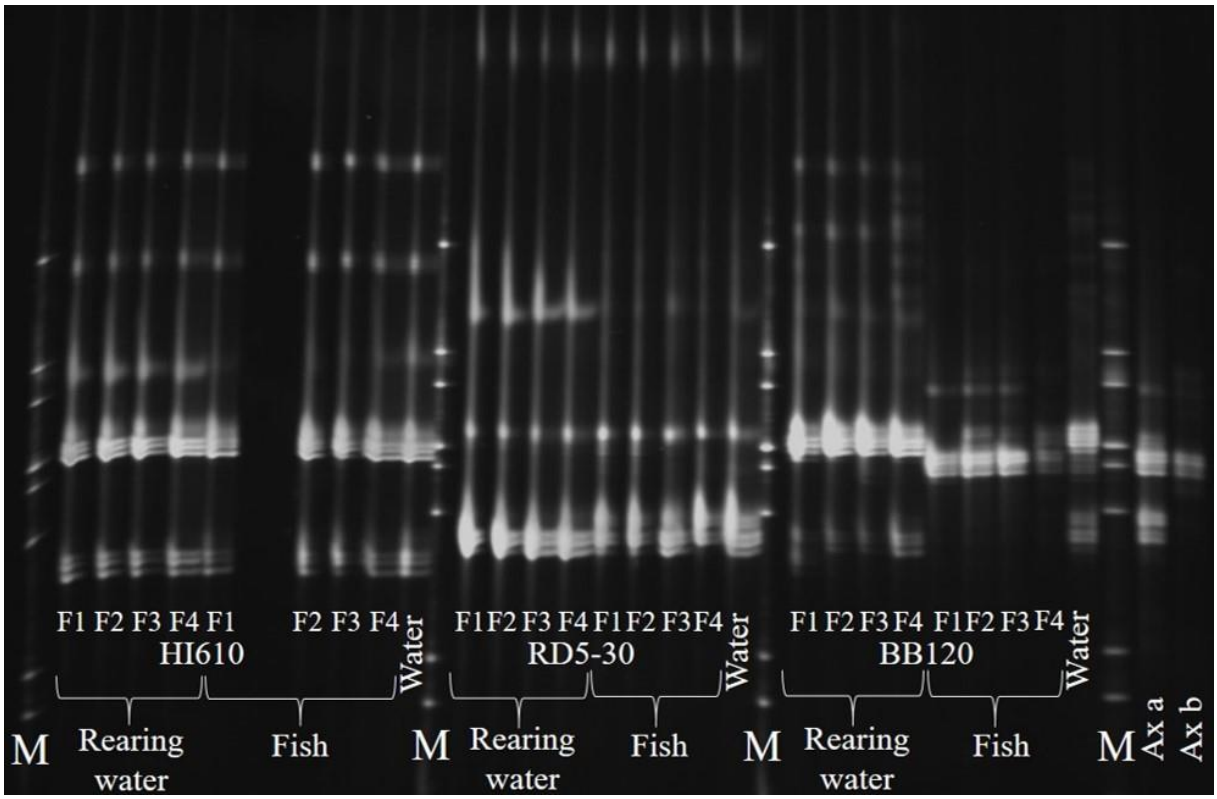


**Figure 3.3:** DGGE gel with 16S rRNA V1-region PCR products amplified from RNA isolated from cod larvae and water samples in the gnotobiotic fish experiment. Bacterial strain and sample type (water or fish) is indicated for each well. F1 through F4 indicates the replicate flask that the samples were obtained from, whereas lanes named just "water" indicates flasks that were added a *Vibrio* strain but no cod larvae. The two rightmost wells contain negative controls from water (Ax a) and fish (Ax b) samples from axenic flasks. M indicates wells with DGGE markers.

In this analysis, there was no clear difference between the band profiles of the fish and water samples within either the HI610 or RD5-30 strain. Once again, fish and water samples from BB120 exhibited different band profiles. The fish band profiles again seem more similar to the axenic fish sample than any other band profiles, indicating the PCR amplification products used for the analysis did not properly represent fish colonized by BB120. Similar to the DNA V1-region DGGE gel, the bands were mostly clustered close together on the gel, indicating only minor deviations in sequence for the different sequence variants the bands assumedly represent.

We further investigated potential intragenomic sequence variation in the V3-region of the 16S rRNA gene, first by using DNA isolated from the gnotobbiotic fish experiment. Earlier DGGE analyses of RD5-30 (see Figure 1.3) indicated such intragenomic variation was present in the V3-region of this *Vibrio* strain.

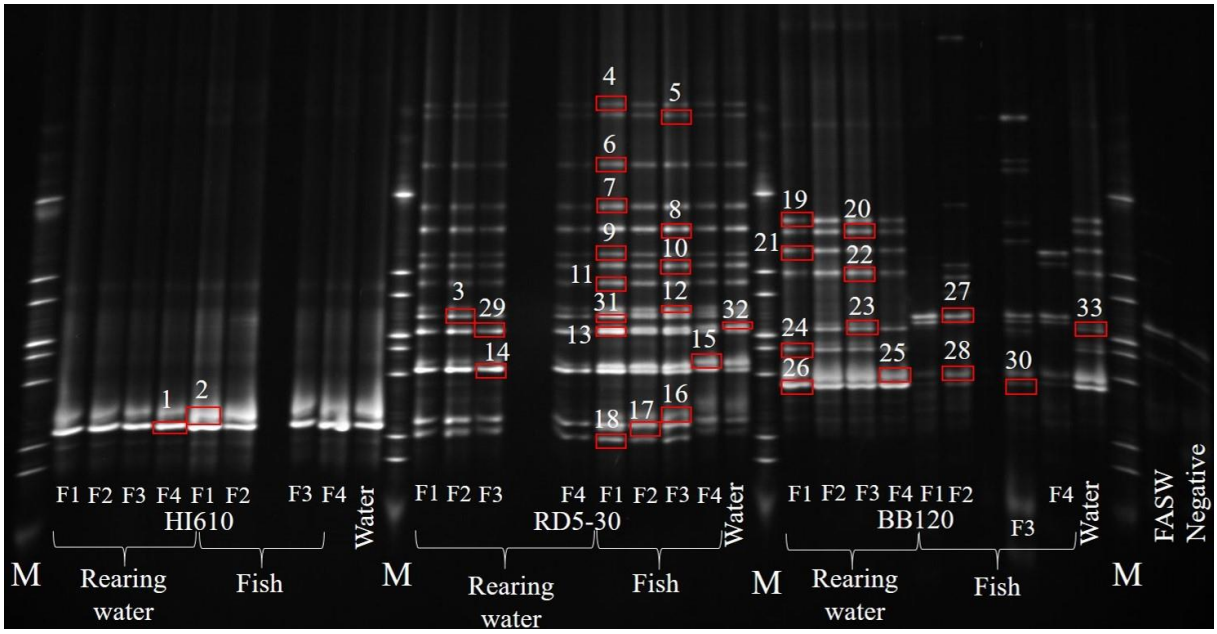


**Figure 3.4:** DGGE gel with 16S rRNA V3-region PCR products amplified from DNA isolated from cod larvae and water samples in the gnotobiotic fish experiment. Bacterial strain and sample type (water or fish) is indicated for each well. F1 through F4 indicates the replicate flask that the samples were obtained from, whereas lanes named just "water" indicates flasks that were added a *Vibrio* strain but no cod larvae. Bands marked with red squares were reamplified and analyzed by Sanger sequencing. The two rightmost wells contain negative controls from FASW and a negative PCR amplification control. M indicates wells with DGGE markers.

The HI610 16S rRNA V3-region PCR products based on DNA isolates all showed two distinct DGGE bands, though they were not far apart in the DGGE profile. The PCR products obtained for the RD5-30 samples showed as many as 15 distinct bands distributed throughout the denaturing gradient. This indicates intragenomic variants within the V3-region of this strain with big differences in nucleotide sequence. Some possible differences between the band profiles of fish and water samples were observed for the RD5-30 samples. The band marked number 15 seemed generally somewhat brighter in fish samples than in the rearing water samples, though it was also bright for the sample from the flask without cod larvae (Figure 3.4).

For the BB120 samples a similar trend as in the V1-region analyses was observed, with different DGGE profiles for water and fish samples (Figure 3.4). All the PCR products representing the water samples each had band profiles containing about eight distinct bands, indicating large V3-region sequence diversity, while the fish samples had entirely different band profiles. Though less clear in this gel than the previous V1-region gels, the BB120 fish samples seem like they might again be more similar to the band profiles of the negative controls. This is more difficult to determine for this gel because of the smiling effect on the edge of the gel skewing the band profiles of the negative samples. This result does again seem to reflect that BB120 did not manage to efficiently colonize the cod larvae.

RNA from all the gnotobiotic fish experiment samples then had their V3-regions amplified by PCR and were then run on a DGGE gel (Figure 3.5) to investigate if the sequence variants were differently expressed in the *Vibrio* strains depending on whether they were colonizing fish or living planktonic in rearing water or seawater.

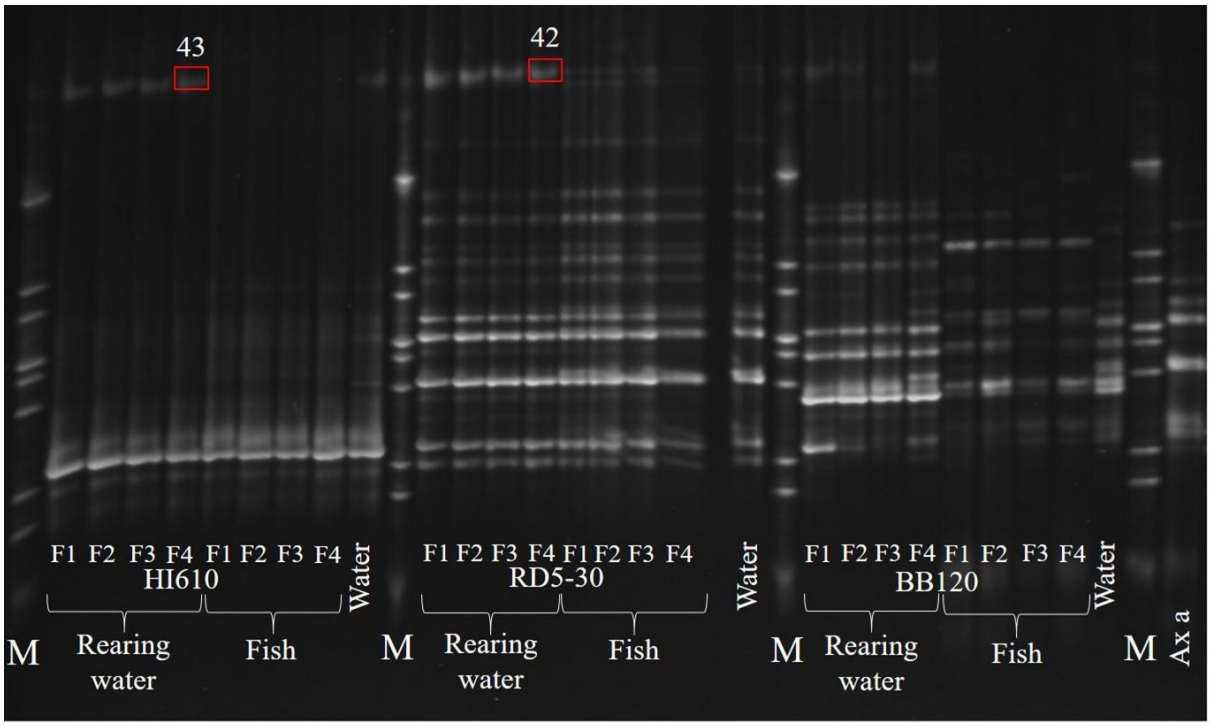


**Figure 3.5:** DGGE gel with 16S rRNA V3-region PCR products amplified from RNA isolated from cod larvae and water samples in the gnotobiotic fish experiment. Bacterial strain and sample type (water or fish) is indicated for each well. F1 through F4 indicates the replicate flask that the samples were obtained from, whereas lanes named just "water" indicates flasks that were added a *Vibrio* strain but no cod larvae. Bands marked with red squares were reamplified and analyzed by Sanger sequencing. The rightmost well contain a negative control from a water sample from an axenic control flask. M indicates wells with DGGE markers.

The resulting DGGE gel again revealed two bands for all HI610 samples (Figure 3.5). The band profiles representing the water samples all seemed to have an additional, weak band higher up on the gel compared to the band profiles representing the fish samples. The band profiles for the RD5-30 samples all showed band patterns similar to those obtained for the DNA-based amplicons. There might also have been some differences in the relative abundance of some of the upper bands when comparing samples from fish and rearing water, which indicates a potential difference in the expression of the different 16S rRNA variants in these two different environments. The V3-region RNA DGGE analysis of BB120 once again revealed very different band profiles for fish and water samples.

A number of bands from the V3-region DGGE profiles were excised from the DGGE gels (see Figure 3.4 and 3.5) and subjected to Sanger sequencing to investigate which sequence variants the different bands represented. The results are presented in Figure 3.6. Bands 17, 27 and 32 (Figure 3.4) yielded sequences that were of too poor quality and were excluded from further analysis.

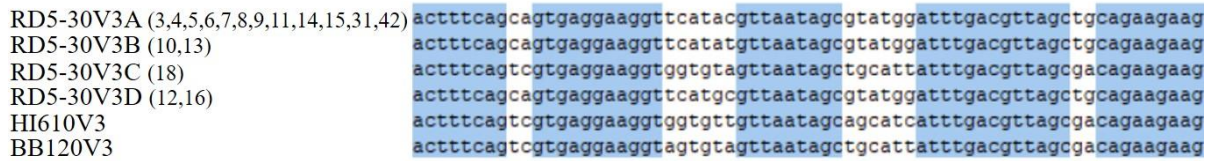


**Figure 3.6:** Alignment of DNA sequences for eluted and reamplified bands from DNA and RNA 16S rRNA gene V3-region DGGE gels. The numbers in the parentheses represent the sequenced bands corresponding to the RD5-30 variants (Figure 3.4 and 3.5).

Surprisingly, many of the DGGE bands that had migrated differently in the DGGE denaturing gradient appeared to represent identical sequences. The RD5-30 sequences yielded only four different sequence variants, while they represented 14-15 distinct bands in the DGGE gels (Figure 3.4 and 3.5). All three HI610 bands sequenced yielded identical sequences, as did all of the BB120 bands.

## 3.2 Growth experiment with RD5-30 and BB120 in exponential and stationary growth

Next, an additional experiment was conducted to investigate the possible differences in abundance of the potential intragenomic 16S rRNA gene sequence variants in different growth phases. The *V. anguillarum* strain HI610 was excluded from further analyses, as there were no indications of extensive intragenomic 16S rRNA gene sequence variation in the DGGE analyses. Samples from this experiment were subsequently subjected to Illumina amplicon sequencing together with samples from the gnotobiotic fish experiment, and the results are presented in Section 3.3.

Firstly, growth curves were established for the *Vibrio* strains RD5-30 and BB120 in the liquid growth medium M65 (Figure 3.7). Growth experiments were then performed to generate samples representing the strains in their exponential and stationary phases of growth for Illumina 16S rRNA amplicon sequencing. OD measurements indicated somewhat slower growth than what was shown by the growth curves (Figure 3.7), and sampling was performed accordingly. For details regarding OD measurements and sampling time points, see Appendix K.

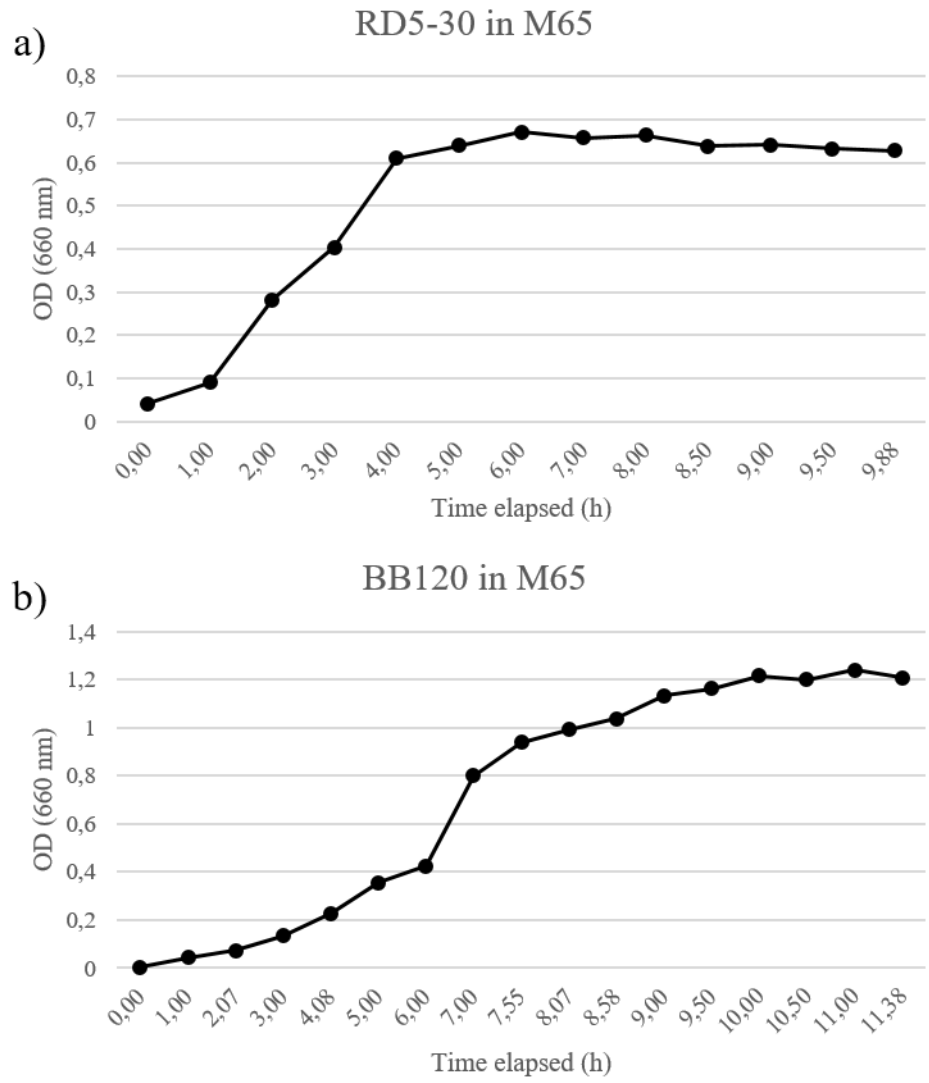


**Figure 3.7:** Growth curves for a) the *Vibrio* strain RD5-30 and b) the *Vibrio* strain BB120, in M65-medium.

## 3.3 Illumina sequencing of 16S rRNA amplicons

After the somewhat surprising results of the DGGE analyses, which showed few rRNA sequence variants for the RD5-30 and BB120 in spite of them having band-rich DGGE profiles, we wanted to investigate the sequence variation in more detail through the use of Illumina amplicon sequencing. Included in the Illumina sequencing were samples from the gnotobiotic cod larvae experiment (Section 3.1), the growth experiment (Section 3.2) and samples from day 13 of the gnotobiotic fish experiment performed prior (2014) to this master project (Section 1.7). The amplicons analyzed covered both the V1- and V3-regions of the 16S rRNA gene. An overview of all samples included in the amplicon sequencing is given in Appendix G.

After quality filtering and chimera removal, the total number of sequence reads varied between 35 075 and 232 434 among samples. A minor fraction of the sequence reads represented *Salmo salar* and chloroplast sequences, and were subsequently excluded from the data set. Most OTUs were classified as bacteria, and specifically *Vibrio* as expected – however, a large number of rare OTUs were classified as other bacterial taxa. These OTUs might be a result of the amplification of bacterial DNA fragments found in FASW.

### 3.3.1 HI610 sequence variants

For all HI610 samples, one OTU dominated, on average making up 99.48% of the sequence reads in each sample. Because of this lack of variation, no further analyses were done for the HI610 samples. The predicted secondary structure of this HI610 OTU is presented in Figure 3.8.

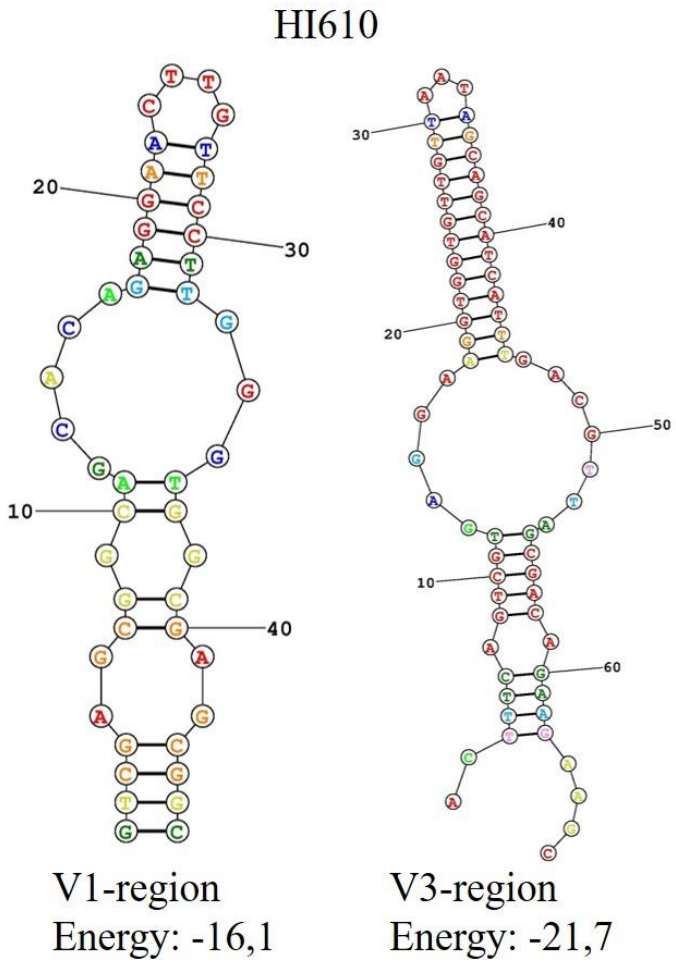


**Figure 3.8:** The secondary structure of the V1- and V3-regions of the 16S rRNA gene for the most common HI610 sequence as predicted by the RNA secondary structure predictor by Mathews group (2017).

### 3.3.2 BB120 sequence variants

For BB120, the nine most common OTUs made up about 92.39% of the average sequence reads per sample. Among these nine sequences, however, were OTU 2 and OTU 1 - the most common sequence variants in HI610 and RD5-30, respectively. As these OTUs also differed a lot in sequence from the remaining BB120 variants, they probably represented a contamination, from either the rearing flask or the PCR reaction. The seven remaining common OTUs constituted an average of 90.48% of the BB120 sequence reads. However, these sequences differed by only a few nucleotides, and very little of this sequence variation was located in the variable regions V1 and V3. Because of the lack of sequence variation in the two variable regions examined, as

well as the problems BB120 had colonizing cod larvae, no further analyses were done for comparisons between different sample groups. The predicted secondary structure in the V1- and V3-regions for the most common sequence variant, OTU 3, is shown in Figure 3.9.

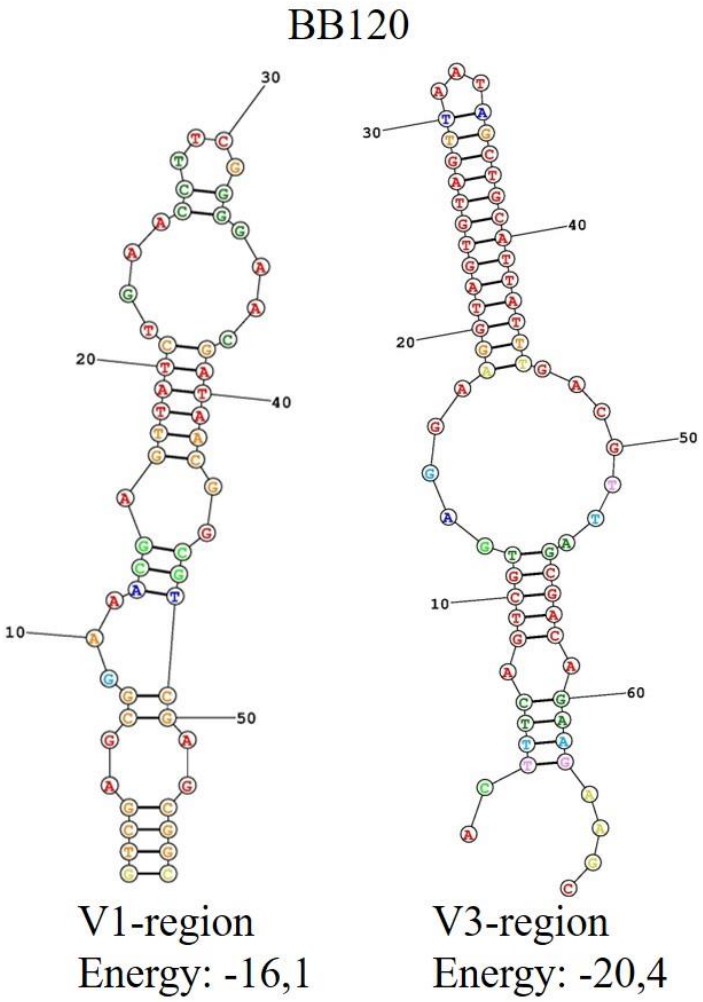


**Figure 3.9:** The secondary structure of the V1- and V3-regions of the 16S rRNA gene for the most common BB120 sequence as predicted by the RNA secondary structure predictor by Mathews group (2017).

### 3.3.3 RD5-30 sequence variants and secondary structure variants

For the RD5-30 samples, 22 sequence variants with a minimum average abundance of at least 0.1% in the samples were identified. These sequences were aligned, and differences in the V1- and V3-regions of these sequences were identified. Most of the sequence variance between these OTUs were located in hypervariable regions of the V1- and V3-regions (Figure 3.10 and Figure 3.11). OTU 1 was by far the most common OTU and constituted on average 50.8% of all sequence reads in the RD5-30 samples (Table 3.1). OTU 2 and OTU 14 had average abundances of 0.4% and 0.1% respectively, and were classified as bacteria other than *Vibrio*. Since they were rare and classified as other species, they were excluded from further analyses. OTU 51 (average abundance 0.12%) was identical to OTU 1 except it lacked a nucleotide stretch of approximately 120 nt. It probably represented a PCR artifact, and was also excluded from further analyses. One of the OTUs (OTU 53) was classified as the genus *Lucibacterium* but was included in further analyses, as it showed high sequence similarity to the other RD5-30 variants.

The sequence variants found for the V1-region of the RD5-30 samples were aligned (Figure 3.10). The hypervariable part of the V1-region in RD5-30 seems to encompass an area from about nucleotide (nt) 76 to 93, according to the *E.coli* 16S rRNA numbering (Figure 1.1). This area of hypervariability corresponds to the stem-loop structure found in the V1-region of *E. coli*. Compensatory mutations were observed on complimentary strands of stem structures for some of the V1-region sequence variants (Figure 3.10). Such compensatory mutations will ensure preservation of the secondary structure of the gene.

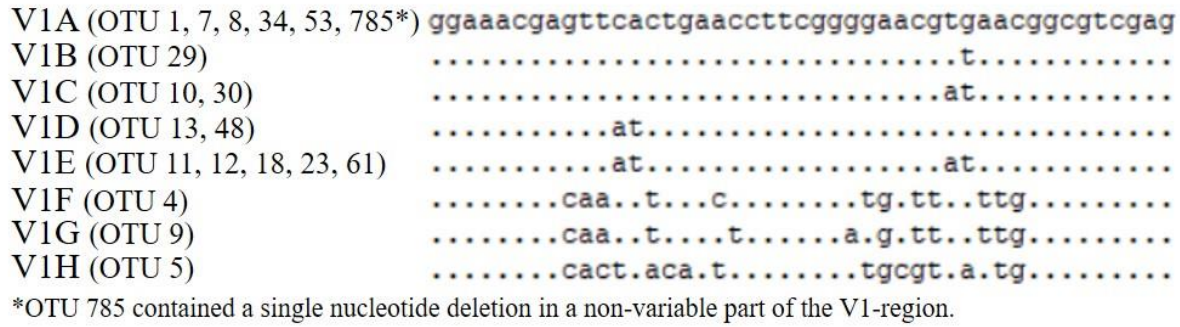


**Figure 3.10:** Sequence variants identified for the V1-region of the 16S rRNA gene in the RD5-30 samples. The OTUs containing the respective variant are listed in parentheses. Dots symbolize nucleotide positions that are identical to the reference sequence.

The different sequence variant found for the V3-region were also aligned (Figure 3.11). The alignment showed that the hypervariable part of the V3-region in the RD5-30 samples includes two stretches of the stem parts of the V3-region stem-loop secondary structure (Figure 1.1).

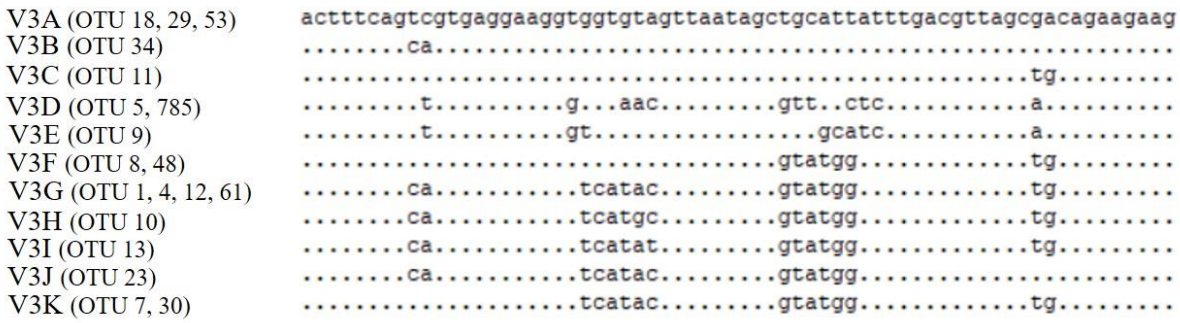


**Figure 3.11:** Sequence variants identified for the V3-region of the 16S rRNA gene in the RD5-30 samples. The OTUs containing the respective variant are listed in parentheses. Dots symbolize nucleotide positions that are identical to the reference sequence.

The combinations of V1- and V3-sequence variants in the most abundant RD5-30 OTUs together with the average OTU abundances are presented in Table 3.1. The highly similar V1A and V1B variants were present in about 66% of the sequence reads (Table 3.1). For the V3-region, the very distinct V3G and V3A sequence variants accounted for approximately 61.1% and 14.3% of the reads, respectively (Table 3.1). OTUs 5 and 785 accounted for 1.66% and

0.32% of the reads in all samples, respectively, but were practically non-existent among the 2016 fish experiment samples.

**Table 3.1:** Common sequence variants for RD5-30 showing OTU name, their V1- and V3-region sequence variants and their average abundance among all RD5-30 samples as well as just among the samples from the 2016 gnotobiotic fish experiment.

| OTU | 16S rRNA sequence variant | Average abundance % (all samples) | Average abundance % (2016 fish experiment) |
|---|---|---|---|
| OTU 1 | V1A-V3G | 50.83 | 49.82 |
| OTU 29 | V1B-V3A | 12.58 | 12.71 |
| OTU 4 | V1F-V3G | 10.14 | 11.21 |
| OTU 10 | V1C-V3H | 7.35 | 7.94 |
| OTU 11 | V1E-V3C | 3.85 | 4.10 |
| OTU 13 | V1D-V3I | 3.53 | 4.58 |
| OTU 12 | V1E-V3G | 2.11 | 2.17 |
| OTU 5 | V1H-V3D | 1.66 | 0.00007* |
| OTU 23 | V1E-V3J | 1.55 | 1.69 |
| OTU 18 | V1E-V3A | 1.44 | 1.41 |
| OTU 34 | V1A-V3B | 0.82 | 0.78 |
| OTU 8 | V1A-V3F | 0.77 | 0.55 |
| OTU 7 | V1A-V3K | 0.55 | 0.51 |
| OTU 785 | V1A**-V3D | 0.32 | 0.00006* |
| OTU 53 | V1A-V3A | 0.26 | 0.23 |
| OTU 61 | V1E-V3G | 0.17 | 0.21 |
| OTU 48 | V1D-V3F | 0.13 | 0.09 |
| OTU 9 | V1G-V3E | 0.12 | 0* |
| OTU 30 | V1C-V3K | 0.10 | 0.07 |

*Low average abundance in this experiment, but relatively abundant in 2014 fish sample, where OTU 5 = 22.7 %, OTU 785 = 4.4 and OTU 9= 1.6

** This OTU had a single nucleotide deletion in a non-variable part of the V1-region

The four V3-region 16S rRNA sequence variants found for RD5-30 through Sanger sequencing of DGGE bands (Figure 3.6) all corresponded to V3-region sequence variants found through Illumina sequencing (Figure 3.11). The Sanger sequencing variants V3I, V3II, V3III and V3IV corresponded to Illumina variants V3G, V3I, V3A and V3H respectively. V3G was found in OTUs 1, 4 12 and 61, which constituted more than 60% of the RD5-30 sequence reads. In addition to V3G, variants V3A and V3H were also found among the four most common RD5-30 OTUs (Table 3.1). V3I was found in OTU 13, which represented 3.5% of the total RD5-30 sequence reads.

Secondary structures were predicted for the most abundant RD5-30 V1- and V3-region sequence variants by the RNA secondary structure predictor by Mathews group (2017). Among the V1-region, mainly four different types of secondary structure were suggested (Figure 3.12). Predicted secondary structures for the remaining V1-region sequence variants are shown in Appendix L. V1A, V1E and V1F were found to have highly similar structures, as were V1B and V1C, and V1H and V1G (Figure 3.12, Appendix L). The predicted secondary structure of V1D is dissimilar to the other variants (Figure 3.12).

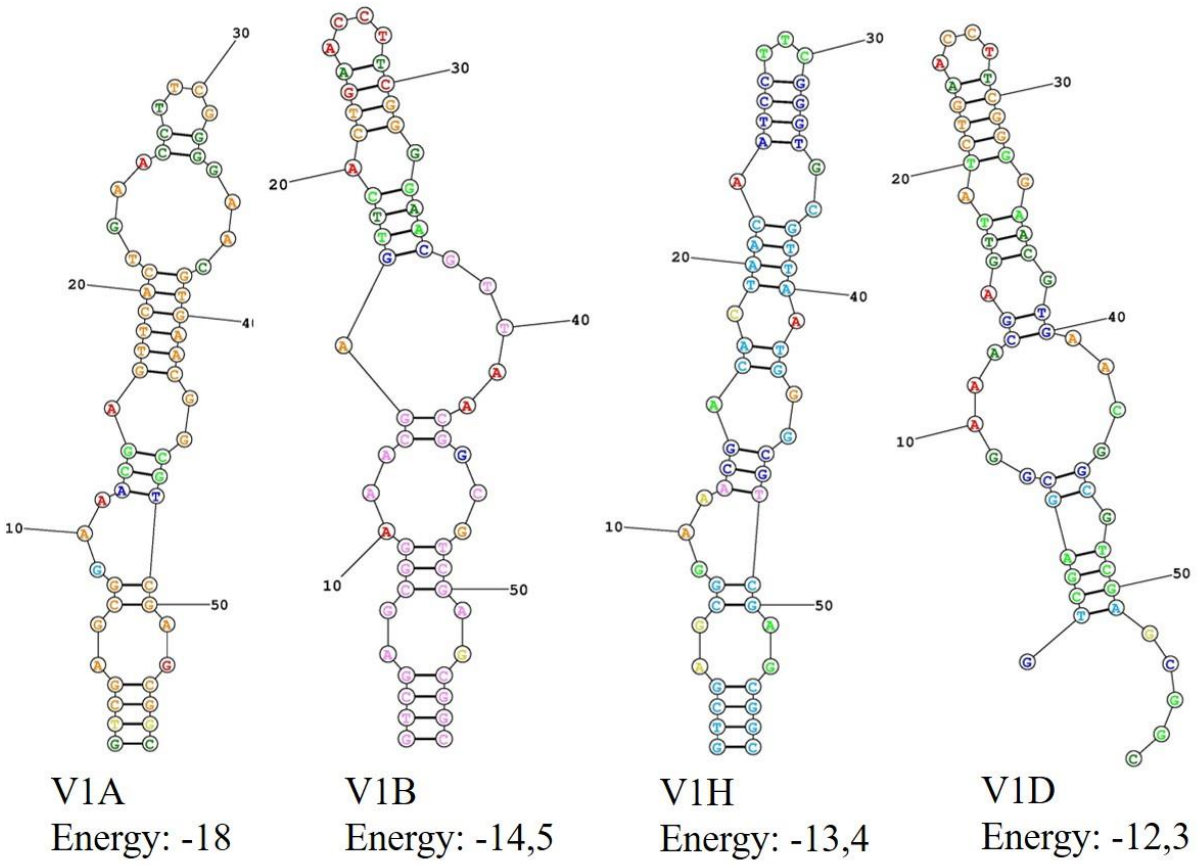


**Figure 3.12:** The four main types of suggested secondary structures suggested by the RNA secondary structure predictor by Mathews group (2017) among the sequence variants from the V1-region of the RD5-30 16S rRNA gene.

For the V3-region secondary structures, most sequence variants fell into one of four categories (Figure 3.13). The secondary structures not displayed in Figure 3.13 can be found in Appendix M. Variants V3G, V3A, V3H, V3I, V3D and V3E were suggested to all have highly similar secondary structures. These were the most energetically favorable structures out of the four different categories of secondary structures (Figure 3.13, Appendix M). V3C and V3K shared

similar suggested secondary structures, as did V3J and V3B, all less energetically favorable than the previously mentioned structures. V3F had a different looking structure than the others (Figure 3.13).

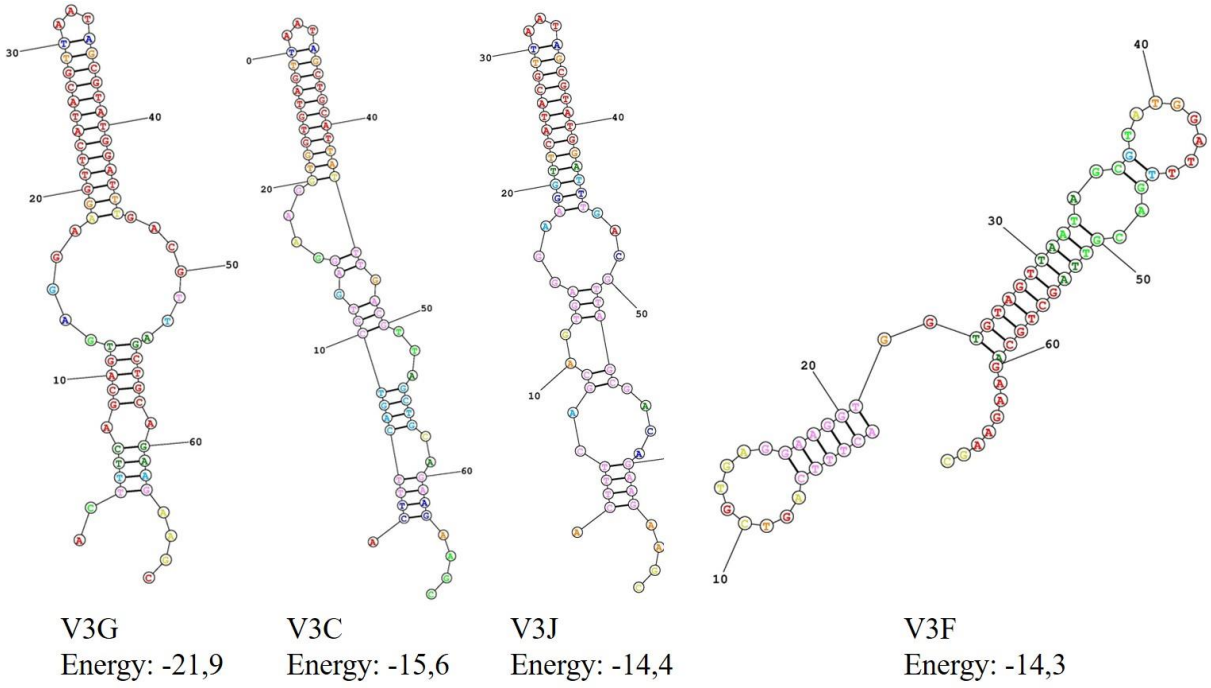


**Figure 3.13:** The main types of secondary structure suggested by the RNA secondary structure predictor by Mathews group (2017) for the sequence variants from the V3-region of the RD5-30 16S rRNA gene.

### 3.3.4 Comparisons of 16S rRNA gene variant profiles between groups of RD5-30 samples

Principal coordinates (PCoA) analysis based on Bray-Curtis similarities was performed to compare the RD5-30 16S rRNA OTU profiles among samples (Figure 3.14a). The fish samples from the 2014 cod larvae experiment were very distinct from all other sample groups in the plot (Figure 3.14a). Therefore, these samples were excluded to allow for better comparisons of the remaining samples in a new PCoA plot (Figure 3.14b). In this plot, all samples seem to cluster together according to sample type, seen most clearly for the 2016 fish experiment DNA and RNA water samples, and M65 cultured samples from the exponential phase.

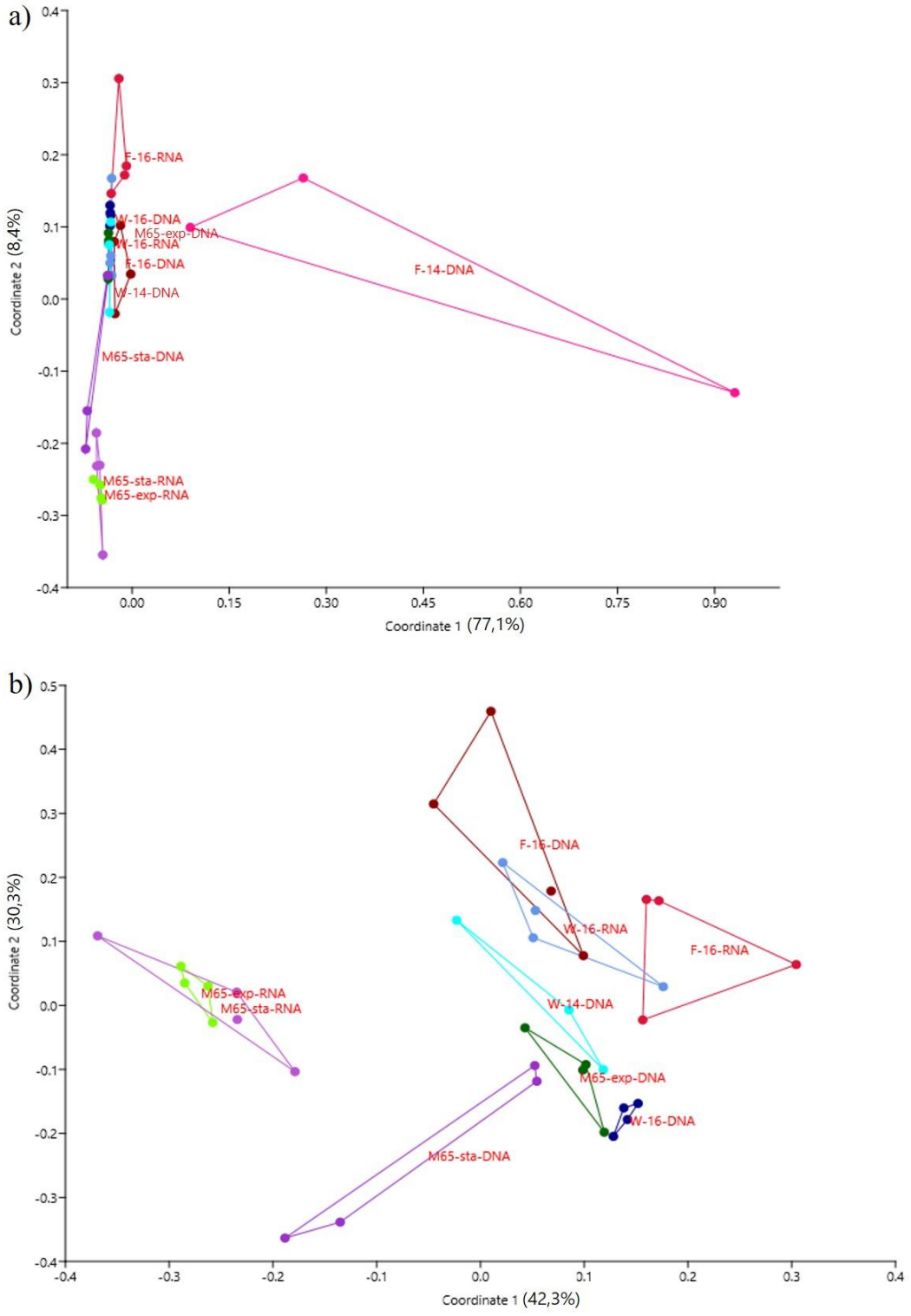


**Figure 3.14:** Principal coordinate analysis (PCoA) plots based on Bray-Curtis similarities for comparison of RD5-30 16S rRNA OTU profiles for a) all samples except water samples from flasks without cod larvae in the 2016 cod experiment, and b) all samples with the exception of the fish samples from the 2014 cod larvae experiment. Samples labels: F: fish; W: rearing water; M65: from M65 cultures (growth experiment), M65-exp: exponential growth phase, M65-sta: stationary growth phase. RNA indicates amplicons based on RNA extracts and DNA indicates amplicons based on DNA extracts.

One-way PERMANOVA tests also using the Bray-Curtis dissimilarity indices were conducted to test whether there were significant differences in the RD5-30 OTU profiles between groups of samples (Table 3.2). These tests revealed significant differences between most groups (P<0.5). These exceptions included DNA extracted from the 2014 cod larvae experiment when compared with several different groups of samples, RNA extracted from M65 stationary phase samples compared with RNA extracted from M65 exponential phase samples, as well as RNA extracted from water samples from the 2016 cod larvae experiment compared with some different sample groups (Table 3.2).

**Table 3.2:** One-way PERMANOVA for testing differences in the Bray-Curtis dissimilarities between groups of samples. P-values indicating significant differences (p<0.05) are written in bold.

| | **F-16-DNA** | **F-16-RNA** | **W-16-DNA** | **W-16-RNA** | **M65-EXP-DNA** | **M65-STA-DNA** | **M65-EXP-RNA** | **M65-STA-RNA** | **W-14-DNA** | **F-14-DNA** |
|---|---|---|---|---|---|---|---|---|---|---|
| **F-16-DNA** | | **0.028** | **0.029** | 0.085 | **0.031** | **0.030** | **0.031** | **0.030** | 0.116 | **0.027** |
| **F-16-RNA** | **0.028** | | **0.029** | 0.056 | **0.029** | **0.026** | **0.029** | **0.029** | 0.059 | **0.032** |
| **W-16-DNA** | **0.029** | **0.029** | | **0.030** | **0.028** | **0.029** | **0.027** | **0.026** | **0.029** | **0.027** |
| **W-16-RNA** | 0.085 | 0.056 | **0.030** | | **0.028** | **0.025** | **0.028** | **0.028** | 0.120 | **0.030** |
| **M65-EXP-DNA** | **0.031** | **0.029** | **0.028** | **0.028** | | **0.033** | **0.029** | **0.029** | 0.086 | **0.027** |
| **M65-STA-DNA** | **0.030** | **0.026** | **0.029** | **0.025** | **0.033** | | **0.028** | **0.028** | 0.172 | **0.028** |
| **M65-EXP-RNA** | **0.031** | **0.029** | **0.027** | **0.028** | **0.029** | **0.028** | | 0.801 | **0.028** | **0.031** |
| **M65-STA-RNA** | **0.030** | **0.029** | **0.026** | **0.028** | **0.029** | **0.028** | **0.801** | | **0.031** | **0.029** |
| **W-14-DNA** | 0.116 | 0.059 | **0.029** | 0.120 | 0.086 | 0.172 | **0.028** | **0.031** | | 0.096 |
| **F-14-DNA** | **0.027** | **0.032** | **0.027** | **0.030** | **0.027** | **0.028** | **0.031** | **0.029** | 0.096 | |

## Comparison of OTU profiles between fish and water samples from the 2016 gnotobiotic cod larvae experiment

The PCoA plot (Figure 3.14b) indicated differences in the RD5-30 OTU profiles between water and fish samples based on DNA from the 2016 cod larvae experiment, and a PERMANOVA test confirmed that there were significant differences (Table 3.2). To further investigate which OTUs contributed the most to this difference, a SIMPER analysis was conducted (Table 3.3). The SIMPER analysis showed that the OTUs 29, 11, 13, 1 and 4 contributed the most to the difference between the two groups of samples. Those five OTUs combined accounted for 81.8%

of all the dissimilarity found between the groups. OTU 29 was more common in water samples (15.9%) than in fish samples (10.2%) and this was the biggest contributing factor to the difference between the two groups (32.1%, Table 3.3). OTUs 11, 13 and 1 all also contributed to more than 12% of the dissimilarity, and while OTU 11 was more common in water samples, OTUs 13 and 1 were more common in fish samples (Table 3.3).

**Table 3.3:** SIMPER analysis based on Bray-Curtis dissimilarities for identification of the OTUs contributing most to the differences in the RD5-30 16S rRNA OTU profiles between fish and water samples from the 2016 gnotobiotic cod larvae experiment.

| Taxon | Average dissimilarity | Contributing % | Cumulative % | Mean abundance % F-16-DNA | Mean abundance % W-16-DNA |
|---|---|---|---|---|---|
| OTU 29 | 2.938 | 32.08 | 32.08 | 10.2 | 15.9 |
| OTU 11 | 1.372 | 14.97 | 47.06 | 2.6 | 5.3 |
| OTU 13 | 1.19 | 12.99 | 60.04 | 5.8 | 3.5 |
| OTU 1 | 1.14 | 12.44 | 72.49 | 51.9 | 49.7 |
| OTU 4 | 0.8529 | 9.312 | 81.8 | 11.4 | 9.7 |

In the fish samples from the 2016 cod larvae experiment, the RNA- and DNA-based samples seemed distinct on the PCoA plot (Figure 3.14b) and showed significant differences in the PERMANOVA tests (Table 3.2). A SIMPER test was conducted to examine the differences between these two groups in more detail (Table 3.4). Between the fish DNA and RNA based samples, OTU 1 contributed the most to the dissimilarity between the groups this time (31.9%), while OTU 29 contributed the second most (18.2%). OTU 1 was more abundant in samples based on DNA than RNA, while the opposite was true for OTU 29 (Table 3.4).

**Table 3.4:** SIMPER analysis based on Bray-Curtis dissimilarities for identification of the OTUs contributing most to the differences in the RD5-30 16S rRNA OTU profiles between 2016 gnotobiotic cod larvae experiment fish DNA and RNA based samples.

| Taxon | Average dissimilarity | Contributing % | Cumulative % | Mean abundance % F-16-DNA | Mean abundance % F-16-RNA |
|---|---|---|---|---|---|
| OTU 1 | 2.374 | 31.88 | 31.88 | 51.9 | 47.4 |
| OTU 29 | 1.358 | 18.23 | 50.11 | 10.2 | 12.2 |
| OTU 13 | 0.7936 | 10.66 | 60.77 | 5.8 | 4.8 |
| OTU 11 | 0.71 | 9.534 | 70.3 | 2.6 | 4.0 |
| OTU 4 | 0.5463 | 7.337 | 77.64 | 11.4 | 11.4 |

**Comparison of OTU profiles between fish and water samples from the 2014 gnotobiotic cod larvae experiment**

PCoA analysis (Figure 3.14b) indicated that the OTU profiles for water and fish samples from the 2014 gnotobiotic cod larvae experiment were different, but according to the PERMANOVA test ($p>0.05$) the two groups were not significantly different (Table 3.2). A small number of replicates (3 each for water and fish) and big differences in OTU profiles within the fish samples could explain this. A SIMPER analysis was conducted on these two sample groups as well (Table 3.5).

**Table 3.5:** SIMPER analysis based on Bray-Curtis dissimilarities for identification of the OTUs contributing most to the differences in the RD5-30 16S rRNA OTU profiles between water and fish samples from the 2014 gnotobiotic cod larvae experiment.

| Taxon | Average dissimilarity | Contributing % | Cumulative % | Mean abundance % W-14-DNA | Mean abundance % F-14-DNA |
|---|---|---|---|---|---|
| OTU 5 | 11.79 | 35.71 | 35.71 | 0 | 22.7 |
| OTU 1 | 8.326 | 25.22 | 60.93 | 52.3 | 36.2 |
| OTU 29 | 2.689 | 8.144 | 69.08 | 13.5 | 8.3 |
| OTU 785 | 2.267 | 6.867 | 75.94 | $2.2 \cdot 10^{-04}$ | 4.4 |
| OTU 4 | 2.187 | 6.623 | 82.57 | 10.2 | 6.7 |

In this SIMPER analysis, OTU 5 contributed most to the difference between the groups (35.7%) and was by far most common in fish samples (22.7%) as it was not observed in the water sample profiles (0%). OTU 1 (25.2%) and OTU 29 (8.1%) contributed the second and third most to the dissimilarity, and were both more abundant in water samples. OTU 785 contributed the fourth most to the difference (6.9%) and was most commonly found in fish samples (4.4%) and only barely found in water samples ($2.2 \cdot 10^{-04}$ %). OTU 4 was most common in water samples and had the fifth biggest contribution to the dissimilarity between the water and fish sample groups (6.6%). Interestingly, OTUs 5 and 785 are almost only present in the fish DNA based samples from the 2014 experiment.

**Comparison of OTU profiles between RNA and DNA based samples from the growth experiment**

In the growth experiment using as the M65 growth medium, RNA and DNA based samples were taken from both the exponential and stationary stages of growth. The PCoA plot (Figure 3.14b) indicated that the two groups of samples from this experiment with the biggest differences were between the OTU profiles based on RNA and DNA. The exponential and stationary samples were therefore grouped together, so that all RNA based samples could be compared with all DNA based samples from the growth experiment by way of SIMPER analysis (Table 3.6). The five OTUs contributing the most to the difference between the two groups are the same as seen in Table 3.3 and Table 3.4. The two most common OTUs, 1 and 29, in the two sample groups are also the two OTUs contributing the most to the difference between the groups with 34.5% and 27.5% respectively. OTU 1 is more abundant in the RNA based samples than the DNA, indicating a higher expression level of this OTU compared to its abundance in the genome.

**Table 3.6:** SIMPER analysis based on Bray-Curtis dissimilarities for identification of the OTUs contributing most to the differences in the RD5-30 16S rRNA OTU profiles between DNA and RNA based samples from the growth experiment.

| Taxon | Average dissimilarity | Contributing % | Cumulative % | Mean abundance % M65-DNA | Mean abundance % M65-RNA |
|---|---|---|---|---|---|
| OTU 1 | 2.857 | 34.47 | 34.47 | 51.6 | 57.2 |
| OTU 29 | 2.279 | 27.49 | 61.96 | 15.2 | 10.7 |
| OTU 11 | 0.7647 | 9.223 | 71.18 | 4.6 | 3.1 |
| OTU 13 | 0.6627 | 7.993 | 79.17 | 2.6 | 1.8 |
| OTU 4 | 0.5545 | 6.688 | 85.86 | 9.8 | 9.4 |

# 4 DISCUSSION

## 4.1 The use of denaturing gradient gel electrophoresis (DGGE) and Illumina sequencing to map sequence heterogeneity

DGGE was chosen as a method to investigate intragenomic heterogeneity in the 16S rRNA gene of *Vibrio* as a previous DGGE analysis suggested extensive sequence variety in the 16S V3-region of *Vibrio* strain RD5-30 (Figure 1.3). In a DGGE analysis, DNA molecules with different sequences are separated according to their denaturation in a denaturing gradient (Muyzer et al., 1993). It was therefore assumed that different bands in the denaturing gradient (Figure 1.3) represented different DNA sequences. The Sanger sequencing of the 16S rRNA V3-region DGGE bands from the 2016 gnotobiotic fish experiment, however, revealed that the three sequenced HI610 bands shared the exact same sequence, as did the eleven bands sequenced for BB120. The 18 RD5-30 bands sequenced yielded only four different sequence variants, with most bands again sharing one identical sequence (Figure 3.6). Even bands located far apart in the denaturing gradient represented the same sequence, such as bands 4 and 14 and bands 19 and 25 (Figure 3.4). The four RD5-30 16S rRNA sequence variants that were found through Sanger sequencing of DGGE bands were also present in Illumina sequencing results, which indicates they are actual V3-region 16S rRNA variants found in RD5-30. DGGE is already known to be an unreliable method for directly estimating species diversity because of intragenomic sequence heterogeneity, which leads to an overestimation of diversity (Dahllöf et al., 2000, Vadstein et al., 2013). The findings of these DGGE analyses – different bands representing the same sequence – might indicate there is an even bigger overestimation of species diversity when using DGGE as a method, caused by DNA molecules of identical sequences giving rise to several DGGE bands. The reason for this is unclear.

The distribution of the bands seem to be replicable, as all band profiles were very similar in both gels based on both DNA and RNA extracts for any given strain, as can be seen in the two V3-region 16S DGGE gels (Figure 3.4 and 3.5). As argued by Dahllöf et al. (2000), this reproducible band pattern indicates that the many bands representing the same sequence was not a result of random PCR-related errors. One possible explanation might be that DNA molecules with a specific sequence might fold into several alternative, stable secondary structures that exhibit different behavior in the denaturing gradient. This has previously been

suggested by Neilson et al. (2013). In summary, the DGGE-profiles of PCR products that can fold into secondary structures do not necessarily reflect sequence variation in the PCR product.

In the case of Illumina amplicon sequencing to map sequence heterogeneity, the problems encountered using DGGE analyses were avoided. The secondary structures of the 16S rRNA sequence variants did not have any impact on the sequences found by Illumina sequencing, and the method gave a detailed description and overview over the different gene variants found within the three *Vibrio* strains. The method does however have its drawbacks, as hundreds of OTUs classified as other taxa than *Vibrio* were also included in the resulting OTU table. These OTUs might occur because of contamination from the water – the FASW used in the fish experiment as well as the M65 medium has been autoclaved, but while this heating process destroys bacteria, it does not necessarily degrade all DNA in the seawater. Some of that DNA or RNA might be extracted along with the desired *Vibrio* DNA or RNA. Rare OTUs that were classified as species other than *Vibrio* were likely to represent such contaminations.

## 4.2 Presence of different 16S rRNA gene variants in the three *Vibrio* strains

The DGGE and Illumina sequencing analysis of the three *Vibrio* strains revealed that only one 16S rRNA gene variant was abundant for both HI610 and BB120. In RD5-30, 8 different V1-region sequence variants and 11 different V3-region variants of the 16S rRNA gene were found (Figure 3.10 and 3.11). There were 19 different OTUs with average abundances above 0.10% in RD5-30, each with a novel V1- and V3-region sequence variant combination (Table 3.1). In all three *Vibrio* strains, the suggested secondary structures included G-T base pairs – while not one of the regular base pairings, the G-U wobble pair is common and important in RNA (Varani and McClain, 2000) and as such does not discredit the validity of these 16S rRNA sequences. These two regions exhibited quite extensive sequence variation between the variants, with up to 18 variable nucleotide positions for both the V1- and V3-region. The conservation of gene copies as divergent as this indicates that they have different biological functions, which has previously been suggested by Jensen et al. (2009). This biological function is likely be related to ribosome assembly and structure, and might be relevant for recruitment of ribosomal proteins (Mayerle and Woodson, 2013). Both the V1- and V3-regions of the 16S rRNA gene have been found to interact with protein S4, which is among the first two proteins to interact with 16S

rRNA during protein assembly (Sapag et al., 1990). This interaction causes conformational changes in the 16S rRNA, potentially affecting its secondary and tertiary structure (Mayerle and Woodson, 2013). In *E. coli* there was not found much sequence variation between 16S rRNA gene copies, yet the promoters of the eight different rRNA operon promoters were still differentially regulated when the bacteria was living with different lifestyles (Maeda et al., 2015).

The *Vibrio* genome has already shown a high degree of plasticity, as well as having two chromosomes, and is often a target of recombination (Thompson et al., 2010). Many species of *Vibrio* have a high 16S rRNA gene copy number (Regents of the University of Michigan, 2017). The genome of RD5-30 has not been sequenced, but since 19 different OTUs were found through Illumina sequencing (Table 3.1), this would indicate 19 copies of the rRNA operon. This is more than has been seen in *Vibrio*, where the highest recorded operon copy number seems to be 14 (Regents of the University of Michigan, 2017). This seems to indicate an overestimation of 16S rRNA copy number in RD5-30, meaning that not all of the sequence variants represent existing functional gene variants, which is something that will be discussed further below. Furthermore, all observed OTUs might not all exist in the same bacterial genome - there might be a different composition of OTUs in different RD5-30 cells - which might especially be true for the OTUs with the lowest abundance, which might not be present in all RD5-30 bacteria. The different 16S rRNA variants in *Vibrio* sp. RD5-30 might have originated from recombination by horizontal gene transfer from other *Vibrio* species mediated by bacteriophages that are abundant in the marine environments they live in, as discussed by Harth et al. (2007).

When looking into the predicted secondary structures of the V1- and V3-region sequence variants, it became apparent that some sequence variants had quite different structures, such as V1-region variant V1A when compared to V1D (Figure 3.12) or even more noticeable in the V3-region with e.g. sequence variant V3G compared to V3C, V3J or V3F (Figure 3.13). From the V3-region secondary structures especially, it became apparent that some sequence variants had deviating structures because they were lacking base pairings in the stem-regions, causing their predicted secondary structures to look unusual, while also being less energetically favorable. When looking closer into the nucleotide sequences of the deviating secondary

structures (Figure 3.10 and Figure 3.11) it became apparent that some of the V1- and V3-region sequence variants found in RD5-30 seemed to be composed of two different template sequences that had combined into one sequence. These potential chimeras could be spotted by their lack of complimentary base pairings in the stem regions of the hypervariable sections of V1 and V3. This can most clearly be seen for sequence variant V3F, where it seems half of the sequence is identical to variant V3A, while the other half is identical to e.g. V3G, leaving this sequence variant with many nucleotides lacking complimentary base pairings in their secondary structure (Figure 3.11 and Figure 3.13). This same trend was apparent in two (V1C, V1D; Figure 3.10) of the V1-region sequence variants and four (V3B, V3C, V3J, V3K; Figure 3.11) other V3-region 16S rRNA gene variants of RD5-30. These potentially chimeric sequence variants were found in 9 (OTU 10, 11, 13, 23, 34, 8, 7, 48, 30; Table 3.1) of the 19 most common OTUs for RD5-30, on average accounting for around 18.7% of the reads.

Chimeras are known to be synthesized during PCR due to the polymerase shifting from one template DNA molecule to another during the elongation step (Smyth et al., 2010). It is worth noting that none of the four sequence variants found during Sanger sequencing of RD5-30 V3-region 16s rRNA corresponded to any of the potentially chimeric sequences, which might mean the differences in PCR protocols combined with the much higher resolution of the Illumina sequencing method led to the chimeras only being observed here. There is a possibility that these are genuine biological sequence variants found in RD5-30, potentially caused by the recombination of 16S rRNA genes between bacterial cells, as mentioned above (Harth et al., 2007). If bacteria benefit from changing the abundance of different 16S sequence variants while living with different lifestyles, the chimeras might just be intermediate recombination products caused by the selection for more beneficial sequence variants in their current environments. All of the OTUs with potentially chimeric sequence variants only have either a chimeric V1-region or V3-region variant, never both. Some of the OTUs with potentially chimeric sequence variants also have a unique non-chimeric variant, such as OTU 10 with a potential chimeric sequence variant V1C and unique sequence variant V3H (Table 3.1). Because of the potential biological legitimacy of these chimeric sequences, all 19 common RD5-30 OTUs were included in all analyses performed.

## 4.3 Distribution of 16S rRNA gene variants in different lifestyles of RD5-30

Statistical analyses showed that there was a significant difference in the distribution of different RD5-30 16S rRNA sequences between different sample types – including differences between samples representing different life strategies and differences between PCR products based on DNA and RNA extractions (Table 3.2). Firstly, there were significant differences at the RNA level between different sample types, such as between 2016 gnotobiotic experiment fish samples and samples for RD5-30 grown in M65 medium (Figure 3.14b). These differences indicate that the bacteria regulate gene expression of the different sequence variants depending on which sequence variants are needed in a given environment or lifestyle.

Differences were also observed in the abundances of different sequence variants when comparing gene variant profiles based on DNA and RNA. This was seen when comparing DNA and RNA based gene variant profiles from the 2016 gnotobiotic fish experiment (Table 3.4) and the growth experiment (Table 3.6). In both cases some OTUs were found to have a higher mean abundance in RNA based profiles than in those based on DNA, indicating upregulation of the expression of these RNAs. This also indicates that the expression of the gene variants are regulated at the RNA level. Different OTUs appeared to be upregulated in these two sample groups, indicating a difference in gene expression in the two different lifestyles. A regulation of 16S rRNA variants at the RNA level has been found in e.g. *E. coli* (Maeda et al., 2015) and will result in a rapid response in the bacterial cell.

An interesting finding was that there were significant differences between sample groups at the DNA level (Table 3.2). For example, a difference in the mean abundance of different sequence variants were found when comparing fish and water DNA based gene variant profiles from the 2016 gnotobiotic fish experiment (Table 3.3). There was at the most about a 5.7% difference in the mean abundance of the same OTU between the two sample groups. This indicates that the *Vibrio* species living in different environments may regulate the abundance of different 16S rRNA gene variants at the DNA level. A possible mechanism for this could be horizontal gene transfer and recombination of the 16S rRNA copies, changing the abundances of the 16S rRNA gene variants of the bacterial cells. This could have happened through gene conversion by homologous recombination within the genome of a RD5-30 bacterium, potentially increasing

the abundance of some sequence variants and decreasing the abundance of others by copying parts of the 16S rRNA gene from one sequence variant to another, making them identical (Gonzalez-Escalona et al., 2005).

Another example of differing gene variant profiles at the DNA level was observed when comparing the rearing water and fish samples based on DNA from the 2014 gnotobiotic fish experiment. Interestingly, the gene variant profiles these fish samples contained high abundances of two OTUs (5 and 785) that were barely found in any other sample group, including the 2016 fish experiment (Table 3.1). There is a possibility that these two OTUs are contaminants. When compared with the other common RD5-30 OTUs, OTU 5 along with OTU 9, which was also almost exclusively found in these fish samples, has as many as twelve nucleotides differing from the remaining OTUs outside of the hypervariable regions of V1 and V3. This makes these OTUs likely to represent contaminating bacteria, which would also mean there has been found 17 and not 19 legitimate, common RD5-30 sequence variants. Disregarding these two OTUs also leaves 6 V1-region variants and 10 V3-region variants, and the highest number of observed variable nucleotide positions is then 16 nucleotides for both the V1- and V3-region variants (Figure 3.10 and Figure 3.11).

OTU 785, on the other hand, is more similar to the other RD5-30 16S rRNA gene variants with a near identical sequence as the other gene variants outside of the variable regions and a V1-region highly similar to the most common V1-variant, but with a single nucleotide deletion. However, it does have a V3-region variant shared only with OTU 5. It is possible that the acquisition of the OTU 785 sequence variant is a result of recombination by RD5-30 with bacteria containing OTU 5, maybe as a form of adaption to its lifestyle of colonizing the fish. There is also a possibility it might be a chimeric PCR artifact, though its unique V1A sequence variant makes this somewhat less likely. The 2014 gnotobiotic cod larvae experiment samples were from day 13 of that experiment, while the 2016 experiment ended at day 6 post-hatch. A longer fish experiment might lead to an increase in the differences in sequence variant abundances between RD5-30 living in fish versus rearing water.

It is worth noting that all experiments conducted and analyzed in this project only used a small number of replicates (3-4) per sample group. Moreover, even when the groups were found to be significantly different these differences were often minor and generally due to relatively small variations in abundances. More replicates would therefore help in verifying the validity of the findings in this thesis.

## 4.4 Future work

The results from this project showed that *Vibrio* sp. RD5-30 has several distinct 16S rRNA gene variants, and indicate that these gene variants may be regulated at both the DNA and RNA level, according to lifestyle. Future work should attempt to confirm these findings. One way to do this is to conduct another gnotobiotic fish experiment with RD5-30 over a longer period of time, which might allow for larger differences in 16S rRNA gene variant abundances between fish and rearing water. In addition, the inclusion of more replicates of each sample group allow for more robust statistical analyses. Other lifestyles could also been studied, such as biofilm growth. More research should be done on why the V1- and V3-regions of the 16S rRNA gene in RD5-30 exhibit such a high level of variability, and what the biological significance of these hypervariable regions is. The possible biological significance of the potentially chimeric sequences should be investigated also. Possible mechanisms to control the expression of 16S rRNA gene variants at both the DNA and RNA level would be interesting to investigate. A potential approach could be to genetically manipulate RD5-30 e.g. by knocking out some 16S sequence variants and then observe how the growth and 16S sequence variant abundance changes in the bacteria. However, this might not be so easy for the *Vibrio* sp. RD5-30, where the appropriate tools for gene manipulation are limited. Furthermore, it could be interesting to investigate how bacteria in genera other than *Vibrio* regulate the expression of their 16S rRNA gene variants. One candidate for this is *Photobacterium profundum* strain SS9, which has been found to have 15 rRNA operons distributed between two chromosomes and a high intragenomic sequence variation within the 16S rRNA genes (Vezzi et al., 2005). Another alternative is *Clostridium paradoxum* DSM 7308, which has 15 V1-region 16S rRNA gene sequence variants (Rainey et al., 1996).

# 5 CONCLUSION

DGGE turned out to be a poor method for investigating intragenomic 16S rRNA sequence variation, as many distinct bands within DGGE profiles represented identical sequences. No intragenomic 16S rRNA V1- or V3-region heterogeneity was observed for HI610 or BB120, neither through Sanger sequencing of DGGE bands nor Illumina amplicon sequencing. The Illumina sequencing of RD5-30 16S rRNA amplicons revealed a large degree of intragenomic variation, located in hypervariable nucleotide stretches in stem-regions of the V1- and V3-regions. Sequencing of RD5-30 revealed 6 V1-region and 10 V3-region sequence variants in the 16S rRNA gene. There were found 17 relatively abundant OTUs (i.e. gene variants) for RD5-30, each with unique combinations of V1- and V3-region sequence variants. Based on inspection of suggested secondary structures, some of the RD5-30 16S rRNA sequence variants appeared to be chimeric combinations of different sequences. It is difficult to determine whether these sequences are biologically relevant or artifacts from the PCR amplification. The various gene variants might have different functions, mediated by variations in secondary and tertiary structure of the 16S rRNA, and in interaction with ribosomal proteins (e.g. S4), which can affect ribosome assembly. If the 16S rRNA gene variants do have different functions, a differential expression in distinct environments and lifestyles seems plausible. There were found significant differences in the abundance of gene variants between sample groups at both the DNA and RNA level. This applied to comparisons between fish and water samples (gnotobiotic fish experiment) and between growth experiment samples. The expression of rRNA sequence variants seemed to be regulated at the RNA level, and seemed to be regulated differently depending on the lifestyle of the bacteria, such as in RD5-30 colonizing fish versus growing in M65 medium. There were also significant differences in gene variant abundances at the DNA level between fish and rearing water samples from the gnotobiotic fish experiments. A possible explanation for this observation is that RD5-30 adjusts the abundance of 16S rRNA variants in a cell population at the DNA level, e.g. through some sort of recombination. Future research should focus on determining the biological function of hypervariable, intragenomic 16S rRNA variation. More research is needed to understand the mechanism for creation of this intragenomic hypervariability and for the regulation of the abundance of the different sequence variants at the DNA and RNA level.

# 6 REFERENCES


ACINAS, S. G., MARCELINO, L. A., KLEPAC-CERAJ, V. & POLZ, M. F. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol,* 186**,** 2629-35.

ALBERTS, B., JOHNSON, A., LEWIS, J., MORGAN, D., RAFF, M., ROBERTS, K. & WALTER, P. 2015. *Molecular Biology of the Cell,* United States of America, Garland Science Taylor & Francis Group.

AMANN, R. I., LUDWIG, W. & SCHLEIFER, K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev,* 59**,** 143-69.

AMATO, K. R. 2017. An introduction to microbiome analysis for human biology applications. *Am J Hum Biol,* 29.

ANDERSON, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology,* 26**,** 32-46.

BAKKE, I., COWARD, E., ANDERSEN, T. & VADSTEIN, O. 2015. Selection in the host structures the microbiota associated with developing cod larvae (Gadus morhua). *Environ Microbiol,* 17**,** 3914-24.

BD BIOSCIENCES 2000. Introduction to Flow Cytometry: A Learning Guide. San Jose, CA, United States of America: Becton, Dickinson and Company.

BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & WHEELER, D. L. 2005. GenBank. *Nucleic Acids Research,* 33**,** D34-D38.

BOUTIN, S., BERNATCHEZ, L., AUDET, C. & DEROME, N. 2012. Antagonistic effect of indigenous skin bacteria of brook charr (Salvelinus fontinalis) against Flavobacterium columnare and F. psychrophilum. *Vet Microbiol,* 155**,** 355-61.

CILIA, V., LAFAY, B. & CHRISTEN, R. 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol,* 13**,** 451-61.

CLARRIDGE, J. E. 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clin Microbiol Rev,* 17**,** 840-62.

CONDON, C., LIVERIS, D., SQUIRES, C., SCHWARTZ, I. & SQUIRES, C. L. 1995. rRNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation. *J Bacteriol,* 177**,** 4152-6.

DAHLLÖF, I., BAILLIE, H. & KJELLEBERG, S. 2000. rpoB-Based Microbial Community Analysis Avoids Limitations Inherent in 16S rRNA Gene Intraspecies Heterogeneity. *Appl Environ Microbiol,* 66**,** 3376-80.

DETHLEFSEN, L. & SCHMIDT, T. M. 2007. Performance of the translational apparatus varies with the ecological strategies of bacteria. *J Bacteriol,* 189**,** 3237-45.

DI BELLA, J. M., BAO, Y., GLOOR, G. B., BURTON, J. P. & REID, G. 2013. High throughput sequencing methods and analysis for microbiome research. *Journal of Microbiological Methods,* 95**,** 401-414.

DORSCH, M., LANE, D. & STACKEBRANDT, E. 1992. Towards a phylogeny of the genus Vibrio based on 16S rRNA sequences. *Int J Syst Bacteriol,* 42**,** 58-63.

DUAN, J., REIMER, L., HEIKKILA, J. J. & GLICK, B. R. 2014. Differential expression of the seven rRNA operon promoters from the plant growth-promoting bacterium Pseudomonas sp. UW4. *FEMS Microbiol Lett,* 361**,** 181-9.

FJELLHEIM, A. J., KLINKENBERG, G., SKJERMO, J., AASEN, I. M. & VADSTEIN, O. 2010. Selection of candidate probionts by two different screening strategies from Atlantic cod (Gadus morhua L.) larvae. *Vet Microbiol,* 144**,** 153-9.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.

FORBERG, T., ARUKWE, A. & VADSTEIN, O. 2011a. A protocol and cultivation system for gnotobiotic Atlantic cod larvae (Gadus morhua L.) as a tool to study host microbe interactions. *Aquaculture,* 315**,** 222-227.

FORBERG, T., ARUKWE, A. & VADSTEIN, O. 2011b. Two strategies to unravel gene expression responses of host–microbe interactions in cod (Gadus morhua) larvae. *Aquaculture Research,* 42**,** 664-676.

FRANS, I., MICHIELS, C. W., BOSSIER, P., WILLEMS, K. A., LIEVENS, B. & REDIERS, H. 2011. Vibrio anguillarum as a fish pathogen: virulence factors, diagnosis and prevention. *J Fish Dis,* 34**,** 643-61.

FRAUNE, S. & BOSCH, T. C. 2010. Why bacteria matter in animal development and evolution. *Bioessays,* 32**,** 571-80.

FUKUDA, K., OGAWA, M., TANIGUCHI, H. & SAITO, M. 2016. Molecular Approaches to Studying Microbial Communities: Targeting the 16S Ribosomal RNA Gene. *J uoeh,* 38**,** 223-32.

GLAESER, S. P. & KAMPFER, P. 2015. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol,* 38**,** 237-45.

GONZALEZ-ESCALONA, N., ROMERO, J. & ESPEJO, R. T. 2005. Polymorphism and gene conversion of the 16S rRNA genes in the multiple rRNA operons of Vibrio parahaemolyticus. *FEMS Microbiol Lett,* 246**,** 213-9.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-351.

GRAY, M. W., SANKOFF, D. & CEDERGREN, R. J. 1984. On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Research,* 12**,** 5837-5852.

HAMMER, Ø., HARPER, D. A. T. & RYAN, P. D. 2001. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontol Electron,* 4**,** 9.

HARTH, E., ROMERO, J., TORRES, R. & ESPEJO, R. T. 2007. Intragenomic heterogeneity and intergenomic recombination among Vibrio parahaemolyticus 16S rRNA genes. *Microbiology,* 153**,** 2640-7.

HEAD, I. M., SAUNDERS, J. R. & PICKUP, R. W. 1998. Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microb Ecol,* 35**,** 1-21.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

HEIDELBERG, J. F., EISEN, J. A., NELSON, W. C., CLAYTON, R. A., GWINN, M. L., DODSON, R. J., HAFT, D. H., HICKEY, E. K., PETERSON, J. D., UMAYAM, L., GILL, S. R., NELSON, K. E., READ, T. D., TETTELIN, H., RICHARDSON, D., ERMOLAEVA, M. D., VAMATHEVAN, J., BASS, S., QIN, H., DRAGOI, I., SELLERS, P., MCDONALD, L., UTTERBACK, T., FLEISHMANN, R. D., NIERMAN, W. C., WHITE, O., SALZBERG, S. L., SMITH, H. O., COLWELL, R. R., MEKALANOS, J. J., VENTER, J. C. & FRASER, C. M. 2000. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature,* 406**,** 477-83.

ILLUMINA. 2016. *An Introduction to Next-Generation Sequencing Technology* [Online]. Illumina, Inc. Available: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf [Accessed 26.09.2017].
JAIN, R., RIVERA, M. C. & LAKE, J. A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A,* 96**,** 3801-6.
JENSEN, S., FROST, P. & TORSVIK, V. L. 2009. The nonrandom microheterogeneity of 16S rRNA genes in Vibrio splendidus may reflect adaptation to versatile lifestyles. *FEMS Microbiol Lett,* 294**,** 207-15.
KIM, H. L., SHIN, E. K., KIM, H. M., RYOU, S. M., KIM, S., CHA, C. J., BAE, J. & LEE, K. 2007. Heterogeneous rRNAs are differentially expressed during the morphological development of Streptomyces coelicolor. *FEMS Microbiol Lett,* 275**,** 146-52.
KIRKUP, B. C., JR., CHANG, L., CHANG, S., GEVERS, D. & POLZ, M. F. 2010. Vibrio chromosomes share common history. *BMC Microbiol,* 10**,** 137.
KITAHARA, K., YASUTAKE, Y. & MIYAZAKI, K. 2012. Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in Escherichia coli. *Proc Natl Acad Sci U S A,* 109**,** 19220-5.
KLAPPENBACH, J. A., DUNBAR, J. M. & SCHMIDT, T. M. 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol,* 66**,** 1328-33.
KOCH, A. L. 2001. Oligotrophs versus copiotrophs. *Bioessays,* 23**,** 657-61.
KOLMSEE, T., DELIC, D., AGYENIM, T., CALLES, C. & WAGNER, R. 2011. Differential stringent control of Escherichia coli rRNA promoters: effects of ppGpp, DksA and the initiating nucleotides. *Microbiology,* 157**,** 2871-9.
KONIG, J., WELLS, J., CANI, P. D., GARCIA-RODENAS, C. L., MACDONALD, T., MERCENIER, A., WHYTE, J., TROOST, F. & BRUMMER, R.-J. 2016. Human Intestinal Barrier Function in Health and Disease. *Clin Trans Gastroenterol,* 7**,** e196.
LAURO, F. M., CHASTAIN, R. A., BLANKENSHIP, L. E., YAYANOS, A. A. & BARTLETT, D. H. 2007. The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl Environ Microbiol,* 73**,** 838-45.
LIAO, D. 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J Mol Evol,* 51**,** 305-17.
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.
LÓPEZ-LÓPEZ, A., BENLLOCH, S., BONFA, M., RODRIGUEZ-VALERA, F. & MIRA, A. 2007. Intragenomic 16S rDNA divergence in Haloarcula marismortui is an adaptation to different temperatures. *J Mol Evol,* 65**,** 687-96.
LÓPEZ-PÉREZ, M., GONZAGA, A., MARTIN-CUADRADO, A. B., LOPEZ-GARCIA, P., RODRIGUEZ-VALERA, F. & KIMES, N. E. 2013. Intra- and intergenomic variation of ribosomal RNA operons in concurrent Alteromonas macleodii strains. *Microb Ecol,* 65**,** 720-30.
MACHADO, H. & GRAM, L. 2015. The fur Gene as a New Phylogenetic Marker for Vibrionaceae Species Identification. *Appl Environ Microbiol,* 81**,** 2745-52.
MACKIE, G. A. 2013. RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Micro,* 11**,** 45-57.
MAEDA, M., SHIMADA, T. & ISHIHAMA, A. 2015. Strength and Regulation of Seven rRNA Promoters in Escherichia coli. *PLoS One,* 10.
MAIDAK, B. L., COLE, J. R., LILBURN, T. G., PARKER, C. T., SAXMAN, P. R., FARRIS, R. J., GARRITY, G. M., OLSEN, G. J., SCHMIDT, T. M. & TIEDJE, J. M.

2001. The RDP-II (Ribosomal Database Project). *Nucleic Acids Research,* 29**,** 173-174.
MAKINO, K., OSHIMA, K., KUROKAWA, K., YOKOYAMA, K., UDA, T., TAGOMORI, K., IIJIMA, Y., NAJIMA, M., NAKANO, M., YAMASHITA, A., KUBOTA, Y., KIMURA, S., YASUNAGA, T., HONDA, T., SHINAGAWA, H., HATTORI, M. & IIDA, T. 2003. Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V cholerae. *Lancet,* 361**,** 743-9.
MATHEWS GROUP. 2017. *RNAstructure: Web Servers for RNA Secondary Structure Prediction* [Online]. Available: https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html [Accessed 01.01.2017].
MAYERLE, M. & WOODSON, S. A. 2013. Specific contacts between protein S4 and ribosomal RNA are required at multiple stages of ribosome assembly. *Rna,* 19**,** 574-85.
MAYNARD, C. L., ELSON, C. O., HATTON, R. D. & WEAVER, C. T. 2012. Reciprocal interactions of the intestinal microbiota and immune system. *Nature,* 489**,** 231-41.
MEIBOM, K. L., BLOKESCH, M., DOLGANOV, N. A., WU, C. Y. & SCHOOLNIK, G. K. 2005. Chitin induces natural competence in Vibrio cholerae. *Science,* 310**,** 1824-7.
MICHIGAN STATE UNIVERSITY. 2016. *The Ribosomal Database Project (RDP-II)* [Online]. Available: https://rdp.cme.msu.edu/ [Accessed 31.10.2017].
MIZRAHI-MAN, O., DAVENPORT, E. R. & GILAD, Y. 2013. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One,* 8**,** e53608.
MOLECULAR PROBES INC. 2001. *SYBR® Green II RNA Gel Stain* [Online]. Available: https://tools.thermofisher.com/content/sfs/manuals/mp07568.pdf [Accessed 01.10.17].
MORENO, C., ROMERO, J. & ESPEJO, R. T. 2002. Polymorphism in repeated 16S rRNA genes is a common property of type strains and environmental isolates of the genus Vibrio. *Microbiology,* 148**,** 1233-9.
MUYZER, G. 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology,* 2**,** 317-322.
MUYZER, G., DE WAAL, E. C. & UITTERLINDEN, A. G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology,* 59**,** 695-700.
MYLVAGANAM, S. & DENNIS, P. P. 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium Haloarcula marismortui. *Genetics,* 130**,** 399-410.
NAYAK, S. K. 2010. Role of gastrointestinal microbiota in fish. *Aquaculture Research,* 41**,** 1553-1573.
NEILSON, J. W., JORDAN, F. L. & MAIER, R. M. 2013. Analysis of Artifacts Suggests DGGE Should Not Be Used For Quantitative Diversity Analysis. *Journal of microbiological methods,* 92**,** 256-263.
NÜBEL, U., ENGELEN, B., FELSKE, A., SNAIDR, J., WIESHUBER, A., AMANN, R. I., LUDWIG, W. & BACKHAUS, H. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in Paenibacillus polymyxa detected by temperature gradient gel electrophoresis. *J Bacteriol,* 178**,** 5636-43.
PAREKH, P. J., BALART, L. A. & JOHNSON, D. A. 2015. The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease. *Clin Trans Gastroenterol,* 6**,** e91.

PEI, A. Y., OBERDORF, W. E., NOSSA, C. W., AGARWAL, A., CHOKSHI, P., GERZ, E. A., JIN, Z., LEE, P., YANG, L., POLES, M., BROWN, S. M., SOTERO, S., DESANTIS, T., BRODIE, E., NELSON, K. & PEI, Z. 2010. Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. *Appl Environ Microbiol,* 76**,** 3886-97.

PORETSKY, R., RODRIGUEZ, R. L., LUO, C., TSEMENTZI, D. & KONSTANTINIDIS, K. T. 2014. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS One,* 9.

RAINEY, F. A., WARD-RAINEY, N. L., JANSSEN, P. H., HIPPE, H. & STACKEBRANDT, E. 1996. Clostridium paradoxum DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology,* 142 ( Pt 8)**,** 2087-95.

RAWLS, J. F., MAHOWALD, M. A., GOODMAN, A. L., TRENT, C. M. & GORDON, J. I. 2007. In vivo imaging and genetic analysis link bacterial motility and symbiosis in the zebrafish gut. *Proc Natl Acad Sci U S A,* 104**,** 7622-7.

RAWLS, J. F., SAMUEL, B. S. & GORDON, J. I. 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 4596-4601.

REGENTS OF THE UNIVERSITY OF MICHIGAN. 2017. *rrnDB* [Online]. Available: https://rrndb.umms.med.umich.edu/ [Accessed 22.10.2017].

RONNESETH, A., CASTILLO, D., D'ALVISE, P., TONNESEN, O., HAUGLAND, G., GROTKJAER, T., ENGELL-SORENSEN, K., NORREMARK, L., BERGH, O., WERGELAND, H. I. & GRAM, L. 2017. Comparative assessment of Vibrio virulence in marine fish larvae. *J Fish Dis*.

ROSENLUND, G. & HALLDÓRSSON, Ó. 2007. Cod juvenile production: Research and commercial developments. *Aquaculture,* 268**,** 188-194.

SALVESEN, I., OIE, G. & VADSTEIN, O. 1997. Surface disinfection of Atlantic halibut and turbot eggs with glutaraldehyde: evaluation of concentrations and contact times. *Aquaculture International,* 5**,** 249-258.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SAPAG, A., VARTIKAR, J. V. & DRAPER, D. E. 1990. Dissection of the 16S rRNA binding site for ribosomal protein S4. *Biochim Biophys Acta,* 1050**,** 34-7.

SEKIROV, I., RUSSELL, S. L., ANTUNES, L. C. & FINLAY, B. B. 2010. Gut microbiota in health and disease. *Physiol Rev,* 90**,** 859-904.

SHAJANI, Z., SYKES, M. T. & WILLIAMSON, J. R. 2011. Assembly of bacterial ribosomes. *Annu Rev Biochem,* 80**,** 501-26.

SKJERMO, J., SALVESEN, I., ØIE, G., OLSEN, Y. & VADSTEIN, O. 1997. Microbially matured water: a technique for selection of a non-opportunistic bacterial flora in water that may improve performance of marine larvae. *Aquaculture International,* 5**,** 13-28.

SMYTH, R. P., SCHLUB, T. E., GRIMM, A., VENTURI, V., CHOPRA, A., MALLAL, S., DAVENPORT, M. P. & MAK, J. 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene,* 469**,** 45-51.

SNUSTAD, D. P. & SIMMONS, M. J. 2012. *Genetics,* Singapore, John Wiley & Sons Singapore Pte Ltd.

STACKEBRANDT, E. & GOEBEL, B. M. 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology,* 44**,** 846-849.

STALEY, J. T. & KONOPKA, A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol,* 39**,** 321-46.
STODDARD, S. F., SMITH, B. J., HEIN, R., ROLLER, B. R. & SCHMIDT, T. M. 2015. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res,* 43**,** D593-8.
SUL, W. J., COLE, J. R., JESUS EDA, C., WANG, Q., FARRIS, R. J., FISH, J. A. & TIEDJE, J. M. 2011. Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. *Proc Natl Acad Sci U S A,* 108**,** 14637-42.
SULLAM, K. E., ESSINGER, S. D., LOZUPONE, C. A., O'CONNOR, M. P., ROSEN, G. L., KNIGHT, R., KILHAM, S. S. & RUSSELL, J. A. 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol Ecol,* 21**,** 3363-78.
SUN, D. L., JIANG, X., WU, Q. L. & ZHOU, N. Y. 2013. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microbiol,* 79**,** 5962-9.
THE NORWEGIAN HIGH THROUGHPUT SEQUENCING CENTRE. 2016. *GUIDELINES for correct completion of SAMPLE SUBMISSION FORM ILLUMINA SEQUENCING* [Online]. Available: http://www.sequencing.uio.no/forms/guidelines-submission-form-(illumina).pdf [Accessed 26.10.17].
THOMPSON, F. L. & KLOSE, K. E. 2006. Vibrio2005: the First International Conference on the Biology of Vibrios. *J Bacteriol,* 188**,** 4592-6.
THOMPSON, F. L., THOMPSON, C. C., VICENTE, A. C. & KLOSE, K. E. 2010. Vibrio2009: the Third International Conference on the Biology of Vibrios. *Mol Microbiol,* 77**,** 1065-71.
TIAN, R. M., CAI, L., ZHANG, W. P., CAO, H. L. & QIAN, P. Y. 2015. Rare Events of Intragenus and Intraspecies Horizontal Transfer of the 16S rRNA Gene. *Genome Biol Evol,* 7**,** 2310-20.
TRUCKSIS, M., MICHALSKI, J., DENG, Y. K. & KAPER, J. B. 1998. The Vibrio cholerae genome contains two unique circular chromosomes. *Proc Natl Acad Sci U S A,* 95**,** 14464-9.
VADSTEIN, O., BERGH, Ø., GATESOUPE, F.-J., GALINDO-VILLEGAS, J., MULERO, V., PICCHIETTI, S., SCAPIGLIATI, G., MAKRIDIS, P., OLSEN, Y., DIERCKENS, K., DEFOIRDT, T., BOON, N., DE SCHRYVER, P. & BOSSIER, P. 2013. Microbiology and immunology of fish larvae. *Reviews in Aquaculture,* 5**,** S1-S25.
VADSTEIN, O., ØIE, G., OLSEN, Y., SALVESEN, I., SKJERMO, J. & SKJÅK-BRÆK, G. 1993. *A strategy to obtain microbial control during larval development of marine fish*.
VAL, M. E., SOLER-BISTUE, A., BLAND, M. J. & MAZEL, D. 2014. Management of multipartite genomes: the Vibrio cholerae model. *Curr Opin Microbiol,* 22**,** 120-6.
VARANI, G. & MCCLAIN, W. H. 2000. The G·U wobble base pair: A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep,* 1**,** 18-23.
VERSCHUERE, L., DHONT, J., SORGELOOS, P. & VERSTRAETE, W. 1997. Monitoring Biolog patterns and r/K-strategists in the intensive culture of Artemia juveniles. *Journal of Applied Microbiology,* 83**,** 603-612.
VESTRUM, R. I. 2009. *Funksjonalitet i bakterie/torskelarve interaksjoner.* Master's thesis, Norges teknisk-naturvitenskapelige universitet (NTNU).

VETROVSKY, T. & BALDRIAN, P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One,* 8**,** e57923.

VEZZI, A., CAMPANARO, S., D'ANGELO, M., SIMONATO, F., VITULO, N., LAURO, F. M., CESTARO, A., MALACRIDA, G., SIMIONATI, B., CANNATA, N., ROMUALDI, C., BARTLETT, D. H. & VALLE, G. 2005. Life at depth: Photobacterium profundum genome sequence and expression analysis. *Science,* 307**,** 1459-61.

WANG, Y. & ZHANG, Z. 2000. Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. *Microbiology,* 146 ( Pt 11)**,** 2845-54.

WIMBERLY, B. T., BRODERSEN, D. E., CLEMONS, W. M., JR., MORGAN-WARREN, R. J., CARTER, A. P., VONRHEIN, C., HARTSCH, T. & RAMAKRISHNAN, V. 2000. Structure of the 30S ribosomal subunit. *Nature,* 407**,** 327-39.

WOESE, C. R. 1987. Bacterial evolution. *Microbiol Rev,* 51**,** 221-71.

WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 5088-5090.

WOESE, C. R., KANDLER, O. & WHEELIS, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A,* 87**,** 4576-9.

WU, L., WEN, C., QIN, Y., YIN, H., TU, Q., VAN NOSTRAND, J. D., YUAN, T., YUAN, M., DENG, Y. & ZHOU, J. 2015. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol,* 15**,** 125.

YANG, Q., PANDE, G. S. J., WANG, Z., LIN, B., RUBIN, R. A., VORA, G. J. & DEFOIRDT, T. 2017. Indole signalling and (micro)algal auxins decrease the virulence of Vibrio campbellii, a major pathogen of aquatic organisms. *Environ Microbiol,* 19**,** 1987-2004.

YARZA, P., YILMAZ, P., PRUESSE, E., GLOCKNER, F. O., LUDWIG, W., SCHLEIFER, K. H., WHITMAN, W. B., EUZEBY, J., AMANN, R. & ROSSELLO-MORA, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol,* 12**,** 635-45.

# List of appendices

# Appendix A Preparation of cultivation media, FASW, buffers and DGGE solutions

Two media were primarily used for cultivation of bacteria. These were the growth media marine broth (MB) (Table A.1) and M65 (Table A.2).

**Table A.1:** Recipe for 1 L Marine Broth (MB) liquid medium and Marine Agar (MA). Ingredients were mixed together, boiled for one minute and autoclaved at 120°C, 20 minutes.

| **Content** | **Amount** |
|---|---|
| Marine broth Difco™, 2216 | 37.4 g |
| Distilled water ($dH_2O$) | 1 L |
| Agar noble* | 15 g/L |

*only included in Marine Agar recipe

**Table A.2:** Recipe for 1 L of M65 liquid and agar media. Ingredients were mixed together and autoclaved at 120°C, 20 minutes.

| **Content** | **Amount** |
|---|---|
| Peptone | 0.5 g |
| Tryptone | 0.5 g |
| Yeast extract | 0.5 g |
| Distilled water ($dH_2O$) | 200 mL |
| FASW | 800 mL |
| Agar noble* | 15 g/L |

*only included in M65 agar recipe

Filtered autoclaved seawater (FASW) was made by vacuum filtrating seawater through two filters – first a glass microfiber filter with a 1.0 µm particle retention (VWR), then an Aqueous Solution Filter Capsule with a 0.2 µm pore size (Polycap 36 AS). The filtered seawater was autoclaved at 120°C, 20 minutes.

The two main buffers used in this project were TE buffer (Table A.3) and TAE buffer (Table A.4). The 50 x TAE buffer was diluted to 1 x TAE or 0.5 x TAE buffer by adding distilled water ($dH_2O$) to the correct dilution was reached. The same was done to make 1:10 TE buffer.

**Table A.3:** Recipe for 500 mL of 1 x Tris-EDTA (TE) buffer. Buffer was autoclaved.

| **Content** | **Amount** |
|---|---|
| 2M Tris-HCL pH 7.5 | 2.5 mL |
| 0.5M EDTA pH 8.0 | 1.0 mL |
| Distilled water ($dH_2O$) | 496.5 mL |

**Table A.4:** Recipe for 500 mL of 50 x TAE buffer. Buffer was autoclaved.

| **Content** | **Amount** |
|---|---|
| Tris-base | 242 g |
| 0.5M EDTA pH 8.0 | 100 mL |
| Distilled water ($dH_2O$) | To 1 L total |

For DGGE analysis, 0% and 80% denatured solutions were made (Table A.5 and Table A.6). The deionized formamide used in the 80% denaturing solution was made by adding 7.5 g DOWEX RESIN AG 501X8 to 200 mL formamide and stirring for 1 hour at room temperature.

**Table A.5:** Recipe for 250 mL 0% denaturing solution. Final solution consists of 8% acrylamide in 0.5 x TAE buffer. Solution was stored at 4°C, shielded from light.

| **Content** | **Amount** |
|---|---|
| 40% acrylamide solution (BioRadLab Inc.) | 50 mL |
| 50 x TAE | 2.5 mL |
| Distilled water ($dH_2O$) | To 250 mL total |

**Table A.6:** Recipe for 250 mL 80% denaturing solution. Final solution consists of 8% acrylamide, 5.6M urea and 32% formamide in 0.5 x TAE buffer. Solution was stored at 4°C, shielded from light.

| **Content** | **Amount** |
|---|---|
| 40% acrylamide solution (BioRadLab Inc.) | 50 mL |
| 50 x TAE | 2.5 mL |
| Urea | 84 g |
| Deionized formamide | 80 mL |
| Distilled water ($dH_2O$) | To 250 mL total |

# Appendix B Recorded temperatures gnotobiotic fish experiment (2016)

Table B.1 shows recorded temperatures of the room where the gnotobiotic fish experiment was conducted. Table B.1 shows the days prior to the Atlantic cod eggs hatching, while Table B.2 shows the recorded temperatures from the day of hatching and the days after.

**Table B.1:** The temperatures recorded during the 2016 gnotobiotic cod larvae experiment before the eggs hatched. All temperatures are given in degrees Celsius (°C). The abbreviations "avg" and "temp" stand for "average" and "temperature", respectively. Min and max temperatures represent minimum and maximum temperatures, respectively.

| | **Days after arrival of eggs** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| Avg temp water | N/A | 4.1 | 5.4 | 5.9 | 5.9 | 5.9 | 5.9 | 5.9 |
| Avg temp air | 6.4 | 5.1 | 5.9 | 6.2 | 6.1 | 6.2 | 6.4 | 6.9 |
| Temp air max | N/A | 6.8 | 7.3 | 7.5 | 7.3 | 7.3 | 7 | 7.3 |
| Temp air min | N/A | 2 | 3.3 | 4.2 | 4.3 | 4.3 | 4.7 | 4.4 |
| Temp water max | N/A | 21 | 5.6 | 6.1 | 6.1 | 6.1 | 6.1 | 6.2 |
| Temp water min | N/A | 3.6 | 4 | 5.4 | 5.7 | 5.7 | 5.7 | 7.2 |

**Table B.2:** The temperatures recorded during the 2016 gnotobiotic cod larvae experiment after hatching of the eggs. All temperatures are given in degrees Celsius (°C). The abbreviations "avg" and "temp" stand for "average" and "temperature", respectively. Min and max temperatures represent minimum and maximum temperatures, respectively.

| | **Days post-hatch** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** |
| Avg temp water | N/A | 6.9 | 7.6 | 8.8 | 9.5 | 10.6 | 10.4 |
| Avg temp air | N/A | 7.7 | 8.4 | 9.9 | 10.5 | 11.4 | 11.5 |
| Temp air max | 11.3 | 8.9 | 10 | 10.3 | 10.9 | 11.8 | 12 |
| Temp air min | 6 | 5.8 | 6.4 | 8.2 | 9.2 | 10.5 | 10.2 |
| Temp water max | 11.3 | 7.2 | 8 | 8.8 | 9.7 | 10.8 | 10.6 |
| Temp water min | 6 | 6.7 | 6.9 | 7.6 | 8.8 | 9.5 | 10.1 |

## Appendix C PowerSoil® DNA isolation protocol

The protocol for the PowerSoil® DNA Isolation Kit by Mo Bio Laboratories Inc.


**EXPERIENCED USER PROTOCOL**
**PowerSoil® DNA Isolation Kit**
Catalog No. 12888-50 & 12888-100

**Please wear gloves at all times**

1. To the **PowerBead Tubes** provided, add 0.25 grams of soil sample.
2. Gently vortex to mix.
3. **Check Solution C1**. If **Solution C1** is precipitated, heat solution to 60°C until dissolved before use.
4. Add 60 µl of **Solution C1** and invert several times or vortex briefly.
5. Secure **PowerBead Tubes** horizontally using the MO BIO Vortex Adapter tube holder for the vortex (MO BIO Catalog# 13000-V1-24) or secure tubes horizontally on a flat-bed vortex pad with tape. Vortex at maximum speed for 10 minutes.

**Note**

If you are using the 24 place Vortex Adapter for more than 12 preps, increase the vortex time by 5-10 minutes.

6. Make sure the **PowerBead Tubes** rotate freely in your centrifuge without rubbing. Centrifuge tubes at 10,000 x g for 30 seconds at room temperature. **CAUTION**: Be sure not to exceed 10,000 x g or tubes may break.
7. Transfer the supernatant to a clean **2 ml Collection Tube** (provided).

**Note**

Expect between 400 to 500 µl of supernatant. Supernatant may still contain some soil particles.

8. Add 250 µl of **Solution C2** and vortex for 5 seconds. Incubate at 4°C for 5 minutes.
9. Centrifuge the tubes at room temperature for 1 minute at 10,000 x g.
10. Avoiding the pellet, transfer up to, but no more than, 600 µl of supernatant to a clean **2 ml Collection Tube** (provided).
11. Add 200 µl of **Solution C3** and vortex briefly. Incubate at 4°C for 5 minutes.
12. Centrifuge the tubes at room temperature for 1 minute at 10,000 x g.
13. Avoiding the pellet, transfer up to, but no more than, 750 µl of supernatant into a clean **2 ml Collection Tube** (provided).
14. Shake to mix **Solution C4** before use. Add 1200 µl of **Solution C4** to the supernatant and vortex for 5 seconds.

15. Load approximately 675 µl onto a **Spin Filter** and centrifuge at 10,000 x g for 1 minute at room temperature. Discard the flow through and add an additional 675 µl of supernatant to the **Spin Filter** and centrifuge at 10,000 x g for 1 minute at room temperature. Load the remaining supernatant onto the **Spin Filter** and centrifuge at 10,000 x g for 1 minute at room temperature.

**Note**

A total of three loads for each sample processed are required.

16. Add 500 µl of **Solution C5** and centrifuge at room temperature for 30 seconds at 10,000 x g.

17. Discard the flow through.

18. Centrifuge again at room temperature for 1 minute at 10,000 x g.

19. Carefully place spin filter in a clean **2 ml Collection Tube** (provided). Avoid splashing any **Solution C5** onto the **Spin Filter**.

20. Add 100 µl of Solution C6 to the center of the white filter membrane. Alternatively, sterile DNA-Free PCR Grade Water may be used for elution from the silica Spin Filter membrane at this step (MO BIO Catalog# 17000-10).

21. Centrifuge at room temperature for 30 seconds at 10,000 x g.

22. Discard the **Spin Filter**. The DNA in the tube is now ready for any downstream application. No further steps are required.

We recommend storing DNA frozen (-20° to -80°C). **Solution C6** contains no EDTA. To concentrate the DNA see the Hints & Troubleshooting Guide.

# Appendix D PowerMicrobiome™ RNA isolation protocol

The protocol for the PowerMicrobiome™ RNA Isolation Kit from Mo Bio Laboratories Inc.


**Important Notes Before Starting**

Solution PM1 must be warmed at 55-60°C for 10 minutes to dissolve precipitates prior to use. Solution PM1 should be used while still warm. Shake to mix Solution PM5 before use.

**Prepare Solution PM1/βME by adding β- mercaptoethanol (βME) to Solution PM1**

Add enough β- mercaptoethanol (βME) to Solution PM1 to produce a final concentration of 10 µl/ml. For best results do not use the PM1/βME mixture if greater than a month old. Since the PM1/βME mixture loses its effectiveness over time, it's best to make it fresh each time. For each prep, you will need 650 µl of PM1/βME solution. Alternatively, you may add 650 µl of solution PM1 and 6.5 µl βME to each Glass Bead Tube, 0.1 mm.

**Note:** Use a fume hood when opening βME to avoid exposure to the chemical.

**DNase I Stock Enzyme and DNase I Solution Preparation and Storage**

A. Prepare **DNase I stock enzyme** by adding **300 µl** of RNase-Free Water (**Solution PM8**) to the DNase I (RNase-Free) lyophilized powder and mix gently. Aliquot the DNase I stock enzyme in 50 µl portions and store at -20°C for long term storage.

   **Note:** The DNase I stock enzyme can be freeze/thawed up to three times without loss of activity.

B. Shortly before proceeding with the protocol, prepare the **DNase I Solution**, by thawing enough DNase I stock enzyme for the number of preps. Per prep, combine **5 µl of DNase I stock enzyme** with **45 µl of Solution PM6**.

Example:

| Number of preps | DNase I stock enzyme | Solution PM6 |
|---|---|---|
| 1 | 5 µl | 45 µl |
| 2 | 10 µl | 90 µl |
| 10 | 50 µl | 450 µl |

## Experienced User Protocol

**Please wear gloves at all times**

**Warm Solution PM1 prior to use at 55°C for 10 minutes. Use Solution PM1 while still warm.**

1. Place 0.25 grams of stool or biosolid sample into the Glass Bead Tube, 0.1 mm provided.
   **Note:** If a phenol-based lysis is desired, add 100 µl of phenol:chloroform:isoamy alcohol pH 6.5-8.0 to the Bead Tube before adding the sample.

2. Add **650 µl of Solution PM1/βME** (see **Important Notes before Starting** section) to the Glass Bead Tube, 0.1 mm. Alternatively, you may add separately 650 µl PM1 and 6.5 µl of βME to the Glass Bead Tube.

3. Secure the Glass Bead Tube, 0.1mm horizontally to a MO BIO Vortex Adapter, catalog number 13000-V1 or 13000-V1-24. The tube caps should be oriented pointing toward the center of the Vortex Adapter.

4. Vortex at maximum speed for 10 minutes.

5. Centrifuge at 13,000 x *g* for 1 minute at room temperature. Transfer the supernatant to a clean 2 ml Collection Tube (provided). The expected volume is ~500-600 µl. If you added phenol:chloroform: isoamyl alcohol, remove the upper aqueous layer and transfer to a clean 2 ml Collection Tube.

6. Add **150 µl of Solution PM2** and vortex briefly to mix. Incubate at 4°C for 5 minutes.

7. Centrifuge the tubes at 13,000 x *g* for 1 minute.

8. Avoiding the pellet, transfer the supernatant to a clean 2 ml Collection Tube (provided). Do not transfer more than 650 µl at this step.

9. Add **650 µl of Solution PM3** and **650 µl of Solution PM4.** Vortex briefly to mix.
   **Note**: To prevent small RNA species (5S, tRNA and degraded RNA) from co-purifying with the mRNA and rRNA, use 650 µl of 70% ethanol instead of Solution PM4.

   For the purification of small RNAs, such as microRNA and siRNA, transfer the lysate to a larger tube to accommodate a higher volume (2.6 ml) and add an additional 650 µl of 100% ethanol to the lysate. You will need to supply 100% ethanol for this step.

10. Load 650 µl of supernatant onto a Spin Filter and centrifuge at 13,000 x *g* for 1 minute. Discard the flow through and repeat until all the supernatant has been loaded onto the Spin Filter.
    **Note:** A total of three loads for each sample processed is required and four loads if an additional volume of 100% ethanol is added for the microRNA and siRNA protocol.

11. Shake to mix Solution PM5. Add **650 µl of Solution PM5** to the Spin Filter and centrifuge at 13,000 x *g* for 1 minute.
    **Note:** Skip steps **12-16** if you want to isolate both RNA and DNA.

12. Discard the flow through and centrifuge again at 13,000 x *g* for 1 minute to remove residual wash.

13. Place the Spin Filter basket into a clean 2 ml Collection Tube (provided).

14. To the center of the Spin Filter, add **50 µl of DNase I Solution** (prepared by mixing **45 µl of Solution PM6** and **5 µl DNase I stock solution.** See **Important Notes Before Starting** section).

15. Incubate at room temperature for 15 minutes.

16. Add **400 µl Solution PM7** and centrifuge the column at 13,000 x *g* for 1 minute.

17. Discard the flow through and add **650 µl of Solution PM5** and centrifuge at 13,000 x *g* for 1 minute.

18. Discard the flow through and add **650 µl of Solution PM4** and centrifuge at 13,000 x *g* for 1 minute.

19. Discard the flow through and centrifuge again at 13,000 x *g* for 2 minutes to remove any residual wash solution.

20. Place the Spin Filter basket into a clean 2 ml Collection Tube (provided).

21. Add **100 µl of Solution PM8 (RNase-Free Water)** to the center of the white filter membrane. Allow the water to sit on the membrane for at least 1 minute.
    **Note:** Eluting with 100 µl of Solution PM8 will maximize RNA yield. For more concentrated RNA, a minimum of 50 µl of Solution PM8 can be used. Do not use less than 50 µl of Solution PM8.

22. Centrifuge at 13,000 x *g* for 1 minute.

23. Discard the Spin Filter basket.

The RNA is now ready for any downstream applications. The RNA in the tube can be stored at -80°C until use.

## Appendix E PrimeScript™ 1st strand cDNA Synthesis protocol

The protocol for the PrimeScript™ 1st strand cDNA Synthesis Kit from TaKaRa Bio Inc. Random 6 mers were used and not an oligo dT primer.

**PrimeScript™ 1st strand cDNA Synthesis Kit** Cat. #6110A v201606Da TaKaRa

### V. cDNA Synthesis Protocol

1. Prepare the following mixture in a microtube.

| Reagent | Volume |
|---|---|
| Oligo dT Primer (50 μM) | 1 μl |
| or Random 6 mers (50 μM) | or 1 μl (0.4 - 2 μl)*1 |
| dNTP Mixture (10 mM each) | 1 μl |
| Template RNA | total RNA : < 5 μg*2<br>polyA+ RNA : < 1 μg |
| RNase free $dH_2O$ | x μl |
| Total | 10 μl |

2. Incubate for 5 min at 65℃, then cool immediately on ice.*3

3. Prepare the reaction mixture in a total volume of 20 μl.

| Reagent | Volume |
|---|---|
| Template RNA Primer Mixture (from step 2) | 10 μl |
| 5X PrimeScript Buffer | 4 μl |
| RNase Inhibitor (40 U/μl) | 0.5 μl (20 units) |
| PrimeScript RTase (200 U/μl) | 1.0 μl (200 units) |
| RNase free $dH_2O$ | x μl |
| Total | 20 μl |

4. Mix gently.

5. Incubate the reaction mixture using the following conditions.

| | |
|---|---|
| 30℃ | 10 min (required when using Random 6 mers) |
| 42℃ (50℃)*4 | 30 - 60 min |

6. Inactivate the enzyme by incubating at 95℃ for 5 min*5, then cool on ice.

*1 : For Random 6 mers, use 0.4 μl (20 pmol) for synthesis of cDNA products over 2 kb long, and use 2 μl (100 pmol) for reverse transcription prior to real-time PCR. If a gene specific primer is used, use a final primer concentration of 0.1 μM.

*2 : For 1st strand cDNA synthesis of templates for real-time RT-PCR, use less than1 μg of total RNA.

*3 : This denaturing step of template RNA is important to improve the reverse transcription efficiency.

*4 : In general, reactions should be performed at 42℃. However, for RT-PCR reactions where the reverse primer for PCR is also used as a reverse transcription primer, we recommend performing the reverse transcription reaction at 50℃ to reduce the possibility of non-specific amplification products.

*5 : For synthesis of longer cDNAs, inactivation at 70℃ for 15 min is recommended to minimize cDNA damage (i.e., nicking).

# Appendix F  QIAquick ® PCR purification protocol

The protocol for the QIAquick ® PCR Purification Kit by Qiagen. One modification was made to the protocol, as 22µL filtered Milli-Q® water was used for the elution in step 7 instead of 50µL EB buffer.

**Quick-Start**Protocol

## QIAquick® PCR Purification Kit

The QIAquick PCR Purification Kit (cat. nos. 28104 and 28106) can be stored at room temperature (15–25°C) for up to 12 months.

For more information, please refer to the *QIAquick Spin Handbook, March 2008*, which can be found at: www.qiagen.com/handbooks.

For technical assistance, please call toll-free 00800-22-44-6000, or find regional phone numbers at www.qiagen.com/contact.

**Notes before starting**

- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional table-top microcentrifuge at room temperature.
- Add 1:250 volume pH indicator I to Buffer PB. The yellow color of Buffer PB with pH indicator I indicates a pH of ≤7.5. If the purified PCR product is to be used in sensitive microarray applications, it may be beneficial to use Buffer PB without the addition of pH indicator I. Do not add pH indicator I to buffer aliquots.

1. Add 5 volumes Buffer PB to 1 volume of the PCR reaction and mix. If the color of the mixture is orange or violet, add 10 µl 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn yellow.
2. Place a QIAquick column in ▲ a provided 2 ml collection tube or into ● a vacuum manifold. For details on how to set up a vacuum manifold, refer to the *QIAquick Spin Handbook*.
3. To bind DNA, apply the sample to the QIAquick column and ▲ centrifuge for 30–60 s or ● apply vacuum to the manifold until all the samples have passed through the column. ▲ Discard flow-through and place the QIAquick column back in the same tube.
4. To wash, add 0.75 ml Buffer PE to the QIAquick column ▲ centrifuge for 30–60 s or ● apply vacuum. ▲ Discard flow-through and place the QIAquick column back in the same tube.
5. Centrifuge the QIAquick column once more in the provided 2 ml collection tube for 1 min to remove residual wash buffer.
6. Place each QIAquick column in a clean 1.5 ml microcentrifuge tube.
7. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) to the center of the QIAquick membrane and centrifuge the column for 1 min. For increased DNA concentration, add 30 µl elution buffer to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge.
8. If the purified DNA is to be analyzed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel.



QIAGEN®

Sample & Assay Technologies

# Appendix G List of Illumina amplicon sequencing samples

**Table G.1:** Sample name and type of all 76 samples sent in for Illumina amplicon sequencing.

| Sample Nr. | Name | Sample type |
|---|---|---|
| 1 | HI610A-V-DNA | DNA water gnotobiotic cod larvae experiment (2016) |
| 2 | HI610B-V-DNA | |
| 3 | HI610C-V-DNA | |
| 4 | HI610D-V-DNA | |
| 5 | RD5A-V-DNA | |
| 6 | RD5B-V-DNA | |
| 7 | RD5C-V-DNA | |
| 8 | RD5D-V-DNA | |
| 9 | BB120A-V-DNA | |
| 10 | BB120B-V-DNA | |
| 11 | BB120C-V-DNA | |
| 12 | BB120D-V-DNA | |
| 13 | HI610MIN-DNA* | DNA water without fish gnoto. fish experiment (2016) |
| 14 | RD5MIN-DNA* | |
| 15 | BB120MIN-DNA* | |
| 16 | HI610A-F-DNA | DNA fish gnotobiotic cod larvae experiment (2016) |
| 17 | HI610B-F-DNA | |
| 18 | HI610C-F-DNA | |
| 19 | HI610D-F-DNA | |
| 20 | RD5A-F-DNA | |
| 21 | RD5B-F-DNA | |
| 22 | RD5C-F-DNA | |
| 23 | RD5D-F-DNA | |
| 24 | RD5-30-REN* | Pure culture |
| 25 | D13-G2-V-DNA | DNA water old gnotobiotic cod larvae experiment day 13 (2014) |
| 26 | D13-G9-V-DNA | |
| 27 | D13-G11V-DNA | |
| 28 | D13-G2-F-DNA | RNA water old gnotobiotic cod larvae experiment day 13 (2014) |
| 29 | D13-G9-F-DNA | |
| 30 | D13-G11F-DNA | |
| 31 | RD5A-V-RNA | RNA water gnotobiotic cod larvae experiment (2016) |
| 32 | RD5B-V-RNA | |
| 33 | RD5C-V-RNA | |
| 34 | RD5D-V-RNA | |
| 35 | BB120A-V-RNA | |
| 36 | BB120B-V-RNA | |
| 37 | BB120C-V-RNA | |
| 38 | BB120D-V-RNA | |
| 39 | RD5MIN-RNA* | RNA water without fish gnoto. fish experiment (2016) |
| 40 | BB120MIN-RNA* | |

| | | |
|---|---|---|
| 41 | RD5A-F-RNA | RNA fish gnotobiotic cod larvae experiment (2016) |
| 42 | RD5B-F-RNA | |
| 43 | RD5C-F-RNA | |
| 44 | RD5D-F-RNA | |

| | | |
|---|---|---|
| 45 | RD5P1EXPDNA | DNA growth experiment, exponential phase |
| 46 | RD5P2EXPDNA | |
| 47 | RD5P3EXPDNA | |
| 48 | RD5P4EXPDNA | |
| 49 | RD5P1STADNA | DNA growth experiment, stationary phase |
| 50 | RD5P2STADNA | |
| 51 | RD5P3STADNA | |
| 52 | RD5P4STADNA | |
| 53 | BB120P1EXDNA | DNA growth experiment, exponential phase |
| 54 | BB120P2EXDNA | |
| 55 | BB120P3EXDNA | |
| 56 | BB120P4EXDNA | |
| 57 | BB120P1STDNA | DNA growth experiment, stationary phase |
| 58 | BB120P2STDNA | |
| 59 | BB120P3STDNA | |
| 60 | BB120P4STDNA | |
| 61 | RD5P1EXPRNA | RNA growth experiment, exponential phase |
| 62 | RD5P2EXPRNA | |
| 63 | RD5P3EXPRNA | |
| 64 | RD5P4EXPRNA | |
| 65 | RD5P1STARNA | RNA growth experiment, stationary phase |
| 66 | RD5P2STARNA | |
| 67 | RD5P3STARNA | |
| 68 | RD5P4STARNA | |
| 69 | BB120P1EXRNA | RNA growth experiment, exponential phase |
| 70 | BB120P2EXRNA | |
| 71 | BB120P3EXRNA | |
| 72 | BB120P4EXRNA | |
| 73 | BB120P1STRNA | RNA growth experiment, stationary phase |
| 74 | BB120P2STRNA | |
| 75 | BB120P3STRNA | |
| 76 | BB120P4STRNA | |

* Not included in the PCoA, PERMANOVA or SIMPER analyses

# Appendix H  Normalization Plate (96) Kit protocol

The protocol for the SequalPrep™ Normalization Plate (96) Kit by Invitrogen™.

invitrogen™

## SequalPrep™ Normalization Plate (96) Kit

**Catalog no: A10510-01** **Store at room temperature (15–30°C)**

### Materials Needed

- PCR reactions containing amplicons of the desired length (see **Generating PCR Amplicon**, above)
- DNase-free, aerosol barrier pipette tips
- *Optional:* automated liquid handling workstation capable of handling and manipulating 96-well plates
- *Optional:* PureLink™ Foil Tape (see previous page)

### Binding Step

1. Transfer the desired volume of PCR product (5–25 µl PCR reaction mix, at least 250 ng amplicon/well) from the PCR plate into the wells of the SequalPrep™ Normalization plate.
2. Add an equivalent volume of SequalPrep™ Normalization Binding Buffer.
   **For example:** To purify 10 µl of PCR product, add 10 µl SequalPrep™ Normalization Binding Buffer.
3. Mix completely by pipetting up and down, or seal the plate with PureLink™ Foil Tape (page 1), vortex to mix, and briefly centrifuge the plate.
4. Incubate the plate for 1 hour at room temperature to allow binding of DNA to the plate surface. Mixing is not necessary at this stage.
   **Note:** Incubations longer than 60 minutes do not improve results. However, depending on your workflow you may perform overnight incubation at room temperature for the binding step.
5. **Optional:** If >25 ng DNA/well yield is desired, transfer the amplicon/Binding Buffer mixture from Step 4 to another, fresh well/plate to sequentially bind more DNA. Perform DNA binding at room temperature for 1 hour.
   **Note:** After binding is complete, you can remove the amplicon/Binding Buffer mixture from the well and store at –20°C for up to 30 days to perform additional purifications at a later time.
6. Proceed to **Washing Step**, next page.

### Washing Step

1. Aspirate the liquid from wells. Be sure not to scrape the well sides during aspiration.
   **Note:** If you wish to store the amplicon/Binding Buffer mixture for additional purifications at a later time, aspirate the liquid from wells into another plate and store at –20°C for up to 30 days.
2. Add 50 µl SequalPrep™ Normalization Wash Buffer to the wells. Mix by pipetting up and down twice to improve removal of contaminants.
3. Completely aspirate the buffer from wells and discard.
   To ensure complete removal of wash buffer and maximize elution efficiency, you may need to invert and tap the plate on paper towels depending on the pipetting technique or instrument used. A small amount of residual Wash Buffer (1–3 µl) is typical and does not affect the subsequent elution or downstream applications.
4. Proceed to **Elution Step**, below.

### Elution Step

Review **Elution Options** (previous page).

1. Add 20 µl SequalPrep™ Normalization Elution Buffer to each well of the plate.
   **Note: Do not** use water for elution. If you need to elute in any other buffer, be sure to use a buffer of pH 8.5–9.0. If the pH of the buffer is <8.5, the DNA will not elute efficiently.
2. Mix by pipetting up and down 5 times or seal the plate with PureLink™ Foil Tape (page 1), vortex to mix, and briefly centrifuge the plate. Ensure that the buffer contacts the entire plate coating (up to 20 µl level).
3. Incubate at room temperature for 5 minutes.
4. Transfer and pool the purified DNA as desired or store the eluted DNA at 4°C (short-term storage) or –20°C (long-term storage) until further use.

### Expected Yield and Concentration

The expected DNA concentration is 1–2 ng/µl when using 20 µl elution volume. The expected DNA yield is ~25 ng/well normalized.

## Appendix I  Amicon® Ultra-0.5 Centrifugal Filter Devices protocol

The protocol for the Amicon® Ultra-0.5 Centrifugal Filter Devices by Merck Millipore Ltd. In step 4, the device was spun for 10 minutes. After of step 4, 500 µL of TE-buffer was added and the device spun at 14 000 x g for 10 minutes and eluate was discarded. This process was repeated once more before continuing on to step 5.

### How to Use Amicon® Ultra-0.5 Centrifugal Filter Devices

1. Insert the Amicon® Ultra-0.5 device into one of the provided microcentrifuge tubes.
2. Add up to 500 µL of sample to the Amicon® Ultra filter device and cap it.
3. Place capped filter device into the centrifuge rotor, aligning the cap strap toward the center of the rotor; counterbalance with a similar device.
4. Spin the device at 14,000 × g for approximately 10–30 minutes depending on the NMWL of the device used. Refer to Figure 1 and Table 2 for typical spin times.
5. Remove the assembled device from the centrifuge and separate the Amicon® Ultra filter device from the microcentrifuge tube.
6. To recover the concentrated solute, place the Amicon® Ultra filter device upside down in a clean microcentrifuge tube. Place in centrifuge, aligning open cap towards the center of the rotor; counterbalance with a similar device. Spin for 2 minutes at 1,000 × g to transfer the concentrated sample from the device to the tube. The ultrafiltrate can be stored in the centrifuge tube.

   NOTE: For optimal recovery, perform the reverse spin immediately.

## Appendix J  Flow cytometry plots

Figures representing trends seen in flow cytometry of 2016 cod larvae experiment water samples can be seen in Figure J.1 and Figure J.2.

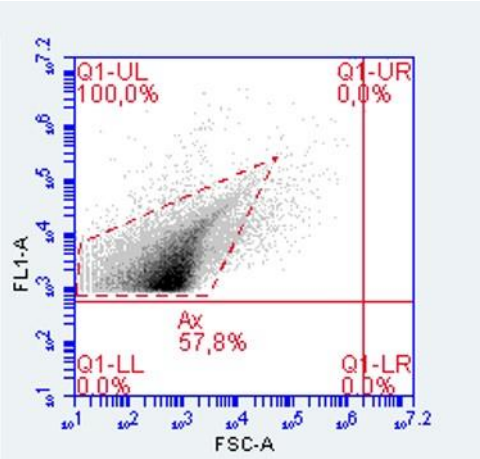


**Figure J.1:** Flow cytometry plot of a bacteria free sample from an axenic flask.

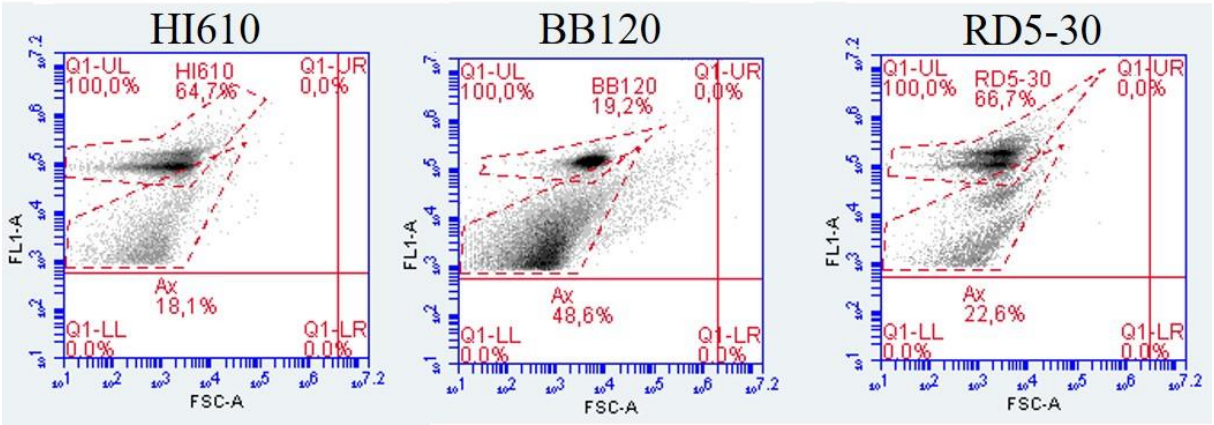


**Figure J.2:** Flow cytometry plots of samples from flasks containing HI610, BB120 and RD5-30.

# Appendix K  Growth experiment OD-measurements for RD5-30 and BB120

In the growth experiment, the overnight culture of RD5-30 had its optical density (OD) measured to 0.516 at 660nm. After 2 hours and 15 minutes the flasks were sampled for the exponential phase samples, but these samples did not form any pellets when spun. New samples were taken from all four flasks, and the OD of these samples measured about 4 hours and 5 minutes after the start of the experiment (Table K.1).

**Table K.1:** OD measurements taken after 4 hours and 5 minutes at 660 nm for each of the four parallel RD5-30 flasks for the exponential phase samples.

| **OD measurement at 660 nm** | | | | |
|---|---|---|---|---|
| **Parallel** | **1** | **2** | **3** | **Average** |
| **P1** | 0.351 | 0.358 | 0.361 | 0.357 |
| **P2** | 0.403 | 0.401 | 0.396 | 0.400 |
| **P3** | 0.353 | 0.351 | 0.347 | 0.350 |
| **P4** | 0.370 | 0.369 | 0.366 | 0.368 |

After 9 hours the four parallel RD5-30 flasks were sampled for the stationary phase samples, and the OD of each sample was measured (Table K.2).

**Table K.2:** OD measurements taken after 9 hours at 660 nm for each of the four parallel RD5-30 flasks for the stationary phase samples.

| **OD measurement at 660 nm** | | | | |
|---|---|---|---|---|
| **Parallel** | **1** | **2** | **3** | **Average** |
| **P1** | 0.622 | 0.610 | 0.606 | 0.613 |
| **P2** | 0.672 | 0.662 | 0.652 | 0.662 |
| **P3** | 0.564 | 0.562 | 0.558 | 0.561 |
| **P4** | 0.584 | 0.576 | 0.562 | 0.574 |

The BB120 overnight culture had an OD of 1.164 when measured at 660 nm. The four BB120 flasks were sampled after 6 hours and 48 minutes for the exponential phase samples, and OD was measured (Table K.3).

**Table K.3:** OD measurements after 6 hours and 48 minutes at 660 nm for each of the four parallel BB120 flasks for the exponential phase samples.

| **OD measurement at 660 nm** | | | | |
|---|---|---|---|---|
| **Parallel** | **1** | **2** | **3** | **Average** |
| **P1** | 0.952 | 0.954 | 0.958 | 0.955 |
| **P2** | 1.052 | 1.052 | 1.050 | 1.051 |
| **P3** | 0.998 | 0.996 | 0.996 | 0.997 |
| **P4** | 0.938 | 0.928 | 0.926 | 0.931 |

After 10 hours and 39 minutes, samples from all BB120 flasks were sampled for the stationary phase samples and the OD of these samples were measured (Table K.4).

**Table K.4:** OD measurements after 10 hours and 39 minutes at 660 nm for each of the four parallel BB120 flasks for the stationary phase samples.

| **OD measurement at 660 nm** | | | | |
|---|---|---|---|---|
| **Parallel** | **1** | **2** | **3** | **Average** |
| **P1** | 1.268 | 1.268 | 1.268 | 1.268 |
| **P2** | 1.232 | 1.232 | 1.232 | 1.232 |
| **P3** | 1.292 | 1.292 | 1.284 | 1.289 |
| **P4** | 1.260 | 1.260 | 1.264 | 1.261 |

## APPENDIX L Suggested secondary structures 16S rRNA V1-region

The remaining RD5-30 V1-region 16S rRNA sequence variant secondary structures as suggested by the RNA secondary structure predicting tool found at https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html are presented in Figure L.1.

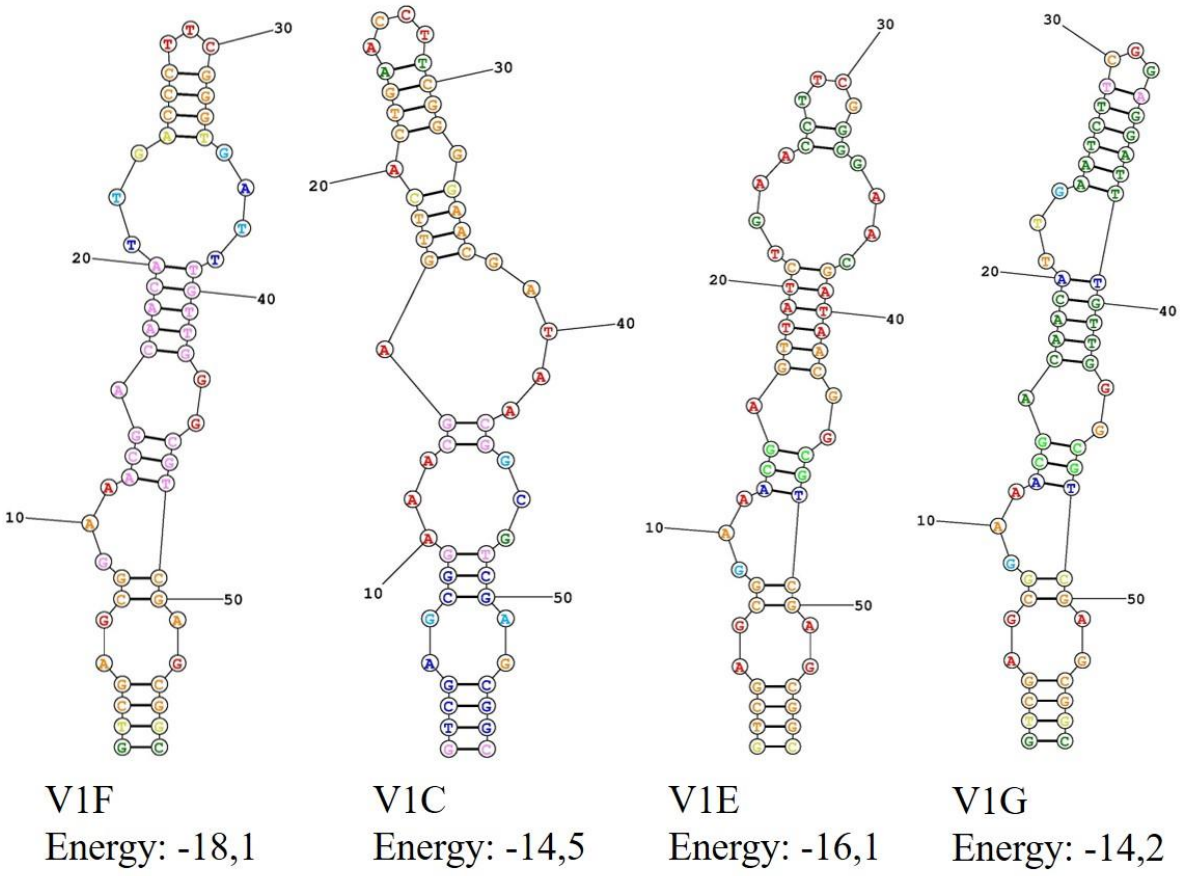


**Figure L.1:** Suggested secondary structures of RD5-30 16S rRNA V1-region sequence variants.

## Appendix M  Suggested secondary structures 16S rRNA V3-region

The remaining RD5-30 V3-region 16S rRNA sequence variant secondary structures as suggested by the RNA secondary structure predicting tool found at the URL https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html are presented in Figure M.1.

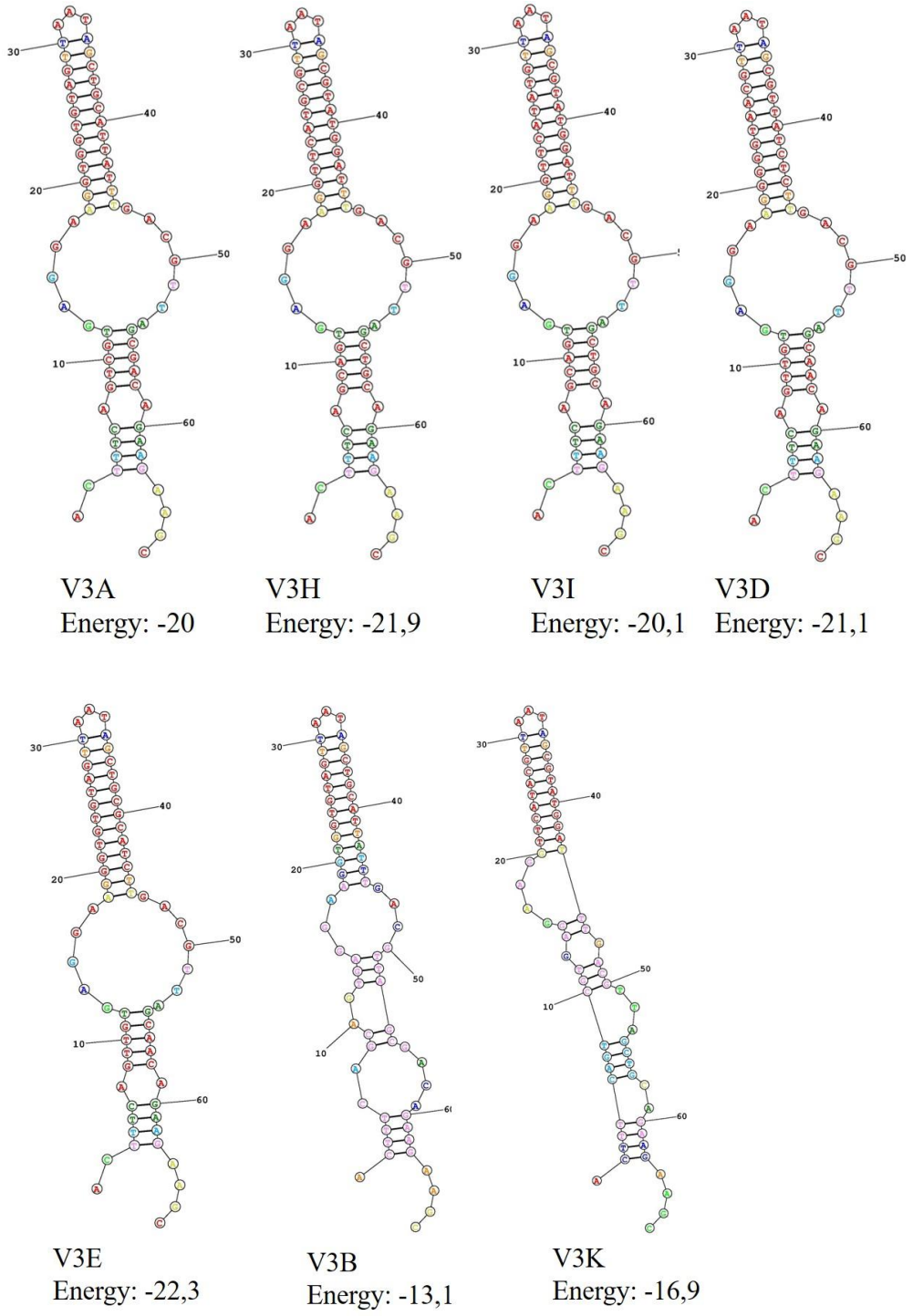


**Figure M.1:** Suggested secondary structures of RD5-30 16S rRNA V3-region sequence variants.