Indication of bacterial species in water samples

Bacterial 16S gene sequencing by Ion Torrent semiconductor technology

Graduation research | *Tímea Kovács*

Van Hall Larenstein University of Applied Sciences

Hogeschool VHL University of Applied Sciences

umcg

30st June 2015
Groningen

Foreword

The research was carried out in the Molecular Microbiology Laboratory of UMCG in Groningen. Wetsus was cooperating with UMCG during the research and provided the waste water samples. UMCG has a high-quality laboratory for research and medical care purposes. Wetsus Academy focuses on Water Technology education and research, with excellent laboratory facilities. I'm very thankful that I could make use of these facilities during my graduation research.
The research was done within the time period of 2nd of February – 30st of June, 2015.

This study would not have been possible without the constant help of Ruud Deurenberg who taught me about the 16S NGS technology and Ilse Verburg who helped our work in the laboratory. Furthermore Jos Theunissen and Marco Verkaik who gave support and supervision during this study.

Tímea Kovács

Groningen, 30st of June 2015

List of terms and definitions

16S NGS = Next Generation Sequencing of the bacterial 16S gene by Ion Torrent semiconductor technique
ABR = Antibiotic resistance, refers to the resistance to antibiotics that occurs in bacteria.
ARG = Antibiotic resistant gene
DWTP = Drinking water treatment plant
NGS = Next Generation Sequencing
Low TE = 10 mM Tris-HCl + 0.1 mM EDTA
Pathogen =refers to bacteria which cause disease or illness to its host.
PCR = Polymerase Chain Reaction
Primer = short nucleic acid sequence which is the starting point for DNA synthesis.
Waterborne disease = pathogens are transmitted by ingestion of contaminated water.
WWTP = Waste water treatment plant

Table of Contents

I. INTRODUCTION

Waterborne diseases are posing a constant threat for people. In developing countries there is a lack of microbiologically safe drinking water and in the developed countries there are occasional waterborne disease outbreaks. These diseases are caused by bacteria or parasites which can be found in various environments, such as drinking water, surface water, waste water etc., when the water is contaminated with human and animal waste.

Waterborne diseases are in correlation with the state of environment – e.g. with surface water quality. The One Health concept focuses on the relationships between the health of humans and the health of animals and the environment (1). There are several examples that show evidence for the influence of the environment on human health. For instance, animals can pass diseases to humans – these diseases are known as zoonotic diseases such as Rabies, Salmonella infection and West Nile virus fever.

In recent years many factors changed in the interactions between humans, animals, and the environment. For example, people migrate to cities where there is no treatment for wastewater (developing countries) or intensive farming with the use of antibiotics (developed countries) releases antibiotic into the environment as well as provides different quality of food products for human consumption. These changes caused the emergence and reemergence of many diseases.

Bacterial species cause several waterborne diseases. These pathogenic bacterial species can also have the property of being resistant to antibiotics. Antibiotic resistance (ABR) refers specifically to the resistance to antibiotics that occurs in bacteria. A microbiologically contaminated water source can cause water-borne diseases because of the presence of pathogens. These pathogens can be resistant to antibiotics – in this case a water-borne disease and in the same time antibiotic resistance are passed over to humans from the water source – see Figure 1.

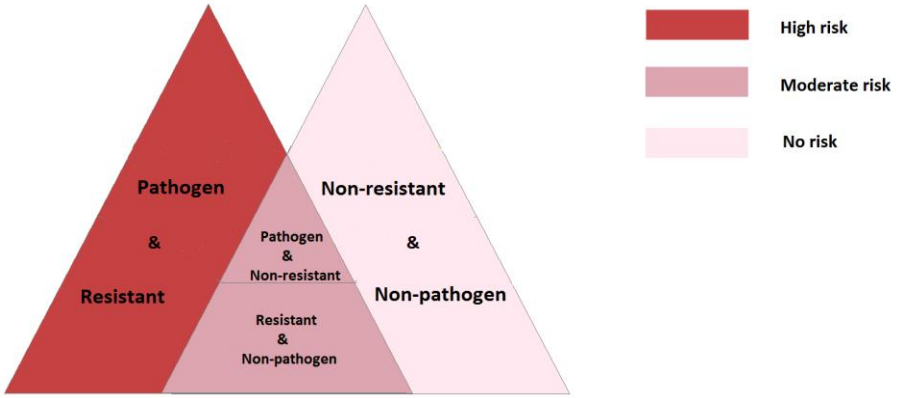


Figure 1: Risks on human health: pathogenic and ABR bacteria (illustration from the author)

Bacterial species which are not pathogenic but resistant to antibiotics pose a threat to human health, since these organisms can pass the ABR to pathogenic bacterial species via horizontal gene transfer (2). In this way non-resistant pathogens become ABR which can cause a hardly treatable human infection. With the spread of ABR, the diseases which were once easily curable with antibiotics become deadly again (3).

I.1. Problem description

To prevent waterborne diseases the microbial composition of water sources have to be known, as well as the ABR of the bacterial species. The microbial contamination of water originates largely from wastewater treatment plants' discharge into surface water, as well as the discharge of untreated wastewater which occurs especially at rural regions where the households are not connected to the sewer system.

The methods which are used to investigate the bacterial composition of the water samples are not accurate – especially when water samples contain several bacterial species and the composition of these bacteria is not known before the analysis. These methods are based on cultivating the bacterial species from water samples on plates, using PCR technique, or using DNA sequencing technologies:

- ***Cultivate bacterial species on plates:***

  It is difficult, because estimates show that less than 2% of the known bacterial species can be cultured in the laboratory. This is among others due to their requirements of special nutrients in the culture medium or their growth is dependent on the presence of another species (4).

- ***PCR technique:***

  The DNA has to be extracted from the bacterial species. Afterwards, specific primers are added to the DNA. These primers can amplify a certain species' genetic material in the PCR. If the genetic material is amplified, it indicates the presence of the given species. Therefore, to indicate the different species in a sample with complex bacterial composition, a wide variety of primers have to be applied (5).

- ***Sequencing techniques:***

  They vary in efficiencies. 16S Sanger-sequencing is a common method to identify bacterial species. The technique investigates only a small part of the genetic material, therefore it is less accurate than Next Generation Sequencing technique which investigates a larger part of the genetic material. Both techniques sequence a special part of the bacterial DNA, called the 16S gene. Based on the sequence of this gene the identification on species level is possible. However there are species whose DNA sequence is almost identical – e.g. *Bacillus globisporus* and *Bacillus psychrophilus* share >99.5% sequence similarity with regard to their 16S genes (6). Therefore the more parts of the DNA are sequenced the more reliable the result of the bacterial identification will be.

The variety as well the accuracy of the techniques are good enough to identify the bacterial composition of (water)samples. The most often used technique is the 16S Sanger-sequencing which is not the most accurate method.

I.2. Research goal

The goal of this study was to develop a method which is fast and accurate to investigate the bacterial composition of water samples. This methodology is based on the sequencing of the bacterial 16S gene by Ion Torrent semiconductor sequencing technology. This study determines the effectiveness of the 16S NGS technique to identify bacterial species using both cultivated bacterial samples from laboratory and water samples.

The technique was tested on water samples taken from untreated and from treated wastewater – as primary sources of the bacterial contamination of surface water. If among the detected bacterial species were any clinically important species present, we determined their antibiotic resistance.
This study determines the effectiveness of Ion Torrent semiconductor sequencing technology on bacterial DNA which originates either from cultivated samples from laboratory or from water samples. Furthermore, this study provides an indication about the bacterial composition of wastewater and ABR from different locations.

I.3. Research questions

The research questions in this thesis are:

1. **Is the sequencing of the bacterial 16S gene by Ion Torrent semiconductor technology applicable to identify the complex bacterial composition of water samples?**

2. **What can be concluded about the water samples which were used during the tests regarding their bacterial compositions and the ABR of clinically important species?**

Sub-questions:

- How is the performance of 16S NGS on cultivated bacterial samples?
- How is the accuracy of 16S NGS influenced by the number of bacterial species present in the cultivated bacterial samples?
- How is the performance of 16S NGS on water samples?
- How is the performance of 16S NGS compared to conventional Sanger-sequencing?
- What kind of clinically important bacterial species are detected in the water samples?
- Which water source contains ABR which are clinically important?

II. BACKGROUND

A vast amount and species of micro-organisms can be found in the aquatic environment. The microbial contamination of water originates from various sources, depending on the type of water:

1. ***Drinking water***
    - Bacteria found in groundwater or in surface water are not completely removed at the DWTP
    - Contamination in the pipeline
2. ***Surface water***
    - Runoff water from agricultural fields
    - Storm water outlet of sewage system
    - Natural fauna
3. ***Recreational water (swimming pool)***
    - Human (faecal) contamination
4. ***Waste water***
    - Sewage from household
    - Sewage from industry
    - Storm water and snow melt

A number of the micro-organisms which are present in the aquatic environment are pathogenic, therefore diseases can be spread via the contaminated water sources. Pathogens on their own pose a threat to human and animal health, however if resistance to antibiotics is developed by these pathogens, it poses an elevated risk to public health. When pathogenic bacteria develop ABR, the treatment of such a medical case is more difficult, takes longer and has a bigger risk of death compared to infections caused by non-resistant pathogenic bacteria (3).

II.1. Water associated diseases

Direct contact with microbial contaminated surface water may cause adverse health effects (gastrointestinal and stomach illnesses), although the health effects are largely dependent on the amount and type of contaminants, the way of exposure, the age and the general health status of the affected person. The diseases which are associated with water are caused either by bacteria, viruses, protozoans, nematodes (roundworm, hookworm) or cestodes (tapeworm). The presence of these organisms is related to the contamination of water by human and animal waste. The water associated diseases are classified in the following categories (7):

1. ***Waterborne diseases:*** pathogens are transmitted by ingestion of contaminated water.

2. **Water-washed diseases***:* caused by the lack of proper water supply for (hand) washing and hygiene.

3. **Water-based infections:** part of the life cycle of pathogenic organisms is spent in other aquatic organisms – all of these diseases are caused by parasitic worms.

4. **Water-related diseases***:* diseases (e.g. malaria) are transmitted by insects which live near to water.

In this study only bacterial species were investigated, therefore this research is related to the category of **waterborne diseases**.

II.2. Antibiotic resistance

Pathogenic bacteria which developed resistance to antibiotics are can pose an elevated risk to public health. The standard medical treatment is often ineffective in the case of infections caused by resistant microorganisms. The patient receives an alternative therapy which is usually longer, more expensive, less effective and has a greater risk of death (3).

Over the last 30 year no new major types of antibiotic have been developed – see Figure 2. Some years after the introduction of a new type of antibiotic, the bacterial resistance against it was identified – see Figure 3. The cost of the development of antibiotics are high and if they are useful only for a short time period it is not worth to invest in their development. This means that diseases which were once curable by antibiotics will again be a serious threat on human (and animal) heath because of the emerging ABR (8).

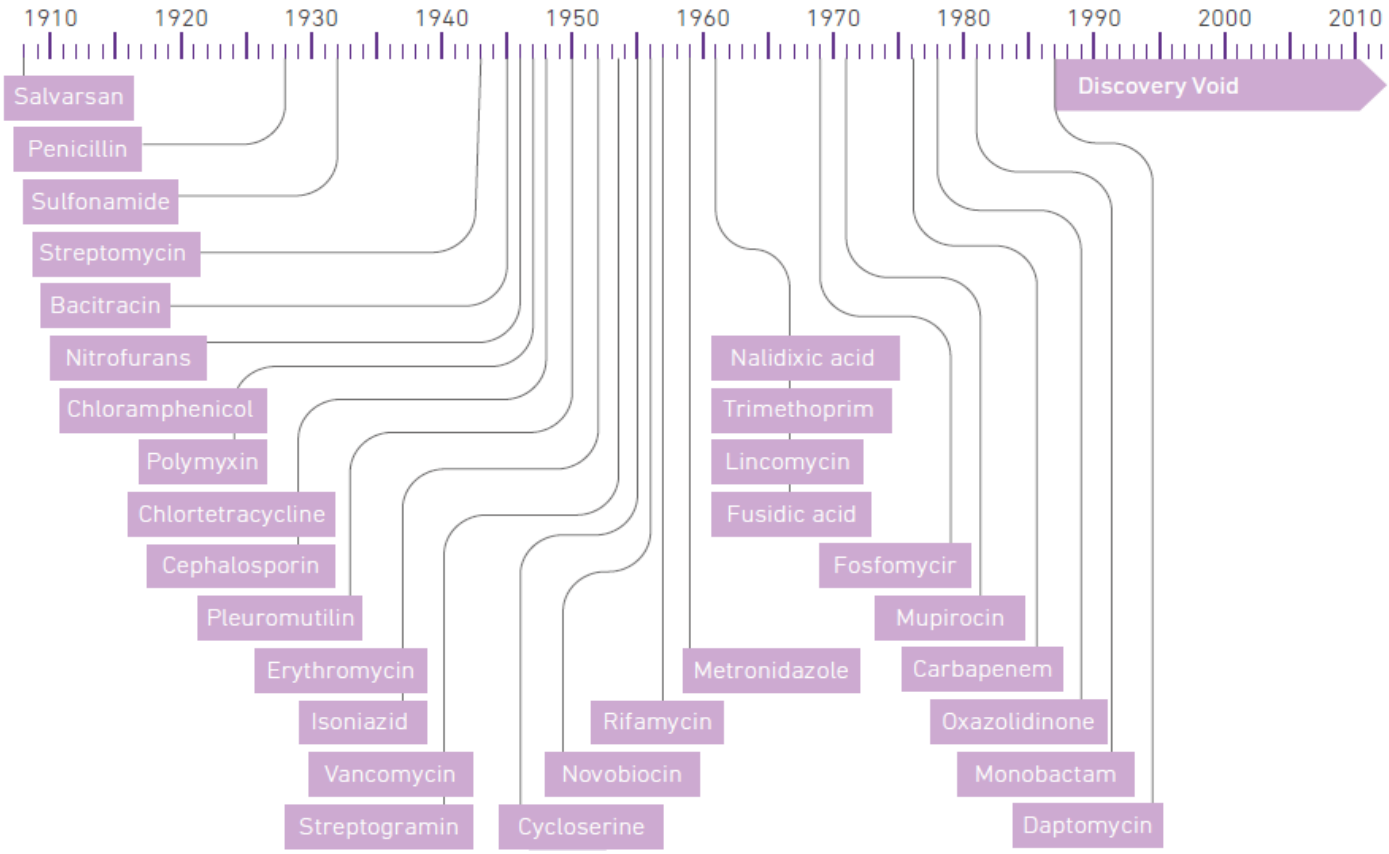


Figure 2: Dates of discovery of distinct classes of antibacterial drugs (8)

Medical treatments, such as organ transplantation and surgical procedures, could not be performed effectively without antibiotics. According to the European Centre for Disease Prevention and Control and the European Medicines Agency, every year approximately 25 000 European citizens (5.1 per 100000 inhabitants) die from infections caused by bacteria that have developed resistance towards antimicrobials (9).

ABR has an effect on economy also, because of the direct costs of increased health-care expenses. Furthermore the indirect costs can be three times as much as the direct costs. Estimates show that losses in GDP can be more than 1% (3).

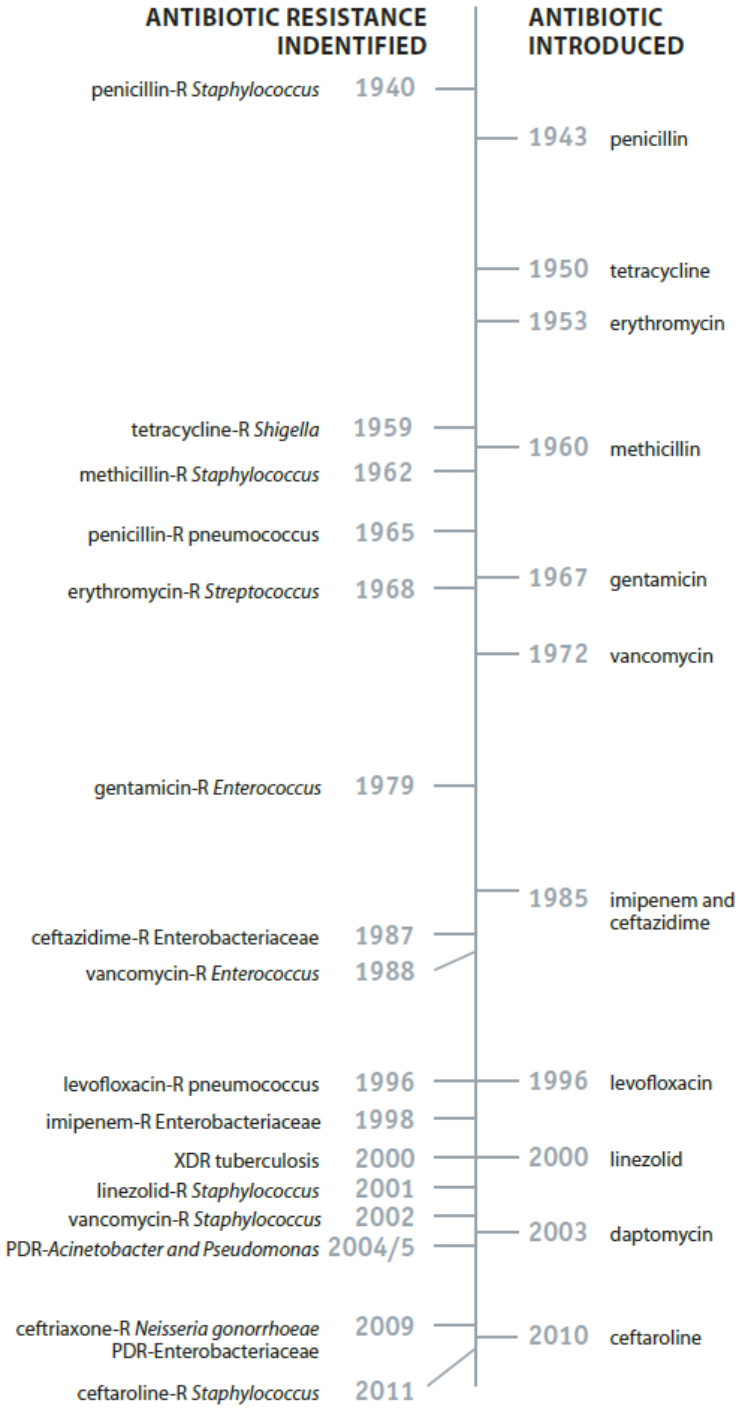


Figure 3: The emergence of antibiotic resistance in time order (10).

II.2.1. Development of antibiotic resistance

The origin of ABR is not linked to the anthropogenic use of antibiotics. Bacterial species developed the ability to synthesize bioactive molecules whether to antagonize or to cooperate with other members of the bacterial population. As a result, they have developed defense systems to protect themselves against these bioactive molecules of others. These molecules are encoded in the genetic material – which part of the genome is called the resistome (9).

There are different mechanisms inside the bacterial cell which protect the bacteria against antibiotics (9):

- exclusion of the antibiotic by the cell membrane
- intracellular modification and/or deactivation of the antibiotic
- reduction in sensitivity of the cellular target
- extrusion from the cell
- intracellular sequestration.

The above mentioned defense mechanisms can evolve through two different ways (9):

1. By mutation and selection
2. By acquiring, from other bacteria, the genetic information that encodes resistance.

In the case of **mutation and selection,** the concentration of antibiotics in the bacterial environment is the key element for the selection. The Mutant Selection Window is the concentration region of antibiotics where the non-resistant bacteria population dies out and the ABR population survives and grow – see Figure 4. This figure indicates that at low antibiotic concentration, the number of ABR mutants will not increase; at high antibiotic concentration both the normal and the ABR population dies out.

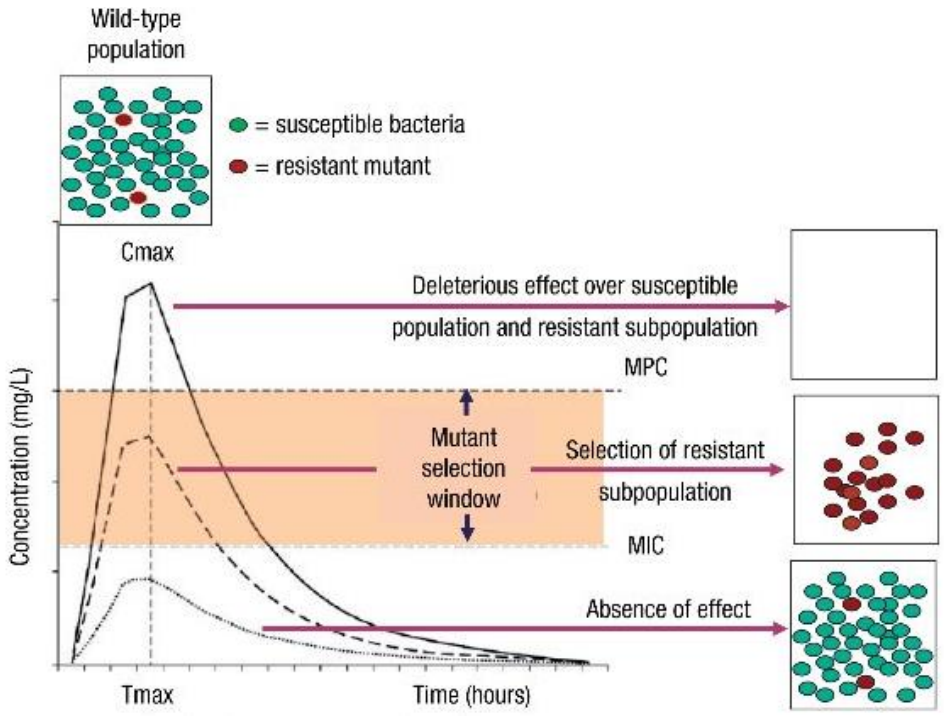


Figure 4: Mutation Selection Window for the development of antibiotic resistance (11).

The development of ABR is a natural phenomenon, however certain human actions contribute to the process, such as inappropriate use of antimicrobial drugs (including in animal husbandry) and poor infection prevention and control practices (3).

II.2.2. Antibiotics and ABR bacteria in the environment

There are two hypotheses to explain the presence of ABR bacteria in the environment:

1. **Antibiotics are released into the environment** – it triggers the selection of resistant populations and thus increasing the number of them.
2. **ABR bacteria are released into the environment** – their genetic material is released and acquired by non-resistant bacteria. The incorporated genetic material from the ABR bacteria transforms the host from a non-resistant into an ABR bacteria.

If the amounts of antibiotics are compared from every sector where they are applied, than it can be concluded that not the human medical application is the main site where antibiotics are used, but it is agriculture – see Figure 5 as an example from the US.

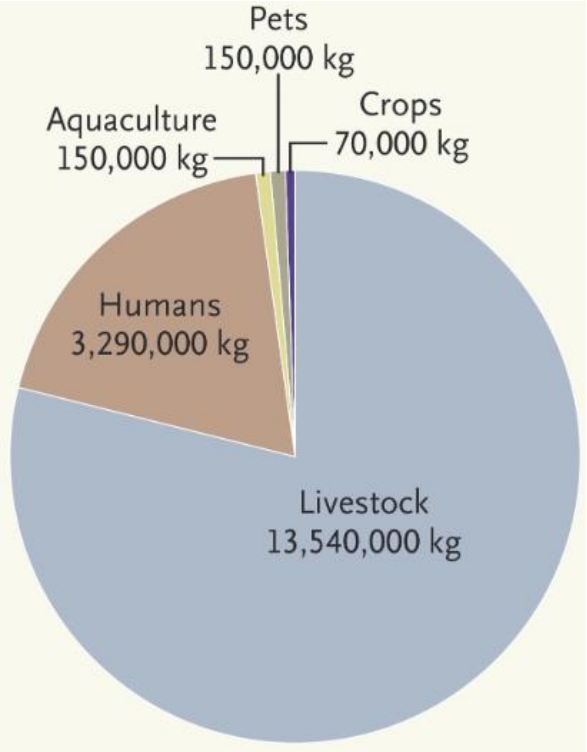


Figure 5: Estimated annual antibiotic use in the United States (15).

Agricultural fields are one of the main site where antibiotics are released into the environment. Antibiotics which are used for the medication of the livestock are excreted via the manure. Therefore the agricultural fields where this kind of manure is spread are the source of antibiotic residues and ABR bacteria. The crops which are grown in these lands can also carry the contamination further.

Waste water treatment plants (WWTPs) represent another main site through which antibiotics are released into the environment. Common biological treatment technologies that are being applied in WWTPs provide an ideal environment (i.e., high bacterial densities, high oxygen, and high nutrient concentrations) for the development of ABR, because bacteria are in continuous direct contact with antibiotics and with ABR bacteria – which gives place for both of the above mentioned hypothesis of the origin of ABR (9).

To prevent the waterborne diseases, water quality is regularly checked. Microbial safety of water is checked by indicator organisms – e.g. *E.coli* is the indicator organism for faecal contamination. There are standards and regulations set about the tolerable number of these microorganisms – e.g. the WHO limit for drinking water is <1 CFU *E.coli* in any 100 mL sample (12).
Figure 6 sums up the spread of ABR in the environment and its relation to human health.

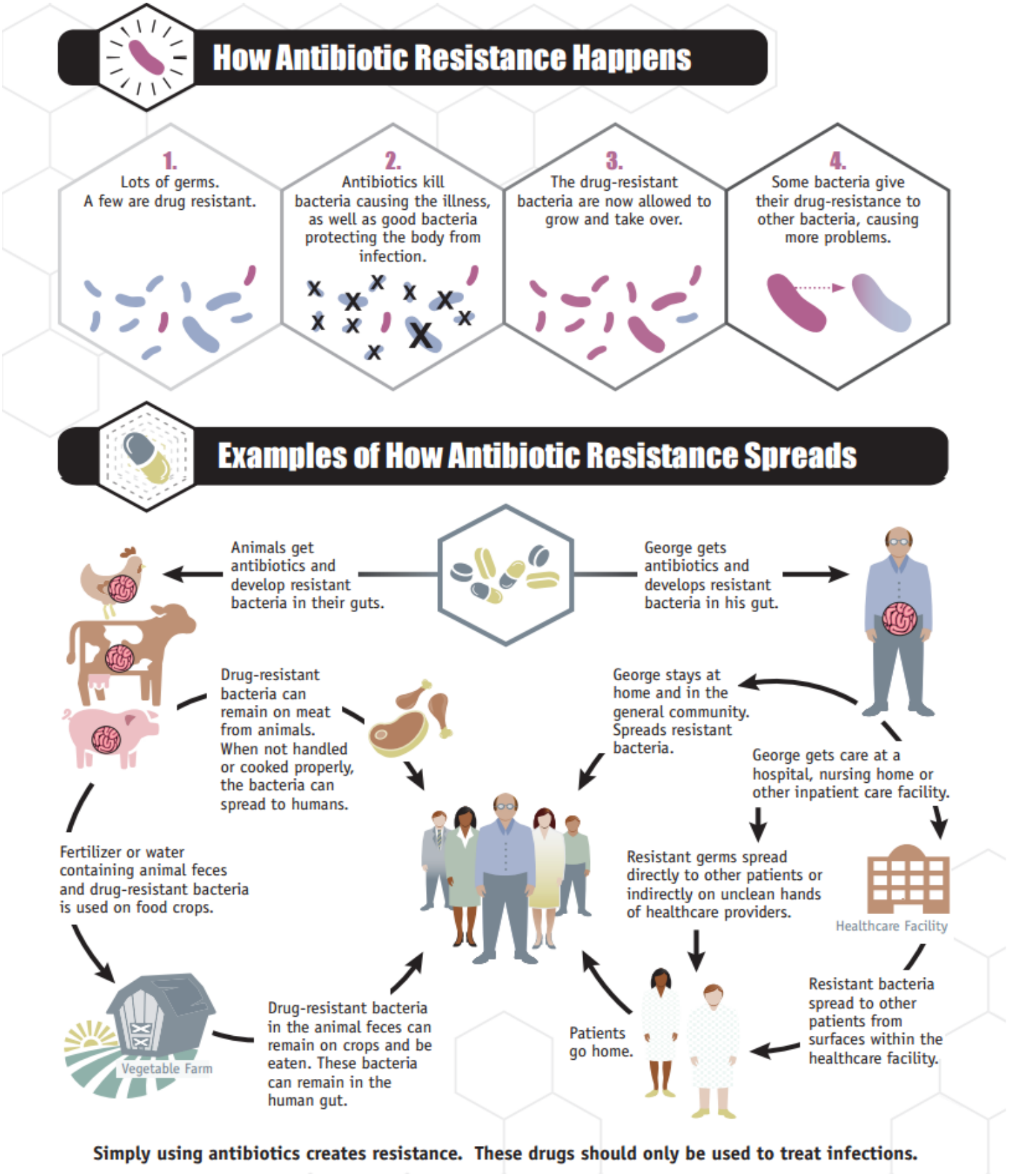


Figure 6: Summary of the spead of antibiotic resistance (10).

II.3. Water treatment plants – removal of bacteria and antibiotics

The water samples in this study are all related to waste water. While waste water going through different treatments in the waste water treatment plant (WWTP), the bacterial composition as well ABR might change. The removal mechanisms of pathogens and antibiotics are listed below.

Surface water is a source applied water in several facilities – e.g. recreational water in swimming pools, drinking water and irrigation water at agricultural fields. If surface water contains pathogens or residues of antibiotics, certain applications (e.g. drinking water production) require sterilization or disinfection of the surface water before it gets in contact with humans. To prevent the waterborne diseases microbiologically safe water is essential. **Sterilization** is the destruction of all the organisms in the final effluent regardless of whether they are pathogenic or not, while **disinfection** is the selective destruction of disease-causing organisms (7).

II.3.1. Bacteria removal

To reduce the amount of pathogens several methods are used in *WWTPs* and in household applications. These techniques are based on the following principles:

1. *Physical removal:*
   Boiling, UV-light, solar radiation, filtration, adsorption, sedimentation, coagulation
2. *Chemical removal:*
   Chlorination, ozonisation
3. *Biological removal:*
   By activated-sludge, macro-invertebrates, digestion

II.3.2. Antibiotics removal

Although WWTPs are not designed to clear antibiotics, there are ways in which the antibiotic concentration is reduced within the WWTPs:

1. *Sorption*: antibiotics are sorbed onto the surface of suspended solids which are removed during sedimentation – see Table 1 for different sorption constants (K_d) of antibiotics in activated sludge.

Compound	K_d (mL/g or L/kg)
Tetracycline	8400 ± 500
Oxytetracycline	3020
Ciprofloxacin	416.9
Sulfamethoxazole	256± 169
Trimethoprim	208± 49
Azithromycin	376± 86
Clarithromycin	262± 93
Estrone	402± 126
17β-estradiol	476± 192
17α-ethinylestradiol	584± 136

Table 1: Estimated sorption constants (K_d) of antibiotics and drugs in activated sludge (13).

2. *Biodegradation*: antibiotics are degraded by bacteria in activated sludge (13).

The biodegradation and sorption rate in WWTPs also largely depended on the chemical properties of the antibiotics.

The normally detected range of antibiotics in different water samples (14):

- higher mg/L range: hospital effluents
- lower mg/L range: municipal wastewater
- ng/L range: surface, sea and groundwater

So far, legal limits have been established for antibiotics in foods: 4-1500 mg/kg for milk and 25-6000 mg/kg for the other food stuffs of animal origin. There is no legislation applied to environmental mediums (14).

Since WWTPs and drinking water treatment plants (DWTPs) are not designed for the removal of antibiotics and drug residues, additional processes are needed to remove the antibiotic content of drinking water and the effluent of the WWTPs. These processes are listed below with their antibiotic removal efficiencies (14).

1. *Oxidation processes*
   Chlorination: >90%
   Ozonation: >76%
   Fenton and photo-Fenton: >80%
2. *Adsorption processes:*
   Active carbon: 90-95%
3. *Membrane processes:*
   Reverse osmosis, Nano filtration, Ultra filtration: >90%
   Ion exchange: 80-90%

Some of the studies mentioned the fact that upon antibiotics degradation the produced metabolites' ecotoxicity was not investigated, or if it was the metabolites' ecotoxicity was largely dependent on the kind of antibiotics – some studies reported enhanced, some studies reported degraded ecotoxicity (14).

II.4. Identification of bacterial species in water samples

II.4.1. Plating technique

To identify the bacterial composition of water samples, a part of the water sample can be applied on cultivating medium. Every species has its own preference for cultivating medium with specific nutrients. Furthermore, they have their specific environment, temperature etc.

The primary limit of this technique is the ability to grow any bacterial species, because only 2% of the bacterial species can be cultured *in vitro*. Therefore, if a water sample is applied on different cultivating media, placed on different temperature, then as a result, distinct bacterial cultures will grow on these plates.

Furthermore if a mixture of bacterial species is cultured in the same plate, one species can overgrow in the plate – it means that a certain species suppress the other ones. Also the way the sample is preserved has an effect on the outcome. For instance, glycerol solution can be added to the water samples to prevent the damage of water crystals to bacteria during the storage period. As an example see Figure 6, which illustrates the difference between the bacterial cultured from water samples with and without glycerol, under the same conditions:

- The bacterial activity is higher in the water sample with glycerol solution, therefore after one day a few species of bacteria dominate in the plate (Figure 7-b).
- The bacterial activity was lower in the water sample without glycerol, however a much diverse culture was observed, because there was no overgrowth of a certain species (Figure 7-a).

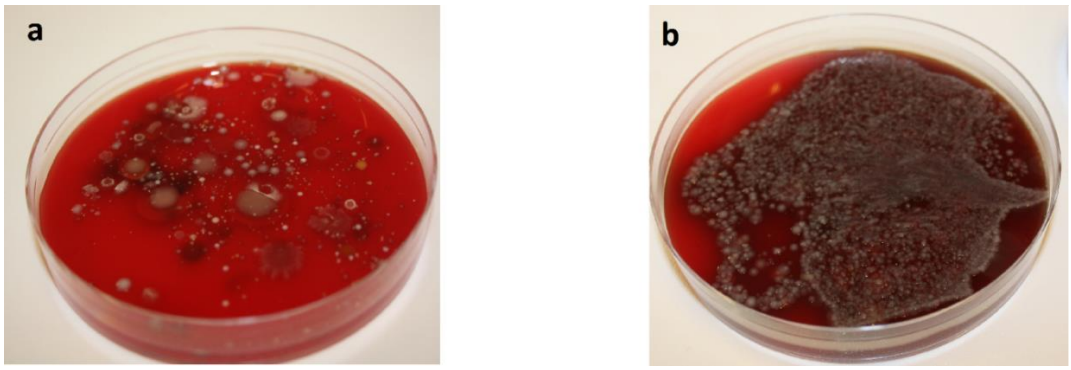


Figure 7: Wastewater samples cultivated on blood-Agar media, after 1 day. The samples were coming from the same source (at the same time). They were stored at -20°C and plated in the same way. **a)** diverse bacterial culture *without glycerol* **b)** overgrown bacterial culture *with 15% glycerol* (illustration from the author)

Therefore to indicate by plating technique the bacterial species in water samples the following have to be taken into consideration:

- Culturable and non-culturable species
- Storage of water samples to preserve bacterial activity
- Time length for culturing – not too long to prevent overgrowth, not too short to provide enough time for growing

II.4.2. Sequencing

The identification of bacterial species is based on the sequencing of their 16S gene. The 16S gene is present in all bacterial species and it consists of 9 DNA regions called V1-9 – see Figure 8. Every species of bacteria have their specific nucleotide sequence in their V1-9 region. Therefore, based on the sequence of the V1-9 regions, it can be determined that from which bacterial species the genetic material originates. Every bacterial species has its specific sequence for their V regions – e.g. the sequence of V3 region is different in *E.coli* than in *S.aureus*.

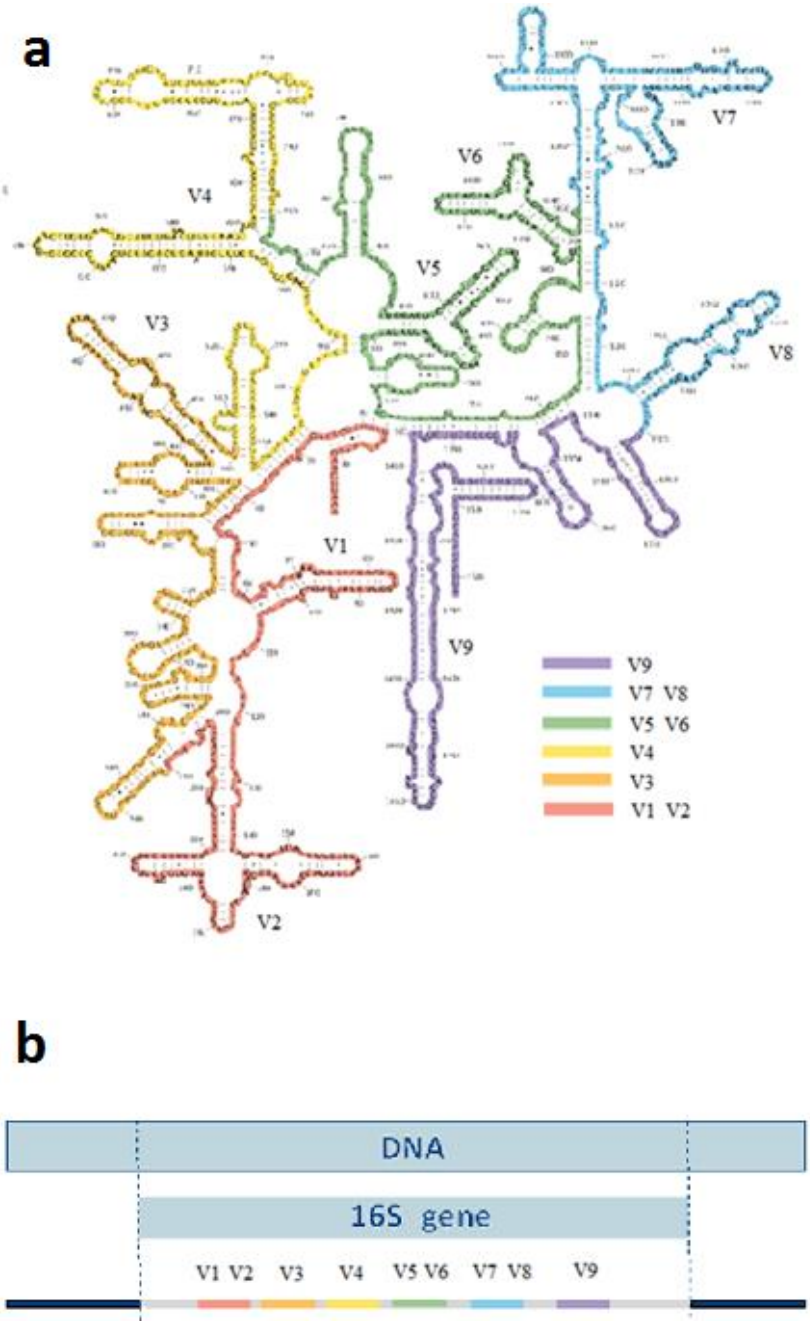


Figure 8: 16S gene and within the V1-9 regions in *E.coli*. **a)** model of the 16S gene(after 15) **b)** schematic representation of the 16S gene

To reveal the nucleotide sequence of the 16S gene, the DNA has to be replicated – see Figure 9. The DNA replication process consists of the following steps:

1. **"Unzip" DNA:** The double stranded DNA is treated in a way that it becomes single-stranded (Figure 8- a).
2. **Replicate**: A polymerase binds to the single-stranded DNA, as it moves along the single strands it incorporates the matching nucleotides to form a complementary strand (Figure 9 – b).
3. **New product:** DNA forms a double strands again, which consists of the original strand and the newly built complementary strand (Figure 8 – c).

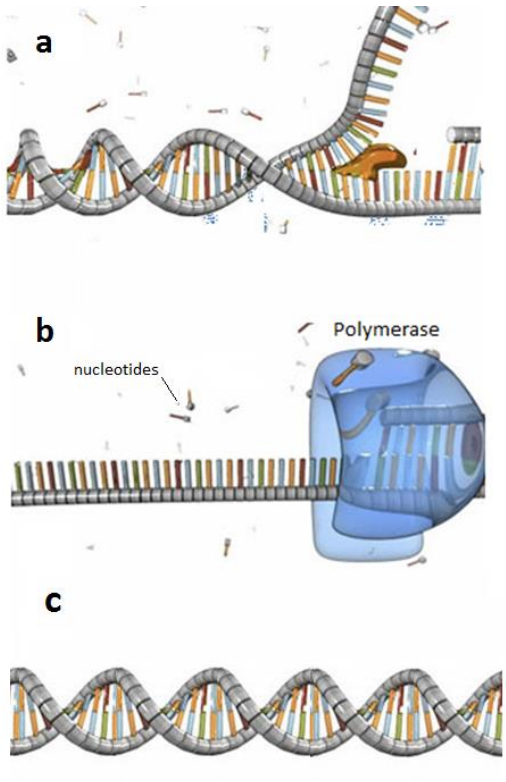


Figure 9: Replication of the genetic material. **a)** *"Unzip" DNA* – it becomes single stranded **b)** *Replication* – the polymerase builds up the complementary strand **c)** *New product* - the newly built complementary strand and the original strand (after 16).

There are 4 types of nucleotides: Adenosine (A), Thymine (T), Guanine (G) and Cytosine (C). The nucleotides are incorporated into the DNA to form the complementary strand during the replication process. The basis of all the sequencing techniques is to register the type of nucleotide as it is being built into the DNA. The different sequencing methods apply different technologies in order to follow this process. Furthermore, the sequencing methods also differ in the regions of DNA what they sequence.

The three main sequencing techniques are presented below. As a short summary Table 2 lists the principles of these techniques.

	Sanger - sequencing	Next Generation Sequencing	
		Illumina	*Ion Torrent*
Sequenced regions of the 16S gene	V3	V3 and V4	V2, V3, V4, V6, V7, V8 and V9
Identification of nucleotides	After gel electrophoresis	Real-time	Real-time
Based on	Fluorescent labelling	Fluorescent labelling	pH change

Table 2: Principles of the different sequencing techniques

Sanger-sequencing

***Sequenced part of the DNA**:* V3 region within the 16S gene.

Based on: fluorescent labelling, electrophoresis (size differentiation)

The DNA extract is divided into 4 parts – to each part a mixture of **nucleotides** (A,T,G and C) and so called **chain terminators** (modified fluorescent nucleotides – ddA, ddT, ddG and ddC) are added. The matching nucleotides are incorporated into the DNA. This process continues until a chain terminator is incorporated, which stops the replication.

Based on how late the replication is stopped, different lengths of DNA pieces are created:

- If the chain terminator incorporated fast → short DNA replicate
- If the chain terminator incorporated later → longer DNA replicate

Based on the type of the chain terminator what was incorporated, these pieces of DNA will end with one of this modified nucleotide: ddA, ddT, ddG or ddC. The DNA extract is placed then on a gel to be separated by size using electric current – this technique called gel electrophoresis. Finally, based on the size differentiation and the fluorescently labelled DNA-ends, the sequence can be read – see Figure 10 for an illustration.

III.4.3. PCR technique

The DNA has to be extracted from the bacterial species. Afterwards, specific primers are added to the DNA. These primers can amplify a certain species' genetic material in the PCR. If the genetic material is amplified, it indicates the presence of the given species. Therefore, to indicate the different species in a sample with complex bacterial composition, a wide variety of primers have to be applied (5).

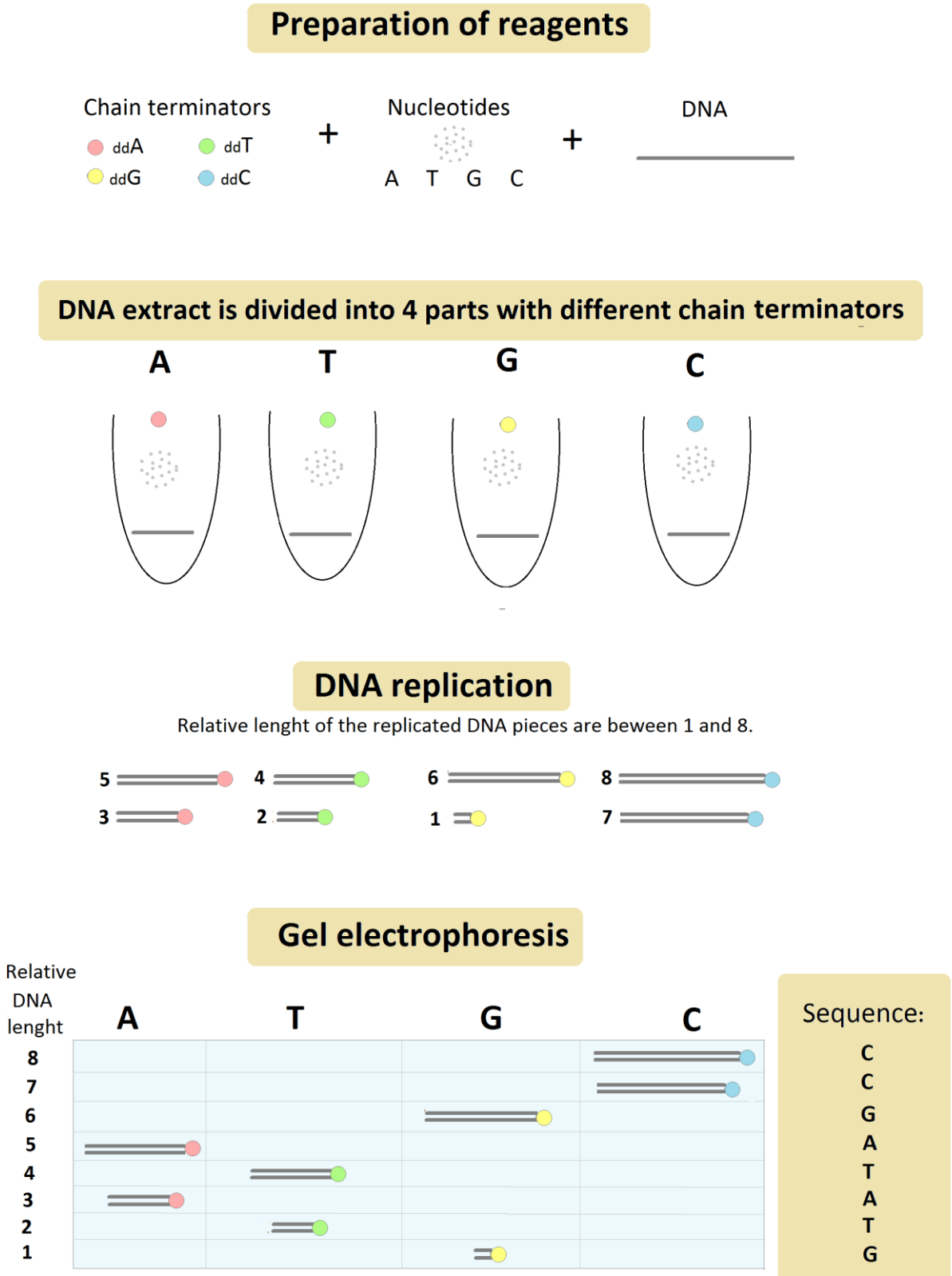


Figure 10: Overview of the Sanger-sequencing method with example (illustration from the author)

Next Generation Sequencing (NGS)

Sequenced part of the DNA: *V2, V3, V4, V6, V7, V8 and V9 regions within the 16S gene.*

The most frequently used application are the Ion Torrent and the Illumina technologies. They sequence larger parts of the 16S-DNA than Sanger-sequencing. Furthermore, the way these devices read the sequence of the DNA is based on different principles.

Ion Torrent - Semiconductor Technology

Based on: *pH change*

When a nucleotide is incorporated into the genetic material, a H^+ ion is released. This causes a change in the pH of the DNA solution. The device notice the pH change and registers the given nucleotide which was incorporated. This process is constantly repeated and the registered nucleotides give the sequence.

Illumina – Sequencing by Synthesis Technology

Based on*: fluorescent labelling*

Fluorescently tagged nucleotides are prepared – each type of nucleotide has a different kind of fluorescence. The device identifies the nucleotides when they are incorporated into the genetic material. The nucleotide identification is based on the nucleotide's fluorescence. This process is constantly repeated and the identified nucleotides are registered to give the sequence.

III. METHODOLOGY

This study determines the effectiveness of Ion Torrent semiconductor sequencing technology on bacterial 16S DNA which originates either from cultivated samples from the laboratory or from water samples. Furthermore, this study provides an indication about the bacterial composition of wastewater and indicates ABR from different locations. The study was carried out in three stages:

- Set up of the methodology
- Phase I: 16S NGS on water samples
- Phase II: Antibiotic resistance test on clinically important species

Set up of the methodology

Bacterial species were cultivated on petri-dishes, and DNA was extracted to test our methodology. The chosen species were *Escherichia coli* (Gram negative), *Staphylococcus aureus* (Gram positive) *Klebsiella pneumoniae* (Gram negative), *Pseudomonas aeruginosa* (Gram negative) and *Enterococcus faecium* (Gram positive).

The following tests were carried out with these bacterial species:

1. ***E.coli and S.aureus cultured:*** separately prepared for sequencing (bacterial samples in duplo, with blank and positive control from the kit (*E.coli*) samples)
2. ***5 bacterial species cultured:*** mixed in a tube, the mixture was sequenced (mixture sample in duplo)

These tests were necessary to adjust the amount of applied reagents, reaction time and settings of the machines in order to assess the accuracy of the method.

Furthermore in this research phase patient related clinical samples were also sequenced. The samples were sequenced by Sanger-sequencing method earlier for diagnosis. These samples were in the form of extracted bacterial DNA:

3. **Bacterial DNA extract:** separately prepared for sequencing

These test were necessary to compare the performance of our method and the conventional 16S Sanger-sequencing technology.

Phase I. – 16S *NGS on water samples*

Water samples were filtered through a 0,45 µm membrane which retained the bacteria on its surface. The bacterial DNA was isolated from the membrane.

4. ***Water samples:*** separately prepared for sequencing

This test was necessary assess the performance of our method on water samples which have complex bacterial composition. The outcome of this test also indicated the bacterial composition of the samples, therefore allowing to recognize clinically important bacterial species.

Phase II. – *Antibiotic resistance test on clinically important species*

100 µL solution form the water samples were applied on cultivating media in Petri-dishes to grow clinically important species which can be found in the samples. Blood-agar medium was applied because it supports the growth of the selected species. During this process many other species were grown next to the clinically important one. The samples were treated by the technician staff in the Medical Microbiology Lab in the UMCG. The clinically important bacterial species were identified by the MALDI-TOF technique, then cultivated separately and finally the antibiotic resistance test was carried out on them by disk-diffusion technique and by VITEK technology.

5. ***Clinically important species**:* bacterial species from water samples were cultivated, the clinically important species selected and antibiotic resistance test was carried out on them

These test were necessary to asses if the clinically important bacterial species are ABR.

Set up of the methodology was done in UMCG, while during Phase I-II the different tasks were divided between the two laboratory of UMCG and Wetsus – see Figure 11.

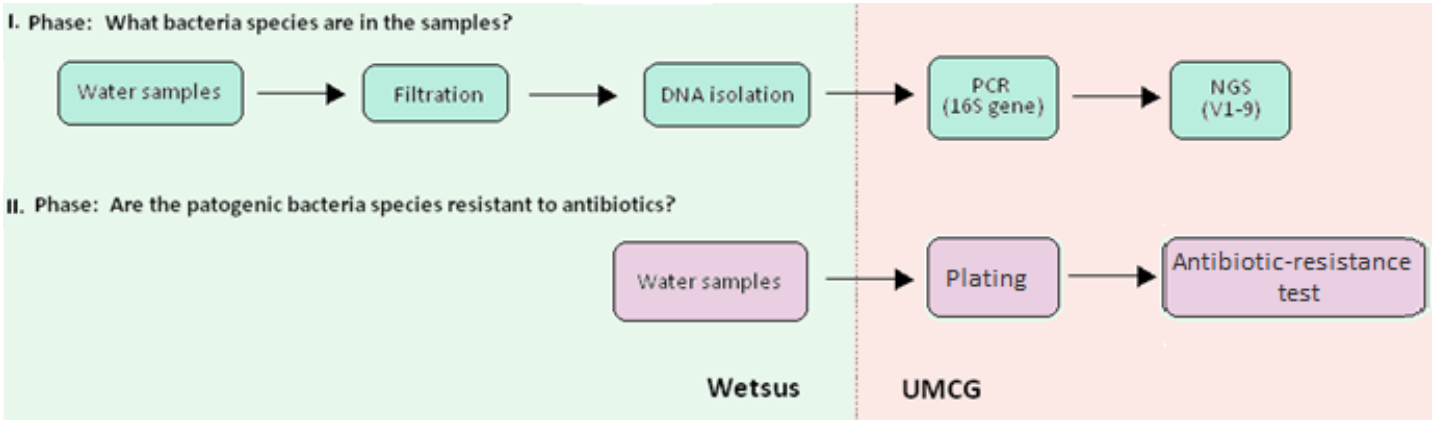


Figure 11: Distribution of tasks between UMCG and Wetsus (illustration from the author)

The outcome of this study is an evaluation of the effectiveness of the 16S NGS technique to identify bacterial species from water samples. The evaluation is based on the outcome of our test which are discussed in the chapter “Results and discussion”.
Furthermore, the bacterial composition of the water samples is presented as a bacterial map which lists the determined species, as well mentions the clinically important species and their ABR.

III.1. Samples

During the different phases of the study three types of bacterial DNA were used for sequencing:

I. DNA from cultivated bacterial species:

Applied: during the set-up of the methodology

DAN was originating from *in vitro* cultured bacterial species.

The bacterial species originated from controlled cultures. These species were applied on blood-agar medium and grown for 24 hours at 37^{o}C. Afterwards, the DNA of the species was extracted.

II. DNA from clinical samples:

Applied: during the set-up of the methodology

DNA was originating from patient related clinical samples, in the form of extracted DNA.

III. DNA from the water samples:

Applied: during Phase I-II

DNA was originating from bacteria from untreated and treated waste water samples. The water samples were filtered through a 0,45 µm membrane which retained the bacteria. Afterwards, the DNA of these bacteria was extracted.

The samples were taken from different sources:

1. Blackwater, Nursing home Noorderhoek, Sneek.

Blackwater means that it is only water from the toilets. The nursing home uses vacuum toilets, so the samples were more concentrated than normal black water samples. This wastewater will be treated at the general WWTP in Sneek.

2. WWTP influent, Sneek

This sample consist of all the wastewater produced by the community of Sneek:

- small part is coming from the nursing home
- wastewater from households
- wastewater from the hospital in Sneek

3. WWTP influent, Grou

Wastewater which is produced by the community of Grou. This water is comparable to WWTP influent, but the main difference is that this sample did not contain any wastewater originating from a hospital.

4. WWTP effluent, Grou

Wastewater after conventional treatment. The biological treatment facilities (daphnia pond, reed filter beds) of the WWTP were not applied.

III.2. From DNA extraction till sequencing

To prepare the samples for sequencing, 13 steps are needed. During these 13 steps, the concentration of the DNA is constantly measured and adjusted. Furthermore, the DNA is purified from the bacterial cell residues and from the applied reagents. The whole procedure takes c.a. 4 days. The principals of each step are presented below. For the detailed information about the applied reagents, reaction times etc. see the detailed protocol in the Appendix.

1. Extraction of DNA

Applied kit: MoBio DNA isolation kit

Devices: Vortex adaptor, centrifuge

To extract the DNA first the bacterial cell wall has to be broken. The cell wall lysis happen due to chemical reactions (addition of reagents to the DNA solution) and by physical forces (vortex adaptor). When the genetic material is not anymore encompassed by the cell wall, it is in a solution together with other building blocks of the cell – e.g. lipids, proteins. The solution has to be cleaned in a way that at the end it contains exclusively DNA dissolved in water. This purification process is done by chemical reactions and filtration.
Samples were treated differently:

- **During the settings of the methodology**: bacteria were taken from the Petri-dishes
- **During Phase I-II:** the water samples were filtered through a 0,45 µm membrane and the bacteria were obtained from the surface of these membranes.

2. Measure DNA concentration

Device: Nanodrop 2000 C Spectrophotometer

One drop (1µL) of DNA solution is placed on the spectrophotometer which measured the DNA concentration (ng/µL) at the wavelength of 260 and 280 nm.

3. Amplify 16 gene

Applied kit: Ion16S Metagenomics Kit

Devices: PCR, centrifuge

The double stranded DNA is “unzipped” – it means that the double strands separate and form single stranded DNA, which called template DNA. The primers are attached to the template DNA and as they are moving along the DNA, they attach the compatible nucleotides to the strand. This process happens repeatedly during the polymerase chain reaction (PCR) in the PCR device: during every cycle the amount of DNA doubles – see Figure 12.

During the study, primers were applied which were connected to the V2, V3, V4, V6, V7, V8 and V9 regions of the 16S gene. Therefore only the V2, V3, V4, V6, V7, V8 and V9 regions of the genetic material were multiplied, while other parts of the DNA weren’t multiplied.

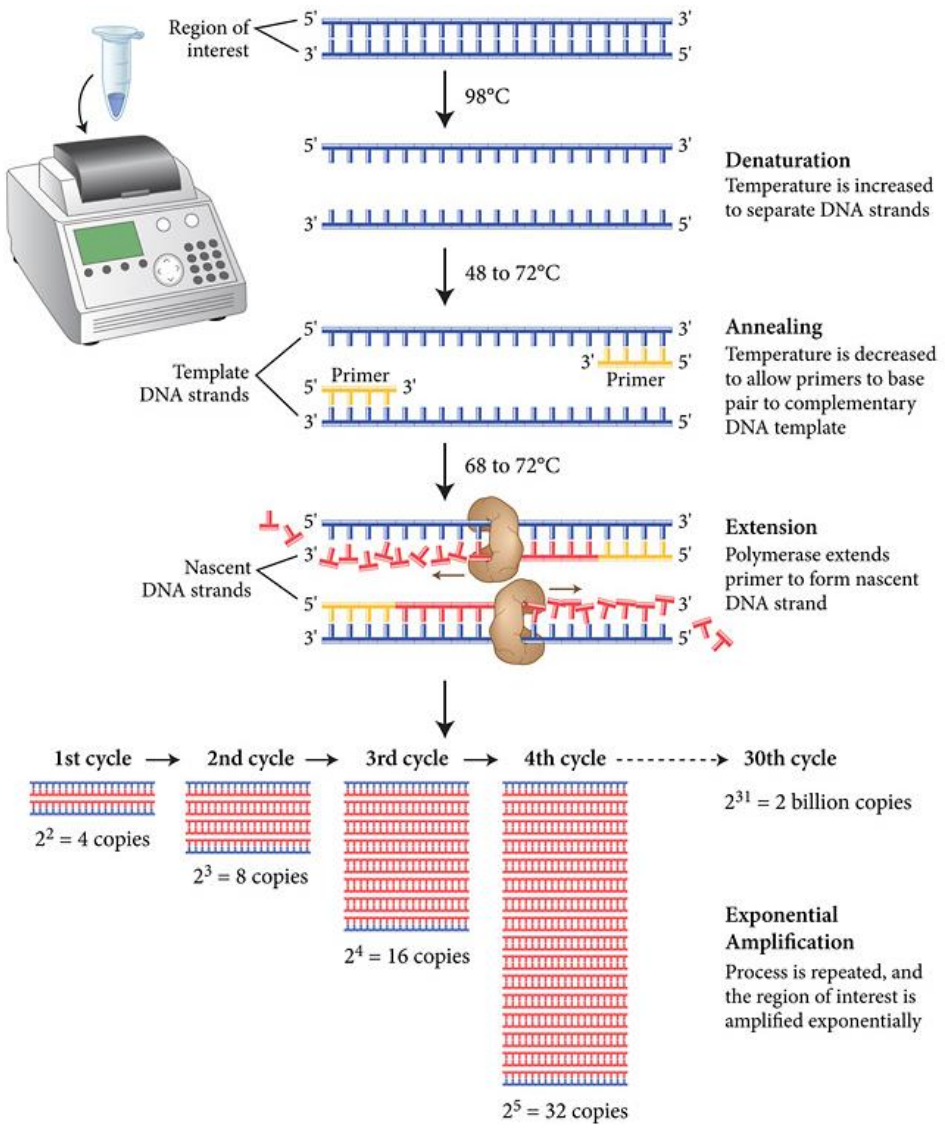


Figure 12: Multiplication of the genetic material by PCR technique (17)

4. Purification

Device: Magnetic rack

The DNA solution has to be purified from the remaining applied reagents. The purifications were done with the use of magnetic beads which binds to the genetic material. The magnetic beads can be separated from the solution by the use of a magnetic rack, which will attract the magnetic particles – see Figure 13. In this way the solution has 2 phase:

- A solid phase with the magnetic beads and with the attached DNA
- A liquid phase with the remaining of the applied reagents and the reaction medium

The liquid phase is removed and the magnetic beads are washed with water. Water dissolves the DNA from the magnetic beads – the solution still has 2 phases:

- A solid phase with the magnetic beads
- A liquid phase with DNA dissolved in water

The liquid phase is transferred into a new tube – the magnetic beads are discarded.

Figure 13: Magnetic rack with the samples: solid phase with the magnetic beads are attracted to the magnetic rack – they are visible as brown spots. The liquid phase can be removed without disturbing the magnetic beads. (18)

5. Measure DNA concentration

Applied kit: Qubit dsDNA HS Assay Kit

Device: Qubit 2.0 Fluorimeter

The device measures the DNA concentration (ng/mL).

6. End repair

Kit: Ion Plus Fragment Library Kit

During the multiplication and purification of the DNA so called “sticky ends” appeared. DNA has a double helix form, the sticky ends are the ends of the DNA which are single stranded. For the further steps (attach adapters), these ends have to be “repaired“, meaning to fill up the missing nucleotides to make the complete DNA double stranded – see Figure 14.

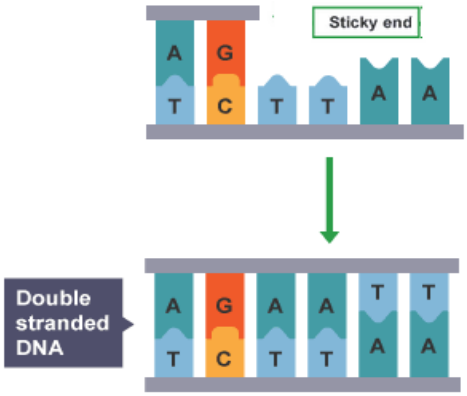


Figure 14: The end repair process: covering the sticky ends of the DNA (after 19)

7. Purification

Device: Magnetic rack

The same purification method with the magnetic beads as described above – the only difference that the DNA is not dissolved from the magnetic beads by water but in Low TE.

8. Labelling the samples

Used kits: Ion Xpress Barcode Adapters 1-16 Kit, Ion Plus Fragment Library Kit

Devices: PCR

During the sequencing procedure the samples are mixed. To know which bacterial DNA is originating from which sample, the DNA is barcoded. It means that to every piece of DNA a barcode will be added – see Figure 15. The barcode itself is a special sequence of nucleotides which will be recognized during the sequencing. Every sample has different barcodes. Next to the barcodes, two kind of adapters are also added to the sample: one adapter will bind the DNA to an Ion Sphere Particle (Step 11) and a primer will bind to the other adapter.

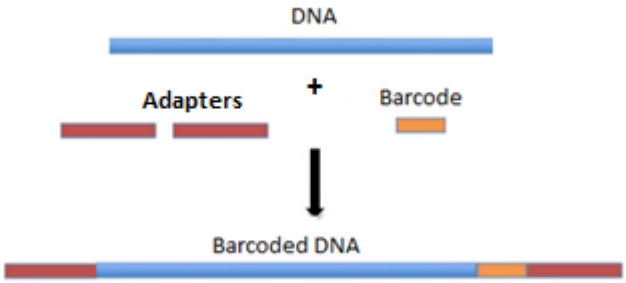


Figure 15: Labelling the samples with barcodes. (after 20)

9. Purification

Device: Magnetic rack

The same purification method with the magnetic beads as described above –the DNA is dissolved from the magnetic beads in Low TE.

10. Measure DNA concentration

Used kit: Ion Universal Library Quantitation Kit

Device: real time PCR (qPCR)

The molar concentration of the DNA has to be determined. Fluorescent probes are added to the DNA solution which binds to the DNA. The fluorescence of the DNA is measured after every DNA multiplication cycles in the qPCR. Based on the fluorescence of the DNA its concentration is determined.

11. Adding Ion Sphere Particles (ISPs)

Used kits: Ion PGM Template OT2 Supplies 400, Ion PGM Template OT2 Solutions 400, Ion PGM Template OT2 Reagents 400

Device: Ion One Touch 2

During the sequencing, the DNA is loaded into a microchip. To bind the DNA to the microchip a carrier is necessary. These carriers are the Ion Sphere Particles (ISPs) which bind both to the DNA and to the microchip – see Figure16. In this way every piece of DNA will have a fixed position which is necessary in order to sequence them. This process was done in the Emulsion PCR device. Many pieces of DNA can bind to one ISP, however in ideal situation only one piece of DNA is attached. If any DNA bound to an ISP it's called a templated ISP.

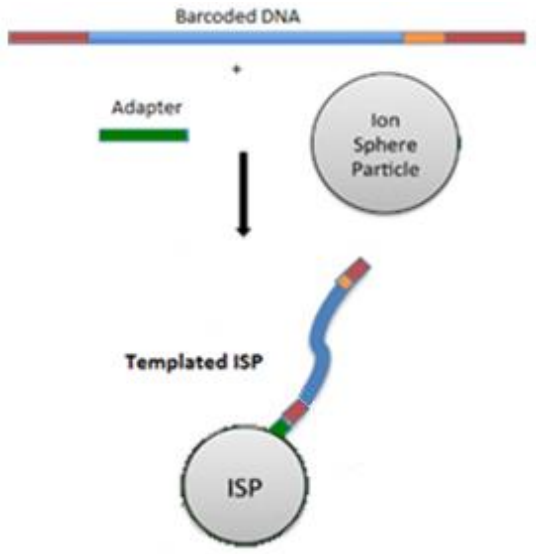


Figure 16: Ion Sphere Particles as carriers of DNA (after 20)

12. Enrichment of templated ISPs

Used kits: Ion PGM Template OT2 Supplies 400, Ion PGM Template OT2 Solutions 400, Ion PGM Enrichment Beads

Device: Ion One Touch ES

To some ISP didn't attach any DNA. The surplus of the ISPs have to be removed from the solution, to leave behind only the ISPs with attached DNA.

13. Sequencing

Used kits: Ion PGM Sequencing Supplies 400, Ion PGM Sequencing Reagents 400, Ion PGM Sequencing Solutions 400, Ion 314™ Chip Kit v2,

Device: Ion Torrent PGM

The DNA sample is loaded into a semiconductor microchip – see Figure 17. The sample will fill the micro wells on the chip's surface. During the research, Ion 314 Chip was used which has 1,2 million micro wells.

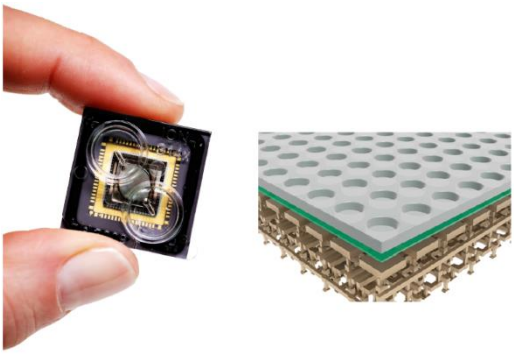
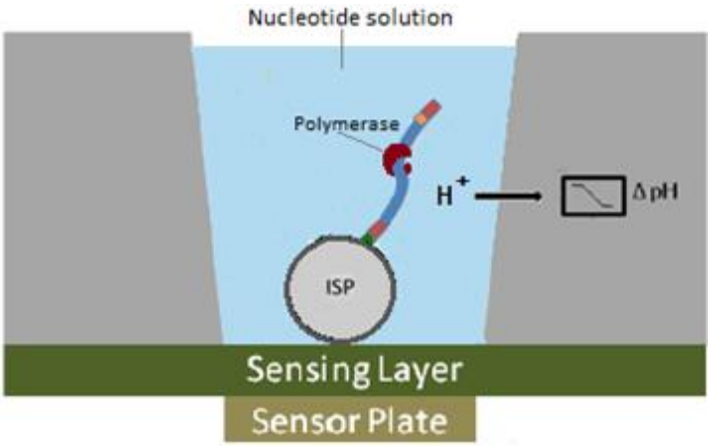


Figure 17: The Ion 314 chip and its surface with the wells (21)

In this way every microcell contains one templated ISP – see Figure 18. The double stranded DNA is "unzipped" and a polymerase is attached to it –see Figure 8. To provide nucleotides for the polymerase, the micro wells are filled up with nucleotide (A, T, G, and C) solutions: in every 15 second another nucleotide solution. As soon as the polymerase incorporates a nucleotide from the solution, the pH in the well changes.

Figure 18: Templated ISP in the microchip's well (after 21)



For example the polymerase has to incorporate a T nucleotide:

- If the ISP is in a G nucleotide solution → it's not a match, therefore nothing happens
- If the ISP is in a T nucleotide solution → it's a match, therefore H^+ is released and the pH change is indicated

The pH change is measured by a sensing layer. Based on the pH change the computer registers which nucleotide is incorporated. As the primers moves along the DNA and incorporates the nucleotides, the

computer builds up the sequence. These sequences from different DNA pieces are aligned by the computer program which finds the same nucleotide sequence within the different DNA pieces and uses them as matching points for the alignments – see Figure 19 as an illustration.

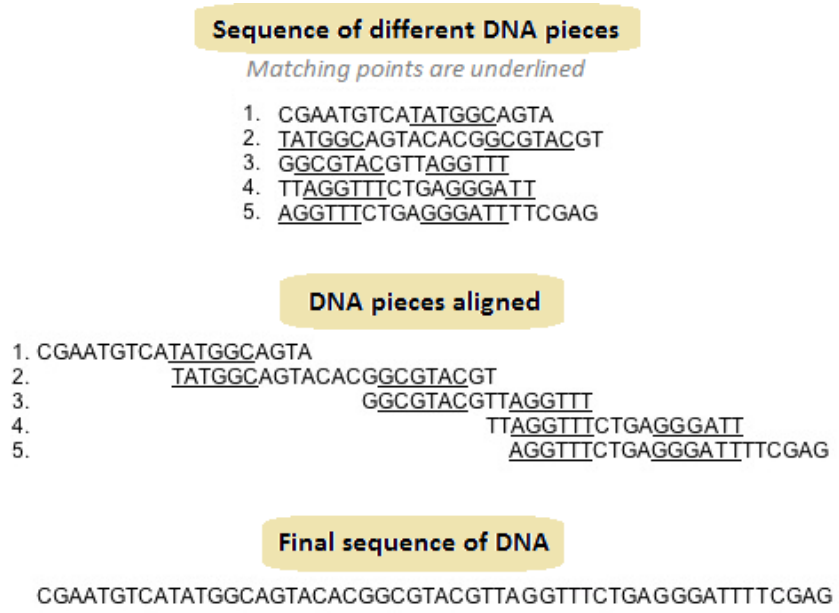


Figure 19: Alignment of the DNA to build up the final sequence (after 22)

After the alignment, the final sequence of the DNA is compared to the sequences found in the "MicroSEQ" and "Greengenes" DNA databases.

The identification of species follows the guidelines of Clinical and Laboratory Standard Institute which requires the following identities:

- ***Family-level:*** <97% identity
- ***Genus-level:*** >97% identity
- ***Species-level:*** >99% identity

Ion Torrent company claims that the accuracy of the sequencing is >99% (23). Although the chances of mistakes are small, the most common mistakes which are made by this technology are:

- ***Substitution***: wrong nucleotide is incorporated
- ***Insertion:*** an extra base pair is incorporated
- ***Deletions:*** the nucleotide is not incorporated

The accuracy of each run (for each microchip) is detected by loading so called Test Fragments together with the samples into the microchip. The Test Fragments are build up from nucleotides (<100 base pair) and they are attached to an ISP just like the DNA from the samples. The exact sequence of the Test Fragments are known, therefore they are suitable for testing the accuracy of the method.

III.3. Antibiotic resistance test

The samples were treated by the technician staff in the Medical Microbiology Lab in the UMCG. The workflow was the following:

1. 100 µL of each sample was applied on blood-agar plates and the species were cultivated. The results was a mixture of bacterial species in the plates.
2. The clinically important bacterial species (*Enterococcus sp.*) from the plates were identified by MALDI-TOF technique.
3. The clinically important species were cultivated separately. The results was one bacterial species per plate.
4. Finally antibiotic resistance test was carried out on them:
    - ***Disk-diffusion technique***: on *Enterococcus aquimarinus*
    - ***VITEK technology***: on *Enterococcus faecium and Enterococcus faecalis*

Four antibiotics were chosen for the test – these antibiotics are used the most often in the case of an *Enterococcus sp.* infection. Different amount of antibiotics were applied by the two technology, Table 3 lists the antibiotics and the applied amounts.

Antibiotic	Applied amounts	
	Disk diffusion	Vitek
Ampicilin	2 µg	2 µg/mL
Gentamicin	30 µg	500 µg/mL
Nitrofurantoin	100 µg	16 µg/mL
Vancomycin	5 µg	1 µg/mL

Table 3: Antibiotics and their applied amounts

IV. RESULTS AND DISCUSSION

Every test that was carried out helped to answer the research questions. Therefore this chapter discusses the research questions (marked with Q), the results of the related tests and the answers for the research questions (marked as A).
The results of the tests 1-4 are a bacterial map what shows the species which were detected. The result of test 5 is a list which indicates that the clinically important *Enterococcus species* are resistant to certain antibiotics or not.

Regarding tests 1-4 information about the accuracy of the runs are compared. Accuracy tests are carried out with Test Fragments which consist of <100 base pairs. The accuracy is measured by the 50AQ17 value (accuracy value) which was measured during every test. These accuracy values are mentioned together with the related bacterial maps, so this value indicates the reliability of the bacterial species' detection.

The definition of the used terms and the calculation of the accuracy value are mentioned below (24):

Q17 : This data gives the length of the DNA (in base pairs) which is sequenced with 2% error – i.e. it allows 1 base error per 50 bases. Thus, for a given read the Q17 length would be one base before the second error. (25)

50AQ17: Test fragments with one or zero errors in the first 50 bases read.

Calculation:

50AQ17 (%) = (number of test fragments with one or zero errors in the first 50 bases read / total number of test fragments)*100

IV.1. Set up of the methodology

Q: How is the performance of 16S NGS on cultivated bacterial samples?

Test 1: E.coli, S. aureus

This test was carried out twice. The first time the result showed Family or Genus-level identification. Therefore, the test was repeated and it showed improved results – Species-level identification. The accuracy values of the tests were representatively 70% and 93%. The higher the accuracy value the more precise the identification of species - Figure 20 shows the results of the test.

Test	Sample	Phylum	Class	Order	Family	Genus	Species
1 70%	Control *E.coli*	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae		
	E.coli	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae		
	S.aureus	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	
1 repeated 93%	Control *E.coli*	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Esherichia	coli
						Raoultella	
						Samlonella	enterica
	E.coli	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Citrobacter	
						Esherichia	coli
							vulnaris
						Raoultella	
						Samlonella	enterica
						Shigella	
						Bacterium	
	S.aureus	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphilococcus	aureus
							epidermis

Figure 20: Bacterial map of Test 1 with the accuracy values (%)

When the test was repeated, the results showed the expected species. Next to the expected species, it also indicated other species which is theoretically not possible, since only one species was applied per sample. The presence of more than one species can be excluded because:

- the origin of the bacteria is controlled
- the cultivation of the species was done with sterile equipments

Furthermore, if the sample would be contaminated by bacterial species during the 4 days of sample-preparation, the DNA of the contaminant bacteria is still not available for sequencing. The bacterial DNA has to be extracted from the cell (see Chapter III.2. Extraction of DNA) before sequencing.

Therefore, the results can be explained by the high similarity between the indicated species' genetic material.

A: ***This test proved that the method is able to detect the cultured bacterial species. Since it also indicated more species that was actually in the samples, the performance was rather an indication of possibilities than an exact detection.***

Q: How is the accuracy of 16S NGS influenced by the number of bacterial species present in the cultivated bacterial samples?

Test 2: *Mixture – E. coli, S. aureus, Klebsiella pneumoniae, Pseudomonas aeruginosa, and Enterococcus faecium*

This test was carried out also twice with the same accuracy values (96%). In both case the results showed Species-level identification for all of the 5 bacteria what were in the mixture – see Figure 21.

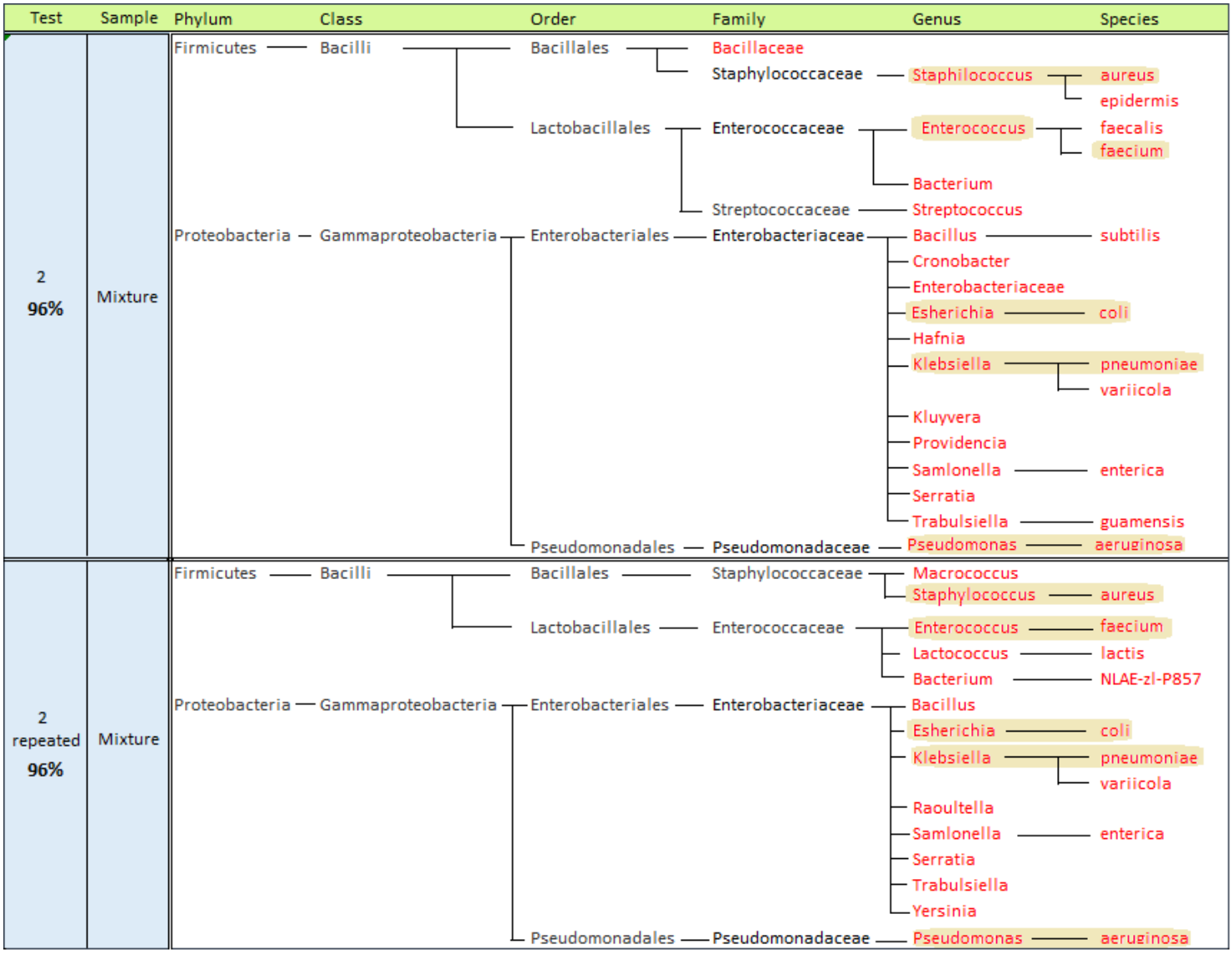


Figure 21: Bacterial map of Test 2 with the accuracy values (%)

A: ***This test proved that the method is able to detect the species from mixed bacterial cultures. Since it also indicated more species that was actually in the samples, the performance was rather an indication of possibilities than an exact detection.***

Q: How is the performance of 16S NGS compared to conventional Sanger-sequencing?

Test 3: Clinical – with S.aureus

Four patient-related clinical DNA extract was sequenced. The samples were sequenced before with Sanger-sequencing. Based on that results it was expected that two of the sample contained *S.aureus* DNA and the other two samples don't have any bacterial DNA. After the Ion Torrent sequencing the four samples showed slightly different result than it was expected – see Figure 22. The samples were containing the followings:

- One sample with not any detectable bacterial DNA → ✓ not any detectable bacterial DNA was expected
- One sample with *Serratia quinivorans* → ✗ not any detectable bacterial DNA was expected
- One sample with *S. aureus* → ✓ S. Aureus was expected
- One sample with *S. epidermis* and 5 other species → ✗ S. Aureus was expected

Test	Sample	Phylum	Class	Order	Family	Genus	Species
3 **96%**	Clinical 1	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	
		Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Serratia	quinivorans
	Clinical 2	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae		
		Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	aureus
		Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae		
			Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Serratia	quinivorans
	Clinical 3	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	graminisolvens
			Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	
		Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	epidermis
				Lactobacillales	Carnobacteriaceae	Trichococcus	
					Enterococcaceae	Enterococcus	aquimarinus
			Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae		
		Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Bacterium	SUTW
				Enterobacteriales	Enterobacteriaceae	Serratia	quinivorans
				Pseudomonadales	Pseudomonadaceae	Pseudomonas	caeni

Figure 22: Bacterial map of Test 3 with the accuracy value (%)

The results of the 2 different types of sequencing is comparable, although it seems the conventional Sanger-sequencing is less reliable in detecting the bacterial species. The difference in the detection sensitivity can be explained by the amount of V regions what is sequenced within the 16S gene. The conventional method sequences V3 region while Ion Torrent sequences the V2, V3, V4, V6, V7, V8 and V9 regions. The larger part of the DNA is sequenced the bigger the more sensitive the method in detection.

A: *The sensitivity to detect bacterial species and the results are different of the two method: Ion Torrent technology is more sensitive to detect the species, because it sequences larger part of the DNA. Therefore Ion Torrent technology is favorable regarding the detection limit of the species.*

IV.2. Phase I. – 16S NGS on water samples

Q: How is the performance of 16S NGS on water samples?

Test 4: Water samples

The results of this test show a variety of bacteria. Many times the identification was not possible at Species-level, but on Family or Genus-level – see Figure 23-26. The identification at Species-level requires 99% identity of the DNA with the database's record, while at Genus-level it's >97% and at Family-level it's <97%. Therefore there is a difference in the reliability of the identification at different levels.
In spite of the 93% accuracy of the test, *E.coli* was not detected at Species-level in none of the sample. *E.coli* is an indicator organism for faecal contamination in the water, therefore its presence is highly expected in wastewater.
Regarding the clinically important species, *Enterococcus aquimarinus* was detected in two of the samples: black water from vacuum toilets and WWTP influent in Sneek. *E. Aquimarinus* belongs to the genus of *Enterococcus* which is clinically important. The diseases which are caused by *Enterococcus* species are: urinary tract infections, bacteremia, peritonitis and infective endocarditis (26).

A: ***The performance of Ion Torrent technology on bacterial DNA originating from water samples is comparable with the results of the tests which were carried out with cultivated bacterial species. Identification of species still has to be optimized, because many times it was possible only at Family or Genus-level.***

Test	Accuracy	Sample
4	95%	Vacuum toilets, Sneek

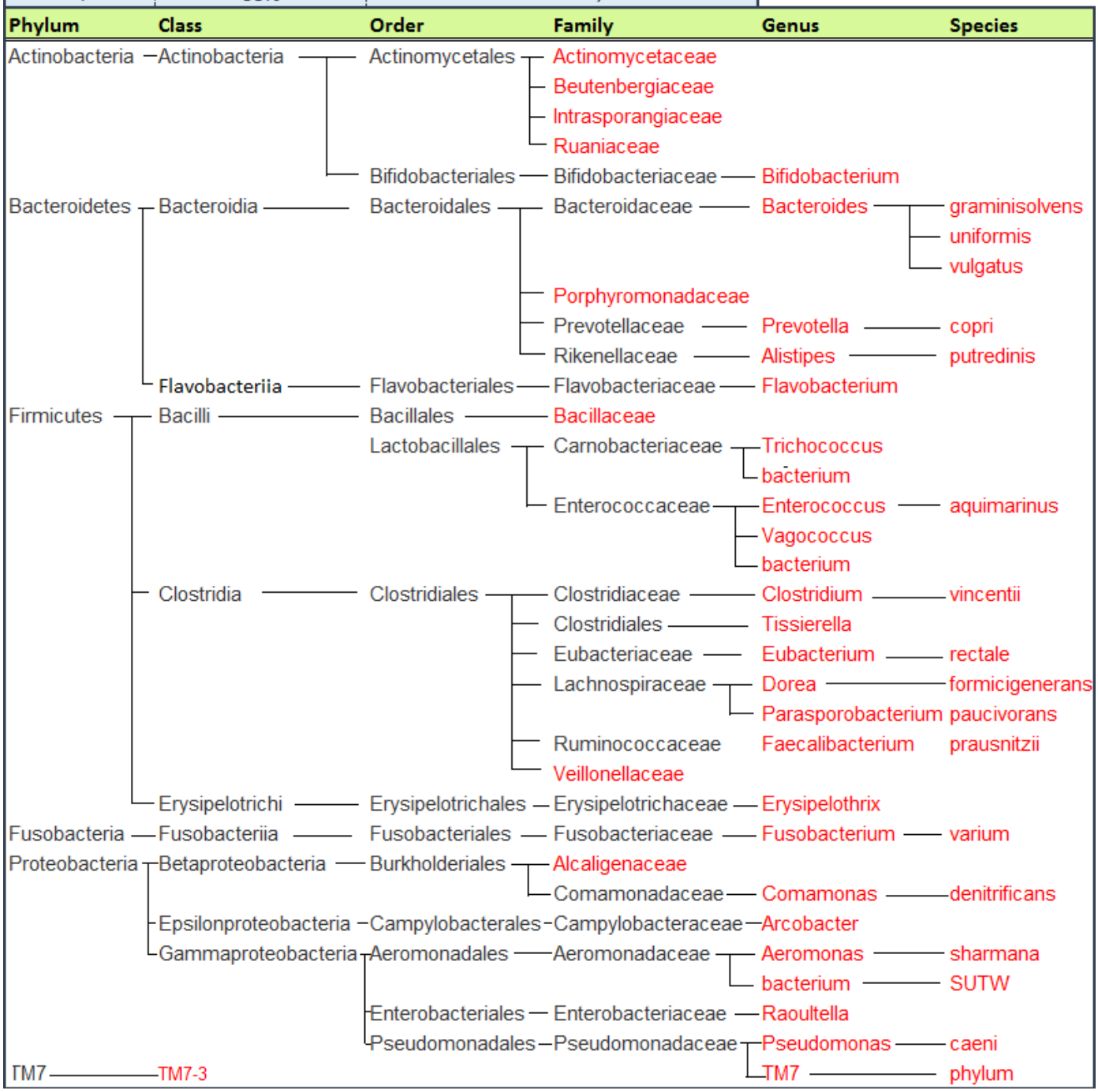


Figure 23: Bacterial map of Test 4 with the accuracy value (%) – black water sample from vacuum toilets, Sneek

Test	Accuracy	Sample
4	95%	WWTP influent, Sneek

Phylum	Class	Order	Family	Genus	Species
Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae	Collinsella	aerofaciens
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	dorei
					graminisolvens
					vulgatus
			Porphyromonadaceae		
			Prevotellaceae	Prevotella	copri
	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	granuli
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Trichococcus	
			Enterococcaceae	Enterococcus	aquimarinus
	Clostridia	Clostridiales	Eubacteriaceae	Eubacterium	rectale
			Lachnospiraceae		
			Peptostreptococcaceae		
			Ruminococcaceae	Faecalibacterium	prausnitzii
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae		
			Leptotrichiaceae		
Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Gemmiger	formicilis
	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	
				Comamonas	denitrificans
				Macromonas	bipunctata
				Simplicispira	metamorpha
				Variovorax	paradoxus
			Oxalobacteraceae		
		Neisseriales	Neisseriaceae		
		Rhodocyclales	Rhodocyclaceae	Zoogloea	ramigera
	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Arcobacter	cryaerophilus
					ellisii
					venerupis
	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas	salmonicida
					sharmana
				Tolumonas	auensis
					osonensis
				bacterium	SUTW
		Alteromonadales	Shewanellaceae	Shewanella	
		Enterobacteriales	Enterobacteriaceae		
		Oceanospirillales	Endozoicimonaceae		
		Pseudomonadales	Moraxellaceae	Acinetobacter	johnsonii
			Pseudomonadaceae	Pseudomonas	caeni

Figure 24: Bacterial map of Test 4 with the accuracy value (%) – waste water sample from WWTP influent, Sneek

Test	Accuracy	Sample
4	95%	WWTP influent, Grou

Phylum	Class	Order	Family	Genus	Species
Actinobacteria	Acidimicrobiia	Acidimicrobiales	Acidimicrobiales	Candidatus Microthrix	parvicella
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	graminisolvens
			Porphyromonadaceae		
			Prevotellaceae	Prevotella	paludivivens
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	
		Lactobacillales	Carnobacteriaceae	Trichococcus	
			Streptococcaceae	Lactococcus	chungangensis
					raffinolactis
	Clostridia	Clostridiales	Veillonellaceae		
				Proteocatella	sphenisci
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae		
			Leptotrichiaceae		
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae		
		Rhodocyclales	Rhodocyclaceae	Zoogloea	
			beta	proteobacterium	
	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Arcobacter	cryaerophilus
					venerupis
	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas	salmonicida
					sharmana
				Tolumonas	auensis
					osonensis
				bacterium	SUTW
		Alteromonadales	Shewanellaceae	Shewanella	
		Enterobacteriales	Enterobacteriaceae	Yersinia	
		Pseudomonadales	Moraxellaceae	Acinetobacter	
		Thiotrichales	Thiotrichaceae	Thiothrix	eikelboomii

Figure 25: Bacterial map of Test 4 with the accuracy value (%) – waste water sample from WWTP influent, Grou

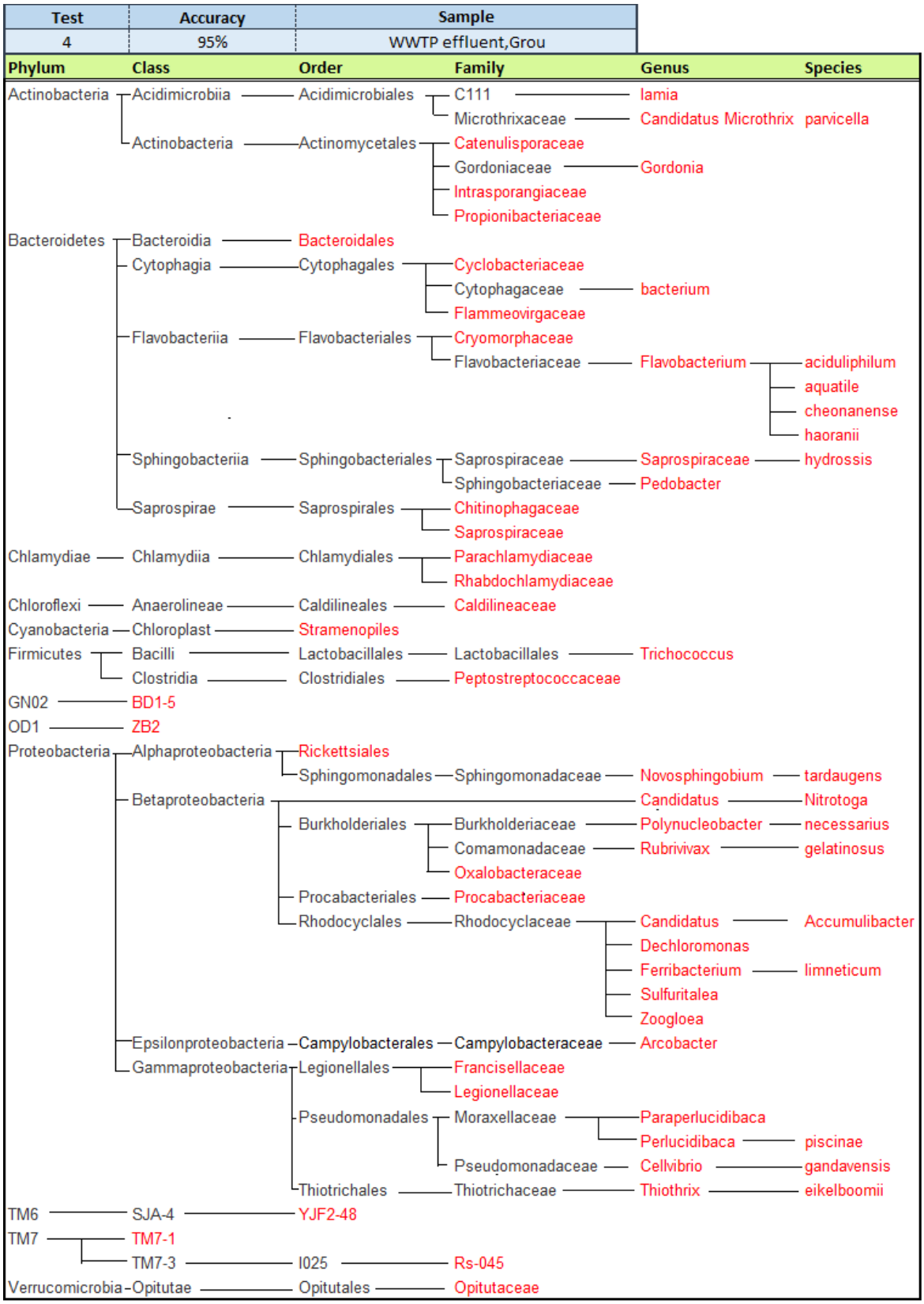


Figure 26: Bacterial map of Test 4 with the accuracy value (%) – treated waste water sample from WWTP effluent, Grou

Q: What kind of clinically important bacterial species are detected in the water samples?

Test 4: Water samples

The list of identified bacterial species, genera and families were compared with the pathogenic bacteria database (26). A small number of indicated bacteria found to be clinically important. The clinically important species, genera or families are listed below and summarized in Table 4.

The list mentions the following data:

→ Family of the clinically important bacteria
→ At which level the bacteria was indicated: species, genus or family
→ In which samples it was indicated
→ Clinically important species from the same family or genus

CAMPYLOBACTERACEAE

Indicated: Clinically important species = *Arcobacter cryaerophilus*

In the following samples:

- **Vacuum toilet, Sneek:** *Arcobacter cryaerophilus*, species-level
- **WWTP influent, Sneek:** *Arcobacter cryaerophilus*, species- level
- **WWTP influent, Grou:** *Arcobacter*, genus-level
- **WWTP effluent, Grou:** *Arcobacter*, genus-level

CLOSTRIDIACEAE

Indicated: Clinically important genus = *Clostridium*
Clinically not important species = *Clostridium vincentii*

In the following sample:

- **Vacuum toilet, Sneek:** *Clostridium vincentii, species-level*

The genus includes important pathogens like *C. tetani, C. perfringens, C. botulinum*and *C. difficile*.

ENTEROBACTERIACEAE

Indicated: Clinically important genus = *Yersinia*
No identified genus or species

In the following sample:

- **WWTP influent, Grou:** *Yersinia*, genus-level

The genus includes important pathogens like *Y. pestis, Y. pseudotuberculosis* and *Y. enterocolitica.*

ENTEROCOCCACEAE

Indicated: Clinically important genus = *Enterococcus*
Clinically not important species = *Enterococcus aquimarinus*

In the following samples:

- **Vacuum toilet, Sneek:** *Enterococcus aquimarinus*, species-level
- **WWTP influent, Sneek:** *Enterococcus aquimarinus*, species-level

The genus includes important pathogens like *E.facalis and E.faecium*.

LEGIONELLACEAE

Indicated: Clinically important family = *Legionellaceae*
No identified genus or species

In the following sample:

- **WWTP effluent, Grou:** *Legionellaceae*, family-level

The family includes important pathogens like *L. pneumophila.*

NEISSERIACEAE

Indicated: Clinically important family = *Neisseriaceae*
No identified genus or species

In the following samples:

- **WWTP influent, Sneek:** *Neisseriaceae*, family-level

The family includes important pathogens like *N. gonorrhoeae* and *N. meningitides.*

PARACHLAMYDIACEAE

Indicated: Clinically important family = *Parachlamidiaceae*
No identified genus or species

In the following samples:

- **WWTP effluent, Grou:** *Parachlamidiaceae*, family-level

The family includes important pathogen like *P. acanthamoebae*

PSEUDOMONACEAE

Indicated: Clinically important genus = *Pseudomonas*
Clinically not important species = *Pseudomonas caeni*

In the following samples:

- **Vacuum toilet, Sneek:** *Pseudomonas caeni*, species-level
- **WWTP influent, Sneek:** *Pseudomonas caeni*, species- level

The genus includes important pathogen like *P. aeruginosa*.

STAPHYLOCOCCACEAE

Indicated: Clinically important genus = *Staphylococcus*
No identified genus or species

In the following samples:

- **WWTP influent, Grou:** *Stapylococcus*, genus-level

The genus includes important pathogens like *S. aureus*.

Family	Vacuum toilets, Sneek			WWTP influent, Sneek			WWTP influent, Grou			WWTP effluent, Grou		
	F	G	S	F	G	S	F	G	S	F	G	S
Campylobacteraceae			✓			✓		✓			✓	
Clostridiaceae		✓										
Enterobacteriaceae								✓				
Enterococcaceae		✓			✓							
Legionellaceae										✓		
Neisseriaceae							✓					
Parachlamydiaceae										✓		
Pseudomonaceae		✓			✓							
Staphylococcaceae								✓				

Table 4: Indication of clinically important bacteria at different levels in the water samples.
The different levels are marked as **F** – family, **G** – genus, **S** – species.

A: *Arcobacter cryaerophilus* species; *Clostridium*, *Yersinia*, *Enterococcus*, *Pseudomonas* and *Staphylococcus* genera; *Legionellaceae, Neisseriaceae* and *Parachlamidiaceae* families are the clinically important bacteria in the water samples. Species were also indicated from the mentioned genera, however they were not clinically important. The recognition of the species by 16S NGS still has to be improved, therefore it needs further investigation to identify which species are present in the water samples from the mentioned bacteria genera and –families.

The most common bacteria from the list is the *Enterococcus*.

IV.4. Phase II. – Antibiotic resistance test on clinically important species

Q: Which water source contains ABR what are clinically important?

Test 5: Antibiotic resistance

The 16S NGS indicated a species from the genus *Enterococcus*, namely *Enterococcus aquimarinus*. Since *Enterococcus* species are clinically important, it was cultivated in Petri-disks from the samples. Although *E. aquimarinus* was indicated by sequencing, other Enterococcus species (*E. faecalis* and *E. faecium*) were also growing in the Petri-disks:

- Sample from vacuum toilets, Sneek:
  E. aquimarinus was indicated by NGS → ✓ *E.aquimarinus* was growing in Petri-disks

- Sample from WWTP influent, Sneek:
  E. aquimarinus was indicated by NGS → ✗ *E. faecalis* and *E. faecium* were growing in Petri-disks

The species from the Petri-disks were identified by MALDI-TOF technology. The difference between the indicated species by 16S NGS and MALDI-TOF can be explained by the followings:

- The genetic material of *E. aquimarinus* is highly similar to the genetic materials of *E. faecalis* and *E. faecium.*
- *E. faecalis* and *E. faecium* couldn't be indicated by 16S NGS because they were present in the samples below the detection limit.

The antibiotic resistance test was done with the species which were growing in the Petri-disks. The results of the antibiotic resistance tests detected not any resistance for the applied antibiotics in none of the samples. However, based on the results it can't be stated that there is not any antibiotic resistant bacteria in the samples: only *Enterococcus sp.* were tested from the many detected species. If other bacterial species are ABR than their resistant gene can be incorporated via horizontal gene transfer into the non-resistant *Enterococcus species*. This induction would transmute the Enterococcus species into ABR.

A: None of the water source contained clinically important species (regarding *Enterococcus* species) which were resistant for 4 types of antibiotics at the given time.

V. CONCLUSIONS AND RECOMMENDATIONS

Based on the tests which were carried out, the followings can be concluded:

- 16S NGS always indicated the species which were tested during Test 1-3.
- In certain cases more species was indicated than it was possible. This is due to the high similarity between the 16S sequence of the given species.
- 16S NGS has a higher sensitivity than the conventional Sanger-sequencing, because it sequences larger part of the 16S gene.
- The preparation of the genetic material for sequencing needs practice. For example, to load the microchip with the prepared 16S DNA is a very delicate work. It depends on the loading of the microchip if the sample can be sequenced – see Figure 27 for an illustration.

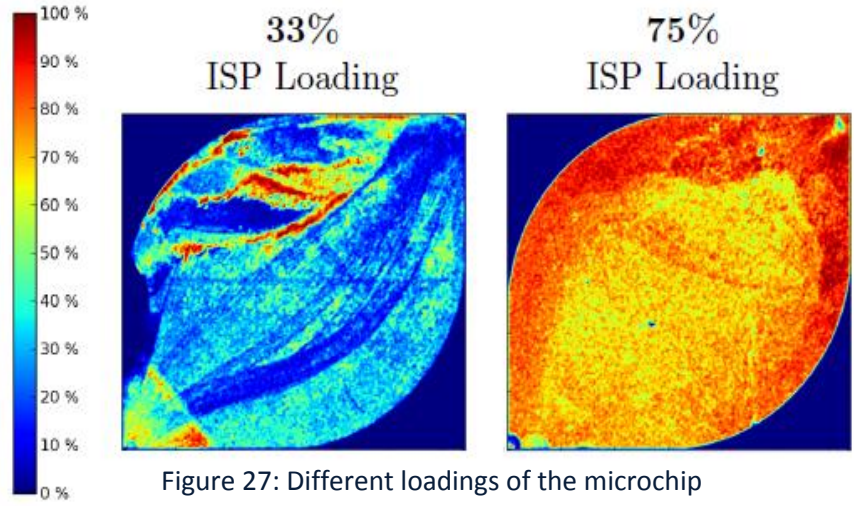


Figure 27: Different loadings of the microchip

- To prepare and sequence the bacterial DNA, it needs a variety of devices and commercial kits, therefore the work can be carried out in well-equipped laboratory or it needs an investment. The list of necessary instruments are the followings:

  Devices: centrifuge, vortex adaptor, Nanodrop 2000 Spectrophotometer, PCR, magnetic rack, Qubit 2.0 Fluorimeter, qPCR, Ion One Touch 2 emulsion PCR, Ion One Touch ES, Ion Torrent PGM sequencing device

  Kits: 15 different kits (mentioned in the chapter II.2.)

  Other: pipettes in the µL range, 200 µL PCR- and 1,5 mL tubes, Nuclease-free and ultrapure water

As a final conclusion the 16S NGS method is applicable on water samples to identify their complex bacterial composition. However, we noticed that a given piece of DNA what was prepared for sequencing, can be classified as more than one species. This is due to the high similarities of the genetic material between different species. The given piece of DNA what was sequenced is identical in these species. Therefore the list of identified species is rather an indication of possibilities, than a strict diagnosis. Furthermore, the water samples contained a vast amount of bacteria and some of them were clinically important. From the clinically important bacteria *Enterococcus* was tested for antibiotic resistance. The result of the antibiotic resistance test shows that *Enterococcus* was not resistant for 4 different types of antibiotics.

To further enhance the species identification and to estimate the limits of the method, the followings are recommended:

- Detect the minimal amount of input DNA which is necessary for sequencing.
  The abundance of different species in the water samples differ. For example one species can be found in vast amount compared to another species which are present only in very small amount. It is recommended to estimate the minimum abundance of species which can be identified by 16S NGS.

- Adjust the identification of species with genetically similar species (e.g. *S. aureus* and *S. epidermis*).
  So far, the list of identified bacterial species is rather an indication of possibilities than an exact diagnosis. The identification of species can be done by using different setting during the sequencing. A variety of adjustments can be done via the Ion Reporter software. Using genetically similar species during these test would help to detect the most suitable setting for the sequencing.

- Optimize sample preparation.
  - → *To have longer sequences from a given piece of DNA would help the identification of species.*
  - → *To save time and expenses the qPCR process (Step 10) could be replaced by Qubit. To do so, a conversion factor has to be calculated: qPCR measures DNA concentration in pmol/L while Qubit measures it in ng/mL. To do the conversion, the molar mass of the DNA has to be known which depends on the number of its base pairs. Since this is not exactly known, the conversion factor has to be based on statistical calculations.*

The sequencing of bacterial 16S gene by Ion Torrent technology is promising. The origin of bacterial DNA (from bacterial cultures, DNA extracts or from water samples) was not influencing the ability of the device to identify the species. However, the technology is new (released in 2014) and certain adjustments has to be made in the sample preparation as well with the setting of the devices.

Ion Torrent Systems Inc. showed interest in this presented study. According to the company, our research group has the most experience with the technology in the Netherlands.

VI. REFERENCES

(1) **CDC US, 2013** *About One Health*
Available: http://www.cdc.gov/onehealth/about.html (Accessed: 22nd June 2015)

(2) **Martinez J L et al., 2009** *A global view of antibiotic resistance*, FEMS Microbiology Reviews, Vol.33 (1), pp. 46-65

(3) **WHO, 2015** *Antimicrobial resistance,* Fact sheet N°194

(4) **Wade W, 2002** *Unculturable bacteria—the uncharacterized organisms that cause oral infections*, Journal of the Royal Society of Medicine, Vol. 95(2), pp. 81–83

(5) **Nature**, *Cells Can Replicate Their DNA Precisely*
Available: http://www.nature.com/scitable/topicpage/cells-can-replicate-their-dna-precisely-6524830 (Accessed: 22nd June 2015)

(6) **Janda J M et al., 2007** *16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls*, Journal of Clinical Microbiology Vol. 45 (9), pp. 2761-2764

(7) **Gray N F, 2005** *Water Technology*, Elsevier Science & Technology Books, pp. 309-344

(8) **WHO, 2014** *Antimicrobial resistance Global Report on Surveillance*

(9) **Marti E et al., 2014** *The role of aquatic ecosystems as reservoirs of antibiotic resistance*, Trends in Microbiology, Vol. 22, pp. 36-41

(10) **CDC US, 2013** Antibiotic resistance threats in the United States, 2013

(11) **Cantón R et al., 2013** *Inappropriate use of antibiotics in hospitals: The complex relationship between antibiotic use and antimicrobial resistance*, Enfermedades Infecciosas Microbiología Clínica, Vol. 31, pp.3-11

(12) **WHO, 2004** *Guidelines for drinking-water quality*, 3rd edition, Vol. 1

(13) **Shah A D, *2008*** *Antibiotics in Water Treatment: The Role of Water Quality Conditions on their Fate and Removal during Chlorination and Nanofiltration*, Georgia Institute of Technology

(14) **Homem V et al., 2011** *Degradation and removal methods of antibiotics from aqueous matrices a review*, Journal of Environmental management Vol. 92 (10), pp. 2304-47

(15) **Yaza P et al., 2014** *Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences*, Nature Reviews Microbiology, Vol.12, pp. 635–645, Figure 1

(16) **Nature**, *Cells Can Replicate Their DNA Precisely*

Available: http://www.nature.com/scitable/topicpage/cells-can-replicate-their-dna-precisely-6524830 (Accessed: 22nd June 2015)

(17) New England Biolabs, *Master mixes*
Available: https://www.neb.com/products/pcr-polymerases-and-amplification-technologies/master-mixes/master-mixes (Accessed: 22nd June 2015)

(18) Diagenode, *DiaMag02 - magnetic rack*
Available: http://www.diagenode.com/en/catalog/diamag-magnetic-beads-more-84/diamag-magnetic-racks-86/product/diamag02-magnetic-rack-361 (Accessed: 22nd June 2015)

(19) BBC, *Recombinant DNA technology - Higher tier*
Available:http://www.bbc.co.uk/schools/gcsebitesize/science/triple_edexcel/using_biology/biotechnology/revision/5/ (Accessed: 22nd June 2015)

(20) Thermo Fisher Scientific, *Everything You Ever Wanted to Know About NGS but Were Afraid to Ask*
Available: http://www.slideshare.net/Lifetech_HID/everything-you-ever-wantedtoknowaboutngs (Accessed: 22nd June 2015)

(21) Life Technologies Corporation, 2011 *Ion Torrent Amplicon Sequencing*, Application note

(22) DNASTAR, *SeqMan NGen Help*
Available: https://www.dnastar.com/seqman_ngen_help/index.html#!Documents/mertags.htm (Accessed: 22nd June 2015)

(23) Life Technologies Corporation, 2011 *Ion Torrent - Ion Personal Genome Machine™ Performance Overview*, Application note

(24) Life Technologies Corporation, 2015 *Torrent Browser Analysis Report Guide*
Available: http://mendel.iontorrent.com/ion-docs/Run-Report-Metrics-Before-Alignment.html (Accessed: 22nd June 2015)

(25) BioLektures,2011 *Ion Torrent – Test Fragment Accuracy*
Available: https://biolektures.wordpress.com/2011/08/01/ion-torrent-test-fragment-accuracy-part-1/ (Accessed: 22nd June 2015)

(26) MOH Key Laboratory of Systems Biology of Pathogens, *Virulence Factors of Pathogenic Bacteria*
Available: http://www.mgc.ac.cn/VFs/main.htm (Accessed: 22nd June 2015)