universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Comparative genomics on three different Acanthamoeba strains and the optimisation of an isothermal amplification assay“

verfasst von / submitted by

Lukas Tegel, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von / Supervisor:	Ao. Univ.-Prof. Dipl.-Biol. Dr. Angela Witte

Statement of authorship

I hereby earnestly declare that this master´s thesis is entirely composed by myself and I have not used any other literature source than those listed in the references and mentioned in the appendices. Furthermore, I declare not to submit my thesis to any institutions to receive another degree.

_____ _____

(Place, Date) (Signature)

Acknowledgement

First of all, I would like to thank the AIT Austrian Institute of Technology for giving me the opportunity to work on the present master thesis in the field of modern research at the Health & Environment department. My particular thanks are dedicated to my supervisor Dr. Ivan Barisic, who heads the H2020 FAPIC project and greatly supported me on all research topics.

Furthermore, I thank all my valuable colleagues for their outstanding teamwork mentality, friendly talks and their tremendous support. I want to thank Elisa DeLlano and Michaela Hendling for helping me on bioinformatics challenges, especially Elisa, who contributed her great knowledge and expertise to the analysis of the NGS data. I highly appreciated all the support and great mindset from Noa Wolff with who I closely worked with in the laboratory on certain parts of this project.

Since the beginning of my work, I received great help from Silvia Schönthaler in the laboratory, who introduced me very well to various operating procedures and taught me the correct handling of individual lab instruments. Moreover, I thank Christa Noehammer very much for reviewing this written assignment.

In esteemed collaboration, I thank Prof. Julia Walochnik from the Medical University of Vienna for providing all *Acanthamoeba* DNA samples.

My sincere gratitude and appreciation is owed to my wonderful family and girlfriend Tamara who made it possible for me to follow my interests and, last but not least, my great circle of friends for their amazing support.

This study was thankfully funded by the European Union and supported by the H2020 FAPIC project under grant agreement number 634137.

Table of content

List of figures

List of tables

List of abbreviations

µg	Microgram
µl	Microliter
A	Adenine
Alu	*Arthrobacter luteus*
AK	Acanthamoeba keratitis
bp	Base pairs
Bst	*Bacillus stearothermophilus*
BLAST	Basic Local Alignment Search Tool
C	Cytosine
ClustalO	ClustalOmega
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide
E. coli	*Escherichia coli*
FU	Fluorescent unit
G	Guanine
GAE	Granulomatous amoebic encephalitis
Gb	Giga base pair
gDNA	Genomic DNA
IDT	Integrated DNA Technologies
ISP	Ion sphere particle
K. pneumoniae	*Klebsiella pneumonia*
LNC	Linear nucleotide chain
Mb	Mega base pair
MB	Molecular beacon
MDA	Multiple displacement amplification
NEB	New England Lab®
NCBI	National center for biotechnology
ng	Nanogram
NGS	Next generation sequencing
PCR	Polymerase chain reaction
PGM	Personal genome machine
pM	Picomolar (pmol/l)
POC	Point-of-care
RCA	Rolling circle amplification
ssp.	Species
T	Thymidine
TBE	Tris-Borat-EDTA

Key words: Next generation sequencing, Whole genome sequencing, Ion Torrent, *Acanthamoeba*, Isothermal amplification, MDA

1. Introduction

1.1. The H2020 FAPIC project - Fast assay for pathogen identification and characterization

The horizon 2020 FAPIC project focuses on the establishment of two cost-effective and economical diagnostic devices that facilitate the rapid identification and characterization of infectious-disease relevant pathogens. The project aims for the development of a portable low-cost Patho-Stick device, which allows the detection of a defined and limited number of pathogens. Moreover, a high-throughput Patho-Doc solution is envisaged, which will provide an automated test system for several hundreds of globally occurring microbes and is considered to be available in clinics and larger hospitals. Both devices are anticipated to allow the detection of pathogenic bacteria, fungi, protozoa, viruses and worms within several hours including sample preparation from human specimens such as blood, saliva, stool or urine. The pathogen detection is realized via a genomic DNA-based ultraplex assay. For the ultraplex assay, an isothermal amplification is applied to target specific sequence regions of microbial DNA without the need of thermo-cycling procedures. Basically, this technique combines a multiple displacement amplification and a rolling circle amplification to create multiple copies of low concentrated target-DNA with high accuracy and great specificity. Using this amplification method, molecular biomarkers including resistance genes, virulence factors and regulatory regions will be amplified applying distinct sets of primers. Furthermore, molecular beacons will allow DNA quantification via target-specific fluorescent probes. Both devices rely on a sensitive hybridization detection system for nucleic acids by using linear chain (LNC) oligonucleotides spotted onto a microarray slide. In the final step, target-specific labeled oligonucleotides will generate target-specific signals, which will be monitored applying suitable readout devices. Together with several other European partners, the different tasks of the H2020 FAPIC project will be accomplished in a five-year period.

During the last decades, microorganisms received various resistances due to the unconventional and uncontrolled use of antimicrobial substances. The highly promoted treatment with antibiotics in combination with vaccines forced microbes to evolve new defense mechanisms against the human immune system. In fact, lateral gene transfer had a major effect on the survival of pathogens by spreading DNA sequences continuously to different species. Spontaneous mutation events in essential gene regions have increased the

spectrum of resistance mechanisms. Accordingly, a rapid and accurate identification of disease-causing agents is crucial to specifically target the medical treatment to the actual pathogen causing infection. Especially in developing countries, where routine diagnostics are often fragile and limited, these devices should provide great help to immediately react on infectious diseases by detecting pathogens in an early stage. By today, only a few multiplex kits are available on the market to distinguish certain infectious disease, such as the PCR-based SeptiFast kit distributed by Roche. The SeptiFast MecA test enables the characterization of methicillin-resistant *Staphylococcus aureus* (MRSA), whereas, the SeptiFast test detects the 25 most common pathogens known to cause blood stream infections (sepsis). Furthermore, Fast Track Diagnostics for instance sells multiplex kits that are capable to determine diseases caused by viral species.

With respect to the grown status in the molecular infectious diseases diagnostics, the within the H2020 FAPIC project newly developed devices will be a major asset for the clinical diagnosis of infectious diseases in laboratories and biomedical institutions, since they will offer a fast, efficient and low-cost detection system for the entire repertoire of human pathogens.

1.2. Objectives

Inspired by the FAPIC project, the present master thesis deals with the genomic comparison of three different *Acanthamoeba* strains that potentially cause severe infectious diseases in humans and animals. Several genotypes of different *Acanthamoeba* species have been determined based on sequence dissimilarities in the 18S rRNA gene. Up to date, only little genomic information is available of this genus. As a consequence, genomes of the worldwide-distributed protist *Acanthamoeba* were investigated by whole genome shotgun sequencing (WGS) to decode their genetic profile. In detail, the genomes of three different Acanthamoeba strains were compared: *Acanthamoeba comandoni* strain Pb30/40, *Acanthamoeba lenticulata* strain 72/2 and *Acanthamoeba castellanii* strain 1BU that refer to different genotypes T4 (1BU), T5 (72/2) and T7 (Pb30/40) and various morphological groups (I, II and III). The genome mapping of strain 1BU to the *A. castellanii* Neff reference strain, both related to T4 genotype, resulted in 62% similarity based on high-quality reads obtained along Ion Torrent sequencing. For strain 1BU, certain housekeeping genes that express regulatory cytoskeleton proteins were matched to 98% for profilin and to 92% for

actin to the annotated Neff reference genome. Moreover, two important virulence factors, the mannose-binding protein (MBP) and serine alkaline protease show high sequence similarities for strain 1BU, but a rare sequence alignment for strain 72/2 and even no coverage for strain Pb30/40. In fact, broad diversity exists among distantly-related genotypes which highly correlates to the sequence dissimilarity (>5%) among the 18 rRNA gene.

The second objective of this master thesis was the development of an isothermal amplification assay to amplify extended spectrum β-lactamase (ESBL) genes from clinical isolates with multiple primers. For technical implementation, multiple displacement amplification was selected as method was implemented as it supports multiplexed reactions with great economy of time at low costs. The main advantage of a MDA reaction is that amplification occurs linearly and therefor provides a tool for precise DNA quantification. All clinical relevant pathogens that will be finally targeted for the FAPIC Patho-Doc device, applying among others MDA, are displayed in figure 1.

1.3. Phylogenic classification and symbiotic relationships of microorganisms

Microbial communities are disseminated around all environmental habitats and play a crucial role in maintaining an intact ecosystem. In nature, microorganisms exist in various biological niches depending on environmental factors such as temperature, salt concentration, pH-level, oxygen content and nutrimental sources. Human health requires bacterial symbionts indispensably as defensive barrier against foreign invaders. Bacterial phyla that occur most frequently in human body include *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria* as reviewed by Morgan [37]. In 2015, Maayan and researchers reported that commensal and mutual bacteria localizing at skin surface, digestive system, respiratory tract and genitourinary system play a remarkable role in host-gene expression and regulation of immune cells [34].

The phylogenetic classification of microbial communities relies on the universal 16S/18S small ribosomal subunit, which is a highly conserved but variable sequence among the three domains of life. 16S profiling with PCR allows the taxonomic differentiation of bacterial species, whereas only little information can be delivered for non-bacterial pathogens within this technique [37]. An *in vitro* cultivation is only feasible for a small number of existing microorganisms due to their extraordinary growth requirements. Metagenomics studies aim to amplify conserved ribosomal regions to announce novel species, if sequence similarity

occurs greater than 98,5%. Genomic “fingerprint” analyses and multiplex-PCR are useful tools for the categorization of microbial species. Moreover, chemotaxonomic methods for taxonomic classification involve the analysis of polar lipids, fatty acids composition, mycolic acids, and composition of lipopolysaccharides, peptidoglycan structure, sugar diversity, quinone systems, polyamines and protein patterns.

A collective and long-termed form of living between microorganisms and host organisms occurs everywhere in nature and is termed symbiosis. Intra- and extracellular interaction of unequal organisms is essential to resist natural selection. The main types of symbiotic relationship can be classified in mutualism, commensalism or parasitism. In mutualistic co-existence both organisms gain profit from nutrient feeding. In the case of a commensal association only one species benefits from this relationship without harming the host cell. Moreover, concerning parasitism, one organism achieves practical advantage and harms its associated host in addition. In human body, most abundant colonization of microbial communities appears in the gastrointestinal tract and provides nutrient digest, protection against pathogens and the synthesis of mucosal immune system. [6]

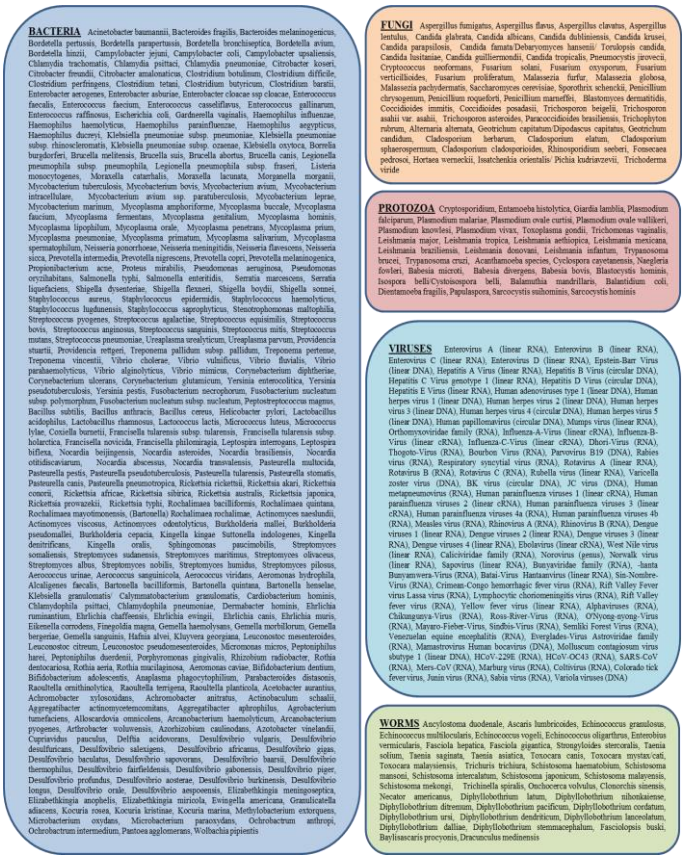


Figure 1. List of clinically relevant pathogens
All microorganisms (MO) that feature a potential threat for human health are represented in subdivided clusters These pathogens will be targeted within the newly developed devices.

1.4. Microbial resistances

Microbial resistances provide a powerful protection for microorganisms to prevent cell damage through antimicrobial substrates and chemicals. A massive spread of antimicrobial resistances has obtained the survival of pathogens. The uncontrolled use of antimicrobial supplements in health care and in the agricultural industry has boosted the rise of highly resistant microorganisms annually. Impervious microbes decrease the medical treatment efficiency tremendously. The establishment of new antimicrobial resistances develops rapidly as reported by the World Health Organization every year. (www.who.int/mediacentre/factsheets/fs194/en/).

The emerge of bacterial resistance mechanisms involves the modification of the membrane composition, overexpression of efflux pumps, overproduction of intrinsic antibodies, variation of certain enzymes and the alteration of the surface for attaching antimicrobial substances [53]. Lateral gene transfer (LGT), also termed horizontal gene transfer (HGF) occurs constantly, which results in which gene acquisition between single cell and multicellular organisms. In literature, different transmission routes describe the genetic exchange to other microorganisms via transduction, transformation, conjugation and virus-associated gene transfer for bacterial species. The relocation of plasmids can easily occur between bacteria species; probably most frequently via cell-cell mediated conjugation [41]. This transfer path of DNA pieces is one major route for the horizontal distribution of antibiotic resistance [18]. Resistance genes are predominantly encoded in plasmids, which are double-stranded, autonomous nucleic acids that occur frequently in all bacterial cells. Individual plasmid sizes vary between less than 2 kilobase pairs to many hundreds kilobase pairs. Concerning the plasmid geometry, most bacteria exhibit circular forms whereas some contain linear structures [8]. Moreover, efflux pumps activate the transportation of toxic molecules to the external space of the cell to overcome antimicrobial therapy [42]. Multidrug resistant bacteria normally obtain efflux pump mechanisms to prevent an intracellular uptake of drugs [53].

1.4.1. β-lactamases

During the evolutionary progress, β-lactam-mediated resistance evolved in Gram-negative and Gram-positive bacteria to prevent them from medical treatment with β-lactam antibiotics

as mentioned by K. Poole [42]. In former times, these antimicrobial substances were commonly used to treat infectious diseases. Due to the rise of numerous bacterial resistances in the past years, β-lactam antibiotics have lost their ability to combat pathogens. β-lactam antibiotics are structurally characterized by a four-membered β-lactam ring to be able to target the penicillin-binding protein (PBP), which is a major component required for the synthesis of bacterial cell wall. Functional β-lactamases open up amide bonds of the ring structure and force inhibition of the antibiotic effect. The molecular classification of β-lactamases is organized in four groups, A, B, C and D, whereas group A and D include enzymes assigned to extended-spectrum β-lactamase (ESBL). ESBLs ensure broader resistance mediated by various point mutations in resistance genes. Since the early 1980s, ESBL resistances constitute a rising threat for the health care sector by providing resistance to penicillin, second generation cephalosporins and third generation oxyimino cephalosporins. Predominantly, broad-spectrum hydrolyzing enzymes are found in subgroups of *Enterobacteriacaea*, especially in *E. coli* and *K. pneumonia*. Representative β-lactamases that belong the ESBL families are class A enzymes TEM, SHV, CTX-M-1 to M-28 as well as recently classified class D enzyme OXA. Distinct mutations in the amino acid progression of TEM and SHV derived ESBLs allow this resistance. OXA, commonly found in *Pseudomonas aeruginosa*, provides resistance to third and fourth generation cephalosporin, cefepime and azetreonam. Plasmid encoded CTX-M derived ESBLs occur in numerous bacteria of the family *Enterobacteriaceae* and have the potential to disrupt cefotaxime. [42]

1.5. General introduction to *Acanthamoeba*

The free-living protist *Acanthamoeba* is a globally distributed eukaryote that can acts as opportunistic agent of rare but severe diseases in humans and animals [36]. The scientific interest on this microorganism has increased in last couple of years by medical doctors and researchers [27]. Plenty of different *Acanthamoeba* species are distributed throughout several natural environments, as they frequently appear in terrestrial surroundings, aquatic reservoirs e.g. swimming areas, tap and fresh water areas, soil and in various dusty air territories [36]. The very first report on an infection caused by an amoeba was released in 1973, when a Texan farmer hurt his eye at work and was affected by cleaning his eye with contaminated water [29]. Up to date, twenty different genotypes (T1-T20) have been described based on more than 5% dissimilarities of the 18S rRNA gene [11]. The most

prevalent strains belong to the T4 genotype to potentially cause harmful keratitis infection, severe granulomatous amoebic encephalitis (GAE), inflammatory diseases and skin irritations [33]. Amoebae can enter human ports via different paths, as they can penetrate through the eye, the nose or the respiratory tract. An eye infection is commonly associated with the degradation of the corneal matrix in immunosuppressed hosts and occurs often in contact lens wearers, but also in non-contact lens wearers in smaller proportion [43]. Additionally, the severity of the protist´s virulence is mediated by carrying a broad spectrum of various bacterial, fungal and viral symbionts [10]. Since the recent years, the infection rate increases year after year due to poor hygiene conditions and can affect patients at any age [29]. In rare cases, however, *Acanthamoeba* can infringe the central nerve system of human very heavily.

In general, the metabolic cycle of *Acanthamoeba* include a vegetative cyst stage on the one hand and an metabolic-active trophozoite form, on the other hand, as illustrated in figure 2 [48]. The amoeba remains in its cyst stage to resist harsh environmental conditions, whereas

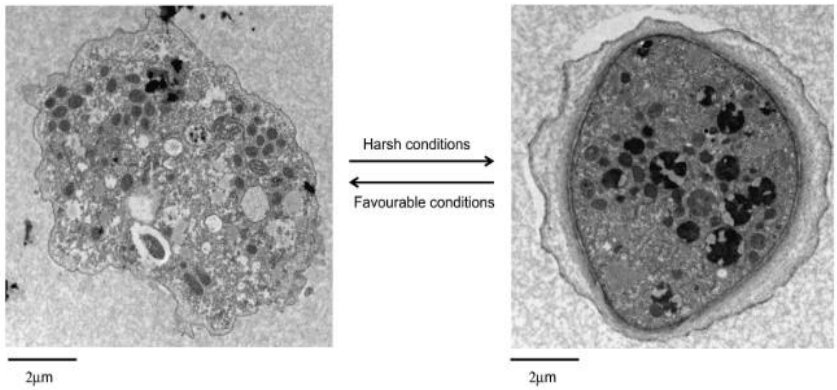


Figure 2. Image of the two metabolic stages of Acanthamoeba [48]
The infectious forms of this protozoa involve the emerged trophozoite stage (left image) under proper circumstances and the protective cyst shape (right image) if inconvenient living conditions are given. Trophozoites are motile as they contain small extension at the outer surface. Mature cysts are characterized by a double wall. Additionally, three morphological groups (I, II and III) of *Acanthamoeba* ssp. are reflected by different cyst structures.

within the trophozoite stage the organisms normally feeds on bacteria, algae and fungi and appears in environmental habitats. Despite, both metabolic stages can infect human body harmfully [48]. The main influence on the level of disease severity is given by the invasiveness and the expression of certain virulence factors. Potential pathogenicity has been declared by the research group of Lorenzo-Morales and involves proteins that cause damage

directly to the host cell and others that ensure the environmental survival of *Acanthamoeba* during an infection [35].

1.5.1. Clinical diagnosis of *Acanthamoeba* infections

The infection rate with *Acanthamoeba* has increased in the last year, mainly in the USA and European countries, due to the raising number of contact-lens wearers [29]. The group of free-living amoebae includes the related genera *Acanthamoeba*, *Balamuthia*, *Sappinia*, *Hartmannella* and *Naegleria* that are potential disease-causing protozoa. In Austria, the main risk factor is poor hygiene conditions [56]. AK is defined as a painful infection of the corneal surface which appears in cyclic turbidity of the eye matrix and damages of the eye tissue and eyelids. In very late stages of the disease it can even lead to blindness. In this aspect, an early diagnosis is crucial to provide an efficient medical treatment [29]. A rapid identification of *Acanthamoeba* is mandatory but very challenging. In routine tests, *Acanthamoeba* species are determined by 18S genotyping with PCR, staining procedures and microscopic examination [56]. Cysts contain a well-structured double cell wall to be clearly distinguishable from the trophozoite stages by professionals under the microscope. Trophozoites possess so-called acanthapodia, tiny appendages to move, by which they can be detected with the phase contrast microscope. Moreover, confocal microscopy is already consulted for an *in vivo* analysis of keratitis infections. Even though *in vitro* cultivation of Acanthamoeba is time-consuming, it still remains the main technique used for proper analysis in the laboratory. An extraction of Acanthamoeba is a sensitive task that is carried out by corneal biopsy or scraping [35].

Interestingly, *Acanthamoeba* is not the only responsible pathogen that can cause severe symptoms of keratitis. Several microbes such as bacteria, viruses and fungi have the potential to harm the corneal epithelium layer. Reported cases of microbial keratitis describe staphylococcal species, *Pseudomonas aeruginosa* and *Streptococcus pneumonia* as dominate infectious agents. An eye infection by Gram-positive and Gram-negative bacteria may result in corneal ulcers with an infection rate of eighty percent [52]. Furthermore, AK can be easily confused with a herpes simplex virus or fungal keratitis [35][29]. Patients that suffer from an *Acanthamoeba* infection require expensive therapy including drugs. Lorenzo-Morales and his group mention useful substrates that can reduce the infectious power of trophozoites and cysts comprising cationic antiseptic, aromatic diamidine, aminoglycoside,

and alcyl phosphocholine. Prolonged therapy is needed to harm the robust cyst stage [35]. Recent studies confirm that high production of tears liquid protect *Acanthamoeba* infection as it contains high content of immunoglobulin A antibodies [39].

1.5.2. Pathogenicity of *Acanthamoeba*

Certain virulence factors allow the amoeba to damage host cells intensively. The ability to adhere to glycoproteins on surface areas on human body initiates the risk to trigger disease. Toward keratitis disease, adhesion proteins enable the prior host-parasite interaction and attachment to the corneal surface. The carbohydrate-recognizing mannose-binding protein (MBP), the laminin-binding proteins (LBP), C-lectin and more have been described as major structures to induce the cytopathic effect (CPE) [39][35]. A certain surface-binding requires pattern recognition receptors (PRR) to activate cell-signaling cascade such as the Rho signaling pathway and Ca++ channels, which play a key role in host damaging. Moreover, host cell-associated amoeba can penetrate the epithelial membrane and lead to the expression of several metalloproteinase and alkaline serine proteinases which furthermore degrade the corneal matrix. Especially the mannose-induced protein MIP-133 (133kDa), which gets activated in the presence of mannose, is known to have strong cytotoxic protease activity and well involved in the CPE [39]. Prior studies confirmed that MIP-133 improves the cytolytic activity in epithelial cells. This cytolytic protein ensures high rates of apoptosis when it is associated to corneal surface and in the absence of inhibiting factors, such as caspase 3 or serine protease inhibitors [22]. Several other proteases, elastases and ligases may additionally be involved in degrading the main unit of the eye. Painful symptoms of a keratitis infections by *Acanthamoeba* are similar to an *Entamoeba histolytica* infection, as both species secrete destructive proteases [9].

1.5.3. Molecular classification of *Acanthamoeba*

Acanthamoeba species are classified into three different morphological groups first reported by Pussard and Pons in 1977: nomenclature of group I, group II and group III is based on the cyst morphology [55]. Group I includes the largest cysts, whereas group II and group III contain cyst sizes of less than 18µm in diameter [30]. All cyst groups are distinguishable by professionals under the microscope. Species of group III are more closely-related to those of morphological group II than to group I [45]. Moreover, 20 different genotypes (T1-T20) of rare pathogenic amoeba have been reported nowadays, while T1 and T20 predominantly

headed to cause GAE [11]. Anyhow, various genotypes are represented in clinical and non-clinical samples [44]. Genetic differentiations rely on dissimilarities of the 18S small ribosomal subunit within a sequence variance of more than 5%. Genus-specific PCR amplification of the 18S rRNA from different *Acanthamoeba* species is performed in routine laboratory work with high sensitivity using the forward primer JDP1 (5′-GGCCCAGATCGTTTACCGTGAA-3′) and the reverse primer JDP2 (5′-TCTCACAAGCTGCTAGGGAGTCA-3′) [46]. The ASA.S1 amplicon contains a length of approximately 500bp depending on individual genotypes [46][44]. Furthermore, the determination of the 16S mitochondrial gene of *Acanthamoeba* can be mediated with following primers to obtain a PCR product of 161bp, the primer Aca16Sf1010 (5´-TTATATTGACTTGTACAGGTGCT-3´) on the one hand and Aca16Sr1180 (´5-CATAATGATTTGACTTCTTCTCCT-3´) on the other hand [58]. However, no correlation between the morphological groups and the various genotypes has been described nowadays. Even though, most commonly the *Acanthamoeba* genotype T4 leads to an outbreak of harsh diseases [48].

The non-pathogenic *Acanthamoeba castellanii* Neff reference strain has been sequenced near its entirety and was published by Clarke in 2013 [10]. This genome belongs to the T4 genotype and includes 15,655 gene annotation, 14,974 proteins and a GC content of 58,4% (AHJI00000000.1). The complete genome of the related and free-living parasite, *Balamuthia mandrillaris* was published recently by Detering et al [12].

1.6. Next generation sequencing technologies

Since the development of Sanger sequencing by Frederick Sanger and DNA sequencing by chemical modification invented by Allan Maxam and Walter Gilbert in 1977, DNA sequencing has become a revolutionary technique in the field of molecular and forensic science, microbial ecology and genetics [25]. Sanger sequencing is an established and reliable technology that researchers still use in many research facilities nowadays. The development of high-throughput sequencing technologies seem almost indispensable in certain research areas these days and has created a broad market of individual sequencing technologies. Some popular NGS providers are Roche supplying the 454-pyrosequencing system, the US-company Illumina with the Solexa sequencing, SOLiD NGS technology from Applied Biosystems and NGS applications from Ion Torrent.

In the field of medical microbiology, 454 GS FLX pyrosequencing technology is mainly used to study human microbial communities by the sequencing-by-synthesis principle, which basically combines PCR and the sequencing applications. Single-stranded DNA is associated to capture beads and placed on a picotiter plate where emulsion PCR is performed. Following sequencing demands specific sequencing primers and a couple of various enzymes: DNA polymerase, ATP sulfurylase, luciferase and apyrase [51]. Pyrosequencing supports paired-end reads with an outcome of approximately 1 million reads per run. The use of many enzymes considers pyrosequencing as a cost-inefficient strategy for sequencing. Furthermore, one can even gain more sequencing output by choosing other NGS approaches such as Illumina, Ion Torrent or SoLiD [21].

As descripted by Hodkinson et al., since 2006, Illumina provides a cost-efficient high-throughput platform that immediately caught attention from various research areas. Various types of applications are accessible for sequencing such as Illumina GS/Hi Seq/Mi Seq/NextSeq. The principle of this NGS technique relies on a reversible dye-termination detection system and the use of oligonucleotide-associated flow cells. Adapter-associated DNA fragments hybridize to oligonucleotides on the flow cells and form clusters. Currently, the highest performance resulted in 4 billion fragments of paired-end reads each in a length of 125bp with the HiSeq 2500 application [21].
Another NGS system, predominantly used for microbiome analysis, is the SOLiD technology offered by Applied Biosystems since 2007. Compared to previous mentioned technologies, SOLiD demonstrates a sequencing-by-ligation procedure. The sample preparation begins with an emulsion PCR. Paired end reads are available with the highest output of 3 billion reads per run with a read length of 75 base pairs. A limiting factor of this instrument is the production of short read lengths [21].

1.6.1. Ion Torrent personal genome machine system

In general, the delivery of new genomic data provides extended understanding on biological mechanisms for diagnostic standards. In the last years, high-throughput sequencing technologies are able to generate billions of sequencing data rapidly with high efficiency at low costs. Since that, scientific research has orientated its focus more on genetics and bioinformatics. The fully automated and unique next-generation sequencing system from Ion Torrent was used to conduct whole genome shotgun sequencing. Approved accuracy,

speed and workflow simplicity ensures a huge output of reads of genomic DNA, RNA or specific amplicons with read lengths of 200 base pairs, 300 base pairs or 400 base pairs.

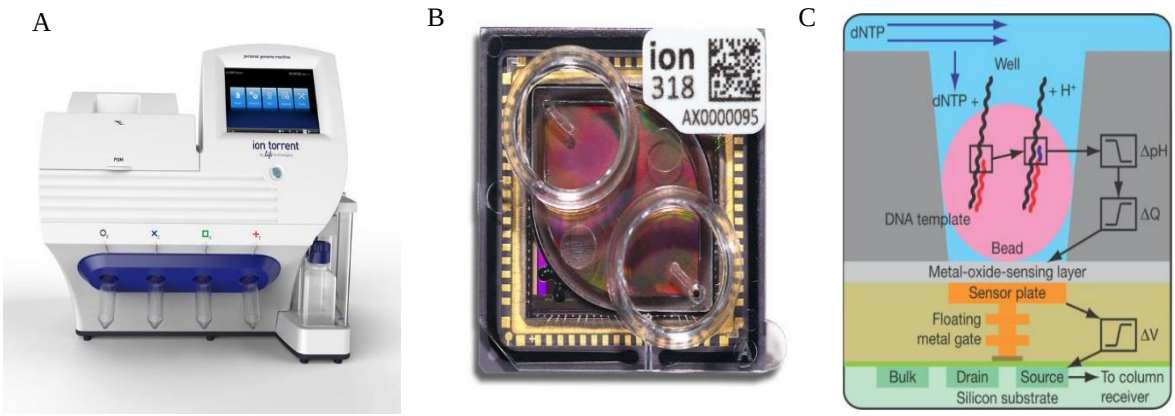


Figure 3. Graphical representation of the Ion Torrent PGM system
The next-generation sequencing procedures with the Ion Torrent PGM system **(A)** requires a semiconductor chip technology to run this process. As soon as the library preparation is finished, the Ion Torrent microchip **(B)** gets loaded with template-associated Ion Sphere Particles. The charged microchip is inserted to the PGM system and launched. Once a dNTP gets incorporated to the sequence, an H+ ion is released and detectable on the metal-sensing layer **(C)**. A signal conversion provides a digital output on the Ion Torrent software.

Technically, the Ion Torrent sequencing system relies on the conversion of a chemical signal into digital data without utilizing any additional optical source to detect incorporated nucleotides (www.thermofisher.com/order/catalog/product/4462921). This means the Ion Torrent detection principle is not based on the imaging of fluorophore-associated nucleotides, but on the accurately defined change of the pH-value when a single nucleotide incorporates into the DNA. The semiconductor chip technology enables the measurement of a single nucleotide, which is based on the release of H+ ion and a diphosphate group as soon as the nucleotide gets integrated during synthesis. Instantly, a released H+ ion creates a biochemical signal that is directly converted into voltage state. In principal, this technology makes it possible to detect signals individually for every single nucleotide. Millions of genomic fragments can be sequenced simultaneously with up to eight different libraries depending on corresponding genome sizes.

1.6.2. Ion Torrent chip technology

The development of innovative next-generation sequencing technologies has improved the efficiency and speed to deliver genetic information. The newly invented power of the semiconductor sequencing system supplied by Ion Torrent enables an affordable high-throughput sequencing method. Certain semiconductor microchips are available in different

read-out sizes with various numbers of wells on its surface. The PGM 314™ chip enables a read-out of 100Mb, the PGM 316™ chip an amount of 1Gb and the PGM 318™ chip can provide an output up to 2Gb.The principle of the microchip is based on a metal-oxide-sensing layer right on the bottom which is directly connected to a sensor plate. After the generation of DNA libraries, sheared fragments are associated to Ion Sphere particles (ISPs) which individually fit to the chip wells. As the incorporation of single nucleotides changes the pH-values the chemical signal is converted into a digital one.

1.7. Nucleotide databases in the field of molecular microbiology

Current information on specific nucleotide sequences for all organisms is united on the GenBank database that is a publicly accessible online system on the national center for biotechnology information (NCBI) website. GenBank basically connects high-throughput nucleotide sequences, expressed sequence tag (EST), whole genome shotgun projects and gene annotation (www.ncbi.nlm.nih.gov/genbank/). NCBI shares a broad spectrum of eligible data banks to simplify the overall access to circulating scientific data. Mainly, the European nucleotide library EMBL and the DNA data bank of Japan DDBJ provide close coordination for regular updates on GenBank. Investigation on genome similarities is most commonly performed with BLAST on GenBank by aligning a query sequence to the sequences of interest [5]. The high quality ribosomal RNA database was created to address bacterial, archaeal and eukaryotic ribosomal RNAs extensively for species-specific and phylogenetic characterization (www.arb-silva.de). Bacterial and archaeal genes as well as microbial genomes are further reported on EmsemblBacteria homepage, which had its release 32 in August 2016 by EMBL-EBI (www.bacteria.ensembl.org). Regarding rare eukaryotic microorganisms, protozoa and worms in particular, only little genomic details are attainable on any nucleotide databases. In contrast, numerous bacteria and viruses have already been sequenced entirely due to relatively small genome structure and most often clinical relevance.

1.8. Bioinformatics analysis

The handling of enormous data sets with bioinformatics tools is often challenging and very time consuming for researchers. The quality of raw sequencing data needs to be determined in order to draft the right path for further analysis. Statistical evaluation was established with FastQC that is defined as a quality control tool for raw high-throughput sequencing data.

This program was established by the Babraham Bioinformatics Institute. FastQC is available as an open source tool for several operating systems. Moreover, FastQC accepts SAM or any FastQ format as input files. This application generates a graphical overview of the quality scores for all reads and reports information on the GC% content, sequence length distribution and base pairs quality per read.

MIRA- Mimicking Intelligent Read Assembly

MIRA 4.0 is defined as a linux-based genome de-novo assembly tool that is able to perform sequence analysis of reads produced by Sanger, 454, Illumina, Ion Torrent or PacBio technologies. This program is available since 1998 with the latest update in 2014. De-novo assembly generates a full-length sequence out of short reads, whereas mapping performances require a backbone reference sequence to be compared. The Ion Torrent Software provides the Assembler (v3.4.2.0) plug-in which was launched for all *Acanthamoeba* sequencing data (www.mira-assembler.sourceforge.net/docs/DefinitiveGuide ToMira.html).

Kraken

Kraken is a comprehensive and rapid system to arrange taxonomic classification of NGS data for bacterial, archaeal and viral species developed by researchers from CCB, Center for computational biology at John Hopkins University. We used the metagenomics program Kraken to filtered out all reads that belong to *Acanthamoeba* and removed those reads which corresponded to any other prokaryotic or eukaryotic organism. Therefore, Kraken requires the download of individual databases for organisms before launching sequence comparison [57].

1.9. Isothermal amplification assay

A fast and accurate identification of disease-causing microorganisms can often not be realized due to limited detection systems, especially in developing and poor developed countries. However, the *in vitro* cultivation of microorganisms is often time-consuming and not applicable for every species. The widely used PCR technique has revolutionized experimental applications in the field of molecular biology and clinical diagnostics to characterize pathogens by amplifying particular gene sequences. This fundamental DNA amplification method is commonly applied to detect infectious diseases and genetic disorders [16]. Classical PCR generates multiple copies of target DNA using specific primers

and a suitable enzyme, the thermostable *Taq* DNA polymerase (*Taq* Pol I), which is isolated from the Gram-negative bacteria *Thermus aquaticus* and exhibits great amplification activity at high temperature rates (93-95°C) [26]. This DNA-dependent DNA polymerase possesses a 5′→ 3′ exonuclease activity but no 3′→ 5′ exonuclease activity. *Taq* DNA polymerase is commonly used for PCR but also suitable for isothermal reactions. Moreover, Klenow *Taq* Polymerase is characterized by the absence of the 5′→ 3′ exonuclease function as the 323 amino acid residues on the N-terminal end are genetically removed and is even active at 68 to 72°C [32]. In comparison to the gold standard PCR technique, during isothermal amplification, a multiple displacement amplification (MDA) enables amplification at constant temperature and results in a linear nature of the amplification process [28]. In the area of point-of-care devices, an isothermal amplification assay demonstrates a robust and powerful technique with the extensive use of thermophilic enzymes [54]. Various types of isothermal applications have been studied over the last years such as the strand-displacement

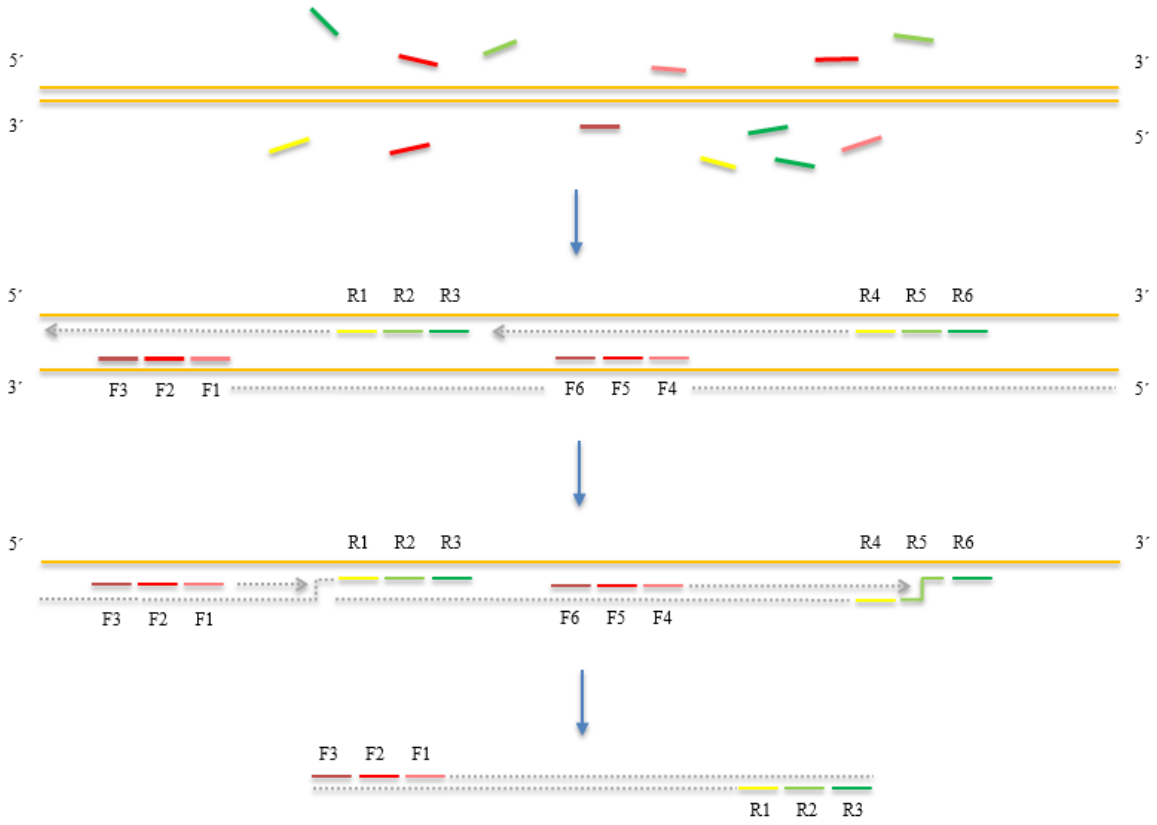


Figure 4. Illustration of a multiple strand displacement amplification (MDA)
This scheme represents the theoretical annealing procedure of two different sets of six primers during the isothermal multiplexed reaction. Primarily, the DNA double helix gets denatured by the strand displacement activity of the *Bst* DNA polymerase. An open conformation forces primers to anneal to the complementary target region. As soon as the DNA polymerase recognizes a hybridized primer (marked by arrow), the newly synthesized strand gets replaced and attached by another set of primer.

amplification (SDA), rolling circle amplification (RCA), helicase- dependent amplification (HDA), loop-mediated isothermal amplification (LAMP) and other procedures [24]. An isothermal reaction presupposes a constant reaction temperature and a thermostable polymerase that owns a strand displacement activity. Moreover, the *Bst* DNA polymerase, large fragment, which was originally isolated from *Bacillus stearothermophilus* is highly suitable for isothermal reactions as it contains a 5′→ 3′ polymerase activity, great 5′→ 3′strand displacement activity and a lack of 3′→ 5′ and 5′→ 3′ exonuclease activity [1, 31]. An important issue for isothermal amplification is, that DNA denaturation, primer annealing and the extension phase are conducted without classical PCR thermo cycling procedures. Moreover, a linear amplification procedure in MDA avoids DNA saturation in late reaction period. In fact, DNA quantification is achievable and strongly benefits the isothermal amplification. Accurate primer design and defined salt concentrations ensure proper strand displacement activity mediated by the enzyme. Besides that, primer-binding specificity is given by the defined reaction temperature. A prudent construction of complementary primers is a mandatory precondition for a successful amplification outcome. Certain parameters should be observed precisely that involves customized GC% content, primer dimerization events and the occurrence of hairpin structures.

1.10. DNA quantification with molecular beacon

Fluorescent detection techniques are commonly used in real-time amplification assays for a fast, reliable and sensitive DNA quantification. Target-specific molecular beacons (MB) are characterized by a short single-stranded oligonucleotide in the size of 15-30nt that forms a stem-loop DNA structure [40]. The stem is considered to contain four to seven nucleotides

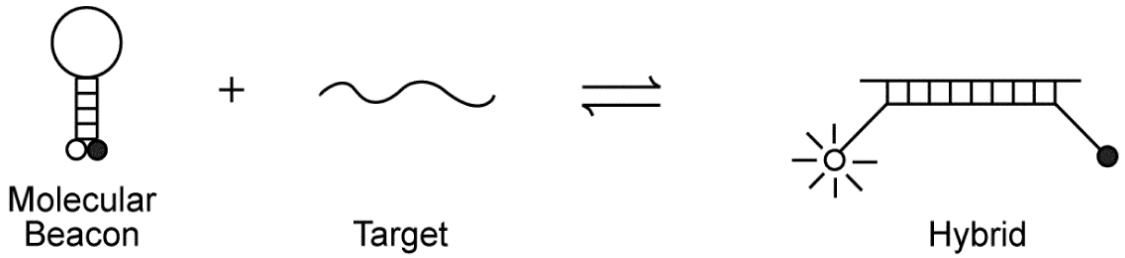


Figure 5. Hybridization mechanism of molecular beacon [31]
The stem of molecular beacon stays closed in absence of target DNA because the quencher inhibits the light emission of the fluorophore. However, when molecular beacon hybridizes to the complementary sequence region forming a double strand, an open conformation of the stem arises and provides fluorescent signal which can be determined with real-time PCR.

depending on the annealing temperature. The stem region consists of the fluorophore 6-carboxyfluorescin (6-FAM) on the 5´ end and a Black Iowa quencher molecule on the 3´ end. 6-FAM probes have their maximum absorbance profile at 495nm and emission spectra at 520nm wavelength. In closed stem conformation, the quencher hinders light emission of the fluorophore by sharing electrons transiently [38]. The recognition of target-DNA forces molecular beacon to emit light. Accurate signal output of the MB occurs at the annealing temperature during amplification. We expect to determine a linear increase of fluorescent signal during the isothermal amplification. A molecular beacon was designed during this thesis to visually follow the isothermal amplification.

2. Materials

Table 1. Applications and laboratory devices

Description	Supplier
2% Agarose Gel Cassettes	Life Technologies (Carlsbad, CA, USA)
1,5mL Eppendorf LoBind® Tubes	Eppendorf (Hamburg, Germany)
0,2ml PCR tubes	STARLAB (United Kingdom)
Agilent® 2100 Bioanalyzer®	Agilent Technologies (Santa Clara, USA)
Light Cycler 480 II	Roche Diagnostics (Switzerland)
DynaMag™ magnetic rack	Invitrogen (Carlsbad, CA, USA)
E-Gel® SizeSelect Agarose Gel	Invitrogen (Carlsbad, CA, USA)
GelAmp® PCR system 2700	Applied Biosystem (Foster City, USA)
Gel electrophorese equipment	VWR (Radnor, Pennsylvania, USA)
Ion Torrent One Touch ES system	Life Technologies (Carlsbad, CA USA)
Ion Torrent One Touch 2 system	Life Technologies (Carlsbad, CA USA)
Ion Torrent PGM system	Life Technologies (Carlsbad, CA USA)
Microcentrifuge	VWR (Radnor, Pennsylvania, USA)
UV imaging system	UVP (Upland, USA)
NanoDrop 2000c spectro photometer	Thermo Fisher Scientific (Waltham, MA, USA)
PCR-hood	Biosan
Picofuge	VWR (Radnor, Pennsylvania, USA)
Pipettors P2, P10, P20, P50, P100, P200, P1000 (µl)	VWR (Radnor, Pennsylvania, USA)
Pipettor tips Safe Seal SurProb	Biozym (Hessisch, Germany)
Qubit®2.0 Fluorometer	Invitrogen (Carlsbad, CA, USA)
Thermocycler T3000	Biometra (Göttingen, Germany)
Thermal duocycler	VWR (Radnor, Pennsylvania, USA)
Vortex mixer	Labnet International, Inc. (Edison, USA)

Table 2. Applied enzymes and chemicals

Description	Supplier
Agencourt® AMPure® XP Kit	Beckman Coulter (Brea, USA)
*Alu*I *(Alu*BI) restriction enzyme (10U/L)	Thermo Fisher Scientific (Waltham, MA, USA)
Breaking solution	Life Technologies (Carlsbad, CA, USA)
Bst DNA Polymerase, large fragment (8000 U/ml)	New England BioLabs® *Inc* (Ipswic, MA, USA)
Deoxyribonucleotides (10mM each)	Thermo Fisher Scientific (Waltham, MA, USA)
DMSO 100%	Thermo Fisher Scientific (Waltham, MA, USA)
E. coli DH10B Control Library	Life Technologies (Carlsbad, CA, USA)
50bp DNA ladder (1µg/µL)	Life Technologies (Carlsbad, CA, USA)

Generuler (100bp) DNA Plus	Fermentas, Thermo Fisher Scientific (Waltham, USA)
Generuler (1kb) DNA Plus (0.5µg/µl)	Thermo Fisher Scientific (Waltham, MA, USA)
Ion P1 Adapter	Life Technologies (Carlsbad, CA, USA)
Ion PGM™ Hi-Q™ View Enzyme Mix	Life Technologies (Carlsbad, CA, USA)
Ion PGM™ Hi-Q™ View ISPs	Life Technologies (Carlsbad, CA, USA)
Ion Shear™ Plus Enzyme Mix II	Life Technologies (Carlsbad, CA, USA)
Ion Torrent OneTouch™ Breaking solution	Life Technologies (Carlsbad, CA, USA)
Ion Torrent OneTouch™ Oil	Life Technologies (Carlsbad, CA, USA)
Ion Torrent OneTouch™ Recovery solution	Life Technologies (Carlsbad, CA, USA)
Ion Torrent TaqMan® qPCR Mix, 2X	Life Technologies (Carlsbad, CA, USA)
Ion Torrent TaqMan® Quantification Assay, 20X	Life Technologies (Carlsbad, CA, USA)
Ion Xpress™ Barcode X	Life Technologies (Carlsbad, CA, USA)
Lambda/HindIII ladder	New England BioLabs® *Inc* (Ipswic, MA, USA)
Library Amplification Primer Mix	Life Technologies (Carlsbad, CA, USA)
$MgSO_4$ (100mM)	New England BioLabs® *Inc* (Ipswic, MA, USA)
Molecular beacon #1 SHV beacon	Integrated DNA technologies (Coralville, IA, USA)
Molzym Mastermix 16S Basic(2,5x)	Molyzm (Bremen, Germany)
Molyzm Mastermix (2,5x) dye	Molyzm (Bremen, Germany)
Molzym MolTaq 16S polymerase	Molyzm (Bremen, Germany)
Nick Repair Polymerase	Life Technologies (Carlsbad, CA, USA)
PEG 4000	Fermentas, Thermo Fisher Scientific (Waltham, USA)
Platinum® PCR SuperMix High Fidelity	Thermo Fisher Scientific (Waltham, MA, USA)
Proteinase K	QIAGEN (Hilden, Germany)
SYBR Safe- DNA Gel Stain	Invitrogen (Carlsbad, CA, USA)
TBE (10X)	Sigma Aldrich (St. Louis, M, USA)

Table 3. Reaction buffers

Description	Ingredients	Supplied by
ACB	100% ethanol added	QIAGEN (Hilden, Germany)
AW1 wash buffer	Not disclosed	QIAGEN (Hilden, Germany)
AW2 wash buffer	"	QIAGEN (Hilden, Germany)
AVE elution buffer	"	QIAGEN (Hilden, Germany)
Ion Shear Plus 10X Reaction Buffer	"	Life Technologies (Carlsbad, CA, USA)
Ion Shear™ Plus Stop Buffer	"	Life Technologies (Carlsbad, CA, USA)
Ion Torrent Low TE	10mM Tris 0.1mM EDTA pH 8.0	Life Technologies (Carlsbad, CA, USA)

QIAGEN 10X PCR buffer	Tris-Cl KCl $(NH_4)_2SO_4$ 15 mM $MgCl_2$ pH 8.7 at 20°C	QIAGEN (Hilden, Germany)
Thermo Pol buffer (10X)	200 mM Tris-HCl 100mM $(NH_4)_2SO_4$ 100mM KCl 20mM $MgSO_4$ 1% Triton® X-100 pH 8.8 at 25°C	New England BioLabs® *Inc* (Ipswic, MA, USA)

3. Methodology

3.1. Sequencing workflow on three different *Acanthamoeba* strains

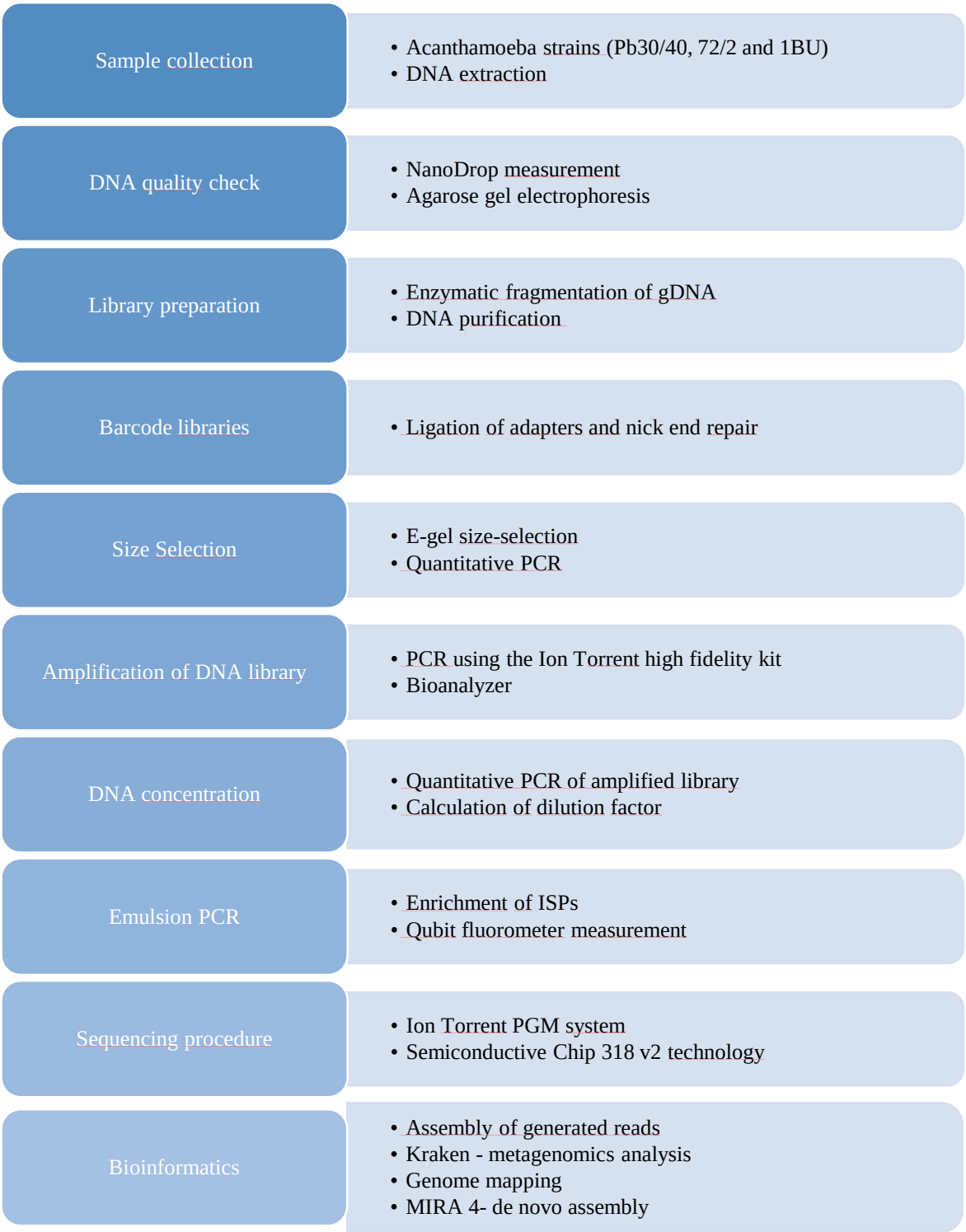


Figure 6. Scheme of whole genome sequencing workflow using the Ion Torrent PGM system

All experimental guidelines, product descriptions and potential troubleshooting are reported in the Ion Torrent manuals. Technical instructions for the library preparation are described in the Ion Xpress™ Plus gDNA Fragment library preparation protocol (Pub. Nr. MAN0009847) and Ion PGM™ template OT2 400 kit manual (Pub. Nr. MAN0007218). Further instructions are mentioned in the Ion PGM™ Hi-Q™ OT2 Kit (Pub. Nr. MAN0010902) manual. The Ion PGM™ Hi-Q™ Sequencing Kit explains the final preparation steps before using the Ion torrent PGM sequencer. All Ion Torrent NGS instruments are supplied by Life technology and include the Ion Torrent One Touch™ 2 system for preparing template-positive Ion Sphere particles, the Ion Torrent™ ES for the enrichment of ISPs and the Ion Torrent personal genome machine (PGM). The resequencing process was guided by the Ion PGM™ template OT2 400 kit protocol (Pub. Nr. MAN0007218), followed by the Ion Torrent™ High-Q™ View OT2 kit (MAN0014579) and Ion Torrent Hi-Q View Sequencing Kit (MAN0014583).

3.1.1. DNA samples

In collaboration with Prof. Dr. Julia Walochnik, who heads the "Molecular Parasitology" department at the Medical University of Vienna, we received 300µl of DNA from the respective *Acanthamoeba* strains, which was isolated by researchers of her group: Pb30/40 (T7 genotype), 72/2 (T5 genotype) and 1BU (T4 genotype) as listed in table 4. After an *in vitro* cultivation, the DNA extraction was performed with the QIAmp DNA isolation kit (Qiagen, Vienna, Austria) guided by the supplier´s manual. As recommended by the supplier, 50-100ng per sample should be used as input DNA for whole genome sequencing performances. All samples were stored at -8 to -30°C. DNA concentrations of obtained samples were determined on Nanodrop c2000 photospectrometer. Moreover, a gel electrophoresis was performed to demonstrate the DNA quality of each strain.

Table 4. DNA samples

Strain	ng/µl	260/280	DNA input (µl)
Pb30/40*	43,5	2,39	2
72/2*	38,3	2,26	2,2
1BU*	44,9	2,29	2

* genomic DNA

3.1.2. Generating DNA libraries

The generation of a robust genomic library is of most importance in order to deliver accurate NGS results with less drawbacks when performing next-generation sequencing. Basically, DNA or RNA molecules are sheared into smaller fragments, which can be performed either mechanically by nebulization or sonication, or by enzymatic fragmentation [43].

In the first sequencing process, the library preparation was generated enzymatically using 100ng DNA of each sample, while in the resequencing run 87ng DNA of Pb30/40, 84,3 ng DNA of 72/2 and 89,2 ng DNA of 1BU were applied to the reaction. Hence, DNA was mixed with nuclease-free water, 10X reaction buffer and 10µl of the mandatory Ion Shear Plus Enzyme mix II to a volume of 50µl. In order to receive fragments in a length of 400 bp, the solutions were incubated for 8 minutes at 37°C on the heating block. The following library purification requires the addition of 99µl Agencourt® AMPure® XP beads and an incubation of five minutes. Subsequent washing workflows were carried out with freshly prepared 70% ethanol. The samples were placed onto the magnetic rack until the solution has cleared. The tubes were twisted twice directly at the magnetic rack and fragment-associated AMPure beads have been washed properly. The supernatant was removed without dispensing the pellet. This washing procedure was repeated. Residual ethanol was removed by pulse-spinning for 30 seconds and collected with a 100µl pipette. The purified library was resuspended in 25µl low TE buffer. Placing the tubes back onto the magnetic rack, 20µl of the desired library was transferred into a new 1,5ml LoBind Eppendorf tube and stored at -20°C.

3.1.3. Ligation of adapters to established libraries and nick end repair

The Ion Torrent sequencing process demands the association of specific adapters (A and P1) to the sheared library fragments. In first sequencing run barcoded adapters were used to differentiate between the three *Acanthamoeba* strains. A 100µl reaction mixture includes 20µl of respective DNA library, 10X ligase buffer, 2µl P1 adapters, 2µl of the dNTP mix and nuclease-free water. Individual barcodes were added to each solution. To avoid cross contamination, gloves were changed immediately after barcode pipetting. Furthermore, 2µl of DNA ligase and 8µl of nick end repair polymerase was added to the final mixture. The solution was vortex, pulse-spinned and placed onto the thermocycler for 30 minutes. The ligation and nick repair procedures require the following thermocycler setting: 25°C for 15

minutes, 72°C for 5 min and 4°C hold. After completion, each solution was transferred into a new 1,5ml LoBind Eppendorf tube and a volume of 100µl of Agencourt® AMPure® XP bead was added as recommended for library lengths of 300 to 400 base pairs. Again, an AMPure™ bead washing step was performed as it is described in the prior chapter "Generating DNA libraries".

3.1.4. Gel-based size-selection

After finishing the construction of adapter-associated fragments, size selection was initialized by loading the samples onto an E-Gel SizeSelect Agarose Gel. Therefore, 1µg/µl of the 50bp DNA ladder has been diluted with Low TE buffer (1:40) and 10µl of the dilution was filled into the M-well on the gel. Furthermore, only middle wells were filled with sheared DNA samples. All empty wells were loaded with 25µl nuclease-free water. As it is recommended in the kit manual, wells at the edge of the gel should not be loaded in order to prevent failures in selecting incorrect fragment sizes. After a 16 minutes gel run the collection well were refilled with 10µl ddH_2O. As soon as the expected 400bp sequences entered the bottom wells of the gel, 15-20µl of the correct fragment sizes were removed from the gel by pipetting. Additionally, 10µl of nuclease-free water was refilled into the bottom wells to reduce a loss of usable fragments. The recovered DNA solution was immediately stored on ice and quantified by real-time PCR.

3.1.5. Quantitative polymerase chain reaction

Real-time PCR was performed with size-selected libraries to ensure adequate DNA concentration for the sequencing approach. Primarily, a serial dilution (1:10) of the standard E. coli DH10B (stock solution: 68pM) was arranged up to a concentration of 0,00068pM. Respective libraries were diluted 1:10 and 1:100 for measurement. Required qPCR master mix includes 210µl of 2x Ion Library *Taq*Man® Quantitation Assay (thawed but not vortexes), 84µl of nuclease-free water and 21µl of the Ion Library *Taq*Man® qPCR Mix. Moreover, 15µl of this solution was aliquoted into 96-well plates. Finally, 5µl of the standard dilution and each sample dilution and 5µl of nuclease-free water as non-template control were added to the plate. Closing up, the 96-well plate was covered with a slide, vortexed, spinned and entered to the light cycler 480 II system and the Ion torrent quantification run was started.

The *Taq*Man probe consists of a linear oligonucleotide that involves a fluorescent reporter at the 5´ end of the probe and a quencher molecule at the 3´ end. Basically, the quencher inhibits fluorophore emission of the reporter. The probes hybridize specifically to single-stranded DNA during the amplification process. As soon as the reporter fluorophore is replaced by the exonuclease function of the *Taq*Man polymerase during the elongation step, it remains in an unquenched form and results in an output of the fluorophore signal. The fluorophore emits direct proportionally to the amount of specifically amplified target sequence.

3.1.6. Amplification of DNA library

Quantitative PCR confirmed too low DNA concentrations for each library to continue the protocol. Therefore at this point, an additional amplification step was carried out with 100µl of supported Platinum® PCR SuperMix High Fidelity mix, 5µl of a universal library amplification primer mix and 22µl of the DNA library in a 0,2ml reaction tube. The library amplification was run with an initial denaturation step at 95°C for 5 minutes followed by nine cycles of 15 seconds denaturation at 95°C, an annealing phase for 15 seconds at 58°C and an extension phase for 1 minute at 70°C. Subsequent purification steps were performed using 130µl of the AMPure® bead solution. The additional purification was performed similarly as described in paragraph “Generating DNA libraries”.

3.1.7. Bioanalyzer analysis

The quality and quantity check of desired fragment sizes was measured with the Agilent 2100 Bioanalyzer which is a sensitive on-chip electrophoresis system. Undiluted and 10-fold dilution of the amplification products for each sample were loaded onto the bioanalyzer chip. Before loading the samples, 9µl of the gel-dye mix was pipetted into intended wells and distributed over the DNA chip using the supplied apertures. Furthermore, 5µl of the marker was added to each well plus 1µl for each sample dilution. The chip was vortexed properly and directly entered to the Bioanalyzer system.

In order to ensure the right DNA concentration for the DNA libraries, a second qPCR run was performed. The standard control of E. coli DH10B was prepared as described in the paragraph above. Due to the very low DNA concentrations within the previous measurement, this time the undiluted library and a 1 to 10 dilutions were analyzed. The next

sequencing procedures require the calculation of individual dilution factors in order to determine an end concentration of 26pM for each library. Based on these values the libraries were diluted with low TE and pooled together into a new LoBind Eppendorf tube.

3.1.8. Emulsion PCR

An emulsion PCR requires the use of supplied forward A-adapters (5´-CCATCTCATCCCTGCGTGTCTCCGACTCAG-template-specific-sequence- 3´) which are primers for initializing the polymerase synthesis as well as a reverse P1-adapters (5´-CCTCTCTATGGGCAGTCGGTGAT-template-specific-sequence- 3´) which bind to the ion sphere particles (ISPs) surface. An accurate preparation of the emulsion PCR with the Ion PGM™ One Touch 2 (OT2) technology demands a prior calculation of the dilution factor for each sample to achieve a final DNA concentration of 26pM for the pooled libraries. During emulsion PCR, polyclonal ISPs are supposed to associate with one particular size-selected DNA sequence. Prior to this, the Ion PGM™ Hi-Q™ Reagent mix was thawed on ice and the Ion PGM™ Hi-Q ISPs placed on room temperature. Following the guidelines to prepare the amplification solution, 15µl of nuclease-free water, 10µl of the calibration standard (recommended for de novo sequencing), 25µl of pooled library, 100µl High Q ISPs and 50µl High Q enzyme mix are directly added to 800µl PGM Hi-Q reagent mix. A total volume of 1000µl is carefully mixed and briefly spinned. Moreover, this 1ml mixture is loaded onto the supplied reaction filter and furthermore 2x 850µl of oil is pipetted equally through the sample port. This filter which now contains the amplification solution is then placed on the top stage of the Ion Torrent OT2 system. After finishing the correct setup for the system, the emulsion PCR is initiated and lasts for 16 hours.

3.1.9. Enrichment of Ion sphere particles

The enrichment of ISPs was enabled by a fully automated process, in which single fragments of the DNA libraries are associated with streptavidin-coupled Dynabeads®. These beads are characterized by that high binding affinity and enlarged surface size. The enrichment of template-positive ISPs was performed with the Ion Torrent OneTouch™ ES system. All reagents were thawed on ice and a melt-off solution was prepared including 280µl Tween® solution and 1M NaOH. The template-associated ISPs were cleaned with the Ion One Touch™ wash solution, heated up at 50°C for 2 minutes and centrifuged for 3 minutes at 15,500x g. For further analysis, 2µl of the ISPs were aliquoted into a new DNA LoBind tube.

Before adding the Dynabeads® to the 8-well strip, they were washed properly with the MyOne™ Beads wash solution. Furthermore, theses beads were added into well 2 of the supplied strip. A volume of 100µl of the template-positive ISPs was pipetted into well 1. As pocket 3 to 5 on the strip were filled with 300µl of Ion OneTouch™ wash solution and well 7 with 300µl of freshly prepared melt-off mixture, well 6 and 8 remained empty. The autonomous enrichment process lasted for 30 minutes.

3.1.10. Qubit fluorometer measurement

The benchtop Qubit fluorometer device provides quick and simple quantification of template-positive ISPs prior to the sequencing process. The detection is based on target-specific fluorescence, which requires photosensitive probes such as Alexa fluor®488 and Alexa fluor®647. The preparation required 2µl of the template-positive ISPs mixed with 19µl annealing buffer and 1µl of the fluorophore probe. The solution was mixed by pipetting up and down before starting the reaction for 2 minutes at 95°C and 2 minutes at 37°C by using the thermocycler. The removal of unbound fluorophores involves a three times washing step with 200µl of Quality Wash buffer. The solution needs to be vortexed and centrifuged at 15.500 x g for 2 minutes. The supernatant was discarded, so that 10µl of fluorophore associated ISPs remain in the tube. After the third washing step, 180µl of the washing solution was added to the ISPs and transferred to a particular 0,2ml Qubit® assay tube. The Qubit® measurement started with 200µl of Quality wash buffer as blank. Moreover, the fluorophore absorbance of purified libraries was determined and should be around 10-30% of template- enriched ISPs.

3.1.11. Ion Torrent PGM sequencing

Before starting the sequencing process, the PGM technology requires several washing and initialization steps as reported in the manual. Previously, the sequencing run must be planned on the Ion Torrent Server. In the final procedure, supplied sequencing primers and the Ion PGM™ Hi-Q™ sequencing polymerase was added to the enriched ISPs and inserted to the semiconductor 318™ microchip technology.

3.1.12. Bioinformatics analysis

Huge amounts of whole genome shotgun reads were analyzed with bioinformatics applications to identify sequence variations compared to the reference genome of the *A. castellanii* Neff strain (AHJI00000000.1). The following paragraphs describe the workflow of biocomputational analysis.

1. The assembly of reads was performed with the AssemblerSPAdes (v5.0.0.0) plug in, which is supported by the Torrent software in order to deliver contiguous sequences. The software output implies FASTA files of assembled contigs and scaffolds. Extended statistical information was reported by the QUAST report and involves the quantity of contigs (≥ 0bp; ≥ 1000bp), total length (≥ 0bp; ≥ 1000bp), N50 and N75 values as well as the GC (%) content.

2. A quality test for generated single-end reads was purchased with FastQC High Throughput Sequence QC report (v0.11.5). In general, FastQC supports data analysis for all NGS reads. Before uploading, UBAM files were converted into FastQ files with bam2fastq function of bamUtial. In principle, FastQC provides statistical information based on the base pair quality scores.

3. Synthetic sequencing adapters (A and P1) were cropped automatically within a adapter-trimming procedure provided by the Ion Torrent software. The removal of bases having low quality scores is mandatory to improve the mapping performance of larger genomes to a reference sequence. Low bp quality scores of reads were evaluated with FastQC and selectively removed.

4. Genome mapping of *Acanthamoeba* strain 1BU to the reference Neff strain was enabled with the rapid and accurate NextGenMap (NGM) program aligning 13-mer sequences and visualized with the Integrative Genome Viewer (IGV) [47]. Depending on the genome size, the mapping quality demands for high coverage of generated reads.

5. Kraken was applied to all trimmed data to perform metagenomics analysis on *Acanthamoeba* species as this specimen obtains a bunch of various symbionts.

6. MIRA 4.0 is an open-source program that was used to perform de novo assembly with raw re-sequencing data for strains Pb30/40 and 72/2 on a Linux operating system. This application provides automatic sequence trimming.

7. BLAST search enables the filtering of similar sequence regions for *Acanthamoeba*.

3.2. Isothermal amplification assay

3.2.1. DNA extraction

Several bacterial isolates were obtained from the academic hospital in Graz and from the general hospital of Vienna. Individual primers ascertained an amplification of specific regions of resistance genes listed in table 5.

Table 5. Clincial isolates

Strain	Resistance gene
Klebsiella pneumoniae 10/134	DHA-1, OXA-1, SHV-11
Klebsiella pneumoniae 10/151	CTX-M-2, TEM-1, SHV-12
Escherichia coli 10/152	OXA-1, CTX-M-1/M-15, TEM-1

Bacteria were grown in 5mL LB-medium containing 5µl cefotaxime antibiotic (0.1µg/ml). The cultivation of these strains was performed at 226rpm at 37°C o/n. Moreover, cells were collected by centrifugation (5´ at 6000rpm) and the pellet was smoothly resuspended twice in 500µl 1X PBS. A ten minute heating step at 95°C inactivates the bacterial organism to isolate DNA subsequently. DNA extraction was conducted using the QIAamp® cador® Pathogen Mini Kit (QIAGEN). This kit includes all reagents required for DNA preparation. First, cells were lysed with protease K and 100µl of VXL buffer added to 200µl of bacteria culture. Lysis was done for 15 minutes at room temperature. After adding 350µl of ACB buffer, the solution was vortexed, loaded onto supported QIAamp® Mini columns and centrifuge for one minute at 8000rpm. The supernatant which includes various cell components was discarded. Subsequent washing steps were performed with 2x 600µl washing buffer AW1 and AW2. Clean DNA material was eluted with AVE buffer into a new LoBind Eppendorf tube. Finally, the columns were spinned for one minute at 13.200rpm and

samples stored at -20°C. The DNA concentrations were evaluated using a Nanodrop UV spectrophotometry.

3.2.2. Primer design

An effective amplification reaction depends on certain salt and buffer concentrations, great polymerase activity and an optimal primer design. Adapted to intended reaction conditions, the creation of primers demands accurate melting temperatures, slight hairpin formation and little self- and hetero dimerization. As the adjustment of the melting temperature is a crucial step along the primer design, primer BLAST (National Center for Biotechnology Information NCBI) as it is reported and recommended as user-friendly and the most precise online tool available [3]. Thus primer design for the MDA reaction was done via the online tool Primer-BLAST (NCBI). Primers listed in table 6 were designed to target different loci of the SHV β-lactamase gene.

3.2.3. Optimisation of the multiple strand displacement amplification

The multiple strand displacement amplification was carried out in a reaction mixture that contained nuclease-free water, 5mM of $MgSO_4$, 0,44mM of each dNTP (dATP, dCTP, dGTP, dTTP), 10X thermo Pol buffer, 0,5µM of each primer, 4 units of *Bst* DNA Polymerase, large fragment (New England Biolabs) and 1,35-18ng/µl of bacterial gDNA in a final volume of 20µl. The melting temperatures of various along optimisation tested primers shown in table 6 range from 70,15± 3,35°C. The reaction was placed on the GelAmp® PCR system 2700 and run at a constant temperature of 65°C for 90 and/or 180 minutes. A negative control was implemented for all experiments by adding 1µl of nuclease-free water instead of template-DNA.

Table 6. Specific primers that target the SHV-12 gene from *K. pneumoniae* strain 10/151

Primer set 1	Sequence (5′ → 3′)	Size (bp)	Product (bp)
SHV F1	TCGCCGGTCAGCGAAAAACA	20	75
SHV R1	GCTCATGGTAATGGCGGCG	19	
SHV F2	CCAGCAGGATCTGGTGGACT	20	119
SHV R2	GCAGATTGGCGGCGCTGT	18	
SHV F3	GCTGGAGCGAAAGATCCACT	20	160
SHV R3	GCCGACGGTGGCCAGNA	17	

Primer set 2			
SHV F4	CTGAATGAGGCGCTTCCCG	19	84
SHV R4	GCTGGTCAGCAGCTTGCG	18	
SHV F5	CTTGACCGCTGGGAAACGG	19	124
SHV R5	AACGGGCGCTCAGACGCT	18	
SHV F6	GACGACAACGTCACCCGC	18	171
SHV R6	CACCATCCACTGCAGCAGCT	20	

Several steps along the MDA procedure were studied in detail to optimise the isothermal multiplexed reaction.

Gradient MDA

A gradient MDA was performed with temperatures ranging from 53°C to 75°C to select the optimal working temperature for the *Bst* DNA polymerase, large fragment. The multiplexed reaction was generated with two sets of primer as listed in table 6. Each reaction comprised 1,35ng of genomic DNA of *K. pneumoniae* strain 10/151, which corresponded to a copy number of 2,5* 10^5 in a 20µl solution (http://cels.uri.edu/gsc/cndna.html). *Bst* DNA polymerase owns an adequate strand displacement activity and polymerase function between 60,4± 5,6°C but drops its activity completely at temperatures higher than 68°C.

Pre-heating step

The formation of single-stranded DNA is required for a highly efficient primer binding to the target sequence. An initial pre-heating step at 95°C for five minutes effects the destabilization of the double helix and breaks up hydrogen bonds between complementary bases. The *Bst* DNA polymerase, large fragment was added after the pre-heating phase because the enzyme does not resist an initial pre-heating step compared to novel *Taq* DNA polymerase mutants [24].

Intercalating DNA oligonucleotides

The usage of short DNA oligonucleotides in an isothermal reaction as pictured in figure 7 is suggested to increase the stability of single-stranded DNA and thereby to improve primer annealing. Newly designed DNA clamps consisted of a 20 thymidine (20T) spacer region that created distance of the DNA double helix close to the SHV target region. Three DNA

clamps listed in table 7 hybridized to complementary sequences and support the *Bst* DNA polymerase activity during the reaction.

Table 7. Target sites of DNA clamps in SHV-12 gene locus

Name	Sequence (5′→ 3′)	Position	Length (bp)
SHV Clamp 214	CTGCGGCGCAGTGCTTTTTTTTTTTTTTTTTTTTCGCGCCCAC CTACG	214-243	48
SHV Clamp 418	CCCGCAGGATTGACTTTTTTTTTTTTTTTTTTTTAACGCGGT CTAGCC	418-452	49
SHV Clamp 630	GGTCGCCGGACCGTTTTTTTTTTTTTTTTTTTTAGGCGAGG CACGACGGCCG	630-665	54

DNA target sites of clamp 1 and 2 are located outside of the primer binding sequences. Intercalating oligonucleotides do not overlap with primer target sites. The online tool OligoAnalyzer 3.1 was used for the design of the intercalating DNA clamps, which is provided by the company Integrated DNA Technology (IDT). The range of the individual melting temperatures of the DNA clamps is between 78,45± 1,35°C having a GC content of 41,25± 4,55%.

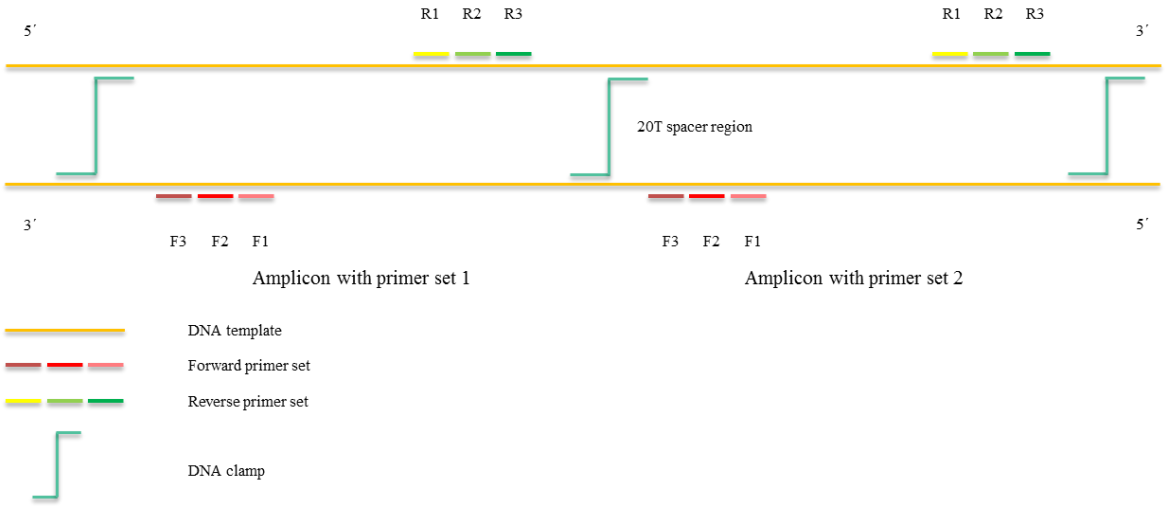


Figure 7. Scheme of intercalating DNA clamps in a multiplexed reaction
Short DNA oligonucleotides support single-strand stability within a 20T spacer region to sustain specific primer annealing to the target region. Those DNA clamps are located nearby the primer binding sites.

Helicase-dependent amplification

A thermostable helicase guarantees higher amount of amplification products in MDA by unwinding the double helix in order to support the *Bst* DNA polymerase during synthesis. The reaction setup was guided by the protocol of the helicase vendor BIOHELIX. The entire solution included nuclease-free water, 10X annealing buffer II, 5mM $MgSO_4$, 40mM NaCl,

0,25µM of SHV F4 and R4 primer in table 6, 1,4µl Iso Amp dNTPs solution and 1,75µl of the Iso Amp enzyme mix. Finally, 0,5µl of SHV-12 PCR product (stock: 1,35ng/µl corresponds to 1,45* 10^9 copies) was added to reach a volume of 20µl per sample. The tubes were placed into the thermocycler. Moreover, the reaction was initiated and run at 65°C for 90 minutes.

Restriction digest

Bacterial DNA was digested with the endonuclease *AluI* to increase the efficiency of primer binding by shortening complementary target regions. The restriction enzyme *AluI*, isolated

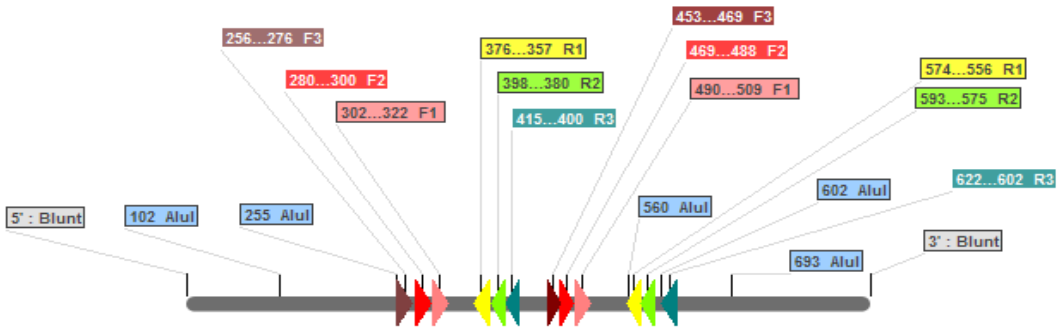


Figure 8. Position of *AluI* restriction sites in the SHV β-lactamase gene
This image represents DNA restriction sites of the restriction enzyme *Alu*I. *Alu*I restriction endonuclease cuts the genomic DNA at the outer edge of the target sequence, thus, it can be amplified with primer set 1 (F1-3; R1-3). Target regions for primer set 2 is cut centrally by this restriction nuclease at the binding position of R1 of primer mix 2.

from *Arthrobacter luteus*, cuts at restriction recognition sites 5´ AG| TC 3´ (3´ TC| AG 5´) and generates blunt-ends at the target sequence as demonstrated with the Serial Cloner 2.5 software. A 20µl digestion reaction included 16µl of nuclease-free water, 10X tango buffer, 1µl of *AluI* endonuclease and 34,1ng of genomic DNA of *K. pneumonia* strain 10/151. Nuclease-free water instead of bacterial DNA was added as negative control. The solution was incubated for one hour at 37°C. To inactivate the restriction enzyme, the reaction was heated up for 10 minutes at 75°C (Thermo Fisher, 2012). Furthermore, a PCR and MDA was performed using the restricted products as template DNA.

3.2.4. Polymerase chain reaction

Classical PCR was carried out with the Mastermix 16S complete (DNA-free) preparation kit supplied by Molzym, life science. The 20µl reaction required nuclease-free water, 10x QIAGEN buffer, 0,437mM each dNTP, 2,5 units of supplied Mol*Taq* 16S DNA polymerase,

0,5µM each primer and approximately 15-30ng of the target-DNA. The PCR reaction started with a 5 minutes initialization step at 95°C, continued by 25 cycles comprising 30 seconds at 95°C for denaturation, an annealing step for 50 seconds at 55°C, a elongation phase for 30 seconds at 72°C and finally a final extension for 5 minutes at 72°C. HotStar*Taq* polymerase was used as this enzyme which requires a 95°C initial activation step for 15 minutes in multiplex reaction and was suggested to display less unspecific primer annealing and fewer primer dimerization (QIAGEN kit). All MDA results were applied to a 1% agarose gel electrophoresis.

3.2.5. Universal primers in PCR and MDA

The yield of the amplification reaction can be increased within a second amplification step by using universal primers in high concentration. Specific primers consist a target-specific sequence at the 3´ end and a universal sequence region at the 5´end as listed in table 9. The universal sequence is unique and has no binding properties to any microbial nucleotide sequence available on the NCBI database (table 8). As DNA template, β-lactamase genes of

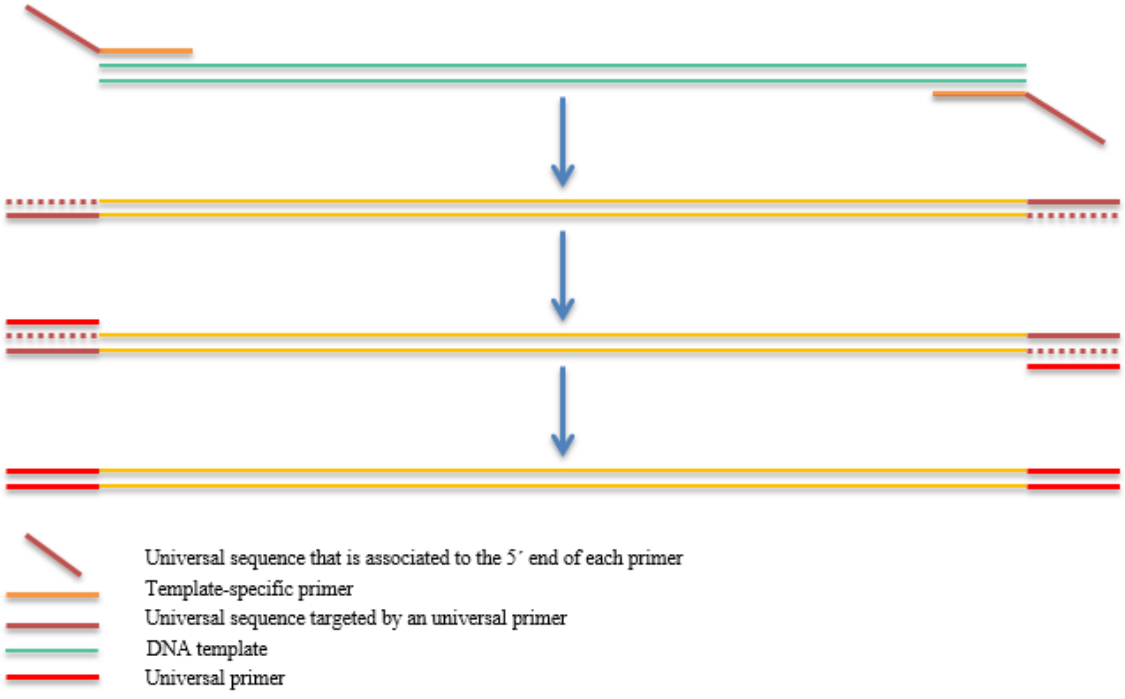


Figure 9. Multiple strand displacement amplification within the use of universal primers
This image demonstrates an amplification process of β-lactamase genes with specific und unspecific primers. Primarily, a target-specific primer that consists of an unspecific sequence on the 5´ end, binds specifically to the DNA target region. In the second amplification step, universal primers in high concentration are considered to hybridize to the complementary unbound sequence (dark red region) to provide an enhanced amplification output. In the end, the amplicon results in a larger fragment size, as it contains an associated sequence of the universal primer.

bacterial organisms that were amplified include resistance genes such as SHV-12, CTX-M-2, CTX-M-1/M-15, DHA-1 and OXA-1. The specific primers consist a melting temperature ranging from 75,85± 2,25°C, whereas universal oligonucleotides with various sequence lengths hold melting temperatures of 51,15± 9,38°C. All primers listed in table 9 were created with the online tool PrimerBLAST.

Table 8. Universal primer sequences

No.	Sequence (5´→ 3´)	Size (bp)	Tm (°C)
1	GCT AAG ATG ATG AGG GTT CAG CTA GGC	27	60,5
2	GCT AAG ATG ATG AGG GTT CAG CTA GG	26	58,5
3	GCT AAG ATG ATG AGG GTT CAG CTA G	25	56,8
4	GCT AAG ATG ATG AGG GTT CAG CTA	24	56,2
5	GCT AAG ATG ATG AGG GTT CAG CT	23	56,6
6	GCT AAG ATG ATG AGG GTT CAG C	22	55,5
7	GCT AAG ATG ATG AGG GTT CAG	21	52,7
8	GCT AAG ATG ATG AGG GTT CA	20	51,7
9	GCT AAG ATG ATG AGG GTT C	19	49,9
10	GCT AAG ATG ATG AGG GTT	18	48,6
11	GCT AAG ATG ATG AGG GT	17	47,8
12	GCT AAG ATG ATG AGG G	16	45,6
13	GCT AAG ATG ATG AGG	15	41,8

Table 9. List of specific primers

No.	Primer (binding position)	Sequence (5´→ 3´)	Size (bp)	Product (bp)
1	SHV-112 F1 (5-23)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GAA ACG GAA CTG AAT GAG G	46	356
	SHV-112 R1 (306-288)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GTA TCC CGC AGA TAA ATC A	46	
2	SHV-112 F2 (44-58)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GAC ACC ACT ACC CCG	42	289
	SHV-112 R2 (278-262)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GCT CTG CTT TGT TAT CC	44	
3	SHV-112 F3 (81-95)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GCA AGC TGC TGA CCA	42	182
	SHV-112 R3 (208-193)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ATC GGC GAT AAA CCA G	43	

1	CTX-M-2 F1 (388-406)	GCT AAG ATG ATG AGG GTT CAG CTA GGC AGT ATA GCG ACA ATA CTG C	46	364
	CTX-M-2 R1 (697-681)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ACT ACC CAT GAT TTC GG	44	
2	CTX-M-2 F2 (520-534)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CCA TTC CAG GCG ACC	42	251
	CTX-M-2 R2 (716-702)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GCT GCC GGT TTT ATC	42	
3	CTX-M-2 F3 (504-518)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CCC ACG CTC AAT ACC	42	251
	CTX-M-2 R3 (700-685)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CCC ACT ACC CAT GAT T	43	
1	CTX-M-1 /15 F1 (160-179)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ATT CGC AAA TAC TTT ATC G	46	450
	CTX-M-1 /15 R1 (555-541)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CAT TGC CCG AGG TGA	42	
2	CTX-M-1 /15 F2 (204-221)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CAG CAC CAG TAA AGT GAT	45	427
	CTX-M-1 /15 R2 (576-559)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CAG ATT ACG CAG AGT TTG	45	
3	CTX-M-1 /15 F3 (140-159)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CAT TGA TTA ACA CAG CAG AT	47	553
	CTX-M-1 /15 R3 (638-622)	GCT AAG ATG ATG AGG GTT CAG CTA GGC TTG CCT TTC ATC CAT GT	44	
1	DHA-1 F1 (576-595)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ATC TCA CAC CTT TAT TAC TG	47	354
	DHA-1 R1 (875-856)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ATG ATC ATA TCT TTC TGC TG	47	
2	DHA-1 F2 (343-359)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CTG GAT CTG GCT ACC TA	44	521
	DHA-1 R2 (809-790)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GTT TTA TAG TAG CGG GTC TG	47	
3	DHA-1 Padlock F3 (305-320)*	GCT AAG ATG ATG AGG GTT CAG CTA GGC CGG AGC TGG CTC TGC	42	776
	DHA-1 Padlock R3 (1011-1027)*	GCT AAG ATG ATG AGG GTT CAG CTA GGC CGC CAC CTG TTT TTC C	43	
1	OXA-1 F1 (202-221)	GCT AAG ATG ATG AGG GTT CAG CTA GGC AAT CAA GAA TTA TCT CAA AG	47	290

	OXA-1 R1 (437-418)	GCT AAG ATG ATG AGG GTT CAG CTA GGC ACA GTT TTG TAC TAT TAT CC	46	
2	OXA-1 F2 (170-187)	GCT AAG ATG ATG AGG GTT CAG CTA GGC CAA GAA ATA ACC CAA AAA	45	359
	OXA-1 R2 (474-457)	GCT AAG ATG ATG AGG GTT CAG CTA GGC GTT CTA TTT GCT GTG AAT	45	
3	OXA-1 Padlock F1 (149-169)*	GCT AAG ATG ATG AGG GTT CAG CTA GGC TTT TCT GTT GTT TGG GTT TC	47	475
	OXA-1 Padlock R1 (552-561)*	GCT AAG ATG ATG AGG GTT CAG CTA GGC AAT TCG ACC CCA AGT TTC	45	

Underlined sequence indicate specific target-binding regions, *Padlock primers used in previous multiplex assays [4]

3.2.6. Agarose gel electrophoresis

Gel electrophoresis is a widely used technique to analyze DNA, RNA and proteins by separating them under the influence of an electric field. Applying this method the sizes of PCR products can be easily determined [20]. For preparing a 1% agarose gel, 0,80g agarose powder was dissolved in 80ml 1xTBE buffer and heated up in the microwave. The solution was cooled down to room temperature before SYBR safe dye was able to stain the gel. The solution which was then poured into the gel aperture. The gel was solidified before samples were loaded. The electrophorese run was set to 180V for 50 minutes. In the end, defined band patterns of the amplification product were displayed via UV-spectrophotometry. The image resolution could be optimised by varying the light settings such as the aperture, zoom and light intensity.

3.3. Amplicon sequencing with the Ion Torrent PGM

For unknown reasons, the multiple displacement amplification assay resulted in false-positive outcomes for most negative controls. In fact, for corresponding control reactions appears with a strong smear on an agarose gel picture. To gain clarity on theses unspecific amplification products, amplicon sequencing was performed with the Ion Torrent PGM system using a 314™ v2 microchip. In this case, six different MDA reactions with six corresponding negative controls were selected for sequencing. All preparation steps were done following the Ion Torrent manuals.

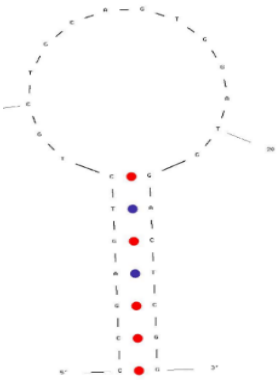
3.4. Design of molecular beacon

A linear amplification of a target of interest within a multiple strand displacement reaction enables DNA quantification in retrospect.

Molecular beacons are characterized by a short single stranded oligonucleotide in the size of 15 to 30 nucleotides that forms a stem-loop DNA structure [40]. The stem consists of a 6-FAM fluorophore on the 5´ end of the DNA strand and an Iowa Black quencher molecule on the 3´ end. In closed stem conformation, the quencher hinders light emission of the fluorophore by sharing electrons transiently [38]. The recognition of target DNA forces molecular beacon to emit light which is detectable. The stem contains 7 nucleotides including 5xGC and 2xAT interaction. The melting temperature of the MB is at 67,1°C, which is slightly higher than the annealing temperature of 65°C. A 20µl reaction included DNA-free water, 2,5x mastermix 16S Basic, 0,05µM per primer, 0,3µM of molecular beacon, 4 units of the *Bst* DNA polymerase, large fragment and 0,3ng DNA of *K. pneumoniae* strain 10/151.

Table 10. Molecular beacon design using OligoAnalyzer 3.1 (IDT)

Name	Probe (5´→ 3´)	Length	Tm (°C)	GC %	Stem (nt)
#1 SHV beacon	5´-6-FAMCCGAGTCTGCTGCAGTGGATGGACTCGG-IBFQ-3´	28bp	67,1	64,3	7



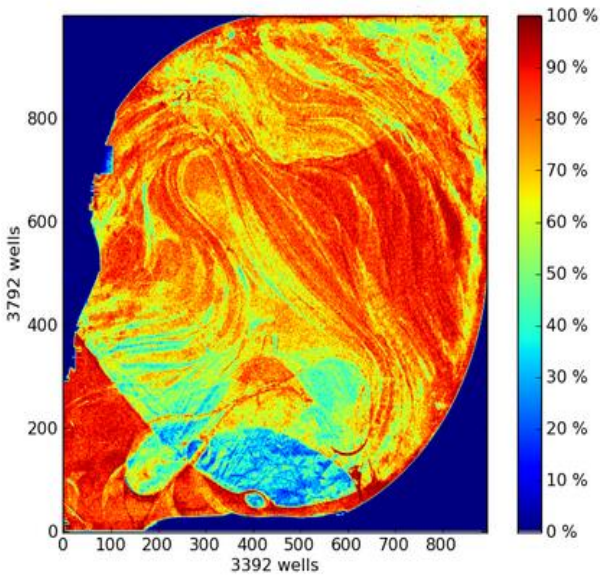
4. Results & Discussion

4.1. Whole genome shotgun sequencing of three *Acanthamoeba* specimen

In this study, the genomic DNA of three different *Acanthamoeba* strains was sequenced with the Ion Torrent PGM system to reveal genetic variances between pathogenic and non-pathogenic species and to reveal two de novo assemblies for strain Pb30/40 and strain 72/2.

The first sequencing outcome that is presented in figure 10 resulted in a chip load density of 69% out of 11 million possible wells onto a 318™ v2 semiconductor chip with template-positive ISPs of genomic material of strains Pb30/40, 1BU and 72/2. Within a seven hours sequencing run, 3.495,210 reads with 847M bases have been delivered, including 45% usable reads for all *Acanthamoeba* samples. Red areas in Fig. 10 indicate a sufficient chip load with target-associated beads, whereas bluish regions represent empty wells on the microchip. On average, one million reads with more than 250 million bases could be

A



B

Barcode Name	Sample	Bases	>=Q20 Bases	Reads	Mean Read Length	Read Length Histogram
No barcode	None	22,103,674	18,608,472	109,458	202 bp	
IonXpress_049	P30_40	283,590,659	245,198,994	1,155,762	245 bp	
IonXpress_050	72_2	283,567,388	245,399,123	1,152,447	246 bp	
IonXpress_051	1BU	257,265,411	222,594,490	1,073,204	240 bp	

Figure 10: Heat map of the 318™ v2 semiconductor microchip
The loading density with Ion Sphere particles on the Ion chip 318™ v2 chip is displayed in this image as well as respective sequencing information. **A)** 318™ v2 chip load density for pooled DNA libraries of Pb30/40, 1BU and 72/2. **B)** The statistical outcome per library involves the read length dissemination, total number of bases and the mean fragment sizes.

generated per sample. The read length histogram in Fig. 10-B underlines the size distribution of the library indicating that the majority of fragments were around 350 bp.

Moreover, the sequencing quality was additionally monitored and checked with an supplied Ion Torrent test fragment.

The following analysis of all shotgun-sequencing data by Blast pointed out that only 3 of 120 largest contigs of *Acanthamoeba* strain Pb30/40 (T7), which had a length between 28.612bp to 9.014bp, could be matched to the *Acanthamoeba castellanii* Neff strain (T4) reference genome. Interestingly, the majority of all contigs belonged to the common endosymbiont *Parachlamydia acanthamoeba* which is known to be an obligate bacterium living within several *Amoeba* species. As earlier reported by Greub, this emerging microbe has the potential to cause respiratory tract infections such as pneumonia and bronchitis in human [17]. For more than one billion year co-evolution of ancestral *Chlamydia* exists within eukaryotic organisms [17]. As presented in table 11, greater sequence comparability to the *A. castellanii* Neff genome could be found for strain 72/2 with 72 out of 91 contigs and for strain 1BU with 84 out of 101 contigs. A sequence consistency for *A. lenticulata* strain 1BU was expected since this species belongs to the genotype T4, equally to the reference genome.

Table 11. Blast results for the largest 120 contigs

Samples	Total hits	*A. castellanii* hits
Pb30/40	87/120	3/87
72/2	91/120	72/91
1BU	101/120	84/101

The genome mapping of *A. c.* strain 1BU to the reference sequence resulted in 44,7% with untrimmed fragments, as for Pb30/40 only 6,68% conformity to the 42,0198Mb genome of *A. castellanii* Neff strain was observed. Moreover, a de novo assembly was provided by MIRA 4 for the sample 1BU that contains a total length of 31,3Mb to serve as mapping template for both other strains. The mapping performance to the new reference genome of 1BU using raw sequencing data lead to 16,75% identity for Pb30/40 and 38,10% for 72/2. In this aspect, the sequencing coverage with *Acanthamoeba*-related reads and the mapping quality was insufficient to declare either large-scaled genomic variations nor any sequence similarities to the reference genome. As a consequence, all samples were re-sequenced using

this time one 318™ v2 microchip per strain to ensure an extended sequencing coverage for the respective sample.

4.2. Results of the resequencing procedure

4.2.1. DNA quality test

Due to an increased DNA contamination of *Acanthamoeba* strain Pb30/40 in the first sequencing process, the growth of this species on bacteria-layered agar involved the pre-

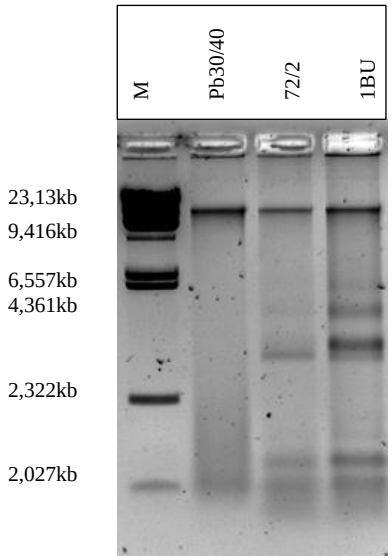


Figure 11. DNA quality test for Acanthamoeba strain Pb30/40, 72/2 and 1BU by gel electrophoresis
The genomic DNA of each strain is detectable at approximately 22 kb. Unpreventable occurring band patterns for sample 72/2 and 1BU may indicate short nucleic acid sequences of symbiotic organisms living within these host cells or partially broken genomic DNA sequences. A Lambda/HindIII gene ruler was loaded as comparative marker.

treatment with 100µg/ml gentamycin (Sigma-Aldrich, St. Louis, MO, US) to remove bacterial endosymbionts. This cleaning assumed to influence the sequencing results in a positive way. In this case, the sample Pb30/40 was treated with the peptidoglycan antibiotic substrate for 10 days, before genomic DNA was performed with the QIAmp DNA isolation kit (Qiagen, Hilden, Germany). Apparently, *Acanthamoeba* strain 72/2 and 1BU require and demand the association with endosymbiotic bacteria for optimal growth, as these strains could not recover from antibacterial pre-treatment. In figure 11, the initial analysis of the DNA quality from all *Acanthamoeba* strains is shown after finishing a 1% agarose gel electrophoresis. Distinct patterns of bands for strain 72/2 and 1BU still indicate DNA contamination probably with bacterial 23 rRNA, certain plasmids at lane 2,1kb or shattered

gDNA of *Acanthamoeba* as displayed in figure 11. A Lambda/ HindIII ladder ranging from 23.130bp to 2.027bp was loaded onto the 1% agarose gel.

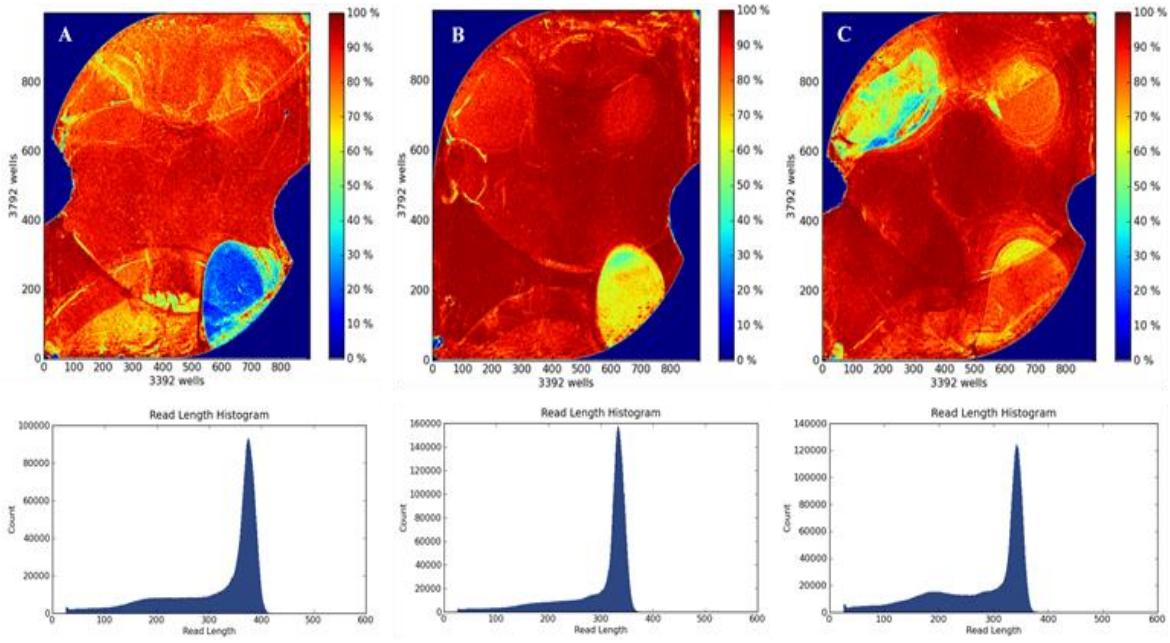


Figure 12. Loading density of Ion sphere particles on 318™ v2 semiconductor chip and respective read length distribution

Template-positive ISPs were filled up to **(A)** 78% for sample Pb30/40, **(B)** 90% for strain 1BU and **(C)** 85% for strain 72/2 of all wells on the microchip. Corresponding fragment size distribution is displayed in the histogram beneath with median length between 320 to 410 base pairs.

One well-loaded 318™ v2 semiconductor chip for each sample was applied to the Ion Torrent PGM system shown in figure 12. The Qubit fluorophore measurement calculated 19,71% of templated-positive ISPs for Pb30/40 and 1BU. The recommended percentage is between 10 to 30%. In contrast to the first sequencing run, an optimised chip load and a substantial yield of sequencing output could be achieved. On average, 5,98M single-end reads were generated per library with 63% of usable reads in total. In table 12, all re-sequencing outcomes for each microchip are listed, including computations from the QUAST report which were calculated by the Spades Assembler plug-in on the Ion Torrent Server. Only a small number of reads were classified as low quality reads. On average, less than ¼ of all sequences were identified as polyclonal ISPs. The sequence distribution of sheared libraries is displayed in the read length histogram in figure 12. Moreover, the genome coverage resulted in a coverage of 44x, on average. Basically, the calculation of the theoretical sequencing depth, mentioned in a publication from Sims et al. is defined as LN/G, where L is the number of reads, N the mean read length and G the genome size (bp) of the reference genome [49].

Table 12. Summary of Ion Torrent PGM re-sequencing performance

Strain	Pb30/40	1BU	72/2
Total bases (G)	1,62	1,85	1,75
Total reads	5,178,749	6,350,920	6,434,049
Usable reads (%)	59	63	67
Enrichment (%)	100	100	100
Clonal ISPs (%)	66	68	72
Low quality reads (%)	10	7	6
Median read length (bp)	358	326	317
Total contigs	31,138	26,712	26,243
Total length (b)	79,100,617	55,046,480	45,471,706
N50	3,585	3,071	2,481
GC (%)	43,48	57,96	57,5

The analysis was continued in FastQC by uploading the adapter-trimmed version of all reads as BAM-file which was downloaded from the Ion Torrent platform and indicates single base pair quality values for each sequence.

4.2.2. Statistical evaluation with FastQC

An upload of individual BAM-files to the FastQC tool allows a quick and accurate analysis of huge sequencing data. High-quality reads for sample Pb30/40 were evaluated ranging from 100 bp to 380 bp in length. As a result, illustrated in figure 13, Pb30/40 (D), the final observed GC-content of 43,48% adapted well to the theoretical value as soon as low-quality reads were trimmed. Certain reads for strain 1BU that contain a length between 100 bp and 365 bp were sorted as high-quality sequences. Significant variances in the GC% distribution may indicate DNA contamination by random nucleic acids as an uncommon peak occurs in addition. Furthermore, the mean GC value of all sequences exceeded the theoretical estimation as shown in figure 13, (D). A prominent contamination was observed for 72/2, as the GC remains disproportional after the selection of high-quality read sizes between 100 bp to 320 bp, which is graphically presented in figure 13, 72/2 (A, D). As evident from the gel electrophoresis picture in figure 11, were *Acanthamoeba* strain 72/2 and strain 1BU showed more DNA bands with any antimicrobial drug not having be treated during growth phase, these species seem to be still associated with various symbiotic organisms that may cause impurity of the extracted DNA. Moreover, the first 20 bp of all high-quality reads were cut

off in addition due to the ambiguous content of nucleotide mismatches (figure 13, Pb30/40 (C), 1BU (C) and 72/2 (C)). The mean quality values of reads across all fragments and strains was found to be very similar as shown in figure 13 (B).

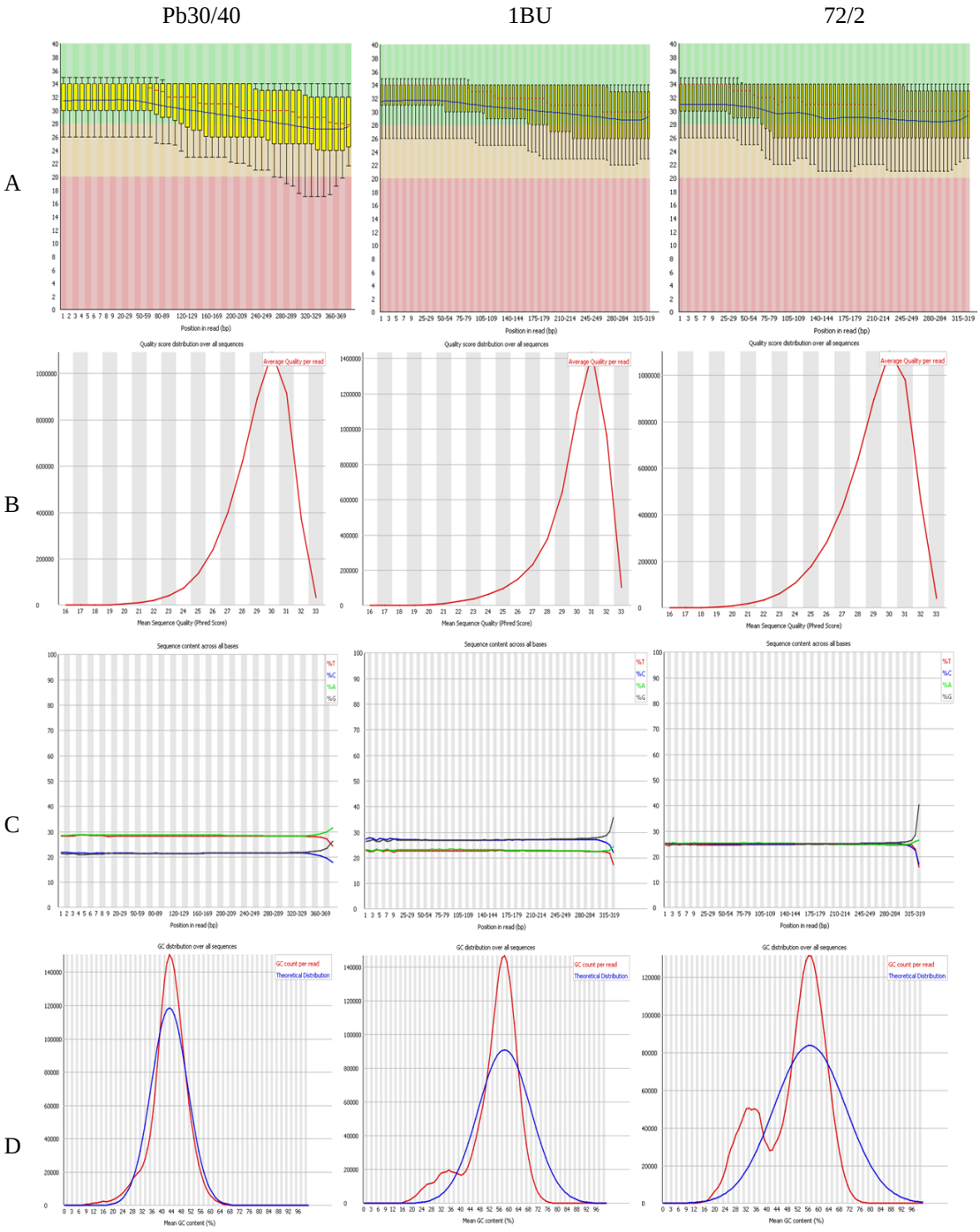


Figure 13. Statistical overview of trimmed reads after the resequencing process
These graphics illustrate the statistical analysis by FastQC of established reads for Pb30/40, 1BU and 72/2 after trimming, stating **(A)** the base pair quality scores for all reads, **(B)** the read length distribution, **(C)** base pair matches per sequence and **(D)** the GC% content.

4.2.3. Metagenomics analysis investigated with Kraken program

The taxonomic clustering of all reads was performed by "Kraken". To start with, we used certain GenBank databases for bacterial, viral, archaeal species, *Mus musculus* and human for metagenomic analysis. The classification of microbial species is based on the mapping of 31-mer sequences to the last common ancestors (LCA) that contains these 31-mer query sequences in the database. Investigations on sequences with 31 bp in length ensure highest accuracy, sensitivity and speed (~1,3Mb/min) compared to other metagenomic programs [57]. Referring to table 13, 487,815 reads of 1BU strain were aligned to the nucleotide database on NCBI, resulting in 57% matched sequences to the reference genome.

Table 13. Metagenomics data on high-quality reads using Kraken

Strain	Pb30/40	1BU	72/2
Trimmed reads	4,852,013	5,247,865	5,243,271
All reads classified by BLAST	564,501	487,815	276,337
A. castellanii str. Neff related reads	4,989	282,039	23,884
Human related hits	140,628	145,630	121,171
P. acanthamoebae related seq.	384,513	35	10
Other bacterial species	34,371	60,111	131,272

Lower alignment hits were determine for the two other samples. Although Pb30/40 was initially treated with antimicrobial substances to remove symbionts, 68,11% of NCBI-classified reads still belong to the bacteria *P. acanthamoebae*. Nevertheless, 89% of reads of Pb30/40 remain unclassified and were used to perform a de novo assembly.

Previously, high-quality sequences were aligned to the GenBank database on NCBI and were arranged into suborders for individual specimen. Well-trimmed reads represented in red columns contribute novel sequences. As pictured in figure 14., mainly sequences of strain 1BU could be matched to *A. c.* Neff genome and was illustrated with IGV (Version 4.2.4.). Consistent sample heterogenicity was observed for Pb30/40, containing sequences of *P. acanthamoebae* UV-7 despite prior antimicrobial therapy. In fact, extracted DNA of strain Pb30/40 contained high quantities of genetic material from endosymbiotic organisms. The distantly-related genotypes of Pb30/40 (T7) and 72/2 (T5) possess low sequence similarity

to the genus *Acanthamoeba* as displayed in figure 14. As a consequence, we performed a de novo assembly for Pb30/40 and 72/2.

4.2.4. Genome de novo assembly for species Pb30/40 and 72/2

Based on the results mentioned in the previous paragraphs, two newly evolved genomes were generated for *Acanthamoeba comandoni* strain Pb30/40 and *Acanthamoeba lenticulata* strain 72/2 conducting a de novo assembly using the Mimicking Intelligent Read Assembly 4 (MIRA). This program requires the initial input of raw sequencing data and sorts high confidence regions automatically as breakpoint to form larger contigs.

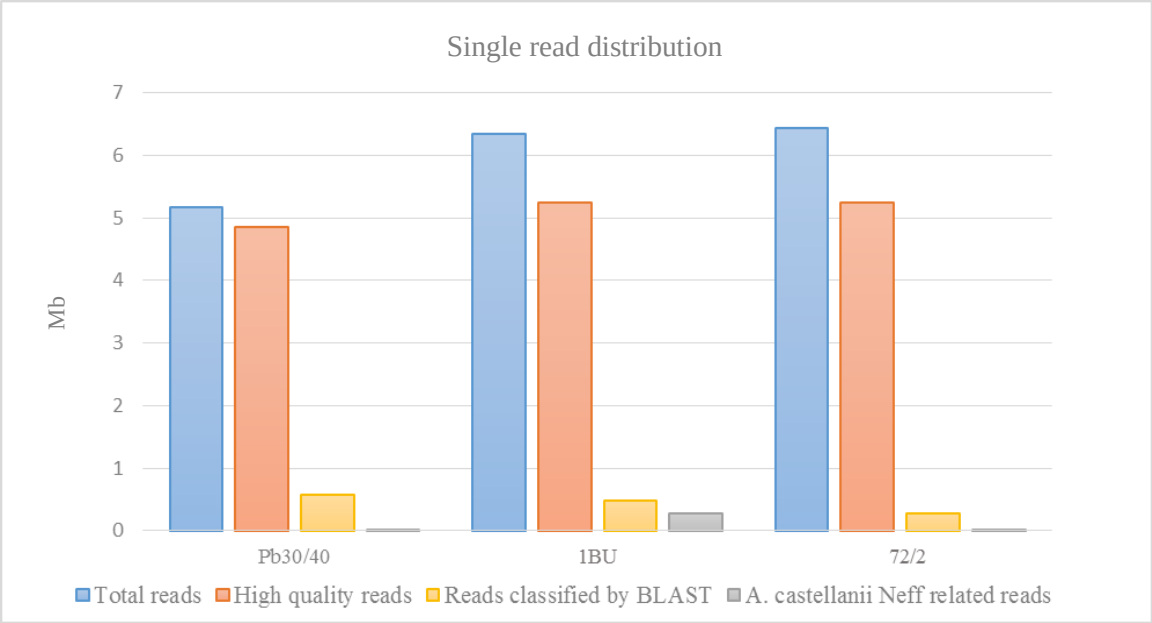


Figure 14. Whole-genome sequences sorted by Kraken according to the NCBI database
This bar chart illustrates the individual read distribution that was aligned to selected databases by Blast. The main number of high-quality reads remained unmatched for all samples. Most *A. castellanii* Neff-related reads were evaluated for strain 1BU.

The complete project of sample Pb30/40 (T7) consisted of a sequence length of 86,262,821 bases with a G+C content of 43,65%. Moreover, the final draft genome of strain 72/2 resulted in a length of 74,699,519 bp with a total G+C content of 56,91%. The evaluation of the N50 statistics define a 50% assembly with sequences that are equal to or larger than the defined N length. In this respect, Pb30/40 presents a N50 value of 2281 bp within a total number of 9583 contigs whereas for strain 72/2, the N50 value appears to be least 2,500 bp with a total of 7508 sequences.

The following image in figure 16 represents the calculated final GC distribution of high-quality contigs that were assembled by MIRA for Pb30/40 and 72/2. As pictured in figure 16, even after trimming, the mean G+C content for strain 72/2 remain increased with approximately 50 to 70% compared to the computed theoretical distribution. The antibiotic treatment inhibited the growth of sub-cultures for strain Pb30/40 and 72/2, whereas a extraction of genomic DNA was not possible without having foreign DNA of intra- and extracellular bacteria co-isolated. In this aspect, we assume no DNA contamination by symbiotic species as we performed trimming of low quality reads. As a result for well-trimming reads of strain 72/2, an equal G+C distribution has been reported for the first de novo assembly of *A. lenticulata* that is available on the NCBI platform.. Anyhow Moreover, the GC value of 45% of all well-distributed sequences for Pb30/40 indicates that the pre-treatment with 100 µg/ml gentamycin had a beneficial impact on DNA quality. In this aspect, a de novo assembly was arranged to create new genomes for Pb30/40 and 72/2.

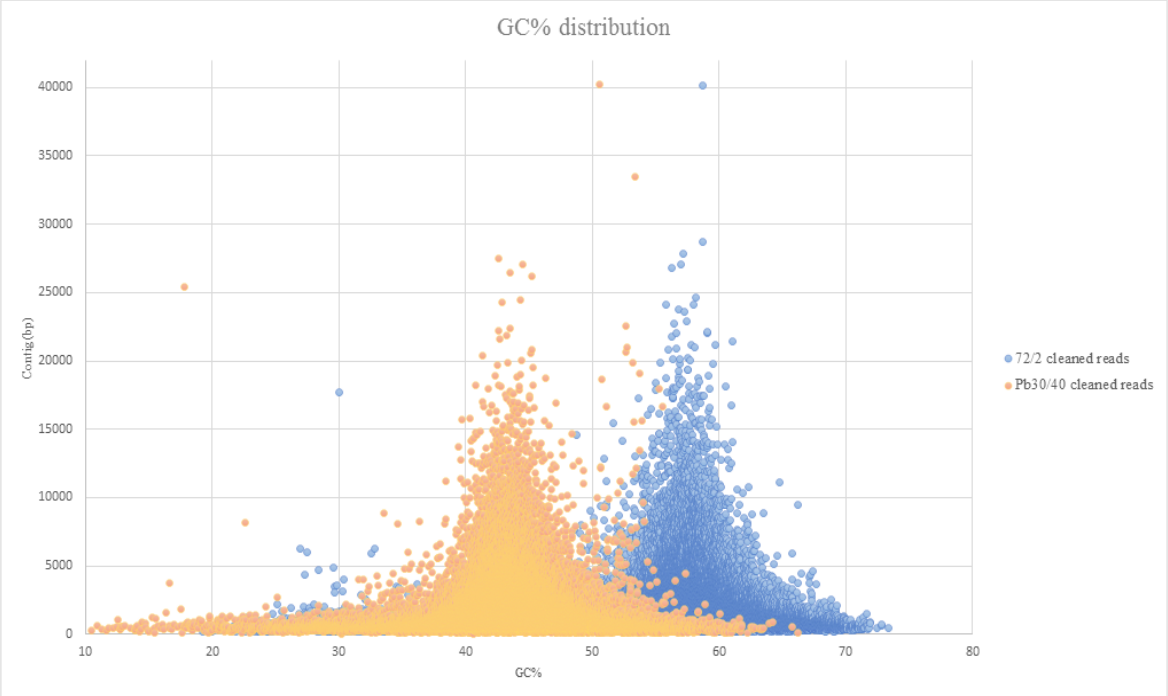


Figure 15. GC% distribution of strain 72/2 (T5) and Pb30/40 (T7)
Each spot in the graph illustrates a single read within its corresponding GC content. Dense spot areas cluster those contigs that consist similar sequence lengths and a highly related GC amount.

In the subsequent chapter, high-quality reads of the pathogenic *A. castellanii* strain 1BU were mapped to the non-pathogenic *A. c.* Neff reference genome.

4.2.5. Visualization of genome mapping using Integrative Genome Viewer (IGV)

In general, genome mapping describes a process in which newly generated NGS data are aligned to an existing template sequence to delineate genome variations such as single nucleotide polymorphisms and mutations.

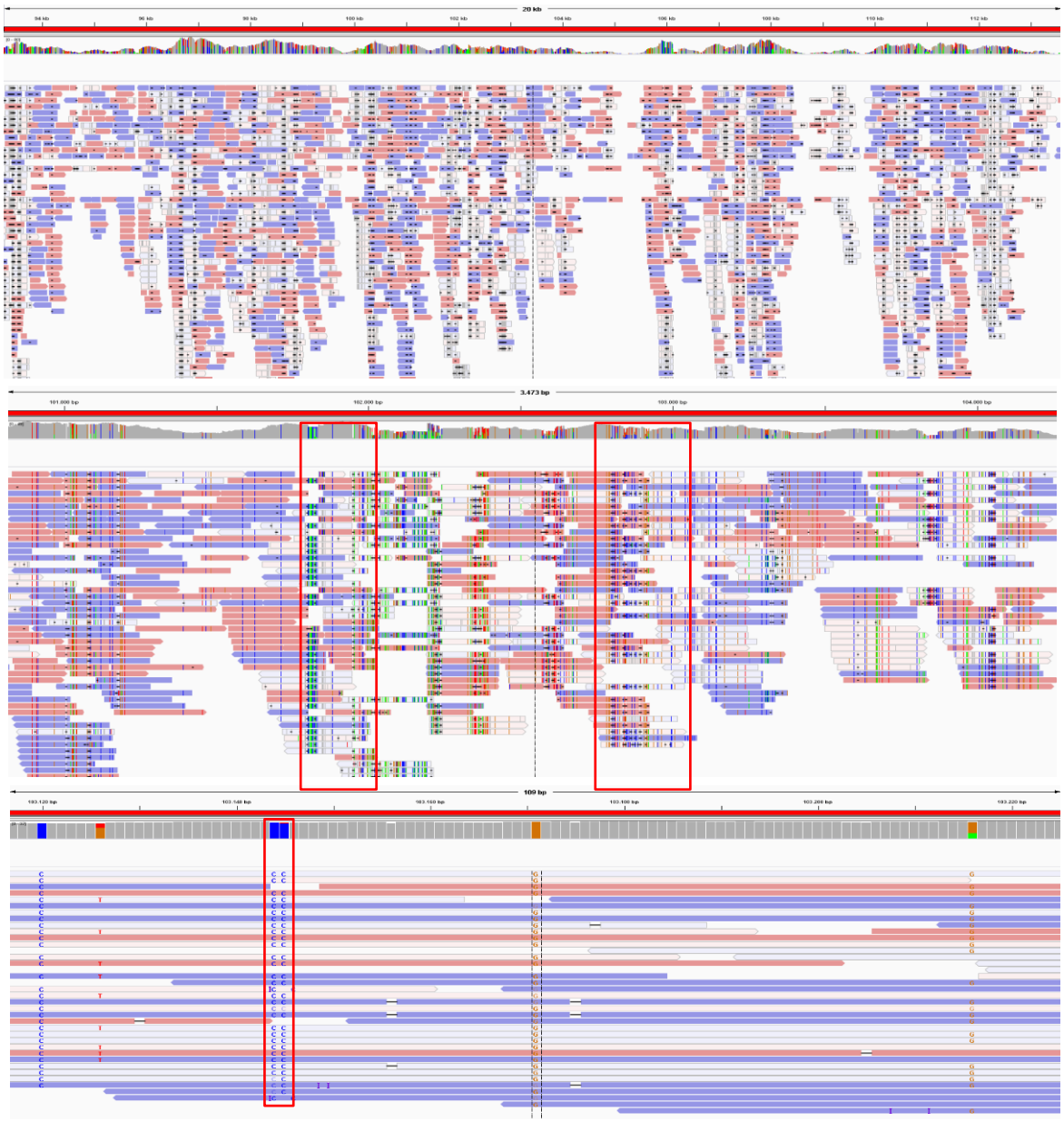


Figure 16. Visualization of genome mapping with IGV

IGV tool features gene annotations, read coverage and visual depths close to single bases. Mean red and blue colored stripes illustrate the respective orientation of mapped reads. The colorful headline demonstrates various nucleotide variations in the spotted gene region. The comparison of high-quality reads of *Acanthamoeba* strain 1BU demonstrates sequence variation in the gene region of the mannose-binding protein (MBP) compared to the reference (results not shown here).

In literature, the mannose-binding protein (MBP) and serine proteases represents an important pathogenic property of *Acanthamoeba* to its capability to associate and live within host cells [35]. These proteins have been annotated and published by K. Clarke in 2013 in full sequence lengths. Genome mapping was performed for all strains using NextGenMap-based (NGM) alignment algorithms and were visualized with the Integrative Genome Viewer. In this aspect, the pathogenic *A. castellanii* strain 1BU was mapped to the non-pathogenic *A. castellanii* Neff strain and accounted for 62% with well-trimmed reads. Both other strains that are classified as different genotypes than the reference resulted in very low coverage of 17,72% for strain 72/2. The sample Pb30/40 belongs to the T7 genotype and covered only 2,49% to the entire genome of *A. c.* Neff sequence. After the analysis of both sequencing runs, phylogenetic distantly-related genotypes result in very low genome alignment to the reference genome. This mapping results correlates well to the high dissimilarities of the 18S rRNA gene among different genotypes.

In 1990, Byers reviewed conserved sequence regions of *Acanthamoeba* that encode cytoskeleton regulatory proteins including actin, profilin and myosin all related to motility [15]. An alignment of those sequences resulted in great coverage and small variation for strain 1BU. The 506bp gene locus of profilin, for instance, a protein that inhibits the polymerization of actin, showed two alignment parts, one with 98% (293/300) sequence similarity and the other with 96% (131/137) for the 1BU strain within 13 single nucleotide substitutions and 69 unaligned bases to the reference locus (ACA1_076370).

A sequence alignment of two species is characterized as homologous if at least 70% of this sequence can be matched to a query [19]. The gene locus of one of the main virulence proteins MBP (ACA1_248600) could be matched with 88% similarity (1352/1530 bp) and 10 gap regions to the T4 reference genotype. Moreover, the genomic region of the alkaline serine protease (ACA1_278460) resulted in 94% identity to the query sequence (1331/1415 bp) and no gaps. However, the sequence coverage for strain Pb30/40 and 72/2 to the Neff strain was insufficient to declare any variation of those gene regions.

4.3. Establishment & optimisation of an isothermal amplification

4.3.1. First experiments on Multiple Strand Displacement Amplification

In this approach, we aimed to develop an isothermal amplification assay that allows multiplexed reactions at one defined reaction temperature. The thermostable *Bst* DNA polymerase (large fragment) is highly suitable for this technique and displays great strand displacement properties to break up the hydrogen bonds of the DNA double helix, if accurate reaction conditions are present [32]. Primer design represents a crucial step in all types of amplification procedures.

The first multiplexed reactions were performed with the *Bst* DNA polymerase, large fragment at 65°C and resulted in similar amplification products for both the positive and the corresponding negative controls (see Fig. 17). This observation strongly indicates sample contamination caused by pipetting or the formation of primer dimers using 0,5µM of each primer set (table 6) and Tm of 70,15± 3,35°C. A 20µl reaction included 1,35ng of gDNA of

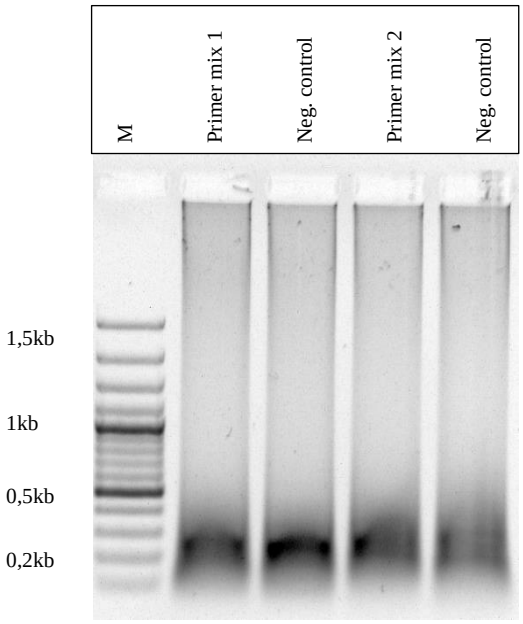


Figure 27. Multiple strand displacement amplification of the β-lactamase gene SHV-12 with two different sets of primer

The product of each isothermal reaction can be detected at a size 180 bp to 250 bp for both the positive and the corresponding negative reaction. Indefinite amplicons in negative control may occur due to insufficient and random primer binding to genomic bacterial DNA, DNA contamination or heterodimerization. PCR products were separated on a 1% agarose gel with 1x TBE buffer for 45 minutes. A 100 bp Plus gene ruler was loaded as ladder.

K. pneumoniae strain 10/151, which corresponds to a calculated copy number of $2{,}5*10^5$ (calculator: http://cels.uri.edu/gsc/cndna.html). The gel image in figure 17 demonstrates a thick smear pattern between 180 bp and 250 bp for primer mix 1 and primer mix 2 for all reactions. The expected amplicon sizes should range from 75 bp to 160 bp for primer mix 1 and from 84 bp to 171 bp for primer mix 2. Furthermore, both negative reactions resulted in false-positive band structures with same amplicon sizes as the positive controls. Paying attention to these pitfalls, various experimental setups were carried out to improve the isothermal reaction.

4.3.2. Gradient multiple displacement amplification

Performing a gradient MDA confirmed that the *Bst* DNA polymerase (large fragment) has its greatest processivity between 58 to 68,5°C. The strand displacement activity and polymerase activity of this enzyme is strongly inhibited at temperatures higher than 68,5°C. This was checked with both sets of primers listed in table 6. Multiplexed reactions at temperatures below 63°C resulted in very unspecific products (result not shown). The optimal condition was at 65°C as it is recommended by the suppliers. However, further optimisation of the multiple strand displacement amplification was performed.

4.3.3. Impact of a restriction digestion with *AluI* endonuclease

By digesting genomic bacterial DNA into smaller fragments using *AluI* endonucleases, we expected primers of set 2 to bind more efficient to the target sequence and thereby increase specificity. Previously digested DNA (34,1ng) was added to the presented PCR/ MDA reaction and resulted in a thick smear pattern at 190 bp compared to the corresponding negative control. Artificial band patterns were detectable for both polymerase chain reactions. In lane 2, an intensive product at 1,35 kb may indicate accumulation of refracted gDNA. Regarding the MDA reaction, digested 10/134 DNA in lane 4 showed a constant band smear over the whole gel lane with a slightly stronger density than the negative control reaction. Even if the specificity is far away from being the same as in PCR, a rising reaction intensity can be recognized in MDA reaction in the range between 150 bp to 500 bp. In fact, this result points out that an initial restriction digest of gDNA has no influence on an accurate primer annealing in MDA as fine as it can be achieved PCR.

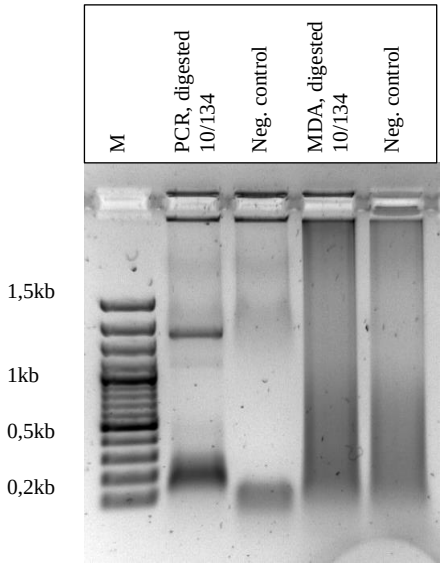


Figure 18. Comparison of PCR and MDA reaction using digested genomic DNA of *K. pneumoniae* strain 10/134

Primer binding in multiplex PCR shown in lane 2 worked out well with a defined band smear for the SHV-12 gene. The use of pre-digested bacterial gDNA in the MDA reaction could not enhance the specific primer annealing. Even though, a tighter smear pattern can be recognized for template-MDA than in the control sample. A 100bp Plus gene ruler was loaded to the 1% agarose gel.

4.3.4. Chemical substrate DMSO supports the multiple displacement amplification

Dimethyl sulfoxide or "DMSO" is an organosulfur solvent, which enhances the efficiency of a PCR reaction by decreasing secondary structures within GC-rich template regions. During the isothermal amplification assay, DMSO is considered to stabilize the denatured DNA conformation and to provide specific primer annealing. Moreover, DMSO is considered to force base pair mismatches and to reduce the fidelity power of polymerases [2].

The following gel picture in figure 19 represents an isothermal reaction at 65°C for 90 minutes with 0,5µM of primer set 2 (table 6.) and different DMSO concentrations. A 20µl reaction included 1,35ng DNA ($2,5*10^5$ copies) of *K. pneumoniae* strain 10/151. The use of 0,5% of DMSO in a 20µl reaction seems to be sufficient to improve short-period reactions as the amplified products have a length between 160 bp to 180 bp. This strand displacement reaction performs a linear amplification with multiple primers and is supposed to result in an intensive smear pattern nearby the respective product lengths. The signal intensity indicates that a low-concentrated amount of DMSO (lane 3 and 4) has provided a beneficial impact on the isothermal amplification compared to the reaction without DMSO in lane 2.

The use of DMSO higher than 1% in a short-termed amplification process (3h) decreases the *Bst* DNA polymerase activity as the density of these band pattern in lane 8, 10 and 12 gets damped. In fact, DMSO concentrations with an above average level reduces the melting temperatures of the primers and inhibit the polymerase activity [2]. A thick pocket smear in lane 2 and 4 indicates that the genomic bacterial DNA was amplified linearly during the reaction and was not inhibited by high DMSO concentrations. Experimental studies confirm that DMSO concentrations greater than 2,5% negatively influence the isothermal amplification. Within the use of 0,5% DMSO, the stability of single-stranded DNA is assured to improve specific primer annealing.

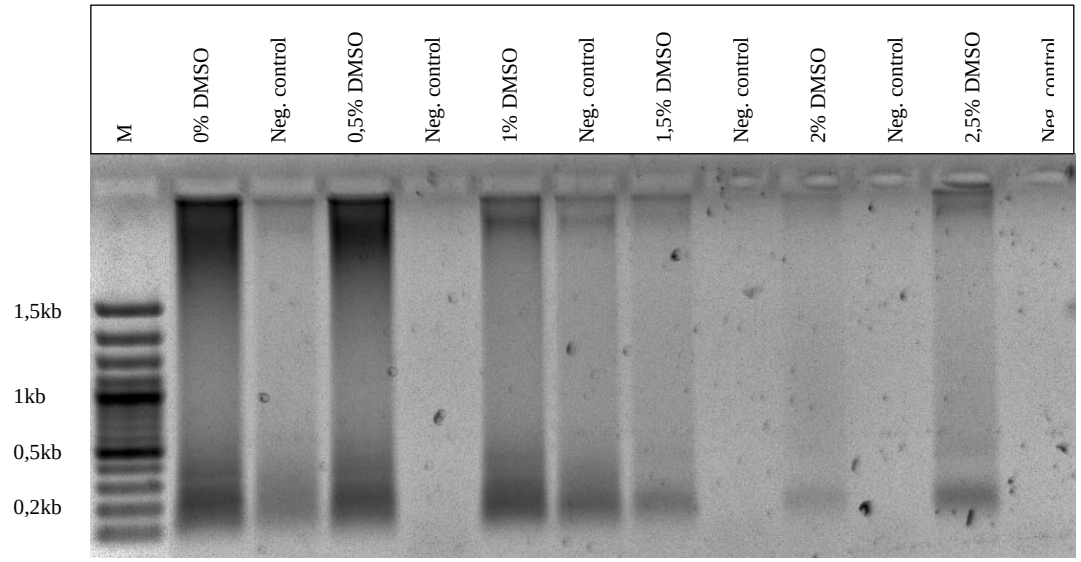


Figure 19. The use of DMSO in a multiple strand displacement amplification
In this experiment different DMSO concentration ranging from 0 to 2,5% were added to the MDA reaction. The specificity of short-termed (90 minutes) MDA reaction increases within the use of 0,5% DMSO. A 100 bp Plus gene ruler was added to the 1% agarose gel.

4.3.5. The use of intercalating DNA clamps in MDA reaction

Specific primer binding to a target-DNA is of most importance to ensure the attachment of the *Bst* DNA polymerase, large fragment in order to defined sequence region and initiate DNA extension.

Three different DNA clamps were designed to stablize the DNA regions outside the primer binding sites of the SHV locus. These clamps consist of 48 to 54 nucleotides with a Tm of 78,45± 1,35°C and contain a centred 20T spacer region displace single-stranded DNA for an efficient annealing step at 65°C. An enhanced Tm of these clamps was supposed to reduce the number of target-associated molecules during elongation to avoid DNA accummulation.

Complementarty sequence regions of the clamps should neither hybridize to primer binding sites nor to any of the other primers. In fact, the clamp design constitutes a true biological challenge. Exothermic procedures force complementary bases to form self and/or hetero dimers, which is diffult to avoid entirely if loads of sequences come together in a single solution. A total concentration of 0,5µM of all DNA clamps was supplied to the reaction. We experimentally compared a standard MDA reaction with primer set 1 and 2 within the use of intercalating clamps, and once with clamps and 0,5% DMSO. Both gel pictures in figure 20 display the noticeable SHV product at a length of 160 bp for primer mix 1 and 180 bp of primer mix 2. The positive reactions demonstrate unspecific signals at 300 bp and 500 bp, which seems to be the result of an uncontrolled elongation of random sequences. Clustering of DNA clamps or primer mismatching seems to be a reason for the occurance of uncertain products. At a later period of the isothermal reaction certain sequence are evenly extended by 50 bp, resulting in a ladder-like structure for three of four negative controls. Basically, these intercalating sequences contain a 3´ hydroxy group which may be a reason to be extended by the polymerase if random mispriming occurred.

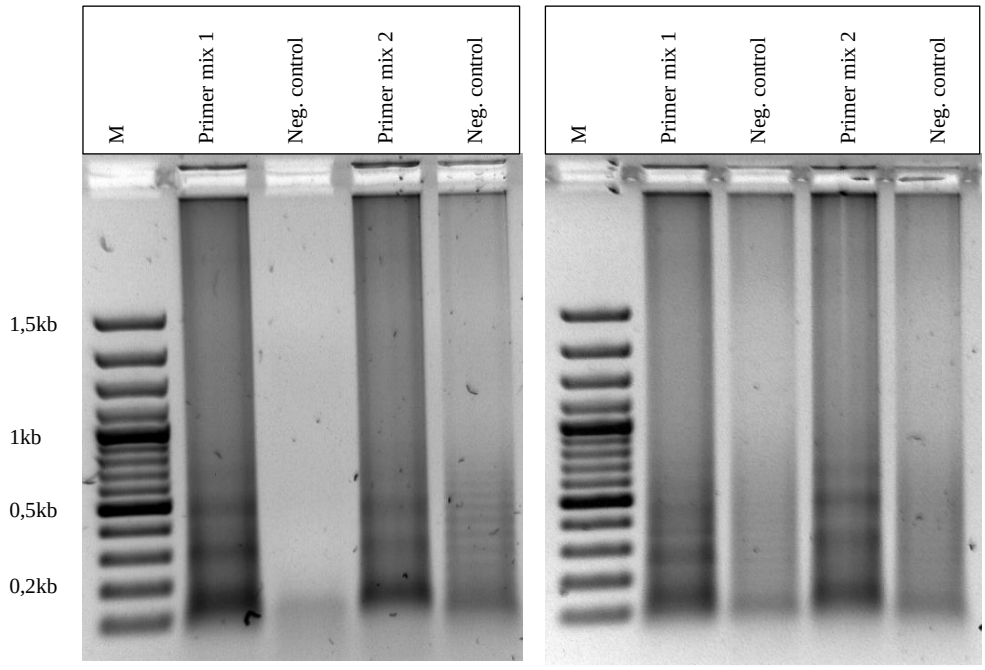


Figure 20. Multiple displacement amplification using DNA clamps and DMSO
Intercalating DNA clamps positively effects the isothermal amplification resulting in a strong band smear in every positive reaction. All reactions that are presented on the left gel picture obtained all three DNA clamps whereas each reaction on the right image, additionally, contained a final amount of 0,5% DMSO. A 100 bp Plus gene ruler is labeled M.

4.3.6. Universal primers applied in MDA and PCR

In order to increase the product output of the amplification, we created target-specific oligonucleotides that contain an universal sequences tail on the 5´ end. The primer specificity of universal primers was tested in singleplex PCR for the β-lactamase genes SHV-112, CTX-M-1/M-15, CTX-M-2, DHA-1 and OXA-1 displayed in figure 21.

Moreover, the reaction components were applied by Molzym. Pictured in figure 21, all designed primers bind properly to the target-DNA at a 55°C annealing step and ensure the

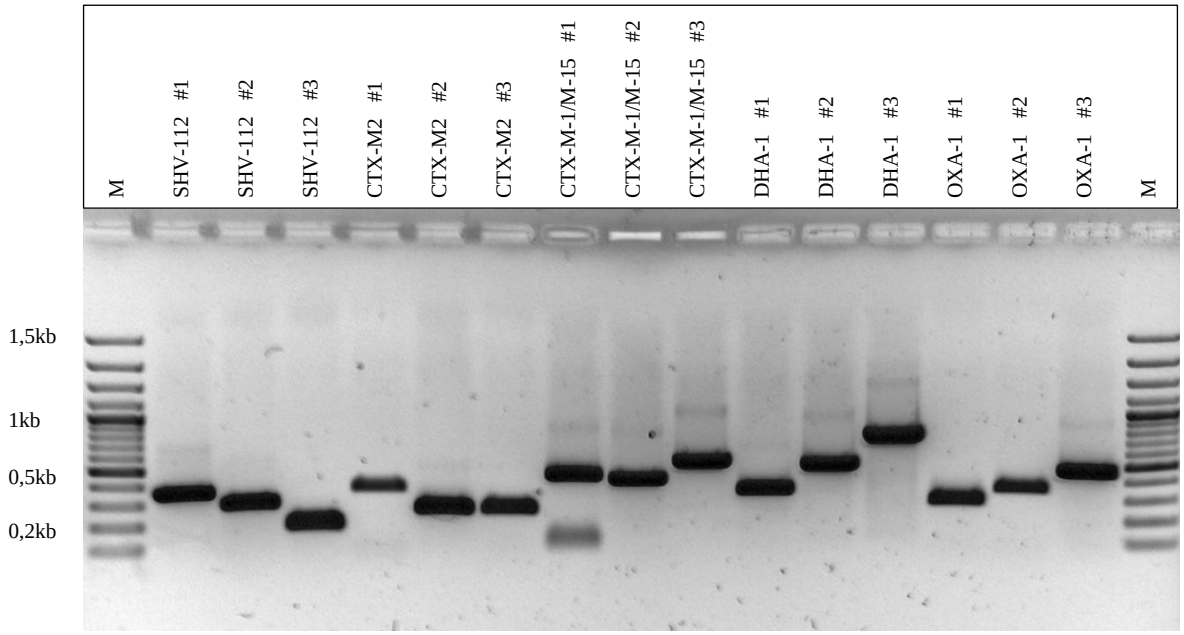


Figure 21. Amplification of beta-lactamase genes with specific primers in PCR
Estimated gene loci of bacterial resistance genes were targeted with specific primer regions and resulted in correct amplicon sizes. A 100 bp Plus gene ruler was pipetted to the 1% agarose gel.

synthesis of desired amplicons consisting the correct fragment lengths (table 9). Few unspecific amplification products can be observed in figure 18, particularly for CTX-M-1/M-15 #1 and #3, and DHA #2 and #3.

The use of universal primers in high concentration (10µM) was meant to amplify the representative SHV-112 #1 sequence (table 21) in an extended yield in a second PCR. Each oligonucleotide sequence is deducted by one nucleotide and ranges from 27 bp to 15 bp. Larger universal sequences that range from 27 bp to 22 bp bind the uniform sequence region of the specific primers more effectively than shorter universal sequences, whereas all primers result in a great amplification output as pictured in figure 22. A single band at 335 bp was detectable on the agarose gel as the SHV-112 #1 amplicon was targeted.

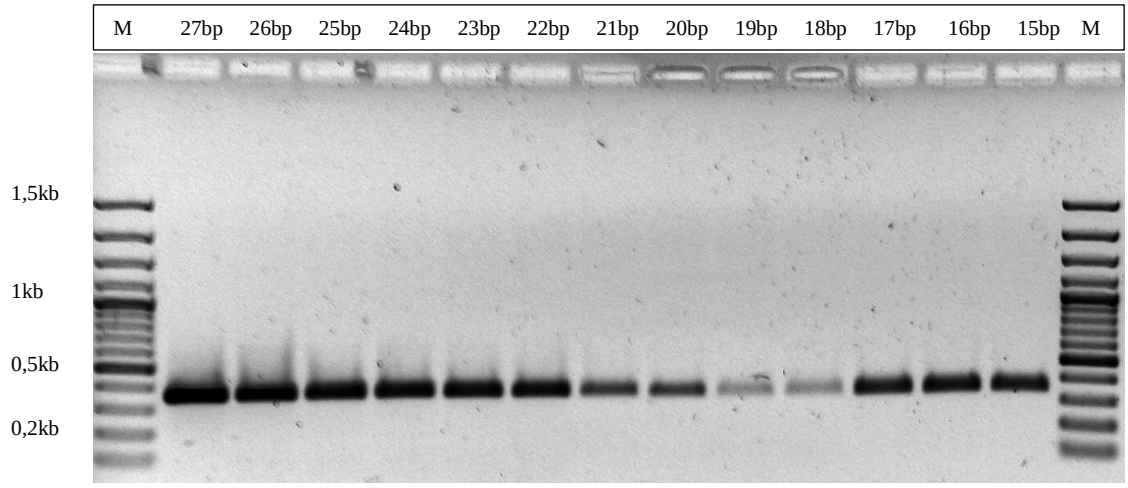


Figure 22. Test of universal primers containing various lengths in PCR
The amplification of the SHV-112 #1 amplicon from previous PCR in figure 21 resulted in a stronger signal with universal sequences consisting of 27 bp to 22 bp represented in lane 2 to 6. A 100 bp Plus gene ruler was added as marker.

4.3.7. Multiplex PCR for the CTX-M-1/M-15 gene

A sensitive multiplex PCR is normally performed in clinical and biological facilities to amplify various targets from one sample using several primers in a single reaction (49, [50]. In this experiment, we aimed to amplify the CTX-M-1/M-15 gene from 5,1ng gDNA of *E. coli* 10/152 ($9{,}45*10^5$ copies) in single- and multiplexed PCR reactions.

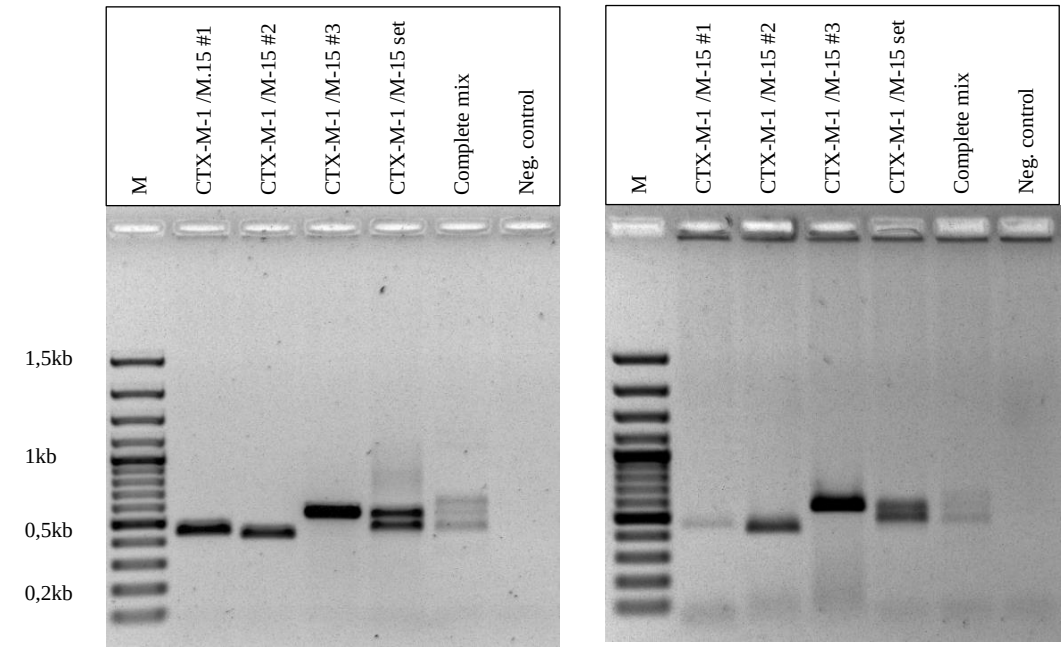


Figure 23. Single- and multiplex PCR for CTX-M-1/M-15 gene with universal primers
The difference of these experiments is that the reactions pictured right involved an isothermal reaction for one hour with target-specific primers instead of a previous cycling procedure with low-concentrated primers.
M: 100 bp Plus gene ladder, lane 2-4: six CTX-M-1/M-15-specific different primers; lane 5: primer set for CTX-M-1/M-15, lane 6: mix of 30 primers for several beta lactamase genes, lane 7: negative control

The first reactions pictured on the left image in figure 20 was initialized with 0,05µM of specific primers (tab. 9) in 10 cycles PCR followed by 25 cycles 0,5µM of a 25 bp universal primer 3 (tab. 8). Subsequent PCR reactions were supported by the Molyzm master mix 16S Basic and Mol*Taq* polymerase. Three single primers were used in individual reactions and resulted in correct amplicon sizes of 450bp with CTX-M-1/M-15 #1, 427 bp with CTX-M-1/M-15 #2 and 553 bp with CTX-M-1/M-15 #3. The reaction in lane 5 that is shown in the left image includes all six primers targeting the CTX-M-1/M-15 locus and resulted in expected distinct band patterns. Moreover, lane 6 demonstrates a multiplexed PCR reaction with 30 primers (tab. 9) involved. A target-specific amplification of all CTX-M-1/M-15 gene regions was achieved successfully without any artificial products obtained. The reactions pictured on the right site were initiated under isothermal condition at 65°C with *Bst* DNA polymerase and Molyzm master mix 16S Basic. After one hour, 0,5µM of universal primers (25 bp) and Mol*Taq* polymerase were added to each tube. The MDA products from the first amplification apparently stick to the gel pockets. Both experiments resulted in clear band patterns for single- and multiplexed reactions.

4.3.8. Sensitivity of the isothermal reaction with specific universal primers

In this procedure, a prior isothermal reaction was combined with 25 cycles of PCR to test the sensitivity of a single- and multiplex reaction with specific primers for CTX-M-1/M-15.

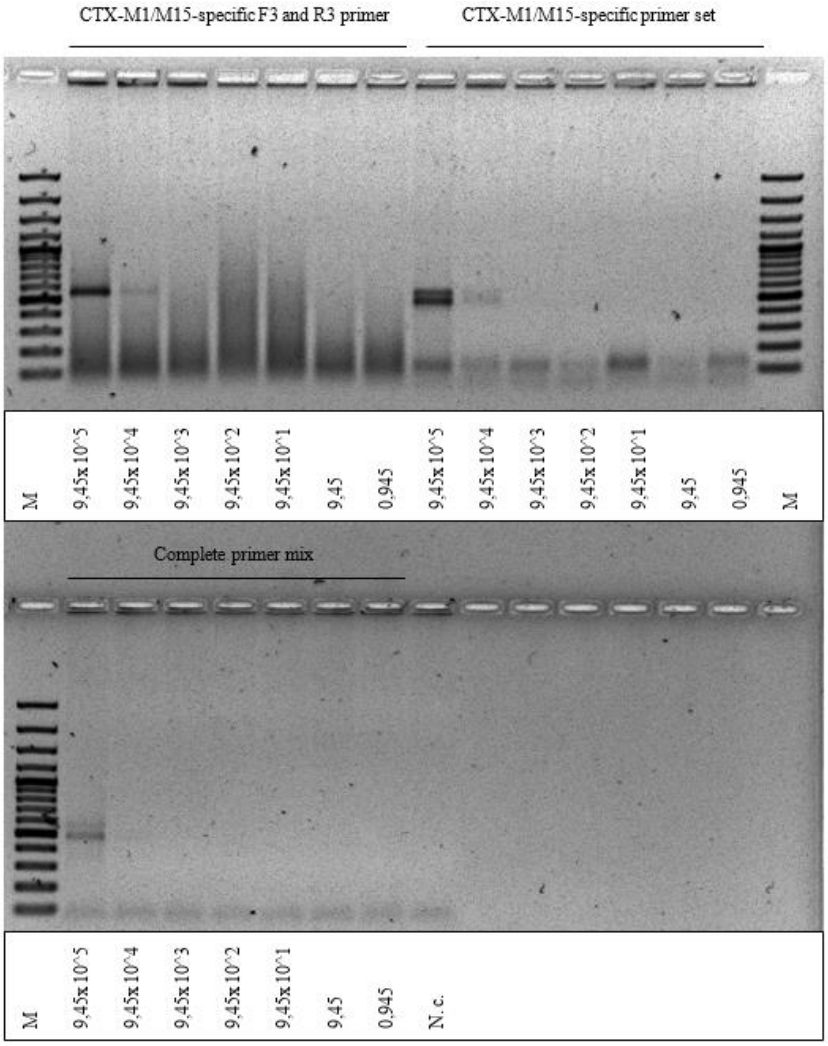


Figure 24. Testing the sensitivity of a MDA reaction for CTX-M-1/M-15
Single- and multiplexed reaction resulted in the correct amplification product with CTX-M-1/M-15-specific primers when adding up to 0,51ng to the reaction. Again, a 100 bp Plus gene ruler was loaded to the a 1% agarose gel.

Each reaction is based on the mastermix 2,5x Basic dye supplied by Molzym and 4 units of the *Bst* DNA polymerase, larger fragment, as the buffer is suitable for this enzyme. Universal primers and *Taq* polymerase were directly added to the reaction after the isothermal amplification.

The CTX-M-1/M-15 gene from genomic bacterial DNA (5,1ng to 5,1fg) amplified in this experiment. A specific product is detectable at 550 bp up to a copy number of $9{,}45*10^4$.

Likewise, in multiplexed reaction that includes six different primers (table 9) for the CTX-M-1/M-15 gene correct band structures were created with an input down to 9,45*10^4 copies. The use of multiple primers principally enhances the specificity of the reaction. The multiplex reaction in the lower part of this gel image contains thirty different primers for five diverse beta-lactamase genes. The amplification of the CTX-M-1/M-15 gene points out products at 550 bp and 480 bp by adding 5,1ng of gDNA to the reaction.

4.3.9. The detection of the SHV-12 amplification with molecular beacon

The designed molecular beacon consists of a 6-carboxyfluorescein (6-FAM) fluorophore attached to the 5´ end and Iowa Black quencher on the 3´ end and has its maximum absorbance at 495nm and emission at 520nm. In quantitative MDA, this molecule hybridized specifically to complementary sequence region of the SHV-12 gene which was

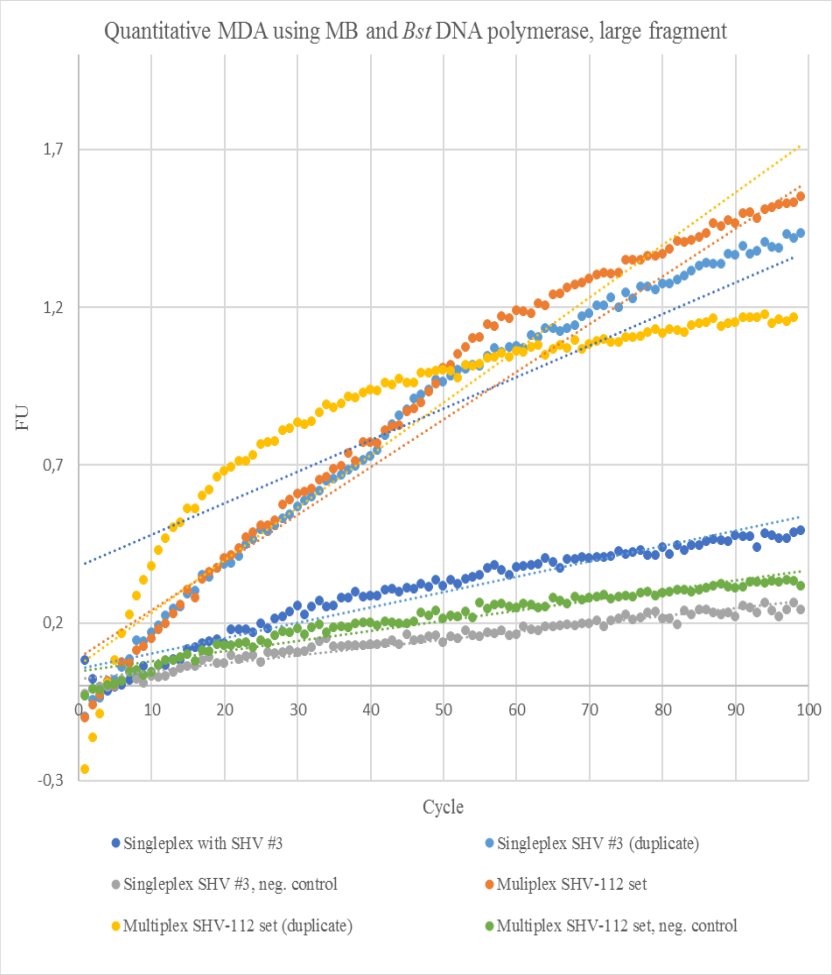


Figure 25. Fluorescent detection of an isothermal multiplex reaction
In qPCR, the SHV-specific molecular beacon led to an increased fluorescent signal after 99 cycles. The corresponding control samples shows no significant amplification product as the fluorescent output stays decreased.

amplified with specific primers listed in table 9. The fluorescent outcome was measured minute after minute in an one hour 39 minutes (99 cycles) process.

Quantitative MDA reaction was performed as duplicate with 0,3M of MB and default components described in 3.2.3. Under isothermal circumstances, single- and multiplexed reactions resulted in an increase of the FU up to 1,5. In figure 25, the corresponding negative control reactions labeled by the green and grey routes exhibit a slight background noise but remained unamplified. Molecular beacon releases a fluorescent signal only if it binds correctly to the target-DNA. An artificial signal can occur if MB gets removed by the 5´→ 3´ exonuclease activity of the polymerase. Due to the fact that *Bst* DNA polymerase does not contain a 5´→ 3´ exonuclease activity, obviously these fluorescent signals are the result of an amplification procedure.

4.3.10. Negative control sequencing

A frequently occurring drawback within the multiple displacement amplification showed false-positive background signals for the control reaction as demonstrated in the gel picture in figure 16. First, the reaction conditions were adjusted minimally by varying the temperature, Mg^{2+} content and primer concentration but that literally, had no influence on reducing the smear in the negative control. In the end, we agreed to perform Ion Torrent

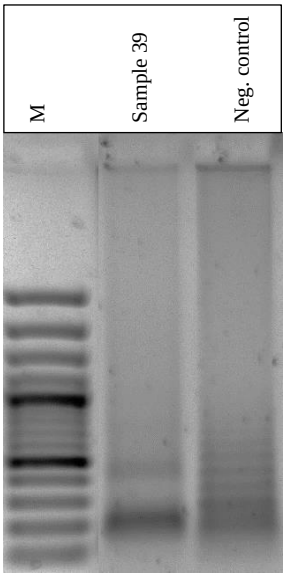


Figure 26. False-positive amplification product in non-template reaction
This image shows the outcome of a multiple displacement amplification at 65°C with *Bst* DNA polymerase, large fragment and primer mix 2 for the SHV gene. As template, 1,35ng of genomic DNA consisting the SHV-12 gene (861bp) was added to the 20µl reaction in sample 39. The corresponding negative control displays unexpected multiple band patterns up to 0,9kb compared to the 100 bp Plus gene ruler. To clarify those random sequences, we performed amplicon sequencing with the Ion Torrent PGM.

sequencing with chosen samples on a 314™ v2 semiconductor chip. In regard to figure 16, the reaction of sample 39 includes genomic bacterial DNA from *K. pneumoniae* strain 10/151 that encodes the SHV gene (861bp), whereas the corresponding negative control contains water instead. The analysis of this troubleshoot includes an assembly with the Ion Torrent SPAdes assembler and subsequent nucleotide alignments using Blast and ClustalO. In this regard, the Ion Torrent sequencing process of barcoded amplicons denotes 2,439 reads consisting a mean size of 111 bp for sample 39 and 6,312 reads with 67-mer mean length for the negative control. The chip load density of enriched ISP concluded in 85% with only 13% of usable reads and a reduced amount of clonal ISPs (sequencing results not shown). Total reads were assembled for every sample to form larger contiguous sequences. Moreover, these fragments were aligned to the NCBI nucleotide database. Interestingly, the largest sequence outcome for sample 39 consisting a length of 443bp out of 861bp confirmed 98% similarity to the query SHV sequence. Eight other contigs that were generated during the sequencing run belong to the SHV gene as well. The usage of multiple oligonucleotides in a MDA reaction probably cause dimerization processes as they get extended randomly but constantly by the DNA polymerase in both reactions. Even though only 51,5% of contig 1 could be matched to the query with high identity, no SHV-related sequences could not be determined in the negative control. Furthermore, all 18 contigs of the negative control could be aligned to all added primer sequences and confirmed no significant accordance to any region of the SHV gene. Therefore, all forward and reverse primers of primer mix 2 were entered to CLUSTAL Omega, an online multiple sequence alignment tool. This comparison pointed out that all six primers cover most parts of the assembled sequences. As a consequence of this result, primer dimerization events and random elongation process may be the reason for causing this strong band signal between 150 bp and 500 bp in our negative samples. This investigation was achieved by aligning the forward orientated primers and the reverse complemented primers of primer set 2 in table 6.

In conclusion, no significant alignment regarding the SHV β-lactamase gene was determined in the negative control which indicates that neither pipetting mistakes nor contamination, but probably spontaneous primer dimerization events arouse in the preliminary preparation.

5. Conclusion & Outlook

5.1. Genome comparison on three different *Acanthamoeba* strains

This master´s study represents the analysis on the genetic profiles of three different *Acanthamoeba* strains, Pb30/40, 72/2 and 1BU which constitute an important and rising threat for human health. The eye infection rate caused by *Acanthamoeba* increased dramatically within the global use of soft contact lenses and bad hygiene as these pathogens potentially attach to the lens surface and lead to corneal degradation [23][56]. In clinical diagnostics, the usual detection is based on PCR amplification of the 18S rRNA gene, but shows high sequence variation among different pathogenic and non-pathogenic *Acanthamoeba* species [11]. As reported before, even among closely-related genotypes high sequence dissimilarity (>5%) exists in the 18S ribosomal subunit [7]. As a consequence, a rapid and clear characterization of a certain amoebic infections is problematic and limited. Most disease-causing strains belong to the most prevalent T4 genotype. In order to investigate genetic properties of certain pathogenic and non-pathogenic strains, whole genome sequencing was conducted with the Ion Torrent semiconductor chip technology. Genome mapping of high-quality simulated reads was performed with NextGenMap, which is highly suitable for whole genome alignments and polymorphic sequences [47]. If we now evaluated the mapping results for all three strains to the reference Neff genome, this study evaluates, that the distantly-related *A. comandonii* strain Pb30/40, genotype T7 represents only 2,49% genome identity than *A. castellanii* strain 1BU, which resulted in 62% as it belongs to the same genotype T4 as the reference. Moreover, well-trimmed sequences for *A. lenticulata*, genotype T5, were mapped 17,72% to the reference strain. Even mapping of untrimmed reads to the reference genome which were generated in the first sequencing run showed similar correlations. In regard to the phylogenetic distance, simulated sequencing data of all genotypes correlate well to the respective sequence dissimilarity (>5%) in the 18S rRNA gene. Moreover, mapping of housekeeping genes resulted in individual nucleotide substation in the actin and profilin gene region of *A. castellanii* strain 1BU, whereas insufficient genome coverage occurred for the other two genotypes. As the MBP is considered as to be initially responsible for causing an infection, pathogenic strain 1BU showed high nucleotide variation across the 1525 bp gene locus. Due to low sequence similarity of simulated single reads to the reference, a first de novo assembly for *A. comandoni* strain Pb30/40 (84,2 Mb; GC% 43,65) was executed, as well as a second draft

genome sequence for *A. lenticulata* strain 72/2 (72,7 Mb; GC% 56,91). High-quality contigs for both strains were sorted based on MIRA algorithms and submitted to the Genbank database.

5.2. Spot on sequencing performance and data output

A scientific revolution was triggered since modern NGS technologies were developed. These devices guarantee simple and rapid processing to gain huge amounts of data at low costs. In concern of technical procedure, billions of sequencing information seems to be easily generated in the laboratory, but truly involves critical preparation steps that possibly lead to sequencing biases and analytical troubleshoots. Difficult parts in sample preparation, for instance, involve the correct size selection of the library, an additional DNA amplification procedure, the resuspension of temple-positive ISPs after emulsion PCR, the chip load and the initialization of the PGM system. Regarding the Ion Torrent technology, DNA amplification by PCR can cause an error rate by increasing the amount of substitutions and indels (insertion/deletion) in the genome and might have an influence on the GC% value of the template as well as the sequencing depth [13][14]. Particular quality values are set for each individual nucleotide that is incorporated during the sequencing process and correspond to error alerts. Ion Torrent PGM supports single reads which makes an whole genome assembly even more complex. In particular, data analysis software and open source tools have to be selected properly for the respective experiment. The statistical analysis by FastQC was suitable for Ion Torrent sequencing data to adjust trimming of low quality bases. As many endobacteria are normally associated to *Acanthamoeba*, random DNA was filtered by Kraken to obtain an accurate alignment of high-quality *Acanthamoeba*-related reads. Moreover, high standardized hardware and great expertise on bioinformatics was necessary to face the challenge of analysing those huge data sets which we accomplished in team. Upcoming studies on this topic should involve a greater number of samples sequenced in order to combine well-sorted datasets to reach an entirely aligned genome. Another genome mapping and sequence comparison, for instance, could be done for *A. lenticulata* strain 72/2, as now two whole draft sequence are publicly available. Additional investigations on universal virulence factors of infectious *Acanthamoeba* might focus on deeper sequencing approaches of pathogenic T4-related species, as this genotype predominantly causes harmful eye infections.

5.3. Improvements on the multiple strand displacement amplification assay

Isothermal reactions are considered to provide a more efficient and rapid amplification of DNA templates with higher stability than standard PCR. The main difference compared to PCR is that the reactions which include DNA denaturation, primer annealing and elongation phase are performed at one defined temperature by a robust, thermostable enzyme. In this approach, *Bst* DNA polymerase, large fragment (NEB) was used to amplify specific gene loci of certain beta-lactamase genes. Prior experiments involved the use of a *φ29* DNA polymerase, well-known for its thermostability. Previous results confirmed that *φ29* DNA polymerase had descendent strand displacement activity in MDA reactions compared to *Bst* DNA polymerase, large fragment. In fact, Bst DNA polymerase, large fragment contains great 5´→ 3´ strand displacement activity at 65°C to open up the double helix in order to ensure primer annealing to the gene region of interest. The enzyme´s 5´→3´polymerase activity was reduced at high temperatures and entirely inhibited at 80°C (acc. NEB specification manual). A total of 20µl is a sufficient reaction volume which requires the adjustment of the optimal amount for each component. Low-concentrated DMSO (0,5%) slightly improved the specificity of the multiplexed isothermal reaction at 65°C whereas high concentration inhibited the polymerase activity. The additional amount of Mg2+ added to the buffer solution seemed to enhance the elongation process of the polymerase by increasing the DNA duplex stability. Moreover, even in classical PCR multiplexed reactions are difficult to adjust and thus, often lead to questionable and unexpected outputs. For this reason, we worked on various strategies for an assessment of an isothermal multiplex reaction. First, primer design is an important issue as they should bind specifically to the target region and initiate the elongation process. The adjustment of melting temperatures, which in theory should be approximately 5°C above the annealing temperature, and a balanced GC content was setup. Inaccurate primer design can cause random and unspecific amplification results. The use of bacterial genomic DNA as target DNA made it more complicated to design specific primers. To keep denatured strands in distance, intercalating DNA clamps were considered to increase the stability of single strands at the gene region of interest and to improve the primer accessibility. In order to increase the amount of the amplified product, specific primers contained an additional universal sequence part tag on their 3´ end.

Designed DNA clamps consist of two specific single strand binding sequences and a thymidine spacer region. Their use in isothermal reaction resulted in a random ladder-like band pattern detectable on a 1% agarose gel. One reason might be that the clamps, which are located nearby the primer binding region, were consistently extended by the polymerase during the amplification process as they contain an unblocked -OH group at the 3´ end. Further improvements of this approach should consider removing and/or blocking the free 3´ hydroxyl molecule which may prevent unspecific amplicons. Another explanation could be that complementary clamp sequences crosslink together randomly and accumulate which may result in artificial band structures. However, a positive signal could be detected in multiplexed strand displacement reaction resulting in expected product sizes (bp) of the desired SHV gene region. Moreover, a pre-digestion of gDNA of *K. pneumoniae* with *AluI* restriction endonuclease could not provide sufficient asset to the isothermal reaction. As DNA was cut into smaller fragments, more effective primer annealing was supposed to occur. Struggling to obtain specific MDA products visualized on the agarose gel, we used molecular beacons in quantitative PCR for more detailed investigation. The design of these molecular beacons was facilitated with the OligoAnalyzer (IDT) and consisted of an accurate stem loop structure as well as 6-FAM as fluorophore and a quencher. For one reaction, specific binding of MB to the SHV gene region resulted in a positive amplification signal for both duplicates within single and multiplexed isothermal reaction. As template, purified gDNA (~5Mb) from *K. pneumoniae* was used. Within the second reaction, the fluorescence emission varied slightly between the duplicates, as one reaction reflected lower fluorescent levels than the other. Anyhow, due to these results we assume that primer annealing happens at the defined target region which under isothermal conditions can be amplified with the *Bst* DNA polymerase. The efficiency of the polymerase activity may be improved by further adjustments and additional experiments. However, the performance of multiplexed reactions with six different primers is challenging but still needs to be continued based on the present results. The combination of strand displacement enzymes, defined DMSO concentrations and the use of properly modified clamps seems to be a fine option to establish this kind of strand displacement amplification. Upcoming experiments will further include tests on other DNA polymerases suitable for that assay and to combine this kind of multiple displacement strand amplification within the in the H2020 developed detection system.

6. References

[1] Aliotta, J. 1996. Thermostable Bst DNA polymerase I lacks a 3′ → 5′ proofreading exonuclease activity. *Genetic Analysis: Biomolecular Engineering* 12, 5-6, 185–195.

[2] Amelia. Investigating the Effects of DMSO on PCR Fidelity Using a Restriction Digest-Based Method.

[3] Bakhtiarizadeh, M. R., Najaf-Panah, M. J., Mousapour, H., and Salami, S. A. 2016. Versatility of different melting temperature (Tm) calculator software for robust PCR and real-time PCR oligonucleotide design: A practical guide. *Gene Reports* 2, 1–3.

[4] Barišić, I., Schoenthaler, S., Ke, R., Nilsson, M., Noehammer, C., and Wiesinger-Mayr, H. 2013. Multiplex detection of antibiotic resistance genes using padlock probes. *Diagnostic Microbiology and Infectious Disease* 77, 2, 118–125.

[5] Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. 2006. GenBank. *Nucleic acids research* 34, Database issue, D16-20.

[6] Dimijian G. G. 2000. Evolving together: the biology of symbiosis, part 1. Volume 13, Number 3, pp. 217-226. In BUMC proceedings 2000;13:217-226

[7] Booton, G. C., Joslin, C. E., Shoff, M., Tu, E. Y., Kelly, D. J., and Fuerst, P. A. 2009. Genotypic identification of Acanthamoeba sp. isolates associated with an outbreak of acanthamoeba keratitis. *Cornea* 28, 6, 673–676.

[8] Casali, N., Ed. 2003. *E. coli Plasmid Vectors*. *Methods and Applications*. Methods in Molecular Biology™ 235. Humana Press, Totowa, NJ.

[9] Clarke, D. W. and Niederkorn, J. Y. 2006. The pathophysiology of Acanthamoeba keratitis. *Trends in parasitology* 22, 4, 175–180.

[10] Clarke, M., Lohan, A. J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., Bertelli, C., Schilde, C., Kianianmomeni, A., Bürglin, T. R., Frech, C., Turcotte, B., Kopec, K. O., Synnott, J. M., Choo, C., Paponov, I., Finkler, A., Heng Tan, C., Hutchins, A. P., Weinmeier, T., Rattei, T., Chu, J. S. C., Gimenez, G., Irimia, M., Rigden, D. J., Fitzpatrick, D. A., Lorenzo-Morales, J., Bateman, A., Chiu, C.-H., Tang, P., Hegemann, P., Fromm, H., Raoult, D., Greub, G., Miranda-Saavedra, D., Chen, N., Nash, P., Ginger, M. L., Horn, M., Schaap, P., Caler, L., and Loftus, B. J. 2013. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biology* 14, 2, R11.

[11] Corsaro, D., Walochnik, J., Köhsler, M., and Rott, M. B. 2015. Acanthamoeba misidentification and multiple labels: redefining genotypes T16, T19, and T20 and proposal for Acanthamoeba micheli sp. nov. (genotype T19). *Parasitology research* 114, 7, 2481–2490.

[12] Detering, H., Aebischer, T., Dabrowski, P. W., Radonic, A., Nitsche, A., Renard, B. Y., and Kiderlen, A. F. 2015. First Draft Genome Sequence of Balamuthia mandrillaris, the Causative Agent of Amoebic Encephalitis. *Genome announcements* 3, 5.

[13] Ekblom, R., Smeds, L., and Ellegren, H. 2014. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC genomics* 15, 467.

[14] Escalona, M., Rocha, S., and Posada, D. 2016. A comparison of tools for the simulation of genomic next-generation sequencing data. *Nature reviews. Genetics* 17, 8, 459–469.

[15] Byers T. J., Hugo E. R., Stewart V. J. 1990. Genes of Acanthamoeba: DNA, RNA and Protein Sequences (A Review). J. Protozool., 37(4), pp. 17S-25S.

[16] Gill, P. and Ghaemi, A. 2008. Nucleic acid isothermal amplification technologies: a review. *Nucleosides, nucleotides & nucleic acids* 27, 3, 224–243.

[17] Greub, G. 2009. Parachlamydia acanthamoebae, an emerging agent of pneumonia. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 15, 1, 18–28.

[18] Grohmann, E., Muth, G., and Espinosa, M. 2003. Conjugative Plasmid Transfer in Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews* 67, 2, 277–301.

[19] Gupta, A. and Sharma, V. K. 2015. Using the taxon-specific genes for the taxonomic classification of bacterial genomes. *BMC genomics* 16, 396.

[20] Hanson, C. 2005. Agarose Gel Electrophoresis. In *Van Nostrand's Scientific Encyclopedia*, G. D. Considine, Ed. John Wiley & Sons, Inc, Hoboken, NJ, USA. DOI=10.1002/9780471743989.vse9792.

[21] Hodkinson, B. P. and Grice, E. A. 2015. Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Advances in wound care* 4, 1, 50–58.

[22] Hurt, M., Neelam, S., Niederkorn, J., and Alizadeh, H. 2003. Pathogenic Acanthamoeba spp. Secrete a Mannose-Induced Cytolytic Protein That Correlates with the Ability To Cause Disease. *Infection and Immunity* 71, 11, 6243–6255.

[23] Ibrahim, Y. W., Boase, D. L., and Cree, I. A. 2009. How Could Contact Lens Wearers Be at Risk of Acanthamoeba Infection? A Review. *Journal of Optometry* 2, 2, 60–66.

[24] Ignatov, K. B., Barsova, E. V., Fradkov, A. F., Blagodatskikh, K. A., Kramarova, T. V., and Kramarov, V. M. 2014. A strong strand displacement activity of thermostable DNA polymerase markedly improves the results of DNA amplification. *BioTechniques* 57, 2, 81–87.

[25] Munish A. 2012. DNA Sequencing- Methods and Applications. InTechOpen, ISBN 978-953-51-0564-0. Section 1, pp. 1-9.

[26] Lawyer F. C., Stoffel S., Saiki R. K., Myambo K., Drummond R., and Gelfand D.H. 1988. Isolation, Characterization, and Expression in Escherichia coli of the DNA Polymerase Gene from Thermus aquaticus. THE JOURNAL OF BIOLOGICAL CHEMISTRY. Vol. 264, No. 11. Issue of April 15, pp. 64274437,1989

[27] Khan, N. A. 2006. Acanthamoeba: biology and increasing importance in human health. *FEMS microbiology reviews* 30, 4, 564–595.

[28] Kim, J. and Easley, C. J. 2011. Isothermal DNA amplification in bioanalysis: Strategies and applications. *Bioanalysis* 3, 2, 227–239.

[29] Walochnik J., Aspöck H. 2014. Entzündliche Augenerkrankungen. Springer Verlag, Pleyer. Chapter 3.5 Akanthamöben-Keratitis (AK) (engl. Akanthamoeba keratitis). pp. 105-116.

[30] Kong, H. H. 2009. Molecular phylogeny of acanthamoeba. *The Korean journal of parasitology* 47 Suppl, S21-8.

[31] Tyagi S, Marras SAE, Vet JAM, Kramer FR. Molecular beacons: hybridization probes for detection of nucleic acids in homogeneous solutions. In: Kessler C, editor. NonradioactiveAnalysis of Biomolecules, 2nd ed. Springer-Verlag, Berlin; 2000. p. 606–16.

[32] Kucera, R. B. and Nichols, N. M. 2008. DNA-dependent DNA polymerases. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 3, Unit3.5.

[33] Ledee, D. R., Iovieno, A., Miller, D., Mandal, N., Diaz, M., Fell, J., Fini, M. E., and Alfonso, E. C. 2009. Molecular identification of t4 and t5 genotypes in isolates from acanthamoeba keratitis patients. *Journal of clinical microbiology* 47, 5, 1458–1462.

[34] Levy, M., Thaiss, C. A., and Elinav, E. 2015. Metagenomic cross-talk: the regulatory interplay between immunogenomics and the microbiome. *Genome medicine* 7, 120.

[35] Lorenzo-Morales, J., Khan, N. A., and Walochnik, J. 2015. An update on Acanthamoeba keratitis: Diagnosis, pathogenesis and treatment. *Parasite* 22, 10.

[36] Marciano-Cabral, F. and Cabral, G. 2003. Acanthamoeba spp. as Agents of Disease in Humans. *Clinical Microbiology Reviews* 16, 2, 273–307.

[37] Morgan, X. C., Segata, N., and Huttenhower, C. 2013. Biodiversity and functional genomics in the human microbiome. *Trends in genetics : TIG* 29, 1, 51–58.

[38] Nguyen, L. V., Giannetti, S., Warren-Smith, S., Cooper, A., Selleri, S., Cucinotta, A., and Monro, T. 2014. Genotyping single nucleotide polymorphisms using different molecular beacon multiplexed within a suspended core optical fiber. *Sensors (Basel, Switzerland)* 14, 8, 14488–14499.

[39] Panjwani N. 2010.Pathogenesis of Acanthamoeba Keratitis. National Institute of Health, Ocul Surf . 2010 April; 8(2): 70–79.

[40] Parish, T., Ed. 2001. *Mycobacterium tuberculosis Protocols*. Methods in Molecular Medicine 54. Humana Press, Totowa, NJ.

[41] Petra F. G. Wolffs. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer.

[42] Poole, K. 2004. Resistance to beta-lactam antibiotics. *Cellular and molecular life sciences : CMLS* 61, 17, 2200–2223.

[43] Poptsova, M. S., Il'icheva, I. A., Nechipurenko, D. Y., Panchenko, L. A., Khodikov, M. V., Oparina, N. Y., Polozov, R. V., Nechipurenko, Y. D., and Grokhovsky, S. L. 2014. Non-random DNA fragmentation in next-generation sequencing. *Scientific reports* 4, 4532.

[44] Pumidonming, W., Koehsler, M., Leitsch, D., and Walochnik, J. 2014. Protein profiles and immunoreactivities of Acanthamoeba morphological groups and genotypes. *Experimental Parasitology* 145, S50-S56.

[45] Schiller, B., Makrypidi, G., Razzazi-Fazeli, E., Paschinger, K., Walochnik, J., and Wilson, I. B. H. 2012. Exploring the unique N-glycome of the opportunistic human pathogen Acanthamoeba. *The Journal of biological chemistry* 287, 52, 43191–43204.

[46] Schroeder, J. M., Booton, G. C., Hay, J., Niszl, I. A., Seal, D. V., Markus, M. B., Fuerst, P. A., and Byers, T. J. 2001. Use of subgenic 18S ribosomal DNA PCR and sequencing for genus and genotype identification of acanthamoebae from humans with keratitis and from sewage sludge. *Journal of clinical microbiology* 39, 5, 1903–1911.

[47] Sedlazeck, F. J., Rescheneder, P., and Haeseler, A. von. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics (Oxford, England)* 29, 21, 2790–2791.

[48] Siddiqui, R. and Khan, N. 2012. Biology and pathogenesis of Acanthamoeba. *Parasites & Vectors* 5, 1, 6.

[49] Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews. Genetics* 15, 2, 121–132.

[50] Sint, D., Raso, L., and Traugott, M. 2012. Advances in multiplex PCR: balancing primer efficiencies and improving detection success. *Methods in ecology and evolution* 3, 5, 898–905.

[51] Siqueira, J. F., JR, Fouad, A. F., and Rocas, I. N. 2012. Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of oral microbiology* 4.

[52] Al-Mujaini A., Al-Kharusi N., Thakral A., Wali U. K. 2009. Bacterial Keratitis: Perspective on Epidemiology, Clinico-Pathogenesis, Diagnosis and Treatment. SQU Med J, August 2009, Vol. 9, Iss. 2, pp. 184-195, Epub 30th June 2009

[53] Tenover, F. C. 2006. Mechanisms of antimicrobial resistance in bacteria. *American journal of infection control* 34, 5 Suppl 1, S3-10; discussion S64-73.

[54] Tong, Y., Lemieux, B., and Kong, H. 2011. Multiple strategies to improve sensitivity, speed and robustness of isothermal nucleic acid amplification for rapid pathogen detection. *BMC Biotechnology* 11, 1, 50.

[55] Walochnik, J., Obwaller, A., and Aspöck, H. 2000. Correlations between Morphological, Molecular Biological, and Physiological Characteristics in Clinical and Nonclinical Isolates of Acanthamoeba spp. *Applied and Environmental Microbiology* 66, 10, 4408–4413.

[56] Walochnik, J., Scheikl, U., and Haller-Schober, E.-M. 2015. Twenty Years of Acanthamoeba Diagnostics in Austria. *Journal of Eukaryotic Microbiology* 62, 1, 3–11.

[57] Wood, D. E. and Salzberg, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15, 3, R46.

[58] Yagi, S., Schuster, F. L., and Bloch, K. 2007. Demonstration of presence of acanthamoeba mitochondrial DNA in brain tissue and cerebrospinal fluid by PCR in samples from a patient who died of granulomatous amebic encephalitis. *Journal of clinical microbiology* 45, 6, 2090–2091.

7. Summary

The free-living protist *Acanthamoeba* is an opportunistic microorganism that potentially causes rare but harmful diseases in humans and animals such as fatal eye infections, brain disorders and skin irritation. To obtain a rapid clinical diagnosis is often challenging and is normally attained by PCR, staining methods or microscopy. Our scientific interest is based on the genetic analysis of pathogenic and non-pathogenic *Acanthamoeba* strains. For that purpose, we performed whole-genome shotgun sequencing on three different *Acanthamoeba* strains: *Acanthamoeba comandoni* strain Pb30/40, *Acanthamoeba lenticulata* strain 72/2 and *Acanthamoeba castellanii* strain 1BU. High-quality reads generated with the Ion Torrent semiconductor chip technology for the pathogenic *A. castellanii* strain 1BU could be mapped to 62% to the non-pathogenic *A. castellanii* Neff reference genome, as both strain belong to the same genotype T4. Homologous gene regions of certain cytoskeleton regulation proteins such as actin and profilin could be mapped with high sequence similarities for this strain. A crucial virulence factor of this protist is reflected by the mannose binding protein (MBP) and serine proteases amongst several others. An alignment of the 1525 bp MBP-related gene locus resulted in 88% sequence similarity for strain 1BU, but low genome coverage occurred for the other two strains. The genomic alignment of the alkaline serine protease for strain 1BU resulted in 94% to the reference sequence (1331/1415 bp). The genome coverage for strain Pb30/40 and 72/2 with well-trimmed reads was low, as they belong to more distantly-related genotypes than the reference genotype T4. As a consequence, after the resequencing procedure we delivered the first draft genome for *A. comandoni* strain Pb30/40, which resulted in 86,2 Mb (GC%: 43,65) and the second draft sequence for *A. lenticulata* strain 72/2 with 74,7 Mb (GC%: 56,91). Sequence submissions are available on the GenBank database.

The second part of the present thesis focused on the optimisation of an isothermal amplification method which is supposed to be used in devices currently in development. In that assay, the traditional thermo cycling procedure is replaced by a non-PCR technique, where a multiple strand displacement amplification allows a linear amplification of the genomic target-DNA. The *Bst* DNA polymerase (large fragment) is intended to displace DNA double strands properly to provide specific primer annealing of multiple primer sets. Several approaches were pursuit to improve the amplification process, including the use of A) dimethyl sulfoxide (DMSO), B) intercalating DNA clamps, C) a restriction digest of

genomic bacterial DNA, and D) the use of universal primers. The chemical substrate DMSO optimised the isothermal reaction at 65°C by adding 0,5% of DMSO to 20µl reactions. DNA clamps intended to intercalate into the genomic DNA were designed to increase the primer binding efficiency but resulted in unspecific background reactions making further optimisation necessary. Moreover, the performance of the isothermal amplification was monitored via target-specific molecular beacons showing a positive fluorescent signal for single- and multiplexed reactions. Additionally, sequencing of negative controls was performed in order to retrieve information on false-positive background signals and the nature of unspecific amplification products.

8. Zusammenfassung

Die vorliegende Masterarbeit beschäftigt sich insbesondere mit der Forschung von freilebenden Akanthamöben, welche heutzutage als signifikante Erreger schwerwiegende Erkrankungen bei Mensch und Tier hervorrufen können. Hierzu zählen unter anderem Augeninfektionen, Gehirnstörungen sowie Hauterkrankungen. Eine rechtzeitige Diagnose erweist sich aufgrund der spät auftretenden Symptomatik als äußert diffizil. Mittels PCR, Färbemethoden oder durch mikroskopische Untersuchungen kann ein Befall jedoch festgestellt werden. Dem Autor der Arbeit war es ein Bedürfnis die genetische Analyse von krankheitserregenden und nicht-krankheitserregenden Akanthamöben-Stämmen mittels Sequenzierung von *Acanthamoeba comandoni* Pb30/40, *Acanthamoeba lenticulata* 72/2 und *Acanthamoeba castellanii* 1BU zu identifizieren. Mit der Ion Torrent Next Generation Sequencing Technologie, für welche ein Mikrochip verwendet wurde, konnten Kurzsequenzen generiert werden. Das sequenzierte Genom des Stammes 1BU kongruiert zu 62% mit dem Referenzgenom, welches dem selbem Genotyp T4 zuordenbar ist. Bei einem Vergleich ausgewählter Genregionen, die für regulatorische Funktionen des Zytoskeletts relevant sind, wurden nur geringe genetische Abweichungen zum *A. castellanii* Neff-Referenzstamm sichtbar. Zu den bekannten Virulenzfaktoren zählen vor allem das Mannose-bindende Protein (MBP) sowie Serinproteasen, welche eine Degradierung der Augenmatrix bedingen. Der Sequenzbereich des MBP (1525 bp) weist eine 88 prozentige – und in Bezug auf der Serinprotease – eine 94% Sequenzübereinstimmung zwischen dem pathogenen Stamm 1BU und dem nicht pathogenen Neff-Referenzgenom auf. Die beiden Stämme Pb30/40 (T7)und 72/2 (T5), gehören unterschiedlichen Genotypen an, wodurch lediglich

eine geringe Konformität zu dem Referenzstamm besteht. Der Autor generierte im Zuge dieser Studie zwei neue De novo-Assemblierungen für *A. comandoni* Pb30/40 mit 86,2 Millionen Basenpaaren (GC%: 43,65) und für *A. lenticulata* 72/2 mit insgesamt 74,7 Millionen Basenpaaren (GC%: 56,91), welche bereits in der Gendatenbank (GenBank) verfügbar sind.

Der zweite Teil dieser Masterarbeit befasst sich mit der Optimierung einer isothermalen DNA-Amplifikation. Hierbei handelt es sich um eine effiziente Methode zur Identifizierung von Mikroorganismen. Die sogenannte Multiple Strand Displacement Amplifikation (MDA) findet im Vergleich zur klassischen PCR-Reaktion bei konstanter Temperatur statt und garantiert ein zügiges, robustes und effizientes Amplifikationssystem. Die *Bst* DNA Polymerase (large fragment) disponiert über eine spezielle enzymatische Funktion, die in der Lage ist doppelsträngige DNA aufzutrennen, um eine spezifische Primerbindung zu konstituieren. Die Optimierung der Primerspezifität sowie der Amplifikation wurde in verschiedenen Experimenten getestet. Hierzu zählen unter anderem die Verwendung von A) Dimethylsulfoxid (DMSO), B) interkalierenden DNA Klammern, C) eines Restriktionsverdaus von bakterieller gDNA, und D) universellen Primern. Die Zugabe von 0,5% DMSO zu einer 20µl Reaktion bei einer Temperatur von 65°C erwies sich als ausreichend, um eine Verbesserung der Primerbindung zu dokumentieren. Die Verwendung von speziellen DNA Strukturen (DNA Klammern) in einer MDA steigerte die Bindung der Primer an ausgewählte Genregionen von Extended Spectrum β-Lactamasen. Die isothermale DNA-Amplifikation wurde mit fluoreszierenden Sonden (Molecular Beacons) mittels qPCR nachgewiesen. Zudem wurden falsch-positive MDA Ergebnisse sowie negative Kontrollen mittels Sequenzierung analysiert, um unspezifische Amplifikationsprodukte ausschließen zu können.

9. Annex

Draft Whole Genome Sequence of the protist *Acanthamoeba comandoni* Strain Pb30/40, morphological group I

AIT Austrian Institute of Technology1, Health & Environment Department, Molecular Diagnostics, Vienna, Austria

Running Head: Draft Whole Genome of *Acanthamoeba comandoni* Strain Pb30/40

The free-living protist Acanthamoeba commonly occurs in aquatic and soil habitats and feeds on bacteria, algae and fungi. However, these amoebae can also cause disease as opportunistic pathogens, on the one hand, and on the other hand function as hosts and vehicles for pathogenic bacteria. In this study, we deliver a novel genome assembly of the *Acanthamoeba comandoni* strain Pb30/40 that contains a sequence length of 86.2 Mb.

In general, organisms belonging to the eukaryotic genus Acanthamoeba globally constitutes a rising threat being the causative agent of a potentially serious eye disease, granulomatous amoebic encephalitis (GAE) and skin infections in humans and animals (1). Up to date, twenty different genotypes (T1-T20) have been characterized based on the DNA sequence analysis of the 18S ribosomal subunit (>5% sequence dissimilarity) (2). In addition to its function as phylogenetic marker in basic research, it is an important biomarker in clinical diagnostics application. Most strains that commonly act as disease-causing agents belong to the most prevalent *A. castellanii* (T4 genotype), for which a full genome is available (3). We explored the entire genome of the non-pathogenic *A. comandoni* strain Pb30/40 which was initially classified as genotype T7 and now, as more genotypes have been established, shows highest identity to genotype T18 (4).

The strain Pb30/40 was originally isolated from a greenhouse in Hamburg, Germany. Approximately 25% of environmental Acanthamoeba isolates possess bacterial endosymbionts. A fraction of these endosymbionts such as Chlamydia have been associated with human diseases (5, 6). Thus, prior to establishing sub-cultures for DNA isolation, the amoebae were treated for 10 days with 100 µg/ml gentamycin (Sigma-Aldrich, St. Louis, MO, US) to avoid sequencing biases due to contamination with bacterial DNA from endosymbionts. Subsequently, genomic DNA was isolated with the QIAmp DNA isolation kit (Qiagen, Hilden, Germany) following the manufacturer's instructions (7). The whole genome shotgun sequencing was performed using the Ion Torrent PGM technology and a 318™ v2 semiconductor microchip. A 400 bp DNA library was created using enzymatic fragmentation of our sample.

The sequencing procedure resulted in 5,178,749 single-end reads comprising 1.62G bases with a mean sequence length of 358 bp and a genome coverage of 37x, on average. The statistical evaluation of well-trimmed reads by FastQC (Version 0.11.5) revealed high base

pair quality scores and a normal G+C distribution. Moreover, BLAST research confirmed that 88% of all high-quality reads were unclassified after being aligned to the GenBank nucleotide database. Additionally, metagenomics clustering was recorded by Kraken to discard bacteria-related sequences based on a 31-mer sequence database (8). The de novo assembly was provided with cleaned sequencing data submitted to MIRA (Version 4.9.5_2). The final draft genome of *A. comandoni* strain Pb30/40 resulted in 86,262,821 bases with a G+C content of 43.65%. Additionally, the N50 value records an equal or larger sequence length of 2281 bp with a total of 9583 contigs. Comparative analysis revealed that strain Pb30/40 shows a genome similarity of 2.49% to the genotype T4 of the *A. castellanii* Neff strain published by Clarke et al. (9). This mapping result corresponds to the dissimilarities of >30% in the 18S rRNA gene between these strains. In this aspect, high genomic variations seem to exist among more distantly-related Acanthamoeba genotypes.

Nucleotide sequence accession number. This Whole Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under the accession MRZZ00000000. The version described in this paper is version MRZZ01000000.

Acknowledgement

This study was supported by the H2020 FAPIC project under grant agreement number 634137. Additionally, we thank Dr. Michel Rolf from the Central Institute of the Federal Armed Forces Medical Services in Koblenz, Germany for providing strain Pb30/40.

References

1. Marciano-Cabral F, Cabral G. 2003. Acanthamoeba spp. as Agents of Disease in Humans. Clinical Microbiology Reviews 16:273–307. doi:10.1128/CMR.16.2.273-307.2003.

2. Corsaro D, Walochnik J, Kohsler M, Rott MB. 2015. Acanthamoeba misidentification and multiple labels: redefining genotypes T16, T19, and T20 and proposal for Acanthamoeba micheli sp. nov. (genotype T19). Parasitol Res 114:2481–2490. doi:10.1007/s00436-015-4445-8.

3. Ledee DR, Iovieno A, Miller D, Mandal N, Diaz M, Fell J, Fini ME, Alfonso EC. 2009. Molecular identification of t4 and t5 genotypes in isolates from acanthamoeba keratitis patients. J Clin Microbiol 47:1458–1462. doi:10.1128/JCM.02365-08.

4. Michel R, Müller KD, Hauröder B, Zöller L. 2000. A Coccoid Bacterial Parasite of Naegleria sp. (Schizopyrenida: Vahlkampfiidae) Inhibits Cyst Formation of its Host but not Transformation to the Flagellate Stage. Acta Protozool. 39: 199-207.

5. Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, Horn M. 2008. Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. Appl Environ Microbiol 74:5822–5831. doi:10.1128/AEM.01093-08.

6. Greub G. 2009. Parachlamydia acanthamoebae, an emerging agent of pneumonia. Clin Microbiol Infect 15:18–28. doi:10.1111/j.1469-0691.2008.02633.x.

7. Walochnik J, Scheikl U, Haller-Schober E-M. 2015. Twenty Years of Acanthamoeba Diagnostics in Austria. Journal of Eukaryotic Microbiology 62:3–11. doi:10.1111/jeu.12149.

8. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. doi:10.1186/gb-2014-15-3-r46.

9. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Burglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Heng Tan CS, Hutchins AP, Weinmeier T, Rattei T, Chu JSC, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick DA, Lorenzo-Morales J, Bateman A, Chiu C-H, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus BJ. 2013. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol 14:R11. doi:10.1186/gb-2013-14-2-r11.

Draft Whole Genome Sequence of the pathogenic *Acanthamoeba lenticulata* Strain 72/2, genotype T5, morphological group III

AIT Austrian Institute of Technology1, Health & Environment Department, Molecular Diagnostics, Vienna, Austria

Running Head: Draft Whole Genome of *Acanthamoeba lenticulata* Strain 72/2

The eukaryote *Acanthamoeba lenticulata* belongs is an opportunistic microorganism that is globally distributed in aquatic and terrestrial environments. It can also act as a host for various bacterial endosymbionts. Here, we performed whole genome shotgun sequencing and a de novo assembly of the *A. lenticulata* strain 72/2 which resulted in a complete draft genome sequence comprising 74.7 Mb.

In recent years, the clinical significance of free-living acanthamoebae has become a growing issue for human health as they potentially can lead to severe eye infections, on the one hand and to fatal encephalitis in immunocomprised patients, on the other hand. The painful so-called Acanthamoeba keratitis has been observed mainly in contact lens wearers but also in non-contact lens wearers (1). Eye infections are predominately caused by the most prevalent *Acanthamoeba castellanii* (genotype T4; group II), but also *A. lenticulata* (genotype T5) is frequently encountered (2, 3). Organisms belonging to the genus Acanthamoeba are clustered in 20 different genotype based on the sequence similarities of their 18S rRNA gene (4). This genotype based classification of Acanthamoeba is mainly used in clinical diagnostic application due to the availability of genotype-associated PCR primers and the simplicity of PCR-based detection. *A. lenticulata* belongs to the morphological subgroup III, which is defined by the cyst morphology and known to include several different genotypes (5, 6). Species of group III are more closely-related to those of morphological group II than to group I (7).

A. lenticulata strain 72/2 had originally been isolated from the nasal mucosa of a healthy individual, but had proven to be highly pathogenic in mice (8). The strain used in this study was re-isolated from the mouse brain. After in vitro cultivation, the total DNA was extracted with the QIAmp DNA isolation kit (Qiagen, Hilden, Germany) guided by the supplier´s manual. We used the Ion Torrent next generation sequencing technology to deliver whole genome sequences of this strain. First, genomic DNA was sheared enzymatically in order to generate a library comprising 400 bp. In the final preparation step, a 318™ v2 semiconductor microchip was loaded with template-positive Ion sphere particles (ISP) and inserted into the Ion Torrent PGM device.

The whole genome sequencing process announced a total of 6,434,049 single-end reads of 1.75 Gb and a genome coverage of 41x, on average. The median length of all raw sequences was 317 bp including 94% of high-quality reads. The subsequent statistical analysis of well-trimmed reads was evaluated by FastQC (Version 0.11.5). As amoebae are known to carry considerable numbers of endosymbiotic bacteria, extracted DNA partially involved foreign DNA and/or plasmids. The addition of 100 µg/ml gentamycin (Sigma-Aldrich, St. Louis, MO, US) to remove the bacterial endosymbionts resulted in non-viable and non-cultivable *A. lenticulata* 72/2 strains. By using the programme Kraken, we rapidly ensured selective filtering and the removal of disturbing sequences by metagenomics execution, which is based on the establishment of a 31-mer database and an automated BLAST research (9). Accordingly, cleaned and adapter-trimmed sequencing data were run with MIRA (Version 4.9.5_2) to access a de novo assembly.

The assembly ultimately revealed an entire genome size of 74,699,519 bp with a G+C content of 56.91%, which equals the Acanthamoeba genome assembly available on the NCBI platform. The N50 value for strain 72/2 reveals 7,508 sequences with at least 2,500 bp. Additionally, the genome mapping of high quality reads of *A. lenticulata* strain 72/2, genotype T5 to the genotype T4 published by Clarke et al. (10) resulted in 17.72% which corresponds to the relatively high sequence dissimilarities of >10% in the 18S rRNA gene.

Nucleotide sequence accession number. This Whole Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under the accession MSTW00000000. The version described in this paper is version MSTW00000000.

Acknowledgement

This study was supported by the H2020 FAPIC project under grant agreement number 634137. Additionally, we thank Dr. Michel Rolf from the Central Institute of the Federal Armed Forces Medical Services in Koblenz, Germany for providing strain 72/2.

References

1. Walochnik J, Scheikl U, Haller-Schober E-M. 2015. Twenty years of acanthamoeba diagnostics in Austria. J Eukaryot Microbiol 62:3–11. doi:10.1111/jeu.12149.

2. Ledee DR, Iovieno A, Miller D, Mandal N, Diaz M, Fell J, Fini ME, Alfonso EC. 2009. Molecular identification of t4 and t5 genotypes in isolates from acanthamoeba keratitis patients. J Clin Microbiol 47:1458–1462. doi:10.1128/JCM.02365-08.

3. Spanakos G, Tzanetou K, Miltsakakis D, Patsoula E, Malamou-Lada E, Vakalis NC. 2006. Genotyping of pathogenic Acanthamoebae isolated from clinical samples in Greece--report of a clinical isolate presenting T5 genotype. Parasitol Int 55:147–149. doi:10.1016/j.parint.2005.12.001.

4. Corsaro D, Walochnik J, Kohsler M, Rott MB. 2015. Acanthamoeba misidentification and multiple labels: redefining genotypes T16, T19, and T20 and proposal for Acanthamoeba micheli sp. nov. (genotype T19). Parasitol Res 114:2481–2490. doi:10.1007/s00436-015-4445-8.

5. BOOTON GC, Visvesvara GS, BYERS TJ, KELLY DJ, FUERST PA. 2005. Identification and distribution of Acanthamoeba species genotypes associated with nonkeratitis infections. J Clin Microbiol 43:1689–1693. doi:10.1128/JCM.43.4.1689-1693.2005.

6. Marciano-Cabral F, Cabral G. 2003. Acanthamoeba spp. as Agents of Disease in Humans. Clinical Microbiology Reviews 16:273–307. doi:10.1128/CMR.16.2.273-307.2003.

7. Schiller B, Makrypidi G, Razzazi-Fazeli E, Paschinger K, Walochnik J, Wilson IBH. 2012. Exploring the unique N-glycome of the opportunistic human pathogen Acanthamoeba. J Biol Chem 287:43191–43204. doi:10.1074/jbc.M112.418095.

8. Jonckheere JF de, Michel R. 1988. Species identification and virulence of Acanthamoeba strains from human nasal mucosa. Parasitol Res 74:314–316. doi:10.1007/BF00539451.

9. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. doi:10.1186/gb-2014-15-3-r46.

10. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Burglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Heng Tan CS, Hutchins AP, Weinmeier T, Rattei T, Chu JSC, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick DA, Lorenzo-Morales J, Bateman A, Chiu C-H, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus BJ. 2013. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol 14:R11. doi:10.1186/gb-2013-14-2-r11.