CHALMERS
UNIVERSITY OF TECHNOLOGY

VÄSTRA GÖTALANDSREGIONEN

Epidemiological typing of *Pseudomonas aeruginosa* – Pulsed-field gel electrophoresis versus next generation sequencing

Master's thesis in Biotechnology

JENNY LINDHOLM

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Abstract

Bacterial infection with *Pseudomonas aeruginosa* is common in immunocompromised individuals, such as patients with cystic fibrosis. These infections are difficult to treat, due to thick mucus in the respiratory tract and antibiotic resistance among the bacterial population is common. This often leads to an increased morbidity and mortality for the colonized patients.

During several years, the gold standard typing method pulsed-field gel electrophoresis was used for the epidemiological typing of *P. aeruginosa* within Västra Götalandsregionen (VGR). Here, *P. aeruginosa* bacterial strains have been whole genome sequenced with the Ion Torrent next generation sequencing technology (NGS), followed by single nucleotide polymorphism analysis or core genome/whole genome multi loci sequencing typing. Our results show that this is a highly discriminative method with an objective interpretation as well as being communicative interlaboratory. In conclusion, surveillance of the Cystic fibrosis patient community in VGR, as well as other immunocompromised patient groups, could be beneficiary by using NGS for bacterial strain typing.

Table of Contents

1. Introduction

Pseudomonas aeruginosa is an opportunistic soil bacterium commonly causing pulmonary infections in immunosuppressed individuals, such as patients with cystic fibrosis (CF) (1), (2). These infections are often difficult to treat due to factors such as thick mucus in the respiratory pathways and bacterial carriage of resistance genes (3). This can in turn lead to long-term infections where some of these bacterial populations have evolved into so-called hypermutator strains, through selective conditions such as immune defenses and antibiotic therapy. In a hypermutator strain a 1,000 fold increase in spontaneous mutations occur (4),(5). These hypermutators give rise to multiple-antimicrobial resistance as well as long-term persistence at colonization, which can increase the morbidity and mortality for the colonized patient. Therefore, it is of great importance for patients with cystic fibrosis or other immunosuppression disorders that *P. aeruginosa* outbreaks are identified and prevented.

There are several molecular genetic methods used for strain surveillance and epidemiological typing of *P. aeruginosa* strains, though varying in discriminatory level (6). Pulsed-field gel electrophoresis (PFGE) analysis has been considered the gold standard typing method for a long time due to its discriminative power (7),(8), (9). Bacterial DNA is cleaved enzymatically with restriction endonucleases and the fragments are then separated on an agarose gel. This results in a strain specific pattern. However, it is a laborious method, sometimes with low reproducibility and with a highly subjective interpretation of the results (10). One of the earliest established sequencing-based methods, using Sanger sequencing, is the traditional multi-locus sequence typing (MLST) method (11). In this method, seven conserved housekeeping genes are sequenced, giving allelic changes generating a sequence type (ST) possible for both national as international comparisons. Although a highly reproducible method with the possibility for an objective interpretation, it lacks in discriminatory power (12).

Lately, next generation sequencing (NGS), a massive parallel sequencing method, has entered the field of Microbiology and various applications are being investigated. NGS methodology has refined DNA sequencing as it is both more cost-effective and

high-throughput, compared to the earlier used Sanger sequencing method (13). The large amount of sequences obtained in one analysis have put bioinformatics in a key role when analysing the data. Depending on the question of interest, different bioinformatical applications can be used for the analysis.
At Clinical Microbiology Laboratories at Sahlgrenska University Hospital in Gothenburg, the Ion Torrent sequencing technology from Thermo Fisher scientific is used. It is of interest for the clinic to evaluate the in-house established whole genome sequencing (WGS) workflow, for epidemiological typing of *P. aeruginosa*, in comparison to the today used PFGE method.

1.1 Aim

The aim of this study is to evaluate whole genome sequencing, using the next generation sequencing platform; Ion torrent, of *P. aeruginosa* as an epidemiological typing method and compare to the traditionally used method pulsed-field gel electrophoresis.

1.2 Specification of issues

Can WGS with NGS technology be used as an epidemiological typing method to identify individual *P. aeruginosa* strains and discriminate between clusters of closely related strains?

Can all *P. aeruginosa* patient clusters be identified and correlated with previous PFGE patterns regardless of bioinformatics analysis application?

How many core genome (cg) MLST/wgMLST alleles or SNPs differ per isolate within a cluster? And is it possible to determine a cut-off for patient related transmissions?

Is it possible to distinguish between reinfection with a new strain of *P. aeruginosa* or if an ongoing infection is caused by a hypermutator strain?

1.3 Limitations

Samples were collected from patients with immunosuppression, both CF and non-CF, but not all PFGE types were included. Samples that did not survive thawing, or of other reasons did not grow, were excluded from the project. If PFGE analysis was

not possible because of degrading DNA for instance, the sample was also excluded. Lastly, the sequences was not analysed regarding genes of resistance or virulence factors.

2. Theory

2.1 *Pseudomonas aeruginosa* infection in cystic fibrosis (CF) patients

P. aeruginosa is a Gram-negative bacterium, commonly found in soil and aquatic environments (1), (2), (11), (14). It has emerged as an opportunistic pathogen due to its high morbidity and mortality rate when infecting immunocompromised individuals, such as cystic fibrosis (CF) patients. A key feature of *P. aeruginosa*, in causing chronic infections, is its tendency to change to mucoid phenotype (3). The change is caused by mutations in the *mucA* gene and the mucoid phenotype results from a production of polysaccharides, both alginate and mucoid exopolysaccharide (MEP) (3).

Together, *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanni, Pseudomonas aeruginosa* and *Enterobacter* species, make up the “ESKAPE bugs”, a small group of bacteria that is highly resistant to antibiotics and cause the major part of nosocomial infections at hospitals worldwide (15). The carbapenem resistant *P. aeruginosa* is also one of twelve bacteria listed, by the World Health Organization (WHO) in 2017, for which there is an urgent need for the development of new antibiotics (16).

Cystic fibrosis is the most common life-threatening autosomal recessive genetic disease in Caucasians with an estimated incidence of one in 2500-4000 and a prevalence of 100,000 globally (2), (3), (17). It is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The gene encodes the CFTR protein which act as a channel in the apical membrane of epithelial cells, figure 2.1 (18).

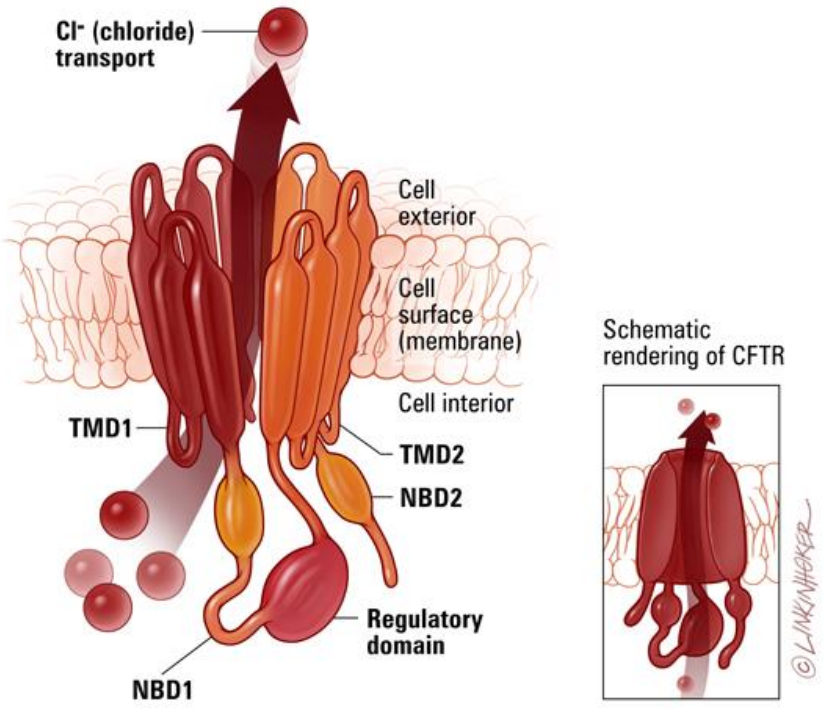

Figure 2.1. The CFTR protein channel located in the outer membrane of the epithelial cell, regulating the flux of chemical molecules e.g. chlorides (18).

Every mutation (~2500 identified) lead to different alterations of the protein domains, which in turn results in a range of CF symptoms and most commonly affected organs, are the skin, pancreas and lungs (18). Due to the dysfunction of the CFTR protein in the lungs of CF individuals, an abnormal accumulation of thick and sticky secretion cause a favourable microenvironment for the colonization and persistence of *P. aeruginosa* among other pathogens (11).

Most commonly, an environmental *P. aeruginosa,* with a nonmucoid phenotype, infects CF patients early in life leading to an initial intermittent colonization (17), (19). Adaptations, such as conversion to mucoidy, loss of flagella, loss of vital components of the lipopolysaccharide (LPS) and emergence of multiresistance to antibiotics, enables the bacteria to survive in the hostile environment of the CF lung, leading to a chronic infection (17), (19), (20). In the mucoid *P. aeruginosa,* the mutation of *mucA* has led to a dysfunctional anti- σ-factor MucA, resulting in a free RNA polymerase σ-factor (σ^{22}), figure 2.2 (17).

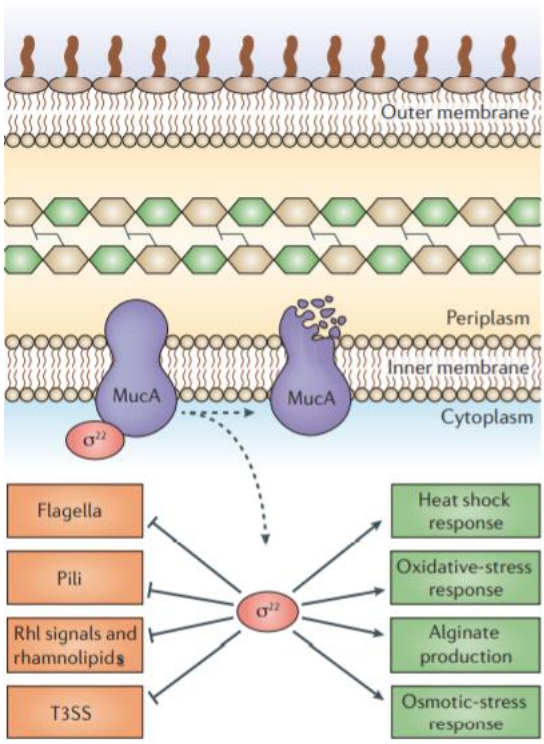

Figure 2.2. The algU regulon encoding the RNA polymerase σ-factor (σ^{22}). In mucoid P. aeruginosa, there is a mutation in the mucA gene, leading to a σ^{22} free to activate transcription of other genes, like those involved in alginate production, and responses to heat shock, osmotic stress and oxidative stress. It also downregulates virulence factors such as flagella, pili, secretion system (T3SS) and rhamnolipids (17).

The free σ^{22}, encoded by the *algU* regulon, upregulates stress responses and alginate production as well as downregulates virulence-associated genes, such as flagella, quorum sensing signals and rhamnolipids (3), (17).

2.2 Epidemiological typing

Epidemiology is the study to find the cause, and its distribution, of diseases in populations, resulting in the public health control (3), (21). Along with epidemiological data e.g. relatedness of patients, strain typing is necessary when out ruling a possible outbreak at a care unit for example. Strain typing can be divided into phenotypic methods and genotypic methods, i.e. based on gene expression or gene structure respectively (22), (23), (24). Either method is characterized by its typeability, reproducibility, and discriminatory power, as well as its ease of performance and interpretation.

2.2.1 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis, is a genotypic typing method using restriction enzymes for cleavage of chromosomal DNA followed by its separation on an agarose gel. Several studies supports PFGE as a highly discriminatory molecular typing method (7), (8), (9). However, it is a laborious method, sometimes with low reproducibility and with a highly subjective interpretation of the results difficult to communicate interlaboratory (10). PFGE, considered the gold standard typing method, dates back to 1984 when first invented by Schwartz and Cantor (25). Schematically visualised in figure 2.3 (26), the method utilizes melted agarose to capture the bacteria, whereafter enzymatic lysis of the bacterial cell wall takes place within the agarose plug and the chromosomal DNA is enzymatically cleaved infrequently by restriction enzymes (27). Thereafter, the agarose plugs with embedded digested DNA are inserted to the wells of an agarose gel and, with an apparatus producing a periodically changing electrical field, the digested DNA fragments resolves into a pattern according to their size. The pattern of each strain can be compared for relatedness studies.

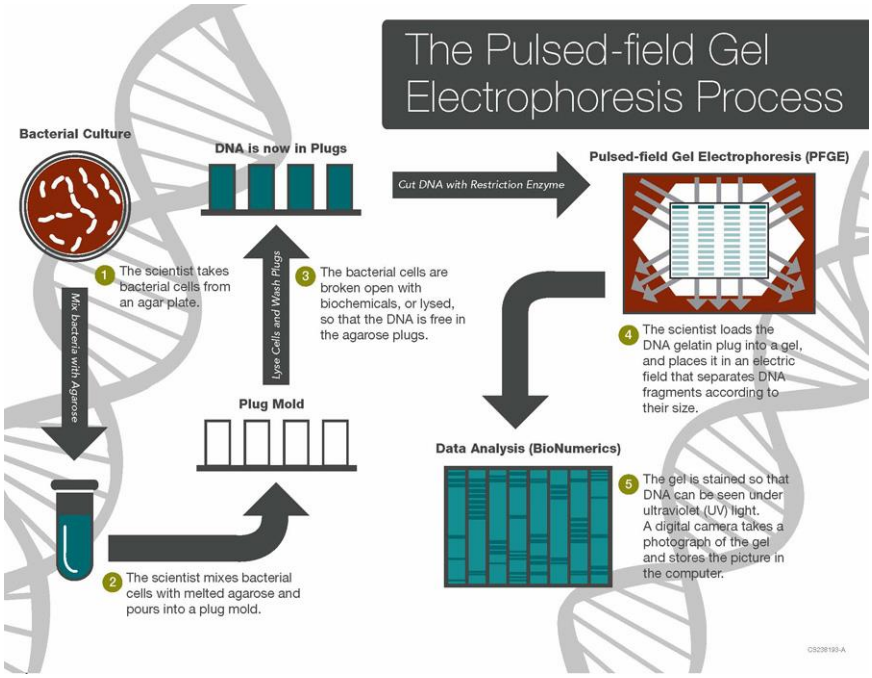

Figure 2.3. A step-by-step illustration of the PFGE method (26). The bacterial culture is mixed with melted agarose gel and poured into a plug mould. Within the plug, the bacterial cell wall is lysed and the chromosomal DNA is digested by a restriction enzyme. The agarose plug, containing the digested DNA, is loaded in a gel, and put in an alternating electric field, that separates DNA according to size.

2.2.2 Next generation sequencing – Ion torrent platform

The technique of sequencing, i.e. determining the nucleotides in the DNA, dates back to the 1970s when Sanger first developed the Sanger method (28). Back then it was costly, time consuming and inefficient, only producing one short DNA sequence (<1000 bases) at the time (29). Further development of techniques, a lot thanks to the human genome project in the 90s-00s (30), have led to next generation sequencing (NGS) methods (28), (29), (31). One of these methods, the Ion torrent platform (Thermo Fisher scientific, MA, USA), utilizes a metal oxide-semiconductor (CMOS) to measure the difference in pH due to the release of a proton (H^+) for every incorporated nucleotide (32). Firstly, DNA is extracted and purified prior to the automated library synthesis (Library Builder™, Thermo Fisher scientific) by fragmentation and adaptor ligation. The two adapters, P1 and barcoded (NNN) A, ligated at each end of the DNA fragment, enables attachment and clonal amplification by emulsion PCR to the surface of magnetic beads. The templated beads are loaded into wells on the semiconductor chip and primed by a sequence of the adapter, the sequencing is carried out by flushing one nucleotide at the time over the chip, figure 2.4.

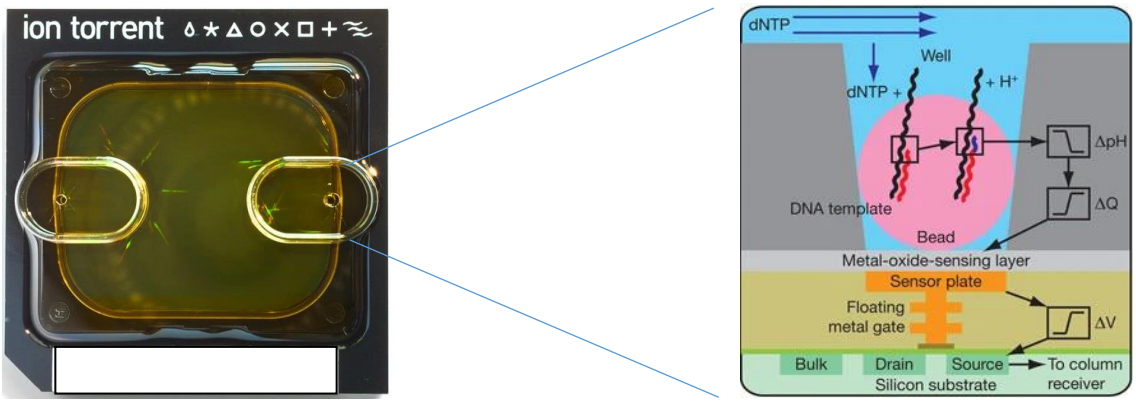

Figure 2.4. A visualisation of the semiconductor technology fitted in a three times three centimetre chip (to the left). To the right: a well containing a templated bead and the underlying technology for each well, sensing the pH change due to the proton release upon incorporation of a nucleotide. The pH change of the well induces a change in surface potential of the metal-oxide-sensing layer and a change in potential (ΔV) in the underlying field-effect transistor (32).

Advantageous are high throughput data at a relatively short time and low cost (29), (33). However, compared to other sequencing technology platforms, the Ion Torrent demonstrates a higher error rate of ~1.5 indels per 100 bases (33), (34) mainly caused by homopolymer regions. It has further been shown an extreme bias when

sequencing AT-rich genomes, such as the *P. falciparum* (80.7% AT), resulting in a deep coverage at GC-rich regions and a low coverage at AT-rich regions, with as much as 30% of the genome having no coverage at all (34).

2.2.2.1 Next generation sequencing - Data analysis

The development of next generation sequencing technologies, resulting in the production of vast amount of data, i.e. raw reads, challenged the data analysis and biological interpretation (35). The field of bioinformatics had to evolve in parallel with the sequencing technologies, which have led to a large number of data analysis software and databases (29). Interpretation of raw read quality is estimated among others by the mean coverage of genomic DNA, i.e. the number of times a unique read is covering a given nucleotide in the reconstructed sequence, evenness of coverage and mean length [bp] of all raw reads for an isolate. Main bioinformatical methods, when it comes to whole genome sequencing, are *de novo assembly* and *map reads to reference.* Single isolate sequences can thereafter be compared to one another by a *single nucleotide polymorphism (SNP, point mutation) analysis* or various types of a *multi locus sequence type (MLST) analysis* (pubMLST, cgMLST and wgMLST).

The *de novo assembly* algorithm aligns overlapping reads to longer contigs, i.e. contiguous sequences, which are ordered into a framework of the sequenced genome, i.e. scaffold, all without using a reference genome (29).
Contrary, the *map reads to reference* process utilizes a common reference for aligning the read to a specific place in the genome.

3. Materials and methods

3.1 Sampling

In this study, 100 clinical *P. aeruginosa* isolates from 72 patients (88 isolates from patients with CF and 12 isolates from non-CF patients) were collected from the Clinical Microbiology laboratory at Sahlgrenska university hospital, Gothenburg. The Clinical Microbiology laboratory is a reference laboratory for bacterial associated infections in CF patients, therefore receiving patient samples from all over the southern part of Sweden. The sample collection for the project was based on three criteria: patients with CF having PFGE type J or B (being the most common and

diverse PFGE types in the Gothenburg region (36)), patients with CF included in a biobank from the early 2000´s, and patients with an immunosuppression other than CF included in a medical ward outbreak. Anonymization of the isolates was done according to –patient number:isolate number:year- e.g. 2:1:2019, the first isolate from year 2019 from patient number two, prior to analysis with next generation sequencing technology and pulsed-field gel electrophoresis.

3.2 PFGE laboratory work

Isolates were cultivated at 37 °C overnight on agar plates supplemented with horse blood. The *P. aeruginosa* CCUG 49694 from Culture Collection University of Gothenburg was used as an internal reference. All isolates were analysed with PFGE based on a validated *in-house* method, as described below, to determine their macro-restriction profile. All substrates are provided from the department of substrate at Clinical Microbiology laboratory if nothing else is implied.
Briefly, bacterial cells were suspended in Pet IV buffer (Sodium chloride, tris(hydroxymethyl)aminomethane (TRIS), pH 7.6) to an optical density (OD) of 0.5 at 600 nm (BioPhotometer, 22331, Eppendorf AG, Hamburg Germany), whereafter the cells were incubated in 56°C for 15 minutes. The bacterial suspension was mixed with an equal volume of 1 % melted agarose (Agarose low EEO, Sigma-Aldrich, Darmstadt, Germany, dissolved in EC-buffer, pH 7,5, (TRIS HCl, Sigma-Aldrich, NaCl, ethylenediaminetetraacetic acid (EDTA), Sigma-Aldrich, polyoxyethylene (20) cetyl ether, Sigma-Aldrich, N-lauryl sarcosine, Sigma-Aldrich, sodiumdeoxycholat, Merck, Darmstadt, Germany, and distilled water pH 7.5)), to produce gelatin plugs, and treated with EC buffer and proteinase K (5 mg/ml, Sigma-Aldrich) for 20 h. After bacterial lysis, the gelatin plugs were washed for several times with TE buffer (TRIS HCl, Sigma-Aldrich, EDTA, Merck and distilled water, pH 7.6) followed by distilled water, pH 7.6. Thereafter, gelatin plugs containing total DNA were cleaved by restriction endonuclease *SpeI* Fast Digest (Life technologies, Carlsbad, CA, USA) in 37°C for 16 h. The gelatin plugs containing the cleaved DNA were loaded on an 1 % agarose gel (Ultra Pure™ Agarose, Invitrogen, Waltham, MA, USA, dissolved in 0.5xTBE buffer (TRIS base, Sigma-Aldrich, H_3BO_3, sodium EDTA, Merck, and distilled water) and placed in an alternating electric field (Chef Mapper™-system, Bio-Rad, Hercules, CA, USA) separating the DNA fragments according to size. Lastly, the gel was stained with ethidium bromide (5 mg/ml, Fisher scientific,

Hampton, NH, USA), visualising the DNA fragments with an UV light, and was photographed with a digital camera (Compact Digimage system UVDI, Major science, Saratoga, CA, USA).

3.2.1 PFGE analysis

Gel photos were imported in Bionumerics software version 7.6.3 with GelCompare II package (Applied maths, Kortrijk, Belgium). Relatedness studies were performed, using cluster analysis (Dice coefficient/UPGMA) and Tenover criteria (27), comparing fragment patterns against the *in-house* pulsotype database for *P. aeruginosa* at the Clinical Microbiology laboratory at Sahlgrenska university hospital. The *in-house* PFGE genotype classification is based on capital letters (e.g. A), and differences of one to six DNA fragments indicated by numerical suffix (e.g. A-1) (14).

3.3 NGS laboratory work

The next generation sequencing workflow on the Ion Torrent platform is described below and visualised in figure 3.1. The left boxes represent, from the top to the bottom, the major experimental steps from the genomic DNA extraction to the sequencing and data analysis. The right boxes representing the Quality Control steps.

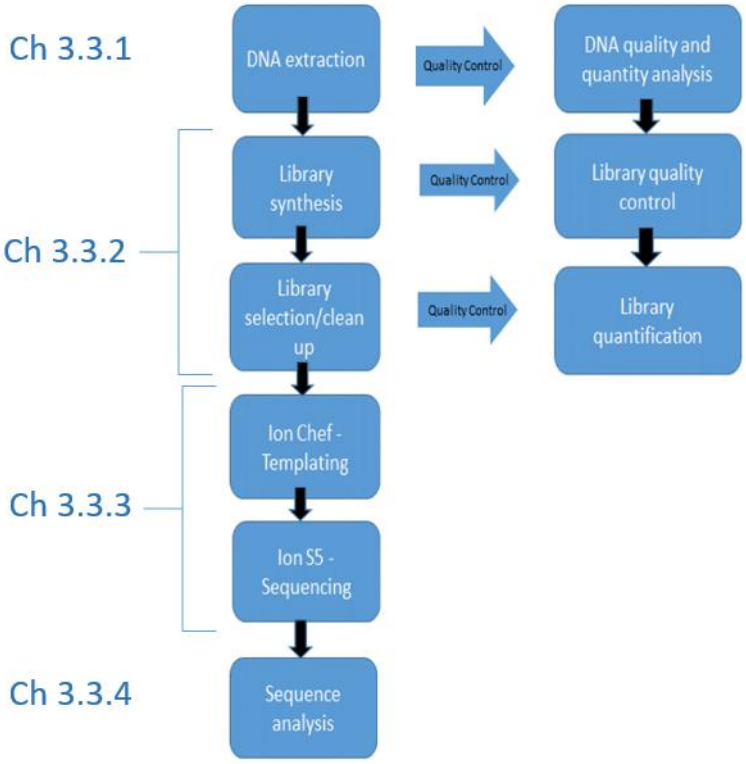

Figure 3.1. A schematic overview of the experimental workflow of whole genome sequencing on the Ion Torrent-platform. From the top left: DNA extraction, followed by library synthesis, library selection, ion chef templating, ion s5 sequencing and lastly sequence analysis. The first three steps are followed by quality control; DNA quality and quantity, library quality and lastly library quantification.

3.3.1 DNA extraction and quality control

All substrates were provided from the department of substrate at Clinical Microbiology laboratory if nothing else is implied. Isolates were cultivated at 37 °C overnight on agar plates supplemented with horse blood. The bacterial cells were suspended in phosphate-buffered saline (PBS) to McFarland 5 and genomic DNA (gDNA) was extracted according to an *in-house* protocol (37). gDNA was quantified and quality controlled to ensure the high quality needed for further NGS analysis. The quantity was measured using the Qubit® double stranded DNA High Sensitivity Assay kit on a Qubit® fluorometer (Thermo Fisher scientific) according to the user's manual MAN0003231. The purity of the gDNA was measured on a Nanodrop 2000 (Thermo Fisher scientific), where the presence of proteins and carbohydrates were measured at absorbance quotient of 260/280 and 260/230, respectively.

3.3.2 DNA library synthesis, purification and quantification

DNA libraries were synthesised in the automated AB Library Builder™ system (Applied Biosystems, Waltham, MA, USA) using Ion Xpress™ Plus fragment Library kit and the Ion Xpress™ Plus Library protocol card illustrated in figure 3.2. The kit was used according to the manufacturer's manual MAN0006946: 300 ng of gDNA was enzymatically sheared using the *automatic shearing and auto size-selection 300 bp* program, theoretically resulting in around 380 bp long fragments after adapter ligation of P1 (CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT),- and barcoded A-adapters (CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNGGTGAT), 40 bp long respectively.

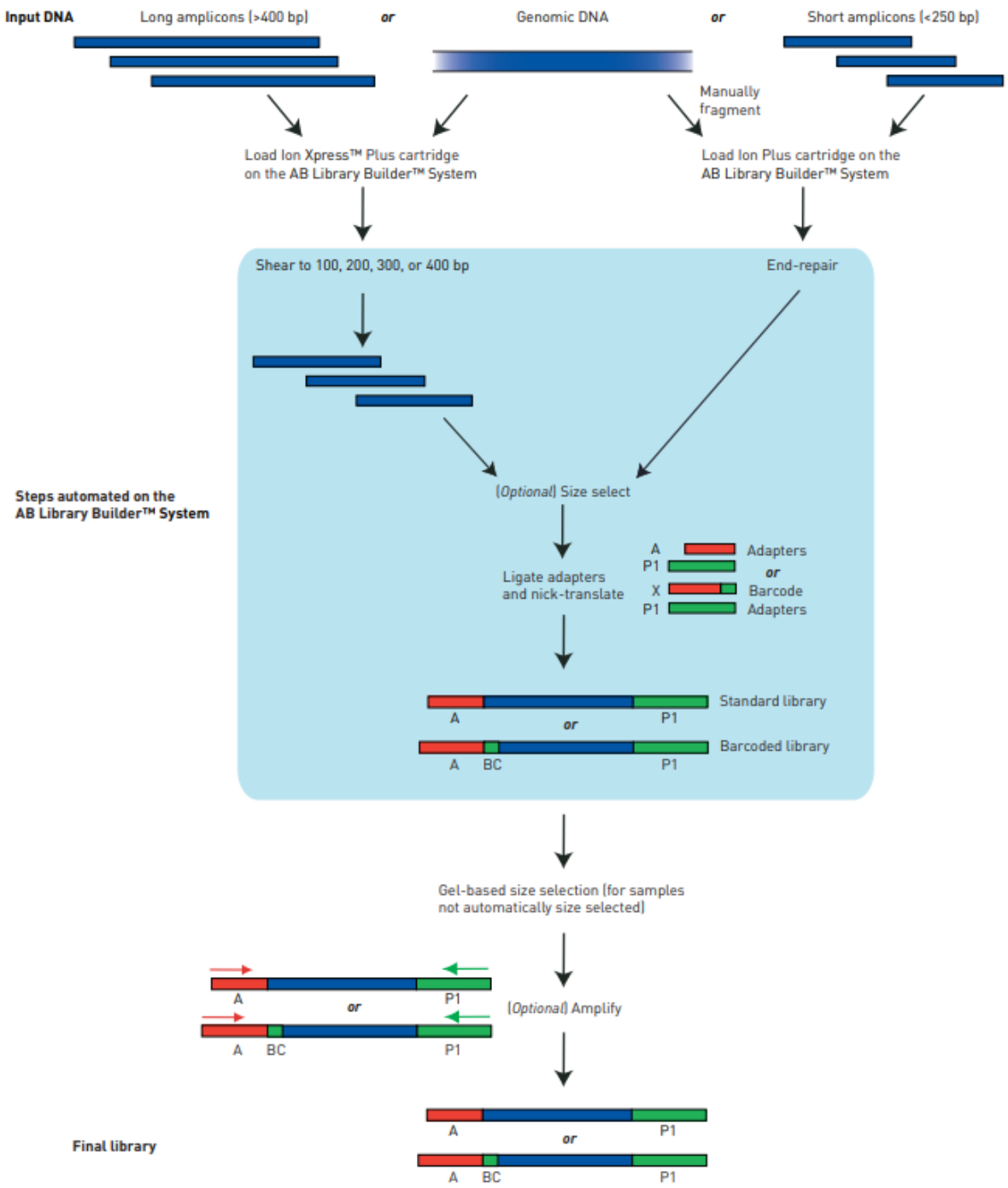

Figure 3.2. Library synthesis in the automated AB Library Builder system (Applied Biosystems, MAN0006946). Genomic DNA is sheared and adapter A and adapter P1 are ligated.

The library fragment length was visualized using 4200 Tape Station with High Sensitivity D1000 ScreenTape assay (Agilent Technologies, Santa Clara, CA, USA). According to the assay manual G2991-90001, libraries and ladder were mixed with sample buffer and run on a Screen Tape, whereafter the fragment length was estimated in TapeStation Analysis Software Ver. 2.2.

Libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) according to manual B37419AB, illustrated in figure 3.3.

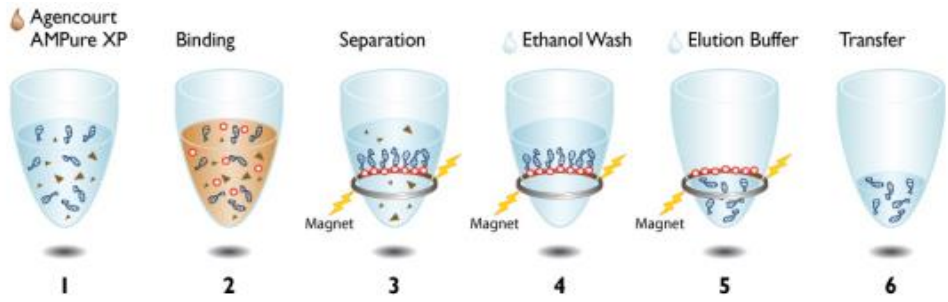

Figure 3.3. The magnetic bead purification using Agencourt AMPure XP beads (Beckman Coulter, B37419AB). Library solution was mixed with an equal volume of magnetic beads, incubated for five minutes before magnetic separation of beads and solution. Following two washing steps with 70% EtOH (99.0% EtOH, from the department of substrates, diluted in distilled water), and finally elution in LowTE buffer (the department of substrates) for five minutes.

After purification, libraries were quantified on a Quant Studio 5 Real-time PCR instrument (Applied Biosystems) with Kapa probe fast mix, 2x, including rox high (Sigma-Aldrich). The library solution was diluted 1:10 000 prior to quantitative PCR using the primer/probe-mix: A-17_Ion-Quant: 5´-CCATCTCATCCCTGCGT-3´ and P1_Ion-Quant: 5´-CCTCTCTATGGGCAGTCGGTGAT-3´ (IDT, Coralville, IA, USA) and MGB probe 6-FAM 5´-CTGAGTCGGAGACACGC-3´ (IDT). *E.coli* DH10B Control 400 Library (Thermo Fisher scientific) and UltraPure™ DNase/RNase-free distilled water (Invitrogen, Waltham, MA, USA) were used as positive and negative controls, respectively. The PCR analysis started with an initiation cycle 50°C for 2 min and 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. The quantity was calculated dependent on CT-value, according to a formula validated *in-house* (valid for CT-values ranging between 20-23), (equation 3.1).

(Eq. 3.1) $= 99448 * 2.718^{(-0.695*CT)}$

3.3.3 Sequencing on Ion S5 system

Prior to sequencing on the Ion Genestudio™ S5 Prime System (Thermo Fisher scientific), *a planned run* i.e. sequence template, was set up on the iontorrent-server connected to the instruments. Dependent on genome size (*P. aeruginosa* ~6.2 Mb (38)), number of genomes to be sequenced and the desired coverage (~60-80, minimum 30x coverage) a chip size was chosen 0.5 GB, 2 GB or 6 GB (Ion 510™, 520™ and 530™ Chip, respectively, Thermo Fisher scientific). Different biological

agents, such as bacteria and viruses, can be pooled together on the same sequencing chip. In theory, isolates are pooled together with respect to the desired coverage and size of the genomes, leading to *P. aeruginosa* calculated almost three times the space as a methicillin resistant *Staphylococcus aureus* (MRSA, 2.8 Mb (39)) (sequenced at a regular basis at Clinical Microbiology). Libraries were diluted and pooled according to a loading of 25 µl of 35 pM solution on the Ion Chef™ instrument (Thermo Fisher scientific). The kit's supplies and solutions (Ion 510™ & Ion 520™ & Ion 530™ Kit) were loaded on the instrument according to protocol MAN0016854.

The Ion GeneStudio™ S5 Prime System was loaded with Ion S5 Sequencing Solutions and Reagents and initialized prior of moving the templated chip from the Ion Chef™ Instrument to the chip holder of the S5.

3.3.4 NGS data analysis

The software in the Ion Genestudio™ S5 instrument performs an automated initial adapter trimming of the P1, - and barcoded A-adapters, before the raw reads are sent to the ion torrent-server. After brief quality control: minimum 30x average coverage and 300 bp mean length, respectively, data analysis was performed on three commercially available software utilizing three different bioinformatics pipelines. 1) CLC genomics workbench 12.0.3 – Microbial module (Qiagen, Hilden, Germany) using single nucleotide polymorphism (SNP) analysis, 2) Bionumerics 7.6 (Applied maths) using pub,- and whole genome multilocus sequence typing (MLST) and 3) 1928 diagnostics (Gothenburg, Sweden) using core genome MLST.

3.3.4.1 CLC genomics workbench 12.0.3 – Microbial module, Qiagen

Sequences were imported in CLC. For SNP analysis, the workflow *Map reads to variable reference*, figure 3.4, was used together with *P. aeruginosa* ASM676v1 reference genome (downloaded at National Center for Biotechnology Information, NCBI, accession NC_002516). All utilized settings listed in Appendix I – Settings in CLC genomics workbench, Qiagen.

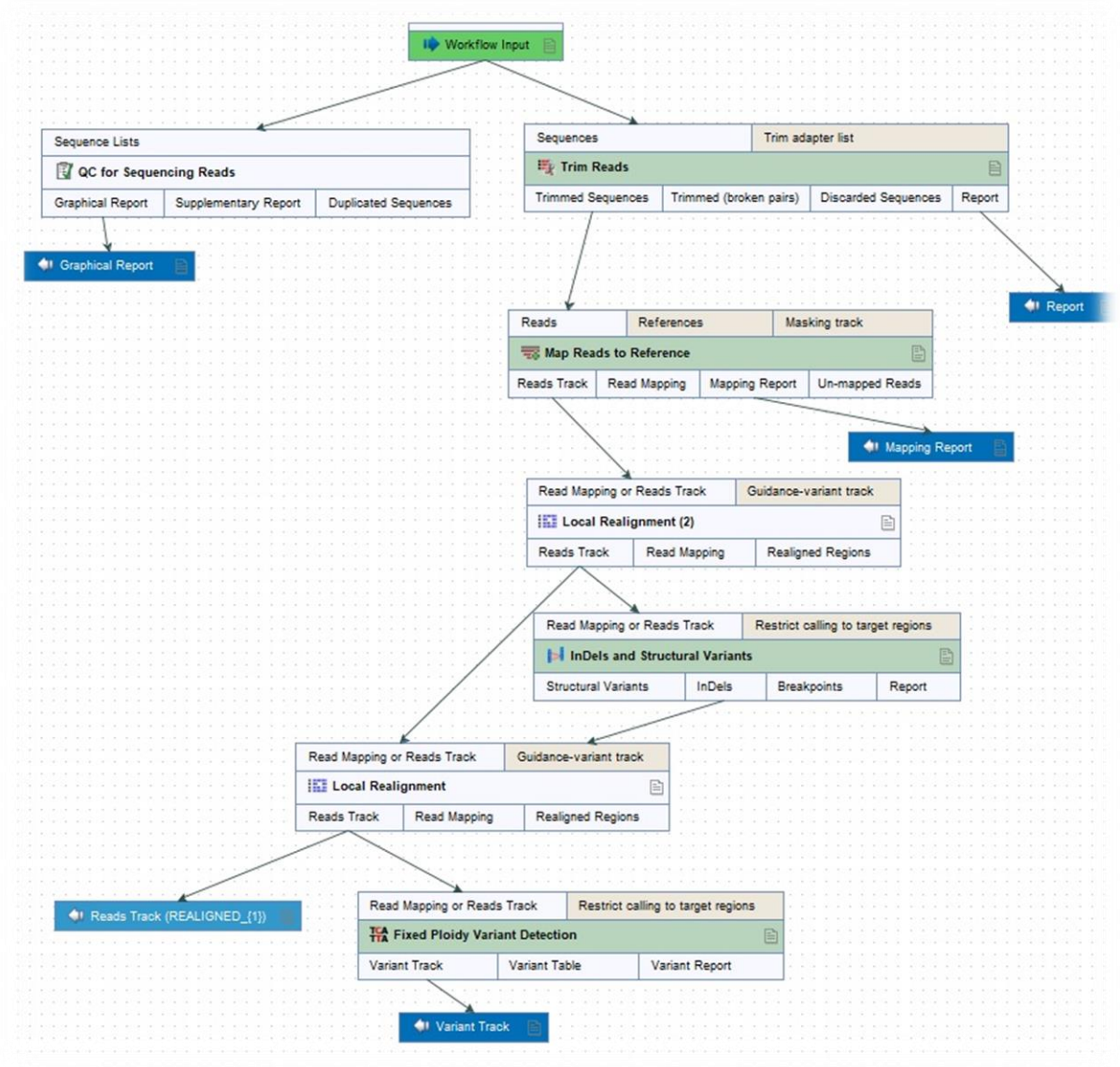

Figure 3.4. Map reads to variable reference- workflow in CLC genomics workbench 12.0.3, Qiagen. Settings utilized in Appendix I – Settings in CLC genomics workbench, Qiagen.

Isolates were pair-wise compared using *SNP analysis* and *pubMLST analysis*. *SNP analysis* settings: minimum of 20 coverage, prune distance of 20 and neighbor joining clustering. Visualised as a phylogenetic tree. The *pubMLST analysis* based on seven housekeeping gene sequences (sequences and MLST profiles downloaded at pubMLST.org), table 3.1 (40).

__Table 3.1__. Seven housekeeping genes included in the MLST scheme downloaded at pubMLST.org (40).

Abbreviation	Gene
acsA	Acetyl coenzyme A synthetase
aroE	Shikimate dehydrogenase
guaA	GMP synthase
mutL	DNA mismatch repair protein
nuoD	NADH dehydrogenase I chain C, D
ppsA	Phosphoenolpyruvate synthase
trpE	Anthralite synthetase component I

The *de novo-workflow*, figure 3.5, with utilized settings for the analysis in Appendix I – Settings in CLC genomics workbench, Qiagen.

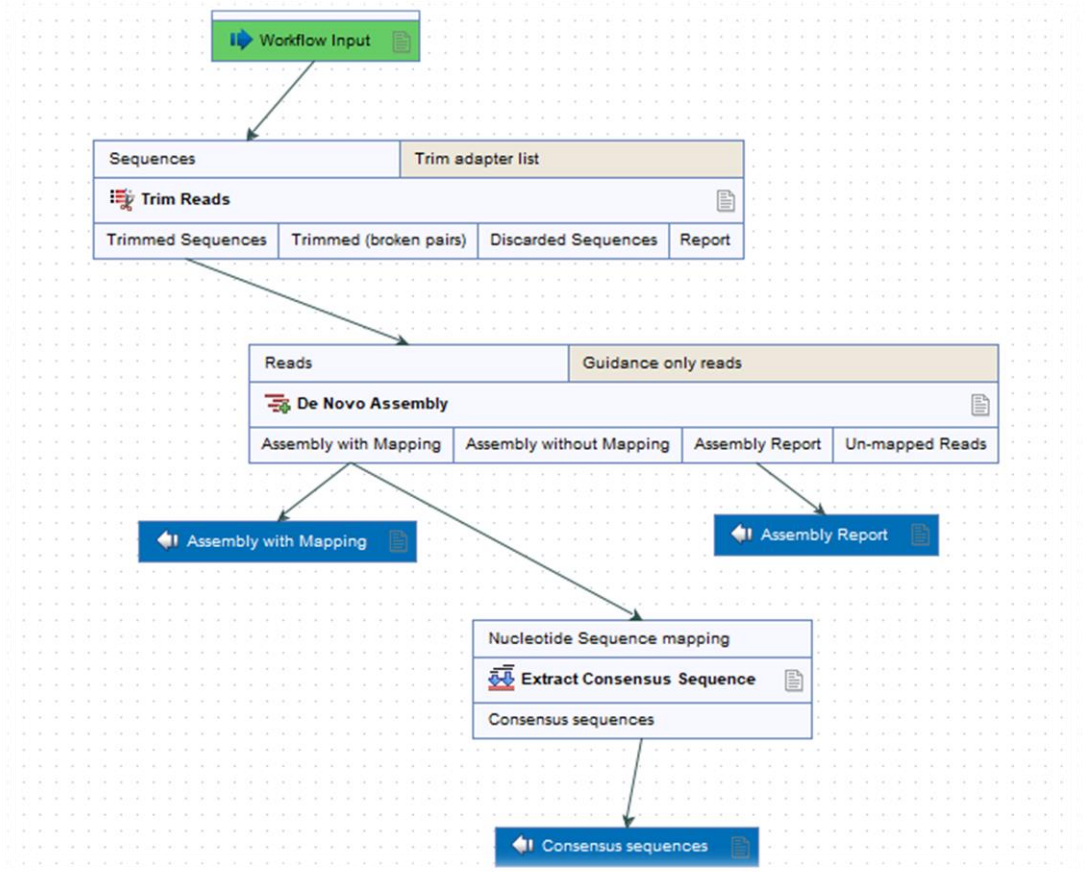

Figure 3.5. De novo assembly-workflow used for P. aeruginosa in CLC genomics workbench 12.0.3 (Qiagen). Settings utilized in Appendix I – Settings in CLC genomics workbench, Qiagen.

3.3.4.2 Bionumerics 7.6 - Applied maths

Sequences were imported in Bionumerics 7.6, both as sequence read sets and *de novo* trimmed assembly files (from CLC genomics). *WgMLST* and *pubMLST analyses* through an automated curated database, *WGS tools plugin* (default

settings), giving the allelic differences. Sequence types (ST) were given based on the pubMLST scheme (the seven previously mentioned housekeeping genes) and the wgMLST is based on 400 reference genomes, with 15,143 loci defined. A neighbor joining clustering was applied for sequence comparisons.

3.3.4.3 1928 diagnostics – Gothenburg, Sweden

Sequences were imported as raw reads and analysed in the *P. aeruginosa* pipeline, with default settings. Coverage cut off of 30x. Thereafter a cgMLST comparison based on a local database including 3927 genes. Sequence types were also generated from pubMLST.

4. Results

There were 100 isolates in total sequenced, generated from 72 patients. All patients, except 10 patients included from the medical ward outbreak, are patients with cystic fibrosis. In total there were 40 different PFGE types of which 20 isolates belonged to PFGE type B and 12 isolates belonged to PFGE type J.

4.1 DNA library synthesis, purification and quantification

Library synthesis resulted in fragment sizes between 200 and 700 bp distributed as a peak at ~410 bp, visualized with Tape Station in left part of figure 4.1. The right part of figure 4.1 shows the purified library. The purification of shorter reads <100 bp changed the mean length to ~425 bp.

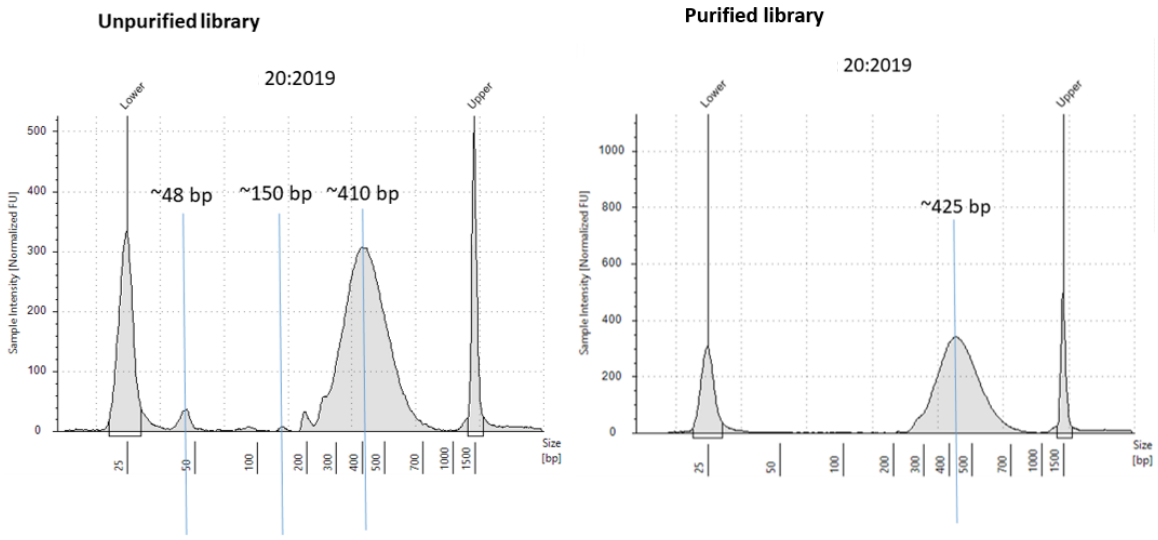

Figure 4.1. Tape Station visualisation of unpurified library on the left, and purified library on the right. The lower and upper peaks are used for aligning the sample relative the ladder as well as calculating the concentration in the sample. The small peak at 48 bp (in the unpurified library) are unbound adapters. Y axis: Sample intensity [Normalised FU] the x-axis: the size distribution peaks of isolate 20:2019. The majority of the library in the size range of 200 and 700 bp. The mean length of the library ~425 bp.

The pooling of *P. aeruginosa* showed to be more complex than earlier experienced, where the contribution of three times as much library of *P. aeruginosa* resulted in a very high coverage of *P. aeruginosa* and a too low coverage of other bacteria and viruses on the same chip. Modifications led to lowering *P. aeruginosa* library input by half. However, pooling with mycobacteria, even with modifications mentioned above, the coverage for mycobacteria were too low, leading to the two bacterial agents not being sequenced on the same chip.

4.2 Data analysis

The data analysis section presents the overall results showed on the three NGS analysis platforms, in comparison to the PFGE data. Further is a closer focus on the clusters, PFGE type B, PFGE type J and the medical ward outbreak.

Fragment patterns from PFGE analysis of a part of all isolates are represented in figure 4.2. To the left is the UPGMA clustering, with branch lengths corresponding to percentage similarity. In the middle are the fragment pattern for each isolate. To the right are PFGE type and isolate ID. Blue circles highlighting PFGE type J and PFGE type B. It can be seen in figure 4.2 that all type J as well as all type B isolates cluster

together, respectively. Further does PFGE type J isolates show a high diversity within the cluster of ~50 % similarity. In comparison does the PFGE type B cluster show a similarity of ~70 %.

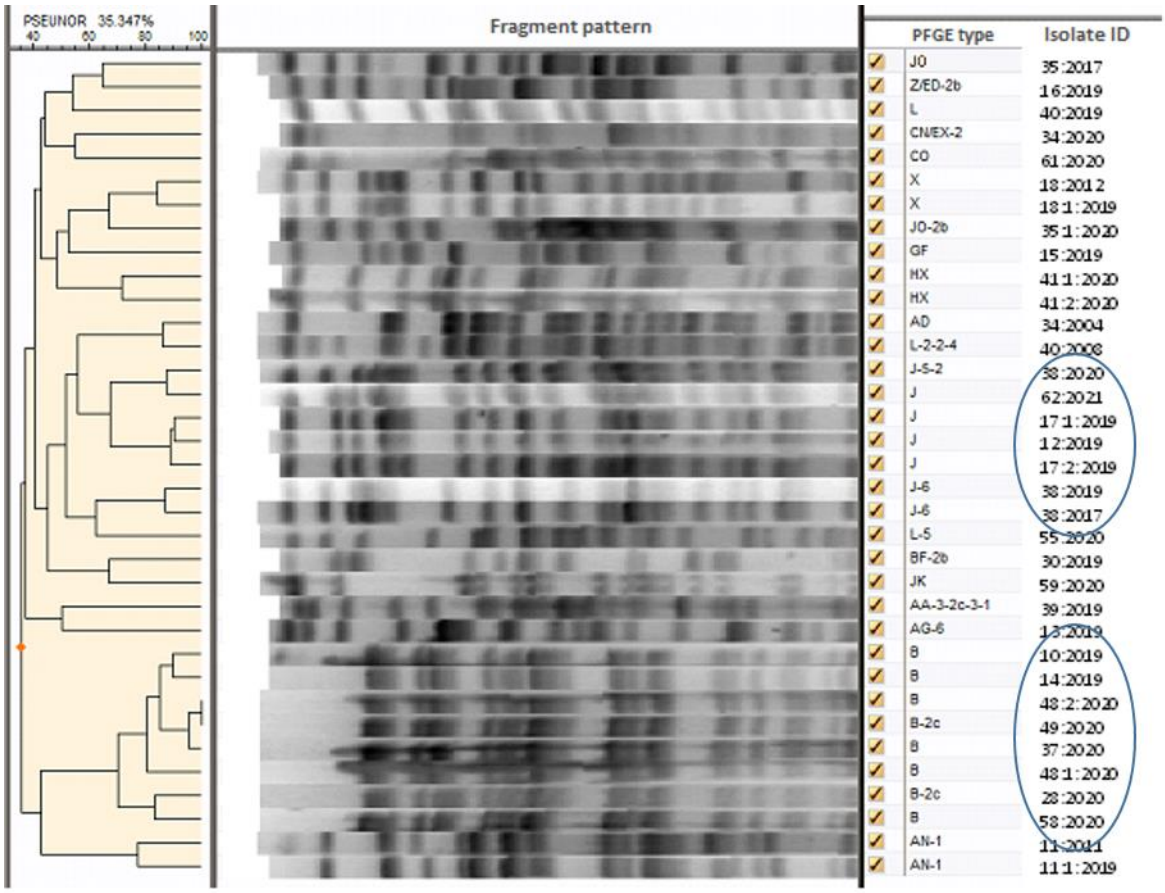

Figure 4.2. Fragment patterns from 35 isolates, analysed with PFGE, visualised as a phylogenetic tree on the Bionumerics platform. To the left: the UPGMA clustering with branch lengths corresponding to percentage similarity, in the middle: aligned gel photos of fragment patterns, and to the right: PFGE type and isolate ID. In blue circles are PFGE type J and PFGE type B.

The 100 sequenced isolates are represented in a SNP analysis in CLC resulting in a phylogenetic tree figure 4.3. All isolates having PFGE type J (marked with a yellow line) and PFGE type B (marked with a blue line) are well organized in separate clusters. Also this analysis shows a high diversity of collected *P. aeruginosa* strains from patients with CF, with forty different PFGE types represented. The isolates collected from the medical ward outbreak, being non-CF patients, are spread out randomly having different PFGE types, except from the three clusters; NM, KD and NF (marked with orange lines).

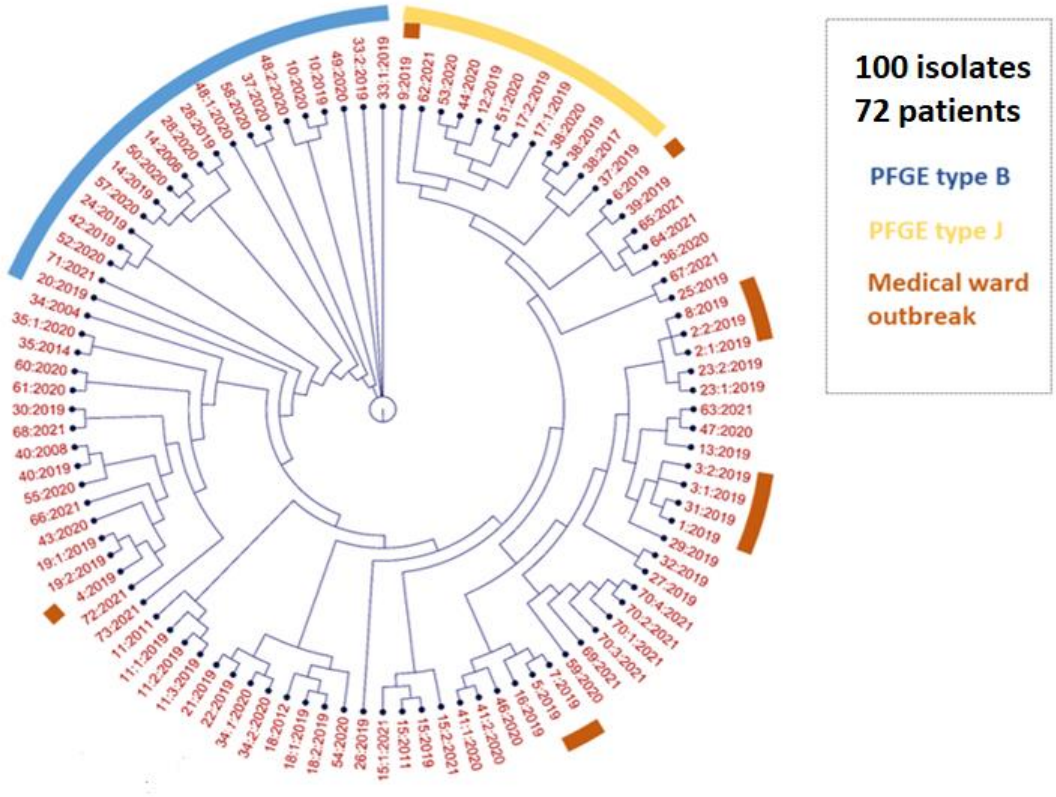

Figure 4.3. 72 patients with a total of 100 isolates were sequenced. The results from the SNP-analysis, in CLC genomics workbench, visualised as a circular cladogram. Isolates having PFGE type B are marked with a blue line, isolates having PFGE type J are marked with a yellow line, and isolates belonging to the medical ward outbreak are marked with the orange lines.

4.2.1 PFGE type J – Cluster analysis

Cluster J is known to be diverse, from previous PFGE data at the Clinical Microbiology laboratory, therefore a clustering analysis of isolates having PFGE type J was performed and compared to analysis of NGS-data. In total it includes twelve isolates from nine different CF patients without additional epidemiological data. Figure 4.4 shows the cluster result from eight out of twelve isolates analysed in Bionumerics. The PFGE type is clearly diverse showing as little as ~50 percent similarity. The two isolates, collected one year apart, from patient 38 show a high diversity with several band differences (PFGE type J-5-2 and J-6 respectively). Also, the two isolates 37:2019 and 38:2019, although both having PFGE type J-6, only show a 50 percent likeness. Contrary, does the two isolates from the same year, from patient 17, show identical band patterns and a 90 percent likeness. The

different PFGE type J isolates (e.g. J, J-5-2, J-2 and J-6) clusters randomly within the J cluster.

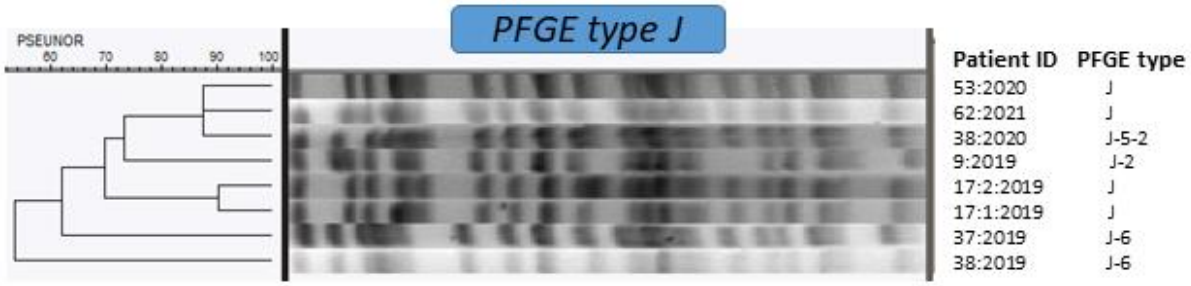

Figure 4.4. Cluster analysis of PFGE data in Bionumerics. Isolates belonging to the PFGE cluster J, having PFGE types J, J-2, J-6 and J-5-2. To the left: the UPGMA clustering with branch lengths corresponding to percentage similarity, in the middle: aligned gel photos of fragment patterns, and to the right: PFGE type and isolate ID.

Figure 4.5 shows the phylogenetic tree of neighbor joining clustering of all isolates having PFGE type J, in CLC.

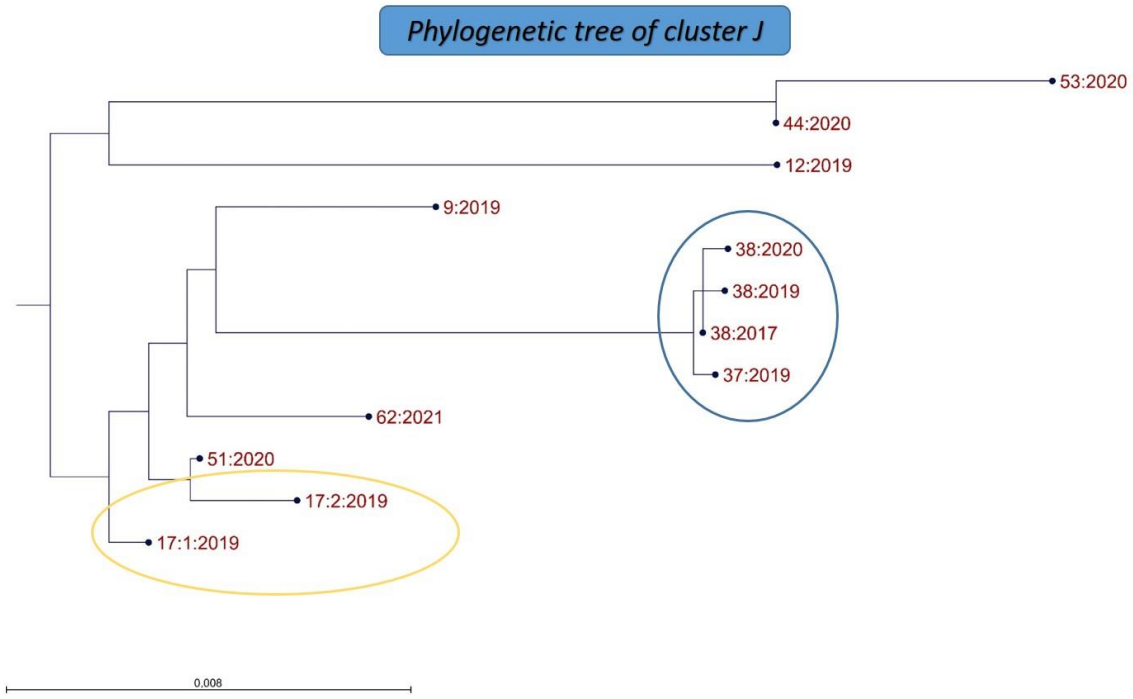

Figure 4.5. A phylogenetic tree of the SNP analysis in CLC genomics workbench. Visualised are the isolates belonging to the PFGE cluster J. In the blue circle: three isolates from patient 38 and one sample from patient 37, less than 10 SNP´s in between. In the yellow circle: two isolates from patient 17, 73 SNP´s in between. (Number of SNP´s difference for cluster J visualised in Appendix III – Bioinformatics analysis results).

Previously mentioned patient 37 and 38 (marked with a blue circle) are on one node in close proximity to one another, while in comparison, the two isolates from patient 17 are more diverge and ended up on two different nodes. Moreover, the cluster is divided into two separate parts, where there are isolates having PFGE type J on both branches, while the PFGE types J-2, J-5-2 and J-6 (patients 9, 37 and 38) are all organized furthest out on the lower branch.

The wgMLST analysis in Bionumerics, figure 4.6, supports the divergence within the cluster, previously showed in both the PFGE analysis and the SNP analysis. To the left in figure 4.6 the two separate parts of the cluster is visualized, with patient 37 and 38 grouping furthest down the branch, contrary to the PFGE data. The wgMLST percentage similarity is visualized to the right, highlighting patient 37 and 38 in the blue square, showing a >99.5% similarity. Also are the two isolates from patient 17 showing only a 98.6% similarity.

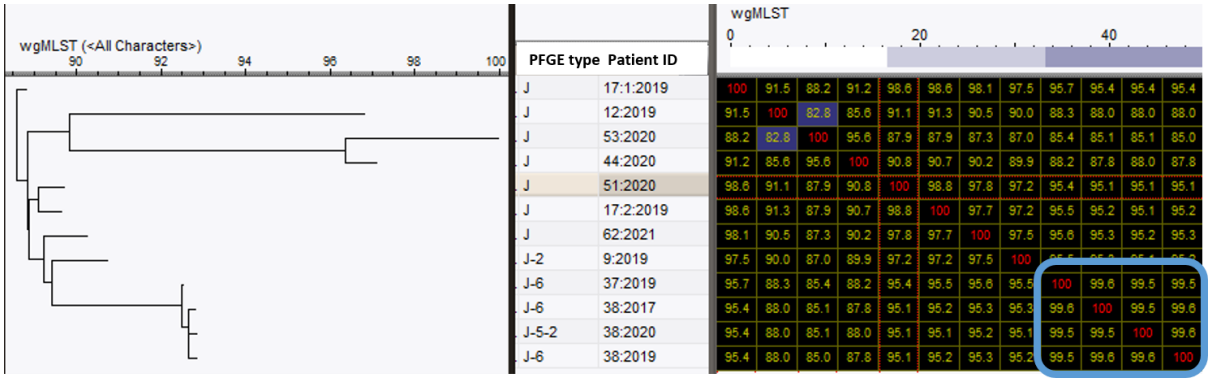

Figure 4.6. A phylogenetic tree of the whole genome MLST analysis in Bionumerics. Visualized is the PFGE cluster J. In the left: the tree branches with lengths corresponding to percentage similarity, in the middle: cluster type and patient ID, in the right: the matrix giving percentage similarity. The blue box corresponds to four different isolates, within 99.5 % similarity, from two different patients.

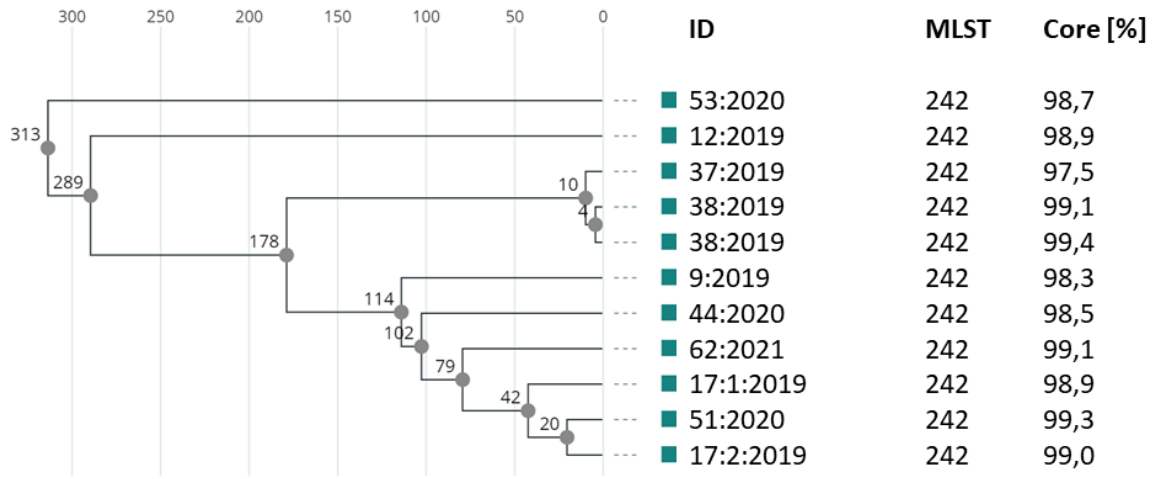

Figure 4.7. A phylogenetic tree of the core genome MLST analysis in 1928 diagnostics. Visualised are the isolates belonging to the PFGE cluster J, except of 38:2020 having too low coverage (<30) to get accepted on the platform.

The cgMLST analysis on the 1928-platform, figure 4.7, shows up to 313 alleles difference within the cluster J, although all having the same pubMLST type 242 (shown on the right in the figure). Also notable on the right is the core percent for each isolate. The number given is the sequenced percent coverage of the core genome. As already visualized previously, there are some differences between the two isolates from patient 17 (42 alleles). Further are patient 37 and 38 grouping as before, only differing in 10 alleles. The 1928-platform has a lower limit of 30x coverage for upload and further analysis, therefore isolate 38:2020 is missing in the cgMLST analysis in figure 4.7.

4.3.2 PFGE type B – Cluster analysis

In figure 4.8 are all isolates having PFGE type B or B-2C, analysed with PFGE in Bionumerics. To the left is the clustering, with as low as 65% similarity between the two main branches of the cluster. Isolates having PFGE type B and B-2C are all spread out randomly. Neither are double isolates from the same patient (14, 28 and 48) clustering together. However, without supporting epidemiological data, are isolates 48:2:2020 and 49:2020 and isolates 42:2019 and 20:2019 respectively,

cluster on a hundred percent similarity.

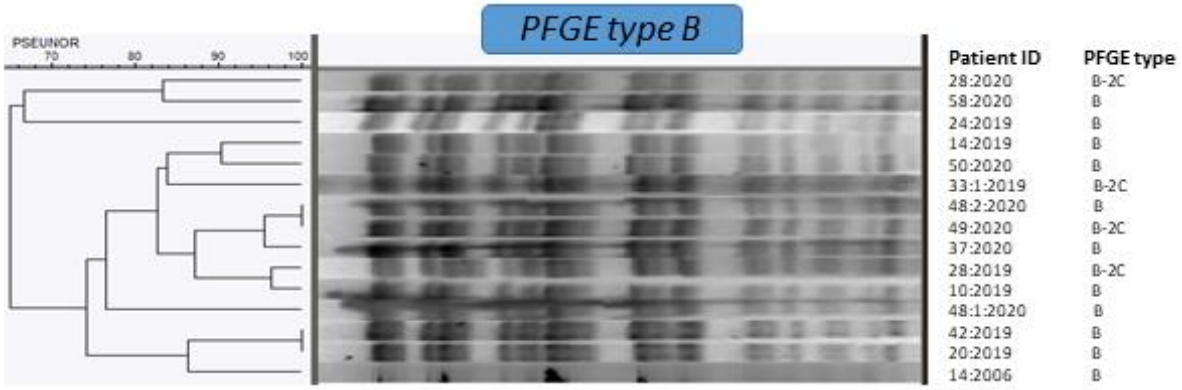

Figure 4.8. Cluster analysis of PFGE data in Bionumerics. Isolates belonging to the PFGE cluster B, having PFGE types B and B-2C. To the left: UPGMA clustering percentage likeness, in the middle: gel photos, to the right: patient ID and PFGE type.

The SNP analysis in CLC, figure 4.9, divides the isolates in to several small clusters. There are four highlighted clusters of which the three blue ones contain double isolates from the same patient, and the yellow one contain two different patients. Contrary to the PFGE data, all double isolates cluster together with as little as one SNP´s difference (patient 33) to as high as 331 SNP´s difference (patient 48) (Appendix III – Bioinformatics analysis results).

The overall divergence within the cluster is ~1600 SNP´s.

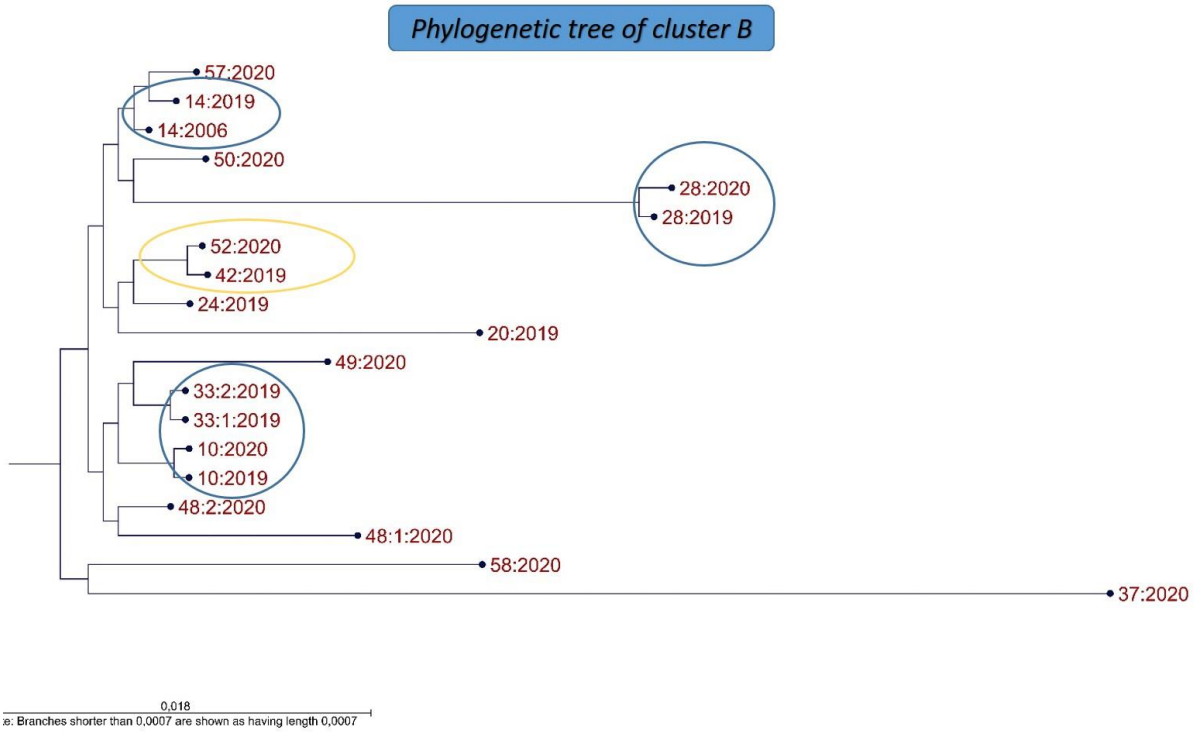

Figure 4.9. A phylogenetic tree of the SNP analysis in CLC genomics workbench. Visualised are the isolates belonging to the PFGE cluster B. In the blue circles are isolates belonging to the same patient, and in the yellow circle are two isolates from different patients. (Number of SNP´s difference for cluster B visualised in Appendix III - Data analysis).

The wgMLST clustering in Bionumerics, figure 4.10, show similar result to the SNP analysis. The double isolates from the same patients cluster with up to 100% identity (figure III.1, Appendix III – Bioinformatics analysis results), as well as the two different patients 42 and 52 (highlighted in blue and yellow respectively). Contrary from the SNP analysis, where patient 37 and 58 diverge alone, the wgMLST analysis clusters them furthest out on one branch. The pubMLST analysis in Bionumerics typed all isolates except 37:2020 and 48:1:2020 (N/A) to ST809.

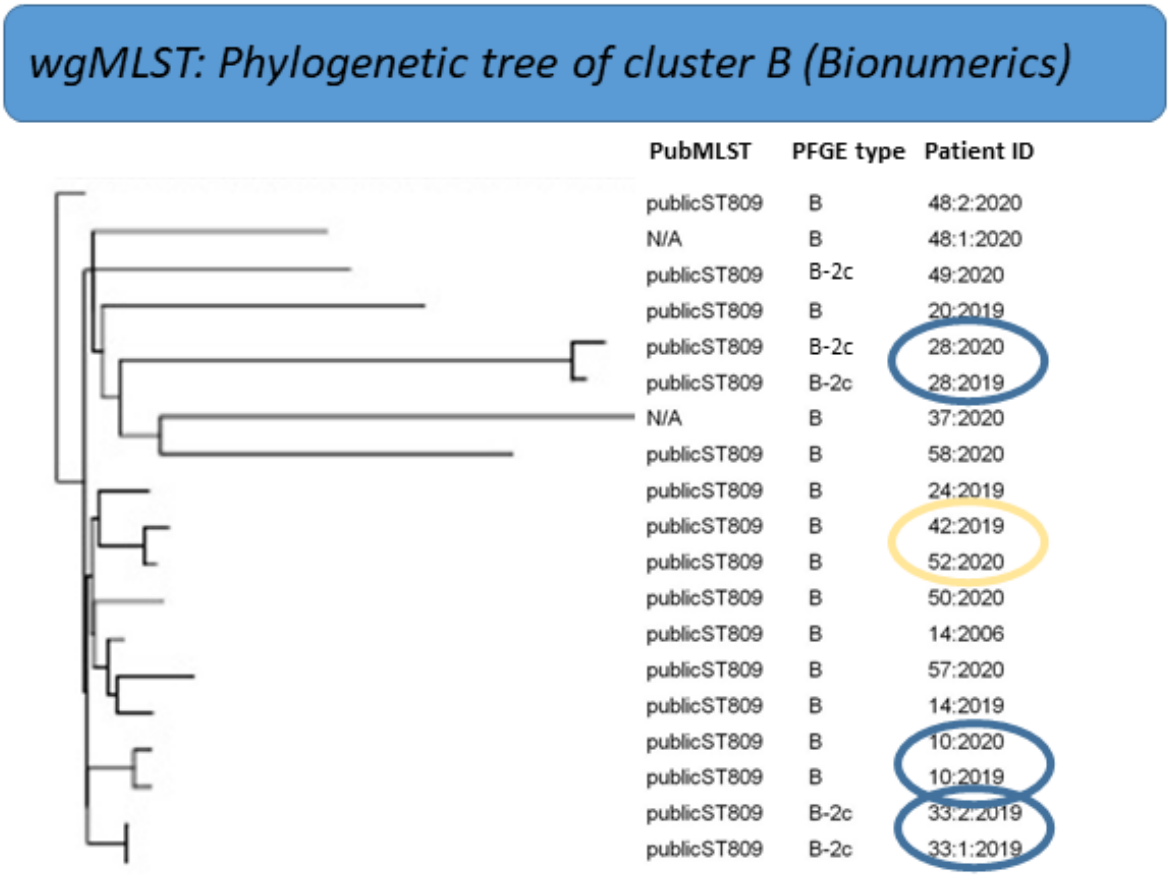

Figure 4.10. A phylogenetic tree of whole genome MLST analysis of cluster B in Bionumerics. Additionally to the wgMLST clustering are the pubMLST type (ST809) and PFGE type (B, and B-2c). In the blue circles are some of the double isolates from the same patient. In the yellow circle is the two isolates from different patients with only 23 SNP´s difference.

Lastly are the results from 1928D, where the cgMLST analysis show similar clustering as before, figure 4.11. All double isolates, except patient 48, cluster closely with between zero and fifty-three alleles difference. The pubMLST type is ST809 for all isolates except 37:2020 (Novel). The isolate 37:2020 is furthest away with as many as 864 alleles difference.

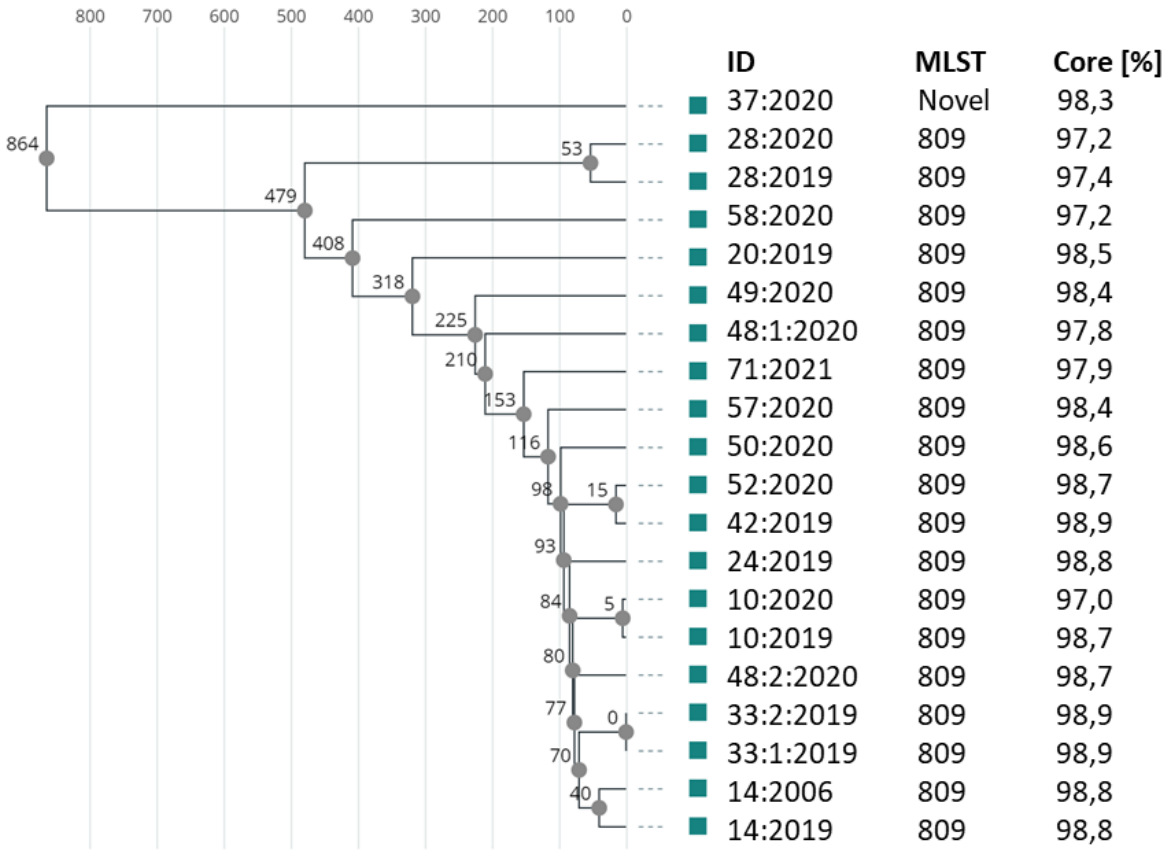

Figure 4.11. A phylogenetic tree of the core genome MLST analysis in 1928 diagnostics. Visualised is the PFGE type B cluster. To the left: the clustering, to the right: the patient ID, pubMLST type and core percent.

4.3.3 Medical ward outbreak – Cluster analysis

In addition to the isolates from patients with CF included in the project, were patient samples from an outbreak at a medical ward in Gothenburg. In comparison, there is an epidemiological relationship for the latter. In figure 4.12 is an UPGMA clustering in Bionumerics of the gel photos resulting from the PFGE analysis of ten patients (in blue circles) involved in the outbreak. Additionally are patient 27 and patient 29, epidemiologically linked to each other but not to the outbreak, included here though because of their pattern similarities.

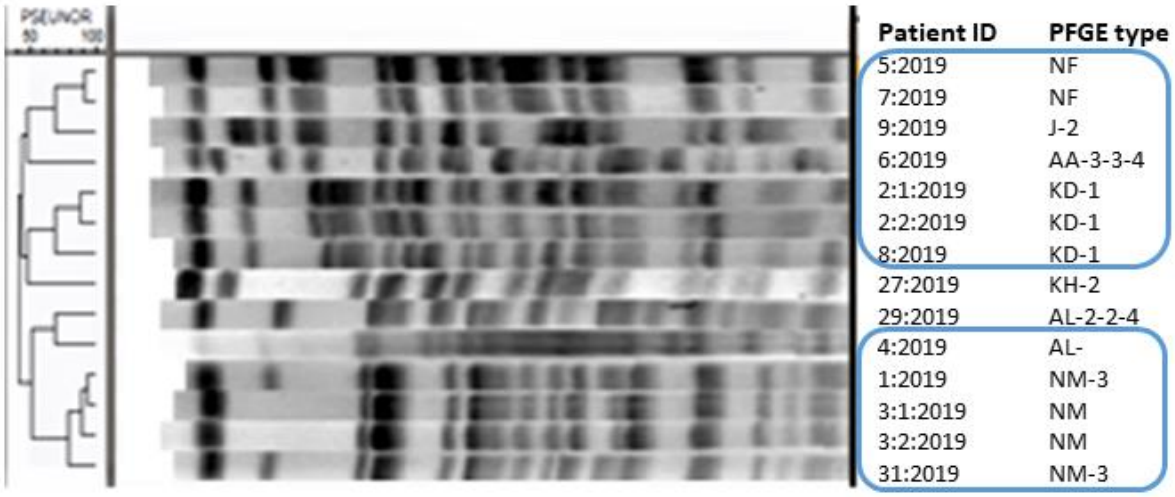

Figure 4.12. UPGMA clustering in Bionumerics of results of the PFGE analysis. To the left: UPGMA clustering, in the middle: gel photos, and to the right: patient ID and PFGE type.

Visualised are two branches, and three clusters containing PFGE types NM, KD and NF. The NM cluster contains three patients (and four isolates), the KD-1 cluster contains two patients (and three isolates) and the NF cluster contains two patients (and two isolates). Isolates within the same cluster are >75% similar.

The phylogenetic tree for the SNP analysis in CLC genomics workbench, figure 4.13, shows similar results to the PFGE analysis previously seen. Three distinct clusters (in blue circles), involving two to three patients each. The SNP matrix (table 3, Appendix III – Bioinformatics analysis results) show zero to eight SNP´s difference within the three clusters and up to ~33 000 SNP´s between isolates with different PFGE types.

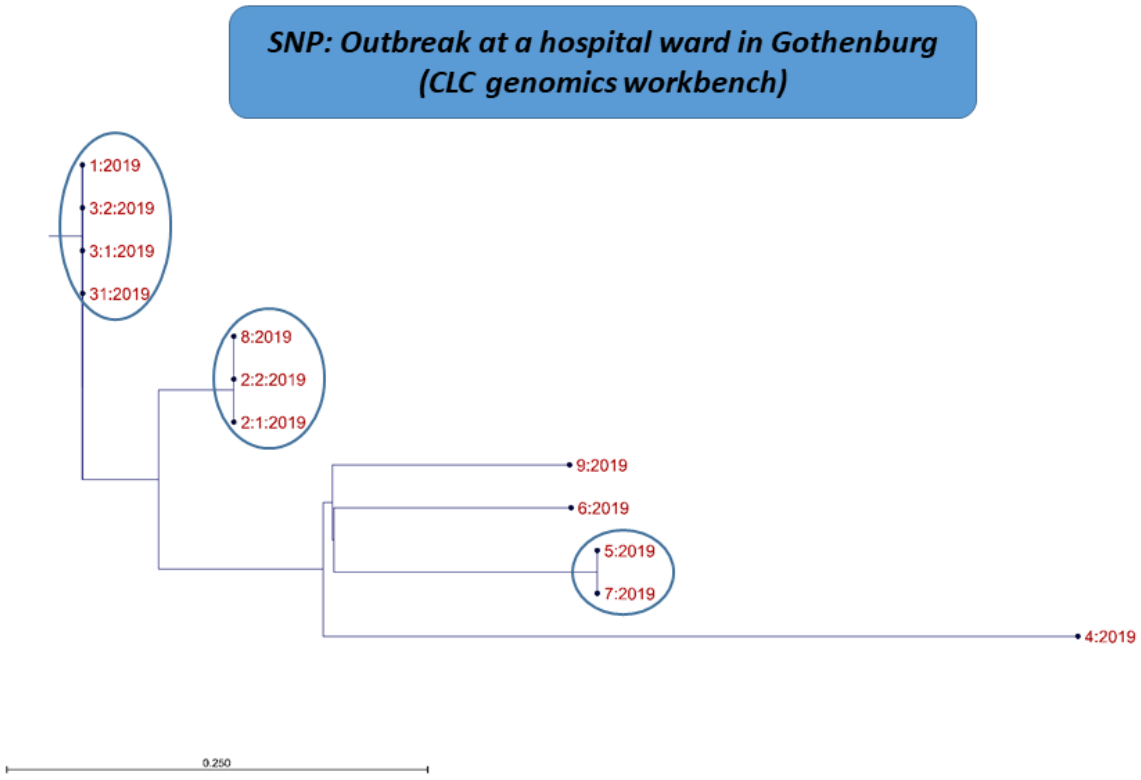

Figure 4.13. SNP analysis, in CLC, of isolates collected from an outbreak at a hospital ward in Gothenburg. Isolates in blue circles differ less than ten SNP´s.

The wgMLST analysis in Bionumerics, figure 4.14, also support the previous results suggesting three distinct clusters, highlighted in blue. The isolates within the clusters are >99.7% similar to each other. It is also clear that there is very little similarity between the different PFGE types.

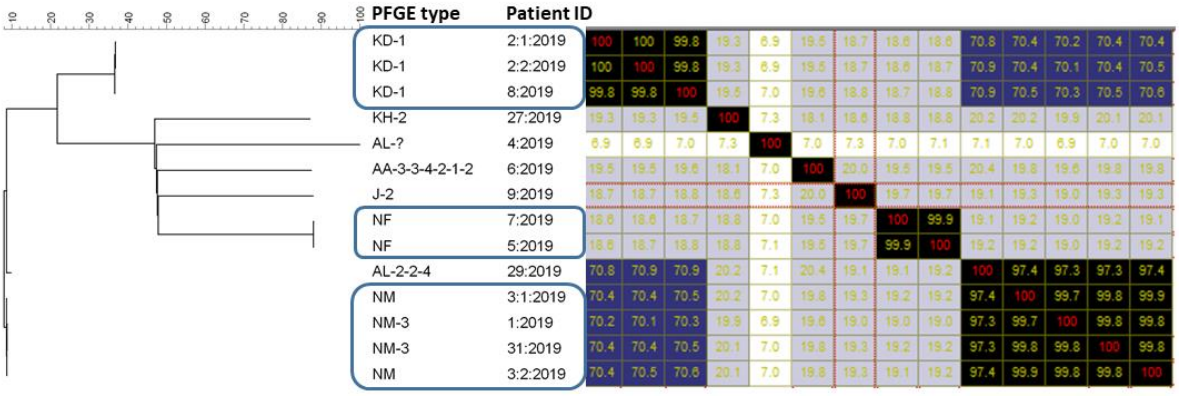

Figure 4.14. Neighbor joining clustering of wgMLST analysis in Bionumerics. Patients involved in a medical ward outbreak in Gothenburg, additionally two patients not linked epidemiologically but with similar PFGE type. In blue circles are isolates from patients belonging to the same clusters.

Lastly, the results from 1928D, in figure 4.15, the three clusters containing two to three patients differ up to seven alleles. Also stated are the pubMLST types, identical

for those isolates within the same cluster, and covered core genome percent. It is a distinct difference between isolates with the same origin (up to seven allele's difference) and isolates derived from different origins (from 1111 to 3597 allele's difference).

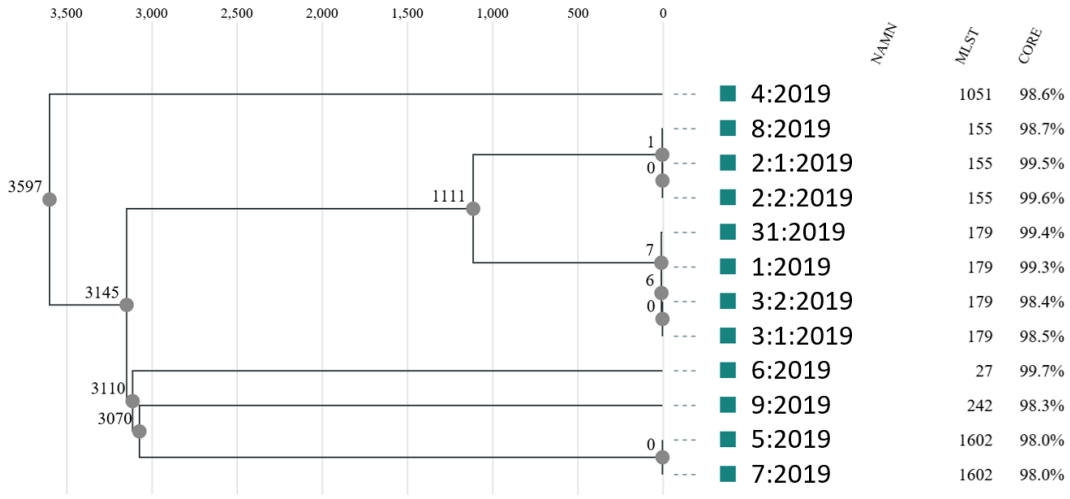

Figure 4.15. A phylogenetic tree of the cgMLST analysis in 1928 diagnostics. Visualised are the isolates included in the medical ward outbreak. To the left the clustering and allele differences, to the right the patient ID, pubMLST type and core percent.

5. Discussion

The already validated laboratory workflow for NGS analysis worked fine for *P. aeruginosa*, until the pooling of libraries of different species origin (mycobacteria, viruses and so on). As already touched upon the pooling is based on genome size and desired coverage, i.e. the bigger genome the bigger space on the chip. *P. aeruginosa* being 2.5 the size of MRSA, when given the corresponding space on the chip, it completely took over the chip though, resulting in a very low to none coverage of other libraries such as from mycobacteria. The high GC-content of *P. aeruginosa* and mycobacteria could possibly lead to the so called GC bias, causing uneven sequencing coverage over the genomes (34), (41). Chen et al looked at Illumina data from six different genomes to examine how GC bias can affect genome assembly, and they saw that genomes with very high or low GC content affects the sequence depth in the extreme GC areas (42).

5.1 Data analysis - PFGE

The subjectivity of PFGE data analysis as well as the difficulty in interlaboratory exchange due to the lack of a standardized nomenclature has previously been stated (7)-(10). The PFGE types has been specified in the Clinical Microbiology laboratory and are not communicable to other laboratories. When analysing PFGE data, a new fragment pattern is included in the database of Bionumerics and compared to other isolates fragment patterns. A huge drawback when analysing the PFGE data is its difficulties in comparison between fragment patterns obtained from different PFGE runs. Although each pattern intensity is normalized to a pattern of a standard strain of the bacteria, the comparison of results obtained from different PFGE runs is difficult.

5.2 J cluster comparison

The PFGE type J cluster, includes isolates with up to six band differences (i.e. J-6) from the original J type (i.e. J). Four to six band differences have previously been set as a maximal cut off for possible relatedness (27). Further is >87% band similarity considered genetically related (43). With that in mind, figure 4.4, visualizing band fragments for isolates belonging to the J cluster, show a lower than expected percentage likeness at around 50%. It is probably a result from making comparisons of band fragments obtained from different PFGE runs due to the light intensity variations in both band,- as well as no-band areas.

All three bioinformatics analysis software showed comparable results to that of PFGE, clustering all isolates having PFGE type J. The discriminatory level was further increased using SNP analysis, or wgMLST analysis compared to PFGE, visualised through the pairwise comparisons of multiple isolates from patient 38 and 17. Every isolate obtained a MLST sequence type 242, compared to PFGE type J, J-5-2, and J-6, showing a higher discriminatory power of PFGE compared to MLST, also supported in literature (43), (44). The divergence within cluster J is clear showing a difference of >500 SNP´s between several isolates.

5.3 B cluster comparison

The PFGE type B cluster includes both type B and B-2C and according to the analysis in Bionumerics the overall similarity is >60%, indicating less of a diversity than the PFGE type J cluster. However when comparing to NGS data, both the SNP

analysis and the cgMLST suggest the opposite, showing ~1600 SNP´s and >800 alleles difference within the B cluster. Further comparisons between duplicate samples (of patient 10, 14, 28, and 33) show great correspondence between all three bioinformatics platforms.

5.4 Medical ward outbreak

In this case, having non-CF patients, with a clear epidemiological relationship, with a new *P. aeruginosa* infection, it is clear that the SNP analysis gives reliable information about disease control. The three different clusters, PFGE types; NF, NM and KD, includes patient samples isolated a month apart and within each cluster the diversity is less than eight SNP´s, indicating a possible patient transmission (45), (46). Alternatively several patients got infected by the same environmental strain of *P. aeruginosa*, previously showed in a hospital outbreak (14), since it is commonly found in water. However, at this medical ward outbreak there were no isolation of environmental samples, e.g. from basin taps, respirators.

5.5 Hypermutators

Several possible hypermutator isolates in the collection made at the Microbiology laboratory with especially one patient, being the only one in the region having PFGE type AN-1, with several hundred SNP´s difference between isolates collected close at time (one year apart) (Appendix IV – Hypermutators). According to literature there is as high prevalence as 36-48% of patients with CF having a hypermutator strain (5), (46). Consequently there is a need for further studies on how to use bioinformatics to detect these hypermutators and to eliminate the risk of missing identical isolates. Therefore we have reached out to Rasmus Marvig, at the Rigshospital in Copenhagen, Denmark, and he has kindly shared two known hypermutant strains, isolated from two CF patients (Data in Appendix IV – Hypermutators).

6. Conclusion

All three bioinformatics platforms showed similar and supportive results to that of the PFGE analysis. Samples having PFGE type J and PFGE type B as well as a longitudinal relation were clustered accordingly when using SNP analysis, cg,- and

wgMLST. Additionally did epidemiologically linked isolates cluster correspondingly to the PFGE analysis, having less than eight SNP´s difference indicating a possible patient transmission (45), (46).
The results support next generation sequencing as a highly discriminative method, that in addition with clinical patient data, works as an epidemiological typing method for *P. aeruginosa.*

However, the abundance of hypermutators (5), (46) within the *P. aeruginosa* population in patients with CF make up a need for further studies on finding additional bioinformatics tools to exclude SNP or allelic changes due to hypermutation and recombination. Additionally would a longitudinal study, including a larger patient group, be beneficiary for determining a SNP or allelic cut off for strain relatedness.

References

1. Chen JW, Lau YY, Krishnan T, Chan KG, Chang CY. Recent Advances in Molecular Diagnosis of *Pseudomonas aeruginosa* Infection by State-of-the-Art Genotyping Techniques. Front Microbiol. 2018;9:1104.
2. Bhagirath AY, Li Y, Somayajula D, Dadashi M, Badr S, Duan K. Cystic fibrosis lung environment and *Pseudomonas aeruginosa* infection. BMC Pulm Med. 2016;16(1):174.
3. Lyczak JB, Cannon CL, Pier GB. Lung infections associated with cystic fibrosis. Clin Microbiol Rev. 2002;15(2):194-222.
4. Macia MD, Blanquer D, Togores B, Sauleda J, Perez JL, Oliver A. Hypermutation is a key factor in development of multiple-antimicrobial resistance in *Pseudomonas aeruginosa* strains causing chronic lung infections. Antimicrob Agents Chemother. 2005;49(8):3382-6.
5. Oliver A, Canton R, Campo P, Baquero F, Blazquez J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science. 2000;288(5469):1251-4.
6. Kidd TJ, Grimwood K, Ramsay KA, Rainey PB, Bell SC. Comparison of three molecular techniques for typing *Pseudomonas aeruginosa* isolates in sputum samples from patients with cystic fibrosis. J Clin Microbiol. 2011;49(1):263-8.
7. Bingen E, Bonacorsi S, Rohrlich P, Duval M, Lhopital S, Brahimi N, et al. Molecular epidemiology provides evidence of genotypic heterogeneity of multidrug-resistant *Pseudomonas aeruginosa* serotype O:12 outbreak isolates from a pediatric hospital. J Clin Microbiol. 1996;34(12):3226-9.
8. Grundmann H, Schneider C, Hartung D, Daschner FD, Pitt TL. Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. J Clin Microbiol. 1995;33(3):528-34.
9. Pfaller MA, Wendt C, Hollis RJ, Wenzel RP, Fritschel SJ, Neubauer JJ, et al. Comparative evaluation of an automated ribotyping system versus pulsed-field gel electrophoresis for epidemiological typing of clinical isolates of *Escherichia coli* and *Pseudomonas aeruginosa* from patients with recurrent gram-negative bacteremia. Diagn Microbiol Infect Dis. 1996;25(1):1-8.
10. van Belkum A, van Leeuwen W, Kaufmann ME, Cookson B, Forey F, Etienne J, et al. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. J Clin Microbiol. 1998;36(6):1653-9.
11. Parkins MD, Somayaji R, Waters VJ. Epidemiology, Biology, and Impact of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. Clin Microbiol Rev. 2018;31(4).
12. Singh A, Goering RV, Simjee S, Foley SL, Zervos MJ. Application of molecular techniques to the study of hospital infection. Clin Microbiol Rev. 2006;19(3):512-30.
13. Krishna B M KMA, and Khan S T. Next-generation sequencing (NGS) platforms: An exciting era of genome sequence analysis. Microbial genomics in sustainable agroecosystems. Singapore: Springer; 2019:89-109.
14. Johansson E, Welinder-Olsson C, Gilljam M. Genotyping of *Pseudomonas aeruginosa* isolates from lung transplant recipients and aquatic environment-detected in-hospital transmission. APMIS. 2014;122(2):85-91.

15. Rice LB. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. J Infect Dis. 2008;197(8):1079-81.
16. "Publishes List of Bacteria for Which New Antibiotics are Urgently Needed". World health organization (WHO). https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed (Accessed 21st of July 2021).
17. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Hoiby N, et al. Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Microbiol. 2012;10(12):841-51.
18. "CFTR". John Hopkins Cystic Fibrosis Center.

https://hopkinscf.org/knowledge/cftr/ (Accessed 29th of May 2021)

19. Lee B, Haagensen JA, Ciofu O, Andersen JB, Hoiby N, Molin S. Heterogeneity of biofilms formed by nonmucoid *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. J Clin Microbiol. 2005;43(10):5247-55.
20. Hoiby N, Ciofu O, Bjarnsholt T. *Pseudomonas aeruginosa* biofilms in cystic fibrosis. Future Microbiol. 2010;5(11):1663-74.
21. In: M LJ, editor. A dictionary of epidemiology. 4th ed. New York, USA: Oxford University Press.
22. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. Infect Control Hosp Epidemiol. 1997;18(6):426-39.
23. Nielsen EM, Engberg J, Fussing V, Petersen L, Brogren CH, On SL. Evaluation of phenotypic and genotypic methods for subtyping *Campylobacter jejuni* isolates from humans, poultry, and cattle. J Clin Microbiol. 2000;38(10):3800-10.
24. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 2007;13 Suppl 3:1-46.
25. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 1984;37(1):67-75.
26. "PulseNet: PFGE". Centers for disease control and prevention (CDC). https://www.cdc.gov/pulsenet/pathogens/pfge.html (Accessed 13th of May 2021)
27. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol. 1995;33(9):2233-9.
28. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
29. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health Microbiology. Pathology. 2015;47(3):199-210.
30. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003;300(5617):286-90.
31. Kulski KJ. "Next-generation sequencing – An overview of the history, tools and “omic” applications" in *Next generation sequencing - Advances, Applications and Challenges*. Jerzy K Kulski, IntechOpen, 2016 DOI: 10.5772/61964. [Online]. Available: https://www.intechopen.com/chapters/49602

32. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348-52.
33. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
34. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341.
35. Lee HC, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D. Bioinformatics tools and databases for analysis of next-generation sequence data. Brief Funct Genomics. 2012;11(1):12-24.
36. Johansson E, Welinder-Olsson C, Gilljam M, Pourcel C, Lindblad A. Genotyping of *Pseudomonas aeruginosa* reveals high diversity, stability over time and good outcome of eradication. J Cyst Fibros. 2015;14(3):353-60.
37. Tang Hallback E, Karami N, Adlerberth I, Cardew S, Ohlen M, Engstrom Jakobsson H, et al. Methicillin-resistant *Staphylococcus argenteus* misidentified as methicillin-resistant *Staphylococcus aureus* emerging in western Sweden. J Med Microbiol. 2018;67(7):968-71.
38. "Reference genome *Pseudomonas aeruginosa* PAO1". NCBI. https://www.ncbi.nlm.nih.gov/genome/?term=Pseudomonas%20aeruginosa%5BOrganism%5D&cmd=DetailsSearch (Accessed 20th of July 2021)
39. "Reference genome methicillin resistant *Staphylococcus aureus* (MRSA) nctc 8325". NCBI. https://www.ncbi.nlm.nih.gov/genome/?term=Staphylococcus%20aureus%5BOrganism%5D&cmd=DetailsSearch (Accessed 20th of July 2021)
40. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.
41. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013;14(5):R51.
42. Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS One. 2013;8(4):e62856.
43. Johnson JK, Arduino SM, Stine OC, Johnson JA, Harris AD. Multilocus sequence typing compared to pulsed-field gel electrophoresis for molecular typing of *Pseudomonas aeruginosa*. J Clin Microbiol. 2007;45(11):3707-12.
44. Maatallah M, Bakhrouf A, Habeeb MA, Turlej-Rogacka A, Iversen A, Pourcel C, et al. Four genotyping schemes for phylogenetic analysis of *Pseudomonas aeruginosa*: comparison of their congruence with multi-locus sequence typing. PLoS One. 2013;8(12):e82069.
45. Snyder LA, Loman NJ, Faraj LA, Levi K, Weinstock G, Boswell TC, et al. Epidemiological investigation of *Pseudomonas aeruginosa* isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. Euro Surveill. 2013;18(42).
46. Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. PLoS Genet. 2013;9(9):e1003741.

Appendix I – Settings in CLC genomics workbench, Qiagen

All settings used in CLC genomics workbench, Qiagen, for the workflows: *Map reads to reference* and *de novo assembly.* All settings are default other than the reference file. When left blank the setting is not used.

Map reads to reference workflow	**Settings**
Trim Reads	
Quality trim	x
Quality limit	0.01
Ambiguous trim	x
Ambiguous limit	2
Trim adapter list	
Automatic read-through adapter trimming	x
Remove 5' terminal nucleotides	
Number of 5' terminal nucleotides	1
Remove 3' terminal nucleotides	
Number of 3' terminal nucleotides	1
Discard short reads	x
Minimum number of nucleotides in reads	15
Discard long reads	
Maximum number of nucleotides in reads	1000
Map Reads to Reference	
References	**GCF_000006765.1_ASM676v1_genomic**
Masking mode	No masking
Masking track	
Match score	1
Mismatch cost	2
Cost of insertions and deletions	Affine gap cost
Insertion cost	3
Deletion cost	3
Insertion open cost	6
Insertion extend cost	1
Deletion open cost	6
Deletion extend cost	1
Length fraction	0.7
Similarity fraction	0.7
Global alignment	
Auto-detect paired distances	x
Non-specific match handling	Map randomly
Local Realignment (2) (Original name: Local Realignment)	
Realign unaligned ends	x
Multi-pass realignment	2
Guidance-variant track	
Maximum guidance-variant length	200

Force realignment to guidance-variants	
InDels and Structural Variants	
P-Value threshold	0.0001
Maximum number of mismatches	3
Minimum quality score	10
Minimum relative consensus coverage	0
Filter variants	
Minimum number of reads	2
Ignore broken pairs	x
Restrict calling to target regions	
Local Realignment	
Realign unaligned ends	x
Multi-pass realignment	2
Guidance-variant track	Defined by: InDels and Structural Variants
Maximum guidance-variant length	200
Force realignment to guidance-variants	
Fixed Ploidy Variant Detection	
Ploidy	1
Required variant probability (%)	80
Ignore positions with coverage above	100000
Restrict calling to target regions	
Ignore broken pairs	x
Ignore non-specific matches	Reads
Minimum read length	20
Minimum coverage	10
Minimum count	10
Minimum frequency (%)	90
Base quality filter	x
Neighborhood radius	5
Minimum central quality	20
Minimum neighborhood quality	15
Read direction filter	
Direction frequency (%)	5
Relative read direction filter	
Significance (%)	1
Read position filter	
Significance (%)	1
Remove pyro-error variants	x
In homopolymer regions with minimum length	4
With frequency below	0.65

De novo assembly work flow	**Settings**
Workflow Input	

Workflow Input	1:2019 (single)
Import Command	
Trim Reads	
Trim using quality scores	x
Quality limit	0,05
Trim ambiguous nucleotides	x
Maximum number of ambiguities	2
Automatic read-through adapter trimming	x
Trim adapter list	
Trim homopolymers from 5'	
Trim homopolymers from 3'	
polyA	
polyC	
polyG	x
polyT	
Remove 5' terminal nucleotides	
Number of 5' terminal nucleotides	1
Remove 3' terminal nucleotides	
Number of 3' terminal nucleotides	1
Trim to a fixed length	
Maximum length	150
Trim end	Trim from 3'-end
Discard short reads	
Minimum length	15
Discard long reads	
Maximum length	1000
De Novo Assembly	
Mapping mode	Map reads back to contigs (slow)
Update contigs	x
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0,5
Similarity fraction	0,8
Alignment mode	local
Match mode	random
Create list of un-mapped reads	
Automatic bubble size	x
Bubble size	50
Automatic word size	x
Word size	20
Minimum contig length	200
Guidance only reads	
Perform scaffolding	x
Auto-detect paired distances	x

Create report	x
Extract Consensus Sequence	
Threshold	0
Action	Insert 'N' ambiguity symbols
Post-remove action	Split into separate sequences
Conflict resolution strategy	Vote
Noise threshold	0,1
Minimum nucleotide count	1
Use quality score	
Add consensus annotations (conflicts, indels, low coverage etc.)	
Keep annotations already on consensus	x
Transfer annotations from reference	

Appendix II – Bioinformatics analysis - overview

Table of all isolates and analyses included in the project. Sorted on PFGE type (39 types and N/A for not typable). Novel meaning not typed to a known sequence type. – meaning that the isolate has not been analysed in Bionumerics (BN).

Isolate ID	*PFGE*	*pubMLST BN*	*cgMLST CLC*	*MLST 1928*
39:2019	AA-3-2c-3-1	ST27	ST27	ST27
6:2019	AA-3-3-4-2-1-2	ST27	ST27	ST27
54:2020	AC-4-3	ST14	ST14	ST14
34:2004	AD	ST2055	ST2055	ST2055
26:2019	AF-2/AL	ST258	ST258	ST258
13:2019	AG-6	ST179	ST179	ST179
32:2019	AH-2	ST649	ST649	ST649
4:2019	AL-	ST1051	ST1051	ST1051
29:2019	AL-2-2-4	ST179	ST179	ST179
11:2011	AN-1	novel	novel	novel
11:1:2019	AN-1	novel	novel	novel
11:2:2019	AN-1	novel	novel	novel
11:3:2019	AN-1	novel	new nuoD	novel
10:2019	B	ST809	ST809	ST809
10:2020	B	ST809	ST809	ST809
14:2006	B	ST809	ST809	ST809
14:2019	B	ST809	ST809	ST809
20:2019	B	ST809	ST809	ST809
24:2019	B	ST809	ST809	ST809
37:2020	B	novel	novel	novel
42:2019	B	ST809	ST809	ST809
48:1:2020	B	novel	ST809	ST809
48:2:2020	B	ST809	ST809	ST809
50:2020	B	ST809	ST809	ST809
52:2020	B	ST809	ST809	ST809
57:2020	B	ST809	ST809	ST809
58:2020	B	ST809	ST809	ST809
71:2021	B	-	ST809	ST809
28:2019	B-2C	ST809	ST809	ST809
28:2020	B-2C	ST809	ST809	ST809
33:1:2019	B-2C	ST809	ST809	ST809
33:2:2019	B-2C	ST809	ST809	ST809
49:2020	B-2C	ST809	ST809	ST809
30:2019	BF-2b	ST1225	ST1225	ST1225
34:2020	CN/EX-2	ST1074	ST1074	ST1074
45:2020	CN/EX-2	ST1074	ST1074	ST1074
21:2019	CN/EX-4-1	ST1074	ST1074	ST1074
22:2019	CN/EX-4-2	ST1074	ST1074	ST1074
60:2020	CO	ST782	ST782	ST782
61:2020	CO	ST782	ST782	ST782

66:2021	DO-2-2b-2	-	ST253	ST253
70:1:2021	FC-3	-	ST1033	ST1033
70:2:2021	FC-3	-	ST1033	ST1033
70:3:2021	FC-3	-	ST1033	ST1033
70:4:2021	FC-3	-	ST1033	ST1033
43:2020	FX-6	ST1248	ST1248	ST1248
36:2020	GA	ST2041	ST2041	ST2041
15:2011	GF	novel	novel	novel
15:2019	GF	novel	novel	novel
15:1:2021	GF	-	novel	novel
15:2:2021	GF	-	novel	novel
41:1:2020	HX	novel	new trpE	novel
41:2:2020	HX	novel	novel	novel
12:2019	J	ST242	ST242	ST242
17:1:2019	J	ST242	ST242	ST242
17:2:2019	J	ST242	ST242	ST242
44:2020	J	ST242	new trpE	ST242
51:2020	J	ST242	ST242	ST242
53:2020	J	ST242	ST242	ST242
62:2021	J	ST242	ST242	ST242
9:2019	J-2	ST242	ST242	ST242
38:2020	J-5-2	ST242	ST242	ST242
37:2019	J-6	ST242	ST242	ST242
38:2017	J-6	ST242	ST242	ST242
38:2019	J-6	ST242	ST242	ST242
59:2020	JK	novel	novel	STnovel
35:2014	JO	ST633	ST633	ST633
35:1:2020	JO-2B	ST633	ST633	ST633
73:2021	JX-2	-	novel	novel
2:1:2019	KD-1	ST155	new trpE	ST155
2:2:2019	KD-1	ST155	ST155	ST155
8:2019	KD-1	ST155	ST155	ST155
27:2019	KH-2	ST385	ST385	ST385
40:2019	L	novel	new trpE	novel
40:2008	L-2-2-4	ST1206	ST1206	ST1206
55:2020	L-5	ST1206	ST1206	ST1206
63:2021	LK-4	ST179	ST179	ST179
64:2021	N/A	ST245	ST245	ST245
65:2021	N/A	ST245	ST245	ST245
5:2019	NF	ST1602	ST1602	ST1602
7:2019	NF	ST1602	ST1602	ST1602
19:1:2019	NH	ST313	ST313	ST313
19:2:2019	NH-2	ST313	ST313	ST313
23:1:2019	NI	ST155	ST155	ST155
23:2:2019	NI-2	ST155	ST155	ST155
25:2019	NJ	ST1320	ST1320	ST1320
3:1:2019	NM	ST179	ST179	ST179
3:2:2019	NM	ST179	ST179	ST179

1:2019	NM-3	ST179	new trpE	ST179
31:2019	NM-3	ST179	ST179	ST179
46:2020	NQ	ST244	ST244	ST244
68:2021	NU	-	new trpE	novel
72:2021	NV	-	new trpE	ST671
67:2021	NX	-	ST348	ST348
69:2021	NY	-	novel	novel
47:2020	Q-2b	novel	new trpE	novel
18:2012	X	ST198	ST198	ST198
18:1:2019	X	ST198	ST198	ST198
18:2:2019	X	ST198	ST198	ST198
16:2019	Z/ED-2b	novel	new guaA	ST683

Appendix III – Bioinformatics analysis results

Table III.1 contains the number of SNP´s difference between each and one of the isolates belonging to cluster J.

Table III.1. Results from SNP analysis visualised as a matrix. The blue square highlighting the difference in SNP´s between patient 37 and 38, a maximum of 18 SNP´s in between patients and 15 SNP´s within patient 38.

Cluster J	Isolate	1	2	3	4	5	6	7	8	9	10	11	12
53:2020	1	0	88	513	434	519	526	527	523	413	360	358	389
44:2020	2	88	0	425	346	431	438	439	435	325	272	270	301
12:2019	3	513	425	0	347	432	439	440	436	326	271	271	300
9:2019	4	434	346	347	0	225	232	233	229	137	108	116	139
38:2017	5	519	431	432	225	0	7	8	10	222	193	201	224
38:2019	6	526	438	439	232	7	0	15	17	229	200	208	231
38:2020	7	527	439	440	233	8	15	0	18	230	201	209	232
37:2019	8	523	435	436	229	10	17	18	0	226	197	205	228
62:2021	9	413	325	326	137	222	229	230	226	0	85	95	118
51:2020	10	360	272	271	108	193	200	201	197	85	0	42	37
17:1:2019	11	358	270	271	116	201	208	209	205	95	42	0	73
17:2:2019	12	389	301	300	139	224	231	232	228	118	37	73	0

Table III.2 contains the number of SNP´s difference between each and one of the isolates belonging to cluster B.

Table III.2. Results from SNP analysis visualised as a matrix. The blue squares highlighting the difference in SNP´s between the five double isolates from patient 10, 14, 28, 33 and 48. The lowest difference of one SNP´s for patient 33 and the highest difference of 331 SNP´s for patient 48. The yellow square highlighting two patient isolates only 23 SNP´s apart.

Cluster B	Isolate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
57:2020	1	0	85	68	170	665	699	154	177	159	506	320	145	144	162	164	374	163	610	1322
14:2019	2	85	0	45	147	642	676	131	154	136	483	297	122	121	139	141	351	140	587	1299
14:2006	3	68	45	0	102	597	631	88	111	91	438	254	77	76	94	96	308	95	542	1256
50:2020	4	170	147	102	0	659	693	154	177	159	504	320	143	142	162	164	374	161	608	1322
28:2019	5	665	642	597	659	0	40	651	674	650	995	817	640	639	657	659	867	658	1103	1819
28:2020	6	699	676	631	693	40	0	685	708	684	1029	851	674	673	691	693	905	692	1137	1851
52:2020	7	154	131	88	154	651	685	0	23	125	472	300	123	122	140	142	354	141	588	1300
42:2019	8	177	154	111	177	674	708	23	0	148	495	323	146	145	163	165	377	164	611	1323
24:2019	9	159	136	91	159	650	684	125	148	0	475	303	126	125	143	145	357	144	591	1303
20:2019	10	506	483	438	504	995	1029	472	495	475	0	648	471	470	490	492	702	489	934	1650
49:2020	11	320	297	254	320	817	851	300	323	303	648	0	263	262	284	286	498	287	734	1448
33:1:2019	12	145	122	77	143	640	674	123	146	126	471	263	0	1	107	109	323	110	555	1265
33:2:2019	13	144	121	76	142	639	673	122	145	125	470	262	1	0	106	108	322	109	554	1264
10:2019	14	162	139	94	162	657	691	140	163	143	490	284	107	106	0	2	342	127	576	1290
10:2020	15	164	141	96	164	659	693	142	165	145	492	286	109	108	2	0	344	129	578	1291
48:1:2020	16	374	351	308	374	867	905	354	377	357	702	498	323	322	342	344	0	331	788	1502
48:2:2020	17	163	140	95	161	658	692	141	164	144	489	287	110	109	127	129	331	0	575	1289
58:2020	18	610	587	542	608	1103	1137	588	611	591	934	734	555	554	576	578	788	575	0	1608
37:2020	19	1322	1299	1256	1322	1819	1851	1300	1323	1303	1650	1448	1265	1264	1290	1291	1502	1289	1608	0

The matrix for the wgMLST analysis in Bionumerics, for the PFGE cluster B, figure III.1.

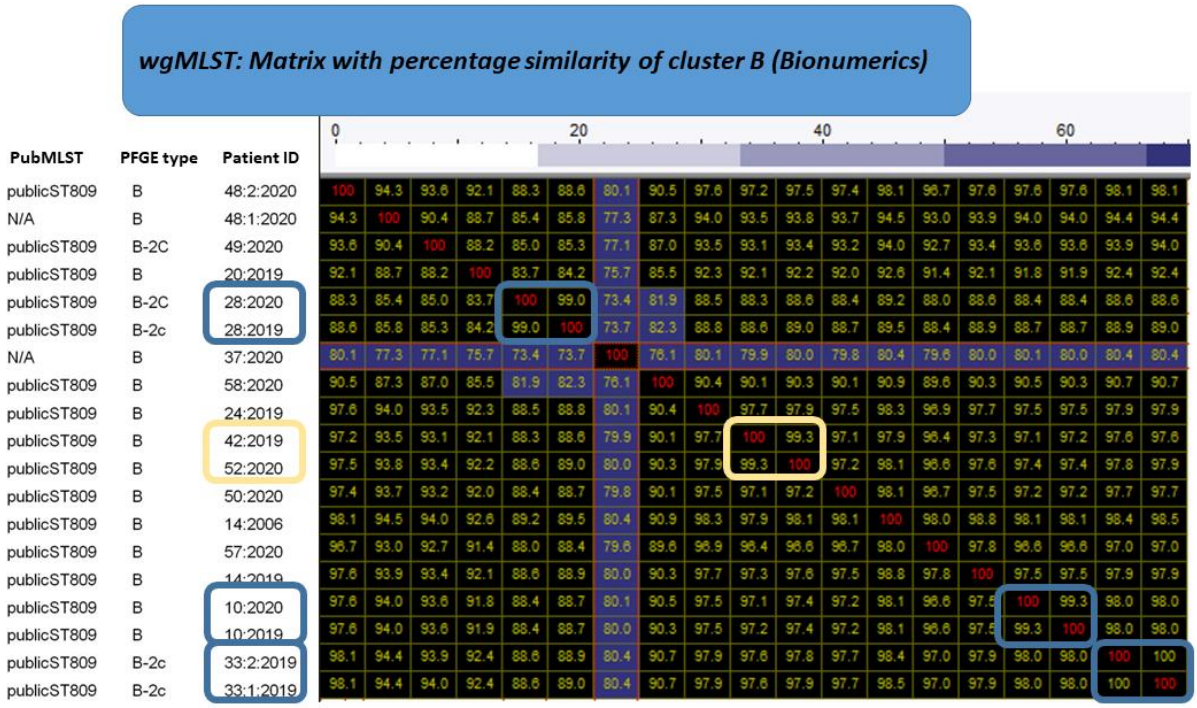

Figure III.1. A matrix of the whole genome MLST analysis in Bionumerics. Visualized is the PFGE cluster B. In the left: pubMLST type, cluster type and patient ID, in the right: the matrix giving percentage similarity. The blue boxes corresponds different isolates from the same patient (99.0-100.0 %), and the yellow box corresponds to isolates from two different patients with a similarity of 99.3 %.

Table III.3 is visualizing the SNP´s difference between isolates included in a medical ward outbreak. There are three distinct clusters highlighted in blue.

Table III.3. SNP matrix, in CLC genomics workbench, for isolates collected from an outbreak at a medical ward in Gothenburg. Blue marked SNP´s visualising patient samples differing less than 8 SNP`s.

Vårdavd.	Isolat	1	2	3	4	5	6	7	8	9	10	11	12
2:1:2019	1	0	3	3	5007	5007	5005	5005	32948	16952	16953	16045	15870
2:2:2019	2	3	0	0	5004	5006	5002	5002	32946	16950	16951	16045	15870
8:2019	3	3	0	0	5004	5006	5002	5002	32946	16950	16951	16045	15870
1:2019	4	5007	5004	5004	0	8	6	6	32805	17061	17062	16276	16163
31:2019	5	5007	5006	5006	8	0	8	8	32804	17061	17062	16274	16161
3:1:2019	6	5005	5002	5002	6	8	0	0	32803	17059	17060	16274	16161
3:2:2019	7	5005	5002	5002	6	8	0	0	32803	17059	17060	16274	16161
4:2019	8	32948	32946	32946	32805	32804	32803	32803	0	33831	33832	32738	33328
5:2019	9	16952	16950	16950	17061	17061	17059	17059	33831	0	1	16531	16555
7:2019	10	16953	16951	16951	17062	17062	17060	17060	33832	1	0	16532	16556
6:2019	11	16045	16045	16045	16276	16274	16274	16274	32738	16531	16532	0	15718
9:2019	12	15870	15870	15870	16163	16161	16161	16161	33328	16555	16556	15718	0

Appendix IV – Hypermutators

In figure IV.1 and table IV.1 are the patient having PFGE type AN-1.

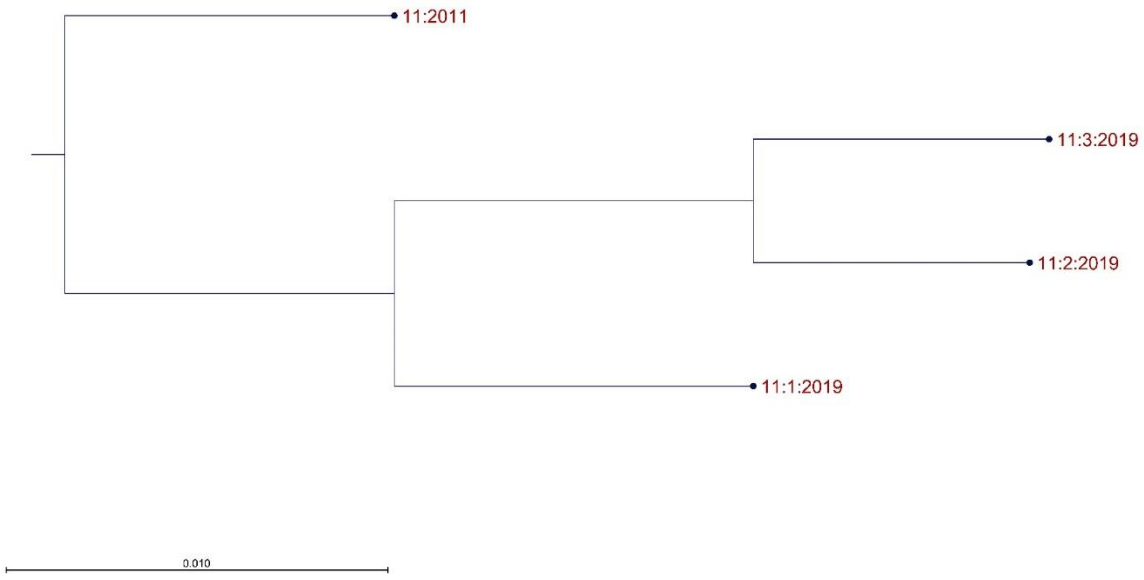

Figure IV.1. SNP tree from CLC genomics workbench, visualising the four isolates derived from one patient being the only one having P. aeruginosa with PFGE type AN-1.

Table IV.1. A matrix, CLC, giving the number of SNP´s difference between the isolates.

AN-1					
Isolat		**1**	**2**	**3**	**4**
11:2011	**1**	0	696	705	752
11:2:2019	**2**	696	0	369	626
11:3:2019	**3**	705	369	0	637
11:1:2019	**4**	752	626	637	0

The SNP analyses in CLC, for the Danish patients, DK01 and DK32, are showed in figure IV.2-3, and table IV.2-3.

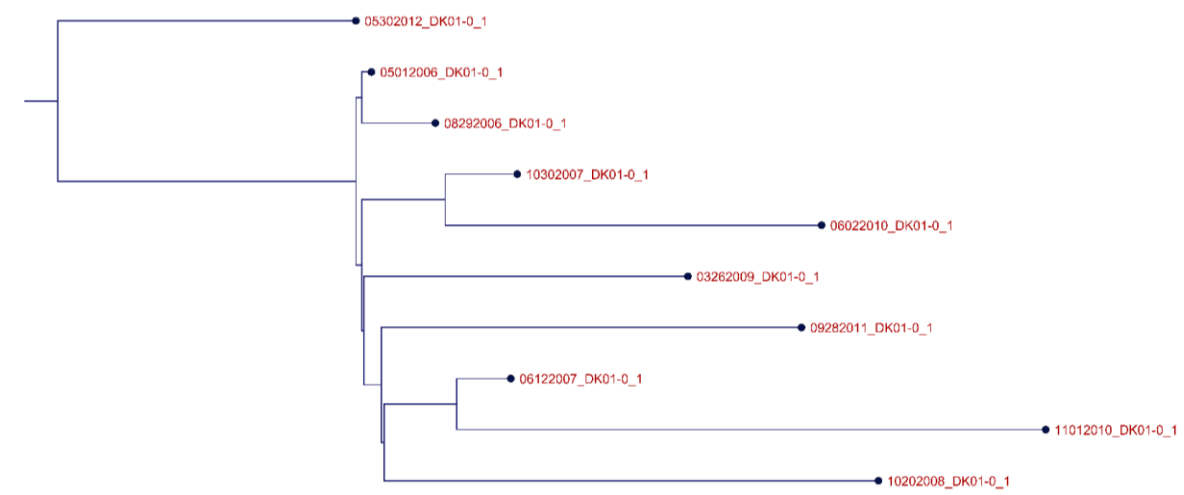

Figure IV.2. SNP analysis, CLC, showing patient from Denmark (DK01). This is a known hypermutator strain.

Table IV.2 SNP matrix, CLC, showing patient from Denmark (DK01). This is a known hypermutator strain.

DK01	Sample number	1	2	3	4	5	6	7	8	9	10
03262009_DK01-0_1	**1**	0	331	879	745	448	392	722	793	458	953
05012006_DK01-0_1	**2**	331	0	576	454	159	79	437	508	165	668
05302012_DK01-0_1	**3**	879	576	0	1012	719	635	991	1068	725	1224
06022010_DK01-0_1	**4**	745	454	1012	0	575	517	851	924	425	1084
06122007_DK01-0_1	**5**	448	159	719	575	0	222	520	589	286	609
08292006_DK01-0_1	**6**	392	79	635	517	222	0	498	571	228	729
09282011_DK01-0_1	**7**	722	437	991	851	520	498	0	869	564	1027
10202008_DK01-0_1	**8**	793	508	1068	924	589	571	869	0	635	1094
10302007_DK01-0_1	**9**	458	165	725	425	286	228	564	635	0	795
11012010_DK01-0_1	**10**	953	668	1224	1084	609	729	1027	1094	795	0

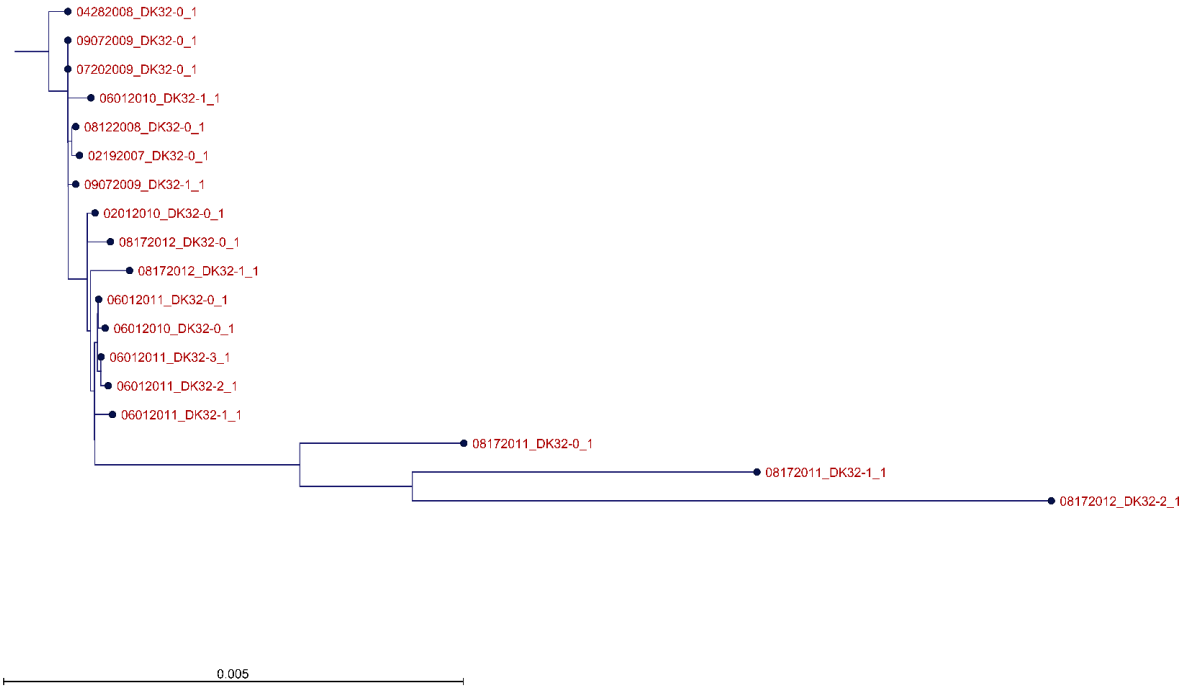

Figure IV.3. SNP analysis, CLC, showing patient from Denmark (DK32). This is a known hypermutator strain.

Table IV.3. SNP matrix, CLC, showing patient from Denmark (DK32). This is a known hypermutator strain.

DK32	Sample number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
02012010_DK32-0_1	1	0	10	17	7	13	5	9	6	6	7	9	100	177	8	13	253	7	9
02192007_DK32-0_1	2	10	0	13	13	9	11	15	12	10	3	3	106	181	14	19	259	3	5
04282008_DK32-0_1	3	17	13	0	20	16	18	22	21	19	10	12	113	190	21	26	266	10	12
06012010_DK32-0_1	4	7	13	20	0	16	2	6	5	3	10	12	99	176	11	14	252	10	10
06012010_DK32-1_1	5	13	9	16	16	0	14	16	17	15	6	8	109	184	17	22	262	6	8
06012011_DK32-0_1	6	5	11	18	2	14	0	6	3	1	8	10	97	174	9	12	250	8	10
06012011_DK32-1_1	7	9	15	22	6	16	6	0	9	7	12	14	101	176	13	16	254	12	12
06012011_DK32-2_1	8	6	12	21	5	17	3	9	0	2	11	11	100	175	12	15	253	11	13

06012011_DK32-3_1	**9**	6	10	19	3	15	1	7	2	0	9	9	98	173	10	13	251	9	11
07202009_DK32-0_1	**10**	7	3	10	10	6	8	12	11	9	0	2	103	180	11	16	256	0	2
08122008_DK32-0_1	**11**	9	3	12	12	8	10	14	11	9	2	0	105	180	13	18	258	2	4
08172011_DK32-0_1	**12**	100	106	113	99	109	97	101	100	98	103	105	0	163	104	107	237	103	105
08172011_DK32-1_1	**13**	177	181	190	176	184	174	176	175	173	180	180	163	0	181	184	256	180	182
08172012_DK32-0_1	**14**	8	14	21	11	17	9	13	12	10	11	13	104	181	0	17	257	11	13
08172012_DK32-1_1	**15**	13	19	26	14	22	12	16	15	13	16	18	107	184	17	0	260	16	18
08172012_DK32-2_1	**16**	253	259	266	252	262	250	254	253	251	256	258	237	256	257	260	0	256	258
09072009_DK32-0_1	**17**	7	3	10	10	6	8	12	11	9	0	2	103	180	11	16	256	0	2
09072009_DK32-1_1	**18**	9	5	12	10	8	10	12	13	11	2	4	105	182	13	18	258	2	0

DEPARTMENT OF BIOLOGY AND
BIOLOGICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se