Marie Duus Dahl

# Construction of vectors for a CRISPR-Cas9 mutagenesis system for *Azotobacter vinelandii*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Helga Ertesvåg
June 2019

**Master's thesis**

**NTNU**
Norwegian University of
Science and Technology

Marie Duus Dahl

# Construction of vectors for a CRISPR-Cas9 mutagenesis system for *Azotobacter vinelandii*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Helga Ertesvåg
June 2019

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

**NTNU**
Norwegian University of
Science and Technology

# Preface

This thesis has been part of the molecular biology group, with Professor Helga Ertesvåg as supervisor, at the Department of Biotechnology and Food Science at the Norwegian University of Science and Technology. The project is a continuation of the specialization project (TBT4500) conducted during the autumn of 2018.

I would like to thank my supervisor Prof. Helga Ertesvåg, for being there with the door to her office open when help and guidance were needed.

I would also like to thank my friends, who I can always count on to be there whether I need a hug or a good laugh. And finally, I would like to thank my family who have supported me and cheered me on when things have not always gone exactly as planned.

Trondheim, June 2019

Marie Duus Dahl

# Summary

*Azotobacter vinelandii* is an alginate producing bacterium of some industrial interest. Alginate is a biopolymer used in many industrial fields because of its gel-forming ability. Bacterial alginate has a different composition from algal alginate, and the production can be regulated and controlled more easily in bacteria. Homologous recombination is used in *A. vinelandii* as a mutagenesis tool to produce alginate with wanted properties. The challenge lies in the bacterium having many copies of its chromosome. The result of homologous recombination is a mixture of wild-type and mutated chromosomes, and a lengthy process to obtain the desired mutants is necessary. CRISPR-Cas9 is a relatively new mutagenesis system that is known to mutagenize several chromosomes at the same time. During this work, vectors for a CRISPR-Cas9 mutagenesis system have been developed for *A. vinelandii*, to look into whether this system can give higher mutation frequencies than homologous recombination, or if it can be used to remove unwanted chromosomes.

CRISPR-Cas9 consists of an endonuclease, Cas9, and a guide RNA that binds to Cas9 and guides it to the target sequence. Binding of Cas9 to the target result in cleavage of the sequence, which can be repaired through homology-directed repair. A cloning plan was developed to construct two plasmids with the necessary CRISPR-Cas9 components - one with the gene encoding Cas9 and a second with a gene encoding the gRNA. The system was designed to test the lethality and the mutagenesis frequency of the system. Plasmid pMDD8 was the final vector for Cas9, and pMDD13, pMDD15, pMDD16, and pMDD17 were four variants of the gRNA vector. *uidA* and *mucA* were chosen as target genes for blue-white plate screening and to yield an alginate overproducing strain, respectively.

The two-plasmid CRISPR-Cas9 mutagenesis system designed in this work could not be tested since the gRNA-encoding vectors would not be conjugated into *A. vinelandii*. There can therefore not be concluded if CRISPR will be an improvement to homologous recombination as a mutagenesis tool in the bacterium. Possible reasons for why the system does not work have been discussed, and suggestions for an altered system based on the vectors designed in this study.

# Sammendrag

*Azotobacter vinelandii* er en alginatproduserende bakterie med en viss industriell inter-
esse. Alginat er en biopolymer benyttet i mange industrielle felt på grunn av dens gel-
dannende egenskap. Bakteriell alginat har en annen sammensetning enn alginat fra alger,
og produksjonen reguleres og kontrolleres enklere i bakterier. Homolog rekombinering er
benyttet i *A. vinelandii* som et muteringsverktøy for å produsere alginat med ønskede
egenskaper. Utfordringen er at bakterien har mange kopier av kromosomet sitt. Resul-
tatet av homolog rekombinering er en blanding av vildtype og muterte kromosom, og det
er nødvendig med en langvarig prosess for å oppnå de ønskede mutantene. CRISPR-Cas9
er et relativt nytt muteringssystem som kan mutere flere kromosom samtidig. I løpet
av dette arbeidet har vektorer for et CRISPR-Cas9 mutetingssystem blitt utviklet for
*A. vinelandii*, for å se om dette systemet kan gi høyere mutasjonsfrekvens enn homolog
rekombinering, eller om det kan benyttes for å fjerne uønskede kromosom.

CRISPR-Cas9 består av en endonuclease, Cas9, og et guide RNA som bindes til Cas9 og
leder det til målsekvensen. Målsekvensen vil bli kuttet når Cas9 binder seg. Kuttet kan
bli reparert ved homolog rekombinering. En kloningsplan ble utviklet for å konstruere
to plasmid med de nødvendige CRISPR-Cas9 komponentene - et med genet som koder
for Cas9 og et annet med et gen som koder for gRNAet. Systemet ble designet for å
teste dødeligheten og mutasjonsfrekvensen til systemet. Plasmid pMDD8 er den endelige
vektoren for Cas9, og pMDD13, pMDD15, pMDD16, og pMDD17 er fire varianter av
gRNA vektoren. *uidA* og *mucA* ble valgt som målgen for blå-hvit fargetest og for å gi en
alginat overproduserende stamme, respektivt.

To-plasmid CRISPR-Cas9 muteringssystemet designet i denne oppgaven ble ikke testet
siden de gRNA-kodende vektorene ikke kunne bli konjugert inn i *A. vinelandii*. Det kan
derfor ikke konkluderes om CRISPR vil være en forbedring for homolog rekombinering
som et muteringssystem i bakterien. Mulige årsaker for hvorfor systemet ikke virker
har blitt diskutert, og forslag for et endret system basert på vektorene designet i dette
arbeidet.

# List of Abbreviations

**(s)gRNA** (single) guide RNA

**Amp** Ampicillin

**Apr** Apramycin

**bp** basepair

**Cas proteins** CRISPR associated proteins

**Cas9** RNA-guided DNA endonuclease

**CRISPR** Clustered Regularly Interspaced Short Palindromic Repeats

**crRNA** CRISPR RNA

**DNA** deoxiribonucleic acid

**Kan** Kanamycin

**nt** nucleotide

**PCR** polymerase chain reaction

**RNA** ribonucleic acid

**SLIC** Sequence and Ligation Independent Cloning

**Smr** Streptinomycin

**Spc** Spectinomycin

**Tet**  Tetracyclin

**tracrRNA**  trans activating RNA

**X-Gluc**  reagent for $\beta$-glucuronidase (product of *uidA*)

# Contents

# 1 Introduction

## 1.1 *Azotobacter vinelandii*

*A. vinelandii* is a gram-negative soil bacterium and a member of the Gammaproteobacteria class, related to the genus *Pseudomonas*[1]. It is a bacterium of some industrial interest, because of its alginate production [2]. Alginate is a polymer used in many different industries. Regulating and controlling the bacterium's alginate production is therefore of great importance.

Alginate is a linear heteropolymer of 1→4-linked $\beta$-D-mannuronic acid and $\alpha$-L-guluronic acid, where some of the $\beta$-D-mannuronic acids are acetylated[3]. Figure 1.1, depicts the synthesis pathway of alginate in *A. vinelandii* from different carbon sources (sucrose, glucose or fructose). The figure shows the major steps in the synthesis, along with the enzymes involved in the reactions.
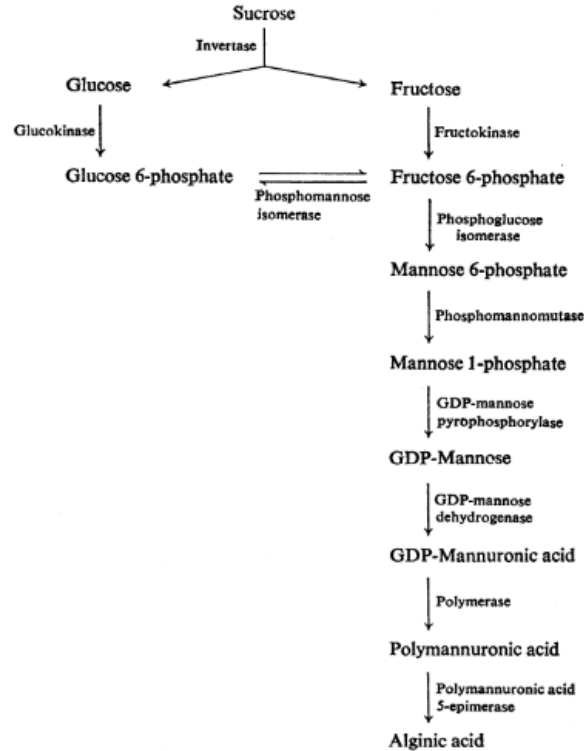
**Figure 1.1: The biosynthetic pathway of alginate synthesis in *A. vinelandii*** (Figure copied from Pindar *et al.* (1975) [4]). The bacteria can use fructose 6-phosphate as a precursor for the synthesis of alginate. The action of different enzymes converts the carbon source (sucrose, glucose or fructose) to fructose 6-phosphate. Fructose 6-phosphate is converted to GDP-mannuronic acid in four steps, through the action of phosphoglucose isomerase (AlgA), phosphomannomutase (AlgC), GDP-mannose pyrospharylase (AlgA), and GDP-mannose dehydrogenase (AlgD). GDP-mannuronic acid is a direct precursor for alginate.

The genes needed for alginate synthesis in *A. vinelandii* are arranged in three operons. One of these transcribes the gene *algD*. *algD* encodes the enzyme GDP-mannose dehydrogenase, which converts GDP-mannose to the alginate precursor GDP-mannuronic acid [5]. AlgU, a $\sigma^E$-like protein, regulates the expression from *algD* through one of three promoters (p2)[6]. *algU* is located in a gene cluster along with *mucABCD*, located downstream from *algU* [7],[8]. The transcription of AlgU is regulated by two promoters; p1, a $\sigma^E$ promoter, and p2, a vegetative $\sigma^D$ promoter. AlgU autoregulates its expression as a $\sigma^E$-factor through the p1 promoter. MucA and MucB play a negative role in the alginate production [9]. MucA is a transmembrane protein in the inner membrane that interacts with MucB. MucB resides in the periplasm between the outer and the inner membrane, to bind and sequester AlgU to the inner membrane [10]. MucB binds to MucA at its periplasmic domain and protects it from degradation by proteolysis [11]. Thereby MucA

is an anti-sigma factor that works as an antagonist to AlgU, by binding to it and inhibiting its interaction with RNA polymerase. Accordingly, MucA represses the expression of AlgU and AlgD, since both expressions are dependent on AlgU. Mutations in *mucA* will yield a mucoid $Alg^+$ strain, because a lack of regulation of AlgU activity will lead to the overproduction of the alginate precursor, AlgD.

*A.vinelandii* also contains seven genes encoding secreted mannuronan C-5 epimerases (*AlgE1-7*), which work to modify the alginate outside the cell [12]. Mannuronan C-5 epimerase changes the conformation of the monomer $\beta$-D-mannuronic acid (M), to $\alpha$-L-guluronic acid (G). By changing the conformation of the monomer, the geometry of the glycosidic linkage in alginate also changes. The linkage between two M residues is equatorial, while the linkage between two G residues is diaxial. The diaxial linkages between two G residues creates cavities or junction zones between the monomers along the chain [13]. Two alginate molecules with consecutive $\alpha$-L-guluronic acid monomers can interact with divalent cations packed between them [14]. This occurs through ionic interactions between the divalent cation and several carboxyl oxygen molecules on two neighboring G residues on each strand. The ability of alginate to form gels makes it a polymer suited for use as stabilizers, thickeners, and gel- and film-forming agents in various industrial fields such as the food and pharmaceutical industry[15]. Brown algae are the main source of commercial alginate. Only two genera of bacteria, *Pseudamonas* and *Azotobacter*, are alginate producing [11]. Bacterial alginate differs in composition from algal alginate, which is heterogeneous in M to G ratio and molecular weight, as a result of variations in the environment and seasons [16]. Since alginate is extracted from algae harvested from the ocean where the growth conditions are varying, it is not possible to control the synthesis of alginate. Therefore, the bacterial production of alginate is of great interest since the growth conditions can be controlled. By altering the conditions, the composition and properties of alginate such as G/M ratio, molecular weight, and viscosity can change. Bacterial alginate is also O-acetylated. An increasing degree of O-acetylation results in increased viscosity[17]. The ability of bacteria to produce alginate with different compositions makes it more suitable in the biomedical industry than algal alginate, since this industry has particular demands.

*A. vinelandii* has a chromosome with a size of around 4,700 kb [18]. It is a highly polyploid bacteria meaning that it has many copies of its chromosome. Maldonado *et al.*[18] showed that the DNA content of *A. vinelandii* increased during the growth cycle using flow cytometry. The ploidy of the cell increased from about four copies to more than 40 in the late exponential phase, more than 80 in the early stationary phase to about 100 in the late stationary phase. The increasing copy number of the chromosome during growth is a problem since the number of targets for homologous recombination also increases, while the mutation rate will stay the same. It is therefore difficult to select for bacteria with only mutated chromosomes, and instead, the result is a mixture of wild-type and mutant chromosomes [19].

## 1.2   Homologous recombination

Homologous recombination is a mechanism that all living organisms use to repair damaged DNA [20]. Genetic engineering utilizes this natural process, which involves the exchange of homologous genetic material, as a mutagenesis tool. The bacteria can recognize a DNA strand with homologous flanking sequences upstream and downstream of the target gene introduced into the bacteria, as similar. It can thus exchange the target with a similar DNA strand. First, the two homologous double-stranded DNA molecules are paired. DNA nucleases such as the RecBCD complex of *Escherichia coli* initiate single-stranded breaks [21]. The helicase activity of the RecBCD complex (subunits RecB and RecD), unwinds the complementary strands in the region next to the single-stranded cuts [22]. RecB unwinds the strand in a 3'-to-5' direction and RecD in the 5'-to-3' direction [23]. RecB degrades both strands as they are unwound. Binding of the subunit RecC at a site called chi, alters the activity of the enzyme preventing the degradation of the 3' strand, but increasing the degradation of the 5' strand. The process creates a 3' single DNA strand, which is bound by the single-stranded binding protein, RecA. The single-stranded DNA-protein complex searches for a homologous DNA molecule and invades it[21]. The single strand is then base paired with the complementary strand of the homologous DNA molecule, which can serve as a template for elongation of the invading strand. Endonucleases separate the complex, and the new recombinant DNA molecules are covalently joined by DNA ligase.

The $\lambda$-Red recombineering system is a RecA independent system from the $\lambda$ bacterio-phage [24]. The mechanism is based on the activity of the exonuclease Exo to create single-stranded overhangs. Another enzyme Beta can anneal two single-stranded homol-ogous overhangs[25] as showed in Figure 1.2. Exo degrades one of the strands of a linear DNA molecule from the 5' end, and so creates 3' single-stranded overhangs that Beta can bind to and protect from degradation [26]. Beta can then anneal two homologous strands to form a recombinant DNA molecule [27]. A third protein of the $\lambda$-Red system binds and inhibits the RecBCD nuclease from degrading linear double-stranded DNA [28].
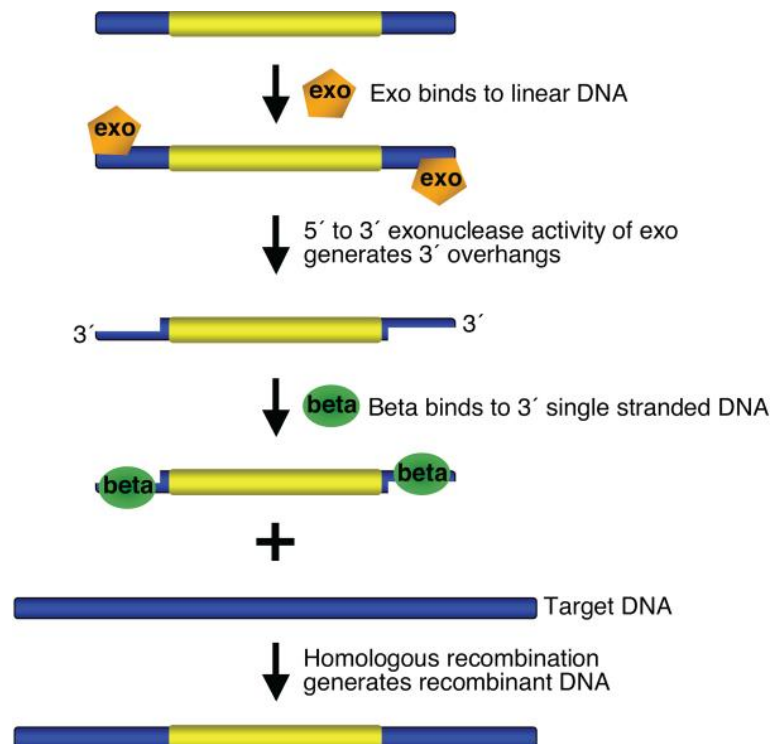


**Figure 1.2: The mechanism of Exo and Beta of the $\lambda$-Red recombineering system** (Figure copied from Sharat *et al.* (2009) [29]). Exo generates 3' single-stranded overhangs by its 5' exonuclease activity, which Beta can bind to and anneal with a homologous single strand to create recombinant DNA.

Homologous recombination in *A. vinelandii*, chromosome editing, occurs through the pro-cess of a double-crossover. Two homologous sequences are then exchanged, or swapped, as compared to homologous recombination through the action of the RecBCD complex, where both strands will get the same sequence [30]. Chromosome editing in *A. vinelandii* requires several selection steps due to the ploidy of the bacterium. A plasmid containing a mutation template must have an antibiotic selective marker and a counter-selective marker. The bacterium must be forced to integrate the conjugated plasmid into its

chromosome. The integration of the plasmid can occur by using a plasmid, where the replication of the plasmid is dependent on the presence of an inducer. If the inducer is absent from the medium, the plasmid cannot replicate independently of the host but requires the host's machinery to replicate. The bacterium is dependent on the plasmid for survival when antibiotics are added to the medium. The bacterium is thus forced to implement the plasmid into its chromosomes. The high ploidy of *A. vinelandii* requires several selection steps for the plasmid to be integrated into all the chromosomes of the bacterium. It is usual to use five cultures inoculated for about three days, each in medium without the inducer and with the antibiotic. The bacterium can be cultivated a sixth time without the antibiotic, to remove non-integrated plasmids and to only select for mutagenized chromosomes. Chromosomes with the entire plasmid integrated are undesirable. These can be selected against by the use of a second selection marker. The expression of the gene *sacB* is lethal in the presence of 5% sucrose. The gene encodes the enzyme levansucrase, which catalyzes the hydrolysis of sucrose to form the high molecular weight fructose polymer levan [31]. The mechanism of toxicity is unclear. It might be a consequence of the accumulation of levan resulting in a burden on the periplasm, or because fructose residues are transferred to unsuitable acceptor molecules yielding toxic effects [31]. If the plasmid is constructed to carry *sacB*, chromosomes with the entire vector would have the gene *sacB* and therefore cannot grow on sucrose. Lastly, antibiotic resistance is tested for colonies that grow on sucrose. Colonies that can grow on sucrose and are sensitive to the antibiotic will only have the mutated chromosomes and not the entire plasmid integrated into its chromosomes.

## 1.3   CRISPR - an adaptive immune system in bacteria

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) together with the CRISPR associated (Cas) enzymes work as an RNA-based adaptive immune system in many bacteria and archaea, by storing a copy of foreign invading phage DNA and plasmids[32]. Figure 1.3 shows how the bacterium cuts the foreign DNA and copies it into its CRISPR DNA sequence. The copy is later used to destroy the DNA of the invader upon re-exposure, through the action of an endonuclease guided by the transcript of the stored DNA copy.
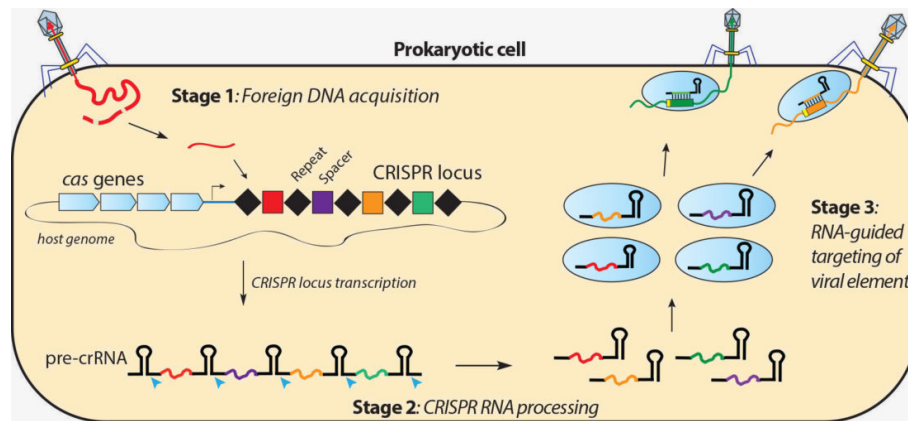
**Figure 1.3: The three stages of the CRISPR mechanism** (Figure copied from Doudna Lab (doudnalab.org)[33]). The foreign phage or plasmid DNA is cleaved and inserted into the CRISPR loci as a memory of foreign invasions. Upon re-exposure, the bacterium transcribes a copy of the invading DNA along with the CRISPR genes. The transcript guides Cas9 to the homologous target sequence, which Cas9 binds and cleaves and thereby eliminating the threat.

There are five types of cited CRISPR systems - type I to V where type II is the most used as a mutagenesis tool [34]. Figure 1.3 shows the three stages of the CRISPR type II system from *S. pyogenes*. In the first stage, a DNA endonuclease recognizes the DNA from an invader as exogenous and snip it into 30 base pair long sequences [35]. The bacterium inserts the sequences into the CRISPR array, which creates a chimeric sequence. The chimeric sequence consists of repeat sequences (black diamond shaped squares) interspaced by foreign DNA (colored squares) called spacer regions [36]. The CRISPR array serves as a memory of past invasions, and as a mean for immunity on subsequent exposure. The CRISPR locus is transcribed and processed by different CRISPR associated (Cas) proteins such as polymerases, nucleases, and helicases[37]. A long primary RNA molecule transcribed from the CRISPR locus, pre-CRISPR RNA (pre-crRNA), is a transcript of repeat (black) and spacer regions (colored) [38]. The pre-crRNA is processed in the second stage by the host's RNaseIII, with the help of a second RNA molecule (not shown in Figure 1.3). This RNA is called the trans-activating CRISPR RNA (tracrRNA), and also contains the direct repeats so that it can bind to the pre-crRNA through homology [39]. The pre-crRNA:tracrRNA duplex is then processed by the host's RNase III into mature crRNA, by cleaving the dsRNA[39]. The mature crRNA contains one spacer and one repeat region [38]. The spacer region provides specificity for different exogenous target sequences because of complementarity with the target. In the third stage, the crRNA:tracrRNA duplex binds to the DNA endonuclease Cas9 via the tracrRNA. The

crRNA guides Cas9 to the target sequence through sequence homology provided by the spacer, and bind to the target through base-pairing to the complementary strand[40]. A Protospacer Adjacent Motif (PAM) downstream (3') of the target sequence is a short (3-5 nucleotide) sequence necessary for nuclease binding [41]. The PAM sequence for Cas9 is NGG, where the N can be any nucleotide [42]. In the absence of the PAM sequence, the nuclease is not able to bind to the target sequence. The absence of the PAM sequence in the spacer region of the endogenous CRISPR array, prevents the nuclease from cleaving and destroying its DNA.

The Cas9 endonuclease from *S. pyogenes* has two nuclease domains, which together cause a double-stranded break (DSB) in the target DNA sequence [40]. One domain is located at the N-terminal and is a RuvC-like nuclease and the second is an HNH-like domain around the center of the protein. After binding at the PAM site, Cas9 undergoes a conformational change which positions the nuclease domains to bind opposite strands [43]. The result is a DSB about three nucleotides upstream of the PAM [44]. The RuvC domain cleaves the strand non-complementary to the spacer in the crRNA, while the HNH domain is responsible for the break in the complementary strand[40].

The endonuclease Cpf1 (CRISPR from *Prevotella* and *Francisellais*) is a Cas9 homolog of the CRISPR type V system, found in various species [34]. Cpf1 only needs one RNA molecule compared to the two needed by Cas9 (tracrRNA and crRNA) [45]. The Cpf1 RNA is about 42 nucleotides long, and hence, much shorter than the long fusion RNA (about 100 nucleotide) needed by Cas9[45]. Cpf1 only has the RuvC-like endonuclease domain of Cas9. Cpf1 is accordingly smaller than Cas9, and is encoded by a smaller gene. The difference in the size of the gene is about 200 nucleotides with *cpf1* of about 3.9 kb and *cas9* of about 4.1 kb [46]. The smaller size of the gene and the shorter guide RNA are advantages of the CRISPR-Cpf1 system, compared to the original type II system, when used as a mutagenesis tool. As a consequence, the system requires vectors with lower capacity, which are easier to construct and deliver. Another advantage of the system is that Cpf1 recognizes a TTN PAM site. The NGG PAM site of Cas9 can be limited if the target gene is an AT-rich region that lacks NGG. In this case, Cpf1 is a better option.

## 1.4   CRISPR as a mutagenesis tool

The use of RNA to specify gene editing in CRISPR makes the system less expensive, more precise, and faster than many other genome engineering systems [35]. CRISPR can also target many loci at the same time. Researchers have developed several variations of the system. Examples are a Cas9 nickase that only cleaves one strand [47], and a dead Cas9 (dCas9) that cannot cleave the target [48]. dCas9 was engineered by mutating the two nuclease domains of Cas9, which yielded a catalytically inactive Cas9. dCas9 can be used to silence a gene by targeting it to the transcriptional repressor domain, so RNA polymerase binding is blocked. Cpf1 of type V CRISPR is another option when using CRISPR as a mutagenesis tool. It has several benefits compared to the type II system since both the gene encoding the endonuclease and the guide RNA is smaller in type V.

CRISPR-Cas9 as a genome engineering tool only requires a single guide RNA (sgRNA or gRNA). The gRNA is a fusion of the tracrRNA and the crRNA. It contains both a nuclease binding domain (a scaffold) and a spacer region of around 20 nucleotides (N20), which provides homology with the target sequence. The only change necessary to alter the target sequence is to exchange the specific N20 sequence of the gRNA.

When designing a CRISPR mutagenesis system for a specific host organism, it is necessary to take several aspects of the system into consideration. The researcher must find a suitable delivery mechanism for the two components of the system (Cas9 and the specific gRNA). Plasmids are the common delivery mechanism. They must replicate in the host and be possible to select for with, for instance, antibiotic selective markers. The plasmids must also be curable; after making the mutations, the plasmids must be removed to avoid subsequent off-target mutations. The researcher must choose a specific N20 sequence and a mutation template designed to introduce desired mutations.

### 1.4.1   The N20 spacer in the guide RNA

The N20 sequence in the gRNA must be homologous to the target sequence that lies directly upstream of a PAM site (target in the form of N20-NGG in *S. pyogenes*). The mutation frequency is independent of whether the N20 sequence targets the template or

non-template strand [49]. Several aspects of the N20 composition have been found to yield higher activity and increase the stability of the gRNA [50]. The study performed by Moreno-Mateos *et al.* (2015) implies that a cytosine or guanine being the N in the NGG PAM site, and depletion of guanine in the nucleotide directly downstream of the PAM, provides stronger binding of Cas9. A general depletion of adenine, a high guanine content in nucleotides 1-14 of the N20 sequence, and a cytosine enrichment in nucleotides 15-18 increases the stability of the gRNA. Guanine directly upstream of the PAM site (nucleotide 20) favors cleavage.

### 1.4.2 Introducing mutations with a mutation template

Two pathways can repair the nick in the target DNA - non-homologous end joining (NHEJ) and homology-directed repair (HDR)(Section 1.2). NHEJ is an error-prone repairing mechanism that can create both deletions and insertions of nucleotides, causing frameshift mutations[51]. Bacteria are usually not good at NHEJ. HDR is a DNA repairing mechanism that uses a homologous strand as a template for repair. When repairing the nicked sequence, mutations can be introduced by providing a template for HDR. The mutation template must have a few changes in the nucleotide sequence around the nick and be without a PAM site. Figure 1.4 shows how Cas9 guided by the gRNA can introduce mutations in the target DNA.
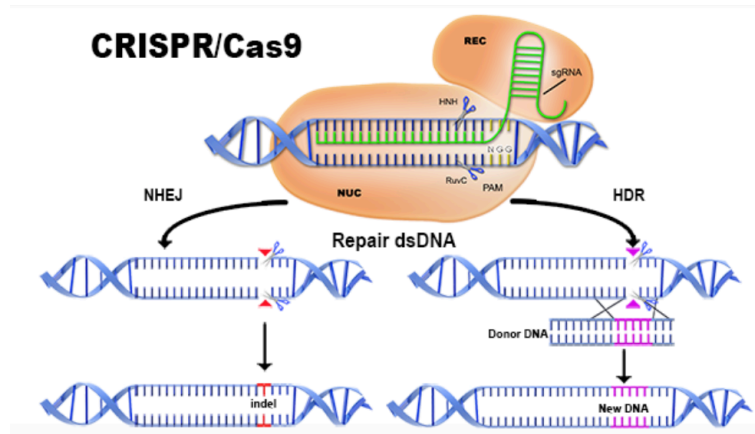
**Figure 1.4: Repairing mechanisms**(Figure copied from Advanced Analytical). The two repairing mechanisms NHEJ and HDR can introduce mutations in the target sequence. NHEJ can cause either deletions or the introduction of nucleotides in the target sequence, but this repairing mechanism usually does not happen in bacteria. HDR creates mutations through a donor DNA with homologous sequences to both sides of the nicked DNA.

## 1.5   Expression systems used in this work

Expression systems regulate the expression of genes, and can be used to tune the expression of genes under varying conditions [21]. There are three kinds of expression systems - inducible systems, repressible systems, and constitutive systems. In constitutive systems the genes are always expressed. The *j23119* promoter from *E. coli* is an example of a promoter of a constitutive system where the expression of the genes it controls, are not regulated [52]. Inducible and repressible systems can be under positive or negative control. The gene expression is under the control of regulatory genes, and their gene products. In positive control mechanisms, the regulator gene product is an activator, since it activates gene expression upon binding to a regulator protein-binding site (RPBS). A repressor is the product of the regulatory gene in a negative control mechanism because it represses gene expression upon binding. The activator and repressor will respectively promote or hinder the binding of RNA polymerase upon binding to their respective RPBS. The presence or absence of an effector molecule may decide whether the regulator binds to the RPBS or not. In inducible systems, an inducer is the name of the effector molecule since it induces transcription. In the case of an inducible system under positive control, the inducer binds to the activator which then binds to the RPBS, while under negative

control the inducer hinders the binding of the repressor. A co-repressor is the effector molecule in repressible systems since binding of it to the regulator protein will repress transcription. In positive control mechanisms, the effector molecule binds to the activator and thereby inactivating it so that it cannot bind to the RPBS. The RNA polymerase is then unable to bind the promoter, and gene expression is turned off. In negative control mechanisms, the co-repressor activates the repressor through binding so that it can bind to the RPBS and promote binding of RNA polymerase and transcription.

The *xylS/Pm* cassette is an inducible system under positive control [53]. The activator XylS binds to the RPBS in the presence of benzoic acids as inducers. The regulator/promoter system functions in a wide range of gram-negative bacteria. Other positive aspects of the system are that the inducer, benzoic acids such as m-toluic acid, are cheap and are taken up passively by the cell [54]. Figure 1.5 shows the XylS/*Pm* regulator/promoter system.
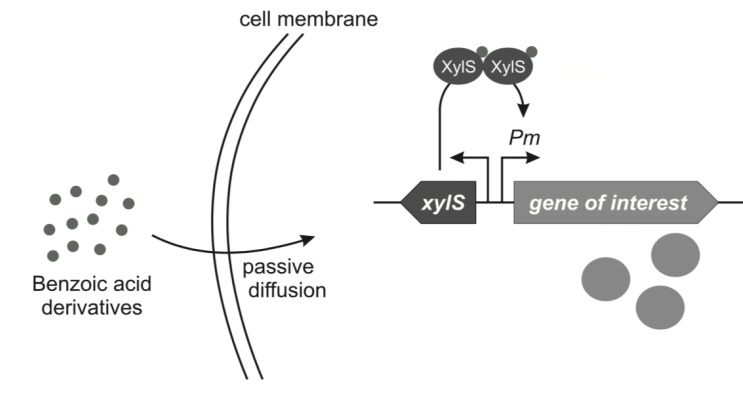


**Figure 1.5: The *xylS/Pm* cassette** (Figure is copied from Brautaset *et al.* (2009)[55]). Benzoic acid derivatives are taken passively up by the cell and bind two molecules of XylS. The inducer/activator complex induces transcription from the *Pm* promoter.

Another expression system is the AraC/$P_{BAD}$ regulator/promoter system controlling the proteins needed in the catabolism of L-arabinose [56]. In the absence of the inducer L-arabinose, the activator AraC binds to two regulatory sites, O and I1, inhibiting transcription. The presence of L-arabinose shifts the binding of AraC to I1 and I2, which stimulate binding of RNA polymerase and the expression of the genes controlled by the promoter $P_{BAD}$ (Figure 1.6) [57]. The expression level from the promoter is very low in the absence of L-arabinose and is further reduced in the presence of glucose [55]. In this way, the cell will only synthesize the proteins needed in the catabolism of L-arabinose

when there is no glucose present. This type of regulation is called catabolite repression. When the concentration of glucose is low, the enzyme adenyl cyclase is activated and converts ATP to cAMP which binds and activates CAP (cAMP activator protein)[58]. CAP stimulates binding of the AraC dimer to the I1 and I2 site.
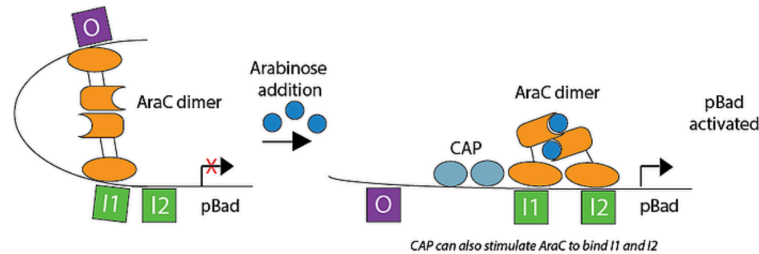


**Figure 1.6: The AraC/$P_{BAD}$ regulator/promoter system** (Figure is copied from Addgene's Blog [59]). When arabinose is added to the medium the AraC dimer change conformation and binds the two promoters, I1 and I2, promoting transcription from $p_{BAD}$. AraC is bound to the repressor site O and the promoter I1 in the absence of arabinose.

## 1.6 Donor strains used in conjugation in this study

*E. coli* S17-1 can transfer, or donate, genes through the process of conjugation. The donor cell is designated F$^+$ by the presence of an F factor. The F factor is a small circular molecule, regulating the synthesis of a cellular extension called F pilus [21]. For the process to occur, the donor and recipient cell must be in direct contact. The donor cell pulls the recipient cell into close contact through the F pilus. The donor can transfer DNA through a channel that forms between the donor and recipient cell. One strand of the F factor is cleaved and passed through the conjugation channel. A site called the origin of transfer (*oriT*) is the initiation site for transfer and replication. When the single strand enters the recipient cell, replication is initiated from *oriT* and results in a double-stranded circular DNA molecule. S17-1 contains a *tra* operon on the chromosomally integrated RP4 plasmid. *tra* contains the essential genes for conjugation, among them the gene encoding the F pili [60]. For conjugation to be possible, the plasmid must include *oriT*, and the donor strain contain the *tra* genes.

*E. coli* S17-1 is naturally resistant to spectinomycin [60]. S17-1 cannot function as a donor strain in conjugation for plasmids with a resistance gene to this antibiotic. In this study, *E. coli* DH5α (pTA19)[61] was utilized as an alternative donor strain for plasmids

giving resistance to spectinomycin. The strain can operate as donor strain since pTA19 contains the *tra* genes and a deficient *oriT* site. Hence, the plasmid can aid in transfer without being transferred itself.

## 1.7   Broad-host-range replicons used in this work

The origin of replication, *ori*, is a sequence from where replication is initiated [23]. Two plasmids in the same bacterium cannot have the same replication system, because the replication of the two will interfere with each other.

The RK2 replicon is a broad-host-range system [62], that belongs to the IncP plasmid group known to function in many gram-negative bacteria and which can be transferred by conjugation [53]. The replicon consists of the initiator protein TrfA and the origin of replication, *oriV*. *trfA* encodes for two versions of the initiation protein because of an internal in-frame translation start [53]. Either form of the proteins encoded by the gene, binds to a region in *oriV* which facilitates an open complex and initiation of replication [62].

In the plasmid pHE95, replication is dependent on the inducer m-toluate since *Pm* regulates TrfA[63]. The addition of m-toluate to the media activates XylS and so induces the expression of TrfA from the promoter *Pm*. Correspondingly, TrfA initiates replication of the plasmid from *oriV*. Since *E. coli* S17-1 can produce TrfA [60] it is an ideal cloning host for plasmids constructed from pHE95 since the addition of m-toluate is not necessary for replication. However, the plasmid copy number increases with increasing concentration of the inducer. pHE95 based plasmids that are transferred by conjugation to a recipient host with S17-1 as a donor, require the inducer to replicate in the new host. In this way, the plasmid can also be removed from the bacterium if desired, by not adding m-toluate to the medium.

pBBR1 is another broad-host-range vector that is known to replicate in several gram-negative bacteria including *A. vinelandii* [64]. The pBBR1 backbone is compatible with IncP plasmids and can be transferred conjugatively via S17-1 [64]. It consists of two cassettes - *rep* encoding the protein Rep for replication, and *mob* encoding the mobilization

protein Mob for conjugation. The vector does not replicate by a rolling circle mechanism, but by theta replication [65]. The initiator protein Rep binds an origin of replication site, *ori*, promoting unwinding of the double strand and creating two single strands which can serve as templates for replication [66]. The replication proceeds in each direction so that when viewed from above the plasmid will resemble the symbol $\theta$. Mob recognizes a 52 bp sequence which contains both the transfer origin (*oriT*) and promoter of *mob* [67]. Mob is a relaxase that binds to *oriT*, where it initiates the transfer by nicking the DNA at a *nic* site.

The pBBR1 vector is not inducible and is therefore not curable such as the RK2 replicon controlled by the XylS/*Pm* cassette. A vector constructed based on the pBBR1 replicon with *sacB* could be removable by adding 5% sucrose to the growth medium, given that *sacB* works in the bacterium. *sacB* is known to function in *A. vinelandii*[63].

# 2  Aim of the study

This study aims to develop a CRISPR-Cas9 mutagenesis system for *A. vinelandii*. Homologous recombination is today used to mutagenize *A. vinelandii*. However, the procedure is hampered by the bacterium having many copies of its chromosome. CRISPR-Cas9 can mutagenize several chromosomes at the same time and with high efficiency. CRISPR-Cas9 may also be utilized to remove wild-type chromosomes.

A two-plasmid recombination system was decided upon to deliver the necessary components of CRISPR (Cas9 and the gRNA) to *A. vinelandii*. The system was based on experiences from Sun *et al.*(2018), where they successfully developed a CRISPR-Cas9 mutagenesis system for *P. putida*, a relative of *A. vinelandii*. The two-plasmid system was chosen because it is easier to cure for each of them. Also one plasmid, with both the *cas9* and *gRNA* genes, might be challenging to construct and transfer because of the size. Two target genes were decided on - *mucA* encoding a negative regulator in the alginate synthesis and *uidA* encoding $\beta$-glucuronidase, an enzyme yielding a blue product in the presence of the substrate X-Gluc. Mutation templates for each target needed to be included to introduce mutations during HDR. A *mucA*$^-$ mutant would result in overproduction of alginate and mucoid colonies. An *A. vinelandii* strain containing *uidA* exists (Ertesvåg, unpublished). A *uidA*$^-$ strain would yield white colonies in contrast to blue colonies if CRISPR does not work.

The plasmids should be cloned using standard cloning techniques with *E. coli* DH5$\alpha$ and S17-1 as cloning hosts. The two-plasmid system wold consist of a Cas9-encoding vector and several variations of a gRNA-encoding vector. The plan was four gRNA vectors; two for each target gene (*mucA* and *uidA*), one with a mutation template and one without. Three strains of *A. vinelandii* would be used - one wild-type *A. vinelandii* ATCC 12518, one *A. vinelandii* ATCC 12518 *uidA* strain, and an *A. vinelandii* ATCC 12518 *mucA*$^-$ strain with both wild-type and mutant *mucA* which would be constructed with homologous recombination. The lethality of CRISPR would be tested with the plasmids without a mutation template, while the two with an additional mutation template would be used to mutagenize the bacterium. The plasmid constructed to mutagenize *mucA*

would also be used with the $mucA^-$ strain as the target, to see if CRISPR could be used to get a pure mutant strain. The goal of the study was a new and improved mutagenesis strategy compared to homologous recombination.

## 2.1 Design of a two-plasmid CRISPR-Cas9 system for *A. vinelandii*

The plasmids must be constructed from replicons and regulator/promoter systems that are known to function in *A. vinelandii*. Antibiotic selective markers for selection of the plasmids, and curing components so that removal of them from the bacteria is possible, were included in the design.

Jiang *et al.* cloned the two plasmids pCas and pTargetF by using the genome of *Streptococcus pyogenes* from which the CRISPR-Cas type II system originates [34]. The two plasmids provided the genes encoding Cas9 (pCas) and the gRNA scaffold (pTargetF). These genes were used to design a two-plasmid system derived from the RK2 plasmid pHE95 and the pBBR1MCS-2 plasmid. Figure 2.1 shows the two planned plasmids. The plasmid to the left is the Cas9 vector, and the plasmid to the right is the gRNA vector.
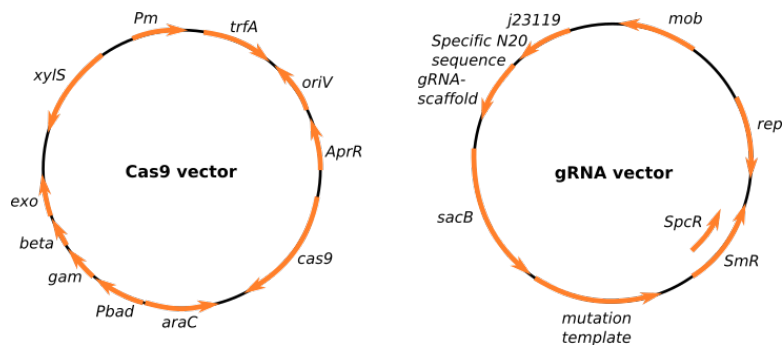


**Figure 2.1: The final plasmids of the two-plasmid CRISPR-Cas9 system.** To the left; the Cas9 vector. To the right; the gRNA vector is shown with the specific N20 sequence and with a mutation template. The vector can be constructed with different N20 sequences depending on which gene is the target, and without the mutation template.

The RK2 replicon with the origin of replication *oriV* and the gene *trfA* encoding the initiation protein would be the basis of the Cas9 vector (figured to the left in Figure 2.1). The replication of the plasmid would be under the control of the XylS/*Pm* regulator/promoter system. The plasmid should only replicate in the presence of the inducer

m-toluate. Consequently, the regulator/promoter system would serve as a curing component so that removal of the plasmid is possible. An apramycin resistance gene would serve as a selectable marker. The expression of the $\lambda$-Red genes *gam*, *beta* and *exo* would be regulated by the inducible AraC/$P_{bad}$ regulator/promoter system. In the presence of the inducer arabinose, the expression of the genes of the $\lambda$-Red system would be induced. When Cas9 executes its nuclease activity, the system could aid the bacteria in HDR of the nicked DNA if a mutation template with homologous regions was present. If arabinose was not present in the medium, the genes would not be expressed, and HDR would not occur. Thus, the CRISPR system could be used both as a mutagenesis system and as a way to remove wild-type chromosomes if the bacterium contains a combination of wild-type and mutant chromosomes. The cloning procedure to construct the Cas9 vector is described in Section 4.1.

Figure 2.1 (to the right) shows the general gRNA vector. The plasmid would be based on the pBBR1 replicon with its genes *mob* and *rep*, and contain the gRNA scaffold and its promoter consecutive *j23119* from pTargetF. The specific N20 sequence would be cloned between the promoter and the scaffold. A spectinomycin resistance gene would serve as a selectable marker and *sacB* as a curing component since the expression would be expected to be lethal to the bacterium in the presence of 5% sucrose. Lastly, a mutation template should be transferred to the plasmid so that the system could work to mutagenize the bacterium in the presence of arabinose. The cloning procedure to construct the gRNA vector is described in Section 4.2.

## 2.2 Plasmids constructed during the TBT4500 specialization project, autumn 2018

This project was started in the autumn of 2018. The plasmids constructed during this master's thesis is cloned from plasmids that were constructed during the specialization project (TBT4500) conducted during the autumn of 2018 (pMDD6, pMDD7 and pMDD9). The flow diagrams of the cloning of pMDD6, pMDD7 and pMDD8 are shown in Figure 4.1 and 4.3.

Plasmid pMDD6 is based on the RK2 replicon pHE95. The XylS/*Pm* regulator/promoter system controls the replication of the plasmid. Since the system is inducible, the plasmid will not replicate in the absence of its inducer, m-toluate. Thereby, the system functions as a curing component for the vector, meaning that the plasmid can be removed from the bacteria when desired. pMDD6 also carries a resistance gene against ampicillin. The ampicillin resistance gene of pMDD6 is replaced by an apramycin resistance gene in pMDD7. As for pMDD6, the *xylS/Pm* cassette regulates the replication of the plasmid from the RK2 replicon.

The precursor of the gRNA-encoding vector, pMDD9, consists of the narrow-host-range *colE1* replicon with the gRNA scaffold controlled by the promoter *j23119*, a spectinomycin resistance gene for selection, and *sacB* for curing.

# 3 Methods

Section 3.1 describes the general cloning techniques used to construct the plasmids, while Section 3.2 and 3.3 describes how the plasmids were transferred to *A. vinelandii.*

Description of methods used in the TBT4500 - specialization project, autumn 2018, are copied from that report.

Recipes for media and buffers are attached in Appendix A.

## 3.1 Standard cloning procedures with *E. coli* as cloning host

### 3.1.1 Bacterial strains used in cloning and growth of bacteria

*E. coli* DH5$\alpha$ was used as the standard host, while *E. coli* S17-1 was used in conjugation and as cloning host for plasmids constructed from the RK2 replicon with the *xylS/Pm* cassette.

*E. coli* was cultivated in Luria-Bertani (LB) medium with the appropriate antibiotics, depending on which plasmid the bacteria contained. For which antibiotic was used for selection of the different plasmids, see Table B.2 in Appendix B. The working concentration of kanamycin, ampicillin, apramycin, spectinomycin and tetracycline was respectively 50 $\mu$g/mL, 100 $\mu$g/mL, 50 $\mu$g/mL, 20 $\mu$g/mL and 10 $\mu$g/mL. Growth of the bacteria and different plasmids needed addition of distinct antibiotics (Table B.2 in Appendix B). The recipe of the LB medium is given in Appendix A. The bacteria were inoculated overnight at 37 °C.

For long term storage, 800 $\mu$L of the cell cultures were mixed with 400 $\mu$L of 60% glycerol in cryotubes and stored at $-80$ °C.

### 3.1.2   Preparation of chemically competent *E. coli* cells

Competent cells are cells with the ability to take up free DNA [21]. *E. coli* is not naturally competent but can be made competent by chemically treating it with buffers containing calcium chloride. Calcium chloride reduces the repulsion between the free DNA and the cell surface so that it can bind and take up DNA [68]. In this study, the buffers that are used (TBF1 and TBF2) contain rubidium chloride in addition to calcium chloride. RbCl works similarly as $CaCl_2$ by establishing an interaction between the free DNA and the cell surface so that the DNA can be taken up.

*E. coli* DH5$\alpha$ was transferred to 10 mL Psi medium in 100 mL shaking flask and incubated over the night at 37 °C with shaking. Early the next morning, 2 mL of the overnight culture was transferred to 200 mL Psi medium and incubated at 37 °C with shaking. Samples were taken out and measured using a spectrophotometer at 600 nm until the OD reached 0.4. The culture was put on ice along with buffers TBF1 and TBF2. After 15 minutes on ice, the culture was transferred to a sterile centrifuge cup and centrifuged at 4000 rpm and 4 °C for five minutes. The supernatant was removed and the cell pellet resuspended in 80 mL TBF1 (80 mL) by carefully shaking the flask. When finally dissolved, the suspension was put on ice for five minutes, and then centrifuged again at 4000 rpm and 4 °C for five minutes. After removal of the supernatant, the remaining cell pellet was carefully dissolved in 6 mL of buffer TBF2. The competent cells were snap-frozen in aliquots of 100 $\mu$L in 1.5 mL Eppendorf tubes, using liquid nitrogen. The tubes were transferred to a cryobox and placed in the $-80$ °C freezer.

The transforming frequency of the competent cells was measured to be $6.8 \cdot 10^7$ cfu/$\mu$g DNA and $1.8 \cdot 10^5$ cfu/$\mu$g DNA for *E. coli* DH5$\alpha$ and S17-1, respectively.

### 3.1.3   Heat-shock transformation of plasmids into chemically competent *E. coli*

Chemically competent *E. coli* can take up exogenous DNA through a process called heat-shock transformation. The calcium- or rubidium-rich environment, in which the cells are made competent, facilitates the interaction between the bacterial membrane and the

exogenous DNA. A sudden increase in temperature helps the free DNA enter the cell [68].

An aliquot of 100 $\mu$L competent *E. coli* DH5$\alpha$ or S17-1 were thawed on ice for 5-10 minutes. 10 $\mu$L of DNA was added to the competent cells and incubated on ice for 30 minutes. The cells were heat-shocked in a water bath at 42 °C for 45 seconds, and immediately put back on ice for two minutes. SOC medium (1 mL) was added to the mixture. The tube was placed in an Erlenmeyer flask and incubated at 37 °C with shaking for one to three hours. 50 and 100 $\mu$L of the transformed cells were plated out on LA plates with appropriate antibiotics. The remaining cell suspension was concentrated by centrifugation at 1200 x g for 30 seconds. 700 $\mu$L of the supernatant was discarded, and the cell pellet resuspended in the remaining fluid and plated out. The three plates were incubated at 37 °C over the night.

### 3.1.4  Preparation of electrocompetent *E. coli*

Electrotransformation is a technique to facilitate the uptake of free DNA into a cell. The protocol involves preparing the cells so that they can survive the electrical pulse by washing them to remove any charged chemicals or salts. An electrical pulse creates pores in the cell membrane and shoots the negative charged DNA towards the positive pole, and into the cell.

An overnight culture of *E. coli* was grown in LB medium. Appropriate antibiotics were added if the strain already contained a plasmid. 5 mL of the culture was transferred to 500 mL sterile LB medium in a 3 L shaking flask the next morning. Antibiotics were added with the same concentration as in the pre-culture. The cells were inoculated at 37 °C with shaking, until the OD$_{600}$ measured between 0.5-0.7. The medium was put on ice for about 20 minutes and then transferred to a chilled 500 mL centrifuge bottle. Everything was kept on ice from this point on. The cell suspension was centrifuged for 15 minutes at 4 °C and 4000 x g. The supernatant was then poured off, and the cells resuspended in 500 mL ice-cold 10% glycerol. These steps were repeated three more times - centrifugation, discarding the supernatant, and resuspension of the cell pellet in first 250 mL, then 20 mL, and at last 1 mL ice-cold 10% glycerol. The cells were frozen in aliquots of 20$\mu$L in 1.5 mL Eppendorf tubes, using liquid nitrogen. The tubes were

transferred to a cryobox and stored in the $-80\,°$C freezer.

### 3.1.5  Electroporation of *E. coli* DH5$\alpha$

20 $\mu$L of electrocompetent *E. coli* DH5$\alpha$ (pTA19) were mixed with 1-2 $\mu$L of plasmid DNA while placed on ice. After incubation on ice for about 1 minute, the mixture was added to a cold 0.1 cm electroporation cuvette. The cells were electroporated by delivering 1.8 kV for approximately 5 ms using Bio-Rad Gene Pulser Xcell. The electroporated cells, now containing the plasmid, were transferred from the cuvette to a 1.5 mL Eppendorf tube and inoculated at $37\,°$C with shaking for 1-2 hours. The cells were plated on LA plates with appropriate antibiotics, as described in Section 3.1.3.

### 3.1.6  Isolation of plasmids

Plasmids from overnight cultures or single colonies on agar plates were isolated with the Monarch Plasmid Miniprep Kit or ZymoPure Plasmid Miniprep Kit. Both kits employ alkaline lysis and color indicators of the same colors.

For low-copy-number plasmids, 10 mL of the overnight culture was used, whereas 1.5 mL for high-copy number. The cells were pelleted by centrifuging at 16000 x g for 30 seconds. After discarding the supernatant, the cell pellet was resuspended in 200 $\mu$L of a buffer (Plasmid Resuspension Buffer or ZymoPure P1). The buffer contains EDTA which works by chelating divalent cations needed for DNase.

Addition of 200 $\mu$L of an alkaline solution (Plasmid Lysis buffer or ZymoPURE P2) containing a detergent and a base, disrupted the cell membrane and chromosomal DNA. The solution was gently mixed until the color changed to dark pink and became transparent and viscous. After incubation at room temperature for 1-2 minute, 400 $\mu$L of a neutralization buffer (Plasmid Neutralization Buffer or ZymoPURE P3) was added to prevent destruction of the plasmid. The buffer also contains RNase that destroys unwanted RNA. The tube was gently inverted. The neutralization of the mixture was complete when the color changed from dark pink to yellow and precipitate formed. The solution was incubated at room temperature for 1-2 minutes. The plasmids were isolated by transferring

the supernatant containing the plasmid, to a spin column, and centrifuging for 4-5 minutes. The plasmids would then bind to the membrane. The flow-through was discarded.

200 $\mu$L of a wash buffer was added (Plasmid Wash Buffer 1 or Endo-Plasmid Wash buffer) to the column to remove unwanted cellular components and salts. After centrifuging the column for 1 minute, 400 $\mu$L of a second wash buffer (Plasmid Wash Buffer 2 or Plasmid Wash Buffer) was added, and the column centrifuged for another minute. The column was transferred to a clean Eppendorf tube, and 30-60 $\mu$L of DNA Elution Buffer was added to the center of the column to elute the plasmid from the membrane. After 1 minute the plasmids were eluted to the Eppendorf tube by centrifuging for 1 minute. The plasmid solution was stored at $-20\,°$C.

### 3.1.7  Restriction cutting of plasmids

Restriction enzymes were used to cut out desired genes for further cloning or to control that the newly cloned plasmid was different from the ones it was derived. The restriction enzymes used during this experiment was provided by New England BioLabs, and come with optimal buffers for each enzyme.

Restriction enzymes type II are endonucleases that recognize short DNA sequences and cut the double-stranded DNA at the site of recognition [69], and therefore, are widely used as a laboratory tool in cloning and gene mapping. Benchling is a digital online tool that is free of charge [70]. It provides visualization of the plasmids so to find cutting sites on each side of the area of the fragment that was to be selected. Restriction enzymes creating complementary sticky ends were chosen so that two sticky ends could be ligated together to form a new plasmid. If two restriction enzymes were used, at least one of them created sticky ends. For a 20 $\mu$L reaction mixture, 150-400 ng DNA was mixed with an optimal buffer (2 $\mu$L) and water. The quantity of water depends on the volume of plasmid and if one or two restriction enzymes were used, but chosen so that the final reaction volume reached 20 $\mu$L. 0.5 $\mu$L of one or two restriction enzymes was then added to the reaction mixture. The cutting reaction was incubated at the optimal temperature for the restriction enzymes, 37$\,°$C, for 1 hour to overnight.

### 3.1.8  Agarose gel electrophoresis

Agarose gel electrophoresis separates nucleic acids of different sizes based on the charge of the molecules. It was used to visualize or separate the DNA fragments produced in the restriction cutting, or PCR products.

A gel was made from 0.8% agarose TAE-buffer GelRed or GelGreen - GelRed for visualization only, and GelGreen when isolation of DNA from the gel was desired. The addition of 10% of 10x loading dye to the sample provides visualization of the DNA. The sample was then loaded onto the gel. By applying an electric field to the gel, the negative charged nucleic acid traveled through the pores of the agarose gel, to the positive pole. The larger fragments were held back so that, the smaller fragments moved faster through the gel. The separation was run at 85-100 V for as long as needed to obtain a good separation. Fragments separated on GelRed were visualized under UV light, while fragments isolated on GelGreen were visualized with a blue light screen converter to avoid UV-induced DNA-damage.

Two DNA ladders, Lambda DNA HindIII and Lambda DNA PstI, with DNA fragments of known molecular sizes, were run on the gel to be compared with the unknown samples.

### 3.1.9  Isolation of DNA restriction fragments from gel

The Monarch DNA Gel Extraction Kit was used to isolate DNA fragments from the agarose gel.

The DNA fragment was excised from the agarose gel and transferred to an Eppendorf tube. The gel was dissolved in four times the gel volume of Gel Dissolving Buffer, by incubating the sample for 5-10 minutes at 54 °C while vortexing regularly.

The sample was transferred to the spin column and centrifuged for 1 minute at 16000 x g and room temperature to isolate the DNA. During this step, the DNA would bind to the membrane of the spin column. After discarding the flow-through, the column with the bound DNA was washed twice by adding 200 $\mu$l of DNA Wash Buffer and centrifuging it for 1 minute at 16000 x g at room temperature. The spin column was transferred to

a clean Eppendorf tube. 6-20 $\mu$L of Elution Buffer was added directly to the center of the column to elute the DNA. After waiting 1 minute, the eluate containing the DNA was spun down for 1 minute at 16,000 x g and room temperature. The eluted DNA was stored at $-20\,°C$.

### 3.1.10   Ligation of DNA fragments

T4 DNA Ligase from New England BioLabs was used in cloning of the restriction fragments. T4 DNA Ligase catalyzes the formation of a phosphodiester bond between the free 5' phosphate and 3' hydroxyl end of the DNA strands. Ligase can catalyze the formation of a phosphodiester bond between both blunt and sticky end DNA fragments.

For the ligation of two DNA fragments (a vector and an insert molecule); 14.5 $\mu$L insert DNA, and 3 $\mu$L vector DNA were mixed. 2 $\mu$L T4 DNA Ligase buffer and 0.5 $\mu$L of T4 DNA Ligase was added to the mixture. The reaction mixture was incubated at $16\,°C$ (optimal temperature for T4 DNA Ligase) for 1 hour to over the night.

### 3.1.11   Amplification of DNA by PCR

The polymerase chain reaction (PCR) is a method where a DNA sequence is amplified *in vitro* by the use of DNA polymerase and two specific DNA primers. The DNA primers are synthetic single-stranded oligonucleotides that are complementary to the DNA strand. The primers used in this master's thesis are provided by Sigma-Aldrich and Invitrogen (Thermo Fisher Scientific), and are listed in Table C.1 in Appendix C. The PCR reaction was carried out using Q5 High-Fidelity DNA polymerase (New England BioLabs inc.) A reaction mixture was set up, as shown below with a final volume of 25 $\mu$L.

5$\mu$L 5X Q5 reaction buffer

0.5 $\mu$L 10mM dNTP

1.25 $\mu$L 10 $\mu$M Forward Primer

1.25 $\mu$L 10 $\mu$M Reverse Primer

<1.000 ng Template DNA

0.25 $\mu$L Q5 High-Fidelity DNA Polymerase

5 $\mu$L 5X Q5 High GC Enhancer

Nuclease free water up to 25 $\mu$L

The sample was placed in a PCR machine, and a program consisting of seven steps was set up (Appendix D). The amplification of the DNA strand occurs in three stages - denaturation, annealing, and extension. The first step in the program was a primary denaturation to break the hydrogen bonds holding the two single DNA strands together, with the use of high temperature of 98 °C. In the second step, the high temperature was held for 10 seconds, before the reaction mixture was cooled down to a temperature of 3 °C lower than the melting temperature ($T_m$) of the primer with the lowest $T_m$ (step 3). At this stage, the *annealing* stage, the primers hybridize with the single strands of the DNA template. Each primer is complementary to one of the complementary strands and at opposing ends. At step 4, the *extension* stage, the temperature was set to 72 °C. The DNA polymerase binds to the free 3' OH of each primer and extends the strand by using the denatured single DNA strand as a template for synthesis. The extension step was set to 30 seconds per 1 kb. Step 2-4 was repeated 25-35 times. A final extension of the temperature was at 72 °C for 2 minutes. The last step in the procedure was to hold the temperature at 4 °C until the sample was collected from the machine and stored in the freezer at −20 °C.

### 3.1.12 Topocloning for blunt end PCR products

Zero Blunt TOPO PCR Cloning Kit from Invitrogen was used to clone blunt-end PCR products. The blunt PCR product is cloned into a plasmid vector by the use of *Vaccinia* virus DNA topoisomerase I. The TOPO vector is supplied linearized containing topoisomerase I attached to the 3' end of each DNA strand. The TOPO vector (0.5 $\mu$L) was mixed with 2 $\mu$L PCR product and 0.5 $\mu$L salt solution from the cloning kit. The reaction mixture was mixed gently and incubated at room temperature for 5-30 minutes, and then placed on ice. Topoisomerase I catalyzes the formation of phosphodiester bonds between the TOPO vector and the PCR product. The introduction of the PCR product disrupts the expression of the lethal fusion gene *lacZα-ccdB*, enabling a very strong positive selection for bacteria containing plasmids with insert after transformation. The

TOPO cloning reaction was transformed into competent cells immediately or stored in the freezer at $-20\,°C$ overnight. The TOPO vector with the PCR insert was selected for on LA plates with kanamycin (50 $\mu$g/mL).

### 3.1.13   SLIC cloning of PCR products

Sequence- and Ligation Independent Cloning is a cloning technique that does not require the use of ligase. It utilizes the 3'-to-5' exonuclease activity of T4 DNA Polymerase to create complementary single-stranded overhangs between the vector and insert DNA molecule [71]. Restriction enzyme digestion linearizes the vector, and the insert is amplified by PCR using primers with 15 bp extension homologous to the ends of the linearized vector. Both the PCR product and the linearized vector were purified with NEB PCR Purification Kit and eluted with 10 $\mu$L 10 mM TrisCl, pH 8.5.

The linearized vector and PCR product were mixed at a molar ratio of 1:2. The mass of the insert was calculated based on the molar ratio times the mass of the vector used (100 ng) and the ratio of the insert to vector length. 1 $\mu$L NEB buffer 2.1 was added to the mixture. For the final reaction volume to reach 10 $\mu$L, an appropriate amount of water was added. 0.5 $\mu$L T4 DNA polymerase was added to the reaction mixture and incubated at 2.5 minutes at room temperature. At this stage, T4 DNA polymerase would use its exonuclease activity to remove bases in the 3'-to-5' direction of both the PCR product and the linearized vector, creating homologous overlaps. The mixture was then put on ice to stop the reaction, and for the two single strands to anneal. The solution was incubated for 10 minutes on ice.

Competent *E. coli* DH5$\alpha$ were thawed on ice, and 2 $\mu$L of the reaction mixture was added to the cells. The transformation was carried out as described in Section 3.1.3. The nick in the DNA was sealed *in vivo* by DNA ligase.

### 3.1.14   NEBuilder HiFi DNA Assembly Reaction

NEBuilder HiFi DNA Assembly Reaction can be used to assemble single-stranded oligonucleotides. The oligonucleotides must be designed to include about 20 nucleotides overlap

homologous to each end of the linearized vector. The vector was first linearized and the restriction enzymes inactivated by heat. A mixture with a molar ratio of vector to insert of 1:10 was added to 10 $\mu$L 2x NEBuilder HiFi DNA Assembly Master Mix. The mass of the insert was calculated based on the molar ratio times the mass of the vector used (100 ng) and the ratio of the insert to vector length. Water was added so that the final volume reached 20 $\mu$L. The reaction mixture was incubated at 50 °C for 1 hour. The master mix includes different enzymes that work together in the same buffer to assemble the oligonucleotide into the linearized vector. An exonuclease creates single-stranded 3' overhangs on the linearized vector so that complementary regions of the oligomer and the vector can anneal. A polymerase fills in the gaps, and a ligase seals the nicks.

The reaction mixture was stored at $-20$ °C or transformed immediately by diluting the sample by 1:4 with water, and adding 2 $\mu$L to the competent *E. coli* DH5$\alpha$ thawed on ice.

### 3.1.15 Sanger sequencing

PCR products cloned into a TOPO vector or SLIC cloned, and plasmids cloned with NEBuilder HiFi DNA Assembly Reaction were controlled by Sanger sequencing, performed by GATC - A Eurofins Genomics Company. The Sanger method can be used to find the nucleotide sequence of a DNA molecule, and thereby to see if any mistakes have been made in cloning a sequence during PCR [21]. The method, also called the 2',3'-dideoxynucleoside triphosphate chain-termination procedure, utilizes *in vitro* DNA synthesis by adding nucleotides that lack the free 3'-OH end (ddXTP; X any of the four bases) needed by DNA polymerase for chain elongation.

A reaction mixture with 400-500 ng template strand to be sequenced was mixed with 0.25 pmol primer and sent into GATC, where they performed the sequencing. Enough primers to cover the whole stretch to be sequenced were chosen, and one primer was added per reaction. First, a DNA polymerization is performed by adding DNA polymerase, all four nucleotides (dGTP, dATP, dTTP, and dCTP), and the chain terminators (ddGTP, ddATP, ddTTP, and ddCTP) labeled with different fluorescent dyes. When a ddXTP (X being any base) attaches to the chain, the synthesis stops since DNA polymerase

cannot elongate the chain, without a free 3' -OH end. The resulting products will be of varying length, each with a fluorescent labeled ddXTP at the end of the chain. The products are separated by polyacrylamide capillary gel electrophoresis based on size so that the shortest chains pass through first. Each chain being one nucleotide longer than the previous, all the nucleotides of the chain are determined in successive order. Their presence in the gel is detected with a scanning laser and a fluorescence detector.

The sequencing results from GATC were aligned with the template using the alignment tool Mview accessed at The European Bioinformatics Institute [72]. The aligned sequencing files were shown as a combination of capital letters and "-". The four bases in the template are colored - blue for thymine (T), yellow for cytosine (C), and green for guanine (G) and adenine (A). The base of the aligned sequence is shown in the same color as the base of the template when they match. A non-colored base of the aligned sequence, indicates a mismatch between the template and the aligned sequence. A "-" spaced between two bases in the aligned sequence indicates that a base is lacking compared to the template, and hence, that the aligned sequence has been shifted. Consecutive "-" represents that there is no alignment between the template and aligned sequence (that the alignment has not started or stopped).

## 3.2 Conjugation of plasmids into *A. vinelandii*

The donor and recipient cells were first cultivated separately. *A. vinelandii*, the recipient cell, was grown in Burks medium for 2-3 days at 30 °C before 1% of the culture was transferred to a new medium and grown over the night. *E. coli* S17-1 or DH5$\alpha$ (pTA19) with the plasmid to be transferred was grown over the night with desired antibiotics at 37 °C. m-toluate (0.5 mM) was added to the medium for plasmids with the *xylS/Pm* cassette. 1% of the culture was transferred to a new medium without antibiotics and inoculated for 2-3 hours with the same conditions. To bring the cells into contact, 4-6 mL of each were mixed and centrifuged. The supernatant was discarded, leaving approximately 100 $\mu$L in which the cell pellet was resuspended. The cells were plated on LA plates and grown over the night at 30 °C. In this step, the plasmid would be transferred from the donor cell to *A. vinelandii*. A dilution series was prepared ($10^0$, $10^{-1}$, $10^{-2}$, $10^{-4}$ and

31

$10^{-6}$) from the bacterial growth on the LA plate, to select for growth of *A. vinelandii* that had received the plasmid. Each dilution was plated on Burks agar (BA) plates with the appropriate antibiotic.

## 3.3   Electrotransformation of plasmids into *A. vinelandii*

Recipient *A. vinelandii* cells were prepared by setting up a pre-culture in Burks medium and inoculating it for 2-3 days at $30\,^{\circ}\text{C}$ with shaking. 1% of the culture was added to new Burks medium, and grown at $30\,^{\circ}\text{C}$ with shaking until $\text{OD}_{620}$ measured 0.4-0.5. The cells were harvested by centrifuging the culture at 5000 rpm at $0\,^{\circ}\text{C}$ for 10 minutes. The supernatant was discarded, and the cells resuspended in 10% ice-cold sterile glycerol to the original culture volume. The 10% glycerol was kept on ice at all time. The suspension was centrifuged again with the same conditions, and the supernatant discarded. Resuspension, centrifugation, and removal of the supernatant were repeated two more times, with 10% glycerol volumes of 50% and 25% of the original culture volume. Finally, the pellet was resuspended in 2-4 mL 10% glycerol.

Aliquots of $40\mu\text{L}$ were snap-frozen with liquid nitrogen and stored at $-80\,^{\circ}\text{C}$. For the electrotransformation, $40\ \mu\text{L}$ of the recipient cells were mixed with $1\ \mu\text{L}$ plasmid ($300\ \text{ng}/\mu\text{L}$) in sterile electroporation cuvette with an interelectrode distance of 0.1 cm. The cells were delivered one pulse of 1.8 kV for about 5 ms using a Bio-Rad Gene Pulser Xcell machine. 1 mL Burks medium was added immediately to the cuvette, and the cells resuspended. The mixture was transferred to a sterile 1.5 mL Eppendorf tube and incubated at 200 rpm at $30\,^{\circ}\text{C}$ for one hour. The solution was plated on selective BA plates.

# 4    Results

Section 4.1 and 4.2 describes the cloning procedure of the Cas9-encoding vector and the gRNA vectors, respectively. All plasmids are listed in Table B.2 in Appendix B. Several elements of the plasmids such as expression systems, selectability, and viability were tested in parallel to the construction of them. This is described in Section 4.3.

## 4.1    Construction of a vector for Cas9

The vector for Cas9 was constructed from pHE95, an RK2 replicon with the regulator/promoter system XylS/*Pm* which regulates and controls the replication of the plasmid. pMDD6 and pMDD7 were cloned during the subject TBT4500 - Biotechnology, specialization project. The cloning procedure of the Cas9-encoding vector is presented in Figure 4.1. The figure illustrates how *sacB* was cut out from the RK2 replicon pHE95 to produce a smaller plasmid, pMDD6. The TrfA/*oriV* replicon and the XylS/*Pm* regulator/promoter system were amplified by PCR from pMDD6. An apramycin resistance gene was isolated from pION100 and cloned into the RK2 backbone. The new plasmid was designated pMDD7. *cas9* and the genes for the $\lambda$-Red recombineering system, were cleaved from pCas, and cloned into pMDD7, to yield pMDD8.
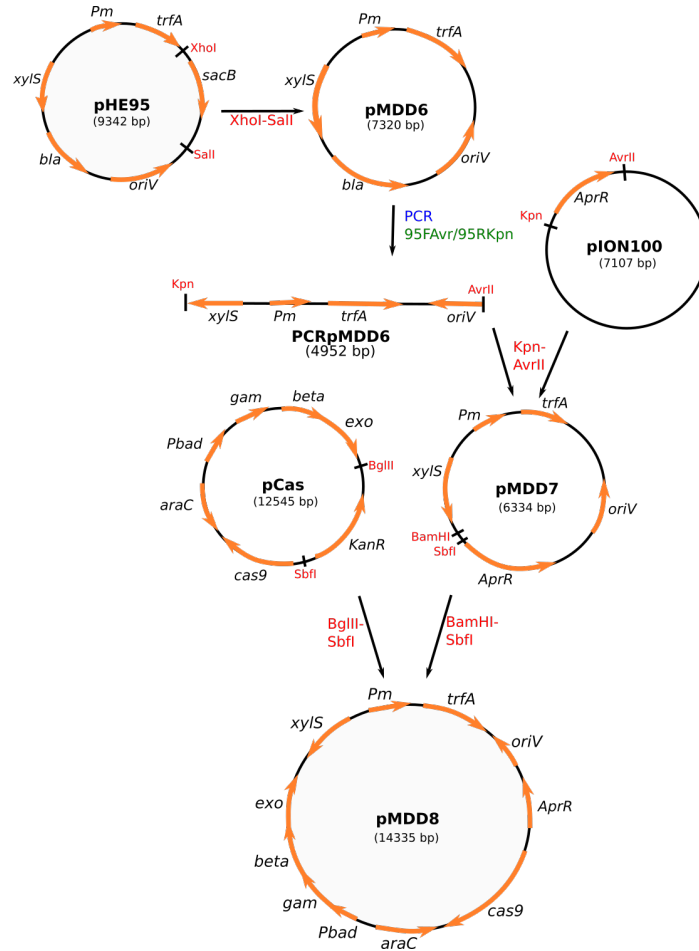
**Figure 4.1: Cloning strategy to produce a vector for the gene *cas9*.** The figure shows the cloning procedure to construct the final Cas9-encoding vector pMDD8 based on the RK2 replicon, regulated by the XylS/*Pm* regulator/promoter system. Restriction enzymes with cleavage sites are colored red, methods are colored blue, and primers are colored green.

The final Cas9-encoding vector was constructed from the RK2 replicon pMDD7. The replication of the plasmid was controlled by the XylS/*Pm* regulator/promoter system and contained an apramycin resistance gene for selection. *cas9* and the genes encoding the $\lambda$-Red recombineering system that is needed to aid in homology-directed repair (HDR), was isolated from BglII-SbfI restricted pCas (8041 bp). The three genes in the recombineering system, *gam-beta-exo*, are under the control of the $P_{bad}$ promoter regulated by the activator AraC with L-arabinose as inducer. The genes were cloned into pMDD7 digested with BamHI-SbfI (6294 bp) to yield the final Cas9 vector. The produced plasmid from four colonies of the transformants were digested with AvrII-BamHI, and all colonies yielded the expected fragments of 2002, 5135 and 7198 bp (Figure 4.2). Colony 1 was designated pMDD8.
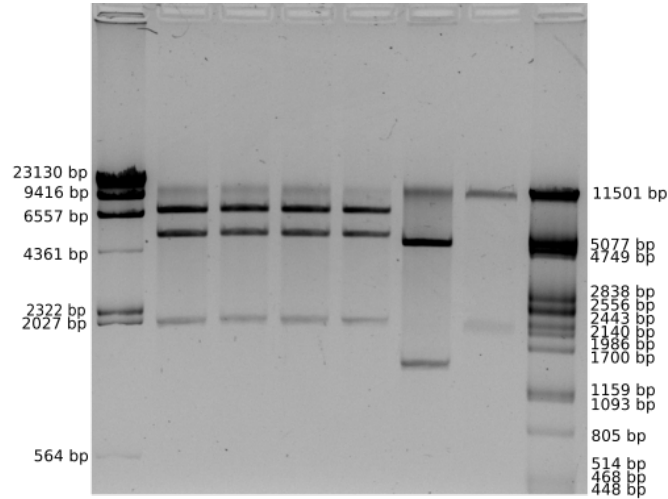
**Figure 4.2: Verification of pMDD8 with gel electrophoresis.** Lane 1 shows Lambda HindIII ladder with known molecular sizes, lane 2-5 shows plasmids from four colonies (1-4) of the transformants cleaved with BamHI and AvrII, lane 6 shows BamHI-AvrII digested pMDD7 (1385 and 4949 bp), lane 7 shows BamHI-AvrII digested pCas, and lane 8 is Lambda PstI ladder with known molecular sizes. The fragments of pMDD8 had the expected sizes, 2002, 5135, and 7198 bp. Colony 1 was designated pMDD8.

## 4.2 Construction of a vector for the gRNA

The last gene needed for CRISPR type II to function is the gene encoding the gRNA. The gRNA needs to have a binding site for Cas9 (a scaffold) and the sequence homologous to the target so that Cas9 can be guided to its target. Figure 4.3 shows the cloning procedure. Plasmid pMDD9 was cloned during the subject TBT4500 - Biotechnology, specialization project. *sacB* was isolated from pAGL2 and cloned into pTargetF, which carried the genes for the gRNA scaffold and spectinomycin resistance. The new plasmid was designated pMDD9. In the final step the genes *sacB*, *SpcR* and *SmR* (giving spectinomycin and streptomycin resistance, respectively) were cleaved out of pMDD9 and inserted into vector pBBR1MCS-2 to yield pMDD10.
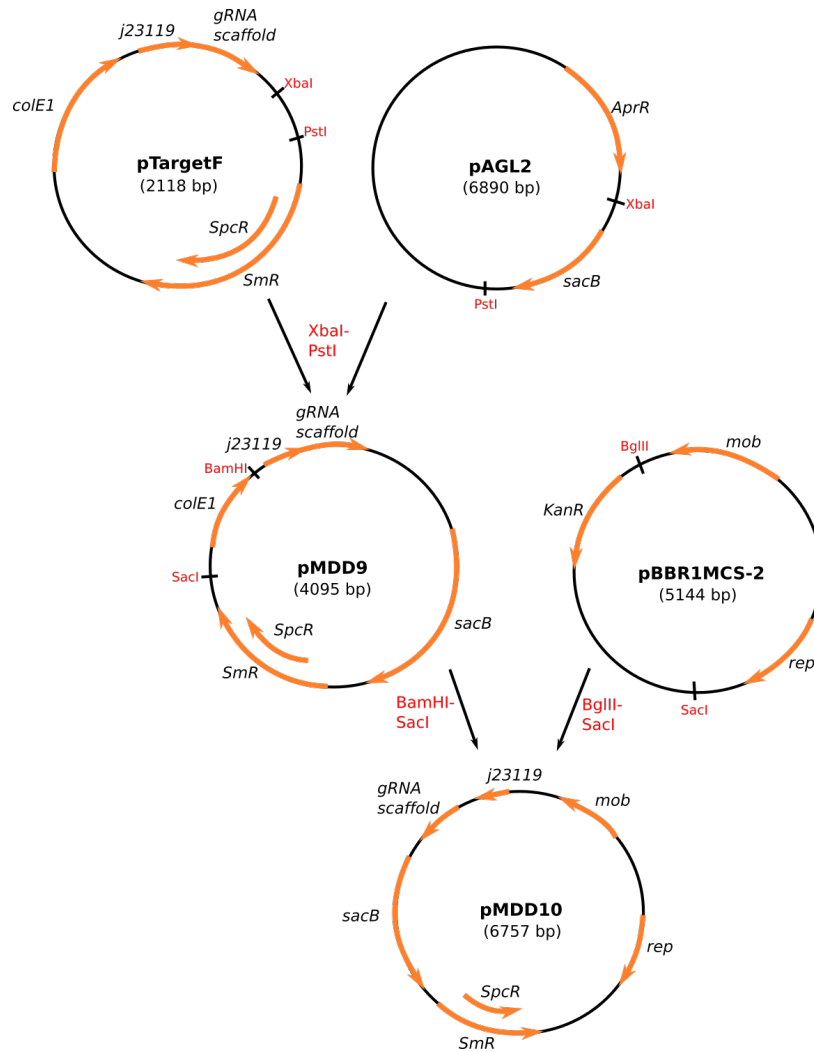
**Figure 4.3: Flow-diagram showing the strategy to clone a gRNA vector.** Construction of the unspecific gRNA-encoding vector based on the pBBR1 replicon. The specific 20N sequence was later cloned into the vector. Restriction enzymes with cleavage sites are colored red.

The pBBR1 replicon from the broad host-range-vector pBBR1MCS-2 was used in the gRNA vector to control replication and conjugal transfer with the genes *rep* and *mob*, respectively [73]. The gRNA scaffold and its promoter *j23119*, *sacB*, and *SmR* and *SpcR* were isolated from BamHI-SacI restricted pMDD9 (3120 bp), and cloned into BglII-SacI digested pBBR1MCS-2 (3645 bp). The new plasmid from four colonies of the transformants, were digested with EcoRI-NotI. Figure 4.4 shows the fragments of sizes 678, 889, and 5190 bp produced when the new plasmid was cleaved with EcoRI-NotI. The four colonies (lane 2-5) all show the expected fragment sizes, and colony 1 was designated pMDD10.
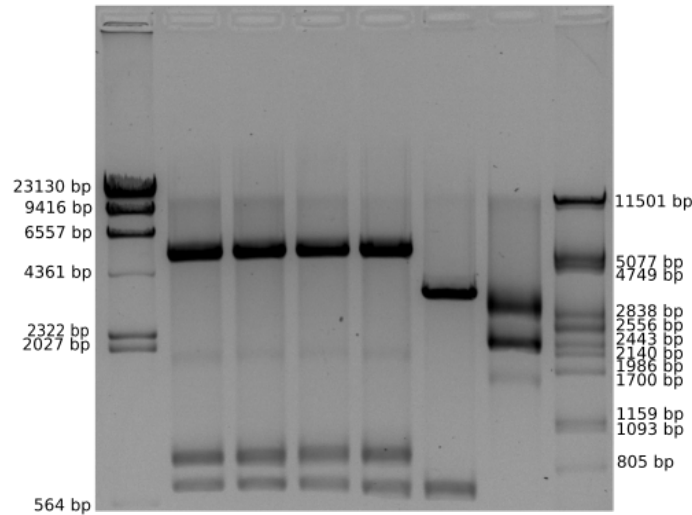
**Figure 4.4: Verification of pMDD10 with gel electrophoresis.** Lane 1 shows Lambda HindIII ladder with known molecular sizes, lane 2-5 show plasmids from four colonies (1-4) of the transformants cut with EcoRI-NotI, lane 6 shows digested EcoRI-NotI pMDD9 (678 and 3417 bp), lane 7 shows EcoRI-NotI digested pBBR1MCS-2 (37, 2194, and 2913 bp)), and lane 7 is Lambda PstI ladder with known molecular standards. All four colonies show the expected fragments of sizes 678, 889, and 5190 bp. Colony 1 was designated pMDD10.

Initially, it was planned to clone the N20 sequences specific for *uidA* and *mucA* directly into pMDD10 by PCR, with primer pairs guidA1441F/guidA1441R and gmucA3651F/ gmucA3651R, respectively. The N20 sequences were chosen based on the theory described in Section 1.4.1. The 20 nucleotide sequence 5'-GAAGCGCAGCCGTAGGCCAG-3' directly downstream of the 5'-GGC-3' PAM site of the coding strand in *uidA*, was chosen as target sequence. It was planned to make a mutation template with a deletion in this region. The N20 3'-GGCGGCTTCGGTGACTGTCG-5' was chosen as the target sequence directly upstream of 3'-CGG-5' in the template strand of *mucA*. This sequence was chosen because there already existed a mutation template for *mucA* with a deletion in this region. Cas9 can target the correct sequence independent of which direction the N20 sequence has, or if it is on the coding or template strand, as long as it is homologous to a sequence next to the PAM site. The forward primers were designed to contain the N20 sequence and a sequence homologous to the constitutive *j23119* promoter in a 5'-to-3' direction. The reverse primers contained the complementary of the N20 sequence and a sequence homologous to the *gRNA scaffold* in a 5'-to-3' direction. Accordingly, the original sequence between the promoter and the scaffold would be replaced by the

37

specific N20 sequences when running PCR. For some reason, the reaction did not result in any product, neither for the gRNA vector with *uidA* nor *mucA* as the target. Thus, a different approach was tested.

A NcoI restriction cutting site was introduced into pMDD10, upstream of the gRNA scaffold by PCR, as shown in Figure 4.5. This was done to make it possible to cleave out the region between the *j23119* promoter (SpeI site) and *gRNA scaffold* (NcoI site), and replace it with the specific N20 sequences.
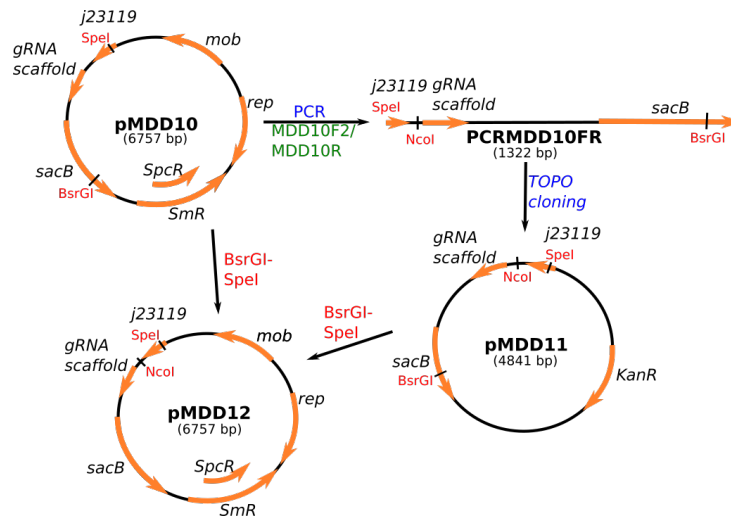


**Figure 4.5:** **Flow-diagram illustrating how a NcoI restriction cleavage site was introduced between the *j23119* promoter and *gRNA scaffold* in pMDD10 by PCR.** The PCR fragment was cloned into a TOPO vector yielding pMDD11. A 1269 bp SpeI-BsrG1 fragment from pMDD11 replaced an identical fragment from pMDD10, except that the new fragment included a NcoI site. Restriction enzymes with cleavage sites are colored red, methods are colored blue, and primers are colored green.

Two PCR were run in parallel with the primer pairs mDD10F/mDD10R and mDD10F2/ mDD10R. mDD10F and mDD10F2 are identical except that mDD10F2 are nine nucleotides shorter than pMDD10F. The two parallels were run with the different primers, to increase the chances of producing the wanted fragment. Both mDD10F and mDD10F2 are complementary to the end of the promoter sequence *j23119* upstream of the gRNA scaffold sequence. The primers would introduce the NcoI site downstream of the SpeI site. mDD10R binds to a region in *sacB* downstream of a BsrG1 site. Both reactions with the two different primer pairs, yielded a product. It was chosen to continue with the product of the reaction with primer pair mDD10F/mDD10R. The PCR product (1322 bp) was cloned into a TOPO vector. However, when plasmids from colonies of the transformants

were controlled, they proved to be wrong. Several colonies were tested but none had the desired insert. This suggested that there was something wrong with primer mDD10F. Therefore, the PCR fragment produced with primer pair mDD10F2/mDD10R, used instead and cloned into a TOPO vector. Twelve colonies of the new plasmid were digested with NcoI to see if the desired restriction site had been introduced. Figure 4.6 shows in lane 2-13 the produced plasmid fragments from twelve colonies of the transformants. Depending on which direction the PCR product had in the TOPO vector the resulting fragment sizes could be one of two possibilities; A) 439, 1641, and 2761 bp, or B) 439, 1487, and 2915 bp.
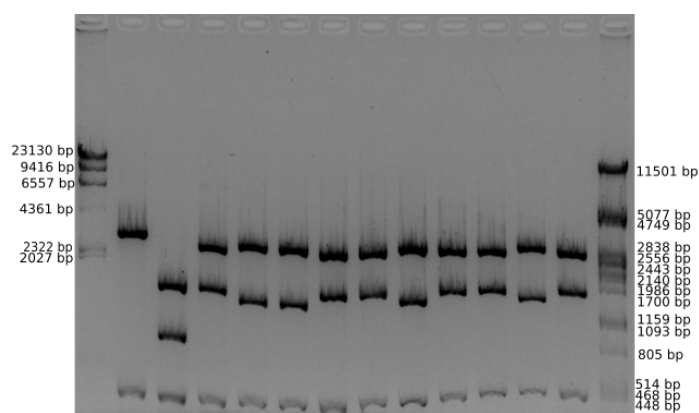


**Figure 4.6: Verification of pMDD11 with gel electrophoresis.** Lane 1 shows Lambda HindIII ladder with known molecular sizes, lane 2-13 show plasmids from twelve colonies (1-12) of the transformants cut with NcoI, lane 14 shows Lambda PstI ladder with known molecular standards. Colony 1 and 2 does not produce the expected fragments. Colony 3, 6, 7, 9, 10, and 12 show the sizes 439, 1641, and 2761 bp (option A), while colony 4, 5, 8, and 11 show the sizes 439, 1487, and 2915 bp (option B).

The gel image suggests that colony 3, 6, 7, 9, 10, and 12 could have orientation option A, and colony 4, 5, 8, and 11 have orientation B. Since there only was a small difference in the fragment sizes depending on which direction the PCR product had, colony 3-12 were also digested with SpeI. Figure 4.7 shows the fragments of the digested plasmids from the ten colonies of the transformants. The fragment sizes would vary depending on which orientation the PCR insert had in the TOPO vector. One direction would yield the fragment sizes of 37 and 4808 bp (option A), while the other 1355 and 3488 bp (option B).
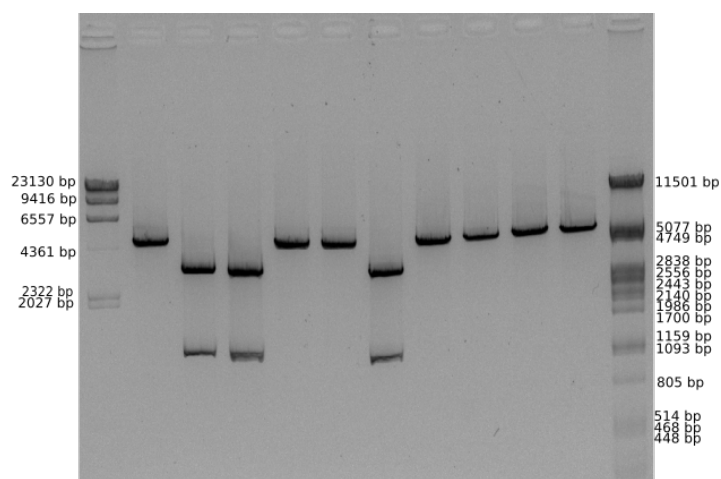
**Figure 4.7: Verification of pMDD11 with gel electrophoresis.** Lane 1 shows the Lambda HindIII ladder with known molecular sizes, lane 2-11 show plasmids from ten colonies (3-12) of the transformants cut with SpeI, lane 12 shows Lambda PstI ladder with known molecular standards. Colony 3, 6, 7, 9, 10, 11, and 12 show the sizes 37 and 4808 bp (option A), while colony 4, 5, and 8 show the sizes 1355 and 3488 bp (option B). Colony 5 was designated pMDD11.

Figure 4.7 clearly shows that colony 3, 6, 7, 9, 10, 11, and 12 have orientation option A, and only colony 4, 5, and 8 have orientation B. Colonies 3-8 were sequenced with primers M13F and M13R. The sequencing of colony 5 showed that it was correct, and this plasmid was designated pMDD11 (Appendix F).

pMDD10 and pMDD11 were restricted with SpeI-BsrG1 resulting in a 5488 bp vector and a 1269 bp insert molecule, respectively. The 5488 bp fragment of pMDD10 and the 1269 bp fragment of pMDD11 were ligated to form a new plasmid, which now included a NcoI site compared to pMDD10. The new plasmid was digested with NcoI-HindIII to control that the fragments were of the correct sizes (74, 113, 1902, and 4668 bp) and that the NcoI site was in place. Figure 4.8 shows the resulting fragments from six colonies of the transformants. Colony 2 was designated pMDD12.
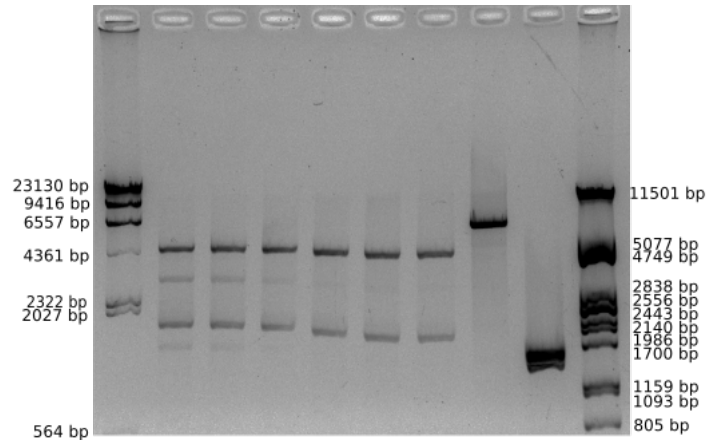
**Figure 4.8: Verification of pMDD12 with gel electrophoresis.** Lane 1 shows Lambda HindIII ladder with known molecular sizes, lane 2-7 show plasmids from six colonies (1-6) of the transformants cut with NcoI-HindIII, lane 8 shows NcoI-HindIII digested pMDD10 (74, 113, and 6570 bp), lane 9 shows NcoI-HindIII digested pMDD11 (439, 1356, 1487, and 1559 bp), and lane 10 is Lambda PstI ladder with known molecular standards. All six colonies show the expected fragments of sizes 74, 113, 1902, and 4668 bp. Colony 2 was designated pMDD12.

### 4.2.1 gRNA vector with N20 sequence targeting and mutating the gene *uidA*

Blue-white plate screening will be used to calculate mutation frequencies. An *A. vinelandii* strain containing a gene *uidA* encoding the enzyme $\beta$-glucuronidase will be one of the targets for CRISPR. When the chemical compound X-Gluc is added to the growth medium the enzyme will cleave it and produce a blue product. *uidA* will in this strain be the target gene of Cas9. Binding of Cas9 to the sequence will result in cleavage and interruption of *uidA*. If the nick is repaired through HDR with the aid of a mutation template, white colonies that can easily be spotted and quantified, will be produced. The cloning procedure to construct a gRNA vector that targets *uidA* is illustrated in Figure 4.9.
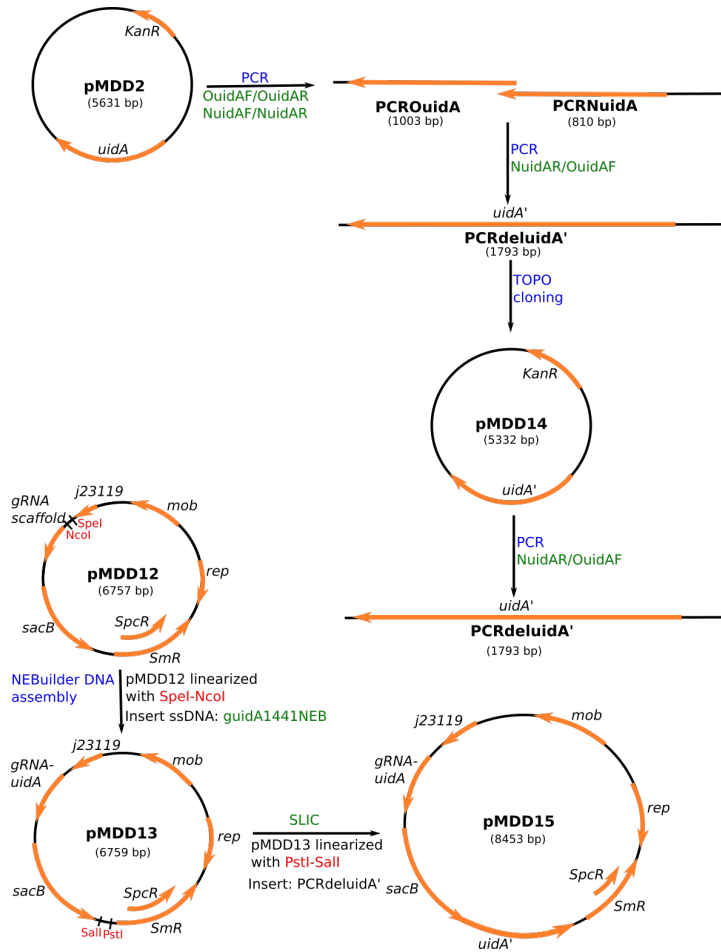
**Figure 4.9: Flow-diagram illustrating the cloning procedure for the gRNA vector targeting *uidA*.** pMDD13 was cloned using NEBuilder DNA Assembly, utilizing the homologous ends of the oligomer guidA1441NEB to the ends of the SpeI-NcoI linearized pMDD12 vector. A mutation template, *uidA'*, was cloned from pMDD2 by PCR and introduced into pMDD13 with SLIC. Restriction enzymes with cleavage sites are colored red, methods are colored blue, and primers and oligomers are colored green.

The N20 sequence 5'-GAAGCGCAGCCGTAGGCCAG-3' directly downstream of the 5'-GGC-3' PAM site of the coding strand in *uidA*, was as mentioned, the target sequence. A spacer homologous to the N20 sequence was cloned between the SpeI-NcoI linearized unspecific gRNA vector pMDD12, using NEBuilder HiFi DNA Assembly with a single-stranded oligomer guidA1441NEB. The oligomer consists of 33 nucleotides complementary to the region around the *j23119* promoter, the N20 sequence, and 25 nucleotides complementary to the region around the *gRNA scaffold*. The new plasmid would have lost the NcoI restriction cleavage site, so it was digested with NcoI-HindIII to control that it was correct. Figure 4.10 shows that colony 3 and 6 had lost the NcoI site, and produced the expected fragments of sizes 74, 113, and 6572 bp.
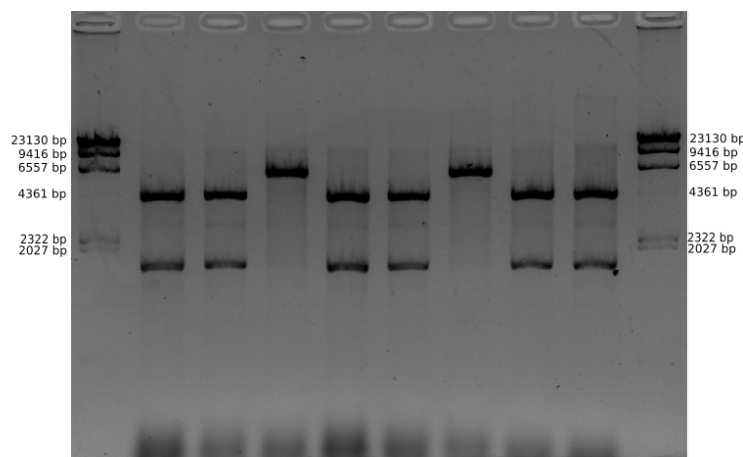
**Figure 4.10: Verification of pMDD13 with gel electrophoresis.** Lane 1 and 10 show
the Lambda HindIII ladder with known molecular sizes, and lane 2-9 show plas-
mids from eight colonies (1-8) of the transformants cut with NcoI-HindIII. The
digestion of pMDD12 is not shown in the picture but would result after cutting
with NcoI-HindIII in fragment sizes of 74, 113, 1902, and 4668 bp (see Figure
4.8). Colony 3 and 6 had the expected fragments of sizes 74, 113, and 6572 bp,
although the 74 bp and 113 bp fragment are too small to show on the gel. Colony
1, 2, 4, 5, 7, and 8 show the fragment sizes of NcoI-HindIII restricted pMDD12
and are therefore not correct. Colony 6 was designated pMDD13.

The plasmids of colony 3 and 6 yielded the expected fragments of 74, 113, and 6572 bp
when restricted with NcoI-HindIII, showing that the N20 sequence successfully had been
implemented between the *j23119* promoter and the gRNA scaffold sequence. Both
colonies were sent to sequencing using the primer uidAF. The sequencing result of colony
6, which was designated pMDD13, is presented in Appendix G, and shows that the spe-
cific N20 sequence homologous to the target in *uidA* had been cloned between the *j23119*
promoter and the gRNA scaffold without any error.

A mutation template for *uidA* was constructed with about 8 kb and 10 kb long stretches
upstream and downstream of the N20 target in *uidA* to introduce mutations when used as
a template in HDR. Two separate PCR reactions were run with two sets of primer pairs
- OuidAF/OuidAR and NuidAF/NuidAR. Primer OuidAR is homologous to the N20
sequence except for 9 nucleotides that are different and thereby will introduce mutations
in the sequence. Primer NuidAF has 20 nucleotides homologous to OuidAR. The resulting
PCR fragments, PCRouidA and PCRnouidA, had a slight overlap and could, therefore,
be used as templates in a second PCR reaction. Equal molar amounts of both templates
were used, and the primers OuidAF and NuidAR. The PCR product was cloned into a

TOPO vector, and the plasmids from four colonies of the transformants, were digested with EcoRI to see if it produced the expected fragments of 1811 and 3501 bp. Figure 4.11 shows that colony 1, 3, and 4 all got the insert, while colony 2 does not.
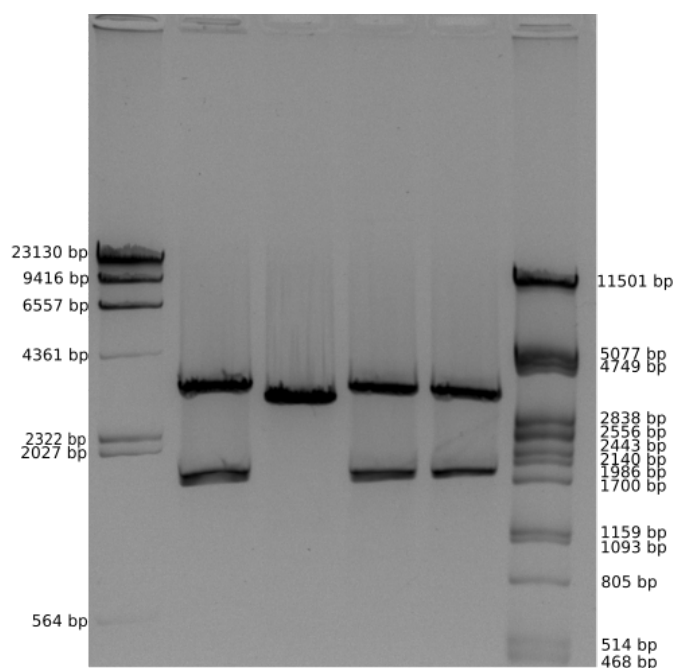


**Figure 4.11: Verification of pMDD14 with gel electrophoresis.** Lane 1 shows the Lambda HindIII ladder with known molecular sizes, lane 2-5 show plasmids from four colonies (1-4) of the transformants cut with EcoRI, lane 6 is Lambda PstI ladder with known molecular standards. Colony 1, 3, and 4 show the expected fragments of sizes 1811 and 3501 bp. Colony 2 does not appear to have the PCR insert. Colony 1 was designated pMDD14.

Colony 1 was designated pMDD14 and sent to sequencing to control that no error had occurred during PCR. The sequencing results (attached in Appendix H) showed that the plasmid had the correct sequence. The mutation template was cloned from pMDD14 by PCR with primers OuidAF and NuidAR, and SLIC cloned into PstI-SalI restricted pMDD13. Plasmids from eight colonies of the transformants, were digested with HindIII. Figure 4.12 shows that only colonies 1 and 6 produced the expected fragments of sizes 113, 1768, and 6572 bp, while the other colonies do not appear to have the insert. Colony 1 was designated pMDD15.
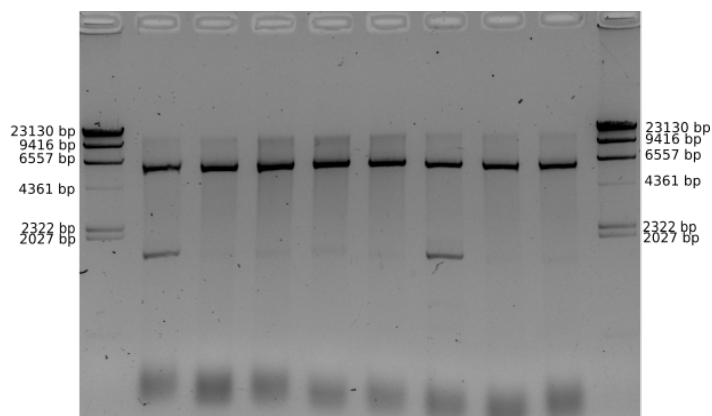
**Figure 4.12: Verification of pMDD15 with gel electrophoresis.** Lane 1 and 10 show the Lambda HindIII ladder with known molecular sizes, and lanes 2-9 show plasmids from eight colonies (1-8) of the transformants cut with HindIII. Colony 1 and 6 show the expected fragments of sizes 113, 1768, and 6572 bp. The other colonies do not appear to have the PCR insert.

Colony 1 and 6 was sent to sequencing. The sequencing results of colony 1, which was designated pMDD15, showed that the plasmid had the correct sequence and are attached in Appendix I.

### 4.2.2 gRNA vector with N20 sequence targeting and mutating the gene *mucA*

Disruption of *mucA* will lead to an increase in AlgU, and thereby an alginate overproducing strain. Such strains grow more slowly, hence they have been proven difficult to obtain (Ertesvåg, pers. corr.). Figure 4.13 illustrates the cloning procedure to produce a gRNA vector that will target *mucA*.
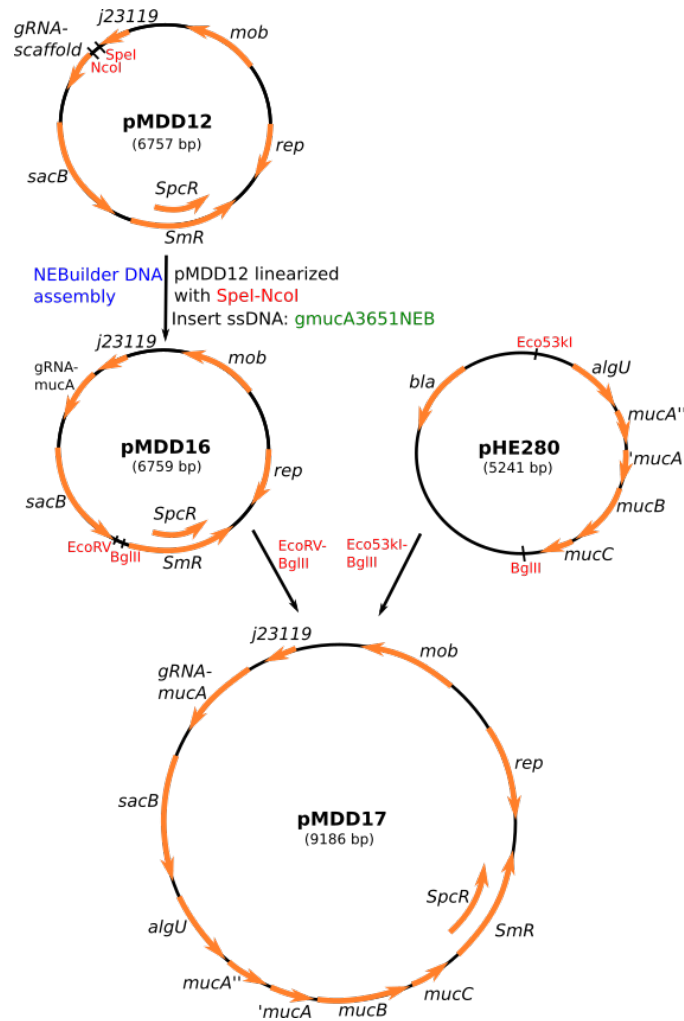
**Figure 4.13: Flow-diagram illustrating the cloning procedure to construct a gRNA vector that targets *mucA*.** pMDD16 was cloned using NEBuilder DNA Assembly, utilizing the homologous ends of the oligomer gmucA3651NEB to the ends of the SpeI-NcoI linearized pMDD12 vector. A mutation template was isolated from Eco53kI-BglII restricted pHE280, and cloned into EcoRV-BglII restricted pMDD16, yielding pMDD17. Restriction enzymes with cleavage sites are colored red, methods are colored blue, and oligomers are colored green.

As already mentioned, the N20 3'-GGCGGCTTCGGTGACTGTCG-5' was chosen as the target sequence directly upstream of 3'-CGG-5' in the template strand of *mucA*. The oligomer gmucA3651NEB was used to clone the N20 sequence into pMDD12. As for guidA1441NEB, the oligomer consists of 33 nucleotides complementary to the region around the *j23119* promoter, the N20 sequence, and 25 nucleotides complementary to the region around the *gRNA scaffold*. The N20 sequence was cloned into SpeI-NcoI restricted pMDD12 with NEBuilder DNA Assembly. Table 4.14 show plasmids from eight colonies of the transformants, digested with NcoI-HindIII.
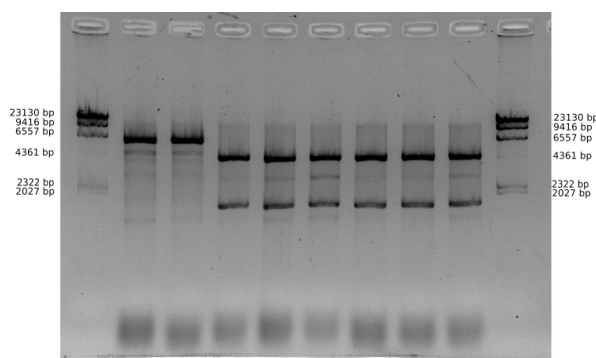
**Figure 4.14: Verification of pMDD16 with gel electrophoresis.** Lane 1 and 10 show the Lambda HindIII ladder with known molecular sizes, and lanes 2-9 show plasmids from eight colonies (1-8) of the transformants cut with NcoI-HindIII. The digestion of pMDD12 is not shown but would result after cutting with NcoI-HindIII in fragment sizes of 74, 113, 1902, and 4668 bp. Colony 1 and 2 yielded the expected fragments of sizes 74, 113, and 6572 bp, although the 74 bp and 113 bp fragment are too small to show on the gel. Colony 3 to 8 shows the fragment sizes of NcoI-HindIII digested pMDD12 and are therefore not correct. Colony 1 was designated pMDD16.

Colony 1 and 2 both yielded the expected fragments of 74, 113, and 6572 bp when restricted with NcoI-HindIII, showing that the N20 sequence successfully had been implemented between the *j23119* promoter and the gRNA scaffold sequence. Both colonies were sent to sequencing with primer uidAF. The sequencing result of colony 1, which was designated pMDD16, is presented in Appendix J and show that the plasmid had the correct sequence.

The mutation template was isolated from pHE280 digested with Eco53kI-BglII (2511 bp) and cloned into pMDD16 cleaved with EcoRV-BglII (6675 bp). Plasmids from four colonies of the transformants, were digested with HindIII. The resulting fragments are shown in Figure 4.15. The final gRNA vector with *mucA* as the target was designated pMDD17 (colony 1).
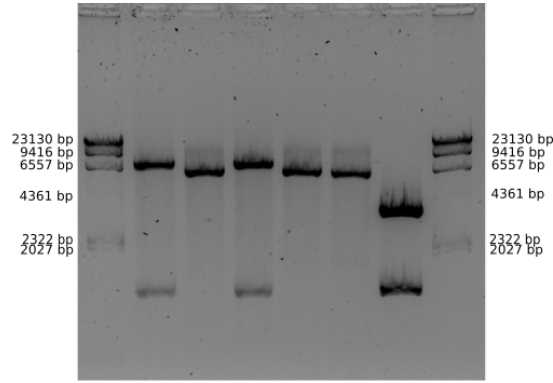
**Figure 4.15: Verification of pMDD17 with gel electrophoresis.** Lane 1 and 8 show the Lambda HindIII ladder with known molecular sizes, and lanes 2-5 show plasmids from four colonies (1-4) of the transformants cut with HindIII. Lane 6 shows HindIII digested pMDD16 (74, 113, and 6772 bp), and lane 7 shows HindIII digested pHE280 (1458 and 3785 bp). Colony 1 and 3 had the expected fragments of sizes 148, 1458, and 7580 bp, although the 148 bp fragment is too small to show on the gel. Colony 2 to 4 show the fragment sizes of HindIII digested pMDD16 and are therefore not correct. Colony 1 was designated pMDD17.

## 4.3   Validation of elements of the planned plasmids

Several elements of the wanted plasmids were tested in parallel with the construction of them. Precursors of the final plasmids were used to test stability, selectability, and viability. pMDD7 was used to test these elements of the final Cas9-encoding vector, which at this time had not been cloned, while pMDD10 was used for the gRNA vectors as they neither were cloned at this stage.

### 4.3.1   The $P_{bad}$ promoter works in *A. vinelandii*

If arabinose is to work as an inducer for the promoter $P_{bad}$, it is essential that *A. vinelandii* does not utilize the added arabinose as a carbon source. If the bacterium consumes the sugar, arabinose will not induce the promoter. Thus, the genes controlled by $P_{bad}$ will not be transcribed.

To test if *A. vinelandii* can grow on arabinose, the bacterium was cultivated for 2-3 days in Burks medium at 30 °C. The culture was plated on Burks agar (BA) plates and incubated for 2-3 days. Bacteria grown on the normal BA with glucose were transferred

to a BA plate with 6 g/L arabinose. There was very little growth on the plates and only after several days, which implies that arabinose is not a good carbon-source for the bacterium, and that it could work as an inducer if it were imported.

The plasmid pSB-B1I carrying the $P_{bad}$ promoter was used to control that the promoter worked in *A. vinelandii*. The promoter regulates the gene encoding luciferase. Luciferase catalyzes the formation of oxyluciferin from the oxidation of luciferin, by using ATP·Mg$^{2+}$ as a co-substrate. A luciferase assay which contains luciferin can be used to provide the substrate, and a luminometer can measure the light converted from the chemical energy produced in the reaction. pSB-B1I was first isolated from DH5$\alpha$ and transformed into S17-1. pSB-B1I was transferred from *E. coli* S17-1 to *A. vinelandii* by conjugation as described in Section 3.2. BA plates with kanamycin (1, 2, and 5 $\mu$g/mL) and m-toluate (0.5 mM) were used to select for the conjugated plasmid. The low concentration of kanamycin was necessary because *A. vinelandii* is highly sensitive to kanamycin. The plates were incubated at 30 °C for 3-5 days.

A pre-culture of the conjugates grown on the plates was prepared for the luciferase assay. It was transferred to Burks medium with kanamycin (5 $\mu$g/mL) and m-toluate (0.5 mM) and cultivated for 1-2 days at 30 °C. A pre-culture of *A. vinelandii* was used as a control. 1% of the pre-cultures were transferred to new media and inoculated over the night until OD$_{600}$ reached 0.2-0.4. Table 4.1 shows the measured OD$_{600}$ for *A. vinelandii* and *A. vinelandii* (pSB-B1I).

**Table 4.1:** OD$_{600}$ was measured for the 1% pre-culture of *A. vinelandii* (pSB-B1I) and the control *A. vinelandii*.

| **Sample** | **OD$_{600}$** |
|---|---|
| *A. vinelandii* | 0.288 |
| *A. vinelandii* (pSB-B1I) | 0.238 |

Then, the culture was induced by adding 6 g/L arabinose to the culture. Arabinose would induce the expression of the luciferase gene since it is under the control of $P_{bad}$. The next day, three samples with 90 $\mu$L of the induced culture was mixed with 10 $\mu$L of 1 M K$_2$HPO$_4$, 20 mM EDTA (pH 7.8). The mixture was frozen at $-80$ °C to kill the cells and release luciferase. Controls were made - one sample of *A. vinelandii* without plasmid pSB-1I, and one with the plasmid that had not been induced with arabinose.

The cell mixtures were thawed at room temperature in a water bath. 300 $\mu$L of freshly prepared lysis mix (1 X CCLR, 2.5 mg/mL BSA and 1.25 mg/mL lysozyme) was added to the cell mixture, and incubated at room temperature for a minimum of ten minutes. 100 $\mu$L of Luciferase assay reagent was aliquoted into five Eppendorf tubes. The luminometer had a two-second delay and a 10-second measurement. 20 $\mu$L of the cell lysate was added and mixed with the Luciferase assay reagent containing the substrate luciferin. A luminometer measured the activity of luciferase for all the five samples - one without the conjugated plasmid, one uninduced, and three induced. The measurements of the luciferase activity is presented in Table 4.2. The sample without the plasmid was used as a control and gave a reading of 94. The uninduced sample with conjugated pSB-B1I, gave a reading of 32,019. The three arabinose induced samples gave a reading of 26,020,568, 26,633,698, and 42,067,032. These high readings show that $P_{bad}$ works in *A. vinelandii*, and that it was inducible.

**Table 4.2:** The luciferase activity measured for *A. vinelandii*, uninduced *A. vinelandii* (pSB-B1I), and three parallels of induced *A. vinelandii* (pSB-B1I).

| Sample | Reading |
|---|---|
| *A. vinelandii* | 94 |
| *A. vinelandii* (pSB-B1I) | 32,019 |
| (1) *A. vinelandii* (pSB-B1I) induced | 26,020,568 |
| (2) *A. vinelandii* (pSB-B1I) induced | 26,638,698 |
| (3) *A. vinelandii* (pSB-B1I) induced | 42,067,032 |

### 4.3.2 The apramycin resistance gene does not give resistance against spectinomycin

The Cas9-encoding vector (pMDD8) and the gRNA-encoding vectors (pMDD13, pMDD15, pMDD16, and pMDD17), have different antibiotic selection markers (apramycin and spectinomycin, respectively) so that it is possible to select for each of the plasmids. After conjugation of both plasmids into *A. vinelandii*, the bacteria will be plated on BA plates with apramycin and spectinomycin. If one of the antibiotic resistance genes gives resistance to the other antibiotic as well, it will not be possible to select for both of them. The apramycin resistance gene used may give resistance to kanamycin and gentamycin. Since spectinomycin belongs to the same group of antibiotics, it was controlled that the

apramycin resistance gene does not give resistance to spectinomycin. pMDD7, carrying the apramycin resistance gene, was conjugated into *A. vinelandii.* The conjugates were selected for on BA plates with apramycin (25 $\mu$g/mL) and m-toluate (0.5 mM), to make sure that pMDD7 successfully had been transferred to *A. vinelandii.* The plates were incubated for 3-5 days at 30 °C before the bacteria were transferred to BA plates with spectinomycin (20 $\mu$g/mL) and m-toluate (0.5 mM). The bacterium was not able to grow on the plates with spectinomycin, which shows that the apramycin resistance gene does not give resistance to spectinomycin.

### 4.3.3   The Cas9 vector is curable

The plasmids must be curable so that when desired mutations have happened, the plasmids can be removed to avoid further off-target mutations. It is also critical that they are curable if the N20 sequence of the gRNA vector must be exchanged to alter the target. pMDD8 contains a different curing component than pMDD13, pMDD15, pMDD16, and pMDD17.

The stability of pMDD8 was controlled using its precursor, pMDD7, which also has the RK2 replicon and the *xylS/Pm* cassette. The plasmid is dependent on the inducer m-toluate to replicate. The plasmid will not replicate in the absence of the inducer, and hence, *A. vinelandii* will lose the plasmid. To control that this is the case, *A. vinelandii* with conjugated pMDD7 was grown for three days in Burks media without apramycin and m-toluate. A dilution series was made from the culture ($10^0$, $10^{-1}$, $10^{-2}$, $10^{-3}$, $10^{-4}$, $10^{-5}$ and $10^{-6}$). Each dilution was plated on BA plates with apramycin (25 $\mu$g/mL) and m-toluate (0.5 mM). The three lowest dilutions ($10^{-4}$, $10^{-5}$ and $10^{-6}$) were plated on BA plates as well. The plates with apramycin and m-toluate would select for the bacteria that had not lost its plasmid, while the BA plates would select for all the bacteria. The number of colony forming units on the BA plates were $8.9 \cdot 10^4$ cfu/mL, while the plates with apramycin and m-toluate had no growth, which shows that the plasmid is curable.

After completion of the final Cas9-encoding vector pMDD8, it was conjugated into *A. vinelandii.* pMDD8 was about 8 kb larger than pMDD7 (14335 bp versus 6334 bp). A significant reduction in the conjugation frequency was observed with only 0.4 cfu/mL. The

conjugation of pMDD8 was only tested once so it is unclear if the decrease in frequency should be expected each time, or if something went wrong in the conjugation.

### 4.3.4   The gRNA vector could not be transferred to *A. vinelandii*

The gRNA vectors, pMDD13, pMDD15, pMDD16, and pMDD17, have a different curing system than pMDD8 - the suicide gene *sacB* which expression is lethal in the presence of 5% sucrose. pMDD10 was used to control the stability of the vectors. Since pMDD10 carry a resistance gene against spectinomycin, the plasmid could not be transferred directly to *A. vinelandii* by conjugation with S17-1 as the donor.

Electransformation with electrocompetent *A. vinelandii* was first tested as described in Section 3.3. Electrotransformation was unsuccessful, and a different approach was tried out. The plasmid was first electrotransformed into *E. coli* DH5$\alpha$ (pTA19) as described in Section 3.1.5. *E. coli* DH5$\alpha$ (pTA19) can act as a donor in conjugation because the plasmid pTA19 carries the essential genes (see Section 3.2 for more information). pMDD10 was first electrotransformed into *E. coli* DH5$\alpha$ (pTA19), and then conjugated into *A. vinelandii.*

A dilution series was made ($10^0$, $10^{-1}$, $10^{-2}$, $10^{-4}$, and $10^{-6}$) and each dilution plated on BA plates with spectinomycin (20 $\mu$g/mL). The plates were incubated for 5-7 days at 30 °C. No growth was observed after seven days. The experiment was repeated several times with pMDD10 and also with the final plasmids (pMDD13, pMDD15, pMDD16, and pMDD17). The conjugation was unsuccessful each time.

### 4.3.5   *E. coli* DH5$\alpha$ (pTA19) appears not to function as a donor cell of pMDD7 in conjugation

It would be interesting to test if *E. coli* DH5$\alpha$ (pTA19) can function as a donor strain in conjugation of pMDD7 and pMDD8 into *A. vinelandii.* If conjugation of pMDD7 and pMDD8 works, it can be excluded that DH5$\alpha$ (pTA19) does not work as a donor strain, and that something is otherwise wrong with the gRNA vectors.

pMDD7 and pMDD8 were electroporated into *E. coli* DH5α (pTA19) and selected for on LA plates with ampicillin (100 μg/mL), apramycin (50 μg/mL), and m-toluate (0.5 mM). Only pMDD7 was successfully electroporated, although the same concentration was used for both plasmids (about 400 ng/μL). A few colonies of pMDD8 were visible after incubation for two days, but these were not used further since it was unsure if they were correct. pMDD7 was conjugated into *A. vinelandii* and selected for on BA plates with apramycin (25 μg/mL) and m-toluate (0.5 mM). After incubation for seven days the plates showed no colonies. The experiment was only performed once.

## 4.4   Construction of *A. vinelandii* ATCC12518 *mucA⁻* strain with homologous recombination

Plasmid pHE437 containing a mutation template for *mucA* was conjugated into *A. vinelandii* to mutagenize the bacteria by homologous recombination and create a strain with a mixture of wild-type and mutant chromosomes. *A. vinelandii* (pHE437) was selected for on BA plates with 10 μg/mL tetracycline and 0.5 mM m-toluate. After incubating the plates at 30 °C for five days, a transconjugant colony was picked and transferred to Burks medium with tetracycline and no m-toluate, and grown for three days at 30 °C. Since m-toluate is absent from the medium, the plasmid is forced to integrate into the chromosome. For the plasmid to integrate into all chromosomes of the bacterium, 1% of the culture first culture was transferred to new medium several times. After the fifth time, a sixth cultivation was prepared without antibiotics to allow for a second recombination to take place. To select against chromosomes were the entire plasmid vector was inserted, a dilution series was made from the cultivation and plated on BA plates with 5% sucrose. Chromosomes with the vector would have the gene *sacB* and therefore not be able to grow on sucrose. Colonies were then transferred from the BA plates with sucrose to BA plates with and without tetracycline. The desired colonies grew on the BA plate but not the tetracycline plate, since they were sensitive to tetracycline.

Initially, it was not planned to construct this *A. vinelandii* ATCC 12518 *mucA⁻* strain, because it was thought that it already existed. Due to the fact that it had been discarded, the strain had to be constructed in this work. However, the construction was started at

the end of the project, and due to limited amount of time on this master's thesis, the experiment was ended at this point. The tetracycline sensitive colonies should further be tested and controlled as a $mucA^-$ strain.

# 5    Discussion

This study aimed to develop a CRISPR-Cas9 mutagenesis system for *A. vinelandii* that would be an improved method compared to homologous recombination. The vectors of the two-plasmid system have been designed as planned (Figure 4.1, 4.3, 4.5, 4.9 and 4.13). The Cas9 vector (pMDD8) based on the RK2 replicon was successfully transferred to *A. vinelandii*, but the transfer of the gRNA vectors based on the pBBR1 replicon were unsuccessful. Since the plasmids were able to replicate in *E. coli* DH5$\alpha$, it is not likely that it is anything wrong with the replicon. The replicon was isolated with classical restriction enzyme-based cloning, and not PCR where errors can be introduced. Hence, it could be excluded that the replicon is faulty.

Since the plasmids include a spectinomycin resistance gene, *E. coli* S17-1 could not serve as a donor in conjugation. The plasmids would neither be transferred directly to electrocompetent *A. vinelandii*. The reason for this is not known. The second approach was to use *E. coli* DH5$\alpha$ (pTA19). pTA19 was constructed by Strand *et al.*(2014) as a precursor to a plasmid, pTA-Mob, that was designed to aid in conjugation of plasmids based on the RK2 replicon. *E. coli* DH5$\alpha$ (pTA-Mob) is known to function as a donor strain in many species, but since it is based on the pBBR1-replicon, it could not act as a donor strain for the pBBR1 replicons pMDD13, pMDD15, pMDD16, and pMDD17. However, pTA19 was based on the RK2 replicon and contains the *tra* genes necessary for conjugal transfer. DH5$\alpha$ (pTA19) was then tested as a donor strain instead. The plasmids were electrotransformed into *E. coli* DH5$\alpha$ (pTA19), but would not be conjugated into *A. vinelandii*. Since DH5$\alpha$ (pTA19) had not been tested as a donor strain before, it is possible that the problem is the choice of donor strain. The strain was therefore tested as a donor in conjugation of pMDD7. The experiment was unsuccessful. However, the experiment was only performed once. To draw a conclusion and say for certain that *E. coli* DH5$\alpha$ (pTA19) does not work as a donor strain in conjugation, the experiment should be performed one more time. Nevertheless, the result could indicate that *E. coli* DH5$\alpha$ (pTA19) does not function as a donor strain, and give a possible reason for why the gRNA vectors could not be conjugated into *A. vinelandii*.

A possible solution would be to change the spectinomycin resistance gene for another, so that S17-1 can serve as a donor strain. S17-1 is in addition to spectinomycin and streptomycin resistant, also resistant to trimethoprim (Tp$^R$). Accordingly, neither of these can work as a selection marker. *A. vinelandii* grows poorly on kanamycin and needs a low concentration to survive, so this is not an optimal candidate. Nevertheless, kanamycin worked as a selection marker when plasmid pSB-B1I was transferred to *A. vinelandii*, even with a concentration of 5$\mu$g/mL. Tetracycline or gentamycin could be other options that has worked in *A. vinelandii*. It would have to be tested if the apramycin resistance gene does not give resistance to either of these, before exchanging the gene on the gRNA vector.

Plasmids that are too large might sometimes provide a problem in conjugation, but this should not be the case with the gRNA vectors since they are only 8453 bp and 9186 bp for pMDD15 and pMDD17, respectively. pMDD10, pMDD13, and pMDD16 were neither possible to transfer to *A. vinelandii* by conjugation although they are smaller (6757 bp, 6759 bp, and 6759 bp, respectively).

Another solution would be to construct a one-plasmid system. A two-plasmid system was decided on since it would be easier to cure for each of the vectors, and because one plasmid might be too large if both Cas9 and the gRNA were included. However, since the Cas9 vector pMDD8 could be transferred by conjugation, it would be worth investigating a possible one-plasmid system. The vector would not have to increase much in size to move only the gRNA to pMDD8. The increase would only be of a few hundred base pairs, as compared to about 2000 bp if the mutation templates were transferred to pMDD8 as well. Still, there was a significant reduction in the conjugation frequency of pMDD8, which has a size of 14335 bp compared to pMDD7 at 6334 bp. Only a few colonies with *A. vinelandii* (pMDD8) grew on the selective plates (0.4 cfu/mL). The question then is if a further increase in size will yield an even more considerable decrease in the frequency. The size of the vectors could be decreased by removing the genes of the $\lambda$-Red recombineering system (*gam*, *beta*, and *exo*) with the AraC/$P_{BAD}$ regulator/promoter system, for the vectors without a mutation template. This would produce an about 3.1 kb smaller vector.

The following section (5.1) describes a cloning plan for the one-plasmid system. If the one-plasmid system proves to be too large to be conjugated into *A. vinelandii*, a final possibility would be to exchange the CRISPR type II system for type V. This would yield a smaller vector since the gene encoding Cpf1 is about 200 nucleotides shorter than the *cas9* gene, in addition to the single guide RNA that is less than half the length (40 versus 100 nt) of the fusion gRNA of CRISPR-Cas9.

## 5.1 Further work

Figure 5.1 shows the plan to construct a one-plasmid delivery mechanism with both Cas9 and the gRNA specific for *uidA*.
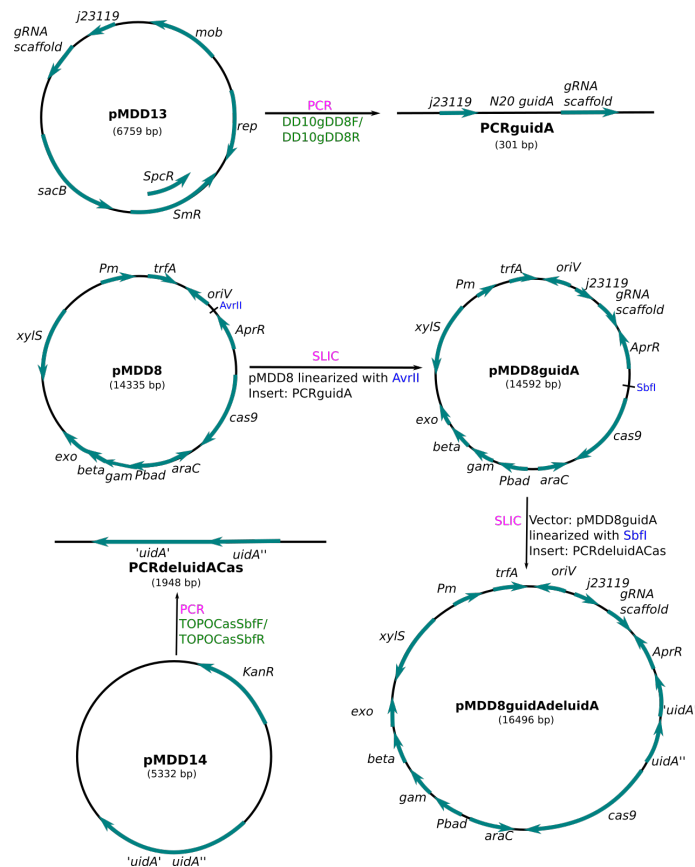


**Figure 5.1: Construction of a one-plasmid delivery system targeting *uidA*.** The plasmid would be based on the RK2 replicon. The *j23119* promoter, the specific N20 guidA sequence, the gRNA scaffold, and the mutation template for *uidA* would be transferred in two steps into the RK2 based plasmid pMDD8. Restriction enzymes with cleavage sites are colored blue, methods are colored pink, and primers are colored green.

The gRNA targeting *uidA* with the promoter *j23119* and scaffold would be amplified from pMDD13 by PCR with the primer pair DD10gDD8F/DD10gDD8R. The primers would be designed to create homologous ends of the PCR product, PCRguidA, with each end of AvrII restricted pMDD8. DD10gDD8F would be partly homologous with the region upstream of the *oriV* gene next to the AvrII cleavage site, and upstream of the *j23119* promoter. DD10gDD8R would be partially homologous with the area downstream of the *AprR* gene next to the AvrII cleavage site and downstream of the gRNA scaffold. Due to the homologous overlaps at each end, PCRguidA could be SLIC cloned between *oriV* and *AprR* at the AvrII restriction site. The new plasmid with guidA is shown as pMDD8guidA in Figure 5.1. PCR could amplify the mutation template in pMDD14 constructed based on *uidA* in pMDD2, with primers TOPOCasSbfF and TOPOCasSbfR. These primers should be designed the same way as DD10gDD8F/R to create homologous regions at both sides of the PCR product to each end of SbfI restricted pMDD8guidA. The PCR product PCRdeluidACas, could then be SLIC cloned into pMDD8guidA digested with SbfI, between *AprR* and *cas9*.

A one-plasmid system could also be constructed with *mucA* as target (Figure 5.2). This plasmid would be designed without a mutation template, and so could only be used to kill the bacterium unless a mutation template is delivered on another plasmid. The plasmid pMDD8gmucA could be cloned the same way as pMDD8guidA, based on a two fragment assembly with overlaps, only with pMDD16 as the template in PCR with primers DD10gDD8F and DD10gDD8R. The same primers could be used since pMDD16 and pMDD13 are the same, except for the difference in the N20 target sequence, and the primers only overlap with the regions on each side of this sequence.
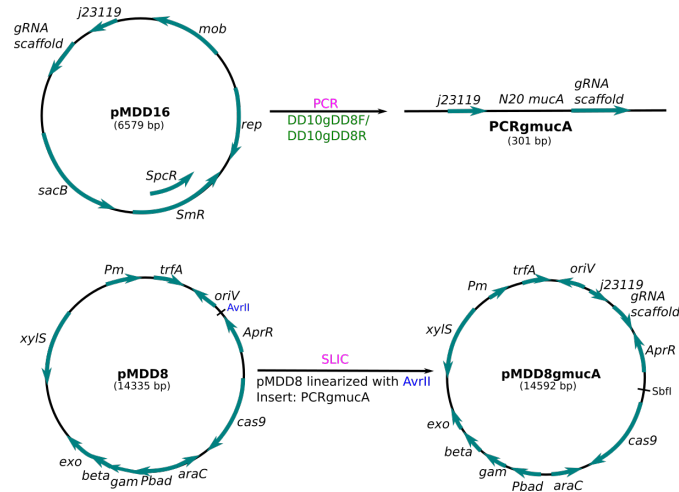
**Figure 5.2: Construction of a one-plasmid delivery system targeting *mucA*.** The plasmid would be based on the RK2 replicon. The *j23119* promoter, the specific N20 gmucA sequence, and the gRNA scaffold would be transferred into the RK2 based plasmid pMDD8. Restriction enzymes with cleavage sites are colored blue, methods are colored pink, and primers are colored green.

After the construction of the plasmids, they could be delivered to *A. vinelandii* by conjugation. Only a single conjugation should be required compared to the two-plasmid system where each plasmid needed to be conjugated in turn. Three *A. vinelandii* strains - wild type, *uidA*, and *mucA⁻* strain - could be used to test lethality and the mutation rate of the system. pMDD8gmucA would be transferred to the wild type strain and the *mucA*-wild type strain, created with homologous recombination. Since pMDD8gmucA should not contain any mutation template, it could be used to test the lethality of the system. The first strain, which only contains the wild type *mucA*, would be killed. The strain with both mutated and wild type copies of *mucA* would survive since some copies of the chromosome (the ones with mutated *mucA*) would stay intact. pMDD8 would be conjugated into both strains as control and counted to see total colony numbers. The lethality of pMDD8guidA would be tested the same way, but with the wild type strain and the *uidA* strain. Since the wild type strains does not contain *uidA*, they would survive. pMDD8 would serve as a control, and therefore, should be conjugated into both strains.

The mutation rate could be calculated by conjugating pMDD8guidAdeluidA into *A. vinelandii uidA* strain. The inducer arabinose should be added to the growth medium and to the selective plates to express the genes of the λ-Red recombineering system.

When Cas9 cleaves *uidA* integrated into the genome of the bacterium, Exo and Beta of the $\lambda$-Red system would repair the nick using the *uidA* mutation template. By adding X-Gluc to the plates, the wild type colonies would produce a product turning the colonies blue. The mutated colonies would not be able to express $\beta$-glucuronidase, and thereby, not produce the blue product. These colonies would be distinguished from the wild type colonies since they would appear white. The mutation rate could be calculated by counting white and blue colonies.

# 6   Conclusion

The goal of this work was to develop a CRISPR-Cas9 mutagenesis system for *A. vinelandii*, to look into whether this system could give higher mutation frequencies than homologous recombination, or if it could be used to remove unwanted chromosomes.

Vectors for a CRISPR type II mutagenesis system have been constructed during this study to find out if CRISPR could solve the problem of homologous recombination. The two-plasmid system with one vector for the Cas9, and the second carrying the gene encoding the gRNA, were constructed as planned. Only the Cas9 encoding vector (pMDD8) based on the RK2 replicon, was successfully conjugated into *A. vinelandii*. However, the conjugation frequency was very low (0.4 cfu/mL). The gRNA vectors (pMDD13, pMDD15, pMDD16, and pMDD17) based on the pBBR1 replicon could not be conjugated into *A. vinelandii* with *E. coli* S17-1 as a donor strain, since the vectors carry a spectinomycin resistance gene, and S17-1 is resistant to this particular antibiotic. Other delivery mechanisms were instead tested - electroporation of the vectors directly into electrocompetent *A. vinelandii*, and conjugation with the use of *E. coli* DH5α (pTA19) as donor strain. Both approaches failed. Since CRISPR needs both Cas9 and the gRNA to function, the system could not be tested and it is not possible to conclude whether the system is an improvement to homologous recombination. Further work must be performed to develop a functional system. Possible solutions would be to replace the spectinomycin resistance gene with another so that S17-1 can act as a donor strain, develop a one-plasmid system with both *cas9* and the gene encoding the gRNA, or to change the CRISPR type II system with a type V system. CRISPR type V consists of the smaller Cas9 homolog Cpf1, and a shorter gRNA. This could yield a smaller vector.

Development of a one-plasmid system is already under construction. After completion it will be possible to say if it functions, and if the system is an improvement to homologous recombination. Otherwise, development of a CRISPR type V system should be investigated.

# References

1. Özen, A. I. & Ussery, D. W. Defining the Pseudomonas Genus: Where Do We Draw the Line with Azotobacter? *Microbial Ecology* **63,** 239–248 (2012).

2. Clementi, F. Alginate production by Azotobacter vinelandii. *Critical Reviews in Biotechnology* **17,** 327–361 (1997).

3. Draget, K. I., Skjåk-Bræk, G. & Smidsrød, O. Alginate based new materials. *International Journal of Biological Macromolecules* **21,** 47–55 (1997).

4. Pindar, D. F. & Bucke, C. The biosynthesis of alginic acid by Azotobacter vinelandii. *Biochemical Journal* **152,** 617–622 (1975).

5. Campos, M. E. *et al.* Characterization of the gene coding for GDP-mannose dehydrogenase (algD) from Azotobacter vinelandii. *Journal of Bacteriology* **178,** 1793–1799 (1996).

6. Guzma, J., Moreno, S., Na, R. & Microbiologı, D. D. Role of Alternative sigma Factor AlgU in Encystment of Azotobacter vinelandii. **180,** 2766–2769 (1998).

7. Schurr, M. J. *et al.* Control of AlgU, a Member of the sigma E-Like Family of Stress Sigma Factors, by the Negative Regulators MucA and MucB and Pseudomonas aeruginosa Conversion to Mucoidy in Cystic Fibrosis. **178,** 4997–5004 (1996).

8. Martínes-Salazar, J. M. *et al.* Characterization of the Genes Coding for the Putative Sigma Factor AlgU and Its Regulators MucA, MucB, MucC, and MucD in Azotobacter vinelandii and Evaluation of Their Roles in Alginate Biosynthesis. *Microbiology* **178,** 1800–1808 (1996).

9. Núñez, C. *et al.* Role of Azotobacter vinelandii mucA and mucC gene products in alginate production. *Journal of Bacteriology* **182,** 6550–6556 (2000).

10. Mathee, K., McPherson, C. J. & Ohman, D. E. Posttranslational control of the algT (algU)-encoded $\sigma 22$ for expression of the alginate regulon in Pseudomonas aeruginosa and localization of its antagonist proteins MucA and MucB (AlgN). *Journal of Bacteriology* **179,** 3711–3720 (1997).

11. Hay, I. D. *et al.* Microbial alginate production, modification and its applications. *Microbial Biotechnology* **6,** 637–650 (2013).

12. Ertesvåg, H. *et al.* A family of modular type mannuronan C-5-epimerase genes controls alginate structure in Azotobacter vinelandii. *Molecular Microbiology* (1995).

13. Ivanova, E. P., Bazaka, K. & Crawford, R. J. *New Functional Biomaterials for Medicine and Healthcare* (Elsevier Science, 2013).

14. Plazinski, W. Molecular basis of calcium binding by polyguluronate chains. Revising the egg-box model. *Journal of Computational Chemistry* **32,** 2988–2995 (2011).

15. Galindo, E. *et al.* Molecular and bioengineering strategies to improve alginate and polydydroxyalkanoate production by Azotobacter vinelandii. *Microbial Cell Factories* **6,** 1–16 (2007).

16. Draget, K. I., Smidsrød, O. & Skjåk-Bræk, G. in *Biopolymers Online* (American Cancer Society, 2005). `https://onlinelibrary.wiley.com/doi/abs/10.1002/3527600035.bpol6008`.

17. Skjåk-Bræk, G., Zanetti, F. & Paoletti, S. Effect of acetylation on some solution and gelling properties of alginates. *Carbohydrate Research* **185,** 131–138 (1989).

18. Maldonado, R., Jimenez, J. & Casadesus, J. Changes of ploidy during the Azotobacter vinelandii growth cycle. *Journal of Bacteriology* **176,** 3911–3919 (1994).

19. Qurollo, B. A., Bishop, P. E. & Hassan, H. M. Characterization of the iron superoxide dismutase gene of Azotobacter vinelandii: sodB may be essential for viability. *Canadian Journal of Microbiology* **47,** 63–71 (2001).

20. Li, X. & Heyer, W.-D. Homologous recombination in DNA reapir and DNA damage tolerance. *Cell Res.* **18,** 99–113 (2008).

21. Snustad, D. P. *Genetics* 6th (Wiley, Singapore, 2012).

22. Phillips, R. J., Hickleton, D. C., Boehmer, P. E. & Emmerson, P. T. The RecB protein of Escherichia coli translocates along single-stranded DNA in the 3 to 5 direction: a proposed ratchet mechanism. *MGG - Molecular & General Genetics* **254,** 319–329 (Apr. 1997).

23. Nelson, D. L. *Lehninger principles of biochemistry* 6th (W.H. Freeman, New York, 2013).

24. Lee, E. C. *et al.* A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73,** 56–65 (2001).

25. Cassuto, E., Lash, T., Sriprakash, K. S. & Radding, C. M. Role of Exonuclease and Beta Protein of Phage Lambda in Genetic Recombination, V. Recombination

of Lambda DNA in Vitro. *Proceedings of the National Academy of Sciences* **68,** 1639–1643 (2006).

26. Karakousis, G. *et al.* The beta protein of phage $\lambda$ binds preferentially to an intermediate in DNA renaturation. *Journal of Molecular Biology* **276,** 721–731 (1998).

27. Kmiec, E. & Holloman, W. K. Beta protein of bacteriophage Lambda promotes renaturation of DNA. *The journal of biological chemistry* **256,** 12636–12639 (1981).

28. Marsic, N. *et al.* In vivo studies on the interaction of RecBCD enzyme and $\lambda$ gam protein. *Journal of Bacteriology* **175,** 4738–4743 (1993).

29. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nature protocols* **4,** 206–223 (2009).

30. Kuzminov, A. Homologous Recombination-Experimental Systems, Analysis and Significance. *EcoSal Plus* **4,** 1124–1129 (2011).

31. Gay, P. *et al.* Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of Bacteriology* **164,** 918–921 (1985).

32. Barrangou, R., Fremaux, C., Deveau, H. & Richards, M. CRISPR Provides Aquired Resistance Against Viruses in Prokaryotes. *Science* **315,** 1709–1712 (2007).

33. *CRISPR Systems - Doudna Lab* `http://doudnalab.org/research_areas/crispr-systems/`.

34. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews Microbiology* **13,** 722–736 (2015).

35. Mary Gearing. CRISPR 101: A Desktop Resource. **2016,** 125 (2016).

36. Martynov, A., Severinov, K. & Ispolatov, I. Optimal number of spacers in CRISPR arrays. *PLoS Computational Biology* **13,** 1–23 (2017).

37. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Molecular Cell* **54,** 234–244 (2015).

38. Karvelis, T. *et al.* crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biology* **10,** 841–851 (2013).

39. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602 (Mar. 2011).

40. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. *Science (New York, N.Y.)* **337,** 816–822 (2012).

41. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).

42. Jiang, W. *et al.* CRISPR-assisted editing of bacterial genome. **31,** 233–239 (2013).

43. Palermo, G. *et al.* CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences* **114,** 7260–7265 (2017).

44. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109,** E2579–E2586 (2012).

45. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. **118,** 6072–6078 (2016).

46. *Multiplex Genome Editing with CRISPR-Cpf1* `https://blog.addgene.org/ multiplex-genome-editing-with-crispr-cpf1`.

47. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

48. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2014).

49. Sun, J. *et al.* Genome editing and transcriptional repression in Pseudomonas putida KT2440 via the type II CRISPR system. *Microbial Cell Factories* **17,** 1–17 (2018).

50. Moreno-mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR/ Cas9 targeting in vivo. **12,** 982–988 (2016).

51. Pitcher, R. S., Brissett, N. C. & Doherty, A. J. *Nonhomologous end-joining in bacteria: a microbial perspective.* 2007.

52. Yan, Q. & Fong, S. S. Study of in vitro transcriptional binding effects and noise using constitutive promoters combined with UP element sequences in Escherichia coli. *Journal of Biological Engineering* **11,** 1–11 (2017).

53. Lale, R. *The xylS/Pm expression cassette: New functional insights and application potentials* PhD thesis (2009).

54. Lambert, R. J. & Stratford, M. Weak-acid preservatives: modelling microbial inhibition andresponse. *Journal of Applied Microbiology* **86,** 157–164 (1999).

55. Brautaset, T., Lale, R. & Valla, S. Positively regulated bacterial expression systems. *Microbial Biotechnology* **2,** 15–30 (2009).

56. Englesberg, E., Irr, J., Power, J. & Lee, N. Positive control of enzyme synthesis by gene C in the L-arabinose system. *Journal of Bacteriology* **90,** 946–957 (1965).

57. Lee, N., Francklyn, C. & Hamilton, E. P. Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proceedings of the National Academy of Sciences of the United States of America* **84,** 8814–8 (1987).

58. Botsford, J. L. & Harman, J. G. Cyclic AMP in prokaryotes. *Microbiological reviews* **56,** 100–22 (1992).

59. Gearing, M. *Plasmids 101: Inducible Promoters* 2018. `https://blog.addgene.org/plasmids-101-inducible-promoters`.

60. Simon, R., Priefer, U. & Pühler, A. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology* **1,** 784 (Nov. 1983).

61. Strand, T. A. *et al.* A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE* **9,** 1–6 (2014).

62. Caspi, R. *et al. A broad host range replicon with different requirements for replication initiation in three bacterial species* 2001.

63. Gimmestad, M. *et al.* Characterization of three new Azotobacter vinelandii alginate lyases, one of which is involved in cyst germination. *Journal of Bacteriology* **191,** 4845–4853 (2009).

64. Khan, S. R., Gaines, J., Roop, R. M. & Farrand, S. K. Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. *Applied and Environmental Microbiology* **74,** 5053–5062 (2008).

65. Fernández-Lópeza, C. *et al.* Mobilizable Rolling-Circle Replicating Plasmids from Gram- Positive Bacteria: A Low-Cost Conjugative Transfer. *Microbiol Spectr.* **2** (2014).

66. Lilly, J. & Camps, M. Mechanisms of Theta Plasmid Replication. *Microbiol Spectr.* **3** (2015).

67. Cédric Y. Szpirer Michel Faelen, M. C. Mobilization Function of the pBHR1 Plasmid, a Derivative of the Broad-Host-Range Plasmid pBBR1. *Journal of Bacteriology* **63,** 9–12 (2008).

68. Panja, S., Saha, S., Jana, B. & Basu, T. Role of membrane potential on artificial transformation of E. coli with plasmid DNA. *Journal of Biotechnology* **127,** 14–20 (2006).

69. Roberts, R. J. How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences* **102,** 5905–5908 (2005).

70. *Benchling* https://www.benchling.com/academic/.

71. Jeong, J.-Y. *et al.* One-Step Sequence- and Ligation-Independent Cloning as a Rapid and Versatile Cloning Method for Functional Genomics Studies. *Applied and Environmental Microbiology* **78,** 5440–5443 (2012).

72. *The European Bioinformatics Institute* http://europepmc.org/abstract/MED/30976793.

73. Kovach, M. E. *et al.* Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* (1995).

74. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Applied and Environmental Microbiology* **81,** 2506–2514 (2015).

# Appendix A   Materials

The growth mediums, buffers and antibiotics used in the experiment were made as described in the following section.

**Burks medium**

For the preparation of Burks medium a 5x concentrated Burks solution containing the salts listed below was prepared.

**5x Burks**

1.00 g/L $MgSO_4 \cdot 7H_2O$

1.00 g/L NaCl

3.20 g/L $K_2HPO_4$

0.80 g/L $KH_2PO_4$

deionized water

5x Burks was diluted with water to 1x stock and autoclaved at 120 °C.

**Glucose solution**

A 120 g/L glucose solution was prepared separately.

6.00 g glucose

50 mL deionized water

**Sterile-filtered stock-solutions**

Stock solutions of $FeSO_4$, $CaCl_2$ and $Na_2MoO_4$ was prepared and sterile filtered.

150 g/L $FeSO_4$

500 g/L $CaCl_2$

21 g/L $Na_2MoO_4$

The autoclaved glucose solution (120 g/L) was transferred sterile to autoclaved 1x Burks medium, giving a final concentration of 20 g/L glucose. Finally sterile-filtered stock solutions of $FeSO_4$, $CaCl_2$ and $Na_2MoO_4$ was transferred to the mixture for a final concentration of 30 mg/L, 50 mg/L and 2.1 mg/L respectively.

**Luria-Bertani (LB) medium**

For the LB medium the following concentrations were used.

10 g/L tryptone

5 g/L yeast extract

5 g/L NaCl

The medium was autoclaved at 120 °C.

For agar plates, bacterial agar was added to the solution for a final concentration of 15 g/L before autoclaving.

**LA medium**

15 g/L agar was added to the LB medium, autoclaved at 120 °C and cooled down to 50 °C.

**Psi medium**

5 g/L yeast extract

20 g/L tryptone

10.24 g/L MgSO$_4$·7 H$_2$O

The pH was adjusted to 7.6 with KOH and then autoclaved.

**TFB1**

2.94 g/L KAc

12.1 g/L RbCl

1.47 g/L CaCl$_2$

10 g/L MnCl$_2$·4H$_2$O

150 mL/L glycerol

The solution was mixed with deionized water (less than the final intended volume). The pH was adjusted to 5.8 using dilute acetic acid, and deionized water was added up to the correct volume.

The buffer was sterile-filtered and 40 mL aliquoted into 50 mL tubes.

**TFB2**

2.1 g/L MOPS

11 g/L $CaCl_2$

1.21 g/L RbCl

150 mL/L glycerol

The solution was mixed with deionized water (less than the final intended volume). The pH was adjusted to 6.5 using dilute NaOH, and deionized water was added up to the correct volume.

The buffer was sterile-filtered.

**SOC medium**

20 g/L tryptone

5 g/L yeast extract

0.5 g/L NaCl

2.5 mM KCl

3.6 g/L glucose

5.08 g/L MgCl2

The solution was sterilized by filtration, aliquoted into 1.5 mL Eppendorf tubes and stored in the freezer at $-20\,°C$.

**Stock-solutions**

Stock-solutions of antibiotics were made by dissolving it in deionized water. The final concentrations of ampicillin, kanamycin, apramycin, spectinomycin, and tetracycline was respectively 100 mg/mL, 50 mg/mL, 50 mg/mL, 20 mg/mL, and 5 mg/mL.

The solutions were filter-sterilized and aliquoted into 1.5 mL Eppendorf tubes and stored in the freezer at $-20\,°C$.

A 0.5 M m-toluic acid solution was prepared.

0.6807 g m-toluic acid

Dissolved in deionized water up to 10 mL.

# Appendix B   Bacterial strains and plasmids

Table B.1 shows the bacterial strains used in the experiment.

**Table B.1:** Bacterial strains used during the experiment.

| Strain | Description | Source |
|---|---|---|
| *A. vinelandii* ATCC 12518 | Target strain of CRISPR-Cas9 system. | American Type Culture Collection |
| *E. coli* DH5$\alpha$ | Used in standard cloning work. | Bethesda Research Laboratories |
| *E. coli* S17-1 | Used in standard cloning work for plasmids with the XylS/*Pm* regulator/promoter system, and in conjugal transfer. | [60] |

Table B.2 shows the plasmids used in the experiment. The vector for *uidA*, the Cas9 vector and gRNA vector are the final plasmids in the cloning procedures that have not yet been cloned.

**Table B.2:** Plasmids used in the experiment.

| Plasmid | Description | Source |
|---|---|---|
| pAGL4 | Contains the gene *uidA*. Kan$^R$. | A. G. Løodeng (unpublished). |
| pMDD2 | Derivative of pAGL4 from which a 6 nt NotI-XhoI fragment was removed. The remaining fragment was blunted with Klenow DNA polymerase and self-ligated. Kan$^R$. | This work. |
| pHE95 | A conditional suicide vector dependent on the XylS/*Pm* regulator/promoter system and with the RK2 replicon. Amp$^R$. | Gimmestad *et al.* [63] |

| | | |
|---|---|---|
| pMDD6 | Derivative of a 7.3 kb SalI-XhoI digested pHE95 fragment that was re-ligated, so that *sacB* was removed from the plasmid. 7.3 kb). Amp$^R$. | This work. |
| pION100 | Apr$^R$. | I. Onsager (un-published). |
| pMDD7 | Derivative of 4.9 kb KpnI-AvrII digested PCR product containing the XylS/*Pm* regulator/promoter system and the RK2 replicon from pMDD6. A 1.4 kb fragment (apramycin resistance gene) from KpnI-AvrII digested pION100 was inserted into the vector. Apr$^R$. | This work. |
| pCas | Contains genes encoding Cas9 and the $\lambda$-Red recombination system. Kan$^R$. | Jiang *et. al.* [74] |
| pMDD8 | Derivative of pMDD7 in which a BamHI-SbfI 6.3 kb fragment was ligated with 8.0 kb fragment from SbfI-BglII digested pCas carrying the genes encoding Cas9 and the $\lambda$-Red proteins. Apr$^R$. | This work. |
| pTargetF | Contains the gRNA scaffold with its promoter *j23119*. Spc$^R$. | Jiang *et al.*[74] |
| pAGL2 | Contains the gene encoding SacB for self-curing. Apr$^R$. | A. G. Lødeng (unpublished). |
| pMDD9 | Derivative of pTargetF digested with XbaI-PstI (2.1 kb fragment), in which a 2.0 kb XbaI-PstI fragment containing *sacB* from pAGL2 was inserted. Spc$^R$. | This work. |

| | | |
|---|---|---|
| pBBRIMCS-2 | A pBBR1 replicon with the genes *mob* and rep. Kan$^R$. | [73] |
| pMDD10 | Derivative of a 3.6 kb fragment containing the pBBR1 replicon and selection marker from pBBRIMCS-2 (BglII-SacI), with an insert of 3.1 kb BamHI-SacI fragment from pMDD9, carrying the gRNA scaffold genes with its promoter, and *sacB*. Spc$^R$. | This work. |
| pMDD11 | TOPO vector with PCR insert cloned from pMDD10, where a NcoI restriction cleavage site was introduced between the gRNA scaffold and promoter *j23119*. Kan$^R$. | This work. |
| pMDD12 | Derivative of pMDD10 where a 1269 bp SpeI-BsrG1 fragment in pMDD10, was exchanged for a 1269 bp SpeI-BsrG1 fragment from the PCR product in pMDD11 with the NcoI restriction cleavage site. Spc$^R$. | This work. |
| pMDD13 | Derivative of pMDD12 where the N20 target sequence of *uidA* was inserted between a SpeI site and the newly introduced NcoI site in pMDD12 (both sites lie between the gRNA scaffold and *j23199*), using NEBuilder DNA Assembly with the oligomer guidA1441NEB. Spc$^R$. | This work. |
| pMDD14 | TOPO vector with the mutation template for *uidA* cloned from pMDD2 with PCR. Kan$^R$. | This work. |

| pMDD15 | Derivative of PstI-SalI restricted and linearized pMDD13 (6702 bp fragment) with the *uidA* mutation template from pMDD14 cloned with PCR, and then SLIC cloned into the linearized vector. Spc$^R$. | This work. |
| --- | --- | --- |
| pMDD16 | Derivative of pMDD12 where the N20 target sequence of *mucA* was inserted between a SpeI site and the newly introduced NcoI site in pMDD12 (both sites lie between the gRNA scaffold and *j23199*), using NEBuilder DNA Assembly with the oligomer gmucA3651NEB. Spc$^R$. | This work. |
| pHE280 | Carrying a mutation template for *mucA*. Amp$^R$. | H. Ertesvåg (unpublished). |
| pMDD17 | Derivative of pMDD16 (6675 bp EcoRV-BglII restriction fragment) with the *mucA* mutation template from Eco53kI-BglII restricted pHE280 (2511 bp) inserted into the vector. Spc$^R$. | This work. |
| pTA19 | A RK2 replicon with a deficiant *oriT* (*oriT*$^-$) and *tra* genes for conjugal transfer. Tc$^R$. | Strand *et al.* [61]. |
| pHE437 | A RK2 replicon dependent on the XylS/*Pm* regulator/promoter system with a *mucA* mutattion template for homologous recombination. Tet$^R$. | H. Ertesvåg (unpublished). |

# Appendix C   Primer sequences

Table C.1 shows the primers used in PCR.

**Table C.1:** Primers used in the experiment and their sequences. "R" in their names indicate that they are primers for the reverse strand, while "F" is for the forward strand. The direction is 5' to 3'.

| Primer | Sequence | Source |
|---|---|---|
| oppPHBBF | CGCATGCACGAACTCGATACAC | Sigma-Aldrich |
| oppPHBBR | TCGAGAACTTCCTAGGTTGTAGAGGCGCCGGCTG | Sigma-Aldrich |
| NedPHBBF | TACAACCTAGGAAGTTCTCGAGCAGATCAG | Sigma-Aldrich |
| NedPHBBR | TTTCCAGACAGCGACGGGTAG | Sigma-Aldrich |
| 95FAvr | TAACCTAGGAGCGCAACCCACTCACTAC | Sigma-Aldrich |
| 95RKpn | TATGGTACCGCTGCATAACCCTGCTTC | Sigma-Aldrich |
| ouidAF | GATATCAAGCTTATCGATACCGCACCGAAGTTCATGCCAGTC | Sigma-Aldrich |
| ouidAR | CGTCACAGCTCCGGTCAGTGGCAGTGAAG | Sigma-Aldrich |
| nuidAF | CACTGACCGGAGCTGTGACGCACAGTTCATAG | Sigma-Aldrich |
| nuidAR | GGGTAATAGATCTAAGCTTCGAGGAGTCCCTTATGTTACG | Sigma-Aldrich |
| gmucA3651F | GGCGGCTTCGGTGACTGTCGACTAGTATTATACCTAGGACTGAG | Sigma-Aldrich |
| gmucA3651R | GTCCTAGGTATAATACTAGTGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAG | Sigma-Aldrich |
| guidA1441F | GACCGGATGCCGACGCGAAGACTAGTATTATACCTAGGACTGAG | Sigma-Aldrich |
| guidA1441R | CTTCGCGTCGGCATCCGGTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG | Sigma-Aldrich |
| mDD10F | ACTAGTCATCGCCGCAGCGGTCCATGGTTTTAGAGCTAGAAATAG | Sigma-Aldrich |
| mDD10F2 | ACTAGTCATCGCCGCAGCGGTCCATGGTTTTAGAGC | Sigma-Aldrich |
| mDD10R | GCGTATGGTTGTCGCCTGAG | Sigma-Aldrich |
| gmucA3651NEB | CAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGGCGGCTTCGGTGACTGTCGGTTTTAGAGCTAGAAATAGCAAGTT | Invitrogen |
| guidA1441NEB | CAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGACCGGATG CCGACGCGAAGGTTTTAGAGCTAGAAATAGCAAGTT | Invitrogen |
| DD10gDD8F | GTAGTGAGTGGGTTGCGCTCGGAAGCCCTGCAAAGTAAAC | Invitrogen |
| DD10gDD8R | CAACAGTTGCGCAGCCTGACA CGGCAGGTATATGTGATGG | Invitrogen |
| TOPOCasSbfF | TACTCTAGCTTCCCGGCACCGCTCGGATCCACTAGTAACG | Invitrogen |
| TOPOCasSbfR | TCTACTCTGACTGCAAACCCGAATTGGGCCCTCTAGATG | Invitrogen |

# Appendix D   PCR program

The PCR reaction mixture was set up as shown below with a final volume of 25 $\mu$L.

5$\mu$L 5X Q5 reaction buffer

0.5 $\mu$L 10mM dNTP

1.25 $\mu$L 10 $\mu$M Forward Primer

1.25 $\mu$L 10 $\mu$M Reverse Primer

<1.000 ng Template DNA

0.25 $\mu$L High-Fidelity DNA Polymerase

5 $\mu$L 5X Q5 High GC Enhancer

Nuclease free water up to 25 $\mu$L

When using two overlapping DNA fragments as templates the same molar amounts was added to the reaction mixture.

The PCR reaction was carried out using a PCR machine. The general steps of the program was as follows.

1. Initial denaturation at 98 °C for 30 seconds
2. 98 °C for 10 seconds
3. Lowest $T_m$ + 3 °C for 30 seconds
4. 72 °C for 30 seconds/kb template DNA
5. Repeat step for 2.-4. 25-35 cycles
6. Final extension at 72 °C for 2 minutes
7. Hold at 4 °C

# Appendix E   Sequencing of pMDD7

The sequencing results of pMDD7 are presented in the next pages. The first line shows the template pMDD7, and the next lines show in successive order the alignment of primer 95RKpn, 385Age, AvPmsjekkB, pmucA4400, pmucA5200, catBspQ1, pmv23, and amsacF. A few errors can be shown in the alignment, but must of them are outside any coding region. pmucA4400 and pmucA5200 cover *trfA*, and amsacF cover *oriV*. A few errors could be found in the alignment of these primers. However, since the plasmid was able to replicate these errors were ignored. "cov" shows how much of the template the aligned sequence covers in percentage. "pid" shows the percentage of identity between the aligned sequence and the template.

Reference sequence (1): pMDD7
Identities normalised by aligned length.
Colored by: identity

```
                    cov     pid    1 [                    .                    .                    :                    .  80
1 pMDD7          100.0%  100.0%     GATTTTTTTCGGTATATCCATCCTTTTTTCGCACGATATACAGGATTTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTA
2 95RKpn          16.5%   99.2%     GATTTTTTTCGGTATATCCATCCTTTTTTCGCACGATATACAGGATTTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTA
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid   81                    .                    1                    .                    .  160
1 pMDD7          100.0%  100.0%     TCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTACTCACTGTCCTTATTCG
2 95RKpn          16.5%   99.2%     TCCAACGGCGTCAGCCAGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCG
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  161                    .                    .                    2                    .  240
1 pMDD7          100.0%  100.0%     CACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCG-TAAGCTCTAAGAAACCATTATTATCATGACATTA
2 95RKpn          16.5%   99.2%     CACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGATAAGCTCTAAGAAACCATTATTATCATGACATTA
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  241                    :                    .                    .                    3  320
1 pMDD7          100.0%  100.0%     ACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAATTCACTGGCCGTCGTTTTACAACGTCGTGAC
2 95RKpn          16.5%   99.2%     ACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAATTCACTGGCCGTCGTTTTACAACGTCGTGAC
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  321                    .                    :                    .                    4  400
1 pMDD7          100.0%  100.0%     TGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGGCCAGCAGATCCACATCCTTGAAGG
2 95RKpn          16.5%   99.2%     TGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTCGCAGCACATCCCCCTTTC-GCCAGCAGATCCACATCCTTGAAGG
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  401                    .                    .                    :                    .  480
1 pMDD7          100.0%  100.0%     CCGCAGCGACGAGCAGAAGGAAACCCTCATTTCGGGAAGTCAGCGAGGCCATCTCGCGCTCCCTGGATGCGCCGCTGACCA
2 95RKpn          16.5%   99.2%     CCGCAGCGACGAGCAGAAGGAAACCCTCATTTCGGGAAGTCAGCGAGGCCATCTCGCGCTCCCTGGATGCGCCGCTGACCA
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  481                    .                    5                    .                    .  560
1 pMDD7          100.0%  100.0%     GCGTGCGAGTGATTATCACGGAGATGGCCAAGGGCCACTTCGGCATCGGCGGCGAACTGGCCAGCAAGGTCAGACGCTGA
2 95RKpn          16.5%   99.2%     GCGTGCGAGTGATTATCACGGAGATGGCCAAGGGCCACTTCGGCATCGGCGGCGAACTGGCCAGCAAGGTCAGACGCTGA
3 385Age          21.8%   95.0%     --------------------------------------------------------------------------------
4 AvPmsjekkB      22.9%   94.9%     --------------------------------------------------------------------------------
5 pmucA4400       16.8%   98.0%     --------------------------------------------------------------------------------
6 pmucA5200       18.6%   96.8%     --------------------------------------------------------------------------------
7 catBspQ1         8.4%   98.2%     --------------------------------------------------------------------------------
8 pmv23            5.8%   95.3%     --------------------------------------------------------------------------------
9 amsacF          11.6%   99.5%     --------------------------------------------------------------------------------

                    cov     pid  561                    .                    .                    6                    .  640
1 pMDD7          100.0%  100.0%     AGTGGAGATGCCCAAGGGCACTTCGGGTCGAGGAACCCGACCTGCATTGGGACGCGGCCACGGAGAGCGCGGGCAAACGC
```

A11

```
2 95RKpn    16.5%  99.2%   AGTGGAGATGCCCAAGGGCACTTCGGGTCGAGGAACCCGACCTGCATTGGGACGCGGCCACGGAGAGCGCGGGCAAACGC
3 385Age    21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid   641        :         .         .         .         .         7         .         . 720
1 pMDD7     100.0% 100.0%  CGGCACTATAGCCAGTGGAGTTTGTAAAACGCTATTTCAGAGCTTGGAGAGTGTCTAAGAAAGCC-GGGCGATGCCAACC
2 95RKpn    16.5%  99.2%   CGGCACTATAGCCAGTGGAGTTTGTAAAACGCTATTTCAGAGCTTGGAGAGTGTCTAAGAAAGCCGGGGCGATGCCAACC
3 385Age    21.8%  95.0%   ------------------------GCGGGCATTTTAGAGCTTGGAGAGTGTCTAAGAAGCCG-GGGCGATGCCAACC
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid   721        .         .         :         .         .         .         .         8 800
1 pMDD7     100.0% 100.0%  CATCCCTTCTTCGGCTACGTTCGTAATCAAGCCACTTCCTTTTTGCATTGACGCAGGGTGTCGGAAGGCAACTCGCCGAA
2 95RKpn    16.5%  99.2%   CATCCCTTCTTCGGCTACGTTCGT-ATCAAGCCACTTCCTTTTTGCATTGACGCAGGGTGTCGGAAGG-----------
3 385Age    21.8%  95.0%   CATCCCTTCTTCGGCTACGTTCGTAATCAAGCCACTTCCTTTTTGCATTGACGCAGGGTGTCGGAAGGCAACTCGCCGAA
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid   801        .         .         .         .         .         .         .         . 880
1 pMDD7     100.0% 100.0%  CGCGCTCCTATAGTTTTCAGCGAAGCGTCCCAAATGTAAGAAGCCGTAGTCTCAGGGCTATCTCAGTTATACTACGCACAT
2 95RKpn    16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age    21.8%  95.0%   CGCGCTCCTATAGTTTTCAGCGAAGCGTCCCAAATGTAAGAAGCCGTAGTCTCAGGGCTATCTCAGTTATACTACGCACAT
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid   881        .         .         9         .         .         .         .         . 960
1 pMDD7     100.0% 100.0%  TGGCACTGGGATCGTTCAAGCAGGCGCGGATGCTTTTCGAGCTTTGCGGTTGCGGATGTAGTTCTTTCGGCGTGGTGCCGGCA
2 95RKpn    16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age    21.8%  95.0%   TGGCACTGGGATCGTTCAAGCAGGCGCGGATGCTTTTCGAGCTTTGCGGTTGCGGATGTAGTTCTTTCGGCGTGGTGCCGGCA
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid   961        .         .         .         0         .         .         .         . 1040
1 pMDD7     100.0% 100.0%  TGCTTCTCGAACAAATTGTAGAGCGAGCGTGGACTCATCATCGCCAGCTCCGCTAACCGCTCAAGGCTGATATTCCGTTT
2 95RKpn    16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age    21.8%  95.0%   TGCTTCTCGAACAAATTGTAGAGCGAGCGTGGACTCATCATCGCCAGCTCCGCTAACCGCTCAAGGCTGATATTCCGTTT
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid  1041        :         .         .         .         .         1         .         . 1120
1 pMDD7     100.0% 100.0%  GAGATTCTCCTCAATGAATTGAACGACTCGCTCGAAAGACGGGTTACCTTTGCTGAAAATTTTCACGGCTGACATTGCTGC
2 95RKpn    16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age    21.8%  95.0%   GAGATTCTCCTCAATGAATTGAACGACTCGCTCGAAAGACGGGTTACCTTTGCTGAAAATTTTCACGGCTGACATTGCTGC
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------

            cov    pid  1121        .         .         :         .         .         .         .         2 1200
1 pMDD7     100.0% 100.0%  CCAGCATTTTCGAGCAGCTTGGAAGCGATGATCCCCGCATAGTGCTCTTGGACCCGAGGCATCGACTTTGTATGTTCCGCT
2 95RKpn    16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age    21.8%  95.0%   CCAGCATTTTCGAGCAGCTTGGAAGCGATGATCCCCGCATAGTGCTCTTGGACCCGAGGCATCGACTTTGTATGTTCCGCT
4 AvPmsjekkB 22.9% 94.9%   --------------------------------------------------------------------------------
5 pmucA4400  16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200  18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1    8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23       5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF     11.6%  99.5%   --------------------------------------------------------------------------------
```

A12

```
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2481          .         5         .         .         .         .         :         .  2560
1 pMDD7       100.0% 100.0%   CTGAAGAACAAATAGTTTTTTTCCTGATCCATAAAGCAGAACGGCCTGCTCCATGACAAATCTGGCTCCCCAACTAATGCC
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   CTGAAGAACAAATAGTTTTTTTCCTGATCCATAAAGCAGAACGGCCTGCTCCATGACAAATCTGGCTCCCCAACTAATGCC
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2561          .         .         .         6         .         .         .         .  2640
1 pMDD7       100.0% 100.0%   CCATGCAGCCAGCATAACCAGCATAAACTGTCCGGTTTGATAGGGATAAGTCCAGCCTTGCAAGAAGCGGATACAGGAGT
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   CCATGCAGCCAGCATAACCAGCATAAACTGTCCGGTTTGATAGGGATAAGTCCAGCCTTGCAAGAAGCGGATACAGGAGT
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2641          :         .         .         .         .         7         .         .  2720
1 pMDD7       100.0% 100.0%   GCAAAAAATGGCTATCTCTAGAAAGGCCTACCCCTTAGGCTTTATGCAACAGAAACAATAATAATGGAGTCATGAACATA
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   GCAAAAAATGGCTATCTCTAGAAAGGCCTACCCCTTAGGCTTTATGCAACAGAAACAATAATAATGGAGTCATGAACATA
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2721          .         .         :         .         .         .         .         8 2800
1 pMDD7       100.0% 100.0%   ATGAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAACTGATCGACGCGGGGTTTTCCGCCGAGGATGCCGAAACCAT
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   ATGAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAACTGATCGACGCGGGGTTTTCCGCCGAGGATGCCGAAACCAT
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2801          .         .         .         .         :         .         .         .  2880
1 pMDD7       100.0% 100.0%   CGCAAGCCGCACCGTCATGCGTGCGCCCCGCGAAACCTTCCAGTCCGTCGGCTCGATGGTCCAGCAAGCTACGGCCAAGA
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   CGCAAGCCGCACCGTCATGCGTGCGCCCCGCGAAACCTTCCAGTCCGTCGGCTCGATGGTCCAGCAAGCTACGGCCAAGA
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2881          .         9         .         .         .         .         .         :  2960
1 pMDD7       100.0% 100.0%   TCGAGCGCGACAGCGTGCAACTGGCTCCCCCTGCCCTGCCCGCGCCATCGGCCGCCGTGGAGCGTTCGCGTCGTCTCGAA
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   TCGAGCGCGACAGCGTGCAACTGGCTCCCCCTGCCCTGCCCGCGCCATCGGCCGCCGTGGAGCGTTCGCGTCGTCTCGAA
6 pmucA5200    18.6%  96.8%   ----------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 2961          .         .         .         0         .         .         .         .  3040
1 pMDD7       100.0% 100.0%   CAGGAGGCGGCAGGTTTGGCGAAGTCGATGACCATCGACACGCGAGGAA-CTATGACGACCAAGAAGCGAAAAACCGCCG
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   CAGGAGGCGGCAGGTTTGGCGAAGTCGATGACCATCGACACGCGAGGAACTATGACGACCAAGAAGCGAAAAACCGCCG
6 pmucA5200    18.6%  96.8%   ------------------------------CGCCGGGAACTCGACGCGGAGGACTATGACGACCAAGAAGCGAAAAACCGCCG
7 catBspQ1      8.4%  98.2%   ----------------------------------------------------------------------------
8 pmv23         5.8%  95.3%   ----------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   ----------------------------------------------------------------------------

                  cov   pid 3041          :         .         .         .         .         1         .         .  3120
1 pMDD7       100.0% 100.0%   GCGAGGACCTGGCAAAACAGGTCAGCGAGGCCAAGCAGGCCGCGTTGCTGAAACACACGAAGCAGCAGATCAAGGAAATG
2 95RKpn       16.5%  99.2%   ----------------------------------------------------------------------------
3 385Age       21.8%  95.0%   ----------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   ----------------------------------------------------------------------------
```

A15

```
  5 pmucA4400   16.8%  98.0%  GCGAGGACCTGG--------------------------------------------------------------------
  6 pmucA5200   18.6%  96.8%  GCGAGGACCTGGCAAAACAGGTCAGCGAGGCCAAGCAGGCCGCGTTGCTGAAACACACGAAGCAGCGATCAAGGAAATG
  7 catBspQ1     8.4%  98.2%  --------------------------------------------------------------------------------
  8 pmv23        5.8%  95.3%  --------------------------------------------------------------------------------
  9 amsacF      11.6%  99.5%  --------------------------------------------------------------------------------

              cov    pid  3121        .         .         :         .         .         .         .      2 3200
  1 pMDD7     100.0% 100.0% CAGCTTTCCTTGTTCGATATTGCGCCGTGGCCGGACACGATGCGAGCGATGCCAAACGACACGGCCCGCTCTGCCCTGTT
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% CAGCTTTCCTTGTTCGATATTGCGCCGTGGCCGGACACGATGCGAGCGATGCCAAACGACACGGCCCGCTCTGCCCTGTT
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3201        .         .         .         .         :         .         .      . 3280
  1 pMDD7     100.0% 100.0% CACCACGCGCAACAAGAAAATCCCGCGCGAGACGCTGCAAAACAAGGTCATTTTTCCACGTCAACAAGGACGTGAAGATCA
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% CACCACGCGCAACAAGAAAATCCCGCGCGAGACGCTGCAAAACAAGGTCATTTTTCCACGTCAACAAGGACGTGAAGATCA
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3281        .         3         .         .         .         .         :      . 3360
  1 pMDD7     100.0% 100.0% CCTACACCGGCGTCGAGCTGCGGGCCGACGATGACGAACTGGTGTGGCAGCAGGTGTTGGAGTACGCGAAGCGCACCCCT
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% CCTACACCGGCGTCGAGCTGCGGGCCGACGATGACGAACTGGTGTGGCAGCAGGTGTTGGAGTACGCGAAGCGCACCCCT
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3361        .         .         .         4         .         .         .      . 3440
  1 pMDD7     100.0% 100.0% ATCGGCGAGCCGATCACCTTCACGTTCTACGAGCTTTGCCAGGACCTGGGCTGGTCGATCAATGGCCGGTATTACACGAA
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% ATCGGCGAGCCGATCACCTTCACGTTCTACGAGCTTTGCCAGGACCTGGGCTGGTCGATCAATGGCCGGTATTACACGAA
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3441        :         .         .         :         .         5         .      . 3520
  1 pMDD7     100.0% 100.0% GGCCGAGGAATGCCTGTCGTGCCTACAGGCGACGGCGATGGGCTTCACGTCCGACCGCGTTGGGCACCTGGAATCGGTGT
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% GGCCGAGGAATGCCTGTCGTGCCTACAGGCGACGGCGATGGGCTTCACGTCCGACCGCGTTGGGCACCTGGAATCGGTGT
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3521        .         .         :         .         .         .         .      6 3600
  1 pMDD7     100.0% 100.0% CGCTGCTGCACCGCTTCCGCGTCCTGGACCGTGGCAAGAAAACGTCCCGTTGCCAGGTCCTGATCGACGAGGAAATCGTC
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% CGCTGCTGCACCGCTTCCGCGTCCTGGACCGTGGCAAGAAAACGTCCCGTTGCCAGGTCCTGATCGACGAGGAAATCGTC
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3601        .         .         .         .         :         .         .      . 3680
  1 pMDD7     100.0% 100.0% GTGCTGTTTGCTGGCGACCACTACACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGACGGCCCGACGGATGTT
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
  3 385Age     21.8%  95.0% --------------------------------------------------------------------------------
  4 AvPmsjekkB 22.9%  94.9% --------------------------------------------------------------------------------
  5 pmucA4400  16.8%  98.0% --------------------------------------------------------------------------------
  6 pmucA5200  18.6%  96.8% GTGCTGTTTGCTGGCGACCACTACACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGACGGCCCGACGGATGTT
  7 catBspQ1    8.4%  98.2% --------------------------------------------------------------------------------
  8 pmv23       5.8%  95.3% --------------------------------------------------------------------------------
  9 amsacF     11.6%  99.5% --------------------------------------------------------------------------------

              cov    pid  3681        .         7         .         .         .         .         :      . 3760
  1 pMDD7     100.0% 100.0% CGACTATTTCAGCTCGCACCGGGAGCCGTACCCGCTCAAGCTGGAAACCTTCCGCCTCATGTGCGGATCGGATTCCACCC
  2 95RKpn     16.5%  99.2% --------------------------------------------------------------------------------
```

A16

```
3 385Age       21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%   CGACTATTTCAGCTCGCACCGGGAGCCGTACCCGCTCAAGCTGGAAACCTTCCGCCTCATGTGCGGATCGGATTCCACCC
7 catBspQ1      8.4%  98.2%   -----------------------------------------ACAATCGGCCTATGTGCGGATCGGATTCCACCC
8 pmv23         5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  3761      .         .         .         8         .         .         .       . 3840
1 pMDD7        100.0% 100.0%   GCGTGAAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGAAGAGTTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTC
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   GCGTGAAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGAAGAGTTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTC
7 catBspQ1       8.4%  98.2%   GCGTGAAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGAAGAGTTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTC
8 pmv23          5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF        11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  3841      :         .         .         .         9         .         .       . 3920
1 pMDD7        100.0% 100.0%   AATGATGACCTGGTGCATTGCAAACGCTAGGGCCTTGTGGGGTCAGTTCCGGCTGGGGGTTCAGCAGCCACCTGCATCTC
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   AATGATGACCCTGTGCATTGGCAAACGCTAGGGC----------------------------------------------
7 catBspQ1       8.4%  98.2%   AATGATGACCTGGTGCATTGCAAACGCTAGGGCCTTGTGGGGTCAGTTCCGGCTGGGGGTTCAGCAGCCACCTGCATCTC
8 pmv23          5.8%  95.3%   -----------------------------------------ACTCTGGCGGGGGTCGCAGCCACCTGCATCTC
9 amsacF        11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  3921      .         .         :         .         .         .         .       0 4000
1 pMDD7        100.0% 100.0%   AAGCTCTGGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTCAGCTATCGCCATGTAAGCCCACTGCAAGCTA
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1       8.4%  98.2%   AAGCTCTGGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTCAGCTATCGCCATGTAAGCCCACTGCAAGCTA
8 pmv23          5.8%  95.3%   AAGCTCTGGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTCAGCTATCGCCATGTAAGCCCACTGCAAGCTA
9 amsacF        11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  4001      .         .         .         .         .         :         .       . 4080
1 pMDD7        100.0% 100.0%   CCTGCTTTCTCTTTGCGCGTTGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGGGGTCAGCACCGTTTC
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1       8.4%  98.2%   CCTGCTTTCTCTTTGCGCGTTGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGGGGTCAGCACCGTTTC
8 pmv23          5.8%  95.3%   CCTGCTTTCTCTTTGCGCGTTGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGGGGTCAGCACCGTTTC
9 amsacF        11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  4081      .         .         .         1         .         .         .       : . 4160
1 pMDD7        100.0% 100.0%   TGCGGACTGGCTTTCTACGTGGCTGCCATTTTTGGGGTGAGGTCGTTCGCGGCCGAGGGGCGCAGCCCCTGGGGGGATGG
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1       8.4%  98.2%   TGCGGACTGGCTTTCTACGTGGCTGCCATTTTTGGGGTGAGGCC------------------------------------
8 pmv23          5.8%  95.3%   TGCGGACTGGCTTTCTACGTGGCTGCCATTTTTGGGGTGAGGCCGTTCGCGGCCGAGGGGCGCAGCCCCTGGGGGGATGG
9 amsacF        11.6%  99.5%   --------------------------------------------------------------------------------

                  cov    pid  4161      .         .         .         2         .         .         .       . 4240
1 pMDD7        100.0% 100.0%   GGTGCCGCGTTAGCGGGCCGGGAGGGTTCGAGAAGGGGGGGCACCCCCCTGTCGGCGTGCGCGGTCACGCGCGC-CAGGGCGC
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1       8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23          5.8%  95.3%   GAG-----------------------------------------------------------------------------
9 amsacF        11.6%  99.5%   --------------------------------------CCCCCCTGTCGGCGTGCGCGGTCACGCGCGCACAGGGCGC

                  cov    pid  4241      :         .         .         .         .         3         .       . 4320
1 pMDD7        100.0% 100.0%   AGCCCTGGTTAAAAACAAGGTTTATAAATATTGGTTTAAAAGCAGGTTAAAAGACAGGTTAGCGGTGGCCGAAAAACGGG
2 95RKpn        16.5%  99.2%   --------------------------------------------------------------------------------
3 385Age        21.8%  95.0%   --------------------------------------------------------------------------------
4 AvPmsjekkB    22.9%  94.9%   --------------------------------------------------------------------------------
5 pmucA4400     16.8%  98.0%   --------------------------------------------------------------------------------
6 pmucA5200     18.6%  96.8%   --------------------------------------------------------------------------------
7 catBspQ1       8.4%  98.2%   --------------------------------------------------------------------------------
8 pmv23          5.8%  95.3%   --------------------------------------------------------------------------------
9 amsacF        11.6%  99.5%   AGCCCTGGTTAAAAACAAGGTTTATAAATATTGGTTTAAAAGCAGGTTAAAAGACAGGTTAGCGGTGGCCGAAAAACGGG

                  cov    pid  4321      .         .         :         .         .         .         .       4 4400
```

```
1 pMDD7       100.0% 100.0%  CGGAAACCCTTGCAAATGCTGGATTTTCTGCCTGTGGACAGCCCCTCAAATGTCAATAGGTGCGCCCCTCATCTGTCATC
2 95RKpn       16.5%  99.2%  --------------------------------------------------------------------------------
3 385Age       21.8%  95.0%  --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%  --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%  --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%  --------------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%  --------------------------------------------------------------------------------
8 pmv23         5.8%  95.3%  --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%  CGGAAACCCTTGCAAATGCTGGATTTTCTGCCTGTGGACAGCCCCTCAAATGTCAATAGGTGCGCCCCTCATCTGTCAGC

              cov    pid  4401    .         .         .         .         :         .         .         . 4480
1 pMDD7       100.0% 100.0%  ACTCTGCCCCTCAAGTGTCAAGGATCGCGCCCCTCATCTGTCAGTAGTCGCGCCCCTCAAGTGTCAATACCGCAGGGCAC
2 95RKpn       16.5%  99.2%  --------------------------------------------------------------------------------
3 385Age       21.8%  95.0%  --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%  --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%  --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%  --------------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%  --------------------------------------------------------------------------------
8 pmv23         5.8%  95.3%  --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%  ACTCTGCCCCTCAAGTGTCAAGGATCGCGCCCCTCATCTGTCAGTAGTCGCGCCCCTCAAGTGTCAATACCGCAGGGCAC

              cov    pid  4481    .         5         .         .         .         .         :         . 4560
1 pMDD7       100.0% 100.0%  TTATCCCCAGGCTTGTCCACATCATCTGTGGGAAACTCGCGTAAAATCAGGCGTTTTCGCCGATTTGCGAGGCTGGCCAG
2 95RKpn       16.5%  99.2%  --------------------------------------------------------------------------------
3 385Age       21.8%  95.0%  --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%  --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%  --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%  --------------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%  --------------------------------------------------------------------------------
8 pmv23         5.8%  95.3%  --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%  TTATCCCCAGGCTTGTCCACATCATCTGTGGGAAACTCGCGTAAAATCAGGCGTTTTCGCCGATTTGCGAGGCTGGCCAG

              cov    pid  4561    .         .         .         6         .         .         .         . 4640
1 pMDD7       100.0% 100.0%  CTCCACGTCGCCGGCCGAAATCGAGCCTGCCCCTCATCTGTCAACGCCGCGCCGGGTGAGTCGGCCCCTCAAGTGTCAAC
2 95RKpn       16.5%  99.2%  --------------------------------------------------------------------------------
3 385Age       21.8%  95.0%  --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%  --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%  --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%  --------------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%  --------------------------------------------------------------------------------
8 pmv23         5.8%  95.3%  --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%  CTCCACGTCGCCGGCCGAAATCGAGCCTGCCCCTCATCTGTCAACGCCGCGCCGGGTGAGTCGGCCCCTCAAGTGTCAAC

              cov    pid  4641    :         .         .         .         .         7         .         . 4720
1 pMDD7       100.0% 100.0%  GTCCGCCCCTCATCTGTCAGTGAGGGCCAAGTTTTCCGCGTGGTATCCACAACGCCGGCGGCCCTACATGGCTCTGCTGT
2 95RKpn       16.5%  99.2%  --------------------------------------------------------------------------------
3 385Age       21.8%  95.0%  --------------------------------------------------------------------------------
4 AvPmsjekkB   22.9%  94.9%  --------------------------------------------------------------------------------
5 pmucA4400    16.8%  98.0%  --------------------------------------------------------------------------------
6 pmucA5200    18.6%  96.8%  --------------------------------------------------------------------------------
7 catBspQ1      8.4%  98.2%  --------------------------------------------------------------------------------
8 pmv23         5.8%  95.3%  --------------------------------------------------------------------------------
9 amsacF       11.6%  99.5%  GTCCGCCCCTCATCTGTCAGTGAGGGCCAAGTTTTCCGCGAGGTATCCACAACGCCGGCGGCCCTACATGGCTCTGCTGT

              cov    pid  4721    .         .         :         ] 4757
1 pMDD7       100.0% 100.0%  AGTGAGTGGGTTGCGCTCCTAGGTCAGGCTGCGCAAC
2 95RKpn       16.5%  99.2%  -------------------------------------
3 385Age       21.8%  95.0%  -------------------------------------
4 AvPmsjekkB   22.9%  94.9%  -------------------------------------
5 pmucA4400    16.8%  98.0%  -------------------------------------
6 pmucA5200    18.6%  96.8%  -------------------------------------
7 catBspQ1      8.4%  98.2%  -------------------------------------
8 pmv23         5.8%  95.3%  -------------------------------------
9 amsacF       11.6%  99.5%  AGTGAGTGGGTTGCGCTCCTAGGTCAGGCTGCGCAAC
```

MView 1.63, Copyright © 1997-2018 Nigel P. Brown

A18

# Appendix F   Sequencing of pMDD11

The sequencing results of pMDD11 are presented in the next pages. The first line shows the template pMDD11, while the two successive lines show the alignment of M13F and M13R. A few errors were found in the alignments of M13F, but since M13R is correct, these were ignored. "cov" shows how much of the template the aligned sequence covers in percentage. "pid" shows the percentage of identity between the aligned sequence and the template.

Reference sequence (1): pMDD11
Identities normalised by aligned length.
Colored by: identity

|   | cov | pid | 1 ... 80 |
|---|-----|-----|---|
| 1 pMDD11 | 100.0% | 100.0% | |
| 2 M13F | 82.2% | 99.7% | |
| 3 M13R | 99.9% | 83.8% | |

A20

```
2 M13F   82.2%  99.7%  TTTAATCACAATTCAGAAAATATCATAATATCTCATTTCACTAAATAATAGTGAACGGCAGGTATATGTGATGGGTTAAA
3 M13R   99.9%  83.8%  TAAATCCAATTCAAAAATATCATATATCTCATTTCCTAATAATAGTGACGGCAGGAATTGGGATGGGTTAAAAGGATCGA

            cov    pid 1281        .         3         .         .         .         .         :         . 1360
1 pMDD11 100.0% 100.0% AAGGATCGATCCTCTAGAGAATTTCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATT
2 M13F    82.2%  99.7% AAGGATCGATCCTCTAGAGAATTTCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATT
3 M13R    99.9%  83.8% TCCCTAAAAAATTCAAAAAAGCCCCGATCCGGGCCTTTTTCAGTTGAAACGGACTACCTATTTTAATTGCTATTCAAGC

            cov    pid 1361        .         .         .         4         .         .         .         . 1440
1 pMDD11 100.0% 100.0% TTAACTTGCTATTTCTAGCTCTAAAACCATGGACCGCTGCGGCGATGACTAGTAAGGGCGAATTCTGCAGATATCCATCA
2 M13F    82.2%  99.7% TTAACTTGCTATTTCTAGCTCTAAAACCATGGACCGCTGCGGCGATGACTAGTAAGGGCGAATTCTGCAGATATCCATCA
3 M13R    99.9%  83.8% TAAAACCTGGACCCTGGGGGGAATAAATTAAAGGGGGAAATTCGGAAATTCCCAAATGGGGGCCCCAAAAATTTAAAGGG

            cov    pid 1441        :         .         .         . ] 1476
1 pMDD11 100.0% 100.0% CACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCC
2 M13F    82.2%  99.7% CACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCC
3 M13R    99.9%  83.8% CCAATTCCCCAAAGGGGGGTTAAAATTTGGGGCCGTT
```

# Appendix G   Sequencing of pMDD13

The sequencing results of pMDD13 are presented in the next pages. The first line shows the template pMDD13 and the next line the alignment of primer uidAF. The alignment show that the N20 sequence is correct.

```
                cov    pid     1 [        .         .         .         .         :         .         .         .  80
1 pMDD13       100.0% 100.0%     GCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTT
2 uidAF        100.0%  99.6%     GCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTT

                cov    pid    81            .         1         .         .         .         .         :         . 160
1 pMDD13       100.0% 100.0%     GCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCCTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGACCGG
2 uidAF        100.0%  99.6%     GCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCCTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGACCGG

                cov    pid   161            .         .         .         2         .         .         .         . 240
1 pMDD13       100.0% 100.0%     ATGCCGACGCGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
2 uidAF        100.0%  99.6%     ATGCCGACGCGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC

                cov    pid   241            :         .         .         .         .         3         .         . 320
1 pMDD13       100.0% 100.0%     GAGTCGGTGCTTTTTTTTGAATTCTCTAGAGGATCGATCCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTA
2 uidAF        100.0%  99.6%     GAGTCGGTGCTTTTTTTTGAATTCTCTAGAGGATCGATCCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTA

                cov    pid   321            .         .         :         .         .         .         .         4 400
1 pMDD13       100.0% 100.0%     GTGAAATGAGATATTATGATATTTTTCTGAATTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAA
2 uidAF        100.0%  99.6%     GTGAAATGAGATATTATGATATTTTTCTGAATTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAA

                cov    pid   401            .         .         .         .         :         .         .         . 480
1 pMDD13       100.0% 100.0%     TACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTTATGATTTTCTATCA
2 uidAF        100.0%  99.6%     TACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTTATGATTTTCTATCA

                cov    pid   481            .         5         .         .         .         .         :         . 560
1 pMDD13       100.0% 100.0%     AACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTT
2 uidAF        100.0%  99.6%     AACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTT

                cov    pid   561            .         .         .         6         .         .         .         . 640
1 pMDD13       100.0% 100.0%     GTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGG
2 uidAF        100.0%  99.6%     GTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGG

                cov    pid   641            :         .         .         .         .         7         .         . 720
1 pMDD13       100.0% 100.0%     TCTTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAA
2 uidAF        100.0%  99.6%     TCTTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAA

                cov    pid   721            .         .         .         .         :         .         .         8 800
1 pMDD13       100.0% 100.0%     AAAAGGAGACATGAACGATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCA
2 uidAF        100.0%  99.6%     AAAAGGAGACATGAACGATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCA

                cov    pid   801            .         .         .         .         :         .         .         . 880
1 pMDD13       100.0% 100.0%     GGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACG
2 uidAF        100.0%  99.6%     GGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACG

                cov    pid   881            .         9         .         .         .         .         :         . 960
1 pMDD13       100.0% 100.0%     CCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAA
2 uidAF        100.0%  99.6%     CCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAA

                cov    pid   961            .         .         .         0         .         .         .         . 1040
1 pMDD13       100.0% 100.0%     ATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCAC
2 uidAF        100.0%  99.6%     ATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCAC

                cov    pid  1041            :         .         .         .         .         1         .         . 1120
1 pMDD13       100.0% 100.0%     GGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGT
2 uidAF        100.0%  99.6%     GGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGT

                cov    pid  1121            .         .         .         .         :         .         .         2 1200
1 pMDD13       100.0% 100.0%     CGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTA
2 uidAF        100.0%  99.6%     CGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTA

                cov    pid  1201            .         .         .         .         :         .         .         . 1280
1 pMDD13       100.0% 100.0%     TCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTCACACTGAT
2 uidAF        100.0%  99.6%     TCCTAAAAGACCAAACCCAAGAATGGTCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTAACCGGAT

                cov    pid  1281            . ] 1292
1 pMDD13       100.0% 100.0%     TTCTCCGGTAAA
2 uidAF        100.0%  99.6%     TTCCCCGGTAAA
```

A23

# Appendix H   Sequencing of pMDD14

The sequencing results of pMDD14 are presented in the next pages. The first line shows the template pMDD14. The next lines show in successive order the alignment of primer M13R, OuidAR, M13F, and NuidAF. Since this sequence was a mutation template designed to implement errors when used as a template for HDR, a few errors could be accepted. A few errors can be shown in one of the aligned sequences at a time, but since the other sequences are correct, these errors can be ignored. "cov" shows how much of the template the aligned sequence covers in percentage. "pid" shows the percentage of identity between the aligned sequence and the template.

```
Reference sequence (1): pMDD14
Identities normalised by aligned length.
Colored by: identity

             cov    pid    1 [          .          .          .          :          .          .          .         . 80
1 pMDD14  100.0% 100.0%   CTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTC
2 M13R     52.7%  99.2%   CTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTC
3 OuidAR   49.3%  93.4%   -------ATACTCAAGCTATGCATCAAGGCTTGGTACCGAGGTCGGATCAACTAGTAACGCGCCAGTGTGCCTGGAATT
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid   81            .          .          1          .          .          .          :         . 160
1 pMDD14  100.0% 100.0%   GCCCTTGATATCAAGCTTATCGATACCGCACCGAAGTTCAT-GCCAGTCCAGCGTTTTTGCAGCAGAAAAGCCGCCGACT
2 M13R     52.7%  99.2%   GCCCTTGATATCAAGCTTATCGATACCGCACCGAAGTTCAT-GCCAGTCCAGCGTTTTTGCAGCAGAAAAGCCGCCGACT
3 OuidAR   49.3%  93.4%   CGCCCTTGATATCAAGCTTATCGATACCGCACGAAGTTCATGGCCAGTCCAGCGTTTTTGCAGCAGAAAAGCCGCCGACT
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  161            .          .          .          2          .          .          .         . 240
1 pMDD14  100.0% 100.0%   TCGGTTTGCGGTCGCGAGTGAAGATCCCTTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCG
2 M13R     52.7%  99.2%   TCGGTTTGCGGTCGCGAGTGAAGATCCCTTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCG
3 OuidAR   49.3%  93.4%   CCGGTTTGCGGTCGCGAGTGAAGATCCCTTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCG
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  241            :          .          .          .          .          3          .         . 320
1 pMDD14  100.0% 100.0%   AAATTCCATACCTGTTCACCGACGACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATA
2 M13R     52.7%  99.2%   AAATTCCATACCTGTTCACCGACGACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATA
3 OuidAR   49.3%  93.4%   AAATTCCATACCTGTTCACCGACGACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATA
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  321            .          .          :          .          .          .          4 400
1 pMDD14  100.0% 100.0%   CTCTTCACTCCACATGTCGGTGTACATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGTGATGATAATCGGCT
2 M13R     52.7%  99.2%   CTCTTCACTCCACATGTCGGTGTACATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGTGATGATAATCGGCT
3 OuidAR   49.3%  93.4%   CTCTTCACTCCACATGTCGGTGTACATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGTGATGATAATCGGCT
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  401            .          .          .          .          .          .          .         . 480
1 pMDD14  100.0% 100.0%   GATGCAGTTTCTCCTGCCAGGCCAGAAGTTCTTTTTTCCAGTACCTTCTCTGCCGTTTCCAAATCGCCGCTTTGGACATAC
2 M13R     52.7%  99.2%   GATGCAGTTTCTCCTGCCAGGCCAGAAGTTCTTTTTTCCAGTACCTTCTCTGCCGTTTCCAAATCGCCGCTTTGGACATAC
3 OuidAR   49.3%  93.4%   GATGCAGTTTCTCCTGCCAGGCCAGAAGTTCTTTTTTCCAGTACCTTCTCTGCCGTTTCCAAATCGCCGCTTTGGACATAC
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  481            .          .          5          .          .          .          :         . 560
1 pMDD14  100.0% 100.0%   CATCCGTAATAACGGTTCAGGCACAGCACATCAAAGAGATCGCTGATGGTATCGGTGTGAGCGTCGCAGAACATTACATT
2 M13R     52.7%  99.2%   CATCCGTAATAACGGTTCAGGCACAGCACATCAAAGAGATCGCTGATGGTATCGGTGTGAGCGTCGCAGAACATTACATT
3 OuidAR   49.3%  93.4%   CATCCGTAATAACGGTTCAGGCACAGCACATCAAAGAGATCGCTGATGGTATCGGTGTGAGCGTCGCAGAACATTACATT
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  561            .          .          .          6          .          .          .         . 640
1 pMDD14  100.0% 100.0%   GACGCAGGTGATCGGACGCGTCGGGTCGAGTTTACGCGTTGCTTCCGCCAGTGGCGCGAAATATTCCCGTGCACCTTGCG
2 M13R     52.7%  99.2%   GACGCAGGTGATCGGACGCGTCGGGTCGAGTTTACGCGTTGCTTCCGCCAGTGGCGCGAAATATTCCCGTGCACCTTGCG
3 OuidAR   49.3%  93.4%   GACGCAGGTGATCGGACGCGTCGGGTCGAGTTTACGCGTTGCTTCCGCCAGTGGCGCGAAATATTCCCGTGCACCTTGCG
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  641            :          .          .          .          .          7          .         . 720
1 pMDD14  100.0% 100.0%   GACGGGTATCCGGTTCGTTGGCAATACTCCACATCACCACGCTTGGGTGGTTTTTGTCACGCGCTATCAGCTCTTTAATC
2 M13R     52.7%  99.2%   GACGGGTATCCGGTTCGTTGGCAATACTCCACATCACCACGCTTGGGTGGTTTTTGTCACGCGCTATCAGCTCTTTAATC
3 OuidAR   49.3%  93.4%   GACGGGTATCCGGTTCGTTGGCAATACTCCACATCACCACGCTTGGGTGGTTTTTGTCACGCGCTATCAGCTCTTTAATC
4 M13F     57.2%  90.2%   --------------------------------------------------------------------------------
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  721            .          .          .          .          :          .          8 800
1 pMDD14  100.0% 100.0%   GCCTGTAAGTGCGCTTGCTGAGTTTCCCCGTTGACTGCCTCTTCGCTGTACAGTTCTTTCGGCTTGTTGCCCGCTTCGAA
2 M13R     52.7%  99.2%   GCCTGTAAGTGCGCTTGCTGAGTTTCCCCGTTGACTGCCTCTTCGCTGTACAGTTCTTTCGGCTTGTTGCCCGCTTCGAA
3 OuidAR   49.3%  93.4%   GCCTGTAAGTGCGCTTGCTGAGTTTCCCCGTTGACTGCCTCTTCGCTGTACAGTTCTTTCGGCTTGTTGCCCGCTTCGAA
4 M13F     57.2%  90.2%   ----------------------TGAGCTTCTGACGCTGTGAACAGTTTCTTACGTGCTGTTGCCCGGATTCGAATCAAT
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------

             cov    pid  801            .          .          .          .          :          .          .         . 880
1 pMDD14  100.0% 100.0%   ACCAATGCCTAAAGAGAGGTTAAAGCCGACAGCAGCAGTTTCATCAATCACCACGATGCCATGTTCATCTGCCCAGTCGA
2 M13R     52.7%  99.2%   ACCAATGCCTAAAGAGAGGTTAAAGCCGACAGCAGCAGTTTCATCAATCACCACGATGCCATGTTCATCTGCCCAGTCGA
3 OuidAR   49.3%  93.4%   ACCAATGCCTAAAGAGAGGTTAAAGCCGACAGCAGCAGTTTCATCAATCACCACGATGCCATGTTCATCTGCCCAGTCGA
4 M13F     57.2%  90.2%   GCCTAAAGAAGAGGTTAAAGCCGACAAGCAGGCAGTTTCATCAATTCACCCACGATGTCATGTCATCTGGTCCCAGTCGA
5 NuidAF   48.3%  98.7%   --------------------------------------------------------------------------------
```

A25

```
                cov     pid  881        .           9         .         .                  .         .         .     960
1 pMDD14 100.0% 100.0%  GCATCTCTTCAGCGTAAGGGTAATGCGAGGTACGGTAGGAG---TTGGCCCCAATCCAGTCCATTAATTGCGTGGTCGTG
2 M13R    52.7%  99.2%  GCATCTCTTCAGCGTAAGGGTAATGCGAGGTACGGTAGGAG---TTGGCCCCAATCCAGTCCATTAATTGCGTGGTCGTG
3 OuidAR  49.3%  93.4%  GCATCTCTTCAGCGTAAGGGTAATGCGAGGTACGGTAGGAG---TTGGCCCCAATCCAGTCCATTAATTGCGTGGTCGTG
4 M13F    57.2%  90.2%  GCATCTTTTCAGCGTAAAGGTATTGCGAGGTACGGTAGGAGAGTTTGGCCCCAAATCCAGTCCATTAATGGCGTGGTCGTG
5 NuidAF  48.3%  98.7%  -------------------------------------------------------------------------------

                cov     pid  961        .           .         .                  0         .         .         .    1040
1 pMDD14 100.0% 100.0%  CACCATCAG-CACGTTATCG-AATCCTTTGCCACGCAAGTCCGCATCTTCATGACGACC-AAAGCCAGTAAAGTAGAACG
2 M13R    52.7%  99.2%  CACCATCAG-CACGTTATCG-AATCCTTTGCCACGCAAGTCCGCATCTTCATGACGACC-AAAGCCAGTAAAGTAGAACG
3 OuidAR  49.3%  93.4%  CACCATCAG-CACGTTATCG-AATCCTTTGCCACGCAAGTCCGCATCTTCATGACGACC-AAAGCCAGTAAAGTAGAACG
4 M13F    57.2%  90.2%  CACCATCAGCCACGTTATCGAAATCCTTTGCCACGCAAGTCCGCATCTTCATGACGACCAAAAGCCAGTAAAGTAGAACG
5 NuidAF  48.3%  98.7%  -------------------------------------------------------------------------------

                cov     pid 1041        :           .         .                  .         1         .         .    1120
1 pMDD14 100.0% 100.0%  GTTTGTGGTTAATCAGGAACTGTTCGCCCTTCACTGCCACTGACCGGAGCTGTGACGCACAGTTCATAGAGATAAACCTTC
2 M13R    52.7%  99.2%  GTTTGTGGTTAATCAGGAACTGTTCGCCCTTCACTGCCACTGACCGGAGCTGGGACCCACAGTTCAAAAGAAGACCCTC
3 OuidAR  49.3%  93.4%  GTTTGTGGTTAATCAGAGCGATTAG---------------------------------------------------------
4 M13F    57.2%  90.2%  G-TTGTGGTTAATCAGGAACTGTTCGCCCTTCACTGCCACTGACCGGAGCTGTGACGCACAGTTCATAGAGATAAACCTTC
5 NuidAF  48.3%  98.7%  ----------------------------------------------------------------AAAGATACGTTC

                cov     pid 1121        .           .         :         .                  .         .         2   1200
1 pMDD14 100.0% 100.0%  ACCCGGTTGCCAGAGGTGCGGATCACCCACTTGCAAAGTCCCGCTAGTGCCTTGTCCAGTTGCAACCACCTGTTGATCCG
2 M13R    52.7%  99.2%  CCCCGGGTG------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  ACCCGGTTGCCAGAGGTGCGGATCACCCACTTGCAAAGTCCCGCTAGTGCCTTGTCCAGTTGCAACCACCTGTTGATCCG
5 NuidAF  48.3%  98.7%  ACCCGGTTGCCAGAGGTGCGGATCACCCACTTGCAAAGTCCCGCTAGTGCCTTGTCCAGTTGCAACCACCTGTTGATCCG

                cov     pid 1201        .           .         .         .                  .         .         .   1280
1 pMDD14 100.0% 100.0%  CATCACGCAGTTCAACGCTGACATCACCATTGGCCACCACCTGCCAGTCAACAGACGCGTGGTTACAGTCTTGCGCGACA
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  CATCACGCAGTTCAACGCTGACATCACCATTGGCCACCACCTGCCAGTCAACAGACGCGTGGTTACAGTCTTGCGCGACA
5 NuidAF  48.3%  98.7%  CATCACGCAGTTCAACGCTGACATCACCATTGGCCACCACCTGCCAGTCAACAGACGCGTGGTTACAGTCTTGCGCGACA

                cov     pid 1281        .           .         3         .                  .         .         .   1360
1 pMDD14 100.0% 100.0%  TGCGTCACCACGGTGATATCGTCCACCCAGGTGTTCGGCGTGGTGTAGAGCATTACGCTGCGATGGATCCCGGCATAGTT
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  TGCGTCACCACGGTGATATCGTCCACCCAGGTGTTCGGCGTGGTGTAGAGCATTACGCTGCGATGGATCCCGGCATAGTT
5 NuidAF  48.3%  98.7%  TGCGTCACCACGGTGATATCGTCCACCCAGGTGTTCGGCGTGGTGTAGAGCATTACGCTGCGATGGATCCCGGCATAGTT

                cov     pid 1361        .           .         .         .                  4         .         .   1440
1 pMDD14 100.0% 100.0%  AAAGAAATCATGGAAGTAAGACTGCTTTTTCTTGCCGTTTTCGTCGGTAATCACCATTCCCGGCGGGATAGTCTGCCAGT
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  AAAGAAATCATGGAAGTAAGACTGCTTTTTCTTGCCGTTTTCGTCGGTAATCACCATTCCCGGCGGGATAGTCTGCCAGT
5 NuidAF  48.3%  98.7%  AAAGAAATCATGGAAGTAAGACTGCTTTTTCTTGCCGTTTTCGTCGGTAATCACCATTCCCGGCGGGATAGTCTGCCAGT

                cov     pid 1441        :           .         .         .                  .         5         .   1520
1 pMDD14 100.0% 100.0%  TCAGTTCGTTGTTCACACAAACGGTGATACGTACACTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGC
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  TCAGTTCGTTGTTCACACAAACGGTGATACGTACACTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGC
5 NuidAF  48.3%  98.7%  TCAGTTCGTTGTTCACACAAACGGTGATACGTACACTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGC

                cov     pid 1521        .           .         .         .                  .         .         6   1600
1 pMDD14 100.0% 100.0%  GTATAGCCGCCCTGATGCTCCATCACTTCCTGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAG
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  GTATAGCCGCCCTGATGCTCCATCACTTCCTGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAG
5 NuidAF  48.3%  98.7%  GTATAGCCGCCCTGATGCTCCATCACTTCCTGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAG

                cov     pid 1601        .           .         .         .                  .         .         .   1680
1 pMDD14 100.0% 100.0%  CACGATACGCTGGCCTGCCCAACCTTTCGGTATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATAT
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  CACGATACGCTGGCCTGCCCAACCTTTCGGTATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATAT
5 NuidAF  48.3%  98.7%  CACGATACGCTGGCCTGCCCAACCTTTCGGTATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATAT

                cov     pid 1681        .           7         .         .                  .         .         :   1760
1 pMDD14 100.0% 100.0%  CTGCATCGGCGAACTGATCGTTAAAACTGCCTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGC
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  CTGCATCGGCGAACTGATCGTTAAAACTGCCTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGC
5 NuidAF  48.3%  98.7%  CTGCATCGGCGAACTGATCGTTAAAACTGCCTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGC

                cov     pid 1761        .           .         .         8         .                  .         .   1840
1 pMDD14 100.0% 100.0%  TGATCAATTCCACAGTTTTCGCGATCCAGACTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTC
2 M13R    52.7%  99.2%  -------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4%  -------------------------------------------------------------------------------
4 M13F    57.2%  90.2%  TGATCAATTCCACAGTTTTCGCGATCCAGACTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTC
5 NuidAF  48.3%  98.7%  TGATCAATTCCACAGTTTTCGCGATCCAGACTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTC
```

A26

```
                cov    pid 1841      :          .           .          .          .           9           .         . 1920
1 pMDD14 100.0% 100.0% TACAGGACGTAACATAAGGGACTCCTCGAAGCTTAGATCTATTACCCAAGGGCGAATTCTGCAGATATCCATCACACTGG
2 M13R    52.7%  99.2% --------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4% --------------------------------------------------------------------------------
4 M13F    57.2%  90.2% TACAGGACGTAACATAAGGGACTCCTCGAAGCTTAGATCTATTACCCAAGGGCGAATTCTGCAGATATCCATCACACTGG
5 NuidAF  48.3%  98.7% TACAGGACGTAACATAAGGGACTCCTCGAAGCTTAGATCTATTACCCAAGGGCGAATTCTGCAGATATCCATCACACTGG

                cov    pid 1921              .            :           .          .           .            . 0 2000
1 pMDD14 100.0% 100.0% CGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACA
2 M13R    52.7%  99.2% --------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4% --------------------------------------------------------------------------------
4 M13F    57.2%  90.2% CGGCCGCTCGAGCATGCATCTAGAGGGCCCAATCGCCGTAATAAATTCG-------------------------------
5 NuidAF  48.3%  98.7% CGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACA

                cov    pid 2001              .            :           .          .           :            . 2080
1 pMDD14 100.0% 100.0% ACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA
2 M13R    52.7%  99.2% --------------------------------------------------------------------------------
3 OuidAR  49.3%  93.4% --------------------------------------------------------------------------------
4 M13F    57.2%  90.2% --------------------------------------------------------------------------------
5 NuidAF  48.3%  98.7% ACGTCGTGACTGGGAAAACCCT-GCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA

                cov    pid 2081          .            1           .          .           . ] 2136
1 pMDD14 100.0% 100.0% GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGG
2 M13R    52.7%  99.2% -------------------------------------------------------
3 OuidAR  49.3%  93.4% -------------------------------------------------------
4 M13F    57.2%  90.2% -------------------------------------------------------
5 NuidAF  48.3%  98.7% GCGAAGAGGCCCGCACCGATCGCCCTTTCCCAACAGTGCGCAGCCTATACGTACGG
```

MView 1.63, Copyright © 1997-2018 Nigel P. Brown

A27

# Appendix I   Sequencing of pMDD15

The sequencing results of pMDD15 are presented in the next pages. The first line shows the template pMDD15. The next lines show in successive order the alignment of primer mDD10R, sacBsekv, sjekksacBF, OuidAR, NuidAF, and sjekkAmdelR. As for pMDD14, this sequence was a mutation template designed to implement errors when used as a template for HDR, and a few errors could be accepted. There are a lot of errors at the beginning of the alignment of primer mDD10R because this primer was designed to be complementary to the corresponding region in pMDD10, where the N20 sequence had not yet been implemented. A few errors can be shown in one of the aligned sequences at a time, but since the other sequences are correct, these errors can be ignored. "cov" shows how much of the template the aligned sequence covers in percentage. "pid" shows the percentage of identity between the aligned sequence and the template.

Reference sequence (1): pMDD15
Identities normalised by aligned length.
Colored by: identity

```
                    cov    pid    1 [                    .                    .                    :                    .                    .                    :                    .                  . 80
1 pMDD15        100.0% 100.0%      TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGACCGGATGCCGACGCGAAGGTTTTAGAGCTAGAAATAGCAAG
2 mDD10R         33.2%  94.4%      --------ACGTTAGTCATTCTTGGTTTAATTTTGTCAGCCCGATCCCACCGGAAGGTTTTGAGCTAGAAATGGCAGT
3 sacBsekv       27.2%  96.4%      --------------------------------------------------------------------------------
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid   81                    .                    1                    .                    .                    .                    :                    .                  . 160
1 pMDD15        100.0% 100.0%      TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGCTGCTTTTTTTGAATTCTCTAGAGGATCGAT
2 mDD10R         33.2%  94.4%      T---AAATAAGCTAGTCCG-TATCAACT--GAAAAGTGCACCGAGTCGGT--CTTTTTTGAATTCTTTAGAGGATCGAT
3 sacBsekv       27.2%  96.4%      --------------------------------------------------------------------------------
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  161                    .                    .                    .                    .                    2                    .                    .                  . 240
1 pMDD15        100.0% 100.0%      CCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGAT
2 mDD10R         33.2%  94.4%      CCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGAT
3 sacBsekv       27.2%  96.4%      --------------------------------------------------------------------------------
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  241                    :                    .                    .                    .                    .                    3                    .                  . 320
1 pMDD15        100.0% 100.0%      TAAAAAGGCAACTTTATG-CCCATGCAACAGAAACTATAAAAAATACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGT
2 mDD10R         33.2%  94.4%      TAAAAAGGCAACTTTATGCCCATGCAACAGAAACTAT-AAAAATACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGT
3 sacBsekv       27.2%  96.4%      -----------ACCTTTGCCCTAGCACAGTAACTTTAAAAAAATTCAGGGAAGGATAAGGAACTGATAGATTTTTTTAGT
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  321                    .                    .                    :                    .                    .                    .                    4 400
1 pMDD15        100.0% 100.0%      TCTTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTGCAATCCAAAC
2 mDD10R         33.2%  94.4%      TCTTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTGCAATCCAAAC
3 sacBsekv       27.2%  96.4%      TCTTTTAGGCCCGTAGTCTGC-AATTCCTTTTATGATTTTCTTTCAAACAAAGGAGGAAAATAGACCAGTTGCAATCCAAAC
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  401                    .                    .                    .                    .                    :                    .                  . 480
1 pMDD15        100.0% 100.0%      GAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAA
2 mDD10R         33.2%  94.4%      GAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAA
3 sacBsekv       27.2%  96.4%      GAGAGTCTAATAGAATGAGGTCGACAAGTAAATCGCGCGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAA
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  481                    .                    5                    .                    .                    .                    .                    :                  . 560
1 pMDD15        100.0% 100.0%      AGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTC
2 mDD10R         33.2%  94.4%      AGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTC
3 sacBsekv       27.2%  96.4%      AGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTC
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  561                    .                    .                    .                    6                    .                    .                  . 640
1 pMDD15        100.0% 100.0%      AAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACATCAAAAAGTTTG
2 mDD10R         33.2%  94.4%      AAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACATCAAAAAGTTTG
3 sacBsekv       27.2%  96.4%      AAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACATCAAAAAGTTTG
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------

                    cov    pid  641                    :                    .                    .                    .                    .                    7                    .                  . 720
1 pMDD15        100.0% 100.0%      CAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAAC
2 mDD10R         33.2%  94.4%      CAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAAC
3 sacBsekv       27.2%  96.4%      CAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAAC
4 sjekksacBF     34.8%  93.5%      --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%      --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%      --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%      --------------------------------------------------------------------------------
```

A29

```
                       cov    pid   721                .                    :                    .                    .                    .             8  800
1 pMDD15        100.0% 100.0%        CAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAA
2 mDD10R         33.2%  94.4%        CAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAA
3 sacBsekv       27.2%  96.4%        CAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAA
4 sjekksacBF     34.8%  93.5%        --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid   801                .                    .                    .                    :                    .             880
1 pMDD15        100.0% 100.0%        TGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACA
2 mDD10R         33.2%  94.4%        TGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACA
3 sacBsekv       27.2%  96.4%        TGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACA
4 sjekksacBF     34.8%  93.5%        --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid   881                .                    9                    .                    .                    .             960
1 pMDD15        100.0% 100.0%        GCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCT
2 mDD10R         33.2%  94.4%        GCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCT
3 sacBsekv       27.2%  96.4%        GCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCT
4 sjekksacBF     34.8%  93.5%        --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid   961                .                    .                    0                    .                    .             1040
1 pMDD15        100.0% 100.0%        AAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGG
2 mDD10R         33.2%  94.4%        AAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGG
3 sacBsekv       27.2%  96.4%        AAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGG
4 sjekksacBF     34.8%  93.5%        --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1041                :                    .                    .                    .                    1             1120
1 pMDD15        100.0% 100.0%        CCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAG
2 mDD10R         33.2%  94.4%        CCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAG
3 sacBsekv       27.2%  96.4%        CCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAG
4 sjekksacBF     34.8%  93.5%        --------------------------------------------------------------------------------
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1121                :                    .                    :                    .                    2             1200
1 pMDD15        100.0% 100.0%        CCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACTGATTCTCCGGTAAACATTACGGCAAACAAACACTGACA
2 mDD10R         33.2%  94.4%        CCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACTGATTCTCCGGTAAACATTACGGCAAACAAACACTGACA
3 sacBsekv       27.2%  96.4%        CCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACTGATTCTCCGGT-AACATTACGGCAAACAAACACTGACA
4 sjekksacBF     34.8%  93.5%        -----------------GGAGGAACCGTTTAATTCTACACTGATTCTCCGGT-AACATTACGGCAAACAAACACTGACA
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1201                .                    .                    .                    :                    .             1280
1 pMDD15        100.0% 100.0%        ACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTGAGAGGATTATAAATCAATCTTTGACGG
2 mDD10R         33.2%  94.4%        ACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTGAGAGGATTATAAATCAATC-TTGACGG
3 sacBsekv       27.2%  96.4%        ACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTGAGAGGATTATAAATCAATC-TTGACGG
4 sjekksacBF     34.8%  93.5%        ACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTGAGAGGATTATAAATCAATCTTTGACGG
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1281                .                    3                    .                    .                    .             1360
1 pMDD15        100.0% 100.0%        TGACGGAAAAACGTATCAAA-ATGTACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGA
2 mDD10R         33.2%  94.4%        TGACGGAAAAACGTATCAAA-ATGTACAGCAGTTCTCGTGACGCCC----------------------------------
3 sacBsekv       27.2%  96.4%        TGACGGAAAAACGTATCAAACATGTACAGCAGTTATCAAACCCC-------------------------------------
4 sjekksacBF     34.8%  93.5%        TGACGGAAAAACGTATCAAA-ATGTACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGA
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1361                .                    .                    .                    4                    .             1440
1 pMDD15        100.0% 100.0%        GATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGG
2 mDD10R         33.2%  94.4%        --------------------------------------------------------------------------------
3 sacBsekv       27.2%  96.4%        --------------------------------------------------------------------------------
4 sjekksacBF     34.8%  93.5%        GATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGG
5 OuidAR         23.6%  94.7%        --------------------------------------------------------------------------------
6 NuidAF         21.8%  92.1%        --------------------------------------------------------------------------------
7 sjekkAmdelR    16.1%  95.0%        --------------------------------------------------------------------------------

                       cov    pid  1441                :                    .                    .                    5                    .             1520
1 pMDD15        100.0% 100.0%        CGAAGAATCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCTGCAAA
2 mDD10R         33.2%  94.4%        --------------------------------------------------------------------------------
3 sacBsekv       27.2%  96.4%        --------------------------------------------------------------------------------
4 sjekksacBF     34.8%  93.5%        CGAAGAATCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCTGCAAA
```

A30

```
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1521   .         .         :         .         .         .       6 1600
1 pMDD15      100.0% 100.0%  GCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACACTGAAAAAA
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  GCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACACTGAAAAAA
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1601   .         .         :         .         .         .         . 1680
1 pMDD15      100.0% 100.0%  GTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAACGTCTTTTAAAATGAACGGCAAATG
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  GTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAACGTCTTTTAAAATGAACGGCAAATG
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1681   .         .       7 .         .         .         .       : . 1760
1 pMDD15      100.0% 100.0%  GTACCTGTTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATG
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  GTACCTGTTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATG
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1761   .         .         .       8 .         .         .         . 1840
1 pMDD15      100.0% 100.0%  TTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGAT
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  TTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGAT
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1841   :         .         .         .         .       9 .         . 1920
1 pMDD15      100.0% 100.0%  GTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACAG
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  GTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACAG
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  1921   .         .         :         .         .       0 .         . 2000
1 pMDD15      100.0% 100.0%  AGGATTCTACGCAGACAAACAATCAACGTTTGCGCCAAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCA
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  AGGATTCTACGCAGACAAACAATCAACGTTTGCGCCAAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCA
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  2001   .         .         .         .         :         .         . 2080
1 pMDD15      100.0% 100.0%  AAGACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATCAAGCTTATCGA
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  AAGACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATCAAGCTTATCGA
5 OuidAR       23.6%  94.7%  ------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  2081   .         .       1 .         .         .         :         . 2160
1 pMDD15      100.0% 100.0%  TACCGCACCGAAGTTCATGCCAGTCCAGCGTTTTTGCAGCAGAAAAGCCGCCGACTTCGGTTTGCGGTCGCGAGTGAAGA
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  TACCGCACCGAAGTTCATGCCAGTCCAGCGTTTTTGCAGCAGAAAAGCCGCCGACTTCGGTTTGCGGTCGCGAGTGAAGA
5 OuidAR       23.6%  94.7%  ------------------------CCAGCGTATTCAGCAGAAAAGAAGTCCAGCTTCGATTGCGGTCACGATGAAGA
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  2161   .         .         .       2 .         .         .         . 2240
1 pMDD15      100.0% 100.0%  TCCCTTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCGAAATTCCATACCTGTTCACCGACG
2 mDD10R       33.2%  94.4%  ------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  ------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  TCCCTTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCGAAATTCCATACCTGTTCACCGACG
5 OuidAR       23.6%  94.7%  TCCATTTCTTGTTACCGCCAACGCGCAATATGCCTTGCGAGGTCGCAAAATCGGCGAAATCCATACCTGTTCACCGACG
6 NuidAF       21.8%  92.1%  ------------------------------------------------------------------------------
7 sjekkAmdelR  16.1%  95.0%  ------------------------------------------------------------------------------

                cov    pid  2241   :         .         .         .         .       3 .         . 2320
1 pMDD15      100.0% 100.0%  ACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATACTCTTCACTCCACATGTCGGTGTA
```

A31

```
                                                                                                                  
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    ACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATACTCTTC-CTCCACATGTCGGTGTA
5 OuidAR        23.6%  94.7%    ACGGCGCTGACGCGATCAAAGACGCGGTGATACATATCCAGCCATGCACACTGATACTCTTCACTCCACATGTCGGTGTA
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2321     .         .         :         .         .         .         .      4 2400
1 pMDD15       100.0% 100.0%    CATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGTGATGATAATCGGCTGATGCAGTTTCTCCTGCCAGGCCA
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    CATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGGGATGAAAATCGGCTGAAGCA-TTCTCCCGGCAGGCAC
5 OuidAR        23.6%  94.7%    CATTGAGTGCAGCCCGGCTAACGTATCCACGCCGTATTCGGTGATGATAATCGGCTGATGCAGTTTCTCCTGCCAGGCCA
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2401     .         .         .         .         .         :         .    . 2480
1 pMDD15       100.0% 100.0%    GAAGTTCTTTTTCCAGTACCTTCTCTGCCGTTTCCAAATCGCCGCTTTGGACATACCATCCGTAATAACGGTTCAGGCAC
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    GAAGTTCTTTTTCAGAACCTTCTTGCCGTTTCAAATCCCCTTTGGAAAAACCTCCCGAAAAACGGGTTCGGCCACGCCA
5 OuidAR        23.6%  94.7%    GAAGTTCTTTTTCCAGTACCTTCTCTGCCGTTTCCAAATCGCCGCTTTGGACATACCATCCGTAATAACGGTTCAGGCAC
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2481     .         5         .         .         .         .         :    . 2560
1 pMDD15       100.0% 100.0%    AGCACATCAAAGAGATCGCTGATGGTATCGGTGTGAGCGTCGCAGAACATTACATTGACGCAGGTGATCGGACGCGTCGG
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    TAAAAAATTCCTTGAGGGGTTGCGGGGGAG--------------------------------------------------
5 OuidAR        23.6%  94.7%    AGCACATCAAAGAGATCGCTGATGGTATCGGTGTGAGCGTCGCAGAACATTACATTGACGCAGGTGATCGGACGCGTCGG
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2561     .         .         .         6         .         .         .    . 2640
1 pMDD15       100.0% 100.0%    GTCGAGTTTACGCGTTGCTTCCGCCAGTGGCGCGAAATATTCCCGAGCACCTTGCGGACGGGTATCCGGTTCGTTGGCAA
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    GTCGAGTTTACGCGTTGCTTCCGCCAGTGGCGCGAAATATTCCCGTGCACCTTGCGGACGGGTATCCGGTTCGTTGGCAA
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2641     :         .         .         .         .         7         .    . 2720
1 pMDD15       100.0% 100.0%    TACTCCACATCACCACGCTTGGGTGGTTTTTTGTCACGCGCTATCAGCTCTTTAATCGCCTGTAAGTGCGCTTGCTGAGTT
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    TACTCCACATCACCACGCTTGGGTGGTTTTTGTCACGCGCTATCAGCTCTTTAATCGCCTGTAAGTGCGCTTGCTGAGTT
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2721     .         .         :         .         .         .         .    8 2800
1 pMDD15       100.0% 100.0%    TCCCCGTTGACTGCCTCTTCGCTGTACAGTTCTTTCGGCTTGTTGCCCGCTTCGAAACCAATGCCTAAAGAGAGGTTAAA
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    TCCCCGTTGACTGCCTCTTCGCTGTACAGTTCTTTCGGCTTGTTGCCCGCTTCGAAACCAATGCCTAAAGAGAGGTTAAA
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2801     .         .         .         .         .         .         .    . 2880
1 pMDD15       100.0% 100.0%    GCCGACAGCAGCAGTTTCATCAATCACCACGATGCCATGTTCATCTGCCCAGTCGAGCATCTCTTCAGCGTAAGGGTAAT
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    GCCGACAGCAGCAGTTTCATCAATCACCACGATGCCATGTTCATCTGCCCAGTCGAGCATCTCTTCAGCGTAAGGGTAAT
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2881     .         9         .         .         .         .         .    . 2960
1 pMDD15       100.0% 100.0%    GCGAGGTACGGTAGGAGTTGGCCCCAATCCAGTCCATTAATGCGTGGTCGTGCACCATCAGCACGTTATCGAATCCTTTG
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    GCGAGGTACGGTAGGAGTTGGCCCCAATCCAGTCCATTAATGCGTGGTCGTGCACCATCAGCACGTTATCGAATCCTTTG
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------

                cov    pid 2961     .         .         .         0         .         .         .    . 3040
1 pMDD15       100.0% 100.0%    CCACGCAAGTCCGCATCTTCATGACGACCAAAGCCAGTAAAGTAGAACGGTTTGTGGTTAATCAGGAACTGTTCGCCCTT
2 mDD10R        33.2%  94.4%    --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%    --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%    --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%    CCACGCAAGTCCGCATCTTCATGACGACCAAAGCCAGTAAAGTAGAACGGTTTGTGGTTAATCAGAACTTCCCGC-----
6 NuidAF        21.8%  92.1%    --------------------------------------------------------------------------------
7 sjekkAmdelR   16.1%  95.0%    --------------------------------------------------------------------------------
```

```
                      cov    pid  3041        .         .         .         .         1         .        . 3120
1 pMDD15      100.0% 100.0%  CACTGCCACTGACCGGAGCTGTGACGCACAGTTCATAGAGATAACCTTCACCCGGTTGCCAGAGGTGCGGATTCACCACT
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  ------------------------------TAATCCCTTGGTATCTTCACCGGTTGCCAGAGGTGCGGATTCACCA
7 sjekkAmdelR  16.1%  95.0%  --------------------------------------------------------------------------------

                      cov    pid  3121        .         .         :         .         .         .        2 3200
1 pMDD15      100.0% 100.0%  TGCAAAGTCCCGCTAGTGCCTTGTCCAGTTGCAACCACCTGTTGATCCGCATCACGCAGTTCAACGCTGACATCACCATT
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  CTTGCAAAGTCCCGCTAGTGCCTTGTCCAGTTGAACGCCTGTTGATCCGCATCACGCAGTTCAACGCTGACATCACCATT
7 sjekkAmdelR  16.1%  95.0%  --------------------------------------------------------------------------------

                      cov    pid  3201        .         .         .         .         :         .        . 3280
1 pMDD15      100.0% 100.0%  G-GCCACCACCTGCCAGTCAACAGACGCGTGGTTACAGTCTTGCGCGACATGCGTCACCACGGTGATATCGTCCACCCAG
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  GAGCCACCACCTGCCAGTCAACAGACGCGTGGTTACAGTCTTGCGCGACATGCGTCACCACGGTGATATCGTCCACCCAG
7 sjekkAmdelR  16.1%  95.0%  --------------------------------------------------------------------------------

                      cov    pid  3281        .         .         3         .         .         .        . 3360
1 pMDD15      100.0% 100.0%  GTGTTCGGCGTGGTGTAGAGCATTACGCTGCGATGGATCCCGGCATAGTTAAAGAAATCATGGAAGTAAGACTGCTTTTT
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  GTGTTCGGCGTGGTGTAGAGCATTACGCTGCGATGGATCCCGGCATACTTAAAGAAATCATGGAAGTAAGACTGCTTTTT
7 sjekkAmdelR  16.1%  95.0%  ----------------GAGCATCACTGATTTGATGGAACCCTGCTTAGTTACAGAAGTCACGGAAGTACGAGAGGTGTTT

                      cov    pid  3361        .         .         .         4         .         .        . 3440
1 pMDD15      100.0% 100.0%  CTTGCCGTTTTCGTCGGTAATCACCATTCCCGGCGGGATAGTCTAGCCAGTTCAGTTCGTTGTTCACACAAACGGTGATAC
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  CTTGCCGTTTTCGTCGGTAATCACCATTCCCGGCGGGATAGTCTAGCCAGTTCAGTTCGTTGTTCACACAAACGGTGATAC
7 sjekkAmdelR  16.1%  95.0%  CATGACGTTCTGGTCGGTATTCACCATTCCCGGCGGGATAGTCTAGCCAGATCAGTTCGTTGTTCACACAAACGGTGATAC

                      cov    pid  3441        .         .         :         .         .         5        . 3520
1 pMDD15      100.0% 100.0%  GTACACTTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGCGTATAGCCGCCCTGATGCTCCATCACTTCC
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  GTACACTTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGCGTATAGCCGCCCTGATGCTCCATCTCTTCC
7 sjekkAmdelR  16.1%  95.0%  GTACACTTTTTCCCGGCAATAACATACGGCGTGACATCGGCTTCAAATGGCGTATAGCCGCCCTGATGCTCCATCACTTCC

                      cov    pid  3521        .         .         :         .         .         .        6 3600
1 pMDD15      100.0% 100.0%  TGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAGCACGATACGCTGGCCTGCCCAACCTTTTCGG
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  TGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAGCACGATACGCTGGCCTGCCCAACCTTTTCGG
7 sjekkAmdelR  16.1%  95.0%  TGATTATTGACCCACACTTTGCCGTAATGAGTGACCGCATCGAAACGCAGCACGATACGCTGGCCTGCCCAACCTTTTCGG

                      cov    pid  3601        .         .         .         .         .         .        . 3680
1 pMDD15      100.0% 100.0%  TATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATATCTGCATCGGCGAACTGATCGTTAAAACTGC
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  TATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATATCTGCATCGGCGAACTGATCGTTAAAACTGC
7 sjekkAmdelR  16.1%  95.0%  TATAAAGACTTCGCGCTGATACCAGACGTTGCCCGCATAATTACGAATATCTGCATCGGCGAACTGATCGTTAAAACTGC

                      cov    pid  3681        .         7         .         .         .         .        : 3760
1 pMDD15      100.0% 100.0%  CTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGCTGATCAATTCCACAGTTTTCGCGATCCAGA
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR       23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF       21.8%  92.1%  CTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGCTGATCAATTCCACAGTTTTCGCGATCCAGA
7 sjekkAmdelR  16.1%  95.0%  CTGGCACAGCAATTGCCCGGCTTTCTTGTAACGCGCTTTCCCACCAACGCTGATCAATTCCACAGTTTTCGCGATCCAGA

                      cov    pid  3761        .         .         .         8         .         .        . 3840
1 pMDD15      100.0% 100.0%  CTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTCTACAGGACGTAACATAAGGGACTCCTCGAA
2 mDD10R       33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv     27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF   34.8%  93.5%  --------------------------------------------------------------------------------
```

A33

```
5 OuidAR        23.6%  94.7%  ----------------------------------------------------------------------------------
6 NuidAF        21.8%  92.1%  CTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTCTACAGGACGTAACATAAGGGACTCCTCGAA
7 sjekkAmdelR   16.1%  95.0%  CTGAATGCCCACAGGCCGTCGAGTTTTTTGATTTCACGGGTTGGGGTTTCTACAGGACGTAACATAAGGGACTCCTCGAA

                    cov   pid 3841   :       .       .       .       .       9       .       . 3920
1 pMDD15       100.0% 100.0%  GCTTAGATCTATTACCCTGTTATCCCTACTCGAGTTCATGTGCAGCTCCATAAGCAAAAGGGGATGATAAGTTTATCACC
2 mDD10R        33.2%  94.4%  --------------------------------------------------------------------------------
3 sacBsekv      27.2%  96.4%  --------------------------------------------------------------------------------
4 sjekksacBF    34.8%  93.5%  --------------------------------------------------------------------------------
5 OuidAR        23.6%  94.7%  --------------------------------------------------------------------------------
6 NuidAF        21.8%  92.1%  GCTTAGATCTATTACCCTGTTATCCCTACTCGAGTTCATGTGCAGCTCCATAAGCAAAAGGGGATGATAAGTTTATCACC
7 sjekkAmdelR   16.1%  95.0%  GCTTAGATCTATTACCCTGTTATCCCTACTCGAGTTCATGTGCAGCTCCATAAGCAAAAGGGGATGATAAG-TTATCACC

                    cov   pid 3921         . ] 3932
1 pMDD15       100.0% 100.0%  ACCGACTATTTG
2 mDD10R        33.2%  94.4%  ------------
3 sacBsekv      27.2%  96.4%  ------------
4 sjekksacBF    34.8%  93.5%  ------------
5 OuidAR        23.6%  94.7%  ------------
6 NuidAF        21.8%  92.1%  ACCGACTATTTG
7 sjekkAmdelR   16.1%  95.0%  ACCGACTA-TTG
```

MView 1.63, Copyright © 1997-2018 Nigel P. Brown

A34

# Appendix J   Sequencing of pMDD16

The sequencing results of pMDD16 are presented in the next pages. The first line shows the template pMDD16 and the next line the alignment of primer uidAF. The alignment show that the N20 sequence is correct. "cov" shows how much of the template the aligned sequence covers in percentage. "pid" shows the percentage of identity between the aligned sequence and the template.

```
              cov    pid     1 [        .         .         .         .         :         .         .         . 80
1 pMDD16 100.0% 100.0%         GCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTT
2 uidAF  100.0%  99.6%         GCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTT

              cov    pid    81          .         .         1         .         .         .         :         . 160
1 pMDD16 100.0% 100.0%         GCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCCTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGGCGGC
2 uidAF  100.0%  99.6%         GCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCCTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTCAGGCGGC

              cov    pid   161          .         .         .         2         .         .         .         . 240
1 pMDD16 100.0% 100.0%         TTTCGGTGACTGTCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
2 uidAF  100.0%  99.6%         TTTCGGTGACTGTCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC

              cov    pid   241          :         .         .         .         .         3         .         . 320
1 pMDD16 100.0% 100.0%         GAGTCGGTGCTTTTTTTTGAATTCTCTAGAGGATCGATCCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTA
2 uidAF  100.0%  99.6%         GAGTCGGTGCTTTTTTTTGAATTCTCTAGAGGATCGATCCTTTTTAACCCATCACATATACCTGCCGTTCACTATTATTTA

              cov    pid   321          .         .         :         .         .         .         .         4 400
1 pMDD16 100.0% 100.0%         GTGAAATGAGATATTATGATATTTTCTGAATTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAA
2 uidAF  100.0%  99.6%         GTGAAATGAGATATTATGATATTTTCTGAATTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAA

              cov    pid   401          .         .         .         .         :         .         .         . 480
1 pMDD16 100.0% 100.0%         TACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCA
2 uidAF  100.0%  99.6%         TACAGAGAATGAAAAGAAACAGATAGATTTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCA

              cov    pid   481          .         5         .         .         .         .         :         . 560
1 pMDD16 100.0% 100.0%         AACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTT
2 uidAF  100.0%  99.6%         AACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTT

              cov    pid   561          .         .         .         6         .         .         .         . 640
1 pMDD16 100.0% 100.0%         GTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGG
2 uidAF  100.0%  99.6%         GTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGG

              cov    pid   641          :         .         .         .         .         7         .         . 720
1 pMDD16 100.0% 100.0%         TCTTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAA
2 uidAF  100.0%  99.6%         TCTTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAA

              cov    pid   721          .         .         :         .         .         .         .         8 800
1 pMDD16 100.0% 100.0%         AAAAGGAGACATGAACGATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCA
2 uidAF  100.0%  99.6%         AAAAGGAGACATGAACGATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCA

              cov    pid   801          .         .         .         .         :         .         .         . 880
1 pMDD16 100.0% 100.0%         GGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACG
2 uidAF  100.0%  99.6%         GGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTCCCATATTACACG

              cov    pid   881          .         9         .         .         .         .         :         . 960
1 pMDD16 100.0% 100.0%         CCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAA
2 uidAF  100.0%  99.6%         CCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAA

              cov    pid   961          .         .         .         0         .         .         .         . 1040
1 pMDD16 100.0% 100.0%         ATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCAC
2 uidAF  100.0%  99.6%         ATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCAC

              cov    pid  1041          :         .         .         .         .         1         .         . 1120
1 pMDD16 100.0% 100.0%         GGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGT
2 uidAF  100.0%  99.6%         GGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGT

              cov    pid  1121          .         .         .         .         :         .         .         2 1200
1 pMDD16 100.0% 100.0%         CGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTA
2 uidAF  100.0%  99.6%         CGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTA

              cov    pid  1201          .         .         .         .         :         .         .         ] 1270
1 pMDD16 100.0% 100.0%         TCCTAAAAGACCAAACACAAGAATTGGTCAGGTTTCAGCCACATTTTACATCTGACGGAAAAATCCGTTTATT
2 uidAF  100.0%  99.6%         TCCTAAAAGACCAAACCCAAGAATTGGTCAGGTTTCAGCCCATTTTACTTCTGACGGAAAAATCCGTTTATT
```

Marie Duus Dahl

# NTNU

Norwegian University of
Science and Technology