Developing Genetic Tools in *Methanococcus maripaludis*

Nathan Innard

MSc by Research

University of York

Biology

January 2018

Abstract

Highly sophisticated genetic tools have been deployed in many eukaryotes and bacteria, yet those available in the third domain of life: the archaea, remain comparatively limited. The constraints this imposes must be overcome if the unique capabilities of this domain are to be tapped. One biological pathway found solely within members of the archaea is methanogenesis: the conversion of simple molecules into methane. The manipulation of methane production through genetic modification of methanogens is highly desirable, since in addition to its potential use as a biofuel, methane contributes significantly to the greenhouse effect. Among the methanogens, *Methanococcus maripaludis* (*M. maripaludis*) is one of the best developed as a model, however many of its available genetic tools are outdated. The CRISPR/Cas9 system is a relatively recently developed and highly versatile genetic technique, and at the outset of this work had not been deployed in any archaeal system. The aim of this project was therefore to provide a proof of principle that the CRISPR/Cas9 system could be deployed successfully in this organism. The experimental approach selected was a plasmid invader assay, in which the activity of the system could be demonstrated by Cas9 mediated plasmid destruction resulting in reduced *M. maripaludis* cell viability. A scheme by which this assay could be conducted using existing genetic tools in two different *M. maripaludis* strains was designed, and the full series of required plasmids was produced. Attempts to use these plasmids to produce the *M. maripaludis* strains required for the invader assay were unsuccessful, and it could therefore not be carried out. However, it is anticipated that the assay system designed and plasmids produced here should enable the rapid testing of this system were this work continued, which should enable the addition of the CRISPR/Cas9 system to the *M. maripaludis* genetic toolbox.

Table of Contents

Abstract 2

Table of Contents 3

List of Figures 9

List of Tables 11

Acknowledgements 12

Declaration 13

Chapter 1: General Introduction 14

1.1.1 Genetic Tools in the Archaea 14

1.1.2 Methanogenesis and Methanogens 14

1.1.3 *M. maripaludis* 14

1.1.4 Existing Genetic Tools in *M. maripaludis* 15

1.1.4.1 Chromosomal modification 15

1.1.4.2 Shuttle vectors 16

1.1.4.3 Reporter genes 17

1.1.4.4 Gene expression 17

1.1.5 The CRISPR/Cas System 18

1.1.5.1 CRISPR/Cas as a genetic tool 18

1.1.5.2 CRISPR/Cas in archaea 19

1.1.5.3 Deploying the CRISPR/Cas9 system in *M. maripaludis* 19

1.1.6 The Plasmid Invader Assay 20

1.1.7 Project Aims 23

Chapter 2: Materials and Methods 24

2.1.1 Molecular Work 24

2.1.1.1 Agarose gel electrophoresis 24

2.1.1.2 Resuspending gBlocks 24

2.1.1.3 Restriction digestion 24

2.1.1.4 PCR purification 24

2.1.1.5 Gel extraction 25
2.1.1.6 Plasmid dephosphorylation 25
2.1.1.7 Ligation 25
2.1.1.8 Golden Gate reactions 25
2.1.1.9 Transformation of ligation products/plasmids into *E. coli* 26
2.1.1.10 Colony PCR 26
2.1.1.11 Plasmid DNA isolation (from *E. coli*) 27
2.1.1.12 Generating chemically competent cells (*E. coli* DH5α) 27
2.1.1.13 *E. coli* culturing 27
2.1.1.14 Long term plasmid storage 28
2.1.1.15 Sequencing reactions 28
2.1.2 Aerobic Media and Antibiotics 28
2.1.2.1 LB medium 28
2.1.2.2 LB agar plates 28
2.1.2.3 Ampicillin and kanamycin 28
2.1.2.4 IPTG 29
2.1.2.5 X-gal 29
2.1.3 *M. maripaludis* 29
2.1.3.1 *M. maripaludis* culturing 29
2.1.3.2 McCas medium 29
2.1.3.3 N-free trace minerals (1000x) 30
2.1.3.4 Vitamin solution (100x) 31
2.1.3.5 N-free general salts solution 31
2.1.3.6 $FeSO_4$ solution 32
2.1.3.7 K_2HPO_4 solution 32
2.1.3.8 Resazurin solution 32
2.1.3.9 Transformation of plasmids into *M. maripaludis* 32
2.1.3.10 *M. maripaludis* outgrowth and selection following transformation 33

2.1.4 Anaerobic Antibiotics and Solutions 33
2.1.4.1 Reducing agent 33
2.1.4.2 Transformation buffer 34
2.1.4.3 PEG solution 34
2.1.4.4 Na_2S solution 35
2.1.4.5 Neomycin 35
2.1.5 In Silico Methods 35
2.1.5.1 Codon optimisation 35
2.1.5.2 gRNA design 35
2.1.5.3 Random sequence generation 36
2.1.5.4 Checking complexity IDT gblocks 36
2.1.5.5 Removal of incompatible restriction enzyme sites 36
2.1.5.6 Construct design and testing 36
2.1.5.7 Primer design 37
2.1.5.8 Sequence analysis 37
Chapter 3: In Silico Planning 38
3.1 Introduction 38
3.1.1 Chapter Aims 38
3.1.2 Component Design 39
3.1.2.1 Promoter and terminator 39
3.1.2.2 Cas9 40
3.1.3 gRNA 41
3.1.3.1 Test gRNA 42
3.1.3.2 Control gRNA 43
3.1.4 Construct Design and Assembly 43
3.1.4.1 Obtaining parts 43
3.1.4.2 Combining parts: Golden Gate 43
3.2 Results 46

3.2.1 Component Design 46
3.2.1.1 Cas9 46
3.2.1.2 gRNA: PurR 48
3.2.1.3 gRNA: eGFP 50
3.2.1.4 gRNA: control 51
3.2.1.5 Phmv and Tmcr 51
3.2.2 Construct Design and Assembly 52
3.2.2.1 Module organisation 52
3.2.2.2 Division into gBlocks 53
3.2.2.3 Fragment assembly 54
3.3 Discussion 57
3.3.1 Codon Optimisation of SpCas9 57
3.3.2 gRNAs 58
3.3.2.1 Off-targeting 58
3.3.2.2 On-targeting 59
3.3.3 Chapter Summary/Next Steps 61
Chapter 4: Molecular Work 62
4.1 Introduction 62
4.1.1 Chapter Aims 62
4.1.2 Delivering CRISPR/Cas9 Assemblies to *M. maripaludis* Transformation Vectors 62
4.1.3 Selection of Chromosomal Target Site for *M. maripaludis* S0001 Strain 63
4.1.4 Golden Gate Compatible *M. maripaludis* Transformation Vectors 64
4.1.5 Modifying pLW40/pAW42 65
4.2 Results 66
4.2.1 Production of Golden Gate Compatible Integrative Vectors pBLPrt/pCRPrtneo 66
4.2.1.1 Modification via PCR 66

4.2.1.2 Modification via gBlock 67
4.2.1.3 Subcloning Dropout Mod gBlock into pUC18 69
4.2.1.4 Generating Golden Gate compatible pCRPrtneo (pCRPrtneo GG) 73
4.2.1.5 Generating Golden Gate compatible pBLPrt (pBLPrt GG) 77
4.2.2 Assembling CRISPR/Cas9 constructs into pCRPrtneo GG/pBLPrt GG 79
4.2.2.1 Subcloning Fragment 1 gBlocks into pHSG298 80
4.2.2.2 Subcloning Fragment 2 gBlock into pHSG298 82
4.2.2.3 Subcloning Fragment 3 gBlock into pHSG298 83
4.2.2.4 Assembling CRISPR/Cas9 constructs into pCRPrtneo GG 84
4.2.2.5 Assembling CRISPR/Cas9 constructs into pBLPrt GG 87
4.2.3 Modifying pLW40/pAW42 to Contain the eGFP Target Site 90
4.2.3.1 Integration site selection 90
4.2.3.2 Modification Strategy 91
4.2.3.3 Subcloning eGFP Mod gBlock into pUC18 92
4.2.3.4 Generating modified pLW40 (pLW40 Mod) 93
4.2.3.5 Generating modified pAW42 (pAW42 Mod) 95
4.2.4 Producing *M. maripaludis* test and control strains 96
4.3 Discussion 99
4.3.1 Construct Assembly Issues 99
4.3.2 Constructs Produced 100
4.3.3 Golden Gate Constructs 101
4.3.4 Production of *M. maripaludis* Test Strains 102
4.3.5 Chapter Summary 103
Chapter 5: General Discussion 104
5.1.1 Conducting the Invader Assay 104
5.1.2 Limitation of the Invader Assay as Designed in this Study 105
5.1.3 Improving the Invader Assay: Decoupling 107
5.1.3.1 Inducible promoter 107

5.1.3.2 Two plasmid system 107
5.1.4 Possible Improvements to Vectors 109
5.1.5 Removal of CRISPR/Cas9 from *M. maripaludis* 109
5.1.6 Possible Applications for the CRISPR/Cas9 system in *M. maripaludis* 110
5.1.6.1 Improvements to *M. maripaludis* transformation 110
5.1.6.2 Improvements to *M. maripaludis* promoters 112
5.1.6.3 Unbiased assessment of off-targeting in organisms with small genomes 114
5.1.7 Summary 115
Appendix 116
Reference List 123

List of Figures

Figure 1.1: The plasmid invader assay. 22

Figure 3.1: Composition of the Cas9 and gRNA modules. 38

Figure 3.2: Golden Gate assembly. 45

Figure 3.3: Comparison of codon usage frequencies between the optimised Cas9 gene (Mmp-Cas9) and *M. maripaludis*. 48

Figure 3.4: Mouse gRNA aligned to invader plasmid PurR gene. 49

Figure 3.5: Minimal Phmv. 52

Figure 3.6: Orientation of the Cas9 and gRNA modules. 53

Figure 3.7: Division of the 4593bp unit into three fragments. 54

Figure 3.8: Fragment assembly scheme. 56

Figure 4.1: Structure of the MCM-C homology region in pAW10. 67

Figure 4.2: Producing Golden Gate compatible pBLPrt/pCRPrtneo vectors scheme. 68

Figure 4.3: Subcloning Dropout Mod gBlock into pUC18. 73

Figure 4.4: Generating Golden Gate compatible pCRPrtneo (pCRPrtneo GG). 76

Figure 4.5: Generating Golden Gate compatible pBLPrt (pBLPrt GG). 79

Figure 4.6: Subcloning Fragment 1 gBlocks into pHSG298. 81

Figure 4.7: Subcloning Fragment 2 gBlock into pHSG298. 83

Figure 4.8: Subcloning Fragment 3 gBlock into pHSG298. 84

Figure 4.9: Assembling CRISPR/Cas9 constructs into pCRPrtneo GG. 87

Figure 4.10: Assembling CRISPR/Cas9 constructs into pBLPrt GG. 90

Figure 4.11: Selected insertion site for eGFP target region in pLW40/pAW42. 91

Figure 4.12: Producing modified pLW40/pAW42 vectors scheme. 92

Figure 4.13: Subcloning eGFP Mod gBlock into pUC18. 93

Figure 4.14: Generating modified pLW40 (pLW40 Mod). 95

Figure 4.15: Generating modified pAW42 (pAW42 Mod). 96

Figure 5.1: Control of gene expression using CRISPRi. 113

Figure 5.2: Control of gene expression using CRISPRa. 114

Figure A1: Nucleotide alignment of Cas9 codon optimised to *M. maripaludis*, *A. thaliana* and *Streptomyces*. 120

Figure A2: Mmp-Cas9 nucleotide sequence. 122

List of Tables

Table 2.1: McCas medium components. 30

Table 2.2: N-free trace minerals (1000x) components. 31

Table 2.3: Vitamin Solution (100x) components. 31

Table 2.4: N-free general salts solution components. 32

Table 2.5: Transformation buffer components. 34

Table 3.1: Targeting predictions for the three selected gRNAs. 51

Table 3.2: Sites modified in the Cas9 gene to remove pre-existing BpiI sites. 55

Table 4.1: Sites modified in pDVA_BC LacZ gene to remove pre-existing BamHI/SacI sites. 69

Table 4.2: S0001 cultures used for transformation. 97

Table 4.3: OD_{600} growth measurements of subcultured transformed cells. 98

Table A1: Primers Details. 116

Table A2: Plasmids produced in this study. 117

Acknowledgements

I would like to thank James Chong for giving me the opportunity to undertake this project, and for his guidance, support and patience throughout. I would also like to thank Anna Alessi and James Robson for helping me settle in. Finally I would like to thank all members of the Chong group for making the lab such a pleasant environment to have worked in.

Declaration

I declare that this thesis is a presentation of original work and I, Nathan Innard, am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1: General Introduction

1.1.1 Genetic Tools in the Archaea

The repurposing of biological systems to benefit man is reliant on tools which allow manipulations at the genetic level. The ease and extent to which this can be achieved is limited by the quality and variety of the tools available (*1*). While significant progress has been made in bacterial and eukaryotic models, the development of genetic tools in the archaea has lagged behind, limiting the extent to which they can be manipulated (*2*). Archaea are unicellular prokaryotes (*3*) which were first recognised as a distinct domain of life in 1977 (*4*). They possess a multitude of genes and pathways not found in eukaryotes or bacteria (*5, 6*). The investigation, understanding and ultimately manipulation of these unique features is currently reliant upon limited archaeal specific genetic tools: providing a clear incentive for further development. One important pathway found uniquely among the archaea is methanogenesis (*7*).

1.1.2 Methanogenesis and Methanogens

Methanogenesis is the process by which simple molecules such as dihydrogen and carbon dioxide are converted into methane (*6, 8*). It is carried out solely by the methanogens: a diverse group of species belonging to five different orders of the domain archaea (*7*). Methanogens are obligate anaerobes and occupy a wide variety of environmental niches including oceans, wastewater and the gastrointestinal tracts of animals (*5, 9*). Methane exhibits 21 times the warming potential of carbon dioxide, making it a potent greenhouse gas and an environmental concern: particularly in industries such as livestock farming and water treatment (*9, 10*). However, methane can also be used as a biofuel (*11*). This illustrates the potential benefit to manipulating methanogens: both in combatting global warming as well as potentially providing a source of energy. Laboratory manipulation of methanogens is challenging since they can only be handled anaerobically (*6*). Despite this, genetic tools have been deployed in several species, of which *Methanococcus maripaludis* (*M. maripaludis*) is one the most well developed as a model (*12*).

1.1.3 *M. maripaludis*

Originally isolated from a salt marsh sediment, *M. maripaludis* grows optimally at 38°C and has a doubling time of just over two hours (*13*). Its 1.6 megabase genome has been sequenced, which has a low GC content and is predicted to contain approximately 1700

protein coding genes (*14*, *15*). *M. maripaludis* is polyploid, with the number of copies of the chromosome fluctuating between approximately 30 and 55 depending on growth phase (*16*). A variety of genetic tools have been developed in this organism, including a PEG-based transformation protocol (*17*), a shuttle vector system for expression of exogenous genes (*18*), and a positive/negative selection system for markerless mutagenesis, modification of endogenous genes and introduction of heterologous material onto the chromosome (*19*). *M. maripaludis* is sensitive to the antibiotics puromycin and neomycin, and their corresponding resistance genes can be used for selection in this organism (*20*, *21*).

1.1.4 Existing Genetic Tools in *M. maripaludis*

1.1.4.1 Chromosomal modification

M. maripaludis was found to recombine naturally when exogenous DNA bearing homology to a region of the genome was introduced on an integrative (i.e. non-replicative) plasmid (*22*), and the efficiency at which chromosomal modifications could be recovered was increased through the development of a PEG-based transformation protocol (*17*). However, these methods were limited since recovery was based on a single integrated positive selection marker, which meant that sequential genetic manipulations were not possible. This was overcome by the development of a positive/negative selection based markerless mutagenesis protocol, in which both markers are removed following chromosomal modification, allowing repeated manipulations (*19*).

This protocol is conducted in the *M. maripaludis* Mm900 strain, in which the hypoxanthine phosphoribosyltransferase (HPT) gene, which confers sensitivity to the base analogue 8-azahypoxanthine, has been deleted. Material to be introduced onto the chromosome is inserted into the backbone of the integrative pCRPrtneo vector, in between approximately 500 bp of 5' and 3' flanking sequence corresponding to each side of the desired chromosomal integration site. The pCRPrtneo backbone contains a neomycin resistance gene (NeoR) for positive selection, and the HPT gene for negative selection, but no archaeal replication origin: so its cargo must be chromosomally integrated to be maintained (*19*). Following transformation (*17*) of Mm900 with pCRPrtneo, NeoR enables the initial selection of cells which have undergone a loop-in recombination event at the target site, resulting in the chromosomal integration of the full transformation vector (i.e. insert cargo and backbone). Subsequent culturing of transformed cells without selection allows a second recombination event between one of the vector flanking regions and the corresponding chromosomal flank to take place. This has two possible outcomes, which

depend on the flanking regions involved and the relative position of the looped-in region to the chromosomal locus. In the first, the vector backbone and intervening region between the chromosomal flanks is lost, resulting in the desired product: the vector cargo incorporated at the target site on the chromosome. The other equally likely possibility is loss of both the vector backbone and insert cargo, resulting in regeneration of the original chromosomal target site. Since both possible loop-out events result in the loss of the vector backbone, which contains the HPT gene, 8-azahypoxanthine is used to purge non-looped-out cells, and screening enables recovery of those which have retained the desired insert (*19*).

A second integrative vector, pBLPrt, was also produced. This vector contains flanking regions to the *M. maripaludis* genomic uracil phosphoribosyltransferase (UPT) locus, which confers sensitivity to the base analogue 6-azauracil, in addition to all of the components present in pCRPrtneo. Material introduced in between these flanking regions in pBLPrt can be incorporated into this site on the *M. maripaludis* chromosome using the protocol described above, but without the need for the introduction of flanking regions (since they are already present within this vector) (*19*).

This protocol is used to produce *M. maripaludis* strains in which all chromosomes bear the same desired modification (*19*). The mechanism by which genetically uniform modified chromosomes can be obtained in *M. maripaludis*, which is highly polyploid, is thought to involve rapid propagation of the DNA element conferring a selective advantage to other chromosomes via gene conversion, under conditions of strong selective pressure (*16*). In terms of the positive/negative selection protocol, the initial loop-in event presumably takes place on a single chromosome, and this intermediate is rapidly propagated to other as yet unmodified chromosomes via gene conversion under strong positive selection. Similarly, the loop-out event likely initially takes place on one chromosome, and this spreads to the others under negative selection.

1.1.4.2 Shuttle vectors

A cryptic plasmid from *M. maripaludis*, pURB500 (*23*), was used to produce a shuttle vector, pDLT44, capable of replication in both *M. maripaludis*, and *E. coli*. This vector contains a puromycin (PurR) and ampicillin (AmpR) resistance gene for selection in *M. maripaludis* and *E. coli*, respectively. It had been hypothesised that a region of pDLT44, ORF1, coded for a protein responsible for plasmid replication, and that one or both of two

regions, ORFLESS1 or ORFLESS2, was the replication origin (*18*). The *M. maripaludis* S0001 strain was produced by transferring the ORF1 region to the chromosomal UPT site in the Mm900 background. S0001 was able to support maintenance of pURB500 derived plasmids in which ORF1 had been deleted, demonstrating that this factor functioned in *trans* (*12*). The smallest plasmid derived from pDLT44 available at the time, pLW40 (*12, 24*), was modified by deleting its ORFLESS2 region, creating pAW42. pAW42, was able to replicate in S0001, indicating that ORFLESS1 was the replication origin. The S0001 strain was found to be transformed with pAW42 at a 7000-fold higher frequency than Mm900 with pLW40. Additionally, S0001 was transformed with pLW40 more efficiently than Mm900. The authors speculated this was because the ORF1 protein, which is chromosomally expressed in this strain, would be present immediately after plasmid transformation, and therefore permit rapid replication of any pURB500 based plasmid (*12*).

1.1.4.3 Reporter genes

While some genetic tools are relatively well developed in *M. maripaludis*, others are still comparatively limited. Reporter genes are particularly problematic since widely used fluorescent proteins such as GFP cannot function in the strict anaerobic environment required to support *M. maripaludis* (*25*). The beta-galactosidase assay can be used in live *M. maripaludis* cells. However, since the colour formation of the substrate X-gal requires oxidised conditions (*26*), this assay involves the transient removal of plates from the anaerobic chamber and spraying with X-gal (*27, 28*), and is therefore only suitable for qualitative analysis of gene expression. Quantitative assays using beta-galactosidase are possible (*29*), but the required oxygen exposure times prohibit the use of living cells and must instead be performed on cell extracts (*8*).

1.1.4.4 Gene expression

This lack of a simple *in vivo* quantitative reporter assay is likely partly responsible for the relatively limited gene expression systems currently available in *M. maripaludis*. These comprise three constitutive promoters derived from closely related species: the widely used Phmv and Pmcr from *Methanococcus voltae (M. voltae) (17, 30)*, and a histone promoter from *Methanococcus vannielii* (*M. vannielii*) (*31*). Inducible gene expression is possible using the endogenous nif promoter (*28, 32*). Expression from this promoter is regulated by repression based on nitrogen availability, with complete repression in cells grown in the presence of ammonium, intermediate repression when grown with alanine, and no repression (i.e. full expression) when grown in medium containing dinitrogen as the sole

nitrogen source (*32*). However, *M. maripaludis* grows somewhat more slowly on alanine, but substantially more slowly on dinitrogen compared with ammonium (*28*), presumably due to the high energy demands associated with nitrogen fixation (*33*). As such, this promoter is induced by growth in alanine in studies requiring inducible gene expression (*34-37*).

The situation with the *M. maripaludis* nif promoter is representative of the majority of other inducible gene expression systems currently available for methanogens, which are primarily based on endogenous promoters regulated by growth substrates, which in addition to making culturing more complicated (*38*) may also lead to unwanted regulation of endogenous genes. These limitations can be avoided by using systems induced by substrates that are unlikely to be used by the recipient organism (i.e. from a distantly related donor organism). A single example of such a system has been developed in several species of the order *Methanosarcina*. The binding site for the bacterial tetracycline-response protein (tetR) was incorporated at different sites within a strong constitutive archaeal promoter. Genes under its control could be induced by tetracycline, and the level of induction was dependent on the position of the tetR binding site within the promoter. The end result was a series of promoters capable of driving different levels of gene expression when induced (*39*). Unfortunately, attempts to transfer this system to *Methanococcus* have been unsuccessful (*8*).

1.1.5 The CRISPR/Cas System

1.1.5.1 CRISPR/Cas as a genetic tool

One tool which has relatively recently been developed is the CRISPR/Cas system. Endogenous CRISPR/Cas systems are a type of adaptive immunity in prokaryotes, and are found in many bacteria and the majority of archaea (*40*). They have been classified into a number of different types, and CRISPR/Cas9, which is a type II system, is the most widely used as a genetic tool (*41*). CRISPR/Cas9 consists of two components: a customisable guide RNA (gRNA) which provides targeting specificity by forming a DNA/RNA interaction, and a Cas9 nuclease, which mediates double strand break (DSB) induction. When both components are expressed in cells, they associate forming a functional complex and cleave DNA complimentary to the 5' most 20 nucleotides (nt) of the gRNA. The targeting specificity of the Cas9 protein can therefore easily be controlled by varying this sequence. The only constraint with respect to target site is that it must be immediately followed by a 3 bp "protospacer adjacent motif" sequence: "NGG" (*42*). The CRISPR/Cas9

system and its derivatives have been used to perform many useful functions, including (but not limited to) genome and epigenome editing, regulation of gene expression, DNA imaging and investigating protein-genome interactions (*43*). CRISPR/Cas9 has been deployed in a wide variety of systems including bacteria, plants, animals and fungi (*43*), and it seems likely that the development of this highly versatile tool for use in *M. maripaludis* would be a useful addition to the existing genetic toolbox.

1.1.5.2 CRISPR/Cas in archaea

At the outset of this study, to our knowledge, no endogenous archaeal CRISPR/Cas9 system had been reported, and it had not been deployed exogenously in any archaeal species. Conversely, type I and III systems, which were known to be endogenous to archaea, had been co-opted for the purpose of transcriptional regulation in *Haloferax volcanii* and *Sulfolobus solfataricus* (*44*, *45*). We envisaged that CRISPR/Cas9 would be a superior tool for several reasons. Firstly, unlike type I and III systems, which are reliant on large multi-component effector complexes (*43*), CRISPR/Cas9 requires just two components (Cas9 and a gRNA) (*46*), and is therefore simpler. Secondly, since CRISPR/Cas9 is by far the most widely used system it is the best characterised (*42*). Finally, as type II systems were thought to be absent from all known archaea (*41*) it would theoretically have bypassed the potential risk of unwanted effects which may be associated with manipulating endogenous systems. During the course of this project, evidence of an endogenous archaeal type II system was discovered in two species of nanoarchaea (*47*). Shortly afterwards, Nayak and Metcalf (2017) reported the deployment of the widely used bacterial CRISPR/Cas9 system form *Streptococcus pyogenes* (*S. pyogenes*) in the archaeon *Methanosarcina acetivorans* (*M. acetivorans*). Using this system, the authors were able to generate chromosomal modifications in less than half the time that had previously been required. The system was also relatively efficient, with chromosomal modifications achieved in approximately 20% of successfully transformed cells (*48*). This demonstration further highlights the potential benefits of developing this genetic tool for use in *M. maripaludis*.

1.1.5.3 Deploying the CRISPR/Cas9 system in *M. maripaludis*

Before the CRISPR/Cas9 system can be used as a genetic tool in *M. maripaludis*, it must first be to be shown to be functional in this organism. This will therefore be the focus of this project. A simple assay to demonstrate functionality of the CRISPR/Cas9 system would be to introduce a plasmid encoding Cas9 and a gRNA targeting the *M. maripaludis*

chromosome. Self-targeting of DSBs would be cytotoxic, and expected to result in lethality (*49*). Lower cell counts in test transformations compared with vector controls would therefore indicate Cas9 activity. However, it was anticipated that the polyploid nature of *M. maripaludis* would be problematic for this assay design since all 30–55 copies of the genome (*16*) would theoretically have to be destroyed to give the test output (i.e. cell death).

1.1.6 The Plasmid Invader Assay

The plasmid invader assay (*44*) is based on a similar principle and could be adapted for use in *M. maripaludis* to overcome the potential problems associated with its polyploidy. Here, DNA encoding both the Cas9 gene and a gRNA targeting the antibiotic resistance gene of a plasmid "invader" could be stably introduced into *M. maripaludis*. This "test" strain would then be transformed with the invader plasmid, and colonies would be plated on selective medium corresponding to the resistance gene carried on this plasmid. Lower colony counts in test transformants (in which the invader plasmid is targeted by Cas9) compared with controls (in which it is not) would indicate that Cas9 is active and able to destroy the invader plasmid, resulting in reduced survival.

The critical requirement of this assay is that the Cas9/gRNA complex expressed in the test strain cells must be able to completely remove the invader plasmid, since cell death resulting from plasmid loss is the only way Cas9 functionality would be detected. In theory, the simplest way of increasing the likelihood of this is by designing the system such that the ratio of functional Cas9 complexes to invader plasmids is as high as possible: i.e. that the Concentration of Cas9 is high relative to the invader. There are two main design parameters which could be used to achieve this. Firstly, both components of the CRISPR/Cas9 system, the gRNA and Cas9 protein, should be expressed as strongly as possible. The simplest approach here would be to place both under the control of the strongest promoter possible. In addition, expression of both components from the chromosome rather than a shuttle vector would also be expected to result in higher expression. This is because the pURB500 plasmid, from which the *M. maripaludis* shuttle vectors are derived (*18*), was estimated to have a low copy number of approximately three per cell (*23*), whereas the *M. maripaludis* chromosome has 30-55 (*16*). Chromosomal expression would also theoretically satisfy the second design parameter: that the CRISPR/Cas9 components should have been expressed and formed functional complexes prior to challenge with the invader plasmid. This would be expected to increase the likelihood of complete invader plasmid clearance since the plasmid, which would

presumably enter the cell in a single copy following transformation, would theoretically be destroyed before it has had a chance to replicate.

This invader assay must be achievable using the genetic tools currently available in *M. maripaludis*. To prove destruction of the invader plasmid is due solely to specific targeting, one or both of the CRISPR/Cas9 components would ideally be placed under the control of an inducible promoter, and cells induced following transformation with invader plasmid shown to exhibit lower survival than un-induced controls. The single inducible promoter available in *M. maripaludis*, nif, is only expressed at intermediate levels when induced by growth with alanine (*28*). Protein expression from nif was found to be lower than from Phmv (*50*), which suggests it is weaker, and therefore does not meet the design constraint that the CRISPR/Cas9 components should be expressed as strongly as possible. Additionally, low level expression from nif under growth on ammonium (i.e. the un-induced state) has been reported in several studies (*35-37*), which could complicate interpretation of the invader assay output. The invader assay could equally be conducted using constitutive promoters. Two separate *M. maripaludis* strains could be produced, both constitutively expressing Cas9 and a gRNA from the chromosome. Critically, the two strains would differ in the targeting specificity of their gRNA: with the test strain expressing a gRNA which targets the invader plasmid, and the control strain expressing a gRNA which does not. When transformed with the plasmid invader, the test strain gRNA would be able to direct Cas9 to the invader, resulting in cleavage, whereas the control strain would not, so the plasmid would be retained. Both transformed strains would be plated in the presence of the antibiotic corresponding to the resistance gene carried by the invader plasmid, and a higher frequency of colonies recovered on control compared with test strain transformants would indicate that the CRISPR/Cas9 system was functional (Figure 1.1).

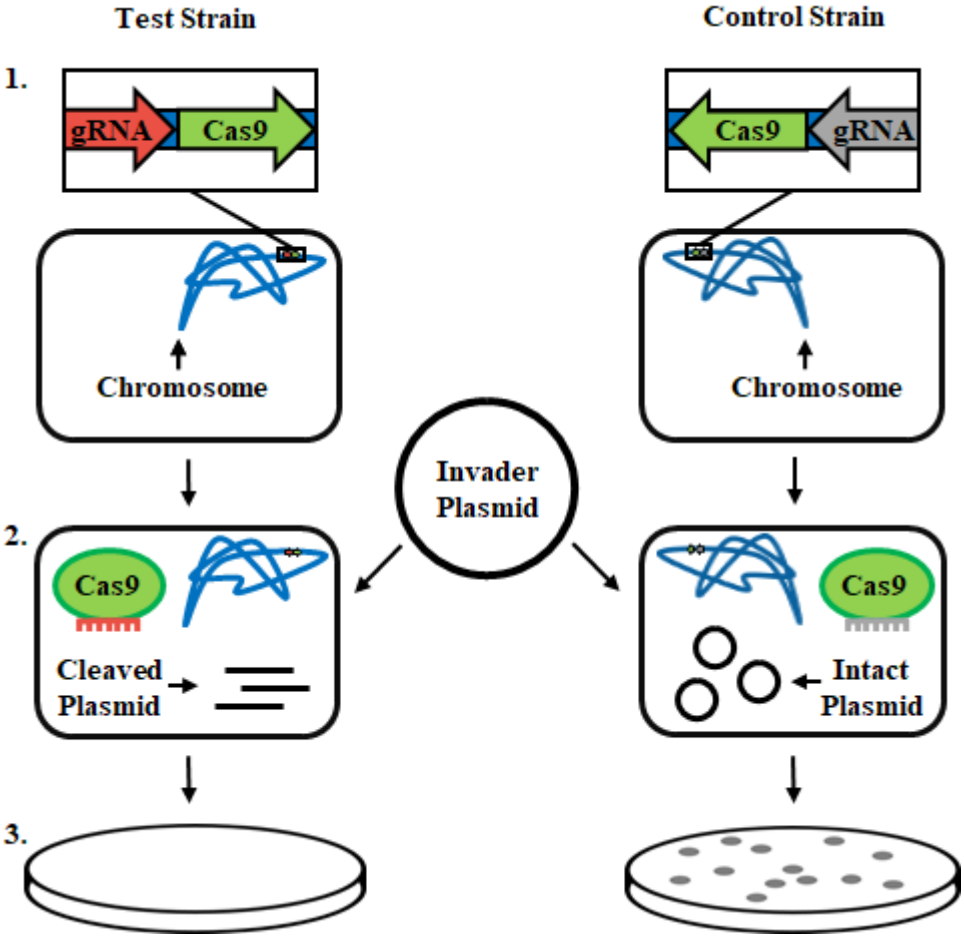

Figure 1.1: The plasmid invader assay.
1: Strains are produced by transforming Cas9 and a gRNA targeting the invader plasmid (left) or a control gRNA (right) onto the *M. maripaludis* chromosome. **2:** Transformation of the test strain with a puromycin resistant invader plasmid results in Cas9 mediated cleavage (left). The control strain cannot target the plasmid, which remains intact (right). **3:** Plating of the test strain on puromycin results in cell death: no colonies are recovered (left). Control plasmid transformants grow on puromycin: colonies are recovered (right).

If CRISPR/Cas9 mediated plasmid destruction were highly efficient, the most suitable *M. maripaludis* background strain/invader plasmid combination would be S0001/pAW42. This is because this combination offers the highest currently available transformation efficiency (*12*), and assuming highly efficient plasmid cleavage in test strain cells, would yield the greatest difference in colonies recovered between the test and control strains, since more cells in total would be expected to be successfully transformed. If, however, Cas9 mediated plasmid cleavage is inefficient, this combination could be problematic since test strain cells may be unable to clear the plasmid invader effectively before it is able to replicate, which is thought to be more rapid in S0001 (*12*). In this scenario, the Mm900

strain and the smallest available plasmid it can be used with, pLW40 (*12, 24*), would be more appropriate. The disadvantage of this approach is that the generally lower transformation efficiency would be expected to result in overall lower colony recovery, which could make detecting differences between the test and control strains more difficult. Since the CRISPR/Cas9 system has not to our knowledge been used in *M. maripaludis*, it is not possible to accurately predict how efficient it will be. For this reason, it would be appropriate to conduct the invader assay using both background strain/invader plasmid combinations.

1.1.7 Project Aims

The primary aim of this project is to determine whether the CRISPR/Cas9 system is functional in *M. maripaludis* through use of a plasmid invader assay (Figure 1.1). This can be divided into a number of sub-aims:

1. A set of constructs carrying the Cas9 and gRNA components required to produce the invader assay *M. maripaludis* test and control strains will to be designed, and an appropriate assembly scheme selected.
2. The components will be assembled in a cloning vector and then transferred into the integrative *M. maripaludis* transformation vectors, which will be used to transform *M. maripaludis* strains S0001 and Mm900. The established positive/negative selection system will then be used to recover the test and control strains carrying the CRISPR/Cas9 assemblies on their chromosomes.
3. The invader assay will be performed by transforming the invader plasmids pLW40/pAW40 into these strains, with lower survival in test compared with control strains indicating functionality of the CRISPR/Cas9 system.

Chapter 2: Materials and Methods

2.1.1 Molecular Work

2.1.1.1 Agarose gel electrophoresis

DNA was separated by size on gels which contained agarose powder (Melford Biolabs) dissolved to a final concentration of 1% in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). 3.5 µl of SYBR® Safe DNA Gel Stain (Thermo Scientific) was added to each 120 ml gel, and 5 µl of GeneRuler 1 kb DNA Ladder (Thermo Scientific) was loaded on each for size estimation. 6x DNA Loading Dye (Thermo Scientific) was added to samples to a final concentration of 1x, with the exception of PCR products which contained premixed loading dye. Gels were run at 90 V until the loading dye had migrated approximately two thirds of the distance down the gel, then imaged.

2.1.1.2 Resuspending gBlocks

Lyophilised gBlocks (IDT) were centrifuged at 3000 g for 5 seconds and the pelleted material was resuspended in TE buffer (10mM Tris, 0.1mM EDTA, pH 8) to a final concentration of 20 ng/µl, then stored at 4°C.

2.1.1.3 Restriction digestion

All digestions were carried out using New England Biolabs (NEB) restrictions enzymes. 1-2 µg of plasmid DNA (pDNA) products were digested in standard 50 µl reactions, or larger/smaller reaction mixtures scaled appropriately, as per the manufacturer's instructions. 100 ng of digested product was first inspected on a gel to check the size was as expected. For plasmid material intended for downstream ligation or dephosphorylation, the remainder of the reaction was then either run on a gel and extracted, or used directly in the dephosphorylation reaction. For gBlocks, 500 ng of those over 500 bp in size, and 250 ng of those under, were digested in 50 µl reactions which were otherwise as per the manufacturer's instructions. 20 ng of product was inspected on a gel, and the remainder was PCR purified.

2.1.1.4 PCR purification

Digested or dephosphorylated products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) as advised by the manufacturer, except the final elution step was carried out in 30 µl of Nuclease-Free Water (Promega).

2.1.1.5 Gel extraction

Desired bands were extracted from agarose gels with the Wizard® SV Gel and PCR Clean-Up System (Promega) as per the manufacturer's instructions, except that in some cases the final elution step was carried out in 30 µl of Nuclease-Free Water (Promega).

2.1.1.6 Plasmid dephosphorylation

40 µl of pDNA (1.6 µg) which had been digested as described above (see "Restriction digestion") was added directly to 5 µl of 1 U/µl Shrimp Alkaline Phosphatase (Takara Bio) and 5 µl of 10X SAP Buffer (Takara Bio). The reaction was mixed by pipetting, and incubated at 37°C for 30 minutes, then at 65°C for 15 minutes, then PCR purified.

2.1.1.7 Ligation

Digested products were ligated using T4 DNA Ligase (M0202S, NEB). 1:1, 3:1 and 6:1 insert:vector molar ratios were used. Reactions were carried out as advised by the manufacturer, except with the heat inactivation step omitted. Most reaction mixtures contained the suggested 50 ng of recipient vector in a 20 µl total volume, but in some cases 25 ng of vector was used. Following incubation at room temperature for 10 minutes, 10 µl of ligation reaction was transformed into *E. coli* (see "Transformation of ligation products/plasmids into *E.coli*"). If this did not yield satisfactory results on plating, the remainder of the reaction which had subsequently been incubated at 4°C overnight was also transformed.

2.1.1.8 Golden Gate reactions

For Golden Gate reactions the "Long protocol in ligase buffer" as explained in: (synbio.tsl.ac.uk/uploads/a94d80d88a4d03acbe8b53da74fdcf49.pdf) was used. For the reaction mixture, the holding vectors containing the fragments to be combined were added to 200 ng of destination vector at a 2:1 (holding:destination vector) molar ratio. 0.5 µl of 400 U/µl T4 DNA Ligase (NEB), 1.5 µl of T4 DNA Ligase Buffer (NEB), 0.5 µl of 10 U/µl BpiI (Thermo Scientific) and 1.5 µl of 10X BSA (NEB) were then added, made up to 20 µl with dH_2O, and mixed by pipetting. Reactions were cycled in the Prime Thermal Cycler (Techne) as follows: 37°C for 20 minutes, 26 cycles of 37°C for 4 minutes, 16°C for 4 minutes and 50°C for 5 minutes, then a final step of 80°C for 5 minutes. Reactions were held at 16°C and 5 µl of each was transformed into *E. coli*.

2.1.1.9 Transformation of ligation products/plasmids into *E. coli*

Material was transformed into chemically competent *E. coli* strain DH5α (prepared as described in "Generating chemically competent cells (*E. coli* DH5α)"), or dam$^-$/dcm$^-$ competent *E. coli* (genotype: *ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)* Tet^S *endA1 rspL136* (Str^R) *dam13::Tn9* (Cam^R) *xylA-5 mtl-1 thi-1 mcrB1 hsdR2*) (C2925H, NEB). For DH5α, 5 or 10 µl of material (either ligation reaction or 1 ng total of intact plasmid) was added to 100 µl of competent cells which had been thawed on ice. The tube was mixed by flicking, incubated on ice for 30 minutes and heat shocked at 42ºC for 1 minute. Cells were incubated on ice for 2 minutes and then 1 ml of LB medium (which had been pre-incubated at 37°C) was added. The sample was incubated at 37°C under 200 rpm shaking for 45 minutes. A 200 µl aliquot was spread on an LB agar plate containing antibiotic, IPTG and X-gal where appropriate (see "LB agar plates"). The remainder of the sample was concentrated by pelleting at 6800 g for 2 minutes, removing the supernatant leaving approximately 100-200 µl, and resuspending the pellet in this. This was spread on an LB agar plate, and plates were incubated at 37°C overnight. For dam$^-$/dcm$^-$ competent *E. coli*, 1 ng of intact plasmid diluted in a final volume of 5 µl with dH_2O was transformed according to the manufacturer's instructions with the following exceptions. The 37°C incubation step was done at 200 rpm shaking. An additional concentrated sample (produced as described above) was plated for each transformation.

2.1.1.10 Colony PCR

Colony PCRs were carried out using DreamTaq Green DNA Polymerase (Thermo Scientific). A master mixture was set up according to the manufacturer's instructions, with the exception that template DNA was not added, and then separated into 10 µl aliquots in PCR tubes. Under sterile conditions, single *E. coli* colonies were selected from transformation plates using pipette tips and inserted into 10 µl reaction mixtures in separate PCR tubes. Tips were left to stand for 2 minutes and then spread onto an LB agar master plate in sectors labelled according to the tube of origin. Master plates were incubated overnight at 37 °C. PCRs were carried out in the Prime Thermal Cycler (Techne). The Thermo Scientific Tm calculator (www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html) was used to select an appropriate annealing temperature according to the primers in each reaction. Reactions

were run for 35 cycles, with cycling conditions otherwise as suggested by the manufacturer. 6 µl of each PCR product was analysed by agarose gel electrophoresis.

2.1.1.11 Plasmid DNA isolation (from *E. coli*)

Plasmid DNA was isolated from 5 ml of overnight *E. coli* culture grown in LB medium supplemented with antibiotic where appropriate (see "*E. coli* culturing"). Either the GeneJET Plasmid Miniprep Kit (Thermo Scientific), or QIAprep® Spin Miniprep Kit (Qiagen), was used according to the manufacturer's instructions, with the exception that plasmids which had previously been recovered at very low concentrations were eluted in 30 µl of the respective manufacturer's Elution Buffer.

2.1.1.12 Generating chemically competent cells (*E. coli* DH5α)

All steps were carried out under sterile conditions. All centrifugations were done for 10 minutes at 3214 g at 4ºC. *E. coli* DH5a was streaked from a glycerol stock onto an LB agar plate and grown overnight at 37ºC. A single colony was then used to inoculate 5 ml of LB medium, which was incubated overnight at 37ºC under 200 rpm shaking. 500 µl of overnight culture was used to inoculate 50 ml of LB medium in a 250 ml flask, which was incubated at 37ºC under 200 rpm shaking for a further two hours. OD_{650} measurements were then taken every 15 minutes until the culture reached OD_{650} 0.5-0.6. For the following steps the cell suspension was kept on ice as much as possible. The culture was transferred to a 50 ml Falcon tube which had been pre-chilled on ice, and incubated on ice for a further five minutes. The culture was then centrifuged, the supernatant discarded, and the pellet resuspended in 10 ml of pre-chilled 0.1M $CaCl_2$. The suspension was incubated for 20 minutes on ice and then spun again. The supernatant was discarded and the pellet resuspended in 0.6 ml of 50% glycerol which had been previously added to 1.4 ml of 0.1 M $CaCl_2$ and pre-chilled. 100 µl aliquots of the resuspension were snap frozen in liquid nitrogen and stored at -80ºC.

2.1.1.13 *E. coli* culturing

E. coli was grown in LB medium supplemented (where appropriate) with antibiotics to a final concentration of 100 µg/ml for ampicillin or 50 µg/ml for kanamycin. Under sterile conditions, a single *E. coli* colony from an LB agar plate, or a piece of frozen material from a glycerol stock was picked on a pipette tip and used to inoculate the medium (usually 5 ml) in a 50 ml Falcon tube. The tube was sealed and incubated at 37ºC, 200 rpm shaking overnight.

2.1.1.14 Long term plasmid storage

Glycerol stocks were produced for all plasmids which had been confirmed to contain the correct insert by sequencing. Stocks were produced by mixing 500 µl of overnight *E. coli* culture containing the sequenced plasmid with 500 µl of 50% glycerol solution under sterile conditions in 2 ml Cryo.s™ screw cap tubes (Greiner Bio-One). Tubes were stored at -80°C.

2.1.1.15 Sequencing reactions

Plasmid inserts were Sanger sequenced using the Lightrun tube service from GATC Biotech. Premixed reactions were produced by adding 400 ng of pDNA to 25 µM of the appropriate primer in a final volume of 10 µl, and sent to the company for sequencing.

2.1.2 Aerobic Media and Antibiotics

2.1.2.1 LB medium

For 1 L, 10 g of NaCl, 10 g of tryptone and 5 g of yeast extract were added to 1 L dH_2O and mixed with a magnetic stirrer for 10 minutes. Medium was autoclaved at 121°C, 15 psi for 20 minutes.

2.1.2.2 LB agar plates

For LB agar, the protocol was as for "LB medium", except that agar powder was added just before autoclaving (15 g per 1 L LB). The medium was cooled to approximately 55°C. Under sterile conditions, antibiotics were added (if required) to a final concentration of 100 µg/ml (ampicillin) or 50 µg/ml (kanamycin) and mixed. The media was then poured onto plates (approximately 25 ml per plate), which were left with the lids partially off for 1 hour to set. Lids were replaced and plates were stored at 4°C. For blue/white screening, 40 µl of 100 mM IPTG and 120 µl of 20 mg/ml X-gal were spread over the surface of the plate just before bacteria were added.

2.1.2.3 Ampicillin and kanamycin

For 100 mg/ml ampicillin stock solution, 1 g of ampicillin powder was dissolved in 10 ml dH_2O. The solution was filter sterilised into 1 ml aliquots, which were stored at -21°C. 30 mg/ml kanamycin stock solution was made in the same way except 0.3 g of powder was used.

2.1.2.4 IPTG

100 mM stock solutions were prepared and stored as described for antibiotics (see "Ampicillin and kanamycin"), except 0.238 g of IPTG powder was used per 10 ml.

2.1.2.5 X-gal

For 20 mg/ml stock solution, 0.2 g of X-gal was added to 10 ml of dimethylformamide, and dissolved by vortexing. 1 ml aliquots were stored in tubes wrapped in aluminium foil at -21°C.

2.1.3 *M. maripaludis*

2.1.3.1 *M. maripaludis* culturing

M. maripaludis cultures were grown in Anaerobic Culture Tubes (Bellco Glass) containing 5 ml of McCas medium (see below) sealed with 20 mm Blue Butyl Rubber Stoppers (GPE Scientific) and crimped with 20 mm Aluminium Crimp Caps (Supelco)**.** To establish cultures, a glycerol stock containing the desired strain was transferred to the anaerobic chamber on dry ice, along with a fresh McCas tube. The stoppers were removed and a piece of frozen culture was transferred from the stock to the McCas tube. The stoppers were replaced and the tubes removed from the chamber. 0.1 ml of 2.5% Na_2S was then transferred anaerobically to the inoculated McCas tube using a needle and syringe which had been pre-gassed with H_2/CO_2. This was then pressurised to 40 psi with H_2/CO_2 (80:20) and incubated at 37°C under 110 rpm shaking. Once dense growth was visible (typically between 24-48 hours), a subculture was made by transferring 0.2-0.5 ml of culture using a needle and pre-gassed syringe into a fresh McCas tube which had had Na_2S added as explained above. This was then incubated as explained above. Cultures were maintained by subculturing continuously in this manner throughout the project.

2.1.3.2 McCas medium

All components (Table 2.1) with the exception of DTT were combined in a 1 L round bottomed flask (Duran) and mixed with a magnetic stirrer for 10 minutes. DTT was then added, and the solution placed under a stream of N_2/CO_2 (30 psi) by inserting a gassing cannula into the flask. A bung was placed in the neck of the flask such that a small gap where it contacted the cannula was left for displaced air to escape. The solution was made anaerobic by heating with a Bunsen burner (blue flame) while swirling continuously. Once the colour had changed from blue to colourless, the vessel was sealed by simultaneously

applying pressure to the bung while withdrawing the cannula. The flask was taken into the anaerobic chamber, and 5 ml aliquots of medium dispensed into Anaerobic Culture Tubes (Bellco Glass), which were then sealed with 20 mm Blue Butyl Rubber Stoppers (GPE Scientific) (both of which had been placed in the chamber at least 2 hours previously). The tubes were removed from the chamber and crimped with 20 mm Aluminium Crimp Caps (Supelco). The headspace in the tubes was exchanged 4 times by pressurising to 20 psi with H_2/CO_2 (80:20), followed by vacuum withdrawal. Tubes were pressurised to 30 psi with H_2/CO_2 (80:20) and autoclaved at 121°C, 15 psi for 20 minutes.

Table 2.1: McCas medium components. See below for recipes of solutions marked "*".

Component	Amount
dH_2O	100 ml
$NaHCO_3$	1 g
NH_4Cl	0.1 g
NaCl	4.4 g
Na Acetate•$3H_2O$	0.28 g
Bacto Casamino Acids	0.4 g
Dithiothreitol (DTT)	0.1 g
N-free trace minerals (1000x)*	0.2 ml
Vitamin solution (100x)*	2 ml
N-free general salts solution*	100 ml
$FeSO_4$ solution*	1 ml
K_2HPO_4 solution*	2 ml
Rasazurin solution*	0.2 ml

2.1.3.3 N-free trace minerals (1000x)

For 100 ml, Na_3Citrate•$2H_2O$ (Table 2.2) was dissolved in 80 ml dH_2O and the pH was adjusted to 6.5. The remaining components (Table 2.2) were dissolved in this, and the solution was made up to 100 ml with dH_2O.

Table 2.2: N-free trace minerals (1000x) components.

Component	Amount (g/100 ml)
$Na_3Citrate•2H_2O$	2.1
$MnSO_4•H_2O$	0.45
$CoCl_2•6H_2O$	0.1
$ZnSO_4•7H_2O$	0.1
$CuSO_4•5H_2O$	0.01
$AlK(SO_4)_2•12H_2O$	0.018
H_3BO_3	0.01
$Na_2MoO_4•2H_2O$	0.1
$NiCl_2•6H_2O$	0.025
$NaSeO_3•5H_2O$	0.3
V(III)Cl	0.01
$Na_2WO_4•2H_2O$	0.0033

2.1.3.4 Vitamin solution (100x)

For 1 L, components (Table 2.3) were dissolved in 1 L dH_2O.

Table 2.3: Vitamin Solution (100x) components.

Component	Amount (g/L)
Biotin	0.002
Folic acid	0.002
Pyridoxine HCL	0.01
Thiamine HCL	0.005
Riboflavin	0.005
Nicotinic acid	0.005
DL-calcium pentothenate	0.005
Vitamin B_{12}	0.0001
p-aminobenzoic acid	0.005
Lipoic acid	0.005

2.1.3.5 N-free general salts solution

For 1 L, components (Table 2.4) were dissolved in 800 ml dH_2O, then made up to 1 L with dH_2O.

Table 2.4: N-free general salts solution components.

Component	Amount (g/L)
KCL	0.68
$MgCL_2•6H_2O$	5.5
$MgSO_4•7H_2O$	6.9
$CaCl_2•2H_2O$	0.28
NH_4Cl	1

2.1.3.6 $FeSO_4$ solution

For 100 ml, 0.19 g of $FeSO_4$ was dissolved in 100 ml of 10 mM HCL.

2.1.3.7 K_2HPO_4 solution

For 1L, 14 g of K_2HPO_4 was dissolved in 1 L dH_2O.

2.1.3.8 Resazurin solution

For 100 ml, 0.1 g Resazurin was dissolved in 100 ml dH_2O.

2.1.3.9 Transformation of plasmids into *M. maripaludis*

A slightly modified version of the PEG-based transformation protocol (*17*) was carried out as explained in: (faculty.washington.edu/leighj/protocols/transformation_procedure.pdf). All centrifugation steps were carried out at 1070 g, 20ºC for 10 minutes. *M. maripaludis* cultures (S0001 strain) to be transformed were grown to an OD_{600} of 0.7-1.0, and then pressurised to 30 psi with H_2:CO_2 (80:20). Cells were centrifuged and the supernatant removed by inverting the tube, inserting a needle through the stopper and using the pressure to pushed it out. 5 ml of transformation buffer was then added using a needle and syringe which had been pre-gassed with H_2/CO_2, and the pellet resuspended by tapping the tube. The tube was repressurised to 30 psi with H_2/CO_2 (80:20), and centrifuged. The supernatant was removed and the pellet resuspended in 0.375 transformation buffer as described above. The tube was taken into the anaerobic chamber, the stopper removed, and 5 µg of pDNA (which had been placed in the chamber with the cap open at least two hours previously) was added and mixed immediately by tapping. The stopper was replaced and the tube removed from the chamber. The tube was flushed with N_2 for 1 minute, and 0.225 µl of PEG solution was added using a needle and syringe (pre-gassed as above). Tubes were pressurised to 30 psi with N_2 and then incubated at 37ºC for 1 hour. A fresh McCas tube was prepared by adding 0.1 ml of 2.5% Na_2S with a pre-gassed needle and syringe, and

pressurised to 30 psi with H_2/CO_2 (80:20). The fresh medium was transferred into the transformation tube by inserting the long end of a Vacutainer® needle (BD Diagnostics) through the stopper of the transformation tube, and just as the pressure was lost pushing the short end through the stopper of the inverted McCas tube. The needle was removed and the transformation tube was mixed by tapping, repressurised to 30 psi with H_2/CO_2 (80:20), and centrifuged. The tube was then inverted carefully so as not to disturb the pellet, the long end of a vacutainer needle was inserted through the stopper, and the pressure used to expel the supernatant. Once the pressure was released, the short end of the needle was pushed through the stopper of a fresh McCas tube (prepared as above), and the two tubes were inverted, resulting in the pressurised transfer of the fresh medium into the transformation tube. The needle was removed and the pellet resuspended by tapping. The tube was flushed with H_2/CO_2 for 1 minute, and then pressurised to 40 psi with H_2/CO_2 (80:20). Tubes were incubated at 37ºC under 110 rpm shaking overnight, and then subcultured into appropriate selection (see below).

2.1.3.10 *M. maripaludis* outgrowth and selection following transformation

All incubation steps were done at 37ºC under 110 rpm shaking. Following overnight growth (see "*M. maripaludis* transformation"), 0.2 ml of transformed culture was used to inoculate a fresh McCas tube which had had 0.1 ml of 2.5% Na_2S, and neomycin to a final concentration of 0.5 mg/ml added using pre-gassed needles. This was then incubated for 2 days. The following steps were not carried out in this study but are included for reference. 0.05 ml of the resulting culture is subcultured into McCas containing neomycin, (prepared as above), and incubated overnight. 0.05 ml of overnight growth is then subcultured into plain McCas tubes (i.e. containing 0.1 ml of Na_2S but no selection), and incubated overnight. These cultures are then spread onto McCas agar plates containing 8-azahypoxanthine to a final concentration of 250 µg/ml, and incubated in a pressure vessel at 15 psi, 37ºC for 3-4 days. Resulting colonies are then screened for the desired modification.

2.1.4 Anaerobic Antibiotics and Solutions

2.1.4.1 Reducing agent

100 ml of dH_2O and 100 ml of 1M Tris base were made anaerobic by transferring them in 250 ml bottles (Duran) to the anaerobic chamber, removing the caps and leaving them for 24 hours with occasional swirling. For 50 ml of reducing agent, 0.5 g of cysteine•HCL•H_2O and 0.39 g of DTT (in tubes that had been opened in the chamber 2 hours previously) were

dissolved in 50 ml anaerobic dH_2O. The solution was then filter sterilised through a 0.22 µm filter unit (Millex®) into a sterile 100 ml anaerobic media bottle which has been placed in the chamber the previous day. The pH was adjusted to 7.5 with the anaerobic Tris base. For storage, the bottle was stoppered, taken out of the chamber, crimped, pressurised to 12 psi with N_2, and then placed at 4°C.

2.1.4.2 Transformation buffer

For 250 ml, all components (Table 2.5) were added to 250 ml dH_2O, mixed with a magnetic stirrer for 10 minutes, and the pH was adjusted to 7.5. The solution was taken into the anaerobic chamber and made anaerobic as explained previously (see "Reducing agent"). 50 ml aliquots were then filter sterilised into sterile anaerobic media bottles which had been placed in the chamber the previous day, and 1 ml or reducing agent (see above) was added to each and mixed. Bottles were left open with occasional agitation until the colour changed from blue to colourless, at which point they were prepared for storage as described (see "Reducing agent").

Table 2.5: Transformation buffer components.
See above for recipe of solution marked "*".

Component	Amount
Tris	1.5 g
Sucrose	30 g
NaCl	5.5 g
$MgCl_2 \cdot 6H_2O$	0.05 g
Resazurin Solution*	150 µl

2.1.4.3 PEG solution

For 100 ml 40% (wt/vol) PEG solution, 40 g of PEG 8000 (Sigma-Aldrich) was added to 75 ml of transformation buffer (which had been prepared as explained above up to the point where the pH was adjusted to 7.5). The solution was mixed at 60°C with a magnetic stirrer for 30 minutes, and then made up to 100 ml with transformation buffer. The solution was taken into the anaerobic chamber and prepared from this point as explained for the transformation buffer, except that it was not filter sterilised.

2.1.4.4 Na_2S solution

For 25% Na_2S solution, 12.5 g of sodium sulfide nonahydrate crystals (Sigma-Aldrich) were weighed out, taken into the anaerobic chamber and dissolved in 50 ml of anaerobic dH_2O in an anaerobic bottle which had been placed in the chamber the previous day. 2.5% Na_2S solution was made by diluting the 25% solution in anaerobic water. Solutions were stoppered, crimped and stored at room temperature.

2.1.4.5 Neomycin

For 25 mg/ml stock solutions, 0.25 g of neomycin sulphate powder was weighed out, taken into the anaerobic chamber and dissolved in 10 ml anaerobic dH_2O. The solution was filter sterilised into an anaerobic media bottle then made anaerobic and stored as explained above (see "Reducing agent").

2.1.5 In Silico Methods

2.1.5.1 Codon optimisation

The SpCas9 amino acid sequence was obtained from UniProt (*51*) (accession number**:** Q99ZW2). For codon optimisation using Jcat (*52*), the *M. maripaludis* S2 strain codon usage table was selected, with all other settings set to default. As the COOL codon optimisation resource (*53*) does not directly support *M. maripaludis*, the codon usage frequencies for the S2 strain were imported as a custom profile. This data along with the total GC and GC3 content values was obtained from the Codon Usage Database (*54*), which sources data from Genbank (*55*). For the final run used to generate the SpCas9 codon optimisation used in this study (Mmp-Cas9) the tool was run using default settings with the following exceptions: optimisation criteria: ICU "maximise", all other options: "ignore"; codon auto correction: 0; 5' folding instability: maximise (to a window of the first 40 bases); total GC content target percentage: 34.01; GC3 content target percentage: 25.17. Output sequences were filtered by ICU score, and of those with the joint highest score one was arbitrarily selected.

2.1.5.2 gRNA design

gRNA target sequences were designed and/or tested using the in-built "Find CRISPR Sites" tool in Geneious 10.0.7 (*56*). To select potential target sites and simultaneously generate their predicted on-target scores, the region to be targeted within the appropriate gene was selected and the tool was run on "score sites through their on-target activity" with default parameters. On-target scores are calculated by Geneious as explained in (*57*). For off-target

analysis, a library was first created containing the *M. maripaludis* S2 genome sequence, the sequences of plasmids pCRPrtneo, pBLPrt, pLW40 and pAW42, and in the case of the eGFP gRNA the eGFP gene sequence. Off-target scores for all sites within the candidate target regions were then generated by selecting "score sites through off-target analysis" using default setting with the following amendments: speed and filter strategy: "slow – score all sites"; maximum mismatches allowed against off-target sites: 23; maximum mismatches allowed to be indels: 4; score against off-target database: set to the respective library for that target region (see above). Off-target scores are calculated by Geneious as explained in (*58*).

2.1.5.3 Random sequence generation

Random control gRNA targeting sequences were generated using the Random DNA Sequence Generator resource (http://www.faculty.ucr.edu/~mmaduro/random.htm).

2.1.5.4 Checking complexity IDT gblocks

Potential gBlocks were tested using the "Test Complexity" function on the IDT website (https://www.idtdna.com/site/Order/gblockentry).

2.1.5.5 Removal of incompatible restriction enzyme sites

Pre-existing restriction enzyme recognition sites incompatible with the designed cloning schemes were removed from gBlocks at the design stage. For sites located within protein coding components, the codon usage frequencies corresponding to the organism in which that component would be expressed (*M. maripaludis* S2 and *E. coli* K12 strains) were obtained from the Codon Usage Database (*54*). For each site, all possible base changes which would result in removal of the restriction site while maintaining the identity of the amino acid designated by the codon they were located in (i.e. silent) were considered, and the change resulting in the use of the next most frequently used codon for that amino acid was selected. For restriction sites located outside of coding regions, a single base was arbitrarily changed to remove them.

2.1.5.6 Construct design and testing

Constructs were designed using Geneious 10.0.7 (*56*). Cloning strategies were checked by in silico simulation of all of the digestion/ligation steps in both Geneious (*56*) and Benchling (www.Benchling.com).

2.1.5.7 Primer design

Primers (Table A1, Appendix) were designed using the in-built primer design tool in Geneious 10.0.7 (*56*).

2.1.5.8 Sequence analysis

Sequence analysis was conducted using Geneious 10.0.7 (*56*). Reads were trimmed manually to remove low quality or ambiguous base calls. These were then aligned to an in silico prediction of the appropriate plasmid containing the expected insert. Finally, the alignment was inspected manually for mismatches.

Chapter 3: In Silico Planning

3.1 Introduction

The object of this project is to determine whether the CRISPR/Cas9 system is functional in *M. maripaludis* through use of a plasmid invader assay (Figure 1.1). The production of the test and control strains which underpin this assay would require the introduction of the two components of the CRISPR/Cas9 system: the Cas9 gene and a gRNA, onto the *M. maripaludis* chromosome. Each would need to be placed into a module containing a suitable promoter and terminator (Figure 3.1). The introduction of these expression modules onto the chromosome would be achieved using the existing *M. maripaludis* transformation protocol (*19*), which has three stages. In the first, material to be introduced (i.e. the two modules) is assembled within a cloning vector in *E. coli*. Next, the portion of the cloning vector containing both modules is subcloned into an *M. maripaludis* transformation vector. Finally, *M. maripaludis* is transformed with this vector, resulting in chromosomal integration of the expression modules, yielding the test and control strains.

Figure 3.1: Composition of the Cas9 and gRNA modules.
Two modules will be transformed into *M. maripaludis* to generate the test and control strains: both are composed of three separate elements. The Cas9 module (left) contains the Cas9 gene and the gRNA module (right) contains either the test or control gRNA. Both are under the control of a suitable *M. maripaludis* promoter and terminator.

3.1.1 Chapter Aims

The first step of implementing the plasmid invader assay, and the aim of this chapter, is to design the two expression modules to be assembled within the *E. coli* cloning vector. This can be sub-divided into two parts. In part one, the individual components of each module need to be designed. In part two, a system by which they can be physically assembled together into a cloning vector must be planned. For the invader assay to confirm that the CRISPR/Cas9 system is functional in *M. maripaludis*, there must be detectably lower survival rates in test strain cells compared with the controls (Figure 1.1). The factor

responsible for invader plasmid destruction and ultimately cell death in the test strain (i.e. the test output) is the functional Cas9/gRNA complex. It is anticipated that greater amounts of this complex will result in higher levels of plasmid destruction, and therefore a more pronounced difference in test and control strain cell survival (see Chapter 1). For this reason the primary design constraint with regard to the module components and their assembly will be ensuring that expression levels of the gRNA and Cas9 protein are as high as possible.

3.1.2 Component Design

3.1.2.1 Promoter and terminator

Both expression modules (Figure 3.1) must be placed under the control of a strong promoter which is functional in *M. maripaludis*. The gRNA promoter has an additional design constraint in that it must have a strict transcriptional start site. This is because the 20 nt portion of the gRNA responsible for targeting in the CRISPR/Cas9 system is transcribed first (*42*). Of the promoters available for *M. maripaludis*, (*17, 28, 30, 31*), only two: nif (*59*) and a minimal version of Phmv (James Chong, personal communication) have had their transcriptional start sites mapped. As discussed previously, nif is not a suitable option for the invader assay (see Chapter 1). The only remaining option, Phmv, is both constitutive and strong under normal growth conditions (*60*), making it an appropriate choice. Some promoters exhibit bias, whereby they have a preference for a certain base to be transcribed first: for example the U6 promoter which is routinely used to express the gRNA component in eukaryotic CRISPR/Cas9 studies requires a G at the +1 position (*42*). It is not known whether Phmv exhibits promoter bias, because the promoter has only been used in one conformation. However, in this conformation the first base transcribed is an adenine. To account for the possibility of promoter bias in the minimal Phmv it would be desirable (if possible) to select gRNA sequences beginning with adenine.

Phmv is known to drive high expression levels (*60*), which also makes it an attractive option for use with the Cas9 module. One potential issue with using the same promoter twice is that it could cause undesirable recombination events, either within the cloning vectors or on the *M. maripaludis* chromosome itself. However, using one of the other remaining promoters available in *M. maripaludis* to drive Cas9 expression also presents potential risks. The *M. vannielii* histone promoter has only been used in a single study (*31*), and is therefore a less reliable option than Phmv because it has been validated less extensively. The only other remaining option, Pmcr from *M. voltae*, is used to express the nemocyin

and puromycin resistance genes in the *M. maripaludis* transformation (*19*) and replicative (*12, 24*) vectors, and would therefore theoretically be subject to the same potential recombination difficulties. At just 73 base pairs (bp) in length, the minimal Phmv is relatively small, which makes the probability of recombination unlikely. Additionally, multiple instances of this promoter have been used to simultaneously drive the expression of several genes in *M. maripaludis* before without issue (James Chong, personal communication). Taken together, this suggests that Phmv is also a suitable promoter choice for the Cas9 module in the invader assay.

Both modules must also contain a transcriptional termination signal. The only functionally characterised terminator sequence in *M. maripaludis* is the *M. voltae* MCR terminator (Tmcr) (*17, 20, 21*), making this the only currently available option.

3.1.2.2 Cas9

Several alternative versions of the Cas9 protein have been used in CRISPR/Cas9 studies in various organisms (*43*). However, by far the most commonly used and therefore widely validated comes from *S. pyogenes* (SpCas9) (*42*), and will therefore be used in this study. When introduced in their native format, exogenous protein coding genes may be expressed poorly in new host organisms due to the presence of codons within the sequence which are rare in host genes (*61*). Such a situation would clearly be undesirable in the invader assay, since a key determinant of its success is high Cas9 expression. To avoid this, the SpCas9 sequence should be codon optimised to *M. maripaludis*. The basic premise of codon optimisation is to modify the codons in the exogenous gene such that they correspond as closely to the usage bias of the host as possible (*62*), thereby achieving a better fit to host expression parameters, for example by making more economical use of its tRNA repertoire (*63*).

Multiple different methods for optimising codons exist, many of which are supported by online tools (*64*). Systems based on the codon adaptation index (CAI) (*65*), such as the Jcat tool, aim to replace all codons in an exogenous DNA sequence with the most frequently used synonymous codon from a subset of highly expressed host genes (*52*). Jcat comes with in-built codon bias data for *M. maripaludis*, calculated from a set of genes predicted to be highly expressed in this organism (*52, 66*). An alternative but similarly simplistic option is to optimise based on host individual codon usage (ICU), which aims to select codons based on their proportional usage in a given set of genes (*64*). ICU based

optimisation is supported by the Codon Optimization On-Line (COOL) program, which offers additional parameters such as customisable target % GC content (*53*). However, unlike Jcat it does not directly support *M. maripaludis*.

3.1.3 gRNA

The invader assay requires the production of two *M. maripaludis* strains, which differ only in that one should be able to target the invader plasmids (test strain) and the other should not (control strain). Target site specificity of the CRISPR/Cas9 system is conferred by the gRNA. This consists of an invariant structural scaffold sequence, and a variable 20 nt targeting sequence. Since this 20 nt sequence is solely responsible for target site specificity (*67*), this stretch of DNA will be the only difference between the test and control strains.

gRNA design must be considered carefully, because different targeting sequences are known to exhibit different levels of on- and off-target activity (*68*). In the invader assay, poor on-target activity in the test strain (i.e. selecting a targeting sequence which does not deliver the Cas9 component effectively to the target site), could result in no or little invader plasmid cleavage. This could mean that even if the CRISPR/Cas9 system was functional in *M. maripaludis*, the level of cell death in the test strain might fall below the level of detection, resulting in a false negative outcome. Poor off-targeting (or a propensity for the selected gRNA targeting sequence to bind at unintended sites) could be similarly problematic. In the worst case, off-target cutting in the cloning or *M. maripaludis* transformation vectors could result in plasmid linearisation and subsequent destruction. Leaky activity of Phmv has been reported in *E. coli* (*30*), which means such activity could prohibit the production of the test strains. Off-targeting to the *M. maripaludis* chromosome could also cause the assay to fail. Cuts resulting in chromosomal destruction could lead to reduced survival, although it should be noted that the highly polyploid nature of *M. maripaludis* (*16*) would theoretically provide some level of buffering against this. Perhaps more significantly, ectopic targeting to the chromosome in the test strain would effectively sequester a certain proportion of Cas9 away from the intended plasmid invader target, which could result in undesirable increases in cell survival.

Algorithms which attempt to assess the quality of candidate gRNA sequences based on estimated on- (*57*) and off- (*58*) target activities have been developed. However, the extent to which these systems can accurately estimate these parameters are limited, particularly in the case of off-targeting (*69*). This means that while such algorithms can serve as useful

tools, their outputs should be regarded with caution. The most reliable method for determining on-target activity is experimental validation (*70*). For this reason, it would be preferential to use a validated gRNA target sequence in the test strain for the invader assay.

3.1.3.1 Test gRNA

The test gRNA must be capable of delivering Cas9 to the plasmid invader. The assay will be conducted in two different *M. maripaludis* strains, Mm900 and S0001, each of which has its own cognate replicative plasmid (plasmid invader), pLW40 and pAW42, respectively (see Chapter 1). The use of a single test gRNA capable of targeting both plasmids would be advantageous for two reasons. Firstly, it would reduce the number of plasmids which need to be produced, and secondly it would allow more reliable comparisons to be drawn between the results obtained in each strain. pAW42 was produced through excision of a portion of pLW40 (*12*), which means any target site on pAW42 would be present on both plasmids. The structural backbone of these plasmids can be ruled out as containing any pre-validated gRNA target sites since both are unique to *M. maripaludis*, and to our knowledge no study using the CRISPR/Cas9 system has been reported for this organism. The protein coding components of these plasmids: an ampicillin and a puromycin resistance gene, are good candidates for containing validated gRNA target sites since these elements are also present in many other cloning vectors. However, AmpR is unsuitable because it is present on the *M. maripaludis* transformation vectors, leaving just the PurR gene. Only a single publication could be found in which a gRNA targeting the PurR gene has been validated (*71*). The authors report their gRNA was able to knock out the PurR gene in mouse cells, resulting in a 30% rate of cell death when plated on puromycin. This gRNA is particularly suitable for the invader assay because in addition to being validated, it has also been successfully used in a similar context (i.e. a cell destruction assay).

Another way of obtaining a validated gRNA target site in a feature shared by the invader plasmids pLW40 and pAW42 would be to modify both such that a new target site was introduced. The advantage of this approach is that it would allow the selection of essentially any gRNA, which means that one which has been extensively validated could be selected. Highly validated gRNAs can be obtained from online resources such as the Addgene validated gRNAs database (www.addgene.org/crispr/reference/grna-sequence), and are typically located in genes which are highly targeted in CRISPR/Cas9 studies, such as the enhanced green fluorescent protein (eGFP) gene (*72-75*).

3.1.3.2 Control gRNA

The purpose of the control strain is to demonstrate that any increased level of cell death observed in the test strain is due solely to the specific targeting of Cas9 to the invader plasmid. As such the key requirement of the control gRNA is that in addition to being unable to target any part of the *M. maripaludis* chromosome or transformation plasmids, (as is the case for the test gRNA), it must also not be able to target the invader plasmids. An approach which could be used to generate the control gRNA would be to "scramble" the targeting sequence of the test gRNA (i.e. generate a random sequence with the same GC content).

3.1.4 Construct Design and Assembly

3.1.4.1 Obtaining parts

Once the individual components have been designed, they will need to be obtained and assembled into the cloning vector. The Cas9 component will have to be obtained as a synthesised piece of DNA since it will be codon optimised and thus consist of a novel sequence. The remaining components (i.e. Phmv, Tmcr and the gRNAs) could theoretically be generated by PCR. However, since they are small relative to the Cas9 gene a simpler option would be to have all the components chemically synthesised. DNA synthesis services such as gBlocks from IDT are now widely available (*76*). gBlocks are offered in sizes ranging from 500 bp to 3 kb, with cost and production times increasing with size (https://www.idtdna.com/pages/products/genes/gblocks-gene-fragments/). It should be noted that using this approach will impact on the construct design processes because the DNA sequences selected must abide by the synthesis constraints of the manufacturer. In addition to size limitations, highly repetitive sequences (particularly at the 5' and 3' ends), and sequences with very low GC content cannot be ordered as gBlocks (https://www.idtdna.com/pages/products/genes/gblocks-gene-fragments/).

3.1.4.2 Combining parts: Golden Gate

The total size of the unit to be assembled in the cloning vector will be too large to be ordered as a single gBlock (the Cas9 gene alone is 4104 bp in length (*77*)), and will therefore have to be split and ordered as separate fragments. This necessitates the use of a strategy for combining these into the cloning vector. The Golden Gate system (*78*) is particularly suitable for this application for a number of reasons. Firstly, unlike standard type II restriction enzyme cloning it is scarless (*79, 80*), which means the boundary which must

occur somewhere within the Cas9 gene by virtue of its size would be easier to select. Secondly, Golden Gate fragments can be easily varied and exchanged (*81*), which means the three different versions of the construct (i.e. harbouring the three alternative gRNAs) could be produced simply by modifying the gBlock containing this component. It is worth noting that this characteristic also makes the construct more versatile, as further modifications could easily be made, for example to incorporate different promoters or additional gRNAs.

The Golden Gate system makes use of type IIS restriction enzymes which cleave outside of their recognition sequences, leaving short overhangs (*78*). For example, the widely used BpiI enzyme (a BbsI isoschizomer) cleaves 2bp downstream of this site leaving a 4bp overhang (*82*). Complementarity between pre-determined overlaps following digestion allows different DNA fragments to be ligated together in a user specified order. This allows the assembly of multiple DNA fragments into a destination vector in a one pot, one step reaction (*83*) (Figure 3.2). While double stranded DNA (dsDNA) fragments can in principle be assembled directly into a destination vector, they are typically first incorporated into intermediate holding vectors to limit assembly errors (*80*).

Three key design constrains must be followed when designing a Golden Gate strategy. First, pre-existing type IIS recognition sites corresponding to the selected enzyme must be removed from elsewhere in the fragments/cloning vectors to prevent unwanted cleavage. Second, overhangs should be non-palindromic and differ from each other by at least two bases to prevent possible unwanted ligation events. Third, the holding and destination vectors should have different antibiotic resistances to allow cells transformed with undigested holding vector to be purged on plating (*79, 83*). Provided these parameters are met, Golden Gate assembly can in principle be carried out in any vector system. However, "MoClo" kits, which contain a series of holding and destination vectors with pre-determined overhangs can be used to streamline the process (*79, 84, 85*). These vectors often contain useful features, for example inclusion of a LacZ reporter gene in a "dropout" region bounded by two type IIS restriction sites allows cells harbouring undigested destination vector to be distinguished by blue/white screening (*84*) (Figure 3.2).

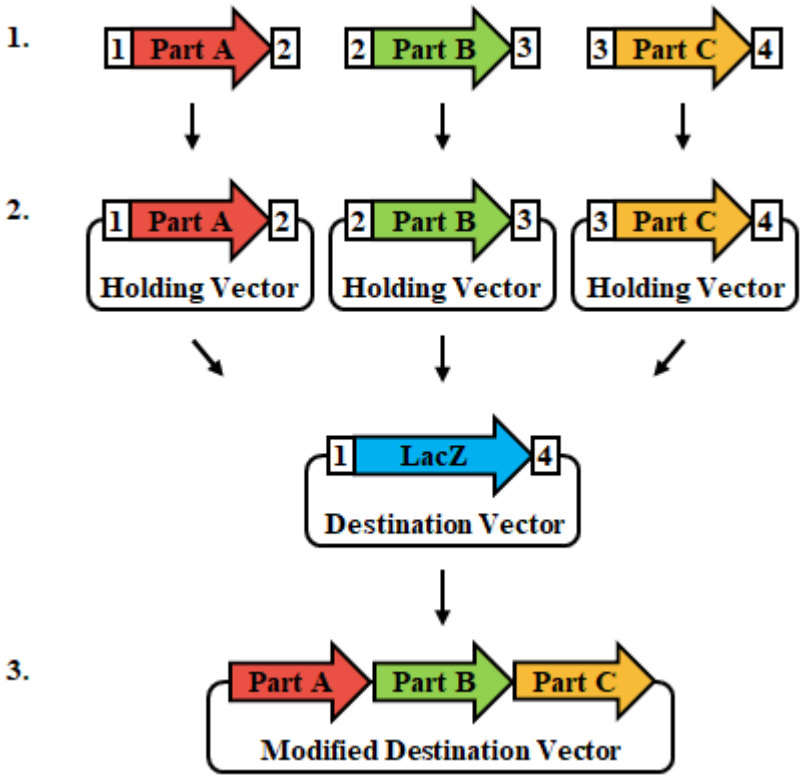

Figure 3.2: Golden Gate assembly.
1: Parts to be assembled (A-C) are first inserted into holding vectors. **2:** Each part contains type IIS restriction enzyme recognition sites (see below) at each end. The identity of the overlaps corresponding to each site are selected such that following IIS restriction digestion in a Golden Gate reaction mix containing each part and the destination vector, the 5' overlap of each part corresponds to the 3' overlap of the preceding part. **3:** The Golden Gate reaction results in the ligation of each part in the correct order into the destination vector, replacing the LacZ dropout region. Type IIS recognition sites are not retained in this vector, so it cannot be cut again in subsequent digestion steps. Following transformation, *E. coli* colonies harbouring the correctly modified destination vector are selected by blue/white screening. Numbered white boxes: type IIS restriction enzyme recognition sites. Numbers indicate the identity of the overlap corresponding to each site.

3.2 Results

3.2.1 Component Design

The first step in the development of the invader assay was to design a construct carrying the elements which when subsequently transformed into *M. maripaludis* would give rise to the test and control strains. These elements comprise two modules (Figure 3.1), which should express the two components of the CRISPR/Cas9 system as highly as possible. The first part of this process was to select/design the individual parts of each module according to the specifications set out previously (See Chapter 3, Introduction).

3.2.1.1 Cas9

To increase the likelihood of achieving high expression levels, the SpCas9 gene was codon optimised for *M. maripaludis*. Since the optimised Cas9 was to be ordered as a synthetic fragment, output sequences also had to meet the IDT gBlocks synthesis requirements. First, the Jcat tool was used to generate an optimised sequence (*52*), however the output contained multiple regions which were too repetitive for synthesis, rendering this sequence unusable.

Next, an optimisation was generated based on ICU in the COOL program (*53*). Since COOL does not directly support *M. maripaludis*, codon usage frequencies for this organism were first imported as a custom profile. In addition to the ICU parameter, the target total % GC and % GC3 (i.e. frequency of a guanine or cytosine at the third codon position) were set to correspond to those present in *M. maripaludis*. The resulting sequence did not contain any complexities that would prohibit gene synthesis. However, on visual inspection it appeared highly repetitive. Repetitive elements within genes can compromise expression levels and/or lead to instability (see Chapter 3, Discussion). To better estimate whether the level of repetitiveness in this output was likely to be problematic, it was aligned to other codon optimised Cas9 sequences which had been used successfully in other organisms: *Arabidopsis thaliana* (*86, 87*) and three different *Streptomyces* species (*88*). It was reasoned that if these sequences were similarly as repetitive, the repetitiveness of the optimisation was unlikely to negatively impact functionality. However, in all cases the *M. maripaludis* optimisation appeared to be more repetitive than the validated sequences (Figure A1, Appendix).

In an attempt to resolve this, additional optimisation runs were conducted in which several parameters intended to lower the repetitiveness of the output were tested in COOL. However, even on their most stringent settings none of these appeared to make a visually detectable impact in terms of repetitiveness. Additionally, some had a negative effect on the ICU score (i.e. resulted in optimisations which were more distant from the target *M. maripaludis* codon usage values). For the final run, those parameters which had negatively affected ICU were reverted. Those which had not were retained since they had not reduced the quality of the optimisation, but may have slightly reduced repetitiveness. The result of this final optimisation, hereafter referred to as Mmp-Cas9, was selected as the Cas9 sequence which would be used in *M. maripaludis* (Figure A2, Appendix). The codon usage frequencies were very close to the average values for native *M. maripaludis* genes (Figure 3.3). The GC3 value for the optimised sequence (25.17%) was identical to the average for *M. maripaludis* genes. The total GC content was also very similar: with a value of 34.04% in the optimisation compared with an average of 34.01% in *M. maripaludis* genes.

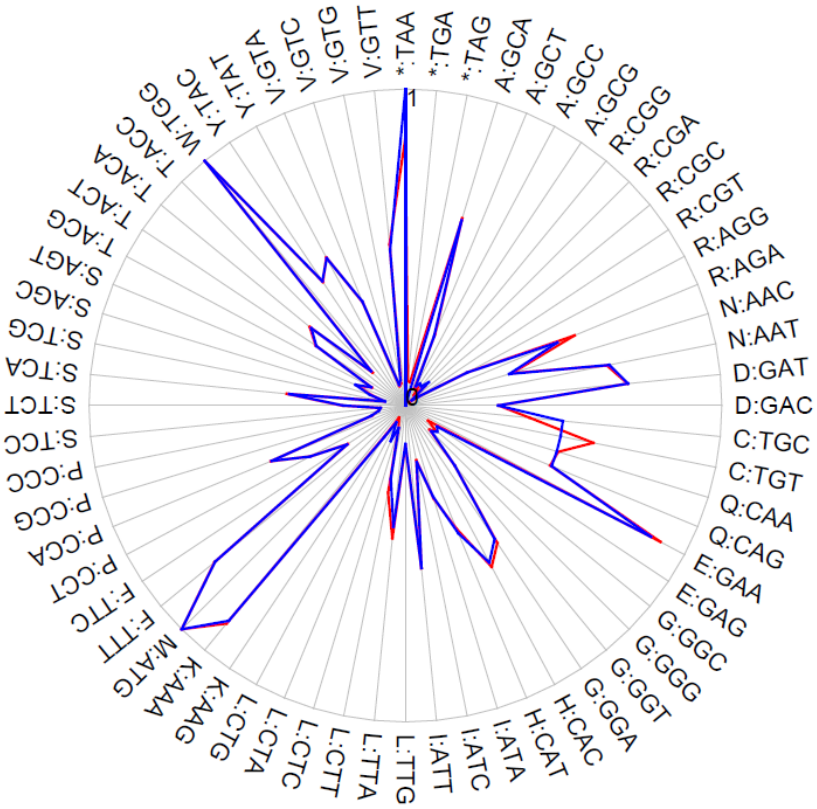

Figure 3.3: Comparison of codon usage frequencies between the optimised Cas9 gene (Mmp-Cas9) and *M. maripaludis*.
Each line radiating from the centre of the circle represents a different codon. The relative frequency at which each codon is used is depicted by a point on each line. A point touching the line at the edge of the circle indicates a frequency of 1 (i.e. always used) and a point touching the base of the line in the centre of the circle indicates a frequency of 0 (i.e. never used). Points are joined by a coloured line: blue for *M. maripaludis* and red for the codon optimised Cas9. Regions in which no red line is visible indicate the usage frequencies between the two are identical.

3.2.1.2 gRNA: PurR

The gRNA component of the test strain must be able to target the Cas9 protein to the invader plasmids. To achieve this, the 20 nt gRNA targeting sequence was to be designed to target PurR, an element present in both invader plasmids. The sequence of a validated gRNA known to be able to successfully target PurR in mouse was already available (*71*). For this validation to be a reliable indicator of functionality in the invader assay, the site it targets must also be present in the pLW40/pAW42 plasmids. However, when it was aligned to the invader plasmid PurR gene there was a single mismatch (Figure 3.4). Its location was particularly unfavourable, since it was within the PAM proximal seed region, which is known to be highly important for the gRNA/target interaction. As explained previously,

the first base in the gRNA should ideally be an adenine due to the potential of promoter bias for this base in Phmv, however the first base in the spacer used in mouse was a cytosine. Since this gRNA did not start with an adenine, and the alignment had revealed it could also not serve as a true validated target due to the presence of a mismatch, it was decided to instead design a gRNA to target PurR in silico.

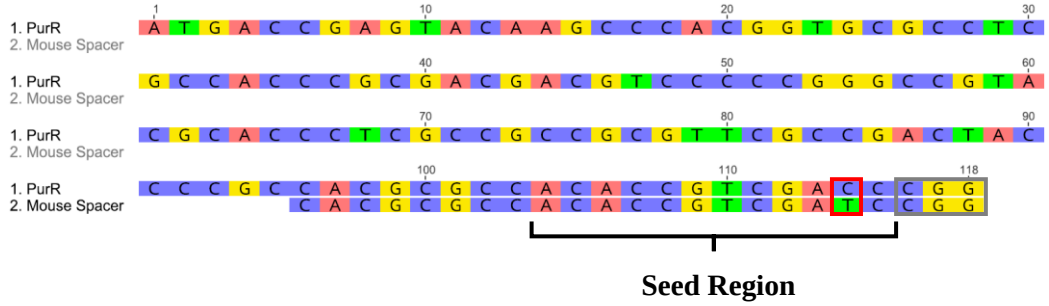

Figure 3.4: Mouse gRNA aligned to invader plasmid PurR gene.
Nucleotide alignment of the validated gRNA from mouse with the 5' portion of the PurR gene from the invader plasmids (pLW40/pAW42). The location of the PAM proximal seed region is indicated below the alignment. Red box = mismatch. Grey box = PAM. Labels to the left indicate the identity of each sequence.

Indels in the 3' portion of genes are typically tolerated more than those at more 5' locations since frameshift mutations closer to the end of the gene disrupt fewer amino acids. For the test strain to perform as intended, it must be capable of completely abolishing PurR function. It was reasoned that since the gRNA used in mouse was able to inactivate PurR, targeting Cas9 to the corresponding target site in the invader plasmid PurR gene, or any other site closer than this to the start codon should also be sufficient to knock out function. As such, the 108 bp 5' region of PurR (up to and including where the mouse gRNA had aligned) was extracted and assessed for potential gRNA target sites using the inbuilt "find CRISPR sites" tool in Geneious. 24 potential target sites were identified in this region, with estimated on-target activity scores ranging from 0.009 to 0.615 (with one being best predicted on-target activity, and zero being worst). Of these potential gRNAs, only one began with adenine (hereafter referred to as "gRNA-pur-A"). It binds in the reverse orientation between positions 66 and 88 of PurR and has a predicted on-target score of 0.251 (Table 3.1).

The second critical requirement of the gRNA sequence used for the invader assay is that, excluding its intended target site, it should not be able to bind to any other genetic feature it could potentially come into contact with: i.e. the *M. maripaludis* chromosome, the

transformation plasmids pBLPrt and pCRPrtneo, and other parts of the invader plasmids. To assess the likelihood of this, the CRISPR site finder tool was used to estimate off-target scores. The lowest off-target score for any of the 24 potential gRNAs was 98.86% (with values closer to 100% being least likely to form unwanted interactions). The score for gRNA-pur-A was 99.86%, and excluding the actual target site on pLW40/pAW42, which was identified as a perfect match, the most likely five off-target sites were predicted to be in the *M. maripaludis* plasmids. Despite this, even the "most likely" off-target site, identified in pAW42, differed to the gRNA at seven positions (Table 3.1). This analysis suggests that off-targeting is extremely unlikely for any of the 24 potential gRNAs, and was as such not a criterion that needed to be considered when deciding between them. gRNA-pur-A was therefore selected for use in the invader assay based on the fact that it was the only gRNA in the potential target region which began with an adenine.

3.2.1.3 gRNA: eGFP

Since it had not been possible to select an experimentally validated gRNA for the PurR gene, a second gRNA was designed to target pLW40/pAW42 which had been validated. There were no more features shared between the invader plasmids which also met the requirement of being absent from the pBLPrt/pCRPrtneo transformation vectors. This meant the only option would be to introduce a new target by modifying the invader plasmids. As explained previously, the eGFP gene has been widely targeted in CRISPR studies, so this was selected as the target. Eight eGFP gRNAs were downloaded from the addgene "validated gRNA" database and aligned to the 717 bp eGFP gene. The same approach was used to identify 119 potential gRNA target sites in eGFP. As expected, this included all eight of the validated eGFP sequences. Similarly to the result for PurR, on-target activity predictions varied greatly: ranging from 0.002 to 0.858, while off-target estimates were all very high (all above 96%). On-target score predictions were disregarded as a selection criterion since all eight of the gRNAs obtained from the database had been experimentally validated, which is considered to be a more reliable estimate of activity. The predicted off-target scores of these eight gRNAs were similar (ranging from 97.54-99.82%), and none of them began with an adenine. Therefore selection was based solely on information reported in the literature. The gRNA selected, hereafter referred to as "gRNA-eGFP", was chosen because unlike the other previously validated candidate gRNAs, it had been used successfully in two different organisms (see Chapter 3, Discussion). It binds the reverse strand of eGFP gene from positions 152-171, has a

predicted on-target score of 0.109, and has nine mismatches to most likely predicted off-target site (Table 3.1).

3.2.1.4 gRNA: control

As is the case for the test strains, the control strain should not be able to target Cas9 to any part of the *M. maripaludis* chromosome or the transformation vectors. However, by definition it should also not be able to target the invader plasmids. By fortunate coincidence, the GC content of both gRNA-pur-A and gRNA-eGFP was 80%, which meant that a single scrambled gRNA could serve as a suitable control for both. To produce the control gRNA, five random 20 bp sequences each with a GC content of 80% were generated. Off-target analyses revealed that all five were predicted to be similarly unlikely to bind to any of the genetic features they could potentially come into contact with (all off-target scores were above 99%). The sequence selected, hereafter referred to as "gRNA-ctrl" was chosen because it began with an adenine. gRNA-ctrl has an off-target score of 99.92%, and differs to the sequence of its most likely predicted off-target site at six positions (Table 3.1).

Table 3.1: Targeting predictions for the three selected gRNAs.
Data was generated using the find CRISPR sites tool in Geneious. PAM sequences are underlined. For the most likely off-target site, mismatches between the selected gRNA and the binding site are highlighted in red.

Name	Sequence (5' - 3')	Estimated On-Target Score	Estimated Off-Target Score	Most Likely Off-Target	
				Sequence (5' - 3')	Location
gRNA-pur-A	AGTCGGCGAA CGCGGCGGCG	0.251	99.86%	TGGCGGCCGCGG GGGCGGCGAGG	pAW42
gRNA-eGFP	GGGCACGGGC AGCTTGCCGG	0.109	99.80%	ATTGCAGGGATG CTTGCCGATGG	*M. maripaludis* chromosome
gRNA-ctrl	ACCGGGCGCA CCACCGCGTC	N/A	99.92%	ACAGTACGCACC ACCAGGACGGG	*M. maripaludis* chromosome

3.2.1.5 Phmv and Tmcr

Both the Cas9 and gRNA modules will be driven by the minimal Phmv promoter (Figure 3.5). For Cas9, the full 73 bp promoter sequence was inserted directly upstream of the start codon. To satisfy the requirement for a strict transcriptional start site for the gRNA module, a shortened version of the promoter was created. This consisted of the 5' portion of the sequence up to and including the base preceding the mapped transcriptional start site (Figure 3.5, grey box). This modified Phmv was placed directly upstream of the gRNA,

and should result in transcription beginning precisely with the first base of the gRNA. To ensure transcriptional termination, the 26 bp Tmcr was inserted at the 3' end of both modules, immediately downstream of the Cas9 stop codon and gRNA scaffold sequence.

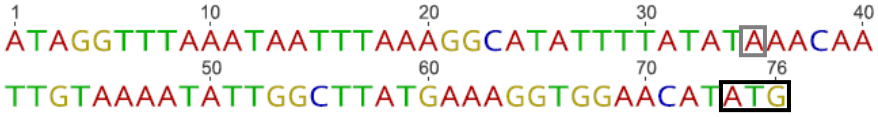

Figure 3.5: Minimal Phmv.
The minimal Phmv promoter is 73 bp in length. The position of the start codon is indicated by the black box. The transcriptional start site has been mapped to position 35 (grey box).

3.2.2 Construct Design and Assembly

With the individual components comprising the two modules required for the invader assay designed, the next step was to work out a scheme by which they could be assembled together into a vector. There were several steps to this process: first the arrangement of the two modules in the vector had to be decided. Next, this region needed to be delineated into fragments orderable as gBlocks. Finally, the gBlocks would have to be modified so that they could be assembled together using the Golden Gate system.

3.2.2.1 Module organisation

While the spacing of the individual components within each module had already been decided (see "Phmv and Tmcr"), the spacing and orientation of the two modules relative to each other had to be considered. The simplest option would have been to place one module directly in front of the other. However, it was possible that using this orientation might cause read through from the 5' module, which could negatively affect expression of the 3' module. In an attempt to avoid this, the two were orientated facing away from each other (Figure 3.6). An unwanted result of this was the creation of a 34 bp region of perfect complementarity between the two modules, caused by the shared region between the Phmv promoters. To minimise potential difficulties this might have caused in terms of secondary structure, a 226 bp "filler" sequence was inserted between the two (Figure 3.6), which was obtained from a region of the pUC18 vector backbone. The result was a 4593 bp sequence containing all the components necessary for the production of the *M. maripaludis* invader assay strains.

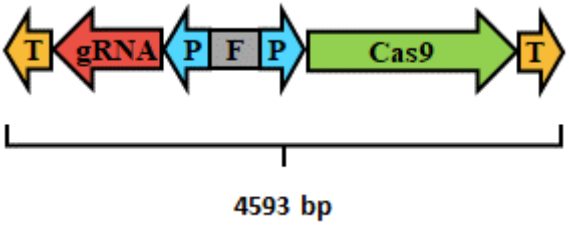

Figure 3.6: Orientation of the Cas9 and gRNA modules.
The gRNA module is orientated in reverse relative to the Cas9 module. The two are separated by a 226 bp filler sequence. T: Tmcr. P: Phmv. F: filler sequence.

3.2.2.2 Division into gBlocks

The 4593 bp unit containing the Cas9 and gRNA modules was too large to be synthesised as a single gBlock. This meant that before it could be ordered, it would have to be divided into at least two fragments. The production of each of the three different invader assay strains would require the production of three different constructs, each differing only by the 20 bp gRNA targeting sequence. This meant that three different versions of the first fragment carrying the gRNA would have to be ordered. It was decided that this fragment should be the smallest (and therefore cheapest) size available, which was 500bp. An attempt was made to include the full gRNA module, plus the filler and Phmv controlling Cas9, in Fragment 1. This would have made the construct more versatile, since the majority of the units could have been easily varied simply by ordering a new 500 bp gBlock. However, this fragment could not be synthesised by IDT due to the presence of the 34 bp region of complementarity between the two Phmv units. Instead the next best option was taken: Fragment 1 was designated as the 387 bp region comprising the full gRNA module and the filler sequence (Figure 3.7).

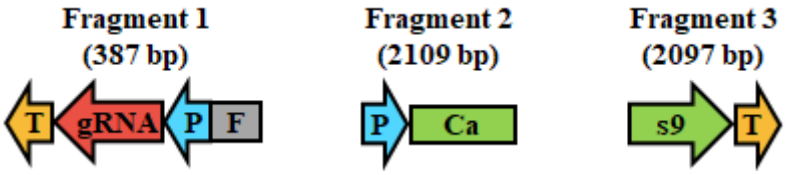

Figure 3.7: Division of the 4593bp unit into three fragments. Schematic overview of the contents of each fragment. Fragment 1 contains the full gRNA module and the filler sequence. Fragment 2 contains the first part of the Cas9 module (Phmv and the 5' portion of Cas9). Fragment 3 contains the last part of the Cas9 module (3' portion of Cas9 and Tmcr). The size of each module is indicated in brackets. T: Tmcr. P: Phmv. F: filler sequence.

The remaining region of the construct, comprising the full 4206 bp Cas9 module, needed to be separated into two fragments (fragments 2 and 3). The only constraint governing the selection of the boundary between these fragments was that once split, both had to abide by gBlocks synthesis constraints. Boundaries approximately in the middle of the 4206 region were arbitrarily selected and the resulting fragments tested for complexity in the IDT gBlocks tool. A boundary yielding fragments that could be synthesised was selected, which resulted in a 2109 bp Fragment 2, and a 2097 bp Fragment 3 (Figure 3.7).

3.2.2.3 Fragment assembly

To allow the simple assembly of the three fragments into a construct, a scheme was designed such that they could be combined in the correct order using the Golden Gate system. There were two stages to this approach. First, each gBlock would be inserted into a holding vector through use of restriction cloning (Figure 3.8, Step 2). In the second, a Golden Gate reaction would be used to assemble the fragments in the correct order in a destination vector (Figure 3.8, Steps 3-4). The Golden Gate vector DVA_BC (*84*) (hereafter pDVA_BC), which contains a dropout region flanked by BpiI sites, was selected as the destination vector. For the holding and destination vectors to be Golden Gate compatible, both must have different antibiotic resistances. Since pDVA_BC contains an ampicillin resistance gene (*84*), the kanamycin resistant pHSG298 plasmid was selected as the holding vector.

As the Golden Gate reaction was to be conducted using BpiI, existing BpiI sites already present within the fragments had to first be removed. Four sites were found, all of which were located within the codon optimised Cas9 gene. For each site, a single base change

was introduced at the design stage to remove it, while maintaining the identity of the amino acid designated by codon that base was present in (Table 3.2).

Table 3.2: Sites modified in the Cas9 gene to remove pre-existing BpiI sites.

Site	Position	Amino	Codon		Host Relative Freq.	
		Acid	Original	Modified	Original	Modified
1	1755	Asp	GAC	GAT	0.29	0.71
2	1842	Asp	GAC	GAT	0.29	0.71
3	2256	Gly	GGA	GGT	0.52	0.26
4	3753	Asp	GAC	GAT	0.29	0.71

To make fragments 1-3 compatible with pDVA_BC, BpiI sites which give rise to 4bp overlaps corresponding to those flanking the pDVA_BC dropout region were appended to the 5' end of Fragment 1 (Bpi site 1), and the 3' end of Fragment 3 (Bpi site 4). Next, the fragments were made compatible with each other by adding BpiI sites which generate matching overlaps to the 3' end of Fragment 1 and 5' end of Fragment 2 (Bpi site 2), and the 3' end of Fragment 2 and 5' end of Fragment 3 (BpiI site 3) (Figure 3.8, Stage 1). The overlaps for BpiI sites 2 and 3 were designed arbitrarily, however it was ensured that they were not palindromic and differed from all other overlaps in the scheme by at least one base. Next, to allow the initial insertion of each of the fragments into the holding vector pHSG298, BamH1 and SacI sites (which are present in the pHSG298 MCS) were appended to the ends of each fragment (Figure 3.8, Step 1). Finally, the full scheme as described above was simulated in two different software packages (Geneious (*56*) and Benchling (www.Benchling.com)) to check for any inconsistencies. Both programs output pDVA_BC containing the three fragments fused together as intended, so the corresponding gBlocks were ordered.

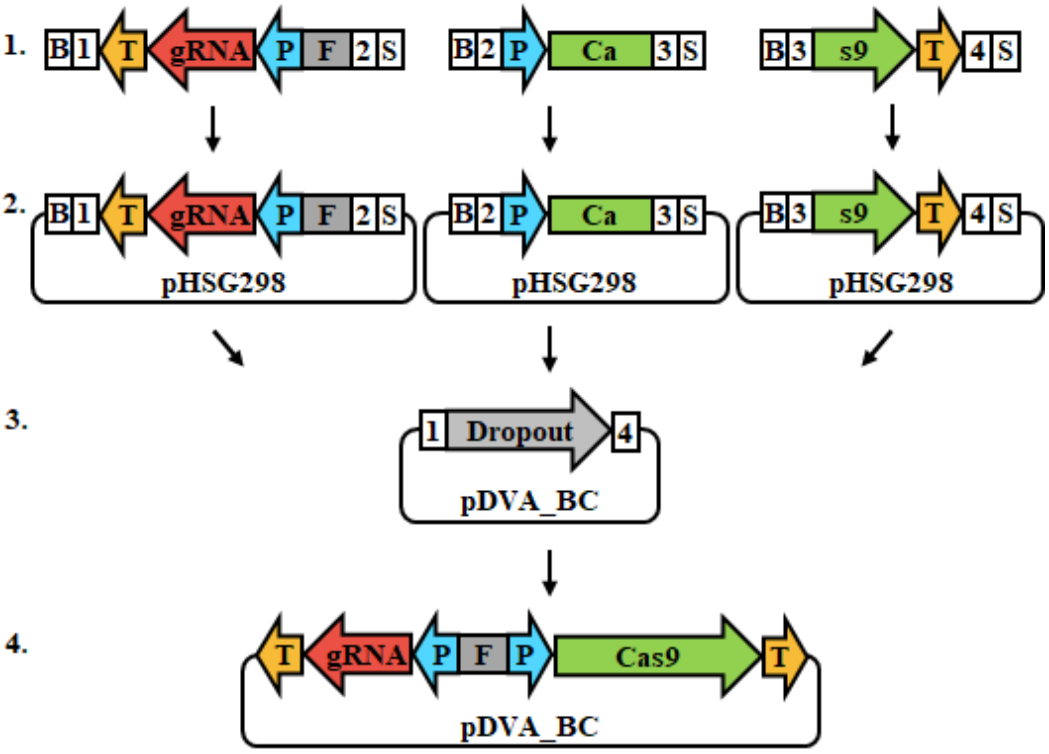

Figure 3.8: Fragment assembly scheme.
1: Restriction enzyme recognition sites are added to the ends of each of the fragments to make them compatible with the assembly scheme. First, BpiI sites (see below) are added to enable one step assembly of the fragments into the destination vector pDVA_BC in the correct order. External to these, BamHI/SacI sites are added to the 5' and 3' end of each fragment respectively to enable the initial insertion of each into the holding vector pHSG298. **2:** Restriction cloning using BamHI/SacI is used to insert each fragment into pHSG298. **3:** A Golden Gate reaction is conducted using each of the holding vectors containing the three different fragments and the destination vector pDVA_BC. **4:** Complementarity between the BpiI overlaps results in the incorporation of the three fragments into pDVA_BC in the correct order. White boxes: restriction sites. B = BamHI. S = SacI. Numbered white boxes: BpiI recognition sites. Numbers indicate the sequence of the overlap corresponding to each: 1 = TACT, 2 = ATAG, 3 = TCTA, 4 = AATG. T: Tmcr. P: Phmv. F: filler

3.3 Discussion

The objective of this chapter was to design the expression modules which when transformed into *M. maripaludis* would give rise to the test and control strains required for the invader assay. This was achieved in two stages. In the first the individual components of the expression modules were designed, and in the second a system by which they could be assembled in the desired order in a cloning vector was planned. However, issues encountered during this design process meant that some alterations to the plan as detailed in the introduction had to be made.

3.3.1 Codon Optimisation of SpCas9

The codon optimised Cas9 sequence initially generated with the Jcat tool was too repetitive to be synthesised by IDT. This can be explained by the fact that this program is based on the CAI method, and simply replaces all codons with their most commonly used alternative in a subset of host genes (*52, 65*). This means that in all cases only one of the possible codons for each amino acid was used: a scenario which by definition would result in an unavoidably repetitive output (*62, 89*). Optimisations generated based on ICU in the COOL program were comparatively less repetitive, presumably because ICU assigns codons based on their relative frequency (*53, 64*), and can therefore make use of all codons. However, while these sequences were orderable as synthetic fragments, they still appeared to be more repetitive than other previously validated Cas9 sequences (Figure A1, Appendix).

Repetitive sequences in genes can result in reduced protein expression (*90, 91*), and in severe cases genetic instability which may result in unwanted excision through recombination (*89*), both of which would clearly have been undesirable for this application. However, attempts to limit the repetitiveness of the outputs by modifying the optimisation parameters in COOL appeared relatively ineffective. This could be explained by the AT rich nature of the *M. maripaludis* genome (*14*). Optimisations to this organism would be expected to favour codons with a high AT content, while making less use of GC rich alternatives, resulting in a less diverse and therefore more repetitive sequence. Additionally, these efforts also had a negative effect on the quality of the optimisation. Since sub-optimal codon content can also hamper expression (*92*), there was a trade-off between the extent of repetitiveness and the fit of the optimisation to the host codon usage bias. Ideally it would have been possible to test a rage of optimised Cas9 sequences prioritising these conflicting factors to different extents. However this would have required

the purchase of additional large gBlocks, and the production of more test strains, and was therefore beyond the scope of this study due to both financial and time constraints. For the Cas9 codon optimisation selected for the invader assay: Mmp-Cas9 (Figure A2, Appendix), the decision was therefore taken to prioritise optimisation quality.

Mmp-Cas9 was produced using codon usage frequencies calculated from protein coding gene entries in Genbank (*54*), which are based primarily on gene predictions (*14*). It is unlikely that this precisely reflects the true codon bias of *M. maripaludis* since some genes may have been predicted incorrectly/not be expressed. Additionally, codon usage frequencies averaged across all genes are likely to differ to some extent to those calculated using only highly expressed genes (*61*). A more robust approach may therefore have been to optimise Cas9 based on codon usage frequencies calculated solely from experimentally validated, highly expressed genes. Despite this, the codon bias data used to generate Mmp-Cas9 is likely to have been a reasonable approximation for this organism. Perhaps most importantly, Mmp-Cas9 does not make excessive use of any codons predicted to be rarely used in the host (Figure 3.3), which is thought to be one of the primary factors responsible for poor expression of foreign genes in different organisms (*61, 62, 91*). It should be noted that there are many examples of Cas9 genes having been used successfully in CRISPR/Cas9 studies in organisms to which they were not initially optimised (*93, 94*). Perhaps most significantly, Nayak and Metcalf (2017) were able to demonstrate Cas9 functionality in a cell destruction based experiment similar to the invader assay using the native *S. pyogenes* Cas9 in the archaeon *M. acetivorans (48)*. Since Mmp-Cas9 has been specifically codon optimised to its host, it should theoretically be expressed at higher levels, and therefore perform comparatively better than the Cas9 used in this example.

3.3.2 gRNAs

3.3.2.1 Off-targeting

Off-target binding has been described as one of the main weaknesses of the CRISPR/Cas9 system (*95-97*), and could compromise the invader assay (see Chapter 3, Introduction). The precise rules which dictate CRISPR/Cas9 target site specificity are still not fully understood, and a multitude of different techniques, some inherently biased, have been used to assess off-target activity in a variety of different systems (*98*). With this in mind, it is perhaps unsurprising that estimates as to the extent to which a gRNA must differ from other potential target sites for it to be considered specific vary. In general however, potential binding sites with higher numbers of mismatches are thought to be less likely to result in

off-target effects (*99*)**.** Positioning is also important: mismatches in the PAM proximal seed region (i.e. the 3' most 12 bases of the gRNA targeting sequence) generally appear to be less well tolerated (*100*). On these bases, the gRNAs designed for use in this study seem unlikely to exhibit off-target effects, since none have fewer than six mismatches to their respective most likely predicted off-target site, including in all cases at least two mismatches to the PAM proximal seed region (Table 3.1).

Additionally, the software used during the design process gave all three gRNAs highly favourable off-target scores (above 99%). In fact, none of the 24 potential target sites for PurR and 119 for eGFP were predicted to have an off-target score below 96 %, suggesting that (at least according to the software) they were all likely to have been highly specific. Since in silico based off-targeting predictions can be unreliable (*69*), it is possible that these values were overly favourable due to incorrect prediction. However, gRNAs targeting these genes might in reality be expected to be highly specific. Firstly, the target regions considered were within genes of non-host origin, and therefore had comparatively higher GC content than the *M. maripaludis* genome, which comprises the vast majority of the sequence to which these gRNAs could potentially bind. Secondly, at just 1.66 Mb in length (*14*), the *M. maripaludis* genome is relatively small compared with organisms for which off-targeting has been widely reported, for example in the 3000 Mb human genome (*101*). Clearly, the odds of a region highly similar to a given target site being present in a genome is proportional to its size. In agreement with this, off-targeting is also thought to be less likely in bacterial compared with eukaryotic systems for the same reason (*102*).

3.3.2.2 On-targeting

The intention had been to use a pre-validated gRNA from mouse (*71*) to target PurR to the invader plasmids. However, due to sequence differences between the two PurR genes (Figure 3.4) the decision was taken to instead design this gRNA in silico. Very low or no on-target activity by gRNA-Pur-A could be particularly problematic if *M. maripaludis* possesses some kind of non-homologous end joining (NHEJ) pathway. This is because such a repair mechanism could permit re-circularisation of cleaved invader plasmids, allowing regeneration of the resistance phenotype even if Cas9 was active. gRNA-Pur-A would be critical in this scenario because Cas9 mediated cleavage within the puromycin resistance gene itself would be resistant to such regeneration. This is because NHEJ often generates indels at the repair site (*93*), which would lead to inactivating mutations and frame shifts within PurR even following re-circularisation. gRNAs targeted elsewhere on the invader

plasmid backbone would lack this in-built NHEJ resistance, which makes the fact that gRNA-Pur-A could not be based on an experimentally validated gRNA undesirable since in silico methods alone provide a less reliable estimate of functionality (*70*).

Fortunately, homologous recombination (HR) appears to be the primary mechanism for DSB repair in archaea (*103*), and while NHEJ has been reported in this domain (*104*), its presence in *M. maripaludis* seems unlikely based on currently available evidence. No gene homologous to Ku, a core component of the NHEJ machinery (*104*) could be found in the sequenced *M. maripaludis* genome (*105*). Additionally, the polyploid nature of *M. maripaludis* (*16*) suggests NHEJ is unlikely to have been required during its evolution, since repair templates for the preferred HR repair mechanism would be readily available, as is the case in rapidly dividing bacterial species such as *E. coli*, which are known to lack NHEJ (*106*). With this in mind, and to provide a backup in case gRNA-Pur-A does not exhibit sufficient on-target activity, an additional highly validated gRNA which targets the eGFP gene was selected.

Of the eight previously validated gRNAs considered (*72-75*), the one chosen as gRNA-eGFP was selected because unlike the others it has been previously validated in two different organisms: human (referred to as "EGFP Site 1" in (*73*)) and zebrafish (referred to as "*egfp* gRNA" in (*74*)). High activity for this gRNA was reported in both organisms: > 80% in Zebrafish (*74*), and ≈70-90% in human cells (*73*). Clearly, high on-target activity of a gRNA in one organism does not necessarily guarantee the same in another. However, it was anticipated that a gRNA which had already been shown to be highly active in two different organisms would be more universal and therefore more likely to also have high on-target activity in *M. maripaludis*. The authors also reported that transfecting the plasmid harbouring gRNA-eGFP at 20-fold lower concentrations, or in combination with 15-fold lower concentrations of the plasmid carrying the Cas9 gene did not result in a substantial loss of activity (*73*). Assuming this gRNA also exhibits this characteristic in *M. maripaludis*, it would be highly desirable for the invader assay as it would provide some buffering in the event that the either of these components is not expressed as highly as intended. It should be noted however that including this additional gRNA in the study will require the production of an extra *M. maripaludis* test strain, as well as the modification of both invader plasmids to include the eGFP target site. This will necessitate more cloning steps, which will be more time consuming.

3.3.3 Chapter Summary/Next Steps

In summary, while some concessions have been made, the aims of this chapter have still been met: a scheme which should be sufficient to produce the test strains required to carry out the invader assay has been produced. The next steps will be to:

1. Assemble the designed fragments via Golden Gate into pDVA_BC.
2. Subclone the complete assemblies into the *M. maripaludis* transformation vectors pBLPrt/pCRPrtneo.
3. Transform *M. maripaludis* with pBLPrt/pCRPrtneo to produce the test strains.
4. Modify the invader plasmids pLW40/pAW42 to contain the eGFP target site.
5. Carry out the invader assay

Chapter 4: Molecular Work

4.1 Introduction

4.1.1 Chapter Aims

In the previous chapter a scheme had been designed by which a set of constructs capable of producing the test and control strains required to carry out the invader assay could be introduced into the *E. coli* cloning vector pDVA_BC, via the Golden Gate system. The aim of this section is to use this design to produce all of the required plasmids and strains such that the invader assay can be carried out. This can be split into several sub-aims. First, the CRISPR/Cas9 constructs must be assembled into vectors in *E. coli*. Next, the CRISPR/Cas9 assemblies need to be transferred into the *M. maripaludis* transformation vectors. Finally, these vectors will be used to transform *M. maripaludis*, which should result in the chromosomal integration of the CRISPR/Cas9 assemblies: producing the invader assay test and control strains. An additional gRNA targeting the eGFP gene had also been designed in Chapter 3. Since the invader plasmids pLW40/pAW42 do not contain the corresponding target site, an additional aim of this chapter will be to introduce this into both of these plasmids, making them compatible with the scheme designed.

4.1.2 Delivering CRISPR/Cas9 Assemblies to *M. maripaludis* Transformation Vectors

Following the assembly of the CRISPR/Cas9 modules into the *E. coli* cloning vector, they must first be transferred into the *M. maripaludis* transformation vectors pBLPrt/pCRPrtneo, before they can be delivered to the *M. maripaludis* chromosome. Material is introduced into the transformation vectors in between 5' and 3' flanking regions to the selected chromosomal target site, resulting in the incorporation of this material at the corresponding locus in *M. maripaludis* via HR, following transformation. The current method for delivering material ultimately intended for incorporation into the *M. maripaludis* chromosome into these transformation vectors is standard restriction enzyme cloning. In the case of pBLPrt this must be done using a single AscI site which is located in between 500 bp flanking regions to the UPT locus, such that on transformation material introduced into this site is delivered at the UPT locus on the *M. maripaludis* chromosome. pBLPrt is suitable for transformation of *M. maripaludis* strain Mm900, which has a wild-type UPT locus (*19*). However, this vector is not suitable for delivery into S0001, since the UPT site in this strain already contains the ORF1 product, which is responsible for the increased transformation efficiency phenotype of this strain (*12*). Integration of the

CRISPR/Cas9 modules here using pBLPrt would result in its replacement, which would effectively defeat the object of using it. This means the CRISPR/Cas9 assembly must be introduced elsewhere on the chromosome in S0001, which can be achieved with the pCRPrtneo transformation vector. pCRPrtneo is structurally similar to pBLPrt, except it does not contain UPT flanks. This vector can be made compatible with the introduction of material at any chromosomal locus in *M. maripaludis* by transferring said material onto its backbone in between approximately 500 bp of 5' and 3' flanking sequence corresponding to the desired target site (*19*). This effectively creates a vector functionally identical to pBLPrt, except with a different target locus. This target locus must therefore first be selected.

4.1.3 Selection of Chromosomal Target Site for *M. maripaludis* S0001 Strain

The primary criterion for selecting a target site for the delivery of the CRISPR/Cas9 assembly into S0001 is that the introduction should not result in any phenotypic effect which could be disruptive to the invader assay: as an extreme example, causing lethality. Such a scenario could best be avoided by selecting a target site outside of functional genomic DNA elements, such as genes. While the locations of many of these elements within the *M. maripaludis* genome have been predicted (*14*), it is likely that some have been missed. Assuming no delays, the *M. maripaludis* transformation process, including the required selection and screening steps, takes several weeks (*107*). Selection of an unsuitable site would therefore be extremely undesirable given the project timeframe. For this reason it would be advantageous to select a pre-validated target site (i.e. one that had been used successfully before).

One such site is the *M. maripaludis* MCM-C locus. Replacement of the MCM-C gene with a short read-through product resulted in a slight increase in *M. maripaludis* growth rate, but no other phenotypic effect was detected (*107*). While use of a locus resulting in any phenotypic effect (particularly one relating to growth) might initially seem undesirable in the case of the invader assay, this could actually be useful. The integration of the three alternative CRISPR/Cas9 assemblies at the MCM-C locus, which would give rise to the two *M. maripaludis* test strains targeting the invader plasmids, and the control which does not, would happen at the same site. This means that all three (critically including the control strain) would be subject to the same increased growth phenotype, allowing comparisons between them to remain valid. Assuming the proportion of cells in these test and control strains subsequently transformed with the invader plasmids remained the same, an

increased growth phenotype would actually be expected to add power to the assay, since a greater number of cells would presumably be recovered, which (if the CRISPR/Cas9 system was functional) would correspond to a greater difference in the number of colonies recovered for test vs control strains: i.e. the desired output of the assay.

The plasmid previously used to achieve the integration at the MCM-C locus, pAW10, contains a short read-through product in between 500 bp of 5' and 3' flanking sequence corresponding to the MCM-C locus, cloned into a single NotI site on the pCRPrtneo backbone (*107*). pAW10 could theoretically be repurposed for delivery of the CRISPR/Cas9 assemblies into the MCM-C locus in S0001 by replacing the read-through product with these assemblies. However, the process of transferring the Golden Gate assembled CRISPR/Cas9 modules from the pDVA_BC plasmid into the *M. maripaludis* transformation vectors using standard restriction enzyme cloning is likely to be challenging. Firstly, at ≈ 4.6 kb in length, the fully assembled CRISPR/Cas9 product is relatively large, and can only be introduced (at least in the case of pBLPrt) using a single restriction enzyme, which requires an additional vector dephosphorylation step to prevent recircularisation (*108*). Additionally, neither pBLPrt nor pCRPrtneo support blue/white screening, which would make the process of identifying correctly integrated products more complicated. For these reasons, it would be desirable to use a more efficient cloning technique to produce the required *M. maripaludis* transformation vectors.

4.1.4 Golden Gate Compatible *M. maripaludis* Transformation Vectors

Both pCRPrtneo and pBLPrt contain an ampicillin and kanamycin resistance gene (*19*), and according to the available sequences (James Chong, personal communication) do not contain any BpiI recognition sites. This means that both could theoretically be modified to act as Golden Gate destination vectors in the place of pDVA_BC, while remaining compatible with the assembly scheme designed in Chapter 3, as they can be selected with a different antibiotic to the pHSG298 holding vectors, and will not be cleaved ectopically by BpiI (*81*). Achieving the required modifications would involve transplanting the pDVA_BC dropout region, which contains a LacZ gene and the BpiI sites (*84*) to which the assembly scheme has been designed, into both. This would allow Fragments 1-3 to be assembled directly into both via the Golden Gate system, which would be expected to streamline the cloning process for several reasons. Firstly, the pDVA_BC dropout region is relatively short, and would only need to be inserted into each of the transformation vectors once. The Golden Gate system, which is highly efficient (*79*), could then be used

to assemble the three different versions of the CRISPR/Cas9 modules (each carrying an alternative gRNA) into both modified transformation vectors, which due to the introduction of the LacZ gene would additionally support blue/white screening. This would effectively reduce the number of times standard restriction cloning techniques (without the facility of blue/white screening) would need to be used from six to two, and should therefore be quicker. Secondly, since the intermediate assembly step in pDVA_BC would be eliminated, there would be no requirement for PCR to add appropriate restriction sites to the CRISPR/Cas9 assemblies to enable subcloning into the transformation vectors. This direct assembly would therefore remove the possibility of PCR induced mutations (*83, 85*), which should simplify screening of the resulting assemblies. To make pBLPrt Golden Gate compatible, the pDVA_BC dropout region will have to be inserted at the AscI site in between the existing UPT flanks. For pCRPrtneo, this could be achieved by inserting the dropout region into pAW10 in between the MCM-C flanks to replace the existing read-through product.

4.1.5 Modifying pLW40/pAW42

Since a suitable pre-validated gRNA targeting the puromycin resistance gene in the invader plasmids had not been described, a highly validated alternative targeting the eGFP gene had also been included in this study (see Chapter 3). The invader plasmids pLW40/pAW42 do not contain the eGFP gene (*12, 24*), so the target site will have to first be introduced into both to enable them to be targeted by *M. maripaludis* test strains carrying this gRNA. Theoretically, introducing the 20 bp sequence to which gRNA-eGFP is designed to bind, plus a PAM sequence into both plasmids should be sufficient to achieve this. To preserve plasmid function, it is critical that this target region be introduced into a location in both which does not disrupt functional elements, such as PurR. Finally, while not essential, it would be desirable to introduce this modification in the same relative location and orientation in both plasmids, as this would make results obtained from each more comparable, while also simplifying the cloning process.

4.2 Results

4.2.1 Production of Golden Gate Compatible Integrative Vectors pBLPrt/pCRPrtneo

To streamline the cloning process, the pBLPrt/pCRPrtneo vectors were to be modified such that they were compatible with direct assembly of Fragments 1-3 via the Golden Gate system. Since the fragments had been designed for assembly into the pDVA_BC vector, these modifications would be achieved by incorporating its dropout region (containing LacZ and BpiI sites 1 and 4 (Figure 4.2, Panel A)) into both pBLPrt and pCRPrtneo. To ensure the resulting assembled Cas9/gRNA constructs would be introduced at the correct genomic location following *M. maripaludis* transformation, the dropout region had to be inserted in between the 5' and 3' flanking regions corresponding to the genomic target site in each integrative vector. In the case of pBLPrt this would involve insertion of the dropout region into the AscI site located in between existing flanking regions to the UPT locus (Figure 4.2, Panel A). However, for pCRPrtneo (Figure 4.2, Panel A) appropriate flanking regions (to the MCM-C locus – see Chapter 4, Introduction) would have to be introduced along with the dropout region.

4.2.1.1 Modification via PCR

The possibility of using a PCR based approach to introduce the modification described above was first investigated. This would have involved PCR amplification of the pDVA_BC dropout region with primers designed to add restriction sites appropriate for insertion into each integrative vector. The products would then have been cloned into the AscI site in pBLPrt, and a pair of restriction sites in pAW10 (a derivative of pCRPrtneo containing the MCM-C 5' and 3' flanks) resulting in replacement of the read through product with the dropout region. While this strategy would have been suitable for pBLPrt, sequence analysis revealed this would not have been possible in pAW10 because the MCM-C 5' flank was found to already contain two BpiI sites (Figure 4.1), which would have impeded correct Golden Gate assembly. While it would theoretically have been possible to address this problem using PCR based methods to modify pAW10, it was decided that this would be too time consuming, so a different a different approach was taken.

Figure 4.1: Structure of the MCM-C homology region in pAW10.
The location of the pre-existing BpiI sites are indicated in the MCM-C 5' flanking region. Relevant features are labelled. Base numbers are indicated above the figure.

4.2.1.2 Modification via gBlock

A gBlock was designed which could be used to concurrently insert the pDVA_BC dropout region into pBLPrt, and the dropout region flanked by MCM-C homology arms in pCRPrtneo. First, the 550 bp pDVA_BC dropout sequence containing the required features (LacZ cassette and BpiI sites 1 and 4) was extracted. AscI sites were appended to each end (Figure 4.2, Panel B, Stage 1) such that subsequent restriction cloning using this enzyme could be used to transfer the dropout region into pBLPrt between the UPT flanks (Figure 4.2, Panel B, Stage 3). Next, the 500 bp MCM-C flanking sequences previously used in pAW10 were added: with the 5' flank placed 5' to the dropout region. NotI sites were then added to each end (Figure 4.2, Panel B, Stage 1) to allow the dropout region and MCM-C flanks to be transferred together into pCRPrtneo in the same location as had been used in pAW10 (Figure 4.2, Panel B, Stage 3). As with Fragments 1-3 designed in Chapter 3, BamHI/SacI sites were added to the extreme ends of the gBlock to enable initial insertion into an *E. coli* cloning vector: in this case pUC18 (Figure 4.2, Panel B, Stage 2). Before the gBlock was ordered, several pre-existing restriction enzyme recognition sites had to be removed to make it compatible with the cloning strategy. A SacI and BamHI site present within the LacZ gene of the pDVA_BC dropout region were removed by altering a single base or group of bases to use the next most frequently used codon for the amino acid in question, in this case using codon usage data from the *E. coli* K12 strain, which DH5α is derived from (*109*) (Table 4.1). Since the MCM-C 5' flank is non-coding, a single base within each of the two BpiI sites was arbitrarily changed to remove them. The same method was used to remove an unwanted NotI site found to be present within a non-coding portion of the gBlock.

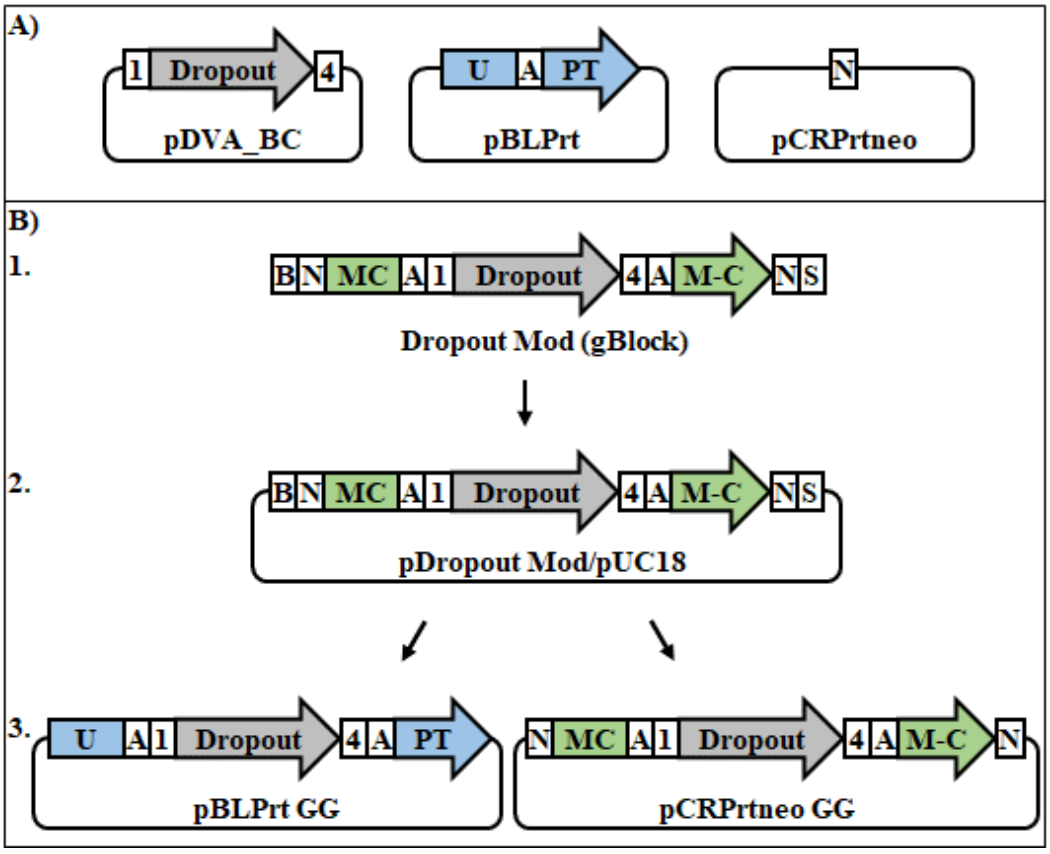

Figure 4.2: Producing Golden Gate compatible pBLPrt/pCRPrtneo vectors scheme. **A)** Relevant features of the vectors used in this scheme. pDVA_BC: Golden Gate vector containing a dropout region carrying a LacZ gene in between BpiI sites 1 and 4. Golden Gate assembly into these sites replaces the Dropout region allowing blue/white screening. pBLPrt: *M. maripaludis* transformation vector containing an AscI site in between flanking regions to UPT. Material introduced into the AscI site is incorporated into the genomic UPT locus in *M. maripaludis* via HR on transformation. pCRPrtneo: *M. maripaludis* transformation vector. Material to be introduced can be inserted into the NotI site flanked by homology regions to the desired genomic target site. **B)** Vector modification scheme. **1:** A gBlock is constructed containing the pDVA_BC dropout region and BpiI sites 1+4. AscI sites are added to each end to enable subcloning of the enclosed region into pBLPrt. NotI sites external to the MCM-C flanking regions as used in pAW10 are then added to the ends of this to allow subcloning into pCRPrtneo. BamHI and SacI sites are added to the ends of the gBlock to enable initial insertion into the cloning vector pUC18. **2:** Restriction cloning with BamHI/SacI is used to insert the gBlock into pUC18. **3:** The AscI flanked insert in pDropout Mod/pUC18 is subcloned into pBLPrt, and the NotI flanked insert into pCRPrtneo. The result is a pair of modified Golden Gate compatible *M. maripaludis* transformation vectors. The presence of the Dropout region allows blue/white screening on assembly using BpiI sites 1 and 4. The position of these sites in between the UPT and MCM-C flanking regions means assembled material can be transferred directly into these loci upon *M. maripaludis* transformation. White boxes: restriction sites. B = BamHI. S = SacI. N = NotI. A = AscI. Numbered white boxes: BpiI recognition sites. U = UPT 5' flank. PT = UPT 3' flank. MC = MCM-C 5' flank. M-C = MCM-C 3' flank.

Table 4.1: Sites modified in pDVA_BC LacZ gene to remove pre-existing BamHI/SacI sites.
Site 1 was formerly a BamHI recognition site, and site 2 SacI. Position refers to the relative position in the LacZ gene of the nucleotide(s) modified to achieve the codon change. "Original" and "modified" refer to the status of the codon before and after modification. Host Relative Freq. is the frequency at which the specified codon is used proportional to all other codons for that amino acid in the *E. coli* K12 strain.

Site	Position	Amino	Codon		Host Relative Freq.	
		Acid	Original	Modified	Original	Modified
1	48	Glu	GAG	GAA	0.3	0.7
2	67-69	Ser	TCG	AGC	0.16	0.33

4.2.1.3 Subcloning Dropout Mod gBlock into pUC18

The first step to producing the Golden Gate compatible pBLPrt/pCRPrtneo vectors was to subclone the Dropout Mod gBlock into pUC18 (Figure 4.3, Panel A). To generate sticky ends compatible with the insert, pUC18 was digested with BamHI/SacI, which yielded a band at the expected size (Figure 4.3, Panel B, Lane 4). As the excised fragment was estimated to be only 19 bp in length, and therefore too small to be visible on a gel, single digestions with each enzyme were carried out to check for complete digestion. Both reactions resulted in a single band at the same size as the double digestion (Figure 4.3, Panel B, Lanes 2-3), indicating complete digestion with both enzymes. The double digested pUC18 product was gel extracted. Next, the gBlock was digested with the same two enzymes to allow ligation into the vector, which resulted in a product at the expected size (Figure 4.3, Panel C, Lane 3). This was indistinguishable in size from the undigested product (Figure 4.3, Panel C, Lane 2), which was as expected since it was predicted to be just 32 bp shorter. The remainder of the gBlock double digestion reaction was PCR purified, ligated with the gel extracted pUC18 product and transformed into *E. coli*. A control reaction containing digested pUC18 and ligase (+ ligase) was also transformed.

12 colonies were recovered on the test plates (i.e. plates containing *E. coli* transformed with ligated vector and insert) following transformation: all were blue as expected since the dropout region of the gBlock contains a LacZ gene. There were a large number of smaller white satellite colonies. Only a single blue colony was present on the control plates which suggested that the majority of the 12 test plate colonies contained an insert. A colony PCR using primers pUC18 CP F1/R1, designed to bind to the pUC18 backbone (Figure 4.3, Panel A) and amplify a 2.6 kb product if the correct insert was present was carried out. All

eight tested colonies produced the same banding pattern (Figure 4.3, Panel D, Lane 2), which unexpectedly contained multiple bands, none of which were at the expected size. In an attempt to identify the origins of these, the PCR was repeated using material from a series of control colonies, as well as to the four remaining blue test colonies (Figure 4.3, Panel E). The product generated from an unmodified pUC18 containing colony was slightly below the expected size of 1 kb (Figure 4.3, Panel E, Lane 5): however the large amount of material present likely caused this band to migrate further on the gel, and as such it was considered to be at the expected size. The same band was present in the pUC18 colony from the + ligase control plate (Figure 4.3, Panel E, Lane 4), which suggested that the unexpected bands in the initial colony PCR (Figure 4.3, Panel D, Lane 2) were not caused by modification to the vector itself. No product was obtained using satellite colony material (Figure 4.3, Panel E, Lane 3), which confirmed that satellite contamination was not the cause of the unexpected bands. The four remaining blue colonies all contained a single band at 1 kb (Figure 4.3, Panel E, Lane 2), which would be expected if they did not contain the insert. However, it seems unlikely that of the 12 colonies tested in total, all eight in the first reaction would have a shared but distinct band from those in the second. A more likely explanation is that stochastic differences between the two reactions may have affected primer binding stringency. Taken together these results suggest that the pUC18 CP F1/R1 primers were able to bind ectopically to some region of the insert, and were therefore unsuitable.

Next a colony PCR using primers pUC18 CP R1 and Dropout Seq F3 (Figure 4.3, Panel A) was carried out on colonies representative of each of the two distinct banding patterns previously observed for test plate colonies. Since Dropout Seq F3 was designed to bind within the MCM-C flanking sequence (which is highly dissimilar to the pUC18 backbone), this PCR was expected to be more specific. All five colonies tested produced a band just under the expected size of 1.1 kb (Figure 4.3, Panel F, Lane 2), suggesting that all contained the correct insert. Unexpectedly, the pUC18 control lane contained a ladder-like banding pattern (Figure 4.3, Panel F, Lane 3). Since there was no product in the water control (Figure 4.3, Panel F, Lane 4), the most likely explanation was contamination of the pUC18 product with ladder during gel loading.

Diagnostic digests were carried out on plasmid DNA isolated from two of these colonies (colonies 1 and 9), both of which produced the same banding pattern as each other (colony 1 products are shown). As expected, vector linearization with BamHI resulted in a single

band at a size corresponding to the sum of the two components: pUC18 and Dropout Mod (Figure 4.3, Panel G, lane 3). Double digestion with BamHI/SacI to liberate the insert from the vector backbone resulted in two smaller bands also at the expected sizes, as well as an additional band and the same size as the BamHI digested product (Figure 4.3, Panel G, Lane 2), indicating incomplete double digestion. Taken together these results revealed that the vector contained an insert of the correct size. Finally, the full inserts in colonies 1 and 9 were sequenced and confirmed to be correct for their full length. Colony 1 and its plasmid cargo (hereafter referred to as pDropout Mod/pUC18) was arbitrarily selected for further use in this study.

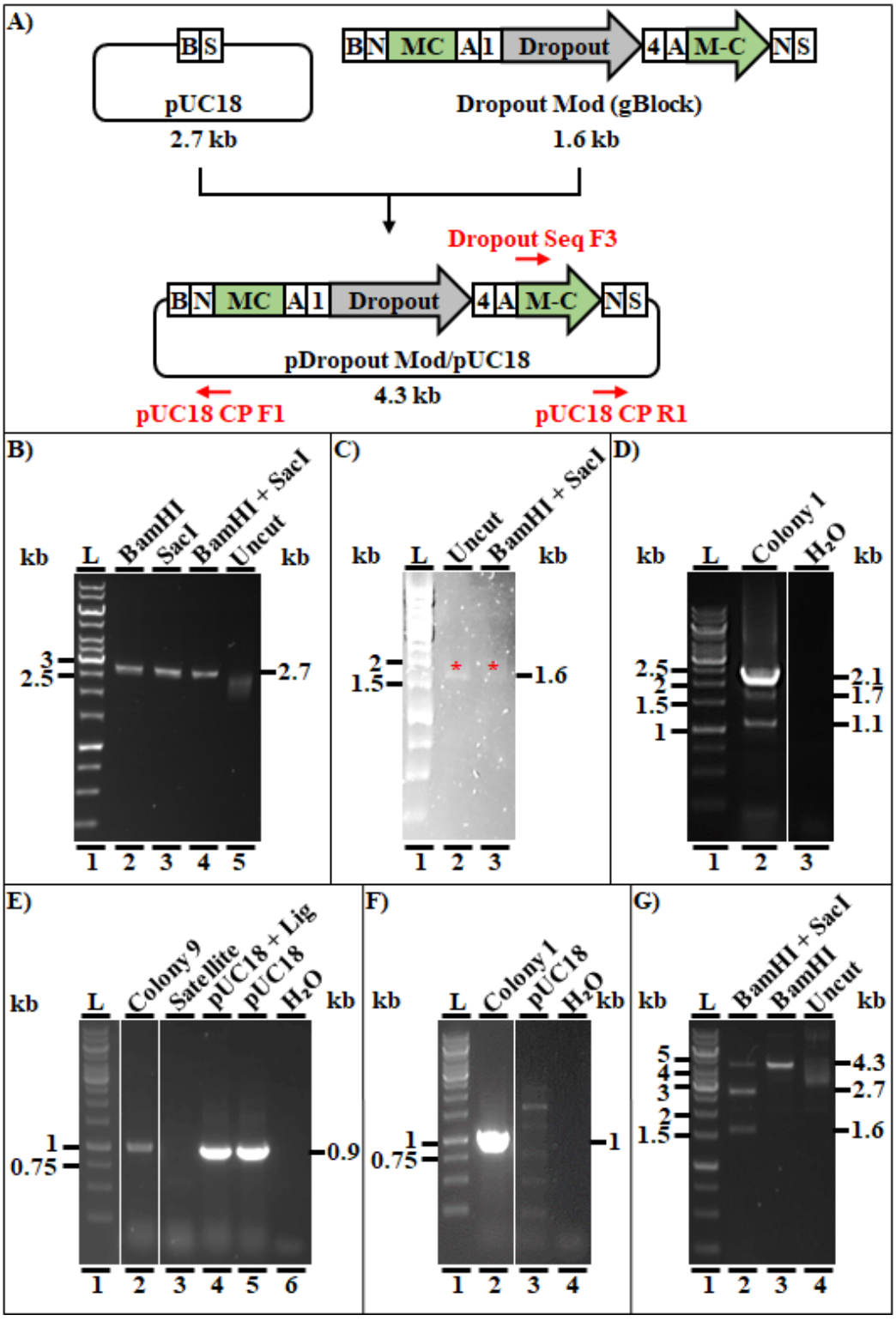
A)
B S
pUC18
2.7 kb
B N MC A 1 Dropout 4 A M-C N S
Dropout Mod (gBlock)
1.6 kb
Dropout Seq F3
B N MC A 1 Dropout 4 A M-C N S
pDropout Mod/pUC18
4.3 kb
pUC18 CP F1
pUC18 CP R1
B)
kb L BamHI SacI BamHI + SacI Uncut kb
3
2.5
2.7
1 2 3 4 5
C)
kb L Uncut BamHI + SacI kb
2
1.5
1.6
1 2 3
D)
kb L Colony 1 H2O kb
2.5
2
1.5
1
2.1
1.7
1.1
1 2 3
E)
kb L Colony 9 Satellite pUC18 + Lig pUC18 H2O kb
1
0.75
0.9
1 2 3 4 5 6
F)
kb L Colony 1 pUC18 H2O kb
1
0.75
1
1 2 3 4
G)
kb L BamHI + SacI BamHI Uncut kb
5
4
3
2
1.5
4.3
2.7
1.6
1 2 3 4

Figure 4.3: Subcloning Dropout Mod gBlock into pUC18.
A) Cloning strategy overview. Double digestion with BamHI/SacI is used to open the pUC18 vector (top left) and produce compatible sticky ends on the Dropout Mod gBlock insert (top right). The digested vector is gel extracted and the gBlock PCR purified before the two are ligated, producing pDropout Mod/pUC18 (middle) containing the Dropout Mod insert inside the pUC18 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. N = NotI. A = AscI. Numbered white boxes: BpiI recognition sites. Green boxes: MCM-C flanks. MC = 5' flank. M-C = 3' flank. Red arrows = primer binding sites. **B, C, G)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI and/or SacI on 1% agarose gels. DNA source: B: pUC18; C: Dropout Mod gBlock; G: colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. **D, E, F)** Reaction products from colony PCRs on 1% agarose gels. Primers used: D, E: pUC18 CP F1/R1; F: pUC18 CP F1/Dropout Seq F3. Template DNA source is shown above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder. Red stars are included above bands which are difficult to see.

4.2.1.4 Generating Golden Gate compatible pCRPrtneo (pCRPrtneo GG)

To produce the Golden Gate compatible pCRPrtneo transformation vector (pCRPrtneo GG), the fragment containing the pDVA_BC dropout region and MCM-C 5' and 3' flanks was subcloned from the pDropout Mod/pUC18 vector into pCRPrtneo using NotI (Figure 4.4, Panel A). The cloning strategy was similar to that used to produce pDropout Mod/pUC18. However, since pCRPrtneo contains only one NotI site, the NotI digested plasmid was first dephosphorylated to prevent recircularisation on ligation, then PCR purified. Next, pDropout Mod/pUC18 was digested with NotI and a fragment corresponding in size to the MCM-C flanked pDVA_BC dropout region insert was gel extracted, ligated to the dephosphorylated recipient vector, and transformed into *E. coli*.

Since the insert carries a LacZ gene, but the recipient vector does not, colonies resulting from the desired ligation event were expected to be easily distinguishable by blue/white screening. A majority of blue colonies were recovered on test plates. There were approximately half the number of colonies on vector + ligase and vector – ligase control plates, but unexpectedly most of these were also blue. Streak plates of the unmodified recipient vector containing X-gal had resulted in only white colonies, as would be expected for a vector lacking LacZ. This, coupled with the presence of blue colonies from the – ligase control indicated the most likely cause was contamination with a second LacZ containing plasmid.

A colony PCR using primers pCRP CP F1/R1 (Figure 4.4, Panel A), which were designed to amplify a 2402 bp fragment in the correctly ligated vector, was used to screen the blue colonies on the test plates. Several colonies from the control plates were also included in an attempt to determine the origins of the ectopic blue colonies. As expected, a white colony from the + ligase control plate yielded a product at the size close to that expected for unmodified pCRPrtneo (Figure 4.4, Panel B, Lane 3). The product from an ectopic blue colony on the – ligase control plate had the same band (Figure 4.4, Panel B, Lane 2). Test colonies yielded three distinct banding patterns: colonies 43, 49 and 50 were representative of each, but none of these contained a product at the expected size (Figure 4.4, Panel B, Lanes 4-6).

To investigate this further, diagnostic digests with BamHI and NotI were conducted on pDNA isolated from all five of these colonies. BamHI was expected to cut the pCRPrtneo backbone once resulting in a linear product. NotI would liberate the insert from the correctly modified vector resulting in two bands, but linearise unmodified vector as with BamHI. As expected, banding patterns for the white colony were consistent with unmodified pCRPrtneo (Figure 4.4, Panel C, Lanes 2-4). Digestion products for the ectopic blue colony (Figure 4.4, Panel C, Lanes 5-7) confirmed the presence of a contaminating vector: linearisation with BamHI (Figure 4.4, Panel C, Lane 5) showed that this was approximately 2.6 kb in size. Colony 49 also produced this banding pattern, suggesting it was derived from the same plasmid, which was unexpected since it had had a distinct banding pattern at colony PCR. Colonies 50 and 43 both contained bands very close in size to those expected for the correct plasmid product (colony 43 bands are shown: Figure 4.4, Panel C, Lanes 8-10). Since only colony 43 had lacked the band expected to correspond to unmodified pCRPrtneo at colony PCR (Figure 4.4, Panel B, Lane 4), its insert region was sequenced. The results confirmed this colony, hereafter referred to as pCRPrtneo GG, contained the full expected product with no mismatches.

It was reasoned that the plasmid contaminant was most likely to have been one used elsewhere in this study. Additionally, since the only antibiotic selection used in this section was kanamycin, the contaminant must have been KanR. In silico analysis of all KanR plasmids used in this work revealed pHSG298 would be expected to give the banding patterns in the diagnostic digest (Figure 4.4, Panel C, Lanes 5-7). Since pCRPrtneo is both KanR and AmpR, replica plating of colonies screened by PCR in this section onto ampicillin plates was expected to result in the death in those containing only pHSG298.

However, all 30 colonies tested (including the five analysed above), grew in this condition. While it was possible that all colonies from this subset had simultaneously been transformed with both plasmids, therefore also gaining ampicillin resistance, it seemed unlikely that there was not a single instance of a colony containing only pHSG298, which would not have been expected to grow in this condition.

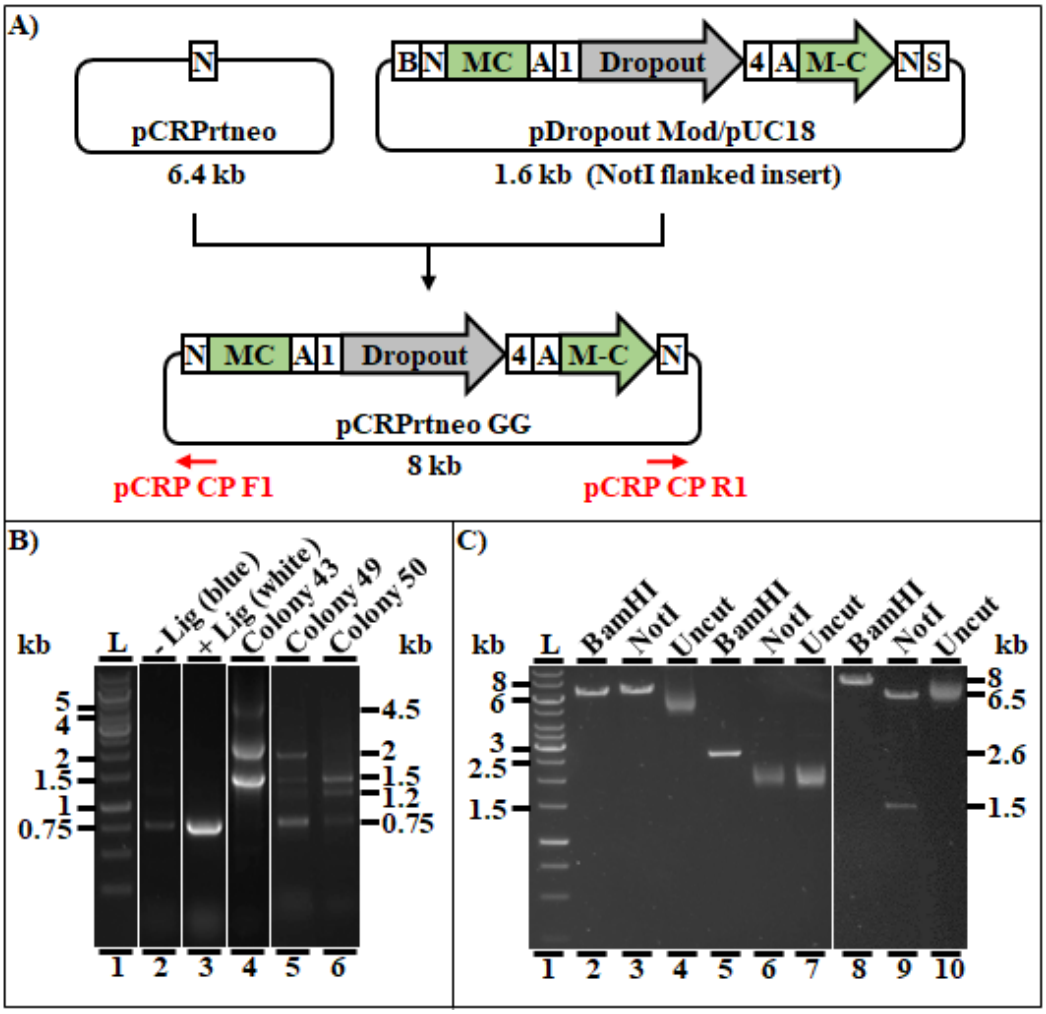

Figure 4.4: Generating Golden Gate compatible pCRPrtneo (pCRPrtneo GG).
A) Cloning strategy overview. Digestion with NotI is used to open the pCRPrtneo vector (top left) and liberate the NotI flanked insert from pDropout Mod/pUC18 backbone (top right). The digested vector is dephosphorylated then PCR purified and the insert gel extracted before the two are ligated, producing pCRPrtneo GG (bottom) containing the pDVA_BC dropout region and MCM-C flanks inside the pCRPrtneo backbone. Approximate relevant expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. N = NotI. A = AscI. Numbered white boxes: BpiI recognition sites. Green boxes: MCM-C flanks. MC = 5' flank. M-C = 3' flank. Red arrows = primer binding sites. **B)** Reaction products from colony PCR on a 1% agarose gel. Primers used: pCRP CP F1/R1. Template DNA source is shown above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. **C)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI or NotI on a 1% agarose gel. DNA source: Lanes 2-4: + Lig (white); Lanes 5-7: -Lig (blue); Lanes 8-10: Colony 43. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.1.5 Generating Golden Gate compatible pBLPrt (pBLPrt GG)

The Golden Gate compatible pBLPrt vector (pBLPrt GG) was produced by subcloning the pDVA_BC dropout region from pDropout Mod/pUC18 into pBLPrt with the same strategy as for pCRPrtneo GG, except using AscI (Figure 4.5, Panel A). Unexpectedly, diagnostic digestion of the initial pBLPrt recipient vector revealed that its total size was approximately 9 kb, which was roughly 600 bp larger than expected. Following ligation of the insert and this vector, and subsequent transformation, ectopic blue colonies were recovered on ligation control plates at similar relative frequencies to test plates as had been the case for pCRPrtneo. This suggested that the same plasmid contaminant was also present here. A colony PCR with primers designed to bind to the vector backbone and amplify a product containing the insert (in this case pBLP CP F1/R1 (Figure 4.5, Panel A)), identified a band at the expected size in some test colonies.

Diagnostic digests were carried out on pDNA isolated from two of these: NdeI was expected to linearise the plasmid, while AscI would liberate the pDVA_BC dropout region insert from the backbone of the correctly ligated product. Both colonies produced the same unexpected banding patterns (colony 44 bands are shown: Figure 4.5, Panel B, Lanes 2-4). NdeI did not appear to cut the product (Figure 4.5, Panel B, Lane 2), but it had successfully linearised the initial pBLPrt plasmid. This may have been due to some change in the plasmid backbone, for example a single nucleotide substitution at this site. Since this site falls outside of functional plasmid regions, such a substitution would not be expected to impede vector function. It is also possible that the failed NdeI digestion was due to a procedural error: for example during pipetting of the enzyme. Digestion with AscI resulted in a single band at approximately 8 kb (Figure 4.5, Panel B, Lane 3), which presumably corresponded to the vector backbone. This product was 400 bp smaller than would have been expected based on the available plasmid sequence, and 1 kb smaller than the expectation based on the size of the pBLPrt vector used here. There was also no smaller band corresponding to the insert, which was unexpected as these colonies had contained a band as predicted for the insert at colony PCR. Sequencing of colony 44, hereafter referred to as pBLPrt GG, revealed that the full insert had been ligated into the vector backbone at the intended site. Since the corresponding insert band had not been recovered at digestion this suggested the bands had migrated further on this portion of the gel, which could explain the discrepancies in vector backbone size.

Digestion of pDNA from two ectopic blue colonies from ligation control plates was also carried out with NdeI and AscI. Both produced the same banding pattern: the product was not cut with either enzyme (Figure 4.5, Panel B, Lanes 5-7). These products were consistent in appearance to the bands presumed to correspond to uncut contaminating plasmid obtained for pCRPrtneo GG (Figure 4.4, Panel C, Lanes 6-7), which strongly suggested this contaminant was the same. Neither AscI nor NdeI would be expected to cleave pHSG298, which had previously been thought to be the plasmid contaminant. It was reasoned that the replica plating test which had suggested the contaminant was ampicillin resistant (which would not be expected for pHSG298), may have given this result due to degradation of the antibiotic on the plates. To test this, colonies containing the pBLPrt and pCRPrtneo plasmids, in addition to the two ectopic blue colonies shown by diagnostic digestion to contain the contaminant from both the pBLPrt GG and pCRPrtneo GG assembly efforts, were inoculated into LB medium containing ampicillin. Following overnight incubation, dense growth was visible for both pBLPrt and pCRPrtneo, as would be expected since both vectors carry AmpR. No growth was obtained for the colonies carrying the contaminating plasmid, suggesting that it was in fact not resistant to ampicillin and therefore could have been pHSG298. To confirm this, pDNA from one of these colonies was sequenced with the primers pHSG298 and M13F, which should have bound to the pHSG298 backbone. However, both sequencing reads were very short and contained only ambiguous, low quality base calls, and were therefore unusable.

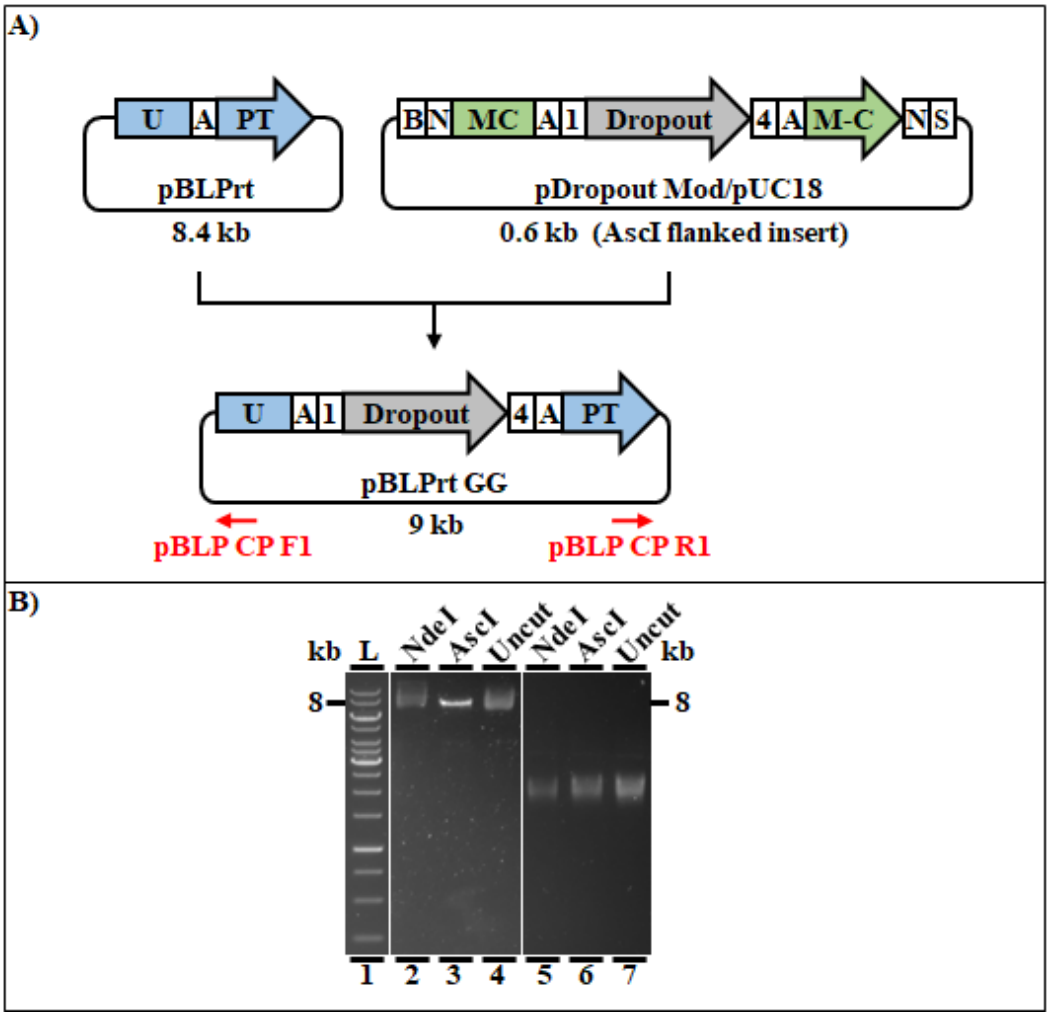

Figure 4.5: Generating Golden Gate compatible pBLPrt (pBLPrt GG).
A) Cloning strategy overview. Digestion with AscI is used to open the pBLPrt vector (top left) and liberate the AscI flanked insert from pDropout Mod/pUC18 backbone (top right). The digested vector is dephosphorylated then PCR purified and the insert gel extracted before the two are ligated, producing pBLPrt GG (bottom) containing the pDVA_BC dropout region inside the pBLPrt backbone. Approximate relevant expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. N = NotI. A = AscI. Numbered white boxes: BpiI recognition sites. Green boxes: MCM-C flanks. MC = 5' flank. M-C = 3' flank. Red arrows = primer binding sites. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with NdeI or AscI on a 1% agarose gel. DNA source: Lanes 2-4: colony 44; Lanes 5-7: - Lig (blue). Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.2 Assembling CRISPR/Cas9 constructs into pCRPrtneo GG/pBLPrt GG

With the Golden Gate compatible *M. maripaludis* transformation vectors (pCRPrtneo GG and pBLPrt GG) produced, the next step was to assemble the three fragments designed in

Chapter 3 into them via the Golden Gate system. Three different assemblies in each transformation vector were to be produced: each carrying one of the alternative versions of Fragment 1 (containing different 20 nt gRNA targeting sequences) along with the invariant Fragments 2 and 3. The resulting set of assembled pCRPrtneo GG and pBLPrt GG transformation vectors would give rise to the test and control strains (see Chapter 3) when transformed into *M. maripaludis* strains S0001 and Mm900, respectively. The first step in this process was to subclone each fragment (Fragments 2 and 3 and the three versions of Fragment 1) into the holding vector pHSG298 (Chapter 3, Figure 3.8, Stage 2). Next, the fragments would be assembled via Golden Gate into both of the modified *M. maripaludis* transformation vectors.

4.2.2.1 Subcloning Fragment 1 gBlocks into pHSG298

The three different versions of the Fragment 1 gBlock differed to each other only in their 20 bp gRNA targeting sequence, which had been designed to target the eGFP and puromycin genes (Fragment 1 eGFP and pur) and a non-targeting control (Fragment 1 ctrl). The same strategy as had been used to generate pDropout Mod/pUC18 was used to subclone all three Fragment 1 gBlocks into separate instances of the kanamycin resistant pHSG298 vector (Figure 4.6, Panel A). Following BamHI/SacI digestion of each gBlock and the recipient vector, ligation, and transformation, a majority of white colonies were recovered on plating, as would be expected since the BamHI/SacI sites in pHSG298 are within a LacZ gene. A colony PCR using primers M13 F and PHSG R1 confirmed the presence of a product at the expected size in all nine white colonies tested for each of the different gBlock ligations. Diagnostic digests were carried out on pDNA isolated from two colonies for each of the different versions of Fragment 1 ligated into pHSG298, all of which produced the same expected banding pattern. (Fragment 1 eGFP colony 2 bands are shown: Figure 4.6, Panel B, Lanes 2-4). All six colonies were sequenced, and the insert in each confirmed to be correct. Fragment 1 eGFP colony 2, pur colony 11 and ctrl colony 26 were selected for further use, hereafter referred to as pFragment 1 eGFP/pHSG298, pur/pHSG298 and ctrl/pHSG298 respectively.

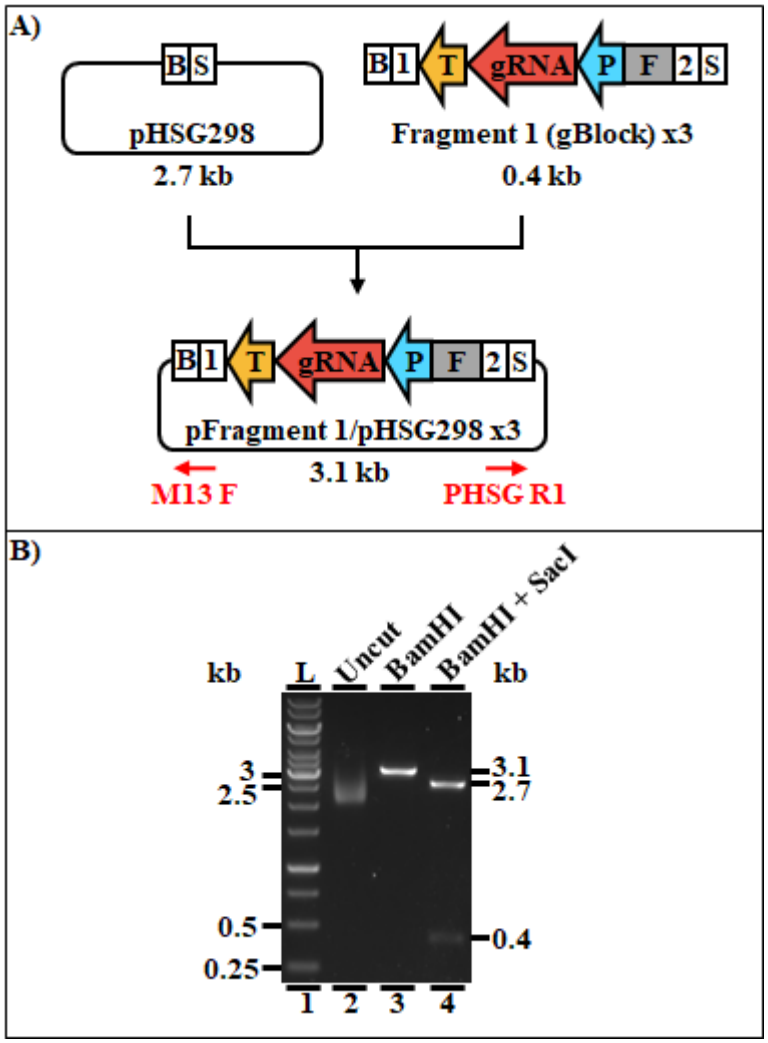

Figure 4.6: Subcloning Fragment 1 gBlocks into pHSG298.
A) Cloning strategy overview. Double digestion with BamHI/SacI is used to open the pHSG298 vector (top left) and produce compatible sticky ends on the three Fragment 1 gBlock inserts (top right). The gBlocks are PCR purified and each is ligated separately into the digested and gel extracted vector, producing the three different pFragment 1/pHSG298 vectors (bottom), each containing a different version of Fragment 1 ligated into the pHSG298 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. Numbered white boxes: BpiI recognition sites. T = Tmcr. P = Phmv. F = filler sequence. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI and/or SacI on a 1% agarose gel. DNA source: Fragment 1 eGFP colony 2. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.2.2 Subcloning Fragment 2 gBlock into pHSG298

The Fragment 2 gBlock was subcloned into the pHSG298 holding vector (Figure 4.7, Panel A) as explained for Fragment 1. Following a ligation at a 3:1 insert:vector ratio, just 11 white colonies were recovered after overnight transformation plate growth. Screening by colony PCR using primers M13 F and PHSG R1, which bind to the vector backbone either side of the insert, produced several different banding patterns, but no band at the expected size was recovered for any colony. The results were similar for a second PCR using the internal Frag2 Seq F4 primer (which was designed to bind within the insert), in combination with PHSG R1. Unexpectedly, diagnostic digestions on colonies representative of each of these distinct banding patterns yielded no products, suggesting they did not contain a plasmid. Since the Fragment 2 insert was similar in size to the recipient vector (2149 and 2675 bp, respectively), it was reasoned that a ligation with a 1:1 insert:vector ratio may be more appropriate. Following transformation, a band at the expected size was recovered in 6 of the 8 white colonies tested by colony PCR with M13 F and PHSG R1. Diagnostic digests on pDNA from two of these colonies (colonies 43 and 46) revealed both produced the expected banding pattern, with band sizes slightly above those expected (colony 46 bands are shown: Figure 4.7, Panel B, Lanes 2-4). Sequencing of these colonies confirmed they both contained the full Fragment 2 insert, and colony 46 (hereafter referred to as pFragment 2/pHSG298) was selected for further use.

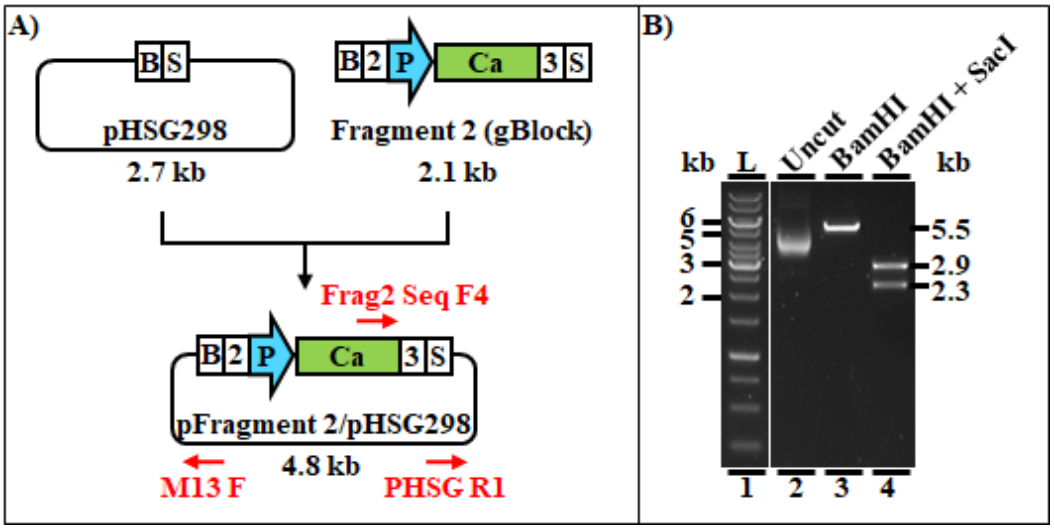

Figure 4.7: Subcloning Fragment 2 gBlock into pHSG298.
A) Cloning strategy overview. Double digestion with BamHI/SacI is used to open the pHSG298 vector (top left) and produce compatible sticky ends on the Fragment 2 gBlock insert (top right). The digested vector is gel extracted and the gBlock PCR purified before the two are ligated, producing pFragment 2/pHSG298 (bottom) containing the Fragment 2 insert inside the pHSG298 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. Numbered white boxes: BpiI recognition sites. P = Phmv. Ca = Cas9 5' end. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI and/or SacI on a 1% agarose gel. DNA source: Fragment 2 colony 46. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.2.3 Subcloning Fragment 3 gBlock into pHSG298

The Fragment 3 gBlock was subcloned into pHSG298 (Figure 4.8, Panel A) as explained for Fragments 1 and 2. Colony PCR with primers M13 F/PHSG R1 identified a single white colony which produced a band at the expected size for the Fragment 3 insert. In a diagnostic digest on pDNA from this colony (colony 1) two bands at the expected sizes were recovered for the double digestion (Figure 4.8, Panel B, Lane 2). Similarly to the case for Fragment 2 colony 46, the linearised product appeared just above the expected size (Figure 4.8, Panel B, Lane 3). Sequencing of this colony, hereafter referred to as pFragment 3/pHSG298, confirmed it contained the correct insert.

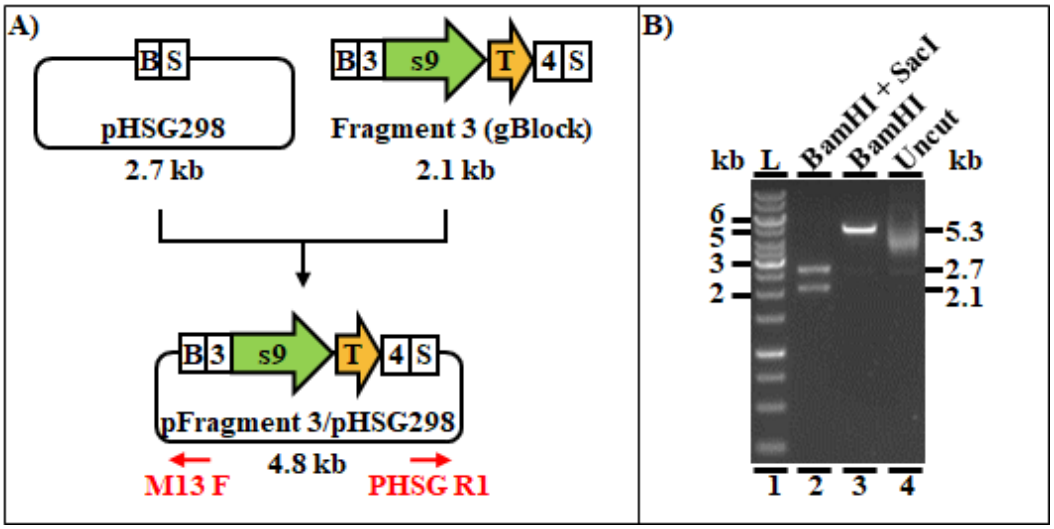

Figure 4.8: Subcloning Fragment 3 gBlock into pHSG298.
A) Cloning strategy overview. Double digestion with BamHI/SacI is used to open the pHSG298 vector (top left) and produce compatible sticky ends on the Fragment 3 gBlock insert (top right). The digested vector is gel extracted and the gBlock PCR purified before the two are ligated, producing pFragment 3/pHSG298 (bottom) containing the Fragment 3 insert inside the pHSG298 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. Numbered white boxes: BpiI recognition sites. S9 = Cas9 3' end. T = Tmcr. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI and/or SacI on a 1% agarose gel. DNA source: Fragment 3 colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.2.4 Assembling CRISPR/Cas9 constructs into pCRPrtneo GG

With all the required vectors constructed, a Golden Gate reaction was conducted to assemble the three versions of the CRISPR construct (each containing an alternative Fragment 1) into pCRPrtneo GG. To achieve this, three separate Golden Gate reactions, each containing the pCRPrtneo GG destination vector, the pFragment 2/pHSG298 and pFragment 3/pHSG298 holding vectors, and each of the three alternative pFragment 1/pHSG298 vectors were done in parallel. The anticipated outcome was a set of three vectors in which the pDVA_BC dropout region had been replaced with Fragments 1-3 fused together at their matching BpiI overlaps, enclosed within the MCM-C flanking regions in the pCRPrtneo GG backbone (Figure 4.9, Panel A).

Following transformation of the Golden Gate reactions hundreds of colonies were recovered, the vast majority of which were white, as was expected since the pCRPrtneo GG dropout region contained a LacZ gene, which should have been replaced on correct assembly. A colony PCR using primers pCRP CP F1/R1 was conducted on nine white

colonies for each of the alternative Fragment 1 assemblies. All but one produced a band at the size expected for a correctly assembled insert. pDNA from two colonies for each of the three Fragment 1 combinations which had given the expected band at colony PCR was isolated for diagnostic digestion. BamHI digestion was expected to linearise the correctly assembled vector, yielding a 12154 bp product, while NotI would liberate the full Fragment 1-3 insert and MCM-C flanks from the pCRPrtneo GG backbone, producing 5714 and 6440 bp products respectively. All six digested plasmids resulted in the same banding patterns (pCRPrtneo GG eGFP colony 1 bands are shown: Figure 4.9, Panel B, Lanes 2-4). While the linearised product was above the size range of the ladder used, its position, taken together with the fact that the bands resulting from the double digestion were at the expected sizes, suggested that all colonies contained the correct insert. One plasmid from a colony for each of the alternative Fragment 1s (pCRPrtneo GG eGFP colony 1, pur colony 10 and ctrl colony 19) was sequenced, and all were found to contain the full correctly assembled insert with no mismatches to what was expected. These are hereafter referred to as pCRPrtneo GG eGFP, pur and ctrl, respectively.

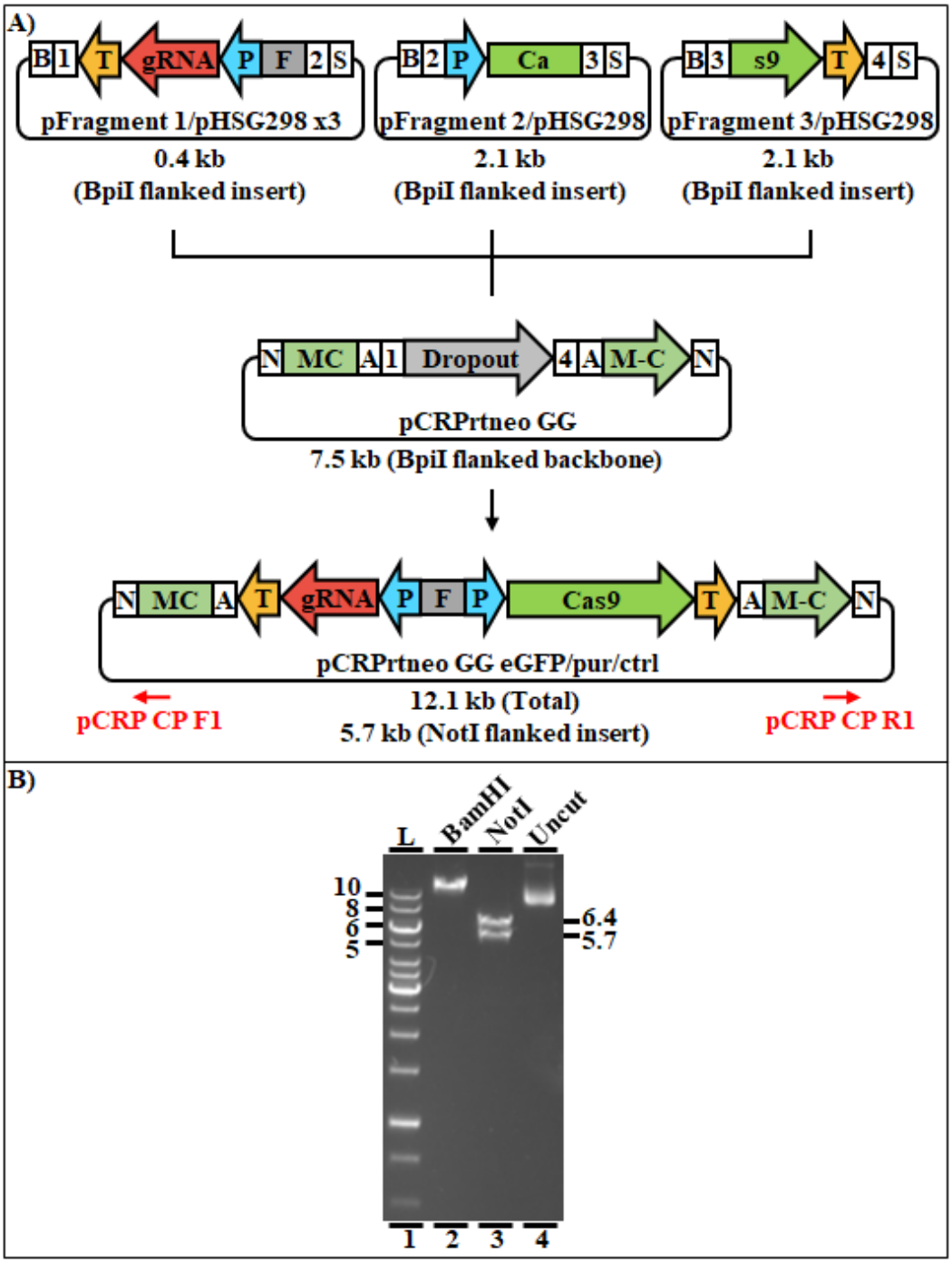
A)
B 1
T
gRNA
P
F
2 S
pFragment 1/pHSG298 x3
0.4 kb
(BpiI flanked insert)
B 2
P
Ca
3 S
pFragment 2/pHSG298
2.1 kb
(BpiI flanked insert)
B 3
s9
T
4 S
pFragment 3/pHSG298
2.1 kb
(BpiI flanked insert)
N
MC
A 1
Dropout
4 A
M-C
N
pCRPrtneo GG
7.5 kb (BpiI flanked backbone)
N
MC
A
T
gRNA
P
F
P
Cas9
T
A
M-C
N
pCRPrtneo GG eGFP/pur/ctrl
12.1 kb (Total)
5.7 kb (NotI flanked insert)
pCRP CP F1
pCRP CP R1
B)
L
BamHI
NotI
Uncut
10
8
6
5
6.4
5.7
1
2
3
4

Figure 4.9: Assembling CRISPR/Cas9 constructs into pCRPrtneo GG.
A) Cloning strategy overview. Three separate Golden Gate reactions are conducted using the holding vectors containing Fragments 1-3, each using one of the alternative versions of Fragment 1 (top) and the destination vector pCRPrtneo GG (middle). The result is a set of three vectors each containing the CRISPR/Cas9 construct with an alternative gRNA, flanked by the MCM-C homology regions in the pCRPrtneo backbone (bottom). Relevant approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. N = NotI. A = AscI. Numbered white boxes: BpiI recognition sites. Numbers indicate the identity of the overlap corresponding to each. T = Tmcr. P = Phmv. F = filler sequence. Ca = Cas9 5' end. S9 = Cas9 3' end. MC = MCM-C 5' flank. M-C = MCM-C 3' flank. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI or NotI on a 1% agarose gel. DNA source: pCRPrtneo GG eGFP colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.2.5 Assembling CRISPR/Cas9 constructs into pBLPrt GG

The three versions of the CRISPR/Cas9 constructs were assembled into pBLPrt GG (Figure 4.10, Panel A) as described for pCRPrtneo GG. Colony PCR with primers pBLP CP F1/R1 identified a band at the expected size for the insert in all but one of the nine white colonies screened for each of the different Fragment 1 assemblies. Diagnostic digests using XmaI and AscI were conducted on pDNA isolated from two colonies for each Fragment 1 combination which had screened positive at colony PCR. XmaI was expected to linearise the 13130 bp vector, while AscI was predicted to cut the Fragment 1-3 insert out of the vector backbone, giving 4607 and 8523 bp bands respectively. pDNA from five colonies gave the same banding pattern (pBLPrt GG eGFP colony 4 bands are shown: Figure 4.10, Panel B, Lanes 5-7). For these five colonies, the XmaI product was indistinguishable from the uncut control (Figure 4.10, Panel B, Lanes 5 and 6), indicating it had not cut. AscI digestion produced a band at the expected size for the liberated insert, but the larger product presumed to correspond to the vector backbone (Figure 4.10, Panel B, Lane 7) was 1 kb larger than the backbone fragment which had previously been obtained for pBLPrt GG (Figure 4.5, Panel B, Lane 3). However, this product did correspond in size to the original pBLPrt vector used in this study, which was found to be about 9kb. This suggests that the unexpected size which had been obtained for the pBLPrt GG backbone fragment (Figure 4.5, Panel B, Lane 3) was most likely due to unequal band migration on the gel.

Unexpectedly, pBLPrt GG eGFP colony 3 pDNA was not cut by either enzyme, and all lanes contained a band resembling uncut plasmid (Figure 4.10, Panel B, Lanes 2-4) at a distinct size to the uncut controls for the other five colonies (Figure 4.10, Panel B, Lane 5).

The fact that no colonies had been recovered on plates transformed with a control Golden Gate reaction using dH_2O instead of component vectors, coupled with the fact that this colony has produced a band at the expected size for the desired insert at colony PCR, made a contaminating vector explanation seem unlikely. An alternative possibility is that the vector in the cell which gave rise to the colony used to generate the digested pDNA had undergone a recombination event, resulting in the loss of a fragment containing both AscI sites prior to overnight growth on the master plate following colony PCR, but before propagation in liquid culture.

Sequencing of one plasmid from a colony for each of the alternative Fragment 1s (pBLPrt GG eGFP colony 4, pur colony 13 and ctrl colony 21) revealed that all contained the correctly assembled insert. They are hereafter referred to as pBLPrt GG eGFP, pur and ctrl, respectively. The XmaI cut site, which is within the insert, was found to be correct in all three colonies, suggesting the failed digestion was likely due to enzyme degradation.

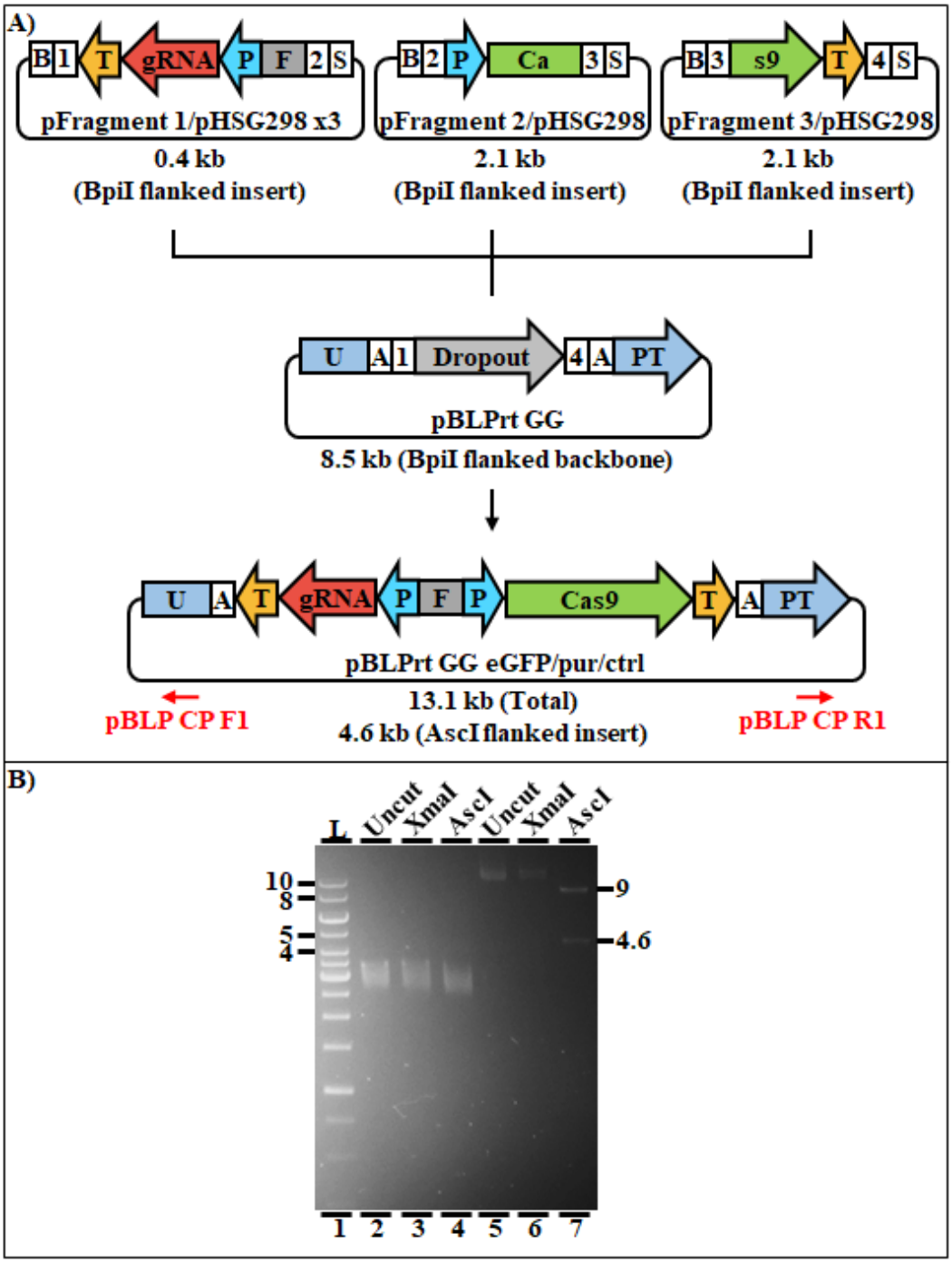
A)
B 1 T gRNA P F 2 S
pFragment 1/pHSG298 x3
0.4 kb
(BpiI flanked insert)
B 2 P Ca 3 S
pFragment 2/pHSG298
2.1 kb
(BpiI flanked insert)
B 3 s9 T 4 S
pFragment 3/pHSG298
2.1 kb
(BpiI flanked insert)
U A 1 Dropout 4 A PT
pBLPrt GG
8.5 kb (BpiI flanked backbone)
U A T gRNA P F P Cas9 T A PT
pBLPrt GG eGFP/pur/ctrl
13.1 kb (Total)
4.6 kb (AscI flanked insert)
pBLP CP F1
pBLP CP R1
B)
L
Uncut
XmaI
AscI
Uncut
XmaI
AscI
10
8
5
4
9
4.6
1 2 3 4 5 6 7

Figure 4.10: Assembling CRISPR/Cas9 constructs into pBLPrt GG.
A) Cloning strategy overview. Three separate Golden Gate reactions are conducted using the holding vectors containing Fragments 1-3, each using one of the alternative versions of Fragment 1 (top) and the destination vector pBLPrt GG (middle). The result is a set of three vectors each containing the CRISPR/Cas9 construct with an alternative gRNA, flanked by the UPT homology regions in the pBLPrt backbone (bottom). Relevant approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. A = AscI. Numbered white boxes: BpiI recognition sites. Numbers indicate the identity of the overlap corresponding to each. T = Tmcr. P = Phmv. F = filler sequence. Ca = Cas9 5' end. S9 = Cas9 3' end. U = UPT 5' flank. PT = UPT 3' flank. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with XmaI or AscI on a 1% agarose gel. DNA source: Lanes 2-4: pBLPrt GG eGFP colony 3; Lanes 5-7: pBLPrt GG eGFP colony 4. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.3 Modifying pLW40/pAW42 to Contain the eGFP Target Site

To make the invader plasmids targetable by *M. maripaludis* test strains carrying the eGFP gRNA, they needed to be modified to contain the appropriate target. This would require several steps. First, a suitable site for the delivery of the insert needed to be found in both vectors. The portion of the eGFP gene to be inserted would then have to be selected, and an appropriate cloning strategy for achieving the desired insertion would have to be designed.

4.2.3.1 Integration site selection

To make any results obtained from invader assays using each of the two invader plasmids more comparable, while also simplifying the cloning process, an integration site in the same relative location in both plasmids was sought. The two vectors were aligned, and the region of homology between them inspected for restriction enzyme sites which could be used for the integration. All single cutters from the list of 618 "commercially available enzymes" available in Geneious (*56*) were annotated onto the alignment. The critical parameter for site selection was that integration at that site should have the lowest possible chance of impeding vector function. As such, all sites present within annotated functional components such as genes were disregarded. Additionally, sites on the vector backbone between linked components: for example a gene and its promoter were also not considered. Just three of the annotated recognition sites met these requirements. Due to their relative positions, directional cloning using any of the possible pairs was not suitable, meaning

integration using a single enzyme was the only option, so the ClaI site was selected (Figure 4.11).

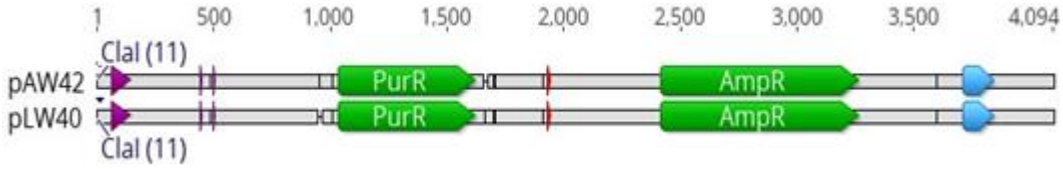

Figure 4.11: Selected insertion site for eGFP target region in pLW40/pAW42. The location of the selected ClaI insertion site is indicated at the 5' end of the region of sequence homology between plasmids pLW40 and pAW42. Grey portions of horizontal bars indicate shared sequence identity, black portions indicate mismatches, and regions without bars show gaps. Functional regions of the sequences and their associated features are indicated by coloured arrows. Green = genes: PurR = puromycin resistance gene; AmpR = ampicillin resistance gene. Purple = promotors. Red = terminators. Blue = replication origins. Labels to the far left indicate the identity of each sequence. Consensus positions are indicated by numbers above the alignment. The 3' portion of the alignment containing the non-homologous regions of the two plasmids has been clipped for clarity.

4.2.3.2 Modification Strategy

With the insertion site selected, the portion of the eGFP gene to be placed into it needed to be chosen. While the 20 bp target region plus an appropriate PAM sequenced was envisaged to be sufficient, a 450 bp portion of the eGFP gene was selected to make it easier to handle. The selected region contained the eGFP target site with 141 bp of 5' and 289 bp of 3' flanking sequence. To facilitate delivery of the selected region of eGFP into the ClaI site in pLW40/pAW42, a gBlock compatible with a cloning strategy similar to that used to generate the Golden Gate *M. maripaludis* transformation vectors was designed. The gBlock (hereafter referred to as eGFP Mod) contained the selected 450 bp region of the eGFP gene flanked by ClaI sites (Figure 4.12, Panel B, Stage 1). BamHI and SacI sites were then appended to each end to enable initial insertion into the pUC18 cloning vector (Figure 12, Panel B, Stage 2). Subcloning using ClaI would then be used to deliver the eGFP target site into the invader plasmids (Figure 4.12, Panel B, Stage 3). The aim had been to generate a pair of modified invader plasmids with the eGFP insert not only in the same relative location, but also in the same relative orientation. However, since the subcloning step was to be conducted using a single enzyme (ClaI), the direction of insertion could not be controlled. The full insert was anticipated to be small enough to be covered by a single sequencing read, so it was decided that several modified pLW40 and pAW42 plasmids would be sequenced to enable the recovery of a pair with the insert in the same relative orientation.

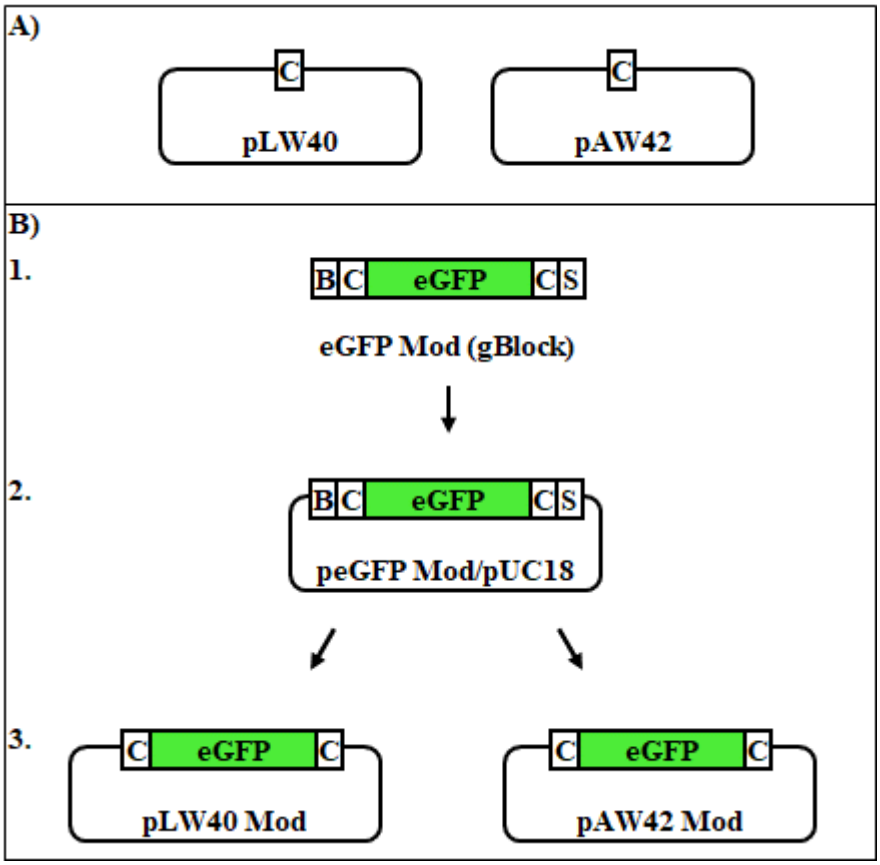

Figure 4.12: Producing modified pLW40/pAW42 vectors scheme.
A) Relevant features of the vectors used in this scheme. pLW40 (left) and pAW42 (right) both contain a ClaI site in a location suitable for insertion at the same relative position. **B)** Vector modification scheme. **1:** A gBlock (eGFP Mod) is constructed containing a portion of the eGFP gene carrying the eGFP gRNA target site. ClaI sites are appended to each end to allow subcloning into pLW40 and pAW42. BamHI and SacI sites are added to the ends of the gBlock to enable initial insertion into the cloning vector pUC18. **2:** Restriction cloning with BamHI/SacI is used to insert the gBlock into pUC18. **3:** The ClaI flanked insert in peGFP Mod/pUC18 is subcloned into both pLW40 and pAW42 producing a pair of modified invader plasmids carrying the eGFP target site, which should be targetable by *M. maripaludis* strains carrying gRNA-eGFP when used as invader plasmids. White boxes: restriction sites. B = BamHI. S = SacI. C = ClaI. Green box = selected portion of the eGFP gene carrying the gRNA target site.

4.2.3.3 Subcloning eGFP Mod gBlock into pUC18

The eGFP Mod gBlock was subcloned into the pUC18 holding vector (Figure 4.13, Panel A) using the same approach as had been used for the Dropout Mod gBlock. Following ligation of the BamHI/SacI digested gBlock and vector, white colonies, which were expected to contain an insert, were screened by colony PCR using the primers M13 F/R.

The majority of these yielded the band expected for the eGFP insert. Diagnostic digestion of one colony (colony 1) revealed that the larger band presumed to correspond to the backbone in the double digestion and the linearised plasmid (Figure 4.13, Panel B, Lanes 2-3) were both larger than expected. Despite this a smaller band close to the expected size for the desired insert was recovered (Figure 4.13, Panel B, Lane 2), and sequencing revealed this insert was correct for its full length. This colony and its plasmid, hereafter referred to as peGFP Mod/pUC18, was selected for further use regardless of the size issues with the vector backbone since this region would not be retained in the final plasmid products pLW40/pAW42 Mod.

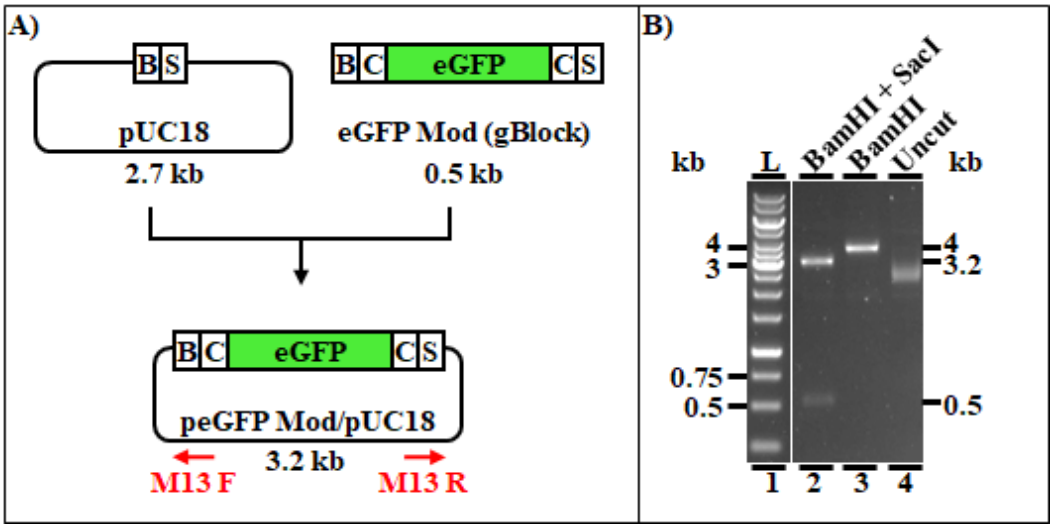

Figure 4.13: Subcloning eGFP Mod gBlock into pUC18.
A) Cloning strategy overview. Double digestion with BamHI/SacI is used to open the pUC18 vector (top left) and produce compatible sticky ends on the eGFP Mod gBlock insert (top right). The digested vector is gel extracted and the gBlock PCR purified before the two are ligated, producing peGFP Mod/pUC18 (bottom) containing the eGFP Mod insert inside the pUC18 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. C = ClaI. Green box = selected portion of the eGFP gene carrying the gRNA target site. Red arrows = primer binding sites. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI and/or SacI on a 1% agarose gel. DNA source: colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.3.4 Generating modified pLW40 (pLW40 Mod)

To produce the modified pLW40 vector, the ClaI flanked insert containing the eGFP target site was subcloned from peGFP Mod/pUC18 into pLW40 using ClaI (Figure 4.14, Panel A). Unexpectedly, initial attempts to cut both pLW40 and peGFP Mod/pUC18 with ClaI were unsuccessful. Further investigation revealed this was because the ClaI site in pLW40

and one of those in peGFP Mod/pUC18 overlapped with Dam sites, which are methylated in DH5α: the *E. coli* strain from which these plasmids had been isolated. To address this, both plasmids were transformed into dam^-/dcm^- competent *E. coli* cells, which cannot methylate these sites. Following purification, both plasmids produced bands at the expected sizes when digested with ClaI, indicating that the unexpected backbone sizes previously obtained for peGFP Mod/pUC18 (Figure 4.13, Panel B, Lanes 2-3) were likely due to abnormal band migration on the gel. The band corresponding to the double digested peGFP Mod/pUC18 insert was gel extracted, ligated to digested and dephosphorylated pLW40, and the ligation product transformed back into *E. coli* DH5α.

Since neither the insert nor recipient vector contained a LacZ gene, blue/white screening was not possible. PCR of 14 blue colonies using primers pLW40 CP F1/pAW42 CP R1 revealed that 11 produced a band at the expected size. pDNA from five of these was generated for diagnostic digestion and sequencing. BamHI was expected to linearise the plasmid. Since these plasmids had been obtained from DH5α, ClaI could not be used for the double digestion. Instead, NaeI and SacI, which were expected to cut close to the insert borders generating 683 and 10081 bp products were used. The same bands were generated from the digested plasmids of all five colonies (colony 1 bands are shown: Figure 4.14, Panel B, Lanes 2-4). The double digested band was the same as for the single digestion (Figure 4.14, Panel B, Lanes 3+4), indicating either that one of the enzyme target sites had mutated/been sequenced incorrectly in the recipient vector, or that one of the enzymes used was not functional. It was not possible to accurately estimate the size of the linearised products since they were beyond the size range of the ladder. However, since their position seemed reasonable given the expected size, and due to time constraints, the inserts in all five colonies were sequenced without conducting further digestions. All colonies were found to contain a single copy of the correct insert. As expected these were a mixture of the two possible orientations.

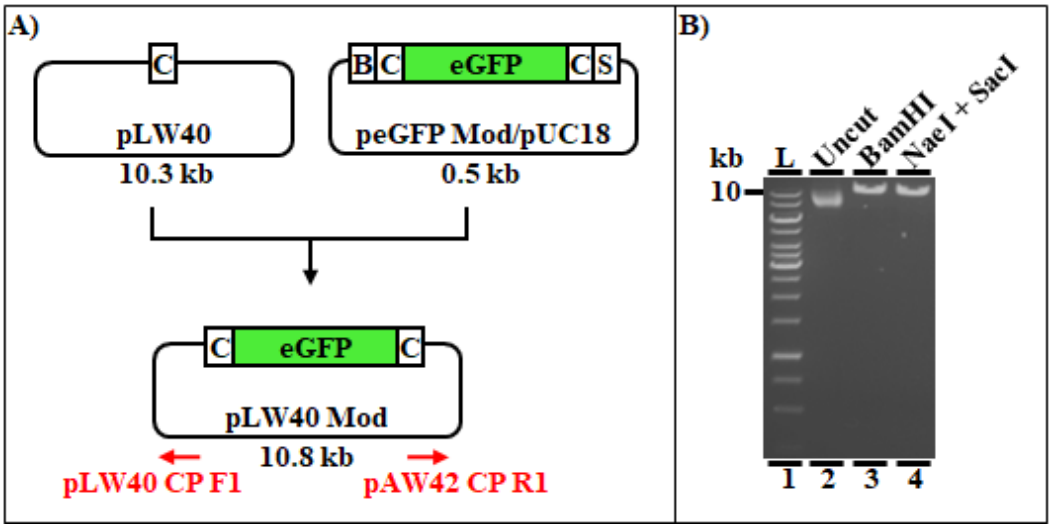

Figure 4.14: Generating modified pLW40 (pLW40 Mod).
A) Cloning strategy overview. Digestion with ClaI is used to open the pLW40 vector (top left) and liberate the ClaI flanked insert from the peGFP Mod/pUC18 backbone (top right). The digested vector is dephosphorylated then PCR purified and the insert gel extracted before the two are ligated, producing pLW40 Mod (bottom) containing the eGFP Mod insert inside the pLW40 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. C = ClaI. Green box = selected portion of the eGFP gene carrying the gRNA target site. Red arrows = primer binding sites. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI or NaeI/SacI on a 1% agarose gel. DNA source: colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left. L = GeneRuler 1 kb ladder.

4.2.3.5 Generating modified pAW42 (pAW42 Mod)

pAW42 Mod was produced using the same approach as for pLW40 Mod (Figure 4.15, Panel A). pDNA from five colonies which had been found to contain the expected band at colony PCR using the pAW42 CP F1/R1 primers was isolated for diagnostic digestion. BamHI was used to linearise the vector, while NaeI and SpeI were used to excise a fragment containing the insert from the vector backbone, with expected sizes of 667 and 4743 bp, respectively. All pDNA samples produced the same banding pattern. As had been the case for pLW40, the double and single digested products were the same (Figure 4.15, Panel B, Lanes 3-4). The most likely cause was therefore inactivity of the NaeI enzyme, which had been used in both double digests. The linearised product was very close to the expected size (Figure 4.15, Panel B, Lane 3), and sequencing revealed that all five plasmids contained one copy of the desired insert. As for pLW40, inserts were recovered in both directions, which meant that a pair with the insert in the same relative orientation had been produced. Due to time constraints further digestions were not carried out.

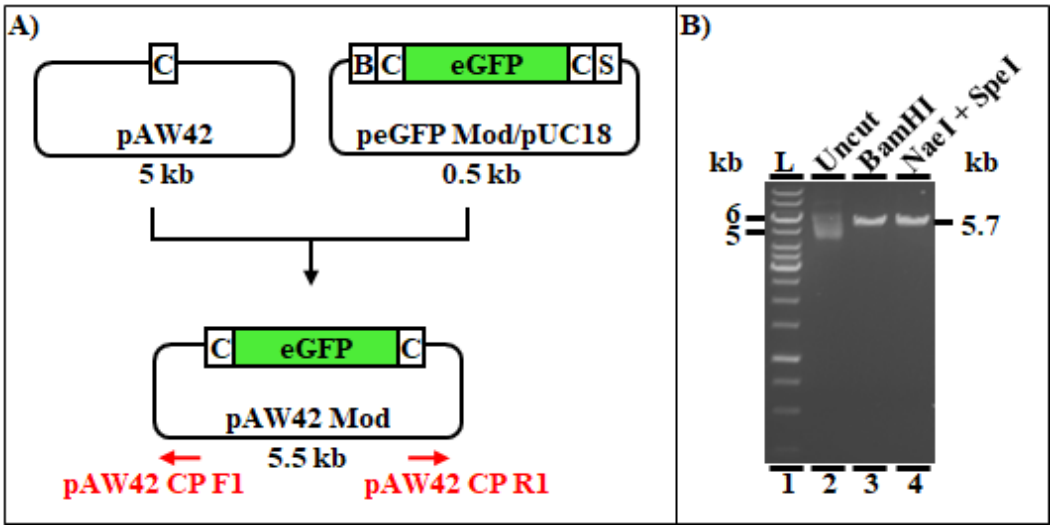

Figure 4.15: Generating modified pAW42 (pAW42 Mod).
A) Cloning strategy overview. Digestion with ClaI is used to open the pAW42 vector (top left) and liberate the ClaI flanked insert from the peGFP Mod/pUC18 backbone (top right). The digested vector is dephosphorylated then PCR purified and the insert gel extracted before the two are ligated, producing pAW42 Mod (bottom) containing the eGFP Mod insert inside the pAW42 backbone. Approximate expected product sizes are shown below each component. White boxes: restriction sites. B = BamHI. S = SacI. C = ClaI. Green box = selected portion of the eGFP gene carrying the gRNA target site. Red arrows = primer binding sites. **B)** 100 ng aliquots of DNA either undigested (uncut) or digested with BamHI or NaeI/SpeI on a 1% agarose gel. DNA source: colony 1. Restriction enzymes used are indicated above and lane numbers below each lane. Band sizes are in kb to the sides of the figure: ladder to the left, products to the right. L = GeneRuler 1 kb ladder.

4.2.4 Producing *M. maripaludis* test and control strains

With all the vectors required for the invader assay produced, the next step was to create the *M. maripaludis* test/control strains which, once complete, could be transformed with the modified invader plasmids to carry out the invader assay. Producing these strains would require transformation of *M. maripaludis* with the three modified integrative vectors, each carrying the assembled CRISPR/Cas9 construct with an alternative version of the gRNA flanked by their respective homology arms. Subculturing of the transformed cells in the appropriate antibiotics would be used to encourage the desired loop-in/loop-out recombination events resulting in replacement of the target *M. maripaludis* genomic locus (UPT for Mm900 and MCM-C for S0001) with the CRISPR/Cas9 constructs. Due to time constraints it was only possible to attempt one set of transformations. Strain S0001 was selected because it was expected to be transformed at a higher rate.

The plasmids pCRPrtneo GG eGFP, pur and ctrl, along with an H_2O control, were transformed separately into S0001 cells. The cultures used should have been at an OD_{600} value of 0.7-1 following overnight growth. However, the S0001 cells used were all growing too slowly to allow for this. Overnight cultures, which had been set up 14 hours prior to transformation all gave OD_{600} readings below 0.4, and even cultures which had been growing for 34 hours gave readings below 0.7. For this reason, the above plasmids/control were transformed into a pair of cultures within the correct OD_{600} range which had been growing for 58 hours, and the two of the cultures which had been set up 34 hours previously that were the most well grown (Table 4.2).

Table 4.2: S0001 cultures used for transformation.

Material Transformed	S0001 Culture Used	
	Hours Grown	OD_{600}
pCRPrtneo GG eGFP	58	0.89
pCRPrtneo GG pur	58	0.78
pCRPrtneo GG ctrl	34	0.57
H_2O	34	0.61

Following transformation and subsequent overnight growth, the transformed cells were each subcultured in duplicate into fresh medium supplemented with neomycin, which should have selected for cells which had undergone the desired loop-in recombination event, and therefore carried the full plasmid construct including NeoR, integrated at the target locus by HR. For each transformation, a control subculture into medium without selection was also carried out. Growth of all cultures was monitored by taking OD_{600} measurements every 12 hours, for the full 48 hours required for this selection step (Table 4.3).

Table 4.3: OD_{600} growth measurements of subcultured transformed cells.

Material Transformed	Culture	Neomycin	OD_{600}			
			12 h	24 h	36 h	48 h
pCRPrtneo GG eGFP	A	+	0.105	0.261	0.302	0.312
pCRPrtneo GG eGFP	B	+	0.098	0.289	0.359	0.362
pCRPrtneo GG eGFP	C	-	0.180	0.444	0.656	0.678
pCRPrtneo GG pur	A	+	0.098	0.185	0.220	0.228
pCRPrtneo GG pur	B	+	0.104	0.174	0.176	0.181
pCRPrtneo GG pur	C	-	0.155	0.307	0.422	0.446
pCRPrtneo GG ctrl	A	+	0.068	0.262	0.266	0.289
pCRPrtneo GG ctrl	B	+	0.087	0.249	0.265	0.272
pCRPrtneo GG ctrl	C	-	0.177	0.414	0.558	0.654
H_2O	A	+	0.110	0.271	0.290	0.293
H_2O	B	+	0.103	0.284	0.342	0.360
H_2O	C	-	0.185	0.424	0.606	0.668

All cultures grew less well when supplemented with neomycin (Table 4.3), as would be expected since even cells which had not taken up the NeoR gene were expected to grow in this condition. However, at all time points the growth of the H_2O mock transformed control cultures was either greater than or approximately the same as each of the cultures transformed with plasmid. This was not as expected since theoretically no cells in the control should have contained NeoR, and therefore should not have been able to survive when supplemented with this antibiotic. In practice some amount of background growth had been expected, however since the level of background (i.e. the mock transformed control cultures) was the same as that observed for the plasmid transformed cultures, this suggested that transformed cultures contained only background cells, and ultimately that the transformation had been unsuccessful. Unfortunately there was insufficient time to attempt further transformations.

4.3 Discussion

The aim of this section was to produce all the plasmid constructs required for the invader assay, and then to use the modified integrative vectors to create the *M. maripaludis* test and control strains which could be used to perform this assay. While all the required plasmids were produced, due to time constrains it was not possible to generate the *M. maripaludis* test and control strains.

4.3.1 Construct Assembly Issues

Due to issues encountered during assembly, the production of many of the constructs required for the invader assay was more challenging, and therefore more time consuming than expected. Perhaps most notable was the contamination encountered during the production of the Golden Gate compatible pBLPrt/pCRPrtneo vectors. The plasmid insert used in both cases carried a LacZ gene, while neither recipient vector contained this feature. It was therefore anticipated that screening would be simple, since only the desired ligation products should have been blue in the presence of X-gal. However, in both cases blue colonies were recovered on plates resulting from a control ligation of recipient plasmid without ligase, which suggested the presence of a LacZ containing contaminating vector. Diagnostic digestion of ectopic blue colonies for both (Figure 4.4, Panel C, Lanes 5-7 and Figure 4.5, Panel B, Lanes 5-7) confirmed the presence of a 2.6 kb plasmid contaminant. The pHSG298 plasmid matched this digestion profile and was therefore suspected as the contaminant. The results of an initial replica plating investigation suggested that this plasmid was ampicillin resistant. However, further testing of contaminated colonies in liquid medium revealed that it was in fact sensitive to this antibiotic. Since pHSG298 does not carry AmpR, this strongly suggested it was the plasmid responsible. However, sequencing of pDNA isolated from an ectopic blue colony with primers expected to bind to pHSG298 was unsuccessful. The possibility of a procedural error having caused this seems unlikely, since all other sequencing reactions set up in parallel were successful. Additionally, both primers used had produced high quality sequence when used elsewhere in this study, demonstrating their functionality. It is therefore also possible that the contaminant was a different plasmid which by chance had the same digestion and resistance profile to pHSG298. As this issue had already caused delays, and since the desired modified pBLPrt GG and pCRPrtneo GG vectors had been recovered, further investigation of this contaminant was not carried out, and its identity remains unknown.

Another issue which complicated the construct production process was the generally poor performance of colony PCR. In many cases, these reactions resulted in multiple bands at unexpected sizes for the corresponding insert, suggesting widespread mispriming. It is likely that longer primers with higher melting temperatures would have been more specific, and therefore performed better. The approach taken in cases of unexpected colony PCR products was to perform diagnostic digestions on pDNA isolated from colonies representative of each banding pattern. The rationale behind this was that these likely represented colonies containing distinct products, although there was at least one example where this does not appear to have been the case: PCR screening of an ectopic blue colony and colony 49 when producing pCRPrtneo GG gave distinct banding patterns (Figure 4.4, Panel B, Lanes 2+5), but their plasmid digestion profiles were the same (Figure 4.4, Panel C, Lanes 5-7). While this approach was ultimately successful, excessive troubleshooting of colony PCR reactions, particularly in the early stages of the construct production process, contributed to delays.

4.3.2 Constructs Produced

While the assembly process took longer than expected, the full complement of constructs designed for the invader assay were produced. In all cases at least one colony was recovered containing a plasmid product with the desired insert. Sequencing of flanking regions confirmed that for all of these plasmids, the integration event had occurred at the intended site. Ideally it would have been possible to fully sequence all of the vectors intended for subsequent use in *M. maripaludis*. This is because their backbones contain features critical to function. For example, the HPT and NeoR genes in the transformation vectors pBLPrt/pCRPrtneo are required for the positive and negative selection steps responsible for the recovery of colonies carrying their integrated insert cargo following transformation (*19*), and loss or disruption of either would be expected to impede vector function. Diagnostic digestions on all *M. maripaludis* vectors had been conducted before and after modification, and with the notable exception of pBLPrt, all had resulted in the expected banding patterns. Based on these observations there is no reason to suspect that any of these plasmid backbones contain issues which might prohibit function, although diagnostic digestion would be unlikely to detect problems such as single nucleotide changes, or small insertions/deletions, which at critical locations could equally impede vector function.

The exception here was pBLPrt. The pBLPrt GG backbone fragment (Figure 4.5, Panel B, Lane 3) was approximately 1 kb smaller than the initial 9 kb pBLPrt vector which had been

used in this study, even though it would have been expected to have been the same size. However, the backbone fragment in diagnostic digestions of pDNA from five colonies following Golden Gate assembly into pBLPrt GG was again 9 kb (Figure 4.10, Panel B, Lane 7). While unlikely, it is possible that this could have been caused by two sequential plasmid recombination events: the first resulting in the loss and the second in the gain of 1 kb. A seemingly more probable explanation, however, is simply that the band for pBLPrt GG (Figure 4.5, Panel B, Lane 3) had run further on the gel than would have been expected for its size: (i.e. that it was in reality 9 kb). Despite this, the pBLPrt vector used here, and therefore the backbone fragment present in the pBLPrt GG colonies, was 600 bp larger than what would have been expected based on the plasmid sequence provided (James Chong, personal communication). This may have been caused by some form of recombination event on the plasmid backbone, after it had been sequenced but prior to use in this study. It is also possible that the size observed here reflects the true size of the original plasmid, and that the available sequence contains an error.

It was not possible to fully sequence the vectors produced here due to financial constraints, and it seems sensible to suggest that backbone sequencing only need be carried out in situations where vectors are shown to be dysfunctional. For example, if no colonies were recovered in a control transformation of wild type *M. maripaludis* cells with the invader plasmid pLW40 following plating on puromycin, it would be sensible at this point to sequence its PurR gene to check whether it contained any inactivating mutation(s), as would be expected in this scenario.

4.3.3 Golden Gate Constructs

Compared with the other cloning steps carried out in this section, particularly those involving vector dephosphorylation, Golden Gate assembly was highly efficient, as has been reported (*83*). Given the difficulties encountered during the production of the Golden Gate compatible transformation vectors using the single restriction enzyme cloning technique required for insertion in between the UPT flanks in pBLPrt (*19*), and at the NotI site in pCRPrtneo as used in pAW10 (*107*), it seems reasonable to speculate that producing all six versions of the assembled CRISPR/Cas9 construct in this manner would have been extremely challenging. While unlikely to have impacted the functionality of the CRISPR/Cas9 components, to allow fair comparison between each set of three test and control strains it was ensured that all CRISPR/Cas9 constructs were integrated in the same relative orientation to their respective homology arms in the transformation vectors. The

use of the modified transformation vectors meant that this could be achieved without the requirement for additional screening, since all inserts were assembled directionally based on the orientation of the dropout region. It is envisaged that the modified transformation vectors produced here should enable more efficient construct production and therefore more rapid *M. maripaludis* chromosomal modification at the UPT and MCM-C loci in future work. It should however be noted that use of pCRPrtneo GG to integrate material at the MCM-C site should be considered carefully, due to the increased growth rate phenotype associated with this integration (*107*), which may be inappropriate for some experiments.

4.3.4 Production of *M. maripaludis* Test Strains

An attempt was made to produce one complete set of test and control *M. maripaludis* strains by transforming S0001 cells with the three alternative assembled pCRPrtneo GG vectors. However, comparative levels of growth in mock compared with plasmid transformed cells subcultured in the presence of neomycin, which should have killed any cells not transformed with the pCRPrtneo backbone, suggested that this had been unsuccessful. One possible explanation for this is that the transformation itself had failed, resulting in no uptake of the desired plasmids, or so little that growth resulting from resistant colonies was undetectable compared with that resulting from background. The most likely explanation for this is that the cells in the cultures used were not growing rapidly enough to be transformed effectively. Transformation of cultures in early stationary phase (corresponding to an OD_{600} value of approximately 0.8) was found to give the greatest yields of stable transformants in *M. maripaludis* (*17*), and the version the protocol used here (see Chapter 2) states cultures to be transformed should have reached an OD_{600} of 0.7-1 following overnight growth. While two of the cultures transformed in this section were within the advised OD_{600} range, it had taken 58 hours for them to reach this point, so the cells were clearly growing more slowly than advised in the protocol. The same is also true for the two other cultures transformed, which even after 34 hours had not reached the minimum suggested OD_{600} value.

Another possibility is that some cells were successfully transformed, but the concentration of neomycin used was too low to enable efficient selection. The concentration of neomycin required to completely inhibit the growth of 10^7 *M. maripaludis* cells in a 5 ml culture was found to be 1 mg/ml (*20*). However, 0.5 mg/ml was used in this study because the higher concentration had previously been found to result in poor recovery of transformed cells (James Chong, personal communication). Neomycin acts by interfering with ribosome

function, causing mistranslation (*110*). APH3'II, the resistance gene used in pCRPrtneo (*19, 20, 111, 112*) imparts resistance by catalysing the addition of an inactivating phosphate group onto the antibiotic molecule (*110*). The previously observed poor transformation efficiency at 1 mg/ml of neomycin was therefore presumably because even cells carrying the resistance cassette were unable to effectively inactivate such high levels of antibiotic. However, it seems that in this study the lower concentration was not sufficient to result in the death of enough untransformed cells to allow growth of successful transformants to be detected against the presumably resultant high level of background, suggesting that an intermediate concentration would have been more appropriate. It should be noted that these two possible explanations need not be mutually exclusive: an inefficient transformation resulting in very few genuine neomycin resistant cells would be expected to be compounded by subsequent ineffective selection for that small population of desired cells.

While the production of the *M. maripaludis* test and control strains as attempted here was ultimately unsuccessful, if either or both of the possibilities discussed above were responsible it is anticipated that they could be overcome with some simple troubleshooting steps. While the growth rates of the *M. maripaludis* cultures maintained throughout this study fluctuated considerably, they were growing unusually slowly when this transformation was attempted: with more time it would have been possible to coincide a transformation with a period of more rapid growth. Additionally, the subculturing steps following the transformation could be conducted using a range of concentrations of neomycin intermediate between 0.5 and 1 mg/ml, which should permit more effective selection.

4.3.5 Chapter Summary

In summary while it was not possible to produce the test and control strains which would have permitted testing of the invader assay due to time constraints, the full suite of vectors thought to be sufficient for conducting this assay were produced. This should enable the project to easily be continued from the point reached at the end of this chapter.

Chapter 5: General Discussion

The object of this study was to improve the genetic tools available in *M. maripaludis* by adding the CRISPR/Cas9 system to the existing toolbox. The aim was therefore to provide an initial demonstration the functionality of CRISPR/Cas9 in this organism, using an experimental approach based on a plasmid invader assay (Figure 1.1). Implementing this assay had several steps. First, all required constructs and an appropriate assembly scheme had to be designed. Second, these constructs had to be assembled. Third, the relevant constructs would be used to produce the *M. maripaludis* test and control strains required to carry out the assay. Finally, the assay would be performed by transforming the invader plasmids into these strains, with lower survival in test vs control strains indicating functionality of the CRISPR/Cas9 system. Unfortunately, the construct design and assembly stages proved to be more challenging, and therefore time consuming, than initially anticipated. As a result, while an attempt was made, it was not possible to produce the *M. maripaludis* strains required for the invader assay, which could therefore not be carried out. Despite this, a full set of vectors thought to be sufficient for conducting the assay as designed in this study was produced, which should enable rapid completion were it to be continued.

5.1.1 Conducting the Invader Assay

Completing the project as designed here would require just two successive *M. maripaludis* transformations. The first would produce the *M. maripaludis* test and control strains required for the invader assay. For the S0001 background, this would involve transformation with the pCRPrtneo GG eGFP/pur/ctrl integrative plasmids, which carry the three alternative versions of the CRISPR/Cas9 assembly. The result would be a set of three *M. maripaludis* strains in this background containing chromosomally integrated CRISPR/Cas9 constructs which differ only in their gRNA component: two test strains which target the eGFP or puromycin genes of the invader plasmid, and a control strain which should not target the invader. The test and control strains in the Mm900 background would be produced in the same way, but instead using the vectors pBLPrt GG eGFP/pur/ctrl. With the required strains produced, the second transformation is essentially the actual invader assay, using the appropriate invader plasmid (pAW42 Mod for the S0001 background and pLW40 Mod for Mm900). In the first instance, the invader assay would be conducted by transforming each set of three *M. maripaludis* strains (grown to a similar OD_{600} value) with the same quantity of the appropriate plasmid invader, and then plating

the transformants on puromycin (to which the invader confers resistance). Since only cells carrying intact invader plasmid should be able to grow on this selection, control strain transformants should be recovered at transformation efficiency. Test strain cells should be recovered at transformation efficiency minis the proportion of transformed cells in which CRISPR/Cas9 activity resulted in loss of the invader plasmid, and a difference in recovery between the two would indicate whether the CRISPR/Cas9 system was functional.

5.1.2 Limitation of the Invader Assay as Designed in this Study

The invader assay assumes the number of test and control strain cells transformed with invader plasmid is equal, and as such any difference in plating efficiency is due solely to CRISPR/Cas9 mediated destruction of the invader plasmid in the test strains. The reliability of this output therefore depends on ensuring the number of cells transformed between cultures is as similar as possible, which is influenced by the amount of invader plasmid transformed, the number of cells used, and the transformation protocol itself. Ensuring equal amounts of invader plasmid is straightforward since it involves simply adding the same amount in each transformation. However, controlling the number of cells used with the system described above is more difficult. This is because *M. maripaludis* strains would have to be grown separately prior to transformation. While growth of cultures to a similar OD_{600} value should yield similar numbers of cells in each, it would clearly not be possible to ensure exactly equal numbers using this approach. For this reason it would be appropriate to use a plating method which provides an estimate of the number of viable cells which were present in each culture prior to transformation. This could be achieved by splitting each of the cultures following transformation and plating serial dilutions with and without selection (*12*). Equivalent numbers of cells at corresponding dilutions for each strain without selection would indicate an equivalent number of cells were present prior to transformation. However, while this gives an indication it still does not confirm the number of viable cells prior to transformation were exactly the same between strains.

In this sense the eGFP gRNA included in this study would be particularly useful. This is because *M. maripaludis* strains carrying this gRNA should not be able to target the unmodified invader plasmids pLW40/pAW42, since they do not carry the corresponding target site (*12, 24*). An *M. maripaludis* culture carrying gRNA-eGFP could be split prior to transformation, and half transformed with the appropriate modified invader plasmid (which it should target), and the other half transformed with the corresponding unmodified invader (which it should not target – effectively serving as the control). This would guarantee that

the same amount of *M. maripaludis* cells were present prior to transformation with the same amount of invader plasmid. However what this still does not guarantee is that the same amount of cells were actually successfully transformed. While use of the same transformation protocol would be expected to result in approximately equal transformation rates, the *M. maripaludis* transformation procedure is relatively complex, and involves several steps where samples must be handled separately (*17*). It is therefore possible that stochastic differences between samples could result in slightly different transformation rates, which could lead to false conclusion based on the invader assay output. For this reason it would be desirable to be able to measure the transformation efficiency of cultures transformed with invader plasmids prior to their CRISPR/Cas9 mediated destruction. However, this is not possible using this assay design.

The inability to decouple transformation efficiency from Cas9 mediated destruction, i.e. the fact that it would not be possible to directly measure what proportion of cells not recovered on selection plates following transformation was due to Cas9 mediated invader plasmid destruction, compared with cells which were simply not transformed, is probably the main weakness of the assay designed here. This issue would be unlikely to be problematic in a situation where the CRISPR/Cas9 system was able to destroy the invader plasmid highly efficiently. As an extreme example, an output resulting in recovery of lawns on control plates compared with no/very few colonies on test plates would clearly be indicative of functionality. However, if the CRISPR/Cas9 system were less efficient, small differences in transformation rates between cultures could mask its effects. For example, if Cas9 activity were only efficient enough to result in invader plasmid destruction in 5% of test strain cells, a 5% higher transformation efficiency in control cultures would result in equivalent recovery. For the same reason, a 5% reduced recovery in test compared with control strain transformations would have to be regarded with extreme caution. It should be noted that there is no reason to expect that this should be the case, since the primary design constraint followed throughout this project was to produce a system whereby CRISPR/Cas9 should have the greatest chance of giving a strong output in the invader assay as possible. Namely, the CRISPR/Cas9 components (including a codon optimised Cas9) were placed under the control of a strong promoter (*60*), which would be integrated onto the *M. maripaludis* chromosome at 30-55 copies per cell (*16*), and should have been present as a functional complex prior to challenge with a single instance of the invader plasmid on transformation.

5.1.3 Improving the Invader Assay: Decoupling

It is noteworthy that any experimental design which enables transformation of the invader plasmid to be decoupled from Cas9 mediated destruction would be expected to be less sensitive. This is because by definition the invader plasmid would have to be stably transformed into cells prior to targeting by the CRISPR/Cas9 system. It would therefore be present at multiple copies (*23*), which means multiple plasmids would have to be destroyed (in close succession) to give the desired output. Indeed, this presumed reduced sensitivity is the primary reason why such a version of the assay was not attempted in the first instance. However if the assay designed here gave a reasonable output, it could be modified for this purpose (i.e. to allow decoupling), which would be expected to provide a more scientifically robust demonstration of CRISPR/Cas9 functionality.

5.1.3.1 Inducible promoter

The simplest way this could be done would be to produce an *M. maripaludis* strain in which the Cas9 and/or gRNA components were placed under the control of an inducible promoter. This strain would then be transformed with the invader plasmid, and a pure culture generated by selecting for puromycin resistance. The pure culture would then be subcultured into medium supplemented with and without the inducer, grown, and then each culture split and plated with and without the selection. Since all cells should have initially contained the plasmid invader, the difference in relative recovery of colonies plated with selection compared to without between induced and uninduced cultures would give a measure of the proportion of cells in which the induction event lead to destruction of all the invader plasmids within them. In principle this would be possible using the nif promoter (*28, 32*). This promoter has been used to drive inducible gene expression in *M. maripaludis* by transitioning cells to medium containing NH_4 (in which the promoter is inactive) or to medium containing alanine as a sole nitrogen source (in which the promoter is active) (*36*). However, it should be noted that in addition to having to be able to drive high enough expression to destroy a greater number of plasmid invaders, protein expression from nif was found to be lower than from Phmv (*50*), suggesting it is weaker, and therefore might not be able to drive expression high enough to yield the test output. Leaky expression in uninduced conditions, which has been detected for nif (*35-37*), could also cause problems.

5.1.3.2 Two plasmid system

Another modified version of the invader assay which would allow transformation to be decoupled from Cas9 activity, while also not relying on a (potentially unsuitable) inducible

promoter would be a two plasmid system. This would involve the successive transformation of two plasmids, the invader and a "killer", each carrying a different selectable marker, into an *M. maripaludis* test strain which constitutively expresses Cas9. The first plasmid transformed into this strain, the invader, would contain a target site for a gRNA cassette carried on the second transformed plasmid, the killer. Following transformation with the invader plasmid, and establishment of a pure culture, the killer plasmid would be transformed. Expression of the gRNA cassette carried on the killer plasmid would allow Cas9 mediated destruction of the invader plasmid. Cultures transformed with the killer would then be split, with half plated on the selections corresponding to both the killer and invader plasmids, and the other half plated only on the selection corresponding to the killer. Colonies recovered on plates corresponding to the killer plasmid selection would indicate transformation efficiency, while those recovered on plates with both plasmid selections would reveal of those cells successfully transformed, what proportion underwent CRISPR mediated invader plasmid destruction.

Unfortunately a two plasmid system would not be possible using existing tools available in *M. maripaludis*. Firstly, both plasmids would need to have different replication origins, since use of the same would lead to plasmid incompatibility issues (*113*). However, all currently available *M. maripaludis* replicative plasmids are derived from the same parent (*18*), and therefore presumably possess the same origin. A two plasmid system would also require use of two different selections. While both puromycin and neomycin selection is possible in *M. maripaludis* (*20, 21*), cells grown in neomycin rapidly develop spontaneous resistance, and plating on this antibiotic is known to result in growth of satellite colonies (*20*), which would make it inappropriate for this application. For this reason a reliable second selection would first be required. The majority of bacterial antimicrobial agents are ineffective against archaea due to physiological differences between the two domains. For example, archaeal cell walls do not contain peptidoglycan, and its biosynthesis is a common target for bacterial antibiotics (*114*). A better potential source of selective markers might be auxotrophic strains. A good candidate in *M. maripaludis* would be tryptophan. Transposon insertions into various regions of its operon was used to generate corresponding auxotrophs in *M. maripaludis* (*115*), and deletion of the full locus followed by complementation on a plasmid vector could potentially serve as a more robust second selection strategy.

5.1.4 Possible Improvements to Vectors

The modified *M. maripaludis* transformation vectors produced in this study, pCRPrtneo GG and pBLPrt GG, should be to some extent useful in future work since material can be assembled into them using the highly efficient Golden Gate system (*83*). However, these vectors are restricted to two *M. maripaludis* genomic integration sites. A more versatile *M. maripaludis* transformation vector could be produced by incorporating a Golden Gate dropout region, such as that used in this study, into the pCRPrtneo vector backbone. Since pCRPrtneo can be repurposed to introduce material into any chromosomal location in *M. maripaludis* by introducing appropriate flanking regions (*19*), the Golden Gate system could be used to incorporate both the material to be delivered and the flanks in this modified vector. The primary reason that such a vector was not produced in this study is that introducing two additional inserts would have required the modification of two of the existing BpiI sites, either on the outer-most insert fragments or within the dropout region itself. In principle this would have been relatively straightforward using PCR methods, but it would also have been more time consuming. Additionally, it was envisaged that keeping the assembly protocol as simple as possible (i.e. with fewer inserts) would be more likely to result in successful assembly, since it was not known how well the *M. maripaludis* transformation vectors, which had not to our knowledge previously been used with the Golden Gate system, would perform. Based on the performance of the pCRPrtneo GG vector produced here, there is no reason to expect that it could not handle additional inserts. Producing a more versatile pCRPrtneo could easily be achieved using material produced in this study: the pDVA_BC dropout region could be amplified while adding appropriate restriction sites from plasmid DNA, and deposited into the pCRPrtneo vector backbone. A similar modification could also be made to the pLW40/pAW42 vectors, making them both compatible with Golden Gate assembly and blue/white screening, perhaps using the ClaI site used in this study.

5.1.5 Removal of CRISPR/Cas9 from *M. maripaludis*

Methanogens such as *M. maripaludis* have multiple potential biotechnological applications (*116*), perhaps most obviously in the production of biofuel (*117*). Due to perceived risks in terms of biosafety, such as the potential transfer of modified DNA to wild-type organisms and its possible environmental consequences (*118*), a system enabling the efficient removal of the CRISPR/Cas9 components from *M. maripaludis* would be desirable, for example to meet potential future legislative requirements. The problem with removing a CRISPR/Cas9

assembly already integrated into the *M. maripaludis* chromosome, as intended in the invader assay designed here, is that all copies would have to be removed simultaneously, which would likely be challenging. One alternative option would be to place the CRISPR/Cas9 assembly onto an integrative vector backbone, which is looped-out following negative selection. The obvious drawback to this approach is that it would make these vectors considerably larger, which may lower transformation efficiency (*12*). Additionally, the CRISPR/Cas9 construct would have to have achieved its desired effect during the short time window in which it was present on the chromosome, (i.e. following the loop-in but prior to the loop-out event after transformation), and as such would need to be highly efficient. Once the CRISPR/Cas9 system is shown to be functional in *M. maripaludis* when expressed from the chromosome, an obvious progression would be to demonstrate its functionality when expressed from a replicative plasmid. With this achieved, a plasmid curing system in which the replicative plasmid carries a negative selection marker could be used to select for cells which had lost this plasmid, similar to that described by Nayak and Metcalf (2017) (*48*). At least two appropriate markers, HPT or UPT, are available in *M. maripaludis* (*19*).

5.1.6 Possible Applications for the CRISPR/Cas9 system in *M. maripaludis*

5.1.6.1 Improvements to *M. maripaludis* transformation

DSBs, as induced by the CRISPR/Cas9 system, dramatically increase the rate of HR at the target site, provided a suitable repair template is available. Perhaps unsurprisingly, the CRISPR/Cas9 system has therefore been broadly implemented to improve chromosomal modification rates in organisms in which this had previously been difficult, particularly eukaryotes (*119*). *M. maripaludis* naturally recombines with introduced material homologous to a given genomic target site (*22*), and development of a PEG-based transformation protocol substantially increased the rates at which genomic modifications could be recovered (*17*). Coupled with the established positive/negative selection system, strains carrying multiple chromosomal manipulations can be generated (*19*). However, even with these developments, chromosomal modification in *M. maripaludis* remains somewhat cumbersome, and further improvements to the existing protocol would be useful. The possibility of achieving modifications via simply inducing a cut at a given target site while providing a suitable repair template is not straightforward in *M. maripaludis* due to its polyploidy. Achieving chromosomal gene deletions/replacements, or stably introducing material onto the chromosome requires that all chromosomes undergo the desired modification. This necessitates the use of a positive selection marker to ensure the complete

propagation of the modification onto all copies of the chromosome, which is thought to be facilitated by gene conversion (*16*). For this reason, use of the CRISPR/Cas9 system to drive chromosomal modification in the absence of selection would not be expected to be appropriate for these applications, since even if highly active it seems extremely unlikely that it would be capable of modifying all chromosomes at once.

However, the CRISPR/Cas9 system might be suitable for augmentation of the existing positive/negative based modification protocol (*19*). Induction of a DSB at the intended chromosomal target site would be expected to boost the frequency of the HR event required for the initial loop-in of the transformation plasmid and its cargo, while concurrently increasing the frequency of gene conversion events of the looped-in plasmid to other as yet unmodified chromosomes. DSBs at this site would also be expected to increase the frequency of loop-out events, and their propagation via gene conversion, once the positive selection initially used to select for looped-in plasmids is withdrawn. As such the CRISPR/Cas9 system could potentially be used to boost the rates of both recombination events required for successful chromosomal manipulation, and therefore improve the efficiency of the process in general. It should be noted that because a portion of the genomic target locus is replaced with the plasmid cargo using the current positive/negative selection procedure (*19*), as long as the CRISPR/Cas9 target site was selected within this region, its loss following the intended modification event would prohibit further Cas9 mediated DSB induction.

Such a system could theoretically be achieved by placing both the Cas9 and gRNA components on the integrative transformation vector backbone. This approach would have the advantage that both components would be removed on successful modification. However, this would also increase the size of the plasmid, which as discussed previously could have a negative effect on transformation efficiency (*12*). Also, since both components would theoretically only initially be present at a single copy following delivery of the plasmid to the cell, expression of both would initially be low. It should be noted, however, that any increase in HR at the target site would be expected to increase modification efficiency, and following the initial loop in event and subsequent gene conversion the copy number of both components would increase, in turn increasing expression, creating a positive feedback loop. Another option would be to create an *M. maripaludis* strain with a constitutively chromosomally expressed Cas9. This component would therefore not be required on the transformation vector backbone, which would both reduce its size while

also providing a large initial amount of Cas9 protein, which means the Cas9 mediated augmentation of the initial loop-in event would only be limited by expression of the gRNA carried on the plasmid. While use of either of these system would require the integration of a different gRNA into the transformation plasmid for each genomic target locus, PCR protocols have been developed for other systems which enable the 20 nt targeting sequence within the gRNA to be easily exchanged (*67*). Alternatively, since the full gRNA is only 100 bp in length (*120*) it could be ordered relatively cheaply as a synthetic DNA fragment, which if coupled with Golden Gate compatible plasmids suggested above, would allow the rapid production of transformation vectors carrying alternative gRNA target site specificities.

5.1.6.2 Improvements to *M. maripaludis* promoters

To our knowledge just four promoters (*17, 28, 30, 31*) have been used for gene expression purposes in *M. maripaludis*, providing a clear incentive for the development of more options. One method which can be used to achieve this is to obtain additional promoters from closely related organisms, as illustrated for example by the use of *M. voltae* promoters in *M. maripaludis* (*17, 30*). Another option is to modulate the function of existing promoters, which could potentially be achieved in *M. maripaludis* using a variant of the CRISPR/Cas9 system known as CRISPR interference (CRISPRi). CRISPRi employs a catalytically inactive Cas9 (dCas9) which is able to target but not cleave DNA. In its simplest sense, dCas9 can repress target genes when delivered to a promoter through steric hindrance of RNA polymerase. In bacteria, dCas9 targeted to different regions of the same promoter was found to result in different levels of transcriptional repression (*46*). A similar system could be developed in *M. maripaludis*: a series of strains carrying constitutively expressed dCas9 and gRNAs targeting distinct regions of a known promoter (for example Phmv) could be developed. Transforming a gene of interest fused to this promoter into these different strains would result in distinct levels of repression, allowing controlled expression (Figure 5.1).

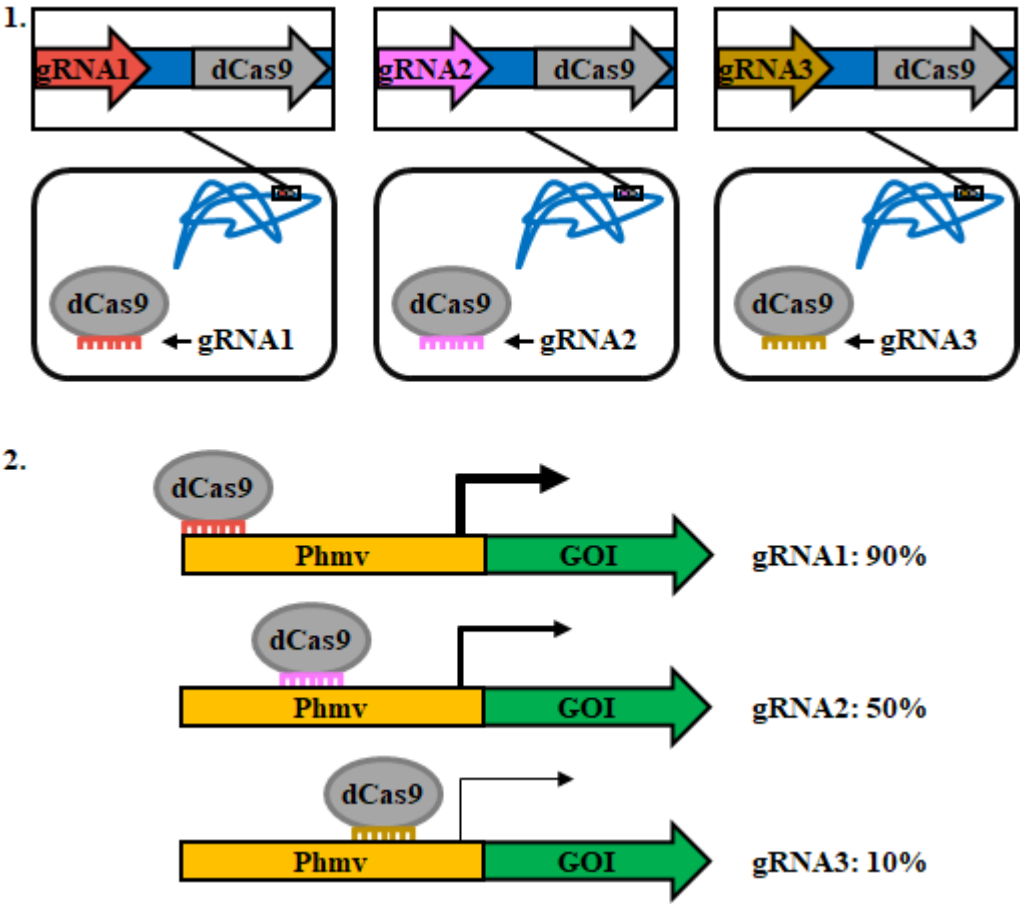

Figure 5.1: Control of gene expression using CRISPRi.
1: dCas9 and a series of different gRNAs under the control of the constitutive Pmcr are stably transformed onto the *M. maripaludis* chromosome (blue line) producing a set of distinct strains. **2:** The previously characterised gRNAs each target a different portion of Phmv, resulting in distinct but stable levels of repression. A gene of interest (GOI) placed under the control of Phmv is transformed into the desired strain depending on the required level of expression. Transformation into the strain carrying gRNA1 yields 90% of the expression level of the standard Phmv promoter, while transformation into the strains carrying gRNA2 and 3 give 50% and 10% expression, respectively.

The obvious limitation of this approach is that the maximal possible level of expression would be that of the native promoter, yet higher levels of expression may be desirable for some applications. dCas9 can also be used to increase gene expression by fusing it to a transcriptional activator, and then using its targeting capabilities to deliver this to a promoter in a system termed CRISPR activation (CRISPRa) (*121*). A transcriptional activator has recently been discovered in *M. maripaludis* (*122*), which means such a system might be possible in this organism (Figure 5.2).

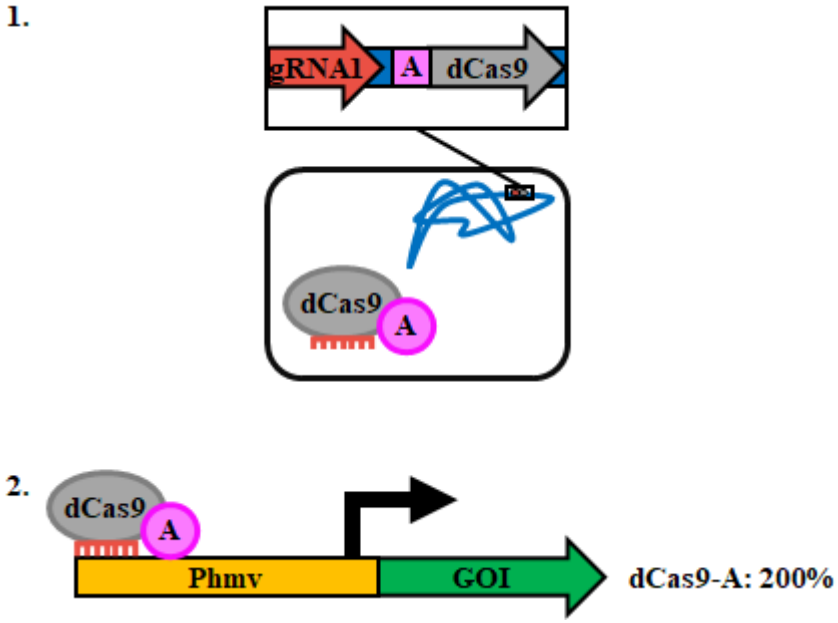

Figure 5.2: Control of gene expression using CRISPRa.
1: dCas9 fused to a transcriptional activator and a gRNAs under the control of the constitutive Pmcr are stably transformed onto the *M. maripaludis* chromosome (blue line) producing an overexpression strain. **2:** The gRNA targets dCas9 and the attached activator to Phmv, resulting in increased levels of expression. A gene of interest (GOI) placed under the control of Phmv is transformed into this strain resulting in 200% of the expression level of the standard Phmv promoter.

5.1.6.3 Unbiased assessment of off-targeting in organisms with small genomes

Off-targeting by the CRISPR/Cas9 system is thought to be low in organisms such as bacteria due to the generally small size of their genomes (*49*, *102*). While this makes logical sense, it is possible that the extent to which off-targeting actually takes place in these organisms has been underestimated. Self-targeting of the CRISPR/Cas9 system to the chromosome has been shown to result in widespread lethality in several bacterial species (*123-125*). It follows that in the case of ectopic targeting, the vast majority of cells in which such events had taken place would not be recovered, effectively destroying the evidence. This may have led to a general overestimation of CRISPR/Cas9 specificity in organisms with small genomes, such as bacteria. Polyploid organisms carry additional copies of the chromosome, which act both as a back-up system and a source of repair templates for HR, making them more resistant to DNA damage (*126*) such as the DSBs induced by Cas9. Since *M. maripaludis* is highly polyploid (*16*) it could present an opportunity for the unbiased assessment of CRISPR/Cas9 specificity in an organism with a small genome (*14*). To achieve this, a strain capable of carrying out NHEJ, and therefore producing detectable

indels at Cas9 cleavage sites, could first be produced by incorporating the Ku and LigD genes onto the chromosome, as was achieved in *M. acetivorans* (*48*). Various gRNAs could then be tested for their ability to cleave throughout the genome using an unbiased approach such as whole genome sequencing (*127*). It should be noted that while this approach could also theoretically be applied in monoploid models such as *E. coli*, cells which undergo perturbations in essential regions of the genome (such as the indels induced by NHEJ) would not be expected to be recovered (*15*), potentially leading to an underestimation of off-targeting. This situation would be much less likely in *M. maripaludis* by virtue of its polyploidy (*16*), since all copies of an essential gene would theoretically have to be inactivated to result in cell death.

5.1.7 Summary

As the only known class of organisms capable of producing methane, methanogens (*7*) such as *M. maripaludis* have clear potential for biotechnological use (*116*). While our understanding of the genetics and biochemical mechanisms employed by these organisms to generate methane has improved (*128*), our ability to utilise this information is likely to be limited by the genetic tools available. While the development of such tools has generally lagged behind in the archaea (*2*), the recent demonstration of the functionality of the CRISPR/Cas9 system in *M. acetivorans* (*48*) is encouraging. As one of the most well developed methanogen models (*12*), *M. maripaludis* is an excellent candidate for receipt of up-to-date genetic tools. Due to time constraints it was not possible to demonstrate the functionality of the CRISPR/Cas9 system in *M. maripaludis* in this project. However, it is hoped that continuation of this work, or work by other groups, should add this highly versatile genetic tool to the *M. maripaludis* toolbox in the near future.

Appendix

Table A1: Primers Details.
All primers excluding those marked * were designed in this study.

Name	Sequence (5'-3')
pUC18 CP F1	ACCTCTGACACATGCAG
pUC18 CP R1	GACAGGTATCCGGTAAGC
pCRP CP F1	TCCCCTTCGAGCAAGTAC
pCRP CP R1	GCCTCTTCGCTATTACGC
pCRP CP R2	GCGAGAAAGGAAGGGAAG
pCRP CP R3	CGTCAAAGGGCGAAAAAC
pBLP CP F1	TCCAGAGATTTCCAACGC
pBLP CP R1	TTGGGAGGACTTACAACCG
pBLP CP R2	GAATCAGGAATGGCAGTTGG
M13 F*	GTAAAACGACGGCCAGT
M13 R*	GGAAACAGCTATGACCATG
PHSG R1	GGAATTGTGAGCGGATAAC
pLW40 CP F1	CCTGCTTGTAAACTTTGCA
pAW42 CP F1	TGCTCTTCTTTCTTTCATGC
pAW42 CP R1	CTTCGATGTGATGGTGATG
Dropout Seq F1	GAATTAATCTCGGGAATGGG
Dropout Seq F2	CTGTTTTCACCTATGGCG
Dropout Seq F3	CGCGCCATAAACTTATTCAC
Frag2 Seq F1	GCGTTGGATGGGCTGT
Frag2 Seq F1 RC	ACAGCCCATCCAACGC
Frag2 Seq F2	AGGCGATCTTAATCCTGAC
Frag2 Seq F2 RC	GTCAGGATTAAGATCGCCT
Frag2 Seq F3	TGGGGGTGCATCTCAAG
Frag2 Seq F4	CAGCTTTCTTGTCTGGCG
Frag3 Seq F1	CGCAGGTATCAGGACAAG
Frag3 Seq F2	ACTAAAGCAGAACGGGGA
Frag3 Seq F3	GTGCTTTCAATGCCACAG
Frag3 Seq F4	TGTTGAACAGCACAAACATTAT

Table A2: Plasmids produced in this study.

Name	Parent Plasmid	Insert Details	Purpose
pDropout Mod/pUC18	pUC18	Dropout Mod gBlock cloned into pUC18 MCS with BamHI/SacI	Generating pCRPrtneo GG and pBLPrt GG
pCRPrtneo GG	pCRPrtneo	NotI flanked region of Dropout Mod gBlock subcloned into pCRPrtneo NotI site	Generating pCRPrtneo GG eGFP/pur/ctrl
pBLPrt GG	pBLPrt	AscI flanked region of Dropout Mod gBlock subcloned into pBLPrt AscI site	Generating pBLPrt GG eGFP/pur/ctrl
pFragment 1 eGFP/ pHSG298	pHSG298	Fragment 1 eGFP gBlock cloned into pHSG298 MCS with BamHI/SacI	Generating pCRPrtneo GG eGFP and pBLPrt GG eGFP
pFragment 1 pur/ pHSG298	pHSG298	Fragment 1 pur gBlock cloned into pHSG298 MCS with BamHI/SacI	Generating pCRPrtneo GG pur and pBLPrt GG pur
pFragment 1 ctrl/ pHSG298	pHSG298	Fragment 1 ctrl gBlock cloned into pHSG298 MCS with BamHI/SacI	Generating pCRPrtneo GG ctrl and pBLPrt GG ctrl
pFragment 2/pHSG298	pHSG298	Fragment 2 gBlock cloned into pHSG298 MCS with BamHI/SacI	Generating pCRPrtneo GG eGFP/pur/ctrl and pBLPrt GG eGFP/pur/ctrl

pFragment 3/pHSG298	pHSG298	Fragment 3 gBlock cloned into pHSG298 MCS with BamHI/SacI	Generating pCRPrtneo GG eGFP/pur/ctrl and pBLPrt GG eGFP/pur/ctrl
pCRPrtneo GG eGFP	pCRPrtneo GG	Fragments 1-3 (Fragment 1 eGFP) assembled by Golden Gate into pCRPrtneo GG BpiI sites	Introducing eGFP targeting CRISPR/Cas assembly at *M. maripaludis* MCM-C locus
pCRPrtneo GG pur	pCRPrtneo GG	Fragments 1-3 (Fragment 1 pur) assembled by Golden Gate into pCRPrtneo GG BpiI sites	Introducing puromycin targeting CRISPR/Cas assembly at *M. maripaludis* MCM-C locus
pCRPrtneo GG ctrl	pCRPrtneo GG	Fragments 1-3 (Fragment 1 ctrl) assembled by Golden Gate into pCRPrtneo GG BpiI sites	Introducing control CRISPR/Cas assembly at *M. maripaludis* MCM-C locus
pBLPrt GG eGFP	pBLPrt GG	Fragments 1-3 (Fragment 1 eGFP) assembled by Golden Gate into pBLPrt GG BpiI sites	Introducing eGFP targeting CRISPR/Cas assembly at *M. maripaludis* UPT locus
pBLPrt GG pur	pBLPrt GG	Fragments 1-3 (Fragment 1 pur) assembled by Golden Gate into pBLPrt GG BpiI sites	Introducing puromycin targeting CRISPR/Cas assembly at *M. maripaludis* UPT locus

pBLPrt GG ctrl	pBLPrt GG	Fragments 1-3 (Fragment 1 ctrl) assembled by Golden Gate into pBLPrt GG BpiI sites	Introducing control CRISPR/Cas assembly at *M. maripaludis* UPT locus
peGFP Mod/pUC18	pUC18	eGFP Mod gBlock cloned into pUC18 MCS with BamHI/SacI	Generating pLW40 Mod and pAW42 Mod
pLW40 Mod	pLW40	ClaI flanked region of eGFP Mod gBlock subcloned into pLW40 ClaI site	Invader plasmid targetable by *M. maripaludis* carrying eGFP CRISPR/Cas assembly
pAW42 Mod	pAW42	ClaI flanked region of eGFP Mod gBlock subcloned into pAW42 ClaI site	Invader plasmid targetable by *M. maripaludis* S0001 strain carrying eGFP CRISPR/Cas assembly

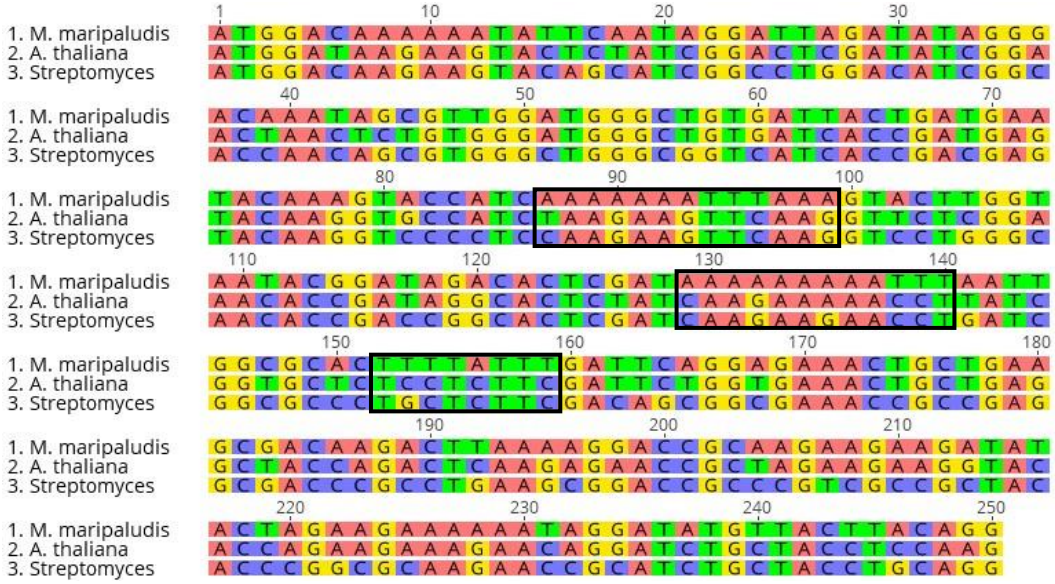

Figure A1: Nucleotide alignment of Cas9 codon optimised to *M. maripaludis*, *A. thaliana* and *Streptomyces*.
Labels on the far left indicate the identity of each sequence. Consensus positions are indicated by numbers above the alignment. Examples of regions which appear more repetitive in the *M. maripaludis* codon optimisation are highlighted by boxes. Only the first 250 bp of the alignment is shown for clarity.

ATGGACAAAAAATATTCAATAGGATTAGATATAGGGACAAATAGCGTTGGAT
GGGCTGTGATTACTGATGAATACAAAGTACCATCAAAAAAATTTAAAGTACT
TGGTAATACGGATAGACACTCGATAAAAAAAAATTTAATTGGCGCACTTTTA
TTTGATTCAGGAGAAACTGCTGAAGCGACAAGACTTAAAAGGACCGCAAGAA
GAAGATATACTAGAAGAAAAAATAGGATATGTTACTTACAGGAAATCTTTTC
TAATGAAATGGCTAAAGTAGATGATTCATTCTTTCACAGACTCGAAGAATCTT
TTCTGGTAGAAGAAGATAAAAAACATGAAAGGCACCCTATTTTCGGAAATAT
AGTAGATGAAGTTGCCTATCACGAAAAATATCCGACTATTTATCACTTACGAA
AAAAATTAGTTGATTCAACCGACAAAGCAGATTTAAGACTTATATATTTAGC
ATTAGCTCATATGATAAAATTCCGAGGACATTTTTTGATTGAAGGCGATCTTA
ATCCTGACAACAGCGATGTTGACAAACTTTTTATACAGTTAGTACAAACTTAC
AATCAGCTCTTTGAAGAAAACCCAATCAATGCTTCAGGAGTTGATGCAAAAG
CAATTCTTAGCGCAAGATTATCGAAATCTAGGAGGTTAGAGAACCTCATCGC
ACAGCTACCTGGAGAAAAAAAAAACGGTTTATTTGGAAACTTGATTGCATTG
AGTTTAGGGTTAACCCCAAATTTTAAATCAAATTTTGATTTAGCGGAGGATGC
AAAACTCCAGTTGAGCAAAGATACATACGACGACGATTTAGACAACCTTCTT
GCTCAAATAGGAGACCAATACGCAGACCTTTTTCTCGCTGCAAAAAACCTAA
GTGACGCAATTTTACTTTCAGATATACTCCGGGTAAATACTGAAATAACGAA
AGCACCTTTATCCGCATCAATGATCAAAAGATACGACGAACACCATCAGGAC
CTTACTTTATTAAAAGCGTTAGTTAGACAACAACTTCCCGAAAAATACAAAG
AAATATTCTTTGATCAGTCGAAAAACGGATATGCTGGATACATCGATGGGGG
TGCATCTCAAGAAGAATTTTATAAGTTCATTAAACCCATACTTGAAAAAATGG
ATGGAACAGAAGAACTTTTGGTGAAGTTAAATAGGGAAGATTTACTGAGAAA
GCAAAGAACTTTTGATAACGGATCAATACCGCACCAAATTCATCTTGGTGAA
CTCCACGCGATTCTTAGGCGTCAAGAAGATTTTTACCCATTTTTAAAGGATAA

TAGAGAGAAAATCGAAAAAATTTTAACATTTAGAATACCGTACTACGTTGGA
CCATTGGCCAGGGGTAATAGCAGATTTGCTTGGATGACGAGAAAATCTGAAG
AAACAATTACTCCTTGGAATTTTGAAGAAGTTGTCGATAAAGGCGCATCCGC
ACAGTCTTTTATTGAAAGAATGACAAATTTTGATAAAAACTTGCCAAACGAA
AAAGTTTTACCGAAGCATTCCCTATTATATGAATATTTTACTGTTTATAACGA
ATTAACTAAAGTCAAATACGTTACGGAAGGAATGAGAAAACCAGCTTTCTTG
TCTGGCGAACAGAAAAAAGCAATAGTTGATTTATTGTTTAAAACAAATAGAA
AGGTCACTGTTAAACAACTGAAAGAAGATTATTTTAAAAAAATTGAATGCTT
TGATTCAGTGGAAATATCAGGAGTAGAAGACCGATTTAATGCTTCATTGGGT
ACCTATCACGATCTGCTGAAAATTATTAAGGATAAAGATTTCTTAGACAATGA
AGAAAACGAAGACATTTTAGAAGATATTGTATTAACCTTAACGCTTTTTGAAG
ATAGAGAAATGATTGAAGAGAGGTTAAAAACATACGCACATTTATTCGATGA
CAAAGTTATGAAACAGCTAAAACGGCGCAGGTATACAGGATGGGGTCGGCTA
TCTAGAAAACTCATTAATGGAATTCGAGATAAACAATCGGGAAAAACTATTC
TAGACTTTCTAAAATCAGATGGTTTTGCAAATAGAAACTTTATGCAATTAATC
CATGACGATTCTCTTACATTCAAAGAGGACATTCAAAAAGCGCAGGTATCAG
GACAAGGTGATAGCCTACATGAACACATAGCAAATTTGGCAGGCTCACCTGC
AATAAAAAAAGGAATCCTACAGACCGTTAAAGTTGTAGATGAATTAGTAAAG
GTTATGGGAAGACACAAACCAGAAAATATCGTCATTGAAATGGCTAGAGAGA
ACCAAACAACGCAAAAAGGACAGAAAAACTCCAGGGAACGAATGAAAAGGA
TAGAAGAAGGAATAAAAGAACTCGGCTCTCAAATTTTAAAGGAACACCCAGT
TGAGAATACTCAGCTACAAAATGAGAAACTTTATTTATACTACCTACAGAAT
GGTCGAGATATGTATGTTGACCAAGAATTAGATATTAACAGATTAAGTGATT
ACGACGTAGACCACATAGTACCACAGAGTTTTCTGAAAGATGATAGTATTGA
TAATAAGGTTCTAACAAGGTCAGATAAAAACAGAGGAAAGTCAGATAATGTA
CCTAGTGAAGAAGTGGTTAAAAAAATGAAAAATTACTGGCGTCAGCTTTTAA
ATGCGAAATTAATTACTCAAAGAAAATTCGATAATCTTACTAAAGCAGAACG
GGGAGGCCTCTCAGAACTAGACAAAGCAGGATTTATAAAAAGACAGCTAGTT
GAAACGAGACAGATTACGAAACATGTAGCCCAGATCTTGGATTCAAGGATGA
ACACCAAATATGATGAGAATGATAAATTAATAAGGGAAGTAAAAGTTATAAC
CCTCAAATCTAAGCTTGTATCGGATTTCAGAAAGGACTTTCAATTTTACAAAG
TTAGGGAAATTAATAATTACCACCACGCACATGATGCATATCTTAATGCAGTA
GTAGGGACTGCATTAATAAAGAAATATCCCAAACTAGAATCAGAGTTTGTAT
ATGGCGACTACAAAGTATACGATGTTAGAAAAATGATTGCTAAGTCCGAACA
AGAAATCGGGAAAGCAACTGCCAAATATTTTTTTTACAGCAACATTATGAATT
TCTTTAAAACAGAAATTACTCTTGCAAACGGTGAGATCCGGAAAAGACCTCT
TATTGAAACAAATGGGGAAACTGGAGAAATTGTTTGGGATAAGGGACGTGAT
TTTGCAACAGTTCGTAAAGTGCTTTCAATGCCACAGGTAAACATTGTAAAAA
AAACCGAAGTTCAAACAGGTGGCTTTTCAAAAGAATCTATATTACCTAAAAG
AAATAGCGATAAACTGATCGCTAGGAAAAAAGATTGGGACCCAAAAAAATA
TGGTGGGTTTGATAGTCCAACAGTTGCTTACTCTGTTCTTGTCGTTGCAAAAG
TTGAAAAAGGAAAAAGTAAGAAATTAAAATCAGTTAAAGAGCTTCTGGGGAT
TACAATTATGGAACGATCATCTTTTGAGAAAAATCCAATTGATTTCTTGGAAG
CCAAAGGATATAAAGAAGTTAAAAAAGACTTGATTATCAAATTGCCAAAATA
CAGTTTATTTGAACTAGAAAATGGTCGGAAACGCATGTTGGCAAGTGCAGGA
GAACTTCAAAAAGGTAACGAATTAGCATTACCTTCAAAATATGTAAATTTTCT
TTATCTTGCAAGTCATTACGAAAAATTAAAAGGATCACCTGAAGACAATGAA
CAGAAACAGCTATTTGTTGAACAGCACAAACATTATTTAGATGAAATTATTG
AACAGATTTCCGAATTCTCTAAAAGAGTAATCTTAGCAGATGCTAATTTGGAT
AAGGTGCTGAGCGCATACAACAAACACAGAGATAAACCAATTCGAGAACAA
GCAGAAAATATTATCCATTTATTTACATTAACAAATCTTGGTGCACCTGCAGC
TTTTAAATATTTTGATACTACTATCGATAGAAAAAGATATACAAGTACCAAAG

AAGTTTTGGATGCAACATTAATTCACCAATCAATTACAGGACTTTATGAAACA
AGAATAGATCTAAGTCAGCTTGGTGGAGACTAA

Figure A2: Mmp-Cas9 nucleotide sequence.
The Cas9 sequence codon optimised to *M. maripaludis* used in this study.

References in the style of the journal *Science*

Reference List

1. H. Atomi, T. Imanaka, T. Fukui, Overview of the genetic tools in the archaea. *Front Microbiol* **3**, 337 (2012).
2. J. A. Farkas, J. W. Picking, T. J. Santangelo, Genetic techniques for the archaea. *Annu Rev Genet* **47**, 539-561 (2013).
3. A. Kloda, B. Martinac, Common evolutionary origins of mechanosensitive ion channels in archaea, bacteria and cell-walled eukarya. *Archaea* **1**, 35-44 (2002).
4. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090 (1977).
5. R. Cavicchioli, Archaea - timeline of the third domain. *Nat Rev Microbiol* **9**, 51-61 (2011).
6. M. Rother, W. W. Metcalf, Genetic technologies for archaea. *Curr Opin Microbiol* **8**, 745-751 (2005).
7. E. Bapteste, C. Brochier, Y. Boucher, Higher-level classification of the archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353-363 (2005).
8. J. A. Leigh, S. V. Albers, H. Atomi, T. Allers, Model organisms for genetics in the domain archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev* **35**, 577-608 (2011).
9. I. Ferrera, O. Sánchez, Insights into microbial diversity in wastewater treatment systems: how far have we come? *Biotechnol Adv* **34**, 790-802 (2016).
10. B. St-Pierre, L. M. Cersosimo, S. L. Ishaq, A. D. G. Wright, Toward the identification of methanogenic archaeal groups as targets of methane mitigation in livestock animals. *Front Microbiol* **6**, 10 (2015).
11. N. Raizada, V. Sonakya, V. Anand, V. C. Kalia, Waste management and production of future fuels. *JSIR* **61**, 184-207 (2002).
12. A. D. Walters, S. E. Smith, J. P. J. Chong, Shuttle vector system for *Methanococcus maripaludis* with improved transformation efficiency. *Appl Environ Microbiol* **77**, 2549-2551 (2011).
13. W. J. Jones, M. J. B. Paynter, R. Gupta, Characterization of *Methanococcus maripaludis* sp. nov., a new methanogen isolated from salt marsh sediment. *Arch Microbiol* **135**, 91-97 (1983).

14. E. L. Hendrickson *et al.*, Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J Bacteriol* **186**, 6956-6969 (2004).
15. F. Sarmiento, J. Mrazek, W. B. Whitman, Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A* **110**, 4726-4731 (2013).
16. C. Hildenbrand, T. Stock, C. Lange, M. Rother, J. Soppa, Genome copy numbers and gene conversion in methanogenic archaea. *J Bacteriol* **193**, 734-743 (2011).
17. D. L. Tumbula, R. A. Makula, W. B. Whitman, Transformation of *Methanococcus maripaludis* and identification of a Pst I-like restriction system. *FEMS Microbiol Lett* **121**, 309-314 (1994).
18. D. L. Tumbula, T. L. Bowen, W. B. Whitman, Characterization of pURB500 from the archaeon *Methanococcus maripaludis* and construction of a shuttle vector. *J Bacteriol* **179**, 2976-2986 (1997).
19. B. C. Moore, J. A. Leigh, Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J Bacteriol* **187**, 972-979 (2005).
20. J. L. Argyle, D. L. Tumbula, J. A. Leigh, Neomycin resistance as a selectable marker in *Methanococcus maripaludis*. *Appl Environ Microbiol* **62**, 4233-4237 (1996).
21. P. Gernhardt, O. Possot, M. Foglino, L. Sibold, A. Klein, Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. *Mol Gen Genet* **221**, 273-279 (1990).
22. K. A. Sandbeck, J. A. Leigh, Recovery of an integration shuttle vector from tandem repeats in *Methanococcus maripaludis*. *Appl Environ Microbiol* **57**, 2762-2763 (1991).
23. A. G. Wood, W. B. Whitman, J. Konisky, A newly-isolated marine methanogen harbors a small cryptic plasmid. *Arch Microbiol* **142**, 259-261 (1985).
24. J. A. Dodsworth, J. A. Leigh, Regulation of nitrogenase by 2-oxoglutarate-reversible, direct binding of a PII-like nitrogen sensor protein to dinitrogenase. *Proc Natl Acad Sci U S A* **103**, 9779-9784 (2006).
25. R. Heim, D. C. Prasher, R. Y. Tsien, Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* **91**, 12501-12504 (1994).

26. J. P. Horwitz *et al.*, Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-Indolyl-β-D-glycopyranosides. *J Med Chem* **7**, 574-575 (1964).
27. T. J. Lie, J. A. Leigh, Genetic screen for regulatory mutations in *Methanococcus maripaludis* and its use in identification of induction-deficient mutants of the eulyarchaeal repressor NrpR. *Appl Environ Microbiol* **73**, 6595-6600 (2007).
28. T. J. Lie, J. A. Leigh, Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source. *J Bacteriol* **184**, 5301-5306 (2002).
29. R. Cohen-Kupiec, C. Blank, J. A. Leigh, Transcriptional regulation in archaea: in vivo demonstration of a repressor binding site in a methanogen. *Proc Natl Acad Sci U S A* **94**, 1316-1320 (1997).
30. W. L. Gardner, W. B. Whitman, Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and beta-galactosidase. *Genetics* **152**, 1439-1447 (1999).
31. K. C. Costa, T. J. Lie, M. A. Jacobs, J. A. Leigh, H2-independent growth of the hydrogenotrophic methanogen *Methanococcus maripaludis*. *MBio* **4**, (2013).
32. T. J. Lie, G. E. Wood, J. A. Leigh, Regulation of nif expression in *Methanococcus maripaludis*: roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. *J Biol Chem* **280**, 5236-5241 (2005).
33. P. S. Kessler, C. Daniel, J. A. Leigh, Ammonia switch-off of nitrogen fixation in the methanogenic archaeon *Methanococcus maripaludis*: mechanistic features and requirement for the novel GlnB homologues, NifI(1) and NifI(2). *J Bacteriol* **183**, 882-889 (2001).
34. Y. Ding, A. Berezuk, C. M. Khursigara, K. F. Jarrell, Phylogenetic distribution of the euryarchaeal archaellum regulator EarA and complementation of a *Methanococcus maripaludis* ΔearA mutant with heterologous earA homologues. *Microbiology* **163**, 804-815 (2017).
35. Y. Ding *et al.*, Identification of genes involved in the biosynthesis of the third and fourth sugars of the *Methanococcus maripaludis* archaellin N-linked tetrasaccharide. *J Bacteriol* **195**, 4094-4104 (2013).
36. D. J. VanDyke *et al.*, Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*. *Mol Microbiol* **72**, 633-644 (2009).

37. B. Chaban *et al.*, Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol Microbiol* **66**, 596-609 (2007).
38. S. Demolli *et al.*, Development of β-lactamase as a tool for monitoring conditional gene expression by a tetracycline-riboswitch in *Methanosarcina acetivorans*. *Archaea* **2014**, 725610 (2014).
39. A. M. Guss, M. Rother, J. K. Zhang, G. Kulkarni, W. W. Metcalf, New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for *Methanosarcina* species. *Archaea* **2**, 193-203 (2008).
40. I. Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *Bmc Bioinformatics* **8**, (2007).
41. K. S. Makarova, E. V. Koonin, Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* **1311**, 47-75 (2015).
42. J. D. Sander, J. K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355 (2014).
43. H. Wang, M. La Russa, L. S. Qi, CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem* **85**, 227-264 (2016).
44. A. E. Stachler, A. Marchfelder, Gene repression in haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas I-B system. *J Biol Chem* **291**, 15226-15242 (2016).
45. Z. Zebec, I. A. Zink, M. Kerou, C. Schleper, Efficient CRISPR-mediated post-transcriptional gene silencing in a hyperthermophilic archaeon using multiplexed crRNA expression. *G3* **6**, 3161-3168 (2016).
46. L. S. Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).
47. D. Burstein *et al.*, New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237-241 (2017).
48. D. D. Nayak, W. W. Metcalf, Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc Natl Acad Sci U S A* **114**, 2976-2981 (2017).
49. K. Selle, R. Barrangou, Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol* **23**, 225-232 (2015).
50. D. B. Nair, K. F. Jarrell, Pilin processing follows a different temporal route than that of archaellins in *Methanococcus maripaludis*. *Life* **5**, 85-101 (2015).

51. The UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169 (2017).
52. A. Grote *et al.*, JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* **33**, W526-531 (2005).
53. J. X. Chin, B. K. Chung, D. Y. Lee, Codon optimization online (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* **30**, 2210-2212 (2014).
54. Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**, 292 (2000).
55. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, GenBank. *Nucleic Acids Res* **35**, D21-D25 (2007).
56. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
57. J. G. Doench *et al.*, Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262-1267 (2014).
58. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).
59. T. J. Lie *et al.*, Overlapping repressor binding sites regulate expression of the *Methanococcus maripaludis* glnK(1) operon. *Mol Microbiol* **75**, 755-762 (2010).
60. D. B. Nair *et al.*, Identification of an additional minor pilin essential for piliation in the archaeon *Methanococcus maripaludis*. *PLoS One* **8**, e83961 (2013).
61. E. Daniel *et al.*, ATGme: open-source web application for rare codon identification and custom DNA sequence optimization. *BMC Bioinformatics* **16**, 303 (2015).
62. P. Puigbò, E. Guzmán, A. Romeu, S. Garcia-Vallvé, OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res* **35**, W126-131 (2007).
63. T. E. Quax, N. J. Claassens, D. Söll, J. van der Oost, Codon bias as a means to fine-tune gene expression. *Mol Cell* **59**, 149-161 (2015).
64. N. Gould, O. Hendy, D. Papamichail, Computational tools and algorithms for designing customized synthetic genes. *Front Bioeng Biotechnol* **2**, 41 (2014).
65. P. M. Sharp, W. H. Li, The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281-1295 (1987).

66. A. Carbone, A. Zinovyev, F. Képès, Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* **19**, 2005-2015 (2003).
67. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308 (2013).
68. J. G. Doench *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
69. C. M. Lee, T. J. Cradick, E. J. Fine, G. Bao, Nuclease target site selection for maximizing on-target activity and minimizing off-target effects in genome editing. *Mol Ther* **24**, 475-487 (2016).
70. D. B. Graham, D. E. Root, Resources for the design of CRISPR gene editing experiments. *Genome Biol* **16**, 260 (2015).
71. P. Ni, Q. Zhang, H. Chen, L. Chen, Inactivation of an integrated antibiotic resistance gene in mammalian cells to re-enable antibiotic selection. *Biotechniques* **56**, 198-201 (2014).
72. O. Shalem *et al.*, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).
73. Y. Fu *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826 (2013).
74. L. E. Jao, S. R. Wente, W. Chen, Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* **110**, 13904-13909 (2013).
75. T. O. Auer, K. Duroure, A. De Cian, J. P. Concordet, F. Del Bene, Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res* **24**, 142-153 (2014).
76. J. K. Rogers, G. M. Church, Multiplexed engineering in biology. *Trends in Biotechnology* **34**, 198-206 (2016).
77. G. J. Gibson, M. Yang, What rheumatologists need to know about CRISPR/Cas9. *Nat Rev Rheumatol* **13**, 205-216 (2017).
78. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
79. E. Weber, C. Engler, R. Gruetzner, S. Werner, S. Marillonnet, A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).
80. A. Casini, M. Storch, G. S. Baldwin, T. Ellis, Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol* **16**, 568-576 (2015).

81. N. J. Patron *et al.*, Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol* **208**, 13-19 (2015).
82. W. Szybalski, S. C. Kim, N. Hasan, A. J. Podhajska, Class-IIS restriction enzymes--a review. *Gene* **100**, 13-26 (1991).
83. C. Engler, R. Gruetzner, R. Kandzia, S. Marillonnet, Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
84. S. V. Iverson, T. L. Haddock, J. Beal, D. M. Densmore, CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth Biol* **5**, 99-103 (2016).
85. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* **4**, 975-986 (2015).
86. F. Fauser, S. Schiml, H. Puchta, Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J* **79**, 348-359 (2014).
87. L. G. Lowder *et al.*, A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol* **169**, 971-985 (2015).
88. R. E. Cobb, Y. Wang, H. Zhao, High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol* **4**, 723-728 (2015).
89. A. Villalobos, J. E. Ness, C. Gustafsson, J. Minshull, S. Govindarajan, Gene designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7**, 285 (2006).
90. A. H. Parret, H. Besir, R. Meijers, Critical reflections on synthetic gene design for recombinant protein expression. *Curr Opin Struct Biol* **38**, 155-162 (2016).
91. C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression. *Trends Biotechnol* **22**, 346-353 (2004).
92. E. Angov, Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* **6**, 650-659 (2011).
93. L. Bortesi, R. Fischer, The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* **33**, 41-52 (2015).
94. G. Song *et al.*, CRISPR/Cas9: A powerful tool for crop genome editing. *The Crop Journal* **4**, 75-82 (2016).

95. X. H. Zhang, L. Y. Tee, X. G. Wang, Q. S. Huang, S. H. Yang, Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
96. R. M. Gupta, K. Musunuru, Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**, 4154-4161 (2014).
97. G. Liang, H. Zhang, D. Lou, D. Yu, Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci Rep* **6**, 21451 (2016).
98. X. Wu, A. J. Kriz, P. A. Sharp, Target specificity of the CRISPR-Cas9 system. *Quant Biol* **2**, 59-70 (2014).
99. Y. Fu, D. Reyon, J. K. Joung, Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs. *Methods Enzymol* **546**, 21-45 (2014).
100. B. Shui, L. Hernandez Matias, Y. Guo, Y. Peng, The rise of CRISPR/Cas for genome editing in stem cells. *Stem Cells Int* **2016**, 8140168 (2016).
101. H. B. Lee, B. N. Sundberg, A. N. Sigafoos, K. J. Clark, Genome Engineering with TALE and CRISPR Systems in Neuroscience. *Front Genet* **7**, 47 (2016).
102. M. L. Luo, R. T. Leenay, C. L. Beisel, Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol Bioeng* **113**, 930-943 (2016).
103. J. K. Blackwood *et al.*, End-resection at DNA double-strand breaks in the three domains of life. *Biochem Soc Trans* **41**, 314-320 (2013).
104. E. J. Bartlett, N. C. Brissett, A. J. Doherty, Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proc Natl Acad Sci U S A* **110**, E1984-1991 (2013).
105. R. P. Parker, A. D. Walters, J. P. Chong, Bacterial and eukaryotic systems collide in the three Rs of *Methanococcus*. *Biochem Soc Trans* **39**, 111-115 (2011).
106. M. R. Lieber, The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**, 181-211 (2010).
107. A. D. Walters, The role of multiple minichromosome maintenance proteins in *Methanococcus maripaludis*, PhD Thesis, University of York, (2009).
108. S. S. Hoseini, M. G. Sauer, Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. *J Biol Eng* **9**, 2 (2015).
109. B. P. Anton, E. A. Raleigh, Complete genome sequence of NEB 5-alpha, a derivative of *Escherichia coli* K-12 DH5α. *Genome Announcements* **4**, e01245-01216 (2016).

110. M. S. Ramirez, M. E. Tolmasky, Aminoglycoside modifying enzymes. *Drug Resist Updat* **13**, 151-171 (2010).
111. T. J. Lie, J. A. Leigh, A novel repressor of nif and glnA expression in the methanogenic archaeon *Methanococcus maripaludis*. *Mol Microbiol* **47**, 235-246 (2003).
112. P. S. Kessler, J. A. Leigh, Genetics of nitrogen regulation in *Methanococcus maripaludis*. *Genetics* **152**, 1343-1351 (1999).
113. R. P. Novick, Plasmid incompatibility. *Microbiol Rev* **51**, 381-395 (1987).
114. S. Khelaifia, M. Drancourt, Susceptibility of archaea to antimicrobial agents: applications to clinical microbiology. *Clin Microbiol Infect* **18**, 841-848 (2012).
115. I. Porat, W. B. Whitman, Tryptophan auxotrophs were obtained by random transposon insertions in the *Methanococcus maripaludis* tryptophan operon. *FEMS Microbiol Lett* **297**, 250-254 (2009).
116. N. Goyal, Z. Zhou, I. A. Karimi, Metabolic processes of *Methanococcus maripaludis* and potential applications. *Microb Cell Fact* **15**, 107 (2016).
117. T. K. Wood, Metabolic manipulation of methanogens for methane machinations. *Microb Biotechnol* **10**, 9-10 (2017).
118. J. A. Noordover *et al.*, Containment in industrial biotechnology within wastewater treatment plants. *J Ind Microbiol Biotechnol* **28**, 65-69 (2002).
119. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
120. A. E. Briner *et al.*, Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339 (2014).
121. A. Lo, L. Qi, Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Res* **6**, (2017).
122. Y. Ding *et al.*, Identification of the first transcriptional activator of an archaellum operon in a euryarchaeon. *Mol Microbiol* **102**, 54-70 (2016).
123. R. B. Vercoe *et al.*, Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* **9**, e1003454 (2013).
124. D. Bikard, A. Hatoum-Aslan, D. Mucida, L. A. Marraffini, CRISPR interference can prevent natural transformation and virulence acquisition during *in vivo* bacterial infection. *Cell Host & Microbe* **12**, 177-186 (2012).
125. A. A. Gomaa *et al.*, Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio* **5**, e00928-00913 (2014).

126. K. P. Schoenfelder, D. T. Fox, The expanding implications of polyploidy. *J Cell Biol* **209**, 485-491 (2015).
127. J. Zischewski, R. Fischer, L. Bortesi, Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv* **35**, 95-104 (2017).
128. E. Purwantini *et al.*, Genetic resources for methane production from biomass described with the gene ontology. *Front Microbiol* **5**, 634 (2014).